4-26-2019 10:00 AM

Understanding hearing aid sound quality for music-listening

Jonathan Matthew Vaisberg
The University of Western Ontario

Supervisor
Macpherson, Ewan A.
The University of Western Ontario Joint Supervisor
Scollie, Susan D.
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

To improve speech intelligibility for individuals with hearing loss, hearing aids amplify speech using gains derived from evidence-based prescriptive methods, in addition to other advanced signal processing mechanisms. While the evidence supports the use of hearing aid signal processing for speech intelligibility, these signal processing adjustments can also be detrimental to hearing aid sound quality, with poor hearing aid sound quality cited as a barrier to device adoption. Poor sound quality is also of concern for music-listening, in which intelligibility is likely not a consideration. A series of electroacoustic and behavioural studies were conducted to study sound quality issues in hearing aids, with a focus on music. An objective sound quality metric was validated for real hearing aid fittings, enabling researchers to predict sound quality impacts of signal processing adjustments. Qualitative interviews with hearing aid user musicians revealed that users' primary concern was understanding the conductor's speech during rehearsals, with hearing aid music sound quality issues a secondary concern. However, reported sound quality issues were consistent with music-listening sound quality complaints in the literature. Therefore, follow-up experiments focused on sound quality issues. An examination of different manufacturers' hearing aids revealed significant music sound quality preferences for some devices over others. Electroacoustic measurements on these devices revealed that bass content varied more between devices than levels in other spectral ranges or nonlinearity, and increased bass levels were most associated with improved sound quality ratings. In a sound quality optimization study, listeners increased the bass and reduced the treble relative to typically-prescribed gains, for both speech and music. However, adjustments were smaller in magnitude for speech compared to music because they were also associated with a decline in speech intelligibility. These findings encourage the increase of bass and reduction of treble to improve hearing aid music sound quality, but only to the degree that speech intelligibility is not compromised. Future research is needed on the prediction of hearing aid music quality, the provision of low-frequency gain in open-fit hearing aids, genre-specific adjustments, hearing aid compression and music, and direct-to-consumer technology.

Keywords

Hearing aids; hearing loss; music; signal processing; speech intelligibility; sound quality

List of Abbreviations

ANFC = adaptive nonlinear frequency compression
ANR = adaptive noise reduction
ANSI = American National Standards Institute
ASHA = American Speech-Language-Hearing Association
B = bass
BSHAA = British Society of Hearing Aid Audiologists
BTE = behind the ear
CAMEQ2-HF = Cambridge Method for Loudness Equalization 2 – High Frequency
CASLPO = College of Audiologists and Speech-Language Pathologists of Ontario
CASP-Q = computational auditory signal processing and perceptional model
CF = cut-off frequency
CR = compression ratio
DAI = direct audio input
DSL = Desired Sensation Level method
DTC = direct-to-consumer
dB = decibel
DSL = desired sensation level
FC = frequency compression
FL = frequency lowering
HA = hearing aid
HAAQI = hearing aid audio quality index
HASQI = hearing aid speech quality index
HL = hearing level
H-M = Hansen model
Hz = hertz
IEC = International Electrotechnical Commission
IEEE = Institute of Electrical and Electronics Engineers
ISD = Itakura-Saito distance
ISO = International Organization for Standardization
ITU = International Telecommunications Union
kHz = kilohertz
LAR = Log-area ratio
LB = low-bass
LLR = Log-likelihood ratio
LPD = Loudness pattern distortion
LTAS = long term average spectrum
M-G = Moore Glasberg
MQ-M = Moore quality model
MR = midrange
MSE = mean square error
MUSHRA = multiple stimulus test with hidden references and anchors
NAL-NL = National Acoustic Laboratory Non Linear method
NAL-R = National Acoustic Laboratory method
NFC = nonlinear frequency compression
openMHA = open source master hearing aid
PEAQ = perceptual evaluation of audio quality

PEMO-Q = perceptual model of quality assessment
PESQ = perceptual evaluation of speech quality
POLQA = Perceptual objective listening quality assessment
PTA3 = 3-frequency pure tone average
RECD = real-ear to coupler difference
RMS = root mean square
rMSE = root mean square error
SD = standard deviation
SII = speech intelligibility index
SPL = sound pressure level
T = treble
VoIP = voice-over-internet-protocol
WDRC = wide dynamic range compression
wRECD = wideband real ear to coupler difference
WSSD = Weighted spectral slope distance

Co-Authorship Statement

This dissertation includes seven chapters. I, Jonathan M. Vaisberg, am responsible for the design of the work presented in this dissertation. I am responsible for study design, data collection, statistical analyses, and results interpretation. I am the lead author on all manuscripts and chapters.

Chapter 1: I am the sole author of the introductory chapter. Ewan Macpherson and Susan Scollie are acknowledged for reviewing drafts of this chapter.

Chapter 2: This chapter was co-authored by Ashley T. Martindale, Paula Folkeard, and Cathy Benedict and published in Journal of the American Academy of Audiology. I was responsible for study ideation and oversaw all aspects of the project. I recruited participants, conducted interviews, analyzed and interpreted results, and wrote the chapter. Ashley T. Martindale supported data analysis. Paula Folkeard attended some participant interviews and verified that audiological procedures were conducted within regulatory guidelines. Cathy Benedict attended all participant interviews and provided mentorship regarding qualitative methodologies. All manuscript co-authors reviewed draft versions prior to submission. Ewan Macpherson and Susan Scollie are acknowledged for reviewing early drafts prior to submission. Amy Wang is acknowledged for her assistance in interview guide development and participant recruitment. Marjorie Leek is acknowledged for sharing her survey questionnaire.

Chapter 3: This chapter was co-authored by Jonathan Pietrobon, Danielle Glista, Vijay Parsa, Ewan Macpherson, and Susan Scollie and will be submitted to a peer-reviewed journal. Jonathan Pietrobon developed the Verifit2 recording software and ensured that recordings were appropriate for analysis. Jonathan Pietrobon also assisted with MATLAB programming. Danielle Glista was consulted on the frequency lowering dataset methodology. Vijay Parsa reviewed draft versions of the HASQI data. Ewan Macpherson and Susan Scollie reviewed the methodology and versions of the draft at various stages of the project. All manuscript co-authors reviewed draft versions prior to submission. James Kates is acknowledged for sharing his HASQI code and for providing helpful comments throughout the analysis. Viji Easwar and

Nazanin Pourmand are acknowledged for contributing the filtered bandwidth and automatic noise reduction datasets, respectively.

Chapter 4: This chapter was co-authored by Paula Folkeard, Vijay Parsa, Matthias Froehlich, Veronika Littmann, Ewan Macpherson, and Susan Scollie and published in AudiologyOnline. For this chapter, I was primarily responsible for statistical analyses, results interpretation, manuscript preparation and contributed in part to study design and data collection. Paula Folkeard helped prepare stimuli and collect data. Vijay Parsa assembled the stimulus recording paradigm and programmed the MUSHRA protocol for data collection. Matthias Froehlich and Veronika Littman of Signia provided music stimuli, provided hearing aids, and contributed to the study design. Susan Scollie contributed to the study design. All manuscript co-authors reviewed draft versions prior to submission. Bilal Sheikh, Adrian Lizzi, and Scott Aker are acknowledged for their assistance with hearing aid recordings and stimulus preparation.

Chapter 5: This chapter was co-authored by Paula Folkeard, Vijay Parsa, Ewan Macpherson and Susan Scollie and will be submitted to a peer-reviewed journal. For this chapter, I was responsible for study ideation and oversaw all aspects of the project. I completed the data reduction, electroacoustic measurement, statistical analysis, results interpretation, and chapter writing. Paula Folkeard helped prepare stimuli and collect data. Vijay Parsa reviewed the electroacoustic measurements. Ewan Macpherson and Susan Scollie reviewed the methodology and versions of the draft at various stages of the project. Bilal Sheikh, Adrian Lizzi, and Scott Aker are acknowledged for their assistance with hearing aid recordings and stimulus preparation.

Chapter 6: This chapter was co-authored by Steve Beaulac, Danielle Glista, Ewan Macpherson and Susan Scollie and will be submitted to a peer-reviewed journal. For this chapter, I was responsible for study ideation and oversaw all aspects of the project. I selected the stimuli and experimental conditions, I programmed the experiment in MATLAB and interfaced it with the open source master hearing aid, I recruited participants, I conducted clinical protocols as required by the project, I collected the data from the participants, I analyzed and interpreted the data, and I wrote the chapter. Steve Beaulac installed the open source master hearing aid on the Linux computer and supported MATLAB integration.

Danielle Glista supervised the clinical aspects of the project. Ewan Macpherson and Susan Scollie reviewed the methodology and versions of the draft at various stages of the project. All manuscript co-authors reviewed draft versions prior to submission. Leonard Cornelisse and Don Hayes of Unitron are acknowledged for their support for experimental implementation. Nancy Bunston, Tina Howard, Carolina Rubiano, and Jesse Sinclair of Unitron are acknowledged for recruitment assistance. Robin O'Hagan, Hasan Saleh, and Maikke Van Eeckhoutte of the National Centre for Audiology are acknowledged for their support piloting the project. Jonathan Pietrobon of Audioscan, A Division of Etymonic Design Incorporated is acknowledged for Verfit2 technical support. Tobias Herzke and Hendrik Kayser of the University of Oldenburg and Hoertech are acknowledged for open source master hearing aid technical support.

Chapter 7: I am the sole author of the concluding chapter. Ewan Macpherson and Susan Scollie are acknowledged for reviewing drafts of this chapter.

Acknowledgments

Funding: This dissertation was supported with funding and/or infrastructure support from the Canadian Academy of Audiology, Canadian Acoustical Association, The Sertoma Foundation of Canada Scholarship, Western University Graduate Research Scholarship, Ontario Research Fund, Ontario Graduate Scholarship, Mitacs Accelerate Award, Audioscan (A Division of Etymonic Designed Incorporated), Unitron Hearing (A Sonova Brand), and Signia.

Supervisors: Thank you Dr. Ewan Macpherson and Dr. Susan Scollie for your mentorship and guidance throughout my PhD. Dr. Macpherson – thank you for taking me on as a graduate student when I entered this program. Dr. Scollie – thank you for joining the team when we reassessed my research interests and objectives. Thank you to both of you for standing by me throughout this journey. With your support, I have grown from a fresh graduate student to a confident early-career academic. I have learned to ask for help, craft my own research and to tackle criticism through an objective and constructive lens. Most of all, I am grateful for the friendships that we have created. Your mentorship will continue to guide me, and I look forward to our interactions as my career progresses.

Committee Members & Administrative Staff: Drs Vijay Parsa and Danielle Glista: Thank you for your insights and suggestions throughout committee meetings and projects that have guided this dissertation, and for helping me see my research from different perspectives.

Dr. Cathy Benedict: Thank you for serving on my comprehensive exam committee. We come from different disciplines and research backgrounds and found a common interest resulting in Chapter 2 of this dissertation. I am grateful to you for what I have learned about qualitative research and music education and am glad to have you as a colleague and friend.

Drs. Marlene Bagatto, Blake Butler, Jessica Grahn, and Todd Ricketts: Thank you for taking the time from your busy schedules to form my defense committee and attend my examination. A special note to Dr. Butler - things have come full circle. You supervised me conducting my first laboratory experiment as an undergrad at McMaster while completing your PhD, and now you are an examiner for my PhD defense. Cool.

NCA, HRS and CSD personnel: This dissertation would not have been completed without any of your help. Dr. Paula Folkeard – thank you for supporting me during the many projects we collaborated on. Steve Beaulac – thank you for your software support operationalizing the master hearing aid, and for putting up with all my emails asking if "you've finished it yet". David Grainger – thank you for building anything we needed in record time. Lucy Kieffer – thank you for your administrative support to acquire new equipment and attend conferences. Dr. Chris Allan – thank you for keeping track of my clinical requirements, especially at the latter end of the dissertation. Thank you to all co-authors and research assistants for playing an integral role supporting the chapters in this dissertation.

Industry Collaborators: This dissertation could not have been completed without the support of those who helped me during my internships. Thank you to Jim Jonkman, Jonathan Pietrobon and John Pumford of Audioscan (A Division of Etymonic Designed Incorporated), and thank you to Nancy Bunston, Doug Baldwin, Leonard Cornelisse, Dr. Don Hayes, Tina Howard, Carolina Rubiano, and Jesse Sinclair of Unitron Hearing.

Classmates & Labmates: Thank you to my clinical classmates for making graduate school more than lectures, placements and research projects. Thank you to my research labmates for your support, ideas, and collective vents.

Family**:** Shelley, David, Mom and Dad – your endless support and encouragement throughout this journey helped get me to where I am today. Thank you for listening (and sometimes politely pretending to listen) to me dive into technical details about my research and for supporting me through all my stresses. And a special acknowledgment belongs to my late grandmother, Lil, who was so happy that I was working on something I loved to do.

Galina: You can take the girl out of Toronto, but you can't take Toronto out of the girl. You took the plunge: you moved to London to be by my side, and you married me - all while I completed my dissertation. I am forever grateful for your patience, love and support.

Table of Contents

Abstract i

List of Abbreviations ii

Co-Authorship Statement iv

Acknowledgments vii

Table of Contents ix

List of Tables xv

List of Figures xvii

List of Appendices xxiii

Chapter 1 1

1 Introduction 1

1.1 Acoustics of speech 2

1.2 Acoustics of music 3

1.3 Attitudes towards hearing aid processed music 4

1.4 Effect of hearing aid parameters on music sound quality 5

1.4.1 Bandwidth 5

1.4.2 Dynamic range compression 6

1.4.3 Feedback cancellation 9

1.4.4 Frequency lowering 10

1.4.5 Noise management strategies 11

1.4.6 Dedicated music programs 12

1.5 Measuring the “success” of hearing aid processed music 13

1.5.1 Music perception tests 14

1.5.2 Subjective sound quality assessment tools 16

1.5.3 Objective sound quality assessment tools 19

1.5.4 An appropriate objective model for hearing aid research 26

1.6 Purpose of the current research 34

1.7 References 36

Chapter 2 43

2 A qualitative study of the effects of hearing loss and hearing aid use on music perception in performing musicians 43

2.1 Introduction 44

2.2 Study Design 48

2.2.1 Sampling 49

2.2.2 Participants 50

2.2.3 Data Collection 54

2.2.4 Analysis 56

2.2.5 Reflexivity 58

2.3 Findings 59

2.3.1 Participants' participatory needs: Hearing the conductor 59

2.3.2 Participants' impressions of hearing aids 61

2.3.3 Effects of hearing loss on music perception 63

2.3.4 Consistency across other literature 66

2.4 Discussion 67

2.4.1 Limitations 70

2.5 Conclusion 71

2.6 References 74

Chapter 3 78

3 Prediction of real hearing aid speech quality using the hearing aid speech quality index (HASQI) 78

3.1 Introduction 78

3.2 Methods 83

3.2.1 Participants 83
3.2.2 Validation data 83
3.2.3 HASQI signal analysis 87
3.2.4 Hearing aid recordings 88
3.2.5 Reference signals 89
3.3 Results 91
3.3.1 Statistical analysis 91
3.3.2 Statistical results 93
3.3.3 Filtered bandwidth 95
3.3.4 Frequency lowering 101
3.3.5 Automatic noise reduction 107
3.4 Discussion 113
3.4.1 Sensitivity to signal processing 114
3.4.2 HASQI across hearing aid users 119
3.4.3 Effects of inaudible noise 119
3.4.4 Implementation considerations 122
3.5 Conclusions and future directions 126
3.6 References 128
Chapter 4 133
4 Comparison of music sound quality between hearing aids and music programs 133
4.1 Introduction 133
4.1.1 The acoustics of speech and music 133
4.1.2 Hearing aids and music programs 135
4.1.3 Study purpose 136
4.2 Methods 136
4.2.1 Participants 136

4.2.2 Test conditions 137
4.2.3 Music genres and recordings 138
4.2.4 Ratings 139
4.3 Results 140
4.3.1 Analysis 140
4.3.2 Effect of model and genre 141
4.3.3 Effect of music programs 143
4.4 Discussion 144
4.4.1 Does music program matter? 144
4.4.2 Does genre matter? 144
4.4.3 Why did sound quality vary? 145
4.4.4 Further thoughts 149
4.5 Conclusion 150
4.6 References 152
Chapter 5 154
5 The relationship between electroacoustic characteristics of default hearing aid music programs and perceived music sound quality 154
5.1 Introduction 154
5.1.1 Summary and purpose statement 158
5.2 Methods 160
5.2.1 Participants 160
5.2.2 Materials 160
5.2.3 Stimulus processing 162
5.2.4 Data reduction 162
5.2.5 Analyses 164
5.3 Results 165

5.3.1 RM-ANOVA results 165
5.3.2 Model results 171
5.4 Discussion 174
5.4.1 Limitations 179
5.5 Conclusion 180
5.6 References 182
Chapter 6 186
6 Preferred frequency shaping for aided music-listening 186
6.1 Introduction 187
6.2 Experiment One 192
6.2.1 Methods 192
6.2.2 Results 200
6.2.3 Discussion 209
6.3 Experiment Two 214
6.3.1 Methods 214
6.3.2 Results 216
6.3.3 Discussion 221
6.4 Summary and conclusions 226
6.5 References 228
Chapter 7 233
7 Discussion 233
7.1 Research aims 233
7.2 Summary of findings 233
7.3 Implications, limitations, and future research 238
7.3.1 Predicting the sound quality of real hearing aid music quality 238
7.3.2 Low-frequency amplification in open versus closed fittings 241

7.3.3 Stimulus-dependent effects 244

7.3.4 Hearing aid compression and music listening 247

7.3.5 Direct-to-consumer hearing devices and the relationship between quality and intelligibility 249

7.4 Concluding statements 251

7.5 References 253

Curriculum Vitae 270

List of Tables

Table 1-1: List of common objective sound quality measures, organized by group..... 21

Table 1-2: Objective sound quality models identified during literature search: Part 1 28

Table 1-3: Average correlation coefficient across validation studies including in review. Group correlations from each validation study were used when available. The coefficients listed above were averaged across each metric's respective studies for normal hearing listeners, hearing impaired listeners and all populations. If multiple coefficients were provided within a study, it would be average to a single value so that all studies would be weighted equally. N = number of studies over which average correlation was calculated ... 31

Table 2-1: Participant characteristics: sex (M = male, F = female), age (years), music experience/ME (years), hearing aid experience/HAE (years), hearing loss type/HLT (SN = sensorineural, M = mixed), hearing aid style/HAS (BTE = behind-the-ear, CIC = completely-in-the-canal, RIC = receiver-in-the-canal, ITE = in-the-ear), binaural or monaural/BM (B = binaural, M = monoaural). 52

Table 2-2: Participant audiometric thresholds. Audiometric thresholds in decibels hearing level. NR= no response. 53

Table 2-3: Interview guide used to prompt dialogue during semi-structured interviews..... 55

Table 3-1: Summary of validation data for correlations on mean values. r = Pearson linear correlation coefficient, rs = Spearman rank correlation coefficient, MSE = mean squared error, rMSE = root mean square error. 94

Table 3-2: Summary of validation data for correlations on individual values. r = Pearson linear correlation coefficient, rs = Spearman rank correlation coefficient, MSE = mean squared error, rMSE = root mean square error. 94

Table 3-3: Summary of sensitivity ANOVA results for dHASQI. df = degrees of freedom. . 115

Table 3-4: Summary of sensitivity ANOVA results for mHASQI. df = degrees of freedom. 116

Table 3-5: Significance of statistical tests across dHASQI, mHASQI, and subjective ratings from original studies. Sig = statistically significant, Non-Sig = statistically non-significant. 125

Table 4-1: Summary of hearing aids and program settings used in this study...... 138

Table 4-2: List of participant's ratings of 2 hearing aid recordings for the folk clip. 146

Table 5-1: Statistical results for repeated-measures analyes of variance for dependent measures of frequency-band specific levels (in dB SPL) and cepstral correlations. DV = depedent variable, IV = independent variables, F = f value, df = degrees of freedom, dfe = error degrees of freedom, p = p-value, η^2 *= effect size.* 170

Table 5-2: Sequence of model fit comparisons for an optimal model which predicts sound quality scores. "Participant" and "Genre" indicate random effects variables for the different levels of those factors. The acronyms for the outcome variable and fixed effects variables are defined as follows: MUSHRA = sound quality ratings using the MUSHRA protocol, , MUSHRA-tx = MUSHRA with transformation, LB = average level (dB SPL) in the low-bass frequency band, B = average level (dB SPL)in the bass frequency band, MR = average level (dB SPL) in the midrange frequency band, T = average level (dB SPL) in the treble frequency band, CC = cepstral correlation...... 172

Table 5-3: Model statistics for the selected iteration of the model...... 172

Table 6-1: Post-hoc simple effects for the interaction of stimulus X frequency band. The difference (dB) column is a difference measure of stimulus 1 relative to stimulus 2. Asterisks indicate statistical significance (p<0.05) 208

*Table 6-2: Pearson correlation coefficients between sound quality descriptors and ratings of overall impressions across individuals. * = p < 0.0001.* 220

List of Figures

Figure 3-1: Scatterplot of mean MUSHRA ratings and dHASQI quality predictions for the filtered bandwidth dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error. 97

Figure 3-2: Scatterplot of average MUSHRA ratings and mHASQI quality predictions for the filtered bandwidth dataset using a recorded hearing aid reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error. 98

Figure 3-3: Scatterplot of individual MUSHRA ratings and dHASQI quality predictions for the filtered bandwidth dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error. 99

Figure 3-4: Scatterplot of individual MUSHRA ratings and mHASQI quality predictions for the filtered bandwidth dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error. 100

Figure 3-5: Scatterplot of average MUSHRA ratings and dHASQI quality predictions for the frequency compression dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error. 103

Figure 3-6: Scatterplot of average MUSHRA ratings and mHASQI quality predictions for the frequency compression dataset using a recorded hearing aid reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error. 104

Figure 3-7: Scatterplot of individual MUSHRA ratings and dHASQI quality predictions for the frequency compression dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error. .. 105

Figure 3-8: Scatterplot of individual MUSHRA ratings and mHASQI quality predictions for the frequency compression dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error… 106

Figure 3-9: Scatterplot of average MUSHRA ratings and dHASQI quality predictions for the automatic noise reduction dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error. 109

Figure 3-10: Scatterplot of average MUSHRA ratings and mHASQI quality predictions for the automatic noise reduction dataset using a recorded hearing aid reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error. 110

Figure 3-11: Scatterplot of individual MUSHRA ratings and dHASQI quality predictions for the automatic noise reduction dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error… 111

Figure 3-12: Scatterplot of individual MUSHRA ratings and mHASQI quality predictions for the automatic noise reduction dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error… 112

Figure 4-1: Frequency-intensity range of speech and music within the audibility of the human auditory system. Adapted from Limb (2011). 135

Figure 4-2: Pure-tone air conduction audiometric thresholds for all participants (n=26). The left panel represents thresholds for participants' left ears and the right panel represents thresholds for the participants' right ears. The light grey lines represent individual audiograms and the dark grey lines represent the mean audiograms across participants for each ear. 137

Figure 4-3: Screenshot of software used to gather the sound quality ratings of hearing aid processed music samples. Clicking on the lettered button played the processed sample randomly assigned to it. Participants adjusted the sliders to indicate their sound quality ratings. 140

Figure 4-4: Boxplots showing the sound quality ratings for all hearing aids in the universal program across genres (on the left) and separated by genre (on the right). Asterisks indicate statistically significant differences between hearing aid models. 142

Figure 4-5: Boxplots showing the sound quality ratings for all hearing aids in the music program across genres (on the left) and separated by genre (on the right). Asterisks indicate statistically significant differences between hearing aids models. 142

Figure 4-6: Boxplots showing the MUSHRA sound quality ratings for both programs for each hearing aid, across all genres. The dark grey boxplots represent the music program and the white boxplots represent the universal programs. Asterisks indicate statistically significant differences between programs within each hearing aid. 143

Figure 4-7: Frequency response curves of the highest and lowest rated hearing aids (left) and hearing aid with the biggest difference between programs (right). Solid lines represent the 0.4-cc coupler long term average spectrum of the output in response to the folk music sample, while dashed lines show the aided peaks and valleys of the signals. The distance between the dotted lines for a single colour is the dynamic range of the hearing aid output. The dark lines are pure tone detection thresholds, plotted in dB SPL in the 0.4-cc coupler. 147

Figure 4-8: Short-term compression ratio as a function of frequency of the highest and lowest rated hearing aids (left) and hearing aid with the biggest difference between programs (right). Solids lines represent compression ratio of each hearing aid. The black dotted line is a reference linear fit. 149

Figure 5-1: Long-term average spectra (left column) and normalized waveforms (right column) for the test stimuli, analyzed using the IEC 60118-15 (2008) recommendation. In the spectra, the top and bottom dashed curves represent the 99th and 30th percentiles, respectively and the solid curve represents the 65th percentile. 161

Figure 5-2: Average sound quality rating across genres and participants (as rated using the multiple stimuli with hidden references and anchors (MUSHRA) method) for each hearing aid for the classical (left) and pop (right) genres. The universal and music program bars are

colored in gray and white, respectively. Error bars represent one standard error of the mean. 166

Figure 5-3: Average ear-simulator real ear aided response levels (REAR) in dB SPL, across participants, for each frequency band (low-bass, 100-200 Hz; bass, 250-800 Hz; midrange, 1000-2500 Hz; and treble, 3000-10000 Hz) for the universal and music programs of each hearing aid. Measurements for classical and pop genres are on top and bottom, respectively. The universal and music program bars are colored in gray and white, respectively. Error bars represent one standard error of the mean. 168

Figure 5-4: Average cepstral correlation (CC) across participants for each hearing aid for the classical (left) and pop (right) genres. The universal and music program bars are colored in gray and white, respectively. Error bars represent one standard error of the mean. 169

Figure 5-5: Scatterplot of average ear-simulator real ear aided response (REAR) levels in dB SPL and average sound quality rating across genres and participants (as rated using the multiple stimuli with hidden references and anchors (MUSHRA) method) per frequency band. Averages were measured across individuals for the conditions: hearing aid, genre and program. Only the low-bass level variable was predictive of sound quality ratings. 173

Figure 5-6: Scatterplot of average cepstral coefficient (CC) and average sound quality rating across genres and participants (as rated using the multiple stimuli with hidden references and anchors (MUSHRA) method). Averages were measured across individuals for the conditions: hearing aid, genre and program. The CC was not predictive of sound quality ratings. 173

Figure 6-1: Mean air conduction pure-tone thresholds for hearing aid user and hearing aid non-user groups for the left ear (left panel) and right ear (right panel). Error bars represent one standard deviation of the mean. 192

Figure 6-2: Long-term frequency spectra measured in 1/3rd octave bands for the stimuli in this study. Black lines represent training stimuli used in both Experiments One and Two, and gray lines represent validation stimuli used in Experiment Two only. 194

Figure 6-3: Visualization of a three-iteration modified simplex procedure run. The x, y, and z coordinates correspond to the parameter settings for the low (0.1-0.8 kHz), mid (1-2.5 kHz) and high (3-10 kHz) gain adjustments, respectively. The cell (0,0,0) represents the initial estimate using prescribed frequency shaping. Each other cell represents shaping that has been adjusted by ±6 dB per step size relative to the initial estimate. In each iteration, the center cell with three larger digits represents the preferred shaping from the previous iteration (or the prescribed shaping in the first iteration). For cells neighbouring the center, the single uppercase digit represents adjusted shaping along that digit's dimension. Paired comparisons are conducted between the center cell and the three adjusted cells. An underlined digit represents the listener's selected preference within a comparison along that digit's dimension. A reversal occurs after the second iteration because the listener prefers the center cell (1,0,-1) relative to adjustments along any of the three dimensions. The run terminates after the third iteration because the center cell (1,0,-1) is preferred twice — after both the second and third iterations. Therefore, this participant's final preferred shaping corresponds to the coordinates: (1,0,-1). 198

Figure 6-4: Cumulative frequency distribution curves illustrating the percentage of listeners (n = 25) who deviated by a given number of step sizes between preferred shaping coordinates from two simplex runs for the low-frequency dimension (A panel), mid-frequency dimension (B panel), high-frequency dimension (C panel), and Euclidean distance across all three dimensions (D panel). Curves were calculated for each stimulus, and for all stimuli pooled. The dotted curve illustrates the cumulative frequency distribution for randomly selected preferred shaping coordinates over 1000 test-retest simulations. 202

Figure 6-5: A (above): Mean coupler responses of the International Speech Test Signal (ISTS; Holube, Fredelake, Vlaming, & Kollmeier, 2010) across listeners and runs at a simulated 70 dB SPL input level as measured via the prescribed and the preferred frequency shaping derived using each stimulus. B (below) Mean observed differences from prescribed gains in dB, measured as the mean difference (dB) across listeners and runs between the preferred and prescribed shaping at audiometric frequencies for a 70 dB SPL input level. 205

Figure 6-6: Mean observed differences from prescribed gains, measured as the mean difference (dB) between the preferred and prescribed shaping averaged across the low (0.1-

0.8 kHz), mid (1-2.5 kHz) and high (3-10 kHz) frequency bands for a 70 dB SPL input level across listeners and runs. Errors bars represent one standard error of the mean. 206

Figure 6-7: Mean sound quality descriptor ratings measured as a function of stimulus status (training/validation), shaping (prescribed/preferred), and stimulus (male speech, female speech, classic music, and pop music) across listeners. Each panel shows ratings for one descriptive dimension. Participants did not rate intelligibility for music. Error bars represent one standard error of the mean. 218

List of Appendices

Appendix A: Ethics approval notice 1 256

Appendix B: Ethics approval notice 2 257

Appendix C: Ethics approval notice 3 258

Appendix D: Copyright permission for Chapter 2 260

Appendix E: Copyright permission for Chapter 4 261

Chapter 1

1 Introduction

Listening to music is an important and enjoyable aspect of many people's lives. Music is considered an art, which combines multiple sounds to produce a type of auditory beauty. However, there are also benefits to music beyond being a form of creativity. Music involvement can improve quality of life through recreational and rehabilitation functions. For instance, music-listening and participation has been associated with enhanced IQ in developing children (Hille, Gust, Bitz, & Kammer, 2011) as well as the mitigation of Alzheimer's disease symptoms in older adults (Simmons-Stern, Budson, & Ally, 2010). Some individuals' auditory systems are compromised, with over a third of Canadian adults aged 20-79 showing some degree of hearing impairment (Feder, Michaud, Ramage-Morin, McNamee, & Beauregard, 2015). Therefore, not all individuals may be able to take full advantage of music's quality of life benefits.

Hearing impairment is a complete or partial reduction in the ability to hear sound. Individuals with hearing impairment typically first report difficulty understanding speech, rather than difficulty perceiving music. This is intuitive, as speech is very important for communication in everyday life. Music, on the other hand, is an art form designed to enrich everyday life. Listeners with hearing impairment often wear hearing aids to improve audibility of sounds. However, hearing aids are typically fitted, programmed, and optimized for speech. Facilitating music-listening in hearing aid users is not fully understood, and making music enjoyable through hearing aids can be challenging. Many researchers have studied music enjoyment using sound quality measurements.

Sound quality is a measure related to the clarity, fidelity, and naturalness of audio output from a hearing aid, and is highly related to overall hearing aid consumer satisfaction (Abrams & Kihm, 2015). Unlike measures of speech intelligibility or music perception tests, sound quality can be measured for any auditory signal including speech and music. Many digital signal processing methods which effectively improve speech intelligibility also distort the amplified signal relative to its original source, potentially signifying a

degradation in sound quality. It is therefore no surprise that a significant portion of hearing aid users are not fully satisfied with hearing aid sound quality (Abrams & Kihm, 2015). Given the number of adjustable hearing aid parameters, and the lack of guidelines to fit them in conjunction with one another, sound quality measurement techniques will afford researchers and clinicians a universal method to compare and optimize combinations of devices, parameter settings, and listening environments. By providing the hearing aid community with accurate and reliable sound quality measurement techniques, researchers can provide new knowledge about hearing aid signal processing parameters, which may help to optimize both sound quality and speech intelligibility.

1.1 Acoustics of speech

Hearing aids optimized for speech may not be optimized as effectively for music. This is likely because speech and music are acoustically quite different. Natural speech will always originate from the same source – the human vocal tract. Speech therefore has well-defined acoustic properties that are fairly consistent across languages (Byrne, 1994) and that contain predictable differences attributed to factors such as age, gender, and vocal effort. Speech tends to have more relative low frequency energy, with a fundamental frequency as low as 100 Hz for males, 160 Hz for females, and 200 Hz for children (Hillenbrand, Getty, Clark, & Wheeler, 1995). Average speech levels can vary between 55 and 66 dBA depending on the speaker vocal effort levels (Olsen, 1998). Additionally its dynamic range is typically 20 – 30 dB (Holube, Fredelake, Vlaming, & Kollmeier, 2010) with a crest factor of 12 dB (Cox, Matesich, & Moore, 1988).

The hardware and software components of a hearing aid have been largely developed to process the acoustic properties of speech. The analog-to-digital converter in the front end of modern hearing aids can handle an input dynamic range of up to 96 dB (Chasin, 2012), which is more than adequate for the 20 – 30 dB dynamic range of speech. In addition, many hearing aid signal processing strategies rely on acoustic properties of speech to improve intelligibility. For example, frequency lowering targets high frequency syllabic information and lowers it to an audible range for the listener (Alexander, 2013). Noise reduction algorithms rely on the amplitude modulation frequencies of speech to suppress the relative intensity of background noise that may be masking the speech signal (Bentler

& Chiou, 2006). Directional microphones rely on a spatial location of the speaker relative to background noise (Ricketts, 2001).

1.2 Acoustics of music

Music is acoustically different from speech. It can originate from a variety of sources, including instruments constructed with many different materials and physical designs in addition to the human voice. The diversity of possible music sound sources expands the possible range of music acoustics compared to speech. Even more variability is introduced when considering different genres and if one listens to recorded or live music.

While the lowest fundamental frequency of speech is about 100 Hz, musical instruments that play in the bass range can have a fundamental frequency as low as 50 Hz (Chasin, 2012). Compared to the relative low-frequency dominance in speech, different frequencies can be distributed differently across the entire music bandwidth depending on the instrumental arrangement and genre of the sample (Kirchberger & Russo, 2016a). Whereas average speech levels vary between 55 and 66 dBA, average music levels can vary between 68 and 108 dBA depending on the instrument played (Chasin, 2012). From a three-meter distance, the sound level of a clarinet ranges from 68-82 dB whereas the trumpet can range from 88-108 dBA. The dynamic ranges of recorded pop, rap, and rock genres are smaller than those of orchestra, piano, and opera genres, which are in turn smaller than the dynamic range of speech (Kirchberger & Russo, 2016a). The dynamic range of live music is much greater than that of speech, being on the order of 80 – 100 dB (Chasin, 2006) with a crest factor up to 20 dB (Chasin, 2012).

The hardware and software components mentioned earlier are likely not appropriate for music. Music with a dynamic range of 80-100 dB can exceed the 96-dB input dynamic range of the analog-to-digital converter of the hearing aid. This upper limit may introduce unwanted distortions of exceedingly loud music. Furthermore, frequency-lowering technology may compromise the integrity of the harmonic structures desired in music, which could introduce unwanted dissonance. Noise reduction may inappropriately identify sustained musical notes as noise and attempt to attenuate them. Feedback suppression systems can cause entrainment, particularly in response to sustained notes with significant

sinusoidal components. These examples represent ways in which music has likely not been considered in the development of hearing technologies. If these technologies are routinely used in modern hearing aid fittings, it is important to understand users' musical experiences using hearing aids.

1.3 Attitudes towards hearing aid processed music

If hearing technologies are predominantly developed to improve speech communication, then it seems logical to assume that hearing aid processing will reduce hearing aid users' enjoyment of music. Furthermore, there are many anecdotal reports suggesting that hearing loss and hearing aid use can be detrimental to music-listening. However, there are few studies which assess the impact of hearing aids on music. This makes it difficult to conclude whether listeners are dissatisfied listening to hearing aid processed music.

Several surveys have documented hearing aid users' enjoyment of music. Leek, Molis, Kubli, & Tufts (2008) surveyed music-listening behaviours in hearing-impaired adults who were hearing aid users. They conducted telephone interviews which addressed questions related to characteristics of hearing loss and hearing aid use, musical habits and practice, sound quality of music, and use of hearing aids when listening to music. They identified that technological developments in hearing aids had reduced music enjoyment concerns, but that almost 30% of respondents still expressed difficulty listening to music. Common complaints included music being too loud or soft, difficulty with melody recognition, difficulty understanding words, and volume changes in music. Madsen & Moore (2014) subsequently investigated whether hearing aids improved or worsened music-listening, as well as the nature and prevalence of listening problems. They conducted an internet-based survey which asked questions similar to those of Leek et al. (2008). Their survey identified that many hearing aid users found their hearing aids to be helpful for both live and reproduced music. However, many respondents identified problems such as distortion, acoustic feedback, irregular gain, unbalanced frequency response, and reduced tone quality. These problems highlight that hearing aid processing can be detrimental to music-listening.

1.4 Effect of hearing aid parameters on music sound quality

Many studies have investigated preferred hearing aid parameters for music enjoyment. Topics of interest have included bandwidth and various forms of compression. Fewer articles investigated higher order algorithms, such as noise reduction and feedback cancellation. Studies investigating the effect of hearing aid signal processing typically measure music preference using sound quality ratings as a dependent variable. Effective measures of hearing aid success for music-listening are discussed later in this chapter. This section discusses literature investigating the effect of hearing aid parameter changes on music sound quality.

1.4.1 Bandwidth

Hearing aid bandwidth often refers to the range of frequencies that a hearing aid can transmit and amplify. Audible bandwidth refers to the frequencies amplified by the hearing aid that are audible for the listener. Perceived sound quality depends in part on the audible bandwidth provided by a hearing aid and can be degraded if sufficient bandwidth is not provided. In general, a wide audible bandwidth seems to be most important for music signals, which tend to contain frequency information across a wide auditory spectrum. Moore & Tan (2003) investigated how the perceived naturalness of music was affected by manipulating the bandwidth of the signal. Normal-hearing listeners judged the perceptual quality of music which varied in upper and lower cut-off frequency. The highest quality music signals were broadband (55-16,854 Hz). Increasing the lower cut-off or decreasing the upper cut-off frequencies resulted in clear quality degradations. Audible bandwidth of a typical hearing aid ranges from 200 – 6000 Hz, which is considerably less than the bandwidth reported above, and so it is not surprising that hearing-impaired listeners are dissatisfied with hearing aid processed music.

Some research has extended the findings from Moore & Tan (2003) to listeners with hearing loss. For example, extended bandwidth beyond the 5-6 kHz limitation in hearing aids has been associated with improved music sound quality in hearing-impaired listeners as well. Listeners with moderate hearing-impairment preferred, on average, music sound

quality with an upper cut-off frequency of 9 kHz over 5.5 kHz (Ricketts, Dittberner, & Johnson, 2008). However, in listeners with more severe hearing losses, there was preference for music with cut-off frequencies ranging from 5-11 kHz (Arehart, Kates, & Anderson, 2011; Brennan et al., 2014; Moore, Füllgrabe, & Stone, 2011). However, these studies have also identified a relationship between audiogram configuration and preference for bandwidth. Higher cut-off frequency has been associated with a shallower audiogram slope, whereas low cut-off frequency has been associated with a steeper audiogram slope (Moore et al., 2011; Ricketts et al., 2008). This suggests that extended high-frequency bandwidth can be a good solution for hearing aid music sound quality but should be considered relative to the patient's audiogram.

Extended bandwidth in the low-frequency range has received considerably less attention than its high frequency counterpart. Chasin (2012) has argued that the perceived quality of deep bass instruments is dominated by harmonic structure, rather than fundamental frequency below 200 Hz. Furthermore, frequencies below 200 Hz could increase the level of amplified noise which would interfere with other listening tasks. However, Revit (2009) has suggested that hearing the fundamental frequency is important for hearing the natural warmth and fullness of low-pitch notes. This second argument is more consistent with Moore & Tan's (2003) finding that frequencies as low as 55 Hz contribute to perceived sound quality. An early study identified that a low cutoff frequency of 90 Hz was preferred to 200 Hz and 650 Hz for the perceived sound quality of hearing aid processed music in hearing-impaired listeners (Franks, 1982). In addition, low-frequency cut-offs as low as 200 Hz were required to avoid quality degradations in hearing aid amplified music (Arehart et al., 2011). Finally, a case study of one listener's quality ratings found that a better quality hearing aid had an additional 10 dB gain in frequencies below 300 Hz (Vaisberg et al., 2017). Overall, it appears that low frequency extended bandwidth should be a strong consideration for hearing-impaired listeners for music-listening.

1.4.2 Dynamic range compression

Dynamic range compression is a method which normalizes the dynamic range of hearing aid output between a listener's auditory thresholds (minimum audible ouput) and uncomfortable listening levels (maximum allowable output). It does so by making level-

dependent gain adjustments as the level of speech entering the hearing aid fluctuates. Typically, a hearing aid applies compression by amplifying soft passages more than loud passages to increase overall audibility. However, there is some flexibility to how compression is applied. Parameters (further described below) including compression ratio, time constants, and number of frequency channels can all be manipulated to apply compression differently. The application of minimal compression ratios and longer time constants is typically ideal for optimal speech quality (Souza, 2002). It is of interest to determine if similar parameter recommendations can be made for music quality.

The compression ratio is the parameter which determines how much gain is applied at each level. A compression ratio of 2:1 would imply that for every doubling of signal level, half the gain would be applied at the higher level. This concept contrasts linear gain (1:1), in which constant gain is applied to the signal regardless of its level, and expansion (0.5:1), in which less gain is applied to very soft inputs to reduce the amplification of processing noise. Several studies have examined the effect of manipulating compression ratio and ratings of amplified music. For example, van Buuren, Festen, & Houtgast (1999) investigated the effect of various compression ratios (0.25:1, 0.5:1, 1:1, 2:1, and 4:1) for several music genres. They found that the highest pleasantness ratings were associated with the reference linear amplification (1:1) condition across all genres. Subsequent studies showed a similar trend. When listening to a variety of music genres processed using multiple compression ratios, linear or linear-like gains were typically preferred relative to minimal compression ratios, which were in turn preferred relative to larger compression ratios (Arehart et al., 2011; Croghan, Arehart, & Kates, 2014; Higgins, Searchfield, & Coad, 2012; Kirchberger & Russo, 2016a). In addition, listeners reported greater clarity of individual instruments when listening using linear amplification compared to wide-dynamic range compression (Madsen, Stone, McKinney, Fitz, & Moore, 2015). It is, however, possible that linear gain can be problematic if it cannot reproduce the signal of interest with fidelity. Exceptionally loud signal peaks can exceed the dynamic range of the analog-to-digital converter of the hearing aid, which can introduce distortions caused by peak-clipping and output limiting. In one study, hearing-impaired listeners rated the quality of music amplified using wide-dynamic range compression or linear amplification with peak-clipping or output limiting. Listeners slightly preferred the compressive settings to

the other settings (Davies-Venn, Souza, & Fabry, 2007). Clinicians fitting hearing aids should be wary of the level of the signal being amplified by a hearing aid and if distortions produced by compression are more detrimental to sound quality compared to peak-clipping or output limiting.

Compression time constants refer to the speed it takes for compression to be applied in response to a level change. When listening to a stimulus with a fixed average level, the level may fluctuate from moment to moment leading to a corresponding change in gain from moment to moment. Attack times are the time required for the gain to adjust to a new signal level, and release times are the time required for the gain to return to its default setting. Large, rapid changes in gain can lead to distortions, which may be disruptive to the listener. When listening to amplified music with release times of 40 ms or 4 seconds, hearing-impaired listeners consistently preferred the longer release time relative to the shorter (Hansen, 2002). However, there were also some genre and level interactions. Moore et al. (2011) identified that slow time constants (50 ms attack time, 3000 ms release time) were preferred to faster time constants for classical and jazz music at 80 dB SPL, while slow time constants were preferred to faster time constants for classical music at 65 dB SPL. In another study, slow time constants (50 ms attack time, 1000 ms release time) were preferred to fast time constants (5/50 ms) for both classical and rock music (Croghan et al., 2014). However, when comparing compressed music with time constants between 10 ms, 70 ms and 200 ms, there were essentially no preferences (Arehart et al., 2011). Together these results encourage the use of long time constants that have an attack and release time of at least 50 ms and 1000 ms, respectively, for optimal music sound quality.

Finally, compression can be manipulated on a channel by channel basis. Multiple channels can allow for more compression at frequencies where listeners may have elevated thresholds relative to other frequencies. While this can help practitioners have greater control of the hearing aid signal-processing, it also has the potential to disrupt spectral peak-to-valley differences which provide balance to the musical spectrum. In general, a single channel, or multiple channels with similar compression ratios, has been recommended for a balanced musical spectrum (Chasin, 2006). When hearing-impaired listeners rated sound quality of music processed by 1, 4 or 16 processing bands, listeners

preferred a single band most frequently (van Buuren et al., 1999). Furthermore, fewer channels (3 vs 18) were preferred for rock music, but not classical music (Croghan et al., 2014). However, in both studies, whether fewer channels were preferred was also a function of whether there was a notably high compression or shorter time constants. Therefore, when considering the number of channels in a compressive hearing aid, it is important to consider it in conjunction with other factors such as compression ratio, time constants, and stimulus.

1.4.3 Feedback cancellation

Feedback cancellation is a form of digital signal processing which aims to cancel the high-pitch “whistling” that is occasionally produced by hearing aids. This “whistling” noise can be annoying to the hearing aid user and listeners around them. Amplified high frequencies are particularly susceptible to generating feedback. As a result, feedback cancellation is typically applied to cancel or filter out the acoustic feedback. However, it is not clear if it is always fully effective. In a recent survey, a third of respondents indicated that they heard acoustic feedback when listening to music (Madsen & Moore, 2014), suggesting that feedback is a significant issue during music-listening. Furthermore, a limitation of feedback cancellation is that it can introduce additional distortions which reduce the sound quality of music. Moore (2016) articulated that these systems can cancel musical tones if tones are steady, and produce “after tones” if a musical tone suddenly stops (known as entrainment). In one study, hearing-impaired listeners were asked to listen to classical music amplified through two hearing aids with feedback cancellation on or off without audible feedback. The results showed that the activation of feedback cancellation did not degrade the sound quality of hearing-aid processed music (Johnson, Ricketts, & Hornsby, 2007). The results support the application of feedback cancellation systems without a noticeable degradation of music sound quality when feedback is absent (at least for classical music). Further research is needed to investigate the effectiveness of feedback cancellation during music-listening when feedback is present, as well as the risk of entrainment for different musical instruments.

1.4.4 Frequency lowering

An alternative to the provision of high-frequency amplification is with the use of frequency lowering. This technology aims to provide high-frequency audibility by moving high-frequency sounds to a lower frequency range where there is more audibility (Alexander, 2013). Its primary purpose is to assist listeners presenting with severe sloping hearing impairments for which amplification of high frequencies can be challenging. Several studies show substantial speech perception benefits in some individuals when frequency lowering is applied (Glista et al., 2009; Glista, Scollie, & Sulkers, 2012; Wolfe et al., 2010). Clinical methods for fitting this processor have been developed using calibrated speech sounds as stimuli (Scollie, Glista, et al., 2016).

Frequency lowering could be disruptive to music perception, as it has the potential to disrupt the harmonic structure of musical notes which gives the sense of tonality and richness to some music instruments. The impact of frequency lowering on music-listening is more variable compared to its impact on speech. Some studies have found a frequency lowering music sound quality impact whereas others have not. Parsa, Scollie, Glista, & Seelisch (2013) investigated the impact of frequency lowering on the sound quality of classical and contemporary music. They did so by manipulating the range of high frequencies that were reduced to a lower bandwidth in hearing-impaired listeners with severe sloping hearing loss. There were no statistically significant differences in sound quality ratings between frequency lowering conditions and standard amplification. Mussoi & Bentler (2015) investigated the impact of frequency compression in hearing-impaired listeners with mild to moderate hearing loss who were trained musicians. Listeners judged the quality of various classical arrangements with varying amounts of frequency compression. Listeners showed preference for conditions with lesser amounts of compression, although there was considerable variability between mild processing and traditional amplification. Brennan et al. (2014) compared quality judgments of music processed using frequency compression, extended bandwidth, or narrow bandwidth processing in listeners with mild to severe hearing impairment. Their results demonstrated a preference for either frequency compression or extended bandwidth over narrow bandwidth processing. Kirchberger & Russo (2016b) developed a novel frequency

lowering algorithm designed to preserve the harmonic relationships found in music. While their algorithm improved attention to musical detail in hearing-impaired listeners, sound quality judgments were statistically unchanged compared to traditional frequency lowering and original processing. Together these studies suggest that there may be an acceptable range of frequency lowering in which speech benefits may be attained without degradation of music sound quality.

1.4.5 Noise management strategies

There are several signal processing mechanisms that are used to minimize the distortive effects of noise on speech intelligibility. Two mechanisms include the directional microphones and noise reduction systems.

Directionality is implemented using two microphones, and this benefit relies on speech and noise being spatially separated. They strive to amplify a signal coming from the direction that the listener is facing, rather than amplifying noises from other directions in their surroundings. Directional microphones are typically preferred for various forms of speech perception when there is noise present (Gnewikow, Ricketts, Bratt, & Mutcher, 2009; Picou, Moore, & Ricketts, 2017; Picou & Ricketts, 2017; Preves, Sammeth, & Wynne, 1999; Ricketts, Henry, & Gnewikow, 2003; Ricketts, 2001). However, in quieter environments, omnidirectional microphones (which are equally sensitive to all directions) are typically preferred (Preves et al., 1999; Surr, Walden, Cord, & Olson, 2002). The impact of directional microphones on music-listening has received less attention. Greasley (2016) hypothesized that directional microphones can help discriminate instruments in the presence of competing sounds. However, directional microphones may also be problematic if equally important sounds are coming from sounds located in separate sources, such as an ensemble in which different instruments are located in different areas.

Noise reduction is a term used to refer to a family of digital signal processing algorithms that are designed to improve hearing aid users' experiences in noisy environments. These systems rely on spectrotemporal differences between speech and noise. While the exact nature of a noise reduction algorithm varies between manufacturers, they generally attenuate noise levels while amplifying speech peaks in the waveform. Some studies have

found sound quality of speech processed using noise reduction systems to be comparable to sound quality of speech processed without noise reduction systems (Bentler, Wu, Kettel, & Hurtig, 2008; Scollie, Levy, et al., 2016). Other studies have found a sound quality preference for noise reduction (Ricketts & Hornsby, 2005) and combined noise reduction and directional microphones (Boymans & Dreschler, 2000) in noisier environments. Theoretically, noise reduction is less likely to serve the same benefit for music perception as it does for speech perception. Musical melodies may occur at different frequencies and may modulate at faster and slower rates compared to speech which may make it challenging for a noise reduction algorithm to recognize the appropriate signal. Furthermore, sustained instruments in a musical excerpt may be recognized as noise and it would not be desirable for a noise reduction algorithm to attenuate them. Unfortunately, these statements are hypothetical, as there is little research that has investigated the effect of noise reduction systems on music-listening. Examining the effect of directional microphones and noise reduction mechanisms on music perception would therefore be of interest.

1.4.6 Dedicated music programs

Many hearing aid manufacturers incorporate dedicated music programs in their hearing aids that are designed to enhance the enjoyment of music-listening. Rather than being a special form of processing on its own, a music program is typically a combination of settings that use the mechanisms described above. Common features in a music program include slow compression time constants, minimal noise reduction, minimal directionality, and reduced feedback cancellation (Moore, 2016), and these settings are consistent with preferred dynamic range compression settings for sound quality cited above. Unfortunately, manufacturers typically do not report their music programs' electroacoustic characteristics, nor their efficacy at improving music-listening. In Madsen & Moore's (2014) survey, 40% of respondents reported having a music program in their hearing aids. However, their satisfaction ratings were no different compared to hearing aid users who did not have a music program. This suggests that, on average, music programs do not improve music sound quality any better than a standard hearing aid program. Users are also

often unsure if their hearing aids include a music program, and they did not use them consistently when they do (Fulford, Ginsborg, & Greasley, 2015).

1.5 Measuring the "success" of hearing aid processed music

The number of combinations of hearing aid parameter adjustments is nearly boundless; this warrants the need for hearing aid evaluation strategies that can evaluate music-listening success in a multi-parameter context. How can hearing aid success be measured for music? In the context of speech, hearing aid success can be defined using a variety of measures. Some behavioural measures include speech recognition thresholds and/or word recognition scores. Speech recognition thresholds determine the level at which speech is just intelligible, and are obtained by asking the listener to repeat words which are systematically adjusted in level (Gelfand, 2009, p 240). Another measure is the word recognition score, in which a percentage of correctly repeated words at a suprathreshold level is indicative of the listener's ability to readily identify speech sounds at a given level (Gelfand, 2009, p 249). Both techniques, which are incorporated into preferred practice guidelines (ASHA, 2006; CASLPO, 2014), can be used to verify hearing aids' benefits relative to unaided listening. Another measure, the Speech Intelligibility Index, is standardized (ANSI, 1997) and calculates a score between 0.0 to 1.0 and is highly correlated with the intelligibility of speech (Amlani, Punch, & Ching, 2002; Honsby, 2004). It generates the score by determining the audibility of speech that is above threshold and noise level per weighted critical band, and then summed over the total number of critical bands. It can be used to evaluate the output of a hearing-aid-amplified speech stimulus. Each of these tests share the feature that they measure or predict performance using words and speech, and the percentage of words (or audibility) measured can change as a function of how hearing aid parameters are changed. Unfortunately, what one hearing aid user may be listening for in a piece of music may vary from the next hearing aid user, raising questions about whether there is truly "intelligibility" in the context of music-listening. There are no standardized procedures designed to evaluate music perception in listeners with hearing loss or hearing aids. This next section will characterize different strategies from the literature that

researchers and clinicians could implement to measure hearing aid benefits for music-listening.

1.5.1 Music perception tests

A music perception test is an assessment tool which can potentially be used to evaluate hearing aid success for music-listening. A music perception test is a procedure which is designed to assess different music perception dimensions that are thought to contribute to the overall music-listening experience. These dimensions consist of music features like meter, harmony, melody, and timbre. While many tests of this nature have been developed for assessment of music perception in listeners with cochlear implants (Kang et al., 2009; Spitzer, Mancuso, & Cheng, 2008), others have been developed more recently for assessment of music perception in hearing aid users.

Uys & van Dijk (2011) developed a music perception test designed to evaluate different aspects of perception of rhythm, timbre, pitch, and melody in hearing aid users. The nature of the test depended on the dimension being evaluated. Rhythm was evaluated by asking listeners to discriminate one temporal rhythm for another, to identify whether a rhythm was a waltz or a march, or whether a melody was played out of time. Timbre was evaluated in two parts. First, timbre was evaluated by asking listeners to identify what instrument was playing. Timbre was then evaluated by asking listeners to indicate how many instruments were playing in a given passage. Pitch was also evaluated in two parts. Pitch was first evaluated by asking listeners to discriminate whether sequential pitches were moving higher or lower in frequency. Next, pitch was evaluated by asking whether two melodies differed by single note manipulations. Melody was evaluated by asking listeners to listen to a melody and to indicate whether it followed specific musical rules, to indicate whether they recognized a familiar melody, and to indicate whether they could identify a familiar melody in the presence of distracting auditory stimuli. The authors administered this test in normal-hearing and hearing-aid-user populations, and identified significant deficits in the hearing aid user population relative to the normal hearing population. They proposed the test as a counselling tool to assist patients in understanding their music perception difficulties. One limitation of this study, in the context of hearing aid evaluation, is that it was not used to evaluate improvement in these music dimensions in hearing-impaired

listeners before and after a hearing aid fitting. Therefore, it is not clear if this test is sensitive to improvements in music perception following amplification.

Kirchberger & Russo (2015) developed an adaptive music perception test that was designed to be more sensitive to differences between normal hearing and hearing-impaired listeners compared to the test proposed by Uys & van Dijk (2011). The test used an adaptive two-alternative-forced-choice method for each dimension evaluation, with the difference between two stimuli in a preceding trial influencing the difference in the following trial. Dimensions assessed consisted of meter, harmony, melody, and timbre. Meter was evaluated by asking listeners to discriminate rhythms that differed in level, pitch or duration. Harmony was evaluated by asking listeners to discriminate chords which varied in dissonance or intonation. Melody was evaluated by asking listeners to determine whether a melody shared the same key as the chords with which it was presented. Timbre was evaluated by asking listeners to discriminate timbres which differentiated by brightness, attack, or spectral irregularity. The authors administered the test on two occasions in both normal hearing and hearing-impaired populations. In general, the hearing-impaired population was poorer at detecting differences in some aspects of meter, harmony, timbre, and melody relative to the normal hearing population. Unfortunately, this test was not administered in a population of hearing aid users, so its applicability in identifying improvements attributed to hearing aid use is not yet understood.

Even if these tests were administered before and after hearing aid fittings, there are some practical considerations before bringing these tests to widespread clinical and research use. For example, within-individual test-retest reliability was not assessed by Uys & van Dijk (2011). Additionally, some of the dimensions tested by Kirchberger & Russo (2015) were confounded with poor test-retest reliability. The tests are also time-consuming, which may increase the administrative burden for the test administrator and the client. Finally, and perhaps most importantly, music perception requires a coherent perception of all these dimensions simultaneously. Therefore, based on the tests, it is unclear how a deficit in one dimension would correspond to an overall deficit associated with a musical-related hearing aid outcome. Some other type of outcome measure, which generalizes across music

dimensions, may be more attractive and efficient for measuring hearing aid success for music.

1.5.2 Subjective sound quality assessment tools

Sound quality is often assessed using listeners' subjective judgments. That is, a listener's auditory description of a hearing aid processed stimulus can be used to inform adjustment of their hearing aid (Jenstad, Van Tasell, & Ewert, 2003; Sabin, Hardies, Marrone, & Dhar, 2011). Some of the early research on sound quality was conducted by Gabrielsson & Sjögren (1979). They described sound quality as a multidimensional phenomenon consisting of descriptors including "clear", "soft", "dark", and "shrill", for example. The authors had the goal of determining which descriptors were associated with perceived sound quality and determining what physical characteristics of the systems were associated with said descriptors. Listeners were required to listen to samples of music, speech, and sounds from daily life played back over loudspeakers, headphones, and hearing aids and indicate how well various descriptors characterized the sound reproduction using a rating from 0 (not associated) to 9 (very associated). The authors then performed a factor analysis to determine broad descriptors which accounted for most of the sound quality associations. The descriptors were interpreted as follows: "clearness/distinctness", which was associated with broad frequency range and flat frequency response, "sharpness/hardness-softness", which was associated with marked resonance peaks with suppressed bass response, "brightness-darkness", which was associated with a frequency response rising towards the treble if bright, "fullness-thinness", which was associated with a broad frequency range with emphasis in the bass range, "nearness", which was moderately associated with sounds being more intense and broadband, "disturbing sounds", which was associated with a presence of high frequency distortions, and "loudness", which was associated with an increased perceived loudness due to the presence of other descriptors. While these descriptors could assist individuals to articulate the characteristic quality of a sound, their use can be challenging if the listener is unable to describe how the sound should improve and simply wants a better overall sound quality experience.

Subjective sound quality measurement assessment techniques that rely on a single measure are frequently used in the literature. These assessments are based on perceptual judgments

of hearing aid processed stimuli by a group of listeners. Listeners are presented with an auditory stimulus that has been processed by one or several forms of distortions, and they are then required to indicate the signal's sound quality using an absolute rating or relative comparison. There are a variety of methods to conduct these assessments.

One sound quality assessment technique is the categorical rating method. This assessment requires listeners to subjectively rank sound quality on a predetermined scale in either one or several perceptual dimensions. A standardized version of this technique is the absolute category rating test (ITU-T, 1996), which has traditionally been used to evaluate sound quality in telecommunications applications. In this test, listeners are asked to rate sound quality of telecommunications-processed passages on a 5-point scale from 1 (bad) to 5 (excellent). The final score is a mean opinion score that has been averaged across a group of listeners who rated the passage. In hearing aid research, categorical rating scales have been used in sound quality assessments for a variety of signal processing parameters.

Another sound quality assessment strategy is the application of paired comparisons. Amlani & Schafer (2009) provided a comprehensive review of paired comparisons and its application in hearing aid research. In this strategy's basic form, a paired comparison involves the presentation of two auditory stimuli and a forced response from the listener on a criterion (i.e., sound quality). This strategy allows a listener to compare devices or electroacoustic parameters of devices in any testing environment. By asking listeners to conduct a series of paired comparisons, it is possible for researchers and clinicians to determine preferred hearing aid settings at population levels. There are several strategies to administer series of paired comparisons. The round-robin procedure is a nonadaptive tournament strategy in which listeners perform paired comparisons between every possible combination of stimuli. These comparisons produce rank-ordered data about the parameters compared which can be used to draw conclusions about preferred parameters at the group level.

The iterative round-robin strategy is an example of an adaptive tournament strategy (Neuman, Levitt, Mills, & Schwander, 1987). This strategy compares a subset of hearing aid parameters compared in the nonadaptive round-robin procedure. The winning condition

from that subset becomes the center of a subsequent iteration of other conditions (also compared in the nonadaptive round-robin procedure). The winner is decided when the winning condition is the same in two iterations.

Convergence strategies are other applications of paired comparisons in which listeners' responses on a previous trial inform the stimuli to be compared in a subsequent trial. They can be more efficient than tournament strategies because they reduce the number of comparisons to be completed while reaching the same goal. This can save administration time. In a simple up/down convergence strategy, a listener compares an initial estimated fitting with a new fitting, and the result determines the next set of comparisons. This process continues until the listener cannot discriminate between two final settings, at which point the optimal setting for a single hearing aid parameter is determined. In a modified simplex convergence strategy (Neuman et al., 1987), the listener performs three comparisons of three stimuli which each differ by one of two possible hearing aid parameters. The optimum condition from that iteration determines the next three stimuli to be compared, with this process continuing until the same optimum condition has been selected after three iterations. The modified simplex procedure mainly differs from the simple up-down strategy because it allows for optimization of a combination of hearing aid parameters.

Finally, the "MUltiple Stimulus test with Hidden References and Anchors" (MUSHRA) protocol (ITU-R, 2015) has recently been proposed as an accurate and reliable paradigm for the evaluation of sound quality for telecommunications applications. On a user interface, a listener can listen to a variety of stimuli by pressing a digital button, and rank the relative quality of each stimulus using a slider under each button. The stimuli consist of an unprocessed, high quality reference stimulus, a range of experimental stimuli, and several poor-quality anchor stimuli. The stimuli are then embedded into the interface by being randomly assigned to the digital buttons. The reference signal provides the listener with an example of high sound quality, while the anchor stimuli provides the listener with examples of poor sound quality. The presence of the hidden references and anchors are to serve as examples of the end points of the rating scale, allowing the listeners to quickly orient themselves to the rating task. This protocol, and variations of it, have been applied

in sound quality assessments for hearing aid compression (Kirchberger & Russo, 2016a), frequency lowering (Glista et al., 2009; Kirchberger & Russo, 2016b; Parsa, 2013) and noise reduction (Scollie, Levy, et al., 2016).

While these subjective ratings methods can exhibit high validity and reliability, there are disadvantages to applying them. These methods can be time-consuming and resource-intensive, limiting time available for a clinician or researcher to adequately perform a battery of tests and assessments. They may also be difficult to administer in young pediatric populations. Additionally, in a clinical environment, a clinician may not have access to a variety of hearing aids and signal processing strategies necessary for a comprehensive evaluation of different products. Therefore, an objective index that is predictive of hearing aid sound quality and that is highly correlated with subjective ratings of sound quality may be more appealing and practical for rapid assessments of sound quality.

1.5.3 Objective sound quality assessment tools

Objective sound quality models are algorithms which produce sound quality scores based on algorithm input parameters such as hearing loss and type of stimulus. Most sound quality models follow a perception-model approach, in which they use a psychoacoustic model of the auditory system to produce sound quality scores. It is best practice to train and validate these models using real listeners' subjective ratings of sound quality. Most existing models were developed for normal hearing listeners using telecommunications applications, although recent models have been developed to evaluate advanced digital signal processing for hearing-impaired listeners using hearing aids. Most of the models are classified as intrusive metrics, in that they compare a modelled representation of a signal under test to the modelled representation of that signal's undistorted reference version. The comparison considers differences between the two signals' quality degradations from the reference, which are in turn used to produce the final quality score. Therefore, these models require a reference signal that represents the optimum sound quality. In hearing aid research, the high-quality reference signal is often the original stimulus amplified to individualized prescriptive speech targets without any additional advanced signal processing. Table 1-1 provides a list of some common intrusive models and their acronyms. Models are grouped into different classes corresponding to their underlying auditory

models and include: technical models, Zwicker-based models, Moore-Glasberg-based models, Dau-based models, and Kates-Arehart models.

Early technical sound quality models based on speech production include the LLR index (Itakura, 1975), the ISD index (Itakura & Saito, 1970), and the LAR index (Quackenbush, Barnwell III, & Clements, 1988). In their original form, these metrics analyze signals in the narrowband, 0-4 kHz, range. The LLR and ISD indices rely on the linear prediction coefficient (LPC), a function that represents the spectral envelope of a signal. The indices then calculate the similarity between the LPCs of the distorted and reference signal to obtain an index value. The LAR quality model relies on the log area ratio, another representation of the LPC, and calculates the distance between ratio coefficients between the distorted and reference signals.

An early quality index using a model of the peripheral auditory system is the WSSD index (Klatt, 1982). In this model, spectral slopes are computed for each critical band, and differentially weighted depending on whether the band is near or far from a spectral peak. The difference between distorted and reference signals' slopes is then computed to produce an index value.

One of the first metrics modelling both the peripheral system and internal representations, which has been ITU-standardized (2001), is the PEAQ index (Thiede et al., 2000). It is primarily used for evaluating wideband (20 Hz-15 kHz) audio codecs with mild quality degradations for normal hearing listeners. The peripheral ear model of the PEAQ uses masked thresholds, and filters the reference and distorted signals using a fast Fourier transform and linear filterbank. It then computes nerve excitation patterns and loudness patterns using Zwicker's auditory model (Zwicker, 1961). These outputs produce eleven signal characteristics that are combined to produce a quality value, known as an objective difference grade.

Table 1-1: List of common objective sound quality measures, organized by group

Group	Model Name	Acronym	Reference
Technical models	Log-likelihood ratio	LLR	Itakura (1975)
	Itakura-Saito distance	ISD	Itakura & Saito (1970)
	Log-area ratio	LAR	Quackenbush et al. (1988)
	Weighted spectral slope distance	WSSD	Klatt (1982)
Zwicker-based models	Perceptual evaluation of audio quality	PEAQ	Thiede et al. (2000)
	Perceptual evaluation of speech quality	PESQ	Beerends et al. (2002)
	Perceptual objective listening quality assessment	POLQA	Beerends et al. (2013)
Moore-Glasberg models	Moore-quality model	MQ-M	Moore et al. (2004)
	Loudness pattern distortion	LPD	Chen et al. (2006)
Dau-based models	Hansen model	H-M	Hansen & Kollmeier (2000)
	Perceptual model of quality	PEMO-Q	Huber & Kollmeier (2006)
	Computational auditory signal-processing and perception model	CASP-Q	Jepsen et al. (2008)
Kates-Arehart models	Hearing aid speech quality index	HASQI	Kates & Arehart (2014)
	Hearing aid audio quality index	HAAQI	Kates & Arehart (2016)

Two other Zwicker-based models are PESQ (Beerends et al., 2002; ITU-T, 2001) and its wideband extension (PESQ-WB; ITU-T, 2007). PESQ is among the most popular quality models, having been used to evaluate a wide variety of telecommunications and hearing aid distortions for normal hearing and hearing-impaired listeners. PESQ and PESQ-WB were developed for assessing speech codecs in the 100-3500 Hz and 50-7000 Hz bands, respectively. These metrics extract a time-aligned, frame-by-frame representation of the original and distorted signal and use the Zwicker model to transform the spectrum to the Bark scale and to transform the amplitudes to “loudness” values. The resulting loudness values are compared to produce the PESQ score. PESQ models weigh added distortions more heavily towards sound quality index values compared to removed signal portions.

PESQ was also modified to incorporate a hearing loss model (Beerends, Krebber, Huber, Eneman, & Luts, 2008). It retains the basic structure of PESQ with the following changes: inclusion of an absolute level indicator, a new loudness scaling algorithm which models the effects of hearing impairment, adaptation of time-frequency integration, and a masking algorithm that more closely resembles the perceptions of hearing-impaired listeners.

The PESQ index was recently succeeded by the POLQA index (Beerends et al., 2013; ITU-T, 2014). This update addressed new wideband voice services in the telecommunications industry that could not be assessed using the PESQ. POLQA retains the basic structure of PESQ but is marked by several key modifications. Firstly, quality estimates for frequency response distortions, additive noise, and room reverberations are produced following signal comparisons after the Zwicker model transformation. These three components are kept separate from the comparison of internal representations to balance the impact of different forms of distortions. Secondly, several variants of internal representations are used to calculate two final disturbance densities representing final disturbances and added degradations. The variants are designed to account for the large range of distortion types in more modern telecommunications technologies. Collectively, these disturbance densities, plus the quality estimates, are used to compute the final score, MOS for Listening Quality Subjective (MOS-LQS). Thirdly, and most notably, POLQA predicts speech quality over a wide range of bandwidths, including narrowband (100-3500 Hz), wideband (50-7000 Hz), and superwideband (50 Hz-14 kHz) signals.

A sound quality model based on the Moore-Glasberg (M-G) loudness model is the LPD index (Chen et al., 2006). In this index, reference and distorted signals are level-normalized and mapped into time frames and filtered using a gammatone filterbank. The resulting frame spectra are filtered through a series of bandpass auditory filters derived from the M-G loudness model (Glasberg & Moore, 1990; Moore & Glasberg, 2004). Auditory filter outputs generate excitation patterns from which loudness patterns are computed. The reference and distorted signals' loudness patterns are then compared to estimate the sound quality of a speech signal.

Another M-G loudness model-based index was proposed by Moore, Tan, Zacharov, & Mattila (2004), reported here as MQ-M. This index is characterized by a linear (D) and nonlinear (R) distortion measure. D is calculated on the basis of M-G auditory filterbank excitation differences between the reference and distorted signals over the long term spectra. R is calculated over short-term M-G auditory filterbank spectra of time-aligned, frame-by-frame representations between the reference and distorted inputs. Although the centre filter bandwidths cover 50-19739 Hz, the components between 500-5000 Hz are most heavily weighted. D and R are then asymmetrically combined to produce a final score of predictive quality, $S_{overall}$ with R more heavily weighted.

Hansen & Kollmeier (2000) proposed a speech quality model based on the Dau model of auditory processing (Dau, Puschel, & Kohlrausch, 1996), reported here as H-M. This model time-aligns the reference and distorted signals and uses a gammatone filterbank to filter them into 19 bands within the telephone bandwidth (350-3800 Hz). A 1 kHz low pass filter captures the temporal fine structure by coding the phase of the signal in the primary auditory fibres. Half-wave rectification allows the envelope of the signal above 1 kHz to be preserved. The index then accounts for dynamic compression and temporal adaptation in the auditory system by passing the signals through a 5-time nonlinear adaptation loop, in which rapid speech onsets are emphasized and sustained speech periods are not. The signal modulations are then low-pass filtered below 8 Hz, and a band importance weighting function is applied to the spectra to emphasize frequencies above 1 kHz. A comparison between the internal representations (characterized by spectra, time windows, and speech modulations) is used to compute the final score.

H-M was updated to PEMO-Q (Huber & Kollmeier, 2006), which has become one of the more popular quality models for evaluations of distorted speech and music. This new measure retains the structure of H-M except for several characteristics. Firstly, PEMO-Q can assess quality for wideband audio signals using 35 auditory filters (235–14500 Hz). Secondly, the band importance weighting function is omitted. Lastly, a modulation filterbank replaces the 8 Hz low pass filter. Like PESQ, an asymmetrical weighting is applied to the signals, so that added distortions are weighted more heavily towards the final score relative to missing elements. The internal representations are represented by time

windows, spectra and modulations, and are compared to produce the final perceptual quality measures (PSM and PSMt). PSM represents the overall linear cross-correlation between the reference and distorted signals, and is applicable for low to intermediate audio qualities. PSMt represents a comparison between the reference and distorted signals' top 5th percentile of instantaneous peak levels. PSMt is designed for estimates of high quality signals with small impairments.

Huber et al. (2014) extended H-M and PEMO-Q to account for the effects of hearing loss (PEMO-Q-HI) by inserting an instantaneous expansion and attenuation stage prior to the adaptation loops, determined by an input audiogram. PEMO-Q-HI interpolates a signal's frequency content using the model's peripheral filterbank, and the total hearing loss is attributed to both inner and outer hair cell loss. The remainder of the indices' structure remain the same.

PEMO-Q was further extended by Harlander, Huber, & Ewert (2014) to become the CASP-Q index. It was developed to account for a wider range of signal distortions and eventually effects of sensorineural hearing loss. This was done by replacing the Dau auditory processing model with the CASP model (Jepsen et al., 2008). The key update in CASP-Q is the replacement of the front end gammatone filter with a dual resonance nonlinear (DRNL) filter. The DNRL filter better mimics human auditory processing because it uses linear and nonlinear input/output level functions to simulate the passive and active mechanism of the level-dependent basilar membrane (Lopez-Poveda & Meddis, 2001). Other changes include an outer- and middle-ear transformation block prior to the auditory filter. Additionally, hair cell output is amplified and nonlinearly expanded. Finally, the integration of realistic hearing thresholds prior to the adaptation loops ensure that inaudible distortions do not affect quality predictions and that the model accounts for quality loss affected by elevated thresholds. Like PEMO-Q, CASP-Q output scores are represented by PSM and PSMt.

Finally, HASQI was developed to predict signal degradations caused by hearing aids including linear and nonlinear distortions and noise (Kates & Arehart, 2010). Its intended primary use was for hearing-impaired listeners. It has since become one of the most popular

quality models for evaluations of hearing aid signal processing. In this index, a reference signal is amplified via the NAL-R prescriptive targets for the listener's audiogram (Byrne & Dillon, 1986) and the test signal is processed using the settings of interest. The signal moves through a middle ear spectrum response and then filtered using a gammatone filterbank with centre frequencies from 80–8000 Hz. The filterbank outputs move through a level-dependent compressive block that simulates the potential outer hair cell loss of the basilar membrane, and then the signals are attenuated by any potential inner hair cell loss. The output of this cochlear model acts as internal representations for two intermediate quality predictors: Q_{linear} and $Q_{nonlinear}$. Q_{linear} addresses excitation differences between the reference and the signal under test. $Q_{nonlinear}$ addresses amplitude envelope by correlating short-term envelope fluctuation between the reference and signal under test for modulations rate below 125 Hz.

In HASQI v.2, $Q_{nonlinear}$ is expanded to also address temporal fine structure information by cross-correlating basilar membrane vibrations in each auditory band between the reference and signal under test (Kates & Arehart, 2014). The final $Q_{nonlinear}$ is the product of the square of the amplitude envelope measurement and temporal fine structure comparisons. The overall quality score, $Q_{combined}$ is the product of Q_{linear} and $Q_{nonlinear}$. HASQI v.2 outperformed its original model for a variety of real and simulated hearing aid distortions for both normal hearing and hearing-impaired listeners.

The HASQI was further expanded to evaluate music and non-speech signals affected by hearing aid distortions, and is known as the hearing aid audio quality index (HAAQI, Kates & Arehart, 2016). It is marked by three key differences from HASQI. Firstly, the $Q_{nonlinear}$ amplitude envelope is compared across a 32 channel modulation filterbank, rather than using the 125 Hz low-pass modulation filter seen in HASQI. Secondly, $Q_{nonlinear}$ is calculated by the sum, rather than product, of the amplitude envelope and temporal fine structure comparisons. Thirdly, the overall quality score, $Q_{combined}$ is a polynomial combination of Q_{linear} and $Q_{nonlinear}$, with nonlinear distortions weighted more heavily than linear distortions. HAAQI outperformed HASQI v.2 for music subjected to simulated linear and nonlinear hearing aid processing in normal hearing and hearing-impaired listeners (Kates & Arehart, 2016).

1.5.4 An appropriate objective model for hearing aid research

A predictive measure of hearing aid sound quality for clinical applications should possess several characteristics. The metric should correlate strongly with subjective sound quality ratings across as many distortions, stimuli, and populations as possible. While it is highly unlikely that any metric would achieve a perfect correlation coefficient with subjective data, a high correlation would suggest good predictive accuracy with a minimal margin of error. In contrast, a low correlation would suggest poor predictive accuracy, which would discourage the use of that metric. Furthermore, the metric should include a model for hearing impairment and a wideband spectrum analysis. The metric should be evaluated for a wide variety of distortions of speech and music including those created by hearing aids, and its output should be correlated with judgments by normal hearing and hearing-impaired populations.

One of the goals of this chapter was to assess the performance of the predictive measures by conducting a literature review of studies that validated objective quality models against subjective sound quality ratings, and then by calculating the average correlation coefficient across studies. Validation studies essentially process speech or music stimuli using a range of distortions, and produce a correlation score by comparing the objective index values with subjective sound quality judgments. The studies identified can be found in Table 1-2. To calculate the average correlation, coefficients were averaged over conditions for each study. Then, the average correlations for each study were averaged to a single value across all studies. This was done independently for normal hearing listeners, hearing-impaired listeners, and then across all listeners. Since cross-study comparisons are difficult to make due to different test conditions and listener populations, these values should be accepted with mild skepticism. However, they do provide a rough benchmark of each model's accuracy across as many validation studies as possible seen in the literature. Note that the technical models LLR, LAR and ISD (Klatt, 1982; Quackenbush et al., 1988) were omitted from the calculations. Because they are based on speech production, they may be sensitive to distortions attributed to the vocal tract that may be inaudible to the listener. Additionally, only PSM scores for PEMO-Q (Huber & Kollmeier, 2006) were used in the calculations due to their popularity when evaluating PEMO-Q. The narrowband, wideband, and

superwideband versions of POLQA were averaged together for its single study (Beerends et al., 2013), as was PESQ (Beerends et al., 2002) and PESQ-WB (ITU-T, 2007) across studies. Finally, M-Q-HI was omitted as it only predicts nonlinear distortions using the $R_{nonlinear}$ intermediate measure of M-Q (Tan & Moore, 2008), and therefore does not account for linear distortions.

Table 1-2: Objective sound quality models identified during literature search: Part 1

NH = normal hearing, HI = hearing impaired

Reference	Populations	Distortions	Stimuli	Measures
Arehart et al. (2011)	NH, HI	Linear, nonlinear distortions identified in Arehart et al. (2011)	Jazz, Haydn female vocal music	HAQSI
Beerends et al. (2002)	NH	Telephone-coding algorithms	Male, female speech	PESQ
Beerends et al. (2008)	NH, HI	Non-optimal presentation levels, circuit noise, single codec packet loss, bit error rates	Male, female speech	PESQ, PESQ-HI
Beerends et al. (2013)	NH	Clean, attenuation, circuit noise, codec packet loss, audio codec, bit rate errors, varied presentation level	Speech	POLQA
Chen et al. (2006)	NH	3-telephone coding algorithms	Male, female speech	LPD, PESQ
Creusere et al. (2008)	NH	Audio codecs from 8-64 kb/s	Rock, classic music	$PEAQ_{basic}$ $PEAQ_{advanced}$
Falk et al. (2015)	NH, HI	Hearing aid frequency lowering, directionality, speech enhancement, speech in noise, peak-clipping, low-pass filtering	Male, female speech	PESQ, HASQI, PEMO-Q, PEMO-Q-HI
Hansen & Kollmeier (2000)	NH	ETSI, ITU 8-kbit, ADPCM simulated, ADPCM real-net codec databases	Speech	H-M
Harlander et al (2014)	NH	Low-bit rate codec, audio source separation, noise reduction used in Hu & Loizou (2008)	Male, female speech	WSSD, PESQ, HASQI, PEMO-Q, $PEMO\text{-}Q_{(ISO)}$ CASP-Q

Table 1-2: Objective sound quality models identified during literature search: Part 2

NH = normal hearing, HI = hearing impaired

Reference	Populations	Distortions	Stimuli	Measures
Hu & Loizou (2008)	NH	Noise reduction, speech in noise	Male, female speech	WSSD, PESQ
Huber & Kollmeier (2006)	NH	Low bit rate codecs	Male, female speech, jazz music	PEMO-Q, PEAQ
Huber et al. (2014)	NH, HI	Hearing aid frequency lowering, peak-clipping, low-pass filtering	Male, female speech	H-M,H-M-HI, PEMO-Q, PEMO-Q-HI, PESQ, MQ-M HASQI, LPD
Kates & Arehart (2010)	NH, HI	Linear and nonlinear distortions in Arehart. et al (2010)	Male, female speech	HASQI
Kates & Arehart (2014)	NH, HI	Linear and nonlinear distortions in Arehart et al (2010), frequency lowering, feedback cancellation, modulated noise, noise vocoder reviewed in Kates & Arehart (2014)	Male, female speech	HASQI, HASQI v.2
Kates & Arehart (2016)	NH, HI	Linear and nonlinear distortions in Arehart et al (2011)	Jazz, Haydn female vocal music	HASQI v.2, HAAQI
Klatt (1988)	NH	Linear and nonlinear distortions	Synthetic vowels	WSSD
Kressner et al. (2013)	NH	Noise reduction used in Hu & Loizou (2008), speech in noise	Male, female speech	WSSD,PESQ, HASQI

Table 1-2: Objective sound quality models identified during literature search: Part 3. NH = normal hearing, HI = hearing impaired

Reference	Populations	Distortions	Stimuli	Measures
Moore et al. (2004)	NH	Linear and nonlinear distortions, simulated and real device distortions	Male, female speech, jazz music	MQ-M
Pourmand et al. (2013)	NH	Noise reduction, speech in noise	Male, female speech	PESQ, LPD HASQI, PEMO-Q, WSSD
Rodenburg et al. (2005)	NH	Noise reduction, speech in noise	Male, female speech	PESQ, PEMO-Q
Sulzle et al. (2013)	HI	Reverberation, noise, multitalker babble, four directionalities	Speech	HASQI
Treurniet & Soulodre (2000)	NH	Audio codecs from 64-192 kb/s	57 music timbres	PEAQ

A list of all average correlations and number of evaluations can be found in Table 1-3. The best average correlation was elicited by HAAQI at 0.945 for normal hearing listeners, 0.978 for hearing-impaired listeners, and 0.962 across all listeners. Note that this average is obtained from one study only. Additionally, this quality measure is specifically designed to measure quality degradations of music. The next best average correlation was elicited by HASQI v.2 at 0.904 for normal hearing listeners, 0.930 for hearing-impaired listeners, and 0.917 across all listeners. This average was taken across two studies. The poorest average correlation for normal hearing listeners was WSSD at 0.635 across five studies. WSSD was not included in any studies with hearing-impaired listeners. The poorest average correlation for hearing-impaired listeners was elicited by LPD at 0.56 for one study.

Table 1-3: Average correlation coefficient across validation studies including in review. Group correlations from each validation study were used when available. The coefficients listed above were averaged across each metric's respective studies for normal hearing listeners, hearing impaired listeners and all populations. If multiple coefficients were provided within a study, it would be average to a single value so that all studies would be weighted equally. N = number of studies over which average correlation was calculated

Metric	Normal-hearing Correlation (n)	Hearing-impaired Correlation (n)	All listeners Correlation (n)
CASP-Q	0.687 (1)	---	0.687 (1)
HAAQI	0.945 (1)	0.978 (1)	0.962 (1)
HASQI	0.799 (7)	0.884 (4)	0.776 (8)
HASQI v.2	0.904 (2)	0.930 (2)	0.927 (2)
H-M	0.879 (2)	0.630 (1)	0.821 (2)
H-M-HI	---	0.845 (1)	0.845 (1)
LPD	0.780 (3)	0.560 (1)	0.765 (3)
MQ-M	0.805 (2)	0.680 (1)	0.785 (2)
PEAQ	0.679 (4)	---	0.679 (4)
PEMO-Q	0.751 (4)	0.790 (2)	0.771 (5)
PEMO-$Q_{(ISO)}$	0.740 (1)	---	0.740 (1)
PEMO-Q-HI	---	0.888 (2)	0.888 (2)
PESQ	0.819 (9)	0.723 (3)	0.804 (10)
PESQ-HI	---	0.870 (1)	0.870 (1)
POLQA	0.922 (1)	---	0.922 (1)
WSSD	0.634 (5)	---	0.634 (5)

Another indicator of success is the number of validation studies performed, provided the average correlation is reasonably high. The frequency of validation studies for a metric is indicative of its popularity and generalizability over a wide range of distortions. PESQ was evaluated the most frequently, with an average correlation of 0.82 across nine normal hearing groups and 0.73 across three hearing-impaired groups, with a total of ten studies. HASQI was evaluated across eight studies, with an average correlation of 0.799 across seven normal hearing groups and 0.884 across four hearing-impaired groups. PEMO-Q was evaluated across five studies, with an average correlation of 0.751 across four normal hearing groups and 0.79 across two hearing-impaired groups. Its hearing loss extension (PEMO-Q-HI) improved the average correlation from 0.79 to 0.888 in the same hearing-impaired groups. This was consistent with other measures updated with hearing-

impairment extensions. For example, average correlations across hearing-impaired listeners improved from H-M, PEMO-Q, and PESQ to H-M-HI, PEMO-Q-HI, and PESQ-HI, respectively.

Finally, HASQI v.2, HAAQI, and POLQA notably elicited average correlations above 0.9. These promising results may reflect that more recent metrics are more technically developed and account for distortions that are more common in modern telecommunications and hearing aid technologies. However, to date, these metrics have only been evaluated in two or fewer studies. Therefore, it is not yet clear how these metrics generalize to other distortions not yet evaluated.

In summary, this section presents an overview of intrusive quality metrics in the literature that predict sound quality for a variety of distortions seen in the telecommunications and hearing aid literature. The results can be used to recommend possible models to be administered in clinical and research contexts. HASQI is among the most frequently evaluated quality measures, it provides a reasonably high average correlation across its validation studies, it incorporates wideband spectral analysis, and it can predict sound quality for both normal hearing and hearing-impaired populations. PEMO-Q fulfils these criteria as well, except that it was not developed with a hearing-impaired population. However, PEMO-Q-HI addresses this limitation and improves quality predictions relative to PEMO-Q for hearing-impaired populations. Similarly, PESQ-HI and PESQ-WB overcome PESQ's lack of wideband measurement and hearing-impaired modelling. Therefore, HASQI, PEMO-Q-HI, and PESQ-HI may be reasonable quality measures to incorporate into a clinical context. Furthermore, recent quality measures such as HASQI v.2, HAAQI, and POLQA exhibit exceptionally high correlations in a limited sample of evaluations. Because these metrics are relatively new, they have not yet been used to evaluate these quality measures for distortions not yet investigated.

HASQI, PEMO-Q-HI, and PESQ-HI are the most reasonable candidates for implementation into clinical and research contexts. Of these metrics, PEMO-Q-HI and PESQ-HI are proprietary to their developers and hard to obtain and embed into software. HASQI, however, is open-source code and freely available from its author (Kates &

Arehart, 2014), allowing simple integration into programming software like MATLAB and Octave. HASQI v.2 retains the structure of HASQI, while improving correlations relative to HASQI for the same signal processing parameters. Since HASQI v.2 is developed specifically for hearing aid distortions, accounts for hearing loss, improves and resembles HASQI (which is among the most frequently evaluated indices and most correlated with subjective judgments), and is open-source code, the HASQI v.2 was selected for further study, as will be described. More research is needed to evaluate its ability to predict hearing aid success using real hearing aids and additional measures based on real-world clinical outcome measures.

1.6 Purpose of the current research

The purpose of this dissertation is to explore factors related to sound quality for hearing aid music-listening. This document is divided into seven chapters, five of which consist of integrated manuscript-style chapters accompanied by an introductory literature review (Chapter 1)[1] and a large-scale discussion which integrates the findings across all dissertation chapters (Chapter 7). This purpose was investigated within the following three objectives:

1) The first research objective was to build on previous literature to understand how hearing aid users experience music-listening in real-world situations. While numerous surveys and research articles point to listener dissatisfaction and negative impacts of hearing aid use, few articles allow listeners to report concerns on their own accord. I addressed this objective using a qualitative methodology in which amateur musicians discussed their concerns regarding hearing loss, hearing aid use, and music-listening during a semi-structured interview (Chapter 2).
2) The second research objective was to explore implementation of objective sound quality metrics and perceptual measurements in hearing aid applications. Hearing aid users are frequently dissatisfied with the sound quality produced by their hearing aids. Unlike measures of speech intelligibility, sound quality can be measured for stimuli such as speech, environmental sounds and music. Unfortunately, standardized metrics of sound quality are not easily available to clinicians in practice. In this work, I investigated the utility of an objective sound quality assessment metric (Chapter 3) and discussed its potential for knowledge

[1] The introductory chapter of this dissertation is based on a literature review that was conducted between May and August of 2016. This literature review provided foundational knowledge that directed research decisions resulting in the findings presented in Chapters 2 through Chapter 6. Research published since 2016 may have altered or advanced findings since that time, but the research design adopted (i.e. selection of electroacoustics parameters, objective sound quality model selection, etc.) was based on the evidence available of that time. Research published since 2016 is discussed in Chapter 7 as it integrates with the new findings from this dissertation.

translation for clinical applications. I implemented aspects of this quality metric to assess whether dynamic range compression was related to sound quality ratings of real hearing aids (Chapter 5). I also implemented an advanced paired comparison strategy to perform a multiparameter optimization search strategy across different hearing aid frequency bands, and discussed its reliability (Chapter 6).

3) The third research objective was to identify electroacoustic changes in hearing aid signal processing that optimize hearing aid music sound quality. Many protocols exist for clinicians to facilitate the setting and fine-tuning of many hearing aid parameters for speech. These protocols are standardized and are a part of preferred practice guidelines. However, such protocols do not exist for music-listening. Therefore, without a comprehensive understanding of the literature, clinicians do not have access to guidelines for hearing aid optimization for music. In this work, I investigated changes in hearing aid signal processing using three strategies. The first strategy was to study sound quality differences between the universal and music programs in real hearing aids (Chapter 4). The second strategy was a top-down strategy, in which I evaluated whether any electroacoustic parameters varied across real hearing aids' default settings, and to determine which of those parameters, if any, were associated with good and poor sound quality (Chapter 5). The third strategy was a bottom-up strategy, in which I investigated whether adaptive variations in parameters identified in Chapter 4 could be used to optimize individual hearing aid settings relative to typical prescriptions for music-listening. (Chapter 6).

Taken together, these chapters will point to key tools, both electroacoustic and perceptual, that have utility in assessing the adequacy of hearing aid processing for music. Overall findings, limitations, and future directions are discussed in detail (Chapter 7).

1.7 References

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

Alexander, J. (2013). Individual variability in recognition of frequency-lowered speech. *Seminars in Hearing*, *34*(2), 86–109.

American National Standards Institute. (1997). *Methods for Calculation of the Speech Intelligiblity Index. ANSI S3.5-1997 (R2017)*. New York: Acoustical Society of America.

Amlani, A. M., Punch, J. L., & Ching, T. Y. C. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification*, *6*(3), 81–129.

Amlani, A. M., & Schafer, E. C. (2009). Application of paired-comparison methods to hearing aids. *Trends in Amplification*, *13*(4), 241–259.

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

American Speech-Language and Hearing Association. (2006). Preferred Practice Patterns for the Profession of Audiology [Preferred Practice Patterns]. Retrieved from www.asha.org/policy

Beerends, J. G., Hekstra, A. P., Rix, A. W., & Hollier, M. P. (2002). Perceptual evaluation of speech quality (PESQ) the new ITU standard for end-to-end speech quality assessment part II: Psychoacoustic model. *Journal of the Audio Engineering Society*, *50*(10), 765–778.

Beerends, J. G., Krebber, J., Huber, R., Eneman, K., & Luts, H. (2008). Speech quality measurement for the hearing impaired on the basis of PESQ. In *Proceedings of the 124th Convention of the Audio Engineering Society* (convention paper 7404).

Beerends, J. G., Schmidmer, C., Berger, J., Obermann, M., Ullmann, R., Pomy, J., & Keyhl, M. (2013). Perceptual Objective Listening Quality Assessment (POLQA), the third generation ITU-T standard for end-to-end speech quality measurement part II-Perceptual model. *Journal of the Audio Engineering Society*, *61*(6), 385–402.

Bentler, R., & Chiou, L. (2006). Digital noise reduction: An overview. *Trends In Amplification*, *10*(2), 67–82.

Bentler, R., Wu, Y.-H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: outcomes from laboratory and field studies. *International Journal of Audiology*, *47*(8), 447–460.

Boymans, M., & Dreschler, W. (2000). Field trials using a digital haering aid with active noise reduction and dual-microphone directionality. *International Journal of Audiology*, *39*(5), 260–268.

Brennan, M. A., McCreery, R., Kopun, J., Hoover, B., Alexander, J., Lewis, D., & Stelmachowicz, P. G. (2014). Paired comparisons of nonlinear frequency compression, extended bandwidth, and restricted bandwidth hearing aid processing for children and adults with hearing loss. *Journal of the American Academy of Audiology*, *25*(10), 983–998.

Byrne, D. (1994). An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America*, *96*(4), 2108.

Byrne, D., & Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new

procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, *7*(4), 257–265.

Chasin, M. (2006). Hearing aids for musicians. *The Hearing Review*, *13*(3), 1–11.

Chasin, M. (2012). Music and hearing aids: An introduction. *Trends in Amplification*, *16*(3), 136–139.

Chen, G., Parsa, V., & Scollie, S. (2006). An ERB loudness pattern based objective speech quality measure. *9th International Conference on Spoken Language Processing*, 2174–2177.

College of Audiologists and Speech-Language Pathologists of Ontario (CASLPO). (2014). Preferred Practice Guideline for the Prescription of Hearing Aids for Children. Retrieved from http://www.caslpo.com/sites/default/uploads/files/PPG_EN_Prescriptions_Hearing_Aids_Children.pdf

Cox, R. M., Matesich, J. S., & Moore, J. N. (1988). Distribution of short-term rms levels in conversational speech. *Journal of the Acoustical Society of America*, *84*(3), 1100–1104.

Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, *35*(5), e170–e184.

Dau, T., Puschel, D., & Kohlrausch, A. (1996). A quantitative model of the "effective" signal processing in the auditory system - I - model structure. *Journal of the Acoustical Society of America*, *99*(6), 3615–3622.

Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, *18*(8), 688–699.

Feder, K., Michaud, D., Ramage-Morin, P., McNamee, J., & Beauregard, Y. (2015). Prevalence of hearing loss among Canadians aged 20 to 79: Audiometric results from the 2012/2013 Canadian health measures survey. *Health Reports*, *26*(7), 18–25. Retrieved from http://www.statcan.gc.ca/pub/82-003-x/2015007/article/14206-eng.pdf

Franks, J. R. (1982). Judgments of hearing aid processed music. *Ear and Hearing*, *3*(1), 18–23.

Fulford, R., Ginsborg, J., & Greasley, A. (2015). Hearing aids and music : the experiences of D / deaf musicians. In *Proceedings of the Ninth Triennial conference for the European Society for the Cognitive Sciences of Music*. Manchester, UK.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033.

Gelfand, S. A. (2009). *Essentials of audiology* (3rd ed.). New York - Stuttgart: Thieme.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1–2), 103–138.

Glista, D., Scollie, S., Bagatto, M., Seewald, R., Parsa, V., & Johnson, A. (2009). Evaluation of nonlinear frequency compression: Clinical outcomes. *International Journal of Audiology*, *48*(9), 632–644.

Glista, D., Scollie, S., & Sulkers, J. (2012). Perceptual acclimatization post nonlinear frequency compression hearing aid fitting in older children. *Journal of Speech Language and Hearing Research*, *55*(6), 1765-1787.

Gnewikow, D., Ricketts, T., Bratt, G. W., & Mutcher, L. C. (2009). Real-world benefit from directional microphones. *Journal of Rehabilitation Research and Development*, *46*(2), 603–618.

Greasley, A. (2016). Effects of advanced hearing aid settings on music perception. Retrieved July 27, 2017, from http://musicandhearingaids.org/effects-of-advanced-hearing-aid-settings-on-music-perception/

Hansen, M. (2002). Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear and Hearing*, *23*(4), 369–380.

Hansen, M., & Kollmeier, B. (2000). Objective modeling of speech quality with a psychoacoustically validated auditory model. *Journal of the Audio Engineering Society*, *48*(5), 395–409.

Harlander, N., Huber, R., & Ewert, S. D. (2014). Sound quality assessment using auditory models. *Journal of the Audio Engineering Society*, *64*(5), 324–336.

Higgins, P., Searchfield, G., & Coad, G. (2012). A comparison between the first-fit settings of two multichannel digital signal-processing strategies: Music quality ratings and speech-in-noise scores. *American Journal of Audiology*, *21*, 13–21.

Hille, K., Gust, K., Bitz, U., & Kammer, T. (2011). Associations between music education, intelligence, and spelling ability in elementary school. *Advances in Cognitive Psychology*, *7*, 1–6.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, *49*(12), 891–903.

Honsby, B. (2004). The Speech Intelligibility Index: What is it and what's it good for? *Hearing Journal*, *57*(10), Vol 5, Issue 10. Retrieved from http://journals.lww.com/thehearingjournal/Fulltext/2004/10000/The_Speech_Intelligibility_Index__What_is_it_and.3.aspx#

Huber, R., & Kollmeier, B. (2006). PEMO-Q - A new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech and Language Processing*, *14*(6), 1902–1911.

Huber, R., Parsa, V., & Scollie, S. (2014). Predicting the perceived sound quality of frequency-compressed speech. *PloS One*, *9*(11), e110260.

Itakura, F. (1975). Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *23*(1), 67–72.

Itakura, F., & Saito, S. (1970). A statistical method for estimation of speech spectral density and formant frequencies. *IEICE Transactions on Electronic Communication of Japan*, *53–A*, 36–43.

International Telecommunications Union Radiocommunication Sector (2001). *Method for objective measurements of perceived audio quality*. *Recommendaion ITU-R BS.1387-1*. Geneva, Switzerland.

International Telecommunications Union Radiocommunication Sector (2015). *Recommendation ITU-R BS.1534-3: Method for subjective assessment of intermediate quality level of audio systems*. . *Recommendation ITU-R BS.1534-3*. Geneva, Switzerland.

International Telecommunications Union Standardization Sector (1996). *Methods for subjective determination of tranmission quality*. *Recommendation ITU-T P.800.* Geneva, Switzerland.

International Telecommunications Union Standardization Sector (2001). *Perceptual evaluation of speech quality (PESQ): An objective method fo end-to-end speech quality assessmnt of narrow-band telephone networks and speech codecs. Recommendation ITU-T P.862.* Geneva, Switzerland.

International Telecommunications Union Standardization Sector (2007). *Wideband extension to Recommendatioren P.862 for the assessment of wideband telephone networks and speech codecs*. *Recommendation ITU-T P.862.2.* Geneva, Switzerland.

International Telecommunications Union Standardization Sector (2014). *Methods for objective and subjective assessment of speech quality - Perceptual objective listening quality assessment*. *Recommendation ITU-T P.863* Geneva, Switzerland.

Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology*, *14*(7), 347–360.

Jepsen, M. L., Ewert, S. D., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *Journal of the Acoustical Society of America*, *124*(1), 422–438.

Johnson, E. E., Ricketts, T. A., & Hornsby, B. W. Y. (2007). The effect of digital phase cancellation feedback reduction systems on amplified sound quality. *Journal of the American Academy of Audiology*, *18*(5), 404–416.

Kang, R., Nimmons, G. L., Drennan, W., Longnion, J., Ruffin, C., Nie, K., … Rubinstein, J. (2009). Development and validation of the University of Washington Clinical Assessment of Music Perception test. *Ear and Hearing*, *30*(4), 411–418.

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

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

Kates, J. M., & Arehart, K. H. (2016). The hearing-aid audio quality index (HAAQI). *IEEE/ACM Transactions on Speech and Language Processing*, *24*(2), 354–365.

Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, *36*(2), 217–228.

Kirchberger, M., & Russo, F. A. (2016a). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, *20*, 1–16.

Kirchberger, M., & Russo, F. A. (2016b). Harmonic frequency lowering : Effects on the perception of music detail and sound quality. *Trends in Hearing*, *20*, 1–12.

Klatt, H. (1982). Predction of perceived phonetic distance from critical-band spectra: A first step. *IEEE International Conference on Acoustics, Speech and Signal Processing*, *7*, 1278–1281.

Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, *19*(6), 519–526.

Lopez-Poveda, E. A., & Meddis, R. (2001). A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, *110*(6), 3107–3118.

Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, *18*, 1–29.

Madsen, S. M. K., Stone, M. A., McKinney, M. F., Fitz, K., & 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.

Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *Journal of the Audio Engineering Sociaty*, *64*(3), 112–123.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, *32*(5), 556–568.

Moore, B. C. J., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, *188*(1–2), 70–88.

Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, *114*(1), 408–419.

Moore, B. C. J., Tan, C. T., Zacharov, N., & Mattila, V. V. (2004). Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion. *Journal of the Audio Engineering Society*, *52*(12), 1228–1244.

Mussoi, B. S. S., & Bentler, R. A. (2015). Impact of frequency compression on music perception. *International Journal of Audiology*, *54*, 627–633.

Neuman, A. C., Levitt, H., Mills, R., & Schwander, T. (1987). An evaluation of three adaptive hearing aid selection strategies. *The Journal of the Acoustical Society of America*, *82*(6), 1967.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, *7*, 1–5.

Parsa, V., Scollie, S., Glista, D., & Seelisch, A. (2013). Nonlinear frequency compression: Effects on sound quality ratings of speech and music. *Trends in Amplification*, *17*(1), 54–68.

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

Picou, E. M., & Ricketts, T. A. (2017). How directional microphones affect speech recognition, listening effort and localisation for listeners with moderate-to-severe hearing loss effort and localisation for listeners with moderate-to-severe hearing loss. *International Journal of Audiology*, *56*, 909–918.

Preves, D. A., Sammeth, C. A., & Wynne, M. K. (1999). Field trial evaluations of a switched directional/omnidirectional in-the-ear hearing instrument. *Journal of the American Academy of Audiology*, *10*(5), 273–284.

Quackenbush, S. R., Barnwell III, T. P., & Clements, M. A. (1988). *Objective measures of speech quality*. Eaglewood Cliffs, New Jersey: Prentice-Hall, Inc.

Revit, L. (2009). What's so special about music? *Hearing Review*, *16*(2), 12–19.

Ricketts, T. A. (2001). Directional hearing aids. *Trends in Amplification*, *5*(4), 139–176.

Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of*

Speech, Language, and Hearing Research, *51*, 160–172.

Ricketts, T. A., & Hornsby, B. W. Y. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction. *Journal of the American Academy of Audiology*, *16*(5), 270–277.

Ricketts, T., Henry, P., & Gnewikow, D. (2003). Full time directional versus user selectable microphone modes in hearing aids. *Ear and Hearing*, *24*(5), 424–439.

Sabin, A. T., Hardies, L., Marrone, N., & Dhar, S. (2011). Weighting function-based mapping of descriptors to frequency-gain curves in listeners with hearing loss. *Ear and Hearing*, *32*(3), 399–409.

Scollie, S., Glista, D., Seto, J., Dunn, A., Schuett, B., & Hawkins, M. (2016). Fitting frequency-lowering signal processing applying the American Academy of Audiology pediatric amplification guideline: Updates and protocols. *Journal of the American Academy of Audiology*, *27*(3), 219–236.

Scollie, S., Levy, C., Pourmand, N., Abbasalipour, P., Bagatto, M., Richert, F., … Parsa, V. (2016). Fitting noise management signal processing applying the American Academy of Audiology pediatric amplification guideline: Verification protocols. *Journal of the American Academy of Audiology*, *27*(3), 237–251.

Simmons-Stern, N. R., Budson, A. E., & Ally, B. a. (2010). Music as a memory enhancer in patients with Alzheimer's disease. *Neuropsychologia*, *48*(10), 3164–3167.

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in Amplification*, *6*(4), 131–165.

Spitzer, J. B., Mancuso, D., & Cheng, M.-Y. (2008). Development of a clinical test of musical perception: Appreciation of music in Cochlear Implantees (AMICI). *Journal of the American Academy of Audiology*, *19*(1), 56–81. https://doi.org/10.3766/jaaa.19.1.6

Surr, R. K., Walden, B. E., Cord, M. T., & Olson, L. (2002). Influence of environmental factors on hearing aid microphone preference. *Journal of the American Academy of Audiology*, *13*(2002), 308–322.

Tan, C.-T., & Moore, B. C. J. (2008). Perception of nonlinear distortion by hearing-impaired people. *International Journal of Audiology*, *47*(5), 246–256.

Thiede, T., Treurniet, W. C., Bitto, R., Beerends, J. G., Olomes, C. C., Keyhl, C. H., … Leidschendam, N. L. A K. (2000). PEAQ-- The ITU standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Sociaty*, *48*(1/2), 3–29.

Uys, M., & van Dijk, C. (2011). Development of a music perception test for adult hearing-aid users. *The South African Journal of Communication Disorders*, *58*, 19–47.

Vaisberg, J. M., Folkeard, P., Parsa, V., Froehlich, M., Littmann, V., Macpherson, E. A., & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline*, *Article 20782*. Retrieved from www.audiologyonline.com

Vaisberg, J. M., Martindale, A. T., Folkeard, P., & Benedict, C. (2018). A qualitative study of the effects of hearing loss and hearing aid use on music perception in performing musicians. *Journal of the American Academy of Audiology*, *Epub*, 1–15.

van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. *The*

Journal of the Acoustical Society of America, *105*(5), 2903–2913.

Wolfe, J., John, A., Schafer, E., Nyffeler, M., Boretzki, M., & Caraway, T. (2010). Evaluation of nonlinear frequency compression for school-age children with moderate to moderately severe hearing loss. *Journal of the American Academy of Audiology*, *21*(10), 618–628.

Zwicker, E. (1961). Subdivison of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, *33*(2), 248.

Chapter 2

2 A qualitative study of the effects of hearing loss and hearing aid use on music perception in performing musicians[2]

Hearing aids are important for the rehabilitation of individuals with hearing loss. While the rehabilitation of speech communication is well-understood, less attention has been devoted to understanding hearing-impaired instrumentalists' needs to actively participate in music. Despite efforts to adjust hearing aid settings for music acoustics, there lacks an understanding of instrumentalists' needs and if those hearing aid adjustments satisfy their needs. The purpose of the current study was to explore the challenges that adult HA-wearing instrumentalists face which prevent them from listening, responding to, and performing music. A qualitative methodology was employed with the use of semi-structured interviews conducted with adult amateur instrumentalists. Twelve hearing aid users who were amateur ensemble instrumentalists (playing instruments from the percussion, wind, reed, brass, and string families) and between the ages of 55 and 83 (seven men & five women) provided data for analysis in this study. Amateur in this context was defined as one who engaged mindfully in pursuit of an activity. Semi-structured interviews were conducted using an open-ended interview guide. Interviews were recorded and transcribed verbatim. Transcripts were analyzed using conventional qualitative content analysis. Three categories emerged from the data: (1) participatory needs, (2) effects of hearing aid use, and (3) effects of hearing loss. Participants primarily used hearing aids to hear the conductor's instructions in order to meaningfully participate in music rehearsals. Effects of hearing aid use fell within two subcategories: hearing aid music sound quality and use of a hearing aid music listening program. The effects of hearing loss fell within three subcategories: inability to identify missing information, affected music components,

---

[2] A version of this chapter has originally been published in the Journal of the American Academy of Audiology, Vol. 29, No. 10. Used with permission (see Appendix D).

and non-auditory music perception strategies. Not surprisingly, hearing-impaired instrumentalists face challenges participating in their music activities. However, while participants articulated ways in which hearing aids and hearing loss affect music perception, which in turn revealed perspectives towards listening using the auditory system and other sensory systems, the primary motivation for their hearing aid use was the need to hear the conductor's directions. These findings suggest that providing hearing-impaired instrumentalists access to musical experience via participation should be prioritized above restoring the perception of musical descriptors. Future research is needed with instrumentalists who no longer listen to or perform music due to hearing loss, so that the relationship between musical auditory deficiencies and participation can be better explored.

2.1 Introduction

The use of hearing aids (HAs) has always been associated with improved health-related quality of life (Chisolm et al., 2007; Contrera et al., 2016), including, but not limited to, "improvements in the social, emotional, psychological, and physical well-being of people" (Said, 2017). While the audiology community effectively understands and addresses listeners' needs for speech communication, less attention has been devoted to understanding hearing-impaired instrumentalists' needs associated with listening, responding to, and performing music. Even with a recent surge in studies investigating how hearing aid signal processing affects hearing music, there still is lack of understanding of hearing impaired instrumentalists' needs to meaningfully listen, respond to, and perform music while wearing HAs. The central focus of this study, therefore, was to explore hearing-impaired instrumentalists' perspectives towards HAs and music such that the audiology community can better cater to hearing-impaired instrumentalists' needs.

HAs have largely been developed with speech in mind rather than other complex auditory information such as music. This is intuitive, as hearing-impaired listeners' first complaint most often relates to speech understanding. Outcome-assessment tools, such as the speech intelligibility index (SII), and HA signal processing mechanics such as wide dynamic range compression (WDRC), frequency lowering, adaptive noise reduction (ANR), and feedback cancellation have primarily been developed to improve speech understanding. For

example, the SII is a metric used during HA fittings which predicts speech intelligibility through a HA using weighted speech-frequency regions that are audible to the wearer (Amlani, Punch, & Ching, 2002). WDRC compresses the speech output dynamic range by providing more gain for quieter sounds and less gain for louder sounds. Frequency lowering is an additional signal processing mechanism that targets high frequency syllabic information and lowers it to within the audible bandwidth for the listener (Alexander, 2013). Another additional feature, adaptive noise reduction (ANR), relies on detecting acoustic modulations typical of speech in order to suppress the relative level of background noise that may be interfering with the speech signal (Bentler & Chiou, 2006). Each of these features have been designed to enhance speech understanding. However, due to the differences in acoustic properties between speech and music (Chasin & Hockley, 2014), it is possible that these same features can have an adverse effect on the perception of music.

Several surveys have addressed hearing impaired listeners' music-related complaints by questioning respondents about HAs and music. Feldmann & Kumpf (1988) relate that 79% of their survey respondents reported that their hearing impairment interfered with music enjoyment, with complaints relating to understanding lyrics as well as pitch and melodic distortions. About two thirds of the respondents reported that HAs improved music listening, but that they still struggled to perceive rapid sound level changes. In 2008, Leek, Molis, Kubli, & Tufts found that almost 30% of their respondents were dissatisfied with music listening, attributing the largest challenges to sound level issues, and the authors attributed the reduction in complaints to advancements in HA technology over the two decades between the studies. Most recently, Madsen & Moore (2014) conducted a survey which specifically identified issues that HA users encountered listening to music. Overall, the most prominent problems identified were distortion, feedback, inappropriate gain, unbalanced frequency responses, and reduced tone quality.

In current HA fittings for music listening, clinicians are encouraged to disable the HA signal processing mechanics described above when fitting HAs for music listening (Moore, 2016; Zakis, 2016). Whether these signal processing adjustments are applied in practice, and if they are associated with improved music listening experiences, remains unknown.

More importantly, and underscoring the need for this current research, is that these surveys only address individuals' experiences listening to music, and not instrumentalists' experiences participating in and performing music.

The challenge when using surveys to understand listeners' needs is linked to music's holistic nature. Improving music is much more than removing negative auditory descriptors such as distortion, feedback, and reduced tone quality. Indeed, Small (1988), in his groundbreaking book, *Musicking*, debunks music as an object in that the "fundamental nature and meaning of music lie not in objects, not in musical works at all, but in action, in what people do" (p. 9). In the context of this research, then, music isn't simply the act of listening. Rather, music exists between and within the participatory relationships that are produced with sounds and others. Indeed, music can produce intrinsic enjoyment, emotional rewards and social fulfilment, among other benefits (Coffman & Ademek, 1999; Fulford, Ginsborg, & Goldbart, 2011). While these benefits may contribute to and enhance instrumentalists' experiences listening to, responding to, and performing music, these benefits may not necessarily be contingent upon removing undesirable auditory descriptors. Bartel et al. (2011) conducted a qualitative case study on cochlear implant users and music appreciation. One of their participants reported high enjoyment of music despite poor self-reports of auditory abilities including poor rhythm, tone, and timbre perception. Thus, while questions relating to music's auditory nature can shed light on the degradation of auditory perception due to hearing loss, these same questions might not relate to a listener's ability to achieve some of music's holistic benefits. James Strachan, former chief executive of the UK's Action on Hearing Loss charity and HA user, succinctly articulated the issue:

> "Hearing speech is a binary phenomenon: either you understand, or you do not. Whereas appreciating or enjoying music is a range phenomenon: just as I do not know how you see the color red, I do not know exactly what you hear when you listen to Adele or Beethoven." (Strachan, 2016).

This anecdote beautifully underscores the ways in which music is a complex and multifaceted concept that cannot be fully understood on the basis of direct auditory questions.

Other challenges and perspectives relating to music's holistic nature may not be accessible through quantitative surveys and are better revealed through the use of qualitative methods. Fulford et al. (2011) conducted semi-structured interviews with hearing-impaired musicians to determine the ways in which their musical experiences were impacted by hearing loss. In their sample, music self-efficacy was motivated by family encouragement from an early age, regardless of hearing status. Furthermore, participants reported a variety of listening styles, including reliance on auditory cues and other sensory and attentional cues as well. These findings point to fulfilling music-listening strategies that operated independently of challenges due directly to hearing impairment. Fulford, Ginsborg, & Greasley (2012) revisited their 2011 interviews to screen for reports related specifically to HAs. Several of the participants were dissatisfied with modern digital HAs and had complaints of distortion which lead to some participants rejecting HAs altogether. Other participants reported that HAs were simply able to give them access to music, using strategies such as a HA music program or adapting to their technology over time. Not only do these studies reveal that qualitative methods produce findings that are consistent with quantitative findings, but they also suggest that qualitative research methods reveal ways in which hearing-impaired musicians enjoy fulfilling musical experiences beyond auditory descriptors, which may not have been identified using quantitative methods. What remains to be known is the relationship between auditory impairments and fulfilling musical experiences, and whether one impacts the other.

The purpose of the current study was to explore the challenges that adult HA-wearing instrumentalists face that prevent them from listening, responding to, and performing music. Our broadly conceived research question was: How do adult instrumentalists report the impact of hearing impairment and HA use on music listening, responding, and performing, and on social participation in an instrumental setting? To that end, we employed conventional content analysis as a research method (Hsieh & Shannon, 2005) in order to analyze short oral histories collected from adult amateur instrumentalists participating in a local community wind band organization. The goal was thus to aggregate

their experiences in order to help explore prominent auditory deficiencies and challenges which inhibit musical participation that might be ameliorated through the use of HA technology and rehabilitation. Given the relative lack of literature reporting positive musical outcomes compared to positive speech outcomes following HA intervention, we chose to frame our research question around the challenges of musical participation so that our findings could set the stage for others to develop targeted rehabilitation strategies for music-based concerns.

2.2 Study Design

The researchers conducted semi-structured interviews with adult amateur instrumentalists to gain better insight into their experiences and perspectives related to the research question. Before addressing the design in more detail, descriptions as to who we are as researchers is necessary in order to frame the positionality and reflexivity found throughout this project. Authors JV[3] and PF have extensive experience in quantitative audiology research methods and relatively less experience conducting qualitative methodologies. PF has an interest in music personally and professionally as it relates to HA user satisfaction and digital signal processing challenges for those with hearing loss. JV has a background as an instrumentalist but studied formally only through high school. Author JV also has a strong background in music cognition theories and research methods from his undergraduate education. Author AM has more experience in qualitative audiology research methods than author JV. Although author CB has little to no experience of audiology, her experience in qualitative research methods, as well as her background in instrumental music performance and music education brought another level of expertise to the team. The diversity of this research team is such that the strength of the study only benefited from the multiple perspectives held between the four authors.

[3] Author initials refer to those listed in the Co-Authorship Statement.

During the interviews each researcher wrote memos that were referred to throughout the research process by each of the other researchers. Once the interviews were transcribed, conventional content analysis was used to describe, analyze, and synthesize categories that emerged from interview transcripts, personal memos, and collective insights. Through the analysis, key thoughts and/or concepts (Hsieh & Shannon, 2005, p. 1279) emerged as findings and categories were drawn from the data. A multi-disciplinary perspective between the researchers helped identify the emergent codes and categories (Hsieh & Shannon, 2005). Given the authors' experiences and the literature reviewed in this study, it was recognized that hearing loss and HA use has the potential to negatively impact music perception and performance. During the interviews, the authors generated dialogue using semi-structured questions and then probed areas of interest. This allowed the construction of categories that emerged as a result of interactions between the authors, the field, and the participants.

2.2.1 Sampling

Purposeful sampling was employed in order to yield "information rich" (Patton, 2002) data from knowledgeable participants. The study participants included in the final data analysis were selected because they had experiences that would "purposefully inform an understanding of the research" (Creswell, 2007, p 125). That is, participants with unique experiences of the phenomenon studied were sought so that the findings could arise from the data generated from the sample studied (Knudsen et al., 2012). Therefore, the authors aimed to recruit experienced hearing-impaired instrumentalists who could confidently articulate their perceptions of music listening and performance.

Participants were recruited from several sources. Initially, the study was advertised to an amateur ensemble band consisting primarily of older adults in London, Ontario, Canada. The majority of participants were retirees. Based on statistics collected by the Canadian Health Measures Survey, it was anticipated that a significant portion of these individuals would exhibit some degree of hearing loss due to age (Feder, Michaud, Ramage-Morin, McNamee, & Beauregard, 2015). Participants with known musical experience were also recruited from the National Centre for Audiology patient database, as were personal colleagues of the authors who fit the preliminary criteria.

2.2.2 Participants

A total of 54 participants were recruited and interviewed. For all participants, a detailed case history, pure tone audiometric thresholds (0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz), word recognition scores at a comfortable listening level, speech recognition thresholds, and tympanometric measurements were collected. One participant's interview was conducted via voice-over-internet-protocol (VoIP), and his most recent hearing assessment was faxed from his local audiologist. Among the 54 participants recruited, 49 presented with some degree of hearing loss (a threshold above 25 dB HL for at least one frequency). Among the 49 participants with hearing loss, 24 presented with a 3-frequency pure tone average threshold (PTA3) across 0.5, 1 and 2 kHz greater than 25 dB HL in at least one ear. There were a total of 15 HA users, 14 of whom had a PTA3 above 25 dB HL in at least one ear, and one whose PTA3 was below 25 dB HL in both ears.

After a brief initial review of the interview transcripts, it was clear that not all of the participants had sufficient musical experience to confidently articulate the phenomenon in question. The adult music group, from which most of the participants had been recruited, turned out to be for many, a place to begin music study at the most basic level. This meant they were engaging in the formal study of music for the first time in their lives: learning how to play a musical instrument; how to read music; and, how to respond to the conductor and other instrumentalists around them. In addition, a portion of the individuals among the 49 participants with hearing loss had minimal high frequency hearing loss. Some of these individuals did not find their hearing loss to impact their day-to-day lives, and as a result did not seek intervention. Therefore, the authors selected a subset of participants whom they believed to have sufficient hearing loss and musical experience for the purposes of this study. Their characteristics are described below.

A total of twelve participants' (seven male, five female) interview transcripts were analyzed for this study[4]. Participants ages ranged from 55 to 83 years of age (mean = 67.8, SD = 9.5) and each participant exhibited some degree of hearing loss. The majority of cases were sensorineural hearing loss ranging from mild to severe. There was one instance of moderately-severe mixed hearing loss with a mild conductive component. All participants wore HAs and had between two and 43 years of HA experience (mean = 18.9, SD = 16.0). In addition, all participants had at least four years of musical experience, with the majority exhibiting 40 or more years (mean = 32.7, SD = 19.9). Musical experience was defined as taking private instrumental lessons, experience performing in an instrumental ensemble, writing, arranging and producing musical content, or a combination of all of the above. A detailed breakdown of the twelve participants' hearing characteristics can be found in Table 2-1 and audiometric thresholds in Table 2-2. All participants were financially compensated for their participation in this study (including those whose interviews were not analyzed). This study was approved by the Western University Health Research Ethics Board.

[4] The author CB is using the full dataset (n = 54) to explore the degree to which amateur musicians (normal hearing or hearing-impaired) focus their listening attention on the conductor rather than the sound of the ensemble and the associated musical education implications.

Table 2-1: Participant characteristics: sex (M = male, F = female), age (years), music experience/ME (years), hearing aid experience/HAE (years), hearing loss type/HLT (SN = sensorineural, M = mixed), hearing aid style/HAS (BTE = behind-the-ear, CIC = completely-in-the-canal, RIC = receiver-in-the-canal, ITE = in-the-ear), binaural or monaural/BM (B = binaural, M = monoaural).

	Sex	Age	ME	HAE	HLT	HAS	BM
P1	M	83.4	4	25	SN	BTE	B
P2	F	60.4	>40	15	SN	CIC	B
P3	F	58.8	>40	2	SN	RIC	B
P4	M	75.5	>50	3.5	SN	ITE	B
P5	M	76.2	15	12	SN	RIC	B
P6	F	75.8	5	6	SN	RIC	B
P7	F	59.2	>50	4	SN	ITC	B
P8	M	72.1	40	4.2	SN	RIC	B
P9	M	74.4	5	4.2	SN	RIC	B
P10	F	56.4	42	31.5	SN	BTE	B
P11	M	55.8	40	43	SN	ITC	B
P12	M	65.5	31	37.5	M	BTE	B

Table 2-2: Participant audiometric thresholds. Audiometric thresholds in decibels hearing level. NR= no response.

		Frequencies (kHz)							
Ear		0.25	0.5	1	2	3	4	6	8
Left Ear	P1	60	65	75	75	80	85	85	NR
	P2	50	55	60	60	60	60	65	75
	P3	10	15	20	45	50	45	30	20
	P4	20	20	25	70	65	70	70	75
	P5	30	35	50	50	55	80	105	NR
	P6	30	25	40	45	50	50	60	75
	P7	15	15	10	30	35	30	40	55
	P8	50	45	45	65	60	65	75	80
	P9	25	30	20	30	45	50	55	55
	P10	70	75	70	60	60	60	55	65
	P11	45	55	65	75	85	85	80	75
	P12	75	70	75	65	55	60	65	65
Right Ear	P1	55	50	45	55	65	65	75	80
	P2	70	60	65	55	60	65	65	80
	P3	0	5	15	40	50	45	30	20
	P4	20	20	20	60	70	70	75	NR
	P5	35	35	50	45	50	75	90	NR
	P6	20	30	45	55	50	50	45	70
	P7	30	20	20	30	30	25	45	45
	P8	25	25	35	55	65	75	85	85
	P9	35	30	25	35	55	60	60	65
	P10	80	80	75	65	65	65	65	65
	P11	50	55	65	65	75	75	95	85
	P12	75	85	80	75	55	55	65	50

2.2.3 Data Collection

The semi-structured interviews with the participants ranged in length from 20 to 60 minutes and were completed in a single session in a quiet laboratory at the National Centre for Audiology either in person or via VoIP. To minimize researcher bias and to maximize interviewer sensitivity, authors JV, PF, and CB took part in developing the interview guidelines. Together, they developed an interview guide, loosely adapted from Leek et al.'s (2008) telephone survey questionnaire investigating hearing-impaired listeners' enjoyment of music. The interview guide was designed to encourage discussion topics ranging from when the participant identified their hearing loss, how long they have participated in music ensembles, whether sound quality was affected, and the effectiveness of their HAs. From the beginning of the study, the researchers were aware of the need to situate biases and beliefs even in the development of the interview questions (Berger, 2015). While semi-structured interviews permitted the use of open-ended questions, there was also the need to provide prompts so that participants could provide and elaborate on accounts of their experiences of the phenomena under study (Knudsen et al., 2012). Two of the researchers had experience as performing instrumentalists, and two were experienced in audiology procedures, thus we were constantly cognizant of our abilities to both promote and perhaps hinder on-going dialogue between the interviewers and participants. After discussion and reflection between the researchers, the topics of interest that emerged included HA listening habits, HA program use, music components such as timbre, dynamics, and melodic recognition and social participation goals and needs. The interview guide was piloted with several participants to verify that it effectively promoted dialogue, encouraged meaningful contributions based on each author's background, and allowed for descriptions to emerge. An example of interview questions can be found in Table 2-3.

Table 2-3: Interview guide used to prompt dialogue during semi-structured interviews

1. Has the sound quality of music changed since you acquired hearing loss?
2. How long have you been playing your instrument?
3. Has your enjoyment listening to and playing music changed since you identified hearing loss?
4. Have you changed the way you listen to/play music? (ex. with headphone use?)
5. How would you describe music in general sounds to you? (ex. tinny, bassy, distorted, too loud, too soft, etc.)
6. Are there specific musical elements you have difficulty with? (ex. timbre, dynamics, melody, intonation, rhythm, harmony, etc.)
7. Do you think there is musical information you are missing when you listen?
8. Are you emotionally moved by music, and if so, how?
9. Why did you start wearing your hearing aids?
10. What do you find useful about your hearing aids?
11. If you have a multiple memory hearing aid, what program do you use for music and why?
12. How would you change the hearing aid to improve the sound quality of music?

Interviews were conducted by authors JV, PF, and CB. Participants chose interview times that were convenient for them. Many of the first round of participants had never had a hearing assessment or visited an audiology laboratory. Recognizing that we could not predict how participants would respond to perceived social, professional, and educational positioning (Berger, 2015; Finefter-Rosenbluh, 2017) great care was taken to make the interview as comfortable as possible. Thus, participants were greeted by one of the authors in the reception area, walked through the building to the lab, and then offered a beverage as they settled in. As described above, each participant then provided a detailed case history and completed an audiological assessment prior to their respective interview.

The data consisted of transcripts and memos collected by the interviewers during the interviews. Interviews in person were recorded using Audacity (version 2.0.6) software using a built-in laptop microphone, and interviews via VoIP were recorded using Skype (version 7.37) software. All files were converted to MP3 format and confidentially transcribed verbatim by a third-party transcriptionist. The authors then verified that the transcripts' contents were consistent with the memos that had facilitated the tracking of ideas and concepts throughout the duration of the research study. While it was important to verify consistency between the interview transcripts, the interviewers' impressions of the transcripts, and the final report (Van Den Hoonard, 2012), the researchers were keenly aware of possible bias. To minimize bias, the researchers, throughout the process, regularly met to cross-check their responses to what was heard in the transcripts, what was remembered in the moment of the interviews, and the notes taken in the form of memos.

2.2.4 Analysis

The authors analyzed the data using conventional qualitative content analysis, as outlined by Hsieh & Shannon (2005). First individually, and then throughout multiple meetings, the researchers thoroughly coded the data for emerging themes that spoke to each of us. However, authors JV and CB had a working "start list" (Miles & Huberman, 1994, p. 58) of themes they suspected might emerge based on their diverse personal, professional, and disciplinary backgrounds. Since author AM had less background as an instrumentalist, she did not begin with a "start list." Rather, she read for themes using her background as a

qualitative audiology researcher and checked for consistency with the data analyzed by authors JV and CB.

After reflection and discussion, we determined a set of themes upon which we were all in agreement. Once this set of themes was decided upon, the authors continued reading for other instances which could be coded into the same themes. Once trending ideas among the codes emerged, the authors formed categories, or broader ideas representing a grouping of codes, consisting of multiple participants' perspectives (Miles & Huberman, 1994). The coding process was repeated for each category. Throughout this process, the researchers were also describing and framing the categories in support of the research question. To exemplify: One participant reported, "Well I think you're going to have more trouble hearing music that's very quiet, for sure." This was coded as "difficulty with very quiet sound": which was subsequently grouped into the category "dynamics" as part of "affected musical components." Another participant commented, "When I was playing [with HAs], certain notes I would hit and I would get feedback." This was coded as "certain musical notes create feedback," and was eventually grouped into the "HA sound quality" category. Both these categories represent challenges that hearing-impaired instrumentalists encounter.

The authors aimed to maintain trustworthiness throughout the analysis. Trustworthiness in qualitative research has been considered analogous to validity and reliability in quantitative research (Golafshani, 2003). Trustworthiness consists of multiple components such as credibility, transferability, and dependability (Guba, 1981; Knudsen et al., 2012; Shenton, 2004; Sikolia, Biros, Mason, & Weiser, 2013). Credibility was achieved by coding data from various sources: interview transcripts, memos, case histories, and the authors' impressions. Dependability was achieved in the current study by reflecting upon and discussing emerging categories at each phase of the coding process. While transferability is not a stated goal of qualitative research, it can occur when some or all of the study findings can be transferred to another similar context (Guba, 1981). We are hopeful that transferability of this study can be achieved due to the trustworthiness of the description of the study, the presentation of the data, and anticipated consistency of the data with other research studies. We also anticipate that our findings relating to instrumentalists and their

experiences with hearing loss could likely be transferred to some degree when studying other kinds of instrumentalists, possibly within other cultural practices, in other geographical areas, and/or of varying sex and age who also have hearing loss.

2.2.5 Reflexivity

As a qualitative methodology was deemed the most appropriate for this study, it is necessary to recognize the issue of subjectivity in the research process. Reflexivity refers to an awareness of subjectivity, or more specifically, of how the authors' own presences influenced the research process (Barry, Britten, Barber, Bradley, & Stevenson, 1999), and is considered an essential component in qualitative research (Watt, 2007). Historically, research in the sciences aimed to rid elements of bias and subjectivity from research designs (Wilkinson, 1988). This stance, however, has been challenged by social psychologists, feminist theorists, and critical race scholars (Gough & Madill, 2012, p. 374). In this study, the authors not only embraced the strengths and possibilities embedded in intersubjectivity but understood the impossibility and falseness of claiming a completely objective stance. In this study, the authors' subjectivity, in essence, was "reviewed as a resource that [was] tapped in order to contextualize and enrich the research process and its products" (p. 375). The authors also sought throughout the process, and now here in this article, to make explicit and build on their "understanding, positions, and approaches" (Gentles, Jack, Nicholas, & Mckibbon, 2014, p. 3) in order to address not only their interactions with the participants, but interactions amongst themselves, their distinct influence on how the data were viewed, and even the influence each of them may have had on the others.

Throughout the study, as has been previously articulated, the researchers met at various times during and after data collection to think out loud as to their reactions and perceptions of the engagements of the participants and other issues found in our memos and transcripts. During meetings, the researchers discussed personal memos, perceptions of the data, and interview transcripts and looked for intersections and commonalities among them. Perhaps seemingly peripheral to this study are the conversations we had that on focused interdisciplinary connections based on valuing and values that were distinct to our disciplines. The kinds of questions we were interested in pursuing became one of the more

powerful focal points as we become aware of our internal reflexivity and we became more comfortable with embracing and sharing that which we thought we knew and that which we came to find we did not. While this study is not categorically labeled "interdisciplinary," we do hail from different disciplines, thus, the recognition of the kind of language each of us used, and our epistemological stances, not just in the interviews, but with each other, helped to keep this study rigorous. Indeed, personal, professional and disciplinary reflexivity (Wilkinson, 1988) was not only present throughout, but facilitated the entire arc of the research process, as well as our own growth and transformation as scholars, researchers, and practitioners.

2.3 Findings

Even to those not grounded in audiology research or the performing arts, it is not surprising to state that one of the central findings that emerged was that hearing-impaired instrumentalists encounter challenges participating in their music activities. Three categories (themes) emerged that help exemplify the findings. The first and most prominent category consisted of the participants' participatory needs; hearing loss mostly interfered with their ability to hear the conductor, which they believed to be necessary in order to participate during rehearsals. The second category consisted of the participants' impressions about HAs: sound quality was influenced by HAs and satisfaction using a HA music program. The third category consisted of the effects of hearing impairment on music perception and included the following subcategories: missing auditory information, affected music components, and non-auditory music perception strategies.

2.3.1 Participants' participatory needs: Hearing the conductor

In his research with adult musicians in a community organization, Jutras (2011) categorizes the benefits of participating in an adult New Horizons Band in Rome, Georgia, finding that skill-related and social/cultural were the two most frequently identified reasons for participation. These findings resonate with this study in that participation did comprise "social interaction, social relationships, and socialization" (p. 67) and skill development, including "skill improvement, skill refinement, technique, musicianship, music theory, music listening, and musical knowledge" (p. 67).

However, our research differs from the Jutras study in that teasing out reasons for participation was peripheral to our goal of understanding from the adult instrumentalists' perspective what it was like to listen, respond to, and perform music. Participation was a given as each of the participants were actively playing in an ensemble(s) of some kind. Thus, one assumption going into this study was that these instrumentalists' motivation to wear HAs would be to better hear and discriminate their music making. However, one of the more prominent difficulties reported by participants in an ensemble was being able to hear the conductor's directions during rehearsal. This was the primary reason that participants chose to wear their HAs.

> "One of the reasons I started to get HAs was so I could just hear [the conductor] while playing."

> "I had to keep my HAs in so I can hear [the conductor], like, when she's making comments and things like that."

Some participants reported that hearing the conductor was the *only* challenge related to their hearing loss and HA use, and that they did not actually experience any difficulties perceiving music itself. The following are examples of these sentiments.

> "Without the HAs, I can't really say, because I have to have my HAs in, you know, in order to, you know, hear the conductor."

> "And then my concern was well, should I wear them when I'm the band, will it be too noisy? But then you've got to hear what the conductor's saying, so I wear them now."

Thus, interestingly, rather than using HAs to hear what is happening musically so that they may respond musically, listeners' primary motivation to wear HAs during rehearsal appears to be listening to and understanding the conductor's instructions. For some, this was the only motivation to wear HAs.

2.3.2 Participants' impressions of hearing aids

Despite the need to understand the conductor, the use of HAs would inherently have some sort of impact on the acoustic content of music processed by HAs. Two subcategories emerged related to impressions of HAs and their effect on music. The first concerned how HAs affected sound quality of music. The second related to the use of a HA music program. Some participants briefly commented on what they believed would improve music listening through HAs, although this was not grouped into a separate subcategory.

The participants expressed highly variable opinions regarding the impact of HAs on music sound quality. Some participants expressed positive views, "When I have the HAs in, the clarinet is louder…it seems brighter and sharper than without." Other participants expressed negative views, "I found that I got the real quality, you know, the real actual feel of the music without my HAs." Some participants also had neutral opinions, "I don't think the HAs make a lot of difference." These examples portray considerable variability in satisfaction of HA-amplified music across individuals.

One participant expressed that previous analog HAs provided better music sound quality compared to more recent digital HAs, because the digital aids limited the amount of acoustic information that was amplified: "My best sounding set of HAs ever were analog. They had no bells or whistles. It was just straight gain and these had a very extended range. I had a big problem with digital aids...because of the hard cap." This anecdote associates fidelity of sound quality with signal processing schemes found in analog HAs, compared to digital HAs, and could, with more research, direct researchers to replicate analog processing strategies in modern digital systems.

Participants were further asked if they had experience using a HA music program[5] to improve music sound quality. Some participants reported no benefits when using a music program, "the music settings…just don't provide any benefit to me, I don't see any

[5] A HA music program is a set of HA processing parameters adjusted with the aim of optimizing a music signal, although the exact adjustments vary across manufacturers.

difference." Other participants reported never using a music program. Some participants found a music program to be helpful because it improved the balance and brightness of sounds. However, one of them mentioned that the effort to change the HA setting was not worth sacrificing the convenience of leaving their HA at one setting, "I think [a music program] improves [music listening], but for me it's…with these HAs it's just easier to leave it set at one thing." These statements suggest that music programs are ineffective at improving music sound quality relative to a typical HA program, and even when they are, the relative improvement in sound quality is not worth the effort to change the HA settings.

When asked what it would take to build a better HA, most participants were unsure what would improve music sound quality. However, there were a few characteristics mentioned. One participant suggested that a wider frequency bandwidth would improve the response. This is a reasonable suggestion, as most HAs amplify only between 200 Hz and 6 kHz, despite optimal speech and music sound quality being associated with wider bandwidth in both high and low frequencies (Moore & Tan, 2003). Other participants suggested that background noise could be lowered relative to the signal of interest. This suggestion is consistent with current technologies, as ANR systems are capable of lowering the background noise level without noticeable speech sound quality degradations (Bentler, Wu, Kettel, & Hurtig, 2008; Scollie et al., 2016). It is worth investigating the effect of ANR systems on music stimuli. In addition, another participant wished that loud and soft sounds could be more effectively balanced. This last suggestion may be interpreted as a dynamic range issue. This is no surprise, as HAs are typically built for the dynamic range of speech, while the dynamic range of live music is much greater (Chasin & Russo, 2004). Together, these findings show that most listeners do not consider strategies to improve the sound quality of music using their HAs. However, those who do consider strategies share insights that are consistent with evidence found in the literature.

2.3.3 Effects of hearing loss on music perception

While the use of HAs certainly impacted music sound quality, we were also interested in how HAs and hearing impairment impacted specific aspects of music. We therefore asked participants more targeted questions in these areas. The participants' responses were grouped into three subcategories: awareness of missing information; affected musical components; and multisensory music perception.

When asked if something about a musical signal was affected or missing, many of the participants responded that they thought something was missing. However, they were unable to identify exactly what it was. In fact, some participants reported being unsure what the music was exactly supposed to sound like: "…how do I know what it should sound like? So I just listen to it according to my hearing deficiency, whatever, not knowing what the real thing might be." Another participant said, "with my hearing, and with wearing HAs, you don't know what you're missing." Some participants described that they did not attempt to identify missing sounds until it was brought up in the interview. When asked about what instruments they may not hear, one participant expressed, "…it's one of those things where I haven't sat down and tried to put my finger on." Some participants reported that they have had hearing loss for so long that they could not remember what "normal" music sounded like. Their sense of normal had implicitly changed and they were not able to describe how music "should" sound. These reports suggest that hearing-impaired listeners generally suspect they are missing musical information. However, they do not consider missing musical content to be a significant concern, nor do they find it a barrier to musical participation.

In order to resolve what musical information might have been missing, participants were probed about specific music components. The components most frequently discussed during the interviews were dynamics, intonation, melody, and timbre. During the interviews, the authors defined dynamics as the relative contrasts of loud and soft levels of music. Intonation was defined as the realization of pitch and whether or not the pitch is in tune. Melody was defined as the principal succession of pitches in a musical composition. Timbre was defined as the characteristics of the sound which allowed the listener to identify what instrument is playing.

Many participants expressed concerns with musical dynamics as a result of hearing loss, HA use, or both. They described having difficulty identifying and contrasting loud and soft occurrences. Some participants expressed difficulty perceiving and playing soft music relative to loud music: "…you're going to have more trouble hearing music that's very quiet." Additionally, another participant mentioned: "the softer instruments … [are] probably harder for me to hear." Other participants suggested that dynamics were easier to perceive without HAs: "in fact it's probably more accurate without my HAs." These reports suggest that hearing loss and HA use may make differentiating and performing loud and soft contrasts in music passages more challenging.

With respect to intonation, several participants reported positive experiences staying in tune despite their hearing loss and HA use once they were proficient at their instrument. For example, one participant reported, "I'm rarely very far out of tune." However, another participant suspected that difficulties with intonation are not crucial for overall music perception, "as far as hearing something in tune or out of tune…I'm not sure those are crucial things." These reports suggest that intonation is an aspect that amateur musicians may not associate with challenges due to hearing impairment and HA use, especially with sufficient training on their chosen instrument.

Melodic recognition was particularly challenging for almost half the participants. They expressed that it was difficult to identify the melody if there was too much noise, if there were too many instrumental parts, or if the melody was playing particularly quietly. One participant expressed difficulty following the melody "especially when it's a softer sound." Another said, "I can't do it if they're all played at the same time, it just sounds like mush." This is consistent with previous surveys in which HA users struggled to listen to musical lines in layered ensembles relative to solo instruments (Leek et al., 2008; Madsen & Moore, 2014). These findings suggest that melodic recognition can be affected by other musical parts and noise generated in a rehearsal space. Background noise, whether it is related noise or musical layering, appears to worsen participants' abilities to recognize melodies. This recognition challenge may be interpreted as being analogous to difficulties understanding speech in noise for hearing-impaired listeners.

Finally, many participants articulated positive experiences related to instrumental timbre discrimination and identification. When asked if they could discriminate and identify instrumental timbres, many participants were confident that they could: “I could pick out the oboe from the clarinets [when listening to a performance].” Participants also reported various acoustic cues that helped them identify different timbres. Such cues included the register in which the instrument plays, the intensity that the instrument typically produces, and the quality of sound. For example, one participant said, “I can hear [the piccolo], but that’s just because you’re an octave higher than anyone.” Difficulties attributed to instrumental timbre perception were related to these cues. Some participants reported difficulty discriminating instruments if they played in a similar register or at a similar level. These findings suggest that hearing-impaired listeners easily discriminate various timbres and rely on acoustic cues to do so. While timbre discrimination and identification deficits have been identified in listeners with moderate-to-severe hearing losses (Emiroglu & Kollmeier, 2008; Looi, McDermott, McKay, & Hickson, 2008; Uys & van Dijk, 2011), deficits have been more variable between individual hearing impaired listeners exhibiting mostly moderate flat losses (Kirchberger & Russo, 2015). Given that a majority of participants in the current study presented with moderate flat hearing losses, these previous quantitative findings are in agreement with some reports identified here.

Given the impact of hearing impairment and HA use on music perception, we asked participants about using non-auditory senses to supplement the musical experience. Some participants expressed that they believed their somatosensory system could supplement their auditory system when performing music in an ensemble whether or not they were hearing impaired or whether or not they were wearing HAs. That is, they could perceive parts of music through touch responses and vibrations against and within their body to enjoy music and to monitor musical intonation. With respect to enjoying music, one participant mentioned, “in a hearing-impaired person, feeling is just as valid a method of hearing as audio perception is.” With respect to monitoring musical intonation, a participant said, “I know which pitches on a flute tend to play flat or sharp, so I adjust…I can feel the vibrations and that’s very helpful to me.” This is also consistent with the participant above who felt that correct intonation comes from “where they feel it.”

While it is not clear if the vibrotactile sensations supplement the perception of some of the musical components mentioned, it is possible that the sensations contribute in some way to music enjoyment beyond the auditory experience. Music has previously been described as a multisensory phenomenon which integrates stimuli from a variety of sensory systems on the basis of cortical evidence in multisensory regions (Zimmerman & Lahav, 2012). The behavioral anecdotes reported here are supportive of the multisensory hypothesis.

2.3.4 Consistency across other literature

Many of the HA-related auditory concerns reported in this study are consistent with findings obtained from both qualitative and quantitative approaches to the effects of hearing impairment and HA use on music perception in other studies. These consistencies support the possible transferability and confirmability of this study's findings, suggesting a trustworthy dataset (Knudsen et al., 2012). Qualitatively, these results were consistent with findings reported by Fulford et al. (2011, 2012). In both studies, participants exhibited attitudinal ambivalence in that descriptors about HAs were both positive and negative. Positive descriptors consisted of participants describing HA amplified music as "brighter" or "crisper." The descriptor "brightness" is considered a positive dimension of sound quality and is associated with a modest increase of the treble portion of the frequency response (Gabrielsson & Sjögren, 1979). Negative feedback consisted of sound quality descriptors included "screeching" and "tinny." The descriptor "tinny" is a common complaint for HA users and is associated with too much gain in the high frequencies (Jenstad, Van Tasell, & Ewert, 2003). The descriptor "screeching" may be related to "squealing", which is often used to describe a distortion known as acoustic feedback. Issues of feedback and unbalanced frequency responses have been quantitatively identified as concerns for HAs music (Madsen & Moore, 2014). The HA music program's inability to improve music listening also trended across multiple studies. In our study, listeners were generally indifferent about a music program's efficacy. Some participants in Fulford et al. (2012) were unsure of whether or not they had a music program, and those who did have a music program did not use it consistently. Madsen & Moore (2014) reported music satisfaction scores that were similar from both users and non-users of a music program, suggesting that the music program did not significantly affect music sound quality.

Vaisberg et al. (2017) found that only two out of five HAs' music programs improved music sound quality, and that the magnitude of improvement was less than the variation across HAs. Together, these results may indicate that further improvement in music programs may be desirable. The last consistency between our study and others concerned preferences between analog and digital HAs. One of the current participants preferred legacy analog HAs relative to modern digital aids. This preference was also found among listeners interviewed by Fulford et al. (2012). This finding made sense, as analog HAs provide mostly linear amplification, which some studies have found improves music listening compared to the WDRC commonly provided in today's digital HAs (Arehart, Kates, & Anderson, 2011; Croghan, Arehart, & Kates, 2014; Higgins, Searchfield, & Coad, 2012; Kirchberger & Russo, 2016; van Buuren, Festen, & Houtgast, 1999).

Many of the auditory concerns due to hearing impairment were also consistent with quantitative literature on similar topics, as discussed above. However, these concerns also related to the dynamic listening styles reported by Fulford et al. (2011). The fact that many of the current study's participants articulated the ways in which their auditory experience was affected highlighted some degree of reliance on hearing for musical participation. However, participants frequently discussed supplementing their hearing by using non-auditory attending strategies, such as vibrotactile sensations to perceive musical intonation, as did participants interviewed by Fulford et al. (2011). Both auditory and non-auditory listening styles allow listeners to negotiate concerns caused by hearing impairment and distorting effects due to HAs (Fulford et al., 2011) and may therefore be considered valid methods of perceiving music.

2.4 Discussion

The purpose of the study was to explore the challenges that adult HA-wearing instrumentalists face which prevent them from listening, responding to, and performing music. The following is a discussion of the findings drawn from the participants' interviews. The categories that emerged are considered as they relate to the research questions and the review of literature. The three main categories were: participatory needs, impressions of HAs, and effects of hearing impairment on various aspects of music perception. The most predominant participatory need was connected to hearing the

conductor. Impressions of HAs were inferred based on how sound quality was affected by HAs and satisfaction using a HA music program. Aspects of music perception included the subcategories: missing auditory information, affected music components, and non-auditory music perception strategies.

Due to our extensive background in audiology research, we were interested, to some degree, in exploring auditory deficits reported by the participants. However, we were also focused on understanding the challenges which instrumentalists face, and the relationship between those challenges and the holistic nature of music perception. Therefore, a qualitative methodology effectively afforded participant the opportunity to share auditory (and participatory) deficits that were important for a holistic music experience. While the participants certainly did discuss concerns related to their auditory experiences, the majority of them first expressed listening needs related to their ability to participate in a musical ensemble. This was a notable finding, as participants were not directly asked about participatory issues prior to auditory issues. The fact that participants discussed participatory needs prior to auditory concerns, without even articulating the connection between them, was an interesting result that was supportive of the purpose of this study.

Participation and participatory needs are not newly expressed phenomena in musical engagements, yet it is an idea that has been too often assumed and taken for granted. Over 25 years ago, Gates (1991) encouraged music scholars to better define participation. He defined what he referred to as a "Typology of Music Participants in Societies" and suggested that participation can be typed as work, serious leisure, and play (p. 16). He underscored that most research to that point had been done by scholars using "positivist research paradigms and quantitative data gathering strategies" (p. 15) and suggested "[getting] beyond [a] surface level of categorization" (p. 17). Since this 1991 article, numerous studies on participation have been conducted, and for the purposes of this article we consulted "participation" studies that have taken place with older adults. For instance, Dabback (2008) discussed the importance of structure, health and well-being that musical organizations provide to adults. He specifically addressed the ways in which music engagement may provide continuance of a musical identity that may have been formed during childhood. He also, however, discovered the importance adult musicians place on

the opportunity to reclaim and develop new musical and social identities. Coffman, a researcher in the area of music education for adults, has, for the most part, primarily relied on quantitative methods to address issues based on (among others) intergenerational engagements (Coffman & Levy, 1997), quality of life, well-being and accomplishment (Coffman & Adamek, 1999; Coffman, 2002a), perceived social support (Coffman & Adamek, 2001), meaningful interpersonal relationships (Coffman, 2002b), spirituality (Rohwer & Coffman, 2006), and the experiences of conductors with adult learners (Coffman, 2009). While this current study, qualitative in nature, did not directly focus on the benefits of participation, it was the major category that emerged from the data. It is perhaps through discussing participation, that the participants interviewed tried to express some of these benefits embodying the holistic nature of music.

Participating in musical contexts does require some degree of musical understanding, skills and auditory awareness. One expectation then, would be the necessity to hear the music that is being produced around you; whether that means listening for a melody line to balance and tune your own playing, or attending to the members in your instrument section in order to play in tune as a section. Interestingly, however, being able to hear other sections of instruments, the melody, or even the person playing next to you did not emerge as a need, and thus, a category. When participants were asked whether they ever played without wearing their HAs, almost all of them admitted to doing so at one time or another, even during concerts. This suggests that participating in a musical context is not always contingent upon rehabilitation of auditory deficits.

While not all of the participants interviewed were dissatisfied with their HAs, some articulated negative concerns regarding HA sound quality and music programs. These findings are consistent with previous quantitative surveys where many HA users were dissatisfied with music sound quality (Feldmann & Kumpf, 1988; Fulford et al., 2012; Leek et al., 2008; Madsen & Moore, 2014). However, the participants in the current study expressed similar concerns to previous studies only after first discussing participatory needs, further highlighting that there are aspects in the holistic music experience that should be rehabilitated prior to restoring negative auditory deficiencies. What is still not yet fully understood is the relationship between auditory deficiencies and music

participation, and the degree to which HA dissatisfaction inhibited the instrumentalists from playing or listening to music.

2.4.1 Limitations

This study only included individuals with hearing loss who were active instrumentalists. These findings do not reflect the experiences of hearing impaired instrumentalists who are no longer musically active. There may be individuals so affected by hearing loss or so disappointed with HA sound quality that they are unable to participate in a music ensemble and have chosen to leave musical ensembles. The recruitment strategy administered here did not allow for the inclusion of such a population. Some participants said that past members may have left the adult music group due to hearing loss. Future studies with a similar methodology should pursue recruitment strategies that allow the inclusion of instrumentalists who, as a result of their hearing loss, no longer perform music. Future studies can also examine the efficacy of non-auditory attending styles reported by Fulford et al. (2011), such as using vibrotactile feedback, and determine if those styles can be advantageous for the rehabilitation of musical participation, and if they produce holistic musical anecdotes similar to those expressed by normal hearing instrumentalists. Additionally, a revised questionnaire should also include questions focused directly on music-related quality of life and participatory benefits, so that the relationship between these topics and auditory concerns due to hearing loss and HA use can be better articulated.

Unfortunately, the majority of the 49 participants with hearing impairment interviewed in this study were not HA users or were not able to articulate their musical experiences in a way that the researchers perceived was informative. This limited useable data to that of only twelve of the participants. Future recruitment strategies should target a larger population of HA users who are instrumentalists but also have experience in areas such as acoustics and hearing science. This may allow for a more informative articulation of auditory deficiencies. Future studies may also wish to incorporate sessions in which hearing impaired participants listen to and perform music and then reflect upon their experiences immediately after the session during an interview. This would allow for personal accounts of recent musical experiences, richer descriptions of data, and even the inclusion of additional participants.

It will also be valuable to reenter the data pool and consider the music education implications emerging out of the findings. For instance, several of the findings of this study, such as carefully listening to those around you (intonation), being able to discern melodic and harmonic lines, as well as the ability to differentiate between instrument timbers, are also challenges for instrumentalists who do not wear HAs. What might the implications be of such findings on how instrumentalists are taught? More revealing than this, however, is that even those who have not gone through the most basic music education program would likely report that when you play an instrument in an ensemble, listening and responding to those around you is integral to the individual agency of musical experience. Clearly it is important to know where to begin playing so one is in the correct place in the music. A more critical read of the data, however, would also reveal that participating musically in an ensemble is too often based on the conductor making most if not all of the musical and artistic decisions.[6]

2.5 Conclusion

In the year 2000, Conrad & Gunter wrote the following:

> The time is right to break through the conventional boundaries that surround disciplinary inquiry, especially boundaries between disciplines, boundaries separating theory and research from practice, and boundaries separating scholars from practitioners. (p. 50)

We came to this study as an interdisciplinary team three years ago after meeting during a weekend seminar that brought multiple disciplines together under the umbrella of Musical Learning Across the Lifespan. We ended up sitting at the same table, thus, in immediate ways, forced to find (or, at the very least, discuss) common ground. It may have been serendipitous that we ended up together at that table, but we were present at that gathering

---

[6] For further reading to substantiate this point see, O'Toole (1994) who uses Foucault to problematize authority in choral rehearsals. See also Allsup & Benedict (2008) who use a similar critical lens to interrogate the dominance of conductors and their methodological control in wind ensembles.

precisely because we *were* desirous to break through "conventional boundaries" and craft a way forward that would afford new ways of thinking for each of us.

Like many interdisciplinary teams we first had to come to terms with assumptions and non-understandings we made about the others' discipline. Issues which included the ways in which a review of literature is constructed in our disciplines as well as favored research paradigms were fascinating and less complex to address. More complex, however, (yet equally as fascinating) were the kinds of questions we had individually been exploring prior to this current study. For instance, perhaps not completely incomprehensible, but clearly in need of explanation (and one that brought us great joy and laughter), was why one would choose a more pragmatic 'what' question above a philosophically grounded 'why' question.

In this study, the use of qualitative methodology and conventional content analysis addressed both the what and why questions allowing us to explore the impact of hearing loss and HA use on music perception and participation. The authors discovered that the most common music-related concern for included participants with hearing loss and HAs was not actually related to the perception of music itself – it was related to hearing the conductor in order to actively participate in music-related activities. Participants' reports of participation were thought to address, at least in part, hearing-impaired instrumentalists' needs, above general auditory complaints. This concern for participation generally took precedent over direct auditory needs. However, many of the auditory concerns reported were consistent with both quantitative and qualitative evidence from the literature. With respect to HA use, some participants reported quality degradations whereas others reported quality improvements. When probed about improving HAs, participants suggested that an extended bandwidth, improved noise reduction strategies, and a large dynamic range were proposed solutions. With respect to music perception, most participants reported that hearing loss worsens the quality of music. The degradations were mainly attributed to issues in music dynamics and melodic identification.

In conclusion, this study expands a growing body of literature articulating the possible effects of hearing impairment and HA use on music perception and highlights what may

be important for amateur instrumentalists to meaningfully participate in music. The study also sets the stage for research focused on the rehabilitation of holistic music experiences in hearing-impaired instrumentalists, rather than a sole focus on the restoration of specific auditory deficiencies. Future research in this area should place greater focus on the relationship between auditory deficits and the benefits of music listening and participation, and the degree to which worsening auditory deficits reduce those benefits.

2.6 References

Alexander, J. M. (2013). Individual variability in recognition of frequency-lowered speech. *Seminars in Hearing*, *34*(2), 215–218.

Allsup, R. E., & Benedict, C. (2008). The problems of band: An inquiry into the future of instrumental music education. *Philosophy of Music Education Review*, *16*(2), 156–173.

Amlani, A. M., Punch, J. L., & Ching, T. Y. C. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification*, *6*(3), 81–129.

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

Barry, C. A., Britten, N., Barber, N., Bradley, C., & Stevenson, F. (1999). Using reflexivity to optimize teamwork in qualitative research. *Qualitative Health Research*, *9*(1), 26–44.

Bartel, L. R., Greenberg, S., Friesen, L. M., Ostroff, J., Bodmer, D., Shipp, D., & Chen, J. M. (2011). Qualitative case studies of five cochlear implant recipients' experience with music. *Cochlear Implants International*, *12*(1), 27–33.

Bentler, R., & Chiou, L. (2006). Digital noise reduction: An overview. *Trends In Amplification*, *10*(2), 67–82.

Bentler, R., Wu, Y.-H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: outcomes from laboratory and field studies. *International Journal of Audiology*, *47*(8), 447–460.

Berger, R. (2015). Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qualitative Research*, *15*(2), 219–234.

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

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

Chisolm, T. H., Johnson, C. E., Danhauer, J. L., Portz, L. J. P., Abrams, H. B., Lesner, S., … Newman, C. W. (2007). A systematic review of health-related quality of life and hearing aids: Final report of the American Academy of Audiology task force on the health-related quality of life benefits of amplification in adults. *Journal of the American Academy of Audiology*, *18*(2), 151–183.

Coffman, D., & Adamek, M. (1999). The contributions of wind band participation to quality of life of senior adults. *Music Therapy Perspectives*, *28*(1), 27–40.

Coffman, D. D. (2002). Banding together: New Horizons in lifelong music making. *Journal of Aging and Identity*, *7*(2), 133–143.

Coffman, D. D. (2002). Music and quality of life in older adults. *Psychomusicology*, *18*(1–2), 76–88.

Coffman, D. D. (2009). Learning from our elders: Survey of New Horizons International Music Association band and orchestra directors. *International Journal of Community Music*, *2*(2–3), 227–240.

Coffman, D. D., & Adamek, M. S. (2001). Perceived social support of New Horizons Band participants. *Contributions to Music Education*, *28*(1), 27–40.

Coffman, D. D., & Ademek, M. S. (1999). The contributions of wind band participation

to quality of life of senior adults. *Music Therapy Perspectives*, *28*(1), 27–40.
Coffman, D. D., & Levy, K. M. (1997). Senior adult bands music's new horizon: Not only do senior adult bands benefit members, the community, and university students in a practical sense, they also bring joy to all involved. *Music Educators Journal*, *84*(3), 17–22.
Conrad, C. F., & Gunter, R. (2000). To be more useful: Embracing interdisciplinary scholarship and dialogue. *New Directions for Higher Education*, *110*, 49–62.
Contrera, K. J., Betz, J., Li, L., Blake, C. R., Sung, Y. K., Choi, J. S., & Lin, F. R. (2016). Quality of life after intervention with a cochlear implant or hearing aid. *Laryngoscope*, 1–6.
Creswell, J. W. (2007). *Qualitative inquiry & research design* (2nd edition). Thousand Oaks, California: Sage Publications.
Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, *35*(5), e170–e184.
Dabback, W. A. (2008). Identity formation through participation in the Rochester New Horizons Band programme. *International Journal of Community Music*, *1*(2), 267–286.
Emiroglu, S., & Kollmeier, B. (2008). Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. *Brain Research*, *1220*, 199–207.
Feder, K., Michaud, D., Ramage-Morin, P., McNamee, J., & Beauregard, Y. (2015). Prevalence of hearing loss among Canadians aged 20 to 79: Audiometric results from the 2012/2013 Canadian health measures survey. *Health Reports*, *26*(7), 18–25. Retrieved from http://www.statcan.gc.ca/pub/82-003-x/2015007/article/14206-eng.pdf
Feldmann, H., & Kumpf, W. (1988). Musikhoren bei Schwerho rigkeit mit und ohne Horgerat [Listening to music in the hard-of-hearing individual with and without hearing aid]. *Laryng Rhinol Otol*, *67*(10), 489–497.
Finefter-Rosenbluh, I. (2017). Incorporating perspective taking in reflexivity: A method to enhance insider qualitative research processes. *International Journal of Qualitative Methods*, *16*(1), 1–11.
Fulford, R., Ginsborg, J., & Goldbart, J. (2011). Learning not to listen: The experiences of musicians with hearing impairments. *Music Education Research*, *13*(4), 447–464.
Fulford, R., Ginsborg, J., & Greasley, A. (2012). Hearing aids and music : the experiences of D / deaf musicians. In *Proceedings of the Ninth Triennial conference for the European Society for the Cognitive Sciences of Music*. Manchester, UK.
Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033.
Gates, J. T. (1991). Music participation: Theory, research, and policy. *Bulletin of the Council for Research in Music Education*, 1–35.
Gentles, S. J., Jack, S. M., Nicholas, D. B., & Mckibbon, K. A. (2014). Critical approach to reflexivity in grounded theory. *The Qualitative Report*, *19*(25), 1–14.
Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, *8*(4), 597–607.
Gough, B., & Madill, A. (2012). Subjectivity in psychological science: From problem to

prospect. *Psychological Methods*, *17*(3), 374–384.
Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, *29*(4), 75–91.
Higgins, P., Searchfield, G., & Coad, G. (2012). A comparison between the first-fit settings of two multichannel digital signal-processing strategies: Music quality ratings and speech-in-noise scores. *American Journal of Audiology*, *21*, 13–21.
Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288.
Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology*, *14*(7), 347–360.
Jutras, P. J. (2011). The benefits of New Horizons band participation as self-reported by selected New Horizons band members. *Bulletin of the Council for Research in Music Education*, *187*.
Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, *36*(2), 217–228.
Kirchberger, M. J., & Russo, F. A. (2016). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, *20*, 1–16.
Knudsen, L. V., Laplante-Lévesque, A., Jones, L., Preminger, J. E., Nielsen, C., Lunner, T., … Kramer, S. E. (2012). Conducting qualitative research in audiology: A tutorial. *International Journal of Audiology*, *51*(2), 83–92.
Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired Listeners. *Journal of the American Academy of Audiology*, *19*(6), 519–526.
Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing*, *29*(3), 421–434.
Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Amplification*, *0*(0), 1–29.
Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd edition). Thousand Oaks, California: Sage Publications.
Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *Journal of the Audio Engineering Sociaty*, *64*(3), 112–123.
Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, *114*(1), 408–419.
O'Toole, P. (1994). I sing in a choir but I have "no voice"! *The Quarterly Journal of Music Teaching and Learning*, *4–5*(5–1), 65–77.
Patton, M. (2002). *Qualitative evaluation and research methods* (3rd edition). Thousand Oaks, California: Sage Publications.
Rohwer, D., & Coffman, D. (2006). Relationships between wind band membership, activity level, spirituality, and quality of life in older adults. *Research Perspectives in Music Education*, *10*(1), 21–27.
Said, E. A. (2017). Health-related quality of life in elderly hearing aid users vs. non-users. *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*, *18*(3), 271–279.
Scollie, S., Levy, C., Pourmand, N., Abbasalipour, P., Bagatto, M., Richert, F., … Parsa,

V. (2016). Fitting noise management signal processing applying the American Academy of Audiology pediatric amplification guideline: Verification protocols. *Journal of the American Academy of Audiology*, *27*(3), 237–251.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*(2), 63–75.

Sikolia, D., Biros, D., Mason, M., & Weiser, M. (2013). Trustworthiness of grounded theory methodology research in information systems. *Proceedings of the Eighth Midwest Association for Information Systems Conference*, 1–5.

Small, C. (1988). *Musicking: The meaning of performing and listening*. Hanover, New Hampshire: Wesleyan University Press of New England.

Strachan, J. (2016). With Resound Enzos I can for the first time appreciate music. Retrieved August 4, 2016, from http://hearmystory.resound.com/resound-enzos-i-can-first-time-appreciate-music/

Uys, M., & van Dijk, C. (2011). Development of a music perception test for adult hearing-aid users. *The South African Journal of Communication Disorders*, *58*, 19–47.

Vaisberg, J. M., Folkeard, P., Parsa, V., Froehlich, M., Littmann, V., Macpherson, E. A., & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline*, *Article 20872*. Retrieved from www.audiologyonline.com

van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. *The Journal of the Acoustical Society of America*, *105*(5), 2903–2913.

Van Den Hoonard, D. K. (2012). *Qualitative research in action: A Canadian primer*. Don Mills, Ontario: Oxford University Press.

Watt, D. (2007). On becoming a qualitative researcher: The value of reflexivity. *The Qualitative Report*, *12*(1), 82–101.

Wilkinson, S. (1988). The role of reflexivity in feminist psychology. *Women's Studies International Forum*, *11*, 493–502.

Zakis, J. A. (2016). Music perception and hearing aids. In G. R. Popelka, B. C. J. Moore, R. R. Fay, & A. N. Popper (Eds.), *Hearing Aids* (pp. 217–252). Cham, Switzerland: Spring International Publishing Swizterland.

Zimmerman, E., & Lahav, A. (2012). The multisensory brain and its ability to learn music. *Annals of the New York Academy of Sciences*, *1252*(1), 179–184.

Chapter 3

3 Prediction of real hearing aid speech quality using the hearing aid speech quality index (HASQI)

This study evaluates the accuracy of an objective speech quality prediction tool that has been developed for use with hearing impaired listeners who wear hearing aids. Listener ratings and recordings were obtained from published studies that investigated the impact of filtered bandwidth, frequency lowering and adaptive noise reduction on speech quality. The speech quality of recordings from each study was predicted using the hearing aid speech quality index (HASQI) and compared to listener ratings. The hearing aid bandwidth, frequency compression, and noise reduction studies contributed ratings from twenty-one adults, eleven adults and children, and thirteen children, respectively. Listeners had hearing loss and were hearing aid users. HASQI ratings were positively associated with listener judgments and were sensitive to the impacts of noise reduction, bandwidth, and frequency compression. A modified reference signal effectively aligned HASQI ratings with listener judgments but was not sensitive to differences between brands. HASQI can be used to predict the effects of varying signal-processing parameters on commercial hearing aid speech quality. Future research should integrate subjective variability within the HASQI model and conduct cross-manufacturer comparisons while controlling for recording noise.

3.1 Introduction

Objective metrics are powerful tools in the context of hearing aid fittings. An objective metric is a tool which characterizes the electroacoustic behavior of a hearing aid, and when standardized or validated, can be used to predict outcomes for device users. In other words, such metrics can predict the perceived difference between hearing aid output resulting from an electroacoustic change, such as a programming change or the difference between two different devices. This could allow a clinician to measure the output of a hearing aid and adjust its electroacoustic characteristics in an informed way to achieve a more optimal fitting. Objective metrics can be used for the measurement of several perceptual characteristics, such as speech intelligibility and sound quality.

One standardized objective metric, the speech intelligibility index (SII, ANSI, 1997), predicts the percentage of audible speech based on a variety of factors including listener thresholds, signal level, and environmental noise in the signal. The SII is a tool that can be used to supplement interpretation of verification measures while fitting a hearing aid (Amlani, Punch, & Ching, 2002; Scollie, 2018). The SII has been implemented in clinical hearing instrument verification systems such as the Audioscan Verifit2 and Otometrics Aurical. Normative data have recently been developed to support clinical interpretation of aided SII values derived from clinical verification equipment (Baker & Jenstad, 2017; Moodie, Scollie, Bagatto, & Keene, 2017). However, limitations are also present. For example, it is possible to achieve an ideal SII score while having poor sound quality (Gabrielsson, Schenkman, & Hagerman, 1988; Preminger & Van Tasell, 1995). Further, the SII is not always sensitive to changes in bandwidth (Gustafson & Pittman, 2010) which is an important factor in speech quality (Moore & Tan, 2003). For these reasons, pairing objective indices of sound quality alongside the aided SII could provide a more robust assessment of the adequacy of the hearing aid fitting.

Sound quality metrics, in contrast to intelligibility metrics, are not clinically available. A validated sound quality metric could be beneficial, because many hearing aid users cite poor speech quality as a major barrier to hearing aid acceptance (Abrams & Kihm, 2015; Wong, Hickson, & Mcpherson, 2003). While the exact definition of sound quality can be debated, it has been described as the overall fidelity and enjoyability of sound (Kondo, 2012). Sound quality has also been defined as a multidimensional construct, and has been rated using a variety of descriptors, some of which include “fullness”, “sharpness”, “loudness”, and “overall impression” (Gabrielsson & Sjögren, 1979a, 1979b). More methodological approaches have defined sound quality using a single, overall measure of quality. For instance, absolute categorical ratings require listeners to rank quality on a single, predetermined scale (ITU-T, 1996). In contrast, paired comparisons force a listener to select the preferred of two stimuli, with the most preferred stimulus after a series of comparisons achieving the highest score (Amlani & Schafer, 2009). The “MUltiple Stimulus test with Hidden References and Anchors” (MUSHRA) protocol (ITU-R, 2015) combines aspects of absolute categorical ratings and paired comparisons, in that listeners are required to rank several test signals relative to each other in addition to predetermined

high-quality reference and low-quality anchor stimuli. In a clinical context, variations of these definitions can be used to inform device adjustments. For instance, patients' descriptions of hearing aid output can be informative in the electroacoustic adjustment of their device (Jenstad, Van Tasell, & Ewert, 2003). Furthermore, practitioners can request patients' preferences between two devices or signal processing settings to decide on their final devices. However, these procedures can be confounded by poor individual reliability (Narendran & Humes, 2003), and can also be resource-intensive and time-consuming, especially in addition to other necessary procedures during a hearing aid fitting. Finally, it is not possible for some patients, including infants and those with cognitive limitations, to make subjective speech quality ratings. This leaves practitioners without a robust, routine strategy to optimize hearing aid sound quality, solidifying the desire for an objective metric that can be derived from electroacoustic verification measures of hearing aid signal processing.

Objective sound quality metrics are not a novel concept. Predictive sound quality models have historically been applied in the development and evaluation of telecommunications systems for speech transmission. Most models follow a perception-based approach, in that they generate a speech quality score corresponding to signal fidelity and distortion by processing the signal through a psychoacoustic model. Many of the models require a comparison between a test signal and its unprocessed reference version. The models are then validated by correlating a battery of scores against subjective ratings made by human listeners for the same stimuli. Models with high positive correlations are considered strong predictors of speech quality. Popular models include the perceptual evaluation of audio quality index (PEAQ, Thiede et al., 2000), the perceptual evaluation of speech quality index (PESQ, Beerends, Hekstra, Rix, & Hollier, 2002), and the perceptual model of quality assessment (PEMO-Q, Huber & Kollmeier, 2006). Models such as these have been used to assess speech quality distortions found in telecommunications systems. Therefore, these models may not represent the impacts of hearing aid signal processing. Nonetheless, these specific models have produced objective quality scores for some hearing aid distortions, and when compared with their corresponding subjective quality ratings, produced correlation coefficients of 0.7 or greater (Beerends et al., 2002; Beerends, Krebber, Huber, Eneman, & Luts, 2008; Chen, Parsa, & Scollie, 2006; Falk et al., 2015;

Harlander, Huber, & Ewert, 2014; Hu & Loizou, 2008; Huber & Kollmeier, 2006; Huber, Parsa, & Scollie, 2014; Kressner, Anderson, & Rozell, 2013; Pourmand, Parsa, & Weaver, 2013; Rohdenburg, Hohmann, & Kollmeier, 2005; Treurniet & Soulodre, 2000), indicating relatively strong predictability of human speech quality ratings.

If an objective speech quality model is to be used clinically, it should include the impact of hearing loss and it should be validated for the impacts of hearing aid signal processing and/or distortion. One model that satisfies both of these criteria is called the Hearing Aid Speech Quality Index (HASQI, Kates & Arehart, 2014). This index assesses the quality of a hearing aid signal by comparing it to an undistorted reference version of the same signal that has been amplified to NAL-R prescriptive targets (Byrne & Dillon, 1986). HASQI implements the impact of hearing loss in an auditory model. The model uses a spectral analysis which considers the interaction of auditory thresholds and signal level, auditory dynamic-range compression, and neural-firing rate adaptation. HASQI then analyzes the degree of linear and nonlinear distortions in the perceptual representation of the test signal to generate a prediction of speech quality. HASQI has been validated for a wide array of signal processing strategies and distortions commonly found in hearing aids, which include frequency compression, noise suppression, feedback cancellation, dynamic range compression, additive noise, quantization, and peak-clipping, as well as telecommunications-based distortions (Arehart, Kates, & Anderson, 2011; Falk et al., 2015; Harlander et al., 2014; Houben, Brons, & Dreschler, 2011; Huber et al., 2014; Kates & Arehart, 2014; Kressner et al., 2013; Pourmand et al., 2013; Suelzle, Parsa, & Falk, 2013).

Transitioning HASQI from a research context to a clinical context may be challenging, partially because many of its investigations were completed using hearing aid simulators, not real devices. Kates, Arehart, Anderson, Muralimanohar, & Harvey (2018) identified this issue and investigated the clinical applicability of HASQI by systematically probing common clinical considerations for hearing aid fittings using commercial hearing aids. They tested the sensitivity of HASQI to listener characteristics (audiogram), environmental conditions (signal-to-noise ratio [SNR], signal level), hearing aid programming (degree of signal processing), and device differences (manufacturer and model with settings

programmed similarly) using stimuli processed through commercial hearing aids. HASQI scores were statistically sensitive to different levels of the factors: SNR, signal level, manufacturer, and degree of processing. Increasing noise, level, and degree of processing produced statistically lower scores. These results suggest that HASQI can reliably predict signal-processing impacts on speech quality. However, several concerns remain prior to implementation of the HASQI in a clinical context. First, Kates et al. (2018) used S2 and N4 standardized audiograms (Bisgaard, Vlaming, & Dahlquist, 2010) representative of a mild-sloping-to-severe hearing loss and a flat moderately severe hearing loss. While these audiograms reflect common real listener audiograms, they are averages and do not reflect the diversity of those belonging to actual hearing aid users. Second, HASQI was sensitive to low levels of instrumental and environmental noise, even though this noise was not sufficiently high in level to function as a masker (ANSI, 1997). This suggests that HASQI would produce unnecessarily low scores for speech signals in which the speech is sufficiently clear and intelligible, relative to background noise (such as a 10 dB SNR signal or a signal whose noise is inaudible) creating an artificial measurement ceiling. Finally, to better understand the relationship between HASQI and listener ratings, the sensitivity of HASQI to hearing aid fitting parameter changes needs to be compared to human sensitivity to those same changes.

The purpose of this study was to examine HASQI's robustness as an objective speech quality metric for wearable, commercially-available hearing aids. The following questions were investigated: (1) What is HASQI's performance in predicting subjective quality scores produced by participants listening with individually-prescribed hearing aids? (2) Does HASQI's sensitivity to differences in signals and/or signal-processing adjustments reflect the sensitivity experienced by human listeners? (3) Does the artificial ceiling due to inaudible ambient noise distortions present in Kates et al. (2018) replicate to our data, and if so, can it be resolved using a separate reference signal strategy? These questions are answered using two HASQI implementations. The first HASQI analysis implementation used a digitally-shaped reference signal and the second HASQI analysis implementation used a recorded reference signal. The datasets of hearing-aid-processed speech signals and subjective quality ratings used to probe these questions were obtained from previous studies conducted at the National Centre for Audiology and included the following

conditions: filtered bandwidth (Easwar, Purcell, Aiken, Parsa, & Scollie, 2015), automatic noise reduction (Scollie, Levy, et al., 2016) and frequency compression (Glista et al., 2018). The datasets are intended to represent variations in sound quality typically produced by hearing aids but are not comprehensive of all conditions found within the industry. The studies' methods are briefly summarized below.

3.2 Methods

3.2.1 Participants

Behavioral data included in this study were obtained from studies that have been previously published by our institution and approved by the Western University Health Sciences Research Ethics Board (Easwar et al., 2015; Glista et al., 2018; Scollie, Levy, et al., 2016). Behavioral data consisted of subjective ratings from both adults and children. However, age was not distributed evenly between all studies. Therefore, the current study did not investigate effects of age (i.e. whether HASQI was more accurate for adults or children). A brief description of each study and its participants is included in the Validation Data section below.

3.2.2 Validation data

3.2.2.1 Filtered bandwidth

Extended bandwidth amplification is ideal for the optimization of perceived sound quality of speech and music (Moore & Tan, 2003; Ricketts, Dittberner, & Johnson, 2008). Many hearing aids manufacturers now market devices with processing bandwidths in excess of 10 kHz. Despite this, Kimlinger, McCreery, & Lewis (2015) reported that the maximum audible hearing aid bandwidths ranged from 3.5 kHz to above 8 kHz, indicating that some listeners only receive partial bandwidth from their devices, despite a wider bandwidth being available in the signal processing of the device. Therefore, one of the selected validation datasets was a series of low-pass filtered speech stimuli amplified via a hearing aid.

Easwar et al. (2015) gathered speech quality ratings for low-pass filtered stimuli amplified via a single hearing aid from twenty-one hearing impaired adults whose thresholds were

75 dB HL or lower at all frequencies between 0.25 and 6 kHz. All participants wore hearing aids for at least three months. All participants were fitted with 20-channel Unitron Quantum hearing aids. An S model was fitted to participants with mild-to-moderate hearing loss and an HP model to those with moderately severe to severe hearing loss. Hearing aids were coupled to participants' ears using custom-made acrylic earmolds. Hearing aids were verified to match within ±5dB of the DSL v5.0 adult targets (Scollie et al., 2005) from 0.25 – 6 kHz at International Speech Test Signal (Holube, Fredelake, Vlaming, & Kollmeier, 2010) input levels of 50 and 65 dB SPL. Speech quality ratings were obtained monoaurally using the ear with better (lower) pure tone thresholds.

The stimulus used was an IEEE male-spoken sentence pair: "Raise the sail and steer the ship northward. The cone costs five cents on Monday." The stimulus conditions included a full-bandwidth, unfiltered condition and low-pass filter conditions with cut-off frequencies 4, 2, 1 and 0.5 kHz. The stimulus was presented to the hearing aid via direct audio input (DAI) at an input level equivalent to 65 dB SPL with a sample rate of 32 kHz with 16-bit resolution.

Speech quality ratings of the stimulus under the different conditions were conducted using the MUSHRA listening test. The full bandwidth condition served as the reference condition, the 0.5 kHz condition served as the anchor condition, and the 1, 2 and 4 kHz conditions served as the experimental conditions. Speech quality ratings of the reference, anchor, and experimental conditions were used to validate HASQI by comparing them to HASQI-generated objective quality scores.

3.2.2.2 Frequency lowering

Frequency lowering technology is an alternative strategy to extended-bandwidth amplification. Extended-bandwidth audibility is sometimes unachievable due to factors including device limitations or degree of hearing loss. Frequency lowering can improve high frequency audibility by shifting high frequency speech sounds to lower frequency regions where listeners have better thresholds (Alexander, 2013). Glista et al. (2018) obtained speech quality judgments of frequency-compressed speech using hearing aids. Two frequency lowering technologies were evaluated: nonlinear frequency compression

(NFC) and adaptive nonlinear frequency compression (ANFC). NFC compresses input speech with high-frequency content into a smaller output bandwidth. It defines a cut-off frequency (CF) above which a frequency compression ratio (CR) is applied. ANFC is a real-time, adaptive form of NFC, in which two cut-off values (CF1) and (CF2) are defined. In ANFC, the spectral energy distribution of the input speech signal is analyzed in real time. A higher-frequency CF is applied for low-frequency dominated sounds to protect them from becoming compressed. In contrast, if the signal contains more high frequency spectral content, then a lower CF will be applied.

Eleven hearing-impaired listeners (six children, five adults) completed the speech quality evaluations. All participants presented with sloping, high-frequency sensorineural hearing loss and were classified as appropriate candidates for NFC. All participants were hearing aid users with at least one full year of experience. Participants were fitted binaurally with Phonak Naida Q SP or UP experimental hearing aids, depending on severity of hearing loss. Hearing aids were programmed to match DSL v5.0 child or adult targets.

The stimuli used were two female-spoken speech passages: “The rainbow is a division of white light into many beautiful colours,” and, “When the sunlight strikes raindrops in the air they act like a prism and form a rainbow.” The stimuli were presented to hearing aids in sound field at an input level of 65 dB SPL at a 25 kHz sampling frequency with 16-bit per sample resolution.

There were nine hearing aid signal processing conditions. The first condition consisted of the original stimulus (FC-off) without any NFC or ANFC processing. The second and third conditions applied individualized fine-tuned NFC and ANFC settings, respectively, and both were fitted using the maximum audible output frequency fitting protocol described by Scollie, Glista, et al. (2016). The fourth through seventh conditions applied ANFC processing and included fine-tuned CRs, fine-tuned lower CFs, and one of four fixed upper CFs depending on condition. The upper CFs were 5920 Hz (ANFC-1), 4000 Hz (ANFC-2), 3040 Hz (ANFC-3), and 2080 Hz (ANFC-4). The eighth and ninth conditions applied maximum strength parameters for NFC (NFC-max) and ANFC (ANFC-max), respectively. For NFC-max, the CF was 1440 Hz and CR was 4:1. For ANFC-max, the lower CF was

285 Hz, the upper CF was 1440 Hz, and the CR was 4:1. Conditions implementing prescribed FC parameters were fine-tuned for the better ear and applied to both ears. Each ear was independently matched to DSL v5.0 targets.

Speech quality ratings were gathered using the MUSHRA protocol. The different conditions were activated via connection to a HiPro2 using custom MUSHRA hearing aid software and manufacturer fitting software. This allowed for the randomization of hearing aid settings in real time, dictated by MUSHRA condition assignments. The reference condition was the unprocessed stimulus matched to DSL v5.0 targets. The anchor conditions consisted of the NFC and ANFC maximum parameter conditions. The remaining settings described above served as the experimental conditions.

3.2.2.3 Automatic noise reduction

Scollie, Levy, Pourmand, et al. (2016) obtained speech quality ratings for noisy speech processed by the noise reduction algorithms of four currently available hearing aids (circa 2016). Thirteen children completed the speech quality evaluations. All listeners presented with sensorineural hearing loss with a three-frequency pure-tone average ranging from 35-55 dB HL, with high-frequency hearing loss exceeding no more than 75 dB HL at 4000 Hz. All listeners except one were experienced hearing aid users. Participants were fitted binaurally with Oticon Alta Pro behind-the-ear (BTE) hearing aids for the study, which were programmed to match DSL v5.0 child targets. Hearing aids were coupled to the participants' custom earmolds. The hearing aid's automatic noise reduction feature was disabled for all listeners.

The stimuli were QuickSIN sentences (Etymotic Research, 2001), "The line where the edges join was clean. A white silk jacket goes within his shoes." The passage was recorded in quiet and mixed with speech-shaped noise or with multitalker babble noise. Both noise types were mixed at a -10, 0, 5, and 10 dB SNR, with the overall noise level normalized between signals. All the different passage mixes were recorded at a presentation level of 85 dB SPL through four anonymized currently-available hearing aids in an HA2 2-cc coupler except for the -10 and 10 dB SNR conditions, which were recorded through only one hearing aid. For the recordings, all the hearings aids were programmed to match DSL

v5.0 child targets for a flat 50 dB HL hearing loss. The passages with the 0 and 5 dB SNRs were recorded through the hearing aids twice: once with noise reduction deactivated and once with noise reduction activated at its maximum setting. The passages with the -10 and 10 dB SNRs were recorded once through the hearing aids with the noise reduction deactivated. Recordings were saved at a 32 kHz sampling frequency with 16-bit resolution.

Speech quality ratings were gathered using the MUSHRA protocol. Recordings were presented to listeners in sound field at a 70-72 dB SPL presentation level. In the MUSHRA protocol, hearing aid recordings (four hearing aids x two noise reduction settings) were compared against each other within a single noise condition (multitalker babble at 0 and 5 dB SNR or speech-shaped noise at 0 and 5 dB SNR). The experimental conditions consisted of the hearing aid recordings at 0 and 5 dB SNRs. Experimental conditions were measured against anchors and references consisting of the same type of noise. The reference stimulus consisted of the recordings measured at +10 dB SNR. The anchor stimulus consisted of the recordings measured at a -10 dB SNR. The authors chose a reference in noise rather than a reference in quiet because previous work had identified that some listeners rank all noisy signals as very poor speech quality if the MUSHRA reference consists of clean speech (Parsa, Scollie, Glista, & Seelisch, 2013).

3.2.3 HASQI signal analysis

Hearing aid speech quality in this study was characterized using the HASQI metric (Kates & Arehart, 2014). All HASQI processing was fully implemented in MATLAB (R2016b, Mathworks). The code was obtained from James Kates (personal communication, June 23, 2014) and left unmodified.

The reader is referred to Kates & Arehart (2014) for a detailed description of the HASQI speech quality metric. A brief description follows: HASQI first processes stimuli through an auditory model. HASQI's auditory model passes the signal through a middle ear filter reducing the response below 350 Hz and above 5000 Hz, then sending the middle ear output to a 32-band gammatone filterbank with center frequencies from 80-8000 Hz. Dynamic-range compression is provided by outer hair cell simulation, which is followed by the model's final stage, neural-firing rate adaption provided by inner hair cell simulation. The

model simulates impaired hearing by widening gammatone filters, lowering compression ratios, increasing upward spread of masking, and reducing two-tone suppression (Kates & Arehart, 2014).

A quality rating is calculated via a comparison between the outputs of this auditory model for the test signal and for the high-quality reference signal. HASQI calculates the test signal's nonlinear distortions by comparing envelope modulations and temporal fine structures between the test and reference signals. HASQI calculates the signal's linear distortions by measuring excitation-pattern differences between the test and reference signals. The linear and nonlinear terms are then combined, and a regression function is used to generate a final quality score. The terms of the regression function were derived from a battery of training data (Arehart, Kates, & Anderson, 2010). A description of the test and reference signals used in our implementation follows. For datasets with binaural quality judgments, quality scores were computed for both right and left ears and then averaged into a single final quality score to be compared to its corresponding subjective rating, consistent with the methodology of Falk et al., (2015).

3.2.4 Hearing aid recordings

The HASQI test signal analysis required hearing aid recordings from each validation dataset for each condition per participant (and per ear, if applicable). Ideally, these recordings would have been performed on-ear using a probe tube during the time of each study. However, this was not the case for any of the datasets described above. Therefore, coupler-based participant-specific hearing aid recordings were newly generated for the purposes of this study.

Recordings were made using the Audioscan Verifit2, operated with custom-developed recording software. The Verifit2 presented stimuli to hearing aids in the textbox and used the custom-developed software to record the hearing aid output and save recordings as .wav files. The hearing aids, fitting software versions, and participant fittings from each of the previous studies were retrieved to generate the recordings. The hearing aids included the Unitron Quantum SP and HP BTE hearing aids from the filtered bandwidth dataset, the Phonak Naida Q SP and UP BTE hearing aids from the frequency lowering dataset, and

Oticon Alta Pro BTE hearing aids from the automatic noise reduction dataset. Hearing aids were programmed using the same software and participant fittings that were used to gather the MUSHRA ratings in each of the validation datasets. Fittings were verified to ensure that they were accurately re-created (i.e., that they matched the targets). Hearing aids were coupled to a 0.4 cc coupler in the Verifit2 sound-isolated testbox. Stimuli were presented to the hearing aids via the testbox speaker. The stimuli belonging to the filtered bandwidth and automatic noise reduction datasets were pre-processed using each study condition and presented to either the better ear hearing aid (for the filtered bandwidth study) or both hearing aids (for the automatic noise reduction study), which were programmed per participant. The stimuli from the filtered bandwidth dataset and automatic noise reduction dataset were presented at hearing aid input levels of 65 and 71 dB SPL, respectively. The frequency lowering study conditions were processed in real time. Therefore, the unprocessed stimuli were presented to both hearings aids per participant multiple times as the NFC or ANFC parameters were changed in real time via the hearing aid software. The frequency lowering stimuli were presented at a hearing aid input level of 65 dB SPL. All recordings were saved at a sampling frequency of 32 kHz with 16-bit resolution.

Recordings were post-processed in MATLAB so that the signals could be transformed from coupler levels to real-ear levels. Individualized 0.4 cc real ear to coupler (RECD) transforms were derived from study records of measured RECDs. HA2 RECDs in the datasets were transformed to HA1 RECDs using Audioscan's proprietary HA2 to HA1 transform. The HA1 to 0.4-cc transform reported in Figure 1 of Vaisberg, Folkeard, Pumford, Narten, & Scollie (2018) was then applied to produce individualized 0.4cc RECDs. A 512 point finite impulse response filter was derived from the 0.4 cc RECD using the frequency sampling method and applied to each recording to transform the recording to predicted ear levels.

3.2.5 Reference signals

The undistorted reference signal for hearing impaired listeners in HASQI is defined as an optimal quality, clean version of the stimulus shaped to the NAL-R frequency response (Kates & Arehart, 2014). To generate the reference signals used here, individualized DSL v5.0 (rather than NAL-R) shaping was applied to each digital signal, accounting for the

input level per study and the individual audiogram per participant. The DSL shaping was selected to match the prescription used to generate the MUSHRA reference signals used in the previous studies. The signal shaping was performed in MATLAB using a finite impulse response filter based on the frequency-gain inputs provided by a software application that generated DSL v5.0 targets. These signals represented the reference signals for the initial implementation of HASQI.

Kates et al. (2018) highlighted several concerns described below which raise questions regarding this reference strategy. First, because hearing aid receivers have response peaks that are not present in a digitally-shaped reference signal, the digitally-shaped reference will typically have a smoother response when compared to that measured from a hearing aid. This minor shaping difference will contribute to the measured distortion and reduce the HASQI score (Kates et al., 2018). Second, background noise may have been introduced into the hearing aid recordings from the recording equipment and from hearing aid processing noise. HASQI is particularly sensitive to ambient noise (i.e., recording noise, hearing aid processing, noise floor) (Kates et al., 2018). The ambient noise was expected to introduce an artificial ceiling into the HASQI metric if the reference signal was obtained digitally and not obtained using a similar recording system. The interaction of response peaks and ambient noise did not likely impact participants' subjective ratings in any of the datasets described in this study, and if present in the recording data, may limit the relationship between HASQI and subjective scores. These concerns motivated a second implementation of HASQI with the use a different type of reference signal.

In this second implementation, we chose a hearing aid recording as the reference rather than generating a digitally-shaped reference. The reference hearing aid recording was the reference condition from the subjective MUSHRA procedure. Recall that the MUSHRA protocol requires a high-quality reference signal for listeners who are asked to compare each test signal to a given reference. Comparing a test signal against a reference signal recorded by the same hearing aid using the same recording apparatus would suppress degradations attributed to background noise, isolating the degradations due to signal-processing adjustments alone. However, this second strategy also means that the reference signal is inherently study-specific. This second strategy also means that the objective score

for the reference signal condition will be 1.00 because the signal is being compared against itself. The reference signals used are described below.

The reference condition from the filtered bandwidth dataset consisted of the wideband hearing aid recordings. The reference condition from the frequency lowering dataset consisted of the FC-off hearing aid recordings. The reference conditions from the automatic noise reduction dataset consisted of the +10 dB SNR speech-shaped noise or multitalker babble hearing aid recordings, depending on what type of noise was present in the signal.

3.3 Results

3.3.1 Statistical analysis

The studies above contributed a total of 771 speech quality ratings (filtered bandwidth = 105, frequency lowering = 162, automatic noise reduction = 504) corresponding to the 771 HASQI scores for the digitally-shaped and recorded reference analyses. Raw individual HASQI scores using a digitally-shaped reference (labelled as dHASQI) ranged from 0.001 to 0.60 (mean = 0.17, SD = 0.16). When broken down by validation dataset, scores ranged from 0.001 to 0.60 (mean = 0.28, SD = 0.19) for the filtered bandwidth dataset, 0.01 to 0.59 (mean = 0.38, SD = 0.14) for the frequency lowering dataset, and 0.002 to 0.25 (mean = 0.08, SD = 0.05) for the automatic noise reduction dataset. Raw individual HASQI scores using the recorded references from the subjective MUSHRA procedures (labelled as mHASQI) ranged from 0.000 to 1.00 (mean = 0.57, SD = 0.23). When broken down by validation dataset, scores ranged from 0.001 to 1.00 (mean = 0.44, SD = 0.38) for the filtered bandwidth dataset, 0.004 to 1.00 (mean = 0.63, SD = 0.27) for the frequency lowering dataset, and 0.17 to 1.00 (mean = 0.58, SD = 0.15) for the automatic noise reduction dataset.

We evaluated the predictive validity of HASQI at mean and individual levels. At the mean level, we measured the correlation between mean speech quality ratings and mean HASQI scores for each dataset. At the individual level, we measured the correlation between individual speech quality ratings and individual HASQI scores for each dataset. For correlations of means, raw scores were averaged across individuals within each factor level of each study condition. The filtered bandwidth scores varied depending on the frequency

at which the low-pass filter cut-off was applied and included five factor levels. The factors for the frequency lowering dataset were degree of processing (which included nine factor levels) and stimulus (which included two factor levels). The factors for the automatic noise reduction dataset were manufacturer (which included four factor levels), noise reduction activation (which included two factor levels), and SNR/noise combination (which included four factor levels).

Correlations were quantified using Pearson's linear correlation coefficient r and Spearman's rank correlation coefficient r_s. Rank correlations were only computed for mean values, and not individual values. Individual values consisted of several ties (i.e., two or more of the same rating from one individual in two or more conditions and/or two or more of the same rating from two or more individuals for one condition) for both HASQI and MUSHRA scores, which rendered the rank correlation coefficient unable to compute p-values for all the datasets. Second, we used the mean square error (MSE) to quantify the estimated difference between the objective scores and subjective ratings, by measuring the mean of the squares of errors between HASQI scores and speech quality ratings listeners within each condition. This was similar to the analysis in Kressner et al. (2013), in which the MSE informed the development of a modified metric that better represented a set of validation scores. We also measured the root mean square error (rMSE) by calculating the square root of the MSE. We tested HASQI's sensitivity to differences between conditions using multiple repeated measures analyses of variance (RM-ANOVAs) for each dataset. Greenhouse-Geisser corrections were applied to protect against potential departures from sphericity (Gray & Kinnear, 1999). The RStudio software (Version 1.0.132; R Core Team, 2017) and "ez" package (Lawrence, 2016) were used to complete the validation and sensitivity analyses. The analyses were performed independently for dHASQI and mHASQI scores.

For the RM-ANOVAs, HASQI scores were treated as the dependent variable. The factors listed in this section were treated as the independent variables. For the mHASQI analysis, the MUSHRA reference recording served as the reference for the HASQI analysis. For that condition, the reference stimulus would have been compared against itself, yielding a mean of 1 with no variability. Therefore, MUSHRA reference conditions for the filtered

bandwidth and frequency lowering were excluded from the RM-ANOVA analyses. The MUSHRA +10 dB SNR reference and -10 dB SNR anchor conditions were excluded from the automatic noise reduction dataset because they were only recorded and rated using one of the four hearing aids used in that study. This would have created an unbalanced level in the RM-ANOVA.

3.3.2 Statistical results

In this section, each validation dataset is presented independently. Within each dataset, the correlations are first presented with accompanying scatterplots (Figures 3-1 through 3-12). Correlation coefficients across all datasets are also summarized in Table 3-1. Within each scatterplot, the x-axes represent the mean HASQI predictions across listeners and the y-axes represent the mean subjective MUSHRA ratings across listeners. Each data point represents a single hearing aid processing condition averaged across listeners and stimuli. While not reflecting the ratings of every individual, averaging across individuals removes between-subject error and is a method that is consistent with previous validations which report on ratings averaged over subject groups (Falk et al., 2015; Huber et al., 2014; Kates & Arehart, 2014; Kressner et al., 2013). The diagonal lines represent perfect agreement between the HASQI scores and MUSHRA ratings. Data points to the left of the line represent HASQI predictions that underestimate MUSHRA ratings whereas data points to the right of the line represent HASQI predictions that overestimate MUSHRA ratings.

Table 3-1: Summary of validation data for correlations on mean values. r = Pearson linear correlation coefficient, rs = Spearman rank correlation coefficient, MSE = mean squared error, rMSE = root mean square error.

	dHASQI				mHASQI			
	r	*rs*	MSE	rMSE	*r*	*rs*	MSE	rMSE
Filtered bandwidth	0.99	1.00	0.099	0.315	0.93	1.00	0.036	0.189
Frequency compression	0.98	0.73	0.033	0.180	0.94	0.73	0.019	0.139
Automatic noise reduction	0.84	0.83	0.234	0.484	0.90	0.83	0.006	0.079

Table 3-2: Summary of validation data for correlations on individual values. r = Pearson linear correlation coefficient, rs = Spearman rank correlation coefficient, MSE = mean squared error, rMSE = root mean square error.

	dHASQI				mHASQI			
	r	*rs*	MSE	rMSE	*r*	*rs*	MSE	rMSE
Filtered bandwidth	0.88	---	0.116	0.342	0.85	---	0.054	0.233
Frequency compression	0.52	---	0.079	0.281	0.52	---	0.070	0.264
Automatic noise reduction	0.42	---	0.282	0.531	0.46	---	0.055	0.234

3.3.3 Filtered bandwidth

Figures 3-1 and 3-2 represent the scatterplots for the dHASQI and mHASQI scores, respectively, for relation between mean HASQI and mean speech quality ratings across bandwidth. Figures 3-3 and 3-4 are similar figures, except that they illustrate individual speech quality ratings and HASQI scores. For the means, dHASQI scores resulted in a very high linear correlation (r = 0.99, p < 0.01) and a perfect rank correlation (r_s = 1.00, p < 0.05) between the HASQI output and subjective ratings. The strength of correlation was higher for mean ratings compared to individual ratings (r = 0.88, p < 0.0001). However, for both mean and individual ratings, all the data points fell left of the line, suggesting that the HASQI scores underestimated the true subjective ratings (means: MSE = 0.099, rMSE = 0.315, individuals: MSE = 0.116, rMSE = 0.342). For example, dHASQI predicted a mean rating of 0.486 for the full bandwidth condition, whereas its mean subjective rating was 0.912. For the anchor condition, dHASQI predicted a mean rating of 0.039 compared to a subjective rating of 0.122. The RM-ANOVA revealed a significant main effect of filter cut-off ($F_{(1.99,\ 39.73)}$ = 446.5, p < 0.0001, η^2 = 0.91) on the dHASQI scores. Post-hoc Bonferonni contrasts revealed significant differences between each low-pass cut-off, except between the 4 kHz and wideband conditions, with scores increasing as the low-pass cut-off increased.

The mean mHASQI scores also resulted in a very high correlation (r = 0.93, p < 0.05) and a perfect rank correlation (r_s = 1.00, p < 0.05) between the HASQI output and subjective ratings, averaged across condition. The strength of correlation was higher for mean ratings compared to individual ratings (r = 0.85, p < 0.0001). For both mean and individual ratings, the data points were closer to the diagonal line compared to the dHASQI analysis, suggesting that mHASQI scores better aligned with the true subjective scores (means: MSE = 0.036, rMSE = 0.189, individuals: MSE = 0.054, rMSE = 0.233). For example, mHASQI predicted a perfect mean rating of 1.000 for the full bandwidth condition compared to a mean subjective rating of 0.912. For the anchor condition, mHASQI predicted a mean rating of 0.039 compared to the mean subjective rating of 0.122. The 4000 Hz low-pass filter condition was subjectively rated 0.826, compared to a mean HASQI prediction of 0.731. The RM-ANOVA revealed a significant main effect of low-pass cut-off ($F_{(2.03,40.7)}$

= 498.0, $p < 0.0001$, $\eta^2 = 0.94$) on the mHASQI scores. Post-hoc Bonferonni contrasts revealed significant differences between each low-pass cut-off, with scores increasing as the low-pass cut-off increased.

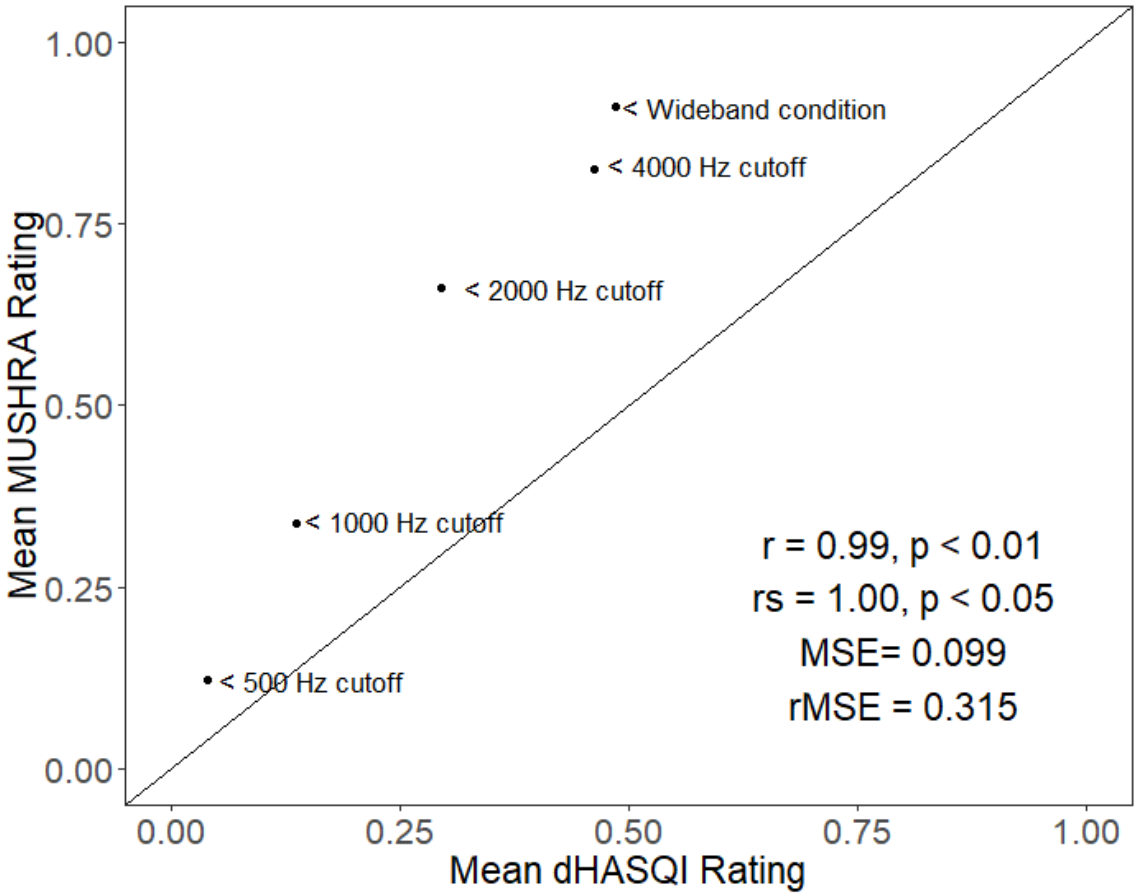

Figure 3-1: Scatterplot of mean MUSHRA ratings and dHASQI quality predictions for the filtered bandwidth dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error.

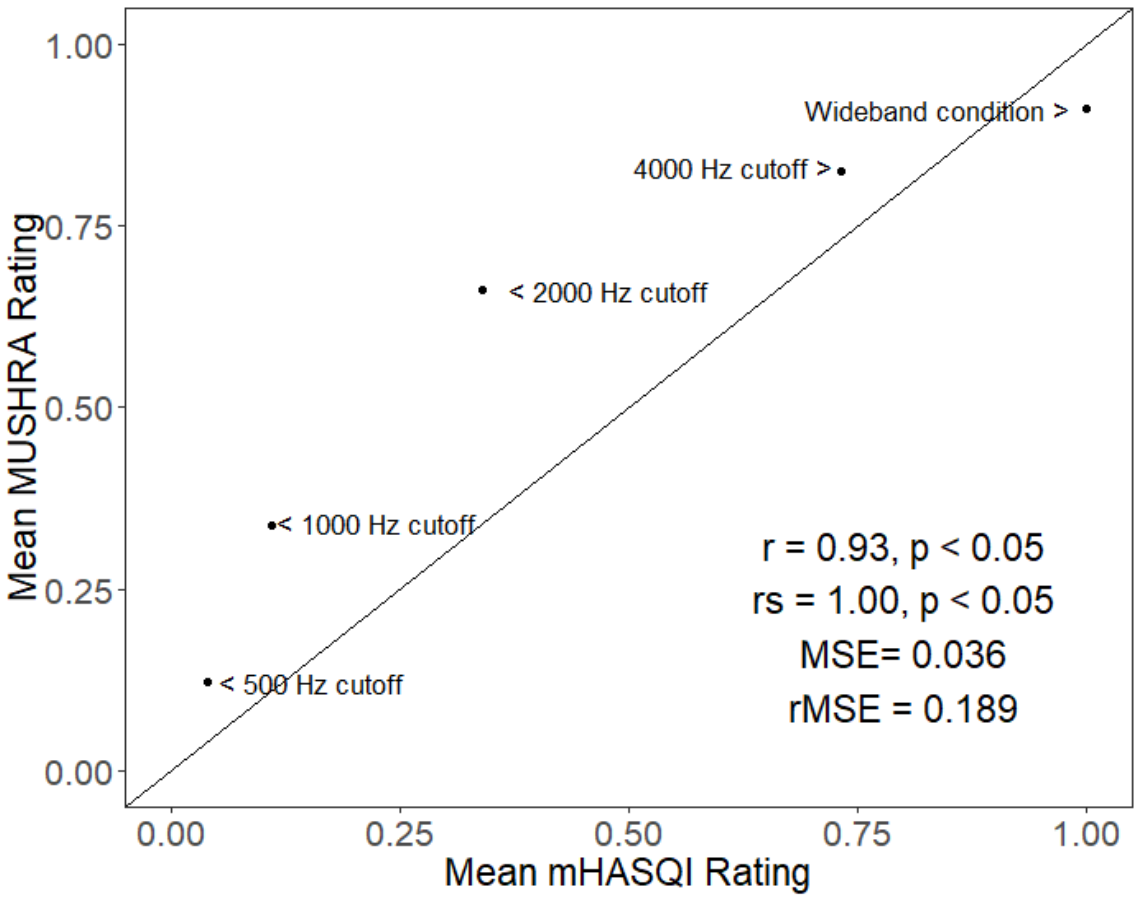

Figure 3-2: Scatterplot of mean MUSHRA ratings and mHASQI quality predictions for the filtered bandwidth dataset using a recorded hearing aid reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error.

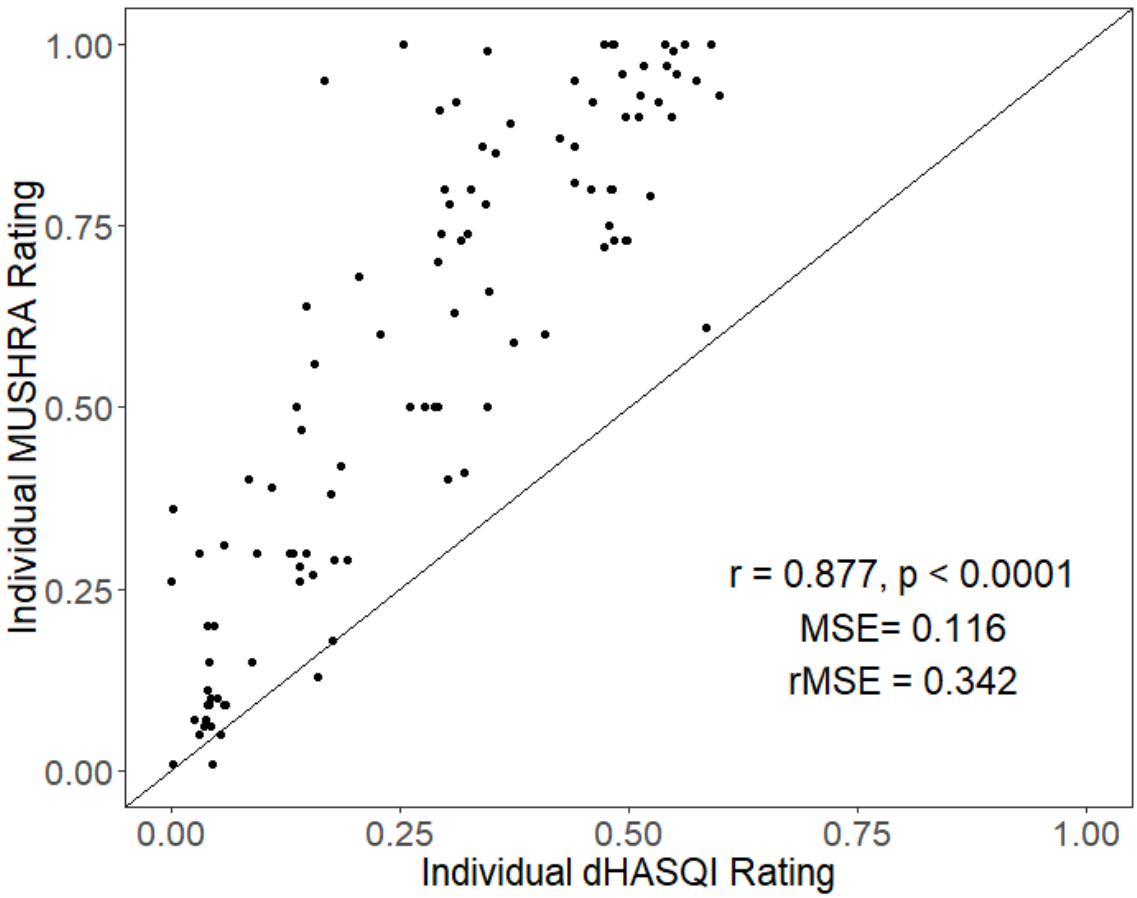

Figure 3-3: Scatterplot of individual MUSHRA ratings and dHASQI quality predictions for the filtered bandwidth dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error.

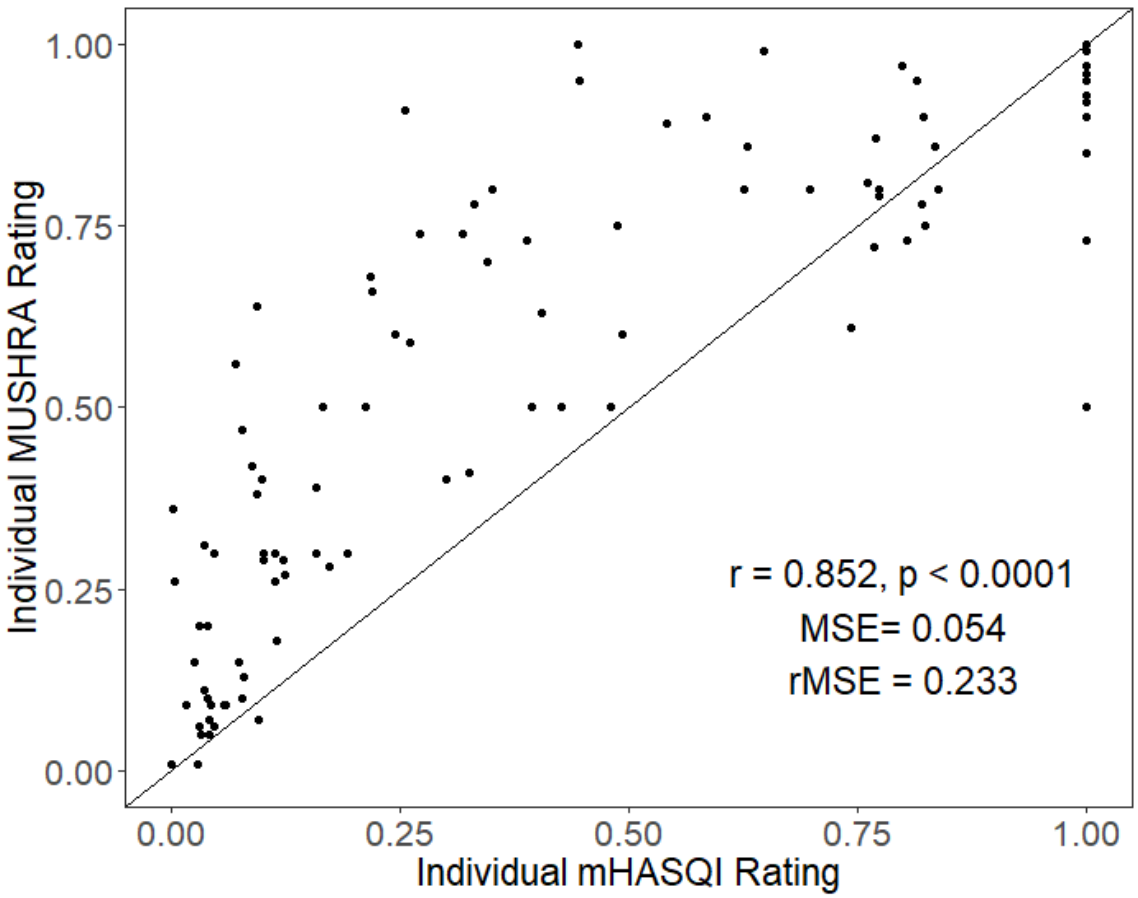

Figure 3-4: Scatterplot of individual MUSHRA ratings and mHASQI quality predictions for the filtered bandwidth dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error.

3.3.4 Frequency lowering

Figures 3-5 and 3-6 represent the scatterplots for the mean subjective ratings and mean dHASQI and mHASQI scores, respectively, for the frequency lowering correlations. Figures 3-7 and 3-8 are similar figures, except that they illustrate individual speech quality ratings and HASQI scores. The mean dHASQI scores resulted in a very high linear correlation ($r = 0.98$, $p < 0.001$) and a moderately high rank correlation ($r_s = 0.73$, $p < 0.05$) between the HASQI output and subjective ratings, although lower rank correlation compared to the filtered bandwidth. The strength of correlation was higher for mean ratings compared to individual ratings ($r = 0.52$, $p < 0.0001$). For both average and individual ratings, all the data points fell left of the line, again suggesting that the HASQI scores underestimated the true subjective scores (means: MSE = 0.033, rMSE = 0.180, individuals: MSE = 0.079, rMSE = 0.281). dHASQI predicted a mean rating of 0.463 for the FC-off condition, which corresponded to a mean subjective rating of 0.688. For the NFC anchor condition, dHASQI predicted a mean rating of 0.226 compared to a mean subjective rating of 0.300. For the ANFC anchor condition, dHASQI predicted a mean rating of 0.162 compared to a mean subjective rating of 0.189. The RM-ANOVA revealed a significant main effect of FC setting ($F_{(1.78,14.2)} = 37.69$, $p < 0.0001$, $\eta^2 = 0.61$) on the dHASQI scores. The main effect of stimulus, and interaction of FC setting x stimulus, were both non-significant. Post-hoc Bonferonni contrasts revealed that both anchor conditions were significantly different from every other condition except each other, and that the ANFC-1 (max FC processing) condition was different from every other condition. Anchor conditions presented lower ratings compared to reference conditions. No other condition was significantly different from any other.

The mean mHASQI scores also resulted in a high correlation ($r = 0.94$, $p < 0.001$) and moderately high rank correlation ($r_s = 0.73$, $p < 0.05$) between the HASQI output and subjective ratings averaged across listeners per condition. The strength of correlation was higher for mean ratings compared to individual ratings ($r = 0.52$, $p < 0.0001$). For both average and individuals ratings, the majority of the data points were slightly right of the diagonal line, suggesting that mHASQI overestimated the true subjective scores (means: MSE = 0.019, rMSE = 0.139, individuals: MSE = 0.070, 0.264), although they better

represented the true subjective scores compared to dHASQI (means: MSE = 0.033, rMSE = 0.180, individuals: MSE = 0.079, rMSE = 0.281). mHASQI predicted a perfect mean rating of 1.000 for the FC-off condition compared to a mean subjective rating of 0.688. For the NFC anchor condition, mHASQI predicted a mean rating of 0.289, compared to a mean subjective rating of 0.300. For the ANFC anchor condition, mHASQI predicted a mean rating of 0.160 compared to a mean subjective rating of 0.189. A mid-quality condition, in which ANFC was fine-tuned to the individual, was subjectively rated as 0.659, compared to a mean mHASQI prediction of 0.722. The RM-ANOVA revealed a significant main effect of FC setting ($F_{(2.75,21.99)} = 131.13$, $p < 0.0001$, $\eta^2 = 0.81$) on the mHASQI scores. Post-hoc Bonferonni contrasts revealed an almost identical pattern of effects as for dHASQI, except that the ANFC-2 and ANFC-4 conditions were also significantly different from one another. There was a significant main effect of stimulus ($F_{(1,8)} = 7.99$, $p < 0.05$, $\eta^2 = 0.03$), suggesting that mHASQI produced significantly different scores for the two speech passages, unlike dHASQI. The interaction of FC setting x stimulus was non-significant. The FC-off condition was excluded from mHASQI analysis.

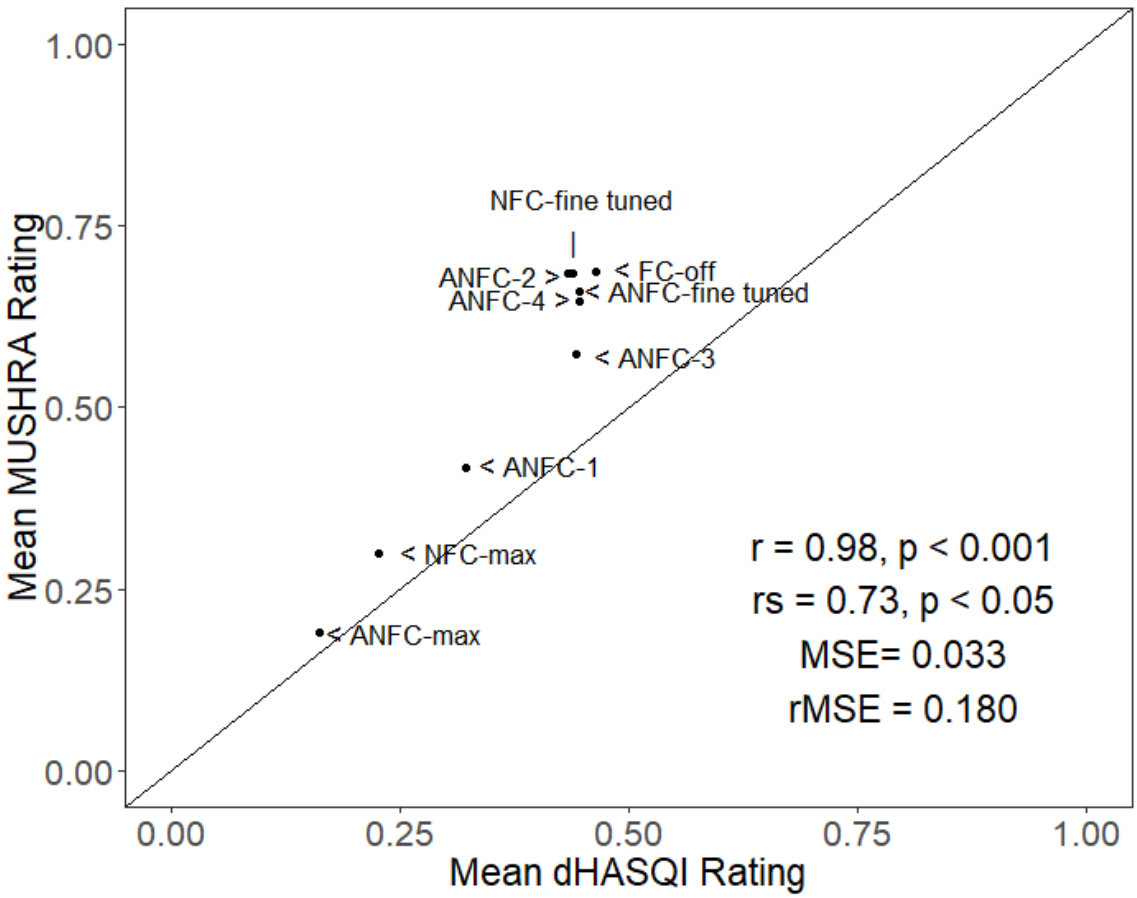

Figure 3-5: Scatterplot of mean MUSHRA ratings and dHASQI quality predictions for the frequency compression dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error.

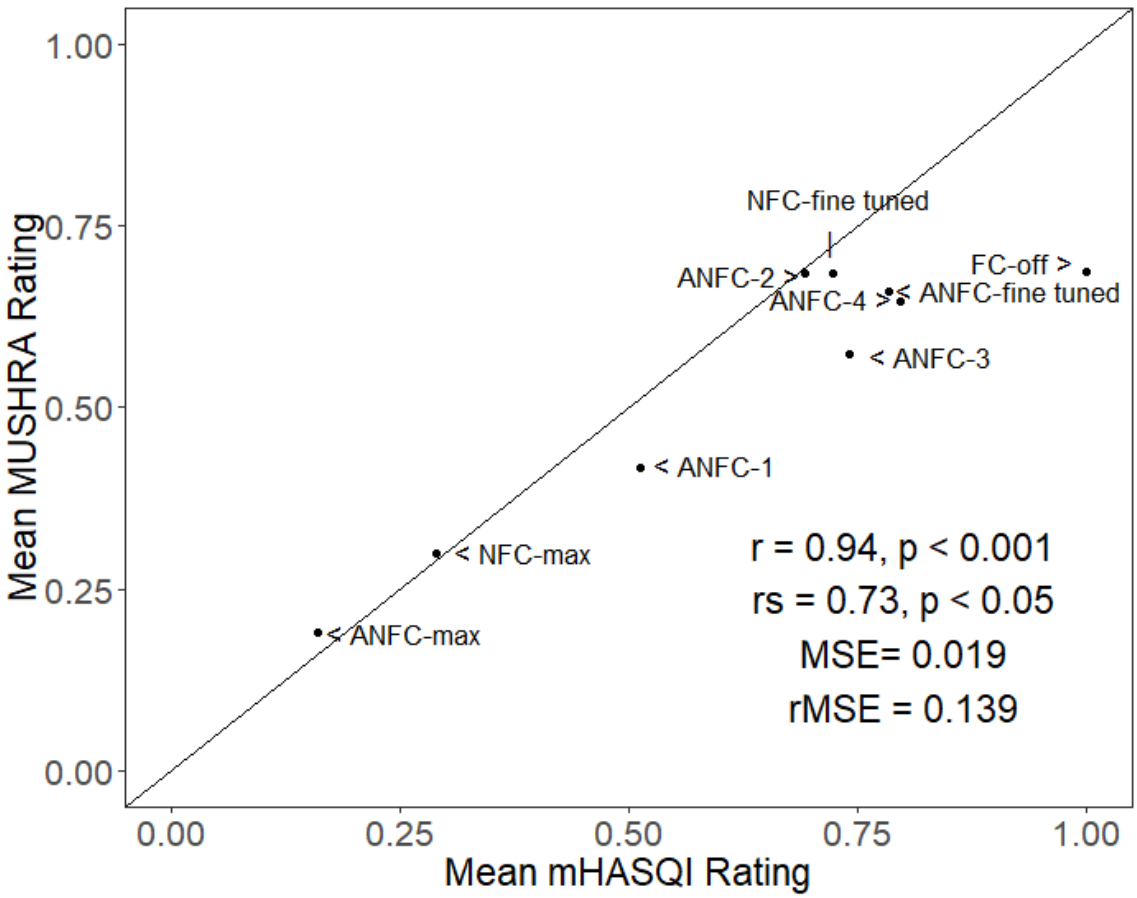

Figure 3-6: Scatterplot of mean MUSHRA ratings and mHASQI quality predictions for the frequency compression dataset using a recorded hearing aid reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error.

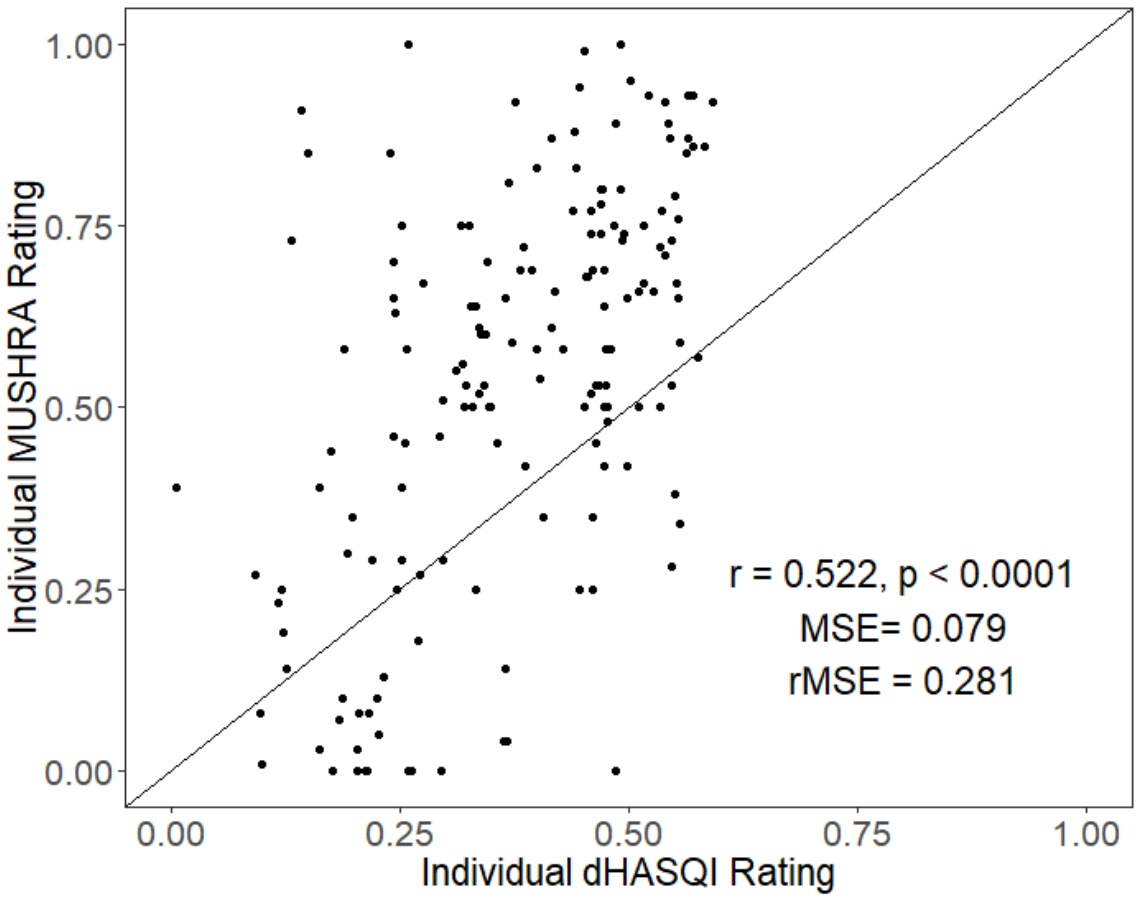

Figure 3-7: Scatterplot of individual MUSHRA ratings and dHASQI quality predictions for the frequency compression dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error.

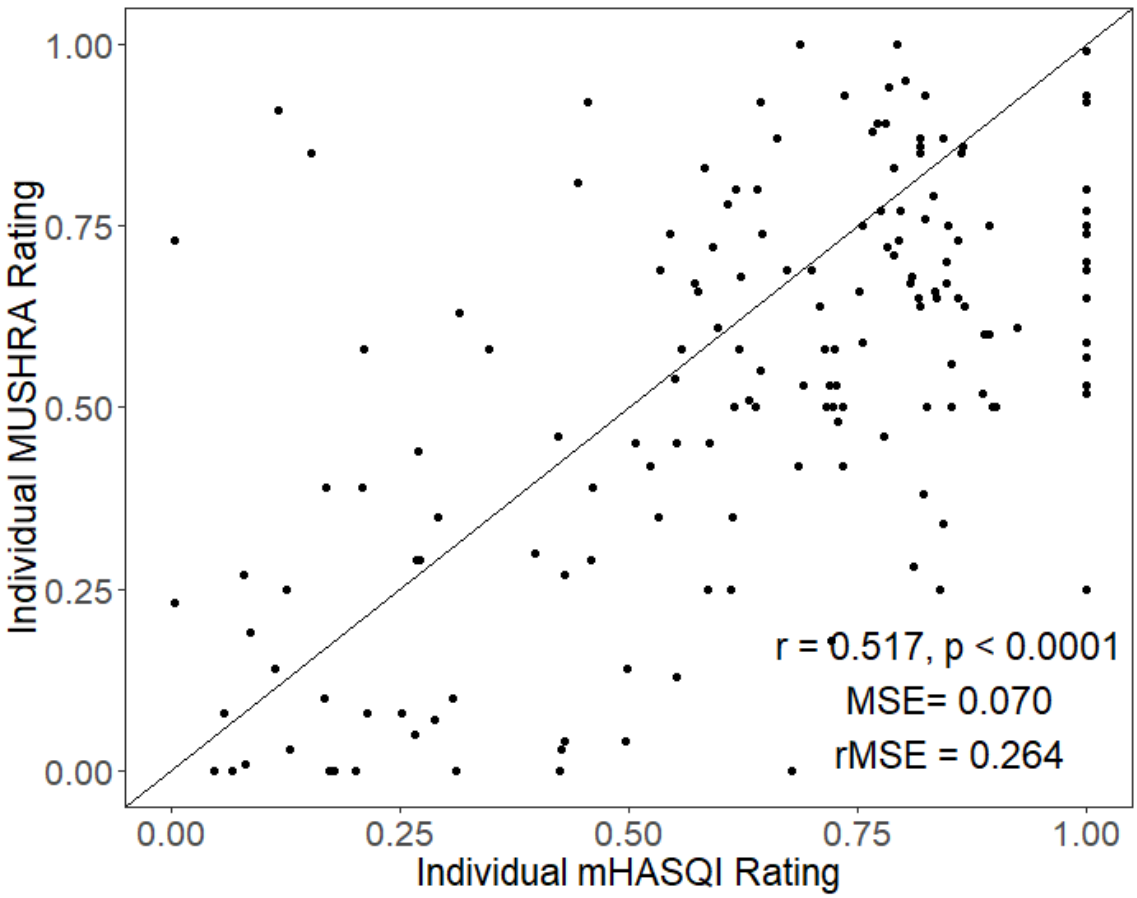

Figure 3-8: Scatterplot of individual MUSHRA ratings and mHASQI quality predictions for the frequency compression dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error.

3.3.5 Automatic noise reduction

Figures 3-9 and 3-10 represent the scatterplots for the mean subjective ratings and the mean dHASQI and mHASQI scores, respectively, for the automatic noise reduction correlations. Figures 3-11 and 3-12 are similar figures, except that they illustrate individual HASQI scores and speech quality ratings. The dHASQI scores resulted in a high linear correlation (r = 0.836, p < 0.001) and a high rank correlation (r_s = 0.826, p < 0.001), when averaged across conditions. The strength of correlation was higher for mean ratings compared to individual ratings (r = 0.42, p < 0.0001). For both average and individual ratings, the data points fell left of the line, suggesting that the HASQI scores underestimated the true subjective scores by a larger magnitude (means: MSE = 0.234, rMSE = 0.484, individuals: MSE = 0.282, rMSE = 0.531), when compared to the other datasets. dHASQI predicted a mean rating of 0.178 and 0.196 for the +10 dB SNR speech-shaped noise and multitalker babble conditions, which were subjectively rated 0.870 and 0.863, respectively. dHASQI predicted a mean rating of 0.008 and 0.011 for the -10 dB SNR speech-shaped noise and multitalker babble conditions, which were subjectively rated 0.080 and 0.084, respectively. The RM-ANOVA revealed main effects of hearing aid ($F_{(1.49, 19.43)}$ = 70.3, p < 0.0001, η^2 = 0.13), listening condition ($F_{(1.12,14.58)}$ = 226.59, p < 0.0001, η^2 = 0.49), and noise reduction ($F_{(1, 13)}$ = 240.65, p < 0.0001, η^2 = 0.08) on the dHASQI scores. The results also revealed significant interactions of hearing aid x listening condition ($F_{(2.32, 30.2)}$ = 104.82, p < 0.0001, η^2 = 0.05), hearing aid x noise reduction ($F_{(1.54, 20)}$ = 73.42, p < 0.0001, η^2 = 0.03), listening condition x noise reduction ($F_{(1.79, 23.24)}$ = 138.02, p < 0.0001, η^2 = 0.03) and hearing aid x listening condition x noise reduction ($F_{(2.76, 35.87)}$ = 109.36, p < 0.0001, η^2 = 0.04). Post-hoc Bonferroni contrasts were performed on the same contrasts reported by Scollie, Levy, et al. (2016). In this study, all four manufacturer's hearing aids were significantly different from one another. Additionally, each hearing aid's noise reduction algorithm yielded significantly higher quality scores when switched from off to on. Post-hoc Bonferroni contrasts were also used to determine if dHASQI produced different scores for each listening condition. Each combination of SNR and noise type was significantly different from one another. dHASQI produced higher scores as SNR increased, and in multitalker babble relative to speech-shaped noise.

The mean mHASQI scores also resulted in a high linear correlation (r = 0.899, p < 0.001) and a high rank correlation (r_s = 0.826, p < 0.001). The strength of correlation was higher for mean ratings compared to individual ratings (r = 0.46, p < 0.0001). For average scores, the data points were closest to the line (MSE = 0.006, rMSE = 0.079) relative to the other datasets, indicating relatively good predictability of subjective scores. For individual scores, the data points were a similar distance from the line (MSE = 0.055, rMSE = 0.234) as observed in the filtered bandwidth dataset for individual mHASQI scores (MSE = 0.054, rMSE = 0.235, Figure 3-4). mHASQI predicted a perfect mean rating of 1.000 for the +10 dB SNR speech-shaped noise and multitalker babble conditions, which were subjectively rated 0.870 and 0.863, respectively. mHASQI predicted a mean rating of 0.216 and 0.291 for the -10 dB SNR speech-shaped noise and multitalker babble (anchor) conditions, which were subjectively rated 0.080 and 0.084, respectively. A mid-quality version, represented by hearing aid-processed speech with multitalker babble at a 5 dB SNR with noise reduction activated, was subjectively rated as 0.482, compared to a HASQI prediction of 0.587. The RM-ANOVA revealed main effects of hearing aid ($F_{(2.16, 27.69)}$ = 217.32, p < 0.0001, η^2 = 0.57), listening condition ($F_{(1.92,24.99)}$ = 642.63, p < 0.0001, η^2 = 0.89), and noise reduction ($F_{(1, 13)}$ = 265.26, p < 0.0001, η^2 = 0.27) on the mHASQI scores. The results also revealed significant interactions of hearing aid x listening condition ($F_{(4.69, 60.97)}$ = 53.02, p < 0.0001, η^2 = 0.28), hearing aid x noise reduction ($F_{(2.46, 31.92)}$ = 111.01, p < 0.0001, η^2 = 0.15), listening condition x noise reduction ($F_{(2.57, 33.42)}$ = 138.02, p < 0.0001 η^2 = 0.22) and hearing aid x listening condition x noise reduction ($F_{(4.54, 58.98)}$ = 62.95, p < 0.0001, η^2 = 0.21). Post-hoc Bonferroni contrasts revealed significant differences identical to those identified by dHASQI, except between two of the manufacturers' hearing aids, which were not significantly different. The +10 dB SNR conditions were excluded from this analysis because they were only measured using one hearing aid to generate references and anchors. A +10 dB SNR condition was not measured for the other three hearing aids.

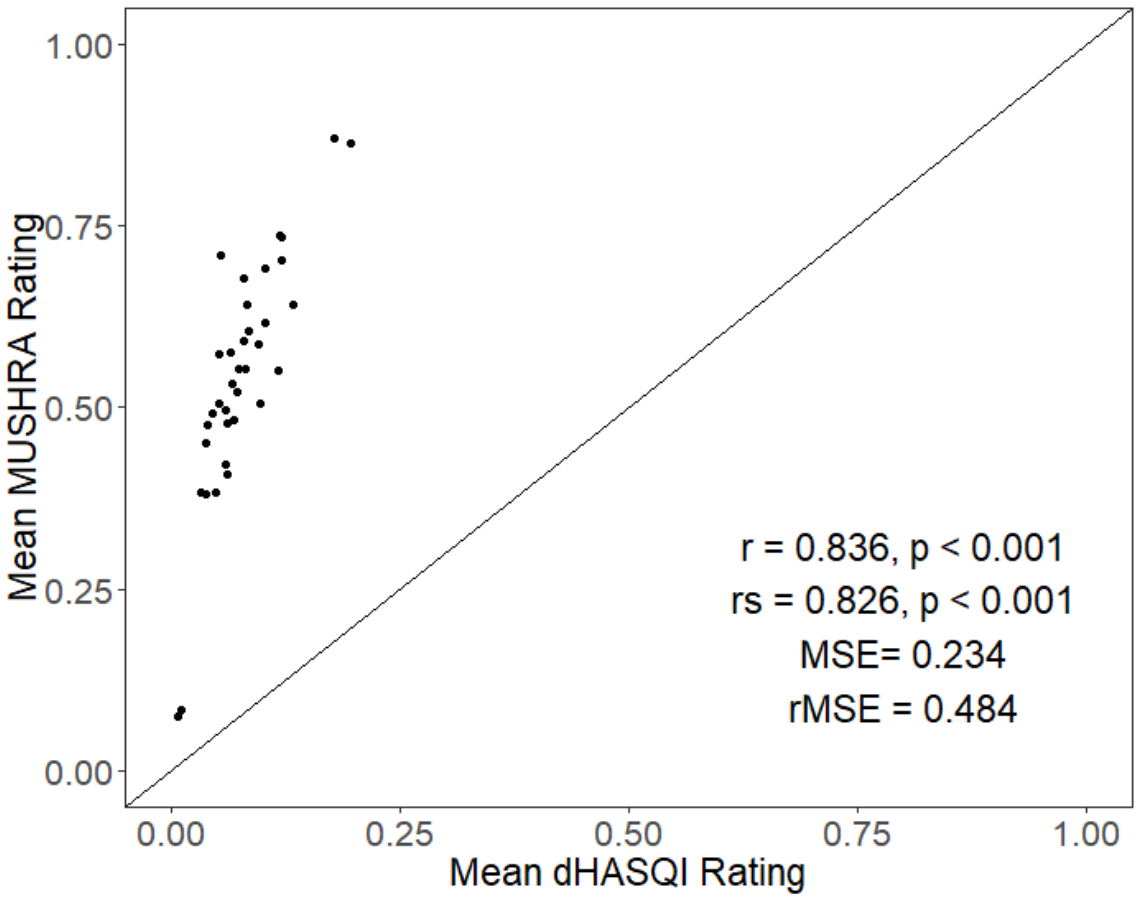

Figure 3-9: Scatterplot of mean MUSHRA ratings and dHASQI quality predictions for the automatic noise reduction dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error.

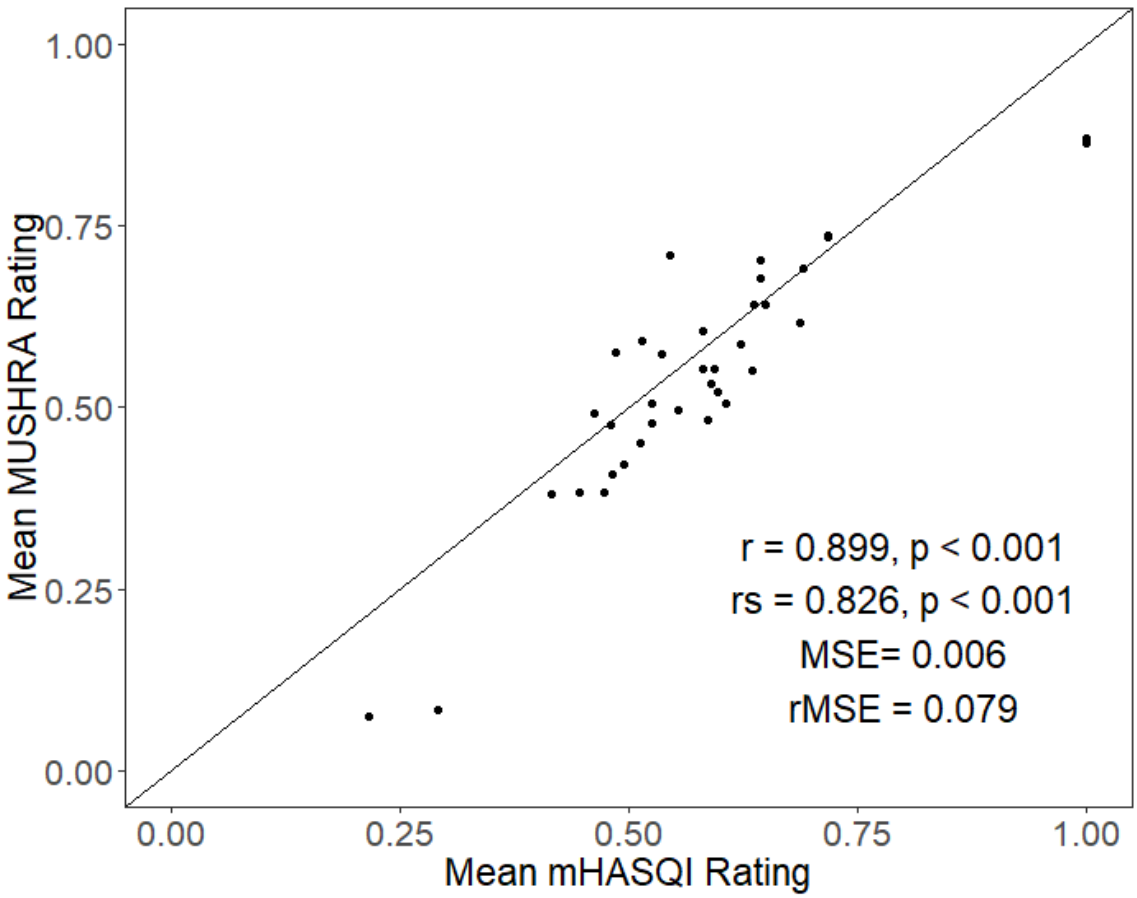

Figure 3-10: Scatterplot of mean MUSHRA ratings and mHASQI quality predictions for the automatic noise reduction dataset using a recorded hearing aid reference signal. r: Pearson linear correlation coefficient, r_s: Spearman rank correlation coefficient, MSE: mean square error, rMSE: root mean square error.

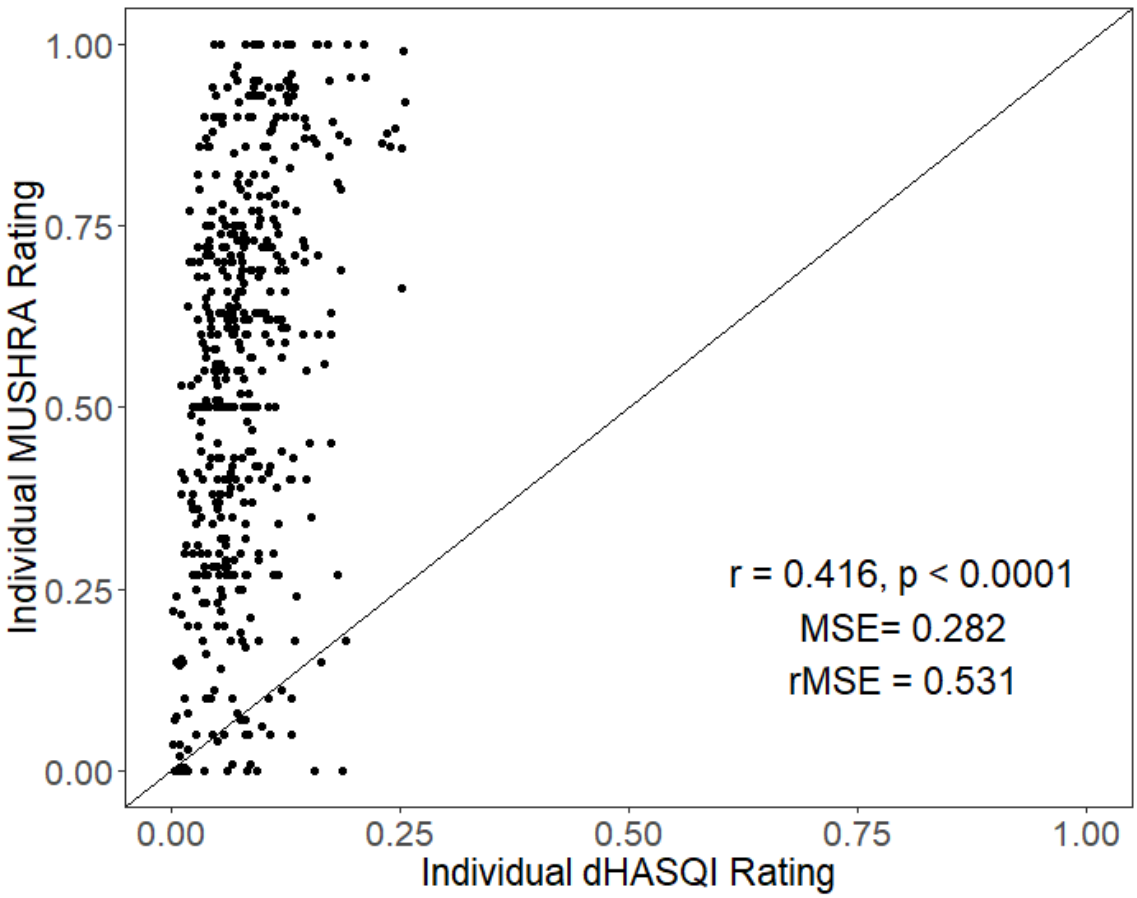

Figure 3-11: Scatterplot of individual MUSHRA ratings and dHASQI quality predictions for the automatic noise reduction dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error.

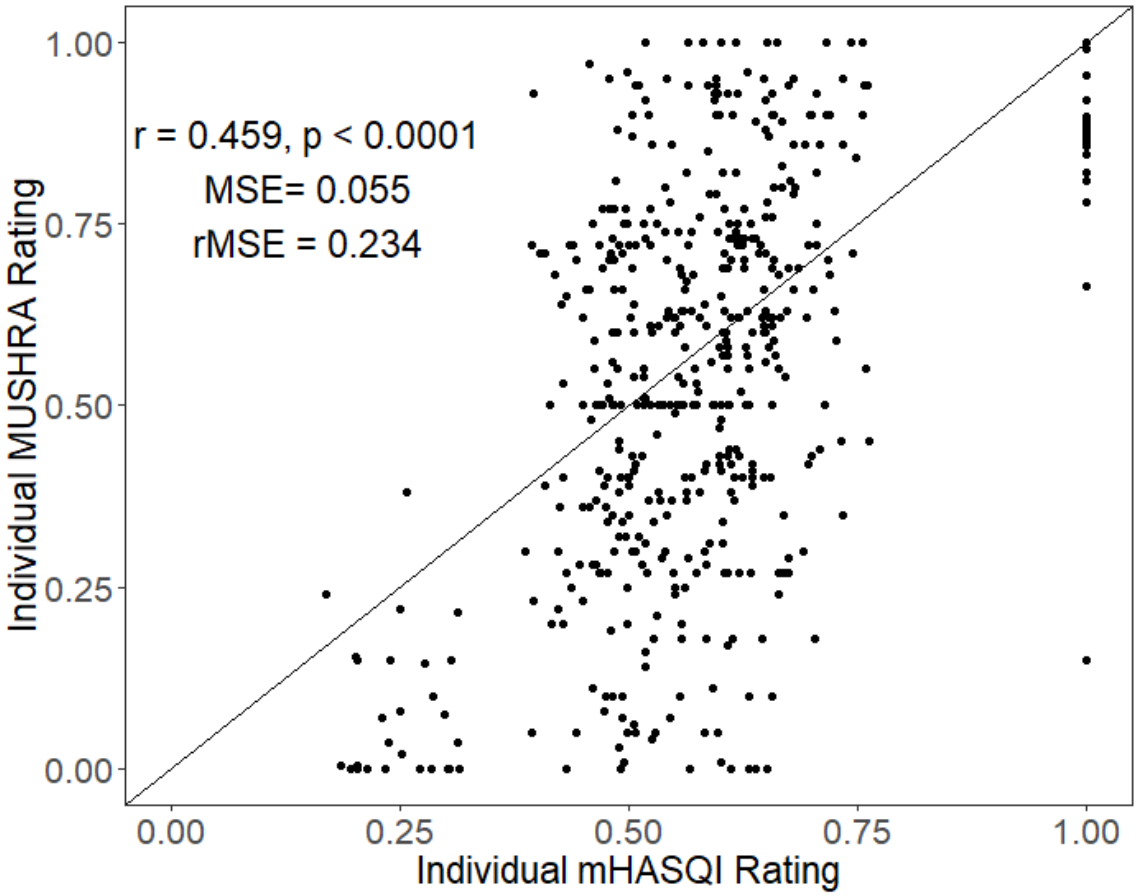

Figure 3-12: Scatterplot of individual MUSHRA ratings and mHASQI quality predictions for the automatic noise reduction dataset using a digitally shaped reference signal. r: Pearson linear correlation coefficient, MSE: mean square error, rMSE: root mean square error.

3.4 Discussion

The goals of this study were to determine: (1) if the HASQI metric predicts hearing aid speech quality rated by hearing-impaired listeners, (2) if HASQI is sensitive to differences in signal-processing adjustments across listeners, and (3) if the artificial ceiling due to inaudible ambient noise and response peaks could be overcome by modifying the HASQI reference signal selection strategy. The results of the study indicate that the HASQI metric is predictive of speech quality, sensitive to differences in signal processing, and that the modified reference strategy was effective. Specifically, the linear correlation coefficients across all three datasets were 0.84 or higher, indicating a strong relationship between HASQI and subjective listener ratings. These findings suggest that HASQI predicted the speech quality from real stimulus adjustments relatively well, and the correlation coefficients were comparable with previous HASQI validation studies for hearing aid signal-processing adjustments as rated by hearing-impaired listeners (Huber et al., 2014; Kates & Arehart, 2014). Further, the average HASQI scores across listeners were sensitive to changes in speech quality due to signal-processing adjustments. Finally, the use of a reference signal that was recorded using the same recording apparatus as the test signal reduced the MSE and rMSE, essentially removing the artificial ceiling caused by inaudible ambient noise in the recording apparatus. This allowed the range of HASQI outputs to be more representative of listeners' ratings.

3.4.1 Sensitivity to signal processing

HASQI was sensitive to signal-processing adjustments present in each of the three validation datasets, as indicated by the values in Tables 3-3 and 3-4. This was the case for both reference signal strategies. In the filtered bandwidth dataset, HASQI scores increased significantly as the low-pass filter cut-off in the stimulus was increased. Interestingly, HASQI did not produce significantly different scores between the 4 kHz cut-off and wideband conditions, while the listeners did detect a difference in the validation dataset. The mean threshold for these listeners at 4 kHz was just under 60 dB HL and increased as frequency increased. Extended high-frequency amplification (especially at 4 kHz and above) is not always beneficial for speech quality (Arehart et al., 2010; Versfeld, Festen, & Houtgast, 1999) especially for listeners with steeply sloping audiograms (Moore, Füllgrabe, & Stone, 2011; Ricketts et al., 2008). HASQI's middle ear simulation also attenuates energy above 5 kHz (Kates, 2013), which would reduce the high frequency energy present in the wideband signal. This reduction may have impacted the wideband condition above 5 kHz sufficiently to produce a comparable score to the 4 kHz condition. In summary, while the sensitivity findings from the filtered bandwidth validation may seem trivial relative to advanced signal-processing characteristics studied in the other datasets, the correlation coefficient confirmed that our HASQI implementation was strongly predictive of subjective ratings, and was sensitive to most stimulus changes that reflect the audibility provided by many hearing aids (Kimlinger et al., 2015).

Table 3-3: Summary of sensitivity ANOVA results for dHASQI. df = degrees of freedom.

Dataset	Factor	num df	denom df	F	p
Bandwidth	Cut-off	1.99	39.73	444.46	< 0.0001
Frequency compression	Frequency compression	1.86	14.89	38.86	< 0.0001
	Stimulus	1	8	3.77	0.09
	Frequency Compression X Stimulus	1.38	19.43	70	< 0.0001
Automatic noise reduction	Manufacturer	1.49	19.43	70	< 0.0001
	Listening Condition	1.12	14.58	226.59	< 0.0001
	Noise reduction	1	13	240.65	< 0.0001
	Manufacturer X Listening Condition	2.32	30.17	104.82	< 0.0001
	Manufacturer X Noise reduction	1.54	20	73.42	< 0.0001
	Listening Condition X Noise Reduction	1.79	23.42	138.02	< 0.0001
	Manufacturer X Listening Condition X Noise Reduction	2.76	35.87	109.36	< 0.0001

Table 3-4: Summary of sensitivity ANOVA results for mHASQI. df = degrees of freedom.

Dataset	Factor	num df	denom df	F	p
Bandwidth	Cut-off	2.03	40.67	497.98	< 0.0001
Frequency compression	Frequency compression	2.75	21.99	131.13	< 0.0001
	Stimulus	1	8	7.99	< 0.05
	Frequency Compression X Stimulus	1.84	14.69	1.83	0.2
Automatic noise reduction	Manufacturer	1.49	27.69	217.32	< 0.0001
	Listening Condition	1.92	24.99	642.63	< 0.0001
	Noise reduction	1	13	265.26	< 0.0001
	Manufacturer X Listening Condition	4.69	60.97	53.02	< 0.0001
	Manufacturer X Noise reduction	2.46	31.90	111.01	< 0.0001
	Listening Condition X Noise Reduction	2.57	33.42	162.1	< 0.0001
	Manufacturer X Listening Condition X Noise Reduction	4.54	58.98	62.95	< 0.0001

The findings from the frequency lowering dataset validation are consistent with those of Kates, Arehart, Anderson, Muralimanohar, & Harvey (2018). We found that HASQI was sensitive to differences in frequency lowering strength, particularly between the ANFC-max versus FC-off conditions and NFC-max versus FC-off conditions. That is, HASQI produced lower scores as the degree of frequency lowering processing was increased to its strongest settings. Kates et al. (2018) found significant differences in HASQI scores between mild, moderate, and maximum frequency lowering processing conditions. Our study did not find a HASQI rating difference between the FC-off and moderate FC-fine-tuned conditions. Recall that this moderate setting was fine-tuned following a protocol designed to provide individualized audibility improvement while minimizing sound quality decrement and that the participants in this study were candidates for FC (Glista et al., 2018). This result may provide objective validation of this fitting method and may also draw a distinction between measurement of FC sound quality effects in participants who are candidates versus participants who are not candidates for this technology. Finally, this study evaluated FC in isolation, not in combination with other signal processing, unlike the study by Kates et al. (2018), which may also explain the difference between these results and those of the previous study. The findings from our study suggest that HASQI did not degrade speech quality as judged by FC candidates when fine-tuning adjustments of frequency lowering were applied. These findings do not address the impact of FC on sound quality when combined with other additional signal processing features.

The findings from the automatic noise reduction dataset validation are also consistent with the findings from Kates et al. (2018). In our study, HASQI scores were significantly different between hearing aids from four manufacturers. This suggests that the combination of hardware and signal-processing between hearing aids is sufficiently different to produce meaningfully different speech quality, which is consistent with the findings of Kates et al. (2018). We also found that increasing noise reduction to its maximum setting improved speech quality when averaged across all four hearing aids, as well as within each hearing aid. This can be interpreted as an outcome that may be beneficial to hearing aid users in the application of noise reduction algorithms. Finally, we found that each combination of SNR and noise type yielded significantly different degrees of quality change. Recall that the listening conditions used various combinations of noise (multitalker babble or speech-

shaped noise) and SNR (0 or 5 dB). HASQI scores increased as SNR increased, and HASQI scores were greater in multitalker babble compared to speech-shaped noise. As SNR increased, HASQI scores increased significantly, and speech passages with a higher SNR better aligned with the high-quality reference signals. These findings are consistent with those of Kates et al. (2018). The finding that HASQI produced significantly greater scores for multitalker babble compared to speech-shaped noise was interesting and novel but contrasts behavioral findings for speech reception threshold differences between types of competing noise. Multitalker babble consists of intermittent acoustic content, pauses, and peaks which may on occasion mask the speech signal, distorting the test signal relative to the reference signal, but may also leave occasional quiet portions in which speech may not be masked. Furthermore, the vocal content in the multitalker babble more closely resembles the vocal content in the test signal, which may reduce the measured differences between the test signal and reference signal compared to other types of non-vocal noise. Speech-shaped noise, on the other hand, consists of the spectral, but not temporal characteristics of acoustic vocal content and masks the entire duration of the primary speech signal. Based on the present results, speech-shaped noise degrades speech quality more than multitalker babble. Behaviorally, normal-hearing listeners can take advantage of the quiet periods in multitalker babble to improve intelligibility (known as “listening in the dips” or “glimpsing” (Cooke, 2006; Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006). However, hearing-impaired listeners repeat sentences no more effectively in competing speech than in steady-state noise, because they lack the temporal fine structure cues to do so (Duquesnoy, 1983; Lorenzi et al., 2006; Moore, 2008). Further research is needed to understand the difference between subjective quality ratings for different types of competing noise and how they relate to HASQI.

3.4.2 HASQI across hearing aid users

One of the interesting aspects of this study design was that stimulus adjustments were treated as within-subjects variables across individuals wearing real hearing aids. Kates et al. (2018) tested the impact of hearing aids programmed for standardized audiograms (Bisgaard et al., 2010) as a within-subjects variable, and found that HASQI produced higher scores for an S2 (mild-sloping-to-severe) versus an N4 (flat moderately severe) audiogram. This was interpreted as the S2 having less hearing loss in the low frequencies, thereby requiring less hearing aid signal-processing, which in turn resulted in minimally-distorted speech passages. Our study investigated the effect of audiogram from the perspective of generalizability of signal-processing adjustments on HASQI scores across listeners with a range of audiograms, rather than across standard audiograms. We found that HASQI was sensitive to most signal-processing adjustments across all the audiograms in their respective datasets, despite the differences in signal-processing changes associated with each audiogram. These individual differences included (a) the varying impact of changes in acoustic bandwidth on audible bandwidth; (b) the individualized setting of frequency compression using a fine-tuning protocol; and (c) the varying impact of fixed SNR and fixed noise reduction strength across the varying auditory dynamic ranges of listeners. These findings suggest that changes in each of these signal parameters elicit changes in speech quality, at least across the audiograms tested here. It may be important to study the generalizability of HASQI for other hearing profiles, such as those with reverse-sloping, conductive, or severe-to-profound hearing losses.

3.4.3 Effects of inaudible noise

HASQI's sensitivity to inaudible ambient noise restricts the highest HASQI value that can be measured (Kates et al., 2018), and this was observed in our study. The dHASQI metric, using a clean digitally-shaped reference, produced a value that underestimated its corresponding speech quality rating. The underestimation likely occurred due to an SNR that was sufficiently large for speech to be subjectively rated as relatively high-quality. For example, the dHASQI scores were relatively low for conditions that had relatively high subjective scores of 0.75 or higher (i.e., the fine-tuned ANFC condition, the +10 dB SNR condition, and the full bandwidth condition) and all scores showed a clear upper limit of

measurement. Estimates of speech audibility only require the peaks of the signal in their calculation, corresponding to an SNR of about 10 dB (ANSI, 1997, 2014; Kates et al., 2018). This implies that intelligibility metrics only measure portions of the signal in which speech content is present and not entirely masked. The SII also includes corrections to account for upward spread of masking and/or in-band masking for when noise is present. HASQI, on the other hand, relies on signal fluctuations for the entire signal duration. Unlike the SII, it does not rely on speech peaks alone and it does not include corrections for masking. This implies that noise during low-level portions of speech will contribute to the overall quality rating. This was apparent for the automatic noise reduction dataset reference condition dHASQI score, in that a +10 dB SNR produced a low score, and for the filtered bandwidth and frequency lowering reference conditions, in that inaudible ambient noise limited the dHASQI score to below 0.5. Kates et al. (2018) reduced the impact of inaudible noise by averaging multiple HASQI analyses for the same condition. This strategy, however, continued to yield a relatively low score.

These impacts of low-level noise were the motivation for investigating the use of an alternative reference signal strategy. The dHASQI analysis, which compared hearing aid recordings to digitally-shaped references, underestimated the corresponding subjective ratings across all three validation datasets, and this was confirmed with the MSE and rMSE measurements. The mHASQI analysis, which compared hearing aid test recordings to hearing aid reference recordings that were recorded using the same apparatus, annulled the impact of inaudible noise and ensured that the analysis was based on the effects of additive noise and/or automatic noise reduction alone. The reference condition in all datasets yielded perfect HASQI scores of 1.00. The mid-level fine-tuned ANFC condition from the FC dataset yielded a HASQI score of 0.685, corresponding to a subjective rating of 0.722. This strategy showed smaller errors between average HASQI scores and average listener ratings. The recorded reference strategy produced mHASQI scores that were more representative of real human subjective ratings for all three validation datasets that we tested.

Another contributor to the artificial ceiling observed in the dHASQI scores, specifically in the current study, may be attributed to the use of the DSL v5.0 formula. HASQI shapes its

reference signal and training data test signals (Arehart et al., 2010; Kates & Arehart, 2014) for listener thresholds using the linear NAL-R prescriptive formula (Byrne & Dillon, 1986). Therefore, the same gains are applied to a signal regardless of level fluctuations within the signal. In the current study, DSL v5.0 was used to amplify the reference and test signals to match the prescriptive formula implemented in the corresponding subjective datasets. DSL v5.0 is a nonlinear prescriptive formula (Scollie et al., 2005) meaning that softer sounds are amplified with more gain compared to louder sounds. In the dHASQI analysis, the digitally-shaped reference signal was linearly shaped using DSL v5.0-prescribed gains that were derived from the average input level of the corresponding stimulus. In other words, both the loud and soft signals of the reference signal were amplified with the same gains. However, like the stimuli in the subjective tests, the test signals were amplified using nonlinear gains and this may have introduced additional distortions relative to the linear digitally-shaped reference. This would have inflated the sound quality degradations, contributing to the artificial ceiling observed in the data. This issue also would have occurred if linear NAL-R shaping was digitally applied to the reference signal, as specified by HASQI (Kates & Arehart, 2014). This issue was subsided using a recorded reference signal, because the reference signal was amplified using nonlinear gains like the test signal. Future HASQI investigations may look to digitally shape the reference signal using a nonlinear prescriptive formula through a hearing aid simulator, such as the University of California: San Diego's Open Speech Platform (Garudadri et al., 2017) or the University of Oldenburg's Open Source Master Hearing Aid (Herzke, Kayser, Loshaj, Grimm, & Hohmann, 2017). This would allow the reference signal to better resemble signals produced by today's nonlinear commercial hearing aids.

3.4.4 Implementation considerations

While mHASQI resolved the issue of degradations in the signal due to inaudible low-level noise, there are still considerations to be addressed. The rationale behind mHASQI was to use a recorded reference signal so that inaudible low-level noise would not artificially inflate the degree of signal degradation. This strategy resulted in HASQI scores that were better aligned with listener ratings. This reference signal strategy, however, limits the use of mHASQI to a single hearing aid and only allows for signal-processing quality measurements of that instrument alone. Different manufacturers' hearing aids have different receiver responses and different amounts of processing noise, and these differences are sufficient to produce statistically different MUSHRA ratings (Scollie, Levy, Pourmand, et al., 2016) and statistically different HASQI scores using a digitally-shaped reference signal (Kates et al., 2018). The recorded reference signal strategy biased mHASQI measurements towards the device used as the reference. For example, in the automatic noise reduction dataset, the reference signal was recorded through only one of the hearing aids. The mean predicted quality value for the reference-based hearing aid was 0.61, whereas the others ranged between 0.54 and 0.57. In the dHASQI analysis, a different hearing aid yielded the highest score. In the mHASQI analysis, the non-reference hearing aids' scores were affected by degradations from multiple sources (i.e., differences in receiver response and differences in processing noise between devices) whereas the reference-based hearing aids' scores were affected by degradations from only one source (signal-processing degradations). This reference signal strategy effectively prevents the use of mHASQI for cross-manufacturer comparisons, which may be problematic if the relative differences in mHASQI were used to select a hearing aid, for example. A future HASQI investigation using the recorded reference-signal strategy could record the reference signal through the output of a high-fidelity earphone that has been matched to prescriptive targets. These recordings would include environmental noise from the measurement system but would omit hearing aid processing noise as there is likely less noise in the earphone transducer compared to a hearing device that includes digital-signal-processing hardware. This strategy would omit ambient noise caused by the recording system and include sound quality degradations due to signal-processing, receiver responses, and processing noise between hearing aids without bias towards the device being used to record the reference

signal. The presence of hearing aid noise may continue to introduce a ceiling effect, but less so compared to a digital reference strategy where the recording noise is also considered.

Another implementation consideration relates to the sensitivity of dHASQI and mHASQI to differences between stimulus conditions. While the two metrics were highly related to subjective human ratings, and successfully detected differences between conditions, the statistical differences within dHASQI and within mHASQI did not always correspond to the corresponding statistical differences between subjective human ratings. Table 3-5 displays a summary of RM-ANOVA effects for dHASQI, mHASQI and the MUSHRA ratings from each of the validation datasets. In the frequency lowering dataset, only mHASQI detected a significant difference between stimuli. In contrast, only dHASQI detected a significant interaction between frequency-compression setting and stimulus. In this dataset, each stimulus was a female-spoken passage with predictable acoustic content by the same speaker. Therefore, the difference in quality scores was likely attributable to spectral differences between the words in the speech passages, despite the two passages yielding promising test-retest intraclass correlations of 0.89 and higher in the behavioral study. Likewise, the main effect of noise reduction in the automatic noise reduction study was not apparent in the corresponding behavioral dataset. Both HASQI implementations found that all four hearing aids' noise reduction algorithms significantly improved speech quality, whereas listeners subjectively found that only one hearing aid's noise reduction algorithm improved perceived speech quality. The HASQI implementations also detected significant differences in speech quality between almost all four manufacturers' hearing aids, whereas listeners were unable to differentiate speech quality between several hearing aid pairings. Furthermore, the ordering of hearing aids from most preferred to least preferred was not the same between predicted and behavioral results. Behavioral ratings are highly variable and this was illustrated by when comparing scatterplots for individual speech quality scores to scatterplots for mean speech quality scores. Across all datasets, correlation strength was weaker and MSE and rMSE values were greater using individual scores, suggesting poorer model accuracy and predictability when using individual scores rather than average scores. Individual variability can be influenced by many factors including acclimatization, cognition, loudness discomfort levels, and intelligibility versus

quality preferences - factors that are not included in computing the HASQI metric (Kates et al., 2018). Behavioral ratings within individuals can also lack reliability between test sessions (Gabrielsson et al., 1988; Narendran & Humes, 2003). In contrast, the HASQI scores for each single condition in this study only varied by differences in prescribed DSL v5.0 reference-shaping between listeners and not individual variability. HASQI may have also been sensitive to imperceptible differences between various signal-processing settings, as it was to inaudible ambient noise. In summary, compared to human listeners HASQI was more sensitive to signal-processing adjustments, despite a strong positive association between predicted and behavioral results. Therefore, clinical use of HASQI would be most appropriate if future studies established the minimum change in HASQI necessary to produce a change in perceived sound quality. Otherwise, small changes to hearing aid fitting parameters could be misinterpreted as clinically important if based on HASQI score changes in isolation.

Table 3-5: Significance of statistical tests across dHASQI, mHASQI, and subjective ratings from original studies. Sig = statistically significant, Non-Sig = statistically non-significant.

Dataset	Factor	dHASQI	mHASQI	Subjective
Bandwidth	Cut-off	Sig	Sig	Sig
Frequency compression	Frequency compression	Sig	Sig	Sig
	Stimulus	Non-Sig	Sig	Not Reported
	Frequency Compression X Stimulus	Sig	Sig	Sig
Automatic noise reduction	Manufacturer	Sig	Sig	Sig
	Listening Condition	Sig	Sig	Sig
	Noise reduction	Sig	Sig	Non-Sig
	Manufacturer X Listening Condition	Sig	Sig	Not Reported
	Manufacturer X Noise reduction	Sig	Sig	Not Reported
	Listening Condition X Noise Reduction	Sig	Sig	Sig
	Manufacturer X Listening Condition X Noise Reduction	Sig	Sig	Not Reported

3.5 Conclusions and future directions

The findings from this article contribute to the research literature on applying the objective speech quality metric, HASQI, to hearing aid processed speech by extending past findings to new hearing aid and signal parameters, by testing across groups of hearing aid users, and by contrasting two strategies for developing a reference signal. The findings show that HASQI is positively associated with behavioral speech quality ratings of hearing aid processed stimuli. The findings also show that HASQI is sensitive to differences in hearing aid signal-processing adjustments across a group of hearing-impaired listeners. The dHASQI implementation (in which HASQI implemented a digital reference) revealed that this metric can detect differences between the hardware and signal-processing combinations belonging to different hearing aids. This analysis also revealed dHASQI's sensitivity to background and processing noise, which creates an artificial measurement ceiling and contributes a significant discrepancy between predicted and behavioral results. The mHASQI implementation incorporated a recorded reference signal strategy and overcame the ceiling issue by using the same recording apparatus for the reference and test signals, so both signals would contain the same inaudible ambient noise. However, the use of a recorded hearing aid reference signal also biased mHASQI towards the test hearing aid that was used to make the recording, removing the utility of cross-device comparisons in this analysis.

There are several other areas to consider in future studies. First, the datasets analyzed here did not encompass all the types of signal processing that occur in hearing aids or even all types of hearing aids, nor did they represent all degradations encountered in realistic listening environments (such as reverberation, different talker's voices, and spatial configuration of talkers and maskers, to name a few). Second, two of the datasets (filtered bandwidth and automatic noise reduction) were gathered from listeners presenting with high frequency thresholds no greater than 75 dB HL on average, so the results do not necessarily transfer to listeners with greater degrees of hearing loss. Third, the validity and sensitivity here were measured over perceptual listener averages, rather than incorporating data points for each individual listener. At the individual level, subjective variability may produce lower correlation coefficients with objective metrics such as HASQI. Finally, the

digital reference signal was amplified using linear amplification, which is not representative of the nonlinear fitting formulas used to prescribed most of today's commercial hearing aids. These considerations highlight many possible directions for future research on objective metrics of speech quality. Future investigations should examine additional signal-processing features (such as feedback cancellation, speech enhancement, different gain models, etc.) across a variety of manufacturers' hearing aids, across a wider a range of audiograms, and in as many realistic listening environments as possible. Future research should also investigate the impact of processing the digital reference signals using nonlinear amplification. Furthermore, this research will be complemented by HASQI analyses at the individual level, behavioral investigations of factors influencing within-individual variability, and the incorporation of those factors in future model developments.

3.6 References

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

Alexander, J. (2013). Individual variability in recognition of frequency-lowered speech. *Seminars in Hearing*, *34*(2), 86–109.

American National Standards Institute. (1997). *Methods for Calculation of the Speech Intelligiblity Index. ANSI S3.5-1997 (R2017)*. New York: Acoustical Society of America.

American National Standards Institute. (2014). *Specification of Hearing Aid Characteristcs. ANSI S3.22-2014*. New York: Acoustical Society of America.

Amlani, A. M., Punch, J. L., & Ching, T. Y. C. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification*, *6*(3), 81–129.

Amlani, A. M., & Schafer, E. C. (2009). Application of paired-comparison methods to hearing aids. *Trends in Amplification*, *13*(4), 241–259.

Arehart, K. H., Kates, J. M., & Anderson, M. C. (2010). Effects of noise, nonlinear processing, and linear filtering on perceived speech quality. *Ear and Hearing*, *31*(3), 420–436.

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

Baker, S., & Jenstad, L. (2017). Matching real-ear targets for adult hearing aid fittings: NAL-NL1 and DSL v5.0 prescriptive formulae. *Canadian Journal of Speech-Language Pathology and Audiology*, *41*(2), 227–235.

Beerends, J. G., Hekstra, A. P., Rix, A. W., & Hollier, M. P. (2002). Perceptual evaluation of speech quality (PESQ) the new ITU standard for end-to-end speech quality assessment part II: Psychoacoustic model. *Journal of the Audio Engineering Society*, *50*(10), 765–778.

Beerends, J. G., Krebber, J., Huber, R., Eneman, K., & Luts, H. (2008). Speech quality measurement for the hearing impaired on the basis of PESQ. In *124th Convention of the Audio Engineering Society* (p. convention paper 7404).

Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, *14*(2), 113–120.

Byrne, D., & Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, *7*(4), 257–265.

Chen, G., Parsa, V., & Scollie, S. (2006). An ERB loudness pattern based objective speech quality measure. *9th International Conference on Spoken Language Processing*, 2174–2177.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, *119*(3), 1562–1573.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America*, *74*(3), 739–743.

Easwar, V., Purcell, D. W., Aiken, S. J., Parsa, V., & Scollie, S. D. (2015). Evaluation of speech-evoked envelope following responses as an objective aided outcome

measure: Effect of stimulus level , bandwidth , and amplification in adults with hearing loss. *Ear and Hearing*, *20*(10), 635–652.

Etymotic Research. (2001). *QuickSIN Speech-In-Noise Test*. 61 Martin Lane, Elk Grove Village, Illinois 60007.

Falk, T. H., Parsa, V., Santos, J. F., Arehart, K., Hazrati, O., Huber, R., … Scollie, S. (2015). Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools. *IEEE Signal Processing Magazine*, *32*(2), 114–124.

Gabrielsson, A., Schenkman, B. N., & Hagerman, B. (1988). The effects of different frequency responses on sound quality judgements and speech intelligibility. *Journal of Speech and Hearing Research*, *31*, 166–177.

Gabrielsson, A., & Sjögren, H. (1979a). Perceived sound quality of hearing aids. *Scandinavian Audiology*, *8*(3), 155–169.

Gabrielsson, A., & Sjögren, H. (1979b). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033.

Garudadri, H., Boothroyd, A., Lee, C., Gadiyaram, S., Bell, J., Sengupta, D., … Rao, B. D. (2017). A realtime, open-source speech-processing platform for research in hearing loss compensation. In *Proceedings of the IEEE 51st Asilomar Conference on Signals, Systems, and Computers* (pp. 1900–1904). Retrieved from http://mesl.ucsd.edu/mesl-website/pubs/Garudadri_ASILOMAR2017.pdf

Glista, D., Hawkins, M., Vaisberg, J. M., Pourmand, N., Parsa, V., & Scollie, S. (2018). Sound quality effects of an adaptive nonlinear frequency compression processor with normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*, *Epub ahead of print*.

Gray, C. D., & Kinnear, P. R. (1999). *SPSS for Windows Made Simple*. Hove: Psychology Press.

Gustafson, S. J., & Pittman, A. L. (2010). Sentence perception in listening conditions having similar speech intelligibility indices. *International Journal of Audiology*, *50*(1), 34–40.

Harlander, N., Huber, R., & Ewert, S. D. (2014). Sound quality assessment using auditory models. *Journal of the Audio Engineering Society*, *64*(5), 324–336.

Herzke, T., Kayser, H., Loshaj, F., Grimm, G., & Hohmann, V. (2017). Open signal processing software platform for hearing aid research (openMHA). In *Proceedings of the Linux Audio Conference* (pp. 35–42). Retrieved from http://musinf.univ-st-etienne.fr/lac2017/pdfs/05_C_G_141775.pdf

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, *49*(12), 891–903.

Houben, R., Brons, I., & Dreschler, W. A. (2011). A nethod to remove differences in frequency response between commercial hearing aids to allow direct comparison of the sound quality of hearing-aid features. *Trends in Amplification*, *1–2*, 77–83.

Hu, Y., & Loizou, P. C. (2008). Evaluation of Objective Quality Measures for Speech Enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(1), 229–238.

Huber, R., & Kollmeier, B. (2006). PEMO-Q - A new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio,*

Speech and Language Processing, *14*(6), 1902–1911.
Huber, R., Parsa, V., & Scollie, S. (2014). Predicting the perceived sound quality of frequency-compressed speech. *PloS One*, *9*(11), e110260.
International Telecommunications Union (2015). *Method for subjective assessment of intermediate quality level of audio systems*. *Recommendation ITU-R BS.1534-3* Geneva, Switzerland.
International Telecommunications Union Standardization Sector (1996). *Methods for subjective determination of tranmission quality*. *Recommendation ITU-T P.800.* Geneva, Switzerland.
Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology*, *14*(7), 347–360.
Kates, J. (2013). An auditory model for intelligibility and quality predictions. In *Proceedings of Meetings on Acoustics* (Vol. 19, p. 050184). Montreal: Acoustical Society of America.
Kates, J. M., & Arehart, K. H. (2014). The hearing-aid speech quality index (HASQI) version 2. *Journal of the Audio Engineering Society*, *62*(3), 99–117.
Kates, J. M., Arehart, K. H., Anderson, M. C., Muralimanohar, R. K., & Harvey, L. O. (2018). Using objective metrics to measure hearing aid performance. *Ear & Hearing*, *36*(6), 1165-1175.
Kimlinger, C., McCreery, R., & Lewis, D. (2015). High-frequency audibility: The effects of audiometric configuration, stimulus type, and device. *Journal of the American Academy of Audiology*, *26*(2), 128–137.
Kondo, K. (2012). Speech Quality. In *Subjective Quality Measurement of Speech: Its Evaluation, Estimation and Applications*. Heidelberg New York Dordrecht London: Springer.
Kressner, A. K., Anderson, D. V, & Rozell, C. J. (2013). Evaluating the generalization of the hearing aid speech quality index (HASQI). *IEEE Transactions on Audio, Speech and Language Processing*, *21*(2), 407–415.
Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. Retrieved from https://cran.r-project.org/package=ez
Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, *103*(49), 18866–18869.
Moodie, S. T. F., Scollie, S. D., Bagatto, M. P., & Keene, K. (2017). Fit-to-targets for the desired sensation level version 5.0a hearing aid prescription method for children. *American Journal of Audiology*, *26*(3), 251–258.
Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, *9*(4), 399–406.
Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, *32*(5), 556–568.
Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, *114*(1), 408–419.
Narendran, M. M., & Humes, L. E. (2003). Reliability and validity of judgments of sound

quality in elderly hearing aid wearers. *Ear and Hearing*, *24*(1), 4–11.

Parsa, V., Scollie, S., Glista, D., & Seelisch, A. (2013). Nonlinear frequency compression: Effects on sound quality ratings of speech and music. *Trends in Amplification*, *17*(1), 54–68.

Pourmand, N., Parsa, V., & Weaver, A. (2013). Computational auditory models in predicting noise reduction performance for wideband telephony applications. *International Journal of Speech Technology*, *16*(4), 363–379.

Preminger, J. E., & Van Tasell, D. J. (1995). Quantifying the relation between speech quality and speech intelligibility. *Journal of Speech and Hearing Research*, *38*, 714–725.

Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*, 160–172.

Rohdenburg, T., Hohmann, V., & Kollmeier, B. (2005). Objective perceptual quality measures for the evaluation of noise reduction schemes. *9th International Workshop on Acoustic Echo and Noise Control*, 169–172.

Scollie, S. (2018). 20Q: Using the aided speech intelligibility index in hearing aid fittings. *AudiologyOnline*, (23707). Retrieved from www.audiologyonline.com

Scollie, S., Glista, D., Seto, J., Dunn, A., Schuett, B., Hawkins, M., … Parsa, V. (2016). Fitting frequency-lowering signal processing applying the American Academy of Audiology pediatric amplification guideline: Updates and protocols. *Journal of the American Academy of Audiology*, *27*(3), 219–236.

Scollie, S., Levy, C., Pourmand, N., Abbasalipour, P., Bagatto, M., Richert, F., … Parsa, V. (2016). Fitting noise management signal processing applying the American Academy of Audiology pediatric amplification guideline: Verification protocols. *Journal of the American Academy of Audiology*, *27*(3), 237–251.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., … Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, *9*(4), 159–197.

Suelzle, D., Parsa, V., & Falk, T. H. (2013). On a reference-free speech quality estimator for hearing aids. *The Journal of the Acoustical Society of America*, *133*(5), EL412-EL418.

R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Compting, Vienna, Austria. URL Https://Www.R-Project.Org/*.

Thiede, T., Treurniet, W. C., Bitto, R., Beerends, J. G., Olomes, C. C., Keyhl, C. H., … Leidschendam, N. L. a K. (2000). PEAQ-- The ITU standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Sociaty*, *48*(1/2), 3–29.

Treurniet, W. C., & Soulodre, G. A. (2000). Evaluation of the ITU-R objective audio quality measurement method. *Journal of the Audio Engineering Society*, *48*(3), 164–173.

Vaisberg, J. M., Folkeard, P., Pumford, J., Narten, P., & Scollie, S. (2018). Evaluation of the repeatability and accuracy of the wideband real-ear-to-coupler difference. *Journal of the American Academy of Audiology*, *29*(6), 520-532.

Versfeld, N. J., Festen, J. M., & Houtgast, T. (1999). Preference judgments of artificial

processed and hearing-aid transduced speech. *The Journal of the Acoustical Society of America*, *106*(3), 1566–1578.

Wong, L. L. N., Hickson, L., & Mcpherson, B. (2003). Hearing aid satisfaction: What does research from the past 20 years say ? *Trends in Amplification*, *7*(4), 117–161.

Chapter 4

4 Comparison of music sound quality between hearing aids and music programs[7]

4.1 Introduction

Listening to music is an important and enjoyable part of many people's lives. Music listening can improve quality of life through its recreational and rehabilitative function. For example, music involvement can enhance IQ in developing children (Hille, Gust, Bitz, & Kammer, 2011), and can mitigate symptoms of Alzheimer's disease in older adults (Simmons-Stern, Budson, & Ally, 2010). Unfortunately, facilitating music listening in hearing aid wearers is not fully understood, and making music enjoyable through hearing aids can be challenging for some fittings. In this article, we will share the results of a sound quality experiment that tested whether music programs in a wide range of leading hearing aid models provide good music quality, including for the listeners' own favorite music passages.

4.1.1 The acoustics of speech and music

We fit hearing aids to improve audibility of sounds, such as speech and music. However, hearing aids are often programmed for listening to speech – and we should remind ourselves that speech and music can be quite different. Speech is produced by the human vocal tract and has well-defined acoustic properties that are fairly consistent across languages (Byrne et al., 1994) and that have predictable variations across genders, ages, and vocal effort levels (Byrne et al., 1994; Olsen, 1998). Average speech levels typically vary between 55 and 66 dBA (Olsen, 1998) with a dynamic range of 20 – 30 dB (Holube,

---

[7] A version of this chapter has been published (See Appendix B): Vaisberg, J. M., Folkeard, P., Parsa, V., Froehlich, M., Littmann, V., Macpherson, E. A., & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline*, *Article 20872*. Retrieved from www.audiologyonline.com

Fredelake, Vlaming, & Kollmeier, 2010). Speech tends to have more low-frequency energy and a fundamental frequency as low as 100 Hz and 160 Hz for males and females, respectively (Cornelisse, Gagne, & Seewald, 1991).

Music, in contrast, can originate from a variety of sources, such as voices and instruments. Music has the potential of having a much larger dynamic range, broader frequency spectrum, and higher overall level (Chasin & Hockley, 2014). We can illustrate these differences between speech and music using displays of the energy, across frequencies, in each type of signal across frequency. The speech range is sometimes called a "speech banana", as shown in Figure 4-1. This is compared to the range of energy in music, with both speech and music overlaid on the dynamic range of the human auditory system. The acoustic differences between speech and music are large, and may pose challenges for designing hearing aid programs that work as well for music as they do for speech.

These characteristics can also depend on the exact instruments and genre, and whether music is listened to live or via a recording. For example, Kirchberger & Russo (2016) found that the dynamic range of recorded classical music was between 20-32 dB while the dynamic range of recorded jazz was 13-23 dB, and that there was more relative low frequency energy in a choir genre versus a pop recording. In fact, previous studies have shown that hearing aid sound quality ratings can be affected by music genre (Arehart, Kates, & Anderson, 2011; Davies-Venn, Souza, & Fabry, 2007). Does this mean that fitting hearing aids for music listening is not a generic problem, but rather one that may need to be tailored to the individual music preferences of the hearing aid wearer?

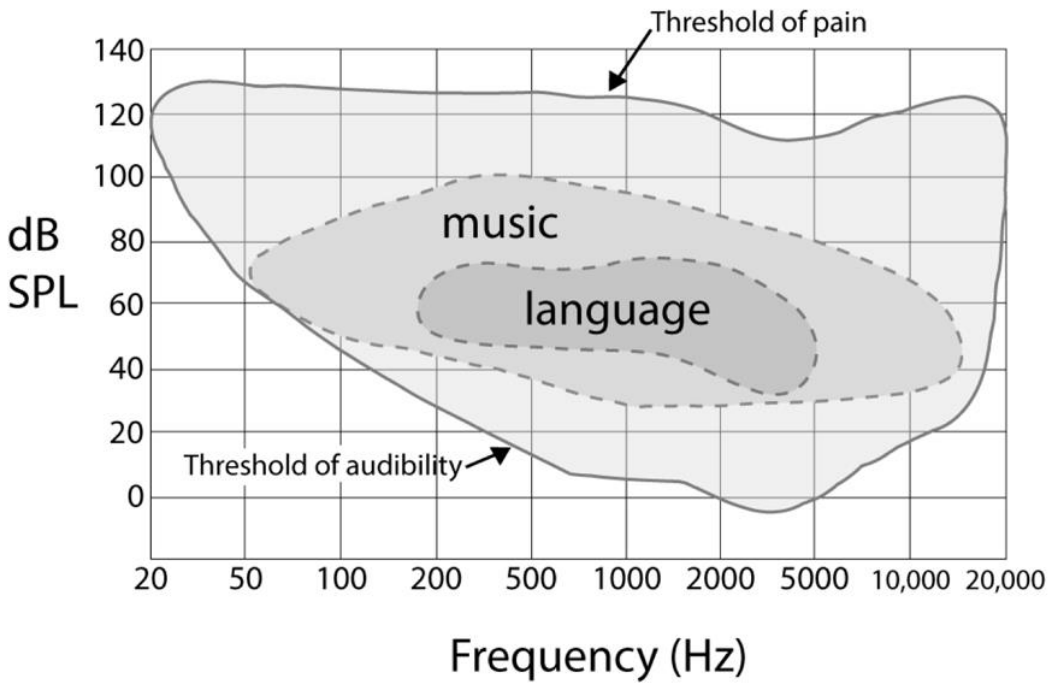

Figure 4-1: Frequency-intensity range of speech and music within the audibility of the human auditory system. Adapted from Limb (2011).

4.1.2 Hearing aids and music programs

Many hearing aid manufacturers have incorporated music programs, designed to improve the sound quality of music, into their products. While the parameters of each music program differ between manufacturers, common features of a music program include slower compression, less noise reduction, reduced directionality, and reduced feedback cancellation, compared to programs intended for use with speech (Moore, 2016). At least one study has shown that different models provide different output levels for music, but also showed that individual preferred listening levels vary considerably across listeners (Croghan, Swanberg, Anderson, & Arehart, 2016; Galster, Rodemerk, & Fitz, 2014).

Previous studies report that many hearing aid users report difficulty listening to amplified music (Leek, Molis, Kubli, & Tufts, 2008). A recent survey found that only 40% of its respondents reported having a music program in their aids (Madsen & Moore, 2014). Those

who did report having a music program reported music listening experiences that were essentially similar to those of users who did not have a music program. This raises the question of whether music programs were effective at improving music listening experiences for these hearing aid users.

Overall, the idea of creating a music program that makes music more enjoyable for hearing aid users is important, but is also complex. The vast differences in music types and genres, hearing aid signal processing, and listener individuality create a highly variable set of possible combinations. For these reasons, it is important to determine whether modern hearing aid music programs produce good sound quality across a wide range of genres and listeners.

4.1.3 Study purpose

The purpose of this study was to examine the sound quality of hearing aid processed music across a wide range of hearing aid models and music genres. Specifically, we compared hearing impaired listeners' subjective sound quality ratings of music processed by several premier hearing aids' respective universal and music programs.

4.2 Methods

4.2.1 Participants

Participants were recruited from the National Centre for Audiology's Translational Research Unit participant database at Western University. A total of 26 adults between ages 20 and 84 (mean = 71 yearrs, standard deviation = 12, 15 males, 11 females) participated in this study. All participants were regular users of hearing aids, and had bilateral symmetrical sensorineural hearing loss. Hearing losses ranged from 35-40 dB in the low frequencies to 65-70 dB in the high frequencies. Figure 4-2 displays the average and individual thresholds for all participants included in this study. This study was approved by the Western University Health Sciences Research Ethics Board. All participants completed informed consent and were compensated for their time.

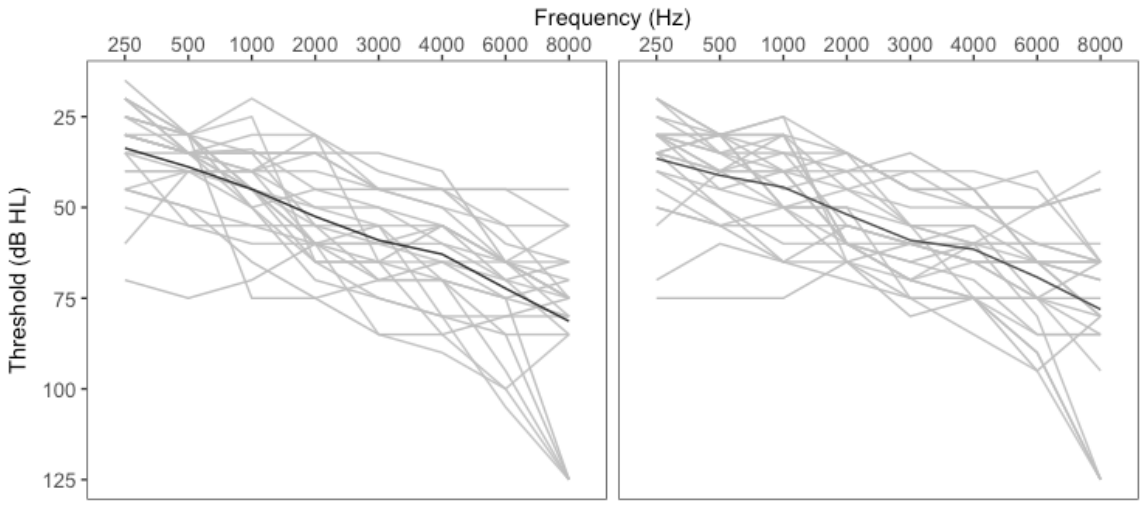

Figure 4-2: Pure-tone air conduction audiometric thresholds for all participants (n=26). The left panel represents thresholds for participants' left ears and the right panel represents thresholds for the participants' right ears. The light grey lines represent individual audiograms and the dark grey lines represent the mean audiograms across participants for each ear.

4.2.2 Test conditions

Hearing aids used in this study included five recently-available premier hearing aids from leading manufacturers. The hearing aids were individually programmed to each participant's thresholds based on each manufacturer's proprietary fitting formula at default settings for all parameters. Participants listened to music clips (described in the next section) processed through two program settings within each hearing aid. The first program was the manufacturer's "first fit" (universal program) and the second program was the manufacturer's proprietary music program. Note that the fifth hearing aid did not offer a music program so recordings were generated using only its universal program. However, these recordings were duplicated to provide a balanced set of stimuli against the other four hearing aids. A summary of the hearing aids and programs can be found in Table 4-1.

Table 4-1: Summary of hearing aids and program settings used in this study.

Hearing Aid	Program 1	Program 2
Hearing Aid 1 (HA1)	Universal	Music
Hearing Aid 2 (HA2)	Universal	Music
Hearing Aid 3 (HA3)	Universal	Music
Hearing Aid 4 (HA4)	Universal	Music
Hearing Aid 5 (HA5)	Universal	Universal

4.2.3 Music genres and recordings

Pre-recorded music samples were selected from four different music genres: classical, jazz, folk, and pop. Samples were 15 to 30 seconds in length, comprising at least a full musical phrase. The classical sample included a full orchestra passage playing at a moderate-to-fast tempo. The pop sample consisted of a female vocalist, drums, electric guitar, piano, and bass guitar playing at a moderate tempo. The folk sample included acoustic guitar, drums, melodic percussion instruments, and bass guitar playing at a fast tempo. The jazz sample included electric guitar, upright bass, and snare drum with brush drumsticks playing at a slow pace. A fifth genre, individualized for each participant, was also included. For this genre, participants chose a favourite song, from which a 15-30 second sample was included.

In order to play the hearing aid processed samples without the listeners knowing which aid was which, we made recordings of each hearing aid and played them back using earphones. Each of the five samples was recorded through the individually-fitted hearing aids and hearing aid programs. This yielded a total of 10 recordings per participant. Hearing aids were fitted using double dome couplings and were recorded on a mannequin (Bruel & Kjaer Head & Torso 4128 C) placed in the center of a double-walled sound-attenuated booth. Music samples were delivered to the mannequin from an Anthony Gallo A'Diva loudspeaker at 0 degrees from 76 cm away. The presentation levels were 60, 72, 73, and

78 dB SPL for the jazz, classical, pop, and folk samples, stimuli, respectively, and 73 dB SPL for the favourite sample.

The hearing aid recordings were compared to high and poor quality samples when the listeners made their ratings (described below). The high quality samples, or "references" were the original samples digitally filtered to match DSL v5.0 adult targets. The poor quality samples, or "anchors" were highly distorted versions of the reference stimuli. Distortions were created by digitally center-clipping the music stimuli at 10% of the peak level. Altogether, there were a total of 70 stimuli per participant.

4.2.4 Ratings

A sound quality rating for each sample was obtained using the "multiple stimulus test with hidden reference and anchors" (MUSHRA) task (ITU-R, 2015). Ratings were made on a continuous scale ranging from "Bad=0" to "Excellent=100", as shown in Figure 4-3. Listeners were seated in a double-walled sound booth, in front of a laptop, wearing insert earphones with a broadband frequency response (Etymotic ER-2). Participants completed a practice run with three stimuli, and adjusted the volume to their most comfortable listening level before ratings began. The ratings were done in groups of seven stimuli at a time (the same sample processed by the five hearing aids plus the corresponding reference and anchor stimuli), and there were 10 of these groups, corresponding to the 10 combinations of genres and program types (universal or music). Presentation order of the 10 groups was randomized between participants. Rating all 10 groups took 30 to 60 minutes to complete.

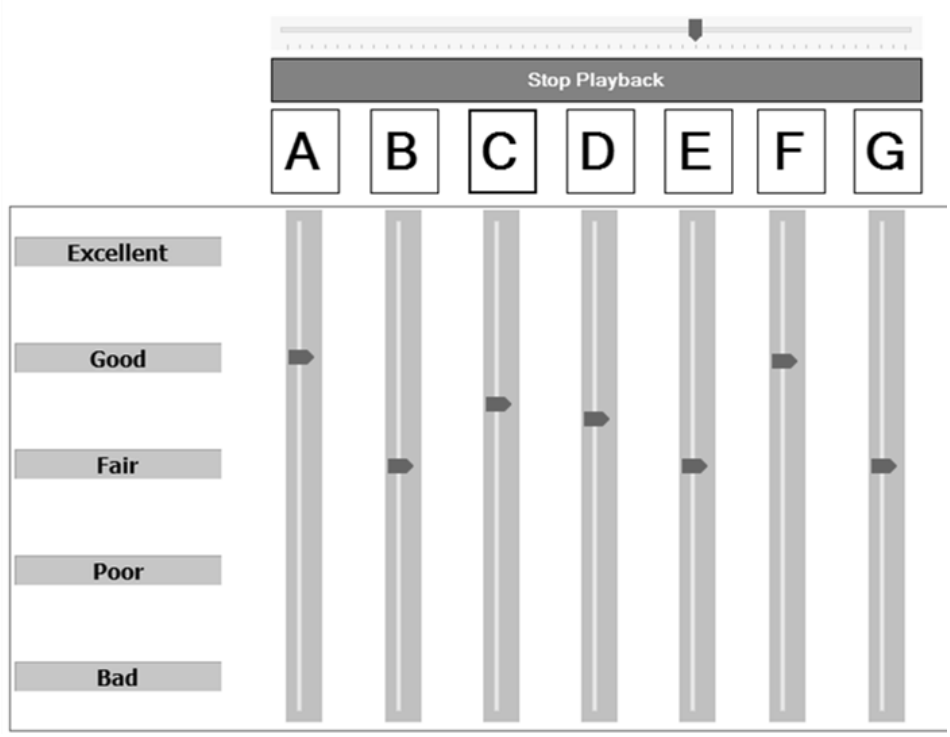

Figure 4-3: Screenshot of software used to gather the sound quality ratings of hearing aid processed music samples. Clicking on the lettered button played the processed sample randomly assigned to it. Participants adjusted the sliders to indicate their sound quality ratings.

4.3 Results

4.3.1 Analysis

The sound quality ratings were analyzed using a three-way repeated measures analysis of variance, using Greenhouse-Geisser corrections, with genre [5], hearing aid [5], and program [2] as within-subjects factors. Post-hoc analyses were completed using Bonferroni corrections when appropriate to locate significantly different contrasts. Sound quality rating data for universal and music programs are presented in Figures 4-4 and 4-5, respectively, and compared in Figure 4-6.

4.3.2 Effect of model and genre

Hearing aid model was a significant factor in sound quality ratings ($F_{(3.2,79.5)} = 19.7$, $p < 0.001$) collapsed across all programs and genres. HA1 was rated significantly higher than all other hearing aids ($p < 0.001$), except for HA2. HA4 was rated significantly poorer than all other hearing aids. Given that a major question of this study was the effect of music program on sound quality, further analysis was done to compare effects between hearing aids within the universal and music programs separately.

For either the universal programs (Figure 4-4) or the music programs (Figure 4-5), some hearing aids had better sound quality for music than others. Comparing specific pairs of hearing aids, we see that HA1 outperformed the other models, but not HA2, among the universal programs. Among the music programs, HA3 improved relative to its own universal program, and both HA1 and HA2 were better than HA4 and HA5. The better-performing models across universal and music programs had average sound quality ratings of about 75%, which roughly corresponds to a rating of "good".

Breaking this down by musical genre, we see that an interaction was also present between model, program type, and genre ($F_{(8.3,207.1)} = 2.0$, $p < 0.05$). This suggests that sound quality ratings differed significantly between hearing aids, and that these differences depended on whether the hearing aids were in a music program for at least some hearing aid models and music genres. The genre-specific ratings are shown in the right panels of Figure 4-4, for the universal programs, and Figure 4-5, for the music programs. HA1 was rated as having better sound quality than other hearing aids seven times in the universal program and eight times in the music program, which was more than other hearing aids. We also saw that there were more significantly different paired comparisons within the Jazz and Pop genres, suggesting that these genres were more sensitive to the differences between models.

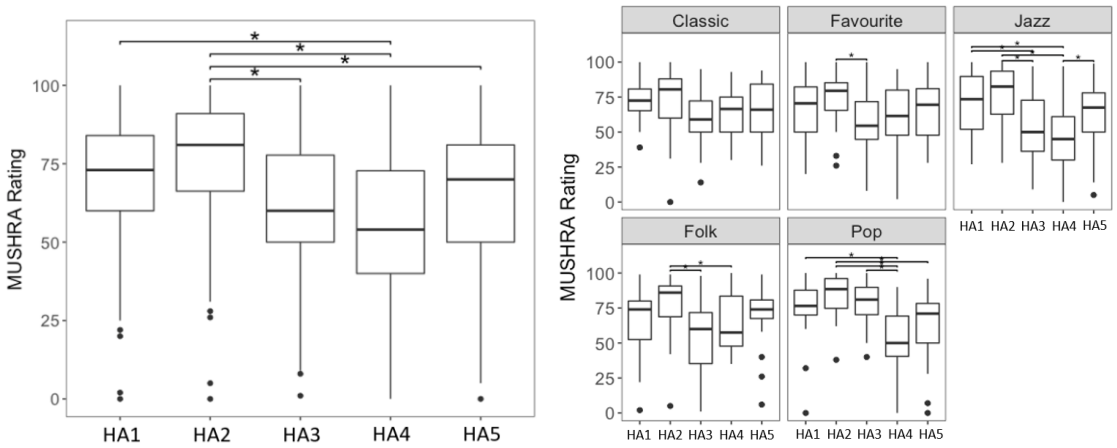

Figure 4-4: Boxplots showing the sound quality ratings for all hearing aids in the universal program across genres (on the left) and separated by genre (on the right). Asterisks indicate statistically significant differences between hearing aid models.

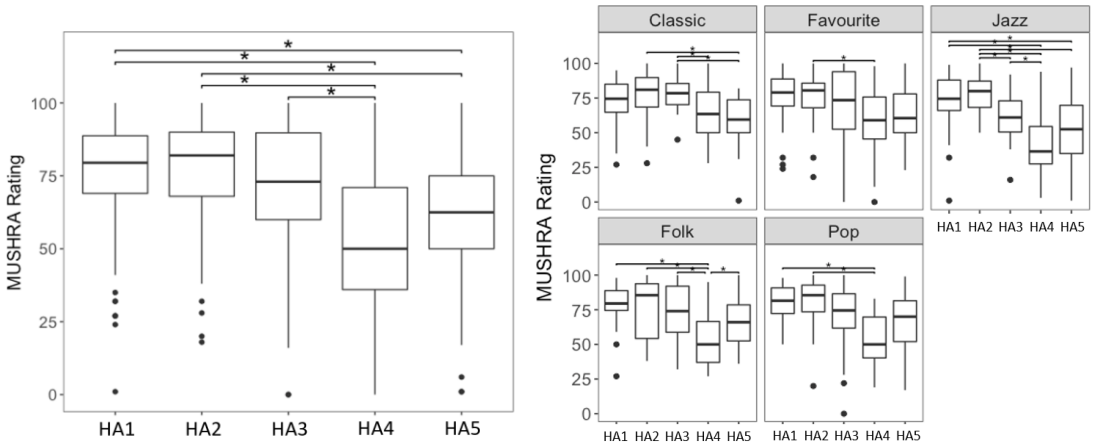

Figure 4-5: Boxplots showing the sound quality ratings for all hearing aids in the music program across genres (on the left) and separated by genre (on the right). Asterisks indicate statistically significant differences between hearing aids models.

4.3.3 Effect of music programs

Across genres, we saw a significant interaction of program by hearing aid ($F_{(4,100)} = 8.3$, $p < 0.001$). This suggests that music programs improved sound quality relative to the universal programs for some models more than others (Figure 4-6; left panels of Figures 4-4 and 4-5 combined). Hearing aid sound quality improved significantly for the music programs offered by HA2 and HA3. For all other hearing aids, the universal and music programs did not result in different sound quality ratings (including HA5, for which the universal and music program recordings were duplicates).

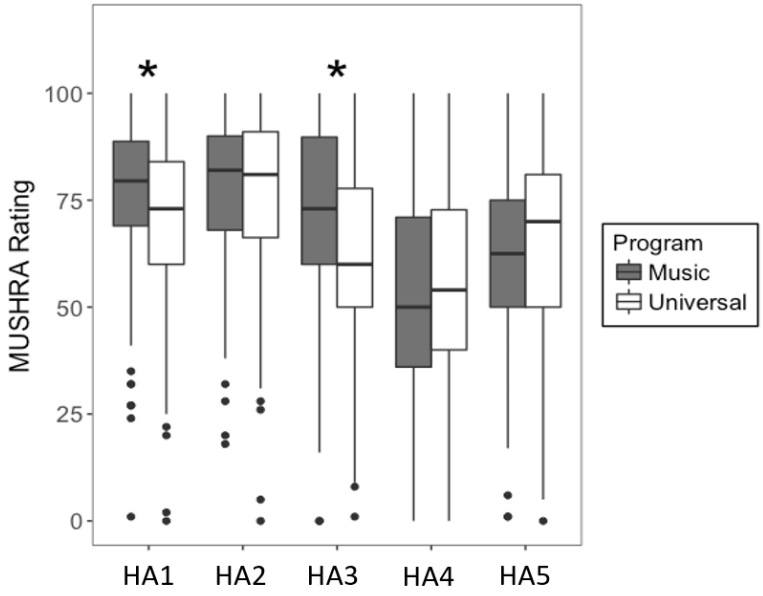

Figure 4-6: Boxplots showing the MUSHRA sound quality ratings for both programs for each hearing aid, across all genres. The dark grey boxplots represent the music program and the white boxplots represent the universal programs. Asterisks indicate statistically significant differences between programs within each hearing aid.

4.4 Discussion

In this study, hearing aid users rated the sound quality of music samples processed by several manufacturers' premier hearing aids, in both universal and music programs. All hearing aids were set to the manufacturer's default settings. Overall, this study revealed significant between-model differences in sound quality for music. HA1 was rated as having the overall best sound quality across programs and genres. In the universal program, HA1 was rated significantly higher than hearing aids HA3, HA4 and HA5. In the music program, HA1 was rated significantly higher than hearing aids HA4 and HA5.

4.4.1 Does music program matter?

Overall, the majority of music programs tested did not significantly improve music sound quality. These results are consistent with recent survey findings of Madsen & Moore (2014), in which perceptions of music did not differ between using versus not using a music program. This was possibly due to hearing aids having good sound quality in either program, leaving little room for improvement for some devices. In other devices (HA3) the music program was effective, with higher ratings in the music program than in the universal program. In other words, sound quality and music program strategies may vary across models, and some combinations of model and program may be more effective than others.

4.4.2 Does genre matter?

Our results show that sound quality differences between hearing aids may be more apparent for some genres of music compared to other genres. This finding is consistent with previous results from simulated hearing aids tested by Arehart et al., (2011). Clinically, this may mean that the most suitable hearing aid, and whether it needs or has a music program, may depend on the type of music that the patient would like to enjoy. In this study, HA1 was rated, in all genres, as having the highest sound quality more frequently than any other hearing aid. Some genres seemed to interact more often with hearing aid model and program. For example, there were only a few noticeable differences for classical music, while the jazz sample elicited many noticeable differences. The individual patient's preferences for music type is an important consideration.

4.4.3 Why did sound quality vary?

The sound quality of hearing aid processed music can be affected by electroacoustic parameters, including some that we can manipulate when fitting hearing aids. For example, some studies have examined bandwidth and music sound quality. Hearings aids typically amplify frequencies between 200 and 5000 Hz. However, Franks (1982) found that listeners preferred music that had additional gain in the low-frequency range below 200 Hz. Moore, Füllgrabe, & Stone (2011) and Ricketts, Dittberner, & Johnson (2008) also found that hearing-impaired listeners preferred music with additional gain in the extended high-frequency range, although listeners with steeply sloping audiograms may prefer a narrower bandwidth. Clinically, this may mean that hearing aids with a more robust bass response may produce better music sound quality.

Other studies have examined the effects of different compression settings on music sound quality. Compression compensates for elevated thresholds by amplifying low-level signals more so than high-level signals, thereby reducing the dynamic range of the hearing aid output. However, for music sound quality, linear settings are frequently preferred to more compressive settings (Arehart et al., 2011; Croghan, Arehart, & Kates, 2014; Kirchberger & Russo, 2016). If compression is used, longer time constants are frequently preferred to shorter time constants (Arehart et al., 2011; Croghan et al., 2014; Moore et al., 2011). Clinically, this may mean that slower-acting compression systems, or systems that are more linear, may produce better hearing aid sound quality.

Given these clear messages in the literature, we examined the results of the present study to determine if low-frequency and compression characteristics seemed to relate to the sound quality ratings we measured. Specifically, we selected one participant whose results were most representative of the group results of the study. This person's ratings can be found in Table 4-2 for the highest and lowest-rated hearing aids, and for the music program that differed the most from the corresponding universal program. In both of these contrasts, the listener rated the better and worse hearing aid conditions very differently, with 40 to 50% improvement for the hearing aid with the better sound quality. This is a substantial change and is likely of clinical importance.

Table 4-2: List of participant's ratings of 2 hearing aid recordings for the folk clip.

	Hearing Aid Comparison		Between-Program Comparison	
Hearing Aid Program	HA1 Universal	HA4 Universal	HA3 Music	HA3 Universal
Rating (%)	100	50	92	58

In order to learn what was different about the hearing aids and programs listed in Table 4-2, we analyzed the spectral and compression characteristics of hearing aid recordings of the folk music clip. This included the aided Long Term Average Spectrum (LTAS) of each hearing aid in 1/3rd octave bands from 100 Hz to 16000 Hz recorded in a 0.4 cc coupler. We compared the LTAS to the participant's 0.4-cc coupler-based thresholds for the same overall level (Figure 4-7). To measure compression, we calculated each hearing aid's short-term compression ratio by dividing the input dynamic range by the output dynamic range across frequencies (Figure 4-8).

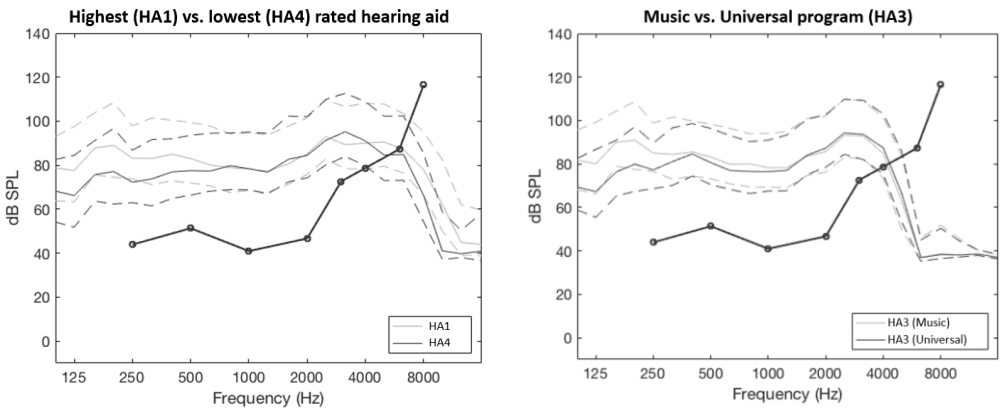

Figure 4-7: Frequency response curves of the highest and lowest rated hearing aids (left) and hearing aid with the biggest difference between programs (right). Solid lines represent the 0.4-cc coupler long term average spectrum of the output in response to the folk music sample, while dashed lines show the aided peaks and valleys of the signals. The distance between the dotted lines for a single colour is the dynamic range of the hearing aid output. The dark lines are pure tone detection thresholds, plotted in dB SPL in the 0.4-cc coupler.

The left panel in Figure 4-7 illustrates the hearing aid outputs for the best-rated hearing aid (HA1) versus the poorest-rated hearing aid (HA4). Clear differences between the shape of the aided frequency response of the two hearing aids can be noted. In the low frequencies (about 100 – 500 Hz), HA1 provided about 10 dB more output for music than did HA4. In the high frequencies (about 8000 – 10,000 Hz) HA1 exceeded HA4 by an average of 16 dB. Overall, the better-rated hearing aid had a broader bandwidth and provided a higher listening level for the bass frequencies of this music sample.

The right panel in Figure 4-7 illustrates the hearing aid outputs for the hearing aid with the largest differences between its universal and music programs (HA3). Similar to the best-worst comparison above, there are differences in low-frequency output. The music program exceeded the universal program by an average of 10 dB, particularly below about 300-400 Hz. Overall, the better-rated program had a higher listening level for the bass frequencies.

We also compared the effective short-term compression ratio for the music signal for the same two pairs of hearing aid recordings. This is not an input/output curve – recall that we

only tested the music response of each hearing aid at one overall level. Instead, we tested whether each hearing aid was compressing the ongoing dynamic range of the music passage. In Figure 4-8, a compression ratio of 1 means that the dynamic range of the aided output was the same as the dynamic range of the unaided input. This implies that the hearing aid applied linear amplification, preserving the loud versus soft changes in level of the music signal in the short term. Comparing the best and worst hearing aids in our sample (left panel, Figure 4-8) we see that both hearing aids provided signal processing that preserved the short-term dynamic range of music to about 6000 Hz. Above this, HA4 provided a higher compression ratio than HA1, nearing a 3:1 compression ratio above 8000 Hz. However, this is likely the result of the distance between the peaks and valleys, which is smaller above these frequencies (as seen in Figure 4-7). It increased the measured compression values, but these higher compression ratios are not likely attributable to the compression system of the hearing aids. This affected both hearing aids, but less so for HA1. A similar characteristic is observed for the universal program versus music program comparison in HA3, which is slightly compressive through most of the frequencies, and has high compression ratios above the roll-off of the frequency response.

Overall, the differences in short-term compression of all four of these hearing aid conditions are less compelling than the clear differences in bass response shown in Figure 4-7. These compression ratios reflect device performance for a stimulus presentation level of 78 dB SPL. Live music, or music across other input level may show different compression effects.

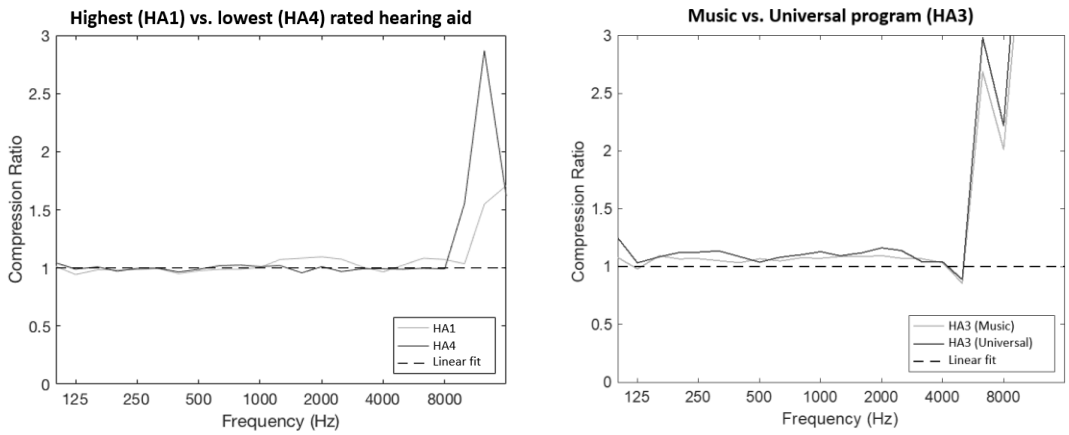

Figure 4-8: Short-term compression ratio as a function of frequency of the highest and lowest rated hearing aids (left) and hearing aid with the biggest difference between programs (right). Solids lines represent compression ratio of each hearing aid. The black dotted line is a reference linear fit.

4.4.4 Further thoughts

As with any study, there are limitations to what we did and how we did it. First, all hearing aid recordings in the current study were recorded using double dome couplings. These couplings significantly occlude the ear canal, resulting in minimal low-frequency leakage but also the potential for occlusion-related complaints that are not related to music. We used this highly occluded option because we wanted the listeners to evaluate the hearing aid response itself, not external signals. In real world fittings, using a fully occluding coupling method may help to retain the low-frequency response for listening to music streamed via a wireless accessory, which may be positive. However, the tradeoff between good low-frequency response for music sound quality versus other listening tasks is a consideration: vented fittings are used to support the sound quality of the wearer's own voice, to improve spatial hearing, and to help to maintain comfort, and are therefore a frequently-chosen option. In vented fittings, the low-frequency gain is often reduced. Readers may wish to consider that this may limit the sound quality for streamed music, although we did not study this directly in this experiment.

Second, we have only evaluated the relationship between sound quality and the hearing aids' frequency responses and compression ratios for one representative participant. This initial analysis certainly leads one to consider the importance of low-frequency response for music, and this is consistent with the literature (Franks, 1982; Tan & Moore, 2003). However, other signal processing features, such as noise reduction, feedback cancellation, frequency lowering, compression speed, distortion, coherence, noise floor, and other characteristics may have also affected sound quality. This short paper is not intended to be a comprehensive evaluation of these characteristics. The interested reader is referred to studies by Arehart et al., (2011), Chasin & Hockley (2014), Madsen & Moore (2014), and Moore (2016) for further information.

Finally, it is important to mention that different stimuli were played in sound field at different presentation levels. Different playback levels may have caused the hearing aids to behave differently. For instance, the hearing aid compressions systems would have been less active at soft levels (i.e. jazz stimulus) relative to loud levels where they would have been more active (i.e. folk stimulus). This may have resulted in different hearing aid preferences at one input level compared to another input level. The study design, in its current form, did not allow for the interpretation of whether differences between stimuli were due to the acoustic differences between genres or differences between their presentation levels. Future iterations of this study should present each genre at the same level to isolate effects of hearing aid, hearing aid program and/or genre alone or present each genre at multiple levels to determine if there is an effect of level and if it interacts with other variables.

4.5 Conclusion

Many people wear hearing aids for music listening (Leek et al., 2008; Madsen & Moore 2014), and premium sound quality is an important factor in satisfaction and outcome for hearing aid users (Abrams & Kim, 2015). However, hearing aid sound quality for music is not as well-understood as hearing aid sound processing for speech. The purpose of this study was to evaluate the sound quality of recorded music samples processed via five premier hearing aids set to a universal or music program. Hearing impaired listeners rated

the sound quality of the samples using a multiple comparison methodology. There were significant differences in quality ratings between hearing aids, with HA1 being rated higher than any other product, regardless of music genre or hearing aid program. A music program improved ratings for two of the five hearing aids tested, although the magnitude of improvement was less than the difference between a high- versus low- rated hearing aid. An analysis of one participant's recordings suggests that a mild gain increase in the low frequencies can enhance music sound quality. Overall, this study suggests that hearing aid selection and programming can be an important factor when music sound quality is a priority of the listener.

4.6 References

Abrams, H., & Kihm, J. (2015). An introduction to MarkeTrak IX: A new baseline for the hearing aid market. *Hearing Review, 22*(6),16.

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

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, W., Cox, R., … & Ludvigsen, C. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America. 96*(4), 2108-2120.

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

Cornelisse, L. E., Gagne, J., & Seewald, R. C. (1991). Ear level recordings of the long-term spectrum of speech. *Ear & Hearing, 12*(1), 47–54.

Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing, 35*(5), e170–e184.

Croghan, N. B. H., Swanberg, A., Anderson, M. C., & Arehart, K. H. (2016). Chosen listening levels for music with and without the use of hearing aids. *American Journal of Audiology*. *25*, 161-166.

Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology, 18*(8), 688–699.

Franks, J. R. (1982). Judgments of hearing aid processed music. *Ear and Hearing, 3*(1), 18–23.

Galster, J., Rodemerk, K., & Fitz, K. (2014, August). *Preferred aided listening levels for music in the sound field.* Poster presented at the 2014 International Hearing Aid Research conference, Tahoe City, CA. Retrieved from https://starkeypro.com/pdfs/research-briefs/Preferred_Aided_Listening_Levels_for_Music_in_the_Sound_Field.pdf.

Hille, K., Gust, K., Bitz, U., & Kammer, T. (2011). Associations between music education, intelligence, and spelling ability in elementary school. *Advances in Cognitive Psychology, 7*, 1–6.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology, 49*(12), 891–903.

International Telecommunications Union (2015). *Method for subjective assessment of intermediate quality level of audio systems*. *Recommendation ITU-R BS.1534-3* Geneva, Switzerland.

Kirchberger, M., & Russo, F. A. (2016). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing, 20*, 1–16.

Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired Listeners. *Journal of the American Academy of Audiology, 19*(6), 519–526.

Limb, C. (2011, October). *Building the musical muscle* [Video file]. Retrieved from

https://www.ted.com/talks/charles_limb_building_the_musical_muscle

Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Amplification*, *0*(0), 1–29.

Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *Journal of the Audio Engineering Society*, *64*(3), 112–123.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, *32*(5), 556–568.

Moore, B. C. H., & Tan, C-T. (2003). Perceived naturalness of spectrally distorted speech and music. *Journal of the Acoustical Society of America, 114*(1). 408-419.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, *7*, 1–5.

Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*, 160–172.

Simmons-Stern, N. R., Budson, A. E., & Ally, B. a. (2010). Music as a memory enhancer in patients with Alzheimer's disease. *Neuropsychologia*, *48*(10), 3164–3167.

Chapter 5

5 The relationship between electroacoustic characteristics of default hearing aid music programs and perceived music sound quality

This study explored the electroacoustic behavior of hearing aids and music programs and determined whether those electroacoustic differences were related to sound quality judgments. We previously obtained sound quality ratings of pop and classical music (Chapter 4) amplified through universal and music programs of different hearing aids. The following parameters were measured for each hearing aid: average decibel sound pressure level across low-bass (100-200 Hz), bass (250-800 Hz), midrange (1000 Hz- 2500 Hz), and treble (3000-10000 Hz) frequency bands, and envelope distortion due to compression. Recordings and sound quality ratings were gathered from 24 hearing aid users from the previous study (Chapter 4). Music programs exhibited higher levels in the low-bass and bass bands and slightly lower levels in the treble band compared to universal programs. Differences between programs were less than differences between the highest- and lowest-rated hearing aids. A linear mixed effects model revealed that greater low-bass band levels were associated with favorable sound quality. Compression varied between hearing aids and programs but was not predictive of sound quality. Except at low-bass frequencies below 250 Hz, hearing aids and music programs at default settings, on average, do not meaningfully vary output levels and compression for music sound quality.

5.1 Introduction

Hearing aids have traditionally been designed for the acoustics of speech, which differ from the acoustics of music (Chasin & Hockley, 2014). This can be especially problematic for the musically inclined hearing aid user, as they may not be as satisfied with amplified music as they are with amplified speech. In a survey of music enjoyment in hearing-impaired listeners, Leek, Molis, Kubli, & Tufts (2008) found that up to 30% of their respondents who wanted to listen to music found it unsatisfying. In addition, Madsen & Moore (2014) conducted a survey on music-listening with hearing aids. They found that many listeners found music to be enjoyable, but also complained of many hearing aid-related sound

quality distortions. Likewise, Vaisberg et al. (2018) reported that hearing aid-wearing instrumentalists were dissatisfied with live music sound quality, particularly noting challenges with dynamic contrasts and melody-following. This evidence highlights the need for hearing aids to incorporate a solution for music-listening. In recent years, most hearing aid manufacturers have recognized this need and many now offer music-listening settings to help optimize the music-listening experience.

Hearing aid music-listening settings can be programmed into a hearing aid within a unique music program. The program itself is a set of signal processing adjustments designed to optimize processing for musical inputs and can be activated using a button on the hearing aid or a remote control, or may be activated automatically by the classification-and-switching functions of the hearing aid. These aspects of hearing aid signal processing are typically developed per manufacturer and may be proprietary in nature. This makes it challenging to study the efficacy of a hearing aid music programs in general.

Relatively few studies have examined the efficacy of hearing aid music programs for music-listening. The general consensus of the limited studies available seems to be that music programs are no more effective than a standard speech program for music-listening. Fulford, Ginsborg, & Greasley (2015) conducted semi-structured interviews with hearing-impaired musicians and reported the listeners' experiences using hearing aids. Many of the participants were unsure of whether they had a music program, and those who did have a music program did not use it consistently. The survey by Madsen & Moore (2014) revealed music satisfaction scores that were no different between users and non-users of a music program. In another study (Bradford, 2014), hearing-impaired listeners were required to indicate their preference between hearing-aid amplified music clips through either a standard program or a music program. The results indicated no difference between the two. Lastly, Chapter 4/Vaisberg et al. (2017) compared sound quality preferences for music between the default settings for both universal speech programs and for music programs for five different hearing aids for a variety of different musical genres. Four of the hearing aids included a manufacturer-programmed music setting. The fifth hearing aid's music program was customized by the researchers for the purposes of the study based on manufacturer recommendations. The music sound quality ratings in Chapter 4/Vaisberg et

al. (2017) were obtained using the method of “MUltiple Stimuli with Hidden References and Anchors” (MUSHRA, ITU-R, 2015). Averaged across all stimuli and programs, all hearing aids were rated as having an average 50% (fair) sound quality or greater. One hearing aid was rated higher more frequently than any others whereas another hearing aid was rated lower more frequently than any others. Only two of the five hearing aids’ music programs improved sound quality ratings relative to their respective universal programs, although the magnitude of improvement was less than the variation observed across hearing aids. Together, these results indicate that music programs were not consistently successful in improving sound quality for music-listening across brands, although some brand-specific programs were effective. These studies also suggest that the electroacoustic differences between the universal and music programs were either not appropriate or too small to elicit a meaningful improvement for listeners.

Fortunately, studies have investigated the impact of individual hearing aid signal processing systems on the perception of music sound quality, which may lend insight into these results. For example, several studies have investigated the effects of hearing aid bandwidth limitations on the sound quality of music. Hearing aid bandwidth can be widened by increasing the relative gain at the extremities of the bandwidth of the device. Moore & Tan (2003) found that optimal music sound quality was associated with a bandwidth of 50 Hz to almost 16 kHz in normal-hearing listeners. However, modern hearing aids have high-frequency cutoffs that vary between 3.6 and 8 kHz (Kimlinger, McCreery, & Lewis, 2015), with some open-fit devices showing usable bandwidth as narrow as 890 Hz to 4.4 kHz (Struck & Prusick, 2017). Ricketts, Dittberner, & Johnson (2008) investigated whether extending the upper bandwidth limit from 5.5 kHz to 9 kHz was associated with an increased preference for music, and whether this was also related to degree or configuration of hearing impairment. Results revealed that listeners who had shallow audiogram slopes (i.e., less severe high-frequency thresholds) had greater preferences for the higher cutoff, whereas those who had steeper audiogram slopes (i.e., more severe high-frequency thresholds) preferred the lower cutoff. Similarly, Moore, Füllgrabe, & Stone (2011) presented hearing-impaired listeners with music samples with increasing cutoffs of 5 kHz, 7.5 kHz, and 10 kHz. Preferences for higher cutoffs were again associated with shallower audiometric slopes. Fewer studies have investigated the impact

of manipulating the lower bandwidth limit. Franks (1982) decreased the lower bandwidth limit by increasing the overall bass response for music stimuli for judgment by hearing-impaired listeners. He found that higher preferences were associated with a cutoff well below the 200 Hz limit often observed in hearing aids. Together, these studies indicate that the bandwidth of a music program may be a predictor of sound quality, with a preference for more bass, and more treble for some listeners.

Hearing aid multichannel amplitude compression is another signal processing system which is known to affect the sound quality of music. Compression compensates for the listener's reduced dynamic range and elevation of detection thresholds by providing more amplification for soft sounds compared to loud sounds, but reduces the output dynamic range of the signal in doing so (Souza, 2002) thus creating envelope distortion. Compression can be manipulated using several parameters: First, the compression ratio manipulations affect the degree to which amplification changes as the input level fluctuates. Second, time constant manipulations affect the amount of time required for the gain to change in response to input level fluctuations. While additional parameters are involved in hearing aid compression systems, these two are the ones that have been more frequently manipulated in sound quality research. In general, hearing-impaired listeners prefer linear or linear-like settings compared to compression systems with a higher compression ratio for a variety of music genres (Arehart, Kates, & Anderson, 2011; Croghan, Arehart, & Kates, 2014; Kirchberger & Russo, 2016a). Furthermore, hearing-impaired listeners typically prefer longer time constants over shorter time constants (Arehart et al., 2011; Hansen, 2002; Moore et al., 2011). For a summary of optimal compression systems for music-listening, the reader is referred to Table 1 in Kirchberger & Russo (2016a).

Advanced hearing aid signal processing systems could also potentially affect the sound quality of music. One such system is frequency lowering, which shifts high-frequency content of a signal to an audible frequency range and has been developed to provide access to high-frequency speech cues. However, frequency lowering may introduce inharmonicity, which may give the sense that a musical instrument is out of tune. In theory, this effect is dependent upon the type and strength of frequency lowering applied (see

review by Alexander, 2013). In studies, hearing-impaired listeners have been less sensitive to moderate amounts of frequency lowering compared to listeners with normal hearing (Kirchberger & Russo, 2016b; Mussoi & Bentler, 2015; Parsa, Scollie, Glista, & Seelisch, 2013). Other advanced signal processing mechanisms may include directionality, automatic noise reduction, or feedback cancellation. Moore (2016) recommends minimizing or disabling advanced systems such as these, including feedback cancellation and noise reduction, in a hearing aid music program.

5.1.1 Summary and purpose statement

The review above summarizes hearing aid signal processing adjustments that may affect music sound quality. Moore (2016) suggests that slow compression systems and an extended low-frequency response are common in hearing aid music programs, both of which are consistent with the presented evidence. Our previous study (Chapter 4/Vaisberg et al., 2017) indicated that variation in music sound quality remains, even when dedicated music programs are used. However, the results reported previously were presented on the basis of perceptual differences across programs, which does not provide insight into whether there were any objective differences among the devices that varied in sound quality for music.

The purpose of this study was, therefore, to explore the relationship between electroacoustic characteristics (band-specific levels and compression characteristics) of the programs and hearing aids and their associated sound quality ratings. To explore these objectives, I analyzed the hearing aid recordings that were made during our previous study and compared them to their associated sound quality ratings (Vaisberg et al., 2017). That is, I observed the band-specific levels and compression characteristics of the hearing aids and programs using measured differences between the electroacoustic characteristics of the previously-generated recordings. Advanced signal processing mechanisms (such as frequency lowering and automatic noise reduction) may have also affected sound quality ratings. However, these processes are proprietary, vary between manufacturers, and are typically applied in conjunction with other signal processing mechanisms. Therefore, I focused on the effects of band-specific levels and compression characteristics alone, because these parameters represent fundamental gain and shaping processes that are

applied in any hearing aid. I hypothesized that the music programs and preferred hearing aids would be more likely to apply strategies consistent with factors supporting good sound quality in the literature, i.e.:

1) A hearing aid music program will exhibit more low-frequency output relative to its universal program
2) A hearing aid music program will exhibit less envelope distortion (i.e. longer time constants and smaller compression ratios) relative to its universal program
3) Additional low-frequency output will be associated with improved sound quality ratings
4) Less envelope distortion (i.e. longer time constants and smaller compression ratios) will be associated with improved sound quality ratings.

5.2 Methods

5.2.1 Participants

Details of the participants and hearing aid fittings have been reported previously (Chapter 4; Vaisberg et al., 2017). Briefly, sound quality ratings were gathered from 24 adults from that study (ages 20 – 79; average = 69.9 years), who were hearing aid users with bilateral symmetrical average mild-sloping-to-severe sensorineural hearing loss. The study was approved by the Western University Health Sciences Research Ethics Board.

5.2.2 Materials

The classical and pop genres from the previous study were selected for this analysis as the two (a) varied in style, and (b) had similar overall presentation levels. The classical sample was provided by Sivantos and included a chamber string section playing at a moderate-to-fast tempo. The pop sample, Linda's Rondstadt's "You're No Good" was downloaded from iTunes and consisted of a female vocalist, drums, electric guitar, and bass guitar playing at a moderate tempo. The classical and pop samples were 14 and 25 seconds in length, respectively. The samples were analyzed using the IEC 60118-15 (2008) recommendation, which was developed to characterize various spectral properties of hearing aid signal processing. As per the recommendation, waveforms were analyzed using 125 ms Hann windows with 50% overlap. A discrete Fourier transformation was performed on each window, and windows were subsequently divided into 1/3rd octave bands from 100 Hz to 10 kHz. Long-term average spectra using this recommendation and waveforms of the stimuli can be found in Figure 5-1.

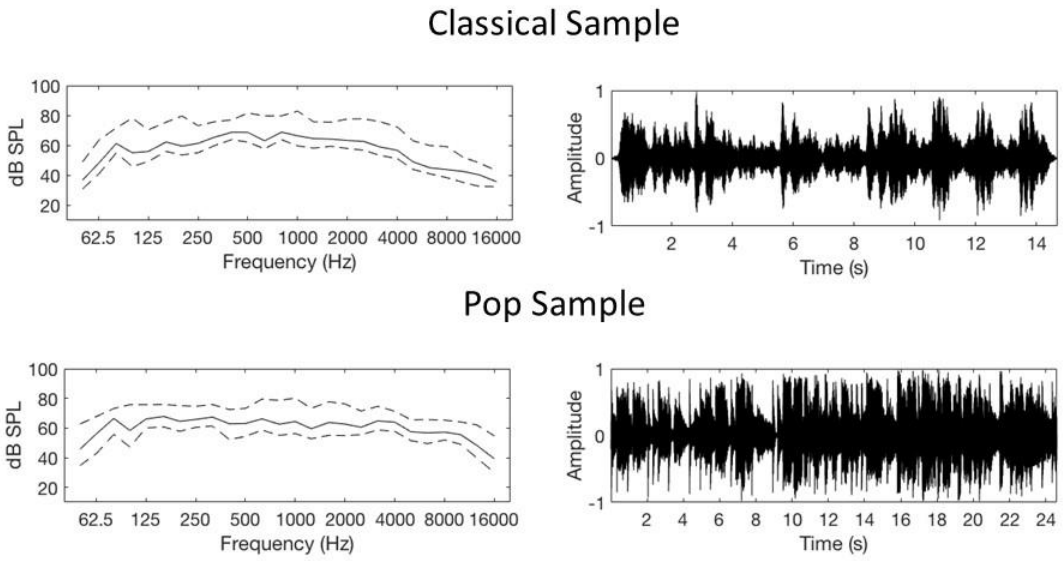

Figure 5-1: Long-term average spectra (left column) and normalized waveforms (right column) for the test stimuli, analyzed using the IEC 60118-15 (2008) recommendation. In the spectra, the top and bottom dashed curves represent the 99th and 30th percentiles, respectively and the solid curve represents the 65th percentile.

The music samples were presented in sound field. Stimuli were processed using 16-bit precision and generated via an Echo AudioFire soundcard using a sampling rate of 48 kHz and were delivered monophonically via an Anthony Gallo A'Diva loudspeaker to five hearing aids in both music and universal programs. The presentation levels were 72 and 73 dB SPL for the classical, and pop music samples, respectively, which is similar to average preferred hearing aid input levels for music listening (Croghan, Swanberg, Anderson, & Arehart, 2016). The hearing aids were programmed per ear using each manufacturer's proprietary fitting formula. Stimuli were presented to the hearing aids at a 0 degree, ear-level distance of 76 cm. The hearing aids were mounted on a Bruel & Kjaer Head & Torso 4128C simulator in a double walled sound-isolated chamber. They were coupled to IEC 60318-4 ear simulators using double dome couplings on the head and torso simulator. The bilateral microphone outputs from the Bruel & Kjaer Head & Torso 4128C simulator were conditioned using a Bruel & Kjaer 2-Channel Nexus amplifier (Type 2690-A). The outputs yielded recordings that were used for bilateral music sound quality ratings in our previous study.

5.2.3 Stimulus processing

Electroacoustic measurements were made on all the recordings for the four hearing aids that offered a proprietary music program via their software defaults. Measurements were averaged between the left and right ears yielding 384 stimuli (24 participants x 2 genres x 4 hearing aids x 2 programs). The waveforms or spectra of the signals were required for the measurements, depending on the measurement of interest. All recordings were analyzed in MATLAB (Mathworks, R2017b). To gather spectra, waveforms were analyzed using the IEC 60118-15 (2008) recommendation as described in the materials section. The stimulus spectra were then subjected to several data reduction analyses, as follows.

5.2.4 Data reduction

5.2.4.1 Band-specific levels

This measurement computed average levels for each hearing aid and for each participant, measured in ear simulator-referenced dB SPL. Levels were grouped into four frequency

bands: low-bass frequencies (100-200 Hz), bass frequencies (250-800 Hz), midrange frequencies (1000-2500 Hz), and treble frequencies (3000 – 10000 Hz). These frequency bands were selected for several reasons. First, they approximate the bass, low midrange, midrange and upper midrange/presence/brilliance regions using in audio equalization in the music production industry ("Audio Spectrum Explained," 2018). Second, the bands are of similar size to frequency bands used in a 3-channel hearing configuration that was associated with preferred music quality (Croghan, Arehart, & Kates , 2014). In that study, the authors found that a 3-channel hearing aid was generally preferred over an 18-channel hearing aid for two music genres. In the 3-channel hearing aid, the range of the first, second, and third bands were 0-748 Hz, 929-1129 Hz, and 2860-11025 Hz, respectively. I decided to split the first band into a bass and a low-bass band. I did this because hearing aid fitting and verification is commonly performed to 250 Hz, with less consistent measurement of hearing aid gain or output below this frequency. Any considerable differences between hearing aids in the range below 250 Hz may therefore provide new information that is not routinely considered in current clinical fitting strategies. Galster, Rodemerk, & Fitz (2014) also grouped preferred aided listening levels into four frequency ranges: low (250, 500 Hz), mid (750, 1000, 1500 Hz), high-mid (2000, 3000, 4000 Hz), and high (6000, 8000 Hz). Our bass and mid range bands are approximately similar, and our high range combines the high-mid and high range reported by Galster, et al. (2014).

5.2.4.2 Cepstral correlation

I was also interested in exploring the compressive characteristics of the hearing aids used in the previous study and in assessing whether or not differences between compression schemes influenced behavioural judgments of perceived music sound quality. Hearing aid compression schemes have the potential to create non-linearities by manipulating the temporal envelope of a waveform, particularly if time constants are fast and/or compression ratios are high (Souza, 2002). Kirchberger & Russo (2016a) found that linear or the least compressive settings with the longest time constants were the most associated with higher music sound quality scores. The cepstral correlation (CC) is a measurement that is sensitive to envelope non-linearities, and is used as part of predictive hearing aid sound quality metrics for speech (Kates & Arehart, 2014) and music (Kates & Arehart, 2016).

The CC quantifies temporal differences between the envelopes of two auditory signals: a distorted test signal and a high-quality reference version. A CC value closer to one indicates relatively similar waveforms, suggesting that minimal envelope distortion has occurred. In the hearing aid audio quality index (HAAQI; Kates & Arehart, 2016), the CC is measured between the hearing aid output of interest (distorted test signal) and a digital version of the hearing aid input that has been equalized to a prescriptive formula (high-quality reference version).

In the current study, I used HAAQI's implementation of the CC to compare the changes in amplitude envelope between recordings of the hearing aid output and reference signals. To generate the reference signals, the digital versions of the stimuli were shaped using DSL v5.0 gains (Scollie et al., 2005) at 1/3rd octave bands for the same input level as the corresponding recordings using a frequency sampling-based finite impulse response filter with MATLAB's FIR2 function. The CC was computed between the test signal and digital versions in MATLAB using HAAQI source code described in Kates & Arehart (2016). The reader is referred to that paper for a detailed review of the CC calculation. A brief explanation of the CC calculation follows: An auditory model processed both the test and reference signals, producing envelopes sampled at 250 Hz that were converted to dB referenced to thresholds across 32 auditory frequency bands. The envelopes were then filtered using an 8-channel modulation filterbank from 0-125 Hz. For each filterbank output, the time-frequency characteristics of the test signal was compared to that of the reference signal using a normalized cross-covariance function. The final CC was the average of the cross-covariances for the four highest modulation frequency bands from 20-125 Hz. I expected that hearing aids with longer time constants and smaller compression ratios would elicit higher CC values and therefore less envelope distortion.

5.2.5 Analyses

Statistical analyses were completed using the R software package (Version 1.0.132; R Core Team, 2017). Repeated measures analysis of variance (RM-ANOVA) was implemented using the ez R package (Lawrence, 2016) to examine the differences between hearing aids and music programs for each of the measurements described above. Greenhouse-Geisser corrections were applied to protect against departures from sphericity (Gray & Kinnear,

1999). Post-hoc contrasts using the Holm correction were applied when appropriate for main effects of hearing aid and music program, simple main effects of hearing aid and music program within levels of genre, and simple main effects of music program within levels of hearing aid and genre (for a total of 29 possible family-wise contrasts). A sequence of linear mixed effects models was used to determine whether the electroacoustic measurements were predictive of subjective sound quality ratings. The models were implemented using the lme4 R package (Bates, Mächler, Bolker, & Walker, 2015). Sound quality ratings were included as the dependent variable, and sound pressure level at each frequency band (low-bass, bass, midrange, and treble) and CC were included as fixed effects predictor variables. A random intercept was included for each subject to account for the within-subjects variance attributed to the repeated-measures design of the original experiment, and for genre to account for between-stimulus variance. The assumption of normally-distributed residuals was assessed at the output of each model iteration by visual inspection (Field, Miles, & Field, 2012, page 870), with violations being corrected for in subsequent iterations. The overall fit of each subsequent model was assessed by comparing it to the most previous model in the sequence with an ANOVA, using the Akaike Information Criterion (AIC) and chi-square test to compare the performance of the models. The AIC measures relative goodness-of-fit between models, with the best-fitting model having the lowest value (Burnham & Anderson, 2004). The chi-square test indicates whether subsequent models are significantly different from one another using the p-statistic.

5.3 Results

5.3.1 RM-ANOVA results

The subjective sound quality results reported by Vaisberg et al. (2017) for four hearing aids are reanalyzed here across only pop and classical genres selected for the current study. The hearing aids were labelled as hearing aid 1 through 4 (HA1, HA2, etc.), in order from most-preferred to least-preferred. The average scores (and standard errors) for the universal and music programs of each hearing aid are illustrated in Figure 5-2. A RM-ANOVA with hearing aid (4 levels), program (2 levels) and genre (2 levels) revealed a significant main effect of hearing aid ($F_{(2.53,58.21)}= 17.43$, $p < 0.0001$, $\eta^2 = 0.16$), signification interaction of

hearing aid by genre ($F_{(2.61,59.99)}$= 4.69, $p < 0.01$, $\eta^2 = 0.04$), and significant interaction of hearing aid by program by genre ($F_{(2.42,55.66)}$= 4.38, $p < 0.01$, $\eta^2 = 0.01$). Across genres, hearing aid 4 was rated lower than all other hearing aids ($p < 0.0001$). Descriptively, HA1 was rated highest, followed by HA2, HA3 and finally by HA4. This pattern of preferences was consistent for the pop genre ($p < 0.0001$). For the classical genre, the difference between HA1 and HA4 approached significance ($p = 0.056$). No other contrasts were significant. The HA3 music program was rated higher than its universal program for the classical genre only ($p < 0.01$).

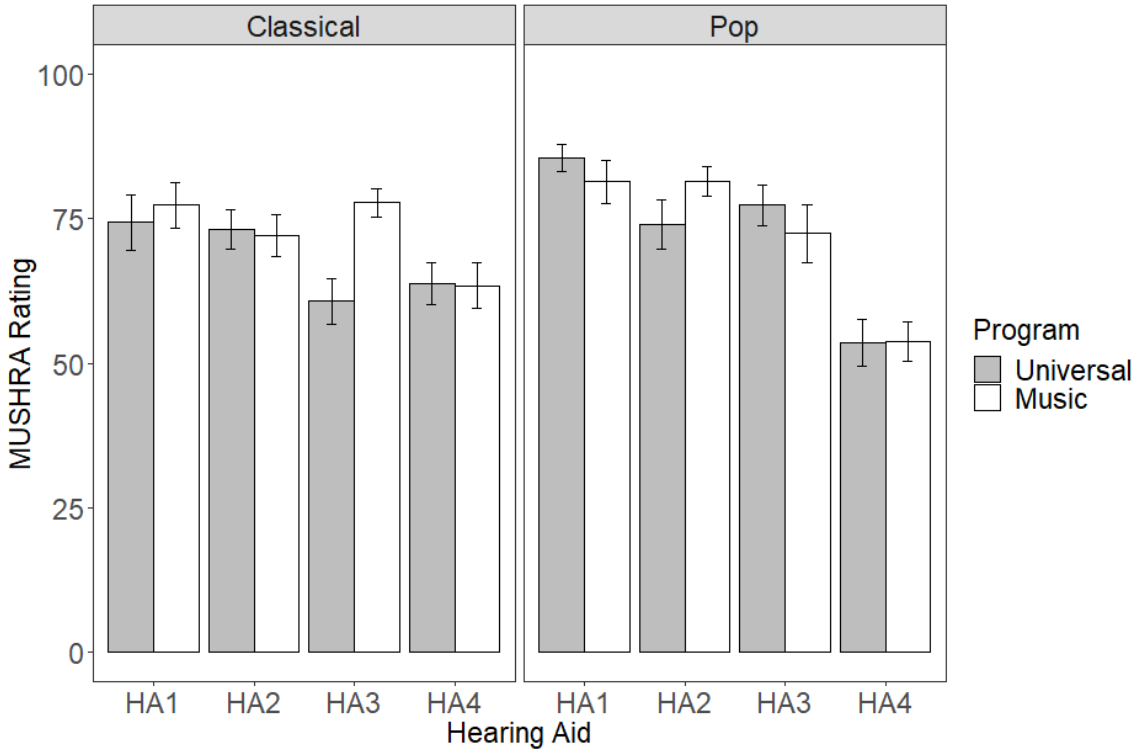

Figure 5-2: Average sound quality rating across genres and participants (as rated using the multiple stimuli with hidden references and anchors (MUSHRA) method) for each hearing aid for the classical (left) and pop (right) genres. The universal and music program bars are colored in gray and white, respectively. Error bars represent one standard error of the mean.

Average levels, reported in ear-simulator referenced dB SPL, for the hearing aids across participants, are illustrated in Figure 5-3 for the classical genre and pop genres. The levels are shown for each frequency band (low-bass: 100-200 Hz; bass: 250-800 Hz; midrange: 1000-2500 Hz; and treble: 3000-10000 Hz). An RM-ANOVA with hearing aid (4 levels), program (2 levels), genre (2 levels) and frequency band (4 levels) as within-subjects variables revealed significant effects for every possible main effect and interaction. Statistical results are reported in Table 5-1. Descriptively, music programs yielded higher low-bass frequency levels versus the universal programs for all four hearing aids. In the low-bass frequency band and averaged across programs, the level was greatest for HA1 and decreased in descending in order from HA2 to HA3 to HA4. In the bass band, the levels followed a similar pattern compared to the low-bass band, although the level difference between the highest-level and lowest-level hearing aids was smaller in the bass band compared to the low-bass band. In the midrange band, levels were greater for the music programs versus the universal programs for HA1, HA2 and HA3, although the differences were 1.5 dB or less. The midrange level for the HA4 music program was lower than the level in the universal program. In the treble band, the average levels were lower for the music programs versus the universal programs for HA1, HA3 and HA4, whereas the level was higher in the music program for HA2.

Average CCs for the hearing aids, across genres and participants, are displayed in Figure 5-4. A RM-ANOVA with hearing aid (4 levels), program (2 levels) and genre (2 levels) as within-subjects variables revealed significant effects for every possible main effect and interaction Statistical results are reported in Table 5-1. Descriptively, HA3 yielded the highest average CC, which was followed by HA2 and HA4 about equally, and finally by HA1 for the classical genre. For the pop genre, HA3 yielded the highest average CC, which was followed by all other hearing aids about equally. HA4's music program yielded a higher CC than its universal program for both genres. The difference in CC between the highest and lowest averages across genres was 0.041.

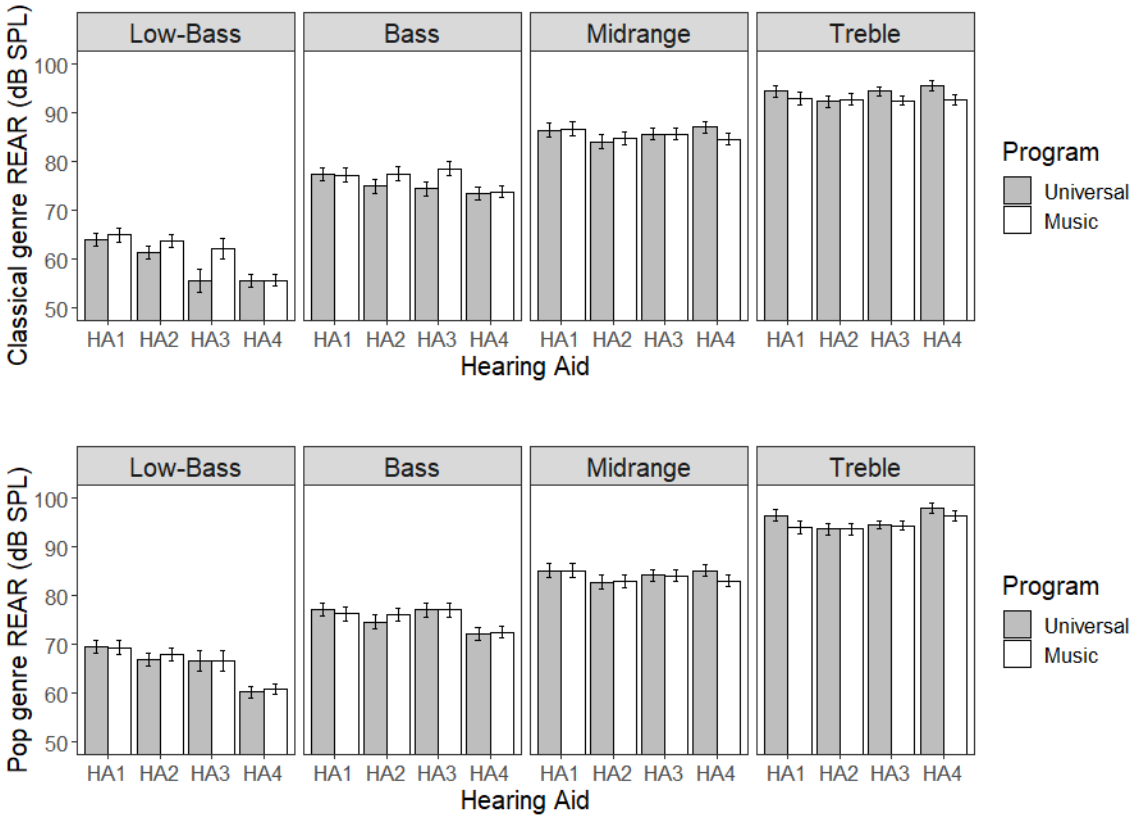

Figure 5-3: Average ear-simulator real ear aided response levels (REAR) in dB SPL, across participants, for each frequency band (low-bass, 100-200 Hz; bass, 250-800 Hz; midrange, 1000-2500 Hz; and treble, 3000-10000 Hz) for the universal and music programs of each hearing aid. Measurements for classical and pop genres are on top and bottom, respectively. The universal and music program bars are colored in gray and white, respectively. Error bars represent one standard error of the mean.

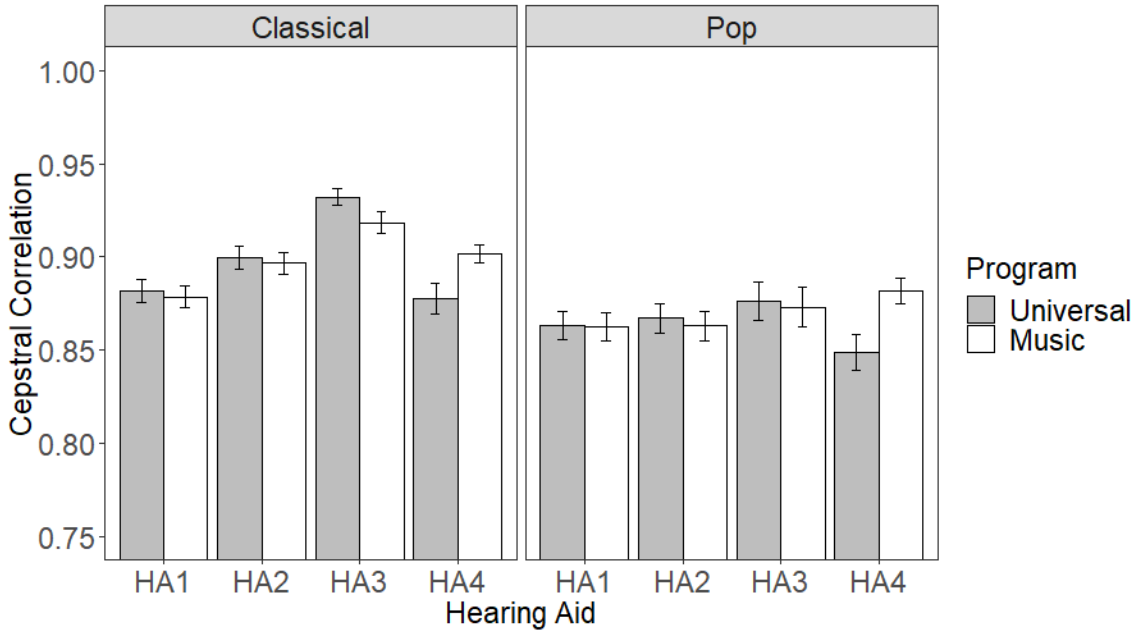

Figure 5-4: Average cepstral correlation (CC) across participants for each hearing aid for the classical (left) and pop (right) genres. The universal and music program bars are colored in gray and white, respectively. Error bars represent one standard error of the mean.

Table 5-1: Statistical results for repeated-measures analyes of variance for dependent measures of frequency-band specific levels (in dB SPL) and cepstral correlations. DV = depedent variable, IV = independent variables, F = f value, df = degrees of freedom, dfe = error degrees of freedom, p = p-value, η^2 = effect size.

DV	IV	F	df	dfe	p	η^2
Frequency-band level (dB SPL)	Hearing aid	28.3	2.21	50.8	< 0.0001	0.03
	Program	18.6	1	23	< 0.001	0.0002
	Genre	349.3	1	23	<0.0001	0.01
	Frequency band	259.0	1.24	28.6	< 0.0001	0.76
	Hearing aid:program	96.5	2.43	55.9	< 0.0001	0.005
	Hearing aid:genre	29.1	2.11	48.6	< 0.0001	0.0009
	Program:genre	166.3	1	23	< 0.0001	0.0009
	Hearing aid:frequency band	37.3	2.09	48.1	< 0.0001	0.06
	Program:frequency band	509.3	1.6	36.8	< 0.0001	0.007
	Genre:frequency band	1412.5	2.14	49.2	< 0.0001	0.04
	Hearing aid:program: genre	71.1	2.7	62.1	< 0.0001	0.001
	Hearing aid:program: frequency band	58.6	2.5	56.4	< 0.0001	0.002
	Hearing aid:genre: frequency band	91.9	3.9	89.7	< 0.0001	0.003
	Program:genre:frequency band	403.6	1.3	29.8	< 0.0001	0.001
	Hearing aid:program:genre: frequency band	227.4	2.2	51.4	< 0.0001	0.003
Cepstral Correlation	Hearing aid	17.1	2.05	47.1	< 0.0001	0.08
	Program	10.2	1	23	< 0.01	0.002
	Genre	63.7	1	23	< 0.0001	0.16
	Hearing aid:program	48.3	1.59	36.5	< 0.0001	0.04
	Hearing aid:genre	44.0	2.3	51.9	< 0.0001	0.03
	Program:genre	48.3	1	23	< 0.0001	0.001
	Hearing aid:program: genre	8.2	2.5	58.1	< 0.001	0.001

5.3.2 Model results

The sequence of linear mixed model iterations used to determine whether the descriptive statistics above were predictive of subjective sound quality is displayed in Table 5-2. An initial simple model incorporating only the per-participant random effects represented the null hypothesis and served as the baseline model for subsequent model comparisons. A random effects per-genre variable was also included to account for between-genre variance. Analysis of the residuals of the initial model output revealed a normal distribution with moderate skewness. In order to reduce the skewness, the sound quality scores in the subsequent model were transformed using a reflect-and-square root transformation (Tabachnik & Fidell, 1996), which improved the quality of the model. Next, all the predictor measurements (levels at all four frequency bands: low-bass, bass, midrange, and treble), and CC were entered into the model as fixed effects, which significantly improved the model relative to the per-subject and per-genre null hypothesis. This model revealed that the low-bass frequency band levels fixed-effects variable was a significant predictor of sound quality judgments. However, this model also revealed moderate multicollinearity between two sets of fixed-effects predictor variables: low-bass by bass frequency band level ($r = 0.84$) and midrange by treble frequency band level ($r = 0.76$). Therefore, the bass and midrange predictor variables were removed from the subsequent iteration. However, this updated model was not significantly different from the prior model ($p = 0.42$).

The selected model included all five fixed-effects variables and participant and genre as random-effects variables. Scatterplots illustrating the relationship between each fixed-effect variable and the dependent variable are illustrated in Figure 5-5 and Figure 5-6. The scatterplots visually show a positive relationship between average MUSHRA ratings and average low-bass levels, a steep positive relationship between average MUSHRA ratings and average bass levels, and a steep negative relationship between average MUSHRA ratings and average treble levels. There are no observable relationships between average MUSHRA ratings and average midrange levels or average CCs. Model statistics are reported in Table 5-3. Only one of the fixed-effects variables, low-bass frequency band levels, was significantly predictive of sound quality ratings. Since the sound quality ratings had been transformed using a reflection, the low-bass frequency band's negative beta

estimate suggests that this predictor was positively related to sound quality scores. That is, as low-bass levels increased, sound quality scores increased.

Table 5-2: Sequence of model fit comparisons for an optimal model which predicts sound quality scores. "Participant" and "Genre" indicate random effects variables for the different levels of those factors. The acronyms for the outcome variable and fixed effects variables are defined as follows: MUSHRA = sound quality ratings using the MUSHRA protocol, , MUSHRA-tx = MUSHRA with transformation, LB = average level (dB SPL) in the low-bass frequency band, B = average level (dB SPL)in the bass frequency band, MR = average level (dB SPL) in the midrange frequency band, T = average level (dB SPL) in the treble frequency band, CC = cepstral correlation.

	Model			Statistics		
	Outcome variable	Random effects variables	Fixed effects variables	AIC	χ^2	p-value
1	MUSHRA	Participant, Genre	~ 1	3402.8	---	---
2	MUSHRA-tx	Participant, Genre	~ 1	1600.3	1801.5	< 0.0001
3	MUSHRA-tx	Participant, Genre	LB, B, MR, T, CC	1540.7	65.59	< 0.0001
4	MUSHRA-tx	Participant, Genre	LB, T, CC	1538.5	1.76	0.4151

Table 5-3: Model statistics for the selected iteration of the model.

Fixed effects variable	**β estimate**	**SE**	**t-value**	**df**	**p-value**
Intercept	6.68	5.15	1.34	97.2	0.185
Low-bass	-0.19	0.03	-5.94	214.5	< 0.001
Bass	0.08	0.05	1.68	102.3	0.096
Midrange	0.01	0.05	0.176	76.2	0.861
Treble	0.02	0.05	0.415	101.1	0.679
Cepstral correlation	1.12	3.58	0.314	196.9	0.754

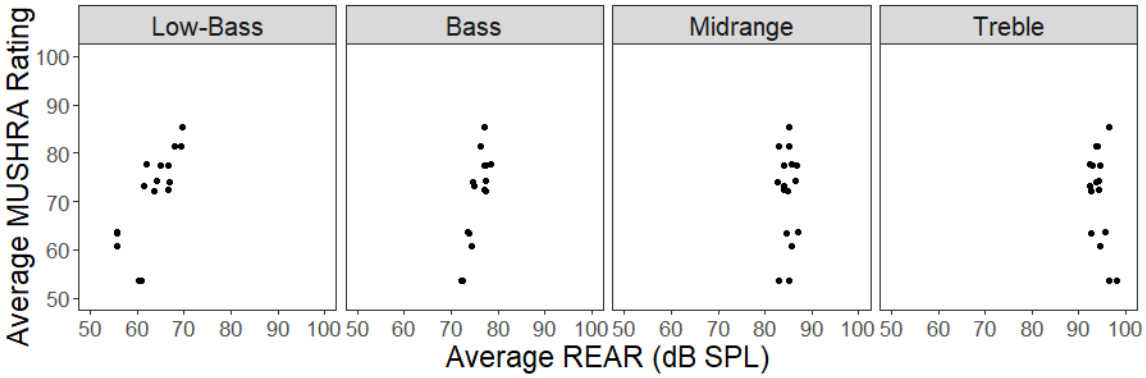

Figure 5-5: Scatterplot of average ear-simulator real ear aided response (REAR) levels in dB SPL and average sound quality rating across genres and participants (as rated using the multiple stimuli with hidden references and anchors (MUSHRA) method) per frequency band. Averages were measured across individuals for the conditions: hearing aid, genre and program. Only the low-bass level variable was predictive of sound quality ratings.

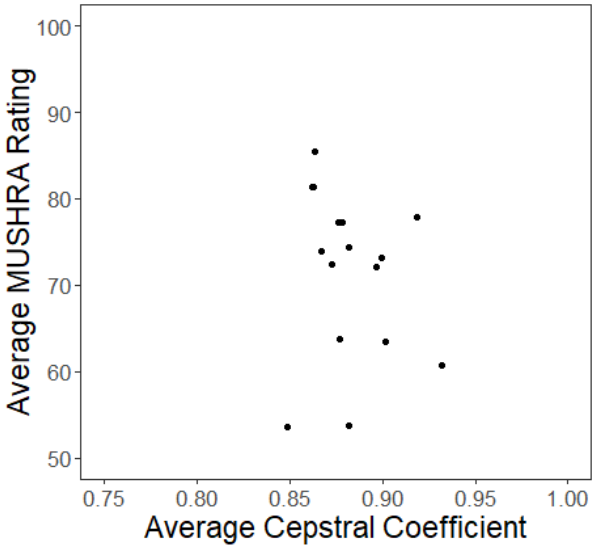

Figure 5-6: Scatterplot of average cepstral coefficient (CC) and average sound quality rating across genres and participants (as rated using the multiple stimuli with hidden references and anchors (MUSHRA) method). Averages were measured across individuals for the conditions: hearing aid, genre and program. The CC was not predictive of sound quality ratings.

5.4 Discussion

The results of this study demonstrated that electroacoustic differences between hearing aids had, on average, a greater impact on sound quality during music-listening than were differences between universal programs and music programs. Levels in the low-bass frequency band were the parameter most significantly associated with sound quality judgments. That is, as the low-bass level increased, sound quality scores increased. For example, the highest-rated hearing aid had a greater level in the low-bass frequency band than that of the lowest-rated hearing aid (average difference = 8.9 dB). The low-bass frequency band's level was also greater in the music programs versus the universal programs for all hearing aids (average difference = 1.4 dB), although this difference was much less than the difference between the highest- and lowest-rated hearing aids. In contrast, the highest-rated hearing aid's level in the treble frequency band was lower than that of the lowest-rated hearing aid (average difference = 1.2 dB), which was comparable to the average difference between treble frequency band levels in the music program versus the universal program (average difference = 1.3 dB). These smaller differences were not significantly related to sound quality judgments. Interestingly, the highest-rated hearing aid's CC was less than the lowest-rated hearing aid's CC (average difference = 0.006). In other words, the highest-rated hearing aid produced more envelope distortion compared to the lowest-rated hearing aid (potentially by faster time constants or higher compression ratios). The highest and lowest CC values were associated with HA3 and HA1 (average difference = 0.028), respectively, and therefore not monotonically related to average sound quality ratings. Only one of the four hearing aids' CC values was greater in the music program compared to its universal program (average difference = 0.029). In other words, only one hearing aid's music program reduced the envelope distortion relative to its universal program, suggesting that the electroacoustic properties in the other three hearing aids may have contradicted compression recommendations from the literature. Differences in CC values across the recordings were not significantly associated with sound quality judgments, suggesting that the differences were likely too minimal to meaningfully elicit different percepts.

The association between more low-frequency output and elevated music sound quality ratings is consistent with the literature. Middle C in the musical scale has a fundamental frequency of 262 Hz, which falls within the bass range, and there is an abundance of musical instruments designed for the bass register below middle C (Chasin, 2012). As discussed above, Franks (1982) found that optimal music sound quality was achieved when sufficient acoustic information below 200 Hz was available. Moore & Tan (2003) also found that lowering the high-pass band limit from 123 Hz to 55 Hz improved ratings of music naturalness. Increased low-frequency output has also been associated with the sound quality descriptor "fullness" (Gabrielsson & Sjögren, 1979; Jenstad, Van Tasell, & Ewert, 2003) which in turn is associated with positive ratings of "overall impression" (Davies-Venn, Souza, & Fabry, 2007). This study adds to this body of evidence because I found that four commercial hearing aids offered a low-frequency boost when switching to the music program versus the universal program, and also that additional low-bass frequency output was related to higher sound quality ratings in some but not all these of hearing aids in either program.

The fact that the low-bass frequency band, rather than a higher frequency band, was most highly associated with sound quality scores was an interesting finding. The low-bass frequency band was in the range of 100-200 Hz. Hearing aids are typically fitted at frequencies well above this frequency band due to available gain and because hearing aids are predominantly developed to work well for speech acoustics. Hearing aid fittings for speech signals are often evaluated using the Speech Intelligibility Index (SII), a metric which calculates the percentage of audible speech based on the level of speech relative to the listener's thresholds (American National Standards Institute, 1997; Amlani, Punch, & Ching, 2002; Scollie, 2018). The SII is calculated using a band-importance function, which characterizes the importance of various frequency bands for speech intelligibility (Pavlovic, 1994). The band-importance, when applied using 1/3rd octave bands function, includes 1/3rd octave bands from 160-8000 Hz and weights the audibility of frequency bands for average speech between 1 and 3 kHz (52%) more than other frequency bands, which includes a 20% weighting between 160 and 500 Hz (Table 1, Pavlovic, 1994). However, this scheme has been developed for prediction of speech intelligibility, not for the sound quality of speech or of music. In the present study, the low-bass frequency band

played a more important role than it does for the SII, relative to frequencies in the 1-3 kHz range, and that the bandwidth of our low-bass frequency band stretched slightly lower (less than 160 Hz) than frequency bands considered in a band importance function. This finding is likely due to the acoustic differences between speech and music. Musical instruments, particularly in the bass range, can have fundamental energy well below the fitting range of traditional hearing aid fitting and verification procedures (Chasin, 2012) and below most fundamental and formant energy found in speech (Hillenbrand et al., 1995). Therefore, for music fittings, it may be important to consider available gain in the frequency range well below 250 Hz. I achieved the observed levels of output in this frequency range using a sealed, unvented hearing aid. Style of hearing aid and venting will affect the availability of low-frequency content in clinical fittings.

The finding that there was no relationship between the treble-frequency band levels and sound quality ratings was unexpected because evidence to support that high-frequency output can be either beneficial or detrimental to sound quality ratings, depending on the study. Extended high-frequency output can be beneficial for music sound quality, at least for listeners who have a relatively shallow audiometric slope (Moore et al., 2011; Ricketts et al., 2008). Therefore, one may expect increased high-frequency output to be associated with enhanced music sound quality ratings. On the contrary, many of the respondents surveyed by Madsen & Moore (2014) indicated that their hearing aids were too “bright” or “shrill” when listening to music, descriptors associated with too much high-frequency output (Jenstad et al., 2003). Those respondents also reported that their hearing aids “worsened the tone quality of music”. Therefore, one may expect that less high-frequency output would be associated with enhanced music sound quality ratings. Descriptively, our results are consistent with the latter argument.

I found that, on average, a music program reduced the treble frequency output by a small amount (average difference = 1.5 dB). In addition, the favourably-rated hearing aid had less treble frequency output compared to the least favourably-rated hearing aid. However, these reductions were not statistically related to sound quality ratings. This likely occurred because there was little variance in output (Figure 5-3, treble panel) across the treble frequency band between the different hearing aids and programs, especially compared to

the variance found in low-bass frequency band. The average difference in treble frequency output between the highest-rated hearing aid and lowest-rated hearing aid was 0.4 dB, while the difference in low-bass frequency output (where the strongest association was observed) was 8.9 dB. On average, a change in 5 dB is recommended to produce meaningful perceptual changes in sound quality (Caswell-Midwinter & Whitmer, 2018). The low-bass frequency band level difference fell well above this recommendation, while the treble frequency band level difference did not. Therefore, sound quality might have been impacted if some of the hearing aids had provided more or less high-frequency amplification relative to other hearing aids.

The finding that compression, as characterized by CC measurements, was not related to sound quality scores was also unexpected. I used the CC to quantify envelope changes due to compression parameters, as the CC is a term used to measure nonlinear distortions as part of popular predictive hearing aid quality metrics (Kates & Arehart, 2014; 2016). I hypothesized that music recordings with the least compressive settings, and therefore highest CC values, would be rated most favourably. While differences in compression between hearing aids and music programs, as indexed by the CC, were observed, they were not monotonically related to average sound quality scores nor were they predictive of sound quality scores using the model. The fact that they were not statistically related to sound quality ratings was also likely due to minimal variance between groups. As discussed, the greatest difference in CC between groups was 0.041. While there is little research that relates differences in CC to perceptual quality differences, the differences observed in this study did not appear to be sufficient to alter sound quality ratings. Like the treble parameter, sound quality might have been impacted if some of the hearing aids had provided significantly more or less compression relative to other hearing aids. Future studies may wish to present stimuli at lower and higher input levels, allowing the hearing aids to apply more compression, which may in turn reveal differences between programs and manufacturers.

This study used a linear mixed effects model to explore the relationship between combined hearing aid parameters and music sound quality ratings in real-world hearing aids. I found that increased relative output in a low-bass frequency band was associated with favorable

sound quality. This study is not the first to investigate the impact of several parameters on single sound quality ratings. For instance, linear and nonlinear distortions are both detrimental to sound quality when used in combination, although nonlinear distortions typically dominate judgments when using hearing aid simulations (Arehart et al., 2011; Moore, Tan, Zacharov, & Mattila, 2004). While these findings contrast those of this study, I did not intentionally apply envelope distortion in a way that could probe such effects, and it would appear that the commercially-available aids in this study may have had less nonlinear distortion compared to what was detrimental in laboratory investigations. Furthermore, the waveform of a signal can be impacted by many parameters in addition to compression ratios and time constants (Jenstad & Souza, 2005, 2007), such as processing noise, output-limiting, and peak-clipping (Arehart et al., 2011; Davies-Venn et al., 2007; Fortune et al., 1994). Therefore, the envelope distortion in our samples may have been caused by a variety of factors, rather than a single parameter. Our study did not provide a structured sample of stimuli to systematically probe combined effects of other hearing aid parameters. Future studies may wish to probe effects of compression and high-frequency output in tandem with low-bass frequency output to determine the relative weighting of one parameter compared to another using real hearing aids.

The electroacoustic differences between universal and music programs observed in this study were only partially consistent with recommendations for music-listening made in the literature. When listening to music via hearing aids, it is suggested to increase the low-frequency gain, disable advanced signal processing mechanisms (Moore, 2016; Zakis, 2016), and minimize compression (Kirchberger & Russo, 2016a). The bass frequency band level differences between programs observed in this study are consistent with the recommendations. However, minimal differences in CC between groups did not reflect perceptual changes of envelope distortion (i.e. compression). Furthermore, I did not systematically isolate whether the small statistical effects were due to the compression and/or other signal processing features.

5.4.1 Limitations

This study only explored the band-specific levels and compression characteristics between universal and music programs, and between hearing aids, rather than the acoustic differences between music genres. However, previous studies suggest that genre is capable of influencing sound quality ratings. Davies-Venn et al. (2007) found that hearing-impaired listeners preferred classical music over vocal music, and attributed this to a preference for reduced high-frequency content and a wider dynamic range in the classical sample. Arehart et al. (2011) suggested that acoustic differences between genres influence the effect of hearing aid signal processing parameters. For example, a constant gain setting for two different musical samples may elicit different quality judgments if the samples differ in spectral content. The interaction between gain and spectral content could impact audibility. In the previous study, although I did not find any genre to be rated statistically higher than any other, I did find that some genres were more sensitive to effects between hearing aids and programs than others. For instance, sound quality judgments were more sensitive to different hearing aids when the genre was pop relative to the classical genre (Chapter 4/Vaisberg et al., 2017). In the current study, the pop genre contained more relative low-frequency spectral content compared to the classical genre (Figure 5-1). It is conceivable that the association between the low-bass frequency band and sound quality judgments would have been stronger for stimuli which contained more low-bass frequency content. In the current study, to observe the effects of hearing aids and programs alone, I removed the between-genre variability by including genre as a random intercept in our model. However, future studies may wish to develop genre-specific models so that researchers can observe the pattern of effects based on the acoustic information that is available.

In addition, our study only explored the effect of hearing aid processing on the sound quality of recorded music – not of live music. Most recorded music has a dynamic range of 25 dB or less, which is below the dynamic range of speech (Kirchberger & Russo, 2016a). Furthermore, the listener can control the listening level of the music through manipulation of the volume control. When asked to adjust aided recorded music to a comfortable listening level, listeners have been found to adjust levels to an average 80.3 dBA at the tympanic membrane (Croghan et al., 2016). If the input dynamic range for many

modern digital hearing aids is 0-96 dB (Chasin, 2012), and the dynamic range of recorded music is 25 dB, then most recorded music could be aided within the input dynamic range of the hearing aid while avoiding substantial peak-clipping or compression limiting. Live music levels, however, are not under the control of the listener and often exceed the maximum power output of the hearing aid. Many styles of music can produce levels above 100 dB SPL with crest factors of up to 20 dB (Chasin, 2006, 2012), exceeding the input dynamic range of a hearing instrument. For instance, Killion (2009) has observed instantaneous live symphonic and jazz music peaks as high as 116 dB SPL which have caused noticeable distortion in hearing aids belonging to both listeners and musicians. Therefore, live music may create additional distortions that may be less frequent for recorded music. Future studies may wish to model hearing aid output against sound quality for live music separately from recorded music.

5.5 Conclusion

The commercial hearing aids analyzed in this article provided music programs that contained some signal processing adjustments consistent with the literature's recommendations for improved music-listening. For the listeners studied, all fitted music programs increased the low-frequency response of the hearing aid by an average 2.4 dB and reduced the high-frequency response of the hearing aid by an average 1.5 dB. Only one of the music programs reduced the envelope distortion of the signal. The measured electroacoustic differences were greater between the different hearing aids than they were between programs within a single hearing aid. Of the signal processing adjustments, output in the low-bass frequency band (particularly in the 100-200 Hz range) was the parameter most predictive of enhanced sound quality scores. Neither differences in output level in the bass (250-800 Hz), midrange (1000-2500 Hz) and treble (3000-10000 Hz) frequency bands, nor amount of envelope distortion, were significantly associated with sound quality scores. The findings reported in this article suggest that neither manufacturers' hearing aids nor programs within each hearing aid differ in their default electroacoustic behaviour in a way that is perceptually meaningful for music at a comfortable listening level, except for low-bass frequency band output levels. The findings also inform researchers and clinicians about the nature of contemporary music program electroacoustic behaviour.

Furthermore, they can inform, indirectly, on how hearing aid style and parameters might be selected in the interest of improving users' music-listening experiences. Specifically, it is important to consider the level of sound well below the typical bandwidth fitting range of a hearing aid if optimal music sound quality is the fitting goal. Vented versus sealed fittings are expected, therefore, to differ in terms of music sound quality due to their different approaches to delivering low-frequency sound. Venting and low-frequency gain adjustments may interact, and this may affect music sound quality. Further research is needed to investigate the relationship between these electroacoustic parameter changes and sound quality judgments in a population including more severe hearing losses and different configurations including vented fittings. Also, stimulus level was not systematically varied within listeners, and therefore cannot specify the required sensation levels for optimal sound quality. In summary, the research here describes the electroacoustic properties of the hearing aids that were related to music sound quality, when hearing aids were fitted at default settings, and not the electroacoustic impacts of systematic parameter adjustments or all commercially-available hearing aids.

5.6 References

Alexander, J. (2013). Individual variability in recognition of frequency-lowered speech. *Seminars in Hearing*, *34*(2), 86–109.

American National Standards Institute. (1997). *Methods for Calculation of the Speech Intelligiblity Index. ANSI S3.5-1997 (R2017)*. New York: Acoustical Society of America.

Amlani, A. M., Punch, J. L., & Ching, T. Y. C. (2002). Methods and applications of the audibility index in hearing aid selection and fitting. *Trends in Amplification*, *6*(3), 81–129.

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

Audio Spectrum Explained. (2018). Retrieved September 27, 2018, from https://www.teachmeaudio.com/mixing/techniques/audio-spectrum/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bradford, K. K. (2014). *The effect of hearing aid program on the perceived sound quality of music* (Doctoral dissertation). Ruston, Louisiana: Louisiana Tech University.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*(2), 261–304.

Caswell-Midwinter, B., & Whitmer, W. M. (2018). Discrimination of frequency-gain curve adjustments. *Poster session presented at the 2018 International Hearing Aid Research Conference*. Tahoe City, California.

Chasin, M. (2006). Hearing aids for musicians. *The Hearing Review*, *13*(3), 1–11.

Chasin, M. (2012). Music and hearing aids: An introduction. *Trends in Amplification*, *16*(3), 136-129.

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

Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, *35*(5), e170–e184.

Croghan, N. B. H., Swanberg, A. M., Anderson, M. C., & Arehart, K. H. (2016). Chosen listening levels for music with and without the use of hearing aids. *American Journal of Audiology*, *25*, 161–166.

Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, *18*(8), 688–699.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: Sage Publications.

Fortune, T. W., Woodruff, B. D., & Preves, D. A. (1994). A new technique for quantifying temporal envelope contrasts. *Ear and Hearing*, *15*(1), 93–99.

Franks, J. R. (1982). Judgments of hearing aid processed music. *Ear and Hearing*, *3*(1), 18–23.

Fulford, R., Ginsborg, J., & Greasley, A. (2015). Hearing aids and music : the experiences of D / deaf musicians. In *Proceedings of the Ninth Triennial conference for the European Society for the Cognitive Sciences of Music*. Manchester, UK.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033.

Galster, J., Rodemerk, K., & Fitz, K. (2014). Preferred aided listening levels for music in the soundfield. *Poster session presented at the 2014 International Hearing Aid Research Conference, Tahoe City, CA*.

Hansen, M. (2002). Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear and Hearing*, *23*(4), 369–380.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

International Electrotechnical Commission. (2008) *Electroacoustics – Hearing aids – Part 15: Methods for characterizing signal processing in hearing aids. IEC 60118-15.* Geneva, Switzerland.

International Telecommunications Union (2015). *Method for subjective assessment of intermediate quality level of audio systems*. *Recommendation ITU-R BS.1534-3* Geneva, Switzerland.

Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech Language and Hearing Research*, *48*(3), 651-667.

Jenstad, L. M., & Souza, P. E. (2007). Temporal envelope changes of compression and speech rate: Combined effects on recognition for older adults. *Journal of Speech Language and Hearing Research*, *50*, 1123–1138.

Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology*, *14*(7), 347–360.

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

Kates, J. M., & Arehart, K. H. (2016). The hearing-aid audio quality index (HAAQI). *IEEE/ACM Transactions on Speech and Language Processing*, *24*(2), 354–365.

Killion, M. (2009). What special hearing aid properties do performing musicians require? *The Hearing Review*, *16*(2), 20–31. Retrieved from http://www.hearingreview.com/2009/02/what-special-hearing-aid-properties-do-performing-musicians-require/

Kimlinger, C., McCreery, R., & Lewis, D. (2015). High-frequency audibility: the effects of audiometric configuration, stimulus type, and device. *Journal of the American Academy of Audiology*, *26*(2), 128–137.

Kirchberger, M., & Russo, F. A. (2016a). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, *20*, 1–16.

Kirchberger, M., & Russo, F. A. (2016b). Harmonic frequency lowering : Effects on the perception of music detail and sound quality. *Trends in Hearing*, *20*, 1–12.

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. Retrieved from https://cran.r-project.org/package=ez

Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, *19*(6), 519–526.

Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, *18*, 1–29.

Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *Journal of the Audio Engineering Sociaty*, *64*(3), 112–123.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, *32*(5), 556–568.

Moore, B. C. J., & Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *The Journal of the Acoustical Society of America*, *114*(1), 408–419.

Moore, B. C. J., Tan, C. T., Zacharov, N., & Mattila, V. V. (2004). Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion. *Journal of the Audio Engineering Society*, *52*(12), 1228–1244.

Mussoi, B. S. S., & Bentler, R. A. (2015). Impact of frequency compression on music perception. *International Journal of Audiology*, *54*, 627–633.

Parsa, V., Scollie, S., Glista, D., & Seelisch, A. (2013). Nonlinear frequency compression: Effects on sound quality ratings of speech and music. *Trends in Amplification*, *17*(1), 54–68.

Pavlovic, C. V. (1994). Band importance functions for audiological applications. *Ear and Hearing*, *15*(1), 100–104.

Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*, 160–172.

Scollie, S. (2018). 20Q: Using the aided speech intelligibility index in hearing aid fittings. *AudiologyOnline, Article 23707*. Retrieved from www.audiologyonline.com

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., … Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, *9*(4), 159–197.

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in Amplification*, *6*(4), 131–165.

Struck, C. J., & Prusick, L. (2017). Comparison of real-world bandwidth in hearing aids vs EarLens light-driven hearing aid system. *The Hearing Review*, *24*(3), 24–29. Retrieved from http://www.hearingreview.com/2017/03/comparison-real-world-bandwidth-hearing-aids-vs-earlens-light-driven-hearing-aid-system/

Tabachnik, B. G., & Fidell, L. S. (1996). Multivariate Normality. In C. Woods (Ed.), *Using Multivariate Statistics* (Third Edit, p. 381). California State University, Northbridge: HarperCollins College Publisher.

R Core Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Compting, Vienna, Austria. URL Https://Www.R-Project.Org/*.

Vaisberg, J. M., Folkeard, P., Parsa, V., Froehlich, M., Littmann, V., Macpherson, E. A., & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline, Article 20872*. Retrieved from www.audiologyonline.com

Vaisberg, J. M., Martindale, A. T., Folkeard, P., & Benedict, C. (2018). A qualitative study of the effects of hearing loss and hearing aid use on music perception in

performing musicians. *Journal of the American Academy of Audiology*, *Epub ahead of print*, 1–15.

Zakis, J. A. (2016). Music perception and hearing aids. In G. R. Popelka, B. C. J. Moore, R. R. Fay, & A. N. Popper (Eds.), *Hearing Aids* (pp. 217–252). Cham, Switzerland: Spring International Publishing Swizterland.

Chapter 6

6 Preferred frequency shaping for aided music-listening

Objectives: Clinical procedures for hearing aid fitting have predominantly been developed for speech signals, with less emphasis on non-speech signals like music. The preferred frequency response for music is therefore not well understood, nor are the perceptual domains of sound quality in hearing-aid-processed music signals. The objectives of this study were to (1) determine whether listeners' preferred frequency shaping differs for music versus speech listening, relative to DSL-prescribed shaping; and (2) assess which sound quality descriptors were related to listener gain adjustments.

Design: Twenty-five adults with mild-sloping-to-moderately severe hearing loss (twelve aid users, thirteen non-users) participated in this study over two experiments. In Experiment One, listeners completed a three-dimensional modified simplex procedure (Neuman, Levitt, Mills, & Schwander, 1987) for pop and classical music, and male and female speech, to optimize preference via gain adjustments in low (0.1-0.8 kHz), mid (1-2.5 kHz), and high (3-10 kHz) frequency bands. In Experiment Two, listeners conducted sound quality ratings of total impression, fullness, loudness, and sharpness for each stimulus amplified using simplex-derived preferred frequency shaping and prescribed frequency shaping. Listeners also provided intelligibility ratings for speech passages. Experiments I and II were implemented using a laboratory hearing aid simulator.

Results: Across stimuli, listeners increased the gain most substantially for the low-frequency band (7.3 dB), followed by a high-frequency decrease (-4.3 dB), and with the smallest adjustment for the mid-frequency band (-1 dB). Low-frequency gain adjustments were affected by stimulus, with music exhibiting larger gain increases. Low-frequency gain was increased more for pop, though not significantly, compared to classical music. Hearing aid users and non-users did not adjust preferred shaping differently. Relative to prescribed shaping, preferred shaping was rated higher in overall impression for pop music and male speech, in loudness for pop music, and in fullness for pop and classical music. Preferred shaping was rated lower in sharpness for male speech, and lower in intelligibility for female speech. Fullness and sharpness ratings were positively and negatively associated with

overall impression ratings, respectively, with fullness being the more strongly associated descriptor of the two. Intelligibility ratings were negatively and most strongly associated with overall impression ratings across speech stimuli.

Conclusions: Listeners increased low-frequency gain relative to their individually-prescribed frequency shaping more substantially than other adjustments, and most substantially for pop music, followed by classical music, and least substantially for speech. The need to improve the sense of fullness likely drove those adjustments. The need to preserve intelligibility likely restricted users' gain adjustments for speech, motivating the implementation of preferred shaping for music and prescribed shaping for speech in hearing aid fittings. Individual factors should be considered if implementing these findings clinically.

6.1 Introduction

Hearing aids are typically fitted using speech signals, rather than with other signals such as music. Standardized, evidence-based prescriptive formulae, like CAMEQ2-HF (Moore, Glasberg, & Stone, 2010), DSLv5.0 (Scollie et al., 2005), and NAL-NL2 (Keidser, Dillon, Flax, Ching, & Brewer, 2011) provide individualized targets with the general goal of improving speech intelligibility, and do so by calculating prescribed gain as a function of hearing loss, using a long-term average speech spectrum (e.g. Holube, Fredelake, Vlaming, & Kollmeier, 2010) as the input spectrum. The use of prescriptive formulae is considered an important part of best-practice guidelines (American Speech-Language-Hearing Association, 1998; British Society of Hearing Aid Audiologists, 2012; Valente et al., 2006). Some of these guidelines also emphasize the need to provide tolerable and comfortable amplification (American Speech-Language-Hearing Association, 1998) which is consistent with studies of the tradeoffs between intelligibility and sound quality in hearing aid fittings (Humes, 2003; Jenstad et al., 2007). Evaluating sound quality for hearing aid fittings that are based upon an intelligibility-driven prescription is of particular importance because poor sound quality remains a significant barrier to device adoption (Abrams & Kihm, 2015).

There have been several evaluations of sound quality for prescriptive formulae in the literature using laboratory-based research designs. Many studies have compared sound quality preferences between prescriptive formulae that have been fitted to optimize intelligibility. For example, comparisons of NAL-NL2 and CAM2 (a variation of CAMEQ2-HF) revealed individual differences in sound quality preferences for either formula (Johnson, 2013; Moore & Sęk, 2012, 2016) which were attributed, in part, to factors including greater high-frequency gain in CAM2, stimulus input level, noise type, or hearing profile. Other studies have identified a range of fittings that deviate from prescribed fittings but that maintain both acceptable intelligibility and quality. For example, Jenstad et al., (2007) investigated the range of audibility in the low- and high-frequency regions that could facilitate an acceptable range of fittings for speech. Listeners preferred a range of frequency responses that varied by about 10 dB in both the low- and high-frequency regions, and that included prescribed targets for the DSL[i/o] formula. Similarly, Dirks, Ahlstrom, & Noffsinger (1993) identified a range of acceptable fittings within 6 dB increments of NAL targets that were not significantly different from one another, as did other studies for other fitting formulas (Kuk & Pape, 1992; van Buuren, Festen, & Plomp, 1995). These studies inform our understanding of how aided speech is perceived when shaped using prescriptive methods.

Music is an interesting signal for the investigation of speech-based prescriptive formula for two reasons. The first reason relates to the acoustic structure of music. The acoustics of music are fairly-well understood, as is hearing aid users' music-listening satisfaction. While speech acoustics originate from the vocal tract, and have well-understood spectral content and levels (Hillenbrand, Getty, Clark, & Wheeler, 1995; Olsen, 1998), music acoustics can originate from a variety of vocal and instrumental sources, which have the potential to produce a much wider range of spectra and more variability in level (Chasin & Hockley, 2014; Kirchberger & Russo, 2016). This signal variability may relate to reports from hearing aid users that they are frequently dissatisfied with hearing aid amplified music (Fulford, Ginsborg, & Greasley, 2015; Leek, Molis, Kubli, & Tufts, 2008; Looi, Rutledge, & Prvan, 2018; Madsen & Moore, 2014; Vaisberg, Martindale, Folkeard, & Benedict, 2018). The second reason relates to listening purposes and the role of intelligibility. When listening to speech, intelligibility is often highly relevant to purpose. Similarly, music often

includes lyrics, so it may be important to optimize lyric-intelligibility. However, many listeners do not to attend to music lyrics (Condit-Schultz & Huron, 2015) and the sound quality of music is also driven by non-lyrical components of song. Therefore, lack of intelligibility in music may not impact listening experience as it would for speech communication. This may mean that sound quality optimization may be the primary goal for hearing-aid-amplified music. If this is the case, a sound quality optimization task centered on music may produce a frequency response which differs from the frequency response that would be selected for speech intelligibility.

Many researchers have investigated the impact of different hearing aid settings for music-listening. The studies by Moore & Sęk (2012, 2016) comparing preferences between prescriptive formulae compared preferences between formulae for music, revealing a slight preference for CAM2 over NAL-NL2. There have also been investigations of single signal-processing parameters. For instance, additional low-frequency gain is often preferred for music (Arehart, Kates, & Anderson, 2011; Franks, 1982; Punch, 1978), as is extended high-frequency audibility, at least for listeners with flat hearing configurations (Moore, Füllgrabe, & Stone, 2011; Ricketts, Dittberner, & Johnson, 2008). Linear processing is also typically preferred compared to compressive nonlinear processing (Davies-Venn, Souza, & Fabry, 2007; Kirchberger & Russo, 2016).

There are several challenges in implementing findings from these studies in the development of hearing aid settings for music. Prescriptive methods vary in multiple parameters (i.e., frequency shaping, compression, kneepoints), so it can be challenging to identify single-parameter differences that should be adjusted for music-listening. Also, the multiple parameters in prescriptions may interact with one another, highlighting the need for a method allowing for the evaluation of hearing aid settings in combination – not in isolation.

A method in which users can incrementally adjust multiple settings simultaneously and relative to their own prescriptions may be more informative for clinically informative results for music-listening. Fortunately, several methods exist in the literature. The modified simplex procedure (hereinafter referred to as “simplex”) is a reliable, systematic,

and time-efficient procedure for multiparameter studies. In simplex, listeners perform a series of paired-comparisons of hearing-aid processed speech differing in fixed increments of multiple fitting parameters such as low- and high-frequency gain relative to prescribed settings (Amlani & Schafer, 2009; Kuk & Pape, 1992, 1993; Kuk & Lau, 1995; Kuk & Lau, 1996; Neuman et al., 1987; Preminger, Neuman, Blakke, Deirdre, & Levitt, 2000; Stelmachowicz, Lewis, & Carney, 1994). The simplex procedure has been applied for a three-channel system consisting of a low-, mid-, and high-frequency gain regions with crossover frequencies varying between conditions (Dirks et al., 1993), and with additional signal processing parameters (i.e., noise reduction, spectral enhancement, spectral adjustment) with adaptive increments (Franck, Dreschler, & Lyzenga, 2004; Franck, Boymans, & Dreschler, 2007). Alternatives to simplex also exist, and typically rely on user self-adjustment, in which users seek an optimal combination of self-adjusted settings using several parameters or similar interfaces corresponding to features which may include treble, bass, and overall level (Boothroyd & Mackersie, 2017; Dreschler, Keidser, Convery, & Dillon, 2008; Keidser & Alamudi, 2013; Mackersie, Boothroyd, & Lithgow, 2018; Nelson, Perry, Gregan, & VanTasell, 2018). To date, these studies have predominantly focused on listening preferences for speech signals in a variety of listening conditions, rather than on music.

This study used the simplex procedure to determine preferred gains for speech and music across three parameters consisting of low-frequency (0.1-0.8 kHz), mid-frequency (1-2.5 kHz), and high-frequency (3-10 kHz) gain relative to a prescribed DSL v5.0 fitting. The three-frequency band hearing aid configuration was most similar to the implementation by Dirks et al., (1993) and was selected for several reasons. First, past implementations of the simplex and many self-adjustment tools have typically allowed listeners to modify only two frequency regions: bass (low frequencies) and treble (high frequencies). Allowing listeners to indicate preferences across three frequency regions enabled them to have greater control over the spectral shape of the aided output. Second, this three-channel configuration has been associated with higher music quality ratings compared to more channels (Croghan, Arehart, & Kates, 2014). Many studies have found listener preferences for more low-frequency gain and/or less high-frequency gain compared to prescribed shaping (Boothroyd & Mackersie, 2017; Boymans & Dreschler, 2012; Kuk & Pape, 1992,

1993; Nelson et al., 2018; Preminger et al., 2000; Zakis, Dillon, & McDermott, 2007), which may inform how listeners would adjust the gain in this study. Third, the middle frequency band corresponds to frequencies that are most heavily weighted for speech intelligibility (SII, American National Standards Institute, 1997; Pavlovic, 1994), and some prescriptive formulae prioritize audibility in this region relative to others. Therefore, it was of interest to determine how listeners would adjust the mid-frequency region relative to prescribed settings, as well as how the low- and high-frequency regions would be adjusted relative to those mid-frequency adjustments. The determination of preferred gains relative to a prescribed fitting for speech intelligibility was done to determine whether listeners preferred different frequency shaping for speech and music.

The primary objectives of this study were (1) to quantify listener's preferred frequency shaping compared to a prescribed frequency shaping for music using overall preference as a criterion, (2) to determine whether unique preferred frequency shaping exists for speech versus music; and (3) to determine whether listeners' preferred frequency shaping produced meaningfully different ratings across different sound quality descriptors. We performed this study using equipment that provided frequency shaping in combination with multichannel dynamic range compression so that any interactions of these would be represented. The objectives were investigated over the course of two experiments. A secondary objective of Experiment One was centered on whether gain adjustments varied as a function of stimulus (which included two music genres and male and female speech).

6.2 Experiment One

6.2.1 Methods

6.2.1.1 Participants

This study included 25 adult listeners between the ages of 34-81 years (mean = 65.8, SD = 10.3). Pure-tone audiometry was conducted using ER-3A insert phones. Participants were tested in a sound-attenuated booth. On average, listeners presented with sensorineural, mild-sloping-to-moderate hearing loss. Figure 6-1 illustrates hearing aid user status and air conduction thresholds for both ears for all participants. Almost half (12) of the participants were experienced hearing aid users, with the duration of hearing aid experience ranging from 2-22 years (mean = 8.33, SD = 6.2) among the user group. A mixed ANOVA with 4-frequency (0.5, 1, 2, and 4 kHz) pure tone average (PTA4) as the dependent variable, user group as the between-subjects factor, and ear as the within-subjects factor revealed no differences between users' and non-users' ($F_{(1,23)}$=1.45, p=0.24, η^2=0.05) ears ($F_{(1,23)}$=0.19, p=0.67, η^2=0.0009), nor any interaction between these variables ($F_{(1,23)}$=0.55, p=0.47, η^2=0.003), suggesting comparable audiograms and balanced user/non-user groups according to PTA4. This study was approved by the Western University Health Science Research Ethics Board. Participants were compensated at a rate of $10/hour for their time and were offered refreshments and breaks during the sessions.

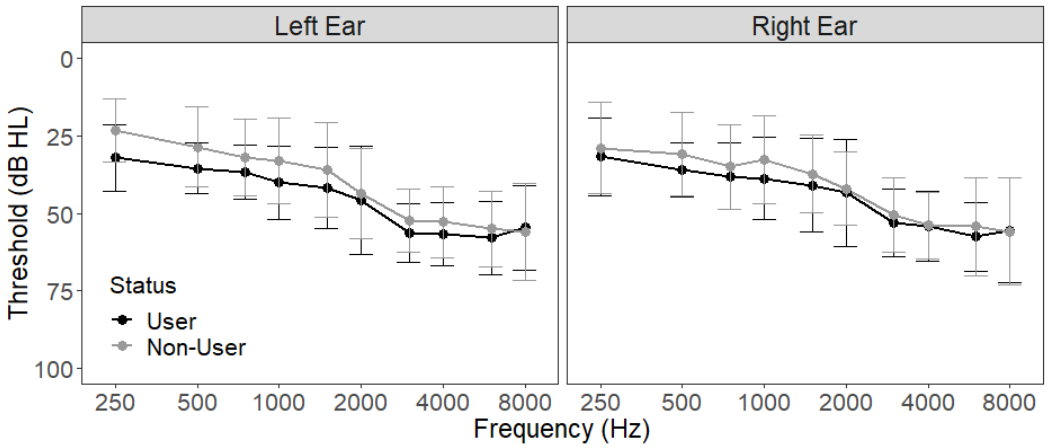

Figure 6-1: Mean air conduction pure-tone thresholds for hearing aid user and hearing aid non-user groups for the left ear (left panel) and right ear (right panel). Error bars represent one standard deviation of the mean.

6.2.1.2 Test materials

The test materials were two IEEE speech passages and two music passages downloaded from iTunes. The speech passages were chosen to represent both genders of voice, and consisted of a male-spoken sentence, “Raise the sail and steer the ship northwards” and a female-spoken sentence “Would you please give us the facts?”. The music samples were chosen to represent genres that may interact differently with signal processing adjustments (Arehart et al., 2011; Davies-Venn et al., 2007). These consisted of a 5.1 second sample from the contemporary/pop song “A Little Help from my Friends” by The Beatles and a 2.6 seconds sample from Mozart’s classical string arrangement “Serenade No. 6, K. 239 “Serenata notturna: III. Rondo. Allegro” by the Franz Liszt Chamber Orchestra & Sandor Frigyes. The lengths of the music samples were long enough to include a full musical phrase, but short enough so that sufficient testing and minimal fatigue occurred during a test session. Stimulus length varied due to tempo and phrasing differences between the two samples. Average long-term spectra of the samples are illustrated in Figure 6-2. Experimental development suggested that two seconds per stimulus was sufficient for participants to confidently judge stimulus preference. The stimuli were digitized at a 44.1 kHz sampling rate at 16 bits per sample. If a stimulus was in stereo format, it was summed to mono format, and duplicated for both channels, so both ears received the same signal.

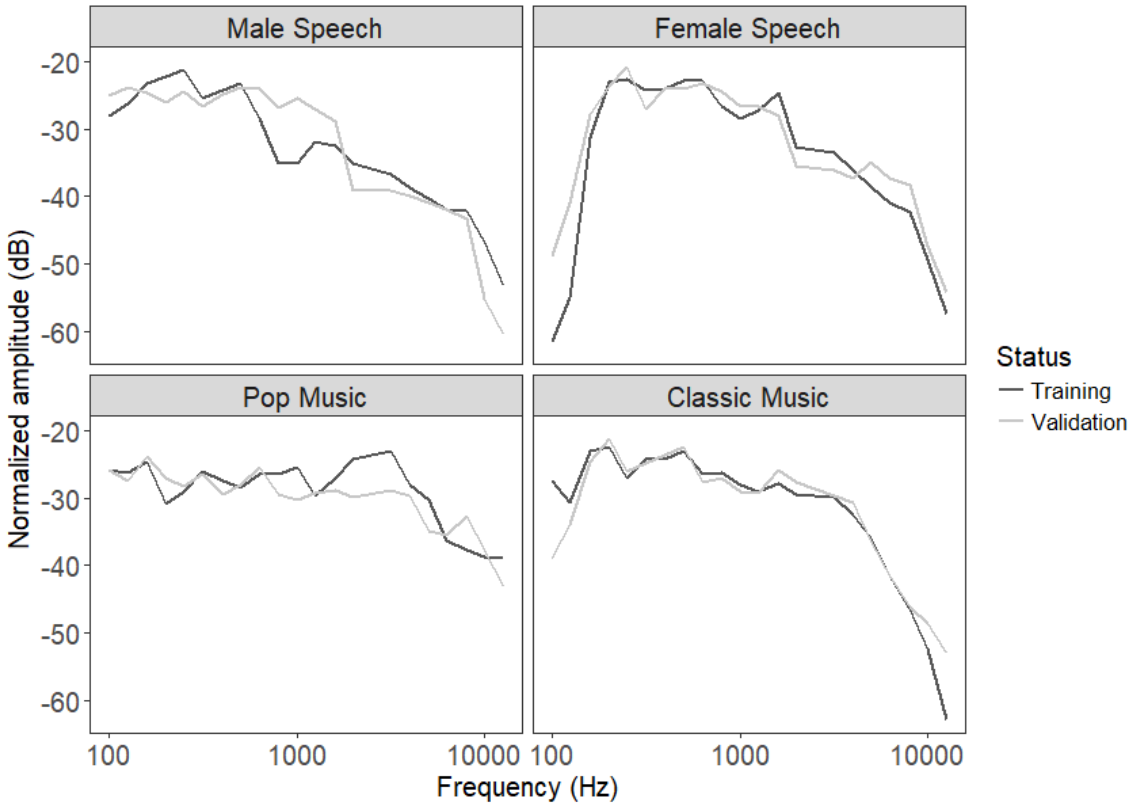

Figure 6-2: Long-term frequency spectra measured in 1/3rd octave bands for the stimuli in this study. Black lines represent training stimuli used in both Experiments One and Two, and gray lines represent validation stimuli used in Experiment Two only.

6.2.1.3 Calibration and fitting to initial targets

This study used the open source master hearing aid (openMHA; Herzke, Kayser, Loshaj, Grimm, & Hohmann, 2017) to process and amplify the test materials. The openMHA was installed on a Linux computer and connected to a low-latency Focusrite Scarlett 18i8 USB soundcard, which sent the signals to two Etymotic Research 4p (ER4p) transducer sets via two different channels. Separate ER4p sets were used because there was significant cross-talk when using a single ER4p set via a single stereo channel. To calibrate the openMHA, the test materials were presented from the ER4p transducers into a Bruel & Kjaer (B&K) Type 4157 occluded ear simulator mounted on a B&K Type 2250 sound level meter. The speech passages and music samples were digitally scaled such that an unamplified (zero-gain) version of each stimulus produced a long-term average of 70 dB SPL over 30 seconds of the repeated stimulus. This scaling allowed for sufficient headroom for listeners to increase the overall level, and modify frequency shaping during the experiment before encountering digital peak-clipping or earphone total harmonic distortion. Pre-testing electroacoustic evaluations indicated that the soundcard provided sufficient amplification to achieve both prescribed and preferred shaping. Daily calibration checks were performed by measuring the output of the International Speech Test Signal (ISTS, Holube, Fredelake, & Vlaming, 2010) from the ER4p transducer in a commercially-available hearing aid analyzer (Audioscan Verifit2) 0.4 cc coupler. The ISTS was scaled to produce the same output as the test stimuli.

To simulate a hearing aid, the openMHA implemented a multiband dynamic compression plugin. In this configuration, the openMHA inputs were the digital stimuli. Next the openMHA applied a reference input peak level of 125 dB (0 dB full scale corresponds to this SPL level) to determine the simulated SPL level of the waveform. The test materials were scaled to a digital -55 dB full scale, so that the openMHA would apply level-dependent gains for a simulated average 70 dB SPL input level. Next, the waveform was processed using a fast-Fourier-transform-filterbank plugin, in which the signal was processed into 21 Hann-window channels centered on 1/3rd octave bands from 0.1-10 kHz. Frequency shaping and dynamic range compression were applied using DSL v5.0-

prescribed gains for 55, 65, and 75 dB SPL input levels at each of the 21 channels. openMHA's default attack and release times were applied. The amplified waveform was then compiled from the channels using a digital reconstruction plugin, after which point, the respective ears' digital outputs were sent to their corresponding ER4p transducers.

Individual openMHA fittings were verified in the Audioscan Verifit2. First, thresholds and wideband real-ear-to-couple difference (wRECD) measurements were entered into the Verifit2, and targets were generated for a 70 dB SPL input level by the Verifit2 software. The wRECD was measured in order to capture individual ear canal resonances across a wide range of frequencies. Second, the openMHA output was routed to the ER4p transducers, which coupled to the Verifit2 0.4 cc couplers. The openMHA output was fine-tuned within 3.5 dB RMS of targets at octave and interoctave frequencies from 0.25–8 kHz across listeners. This is within the tolerance associated with ideal SII values for various degrees of hearing loss in clinical hearing aid fittings (Folkeard, Saleh, Glista, & Scollie, 2018; McCreery, Bentler, & Roush, 2013).

6.2.1.4 Modified simplex procedure

This study implemented the simplex in a three-dimensional space. This permitted listeners to listen to an initial frequency shaping and compare it with subsequent versions generated by making preference-based gain adjustments in three frequency bands. The initial shaping in a simplex procedure is typically determined using a prescriptive method (Kuk & Lau, 1995; Kuk & Lau, 1996; Preminger, Neuman, Blakke, Deirdre, & Levitt, 2000; Stelmachowicz, Lewis, & Carney, 1994). The initial shaping in this study was amplified using DSL v5.0 gains at a simulated hearing aid input level of 70 dB SPL. This level has been identified as an average comfortable listening input level for aided music listening, and has been recommended for use in music-based hearing aid research (Croghan, Swanberg, Anderson, & Arehart, 2016). The step size of gain adjustment was ±6 dB across a subset of the 21 openMHA channels which fall into the following frequency bands: low-frequency (0.1-0.8 kHz), mid-frequency (1-2.5 kHz), and high-frequency (3-10 kHz). A ±6 dB adjustment step was also selected in the previous, three-channel implementation of the simplex procedure (Dirks et al., 1993) and this step was comparable to the minimum value in which participants detected differences in frequency gain characteristics for speech-

shaped noise (Caswell-Midwinter & Whitmer, 2019), and to adjustments used to detect differences in sound quality descriptors (Caswell-Midwinter & Whitmer, 2018).

The simplex procedure determines preferred frequency shaping using a series of iterations (Figure 6-3), following methods for a two-dimensional simplex originally described by Neuman et al. (1987). The center coordinate (0,0,0) represented the prescribed initial estimate, and each step along the x, y, and z axes represented a ±6 dB adjustment in either the low-, mid-, or high-frequency gain. Each iteration consisted of three paired comparisons. In the first iteration, the listener compared prescribed frequency shaping (0, 0, 0) with a ±6 dB low-frequency (1, 0 ,0), mid-frequency (0, -1, 0) or high-frequency (0, 0, -1) gain adjustment. Listeners could repeat each stimulus pair once if they wished to hear the stimuli again before indicating which of the pair was preferred. The preferred stimuli from each comparison determined the center coordinate shaping for the subsequent iteration. That center coordinate would either be compared with additional gain (if the listener preferred the adjustment), or with less gain relative to the initial estimate (if the listener did not prefer the adjustment and preferred the estimate instead). In the Figure 6-3 examples, the preferred shaping after the first iteration was (1,0,-1) and was compared to (2,0,-1) low-frequency, (1,-1,-1) mid-frequency, and (1,0,-2) high-frequency gain adjustments in the second iteration. During the second iteration, the listener preferred the center coordinate estimate after all three comparisons. Therefore, cell (1,0,-1) continued to be the center coordinate estimate during the third iteration, and each dimension was compared to the reflected (0,0,-1) low-frequency, (1,1,-1) mid-frequency, and (1,0,0) high-frequency gain adjustments. These reflected adjustments across all three parameters was known as a reversal, and completion of two reversals was used as a stopping rule. During the third iteration, the listener continued to prefer the center coordinate estimate indicating that another reversal would occur. After two reversals, the listener had compared the center cell to every possible parameter manipulation and had not preferred any adjustments. Therefore, testing terminated and the center coordinate at the final iteration was considered the listener's preferred frequency shaping. If after 18 iterations (54 comparisons), the listener did not complete two reversals, then the final preferred shaping after the 18th iteration was considered the listener's preferred shaping. This was known as a timeout. In

past three-dimensional adaptive procedures similar to simplex, a set of 54 paired comparisons has been considered a reasonable amount of testing to minimize fatigue and reach an optimum (Franck, Dreschler, & Lyzenga, 2004; Franck, Boymans, & Dreschler, 2007).

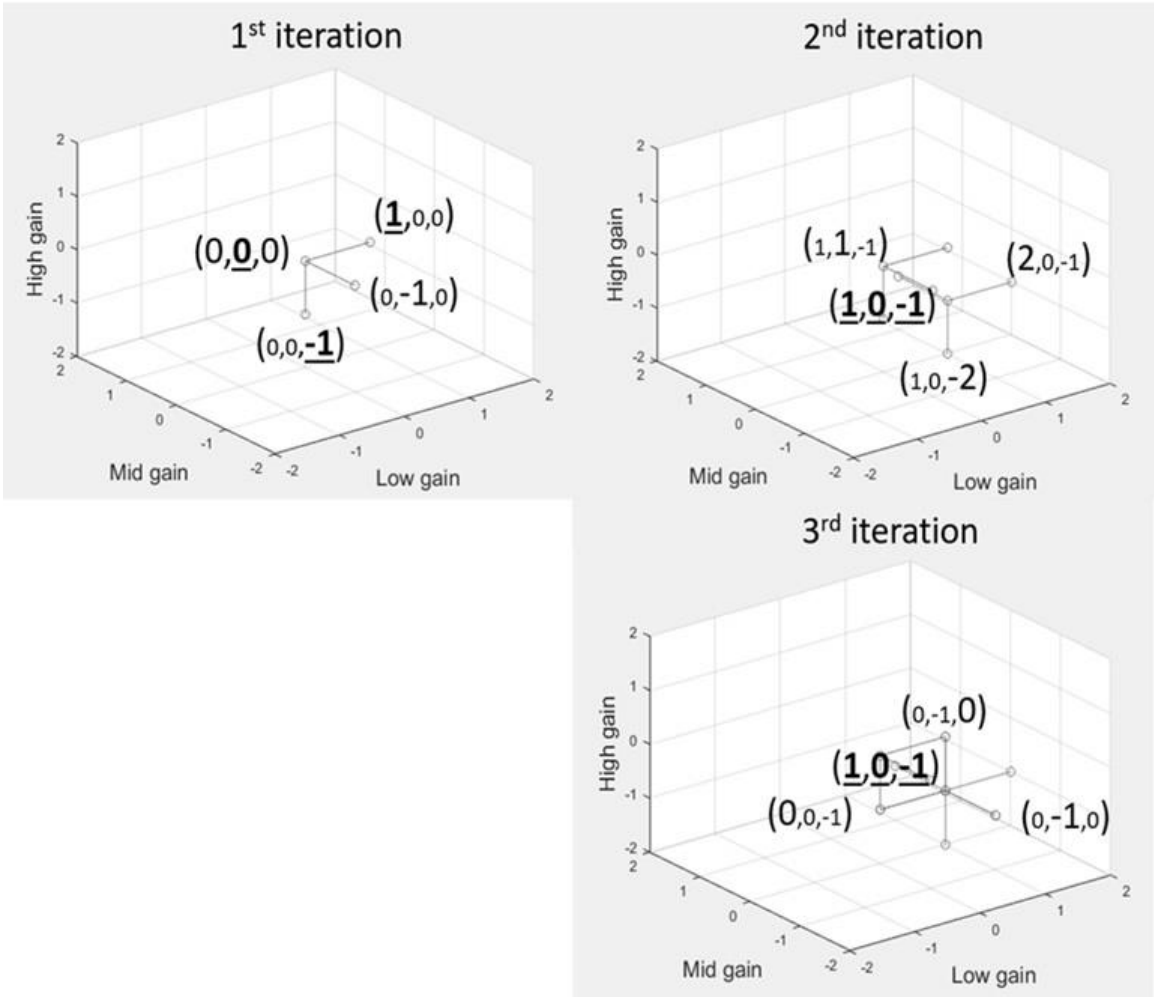

Figure 6-3: Visualization of a three-iteration modified simplex procedure run. The x, y, and z coordinates correspond to the parameter settings for the low (0.1-0.8 kHz), mid (1-2.5 kHz) and high (3-10 kHz) gain adjustments, respectively. The cell (0,0,0) represents the initial estimate using prescribed frequency shaping. Each other cell represents shaping that has been adjusted by ±6 dB per step size relative to the initial estimate. In each iteration, the center cell with three larger digits represents the preferred shaping from the previous iteration (or the prescribed shaping in the first iteration). For cells neighbouring

the center, the single uppercase digit represents adjusted shaping along that digit's dimension. Paired comparisons are conducted between the center cell and the three adjusted cells. An underlined digit represents the listener's selected preference within a comparison along that digit's dimension. A reversal occurs after the second iteration because the listener prefers the center cell (1,0,-1) relative to adjustments along any of the three dimensions. The run terminates after the third iteration because the center cell (1,0,-1) is preferred twice — after both the second and third iterations. Therefore, this participant's final preferred shaping corresponds to the coordinates: (1,0,-1).

Past simplex studies (Kuk & Lau, 1995; Kuk & Lau, 1996; Neuman, Levitt, Mills, & Schwander, 1987; Preminger et al., 2000; Stelmachowicz et al., 1994) have implemented optimization in two-dimensional parameter spaces and terminated them after three reversals. However, our three-dimensional implementation required a greater number of comparisons to be made compared to a two-dimensional version. Therefore adopting two reversals, rather than three, as a stopping rule also increased the likelihood that a listener would finish testing within the desired 18 iterations, rather than encountering a timeout, and would also reduce listener fatigue during the experiment. Pilot evaluations confirmed that two reversals were more effective than three for the purposes of the procedure.

6.2.1.5 Experimental implementation

The simplex procedure was written and administered using MATLAB on a Windows computer. The Windows computer was connected to a touchscreen monitor inside the sound booth and to the Linux MHA via an ethernet connection. This allowed listener judgments to trigger stimulus presentations and shaping iterations. The aided response was delivered via ER4p transducers and they were coupled to the participants' ears using foamtips.

Before experimental testing, listeners completed a practice run of non-adaptive paired comparisons also programmed with MATLAB and implemented via the openMHA. Their instructions were to listen to each of two stimuli and choose the one they preferred "as if it were their own hearing aid" without basing their judgment on loudness differences. The

practice run conditions were defined using the simplex parameters and consisted of six fixed and pre-defined stimulus pairs comparing prescribed frequency shaping (0,0,0) with either: a simulated low-pass filter (1,-4,-4) in which the low-frequency band gain was increased by 6 dB and the mid- and high-frequency band gains were decreased by 24 dB; or a simulated high-pass filter (-4,-4,1) in which the high-frequency band gain was increased by 6 dB and the mid- and low-frequency band gains were decreased by 24 dB. Listeners were expected to prefer the prescribed shaping and results of the practice round were examined for each listener. If the results were not consistent with the practice round expectations, then the listener was instructed to repeat the practice round.

During experimental testing, each listener completed two simplex runs for each of the four stimuli (male speech, female speech, pop music, and classical music), yielding a total of eight simplex runs. The direction of adjustment (increasing or decreasing gains) during the initial simplex iteration, and the ordering of parameter comparisons within each simplex iteration, were randomized. Presentation order of all eight runs was also randomized, except that simplex runs for the same stimulus did not occur in adjacent trials. Each listener's preferred frequency shaping was determined by calculating the average for each x-, y-, and z- coordinate across the two simplex runs for the same stimulus.

6.2.2 Results

Listeners completed an average of 9.8 iterations (SD = 5.3) per simplex run, corresponding to an average of 29.4 stimulus pairs. 9.8 iterations corresponded to an average 4.4 minutes (SD = 2.4) per simplex run, or 35.2 minutes of total simplex testing not including breaks. The number of iterations completed for each stimulus were within one SD of each other, indicating that the duration of testing was similar across stimuli. On average, listeners repeated 15% of the stimulus pairs per simplex run before finding the preferred frequency shaping.

6.2.2.1 Reliability

Reliability was assessed by measuring the Euclidean distance in step sizes between listeners' preferred shaping coordinates obtained in two simplex runs for the same stimulus. Distances were computed within each dimension and also across all three dimensions simultaneously. Within each dimension, the final coordinate from the first run was subtracted from the final coordinate from the second run. Across all dimensions, the three-dimensional distance was calculated between the final coordinate from the run first and the final coordinate from the second run by measuring the square root of the sum of squares of the differences across all three dimensions. Cumulative frequency-distribution curves representing the percentage of listeners whose distance between runs fell below a given number of step sizes are illustrated in Figure 6-4. This method was administered by Kuk & Pape (1992) for measuring the reliability of a two-dimensional simplex procedure with low- and high-frequency gain as the two dimensions. Reliability along each dimension and across all dimensions was also compared to cumulative frequency distribution curves for 1000 pairs of randomly selected preferred shaping coordinates. Randomly selected frequency distribution curves have been used to assess the reliability of adaptive three-dimensional multiparameter search strategies (Franck et al., 2004; Franck et al., 2007). In this study, randomly selected preferred coordinates were drawn from the range of coordinates within the 95% confidence interval of listeners' preferred coordinates for each dimension. The ranges corresponded to -2 to 5 for the low-frequency band, -4 to 3 for the mid-frequency band, and -6 to 4 for the high-frequency band.

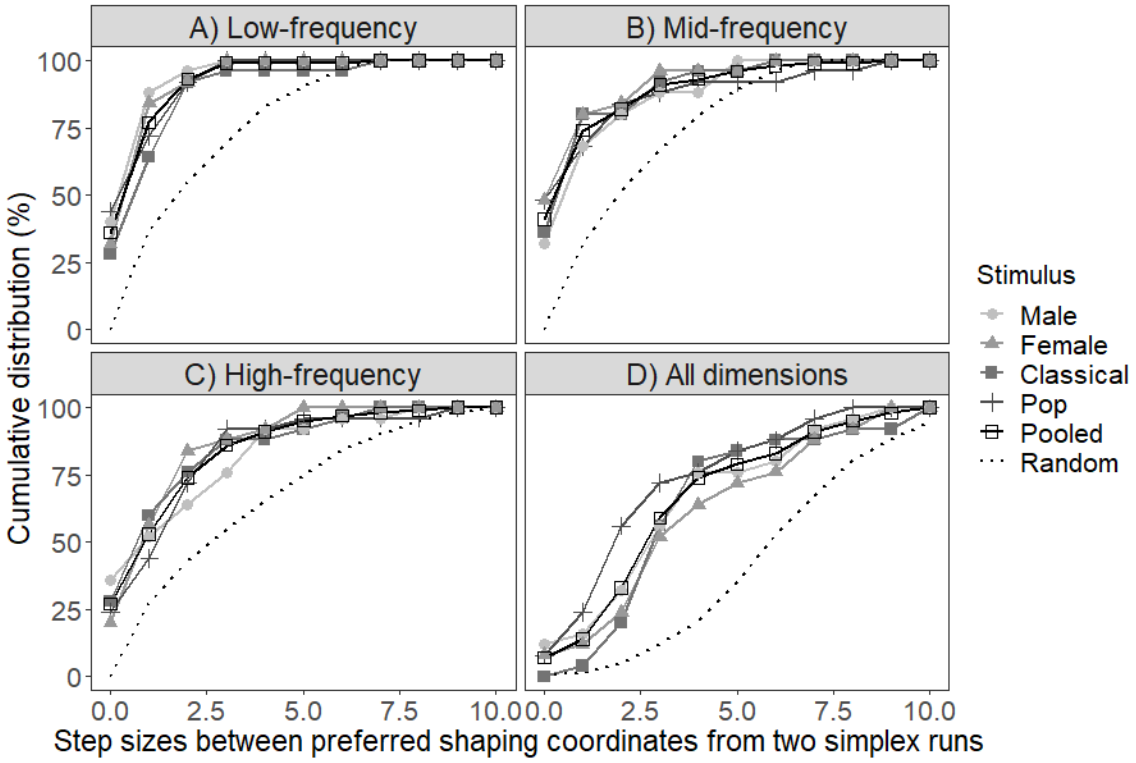

Figure 6-4: Cumulative frequency distribution curves illustrating the percentage of listeners (n = 25) who deviated by a given number of step sizes between preferred shaping coordinates from two simplex runs for the low-frequency dimension (A panel), mid-frequency dimension (B panel), high-frequency dimension (C panel), and Euclidean distance across all three dimensions (D panel). Curves were calculated for each stimulus, and for all stimuli pooled. The dotted curve illustrates the cumulative frequency distribution for randomly selected preferred shaping coordinates over 1000 test-retest simulations.

Figure 6-4 illustrates that listeners' frequency distribution curves, for both one- and three-dimensional distances, fell above and to the left of the random frequency distribution curve, indicating that listener decisions were more reliable than random performance. Figure 6-4 also illustrates that 7% of listeners selected the same preferred shaping coordinates across all stimuli and all dimensions (bottom right). When examining Figure 6-4 per dimension, 36% of listeners selected the same low-frequency coordinate (top left), 41% of listeners selected the same mid-frequency coordinate (top right), and 27% of listeners selected the same high-frequency coordinate (bottom left). For the low-frequency dimension, 93% of listeners were within two step sizes between simplex runs. For the mid-frequency dimension, 91% of listeners were within three step sizes between simplex runs. For the high-frequency dimension, 91% of listeners were within four step sizes between simplex runs. Across all dimensions, 91% of listeners were within seven three-dimensional step sizes between simplex runs. When examining Figure 6-4 per stimulus within each dimension, each stimulus had a frequency distribution curve that was similar to the pooled frequency distribution curve across stimuli. However, when examining Figure 6-4 per stimulus across all dimensions, more listeners selected frequency shaping coordinates that were within three step sizes between simplex runs for the pop stimulus compared to the other stimuli.

6.2.2.2 Observed differences from prescribed shaping

Listeners' preferred low-, mid-, and high- frequency-gain parameters per stimulus were assessed by measuring the observed differences in dB between their individually preferred shaping (averaged between simplex runs) and prescribed shaping for each parameter. However, the simplex procedure did not produce dB difference values directly. Rather, the simplex procedure produced sets of coordinates representing the distance between prescribed and preferred shaping along the x, y, and z dimensions for each simplex run. To determine the observed difference in dB between preferred and prescribed shaping each listener's mean coordinates per stimulus were programmed into the openMHA such that the openMHA recreated the stimulus-specific preferred shaping. The output of the openMHA was then measured by recording the output of the ER4p in the Verifit2 0.4 cc coupler as a waveform. The shaping for each stimulus was measured using the ISTS, and

not each respective stimulus, as the test signal so that the results would be relative to a standard stimulus; this allowed the shaping differences to be compared directly. A recording of the ISTS under prescribed DSL v5.0 gains at a simulated input level of 70 dB SPL was also measured as the reference condition. The Verifit2 produced frequency-specific (.25, .5, .75, 1, 2, 3, 4, 6, 8 kHz) simulated real ear response spectral values in dB SPL for each waveform, which were used to generate the aided spectra illustrated in Figure 6-5a.

Observed differences from prescribed shaping, illustrated in Figure 6-5b, were measured by calculating the difference between the preferred shaping and prescribed shaping dB at octave and interoctave frequencies (.25, .5, .75, 1, 2, 3, 4, 6, 8 kHz) and then averaged within the low- (0.1-0.8 kHz), mid- (1-2.5 kHz), and high- (3-10 kHz) frequency bands for statistical analysis. A 2x4x3 mixed analysis of variance (ANOVA) was used to test the effect of status (user/non-user) x stimulus (male speech/female speech/pop music/classical music) x frequency band (low-frequency/mid-frequency/high-frequency) on observed differences (dependent variable). Greenhouse-Geisser corrections were applied to adjust for departures from sphericity. Post-hoc contrasts were performed, when appropriate, using the Holm correction. Statistical analyses were completed using RStudio (Version 1.0.132; R Core Team, 2017) and ez package (Lawrence, 2016).

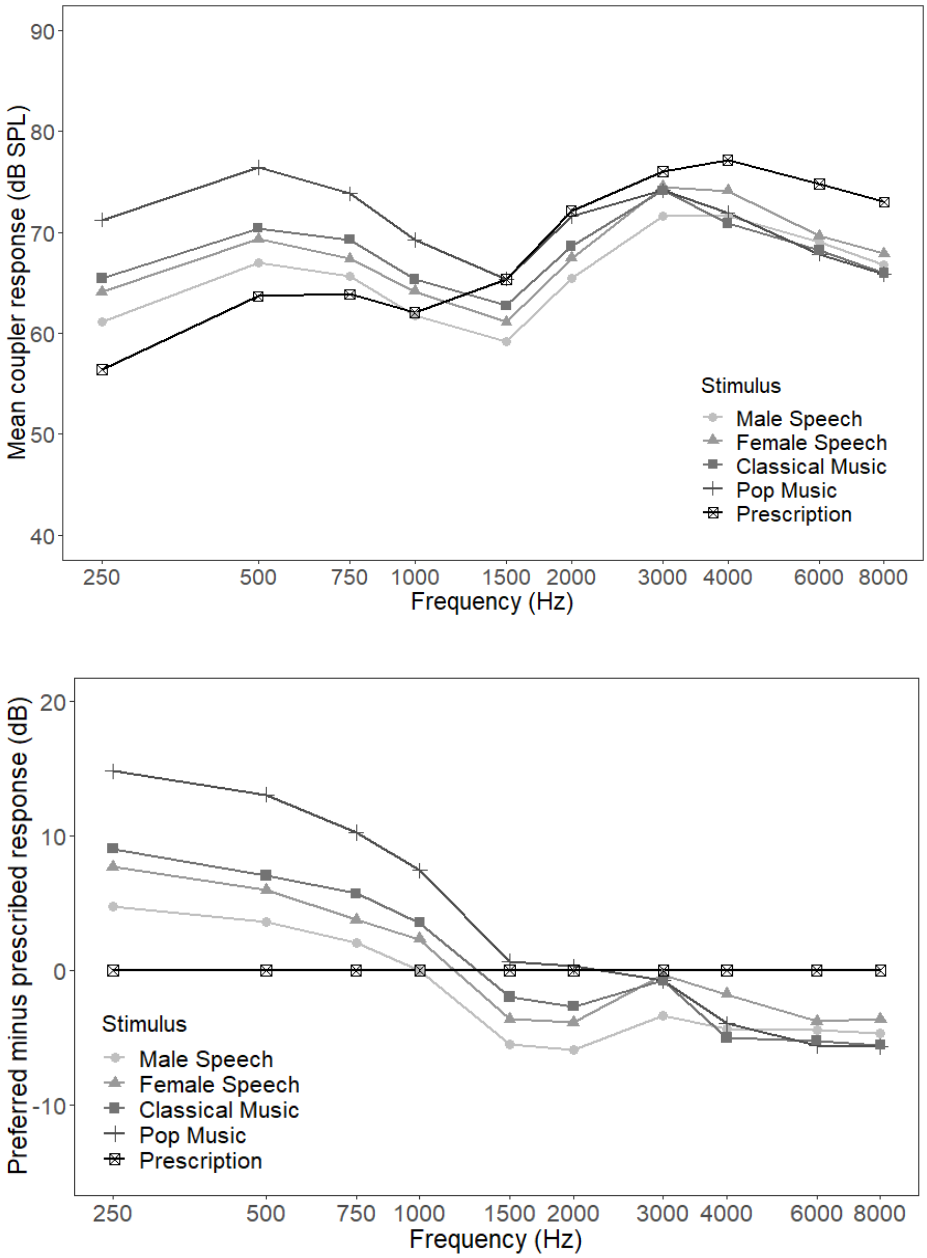

Figure 6-5: A (above): Mean coupler responses of the International Speech Test Signal (ISTS; Holube, Fredelake, Vlaming, & Kollmeier, 2010) across listeners and runs at a simulated 70 dB SPL input level as measured via the prescribed and the preferred frequency shaping derived using each stimulus. B (below) Mean observed differences from prescribed gains in dB, measured as the mean difference (dB) across listeners and runs between the preferred and prescribed shaping at audiometric frequencies for a 70 dB SPL input level.

The average observed differences from individuals' prescribed shaping are illustrated in Figure 6-6. The mixed ANOVA revealed significant main effects of stimulus ($F_{(2.27,52.29)}$ =7.14, $p<0.001$, $\eta^2=0.04$) and frequency band ($F_{(1.46,33.55)}$=31.00, $p<0.0001$, $\eta^2=0.23$). There was also a significant interaction of stimulus by frequency ($F_{(3.75,86.26)}$=4.93, $p<0.01$, $\eta^2=0.03$). The effects of user, user by stimulus, user by parameter adjustment, and user by stimulus by frequency band were all non-significant, suggesting that user status was not associated with gain adjustments.

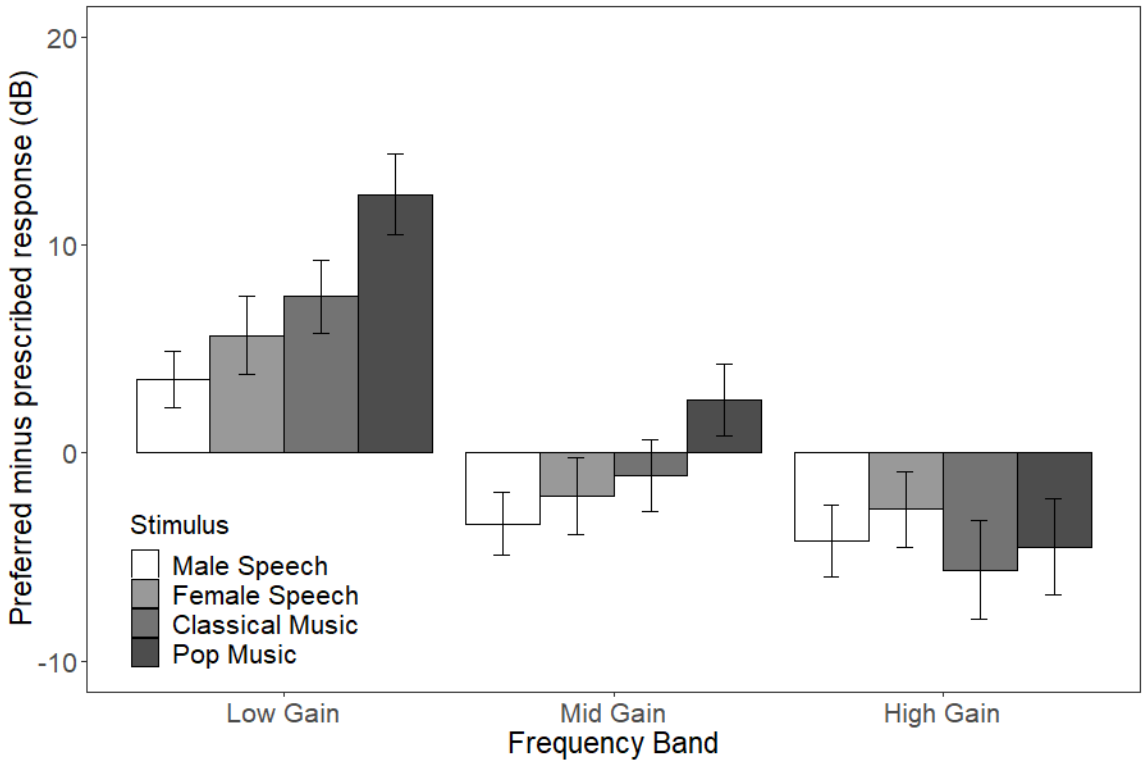

Figure 6-6: Mean observed differences from prescribed gains, measured as the mean difference (dB) between the preferred and prescribed shaping averaged across the low (0.1-0.8 kHz), mid (1-2.5 kHz) and high (3-10 kHz) frequency bands for a 70 dB SPL input level across listeners and runs. Errors bars represent one standard error of the mean.

The sources of the main effect of stimulus occurred between the pop music and male speech stimuli. Across listeners and frequency bands, the pop stimulus was adjusted on average 4.7 dB higher than the male speech stimulus ($p < 0.001$). The pop stimulus was also adjusted 3.1 dB higher than both the classical music ($p = 0.08$) and female speech stimuli ($p = 0.08$), though neither difference reached statistical significance at the 0.05 alpha level. The remaining contrasts were all 1.7 dB or less.

The sources of the main effect of frequency band occurred between low-frequency gain adjustments and the other bands. Descriptively, low-frequency gain was increased by a mean of 7.3 dB above individuals' prescribed gains, and low-frequency gain was adjusted 8.3 dB and 11.6 dB higher than mid- ($p < 0.0001$) and high-frequency gain ($p < 0.0001$), respectively. High-frequency gain was decreased by a mean of 4.27 dB relative to prescribed gains, which was 3.27 dB less than the 1.0-dB mid-frequency gain decrease, although this difference was not statistically significant ($p=0.55$).

The sources of the interaction of stimulus x parameter adjustment are illustrated in Table 6-1. Post-hoc comparisons within each frequency band revealed that gains for pop-music were adjusted higher than for other stimuli than other stimulus pair comparisons. The low-frequency gain was increased more for pop music compared to classical music ($p < 0.05$), female speech ($p < 0.001$), and female speech ($p < 0.0001$). The mid-frequency gain was increased more for pop music versus male speech ($p < 0.01$).

Table 6-1: Post-hoc simple effects for the interaction of stimulus X frequency band. The difference (dB) column is a difference measure of stimulus 1 relative to stimulus 2. Asterisks indicate statistical significance (p<0.05)

Parameter Adjustment	Stimulus 1	Stimulus 2	Difference (dB)	p-value
Low-frequency gain	Male	Female	-2.09	1.00
	Male	Pop	-8.88*	<0.0001
	Male	Classical	-3.95	0.23
	Female	Pop	-6.79*	<0.001
	Female	Classical	-1.86	1.00
	Pop	Classical	4.92*	<0.05
Mid-frequency gain	Male	Female	-1.32	1.00
	Male	Pop	-5.89*	<0.01
	Male	Classical	-2.29	1.00
	Female	Pop	-4.57	0.08
	Female	Classical	-0.97	1.00
	Pop	Classical	3.60	0.39
High-frequency gain	Male	Female	-1.61	1.00
	Male	Pop	0.41	1.00
	Male	Classical	1.38	1.00
	Female	Pop	2.03	1.00
	Female	Classical	2.99	0.85
	Pop	Classical	0.97	1.00

6.2.3 Discussion

Experiment One investigated whether listeners adjusted hearing aid frequency responses away from prescribed shaping for different music and speech stimuli using the simplex procedure, and if they did so consistently. In general, listeners increased low-frequency gain by the greatest magnitude, followed by a high-frequency decrease of a smaller magnitude, followed by minimal adjustments to the mid-frequency gain. On average, listeners increased the low-frequency gain by 9.99 dB across music stimuli and by 4.60 dB across speech stimuli and decreased the high-frequency gain to 4.28 dB below the prescribed targets across all stimuli. The low-frequency gain was increased most substantially for the pop music, followed by classical music, followed by the female and male speech passages. The mid-frequency gain was increased for pop music, and decreased for the remaining stimuli, but mean adjustments were within 3 dB of the prescribed targets across all stimuli.

6.2.3.1 Reliability

The reliability of listener performance using the simplex procedure was better than the reliability of randomly selected preferred coordinates. In Figure 6-4, the cumulative frequency-distribution curves within each dimension and across all dimensions clearly deviated to the upper-left of the random cumulative frequency distribution curves, suggesting that there were fewer step sizes between listener's preferred coordinates than random preferred coordinates. This implies that listener judgments were more reliable than randomly selected coordinates. The reliability of the simplex procedure was also moderately related to the parameters and stimuli used in this study. Within each dimension, listeners selected low-frequency gain adjustments with the most consistency, followed by mid-frequency gain adjustments, followed by high-frequency gain adjustments with the least consistency. This pattern of consistency may reflect the fact that listeners' lowest and highest thresholds were in the low frequencies and highest frequencies, respectively. Listeners may have adjusted the low-frequency gain more consistently than high-frequency gain because they had better hearing in that frequency range. Across dimensions, listener gain adjustments were similar for all stimuli, except for pop music, which yielded more similar gain adjustments between simplex runs compared to the other stimuli. The pop

stimulus had a more even distribution of spectral content relative to the other stimuli. It is possible that listeners were more sensitive to the interaction between frequency-band adjustments and spectral content in the pop stimulus.

Aspects of the reliability of the three-dimensional simplex procedure are similar to past simplex procedures. In the current study, between 25% to 50% of listeners selected the same frequency gain adjustment within each dimension between simplex runs across all stimuli. Figures 5 and 6 in Kuk & Pape, (1992) show that a similar percentage of listeners selected the same frequency gain adjustment within each dimension across listening conditions. However, 7% of listeners in the current study selected the same preferred shaping (i.e. the same gain adjustment simultaneously in all three dimensions) between simplex runs, which contrasts the 30% of listeners in Kuk & Pape (1992) who selected the same parameters between simplex runs. Furthermore, 14% of listeners in the current study selected preferred shaping that was within one three-dimensional step size between simplex runs, which contrasts the 80% of listeners in Kuk & Pape (1992) who selected parameters within one cell between simplex runs.

The fact that listeners were more consistent in the study by Kuk & Pape (1992) likely reflects the differences between a two-dimensional and a three-dimensional simplex procedure and differences between studies in how frequency responses were modified. First, any uncertainty in participants' paired-comparison responses would lead to a lower probability of selecting the same shaping between runs in a three-dimensional design than in a two-dimensional design. As a demonstration of this, consider one run in which the participant selected the truly optimal frequency shaping, and a second run in which the simplex procedure reaches that shaping before stopping. To select the same final optimum coordinate in the second run, a listener would need to prefer that cell to each of its adjacent cells. In a two-dimensional design, there are four testable cells adjacent to each coordinate, and in a three-dimensional design there are six. If there is some probability (p) that the listener chooses the optimum cell in each comparison with adjacent cells, then the overall probability of selecting the same optimum is p^4 in two dimensions but p^6 in three dimensions. If the listener merely guessed for each paired-comparison (p=1/2), the overall probability of making the same four selections would be $(1/2)^4$, or 6.25%, in two

dimensions, but only $(1/2)^6$, or 1.56%, in three dimensions. Since listeners were not guessing (as indicated by Figure 6-4), p must have been greater than ½. For example, if the probability of choosing the optimum correctly was 0.75, there would be a 31.6% probability of finishing on the same preferred cell for a 2D simplex design, but only 17.8% for a 3D design. Second, Kuk & Pape (1992) varied the low- and high-frequency parameters by manipulating the cut-off for both frequency ranges. In the current study, however, the low- and high-frequency cut-offs were fixed due to the channel structure of the openMHA, and the level of each frequency band was increased or decreased. Listeners may have been more sensitive to perceptual changes in cut-off compared to perceptual changes in level. The impact of factors such as number of dimensions, parameter manipulation, and stimulus on consistency requires further evaluation.

6.2.3.2 Observed differences from prescribed shaping

The finding that listeners on average increased the low-frequency gain and decreased the high-frequency gain is consistent with past literature when listening preference was the criterion. Hearing aid users have reported insufficient low-frequency gain and/or excessive high-frequency gain for music-listening (Madsen & Moore, 2014) and so it seems appropriate that listeners adjusted the gain in these directions. Previous studies on music-listening and gain adjustments have shown that listeners preferred additional low-frequency gain compared to stimuli shaped with less low-frequency gain (Franks, 1982; Punch, 1978). Furthermore, an extended low-frequency response has been recommended for special hearing aid music programs (Moore, 2016). Past speech-based studies have also found that some listeners prefer more low-frequency gain and less high-frequency gain for speech quality preferences relative to prescribed NAL-based fittings (Kuk & Pape, 1992, 1993; Nelson et al., 2018; Preminger et al., 2000).

The finding that the mid-frequency range was the least adjusted parameter supports the importance of audibility in that region for speech intelligibility (Pavlovic, 1994). Hearing

aid prescriptions aim to provide enough audibility so that listeners can find speech intelligible. The amount of gain prescribed for that region was more satisfying for listeners than the amount of gain for either of the other regions, supporting the use of the prescribed gains applied in this study, even when overall preference was the criterion. However, the gain adjustments in the mid-frequency region were statistically comparable to the gain adjustments in the high-frequency region. This observation suggests that a two-dimensional simplex with low- and high-frequency gain may be sufficient for studies of this nature, or that a more advanced signal-processing feature could be substituted as a third signal processing parameter in future three-dimensional simplex procedures (in addition to low- and high-frequency gain).

6.2.3.3 Stimulus dependencies

The finding that there were different preferences for music compared to speech may be due to the role of listening criterion. Recall that in the current study, listeners were instructed to select the preferred hearing aid "as if it were their own". It is possible that listeners wanted to optimize sound quality, but without degrading speech intelligibility. The study by Kuk & Pape (1992) required listeners to complete the simplex based on clarity judgments or consonant recognition. The preferred frequency responses for the consonant recognition tasks consisted of more high-frequency gain compared to those for the clarity judgments. Similarly, the "Goldilocks" self-adjustment strategy incorporated a speech perception test, and listeners ultimately self-adjusted to have less low-frequency and more high-frequency gain compared to their initial estimate following the speech task (Boothroyd & Mackersie, 2017). Preminger & Van Tasell (1995) found that as long as speech intelligibility was ideal, sound quality ratings could vary. However, if speech became less intelligible, then the pattern of sound quality ratings would follow ratings of intelligibility. In Experiment One, it is possible that listeners conducted quality-based adjustments for speech, but only until intelligibility became compromised. Adjustments with excessive low-frequency gain may have enhanced upward spread of masking, which could have been detrimental for speech intelligibility. Similarly, reductions to the high-frequency gain may have stopped if intelligibility was affected. This may have also explained the smaller low-frequency gain increases for speech versus music, and for male

speech (which contained more low-frequency content leading to a greater upward spread of masking) versus female speech (which contained less relative low-frequency content).

Differences in the acoustic content between stimuli may have also driven stimulus-specific gain adjustments. In other words, the same parameter adjustment may have had a different perceptual impact due to differences in acoustic content between the signals. Davies-Venn et al. (2007) found that listeners rated popular music as sharper compared to classical music because the popular music contained more high-frequency energy. The pop music sample in the current study contained more high-frequency content compared to the classical sample, and so it is possible that listeners increased the low-frequency gain to distribute energy evenly over the entire spectrum. Arehart et al. (2011) also found that different genres interacted differently with various types of hearing aid distortions, again indicating that acoustic differences between stimuli will have different perceptual consequences as they interact with the same parameter adjustments.

Stimulus-specific gain adjustments, particularly in the low-frequency band may have been related to the amount of the acoustic energy present in the stimulus. An examination of Figure 6-2 shows that the pop contains spectral energy that is the most evenly distributed across frequencies compared to the other stimuli. In contrast, the male speech stimulus is mostly dominated by low-frequency content. As a percentage of total frequency content, 54% of the pop stimulus, 82% of the classical stimulus, 79% of the female speech stimulus, and 94% of the male speech stimulus was dominated by low-frequency content (in the 100-800 Hz range). The fact that low-frequency gain was increased most substantially for the pop genre and least substantially for the male speech stimulus suggests that the relative contribution of low-frequency content may be related to the degree to which listeners will adjust the low-frequency gain. That is, listeners may increase the low-frequency gain more if there is less relative contribution of low frequencies towards the overall spectral energy in the stimulus. Future research is needed to understand the relationship between acoustic features and self-adjustment preferences.

6.3 Experiment Two

The question of whether preferred shaping, as determined by the simplex, produced meaningfully different outcomes relative to prescribed shaping remains. This was the central question of Experiment Two. Past studies have investigated field-based listener outcomes between prescribed and simplex-selected gains on the basis of overall satisfaction (Kuk & Pape, 1992) and intelligibility (Preminger et al., 2000), revealing that only some listeners benefit from the simplex-selected settings. This study also investigated listener outcomes but did so in the laboratory rather than in the field. Hearing-aid-processed music can be measured subjectively using a sound quality task, often for multiple descriptors such as "sharpness" and "fullness" (Gabrielsson, Schenkman, & Hagerman, 1988). Some of these descriptors correspond to level differences between frequency regions addressed in Experiment One. Therefore, Experiment Two focused on whether listeners' preferred shaping produced meaningfully different sound quality ratings compared to prescribed shaping, which would also inform what descriptors may have driven listener judgments when overall preference was the criterion in Experiment One. A secondary objective addressed whether stimulus-specific shaping would produce sound quality effects that generalized to other spectrally-similar stimuli that were not used in Experiment One.

6.3.1 Methods

6.3.1.1 Participants

All participants who participated in Experiment One also participated in Experiment Two. Testing was completed during the same session.

6.3.1.2 Test materials

The test materials consisted of the speech and music passages from Experiment One, referred to here as "training" stimuli, as well as new speech and music passages that were spectrally-similar to their genre-matched counterparts, referred to here as "validation" stimuli. The long-term spectra of the training and validation stimuli are illustrated in Figure 6-2.

The IEEE training sentences included the male-spoken, “Raise the sail and steer the ship northwards. A cone costs five cents on Mondays”, and female-spoken “Would you please give us the facts? He arrived home every other night”. The Beatles and Mozart music passages were 8.7 and 5.1 seconds, respectively, in duration. The validation speech passages were the male-spoken, “The ramp led up to the wide highway. Beat the dust from the rug onto the lawn” and female-spoken, “They could laugh, although they were sad. Farmers came in to thresh the oat crop” IEEE sentences. The talker for each gender differed between the training and validation stimuli. The validation music passages were downloaded from iTunes and included a 6.4 second clip of “New Orleans is Sinking” by The Tragically Hip for the pop genre and a 7.3 second clip of Beethoven’s “String Quartet No. 4 in C Minor, Op. 18: III. Menuetto: Allegretto” by the Emperor String Quartet for the classical string genre. Stimulus duration was selected so that each sample consisted of a full musical phrase. Stimulus duration was longer in Experiment Two compared to Experiment One. In Experiment One, listeners were required to listener to the entire sample. In Experiment Two, the stimulus was looped until the listener was ready to finalize a rating. Digital scaling and calibration of Experiment Two test materials followed the same procedure described in Experiment One.

6.3.1.3 Sound quality ratings

Participants rated several dimensions of sound quality for each stimulus. Participants rated two versions of each stimulus: one processed using their prescribed frequency shaping and one processed using their preferred frequency shaping determined using the simplex procedure from Experiment One. The sound quality dimensions were adapted from those used by Gabrielsson, Schenkman, & Hagerman (1988) and consisted of “Overall Impression”, “Loudness”, “Fullness”, “Sharpness”, and “Intelligibility” (for speech only). Participants rated each dimension using a continuous horizontal scroll-bar with five descriptors from lowest to highest, which produced a number from 0 (lowest)-10 (highest). The participants were blind to the numerical rating. The descriptors, from lowest to highest, for “Overall Impression” were: 'Very Bad', 'Rather Bad', 'Midway', 'Rather Good' and 'Very Good'. The descriptors for “Loudness” were 'Very Soft', 'Rather Soft', 'Midway', 'Rather Loud' and 'Very Loud'. The descriptors for 'Fullness' were 'Very Thin', 'Rather Thin',

'Midway', 'Rather Full' and 'Very Full'. The descriptors for “Sharpness” were 'Very Gentle', 'Rather Gentle', 'Midway', 'Rather Shrill' and 'Very Shrill'. The descriptors for “Intelligibility” were 'Very Unclear', 'Rather Unclear', 'Midway', 'Rather Clear' and 'Very Clear'.

The sound quality rating procedure was written and administered in MATLAB and used the same hardware setup as Experiment One. Overall, there were four stimulus categories (male speech, female speech, pop music, and classic music) and each consisted of a training and validation version. The training version was the same stimulus used to determine the preferred shaping, and the validation stimulus was the stimulus for which the preferred shaping for that category was applied. Each of the stimuli were processed using each participant’s prescribed and preferred shaping. This yielded a total of 16 stimuli to be rated (four categories x two training/validation x two prescribed/preferred shaping). A total of four descriptors for each music stimulus and five descriptors for each speech stimulus meant that a total of 144 ratings were completed. Stimulus order, as well as descriptor order within each stimulus, was randomized. Each stimulus was presented at a simulated hearing aid input level of 70 dB SPL.

6.3.2 Results

Sound quality ratings are illustrated in Figure 6-7 and grouped by descriptor. In order to choose an appropriate analysis, we considered the objective, which was to measure whether listeners were sensitive (as measured by changes in descriptive ratings) to differences between their preferred shaping from Experiment One, and whether preferences would generalize to other stimuli. The objective was not focused on differences in ratings between stimuli, partly because different stimuli were rated using different shaping. Therefore, a series of two (training/validation) x two (prescribed/preferred) repeated measures ANOVAs were used to analyze ratings independently for each stimulus. Post-hoc contrasts were performed when appropriate using the Holm correction for main effects of prescribed/prescribed, main effects of training/validation, and interactions of prescribed/preferred within each level of training or validation.

It was also of interest to determine which descriptors were most strongly associated with each listener's preferred shaping in Experiment One. Therefore, the "Overall Impression" descriptor was used to characterize listeners' preferences between stimuli in Experiment One and Pearson correlation coefficients were calculated between ratings for each descriptor and ratings of overall impression across individuals. Because speech preferences may have been influenced by intelligibility in a way that music would not be, correlations were calculated across all speech stimuli and across all music stimuli separately.

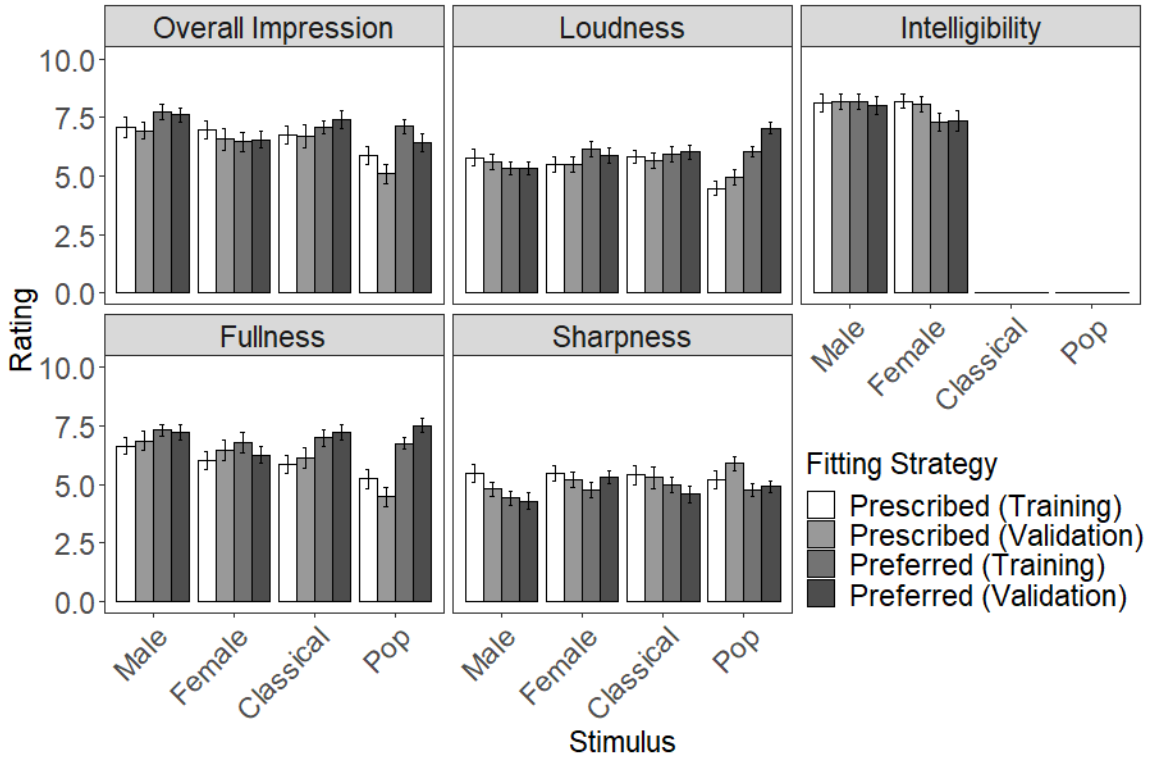

Figure 6-7: Mean sound quality descriptor ratings measured as a function of stimulus status (training/validation), shaping (prescribed/preferred), and stimulus (male speech, female speech, classic music, and pop music) across listeners. Each panel shows ratings for one descriptive dimension. Participants did not rate intelligibility for music. Error bars represent one standard error of the mean.

6.3.2.1 Overall impression

For overall impression, preferred shaping was rated higher than the prescribed shaping for both male speech ($F_{(1,24)}$= 4.36, $p<0.05$, η^2=0.03) and pop music ($F_{(1,24)}$= 16.76, $p<0.001$, η^2=0.11). The pop music training stimulus was also rated higher than the pop music validation stimulus ($F_{(1,24)}$=6.07, $p<0.05$, η^2=0.11). There were no statistical effects of overall impression for classical music or female speech. Descriptively, ratings of overall impression increased from the prescribed to preferred shaping, except for female speech.

6.3.2.2 Loudness

For loudness, the preferred shaping was rated louder than the prescribed shaping for pop music ($F_{(1,24)}$= 24.05, $p<0.0001$ η^2=0.30), as was the validation stimulus versus training stimulus ($F_{(1,24)}$= 10.05, $p<0.01$, η^2=0.07). There were no statistical effects of loudness for the other stimuli. Descriptively, loudness ratings were comparable between the preferred and prescribed shaping for all other stimuli.

6.3.2.3 Fullness

For fullness, the preferred shaping was rated fuller than the prescribed shaping for both pop ($F_{(1,24)}$= 39.59, $p<0.0001$, η^2=0.30) and classical music ($F_{(1,24)}$= 12.15, $p<0.01$ η^2=0.09). There were no other statistical effects of interest for the music stimuli, nor any statistical effects for male or female speech. Descriptively, fullness ratings increased from the prescribed to preferred shaping, except for female speech.

6.3.2.4 Sharpness

For sharpness, the preferred shaping was rated less sharp than the prescribed shaping for male speech ($F_{(1,24)}$= 4.79, $p<0.05$, η^2=0.05) as was the validation stimulus versus training stimulus ($F_{(1,24)}$= 4.78, $p<0.05$, η^2=0.01). There were no statistical effects of sharpness for the other stimuli. Descriptively, sharpness ratings decreased from the prescribed to preferred shaping, except for female speech.

6.3.2.5 Intelligibility

For intelligibility, the prescribed shaping was rated as more intelligible compared to the preferred shaping for female speech ($F_{(1,24)}$=9.14, p<0.01, η^2=0.05). There were no statistical effects of intelligibility for male speech.

6.3.2.6 Relationship between sound quality descriptors and overall impression

The correlation coefficients describing the relationships between sound quality descriptors and ratings of overall impression are listed in Table 6-2. For music stimuli, ratings of fullness and sharpness were positively and negatively associated with ratings overall impression, respectively, with fullness being more strongly associated. Correlations for loudness were negligible. For speech stimuli, ratings of intelligibility were positively and most strongly related with ratings of overall impression. Ratings of fullness and sharpness followed the same trend for speech as they did for music, although the associations were less strong. Ratings of loudness were negatively, and least associated with ratings of overall impression.

*Table 6-2: Pearson correlation coefficients between sound quality descriptors and ratings of overall impressions across individuals. * = p < 0.0001.*

Descriptor	Stimulus	
	Music	Speech
Loudness	-0.05	-0.31
Fullness	0.57*	0.42*
Sharpness	-0.46*	-034*
Intelligibility	---	0.50*

6.3.3 Discussion

The results of Experiment Two suggest that listeners, on average, preferred the music stimuli and male speech stimuli processed under the preferred frequency shaping compared to prescribed frequency shaping. Most sound quality ratings for preferred shaping generalized to spectrally similar stimuli, which were not used in determining the preferred shaping. These laboratory findings are not entirely consistent with field-based self-adjustment studies, in which listeners showed mixed preferences between preferred and prescribed frequency shaping (Keidser & Alamudi, 2013; Kuk & Pape, 1993; Preminger et al., 2000). The fact that fullness was most strongly associated with ratings of total impression suggests that prescribed gains lack low-frequency gain, and that listeners will increase low-frequency gain when adjusting towards a preferred frequency shaping.

6.3.3.1 Sound quality descriptor ratings

Ratings of fullness increased significantly from prescribed to preferred shaping for the music stimuli, but this was not significant for the speech stimuli. Increased ratings of fullness correspond to more relative emphasis in the low-frequency region (Gabrielsson & Sjögren, 1979), and this was consistent with the observed preferred differences from prescribed shaping from Experiment One. Fullness ratings were positively and most strongly related to ratings of overall impression for music, and fullness was the second-most strongly related descriptor for speech (after intelligibility), which is consistent with the relationships between ratings of fullness and overall impression in past literature (Davies-Venn et al., 2007; Gabrielsson et al., 1988). The strength in this relationship is likely related to the finding that the largest gain difference between prescribed and preferred frequency shaping was in the low-frequency region.

The finding that loudness increased significantly from prescribed to preferred frequency shaping for pop music can be explained by that stimulus's acoustic properties and preferred shaping. Pop music contained more low-frequency energy compared to the other stimuli, as well as the largest low-frequency gain increase between frequency responses. Low-frequency energy (below 1 kHz) is considered to contribute more to perceived loudness than high-frequency energy (Keidser et al., 2002), and likely explains this effect. The low-

frequency gain adjustments for the other stimuli were substantially lower, suggesting that loudness differences between prescribed and preferred responses may have been less noticeable. This would explain why the other stimuli's preferred shaping did not yield different loudness ratings, and likely also why loudness ratings were least associated with ratings of overall impression.

Ratings of sharpness descriptively decreased from prescribed to preferred shaping for almost all stimuli. Decreased ratings of sharpness occur due to less relative emphasis in the high-frequency region (Gabrielsson & Sjögren, 1979), and this was consistent with the differences observed between responses. However, these differences failed to reach statistical significance for almost all stimuli. The mean high-frequency gain decrease was -4.3 dB, which was smaller than the minimum of ±5 dB recommended to produce a meaningful perceptual difference for speech-shaped noise (Caswell-Midwinter & Whitmer, 2019). Despite the lack of significance, sharpness ratings were negatively related to ratings of overall impression for both speech and music, and were the most associated descriptor after fullness. The strength in this relationship was likely related to the finding that the second-largest gain adjustment occurred in the high-frequency region. Gabrielsson et al. (1988) found that sharpness was negatively related to overall impression in normal hearing listeners but not hearing-impaired listeners. This discrepancy may have occurred because the hearing-impaired listeners in their study presented with poorer high-frequency thresholds compared to this study's listeners. Their listeners may have encountered challenges distinguishing stimuli with different high-frequency content, especially if there was insufficient audibility in that region. Similarly, Davies-Venn et al. (2007) did not find a strong relationship between ratings of sharpness and overall impression, which may have been confounded by audibility issues due to more elevated high-frequency thresholds. However, their experimental conditions did not systematically probe differences in high-frequency content, suggesting that listeners simply may not have perceived sharpness differences between stimulus-processing conditions. This latter explanation may be more plausible, as the authors attributed a significant main effect of genre for sharpness to differences between high-frequency content between popular and classical music.

Ratings of intelligibility significantly decreased from the prescribed to the preferred shaping for female speech and were most strongly associated with ratings of overall impression across speech stimuli. Intelligibility has previously been the descriptor most related to overall impression (Davies-Venn et al., 2007). This supports the hypothesis from Experiment One, in that intelligibility may have influenced overall preferences for speech, which in turn yielded smaller gain adjustments compared to music. The acoustic differences between male and female speech may have explained the female-speech-specific intelligibility reduction. The female voice contained more high-frequency energy, which if reduced (as indicated by lower ratings of sharpness), would yield lower ratings of intelligibility. For example, during the simplex paired comparisons, one of the participants anecdotally commented that the plurality of the word "facts" changed between the stimuli. Plurality is attributed to the word-final fricative sound /s/ which is predominantly composed of high-frequency energy above 4 kHz (Glista & Scollie, 2012). The participant's anecdote presumably occurred during a stimulus pairing in which the interaction between high-frequency energy of /s/ and participant high-frequency thresholds rendered the /s/ inaudible in one of the conditions (Stelmachowicz, Pittman, Hoover, Lewis, & Moeller, 2004). Intelligibility may have also been impacted by an increased upward spread of masking due to greater low-frequency gain in the preferred condition. The association between prescribed shaping and improved intelligibility supports the use of hearing aid prescriptions if speech communication is the fitting goal.

6.3.3.2 Generalizability to validation stimuli

The fact that preferred shaping yielded comparable ratings across most of the training and validation stimuli suggests that preferred shaping trained on one stimulus will have a similar impact on other stimuli sharing a similar spectral profile. There were, however, differences between pop music stimuli for ratings of overall impression and loudness, and for male speech stimuli for ratings of sharpness. With respect to overall impression, it is possible that listeners simply preferred listening to the training stimulus rather than to the validation stimulus. With respect to loudness, the validation stimulus contained more low-frequency content, which could have potentially yielded higher ratings of loudness. With respect to sharpness, the male speech validation stimulus contained less high-frequency (3-

10 kHz) content and was rated as less sharp. As determined in Experiment One, acoustic differences between stimuli were likely responsible for determining different stimulus-dependent frequency shaping. It is possible that, similarly, acoustic differences between the spectrally-similar samples belonging to the same genre in Experiment Two interacted differently to produce different ratings between the training and validation stimuli.

6.3.3.3 Listener considerations

The data presented here reflected findings averaged across individuals. Even though listeners preferred a mean low-frequency gain increase across stimuli relative to prescribed gains, 12% of listeners preferred a low-frequency gain decrease. Likewise, 40% and 32% of listeners preferred a mid- and high-frequency gain increase, despite those ranges exhibiting mean decreases relative to prescribed gains. Preferences may be impacted by individual factors such as experience, cognition, and loudness discomfort levels. Therefore, it is important to consider the findings from this article in tandem with individual factors if implementing these findings in a clinical context.

The pursuit of threshold-matched user and non-user groups may have also influenced the lack of statistical effects between groups for gain adjustments. Mackersie et al. (2018) compared gain adjustments between users and non-users following a multiparameter self-adjustment strategy and found that users increased high-frequency output relatively more compared to non-users. While thresholds between their groups were statistically comparable, their user group presented with more elevated high-frequency thresholds (8 kHz mean = 75 dB HL) compared to non-users (8 kHz mean = 60 dB HL). This difference contrasts with the current study, in which mean thresholds at 8 kHz for both groups were 55 dB HL. Furthermore, the current study's user group had more elevated (though not significantly different) low-frequency thresholds, but they did not adjust low-frequency gain differently compared to the non-user group. An alternative approach to explore differences in gain adjustments would be to explore the role of hearing aid experience. Differences in gain adjustments between experienced and inexperienced users for speech tend to be minimal (Convery, Keidser, & Dillon, 2005), although inexperienced users with mild loss tend to prefer less gain compared to experienced users (Keidser, O'Brien, Carter,

McLelland, & Yeend, 2008). Future research may probe interactions between length of experience, hearing loss severity and differences between speech and music.

6.3.3.4 Acoustic considerations

The low-frequency gain adjustments in the current article were achieved using a fully-occluded transducer. A fully-occluded transducer, rather than an open fitting, was ideal because it did not allow for the leakage of low-frequency content from the ear canal, thus allowing substantial upwards adjustments of low-frequency gain (Ricketts, Bentler, & Mueller, 2019). However, most of the participants included in the study had normal or near-normal low-frequency thresholds and would either be candidates for or were already users of open-fit receiver-in-the-ear devices. Therefore listeners could have had negative reports of own-voice perception and the occlusion effect while using the experimental transducer (Ricketts et al., 2019). Future studies could replicate the simplex procedure for stimuli in which considerable low-frequency gain is desired and compare the low-frequency adjustment between a closed and open fitting. These studies could also conduct subjective measures of occlusion in each type of fitting.

Finally, this study used studio-compressed music recordings at a fixed level of 70 dB SPL. Previous studies have found that higher input levels can influence the impact of hearing aid compression on music sound quality (Davies-Venn et al., 2007; Moore et al., 2011). It is conceivable that higher input levels may also influence gain adjustments. For example, the same low-frequency boost at a higher input level may have a relatively higher upward spread of masking, which could negatively impact sound quality, thus leading listeners to minimize gain adjustments at higher input levels. Furthermore, loudness growth is larger for lower frequencies at higher input levels (ISO-226, 2003). Therefore, an overall level increase may be perceived as a low-frequency increase, which may yield smaller low-frequency gain adjustments. In addition, while studio compression has not been found to negatively impact music sound quality, it results in lower crest factors which enables listeners to listen to higher overall listening levels (Croghan et al., 2016). Live music consists of much larger crest factors (Chasin & Hockley, 2014), which produces music peaks that may cause hearing aid output limiting or peak-clipping, both of which can be detrimental to music listening (Davies-Venn et al., 2007). Therefore, an important research

question will be to understand the relationship between listening levels, crest factors and hearing aid circuity, and how that relationship interacts with listener gain preferences.

6.4 Summary and conclusions

This study aimed to determine the degree to which hearing-impaired listeners self-adjusted hearing aid amplification relative to prescribed gains for music and speech, and what sound quality characteristics were associated with those adjustments. This study pursued those research questions using a three-dimensional simplex procedure, using low-frequency, mid-frequency and high-frequency gain adjustments as the three dimensions, implemented with the openMHA. In Experiment One, listeners completed simplex runs for pop music, classical music, female speech, and male speech. Listeners generally increased low-frequency gain (7.3 dB) by the greatest magnitude and decreased mid- (-1 dB) and high-frequency (-4.3 dB) gain by smaller magnitudes. Low-frequency gain was increased most substantially for pop music (12.5 dB), which was significantly more gain than for both male (3.5 dB) and female speech (5.7 dB). Mid-frequency gain was increased for pop music (2.6 dB), which was significantly different than the decrease for male speech (-3.4 dB). The remaining comparisons were statistically similar. In Experiment Two, listeners provided ratings of overall impression, loudness, fullness, and sharpness for each of the stimuli amplified using both prescribed frequency shaping and preferred frequency shaping. Intelligibility ratings were performed for speech stimuli only. Relative to prescribed shaping, preferred shaping statistically increased ratings of overall impression for pop music and male speech, loudness for pop music, and fullness for pop music and classical music. Prescribed frequency shaping was rated as sharper than preferred shaping for male speech, and more intelligible than preferred shaping for female speech. For the most part, these effects generalized to spectrally-similar validation stimuli. Across music stimuli, ratings of fullness were positively and most strongly associated with ratings of overall impression, followed by sharpness, which was negatively associated with ratings of overall impression. Across speech stimuli, intelligibility was positively and most strongly associated with ratings of overall impression. Ratings of fullness and sharpness followed suit for speech as it did for music, after ratings of intelligibility.

Taken together, these experiments suggest that increasing the bass response (which is associated with the descriptor fullness) from prescribed gains, is the primary driver in music quality preferences. An increased speech bass response was also associated with improved speech quality, although less so compared to music, as listeners were likely attempting to preserve speech intelligibility. Optimal speech intelligibility was associated with prescribed, and not preferred gains, supporting the use of hearing aid prescriptions if speech intelligibility is the listening goal. The magnitude of this effect was greatest for female speech. Further research is necessary to understand the degree to which the sound quality benefits of using an increased bass response outweigh any potential discomfort from either loudness or from the occlusion effect, and whether sufficient low-frequency amplification can be achieved in a vented fitting. Further research is also necessary to understand the relationship between gain adjustments and genre, and whether stimulus characteristics can predict the degree to which gains will be adjusted.

6.5 References

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

American National Standards Institute. (1997). *Methods for Calculation of the Speech Intelligiblity Index. ANSI S3.5-1997 (R2017)*. New York: Acoustical Society of America.

Amlani, A. M., & Schafer, E. C. (2009). Application of paired-comparison methods to hearing aids. *Trends in Amplification*, *13*(4), 241–259.

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

American Speech-Language and Hearing Association (1998). Guidelines for hearing aid fitting for adults. *American Journal of Audiology*, *7*(1), 5–13.

British Society of Hearing Aid Audiologists (2012). Guidance on professional practice for Hearing Aid Audiologists. Retrieved from https://www.bshaa.com/write/MediaUploads/BSHAA Publications/bshaa_guidance_14%5B1%5D.pdf

Boothroyd, A., & Mackersie, C. (2017). A “goldilocks” approach to hearing-aid self-fitting: User interactions. *American Journal of Audiology*, *26*(3S), 430–435.

Boymans, M., & Dreschler, W. A. (2012). Audiologist-driven versus patient-driven fine tuning of hearing instruments. *Trends in Amplification*, *16*(1), 49–58.

Caswell-Midwinter, B., & Whitmer, W. M. (2018). Subjective judgments on frequency-gain curve adjustments. *Poster session presented at the 2018 International Hearing Aid Research Conference*. Tahoe City, California.

Caswell-Midwinter, B., & Whitmer, W. M. (2019). Discrimination of gain increments in speech-shaped noises. *Trends in Hearing*, *23*, 1–12.

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

Condit-Schultz, N., & Huron, D. (2015). Catching the lyrics: Intelligiblity in twelve song genres. *Music Perception: An Interdisciplinary Journal*, *32*(5), 470–483.

Convery, E., Keidser, G., & Dillon, H. (2005). A review and analysis: Does amplification experience have an effect on preferred gain over time? *The Australian and New Zealand Journal of Audiology*, *27*(1), 18–32.

Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, *35*(5), e170–e184.

Croghan, N. B. H., Swanberg, A. M., Anderson, M. C., & Arehart, K. H. (2016). Chosen listening levels for music with and without the use of hearing aids. *American Journal of Audiology*, *25*, 161–166.

Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, *18*(8), 688–699.

Dirks, D. D., Ahlstrom, J., & Noffsinger, P. D. (1993). Preferred frequency response for two-and three-channel amplification systems. *Journal of Rehabilitation Research and Development*, *30*(3), 305–317.

Dreschler, W. A., Keidser, G., Convery, E., & Dillon, H. (2008). Client-based

adjustments of hearing aid gain: The effect of different control configurations. *Ear and Hearing*, *29*(2), 214–227.

Folkeard, P., Saleh, H., Glista, D., & Scollie, S. (2018). Fit-to-target and SII normative data for DSl v5.0 adult fittings. *Poster session presented at the 2018 International Hearing Aid Research Conference*. Tahoe City, California.

Franck, B. A. M., Boymans, M., & Dreschler, W. A. (2007). Interactive fitting of multiple algorithms implemented in the same digital hearing aid. *International Journal of Audiology*, *46*(7), 388–397.

Franck, B. A. M., Dreschler, W. A., & Lyzenga, J. (2004). Methodological aspects of an adaptive multidirectional pattern search to optimize speech perception using three hearing-aid algorithms. *Journal of the Acoustical Society of America*, *116*(6), 3620–3628.

Franks, J. R. (1982). Judgments of hearing aid processed music. *Ear and Hearing*, *3*(1), 18–23.

Fulford, R., Ginsborg, J., & Greasley, A. (2015). Hearing aids and music : the experiences of D / deaf musicians. In *Proceedings of the Ninth Triennial conference for the European Society for the Cognitive Sciences of Music*. Manchester, UK.

Gabrielsson, A., Schenkman, B. N., & Hagerman, B. (1988). The effects of different frequency responses on sound quality judgements and speech intelligibility. *Journal of Speech and Hearing Research*, *31*, 166–177.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of hearing aids. *Scandinavian Audiology*, *8*(3), 155–169.

Glista, D., & Scollie, S. (2012). Development and evaluation of an English language measure of detection of word-final plurality markers: The University of Western Ontario Plurals Test. *American Journal of Audiology*, *21*, 76–81.

Herzke, T., Kayser, H., Loshaj, F., Grimm, G., & Hohmann, V. (2017). Open signal processing software platform for hearing aid research (openMHA). In *Proceedings of the Linux Audio Conference* (pp. 35–42). Universite Jean Monnet, Saint-Etienne.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, *49*(12), 891–903.

Humes, L. E. (2003). Modeling and predicting hearing aid outcome. *Trends in Amplification*, *7*(2), 41–75.

International Organization for Standrdization (2014). *Acoustics – Normal equal-loudness-level contours*. *ISO 226:2003.* Geneva, Switzerland.

Jenstad, L. M., Bagatto, M. P., Seewald, R. C., Scollie, S. D., Cornelisse, L. E., & Scicluna, R. (2007). Evaluation of the desired sensation level [input/output] algorithm for adults with hearing loss: The acceptable range for amplified conversational speech. *Ear & Hearing*, *28*(6), 793–811.

Johnson, E. E. (2013). An initial-fit comparison of two generic hearing aid prescriptive methods (NAL-NL2 and CAM2) to individuals having mild to moderately severe high-frequency hearing loss. *Journal of the American Academy of Audiology*, *24*(2), 138–150.

Keidser, G., & Alamudi, K. (2013). Real-life efficacy and reliability of training a hearing aid. *Ear and Hearing*, *34*(5), 619–629.

Keidser, G., Dillon, H. R., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, *1*(e24), 88–80.

Keidser, G., Katsch, R., Dillon, H., & Grant, F. (2002). Relative loudness of low- and high-frequency bands of speech-shaped babble, including the influence of bandwidth and input level. *Journal of the Acoustical Society of America*, *111*(2), 669–671.

Keidser, G., O'Brien, A., Carter, L., McLelland, M., & Yeend, I. (2008). Variation in preferred gain with experience for hearing-aid users. *International Journal of Audiology*, *47*(10), 621–635.

Kirchberger, M., & Russo, F. A. (2016). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, *20*, 1–16.

Kuk, F. K., & Lau, C. C. (1995). Effect of initial setting on convergence to optimal hearing aid setting using a simplex method. *British Journal of Audiology*, *29*(5), 263–269.

Kuk, F. K., & Pape, N. M. C. (1992). The reliability of the modified simplex procedure in hearing aid frequency response selection. *Journal of Speech and Hearing Research*, *35*(2), 418–429.

Kuk, F. K., & Pape, N. M. C. (1993). Relative satisfaction for frequency responses selected with a simplex procedure in different listening conditions. *Journal of Speech and Hearing Research*, *36*(1), 168–177.

Kuk, K., & Lau, C. (1996). Comparison of preferred frequency gain settings obtained with category rating and modified simplex procedure. *Journal of the American Academy of Audiology*, *7*, 322–331.

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. Retrieved from https://cran.r-project.org/package=ez

Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, *19*(6), 519–526.

Lindley, G. A., Palmer, C. V, Durrant, J., & Pratt, S. (2001). Audiologist- versus patient-driven hearing aid fitting protocols. *Seminars in Hearing*, *22*(2), 139–160.

Looi, V., Rutledge, K., & Prvan, T. (2018). Music appreciation of adult hearing aid users and the impact of different levels of hearing loss. *Ear and Hearing*, *XX*(XX), 00–00.

Mackersie, C., Boothroyd, A., & Lithgow, A. (2018). A "Goldilocks" approach to hearing aid self-fitting. *Ear and Hearing*, *40*(1), 1.

Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, *18*, 1–29.

McCreery, R. W., Bentler, R. A., & Roush, P. A. (2013). Characteristics of hearing aid fittings in infants and young children. *Ear & Hearing*, *36*(6), 701–710.

Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *Journal of the Audio Engineering Society*, *64*(3), 112–123.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, *32*(5), 556–568.

Moore, B. C. J., Glasberg, B. R., & Stone, M. A. (2010). Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *International Journal of Audiology*, *49*(3), 216–227.

Moore, B. C. J., & Sęk, A. (2012). Comparison of the CAM2 and NAL-NL2 hearing aid fitting methods. *Ear and Hearing*, *34*(1), 83–95.

Moore, B. C. J., & Sęk, A. (2016). Comparison of the CAM2A and NAL-NL2 hearing-aid fitting methods for participants with a wide range of hearing losses. *International Journal of Audiology*, *55*, 93–100.

Nelson, P. B., Perry, T. T., Gregan, M., & Van Tasell, D. (2018). Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. *Trends in Hearing*, *22*, 1–13.

Neuman, A. C., Levitt, H., Mills, R., & Schwander, T. (1987). An evaluation of three adaptive hearing aid selection strategies. *The Journal of the Acoustical Society of America*, *82*(6), 1967.

Olsen, W. O. (1998). Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology*, *7*, 1–5.

Pavlovic, C. V. (1994). Band importance functions for audiological applications. *Ear and Hearing*, *15*(1), 100–104.

Preminger, J. E., Neuman, A. C., Blakke, M. H., Deirdre, W., & Levitt, H. (2000). An examination of the practicality of the simplex procedure. *Ear and Hearing*, *21*(3), 177–193.

Preminger, J. E., & Van Tasell, D. J. (1995). Quantifying the relation between speech quality and speech intelligibility. *Journal of Speech and Hearing Research*, *38*, 714–725.

Punch, J. L. (1978). Quality judgments of hearing aid-processed speech and music by normal and otopathologic listeners. *J Am Audiol Soc*, *3*(4), 179–188.

Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*, 160–172.

Ricketts, T., Bentler, R. A., & Mueller, G. H. (2019). Chapter 9: Ear Impressions, Earmolds, and Associated Plumbing. In *Essentials of Modern Hearing Aids: Selection, Fitting, and Verification* (pp. 273–306). San Diego, CA: Plural Publishing.

Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., … Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, *9*(4), 159–197.

Stelmachowicz, P. G., Lewis, D. E., & Carney, E. (1994). Preferred hearing aid frequency responses in simulated listening environments. *Journal of Speech and Hearing Research*, *37*(3), 712–719.

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., & Moeller, M. P. (2004). The importance of high-frequency audibility in the speech and language development of children with hearing loss. *Arch Otolaryngol Head Neck Surg*, *130*, 556–562.

R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Compting, Vienna, Austria. URL Https://Www.R-*

Project.Org/.

Vaisberg, J. M., Martindale, A. T., Folkeard, P., & Benedict, C. (2018). A qualitative study of the effects of hearing loss and hearing aid use on music perception in performing musicians. *Journal of the American Academy of Audiology*, *Epub*, 1–15.

Valente, M., Abrams, H., Benson, D., Chisolm, T., Citron, D., Hampton, D., … Sweetow, R. (2006). Guidelines for the audiologic management of adult hearing impairment. *Audiology Today*, *18*(5), 1–44. Retrieved from http://audiology-web.s3.amazonaws.com/migrated/haguidelines.pdf_53994876e92e42.70908344.pdf

van Buuren, R. A., Festen, J. M., & Plomp, R. (1995). Evaluation of a wide range of amplitude-frequency responses for the hearing impaired. *Journal of Speech and Hearing Research*, *38*(1), 211–221.

Zakis, J. A., Dillon, H., & McDermott, H. J. (2007). The design and evaluation of a hearing aid with trainable amplification parameters. *Ear and Hearing*, *28*(6), 812–830.

Chapter 7

7 Discussion

7.1 Research aims

The goals of this dissertation were to measure and optimize hearing aid sound quality for music listening, and to study innovative sound quality measurement methods to assist in understanding hearing aid sound quality. Chapter 2 described a qualitative investigation in which hearing aid users were interviewed and reported their own concerns listening to and participating in a music ensemble in a live rehearsal setting. Chapter 3 assessed the validity and sensitivity of a predictive hearing aid speech quality metric for real commercial hearing aids. Chapter 4 compared music sound quality between universal speech programs and music programs for music listening between five commercially available hearing aids, as rated by hearing-impaired listeners. Chapter 5 measured electroacoustic characteristics of the hearing aids studied in Chapter 4 and modelled the relationship between those characteristics and listener sound quality ratings gathered in Chapter 4. Chapter 6 implemented a multiparameter paired comparison search strategy to optimize music sound quality, using the most predictive electroacoustic characteristics from Chapter 5 to inform parameter selection.

7.2 Summary of findings

Hearing aid users are frequently dissatisfied with music listening (Feldmann & Kumpf, 1988; Leek, Molis, Kubli, & Tufts, 2008; Looi, Rutledge, & Prvan, 2018; Madsen & Moore, 2014). This is understandable, as hearing aids are typically developed with speech in mind, with other auditory percepts like music a secondary concern. Numerous surveys have shed light on the ways in which hearing aid users are dissatisfied with hearing aids and music listening (Feldmann & Kumpf, 1988; Leek et al., 2008; Looi et al., 2018; Madsen & Moore, 2014). However, these surveys have typically defined music using a closed set of descriptors, rather than an open-ended set of questions allowing respondents to report their own individual experiences. Therefore, Chapter 2 introduced the difficulties that amateur instrumentalists (who were hearing aid users) face while listening to and performing music through qualitative semi-structured interviews. The primary challenge

reported by participants was not related to the perception of music. The primary challenge was the need to hear the conductor's instructions so that participants could effectively participate in the ensemble. Music-listening concerns were generally secondary to participation challenges. In terms of music-listening concerns, participants felt that their hearing loss (1) made them unaware of acoustic information that may have been present, (2) caused some deficiencies perceiving aspects of music like dynamics and melody recognition, and (3) encouraged them to use senses beyond their hearing to perceive music. Participants had mixed attitudes toward hearing aid music sound quality and whether hearing aid music programs effectively improved their music-listening experiences. The remainder of the dissertation built upon this last result by measuring real hearing aid sound quality and exploring what factors between hearing aids may have explained the mixed preferences.

One of the strategies used to measure hearing aid sound quality in the literature has been the use of predictive hearing aid speech quality metrics. The metric used in this dissertation was the Hearing Aid Speech Quality Index (HASQI; Kates & Arehart, 2014), which has been validated for a variety of simulated hearing aid signal-processing features (Falk et al., 2015; Harlander, Huber, & Ewert, 2014; Houben, Brons, & Dreschler, 2011; Huber, Parsa, & Scollie, 2014; Kates & Arehart, 2014; Kressner, Anderson, & Rozell, 2013; Pourmand, Parsa, & Weaver, 2013; Suelzle, Parsa, & Falk, 2013), with relatively high agreement between predictive scores and subjective ratings. HASQI scores have also been benchmarked for various strengths of real hearing aid signal-processing (Kates, Arehart, Anderson, Muralimanohar, & Harvey, 2018). Chapter 3 took the novel approach of generating HASQI scores for real hearing aid recordings of speech and compared them with corresponding subjective speech quality ratings. The results showed that HASQI produced scores that were more sensitive to stimulus adjustments compared to humans, and that those scores were highly associated with listener ratings of speech subjected to various types and degrees of hearing aid signal processing and brands of hearing aid processing. This strong association suggested that HASQI was highly predictive of human judgments of real hearing aid speech quality. However, the digital HASQI reference strategy was sensitive to test signal recording noise, which resulted in an artificial

measurement ceiling, potentially creating interpretability challenges if HASQI was implemented in clinical contexts. The measurement ceiling was resolved by adjusting HASQI's reference signal strategy with the use of a recorded reference signal, and it resulted in HASQI scores that better resembled listener ratings. However, adjusting HASQI's reference signal strategy also biased HASQI scores towards the hearing aid used to record the reference signal and prevented a meaningful cross-comparison of hearing aid models, limiting the model's utility in more practical contexts. Altogether, Chapter 3 validated HASQI for a set of signal-processing adjustments using real hearing aids and highlighted areas for further developments prior to clinical and commercial implementation. One of the next goals was to expand understanding of predictive sound quality metrics to hearing aid signal processing on amplified music sound quality. However, to validate a predictive sound quality metric for music sound quality, it was first necessary to collect a series of music sound quality ratings.

To understand how music sound quality varies across the hearing aid industry, Chapter 4 investigated hearing aid users' judgments of the sound quality of hearing aid processed music between the universal and music programs of five commercial hearing aids. A music program is a special set of signal-processing adjustments designed to improve music-listening in hearing aids (Moore, 2016), although the specific adjustments vary between manufacturers. Many hearing aid users are unsure if their devices include a music program, and those whose devices do include a music program either do not use it consistently or are unsure if it enhances music-listening compared to a speech program (Fulford, Ginsborg, & Greasley, 2015; Madsen & Moore, 2014; Vaisberg, Martindale, Folkeard, & Benedict, 2018). Chapter 4 revealed that only two out of five of the hearing aids tested implemented music programs that significantly improved sound quality ratings, and the magnitude of improvement was less than the difference between sound quality ratings belonging to high- versus low-rated hearing aids. A case study analysis of one of Chapter 4's participant's recordings was conducted to explore hearing aid electroacoustic characteristics that may have explained the sound quality preferences. Overall, the most preferred hearing aid provided more low- and high-frequency gain compared to the least preferred hearing aid. The most beneficial music program also provided more low-frequency gain compared to

the universal program of the same hearing aid. Short-term compression between hearing aids and programs was comparable, suggesting that compression was less likely to explain the differences between high and low ratings. This case study set the foundation for the primary research question in Chapter 5.

The objectives of Chapter 5 were to explore the electroacoustic characteristics of the hearing aids and music programs in Chapter 4 and to assess whether the differences observed between hearing aids and programs were associated with listener ratings. Optimal music sound quality has been associated with extended bandwidth in the low-frequency region (Franks, 1982), the high-frequency region, at least for some listeners (Moore, Füllgrabe, & Stone, 2011; Ricketts, Dittberner, & Johnson, 2008), and with minimal hearing aid compression (Arehart, Kates, & Anderson, 2011; Croghan, Arehart, & Kates, 2014; Hansen, 2002; Kirchberger & Russo, 2016; Moore et al., 2011). Bandwidth characteristics were explored by measuring the mean level across the ultra-low (100-200 Hz), low (250-800 Hz), mid (1000 Hz- 2500 Hz) and high (3000-10000 Hz) frequency bands. Hearing aid compression was inferred by measuring envelope distortion via the cepstral correlation measurement. Hearing aids differed most substantially in output level in the ultra-low frequency region, followed by the output level in the low frequency region (with the preferred settings providing more output). Preferred hearing aid settings provided less high-frequency output compared to lower-rated conditions, although the difference in high-frequency gain was less than the difference in low-frequency gain. Mid-frequency gain was comparable across hearing aids and music programs. Compression did differ between hearing aids, although compression differences were not monotonically related to the ordering of preferences between hearing aids. Of these characteristics, only hearing aid output in the ultra-low frequency region was predictive of sound quality ratings. These findings showed that hearing aids and music programs did not vary output levels and compression meaningfully for music sound quality, except for frequencies below 250 Hz.

The next questions centered on how listeners would adjust electroacoustic characteristics relative to prescribed gains when asked to optimize hearing aid sound quality, and whether their optimally-selected characteristics meaningfully differed in sound quality compared to

prescribed gains. Given that chapter 5 showed that commercial hearing aids differed most significantly in gain in different frequency bands, part 1 of Chapter 6 asked listeners to optimize hearing aid gain across low (0.1-0.8 kHz), mid (1-2.5 kHz), and high (3-10 kHz) frequency bands relative to their prescribed gains while listening to speech and music. They did so using a multiparameter adaptive paired comparison strategy (simplex) that was interfaced with a hearing aid simulator and that presented output through high-power occluding earphones. Listeners on average increased gain in the low frequency band most substantially, followed by a smaller-magnitude decrease in the high frequency band, with minimal adjustments in the mid-frequency band. Adjustments were typically larger when listening to music compared to speech. In part 2 of Chapter 6 listeners were instructed to compare ratings of total impression, fullness, loudness and sharpness for speech and music, plus intelligibility for speech. Ratings for each scale were conducted for prescribed and preferred settings determined during part 1. For music stimuli, preferred settings were typically rated as fuller, louder, and less sharp compared to prescribed settings, with ratings of fullness being most strongly associated with ratings of overall impression. For speech stimuli, the ratings followed a similar pattern of responses across the descriptors, although the differences between preferred and prescribed settings were smaller in magnitude compared to music. Similarly, the correlations between overall impression ratings and other descriptors were not as strong for speech as they were for music. For speech stimuli, ratings of intelligibility were meaningfully higher for prescribed settings compared to preferred settings and were most strongly associated with ratings of overall impression. These results suggested that listeners prioritized increasing the low-frequency gain to improve music sound quality, with a similar trend followed for speech quality, until intelligibility became compromised. The results also support fitting hearing aids using prescriptive methods so long as speech intelligibility is the goal.

7.3 Implications, limitations, and future research

7.3.1 Predicting the sound quality of real hearing aid music quality

This dissertation did not investigate whether a predictive sound quality model, such as HASQI, was predictive of hearing aid processed music. This omission is of interest because a predictive sound quality model - the hearing aid audio quality index (HAAQI; Kates & Arehart, 2016) has been developed for sound quality evaluations of hearing aid processed music samples, and is in high agreement with subjective ratings of simulated hearing aid signal-processing (Arehart et al., 2011) with a correlation coefficient of r = 0.97 (Kates & Arehart, 2016). Evaluations of HAAQI were not conducted in this dissertation due to (1) the inability to conduct between-hearing aid comparisons using a recorded reference strategy, (2) the limited range of hearing aid sound quality distortions available for model validation and (3) the question of whether a speech-based reference signal is appropriate for the prediction of music-based sound quality.

Chapter 3 revealed that HASQI was not appropriate for a between-hearing aid comparison because the recorded reference signal strategy labelled mHASQI (which aligned the range of objective scores with listener ratings) biased predicted sound quality scores towards the hearing aid used to record the reference. The samples of hearing aid processed music in Chapter 4 would have suffered from this bias. Chapter 4 provided a set of music samples that were processed using the universal and music programs of five commercial hearing aids, which prior to the investigation in Chapter 3, would have seemed like a reasonable set of recordings upon which to validate HAAQI for real hearing aids. Like HASQI, HAAQI requires a high-quality reference signal that has been amplified to match individualized prescriptive targets. The only stimulus in the Chapter 4 dataset that was shaped to closely match a prescribed set of targets was the digital file used as the reference signal in the subjective experiment. As shown in Chapter 3, the use of a digital reference would have introduced an artificial ceiling in the output scores, which would have restricted the range of scores and impacted the interpretability of the results. The artificial ceiling was resolved with the incorporation of a recorded reference signal through a hearing

aid that was also matched to prescribed targets. However, in a between-hearing aid comparison, this strategy would not be appropriate, because only one hearing aid can serve as the reference in this strategy; any results would therefore be biased towards the hearing aid used as the reference signal. If a HAAQI model was investigated using the hearing aid recordings in Chapter 4, reference signal biases would have also been present and this would have impacted the validity of the findings in a manner similar to what we have observed with the HASQI model in Chapter 3.

Another reason that a HAAQI analysis was not conducted lies in the range of hearing aid settings included in the data in Chapter 4. In early HASQI validations, signal-processing adjustments consisted of a range of settings to which both listeners and HASQI were sensitive. This was also the case in the Chapter 3 HASQI validation. Many of the signal-processing adjustments consisted of simulated hearing aid distortions used to probe sensitivity measurements, but some of the settings were so extreme that they would likely never be incorporated in real-world hearing aid fittings. This allowed HASQI scores and subjective ratings to be nearly monotonically related, which helped to both develop and validate the model. The range of settings in Chapter 4, however, included the hearing aids' default programs and did not consist of a range of high- to low-quality settings used to probe HAAQI or listener sensitivity. In a competitive hearing aid industry, it would be reasonable to expect sound quality to be relatively similar between different manufacturers, because settings with highly unacceptable sound quality are unlikely to remain as commercially viable products. Unacceptable settings that are included in lab studies for experimental reasons are less likely to occur in commercial products. Consistent with this speculation, while Chapter 4 revealed preferences between some hearing aids, all the hearing aids were rated well, and sound quality was comparable between some of the five brands and their universal and music programs. This raises the question of whether HAAQI scores would be sensitive to differences between hearing aids, and if the scores would be monotonically related to listener ratings. Chapter 5 revealed that HAAQI's cepstral correlation term was sensitive to differences in nonlinear distortion between hearing aids. However, chapter 5's mixed effects model revealed that differences in nonlinearity between hearing aids were not perceptual meaningful for sound quality. One can speculate

from this that predicted HAAQI quality scores would be poorly associated with listener sound quality ratings, at least for the range of scores and hearing aid settings included in Chapter 4.

The final reason that the HAAQI metric was not investigated for music-listening is due to a combination of the preferred-gain findings in Chapter 6 and the HAAQI reference signal requirements. Recall that the HAAQI metric calculates a sound quality score by comparing linear and nonlinear aspects of a test signal to those of a high-quality reference signal. In the context of hearing aid research, a high-quality reference signal is a stimulus that has been amplified to match prescriptive targets, and the highest HAAQI score is achieved when the test signal exactly matches the reference signal. As indicated by the findings in Chapter 6, the most preferred gains for music-listening were rated statistically higher than prescribed gains, indicating that a high-quality reference signal for a predictive music sound quality metric should be matched to a set of gains that have not been prescribed.

The question of using a prescribed-gain reference signal is investigated here by revisiting the sound quality ratings gathered in Chapter 4. In Chapter 4, listeners completed the MUSHRA protocol to compare processed music between commercial hearing aids and high-quality references and low-quality anchors. Like the reference signal in HAAQI or HASQI approaches, the high-quality MUSHRA reference signal was shaped to match prescribed targets. It was of interest to know whether Chapter 4's MUSHRA reference signal was rated favourably compared to Chapter 4's highest rated hearing aid. A follow-up repeated measures ANOVA was conducted to compare sound quality ratings between the MUSHRA reference signals and the hearing aid signals (collapsed across the levels of genre and program). The ANOVA revealed a main effect ($F_{(1,25)} = 36.74$, $p < 0.0001$, $\eta^2 = 0.40$) of stimulus-processing, with the preferred hearing aid rated 20 points higher than the reference signal. This indicates that the prescribed reference signal was not favourable compared to the highest-rated hearing aid. The data in Chapter 4 and Chapter 6 suggest that prescribed gains do not constitute a universally-preferred signal at least for music, indicating that a prescribed reference signal may not be the appropriate reference signal for the HAAQI metric for music signals.

There are two possible strategies to improve predictive hearing aid sound quality metrics for music-listening. The first strategy would be to benchmark preferred hearing aid gain settings for music-listening. These preferred settings may then be used serve as the reference signal used in the metric. Chapter 6 revealed that a boost of about 10 dB was preferred in the bass region and reduction of about 4 dB was preferred in the treble region, relative to DSL v5.0 targets for recorded music-listening at a simulated input level of 70 dB SPL. Further research is needed to determine preferred gains relative to prescribed targets for other input levels, other genres, and for live music. Further research is also needed across different degrees of hearing loss, durations of hearing aid experience, and form factors (i.e., open-fit domes, closed-fit domes, vented earmolds, etc.). Such findings can inform the development of prescribed gains for music-listening, and could contribute to the development of a reference signal for comparison-based predictive sound quality metrics for music-listening. The second strategy may be to abandon the reference signal concept altogether. A different class of predictive sound quality models known as single-ended models do not require a reference signal. Some reference-free models have been standardized and are used to evaluate the sound quality of narrowband telephone speech quality (International Telecomunnications Union, 2004), with more recent models applied to hearing aid processed speech (Suelzle et al., 2013). Sound quality analysis of hearing aid processed music using reference-free metrics may resolve some of the reference signal concerns discussed in this section and is a ripe area for future research on this topic.

7.3.2 Low-frequency amplification in open versus closed fittings

A central finding from this dissertation is that hearing-impaired listeners preferred hearing aid amplified music with more low-frequency content relative to hearing aid amplified music with less low-frequency content. This finding was investigated using both top-down (Chapters 4 and 5) and bottom-up approaches (Chapter 6). In Chapters 4 and 5, listeners compared the music sound quality of five different commercial hearing aids. On average, listeners preferred the hearing aid that amplified low-frequency energy up to about 10 dB more compared to the least preferred hearing aid. In Chapter 6, listeners increased the low-frequency content for music by an average 10 dB relative to their prescribed settings. A

follow-up investigation revealed that listeners' overall impression of the music amplified by the preferred gain was higher than that of music amplified by the prescribed gain. Overall impression ratings were most associated with ratings of "fullness" (i.e., low-frequency presence (Gabrielsson & Sjögren, 1979)) relative to other sound quality descriptors corresponding to other electroacoustic characteristics. The provision of additional low-frequency gain is consistent with recommendations for fitting hearing aids for music (Moore, 2016; Zakis, 2016). Despite these findings, there are practical challenges to consider when low-frequency amplification is desired in hearing aid fittings.

Many hearing aids are coupled to the ear using an open fitting, in which the "ear canal is…open for directly receiving ambient sounds" (Fretz, Stypulkowski, & Woods, 2001; Winkler, Latzel, & Holube, 2016), which contrast with closed (or occluded) fittings, in which the ear canal is fully sealed. Open-fit hearing aids are typically prescribed for first-time hearing aid users with mild losses and, relative to closed-fits, are often preferred for speech quality and own-voice perception (Winkler et al., 2016). However, open-fits also leak low-frequency energy (Ricketts, Bentler, & Mueller, 2019) which can degrade the sound quality of signals like music, raising the question if sufficient low frequency can be provided for ideal music sound quality. Closed-fits can be advantageous because they do not allow low frequency content to leak, thereby transmitting them directly to the auditory system and preserving music sound quality, in addition to other signal-processing benefits (Winkler et al., 2016). However, the trapping of low-frequency energy in the ear canal also leads to complaints of one's voice sounding too loud or "boomy" in what is known as the occlusion effect (Ricketts, Bentler, & Mueller, 2019). This can be problematic especially for listeners with mild losses who present with normal low-frequency thresholds and are candidates for open fittings. The low-frequency benefits determined in Chapter 4 through 6 were collected from listeners with mild to moderate hearing losses while wearing closed-fit configurations, raising questions about the generalizability of those studies' findings to open hearing aid fittings seen outside the laboratory.

The preference for low-frequency amplification and prevalence of open-fit configurations outside the laboratory raises the following research question: Can the provision of low-

frequency amplification that is satisfactory for hearing aid music sound quality be achieved using open-fit hearing aids? If sufficient low-frequency amplification can be achieved in open fittings, listeners will likely be satisfied with hearing aid music sound quality without concerns of the occlusion effect. For instance, Lundberg, Ovegård, Hagerman, Gabrielsson, & Brändström (1992) compared listener sound quality ratings of speech and music between a closed earmold and a vented earmold whose frequency response was digitally equalized by compensating for the loss of low frequency energy to match that of the closed earmold. Listener ratings were comparable across both configurations, provided that listeners' perceptions of loudness were the same for both configurations. A future research direction could be to replicate the findings of Lundberg et al., (1992) in a methodology similar to that of Chapter 6.

A possible research design could be to instruct listeners to perform the modified simplex procedure for music stimuli and compare the preferred gains using different earmold configurations ranging from fully-occluded, to partially-occluded, to fully-open. In order to conduct this study, a set of earmolds would need to be constructed for each individual participating in the study, and vents of increasing diameter would be built within the set of earmolds for each listener. Prior to simplex testing, the same stimulus processed using prescribed gains should be presented using each earmold configuration, and real-ear measurements should be conducted to determine the relative loss of low-frequency content in each earmold of increasing vent size. It is of interest to determine if listeners can achieve sufficient low-frequency gain in an open fitting within device limitations because the device would require considerably more low-frequency amplification in an open-fit compared to a closed-fit to compensate for the low-frequency leakage associated with the open-fit configuration.

It is possible that listeners may not be able to achieve satisfactory electroacoustic low-frequency gain for music sound quality using an open fitting. For instance, additional bass enhancement for music sound quality may accelerate battery consumption and/or may not be achievable depending on the output limit of the hearing aid receiver. Therefore, if ideal sound quality is the fitting goal, it may be attractive to fit the hearing aid using a closed-

fit. Given the issue of the occlusion effect in closed fittings, it would be of interest to measure the ideal sound quality/occlusion effect tradeoff with each earmold configuration. Kuk, Keenan, & Lau (2005) measured the objective and subjective occlusion effect in listeners while wearing a set of earmolds with increasingly larger vent diameters. The objective occlusion effect was analyzed by measuring an occluded ear response while the listeners vocalized the syllable /i/ and the subjective occlusion effect was measured while listeners repeated the phrase "Baby Jeannie is teeny tiny" while rating their own voice. The objective and subjective occlusion effects were significantly correlated with one another. It would be of interest to determine how much subjective occlusion is tolerable in the proposed study design. This is because if the listener (even with normal low-frequency thresholds) can tolerate a partially occluded earmold, then additional low-frequency audibility can be provided acoustically, and this would limit the power consumption needed to electroacoustically deliver low-frequency amplification.

7.3.3 Stimulus-dependent effects

Hearing aid sound quality for music can be affected depending on the genre that the user is listening. In Chapter 4, listeners were more sensitive to differences between hearing aids for the pop and jazz genres relative to the classic genre. In Chapter 6, listeners increased the low-frequency gain more substantially for pop music relative to classical music. Observations of such genre dependencies are not limited to this dissertation. Davies-Venn, Souza, & Fabry (2007) evaluated music ratings for different styles of hearing aid circuity for popular vocal and classical music genres. The authors found that sound quality ratings for classical music were higher than ratings for popular vocal music. The authors commented on the acoustic differences between stimuli, and that the classical piece had less "sharpness" (i.e., high-frequency energy (Gabrielsson & Sjögren, 1979; Jenstad, Van Tasell, & Ewert, 2003)) compared to the popular vocal music, and that sharpness may have been detrimental for music sound quality. Arehart et al., (2011) evaluated sound quality ratings for Haydn, vocal and jazz samples, which were processed using a variety of linear and nonlinear distortions and some genres were more sensitive to specific types of distortions. The authors attributed the difference in sensitivity to the different acoustic characteristics present in each of the signals. Therefore, one can speculate how signal

processing may interact with stimulus characteristics in determining sound quality. For instance, linear low-pass filters may more greatly impact music with high frequency content above the cutoff compared to music with less high frequency content. Likewise, large compression ratios may be more detrimental for a music with faster modulations and higher crest factors compared to music without.

Acoustic differences between stimuli in Chapters 4 and 6 may have explained why some genres were more sensitive to sound quality ratings than others. First, in Chapter 4, the pop and jazz samples consisted of more low-frequency energy compared to the classical samples. Chapter 5 revealed that the hearing aids differed most in the provision of low-frequency gain, whereas other electroacoustic parameters were similar across devices. Therefore, listeners may not have detected differences between hearing aids for classical music because there was not enough low-frequency content compared to the other genres to hear those differences. Second, Chapter 5 revealed that the cepstral correlation coefficient was different between hearing aids for classical music, but comparable across hearing aids for pop music. The cepstral correlation coefficient is a measurement that is sensitive to nonlinear distortion and was therefore used to infer the compressive characteristics of the hearing aids. The classical sample consisted of a much larger dynamic range (indicated by taller peaks and deeper valleys in the waveform) compared to the pop sample. Therefore, the hearing aids implemented more short-term compression by adjusting the gains for a larger range of short-term input levels in the classical sample compared to the pop sample. While the differences in compression between hearing aids were not sufficient to produce perceptual differences, this acoustic difference was likely related to the different cepstral correlation coefficients between hearing aids for that genre. Similarly, in Chapter 6, the pop genre consisted of substantially less relative low-frequency content compared to its overall spectrum versus the other stimuli. It is possible that degree of low-frequency gain increase was related to the proportion of low frequency energy present in the stimulus.

The effects of genre observed (Chapters 4, 5, 6, Arehart et al., 2011; Davies-Venn et al., 2007) highlight the fact that researchers and clinicians should be mindful that there is no

single solution that will improve music quality ratings for all genres. Fortunately, some hearing aid manufacturers offer music programs that include sub-options to specify music genres. Presumably, these music program sub-options would adjust the hearing aid characteristics to favourably amplify the acoustic characteristics of different genres. Unfortunately, these technologies are proprietary in nature and it is not clear what evidence manufacturers use to develop genre-specific music programs. Future research is needed to understand hearing aid signal processing behavior for genre-specific music programs across manufacturers and whether it is appropriate for different genres of music.

To study the impact of different genres on hearing aid processing, it is important to consider the different acoustic characteristics between genres. However, this can be challenging due to the many acoustical differences between even two types of genres. For example, different genres are associated with different average spectra, dynamic ranges (Kirchberger & Russo, 2016), crest factors (Chasin & Hockley, 2014), as well as different instruments which produce different levels and frequency ranges (Chasin & Hockley, 2014). A significant effect of genre on sound quality would not explain whether a difference in frequency range, dynamic range, or crest factor caused that effect. A better approach than studying the impact of different genres on sound quality would be to study the interaction of different acoustic features and hearing aid processing on sound quality. To do so, one should design a study in which variations of a stimulus are generated in which only a single feature is manipulated. For example, different stimuli can be digitally filtered with different increments of low-frequency cuts. This would enable a researcher to understand the impact of changing that feature alone on hearing aid processing. A large-scale research program focusing on the impact of different acoustic features in music (i.e., frequency range, dynamic range, modulation rate, etc.) on hearing aid processing would produce a body of evidence that would inform the development of stimulus-dependent music programs that would better specify the fitting of hearing aid music programs for specific music genres.

7.3.4 Hearing aid compression and music listening

A relatively minor and unexpected result from this dissertation was that hearing aid compression was not related to listener preferences of music sound quality. One of the hypotheses for the sound quality preferences between hearing aids was that preferences may have been explained by differences in compression settings between hearing aids. The goals of Chapter 5 were to determine if compression settings, among other electroacoustic measurements, were predictive of listener sound quality judgments using a linear mixed model. In Chapter 5, compression strength was estimated by analyzing nonlinear distortion in the hearing aid recordings using the cepstral correlation coefficient within the HAAQI metric (Kates & Arehart, 2016). The results revealed significant differences in nonlinearity between hearing aids. However, the model did not reveal a relationship between the cepstral correlation coefficient measurements and sound quality preferences compared to other electroacoustic predictors. These findings were surprising, as they disagree with background knowledge on the topic from previous literature.

As discussed in Chapter 1, the relationship between hearing aid compression and music-listening preferences is well understood. In general, less compression seems to be ideal for music-listening. Hearing-impaired listeners typically prefer linear or linear-like gains compared to larger compression ratios (Arehart et al., 2011; Croghan et al., 2014; Higgins, Searchfield, & Coad, 2012; Kirchberger & Russo, 2016; Madsen, Stone, McKinney, Fitz, & Moore, 2015) which compress the dynamic range of the amplified signal. Furthermore, hearing-impaired listeners typically prefer longer time constants over shorter time constants (Arehart et al., 2011; Hansen, 2002; Moore et al., 2011). Therefore, one would have predicted Chapter 5's cepstral correlation coefficient to be positively and significantly associated with sound quality judgments. That is, as the amount of hearing aid compression decreased, the nonlinearity in the test signal would decrease, and the cepstral correlation value would increase, which would finally be associated with a greater sound quality preference. This result was not the case, and this is discussed further below.

The results from Chapter 5 may not necessarily contradict the background knowledge on hearing aid compression and music-listening. The findings can likely be explained by the

research design of the study. The original research plan was to conduct a cross-manufacturer comparison of hearing aid sound quality music, which was reported in Chapter 4. The initial research plan was not to develop the exploratory linear mixed-model comparing electroacoustic measurements and sound quality ratings. Therefore, the stimuli were not designed in a way that would probe significantly different electroacoustic differences that would yield different perceptual ratings. Rather, the stimuli were selected to assess between-brand differences at a typical listening level. Therefore, the stimuli (investigated in Chapter 5) were all presented to the hearing aids at a short-term input level of about 70 dB SPL, which may not have fully engaged the compressive behaviour of the different hearing aids. For instance, the highest and lowest cepstral correlation coefficient measurements were between 0.85 and 0.95, which based on the results of Chapter 5, did not yield perceptual differences. These reflections have implications for future research designs in this area.

There are several ways to consider the role of hearing aid compression in music sound quality using research designs similar to those implemented in this dissertation. In Chapter 5, a linear mixed model was used to analyze stimuli recorded from hearing aids that implemented compression and that were stimulated at a fixed moderate input level. Future studies may wish to gather ratings and recordings across a wider range of input levels in order to engage level-dependent hearing aid behaviours, such as the effects of compression kneepoints and/or output limiting characteristics. This would generate a larger range of measurements which may probe perceptual differences in sound quality ratings more effectively. In Chapter 6, a three-dimensional modified simplex procedure determined the optimal gain adjustments across low, middle, and high frequency band levels relative to prescribed settings. The middle and high frequency band adjustments were statistically similar, suggesting that the two frequency bands could be combined in a future simplex design, leaving the option to include an additional electroacoustic parameter. Hearing aid compression ratio could be a suitable candidate parameter to fill the third electroacoustic parameter dimension. It would first be necessary to determine a just noticeable difference step size for a compression ratio/time constant combination, prior to optimizing compression in conjunction with low and high frequency bands using a simplex procedure.

7.3.5 Direct-to-consumer hearing devices and the relationship between quality and intelligibility

Direct-to-consumer (DTC) hearing devices will soon become available in the assistive hearing device market (Manchaiah et al., 2017). DTCs are assistive hearing devices that can be purchased directly by the consumer, bypassing the involvement of a hearing health care professional, and some are sold at a fraction of the cost of traditional hearing aids. Many DTC technologies will enable listeners to program them on their own using app-based hearing tests, self-adjustable settings and parameters, or a combination of both. The lack of professional involvement in the self-adjustment process implies that nobody will be available to manage consumer expectations from the devices once they are fully programmed, raising questions about the adequacy of DTC fitting outcomes. Recent research has shown that some DTC products show similar, but slightly poorer outcomes compared to traditional hearing aids and service delivery models (Humes et al., 2017; Reed, Betz, Kendig, Korczak, & Lin, 2017). Some DTC models do not improve speech understanding in noise and others even make speech understanding in noise worse (Reed et al., 2017).

In, Chapter 6 hearing-impaired listeners self-adjusted frequency band parameters to optimize sound quality. They were instructed to choose processing conditions which they preferred "as if it were their own [hearing aid]." Listeners consistently chose to increase the low frequencies and reduce the high frequencies. While these adjustments improved ratings of overall impression, they also reduced ratings of intelligibility for female speech. Boymans & Dreschler (2012) compared listeners' outcomes for hearing aids fitted using the NAL-NL1 prescription to hearing aids fitted using patient-driven fine tuning. Listeners, on average, selected lower high frequency gains, which yielded poorer intelligibility scores. This tendency to increase low-frequency gain and/or reduce high frequency gain relative to hearing aid prescriptions, which is associated with poorer intelligibility, is fairly common in self-adjustment hearing aid studies (Boothroyd & Mackersie, 2017; Kuk & Pape, 1993, 1992; Nelson, Perry, Gregan, & VanTasell, 2018; Ricketts, 1996).

As discussed above, quality-based self-adjustment protocols may not be ideal for optimal intelligibility. Several studies investigating the relationship between speech intelligibility and sound quality show that sound quality will only be optimal as long as speech intelligibility is ideal. This was indirectly observed in Chapter 2 of this dissertation. Through qualitative interviews, instrumentalists who were also hearing aid users reported that their primary concerns during rehearsals were understanding the conductor's speech. This contrasted with the expectation in which listeners would primarily be concerned with amplified music quality. In Chapter 6, ratings of fullness (low-frequency content) were the parameter most associated with ratings of total impression if the stimulus was music. However, as long as speech was involved, ratings of intelligibility become most associated with ratings of total impression, which was then followed by fullness ratings. Gabrielsson, Schenkman, & Hagerman (1988) instructed listeners to rate the importance sound quality dimensions for speech and music under various frequency responses. Listeners prioritized descriptors like clarity, loudness and nearness for speech and fullness and spaciousness (at least for normal hearing-listeners) for music. Davies-Venn et al. (2007) also instructed listeners to rate various sound quality dimensions of speech and music subjected to different forms of compression, and correlated subdimensions with ratings of overall impression. For speech, ratings of intelligibility were most strongly associated with ratings of overall impression. For music stimuli, ratings of fullness were most strongly associated with ratings of overall impression. Preminger & Van Tasell (1995) also examined the relationship between intelligibility and other sound quality dimensions while participants listened to speech under different processing conditions. When intelligibility was ideal and held constant, other sound quality ratings varied in predictable ways based on the listening conditions. When intelligibility varied, sound quality ratings were no longer predictable. Together, these findings highlight the need to preserve intelligibility in self-adjustment methods to support improvement in speech communication.

For these reasons, self-adjustment protocols that include a speech intelligibility task may be best suited for successful self-fitting outcomes. Kuk & Pape (1992) instructed listeners to perform the simplex procedure in which parameters varied by low and high gain using two criteria: consonant detection in noise and listening preference for discourse. Listeners

preferred more high-frequency gain using the consonant detection criterion task compared to a preference criterion. In Boothroyd & Mackersie, (2017), listeners were required to make a shaping adjustment while listening to discourse, followed by a speech recognition task, followed by another shaping adjustment. Most listeners increased the overall level and some listeners increased high-frequency content only after the speech recognition task, which led to improved speech recognition scores following the second adjustment. These two studies suggest that listeners can optimize intelligibility using a self-adjustment protocol if the protocol includes a speech intelligibility task.

In summary, DTC self-adjustment protocols need to be carefully developed if successful DTC outcomes are desired. Without correct instructions, listeners will likely self-adjust DTC parameters to settings which optimize sound quality. However, the evidence suggests that listeners need to prioritize optimal intelligibility. Only once intelligibility is ideal should listeners self-adjust to improve sound quality. Therefore, DTC self-adjustment protocols should incorporate speech recognition tasks, so that listeners can fit the device to improve their ability to understand speech. Future research is needed for the development of protocols which allow listeners to optimize both speech intelligibility and sound quality across multiple listening environments and for different types of signals that they encounter in their everyday lives.

7.4 Concluding statements

Hearing aids assist individuals with hearing loss to listen and communicate. Hearing aids provide this assistance primarily through improving speech intelligibility using evidence-based prescribed gains. Many hearing aid signal processing parameters are programmed to improve intelligibility but can also have negative impacts on sound quality, particularly for music signals. These sound quality degradations can be a significant barrier for device adoption in hearing aid users. This dissertation focused on several areas relevant to sound quality issues in hearing devices. In one area, a predictive sound quality metric, known as the hearing aid speech quality index, was validated for real hearing aid fittings. This metric will enable researchers and clinicians to understand the relationship between hearing aid

signal parameter adjustments and perceptual changes in real hearing aids, allowing them to optimize sound quality in experimental designs and perhaps even clinical fittings. In another area, novel investigations of hearing aids and music allowed for the evaluation of sound quality and these investigations enabled listeners to compare hearing aids or adjust hearing aids without intelligibility being a factor. These investigations revealed that increasing the sense of fullness and reducing the sense of sharpness will improve the sound quality of hearing aid fittings relative to how they are typically prescribed. However, qualitative interviews suggested that listeners prioritize speech intelligibility before sound quality. Furthermore, listeners' gain adjustments for speech were, on average, associated with only a modest improvement of overall impression, but also with a significant decline in intelligibility. Individual ratings of intelligibility were positively correlated with individual ratings of overall impression. These results suggest that quality should only be optimized so long as speech intelligibility is not compromised. The findings from this dissertation have future research and technological implications for music-based sound quality metrics, the provision of low frequency gain, genre-dependent music programs, modelling hearing aid compression and music, and DTC technologies.

7.5 References

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

Boothroyd, A., & Mackersie, C. (2017). A “goldilocks” approach to hearing-aid self-fitting: User interactions. *American Journal of Audiology*, *26*(3S), 430–435.

Boymans, M., & Dreschler, W. A. (2012). Audiologist-driven versus patient-driven fine tuning of hearing instruments. *Trends in Amplification*, *16*(1), 49–58.

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

Croghan, N. B. H., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, *35*(5), e170–e184.

Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, *18*(8), 688–699.

Falk, T. H., Parsa, V., Santos, J. F., Arehart, K., Hazrati, O., Huber, R., … Scollie, S. (2015). Objective quality and intelligibility prediction for users of assistive listening devices: Advantages and limitations of existing tools. *IEEE Signal Processing Magazine*, *32*(2), 114–124.

Feldmann, H., & Kumpf, W. (1988). Musikhoren bei Schwerho rigkeit mit und ohne Horgerat [Listening to music in the hard-of-hearing individual with and without hearing aid]. *Laryng Rhinol Otol*, *67*(10), 489–497.

Franks, J. R. (1982). Judgments of hearing aid processed music. *Ear and Hearing*, *3*(1), 18–23.

Fulford, R., Ginsborg, J., & Greasley, A. (2015). Hearing aids and music : the experiences of D / deaf musicians. In *Proceedings of the Ninth Triennial conference for the European Society for the Cognitive Sciences of Music*. Manchester, UK.

Gabrielsson, A., Schenkman, B. N., & Hagerman, B. (1988). The effects of different frequency responses on sound quality judgements and speech intelligibility. *Journal of Speech and Hearing Research*, *31*, 166–177.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of hearing aids. *Scandinavian Audiology*, *8*(3), 155–169.

Hansen, M. (2002). Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear and Hearing*, *23*(4), 369–380.

Harlander, N., Huber, R., & Ewert, S. D. (2014). Sound quality assessment using auditory models. *Journal of the Audio Engineering Society*, *64*(5), 324–336.

Higgins, P., Searchfield, G., & Coad, G. (2012). A comparison between the first-fit settings of two multichannel digital signal-processing strategies: Music quality ratings and speech-in-noise scores. *American Journal of Audiology*, *21*, 13–21.

Houben, R., Brons, I., & Dreschler, W. A. (2011). A nethod to remove differences in frequency response between commercial hearing aids to allow direct comparison of the sound quality of hearing-aid features. *Trends in Amplification*, *1–2*, 77–83.

Huber, R., Parsa, V., & Scollie, S. (2014). Predicting the perceived sound quality of frequency-compressed speech. *PloS One*, *9*(11), e110260.

Humes, L. E., Rogers, S. E., Quigley, T. M., Main, A. K., Kinney, D. L., & Herring, C.

(2017). The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: A randomized double- blind placebo-controlled clinical trial. *American Journal of Audiology*, *26*, 53–79.

Intenational Telecommunication Unions (2004). *Single-ended method for objective speech quality assessment in narrow-band telephony applications*. *Recommendaion ITU-T P.563*. Geneva, Switzlerland.

Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *Journal of the American Academy of Audiology*, *14*(7), 347–360.

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

Kates, J. M., & Arehart, K. H. (2016). The hearing-aid audio quality index (HAAQI). *IEEE/ACM Transactions on Speech and Language Processing*, *24*(2), 354–365.

Kates, J. M., Arehart, K. H., Anderson, M. C., Muralimanohar, R. K., & Harvey, L. O. (2018). Using objective metrics to measure hearing aid performance. *Ear & Hearing*, *36*(6), 1165-1175.

Kirchberger, M., & Russo, F. A. (2016). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, *20*, 1–16.

Kressner, A. K., Anderson, D. V, & Rozell, C. J. (2013). Evaluating the generalization of the hearing aid speech quality index (HASQI). *IEEE Transactions on Audio, Speech and Language Processing*, *21*(2), 407–415.

Kuk, F. K., & Pape, N. M. (1993). Relative satisfaction for frequency responses selected with a simplex procedure in different listening conditions. *Journal of Speech and Hearing Research*, *36*(1), 168–177.

Kuk, F. K., & Pape, N. M. C. (1992). The reliability of the modified simplex procedure in hearing aid frequency response selection. *Journal of Speech and Hearing Research*, *35*(2), 418–429.

Kuk, F., Keenan, D., & Lau, C.-C. (2005). Vent configurations on subjective and objective occlusion effect. *Journal of the American Academy of Audiology*, *16*, 747–762.

Leek, M. R., Molis, M. R., Kubli, L. R., & Tufts, J. B. (2008). Enjoyment of music by elderly hearing-impaired listeners. *Journal of the American Academy of Audiology*, *19*(6), 519–526.

Looi, V., Rutledge, K., & Prvan, T. (2018). Music appreciation of adult hearing aid users and the impact of different levels of hearing loss. *Ear and Hearing*, *XX*(XX), 00–00.

Lundberg, G., Ovegård, A., Hagerman, B., Gabrielsson, A., & Brändström, U. (1992). Perceived sound quality in a hearing aid with vented and closed earmould equalized in frequency response. *Scandinavian Audiology*, *21*(2), 87–92.

Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, *18*, 1–29.

Madsen, S. M. K., Stone, M. A., McKinney, M. F., Fitz, K., & 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.

Manchaiah, V., Taylor, B., Dockens, A. L., Lane, K., Castle, M., & Grover, V. (2017).

Applications of direct-to-consumer hearing devices for adults with hearing loss : a review. *Clinical Interventions in Aging*, *12*, 859–871.

Moore, B. C. J. (2016). Effects of sound-induced hearing loss and hearing aids on the perception of music. *Journal of the Audio Engineering Society*, *64*(3), 112–123.

Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, *32*(5), 556–568.

Nelson, P. B., Perry, T. T., Gregan, M., & VanTasell, D. (2018). Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. *Trends in Hearing*, *22*, 1–13.

Pourmand, N., Parsa, V., & Weaver, A. (2013). Computational auditory models in predicting noise reduction performance for wideband telephony applications. *International Journal of Speech Technology*, *16*(4), 363–379.

Preminger, J. E., & Van Tasell, D. J. (1995). Quantifying the relation between speech quality and speech intelligibility. *Journal of Speech and Hearing Research*, *38*, 714–725.

Reed, N. S., Betz, J., Kendig, N., Korczak, M., & Lin, F. R. (2017). Personal sound amplification products vs a conventional hearing aid for speech understanding in noise. *Journal of the American Medical Association*, *318*(1), 4–5.

Ricketts, T. A. (1996). Fittings hearing aids to individual loudness-perception measures. *Ear & Hearing*, *17*(2), 124–132.

Ricketts, T. A., Dittberner, A. B., & Johnson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*, 160–172.

Ricketts, T., Bentler, R. A., & Mueller, G. H. (2019). Chapter 9: Ear Impressions, Earmolds, and Associated Plumbing. In *Essentials of Modern Hearing Aids: Selection, Fitting, and Verification* (pp. 273–306). San Diego, CA: Plural Publishing.

Suelzle, D., Parsa, V., & Falk, T. H. (2013). On a reference-free speech quality estimator for hearing aids. *The Journal of the Acoustical Society of America*, *133*(5), EL412-EL418.

Vaisberg, J. M., Martindale, A. T., Folkeard, P., & Benedict, C. (2018). A qualitative study of the effects of hearing loss and hearing aid use on music perception in performing musicians. *Journal of the American Academy of Audiology*, *Epub*, 1–15.

Zakis, J. A. (2016). Music perception and hearing aids. In G. R. Popelka, B. C. J. Moore, R. R. Fay, & A. N. Popper (Eds.), *Hearing Aids* (pp. 217–252). Cham, Switzerland: Spring International Publishing Swizterland.

Appendices

Appendix A: Ethics approval notice 1

Western Research

Research Ethics

Western University Health Science Research Ethics Board
HSREB Amendment Approval Notice

Principal Investigator: Dr. Prudence Allen
Department & Institution: Health Sciences\Communication Sciences & Disorders, Western University

Review Type: Expedited
HSREB File Number: 7059
Study Title: Evaluation of hearing aid technology in adults (NCA Translational Research Unit 1) - 17091E

HSREB Amendment Approval Date: February 05, 2016
HSREB Expiry Date: May 19, 2016

Documents Approved and/or Received for Information:

Document Name	Comments	Version Date
Revised Western University Protocol		2016/02/05
Instruments		2016/02/05
Revised Letter of Information & Consent		2016/01/07

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the amendment to the above named study, as of the HSREB Initial Approval Date noted above.

HSREB approval for this study remains valid until the HSREB Expiry Date noted above, conditional to timely submission and acceptance of HSREB Continuing Ethics Review.

The Western University HSREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guideline for Good Clinical Practice Practices (ICH E6 R1), the Ontario Personal Health Information Protection Act (PHIPA, 2004), Part 4 of the Natural Health Product Regulations, Health Canada Medical Device Regulations and Part C, Division 5, of the Food and Drug Regulations of Health Canada.

Members of the HSREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB.

The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Ethics Officer, on behalf of HSREB Vice Chair

Ethics Officer to Contact for Further Information

This is an official document. Please retain the original in your files

Western University, Research, Support Services Bldg., Rm. 5150

Appendix B: Ethics approval notice 2

Western

Office of Research Ethics

The University of Western Ontario

Website: www.uwo.ca/research/ethics

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. P. Allen
Review Number: **17091E** **Review Level:** Expedited
Review Date: May 05, 2010 **Approved Local # of Participants:** 500
Protocol Title: Evaluation of hearing aid technology in adults (NCA Translational Research Unit 1)
Department and Institution: Communication Sciences & Disorders, University of Western Ontario
Sponsor: ONTARIO RESEARCH FUND #RE-03-009
Ethics Approval Date: May 19, 2010 **Expiry Date:** August 31, 2014
Documents Reviewed and Approved: UWO Protocol, Letter of Information and Consent. Participation in Research Opportunities, Telephone Script. Poster.

Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced study on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:

a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
b) all adverse and unexpected experiences or events that are both serious and unexpected;
c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. Joseph Gilbert
FDA Ref. #: IRB 00000940

Ethics Officer to Contact for Further Information

This is an official document. Please retain the original in your files.

cc: ORE File

Appendix C: Ethics approval notice 3

Western Research

Date: 19 March 2018

To: Dr. Susan Scollie

Project ID: 111154

Study Title: Music preferences in hearing aids

Application Type: HSREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 03APR2018

Date Approval Issued: 19/Mar/2018 12:50

REB Approval Expiry Date: 19/Mar/2019

Dear Dr. Susan Scollie

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study as described in the WREM application form, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
_2AFC Screen	Other Data Collection Instruments	29/Jan/2018	1
_MUSHRA Screen	Other Data Collection Instruments	31/Jan/2018	1
Email Script _03-06-2018 - NCA - R1 - CLEAN	Email Script	06/Mar/2018	2
Email Script _03-06-2018 - Unitron - R1 - CLEAN	Email Script	06/Mar/2018	2
Participant Information and Consent Form for Future Research Studies	Written Consent/Assent	10/Apr/2017	1
protocol	Protocol	06/Mar/2018	1
Study Debriefing	Debriefing Script	01/Feb/2018	
StudyPoster - R1 - CLEAN	Recruitment Materials	06/Mar/2018	2
Telephone_Script - NCA - R1 - CLEAN	Telephone Script	06/Mar/2018	2
Telephone_Script - Unitron - R1 - CLEAN	Telephone Script	06/Mar/2018	2
WREM_LOI - R1 - CLEAN	Written Consent/Assent	13/Mar/2018	2

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
Flow Diagram - R1 - CLEAN	Flow Diagram	06/Mar/2018	2

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB , except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C,

Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix D: Copyright permission for Chapter 2

From:
Sent: Monday, March 4, 2019 6:00 PM
To:
Cc:
Subject: RE: Copyright inquiry

Hi, Jonathan Vaisberg,

Thank you for reaching out to the Academy for your request.

We do not use copyright release forms, however, we allow authors to repurpose their articles that were published in JAAA with the following language:

©2019 American Academy of Audiology. Originally published in the Journal of the American Academy of Audiology, Vol. 29, No. 10. Used with permission.

Thank you for choosing JAAA!

--

Senior Manager of Communications and Publications

American Academy of Audiology
11480 Commerce Park Drive, Suite 220
Reston, VA 20191
www.audiology.org

Appendix E: Copyright permission for Chapter 4

AudiologyOnline CEU Requirements Form
& Author Agreement

1. Course Title: Comparison of music sound quality between hearing aids and music programs

2. Course Author(s): Jonathan M Vaisberg, Paula Folkeard, Vijay Parsa, Ewan Macpherson, Matthias Froehlich, Veronika Littmann, Susan Scollie

3. Short paragraph professional bio of author(s):

Jonathan Vaisberg is a PhD/MClSc Candidate at the National Centre for Audiology at Western University, London, ON, Canada. Jonathan's research interests include the impact of hearing loss on music perception, and optimizing hearing aid sound quality for music.

Paula Folkeard, AuD, is the Research Audiologist/Project Coordinator of the Translational Research unit for the National Centre for Audiology at Western University. Her areas of interest include adult amplification, audiological equipment and assistive listening device evaluation, and outcome measures.

Vijay Parsa, PhD is an Associate Professor at the National Centre for Audiology/Faculties of Engineering and Health Sciences, Western University. His research interests are in speech signal processing with applications to hearing aids, assistive listening devices and augmentative communication devices.

Ewan Macpherson, PhD is the director of the Spatial and Prosthetic Hearing Laboratory at the National Centre for Audiology at Western University, and is an Associate Professor in the School of Communication Sciences and Disorders. His current research program focuses on the effect of head movement and multimodal sensory integration on spatial hearing, the effects of assistive devices on spatial hearing in complex environments, and virtual auditory space methods.

Matthias Froehlich, PhD, is the head of Corporate Marketing Audiology for Sivantos in Erlangen, Germany. He received his PhD in Physics from Goettingen University, Germany.

Veronika Littmann, PhD, is the team lead of R&D Audiology System Development Team for Sivantos in Erlangen Germany. She holds a PhD in Neurophysiology from Cambridge University.

Susan Scollie, PhD is an Associate Professor at the National Centre for Audiology, University of Western Ontario. With colleagues, she developed version 5.0 of the DSL method for hearing aid fitting. Her current research focuses on frequency compression signal processing, and outcomes of hearing aids for infants, children and adults.

4. Two-sentence course description: While hearing aid users often wear their hearing aids for music listening, they are frequently dissatisfied with the sound quality of music. This study describes music sound quality ratings between the universal and music programs of five premier market hearing aids, and which hearing aid achieves the best overall sound quality ratings.

5. Instructional Level (Introductory, Intermediate, or Advanced):
Intermediate

6. Learner Outcomes: Please include 3 learner outcomes that are measurable and complete the sentence below. Please do not use words such as "learn" and "understand" since these outcomes cannot be measured.
As a result of this course, participants will be able to:

1) As a result of this Continuing Education Activity, participants will be able to consider the acoustic differences between speech and music as they relate to hearing aid fittings.

2) As a result of this Continuing Education Activity, participants will be able to describe the diversity of hearing aid music sound quality and hearing aid music programs across the industry.

3) As a result of this Continuing Education Activity, participants will be aware of hearing aid selection choices and fine-tuning adjustments that may improve hearing aid sound quality for music.

7. Ten Multiple Choice Questions with 4 possible answers for each and the number of the related course learning outcome. No True/False questions. Please follow format below and add the additional questions following

Question 1.
Q1: Which statement about speech is not true?
a) Average speech levels vary between 55 and 66 dBA.
b) The female fundamental frequency can be as low as 100 Hz.
c) Speech acoustics are fairly consistent across languages.
d) The typical speech dynamic range is 20-30 dB.
Correct Answer: b
Learner Outcome number: 1

Question 2.
Q2: In what ways can music differ acoustically compared to speech?
a) Larger dynamic range
b) Broader frequency spectrum
c) Higher overall level
d) All of the above
Correct Answer: d
Learner Outcome Number: 1

Question 3.
Q3: Which statement about music genres is not true?
a) Acoustically, all genres are the same.
b) The acoustics of each genre depends on the instruments played.
c) Music genres can affect hearing aid sound quality differently.
d) Music genres can be acoustically similar to speech.
Correct Answer: a.
Learning Outcome Number: 1.

Question 4.
Q4: Which of the following statements best describe the results of a survey of hearing aid users regarding their own music programs?
a) Common features of a music program include strong compression and noise reduction settings.
b) Most hearing aid users have a music program and they find it improves music sound quality.
c) Music program parameters are typically the same between manufacturers.
d) A minority of hearing aid users have a music program and they do not find that it improves music sound quality.
Correct Answer: d
Learner Outcome Number: 1

Question 5.
Q5: What is the typical audible bandwidth of a hearing aid?
a) 300 Hz – 3400 Hz
b) 200 Hz – 5000 Hz
c) 100 Hz – 12500 Hz
d) 100 Hz – 2000 Hz
Correct Answer: b
Learning Outcome Number: 1

Question 6.
Q6: Is hearing aid selection an important consideration for hearing aid users who listen to music?
a) Yes. Some hearing aids have better music sound quality compared to others.
b) Yes. Only a hearing aid with a music program should be selected.
c) No. All hearing aids improve music sound quality equally.
d) No. All hearing aids distort music sound quality equally.
Correct Answer: a
Learner Outcome Number: 2

Question 7.
Q7: How does a hearing aid music program affect music sound quality?
a) It improves music sound quality, regardless of genre or brand
b) It improves music sound quality, depending on the brand but not genre.
c) It improves music sound quality, depending on the genre and brand.
d) It does not improve music sound quality for any genre or brand.
Correct Answer: c
Learner Outcome Number: 2

Question 8.
Q8: In the study, how did genre affect hearing aid music sound quality?
a) Some genres were more sensitive to differences between hearing aid brands.
b) Some genres were more sensitive to differences between music programs.
c) There was no relationship between genre and hearing aid music sound quality.
d) A & B
Correct Answer: d
Learner Outcome Number: 2

Question 9.
Q9: Based on the literature, what are some electroacoustic parameters that best improve music sound quality?
a) Lowering the gain in the low-frequency range.
b) Using the least compressive settings possible.
c) Using fast-acting, non-linear compression settings.
d) Using short-acting, non-linear compression settings.
Correct Answer: b

Learner Outcome Number: 3

Question 10.
Q10: What was the difference between frequency-gain curves for a high-rated music clip vs. a low-rated music clip?
a) More gain the in the low frequencies for the high-rated clip.
b) Less gain in the high frequencies for the high-rated clip.
c) Less gain in the low frequencies for the high-rated clip.
d) There was no difference.
Correct Answer: a
Learner Outcome Number: 3

8. Transparency. Continuing education course content must focus on the science and/or contemporary practice of audiology and/or hearing healthcare. Courses about products or services must provide information in a scholarly manner regarding theoretical aspects and/or the details of operation of the product or service. It must be disclosed prior to such courses if there will be limited or no information provided about similar products or services. No promotions or offers of any kind may be associated with your AudiologyOnline course.
Please initial the description that best describes the content of your course:

_____ This learning event does not focus exclusively on any specific product or service.

__JV_ This learning event focuses on the following specific products or services and will not include info on other products or services: Signia Primax & Binax

9. Disclosures. Conflicts of interest may be real or perceived, and any potential for financial or personal gain deriving from the content of your course may constitute a conflict of interest and must be disclosed. Examples of financial disclosures may include but are not limited to: your employer, any grant/research support you are receiving and/or have received, stockholder/ownership in any company or organization, consulting services you provide, royalties you receive and/or patents you own. Non-financial disclosures may include, but are not limited to participation in an advisory board, review panel, membership in an organization, etc. Please list all disclosures for each author or indicate "None" if applicable.

- Financial disclosures: Funding for this project was provided by Sivantos-Signia.
- Non-financial disclosures:

Additional Terms and Conditions

Term: This Agreement is applicable for the above named course to be made by the Author. The Author may be referred to as you or your throughout this Agreement.

Copyright Policy: AudiologyOnline and its Affiliates are for-profit companies and your article and course materials do not fall under the "fair use" exception of the Copyright Act. Therefore, by agreeing to these terms and conditions, you acknowledge that you are required to obtain permission to use any copyrighted material in your article, and this permission must include publishing the material on AudiologyOnline under the terms of this Agreement. Such permission shall be made available to AudiologyOnline upon request.

Indemnification: By signing this Agreement, you attest that all articles are original works, free of any copyright restrictions, and/or that you have secured appropriate permissions to publish all contents of the article on AudiologyOnline and/or its Affiliates including, but not limited to images, graphs, charts, and videos. By signing this form, you assume all liability should a claim or dispute arise and you indemnify and hold harmless AudiologyOnline, its Affiliates, directors, officers, employees and agents and their respective successors, heirs and assigns against any and all claims, damages and expenses (including without limitation attorneys' fees) arising out of, without limitation, any breach of terms of this Agreement, misrepresentation by you, whether or not intentional and/or resulting from or related to any claim of copyright infringement, claims of invasion of privacy, unauthorized use of names or pictures, trademark infringement, libel and/or misrepresentation brought by a third party claim or proceeding involving use of any of the materials provided by you.
In the event of a claim (including third party claims) or proceeding for any of the above stated causes of action, AudiologyOnline will notify you, in writing, promptly after AudiologyOnline becomes aware of the facts giving rise to such claim(s) for indemnification. If you do not take control of such claim(s) within thirty (30) days of notification, AudiologyOnline retains the control to respond, defend, litigate, settle or take any other necessary means to handle the third party claim or proceeding and you will be liable to AudiologyOnline for all costs, expenses and damages (including but not limited to attorneys' fees).

Permission. The Author hereby grants Audiology Holdings, LLC d/b/a AudiologyOnline to publish my article as a continuing education (CE) course on AudiologyOnline. Author shall maintain ownership rights in the content of the CE course. However, all rights in the online CE course shall vest in AudiologyOnline. AudiologyOnline shall be entitled to market, promote, assign, license, store and/or use the CE course as it sees fit based on its sole discretion. This includes offering the CE course in various formats (including, but

not limited to video, recorded, transcription, podcast, modular) and offering the CE course on its affiliate sites OccupationalTherapy.com, SpeechPathology.com, PhysicalTherapy.com, and any future affiliate sites ("Affiliates").

Patient Information. Any names, photographs, videos, or other patient/client identifiers in course materials must be used only with the patient/client's knowledge and written authorization, which shall be made available to AudiologyOnline upon request.

Evidence Based. All CE courses on AudiologyOnline must be evidence-based. In evidence-based continuing education sessions, the Author provides:

- the best available, current, scientific evidence to support a particular practice or guideline which can include exploratory or pilot study data and the related evidence
- information about his/her clinical expertise, expert opinion, and experience in this topic area
- information related to client/patient/caregiver values.

For additional information regarding evidence-based requirements, please see the ASHA resource: http://www.asha.org/ce/for-providers/EBCETutorialIntro.htm

Nondiscrimination Policy: It is the policy of AudiologyOnline to create and maintain a rich online learning environment that values diversity, respects human dignity, is hospitable, equitable, tolerant, and one in which all persons are free from all forms of discrimination or discriminatory harassment. AudiologyOnline requires its authors to refrain from use of derogatory or disparaging language and other forms of expression that insults persons on the basis of race, ethnicity, national origin, ancestry, citizenship, religion, creed, gender, sexual orientation, marital status, age, disability, or veteran status.

Honorarium: N/A

Cancellation: Author and/or AudiologyOnline may cancel the article with written notice to the other party prior to the submission of the article without penalty. However, if the article is cancelled before it is published, the Honorarium, if applicable, shall not be paid.

No Waiver: No failure or delay by AudiologyOnline in exercising any right, power or privilege given by any provision of this Agreement shall operate as a waiver of the provision. Additionally, no single or partial exercise of any right, power or privilege shall preclude any other or further exercise of that or any other right, power or privilege.

Severability: If any part of this Agreement is found to be prohibited by law, the remainder shall remain valid and enforceable.

Counterpart Execution: This Agreement may be executed in one or more counterparts, each of which shall be deemed an original and part of one and the same document. Delivery of an executed copy of this Agreement by facsimile or email transmission shall constitute effective and binding execution and delivery hereof and the signatures thereon shall be deemed to be original signatures for all purposes.

Successors and Assigns: AudiologyOnline, without the Author's consent, may transfer or assign any of its rights and obligations under this Agreement.

ENTIRE AGREEMENT: The Author acknowledges that he/she has read this Agreement, understands it and agrees to be bound by its terms and conditions. Further, the Author agrees that this Agreement and its attachments constitute the complete and exclusive agreement between the parties with respect to the matter contained herein, and that such Agreement supersedes and replaces all other documents including, without limitation, oral or written agreements and all other communications between the parties as of the date written above. Any changes to the Additional Terms and Conditions in this agreement without the prior express written consent of AudiologyOnline shall render this entire agreement null and void.

Fron.
Sent.
Tc
Subject: Re: AudiologyOnline article ref #20872 (permission to reprint)

H.

Yes permission is granted to publish the article as part of your dissertation. Please print the following on each page of the article:

Congratulations in advance on your PhD! Good luck getting to the finish line!
Best,

Reference
Vaisberg, J.M., Folkeard, P., Parsa, V., Froehlich, M., Littmann, V., Macpherson, E.A., & Scollie, S. (2017, August). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline*, Article 20872. Retrieved from www.audiologyonline.com

CONTINUED

Curriculum Vitae

Name:	Jonathan Matthew Vaisberg
Post-secondary Education and Degrees:	McMaster University Hamilton, Ontario, Canada 2008-2013 B.Sc. (Hons) The University of Western Ontario London, Ontario, Canada 2013-2019 M.Cl.Sc. Candidate The University of Western Ontario London, Ontario, Canada 2013-2019 Ph.D. Candidate
Honours and Awards:	Canadian Acoustical Association: Alexander Graham Bell Prize in Speech Communication and Hearing 2015 Canadian Academy of Audiology: Outstanding Research Award 2015 Mitacs Accelerate Award 2016-2017 Sertoma Foundation of Canada Scholarship 2015, 2016, 2017, 2018 International Hearing Aid Research Conference: Graduate Student Scholarship 2018 Province of Ontario Graduate Scholarship 2017-2018, 2018-2019
Related Work Experience	Teaching Assistant Communication Sciences and Disorders The University of Western Ontario London, ON 2016-2019

Audiology Research Intern
Audioscan, A Division of Etymonic Design Incorporated
Dorchester, ON
2016-2017

Audiology Research Intern
Unitron Hearing
Kitchener, ON
2018

Publications in Peer-Reviewed Journals:

Schutz, M. & **Vaisberg, J.M.** (2014). Surveying the temporal structure of sounds used in Music Perception. *Music Perception.* 31(3), 288-296.

Vaisberg, J.M., Macpherson, E.A., & Scollie, S. (2016). Extended bandwidth real-ear-measurement accuracy and repeatability to 10 kHz. *International Journal of Audiology.* 55(10), 580-586.

Vaisberg, J.M., Folkeard, P., Pumford., J., Narten, P., & Scollie. S. (2018). Evaluation of the repeatability and accuracy of the wideband real-ear-to-coupler difference. *Journal of the American Academy of Audiology.* 29(6), 520-532.

Glista, D., Hawkins, M., **Vaisberg, J.**, Pourmand, N., Parsa, V., & Scollie S. (2018). Sound quality effects of an adaptive nonlinear frequency compression processor with normally hearing and hearing impaired listeners. *Journal of the American Academy of Audiology.* Epub ahead of print.

Vaisberg, J.M., Martindale, A. T., Folkeard, P., & Benedict, C. (2018). A qualitative study of the effects of hearing loss and hearing aid use on music perception in performing musicians. *Journal of the American Academy of Audiology.* Epub ahead of print.

Publications in Non-Peer-Reviewed Journals:

Vaisberg, J.M., Macpherson, E.A., & Scollie, S. (2016). Comparing probe tube placements and frequency averaging in the ear canal up to 10 kHz. *Canadian Audiologist.* 3(5). Retrieved from: http://www.canadianaudiologist.ca/ comparing-probe-tube-placements-feature/

Froehlich, M., Littman, V., **Vaisberg, J.**, Folkeard, P., Parsa, V., & Scollie, S. (2017). Signia rated superior to competing products for music sound quality [White Paper]. Retrieved from: https://pro.signiausa.com/scientific_marketing/ signia-rated-superior-to-competing-products-for-music-sound-quality/

Vaisberg, J. M., Folkeard, P., Parsa, V., Macpherson, E., Froelich, M., Littman, V. & Scollie, S. (2017). Comparison of music sound quality between hearing aids and music programs. *AudiologyOnline*, Article 20872. Retrieved from www.audiologyonline.com

Relevant Presentations

Vaisberg, J. M., Scollie, S., Macpherson, E.A. (2015, October). High-frequency measurement repeatability using the probe tube method. Poster session presented at the 18th Annual Meeting of the Canadian Academy of Audiology. Sheraton on the Falls, Niagara Falls, ON, Canada.

Folkeard, P., Pumford, J., Narten, P., **Vaisberg, J.M.** & Scollie, S. (2016, April). A comparison of wRECD and RECD values and test-retest reliability. Poster session presented at the American Academy of Audiology AudiologyNOW! Conference. Phoenix Convention Center, Phoenix, AZ, USA.

Vaisberg, J., Folkeard, P., Macpherson, E., Parsa, V., & Scollie, S. (2017, September). Electroacoustic correlates of subjective sound quality for hearing aid processed music. Poster session presented at the Music for Hearing Aids Conference. University of Leeds School of Music, Leeds, UK.

Vaisberg, J. M. (2017, September). Hearing aids, music, and sound quality. Centre for Hearing Research, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany.

Vaisberg, J. M. (2017, September). Hearing aids, music, and sound quality. Medical Research Council: Institute of Hearing Research, University of Nottingham (Glasgow site), Glasgow, UK.

Vaisberg, J. M. (2017, September). Sound quality judgments of hearing aid processed music: Implications for the selection, prescription, and fine-tuning of hearing devices. Professional Development Day at the British Society of Hearing Aid Audiologists. London, UK.

Vaisberg, J., Folkeard, P., Macpherson, E., Parsa, V., & Scollie, S. (2017, October). Electroacoustic correlates of subjective sound quality for hearing aid processed music. Poster session presented at the 20th Annual Meeting of the Canadian Academy of Audiology. Delta Hotels Ottawa City Centre, Ottawa, ON, Canada.

Vaisberg, J., Folkeard, P., Macpherson, E., Parsa, V., & Scollie, S. (2017, November). Electroacoustic correlates of subjective sound quality for hearing aid processed music. Poster session presented at the 13th Annual NeuroMusic Conference. McMaster University, Hamilton, ON, Canada.

Vaisberg, J. M., Beaulac, S., Glista, D., Van Eeckhoutte, M., Macpherson, E., & Scollie, S. (2018, August). Preferred hearing aid gain settings for music-listening using a 3D modified simplex procedure implemented with the Open Source Master Hearing Aid platform. 2018 International Hearing Aid Research Conference (IHCON). Granlibakken Tahoe, Tahoe City, CA, USA.