Client-Based Adjustments of Hearing Aid Gain: The Effect of Different Control Configurations

Wouter A. Dreschler,[1] Gitte Keidser,[2] Elizabeth Convery,[2] and Harvey Dillon[2]

Objectives: **Facilitating the fine-tuning of advanced hearing aids requires information about the acoustic environment. The concept of a "trainable" hearing aid may provide a more direct approach to hearing aid fine-tuning if the aid user is allowed to control the most important fitting parameters in his/her own acoustic environments.**

Design: **In a laboratory study, the concept of self-adjustment of the gain-frequency response was tested by 24 hearing aid users using four different controllers with a limited number of control functions. Research questions focused on the reproducibility of the fine-tuned responses, the efficiency of the control configurations, and the effects of the control configuration on the end results of the fine-tuning process.**

Results: **The subjects were able to provide systematic and reproducible feedback with respect to their preferences in different acoustic conditions presented audiovisually, achieving an average short-term test/retest standard deviation value of 2.8 dB. Two of the control configurations, featuring volume/slope and volume/bass/treble keys, were found to be more time-efficient and reliable, and were also preferred by 86% of the subjects. Although the control configuration did not have a strong influence on the end result, the gain-frequency response from which the subjects started their adjustments was found to have a significant effect on their preferred settings.**

Conclusions: **Client-based adjustments of hearing aid gain provide a reliable method of individual fine-tuning. The results also showed that a biased correction of amplification is reached via self-adjustment within one session, which reduces the effectiveness of fine-tuning in a traditional clinical setting.**

(Ear & Hearing 2008;29;214–227)

INTRODUCTION

The fitting of modern hearing aids is becoming increasingly complex. One reason for this trend is that modern hearing aids aim to compensate for more than the loss of audibility. New techniques have been introduced that try to compensate for suprathreshold deficits using nonlinear approaches in multiple bands, either to enhance the speech signal or to reduce the amount of background noise. Traditional prescriptive formulae for determining the gain-frequency response of a hearing aid, such as POGO (McCandless & Lyregaard, 1983) and NAL-RP (Byrne & Dillon, 1986), have been replaced by such nonlinear formulae as NAL-NL1 (Byrne, et al., 2001) and DSL (i/o) (Cornelisse, et al., 1994). Evidence-based procedures for fitting noise reduction or signal enhancement algorithms are not yet available, and consequently fitting of these features is guided either by trial and error, or by proprietary fitting rules for specific devices that are not publicly available.

Another reason for the increase in hearing aid fitting complexity is that modern hearing aids allow activation of different processing strategies for different acoustical environments, as it is widely accepted that hearing aid users prefer different responses for different listening situations (Keidser, 1995; Keidser, et al., 2005). Unlike the gain-frequency response of a hearing aid, which is traditionally determined by an individual's audiometric data, it has been suggested that modern hearing aid features should be fitted according to the characteristics of the acoustic environment (Kiessling, 2001). The adaptation of the device to different acoustical environments is implemented either under user control (Fabry, 1996; Goldstein, et al., 1991; Sandlin & Meltsner, 1989; Stypulkowski, et al., 1992) or automatically (Büchler, et al., 2005; Powers & Hamacher, 2002; Schum, 2005).

Because of the incompleteness of prescriptive formulae for nonlinear processing and a lack of knowledge of how amplification should optimally adapt to different acoustical "soundscapes," active feedback from the hearing aid user with respect to his/her amplification preferences for different acoustical environments is required. Several approaches have been presented, each of which has a number of advantages and disadvantages.

For most hearing aids, the fitting process starts with a set of "first fit" parameters that are based on the audiometric data, a rough definition of the acoustical environments that the hearing aid user will probably encounter, and some proprietary fitting

---

[1] Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; and [2] Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovations, National Acoustic Laboratories, Sydney, Australia.

0196/0202/08/2902-0214/0 • Ear & Hearing • Copyright © 2008 by Lippincott Williams & Wilkins • Printed in the U.S.A.

rules designed by the manufacturer that use fixed deviations from the default program for listening environments other than a speech-in-quiet condition. After a trial period that may last several weeks, the audiologist may use a "complaint-driven" fine-tuning procedure to reduce any negative experiences encountered during this time (e.g., Jenstad, et al., 2003; Kuk, 1999; Nelson, 2001). The advantage of this approach is a uniform starting point that can be improved upon if the manufacturer actively seeks to use the experiences of large numbers of hearing aid users to refine the proprietary fitting strategy. The disadvantages are the large interindividual variation across subjects, and the poor definition of acoustic environments that makes the method insensitive to subtle individual needs for specific situations. New features in hearing aids allow datalogging to refine the individual feedback from the trial period (e.g., Flynn, 2005), but a complaint-driven approach still works on the basis of complaints. New users of hearing aids may be disappointed by the difficulties they encounter during the trial period and may give up on amplification altogether. They also need to return to the hearing clinic to have their complaints rectified.

Another approach to hearing aid fine-tuning uses well-defined sets of speech and/or noise stimuli that can be used during the initial fitting session to improve the first fit parameters provided by a generic or proprietary prescription (e.g., Moore, et al., 1998; Russ & Olsen, 2001). The stimuli can be recorded on CD or DVD (to add visual cues) to simulate specific listening environments. The advantage of this fitting strategy is a better tailoring of the hearing aid settings to specific sounds. However, the approach is time-consuming and it is not easy to reproduce realistic situations in a clinical setting, although this may change in the near future by using multiple loudspeakers to create a spatial sound presentation (Howes & Olsen, 2006; Nilsson, et al., 2005; Revitt, et al., 2002). Still, clinical practice shows that it is impossible to reproduce in the clinic or laboratory all situations that are relevant for the individual user. Therefore, this method is also relatively indirect.

A complication of both approaches is that the number of fitting parameters is large, and it is difficult to make systematic comparisons across several dimensions simultaneously. Research on multidimensional pattern search techniques may result in clinically applicable fine-tuning methods (e.g., Franck, 2004; Franck, et al., 2004), but the precise fitting of additional programs will need considerable time and effort.

Future developments may facilitate a more direct approach to hearing aid fine-tuning if the aid user is allowed to control the most important hearing aid parameters in his/her own acoustic environments. This requires a more refined control of the amplification parameters; that is, more than overall gain and the choice between different programs as in existing devices. Such an extension needs control functions that allow the subject to adjust the many hearing aid parameters in a practical way. Whether relatively naive hearing aid users will be able to use such control functions in a reproducible way, and whether self-adjustments in fact lead to more benefit from the hearing aid, are current topics for investigation.

A further step is the concept of a "trainable" hearing aid (Dillon, et al., 2006; Zakis, et al., 2007). If technology allows, hearing aid users may be equipped with a training unit for the first few weeks of aid use to train the hearing aid to their individual listening preferences in a wide range of acoustic environments. Such a concept may circumvent a number of the disadvantages inherent in the current fitting approaches, and includes the user, his/her personal preferences, and his/her specific listening situations in the fine-tuning process of the hearing aid. The same technique may be used if the hearing aid setting must be changed, due either to a change in the hearing loss or in the individual's most frequently encountered acoustical environments (e.g., after a change in employment). Important prerequisites of such an approach are that the hearing aid user will be able to give detailed feedback about his/her listening preferences, that the feedback is consistent, and that the individually fine-tuned solutions are significantly better accepted than a hearing aid whose parameters have been set according to a fixed fitting rule followed by a dispenser-driven fine-tuning process in the clinic.

This study was conducted as part of a research line within the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation and the National Acoustic Laboratories (NAL) to develop a trainable hearing aid and to test a number of aspects related to the applicability of the concept. In this laboratory study, the listener was given access to different sets of gain controls, and, while listening to selective sound samples that represent real-life listening situations, the subject was able to adjust the amplification characteristics from different baseline responses to the preferred settings. The experimental questions of this study are as follows:

1. Is the hearing aid user able to provide systematic and reproducible feedback with various control configurations with respect to his/her preferences in different acoustical conditions as simulated in a laboratory environment?

2. Given a limited number of control functions, which of a set of four different controllers is most efficient, based on subjective reports and the fine-tuning effort required to reach the preferred response?
3. Is there an effect of the control configuration on the end result of the fine-tuning process?

MATERIALS AND METHODS

Subjects

Twenty-four subjects with symmetrical (<10 dB difference between left and right ear thresholds between 500 and 6000 Hz, with a 15 dB difference at one frequency above 1500 Hz accepted) mild to moderate sensorineural hearing loss participated in the experiment. The subjects were selected on a voluntary basis and were paid for their participation. The three frequency average (average of hearing threshold levels at 500, 1000, and 2000 Hz) ranged from 17 to 57 dB HL, with a mean three frequency average of 39 dB HL. To test the effect of audiometric configuration, the subject group comprised seven subjects with flat audiograms (audiometric slopes between 500 and 4000 Hz shallower than 6 dB/octave), nine subjects with gently sloping audiograms (slopes between 6 and 15 dB/octave), and eight subjects with steeply sloping audiograms (slopes steeper than 15 dB/octave). Figure 1 shows the average threshold data for the three groups of subjects.

The ages of the 16 male and eight female subjects ranged from 38 to 85 yrs, with a median age of 74 yrs. For the purpose of testing for age effects, "younger" subjects are defined as being ≤74 yrs of age (N = 12), whereas "older" subjects are above 74 yrs of age (N = 12). Eighteen subjects were experienced users of amplification (>1 yr of experience), whereas six were relatively new to hearing aids (<1 yr of experience).

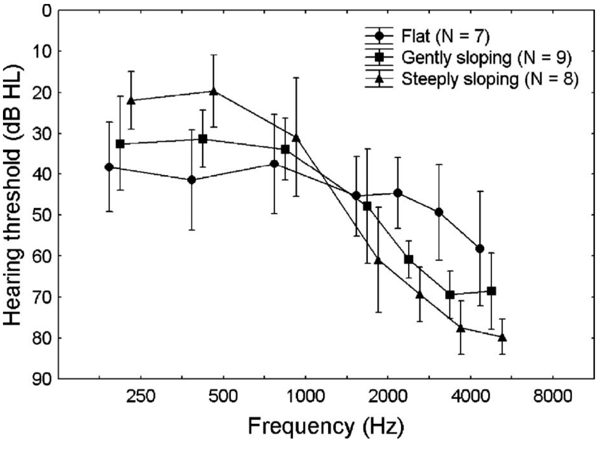

Fig. 1. Mean hearing threshold levels and standard deviation values of the three audiometric subgroups: flat audiograms (N = 7), gently sloping audiograms (N = 10), and steeply sloping audiograms (N = 7). Data are shown slightly shifted relative to the x axis for clarity.

Experimental Set-up

In a laboratory setting, the subjects were instructed to fine-tune the gain-frequency responses of a variety of stimuli according to their individual preferences. The specific instructions are shown in Appendix A. The stimuli were presented audiovisually. Testing was conducted unaided to avoid the confounding effects of the signal processing in the subjects' hearing aids and the acoustic properties of the earmolds. Subjects were seated at a table directly opposite to an 80 cm Sony television monitor, which was positioned at head level 1.8 m away. Two B&W 600 series loudspeakers were placed on opposite sides of the television monitor, facing the subject from an angle of approximately ±20°. The test stimuli originated from a Panasonic AG-DV2700 digital video cassette player and were directed through a dual channel preamplifier built in-house, three 1:1 transformers, two Behringer Ultra-Curve Pro DSP 8024 digital equalizers, two remotely controlled attenuators built in-house, and a Sony TA-F242 integrated stereo amplifier to the loudspeakers. The digital equalizers each enabled gain changes of ±16 dB in 0.5 dB steps overall (master gain) and at each one-third octave frequency from 20 to 20,000 Hz. One equalizer was used to implement the individually calculated baseline responses (see below) corrected for the overall equipment response, and the other digital equalizer was remotely controlled using a Swann Multi keypad. The keypad could be configured to control different gain parameters via a Pentium II 2.66 MHz (Windows 98) computer and a software program that was developed specifically for the experiment.

The equipment was calibrated before each test appointment by ensuring that the combined speaker output was ±1 dB of the real-life sound pressure levels of the stimuli when measured from the subject's head position.

Controllers

Four control configurations (A, B, C, and D) were implemented in the Swann Multi Keypad and evaluated. The parameters that could be controlled by the subjects are listed for each of the controllers in Table 1. Controller A enabled the subjects to control two parameters, whereas the remaining controllers allowed control of three parameters. The controller functions changed the amplification curve in dis-

TABLE 1. Summary of the control functions implemented in the four experimental control configurations (A, B, C, and D)

	Controller			
	A	B	C	D
Volume	x	x		x
Tone balance 1	x	x		
Tone balance 2		x		
Bass			x	x
Mid			x	
Treble			x	x

Note that controller A is a two-function controller, while controllers B, C, and D are three-function controllers. The descriptions of the controllers' functions reflect the controller labels that were provided to the subjects.

crete steps at 400, 1250, and 4000 Hz. For frequencies below 400 or above 4000 Hz, the gain change was equal to the gain changes at 400 or 4000 Hz, respectively. For frequencies between 400 and 1250 Hz and between 1250 and 4000 Hz, the gain changes were interpolated linearly (in decibel) on a logarithmic frequency scale.

For each labeled controller function, the effect could be changed in discrete steps in two directions using plus and minus keys. Each key press of the volume control (controllers A, B, and D) resulted in an overall gain increase or decrease of 2 dB. Each key press of the "tone balance 1" control (a slope control used in controllers A and B) resulted in a gain reduction of 2 dB at 400 Hz and a simultaneous gain increase of 2 dB at 4000 Hz (or vice versa). In controller B, a second response shape control ("tone balance 2") was used to reduce the gain by 1 dB at 400 and 4000 Hz, while simultaneously increasing the gain by 2 dB at 1250 Hz (or vice versa). The bass, midfrequency, and treble functions used in controllers C and D decreased or increased the gain by 2 dB per key press at 400, 1250, and 4000 Hz, respectively. There was no way of informing the subject if the limit of 16 dB in either direction had been reached, but no further changes were made to the gain-frequency response to key presses beyond these boundaries.

Test Signals

Six test signals were selected to represent a variety of acoustic environments, with their real-life long-term equivalent root-mean-square (Leq) presentation levels shown in parentheses below. Three of the test signals were recordings containing predominantly speech, but in different conditions:

1. A female talker reading a passage from a book in a quiet room with low reverberation (65 dB SPL);
2. A dialogue between a male and a female talker using raised vocal effort in a room with relatively high reverberation, with a large variation in intensity level over the course of the sample (75 dB SPL); and
3. A female talker speaking from the back seat of a car that was in motion (73 dB SPL).

The other three signals were nonspeech situations, selected on the basis of their different frequency spectra:

4. Ocean waves, producing a relatively constant sound with dominant low-frequency components (74 dB SPL);
5. A small group of people applauding in a medium-sized lecture room, producing a sound with dominant middle frequencies (73 dB SPL); and
6. Vegetables sizzling in a frying pan, characterized by relatively strong level variations and a frequency spectrum with dominant high frequencies (73 dB SPL).

The spectrum levels of the speech and nonspeech signals are presented in Figure 2.

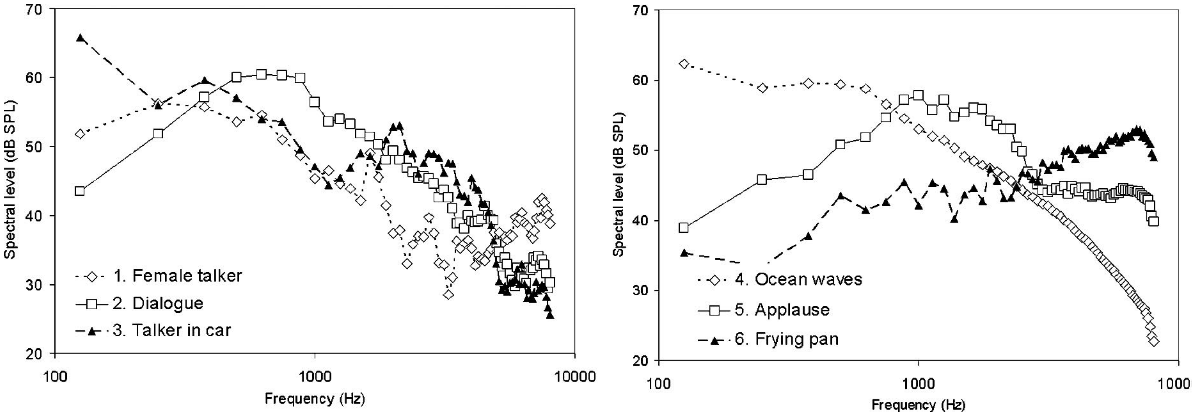

Fig. 2. Spectral levels of the test signals. In the left panel, the average speech signals are presented: speech in quiet, dialogue between male and female talkers, and speech in car noise. The right panel represents the average spectra of the non-speech test signals: ocean waves, applause, and frying.

The Starting Gain-Frequency Response

For each subject, a linear gain-frequency correction was applied to the stimuli using the NAL-RP prescriptive formula (Byrne & Dillon, 1986). Given the fact that modern hearing aids usually apply some form of automatic gain control, gain differences from NAL-RP were applied based on the overall level of the test stimuli and an overall compression ratio of 2:1. A baseline presentation level of 65 dB SPL was chosen, with gain reduced for videos with a default presentation level above 65 dB SPL, and increased for videos with a default presentation level below 65 dB SPL.

To prevent systematic bias in the fine-tuning results caused by the baseline response, a technique of roving that changed the overall gain and the overall slope of the frequency-gain curve was applied. Each subject was tested with two initial gain-frequency responses: one with a slope shallower than the prescribed NAL-RP response, and another with a slope steeper than NAL-RP. For roving of the initial overall slope, the NAL-RP prescribed slopes were corrected at random with gain per octave (dB/oct) values of 2.4, 3.6, or 4.8 dB/oct in both directions. Because of equipment limitations, the slopes ranged from 1.2 to 3.6 dB/oct in the steeper direction for subjects, with steeply sloping hearing loss, who were prescribed with more gain in the higher frequencies than other subjects. In addition, a roving level for the overall gain was applied that attenuated the signal by 0, 2, 4, or 6 dB relative to the NAL-RP prescription.

All gain corrections (subject-dependent NAL-RP prescription, video-dependent overall level compression, baseline-dependent roving of the amplification slope, and overall gain roving relative to NAL-RP) were implemented in the digital equalizer not under the subjects' control, and thus were applied to the test signals before they were presented in the free field.

Procedure

The four control configurations were tested independently during four test sessions, each of which lasted 30 to 60 min. With a few exceptions, the test sessions were conducted on separate days.

In session 1, the subject's uncomfortable loudness level (UCL) was measured using the seven-point categorical scale introduced by Cox et al. (1997). The arguing and frying stimuli that contained the highest intensity levels in the low and high frequencies, respectively, were presented three times in an ascending manner. The lower of the two highest levels across the two stimuli that the subject rated "loud, but OK" at the third repetition determined the maximum output level of all signals to be presented in further test sessions with that particular subject. Eighteen subjects required the gain decreased for one or more videos. Of these subjects, four required gain decreased by 2 dB for only the two loudest videos. The maximum gain reduction required for the loudest video was 7 dB. These gain reductions affected only the initial presentation level and did not prevent the subjects from choosing higher levels during the fine-tuning of the responses.

Following the UCL measurement and in each subsequent session, the subject was introduced to the control configuration to be tested and was trained on the effects of the control keys using speech and nonspeech videos that were not included in further testing. The training lasted as long as the subject needed to become acquainted with the control keys of the particular control configuration under test. The control configuration was then used to fine-tune the amplification settings for the six videos starting from shallower or steeper than NAL-RP frequency curves, both in test and retest (6 videos × 2 slopes × 2 test runs = 24 settings). The baseline characteristics used as the starting points for the tests had identical gain roving in both test and retest conditions. Test and retest never followed each other directly. Each test session ended with a questionnaire designed to elicit subjective reactions to the controller's key configuration and ease of use (Appendix B). Both the order of test conditions within a test session and the order of controller configurations across test sessions were balanced according to Latin squares.

At the end of the final test session, subjects were shown diagrams of the four controller configurations. A questionnaire that was designed to elicit information about subjects' configuration preferences and interest in using such a controller with their own hearing aids was then administered (Appendix C).

RESULTS

Given that the adjustment range under the subjects' control was limited to ±16 dB around the baseline response, the subjects may have encountered ceiling or floor effects in their attempts to fine-tune the response. Such effects may have contaminated the end results in that subjects who reached the ceiling or floor may not have reached their true preferred frequency response. All individually selected settings were therefore checked, and if the maximum or minimum gain was reached for any frequency band in more than 15% of the cases, the subjects were excluded from further analysis. This proved to be the case for two subjects with flat hearing loss who were both new to amplification (341 and 342 reached maximum values in 45% and

24% of the settings, respectively). Therefore, further analysis will be based on 22 subjects. Not surprisingly, the limit of the equalizers was reached most often when using controller C, the only controller that lacked a direct method of controlling the overall volume.

The main results will be presented in five sections. As a whole, they form the basis for the Discussion section, in which the three main experimental questions will be answered.

Efficiency of Fine-Tuning for Different Controllers

The efficiency of the fine-tuning process was investigated by measuring both the average time and the number of key presses required to reach the preferred response shape with each controller. These data are available for each of the 2112 trials performed (22 subjects × 4 controllers × 6 videos × 2 baseline characteristics × test/retest). It was evident that some subjects spent more time than did others exploring the individual control buttons. Among the subjects, the average response time ranged from 16.2 to 130.9 sec, whereas the average number of key presses ranged from 3.3 to 47.9. For both parameters, the observations were transformed using the natural logarithmic function to meet the assumption of normal distribution. As the baseline response was accepted without further adjustment in 84 trials, which resulted in zero key presses, a value of one was added to this parameter before the natural logarithmic transformation so that the key presses ranged from 1 to 123, instead of from 0 to 122. As expected, response time and number of key presses were moderately but highly significantly correlated with each other ($r = 0.59, p < 0.001$). That is, the greater the number of key presses, the longer it took to reach the preferred response. Consequently, only the number of key presses has been used in further repeated measures analyses. This variable was selected because the number of key presses better demonstrates the extent of fine-tuning effort required to obtain the preferred response, whereas the time parameter also includes the time each subject spent listening passively to a particular setting between key presses.

A repeated measures analysis of variance (ANOVA) of key presses was conducted using video, controller, baseline characteristic, and repetition as repeated measures and hearing loss as the between-group variable. Of the main factors, video ($F_{5,95} = 3.7, p = 0.004$) and repetition ($F_{1,19} = 11.5, p = 0.003$) were highly significant ($p < 0.01$). There was also a highly significant interaction between video and baseline characteristic ($F_{5,95} = 6.8, p = 0.00002$). On average, subjects required fewer key presses (and hence less time) for the second repetition (13.7 key presses/42.4 sec) than for the first (15.6 key presses/45.5 sec). When starting from a baseline response that was flatter than NAL-RP, the subjects required, on average, significantly fewer key presses to reach their preferred response when listening to applause (11.5 key presses) than when listening to waves or to any of the three videos that contained a speech component (14.4 to 16.9 key presses). However, more key presses were required, on average, to reach the preferred response for frying when starting from the baseline response that was steeper than NAL-RP (16.8 key presses) than when listening to female reading, waves, or applause (13.7 to 14.5 key presses). Although not significant ($F_{2,19} = 1.5, p = 0.25$), there was a trend for subjects with a flat hearing loss to need fewer key presses (7.4 key presses, on average) than did subjects with gently or steeply sloping hearing loss (15.7 and 17.3 key presses, on average, respectively) to reach their preferred responses.

Similar repeated measures analyses using age group and experience with amplification as between-group variables revealed no significant effect of age ($F_{1,20} = 0.03, p = 0.85$; 13.7 and 15.8 key presses, on average, by younger and older subjects, respectively), or experience ($F_{1,20} = 0.7, p = 0.41$; 10.4 and 15.6 key presses, on average, by subjects who were inexperienced and experienced with amplification, respectively).

Reproducibility of the Preferred Responses

In further analyses, the focus will be on the gain corrections (re: NAL-RP) for 400, 1250, and 4000 Hz because these were the independent gain parameters that determined the preferred gain-frequency response. The gains at other frequencies were derived, either by interpolation or extrapolation, from these gain values. If we compare the result from a test and retest measurement for a given condition, a test/retest standard deviation (SD) can be calculated from the gain corrections selected at 400, 1250, and 4000 Hz.

The design of the study was such that 1056 test conditions (22 subjects × 4 controllers × 6 videos × 2 baseline characteristics) were measured twice. The test/retest SD across all subjects and conditions ranged from 0 to 13.9 dB, with a mean value of 2.8 dB. As the SDs did not follow a normal distribution, four separate Friedman ANOVAs were conducted to investigate the effect of subject, video, controller, and baseline response. Of these factors, subject, controller, and baseline response showed a highly significant effect ($p < 0.000001$, $p < 0.000001$, and

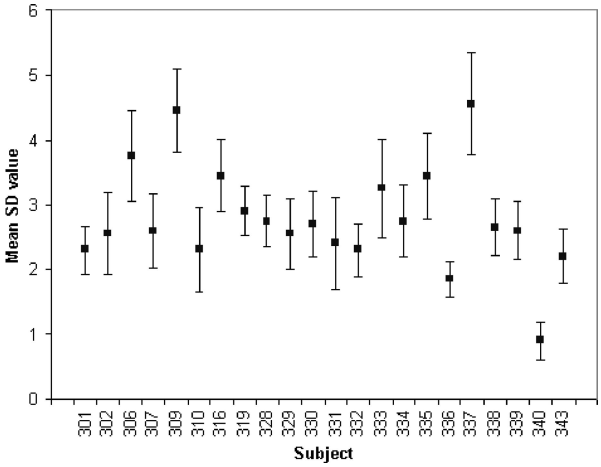

Fig. 3. The mean test/retest standard deviation (SD) produced by each subject across video, controller, and baseline response. The spreads show the 95% confidence intervals.

$p = 0.0001$, respectively). Figure 3 illustrates that while the majority of subjects produced a test/retest SD between 2 and 3 dB, two subjects (309 and 337) were relatively inconsistent in choosing their preferred responses (SD = 4.4 and 4.5 dB, respectively) and one subject (340) was very consistent (SD = 0.9 dB). On average, the consistency with which the preferred responses were selected decreased as the hearing loss configuration became steeper (mean SD = 2.2, 2.8, and 3.0 dB for flat, gently sloping, and steeply sloping loss, respectively). According to a Kruskal-Wallis ANOVA by Ranks test, the difference in consistency between hearing loss groups was significant ($p < 0.00001$). There was, however, no significant effect of age ($p = 0.30$; mean SD = 2.8 for both younger and older subjects) or experience with amplification ($p = .11$; mean SD = 2.7 and 2.9 for experienced and inexperienced hearing aid users, respectively).

With respect to the controllers, the test/retest SD of controller B was much higher (3.6 dB) than those of controllers A, C, and D (2.5, 2.7, and 2.3 dB, respectively). Controller B was the only controller who had two tone balance controls that each changed the shape of the gain-frequency response across all frequencies, which may have made it more difficult for the subjects to reach their preferred response in a consistent manner. Although the difference in test/retest SDs across baseline responses was relatively small, it was significantly larger for the baseline response that was steeper than NAL-RP (3.0 dB) than for the baseline response that was flatter than NAL-RP (2.6 dB).

Preferred Gain Characteristics

To determine the effects of fine-tuning on the preferred gain characteristics, the data were analyzed in terms of the preferred overall gain relative to NAL-RP (ΔG-av) and the slope of the preferred gain characteristic relative to the slope prescribed by NAL-RP (ΔG-sl). The overall gain was calculated as the average gain across 400, 1250, and 4000 Hz (inclusive of the volume control), and the slope was calculated in decibel per octave from 400 to 4000 Hz. Across all subjects, videos, controllers, and baseline responses, the preferred overall gain was 6.3 dB lower and the preferred slope was 1.43 dB/oct flatter than the NAL-RP prescription. This effect resulted more from a decrease of gain in the high frequencies than an increase of gain in the low frequencies.

Two separate repeated measures ANOVAs were conducted using the ΔG-av and ΔG-sl values as observations. In both analyses, video, controller, baseline response, and repetition were used as repeated measures and hearing loss was used as the between-group factor.

The only factor that had a highly significant effect on overall gain was video ($F_{5,95} = 12.7$, $p < 0.0000001$). Post hoc Bonferroni testing revealed that the ΔG-av selected for speech in car (−3.8 dB) was significantly higher than the ΔG-av selected for the other videos (−5.9 to −8.3 dB). In addition, a significantly lower ΔG-av was selected for frying than for female reading and applause (Fig. 4A).

Video also interacted significantly with the baseline response ($F_{5,95} = 20.2$, $p < 0.0000001$), with controller and hearing loss ($F_{30,285} = 2.1, p = 0.001$), and with controller, repetition, and hearing loss ($F_{30,285} = 2.0$, $p = 0.002$). For two videos (female reading and waves), the subjects, on average, selected significantly higher overall gain when starting from a baseline response that was steeper than NAL-RP. For the frying video, subjects selected significantly higher overall gain when starting from a baseline response that was flatter than NAL-RP. With respect to video, there were no systematic differences in the selected overall gain across controller, repetition, and hearing loss. The preferences of subjects with a gently sloping loss tended to be somewhat less consistent across these factors than those of the subjects with a flat or steeply sloping loss.

The preferred slope was significantly highly affected by video ($F_{5,95} = 5.2, p = 0.0003$) and baseline response ($F_{1,19} = 260.6$, $p < 0.0000001$). No other main factors or interactions reached significance at the 0.01 significance level. Figure 4B shows that, on average, significantly flatter ΔG-sl values were selected for the arguing and frying videos (−1.8 dB/oct and −2.0 dB/oct, respectively) than for the speech in car and waves videos (−1.0 dB/oct for both stimuli). When subjects started from a baseline curve that was flatter than NAL-RP, the average preferred

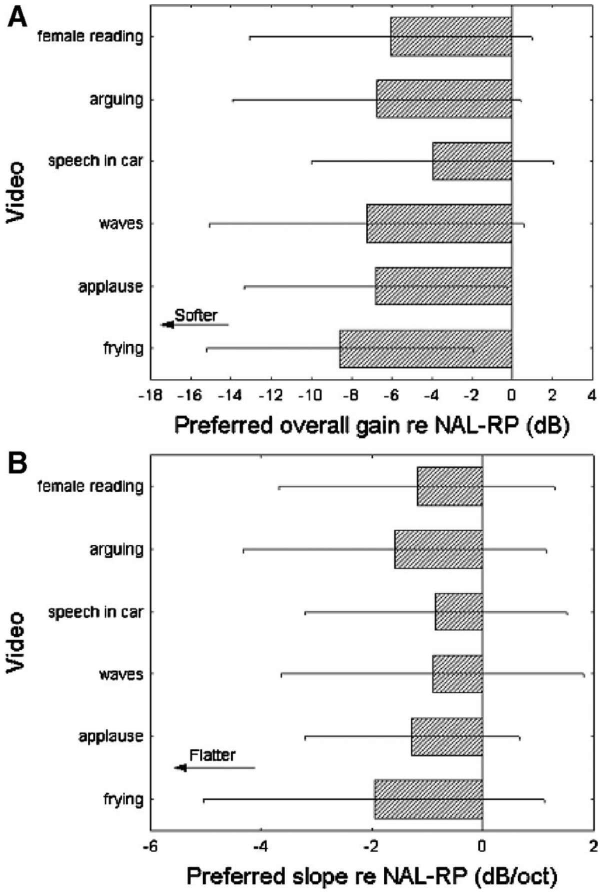

Fig. 4. The average preferred gain (A) and slope (B) relative to the NAL-RP prescription for each video. The spreads show the 95% confidence intervals.

ΔG-sl was −3.91 dB/oct, whereas the average preferred slope was +1.05 dB/oct (i.e., steeper than NAL-RP) when subjects started from a baseline curve that was steeper than NAL-RP.

The Effect of Baseline Characteristics

Given the rather large effect of the baseline response on the selected slope variation, this parameter warranted further investigation. For this purpose, the effect of gain and slope roving on the preferred overall gain and slope deviations was considered. As the roving categories were not balanced across video, controller, baseline, and hearing loss, the effect of roving was tested using a multivariate test of significance and type III errors.

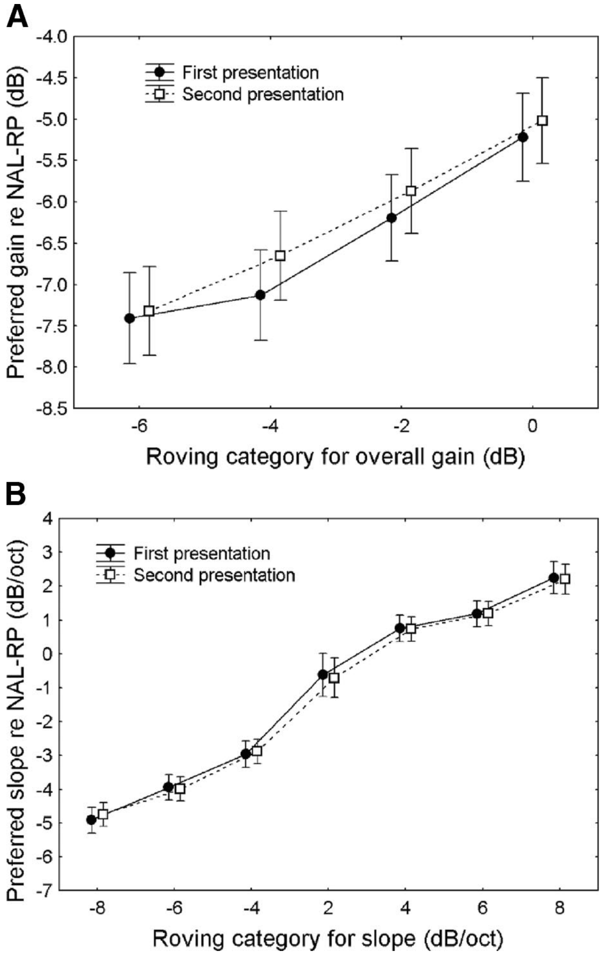

Fig. 5. The average preferred overall gain (A) and slope (B) relative to the NAL-RP prescription as a function of the roving category applied. The solid circles and the open squares show the results obtained during the first and second presentation, respectively. The spreads show the 95% confidence intervals.

When using the overall gain deviations as observations, gain roving (0, −2, −4, and −6 dB) as the between-group parameter, and repetition as the repeated measure, a significant effect of roving was found ($F_{6,2102} = 7.7, p < 0.0000001$; Fig. 5A). According to a post hoc Bonferroni test, the selected overall gain differed significantly between the 0 dB roving category and the −4 and −6 dB roving categories, as well as between the −2 and −6 dB roving categories. This was true for the results obtained at both the first and second repetitions. The lower the gain from which the subjects started, the lower the final overall gain they selected. It should also be noted that there was a small tendency for the subjects to choose slightly higher gain levels during retest (Fig. 5A).

When the selected slope deviations were used as observations, the effect of slope roving (4.8, 3.6, 2.4, 1.2, −2.4, −3.6, and −4.8 dB/oct) was again significant ($F_{12,2096} = 96, p < 0.0000001$; Fig. 5B). A post hoc Bonferroni test revealed that for both repetitions, the selected slope was significantly different between all roving categories, except between the 2.4 and 3.6 dB/oct categories. That is, a steeper response was selected when starting from a steeper baseline response, whereas starting from a flatter baseline response resulted in a flatter preferred response.

Subjective Preferences for Different Controllers

The main results of the questionnaires that were administered at the end of each test session are presented in Figs. 6A–E. Figure 6A shows that the subjects generally had no difficulty in using the different control configurations. Overall, the scores were most favorable for controllers A and D, whereas eight subjects found only controllers B and C “somewhat easy” or “difficult” to use. Figure 6B shows that the change in gain resulting from each key press was often judged to be “somewhat subtle,” especially for controller C, a configuration in which each key press resulted in a gain change in only one frequency band. For two-button controller A, with which subjects could change volume and slope, half the subjects found the step size “just right.”

Figures 6C–E illustrate the subjects’ perception of their ability to improve different aspects of the sound, such as the sound quality, speech clarity (if applicable), and comfort. Note that this is not a direct comparison between the performance of the fine-tuned setting relative to the baseline setting, but rather a subjective impression about the effectiveness of the fine-tuning process. The majority of subjects found that it was possible to improve all three qualities “all” or “most” of the time. One subject indicated that speech clarity could “never” be improved by controller C. With respect to quality improvement, there was a trend for controller C, which had no volume control function, to be rated least favorable.

In the questionnaire conducted at the end of the study, the majority (86%) of subjects showed a preference for controller A (50%) or D (36%), which is in agreement with the findings in Figure 6. That is, there was an overwhelming preference for the controller that changed the volume and the slope, as

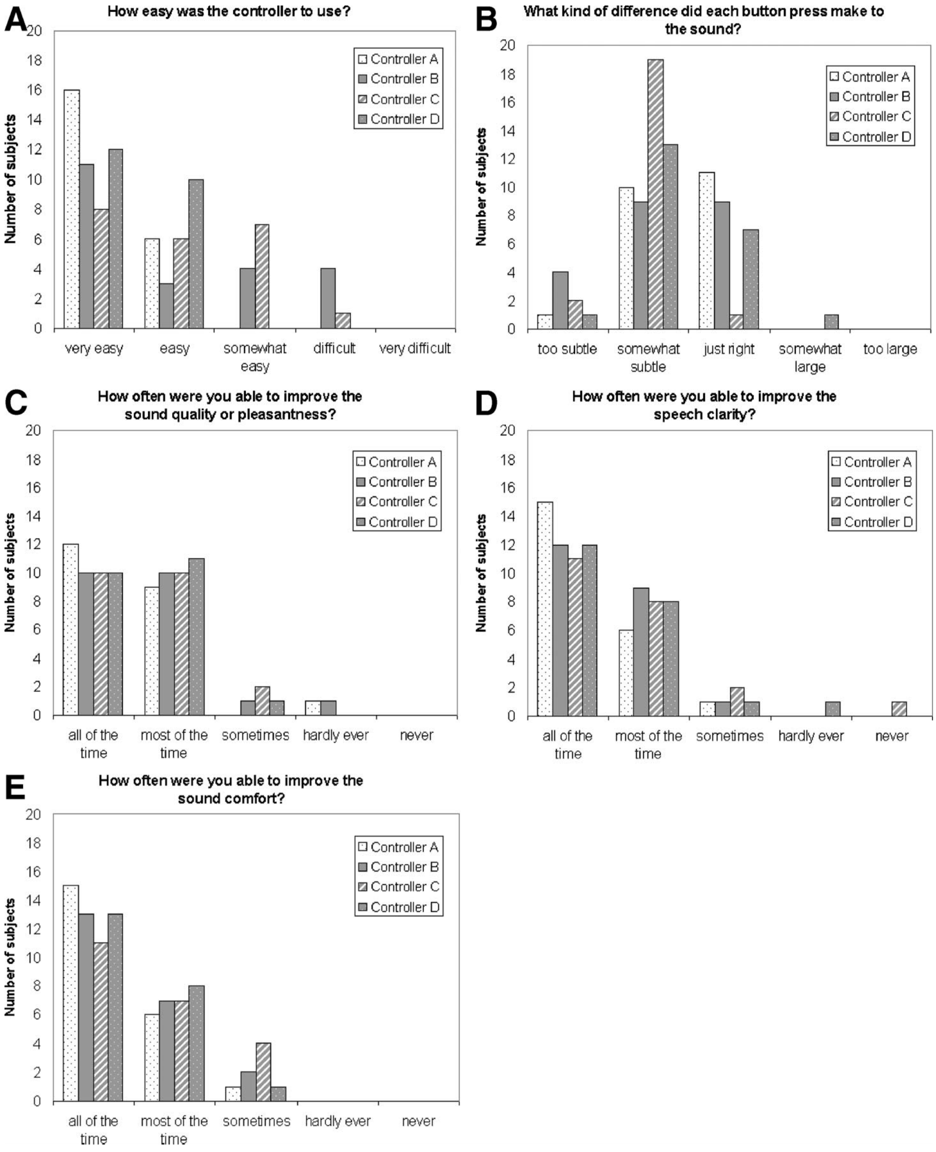

Fig. 6. Subjective judgments of the control configurations: A (volume/slope), B (volume/contrast/slope), C (bass/mid/treble), and D (volume/bass/treble).

well as for the three-button controller that changed the volume, bass, and treble. Controller B was preferred by 5% of subjects, whereas 9% preferred controller C. Another interesting result from the final questionnaire was that 17 of 22 subjects indicated it was "very likely" that they would use the preferred controller with their own hearing aids, if it were available.

DISCUSSION

Reproducibility

The results of this study show that the subjects were able to provide systematic and reproducible gain-frequency response preferences in different acoustical conditions as simulated in a laboratory environment. Across controllers, subjects, videos, and baseline configurations, the average test/retest SD was 2.8 dB. In comparison, mean test/retest SDs that range from about 1 to 6 dB have been observed for hearing-impaired listeners performing such behavioral tests as aided thresholds (Humes & Kirn, 1990), speech reception thresholds in quiet and in noise (Hagerman & Kinnefors, 1995), and loudness perception (Cox, et al., 1997). The larger SD values were observed in tests using greater step sizes between stimulus levels. However, there was a significant variation in the test/retest SD produced by the 22 subjects in this study (Fig. 3), indicating that some subjects were more consistent in selecting their preferred responses than others. Whether inconsistent data are a sign that a large range of response shapes are equally acceptable or that the subjects lacked the ability to perform the task in a reliable way is currently unknown and should be further explored.

A statistical analysis revealed that controller B produced a significantly greater test/retest SD (3.6 dB) than did the other controllers. As a few subjects commented that they perceived very little change to the sound when using the tone control 2 button, or asked during the test sessions whether the button had any effect on the sound, it is likely that this particular function contributed to the larger discrepancies between the final responses of the test and retest trials for this control configuration. Similar comments were made about the key on controller C that altered gain in the mid frequencies, which produced the second largest SD value (2.7 dB). The statistical analysis also showed that a significantly larger SD was produced when starting from the baseline response that was steeper than NAL-RP (3.0 dB) than from the baseline response that was flatter than NAL-RP (2.6 dB). As the subjects generally preferred a response flatter than NAL-RP (Fig. 4B), the greater reproducibility for the flatter starting slope presumably arose from the lower number of key presses needed to reach an acceptable response. Overall, these observations suggest that any control function should produce clearly perceived changes to the sound (albeit small enough to select subtle changes), and that the starting point should not be hugely inappropriate for the particular hearing loss, otherwise less consistent adjustments may be made.

It has been found that when asking hearing-impaired listeners to assign a loudness category to different input levels, repeated presentations are needed to reach a stable outcome (e.g., Beattie, et al., 1997; Keidser, et al., 1999; Robinson & Gatehouse, 1996). Although there was no significant effect of repetition on the preferred response slope ($p = 0.83$), repetition had a statistically significant ($p = 0.02$), though clinically insignificant, effect on the preferred overall gain, as the subjects, on average, selected slightly higher levels on retest (−6.2 dB versus −6.5 dB re: NAL-RP). The change of 0.3 dB seems sufficiently small that there would be no need for a repetition to overcome a systematic change of preference. It may, however, be desirable to repeat the fine-tuning process to average out random changes.

Efficiency

The majority of subjects judged the controllers to be easy to use, despite the fact that the subjects were not instructed in detail about the function of each control key. Apart from a brief label above each key, subjects discovered the effects of the different keys themselves during a short training period and reported at the end of a test session that they generally had been able to improve important characteristics of the sound, such as sound quality, speech clarity, and sound comfort (Fig. 6). Nevertheless, subjective preferences revealed clear differences between the controllers. The final questionnaire showed that 11 subjects (50%) preferred controller A (using two keys to alter volume and slope), whereas eight subjects (36%) preferred controller D (using three keys to alter volume, bass, and treble). These subjective results correspond well with such objective results as the significantly poorer reproducibility measured for controller B, as reported earlier. In terms of the number of key presses, and hence the response time needed to reach the preferred response, the effect of the controllers was moderately significant ($p = 0.02$), showing that the subjects reached their preferred response sooner with controllers A and D (13 key presses per repetition, on average), than with controllers B and C (17 key presses per repetition, on average).

One aspect that may have biased the apparent preference for controllers A and D is the size of the perceptual change corresponding to each key press. Figure 6B shows that the implemented step size was judged to be "somewhat subtle," especially for controller C (using three keys for bass, mid, and treble). A typical pattern followed by the subjects in selecting their preferred response was first to use the volume control to find their preferred overall loudness and then to use the other available keys to shape the response. Controller C was the only one that did not have a volume control button, which means that the subjects would need to adjust all three keys in the same direction (which could not be done simultaneously) to achieve an equivalent change in loudness. This may be the reason why controller C was judged more subtle and less efficient than were the other controllers. In addition, as mentioned earlier, both controllers B and C had a key function (tone control 2 and mid frequency, respectively) that for some subjects contributed little to the overall percept of the sound.

Preferred Responses

Across videos, there was a significant difference in both the preferred overall gain and slope relative to NAL-RP. One reason for this may be that NAL-RP is based on a speech stimulus at a conversational level, whereas most of the stimuli used were nonspeech and speech with raised vocal effort. The subjects tended to select lower gain levels for louder sounds and for the noise-only situations (Fig. 4A), which is in agreement with findings of Smeds et al. (2006). With respect to the preferred slope, there was a tendency for the subjects to select relatively flatter responses for stimuli with a weighting of energy across the mid and high frequencies than for low-frequency weighted stimuli (Fig. 4B). This is in agreement with findings of Keidser et al. (2005). Although there were no significant effects of controller on the selected slope ($p = 0.67$), there was a moderately significant effect of controller on the selected overall gain ($p = 0.01$). On average, the subjects chose significantly higher overall gain levels with controller C, the controller that lacked a volume control (−5.4 dB), than with controllers A and B, both of which had a direct method of controlling the volume (−6.8 dB). Although significant, the difference in overall gain between these controllers is smaller than the difference in overall gain selected across videos and the average test/retest standard deviation value. The findings therefore suggest that the level of precision obtained with each controller was high enough to allow significantly different settings in the gain-frequency responses for the different acoustical situations included in this study, and that these differences were found for each of the controllers to be about the same. Note that because the stimuli used in this experiment were presented in the free field to unaided subjects, the results are not influenced by such acoustical uncertainties as earmold and tubing properties.

Baseline Bias

A somewhat unexpected finding in this study was the highly significant effect of the baseline configuration that has not been observed when using more structured approaches, such as the simplex method, in finding the optimal hearing aid setting for individuals (Kuk & Lau, 1995). The results presented in Figure 5A show that the subjectively preferred final setting for the overall gain relative to the individual NAL-RP prescription was consistently dependent on the gain roving relative to the baseline characteristic at the start of the fine-tuning session. The slope of the curve in Figure 5A suggests that a gain deviation at the beginning of the session was reduced to about half its size during the fine-tuning process. A similar roving effect is seen in Figure 5B for the preferred gain slope. The slope deviation was also reduced to about half its size during fine-tuning. These curves were averaged across subjects, videos, and controllers. A closer inspection of the data revealed that the influence of the degree of roving on the baseline configuration was consistently present in the final settings chosen by each subject. These results may point to a certain degree of conservatism in the fine-tuning process. That is, the subjects may only have adjusted the controller enough to make a small change to the response in the preferred direction, fearing that if they experimented too much with the keys and reached a less optimum response, they would be unable to return to an acceptable setting. Alternatively, there may be a range of equally good response shapes, and the subjects ceased making adjustments once they entered the nearest part of that range. Adjustment procedures that require subjects to fully traverse the acceptable range would be less affected by the starting value. Note that the effect of the baseline response did not prevent the subjects from choosing different responses for different acoustical environments (Fig. 4).

The large impact of the baseline condition on the preferred frequency response seen in this study was measured within one test session using one repetition. The finding implies that a biased correction of amplification is reached within one session, which means that the "real" optimum may not be reached in a single fine-tuning session. What is currently unknown is to what extent the fine-tuned setting may converge toward an optimum setting when the baseline characteristic is adaptively changed over a period of time based on past adjustments, and what happens when the adjustments are done in real life. An adaptive adjustment procedure is more likely to form part of a daily "training program," so from this perspective, the concept of a trainable hearing aid seems worth pursuing.

Effects of Hearing Loss, Age, and Hearing Aid Experience

The data suggest that subjects with flat loss were relatively faster and more consistent in reaching their preferred responses than were subjects with sloping loss, especially subjects whose loss slopes were steep. The former observation may mean that this subject group reached their preferred responses sooner, or that they did not explore the keys as much as subjects with other hearing loss configurations. Both explanations would lead to more consistent selections. The observations somewhat disagree with findings of Keidser et al. (2005), who found that subjects with flat loss were less reliable in choosing their preferred response slope than were subjects with steeply sloping loss. The observation in the present study, however, should be regarded with some caution, because only five subjects with a flat hearing loss were included in the final analyses. Overall, there is nothing in the data suggesting that self-adjustments are more beneficial to or better managed by individuals with some hearing loss configurations than others, nor did age or hearing aid experience discriminate in any significant way between the subjects' performance.

Procedural Issues

Because of ceiling effects, we excluded two subjects from this study. Other subjects also reached the limit of the equalizers; that is, changed gain by 16 dB in either direction at at least one discrete frequency, but much less often than did the two subjects excluded from the analyses. These results were accepted to retain data at the most extreme levels that could be selected in this study, and thus prevent the introduction of bias toward a narrower range of preferred responses. As indicated in the results section, the ±16 dB limit of the subject-controlled equalizer was most often reached for controller C, presumably because of the lack of a volume control. The controller with which the limit was reached the least was controller A, the most preferred.

For about two-thirds of the subjects, the UCL measurements resulted in further reductions to the starting level of the overall gain. A conservative approach was adopted to determine the starting level for each video to avoid any loudness discomfort during the testing, but the starting level did not prevent the subjects from choosing higher gain levels during the fine-tuning process. However, the choice to reduce the starting level of the signals based on the UCL outcome may have influenced the number of key presses (and thus the response time) needed to reach the preferred responses. As the preferred overall gain levels were usually lower than the gain at baseline, this may have produced a slight underestimation of the fine-tuning effort required. No other adverse effects from the decreased starting levels are likely.

In this study, the subjects were presented with various listening environments on video using a stereo sound path and were asked to make linear adjustments from two different baseline responses twice, using different controllers that each manipulated a different set of gain parameters. This method, conducted in the laboratory, is far from the potential real-world application of a trainable state-of-the-art hearing aid. However, the current data show that in a laboratory setting, hearing-impaired individuals are able to use simple controls to manipulate the hearing aid response in a reliable and efficient way when listening to recordings of real-life environments. We note that if the linear gain adjustments made for different listening environments that vary in intensity level are related to the overall level of the environment, this is equivalent to training the compression ratio. By relating even overall gain adjustments to the input levels in each frequency band, the compression ratio can be trained independently in different frequency bands. Alternatively, the actual compression ratio could be one of the control functions. Future work should focus on resolving the baseline response bias, and testing the validity of self-adjustments in real life. Further, the reliability and efficiency need to be tested for such other signal processing parameters as compression, noise reduction, and microphone directionality.

CONCLUSIONS

This study investigated the reliability and efficiency of the use of four different control configurations to adjust the gain-frequency response to a preferred setting, a small step toward the concept of a trainable hearing aid. The evaluation was conducted in the laboratory using a limited number of acoustic environments and a limited number of hearing aid parameters, and showed encouraging results. The subjects produced reproducible results with clear and consistent differences between specific acoustic environments using various simple control keys, such as volume and tone balance. The selected responses did not strongly depend on the combination of control keys; however, the gain response at baseline turned out to strongly influence the preferred response. This would be a problem in a single fine-tuning session, but may be solved in the concept of the trainable hearing aid where the baseline response changes adaptively throughout the training period. Although all the key function combinations tested in this study were thought to be easy to use, the subjects showed a clear preference for two controllers (one with volume and slope keys, and one with volume, bass, and treble keys) that also seemed to produce more reliable results and be the most efficient in terms of number of key presses, and hence the response time, needed to reach the preferred response. The least accepted controllers each had a key function that was thought to have little effect on the percept of sound. These preferences suggest that one of the key functions must be a volume control, and that each function must provide sufficiently audible changes to the sound. Because the subjects were able to handle different control configurations without knowing exactly what the effects were, the control parameters may be extended toward more complex hearing aid functions.

ACKNOWLEDGMENTS

The authors thank Scott Brewer for his programming activities and technical support. We owe many thanks to the test subjects for their patience and enthusiastic participation. A patent on trainability in hearing aids as discussed in this article has been submitted under the Co-operative Research Centre for Cochlear Implant and Hearing Aid Innovation (Australia), which is likely to be licensed to a major hearing aid manufacturer. However, none of the authors has a personal financial interest in or association with any commercial product that may incorporate trainability as described in the patent.

Portions of this work were presented at the International Hearing Aid Research Conference, Lake Tahoe (August 2004) and in "The 'trainable' hearing aid: concepts, design, and candidacy," session at the 17th Annual Convention and Exposition of the American Academy of Audiology, Washington, DC (March 2005).

Address for correspondence: Wouter Albert Dreschler, PhD, Academic Medical Center, Meibergdreef 9, D2-240, 1105 AZ Amsterdam, The Netherlands. E-mail: w.a.dreschler@amc.nl.

Received December 7, 2006; accepted August 23, 2007.

APPENDIX A

Instructions

You will see and listen to a number of video recordings of common listening situations. Your task is to use the buttons on the controller in front of you to adjust the sound until your preferred setting is reached.

Each videotape will run continuously until you have made your choice. Spend some time experimenting with the controller and listen to the changes it can make to the sound. There are no "right" or "wrong" settings, and you can take as long as you need to make the adjustments.

APPENDIX B

Please read the following statements about the use of the controller carefully and choose the answer that best matches your opinion.

1. The controller I used to adjust the amplification and sound quality was very easy to use/easy to use/somewhat easy to use/difficult to use/very difficult to use
2. When using the buttons on the controller, I found that the changes to the sound for each button press were too subtle/somewhat subtle/just right/somewhat large/too large
3. If you could choose your own label for button 1, what would it be? Or, how would you describe the effect it had on the sound?
4. If you could choose your own label for button 2, what would it be? Or, how would you describe the effect it had on the sound?
5. If you could choose your own label for button 3, what would it be? Or, how would you describe the effect it had on the sound?
6. If I were to rank the buttons from most to least important, I would place them in the following order from most to least important:
7. Overall, I felt that I was able to improve the sound quality or pleasantness all of the time/most of the time/sometimes/hardly ever/never
8. Overall, I felt that I was able to improve the speech clarity all of the time/most of the time/sometimes/hardly ever/never
9. Overall, I felt that I was able to improve the sound comfort all of the time/most of the time/sometimes/hardly ever/never

Comments

APPENDIX C

The diagrams in front of you show the different styles of sound controllers that you used in the last four sessions to adjust a variety of listening situations to your preferred settings. Please feel free to refer to these diagrams when answering the following questions.

1. Which of the four styles of controllers did you prefer? Controller A/B/C/D

2. If the controller you chose in question 1 were available for use with your own hearing aids in real-life listening situations, how likely would you be to use it? Very likely/likely/somewhat likely/not very likely/unlikely
3. If you were able to design your own controller with any number or combination of buttons, what would it look like? What labels would it have?

REFERENCES

Beattie, R. C., Huynh, R. C., Ngo, V. N., et al. (1997). IHAFF loudness contour test: reliability and effects of approach mode in normal-hearing subjects. *J Am Acad Audiol, 8*, 243–256.

Büchler, M., Allegro, S., Launer, S., et al. (2005). Sound classification in hearing aids inspired by auditory scene analysis. *EURASIP J Appl Signal Processing, 18*, 2991–3002.

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*, 257–265.

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

Cornelisse, L., Seewald, R., & Jamieson, D. (1994). Fitting wide-dynamic range compression hearing aids: the DSL I/O approach. *Hear J, 47*, 23–29.

Cox, R. M., Alexander, G. C., Taylor, I. M., et al. (1997). The contour test of loudness perception. *Ear Hear, 18*, 388–400.

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

Fabry, D. A. (1996). Clinical applications of multimemory hearing aids. *Hear J, 49, 10*, 53–56.

Flynn, M. C. (2005). Data logging: a new paradigm in the hearing instrument fitting process. *Hear Rev, 12*, 52–57.

Franck, B. A. M. (2004). Hearing-aid fitting in interaction—on optimal combinations of multiple signal-processing strategies. PhD thesis, University of Amsterdam.

Franck, B. A. M., Dreschler, W. A., & Lyzenga, J. (2004). Methodological aspects of an adaptive multidirectional pattern search to optimize speech perception using three hearing-aid algorithms. *J Acoust Soc Am, 116*, 3630–3628.

Goldstein, D. P., Shields, A. R., & Sandlin, R. E. (1991). A multiple memory, digitally-controlled hearing instrument. *Hear Instrum, 42*, 18–21.

Hagerman, B., & Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scand Audiol, 24*, 71–77.

Howes, C., & Olsen, L. (2006). The role of virtual reality in hearing instrument fittings. *Hear Rev, 13*, 60–63.

Humes, L. E., & Kirn, E. U. (1990). The reliability of functional gain. *J Speech Hear Disord, 55*, 193–197.

Jenstad, L. M., Van Tasell, D. J., & Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *J Am Acad Audiol, 14*, 347–360.

Keidser, G. (1995). The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids. *Ear Hear, 16*, 575–586.

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.

Keidser, G., Seymour, J., Dillon, H., et al. (1999). An efficient, adaptive method of measuring loudness growth functions. *Scand Audiol, 28*, 3–14.

Kiessling, J. (2001). Hearing aid fitting procedures—state-of-the-art and current issues. *Scand Audiol Suppl, 52*, 57–59.

Kuk, F. K. (1999). How flow charts can help you troubleshoot hearing aid problems. *Hear J, 52*, 46–52.

Kuk, F. K., & Lau, C. (1995). Effect of initial setting on convergence to optimal hearing aid setting using a simplex method. *Br J Audiol, 29*, 263–269.

McCandless, G., & Lyregaard, P. (1983). Prescription of gain/output (POGO) for hearing aids. *Hear Instrum, 1*, 12–16.

Moore, B. C. J., Alcantara, J. I., & Glasberg, B. R. (1998). Development and evaluation of a procedure for fitting multi-channel compression hearing aids. *Br J Audiol, 32*, 177–195.

Nelson, J. A. (2001). Fine tuning multi-channel compression hearing instruments. *Hear Rev, 8*, 30–35, 58.

Nilsson, M., Ghent, R. M., Bray, V., et al. (2005). Development of a test environment to evaluate performance of modern hearing aid features. *J Am Acad Audiol, 16*, 27–41.

Powers, T. A., & Hamacher, V. (2002). Three-microphone instrument is designed to extend benefits of directionality. *Hear J, 55*, 38–45.

Revitt, L. J., Schulein, R. B., & Julstrom, S. D. (2002). Toward accurate assessment of real-world hearing-aid benefit. *Hear Rev, 9*, 36–38.

Robinson, K., & Gatehouse, S. (1996). Test/retest reliability of loudness scaling. *Ear Hear, 17*, 120–123.

Russ, D. M., & Olsen, G. (2001). Audio verification environment: how to present and assess real world performance in the office. *Audiology* [Online].

Sandlin, R. E., & Meltsner, R. (1989). Clinical trials with a remote control, programmable hearing instrument. *Hear Instrum, 40*, 34–39.

Schum, D. J. (2005). Integrating new technology: addressing old problems. *Audiology* [Online].

Smeds, K., Keidser, G., Zakis, J., et al. (2006). Preferred overall loudness. I. Sound field presentation in the laboratory. *Int J Audiol, 45*, 2–11.

Stypulkowski, P. H., Hodgson, W. A., & Raskind, L. A. (1992). Clinical evaluation of a programmable multiple memory ITE. *Hear Instrum, 43*, 25–29.

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