Real-life efficacy and reliability of training a hearing aid

Gitte Keidser PhD and Karima Alamudi MClAud

National Acoustic Laboratories and the HEARing Cooperative Research Centre

Corresponding author:

Gitte Keidser, NAL, 126 Greville St, Chatswood, NSW 2067, Australia

Phone: +61 2 9412 6800

Fax: +61 2 9411 8273

Email: gitte.keidser@nal.gov.au

Conflicts of Interest and Source of Funding:

All intellectual property (IP) vested in the trainable hearing aid is legally owned by HEARworks Pty Ltd. HEARworks Pty Ltd was founded in 1999 by the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation (CRC HEAR) as the IP holding and commercialization arm of the current CRC (The HEARing CRC), which funded this study. Siemens Instruments is a partner of the HEARing CRC and provided the test devices and fitting platform for the study. The first author is a co-author of the patent on trainable hearing aids but is obtaining no personal gain from the patent. The second author has nothing to declare.

Summary

We investigated if training of hearing aids result in effective and reliable outcomes. Hearing-impaired adults trained the compression characteristic from the NAL-NL2 prescription for three weeks and compared their trained response with the prescription for two weeks. The training and comparison trials were repeated. We found that for 80% of those who trained the devices away from the prescription, training was effective. Significant correlations were seen between trained variations for 63% of participants, with consistent trained variations leading to consistent preferences across trials, and vice versa. Based on our data we present a guideline for clinically managing training with clients.

Abstract

Objectives: Commercial trainable hearing aids (i.e. devices that for a period are adjusted by the user in different acoustic environments and that subsequently with changing environments automatically adapt to the user's preferred settings), are readily available; however, little information exists about the efficacy of training a hearing aid. The purpose of this study was to investigate the efficacy and reliability of training a hearing aid in everyday environments.

Design: The participants were 26 hearing-impaired volunteers with a median age of 79 years and an average pure tone average of 53 dB HL. Test devices were commercial, multi-memory, prototype devices that enabled training of the compression characteristics in four frequency bands and in six sound classes. Participants wore the NAL-NL2 prescription for three weeks and trained the devices from the prescribed response for three weeks, before comparing their trained response with the prescription for two weeks. The devices were reset to the prescription, and 19 participants repeated the training and comparison trials. During the comparison trial, participants made daily diary ratings of their satisfaction with the programs, and a structured interview was completed at the end of the comparison trial.

Results: The participants displayed different needs for changing the prescription, with more daily adjustments leading to training across more sound classes. Unreliable observations were obtained from eight participants after each of the test and retest comparison trials. Of the 10 participants who made sufficient changes to the prescription during the first trial, 80% preferred their trained response. The eight "low trainers" reported no preference, and also reported lower overall satisfaction with the device. Fewer adjustments were made during the repeat trial, resulting in less training. Significant correlations between trained variations were seen for 63% of 19 participants. Of the 10 participants who provided valid data after both comparison trials, those who trained the device consistently generally showed consistent preferences, and vice versa.

Conclusions: For those who wanted a change to the prescription, training was mostly effective. Limited data on reliability showed reasonable consistency in training outcomes and preferences. Findings, in particular on reliability, should be verified in larger populations. A guideline for how to clinically manage training with clients is presented.

Key words: hearing aids; amplification; trainable hearing aids; trainability; efficacy; reliability; speech recognition in noise; satisfaction

Introduction

A trainable hearing aid is a device equipped with user controls, which enable the wearer to manipulate one or more of the amplification characteristics (e.g. overall gain) while wearing the device in real life, and an algorithm that relates the selected amplification variation to some acoustic characteristics of the environment (e.g. sound pressure level) at the time the changes to the amplification were made. Over time, the algorithm will use the accumulated information to automatically adjust the hearing aid settings according to the wearer's preferences whenever the acoustics of the environment change. This concept, introduced by Dillon et al. (2006), overcomes problems associated with fine-tuning the hearing aid in a clinic, such as interpretation by the clinician of the wearer's complaints (Elberling & Vejby Hansen 1999; Nelson 2001) and generation of real-life situations for validation of the adjustments made. Greater satisfaction with the hearing aid, and hence more wear time, is further presumed to result from training hearing aids, as the wearer is more involved in the fitting process and hence may feel a greater sense of psychological ownership (e.g. Pierce et al. 2003) of his/her rehabilitation.

Laboratory tests have suggested that hearing aid wearers are able to use different control configurations to manipulate gain and that different gain-frequency responses can be reliably selected for different listening situations (Keidser et al. 2005; Dreschler et al. 2008). It has also been established that the selected response shape depends highly on the baseline response from which self- adjustments are begun (Dreschler et al. 2008; Keidser et al. 2008). Consequently, training must start from an appropriately prescribed response shape.

A training algorithm was first implemented in a body-worn, digital research device and evaluated in the field by Zakis et al. (2007). The device had one user control with which the wearer trained overall gain, the slope of the response, and the gain at mid frequencies relative to gain at low and high frequencies in a cyclic manner. The user pressed a voting button when the preferred setting was obtained. In three independent frequency bands, selected gain was related to the input level and the estimated modulation depth of the input signal to train gain below the compression threshold (CT), the CT, the compression ratio (CR), and the noise suppression strength. Information about the acoustic parameters and gain levels were recorded in a data logger every time the voting button was pressed. This implementation was evaluated by 13 hearing-impaired adults. Fitted initially with the NAL-NL1 prescription (Dillon 2001), the participants first trained the device in their everyday environments until 300 adjustments had been logged, a period of from one to three weeks. They subsequently compared the trained response with the NAL-NL1 response in the field for as long as it took them to log a minimum of 50 comparisons. The comparison trial was performed double-blinded as the two responses were randomly assigned to one of the two listening programs every time the device was turned on. After each comparison, the participant left the device on the preferred program and pressed the voting button. Ten out of 13 participants logged significantly more votes for the trained response than for the prescribed response, while only one participant showed a significant preference for the NAL-NL1 prescription. The findings suggested that training a hearing aid in the field, using a sophisticated paradigm, can be done effectively. However, one potential caveat in this study is that, apart from age, the participants, selected primarily for their capability and willingness to use the body-worn research device in the field, could hardly be considered representative of a typical clinical population.

Hearing aids with some training or learning capacity have been commercially available for several years. Early algorithms used information about the wearer’s volume control (VC) adjustment patterns to gradually change the default VC setting at power up (e.g. Chalupper 2006; Hayes 2007; Groth et al. 2008). The training of overall gain could be done independently for different environments through multiple memories or an environmental classifier. However, this approach is only effective if a person consistently wants the volume increased or decreased in a given environment. Changes of the same magnitude made in both directions would result in a net gain adjustment of zero. More recently, more sophisticated algorithms have been introduced to include training of the comfort-clarity balance, which affects the shape of the gain-frequency response (Unitron 2011), and training of the compression characteristics and response shape by relating preferred gain settings to the input level in four independent frequency bands (Chalupper et al. 2009).

No studies have been published that investigated whether training of the VC resulted in better outcomes, greater perceived benefit, or higher satisfaction than that obtained with the prescribed setting. Chalupper (2006) showed data on 19 experienced hearing aid users, who had trained the VC setting in different environments using a commercial device, to illustrate that individuals prefer different amounts of overall gain relative to that prescribed, and that some individuals prefer different VC settings in different environments. Mueller et al. (2008) used the same device to study overall gain preferences in the real world. In this study, 22 experienced hearing aid users were fitted with two different responses. Both responses provided the NAL-NL1 prescribed gain-frequency response shape, while one response provided 6 dB more overall gain and the other 6 dB less. The main conclusion from this study was that most hearing aid users selected an overall gain setting that differed from that prescribed, and that the selected gain was influenced by the starting level. The latter finding supports data discussed above, although it should be noted that in the Mueller et al. (2008) study, this result could be influenced by the design. In their study, the VC range was 16 dB, enabling the participant to turn gain up and down from the baseline response by 8 dB. As the two starting points were 12 dB apart, there was a common gain range of only 4 dB, which was the range of ±2 dB around the NAL-NL1 prescribed gain. Any participant for whom the preferred overall gain fell outside this range would be able to reach the preferred gain level from only one baseline response. The selected gain from the other baseline response would be much lower or higher, and as a result would appear to be influenced by the starting point.

In Chalupper et al. (2009), there is a reference to an evaluation of a commercial device that enabled independent training of the compression characteristics in three frequency bands. Outcomes from this study suggested that training was largely completed after one week, and that apart from one person who reduced gain drastically across the high frequencies, the trained response was preferred to the prescribed response, and training had no effect on speech intelligibility in quiet. The presentation, however, does not provide information about the methodology or statistical analyses of the results. Consequently, apart from the study by Zakis et al. (2007), there is currently little published data to support the efficacy of trainable devices, especially the implementations used in commercial devices.

In this study, 26 hearing-impaired participants were recruited to train a commercial prototype behind-the-ear (BTE) device equipped with a training algorithm that enabled independent training of the compression characteristics in four frequency bands and in six environmental sound classes (ESCs).   The study was designed to address the following two main questions:

1) Is training effective? That is, if a person makes consistent changes to the prescribed response through training such that the trained response varies noticeably from that prescribed, does the person prefer the trained over the prescribed response for listening in everyday environments?
2) Is training reliable? That is, if a person starts training from the same baseline response twice, will the same trained outcome be achieved and will the person's preference remain constant?

Materials and Methods

Participants

The test participants comprised five females and 21 males with bilateral sensorineural hearing loss. The total number of recruited participants was directed by the number of available trainable devices and the allocated time frame for the study. The four-frequency average (4FA) hearing loss, measured across 0.5, 1, 2, and 4 kHz, ranged from 40 to 64 dB HL, with an average of 53 dB HL. People with moderate hearing loss were targeted to reduce the number of participants who may have required fitting with large vents and for whom training consequently could be ineffective across a wide range of low frequencies. All participants had symmetrical hearing loss; defined as no more than 10 dB difference between left- and right-ear 4FA hearing loss, with a difference of 20 dB allowed at 4 kHz. The group presented with a median age of 79 years, ranging from 67 to 89 years. All participants had at least one year of experience with amplification and, on average, used their devices 4-8 hours a day.

Test devices

The test devices were prototype multi-program BTE devices from Siemens that included multi-channel wide dynamic range compression, multi-channel output limiting, feedback cancellation, noise reduction, an automatic adaptive directional microphone, environmental classification (Speech in Quiet, Speech in Noise, Music, Noise, Car Noise, Quiet), and training. The training algorithm was a revised version of SoundLearning (Chalupper et al., 2009) that enables independent training of gain below the CT and the CRs in four frequency bands (0-375 Hz, 375-1375 Hz, 1375-4635 Hz, and 4635-8000 Hz) by relating selected gain to the input level in each band (see Dillon et al., 2006). This means that the gain-frequency response shape is also trained. The compression parameters can be further trained independently in each of the six ESCs. Data logging enabled access to information about hours of usage, number of gain adjustments made, number of program changes made, percentage of time spent in each ESC, and the trained gain offset in each of the four frequency bands for each of three input levels (40, 65, and 90 dB) for each ESC.

A remote control (RC) was used to change gain by manipulating a VC, affecting gain simultaneously in all four frequency bands, or a treble control (TC), affecting only gain in the two highest frequency bands. Any adjustments made with the RC affected both aids. By default, the adjustment range was 16 dB and the step size 2 dB. The step size was reduced to 1 dB, and hence the adjustment range to 8 dB, for the second highest frequency band when using the TC. For this study, a default VC position of 75% was used, meaning that gain could be increased by up to 4 dB and reduced by up to 12 dB relative to the baseline response. The adjustment range moved with the trained settings, and was at the upper end only limited by critical gain. This means that gain could effectively be increased by more than 4 dB, or reduced by more than 12 dB, during training. Prior to this study, the expected performance of the training algorithm was verified by simulating various training scenarios and measuring the results in a coupler with all adaptive features switched off.

Custom earmolds were used with the test devices. The average vent size fitted to participants was 1 mm, with four participants fitted with vent sizes greater than 2 mm and one participant fitted with an fully occluded mold.

Fitting

Fitting took place in a sound treated room. The test devices were individually adjusted according to the NAL-NL2 prescription (Keidser et al. 2011). The gain curves for 50, 65 and 80 dB input levels were verified using the MedRx REM system and a speech-shaped noise as input. To determine whether the overall gain was acceptable and gain was balanced across the two devices, the participants were presented with recorded male discourse at 65 dB SPL from a Tannoy 800 loudspeaker positioned 1 m from the participants at 0° azimuth. A Matlab script supplied by Siemens further enabled adjustment of the maximum power output (MPO) in four frequency bands, activation and deactivation of all adaptive features at default values, training, multiple programs, downloading of logged data, and uploading of trained gain offsets. All commands in the Matlab script were applied simultaneously to both aids. After activation of the adaptive features, which could not be adjusted, the MPO setting was verified by presenting recorded stimuli consisting of noises that were low- and high-frequency weighted (traffic and a transient "pling/clang" sound), and broad-band (applause). Only two participants asked to have the MPO changed. Apart from adjustments to the overall gain or MPO, no further fine-tuning was performed.

Test protocol

Figure 1 shows an overview of the protocol. Initially, all participants had otoscopy, tympanometry, and pure-tone air- and bone-conduction audiometry completed. Impressions were taken bilaterally and a vent size common to both ears was selected. Participants were then asked to volunteer different listening conditions that they experienced on at least a weekly basis, of which two should be 'speech in quiet' and 'speech in noise' situations.

The protocol had two parts. In the first part that investigated efficacy of training, the hearing aid was fitted to the NAL-NL2 prescription. For half the participants, training was activated and the training concept explained. A demonstration of how to use the RC to train the devices in the field was completed using a series of sounds presented at different levels from a loudspeaker. The other half of the participants left with the prescribed response and training deactivated. These participants were also equipped with the RC, which gave them access to a VC if needed.

Three weeks later, the participants returned to the laboratory. For participants who had trained the devices, logged data were retrieved and saved. For these participants, the data register was subsequently cleared, the NAL-NL2 program downloaded, and training deactivated. Originally, the training period was extended for participants for whom logged data showed < 150 adjustments. However, this procedure was abandoned after the first seven participants to shorten the data collection period. Participants who wore NAL-NL2 had training activated and the training procedure demonstrated.

After another three weeks in the field, logged data were retrieved and saved for those participants who had trained the devices. All participants then had their trained gain offsets reloaded into program one (P1) and NAL-NL2 loaded into program two (P2). Only P1 of the device was equipped with the memory capacity required for the environment-specific training, so randomization of the two settings was not possible. Participants were not told the difference between P1 and P2, and were instructed to alternate between the two programs on a daily basis during the comparison trial. To help them keep track of the program changes, they were provided with a diary that showed the schedule of the days for which each program was to be worn. The diary included forms on which the participants were asked to make daily ratings of their satisfaction with the current program on a scale from 0 (not at all) to 10 (very much), overall and in individually selected situations. The approach of alternating between the two programs on a daily basis was chosen because it took the devices up to 15 seconds to settle down in the correct ESC with the trained gain offsets applied when switching to the trained response, making direct comparisons difficult, if not impractical. During the comparison trial, the RC was not accessible, and hence VC adjustments were not possible. Program changes were achieved through a button switch on the device. Wireless communication between devices meant that a press on one device activated program changes in both. One or two brief tones alerted the participants to which program they were on.

The participants returned to the laboratory two weeks later, where the logged data were retrieved and the completed diary forms were handed in. The participants also filled in an exit interview that probed estimated usage of the devices, preferred program in the field, strength of preference, satisfaction with preferred program, impression of training the devices, and the general experience with the devices.

Due to time constraints only 19 of the original 26 recruits participated in part two of the study to investigate reliability of training. These participants had the data registers cleared and NAL-NL2 downloaded in P1, and were instructed to train the devices again. After three weeks, the participants returned to the laboratory to have the device set up to compare the newly trained response with the prescribed NAL-NL2 response. Participants were again instructed to alternate between the two programs on a daily basis and were given a new schedule and a new set of diary forms. At the final appointment two weeks later, the participants had the logged data read, filled in the exit interview, and handed in the completed diary forms and test devices. Participants were paid a small gratuity for their participation.

Data processing and analysis

To study the result of training, the trained gain offsets downloaded from the data logger, corrected for real-life vent effects, were used. To determine the real-life vent effects for each individual, REIG measurements of the trained responses with all adaptive features activated were obtained using the Aurical FreeFit system and the International Speech Test Signal (ISTS; Holube et al. 2010) and a speech-shaped random noise as input. These two test stimuli were classified 100% of the time by the devices as ‘speech’ and ‘noise’, respectively. For each of the two test signals, a sequence was implemented that played the test signal at 65 dB SPL for 30 seconds, to ensure that all adaptive features, the classifier, and trained offsets were settled, immediately followed by 5 sec presentations at 65, 80, and 50 dB SPL. These REIG measurements demonstrated that for most participants (88%), vent-transmitted sound dominated the lowest frequency band, and that for a few participants (19%), the second band was dominated by vent-transmitted sound at the highest input level. For nearly all participants, the effects of the vent were the same for the two stimuli, and consequently, the combined vent effect pattern from these two measurements was applied to the gain offsets obtained in all six ESCs. That is, any recorded trained gain offset in the relevant bands was adjusted to zero to show the effect of training at the output of the hearing aid.

Three sets of data were considered in deciding whether a participant had a preference for one response over the other. First, the preferred response selected in the exit interview was noted. Second, the average daily overall satisfaction score from the diary forms was calculated for each program. A “no preference” was noted if the average scores did not differ by more than 0.2 units (arbitrary choice), otherwise the highest rated program was noted. Third, the average satisfaction score was calculated for each program across individual listening situations identified as belonging to ESCs for which the participant had trained the response. Again, a “no preference” was noted if the average scores did not differ by more than 0.2 units; otherwise, the highest rated program was noted. Where there were inconsistencies in the selected preferences across these measures, comments in the exit interview and satisfaction ratings in other ESCs were consulted to obtain a consolidated preference.

During the comparison trial, the trained response was to be worn and evaluated every second day and the NAL-NL2 response worn and evaluated every other day. For six participants, diary entries supported this behavior while data logging suggested that they had worn the device on P1 (i.e. the trained response) more than 82% of the time. As it is uncertain how the diary entries for these six participants relate to the two different programs, their data have been excluded from analyses related to the comparison trials, which means that preference data were only available from 20 and 15 participants in parts one and two of the study, respectively (cf. Figure 1).

For data analyses, non-parametric tests were generally used due to small number of observations and data being on a rating scale.

Results

Figure 2 shows the average difference between prescribed and achieved gain in the test devices for 50, 65, and 80 dB input levels. It can be seen that, on average, training started from a baseline response that matched the prescribed target well up to and including 4 kHz at the 65 dB input level. Variations at the 50 and 80 dB input levels suggest that slightly lower compression ratios than prescribed were achieved.

Training

All 26 participants completed the first training period. The average training period was 27 days, varying from 20 to 48 days, with an average wear time of 8.7 hours/day, ranging from 3.3 to 15.4 hours/day. On average, the participants made a total of 7.9 adjustments per day, with roughly twice as many VC (5.3/day, on average) than TC (2.6/day, on average) adjustments. Participants who wore their devices more hours during the day tended to make more daily adjustments, although according to Pearson's product-moment correlation analysis, the relationship did not quite reach significance (r = 0.38; p = 0.05).

From the vent-adjusted logged trained data, four parameters were extracted for each participant and ESC. They were: 1) the average gain change for a 65 dB input across the two lowest frequency bands (LF gain), 2) the average gain change for a 65 dB input across the two highest frequency bands (HF gain), 3) the average difference in gain offsets measured at 90 dB and at 40 dB input across the two lowest frequency bands (LF CR), and 4) the average difference in gain offsets measured at 90 dB and at 40 dB input across the two highest frequency bands (HF CR). Figure 3 shows the average trained variation by the participants for each of these parameters in each of the six ESCs. From this figure it can be seen that the device was trained differently in different environments. In particular, according to Kruskal Wallis tests, the trained HF gain varied significantly across the six ESCs (df = 5; p = 0.004), while the selected differences in LF gain and HF CR across environments were nearing significance (df = 5; p = 0.06). Larger variations were noted across the high frequencies, which is primarily due to vents and other leakages limiting the effect of gain adjustments at lower frequencies.

Arbitrarily, it was decided that noticeable changes were made to the baseline response (i.e. the devices were trained) if gain changes exceeded or equalled 2 dB in either frequency band, or if the difference in gain offset between the 90 and 40 dB input exceeded or equalled 4 dB. Using these criteria, it was found that two participants had not trained the devices in any of the environments, and that five had made changes in just one environment. No participants trained the response in all six environments, while three, six, seven, and three participants trained the devices in five, four, three, and two environments, respectively. Time using the device or time spent in specific environments did not affect these numbers, but those who made more daily adjustments were likely to train the device in more environments (Spearman R = 0.49, $p < 0.05$). More participants had trained the devices in speech in noise and music (18) than in noise (12) and in quiet (11). Only six participants trained the devices in speech in quiet and in car noise. Overall, participants showed different needs to train the devices in general and in specific environments.

Efficacy of training

For the 20 participants who adhered with the protocol for the comparison trial, inconsistent reports were obtained for two participants (1 and 24) for whom no conclusion about preference could be made (see Table I, Supplemental Digital Content 1, which shows the individual preferences according to the exit interview and diary forms from the comparison trial, and the comments leading to consolidated preferences). These two participants made relatively few adjustments during the training period (2.2 and 2.3 adjustments per day, respectively) and only trained the response in one ESC (speech in noise and quiet, respectively). Among the remaining 18 participants, eight preferred the trained response, two preferred NAL-NL2, and eight showed no preference. Those who reported a preference had, on average, made more daily adjustments (8.7 adj./day vs 5.1 adj./day) and had trained the devices in more ESCs (3.5 vs 2.4) than those who reported no preference.

Figure 4 shows that across environments, the 8 participants who reported no preference had, on average, made no or small changes to the trainable parameters, while the 10 participants who had a preference, either for the trained or the prescribed response, selected significant variations to the baseline response. According to Kruskal Wallis tests, the differences in trained changes were significant for LF gain (df = 2; $p < 0.0001$), HF gain (df = 2; $p = 0.0001$), and LF CR (df = 2; $p = 0.0002$), but not for HF CR (df = 2; $p = 0.46$). These observations suggest that for the 10 participants who trained the devices sufficiently away from the prescribed response and who demonstrated a preference, training was effective for 80%. A binomial test revealed that the distribution of eight preferences to the trained and two to the prescribed response was not quite significant ($p = 0.06$). It is notable that the two participants who preferred NAL-NL2 made the largest changes to the response (cf. Figure 4). They both made a very high number of adjustments during the training period (17.1 and 14.3 adjustments per day, respectively) and had trained the response in five of the six ESCs.

Between the eight participants who preferred the trained response and the eight participants who had no preference, there was a tendency for those who showed no preference to have a higher degree of hearing loss (55.7 dB HL vs 50.8 dB HL; $p = 0.13$) and to be less satisfied with the test device (a rating of 7.0 vs a rating of 8.1; $p = 0.13$). However, neither factor, together with age ($p = 0.65$), fit to target ($p = 0.51$), and hours of hearing aid use ($p = 0.57$), reached significance according to a Mann Whitney U test.

Reliability

The training period in part two was, on average, about one week shorter than the training period in part one. For the 19 participants who completed the second training period, the logged hours/day device use during the two training periods were significantly correlated (Spearman R = 0.89, $p < 0.01$). With 3.6 VC and 1.8 TC adjustments per day, on average, fewer daily adjustments were made during the second than during the first training period, and as a result the device was trained in only 32% of potential cases (participants by ESC) compared to 46% in the first training period.

For each participant and ESC, the trained parameters LFgain, HFgain, LF CR, and HF CR from the two training periods were, together with slope (the difference in gain change between the highest and lowest frequency bands for a 65 dB input), compared to each other. Examples of the correlation between trained parameters obtained in each ESC in the two training trials for four individuals are shown in Figure 5. Among the 19 participants who repeated training, significant correlations ($p < 0.01$) across ESCs and trained parameters were seen in 63%, explaining between 81% (Figure 5a) and 29% (Figure 5b) of test-retest variation. For the remaining 37% of participants, correlations were not significant. Four of these participants showed very different trained results across trials (e.g. Figure 5c), and for the remaining three participants, training was inconsistent because no effective training resulted from the second training period (e.g. Figure 5d). There was a significant and moderate correlation between the number of ESCs in which the participants had successfully trained the response in the two training periods (Spearman R = 0.52; $p = 0.03$), and as in the first training period, the number of ESCs in which the participants had trained the response was significantly correlated to the average daily number of adjustments made (Spearman R = 0.56; $p = 0.02$).

Of the fifteen participants who provided valid preference data (cf. Figure 1), four presented inconclusive preferences. Two of these were high trainers (9 and 20) while the other two were low trainers (16 and 26). Among the remaining 11 participants, three preferred the trained response, one preferred NAL-NL2, and seven had no preference (see Table II, Supplemental Digital Content 2, which shows the individual preferences according to the exit interview and diary forms from the second comparison trial). The proportion of participants who selected the trained response among those demonstrating a preference in the repeat trial was 75%, which is similar to the 80% observed after the first training trial. Note that the number of observations in the repeat trial was too small to statistically test the frequency of preferences. In agreement with the first trial, participants who had a preference, on average, trained the devices in more ESCs (2.8 vs 1.1) and had obtained greater variations of the trainable parameters from the baseline response, although a significant difference was this time only observed for the HF CR parameter (Mann-Whiney U test; p = 0.02).

Ten participants provided valid preferences after both comparison trials. Of these, six participants showed consistent preferences (four no preferences, one trained response, and one NAL-NL2 response). Five of these participants had produced significant correlations between the trained outcomes. Of the four who changed preferences, three produced non-significant correlations in their trained data. Specifically, three participants (11, 15, and 17) changed from preferring the trained response to having no preference while one participant (12) changed preference from NAL-NL2 to the trained response.

Discussion

A prototype device that enables training of the compression characteristics independently in four frequency bands and in each of six ESCs was fitted to 26 participants to study the effectiveness and reliability of training.

Training outcomes

Generally, the logged trained gain offsets adjusted for vent effects showed sensible outcomes. In agreement with previous observations (e.g. Keidser et al. 2005; Dreschler et al. 2008), participants made different changes to the gain-frequency response in different environments, which suggests that allowing for independent training in separate environments is a reasonable approach. Further, more daily adjustments resulted in trained outcomes across more environments. This is a sensible finding for two reasons: 1) participants who experience less variation in their daily environments would probably need to make fewer adjustments during a day, and therefore would only be able to train the device in few environments, at least in the shorter term; and 2) the training algorithm relies on a sufficient number of repeated adjustments being made in each environment to converge, over time, to its fully trained stage in a wide range of ESCs. The slow convergence ensures that the trained response is not incorrectly influenced by sudden large changes made to gain.

Across participants and ESCs, the trained variations to gain for a 65 dB input and the CR were, on average, smaller in this study than reported in Zakis et al. (2007). This may be because the NAL-NL2 prescription, used as baseline response in this study and adjusted based on the average trained gain variation from their study, among others (Keidser et al. 2012), was a more satisfactory starting point. Of interest is that the participants in this study, for whom the achieved CRs were, on average, lower than prescribed, often reduced the CRs further through training, which is the opposite of the trained changes reported in Zakis et al. (2007). The explanation may be in the compressor speed. Neuman et al. (1998) showed that higher CRs were better accepted when the release time was slow than fast; fast time constants were used in the test devices in this study whereas slow time constants were implemented in the research device evaluated previously. Differences in protocols and training algorithms used in the two studies may also have contributed to the different training outcomes. For example, while most participants in this study were given a fixed period in which to train the devices, all participants in Zakis et al. (2007) had to make at least 300 adjustments (votes) before proceeding to the comparison phase. This approach ensured that the convergence of the trained response was based on a sufficient number of data points. Finally, the training algorithm evaluated previously could have been more efficient as training occurred over a continuum of acoustic parameters, meaning that every adjustment contributed to training across all environments (e.g. Dillon et al. 2006; Keidser 2009), instead of each adjustment being related only to one of six environments, as was the case in the test devices used in this study.

Efficacy of training

About half of those who provided valid comparison data reported no preference for either response. This is a higher proportion than reported in Zakis et al. (2007), presumably because the protocol in their study ensured sufficient training of the devices (300 adjustments were required) as well as comparison of the trained and prescribed responses (at least 50 ratings were required). Participants reporting no preferences in this study had made fewer adjustments, resulting in trained outcomes across fewer ESCs, and hence small changes to the prescription across environments. Many of the recruited participants were pensioners who wore government-funded devices when entering the study, and many of them reported early in the study that the test devices were better than their own, although such comments should be regarded with caution (Dawes et al. 2011). It is, however, likely that some participants were genuinely satisfied with the test devices in their range of everyday environments and consequently experienced little need to deviate from the prescription. It is also possible that some participants, if given more time, would have trained the devices across more environments, which may have shifted them to one of the two preference categories.

In this study it was found that of those participants who had trained the devices sufficiently across several environments, training was effective for 75-80%, a proportion that is only slightly lower than that reported by Zakis et al. (2007). It is, however, much higher than the 18% of participants who preferred a fine-tuned response that was patient-driven, as opposed to audiologist-driven, in Boymans & Dreschler (2012). In their study, 67% preferred the response fine-tuned to the prescription by the audiologist. The discrepancy in findings may be explained by the fine-tuning procedure in this study and in Zakis et al. (2007) taking place in the participants' real-life environments, while fine-tuning in Boymans & Dreschler (2012) was performed in the clinic using video clips of situations considered relevant to the participants. Of great interest is the fact that the participants who preferred their trained responses in this study rated their satisfaction with the test devices (non-significantly) higher than did those who reported 'no preference', despite many of those reporting 'no preference' also reporting that they liked the test devices better than their own. Among those who had trained the devices, the high proportion of participants for whom training was effective is encouraging, especially given the older age of the test participants, and would suggest that a training feature is an efficient way of fine-tuning the devices away from the clinic.

It is of concern that two participants trained the devices to produce gain-frequency responses that were inferior to those prescribed. The two participants had made many adjustments during the training period and had both trained the devices away from the prescription in five of the six ESCs. The instructions to participants before the training period were to try out different settings when encountering different listening environments rather than to only make adjustments to the response if they felt a need to do so. The rationale for this approach was that if the prescribed responses were in fact preferred, then the participants would either get back to the default positions of the controls when making adjustments, or, on average, achieve zero gain offsets because no consistent variation was preferred. However, this did not happen for these two participants. It is possible that they could not clearly distinguish between responses, as changes were applied gradually, and were able to make a conscious decision about preference only when comparing two (very) different responses (e.g. Keidser et al. 2008). An unknown, but important, factor in these cases is whether the devices would have been trained at all if the participants had been asked to only make adjustments as needed. A less likely explanation is that these two participants trained the devices in environments very different to those experienced during the comparison trial. Interestingly, one of the participants (22) trained the devices in a very similar way during the second trial (the correlation between trained parameters obtained in each environment was 0.78) and again preferred the prescribed response, while the other participant (12) trained the devices differently (r = 0.39) and this time preferred the trained response. The reasons for ineffective training should be further explored in future studies.

Reliability

Some participants were very reliable in their training, others less so (cf. Figure 5). Differences in environments encountered during the two periods may have influenced the reliability of training outcomes, although participants were encouraged to focus on individually nominated situations that they commonly experienced. However, if the characteristics of background noises experienced during the first trial differed from those experienced in the second trial, then a different training outcome is plausible. Half of the inconsistent training outcomes was due to large changes being made to the response in some environments during the first trial, while no changes were evident after the second trial. Study fatigue among participants or diminished novelty of experimenting with the controls may have contributed to the lower number of daily adjustments observed during the second training period. Additionally, for many participants, the second training period took place around or immediately after Christmas, when participants may have been less focused on their tasks and may have missed their nominated environments.

Despite less training during the second training trial in general and few participants producing reliable preference data after both trials, it was encouraging to observe that those who produced consistent training outcomes also generally showed consistent preferences, just as those who produced inconsistent training outcomes mainly showed inconsistent preferences. Only two participants did not fit this model. One participant (17) had consistent training outcomes (r = 0.69) but inconsistent preferences, expressing a preference for the trained response after the first trial and no preference after the second trial. This person had a strong interest in music and had only trained the devices significantly in this ESC. Although the person trained the response in the same direction (reduced gain and compression), the variations were larger after the first trial despite similarities in the length of the training period and the number of daily adjustments across trials. Another participant (6) produced inconsistent training outcomes (r = 0.35) but expressed a consistent preference for the trained response. For this person the correlation between training outcomes would have reached significance had it not been for two outlying data points that both related to training in one ESC (speech in noise), for which the participant reduced HF gain during the first trial, but increased HF gain and HF CR during the second trial.

Overall, reliability of training should be further investigated in a larger population. It is recommended that in such studies, only participants who trained the devices sufficiently to have a preference for either response in the first trial are invited to repeat the training, and that the training period is defined by the number of adjustments rather than time. A larger and more diverse population should also be targeted in future studies, with the aim of determining a profile of those who obtain reliable training and preference results and hence would be candidates for trainable devices.

Procedural factors

A potential pitfall in this study was the inability to randomize the trained and prescribed responses to the two programs during the comparison trial. Only P1 in the devices had the capacity to contain the set of trained gain offsets for six ESCs, and therefore, the trained response was always in P1. The participants were not specifically alerted to this fact and they were assumed blinded to the setup of the two programs. However, it could be argued that the trained response was selected more often as the preferred response than NAL-NL2 simply because it was convenient to nominate ‘P1’ as the preferred program. One suggestion that preferences were not directed this way is the fact that consistent or inconsistent preferences across the two trials were in agreement with whether the devices were trained in a consistent or inconsistent manner across trials. Also, previous research has shown that preferences did not depend on which program the feature was assigned to (e.g. Surr et al., 2002).

After each comparison trial a number of observations were discarded because logged data suggested that some participants did not follow the protocol correctly, while others reported inconsistent preferences. For logistical reasons, the inconsistency between program usage and diary entries during the comparison trial was not discovered until the study was completed. The correct functionality of the logged feature was verified in the devices fitted to the implicated participants. Further, any effects of an age-related cognitive decline were dismissed. All participants, who had a long-standing record as research volunteer at our laboratory, demonstrated their cognitive fitness through the performance of many and varied tasks in the laboratory, including challenging objective tests not described in this paper. Therefore, the most likely explanation for the discrepancy in reported time on each program between diaries and logged data is that the participants inadvertently reverted to P1 on days where they were supposed to evaluate P2. Although all participants were instructed to only press the button on one device to effect program changes in both devices and to listen for the number of beeps, it is possible that some participants pressed the button on each device, changing first from P1 to P2 and then back to P1. Also, not all participants wore the test devices all day. It is possible that some of these participants started the day on the correct program, but forgot to change to P2 later in the day after the devices had been switched off. It should be noted that for participants who spent significantly more time on P1 according to their logged data, the behavior was consistent in both comparison trials.

Inconsistent preferences in the exit interview and in the diary forms may result from participants experiencing cognitive dissonance (e.g. Festinger 1957). That is, they could not really tell the two programs apart, but felt obliged to express a preference for a response. Four of the six participants who provided inconsistent reports during this study had trained the devices sufficiently in only one ESC, or not at all. It is therefore likely that they had no real preference, as noticeable differences between programs would have been limited to very specific situations, if any.

A relatively high proportion of reported 'no preferences' due to insignificant training further contributed to the final number of valid observations being smaller than anticipated. The number of recruits for this study was primarily directed by the number of available devices and allocated time frame. An a priori power analysis was not performed as no prior knowledge allowed for an estimation of the proportion of participants who would actually train the amplification to be sufficiently different to that prescribed across a range of environments. If future studies wish to verify that 80% is the true proportion of the population for whom training is effective, then power calculations on binomial distributions, using a target power of 0.8, suggest that 23 participants who will train the device are needed. If it is assumed that the proportion of recruited participants who will train the device sufficiently is 10/18, as indicated in this study, then 23*18/10 = 42 participants are required. Note that the proportion of low trainers may be reduced if more flexible periods of training and comparison are allowed for.

Managing trainable devices in the clinic

Although the number of observations in this study was small, the findings were meaningful, leading to the following recommendations for managing trainable devices in the clinic: Starting with the assumption that there are no exclusion criteria for having trainability activated, except for poor manual dexterity and low cognitive function (e.g. Erber 2003), and perhaps the fitting with open devices, with which gain manipulations could be significantly limited, it is recommended that a follow-up appointment is scheduled four to six weeks post-fitting with clients who are interested in having the training feature. At the follow-up appointment, the clinician should investigate to what extent the trained response differs from the prescribed response and determine the client's level of satisfaction with the devices. A satisfied client who has obtained significant variations to the prescribed response is likely to have done a good job training the devices and should continue wearing the trained devices as-is. Those who are satisfied, but who have made negligible changes to the prescription, are assumed to be happy with the prescription and could have training deactivated to avoid allowing inadvertent changes to the prescription in the future. As the first few adjustments can be rather explorative without a real direction, especially if the prescription is, in fact, acceptable, it is also recommended that the response is reset for these clients so that if they wish to take up training some time later, future training will start from a pure prescribed response. Dissatisfied clients who have obtained negligible variations to the prescribed response should be encouraged to continue training the devices for longer to ensure that sufficient adjustments are available to enable the algorithm to converge to a fully trained state. A second follow-up appointment should be scheduled with these clients. For those who are dissatisfied and who have made significant variations to the baseline response, training should be deactivated and the response reset to the prescription, at least until it is understood why inferior results are obtained by some hearing aid wearers.

Although the data in this study were obtained using experienced hearing aid users, there is no reason to think that new users could not as aptly manage the trainable feature under the same clinical management scheme. However, activation of trainability could be delayed for any new user who appeared overwhelmed with the hearing aid at the initial fitting appointment.

Conclusion

Participants were fitted with the NAL-NL2 prescription in a commercial prototype trainable device. Through manipulation of overall gain and gain across high frequencies, they could train the compression characteristics independently in four frequency bands and in six ESCs. The trained response was subsequently compared to the prescription, and the training and comparison trials were repeated. The following conclusions can be drawn:

1) More daily adjustments led to training across more ESCs, and different variations were, on average, made to the prescribed response in different environments.

2) The need to train the devices varied among participants. About half made insufficient changes to the responses across environments and consequently could not distinguish between the prescribed and trained responses.

3) Of those who made sufficient changes to the prescribed response in their everyday environments, training was effective for 75-80% and tended to result in higher overall satisfaction with the devices.

4) Outcomes from repeated training sessions were significantly correlated for two-thirds of the participants, with those who made consistent changes to the responses also showing consistent preferences and those who could not repeat the training outcome showing inconsistent preferences.

5) Training can be ineffective for a small proportion of people and therefore training should be clinically managed by scheduling a follow-up appointment with those who are interested in training their devices in their everyday environments.

Findings in this study, on reliability in particular, need to be verified in a larger population.

Acknowledgements

The authors would like to thank Ronny Hannemann from Siemens (Erlangen, Germany) for verifying the trainable algorithm and for providing support with the prototype devices and proprietary fitting platform. We would also like to thank our colleagues Elizabeth Convery, for assisting with the implementation of the study and for valuable comments to a previous draft of this manuscript, and Ingrid Yeend, for organising the myriad of data collected in this study.

This work was financially supported by the HEARing Cooperative Research Centre established and supported under the Cooperative Research Centres Program – an Australian Government initiative.

Preliminary findings from this study were presented at the XXXI World Congress of Audiology in Moscow, May 2012; the XX Audiology Australia National Conference in Adelaide, July 2012; and the International Hearing Aid Research Conference at Lake Tahoe, August 2012.

References

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

Chalupper, J., Junius, D., Powers, T. (2009). Algorithm lets users train aid to optimize compression, frequency shape, and gain. *Hear J, 62(8),* 26-33.

Chalupper, J. (2006). Changing how gain is selected: The benefits of combining datalogging and a learning VC. *Hear Rev, 13(13),* 46-55.

Dawes, P., Powell, S., Munro, K.J. (2011). The placebo effect and the influence of participant expectation on hearing aid trials. *Ear Hear, 32(6),* 767-74.

Dillon, H. (2001). Prescribing hearing aid performance. In H. Dillon (Ed.) *Hearing Aids* (pp, 249-61). Sydney: Boomerang Press.

Dillon, H., Zakis, J.A., McDermott, H., et al. (2006). The trainable hearing aid: What will it do for clients and clinicians? *Hear J, 59(4),* 30-36.

Dreschler, W. A., Keidser, G., Convery, E., et al. (2008). Client-based adjustments of hearing-aid gain: the effect of different control configurations. *Ear Hear, 29*, 214-227.

Elberling, C., & Vejby Hansen, K. (1999). Hearing instruments: Interaction with user preference. In A.N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (Eds) *Auditory Models and Non-Linear Hearing Instruments* (pp, 341–357). Denmark: Holmens Trykkeri.

Erber, N.P. (2003). Use of hearing aids by older people: Influence of non-auditory factors (vision, manual dexterity). *Int J Audiol, 42(2),* S21-S25.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Row, Peterson.

Groth, J., Nelson, J., Jesperson, C.T., et al. (2008). Automatic hearing instrument adjustments based on environmental listening situations. *Hear Rev, 15(4),* 40-48.

Hayes, D. (2007). Empowering the hearing aid wearer through logging plus learning. *Hear J, 60(12),* 20-25.

Holube, I., Fredelake, S., Vlaming, M., et al. (2010). Development and analysis of an International Speech Test Signal (ISTS). *Int J Audiol, 49(12),* 891-903

Keidser, G. (2009). Many factors are involved in optimizing environmentally adaptive hearing aids. *Hear J, 62(1),* 26-32.

Keidser, G., Dillon, H., Carter, L., et al. (2012). NAL-NL2 empirical adjustments. *Trends Amp, 16(4),* 211-223.

Keidser, G., Dillon, H., Flax, M., et al. (2011). The NAL-NL2 prescription procedure. *Audiol Res, 1(e24),* 88-90. Retrieved July 27, 2012 from PAGEPress, Pavia, Italy.

Keidser, G., Convery, E., Dillon, H. (2008). The effect of the baseline response on self-adjustments of hearing aid gain. *J Acoust Soc Am, 124(3),* 1668-1681.

Keidser, G., Brew, C., Brewer, S., et al. (2005). The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input. *Int J Audiol, 44,* 656–670.

Mueller, H.G., Hornsby, B.W.Y., Weber, J.E. (2008). Using trainable hearing aids to examine real-world preferred gain. *J Am Acad Audiol, 19,* 758-773.

Nelson, J.A. 2001. Fine Tuning Multi-Channel Compression Hearing Instruments. *Hear Rev, 8,* 30–35, 58.

Neuman, A.C., Bakke, M.H., Mackersie, C., et al. (1998). The effect of compression ratio and release time on the categorical rating of sound quality. *J Acoust Soc Am, 103(5),* 2273-2281.

Pierce, J.L., Kostova, T., Dirks, K.T. (2003) The state of psychological ownership: integrating and extending a century of research. *Rev Gen Psychol, 7,* 84-107

Surr, R.K., Walden, B.E., Cord, M.T., & Olson, L. (2002) Influence of Environmental Factors on Hearing Aid Microphone Preference. *J Am Acad Audiol, 13,* 308-322.

Unitron (2011). *Next. Everything you need to succeed,* p.6. Unitron hearing, publication 08-020 028-5274-02.

Zakis, J. A., McDermott, H. J., & Dillon, H. (2007). The design and evaluation of a hearing aid with trainable amplification parameters. *Ear Hear, 28,* 812-830.

Table I: Overview of the participants' reported preference in the exit interview and in diaries from the first training trial. The diaries revealed an overall satisfaction score, and the satisfaction score in individually trained sound classes. The observations that led to the consolidated preference shown in the last column are outlined in the comment column.

Participant	Exit interview	Overall satisfaction	Satisfaction in trained classes	Comment	Preference
1	Trained	NAL-NL2	NAL-NL2	Inconsistent data	?
2	No preference	No preference	No preference	Consistent data	No preference
3	Trained	Trained	Trained	Consistent data	Trained
4	No preference	No preference	No preference	Consistent data	No preference
5	No preference	No preference	No preference	Consistent data	No preference
6	Trained	Trained	No preference	Trained response was also rated highest in trained classes, but only by 0.2 units	Trained
7	No preference			No diary entries	No preference
9	Trained	Trained	No preference	Trained response rated highest in marginally trained classes not included in the 'satisfaction in trained classes' category	Trained

10	No preference	NAL-NL2		No noticeable training in any sound classes	No preference
11	No preference	Trained	Trained	Clear preference for trained response in trained classes, plus exit interview mentions marginal preference for P1 before participant reported a ‘no difference’ between programs	Trained
12	NAL-NL2	NAL-NL2	NAL-NL2	Consistent data	NAL-NL2
13	No preference	No preference	Trained	Marginal preference for trained response in one class	No preference
15	Trained	Trained	Trained	Consistent data	Trained
16	No preference	No preference	NAL-NL2	Preference for NAL-NL2 in individual classes, including those with no training (?)	No preference
17	Trained	No preference	Trained	Preference for trained response in one trained class (Music) – consistent with report in exit interview	Trained
18	Trained	Trained	Trained	Consistent data	Trained
20	Trained	No preference	Trained	Trained response generally preferred in trained classes where the HF gain was reduced – consistent with reports in exit interview	Trained
22	NAL-NL2	NAL-NL2	No preference	Trained response rated highest in some trained classes, but not	NAL-NL2

				overall	
24	Trained	NAL-NL2		Inconsistent data	?
26	No preference	No preference	No preference	Consistent data	No preference

Table II: Overview of the participants' reported preference in the exit interview and in diaries from the repeat training trial. The diaries revealed an overall satisfaction score, and the satisfaction score in individually trained sound classes. The observations that led to the consolidated preference shown in the last column are outlined in the comment column**.**

Participant	Exit interview	Overall satisfaction	Satisfaction in trained classes	Comment	Preference
1	No preference	No preference	Trained	Overall preferences were marginally for trained response, but trained response clearly preferred in trained, and marginally trained sound classes	Trained
2	No preference	No preference	No preference	Consistent data	No preference
6	Trained	No preference	No preference	Diary shows marginal preferences (0.2 units) for trained response overall and in trained sound classes	Trained
7	No preference	No preference			No preference
9	No preference	No preference	NAL-NL2	Predominant preference for NAL-NL2 in diary ratings not supported by comments in exit interview	?
10	No preference	No preference		Did not train devices	No preference
11	No preference	No preference	No preference	Consistent data	No preference

12	Trained	Trained	Trained	Consistent data	Trained
13	No preference	No preference	Trained	Trained preferred in the one sound class the response had been trained in, but overall this made no difference	No preference
15	Trained	No preference	No preference	No clear reasons for preference for trained response in exit interview	No preference
16	NAL-NL2			No supportive data from diary + trained result does not support argument for NAL-NL2 preference in exit interview	?
17	No preference	Trained	NAL-NL2	Exit interview suggests that participant could not clearly distinguish between responses	No preference
20	Trained	NAL-NL2	NAL-NL2	Inconsistent data	?
22	No preference	NAL-NL2	No preference	Marginal preference for NAL-NL2 in trained class, plus exit interview suggests a leaning towards NAL-NL2	NAL-NL2
26	No preference	NAL-NL2		Inconsistent data: Strong preference for NAL-NL2 in diary, while comments in exit interview suggest a leaning towards trained response	?

Figure legends

Figure 1: Schematic overview of the test protocol used in the two parts to study efficacy and reliability of training.

Figure 2: Average difference between achieved and NAL-NL2 target REIG for 50 dB (open triangles), 65 dB (full circles), and 80 dB (open squares) input levels measured on the 26 participants. The bars show plus and minus one standard deviation. Note that points for the three input levels are shown shifted for clarity.

Figure 3: The average trained deviations from the baseline response across participants for each of six sound classes. Positive and negative values mean that participants preferred more or less, respectively, of each parameter than prescribed. LF gain = Average gain offsets across the two lowest frequency bands; HF gain = Average gain offsets across the two highest frequency bands; LF CR = Difference in gain offsets at 40 and 90 dB input levels averaged across the two lowest frequency bands; and HF CR = Difference in gain offsets at 40 and 90 dB input levels averaged across the two highest frequency bands. The boxes show plus and minus one standard error and the bars show plus and minus 0.95 times the standard deviation.

Figure 4: The average trained deviations from the baseline response across environments by those who had no preference (N = 8), preferred the trained response (N = 8), and preferred NAL-NL2 (N = 2). Positive and negative values mean that participants preferred more or less, respectively, of each parameter than prescribed. LF gain = Average gain offsets across the two lowest frequency bands; HF gain = Average gain offsets across the two highest frequency bands; LF CR = Difference in gain offsets at 40 and 90 dB input levels averaged across the two lowest frequency bands; and HF CR = Difference in gain offsets at 40 and 90

dB input levels averaged across the two highest frequency bands. The boxes show plus and minus one standard error and the bars show plus and minus 0.95 times the standard deviation.

Figure 5: The relationship between the test and retest trained variations from the baseline response for the trainable parameters in six sound classes for four different individuals, who showed: a) a strong significant correlation; b) a moderate, but significant correlation; c) a non-significant correlation; and d) a non-significant correlation due to a low number of adjustments being made during the second training period. The full line shows the regression line and the broken lines show the 95% confidence bands.

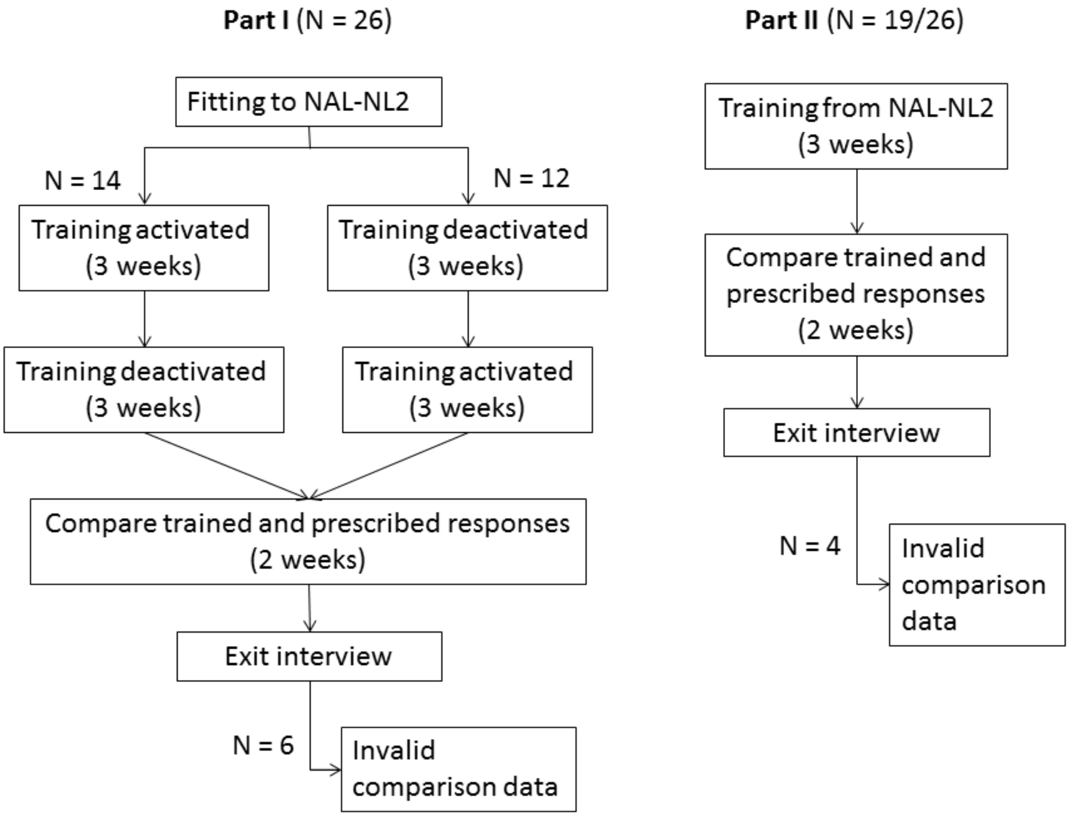

Part I (N = 26)
Fitting to NAL-NL2
N = 14
N = 12
Training activated (3 weeks)
Training deactivated (3 weeks)
Training deactivated (3 weeks)
Training activated (3 weeks)
Compare trained and prescribed responses (2 weeks)
Exit interview
N = 6
Invalid comparison data
Part II (N = 19/26)
Training from NAL-NL2 (3 weeks)
Compare trained and prescribed responses (2 weeks)
Exit interview
N = 4
Invalid comparison data

Figure 1

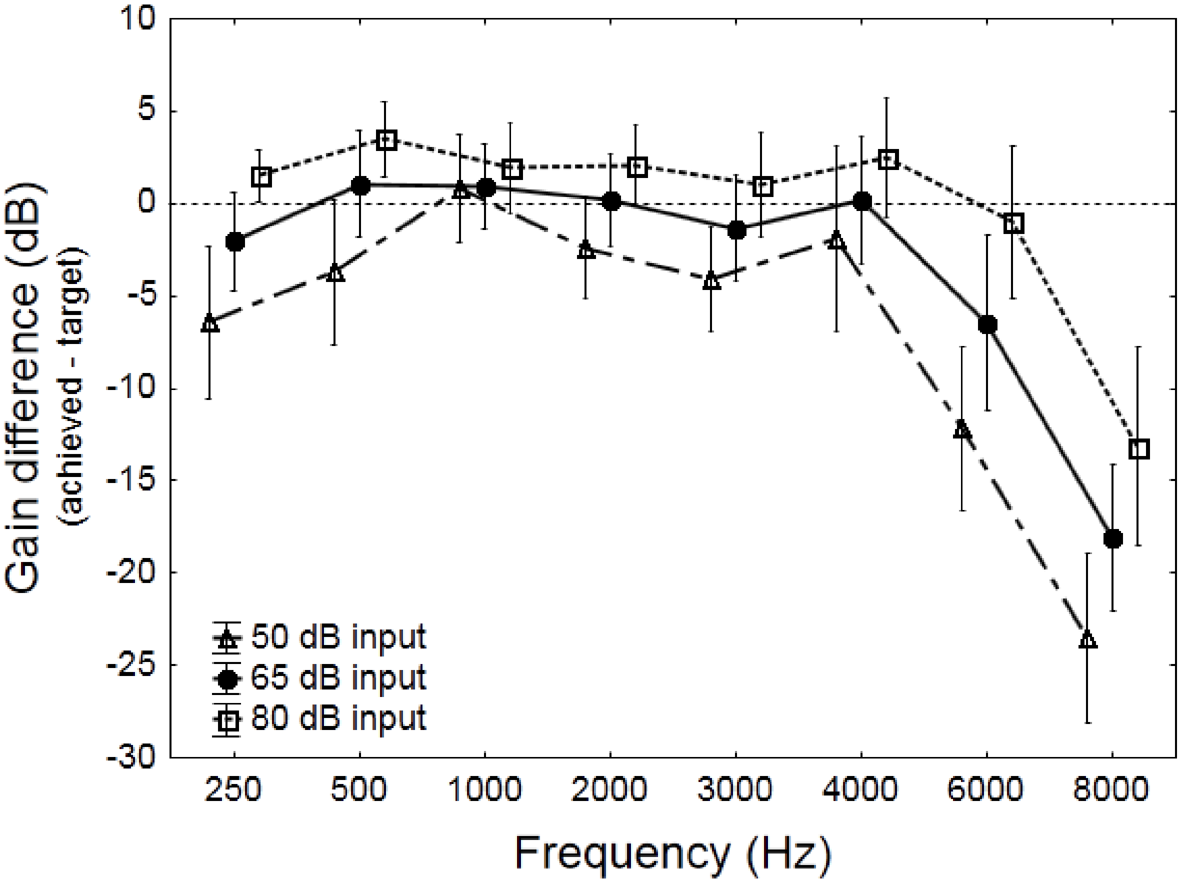
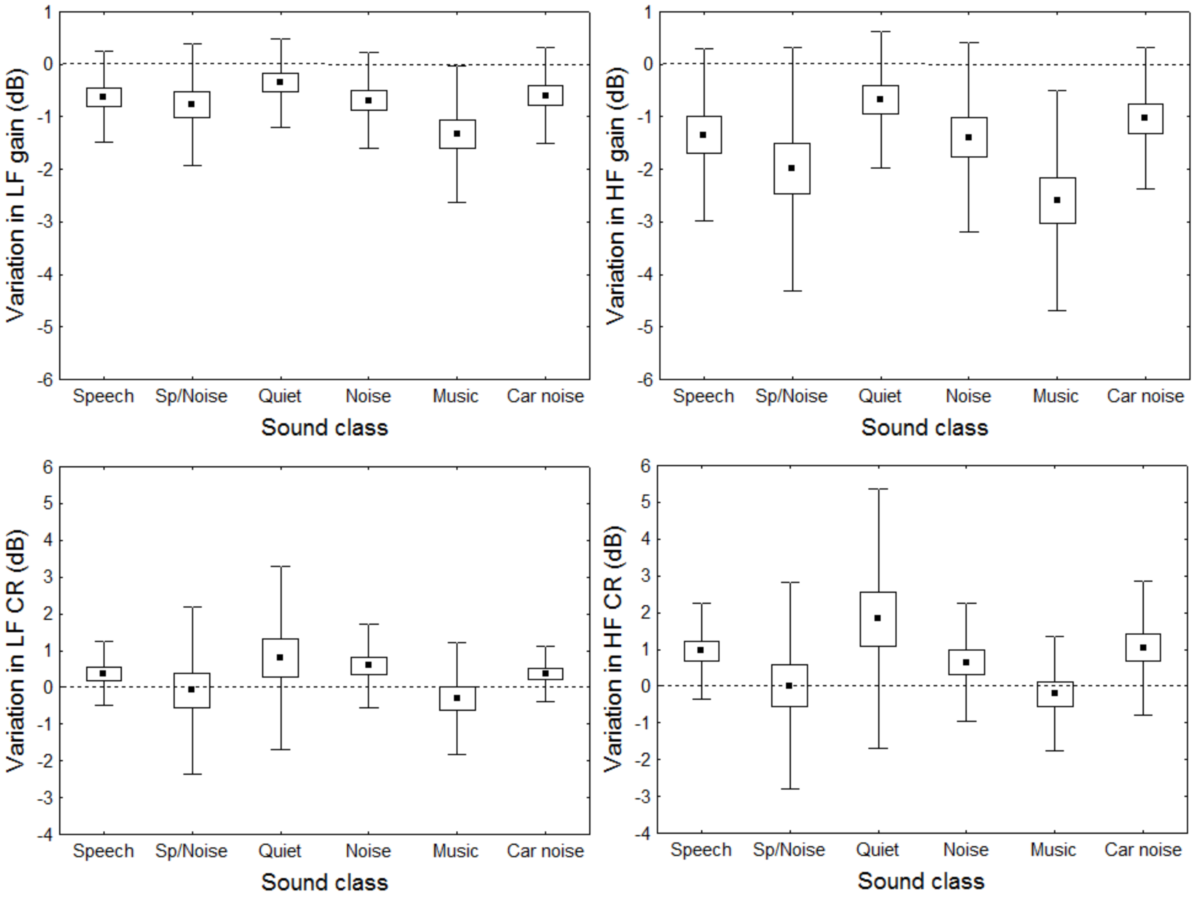

10
5
0
-5
-10
-15
-20
-25
-30
Gain difference (dB)
(achieved - target)
50 dB input
65 dB input
80 dB input
250
500
1000
2000
3000
4000
6000
8000
Frequency (Hz)

Figure 2



Variation in LF gain (dB)
1
0
-1
-2
-3
-4
-5
-6
Speech
Sp/Noise
Quiet
Noise
Music
Car noise
Sound class
Variation in HF gain (dB)
Variation in LF CR (dB)
6
5
4
3
2
1
0
-1
-2
-3
-4
Variation in HF CR (dB)

Figure 3

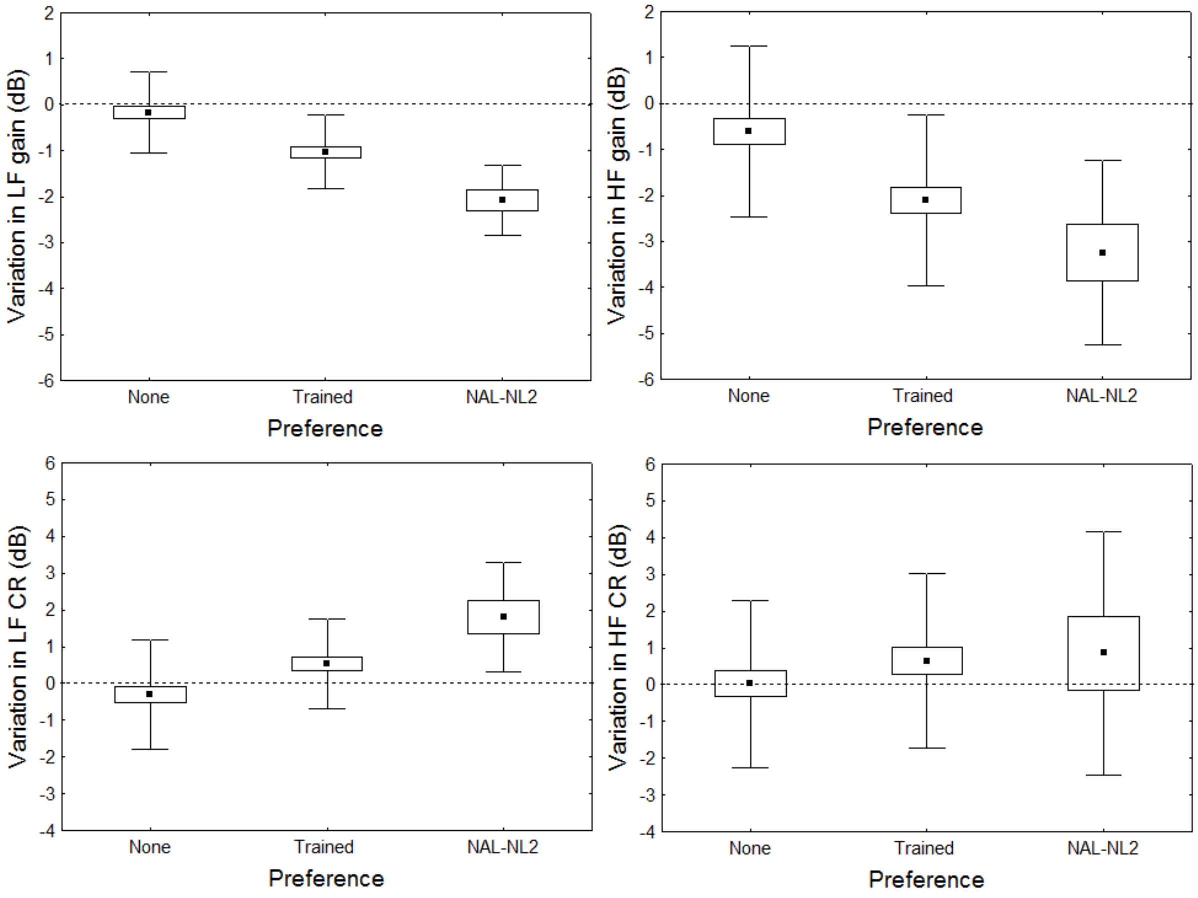

Variation in LF gain (dB)
2
1
0
-1
-2
-3
-4
-5
-6
None
Trained
NAL-NL2
Preference
Variation in HF gain (dB)
2
1
0
-1
-2
-3
-4
-5
-6
None
Trained
NAL-NL2
Preference
Variation in LF CR (dB)
6
5
4
3
2
1
0
-1
-2
-3
-4
None
Trained
NAL-NL2
Preference
Variation in HF CR (dB)
6
5
4
3
2
1
0
-1
-2
-3
-4
None
Trained
NAL-NL2
Preference

Figure 4

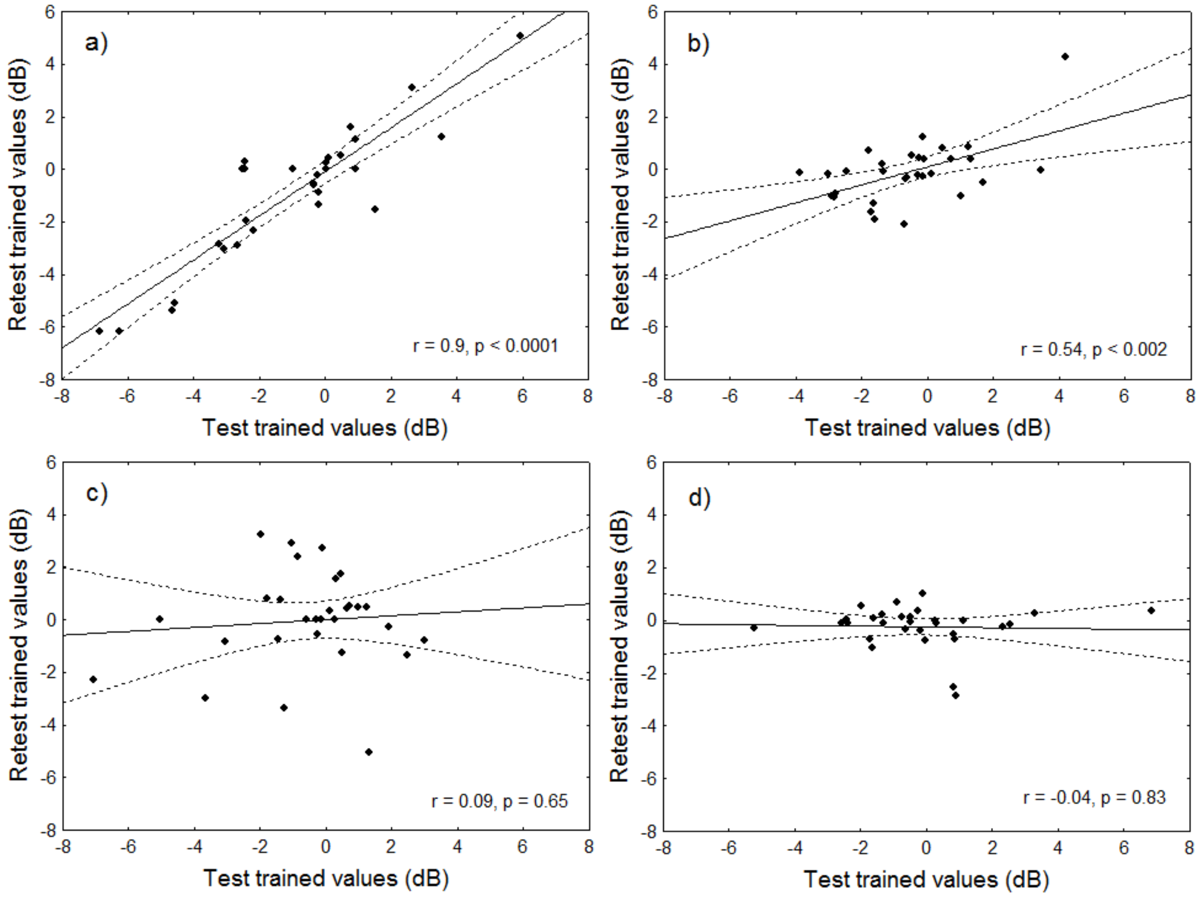

a)
b)
c)
d)
r = 0.9, p < 0.0001
r = 0.54, p < 0.002
r = 0.09, p = 0.65
r = -0.04, p = 0.83
Retest trained values (dB)
Test trained values (dB)
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8

Figure 5