Evaluating sound quality in hearing aids with reference test audiograms

CARSTEN DAUGAARD[1], SØREN L. JØRGENSEN[1], CHRIS V. JØRGENSEN[1], SØREN V. LEGARTH[2] AND NICK V. ZACHAROV[2]

1 DELTA Technical-Audiological laboratory, Edisonvej 24, 5000 Odense C, Denmark

2 DELTA SenseLab, Venlighedsvej 4, 2970 Hørsholm, Denmark

In the draft of the IEC 60118-15 [2008] standard a set of reference test audiograms are described. These audiograms are defined in order to give a common set of fittings for hearing aids, exposed to broadband measurements.

The audiograms also provide a fine basis for the purpose of evaluating the perceived sound quality from various fitting strategies in hearing aids.

Evaluating sound quality involves creating a number of sound stimuli to evaluate the performance of the fitted hearing aids. By defining and recording sound stimuli from sound scenarios that are expected to be challenging to hearing aids, as many sound attributes as possible are elicited for data analysis.

For the evaluation of sound quality, descriptive analysis, a method known from the food industry, is applied. For sound evaluation a panel of normal hearing trained assessors is used.

Results from the test indicate that differences between the hearing aids fitted with the standard audiograms are clearly audible for a number of sound attributes.

INTRODUCTION

In recent years, open fitted hearing aids with relatively low gain have been a commercial success. Although these undoubtedly provide the basic benefits of hearing aids; level and frequency dependent amplification, which in turn eases the listening effort as well as provide a better speech understanding in noisy situations, they will change the perceived sound from that of the un-amplified ear. In most situations this coloration is desirable at least for the speech understanding, but how does it affect the perceived quality of the sound from the hearing aid?

In order to investigate sound quality across different hearing aids, a reproducible and yet realistic setting of the instruments are needed. Since the fitting algorithm is an important part of a hearing aid today a natural place to start is to fit to a predetermined audiogram. To preserve objectivity, and avoid unintentional bias the collection of data from the sound, evaluation needs to be performed in a structured and objective manner. In this article a method for preparing and evaluating the hearing aids is presented.

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

METHOD

A hearing aid is a very complex sound processor with several partly independent factors affecting the quality of the sound. The choice of hardware will affect the sound quality, as well as the number of algorithms which are enabled in the instrument. Where the hardware and to some extend the fitting settings are stationary, some algorithms i.e. noise suppression add a time-variant element to the case. This increases the complexity of the evaluation and thus demands for a careful selection of sound stimuli to exercise all the realistic combinations of features in the hearing aid.

The sound scenarios for this test were selected from common experience with typical hearing aid challenging events, like traffic noise, at a mall and TV sound. Also scenarios including speech perception were used (Dantale Samsoe story, recording of a 3 persons meeting). Other more inventive and perhaps more selective stimuli could have been presented but the chosen set seems to cover the most obvious scenarios. The selected sound scenarios can be found in Table 1.

Stimuli no.	*Stimulus name:*	*English title:*	*Original recording*
1	DANTALE, Samsøhistorien	DANTALE, Samsoe History	Stereo/CD
2	Pausesnak, overvejende kvinder	Pause talk, mainly by women	Sound Field
3	Musik: Jazz, sangerinde	Music with female Jazz singer	Stereo/CD
4	“Møde”, 1 mand, 2 kvinder	Meeting, 1 man, 2 women	Sound Field
5	Supermarked	Supermarket	Sound Field
6	Keyboard tasteri	Typing on PC keyboard	Sound Field

Table 1: List of sound scenarios used to test the hearing aid sound reproduction.

Finding the right adjustment for the hearing aids is not as simple a task as it appears to be at a first look. Traditionally, when evaluating hearing aids these are fitted to the user, and the user is asked to answer a questionnaire. Typically, this inventory covers benefits and limitations of the user’s new hearing aids.

Since the fitting is individual, a comparison of the instruments performance rather than the effect from the instrument on the user is difficult to obtain. Historically, in evaluation of hearing aid coupler measurements either a full on gain setting or a reference test gain (RTG) setting has been applied. While the full on gain was an easy technical fix point with all trimmers and volume control in max, the RTG was designed as a technically well defined setting bringing the hearing instrument in a realistic user setup. However, since the initial fitting of a hearing aid has been an increasingly more integral part of the hearing aid, technically defined settings are not able to account for all the possibilities in the fitting software.

One approach to align the hearing aid fittings is to use the adjustment possibilities in the fitting software to reach the same insertion gain (IG) or coupler gain for all hearing aids to be compared. But to match all the IG curves completely will be troublesome and to keep control over the setting of other features in the hearing aid could prove difficult. The clear advantage of this approach of course is that no frequency coloration between the hearing aids appears. Differences in other attributes are not blurred i.e. by a dominating treble. Another matter is what target to aim for when matching the IG curves? A number of generic fitting prescriptions still exist, but will the performance of the hearing aid be reliable when it is forced to emphasis frequency areas that are not natural parts of the strategy for that product?

Rather than evaluating on a complete hearing aid system, another approach could be to focus on a specific part of the hearing aid i.e. the transducer. But although a breakdown in several details of the sound processing in the hearing aid could be very interesting work, it is the sum of the contributions from all parts which results in the sound quality that the user in the end will be listening to. A complex interaction between different algorithms and hardware will only be properly exited in a complex setup where all factors are as realistic as possible.

The third approach to be considered is to use a predefined audiogram. The audiogram provides the basis for a complete and repeatable setting of the hearing aid, along with the entering of a few other options (i.e. user experience, cognitive capabilities and ear mould type).

In the typical clinical fitting situation an audiogram is applied and the fitting software calculates the settings for the instrument for this particular hearing loss. Although some adjustments will be made in the fine tuning of each hearing aid which probably also are of importance to the sound quality, it is the initial settings most hearing aid users set out their listening experience with.

Due to the impact of the initial fitting on the sound quality it is fair in this experiment to consider the hearing aid to be a black-box, filling in the audiogram and sound stimuli in one end, receiving sound in a given quality at the other.

This approach is also used as the basis for the IEC 60118-15 draft standard “signal processing in hearing aids” (IEC 60118-15, 2008). In this standard a set of “standard audiograms” are proposed. One Table is covering a range of 7 hearing losses ranging from a very mild to a very profound hearing loss. A special table includes 3 steeply sloped hearing losses, appropriate for the open fitted hearing aids. (All 10 hearing losses are shown in Fig. 1). The scope of the IEC standard is to obtain level vs. frequency measurements of the standard fitted hearing aids to evaluate hearing aid performance. The measurements commonly referred to as speech gain are carried out with a specific speech-like signal (ISTS) in a coupler. The method is described in the IEC 60118-15 draft-standard. The International Speech Test Signal (ISTS) (Holube, and EHIMA-ISMADHA working group, 2008) is developed in Oldenburg Hörcentrum as an artificial speech comprised of phonemes of several languages.

When the scope is to evaluate the performance of hearing aids in terms of sound quality, the standard audiogram approach seems quite straight forward, fitting the hearing aids with one or more of the audiograms, expose the instruments to a selection of sound scenarios and evaluate the outcome. The major drawback from this approach is that the differences in frequency- response as a result of different fitting strategies can be very dominating and that might be a problem when evaluating other attributes of the sound not directly related to this coloration.

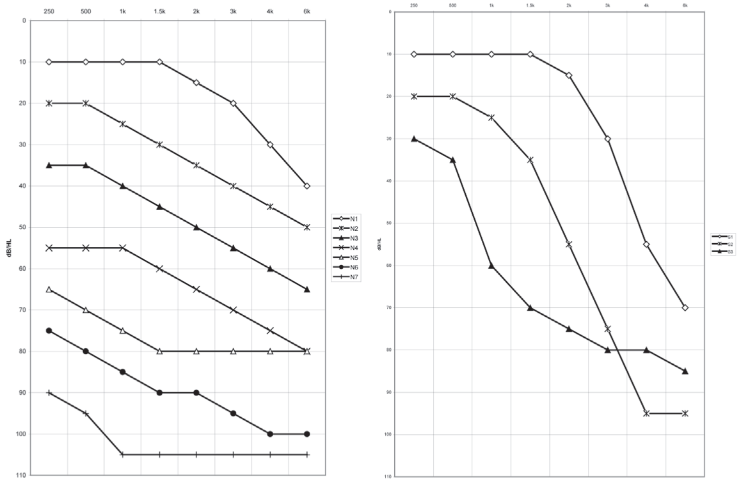

Fig. 1: Standard audiograms for hearing aid fitting, as described in IEC 60118-15 (2008) . Left pane shows N1 to N7 gradually increasing hearing losses, right pane shows S1 to S3 steeply sloped hearing losses.

With the correctly fitted hearing aids and a good selection of sound scenarios the recordings for the evaluation were made. The fitted hearing aids were placed on a B&K head and torso simulator (HATS B&K type 4128C) in the center of a surround sound setup in an anechoic room. The sound scenarios were played back trough the surround sound setup. The resulting HATS recordings were shortened and presented as stimuli in headphones controlled by an evaluation program running on a computer.

Selecting the persons for the evaluation of the sound stimuli is also worth consideration. In the field of food sensory a trained assessor panel is often used. The panel simply exercises in tasting or smelling very small differences in an attribute for a given product. In similar manner a panel of trained listeners is established in DELTA, and a subset of this panel was used for the evaluation of the recorded stimuli. The seven members of the listening panel are screened for normal hearing (better than 15 dB HL across all frequencies), as it is assumed that normal hearing listeners will be at least as capable of evaluating small changes in quality as the slightly hearing impaired listeners who normally wear the products. Furthermore the assessors are trained in sensory evaluation, improving their ability to consistently judge sound attributes.

The attributes to be assessed were selected by the panel themselves in a so-called word elicitation process, where words associated with listening to a set of the sound stimuli were noted and through a discussion in the panel grouped into a set of attributes characteristic for hearing aid sound perception.

RESULTS

From an initial study where hearing aids were fitted with two different audiograms, the N1 and S1, (see the upper curve for both panels in Fig.1) it was learned that naturally the steeper audiogram shape of S1, compared to N1 gave rise to changed frequency responses in the hearing aid insertion gain, and thus an impression of a high frequency emphasis for recorded stimuli with the S1 fitting. The initial study did not show any particular differences between other attributes from the two stimuli set, and thus only the most radical setting, S1, is used in the main experiment. This choice is made based on the argument that the heavier hearing loss must have the highest probability of eliciting attributes as it simply contributes more to the resulting sound.

From an earlier pilot study we learned that significant differences between hearing aids fitted with a standard audiogram could be perceived, even with non-trained listeners. The main test confirmed this and shows that trained listeners can perform consistently in terms of discriminability, panel agreement, repeatability of judgments and a fair use of the scale for the majority of the tested attributes. However, it must be expected that some attributes are linked to specific sound scenarios, and maybe even to specific hearing aids. Most obvious is the attribute “speech reproduction”, which is only meaningful when evaluated in the two scenarios with understandable speech. The attributes evaluated in the project are found in Table 2.

Generally, in the main study for each attribute 2-3 hearing aids with significantly different average scaling can be found. Which hearing aids that differ may change across the attributes but there seems to be a trend towards forming some groups of hearing aids with similar sound characteristics. As could be expected, the attributes more directly linked with the properties of the spectral and gain adjustment of the fitting process such as “Baggrundsstøj (Background noise)”, “Diskant (Treble)” and “Lydstyrke (Loudness)” has a high panel agreement. But the panelists are in good agreement with the attributes “Rumklang (Reverberation)”, “Overstyring (Overload/ distortion)”, and “Detaljer (Details)” as well. Also for “Baggrundsstøjens klangfarve (Background noise tone color)”, there is agreement if one of the panelists is excluded. Thus it seems that it is possible to focus on and evaluate other attributes than those directly affecting the spectral properties of the sound. In other words: Even if the use of a reference audiogram causes different sound colorations due to different fitting algorithms, it is still possible to hear other attributes presumably not directly linked to the fitting process, and rate them independently.

Danish	***English***
Baggrundsstøj (Svag – Kraftig)	Background noise (Soft – Loud)
Baggrundsstøjens klangfarve (Mørk – Lys)	Background noise tone color (Dark – Bright)
Detaljer (Få – Mange)	Details (Few – Many)
Diskant (Svag – Kraftig)	Treble (Soft – Loud)
Dynamikområde (Lille – Stor)	Dynamic range (Small – Large)
Lydstyrke (Svag – Kraftig)	Loudness (Soft – Loud)
Overstyring (Lidt – Meget)	Overload/ Distortion (A little – A lot)
Resonans (Lidt – Meget)	Resonance (A little – A lot)
Rumklang (Tør – Rungende)	Reverberation (Dry – Booming)
Talegengivelse (Uklar – Klar)	Speech reproduction (Unclear – Clear)

Table 2: List of attributes and their translation into English.

The most fruitful way of looking at the sensory profile of a hearing aid is probably to look at the spider web plots as can be seen in Fig. 2. The spider web plot offers a graphical representation of the mean scaling of each hearing aid on each of the attributes. The pattern of each hearing aid in the web depicts which attributes that have the most impact on the sound quality of the given hearing aid.

As an example of reading the spider web plot it is seen in Fig. 2 that the profiles of the hearing aid represented with a “X” scores quite high on “Diskant (Treble)” “Lydstyrke (Loudness)” “Overstyring (Overload/Distortion)” and “Baggrundsstøjens klangfarve (Tonecolour of backgroundnoise)”. The hearing aid represented with an “O” have high scores on “Talegengivelse (Speech reproduction)”, “Dynamik (Dynamic)” and ”Detaljer (Details)”. Clearly these two hearing aids have very different sound, one of them probably more preferable than the other. A spider web plot for all hearing aids are drawn for each scenario, so the same profile may not be expected from the same hearing aid, if the different scenarios cause the hearing aid to “react” differently to the sound stimuli.

Another way of evaluating the sensory profile is to look at the actual scaling pr. scenario for each hearing aid, and then consider the ranking of hearing aids that are scaled significantly different (see Fig. 3). Looking at such plots reveals some interesting points:

1) For some attributes one hearing aid seems to “stand out”, with scores that clearly differ from the rest.

2) The hearing aids in this test are quite consistently ranked across sound scenarios.

3) For all attributes it is possible to group the hearing aids in 2-3 groups which are scaled significantly different.

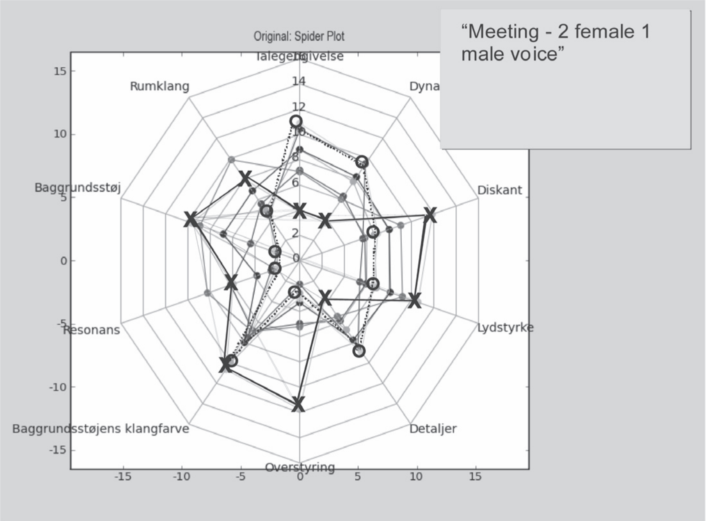

Fig. 2: Spider web plot of the attributes for each of the 7 hearing aid (7 shades of gray) for scenario 4 "meeting". Two very different hearing aid profiles are enhanced and marked with "O" and "X". See Table 2 for an English translation of the Danish attributes.

Indeed there are significant differences on single attributes for one of the hearing aids. Looking at Fig. 3, in the attribute "Diskant (Treble)" especially HA5 stands out providing much more high frequency gain than the rest across all scenarios, which is particularly obvious in the scenario 4 "meeting" and the scenario 5 "supermarket". This is interesting as there seems to be very little in common between these scenarios. Even more significant is the evaluation of the same hearing aid on the attribute "Overstyring (Overload/distortion)" where it is evaluated much higher across all scenarios. Other hearing aids also deviate in the evaluation across scenarios although not that strong. The group comprising of HA1, HA3 and HA4, is performing differently for the scenario 6 "type on a keyboard" on the attributes "Rumklang (Reverberation)", "Resonans (Resonance)", "Baggrundsstøj (Background noise)". It is interesting to note that HA1 is scaled significantly higher in "Baggrundsstøj (Background noise) for this scenario. This can be explained from Fig. 3 by looking at the scaling across all scenarios in "background noise". It seems like the differences between the scaling of the 3 hearing aids HA1, HA3 and HA4 decrease in the rather noisy scenarios 3 and 5 (Jazz and supermarket), but increase in the quieter scenarios (i.e. scenario 1, 4, 6). Therefore it could be speculated that the two hearing aids HA3 and HA4 apply different use of compression/expansion than HA1.

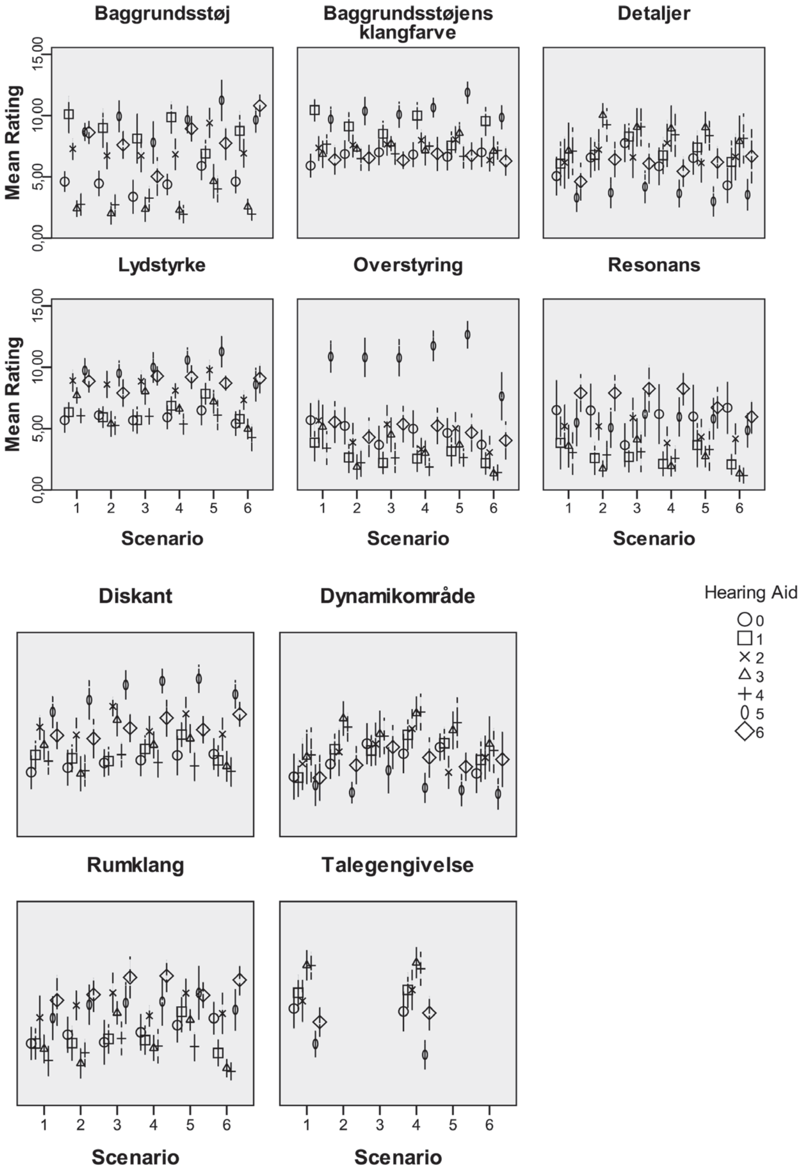

Fig. 3: Mean ratings for all hearing aids per attribute and scenario. 95% confidence intervals are shown. The number on the x axis refer to the number of the scenario as described in Table 1.

It is interesting to note that the difference in ranking of the hearing aids across scenarios is rather small. It can be seen as a display of robustness of the attributes that they are evaluated uniformly in all scenarios. But if it is assumed, that special scenarios can trigger certain attributes in certain hearing aids, it could also be an indication that the selection of attributes and/or stimuli is not comprehensive enough.

To be able to rank the hearing aids is an important tool in investigating their performance. Looking at the relative placement of a hearing aid on different attributes and in different scenarios, makes it possible to investigate the rationale behind the fitting. In some situations the ranking of the hearing aids can be explained by guessing on the behavior of the fitting rationale. As an example from Fig. 3: In background noise for the stimuli “Dantale” and “meeting” the hearing aid labeled HA1 seems to be more noisy, while in the scenario 2 (“girls talk”) and scenario 5 “supermarket” the hearing aid HA5 tends to be the noisier. Since the two first mentioned scenarios are closely linked to speech understanding, it could be speculated that HA1 applies much amplification in these scenarios, but in situations that are detected as more noisy, it reduces its amplification, where perhaps HA 5 keeps a rather large gain in all 4 situations, maybe because it detects some speech in all 4 situations.

CONCLUSION

7 hearing aids have been fitted with a standard audiogram and perceptually evaluated. The main result is that hearing aid sound quality can be consistently evaluated with a panel of trained listeners.

Evaluation of all attributes result in groups of significantly differently scaled hearing aids. For the majority of attributes there is a good panel consensus of the judgments. All scenarios seems to elicit the attributes and results in a fairly consistent ranking of groups of hearing aids. The lack of extremely scaled hearing aids for specific scenarios contradict the assumption that some hearing aids react to specific combinations of stimuli and attributes, or indicates that the set of stimuli used for this test has not been selected for that purpose.

The approach of applying standard audiograms does allow for evaluation of a number of attributes, also some not directly connected with the spectral shaping of the fitting process.

The changes in ranking of the hearing aids across attributes and scenarios inspires to explanations related to fitting concepts analogue to the exercise of the technical measurements described in IEC 60118-15 (2008), which indicates that the fitting approach also applies nicely to perceptual evaluations.

REFERENCES

Holube, I., and EHIMA-ISMADHA working group (**2008**). ”Short description of the International Speech Test Signal (ISTS)”. EHIMA - European Hearing Instrument Manufacturers Association.

Holube, I., and EHIMA-ISMADHA Working Group (**2007**). “Short description of the international speech test signal (ISTS),” EHIMA - European Hearing Instrument Manufacturers Association.

IEC 60118-15 (**2008**). “IEC TC43/WG13: Committee draft 1: IEC60118-15 Signal Processing in hearing aids”.