Differences in Word and Phoneme Recognition in Quiet, Sentence Recognition in Noise, and Subjective Outcomes between Manufacturer First-Fit and Hearing Aids Programmed to NAL-NL2 Using Real-Ear Measures

Michael Valente*
Kristi Oeding*
Alison Brockmeyer*
Steven Smith*
Dorina Kallogjeri*

Abstract

Background: The American Speech-Language-Hearing Association (ASHA) and American Academy of Audiology (AAA) have created Best Practice Guidelines for fitting hearing aids to adult patients. These guidelines recommend using real-ear measures (REM) to verify that measured output/gain of hearing aid(s) match a validated prescriptive target. Unfortunately, approximately 70–80% of audiologists do not routinely use REM when fitting hearing aids, instead relying on a manufacturer default "first-fit" setting. This is problematic because numerous studies report significant differences in REM between manufacturer first-fit and the same hearing aids using a REM or programmed-fit. These studies reported decreased prescribed gain/output in the higher frequencies for the first-fit compared with the programmed fit, which are important for recognizing speech. Currently, there is little research in peer-reviewed journals reporting if differences between hearing aids fitted using a manufacturer first-fit versus a programmed-fit result in significant differences in speech recognition in quiet, noise, and subjective outcomes.

Purpose: To examine if significant differences were present in monosyllabic word and phoneme recognition (consonant-nucleus-consonant; CNC) in quiet, sentence recognition in noise (Hearing in Noise Test; HINT), and subjective outcomes using the Abbreviated Profile of Hearing Aid Benefit (APHAB) and the Speech, Spatial and Qualities of Hearing (SSQ) questionnaires between hearing aids fit using one manufacturer's first-fit and the same hearing aids with a programmed-fit using REM to National Acoustic Laboratories Nonlinear Version 2 (NAL-NL2) prescriptive target.

Research Design: A double-blind randomized crossover design was used. Throughout the study, one investigator performed all REM whereas a second investigator measured speech recognition in quiet, noise, and scored subjective outcome measures.

Study Sample: Twenty-four adults with bilateral normal sloping to moderately severe sensorineural hearing loss with no prior experience with amplification.

Data Collection and Analysis: The hearing aids were fit using the proprietary manufacturer default first-fit and a programmed-fit to NAL-NL2 using real-ear insertion gain measures. The order of the two fittings was randomly assigned and counterbalanced. Participants acclimatized to each setting for four weeks and returned for assessment of performance via the revised CNC word lists, HINT, APHAB, and SSQ for the respective fitting.

Results: (1) A significant median advantage of 15% ($p < 0.001$; 95% CI: 9.7–24.3%) for words and 7.7% ($p < 0.001$; 95% CI: 5.9–10.9%) for phonemes for the programmed-fit compared with first-fit at 50 dB

*Department of Otolaryngology – Head and Neck Surgery, Division of Adult Audiology, Washington University in St. Louis School of Medicine, St. Louis, MO

Corresponding author: Michael Valente, Department of Otolaryngology, Washington University in St. Louis School of Medicine, St. Louis, MO 63110; Email: valentem@wustl.edu

sound pressure level (SPL) and 4% ($p < 0.01$; 95% CI: 1.7–6.3%) for words at 65 dB SPL; (2) No significant differences for the HINT reception threshold for sentences (RTS); (3) A significant median advantage of 4.2% [$p < 0.04$; 95% confidence interval (CI): −0.6–13.2%] for the programmed-fit compared with the first-fit for the background noise subscale problem score for the APHAB; (4) No significant differences on the SSQ.

Conclusions: Improved word and phoneme recognition for soft and words for average speech in quiet were reported for the programmed-fit. Seventy-nine percent of the participants preferred the programmed-fitting versus first-fit. Hearing aids, therefore, should be verified and programmed using REM to a prescriptive target versus no verification using a first-fit.

Key Words: hearing aid, manufacturer first-fit, programmed-fit, real-ear insertion gain (REIG), real-ear measures (REM), reception threshold for sentences (RTS)

Abbreviations: APHAB = Abbreviated Profile of Hearing Aid Benefit; AV = aversiveness of sounds; BN = background noise; CI = confidence interval; CNC = consonant-nucleus-consonant; EC = ease of communication; GoF = goodness of fit; HINT = Hearing in Noise Test; ISTS = international speech test signal; LDL = loudness discomfort level; NAL-NL1 = National Acoustic Laboratories Nonlinear Version 1 prescriptive target; NAL-NL2 = National Acoustic Laboratories Nonlinear Version 2 prescriptive target; NAL-R = National Acoustic Laboratories Revised prescriptive target; REAR = real-ear aided response; REIG = real-ear insertion gain; REM = real-ear measures; REUG = real-ear unaided gain; RIC/RITC = receiver-in-the-canal; RTS = reception threshold for sentences; RV = reverberation; SD = standard deviation; SNR = signal-to-noise ratio; SPL = sound pressure level; SSQ = Speech, Spatial, and Qualities of Hearing; WRS = word recognition score

INTRODUCTION

Real-ear measures (REM) are a tool audiologists use to program and verify amplification produced by hearing aids in patients' ear canals. When using REM to fit hearing aids, the patient's audiogram is entered into a real-ear analyzer and a real-ear insertion gain (REIG) or real-ear aided response (REAR) prescriptive target is generated based on the patient's hearing loss and corrected for possible air-bone gap(s), a unilateral or bilateral (binaural summation) fitting, and the number of channels of signal processing (channel summation). The prescriptive target calculates the amount of gain or output the patient requires from 250 to 8000 Hz from the hearing aid(s) to provide appropriate amplification for his/her hearing loss.

REM have been recommended as the "gold standard" for fitting hearing aids in Best Practice Guidelines by the American Speech-Language-Hearing Association (ASHA, 1998) and the American Academy of Audiology (AAA, 2006). In addition, the British Society of Audiology (BSA, 2007) recommends that measured REIG should be within ±5 dB of target to 2000 Hz and within ±8 dB at 3000–8000 Hz. The *slope* at each octave should also be within ±5 dB of the target slope. Despite these recommendations, approximately 70–80% of dispensed hearing aids are not routinely verified and programmed ("always" or "nearly always") using REM (Consumer Reports, 2009; Mueller and Picou, 2010). Rather, the manufacturer's first-fit is only utilized. A first-fit is typically based on a prescriptive target or can be proprietary, but usually includes modifications based on a hearing aid's features. When a first-fit alone is utilized,

the patient's hearing loss is placed into the programming software and the manufacturer predicts the output/gain of the hearing aid(s) using a proprietary algorithm or a modified prescriptive target. The output/gain of the hearing aid(s) is not verified by REM to ensure the hearing aids are providing the appropriate amount of amplification. The concern with using a manufacturer first-fit instead of verifying and programming hearing aids to a prescriptive target, such as National Acoustic Laboratories Nonlinear Version 2 (NAL-NL2; Keidser et al, 2012) is that first-fits are not validated by independent research nor do they typically provide appropriate gain/output when compared with a validated prescriptive target. If the gain/output is not verified using REM to closely match a prescriptive target, the patient may not receive optimum audibility of speech.

Numerous investigations (Swan and Gatehouse, 1995; Aarts and Caffee, 2005; Aazh and Moore, 2007; Aazh et al, 2012; Abrams et al, 2012; Boymans and Dreschler, 2012; Leavitt and Flexer, 2012; Sanders et al, 2015; Munro et al, 2016) have reported that the manufacturer first-fit (whether proprietary or a modified prescriptive algorithm) provided significantly less gain/output in comparison to when hearing aids were programmed-fit to a prescriptive target using REM. The greatest difference in gain/output was in the higher frequencies.

Several studies have examined differences in gain/output of a prescriptive target using REM between a manufacturer-first fit and programmed-fit. Sanders et al (2015) fit five premium mini receiver-in-the-canal (RIC) digital hearing aids from five manufacturers on eight participants with sloping high-frequency hearing loss. The programmed-fit was verified and programmed

using REAR at 55, 65, and 75 dB sound pressure level (SPL) using NAL-NL2. Results revealed the first-fit for 55 dB SPL was below NAL-NL2 in 74% of the cases. For 65 dB SPL, 55% of the cases were 10 dB or more below target in at least one frequency between 250 and 4000 Hz. Most fits were 7–10 dB below NAL-NL2 above 2000 Hz. For the 75 dB SPL input, first-fit was equal or greater than NAL-NL2.

In a similar study, Aarts and Caffee (2005) investigated the predictive accuracy of one manufacturer's first-fit in a nine-channel hearing aid in 41 participants (N = 79 ears) using two audiometric configurations (flat and sloping). REAR was used to verify the output/gain using a 50 and 90 dB SPL signal. Using a "passing" criterion of within ±4 dB, the first-fit differed from the prescriptive REAR target (unspecified prescriptive target) in 88–98% of the ears and the differences were greater in males. For both audiometric configurations and input levels, one-third to one-half of first-fits were below the REAR prescriptive target by ≥4 dB for the 50 dB SPL input (i.e., under-amplifying soft sounds). For the 90 dB SPL input, first-fits exceeded the REAR target ≥4 dB for 19–26% of the measurements (i.e., overamplifying loud sounds). Most first-fits were below REAR target between 1500 and 4000 Hz for 50 dB SPL and were 10–15 dB below REAR target at 4000 Hz for the two input levels and the two configurations.

In a similar study examining differences in REIG, Swan and Gatehouse (1995) investigated 319 new hearing aid users with analog technology and reported that 57% of the fittings using the manufacturer first-fit were not within ±10 dB of NAL between 250 and 3000 Hz. After fine-tuning and changing the earmold configurations, 85% achieved satisfactory gain. Aazh and Moore (2007) investigated first-fit versus programmed-fit REIG to National Acoustic Laboratories Nonlinear Version 1 prescriptive target (NAL-NL1; Byrne et al, 2001) in open fit and custom earmold fittings in 42 ears (24 participants) using four digital hearing aids with four or seven "gain handles." It was reported that 64% (73% for females and 60% for males) of the first-fits failed to arrive within ±10 dB of NAL-NL1 at one or more frequencies at 250–4000 Hz and the percent of failures were greatest in the high frequencies (19–40% failure at 1500–4000 Hz). Also, the first-fit open fittings had greater failures (68%) than the closed fittings (60%). After programming to NAL-NL1, 83% arrived within ±10 dB of NAL-NL1. Aazh et al (2012) also investigated manufacturer first-fit versus programmed-fit REIG to NAL-NL1 in an open-fit hearing aid with eight channels/eight bands in 51 fittings for 30 participants. It was reported that 71% of the first-fit fittings failed to be within ±10 dB of NAL-NL1 at one or more frequencies at 250–4000 Hz and after programming, 82% arrived within ±10 dB of NAL-NL1.

Finally, in a study examining Goodness of Fit (GoF; Hostler et al, 2004) Munro et al (2016) investigated 100

bilateral fits (49 open-fit and 51 closed-fit with custom earmolds) and reported only on results from one ear (50 right; 50 left). The hearing aids were programmed using the manufacturer first-fit and programmed-fit to NAL-NL1 using REIG. The investigators examined GoF calculation where "0.0" represented a very poor fit and "1.0" represented a perfect match to prescribed REIG at all measured frequencies. The calculated GoF is the culmination of three measures: (a) "close fit" (difference between target and programmed REIG at each test frequency), (b) "similar shape" (difference in the shape of the target and the programmed REIG), and (c) "adequate gain" (difference between target and the total programmed REIG). Before programming 18%, 50%, and 61% of the open-fit and 20%, 63%, and 67% of the custom earmold fit were within 10 dB of NAL-NL1 at 50, 65, and 80 dB SPL, respectively. After programming, >85% of open-fit and >90% of custom earmold fits were within 10 dB of NAL-NL1 for all input levels. The investigators reported an average GoF for the open-fit between 0.5 and 0.7 at 50, 65, and 80 dB SPL using the first-fit setting and 0.7–0.8 for the programmed-fit setting. The investigators also reported an average GoF for the custom earmold fits of 0.6–0.8 at 50, 65, and 80 dB SPL using the first-fit setting and 0.8–0.9 for the programmed-fit setting.

As these studies report, a manufacturer first-fit frequently fails to match a prescriptive target, particularly for high frequency signals. This implies that with a first-fit, a patient may not be obtaining the audibility he/she requires to recognize soft speech or possibly even average speech. To determine the negative effects of this decreased output/gain, further research with objective measurements of speech recognition are required. Only one study thus far has examined the impact of using a first-fit versus a programmed-fit on speech recognition in noise. Leavitt and Flexer (2012) in a nonpeer-reviewed article evaluated five participants and compared Quick-SIN signal-to-noise ratio (SNR) hearing loss (Killion et al, 2004) between manufacturer first-fit and the same hearing aids using programmed-fit to National Acoustic Laboratories Revised prescriptive target (NAL-R; Byrne and Dillon, 1986) using REM for six premium digital hearing aids from six different manufacturers. Participants performed poorer with all six hearing aids when fit using the manufacturer first-fit compared with the programmed fit to NAL-R. On average, the SNR loss improved by 6.6 dB when the hearing aids were programmed-fit to NAL-R using REM compared with first-fit.

In addition to understanding the impact of a first-fit on objective measures, subjective measures of patient preferences are also of interest to determine if patients prefer a first-fit or a programmed-fit. Only one study has examined perceived/subjective benefit of a first-fit compared with a programmed fit. Abrams et al (2012) investigated self-perceived hearing aid benefit by fitting half of 22

participants using a manufacturer first-fit and the other half were fit to NAL-NL1 using REM. The Abbreviated Profile of Hearing Aid Benefit (APHAB; Cox and Alexander, 1995) was completed after four weeks of hearing aid use. The APHAB benefit score with the programmed-fit setting was significantly better than first-fit in 75% of the fits for the ease of communication (EC), background noise (BN), and reverberation (RV) APHAB subscales. Of the 22 participants, seven preferred the first-fit and 15 preferred the programmed-fit using REM.

As can be seen, there are limited data examining the impact of decreased output/gain for high frequencies using a first-fit compared with a programmed-fit on objective and subjective outcomes. The data available suggest that the first-fit may have a negative impact on speech recognition and also subjective impressions of amplification fit using a first-fit because of reduced audibility of sound. As stated earlier, it is well documented that most dispensed hearing aids (Consumer Reports, 2009; Mueller and Picou, 2010) are not verified and programmed using REM and instead rely on manufacturer first-fit. These findings may help explain the poor satisfaction rate with hearing aids that is often discussed at conferences, in classroom, and journals. That is, the majority of patients may not be receiving sufficient gain/output for soft sounds, especially in the critically important high-frequency region and may be receiving excessive gain/output for loud sounds that are judged as being uncomfortable. Because of the limited evidence examining the impact of reduced output/gain of a first-fit on objective and subjective measures compared with a programmed-fit, the following study was conducted to determine the impact of the decreased first-fit output/gain on objective and subjective measures.

The primary purpose of this study was to determine if significant differences were present in speech recognition in quiet, noise, and subjective outcomes assessed using questionnaires between hearing aids fit using one manufacturer's first-fit and hearing aids programmed-fit to the NAL-NL2 prescriptive target using REM.

The results were examined to investigate the following:

- Are significant differences present in consonant-nucleus-consonant (CNC; Peterson and Lehiste, 1962) word and phoneme recognition in quiet presented at input levels of 50, 65, and 80 dB SPL between the two fitting settings?
- Are significant differences present in Hearing in Noise Test (HINT; Nilsson et al, 1994) sentence recognition in noise between the two fitting settings?
- Are significant differences present between the four subscales of the APHAB and the two fitting settings?
- Are significant differences present between the three subscales of the Speech, Spatial and Qualities of Hearing (SSQ; Gatehouse and Noble, 2004) and the two fitting settings?

METHODS

Participants

Twenty-five participants were recruited from the Division of Adult Audiology at Washington University in St. Louis School of Medicine via personal communication in the clinic, telephone, or a letter approved by the Human Research Protection Office. Using the results from Leavitt and Flexor (2012); mean = 3.3 dB, SD = 3.0 dB, an a priori power analysis and sample size calculation utilizing G*Power 3.0.10 (G*Power3, 2016) calculated that ten participants were needed to investigate with 80% power at the two-sided alpha level of 0.05 differences in speech recognition between hearing aids fit using manufacturer first-fit and hearing aids programmed-fit to a prescriptive target using REM. Assuming that the large effect size of 1.1 will not be observed in either outcome measures, and taking into consideration withdrawals from the study, the investigators enrolled 25 participants. The sample size of 25 participants would allow the investigators to detect an effect size of 0.58 or larger. During the study, one participant dropped out because of decreased hearing leaving a final participant count of 24.

Each participant signed an informed consent form approved by Human Research Protection Office. Inclusion criteria included (a) adults who were 18 yr or older; (b) bilateral sensorineural hearing loss; (c) hearing thresholds between 0–65 dB HL at 250–1000 Hz and 25–80 dB HL at 2000–6000 Hz; (d) unaided word recognition score (WRS) of at least 60% in each ear; (e) no prior experience with hearing aids; (f) able to keep head stable during measurements; and (g) able to comprehend basic maintenance and use of hearing aids. Participants were excluded if they had a history of (a) chronic or acute middle ear pathology (otitis media, etc.); (b) otosurgery; (c) hyperacusis or recruitment (loudness discomfort levels [LDLs] below 85 dB HL); (d) were nonambulatory; (e) could not commit to the time requirements of the study, and (f) did not meet the inclusion criteria.

Mean hearing thresholds (dB HL) and ± one SD for the 24 participants is reported in Figure 1. All participants had symmetrical sensorineural hearing loss, with the mean hearing threshold revealing a normal to slight sensorineural hearing loss sloping to moderately severe sensorineural hearing loss. The mean WRS was 85.9% (SD = 9.6%) for the right ear and 84.3% (SD = 11.2%) for the left ear. Sixteen participants were male (64%) and the mean age was 72.1 yr (SD = 10.4 yr).

Procedures

The study was a double-blinded crossover design to avoid bias where Investigator #1 performed all hearing aid programming, whereas Investigator #2 assessed all the speech tests in quiet and noise and the two subjective questionnaires. The study required four visits. In the
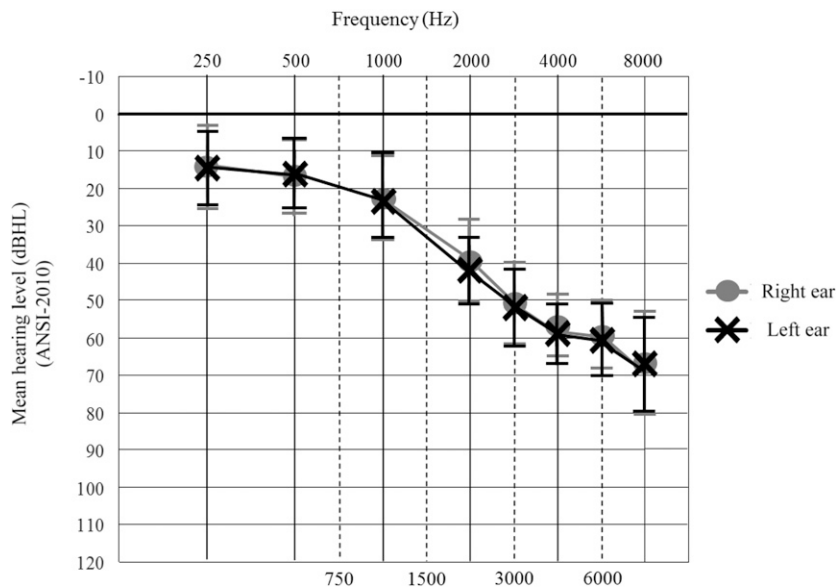
Figure 1. Audiogram reporting the mean and ±1 SD for hearing thresholds (dB HL) in the right (●) and left (x) ear.

first visit, participants were seated in a double-walled sound-treated booth, and otoscopy was completed to ensure the ear canals were clear bilaterally. A comprehensive audiological evaluation was performed if the participant had not had an audiological evaluation within the last six months to ensure that the participant was eligible to participate in the study. Air conduction thresholds were measured in octave and midoctave intervals from 250 to 8000 Hz, and bone conduction thresholds were measured at octave and midoctave intervals from 250 to 4000 Hz. Speech recognition threshold testing was performed using recorded spondee words of a female talker and WRS were measured utilizing the compact disc female-talker version of the Northwestern University Auditory Test Number 6 (Tillman and Carhart, 1966) at the participant's most intelligible level. The most intelligible level was measured using monitored-live voice presentation (voice peaking at 0 dB on the VU meter) by asking the participant to indicate when the presentation level was comfortably loud and most intelligible.

LDLs were measured to ensure the participant did not have hyperacusis/recruitment. The participant was asked to assign a loudness judgment ranging from "very soft" to "uncomfortably loud" to pure-tone stimuli at 500, 1000, 2000, and 4000 Hz. The pure-tone was initially presented at 50 dB HL or 10 dB above the participant's pure-tone threshold if the hearing loss exceeded 50 dB HL and was increased in 5 dB increments. The decibel level described as "Loud, but OK" was recorded. If the LDL was below 85 dB HL, the participant was excluded from the study.

Hearing Aids and Earmolds

Next, if the participant fulfilled all the inclusion criteria, bilateral earmold impressions were taken, and

custom RIC earmolds were ordered for each participant. The participant chose the preferred color for the hearing aid case, the appropriate receiver wire length was measured, and bilateral RIC hearing aids were ordered. The experimental hearing aids were "premium" hearing aids having 16 channels/bands of signal processing and automatic adaptive multichannel directional microphones. At this visit, participants were randomly assigned into two groups: manufacturer first-fit or programmed-fit to NAL-NL2 using REM. This initial visit lasted approximately one hour and the participant was scheduled to return in two weeks for the fitting of the hearing aids using one of the two fitting settings.

Hearing Aid Fitting

Before the second visit, the electroacoustic analysis of the hearing aids (ANSI, 2009) and the directional microphone performance were verified using the Frye 8000 analyzer to ensure proper function of the hearing aids.

At the second visit, otoscopy was performed to ensure clear ear canals bilaterally. The hearing aids and earmolds were inserted to ensure a comfortable fit. The participant's audiometric thresholds from 250 to 8000 Hz in octave and midoctave intervals were entered into the Aurical Hearing Aid Analyzer (V 4.78.00), and the NAL-NL2 REIG prescriptive targets for input levels of 50, 65, and 80 dB SPL were generated. The Aurical hearing aid analyzer was configured as follows—Target rule: NAL-NL2; Fitting mode: real-ear; Applied real-ear unaided gain (REUG): measured REUG; Hearing aid type: behind-the-ear; Venting: diameter based on low-frequency hearing levels; Amplification: binaural; Tubing: RITC; Limiting: wideband; No. of channels:

16; Compression: 52; Compression speed: fast; Fitting depth: standard; Target type: REIG; REUG orientation: 0°; Experience: new; and Language type: nontonal.

For participants who started with manufacturer first-fit, the participant's hearing thresholds were entered into NOAH (V 4.5.0). The test hearing aids were connected to the most current manufacturer fitting software. The manufacturer software calculated/estimated gain based on the participant's entered hearing thresholds and using the manufacturer's default proprietary fitting algorithm. The default settings from the manufacturer software were downloaded to the hearing aids. The default settings were Experience Manager: 3; Noise Control: adaptive; and Fitting Rule: default. The reference microphone and probe microphone were calibrated. Then the probe-tube was advanced until the measured REUG was as close as possible to 0 dB at 6000 Hz using ANSI white noise presented at 65 dB SPL from the Aurical loudspeaker placed 12″ from the center of the head. Then the probe tube was taped in place. Next, the hearing aid and earmold were inserted into the ear canal and the real-ear aided gain and resulting REIG were measured using the international speech test signal (ISTS; Holube et al, 2010) presented for 14 sec at 50, 65, and 80 dB SPL. Because the participant was fit using the manufacturer first-fit, no programming changes were made. Therefore, in the first-fit fitting, REM were used only as a tool to determine the amount of output/gain the hearing aid was providing compared with NAL-NL2 and not as a tool to program the hearing aids. The NAL-NL2 targets, measured real-ear aided gain and REIG for 50, 65, and 80 dB SPL as well as the measured REUG were recorded at octave and midoctave frequencies from 250 to 8000 Hz. Loudness scaling was completed using the ISTS signal. The participant was asked to assign a loudness judgment ranging from "very soft" to "uncomfortably loud" to input levels of the ISTS signal randomly presented at 50, 60, 70, and 80 dB SPL. Finally, the participant was instructed on the use and maintenance of the hearing aids and earmolds and provided with instructional handouts.

For the participants where the hearing aids were first programmed-fit to NAL-NL2, the same methodology as described earlier was used, but instead the manufacturer software was used to program the hearing aids to match as close as possible to the NAL-NL2 targets at 50, 65, and 80 dB SPL. Early in the investigation, it became apparent that the first-fit algorithm for the manufacturer software provided essentially linear signal processing (i.e., minimal change in measured REIG as the input level changed from 50 to 80 dB SPL). When programming the hearing aids, the REIG was programmed to match NAL-NL2 within ±5 dB at 500–4000 Hz using the 65 dB SPL input level. Matching NAL-NL2 at 500–4000 Hz to be within ±5 dB was also the goal for the input levels of 50 and 80 dB SPL.

Attempting to match NAL-NL2 for the 50 and 80 dB SPL input levels, however, resulted in changes to the previously matched NAL-NL2 for the 65 dB SPL input level. According to the manufacturer, the compressor within the hearing aids allows the user to adjust gain for soft, average, and loud input levels independently by using two compression threshold kneepoints and two compression ratios. There were occasions, however, when Investigator #1 observed changes in the other previously programmed frequency response(s) when programming to another input level because the processor algorithm included safeguards to ensure the compression ratio did not exceed 3:1 or be less than 1:1. Because of the presence of this algorithm, a decision was made by the investigators to program the 50 and 80 dB SPL input levels as closely as possible to match the respective NAL-NL2 target but without changing the previously measured REIG for 65 dB SPL.

This visit lasted approximately one and a half hours. All participants in the two groups were contacted one week following the initial hearing aid fitting. Questions regarding the care and use of the hearing aids were answered, and the participant was invited to return for reinstruction or clarification if needed. No fine-tuning adjustments, however, were made to hearing aid programming. The participant then returned for the third visit in four weeks.

CNC

At the third visit, aided speech testing in quiet and noise was performed in a double-walled sound-treated booth containing the R-Space™ system described elsewhere by Investigator #2. For listening in quiet, the participant repeated CNC words presented at 50, 65, and 80 dB SPL from a loudspeaker located at 0° azimuth two feet from the participant. The participant and the investigator wrote each word on an answer sheet to avoid scoring errors. There are ten CNC word lists with 50 words per list. Lists 1–8 and 10 were used because of better interlist equivalency (Skinner et al, 2006). One complete list was presented at each presentation level and the word lists were randomized among participants. Word and phoneme scoring were completed. The whole word needed to be repeated correctly to obtain a correct response for each of the 50 words. For phoneme scoring, the correct phoneme needed to be repeated to obtain a credit for each of the 150 phonemes.

HINT

Next, the HINT was performed using noise presented in the R-Space system. The R-Space system consists of eight Boston Acoustics CR-65 loudspeakers in a circular array, with each loudspeaker separated by 45° in a 1.97 × 2.54 × 2.73 m double-walled sound suite

(volume = 14.05 m^3) with a reported RV time of 0.19 sec (personal communication with Industrial Acoustics Company [Bronx, NY]). The radius of the circle is two feet plus the depth of the loudspeaker (200 mm; see Oeding et al, 2010 for a detailed description of the R-Space system, daily calibration, and recording of the uncorrelated restaurant noise). The R-Space was calibrated daily before testing.

The HINT consists of 250 sentences (25 lists of ten sentences per list) read by a male speaker that are of approximately equal length (six to eight syllables) and difficulty (first-grade reading level) and have been digitally recorded for standardized presentation. The first 240 sentences (24 lists) were utilized in this study. The HINT uses an adaptive step procedure to estimate the reception threshold for sentences (RTS) at which sentences, embedded in uncorrelated Lou Malnati's restaurant noise, can be repeated correctly 50% of the time. The administration of the HINT requires presentation of two lists (20 sentences) for the determination of a RTS. The first sentence was presented 10 dB below the attenuator setting necessary for the noise to be presented at 65 dBA. The first sentence was presented repeatedly, increasing the level of the presentation by 4 dB, until repeated correctly by the participant. Subsequently, the intensity level was decreased by 4 dB, and the second sentence was presented. The stimulus level was raised (incorrect response) or lowered (correct response) by 4 dB after the participant's responses to the second and third sentences. The step size was reduced to 2 dB after the fourth sentence and a simple up-down stepping rule was continued for the remaining 16 sentences. The calculation of the RTS necessary for 50% sentence recognition was based on averaging the presentation levels of sentences five through 20, and the intensity of the 21st presentation, which is estimated using the response to sentence 20. For testing in noise, the R-Space restaurant noise was presented at 65 dBA. In this study, the R-Space restaurant noise was presented from all eight loudspeakers whereas HINT sentences were presented from the loudspeaker at 0° azimuth. The participant was seated in the center of the R-Space system facing the front (0°) loudspeaker, and head placement was level with the loudspeakers. Each participant was instructed to face the front loudspeaker throughout the test session and to keep his/her head in the same position. The participants were asked to repeat the test sentence exactly as heard and if unsure, take a guess. The sentence lists were randomized among participants. Participants were tested using four lists (forty sentences) for each fitting, and an average of the four lists was scored as the participant's RTS in dB.

APHAB

At this third visit, the participants also completed the APHAB and SSQ questionnaires. The APHAB questionnaire measures a participant's perception of how well he/she performs in 24 listening environments divided into four subscales (six listening conditions per subscale): EC, BN, RV, and aversiveness of sounds (AV). The participants rated the percent of problems he/she experienced in each condition using pen and paper on a seven-point assessment scale for the first-fit and programmed-fit settings. The programmed-fit problem score (in %) was compared with the first-fit problem score in each of the four subscales to determine if benefit (i.e., benefit score) was perceived by the participant from the programmed-fit compared with the first-fit.

SSQ

The SSQ was also completed and contains 49 questions concerning the hearing of speech (i.e., hearing in the presence of competing noise, hearing in the presence of multiple talkers, etc.), spatial hearing (i.e., directional and distance judgments), and other qualities (i.e., segregation of sounds, listening effort). A score was calculated for each subscale where ten was equated to less perceived disability.

Interview Questions

The participants were questioned to determine what he/she liked and disliked about each setting. The questions asked at the first and second test sessions were (a) What did you like about the hearing aids and in what specific environment(s) did you find them helpful? (b) What did you not like about the hearing aids and were there environment(s) where the hearing aids were not helpful? The following questions were asked only at the second test session to determine if each participant had a preferred setting: (a) Did you notice any difference in sound quality or performance between the current and previous settings? If yes, please describe. (b) Do you prefer one setting over another? If so, please describe.

Upon completion of the speech and subjective measures, the participant was once again seen by Investigator #1 to be fit using the other fitting (either first-fit or programmed-fit). After the fitting, the patient was scheduled to return in four weeks to repeat the same battery of tests with the different fitting setting. This visit lasted about one and a half hours.

At the fourth visit, Investigator #2 once again completed the CNC, HINT, APHAB, SSQ, and interview questionnaire for the second fitting (either first-fit or programmed fit). It was at the end of this visit that the participant could elect to purchase the hearing aids at a significantly reduced cost of $3475, and the hearings aids were then programmed using the participant's reported preferences. The cost of the hearing aids, earmolds, and remote control, if purchased in the clinic, would have been $8100. If the participant elected not

to purchase the hearing aids, he/she received $100 for his/her participation.

Data Analysis

Data distribution was explored through histograms and analyzed through the Kolmogorov–Smirnov test. Because the data were found not to be normally distributed, the median and range (Minimum and Maximum scores) were used to describe distribution of the scores for each test at each condition. Wilcoxon signed rank test was explored and used to test for significant differences between the test conditions. Effect size was explored and presented as the median of the pairwise differences and the 95% confidence interval (CI) around that. Box and whisker plots were used to graphically display and compare the distribution of scores. In each box whisker plot, the black horizontal line represents the median, and the upper and lower boundaries of the box represent the 75th and 25th percentiles, respectively. The whiskers expand 1.5 box lengths below and above the interquartile range. The circles and stars are outliers.

RESULTS

REIG Measures

Figure 2 reports the scatterplot of the individual REIG using first-fit (Figure 2A, C, and E) and programmed-fit (Figure 2B, D, and F) settings for a 50 dB SPL (Figure 2A and B), 65 dB SPL (Figure 2C and D), and 80 dB SPL (Figure 2E and F) input level and the median value (X). A value of 0 dB represents no difference between NAL-NL2 and REIG at 500–8000 Hz, below 0 dB represents underamplification, and above 0 dB represents overamplification. As can be seen in Figure 2, the median programmed-fit REIG for 50 dB SPL (Figure 2B), although within the criteria of ±5 dB at 500–4000 Hz, is below NAL-NL2 compared with the median programmed-fit REIG for 65 dB SPL (Figure 2D). On viewing the scatterplots in Figure 2A–D, it is clear that some participants were as much as 9 dB below NAL-NL2 at 4000 Hz, whereas the mismatch between NAL-NL2 for the programmed-fit settings at 65 dB SPL (Figure 2D) was no >2 dB. The inability to achieve as close a match for 50 dB SPL as the investigative team was able to achieve for 65 dB SPL could serve to underestimate the resulting difference in performance between 50 and 65 dB SPL for the two fitting strategies. That is, if the investigators could have met NAL-NL2 for 50 dB SPL as was achieved for 65 dB SPL, then the resulting median difference in performance between the default first-fit and programmed-fit might have been even greater than what was found in this study. Overall, at 50 dB SPL and 65 dB SPL, the programmed-fit more closely matched NAL-NL2 target compared with first-fit, which underamplified sound

in the mid to high frequencies. At 80 dB SPL, the programmed-fit tended to overamplify sound by about 5 dB in the mid frequencies, whereas first-fit underamplified sound in the higher frequencies.

Participant Preference between the Two Fitting Strategies

Of the 24 participants, 19 (79%) preferred the programmed-fit. Also, of the 24 participants, 16 (67%) elected to purchase the hearing aids at the conclusion of the study. Of the five who did not prefer the programmed-fit, three preferred the first-fit; however, in time, the remaining two eventually stated a preference for and used the programmed-fit. Eight elected to receive the $100 compensation. Of these eight, five preferred the programmed-fit, but felt the cost of the hearing aids, even at the reduced cost, was prohibitive.

CNCs in Quiet at 50, 65, and 80 dB SPL

Words

The median, and min and max CNC word score (in %) at 50, 65, and 80 dB SPL for the two fitting strategies is reported in the whisker box plot in Figure 3. As expected, as the input level increased for both strategies, the CNC word score increased, but for the 50 and 65 dB SPL input levels, the median CNC word score for the programmed-fit strategy was significantly greater than for the first-fit strategy. For 50 dB SPL, a median score of 64.0% (min = 22%; max = 88%) was revealed for first-fit and the programmed-fit revealed a median of 79.0% (min = 48%; max = 94%) or a median programmed-fit advantage of 15.0%. For 65 dB SPL, a median score of 90.0% (min = 60%; max = 98%) was revealed for first-fit and the programmed-fit revealed a median score of 92.0% (min = 72%; max = 100%) or a median programmed-fit advantage of 4.0%. For 80 dB SPL, a median score of 93.0% (min = 74%; max = 98%) was revealed for first-fit, and the programmed-fit revealed a median score of 94.0% (min = 72%; max = 100%) or a median programmed-fit advantage of 1.0%. A Wilcoxon signed rank test for the 50 dB SPL input level ($p < 0.001$; 95% CI of the difference between medians = 9.7–24.3%) and the 65 dB SPL input level ($p < 0.01$; 95% CI of the difference between medians = 1.7–6.3%) revealed these differences to be statistically significant.

Phonemes

In addition to words, differences in CNC phonemes were examined between first-fit and programmed-fit. Phonemes were analyzed because previous research (Swan and Gatehouse, 1995; Aarts and Caffee, 2005; Aazh and Moore, 2007; Aazh et al, 2012; Boymans and Dreschler,
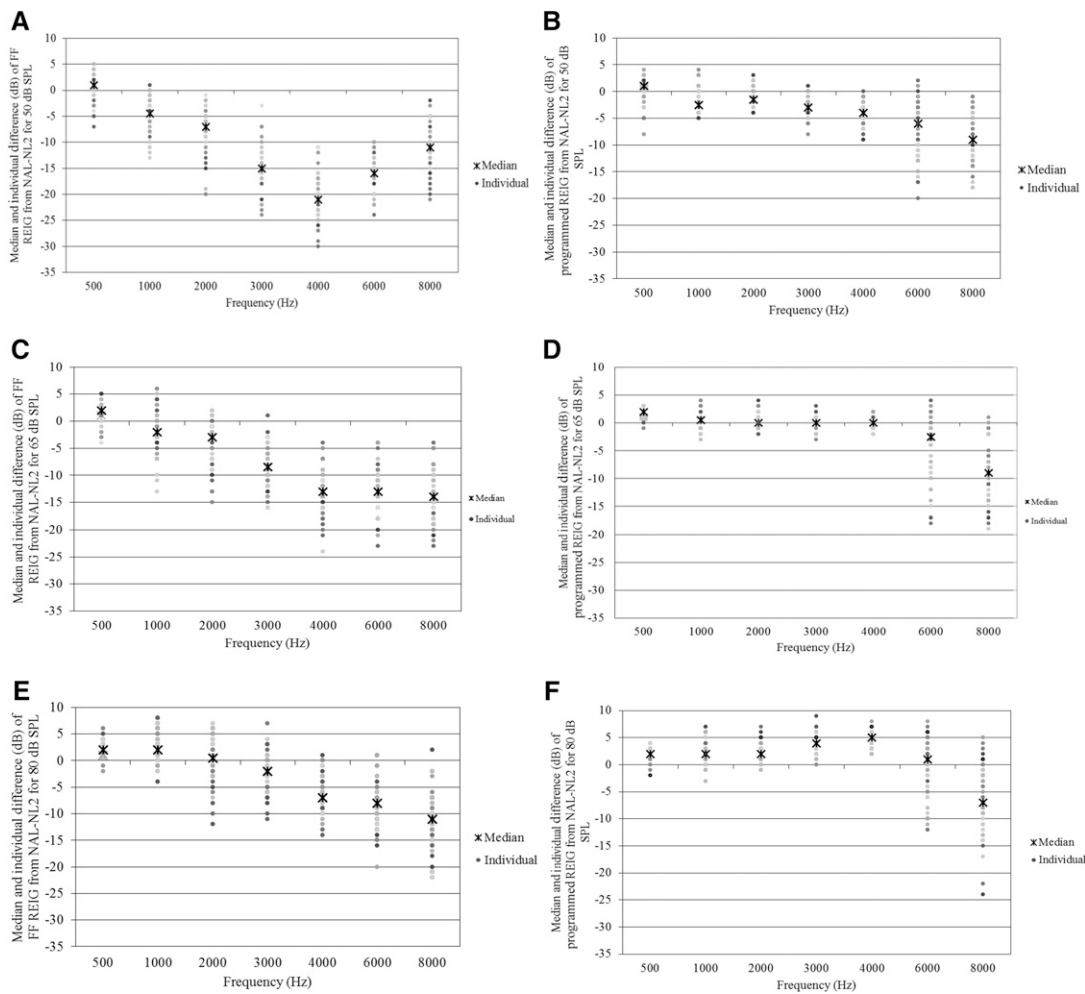
Figure 2. (A) Scatterplot of the individual REIG using the first-fit (FF) setting for a 50 dB SPL input (A), 65 dB SPL input (C), and 80 dB SPL input (E) and for the programmed-fit setting for a 50 dB SPL input (B), 65 dB SPL input (D), and 80 dB SPL input (F). A value of 0 dB represents no difference between NAL-NL2 and REIG at 500–8000 Hz, symbols below 0 dB represent underamplification, above 0 dB represent overamplification, and X represents the median REIG.

2012; Sanders et al, 2015) has reported decreased gain in the high frequencies for the first-fit, which is important for phoneme recognition. The median, and min and max CNC phoneme score (in %) at 50, 65, and 80 dB SPL for the two fitting strategies is reported in the whisker box plot in Figure 4. As expected, as the input level increased for both strategies, the CNC phoneme score increased, but for the 50 and 65 dB SPL input levels, the median CNC word score for the programmed-fit setting was significantly greater than for the first-fit strategy. For 50 dB SPL, a median score of 84.4% (min = 48%; max = 95%) was revealed for first-fit and the programmed-fit revealed a mean of 92.0% (min = 76%; max = 98%) or a median programmed-fit advantage of 7.7%. For 65 dB SPL, a median score of 96.0% (min = 83%; max = 99%) was revealed for first-fit and the programmed-fit revealed a median score of 96.7% (min = 89%; max = 100%) or a median programmed-fit advantage of 1.3%. For 80 dB SPL, a median score of 96.7% (min = 87%; max = 99%)

was revealed for first-fit, and the programmed-fit revealed a median score of 97.6% (min = 87%; max = 100%) or a median programmed-fit advantage of 1.0%. A significant median advantage of 7.7% (p < 0.001; 95% CI: 5.9–10.9%) for phonemes for the programmed-fit compared with the first-fit at 50 dB SPL was revealed by the Wilcoxon signed rank test.

95% CI for CNC Performance Differences

Figure 5 reports the 95% CI of the difference between the median for differences found to be significant in the present study. The value "0" represents no difference in percent correct (%) between first-fit and programmed-fit settings. Values to the left of "0" indicate performance was significantly better using the programmed-fit. Values to the right of "0" would have indicated better performance using the first-fit approach. Notice in Figure 5 that for the three test conditions (CNC words at 50 and 65 dB SPL and CNC phonemes at 50 dB SPL), no median
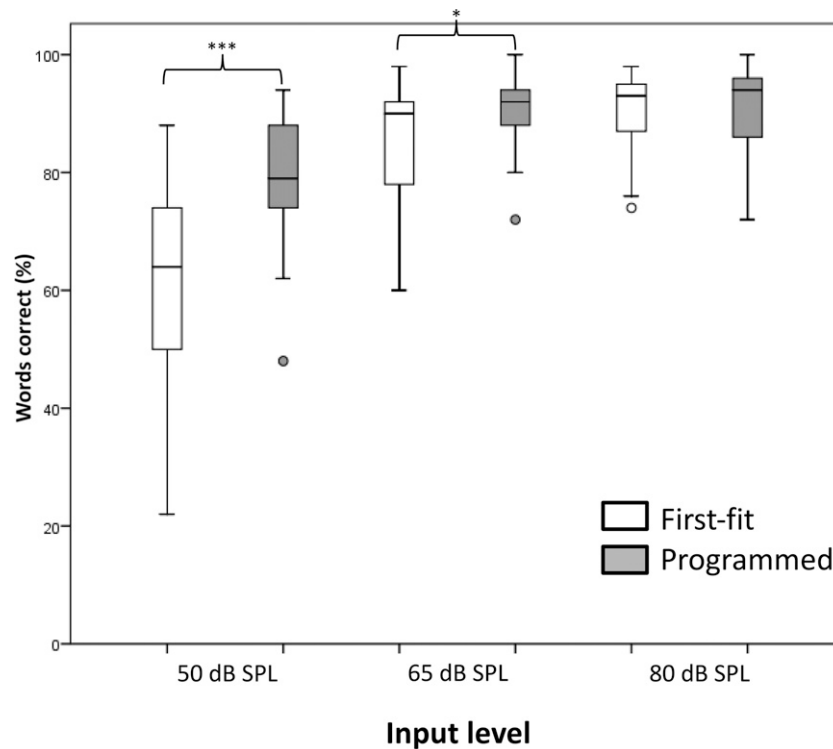
Figure 3. Whisker box plots for CNC word scores at 50, 65, and 80 dB SPL for the FF and programmed-fit settings. Note: *$p < 0.05$; ***$p < 0.001$.

difference crossed "0," the median value (symbol) was to the left of "0" as was the line representing the 95% CI surrounding the median. The fact that the symbol and line were to the left of "0" indicates that for these three conditions, 95% of scores reported better performance using the programmed-fit approach.

HINT

The median, and min and max for the HINT RTS for the two fitting strategies are reported in the whisker box plot in Figure 6. For first-fit, a median RTS of 1.94 dB (min = −2.59 dB; max = 8.59 dB) was revealed, and for the programmed-fit, a median RTS of 1.83 dB (min = −2.36 dB; max = 5.65 dB) was revealed or a median programmed-fit advantage of 0.88 dB. This difference was not statistically significant.

APHAB

The median, and min and max problem score (%) for the four subscales of the APHAB for the first-fit and programmed-fit are reported in a whisker box plot in Figure 7. The higher the bar, the more problems were perceived for the respective fitting strategy for the subscale. Of the four subscales, only the BN subscale revealed a significant difference. For the BN

subscale, the first-fit revealed a median problem score of 34.4% (min = 5.4%; max = 64.3%) whereas the programmed-fit revealed a median problem score of 22.8% (min = 6.5%; max = 83%) resulting in a median benefit score of 4.2% (min = −18.7%; max = 33.5%). A Wilcoxon signed rank test for the BN subscale ($p < 0.04$; 95% CI of the difference between medians = −0.6 to 13.2%) revealed these differences to be statistically significant. All other comparisons were not found to be significantly different.

SSQ

The median, and min and max first-fit and programmed-fit SSQ Hearing Disability Score for the SSQ subscales are reported in a whisker box plot in Figure 8. The higher the bar, the less disability was reported for the respective fitting strategy for that subscale. None of these differences were found to be statistically significant.

DISCUSSION

The primary purpose of this study was to determine if significant differences were present in speech recognition in quiet, noise, and subjective outcomes between hearing aids fit using one manufacturer's first-fit
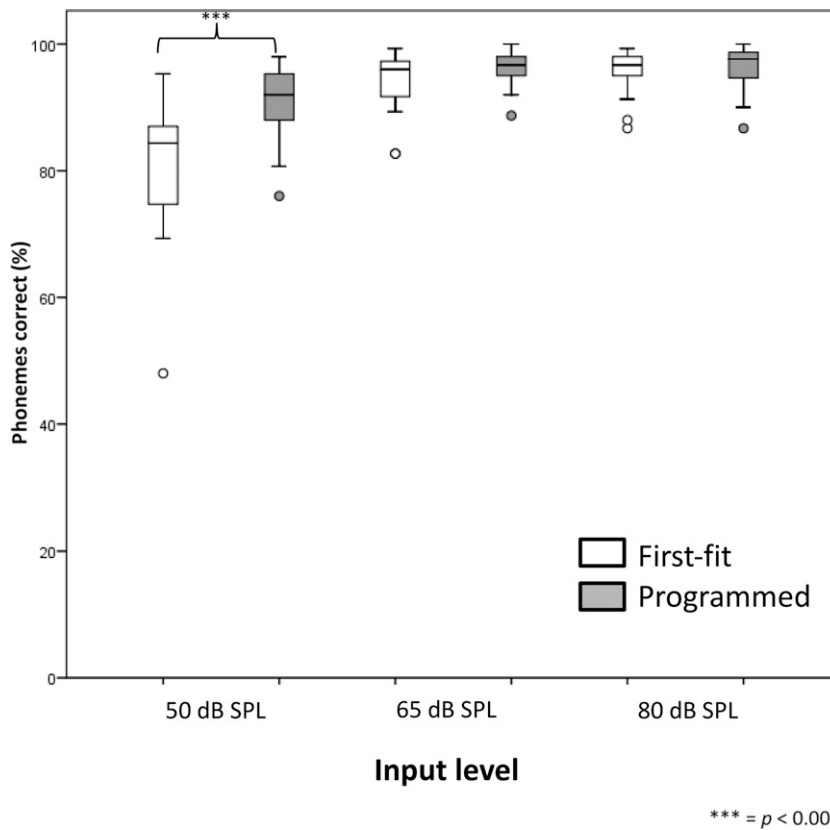
Figure 4. Whisker box plots for CNC phoneme scores at 50, 65, and 80 dB SPL for the FF and programmed-fit settings. Note: ***$p < 0.001$.

and hearing aids programmed-fit to NAL-NL2 using REM. Previous research (Swan and Gatehouse, 1995; Aarts and Caffee, 2005; Aazh and Moore, 2007; Aazh et al, 2012; Abrams et al, 2012; Boymans and Dreschler, 2012; Leavitt and Flexer, 2012; Sanders et al, 2015; Munro et al, 2016) suggests that a first-fit reduces the audibility of high-frequency signals, which could negatively impact speech recognition and patient satisfaction. Significant differences were reported on some of the current study's measures, with the first-fit performing poorer than the programmed-fit. The implications of these finding are discussed further.

CNC

Results on the CNC words revealed a significant median programed-fit advantage of 15% for a 50 dB SPL input level and a median programmed-fit advantage of 4% for the 65 dB SPL input level. Results for CNC phonemes revealed a significant median programmed-fit advantage of 7.7%. In addition, for CNC words at 50 and 65 dB SPL and CNC phonemes at 50 dB SPL, the 95% CI of the difference between medians indicated that for these three conditions, 95% of the scores revealed better performance using the programmed-fit approach. Although no previous studies have examined differences in

speech recognition for words in quiet between first-fit compared with programmed-fit, the results reported here are not surprising. When examining the scatterplots in Figure 2, it is clear that some participants were as much as 4–24 dB below NAL-NL2 at 4000 Hz for first-fit at 65 dB SPL, whereas the mismatch between NAL-NL2 for the programmed-fit setting at 65 dB SPL was no >2 dB. For 50 dB SPL, the difference at 4000 Hz ranged from 11 to 30 dB SPL below prescriptive target for the first-fit compared with a range of 0–9 dB SPL for
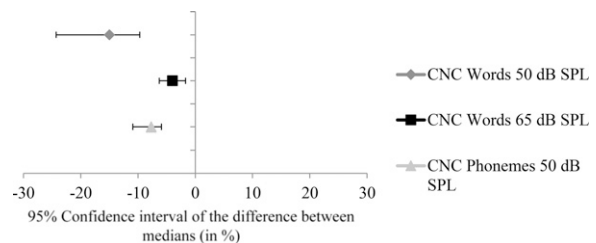

Figure 5. Ninety-five percent confidence interval of the difference between the median for differences found to be significant. The value "0" represents no difference in percent correct (%) between first-fit and programmed-fit settings. Values to the left of "0" indicate performance was significantly better using the programmed-fit setting. Values to the right of "0" would have indicated better performance using the first-fit setting.
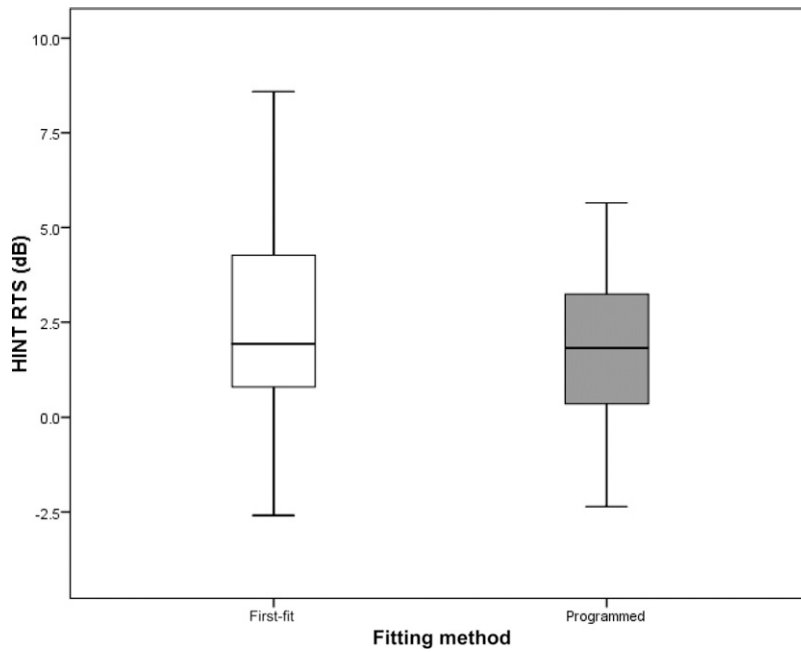
Figure 6. Whisker box plots for HINT RTS (dB) for the FF and programmed-fit settings.

programmed. It is well established that high frequencies are important for speech recognition because of access to consonants. Therefore, the programmed-fit provided improved audibility of consonants compared with the first-fit. These results are similar to previous studies (Swan and Gatehouse, 1995; Aarts and Caffee, 2005; Aazh and Moore, 2007; Aazh et al, 2012; Boymans and Dreschler, 2012; Sanders et al, 2015) in which the
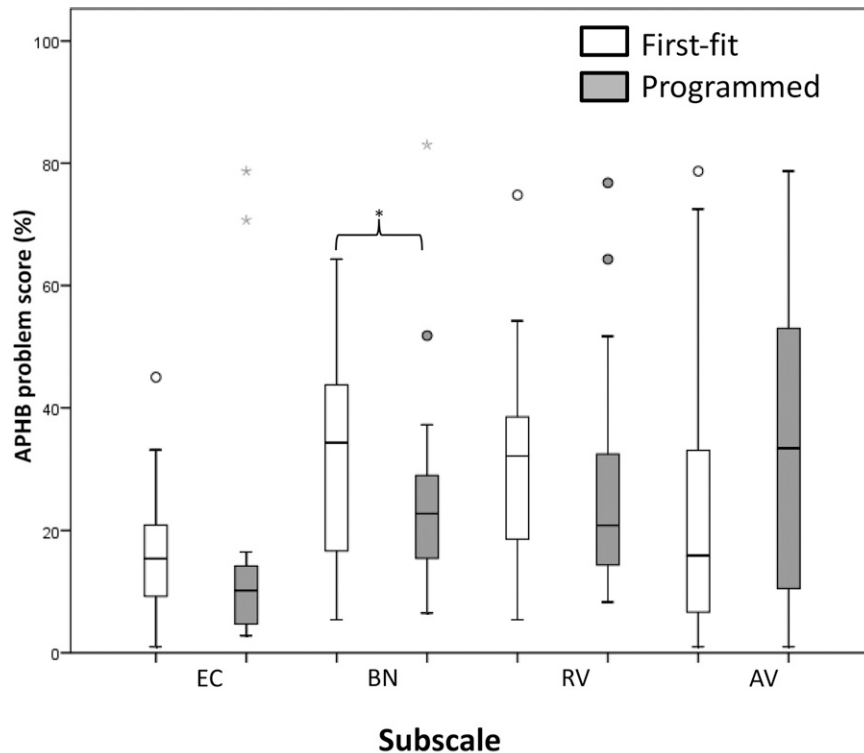
Figure 7. Whisker box plots for the FF and programmed-fit settings across the EC, BN, RV, and AV subscales for the APHAB. A higher problem score indicates more perceived problems for the respective setting. *$p < 0.05$.
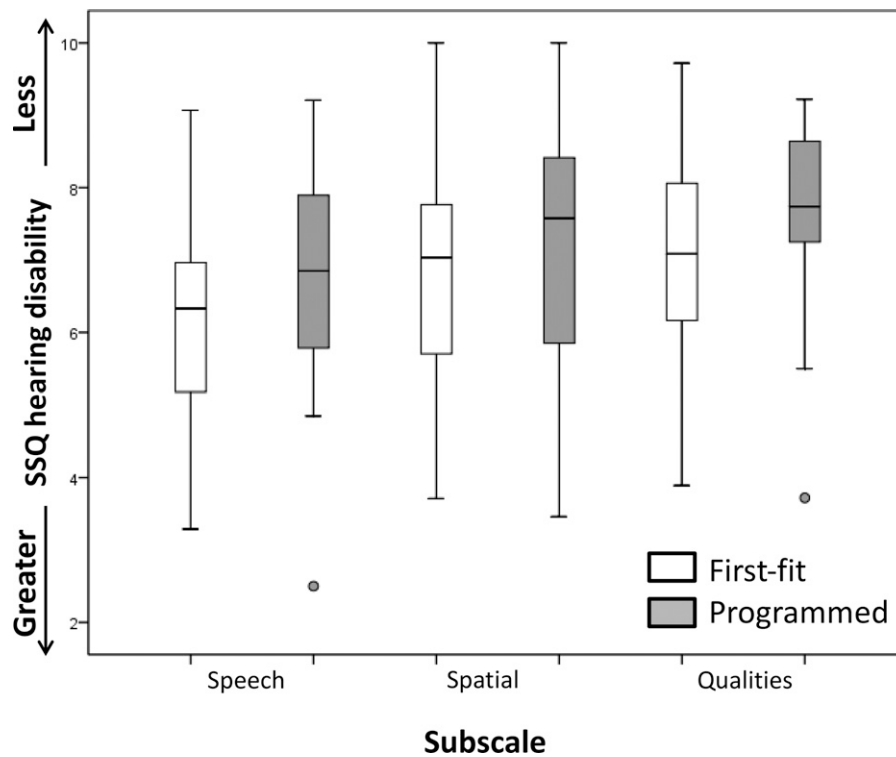
Figure 8. Whisker box plots for the FF and programmed-fit settings for hearing disability for the three subscales of the SSQ for the FF and programmed-fit settings. A higher score indicates less perceived disability.

REAR or REIG measures were below a prescriptive target for a first-fit compared with a programmed-fit.

HINT

No significant differences were reported for HINT testing between the first-fit and programmed fit, with the programmed-fit providing a 0.88 dB median nonsignificant advantage compared with first-fit. This result differs from Leavitt and Flexer (2012), who reported a mean 6.6 dB improvement in SNR using six hearing aids from six manufacturers. The results from the present study may differ because of different tests as the HINT was used in this study and the QuickSIN was used by Leavitt and Flexer (2012). Also, only one manufacturer was used in this study, and the difference between first-fit and programmed-fit may not be as robust compared with the hearing aids from the six manufacturers examined in Leavitt and Flexer (2012). Also, there may have been differences in the magnitude and configuration of hearing loss between the two studies. It is also possible that the differences in output/gain between the first-fit and programmed-fit are minimized when BN is present because of the difficulty of the listening environment and that both the noise and speech are amplified. It could be hypothesized that greater audibility of speech in quiet would be expected to provide improved speech recognition in BN, but this was not the case in the current study. The results will be further discussed in the limitations section.

Subjective Outcomes

Results for the APHAB revealed a statistically significant median advantage benefit score of 4.2% for the programmed-fit compared with the first-fit for the BN subscale. No significant differences were noted for the remaining three subscales. The SSQ did not report significant differences between the first-fit and programmed-fit. Abrams et al (2012) reported that the benefit score with the programmed-fit was significantly better than first-fit in 75% of the fits for the EC, BN, and RV subscales. This result differs in that only BN was reported to be statistically significant in the current study. It should be noted that experienced hearing aid users were examined in Abrams et al (2012), and the two aided conditions were not compared with each other, but rather compared with unaided problem score. Of the 22 participants in Abrams et al (2012), seven preferred the first-fit, and 15 preferred the programmed-fit, which is similar to what was reported in the current study.

Limitations

First, in viewing the mean hearing thresholds at 250–1000 Hz in Figure 1, it is clear that the average participant had normal low-frequency hearing. The presence of normal hearing in the low-frequency region allowed "aided" listening to be combined with unaided listening.

That is, it was not possible to differentiate unaided from aided performance. In an attempt to arrive at a better estimate of the aided difference between the two fitting strategies, the impact of low-frequency hearing is being assessed in a future study. Early findings from data in the current study appear to suggest that as hearing loss increased at 250–1000 Hz, the difference in performance between first-fit and programmed-fit increased with better performance favoring the programmed-fit setting. A test hearing aid was reprogrammed assuming a 30 dB HL hearing loss at 250–1000 Hz but maintaining the high-frequency mean hearing loss at 2000–8000 Hz reported in Figure 1. The separation between the front and back response of the directional microphones for the programmed-fit was significantly greater than the first-fit suggesting a possible relationship between the magnitude of low-frequency hearing loss and improved speech recognition in noise.

Second, only one hearing aid model from one manufacturer was assessed. Results might have been different if the investigators used another model(s) from the same manufacturer or different models from different manufacturers. The investigators, however, routinely fit hearing aids from a variety of manufacturers and it is very common to observe that the manufacturer default first-fit almost always falls below the NAL-NL1 or NAL-NL2 prescriptive target. This is especially true in the higher frequencies. Also, the inability of default first-fit to match a prescriptive target using a wide variety of hearing aids has been reported in a large number of studies as reported earlier (Swan and Gatehouse, 1995; Aarts and Caffee, 2005; Aazh and Moore, 2007; Aazh et al, 2012; Boymans and Dreschler, 2012; Sanders et al, 2015). Thus, the authors believe that the same relationship between default first-fit and programmed-fit would have been similar to that reported in the present study regardless of almost any current hearing aid model.

Third, the reader might suggest that if the investigators had used the manufacturer NAL-NL2 version of first-fit instead of the default (proprietary named first-fit) then the difference between default (manufacturer version of NAL-NL2) would have been smaller or nonexistent. While the answer is unknown, Figure 9 added the median REIG for the manufacturer first-fit version of NAL-NL2 to what was previously reported in Figure 2. It is clear in Figure 9 that even if the manufacturer version of NAL-NL2 had been used as the "default" first-fit setting, differences in audibility between this setting and programmed-fit, although maybe not as poor as the proprietary first-fit algorithm, would still have been present.

Fourth, it was somewhat surprising to find no significant difference between first-fit and programmed-fit in the HINT RTS. To investigate this further, the investigators reprogrammed the test hearing aid using the average audiogram in Figure 1 and programmed the test hearing aid using the first-fit and programmed-fit set-

tings. For both settings, the hearing aid was evaluated using the directional microphone test available with the Verifit 2 (Version 4.7.60) hearing aid analyzer using an input level of 65 dB SPL (i.e., input level of the R-Space noise used in the study) and a 0 dB SNR. Results clearly illustrated an omnidirectional front-to-back pattern. This may help explain why no difference was reported between the two fitting settings. In fact, separation between the front and back response did not occur until the input level was increased to 70 dB SPL. When the input level was increased to 75 and 80 dB SPL, the first-fit setting reverted back to an omnidirectional response whereas the programmed-fit setting maintained and increased the separation between the front and back response. In addition to varying input levels impacting HINT results, other factors may include compression differences between the first-fit and programmed-fit settings.

Fifth, no adjustments were made to first-fit or programmed fit. One could argue that over time, participants may have preferred the gain/output to be increased or decreased with first-fit. Boymans and Dreschler (2012) used a blinded randomized controlled trial to assess an "audiologist-driven fit" (i.e., fit to NAL-NL1) versus a "participant-driven fit" (i.e., fitting based on subjective preferences) fitting strategy on speech recognition in quiet and noise in 73 participants. These investigators reported greater REIG was provided using the audiologist approach, but even this was ∼9 dB below target at 4000 Hz. On the other hand, the participant driven approach yielded measured REIG that was significantly below target at almost all frequencies. The authors reported that the audiologist approach resulted in better word recognition in quiet and better subjective preferences. On the other hand, perception of loud sound was judged to be better using the participant approach. In 67% of the cases, overall preference favored the audiologist approach. In addition, Mueller et al (2008) examined the impact of the starting level of a trainable hearing aid. Participants were started either 6 dB above or below NAL-NL1 target. Results revealed that when participants were allowed to adjust the volume, those starting above NAL-NL1 target reduced the gain by 3 dB and those starting below target usually left the volume as is. Participants also preferred the loudness of the lower starting level. This suggests that patients may prefer gain at or below a prescribed target. In the current study, 79% of participants preferred the programmed-fit gain compared to the first-fit gain. Future research is needed to investigate if participants would eventually acclimatize to the prescribed target or if a reduced gain setting is preferred long-term.

CONCLUSION

Results from the present study report significantly improved audibility and performance using the
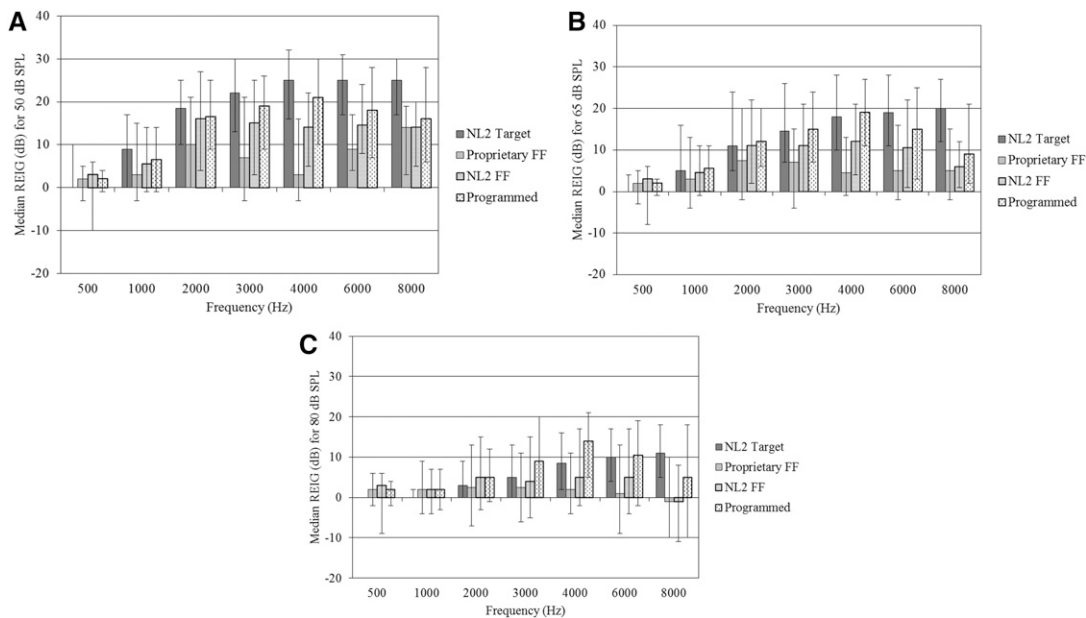
Figure 9. Median, and min and max for NAL-NL2 (NAL-NL2 target), measured REIG using manufacturer proprietary first-fit (proprietary FF), measured REIG using manufacturer proprietary NAL-NL2 (NL2 FF), and programmed-fit REIG to match NAL-NL2 at 500–8000 Hz. (A) 50 dB SPL; (B) 65 dB SPL; and (C) 80 dB SPL.

programmed-fit setting for soft and average input levels in comparison to the first-fit approach as well as 79% of the participants preferred the programmed-fit when given the opportunity to select one fitting from the other. Research has shown that lack of verification will likely lead to underamplification in the higher frequencies, which may result in less than ideal audibility of soft and average speech. This, in combination with the fact that the typical audiogram seen clinically in adult patients is a sloping configuration where hearing is poorer in the higher frequencies and where the greatest amount of gain/output is prescribed. With these factors in mind, consider the large number of patients currently wearing hearing aids fit using the first-fit, that if given the opportunity based on the results in this study, would probably elect to have the programmed-fit.

Using REM to verify hearing aid performance and adjust the gain/output to match a prescriptive target is recommended as part of Best Practice Guidelines by ASHA (1998) and AAA (2006). Based on past and the results of the current study, it is difficult to understand why 70–80% of dispensed hearing aids (Mueller and Picou, 2010) are not fit using REM to verify and match a prescriptive target. Using a first-fit alone without reprogramming using REM is fitting hearing aids blindly and not determining whether the patient is being provided appropriate audibility for soft and average speech. This is concerning as the majority of patients who are seen clinically report that speech is mumbled and unclear. The audiologist or hearing aid dispenser can never be certain if he/she is appropriately fitting the patient without

using REM. Consider an analogy of patients pursuing eyewear to obtain a prescription. To determine the optimal prescription, the specialist uses a phoropter to determine the prescription required to bring the patient's corrected vision to as close to 20/20 as possible. Patients probably would never consider purchasing eyewear without using a phoropter, and no specialist would consider prescribing lenses without using a phoropter. Just as a phoropter is expected to be used as part of best care when purchasing eyeglasses or contacts, patients provided care by an audiologist should expect REM to be used when purchasing hearing aids.

REFERENCES

Aarts NL, Caffee CS. (2005) Manufacturer predicted and measured REAR values in adult hearing aid fitting: accuracy and clinical usefulness. *Int J Audiol* 44(5):293–301.

Aazh H, Moore BC. (2007) The value of routine real ear measurement of the gain of digital hearing aids. *J Am Acad Audiol* 18(8): 653–664.

Aazh H, Moore BC, Prasher D. (2012) The accuracy of matching target insertion gains with open-fit hearing aids. *Am J Audiol* 21(2):175–180.

Abrams HB, Chisolm TH, McManus M, McArdle R. (2012) Initial-fit approach versus verified prescription: comparing self-perceived hearing aid benefit. *J Am Acad Audiol* 23(10):768–778.

American Academy of Audiology (AAA). (2006) Guidelines for the audiologic management of adult hearing impairment. Retrieved from: http://audiology.org/resources/documentlibrary/documents/haguidelines.pdf. Accessed January 4, 2016.

American National Standards Institute (ANSI). (2009) *Specifications of Hearing aid Characteristics, ANSI S3.22–2009*. New York, NY: ANSI.

ASHA Ad Hoc Committee on Hearing Aid Selection and Fitting. (1998) Guidelines for hearing aid fitting for adults. *Am J Audiol* 7(1):5–13.

Boymans M, Dreschler WA. (2012) Audiologist-driven versus patient-driven fine tuning of hearing instruments. *Trends Amplif* 16(1):49–58.

British Society of Audiology (BSA). (2007) Guidance on the use of real ear measurements to verify the fitting of digital signal processing hearing aids. http://www.thebsa.org.uk/wp-content/uploads/2014/04/REM.pdf.

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

Byrne D, Dillon H, Ching T, Katsch R, Keidser G. (2001) NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. *J Am Acad Audiol* 12(1):37–51.

Consumer Reports. (2009) Hear well in a noisy world: hearing aids, hearing protection, and more. *Consum Rep* 74(7):32–37.

Cox RM, Alexander GC. (1995) The abbreviated profile of hearing aid benefit. *Ear Hear* 16(2):176–186.

Gatehouse S, Noble W. (2004) The speech, spatial and qualities of hearing scale (SSQ). *Int J Audiol* 43(2):85–99.

G-Power[3]. (2016) G*Power:statistical power analyses for windows and mac. Retrieved from: http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3.

Holube I, Fredelake S, Vlaming M, Kollmeier B. (2010) Development and analysis of the international speech test signal (ISTS). *Int J Audiol* 49(12):891–903.

Hostler M, Hostler J, Bamford J, Whitehouse H. (2004) A "goodness of fit calculator". *BSA News* 43:32–35.

Keidser G, Dillon H, Carter L, O'Brien A. (2012) NAL-NL2 empirical adjustments. *Trends Amplif* 16(4):211–223.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116(4 Pt 1):2395–2405. Erratum (2006), 119(3):1888

Leavitt R, Flexer C. (2012) The importance of audibility in successful amplification of hearing loss. *Hear Rev* 19(13):20–23.

Mueller HG, Hornsby BW, Weber JE. (2008) Using trainable hearing aids to examine real-world preferred gain. *J Am Acad Audiol* 19(10):758–773.

Mueller G, Picou E. (2010) Survey examines popularity of real-ear probe-microphone measures. *Hear J* 63(5):27–28, 30, 32.

Munro KJ, Puri R, Bird J, Smith M. (2016) Using probe-microphone measurements to improve the match to target gain and frequency response slope, as a function of earmould style, frequency, and input level. *Int J Audiol* 55(4):215–223.

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95(2):1085–1099.

Oeding K, Valente M, Kerckhoff J. (2010) Effectiveness of the directional microphone in the Baha® Divino™. *J Am Acad Audiol* 21(8):546–557.

Peterson GE, Lehiste I. (1962) Revised CNC lists for auditory tests. *J Speech Hear Disord* 27(1):62–70.

Sanders J, Stoody T, Weber J, Mueller HG. (2015) Manufacturers' NAL-NL2 fittings fail real-ear verification. *Hear Rev* 21(3):24–26, 28, 30, 32.

Skinner MW, Holden LK, Fourakis MS, Hawks JW, Holden T, Arcaroli J, Hyde M. (2006) Evaluation of equivalency in two recordings of monosyllabic words. *J Am Acad Audiol* 17(5):350–366.

Swan IR, Gatehouse S. (1995) The value of routine in-the-ear measurement of hearing aid gain. *Br J Audiol* 29(5):271–277.

Tillman T, Carhart R. (1966) An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University auditory test No. 6. *Report SAM-TR-66-55*.