Research Article

A “Goldilocks” Approach to Hearing-Aid Self-Fitting: User Interactions

Arthur Boothroyd^a and Carol Mackersie^a

Purpose: The purpose of the current study was to evaluate user reactions to custom software designed for self-adjustment of amplification.

Method: “Goldilocks” software was developed to allow user exploration and selection of preferred levels of overall output, low-frequency cut, and high-frequency boost while listening to preprocessed speech. Thirteen hearing-aid users and 13 nonusers self-adjusted before and after taking a speech perception test incorporated into the software.

Results: All 26 participants were able to complete the 2 adjustments and the intervening test in an average of 6.5 min—20 of them from on-screen instructions without experimenter help. Relative to a generic starting condition,

the average participant opted to increase overall output, reduce low-frequency cut, and increase high-frequency boost. The first and second self-selected values were highly correlated, but there was evidence of further increases of overall volume and high-frequency boost after speech perception testing with the initial adjustment. There was no evidence that prior hearing-aid experience affected the ability to understand or complete the self-fitting process.

Conclusions: This approach to hearing-aid self-fitting can be a speedy, reliable, and feasible alternative to, or supplement to, conventional fitting procedures, but many questions remain to be answered.

This special issue contains papers from the 2016 Hearing Across the Lifespan (HEAL) conference held in Cernobbio, Italy.

The availability of low-cost direct-to-consumer hearing aids and Personal Sound Amplification Products brings the issue of user self-fitting, without audiologist or dispenser involvement, to the fore (Convery, Keidser, Seeto, & McLelland, 2017; Dillon & Keidser, 2011; Keidser & Convery, 2016). Even in an audiology clinic, user self-fitting could save time, provide useful information, and/or provide the opportunity for increased attention to other aspects of aural rehabilitation. The authors are currently engaged in a study of the efficacy of, optimization of, and candidacy for hearing-aid self-fitting. A first need for this work was the development of self-fitting software that would be usable by adults with hearing impairment and without hearing-aid experience while generating useful research data. The present article describes the software and reports on the ways in which participants interacted with it during an initial study.

The self-fitting procedure incorporated into the software was modeled on the second most user-preferred configuration in the work of Dreschler, Keidser, Convery, and Dillon (2008) and provided for adjustment of three parameters: overall level, high-frequency boost, and low-frequency cut. Although a two-parameter configuration was the most preferred in the Dreschler et al. study, we decided to retain separate high- and low-frequency adjustments because of the potential value of the latter during self-adjustment in noise.

Research on self-fitting has generally followed the clinical pattern of threshold-based prescription followed by fine-tuning (Convery, Keidser, Seeto, Yeend, & Freeston, 2015; Keidser & Convery, 2016; Keidser, Dillon, Zhou, & Carter, 2011; Keissling et al., 2015; Wong, 2012). In the present study, however, we bypassed individual prescription and gave participants the opportunity to make changes from a generic starting condition. We also included a speech recognition test (potentially self-administered) in the self-fitting protocol.

The goals of this study were to determine the following: the proportion of participants able to complete the self-fitting process and to do so without researcher help; the number of adjustments made and the time taken; the self-selected values of the three parameters; the number of word recognition errors made after self-adjustment; replicability; and the effects, if any, of the speech recognition test.

^aSan Diego State University, CA

Correspondence to Arthur Boothroyd: aboothroyd@sdsu.edu

Editor-in-Chief: Sumitrajit Dhar

Received December 14, 2016

Revision received May 4, 2017

Accepted May 16, 2017

https://doi.org/10.1044/2017_AJA-16-0125

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Software

Design Criteria

The design of the self-adjustment interface followed four principles:

1. There should be no more than three adjustable parameters. The purpose was to minimize cognitive demands on the target population of elderly acquirers of hearing aids.
2. Each parameter adjustment should be presented, initially, in isolation. We felt that users should have the opportunity to learn the association between each parameter and the resulting sensation in isolation before being exposed to all three in combination.
3. Users should experience the impact of their adjustments on speech recognition—perhaps increasing the emphasis on intelligibility rather than on comfort.
4. Each step in the adjustment process should be logged and timed for research purposes.

On the assumption that an adjustable parameter produces sensations that are “too little,” “too much,” or “just right,” the resulting software was named Goldilocks. The three adjustment parameters used in this first iteration of the software were:

1. **Loudness**—overall root-mean-square (rms) coupler amplitude at 11 levels from 2 to 13 (60 to 100 dB SPL).
2. **Crispness**—high-frequency boost above 1 kHz in five steps from 0 to 4 (0 to 8 dB/octave).
3. **Fullness**—low-frequency cut below 500 Hz at five levels from 0 to 4 (0 to 20 dB).

The software was written in ToolBook Instructor (SumTotal Systems Inc., Gainesville, FL, USA) and implemented on a Microsoft Surface Pro3 computer.

Speech Materials

The Goldilocks software incorporated prerecorded speech for two purposes. The first was to provide users with amplified speech on which to base judgments. The second was to provide material for a speech recognition test to be given after the initial adjustment.

The speech material consisted of 100 four-word sentences arranged in 10 lists of 10 sentences each, developed and recorded by Charles Ruby for an earlier research project (Ruby, 2015). The first sentence list was repeated continuously during user self-adjustment. The remaining nine lists were used for testing.

Spectral Alternatives

For use in Goldilocks, the sentences were preprocessed using DADiSP array-processing software (DSP Corporation, Newton, MA, USA). The required levels of low-frequency cut and high-frequency boost were approximated in five frequency bands centered at octave intervals from 354 to

5657 Hz. The result was a matrix of 25 spectra for each of the 100 sentences. The user's selection of “Crispness” and “Fullness” determined which of the 25 versions of the current sentence set was being heard. The user's selection of “Loudness” set the gain of the computer's sound card.

Starting Level and Spectrum

For purposes of the present study, participants started with a generic combination of spectral shape and level. This was based on the National Acoustics Laboratory's second prescriptive formula for nonlinear hearing aids (NAL-NL2) using a sound-field rms input of 65 dB SPL and a mild-to-moderate hearing loss of around 30 dB in the lower frequencies, sloping to 50 dB at 4 kHz. The starting “loudness” level (#4) gave an average coupler output of 67 dB SPL; the starting fullness level (#2) gave 10 dB of attenuation below 500 Hz; the starting “crispness” level (#2) gave a 4-dB/octave boost above 1 kHz. Note that NAL-NL2 prescribes a nonlinear relationship between input level and output. In this study, however, there was no compression of the speech signal heard by the participants.

Installation

As indicated earlier, Goldilocks was installed on a Microsoft Surface Pro3 computer (Microsoft, Redmond, WA, USA). Audio output was taken from the computer's docking station and delivered monaurally using an HF5 receiver (Etymotic Research, Inc., Elk Grove Village, IL, USA). The upper panel in Figure 1 shows the three user touch-screen adjustments. The lower panel shows a single logged adjustment record. In this case, the participant explored a range of levels for each parameter before returning to the starting spectral shape but with an overall amplitude increase of one step (approximately 4 dB).

Method

Participants

Twenty-six adults with mild-to-moderately-severe hearing loss participated in the study. Ages ranged from 52 to 89 years with a mean of 73 years. Average hearing loss at 500, 1000, 2000, and 4000 Hz ranged from 45 to 80 dB with a mean of 57 dB. Scores on the Montreal Cognitive Assessment Test (administered as an exclusionary criterion) ranged from 21/30 to 30/30 with a mean of 26/30. Fifteen participants were women and 11 were men. Thirteen of the participants were experienced hearing-aid users. The other 13 had never used hearing aids. Although not planned, the mean thresholds of the two groups were not significantly different between 250 and 4000 Hz, as shown in Figure 2.

Procedure

After consent to participate and collection of demographic data, cognitive screening and pure-tone audiometry were completed. Participants then heard, but were not asked to repeat, Set 1 of the short sentences, concatenated and

Figure 1. The upper panel shows the three user controls for low-frequency cut (Fullness), overall amplitude (Loudness), and high-frequency boost (Crispness). These were initially presented one at a time. The lower panel shows the adjustment record for an individual user while listening to Set 1 of the preprocessed sentence materials.

Figure 2. Mean thresholds (± 1 SD) of 13 experienced hearing-aid users and 13 nonusers.

played repeatedly. While listening, the order of adjustments and the accompanying on-screen instructions were as follows:

1. Overall amplitude (Loudness)—“*You are going to hear a man talking. You are encouraged to explore the effect of making the sound louder and softer. Adjust the loudness to your liking. Hit OK when you are satisfied. Tap here to Start.*”
2. High-frequency boost (Crispness)—“*Now repeat the same for the crispness (or detail) of the speech.*”
3. Overall amplitude (after high-frequency adjustment)—“*Readjust if you want to, or just hit OK.*”
4. Low-frequency cut (Fullness)—No instruction.
5. Final adjustment of all three parameters—“*Make any final adjustments. Click OK when you are satisfied.*”

For Steps 1 to 4, each adjustment was shown in isolation. Only at Step 5 did the user see all three at the same time. Participants were instructed to adjust the sound to their liking. After Step 5, participants were presented with 30 novel sentences and were asked to repeat them. Performance was scored by the researcher in terms of percent words correct. No feedback on performance was provided to the participants but it was assumed they would know when they were having difficulty. After the test, self-adjustment was repeated, starting again from the generic condition, and followed by a second speech recognition test with the new settings. Note that the first speech test was part of the self-fitting protocol. The second was added for purposes of research.

Participant Responses

Completion of Self-Fitting

All participants were able to complete the self-fitting process. Twenty of the 26 (77%) followed the on-screen instructions without researcher help, but three hearing-aid users and three nonusers required a little help—either by pointing out the location of the “Start” instruction or by encouraging exploration. Three of those needing help were men and three were women. In simple t tests, there was no evidence that the six participants needing help differed significantly from the other 20 in terms of mean age or Montreal Cognitive Assessment Test score. They did, however, have significantly higher mean thresholds at 2 kHz, 65 versus 53 dB; $t(24) = 3.01$; $p = .006$, and 4 kHz, 76 versus 60 dB; $t(24) = 2.65$; $p = .014$.

Number of Changes and Adjustment Time

The average hearing-aid user took 66 s to explore nine changes from the starting condition before reaching a first setting. After taking the speech recognition test with this setting, the second self-adjustment involved 13 changes from the starting condition and took 64 s.

The average nonuser took 106 s to explore 12 changes at the first self-adjustment but only 70 s to explore 16 changes

at the second. A mixed analysis of variance was completed using hearing-aid experience as a grouping factor and adjustment session as a repeated measure. Post-hoc analysis of the interaction, using the least significant difference test, showed that the nonusers took significantly longer for the first self-adjustment ($p = .001$) but reduced the time significantly for the second ($p = .001$), at which point they were not significantly slower than the hearing-aid users. Analysis of the number of changes, however, failed to reveal significant main effects or interactions.

During the first self-adjustment, two of the hearing-aid users and four of the nonusers took advantage of the opportunity to explore changes in overall volume after adjusting high-frequency boost. Two members of each group also took advantage of the opportunity to adjust one or more of the parameters before indicating final acceptance. After taking the speech recognition test, nine of the hearing-aid users and nine of the nonusers explored changes in overall volume after adjusting high-frequency boost but only one hearing-aid user and two nonusers made further final adjustments before indicating acceptance.

Word Recognition

Using the first self-adjustment, the average hearing-aid user recognized 95.3% of the words in 30 sentences (range 81.7% to 100%). The second self-adjustment, after the speech recognition test, resulted in a score of 97.5% (range 90.8% to 100%). For the nonusers, the corresponding scores were 94.8% (range 65% to 100%) and 97.6% (range 91.7% to 100%). Arcsine transformation and analysis of variance failed to show significant effects, at the .05 level, of hearing-aid experience, replication, or their interaction. Note, however, that there was a serious ceiling effect with nine of the participants scoring 100% on one or both tests.

Parameter Selections

Figure 3 shows the distributions of individual parameter selections together with group means. The panels on the left show the first adjustment in relation to the generic starting value. The center column shows the second adjustment, also in relation to the starting value. The right panels show the second self-adjustment in relation to the first. The top, middle, and bottom rows show data for low-frequency cut, overall level, and high-frequency boost, respectively. Significant within-group changes based on simple t tests are shown by asterisks.

Both groups made significant reductions in low-frequency cut from the starting condition during the first and second self-adjustments. Both groups also showed significant increases of the overall level but only after the speech recognition test. For the hearing-aid users, but not the nonusers, there was also evidence of increased high-frequency boost after the second self-adjustment. Differences between the first and second mean self-adjustments, however, failed to reach the .05 level of confidence.

The final selected overall level was well predicted in the nonusers by the average pure-tone threshold at 1, 2, and 4 kHz, $r(13) = .77$, $p = .002$. This strong relationship was not found for the hearing-aid users, $r(13) = .47$, $p = .11$.

The data in the right-hand panels of Figure 3 provide evidence of repeatability. The first and second selections differed by no more than one step for 12 of the 13 nonusers and all three parameters. In the case of the hearing-aid users, all 13 differences were within one step for low-frequency cut and 12 met this criterion for high-frequency boost. For overall level, however, only 10 of the hearing-aid users met this criterion, whereas three users increased output by two or three steps. In a separate analysis of data for the 26 participants, significant correlations were found between the first and second adjustments for all three parameters, $r(26) = .59$, $p = .002$; $r(26) = .74$, $p < .001$; $r(26) = .67$, $p < .001$ for low-frequency cut, overall level, and HF boost, respectively. In separate mixed analyses of variance with hearing-aid experience as a grouping variable and replication as a repeated measure, the main effect of group failed to reach the .05 level of significance. These findings offer no convincing evidence to suggest that hearing-aid experience is necessary for successful completion of this approach to user self-adjustment. In post-hoc analysis, using the least significant difference test, the replication effect for the volume setting approached the .05 level of significance ($p = .075$) for the hearing-aid users but not for the nonusers ($p = .49$).

Discussion

Six of these 26 participants needed some help in following the on-screen instructions in the Goldilocks program. Failure to follow the "Start" instruction can be addressed with simple software modification. Failure to explore a wide range of settings would require major changes to the self-fitting protocol. It will be important to determine, however, whether failure to explore means that the generic starting condition already falls within a range of optimal settings for that individual, as opposed to other factors such as personality or difficulty with computers. The fact that the group needing help had higher thresholds at 2 and 4 kHz is difficult to interpret in light of the fact that instructions were given textually and not verbally.

The wide range of adjustment changes and times, and the frequent acceptance of starting conditions without exploration, are in keeping with the results reported in Dreschler et al. (2008). That study, however, used several input stimuli and fitting protocols, precluding a direct comparison. In the present study, the total time spent on the two adjustments averaged 86 s for the participants without hearing-aid experience. Time for the speech recognition test ranged from 3 to 6 min with an average of just under 4 min. The total time for this protocol, therefore, averaged around 6.5 min. If further studies show that replication without testing (or with a shorter test) is adequate, completion time will decrease. Even at 6.5 min, it seems unlikely that the time spent on self-adjustment would be a deterrent to practical application. Note, however, that this study

Figure 3. Distribution of 13 hearing-aid users and 13 nonusers in terms of changes in three parameters. Left-hand and center columns show change from the generic starting condition to the first and second self-adjusted conditions, respectively. The right-hand column shows the change from the first to the second self-adjustment. The solid vertical lines show starting values. The dashed vertical lines show group means.

dealt only with adjustment of the basic parameters of amplitude and spectral shape. Many other steps are involved in the self-fitting of direct-to-consumer hearing aids (Convery et al., 2017; Keidser & Convery, 2016).

The average self-adjusted low-frequency attenuation was one step (5 dB), suggesting that two steps (10 dB) was too much for a generic starting condition. Further research will be needed to determine whether the choice of more low-frequency content was affected by sound leakage from the flanged ear-tips used in this study. The average participant selected an overall level that was one step (approximately 4 dB) higher than was used for the generic starting condition. This change reached the .05 level of significance

after, but not necessarily because of, the speech recognition test. Further research will determine whether simple replication, without an intervening test, produces similar changes. There was also evidence of an increase in high-frequency boost after the test (by 2 dB/octave) but only for the hearing-aid users. Further research will explore the possibility that the first self-adjustment was influenced by the characteristics of their existing hearing aids.

The fact that these individuals volunteered to participate in this study suggests a high level of interest and motivation. The positive findings reported here may not apply to a wider population of potential hearing-aid users. Similarly, the findings so far apply to a specific self-fitting approach

that gives users the control of amplitude and spectral shape only. Applicability of this approach to issues such as amplitude compression, directionality, noise management, or frequency shifting has yet to be determined. Note, also, that participants made their adjustments while listening to clear speech, from a male talker, produced with constant effort, in quiet. We have yet to explore the applicability of the Goldilocks approach in more ecologically valid conditions, using a wearable system with microphone input and real-time processing.

Participants in this study were given no guidance about the criteria for adjustment other than “adjust to your liking.” As indicated, the speech recognition test was included on the assumption that it would increase emphasis on intelligibility.

Unfortunately, the test used for this first study was too easy to show any increase. Further work will seek self-report on individual adjustment criteria and will use more sensitive tests that avoid ceiling effects.

Much of the research on hearing-aid fitting has been based on the concept of a single optimal combination of parameters. Departure from this combination is assumed to result in poorer outcomes. It is likely, however, that there is a multidimensional range of parameter combinations for which outcomes are equally acceptable. It could then be argued that the ideal fitting is one that places the listener in the center of this range, thereby allowing for the most changes in signal and listening conditions. Further studies will attempt to provide insights into the acceptable range of parameter combinations and of participants’ self-placements within this range.

Several researchers have shown that starting conditions influence the self-selection of level and/or spectrum (Dreschler et al., 2008; Keidser, Dillon, Carter, & O’Brien, 2012; Mueller, Hornsby, & Weber, 2008). Further work will determine whether starting from an individual threshold-based prescription as opposed to a generic condition, as used here, has a clinically significant effect on self-fitting outcome. We also plan to determine the traits of people who are effective at, and desirous of, self-fitting.

Conclusion

The pattern of participant interactions with the self-adjustment program used here supports the conclusion that this approach to hearing-aid self-fitting can be a speedy, reliable, and feasible alternative to, or supplement to, conventional fitting procedures, but many questions remain to be answered.

Acknowledgment

Research funded by NIH Grant 1R21DC015046 to San Diego State University, Dr. Carol Mackersie, principal investigator, with subaward to University of California San Diego, Dr. Hari Garudadri, principal investigator.

AuD students Alex Lithgow and Mark Datuin assisted with participant testing. AuD students Gregory Hobbs and Jason Duda assisted with analysis of user adjustment steps.

References

- Convery, E., Keidser, G., Seeto, M., Yeend, I., & Freeston, K. (2015). Factors affecting reliability of self-directed automatic in-situ audiometry: Implications for self-fitting hearing aids. *Journal of the American Academy of Audiology*, 26, 5–18. <https://doi.org/10.3766/jaaa.26.1.2>
- Convery, E., Keidser, G., Seeto, M., & McLelland, M. (2017). Evaluation of the self-fitting process with a commercially available hearing aid. *Journal of the American Academy of Audiology*, 28, 109–118. <https://doi.org/10.3766/jaaa.15076>
- Dillon, H., & Keidser, G. (2011). Cost-effective hearing rehabilitation—A role for self-fitting hearing aids? *Trends in Amplification*, 15, 155–156. <https://doi.org/10.1177/1084713812439798>
- Dreschler, W., Keidser, G., Convery, E., & Dillon, H. (2008). Client-based adjustment of hearing aid gain: The effect of different control configurations. *Ear and Hearing*, 29, 214–227. <https://doi.org/10.1097/AUD.0b013e31816453a6>
- Keidser, G., Dillon, H., Carter, L., & O’Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends in Amplification*, 16, 211–223. <https://doi.org/10.1177/1084713812468511>
- Keidser, G., Dillon, H., Zhou, D., & Carter, L. (2011). Threshold measurement by self-fitting hearing aids: Feasibility and challenges. *Trends in Amplification*, 15, 167–174. <https://doi.org/10.1177/1084713812438700>
- Keidser, G., & Convery, E. (2016). Self-fitting hearing aids: Status quo and future predictions. *Trends in Hearing*, 20, 1–15. <https://doi.org/10.1177/2331216516643284>
- Keissling, J., Leifholz, M., Unkel, S., Pons-Kuhnemann, J., Jespersen, C., & Pedersen, J. (2015). A comparison of conventional and in-situ audiometry on participants with varying levels of sensorineural hearing loss. *Journal of the American Academy of Audiology*, 26, 68–79. <https://doi.org/10.3766/jaaa.26.1.8>
- Mueller, H., Hornsby, B., & Weber, J. (2008). Using trainable hearing aids to examine real-world preferred gain. *Journal of the American Academy of Audiology*, 19, 758–773. <https://doi.org/10.3766/jaaa.19.10.4>
- Ruby, C. (2015). *Towards a clinically viable measure of sentence context usage*. Unpublished manuscript, School of Speech, Language, Hearing Sciences, San Diego State University, San Diego, CA.
- Wong, L. (2012). Evidence on self-fitting hearing aids. *Trends in Amplification*, 15, 215–225. <https://doi.org/10.1177/1084713812444009>