Predicting individual hearing-aid preference in the field using laboratory paired comparisons

MARTIN DAHLQUIST*, JOSEFINA LARSSON, SOFIA HERTZMAN, FLORIAN WOLTERS, AND KAROLINA SMEDS

ORCA Europe, Widex A/S, Sweden

Two gain settings were compared in two hearing-aid programs. Twenty participants with impaired hearing evaluated the settings during a two-week field trial period using a double-blind design. During the field test, the participants used a diary to report which program they preferred in various self-selected situations. After the field trial, the participants stated their overall preferred setting in an interview and answered questions about their preferred settings in various predefined sound scenarios. In the laboratory, the participants made paired comparisons of preference, speech intelligibility, comfort, and loudness. The analysis focused on whether the laboratory test could predict the results obtained in the field. On a group level, it looked as if the results from the diary and questionnaire (data from the field) agreed well with the laboratory paired comparisons. However, on an individual level, the laboratory paired comparisons were not effective in predicting real-life preference. Potential reasons for this result and the consequences of the result are discussed.

BACKGROUND

Currently, there is a certain focus in the hearing-device research community on how more realistic laboratory tests should be designed. Another question is how we can collect data that is more sensitive to small signal-processing differences from the test participants' real life.

Testing using paired comparisons (PCs) is often advocated as a sensitive measure when small differences in for instance signal processing are studied (Amlani and Schafer, 2009; Kuk, 2002). However, the correlation between PCs performed in the laboratory and the preference experienced in the field is not commonly documented. The purpose of the current study was to see if laboratory PCs could predict the preference experienced in the field.

METHOD

Twenty participants compared two hearing-aid gain settings in the field and in the laboratory using a double-blind design. In a two-week field trial, the participants compared the two settings in two hearing-aid programs in a balanced design. The following outcome measures were used: An interview that focused on the preferred

*Corresponding author: martin.dahlquist@orca-eu.info

Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies. 5th symposium on Auditory and Audiological Research. August 2015, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C. Dalsgaard, L. Tranebjærg, and T. Andersen. ISBN: 978-87-990013-5-4.

hearing-aid program, a diary for paired comparisons in the field, a questionnaire answered after the field test, and the hearing aid log data. In the laboratory, the participants made paired comparisons of preference, speech intelligibility, comfort, and loudness.

Participants

Twenty experienced hearing-aid users, 8 females and 12 males (average age: 74 years) were recruited from the ORCA Europe database. They all had symmetrical hearing losses and all had experience with hearing aids of other brands than Widex. Measured pure-tone thresholds are found in Fig. 1.

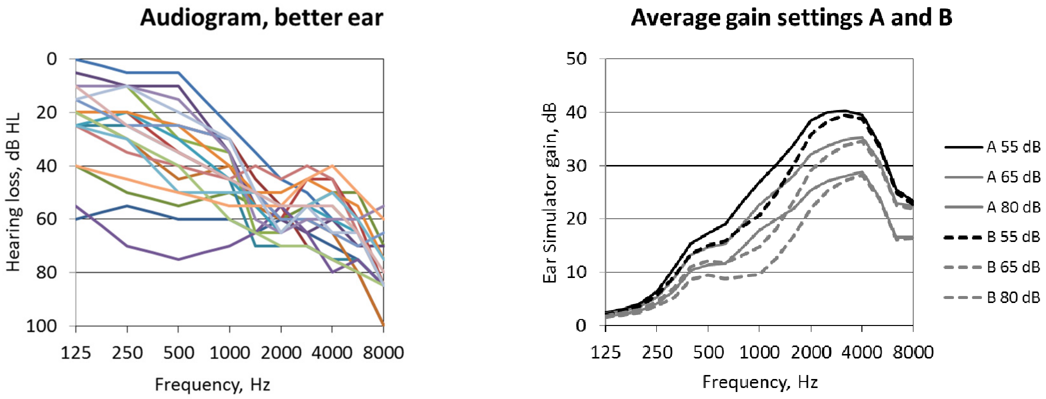

Fig. 1: Left: Better ear audiometric thresholds for all participants. Right: Average hearing-aid gain (Ear simulator) for the two prescriptions A and B measured using speech at three input levels (55, 65, and 80 dB SPL).

Hearing aids and prescriptions

A research receiver-in-the-ear (RITE) hearing aid was fitted using custom earmoulds (15 participants) or standard domes (5 participants) with varying degree of ventilation (selected based on audiogram configuration and type of currently used hearing aids). A directional microphone system and an SII-based noise reduction were switched on. One hearing-aid program was programmed using Widex' general prescription, and the other prescription differed from the Widex prescription by reduction of the gain in a fairly broad region around 1 kHz (Fig. 1, right panel). The hearing-aid fittings were verified and documented using real-ear measurements (Interacoustics Equinox REM440). The hearing-aid settings were also documented using box measurements (Interacoustics Equinox HIT440, equipped with test box TBS25 and an ear simulator G.R.A.S. RA0045).

Field – Diary

The purpose of the diary was to give the participants an opportunity to make direct paired comparisons of the two hearing-aid programs in real-life situations. Seven sound scenario categories were described in the booklet that constituted the diary. Short descriptions can be found in Fig. 2. These sound scenario categories focused on activities and the participants' intent rather than purely on the acoustical environment. Prior to the field test, the participants suggested at least one example of a situation from their everyday life matching each scenario category, and the participants were encouraged to make the evaluation in this specific situation as well as in other situations.

The participants were asked to categorize experienced sound situations into one of the seven pre-defined scenarios, to make a paired comparison of the two programs and write a comment if they wanted to describe why a particular program was chosen.

Field – Interview

When the participants returned to ORCA Europe after the field trial, a structured interview was performed. Overall preference for one of the hearing-aid programs was the main outcome of the interview.

Field – Sound scenario questionnaire

Together with the test leader, the participants also filled out a questionnaire with questions about preference in a number of presented scenarios. For each of the seven main sound scenario categories used for the diary (except category 4), one or two more specific examples, assumed to be encountered by a majority of the participants, were presented. The participants assessed the occurrence of each sound scenario example, which hearing-aid program they preferred in the scenario, and the strength or certainty of their choice. The participants also had the possibility to add four own sound scenario examples that were not covered by the ten pre-defined scenarios. Short descriptions of the scenarios can be found in Fig. 3.

Field – Data log

The hearing aid logged data (active program, sound level, and volume control setting) in 24.4-min intervals during 104 hours.

Lab – Paired comparisons (PC)

In the laboratory, sound field paired comparisons were made for four attributes (preference, speech intelligibility, comfort, and loudness) for a number of stimuli (Table 1). Six loudspeakers (KRK R6, powered by two Rotel RMB-1075 amplifiers) were placed at 1.0 m distance from the reference point (in the middle of the listener's head) at 0, 45, 90, 180, 270, 315 degrees azimuth in a sound-treated test booth.

During the test, the hearing aids were connected to a computer via a USB link. The volume control was set to default (the fitted gain). A Matlab script controlled playback of the test stimuli, the program switching, and the storage of the responses. The sound files were looped and played back continuously, while the participants used a graphical user interface to control which hearing-aid program was active and to indicate their choice of program. They also indicated the magnitude of difference between the two programs (small/moderate/large difference).

For each sound example, the order of the hearing-aid programs was randomized. A pre-conditioning time of 15 seconds was used at the beginning of each new sound stimulus in order for the hearing aids to stabilize their performance. During this time, the program selection buttons were locked. Generally, one round of comparison was made for each rating attribute and sound stimulus, but preference was also assessed at the end of the visit in order to collect retest data.

Stimulus	Level, dB	# Loud-speakers	Prefe-rence	Speech Intell.	Com-fort	Loud-ness
Outdoors with birds	51	2	X			
Speech in Quiet	55	1	X	X		X
Speech in Quiet	65	1	X	X		X
Speech in Cafeteria noise	75/71	1+5	X	X	X	X
Music,string quartet	75	2	X			X
Music, piano	75	2	X			
Soccer chant	85	4	X		X	

Table 1: Paired comparisons: Sound stimuli, presentation levels, loudspeaker setup, and rating attributes.

Relationships between measures

Individual results from field and laboratory tests were collected and processed to allow for correlation analysis. From the field test, the interview gave 1 variable, the diary 7 variables, and the questionnaire 10 variables. From the laboratory test, there were 16 variables from the PCs. The outcomes were transformed into difference measures. Both Spearman's rank correlation and a binomial method were applied.

RESULTS

Field – Interview

During the interview 9 participants stated that they preferred setting A and 11 that they preferred setting B, with varying degree of confidence.

Field – Diary

The participants made paired comparisons in a variety of relevant sound environments, performing various activities. In median, the participants had 27 entries (range 4-80) and these entries were often described in detail. For each participant and each sound scenario category, the preference for A and B was calculated as percentages. Then an average was calculated across all participants. Fig. 2 shows close to equal preference, but setting A (providing more gain) was slightly preferred for live focused listening, whereas B (providing less gain) was slightly preferred for sound monitoring and passive listening.

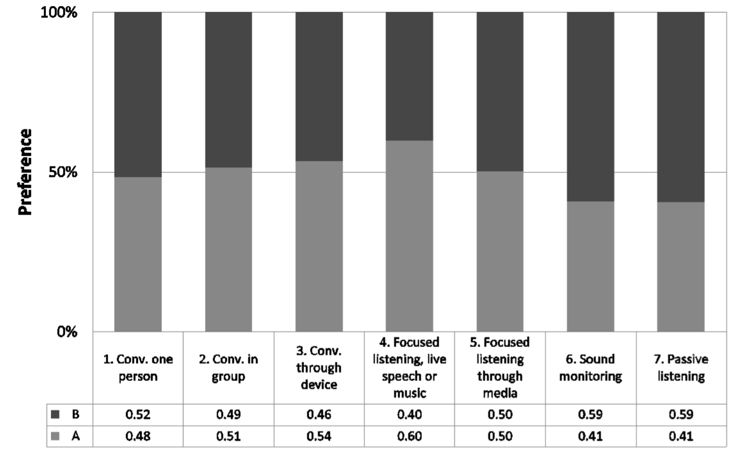

Fig. 2: Diary. Average preference for setting A (light grey) and B (dark grey) for the seven sound scenario categories used in the diary.

Field – Questionnaire

For all participants who had experienced the various predefined scenarios described in the questionnaire, the preference is indicated in Fig. 3. This shows close to equal preference, but setting A (providing more gain) was slightly preferred for speech communication scenarios whereas setting B (providing less gain) was slightly preferred for passive listening scenarios. Only the difference for the last scenario ("Resting on a train") was statistically significant ($p<0.05$, sign test).

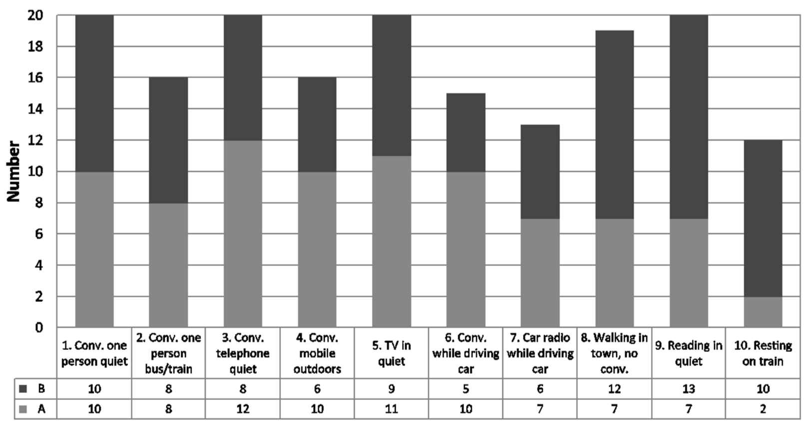

Fig. 3: Number of participants who preferred setting A (light grey) and setting B (dark grey) in the ten sound scenarios described in the questionnaire.

Field – Data log

The data log showed that the hearing-aids were used 11.4 hours per day in median. The preferred setting was generally used more than the non-preferred. The volume control was on average changed equally often up and down for both programs, but the fitted gain was used on average 70% of the time.

Lab – Paired comparisons (PC)

When laboratory PC data for all sound stimuli were pooled (Fig. 4), there was about equal preference for settings A and B. Setting A was preferred for speech intelligibility and setting B for comfort, and A was judged as louder than B (these differences were statistically significant). Differences in the preference pattern were seen for the various stimuli (not shown in the figure): For soft speech there was a preference for A, while B was preferred for speech in cafeteria noise (both differences statistically significant). The interpretation was that preference in the latter scenario was more related to comfort than to speech intelligibility. Statistical testing was done using Wilcoxon signed ranks test with $p<0.05$.

Relationships between measures

Correlation analyses were made with two methods, whose results agreed well. Generally, the results from the correlation analyses showed that a large number of the field outcome variables correlated with each other. Specifically, there were statistically significant correlations ($p<0.05$) with the interview question about preferred setting after the field trial for 6 out of 7 diary sound scenario categories and for 6 out of 10 questionnaire sound scenarios. On the other hand, none of the laboratory PC results correlated with the result of this interview question about overall preference.

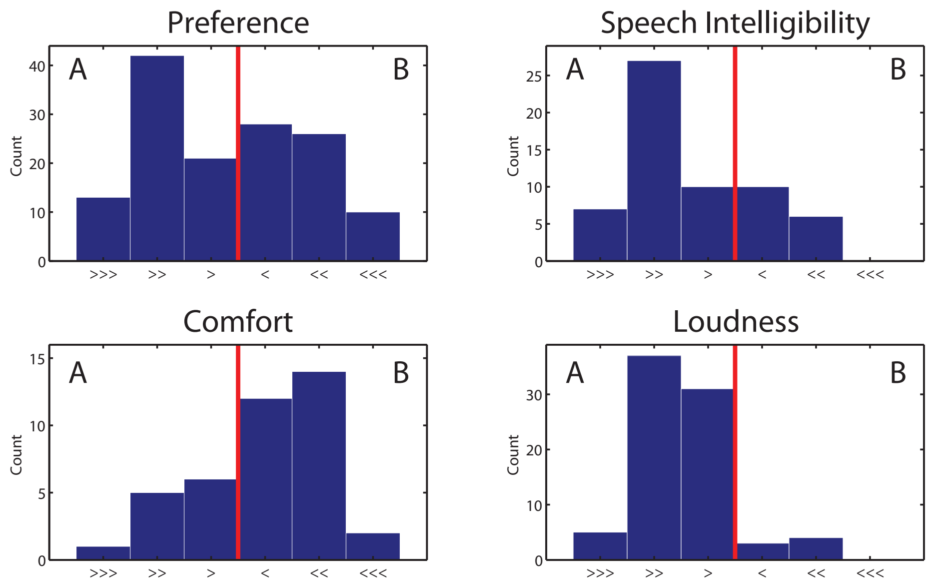

Fig. 4: Laboratory PC results for the attributes when all sound stimuli were pooled. Bars to the left of the vertical line indicate that A was selected more often, bars to the right that B was selected more often. The y-axes represent the total number of ratings. The symbols on the x-axes indicate the magnitude of difference between the two programs.

DISCUSSION AND CONCLUSIONS

The laboratory PC group data seemed to agree with the group data from the diary and the questionnaire. The clear difference between the settings found in the laboratory for speech intelligibility and comfort (Fig. 4) were mirrored by similar tendencies in the field (Figs. 2 and 3). But, on an individual level, the laboratory PC data only correlated with very few of the field test outcomes. Specifically, the laboratory PC data could not predict the overall preference in the field. Potential explanations for this prediction mismatch will be discussed.

If the difference between compared settings is so small that the participants have difficulties detecting the difference, the results found in this study would be expected. But, that did not seem to be the case here. The audiologists who performed the testing reported that the participants did not seem to have any difficulties hearing the difference between the two programs, neither during the field test, nor during the laboratory testing. The ratings done in connection to both field and laboratory paired comparisons also showed that the participants often rated the difference to be at least “moderate”.

Laboratory paired comparisons are sensitive when small differences are evaluated, but there are a number of problems associated with these tests when the results are compared to real-life performance.

One limitation with the traditional laboratory PC setup used in this study is the artificial situation and task. In particular, the participants only listened to speech, instead of participating in a dialogue. This means that potentially important aspects of the signal processing might have been lost. These aspects could be related to the sound quality of the participants’ own voice and to the changes in signal processing perceived when the speech levels change during a dialogue. This could create a difference between the field and the laboratory. At ORCA Europe, we have subsequently tried a dialogue-based paired comparison task. Own voice is included and a more complex activity is created. Initial testing has indicated that this might be a possible way forward.

Another aspect of the difference between the laboratory and the field is the selection of sound stimuli. For the laboratory PC, a fairly large range of presentation levels and sound types were selected (Table 1). These stimuli might not be representative of the situations the participants encountered in real life. This listening situation mismatch could perhaps be overcome by selecting stimuli in the laboratory test based on the situations encountered in the field (reported in the diary). However, a separate analysis of the current laboratory PC data, only including the most commonly experienced stimuli, did not show a better correlation between laboratory and field outcomes.

Further, the loudness difference between the two settings probably play a larger role in the laboratory PCs than in the field, where the volume control could be used. It is also possible that the participants in the laboratory PC focused on some easily identified details, perhaps specific to the recording or talker, in a way that is not done in the field.

In addition to these general shortcomings of the laboratory PCs, it turned out that the test-retest reliability for the laboratory PCs was poor for the speech stimuli, but acceptable for the music stimuli. The number of repetitions was too limited, but there also seemed to be some other methodological difficulties with the test when speech was used. Especially when the speech was easily understood, the participants seemed to find it difficult to judge preference and speech intelligibility.

Some participants also indicated a confusion when using the preference attribute. First, PCs for preference was measured, then followed the three other attributes before a retest for the preference was performed. During the first preference measurements, the participants did not seem to find the task difficult, whereas some of them commented things like “Do you want me to concentrate on speech intelligibility or comfort?” when they were asked to rate preference the second time. That “attribute confusion” did not take place for the music stimuli, for which only preference and loudness were measured, and the preference judgments seemed less complex.

Development of more realistic laboratory tests is one way of improving the evaluation of hearing-aid characteristics. Another strategy will probably be to perform more controlled and sensitive field tests. Advanced logging of field preference using smartphones (e.g., Kissner *et al.*, 2015) and semi-controlled field tests, for instance inspired by “Soundwalks” (e.g., Adams *et al.*, 2008), seem promising.

In conclusion, the current laboratory paired comparisons could not predict outcomes in real life. Suggestions for improving the laboratory paired comparisons (including both basic methodological questions to improve test-retest reliability, and more substantial changes to the task included in the comparisons) have been presented and alternative methods for collecting real-life sensitive data have been mentioned.

REFERENCES

Adams, M.D., Bruce, N.S., Davies, W.J., Cain, R., Jennings, P., Carlyle, A., Cusack, P, Hume, K., and Plack, C. (**2008**). “Soundwalking as a methodology for understanding soundscapes,” Proc. Ins. Ac., **30**.

Amlani, A.M. and Schafer, E.C. (**2009**). “Application of paired-comparison methods to hearing aids,” Trends Amplif., **13**, 241-259.

Kissner, S., Holube, I., and Bitzer, J. (**2015**). “A smartphone-based, privacy-aware recording system for the assessment of everyday listening situations,” In Proc. ISAAR 2015: *Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*. 5th International Symposium on Auditory and Audiological Research, Nyborg, Denmark. Eds. S. Santurette, T. Dau, J.C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN 978-87-990013-5-4. The Danavox Jubilee Foundation, 2015.

Kuk, F.K. (**2002**). “Paired-comparisons as a fine-tuning tool in hearing aid fitting,” in *Strategies for selecting and verifying hearing aid fittings*, 2nd ed. Ed. M. Valente, New York, NY, USA: Thieme medical publishing, pp. 125-150.