


Consistency of Hearing Aid Setting Preference in Simulated Real-World Environments: Implications for Trainable Hearing Aids

Trends in Hearing
Volume 24: 1–19
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2331216520933392
journals.sagepub.com/home/tia


Els Walravens^{1,2,3} , Gitte Keidser^{1,2,3}, and Louise Hickson^{1,3}

Abstract

Trainable hearing aids let users fine-tune their hearing aid settings in their own listening environment: Based on consistent user-adjustments and information about the acoustic environment, the trainable aids will change environment-specific settings to the user's preference. A requirement for effective fine-tuning is consistency of preference for similar settings in similar environments. The aim of this study was to evaluate consistency of preference for settings differing in intensity, gain-frequency slope, and directionality when listening in simulated real-world environments and to determine if participants with more consistent preferences could be identified based on profile measures. A total of 52 adults (63–88 years) with hearing varying from normal to a moderate sensorineural hearing loss selected their preferred setting from pairs differing in intensity (3 or 6 dB), gain-frequency slope (± 1.3 or ± 2.7 dB/octave), or directionality (omnidirectional vs. cardioid) in four simulated real-world environments: traffic noise, a monologue in traffic noise at 5 dB signal-to-noise ratio, and a dialogue in café noise at 5 and at 0 dB signal-to-noise ratio. Forced-choice comparisons were made 10 times for each combination of pairs of settings and environment. Participants also completed nine psychoacoustic, cognitive, and personality measures. Consistency of preference, defined by a setting preferred at least 9 out of 10 times, varied across participants. More participants obtained consistent preferences for larger differences between settings and less difficult environments. The profile measures did not predict consistency of preference. Trainable aid users could benefit from counselling to ensure realistic expectations for particular adjustments and listening situations.

Keywords

hearing loss, normally hearing, fine-tuning

Received 23 September 2018; Revised 4 March 2020; accepted 12 March 2020

Changes to hearing aid (HA) settings after the initial fitting, referred to as fine-tuning, are often requested as not everyone is happy with the prescribed response (e.g., Valentine et al., 2011). Fine-tuning can be done by a clinician or by the HA user themselves. An increasing number of HA settings are available to clinicians for adjustment: gain for different frequencies and levels; attack and release times for compression; and features affecting directionality, noise reduction, and speech enhancement. The most commonly adjusted settings in clinical practice are changes in gain for different frequencies and levels, for which there is no consensus but are often based on the experience of the clinician (Anderson et al., 2018; Thielemans et al., 2017). The HA user can perform fine-tuning themselves, for example, using a

trainable algorithm (Dillon et al., 2006). Trainable algorithms use as input acoustical information from the user's listening environment, such as the type of background noise and its level, and the listener's adjustments made to the HA controls in those environments.

¹The HEARing Cooperative Research Centre, Carlton, Victoria, Australia

²National Acoustic Laboratories, Macquarie Park, New South Wales, Australia

³School of Health and Rehabilitation Sciences, The University of Queensland, St Lucia, Queensland, Australia

Corresponding Author:

Els Walravens, WS Audiology, Widex A/S, Nymøllevej 6, 3540 Lyngø, Denmark.

Email: els.walravens@wsa.com



Adjustments can be made using the controls on the HA, if available, or using a remote control or smartphone app. Based on any consistent user-adjustments in the same or similar acoustic environments, the trainable algorithm will modify the HA settings for that listening situation to the user's preference. For example, if the HA user reduces the volume every time they go for a walk in a busy street, the trainable algorithm will over time reduce overall gain for that situation, so that the user will have less need to make adjustments in that situation. Inconsistent adjustments on the other hand will result in settings only marginally changed from the original. For example, a HA user might listen to classical music on the radio set at a particular level, reducing the HA volume some of the time and increasing it at other times depending on the type of music and/or how much they like the piece. If the HA user makes a similar amount of changes in volume in opposite directions and of a similar magnitude, this will result in HA settings essentially unchanged from the original.

For the trainable algorithm to effectively fine-tune the HA settings, adjustments need to be made that result in similar HA settings in similar acoustic environments. This assumes that listeners have a preference for a given HA setting and can select it consistently every time they are in similar listening environments. However, previous research has suggested that hearing-impaired people do not always have a preference when comparing different HA settings, and that consistency of preference varies. In several studies, hearing-impaired people have been tasked with comparing different response shapes (e.g., Byrne & Cotton, 1988), HA prescription procedures (e.g., Keidser & Grant, 2001; Moore & Sek, 2003), or microphone modes (e.g., Walden et al., 2005; Wu, 2010) and asked to indicate their preferred setting for a range of listening conditions. Invariably not all participants demonstrated a preference in all listening conditions. While Moore and Sek (2013) suggested that weak preferences could be related to small perceptual differences between prescriptions compared in their study, the difference between response shapes in Byrne and Cotton (1988) was selected to be perceptually different, and Keidser and Grant (2001) established that perceptual differences between prescriptions compared in their study were "slight to distinct." In all studies, the proportion of preferences tended to vary with both participant and listening condition.

At least three studies have further reported consistency of preference to be variable among hearing-impaired people. Kuk and Pape (1992) used repeated paired comparisons between the National Acoustic Laboratories Revised (NAL-R) prescription and a low- or high-frequency cut to determine hearing-impaired listeners' preferred gain-frequency response slope for clarity for four different stimuli. Dependent on the stimulus,

between 9 and 18 of 20 participants showed inconsistency in preferences across three trials within the same session by selecting different combinations of low- and high-frequency cuts for each trial. Keidser et al. (2005) asked 27 participants with varied audiometric configurations to select their preferred frequency response slope while listening in 20 different listening conditions. Using a parameter adjustment and selection procedure, the response slope was selected adaptively among settings tilting in ± 2 dB steps across low and high frequencies around the prescribed NAL-RP response up to ± 28 dB. For each listening condition, three adaptive trials were completed with the starting point selected at random. In total, 27% of selected responses, for which the intraparticipant standard deviation across trials exceeded 5 dB, were considered inconsistent. Third, using a two-alternative forced-choice task and 10 repetitions, Keidser et al. (2008) asked 12 participants to select a preferred response for six different stimuli among pairs of gain-frequency slopes differing in root-mean-square (rms) value, calculated from gain differences measured at three frequencies, from 1 to 10 dB. They found that, in 77% of cases, participants demonstrated inconsistency in their preference; that is, they selected the same response less than 9 out of 10 times. In all three studies, inconsistent responses could not generally be explained by difficulties discriminating between settings, and the proportion of inconsistent responses varied with the stimulus. Although these studies have provided some insights into the consistency of listeners' preferences, participant numbers were small, linear amplification was applied, stimuli were rather artificial, and only gain-frequency slope differences were evaluated for consistency of preference. This study set out to address these identified shortcomings of previous research.

Another limitation of research in this area is that little information is available about factors that may influence consistency of auditory preference. So far, an association between greater high-frequency hearing loss and more consistent preferences for gain-frequency differences have been observed in both Keidser et al. (2005) and Keidser et al. (2008). Keidser et al. (2008) suggested that this association was due to the narrower audible dynamic range at high frequencies of this population that makes them very sensitive to changes in high-frequency gain. Furthermore, people with greater high-frequency hearing loss have been found to show stronger preference for directionality in a laboratory setting (Wu, 2010). A requirement to consistently select a preference is an ability to discriminate between the available responses. Therefore, it would be predicted that those demonstrating better intensity discrimination would be more sensitive to stimuli varying in overall gain, those with better frequency selectivity would be more sensitive to changes in the frequency characteristics of

stimuli when the gain-frequency response is varied, and those with better temporal resolution would be more sensitive to stimuli in which changing the gain-frequency response shape alters the relative prominence of gaps in noise across frequencies (Lister et al., 2011). Consequently, measures of low- and high-frequency average hearing loss, dynamic range, intensity discrimination, frequency selectivity, and temporal resolution were included as potential predictors of consistency of preference in this study.

There is further some evidence about the likely importance of cognitive factors for demonstrating an auditory preference. For example, Lunner (2003) found that listeners with poorer working memory recall showed a preference for the same HA program when listening in different environments, whereas those with better working memory recall showed a preference for different HA programs. Similarly, listeners with poorer working memory recall showed a preference for the highest degree of noise reduction irrespective of the listening situation, whereas those with better working memory recall preferred different degrees of noise reduction (Neher et al., 2014), though not when adding directionality to noise reduction (Neher, 2014). In addition, participants with better results on an executive function task showed more selective preferences for strong noise reduction when listening using an omnidirectional microphone (Neher, 2014). Accordingly, as measures of working memory and executive function seem to influence preference for HA settings, they were also included in this study. Furthermore, because of the relationship between executive function and preference for HA settings, working memory updating, a component of executive function (Miyake et al., 2000), was included. Working memory updating tracks new and discards unnecessary information, a process that seems inherent to the fine-tuning process when comparing different HA settings. Finally, a measure of personality was included to examine if consistency of preference might be influenced by how the task is approached rather than by underlying psychoacoustic and cognitive abilities.

In summary, this study set out to investigate consistency of preference for HA settings differing in intensity, gain-frequency slope, and directionality, when listening in four simulated real-world environments using nonlinear amplification. The intensity, gain-frequency slope, and directional differences in HA settings were chosen, as they represent some of the most commonly adjusted parameters in clinical practice and are increasingly available to HA users as adjustments. As well as evaluating consistency of preference among hearing-impaired people, its dependency on the environment and HA setting was evaluated. In addition, a range of factors were included that might predict which listeners were more likely to obtain consistent preferences. Psychoacoustic

and cognitive measures were included, along with a personality screening test.

Method

Participants

On the basis of preliminary data, a power analysis was conducted to determine the number of participants. Power calculations were based on the test of the null hypothesis of no effect of environment on consistency of preferences, with environment being a categorical variable with four categories in a mixed-effects logistic regression model for consistency. The power calculations employed a simulation approach, with a significance level of 5% and target power of 80%. Using one environment as a reference category, the power calculations were based on the assumption that relative to the reference environment, the odds ratios for consistency in the other three environments were r , r^2 , and r^3 (equivalent to constant differences on the log-odds scale), respectively, for some specified value of r . For $r = 1.50$, the required sample size for 80% power was approximately 70, and for $r = 1.65$, the required sample size for 80% power was approximately 44. Because of budget and time constraints, a sample size of 70 was not feasible and instead a sample size of 50 was aimed for. The power analysis estimated that for $r = 1.65$, the sample size of 50 would give high power (85%), and for $r = 1.50$, the power would still be moderately high (66%).

In total, 52 adult volunteers (23 women) with an average age of 73 years (63–88 years) participated. While three participants were nonnative English speakers, all were fluent in English. Their hearing was symmetrical, defined as the difference between ears in four-frequency average hearing loss not exceeding 10 dB, and the difference between thresholds at individual frequencies up to 4000 Hz not exceeding 20 dB. Participants were selected to represent a range of degrees of hearing (i.e., average binaural four-frequency average hearing loss from normal to moderate) and slopes of hearing loss (binaural average difference between thresholds across 250, 500, and 1000 Hz (low-frequency average) and across 2, 3, and 4 kHz (high-frequency average), see Figure 1. A total of 35 participants with a four-frequency average hearing loss over 40 dB owned HAs.

Participants were initially assessed using the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005); 41 participants performed within the normal range of the test (score ≥ 26), 10 displayed a mild cognitive impairment ($21 \leq \text{score} < 26$), and 1 participant produced a score of 19. Participants relying on HAs wore their own devices during this paper-and-pen cognitive screening measure to control for any hearing difficulty.

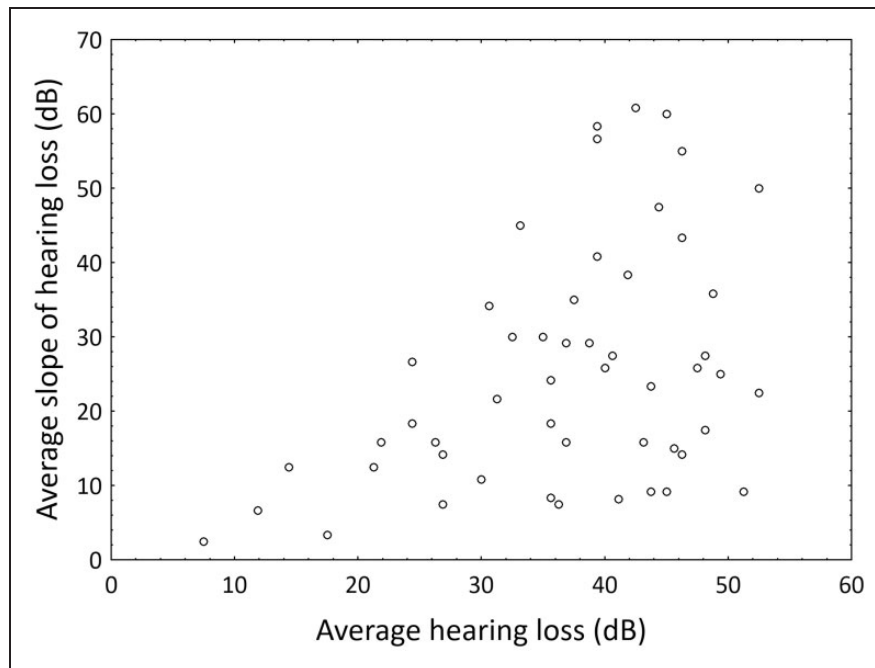


Figure 1. The Spread in Binaural Average Hearing Loss Represented by the Four-Frequency Average Hearing Loss, and the Slope, That Is, the Difference Between the Average Thresholds Across 250, 500, and 1000 Hz (Low-Frequency Average) and Across 2, 3, and 4 kHz (High-Frequency Average).

All participants provided written informed consent. The research was approved by the Australian Hearing Human Research Ethics Committee (AHHREC2016-3) and the Behavioural & Social Sciences Ethical Review Committee of The University of Queensland (2011000857). Participants were offered a small gratuity at the end of their final appointment to offset their transport costs.

Profile Measures

Psychoacoustic Measures

Average Low- and High-Frequency Thresholds. Based on the audiogram, obtained using insert earphones, the low-frequency average (250, 500, and 1000 Hz) and high-frequency average (2000, 3000, and 4000 Hz) were calculated. Measures for both ears were averaged to obtain one measure of each.

The remaining psychoacoustic measures were presented using a computer and included practice trials, except for the comfortable dynamic range measure. Measures were completed using Sennheiser HD 215 headphones unless indicated otherwise.

Intensity Discrimination. Individual ear discrimination thresholds were obtained for 500 and 3000 Hz pure tones of 600 ms duration with reference tones presented at 30 dB Sensation Level (SL), using a three-interval forced-choice task, with participants selecting the interval that

contained the louder pure tone. The step-size of the 1-up 2-down procedure varied adaptively from an initial difference of 4 dB and completed when 71% correct detection was reached. The threshold was calculated as the average of the levels of the last eight reversals at the final step size of 1 dB (Hansen, 2006; Jepsen & Dau, 2011). Each frequency was assessed twice in each ear, and the result was averaged per ear; the better-ear result at each frequency was used for further analysis.

Comfortable Dynamic Range. Using the Contour Test of Loudness Perception (Cox et al., 1997), participants were asked to report which category best described the loudness of a 5-s fragment of a monologue when listening unaided in the sound field. Starting from 35 or 50 dB SPL, depending on the participant's hearing loss, the level was increased in 3-dB steps until the speech fragment was reported to be "loud but ok," or a level of 83 dB SPL was reached. The median level difference between the levels perceived as "comfortable" and "loud but ok" was calculated based on three trials.

Spectral and Temporal Resolution. Ear-specific detection thresholds for pulsed pure tones of 500 and 3000 Hz with a 275 ms duration were obtained using a Békésy technique. Presentation levels were derived from the one-third gain formula as recommended by Athalye (2010). These were varied by 3 dB/s and presented in octave-band noise (a) without gaps, (b) with continuous

half-octave spectral gaps around the test frequency, or (c) with 50 ms temporal gaps. The threshold was defined by the average value of six upper and six lower reversals after two initial turning points. The difference in threshold obtained when listening to pulsed tones in these different conditions quantified the listener's spectral and temporal resolution (Larsby & Arlinger, 1998; van Esch et al., 2013). Each threshold was measured twice and averaged; a third was completed when the difference between the initial pair was 5 dB or more. Any trial with a trace exceeding a range of 20 dB after the first reversal was discarded and repeated. The better-ear result for each of the four conditions (500 and 3000 Hz, spectral and temporal resolution) was used for analysis.

Cognitive Measures. Cognitive measures were presented visually only, using a computer, and practice trials were provided.

Working Memory Recall. The Reading Span test, adapted from Daneman and Carpenter (1980) and Rönnerberg et al. (1989), was used to measure recall in working memory. Sentences were presented in three parts on a computer screen for 800 ms each: the subject, verb, and an object or descriptor. At the end of each sentence, the participant was asked to indicate verbally whether that sentence made sense, and this question was displayed on the screen for 5 s. An interval of 3 s was presented between each sentence. Two sets of three, four, and five sentences were presented. After a set of sentences was presented, participants were asked to recall the first or last words of as many of the sentences as they could. The final score was the percentage words recalled correctly out of a total of 24, independent of order.

Executive Function. The Executive Control subtest of the Test of Attentional Performance—Mobility version (Zimmermann & Fimm, 2014) measured executive function. Participants were presented with letters or numbers shown in red or blue on a computer screen and instructed to push the left button when they saw a red number and the right button when they saw a blue letter as fast as they could, while ignoring the red letters and blue numbers. In total, 80 items were presented with a duration of 0.5 s and an interstimulus interval between 2 and 3 s. The buttons registered responses and reaction time; the median response time of the correct responses was used for further analysis.

Working Memory Updating. The Letter Memory Task was adapted from Morris and Jones (1990) and Miyake et al. (2000). Sequences of 5, 7, 9, or 11 consonants were presented on a computer screen one at a time, in large font. Each letter was presented for 2 s. Blinded to the length of a sequence, participants were asked to

recall the last four letters of 12 trials. The number of letters recalled correctly, independent of order, was used as a measure of working memory updating.

Personality Measure. The Ten-Item Personality Inventory (Gosling et al., 2003) evaluated the “Big Five” personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experiences. Using pen and paper, participants scored 10 statements on a seven-point Likert scale, with two statements for each trait. The average score for each trait was used for further analysis.

Hearing Devices and Test Settings

An in-house real-time master HA was used for this study. The master HA contained microphones and receivers embedded in behind-the-ear shells wired to a sound card and a computer, performing all the signal processing. The HA parameters were manipulated via a GUI, providing 16 independent gain and compression channels, with a center frequency of 62.5 Hz (125 Hz bandwidth), 250, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250 Hz (250 Hz bandwidth), 2625 Hz (500 Hz bandwidth), 3250, 4000 Hz (750 Hz bandwidth), 5000 Hz (1250 Hz bandwidth), 6375 Hz (1500 Hz bandwidth), and 9562.5 Hz (4875 Hz bandwidth). The compression was fast-acting ($t_a = 10$ ms and $t_r = 100$ ms) and matched the NAL-NL2 prescription (Keidser et al., 2011). No other sound processing features were activated for the baseline setting in the master HA, but when feedback was detected, measurements were done to estimate and add a filter to reduce feedback. The behind-the-ear HAs were coupled to participants' ears using HAL-HEN 2602 occluding foam ear tips. Real-ear insertion gain using the International Speech Test Signal (Holube et al., 2010) as input was used to adjust the HA gain to match targets, with participants with normal hearing to minimal loss all fitted to a 25 dB HL loss across all frequencies. A minimum amplification of 5 dB (measured by insertion gain) was provided at any frequency with targets below this level, to ensure amplification dominated the signal so differences in HA settings (see later) could be achieved. A monologue was presented at 60 dBA to ensure the amplification was comfortable to the participant. Using the Contour Test of Loudness Perception scale (Cox et al., 1997), the overall gain was adjusted until the participant indicated the setting was “comfortable,” or “comfortable but slightly loud” for those with normal hearing and minimal hearing loss. Both the minimum amplification and listening comfort criteria were met for all participants. Although the NAL-NL2 target was used to set the HA gain, the adjustments made to meet the aforementioned criteria could modify the participant's baseline response from the prescription.

Table 1. Description of the Different Comparison Pairs, Including Their Variation From the Baseline and Root-Mean-Square Difference.

Comparison pairs	Difference from baseline		Measured rms difference mean (SD)
	Response 1	Response 2	
Directionality		Cardioid	1.2 (0.5)
Intensity—large difference	+2 dB	−4 dB	5.6 (0.5)
Intensity—small difference	+1 dB	−2 dB	2.9 (0.3)
Gain-frequency slope—large difference	+2.7 dB/oct	−2.7 dB/oct	5.6 (0.6)
Gain-frequency slope—small difference	+1.3 dB/oct	−1.3 dB/oct	2.9 (0.3)

rms = root-mean-square.

Based on the participant's baseline response, five pairs of HA settings were created, differing in directionality, intensity, or gain-frequency slope (Table 1). The directionality pair was composed of the omnidirectional baseline and a fixed cardioid microphone response with a Directivity Index of 5.4 dB (measured using white noise presented from all loud speakers with the master HA positioned on a Head and Torso Simulator). The cardioid setting had the same gain-frequency response as the omnidirectional baseline setting at 0° azimuth (i.e., compensating for the low-frequency roll-off). Two pairs differing in intensity were created by changing the overall level of the baseline response to create a 6 dB (+2 dB and −4 dB from baseline) and 3 dB (+1 dB and −2 dB from baseline) overall gain difference. Pairs differing in gain-frequency slope had an overall loudness presumed equal to that of the baseline response, but different slopes, created by increasing the gain at 500 Hz by 4 or 2 dB and decreasing the gain by a similar amount at 4000 Hz and vice versa, using 1500 Hz as the cross-over frequency, resulting in a slope of ± 2.7 dB/oct or ± 1.3 dB/oct compared with the baseline response, see Table 1. Using gain differences obtained at 1/24 octave frequencies from 250 to 6000 Hz, the implemented differences in slope resulted in rms differences between the pairs of gain-frequency responses of 6 and 3 dB. An rms difference of 6 and 3 dB between intensity and gain-frequency responses was chosen to represent differences that have been demonstrated to be discernible in both dynamic signals and complex tones by hearing-impaired people (Caswell-Midwinter & Whitmer, 2019; Keidser et al., 2008; Lentz & Leek, 2003), but not necessarily large enough to reveal an actual preference for or benefit from one response over the other (Keidser et al., 2008; McShefferty et al., 2015). We also note that these differences exceed the step sizes typically implemented in user controls and often used during fine-tuning procedures.

Equipment and Stimuli

The listening environments were presented in a horizontal ring (radius of 1.2 m) of 16 Genelec 8020 C

loudspeakers, situated in a test booth with a reverberation time of 0.3 s. The loudspeakers, spaced uniformly at 22.5° intervals, were driven by an RME Fireface UFX interface (44.1 kHz output) and two ADI-8 DS digital-to-analogue converters.

Two target and two noise recordings were combined to create four listening environments: traffic noise (termed *Traf*); a monologue in traffic noise at 5 dB signal-to-noise ratio (SNR; *MonTraf5dB*); and a dialogue in café noise at both 5 dB (*DiaCafe5dB*) and at 0 dB SNR (*DiaCafe0dB*). The target recordings were two monologues and two dialogues from the NAL Dynamic Conversations Test described in Best et al. (2016). This material is considered to approximate natural speech as talkers were instructed to play out transcripts rather than read them out loud, so it contained variations in speed, pauses, dysfluencies, and interjections. The two monologues were by a female speaker, and the two dialogues were between a male and a different female speaker. Each chosen passage was about 5 minutes, resulting in almost 10 minutes of continuous speech for both the monologue and dialogue. The monologue was presented from 0° azimuth, the dialogue with the two talkers spatially separated at +22.5° and −22.5° azimuth.

The background noises were recordings of real-life acoustic environments obtained using a three-dimensional 62-channel hard-sphere microphone array built in-house. The recorded signals were transformed into loudspeaker signals using the higher order Ambisonics method (Oreinos, 2015; Oreinos & Buchholz, 2016). Only the horizontal components were taken into account (up to an Ambisonics order of $N=7$) in the sound reproduction process (e.g., Oreinos, 2015), which has been shown to be adequate for HA settings for sounds arriving from the horizontal plane (Oreinos & Buchholz, 2015). Recorded sounds arriving from above or below were reproduced with a decreased spatial resolution.

Measured at the center of the array using a Brüel & Kjær sound level meter with a model 4166 microphone, the traffic noise, coming from all 16 loudspeakers, was presented at 67.3 dBA, and the café noise at 67.6 dBA

long-term average. Speech was presented at 5 dB SNR for both the MonTraf5dB and DiaCafe5dB environments, based on the SNRs regularly experienced by HA users in Smeds et al. (2015). The dialogue in cafeteria noise was also presented at 0 dB SNR, reflecting the SNR experienced by normal-hearing researchers in Pearsons et al. (1977), assumed to be a rather challenging environment for those with a hearing loss. The third-octave band levels of the speech and noise stimuli across the four listening environments are shown in Figure 2.

The average Speech Intelligibility Index (American National Standards Institute, 1998) measured across participants' intensity and gain-frequency slope responses was used as an indication of the difficulty of listening in that environment. The mean Speech Intelligibility Index across HA settings was 0.55 ($SD=0.03$), 0.50 ($SD=0.04$), and 0.36 ($SD=0.03$) for the MonTraf5dB, DiaCafe5dB, and DiaCafe0dB environments, respectively. The Traf environment was considered the easiest environment, as it was a relatively steady sound with no speech target present.

Procedure

Each participant attended three appointments ranging in duration from 1 to 2.5 hr. All participants were offered a break mid-way through each appointment. Participants

completed their final appointment, on average, 30 days after the first (range = 3–146 days). Across the three appointments, participants completed nine assessments (as described earlier) and the preference tasks. All written instructions were provided in large print, and those written specifically for this project did not exceed Flesch-Kincaid Grade 6 reading level (Kincaid et al., 1975).

Using a two-alternative forced-choice paradigm, participants selected their preference between five pairs of HA settings differing in directionality (one pair), intensity (two pairs), or gain-frequency slope (two pairs, Table 1). Preference measures for intensity and gain-frequency comparisons were completed across all four listening environments. In total, 19 measures were completed as preference for directionality was not evaluated in the Traf environment, as the difference between the omnidirectional and directional setting was considered to most noticeably be a small overall level difference, already assessed in the intensity condition. At the start of the first preference task of every appointment, participants were provided with written instructions (see Appendix) and advised, "Your task is to choose which setting you would prefer [. . .] for listening to each situation," and any questions were addressed. Using a small keypad of which three buttons were labelled "A," "B," and "VOTE," participants listened to setting A first, would then push B and listen to B; they could go

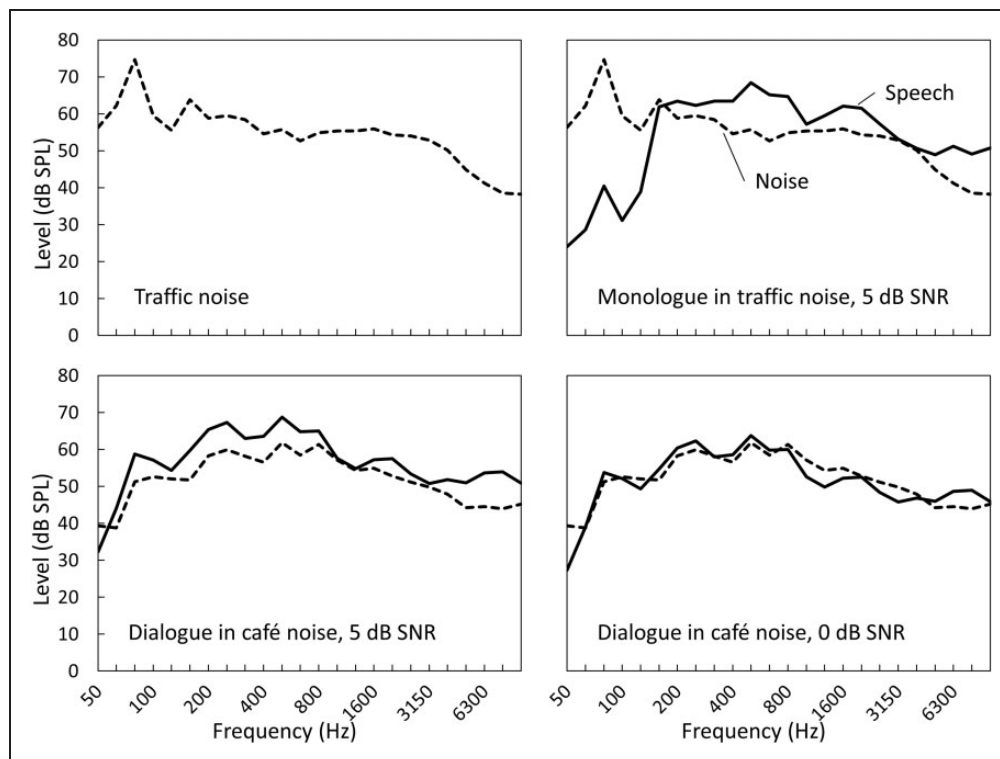


Figure 2. The Levels of Speech (Full Line) and Noise (Dashed Line) Across the Four Listening Environments in Third Octave Bands (dB SPL Long-Term Average): Traffic Noise, Monologue in Traffic Noise at 5 dB SNR, Dialogue in Café Noise at 5 dB SNR and at 0 dB SNR.

back and forth as often as they liked and listen for as long as they liked. They were instructed to ensure they were listening to their preferred setting and then to press "VOTE" for their preference to be registered. As soon as they pressed "VOTE," the next comparison would start with A. This process would be repeated until they had completed all comparisons. Both the environments and the pairs of settings were presented in a randomized order, and the recordings were looped so participants could listen for as long as they liked. Participants selected their preference between each pair 10 times in each of the four environments, with settings for each presentation randomly assigned to the A and B buttons. Participants were advised which listening environment would be presented, including, if applicable, the number of talkers and where they were located. This was done to avoid participants waiting for speech signals when none would be presented or spending time to try to localize the talkers.

The preference task was automated so that the duration of each vote was recorded as the duration from the start of the first presentation of the "A" stimulus of the pair until the participant pressed "VOTE." Any trial where the participant selected "VOTE" accidentally before changing to "B" was repeated at the completion of that preference task. Such accidental votes occurred in 1.4% of comparisons and were spread across 33 participants. When asked, all but two participants reported being able to follow the target speech during the preference task: one when listening to the MonTraf5dB and one participant when listening to the DiaCafe0dB environment. Participants completed the preference task for one environment in the first and final appointment and two environments in the second appointment, except for two participants completing two preference tasks in the second and final appointment.

Analysis

Analyses of data that were not normally distributed were completed using the Mann–Whitney U test (group differences) or the Friedman's ANOVA (repeated measures). A mixed-effects logistic regression analysis was conducted to evaluate the difference between the number of participants with a consistent preference across the 19 different conditions, with the environment, the difference between HA settings, and their interaction as fixed effects and a subject-specific intercept as the random effect. In addition, pairwise comparisons were completed to further investigate the influence of the environment and the difference between HA settings on consistency of preference, quantified using the odds ratio (OR) and 95% confidence interval (CI). The p values were adjusted for multiple comparisons using a simultaneous inference procedure (Hothorn et al., 2008).

Before investigating what profile measures may predict the number of consistent preferences for large and small differences in HA settings, missing values of measures were filled and measures with nonnormal distribution were transformed. Three participants could not complete the task to measure their spectral and temporal resolution at 3000 Hz in either ear as the reference level was inaudible to them, and five participants were unable to complete the intensity discrimination task at 3000 Hz in either ear as the reference level of 30 dB SL created loudness tolerance problems. For both spectral and temporal measures, the hearing profile of participants with missing values were most closely associated with the poorest average high-frequency hearing losses, and thence missing values were substituted with the highest thresholds measured with that test. The same approach was used for the missing values of the intensity discrimination task at 3000 Hz. Variables that displayed a non-normal distribution were transformed to improve linearity as assessed using the Shapiro-Wilk test: The intensity discrimination at 500 Hz and 3000 Hz were transformed by using the logarithmic value. Next, the profile measures were used as independent variables in multiple regression analyses to evaluate the influence of the profile measures (separately for the psychoacoustic, cognitive, and personality measures) on the number of consistent preferences for large and small differences obtained by each participant across all environments.

Results

Consistency of Preferences

The main aim of this study was to evaluate consistency of preference for five pairs of HA settings differing in intensity (two pairs: 3 dB, small or 6 dB, large, difference in overall gain), gain-frequency slope (two pairs: ± 1.3 dB/oct, small, or ± 2.7 dB/oct, large), or directionality (one pair: omnidirectional vs. cardioid), when listening in varied simulated real-world environments. If the participant selected the same setting of a pair 9 or 10 out of 10 times, the choice of setting was considered consistent. This criterion was arbitrary and chosen to be high as clinicians often perform fine-tuning based on a single comparison, relying on a high level of consistency. Across HA settings and environments, a total of 19 measures of consistency of preference were obtained for each participant. The number of consistent preferences across participants was varied, ranging from two participants with three consistent preferences to three with 17 consistent preferences (mean and median = 11).

To examine the effect of the magnitude of change to the HA setting, Figure 3 shows the distribution of the number of participants obtaining consistent preferences

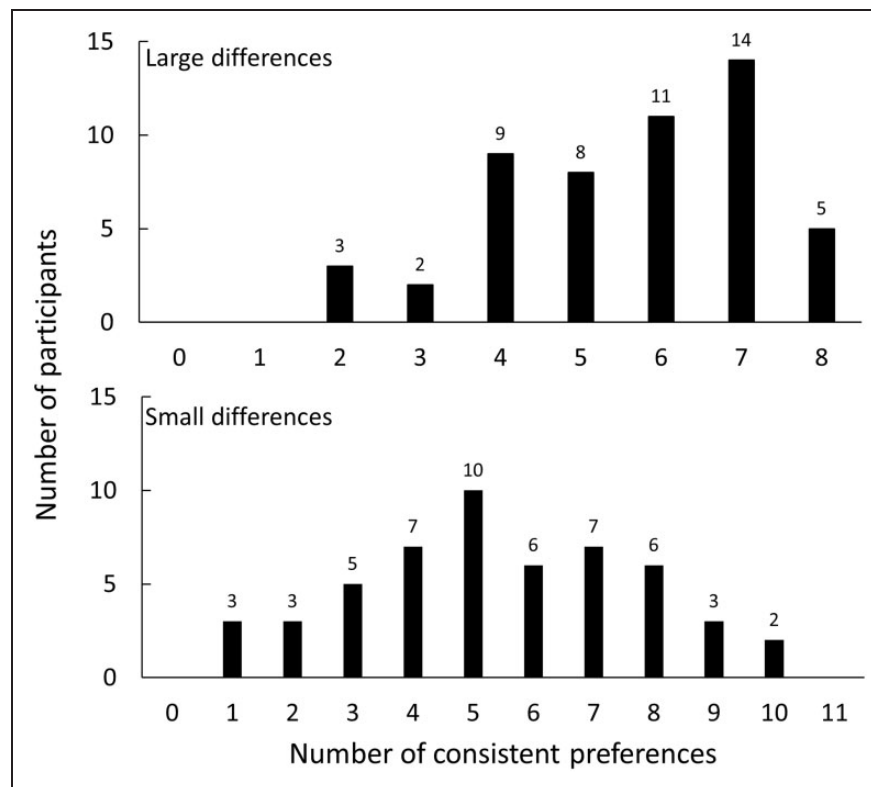


Figure 3. The Number of Consistent Preferences Obtained by Participants ($n = 52$) Across the Large Intensity and Gain-Frequency Slope Differences (Top) and Small Directionality, Intensity, and Gain-Frequency Slope Differences (Bottom) in the Different Listening Environments.

separately for large (top) and small (bottom) differences in stimuli. This distinction highlights the potential influence of discrimination on consistency of preference, with large differences considered to be more easily discriminated than small differences. For this purpose, the difference between microphone modes was considered a small difference, resulting in a maximum of 8 and 11 consistent preferences for large and small differences, respectively. While 30 participants had a consistent preference for 80% or more of the stimuli with a large difference, only 5 participants had consistent preferences for 80% or more of the stimuli with a small difference.

Results for participants who did not pass the MoCA screening measure, for those who did not have English as a first language but were fluent in English, and for those who did not own HAs were examined in more detail. There was no significant difference in the number of consistent preferences for participants with a MoCA score outside the normal range ($n = 11$; median = 12) and within the normal range ($n = 41$; median = 11; $U = 166$; $p = .2$). Of the three participants who did not have English as a first language, all obtained MoCA scores within the normal range, and one had a consistent preference for 14 out of 19 conditions and the

other two obtained six consistent preferences each. In addition, there was no significant difference in the number of consistent preferences between those who owned ($n = 35$; median = 11) and did not own ($n = 17$; median = 11; $U = 234.5$, $p = .2$) HAs. Thus, there was no evidence of mild cognitive impairment, language ability, or HA use impacting on the preference task, and hence, all data were included in further analyses.

Figure 4 shows the number of participants with a consistent preference for each of the five pairs of HA settings across the four environments ranked by increasing difficulty. Although the consistency of preference was variable, trends are visible across environments and HA settings. The proportion of consistent preferences decreased as the difficulty of the environment increased, with a consistent preference for, on average, 78% of the comparisons in traffic noise, decreasing systematically to 38% in the dialogue in café noise at 0 dB SNR. Participants obtained more consistent preferences for HA settings differing in intensity (on average 65%), than gain-frequency slope (58%) and directionality (40%). The finding that the number of consistent preferences decreased with an increase in the difficulty of the environment was also visible in the average time

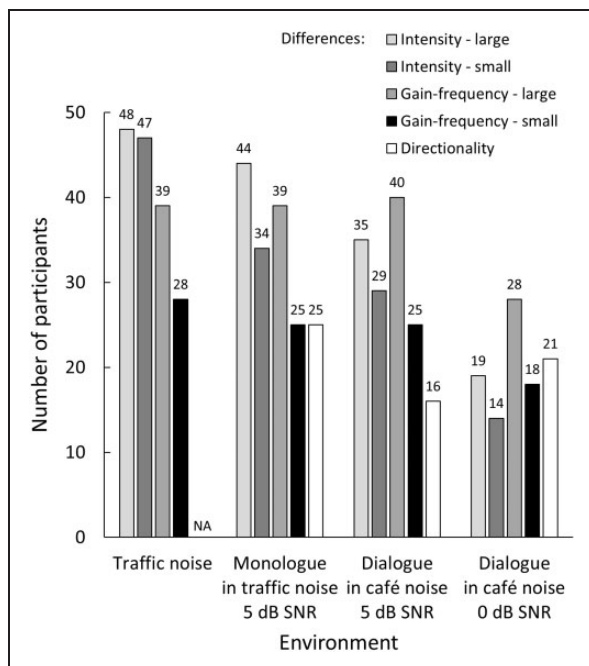


Figure 4. The Number of Participants With a Consistent Preference for the Five Pairs of HA Settings Across the Four Listening Environments From Easiest to Most Difficult: Traffic Noise, Monologue in Traffic Noise, Dialogue in Café Noise at 5 dB SNR and 0 dB SNR.

participants took to complete the preference trials. Across HA settings, the average duration for each vote increased significantly with increasing difficulty of the environment (Friedman's ANOVAs for all HA settings $p < .0001$), from 14.7 s ($SD = 9.8$) for a trial in Traf to 17.3 s ($SD = 10.1$) for the MonTraf5dB, to 21.9 s ($SD = 11.5$) for the DiaCafe5dB environment, reaching 24.5 s ($SD = 12.1$) for the DiaCafe0dB environment.

The variability in the number of consistent preferences each participant obtained also extended to the distribution of consistent preferences across the different environments and differences between the HA settings. For example, the number of participants with a consistent preference for large gain-frequency differences was similar, with 39, 39, and 40 participants showing consistent preferences for the Traf, MonTraf5dB, and DiaCafe5dB environments, respectively (Figure 4). However, only 24 of those participants had a consistent preference for large gain-frequency differences across all three environments.

A mixed-effects logistic regression showed that consistency of preference depended on both the environment and the difference between HA settings, with both main effects and their interaction being statistically significant ($p < .001$). Comparisons between the different conditions are listed in Tables 2A to E and 3A to D, showing the influence of environment (across the differences between HA settings) and difference between HA settings (across the different

environments), respectively. For example, Table 2A shows that the odds ratio (OR) associated with a consistent preference for large intensity differences in the Traf compared with the DiaCafe5dB environment was estimated to be 7.20 (CI [1.05, 49.5]), meaning that participants were more likely to have a consistent preference for large intensity differences when listening to the Traf environment than when listening to DiaCafe5dB environment ($p = .04$).

Only when comparing HA settings differing in intensity were participants significantly more likely to have a consistent preference between different environments, with an OR ranging from 4.25 (CI [1.05, 17.20] to 41.90, CI [6.6, 265.91]) see Table 2A and B). The largest significant OR was for a consistent preference for small intensity differences in the Traf compared with the DiaCafe0dB environment ($p < .001$).

Although the difference between HA settings also had a significant influence on consistency of preference, fewer comparisons reached significance across environments compared with the influence of the environment (Table 3A to D). Significant ORs ranged from 4.48 (CI [1.03, 19.42]; a consistent preference was more likely for large than small gain-frequency slope differences listening to the DiaCafe5dB environment) to 10.63 (CI [2.32, 48.68]; a consistent preference for large gain-frequency slope differences was more likely than for directionality differences in DiaCafe5dB environment). No significant difference in probability was measured between any of the HA settings when listening in the most difficult DiaCafe0dB environment.

The interaction between environment and difference between HA settings is visible in the different patterns of consistent preferences for the different HA settings across environments (Figure 4). Both the patterns for intensity and gain-frequency slope differences show a reduction in the number of consistent preferences with increasing difficulty of environment. While the reduction of preferences is systematic for the intensity pairs, it is similar across the three least difficult environments for the gain-frequency slope pairs, before dropping in the dialogue in café noise at 0 dB SNR. As shown in Table 2A to E, participants were significantly more likely to obtain a consistent preference for intensity differences when listening in less rather than more difficult environments, while no differences in consistent preferences reached significance between any environments for large or small gain-frequency slope differences. The interaction between environment and HA setting was different for directionality: The number of participants with a consistent preference dropped from the MonTraf5dB to the DiaCafe5dB environment but increased from the DiaCafe5dB to the DiaCafe0dB environment (comparisons not reaching significance; Figure 4).

Table 2. Pairwise Comparisons With the Odds Ratio (OR) and the 95% Confidence Interval (CI) Quantifying the Influence of Environment on Consistency of Preference for (A) Large Intensity Differences, (B) Small Intensity Differences, (C): Large Gain-Frequency Response Differences, (D) Small Gain-Frequency Response Differences, and (E) Directionality.

		Traf	MonTraf5dB	DiaCafe5dB	DiaCafe0dB
A	Traf	OR [95% CI]	2.34 [0.30, 18.59]	7.20 [1.05, 49.53]	32.38 [4.70, 223.33]
		<i>p</i>	.98	.04	<.001
	MonTraf5dB	OR [95% CI]	0.43 [0.05, 3.39]	3.07 [0.63, 15.05]	13.82 [2.83, 67.63]
		<i>p</i>	.98	.47	<.001
	DiaCafe5dB	OR [95% CI]	0.14 [0.02, 0.96]	0.33 [0.07, 1.59]	4.50 [1.13, 17.86]
	<i>p</i>	.04	.47	.02	
	DiaCafe0dB	OR [95% CI]	0.07 [0.01, 0.35]	0.22 [0.06, 0.88]	
		<i>p</i>	<.001	.02	
B	Traf	OR [95% CI]	6.12 [1.02, 36.82]	9.85 [1.66, 58.35]	41.90 [6.6, 265.91]
		<i>p</i>	<.05	.002	<.001
	MonTraf5dB	OR [95% CI]	0.16 [0.03, 0.98]	1.61 [0.42, 6.19]	6.85 [1.64, 28.50]
		<i>p</i>	<.05	1.00	.001
	DiaCafe5dB	OR [95% CI]	0.10 [0.02, 0.60]	0.62 [0.16, 2.39]	4.25 [1.05, 17.20]
	<i>p</i>	.002	1.00	.04	
	DiaCafe0dB	OR [95% CI]	0.15 [0.04, 0.61]	0.24 [0.06, 0.95]	
		<i>p</i>	.001	.04	
C	Traf	OR [95% CI]	1.00 [0.23, 4.44]	0.89 [0.20, 4.01]	3.02 [0.74, 12.36]
		<i>p</i>	1.00	1.00	.30
	MonTraf5dB	OR [95% CI]	1.00 [0.23, 4.44]	0.89 [0.20, 4.01]	3.02 [0.74, 12.36]
		<i>p</i>	1.00	1.00	.30
	DiaCafe5dB	OR [95% CI]	1.13 [0.25, 5.11]	1.13 [0.25, 5.11]	3.41 [0.82, 14.24]
	<i>p</i>	1.00	1.00	.18	
	DiaCafe0dB	OR [95% CI]	0.33 [0.08, 1.36]	0.29 [0.07, 1.23]	
		<i>p</i>	.30	.18	
D	Traf	OR [95% CI]	1.32 [0.35, 4.89]	1.31 [0.35, 4.89]	2.54 [0.66, 9.77]
		<i>p</i>	1.00	1.00	.50
	MonTraf5dB	OR [95% CI]	0.76 [0.20, 2.83]	1.00 [0.27, 3.72]	1.93 [0.50, 7.41]
		<i>p</i>	1.00	1.00	.92
	DiaCafe5dB	OR [95% CI]	0.76 [0.20, 2.83]	1.00 [0.27, 3.72]	1.93 [0.50, 7.41]
	<i>p</i>	1.00	1.00	.92	
	DiaCafe0dB	OR [95% CI]	0.39 [0.10, 1.51]	0.52 [0.13, 1.98]	
		<i>p</i>	.50	.92	
E	MonTraf5dB	OR [95% CI]		2.37 [0.61, 9.28]	1.45 [0.39, 5.44]
		<i>p</i>		.65	1.00
	DiaCafe5dB	OR [95% CI]	0.42 [0.11, 1.65]		0.61 [0.15, 2.41]
	<i>p</i>	.65		1.00	
	DiaCafe0dB	OR [95% CI]	0.69 [0.18, 2.60]	1.64 [0.42, 6.47]	
		<i>p</i>	1.00	1.00	

Note. The pairwise comparisons reaching significance are shown in bold. SNR = signal-to-noise ratio; Traf = traffic noise; MonTraf5dB = monologue in traffic noise at 5 dB SNR; DiaCafe5dB = dialogue in café noise at 5 dB SNR; DiaCafe0dB = dialogue in café noise at 0 dB SNR.

Relationship Between Profile Measures and Consistency of Preferences

To evaluate the influence the profile measures had on the number of consistent preferences obtained by each

participant across all environments, multiple regression analyses were conducted separately for the HA settings with large and small differences to distinguish between conditions that are considered more easy or more difficult to discriminate, respectively. A separate analysis

Table 3. Pairwise Comparisons With the Odds Ratio (OR) and the 95% Confidence Interval (CI) Quantifying the Influence of the Difference Between HA Settings on Consistency of Preference for (A) Traffic Noise, (B) Monologue in Traffic Noise, (C) Dialogue in Café Noise at 5 dB SNR, and (D) Dialogue in Café Noise at 0 dB SNR.

		Intensity, large	Intensity, small	Gain-frequency, large	Gain-frequency, small	Directionality
A	Intensity, large	OR [95% CI]	1.30 [0.13, 12.85]	4.65 [0.62, 34.85]	14.05 [1.99, 99.31]	
		<i>p</i>	1.00	.36	<.001	
	Intensity, small	OR [95% CI]	0.77 [0.08, 7.60]	3.58 [0.54, 23.55]	10.80 [1.75, 66.80]	
		<i>p</i>	1.00	.57	.001	
Gain-frequency, large	OR [95% CI]	0.21 [0.03, 1.61]	0.28 [0.04, 1.84]		3.02 [0.71, 12.80]	
	<i>p</i>	.36	.57		.35	
Gain-frequency, small	OR [95% CI]	0.071 [0.01, 0.50]	0.093 [0.01, 0.57]	0.33 [0.08, 1.40]		
	<i>p</i>	<.001	.001	.35		
B	Intensity, large	OR [95% CI]	3.40 [0.67, 17.19]	1.99 [0.37, 10.59]	7.89 [1.59, 39.14]	7.89 [1.59, 39.14]
		<i>p</i>	.38	.99	.002	.002
	Intensity, small	OR [95% CI]	0.29 [0.06, 1.49]	0.58 [0.13, 2.54]	2.32 [0.59, 9.20]	2.32 [0.59, 9.20]
		<i>p</i>	.38	1.00	.73	.73
Gain-frequency, large	OR [95% CI]	0.50 [0.09, 2.69]	1.71 [0.39, 7.43]		3.97 [0.94, 16.85]	
	<i>p</i>	.99	1.00		.08	.08
Gain-frequency, small	OR [95% CI]	0.13 [0.03, 0.63]	0.43 [0.11, 1.71]	0.25 [0.06, 1.07]		1.00 [0.26, 3.84]
	<i>p</i>	.002	.73	.08		1.00
Directionality	OR [95% CI]	0.13 [0.03, 0.63]	0.43 [0.11, 1.71]	0.25 [0.06, 1.07]	1.00 [0.26, 3.84]	
	<i>p</i>	.002	.73	.08	1.00	
C	Intensity, large	OR [95% CI]	1.78 [0.44, 7.14]	0.57 [0.13, 2.56]	2.57 [0.64, 10.28]	6.09 [1.44, 25.78]
		<i>p</i>	.99	1.00	.56	.003
	Intensity, small	OR [95% CI]	0.56 [0.14, 2.25]	0.32 [0.07, 1.39]	1.44 [0.37, 5.56]	3.42 [0.84, 13.92]
		<i>p</i>	.99	.33	1.00	.16
Gain-frequency, large	OR [95% CI]	1.75 [0.39, 7.80]	3.11 [0.72, 13.48]		4.48 [1.03, 19.42]	10.63 [2.32, 48.68]
	<i>p</i>	1.00	.33		.04	<.001
Gain-frequency, small	OR [95% CI]	0.39 [0.10, 1.56]	0.69 [0.18, 2.67]	0.22 [0.05, 0.97]		2.37 [0.59, 9.60]
	<i>p</i>	.56	1.00	.04		.72
Directionality	OR [95% CI]	0.16 [0.04, 0.69]	0.29 [0.07, 1.19]	0.09 [0.02, 0.43]	0.42 [0.10, 1.70]	
	<i>p</i>	.003	.16	<.001	.72	
D	Intensity, large	OR [95% CI]	1.68 [0.40, 7.14]	0.43 [0.11, 1.71]	1.10 [0.27, 4.46]	0.83 [0.21, 3.28]
		<i>p</i>	1.00	.74	1.00	1.00
	Intensity, small	OR [95% CI]	0.59 [0.14, 2.52]	0.26 [0.06, 1.08]	0.66 [0.15, 2.80]	0.49 [0.12, 2.06]
		<i>p</i>	1.00	.08	1.00	.93
Gain-frequency, large	OR [95% CI]	2.30 [0.58, 9.08]	3.88 [0.93, 16.20]		2.54 [0.64, 10.11]	1.90 [0.49, 7.41]
	<i>p</i>	.74	.08		.57	.95
Gain-frequency, small	OR [95% CI]	0.91 [0.22, 3.67]	1.53 [0.36, 6.51]	0.39 [0.10, 1.56]		0.75 [0.19, 2.99]
	<i>p</i>	1.00	1.00	.57		1.00
Directionality	OR [95% CI]	1.21 [0.31, 4.81]	2.04 [0.49, 8.55]	0.53 [0.14, 2.04]	1.34 [0.33, 5.35]	
	<i>p</i>	1.00	.93	.95	1.00	

Note. The pairwise comparisons reaching significance are shown in bold.

was further conducted for the psychoacoustic (average low- and high-frequency thresholds, comfortable dynamic range, and intensity discrimination and spectral and temporal resolution at 500 and 3000 Hz), cognitive (working memory recall, executive function, and working memory updating), and personality measures (extraversion, agreeableness, conscientiousness, emotional stability, and openness to experiences). The analyses revealed no significant model, suggesting that none of the psychoacoustic (large differences: $F(9, 42) = 0.89$; $p = 0.54$; small differences: $F(9, 42) = 0.60$; $p = 0.79$), cognitive (large differences: $F(3, 48) = 1.09$; $p = 0.36$; small differences: $F(3, 48) = 1.22$; $p = 0.31$) or personality measures (large differences: $F(5, 46) = 1.65$; $p = 0.17$; small differences: $F(5, 46) = 1.30$; $p = 0.28$) could significantly predict the number of consistent preferences across environments.

Discussion

The prevalence of consistent auditory preferences of adults with normal hearing to a moderate sensorineural hearing loss and different audiogram configurations was found to be variable across participants and dependent on the listening environment, the difference between HA settings, and their interaction. Participants gave more consistent preferences for large rather than small differences in HA settings, and in less difficult listening environments. However, this tendency differed across the environments, and these overall results were not systematically reflected at the individual level. None of the included psychoacoustic, cognitive, and personality profile measures could predict consistency of preference for large or small differences in HA settings.

Consistency of preference was variable across participants, with the total number of consistent preferences ranging from 3 to 17 out of 19. This finding is in agreement with Kuk and Pape (1992), Keidser et al. (2005), and Keidser et al. (2008). The variability across participants highlights two of the necessary components for showing consistency of preference: discriminating between settings (noticing a difference) and having a preference (one setting is clearly preferred over the other). First, the influence of discrimination ability on consistency of preference is evident in the finding that a greater number of participants obtained a consistent preference when listening to settings with large rather than small differences (Figure 3), which was also observed by Keidser et al. (2008) for gain-frequency slope differences. Second, the influence of discrimination on consistency of preference was not uniform indicating that not only is a noticeable difference between HA settings or discrimination needed but also the difference has to be meaningful to the participants in order for them to select a preference. This is visible, for example, in the finding that 90% of participants were able to

discriminate between responses with an rms difference of 3 dB as they had obtained a consistent preference for small intensity differences for at least one environment (Figure 4). However, consistency of preference did not extend to small gain-frequency differences with the same rms difference in the same environment, for which only 54% of participants obtained a consistent preference. This latter finding supports Keidser et al. (2008) who found that 10 of 12 participants with a hearing loss were able to indicate an increasing perceptual difference with increasing rms difference (from 1 to 10 dB) between gain-frequency responses. However, only three of the 10 participants could consistently select a preferred response for most listening conditions, including some differing by less than 3 dB. Similarly, results from McShefferty et al. (2016) allude to this distinction between discrimination and preference with listeners selecting different SNR changes for different purposes. In that study, participants were not asked to indicate if they noticed a difference, but whether the second sentence of a pair was better, the same, or worse than the first; that is, participants had to be able to discriminate and if they did, evaluate if this had the effect of being better, the same or worse. A mean SNR difference of 3 dB was needed to complete the task; however, a higher SNR difference of 6 to 8 dB was needed for participants to take action to obtain an SNR improvement by going to the clinic for a change in SNR (McShefferty et al., 2016). The findings from McShefferty et al. suggest that although participants could likely discriminate between the intensity and gain-frequency slope rms differences of 3 and 6 dB in the paired-comparison task, some may not have had a preference. Presumably, those without a preference for some or all comparisons simply found a large range of HA settings acceptable and thus have less need for fine-tuning.

The complexity of the cognitive processes involved in obtaining a consistent preference could be another potential reason for the variation in consistent preferences. In each environment, participants not only had to discriminate between the pairs of HA settings but also establish a criterion for their preference, and apply one or more criteria across the different comparisons when pairs of HA settings were presented in a randomized order. A change of criterion used to select a preference (e.g., naturalness, ease of understanding) within the same environment may influence the participant's preference (Keidser, 1995), and consequently the consistency of their responses. In view of the number of preferences to be completed and the unlimited time provided, it is also possible some participants lost motivation or changed their self-chosen criterion part-way through due to boredom and/or fatigue (De Beuckelaer et al., 2013). However, the consistency of preference of participants who did not pass the MoCA screening measure,

or those who were nonnative speakers of English, did not stand out from other participants, suggesting the cognitive processes necessary to obtain consistent preferences are unaffected by such characteristics.

As previously reported by, for example, Kuk and Pape (1992) and Keidser et al. (2008), consistency of preference was dependent on the environment as shown by the results of the mixed-effects logistic regression analysis. The dependency of consistency of preference on the environment suggests an influence of the degree of difficulty of the environment. As the environment got more difficult in terms of accessing speech, it also became more difficult as the number of target talkers increased, the SNR became poorer, and the noise more fluctuating. For example, more fluctuation was present in the café noise, which comprised multiple speech signals, than in the traffic noise. As speech is a very dynamic signal, the SNR will also fluctuate over time, with greater changes possible from moment to moment in the more fluctuating café noise (e.g., Bentler & Chiou, 2006; Edwards et al., 1998). In a given trial when switching back and forth between settings, the preferred setting could depend on the actual SNRs and the quality of the target voice heard in each setting. If these factors change between settings across trials, then that could influence the participant's ability to obtain consistent preferences. It would be expected that numerous real-world listening situations, especially those containing speech-in-speech, would contain a similar variation, which means that it is potentially very challenging to select a consistent preference.

Overall, our results suggest obtaining a consistent preference for intensity differences was easier than for gain-frequency and directionality differences. Support for this finding can be found in the study by Keidser et al. (2008), in which participants were asked to adjust the volume and gain-frequency slope of a response to reach a preferred setting. Keidser et al. found that more participants made changes to overall gain than to the slope of the response, suggesting reaching a preferred volume level may be easier than a preferred gain-frequency slope. The low number of consistent preferences for HA settings that varied in directionality compared with intensity and gain-frequency slope differences in this study was expected because of the smaller perceptual difference between the settings. Both directionality settings had the same gain-frequency response for targets presented at 0° azimuth (Table 1).

For 85 comparisons, participants obtained a consistent preference for large differences only. For 37 of these comparisons, participants had selected the same preference for the corresponding small difference eight times, one vote short of being consistent. Comparisons where participants obtained a consistent preference for the small difference only were less common, with 9 out of 13 such comparisons being one vote short of a

consistent preference for the corresponding large difference. If an 8/10 criterion for consistency had been used throughout, this would have increased the number of consistent preferences (from mean and median = 11 for 9/10 to mean = 13 and median = 14 for 8/10); however, no change was seen to the overall pattern of responses shown in Figure 4.

The significant interaction between environment and difference between HA settings highlights the exceptions to the main findings. Although participants were more likely to have a consistent preference for intensity than for the gain-frequency slope differences, this was only the case for the less difficult environments; and although there was a trend for fewer consistent preferences as environments became more difficult, this was not the case for directionality (Figure 4). The pattern of the consistent preferences for the directionality pair may be influenced by the effectiveness of the directional microphone in improving intelligibility in the different environments. The largest number of consistent preferences for the monologue in traffic noise at 5 dB SNR was expected, as the directional microphone would be most effective in improving the SNR in this situation by decreasing the section of low-frequency dominant traffic noise present behind the participant. The increase in the number of consistent preferences from the dialogue in café noise at 5 dB SNR to the same environment at 0 dB SNR is similar to findings of Walden et al. (2005), who asked 31 participants to select a preference between omnidirectional and hypercardioid responses when listening to sentences presented in speech-shaped noise. When they changed the SNR from 6 to 0 dB, the percentage of preferences for the directional microphone increased from around 55% to 80%. These findings are in line with the expectation of a nonlinear relationship between preference for directionality and SNR, with directional microphones being effective in a small range of SNRs, but no more effective than omnidirectional microphones at very large and small SNRs due to the dominance of the target and inability to effectively improve the SNR, respectively (e.g., Walden et al. 2005).

The individual participant profile measures used in this study (psychoacoustic, cognitive, and personality) could not predict who obtained consistent preferences for HA settings with large or small differences across environments. Future research could explore other possible predictors of consistency of preference.

Study Limitations

Some methodological choices may limit the extension of these findings. Despite the aim of simulating real-world test environments, implementations of the speech target and HA amplification resulted in reduced realism of the listening condition. First, although realistic background

stimuli and speech signals were used, the speech signals lacked the influence of the background noise on the speaker's voice (Lombard effect) and reverberation, potentially limiting the applicability of the findings to similar real-life situations. The effect of the presence of the Lombard effect and reverberation on consistency of preference is unknown. When listening in background noise, Lombard speech is expected to be more easily understood than speech recorded in quiet (Pichora-Fuller et al., 2010). On the other hand, the inclusion of reverberation in the speech signal would reduce speech understanding (e.g., Helfer & Wilber, 1990).

Second, the signal processing implemented in the master HA was less sophisticated than what is available in most modern commercial HAs. It is possible additional signal processing could increase the difference between the HA settings beyond the differences of intensity, gain-frequency slope, and directionality introduced in this study. This increased difference between the HA settings could increase the number of consistent preferences, as participants obtained more consistent preferences for large rather than small differences.

Third, the amount of low-frequency gain provided was more than what is prescribed for participants with normal and near-normal hearing in the low frequencies. All participants were provided with a minimum amount of gain to ensure the difference between the HA settings was achieved for all participants. However, when clinically fitted, HA users with normal low-frequency hearing and a high-frequency hearing loss would be provided with venting, reducing low-frequency gain and consequently reducing the contrast between the pairs of HA settings. The use of gain and venting matching the low-frequency thresholds is expected to result in fewer consistent preferences in our experiment, as participants had fewer consistent preferences for small rather than large differences between the pairs of HA settings.

Finally, when evaluating consistency of preference, comparisons were presented in a randomized order across the different HA settings, contrary to approaches followed when fine-tuning, whether done by the clinician or by the user in their own listening environment, where complaints would be addressed successively. This presentation mode was chosen to reduce a possible order effect, as participants completed 40 to 50 comparisons for each environment. This randomization within each environment may have resulted in fewer consistent preferences for participants who selected different preference criteria (e.g., comfort, speech perception) for the comparisons of different HA settings (intensity, gain-frequency slope, and directionality).

A final methodological choice that may impact the findings was the use of a self-paced paradigm. When selecting their preference, participants could listen for as long as they liked and switch back and forth between

settings as many times as they liked. The average duration of the 10 comparisons for each HA setting varied from 3 to 88 s. A self-paced paradigm was chosen over a more structured paradigm, where participants would listen for a set duration to each HA setting with a predefined number of switches between settings before being asked to vote, as it would both reduce session times and listening fatigue when participants could make a quick decision. A potential downfall of the self-paced paradigm is that participants may be less thorough in their evaluation, leading to inconsistent preferences. However, shorter vote durations in this study were associated as one might expect with the easier environments and perceptually more different HA settings, conditions for which the largest proportions of consistent preferences were observed. In addition, a self-paced paradigm more closely parallels how a person will go about evaluating different settings to select a preference in real life.

Implications and Future Directions

These findings suggest listeners may not have a consistent preference in all listening situations when choosing between two alternative HA settings, challenging the effectiveness of typical fine-tuning approaches in clinical practice. Completing multiple paired-comparisons in the clinic to ensure consistency of preference would be time-consuming, but possible. However, the effectiveness of performing multiple paired-comparisons in the clinic is limited due to the dependency of consistency of preference on the listening environment and the HA settings selected for comparison. The dependency on listening environment is particularly problematic due to difficulties in identifying (Valentine et al., 2011) and recreating (Dreschler et al., 2008) the same, potentially complex, listening environments in the clinic that cause problems in the field. Alternatively, the user may fine-tune the amplification characteristics themselves in their own listening environment. Today, many hearing devices are app controlled (Chasin, 2017), with the app giving the user access to rather sophisticated controls for manipulating the amplification characteristics. Most commonly, the in-situ changes made by the user to the HA setting are temporary, meaning the changes will be undone when the device is next turned off. Permanent fine-tuning is possible by either allowing the user to create an additional listening program for a particular situation, or by providing them with trainable aids. Trainable HAs allow for those requiring extensive fine-tuning to complete multiple paired-comparisons between the current HA settings and any adjustments made. In view of the dependency of consistency of preference on environment and HA setting, trainable HA users would benefit from counselling. Counselling would help establish realistic expectations of the technology, as their efforts may be less effective in more difficult listening

environments and when altering directionality and the slope of the gain-frequency response.

The relationship between consistency of preference in simulated real-world environments and trainable HA outcomes in the real world remains to be investigated: Can those with more consistent preferences in the laboratory make more consistent adjustments to the HA settings, and in the process fine-tune their HAs? In parallel, investigation is required to establish if those with fewer consistent preferences make inconsistent adjustments of trainable devices resulting in undesirable settings, a concern held by 73% of clinicians who reported not to activate training, in response to a survey about their use and perception of trainable HAs (Walravens et al., 2016).

Conclusion

The findings from this study showed variability in consistency of preference for adults with hearing ranging from normal to moderate sensorineural hearing loss, depending on the difference between the HA settings, the environment and their interaction. Furthermore, the study showed that some common psychoacoustic and cognitive measures, plus measures of the “Big Five” personality traits did not predict consistency. These findings challenge the effectiveness of fine-tuning procedures, as they are commonly performed in the clinic and suggest that users who are training their own HAs could benefit from counselling to ensure they have realistic expectations of the technology.

Appendix: Instructions for the Preference Task

You will be listening to different situations that you may experience in real life. In each situation, you will listen to different HA settings in pairs. The settings can be different in volume, pitch, or direction.

Imagine you are given new HAs that can be set up with different settings for different situations. Your task is to choose which setting you would prefer in your new HAs for listening to each situation.

The situations you will be listening to are:

- traffic noise;
- one woman talking in traffic noise;
- two people talking in café noise.

The levels are based on the levels experienced during the recording. In the test booth, they seem to be louder than in real life, but they aren't.

Listen to each situation and compare two settings by pushing buttons A and B on the controller in front of you. Listen to settings A and B for as long as you like.

Once you have chosen your preferred setting, push the button for that setting and then press VOTE. The settings of A and B will differ from trial to trial.

Please consider your choice carefully. You can listen for as long as you like. You will listen to 50 pairs of settings.

Do you have any questions?

Notes

Preliminary findings from this study were included in a poster presented at the International Hearing Aid Research Conference, Tahoe City, California, on August 13, 2016, for which the first author received a student scholarship. Findings from this study were presented at the B-Audio session of the Annual Congress 2017 of the Royal Belgian Society for Ear, Nose, Throat, Head and Neck Surgery, Louvain-la-Neuve, Belgium, on November 24, 2017, and at the Audiology Australia National Conference 2018, Sydney, Australia, on May 22, 2018.

Acknowledgments

The authors thank the volunteers who spent hours participating in this study, Jörg Buchholz for implementing the playback of the recordings of real-life acoustic environments, and Cong-Van Nguyen for creating the automated voting system. The authors also thank James Galloway for setting up the test area and implementing the psychoacoustic tests, Mark Seeto for statistical analysis and advice, and Benjamin Steves and team at Psytest for use of the Test of Attentional Performance—Mobility version. And finally, the authors thank Scott Brewer for his IT wizardry.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and publication of this article: This work was conducted by the HEARING CRC, established and funded through the Cooperative Research Centres Program—Business Australia. The authors acknowledge the financial support of the Australian Government through the Department of Health. The first author received support from an Australian Government Research Training Program Scholarship.

ORCID iD

Els Walravens  <https://orcid.org/0000-0002-1260-128X>

Supplemental material

Supplemental material for this article is available online.

References

- American National Standards Institute. (1998). Methods for calculation of the speech intelligibility index (ANSI S3.5-1997).
- Anderson, M. C., Arehart, K. H., & Souza, P. E. (2018). Survey of current practice in the fitting and fine-tuning of common signal-processing features in hearing aids for adults. *Journal of the American Academy of Audiology*, 29(2), 118–124. <https://doi.org/10.3766/jaaa.16107>
- Athalye, S. P. (2010). *Factors affecting speech recognition in noise and hearing loss in adults with a wide variety of auditory capabilities* [PhD thesis]. Institute of Sound and Vibration Research, University of Southampton, <https://eprints.soton.ac.uk/191083/>
- Bentler, R. A., & Chiou, L.-K. (2006). Digital noise reduction: An overview. *Trends in Amplification*, 10(2), 67–82. <https://doi.org/10.1177/1084713806289514>
- Best, V., Keidser, G., Freeston, K., & Buchholz, J. M. (2016). A dynamic speech comprehension test for assessing real-world listening ability. *Journal of the American Academy of Audiology*, 27(7), 515–526. <https://doi.org/10.3766/jaaa.15089>
- Byrne, D., & Cotton, S. (1988). Evaluation of the National Acoustic Laboratories' new hearing aid selection procedure. *Journal of Speech and Hearing Research*, 31(2), 178–186. <https://doi.org/10.1044/jshr.3102.178>
- Caswell-Midwinter, B., & Whitmer, W. M. (2019). Discrimination of gain increments in speech-shaped noises. *Trends in Hearing*, 23, 2331216518820220. <https://doi.org/10.1177/2331216518820220>
- Chasin, M. (2017). Overview of smartphone control of hearing aids. *Canadian Audiologist*, 4(4). <http://canadianaudiologist.ca/issue/volume-4-issue-4-2017/smartphone-overview-feature/>
- Cox, R. M., Alexander, G. C., Taylor, I. M., & Gray, G. A. (1997). The contour test of loudness perception. *Ear and Hearing*, 18(5), 388–400. <https://doi.org/10.1097/00003446-199710000-00004>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466. [https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- De Beuckelaer, A., Kampen, J. K., & Van Trijp, J. C. M. (2013). An empirical assessment of the cross-national measurement validity of graded paired comparisons. *Quality & Quantity*, 47(2), 1063–1076. <https://doi.org/10.1007/s11135-011-9583-1>
- Dillon, H., Zakis, J. A., McDermott, H. J., Keidser, G., Dreschler, W. A., & Convery, E. (2006). The trainable hearing aid: What will it do for clients and clinicians? *Hearing Journal*, 59(4), 30–34, 36. <https://doi.org/10.1097/01.HJ.0000286694.20964.4a>
- Dreschler, W. A., Keidser, G., Convery, E., & Dillon, H. (2008). Client-based adjustments of hearing aid gain: The effect of different control configurations. *Ear and Hearing*, 29(2), 214–227. <https://doi.org/10.1097/AUD.0b013e31816453a6>
- Edwards, B. W., Hou, Z., Struck, C. J., & Dharan, P. (1998). Signal processing algorithms for a new, software-based, digital hearing device. *Hearing Journal*, 51(9), 44–52.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Hansen, M. (2006). *Lehre und Ausbildung in Psychoakustik mit psylab: freie Software für psychoakustische Experimente* [Education in psychoacoustics with psylab: free software for psychoacoustic experiments]. Paper presented at the Fortschritte der Akustik - DAGA 2006, Braunschweig.
- Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech and Hearing Research*, 33(1), 149–155. <https://doi.org/10.1044/jshr.3301.149>
- Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, 49(12), 891–903. <https://doi.org/10.3109/14992027.2010.506889>
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Jepsen, M. L., & Dau, T. (2011). Characterizing auditory processing and perception in individual listeners with sensorineural hearing loss. *Journal of the Acoustical Society of America*, 129(1), 262–281. <https://doi.org/10.1121/1.3518768>
- Keidser, G. (1995). The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids. *Ear and Hearing*, 16(6), 575–586. <https://doi.org/10.1097/00003446-199512000-00004>
- Keidser, G., Brew, C., Brewer, S., Dillon, H., Grant, F., & Storey, L. (2005). The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input. *International Journal of Audiology*, 44(11), 656–670. <https://doi.org/10.1080/14992020500266803>
- Keidser, G., Dillon, H., & Convery, E. (2008). The effect of the base line response on self-adjustments of hearing aid gain. *Journal of the Acoustical Society of America*, 124(3), 1668–1681. <https://doi.org/10.1121/1.2951500>
- Keidser, G., Dillon, H., Flax, M., Ching, T. Y. C., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, 1(e24), 88–90. <https://doi.org/10.4081/audiores.2011.e24>
- Keidser, G., & Grant, F. (2001). Comparing loudness normalization (IHAF) with speech intelligibility maximization (NAL-NL1) when implemented in a two-channel device.

- Ear and Hearing*, 22(6), 501–515. <https://doi.org/10.1097/00003446-200112000-00006>
- Kincaid, J. P., Fishburne, R. P., Jr., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*. <http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf>
- Kuk, F. K., & Pape, N. M. C. (1992). The reliability of a modified simplex procedure in hearing aid frequency-response selection. *Journal of Speech and Hearing Research*, 35(2), 418–429. <https://doi.org/10.1044/jshr.3502.418>
- Larsby, B., & Arlinger, S. (1998). A method for evaluating temporal, spectral and combined temporal-spectral resolution of hearing. *Scandinavian Audiology*, 27(1), 3–12. <https://doi.org/10.1080/010503998419641>
- Lentz, J. J., & Leek, M. R. (2003). Spectral shape discrimination by hearing-impaired and normal-hearing listeners. *The Journal of the Acoustical Society of America*, 113(3), 1604–1616. <https://doi.org/10.1121/1.1553461>
- Lister, J. J., Roberts, R. A., & Lister, F. L. (2011). An adaptive clinical test of temporal resolution: Age effects. *International Journal of Audiology*, 50(6), 367–374. <https://doi.org/10.3109/14992027.2010.551218>
- Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, 42(Suppl 1), S49–S58. <https://doi.org/10.3109/14992020309074624>
- McShefferty, D., Whitmer, W. M., & Akeroyd, M. A. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends in Hearing*, 19, 1–9. <https://doi.org/10.1177/2331216515572316>
- McShefferty, D., Whitmer, W. M., & Akeroyd, M. A. (2016). The just-meaningful difference in speech-to-noise ratio. *Trends in Hearing*, 20. <https://doi.org/10.1177/2331216515626570>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Moore, B. C. J., & Sek, A. (2013). Comparison of the CAM2 and NAL-NL2 hearing aid fitting methods. *Ear and Hearing*, 34(1), 83–95. <https://doi.org/10.1097/AUD.0b013e3182650adf>
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, 81, 111–121. <https://doi.org/10.1111/j.2044-8295.1990.tb02349.xronn>
- Nasreddine, Z. S., Phillips, N. A., Bedirian, 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>
- Neher, T. (2014). Relating hearing loss and executive functions to hearing aid users’ preference for, and speech recognition with, different combinations of binaural noise reduction and microphone directionality. *Frontiers in Neuroscience*, 8, 391. <https://doi.org/10.3389/fnins.2014.00391>
- Neher, T., Grimm, G., Hohmann, V., & Kollmeier, B. (2014). Do hearing loss and cognitive function modulate benefit from different binaural noise-reduction settings? *Ear and Hearing*, 35(3), e52–62. <https://doi.org/10.1097/AUD.0000000000000003>
- Oreinos, C. (2015). *Virtual acoustic environments for the evaluation of hearing devices (PhD Thesis)*. Macquarie University.
- Oreinos, C., & Buchholz, J. M. (2015). Objective analysis of ambisonics for hearing aid applications: Effect of listener’s head, room reverberation, and directional microphones. *The Journal of the Acoustical Society of America*, 137(6), 3447–3465. <https://doi.org/10.1121/1.4919330>
- Oreinos, C., & Buchholz, J. M. (2016). Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids. *Journal of the American Academy of Audiology*, 27(7), 541–556. <https://doi.org/10.3766/jaaa.15094>
- Pearsons, K. S., Bennett, R. L., & Sanford, F. (1977). *Speech levels in various noise environments*. <https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockkey=P100CWGS.TXT>
- Pichora-Fuller, M. K., Goy, H., & Van Lieshout, P. (2010). Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments. *Seminars in Hearing*, 31(02), 077–094. <https://doi.org/10.1055/s-0030-1252100>
- Rönneberg, J., Arlinger, S., Lyxell, B., & Kinnefors, C. (1989). Visual evoked potentials: Relation to adult speechreading and cognitive function. *Journal of Speech and Hearing Research*, 32(4), 725–735. <https://doi.org/10.1044/jshr.3204.725>
- Smeds, K., Wolters, F., & Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, 26(2), 183–196. <https://doi.org/10.3766/jaaa.26.2.7>
- Thielemans, T., Pans, D., Chenault, M., & Anteunis, L. (2017). Hearing aid fine-tuning based on Dutch descriptions. *International Journal of Audiology*, 56(7), 507–515. <https://doi.org/10.1080/14992027.2017.1288302>
- Valentine, S., Dundas, J. A., & Fitz, K. (2011). Evidence for the use of a new patient-centered fitting tool. *Hearing Review*, 18(4), 28–34. <http://www.hearingreview.com/2011/04/evidence-for-the-use-of-a-new-patient-centered-fitting-tool/>
- van Esch, T. E. M., Kollmeier, B., Vormann, M., Lyzenga, J., Houtgast, T., Hällgren, M., Larsby, B., Athalye, S. P., Lutman, M. E., & Dreschler, W. A. (2013). Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *International Journal of Audiology*, 52(5), 305–321. <https://doi.org/10.3109/14992027.2012.759665>
- Walden, B. E., Surr, R. K., Grant, K. W., Van Summers, W., Cord, M. T., & Dyrland, O. (2005). Effect of signal-to-noise ratio on directional microphone benefit and preference. *Journal of the American Academy of Audiology*, 16(9), 662–676. <https://doi.org/10.3766/jaaa.16.9.4>

- Walravens, E., Keidser, G., & Hickson, L. (2016). Provision, perception and use of trainable hearing aids in Australia: A survey of clinicians and hearing impaired adults. *International Journal of Audiology*, 55(12), 787–795. <https://doi.org/10.1080/14992027.2016.1219776>
- Wu, Y.-H. (2010). Effect of age on directional microphone hearing aid benefit and preference. *Journal of the American Academy of Audiology*, 21(2), 78–89. <https://doi.org/10.3766/jaaa.21.2.3>
- Zimmermann, P., & Fimm, B. (2014). Test of Attentional Performance (Mobility version) Version 1.3. Herzogenrath: Psytest.