Perceptual evaluation of noise reduction in hearing aids

Brons, I.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
A method to remove differences in frequency response between commercial hearing aids to allow direct comparison of the sound quality of hearing aid features

Rolph Houben, Inge Brons, Wouter A. Dreschler

Adapted from *Trends in Amplification, 2011, 15(1), pp.77-83.*
2.1 Introduction

Most hearing aids currently marketed have advanced signal processing schemes implemented, such as noise reduction. In our experience, many clinicians do not actively select such techniques or their fitting options to meet the requirements of an individual hearing-impaired listener. One reason for this is a lack of knowledge about the processing details and their perceptual effects for the user. For instance, most research into noise reduction in (commercial) hearing aids was done by comparing different settings within the same hearing aid (e.g. Bentler 2005; Boymans and Dreschler 2000; Mueller et al. 2006). However, a clinician needs to be able to choose also between devices. Unfortunately, direct perceptual comparisons of the sound quality between different devices are uninformative because the perceptual effects are largely determined by other parameters not related to the signal processing under investigation. For instance, the frequency-dependent hearing aid gain can differ substantially across hearing aids, even for hearing aids fitted to the same hearing loss (e.g., Mueller et al. 2008, showed differences up to 15 dB). There can also be large differences in perceived sound quality. Legarth et al. (2010) fitted four hearing aids according to the same fitting rule and found that for normal-hearing listeners these four aids differed markedly in subjective sound quality (ranging from between “poor” and “fair” to “good” on a mean opinion scale). These examples illustrate clearly that audible differences between hearing aids cannot be removed just by fitting them to the same hearing loss. Spectral characteristics can strongly influence a sound quality percept (Gabrielsson et al. 1988, Davis and Davidson 1996). For instance, Gabrielson and Sjögren (1979) did an experiment in which subjects had to describe the sound of eight different headphones. They found that the headphone with a 10-dB peak in the frequency response at 3 kHz scored strongly on adjectives related to “harp/hard/loud” and on adjectives related to “disturbance”. In general, smoother frequency responses lead to better sound quality judgements (Arehart et al. 2010) and can improve the threshold of discomfort (Warner and Bentler 2002).

In conclusion, there is need for a method that allows for perceptual comparison between features of hearing aids by removing the (usually large) differences in frequency response between devices. In this paper we will therefore answer the following research question:

Q1. Is it possible to reduce the perceptual differences between a set of hearing aid recordings so that the recordings are indistinguishable from each other, with the following three successive steps:
a. careful manual adjustment of the insertion gain of the hearing aids;

b. limitation of bandwidth of hearing aid recordings;

c. application of an inverse filter on the hearing aid recordings?

To answer this, we recorded the output of a selection of hearing aids and these recordings were processed in three varying degrees (careful adjustment of the insertion gain, adjustment with bandwidth limitation, and inverse filtering with bandwidth limitation) to minimize differences between them. A sound quality model was used to determine objective differences in quality between the hearing aids in each set. Additionally, we did two listening experiments with six normal-hearing subjects. In the first experiment the subjects had to detect which sound sample differed from two other identical samples. The outcome was the percentage of times the subjects could detect differences between the hearing aids, within each set of stimuli. Finally, we did a paired-comparison test in which the subjects had to indicate which sample they would prefer for long-time listening. This test was meant to measure the effect of our processing on the sound quality of the recordings.

2.2 Methods

2.2.1 Experimental setup

All recordings and experimental validations were done in a sound-treated double-walled booth (2.20 x 2.53 x 2.0 m). The recording system consisted of a B&K Head and Torso Simulator (HATS Type 4128C) fitted with a custom made tight-fitting ear mould without venting. Sound signals were generated and recorded monaurally at a 44100-Hz sample rate with a resolution of 24 bits. The digital signals were converted to the analogue domain with a RME Fireface 800 sound card, and were presented to the hearing aid via a Samsung Servo 120 amplifier connected to a Tannoy Reveal 6 near-field monitoring speaker that was placed at 62 cm in front of the recording microphone (on axis). All free-field hearing aid input signals were corrected for the speaker response and all signals were presented within the direct sound field to minimize the influence of room reflections.

The hearing aids used in this study were five frequently used BTE hearing aids from different brands (Oticon Vigo Pro, Phonak Exélia M, ReSound Azure AZ80-DVI, Starkey Destiny 1200, Widex Mind 440), randomly coded as HA1 to HA5. All signal processing features (directionality, feedback control, noise reduction, compression, frequency transposition, etc.) were turned off.
2.2.2 Stimuli

We recorded the hearing aid output for speech (Versfeld et al. 2000) in speech babble (Luts et al. 2010). We used speech in noise because (a) this is the target signal for most signal processing features in hearing aids, and (b) possible remaining differences in both the target speech and the background noise can be taken into account. The signal-to-noise ratio was chosen to be +10 dB because this is a relevant ratio for speech in noise experiments and it is high enough to allow perception of possible distortions and colouring to both speech and noise. Note that all hearing aids add noise to the signal. In our selection of hearing aids, the specified equivalent noise input level was between 20 to 30 dB SPL, and this resulted in a noise level about 45 dB lower than our average speech level. This was assumed not to influence the quality of the recordings of our speech in speech-shaped babble noise (at +10 dB), as the low-level noises will be masked by the background noise.

Three sets of stimuli were made to answer the three parts of the research question. Set 1 consisted of the unprocessed hearing aid recordings that were made after manual adjustment of the insertion gain (i.e. the difference between aided and unaided response). Set 2 was based on the same recordings, but the signals were limited in bandwidth, and in set 3 these bandwidth limited recordings were also filtered with an inverse filter to remove differences in frequency response.

2.2.3 Stimulus set 1: Manual adjustment of the insertion gain

During the hearing aid fitting, the insertion gain was measured in-situ with pink noise. In order to simulate a realistic condition we selected a conductive hearing loss of 30 dB at 500 Hz and 15 dB at 2 kHz, that resulted in a NAL-RP prescription (Dillon 2001) of about 10 dB insertion gain in the low and mid frequencies. More precisely, the target insertion gain was 4 dB between 100 Hz and 125 Hz; 10 dB between 125 Hz and 2 kHz; decreasing to 0 dB at 2 kHz, and it was 0 dB from 4 to 6 kHz. This frequency range (100 Hz to 6 kHz) was within the specified operational frequency range for all hearing aids except for HA5 (its specified operational low-end frequency is 200 Hz, but the aid was verified to give reliable output to at least as low as 100 Hz). Although the fittings were carefully adjusted to obtain the same insertion gain for all hearing aids, several peaks and valleys remained in the responses, making them different from each other and from a flat frequency response. These remaining differences in gain between the devices can be seen from the top panel in Figure 2.1 and were smaller than 4.5 dB up to 2 kHz and smaller than 12 dB between 2 and 6 kHz.
Figure 2.1: Narrowband analyses of the hearing aid output for an input of pink noise at 70 dB SPL. The top panel shows the spectra of the raw recordings for the five hearing aids, the bottom panel shows the spectra for the recordings that were filtered with the inverse filter and bandwidth limited.

2.2.4 Stimulus set 2: Bandwidth limited recordings

During the fitting we selected a linear setting (no dynamic range compression) with the devices’ fitting software for input sound levels between 50 and 95 dB SPL and we verified the linearity of the gain by electro-acoustical measurements. For input levels below 55 dB SPL, the response of HA3 turned out to be compressed above 6 kHz. To remove this nonlinearity, we limited the frequency range of all devices to 5.8 kHz. Additionally, we used a high-pass filter to remove frequencies lower than 100 Hz to limit the frequency response to those frequencies that are clinically relevant (100 Hz to 5.8 kHz). The band limitation filters were designed with Matlab (function “ellip” and were elliptical low-pass and high-pass filters of the 7th order with a pass-band ripple of 0.1 dB, a stop band attenuation of >50 dB, and low and high frequency knee points at 100 Hz and 5800 Hz, respectively.

2.2.5 Stimulus set 3: Fully filtered recordings

Inverse filters were designed to remove the remaining irregularities (after careful manual adjustment and bandwidth limiting) in the frequency response. For each hearing aid, one filter was calculated. The goal of the filter was to remove perceptually disturbing effects (sound coloration), and not to compensate for hearing aid processing delay and the phase response. Therefore, the required transfer function was determined with linear system identification (Bendat and Piersol 2010). Since our recordings are intended to be used for speech-in-noise measurements, it sufficed to estimate the transfer function by simply dividing the output spectrum by the input spectrum. The
frequency response was measured with pink noise, because this resembles the speech spectrum as a first-order approximation. The required filter response of the inverse filter was obtained by comparing the hearing aid output to that of a measurement microphone (B&K 2260) at the location of the hearing aid microphone. The coefficients of the inverse filter were calculated with the Matlab function “fir2”. The constructed filter had 500 taps and was designed for non-causal application (Smith 1997) to correct for group delay and phase distortion introduced by the filter. The maximally required correction (difference between highest unwanted peak and lowest unwanted valley) was 22 dB and the maximal slope was 50 dB/octave and occurred around 4 kHz. These requirements were met by the digital filter. The resulting time-domain impulse response was windowed with a Hamming window. Other windows (e.g. a Bartlett window) might be more suitable if an accurate low-frequency response is important, but this was not necessary now since our signals were limited to frequencies above 100 Hz. Figure 2.1 shows the response to pink noise for each hearing aid prior and post filtering (excluding the band limitation). As expected, the inverse filter reduced the differences in frequency responses between hearing aids. To remove differences in bandwidth, all stimuli were bandwidth limited with the same filters as used on the previous set of stimuli.

2.2.6 Evaluation methods

To assess the homogenization of the recordings in the three stimulus sets, an objective quality metric was used and two listening tests were done.

**Objective evaluation**

We calculated the objective hearing aid speech quality index (HASQI, Kates and Arehart 2010) for all stimuli. This index estimates the quality of a target signal by comparing it to a reference signal. HASQI provides two outcome indices, one for linear effects and one for non-linear effects. The calculation for linear effects considers the change in the long-term spectral shape caused by the processing, while ignoring any changes to the signal envelope modulation. The calculation of non-linear effects, by contrast, considers the change in signal envelope modulations caused by the processing, while ignoring any long-term spectral changes. This non-linear measure is sensitive to the effects of noise, distortion, and nonlinear signal processing, and is expected to be rather insensitive to our non-causal inverse filtering. The reference signal was the original unfiltered digital input signal (i.e. the original speech-in-noise wave file that was not processed by the hearing aids). The reason for using speech in noise as reference is that we want to detect any differences caused by the filter, irrespective of whether the differences occur in the speech or in the noise. An additional calculation using the clean speech signal as reference gave the same linear HASQI scores and lower non-linear
HASQI scores (with an average of 0.19) due to the fact that now the noise is not part of the reference but considered a distortion. An important observation for the validity of our approach with speech in noise was that the ranking of the hearing aids was the same for clean and noisy speech as reference signal. The target signals consisted of the three sets of stimuli. The calculation was done on the same three sentences that were used in the subjective measurements (see next section). Calculation with 50 sentences gave near identical results and will therefore not be shown.

**Listening test: Detection**

To investigate whether listeners can distinguish between the hearing aid recordings, we conducted a listening experiment with six normal-hearing (ANSI, 2004) subjects. Although different from the target group, we chose normal-hearing listeners because they are assumingly better at detecting differences between stimuli than hearing-impaired listeners. Listeners with a sensorineural hearing deficit may be expected to have not only poorer hearing sensitivity, but also poorer suprathreshold processing like frequency resolution (Moore 1995), and modulation detection (Grant et al. 1998). If differences cannot be detected by normal-hearing subjects, we can be quite confident that these differences will also be unnoticeable for hearing-impaired subjects. Subjects were presented with three stimuli of which two were from the same hearing aid (standard) and one was from another aid (target). The subjects’ task was to select the hearing aid recording that differed from the other two (i.e. an odd-ball paradigm). To limit the duration of the experiment, only set 2 (bandwidth limited) and set 3 (fully filtered) were included and set 1 (the raw recordings, based on a manually optimized insertion gain) was omitted. The stimulus duration was on average 2.7 s (i.e. one sentence of 1.7 s with a 0.5 s lead-in and a 0.5 s lead-out). The stimuli were presented diotically with Sennheiser HDA200 headphones at 70 dB SPL. All combinations of hearing aids and filter conditions were presented at random in one session. Standard and target were always from the same stimulus set (i.e. bandwidth limited or fully filtered). Recordings from each hearing aid were used as target with standards of the recordings of all other hearing aids and vice versa. In total 20 distinct stimulus pairs were included (5 x 4, including AAB and BBA) and each stimulus pair was tested 3 times, leading to 60 trials per filter condition and thus 120 trials per subject. Three different sentences were used for the 3 repeats. Directly after the subjects had given their response, they received feedback on whether they had chosen the correct stimulus and if not, which one they should have chosen.

**Listening test: Preference judgement**

To determine if the inverse filtering influenced the sound quality of the signals, we also did a paired-comparison test in which the same subjects were asked to choose the
sound sample they preferred. The subject’s task was to make a choice based on the question: “Imagine that you will have to listen to these signals all day. Which sound would you prefer for prolonged listening?”. The choice was between the fully filtered stimulus (set 3) and its counterpart from the same hearing aid that was only bandwidth limited (set 2). The stimuli were identical to those from the previous experiment (3 comparisons per hearing aid and $5 \times 3 = 15$ comparisons per subject).

### 2.3 Results

#### 2.3.1 Objective evaluation

The results of the calculations with the HASQI model are shown in Figure 2.2. The mean linear index of the unfiltered signal (set 1) of the five hearing aids was 0.865 (with a range of 0.853 to 0.872). For the bandwidth limited signals (set 2) it was 0.863 (with a range of 0.849 to 0.871), and for the fully filtered signals (set 3) it was 0.945 (with a range of 0.941 to 0.947). Bandwidth limiting did not reduce the maximum difference between two hearing aids signals (0.02, for both sets 1 and 2), but applying the full filter reduced the maximum difference to 0.006. For the non-linear index the average indices were 0.752 (with a range of 0.697 to 0.798) for the unprocessed, 0.759 (with a range of 0.685 to 0.793) for the bandwidth limited signals and 0.790 (with a range of 0.731 to 0.814) for the fully filtered signals. Thus, bandwidth limitation increased the maximum difference in the nonlinear index between two hearing aid stimuli from 0.10 (set 1) to 0.11 (set 2) and additional application of the inverse filter reduced the maximum difference to 0.08 (set 3).

![Figure 2.2: Results of the HASQI objective quality model for the three stimulus sets.](image-url)
### 2.3.2 Listening tests

**Detection task**

Figure 2.3 shows the percentages of correct detection averaged over all subjects. The average detection score for the bandwidth limited signals was 87% and for the fully filtered signals it was 39%. A two-way analysis of variance with subject (6 levels) as random effect and hearing aid (5 levels) and stimulus set (2 levels) as fixed effects indicated that the main effect of stimulus set (fully filtered versus bandwidth limited) was highly significant (F[1,20] = 90, p < 0.0005). The interaction between subject and filter type was significant as well (F[5,20] = 6, p < 0.005). The other main and interaction effects were statistically insignificant (p > 0.1). To determine if the detection rate of any of the hearing aid signals was higher than chance (33%), one-sided t-tests were used with Bonferroni correction. For the bandwidth limited set all results were significant (p = 0.001). For the fully filtered stimuli, none of the results were significant (p > 0.13). A one-sided t-test on the pooled data of this set showed that the detection of the group of hearing aids was slightly higher than chance: 39% with p < 0.002 (for this no Bonferroni correction was required).

![Figure 2.3](image.png)

**Figure 2.3:** Percentage of times the subjects selected the correct stimulus as deviant from the other two. Signals were only compared to others of the group they belonged to, i.e. bandwidth filtered only (open circles), or fully filtered (bandwidth limited and inversely filtered, filled circles). Chance level was 33% and error bars denote 95% confidence intervals.

**Preference judgement**

Five of the six subjects preferred the fully filtered signals over the bandwidth limited signals in all (100%) of the sound samples, the sixth subject preferred the fully filtered signals in 73% of the sound samples.
2.4 Discussion

The results indicate that in order to reduce the perceptual differences between hearing aid recordings

Q1a. it was not sufficient to carefully adjust the insertion gain of the hearing aids;

Q1b. it was not sufficient to limit the bandwidth of the recordings to that of the smallest device;

Q1c. it was sufficient to apply a hearing aid specific inverse filter on the bandwidth-limited recordings.

2.4.1 Objective evaluation

For both set 1 and 2, the difference in score between the hearing aids was larger than for set 3 (0.02 compared to 0.006). This indicates that both manual adjustment of insertion gain (set 1) and bandwidth limitation (set 2) were not sufficient to make the hearing aid recordings indistinguishable from each other, and additional application of the inverse filter (set 3) was required. Moreover, the linear HASQI score was improved by the inverse filtering, which suggests that the filter actually may improve sound quality.

The range of scores for the non-linear HASQI metric was similar for all three sets. As expected, the bandwidth limitation and the inverse filters did not greatly influence the non-linear HASQI score. Therefore these results indicate that the inverse filters did not add non-linear distortions (at least for those aspects for which HASQI non-linear is sensitive). HA3 and HA4 had lower scores than the other aids, but this does not necessarily mean that the sound quality of these hearing aids is lower. The lower scores for HA3 and HA4 indicate that these aids were perhaps not operating completely linear, although all non-linear processing was switched off. Indeed, HA3, was shown to be compressive above 6 kHz (see Methods) and the non-linear index increased after band-pass limiting. The reason for this is that the bandwidth limiting removed those frequencies that fell outside the linear range of the hearing aid: HA3 was the aid that limited the bandwidth in the high frequencies. The reason for the lower score for HA4 is unknown and falls beyond the scope of this paper.

2.4.2 Listening tests

The fact that the detection of the “oddball” was much poorer for the fully-filtered signals than for the signals that were bandwidth limited only, indicates that the inverse filtering increased the similarity between the hearing aid signals. The detection for the inverse filter for each of the five hearing aids did not deviate significantly from chance.
The result for the pooled dataset was slightly, but significantly, above chance (detection was 39%). The larger number of comparisons, coupled with the fact that a Bonferroni correction was not necessary here, gave larger statistical power. However, the influence of this detection rate on perceptual comparisons is expected to be only small since one will be primarily interested in differences between single pairs and thus have access to only a smaller number of comparisons than was used for the pooled data set. The higher than chance detection-rate was probably caused by small residual differences in frequency response between hearing aids. These small differences are unlikely to lead to differences in preference judgements.

There was a significant interaction between subject and filter type: the difference in detection rate between the fully filtered and the bandwidth limited signals depended on the subject. The reason for this is that some subjects performed worse at the detection of the bandwidth limited signals, while the detection of the fully filtered signals was around chance for all subjects. The interaction thus reflects that subjects differ in the discrimination of the bandwidth limited signals and not in the discrimination of the fully filtered signals. This interaction will therefore not be relevant for use of the inverse filter.

The second listening experiment showed that all subjects preferred the fully filtered signals over the bandwidth limited signals. This supports the results from the objective quality model and indicates that the filtering did not degrade the sound quality and in fact improved it for all hearing aids. This leads to two conclusions. First, the fact that the filter did not lower the quality shows that the filter did not add distortions while reducing the differences between hearing aids. Second, it shows that the quality of the recordings could be easily improved by flattening the frequency response. This agrees with results from previous research that a smoother frequency response leads to better sound quality judgments (Arehart et al. 2010). It supports the implication of this study that quality judgment tests across hearing aids should not be based on raw recordings because this can mask the effect of the processing under investigation, but that additional filtering is required.

2.4.3 Application of the inverse filter

An inverse filter has been shown to be able to compensate for the response of the hearing aids included in this study. This compensation also works after an additional signal processing feature is turned on. The filter does not influence the noise-reduction algorithm itself because the filtering acts at the output of the hearing aid and only corrects for the characteristics that remain equal with or without the noise reduction. However, filters cannot transparently correct for compression. In case a noise reduction is imple-
Method to compare noise reduction in hearing aids

mented such that it depends on a compression stage, one would need to investigate compression and noise reduction in interaction. An inverse filter is then still required to remove differences in frequency response between hearing aids. The intended use of this research is to facilitate research into hearing aids. Application of the inverse filter in a clinical setting, (e.g. to allow clients to directly compare the effect of noise reduction between different devices) is cumbersome since the technique requires a specific filter for each device.

The normal-hearing subjects preferred the recordings with a flattened frequency response. Perhaps this result carries over to listeners with hearing loss, especially for subjects with mild conductive loss. If this would also hold for other hearing losses one might contemplate to add a simplified version of the inverse filter to a hearing aid.

Instead of focussing on group results, recently the individualisation of noise reduction in hearing aids has gained attention. The few available studies (Zakis et al. 2009, Houben et al. 2012) are inconclusive. The current approach might stimulate research that focuses on individuals rather than on the group they belong to.

2.5 Conclusion

We conclude that the perceptual differences between recordings of different linearly fitted hearing aids can be removed by application of an inverse filter in combination with a band-pass filter. Application of such a filter might even improve the sound quality of the recordings. However, the main objective is to remove large differences in frequency response between hearing aids thereby facilitating the comparison of more subtle differences between hearing aids due to nonlinear processing. Once an inverse filter is designed for a specific hearing aid, it can also be applied on recordings with (nonlinear) processing, such as noise reduction, turned on.