Simultaneous testing of multiple subjects in ecologically valid assessments of hearing aids and assistive listening for speech in noise and music

SIMULTANEOUS TESTING OF MULTIPLE SUBJECTS IN ECOLOGICALLY VALID ASSESSMENTS OF HEARING AIDS AND ASSISTIVE LISTENING FOR SPEECH IN NOISE AND MUSIC

BY
LARISSA ANNE TAYLOR, MASc, (Electrical Engineering)
McMaster University, Hamilton, Canada, B.Eng., (Electrical and Biomedical Engineering)
McMaster University, Hamilton, Canada

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2022) — McMaster University
(Electrical & Computer Engineering) — Hamilton, Ontario, Canada

TITLE:	Simultaneous testing of multiple subjects in ecologically valid assessments of hearing aids and assistive listening for speech in noise and music
AUTHOR:	Larissa Anne Taylor MASc, (Electrical Engineering) McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Ian Bruce

I dedicate this work to my mother, Carol Taylor, for all her love and support.

Lay Abstract

While hearing aids are optimized for listening to speech, they still face challenges in noisy environments and when listening to music. The goal of the studies described in this thesis was to test hearing aids in realistic noisy environments as well as in a live concert setting. Our results testing hearing aids in different levels of background noise in normal hearing listeners showed effects on the effort subjects felt they required to listen, changes in heart rate, and looking behaviour as they responded to the additional demands of listening.

From studies in the LIVELab and the concert experiment, it is clear that judgements of music sound quality are highly subjective and varied between people. Overall the participants in the concert experiment were very satisfied with the hearing aids and assistive listening, but we did find that those with more experience with music were more critical and consistent in their ratings.

Abstract

Listening effort, or the amount of cognitive effort required to listen to a sound of interest, is an important measure of hearing performance, especially for hearing aid users. Hearing loss leads to increased listening effort in noisy situations and ideally hearing aid processing would reduce this effort. The goal of the two listening effort studies were to collect multiple measures of listening effort in an ecologically valid scenario, testing the effects of background noise, reverberation, and hearing aid directional processing on listening effort and head movement. To avoid the variability introduced due to age and varying degrees of hearing loss, for this initial study young normal hearing listeners were used. Two types of directional hearing aid processing were compared to the unaided condition.

Our results show an effect of background noise level and reverberation on subjective listening effort, an effect on physiological listening effort, as well as a right ear bias for head direction in increased background noise and reverberation. Hearing aid type showed a significant effect on deviation angle from the speaker on stage, that is the difference between where the subject was looking and the location of the actor speaking on stage. There was also a pattern of speech intelligibility changes with changing signal-to-noise ratio, which was different based on the type of hearing aid directional processing.

In addition to listening effort and speech intelligibility, music sound quality can be greatly affected by hearing aid processing. Live music has additional challenges compared to recorded music, so ecologically valid studies during live performances are essential to fully characterize sound quality. Preliminary studies in the LIVELab and an experiment conducted during an orchestra concert showed that while music sound quality judgments are subjective and variable between subjects, those with high musical sophistication are more critical and consistent in their judgments.

Acknowledgements

I would like to thank my supervisor, Dr. Ian Bruce, for his support and guidance throughout this project. I would also like to thank the staff in the LIVELab, especially Dan Bosnyak, Laurel Trainor, and Dave Thompson, for their help and advice on all experiments done in the LIVELab. Thanks to Steve Armstrong for his collaboration on the telecoil loop project. Support and funding from Unitron and Phonak made the listening effort studies and HPO concert experiment possible. Thanks to Brendan Tao, Daniels Shields, and Michael Bagnowski for their work on the physiological and motion capture data analysis and Ben Cinq Mars for his work on the HPO project. Thanks to the team at the CI Hackathon for organizing this challenge. And thanks to my supervisory committee, Henry Luo, Michael Noseworthy and Aleksandar Jeremic, for their time and feedback on these projects.

Notation and Abbreviations

CI - Cochlear Implant

dB SPL - Decibel Sound Presentation Level

dB HL - Decibels of Hearing Loss

ECG - Electrocardiography

EEG - Electroencephalography

H3fa - Three frequency average hearing loss

HA - Hearing Aid

HASQI - Hearing aid speech quality index

HF HRV - High frequency heart rate variability

HI - Hearing Impaired

HL - Hearing Loss

HPO - Hamilton Philharmonic Orchestra

HRV - Heart rate variability

KEMAR - Knowles' Electronics Manikin for Acoustic Research

LIVELab - Large Interactive Virtual Environment Lab at McMaster University `http://livelab.mcmaster.ca/`

LE - Listening effort

OLE - Objective Listening Effort

MUSHRA - MUlti stimulus test with hidden reference and anchor

NR - Noise Reduction

RMSSD - Root mean square successive difference

SNR - Signal to noise ratio

SPL - Sound pressure level

STI - Speech transmission index

Declaration of Academic Acheivement

Larissa Taylor designed and performed the listening effort studies described in Chapters 3–4 in collaboration with Ian Bruce and Henry Luo, with assistance from Unitron audiologists in setting up and fitting the hearing aids on participants. Data analysis for the listening effort experiments was done by Larissa Taylor with the assistance of undergraduate research students, with feedback from Ian Bruce and Henry Luo.

Setup of the telecoil loop system for HPO and LIVELab experiments in Chapters 5–6 was done by Larissa Taylor, Dan Bosnyak, and Steve Armstrong. The LIVELab concert experiment was designed by Larissa Taylor with feedback from Dan Bosnyak, Laurel Trainor, and Ian Bruce. The HPO concert experiment was designed by Larissa Taylor with feedback from Laurel Trainor, Dan Bosnyak, and Ian Bruce. Data analysis of the data from LIVELab and HPO concerts was done by Larissa Taylor. The HPO follow up study telecoil setup was done in collaboration with Dan Bosnyak and Steve Armstrong. The follow up listening experiment was designed by Larissa Taylor with feedback from Laurel Trainor and Ian Bruce.

Statement Regarding the Effect of COVID Restrictions

Follow up studies for both listening effort and HPO concert study were planned as part of this thesis work. COVID restrictions have prevented either of these follow up studies being performed at this time.

Contents

List of Figures

Chapter 1

Introduction

Hearing loss is a common issue, especially with an aging population. According to a Statistics Canada report in 2015, 19.2% of those age 20–79 had measured hearing loss in at least one ear (four frequency pure tone average threshold of 25 dB or worse) (Feder et al., 2015). This percentage jumps to 65% for those 70–79 (Feder et al., 2015). Hearing clearly, especially in situations with challenging background noise, becomes much more difficult with hearing loss. Hearing aids (HAs) have been designed to help, and have come a long way since simple analog amplifiers, but they do have remaining limitations and challenges in fully restoring hearing.

1.1 HA processing

In order to make sounds audible for those with hearing loss, hearing aids have several algorithms designed to provide gain but also other algorithms for more comfortable listening. They typically include compression to avoid sounds being amplified too much, noise reduction and along with it directionality, and multiple program settings for different listening scenarios such as quiet versus background noise, and for music. The microphone, analog to digital conversion, and other receiver inputs such as telecoil loops, as well as the digital signal processing all contribute to the overall sound quality of a HA and each have their own design decisions Dillon (2012). Consideration for design limitations such as low power use and minimal physical space must go into the HA design as well.

A great deal of progress has been made in improving noise reduction in hearing aids using sophisticated algorithms that predict background noise levels, and predict the speakers location to steer a beamformer created using the multiple HA mics. In fact some hearing aids have been shown in studies to improve speech perception in noise by reducing listening effort even in normal hearing listeners (van den Tillaart-Haverkate et al., 2017). User satisfaction with HAs has improved over time, especially as digital technology was incorporated as a standard (Kochkin, 2000, 2005, 2010; Abrams and Kihm, 2015) but it is still not 100%. A common reason selected for not using HAs in the series of MarkeTrak surveys was technology not working well enough in background noise, or poor sound quality (Kochkin, 2007; Abrams and Kihm, 2015).

Current hearing aids do perform well for speech in noise or other speech perception tests in controlled lab environments, however this performance does not always translate to more complex listening tasks or in real life scenarios. Another issue with hearing loss and HA benefit is not just making speech intelligible, but reducing the fatigue that some hearing impaired (HI) listeners have when listening in noisy environments (Hornsby, 2013). This has lead to an increased interest in not just speech intelligiblity but also other measures of hearing performance that could better test the effectiveness of algorithms at helping the end HA user, such as subjective and objective measures of listening effort.

1.2 Measures of Listening Effort

The review article by Pichora-Fuller et al. (2016) describes the best current understanding of what listening effort (LE) is, and descriptions of different metrics than can be used to measure it in listening studies, including subjective ratings and several objective measures. The definition of listening effort is the cognitive effort required to listen to a stimulus of interest (most often speech) at a given time. The idea behind measuring listening effort is that if a task requires greater listening effort, like any other effort this uses more cognitive resources and can lead to fatigue. Listening effort has become of great interest for those with HI and for HA processing due to the increased fatigue HI tend to report when listening compared to normal hearing individuals

Literature on listening effort uses many different objective measures. Secondary task performance is used commonly in literature, but since the goal of the LIVELab studies was to measure listening effort in as ecologically valid a way as possible, it was not used in the studies discussed here. Physiological measures of listening effort in literature are also varied, and include heart rate variability (HRV) using two different metrics, and alpha band power and instantaneous phase measures from EEG. These measures are based on the body's response to the demands and required resources for listening. More details on LE measures and their use in studies can be found in Chapter 2.1.

1.3 Speech sound quality and intelligibility metrics

Speech intelligibility is dependent on signal to noise ratio (SNR), but not linearly. There are also other types of distortions that affect speech intelligibility. In situations where lack of a clean source signal makes calculating SNR, a predictive measure of speech signal quality is the Speech Transmission Index (STI). This was developed by Houtgast and Steeneken in 1971 (Houtgast and Steeneken, 1971; Houtgast et al., 2002). It uses speech shaped noise as a test signal and the metric is calculated based on effective signal to noise ratios and other modulation factors in several frequency bands. The modulation of the test signal and frequency band weightings are designed to best capture changes to speech signals that would affect intelligibility.

Another prediction of speech quality is the hearing-aid speech quality index (HASQI) (Kates and Arehart, 2010, 2014). This metric has been designed specifically for sound processed through HAs, and is based on a cochlear model that also includes hearing impairment. This allows the metric to capture the nonlinear aspects of hearing.

1.4 Music and HAs

The benefits of improved speech understanding with HAs is essential for those with hearing impairment, but speech is not the only thing that we listen to in everyday life. Music is something that many people enjoy listening to or participating in, and the benefits of listening to and participating in musical activities can be seen in survey results in studies discussed in Coffman (2002); Črnčec et al. (2006). There is unfortunately also the possibility that participating in music for long periods of time could lead to increased risk of noise induced HI (Jansen et al., 2008), so professional musicians may find themselves in need of HA and would obviously want them to have good sound quality for music. While speech has been the focus of the majority of HA studies and algorithm development, there are also studies on the affect of HI and HA processing on music perception and sound quality.

There are several aspects of music that make it different from speech, and make amplification to within a comfortable and audible range with minimal distortion quite difficult. The dynamic range, frequency range, and temporal properties of music such as varying attack times between instruments, sustained and vibrato notes, all create unique challenges for hearing aid processing (Chasin and Hockley, 2014). The dynamic range of music is much greater than normal speech. A solo performance could be as quiet as 20-30 dB SPL, around the internal noise level of most hearing aid microphones (Zakis, 2016), or peak levels as loud as 120 dB SPL for an orchestral performance, which is above the input range of 16 bit analog to digital converters commonly used in hearing aids (Hockley et al., 2012). There has been progress in using larger numbers of bits to represent audio and represent these louder signals without clipping or distortion, or alternatively by shifting the input range of a 16 bit analog to digital converter up enough to be able to represent most levels that would be present in music.

The frequency range of music is also an important consideration. Speech has most of its important frequency and harmonic information in mid range frequencies (250 - 6000 Hz), whereas musical instruments have a much broader range of fundamental frequencies they are capable of producing as well as complex harmonic structures that give each instrument its characteristic timbre or sound. HAs have addressed this frequency range issue by increasing the frequency range from 4 kHz up to 10 kHz, although typically only minimal gain is possible at the highest frequencies to avoid issues with microphone feedback. Another difference between music and speech is that since speech is produced by the vibrations of a system that is relatively similar across all people, there is a defined long term average spectrum of speech that can be used to optimize any HA processing. Music is so varied in terms of styles of music, and also characteristics of different types of instruments from wind instruments such as the clarinet to sting instruments like violins that lead to different harmonic structures in the musical tones they produce, there is no well defined "average" spectrum for music (Chasin and Russo, 2004).

Any algorithms on the hearing aid that manipulate the signal in a nonlinear fashion could distort the frequency or timing information in music. This includes compression, or providing gain based on the level of the sound in a given frequency band as well as the required gain based on hearing loss (HL). Preferences for older linear amplification prescriptions over compressive prescriptions have been shown across different styles of music in Kirchberger and Russo (2016). This is a typical feature of hearing aids to ensure comfortable listening for the wearer. Limiting the gain on loud sounds serves to prevent any hearing damage, but also ensures it does not reach an uncomfortable level for the wearer. The uncomfortable sound level for someone with hearing loss is not increased with their increased sound threshold, so it is important to take this into consideration when providing gain with a HA. However, this compression will limit the dynamic range of music which spans from very low to very high sound levels, and also distort frequency bands relative to each other if this compression is done by individual frequency bands.

Another challenge for hearing impaired individuals listening to music is that they potentially have a decreased frequency and temporal resolution due to hearing loss (Chasin and Hockley, 2014; Oxenham, 2008), so listening to music is more challenging even with amplification. This makes accurate representation of all musical elements in the processed audio even more important to allow for better music perception. Improved audibility and sound clarity are needed to be able to separate music streams,

Figure 1.1: The LIVELab main auditorium, with KEMAR seated in the audience

hear melody/harmony separately and identify instruments which are all key elements of music.

Listening to music with hearing aids is still a challenge that may limit hearing aid use or adoption for some people. Solutions have been suggested in the past, as in Chasin and Hockley (2014), but many of them are less than ideal. For example, one suggestion is to simply remove the HAs. This does work for most, since music is typically loud enough that the HA would be providing minimal gain anyways and the persons natural hearing may provide less distorted sound. However this means that the HA would not be providing amplification when it was needed for speech, say in a music rehearsal scenario or at the intermission of a concert when speaking to others. Ideally, the HAs themselves would give high sound quality for music without requiring sacrificing quality or audibility of other sounds.

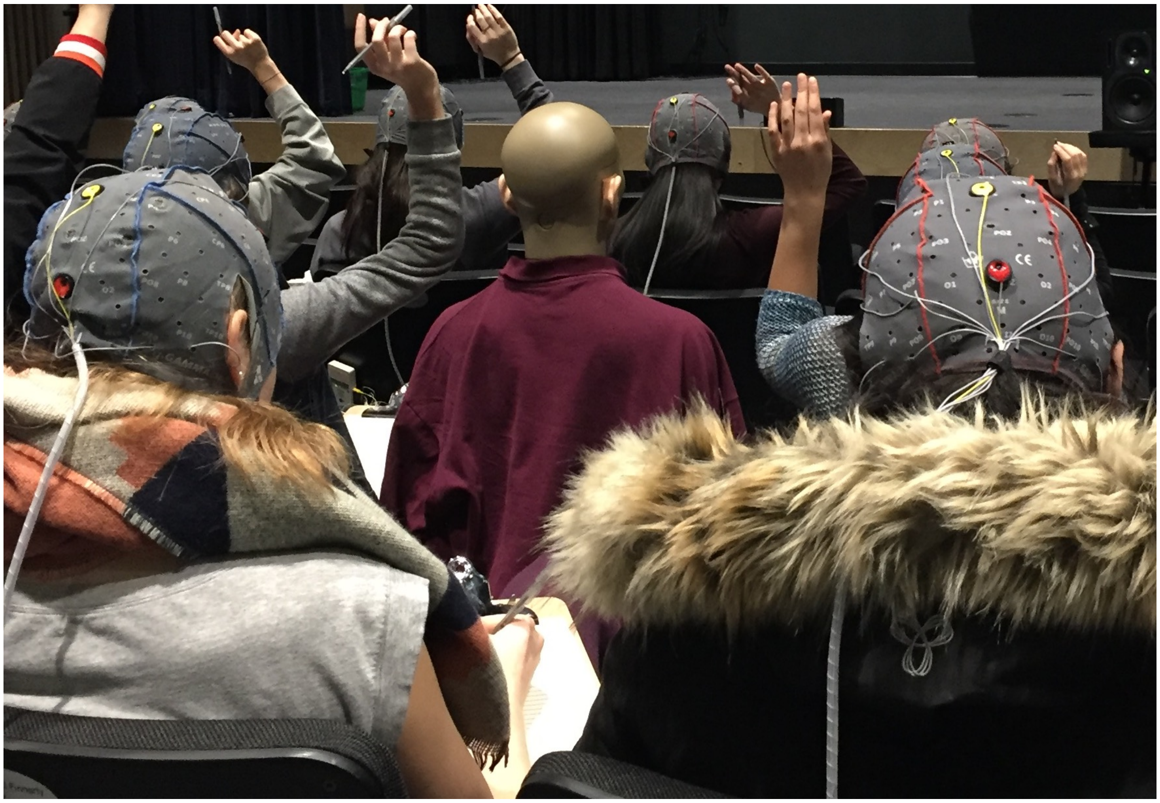


Figure 1.2: EEG caps as worn during the listening effort experiment described in Chapter 4

1.5 McMaster LIVELab

The LIVELab is a performance venue designed for large scale experiments at McMaster University. The main auditorium (shown in Fig. 1.1) seats just over 100 audience members and is equipped with measurement systems for electrophysiology such as ECG and EEG, tablets for continuous audience responses, a motion capture system, a live active acoustic system, and high quality microphones as well as a KEMAR manikin for audio recordings. It is possible to coordinate and use many of these systems at once for both data capture and creating experimental conditions (background noise, room acoustics etc), so it has made it possible to do the ecologically valid studies presented here in this thesis.

1.5.1 Physiology

The LIVELab has a g-Tec (Austria) g.USBamp system for measuring physiological signals. This can be used with ECG electrodes, and with low density EEG caps

Figure 1.3: Motion capture during the listening effort experiment described in Chapter 3

(shown in Fig. 1.2). Typical sampling rate is 512 Hz but this can be changed based on experimental requirements, and up to 32 subjects with EEG can be measured simultaneously in the main auditorium. Data is recorded on the control room computers as a .mat file that can be imported into matlab for analysis. Triggers from all LIVELab systems during the experiment can be recorded along with the physiological data for exact time matching of data and events during the experiment.

1.5.2 Motion capture

Motion capture can be done using the Qualysis Opus 5+ Motion Capture System with 28 IR cameras capturing both audience and performers on stage. Audience and performers can wear caps with four motion capture markers on a stationary cap to allow for calculation of head direction, as shown in Fig. 1.3. Individual motion capture dots can also be attached on the body for more customized motion capture setups. Labeling of each person and exporting the data is done using Qualysis software, and location and angle data can then be exported for further analysis done in Matlab

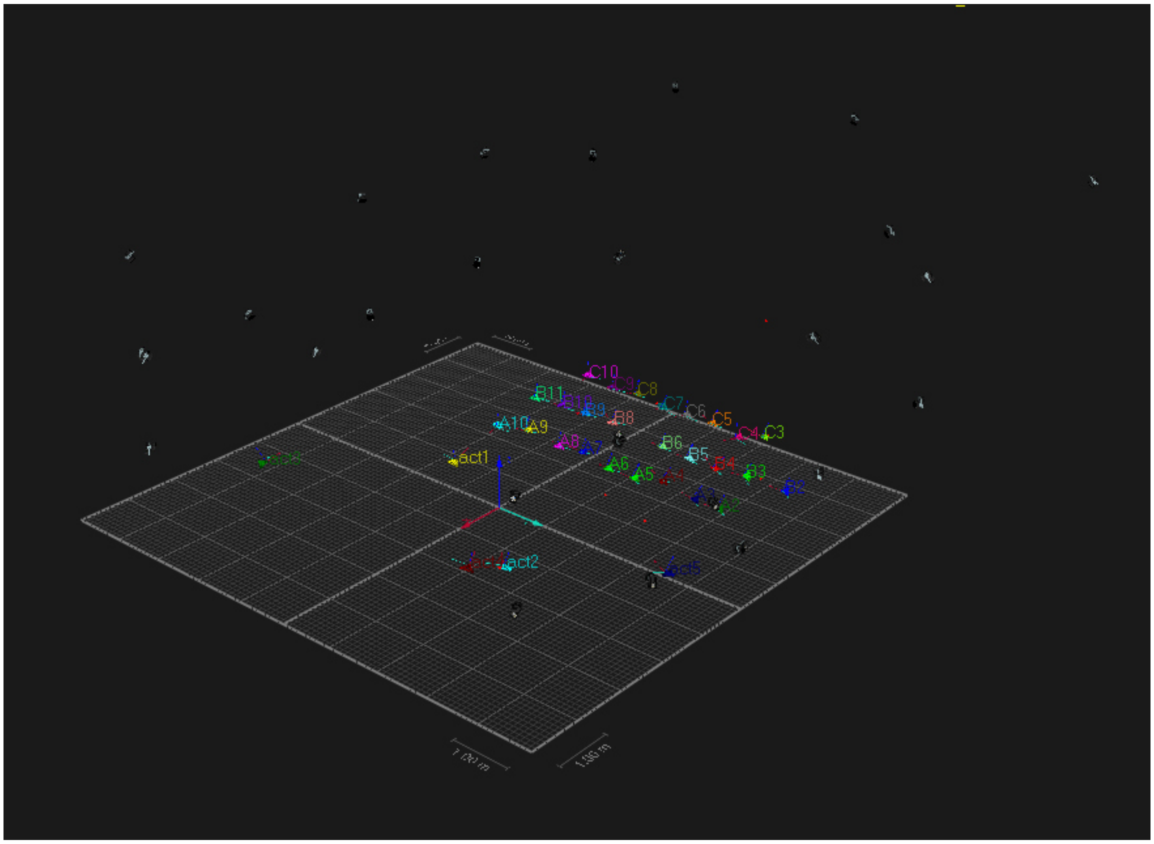


Figure 1.4: Example screenshot of the motion capture data from Qualysis software.

or other software. Motion capture along with video recording can be used to match motion capture with events during the experiment, such as which person is speaking on stage as used during the listening effort experiment performed with actors on stage as discussed in Chapter 3.

1.5.3 Meyer Constellation Active acoustic system

The LIVELab has a Meyer Constellation Active Acoustic system with 28 microphones and 75 speakers and subwoofers mounted on the wall and ceiling to simulate reverberation and also play sounds in the room such as background noise from all directions. The Variable Room Acoustic System (VRAS) algorithm (Poletti, 1999, 2000) is used to create the reverberation. This allows for the controlled simulation of different listening environments by changing the room reverberation and adding background noise from all directions, ideal for hearing aid testing.

Figure 1.5: The LIVELab KEMAR manikin, shown wearing a BTE HA.

1.5.4 KEMAR

The KEMAR manikin is a tool for audio research made by GRAS Sound & Vibrations (Denmark) (GRAS Sound & Vibrations, 2022). It is a manikin based on the average adult male head and torso with microphones in the ears that allow for recording environmental audio or audio while wearing hearing aids as it would be heard in the human ear. With recording and calibration, the KEMAR recordings can also be used for experiments using recordings with HAs, or to record reference data for later analysis and playback from an experiment with human participants.

1.6 Telecoil loops

Telecoil loops or T-coils are systems that can stream single channel audio to any hearing aids with enabled telecoil receivers through a electromagnetic field within the loop. The development of this technology is credited to Sam Lybarger in 1947 (Yanz and Preves, 2003; Levitt, 2007). They are commonly used in public venues and places of worship as assistive listening systems. This technology has been around for many years in hearing aids and was also used for hearing aid users to listen to phones with magnetic receivers. The basic components of the loop are a powerful amplifier, which needs to be able to drive significant current through a very low resistance load, the wire loop or loops and potentially a phase delay circuit to ensure accurate relative phase delays. An overlapping loop system can be used to reduce the effect of any metal or other interference in the magnetic field of the loop, and in this case a phase delay circuit is required to ensure that the loops have an appropriate relative phase delay and remain perfectly out of phase with each other. The telecoil loop field is then received by a telecoil receiver on the hearing aids, which can either be the sole input to the HA or be mixed with HA microphone signal depending on the program setup for telecoil loops on the HA itself. With amplification to account for passive inductance being positively correlated with frequency, the transfer function of the HA for telecoil loops and the HA microphones are not significantly different when averaged across all frequencies (Putterman and Valente, 2012; Yanz and Preves, 2003). In their comparison of HA telecoil receivers and HA microphones, Putterman and Valente (2012) did find differences at low and high frequencies. At 200 and 400 Hz, the telecoil signal was less but at 4000, 5000, and 6300 Hz the telecoil signal was greater than the microphone signal. There are both advantages and disadvantages to telecoil loops compared to other pairing or streaming technology, such as Bluetooth. Since the telecoil loop uses an electromagnetic field to send a signal to the HA there is no acoustic noise when listening purely to the loop, so in a noisy room this could be a great advantage. This explains why conference rooms, classrooms, and places of worship were some of the top venues where telecoil loops are used (Kochkin et al., 2014). In their survey on the perception of telecoil loops in venues, Kochkin et al. (2014) found that the majority of respondents agreed that telecoil loops increased their satisfaction with their HA or cochlear implant (CI), and the majority also indicated that the telecoil loops improved sound quality, speech intelligibility, concentration, and reduced background noise. Most of the complaints their respondents had were in regards to improper microphone use or sound mixing rather than the presence of the loop itself. One concern for using telecoil loops in venues such as classrooms where another loop may be located nearby is overspill from the electromagnetic field of the loop (Levitt, 2007). In addition to other loops, sources of electromagnetic noise are an issue for telecoil loop receivers on the HA. This includes computers, as well as motorized equipment (Yanz and Preves, 2003) which could be a challenge in an office environment with many such sources of electromagnetic interference. Orientation of the HA telecoil receiver relative to the telecoil loop field is an essential part of receiving a strong signal (Yanz and Preves, 2003), so loops must be setup with the orientation of the users in mind. Finally, while telecoil receivers on HAs were originally developed for use with phones, cellphones that use more modern piezoelectric transducers do not create the electromagnetic field required. This is why Bluetooth has become the more commonly used pairing technology for HAs and phones. Another difference between telecoil loops and Bluetooth is that since the signal transmission for telecoil loops is done using a changing magnetic field, the signal to both HAs will be the same. This means that only mono audio is possible. In contrast, with bluetooth it is possible for different audio signals to be sent to each HA so stereo audio is possible.

1.7 Summary of thesis chapters

A review of relevant literature is given in Chapter 2.

Chapters 3–4 describe two listening effort studies. Chapter 3 is soon to be submitted as a manuscript to Frontiers in Psychology.

Chapter 5 describes a series of preliminary studies on HA processing and assistive listening technology for live music performed in the LIVELab. Chapter 6 describes a HA and assistive listening experiment during a Hamilton Philharmonic orchestra concert.

Appendix A on the Cochlear Implant Hackathon was included as an appendix since it was a project Larissa Taylor took the lead role on though it was not directly related to the main research projects described in the main thesis chapters.

Chapter 2

Literature review

2.1 Listening effort

Listening effort has been described as the amount of effort or cognitive resources required for a listening task (Pichora-Fuller et al., 2016). Findings from both subjective and objective measures have found that older adults require more cognitive resources than younger adults to understand speech in noise. This is also true of those with hearing loss (Hornsby, 2013). The amount of resources a person requires for a given listening task depends on several factors related to hearing. These include cognitive factors such as working memory, which is related to speech comprehension, and speed of processing (Pichora-Fuller et al., 2016; Peelle, 2018). Attention to a task also determines the allocation of resources and thus the amount of effort being used for a task. Available context and the quality of sensory input both play a part in listening effort. Those with hearing loss experience lower quality sensory input, and older adults typically experience both sensory and cognitive decline (Pichora-Fuller et al., 2016; Tremblay and Backer, 2016). Age related structural changes in the brain are thought be a disadvantage when trying to selectively attend to a speaker in competing noise, leading to increased effort even without hearing loss (Tremblay and Backer, 2016). Reduced processing speed and memory with age also reduce the available context when listening (Tremblay and Backer, 2016; Lemke and Besser, 2016). The combination of these factors mean that both aging and hearing loss lead to increased listening effort compared to younger or normal hearing listeners.

Many studies have used listening effort to measure listening performance in different conditions, as well as the effectiveness of different HA processing algorithms. Listening effort can be measured using subjective ratings, but several objective measures have been used in addition to ratings.

2.1.1 Subjective listening effort

Results from subjective and dual task measures of listening effort in Hornsby (2013) suggest that sustained speech processing can lead to mental fatigue in those with hearing loss. Their study found no additional benefit from directional processing in hearing aids. They also found that subjective and objective measures of listening effort were not strongly associated, speculating that the different measures may assess different aspects of listening effort and fatigue.

Listening effort increases with the addition of background noise or reverberation have been shown even in normal hearing listeners (Visentin and Prodi, 2017; Rennies et al., 2014). Rennies et al. (2014) also looked at the combined effect of noise and reverberation by looking at the speech transmission index of their stimulus. Subjective listening effort increased with decreasing speech transmission index in their study, whether it was due to noise or reverberation. They did find that some material was more robust to the effect of reverberation, so reverb time alone was not a true indication of listening effort. This robustness of certain material to reverberation may have been a factor in the results of Picou et al. (2019), who found that background noise but not reverberation increased behavioural and subjective listening effort in normal hearing children. Sato et al. (2012) also found that subjective listening difficulty was correlated with speech transmission index, and that listening difficulty could distinguish between sound fields where speech intelligibility could not. Visentin and Prodi (2017) found that the shape of the added background noise also had an effect on the perceived effort. They found that fluctuating noise was more effortful compared to steady state speech shaped noise. Other factors that may not affect comprehension have also been shown to increase listening difficulty, such as voice hoarseness for children listening in babble background noise (Rudner et al., 2018). All of these studies show that a non ideal listening environment can have negative effects on listening effort for normal hearing individuals.

The benefits of various hearing aid noise reduction algorithms have also been studied in both normal hearing and hearing impaired listeners. What van den Tillaart-Haverkate et al. (2017) showed in their study is that both ideal and realistic noise reduction can reduce response time in normal hearing listeners, a behavioural measure of listening effort, even at SNRs where speech intelligibility was close to 100%. A multi site European study by Luts et al. (2010) in both normal hearing and hearing impaired listeners showed differences in listening effort ratings with five different noise reduction algorithms with no difference in speech reception threshold. They also showed a preference for most of the noise reduction algorithms over unprocessed sound, especially at SNRs greater then 0 dB. Brons et al. (2014) similarly found that none of their tested noise reduction algorithms improved speech intelligibility, but they all reduced noise annoyance. They pointed out a potential trade off between intelligibility and listening comfort as noise reduction algorithms attenuate signal assumed to be noise. Another tradeoff was found in Brons et al. (2013) between speech naturalness and noise annoyance. Both factors were determining factors in overall preference for noise reduction algorithms, but the weighting differed between people even in their normal hearing group.

Objective and subjective measures of listening effort are not always found to be strongly associated (Hornsby, 2013) and have been suggested as measuring different aspects of listening effort. An example of this was found in Picou et al. (2017). They used a dual task paradigm along with subjective ratings to test speech recognition and effort in background noise using HI adults. They found that while subjective ratings of work and tiredness were more closely related to word recognition performance than objective listening effort, subjective ratings of control were more strongly related to objective listening effort than word recognition. Some studies with subjective ratings and objective measures of listening effort do find significant correlations between the two measures, such as in Holube et al. (2016). Holube et al. (2016) found a significant correlation between electrodermal activity and subjective ratings of effort in their normal hearing group. They also found a significant difference between the easy and hard listening condition for subjective ratings. There was a similar trend in the electrodermal activity difference between easy and hard listening conditions but it was not significant.

While studies often use more than one measure of listening effort, typically subjective ratings along with one objective measure, another method to look at the multi dimensional problem of listening effort and speech intelligibility suggested by Prodi et al. (2010) is to combine listening effort and speech intelligibility into a new measure, which they called listening efficiency. In this way you could capture the different aspects of effort from each measure but also the benefits of measuring both listening effort and speech intelligibility. A combined measure would capture the sensitivity of listening effort to changes where speech intelligibility remains high, and the changes in speech intelligibility where effort is high but intelligibility may be improving. There are also physiological based listening effort measures that can be used in place of task performance and response times.

2.1.2 Physiological listening effort measures

Physiological measures of listening effort have been developed with the assumption that they are perhaps less affected by motivation or other external factors like secondary task performance or response times could be. They are based on the body's response to stress and parasympathetic activity in the body. Measures of listening effort that are affected by the parasympathetic nervous system are heart rate variability and pupillometry. Peak pupil dilation has been associated with cognitive effort, and it results from the relaxation of the parasympathetic pupil dilation (Pichora-Fuller et al., 2016; Francis and Oliver, 2018). Peak pupil dilation has been shown to change with changing SNRs in listening effort studies (Ohlenforst et al., 2017, 2018). McGarrigle et al. (2017) showed a similar change in pupil dilation with increased SNR.

The EEG based metrics of listening effort are based on differences between passive versus effortful processing of speech in the brain, and the effect that has on the synchronization of EEG activity. EEG studies of speech perception along with fMRI studies have continued to try and determine the best location and frequency bands of interest for speech processing (Evans and McGettigan, 2017; Peelle, 2018; Bernarding et al., 2012).

Heart rate variability

Heart rate variability is used as a measure of listening effort based on it measuring activity in the parasympathetic activity (Mackersie and Calderon-Moultrie, 2016; Berntson et al., 2017). Berntson et al. (2017) describes the physiological basis within the parasympathetic nervous system activity as well as the calculation of root mean square successive difference (RMSSD) and high frequency heart rate variability (HF HRV), two different measures of HRV.

There have been differences in HRV shown in listening studies, but they are varied in whether decreased SNR or hearing loss lead to increased, decreased, or no change in HRV (Mackersie and Calderon-Moultrie, 2016; Mackersie et al., 2015; Cvijanović et al., 2017). Mackersie and Calderon-Moultrie (2016) increased task demand by increasing speaking rate as well as adding background noise. The background noise levels for normal and fast speech were set individually between -3 and 6 dB for each subject based on 80 % correct word repetition. The study found that HF HRV was reduced with speech at a fast rate compared to a normal speaking rate. They used a baseline period of HF HRV and log transformed HF HRV values. Mackersie et al. (2015) compared physiological stress for normal hearing adults and adults with hearing loss using HF HRV and found an increase in autonomic nervous system activity in those with hearing loss compared to normal hearing even when recognition was the same. Cvijanović et al. (2017) tested normal hearing participants in a 'skype' video conversation with noise and three different levels; -6, 0, and 6 dB SNR, and found no change in HRV across noise levels. They theorize that this is due to HRV not being sensitive enough in their test scenario.

EEG measures

The electrical activity of the brain is complex but there are measurable changes from baseline due to external stimulus that can used in studies, such as those on hearing and listening effort. There are two main types of changes that can occur in brain oscillations, a more transient change evoked by a single stimulus or shifts in the baseline frequencies of oscillation due to an external periodic stimulus (Thut et al., 2011). The magnitude and frequency of these changes can indicate the amount of response to a particular stimulus, and the state of current brain activity.

The left inferior frontal area is of particular interest for listening as it is the location of Broca's area, and the primary region of the cortex for speech processing. A review by Evans and McGettigan (2017) gives an thorough overview of the fMRI studies that have led to the focus on the left Inferior Frontal Gyrus (IFG) for measurement of effort or resource use during speech processing. Alain et al. (2018) similarly showed through meta analysis of over 50 fMRI studies that there was an increase in IFG activity for speech in noise as well as for more liguistically complicated speech. They also speculated that the prefrontal activity found in listening effort fMRI studies could reflect increased use of predictive processing to offset poor encoding of speech or poor working memory. Johnsrude and Rodd (2016) also describes increases in activity in the Inferior Frontal Gyrus when listening to speech with more ambiguous words and context, confirming again this regions importance in speech processing especially for more effortful listening. The same review by Johnsrude and Rodd (2016) also describes an fMRI study that shows changes in superior temporal suculus and frontal regions, both part of the cortex associated with higher auditory processing, were modulated by attending to speech being played compared to attending to a distractor or secondary task. While fMRI is a powerful tool for identifying areas of the brain activated by effortful listening, it is not practical in a more realistic scenario so more indirect methods of measuring brain activity are used, such as EEG.

Similar to ECG measures of parasympathetic activity, several EEG measures of listening effort have been used in studies. The validity of this measure has been shown by the fact that alpha activity does increase with active listening versus passive listening (Dimitrijevic et al., 2019). While Dimitrijevic et al. (2019) found that for speech alpha activity did increase with passive versus active listening for speech, in a previous study with digits in noise there was no difference in alpha event related desyncronization (Dimitrijevic et al., 2017), indicating that this alpha change may be specific for more complex speech stimulus processing. A study by Wisniewski (2017) with tones also indicates that increased power in EEG likely reflects increased utilization of cognitive control processes, which would be a measure of listening effort based on using increased resources for the task (Pichora-Fuller et al., 2016). Cartocci et al. (2019) also found less lateralization of EEG activity in the hearing impaired group in their study, indicating that additional areas are being used for speech processing in noise compared to the normal hearing group. They found that this difference was most characterized by alpha and theta activity in the inferior frontal areas (Cartocci et al., 2019). This again goes with previous findings that hearing impairment leads to increased listening effort, and also other studies that have found listening effort being best characterized by alpha activity in the inferior frontal area.

EEG alpha band power and instantaneous phase in particular have been used as a parasympathetic activity or resource use based measure of listening effort. The band power of EEG is a measure of the overall processing and top down inhibition, so an evoked response or reorganization of the ongoing neural activity (Strauss et al., 2010). Strauss et al. (2010) proposed the use modeling event related potentials to measure attention and listening effort, specifically the N1 response magnitude of the auditory late responses. Miles et al. (2017) measured EEG alpha power and pupilometry in normal hearing adults listening to degraded speech to determine if the measures would change with increased degradation of the speech, and if the measures would be correlated. The measures were not correlated, and SNR was also not a significant predictor of either listening effort measure. This could be due to individual differences, but also due to these physiological measures measuring different aspects of effort. They note that the EEG data was only analyzed for the parietal region, which from previous studies may not be the best representation of listening effort in EEG (Dimitrijevic et al., 2019). The change that Miles et al. (2017) found in alpha power for the different amount of degradation was opposite to the increased alpha power with increased listening effort seen in other studies. Although alpha band power has been most commonly used to measure listening effort from EEG, other frequency bands have also shown effects of effortful listening. Cartocci et al. (2019) showed significant differences in beta band lateralization in both C3–C4 and F7–F8 electrode pairs for speech in noise compared to speech in quiet, as well as showing significantly different asymmetry in the gamma band in the F7–F8 electrode pair.

Another measure of listening performance that has been used is coherence between EEG and the speech signal being listened to. This is a measure of speech intelligibility rather than listening effort, as it corresponds to how well the speech is being represented at the cortical level. This coherence measurement has been done previously for speech sources using both EEG and MEG data (Gross and Ioannides, 1999; Gross et al., 2001). Dimitrijevic et al. (2019) used the speech envelope coherence in the 2-5 Hz range to measure speech intelligibility, while they found that alpha power in the left inferior frontal cortex could predict subjective listening effort ratings. The theta band of EEG is most often used in coherence calculations as it aligns with the frequencies of the speech envelope (2-5 Hz). Their previous study with digits in noise also indicated that temporal alpha band event related desyncronization was related to correct digit identification, or speech intelligibility (Dimitrijevic et al., 2017). This is similar to findings that the frontal areas contained more differences in listening effort between normal hearing and hearing impaired groups in Cartocci et al. (2019).

Apart from using band powers, instantaneous phase of EEG is also used for measuring effort while listening to speech. Strauss et al. (2010) used the instantaneous phase of auditory late response during a detection of syllable in noise or musical tone detection task. They found that it did match listening effort model predictions, and could be a potential method for measuring listening effort. These are based on auditory responses to short stimuli though and not full segments of speech. Bernarding et al. (2012) described a proposed method of measuring LE from EEG data based on phase distribution from a wavelet transform. The wavelet used was a 6th derivative Gaussian wavelet, and they found that the frequency of wavelet most correlated with effort was 7.68 Hz, close to the alpha-theta band border. They did an initial feasibility study in normal hearing undergraduate students (Bernarding et al., 2012) followed by studies with experienced HA users (Bernarding et al., 2014, 2017). They found in both experiments that the objective EEG measure mapped to subjective listening effort ratings, and in Bernarding et al. (2017) using a wavelet with a pseudo frequency on the alpha theta band border they found significant effects of HA setting on listening effort for both EEG and subjective ratings. They also found that the objective results were highly correlated with the subjective ratings, and also negatively correlated with speech intelligibility.

Schafer et al. (2015) also used instantaneous phase of EEG from a wavelet transform on the alpha theta border. Their study was with hearing impaired listeners, and they collected EEG and subjective ratings for different HA settings. They found that the EEG measures matched the subjective data well, and may be more sensitive to small variances in listening effort. They did find different results between HA settings for objective ratings versus EEG measures of listening effort. Mortezapouraghdam et al. (2017) used the instantaneous phase to predict perceived listening effort. Their prediction did improve when instantaneous phase was included compared to just using intelligibility scores to predict effort.

2.2 Music Perception

Digital hearing aids have had many advances in the last decade to increase audibility and comfort for listening to speech, but little has been done to tackle the problem of music. With the proven cognitive and social benefits of listening to and creating music, it is essential that we begin to understand how to improve the music listening experience for hearing aid (HA) users. Music has been shown to have benefits in both children, and in older adults (Črnčec et al., 2006; Coffman, 2002). Music performance provides a form of entertainment for older adults, but also provides other benefits such as the opportunity to form relationships with others, and an outlet for creative expression. If hearing aid users are discouraged from or unable to participate due to poor quality or perception of music through hearing aids, they cannot receive these potential benefits. The 2012-2013 Canadian Health Measures Survey indicated that 47% of those 60-79 years old had at least mild hearing loss (Feder et al., 2015), so the population of older adults that could benefit from hearing aids is potentially quite large. Another population that would benefit from improved hearing aid processing for music is musicians who develop noise induced hearing loss due to their profession. A study of university students at the University of North Carolina in 2010 found that 45% of the student musicians had noise induced hearing loss in at least one ear (Phillips et al., 2010). This is similar to the prevalence of hearing loss in industrial workers, and matches results from studies of older classical musicians. With continued exposure, some of these musicians may require hearing aids.

2.2.1 Music perception and hearing loss

An effect of hearing loss on aspects of music perception such as melodic interval perception, melodic contour perception, and instrument timbre, has been shown (Galvin et al., 2008; Luo et al., 2014), but much less work has been done on practical solutions for hearing aid users who want to listen to music using their hearing aids (Chasin and Hockley, 2014; Zakis, 2016). Solutions proposed by Chasin and Hockley (2014) include covering the hearing aid microphone with tape to dampen the sound, removing the hearing aids entirely, or switching to analog hearing aids to avoid the issues with the analog to digital converter. These solutions might solve the problem while listening to music, but are not entirely practical. The hearing aid user needs to enjoy the music, but also be able to listen to conversation happening between playing and possibly with a large amount of background noise. Other studies have looked at preferred amplification settings, bandwidth, and number of frequency channels for music, but did not determine the effects on rhythm, timbre, pitch, or melody perception, which are all basic elements of music (Zakis, 2016).

The ability to interpret the different parts of music separately is essential to the appreciation of music. The perception of differences in frequency, localisation cues, loudness, and temporal or spectral features all contribute to the ability to segregate streams of music (Innes-Brown et al., 2011). Unfortunately, hearing loss leads to the degradation of many of these cues, reducing the ability to segregate auditory streams. Increasing the audible bandwidth, or having a higher target sensation level was shown to give small improvements in spatial release from masking for speech (Jakien et al., 2017). In this study, they suggest that the audibility enhances access to the interaural level and timing difference cues that are necessary for separating competing talkers. No study was found describing the benefits of this for separating non speech signals, but since the high frequency harmonic information is also used for music stream separation, and this would also be increased with increased audibility, it would likely be beneficial.

2.2.2 HAs and music perception, sound quality

Though many hearing aid manufacturers offer a music program for their hearing aids, in a study by Madsen and Moore (2014) only 40% of hearing aid users reported having one on their hearing aid. This same study found that the music program did not lead to improved loudness, clarity, or tonal balance and only gave small improvements in the user's ability to detect individual instruments. It is generally understood that most hearing aid algorithms, such as directionality, frequency lowering, feedback reduction, and noise reduction, need to be adjusted or turned off for listening to music to avoid distorting the signal. However, there is little literature showing the benefit of doing so (Zakis, 2016). Compression of loud sounds by the hearing aid is another feature that may need to be turned off for music, to avoid distorting louder parts of the song, but a study by Arehart et al. (2011) found that a small amount of compression did not significantly affect the sound quality of music. A study was done that showed hearing aid users preferred Adaptive Dynamic Range Optimization (ADRO) over the more commonly used Wide Dynamic Range Compression, but it was not clear if this effect came from the number of channels used, slow time constants, ADRO processing, or some other factor (Zakis, 2016). Some possible improvements to hearing aid processing included in Zakis (2016) are a preamplifier that dynamically adjusts gain to increase the input range of the analog to digital converter, and higher sampling rates to more accurately represent the high frequency components of music. The input range of the hearing aid could also be increased by switching to a 20 or 24 bit analog to digital converter, but this is not commonly done in practice since it would require additional power. The 10 kHz frequency bandwidth of the hearing aid is also a potential issue, but improvements are limited by size constraints on the receiver (the miniature loudspeaker in the hearing aid). Hearing aid microphones also have a roll off in gain at frequencies below 1 kHz. Increased digital gain accommodates for this roll off, but could distort or suppress very soft music. Further study is clearly needed to determine what an ideal music program on a hearing aid consists of. Studies have been done on speech recognition in reverberant environments, and have found that older individuals and individuals with hearing impairment are more susceptible to reverberant distortions of speech (Srinivasan et al., 2016). It was also found that there was less spatial release from masking of speech with hearing impairment (Marrone et al., 2008), and age was found to be a significant factor, which could be due to age related hearing impairment. Orchestra music is often performed in large concert halls, so that the reverberation adds character to the music and makes it sound fuller. A study on the effect of reverberation time on music enjoyment with cochlear implants by Certo et al. (2015) found a preference for mid range reverberation time (measured as T60, or time for the signal to decay 60 dB) of 3 s in the original music samples, but a short T60 of 0.2 s was preferred in the cochlear implant processed samples. They did show a significant effect of reverberation time on music enjoyment, however they were using simulated music generated by notation software and simulated cochlear implant processing being played for normal hearing subjects, so the results may differ in hearing aid users. No study was found that examined this effect on music with hearing impairment or hearing aids, but it is an important factor to look at.

2.3 Sound quality measurements

Various sound quality studies have been done with HA processing algorithms, typically for speech. The majority of these studies are done with recordings, listened to over headphones, to allow for more direct comparisons of HA processing (Muralimanohar et al., 2013; Wu et al., 2019; Schoeffler et al., 2018; Cubick et al., 2014; Simonsen and Legarth, 2010). Common features between these speech quality studies are the use of a reference and hidden anchor based test, with multiple sentence stimuli to compare each of the processing conditions. This MUlti stimulus test with Hidden Reference and Anchor (MUSHRA) is a commonly used test for judgements of sound quality in audio systems, and has been adapted for use in audiology as it is also well suited to judgements of HA sound quality (Völker et al., 2016; Simonsen and Legarth, 2010). In Muralimanohar et al. (2013) there was a significant correlation between MUSHRA sound quality ratings and objective speech quality predictions using HASQI. The unprocessed reference signal was rated highest, and the anchor of speech with -5db SNR with babble noise was rated lowest. They also found a significant effect of processing on MUSHRA ratings. Wu et al. (2019) also found a significant correlation between subjective ratings in a MUSHRA test and speech scores. Significant effects of both HA processing strategy and interactions with test SNR were shown for both subjective ratings and speech scores. They also showed a difference in mean test SNR for different auditory profiles (HL severity and shape). The test SNRs were individually selected based on SRT_{50} speech reception threshold.

The choice of the reference and anchor in MUSHRA tests vary by study, for example if done with normal hearing listeners no HA can be used as the reference (Simonsen and Legarth, 2010). For the anchor, a signal with an artificially worse SNR can be used as in Wu et al. (2019), or a signal low pass filtered at 3.5 kHz to replicate older HA technology (Schoeffler et al., 2018), or in the case of Simonsen and Legarth (2010) actual older technology HAs. One of the benefits of the MUSHRA test is easy paired comparisons of all the processing conditions for multiple test stimuli.

There has even been a web based MUSHRA test developed for remote experiments (Schoeffler et al., 2018). A similar multi comparison approach like MUHSRA can also be adapted for other sound judgements, such as the perceived distance of a sound like in Cubick et al. (2014). In their study, Cubick et al. (2014) note that they found large inter subject variability in HI subjects, something important to consider with any HA or HI study. Simonsen and Legarth (2010) also found difference in HA technology preferences based on HL profile.

In addition to subjective assessments of speech quality, MUSHRA tests are also well suited to subjective measurements of music sound quality. D'Onofrio et al. (2019) did listening tests for both music and speech stimuli with musicians and non musicians. Through repeated trials, they found that musicians were more consistent trial to trial. This difference between musicians and non musicians is important for music sound quality judgements in particular, since it is so subjective. This is why surveys such as the Goldsmith survey of musical sophistication (Müllensiefen et al., 2014) have been developed to objectively separate those with musical experience from those without. One difference between music and speech MUSHRA tests is the length of the samples. While for speech, single sentences are typically used as the audio stimuli, for music a longer sample is needed. For example, in D'Onofrio et al. (2019) 45 s segments of music were used, and in Croghan et al. (2012) 13 s long segments of music were used. The length of the music segments is determined by musical phrasing, explaining any differences in sample length.

Along with needing to take into consideration musical experience, another complexity of music sound quality is that there are many measures of quality that can be taken into account for music. This is true in Croghan et al. (2012) where they discuss eight dimensions of sound quality in their study: clearness, sharpness, brightness, fullness, feeling of space, nearness, loudness, and disturbing sounds. They also asked about preference and pleasantness separately from quality. Croghan et al. (2012) was based on paired comparisons of two samples, rather than the MUSHRA test with all stimulus presented and rated together. Like Cubick et al. (2014), Croghan et al. (2012) found substantial individual variability across subjects. Across all the factors they tested (loudness, dynamic range, pleasantness, preference) low levels of compression had little effect, but compression greater then -12DbFS lead to decreases in all measures of quality. Another study on compression, Madsen et al. (2015), focused more on speed of compression. They also used paired comparisons, focusing on the clarity of the music in short 4–5 s excerpts. Their results showed that the linear condition was rated more clear than any of the compression conditions, and the subjects showed a preference for slow compression, though this was not a significant difference from fast compression.

Chapter 3

Comparing listening effort measures in multiple participants simultaneously during a live performance

3.1 Introduction

Listening effort studies with hearing aids show varied results and use a variety of listening effort measures, speech material, and listening test setup. There are numerous methods to measure listening effort and most studies select one or two as measures for their study. In order to better evaluate the effects of hearing aid algorithms on listening effort, our goal in this study was to design an experiment that is as ecologically valid as possible while controlling factors such as background noise level and reverberation. In this environment a listening task was performed to measure subjective listening effort as well as multiple objective measures for comparison. Normal hearing subjects were used to test the effects of background noise and reverberation on listening effort, the correlations between listening effort measures, and the feasibility of doing a future large scale listening effort study. Our plan is to refine and repeat this study with hearing impaired listeners based on our results with normal hearing listeners.

Subjective listening effort changes with different reverberation times and SNR have been demonstrated in both normal hearing subjects (Rennies et al., 2014; van den Tillaart-Haverkate et al., 2017; Rudner et al., 2018) and subjects with hearing loss alongside normal hearing subjects (Sato et al., 2012; Holube et al., 2016). Rennies et al. (2014) used subjective listening effort and different levels of Signal to Noise

Ratios (SNRs) and reverberation on German sentences to demonstrate how listening effort varied with speech transmission. Their data showed that listening effort increased with decreasing speech transmission. They also measured speech intelligibility and found that while intelligibility was more sensitive for SNRs below 2 dB, listening effort was a more sensitive measure at higher SNRs. van den Tillaart-Haverkate et al. (2017) examined the effect of noise reduction on speech intelligibility and listening effort at 4 different SNRs in normal hearing subjects. They found that noise reduction did reduce response time even at SNRs where speech intelligibility was close to 100%. Rudner et al. (2018) looked at the effect of background noise and a simulated hoarse voice on comprehension and perceived difficulty listening in normal hearing children. The babble background noise did have an effect on both comprehension and perceived difficulty listening though this was reduced by seeing the face of the talker. The hoarse voice did not reduce comprehension but did increase perceived difficulty listening. This study indicates that advantages that people use in a difficult listening scenario such as seeing the talker's face, and additional factors that might affect speech such as hoarseness should also be taken into account in listening studies. Sato et al. (2012) found that their older hearing impaired subjects rated listening difficulty lower than young normal hearing subjects, however for the same perceived listening difficulty the older subjects word intelligibility scores were below 100%, whereas the younger subjects achieved 100%. Subjective listening effort ratings are useful, but they are biased and can be influenced by something as simple as changing the wording of the question being asked of them (Picou and Ricketts, 2018). Picou and Ricketts (2018) found varying degrees of correlation between subjective responses and dual task listening effort measures for different questions related to listening effort. Picou et al. (2019) found that perception of time passed was correlated with dual task performance. The results in Picou et al. (2019) showed that for normal hearing children, background noise increased both subjective and behavioural listening effort but the addition of reverberation did not. Holube et al. (2016) examined the effect of background noise and reverberation in normal hearing and hearing impaired subjects. The difference in subjective listening effort rating was significant between an easy and a hard condition, and electrodermal response showed a similar but non significant trend. Holube et al. (2016) did find a significant correlation between subjective rating and electrodermal activity in the normal hearing subjects. Francis et al. (2016) compared listening effort for masked and unmasked speech. They found that the masked condition evoked stronger physiological (electrodermal) response than unmasked speech even when behavioural and subjective listening effort were comparable. This is where physiological listening effort measures may be more reliable, being based on physiological responses of the subjects rather than subjective opinions.

Electrocardiogram (ECG) measures of listening effort are one of several physiological measures discussed in Pichora-Fuller et al. (2016). One of two measures of heart rate variability (HRV) is typically used, both resulting from autonomic nervous system activity (Berntson et al., 2017). The measures of heart rate variability used in studies are root mean square successive difference (RMSSD), and high frequency heart rate variability (HF HRV). Both are based on the variation in intervals between heart beats and are further described in the methods section of this paper. Low frequency HRV is also sometimes used as a measure. The results of studies using HRV with both normal hearing and hearing impaired subjects vary, and a second measure of listening effort is typically used and compared to HRV. In Mackersie et al. (2015) they compared subjective and HRV listening effort measures in both normal hearing and hearing impaired listeners. They found that the hearing impaired group did show decreased HF HRV at lower SNRs, but the normal hearing group did not show the same decrease. They also found no significant effect of the changing SNR on subjective listening effort, and no significant correlation between the subjective rating and HRV measures. They predicted that more acoustically complex noise with spatial separation of sound sources could increase sensitivity to changes in SNR. In a following study by Mackersie and Calderon-Moultrie (2016) they once again compared normal hearing and hearing impaired listeners. In this study the measures used were HRV, skin conductance, and sentence repetition scores. They did find a decrease in HF HRV along with an increase in skin conductance, both indicative of increased listening effort, with increased task demand from increased speaking rate. The increase in skin conductance was greater for the hearing impaired group compared to the normal hearing group, despite similar sentence recognition scores. This is in agreement with the understanding that hearing impaired listeners experience greater listening effort than normal hearing listeners. Cvijanović et al. (2017) used a two person communication task to examine differences in responses for various noise conditions to establish the sensitivity of physiological measures. They used LF HRV and the LF/HF HRV ratio as their measures of listening effort. They found no significant effect of different noise levels on their listening effort measures.

One factor that may affect speech comprehension in realistic difficult listening scenarios is head movement (Brimijoin et al., 2010). Brimijoin et al. (2010) also found differences in head movement latency for hearing impaired listeners compared to normal hearing listeners. Brimijoin et al. (2010) found that hearing impaired listeners were slower to change head position to a new speaker location, however Ricketts and Galster (2008) found that children with hearing loss showed increased visual monitoring to sounds compared to normal hearing children. Previous studies have found that looking behaviour was a significant predictor of comprehension during a story listening task (Valente et al., 2012), and that a right ear bias has been observed for listening tasks Marzoli and Tommasi (2009). Grange and Culling (2016) demonstrated that normal hearing subjects will turn their head as listening gets difficult to improve their ability to hear speech in background noise. This was done without instruction to do so and demonstrates a natural behaviour. While none of these studies directly examined listening effort, attention during a listening scenario is the theory behind using secondary task performance as a measure of listening effort, as increased attention to the listening task will reduce the resources for secondary task performance (Pichora-Fuller et al., 2016). Speech comprehension or speech recognition is often used alongside listening effort in studies, and shows similar changes with increasing task difficulty for more challenging conditions (Rennies et al., 2014).

Listening effort studies often do not reflect performance in real world, and the effect of challenging listening conditions such as background noise and reverberation on listening effort may be greater for hearing impaired listeners. This could be in part due to a lack of complexity in the background noise used in lab studies, as predicted by Mackersie et al. (2015), or a focus on only speech intelligibility as a metric. Our experiment did not test speech intelligibility or comprehension, only listening effort. In their study with hearing impaired listeners, Brons et al. (2014) found that noise reduction available on three commercially available HAs did not affect speech intelligibility but did reduce noise annoyance, evidence that listening effort might be a more sensitive measure than speech intelligibility. Our study was designed using complex, recorded background noise played from all directions and used multiple measures of listening effort to be able to compare any similar effects or correlations between listening effort measures.

In order to capture visual attention behaviour simultaneously with listening effort, multiple measures were used in this study. Both HF HRV and RMSSD were used as physiological measure of listening effort along with subjective ratings of listening effort, and the motion capture system was used on both actors and audience members. The purpose of adding motion capture to this experiment was to examine the audience members tracking behaviour and see where they were looking relative to the actor currently speaking. This tracking behaviour was then compared to their listening effort from physiological and subjective measures.

3.2 Methods

This experiment was designed to evaluate listening effort during a live play performance in background noise and reverberation. The performers were an improv group from The Second City in Toronto. Each act incorporated actors moving around and switching talkers naturally on stage. The content and style of the improvised play were aimed at a younger audience and sophisticated enough to make listening effort fairly high in the background noise and reverberation, even for normal hearing listeners. The three different hearing aid conditions tested were: no hearing aid, a hearing aid in standard forward-pointing beamformer mode, and a hearing aid using Unitron's target-following directionality. Participants were aware if they were not wearing hearing aids, but the experimenter and the subjects were blinded to the type of processing on the hearing aids during the experiment. The acts were each sixteen minutes long. There was one practice block (Act 1) which allowed both the actors and the subjects to familiarize themselves with the task and the listening environment. At the end of this practice act, the actors were given feedback to keep their speech at an appropriate level to maintain a challenging signal-to-noise ratio, and the hearing aid processing condition was switched for the subjects before starting Act 2. Data collected during Act 2 is reported on in this paper. Continuous subjective listening effort ratings and ECG data was collected from the subjects during the performance, and motion capture of head position and movement was collected for both subjects and actors.

Six male, and 17 female undergraduate students were recruited, all fluent in English and with self reported normal hearing. Partial course credit was given to students for participation. The subjects were age 18-22 years, with a median age of 18. This work was approved by the McMaster University Research Ethics Board, Protocol 2014 125. The layout of the LIVELab is shown in Fig. 3.1. The subjects were seated in rows A-C in the LIVELab to maximize the necessity of head movement to track the actors on stage, and so that the binaural processing on the HA would detect a change in speaker location when the actor speaking changed during the performance. From seat B8, if the actors were close to stage front, this created an azimuth range of approximately $\pm$ 30-degrees. For a subject seated in seat A4, if the actor was at stage front on the far side of the stage, they would be at approximately a 60-degree azimuth. A KEMAR manikin was placed in seat B8 to measure sound levels and record the performance.

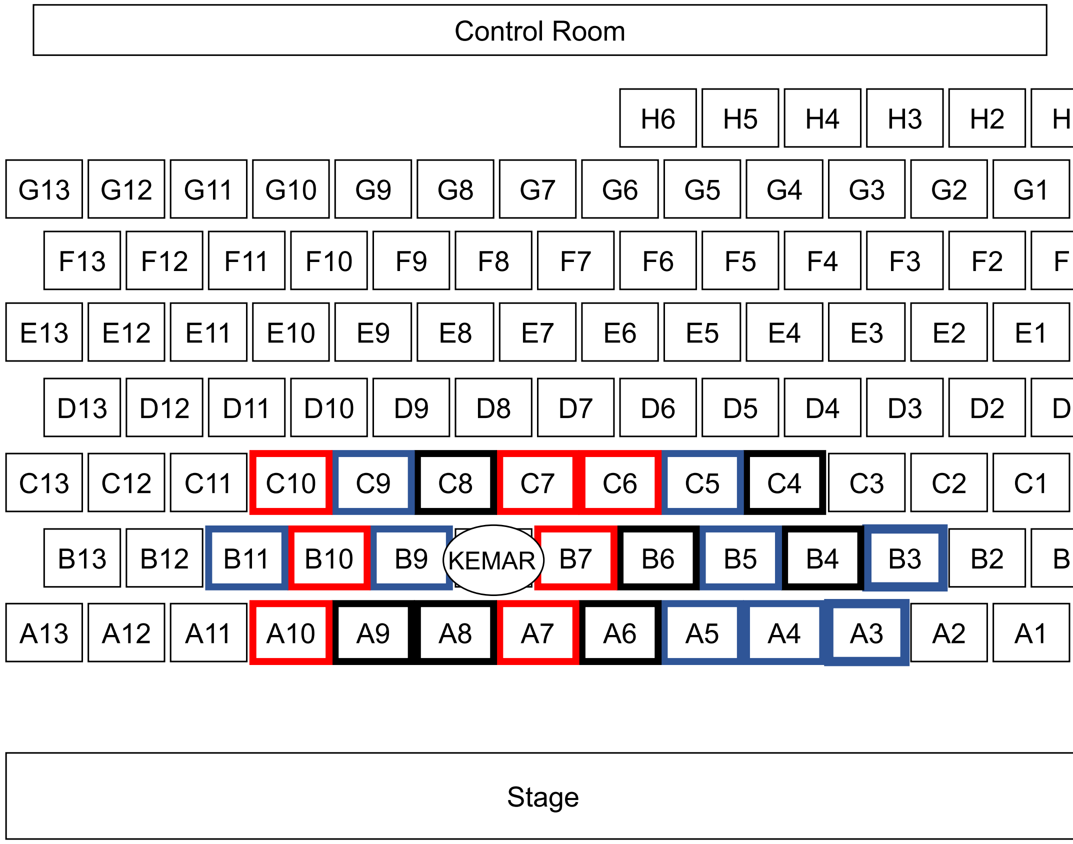


Figure 3.1: Layout of the LIVELab and location of KEMAR for Experiment 1. Highlighted seats indicate where participants were seated. Blue : no HA, Red : Forward-pointing HA, Black : Target-following HA

The subjects were randomly assigned to one of three HA groups for the practice act, and a different HA group for the experimental act. For each act there were 9 subjects in the no HA group, 7 in the standard forward-pointing beamformer HA group, and 7 in the target following HA group. The HAs used were Unitron Moxi Fit RIC. The HAs were programmed with minimal gain, shown in Fig. 3.2, so that the hearing aids processing would be active without providing excess gain to our normal hearing subjects. HAs were fit by audiologists from Unitron, who ensured that the RICs fit snugly into each participant's ear canal.

The target-following directionality is part of the Unitron Tempus platform. The algorithm detects which direction speech is coming from using binaurally linked microphones (referred to as Speech Locator) and adaptively steers beamforming to the target (Speech Focus). It incorporates guided back beamforming with front awareness when speech is detected from the back, and asymmetrical responses when speech is detected from one side of the wearer. These two algorithms together are referred

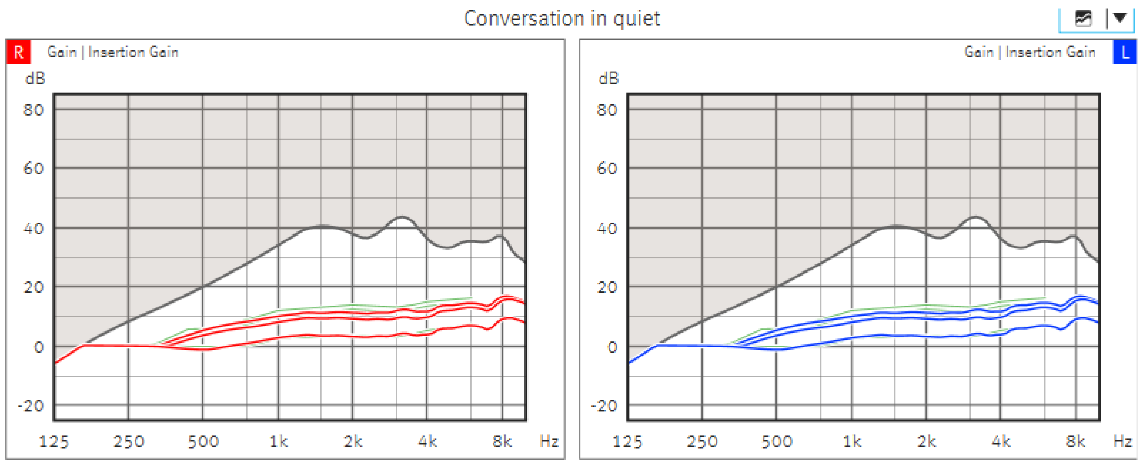


Figure 3.2: HA gain taken from Unitron Trufit software. The curves (from bottom to top) represent the HA insertion gain for 80, 65, and 50 dB speech. Green curves are the target gains, and the red and blue curves (left and right panels respectively) represent the achieved gains in the right and left ear.

to in the Unitron platform as Speech Pro. The platform also has an algorithm that applies separate frequency shaping to the left and right hearing aid input signal to adaptively restore natural localization cues degraded by previous processing (referred to as Dynamic Spatial Awareness). In this study, the target-following directionality is compared with a forward-pointing beamformer that focuses on sounds within the ±30-degree azimuths.

During the performance, participants were asked to continuously rate how difficult it was to follow the conversation on stage using a slider on an ASUS Nexus 7 tablet. The question displayed on the tablet, along with the slider range and corresponding data values, are shown in Fig. 3.3. Subjects were told to continuously move the slider during the performance and the slider value reset to neutral (50) every two minutes, corresponding to when the acoustics in the lab changed. The acoustics and background noise level in the LIVELab changed every two minutes during each act to give a variety of listening difficulty for the participants.

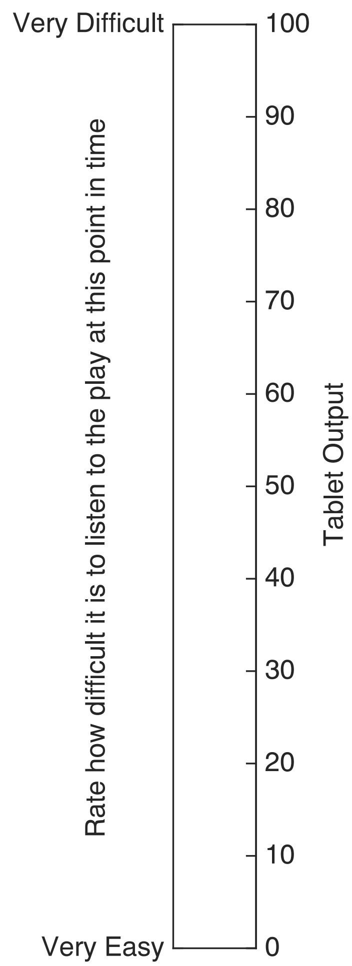


Figure 3.3: The question displayed on the tablets and slider values for Experiment 1

The "food-court" noise, long term spectrum shown in Fig. 3.4, was played continuously from the sound system at either 69 dB SPL or 75 dB SPL as measured by KEMAR. The "food-court" noise recording was supplied by Unitron, and was recorded in a moderately busy, relatively reverberant food-court at the Conestoga Mall in Waterloo using a Zoom Handy model H2n portable digital recorder. The four channels of "food-court" noise were distributed to speakers throughout the LIVELab to create an acoustic environment with noise coming from all directions.

The reverberation was set to classroom ($T_{60} = 0.6s$) or concert hall ($T_{60} = 2.15s$)

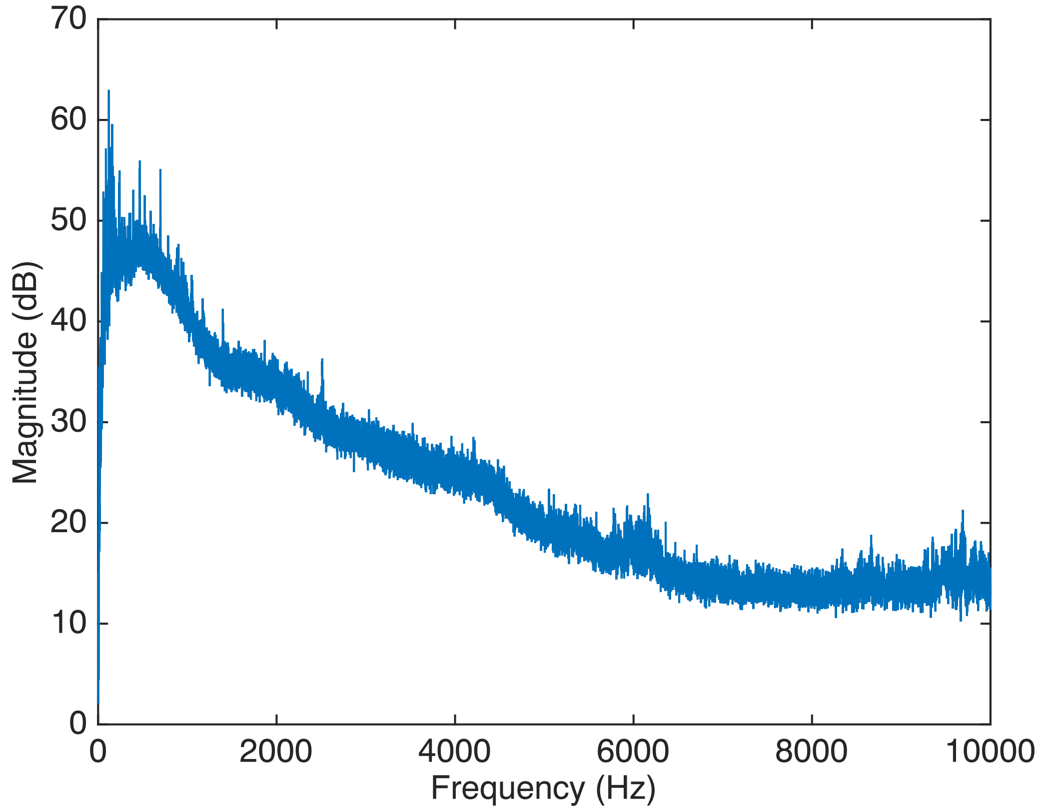


Figure 3.4: Frequency spectrum of the "food court" noise.

in the Meyer Sound Constellation Active Acoustics. The Constellation system uses 28 microphones and 75 speakers and subwoofers mounted on the wall and ceiling to simulate reverberation and also play the background noise. The Variable Room Acoustic System (VRAS) algorithm (Poletti, 1999, 2000) is used to create the reverberation. The order of the background noise level and reverberation settings were balanced using a partial latin square design.

ECG data was collected using a g-Tec (Austria) g.USBamp system using one electrode placed on the left chest, approximately ECG lead position V5. The sampling rate was 512 Hz and the raw data was filtered from 0.1 to 100 Hz with a notch at 60 Hz to remove noise. Two measures of heart rate variability were calculated from the ECG data collected during the experiment: root mean square successive difference (RMSSD) and high frequency variability. The equation for RMSSD is given in 3.1, where $R-R$ is the time interval between two QRS peaks in the ECG. High frequency HRV is the power in the 0.15 – 0.4 Hz frequency band of the ECG R-R interval series, estimated using the plomb() function in Matlab (Tarvainen et al., 2014) to account for uneven spacing between QRS peaks. Both RMSSD and HF HRV analysis were done in 30 second windows and then averaged for the two minutes of background noise and reverberation conditions.

$$\text{RMSSD} = \sqrt{\frac{1}{N-1}(\sum_{i=1}^{N-1}((R-R)_{i+1}-(R-R)_i)^2)} \tag{3.1}$$

Motion capture was done using the Qualysis Opus 5+ Motion Capture System with 28 IR cameras capturing both audience and actors on stage. Audience and actors wore caps with four motion capture markers on a stationary cap to allow for calculation of head direction. Labeling of each person and exporting the data was done using Qualysis software, with further analysis done in Matlab. A video camera pointed at the stage recorded the whole performance for the purpose of determining which actor was talking for the motion capture analysis.

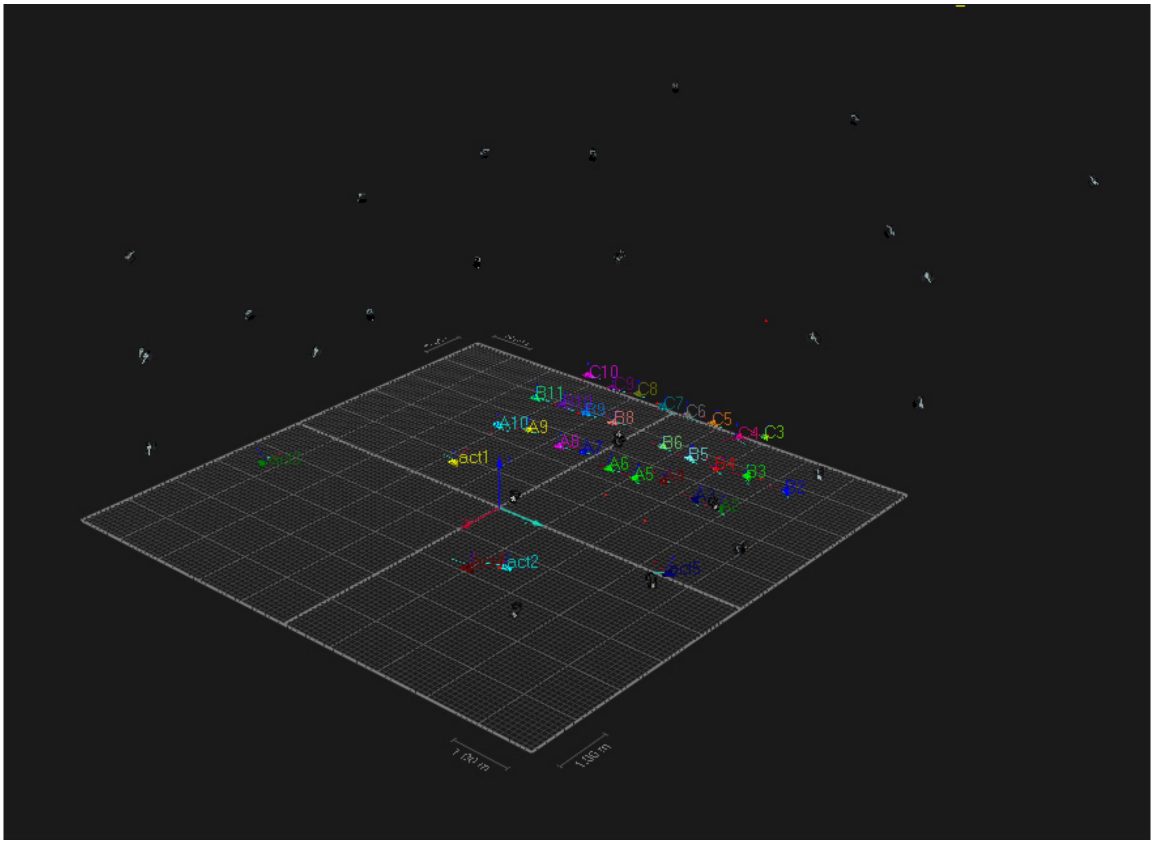


Figure 3.5: Example screenshot of the motion capture data from Qualysis software showing the location of motion capture camera network in the ceilings as well as audience members labeled by their seat location and the actors labeled act1-5.

3.3 Results

A summary of main effects is given in Table 3.1.

3.3.1 Subjective ratings

The mean value of the tablet slider was taken for each two minute window, corresponding to when the reverberation and background noise changed. The group effect of reverberation setting (classroom or concert hall), background noise level (69 or 75 dB SPL, as measured by KEMAR), HA type (no HA, forward-pointing beamformer HA or target-following HA), and the interaction between these variables, were examined using R. Significant main effects were found for seat row (p=0.041), reverberation

Table 3.1: ANCOVA p values from main effect analysis for all listening effort measures. Bold values indicate significant main effects.

Measure **Variable**	Subjective Rating	RMSSD	HF HRV	Deviation Angle
HA Type	0.72	0.567	0.497	0.542
Seat row	**0.041**	0.746	0.548	0.189
Seat number	0.492	0.972	0.635	0.775
Reverberation setting	**0.005**	0.119	**0.013**	**0.006**
Background noise	**0.0008**	0.823	0.374	**0.0003**
Background noise: Reverberation setting	**0.001**	0.352	0.509	0.943
HA Type : Reverberation setting	0.959	0.98	0.81	0.241
Seat row : Reverberation setting	0.9	0.79	0.63	0.097
Seat number : Reverberation setting	0.998	**0.012**	**0.004**	0.174
HA Type : Background noise	0.96	0.183	0.377	**0.031**
Seat row : Background noise	0.549	0.159	0.221	0.132
Seat number : Background noise	0.218	0.365	0.241	0.177
Resting HRV	NA	0.084	**0.016**	NA

setting (p = 0.0048), background noise level (p < 0.001), and the interaction between reverberation and background noise (p = 0.0014). Results are shown in Fig. 3.6 and Fig. 3.7.

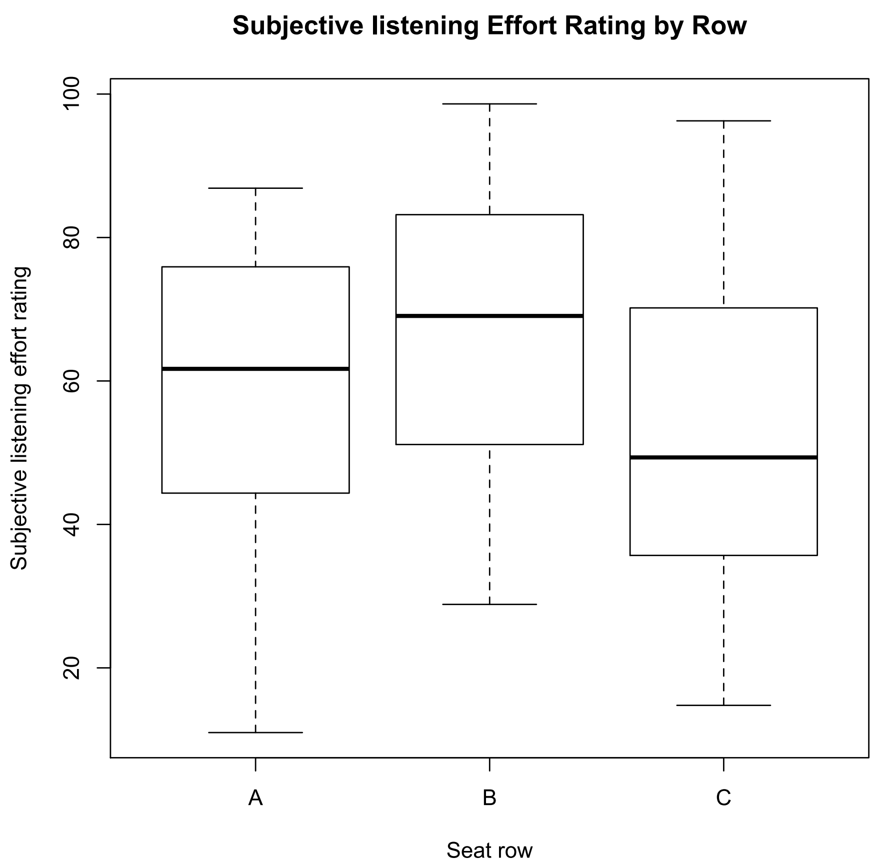
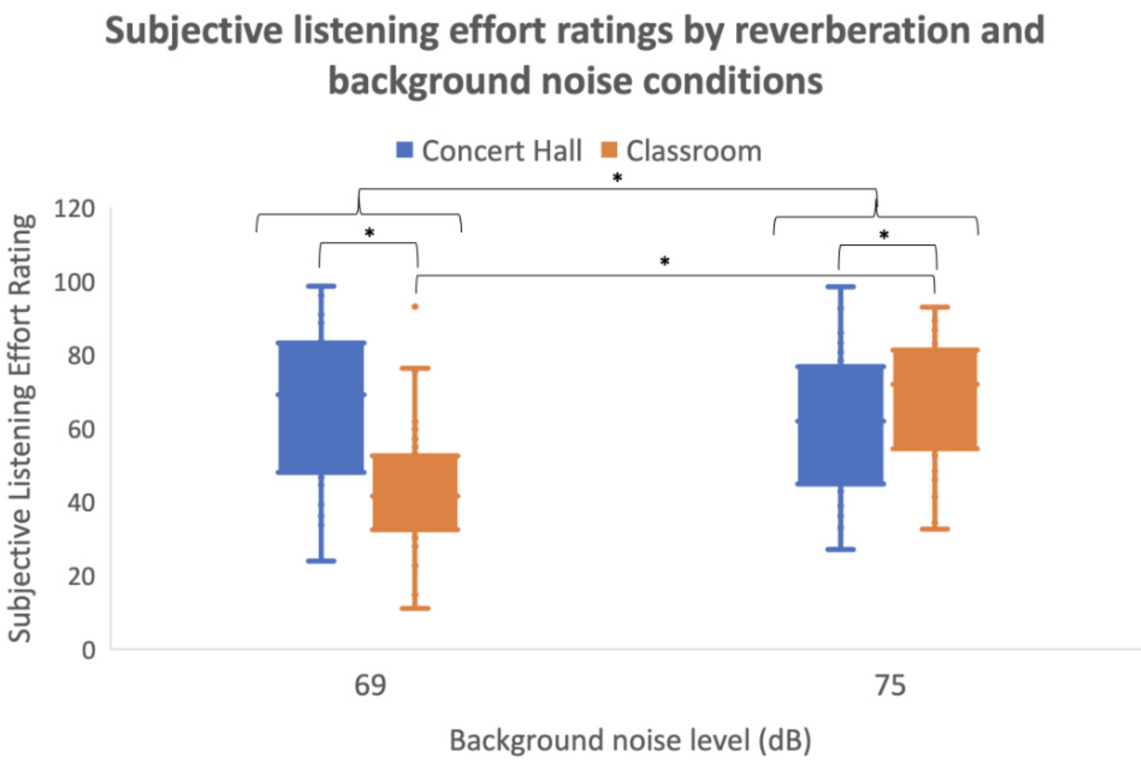


Figure 3.6: Subjective listening effort by seat row. A comparison between rows B and C gives a p value of 0.12 after correction.

3.3.2 Heart rate variability

Each HRV measure was calculated for the same 2 minute windows as the subjective ratings. Resting values for each measure were calculated using 4 minutes of data from prior to the experiment starting, while participants were seated. HRV analysis was done using Matlab and then exported to R for further statistical analysis.



![h]

Figure 3.7: Subjective listening effort by reverberation (concert hall (blue) and classroom (orange)) and background noise level. * $p < 0.001$

RMSSD

An ANCOVA was performed on the RMSSD values with resting RMSSD as a covariate, between subject factor of HA type, seat row and seat number, and within subject factors of reverb, background noise, and the interaction between reverb and background noise. Resting RMSSD had a p value of 0.08, and the interaction between reverberation and seat number had a p value of 0.012. Results for reverberation setting are shown in Fig. 3.8 for comparison with the HFHRV results.

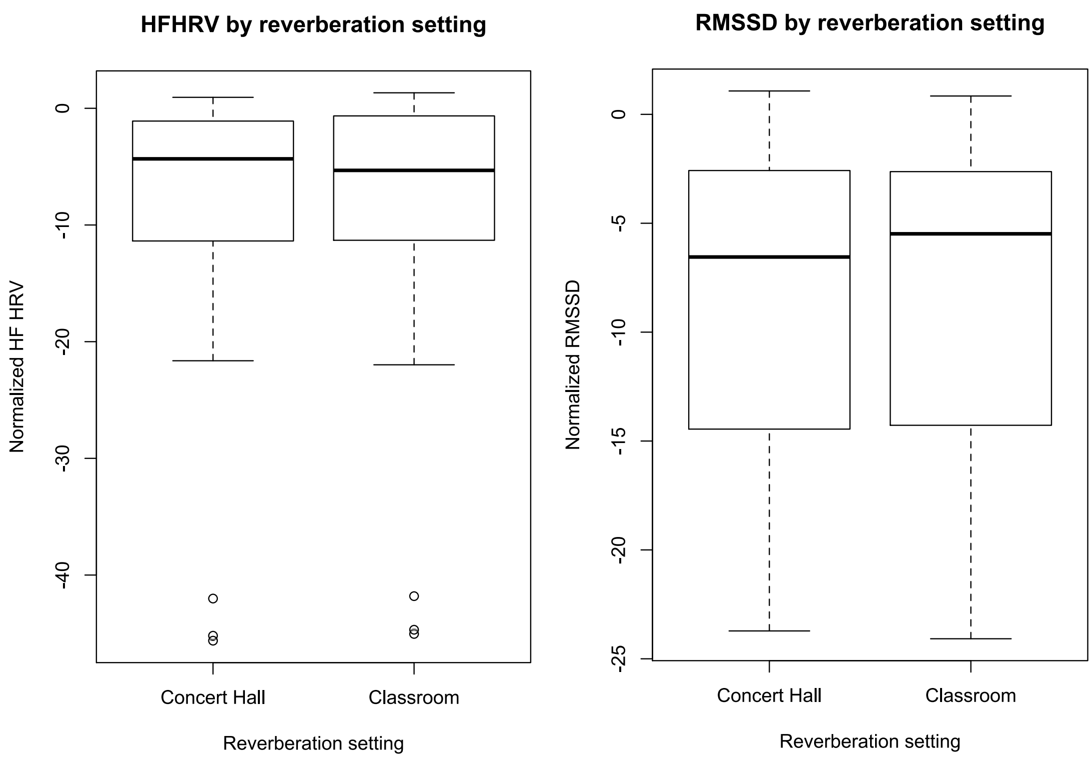


Figure 3.8: Normalized HF HRV and RMSSD by reverberation setting. Reverberation main effect p value for HF HRV p = 0.013, for RMSSD p = 0.119

HF HRV

An ANCOVA was performed on the HF HRV values with resting HF HRV as a covariate, between subject factor of HA type, seat row and seat number, and within subject factors of reverb, background noise, and the interaction between reverb and background noise. Resting HF HRV had a p value of 0.016, reverberation had a p value of 0.012 and the interaction between reverberation and seat number had a p value of 0.004. Results for reverberation setting are shown in Fig. 3.8.

3.3.3 Motion capture

For the whole performance, the deviation angle of each participants head position from the current actor speaking on stage was calculated from motion capture data tagged with actor location after the experiment. This was done using the position and angle data in Matlab, and averaged for each 2 minute period of changing background noise and reverberation. The direction of the current talker was taken as zero degrees, with the deviation angle calculated based on the subjects head direction relative to the talker on a 2D plane. The deviation angle data was then imported to R for further statistical analysis. An ANOVA was performed with between subject factors of HA type, seat row and seat number, and within subject factors of reverberation time, background noise, and the interaction between reverberation time and background noise. Reverberation time had a p value of 0.006, background noise had a $p < 0.001$. The interaction between HA type and background noise had a p value of 0.03.

Increasing background noise did increase the deviation angle, as shown in Fig. 3.9. The classroom reverberation setting lead to greater deviation angle compared to the concert hall reverberation.

In Fig. 3.10 the target following and forward pointing HA groups both had an increase in head deviation angle when background noise levels increased from 69 to 75 dB whereas the no HA group did not.

3.3.4 Correlation between measures

The only significant correlations after correcting for the number of comparisons is between the two ECG measures, HF HRV and RMSSD. This is likely due to the measures being from the same data.

Since both the physiological and subjective data give significant results similar to those found in previous studies it seems that they are all measuring listening effort, but different aspects of effort.

3.4 Discussion

From the subjective rating results, shown in Fig. 3.7, there is a clear increase in perceived subjective listening effort for increased background noise and greater reverberation. The interaction between reverberation and background noise, shown in Fig. 3.7, shows that the increase in subjective listening effort from reverberation is greater for the lower background noise level. This could be an indication that 75 dB

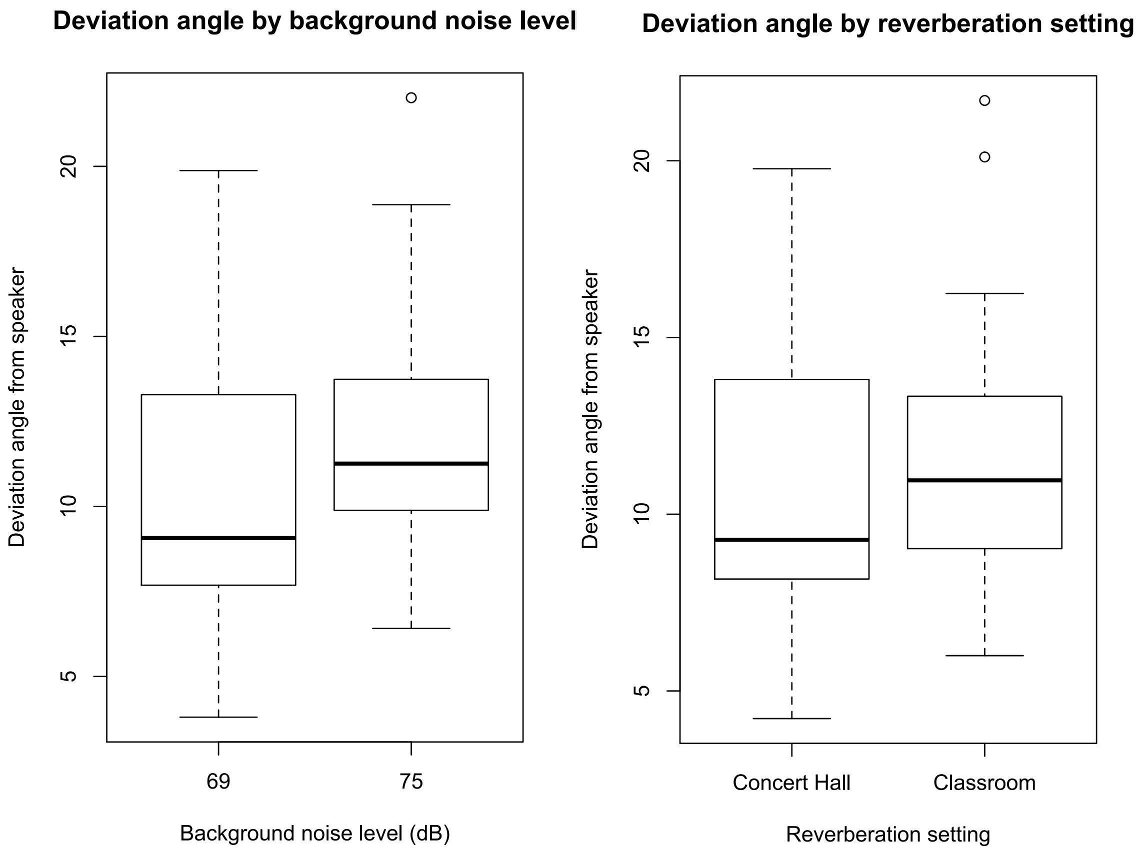


Figure 3.9: Deviation angle from the actor speaking on stage by background noise level and by reverberation setting. Both are significant main effects.

of background noise makes listening effortful enough in this scenario that increased reverberation does not cause an additional increase in listening effort. It could also be true that the 75 dB of background noise is enough for subjects to find it difficult enough that they start to give up on listening, though from Fig. 3.6 and Fig. 3.7 it appears that most subjects gave moderate to high subjective ratings and not the maximum listening effort.

Both HRV measures had significant interaction between reverberation setting and seat number, and HF HRV also had a significant main effect of reverberation, shown in Fig. 3.8. The seat number interaction could be driven by a few subjects due to the small number of subjects in each seat number, and the reverberation effect is not as clear as with subjective ratings and could also be driven by a small number of subjects. It is however interesting to find the same significant main effect of reverberation in

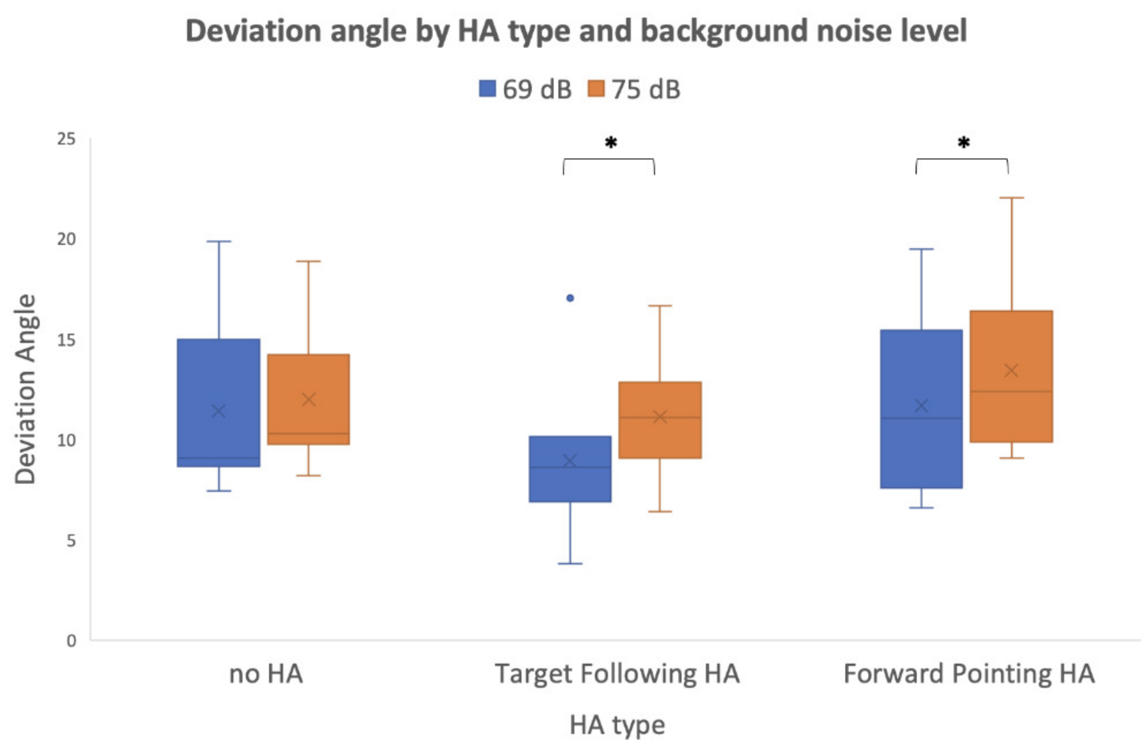


Figure 3.10: Deviation angle from the actor speaking on stage by HA type and background noise level. $*p \leq 0.01$

both subjective rating and HRV.

From Fig. 3.9 there is an increase in deviation angle with increasing background noise, which indicates that the subjects were turning their head to the left and putting their right ear more towards the stage at the front of the room. This head turn is similar to the natural behaviour found in Grange and Culling (2016). Subjects were not given any instructions on head movement so this is natural behaviour, but they were aware of motion capture markers on hats they wore during the experiment.

Fig. 3.10 shows this increase for both the forward pointing HA and target following HA groups but not the unaided group. This head turning behaviour could be due to multiple factors. Firstly, it is possible that the unfamiliar hearing aid processing may have been introducing some small increase in the listening effort that was not discernable in our subjective or objective measures of listening effort, and the increased listening effort in the two aided conditions led to the participants exhibiting a stronger right ear preference. That is, they naturally turned their head in an attempt to achieve a more favorable SNR in the right ear as the background noise increased in level. Secondly, this experiment was the only experience that the participants had with HA processing since they were young NH listeners, and they changed HA groups between the training session (Act 1) and the actual experiment (Act 2). Thus, the head turning behavior may have been an attempt by the participants to adjust to the unfamiliar directional processing of the HAs. However, it is unclear in this case why the group with the target following HA processing would exhibit a robust right-ear preference, when the algorithm works to apply the directional processing in the estimated direction of the target talker. In addition, the target-following directionality does take a small amount of time to adjust to the new estimated direction when the talker switches. Thirdly, the right-ear bias could also be due to the angles covered by the directional processing compared to the range of the actors on the stage. The actors could be at the 60 degree azimuth if they are on the furthest side of the stage from a participant in the outermost seats in each row, but the forward pointing directionality has a beam width of approximately ± 30 degrees, so the participants would not need to turn their heads completely in the direction of the talker for them to fall within the beam. However, it is uncertain why the two HA groups would exhibit this behaviour more to one side than the other. This could be evidence of an effect lead by individual subjects since there is no within subject comparison.

Similar to previous studies on listening effort, our study shows significant effects for some measures and factors but not consistently for all listening effort measures. There is also no correlation between the subjective and physiological measures, which has also been true in some previous studies (Mackersie et al., 2015). From the subjective ratings there was a significant effect of both background noise level and reverberation, similar to the results found in Rennies et al. (2014). We did also find evidence of the right ear bias discussed in Marzoli and Tommasi (2009) from the deviation angle results.

3.5 Limitations and Future Directions

There are limitations with the current study that we hope to improve with a follow up study using hearing impaired listeners. First, the participants were all normal hearing, so the task may have been too easy to show significant differences between the HA conditions. There was also only one act in the experiment, so the subjects only experienced one HA condition and there is no within subject comparison. We also found that, when using live performers, having a listening effort that changed appropriately with changes in the background noise was challenging. During our practice session (Act 1) the actors tended to compensate for additional background noise rather than keeping a consistent speech volume regardless of background noise level. The results of our experiment done during Act 2 demonstrate that the actors were better able to keep their speech levels consistent such that the listening effort did change appropriately with changes in the reverberation and background noise. The instructions to the actors and their ability to avoid compensating for background noise changes will remain an issue for future studies.

3.6 Conclusions

The significant effects of reverberation and background noise on subjective listening effort as well as the significant effect of reverberation on HF HRV show measurable differences in listening effort with within subject comparison. Within subject comparison of HA type and seat location could not be done for this experiment but would be in future experiments and could reveal additional effects on listening effort seen in other listening effort studies with hearing aids.

This realistic live performance based procedure does allow for the measurement of listening effort in multiple subjects simultaneously with results that show similar effects to those shown in individual lab tests of previous studies. This method is useful for the testing of different acoustic environments and their effect on listening effort. With modifications to ensure consistent actor speech levels and within subject comparison it could also be useful for testing hearing aid algorithms in a realistic environment.

Chapter 4

Comparing listening effort and speech intelligibility measures using multidirectional speech in background noise

4.1 Introduction

Listening effort is an important indication of listening performance that can be measured in many ways from subjective ratings to physiological data. Pichora-Fuller et al. (2016) has an extensive review of many listening effort measures and how they have been used in studies, but most studies only consider one or two measures together in the same experiment. Studies using different measures of listening effort show similar effects of SNR and speech transmission index on listening effort but results are not always significant, and when used in the same experiment different listening effort measures are typically uncorrelated.

Both Rennies et al. (2014) and Sato et al. (2012) tested listening effort and speech intelligibility in reverberation and found that listening effort could change even when intelligibility was high. Both used varied artificial reverberation and signal to noise ratios to have a range of listening effort and intelligibility. Rennies et al. (2014) concluded that listening effort did increase with decreasing speech transmission index for both SNR and reverberation. They also found that intelligibility was more sensitive for lower SNRs, whereas listening effort was a more sensitive measure at higher SNRs. Sato et al. (2012) found similar results in their study. In their young normal hearing group, listening difficulty varied from 0 to 60% while intelligibility remained close to 100%. In their older group with mild to moderate hearing loss intelligibility was still high, between 80-100%, with listening difficulty varying from 0 to 60% similar to the younger group.

Subjective ratings are a commonly used measure of listening effort, but are often combined with an objective measure in studies. One such measure is heart rate variability (HRV) calculated from ECG data. Various measures of heart rate variability have been used, all based on the variability in times between QRS spikes in ECG. The use of these measures is based on the physiological source of the variability from the parasympathetic nervous system, as detailed in Berntson et al. (2017). These include a time based measure, root mean square succesive difference (RMSSD), as well as frequency based measures of low frequency and high frequency HRV. Cvijanović et al. (2017) compared LF HRV and the LF/HF ratio to measure the difference between various noise conditions to establish the sensitivity of the measures to changes in listening effort. The study compared communication between two participants at three different SNRs: no noise, 6 dB, and -6 dB. They found no significant effect for LF HRV or LF/HF HRV ratio. Mackersie et al. (2015); Mackersie and Calderon-Moultrie (2016) both compared high frequency HRV in normal hearing and hearing impaired (HI) individuals while changing listening difficulty by altering the SNR (Mackersie et al., 2015) or the speaking rate (Mackersie and Calderon-Moultrie, 2016). Both showed decreased HF HRV for the HI group compared to the normal hearing group for lower SNRs. Mackersie and Calderon-Moultrie (2016) also showed an increase in skin conductance in noise for HI compared to normal hearing, despite similar sentence scores. Mackersie et al. (2015) found no significant effect of SNR on subjective measures, and no significant correlations between listening effort measures. Mackersie and Calderon-Moultrie (2016) did show a decrease in HF HRV and increase in skin conductance with increasing speaking rate however they do not state whether this was a significant correlation.

Another physiological measure used for listening effort that is not related to the respiratory or circulatory system is using EEG activity to measure resource use or parasympathetic activity. This is typically done using the alpha band of EEG, most associated with rest and relaxation. For example, Bernarding et al. (2014) use phase distribution using a wavelet transform at a frequency on the alpha/theta band border to measure listening effort. This frequency was determined to best represent effort based on their previous study looking at phase distribution at various frequencies (Bernarding et al., 2012). Phase distribution that is less variable corresponds to more synchronized firing of neurons in that area of the cortex, and therefor a greater predicted effort. Using phase distribution of EEG in the frontal region, Bernarding et al. (2014) found that strong directional processing on HAs did reduce listening effort compared to weak or no directional processing. Bernarding et al. (2017) similarly found a significant effect of HA settings on subjective listening effort. They also found that the same EEG phase distribution measure from Bernarding et al. (2012, 2014) was highly correlated with subjective effort ratings in their follow up study

Bernarding et al. (2017).

Dimitrijevic et al. (2017) also used the alpha band, but this time alpha power. They examined the possibility of using alpha power to predict both speech intelligibility (digit in noise identification), and differences between active and passive listening. They found a correlation between alpha band power in the temporal region and speech intelligibility. There was also more change in alpha band activity with active listening versus passive listening. Alpha band is thought to correspond to inhibition of higher level processing, so a greater alpha band power corresponds to lower predicted effort. In a follow up study (Dimitrijevic et al., 2019) they again looked at alpha power as a predictor of listening effort, but also coherence between EEG and the target speech. They found a correlation between alpha power and subjective listening effort in the left frontal region after mapping EEG electrode signals. They also found a significant correlation between coherence in the 2-5 Hz band and subjective listening effort in the left temporal region. Coherence also had significant correlation with correct digit identification (speech intelligibility) in the left frontal region. Coherence is an EEG based metric that has been shown in modern studies to be sensitive to listening effort, but it was not used here due to it requiring a clean version of the exact stimulus heard by the subject.

By collecting the EEG data in all subjects simultaneously, such as in the study discussed in the Chapter, it is also possible to capture the effects of social interaction between people on their brain activity, as discussed in papers on hyperscanning or the collection of simultaneous EEG (Czeszumski et al., 2020; Ahn et al., 2017). Though this effect was not analyzed it is accepted as a normal part of ecologically valid scenarios involving listening to speech.

4.2 Methods

This experiment was designed to evaluate listening effort using speech from three directions in background noise and reverberation. Two male, and 22 female undergraduate students were recruited, all fluent in English and with self reported normal hearing. Partial course credit was given to students for participation. The subjects were age 18-21 years, with a median age of 18. This work was approved by the McMaster University Research Ethics Board, Protocol 2014 125.

The layout of the LIVELab is shown in Fig. 4.1. The subjects were seated in rows C-E in the LIVELab with speakers located to the front, left side, and behind the subjects as shown in Fig. 4.1. All speakers were the same distance from the KEMAR.

The “food-court” noise, long term spectrum shown in Fig. 4.2, was played continuously from the sound system at 75 dB SPL as measured by KEMAR. The “food-court” noise recording was supplied by Unitron, and was recorded in a moderately busy, relatively reverberant food-court at the Conestoga Mall in Waterloo using a

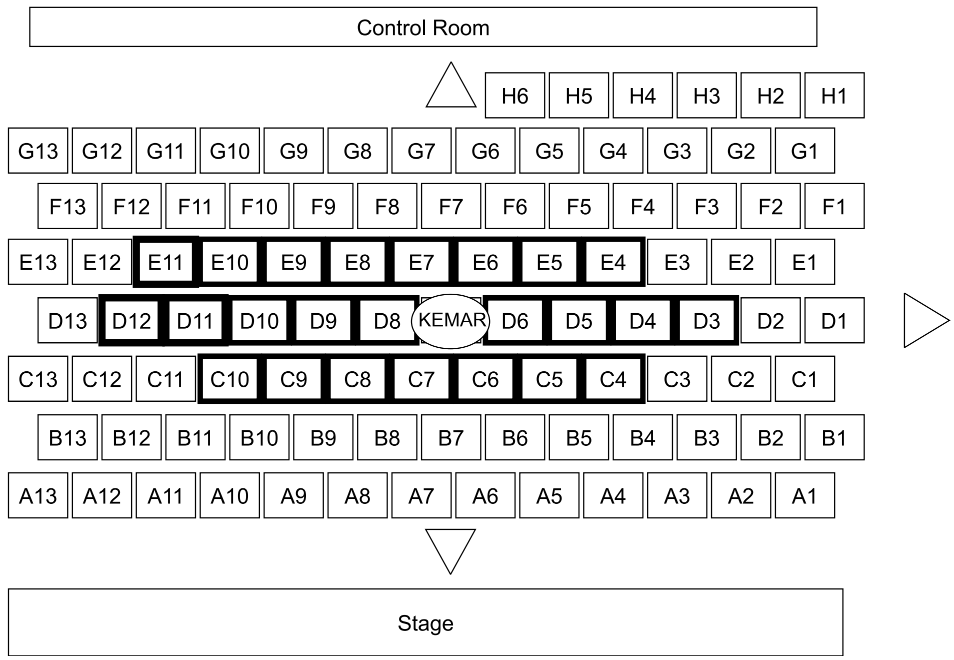


Figure 4.1: Layout of the LIVELab and location of KEMAR for Experiment 2. Highlighted seats indicate where participants were seated and triangles indicate speaker location.

Zoom Handy model H2n portable digital recorder. The four channels of "food-court" noise were distributed to speakers throughout the LIVELab to create an acoustic environment with noise coming from all directions. The LIVELab reverberation was set to classroom ($T_{60} = 0.6s$).

The subjects were randomly assigned to one of three HA groups which determined the order they experienced the three HA conditions: no hearing aid, a hearing aid in standard forward-pointing beamformer mode, and a hearing aid using Unitron's target-following directionality. The HAs used were Unitron Moxi Fit RIC. Participants were aware if they were not wearing hearing aids, but the experimenter and the subjects were blinded to the type of processing on the hearing aids during the experiment. The HAs were programmed with minimal gain, shown in Fig. 4.3, so that the hearing aids processing would be active without providing excess gain to our normal hearing subjects. HAs were fit by audiologists from Unitron, who ensured that the RICs fit snugly into each participant's ear canal.

The target-following directionality is part of the Unitron Tempus platform. The algorithm detects which direction speech is coming from using binaurally linked microphones (referred to as Speech Locator) and adaptively steers beamforming to the

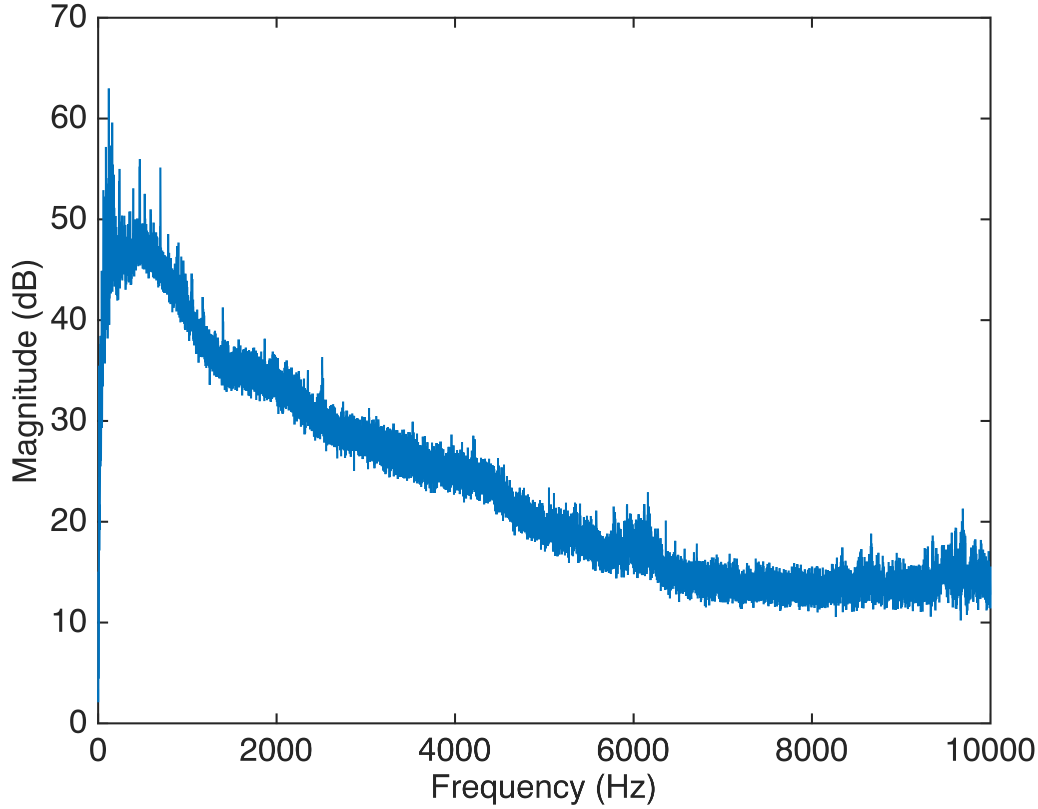


Figure 4.2: Long term frequency spectrum of the “food-court” noise.

target (Speech Focus). It incorporates guided back beamforming with front awareness when speech is detected from the back, and asymmetrical responses when speech is detected from one side of the wearer. These two algorithms together are referred to in the Unitron platform as Speech Pro. The platform also has an algorithm that applies separate frequency shaping to the left and right hearing aid input signal to adaptively restore natural localization cues degraded by previous processing (referred to as Dynamic Spatial Awareness). In this study, the target-following directionality is compared with a forward-pointing beamformer that focuses on sounds within the ±30-degree azimuths.

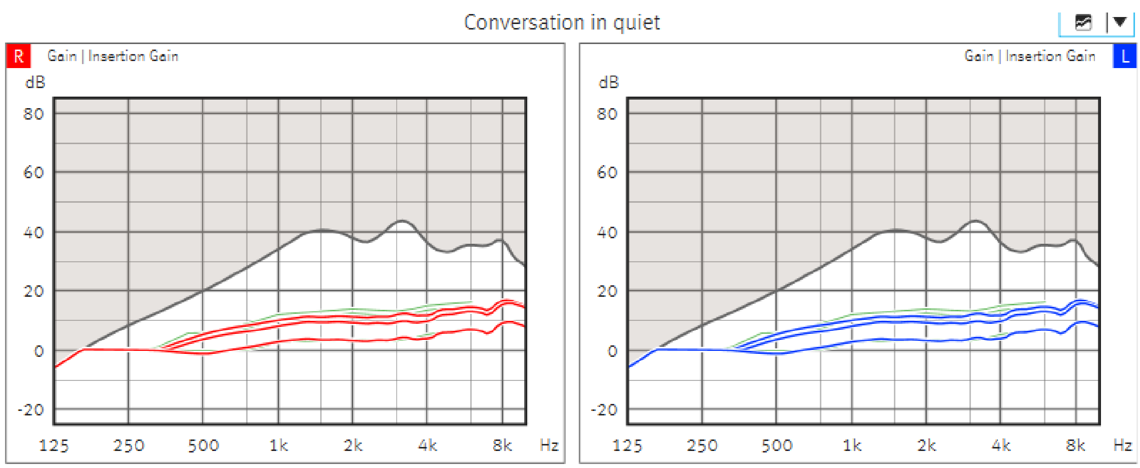


Figure 4.3: HA gain taken from Unitron Trufit software. The curves (from bottom to top) represent the HA insertion gain for 80, 65, and 50 dB speech. Green curves are the target gains, and the red and blue curves (left and right panels respectively) represent the acheived gains in the right and left ear.

The experiment was done in three blocks. During each block, twenty sets of sentences were played randomly from one of the three speakers marked in Fig. 4.1 and participants were asked to write down the last sentence in each set. Only the last sentence was written down to avoid the task becoming more of a memory task than a listening task, and to give the hearing aid processing time to detect and switch to a new speech location. Each set consisted of seven sentences from the Trainor et al. (2004) r-HINT-e recordings of HINT sentences. Sentences were played at 69 dB SPL as measured by KEMAR. This gave a range of SNRs based on where the participants were seated relative to the speakers. Participants also rated their subjective listening effort at the end of each block. In between blocks participants remained seated while audiologists switched the HAs they were wearing for the next block.

4.2.1 ECG and EEG data analysis

ECG data was collected using a g-Tec (Austria) g.USBamp system using one electrode placed on the left chest, approximately ECG lead position V5. The sampling rate was 256 Hz and the raw data was filtered from 0.1 to 100 Hz with a notch at 60 Hz to remove noise. Two measures of heart rate variability were calculated from the ECG data collected during the experiment: root mean square successive difference (RMSSD) and high frequency variability. The equation for RMSSD is given below in Equation 4.1, where $R-R$ is the time interval between two QRS peaks in the ECG.

$$\mathrm{RMSSD} = \sqrt{\frac{1}{N-1}(\sum_{i=1}^{N-1}((R-R)_{i+1}-(R-R)_i)^2)} \tag{4.1}$$

High frequency HRV is the power in the 0.15 – 0.4 Hz frequency band of the ECG R-R interval series, estimated using the plomb() function in Matlab (Tarvainen et al., 2014) to account for uneven spacing between QRS peaks. Both RMSSD and HF HRV analysis were calculated for the time that sentences were being played from the speakers, plus one second before and after.

EEG data was collected using the same g-Tec (Austria) g.USBamp system, with electrodes placed at Pz, Cz, F3, F4, FP1, FP2, and Oz, and a clip electrode on the ear for reference. The same sampling rate and filtering were used as for the ECG. Initial analysis of the EEG data was done using EEG lab to remove segments of EEG with large artifacts and noise. EEG analysis was done using 1.5s epochs. Band power was calculated using pwelch() in Matlab for all channels, and taking the average for the epochs corresponding to each set of sentences with an additional 1s before and after. No baseline value was used as the EEG data prior to starting the experiment was too noisy in many subjects.

The instantaneous phase distribution was calculated according to Bernarding et al. (2012, 2014, 2017). First the Hilbert transform was used to get the envelope of the signal, then a continuous wavelet transform with a sixth derivative Gaussian wavelet with a frequency on the alpha-theta band border that previous papers determined to be the best frequency for measuring listening effort (Bernarding et al., 2012, 2014, 2017). A Rayleigh test was then applied to the instantaneous phases. This phase distribution was normalized using the difference between the maximum and minimum Objective Listening Effort (OLEosc or OLE) (Bernarding et al., 2017) values for each subject.

4.3 Results

A summary of the statistical analysis for the main effects on each listening effort or speech intelligibility metric are given in Table 4.1.

Table 4.1: ANOVA and ANCOVA results from experiment 2. p values shown for main effects. * $p < 0.05$

Factor	HFHRV	RMSSD	Alpha Power	Norm OLE	Beta Power	Gamma Power	Subjective Ratings	Sentence Score
Seat Row	0.23	0.25	0.63	0.053.	0.86	0.70	0.32	0.015 *
Seat Number	0.027 *	0.12	0.97	0.87	0.67	0.90	0.081 .	0.26
HA Type	0.079 .	0.58	0.40	0.22	0.81	0.96	0.00036 *	0.0016 *
Seat Row: HA Type	0.63	0.60	0.99	0.60	0.46	0.33	0.36	0.46
Seat Number: HA Type	0.024 *	0.12	0.72	0.89	0.89	0.92	0.65	0.27
Speaker	0.33	0.19	0.90	0.55	0.88	0.77	N/A	9.9e-12 *
Seat Row: Speaker	0.65	0.31	0.98	0.80	0.74	0.12	N/A	4.9e-07 *
Seat Number: Speaker	0.029 *	0.10	0.96	0.76	0.82	0.47	N/A	0.00044 *

4.3.1 Subjective ratings

The significant main effect found for the subjective ratings was by HA type. As seen in Fig. 4.4, the forward pointing HA condition was rated as being more effortful than either the no HA or target following HA condition. The no HA and target following HA subjective effort ratings were not significantly different from each other.

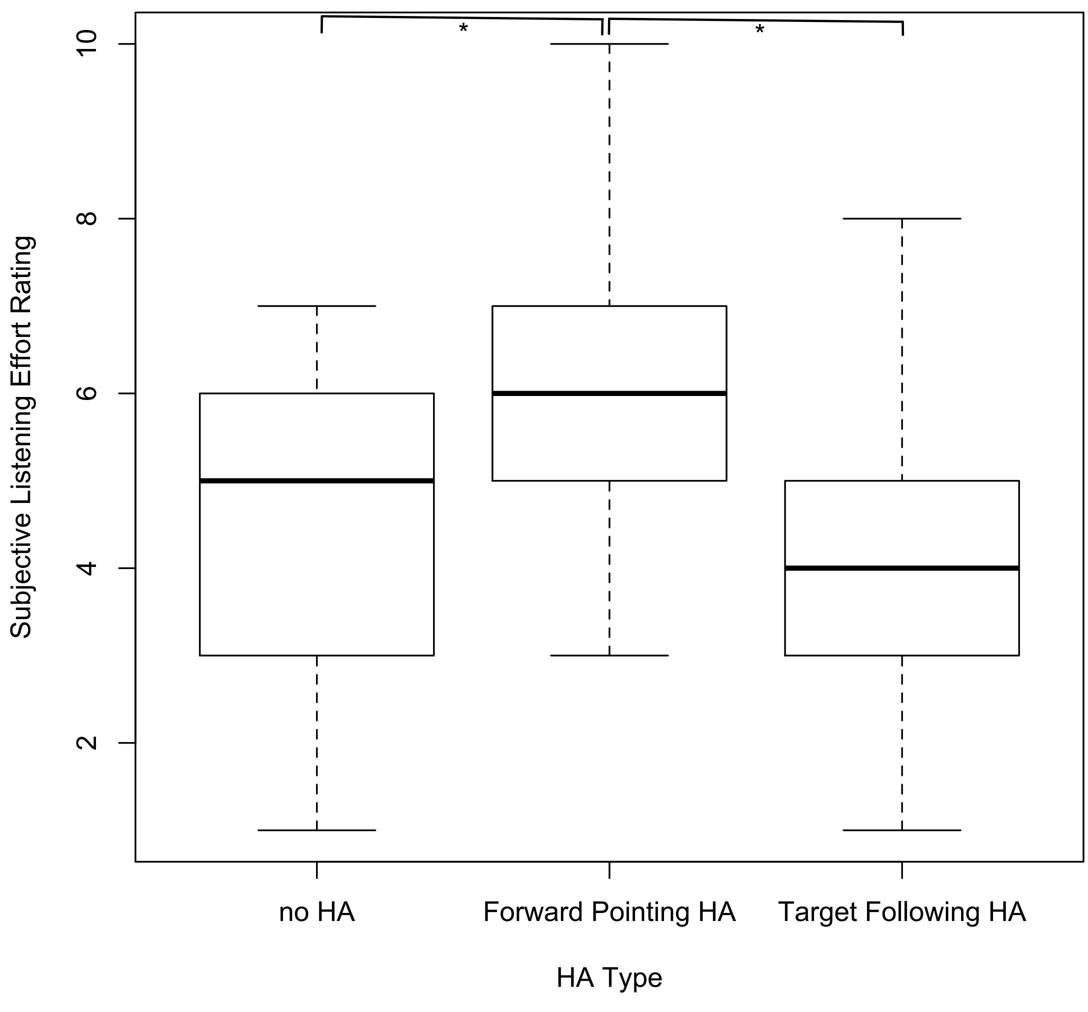


Figure 4.4: Subjective listening effort rating by HA Type. * p = 0.06.

4.3.2 Sentence Recognition Scores

Seat row, HA type, speaker location, and interactions of seat row with speaker location as well as seat number with speaker location all had significant main effects for sentence score. Figure 4.5 shows sentence scores by HA type. To visualize the interaction between seat location and speaker location, Figure 4.6 shows polynomial fits for the data for each HA type. For reference, the side speaker is located closest to seat 1, the back speaker behind row E and the front speaker in front of row C.

4.3.3 Heart rate variability

Each HRV measure was calculated for the duration of each sentence block, on average 12s, plus 1s before and after the stimulus is played. Resting values for each measure were calculated using 4 minutes of data from prior to the experiment starting, while participants were seated.

A main effect of seat number and the interaction between seat number and HA type was found, but from the data this appears to be driven by a small number of subjects so it was not further explored.

RMSSD

An ANCOVA was performed on the RMSSD values with resting RMSSD as a covariate, between subject factor of seat row and seat number, and within subject factors of reverb, background noise,HA Type, speaker location, and the interaction between reverb and background noise. No factors had significant main effects for RMSSD.

HF HRV

An ANCOVA was performed on the HF HRV values with resting HF HRV as a covariate, between subject factor of seat row and seat number, and within subject factors of reverb, background noise, HA Type, speaker location, and the interaction between reverb and background noise. Seat number and the interaction between seat number and HA type both had significant main effects for HF HRV. The seat number effect appears to be driven by a small number of outliers as previously stated, so only the effect of HA type was examined further, as it was close to significant. This is shown in Fig. 4.7.

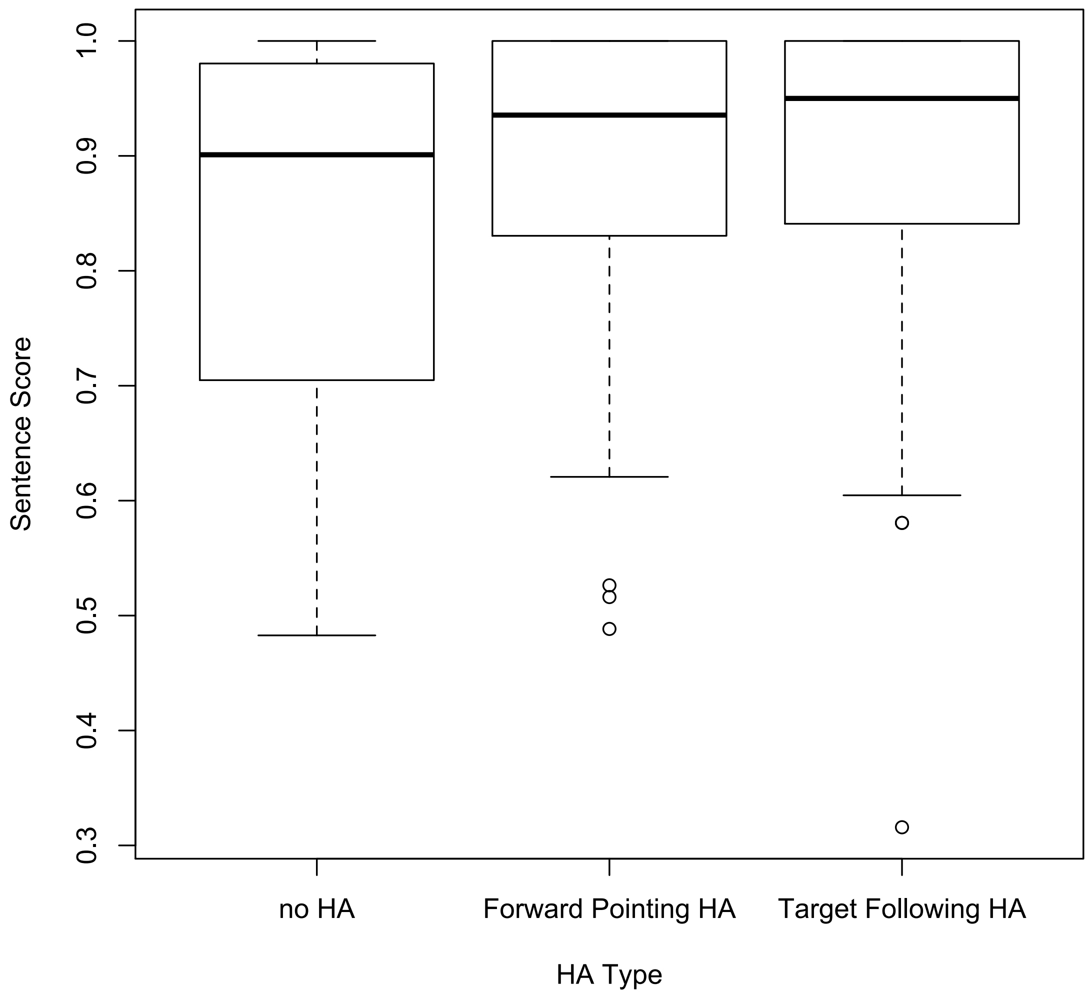
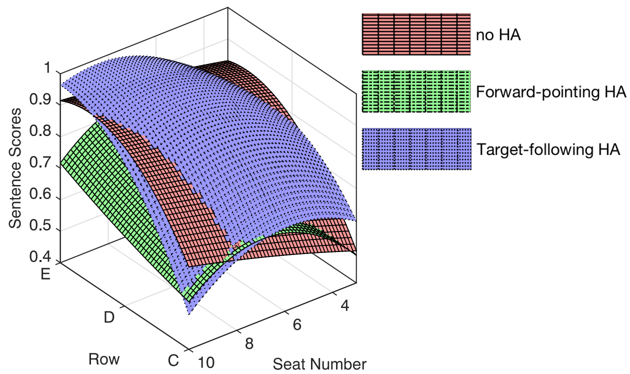
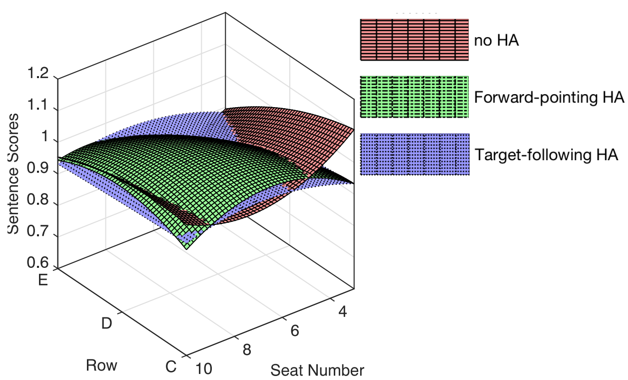


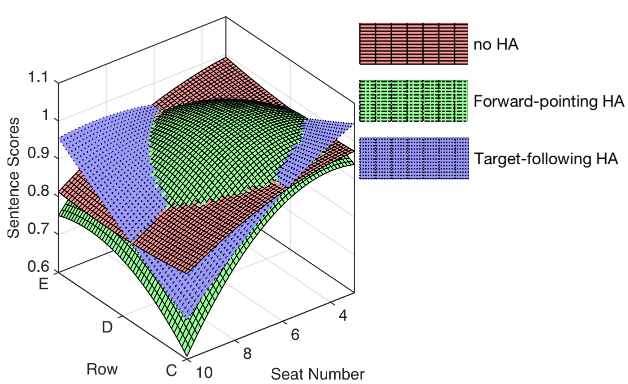
Figure 4.5: Sentence scores by HA type for all speaker locations.



(A) Sentence Scores by Seat : Back Speaker



(B) Sentence Scores by Seat : Front Speaker



(C) Sentence Scores by Seat : Side Speaker

Figure 4.6: Sentence scores by seat location and HA type shown for speech coming from the back speaker (A), front speaker (B), and side speaker (C).

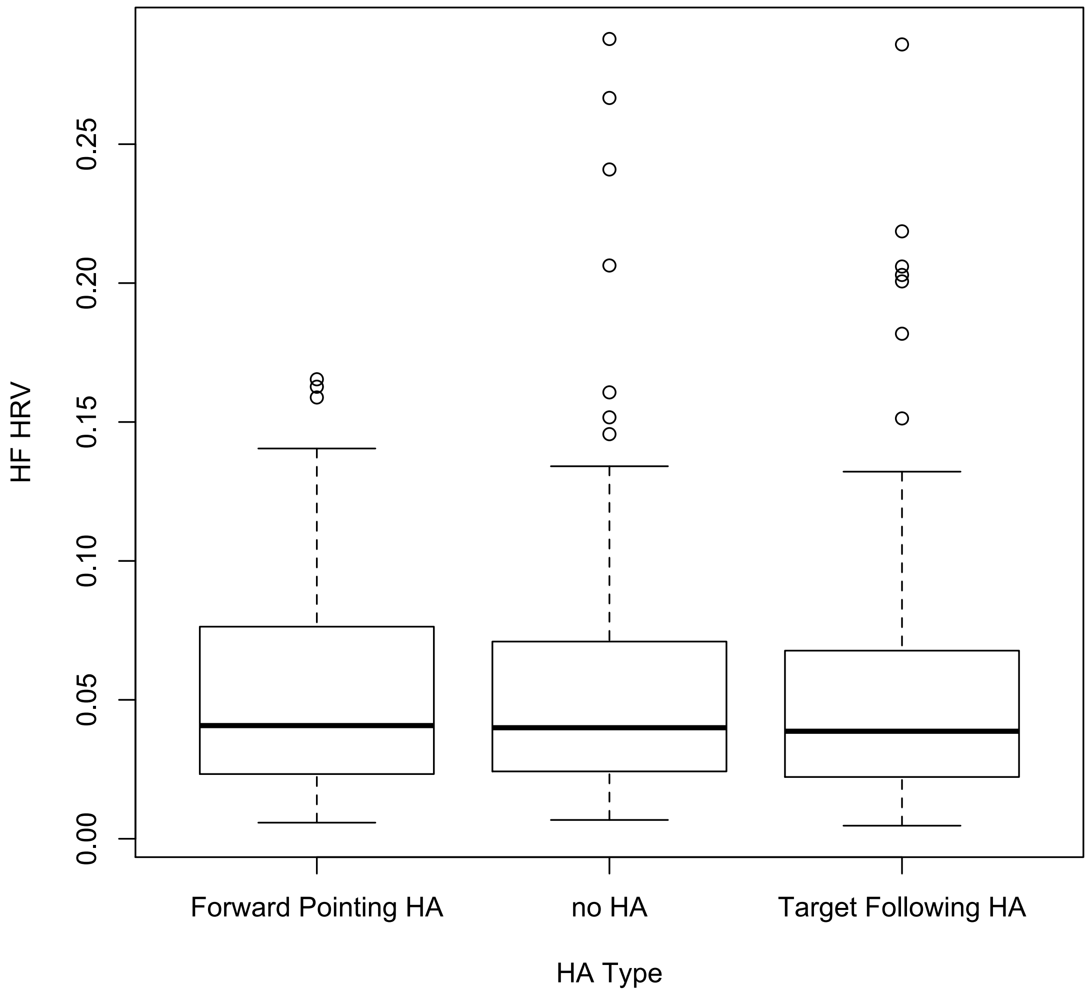


Figure 4.7: High Frequency HRV by HA Type.

4.3.4 EEG Measures

ANCOVAs were performed for each EEG measure with between subject factor of seat row and seat number, and within subject factors of reverb, background noise, HA type, speaker location, and the interaction between reverb and background noise. Likely due to intersubject and intrasubject variability, EEG alpha (8-13 Hz), beta (13-25 Hz), and gamma (25-100 Hz) band power all had no significant main effects. Only the Normalized OLE had a main effect that was close to significant.

Normalized OLE

The close to significant main effect found for phase distribution (normOLE) was by seat row, as shown in Fig. 4.8.

4.3.5 Correlation between measures

Apart from between the two ECG measures, there were no significant correlations between measures consistent across participants.

4.4 Discussion

From subjective ratings shown in Fig. 4.5, the HA conditions did produce significantly different listening effort. Given that these are normal hearing subjects, it is not surprising that the no HA and target following HA condition do not show a difference in listening effort since the listeners ability to focus on their better ear should give a similar benefit to the hearing aid directionality for speech from the front and side speaker conditions. The main difference for the forward pointing HA is likely driven by the cases where speech was coming from the back speaker. While in many scenarios a talker is in front of the person listening to them, there are certainly many realistic scenarios where this is not true. For instance, in a car or when seated beside someone at a table you may not be able to be directly facing the person who is speaking.

Although the listening effort ratings were not compared for each speaker location, speech intelligibility was. Given some previously found correlations between speech intelligibility, speech transmission index, and listening effort in Picou et al. (2017); Sato et al. (2012) it would not be unreasonable to speculate that it would show a similar effect of speaker location, depending on how participants interpreted the listening effort rating question. It may in fact even be more sensitive to the changes than intelligibility in scenarios such as this study where intelligibility is already high. This is something essential to investigate in follow up studies.

In Fig. 4.6 the pattern of benefit for the target following hearing aid for different directions and SNRs can be seen. For speech coming from the front, there is little

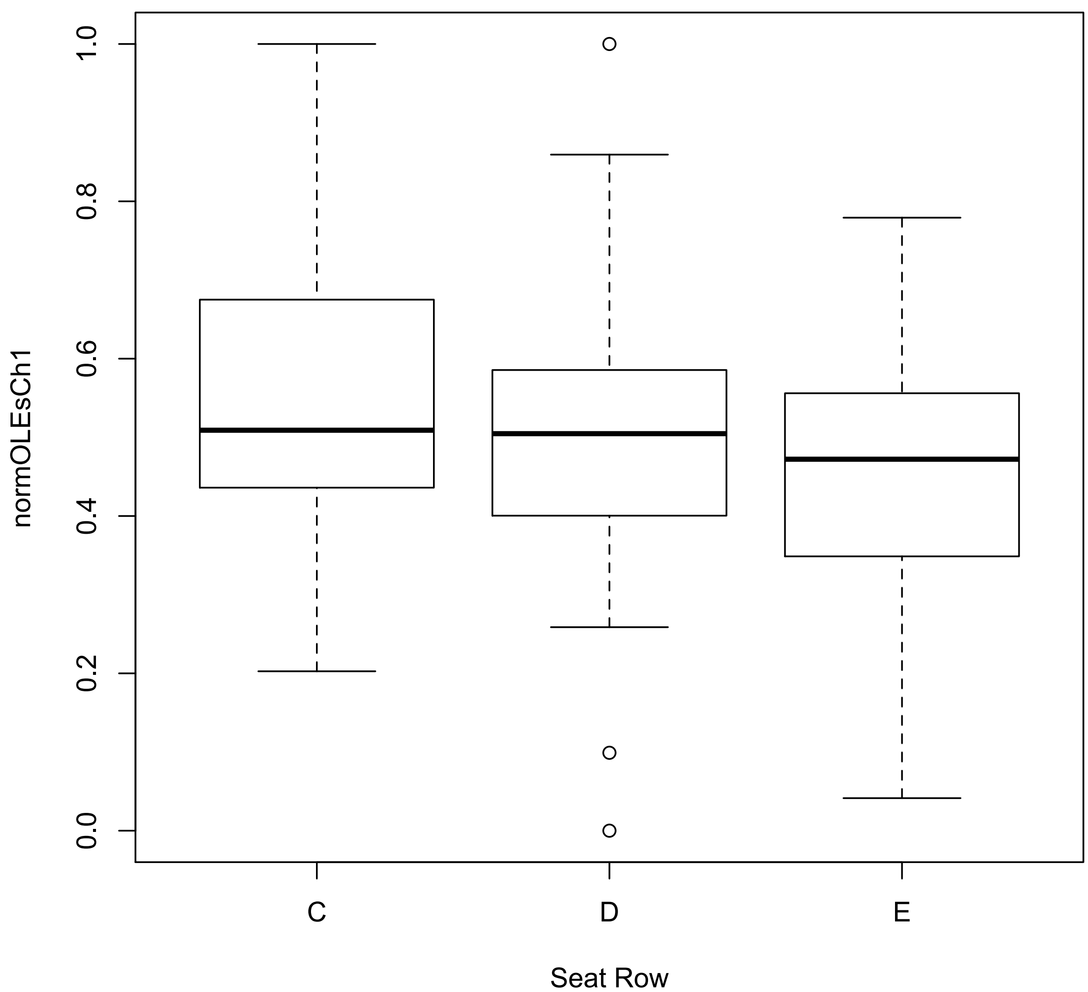


Figure 4.8: Normalized OLE by seat row.

difference between the three listening conditions and little difference based on where the subjects were seated. This is as predicted, since for this direction the target following HA, forward pointing HA, and no HA condition all have directional or no processing that would be steered towards the front. For the side speaker, the effect of SNR can start to be seen. For those seated closest to the speaker the three conditions have similar speech intelligibility which is indicating that the SNR in these seats is such that any additional processing is providing little or no benefit. For those seated further from the speaker, particularly those in the back row E, you can see an increased speech intelligibility for the target following HA compared to the other two conditions. For those seated further away and in front of the speaker, the no HA condition provides the highest speech intelligibility. While the target following HA does steer directional processing to behind the user, it does not steer backwards as strongly as forwards so this could explain the no HA condition being most intelligible here. For the forward pointing HA there is a clear pattern of decreasing speech intelligibility as SNR decreases (as the seat is further from the speaker). This effect of changing SNR or speech transmission index due to background noise does show the same increases in listening effort as Visentin and Prodi (2017); Rennies et al. (2014); Picou et al. (2019); Sato et al. (2012). Seat number and the interaction between seat number and HA type did also show a significant main effect for HF HRV, although as noted this was possibly caused by outliers in certain seats. It is possible that with within subject comparison across different seat locations for ECG and EEG measures this would prove to be a significant effect for these listening effort measurements as well. The lack of EEG baseline values for band power analysis is also a potential limitation of this study, although with within subject comparisons it is possible band power would show changes in listening effort within an individual. Phase distribution of the EEG, which was normalized for each subject, did not show significant main effects either so it is also possible that the average SNR here was such that the physiological listening effort metrics are not sensitive enough to measure differences, something also found in Cvijanović et al. (2017). The goal in this study was to keep to a realistic range of SNRs, which is why more extreme values with low speech intelligibility were not chosen.

4.5 Conclusions

Focusing solely on speech coming from the front for listening effort or speech intelligibility studies may not be sufficient to show differences in HA processing algorithms, as seen in Fig. 4.6. The difference in HA algorithms can be seen here as SNR changes based on seat location for the side and back speaker, even as the overall speech intelligibility appears to have reached a saturation point. Secondaly, within subject comparison is essential for physiological measures of listening effort, as individual variability in these measures can be large enough to drive a main effect based on outlier subjects.

The speech intelligibility scores do appear to reach a saturation point in this experiment, something that with HI listeners would be expected to change. It would also be expected that with HI listeners the range of LE would be greater, and thus possibly more sensitive to changes that would affect LE.

Chapter 5

Preliminary studies on assistive listening and music processing for HAs

5.1 Methods

The following experiments were approved by the McMaster University Research Ethics Board, MREB 1975.

5.1.1 LIVELab Concert for the hearing impaired

The first stage of the assistive listening and music processing for HAs study was a public concert in the LIVELab. Hearing aid users were invited to attend the concert and take part in the assistive listening study. The goal was to test different assistive listening technology for accessibility and preference, but also to build a database of hearing impaired listeners interested in music studies for future experiments.

Two string quartet concerts were held as part of the Neuromusic conference held at McMaster University with HA users invited to attend the concert and give feedback on the available assistive listening technologies. 39 HA users and 4 CI users attended and data was collected on which technologies and programs they had on their HAs. This was done to determine which technologies might have the best application for future assistive listening concerts. The goal was to test several assistive listening technologies to see if they were beneficial for a live music performance, and to determine what sort of environmental reverberation HA users prefer. The technologies tested in these concerts were an induction loop, a Sennheiser Mobileconnect system with 4 different channels of audio, and wireless headphones connected to an FM system. The four channels on the Sennheiser Mobileconnect had different gain attack and release times

Table 5.1: Sennheiser Mobileconnect Automatic Gain Control settings

Channel	Automatic Gain Control	Attack Time (ms)	Release Time (ms)
1	On	50	3000
2	On	10	200
3	On	30	1000
4	Off	N/A	N/A

to determine if there was a preference among the HA users, shown in Table 5.1.

Each instrument in the quartet was miced directly on the instrument and the audio was used for the assistive listening feeds. The induction loop was a mono audio feed while the wireless headphones and Sennheiser Mobileconnect were stereo feeds. The induction loop feed was equalized to adjust for the low and high frequency attenuation in the HA processing. The Sennheiser Mobileconnect feed processing was done using the Sennheiser Mobileconnect automatic gain control, as shown in Table 5.1 for the different channels.

5.1.2 HA processing for live music

Following the concert for the hearing impaired a follow up study was performed during a MITACS internship with Unitron. The goal of this project was to test changes to the hearing aid processing such as the amount of gain at low and high frequencies, the amount of compression, and compression attack and release times, with the goal of improving the processing for music. Through initial listening tests, increasing the gain at low and high frequencies was thought to have a positive effect for listening to live music. Further testing was done in the LIVELab with seven hearing aid (HA) users testing the modified music program compared to the default Unitron music program and the default conversation in quiet program.

The gain changes made to the modified music program were based on the difference in frequency spectrum power between music and speech, shown in Fig. 5.1. Recordings from two string quartet concerts in the LIVELab in November designed for people with hearing aids were compared to a set of 3000 HINT sentence recordings with 6 male and 6 female speakers. The music had less energy in the low frequency (<150 Hz) and high frequency (>7000 Hz), but more in the mid frequency range.

Hearing aid gain prescription formulas typically aim to maximize speech intelligibility by using amplification to correct for hearing impairment. One linear amplification formula, NAL, tries to maximize speech intelligibility by maintaining comfortable loudness across all frequencies, taking into account the frequency spectrum of speech (Dillon, 2012). Another linear gain program, DSL, attempts to reach an optimal sensation level for each frequency band (Dillon, 2012). This equal loudness or sensation

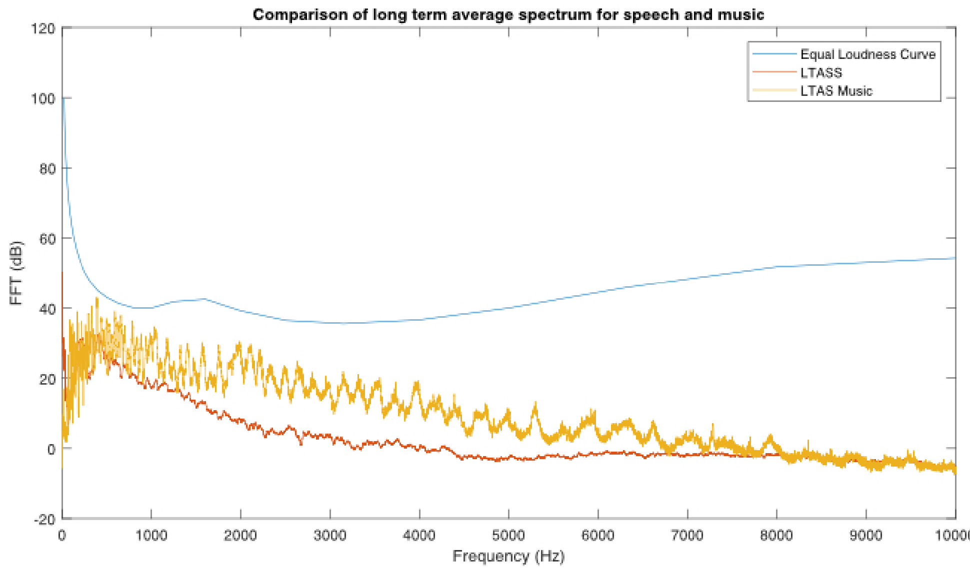


Figure 5.1: Comparison of the long term average spectrum for speech (red) and music (yellow).

level across all frequencies is desirable for music to keep the balance between low and high frequency instruments, so the conclusion was that increasing the gain at high and low frequencies and decreasing it at mid frequencies compared to what is designed for speech would improve the balance of music.

The factors selected for the subjects to rate for each HA program were based on previous studies that considered naturalness, pleasantness, and volume of music, but also to determine if the gain changes are beneficial for all types of instruments, and for hearing individual instrument parts. Statistical analysis was done on the ratings for each program but also which program subjects rated highest for each factor, since this may be more revealing than one overall preference.

5.2 Results

5.2.1 LIVELab Concert for the hearing impaired

Prior to the concert begining, particpants filled out a survey about the technology that they had on their HAs and whether they found them useful. Results are shown in Fig. 5.2. This was done to give an idea of what technologies were accessible to most HA users using the devices they already had.

A post concert survey was sent to participants to determine which technologies

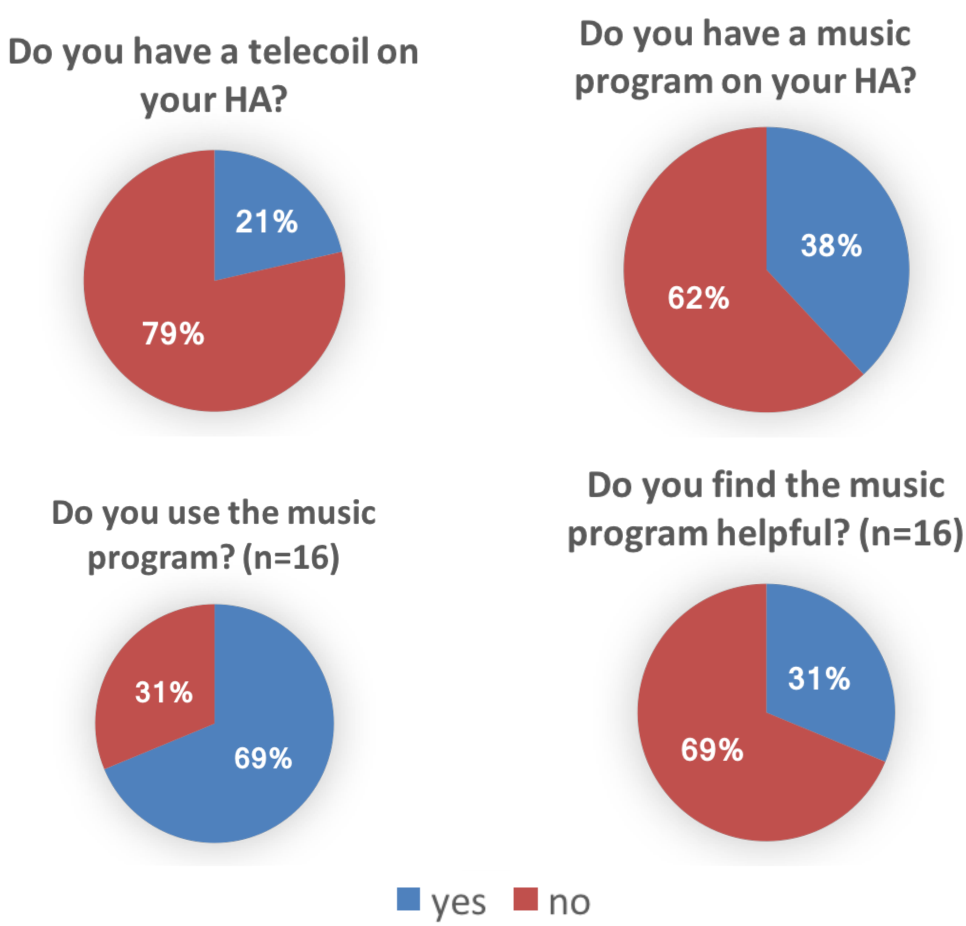


Figure 5.2: Pre concert survey results to determine what technology participants had and used on their HAs.

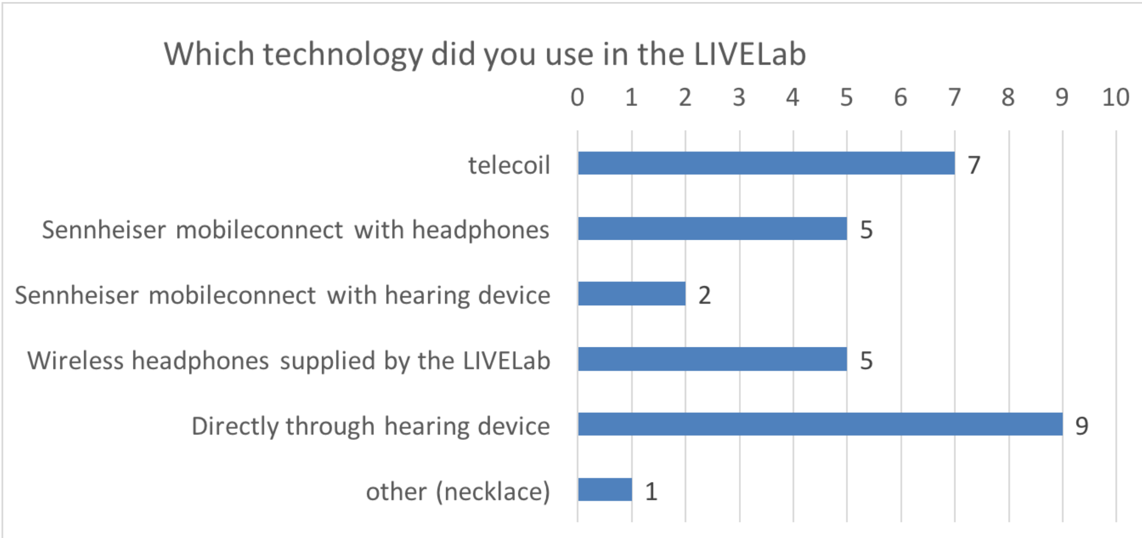


Figure 5.3: Technology used by participants at the LIVELab Concert for the hearing impaired.

they used during the concert, and which they felt were best and worst for listening to music. Survey results are shown in Fig. 5.3 and Fig. 5.4.

5.2.2 HA processing for live music

The number of participants in this experiment was small (n = 7), however when looking at the percentage of times where each program was rated as highest (or closest to neutral for loudness), the modified music program was rated highest more than chance after excluding tied ratings. Results are shown in Fig. 5.5.

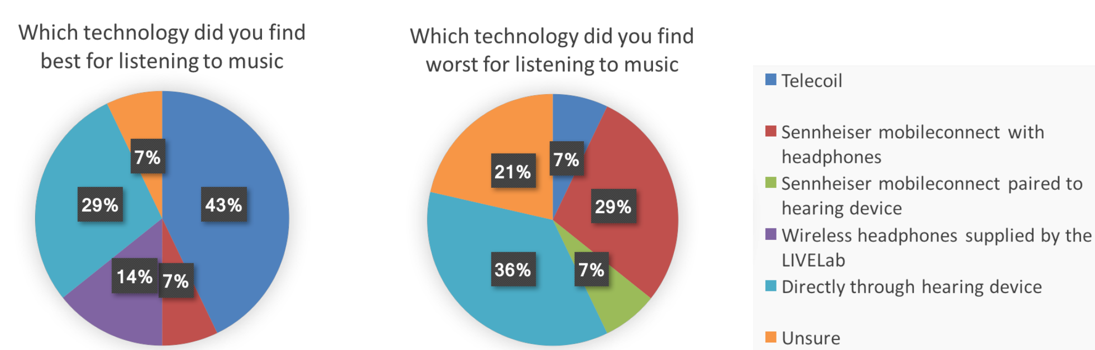


Figure 5.4: Post concert survey results on which technology participants selected as best or worst for listening to the live music.

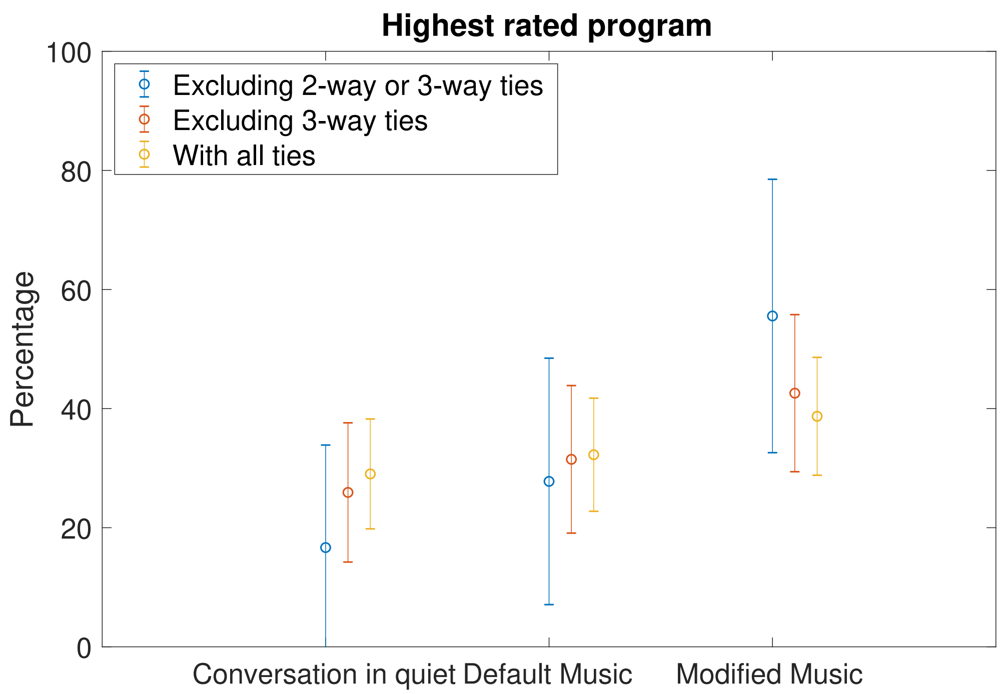


Figure 5.5: Percentage of times where each program was rated the highest by a participant, including or excluding tied ratings as noted.

Chapter 6

HPO assistive listening concert

6.1 Methods

At the Hamilton Philharmonic Orchestra's performance of Mahler's Fifth Symphony on May 11, 2019, at the First Ontario Place in Hamilton, Ontario, an experiment was conducted to investigate the usefulness of assistive listening systems for HA users during a live orchestra performance. The following experiment was approved by the McMaster University Research Ethics Board, MREB 1975.

For hard of hearing individuals to benefit from the social interaction of a live musical experience as an audience member, it is essential to deliver amplification through hearing aids that makes the music audible and clear. Unfortunately, hearing aids that work reasonably well for speech often provide poor quality for live music. With our collaboration with Hamilton Philharmonic Orchestra we are examining the benefits of assistive listening for improving sound quality in a large scale concert venue.

47 subjects participated in the experiment. They were divided into three technology groups according to what hearing aids they used:

- TV Connector (Unitron HAs) (n=15)
- Telecoil (Phonak HAs) (n=15)
- Own HAs (n=17)

The Unitron and Phonak participants were fitted with hearing aids by Unitron and Phonak audiologists at least 16 days prior to the concert so they could get used to them. The hearing aids contained programs to enable the use of the assistive listening systems during the concert. Within each of the Unitron and Phonak groups, the hearing aids were identical. In addition to getting sound as usual through the air to their hearing aids, the Unitron users were given additional sound directly to their hearing aids through a TVConnector and will be referred to as the TVConnector group. In addition to getting sounds as usual through the air to their hearing aids, the Phonak HA users were given additional sound directly to their hearing aids through a telecoil loop and will be referred to as the Telecoil Loop group. The group wearing their own HAs served as a control group. All participants had symmetrical moderate to severe hearing loss.

Eight days prior to the concert, participants came to the LIVELab for a practice session and also filled out the Goldsmith Musical Sophistication questionnaire (Müllensiefen et al., 2014) and the Abbreviated Profile of Hearing Aid Benefit (Cox and Alexander, 1995).

Microphones were placed amongst the musicians on stage during the performance (Figure 6.1) to use for the TVConnector and Telecoil Loop Assistive Listening Feeds. The location of the participants' seats and the assistive listening systems are shown in Figure 6.2. The feed mixes from the microphones were designed by a sound designer who listened to output sounds from a KEMAR recording manikin (Figure 6.3) wearing loop-enabled hearing aids during a rehearsal of Mahler's Fifth. This enabled the creation of sound feeds that were expected to sound good through hearing aids.

Four different Assistive Listening Feed enhancements were created:

- Proximal (using microphones close to the musicians: microphone numbers 11-20 in Figure 6.1)
- Proximal with string enhancement
- Distal (including microphones farther from the musicians: microphone numbers 1-10 and 21-23 in Figure 1)
- Distal with string enhancement
- No sound enhancement feed

During the performance of Mahler's Fifth, the feed changed every 5 minutes and participants were prompted on tablets to rate the sound quality and loudness for the previous 5 minutes (Figure 6.4). Each condition repeated 3 times.

While we did not control for when in the performance each feed was presented, the feeds repeated in a different order and an average rating was used to help reduce the bias based on what part of the piece was being played during which feeds.

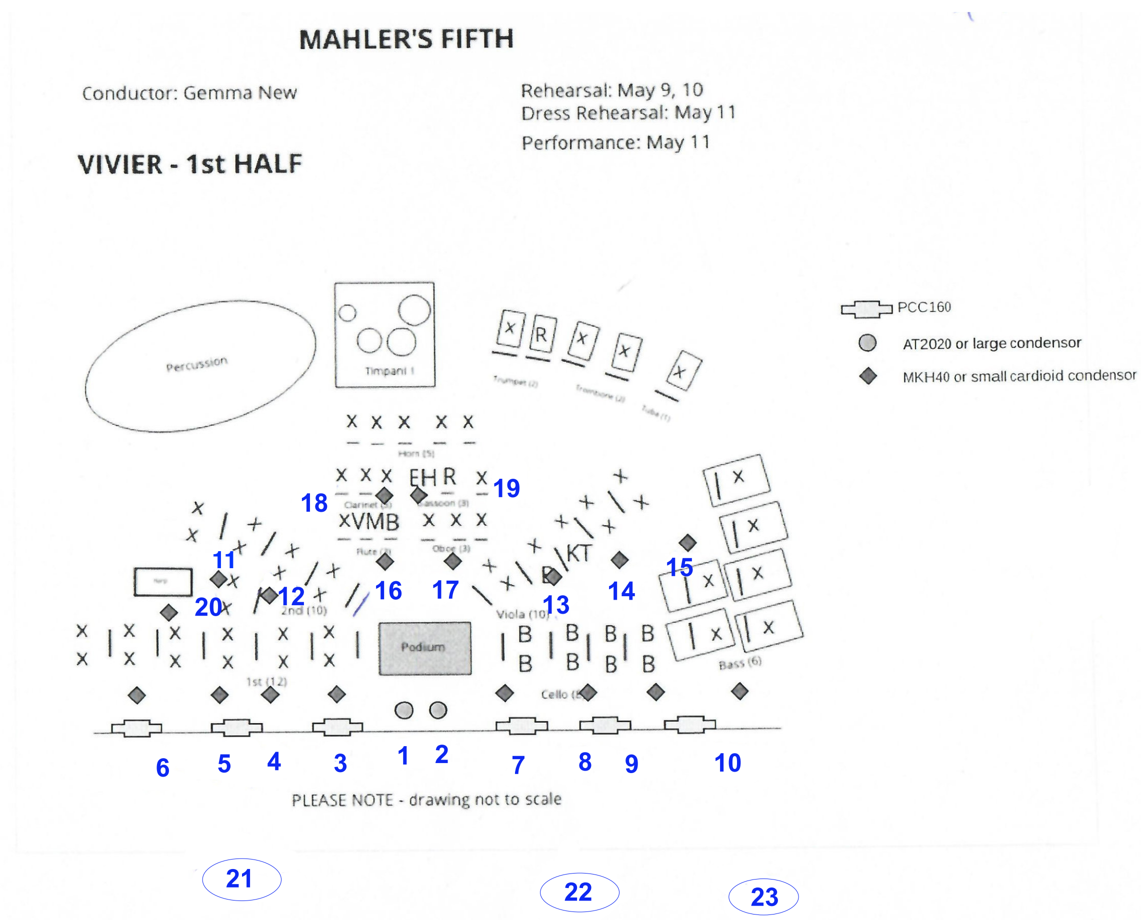


Figure 6.1: Musician and microphone layout. Blue numbers indicate a microphone location. Microphones 21-23 are located on the balcony face.

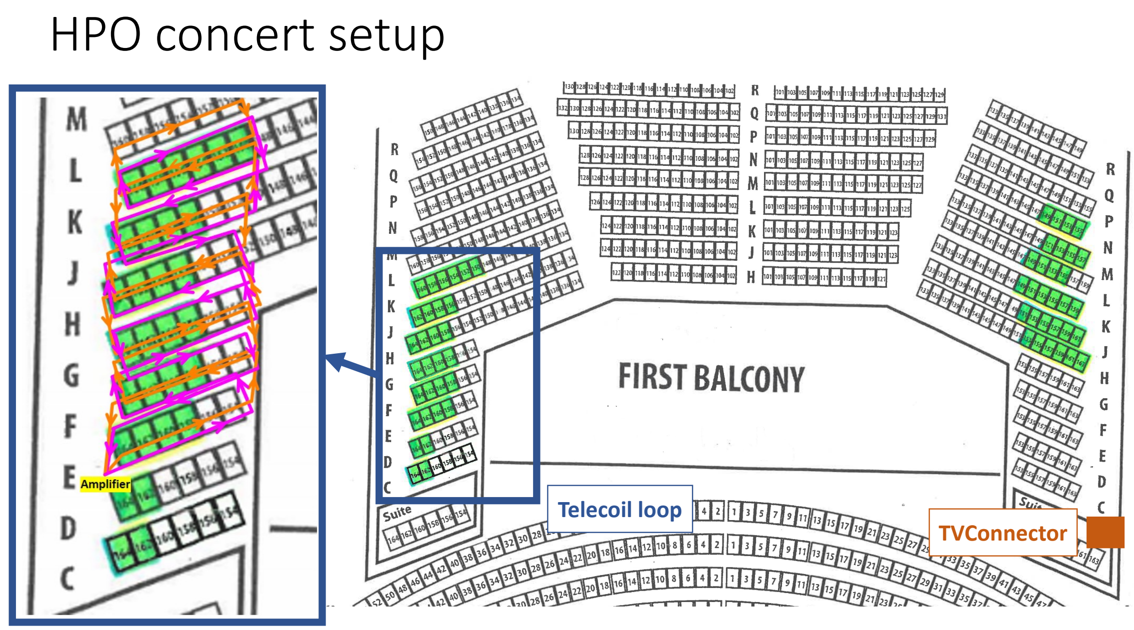


Figure 6.2: Assistive listening system setup for the HPO concert. Highlighted seats indicate where participants were seated, with the location and layout of the TVConnector and telecoil loop shown in blue and orange respectively. The layout of the amplifier and double loop is shown in the expanded view.

Figure 6.3: KEMAR manikin used for recording and feed setup, with researcher Larissa Taylor.

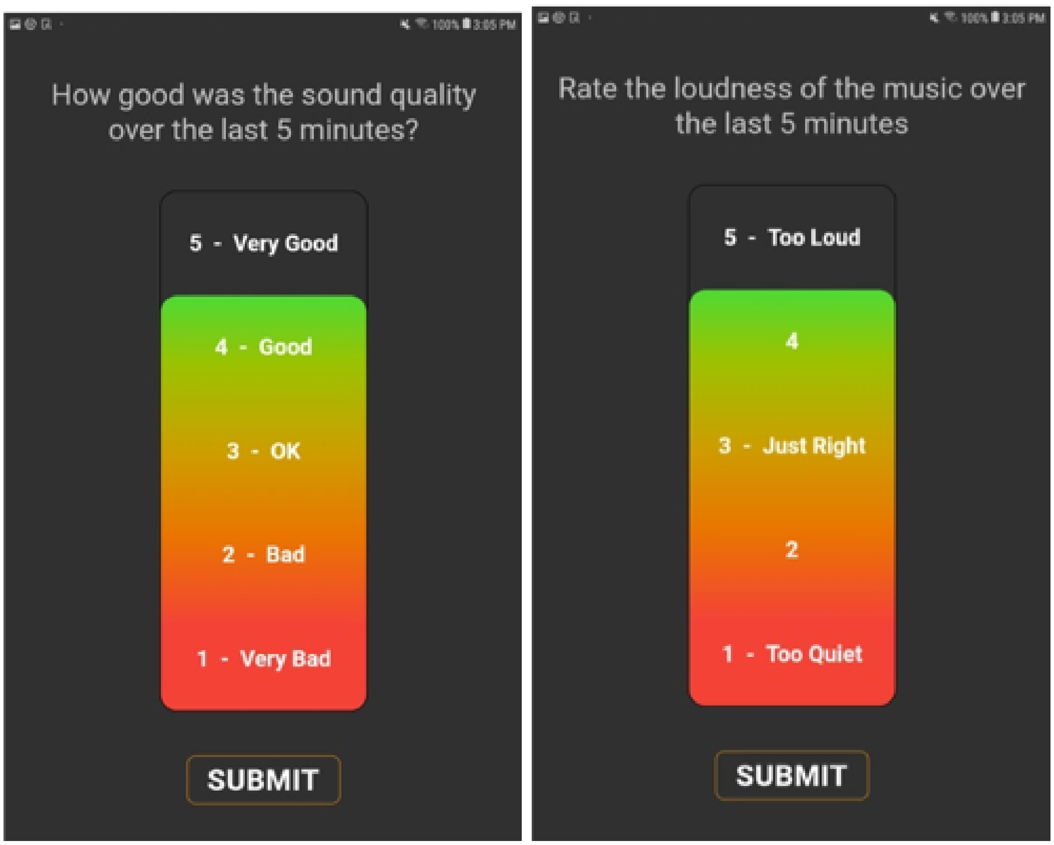


Figure 6.4: Tablet questions during the performance

6.2 Results

6.2.1 Anecdotal responses

Overall the response to the hearing aids and assistive listening setup for the concert were very positive. Several participants asked about purchasing the Unitron and Phonak hearing aids used during the experiment or having their own HAs set up with the same fitting parameters. Several participants commented that the HAs they were fitted with for the concert were better than their current devices. Several participants also commented that they could hear individual instruments better using the provided hearing aids and assistive listening; one participant was pleasantly surprised when they were able to hear even the piccolo clearly.

6.2.2 Participant characteristics

Participants were sorted into groups for analysis purposes based on their responses to the musical sophistication and HA benefit questionnaires, average hearing loss, and technology listening group. Musical sophistication scores were divided into three groups based on 33 and 66 percentiles of the scores (high, medium, or low), HA benefit and three frequency (0.5, 1, and 2 kHz) average hearing loss were divided into two groups again based on percentiles (high or low). The 50th percentile for average hearing loss was 49.5 dB HL. Percentiles were calculated before eliminating participants based on missing data. Numbers in each grouping are given in Table 6.1, with the numbers for each technology group given in Table 6.2 and Table 6.3.Participants in the two assistive listening groups were randomly assigned to either loop or TV Connector. The exception was the preidentified music experts, who were divided evenly between the two assistive listening groups. Participants missing musical sophistication or HA benefit data, or who did not have at least one response for each of the feed conditions were excluded from further analysis.

Table 6.1: Subject group numbers

	Low	Medium	High
Musical Sophistication (Goldsmith)	13	14	13
HA Benefit	19		21
Average Hearing Loss Group	21		19
	Own device (control)	Telecoil loop	TVConnector
Technology group	13	14	13

Table 6.2: Subject group numbers - Telecoil loop

	Low	Medium	High
Musical Sophistication (Goldsmith)	2	8	4
HA Benefit	7		7
Average Hearing Loss Group	7		7

Table 6.3: Subject group numbers - TVConnector

	Low	Medium	High
Musical Sophistication (Goldsmith)	5	4	4
HA Benefit	7		6
Average Hearing Loss Group	7		6

6.2.3 Tablet sound quality ratings

Across all times a response was requested across all participants, but some responses were missing. Existing responses for all trials for each Assistive Listening Feed type for each participant were averaged to yield five ratings per participant. An ANOVA was performed using the statistical program R with Assistive Listening Feed (5 Types, defined above) as a within subject factor, and technology group (TVConnector, Loop), musical sophistication group (low, medium, high), hearing loss group (low, high), and HA benefit group (low, high) as covariate factors. The control group wearing their own HAs were not included in this analysis since they did not experience the five different Assistive Listening Feeds. Musical sophistication had a significant main effect ($p = 0.011$). No other factors or interactions were significant. Figure 6.5 shows the sound quality ratings for each feed condition. See 6.2.4 for more details on the effect of musical sophistication on sound quality ratings.

A linear contrast comparing each participant's sound quality rating for no feed to their scores for the four feed conditions (with a weight of -4 for the no feed, and 1 for each of the feed conditions) was performed. The goal was to examine if ratings when there was a feed were higher than when there was no feed (for TVConnector and loop participants), that is, if average ratings for feed present minus average ratings for feed absent was significantly greater than zero. An ANOVA was performed to see if any of the factors had a significant effect on the difference in ratings. There were no significant main effects from any of the factors included: hearing loss group, musical sophistication group, HA benefit group, and technology group.

The control group (using their own hearing aids) rated all the conditions very similarly with difference ratings ranging between -7 and 4 and their difference scores did not differ significantly from zero ($p = 0.47$), which would be expected because they could not receive the different feeds with their own hearing aids (Figure 6.6).

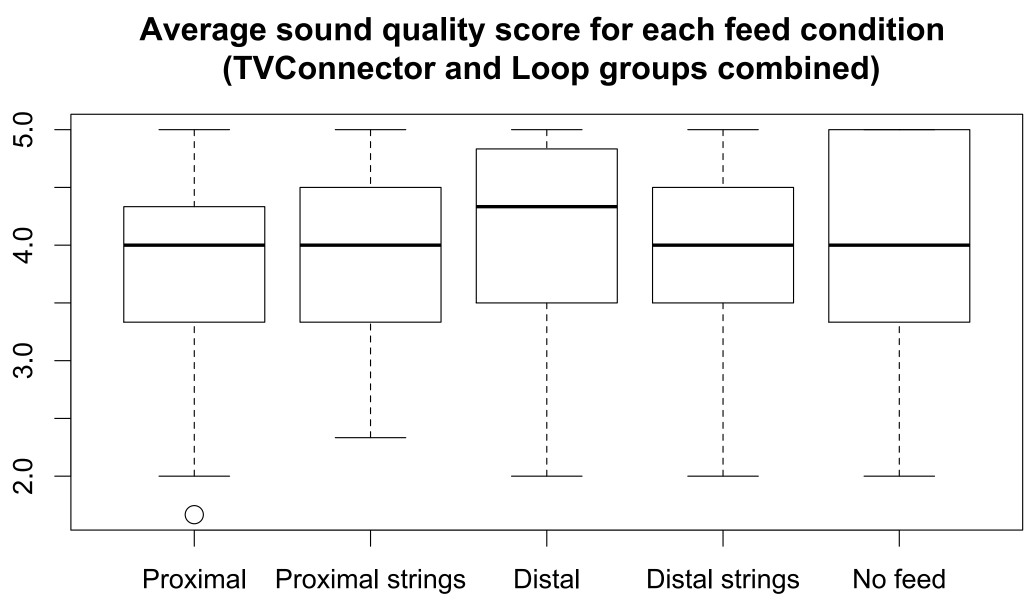


Figure 6.5: Sound quality ratings for each feed condition. Horizontal lines represent quartiles.

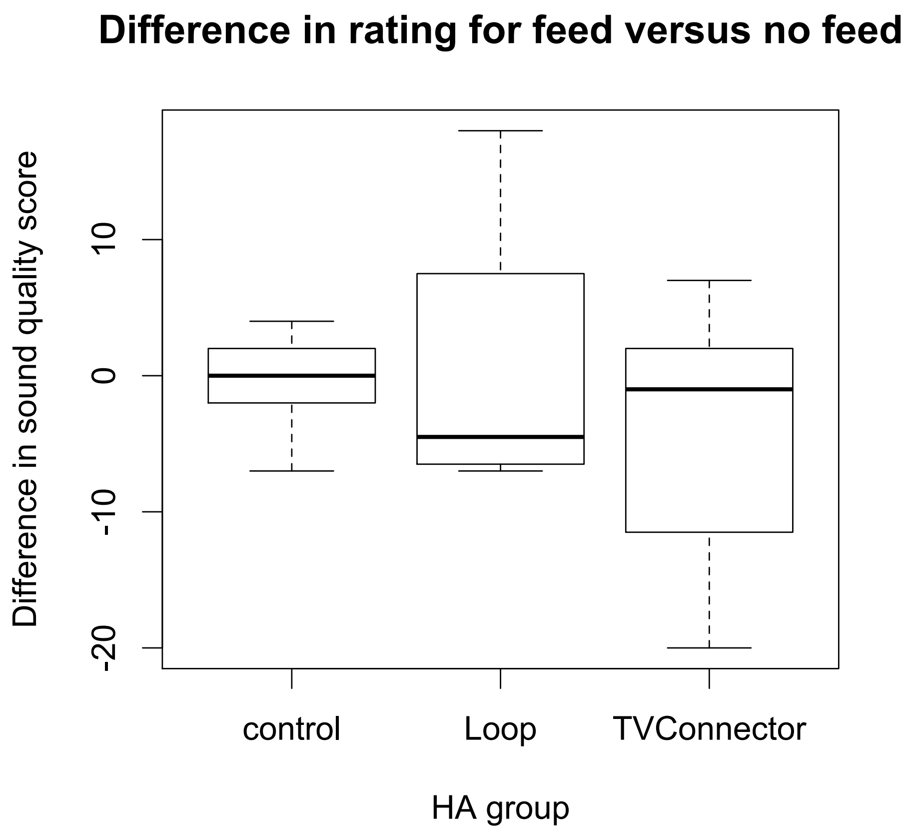


Figure 6.6: Difference in sound quality rating for feed vs no feed. The control group was not included in the trend analysis but is shown for reference. Horizontal lines represent quartiles.

The TVConnector group does look to have mainly negative scores(rating the feed conditions as having lower sound quality compared to no feed), but a one-sided t-test shows that it is not significantly less than zero (p=0.12). This outcome may be partly due to the TVConnector feed being fairly quiet, so there is not as much input to the hearing aid from the assistive listening feeds, leading to the music being mainly heard from the hearing aid microphone. The TVConnector was primarily designed for use in a home with a TV, which is why it is not able to be as loud as the telecoil loop which is mainly used in public spaces.

In conclusion, there does not appear to be a consistent trend in how sound quality was rated with an Assistive Listening Feed present or absent. However, it can be seen in Figures 6.5 and 6.6 that the variance across participants was very large. Thus, in the next sections we explore how differences in musical sophistication affected participants' ratings.

6.2.4 Effects of Musical Sophistication on Assistive Listening Sound Feed Ratings

Making a single rating of sound quality based on the last five minutes during a performance of Mahler's Fifth Symphony might be a difficult task, especially for listeners with little musical experience. Therefore, we were interested in examining how the group with the highest music sophistication scores rated the different feeds. Musical Sophistication had a significant main effect on overall ratings (p=0.011), with the higher music sophistication group rating sound quality lower than the medium and low music sophistication groups (see Figure 6.7). There also appears to be a ceiling effect happening in the low and medium music sophistication group.

Figure 6.8 shows that the high musical sophistication group rated the feed conditions higher than the no feed condition to a greater extent than the low or medium musical sophistication groups.

To better illustrate the differences in ratings by musical sophistication group and score, the interaction between musical sophistication group and feed is shown in Figure 6.9 Only those with high musical sophistication rated the No Feed condition lower than the others, though this pattern was not significant. The linear trend of average sound quality rating and musical sophistication is shown in Figure 6.10.

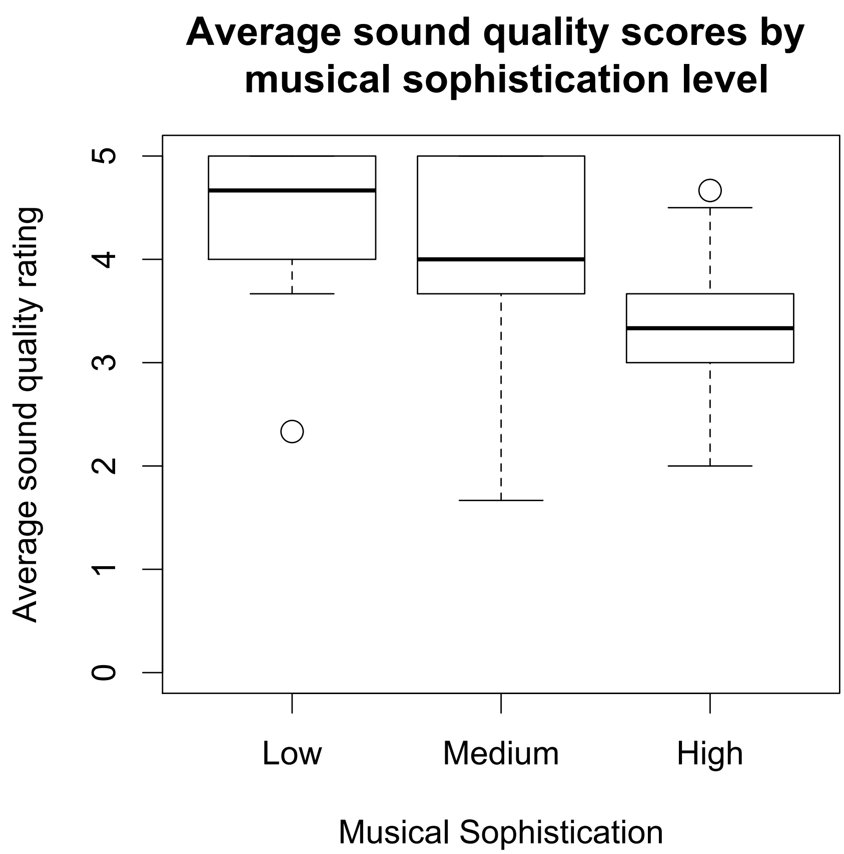


Figure 6.7: Average sound quality ratings by musical sophistication. Error bars represent quartiles.

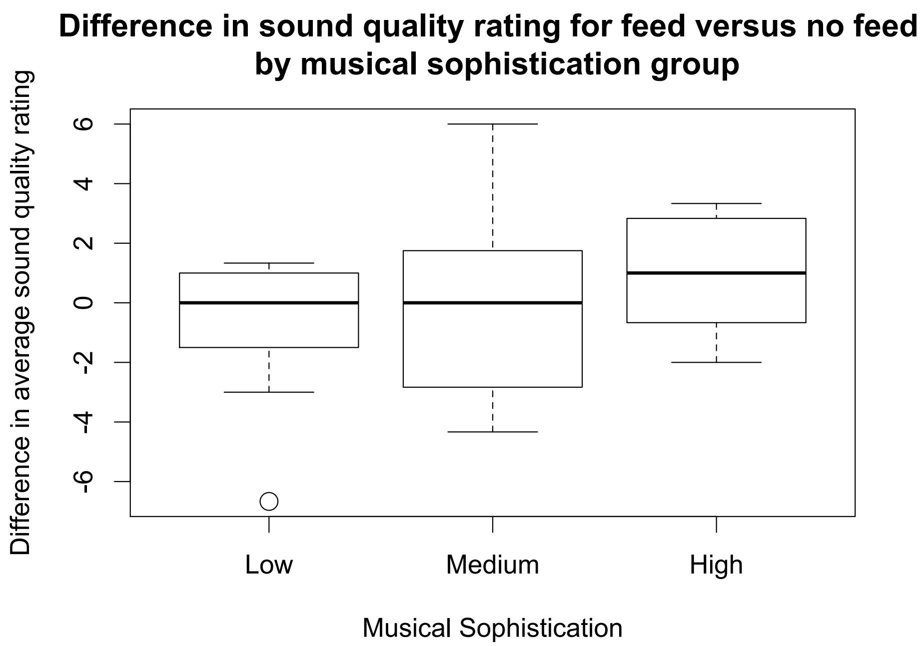


Figure 6.8: Difference in sound quality rating for feed vs no feed by musical sophistication. Error bars represent quartiles.

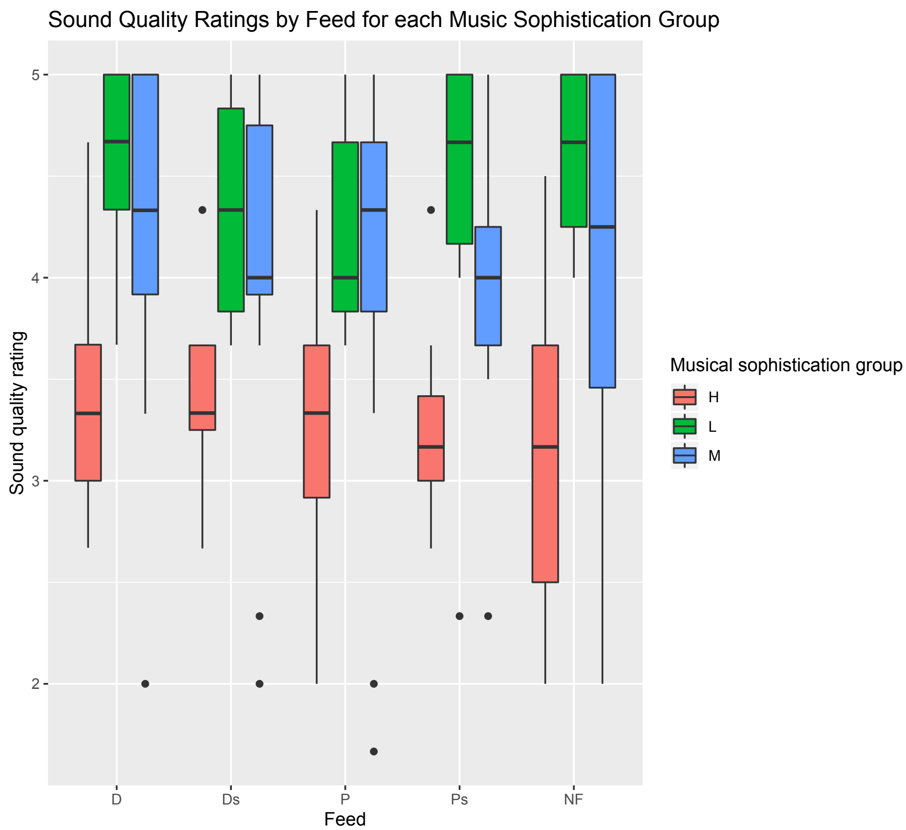


Figure 6.9: Interaction between musical sophistication and feed. D – Distal, Ds – Distal strings, P – Proximal, Ps – Proximal strings, NF – no feed

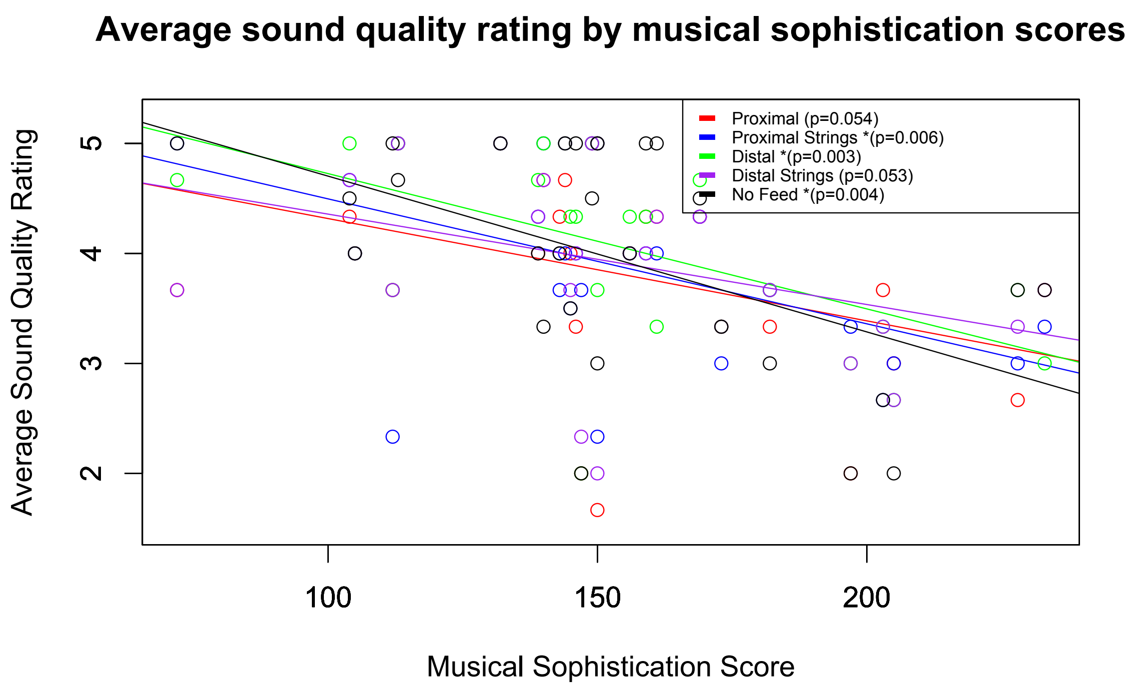


Figure 6.10: Linear trend of average sound quality rating by musical sophistication score for each feed condition

All groups appear to rate the distal feeds higher than the proximal ones. From participant feedback at the concert this could be due to the desire for the feed to sound like the audio the participant should be hearing based on where they are sitting, which is closer to the location of the distal microphones on the balcony.

High Musical Sophistication Group

To further examine the pattern of ratings within the high music sophistication group, a separate ANOVA was performed on just that group with feed as a within subject factor, and HL group, HA benefit group, and technology group as factors. Hearing loss group was close to having a significant effect ($p = 0.08$). This indicates that participants may have different opinions on the feeds based on the degree of their hearing loss. Figure 6.11 shows this interaction for all participants.

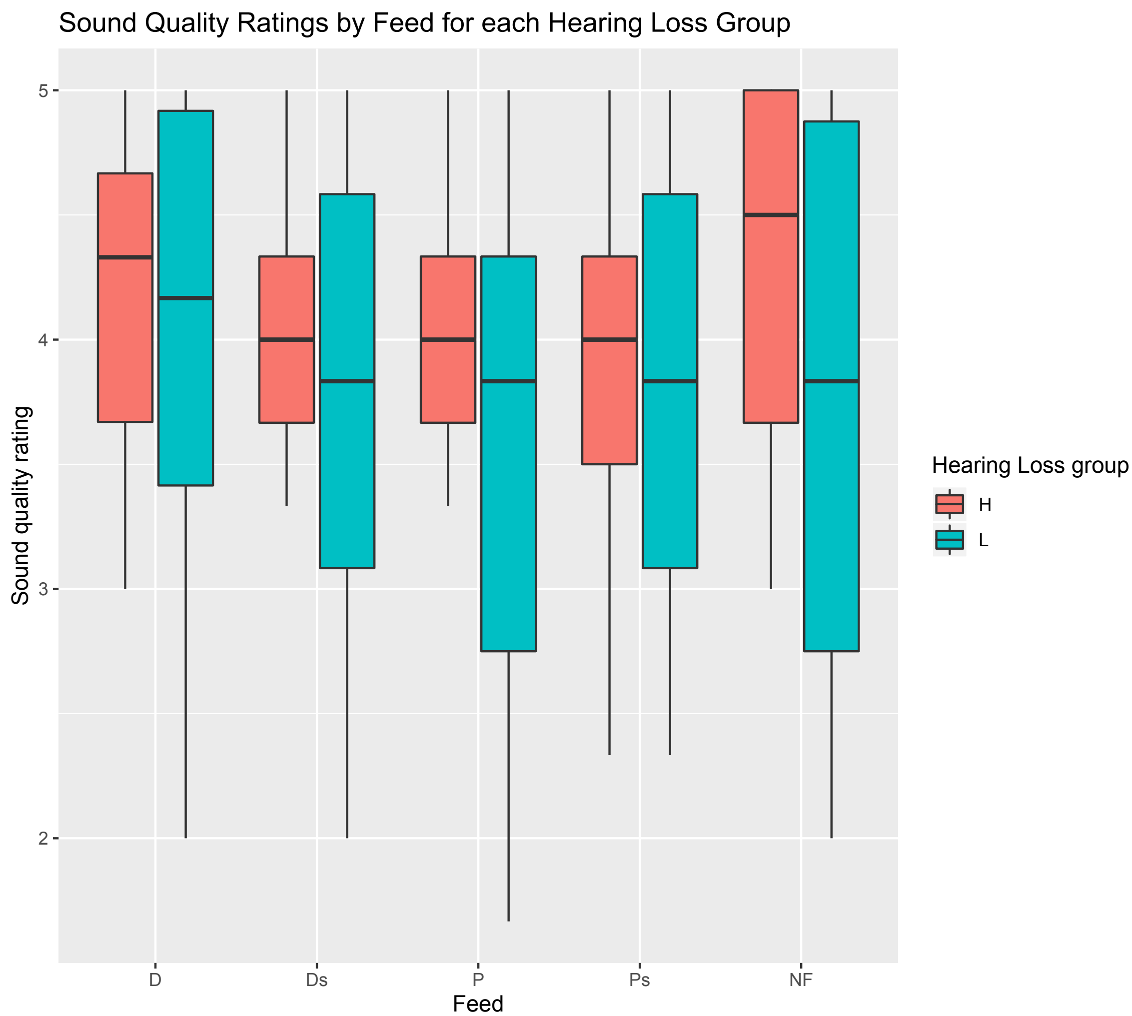


Figure 6.11: Interaction between HL and feed for sound quality ratings, loop and TVConnector groups. D – Distal, Ds – Distal strings, P – Proximal, Ps – Proximal strings, NF – no feed

6.2.5 Loudness Ratings

Participants also rated the loudness for each five minute interval. Averages for each participant were calculated the same as for sound quality. An ANCOVA with musical sophistication group, hearing loss group, HA benefit group, and technology group as covariates and feed as a within subject factor. HA benefit group ($p = 0.04$) and the interaction between feed and musical sophistication ($p = 0.041$ after adjustment for sphericity) were significant. The loudness ratings by hearing aid benefit group are shown in Figure 6.12.

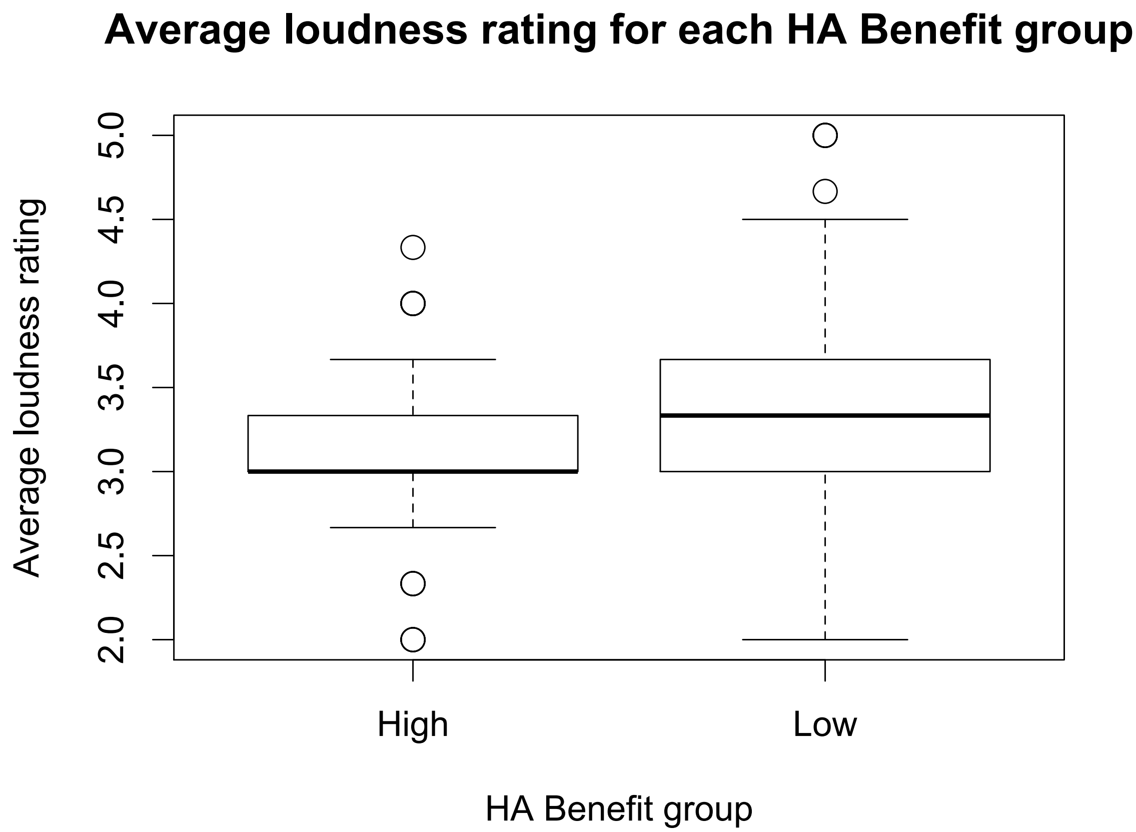


Figure 6.12: Loudness rating by hearing aid benefit group.

The interaction between feed and musical sophistication group is shown in Figure 6.13. It shows that the high musical sophistication group very consistently rated the feeds equally loud, whereas the other two groups had more variation in their ratings.

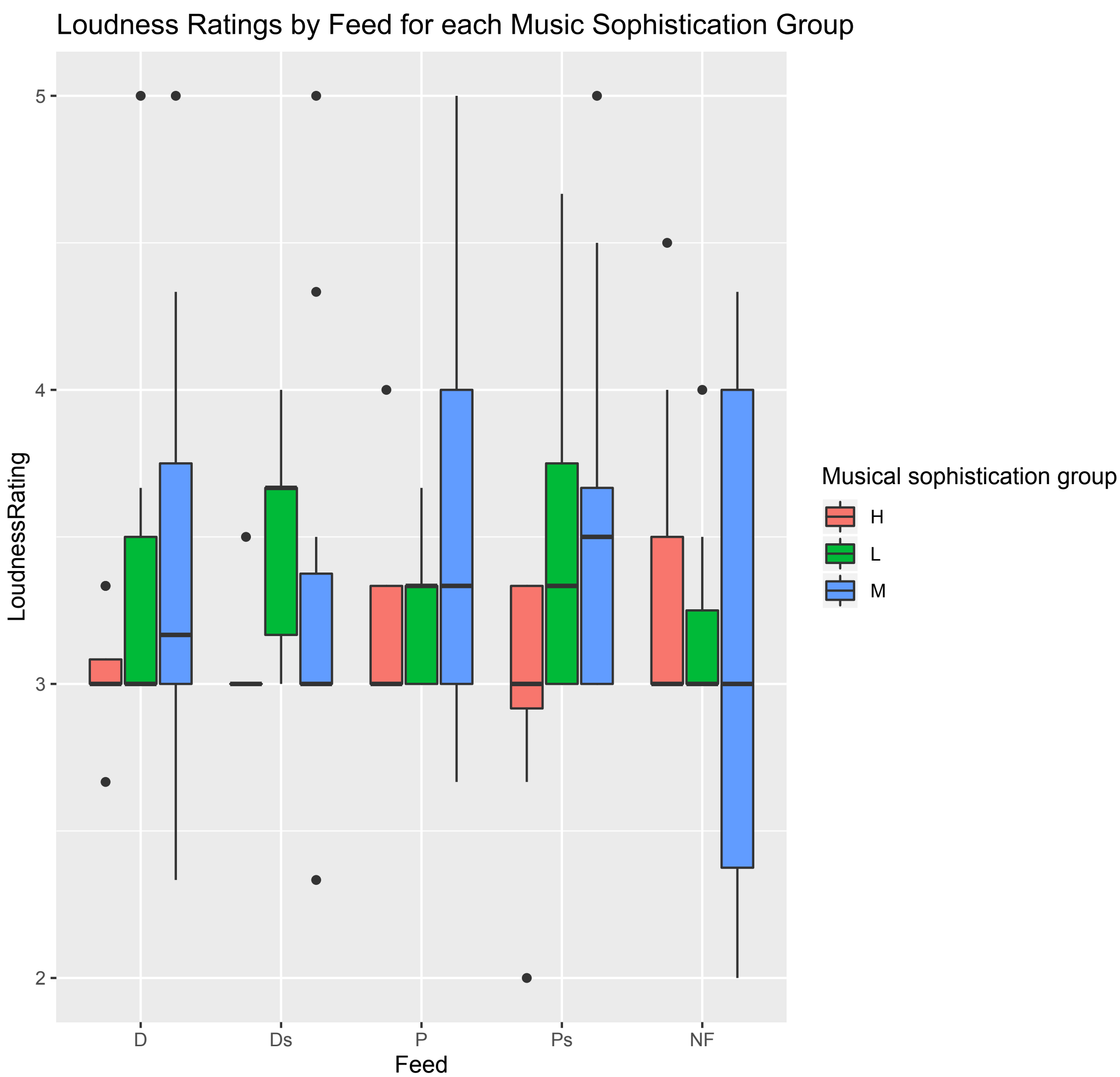


Figure 6.13: Interaction between feed and musical sophistication on loudness ratings.

Chapter 7

General Discussion and Conclusions

7.1 Discussion

While realistic HA studies for both speech and music have challenges and additional considerations compared to individual studies in sound booths, in Chapters 3–6 there have been significant effects shown in large scale, ecologically valid studies. These studies were more complex than typical listening effort studies performed in a sound booth, with multi directional noise, reverberation, and more demanding tasks than just single sentences or words to recognize. Particularly from withing subject measures, the effects of background noise levels and SNR have to be shown to have significant effects on listening effort, as would be expected. Baseline values and subject variability particularly make physiological measures of listening effort less consistent in showing listening effort changes but some effects were still found. It could also be true that in this case the HRV measures were not sensitive enough for the range of background noise and reverberation used in these studies, similar to results in Cvijanović et al. (2017). Similarly, no significant effects were found in EEG measures of listening effort which could also be due to the relatively small size of this effect. Previous studies have used high and low effort conditions where the worse condition is at an SNR where word recognition is 50%. The speech intelligibility in the studies here was much higher, and closer to a realistic SNR one might experience regularly. It is also possible that since these are normal hearing listeners with no previous experience with HA processing that this had an effect on the listening effort results, since they would need to adjust to the unfamiliar processing. This is particularly true in the live performance experiment since we did not have within subject comparisons between HA types, so two groups had this unfamiliar processing while the no HA group did not. There is also some subjectivity in how a person perceives what effort is that could also partly explain the varied ratings of listening effort in Chapters 3 and 4.

Music is a subjective art form, and this leads to great variability in how people rate sound quality. Those with musical training do appear to be more selective and consistent in rating sound quality of loop mixes from the HPO concert results in Chapter 6. There were limitations on how much the different loop conditions could be compared due to the experiment being done during a live concert, but anecdotal feedback as well as results from the high musical sophistication group do point towards improved sound quality with the telecoil loop.

7.2 Future Directions

7.2.1 Follow up listening effort studies in the LIVELab

Given the results in Chapters 3 and 4, follow up listening effort studies with HI listeners wearing HAs in the LIVELab are planned using a modified procedure from the studies with normal hearing listeners. For both, the no HA condition would be replaced with omnidirectional HA processing in order to provide the amplification required for HI listeners in all conditions.

Firstly, to do a follow up study similar to the study with live actors on stage a more consistent level of speech from the stage is required. This could be done by pretraining the actors beforehand to keep a consistent speaking volume in the changing background noise and reverberation and for all acts, or by having a performance more geared towards a consistent delivery such as acts of a play. To allow for within subject comparison of all conditions, enough acts would need to occur so that all subjects experienced all hearing aid conditions.

For a follow up study for the experiment with recorded sentences played from different directions more minor modification would be made. First, the SNR of -5 dB may be too difficult for some HI listeners so this may need to be adjusted based on pilot trials. The method of playing the sentences and speaker location would remain the same. Rather than rating subjective effort after each block of the experiment, it would be rated for each sentence written down for speech intelligibility measurement. Alternatively, the listening effort rating could be done continuously using the tablets the same as in the first experiment. HI participants would need to have symmetrical HL to ensure the assumption about processing being symmetrical for left and right holds. In order to have additional within subject comparison of the different seat locations, all or a subset of participants would redo the experiment sitting in several different seat locations. This could also be done for the experiment with actors on stage, although the same issues of repeatability with live actors maintaining a consistent speaking level would apply for any repetition of this type of study.

7.2.2 Music quality follow up study

As a follow up to the HPO sound quality study discussed in Chapters 5 and 6, a more controlled and specific comparison between different loop feeds and processing strategies has been designed. This is based on a reference and hidden anchor listening test (Völker et al., 2016). A loop has been installed in the LIVELab for use in this and future experiments. Using KEMAR and a selection of excerpts from the HPO concert experiment discussed in Chapters 5 and 6 a MuSRHA listening test will be created for each participant. Different combinations of the on stage microphone recordings will be combined, along with different telecoil loop programs on the HA, to create the different processing conditions for the listening test. The goal is to determine if there are best combinations of loop and microphone signal on the HA for a live concert, and if different stage microphone combinations make better quality audio for loop listening during a concert.

Based on the results from the initial HPO study and follow up studies, a follow up concert experiment should be done to confirm results in a realistic concert environment. Results from this should be useful for providing guidelines for the setup of telecoil loops and HA telecoil programs for live music listening in concert halls or similar environments.

7.3 Conclusions

This series of studies did show significant results in both listening effort experiments, and shows the possibility for follow up studies for additional results in HI listeners with directional HA processing. Within subject comparisons gave more significant results, but even with the challenges of a live performance listening task there were significant effects on listening effort ratings that would be expected from background noise and reverberation. It was also possible to see effects of SNR on different directional processing algorithms on speech intelligibility.

The music concert studies show less consistent results, but do indicate some key factors to consider when recruiting subjects for future music sound quality studies. Musical sophistication was a key factor in how critical and consistent sound quality ratings were in the HPO study in Chapter 6.

Appendix A

CI Hackathon project

A.1 Introduction

This has been included as an appendix rather than a full chapter. The project is small and related to but not part of the main studies discussed in the thesis chapters. Larissa Taylor took the lead role in the Hackathon, and designed and coded all signal processing changes made to the base cochlear implant code provided. Other team mebers included Micheal Wirtzfeld, Kekio Guttierez, Melih Yahli, Brendan Tao and Daniel Shields. Michael performed the speech quality and intelligibility tests on the processed audio from proposed algorithms. Other team members provided subjective feedback on the processed music samples for each potential processing algorithm.

The purpose of the CI Hackathon run by UCSF, University of Minnesota, and Advanced Bionics was for competing teams to use the supplied starting CI code to implement processing strategies that would improve sound quality for four categories of audio stimuli: CNC words, speech, speech in noise, and music. Basic cochlear implant processing code was provided as a starting point, along with a library of audio stimuli for testing and developing new algorithms. This included basic filtering, AGC, envelope and frequency estimation and mapping and stimlulation generation to generate an output spike pattern for the cochlear implant electrodes. A cochlear implant vocoder was also provided to allow for listening and other tests on the output from the cochlear implant processing.

Due to the complexity of the project and the short timeline, our team decided to focus on a strategy that would improve sound quality for music. Our approach was to optimize the filtering supplied in the CI base code for music by making it more broadband, and to reduce the complexity of the input signal. The complexity reduction was chosen as a processing strategy based on Nagathil et al. (2015), which revealed a preference for complexity reduction in terms of melody clarity and ease of listening. This study was done with CI users listening to preprocessed audio, so the goal here was to implement into the cochlear implant processing itself.

A.2 Filtering

Two filters were designed to be used in series for the CI processing. A high pass filter with a cutoff frequency of 100 Hz, and a low pass filter with a cutoff frequency of 7 kHz. Filter magnitude and phase response are shown in Fig. A.1 The high pass cutoff was selected to increase the amount of low frequency content in the signal used for simulation mapping in the CI. More low frequency signal was desired because rhythm is the element of music most perceptible for CI users (Looi et al., 2008; McDermott, 2004), and this is normally contained in the lower frequency part of music. The low pass filter was designed to reduce the excess brightness in the music after CI processing. Both filters were kept to a low order of 2 to avoid any excess time delay from input to simulation. The final filter cutoff frequencies were selected based on subjective feedback from blind listening tests.

A.3 Reducing complexity

Nagathil et al. (2015) suggests that reducing the complexity of music may be preferred for CI listeners, so this was implemented using a function in the code dedicated to reducing the complexity of the signal after initial filtering and AGC. The input to the function were the signal to be reduced in complexity and an n value for the PCA. To reduce the complexity of the signal, the same method from Nagathil et al. (2015) was used. First, the constant Q transform of the signal was calculated using Matlab cqt(). The covariance matrix of the constant Q transform is then calculated. The n largest eigenvalues and corresponding eigenvectors of this covariance matrix are used to reconstruct the signal with an inverse constant Q transform. The selected n value for complexity reduction was optimized based on absolute error and relative energy error between the original and reconstructed signals for the given music stimulus library. The tradeoff curve is shown in Fig. A.2 and Fig. A.2. The n value marked in Fig. A.2 was used for all processing going forward.

A.4 Testing

Several combinations of filter designs, PCA parameters , and gain and compression parameters were tested to optimize the final processing strategy, shown in Table A.1. This was done using a blind listening test as well as speech quality and intelligibility metrics. The speech metrics used were Perceptual Evaluation of Speech Quality

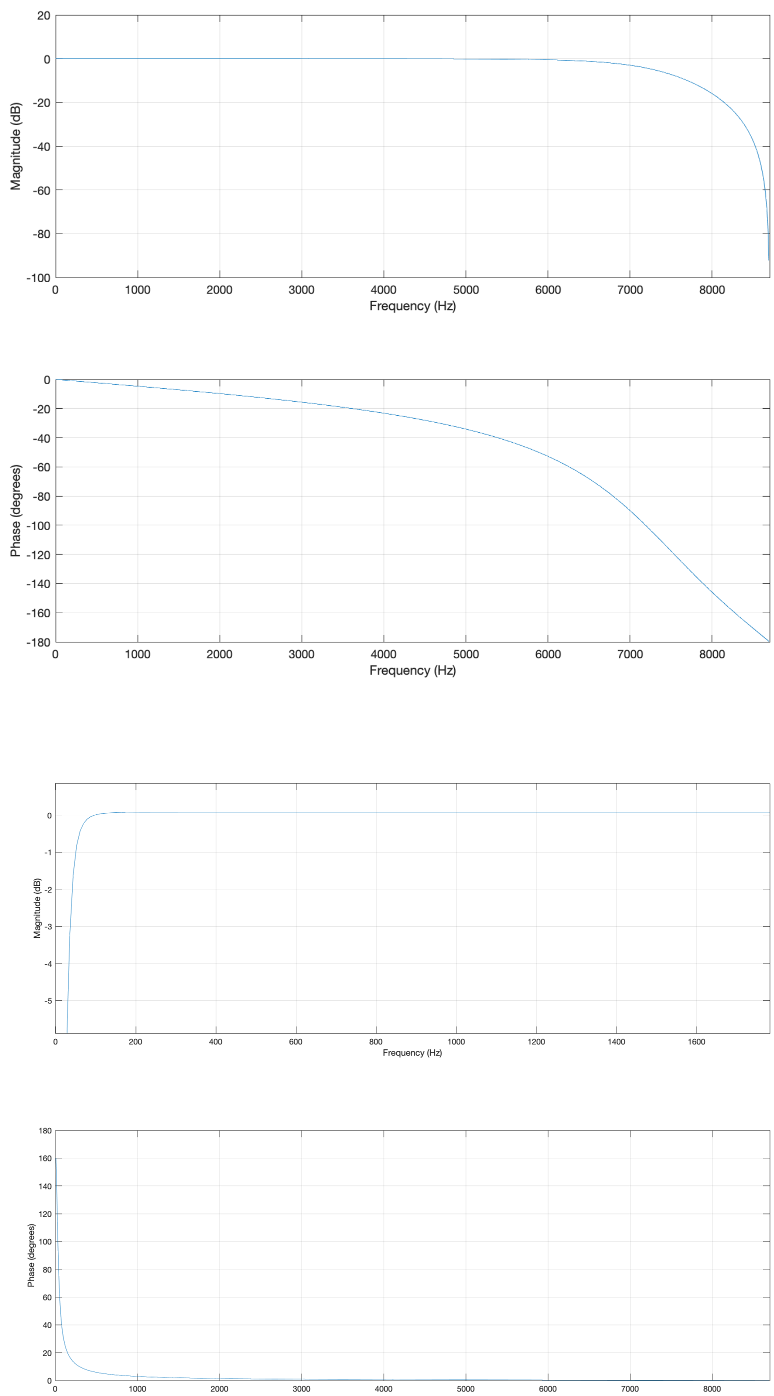
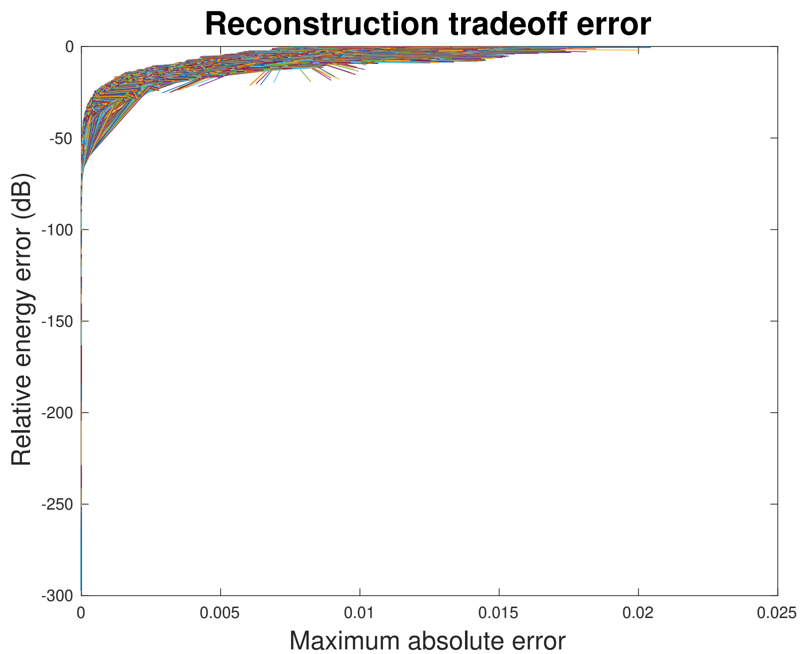
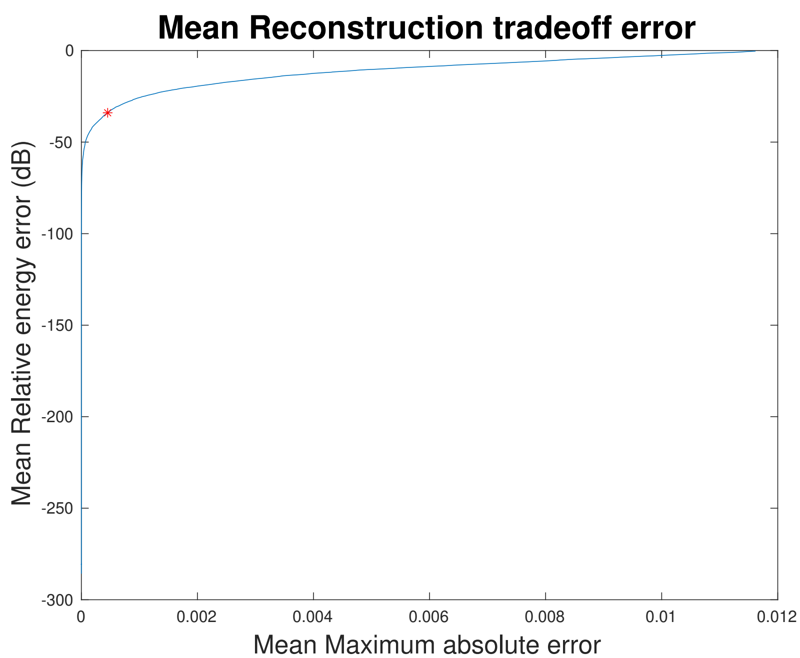


Figure A.1: Filter magnitude and phase response for lowpass and highpass filters designed for CI processing.



(A)



(B)

Figure A.2: Tradeoff for reconstruction using PCA for different n values. (A) Tradeoff curves shown for all provided music samples. (B) Mean PCA reconstruction errors for different n values. Red * marks errors for selected n value. n = 173

(Narrowband and Wideband Implementations) and Short-time Objective Intelligibility (STOI). The results are shown in Fig. A.3

Table A.1: Filter and complexity reduction parameters for test cases of CI processing

Case	HP cutoff	LP cutoff	g0	compRatio	n
A	100		6	12	
B	100	3000	0	12	
C	2500		0	12	
D	100	7000	0	12	
E	100	7000	0	1	
F	100	7000	0	6	
G	100	7000	1.661	6	
H	100	7000	1.661	1	
I	100	7000	1.661	1.5	
J	100	7000	0	12	236
K	100	7000	0	6	236
L	2500		0	12	236
M	100	7000	0	12	168
N	100	7000	0	6	168
O	100		0	12	168

A.5 Conclusions

The reduced complexity and filtering did appear to be subjectively preferred compared to the default processing from the initial code. This was done using a vocoder however, so further listening tests with CI users would be required to confirm this preference. The strategy did also beat the gold stardard Advanced Bionics algorithm in judging by other hackathon participants for music and also some of the speech categories, so it does not appear reducing the complexity of the signal has a negative impact on speech quality. Due to the reduced complexity of the signal there is also less spiking activity on the cochlear implant electrodes in the electrodogram, which could potentially lead to less fatigue listening with this processing strategy compared to the default processing. For comparison, the two electrodograms are shown in Fig. A.4.

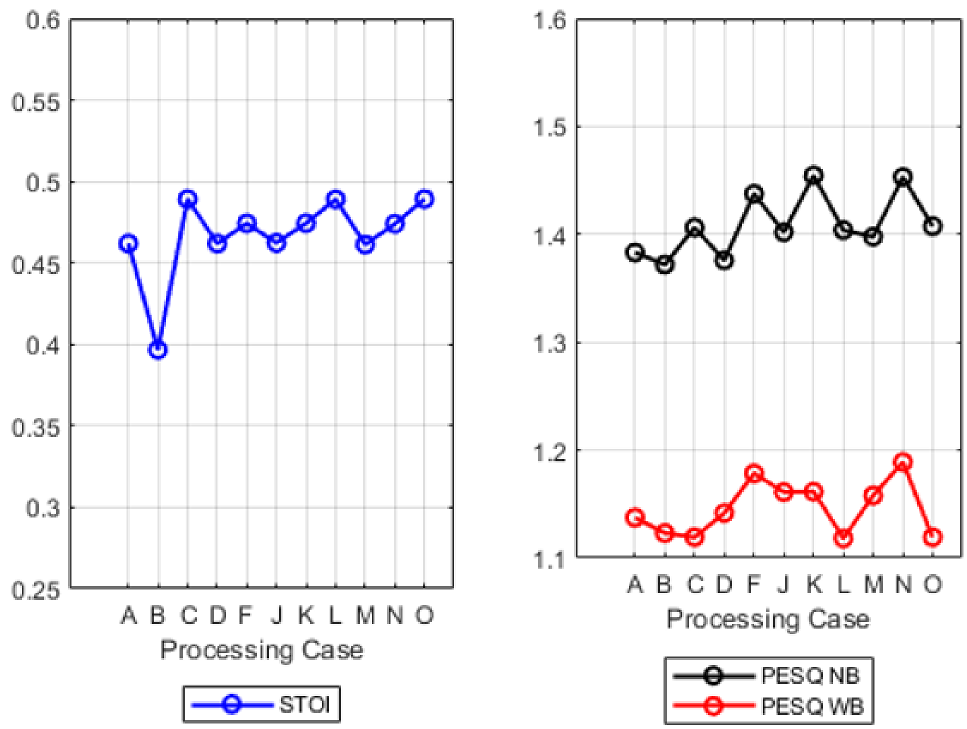
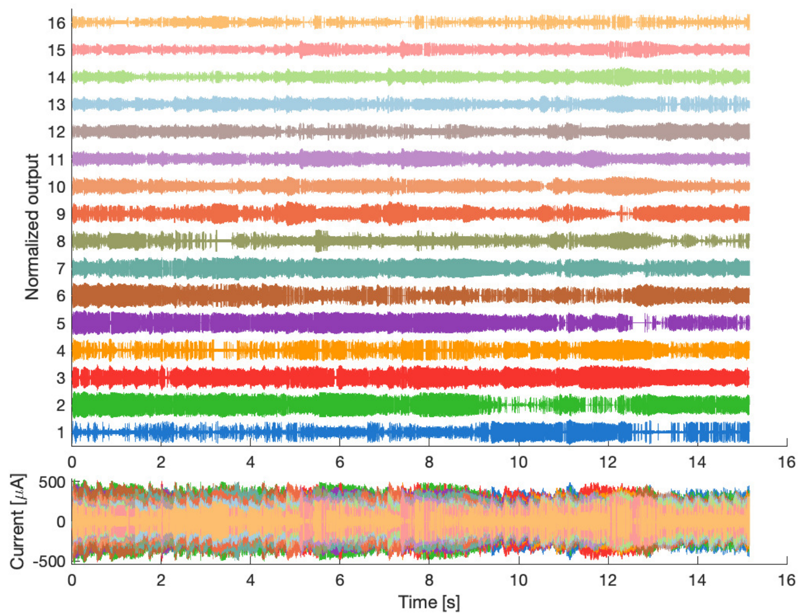
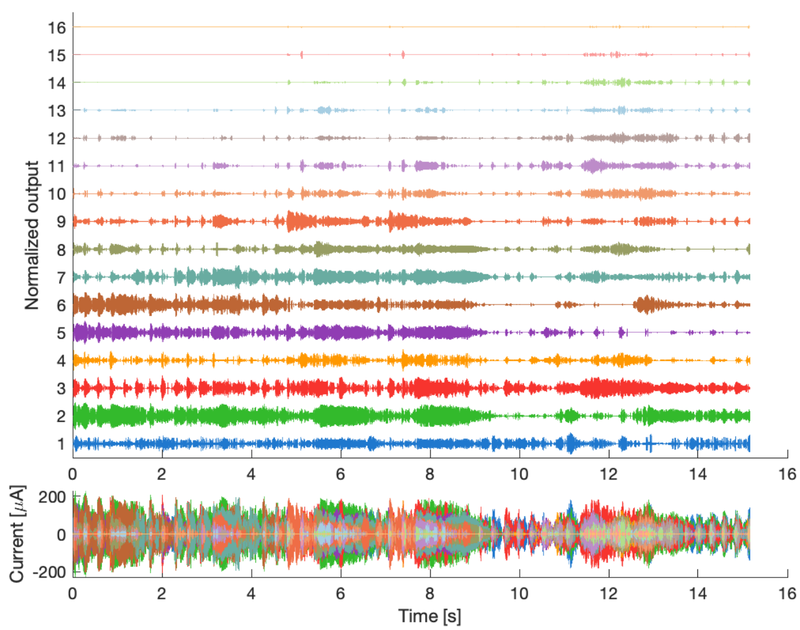


Figure A.3: CI vocoder output audio for music stimuli run through speech intelligibility and quality metrics.



(A)



(B)

Figure A.4: Electrodograms for a sample music stimuli. (A) Default processing (provided code) (B) Processing with additional filtering and reduced complexity.

Bibliography

Abrams, H. B. and Kihm, J. (2015). An introduction to marketrak ix: A new baseline for the hearing aid market. *Hearing Review*, 22(6):16.

Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., Chang, W. S., Chang, J. W., and Jun, S. C. (2017). Interbrain phase synchronization during turn-taking verbal interaction-a hyperscanning study using simultaneous EEG/MEG. *Human Brain Mapping*, 39(1):171–188.

Alain, C., Du, Y., Bernstein, L. J., Barten, T., and Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, 39(7):2695–2709.

Arehart, K. H., Kates, J. M., and Anderson, M. C. (2011). Effects of noise, nonlinear processing, and linear filtering on perceived music quality. *International Journal of Audiology*, 50(3):177–190.

Bernarding, C., Strauss, D. J., Hannemann, R., and Corona-Strauss, F. I. (2012). Quantification of listening effort correlates in the oscillatory EEG activity: A feasibility study. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE.

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., and Corona-Strauss, F. I. (2014). Objective assessment of listening effort in the oscillatory EEG: Comparison of different hearing aid configurations. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE.

Bernarding, C., Strauss, D. J., Hannemann, R., Seidler, H., and Corona-Strauss, F. I. (2017). Neurodynamic evaluation of hearing aid features using eeg correlates of listening effort. *Cognitive Neurodynamics*, 11(3):203–215.

Berntson, G. G., Quigley, K. S., Norman, G. J., and Lozano, D. L. (2017). *Cardiovascular Psychophysiology*. Cambridge University Press.

Brimijoin, W. O., McShefferty, D., and Akeroyd, M. A. (2010). Auditory and visual orienting responses in listeners with and without hearing-impairment. *The Journal of the Acoustical Society of America*, 127(6):3678–3688.

Brons, I., Houben, R., and Dreschler, W. A. (2013). Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort. *Ear and Hearing*, 34(1):29–41.

Brons, I., Houben, R., and Dreschler, W. A. (2014). Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. *Trends in Hearing*, 18:233121651455392.

Cartocci, G., Scorpecci, A., Borghini, G., Maglione, A. G., Inguscio, B. M. S., Giannantonio, S., Giorgi, A., Malerba, P., Rossi, D., Modica, E., Aricò, P., Flumeri, G. D., Marsella, P., and Babiloni, F. (2019). EEG rhythms lateralization patterns in children with unilateral hearing loss are different from the patterns of normal hearing controls during speech-in-noise listening. *Hearing Research*, 379:31–42.

Certo, M. V., Kohlberg, G. D., Chari, D. A., Mancuso, D. M., and Lalwani, A. K. (2015). Reverberation time influences musical enjoyment with cochlear implants. *Otology & Neurotology*, 36(2):e46–e50.

Chasin, M. and Hockley, N. S. (2014). Some characteristics of amplified music through hearing aids. *Hearing Research*, 308:2–12.

Chasin, M. and Russo, F. A. (2004). Hearing aids and music. *Trends in Amplification*, 8(2):35–47.

Coffman, D. D. (2002). Music and quality of life in older adults. *Psychomusicology: A Journal of Research in Music Cognition*, 18(1-2):76.

Cox, R. M. and Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, 16(2):176–186.

Črnčec, R., Wilson, S. J., and Prior, M. (2006). The cognitive and academic benefits of music to children: Facts and fiction. *Educational Psychology*, 26(4):579–594.

Croghan, N. B. H., Arehart, K. H., and Kates, J. M. (2012). Quality and loudness judgments for music subjected to compression limiting. *The Journal of the Acoustical Society of America*, 132(2):1177–1188.

Cubick, J., Santurette, S., Dau, T., and Laugesen, S. (2014). Influence of high-frequency audibility on the perceived distance of sounds. In *7th Forum Acusticum.*

Cvijanović, N., Kechichian, P., Janse, K., and Kohlrausch, A. (2017). Effects of noise on arousal in a speech communication setting. *Speech Communication*, 88:127–136.

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., and König, P. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in Human Neuroscience*, 14.

Dillon, H. (2012). *Hearing aids*. Boomerang Press, Sydney, Australia, 2nd edition edition.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Frontiers in Human Neuroscience*, 11.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*, 9(1):1–10.

D'Onofrio, K. L., Gifford, R. H., and Ricketts, T. A. (2019). Musician and nonmusician hearing aid setting preferences for music and speech stimuli. *American Journal of Audiology*, 28(2):333–347.

Evans, S. and McGettigan, C. (2017). Comprehending auditory speech: previous and potential contributions of functional MRI. *Language, Cognition and Neuroscience*, 32(7):829–846.

Feder, K., Michaud, D., Ramage-Morin, P., McNamee, J., and Beauregard, Y. (2015). Prevalence of hearing loss among canadians aged 20 to 79: Audiometric results from the 2012/2013 canadian health measures survey. https://www150.statcan.gc.ca/n1/pub/82-003-x/2015007/article/14206-eng.htm. Last Accessed: 2022-02-09.

Francis, A. L., MacPherson, M. K., Chandrasekaran, B., and Alvar, A. M. (2016). Autonomic nervous system responses during perception of masked speech may reflect constructs other than subjective listening effort. *Frontiers in Psychology*, 7:263.

Francis, A. L. and Oliver, J. (2018). Psychophysiological measurement of affective responses during speech perception. *Hearing Research*, 369:103–119.

Galvin, J. J., Fu, Q.-J., and Oba, S. (2008). Effect of instrument timbre on melodic contour identification by cochlear implant users. *The Journal of the Acoustical Society of America*, 124(4):EL189–EL195.

Grange, J. A. and Culling, J. F. (2016). The benefit of head orientation to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 139(2):703–712.

GRAS Sound & Vibrations (2022). Kemar. `https://www.grasacoustics.com/industries/audiology/kemar`. Accessed: 2022-02-09.

Gross, J. and Ioannides, A. A. (1999). Linear transformations of data space in MEG. *Physics in Medicine and Biology*, 44(8):2081–2097.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, 98(2):694–699.

Hockley, N. S., Bahlmann, F., and Fulton, B. (2012). Analog-to-digital conversion to accommodate the dynamics of live music in hearing instruments. *Trends in Amplification*, 16(3):146–158.

Holube, I., Haeder, K., Imbery, C., and Weber, R. (2016). Subjective listening effort and electrodermal activity in listening situations with reverberation and noise. *Trends in Hearing*, 20:233121651666773.

Hornsby, B. W. (2013). The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear and Hearing*, 34(5):523–534.

Houtgast, T., Steeneken, H., Ahnert, W., Braida, L., Drullman, R., Festen, J., Jacob, K., Mapp, P., McManus, S., Payton, K., et al. (2002). *Past, present and future of the Speech Transmission Index*. Soesterberg: TNO.

Houtgast, T. and Steeneken, H. J. (1971). Evaluation of speech transmission channels by using artificial signals. *Acta Acustica united with Acustica*, 25(6):355–367.

Innes-Brown, H., Marozeau, J., and Blamey, P. (2011). The effect of visual cues on difficulty ratings for segregation of musical streams in listeners with impaired hearing. *PLoS ONE*, 6(12):e29327.

Jakien, K. M., Kampel, S. D., Gordon, S. Y., and Gallun, F. J. (2017). The benefits of increased sensation level and bandwidth for spatial release from masking. *Ear & Hearing*, 38(1):e13–e21.

Jansen, E. J. M., Helleman, H. W., Dreschler, W. A., and de Laat, J. A. P. M. (2008). Noise induced hearing loss and other hearing complaints among musicians of symphony orchestras. *International Archives of Occupational and Environmental Health*, 82(2):153–164.

Johnsrude, I. S. and Rodd, J. M. (2016). Factors that increase processing demands when listening to speech. *Neurobiology of language*, pages 491–502.

Kates, J. M. and Arehart, K. H. (2010). The hearing-aid speech quality index (hasqi). *Journal of the Audio Engineering Society*, 58(5):363–381.

Kates, J. M. and Arehart, K. H. (2014). The hearing-aid speech quality index (hasqi) version 2. *Journal of the Audio Engineering Society*, 62(3):99–117.

Kirchberger, M. and Russo, F. A. (2016). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, 20:2331216516630549.

Kochkin, S. (2000). Marketrak v: "why my hearing aids are in the drawer" the consumers' perspective. *The Hearing Journal*, 53(2):34–36.

Kochkin, S. (2005). Marketrak vii: Customer satisfaction with hearing instruments in the digital age. *The Hearing Journal*, 58(9):30–32.

Kochkin, S. (2007). Marketrak vii: Obstacles to adult non-user adoption of hearing aids. *The Hearing Journal*, 60(4):24–51.

Kochkin, S. (2010). Marketrak viii: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal*, 63(1):19–20.

Kochkin, S., Sterkens, J., Compton-Conley, C., Beck, D. L., Taylor, B., Kricos, P., Caccavo, M., Holmes, A., and Powers, T. A. (2014). Consumer perceptions of the impact of inductively looped venues on the utility of their hearing devices. *The Hearing Review*, 35(5):16–26.

Lemke, U. and Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear & Hearing*, 37(1):77S–84S.

Levitt, H. (2007). Historically, the paths of hearing aids and telephones have often intertwined. *The Hearing Journal*, 60(11):20–24.

Looi, V., McDermott, H., McKay, C., and Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear & Hearing*, 29(3):421–434.

Luo, X., Masterson, M. E., and Wu, C.-C. (2014). Melodic interval perception by normal-hearing listeners and cochlear implant users. *The Journal of the Acoustical Society of America*, 136(4):1831–1844.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., et al. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America*, 127(3):1491–1505.

Mackersie, C. L. and Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks. *Ear and Hearing*, 37:118S–125S.

Mackersie, C. L., MacPhee, I. X., and Heldt, E. W. (2015). Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. *Ear and Hearing*, 36(1):145–154.

Madsen, S. M. K. and Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, 18:233121651455827.

Madsen, S. M. K., Stone, M. A., McKinney, M. F., Fitz, K., and Moore, B. C. J. (2015). Effects of wide dynamic-range compression on the perceived clarity of individual musical instruments. *The Journal of the Acoustical Society of America*, 137(4):1867–1876.

Marrone, N., Mason, C. R., and Kidd, G. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *The Journal of the Acoustical Society of America*, 124(5):3064–3075.

Marzoli, D. and Tommasi, L. (2009). Side biases in humans (homo sapiens): three ecological studies on hemispheric asymmetries. *Naturwissenschaften*, 96(9):1099–1106.

McDermott, H. J. (2004). Music perception with cochlear implants: A review. *Trends in Amplification*, 8(2):49–82.

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., and Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *Journal of Experimental Child Psychology*, 161:95–112.

Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., and Lyxell, B. (2017). Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends in Hearing*, 21:233121651770639.

Mortezapouraghdam, Z., Bernarding, C., and Strauss, D. J. (2017). Objective assessment of perceived effort in listening by employing EEG feature. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

Muralimanohar, R. K., Kronen, C., Arehart, K., Kates, J., and Pichora-Fuller, M. K. (2013). Quality of voices processed by hearing aids: Intra-talker differences. In *Proceedings of Meetings on Acoustics ICA2013*, volume 19, page 060112. Acoustical Society of America.

Müllensiefen, D., Gingras, B., Musil, J., and Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, 9(2):e89642.

Nagathil, A., Weihs, C., and Martin, R. (2015). Spectral complexity reduction of music signals for mitigating effects of cochlear hearing loss. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(3):445–458.

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (2018). Impact of snr, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hearing research*, 365:90–99.

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., and Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing*, 38(3):267.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants. *Trends in Amplification*, 12(4):316–331.

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

Phillips, S. L., Henrich, V. C., and Mace, S. T. (2010). Prevalence of noise-induced hearing loss in student musicians. *International Journal of Audiology*, 49(4):309–316.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel). *Ear and Hearing*, 37:5S–27S.

Picou, E. M., Bean, B., Marcrum, S. C., Ricketts, T. A., and Hornsby, B. W. Y. (2019). Moderate reverberation does not increase subjective fatigue, subjective listening effort, or behavioral listening effort in school-aged children. *Frontiers in Psychology*, 10.

Picou, E. M., Moore, T. M., and Ricketts, T. A. (2017). The effects of directional processing on objective and subjective listening effort. *The Journal of Speech, Language, and Hearing Research*, 60(1):199–211.

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

Poletti, M. (1999). *The performance of multichannel sound systems*. PhD thesis, ResearchSpace Auckland.

Poletti, M. (2000). The stability of single and multichannel sound systems. *Acta Acustica united with Acustica*, 86(1):163–178.

Prodi, N., Visentin, C., and Farnetani, A. (2010). Intelligibility, listening difficulty and listening efficiency in auralized classrooms. *The Journal of the Acoustical Society of America*, 128(1):172–181.

Putterman, D. B. and Valente, M. (2012). Difference between the default telecoil (t-coil) and programmed microphone frequency response in behind-the-ear (BTE) hearing aids. *Journal of the American Academy of Audiology*, 23(05):366–378.

Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014). Listening effort and speech intelligibility in listening situations affected by noise and reverberation. *The Journal of the Acoustical Society of America*, 136(5):2642–2653.

Ricketts, T. A. and Galster, J. (2008). Head angle and elevation in classroom environments: Implications for amplification. *Journal of Speech, Language, and Hearing Research*, 51(2):516–525.

Rudner, M., Lyberg-Åhlander, V., Brännström, J., Nirme, J., Pichora-Fuller, M. K., and Sahlén, B. (2018). Listening comprehension and listening effort in the primary school classroom. *Frontiers in Psychology*, 9.

Sato, H., Morimoto, M., and Wada, M. (2012). Relationship between listening difficulty rating and objective measures in reverberant and noisy sound fields for young adults and elderly persons. *The Journal of the Acoustical Society of America*, 131(6):4596–4605.

Schafer, P. J., Serman, M., Arnold, M., Corona-Strauss, F. I., Strauss, D. J., Seidler-Fallbohmer, B., and Seidler, H. (2015). Evaluation of an objective listening effort measure in a selective, multi-speaker listening task using different hearing aid settings. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

Schoeffler, M., Bartoschek, S., Stöter, F.-R., Roess, M., Westphal, S., Edler, B., and Herre, J. (2018). webMUSHRA — a comprehensive framework for web-based listening tests. *Journal of Open Research Software*, 6.

Simonsen, C. S. and Legarth, S. (2010). A procedure for sound quality evaluation of hearing aids. *Hearing Review*, 17(13):32–37.

Srinivasan, N. K., Molis, M. R., and Gallun, F. J. (2016). Speech recognition thresholds in reverberation: Effects of age and hearing loss. *The Journal of the Acoustical Society of America*, 139(4):2048–2048.

Strauss, D. J., Corona-Strauss, F. I., Trenado, C., Bernarding, C., Reith, W., Latzel, M., and Froehlich, M. (2010). Electrophysiological correlates of listening effort: neurodynamical modeling and measurement. *Cognitive Neurodynamics*, 4(2):119–131.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-Aho, P. O., and Karjalainen, P. A. (2014). Kubios hrv–heart rate variability analysis software. *Computer methods and programs in biomedicine*, 113(1):210–220.

Thut, G., Schyns, P. G., and Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Frontiers in Psychology*, 2.

Trainor, L., Sonnadara, R., Wiklund, K., Bondy, J., Gupta, S., Becker, S., Bruce, I. C., and Haykin, S. (2004). Development of a flexible, realistic hearing in noise test environment (r-hint-e). *Signal Processing*, 84(2):299–309.

Tremblay, K. L. and Backer, K. C. (2016). Listening and learning: Cognitive contributions to the rehabilitation of older adults with and without audiometrically defined hearing loss. *Ear & Hearing*, 37(1):155S–162S.

Valente, D. L., Plevinsky, H. M., Franco, J. M., Heinrichs-Graham, E. C., and Lewis, D. E. (2012). Experimental investigation of the effects of the acoustical conditions in a simulated classroom on speech recognition and learning in children. *The Journal of the Acoustical Society of America*, 131(1):232–246.

van den Tillaart-Haverkate, M., de Ronde-Brons, I., Dreschler, W. A., and Houben, R. (2017). The influence of noise reduction on speech intelligibility, response times to speech, and perceived listening effort in normal-hearing listeners. *Trends in Hearing*, 21:2331216517716844.

Visentin, C. and Prodi, N. (2017). Effects of the noise type on listening effort: relationship between subjective ratings and objective measurements. In *12th ICBEN Congress, Zurich, Switzerland*.

Völker, C., Bisitz, T., Huber, R., Kollmeier, B., and Ernst, S. M. A. (2016). Modifications of the MUlti stimulus test with hidden reference and anchor (MUSHRA) for use in audiology. *International Journal of Audiology*, 57(sup3):S92–S104.

Wisniewski, M. G. (2017). Indices of effortful listening can be mined from existing electroencephalographic data. *Ear and Hearing*, 38(1):e69–e73.

Wu, Y.-H., Stangl, E., Chipara, O., Hasan, S. S., DeVries, S., and Oleson, J. (2019). Efficacy and effectiveness of advanced hearing aid directional and noise reduction technologies for older adults with mild to moderate hearing loss. *Ear and Hearing*.

Yanz, J. L. and Preves, D. (2003). Telecoils: Principles, pitfalls, fixes, and the future. In *Seminars in Hearing*, volume 24, pages 029–042. Thieme Medical Publishers, Inc., New York, USA.

Zakis, J. A. (2016). Music perception and hearing aids. In *Hearing Aids*, pages 217–252. Springer International Publishing.