Can individualised acoustical transforms in hearing aids improve perceived sound quality?

SØREN LAUGESEN*, NIELS SØGAARD JENSEN, FILIP MARCHMAN RØNNE, AND JULIE HEFTING PEDERSEN

Eriksholm Research Centre, Oticon A/S, Snekkersten, Denmark

This paper presents an experiment which aimed to clarify whether benefits in terms of perceived sound quality can be obtained from fitting hearing aids according to individualised acoustical transforms instead of average transforms. Eighteen normal-hearing test subjects participated, and hearing-aid sound processing with various degrees of individualisation was simulated and applied to five different sound samples. Stimuli were presented over insert phones and evaluated in an A/B test paradigm. Data were analysed with the Bradley-Terry-Luce model. The key result is that hearing aids individualised according to a real-ear insertion gain (REIG) target were preferred over hearing aids individualised according to a real-ear aided response (REAR) target.

INTRODUCTION

When listening with open ears, the sounds that arrive at the eardrum are coloured by the presence of the body, the head, and the detailed structure of the pinna and the ear canal. This colouration is unique to the individual ear. When a hearing aid (HA) is fitted to a person's ear, this colouration is changed, and these changes are taken into account in the hearing aid's amplification in terms of so-called acoustical transforms (ATs). The ATs are typically described in terms of three components: the microphone location effect (MLE), the open ear gain (OEG), and the real ear to coupler difference (RECD). In spite of the aforementioned individual variation, most hearing aids are fitted using average acoustical transforms. By using standardised measures, the individual variation in all three components of the ATs is disregarded. This variation can be quite large, especially at high frequencies. For example, Saunders and Morgan (2003) showed that for the RECD alone, deviations of more than 10 dB are very common at high frequencies. The combined effect of the variation in all three components of the ATs may thus be a substantial difference between the prescribed gain of a HA and the fitting's target.

The individual ATs may be taken into account in the HA fitting by means of real-ear measurements (REMs), but the use of REMs is not widespread (Dillon and Keidser, 2003; Mueller and Picou, 2010). It should be noted that there are (at least) two schools of thought regarding individualisation of HA fittings using REMs. For example, the NAL family of prescriptions (Dillon, 2012) is defined with a real-ear

*Corresponding author: slau@eriksholm.com

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. The Danavox Jubilee Foundation, 2015.

insertion gain (REIG) target. Thus, an REM-based NAL fitting set to a nominal 0-dB insertion gain will seek to recreate the exact same sound pressure level (SPL) in front of the eardrum as would be found in free-field listening with a sound source directly in front of the listener. In contrast, e.g., the DSL prescription (Seewald *et al.*, 2005) is defined with a real-ear aided response (REAR) target. This means that the individual OEG component of the total individual AT will be deliberately ignored and replaced by the standard OEG assumed by DSL. The REAR-target strategy is especially relevant when fitting small children or people with surgically modified ears, that is, when it can be argued that an extreme OEG exists for other reasons than audition (Dillon, 2012).

There are several studies showing that the use of REMs improves on a HA fitting's match to target, (e.g., Aazh *et al.*, 2012; Nelson, 2013). However, as highlighted by Humes (2012) and Mueller (2014), there has been very little research into whether or not using REM leads to any self-perceived benefits for the HA user. In fact, the present authors have identified only one public article (Abrams *et al.*, 2012) that demonstrates a self-perceived end-user benefit of using individually measured ATs in HA fitting. In that study, the APHAB questionnaire and overall preference were used to evaluate two HA fittings tested in the field: one based on standardised ATs and one based on individualised ATs (with REAR targets).

A couple of studies have examined the sound-quality disruptions perceived by listeners due to generic modifications to the frequency response of a reproduction system (van Buuren *et al.*, 1996; Moore and Tan, 2003). Their results indicate that the expected magnitude of differences between individualised and standardised ATs should be perceptible by both normal-hearing and hearing-impaired listeners.

Overall approach

The goal of the experiment reported in this paper was to investigate whether or not it would be possible to measure a sound-quality benefit from using individualised ATs, in the sense that listening through a HA programmed according to individualised ATs would be consistently preferred over listening through a HA programmed according to standardised ATs.

Being a first step, the most advantageous conditions for finding such a benefit were sought for. This involved using test subjects not requiring amplification (implying that the question of selecting gain rule for hearing-loss compensation could be neglected), using laboratory-grade equipment to measure the total individual AT, and ignoring the direction-dependence of the MLE-component of the AT by considering only sound presentation corresponding to the frontal direction. Finally, the experimental stimuli were delivered from Matlab to the test subjects' ears through insert phones.

METHOD AND MATERIAL

Test subjects

N = 18 test subjects (12 female, 6 male; age-range 22-55 years, mean age 37) were recruited. All test subjects had hearing threshold levels (HTLs) at 25 dB HL or better at all audiometric frequencies up to and including 8 kHz, except one test subject who had HTLs of 30 and 35 dB HL at 6 kHz and two test subjects who had 8-kHz HTLs of 35 dB in one ear.

Sound quality experiment

The procedure used for the sound quality assessment was based on an A/B paired-comparison approach (Bramsløw, 2010), where all pairs of the five processing conditions were compared. For each comparison (trial), the test subject had to listen to a stimulus of about 15 seconds duration, which was played back in a continuous loop. The test subject could switch between settings A and B as much and as frequently as he or she desired, using a touch screen. The test subject's task was to determine the preferred setting. When the preferred setting had been indicated, the test subject could start the next trial by pressing the 'Next' button. The user interface also included a 'Pause' button, allowing the test subject to take a break (at any time). This option was chosen by some (but not all) subjects during the main 150-trial test. There was no 'Don't know' option. Thus, the test subjects were instructed to make an arbitrary choice in cases where they had no preference.

Real-ear measurements

The laboratory-grade REM set-up was built around the Brüel&Kjær PULSE audio analyzer system, which was set to carry out two-channel FFT spectrum averaging (6400 spectral lines, 20-kHz bandwidth). The measurements were performed in an anechoic room with the test subject seated in an adjustable chair and sound delivered from a Genelec 8030A loudspeaker. Sound was recorded in the test subjects' ears through probe microphones. The probe microphones were taken from a modified Interacoustics Affinity system, which also served as power supply and conditioning amplifier for the probe microphones. From the Affinity system the microphone signals were routed through a Brüel&Kjær 5935 Dual Microphone Supply to the PULSE system. The measurements comprised loudspeaker free-field response, individual probe-microphone free-field calibration, and individual measurements of open-ear responses as well as aided responses, as described below.

In addition, a standard Interacoustics Affinity system was used for REMs in the clinic, which were used to obtain the $\text{HA0}_{\text{REAR1}}$ setting, see below.

Processing conditions

Five different processing conditions were created, representing different degrees of AT individualisation in a hypothetical HA prescribed to deliver linear amplification with 0-dB insertion gain at all frequencies. The conditions are described in Table 1.

$HA0_{REIG}$	Mimicking a 0-dB insertion gain HA fitted with individual ATs according to a REIG target.
$HA0_{avg}$	This condition was meant to mimic a 0-dB insertion gain HA fitted according to average ATs. However, due to a programming mistake the results from this condition are disregarded.
$HA0_{REAR1}$	Mimicking a 0-dB insertion gain HA fitted with individual ATs according to a REAR target. Based on the automatic AutoFit function in the Genie HA-fitting software together with the Affinity REM system.
$HA0_{REAR2}$	Mimicking a 0-dB insertion gain HA fitted with individual ATs according to a REAR target. Derived directly from the standard OEG used in the Genie fitting software.
$HA0_{REIGlowres}$	Similar to $HA0_{REIG}$, except that the individual ATs were realised with a frequency detail similar to what is available in the Genie fitting software.

Table 1: Labels and description of processing conditions.

Stimuli

The chosen sound samples were recordings of ‘Classical’, ‘Rock’, and ‘Jazz’ music, ‘Speech’ in quiet, and a dialogue in a ‘Canteen’ background. The samples were cut to allow seamless looping and they were scaled to produce reasonable playback levels ranging from 70 to 78 dB SPL (predicted free-field levels). To produce the individual stimuli, the sound samples were convolved with the processing-condition filters described above. In addition, the stimuli were shaped to compensate for the individually measured response of the Etymotic Research ER-2 insert phones used in the experiment. Then, the magnitude-smoothing approach suggested by Schärer and Lindau (2009) was applied using a ¼-octave band filter. Finally, the five condition-specific stimuli for each test person and each sound sample were scaled to have the same predicted A-weighted free-field level, in order to remove any loudness differences which are known to dominate sound-quality evaluations if present.

Test design and protocol

A test design based on counterbalancing of condition pairs and sound samples and with three repetitions was used, which amounts to a total of $10\times5\times3 = 150$ trials. Prior to the actual test trials, 20 practice trials were performed. In each trial, the assignment of the two conditions to the A and B buttons was random.

The experiment comprised one visit for each test subject, starting with the clinical HA fitting session, then the PULSE-based REM session, and finally the sound-quality session.

RESULTS

The A/B preference data were analysed with the Bradley-Terry-Luce (BTL) model (Bradley and Terry, 1952; Luce, 1959), basically following the approach described by Wickelmaier and Schmid (2004). In addition, the individual processing-condition filters were analysed in various ways for the purpose of the correlational analysis.

A/B test

The main outcome of the A/B testing is presented in Fig. 1, which shows the normalised BTL scores for each processing condition together with estimated 95% confidence intervals, based on the data from all test subjects and all sound samples. A high BTL score indicates that the condition is more likely to be preferred in a comparison against another randomly selected condition, and non-overlapping confidence intervals are taken as an indication of a statistically significant difference.

The main result is that $HA0_{REIG}$ was preferred over both $HA0_{REAR1}$ and $HA0_{REAR2}$, which suggests that in consideration of sound quality a REIG-target approach should be preferred over a REAR-target approach. In addition, it is seen that the best preference rating was given to the low-resolution $HA0_{REIGlowres}$. This result was unexpected and is further investigated below.

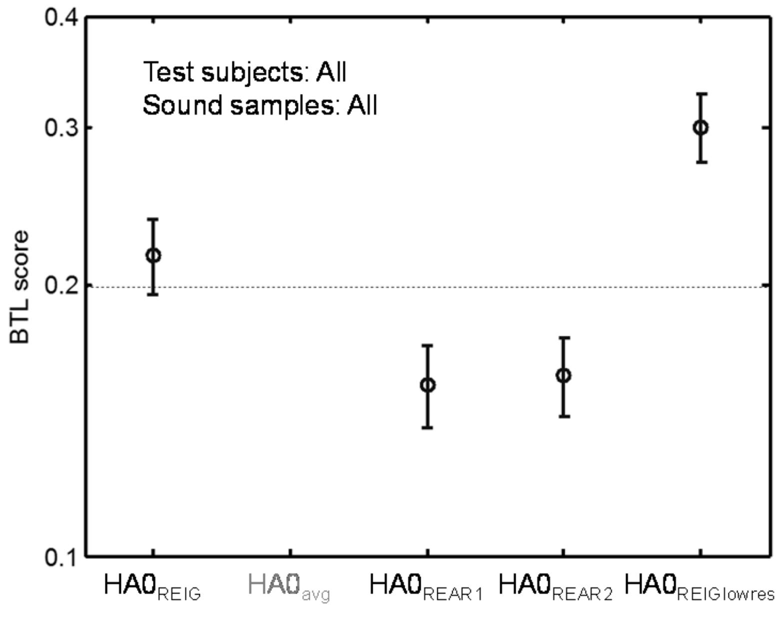

Fig. 1: Overall BTL scores for the five processing conditions with 95% confidence intervals indicated. The dashed line indicates chance level.

A more detailed analysis indicates that the results in Fig. 1 are more pronounced if data obtained only with the music samples are considered, as shown in Fig. 2(a). In contrast, the pattern of results is somewhat different for the speech samples, as indicated in Fig. 2(b). This observation agrees well with the comments from several test subjects, who stated that they applied different criteria for the music and speech samples, mainly because speech intelligibility – rather than sound quality – became the prevailing criterion for the samples including speech.

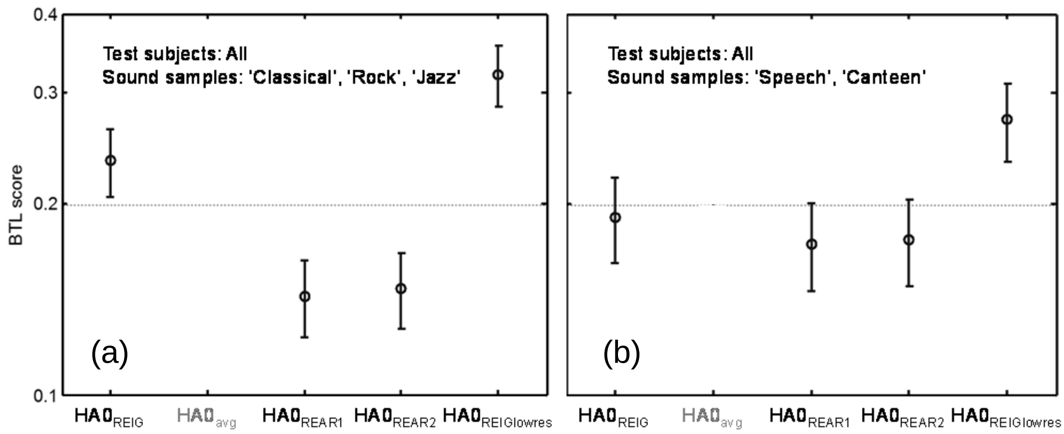

Fig. 2: BTL scores and confidence intervals for (a) the music samples alone, and (b) the samples containing speech.

Furthermore, it should be noted that the individual patterns of preference for some test subjects deviate considerably from the all-test-subjects patterns shown in Figs. 1 and 2. This is not surprising, due to the individual nature of the ATs, which means that the contrasts among the different processing conditions were not the same for all test subjects.

Correlational analysis

The most surprising result from the above BTL analysis was that $HA0_{REIGlowres}$ was preferred over $HA0_{REIG}$. Therefore, it was investigated whether this preference could be related to specific characteristics of the respective processing-condition filter responses. Figure 3(a) shows the mean 'colouration responses' for the two relevant conditions. The colouration responses are similar to the processing-condition filter responses except that the individual insert-phone compensation was removed. The results in Fig. 3(a) show slight systematic differences between the $HA0_{REIG}$ and the $HA0_{REIGlowres}$ conditions (unintended artefacts of the lowres procedure), e.g., a 2-dB boost above 6 kHz and an attenuation in other frequency ranges. The colouration-response differences were averaged across the 6-8 kHz frequency band and across the two ears of each test subject to describe the amount of high-frequency boost. Similar quantities were computed for the attenuation bands. These measures were then used as predictors of the percentage of comparisons where $HA0_{REIG}$ was preferred over $HA0_{REIGlowres}$. None of these predictors turned out to be significant on a 5% level. Nevertheless, the most striking relation (involving the 6-8 kHz high-frequency boost) is illustrated in Fig. 3(b). Although the correlation ($r = -0.40$) is not significant ($p = 0.10$), a larger high-frequency boost ($HA0_{REIGlowres}$ over $HA0_{REIG}$) seems to be associated with stronger preference for $HA0_{REIGlowres}$ (low % values in Fig. 3(b)).

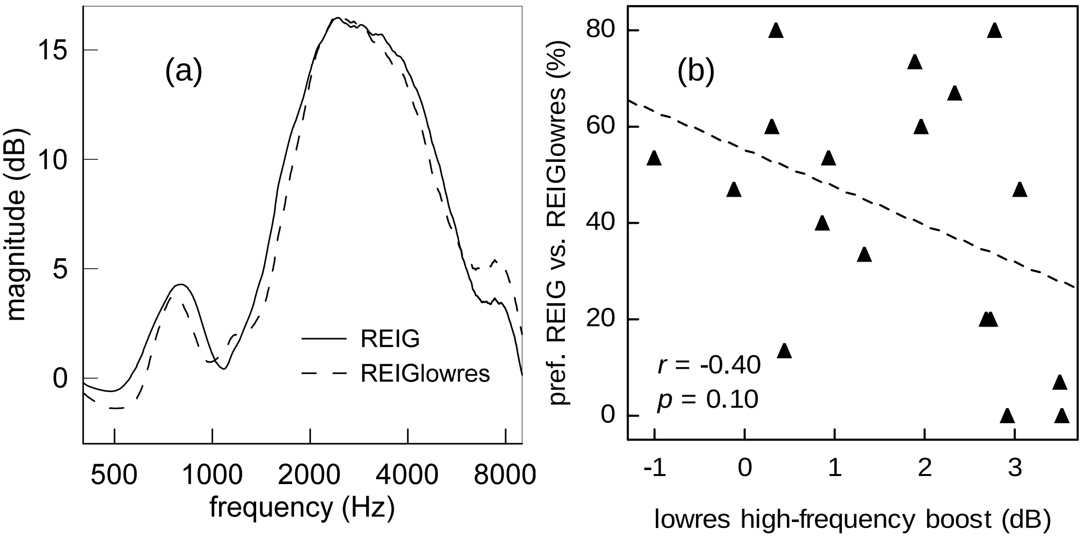

Fig. 3: (a) Mean colouration responses, as indicated. (b) Relation between the lowres high-frequency boost predictor variable and the preference for $HA0_{REIG}$ over $HA0_{REIGlowres}$. The preference data from the 18 test subjects were averaged across the five sound samples and three repetitions.

DISCUSSION

This study investigated the potential sound-quality benefit from fitting hearing aids according to individualised acoustical transforms instead of average transforms.

The key result was that (simulated) hearing aids individualised according to a real-ear insertion gain (REIG) target were preferred over hearing aids individualised according to a real-ear aided response (REAR) target. An earlier study (Abrams *et al.*, 2012) found benefits from individualising to a REAR target relative to a non-individualised approach. However, the main outcome measure used in the Abrams study (the APHAB questionnaire) assesses different benefit domains than sound quality – The APHAB's predefined sub-scales are: Ease of Communication, Reverberation, Background Noise, and Aversiveness of Sounds.

In addition, representing the individualised transforms in lower frequency resolution was preferred over the representation in fine spectral detail. The analysis suggests that this may be because of an artefact of the low-resolution representation which added a slight boost in the 6-8 kHz frequency range. Recall that the test subjects in this study had normal hearing, and therefore might appreciate the brighter timbre brought about by the lowres high-frequency boost. A similar outcome is not expected for test subjects with high-frequency hearing loss, who are less likely to appreciate additional amplification in the frequency range above 6 kHz (Dillon, 2012). The present study was limited to normal-hearing test subjects and simulated hearing-aid processing. Hence, the next step is to perform a similar investigation with hearing-impaired test subjects listening through real hearing aids.

REFERENCES

Aazh, H., Moore, B.C., and Prasher, D. (**2012**). "Real ear measurement methods for open fit hearing aids: Modified pressure concurrent equalization (MPCE) versus modified pressure stored equalization (MPSE)," Int. J. Audiol., **51**, 103-107.

Abrams, H.B., Chisolm, T.H., McManus, M., and McArdle, R. (**2012**). "Initial-fit approach versus verified prescription: Comparing self-perceived hearing aid benefit," J. Acoust. Soc. Am., **23**, 768-778.

Bradley, R.A. and Terry, M.E. (**1952**). "Rank analysis of incomplete block designs: I. The method of paired comparisons," Biometrika, **39**, 324-345.

Bramsløw, L. (**2010**). "Preferred signal path delay and high-pass cut-off in open fittings," Int. J. Audiol., **49**, 634-644.

Dillon, H. (**2012**). *Hearing Aids*. 2nd ed., Stuttgart, Germany: Thieme.

Dillon, H. and Keidser G. (**2003**). "Is probe-mic measurement of HA gain-frequency response best practice?," Hear. J., **56**, 28-30.

Humes, L.E. (**2012**). "Verification and validation: The chasm between protocol and practice," Hear. J., **65**, 8-10.

Luce, R.D. (**1959**). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Moore, B.C.J. and Tan, C.-T. (**2003**). "Perceived naturalness of spectrally distorted speech and music," J. Acoust. Soc. Am., **114**, 408-419.

Mueller, H.G. (**2014**). "20Q: Real-ear probe-microphone measures – 30 years of progress?" Audiology Online, article 12410.

Mueller, H.G. and Picou, E.M. (**2010**). "Survey examines popularity of real-ear probe-microphone measures," Hear. J., **63**, 27-28.

Nelson, S.R. (**2013**). "The impact of using real ear measures to calculate prescriptive targets on hearing aid follow-up visits," Retrieved from: http://digitalcommons.wustl.edu/pacs_capstones/669/

Saunders, G.H. and Morgan, D.E. (**2003**). "Impact on hearing aid targets of measuring thresholds in dB HL versus dB SPL," Int. J. Audiol., **42**, 319-326.

Schärer, Z. and Lindau, A. (**2009**). "Evaluation of equalization methods for binaural signals," Proc. 126th Audio Engineering Society Convention, **1**, 15-31.

Seewald, R., Moodie, S., Scollie, S., and Bagatto, M. (**2005**). "The DSL method for pediatric hearing instrument fitting: Historical perspective and current issues," Trends Ampl., **9**, 145-157.

van Buuren, R.A., Festen, J.M., and Houtgast, T. (**1996**). "Peaks in the frequency response of hearing aids: Evaluation of the effects on speech intelligibility and sound quality," J. Speech Lang. Hear. Res., **39**, 239-250.

Wickelmaier, F. and Schmid, C. (**2004**). "A Matlab function to estimate choice model parameters from paired-comparison data," Behav. Res. Meth. Instr. Comp., **36**, 29-40.