

Research Article

Toward a New Evidence-Based Fitting Paradigm for Over-the-Counter Hearing Aids

Dana Urbanski,^a Helin Hernandez,^b Jacob Oleson,^b and Yu-Hsiang Wu^a

Purpose: This dual-aim study was intended to develop and validate a new fitting paradigm for over-the-counter (OTC) hearing aids. Aim 1 was undertaken to create a limited set of evidence-based preconfigured gain–frequency responses (“presets”) for use in OTC devices. Aim 2 tested the efficacy of the presets relative to best-practice verification and determined the best method(s) for older adults to self-select amplification from the set.

Method: In Aim 1, audiometry data from 267 older adults with mild-to-moderate sensorineural hearing loss were obtained from the National Health and Nutrition Examination Survey database. Using these data and the National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) prescriptive formula, a set of four presets were created that can fit a large percentage of older Americans with presbycusis. In Aim 2, 37 older adults with hearing loss used five methods to select presets. The selection methods were select-by-audiogram, select-by-

self-test, select-by-trying, select-by-questionnaire, and random assignment. Using a crossover design, each participant completed speech recognition testing and sound quality ratings in quiet and noise for all selection methods and a verified NAL-NL2 condition.

Results: The set of presets can fit 67.9% of older Americans with mild-to-moderate hearing loss (Aim 1). Controlling for hearing thresholds and sound quality ratings, liner mixed-effects models indicated that speech recognition scores for select-by-audiogram, select-by-self-test, and select-by-trying were not statistically different from the NAL-NL2 condition. Statistical analysis indicated that select-by-self-test produced outcomes most consistent with individual outcomes for the NAL-NL2 condition (Aim 2).

Conclusion: This newly developed fitting paradigm may provide efficacy comparable to best-practice verification and could be used in OTC hearing aids.

In recent years, untreated hearing loss has been recognized as a significant public health concern for older Americans (The President’s Council of Advisors on Science and Technology, Executive Office of the President, 2015). An estimated two thirds of Americans over the age of 70 years are impacted by hearing loss, yet hearing aid adoption rates remain low in this population (Goman & Lin, 2018). Among Americans over the age of 70 years with hearing impairment, only around 30% seek out and use hearing aids (Bainbridge & Ramachandran, 2014). Adoption rates are even lower for older adults of low socioeconomic status and for those of racial and ethnic minorities (Bainbridge & Ramachandran, 2014). These figures raise concern about a

large population of older adults living with negative effects of untreated hearing loss including depression (Nachtegaal et al., 2009), anxiety (Contrera et al., 2017), social isolation (Nachtegaal et al., 2009), accelerated cognitive decline (Deal et al., 2018; Lin et al., 2013), dementia (Gallacher et al., 2012; Lin et al., 2011), increased risk of falls (Jiam et al., 2016), elevated risk of hospital readmission (Chang et al., 2018), and overall higher health care expenditure (Foley et al., 2014).

In 2017, Congress passed the Over-the-Counter Hearing Aid Act to promote affordable, accessible hearing health care. Proponents of the legislation believe it will mitigate widespread adverse effects of untreated hearing loss by increasing hearing aid adoption among individuals unserved by the traditional service delivery model. This aim, however, requires that over-the-counter (OTC) hearing aids be capable of producing rehabilitative outcomes, namely, that they can approximate outcomes of traditional best-practice hearing aid fittings. In other words, it is not enough for OTC hearing aids to be affordable and accessible; they must support the provision of affordable, accessible, *quality* hearing health care.

^aDepartment of Communication Sciences and Disorders, The University of Iowa, Iowa City

^bDepartment of Biostatistics, The University of Iowa, Iowa City

Correspondence to Dana Urbanski: dana-urbanski@uiowa.edu

Editor-in-Chief: Ryan W. McCreery

Editor: Erin M. Picou

Received June 6, 2020

Revision received August 30, 2020

Accepted September 17, 2020

https://doi.org/10.1044/2020_AJA-20-00085

Disclosure: The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dana Urbanski is a current employee of Novidan, Inc.

Quality OTC outcomes will hinge on the development of effective direct-to-consumer fitting methods. Although some research supports the effectiveness of a direct-to-consumer service delivery model (Humes et al., 2017), there are currently limited data to guide creation of an evidence-based OTC fitting paradigm. Innovation of effective, quality OTC fitting methods will depend on at least two factors: (a) whether noncustom fittings can provide sufficient benefit relative to best-practice clinical verification methods and (b) whether older adults are able to self-select appropriate amplification. The present investigation is a two-part study designed to examine these questions through development and laboratory testing of a new OTC fitting paradigm. The study is composed of two aims. The purpose of the first aim was to optimize noncustom amplification by developing a limited set of evidence-based preconfigured gain–frequency responses that can fit a large percentage of older Americans with mild-to-moderate age-related hearing loss. The second aim was undertaken to test the efficacy of the newly developed set of gain–frequency responses relative to best-practice verification and determine the best method(s) for older adults to self-select appropriate amplification from the set.

Aim 1

Introduction

In order to produce quality outcomes, OTC hearing aids must have appropriate gain characteristics for the OTC target population—namely, older adults with mild-to-moderate age-related hearing loss. Inappropriate gain could result in poor audibility and, by extension, suboptimal outcomes. Research on the electroacoustic properties and gain–frequency responses of currently available personal sound amplification products (PSAPs) and OTC devices suggest cause for concern. Results from previous studies indicate that many OTC devices produce excessively high levels of low-frequency gain with little to no usable high-frequency gain, rendering them most appropriate for rising hearing losses (Chan & McPherson, 2015; Cheng & McPherson, 2000). This pattern of gain is wholly inappropriate for the treatment of presbycusis, which is most commonly characterized as a sloping high-frequency hearing loss. Such devices could produce poor outcomes by failing to provide adequate gain for high frequencies where hearing loss is present.

While many OTC hearing aids are inappropriate for presbycusis, recent research suggests some direct-to-consumer devices may be a viable solution for older adults with mild-to-moderate hearing loss. Reed et al. examined the gain–frequency responses of a range of currently available direct-to-consumer devices (Reed et al., 2017). The authors found that, for a common configuration of mild-to-moderate presbycusis, nine of 10 devices tested met target within ± 10 dB across at least five of nine prescriptive targets from 250 to 6000 Hz. Only three devices, however, met at least five prescriptive targets within a stricter criterion

of ± 5 dB (Reed et al., 2017). Another study found that some high-end direct-to-consumer devices can be appropriately fit to meet National Acoustic Laboratories–Non-Linear 2 (NAL-NL2; Keidser et al., 2011) prescriptive targets for up to a moderate hearing loss (Smith et al., 2016). Broadly, these results demonstrate that OTC amplification can provide suitable gain for older adults with mild-to-moderate hearing loss.

Since OTC devices are distributed direct to consumer, their outcomes depend on the provision of effective self-fitting strategies. Previous research in the area of self-fitting has focused primarily on the development, validation, and refinement of self-configuring methods for fitting direct-to-consumer hearing aids (Boothroyd & Mackersee, 2017; Convery et al., 2011, 2017). Self-configuring methods allow the end user to program and fine-tune the gain–frequency response, generally through use of tablets, smartphones, or other computer interfaces. In some cases, the self-fitting process begins with in situ audiometry to facilitate programming from scratch. In other cases, users fine-tune preconfigured gain–frequency responses designed to provide an appropriate starting point from which to make adjustments (Edwards, 2020).

Although data indicate that, in general, older adults can adjust self-configuring OTC hearing aids to compensate for their hearing loss (Brody et al., 2018), these devices are not a feasible solution for adults of low socioeconomic status and/or those who do not own a smartphone or tablet. Recall that low-income older adults are less likely to pursue traditional hearing aids. Evidence also suggests they may be less likely to own smartphones or tablets necessary for self-adjustment of hearing aids. Indeed, a recent study indicates that only 16%–27% of older adults with lower incomes (< \$30,000 per year) own a smartphone or tablet (Pew Research Center, 2017). Self-configuring hearing aids, therefore, are suboptimal for the realistic OTC target demographic. Preconfigured devices may be an effective alternative. A recent clinical trial found that a limited set of preconfigured, noncustomizable OTC devices was efficacious for older adults with mild-to-moderate sensorineural hearing loss (Humes et al., 2017).

The purpose of Aim 1 was to optimize noncustom amplification by developing quality evidence-based preconfigured gain–frequency responses for use in OTC hearing aids. Since audiometric thresholds vary widely between mild hearing losses (thresholds 26–40 dB HL) and moderate hearing losses (thresholds 41–55 dB HL), it is unlikely that only one gain–frequency response can provide optimal outcomes for every adult in this category of hearing loss. Accordingly, this study was designed to develop multiple gain options while also keeping the number of choices manageable for older adult consumers. Using audiometric data from a large-scale national health study and a rigorous methodology, Aim 1 was undertaken to generate a limited set of evidence-based preconfigured gain–frequency responses capable of fitting a large percentage of older Americans with mild-to-moderate age-related hearing loss.

Method

Study Population

The National Health and Nutrition Examination Survey (NHANES) audiometry database was used to gather data on the various configurations of mild-to-moderate hearing loss in American older adults. NHANES is a rich source of audiometry data, as the data are collected using methodology that allows for generalization to the noninstitutionalized U.S. civilian population. The NHANES survey is conducted every 2 years on noninstitutionalized U.S. residents to collect data on key markers of overall health status. NHANES participants are selected using an intricate multistage probability using many levels of stratification and sample groups.

The NHANES audiometry data from 1999 to 2012 were filtered for adults aged 55 years and over with mild-to-moderate sensorineural hearing loss bilaterally. Unilateral hearing losses were excluded because unilateral hearing loss suggests underlying etiology that may be inappropriate for OTC amplification. Mild-to-moderate hearing loss was defined as pure-tone average (PTA) at 500, 1000, and 2000 Hz of ≥ 25 dB HL and ≤ 55 dB HL with no threshold poorer than 65 dB HL from 500 to 6000 Hz (Clark, 1981). All selected individuals had tympanograms within normal limits bilaterally, defined as ear canal volume from 0.5 to 1.0 ml, peak pressure from -100 to $+50$ daPa, and peak compliance from 0.25 to 1.5 ml. For each selected individual, audiometric thresholds from 500 to 6000 Hz were obtained for each ear. The NHANES test protocol did not include testing at 250 Hz. In order to include 250 Hz in subsequent analyses, the 500-Hz threshold for each ear was copied at 250 Hz. Thus, each selected individual contributed audiometric thresholds from 250 to 6000 Hz bilaterally.

The NHANES audiometry data yielded 267 individuals (68% female, 32% male) and 534 ears that met the above inclusion criteria. Included participants had a mean age of 70.1 years (range: 55–85 years, $SD = 8.3$). Figure 1A shows the averaged audiogram of included ears.

Next, threshold data were converted to their corresponding NAL-NL2 real-ear aided response (REAR) prescriptive targets for a 65-dB SPL speech signal. This step was designed to gather a large sample of gain–frequency responses known to fit a generalizable sample of mild-to-moderate age-related hearing losses. The NAL-NL2 fitting rationale was chosen because it is a validated prescriptive formula widely used for best-practice clinical verification of hearing aids (Keidser et al., 2012, 2011).

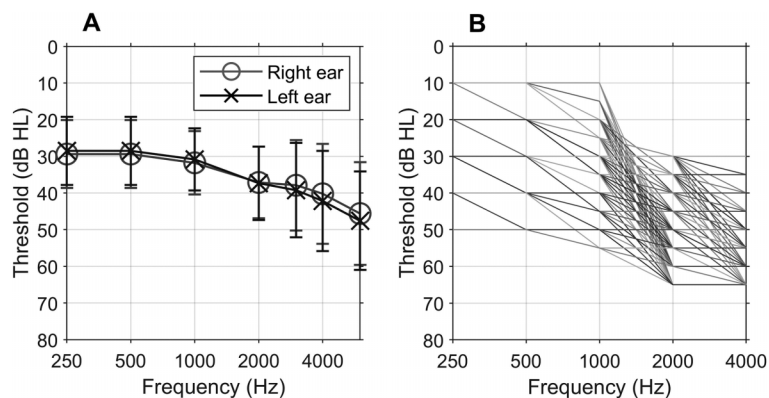
For each NHANES individual, NAL-NL2 REAR prescriptive targets were obtained for the octave frequencies from 250 to 4000 Hz (250, 500, 1000, 2000, and 4000 Hz) for each of three fitting conditions: (a) unilateral right ear fitting, (b) unilateral left ear fitting, and (c) bilateral fitting. The targets were obtained using an Audioscan (Dorchester) Verifit 1 hearing aid analyzer with age set to adult, transducer set to headphone, nontonal language, and average uncomfortable level and real-ear-to-coupler difference. Regardless of fitting condition, the left and right audiometric thresholds from 250 to 6000 Hz were always entered into the Verifit system before obtaining REAR target values. This procedure resulted in four gain–frequency responses per NHANES individual—one for each unilateral fitting condition and two for the bilateral fit condition.

Creation of Candidate Gain–Frequency Responses

As a first step in developing a limited set of preconfigured gain–frequency responses for hearing aids, a large sample of possible—or candidate—gain–frequency responses was created. These candidate gain–frequency responses represented possible preconfigured gain–frequency responses that could be implemented in an OTC hearing aid.

First, a set of audiograms (referred to as base audiograms in the present article) was created to represent the range of possible degrees and configurations of mild-to-moderate hearing loss. The base audiograms were created by placing nodes on the audiogram to represent meaningful increases in threshold within the range of mild-to-moderate

Figure 1. (A) Mean hearing thresholds for audiograms obtained from the National Health and Nutrition Examination Survey database in Aim 1. Error bars indicate 1 SD . (B) Six hundred forty-two base audiograms used to create candidate gain–frequency responses.



hearing loss. This method is similar to procedures previously described in the literature (Bisgaard et al., 2010; Dillon et al., 2014).

Nodes were placed on the audiogram at each octave frequency from 250 to 4000 Hz. At 250 and 500 Hz, nodes were placed at 10-dB HL intervals from 10 to 50 dB HL, representing the range of thresholds of mild-to-moderate hearing loss at these two frequencies. At 1000 Hz, nodes were placed at 5-dB HL intervals from 10 to 50 dB HL. At 2000 and 4000 Hz, nodes were placed at 5-dB HL intervals from 30 to 65 dB HL. The threshold ranges selected for each frequency were consistent with the hearing loss criteria used to filter the NHANES audiometry database. Note that nodes were placed farther apart at 250 and 500 Hz because the NAL-NL2 prescriptive formula assigns a narrower range of REAR target values to low frequencies. In other words, for low frequencies, larger changes in threshold result in smaller changes in prescriptive target value than they do from 1000 to 4000 Hz.

After the nodes were placed, they were then connected in all possible combinations using the method described by Dillon et al. (2014). This procedure generated 642 base audiograms. The resulting base audiograms are shown in Figure 1B. Each base audiogram was then converted to its corresponding NAL-NL2 REAR targets using the Audioscan Verifit 1. This step was undertaken to create a sample of possible gain–frequency responses that might be implemented in OTC hearing aids. For each base audiogram, unilateral NAL-NL2 REAR targets were obtained for the octave frequencies from 250 to 4000 Hz for a 65-dB SPL male talker in the same manner used to obtain REAR targets for the NHANES individuals. Note that the base audiogram thresholds were entered for both the right and left ears prior to recording the unilateral REAR targets. This process resulted in the creation of 642 candidate gain–frequency responses, each of which represented a possible preconfigured gain–frequency response for use in an OTC device.

Development of Preconfigured OTC Gain–Frequency Responses

A MATLAB (MathWorks) algorithm was created to select a set of gain–frequency responses—termed *presets*—from the pool of 642 candidate gain–frequency responses. The purpose of this step was to select a limited set of presets that can fit a large percentage of the NHANES individuals and, by extension, American older adults.

The algorithm first determined every possible combination of four gain–frequency responses from the set of 642 candidate gain–frequency responses. The number four was chosen because it was believed that this number of presets would be manageable for older adults while also providing enough flexibility to fit a wide range of mild-to-moderate hearing losses. For each combination of four gain–frequency responses, the algorithm calculated an overall fit percentage (described below).

Recall that the NHANES data are generalizable to the noninstitutionalized U.S. civilian population. This is realized through use of sample weights assigned to each

person in the NHANES data set. When applied to each individual, the sample weight is a multiplier designed to account for oversampling, survey nonresponse, and post-stratification. To ensure the present NHANES sample was in fact generalizable to the U.S. population, each individual was multiplied by the appropriate sample weight before subsequent analyses. The fit percentage, therefore, was calculated as the sample-weight adjusted percentage of NHANES individuals who could be appropriately fit by at least one preset of the four in each of the fitting scenarios previously mentioned: (a) unilateral right fitting, (b) unilateral left fitting, and (c) bilateral fitting.

For each NHANES subject, an appropriate fit was defined as a gain–frequency response that approximated the individual’s NAL-NL2 REAR target within a ± 5 -dB criterion at all octave frequencies from 250 to 4000 Hz in a given fitting condition. This criterion was selected because it represented a strict match to validated prescriptive targets on par with best-practice clinical verification, which has been shown to match prescriptive targets within a tolerance range of approximately ± 5 dB regardless of manufacturer or number of gain handles (Baker & Jenstad, 2017; Polonenko et al., 2010).

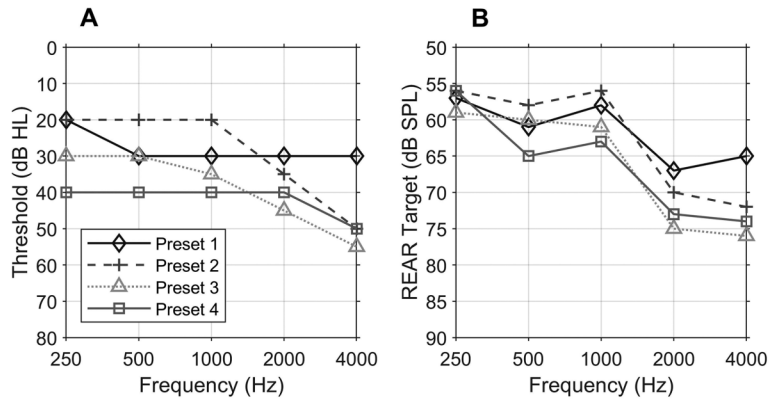
Results

The set of four presets with the highest overall fit percentage was selected as the winning set. The winning set fit 67.9% of the NHANES sample in all three fitting conditions (see Appendix A for distribution of fit percentages). Figure 2A shows the base audiograms for the winning set, and Figure 2B shows their corresponding NAL-NL2 REAR targets. The presets were numbered roughly in order of the amount of gain prescribed by each setting from the lowest gain (Preset 1) to the highest gain (Preset 4). Audiograms not appropriately fit by the set generally fell into one of three categories: (a) low-frequency rising audiograms, (b) notched audiograms characterized by a sharp dip at any frequency followed by sharp recovery, and (c) steeply sloping audiograms. See Appendixes B and C for the audiograms that could be fit by each preset and the audiograms that could not be appropriately fit, respectively.

Discussion

The results demonstrate that a set of four gain–frequency responses can fit approximately 70% of older adults with mild-to-moderate hearing loss in the audiometric range defined by this study (PTA at 500, 1000, and 2000 Hz of ≥ 25 and ≤ 55 dB HL with no threshold poorer than 65 dB HL from 500 to 6000 Hz). This fit percentage holds regardless of whether the individual selects a unilateral or bilateral fitting. Broadly, this result validates the feasibility of using a limited set of preconfigured gain–frequency responses to fit a large percentage of older Americans with mild-to-moderate presbycusis. Specifically, the gain–frequency responses developed in this study may be used in future OTC fitting paradigms to support the provision of quality, evidence-based direct-to-consumer amplification.

Figure 2. (A) Base audiograms of the winning set of four presets. (B) National Acoustic Laboratories–Non-Linear 2 real-ear aided response (REAR) targets of the winning set of four presets.



The resulting set of four presets is particularly promising because it can fit most older adults in the study-defined audiometric range within a stringent ± 5 -dB fit criterion. Previous research has often employed a ± 10 -dB criterion for evaluating match to target (Aazh & Moore, 2007; Jenstad et al., 2007; Smith et al., 2016). When a 10-dB tolerance range is used, the percentage of PSAPs and OTC devices classified as appropriate for a given configuration of mild-to-moderate hearing loss is significantly increased relative to the use of a 5-dB criterion (Reed et al., 2017; Smith et al., 2016). Employing a more relaxed fit standard, however, may increase the likelihood of suboptimal OTC outcomes. If direct-to-consumer devices are intended to produce outcomes comparable to traditional hearing aids, they should be evaluated relative to fit standards representative of clinical best-practice verification. Evidence shows traditional hearing aids custom-programmed to match REAR targets consistently approximate target values within ± 5 dB across the frequency range (Baker & Jenstad, 2017; Polonenko et al., 2010).

The resulting gain–frequency responses give important flexibility for different fitting configurations. Indeed, the winning set can fit approximately 70% of older adults in the study-defined audiometric range within a strict fit criterion regardless of whether the user selects one or two hearing aids. This flexibility is particularly important for the OTC demographic. Evidence suggests OTC users may elect to purchase only one hearing aid, even if they have bilateral hearing loss. A study by Kochkin found that consumers of direct-mail amplification were more likely to purchase one hearing aid (45%) relative to traditional hearing aid users (22%; Kochkin, 2014). Unilateral and bilateral fittings, however, require different gain–frequency responses. To account for binaural loudness summation, the NAL-NL2 prescriptive formula selectively reduces gain for bilateral fittings. Conversely, unilateral fittings are prescribed slightly more gain to account for the lack of summation (Keidser et al., 2012). If OTC hearing aid users opt to purchase only one hearing aid, it is imperative that OTC devices can provide evidence-based gain for both unilateral and bilateral

fittings. The set of gain–frequency responses developed in this study allows consumers to select their preferred fitting configuration without sacrificing fit quality.

While most older Americans with mild-to-moderate presbycusis can be appropriately fit by the resulting set of presets, a small percentage cannot. As shown in Appendix C, most audiograms that could not be fit fell into one of three categories: (a) rising, (b) notched, and (c) steeply sloping. Rising hearing loss constitutes a rare configuration for presbycusis, and notched audiograms are suggestive of noise damage, in contrast to age-related hearing loss, which generally presents as a sloping high-frequency hearing loss (Demeester et al., 2009; Krishnamurti, 2009). Thus, these configurations are not part of the hearing loss profile intended to be fit by OTC amplification. Steeply sloping audiograms, on the other hand, are observed in about a quarter of individuals with age-related hearing loss (Demeester et al., 2009). The resulting gain–frequency responses may not have fit steeply sloping audiograms because this configuration of hearing loss constituted only a small number of audiograms in the NHANES data set. This may have been the result of the filtering criteria, which stipulated that no threshold could be poorer than 65 dB HL from 500 to 6000 Hz. This criterion excluded many steeply sloping audiograms with more severe high-frequency hearing loss, leaving only a small number of steeply sloping losses eligible for inclusion in the NHANES sample. Future research should consider the extent to which steeply sloping losses can be appropriately fit using preconfigured gain–frequency responses. Relatedly, future work might consider broadening the definition of mild-to-moderate hearing loss employed by this study to develop an even more comprehensive set of preconfigured gain–frequency responses for OTC amplification.

Aim 2

Introduction

OTC hearing aids are distributed using direct-to-consumer service delivery models. In the traditional

clinic-based model, a licensed dispenser customizes the gain–frequency response to provide optimal audibility across frequencies, while maintaining appropriate loudness and sound quality. By contrast, in existing OTC fitting paradigms, older adults take on the responsibility of selecting an appropriate gain–frequency response. If the presets developed in Aim 1 are implemented to, say, a family of preconfigured OTC devices, older adults would have to choose which of the four options to purchase. One could imagine that even a set of effective, evidence-based preconfigured gain–frequency responses could produce poor outcomes if older adults systematically selected inappropriate presets for their hearing loss. The purpose of Aim 2 was to identify a method by which older adults will be able to select appropriate amplification from the set of four preconfigured gain–frequency responses developed in Aim 1.

Some evidence supports the efficacy of an OTC fitting paradigm based on self-selection of preconfigured amplification. Humes et al. conducted a placebo-controlled randomized trial designed to compare outcomes for audiology best-practice (AB) and a consumer-decides (CD) service delivery model. The CD arm was designed so that older adults selected their preferred preconfigured gain–frequency response from a set of responses programmed to fit three common configurations of mild-to-moderate presbycusis. Though the participants had access to overall gain adjustments (analogous to a volume control), they conducted no further self-configuration or fine-tuning of the preconfigured gain–frequency responses. The CD and AB service delivery models were both found to be efficacious relative to the placebo group (Humes et al., 2017).

The results from the Humes et al. (2017) study suggest the feasibility of a direct-to-consumer service delivery model in which older adults self-select preconfigured, non-customizable gain–frequency responses. However, Humes et al. only investigated one type of selection model in which users selected their preferred gain–frequency response by listening to the various choices. Though the results showed the CD model to be efficacious, it is unclear if this selection method is the optimal way for older adults to self-select amplification. Given the importance of effective self-selection for this type of service delivery model, it is essential to investigate the outcomes of a variety of plausible selection methods.

In this study, older adults selected amplification from the set of four presets using a variety of selection methods in the laboratory. The study was designed to achieve two interrelated aims: (a) test the efficacy of the set of gain–frequency responses developed in the prior experiment compared to best-practice verification and (b) identify the best way for older adults to select amplification from the set of presets.

Methods

Overview of Design

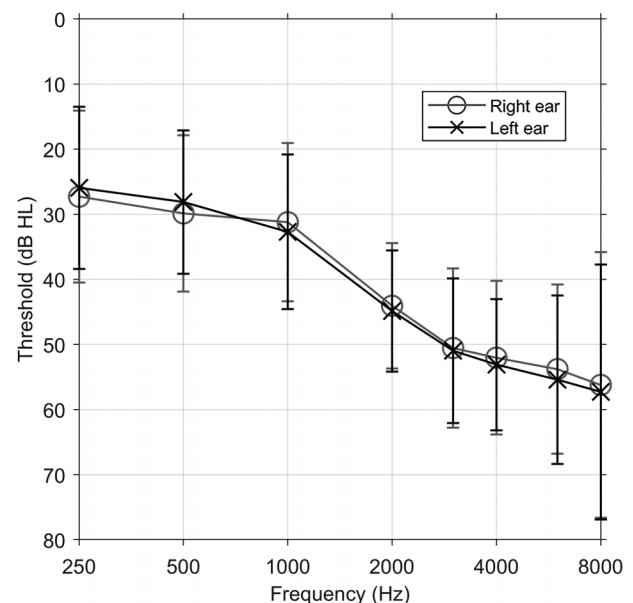
In this study, a group of older adults with bilateral mild-to-moderate sensorineural hearing loss used five selection

models to choose from the set of four preset gain–frequency responses, which were stored in the program memories of basic-level hearing aids. The selection models were as follows: (a) select-by-audiogram, in which presets were assigned using audiometric thresholds; (b) select-by-self-test, in which presets were assigned using predicted thresholds from an online self-hearing test; (c) select-by-trying, in which participants selected presets by listening to them; (d) select-by-questionnaire, in which presets were assigned using the results of a self-assessment; and (e) random assignment. Additionally, participants were fit with the same devices custom-programmed to match their NAL-NL2 REAR targets. Using a crossover design, each participant completed speech recognition testing and sound quality ratings in quiet and noise for all selection methods and the custom-programmed NAL-NL2 devices. Outcomes by selection model were analyzed to evaluate whether the four presets produced outcomes comparable to clinically fit hearing aids and to determine the best method for older adults to select amplification.

Study Participants

Thirty-seven adults (20 men and 17 women) were recruited from the community and completed the study. Participants were eligible for inclusion if they met the following criteria: (a) bilateral mild-to-moderate sensorineural hearing loss defined as PTA at 500, 1000, and 2000 Hz of ≥ 25 and ≤ 55 dB HL; (b) no threshold poorer than 65 dB HL up to 6000 Hz bilaterally; and (c) the ability to understand study directions and perform experiment-related tasks. Figure 3 shows the mean pure-tone thresholds of study participants. Participants ranged in age from 55 to 88 years, with a mean age of 68.5 years ($SD = 8.2$). All participants were screened

Figure 3. Mean hearing thresholds for the participants in Aim 2. Error bars indicate 1 *SD*.



for cognitive impairment using the Montreal Cognitive Assessment (score: $M = 27.4$, $SD = 1.8$; score range: 25–30; Nasreddine et al., 2005). Thirteen participants were experienced hearing aid users defined by any hearing aid use for at least 6 months prior to the beginning of the study; 24 were new users.

Hearing Aids and Programming

The study devices were basic-level behind-the-ear hearing aids with slim tubes and dome tips. These devices were selected because, at the time of the study, they were available for purchase through a direct-to-consumer hearing aid distributor. Additionally, the slim-tube behind-the-ear style device is a popular form factor of many currently available PSAPs and OTC devices, as it is relatively durable and easy to maintain.

The hearing aids were equipped with wide dynamic range compression (WDRC; eight channels) and four program memories. They contained several features including adaptive directional microphones, digital noise reduction algorithms, and impulse sound reduction; however, for the purpose of this study, these features were turned off. Hearing aids were programmed to use only the omnidirectional microphone mode, and noise reduction and feedback cancellation were set to the weakest setting in the programming software. The volume control and push button functions were disabled. Prior to programming, all hearing aids were tested in the test box of a Verifit 1 hearing aid analyzer to ensure they met all manufacturer specifications.

A pair of study hearing aids was programmed using the manufacturer fitting software such that the program memories contained the four gain–frequency responses developed in the prior experiment. The program memories were configured so that Program 1 (P1) contained Preset 1, P2 contained Preset 2, and so on (see Figure 2B for the preset REAR target values). Recall that the presets were numbered roughly in order of their gain from the lowest gain (P1) to the highest gain (P4). The presets were verified by REAR measurements using a 65-dB SPL male talker in the ear of a Knowles Electronics Manikin for Acoustic Research with G.R.A.S. RA0045 ear simulators. The hearing aids were coupled using a size-small slim tube and a tulip dome. Tulip domes were selected because they represent a one-size-fits-all coupling option that also provides enough occlusion to ensure the integrity of the programmed gain–frequency response.

During programming, the specified preset REAR value was matched exactly whenever possible; however, up to ± 3 -dB deviation was tolerated. Recall that the Aim 1 preset development rationale only considered 65-dB SPL NAL-NL2 REAR targets. As a result, during programming, verification was only conducted for 65-dB SPL speech; no verification was conducted for soft and loud speech signals (e.g., 55 and 75 dB SPL). Rather, the manufacturer fitting software was allowed to set the WDRC parameters based on the entered preset base audiogram. In order to match the specified preset REAR value, only overall gain adjustments were applied in the fitting software.

Following programming, of the 20 total REAR values specified by the four presets, 90% were matched exactly or deviated by only 1 dB. Remote controls were used to change programs (i.e., presets) and configure the hearing aids to the desired gain–frequency response. Each hearing aid was individually paired to its own remote control so its gain–frequency response could be controlled independent of the other device.

Another pair of study hearing aids was custom programmed for each participant's audiometric thresholds using clinical verification methods. This condition was designed to represent an audiology best practice model, in which the clinician makes programming adjustments to match REAR targets specified by a validated prescriptive formula (NAL-NL2). Prior to programming, the hearing aids were configured with the participant's appropriate slim tube size and a tulip dome tip.

Using a Verifit 1 hearing aid analyzer and the manufacturer fitting software, the researcher adjusted overall gain of each gain handle to match the participant's NAL-NL2 REAR targets for a 65-dB SPL male talker. Following the same procedure as the preset programming, compression parameters were determined by the fitting software based on the participant's audiometric thresholds, and only overall gain adjustments were applied to match 65-dB SPL targets. For the five octave REAR targets from 250 to 4000 Hz (250, 500, 1000, 2000, and 4000 Hz), the measured REAR was required to match the prescribed value within a ± 3 -dB criterion. For interoctave frequencies from 250 to 4000 Hz and frequencies above 4000 Hz, the researcher matched the prescribed REAR target within a ± 5 -dB criterion. For all participants, all measured REAR values at the octave frequencies from 250 to 4000 Hz met the ± 3 -dB criterion. Occasionally, one of the remaining frequencies deviated by more than 5 dB from the prescribed REAR value, but this was unusual. Following programming, no fine-tuning was completed regardless of participants' comments or feedback.

Selection Models

Study participants selected or were assigned gain–frequency responses from the set of four presets in each of five different selection models. In each selection model, participants were fit with bilateral amplification. The selection models are described below.

Select-by-audiogram. In this selection model, the participant's audiogram was used to select presets. This selection model was designed to simulate the scenario in which an older adult obtains an audiogram in a traditional clinical setting and uses the results to purchase direct-to-consumer amplification elsewhere. In this model, a preset was assigned to each ear individually based on the deviation between the participant's audiogram and the base audiogram used to generate each preset gain–frequency response (see Figure 2A for the corresponding base audiograms). The preset with corresponding base audiogram thresholds within ± 5 dB of the participant's audiometric thresholds at all octave frequencies from 250 to 4000 Hz (250, 500, 1000, 2000, and 4000 Hz) was assigned. If more than one preset or no preset met this

criterion, the preset with the lowest overall summed absolute value deviation across frequencies was selected. Following preset assignment, each participant was assigned one preset per ear. Note that presets could be the same or different across ears depending on the participant's audiometric thresholds, as assignments were made for each ear independent of the other.

Select-by-self-test. In recent years, several self-administered, automated hearing tests have been developed. These tests are typically delivered to individuals via landline or cellular telephones, the Internet, or handheld consumer electronic devices such as smartphones and tablets. It has been suggested that the results of self-administered hearing tests can be used to determine candidacy for OTC hearing aids (Garrison & Bochner, 2017). In the select-by-self-test selection model, presets were assigned using the results of a self-administered automated online speech-based hearing test—the *National Technical Institute for the Deaf Speech Recognition Test (NSRT)*—developed by researchers at the National Technical Institute for the Deaf (Bochner et al., 2015). The NSRT was chosen because this test estimates thresholds at different frequencies, generating a “pseudo audiogram” that gives predicted thresholds at 500, 1000, 2000, 4000, and 8000 Hz. The data suggest that, when NSRT-predicted hearing thresholds are compared with clinical pure-tone audiometry, the sensitivity and specificity of the NSRT is 95% and 87%, respectively, and the diagnostic accuracy is 91% (Bochner et al., 2015; Garrison & Bochner, 2015).

Participants conducted the NSRT in quiet on a desktop computer with Apple EarPods earphones in the laboratory. Each ear was tested individually, and the nontest ear was occluded using an earplug. The order in which ears were tested was randomized across participants. For each participant and each ear, a separate NSRT user account was created to ensure results were not influenced by prior testing, either within or across participants. Prior to speech testing, participants answered the following NSRT pretest survey questions: age, gender, “Do you regularly have difficulty understanding speech?”, and “Do you currently use a hearing aid?” After completing the pretest survey questions, participants self-administered the test without assistance.

Following testing, the pseudo-audiogram thresholds from 500 to 8000 Hz were recorded for each ear. For each ear, the threshold at 500 Hz was duplicated at 250 Hz and the 8000-Hz threshold was dropped. Thus, each ear was assigned pseudo-audiogram thresholds at 250, 500, 1000, 2000, and 4000 Hz. These predicted thresholds were used to assign the best-fit preset in the same manner as the select-by-audiogram selection model. Each participant was assigned one preset per ear, which could be the same or different across ears depending on the NSRT results for each ear.

Select-by-trying. In the select-by-trying selection model, participants selected their preferred presets by listening to them in quiet and noise in a sound booth. This model was designed to simulate future OTC fitting paradigms that may utilize a kiosk or listening station to facilitate self-selection.

Recall that the four presets were stored in the program memories of a pair of study devices, which were independently paired to separate remote controls that were used to change programs. Recall, too, that the program memories were organized such that P1 contained the first preset, P2 contained the second preset, and so forth, and the presets/programs were roughly ordered from the lowest gain (Preset 1) to the highest gain (Preset 4). Program numbers (and thus preset numbers) were indicated by a corresponding number of beeps upon changing programs. For all participants, study hearing aids were configured with the participant's appropriate slim tube size, as measured by the researcher, along with a one-size-fits-all tulip dome. The participant's appropriate slim tube size was used so that his or her self-selections would reflect the efficacy of the selection model and the gain–frequency response options rather than issues related to device fit and comfort.

Participants were seated in the center of a sound booth with all speech materials presented at 0° azimuth. First, the participant was tasked with selecting their preferred presets in quiet. The researcher gave the participant the left and right remote controls and inserted the study hearing aids. This was done to ensure proper placement of the devices, as the purpose of this selection model was to examine the efficacy of self-selection by trying rather than participants' ability to handle the devices.

Participants then listened to concatenated Connected Speech Test (CST; Cox et al., 1987) sentences in quiet presented at 55 dB SPL at 0° azimuth. This presentation level was selected because data show that, in the real world, speech levels in quiet listening environments are approximately 55 dBA (Pearsons et al., 1977; Wu et al., 2018). While listening in quiet, participants were instructed to use the remote controls to sample the four presets and select their preferred setting for each ear. They were advised to use the CST speech material to aid their own selection process, but they were not asked to repeat sentences or otherwise complete speech recognition testing. Since the procedure was designed to elicit older adults' independent selection process and criteria, participants were not given further guidance or strategy for selecting presets. Note that participants were free to choose any preset for each ear; they could choose the same preset across ears or could select different presets. Participants were given as much time as needed to make their selection. When the participant arrived at a final selection, he or she informed the researcher of the numbered preset chosen for each ear.

After participants chose their preferred presets in quiet, they were asked to listen to their selection in noise. This step was undertaken to ensure that selections made in quiet were satisfactory in noise. Participants remained seated in the center of the sound booth and listened to concatenated CST sentences in four-talker babble noise at +5 SNR (speech = 65 dB SPL) at 0° azimuth. The presentation levels of the speech and noise were selected to represent a plausible real-world listening situation based on data collected from recordings of older adults' everyday listening environments (Wu et al., 2018).

If, after listening in noise, participants were pleased with their selection, the self-selection procedure was complete, and the researcher recorded the final preset selection. If participants were dissatisfied with their prior selection, they used the remote controls to find a preferred setting in noise in the same manner used to make their selection in quiet.

If the participant made different selections in quiet versus noise, he or she listened to the noise selection in quiet to ensure it remained suitable for a quiet environment. Most of the time, participants were satisfied at this point and their preset selections were finalized. In the unusual case where a participant was still dissatisfied, he or she was asked to continue using the remote controls to choose presets, alternating between listening in quiet and noise until he or she converged on a final selection. For all participants, the researcher recorded the final preset selection for each ear along with the time taken to complete the select-by-trying procedure. Participants were not informed that the select-by-trying procedure was timed.

Select-by-questionnaire. In the select-by-questionnaire selection model, presets were assigned using the results of a validated self-assessment. This selection model was designed to probe whether older adults can select appropriate amplification based on their self-perceived hearing handicap rather than any type of objective hearing assessment. If this were feasible, it could promote accessibility of hearing health care by obviating the need for a traditional audiogram and circumventing challenges associated with developing reliable, valid self-administered hearing tests.

Participants completed the Better Hearing Institute (BHI) Quick Hearing Check (QHC) questionnaire (Kochkin & Bentler, 2010). The BHI QHC is a 15-item, 5-point, Likert-scaled questionnaire that asks respondents to rate their agreement with statements of common hearing loss problems such as “I have trouble hearing things on TV” and “I have to strain to understand conversations.” It was selected because it has been shown to have a high correlation with audiometric thresholds (Kochkin & Bentler, 2010). As such, for every raw score, the BHI QHC gives a predicted five-frequency PTA (5PTA) for both ears (500, 1000, 2000, 3000, and 4000 Hz). In order to translate the BHI QHC-predicted 5PTA to a preset assignment, each preset was assigned a 5PTA based on its corresponding base audiogram thresholds (see Figure 2A). As the base audiogram did not include a 3000-Hz threshold, the average of the 2000- and 4000-Hz thresholds was used in its place. The full range of BHI QHC scores (1–36) was then subdivided into score ranges corresponding to each of the four presets. The score range was divided so that, for each preset, the BHI QHC 5PTA approximated the preset’s 5PTA. Raw scores from 1 to 9 were assigned to Preset 1, scores of 10–18 were assigned to Preset 2, scores of 19–27 were assigned to Preset 3, and scores of 28–36 were assigned to Preset 4.

For each participant, the BHI QHC raw score was calculated and a preset value was assigned according to the above criteria. Note that the resulting preset was assigned

to both ears, as the BHI QHC questionnaire does not make ear-specific predictions.

Select-by-random-assignment. In this selection model, each participant was randomly assigned one of the four presets to both ears. This selection model was designed to simulate the possibility that, at least in some cases, family members or older adults with hearing loss may select non-custom amplification without any particular rationale.

Each participant was randomly assigned a number 1–4 corresponding to the four presets. The assigned preset was used in both ears, as individuals who randomly select OTC devices are unlikely to select different settings for each ear.

Outcome Measures

Speech Intelligibility Index. The Speech Intelligibility Index (SII; American National Standards Institute, 2012) was used to quantify aided audibility. Following assignment of presets and fitting of NAL-NL2 devices, SII was obtained for each selection model and the NAL-NL2 condition using probe microphone measurements and a Verifit 1 hearing aid analyzer with a 65 dB SPL male talker speech input. The SII was measured for each ear individually. Additionally, the REAR was obtained for the same 65 dB SPL signal. For comparison purposes, unaided SII and real-ear unaided response (REUR) were also recorded.

Office of Research in Clinical Amplification Nonsense Syllable Test. The Office of Research in Clinical Amplification Nonsense Syllable Test (referred to as the NST for the remainder of the article) female short list was used to assess the effect of the gain–frequency responses from each selection model and the custom-programmed NAL-NL2 condition. Developed by Kuk et al. (2010), the NST short list tests recognition of 32 randomized, distinct CVCVC (C = consonant, V = vowel) nonsense words spoken by a female native English talker. The items included all 25 American English consonant sounds and five vowel sounds. Each token was preceded by a carrier phrase “Please say the word....” Participants were instructed to repeat as many sounds as they heard and were encouraged to guess. The results were split into a percentage correct for consonants and a percentage correct for vowels. For each selection model, testing was completed in both quiet (55 dB SPL) and four-talker babble noise (+5 SNR, speech = 65 dB SPL) at 0° azimuth.

The NST was selected over more conventional speech recognition tests in order to isolate the effect of audibility. Evidence shows individuals with mild hearing impairment can perceptually restore inaudible speech segments about as effectively as individuals with normal hearing (Başkent et al., 2010). For this reason, speech recognition scores on word or sentence materials may overestimate an individual’s access to speech sounds by engaging restorative top-down processing.

Sound quality rating. Although understanding speech is arguably the most important measure of hearing aid benefit, perceived sound quality may also be an important mediator of device success. Indeed, reports of poor sound quality may be associated with hearing aid nonuse and dissatisfaction (Solheim et al., 2018). Thus, for each selection

model and the NAL-NL2 condition, perceived sound quality was measured in three domains (clarity, pleasantness, and loudness) in both quiet and noise using rating scales adapted from Neuman et al. (1998).

Participants listened to concatenated CST sentences in quiet (55 dB SPL) and noise (+5 SNR, speech = 65 dB SPL) at 0° azimuth. In quiet and noise, for each selection model and the NAL-NL2 condition, participants were given a set of three scales to rate clarity, pleasantness, and loudness (see Appendix D for the rating scales). Participants were instructed to listen to the sentence material and mark a vertical line on each scale corresponding to their rating of what they heard, using the anchor points as a guide. Ratings were scored to the nearest tenth of a point based on where participants had marked their response. Ratings ranged from 0 to 10.

The NST and sound quality rating, as well as the procedure used in the select-by-trying selection model, were conducted in a sound-treated booth. A Tannoy Di5t loudspeaker was used to present stimuli to the participants at 0° azimuth. The distance between the loudspeaker and the participant was 1.2 m.

Procedure. The study was approved by the institutional review board of the University of Iowa. After agreeing to participate and signing the consent form, participants' hearing thresholds were measured using pure-tone audiometry. If the participant met all required inclusion criteria, audiometry results were recorded but not disclosed. The researcher then measured the participant for his or her appropriate slim tube size, and all study hearing aids were configured accordingly. One-size-fits-all tulip dome tips were used for all participants.

Participants then completed the BHI QHC questionnaire followed by the select-by-trying procedure and administration of the NSRT online self-hearing test. At this point, presets were either selected or assigned for each of the five selection models: select-by-audiogram, select-by-self-test, select-by-trying, select-by-questionnaire, and select-by-random-assignment. After preset assignment, the researcher completed the custom device fitting and on-ear verification for the NAL-NL2 condition. Following custom programming, SIIs and REARs were obtained for each selection model and the NAL-NL2 condition.

The NST was completed in quiet and in noise for each of the five selection models and the NAL-NL2 condition for a total of 12 NST conditions. The order of the conditions was randomized across participants. Sound quality ratings were completed for the same 12 randomized conditions. Occasionally, the same left and right ear gain-frequency responses were assigned to more than one selection model. To ensure length of the experiment did not vary by participant, all participants were tested in 12 conditions regardless of duplicate preset selections. This meant that some participants were tested more than once with the same device settings. When this occurred, the test results were assigned to the corresponding randomized device condition, even if the results differed from those obtained using the same settings for a different condition. This was done to maintain appropriate randomization, which

was intended to control for learning, fatigue, or other systematic effects.

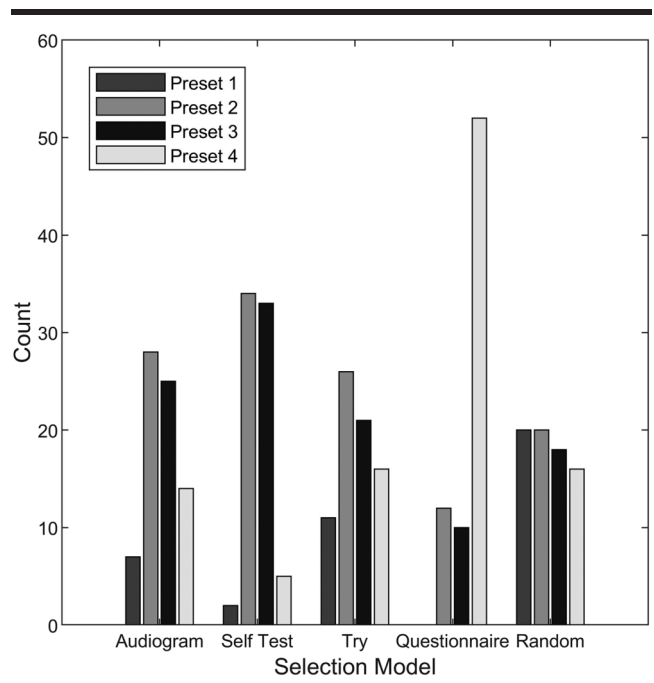
Prior to testing, the devices were inserted into the participant's ears by the researcher to ensure they were inserted correctly. Participants were blinded to the preset settings during testing; however, researchers were responsible for configuring the devices and therefore were not blinded to the condition or preset settings when scoring test materials (single-blinded design). For all testing, a practice condition was administered to confirm participants' understanding of the task. Testing was completed in two study sessions of between 2 and 2.5 hr each. Monetary compensation was provided to the participants following completion of the study.

Results

Selection of Gain-Frequency Responses

Figure 4 shows the number of ears assigned to each gain-frequency response (i.e., preset) in each selection model. Select-by-self-test, select-by-audiogram, and select-by-trying resulted in preset distributions favoring Presets 2 and 3, with fewer ears assigned to Presets 1 and 4. Notably, select-by-questionnaire resulted in a high number of ears assigned to Preset 4 and no ears assigned to Preset 1. A Pearson chi-squared test was conducted and indicated that the distribution of preset selection was significantly different across the five selection models ($\chi^2 = 120.8, p < .0001$). Post hoc Pearson chi-squared pairwise comparisons with Benjamini–

Figure 4. Number of ears assigned to each preset in each selection model. Audiogram = select-by-audiogram; Self Test = select-by-self-test; Try = select-by-trying; Questionnaire = select-by-questionnaire; Random = select-by-random-assignment.



Hochberg adjustment further indicated that all pairwise comparisons of selection models were significantly different from each other, except for select-by-trying versus select-by-random-assignment and select-by-trying versus select-by-audiogram.

All participants were able to self-administer the NSRT online hearing test successfully (select-by-self-test). Figure 5 shows the mean thresholds of the NSRT pseudo audiogram compared to participants' clinically measured audiometric thresholds. A two-way repeated-measures analysis of variance (ANOVA) was conducted to examine the effect of audiogram type (pseudo audiogram vs. clinically measured), test frequency, and their interaction on hearing threshold (data from both ears combined). The result first indicated that there was an interaction between audiogram type and frequency, $F(5, 196) = 5.8, p < .0001$. Hearing thresholds of the two audiogram types were compared at each frequency, and the Benjamini–Hochberg adjustment was used to account for multiple comparisons. The results indicated that the mean thresholds of the NSRT pseudo audiogram and the clinically measured audiometric thresholds differed significantly at 2000 and 8000 Hz.

In the select-by-trying model, all participants were able to self-select their preferred gain–frequency responses following the study procedure. Approximately 90% of participants (33/37) completed the select-by-trying procedure in 15 min or less; the remaining participants completed the process in under 30 min.

REAR

Figure 6 shows the mean REAR for a 65-dB SPL speech signal averaged across all participants, along with the REUR for reference. Mean NAL-NL2 prescriptive targets averaged across participants are also shown in the figure as asterisk symbols. The custom-programmed NAL-NL2 condition was closest to prescribed targets, while the selection models were slightly over target in the mid to high frequencies. To quantify deviation from target, the root-

mean-square (RMS) error in REAR between each condition and the prescribed NAL-NL2 target was calculated at 250, 500, 1000, 2000, and 4000 Hz. The mean RMS differences were 4.8 dB for select-by-audiogram, 5.4 dB for select-by-self-test, 5.4 dB for select-by-trying, 5.2 dB for select-by-random-assignment, and 5.6 dB for select-by-questionnaire. The NAL-NL2 condition came closest to prescribed targets with an RMS error of 1 dB. For comparison, the RMS difference between the REUR and the target REAR was 11.5 dB. A repeated-measures ANOVA test was conducted to determine the effect of selection model (including the NAL-NL2 condition) on RMS error (data from both ears combined). Results indicated that the RMS error differed significantly across selection models, $F(5, 160) = 89.6, p < .0001$. Post hoc pairwise comparisons with Benjamini–Hochberg adjustment further indicated that (a) the RMS error of the NAL-NL2 condition was lower than all five selection models and (b) the RMS error of select-by-audiogram was lower than those of select-by-self-test, select-by-trying, and select-by-questionnaire.

SII

Figure 7 shows the box plot of better-ear SII for each participant in each selection model and the NAL-NL2 condition. The unaided condition is shown for comparison. Repeated-measures one-way ANOVA across all experimental conditions and the unaided condition revealed a significant difference across groups, $F(6, 216) = 137.4, p < .0001$. Pairwise comparisons showed that all selection models and the NAL-NL2 condition were significantly different from the unaided condition at $p < .05$; all other paired comparisons were not significant at $p = .05$.

NST and Sound Quality Ratings

Figure 8 shows NST consonant scores and sound quality ratings for each of the selection models and the NAL-NL2 condition. Note that, for the NST, only consonant scores were analyzed due to ceiling effects for the

Figure 5. Mean hearing thresholds of the participants compared to mean hearing thresholds obtained from National Technical Institute for the Deaf Speech Recognition Test (NSRT) self-test pseudo audiogram for the right and left ears. Error bars indicate 1 SD.

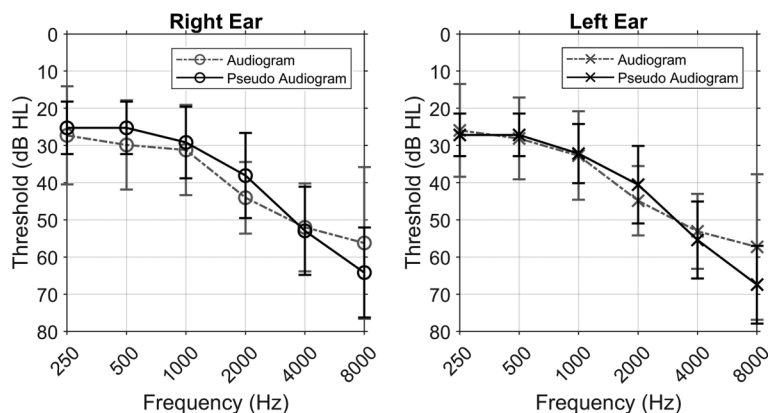
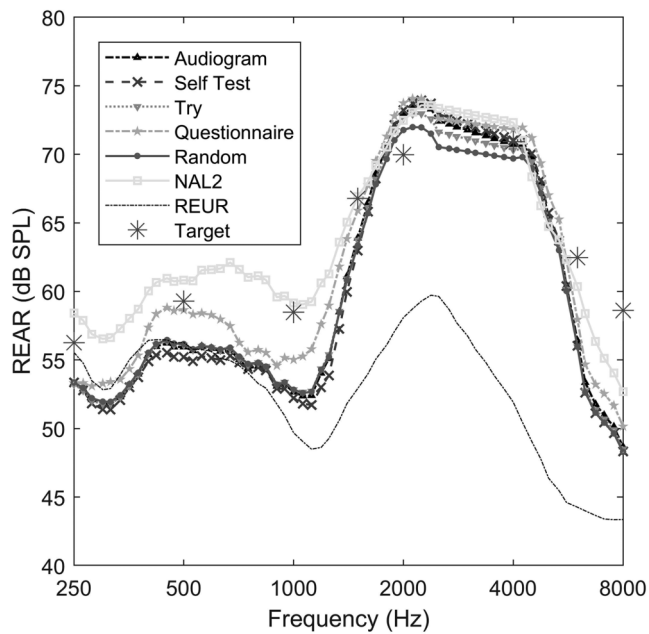


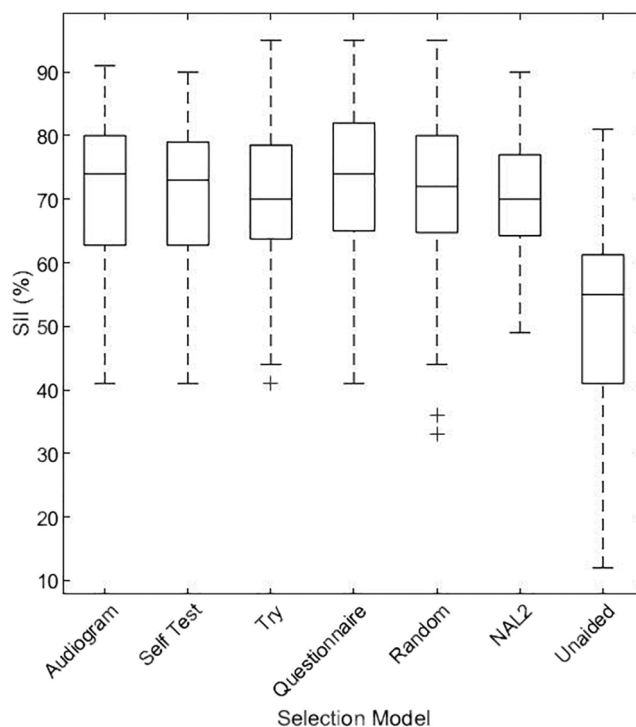
Figure 6. Mean real-ear aided response (REAR) as a function of frequency for each selection model and the National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) condition using a 65-dB SPL male speech input. Real-ear unaided response (REUR) was obtained using the same speech signal. Mean NAL-NL2 targets averaged across participants are plotted as asterisk symbols. Audiogram = select-by-audiogram; Self Test = select-by-self-test; Try= select-by-trying; Questionnaire = select-by-questionnaire; Random = select-by-random-assignment; NAL2 = the custom-programmed NAL-NL2 condition; Target = NAL-NL2 prescriptive target.



vowel score across participants. Preliminary analysis of sound quality judgments showed high correlation between clarity and pleasantness ratings in both quiet and noise ($r^2 = .53$ and $.5$, respectively). Accordingly, clarity and pleasantness ratings were averaged for each participant to create a composite clarity/pleasantness rating. This value was used in subsequent analyses in place of separate clarity and pleasantness ratings. The loudness rating was not highly correlated with either clarity or pleasantness; therefore, it was retained as a separate domain of sound quality for data analysis.

To determine the effect of selection model (including the custom-programmed NAL-NL2 condition) on NST consonant scores, two linear mixed-effects models were used (one for the quiet condition and another for the noise condition). Both models controlled for participant-specific better-ear PTA, sound quality clarity/pleasantness rating, and sound quality loudness rating. Sound quality ratings were included in the model to account for the trade-off between sound quality and speech intelligibility in hearing aid outcomes. In the hearing aid literature, it is well established that gain–frequency responses yielding high audibility for speech may result in poor subjective sound quality (Souza, 2002). If NST consonant scores were considered without regard to sound quality ratings, device settings

Figure 7. Box plots of better-ear Speech Intelligibility Index (SII) measure for each selection model, the verified National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) condition, and the unaided condition. The boundaries of the box represent the 25th and 75th percentiles, and the line inside the box represents the median. Error bars indicate the 10th and 90th percentiles. Plus signs represent outliers beyond the bounds of the error bars. Audiogram = select-by-audiogram; Self Test = select-by-self-test; Try= select-by-trying; Questionnaire = select-by-questionnaire; Random = select-by-random-assignment; NAL2 = the custom-programmed NAL-NL2 condition; Target = NAL-NL2 prescriptive target.

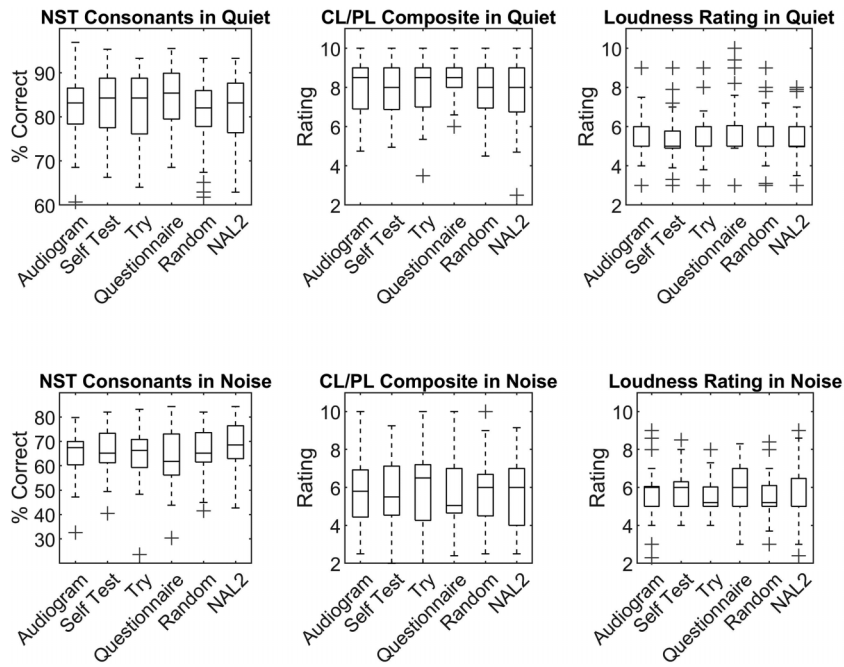


that overamplified speech might be deemed optimal despite unacceptable sound quality.

Figure 9A shows mean and 95% confidence interval for the linear mixed-effects model predicted NST consonant score of each selection model and the NAL-NL2 condition in quiet, controlling for better-ear PTA and sound quality ratings. Higher scores represent a greater percentage of NST consonants repeated correctly and, therefore, better outcomes. Pairwise comparisons were then conducted using model statistics. The results revealed that select-by-questionnaire performed significantly better than select-by-random-assignment (difference = 2.9%, $p = .007$). All other comparisons were not significant at $p = .05$. Results in noise are shown in Figure 9B. For the noise condition, the pairwise comparison between select-by-questionnaire and the NAL-NL2 condition was significant (difference = 4.5%, $p = .02$); all others were not significant at $p = .05$. See Appendixes E and F for detailed statistics.

The previous two linear mixed-effects models were used to characterize the effect of selection model on NST consonant scores. These models, however, did not

Figure 8. Box plots of Nonsense Syllable Test (NST) consonant scores (percent correct) and sound quality ratings (clarity/pleasantness [CL/PL] composite and loudness ratings) in quiet (top row) and noise (bottom row). Audiogram = select-by-audiogram; Self Test = select-by-self-test; Try= select-by-trying; Questionnaire = select-by-questionnaire; Random = select-by-random-assignment; NAL2 = the custom-programmed National Acoustic Laboratories–Non-Linear 2 condition.

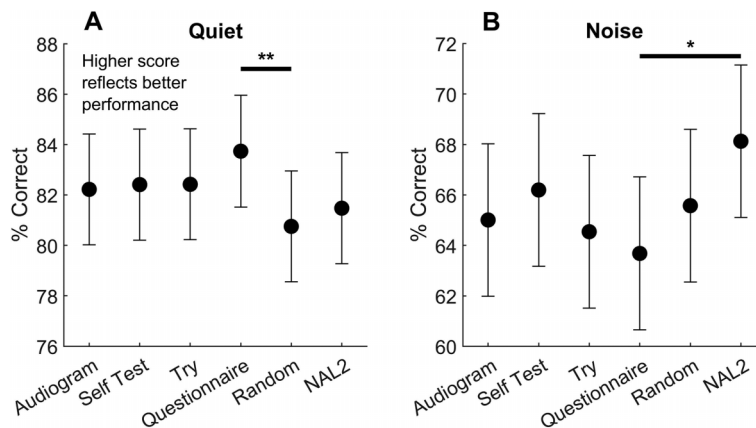


specifically examine the extent to which the five selection models approximated outcomes for the custom-programmed NAL-NL2 condition on the individual participant level. If best-practice verification is considered the gold standard of hearing aid fitting, noncustom outcomes might be judged relative to each individual’s best-practice outcomes. Toward this end, two linear mixed-effects models were fit to model the deviation of selection models from NAL-NL2

speech recognition and sound quality outcomes in quiet and noise. These analyses were modeled on the premise that a custom match to validated prescriptive targets is the current gold standard for hearing aid fitting; thus, the efficacy of OTC outcomes should be gauged relative to best-practice verification.

To account for best-practice outcomes, each participant’s NST consonant scores and sound quality ratings

Figure 9. Estimated mean and 95% confidence interval of percent correct Nonsense Syllable Test (NST) consonants in quiet (A) and noise (B) of each selection model, controlling for better-ear thresholds and sound quality ratings. *Significance at $p < .05$. **Significance at $p < .01$. Audiogram = select-by-audiogram; Self Test = select-by-self-test; Try = select-by-trying; Questionnaire = select-by-questionnaire; Random = select-by-random-assignment; NAL2 = the custom-programmed National Acoustic Laboratories–Non-Linear 2 condition.



(clarity/pleasantness composite score and loudness rating) in each of the five selection models were centered by subtracting the corresponding score of the NAL-NL2 condition. For each participant, this calculation yielded three deviation values—one for each score (NST consonant, clarity/pleasantness, and loudness). To account for the relationship between the three outcomes of interest, the squared Mahalanobis distance was calculated for the centered NST consonant, sound quality clarity/pleasant, and sound quality loudness test scores. The Mahalanobis distance is the distance between two points in multivariate space for correlated items (Mahalanobis, 1930). A log transformation was done on the squared Mahalanobis distance because the raw values did not meet the assumptions for homoscedasticity and normality. Linear mixed-effects models with random intercepts for participants were then created. The dependent variable is log-transformed Mahalanobis distance, and the independent variable is selection model. Each participant's better-ear PTA was accounted for in the models.

Figure 10A shows the mean and 95% confidence interval for the linear mixed-effects model predicted deviation of each selection model from the NAL-NL2 condition in quiet, controlling for better-ear PTA. In this analysis, smaller deviation scores suggest greater consistency with NAL-NL2 outcomes on the individual level and thus better outcomes. Pairwise comparisons were then conducted using model statistics. The results revealed a significant difference between select-by-audiogram and select-by-self-test (difference = 0.6, $p = .02$) and select-by-questionnaire and select-by-self-test (difference = 0.57, $p = .03$). These results indicate that select-by-self-test produced smaller deviations from NAL-NL2 outcomes on the individual participant level than select-by-questionnaire and select-by-audiogram. All other comparisons were not significant at $p = .05$. Figure 10B shows the results in noise. For the noise condition, all pairwise comparisons were not significant. See Appendixes G and H for detailed statistics.

Discussion

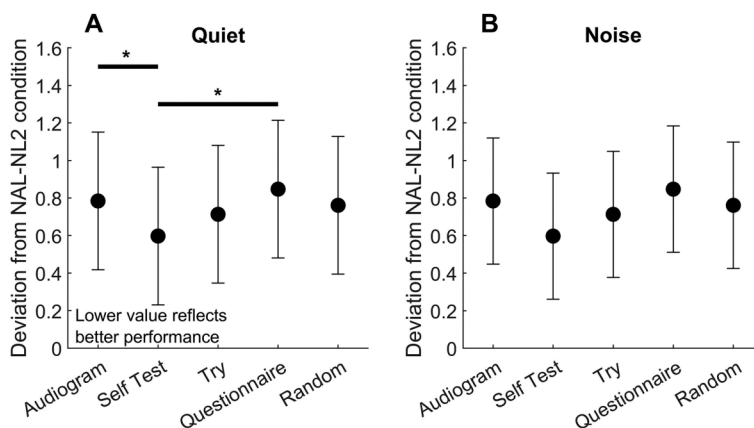
In this study, older adults selected amplification from the set of four preconfigured gain–frequency responses developed in Aim 1 using a variety of self-selection methods (selection models) in the laboratory. For each selection model, outcome measures were obtained in several domains including audibility, speech recognition, and sound quality. These outcomes were compared to a clinical best-practice condition in which participants wore the same devices custom fit to approximate their NAL-NL2 REAR prescriptive targets. The purpose of the study was twofold. The first aim was to validate the efficacy of the set of four gain–frequency responses developed in the previous experiment, and the second aim was to identify a method by which older adults can self-select appropriate amplification from the set.

Selection of Gain–Frequency Responses

Patterns of preset assignments differed across selection models (see Figure 4), suggesting that the selection models represented qualitatively different criteria for selecting noncustom amplification. In other words, the selection method influences the gain–frequency responses selected by older adults, at least given the set of four gain–frequency responses used in this study. This result underscores the importance of determining the efficacy of different modes of self-selection and validates the purpose of this experiment—to determine the best way for older adults to select from the set of four preconfigured gain–frequency responses.

Recall that select-by-questionnaire produced a markedly different pattern of preset assignments than the other selection models. Specifically, select-by-questionnaire assigned a large number of ears to higher preset numbers (more gain). This result indicates that many participants self-rated their hearing loss burden as quite severe, despite having mild-to-moderate hearing loss. It is possible that participants with

Figure 10. Estimated mean and 95% confidence interval for the deviation of each selection model from the National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) condition in quiet (A) and in noise (B). *Significance at $p < .05$. Audiogram = select-by-audiogram; Self Test = select-by-self-test; Try = select-by-trying; Questionnaire = select-by-questionnaire; Random = select-by-random-assignment.



a high degree of self-perceived hearing loss were more likely to participate in the study, or it could be that mild-to-moderate hearing loss spans a wider range of self-perceived handicap than traditionally thought (Humes, 2020). This result is consistent with previous studies that have demonstrated poor sensitivity and specificity of validated self-assessments for detecting criterion levels of hearing loss in population-based samples of older adults (Cox et al., 2003; Gates et al., 2003; Nondahl et al., 1998). Future research should explore the relationship between degree and configuration of hearing loss and self-perceived hearing loss, particularly among the OTC target demographic.

Participants demonstrated the ability to complete the select-by-trying and select-by-self-test procedures. In both selection models, most participants independently completed each self-selection procedure within a short time period (less than 15 min) in the lab. This demonstrates the potential feasibility of OTC fitting paradigms in which older adults take an active role in determining their subjective preferences for amplification and/or self-administering objective hearing assessments. Results showed, however, that select-by-self-test and select-by-trying produced different patterns of preset assignments. In designing OTC fitting paradigms, manufacturers should pay careful attention to the implications of utilizing either a self-hearing test or select-by-trying procedure and should not assume these methods will generate the same selections. Future research should investigate factors that impact self-hearing test accuracy and subjective gain preference in the context of preconfigured, noncustom amplification.

REAR and SII

Unsurprisingly, the verified NAL-NL2 condition produced REAR values significantly closer to NAL-NL2 targets than the other selection models. Among selection models, select-by-audiogram came closer to meeting prescribed REAR targets than select-by-self-test, select-by-trying, and select-by-questionnaire. This result is equally unsurprising, as select-by-audiogram utilized clinically measured thresholds to assign presets—the same thresholds used by the NAL-NL2 fitting formula to assign REAR target values.

These results could be construed as evidence that custom programming and clinically measured audiometric thresholds are likely to produce better outcomes. SII data (see Figure 7), however, indicate that differences in REARs across selection models and the NAL-NL2 condition did not yield significant differences in audibility. All selection models and the NAL-NL2 condition produced comparable SII values, which were all significantly higher than the unaided SII. Thus, all conditions represented a significant improvement over the unaided condition, but they were not significantly different from each other. This finding is consistent with a recent study by Brody et al., which demonstrated that some amplification is better than no amplification and that, across different frequency responses, differences in audibility are small (Brody et al., 2018).

NST and Sound Quality Ratings

The first pair of linear mixed-effects models characterized the effect of selection model on NST consonant scores, controlling for participant-specific better-ear PTA, sound quality clarity/pleasantness rating, and sound quality loudness rating. NST consonant scores were used because participants showed ceiling effects for vowels, particularly in quiet. This result was not unexpected, as individuals with mild-to-moderate hearing loss traditionally show higher performance for vowels than consonants in phonemically scored speech tests (Kuk et al., 2010; Owens & Schubert, 1977).

In quiet, select-by-random-assignment produced statistically poorer NST consonant scores than select-by-questionnaire (see Figure 9A). This is likely because many participants were underamplified in the select-by-random-assignment model due to about a quarter of participants randomly assigned to Preset 1, which provided the lowest gain of the four presets. Conversely, many participants were overamplified in the select-by-questionnaire model due to the large number of participants assigned to Preset 4, which provided the highest gain of the four presets. Regardless, this result suggests that select-by-random-assignment underperforms in quiet and thus should be considered a suboptimal selection model for older adults. Aside from random selection, the other four selection models—select-by-audiogram, select-by-trying, select-by-self-test, and select-by-questionnaire—produced good speech recognition outcomes in quiet.

In noise, select-by-questionnaire yielded significantly poorer speech recognition scores than the verified NAL-NL2 condition (see Figure 9B). This could have been the result of overamplification in the select-by-questionnaire model, which may have led to adverse effects on speech recognition and sound quality ratings in noise. In any case, select-by-questionnaire produced poor outcomes in noise. All other selection models, however, produced comparable outcomes in noise to clinically fit hearing aids.

The next pair of linear mixed-effects models examined the extent to which outcomes of the five selection models deviated from the NAL-NL2 condition on the individual participant level. Results in noise (see Figure 10B) indicated that all selection models produced comparable deviations from NAL-NL2 outcomes; however, results in quiet (see Figure 10A) indicated that select-by-self-test produced individual outcomes most consistent with the verified NAL-NL2 condition. This outcome was not expected and is not easily explained by the data. It is possible, however, that the NSRT self-hearing test produced systematic deviations from clinically measured audiometric thresholds that led to preset assignments more consistent with the NAL-NL2 condition than those assigned in the other selection models. Further analysis should investigate the relationship between the results of self-administered hearing tests, clinically measured audiometric thresholds, and prescribed NAL-NL2 REAR targets.

The results from the Aim 2 study indicate that select-by-random-assignment and select-by-questionnaire are suboptimal, as they produce poor outcomes in quiet and noise,

respectively. Select-by-audiogram, select-by-self-test, and select-by-trying produced comparable outcomes to the NAL-NL2 condition. When considering outcomes on the individual level, select-by-self-test produced outcomes most consistent with best-practice clinical verification. This result, however, should be taken with caution as it is not easily explained by the data.

Limitations

Aim 1

The Aim 1 study was intended to develop a set of four presets capable of fitting a large percentage of older adults with mild-to-moderate presbycusis within a tight fit criterion. In order to pursue this aim, a number of limiting decisions and definitions were employed. Several limitations resulted from the choices made in preset creation.

One key limitation is that Aim 1 employed a strict definition of mild-to-moderate hearing loss for preset development (PTA at 500, 1000, and 2000 Hz of ≥ 25 and ≤ 55 dB HL with no threshold poorer than 65 dB HL from 500 to 6000 Hz). While the study-defined audiometric range is representative of many older adults with mild-to-moderate presbycusis, some might advocate for the use of different criteria for defining mild-to-moderate hearing loss. In particular, older adults' decision to purchase OTC amplification may rely on perceived hearing loss rather than well-defined audiometric threshold criteria. As a result, the eventual OTC demographic could reflect a diversity of audiometric profiles. Future work might examine the effect of broadening the definition of mild-to-moderate hearing loss on the development of OTC presets.

Another limitation is that the Aim 1 preset development rationale only considered 65-dB SPL NAL-NL2 REAR targets. Soft and loud targets (e.g., 55 and 75 dB SPL) were not included. As a result, the presets created in Aim 1 only define gain settings for moderate speech (i.e., 65 dB SPL); they do not give a rationale for setting WDRC parameters. Future work in OTC preset development should extend the current study by incorporating a rationale for WDRC settings.

Finally, Aim 1 did not consider the effect of volume control. In the real world, OTC users may utilize volume control to adjust gain for a given preset frequency response. In the context of the Aim 1 study, inclusion of volume control might reduce the number of presets needed and/or increase the fit percentage of the existing set of four presets. That said, adding volume control introduces an additional variable into self-selection of presets, particularly for the select-by-trying model. Future research might investigate the impact of volume control on users' self-selection of OTC presets.

Aim 2

Aim 2 was intended to serve as a laboratory efficacy study for the set of four preconfigured gain–frequency responses developed in the first aim. Accordingly, the study was tightly controlled to isolate the effect of gain–frequency

response on participants' self-selections and outcomes. The results, therefore, cannot be easily generalized to the real world.

One key limitation is that study participants were not expected to master hands-on hearing aid skills. Participants were not asked to identify the left versus right hearing aid or insert the devices; the researcher always inserted study hearing aids to ensure proper placement. Additionally, participants were not asked to configure the physical fit of their hearing aids. Instead, the researcher measured and installed the appropriate slim tube size for each participant. The researcher also maintained the study hearing aids, keeping them clean and in good working order. In the real world, end users will be responsible for these and other hands-on hearing aid skills that may affect device performance and outcomes.

Another important limitation is that all hearing aid features were turned off for the Aim 2 study. This control was important for investigating the effect of gain–frequency response on self-selection and laboratory outcomes; however, it is not particularly representative of the real world where hearing aid features such as directional microphones and digital noise reduction may improve the user experience in noisy listening environments. In the real world, hearing aid features may play an important role in OTC outcomes.

Additionally, it should be noted that laboratory speech materials may not approximate an individual's real-world listening experience. In the real world, end users may experience dynamically changing listening environments and/or may have listening goals beyond that which can be measured in a laboratory efficacy study.

A real-world field trial is needed to address many of the above limitations. In conducting an OTC field trial, particular attention should be paid to inclusion of low-income and minority participants. In the Aim 2 study, data regarding education and socioeconomic status were not collected; however, the Aim 2 participants likely did not reflect the racial and socioeconomic diversity of the target OTC population. Self-selection and real-world OTC outcomes might differ by income, education, or socioeconomic status. Future studies, therefore, should strive to enroll diverse populations representative of those who are unserved/underserved by the traditional clinical model.

Conclusion

The data indicate that a set of four OTC presets could produce comparable outcomes to best-practice verification in a laboratory setting for individuals with mild-to-moderate hearing loss. Importantly, the data also show that older adults can self-select appropriate amplification using several selection methods. Taken together, the results of this two-part investigation provide empirical evidence for the efficacy of a new OTC fitting paradigm. The data obtained from this study may inform development of evidence-based OTC amplification in support of public health initiatives aimed at promoting affordable, accessible *quality* hearing

health care for older adults with mild-to-moderate age-related hearing loss.

Acknowledgments

This study was supported by the Retirement Research Foundation (RRF 2017-163) and the National Institute on Deafness and Other Communication Disorders (R01DC015997). Portions of this article were presented at the annual meeting of the American Auditory Society, March 2019, Scottsdale, Arizona, and the annual meeting of the American Academy of Audiology, March 2019, Columbus, Ohio.

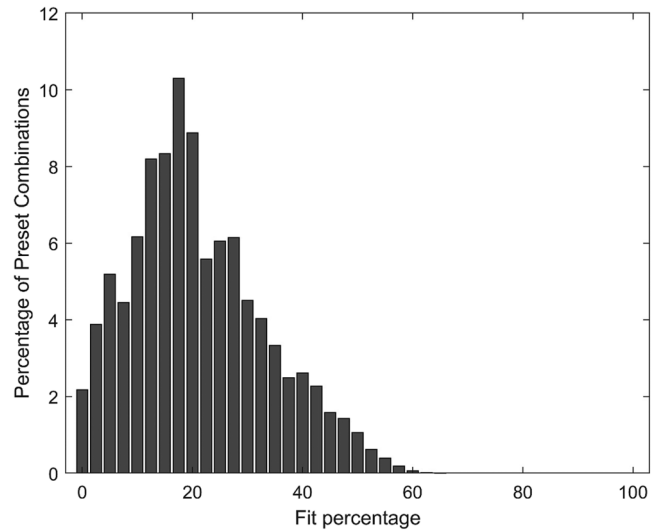
References

- Aazh, H., & Moore, B. C. J. (2007). The value of routine real ear measurement of the gain of digital hearing aids. *Journal of the American Academy of Audiology*, 18(8), 653–664. <https://doi.org/10.3766/jaaa.18.8.3>
- American National Standards Institute. (2012). *Methods for calculation of the Speech Intelligibility Index* (ANSI S3.5-1997) (R2012).
- Bainbridge, K. E., & Ramachandran, V. (2014). Hearing aid use among older United States adults: The National Health and Nutrition Examination Survey, 2005–2006 and 2009–2010. *Ear and Hearing*, 35(3), 289–294. <https://doi.org/10.1097/01.aud.0000441036.40169.29>
- Baker, S., & Jenstad, L. (2017). Matching real-ear targets for adult hearing aid fittings: NAL-NL1 and DSL v5.0 prescriptive formulae. *Canadian Journal of Speech-Language Pathology*, 41(2), 227–235.
- Başkent, D., Eiler, C. L., & Edwards, B. (2010). Phonemic restoration by hearing-impaired listeners with mild to moderate sensorineural hearing loss. *Hearing Research*, 260(1–2), 54–62. <https://doi.org/10.1016/j.heares.2009.11.007>
- Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, 14(2), 113–120. <https://doi.org/10.1177/1084713810379609>
- Bochner, J. H., Garrison, W. M., & Doherty, K. A. (2015). The NTID speech recognition test: NSRT. *International Journal of Audiology*, 54(7), 490–498. <https://doi.org/10.3109/14992027.2014.991976>
- Boothroyd, A., & Mackerse, C. (2017). A “Goldilocks” approach to hearing-aid self-fitting: User interactions. *American Journal of Audiology*, 26(3S), 430–435. https://doi.org/10.1044/2017_AJA-16-0125
- Brody, L., Wu, Y.-H., & Stangl, E. (2018). A comparison of personal sound amplification products and hearing aids in ecologically relevant test environments. *American Journal of Audiology*, 27(4), 581–593. https://doi.org/10.1044/2018_AJA-18-0027
- Chan, Z. Y. T., & McPherson, B. (2015). Over-the-counter hearing aids: A lost decade for change. *BioMed Research International*, 2015, 1–15. <https://doi.org/10.1155/2015/827463>
- Chang, J. E., Weinstein, B., Chodosh, J., & Blustein, J. (2018). Hospital readmission risk for patients with self-reported hearing loss and communication trouble. *Journal of the American Geriatrics Society*, 66(11), 2227–2228. <https://doi.org/10.1111/jgs.15545>
- Cheng, C. M., & McPherson, B. (2000). Over-the-counter hearing aids: Electroacoustic characteristics and possible target client groups. *Audiology*, 39(2), 110–116. <https://doi.org/10.3109/00206090009073062>
- Clark, J. G. (1981). Uses and abuses of hearing loss classification. *ASHA*, 23(7), 493–500.
- Contrera, K. J., Betz, J., Deal, J., Choi, J. S., Ayonayon, H. N., Harris, T., Helzner, E., Martin, K. R., Mehta, K., Pratt, S., Rubin, S. M., Satterfield, S., Yaffe, K., Simonsick, E. M., Lin, F. R., & Health ABC Study. (2017). Association of hearing impairment and anxiety in older adults. *Journal of Aging and Health*, 29(1), 172–184. <https://doi.org/10.1177/0898264316634571>
- Convery, E., Keidser, G., Dillon, H., & Hartley, L. (2011). A self-fitting hearing aid: Need and concept. *Trends in Amplification*, 15(4), 157–166. <https://doi.org/10.1177/1084713811427707>
- Convery, E., Keidser, G., Seeto, M., & McLelland, M. (2017). Evaluation of the self-fitting process with a commercially available hearing aid. *Journal of the American Academy of Audiology*, 28(2), 109–118. <https://doi.org/10.3766/jaaa.15076>
- Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing*, 8(Suppl. 5), 119S–126S. <https://doi.org/10.1097/00003446-198710001-00010>
- Cox, R. M., Alexander, G. C., & Gray, G. A. (2003). Audiometric correlates of the unaided APHAB. *Journal of the American Academy of Audiology*, 14(7), 361–371. <https://doi.org/10.1055/s-0040-1715755>
- Deal, J. A., Goman, A. M., Albert, M. S., Arnold, M. L., Burgard, S., Chisolm, T., Couper, D., Glynn, N. W., Gmelin, T., Hayden, K. M., Mosley, T., Pankow, J. S., Reed, N., Sanchez, V. A., Sharrett, A. R., Thomas, S. D., Coresh, J., & Lin, F. R. (2018). Hearing treatment for reducing cognitive decline: Design and methods of the Aging and Cognitive Health Evaluation in Elders randomized controlled trial. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4(1), 499–507. <https://doi.org/10.1016/j.trci.2018.08.007>
- Demeester, K., van Wieringen, A., Hendrickx, J. J., Topsakal, V., Franssen, E., van Laer, L., van Camp, G., & Van de Heyning, P. (2009). Audiometric shape and presbycusis. *International Journal of Audiology*, 48(4), 222–232. <https://doi.org/10.1080/14992020802441799>
- Dillon, H., Keidser, G., Ching, T., Flax, M., & Brewer, S. (2014). *Development of NAL-NL2*. <https://hearworks.com.au/wp-content/uploads/2014/06/nal-nl2-presentation.pdf>
- Edwards, B. (2020). Emerging technologies, market segments, and MarkeTrak 10 insights in hearing health technology. *Seminars in Hearing*, 41(1), 37–54. <https://doi.org/10.1055/s-0040-1701244>
- Foley, D. M., Frick, K. D., & Lin, F. R. (2014). Association between hearing loss and healthcare expenditures in older adults. *Journal of the American Geriatrics Society*, 62(6), 1188–1189. <https://doi.org/10.1111/jgs.12864>
- Gallacher, J., Ilubaera, V., Ben-Shlomo, Y., Bayer, A., Fish, M., Babisch, W., & Elwood, P. (2012). Auditory threshold, phonologic demand, and incident dementia. *Neurology*, 79(15), 1583–1590. <https://doi.org/10.1212/WNL.0b013e31826e263d>
- Garrison, W. M., & Bochner, J. H. (2015). Applications of the NTID speech recognition test (NSRT). *International Journal of Audiology*, 54(11), 828–837.
- Garrison, W. M., & Bochner, J. H. (2017). *Accommodating the OTC hearing aid discussion: Hearing screening in the absence of audiometry—Part 2*. <http://hearinghealthmatters.org/waynesworld/2017/hearing-screening-for-otc-hearing-aids/>
- Gates, G. A., Murphy, M., Rees, T. S., & Fraher, A. (2003). Screening for handicapping hearing loss in the elderly. *Journal of Family Practice*, 52(1), 56–62.
- Goman, A. M., & Lin, F. R. (2018). Hearing loss in older adults—From epidemiological insights to national initiatives. *Hearing Research*, 369, 29–32. <https://doi.org/10.1016/j.heares.2018.03.031>

- Humes, L. E. (2020). What is “normal hearing” for older adults and can “normal-hearing older adults” benefit from hearing care intervention? *Hearing Review*, 27(7), 12–18.
- Humes, L. E., Rogers, S. E., Quigley, T. M., Main, A. K., Kinney, D. L., & Herring, C. (2017). The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: A randomized double-blind placebo-controlled clinical trial. *American Journal of Audiology*, 26(1), 53–79. https://doi.org/10.1044/2017_AJA-16-0111
- Jenstad, L. M., Bagatto, M. P., Seewald, R. C., Scollie, S. D., Cornelisse, L. E., & Scicluna, R. (2007). Evaluation of the desired sensation level [input/output] algorithm for adults with hearing loss: The acceptable range for amplified conversational speech. *Ear and Hearing*, 28(6), 793–811. <https://doi.org/10.1097/AUD.0b013e318157670a>
- Jiam, N. T. L., Li, C., & Agrawal, Y. (2016). Hearing loss and falls: A systematic review and meta-analysis. *The Laryngoscope*, 126(11), 2587–2596. <https://doi.org/10.1002/lary.25927>
- Keidser, G., Dillon, H., Carter, L., & O'Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends in Amplification*, 16(4), 211–223. <https://doi.org/10.1177/1084713812468511>
- Keidser, G., Dillon, H., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, 1(e24), 88–90. <https://doi.org/10.4081/audiore.2011.e24>
- Kochkin, S. (2014). A comparison of consumer satisfaction, subjective benefit, and quality of life changes associated with traditional and direct-mail hearing aid use. *Hearing Review*, 21(1), 16–26.
- Kochkin, S., & Bentler, R. (2010). The validity and reliability of the BHI Quick Hearing Check. *Hearing Review*, 17(12), 12–28.
- Krishnamurti, S. (2009). Sensorineural hearing loss associated with occupational noise exposure: Effects of age-corrections. *International Journal of Environmental Research and Public Health*, 6(3), 889–899. <https://doi.org/10.3390/ijerph6030889>
- Kuk, F., Lau, C. C., Korhonen, P., Crose, B., Peeters, H., & Keenan, D. (2010). Development of the ORCA nonsense syllable test. *Ear and Hearing*, 31(6), 779–795. <https://doi.org/10.1097/AUD.0b013e3181e97bfb>
- Lin, F. R., Metter, E. J., O'Brien, R. J., Resnick, S. M., Zonderman, A. B., & Ferrucci, L. (2011). Hearing loss and incident dementia. *Archives of Neurology*, 68(2), 214–220. <https://doi.org/10.1001/archneurol.2010.362>
- Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., Simonsick, E. M., & Health ABC Study Group. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, 173(4), 293–299. <https://doi.org/10.1001/jamainternmed.2013.1868>
- Mahalanobis, P. C. (1930). On tests and measures of groups divergence. *Journal of Asiatic Sociology of Bengal*, 26, 541–588.
- Nachtegaal, J., Smit, J. H., Smits, C., Bezemer, P. D., van Beek, J. H. M., Festen, J. M., & Kramer, S. E. (2009). The association between hearing status and psychosocial health before the age of 70 years: Results from an internet-based national survey on hearing. *Ear and Hearing*, 30(3), 302–312. <https://doi.org/10.1097/AUD.0b013e31819c6e01>
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>
- Neuman, A. C., Bakke, M. H., Mackersie, C., Hellman, S., & Levitt, H. (1998). The effect of compression ratio and release time on the categorical rating of sound quality. *The Journal of the Acoustical Society of America*, 103(5), 2273–2281. <https://doi.org/10.1121/1.422745>
- Nondahl, D. M., Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, R., & Klein, B. E. K. (1998). Accuracy of self-reported hearing loss. *Audiology*, 37(5), 295–301. <https://doi.org/10.3109/00206099809072983>
- Owens, E., & Schubert, E. D. (1977). Development of the California consonant test. *Journal of Speech and Hearing Research*, 20(3), 463–474. <https://doi.org/10.1044/jshr.2003.463>
- Pearsons, K. S., Bennett, R. L., & Fidell, S. (1977). *Speech levels in various noise environments (Report No. EPA-600/1-77-025)*. U.S. Environmental Protection Agency.
- Pew Research Center. (2017). *Technology use among seniors*. <http://www.pewinternet.org/2017/05/17/technology-use-among-seniors>
- Polonenko, M. J., Scollie, S. D., Moodie, S., Seewald, R. C., Laurnagaray, D., Shantz, J., & Richards, A. (2010). Fit to targets, preferred listening levels, and self-reported outcomes for the DSL v5.0 a hearing aid prescription for adults. *International Journal of Audiology*, 49(8), 550–560. <https://doi.org/10.3109/14992021003713122>
- Reed, N. S., Betz, J., Lin, F. R., & Mamo, S. K. (2017). Pilot electroacoustic analyses of a sample of direct-to-consumer amplification products. *Otology and Neurotology*, 38(6), 804–808. <https://doi.org/10.1097/MAO.0000000000001414>
- Smith, C., Wilber, L. A., & Cavitt, K. (2016). PSAPs vs hearing aids: An electroacoustic analysis of performance and fitting capabilities. *Hearing Review*, 23(7), 18–24. <https://doi.org/10.1097/01.HJ.0000484547.75301.11>
- Solheim, J., Gay, C., & Hickson, L. (2018). Older adults' experiences and issues with hearing aids in the first six months after hearing aid fitting. *International Journal of Audiology*, 57(1), 31–39. <https://doi.org/10.1080/14992027.2017.1380849>
- Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in Amplification*, 6(4), 131–165. <https://doi.org/10.1177/108471380200600402>
- The President's Council of Advisors on Science and Technology, Executive Office of the President. (2015). *Aging America & hearing loss: Imperative of improved hearing technologies*. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_hearing_tech_letterreport_final.pdf
- Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J. (2018). Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. *Ear and Hearing*, 39(2), 293–304. <https://doi.org/10.1097/AUD.0000000000000486>

Appendix A

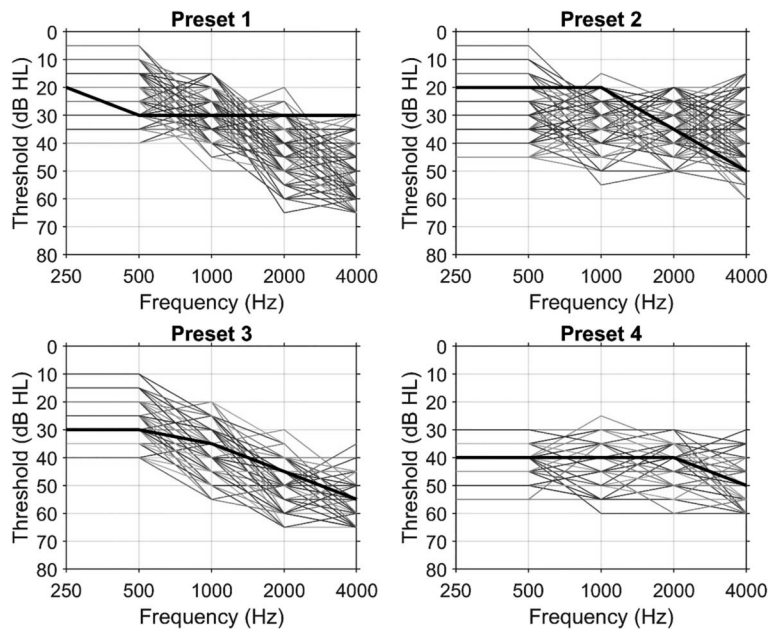
Histogram of Fit Percentages for All Possible Combinations of Four Gain–Frequency Responses From the Set of 642 Candidate Gain–Frequency Responses in Aim 1



Appendix B

National Health and Nutrition Examination Survey (NHANES) Audiograms Fit by Each Preset

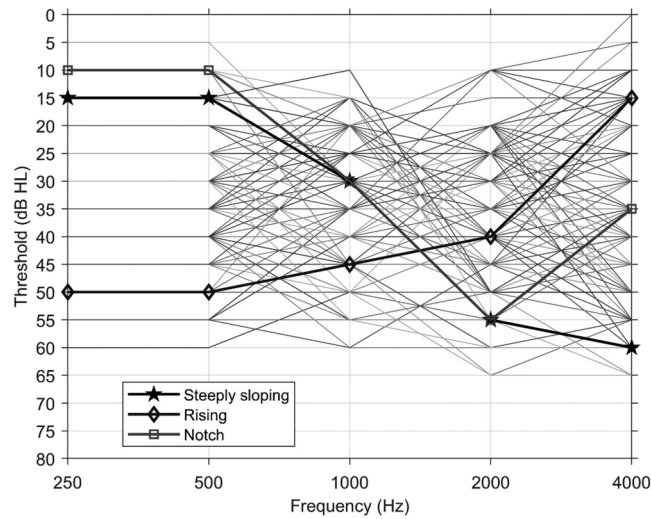
An audiogram was classified as “fit” by a given preset if the audiogram’s National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) real-ear aided response (REAR) target was within ± 5 dB of the preset’s REAR target for all octave frequencies from 250 to 4000 Hz. The bold lines represent the base audiogram of each preset.



Appendix C

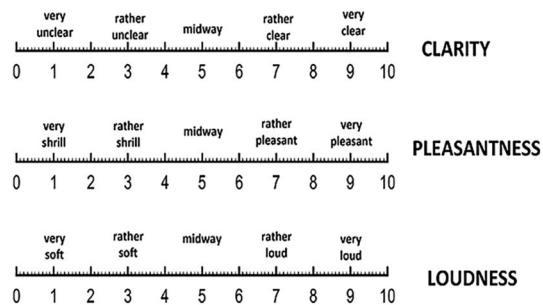
National Health and Nutrition Examination Survey (NHANES) Audiograms That Could Not Be Fit by the Presets

An audiogram was classified as “fit” by a given preset if the audiogram’s National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) real-ear aided response (REAR) target was within ± 5 dB of the preset’s REAR target for all octave frequencies from 250 to 4000 Hz. The bold lines show examples of the three primary audiogram types not able to be fit: a) steeply sloping; b) rising; c) notched.



Appendix D

Rating Scales Used by Participants to Rate Perceived Sound Quality in Both Quiet and Noise



Appendix E

Statistics of Linear Mixed-Effects Models Examining the Effect of Selection Model (Including the National Acoustic Laboratories–Non-Linear 2 [NAL-NL2] Condition) on Nonsense Syllable Test (NST) Consonant Scores in Quiet, Controlled for Participant-Specific Better-Ear Pure-Tone Average (PTA), Sound Quality Clarity/Pleasantness Rating, and Sound Quality Loudness Rating

Mean and 95% confidence interval (CI) estimation

Selection model	LS means ^a	df ^b	95% CI
Audiogram	82.2	58.9	[80.0, 84.4]
Self test	82.4	59.3	[80.2, 84.6]
Try	82.4	59	[80.2, 84.6]
Questionnaire	83.7	60.9	[81.5, 86.0]
Random	80.8	59	[78.6, 83.0]
NAL2	81.5	59.7	[79.3, 83.7]

Note. LS = least square.

^aBetter-ear PTA fixed at 33, loudness rating fixed at 5.37, and clarity/pleasant rating fixed at 7.98. ^bSatterthwaite degrees of freedom.

Pairwise comparison

Contrast	Difference ^b	Test (df ^a)	p
Audiogram – Self test	–0.19	$t = -0.228$ (179)	1.000
Audiogram – Try	–0.20	$t = -0.242$ (178)	1.000
Audiogram – Questionnaire	–1.51	$t = -1.793$ (180)	.473
Audiogram – Random	1.47	$t = 1.773$ (178)	.486
Audiogram – NAL2	0.75	$t = 0.895$ (179)	.947
Self test – Try	–0.01	$t = -0.014$ (179)	1.000
Self test – Questionnaire	1.32	$t = 1.544$ (181)	.636
Self test – Random	–1.66	$t = -2.001$ (178)	.346
Self test – NAL2	–0.94	$t = -1.124$ (179)	.871
Try – Questionnaire	1.31	$t = 1.564$ (179)	.623
Try – Random	–1.67	$t = -2.01$ (179)	.341
Try – NAL2	–0.95	$t = -1.128$ (179)	.869
Questionnaire – Random	2.98	$t = 3.507$ (180)	.007
Questionnaire – NAL2	2.26	$t = 2.615$ (181)	.099
Random – NAL2	0.73	$t = 0.870$ (179)	.953

^aSatterthwaite degrees of freedom. ^bBetter-ear PTA fixed at 33, loudness rating fixed at 5.37, and clarity/pleasant rating fixed at 7.98.

Appendix F

Statistics of Linear Mixed-Effects Models Examining the Effect of Selection Model (Including the National Acoustic Laboratories–Non-Linear 2 [NAL-NL2] Condition) on Nonsense Syllable Test (NST) Consonant Scores in Noise, Controlled for Participant-Specific Better-Ear Pure-Tone Average (PTA), Sound Quality Clarity/Pleasantness Rating, and Sound Quality Loudness Rating

Mean and 95% confidence interval (CI) estimation

Selection model	LS means ^a	df ^b	95% CI
Audiogram	65.0	74.5	[62.0, 68.0]
Self test	66.2	74.5	[63.2, 69.2]
Try	64.5	74.6	[61.5, 67.6]
Questionnaire	63.7	75.2	[60.7, 66.7]
Random	65.6	74.7	[62.6, 68.6]
NAL2	68.1	74.6	[65.1, 71.2]

Note. LS = least square.

^aBetter-ear PTA fixed at 33, loudness rating fixed at 5.77, and clarity/pleasant rating fixed at 5.77. ^bSatterthwaite degrees of freedom.

Pairwise comparison

Contrast	Difference ^b	Test (df ^a)	p
Audiogram – Questionnaire	1.33	$t = 0.964 (178)$.929
Audiogram – NAL2	-3.12	$t = -2.276 (178)$.209
Audiogram – Random	-0.57	$t = -0.414 (178)$.998
Audiogram – Self test	-1.19	$t = -0.868 (178)$.954
Audiogram – Try	0.47	$t = 0.339 (178)$.999
Self test – Try	1.65	$t = 1.205 (178)$.834
Self test – Questionnaire	-2.52	$t = -1.834 (178)$.447
Self test – Random	-0.62	$t = -0.452 (178)$.998
Self test – NAL2	1.93	$t = 1.406 (178)$.723
Try – Questionnaire	-0.86	$t = -0.624 (178)$.989
Try – Random	1.03	$t = 0.754 (178)$.975
Try – NAL2	3.59	$t = 2.614 (178)$.099
Questionnaire – Random	-1.89	$t = -1.368 (178)$.746
Questionnaire – NAL2	-4.45	$t = -3.221 (178)$.019
Random – NAL2	2.55	$t = 1.863 (178)$.428

^aSatterthwaite degrees of freedom. ^bBetter-ear PTA fixed at 33, loudness rating fixed at 5.77, and clarity/pleasant rating fixed at 5.77.

Appendix G

Statistics of Linear Mixed-Effects Models Examining the Effect of Selection Model on Deviation of the Selection Model From National Acoustic Laboratories–Non-Linear 2 (NAL-NL2) Condition in Quiet, Controlled for Participant-Specific Better-Ear Pure-Tone Average (PTA)

Mean and 95% confidence interval (CI) estimation

Selection model	LS means ^a	df ^b	95% CI
Audiogram	0.79	101	[0.43, 1.16]
Self test	0.19	101	[-0.17, 0.56]
Try	0.50	101	[0.14, 0.87]
Questionnaire	0.77	101	[0.40, 1.13]
Random	0.52	101	[0.16, 0.89]

Note. LS = least square.

^aBetter-ear PTA fixed at 33. ^bSatterthwaite degrees of freedom.

Pairwise comparison

Contrast	Difference ^a	Test (df ^b)	p
Audiogram – Questionnaire	0.03	$t = 0.15$ (144)	.9999
Audiogram – Random	0.27	$t = 1.381$ (144)	.6409
Audiogram – Self test	0.60	$t = 3.047$ (144)	.0227
Audiogram – Try	0.29	$t = 1.47$ (144)	.5834
Self test – Try	-0.31	$t = -1.577$ (144)	.5146
Self test – Questionnaire	0.57	$t = 2.897$ (144)	.0347
Self test – Random	0.33	$t = 1.666$ (144)	.458
Try – Questionnaire	0.26	$t = 1.321$ (144)	.6789
Try – Random	0.02	$t = 0.089$ (144)	1.000
Questionnaire – Random	0.24	$t = 1.231$ (144)	.7332

^aBetter-ear PTA fixed at 33. ^bSatterthwaite degrees of freedom.

Appendix H

Statistics of Linear Mixed-Effects Models Examining the Effect of Selection Model on Deviation of the Selection Model From the Condition in Noise, Controlled for Participant-Specific Better-Ear Pure-Tone Average (PTA)

Mean and 95% confidence interval (CI) estimation

Selection model	LS means ^a	df ^b	95% CI
Audiogram	0.78	107	[0.45, 1.12]
Self test	0.60	107	[0.26, 0.93]
Try	0.71	107	[0.38, 1.05]
Questionnaire	0.85	107	[0.51, 1.18]
Random	0.76	107	[0.43, 1.10]

Note. LS = least square.

^aBetter-ear PTA fixed at 33. ^bSatterthwaite degrees of freedom.

Pairwise comparison

Contrast	Difference ^a	Test (df ^b)	p
Audiogram – Questionnaire	-0.06	<i>t</i> = -0.342 (144)	.997
Audiogram – Random	0.02	<i>t</i> = 0.124 (144)	1.000
Audiogram – Self test	0.19	<i>t</i> = 1.005 (144)	.853
Audiogram – Try	0.07	<i>t</i> = 0.378 (144)	.996
Self test – Try	-0.12	<i>t</i> = -0.626 (144)	.971
Self test – Questionnaire	0.25	<i>t</i> = 1.347 (144)	.663
Self test – Random	0.16	<i>t</i> = 0.88 (144)	.904
Try – Questionnaire	0.13	<i>t</i> = 0.72 (144)	.952
Try – Random	0.05	<i>t</i> = 0.254 (144)	.999
Questionnaire – Random	0.09	<i>t</i> = 0.466 (144)	.990

^aBetter-ear PTA fixed at 33. ^bSatterthwaite degrees of freedom.