Self-adjustment of Hearing Aid Amplification: Listener Preferences and Speech Recognition Performance

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Trevor Thomas Perry

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Peggy B. Nelson, Ph.D.

July 2019

Acknowledgements

It is difficult to find the words to properly express my gratitude, appreciation, and respect for Dr. Peggy Nelson. This dissertation would not exist save for her mentorship, ideas, encouragement, and support. Her seemingly-unending generosity, positivity, and creativity inspire me. Thank you, Peggy, for being my advisor.

I have been fortunate to have had incredible mentors throughout my graduate studies. In particular I am grateful for the members of my committee, Drs. Bert Schlauch, Benjamin Munson, and Evelyn Davies-Venn, and for their thoughtful guidance and respectful critiques, as well as for their patience, humor, honesty, and solidarity.

I would like to thank Alix Klang, Kendra Day, Danyi Ma, as well as Drs. Dianne Van Tasell, Melanie Gregan, Andrew Sabin, Coral Dirks and Joseph Hinz for their contributions to study design, data collection, and technological assistance. This research would not have been possible without them.

This dissertation project was made possible by funding from the Leslie E. Glaze fellowship, the CAPCSD PhD fellowship, and the National Institutes of Health (NIDCD R01-DC008306 to Peggy Nelson).

Many thanks to the staff of Shevlin and CATSS. I'd like to extend a special thank-you to Amanda Greenhart, Liz Anderson, and Andrew Byrne for logistical and technical support.

I am extremely appreciative of the charity of the people who volunteered their time to participate in the research. I am humbled by their compassion and selflessness.

Thank you to all the graduate students – past and present – who cheered and encouraged me, commiserated with me, and gave me advice and friendship. Thank you in particular to Drs. Adam Svec and Tess Koerner for patiently answering my anxious questions and listening to my rambling thoughts.

I am also grateful to my first mentor in science, Dr. Bomjun Kwon, and how fiercely he advocates for me and my efforts. I am lucky to have such a tireless champion and friend.

I would be remiss if I did not also thank my family and friends for supporting me through the challenging times and celebrating my successes. Thank you, most of all, to my marvelous fiancé, Joshua Jung, for being my bright beacon of joy, my shoulder to cry on, my goofy comedian, and my unshakeable bedrock of support and understanding. As I finish writing each chapter of this dissertation, I am excitedly anticipating starting a new chapter of our life together.

Abstract

Self-adjustment of amplification parameters is a potential method for improving satisfaction with hearing aids, particularly in noisy environments. People with mild-to-moderate hearing loss adjusted gain parameters in quiet and in several types of noise by using a simple touchscreen interface to control a research device which emulated the basic functionality of a digital hearing aid in real time. Results of self-adjustment indicated reliable individual preferences but a great deal of between-listener variability, indicating that people have stable preferences for amplification and are able to select preferred parameters consistently.

The large individual differences suggest that preferred gain configurations can differ greatly from prescriptive settings in both quiet and in noise and underscore the need for an efficient method of customizing amplification parameters beyond prescribed settings. Audiological listener factors such as age, hearing loss, and experience using hearing aids, predicted little of the between-listener variability. It is unlikely that modifications to prescriptive fitting formulae based on the factors examined here would result in amplification parameters that are similar to user-customized settings. Most self-adjustments were completed in only a minute or two, demonstrating that self-adjustment is a rapid and efficient method for matching hearing aid output to preferred settings.

When self-adjustments were made with speech presented at average conversational levels, gain adjustments did not strongly affect speech recognition within the range of signal-to-noise ratios tested. For speech at a lower presentation level, preferences for amplification were related to speech recognition performance, suggesting

that listeners include their subjective sense of speech clarity among their criteria for selecting amplification parameters during self-adjustment. Self-adjusted amplification was overwhelmingly rated as satisfactory or very satisfactory and as producing a comfortable loudness. Taken together, the results of these experiments support the conclusion that for people with mild-to-moderate hearing loss, self-adjustment is likely to produce satisfactory and comfortable amplification that provides speech recognition comparable to that of hearing aids fit according to current clinical best practices.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

I. Overview

A 2016 study organized by The National Academies of Sciences, Engineering, and Medicine reports that approximately 30 million Americans have hearing loss, but between 67 to 86 percent of adults who could benefit from hearing aids do not use them (National Academies of Sciences, 2016). Hearing loss impedes communication by causing portions of the speech signal to be inaudible to the person with the loss and increasing the difficulty of communication. Hearing aids provide amplification in the frequency regions of the wearer's hearing loss, thereby increasing the audibility of speech. Amplification from hearing aids can benefit people with hearing loss by increasing ease of communication and mitigating the negative consequences associated with hearing loss, including reduced quality of life, social withdrawal and isolation, and an increased risk of dementia, depression, and falls (Mulrow et al., 1990; Rumalla et al., 2015; Rutherford et al., 2017).

In light of the benefits of amplification and the high proportion of people with hearing loss who do not use hearing aids, addressing the reasons for the low rates of hearing aid adoption and use is an important health care goal. Multiple factors contribute to the low rate of hearing aid non-adoption (Knudsen et al., 2010). Some factors relate to access and provision of hearing health care, such as the low rates of adults 65 years and older who receive hearing screenings (Kochkin, 2009) and a need for audiological services that exceeds the current capacity of the hearing health care system, especially for historically-underserved populations (Margolis and Morgan, 2008; Swanepoel et al., 2010). Other factors are centered on the person with hearing loss and their beliefs, such as questions or concerns about the cost and effectiveness of hearing aids, stigma surrounding hearing aids, as well as the individual's perception of their hearing loss and hearing ability (Amlani, 2016). Approaches to increasing hearing aid adoption will likely need to be as multifaceted as the contributing causes of non-adoption.

Among people who own hearing aids, not all regularly wear them. Survey results suggest that as many as 1 in 5 to as few as 1 in 20 hearing aid owners never wear their hearing aids (Hougaard and Ruf, 2011; Sorri et al., 1984). Among a group of older Australian adults, about 40% reported that they use their hearing aids for less than 1 hour per day (Hartley et al., 2010). McCormack and Fortnum (2013) conducted a scoping study and highlighted several reasons why people who own hearing aids do not use them. Across the reviewed literature, the most consistently and commonly cited reasons for non-use involved dissatisfaction with the speech clarity, sound quality, and perceived benefit from wearing the hearing aid. In particular, dissatisfaction with hearing aids in noisy situations was common.

The issue of reduced hearing aid benefit and satisfaction in background noise is a long-standing one. Plomp (1986) presented a mathematical model that described changes in the speech recognition performance of listeners with hearing loss as the level of noise and amount of amplification gain varied. The model demonstrates that for people with sensorineural hearing loss, hearing aids provide the most benefit to speech recognition in quieter backgrounds, and as the signal-to-noise ratio (SNR) worsens, the benefit also decreases. Hearing aids are most suited for amplifying low-intensity speech in quiet or highly favorable SNRs. In such cases, audibility is limited by the hearing acuity of the listener, and the amplification improves audibility by placing more of the speech energy in the dynamic range of the auditory system. In noisy backgrounds, however, as the SNR approaches parity, noise energy (rather than the listener's hearing thresholds) becomes the primary constraint on audibility, and amplification does not improve audibility.

Modern digital hearing aids often include signal processing strategies, such as digital noise reduction or directional microphones, to try to improve the performance of the device in noisy situations. There is little evidence that existing noise-reduction strategies produce meaningful improvements in speech intelligibility, but they may reduce the annoyance of noise (Brons et al., 2014; Hu and Loizou, 2007). Directional microphones can result in large improvements in speech recognition in controlled laboratory testing, but in daily life the benefits to directional modes might be more modest (Bentler, 2005; McCreery et al., 2012). Despite the wide availability of these features in digital hearing aids, dissatisfaction with listening in noise persists among people with hearing loss.

One proposed method for improving the satisfaction with hearing aids is self-adjustment of amplification gain and compression (Elberling and Hansen, 1999; Schweitzer et al., 1999). Self-adjustment of amplification parameters is part of a general category of approaches based on increasing the involvement of the hearing aid wearer in determining the functioning and output of the hearing aid (Boymans and Dreschler, 2012; Keidser and Alamudi, 2013; Kuk and Pape, 1993; Punch and Parker, 1981; Zakis et al., 2007). The rationale is that a portion of the dissatisfaction with hearing aids could be attributed to a mismatch between the output of the device and the preferred listening levels of the wearer. Through self-adjustment, the wearer could change the gain-frequency response and compression parameters to better suit their subjective preferences. Hearing aids are typically fit according to a prescriptive formula designed to increase the audibility of speech in quiet backgrounds, but this might not produce desirable settings for some listeners when in noisy environments.

Based on Plomp's (1986) SNR model of speech reception thresholds, it is unlikely that self-adjustments made in noise would result in improved speech recognition compared to an audiologist-fit hearing aid. Without a reliable method of isolating and suppressing the interfering background sounds, changes to amplification alone are not predicted to result in higher speech clarity. However, self-adjustments made in noise (and in quiet) may still lead to improved satisfaction by enabling the wearer to address other subjective features of the hearing aid output, such as sound quality, comfort, and annoyance.

Beyond addressing dissatisfaction with hearing aid performance in noisy situations, self-adjustment could be used as a method for individualizing the amplification characteristics of self-fitting hearing aids. Self-fitting hearing aids are hearing aids sold as stand-alone products and are intended to be configured by the user without the help of a hearing care professional (Convery et al., 2011). Such devices have the potential to increase hearing aid adoption rates and usage by increasing the wearer's sense of hearing self-efficacy and reducing the cost and access barriers which may deter some people from obtaining hearing aids (Keidser and Convery, 2016). According to a survey of adults with hearing loss who chose not to adopt hearing aids, 1 out of 5 non-adopters would be more likely to purchase a hearing aid in the future if they could fit or adjust the hearing aid themselves (Kochkin, 2007). In the United States, the potential user base of self-fitting hearing aids is likely to be largely comprised of people with mild-to-moderate sensorineural hearing loss (Congress, 2017).

The initial fitting of self-fitting hearing aids can be accomplished using automated, self-contained in situ pure tone audiometry, the results of which are used to set amplification parameters to estimates of prescriptive targets. Afterward, self-adjustment of these parameters would be used to customize the fit to match the preferred gain-frequency response of the wearer (Keidser and Convery, 2016). It is also possible that self-adjustment could be used with the device initially set to a generic default fit without the need for audiometry and a prescriptive fit.

One concern about the use of self-adjustment of hearing aid gain is that individuals with hearing loss will select levels of gain that cause decreased satisfaction, reduced hearing aid usage, and/or reduced benefit compared to a hearing aid fit according to a widely-used, evidence-based prescriptive formula such as NAL-NL2 (Keidser et al., 2011). In one such scenario a user of self-adjustment may naively believe that to make speech clear it needs to be made loud, and thus self-adjust the gain to be too high, causing fatigue and discomfort which could ultimately lead to rejection of the hearing aid. In another scenario, self-adjustment might be used to pick insufficient gain to restore the audibility of speech. In this case the wearer has sacrificed some potential communication benefits to achieve a more subjectively comfortable or desirable sound from the device.

However, this could result in rejection of the hearing aid due to a lack of perceived benefit. Of course, if the wearer has ongoing access to self-adjustment tools, they could change the amplification parameters to address gain that is set either too high or too low. Furthermore, for speech in noise, changes to gain are unlikely to result in large changes in speech audibility (Plomp, 1986).

There is a body of literature describing preferred listening levels as evaluated through a variety of research methods, such as paired comparisons, individual unpaired ratings, inspection of hearing aid output after volume control adjustment, and self-adjustment of amplification (Boymans and Dreschler, 2012; Keidser et al., 2005; Kuk and Pape, 1993; Mackersie et al., 2018; Smeds et al., 2006; van Buuren et al., 1995). However, existing studies have a number of limitations that may reduce their relevance to how self-adjustment is likely to be used in a commercially-available application. Some of these limitations include a constrained range of gain-frequency response changes available to the user, a lack of explicit assessment of preference for self-adjusted parameters, and dissimilarities between the tools and methods used in previous research and how self-adjustment is most likely to be implemented clinically, namely, real-time control of a portable, on-the-ear compression-amplification system using a simple visual user interface on a touchscreen device.

In order to better understand the consequences of self-adjustment of hearing aid gain and its potential to increase hearing aid satisfaction and use, there remains a need to understand what amplification settings are likely to be selected, what motivates or underlies the selection of settings, and how the self-adjusted settings compare to prescriptive settings in terms of preference and speech recognition performance. The purpose of the current work is to describe the use of self-adjustment of amplification by people with mild-to-moderate hearing loss and compare self-adjusted amplification settings with prescriptive settings in terms of gain, speech recognition performance, and listener preference. This information has clinical implications for the use of self-adjustment of amplification in both conventional and self-fitting hearing aids. The specific research questions of each study are described below.

II. Research Questions

Study 1: Self-adjustments of Amplification in Noisy Backgrounds

1. Across listeners with mild-to-moderate hearing loss, does the average gain of a self-adjusted fit differ from that of a prescriptive fit for adjustments made while listening to conversational-level speech in a quiet background?
2. Upon retest within the same listening conditions, how consistent are self-adjustments of amplification?
3. What is the magnitude of the inter-subject variability of self-adjusted gain relative to prescribed gain?
4. Do self-adjusted fits result in reduced speech recognition performance compared to audiologist-fit settings?
5. Does the addition of background noise influence the amount of gain selected during self-adjustment?

6. What effects, if any, do the noise level and long-term average spectrum have on how much gain is selected during self-adjustment?

Study 2: Between-participant variability and listener factors in self-adjustment of amplification

1. How much time does it take for the average participant to make a self-adjustment?
2. What influence does the level of background noise have on the time taken to complete self-adjustment?
3. Can listener characteristics or listener engagement predict how self-adjusted gain differs from prescribed settings?

Study 3: Self-adjustments for low-level speech in quiet and their relations to preference and speech understanding

1. What effect, if any, does the intensity level of the speech presented during self-adjustment have on the amount of gain in self-adjusted fits?
2. Do listeners prefer their self-adjusted settings more than prescriptive settings?
3. Do listeners prefer their self-adjusted settings more than settings that provide less speech audibility?
4. Under conditions in which changes to gain can produce substantial changes to speech audibility, what relationship, if any, is there between the preference for amplification settings and the speech recognition performance achieved when using those settings?

Chapter 2: Self-adjustments of amplification in noisy backgrounds

Sections I-V are reprinted from:
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, 2331216518798264. doi:10.1177/2331216518798264.

I. Introduction

Hearing-aid fitting formulae typically have been designed to improve the audibility of speech in quiet settings (e.g., Johnson, 2013) by applying a gain prescription formula based on hearing thresholds. Recent surveys (e.g., Kochkin, 2012) show hearing-aid users are very satisfied with their hearing aids for understanding speech in quiet. However, it is unusual for a hearing-aid user's experience to only include quiet listening situations. Noisy environments are ubiquitous, and the same surveys that show satisfaction with quiet performance also show that there is much room for improvement in noisy surroundings (Kochkin, 2012). Modern hearing aids employ multiple programs and "noise reduction" algorithms in an attempt to increase comfort in noisy situations, by decreasing the amount of gain in frequency bands where noise dominates. A variety of algorithms— including Wiener filtering—is used to estimate signal-to-noise ratio (SNR) in different frequency bands. The effects of these gain changes are not fully understood.

Preferred gain-frequency responses for hearing aids have previously been investigated using a variety of paradigms. Comparison or rating methods require listeners to make judgments about their preference for or perception of sounds after amplification with different gain-frequency responses, either as paired comparisons (Amlani and Schafer, 2009; Byrne, 1986; Keidser et al., 1995, 2005; Kuk et al., 1994, 1994; Kuk and Lau, 1995, 1996b; Kuk and Pape, 1992, 1993; Moore et al., 2011; Neuman et al., 1987;

Preminger et al., 2000; Punch et al., 2001; Punch and Howard, 1978; Punch and Parker, 1981; Smeds, 2004; Stelmachowicz et al., 1994) or individual, unpaired ratings (Kuk and Lau, 1996a; van Buuren et al., 1995). Another approach is to use adjustment methods which entail assessing the output of a hearing aid after it has been adjusted—often using the volume control—to better match the preferred listening level (Boothroyd and Mackersie, 2017; Boymans and Dreschler, 2012; Cox and Alexander, 1991, 1992; Dreschler et al., 2008; Hornsby and Mueller, 2008; Horwitz and Turner, 1997; Humes et al., 2002; Keidser, Dillon, et al., 2008; Marriage et al., 2004; Polonenko et al., 2010; Smeds et al., 2006; Souza and Kitch, 2001) or by analyzing the output of trainable hearing aids after completion of a training regime (Keidser and Alamudi, 2013; Mueller et al., 2008; Zakis et al., 2007).

Although a few studies indicate that average preferred gain is similar to gain fit according to a clinical formula (Hornsby and Mueller, 2008; Polonenko et al., 2010), a common trend in the literature is that listeners with hearing loss generally prefer less overall gain than their formula-fitted settings (Boymans and Dreschler, 2012; Humes et al., 2002; Keidser and Alamudi, 2013; Smeds, 2004; Smeds et al., 2006). For studies in which gain in the high and low frequencies were varied separately, a common pattern is that listeners typically prefer less high frequency (>1000 Hz) gain than their fitted settings and more low-frequency gain than their fitted settings (Boothroyd and Mackersie, 2017; Kuk and Pape, 1992, 1993; Moore et al., 2011; Preminger et al., 2000; Zakis et al., 2007). However, the opposite of this trend has also been reported, with listeners preferring less gain in the low frequencies and more gain in the high frequencies

(Punch et al., 2001). For listening in noise, the spectral characteristics of preferred gain may depend on the spectrum of sound in which preference is assessed, such that listeners tend to prefer the gain-frequency responses which reduce gain in spectral regions containing relatively higher levels of noise (Keidser et al., 1995, 2005).

The presence or absence of competing sounds may itself influence listener preferences for hearing-aid gain. Due in part to differences in study methods and reporting, the relationship between preferred gain in noise and preferred gain in quiet is somewhat unclear. Cox and Alexander (1991) found that listeners preferred less gain in a noisy or reverberant environment than in a quiet environment, but the level of speech and the overall level in each environment differed substantially, introducing a potential confound. Likewise, Keidser et al. (2005) assessed the preferred spectral tilt of the gain-frequency response across a variety of noise conditions and found that listeners preferred less gain when SNRs were poor. However, the SNRs used did not vary independently of the presentation level. Both studies are consistent with a preference for lower gain at higher listening levels. Indeed, studies which include some variation of input level generally find that listeners prefer less gain as levels increase, which is consistent with the normal operation of compression gain (Kuk and Pape, 1993). Other studies which vary the characteristics of the listening environment without large changes on the input listening level have reported only small differences in gain preference for environments that differ in amount of reverberation or noise (Kuk and Pape, 1992, 1993; Stelmachowicz et al., 1994).

When assessed and reported, the stability, or test–retest reliability, of gain-frequency response preferences appears to differ with the method and materials used. Several studies report that listeners with hearing loss show better consistency in preferred gain-frequency response when listening in noise than in quiet (e.g., Keidser et al., 2005; Kuk and Pape, 1992: 2; Stelmachowicz et al., 1994). Byrne (1986) found that the reliability of intelligibility and pleasantness judgments made by listeners with hearing loss depended on the presence of noise during evaluation, with greater reliability of intelligibility judgments in quiet than in noise and greater reliability of pleasantness in noise than in quiet. In general, the literature indicates that most individuals are moderately consistent in their judgments and preferences for gain-frequency responses across repeated assessments, with most listeners arriving at the same or a similar result upon retest (Kuk and Pape, 1992) or preferring a single-frequency response over nearly all others (Byrne, 1986). Test–retest Pearson correlation coefficients, when reported in the literature, range from about 0.6 to about 0.8 (Boothroyd and Mackersie, 2017; Punch and Parker, 1981) and within-subjects test–retest standard deviations (when reported) are typically 5 dB or less (Dreschler et al., 2008; Keidser et al., 2005). Many different gain-frequency responses produce similar speech recognition outcomes (van Buuren et al., 1995), and estimates of within-subject consistency may reflect that some listeners are willing to accept many gain-frequency responses as preferable (Keidser et al., 2005; Kuk and Lau, 1996a).

When between-subjects variability for preferred gain is reported, it can be substantial (e.g., Hornsby and Mueller, 2008). The preferences of individual listeners can deviate greatly from average trends, and a need to match hearing-aid gain to the preferences of each hearing-aid user provides strong motivation for using self-adjustment technology to investigate preferences for hearing-aid amplification.

It is not clear that hearing professionals know what a hearing-aid user would choose as a gain profile in quiet and noisy situations. That is, would a user choose to have the gain decreased for added comfort in noise but reduced audibility and speech understanding? Alternatively, would they prefer to tolerate a bit more noise in the hopes of improving their understanding of the speech signal? Along those lines, would their preferred adjustment for different environments be the same across hearing-impaired (HI) listeners or would it vary across listener? Given the heterogeneous nature of the HI population, it is hypothesized that the latter would be true, but this has not been directly tested.

This study used self-adjusting simulated hearing aids to determine user-selected gain settings for a group of HI listeners with varying degrees of hearing loss in several noisy settings. The self-fitting process has been used previously to determine listeners' ability to select gain (e.g., Keidser and Alamudi, 2013; Wong, 2011). While self-fitting hearing aids have been tested as a means of getting much-needed amplification to HI listeners in developing countries (e.g., Convery et al., 2011), this article examines the use of self-adjustment or fitting to find preferred settings in varying quiet and noisy conditions. In the current experiment, self-adjustment was used to determine listener preference and performance specifically for listening to speech in quiet and in noise. Findings can inform audiology practice. If most listeners set their gain to a lower (or higher) level in the presence of noise, it would argue that automatic gain changes would be satisfactory, and preset noise-reduction algorithms would satisfy most users. The data, then, could inform the details of proposed gain settings for noisy conditions. However, if different listeners set their gain differently for a given listening condition, this would suggest that preset noise-reduction programs are not ideal, and that self-adjustment is a valuable tool with which to quickly and accurately find a HI listener's uniquely preferred settings. In addition, the results of speech recognition testing can provide information about listeners' potential trade-off between comfort in noise and intelligibility. An important secondary question is whether listeners sacrifice intelligibility for comfort when given self-adjustment options.

II. Methods

Data are presented from 30 listeners aged 59 to 78 years with mild to moderate hearing loss who self-adjusted amplification parameters in laboratory-simulated restaurant environments.

A. Recordings (Background Noise)

Noisy restaurant conditions were chosen because they are among the most challenging environments for hearing- aid users and are the source of dissatisfaction for many (e.g., Kochkin, 2012). Therefore, recordings were made of three local area restaurants during the lunch hour, along with a quiet conference room to mimic a "living room" setting. Stereo recordings approximately 5 min in length were made of the background noise in each restaurant. The set-up for each set of recordings was a Schoeps CMC6 MK4 stereo cardioid microphone in an ORTF configuration (two cardioid microphones spread to 110°, after Killion, 1979) and a Roland R-4 portable sound recorder with 24-bit quantization and a 48 kHz sampling rate. Long-term spectra of the restaurant noises are shown in Figure 2.1. Sound levels were naturally varying.

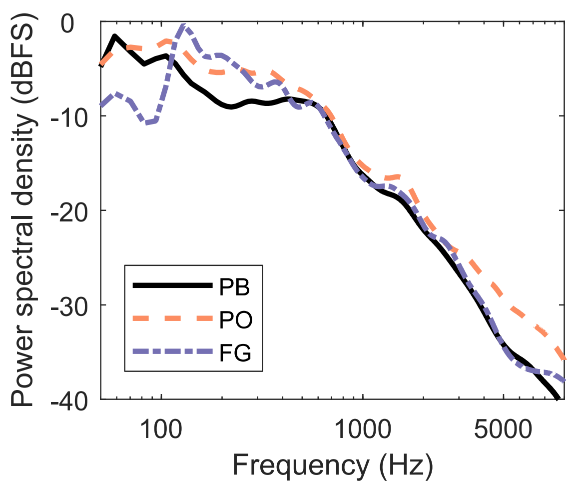

Figure 2.1. Long-term average spectra of the three restaurant recordings. The steady PB noise had the same long-term spectrum as the PB recording.

B. Laboratory

Laboratory characteristics include a 10′ × 13′ × 8.5′ double-walled sound chamber, 48 speaker array (Anthony Gallo Acoustics—A'Diva ti speakers), 24 Crown XLS 1500 power amplifiers, and 3 Lynx Aurora 16 D/A converters. The experimental routine was run on a Dell desktop computer running Matlab.

C. Stimuli (Room Noise)

The recorded binaural room recordings were spatialized to the 48-channel loudspeaker system by presenting the left portion of the signal to all loudspeakers on the left hemisphere and vice versa for the right portion of the signal. For six loudspeakers along the interaural axis, both the left and right signals were summed, and the resulting amplitude divided by half before it was presented to these six loudspeakers. A steady noise with the same frequency spectrum as the PB restaurant was included as an additional noisy environment. In this condition, the PB steady-state noise was played through the entire 48-channel loudspeaker array but without any spatial processing applied in order to approximate a diffuse noise environment. Stimuli (Connected Speech Test—Target Speech) Recordings from the Connected Speech Test (CST; Cox et al., 1987) were used in this study. During gain adjustment, the speech stimuli consisted of 30-s CST passages spoken by a female talker presented on a loop. To make the speech stimuli seem to originate from the same room as the background noise, the speech stimuli were spatialized to match the measured restaurant sizes (see later) and the estimated reverberation times of the recorded rooms.

This gave the desired effect in that a listener in that soundfield was surrounded by the restaurant- (or living room-) recorded stimuli, similar to how they would be if they were actually seated in the middle of the restaurant (or quiet room). The room dimensions used for spatialization were as follows:

- Restaurant 1 (FG): 58′ × 24′ × 9′
- Restaurant 2 (PB): 38′ × 30′ × 25′
- Restaurant 3 (PO): 80′ × 56′ × 13′
- Conference room (“living room”): 16′ × 14′ × 9′

This process used a virtual room model to simulate reflections in an acoustic space. This room simulation used an image-based model (Allen and Berkley, 1979) to calculate 10,000 individual reflections for each room based on a source to receiver distance of 1.3 m. Custom software then assigned the calculated reflections to an appropriate loudspeaker in the sound booth using appropriate timing and power adjustments based on the inverse-square law. The resulting set of 48 impulses responses was then convolved with the speech stimuli, combined with the matching background noise, and played out through a 48-channel loudspeaker array. This processing strategy attempted to approximate a listener’s experience of being seated in the middle of a restaurant (or quiet room) and listening to the female talker at a short distance.

In other words, the CST spatialized to match the PB restaurant was only used when samples of the PB noise were played, and the same was true for the other background noises. A steady noise with the same frequency spectrum as the PB restaurant was included as an additional noisy environment. The PB steady-state noise was played through the entire speaker array but without spatialization.

D. Simulated Hearing Aid—Ear Machine

Self-adjustment was used to determine the preferred settings of the listeners. For our study, listeners used a mobile application developed by Ear Machine LLC, running on the Apple iOS platform and implemented on an iPod Touch (fourth generation). The device was coupled to the listeners' ears using Etymotic foam ear tips. The sound in the booth was picked up by the microphone of the iPod and delivered diotically to both ears. The iPod was held in front of the listener at approximately chin height. The application was designed to simulate a nine-channel hearing aid with slow-acting compression. Specifically, the application used a nine-channel multiband wide-dynamic range compressor or limiter with fast attack (approximately 1ms) and slow release (approximately 500ms) times. Compression center frequencies were as follows: 125, 500, 1000, 1500, 2250, 3250, 4625, 6750, and 15025 Hz. Compression ratios ranged from 1:1 to 2.3:1 for 90% of the possible settings (maximum compression ratios up to 5:1 were possible). The app included a 12-band equalizer, and the signal processing (proprietary) was designed to provide a close match to a commercial hearing aid.

There were two controllers on the screen of the iPod that functioned as wheels which the listeners could turn up or down (earmachine.com). One controller was labeled "loudness," and the other was labeled "fine tuning." The loudness controller changed gain, compression, and limiting parameters in all nine compression channels simultaneously based on fits to multiple audiograms from the NHANES database (https://wwwn.cdc. gov/nchs/nhanes/Default.aspx). The fine tuning controller changed frequency tilt by changing overall frequency response in the 12 equalization bands: As the controller wheel was moved upwards, frequencies above 2 kHz were emphasized and frequencies below 2 kHz were de-emphasized. Moving the wheel downward had the opposite effect.

E. Calibration

Calibration was done using noises that were equivalent to the long-term average spectra (digital RMS) of the various background noises. They were not spatialized. In addition, the CST noise (included on the CD) was used to calibrate the CST passages. A chair in the laboratory was set approximately 3 feet from the 0° azimuth speaker (where the CST was played). A sound level meter was held in the approximate head position (head absent) as each calibration noise was set to 65 dBC.

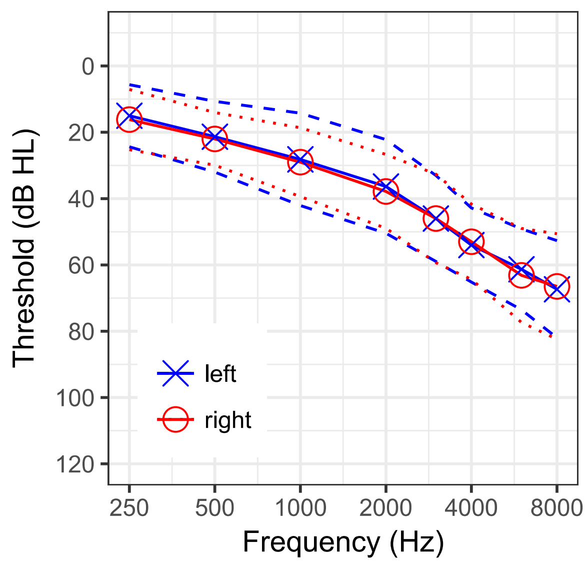

Figure 2.2. Mean participant audiograms for left and right ears of participants in Studies 1 and 2. The dashed blue lines and dotted red lines indicate 1 standard deviation from mean thresholds for left and right ears, respectively.

F. Participants

Listeners who participated in the study included 30 adults with symmetric mild to moderate sensorineural hearing loss (see Figure 2.2 for the average audiograms). The average age was 70 years, with ages ranging from 59 to 78 years. Seventeen of the participants were male. Twelve participants were new users of hearing aids; the others had used hearing aids for varying durations from 1 year to 25 years.

G. Listener instructions

Each listener was instructed that the goal of the task was to turn the wheels on the iPod until the talker's voice (i.e., CST passages) was as clear as possible in the background noise. They were asked to adjust each wheel separately but were told they could adjust each wheel as much as they wanted to and in any order.

Several listeners had participated in previous trials during pilot testing, so no practice was required for them. New listeners practiced in the booth during seven actual trials, with an experimenter standing close by to answer any questions.

Each experimental trial began with the user's custom prescriptive settings of National Acoustic Laboratories (NAL)-NL2, derived from the stand-alone clinical software and verified using real-ear speech mapping techniques. This was chosen as the default position because it has been shown more than once that the starting settings of a self-adjust device dictate to some extent the end configuration (e,g, Dreschler et al., 2008; Keidser et al., 2008; Mueller et al., 2008). Once the experimental trials began and the listener had adjusted the wheels such that they were satisfied that they could hear the

CST voice as clearly as possible, they indicated this by tapping a star-shaped virtual button on the user interface, ending the current trial. This stopped sound playback momentarily and sent the preferred settings via Internet to a virtual server system. There were a total of 34 trials with the CST stimuli set to 65 dBC (2 repetitions per 4 noises; 4 SNRs: −10, −5, 0, and +5; and 2 quiet living room repetitions). Noise levels varied in order to achieve the desired SNR for each condition.

H. Speech recognition testing

Speech understanding was assessed using Harvard or IEEE sentences (IEEE, 1969) spoken by a female talker. The talker's voice was presented at 65 dBC from the front speaker in the presence of diffuse steady noise (i.e., presented through the entire 48 speaker array) which had the same long-term spectrum as the PB restaurant noise. Subjects listened through the iPod running the Ear Machine app, as they did during the self-adjustment trials. However, the gain and compression was locked at either that subject's NAL fit or at the self-adjusted settings which had been previously selected by each subject for the corresponding listening condition. Conditions included quiet (living room), −10, −5, 0, and +5 dB SNR. Participants repeated the sentences and they were scored by the experimenter. Two lists were presented per condition, for a total of 100 key words per condition.

III. Results

A. Gain Adjustment Results

Self-adjustment resulted in a wide range of insertion gains for a 65 dB SPL input. Large intersubject variability in self-adjusted fits was seen in each listening environment, including the quiet (living room) environment. Two examples of NAL fits and the self-adjusted fits selected by subjects in the quiet environment are shown in Figure 2.3. Gain for each band is shown as prescribed by NAL (dashed lines) and as selected by each individual user (solid lines).

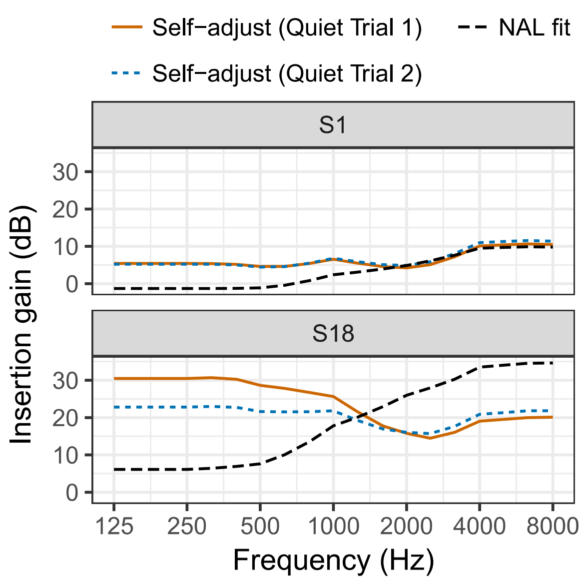

Figure 2.3. Examples of insertion gain resulting from self-adjustment in quiet for two subjects (S1 and S18). Black dashed lines indicate insertion gain for the subject’s NAL fit. Solid orange lines are insertion gain resulting from the first self-adjustment in quiet. Dotted blue lines indicate gain resulting from a second self-adjustment in quiet. Data from S1 (upper panel) exemplify pattern of consistency in self-adjustment between repetitions of self-adjustment, while data from S18 (lower panel) reflect common pattern for subjects to reduce high-frequency gain relative to their NAL fit.

To summarize the data for all subjects, insertion gain for the NAL fit for each subject was averaged into a low-frequency band (125, 250, 500, and 1000 Hz) and a high-frequency band (2000, 3000, 4000, 6000, and 8000 Hz). Insertion gain for each self-adjusted fit was also averaged into the same low- and high-frequency bands. Figure 2.4 shows the NAL fit and self-adjusted fits obtained in quiet for each subject, averaged according to frequency. Most subjects had little or no hearing loss in the low frequencies, and thus most of the NAL fits had little or no insertion gain at low frequencies (mean = 2.4 dB). In contrast, in the quiet environment, subjects often selected some insertion gain at low frequencies (mean = 6.8 dB). Second, for high frequencies, many subjects selected fits with less insertion gain than the NAL fit. Of the 60 self-adjustment trials completed in the quiet environment (two trials per subject), 19 resulted in more high-frequency insertion gain than the NAL fit, while 41 resulted in less gain.

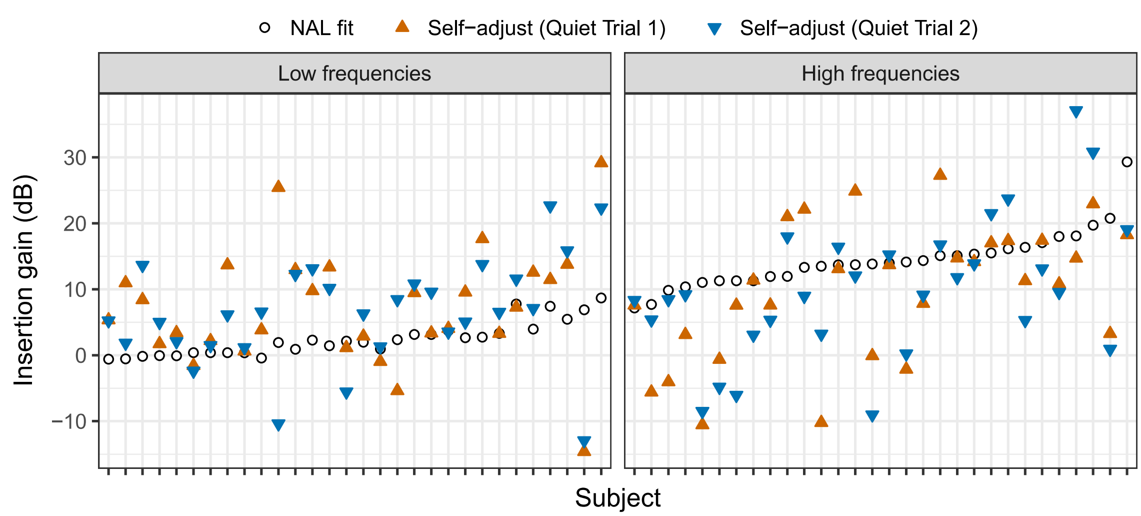

Figure 2.4. Insertion gain for NAL fit and self-adjusted fits obtained in quiet. The left panel displays average gain for frequencies up to 1000 Hz, while the right panel displays average gain for frequencies between 1000 and 8000 Hz. Subjects are ordered from left-to-right on the abscissa according to the average high-frequency insertion gain in their NAL fits. Orange and blue triangles indicate, respectively, the average gain resulting from the first and second trials of self-adjustment in quiet.

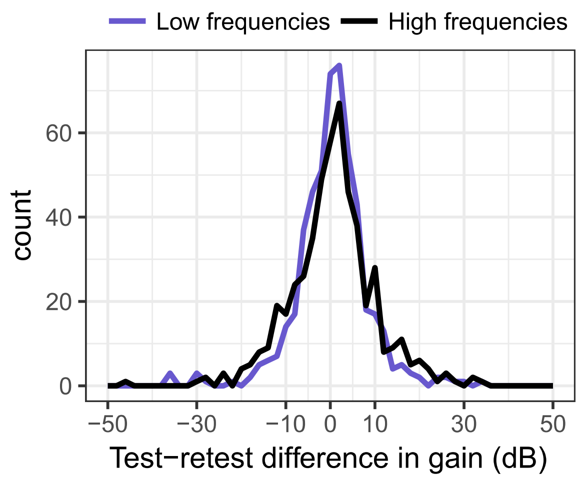

Figure 2.5. Test–retest histogram for insertion gain in the low- and high-frequency bands. Counts of absolute test–retest differences are shown in blue for the low-frequency band (125–1000 Hz), and in black for the high-frequency band (2000–8000 Hz).

Consistency in self-adjusted fits within subjects was assessed by examining test–retest reliability. The test– retest difference was defined as the difference in insertion gains between the first self-adjusted fit in a given listening condition and the second self-adjusted fit in that same listening condition. The absolute test–retest difference averaged across all trials was small (low-frequency band: 5.6 dB, high-frequency band: 6.9 dB) (see Figure 2.5, showing the test–retest histogram for low- and high-frequency bands). The test–retest correlation coefficient across both frequency bands indicated a moderately high degree of reliability, $r(1018) = 0.64$, $p < .001$. The median within-subjects standard deviation was 3.1 and 2.3 dB in the high- and low-frequency bands, respectively, for testing in the quiet condition. Median within-subjects standard deviations were similar for the noise conditions, ranging from 3.1 to 4.0 dB in the high-frequency band and from 2.2 to 3.5 dB in the low-frequency band. Across all retests, 54.6% were within 5 dB of the first self-adjusted fit, while 80.7% were within 10 dB. For subsequent analysis, insertion gains were averaged between the first and second trials in each listening condition.

To describe how self-adjusted fits differed from NAL fits, gain deviation was calculated separately in the low- and high-frequency bands by subtracting the insertion gain of the NAL fit from the insertion gain of each self-adjusted fit. A positive gain deviation indicates that the self-adjusted fit resulted in more insertion gain than the NAL fit. A negative gain deviation indicates the opposite, that is, the self-adjusted fit resulted in less insertion gain than the NAL fit. Figure 2.6 shows the deviations from NAL for gain adjustments made in quiet and in each noise environment and SNR (after averaging first- and second-trial repetitions). The data obtained in quiet are replotted in Figure 2.7 on the rightmost column of each panel.

Figure 2.7 shows the deviation from NAL for gain adjustments made at each SNR, averaged across noise types. Data from individual subjects are connected with lines, and the thick, black line indicates the average across subjects. Again, intersubject variability was notably large. Considering all noise levels and types, subjects adjusted insertion gain relative to NAL over a wide range, in both the low-frequency band (−23.3 – +24.2 dB) and the high-frequency band (−37.7 – +15.8 dB). Gain adjustments made in noise followed similar overall trends as those made in quiet. In the low-frequency band, most subjects chose more gain than NAL in the quiet environment, but as noise was added and as the level of noise was increased, self-adjusted fits tended to result in less gain with increasing SNR. On average, gain deviation from NAL in the high-frequency

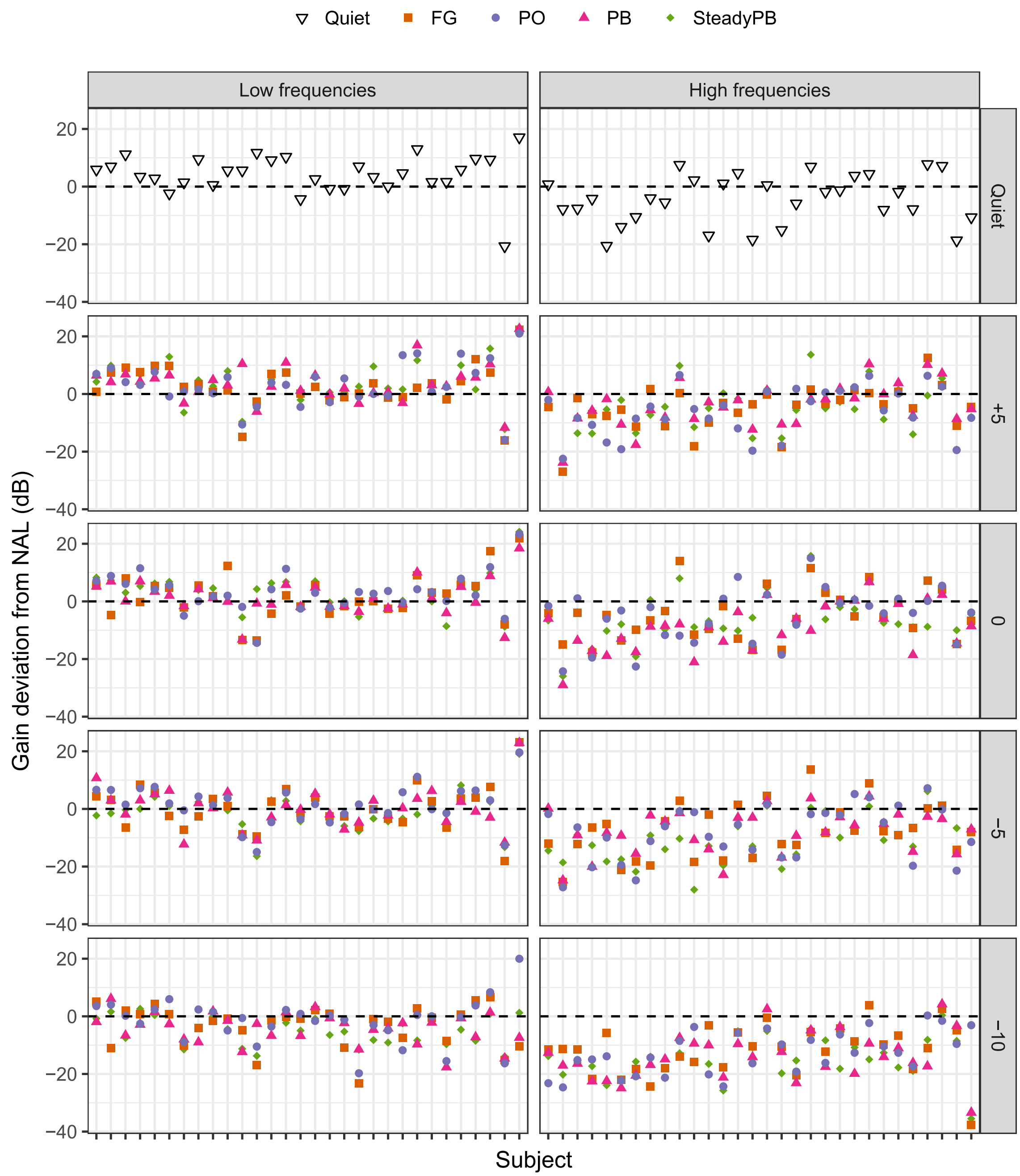

Figure 2.6. Gain deviation from NAL (self-adjusted gain minus NAL gain) for the listening conditions. Positive deviations indicate more insertion gain in the self-adjusted fit than the NAL fit. Negative deviations indicate less gain in the self-adjusted fit than in the NAL fit. Rows of plots correspond to different SNR conditions, while noise environments are indicated by marker shape and color. Gain from self-adjustment was averaged across repetitions within each condition. Subjects are ordered from left-to-right on the abscissa according to the average high-frequency insertion gain in their NAL fits (as in Figure 2.4).

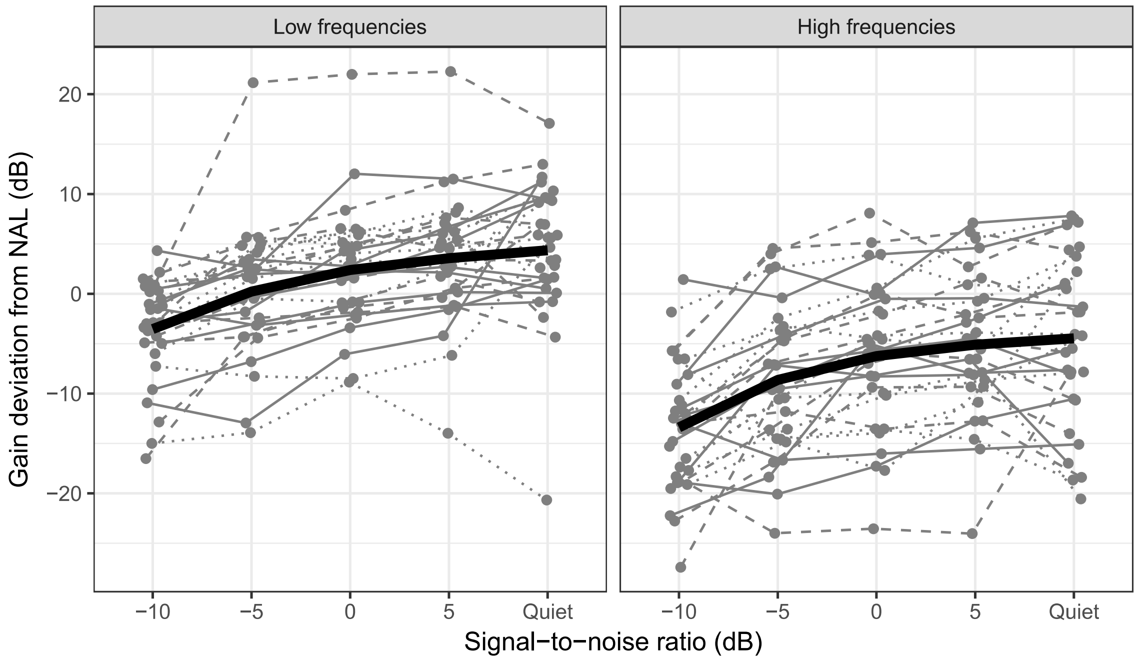

Figure 2.7. Gain deviation from NAL for individual listeners making self-adjustments in varying levels of noise. Deviations have been averaged across noise environments (within the same SNR) and repetitions. Data from low frequencies are shown in the left column; high frequencies are shown in the right. Data from individual subjects are connected with lines of varying line type; the thick, black line indicates the average deviation from NAL across subjects.

band was negative, and with increasing SNR, self-adjusted fits resulted in less insertion gain compared with NAL fits (i.e., increasingly negative deviation from NAL).

To systematically examine the influence of SNR and noise type on gain adjustment, two linear-mixed models were fit to the data in R (R Core Team, 2016) via the lme4 package using restricted maximum likelihood. One model was fit to deviation from NAL in the high-frequency band, and a second model was fit to deviation from NAL in the low-frequency band. Both models included the within-subjects factors of SNR, noise type, and repetition, and a random intercept for subject as well as a random slope for SNR per subject (included to account for any differences in the effect of SNR between subjects).Inspection of the residuals did not indicate violations of the assumptions of normality and homoscedasticity. For each model, an analysis of variance table (Type II sums of squares) and post hoc contrasts were calculated using the Kenward-Roger approximation for degrees of freedom using the lmerTest, pbkrtest, and multcomp packages. For both models, the main effect of SNR was statistically significant, high frequency: $F(4, 34.39) = 11.13$, $p < .001$; low frequency: $F(4, 34.79) = 19.25$, $p < .001$, and post hoc tests of contrasts gave evidence for statistically significant differences (all $p < .01$ for both the high- and low-frequency models) in deviation from NAL between proximal SNR conditions (i.e., between quiet and +5 dB SNR, between +5 dB SNR and 0 dB SNR, and so on). This confirms that subjects tended to select less and less gain as the noise level increased, across the different listening environments, as seen in Figure 2.7.

The main effect of repetition was not statistically significant in both the high-frequency, $F(1, 866) = 1.64$, $p = .20$, and low-frequency models, $F(1, 866) = 1.39$, $p < .24$, indicating no detectable bias across listeners between first- and second-trial repetitions.

The main effect of noise type was statistically significant for both the high-frequency, $F(3, 866) = 3.79$, $p = .01$, and low-frequency models, $F(3, 866) = 3.33$, $p = .02$. Post hoc tests of specified contrasts indicated no difference in deviation from NAL between the PB restaurant noise and the steady PB-spectrum noise in both the high-frequency ($p = .90$) and low-frequency ($p = 1.00$) bands. Gain changes in the FG restaurant noise were significantly different from changes made in the PB, PO, and steady noise conditions in the high-frequency band only ($p = .04$), indicating the subjects tended to make smaller magnitude adjustments to high-frequency gain (i.e., have negative deviation from NAL that is closer to 0) in the FG noise than in the other noises by about 1.4 dB, averaged across SNR, repetition, and subject. Gain changes in the PO restaurant noise were significantly different from changes made in the PB restaurant and steady noises in the low-frequency band only ($p = .02$), indicating that subjects tended to increase low-frequency gain more in the PO noise by about 1.3 dB, averaged across SNR, repetition, and subject. In general, the magnitude of the differences between noise types was small in terms of dB, suggesting that subjects made small, yet consistent, alterations to their gain in response to the listening environment. The finding of no statistically significant difference between the two noise conditions with the same long-term average spectrum (PB restaurant and PB-spectrum steady noise) is consistent with the notion that gain preferences in noisy environments are related to the noise spectra. Because variation in gain adjustments across noise types was small, and in order to generalize across noise environments, for subsequent analyses, the data from the four noise types were averaged together.

To evaluate whether subjects were consistent in their gain adjustments as noise levels changed, bivariate correlations were calculated between gain deviation from NAL within each frequency band for adjustments made in different SNRs, and p values were corrected for multiple comparisons using the Benjamini-Hochberg procedure. The results are displayed in Table 1. Correlations were robust overall, indicating that subjects tended to make consistent gain adjustments across differing levels of noise. Correlation coefficients were highest between conditions in which noise levels were most similar, and when the SNRs represented moderately noisy environments (0 and +5 dB SNR).

Table 2.1. Correlation coefficients for gain adjusted at different noise levels.

SNR (dB)	5	0	−5	−10
High Frequency Band				
Quiet	0.72	0.66	0.63	0.44
5		0.92	0.86	0.47
0			0.91	0.54
-5				0.68
Low Frequency Band				
Quiet	0.68	0.58	0.53	0.31
5		0.95	0.92	0.69
0			0.93	0.70
-5				0.72

Note: SNR = signal-to-noise ratio. All correlation coefficients statistically significant at $p < .05$

B. Speech Intelligibility Results

In some cases, self-adjusted fits resulted in much less insertion gain than the NAL fits, and the influence of this reduction in gain on speech understanding in noise was unknown. Speech intelligibility was assessed to compare subjects’ speech understanding when using their self-adjusted fits with their performance using their NAL fits. Speech understanding was evaluated in a quiet environment and in the steady noise which had the same long-term spectrum as the PB restaurant noise. Sentences were presented in four different SNRs (−10, −5, 0, and +5 dB). Out of the 30 subjects that completed the gain adjustment task, a subset of 17 subjects was able to return to the lab to complete the speech intelligibility task. The first three subjects who completed the speech intelligibility task (S7, S8, and S19) did not complete the −10 dB SNR condition as this condition was added to the protocol after they had finished.

Speech intelligibility was computed as the percentage of the total number of sentence key words that were correctly identified by the listener. Figure 2.8 displays speech intelligibility as a function of SNR for both NAL and self-adjusted fits. Mean performance was similar between NAL and self-adjusted fit conditions at each SNR. Subject S10 showed unusually poor performance for self-adjusted settings in the +5 dB SNR condition. This was the first block of sentences presented, and this subject omitted responses to 6 of the 20 sentences in this block, which suggests that when testing first began, this subject did not initially understand the instructions for responding. This did not occur at any other time.

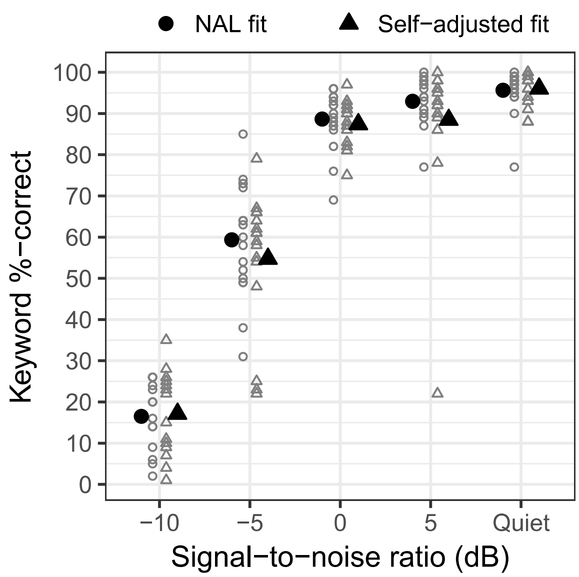

Figure 2.8. IEEE key word recognition achieved using NAL (circles) and self-adjusted (triangles) fits. Large, filled symbols indicate average key word recognition across subjects. Smaller, open symbols are data from individual subjects.

To visualize how adjustments to the gain and compression settings affected speech intelligibility, intelligibility performance using the NAL fit was subtracted from the performance using the self-adjusted fit. A positive intelligibility difference indicates better performance with the self-adjusted fit. Figure 2.9 shows the intelligibility difference plotted as a function of gain adjustment in the low-frequency band (bottom row) and high-frequency band (top row), with plots in each column displaying data from a particular SNR condition. Solid lines in each panel indicate linear fits to the data, excluding two outliers discussed in the next paragraph. Visual examination of the scatterplots and fitted lines suggest that listeners can adjust insertion gain throughout a wide range relative to their NAL fit without greatly altering their speech understanding.

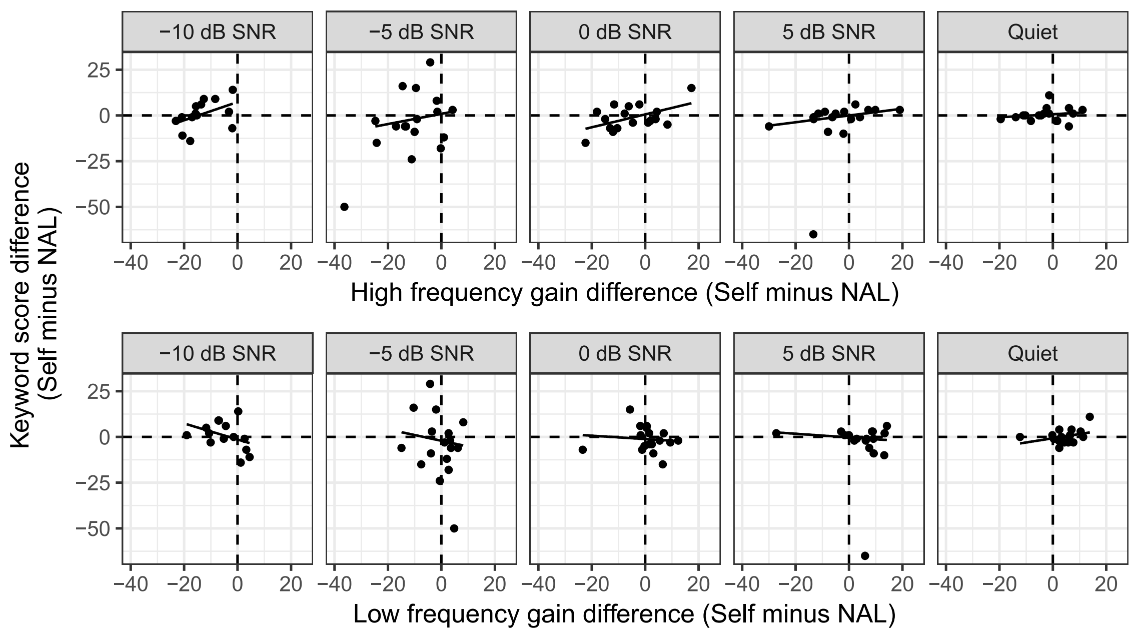

Figure 2.9. Difference in speech recognition performance between self-adjusted and NAL fits plotted with respect to gain deviation from NAL in the low-frequency band (bottom row) and high-frequency band (top row). Each column of panels shows data from a different SNR condition. A positive score difference indicates better performance with the self-adjusted fit than the NAL fit, while a negative score difference (lower on the ordinate) indicates worse performance with the self-adjusted fit.

Data from two subjects stand out as likely outliers. As previously discussed, subject S10's performance in the +5 dB SNR condition was unusually poor, and this might have been due to initial confusion about how to use the interface to respond. Second, subject S20 showed very poor intelligibility in the −5 dB SNR condition with self-adjusted gain settings, which is likely due to the extreme reduction in gain in the high frequencies. The extreme reduction of gain was not replicated in any other trial. For example, the other self-fit completed by S20 in steady noise at −5 dB SNR resulted in 16.6dB more high-frequency insertion gain than the fit that was used to assess speech intelligibility. Further, the self-adjusted fit selected by S20 which was used to assess speech understanding at −5 dB SNR resulted in the least high-frequency insertion gain of all self-adjusted fits.

To systematically evaluate the dependence of intelligibility difference on gain adjustment, for each frequency band and SNR bivariate correlations were computed between the intelligibility difference and gain deviation from NAL, and p values were corrected for multiple comparisons using the Benjamini-Hochberg procedure. The data described earlier as potential outliers were included in this analysis. Although there were trends in the 0 dB and −5 dB SNR conditions for subjects to have poorer intelligibility with the self-adjusted fit as they decreased gain relative to the NAL fit, none of the correlations were statistically significant (all adjusted $p > .05$). Listener adjustments of gain and compression settings did not appear to have a systematic impact on speech understanding. Figure 2.10 shows a histogram of the difference in key word recognition between NAL gain and self-fit gain for each SNR. In over 80% of cases, intelligibility performance with self-adjusted gain was within 10 %-points of performance with NAL gain in the same SNR.

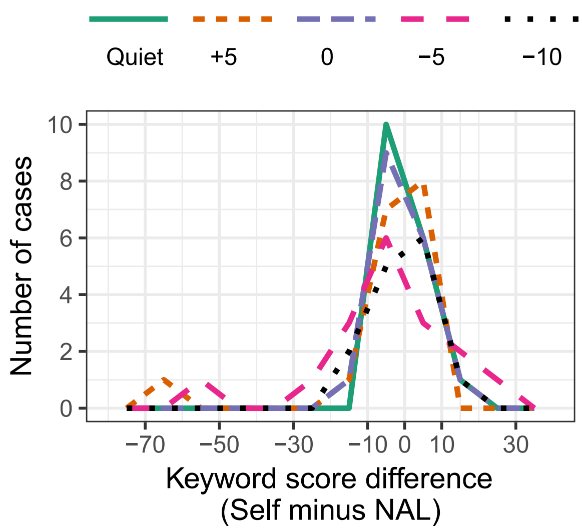

Figure 2.10. Histogram of IEEE key word score differences (self-adjusted fit minus NAL fit) across all subjects that completed the speech recognition assessment. Data from different SNR conditions are displayed as separate lines.

IV. Discussion

The current study investigated users' self-selected gain using the Ear Machine algorithm as a tool to allow listeners to self-adjust hearing-aid gain or compression parameters to select gain for speech understanding in a variety of noise conditions. A different self-adjustment algorithm may have produced different results. Outcomes from self-adjusted gain and audiologist-fit gain (based on NAL-NL2 prescriptive targets) were compared. While listeners showed good test–retest results using the self-adjustment algorithm, indicating consistent performance across days and trials, the variability among participants was striking. Gain changes (differences between NAL-prescribed and self-adjusted gain) were as large as 24 dB in the low frequencies and as high as 37 dB in the high frequencies. Most listeners chose more gain than prescribed in the low frequencies, while others chose less. Some listeners preferred up to 15 dB more gain in the high frequencies, while most preferred less high-frequency gain than prescribed. On average listeners chose more low-frequency gain than prescribed by NAL-NL2. This may not be too surprising, as most listeners were prescribed 0 dB gain in the low frequencies. This may be an outcome of the configuration of the algorithm. It may be noted that listeners were not asked to talk for long periods of time during the fitting process, and so the effect of listeners' own-voice experience may be minimized by the methodology. Note that there are two obvious outliers for low-frequency gain seen in the left panel of Figure 6. One chose significantly less low-frequency gain than prescribed across all SNRs (dotted line) while the other chose significantly more (dashed line), except at +10 dB SNR. The data from the remaining 28 subjects showed a rather tight cluster for self-adjusted low-frequency gain, but it should be noted that even within that group, excluding the two extreme cases, the data show about a 20 dB difference from most to least gain selected.

In the high frequencies, the variability was surprising. While most settings chosen (41 out of 60 fittings) indicated that listeners preferred less high-frequency gain than prescribed, still a number of listeners self-selected 5 to 15 dB more gain than their NAL prescription for quiet to moderately noisy conditions (up to −5 dB SNR). Only one subject selected more gain than prescribed at +10 dB SNR. Selecting gain higher than NAL-NL2 in high-frequency regions seems surprising in light of common reports that listeners typically prefer less high-frequency (>1000 Hz) gain than their fitted settings (Boothroyd and Mackersie, 2017; Kuk and Pape, 1992, 1993; Mackersie et al., 2018;

Moore et al., 2011; Preminger et al., 2000; Zakis et al., 2007). Others have reported significant differences between self-selected and audiologist-fit gain. Hornsby and Mueller (2008) reported gain deviations about half the size of the current results (approximately 8 dB). It is difficult to determine whether the size of the variability noted here is different from many other previous reports, as individual data are not always published. Boothroyd and Mackersie, (2017) report that average self-adjusted gain was within approximately 5 to 10 dB different from NAL-based gain. Overall, we see that individuals vary in their preference for gain-frequency response.

The largest between-subject differences were seen in quiet conditions. Noise level, as predicted, did have some effect on self-adjusted gain parameters. As noise levels increased, preferred gain decreased slightly, even though the NAL-NL2 prescriptions themselves were compressive and resulted in less overall gain with increasing level.

Notably, for the most part, those adjustments did not significantly reduce speech intelligibility in quiet or in noise. At first, this finding was somewhat surprising, due to the large range of gains selected by the subjects. However, because hearing losses were mild to moderate, and in light of the results seen in Figure 8, it can be inferred that in the noise conditions, speech audibility (and resulting intelligibility) was driven mostly by the noise levels. For the quiet conditions, listeners' audibility estimated using the articulation index was greater than 0.5 for all aided conditions (audiologist or self-adjusted) and so intelligibility of speech was near 100%. This phenomenon will be investigated further in a future study.

V. Conclusion

Individuals were largely consistent in their adjustments across SNRs for moderate noise levels ($r \approx 0.9$ for SNRs between −5 and +5 dB), demonstrating that adjustments in moderate noise were not made arbitrarily, and that generally if a listener preferred more gain for one condition, that listener preferred more gain for all conditions. Gain adjustments were more variable in the quiet background and in the most unfavorable noise (−10 dB SNR), suggesting that individuals might weight criteria (e.g., comfort, sound quality) differently when speech is trivially easy or extremely challenging to understand. These findings imply that allowing self-adjustment of gain provides listeners with the opportunity to significantly and uniquely fine-tune their hearing-aid amplification settings.

Future evaluation will obtain qualitative ratings when listening with self-adjusted gain or audiologist-fit NAL-NL2- based gain prescriptions. Results of these new experiments will inform us as to the potential for self-adjustment to result in greater user preference and satisfaction with hearing aids.

Chapter 3: Between-participant variability and listener factors in self-adjustment of amplification

Sections I-VI are reprinted from:
Perry, T. T., Nelson, P. B., & Van Tasell, D. J. (2019). Listener factors explain little variability in self-adjusted hearing aid gain. Trends in Hearing, 23. doi:10.1177/2331216519837124.

I. Introduction

Listener satisfaction with hearing aids has not been well predicted and depends on the complexity of the acoustic environment (Kochkin, 2011). Noisy restaurants are among the most challenging environments for people with hearing aids, but it is not clear from existing data how hearing aid fitting could be changed to improve satisfaction in noisy rooms. Hearing aid amplification parameters are typically set according to prescriptive formulae, such as the National Acoustics Laboratories' non-linear fitting procedure (NAL-NL2) (Keidser et al., 2011). These formulae are intended to increase speech audibility in quiet for people with hearing loss, sometimes in addition to achieving other goals such as normalizing loudness perception (e.g., Moore et al., 2010). This approach may not always be appropriate for selecting amplification gain for all environments. Noise reduction or beamforming algorithms in modern hearing aids can modestly improve subjective aspects of listening in noise for some hearing aid wearers (Bentler et al., 2008; Boymans and Dreschler, 2000), but these algorithms might have side effects that are undesirable, such as reduced speech intelligibility (Brons et al., 2014). Self-adjustment of hearing aid gain (e.g., Keidser et al., 2005) enables listeners to pick amplification settings according to their individual preferences, which could potentially increase listener satisfaction, especially in noise.

The idea of incorporating user feedback or adjustments into the process of fitting hearing aids is not new (Neuman et al., 1987). Although hearing aid wearers sometimes have access to a volume control or may be able to switch between different pre-programmed settings, if hearing aid wearers want substantial changes to the gain and compression characteristics, this typically requires them to ask their audiologist or hearing aid dispenser to fine-tune the hearing aid based on a verbal description of what they want and hope that the resulting changes match their preference (Jenstad et al., 2003). This approach is burdensome to both audiologist and client and might not provide the hearing aid wearer with the gain settings they desire, especially in cases where there are communication difficulties between the audiologist and the client or when the desired change in gain is large.

Greater inclusion of the hearing aid wearer into the fitting process was identified as a potential method for accommodating individual differences in preferred gain by allowing the wearers themselves to strike the balance between settings optimized for speech understanding and settings optimized for comfort or other subjective concerns (Kuk and Pape, 1992). Self-adjustment was advocated by Schweitzer et al (1999), who framed the approach as fitting ‘‘not by prescription, but by perception.’’ The feasibility of using self-adjustment to match hearing aid gain to listeners’ preferred settings has been explored in the past, such as by Elberling and Hansen (1999) who pointed out potential limitations of audiologist-driven fine tuning and implemented an experimental self-adjustment interface on a PC that enabled control over gain in low-, mid-, and high-frequency regions.

In fact, preferred hearing aid gain has been investigated using a number of techniques, including paired comparisons (Amlani and Schafer, 2009; Byrne, 1986; Keidser et al., 1995; Kuk et al., 1994; Moore et al., 2011; Preminger et al., 2000; Punch and Howard, 1978), unpaired ratings (e.g., van Buuren et al., 1995), observation of the gain-frequency response when the volume control is set to the level the wearer typically uses in daily life (e.g., Humes et al., 2002; Smeds et al., 2006), self-adjustment of gain characteristics (e.g., Boothroyd and Mackersie, 2017; Dreschler et al., 2008; Keidser et al., 2005), and trainable hearing aids (Keidser and Alamudi, 2013; Mueller et al., 2008; Zakis et al., 2007). Listener preferences for hearing aid gain have generally been shown to be stable with good within-subject reliability (Dreschler et al., 2008; Elberling and Hansen, 1999; Keidser et al., 2005; Kuk and Pape, 1992; Nelson et al., 2018).

Newly available self-adjustment technology, such as Ear Machine©, allows users a wide degree of control over gain and compression characteristics via a visual interface implemented on a smartphone or similar portable touch screen device. This approach allows users to make quick and potentially substantial adjustments to the function of their hearing aid based on their needs and preferences in real time. Although gain preferences and self-adjustment have been studied in the past, a more recent and direct investigation of a modern self-adjustment method and technology that is available to the general public is warranted.

To that end, we completed a study of self-adjustment of amplification parameters using an example of such portable, real-time technology and reported on the range of gains selected by participants with mild-to-moderate sensorineural hearing loss(Nelson et al., 2018). That report describes the reliability of listeners' selections over repeated self-adjustments, the influence of noise on the amount of gain selected, and the impact of self-adjusted gain on speech recognition performance. Participants were seated in a sound-treated room and used the Ear Machine algorithm to adjust gain and compression characteristics while listening to speech presented either in a quiet background or in simulated restaurant noise environments that were created from recordings made in restaurants. The level of the noise and the restaurant environment were varied to assess how the signal-to-noise ratio (SNR) and listening background influenced the self-adjustments made by the participants.

Participants demonstrated good within-subject reliability, and the insertion gains that resulted from self-adjustment were most strongly affected by the SNR, with listeners selecting less gain as the noise level increased. Gain adjustments made in the various noise environments differed by a small but statistically significant amount; spectral differences among noise environments appeared to have a greater influence on gain adjustment than temporal fluctuations in noise energy. Notably, a wide range of selected gains was observed across participants, spanning about 40 dB. Some listeners opted to give themselves 10 to 15 dB more gain than their audiologist-fit settings, while others chose to reduce the gain by 20 to 25 dB. Despite this large range between listeners, speech recognition did not systematically differ between audiologist gain settings and self-adjusted gain settings.

The amount of variability observed in the previous findings was striking. The present report investigates that variability further and describes the between-listener variability of the listeners' engagement with the self-adjustment technology. The finding of large variability between listeners in preferred gain has important clinical implications for hearing aid adoption and satisfaction. Listeners whose first fits are far from their desired gain characteristics are likely to reject hearing aids. Despite the need to match gain to individual preferences rather than the average preference, in much of the prior literature on self-adjustment of amplification, between-listener variability in selected gain is presented in only a limited fashion. If between-listener variability in gain preference can be predicted from listener characteristics that are readily available to audiologists (factors such as age, hearing thresholds, or prior experience using hearing aids), then fitting methods could be updated to provide more desirable listening levels for hearing aid users. To that end, the NAL-NL2 fitting formula includes options for adjusting fitting targets based on several listener characteristics (Keidser et al., 2012). A primary goal of the present study is to determine whether there are any meaningful relationships between listener characteristics and self-adjusted gain when listeners use a modern, commercially available user interface to adjust the gain and compression characteristics in real time.

The degree to which gain preferences vary between listeners depends, in part, upon the range of possible gain-frequency responses listeners are able to choose from. Possible relationships between listener characteristics and self-adjusted gain may be weakened if the listeners' desired gain settings are outside the limited number of gain-frequency responses offered to them by the adjustment method. In previous studies, the range of gain-frequency responses was sometimes narrow. Mueller et al. (2008) reported results of gain adjustments made by listeners using a range of 16 dB centered around two baseline levels and noted that many listeners reached the limits of the range during self-adjustment. Dreschler et al. (2008) used technology allowing self-adjustments within a 32 dB range and also noted that some people reached the limit during adjustment. Keidser et al. (2008) observed a range of gain preferences (relative to gain prescribed by an earlier version of the National Acoustics Laboratories' non-linear gain fitting procedure, NAL-NL1) that spanned a range of about 20 dB. Participants in a study by Hornsby and Mueller (2008) made gain adjustments in the entire 16 dB range available to them. One impetus for revisiting the issue of variability between listeners is that the range of gain-frequency responses available to participants in the current study was wider than in many previous studies, and the range in which participants selected gain was wider as well (about 40 dB). This wider range of variability may better capture the influence of listener characteristics and deserves closer inspection.

A. Listener Characteristics and Preferred Amplification

Hearing aid experience, or adaptation to amplification, has previously been investigated as an explanation for variation in preferred gain among hearing aid users. Although some reports indicate no statistically significant difference in preferred gain between experienced and new hearing aid users (Cox and Alexander, 1992; Horwitz and Turner, 1997; Humes et al., 2002), other evidence supports the hypothesis that new hearing aid users prefer less gain than experienced users (Boymans and Dreschler, 2012; Keidser, O'Brien, et al., 2008; Marriage et al., 2004). NAL-NL2 includes an adjustment on the basis of hearing aid experience that is also dependent on hearing thresholds

(Keidser et al., 2012). Based on the clear clinical and theoretical questions raised by adaptation to amplification, hearing aid experience was included as a listener characteristic of interest in the current study to examine this relationship using a methodology that gives listeners more direct control over the gain-frequency response of the hearing aid across a wider range.

A common principle of hearing aid fitting is that the gain provided by the aid ought to increase the audibility of speech in the frequency region(s) of the hearing loss. Modern fitting methods for hearing aids incorporate information about the user's pure tone thresholds, but it is not clear whether the difference in gain between self-adjusted fits and prescribed fits is related to hearing thresholds. Mueller et al. (2008) reported no correlation between pure tone average and deviation of self-adjusted gain from prescribed gain, but this finding could have been influenced by ceiling/floor effects based on the limited range of gain in which participants made adjustments (as noted earlier). Keidser et al. (2005) presented evidence that the shape of the gain-frequency response selected by participants with hearing loss was related, in part, to the configuration of their hearing loss, which suggests that listeners were guided by their hearing thresholds as they adjusted gain. Keidser, O'Brien, et al. (2008) found that after user selection of a preferred gain-frequency response and volume control setting, gain deviation from prescribed fit was moderately related to hearing thresholds only for the participants who were new users of hearing aids, which indicates that among new hearing aid users, people who have more severe losses prefer gain that is similar to what people with less severe losses prefer.

To understand the behavior of listeners when self-adjusting gain, two relationships involving hearing thresholds were investigated in the current study. First, the relationship between hearing thresholds and insertion gain was examined to understand if listeners select gain settings that would improve speech audibility in the frequency region(s) of their hearing loss. If it is the case that hearing thresholds have only a weak relationship with the insertion gain from self-adjusted fits, this would suggest that listeners are primarily using criteria other than audibility to guide their selection of gain, contrary to a primary principle of modern prescriptive formulae. Second, the relationship between hearing thresholds and the deviation of the self-adjusted gain from NAL-NL2 fitted gain was investigated to determine the potential utility of modifying a prescriptive formula based on listener thresholds to better match desired gain.

Toward the goal of supplying sufficient gain to overcome the listener’s hearing loss, part of the amplification from a hearing aid compensates for the loss of the resonant energy of the external ear (real-ear unaided gain or REUG) that occurs when the hearing aid is inserted into the ear canal (Upfold and Byrne, 1988). There is substantial variability in the REUG between individuals, and this variability could affect the perceived sound quality of the hearing aid, particularly if there is a meaningful mismatch between the characteristics of the listener’s REUG (such as the peak frequency of the resonant energy) and the gain applied by the hearing aid to compensate for the loss of the REUG (Valente et al., 1991). It is possible that some of the variability in self-adjusted gain is due to listeners attempting to bring the gain provided by the device into better agreement with the particular acoustic characteristics of their ears. Based on this hypothesis, real-ear characteristics were included as predictors of interest.

Keidser, O'Brien, et al. (2008) reported that female listeners tended to prefer less gain than male listeners. The NAL-NL2 formula includes an option for modifying prescribed gain based on gender. When gender is provided to the algorithm, gain is modified by +1 dB for males and −1 dB for females, creating a 2 dB difference in overall gain. The magnitude of this difference reflects the finding that female participants preferred about 2 dB less gain than male participants on average and gives consideration to the trend in the literature for female participants to select lower most comfortable levels (MCL) than male listeners (Keidser et al., 2012). Listener gender was included as a characteristic of interest to evaluate whether further gain modifications based on gender might be appropriate to better match prescribed gain to desired gain.

Although MCL describes the level at which a listener finds speech the most comfortable, acceptable noise level (ANL) describes the maximum level of background noise a listener will tolerate when listening to speech at MCL (Nabelek et al., 2004, 2006). ANL is, in essence, an SNR computed by subtracting the highest level of background noise a listener can accept (in dB) from the MCL. Tolerance of noise as quantified by ANL could potentially explain variability in hearing aid gain preference in noisy conditions. Even though hearing aids amplify speech and noise equally within a processing channel, the gain affects the absolute level of the noise at the output of the hearing aid. Given previous findings that ANL increases (i.e., listeners tolerate less noise) as the overall presentation level increases (Franklin et al., 2006; Tampas and Harkrider, 2006), it is possible that by lowering the gain during self-adjustment, listeners are making the noise level more tolerable even if they are not changing the SNR within individual processing channels. Based on this hypothesis, ANL was included as a variable of interest to understand if ANL could be used to better match prescribed gain to desired gain.

However, it is not clear that self-adjustment will relate to ANL, in part because noise tolerance appears to depend on both the overall sound level as well as how that tolerance is measured. Recker and Edwards (2013) assessed tolerance for noise using the typical ANL procedure as well as the minimum acceptable speech level procedure in which listeners adjust the speech level while the noise level is fixed, in contrast to the ANL procedure in which speech is at a fixed level and the noise level is adjusted. They found that the overall presentation level had opposite effects on noise tolerance depending on whether it was the noise or the speech that was adjusted. As presentation level increased, ANL values also increased (representing less tolerance for noise at higher overall levels), but minimum acceptable speech level values decreased (representing more tolerance for noise at higher overall levels). It is an open question whether between-listener variability in ANL could be used to better match prescribed gain to self-adjusted gain. The hypothesis that lower tolerance for noise, as measured by ANL, will lead listeners to reduce gain relative to NAL-NL2 when listening in noisy environments will be examined.

B. Listener Engagement

Differences in how listeners interact with self-adjustment technology that affords the listener liberal control over the amplification characteristics could also be a potential predictor of variability in self-adjusted gain. The relationship between a listener's engagement with the technology and the resulting gain is not well understood. For example, it is not known whether people who take more time to self-adjust gain or who explore a greater number of alternative gain-frequency responses are more likely to select gain that deviates further from the baseline (which is their NAL-NL2 fit in the present study). Furthermore, listener characteristics might be related to measures of listener engagement, and characterizing these relationships would help to clarify differences in how self-adjustment technology is likely to be used.

Few studies in the literature report details about user interaction with self-adjustment technology, particularly with regard to listener characteristics. In an examination of different controller types for self-adjustment of amplification, Dreschler et al. (2008) noted that participant age, hearing aid experience, or slope of hearing loss were all not statistically significant predictors of the number of key presses needed to reach preferred gain during self-adjustment. In addition, neither age nor hearing aid experience appeared to have systematic effects on test– retest reliability of self-adjustments, suggesting that self-adjustment technology can be used by many listeners to select preferred gain (Boothroyd and Mackersie, 2017; Dreschler et al., 2008). It is not clear that people will interact with any particular implementation of self-adjustment in a similar way to any other particular implementation, and how long it takes users to select their preferred gains could be impacted by the details of the device or algorithm as well as listener factors. Listener engagement will be described and variability in listener engagement will be investigated to understand whether listener characteristics predict how self-adjustment tools might be used by a variety of people, and not just the average user.

C. Research Aims

In this report, listener age, gender, hearing thresholds, hearing aid experience, real-ear characteristics, ANL, and time taken to complete the self-adjustment session will be evaluated as potential predictors of self-adjusted gain relative to NAL-NL2. In addition, known listener characteristics will be evaluated as potential explanations for differences between listeners in the amount of time taken to complete self-adjustment.

The primary research aims are as follows:

1. Report aspects of listener engagement with the self-adjustment technology, including how listener engagement changes with increasing noise level.
2. Investigate the possibility that known listener characteristics can predict between-subject variability in engagement with the self-adjustment technology.
3. Describe the relationship between hearing thresholds and the insertion gain from self-adjusted fits.
4. Evaluate to what degree known listener characteristics and listener engagement predict how much NAL-NL2 fits differ from self-adjusted fits made in a variety of noise environments and SNRs.

II. Methods

A. Subjects

Thirty adults with symmetric sensorineural hearing loss, generally with a sloping configuration and ranging in degree from mild to moderate, participated in the study. Average hearing thresholds of the subjects are shown in Figure 3.1. Subject ages ranged from 59 to 79 years (mean = 70 years). Thirteen subjects were female. Eighteen subjects had prior hearing aid experience, and of that group, 14 subjects had at least 2 years of experience using hearing aids. For all but 2 subjects, REUG and real-ear-to-coupler difference (RECD) were measured during the same audiological evaluation for inclusion in the study. The use of human subjects was approved by the institutional review board of the University of Minnesota. All subjects provided written informed consent.

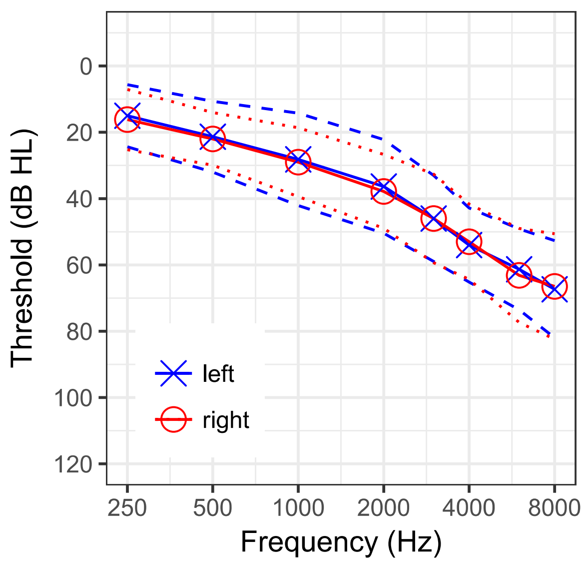

Figure 3.1. Mean participant audiograms for left and right ears of participants in Studies 1 and 2. The dashed blue lines and dotted red lines indicate 1 standard deviation from mean thresholds for left and right ears, respectively.

B. Equipment

Amplification and self-adjustment of amplification parameters was achieved using an application developed by Ear Machine LLC (www.earmachine.com), running on the Apple iOS platform on an iPod touch (fourth generation). Sound was received by the microphone on the iPod, processed by the Ear Machine algorithm according to user adjustments to two software controllers, and delivered to the listeners' ears using Etymotic ER38- 14F foam eartips. The device was designed to simulate a nine-channel multiband wide-dynamic range compressor/ limiter with fast attack and slow release times and output limiting. The proprietary signal processing includes a 12-band equalizer and is similar to a commercial hearing aid.

Listeners adjusted the gain using two virtual wheels: one wheel labeled Loudness which changed gain and compression in all 9 compression channels, and one wheel labeled Fine Tuning which changed the overall frequency response in the 12 equalization bands. Movements to the Loudness wheel simultaneously adjusted the gain values, compression ratios, and output limiter thresholds in each of the nine compression bands. The mapping from controller to parameters was designed to approximate the fit-to-prescriptive-target gains for typical hearing losses from mild (lowest wheel position) to severe (highest wheel position). Therefore, as the wheel was moved upward, the gain in the high-frequency bands increased faster than the gain in the low-frequency bands. Movements of the Fine Tuning wheel controlled the degree of spectral tilt by applying an additional adjustment to the gain values in each of the 12 bands, around a pivot point located near 1 kHz. Increases to high-frequency gains therefore also resulted in decreases to low-frequency gains (and vice versa). The positions of the two wheels interacted to produce the final gain-frequency response. The device was capable of producing a wide range of gain-frequency responses, with up to 40 dB of insertion gain in the low frequencies and 50 to 60 dB of insertion gain in the high frequencies, although in practice the achievable gain is be limited by feedback, based on the individual fit of the earphone.

Figure 2 shows calculated insertion gains for a 65 dB sound pressure level (SPL) speech-shaped input at low, mid, and high positions of the Loudness and Fine Tuning wheels. When the Fine tuning wheel is in a neutral position (when no frequency-specific gain changes are being made in addition to the parameters set by the loudness wheel), the gain effects of the Loudness wheel are clear: At the lowest position, the gain is relatively flat as a function of frequency, but at the highest position, the high-frequency gain has increased more relative to the low-frequency gain, reflecting the increase in high-frequency versus low-frequency hearing loss observed on average as hearing loss becomes more severe.

The Ear Machine controllers constitute a self-fitting method that goes beyond a volume control or even a bass, mid-range, and treble fine tuning. The Loudness wheel adjusts all compression parameters simultaneously in all compression bands to achieve prescriptive fits based on commonly observed audiogram shapes, while the Fine Tuning wheel allows additional gain adjustments beyond the initial prescriptive fit.

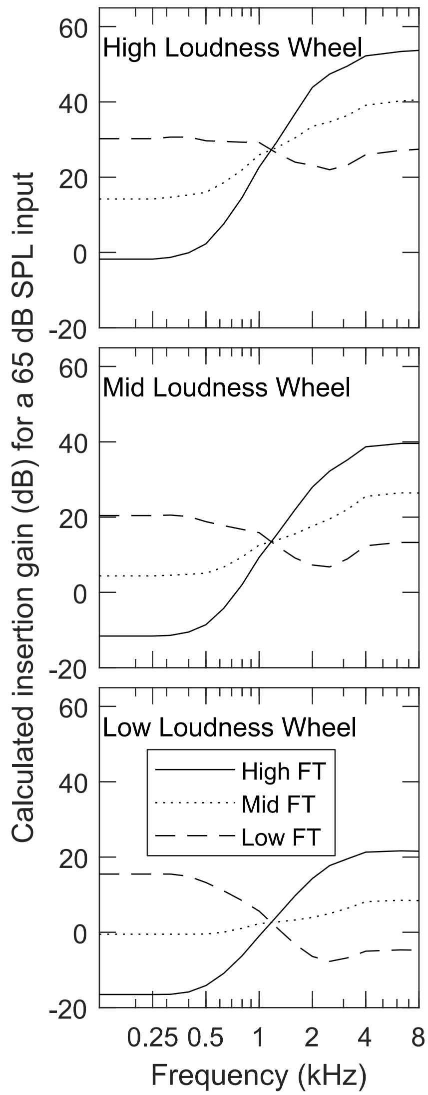

Figure 3.2. Insertion gains from the research device, calculated for a 65 dB SPL speech-shaped input at several positions of the Loudness and FT wheels. FT = Fine Tuning; SPL = sound pressure level.

C. Self-Adjustment Procedure

This study analyzed previously reported gain adjustment data (Nelson et al., 2018), and additional details can be found in that report. Prior to self-adjustment, an audiologist fits the research device to the listener's NAL-NL2 real-ear aided response (REAR) targets (within ±5 dB) using an Audioscan Verifit version 3.16, which does not include the NAL-NL2 empirical adjustments for gender and assumes an experienced hearing aid user. This served as the baseline gain-frequency response that the device was reset to before each self-adjustment trial began. Afterward, the participant was seated in the center of a double-walled sound chamber with a 48-speaker array (Anthony Gallo Acoustics—A'Diva ti speakers) driven by 24 Crown XLS 1500 power amplifiers and 3 Lynx Aurora 16 D/A converters and controlled using MATLAB (MathWorks). During each self-adjustment trial, recordings of a female voice speaking 30-second passages from the Connected Speech Test (CST; Cox et al., 1987) were presented on a loop through a speaker in front of the listener at 65 dBC. Subjects used the Ear Machine software running on an iPod touch (fourth generation) to adjust hearing aid gain and compression. Subjects held the iPod in front of them, and the microphone on the iPod received the sound from the speaker array, after which it was processed by the Ear Machine software according to the adjustments made by the subject using the software wheels on the iPod's touch screen. The processed sound was delivered to the listeners' ears via Etymotic ER38-14F foam eartips.

Participants were instructed to turn the wheels on the iPod until the female talker's voice (i.e., the CST passage) was as clear as possible. They were asked to adjust the wheels one at a time but were told they could adjust each wheel as much as they wanted and in any order. They were also encouraged to go back and forth and adjust the two wheels until they were satisfied that they had found the best settings. To end a trial and confirm the current settings as their self-adjusted fit, the subjects tapped an icon on the iPod screen. Self-adjustments were made while listening to the speech in quiet and in four noisy environments (three simulated restaurants and one steady noise with the same average long-term spectrum as one of the restaurants). The long-term spectra of the noise environments were generally similar to that of conversational speech. The level of the noises was varied to evaluate gain adjustment at 4 SNRs: −10, −5, 0, and +5 dB. Self-adjustment in each condition (i.e., each combination of noise type and noise level) was repeated once. The order of conditions was randomized for each subject.

D. Unaided ANL Procedure

Unaided ANL values were obtained in a separate session following instructions published online by Frye Electronics, Inc. based on the description by Nabelek et al. (2006). The subject was seated in a sound-treated chamber in front of a loudspeaker controlled by an audiometer. The subject was instructed to verbally respond louder or quieter to indicate the direction that the sound level should be changed by the experimenter. A running speech passage was presented using the audiometer. Following the verbal feedback of the subject, the experimenter adjusted the level of running speech in 5 dB steps to reach the levels representing first, too loud, then too soft, and finally the MCL. With the speech passage set at MCL, a noise with the same long-term spectrum as speech was then introduced. The level of the noise was adjusted based on subject feedback until the subject reported that the target voice was incomprehensible. The level was adjusted again until the subject reported that the target voice was clear and easy to hear. Finally, the level was increased up to the point that the subject indicated that it was the most noise that they could put up with while listening for a long period of time. This noise level (in dB) was recorded as the background noise level, and ANL was calculated by subtracting the background noise level from the MCL. Only 21 of the 30 subjects were able to return to the lab for this additional session.

E. Data Description and Analyses

After completion of each gain adjustment trial, the software delivered information about the trial to a data server. This included information about the amplification characteristics as well as listener engagement: trial duration in seconds, number of movements of the Loudness wheel, and number of movements of the Fine Tuning wheel. Trial duration started at the point in time that the software wheels appeared onscreen and ended when the subject tapped an icon on the touch screen to indicate that they have completed the adjustment. The onset of sound presentation was not linked to the software's demarcation of the start of a trial, so the initial period of the trial duration as recorded by the software could include some time in which the subject was waiting for the sound presentation to begin. A wheel movement indicates a single touch and release of a software wheel on the touch screen of the iPod. During a single wheel movement, the wheel can be turned up or down (or both) by varying amounts so long as the finger remained on the wheel. What the software records as a single wheel movement could, in reality, represent a user exploring many different gain-frequency responses.

The software also saved the gain and compression parameters for the self-adjusted fit. From these parameters, insertion gain was automatically estimated for a 65 dB SPL speech-shaped noise input, assuming average adult REUG and RECD values and using coupler-calibrated values to convert from voltage to sound pressure in dB.

To simplify analysis, the self-adjusted estimated insertion gain for each trial was averaged into a low-frequency band (125, 250, 500, and 1000 Hz) and a high-frequency band (2000, 3000, 4000, 6000, and 8000 Hz). Calculated insertion gain (using the same 65 dB SPL speech-shaped stationary noise as input and assuming the same average adult REUG and RECD values) for each subject's NAL-NL2 fit was also averaged into the same low- and high-frequency bands. A low-frequency pure tone average (LFPTA) and a high-frequency pure tone average (HFPTA) were calculated for each subject using the same division of frequencies, averaged across left and right ears to compare the self-adjusted gain to the listener's thresholds in the same frequency region. As a general summary of hearing thresholds, a four-frequency pure tone average (4FPTA) was calculated from thresholds at 500, 1000, 2000, and 4000Hz. NAL-NL2 includes an adjustment for hearing aid experience that depends upon the 4FPTA.

To summarize how self-adjusted fits differed from NAL-NL2 fits, deviation of the self-adjusted gain from NAL-NL2 was calculated by subtracting each subject's NAL-NL2-based insertion gain from the self-adjusted insertion gain (in the two frequency bands). A positive deviation from NAL-NL2 indicates more gain than the NAL-NL2 fit, while a negative deviation indicates less gain than the NAL-NL2 fit.

One trial was excluded from analysis. For a single trial in −5 dB SNR noise for subject S12, the digital record indicated that the subject took over 10 min to finish the trial and did not move either the loudness wheel or the wheel, which suggests that the subject was off-task for this trial. This trial was excluded from all statistical analyses. Of the remaining 1,019 included trials across all subjects, every trial was shorter than 4 min, and every trial but one was shorter than 3 min.

Keidser et al. (2012) presented evidence that suggests that the preference for reduced gain seen in new hearing aid users might change over time, such that at 2 years of hearing aid use, hearing aid user's gain preferences had increased to match the NAL-NL1 prescriptive targets. Accordingly, subjects were sorted into groups according to whether they had at least 2 years of hearing aid use. Using this criterion, 14 subjects were experienced users, and 16 subjects were inexperienced users. Of the inexperienced users, 12 had no experience with hearing aids.

ANL values were obtained for 21 of the 30 subjects. The average age of the 21 subjects that completed the ANL procedure was 69.9 years. Twelve were female, and 11 had any prior experience using hearing aids; of those 11 subjects with any hearing aid experience, 8 people had at least 2 years of prior hearing aid experience. Visualizations of the ANL data are restricted to these 21 subjects, and statistical models that include ANL as a variable were restricted to this subset of subjects. Similarly, because real-ear measures were obtained for 28 of the 30 subjects, statistical analysis of the effect of variability in real-ear acoustics excluded the 2 subjects missing real-ear measures. In short, unless the analysis involved ANL or real-ear measures, data from all 30 subjects were included.

Statistics were computed using the R statistical language. The Benjamini–Hochberg method was used to correct p values to control the false discovery rate. The linear mixed models were created using the lme4 package and the restricted maximum likelihood method (Bates et al., 2015; R Core Team, 2016), and then analysis of variance (ANOVA) tables were calculated using the Kenward–Rogers method for estimating degrees of freedom, via the lmerTest package (Kuznetsova et al., 2017). Several mixed models were created this way: One model was fit to the trial duration, and nested models were fit to the gain deviation from NAL-NL2 data in the low frequencies and, separately, to the gain deviation from NAL-NL2 data in the high frequencies. Models included within-subjects fixed effects (SNR, noise type, and repetition, and a random intercept for subject as well as a random slope for SNR per subject, included to account for any differences in the effect of SNR between subjects) as well as between-subjects fixed effects.

For the model fit to trial duration, the between-subjects fixed effects were age, gender, 4FPTA, and hearing aid experience group. The models fit to the gain deviation data included these same effects as well as trial duration and the interaction between 4FPTA and hearing aid experience group as effects. To evaluate ANL and real-ear characteristics as predictors of variability, subjects with missing data were excluded, and then, the respective effect(s)—either ANL, or REUG and RECD—were added to the model. After fitting, residuals were inspected to verify that there were no violations of test assumptions, including homoscedasticity and normality.

To quantify the amount of variance in deviation from NAL-NL2 gain that is accounted for by listener characteristics, the marginal R^2_{GLMM} and conditional R^2_{GLMM} were calculated using the MuMIn package in R (Bartoń, 2018; Johnson, 2014; Nakagawa and Schielzeth, 2013). The marginal R^2_{GLMM} describes the percentage of variance accounted for by the fixed effects in the model, while the conditional R^2_{GLMM} describes the total percentage of variance accounted for by the model (i.e., by both fixed and random effects). Because the primary interest is the between-subject variability, two reduced models (one each for the high- and low-frequency gain deviation data) that included the within-subjects fixed effects but excluded the between-subjects fixed effects (trial duration, age, gender, 4FPTA, hearing aid experience group, and the interaction between 4FPTA and hearing aid experience group) were fit to the data. The difference in marginal R^2_{GLMM} between full and reduced models indicates the variance accounted for by the between-subjects fixed effects in the full models.

III. Results

A. Listener Characteristics

Six listener characteristics were evaluated as potential predictors of variation in gain adjustment: age, gender, duration of hearing aid use, hearing thresholds (i.e., LFPTA and HFPTA), and ANL. Bivariate correlations were computed between each predictor variable (except gender) and each other predictor variable. To determine the relationship between gender and the other predictors, independent samples t tests were computed between male and female groups for each of the other predictors. Consistent with typical age-related sensorineural hearing loss, age was significantly correlated with HFPTA, $r(28) = .53$, $p = .02$. For this subject sample, years of hearing aid use were significantly correlated with both LFPTA, $r(28) = .64$, $p < .01$, and HFPTA, $r(28) = .51$, $p = .02$. All other correlations were not statistically significant, and no statistically significant differences were observed between male and female subjects on any listener characteristics (all $p > .05$).

B. Listener Engagement

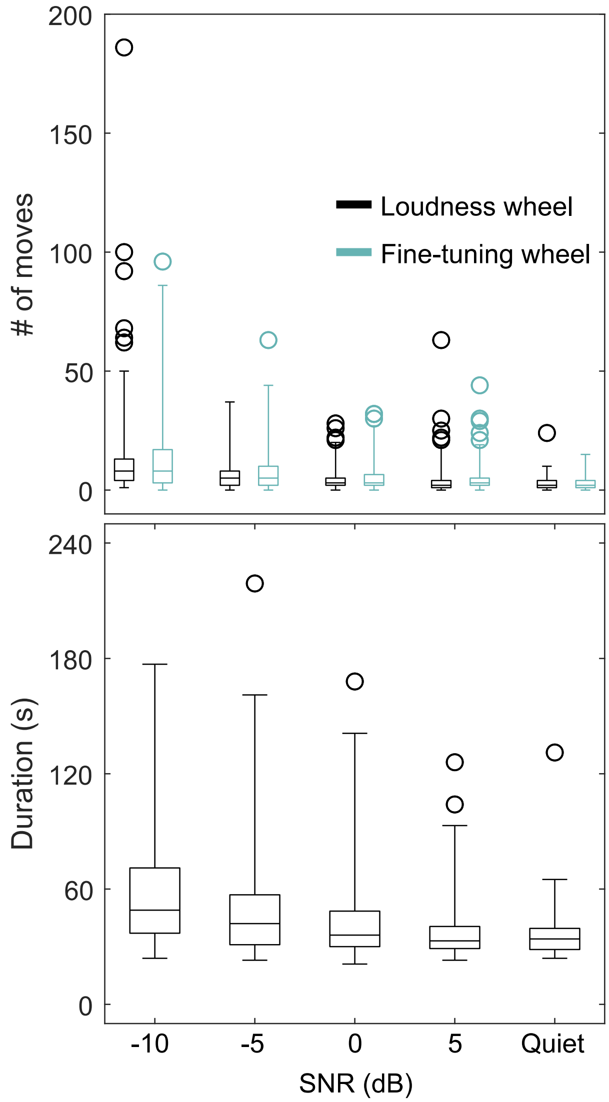

Figure 3.3. Boxplots showing the duration of self-adjustment trials and number of wheel movements across all included trials for all subjects. Whiskers extend up to 5 times the range between the 25th and 75th percentiles of the data.

Listener engagement with the self-adjustment technology was quantified with three metrics: duration of self-adjustment trial (in seconds), number of movements of the Loudness wheel, and number of movement of the Fine Tuning wheel. Figure 3.3 summarizes the distribution of each metric at different noise levels. With increasing noise level, subjects tended to make more wheel movements and spent more time making adjustments.

The three listener engagement variables were strongly correlated with each other (ranging from $r = .68$ to $r = .86$), which suggests that three metrics were consistent in capturing listener engagement during the self-adjustment process. Due to the collinearity between the listener engagement variables, for further analyses, trial duration was taken as a representative measure of listener engagement with the self-adjustment software.

According to the type II sum of squares ANOVA table calculated from the linear mixed model fit to the trial duration data, the only statistically significant fixed effect was SNR, $F(4, 981.0) = 58.19$, $p < .001$; all other $p > .05$. Post hoc tests of contrasts between proximal SNR conditions (i.e., between quiet and +5 dB SNR, between +5 and 0 dB SNRs, and so on) indicated that trial duration progressively increased as the SNR became poorer ($p < .001$ for all SNR contrasts), consistent with the overall pattern seen in Figure 3.3. As the listening situation became more difficult, subjects spent more time before making their final selection, suggesting that listener interaction with the technology followed an understandable pattern. However, listener engagement appears not to depend on the listener’s age, gender, hearing thresholds, or prior experience with hearing aids—at least within the ranges represented in the current sample of 30 subjects.

C. Gain Adjustment and Listener Characteristics

Prescriptive gain fitting methods typically prescribe gain based on the user's hearing thresholds. This enables the hearing aid to provide amplification only where it is appropriate for the goals of prescriptive formula (such as increasing speech audibility or normalizing loudness). Therefore, it was of interest whether the insertion gain selected by subjects using self-adjustment would also relate to their hearing thresholds. Figure 3.4 shows the insertion gain from the self-adjusted fits (averaged across noise environments and trial repetitions) plotted with respect to the LFPTA (for low-frequency insertion gain) and HFPTA (for high-frequency insertion gain) of the subjects. The self-adjusted insertion gain showed statistically significant correlations with the pure tone thresholds in the matching frequency region. For insertion gain and pure tone thresholds in the high frequencies, correlation coefficients ranged from $r = .54$ in the quiet environment to $r = .67$ in the −10 dB SNR condition. For reference, the correlation coefficients between NAL-NL2 gain and pure tone average in this subject sample were $r = .75$ and $r = .60$ for the high and low frequencies, respectively.

The robust correlations between self-adjusted insertion gain and pure tone thresholds indicates that the people who would be prescribed more gain from a hearing aid due to higher thresholds were generally using self-adjustment to achieve more gain than people who had lower thresholds. This indicates that subjects adjusted gain in a meaningful manner that takes into account their hearing thresholds. However, self-adjusted fits showed deviations from NAL-NL2 gain. Explaining the large between-subject variability in deviation from prescribed gain (rather than just the insertion gain of the self-adjusted fit) is a primary goal of the present study.

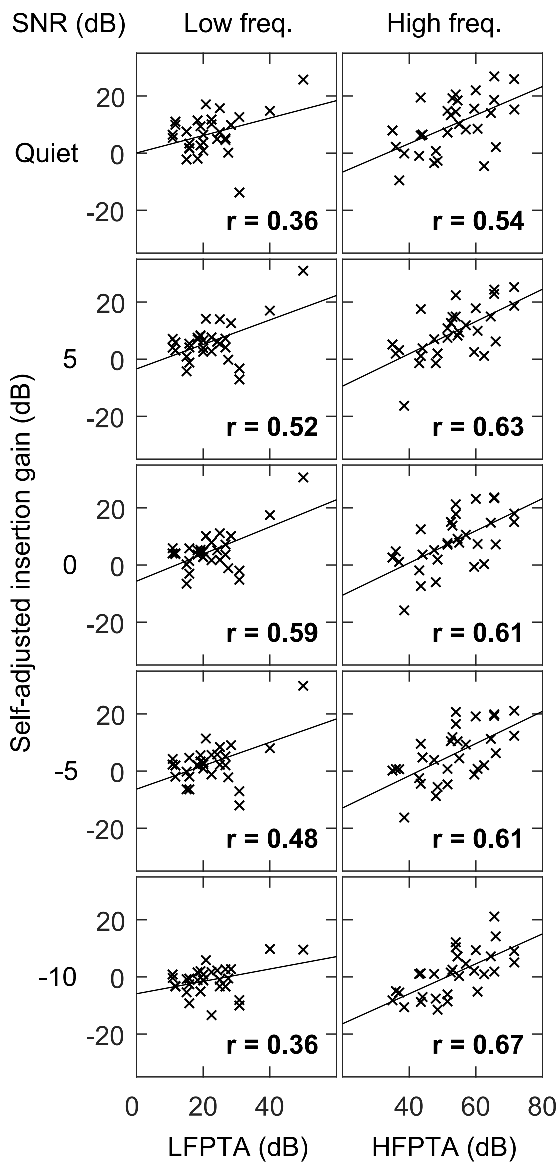

Figure 3.4. Self-adjusted insertion gain plotted with respect to subjects' hearing thresholds. Insertion gain is averaged across noise types and repetitions. Average low-frequency insertion gain and LFPTA are on the left, while average high-frequency insertion gain and HFPTA are on the right. Rows of plots are labeled along the left side by the SNR in which the adjustments were made. Correlation coefficients for the linear fits are shown at the bottom of each plot (all correlations significant at p<.05). LFPTA = low-frequency pure tone average; HFPTA = high-frequency pure tone average.

To inspect the data for possible relationships between listener characteristics and the degree to which self-adjusted gain changed from the prescribed baseline, deviation from NAL-NL2 gain (averaged across noise types and trial repetitions) was plotted with respect to the listener characteristics of LFPTA, HFPTA, age, years of hearing aid use, trial duration (averaged across noise types and trial repetitions), years of hearing aid use, and gender. Figures 3.5 and 3.6 display the resulting scatterplots of listener characteristics and average deviation from NAL-NL2 in the high and low frequencies, respectively. Visually, there appears to be little evidence of relationships between deviation from NAL-NL2 and these listener characteristics.

For the mixed model fit to the high-frequency data, none of the effects of listener characteristics (including the interaction between hearing aid experience and 4FPTA) were statistically significant (all $p > .05$). However, the main effect of trial duration was statistically significant for the low-frequency model, $F(1, 768.66) = 3.88$, $p = .049$. Based on the model coefficient for trial duration, for every additional minute spent using the self-adjustment device, the resulting self-adjusted fit was expected to have 1.2 dB more low-frequency gain than the subject’s NAL-NL2 fit, after controlling for the other effects included in the model. The 95% confidence interval, calculated using a percentile bootstrap method, indicates that the true effect of additional time spent adjusting gain could be as little as 0.03 dB to as much as 2.6 dB per minute. Given the large uncertainty about the effect of trial duration, as evidenced by the confidence interval that spans several orders of magnitude, this finding should be interpreted carefully. Of course, simply sitting with the experimental device in hand will not in itself result in changes to gain—as a reminder, trial duration is used here as a proxy for listener engagement with the device.

According to the statistical models, deviation from NAL-NL2 was not reliably predicted from hearing thresholds and hearing aid experience. However, these two variables are confounded in the subject sample, and these statistical inferences should be interpreted with caution. This is underscored by the fact that when 4FPTA is dropped from the model fit to the high-frequency gain deviation from NAL-NL2, the effect of hearing aid experience is statistically significant, $F(1, 39.27) = 5.54$, $p = .02$, and when

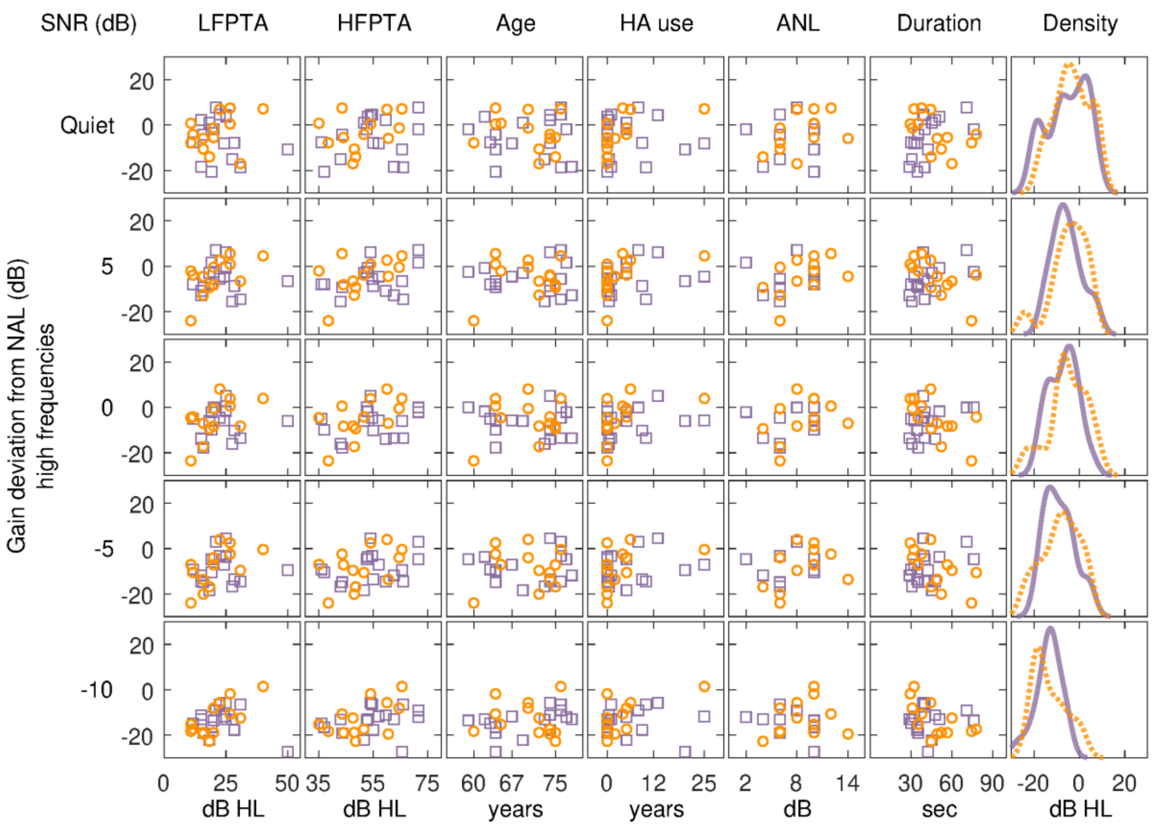

Figure 3.5. Deviation from NAL-NL2 high-frequency gain. Each row contains data from a different SNR condition, averaged across noise types and repetition. Circles and squares represent female and male subjects, respectively. The rightmost column shows smoothed kernel density estimates for the deviation from NAL-NL2 for male (solid line) and female subjects (dashed line). ANL = acceptable noise level; HA use = hearing aid use; HFPTA = high-frequency pure tone average; LFPTA = low-frequency pure tone average.

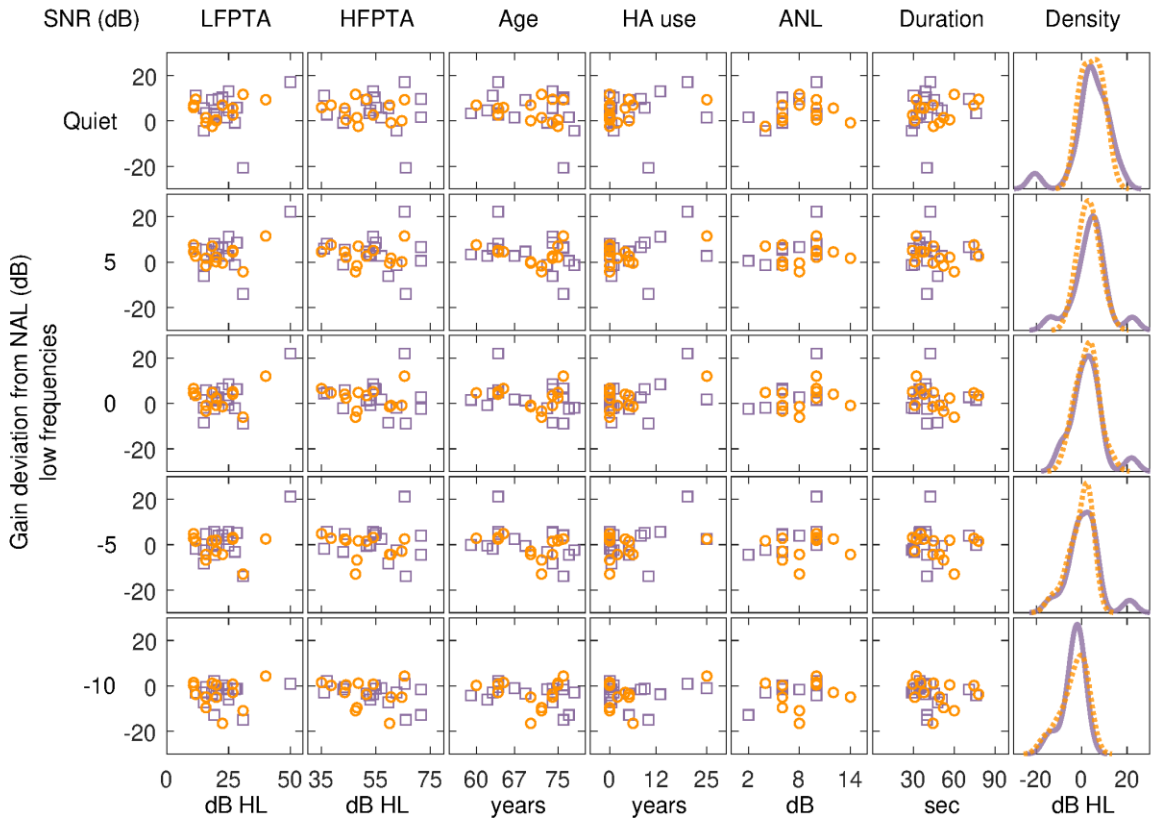

Figure 3.6. Same as Figure 3.5, but for deviation from NAL-NL2 gain in the low frequencies. ANL = acceptable noise level; HA use = hearing aid use; HFPTA = high-frequency pure tone average; LFPTA = low-frequency pure tone average.

hearing aid experience is dropped, the effect of 4FPTA is statistically significant, $F(1, 36.18) = 4.52$, $p = .04$. Briefly setting aside the consideration of statistical controls, inexperienced subjects tended to select less high-frequency gain than experienced subjects. Across all SNRs, including quiet, the average difference in high-frequency gain selected by experienced and inexperienced subjects was about 5 dB. In the +5dB SNR condition, average high-frequency deviation from NAL-NL2 was −2.7 dB for experienced users and −9.4 dB for inexperienced users. In the 0 dB SNR condition, these values were −1.3 and −8.4 dB, respectively. Thus, when noise levels were mild or moderate, both experienced and inexperienced users reduced high-frequency gain relative to NAL-NL2, but the inexperienced users reduced the high-frequency gain by an additional 7 dB, on average. However, due to the confound of hearing threshold and hearing aid experience in this subject sample, it is not possible to determine whether these differences could most accurately be attributed to hearing threshold, hearing aid experience, to neither characteristic, or to some combination of the two.

For the high-frequency data, the reduced model's marginal and conditional R^2_{GLMM} were .11 and .55, respectively. The full model's marginal and conditional R^2_{GLMM} were .21 and .56. Thus, the between-subjects fixed effects, when added to the model, accounted for 10% of the variance in deviation from NAL-NL2 for high-frequency gain. For the low-frequency data, the reduced model's marginal and conditional R^2_{GLMM} were .11 and .58, while the full model's marginal and conditional R^2_{GLMM} were .16 and .59, which indicates that the between-subjects fixed effects accounted for 5% of the variance in the low-frequency data. The fact that the conditional R^2_{GLMM} changed very little by the addition of the between-subjects predictors is likely due to the inclusion of subject-related random effects in the reduced model.

D. ANL and Gain Adjustment

Of the 30 subjects that completed self-adjustments, 21 were able to return for ANL measurement. Figures 3.5 and 3.6 show the deviation from NAL-NL2 gain (averaged across noise type and repetition) for these 21 subjects, plotted with respect to their ANL values. As described earlier, two full models, one per frequency band, were fit to the deviation from NAL-NL2 data for these 21 subjects. These models were the same as the full models described previously, with the addition of a fixed effect of ANL. ANOVA tables (type III sum of squares) were calculated in the same manner as before. The main effect of ANL was not statistically significant in either model—high frequency: $F(1, 20.02) < 0.01$, $p = .98$; low frequency: $F(1, 27.93) = 2.96$, $p = .10$. Calculation of marginal R^2_{GLMM} for the models that included ANL and two reduced models excluding ANL indicates that the inclusion of ANL accounted for less than 1% of the variance in either frequency band. After controlling for the other effects in the model, ANL had almost no relationship with the degree to which the self-adjusted gain deviated from the subjects' NAL-NL2 based fits.

E. Real-Ear Variability

Real-ear measures (REUG and RECD) were obtained for 28 of the 30 subjects. The Ear Machine software assumes an average adult REUG and RECD to estimate insertion gain. Because deviation from NAL-NL2 is a difference measure between two insertion gain values, the REUG and RECD values used in calculating those insertion gains are subtracted out. However, it is still possible that individual variability in real-ear characteristics could have influenced how participants adjusted gain. To evaluate this possibility, the two linear mixed models (one per frequency region) were fit, in the same manner as above, to the deviation from NAL-NL2 data from the 28 subjects for which REUG and RECD were obtained. These models included all the fixed and random effects previously considered except for ANL, as well as two additional between-subjects fixed effects each subject's REUG and RECD, averaged separately within the same high- and low-frequency regions as the gain data. Results of the mixed ANOVAs indicated that, for both models, the main effects of REUG, high frequency: $F(1, 21.21) = 0.61$, $p = .44$; low frequency: $F(1, 20.94) = 0.11$, $p = .74$, and RECD, high frequency: $F(1, 21.21) = 0.29$, $p = .60$; low frequency: $F(1, 21.01) = 0.35$, $p = .55$, were not statistically significant, and variability in real-ear acoustics accounted for less than 1% of the influence on the deviation of self-adjusted gain from NAL-NL2 fits, after controlling for the other included effects.

IV. Discussion

This study analyzed gain self-adjustment data to determine if the large between-subject variability in gain adjustment (about a 40 dB range) could be predicted by known listener characteristics or by listener engagement with the self-adjustment technology. Estimated self-adjusted insertion gain showed strong correlations with listener pure tone thresholds, and self-adjusted insertion gain generally decreased as noise levels increased. In contrast, listener characteristics, including pure tone thresholds, explained little of the between-subject variance in the deviation of self-adjusted gain from NAL-NL2 based gain. Listener characteristics were estimated to account for 10% of between-subject variance in deviation from NAL-NL2 in the high frequencies (>1000 Hz) and 5% of the variance in low frequencies. Using the self-adjusted gain data and the known listener characteristics examined in this study to modify NAL-NL2 or other similar prescriptive formulae is unlikely to result in the preferred gain in noise for many hearing aid users.

Of the characteristics examined (age, gender, prior hearing aid experience, 4FPTA, duration of self-adjustment, ANL, and real-ear acoustics), none showed strong relationships with deviations from NAL-NL2 gain in the high frequencies, and only trial duration had a statistically significant relationship with deviations from NAL-NL2 gain in the low frequencies. In the current sample, each additional minute with the self-adjustment technology was associated with an increase in low-frequency gain of about 1.2 dB. It is not clear from the data why longer self-adjustment trials would tend to result in more low-frequency gain.

Listeners tended to take more time to adjust gain and made more wheel movements as noise levels increased, demonstrating that listeners spent more time exploring the gain settings when listening conditions were more challenging. These results indicate that subjects used the self-adjustment technology in an understandable manner, taking more time as listening conditions became more difficult. However, the differences between self-adjusted gain and NAL-NL2 based gain were not strongly related to known listener characteristics.

The NAL-NL2 baseline as implemented in the Audioscan Verifit system, which was used to fit NALNL2 to subjects in this study, treats all listeners as experienced hearing aid users when calculating REAR targets. When subjects were sorted into two groups based on their years of hearing aid use, as per the findings of Keidser et al. (2012), inexperienced users (< 2 years) generally reduced the high-frequency gain further from the NAL-NL2 baseline than the experienced users (≥ 2 years) did. However, after controlling for hearing threshold, this difference was not statistically significant, which may be because in the current subject sample, subjects with greater losses also tended to be experienced hearing aid users. Although a finding that inexperienced hearing aid users prefer less gain than those with 2 or more years of hearing aid use would be consistent with other reports (Boymans and Dreschler, 2012; Keidser, O'Brien, et al., 2008; Marriage et al., 2004), it was not possible to untangle the effects of hearing aid experience and hearing thresholds in the current subject sample. Furthermore, within-group variability was substantial. Some experienced users preferred 5 dB or more high-frequency gain than NAL-NL2 baseline, while other experienced users preferred substantially less gain than prescribed by the formula (e.g., 15 dB less). Providing a description of this within-group variability, in addition to reporting group averages, is crucial for a full understanding of the influence of hearing aid experience on amplification preferences.

Keidser et al. (2012) reported a gender difference of 2.4 dB in preferred gain between male and female subjects. The NAL-NL2 formula prescribes a 2 dB difference in overall gain when gender is specified, with males receiving a 1 dB boost and females a 1 dB cut (although this gain modification is not implemented on the Audioscan Verifit system that was used to fit NAL-NL2 REAR targets in this study). In the present data, males tended to reduce high-frequency gain more than females (1.3 dB average difference across conditions). According to the coefficient for gender in the linear mixed model fit to the high-frequency data from all 30 subjects, after controlling for the other effects included in the model, males were estimated to have selected 1.6 dB less high-frequency gain than females. The coefficients for gender were not statistically significant in either of the models, suggesting that the true effect of gender in the population on deviation from NAL-NL2 could be 0 dB. The male–female difference in this sample is in the opposite direction of the NAL-NL2 gender correction, which was not applied to the NAL-NL2 fits in this study.

ANL ostensibly reflects the least favorable SNR a person is willing to tolerate when listening to speech and was assessed as a potential predictor for gain adjustment variability in noise to determine if preference for gain in noise was related to noise tolerance as measured by ANL. However, ANL was not predictive of variability in gain adjustment. In this sample, only three subjects produced ANL values within the range of SNRs tested (i.e., ANL values of 5 or lower), meaning that even the most favorable SNR condition tested was an unacceptable level of noise to most of the subjects. Although ANL has been reported to improve (i.e., decrease) when the overall presentation level is reduced (Recker and Edwards, 2013), there was no evidence in the present data that individuals with higher ANL values were more likely than those with lower ANL values to reduce the gain to improve the acceptability of the noise.

Listeners were successful overall in using the software interface to quickly adjust gain and compression parameters across frequencies. Out of 1,020 gain adjustments, only 2 took longer than 3 min for the listener to complete (with an average duration of less than 1 min), and the median number of movements of each of the software wheels was 4. Any single wheel movement could represent the exploration of multiple gain-frequency responses so long as the participant did not remove their finger from the touch screen. These results are similar to those reported by Boothroyd and Mackersie (2017), who found an average time for their self-adjustment method of less than 2 min. None of the known listener characteristics robustly predicted how long subjects took to self-adjust gain. These data suggest that people will be able to quickly adjust hearing aid gain and compression parameters using an appropriately designed interface, regardless of hearing thresholds, age, or other personal characteristics (assuming demographic characteristics similar to the current subject sample). Incorporating self-adjustment into the process of fitting a hearing aid is unlikely to be a substantial time investment if the interface is simple and intuitive and allows users to arrive quickly at appropriate settings.

Individuals are relatively stable in their gain adjustments across noise environments, but variability in gain adjustment across listeners is large (Nelson et al., 2018). That is, if a person tends to use self-adjustment to reduce gain in one noise environment, they are likely to also reduce gain in other noise environments. However, it is not currently possible to predict a priori whether any specific individual will tend to prefer more or less gain than what they are prescribed by NAL-NL2. Listener characteristics and interaction with the self-adjustment technology were ineffective at predicting the magnitude of gain adjustments that listeners made. Considering the wide range over which self-adjusted fits deviated from the NAL-NL2 baselines as well as the speed at which self-adjustment is typically completed, self-adjustment may be the most straightforward and effective way to match hearing aid gain with listener's preferred levels, including when listening in noise.

V. Future Directions

Self-adjustment is a useful tool for investigating preferences for amplification characteristics. In the self-adjustment paradigm, listeners select gain, and their selection is assumed to reflect their preferred gain settings. This assumption will be examined in a future study. Beyond establishing whether listeners prefer their self-adjusted settings to audiologist-fit settings, an important avenue of future research will be to evaluate whether customization of amplification parameters via self-adjustment results in measurable improvements in factors relating to quality of life, such as increased social participation or improved emotional well-being.

Additional work is needed to assess the role of perceived speech intelligibility during self-adjustment and subjective evaluation of hearing aid gain. In the present study, speech was presented at 65 dBC, which represents an average conversational level. For people with mild-to-moderate hearing loss, most of the speech spectrum at this level is above their hearing thresholds, and if noise is present, the audibility of speech is likely to be primarily limited by the level of noise (Plomp, 1986). In such situations, changes to the gain-frequency response are unlikely to have large consequences for speech recognition. While this bolsters the argument that self-adjustment can be used to achieve similar speech recognition outcomes as clinically prescribed gain for conversational-level speech in noise (for people with mild-to-moderate hearing loss), it also means that in the present research, most of the gain-frequency responses available to the subjects through the self-adjustment technology provided similar speech audibility, so speech intelligibility might not have played a large role in the subjects' decisions. Further study of how self-adjustment is used when circumstances permit gain to have a larger influence on speech audibility—such as when speech is at lower levels in quiet—will clarify to what extent people with mild-to-moderate hearing loss are willing to trade speech intelligibility for improved sound quality, comfort, or other subjective factors.

Self-adjustment may one day play an important role in over-the-counter or self-fitting hearing aids, which present a new problem of how to set gain and compression parameters without the direct help of hearing health professionals. Understanding the relationship between gain that is fit according to widely used clinical formulae and gain that is fit using self-adjustment is an important step in understanding the consequences of this new approach. In particular, it will be important to evaluate the many self-adjustment methods (in addition to the Ear Machine method that was used in this study) in terms of their ease of use and effectiveness, because not all self-adjustment methods will produce equivalent results. The present data provide evidence that listeners self-adjust hearing aid gain using the Ear Machine interface according to idiosyncratic preferences that are not easily predicted from known listener characteristics, and it is unlikely that prescriptive formulae can be modified according to demographic information to provide the same degree of personal customization.

VI. Conclusions

The variability in self-selected hearing aid gain that was noted by Nelson et al. (2018) cannot be predicted by known listener factors in this group of 30 subjects. Six listener factors were evaluated as predictors of variation in gain adjustment: age, gender, duration of hearing aid use, hearing thresholds, ANL, and real-ear characteristics. Specifically, we found the following:

1. Listener engagement with the interface was successful in that participants required little time to complete self-adjustment. Subjects took an average of less than 1 min to complete adjustments, and all but 2 adjustments were completed in less than 3 min.
2. Duration of self-adjustment was not related to other known listener characteristics, and while duration was statistically associated with greater reductions in gain relative to NAL-NL2 in the low frequencies, calculation of the confidence intervals for this effect suggest that this association might not be clinically meaningful.
3. Self-adjusted insertion gain was significantly and strongly correlated with high-frequency hearing thresholds.
4. Listener age was significantly correlated with high-frequency hearing thresholds but explained little between-subject variability in the deviation of self-adjusted gain from NAL-NL2 fitted gain.
5. No statistically significant differences between the gain selected by male and female participants were observed. However, a small trend was noted in the opposite direction of the NAL-NL2 gender corrections in that men tended to reduce the gain further than women, relative to their NAL-NL2 fits.
6. Neither ANL nor between-subject variability in real-ear characteristics (REUG and RECD) predicted gain changes relative to NAL-NL2 in the conditions tested here.
7. Due to the significant correlations between hearing thresholds and years of hearing aid use in the current subject sample, it was not possible to determine with statistical

rigor the effects of hearing thresholds and hearing aid experience on deviation of self-adjusted gain from NAL-NL2 fitted gain, but on average, the listeners who had less than 2 years of hearing aid experience (and who also had better pure tone thresholds) reduced the gain more than listeners who had 2 or more years of hearing aid experience (and who had poorer pure tone thresholds).

These findings suggest that, when given the opportunity, individual listeners will choose hearing aid gain characteristics that relate to their hearing thresholds (when starting from a threshold-based prescription) but which may deviate from formula-prescribed gain in ways that are poorly predicted by known factors such as age, gender, hearing loss, or hearing aid experience. This supports the idea that giving people with hearing loss control over hearing aid gain allows them to choose custom parameters that otherwise might not be available when using conventional methods of hearing aid fitting.

Chapter 4: Self-adjustments for low-level speech in quiet and their relations to preference and speech understanding

I. Introduction

A 2016 study organized by The National Academies of Sciences, Engineering, and Medicine reports that approximately 30 million Americans have hearing loss, but as much as 86 percent of adults who could benefit from hearing aids do not use them (National Academies, 2016). Problems identified as barriers to hearing aid use included concerns about comfort, effectiveness, satisfaction, and financial cost. Reducing these barriers has been a long term goal of hearing health professionals, the auditory research community, and hearing aid manufacturers.

One proposed method for increasing hearing aid satisfaction is to give the wearer greater control over the amplification gain and compression characteristics of the device. This approach is sometimes called client-directed hearing aid fitting or self-adjustment of amplification parameters. Preferred listening levels and listener selection of amplification and have been investigated previously (e.g., Boymans and Dreschler, 2012; Keidser et al., 2005; Kuk et al., 1994; Smeds et al., 2006), and the concept of incorporating listener feedback or self-adjustments into the fitting process to improve satisfaction with hearing aids is now several decades old (Elberling and Hansen, 1999; Kuk and Pape, 1992; Punch and Parker, 1981; Schweitzer et al., 1999).

A. *Overview of self-adjustment of hearing aid amplification*

The incorporation of user-guided adjustments can be accomplished in a variety of ways, from volume controls to fully self-fitting hearing aids. Typically, hearing aids are programmed by an audiologist or hearing aid specialist using a prescriptive formula, such as NAL-NL2 or DSL v5 (Keidser et al., 2011; Scollie et al., 2005), or the device manufacturer's own proprietary formula (Byrne, 1996; Keidser et al., 2003; Killion, 2004). The gain is set in one or more frequency bands with the aim of restoring the audibility of lower-intensity signals, and compression parameters are selected to limit the overall loudness of the output of the hearing aid, particularly for higher-intensity inputs.

During programming, it's possible to use feedback from the wearer to fine-tune the gain and compression characteristics using informal verbal feedback or more formalized methods such as paired comparisons (Amlani and Schafer, 2009; Jenstad et al., 2003; Kuk and Pape, 1992; Punch et al., 2001). Formalized methods for user-directed customization of gain and compression parameters during hearing aid fitting have not been widely adopted, and the success of informal methods depend on both the ability of the wearer to adequately express their desired adjustments as well as the ability (and inclination) of the hearing professional to translate the verbal description into modifications of the gain and compression parameters.

After the hearing aid programming is completed, user adjustments to amplification, and in particular the gain-frequency response, are constrained by the programming and specifications of the device. Many hearing aids have a volume control which allows the wearer to make changes to the overall hearing aid output across all frequencies. The maximum range in which the output can be increased or decreased depends on the specifications and programming of the device. Banerjee (2011) used a commercially-available digital hearing aid to investigate real world usage of volume controls, and the volume control on that device provided a 25 dB range (+10 to -14 dB). Compared to some implementations of self-adjustment of amplification, volume controls offer less opportunity for user-directed customization of amplification due to the lack of frequency-specificity in their functioning.

Some hearing aids may allow the wearer to switch between two or more different programs or “memories”, and depending on the device and how it is programmed, the gain-frequency responses of the different memories could differ in meaningful ways. A hearing aid so programmed would provide the wearer some limited choice in the gain-frequency response of the device (Keidser et al., 1995). However, it’s also possible that the hearing aid is programmed so that the memories share similar gain and compression profiles and differ primarily in which additional features are active (e.g., noise reduction and directionality).

Both volume controls and hearing aid memories offer partial control over the operation and output of the hearing aid, but this may not be sufficient to improve listening satisfaction for some wearers. People who desire gain changes beyond what can be achieved with a volume control or who would prefer a different gain-frequency response would need to visit a hearing professional to re-program the hearing aid. When barriers to addressing hearing aid dissatisfaction are high (such as difficulty accessing audiological services due to issues with cost, travel, or availability), some individuals may simply choose not to wear the hearing aid at all, eliminating the opportunity for obtaining the quality of life benefit that the hearing aid could otherwise provide (McCormack and Fortnum, 2013).

Due to the wide availability of mobile computing technology and standardized wireless communication protocols, it is now possible to incorporate self-adjustment of hearing aid gain and compression characteristics into modern hearing aids, providing the user with real-time control over a much larger range of parameter values than traditional volume controls or memories. With such technology, the wearer uses a physical or virtual control interface (such as virtual buttons on a touch screen) to adjust, directly or indirectly, one or more parameters along the signal processing chain of the hearing aid. Hearing aid manufacturers have begun to release smartphone apps that connect wirelessly to hearing aids in order to control their operation. The specific parameters that can be adjusted, the range of values the parameters can take, and the interface used to make the adjustment will all depend on the details of how self-adjustment is implemented in a hearing aid (Dreschler et al., 2008). Potentially, the wearer could have nearly as much freedom in selecting a gain-frequency response and overall output level as audiologists have in the clinic when programming the hearing aid.

Such self-adjustment technology might lead to greater individual customization in order to improve satisfaction, and could be used to program over-the-counter hearing aids, also known as direct-to-consumer hearing aids or fully self-fitting hearing aids. For such devices, self-adjustment could be used to alter the gain and compression characteristics of the hearing aid, either after it has been set to a generic gain-frequency response or set to a prescribed gain-frequency response derived from in-situ pure-tone threshold measurements (Convery et al., 2011; Keidser and Convery, 2016). Self-adjustment technology could be used in the audiology clinic as well to individualize the hearing aid fitting after an initial programming by an audiologist. It’s also possible that self-adjustment could be used to customize the amplification parameters starting not from an individualized fit but from a fit to an average audiogram for people with mild-to-moderate sensorineural hearing loss.

Investigations of self-selected gain have shown that there is substantial variability between adults in the gain selected by people with sensorineural hearing loss when listening to speech at average conversational levels. If a method of self-adjustment is used which permits the listener to select from a large range of possible gains, the insertion gains selected by different listeners with relatively similar hearing thresholds can span about 30 dB (Keidser et al., 2012; Perry et al., 2019). In contrast, within-subject variability tends to be small as test/retest standard deviations are reported to be around 2 to 4 dB, depending on the adjustment methods and listening conditions (Dreschler et al., 2008; Nelson et al., 2018). Self-adjustment tools can be used reliably and each listener has a consistent preference. However any one person’s desired gain may differ substantially from any other person’s.

An important trend is that self-adjusted gain tends to be lower than gain prescribed by fitting formulae like NAL-NL2 or DSLv5 (Boothroyd and Mackersie, 2017; Keidser, O’Brien, et al., 2008; Moore et al., 2011; Nelson et al., 2018; Preminger et al., 2000). On average, the difference between self-adjusted gain and prescription-based gain may be small (e.g., 5 dB), but for individual listeners the difference can be substantial. Nelson et al. (2018) reported multiple cases of self-adjustment resulting in gain that was 20 dB less than prescription-based fittings. Such a large reduction in gain from prescriptive settings invites concern that self-adjustment will result in the hearing aid providing inadequate amplification to restore the audibility of speech.

B. Research questions

Given that substantial scientific effort has been invested into developing prescriptive formulae for setting hearing aid gain and compression, caution and skepticism about hearing aid self-adjustment is justified. A chief concern is that the hearing aid wearer may not be able to efficiently find settings that increase speech audibility while controlling the overall loudness of sound to produce a satisfactory and comfortable output from the hearing aid. New hearing aid users might choose levels of gain that do not sufficiently amplify speech to provide a communication benefit. Or it's possible that some wearers might naively think that louder is always better and, based on this assumption, select excessive levels of gain that induce fatigue and dissatisfaction which could lead to less time spent wearing the aid. The worst of both worlds is also possible, with listeners choosing too little gain in frequency regions with greater hearing loss and too much gain in frequency regions with normal or nearly-normal hearing thresholds.

It's important to address these and similar clinical concerns, which can be generalized into the following research questions.

1. What are the amplification parameters produced by a client-driven approach to fitting?
2. Do hearing aid wearers prefer listening with amplification parameters fit to the targets of a prescriptive fitting formula (such as NAL-NL2), or do they prefer listening with the amplification parameters that they've played a role in selecting?
3. Are there systematic differences in speech recognition performance between audiologist-fit and self-adjusted amplification parameters?
4. What role does a person's subjective sense of speech clarity play in determining preferences for amplification and the selection of gain during self-adjustment?

If it were the case that client-driven fitting methods produced similar amplification as current clinical methods, the incentive to increase the client's involvement in fitting would be weak. Likewise, if it were the case that people do not tend to prefer their self-adjusted fit over an audiologist fit, then it would seem unlikely that self-adjustment would result in greater satisfaction. And finally, if self-adjustment tended to result in poorer speech recognition performance, this could undermine a primary motivation for using amplification in the first place, namely, to increase the wearer's ability to communicate and participate in daily life.

To varying degrees, previous research has addressed these questions, but further investigation is still warranted. First, it has often been assumed in the literature that the gain which results from self-adjustment reflects the listener's true preferred listening level. This assumption is reasonable but deserves to be evaluated explicitly. It's possible that self-adjustment can result in amplification that is dissatisfying or less desirable than audiologist-fit settings.

Second, the end point of self-adjustment is a single set of amplification parameters, but it is possible that listeners may find a range of amplification parameters acceptable. If self-adjustment is used to pick a gain-frequency response from among a set of responses that the listener finds equally acceptable (or roughly so), then characterizing the extent of that set of acceptable responses will be helpful for predicting the results of self-adjustment. However, few published studies present data that addresses this issue. One approach to investigating whether a listener's set of acceptable gain-frequency responses is broad or narrow is to ask listeners to make paired preference judgments comparing their self-adjusted fit with a fit that has more or less gain than the self-adjusted fit. Paired comparisons can also be used to explicitly determine whether listeners prefer their self-adjusted settings over those provided by an audiologist following a prescriptive fitting formula.

Third, the research questions about self-adjustment of hearing aid amplification listed above have not been investigated concurrently with the same listeners. Doing so would help determine whether speech recognition influences the preference for either self-adjusted gain or audiologist-fit gain, and in turn, help answer the question of whether people with hearing loss are likely to prefer amplification settings that result in reduced communication benefit relative to audiologist-fit amplification. It is also worthwhile to have measurements of speech recognition rather than estimates derived from models such as the Speech Intelligibility Index (ANSI, 1997). Especially for people with sensorineural hearing loss, measured speech recognition can differ widely from predicted speech recognition (Akeroyd, 2008; Humes et al., 1994; Pavlovic, 1984) .

Fourth, several prior studies of self-adjustment of amplification had technological limitations that might constrain their relevance to modern implementations and uses of self-adjustment. For example, among several studies of self-adjustment of hearing aid amplification, the range of gain adjustments available to study participants was often small and some participants reached the limits of this range during self-adjustment. This limitation may artificially constrain the extent to which self-adjusted gain might vary from a prescriptive fit. For example, previous literature has examined gain adjustments within a 17 dB range, i.e., ± 8 dB (Hornsby and Mueller, 2008) and a 33 dB range, i.e., ± 16 dB ((Dreschler et al., 2008). In both cases, some participants reached the maximum or minimum of the permitted range when adjusting the gain. In addition, few studies of self-adjustment have used a modern self-adjustment user interface that is available to the public. It’s worthwhile to understand how currently-available tools for self-adjustment of amplification are likely to be used by people with mild-to-moderate hearing loss by incorporating these tools into the research methods.

Finally, to investigate the connections between amplification, speech recognition, and listener preferences, the listening conditions should be carefully selected so that changes to gain are likely to have a greater influence on speech audibility. Plomp (1986) modeled the speech recognition benefit of amplification for people with hearing loss and demonstrated that when speech audibility is limited primarily by the auditory thresholds of the listener rather than by a masking sound, amplification is predicted to improve speech recognition. However, in situations where speech audibility is limited primarily by noise, the benefit of amplification is reduced. Crucially, the regime in which gain has the most influence on speech audibility is when speech is at lower intensity levels so that at least some of the unaided speech energy is inaudible due to elevated hearing thresholds, and the speech is presented in a quiet background. These are the conditions that should be met if the research goal is to understand if listeners with mild-to-moderate hearing loss are likely to sacrifice speech recognition benefit in order to meet other amplification goals (such as increased comfort) using self-adjustment.

The present study attempts to answer the aforementioned research questions while addressing the limitations and gaps in the existing literature. Older adults with mild-to-moderate hearing loss of likely sensorineural origin completed self-adjustment of amplification using a simple control interface implemented on a portable touchscreen device while listening to speech in a variety of conditions. Then participants completed provided subjective preference feedback about their self-adjusted settings, their NAL-NL2 settings, and additional settings with less gain. Finally, speech recognition was measured with each of these amplification settings to quantify how the communication benefits of hearing aids fit using self-adjustment may differ from hearing aids programmed to match prescriptive targets.

II. Methods

A. Procedure overview

For each participant, completion of the study required 2 separate visits to the University of Minnesota campus, with each visit lasting about 2 hours. During the first visit, informed consent was obtained and a brief audiological evaluation was completed, consisting of otoscopy and pure tone audiometry. If the participant met inclusion criteria (see Participants below), then real ear measures would be made and the research amplification device would be fit to the targets prescribed by NAL-NL2 based on the participant's air conduction thresholds. As the last part of the first visit, the participant would complete self-adjustment using the research device while listening to speech in a sound-treated booth.

Upon return for the second visit, each participant rated the amplification settings (as provided by the research device) separately, and then completed blinded paired comparisons of the settings in 5 categories. Finally, the last study task was a sentence recognition task in which participants listened to sentences presented one by one and attempted to verbally repeat back as much of each sentence as possible. Figure 4.1 provides an outline of the study tasks across the three visits as well as the amplification settings and presentation levels used in each task.

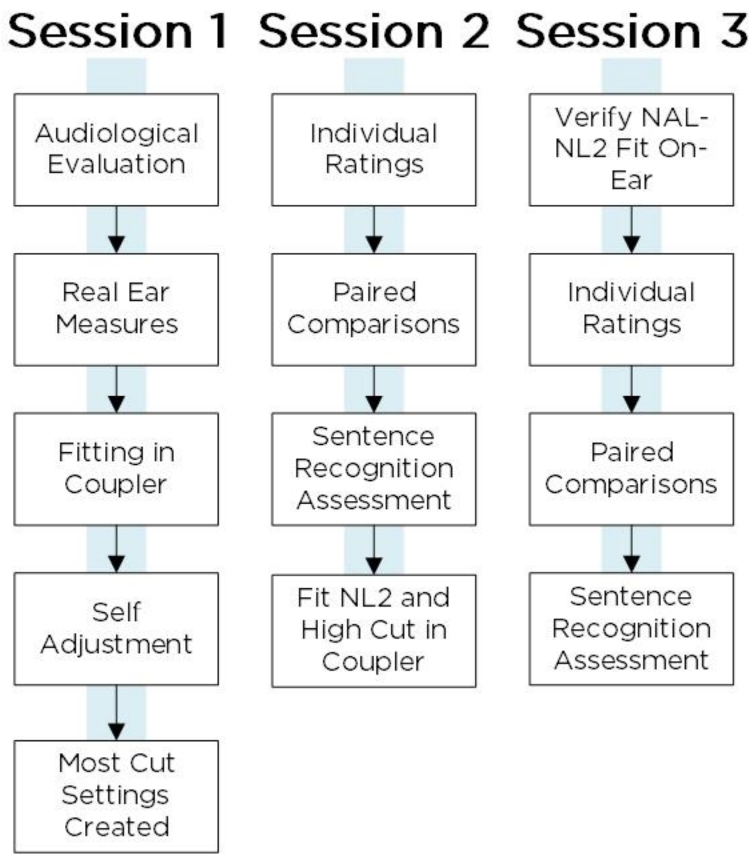

Figure 4.1. Order of study tasks during the three laboratory visits for Study 3.

After 15 participants had completed all study tasks, it was discovered that due to experimenter error the NAL-NL2 fits made during each participant's first visit were often far below prescriptive targets, particularly for low frequencies. Calibration and calculation errors were made when measuring the coupler response that was used to compute the initial real-ear-to-coupler differences (RECD) used when fitting to NAL-NL2 (see Real ear measurements and NAL-NL2 fitting, below).

To remedy this error, participants were invited back in to be properly fit with NAL-NL2 (including on-ear verification) and complete a new set of individual ratings, paired comparisons, and speech recognition testing. During the first visit, one participant was fit with NAL-NL2 solely using on-ear verification rather than with coupler measurements, and thus received an appropriate NAL-NL2 on their first visit. That participant was not contacted to return for further participation. Eleven other participants were able to return to the lab for additional testing and received proper NAL-NL2 fits, resulting in a total of 12 participants who had NAL-NL2 fits within appropriate ranges of the prescriptive targets. Throughout this report, the erroneous first NAL-NL2 settings are referred to as the LowCut settings as they tended to have high frequency gain that was similar to NAL-NL2 targets but much a low frequency response that was far lower than targets.

B. Participants

Fifteen adults with symmetric, mild-to-moderate hearing loss of likely sensorineural origin completed the study tasks. Prior to study participation, otoscopy and pure tone audiometry was completed to ensure the participant met inclusion guidelines, which were as follows: no air-bone gaps larger than 15 dB (and no other observed evidence of a substantial conductive component to the hearing loss), no asymmetries between ears in air conduction thresholds larger than 15 dB at more than 1 frequency, air conduction thresholds 65 dB or better for frequencies up to and including 4 kHz. Figure 4.2 shows the air conduction pure tone thresholds for the 15 participants as well as the average thresholds for left and right ears. Most participants were older adults, with ages ranging from 55 to 79 and an average age of 72 years. 5 participants were male, and 10 were female. 11 participants owned and wore hearing aids daily, while 4 did not. The use of human subjects was approved by the institutional review board of the University of Minnesota. All participants provided written informed consent.

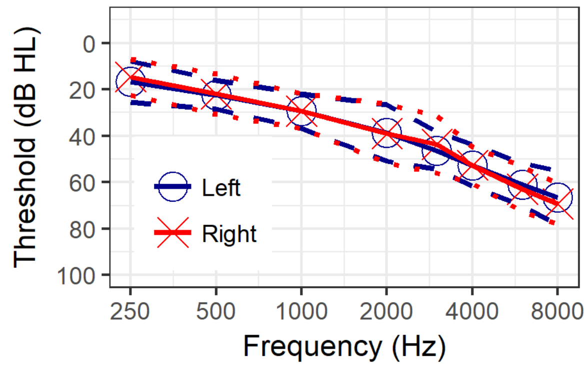

Figure 4.2. Average pure tone thresholds of participants in Study 3. Dashed and dotted lines indicate 1 standard deviation from the mean (solid line) for each ear.

C. Equipment

The study was conducted in a sound treated booth at the University of Minnesota. Sounds were presented through a loudspeaker located 1 meter directly in front of the seated participant. Instead of using conventional hearing aids to provide amplification, an Apple iPod Touch (4th generation) running experimental software developed for iOS by Ear Machine LLC was used to achieve the basic functionality of a wide dynamic range compression hearing aid. A set of modified Bose earphones, with silicone ear tips, were connected to the iPod. These earphones had microphones which received the sound at the pinnae and transmitted the signal to the iPod via Bluetooth. The Ear Machine software processed and amplified the sound according to the position of two adjustable virtual software controllers ("wheels") shown on the touchscreen of the iPod, and sent the output sound back to the earphones to be relayed to the listener's ears. The device was designed to simulate a nine-channel multiband wide-dynamic range compressor/limiter with fast attack and slow release times and output limiting. The proprietary signal processing includes a 12-band equalizer and is similar to a commercial hearing aid.

Participants could modify the output of the device by moving the two virtual wheels up or down. Wheel movements indirectly resulted in changes to gain and compression by changing the input parameters to a proprietary algorithm developed by Ear Machine. One of the wheels was labelled Loudness on the touchscreen, and this wheel changed gain values, compression ratios, and output limiter thresholds simultaneously in all 9 compression channels. The other wheel was labeled Fine Tuning and changed the overall frequency response in the 12 equalization bands.

The mapping from controller to parameters was designed to approximate the fit-to-prescriptive-target gains for typical hearing losses from mild (lowest wheel position) to severe (highest wheel position). Therefore, as the Loudness wheel was moved upward, the gain in the high-frequency bands increased faster than the gain in the low-frequency bands. Movements of the Fine Tuning wheel controlled the degree of spectral tilt by applying an additional adjustment to the gain values in each of the 12 bands, around a pivot point located near 1 kHz. When using the Fine Tuning wheel, increases to high-frequency gains therefore also resulted in decreases to low-frequency gains (and vice versa).

The positions of the two wheels interacted to produce the final gain-frequency response. The device was capable of producing a wide range of gain-frequency responses, with up to 40 dB of insertion gain in the low frequencies and 50 to 60 dB of insertion gain in the high frequencies, although in practice the achievable gain is be limited by feedback, based on the individual fit of the earphone. Figure 4.3 shows calculated insertion gains for a 65-dB sound pressure level (SPL) speech-shaped input at low, mid, and high positions of the Loudness and Fine Tuning wheels. When the Fine tuning wheel is in a neutral position (i.e., when no frequency-specific gain changes are being made in addition to the parameters set by the Loudness wheel), the consequences of changes to the Loudness wheel are clearest: At the lowest position, the gain is relatively flat as a function of frequency, but at the highest position, the high-frequency gain has increased more than the low-frequency gain. This is in line with the average patterns of hearing loss as losses become more severe. High frequencies tend to show more severe losses sooner than low frequencies.

The Ear Machine controllers constitute a self-fitting method that goes beyond a volume control or even a bass, mid-range, and treble fine tuning. The Loudness wheel adjusts all compression parameters simultaneously in all compression bands to achieve prescriptive fits based on commonly observed audiogram shapes, while the Fine Tuning wheel allows additional gain adjustments.

Coupler measurements, real ear measurements, and fitting of NAL-NL2 were accomplished using an Audioscan Verifit1 machine prior to completion of any study tasks. A 2-cc HA1 coupler was used with the Verifit to measure the output of the earphones during fitting of NAL-NL2. After participants had completed all research tasks and left the lab, the Verifit real ear analyzer and coupler were used as to measure the insertion gain of each fit produced during the study.

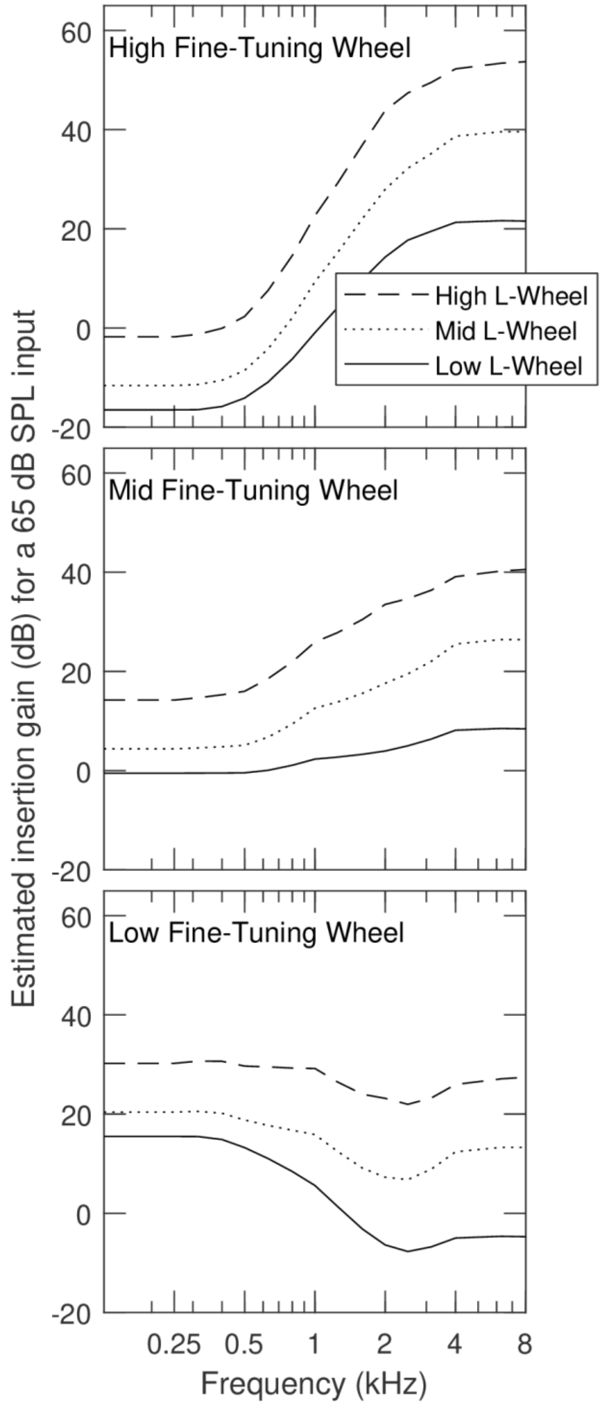

Figure 4.3. Estimated insertion gains for a 65 dB SPL input with a speech-shaped spectrum for combinations of low, moderate, and high wheel positions of the Ear Machine software. Insertion gains were calculated using average adult values for the unaided response. L-Wheel = Loudness Wheel.

Presentation of sound was controlled by custom software running on a Windows PC. Recordings were stored on the computer as .wav files. For the speech recognition task, lists of IEEE sentences were randomized and presented using custom software called Token which uses the AUX scripting language to specify sound processing and presentation parameters (Kwon, 2012). All other tasks were controlled using custom software created in MATLAB (MathWorks). In short, for all listening tasks signals were generated by a PC and sent to a loudspeaker located in front of the participant. The participant was wearing the modified Bose earphones connected to the iPod. The iPod and Ear Machine software processed the sound and transmitted the amplified sound to the participant's ears.

D. Calibration of speech and noise signals

A subset of recordings of the Harvard sentences (also known as the IEEE sentences) spoken by a female talker were used for the speech recognition task (IEEE, 1969). For all other tasks, recordings of the Connected Speech Test (CST) passages were used (Cox et al., 1987). Recordings were stored digitally with a sample rate of 48 kHz. The sentences and CST passages were each calibrated separately, using the same process. First, a calibration noise was created which had the same long-term spectral magnitudes as the original speech files. This was done by concatenating all the digital sound files representing a set of recordings (e.g., all the CST passages or all the Harvard sentences) and applying a Fourier transform to this long waveform. Next, the phases of all spectral components were randomized, and then the inverse Fourier transform was computed and the imaginary portion of the result was discarded. The spectrally-shaped noise was then adjusted to have the same digital long term root mean square (RMS) as the source sound files and truncated to be 5 minutes in duration.

Next, the calibration noise was played through the loudspeaker via the PC. A sound level meter set to slow A-weighting was used to measure the output level of sound at the expected location of a person's head when sitting 1 meter away from the loudspeaker. Based on this measurement, the correspondence between the digital RMS and the sound level in dBA at the location of the listener's head was established and subsequently the presentation level of the sound files could be varied by modifying the digital RMS before playback. This correspondence was verified multiple times over the course of the study's completion to ensure that the equipment stayed in calibration.

In some experimental conditions a background noise was used. This noise was the same noise that has been used previously in our lab during studies of hearing aid self-adjustment and had the same long term spectrum as a recording made from a restaurant near the university during a regular business day (Nelson et al., 2018). The noise used in the experiment was created from the source restaurant recording in a similar manner to the calibration noise described above, i.e., by randomizing the phases of the spectral components of the source restaurant recording. The resulting background noise had a flat amplitude envelope and a spectrum that was approximately similar to the long term average spectrum of speech. Calibration of this background noise was completed in the same manner as described above for the speech signals.

E. Real ear measurements and NAL-NL2 fitting

After the participant was determined to meet the audiological inclusion criteria (see Participants above), probe microphone recordings were completed to enable the research amplification device (i.e., the iPod running the Ear Machine software) to be programmed according to the fitting targets provided by the NAL-NL2 prescription formula as implemented on the Verifit1 real ear analyzer. The real ear measurements occurred in the following order: first, a real ear unaided response (i.e., with the probe microphone placed in the open ear canal) was measured using pink noise at 50 dB SPL. Then the earphones, turned off, were placed in the ear canal and a real ear occluded response was measured, also with pink noise at 50 dB SPL. Next, the earphones were turned on and the software on the iPod was configured to provide a nominal 10 dB linear gain across all frequency bands. With these amplification settings on, a real ear aided response was recorded with the same pink noise input signal as well as with the International Speech Test Signal (ISTS; Holube et al., 2010) at 50 dB SPL.

The real ear recordings made with pink noise were then inspected for signs of an incomplete seal of the silicone earphone tip on the participant's ears. In the case of evidence of low-frequency energy leakage, the silicone tips on the earphones would be replaced with a different size and the real ear recordings would be redone.

The aided response recording using the nominal 10 dB linear gain settings was compared with a reference coupler recording made using the same input and same nominal 10 dB linear gain settings in order to calculate the RECD for each participant. The RECD and the participant's hearing thresholds were entered into the Verifit system for the left and right ears. The system was configured to provide NAL-NL2 targets for a binaural fitting and a non-tonal language. Although the NAL-NL2 standalone software includes the option for adjustments to the prescribed targets based on the wearer's gender (a 1 dB increase for men and a 1 dB decrease for women) and hearing aid experience (a decrease in prescribed gain for new hearing aid users, dependent on their hearing thresholds), the Verifit system does not include an option for these adjustments. Thus the NAL-NL2 targets used in this study did not include the adjustment for gender and assumed that all participants had experience with hearing aids for the purposes of prescribing amplification (Keidser et al., 2012).

For a single participant, NAL-NL2 was fit completely using on-ear measurements and verification of fit-to-target. For all others, NAL-NL2 was fit using coupler measurements corrected with a (RECD). This was done by attaching an earphone from the research device to an HA1 coupler with putty and entering the manually-calculated RECD, described above, into the Verifit system in an attempt to bring the coupler measurements into alignment with the actual pressure levels in each participant's ear canal. Fitting to prescriptive targets was accomplished using a separate user interface on the iPod rather than the two virtual wheels. In this separate fitting mode, the gain and compression ratios could be individually set in 12 frequency bands with center frequencies spanning from 250 Hz to 7100 Hz. First, compression ratios were set to 2 for all bands with center frequencies higher than 1000 Hz and set to 1 for the lower frequency bands. Next, the gain in each band was adjusted until the recorded aided response for the ISTS at 50 dB SPL matched the targets within ± 5 dB. Then the compression ratios were adjusted until the aided response for the ISTS at 70 dB SPL was within ± 5 dB of the prescribed targets as measured in the coupler with an RECD correction. Lastly, a recording with the ISTS at 50 dB SPL was redone to verify that the aided response was still within the tolerance range of the prescribed targets. All amplification settings, including those of the NAL-NL2 fits, were symmetric. The hearing thresholds of the ear (left or right) with less hearing loss in the high frequencies (2 to 4 kHz) were used to calculate fitting targets.

As mentioned above, due to experimenter errors in obtaining the reference coupler response that was used to calculate the RECD, the initial NAL-NL2 fits of 14 of the 15 participants had poor fit-to-targets. These erroneous settings are labeled as the LowCut settings throughout the paper as these settings were often close to prescriptive targets at 2, 3, and 4 kHz, but substantially below targets for low frequencies, typically resulting in a negative insertion gain. When participants returned for a third visit, they were re-fit with NAL-NL2 using properly calculated RECD values and the quality of the fit was always verified using on-ear measurements. Figure 4.4 shows the aided response plotted with respect to prescriptive targets for the on-ear verified NAL-NL2 fits at 6 frequencies.

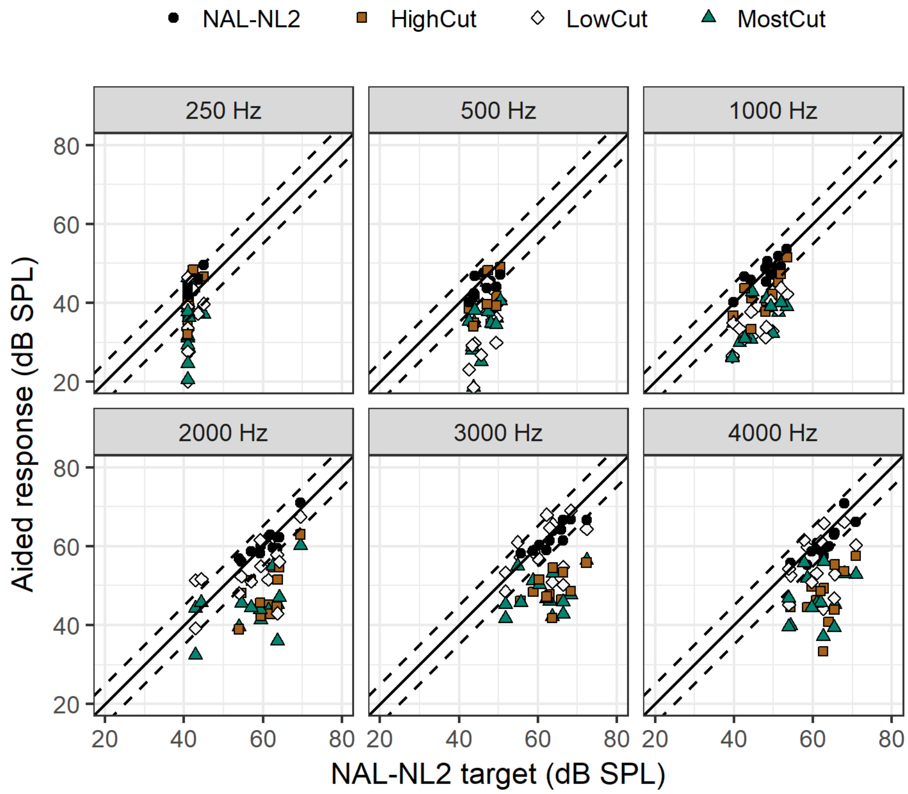

Figure 4.4. Aided response of experimenter-created settings, including NAL-NL2 fits, plotted with respect to NAL-NL2 targets. Dashed diagonal lines indicate a ± 5 dB range around an exact match to target (solid line).

F. Self-adjustment procedure

During self-adjustment, each participant was seated in the sound-treated booth with the loudspeaker 1 meter directly in front of them. The experiment administrator placed the earphones of the research amplification device in the participant's ears and each participant verbally verified that the earphones fit comfortably and securely. The participant held the iPod which was running the Ear Machine software. Participants were told the following instructions: "You will hear a woman talking; her voice will come from the speaker in front of you. Sometimes, there will be background noise, and sometimes the background will be quiet. Using the touchscreen, adjust the wheels until you can understand the woman's voice as clearly as possible. Go back and forth between the wheels until you are satisfied that you have the best setting that you could use for a long period of time. Once you think you have found the best setting possible, tap the star icon on the touchscreen one time."

The voice being played through the loudspeaker was the recording of the female talker reading the CST passages. For each trial, a random passage was selected and played on a loop, giving the participant as much time as they wanted to complete self-adjustment using the touch screen on the research device. The presentation level of the CST passage was determined by the listening condition of each self-adjustment trial. The star icon on the touch screen, when tapped, stored the positions of the Loudness and Fine Tuning wheels in the memory of the iPod for later retrieval.

Prior to data collection, each participant was given an informal familiarization session during which the participant was told that they would have an opportunity to freely explore how moving the software wheels up and down on the touch screen influenced the sound coming out of the earphones. Participants were encouraged to move the wheels as much as they liked during familiarization, and were cautioned that the first time they moved the Loudness wheel up, they should do so slowly to avoid quickly raising the loudness up to an uncomfortable level, but after they had explored using the Loudness wheel, they should feel free to move the two wheels however they liked. During the familiarization session, a random CST passage was played on a loop while the participant explored moving the wheels. When the participant was satisfied that they were familiar with how to use the research device, they told the experiment administrator. The familiarization session then ended and data collection began.

Each participant completed 12 self-adjustments total. There were 6 listening conditions, with each condition being repeated once for the purpose of evaluating reliability. For 3 conditions background noise was presented with the speech, and for 3 the speech was presented alone. For the conditions with background noise, the CST passage was presented at 65 dBA and the level of the background noise was varied to create signal-to-noise ratios (SNRs) of +5, 0, and -5 dB. For the conditions without background noise, the presentation level of the CST passage was 65, 55, or 45 dBA. The order in which these conditions were presented was randomized for each participant with the constraint that no condition would be repeated before the participant had first made a self-adjustment in all 6 conditions.

Before the start of each self-adjustment, it was intended that the wheels be reset to a starting position which most closely matched the NAL-NL2 fit for the participant (a process completed automatically by the Ear Machine software). However, in reality the baseline fit from which each self-adjustment was made was the LowCut settings for 14 of the 15 participants, not the proper NAL-NL2 fit.

G. Amplification settings for further tasks

The self-adjusted settings used for the remaining research tasks (individual ratings of settings, paired comparisons of settings, and speech recognition assessment) were the settings resulting from each participant's second self-adjustment made while speech was presented at 55 dBA in quiet. These settings were used as the self-adjusted settings, also called "Self" settings, regardless of the speech presentation level used in the remaining tasks. That is, even when the presentation level was 45 dBA during ratings and speech recognition, the self-adjusted settings used were still the gain and compression parameters that had been selected by the participant when the speech was presented at 55 dBA during self-adjustment. This was done to both streamline administration of the experiment and to evaluate self-adjusted settings selected under different listening levels than the levels used during preference and speech recognition assessments.

One of the research goals was to investigate the range of gain-frequency responses that listeners might deem to be as acceptable as their self-adjusted or NAL-NL2 settings. In other words, could settings with less gain than the self-adjusted settings and the NAL-NL2 settings be appraised as satisfactory? Therefore, in the remaining study tasks, additional amplification settings were created for each participant after self-adjustment was completed. These additional amplification settings are called the Cut settings (MostCut and HighCut) and were intended to have less gain than either the NAL-NL2 or self-adjusted settings.

Cut settings were created by first determining, via coupler measurements, whether the NAL-NL2 settings or the self-adjusted settings (from the 2nd adjustment trial with 55 dBA speech) had less gain above 1000 Hz for a 50 dBA input. Whichever settings had less high-frequency gain were taken as the baseline settings from which high-frequency gain was reduced to create the new setting. From the baseline, the gain in the high-frequency bands (those with center frequencies higher than 1000 Hz) was reduced by 7 dB (± 1 dB). This reduction was confirmed with coupler recordings. The compression ratios of the Cut settings were not changed from the baseline settings.

During the two lab visits that originally comprised the study, only the MostCut settings were created and used during study tasks. After the NAL-NL2 fitting error was discovered, when participants returned for a third visit, the HighCut settings were created using the same method described above for use during the third visit, and the MostCut settings were not used in the third visit. Therefore, only 12 participants (the 11 that returned and the 1 that was fit with NAL-NL2 on-ear only) made preference judgments involving the HighCut settings or completed sentence recognition testing with the High Cut settings. For some participants, the HighCut settings were similar to the MostCut settings, usually when the MostCut settings were derived from the self-adjusted settings rather than the LowCut settings.

Figure 4.4 shows the aided response of the amplification settings which were not the result of self-adjustment, including the Cut settings. To summarize, depending on whether the participant returned for a third visit to the lab, a total of up to 5 settings were evaluated after self-adjustment was completed: their self-adjusted (Self) settings taken from the second self-adjustment made with 55 dBA speech, NAL-NL2 settings, LowCut settings, MostCut settings, and HighCut settings. The LowCut settings typically produced output somewhat close to NAL-NL2 targets in the high frequencies, but much less than targets in the low frequencies. The HighCut settings had low frequency gain that was similar to Self and NAL-NL2 settings while having reduced high frequency gain, and the MostCut settings typically had the least total gain out of all the settings.

H. Speech presentation levels for further tasks

Based on the motivation of assessing amplification preferences and speech recognition outcomes in a regime where changes to gain result in meaningful changes to speech audibility, two speech presentation levels (45 and 55 dBA) were used for the individual ratings, paired comparisons, and speech recognition assessment. A speech level of 55 dBA represents an average conversational level for speech in quiet backgrounds (Olsen, 1998), while 45 dBA is a realistic level that is also sufficiently low that unaided speech recognition is expected to be challenging for nearly all people with to mild-to-moderate hearing loss. No background noise was used except in the self-adjustment task.

I. Individual ratings of amplification settings

While seated in the sound-treated booth and wearing the research amplification device, participants made ratings of the NAL-NL2, Self, and various Cut settings while listening to speech in quiet. A total of 6 sets of ratings were completed per participant, with 3 ratings completed when speech (a CST passage chosen at random) was at 45 dBA and 3 ratings completed when speech was at 55 dBA. The order of settings and presentation levels was randomized for each participant. The settings were rated on three attributes: satisfaction, subjective speech understanding, and loudness. Ratings were completed using a pen and paper form.

For rating satisfaction, the form asked “How satisfied are you with the sound from the earphones?” and the response options were a 5-item Likert scale: very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied, or very dissatisfied. For rating subjective speech understanding, the question was “How much of a quiet conversation with a friend would you be able to understand if you were listening with these settings?”, and the response options were also a 5-item Likert scale: almost all, most, some, little, or almost none of the conversation. For loudness ratings, the question was “How loud is the sound from the earphones?” and the response options were the labels from the Contour Test of Loudness Perception (Cox et al., 1997): very soft, soft, comfortable but slightly soft, comfortable, comfortable but slightly loud, loud but OK, and uncomfortably loud

The experimenter always walked the participant through the ratings form prior to data collection and verified that the participant understood the task and how to respond. Verbal instructions also included these additional clarifications. First, the participant was told that “quiet conversation” meant a conversation at an everyday volume in a quiet room. Second, the participant was told that an uncomfortably loud sound represented a loudness that they would never choose to listen to on the car radio, no matter what mood they were in. They were also instructed that if they were to experience an uncomfortably loud sound that they should immediately remove the earphones from their ears and get the attention of the experiment administrator. Lastly, participants were told that there were no right or wrong answers and to respond with whatever they personally thought was the best answer to each question.

J. Paired comparisons

During a single visit to the lab, each participant made 4 paired comparisons, 2 with speech presented at 45 dBA and 2 with speech presented at 55 dBA. For those 11 participants who returned for a third visit, an additional 4 paired comparisons were made. The Self setting was always included as one of the settings in the pair. Therefore, the pairs of settings that were included were Self with NAL-NL2, Self with MostCut, Self with HighCut, and Self with LowCut. The order of conditions (i.e., settings pair and speech level) was randomized for each participant. Blinding of participants was achieved through experiment software on the research amplification device. Two buttons were displayed on the iPod's touchscreen, one labeled A and one labeled B. Tapping either button would switch the gain values and compression ratio settings used by the device to match the values that had been randomly assigned to that label. Participants were blind as to what settings were represented by A and B. The assignment of amplification settings to the A and B labels was randomized for each pair of settings.

Participants were asked to rate their relative preference for settings "A" or settings "B" using a pen and paper form. The settings were compared on 5 attributes: overall preference, comfort, sound quality, speech clarity, and loudness preference. On the form were 5 lines, each labeled by one of these attributes, running across the width of the paper. Under each line in smaller print were labels equally spaced from left to right which read as follows: Strongly Prefer A, Slightly Prefer A, No Preference, Slightly Prefer B, and Strongly Prefer B. The participants were instructed to use the pen to make a mark on each line to indicate the strength of their preference with regard to each attribute, with pen marks closer to the left end of the line indicating a preference for A over B, and pen marks closer to the right end of the line indicating a preference for B over A. The experiment administrator demonstrated this response method using an example form. Participants were also instructed that the settings would change for each form and that new settings would be assigned to A and B.

Unless the participant asked for clarification, descriptions of the rating categories were not provided, with the exception of loudness preference. As part of the routine instructions, participants were told the following: "Loudness preference is about what you prefer in terms of the loudness of the sound from the earphones. It might be that you prefer the softer sound, or you might prefer the louder sound. Either way, make your mark based on which one has the loudness you prefer."

Beyond loudness preference, if a participant asked for clarification about the meaning of an attribute, the experiment administrator would provide the following information for each attribute. For overall preference, the participant would be told to consider everything about their listening experience with settings A and settings B and make a judgment between A and B based on whatever factors about their listening experience they found most important. For comfort, the participant would be told to consider how comfortable or uncomfortable they felt when listening with each setting. For sound quality, the participant would be told to focus on how natural the speech sounded using each setting. For speech clarity, the participant would be told to focus on how much of the speech they think that they could understand using each setting. Although it was not strictly tracked, few participants – perhaps 2 or 3 – asked for clarification about any preference category.

After participants had completed the research tasks and left the lab, a ruler was used to measure how far along each line the pen marks were in order to quantify the relative preference for each pair of settings. In addition, data was retrieved from the Ear Machine software which recorded the assignment of settings to the A and B labels. With this information, and the total length of the line on the page, the preference ratings were converted into a numerical scale from 100 to -100, with 100 representing the strongest preference for the Self settings over the comparison settings (which would be either NAL-NL2 or Cut) and -100 representing the strongest preference for the comparison settings over the Self settings. Values near 0 would indicate little or no preference one way or the other.

K. Speech recognition assessment

The final study task was a speech recognition assessment using the NAL-NL2, Self, and various Cut amplification settings. Speech recognition was assessed at 45 dBA and at 55 dBA in a quiet background. Participants were seated in the sound-treated booth, approximately 1 meter from the loudspeaker, and wore the research amplification device. IEEE/Harvard sentences were presented one at a time (and only once) in blocks of 25 sentences. The assignment of sentences to listening condition (amplification setting and presentation level) was randomized for each participant. The order of conditions was also randomized.

Before each sentence, a short piano melody, 800 ms in duration, was presented to notify the participant to prepare to listen. After the melody, a 500 ms silence occurred before the sentence was presented. Before testing began, participants were given instructions about the speech recognition assessment. They were told that a sentence would be presented and that they were to verbally repeat back as much of the sentence as possible even if they weren't completely sure of what they heard. Verbal responses were recorded using Audacity and a lavalier microphone connected to the experiment PC.

Each sentence had 5 words designated as keywords. After each response, the experiment administrator would use the Token software to mark which keywords were correctly recognized. After each block of 25 sentences the software saved a digital record of each sentence presented as well as which keywords were correctly identified. Recognition scores in each listening condition were then calculated as the total number of keywords correctly identified out of the 125 keywords presented in that condition. The recorded waveform of their verbal responses was also saved.

L. Data analysis

To quantify the results of self-adjustment, the sound output (i.e., the aided response) for each amplification setting was measured in a 2 cc coupler with the appropriate RECD applied for each participant. The input signal was the ISTS presented at the same nominal dB level as was used when the self-adjustment was made. For self-adjustments made with speech at 45 dBA, the response was measured for a 50 dB SPL input which was the lowest level input available on the Verifit system. The resulting gain should be similar to the gain for a 45 dBA input as both levels are below the compression kneepoint of the research device. Aided gain was calculated by measuring the spectrum levels at the location of the on-ear reference mic and subtracting this from the aided responses. Insertion gain, which represents the net change in sound level between the unaided and aided responses, was also calculated.

To summarize the gain-frequency responses of the amplification settings, average gain (or average aided response, as applicable) was calculated in two frequency bands by averaging the data at 250, 500, and 1000 Hz for the low frequency band and 2000, 3000, and 4000 Hz for the high frequency band. As a metric to determine how similar the various settings were to the NAL-NL2 fits, deviation of the self-adjusted settings from NAL-NL2 settings was calculated by subtracting each subject's NAL-NL2-based aided response from the aided responses of each other setting, measured using the same speech input level. This deviation from NAL-NL2 metric was calculated separately in each of the two frequency bands. For the 3 participants who did not return to receive a proper NAL-NL2 fitting, the NAL-NL2 prescriptive targets were subtracted from the self-adjusted response instead. Settings with positive deviations from NAL-NL2 have more gain than the prescriptive fit, while negative deviations indicate less gain than NAL-NL2.

A deviation from Self metric was calculated in a similar manner to compare the gain of the self-adjusted setting used in the preference and speech recognition tasks with the gain of other settings in the two frequency bands. The response of the second self-adjusted setting made in the 55 dBA presentation level condition was subtracted from the aided response of the other settings. A positive deviation from Self indicates that the setting has more gain than the self-adjusted setting, while a negative deviation from Self indicates that the self-adjusted settings had more gain.

III. Results

A. Self-adjustment retest reliability

Each participant completed two self-adjustments in each listening condition so that the reliability of their adjustments could be assessed. Retest reliability was good; when averaged across participants (in order to inspect reliability for each listening condition) the within-subject standard deviations ranged from 1.5 to 2.8 dB for low frequency insertion gain and from 2.0 to 4.9 dB for high frequency insertion gain. The only case in which the average within-subject standard deviation exceeded 3 dB was for the high frequency insertion gain in the condition with the most background noise (-5 dB SNR). Across all conditions, the average within-subjects standard deviations were 2.1 and 2.9 dB for gain in the low- and high-frequency bands. Due to the good retest reliability, for subsequent analysis and presentation, only the second self-adjustment each participant made in each condition was included, while the first self-adjustment was discarded.

B. Self-adjusted aided responses

Figures 4.5 and 4.6, show the aided responses for the self-adjusted settings plotted with respect to the NAL-NL2 targets for each participant at each frequency. Figure 4.5 shows the aided responses for the self-adjustments made in quiet backgrounds, while Figure 4.6 shows the aided responses for the self-adjustments made with speech at 65 dBA and in noise. The results of self-adjustments made when speech was presented at 65 dBA in quiet are shown in both figures. Responses plotted between the dotted lines are within ± 5 dB of the NAL-NL2 target, indicated by the solid line.

Because participants tended to have audiograms with little or no hearing loss at low frequencies the NAL-NL2 prescribed little or no gain at 250 or 500 Hz for a 65 dB SPL input, resulting in similar targets across participants at those frequencies. Despite this similarity, there was a large spread in the amount of gain in the self-adjusted conditions made in conditions with 65 dBA speech presented, with a range of about 20 to 30 dB between the aided responses of self-adjusted settings, depending on the SNR. As has been found previously, between-listener differences in self-adjusted settings were large (Perry et al., 2019).

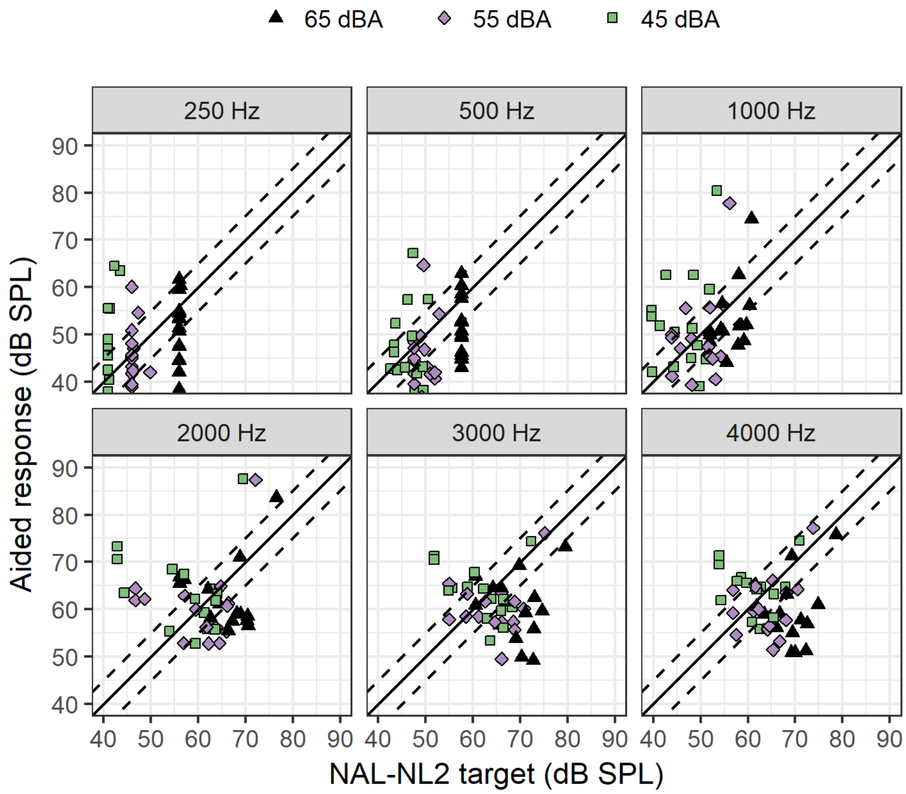

Figure 4.5. Same as Figure 4.4 but for self-adjusted settings made when speech was presented in quiet.

Figure 7 shows how the self-adjusted settings differed from the NAL-NL2 fits (or in the case of the 3 participants who did not receive proper NAL-NL2 fits, the NAL-NL2 targets) in the low frequencies (250, 500, and 1000 Hz) and the high frequencies (2000, 3000, and 4000 Hz). The black squares indicate the average deviation from NAL-NL2 across subjects. Between-subject variability in the deviation from NAL-NL2 gain for self-adjusted settings was large. According to the across-subject averages, self-adjusted settings tended to have more gain than the NAL-NL2 settings when self-adjustments were made listening to speech in quiet at the lowest presentation level used, 45 dBA. However, for the higher presentation levels (55 and 65 dBA in quiet), the average self-adjusted settings had less gain than the NAL-NL2 fits. Self-adjusted fits were most similar to NAL-NL2 fits when speech was at 55 dBA for self-adjustment. When background noise was present during self-adjustment, the average self-adjusted settings had less gain than NAL-NL2. The results of self-adjustments made with speech at 65 dBA (either in noise or in quiet) replicate previous findings using similar equipment and methods (Nelson et al., 2018).

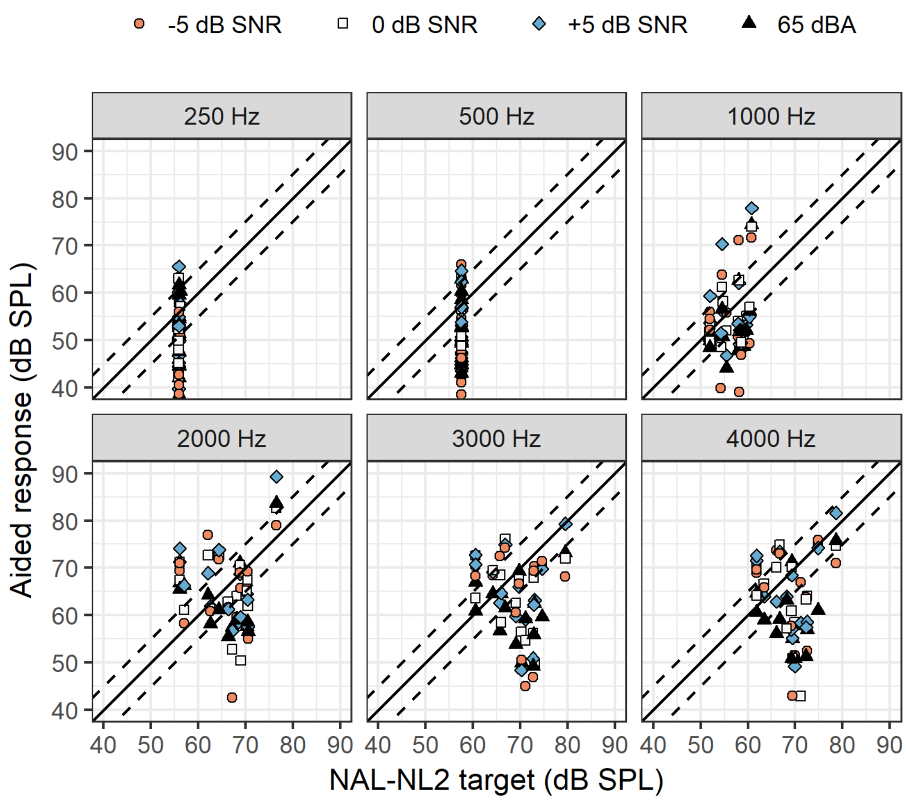

Figure 4.6. Same as Figures 4.4 and 4.5 but for self-adjusted settings made when speech was presented at 65 dBA in noise and in quiet. The aided responses for self-adjusted settings made for 65 dBA in quiet are duplicated from Figure 4.5.

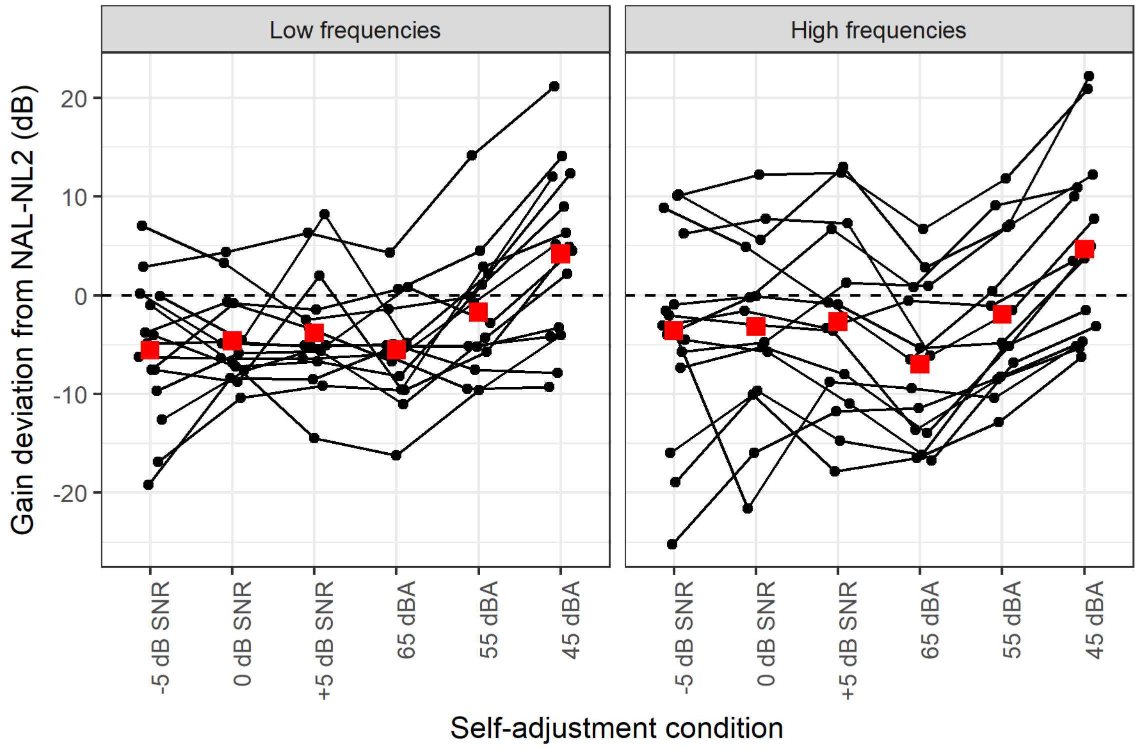

Figure 4.7. Deviation from NL2 for each self-adjusted setting, averaged in low and high frequencies. Each black line represents data from a single participant, while average results are shown as red squares. For conditions in noise, gain deviation was calculated using the ISTS at 65 dB SPL as input. For conditions in quiet, gain deviation was calculated using the ISTS at the same nominal level used during self-adjustment, except in the case of the 45 dBA condition for which the ISTS was set to 50 dB SPL

To evaluate whether subjects were consistent in their gain adjustments as listening conditions changed, bivariate correlations were calculated between gain deviation from NAL-NL2 within each frequency band for each unique pair of conditions, and p values were corrected for multiple comparisons using the Holm method. Similar to a previous study from our lab (Nelson et al., 2018), all correlations between conditions for high frequency values were statistically significant (all adjusted $p<0.05$) and correlation coefficients ranged between $r = 0.56$ and $r = 0.93$. Correlations were highest between conditions without background noise. Listener selections of self-adjusted high-frequency gain are fairly consistent across listening conditions, after accounting for average trends such as the tendency to reduce gain relative to NAL-NL2 as the speech level increases.

For the low frequency band, correlation coefficients were lower and only 3 of the 15 correlations reached statistical significance ($p < .05$). This suggests that participants were more consistent across listening conditions in their selection of high-frequency gain compared to their selection of low-frequency gain, which could be related to the constraints imposed by greater hearing loss in the high-frequencies (in contrast to thresholds in the low frequencies).

C. Individual ratings

Participants rated amplification settings while listening to speech at 2 different levels, 45 and 55 dBA. Figure 4.8 shows the distribution of ratings for the three categories, (overall satisfaction, subjective speech understanding, and loudness). During the first round of ratings, 1 listener rated their NAL-NL2 setting, their Self setting, and their HighCut setting. The other 14 participants rated the LowCut settings, their Self settings, and their MostCut settings. For the 11 participants who returned to be properly fit with NAL-NL2 during a third visit, they completed the rating again, this time rating their NAL-NL2 settings, their Self settings a second time, and their HighCut settings.

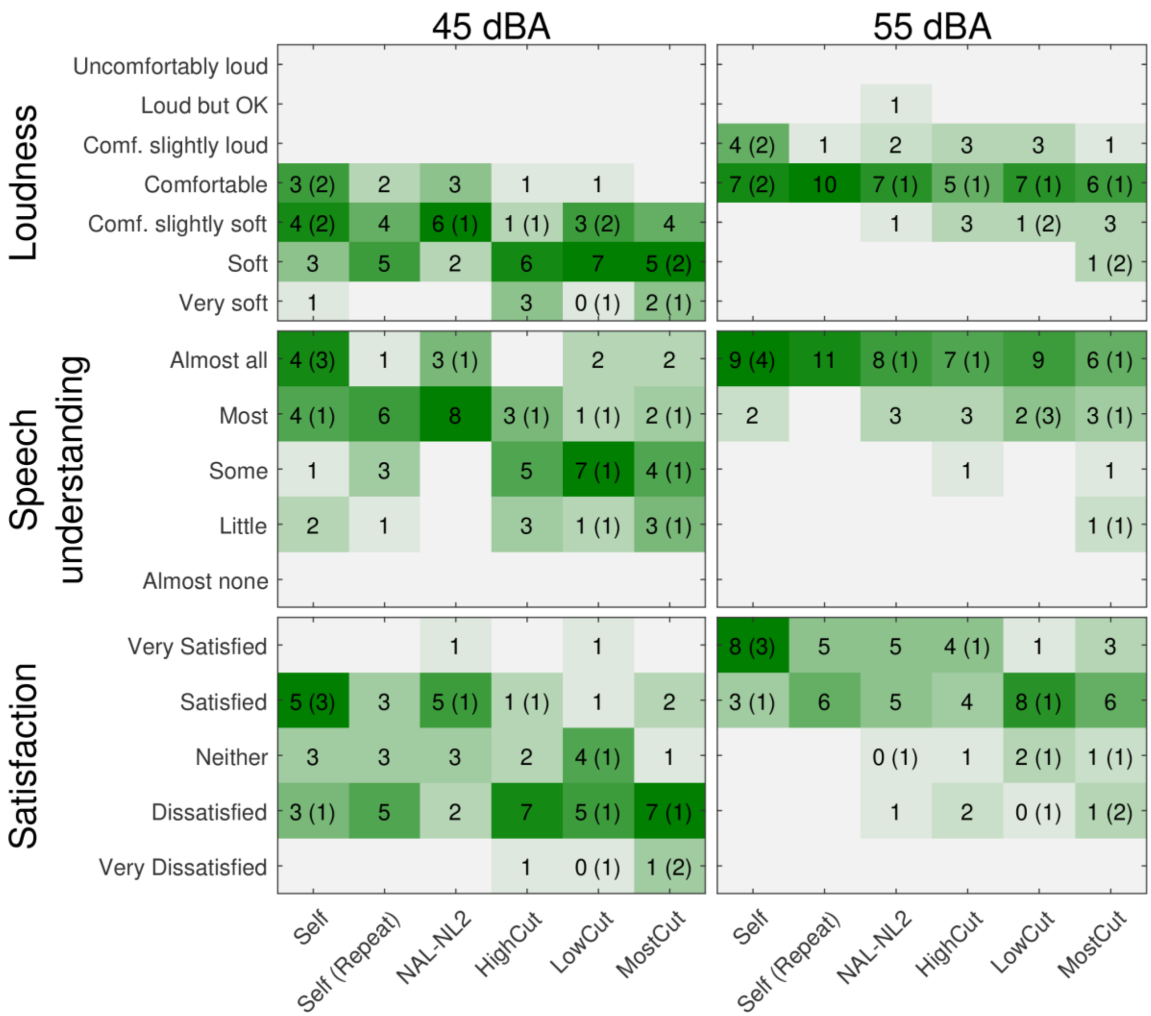

Figure 4.8. Heatmap of ratings of individual gain settings. Cells are colored according the total count of observations for each combination of setting and response, which is also shown as the values in each cell. Values in the parentheses are for the 4 participants who did not return for a third experiment session, and values outside of the parentheses are for the 11 participants who did return.

The speech presentation level had a clear effect on the ratings of loudness. Ratings in the 45 dBA condition showed reduced loudness than ratings in the 55 dBA condition. Ratings of loudness were fairly similar across the amplification settings, although there was a tendency for the Self settings and the NAL-NL2 settings to be reported as louder than the Cut settings, which is sensible considering that the Cut settings had less gain. The loudness judgments made in the 55 dBA condition demonstrate that participants are using self-adjustment to pick settings that have a comfortable loudness.

Ratings of estimated speech understanding show that participants felt they could understand most or almost all of speech using any setting so long as the input speech level was 55 dBA. For the 45 dBA presentation level, ratings of understanding were more variable. Following a similar pattern as the loudness ratings, the Self settings and NAL-NL2 settings were more frequently given higher ratings of speech understanding than the Cut settings. These data illustrate that participants selected settings which they believe would allow them to achieve a high level of speech understanding in quiet backgrounds.

The self-adjusted settings and NAL-NL2 fits produced the highest proportions of "satisfied" or "very satisfied" responses, with the self-adjusted settings receiving the highest proportion of "very satisfied" ratings. The settings with less gain usually had a greater proportion of "neither satisfied nor dissatisfied" or "dissatisfied" ratings. Satisfaction was lower in the 45 dBA condition than in the 55 dBA condition. The presentation level appeared to have a more pronounced effect on satisfaction with the Cut settings which were each deemed by a majority as dissatisfactory or very dissatisfactory in the 45 dBA condition despite being reported by a majority of participants as satisfactory or very satisfactory in the 55 dBA condition. This speech level effect highlights that for listening conditions which require less gain to achieve speech understanding, amplification settings with less gain can be seen as acceptable, but when greater gain is needed, the same settings can be perceived as undesirable.

As a reminder, the self-adjusted settings rated by participants were the settings taken from the second self-adjustment each participant made while listening to speech presented at 55 dBA. The self-adjusted settings made with speech presented at 45 dBA were not rated or otherwise assessed further. This means that for ratings made in the 45 dBA speech condition there was a mismatch between the conditions in which the participants made the self-adjustment and the level of speech presented during the ratings. As can be seen in Figure 4.7, the presentation level of the speech had an influence on the gain of the self-adjusted settings. In particular, self-adjustments made with speech at 45 dBA tended to produce fits which had more gain than both the NAL-NL2 settings and the self-adjusted fits made when speech was at 55 dBA. Because the self-adjusted settings made with speech at 45 dBA were not rated by participants, it is unknown if such settings with more amplification would have been rated as more satisfactory, louder, or as providing greater speech understanding. That said, given the trends seen in the ratings that were made, it would be surprising if the effect of speech level on gain selection and the effect of speech level on ratings of amplification settings were unrelated.

As the study was originally designed, participants were not expected to perform repeated ratings of their self-adjusted settings in the same listening conditions. However, because of experimenter error when fitting NAL-NL2 (described above in Methods), participants were invited to return to complete new ratings. This presented an opportunity to have participants repeat their ratings of their Self settings in order to assess the stability of their judgments over time. Furthermore, it was important to have participants rate their

Self settings in the same lab visit during which they rated their NAL-NL2 settings so as to mitigate possible anchoring biases that could influence the ratings. Thus, 11 participants gave ratings for their self-adjusted fits twice, with 1 to 6 months between the first and second ratings.

Across these participant's 66 repeat ratings of their Self settings (3 rating categories per 2 speech levels per 11 participants), exactly half (33) of all ratings remained the same as during the first round of ratings, while 30 differed by only one response step. One response step is the difference between "very satisfied" and "satisfied", for example, or the difference between "soft" and "very soft". The second ratings differed from the first ratings by two response steps in only 3 cases. Changes upon retest were more common for ratings done with the 45 dBA presentation level than with the 55 dBA presentation level. Compared to their first set of ratings, there was a weak tendency for the 11 returning participants to rate the self-adjusted settings lower on satisfaction and estimated speech understanding. Ratings of loudness were more similar on retest, especially for the 55 dBA presentation level. One possible explanation for these trends upon retest are that the NAL-NL2 settings could have provided a more favorable mental comparison (i.e., more satisfactory, greater subjective sense of speech understanding) than the MostCut and LowCut settings that were rated alongside the Self settings during the first rating session.

To summarize the individual ratings of amplification settings, ratings of settings made with a lower speech presentation level (45 dBA) were typically less favorable than when ratings were made with speech at 55 dBA. The settings with the least amount of gain were rated more often as being less satisfactory and facilitating less speech understanding than the settings with more gain. NAL-NL2 and self-adjusted fits were frequently rated as having a comfortable loudness, being satisfactory or very satisfactory, and providing high degrees of subjective speech intelligibility.

D. Paired comparisons

Participants made blinded paired comparisons between their Self settings and the other settings as applicable (NAL-NL2, LowCut, HighCut, and MostCut)[1]. Figure 4.9 shows the distributions of paired comparison responses for each setting compared to the self-adjusted fit, with comparisons made using a 45 dBA speech level on the bottom row of plots. Focusing on the overall preferences for the Self settings over comparison settings, Self settings were frequently preferred over the Cut settings, especially when the speech presentation level was 45 dBA. When participants compared Self with NAL-NL2 settings, preferences were split. Seven out of 12 participants reported an overall preference for their Self settings when listening to 55 dBA speech, and 5 out of 12 preferred their Self settings over their NAL-NL2 settings when listening to 45 dBA speech. It should be noted in the 55 dBA condition preferences for the Self settings tended to be stronger, while preferences in the 45 dBA condition were more often slight.

For a given pair of settings, paired comparison preferences were usually consistent across the rating categories. All bivariate correlations computed between each

[1] Due to experimenter error, three participants were never given an appropriate NAL-NL2 fit and thus never made a paired comparison between their self-adjusted settings and their NAL-NL2 settings. Four participants, including the three who did not rate or compare an appropriate NAL-NL2 fit, did not receive a HighCut setting or make paired comparisons involving a HighCut setting.

possible pair of rating categories for the same pair of settings were statistically significant after adjusting for multiple comparisons using the Holm method (all adjusted $p < .05$) and correlation coefficients ranged from $r = 0.65$ to $r = 0.87$. Ratings of loudness preference and speech clarity showed the strongest correlations with the overall preference ($r = 0.87$ and $r = 0.86$, respectively), though sound quality and comfort also had strong correlations with overall preference ($r = 0.75$ and $r = 0.79$, respectively). The consistency of judgments across categories makes it difficult to determine whether some attributes contributed more to the overall preference, an issue that has previously been reported regarding relationships between judgments of sound quality (Preminger and Van Tasell, 1995).

To summarize, participants frequently preferred their self-adjusted settings over settings that had less gain, and this preference was often strong. The presentation level of speech appears to influence the preference judgments by biasing the preference toward the setting with more gain when the input speech level is lower. When participants compared their self-adjusted settings with NAL-NL2 settings, overall preferences were split with about half the participants preferring their self-adjusted setting and about half preferring their NAL-NL2 settings. This result is consistent with the participants' individual ratings of settings reported in the previous section. NAL-NL2 and the self-adjusted settings were given similar ratings in terms of satisfaction, estimated speech understanding, and loudness, so it is not surprising that the results of paired comparisons show a split among participants for which setting is preferred more overall.

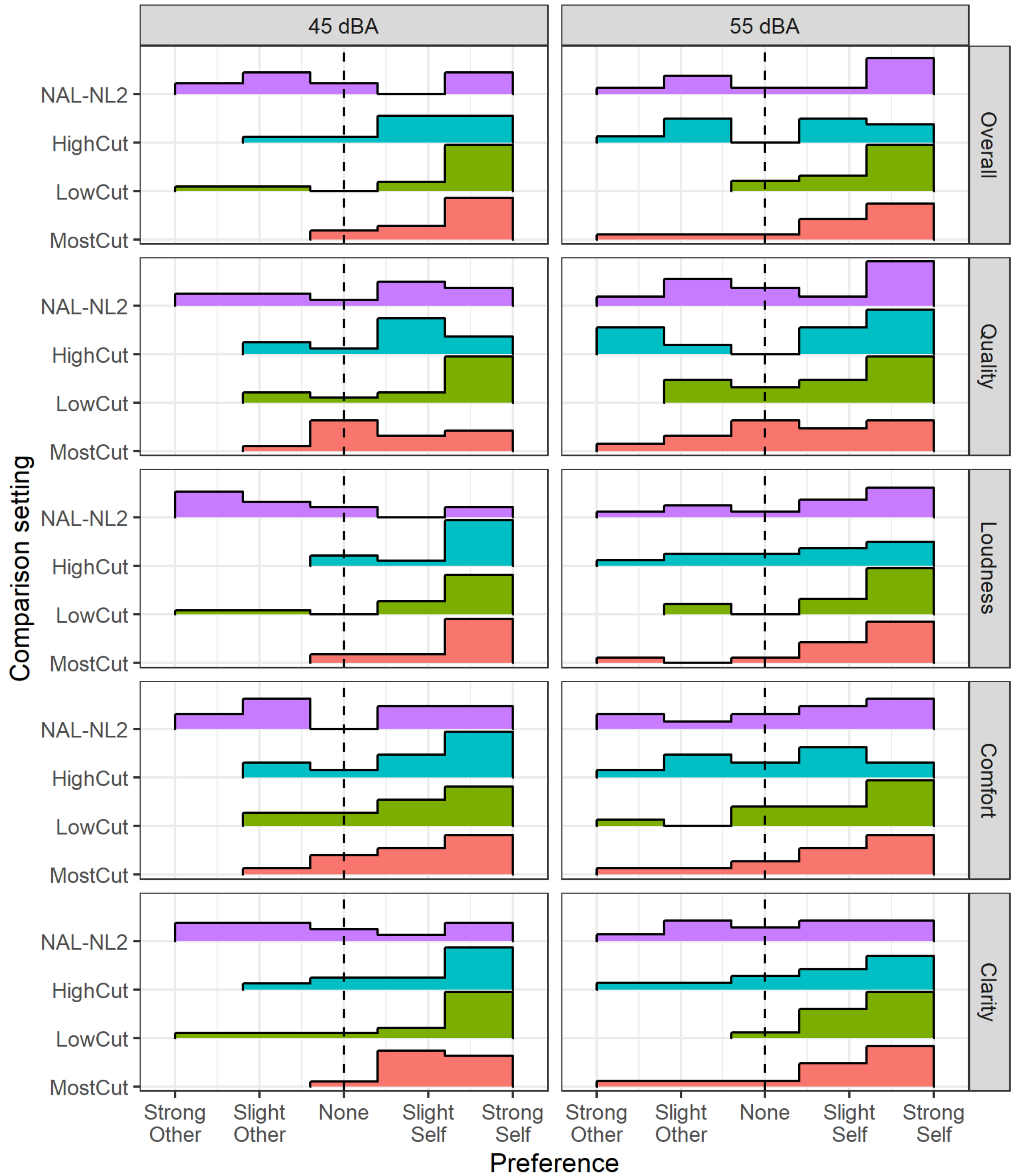

Figure 4.9. Binned distributions of responses for paired comparisons. In each plot, the strength of preference for either the comparison setting or the Self setting is indicated by relative position left to right. Bin heights represent the counts of preferences across all participants.

E. Speech recognition

Speech recognition was assessed at two presentation levels, 45 and 55 dBA. Figure 4.10 shows the percent of sentence keywords correctly identified by each subject, plotted with respect to the amplification setting used during testing. When sentences were presented at 55 dBA, all participants correctly recognized at least 90% of the keywords when using their self-adjusted settings and when using their NAL-NL2 settings. Speech recognition performance was also high with the other amplification settings for the 55 dBA presentation level, although in a few cases performance with the MostCut and HighCut settings was much worse. This result illustrates the point that for many listeners with mild-to-moderate hearing loss, settings with less gain than NAL-NL2 can be sufficient to achieve a high degree of speech recognition when listening to speech in quiet at an average conversational level.

In contrast, when sentences were presented at the lower level, 45 dBA, gain differences between amplification settings produced larger variability in speech recognition. The best average performance, 91.7% of keywords identified correctly, was achieved using NAL-NL2 settings, while performance with self-adjusted settings was about 12 %-points worse at 79.8% correct. Performance with the Self setting was meaningfully better than NAL-NL2 in only a single case, though in several other cases performance was similar between the two settings. The other three settings each resulted in worse average performance than either NAL-NL2 or self-adjusted settings.

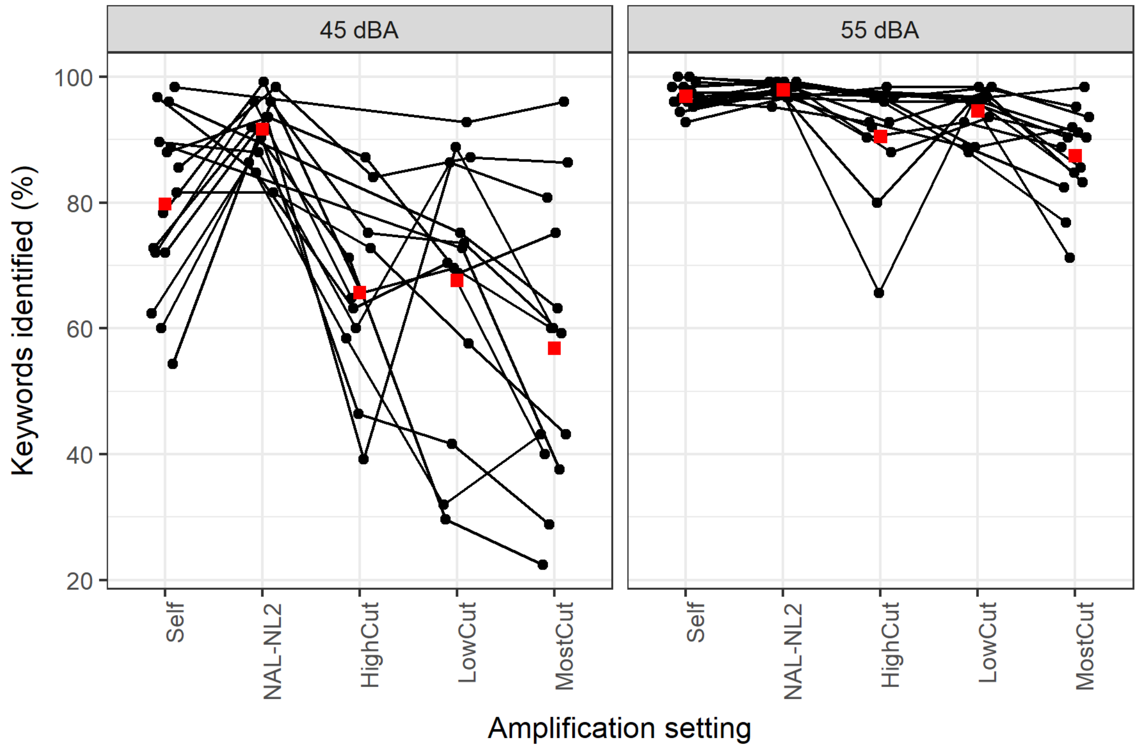

Figure 4.10. Sentence keyword recognition scores. Each black line represents data from a single participant, while average results are shown as red squares.

F. Relating preference, gain, and speech recognition

The relationships between speech recognition performance, overall preference, and average insertion gain (in the two frequency bands as well as averaged across both bands) can be seen in Figure 4.11. In these plots the horizontal location, left to right, is given by the overall preference from the paired comparison data, with points on the left of the center line indicating that the comparison condition (either NAL-NL2, LowCut, HighCut, or MostCut) was preferred over the Self setting, and points to the right of the center line indicating that the Self setting was preferred. The distance from the horizontal midline indicates the strength of the preference. The vertical location of each point indicates the speech recognition advantage of the Self settings over the comparison settings. The higher above the vertical midline, the better the speech recognition performance when using the Self settings compared to when using the comparison setting. Points below the vertical midline are cases in which the Self settings resulted in worse speech recognition than that of the comparison setting. Therefore, points plotted in the top right quadrant of each plot are cases in which speech recognition was better with the Self setting and the participant also indicated that they preferred the Self setting. Points in the bottom left quadrant of each plot represent cases in which speech recognition was poorer with the Self setting and the participant did not prefer the Self setting. The color of the points represents the difference in gain between the Self setting and the comparison setting (i.e., the deviation from Self). Negative deviations from Self are shown as red colors indicating that the comparison setting had less gain than the Self setting. Positive values are assigned to blue colors and indicate that the comparison setting had more gain than the Self setting.

When speech was presented at 55 dBA, there was little difference in speech recognition performance between the settings so most points are clustered near the midline. However, for the cases in which speech was presented at 45 dBA, points are plotted primarily either in the top right or bottom left quadrants, indicating a congruence of preference and speech recognition performance. In other words, the trend shown in this figure is that listeners prefer the self-adjusted settings when there is an advantage to speech recognition, and prefer the comparison settings when speech recognition is better with the comparison settings. This observation is supported by the correlation between the two values: the correlation between the overall preference values and the speech recognition advantage for self-adjusted fits is moderately strong and statistically significant (r = .63, p<.001) for the 45 dBA conditions, but not for the 55 dBA conditions (r = .16, p = .25).

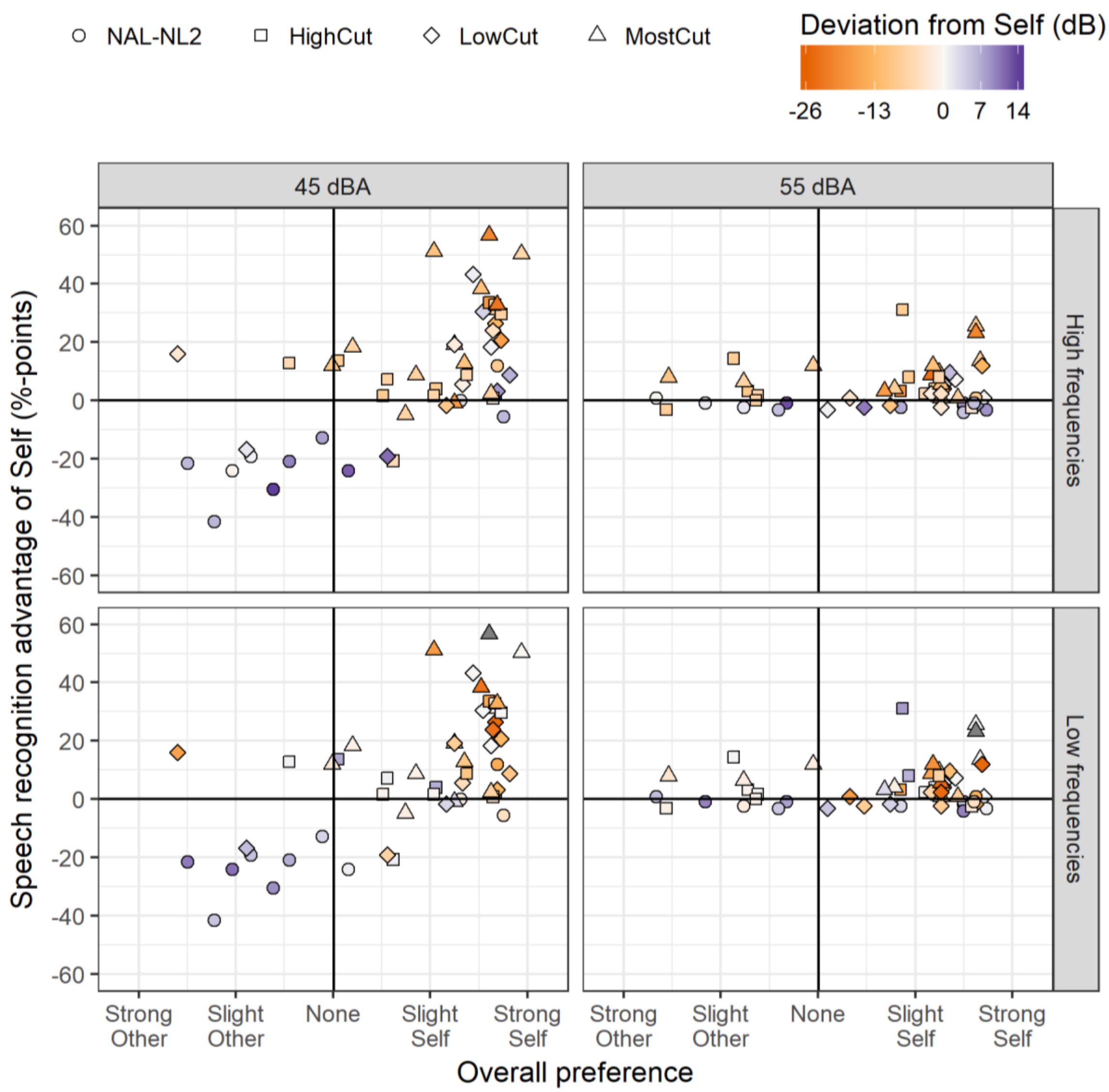

Figure 4.11. Speech recognition advantage of the Self settings over comparison settings plotted with respect to the overall preference from the paired comparisons. Each point represents a comparison setting for a single participant, and the shapes are filled according to the gain deviation of the comparison setting in the low frequencies or high frequencies, with more positive deviations indicating that the comparison setting had more gain than the Self setting.

As discussed above, the only amplification setting which resulted in higher average speech recognition performance than the Self settings was the NAL-NL2 settings at the 45 dBA presentation level. The NAL-NL2 settings also tended to have more gain. Across the non-Self settings, bivariate correlations confirm that there are statistically significant relationships between the gain deviation from Self and overall preference (r = -.52, p<.001) as well as the gain deviation from Self and speech recognition advantage for Self settings (r = -.72, p<.001) at the 45 dBA presentation level. These correlations are consistent with the premise that people with mild-to-moderate hearing loss would prefer to obtain a communication benefit when using amplification, and that benefit depends on the hearing aid applying sufficient gain to meaningfully increase speech audibility when the speech level is low.

This interpretation is supported by another trend in the data presented here. Under conditions where differences in gain can have meaningful effects on speech understanding, and in particular, under conditions (like a 45 dBA presentation level) where NAL-NL2 settings result in less than full speech recognition on average, self-adjustment tends to result in more gain than NAL-NL2 fits, as seen in Figure 4.7. Speech recognition was not assessed using the self-adjusted fits made when participants were listening to speech at 45 dBA, but based on the gain values of those fits, it’s likely that speech recognition performance with the self-adjusted settings selected while listening to speech at 45 dBA would have resulted in speech recognition that was at least as good as the performance with NAL-NL2 fits. Put another way, these data suggest that preference for gain is guided by the listener’s ability to understand speech when using that gain, and that depending on the listening conditions under which the self-adjustments are made, there appears to be a strong tendency for people to use self-adjustment to choose amplification settings which provide a communication benefit comparable to what would be achieved through an audiologist-fit hearing aid.

IV. Discussion

The findings of the present study replicate and extend existing literature supporting the impetus for the use of self-adjustment of hearing aid gain for people with mild-to-moderate sensorineural hearing loss (Boothroyd and Mackersie, 2017; Boymans and Dreschler, 2012; Keidser et al., 2012; Mackersie et al., 2018; Nelson et al., 2018). Specifically, self-adjusted settings showed large variability between listeners in a range of about 30 dB, indicating that self-adjustment can result in substantially different amplification parameters than NAL-NL2. In addition, self-adjusted settings were reported to be at least as satisfactory as NAL-NL2 settings. When the speech level used to evaluate preferences matched the level used during self-adjustment, self-adjusted settings that were often strongly preferred over NAL-NL2 settings in paired comparisons. This demonstrates that the individual customization provided by self-adjustment can improve satisfaction with amplification.

Two of the most commonly-offered reasons that hearing aid owners provide for their non-use of hearing aids fit by hearing care professionals is that they feel the hearing aids have undesirable sound quality and do not provide sufficient speech recognition benefit (McCormack and Fortnum, 2013). A feature of the present study is that self-adjusted gain, preference for gain, and speech recognition were each evaluated using the same participants which enables any relationships between the three concepts to be identified. Relationships between all three measures were found which indicate that when speech recognition differs between amplification settings, listeners preferred the settings that facilitated greater speech recognition. When speech recognition did not differ meaningfully between settings, self-adjusted settings were generally preferred. In addition, when speech presentation levels were low, self-adjustment resulted in more gain on average than the settings based on the NAL-NL2 prescription. This suggests that when participants adjusted amplification parameters they were guided by criteria that included their subjective sense of how well they could understand the speech.

Altogether these results support the assertion that people with mild-to-moderate hearing loss are capable of using a modern, commercially-available self-adjustment tool to adjust the amplification parameters of a compression-amplification device, in real time, in order to achieve a satisfactory fit that benefits speech understanding. The data support an even stronger statement about self-adjustment of amplification: when self-adjustment could possibly result in a reduced speech recognition benefit, the most likely predicted outcome is that settings that provide insufficient audibility will be avoided. There is good reason to take the complaints of hearing aid owners at face value. The findings of this study confirm that people with hearing loss want their hearing aids to provide a meaningful communication benefit, and given the chance to alter the functioning of their hearing aids, they will likely choose parameters that satisfy their subjective sense of sound quality and loudness while avoiding settings that provide inferior speech audibility.

The results of this study are also relevant to understanding gain preferences for audiologist-fit hearing aids with no self-adjustments. Averaged across participants, self-adjusted gains were most similar to NAL-NL2 settings at the level of 55 dBA, an average level of speech in quiet locations. Averaged across listeners, self-adjustments appear to approximate the gain prescribed by NAL-NL2, which validates the formula as a useful method for fitting an average adult with mild-to-moderate sensorineural hearing loss.

In the current study NAL-NL2 was not used as the baseline starting settings from which self-adjustments were made. Rather, the starting settings were the LowCut settings which most often had gain similar to NAL-NL2 targets in the high frequencies and much less than NAL-NL2 targets in the low frequencies. Despite the generally undesirability (and audiological inappropriateness) of the baseline setting, participants were still able to achieve highly satisfactory fits using self-adjustment. Self-adjustments are often completed quickly (Perry et al., 2019). Starting from a NAL-NL2 baseline to which the user can return as desired might improve the efficiency of self-adjustment. It's also possible that a generic fit to an average audiogram might be sufficient as well, though the appropriateness of this approach might depend on the slope of the individual's hearing loss (Keidser, Dillon, et al., 2008).

NAL-NL2 appears to prescribe more gain than listeners desire for higher level inputs such as for speech at 65 dBA. For people with mild-to-moderate hearing loss, 65 dBA is a high enough level that unaided speech audibility is likely to be high, which lessens the usefulness of amplification, consistent with the model predictions describe by Plomp (1986). This is pattern is consistent regardless of whether speech is in quiet or accompanied by noise at a moderate SNR (Nelson et al., 2018).

Although the deviation from NAL-NL2 gain observed for 45 dBA might be interpreted as evidence that NAL-NL2 prescribes too little gain for lower level inputs, that interpretation requires several caveats. First, preferences and speech recognition were not evaluated for the self-adjusted settings participants chose when listening to a 45 dBA input; only the self-adjusted settings from the second self-adjustment in the 55 dBA condition were evaluated. Second, the conditions under which listeners completed the study tasks generally did not involve much exposure to non-speech environmental sounds. While increased gain for lower level inputs might help speech understanding, it is also likely to increase the annoyance of environmental sounds which are often reported as undesirable. Studies of expansion amplification report a trade-off between speech intelligibility and annoyance for different amounts of gain for lower level inputs. Gain preferences for lower level inputs likely depend upon the desirability of the sounds that are used to judge preferences. Listeners prefer more amplification when listening to low level speech but prefer less amplification when listening to undesirable stimuli (see Brennan and Souza, 2009). Because self-adjustments for low level inputs were made only with speech stimuli, the current data are insufficient for evaluating the real-world desirability of NAL-NL2 prescribed gain for low level inputs.

Some potential limitations of the current study stem from the recruited participant group and study methodology. Data from only 12 participants were available to make comparisons between NAL-NL2 settings and self-adjusted settings. Further, it's possible that the present cohort of participants may not be wholly representative of the population of potential users of self-adjustment technology. Based solely on their willingness to volunteer to participate in hearing research, the study participants could possibly have higher-than-average interest in their own hearing loss or perhaps higher hearing health self-efficacy than the average adult. These factors were not assessed, so it's not possible to evaluate the representativeness of the participants, let alone what influence, if any, such individual differences might have. Future studies could be designed to provide clearer answers to these questions.

Another limitation of the current study is that it was an acute laboratory study and thus it was not possible to determine the effects of self-adjusted amplification on broader health outcomes such as social participation and quality of life. Based on the results of speech recognition benefit, satisfaction, and overall preference presented here, it's reasonable to speculate that the use of hearing aids that are customized using self-adjustment will likely produce outcomes at least as good as those produced by the use of audiologist-fit hearing aids. Future research should identify the long-term effects of self-adjustment on quality of life and social participation as well as what influence the availability of self-adjustable hearing aids might have on hearing aid adoption and use.

V. Conclusion

People with mild-to-moderate hearing loss used self-adjustment of hearing aid gain and compression to achieve satisfactory amplification settings. Speech recognition using these settings was generally near maximum when sentences were presented at a sound field level of 55 dBA. At a lower speech presentation level, amplification settings with more gain tended to result in better speech recognition performance than settings with less gain, and these higher-gain settings were more frequently preferred by listeners in blinded paired comparisons. The present data imply that listeners use their subjective sense of speech clarity as a criterion for selecting amplification settings during self-adjustment and as a criterion for evaluating the satisfactoriness of amplification settings. Concerns that self-adjustment will result in settings that provide reduced communication benefit or reduced satisfaction compared to audiologist-fit hearing aids are not supported by these findings. Instead, allowing people with hearing loss a wide degree of control over the gain and compression parameters of their hearing aids is likely to result in the selection of satisfactory settings which have a comfortable loudness. Although listeners also found their NAL-NL2 settings to generally be satisfactory, the gain of self-adjusted settings differed from the gain of NAL-NL2 fits within a large range of 30 dB. Self-adjustment can be a useful method for individualizing hearing aid parameters according to the personal sound quality goals of the wearer without reducing the speech perception benefits of amplification.

Chapter 5: General Conclusions

Dissatisfaction with hearing aids is a long-standing issue which deters hearing aid adoption and use (Kochkin, 2007). Providing hearing aid wearers greater control over amplification parameters is one approach that could increase satisfaction. The projects described in these chapters indicate that preferred listening levels can differ considerably from the output prescribed by widely-used prescriptive formula such as NAL-NL2. Large individual differences in preferred hearing aid gain demonstrate a need for increased customization of amplification parameters. Fine-tuning of gain and compression across multiple frequency channels can be accomplished rapidly and reliably using self-adjustment tools that provide access to a large parameter space via a simple touchscreen interface. These three related projects describe the use of self-adjustment by participants with mild-to-moderate hearing loss and compared self-adjusted settings to NAL-NL2 prescribed settings across a number of measures, including speech recognition and subjective preference. Core findings are listed below.

Study 1 – Self-adjustments of Amplification in Noisy Backgrounds:

1. According to the group average, NAL-NL2 prescriptions were close to what was selected by participants. Self-adjustments made when speech was presented at 65 dBC in a quiet background had an average of about 5 dB less high frequency (2 to 8 kHz) gain, and about 5 dB more low frequency (0.125 to 1 kHz) gain compared to NAL-NL2 fits. Using the common clinical criterion for a high quality fit is that the measured output is within ± 5dB of

the prescribed target (e.g., Hashir et al., 2012), the gain of the average self-adjusted fit can be considered to be similar to the average NAL-NL2 fit for these participants.

2. Retest reliability was high. Participants were consistent in their self-adjustments within the same listening conditions as well as across similar listening conditions.
3. Inter-subject variability in self-adjusted gain was large, indicating a need for self-adjustment in order to better match amplification settings to individual preferences, particularly for people whose preferred gain differs greatly from NAL-NL2 prescribed settings.
4. No systematic differences were found in the speech recognition performance achieved with NAL-NL2 settings and self-adjusted settings for speech presented at 65 dBC in quiet and in noise at a range of SNRs.
5. At the most favorable SNR considered (+5 dB) the presence of noise had only a small impact on the average gain of self-adjusted fits, differing by less than 1 dB from the average gain selected in the quiet background.
6. As the noise level increased and the SNR became less favorable, participants selected progressively less gain across all frequencies, even after considering the effects of compression. Self-adjustments made in different simulated restaurant noise environments were largely similar when made at the same SNR. The spectral attributes of the noisy backgrounds had a small influence

on the selected gain, while the temporal attributes had little or no effect on self-adjustment.

A. Study 2 – Between-participant variability and listener factors:

4. Listeners completed self-adjustment rapidly using the research device. The median duration of self-adjustment was less than 1 minute and all but one self-adjustment took less than 3 minutes to complete.
5. As noise levels increased, the duration of self-adjustments increased, indicating that as the listening conditions became more challenging, participants took more time to select amplification settings.
6. The insertion gain of self-adjusted fits was related to the hearing thresholds of the participants; people with more hearing loss tended to select more gain in the frequency region of their hearing loss. However, the deviation of self-adjusted gain from the participants' NAL-NL2 fits was poorly explained by listener factors. Participants with better thresholds and less experience using hearing aids tended to deviate further from NAL-NL2 settings, but due to a confound between hearing thresholds and hearing aid experience, it's unclear which attribute contributed most to this effect. Self-adjustment is likely to be a more straightforward and feasible way to customize hearing aid gain than making empirical adjustments to prescriptive fitting formula based on listener factors.

B. Study 3 – Self-adjustments for low-level speech in quiet

5. Findings from Study 1 in which audibility was limited by noisy backgrounds were confirmed for low-level speech in quiet. The average deviation of self-adjusted gain from NAL-NL2 fits depended on the presentation level of speech when in a quiet background. Listeners selected more gain when listening conditions required more amplification to achieve speech audibility.
6. Listeners frequently rated both their self-adjusted fits and their NAL-NL2 as satisfactory. At the higher speech presentation level, self-adjusted settings tended to be more strongly preferred over NAL-NL2 settings. At the lower speech presentation level, overall preferences between the two settings were weaker and were related to the relative difference in speech recognition achieved when using the settings.
7. Listeners did not seem to select or prefer gain settings that resulted in suboptimal performance. Listeners very frequently rated settings with reduced gain as being less satisfactory. In paired comparisons, self-adjusted settings were overwhelmingly preferred over settings with reduced gain.
8. Under conditions where speech recognition performance was not at the upper limit, preference for self-adjusted fits over other settings was correlated with the relative speech recognition advantage of self-adjusted fits as well as the difference in gain between self-adjusted fits and comparison settings. Settings with more gain tended to facilitate better speech recognition and participants tended to prefer those settings.

I. Clinical Implications and Future Directions

Altogether, the results of these experiments provide evidence that self-adjustment of hearing aid parameters produces satisfactory settings with speech recognition benefits comparable to those of hearing aids fit using clinical best practices. Best practices for fitting hearing aids include setting the gain and compression parameters across the frequency range to meet prescriptive targets provided by an empirically-validated fitting formula. Probe microphone measures of the sound levels in the ear canal are often recommended to verify the accuracy of the fit. However, prescriptive targets reflect the average preferred listening level, and preferences for hearing aid gain vary greatly from person to person. As shown in Studies 1 and 3, some people prefer to have hearing aids with settings that would be considered a bad clinical fit.

Concerns about the goodness of fit typically stem from models of speech audibility and intelligibility in quiet as well as models of loudness perception. The communication benefit of using hearing aids depends upon increasing the audibility of speech energy through amplification. Prescriptive targets are set so as to increase the audibility of speech in the frequency regions of the hearing loss while controlling the overall loudness of sound. If a hearing aid is configured to have less amplification than the prescribed targets, the benefit provided by the hearing aid may be reduced. If the hearing aid is configured to have more amplification than the prescribed targets, the sound from the hearing aid may be perceived as overly loud and unsatisfactory. Following this logic, self-adjustment may be viewed skeptically by some clinicians who wonder whether adjustments will result in settings that provide less benefit or satisfaction than settings produced by following best practices.

The findings of this dissertation argue against overly-strict adherence to fitting to targets and should assuage concerns that giving the hearing aid wearer a large degree of control over the fit of their device will lead to poorer speech recognition or satisfaction. Specifically, the finding from Study 3 that relative preference for gain has a moderately-strong positive relationship with relative speech recognition performance suggests that self-adjustment of gain and compression is guided, at least in part, by the listener's subjective sense of speech clarity. In addition, self-adjusted settings were at least as satisfactory to listeners as NAL-NL2 settings and were reported as having a comfortable loudness. People with hearing loss want their hearing aids to help them understand speech, and they are likely to reject settings that provide noticeably inferior speech audibility or that are uncomfortably loud.

The source of individual variability in the difference between preferred gain and prescribed gain remains unclear. The listener factors examined in Study 2, including age and hearing thresholds, explained less than 10% of this variability. Provided that self-adjustment can be used to individualize amplification parameters effectively, understanding the source of the variability may be an ancillary research goal. The loudness ratings of self-adjusted settings reported in Study 3 hint at one possible explanation. When the speech presentation level used to rate loudness was the same as the level during self-adjustment, the loudness of self-adjusted settings were always reported as either comfortable or comfortable but slightly loud. Ratings of the loudness of

NAL-NL2 ratings were more variable with comfortable being the most frequently selected descriptor. This result is curious since self-adjusted gain spanned a range of nearly 30 dB relative to the gain of NAL-NL2 settings. In light of reports of similarly large variability among people with hearing loss in the loudness perception of broadband, binaural sounds (Oetting et al., 2016), it’s possible that the variability seen in the deviation of self-adjusted gain from NAL-NL2 gain reflects loudness perception differences between listeners. If this is the case it could have clinical implications for the design of compression systems in hearing aids (Oetting et al., 2018).

Another avenue for future research entails assessing the impact of self-adjustment on hearing aid adoption and use, quality of life, social participation, and other important health outcomes. The acute measures of speech recognition and satisfaction reported in the current studies do not directly address the effects that self-adjustment may have when available for everyday use, but at the very least the results suggest that hearing aids fit with self-adjustment are not likely to lead to worse health outcomes. Additional information is needed to determine the broader consequences of self-adjustment of amplification parameters.

Bibliography

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*, S53–S71. doi:10.1080/14992020802301142.

Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, *65*, 943–950.

Amlani, A. M. (2016). Application of the Consumer Decision-Making Model to Hearing Aid Adoption in First-Time Users. *Seminars in Hearing*, *37*, 103–119. doi:10.1055/s-0036-1579706.

Amlani, A. M., & Schafer, E. C. (2009). Application of paired-comparison methods to hearing aids. *Trends in Amplification*, *13*, 241–259. doi:10.1177/1084713809352908.

ANSI, A. (1997). S3. 5-1997, Methods for the calculation of the speech intelligibility index. *New York: American National Standards Institute*, *19*, 90–119.

Banerjee, S. (2011). Hearing aids in the real world: Use of multimemory and volume controls. *Journal of the American Academy of Audiology*, *22*, 359–374. doi:10.3766/jaaa.22.6.5.

Bartoń, K. (2018). *MuMIn: Multi-Model Inference*. R package version 1.40.4. Retrieved from https://CRAN.R-project.org/package=MuMIn.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi:10.18637/jss.v067.i01.

Bentler, R., Wu, Y.-H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: Outcomes from laboratory and field studies. *International Journal of Audiology*, *47*, 447–460. doi:10.1080/14992020802033091.

Bentler, R. A. (2005). Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence. *Journal of the American Academy of Audiology*, *16*, 473–484.

Boothroyd, A., & Mackersie, C. (2017). A "Goldilocks" approach to hearing-aid self-fitting: User interactions. *American Journal of Audiology*, *26*, 430–435. doi:10.1044/2017_AJA-16-0125.

Boymans, M., & Dreschler, W. A. (2000). Field trials using a digital hearing aid with active noise reduction and dual-microphone directionality. *Audiology*, *39*, 260–268. doi:10.3109/00206090009073090.

Boymans, M., & Dreschler, W. A. (2012). Audiologist-driven versus patient-driven fine tuning of hearing instruments. *Trends in Amplification*, *16*, 49–58. doi:10.1177/1084713811424884.

Brennan, M., & Souza, P. (2009). Effects of expansion on consonant recognition and consonant audibility. *Journal of the American Academy of Audiology*, *20*, 119–127.

Brons, I., Houben, R., & Dreschler, W. A. (2014). Effects of noise reduction on speech intelligibility, perceived listening effort, and personal preference in hearing-impaired listeners. *Trends in Hearing*, *18*, 2331216514553924. doi:10.1177/2331216514553924.

Byrne, D. (1986). Effects of frequency response characteristics on speech discrimination and perceived intelligibility and pleasantness of speech for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *80*, 494–504. doi:10.1121/1.394045.

Byrne, D. (1996). Hearing aid selection for the 1990s: where to? *Journal of the American Academy of Audiology*, *7*, 377–395.

Congress, U. S. (2017). *HR 2430: FDA Reauthorization Act of 2017. August 29, 2017*.

Convery, E., Keidser, G., Dillon, H., & Hartley, L. (2011). A self-fitting hearing aid: Need and concept. *Trends in amplification*, *15*, 157–166.

Cox, R. M., & Alexander, G. C. (1991). Preferred hearing aid gain in everyday environments. *Ear and hearing*, *12*, 123–126.

Cox, R. M., & Alexander, G. C. (1992). Maturation of hearing aid benefit: Objective and subjective measurements. *Ear and Hearing*, *13*, 131–141. doi:10.1097/00003446-199206000-00001.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing*, *8*, 119s. doi:10.1097/00003446-198710001-00010.

Cox, R. M., Alexander, G. C., Taylor, I. M., & Gray, G. A. (1997). The Contour Test of Loudness Perception. *Ear and Hearing*, *18*, 388–400.

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*, 214. doi:10.1097/AUD.0b013e31816453a6.

Elberling, C., & Hansen, K. V. (1999). Hearing instruments: Interaction with user preference. In: *AN Rassmussen, PA Osterhammel, T. Andersen, et al. Auditory Models and Non-Linear Hearing Instruments, Proceedings of the 18th Danavox Symposium*,.

Franklin, C. A., Thelin, J. W., Nabelek, A. K., & Burchfield, S. B. (2006). The effect of speech presentation level on acceptance of background noise in listeners with normal hearing. *Journal of the American Academy of Audiology*, *17*, 141–146. doi:10.3766/jaaa.17.2.6.

Hartley, D., Rochtchina, E., Newall, P., Golding, M., & Mitchell, P. (2010). Use of hearing aids and assistive listening devices in an older Australian population. *Journal of the American Academy of Audiology*, *21*, 642–653.

Hashir, A., Moore, B. C. J., & Deepak, P. (2012). The Accuracy of Matching Target Insertion Gains With Open-Fit Hearing Aids. *American Journal of Audiology*, *21*, 175–180. doi:10.1044/1059-0889(2012/11-0008).

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an international speech test signal (ISTS). *International journal of audiology*, *49*, 891–903.

Hornsby, B. W. Y., & Mueller, H. G. (2008). User preference and reliability of bilateral hearing aid gain adjustments. *Journal of the American Academy of Audiology*, *19*, 158–170. doi:10.3766/jaaa.19.2.6.

Horwitz, A. R., & Turner, C. W. (1997). The time course of hearing aid benefit. *Ear and Hearing*, *18*, 1–11. doi:10.1097/00003446-199702000-00001.

Hougaard, S., & Ruf, S. (2011). EuroTrak 1: A consumer survey about hearing aids in Germany, France, and the UK. *Hearing Review*,.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America*, *122*, 1777–1786. doi:10.1121/1.2766778.

Humes, L. E., Watson, B. U., Christensen, L. A., et al. (1994). Factors Associated With Individual Differences in Clinical Measures of Speech Recognition Among the Elderly. *Journal of Speech, Language, and Hearing Research*, *37*, 465–474. doi:10.1044/jshr.3702.465.

Humes, L. E., Wilson, D. L., Barlow, N. N., & Garner, C. (2002). Changes in hearing-aid benefit following 1 or 2 years of hearing-aid use by older adults. *Journal of Speech, Language, and Hearing Research*, *45*, 772–782. doi:10.1044/1092-4388(2002/062).

IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE transactions on audio and electroacoustics*, *17*, 225–246.

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*, 347–360.

Johnson, E. E. (2013). Modern prescription theory and application: Realistic expectations for speech recognition with hearing aids. *Trends in amplification*, *17*, 143–170.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*, 944–946. doi:10.1111/2041-210X.12225.

Keidser, G., & Alamudi, K. (2013). Real-life efficacy and reliability of training a hearing aid. *Ear and Hearing*, *34*, 619–629. doi:10.1097/AUD.0b013e31828d269a.

Keidser, G., & Convery, E. (2016). Self-Fitting Hearing Aids: Status Quo and Future Predictions. *Trends in Hearing*, *20*, 2331216516643284. doi:10.1177/2331216516643284.

Keidser, G., Dillon, H., & Byrne, D. (1995). Candidates for multiple frequency response characteristics. *Ear and hearing*, *16*, 562–574. doi:10.1097/00003446-199512000-00003.

Keidser, G., Brew, C., & Peck, A. (2003). Proprietary fitting algorithms compared with one another and with generic formulas. *The Hearing Journal*, *56*, 28–32.

Keidser, G., Brew, C., Brewer, S., et al. (2005). The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input.

International Journal of Audiology, 44, 656–670. doi:10.1080/14992020500266803.

Keidser, G., Dillon, H., & Convery, E. (2008). The effect of the base line response on self-adjustments of hearing aid gain. *The Journal of the Acoustical Society of America*, *124*, 1668–1681. doi:10.1121/1.2951500.

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*, 621–635. doi:10.1080/14992020802178722.

Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, *1*, 88–90. doi:10.4081/audiores.2011.e24.

Keidser, G., Dillon, H., Carter, L., & O'Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends in Amplification*, *16*, 211–223. doi:10.1177/1084713812468511.

Killion, M. C. (1979). Equalization filter for eardrum-pressure recording using a KEMAR manikin. *Journal of the Audio Engineering Society*, *27*, 13–16.

Killion, M. C. (2004). Myths about hearing aid benefit and satisfaction. *Hearing Review*, *11*, 14–21.

Knudsen, L. V., Öberg, M., Nielsen, C., Naylor, G., & Kramer, S. E. (2010). Factors Influencing Help Seeking, Hearing Aid Uptake, Hearing Aid Use and Satisfaction With Hearing Aids: A Review of the Literature. *Trends in Amplification*, *14*, 127–154. doi:10.1177/1084713810385712.

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

Kochkin, S. (2009). MarkeTrak VIII: 25-year trends in the hearing health market. *Hearing review*, *16*, 12–31.

Kochkin, S. (2011). MarkeTrak VIII Patients report improved quality of life with hearing aid usage. *The Hearing Journal*, *64*, 25. doi:10.1097/01.HJ.0000399150.30374.45.

Kochkin, S. (2012). MarkeTrak VIII: The key influencing factors in hearing aid purchase intent. *Hearing Review*, *19*, 12–25.

Kuk, F. K., & Lau, C. (1995). The Application of Binomial Probability Theory to Paired Comparison Judgments. *American Journal of Audiology*, *4*, 37–42. doi:10.1044/1059-0889.0401.37.

Kuk, F. K., & Lau, C. (1996a). Comparison of preferred frequency gain settings obtained with category rating and modified simplex procedure. *Journal of the American Academy of Audiology*, *7*, 322.

Kuk, F. K., & Lau, C. (1996b). Effect of hearing aid experience on preferred insertion gain selection. *J Am Acad Audiol*, *7*, 274–281.

Kuk, F. K., & Pape, N. M. C. (1992). The reliability of a modified simplex procedure in hearing aid frequency-response selection. *Journal of Speech, Language, and Hearing Research*, *35*, 418–429. doi:10.1044/jshr.3502.418.

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, Language, and Hearing Research, *36*, 168–177. doi:10.1044/jshr.3601.168.

Kuk, F. K., Harper, T., & Doubek, K. (1994). Preferred real-ear insertion gain on a commercial hearing aid at different speech and noise levels. *Journal of the American Academy of Audiology*, *5*, 99–109.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. doi:10.18637/jss.v082.i13.

Kwon, B. J. (2012). AUX: A scripting language for auditory signal processing and software packages for psychoacoustic experiments and education. *Behavior research methods*, *44*, 361–373.

Mackersie, C., Boothroyd, A., & Lithgow, A. (2018). A "Goldilocks" Approach to Hearing Aid Self-Fitting: Ear-Canal Output and Speech Intelligibility Index. *Ear and Hearing*, *Publish Ahead of Print*, . doi:10.1097/AUD.0000000000000617.

Margolis, R. H., & Morgan, D. E. (2008). Automated Pure-Tone Audiometry: An Analysis of Capacity, Need, and Benefit. *American Journal of Audiology*, *17*, 109–113. doi:10.1044/1059-0889(2008/07-0047).

Marriage, J., Moore, B. C. J., & Alcántara, J. I. (2004). Comparison of three procedures for initial fitting of compression hearing aids. III. Inexperienced versus experienced users. *International Journal of Audiology*, *43*, 198–210. doi:10.1080/14992020400050028.

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *International Journal of Audiology*, *52*, 360–368. doi:10.3109/14992027.2013.769066.

McCreery, R. W., Venediktov, R. A., Coleman, J. J., & Leech, H. M. (2012). An evidence-based systematic review of directional microphones and digital noise reduction hearing aids in school-age children with hearing loss. *American Journal of Audiology*, *21*, 295–312. doi:10.1044/1059-0889(2012/12-0014).

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*, 216–227. doi:10.3109/14992020903296746.

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*, 556–568. doi:10.1097/AUD.0b013e31820b5f4c.

Mueller, H. G., Hornsby, B. W. Y., & Weber, J. E. (2008). Using trainable hearing aids to examine real-world preferred gain. *Journal of the American Academy of Audiology*, *19*, 758–773. doi:10.3766/jaaa.19.10.4.

Mulrow, C., Aguilar, C., Endicott, J., et al. (1990). Quality-of-Life Changes and Hearing Impairment: A Randomized Trial. *Annals of Internal Medicine*, *113*, 188–194. doi:10.1059/0003-4819-113-3-188.

Nabelek, A. K., Tampas, J. W., & Burchfield, S. B. (2004). Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *Journal of Speech, Language, and Hearing Research*, *47*, 1001–1011. doi:10.1044/1092-4388(2004/074).

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *Journal of the American Academy of Audiology*, *17*, 626–639. doi:10.3766/jaaa.17.9.2.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. doi:10.1111/j.2041-210x.2012.00261.x.

National Academies of Sciences, E. (2016). *Hearing health care for adults: Priorities for improving access and affordability*. National Academies Press,.

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*, 2331216518798264. doi:10.1177/2331216518798264.

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*, 1967–1976. doi:10.1121/1.395641.

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., & Ewert, S. D. (2016). Spectral and binaural loudness summation for hearing-impaired listeners. *Hearing Research*, *335*, 179–192. doi:10.1016/j.heares.2016.03.010.

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., & Ewert, S. D. (2018). Restoring Perceived Loudness for Listeners With Hearing Loss. *Ear and Hearing*, *39*, 664. doi:10.1097/AUD.0000000000000521.

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*, 21–25. doi:10.1044/1059-0889(1998/012).

Pavlovic, C. V. (1984). Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment. *The Journal of the Acoustical Society of America*, *75*, 1253–1258. doi:10.1121/1.390731.

Perry, T. T., Nelson, P. B., & Van Tasell, D. J. (2019). Listener factors explain little variability in self-adjusted hearing aid gain. *Trends in Hearing*, *23*, 2331216519837124. doi:10.1177/2331216519837124.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech, Language, and Hearing Research*, *29*, 146–154. doi:10.1044/jshr.2902.146.

Polonenko, M. J., Scollie, S. D., Moodie, S., et al. (2010). Fit to targets, preferred listening levels, and self-reported outcomes for the DSL v5.0a hearing aid prescription for adults. *International Journal of Audiology*, *49*, 550–560. doi:10.3109/14992021003713122.

Preminger, J. E., & Van Tasell, D. J. (1995). Quantifying the Relation Between Speech Quality and Speech Intelligibility. *Journal of Speech, Language, and Hearing Research*, *38*, 714–725. doi:10.1044/jshr.3803.714.

Preminger, J. E., Neuman, A. C., Bakke, M. H., Walters, D., & Levitt, H. (2000). An examination of the practicality of the simplex procedure. *Ear and Hearing*, *21*, 177. doi:10.1097/00003446-200006000-00001.

Punch, J. L., & Howard, M. T. (1978). Listener-assessed intelligibility of hearing aid-processed speech. *Journal of the American Auditory Society*, *4*, 69–76.

Punch, J. L., & Parker, C. A. (1981). Pairwise Listener Preferences in Hearing Aid Evaluation. *Journal of Speech, Language, and Hearing Research*, *24*, 366–374. doi:10.1044/jshr.2403.366.

Punch, J. L., Rakerd, B., & Amlani, A. M. (2001). Paired-comparison hearing aid preferences: evaluation of an unforced-choice paradigm. *Journal of the American Academy of Audiology*, *12*, 190–201.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing,.

Recker, K. L., & Edwards, B. W. (2013). The effect of presentation level on normal-hearing and hearing-impaired listeners' acceptable speech and noise levels. *Journal of the American Academy of Audiology*, *24*, 17–25. doi:10.3766/jaaa.24.1.3.

Rumalla, K., Karim, A. M., & Hullar, T. E. (2015). The effect of hearing aids on postural stability. *The Laryngoscope*, *125*, 720–723. doi:10.1002/lary.24974.

Rutherford, B. R., Brewster, K., Golub, J. S., Kim, A. H., & Roose, S. P. (2017). Sensation and Psychiatry: Linking Age-Related Hearing Loss to Late-Life

Depression and Cognitive Decline. *American Journal of Psychiatry*, *175*, 215–224. doi:10.1176/appi.ajp.2017.17040423.

Schweitzer, C., Mortz, M., & Vaughan, N. (1999). Perhaps not by prescription, but by perception. *High Performance Hearing Solutions*, *3,* 58–62.

Scollie, S., Seewald, R., Cornelisse, L., et al. (2005). The desired sensation level multistage input/output algorithm. *Trends in amplification*, *9*, 159–197.

Smeds, K. (2004). Is normal or less than normal overall loudness preferred by first-time hearing aid users? *Ear and Hearing*, *25*, 159–172.

Smeds, K., Keidser, G., Zakis, J., et al. (2006). Preferred overall loudness. II: Listening through hearing aids in field and laboratory tests. *International Journal of Audiology*, *45*, 12–25. doi:10.1080/14992020500190177.

Sorri, M., Luotonen, M., & Laitakari, K. (1984). Use and non-use of hearing aids. *British Journal of Audiology*, *18*, 169–172.

Souza, P. E., & Kitch, V. J. (2001). Effect of preferred volume setting on speech audibility in different hearing aid circuits. *Journal of the American Academy of Audiology*, *12*, 415–422.

Stelmachowicz, P. G., Lewis, D. E., & Carney, E. (1994). Preferred Hearing-Aid Frequency Responses in Simulated Listening Environments. *Journal of Speech, Language, and Hearing Research*, *37*, 712–718. doi:10.1044/jshr.3703.712.

Swanepoel, D. W., Clark, J. L., Koekemoer, D., et al. (2010). Telehealth in audiology: The need and potential to reach underserved communities. *International Journal of Audiology*, *49*, 195–202. doi:10.3109/14992020903470783.

Tampas, J. W., & Harkrider, A. W. (2006). Auditory evoked potentials in females with high and low acceptance of background noise when listening to speech. *The Journal of the Acoustical Society of America*, *119*, 1548–1561. doi:10.1121/1.2167147.

Upfold, G., & Byrne, D. (1988). Variability of ear canal resonance and its implications for the design of hearing aids and earplugs. *Australian J. Audiol*, *10*, 97–102.

Valente, Michael, Valente, Maureen, & Goebel, J. (1991). Reliability and intersubject variability of the real ear unaided response. *Ear and hearing*, *12*, 216–220. doi:10.1097/00003446-199106000-00009.

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, Language, and Hearing Research*, *38*, 211–221. doi:10.1044/jshr.3801.211.

Wong, L. L. N. (2011). Evidence on Self-Fitting Hearing Aids. *Trends in Amplification*, *15*, 215–225. doi:10.1177/1084713812444009.

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*, 812–830. doi:10.1097/AUD.0b013e3181576738.