Perceptual evaluation of noise reduction in hearing aids

Brons, I.

Publication date
2013

Document Version
Final published version

Link to publication

Citation for published version (APA):
Brons, I. (2013). *Perceptual evaluation of noise reduction in hearing aids*. [Thesis, fully internal, Universiteit van Amsterdam].

Perceptual evaluation of noise reduction in hearing aids

Inge Brons

Perceptual evaluation of noise reduction in hearing aids
Inge Brons

Perceptual evaluation of noise reduction in hearing aids

The work presented in this thesis was supported

by grants of the Heinsius-Houbolt Fund.

ISBN: 978-94-6191-969-4

Printed by Ipskamp Drukkers, Enschede

Perceptual evaluation of noise reduction in hearing aids

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 17 december 2013, te 14:00 uur

door

Ingeborg Brons

geboren te Vlissingen

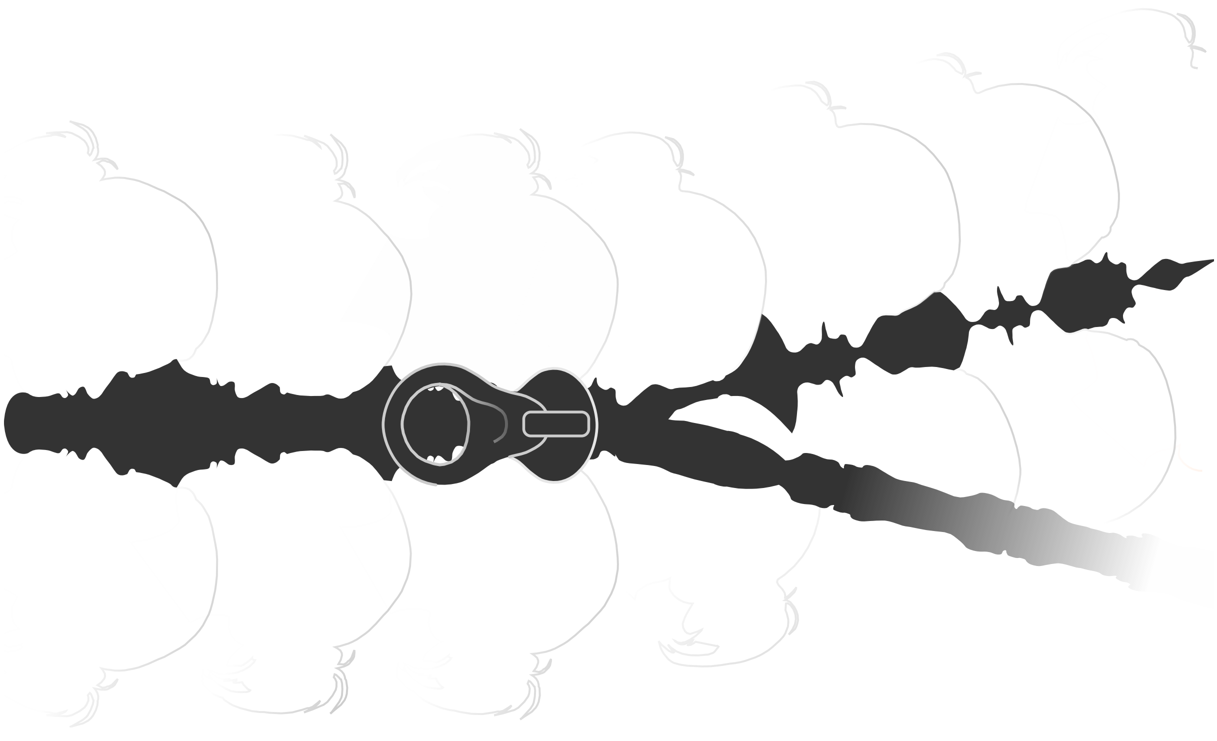
Promotiecommissie

Promotor:	Prof.dr.ir. W.A. Dreschler
Copromotor:	Dr.ir. A.C.H. Houben
Overige leden:	Prof.dr. D. Başkent
	Prof.dr. W.J. Fokkens
	Prof.dr.ir. C.A. Grimbergen
	Dr. N.J. Versfeld
	Prof.dr.ir. B. de Vries

Faculteit der Geneeskunde

Contents

1 Introduction



1.1 Introduction

During the last decades, digital hearing aids have replaced their analogue predecessors. Whereas the size of the hearing aids decreased with time, the capabilities for signal processing increased. A wide variety of signal-processing strategies has been developed for hearing aids in order to compensate for different aspects of hearing loss. Straight-forward amplification can largely compensate for deficits in the outer and middle ear (conductive hearing loss), but compensation for deficits in the cochlea (sensorineural hearing loss) requires a more sophisticated approach. Sensorineural hearing loss causes not only reduced sensitivity for soft sounds, but also a reduced dynamic range as well as a reduced spectral and temporal resolution (Moore 1996). The resulting distortions in the perceived sound make it more difficult to understand speech, especially in noisy environments. For a given level of background noise, a listener with sensorineural hearing loss needs a higher speech level than a normal-hearing listener to obtain the same performance, even if the sound is amplified (Plomp 1986). In noisy situations the hearing aid should therefore amplify speech more than background noise to improve the signal-to-noise ratio (SNR).

Improving the signal-to-noise ratio in noisy environments is one of the most difficult challenges for digital signal processing in hearing aids. For that purpose, the hearing aid should estimate from the incoming signal whether the user is in an environment with speech, noise, or both speech and noise and adjust the gain in each frequency channel accordingly. To distinguish between speech, noise, or both, the hearing aid uses signal properties that are generally different between speech and noise. Hearing aids with multiple microphone inputs can use spatial differences between speech and noise. These hearing aids are able to amplify sounds that enter the hearing aid from one direction (usually from the front) while attenuating incoming sounds from other directions. In situations where speech and noise signals enter the hearing aid from separate directions, these directional microphones effectively improve the SNR (Bentler 2005). However, speech and noise sources are not always spatially separated and even if they are, both signals can be mixed before they enter the hearing aid, for instance due to reverberations. In that case the hearing aid can no longer use spatial differences to separate between speech and noise and should use other cues to recognize speech and noise, which is more difficult because speech and noise enter the hearing aid as a mixed signal. For this purpose, hearing aids have single-microphone noise-reduction algorithms which use temporal and spectral properties to separate between speech and noise. Such algorithms are implemented both in hearing aids with one microphone input as well as in hearing aids with multiple microphone inputs where it is supplementary to directionality.

This thesis focuses on single-microphone noise reduction. The use of spatial differences between speech and noise is thus beyond the scope of this work.

Single-microphone noise-reduction algorithms generally consist of two parts. First, they classify the sound environment into speech, noise, or both. This step generally results in an estimate of the actual input SNR per frequency channel. Second, the algorithm should reduce the hearing aid gain at the right time and for the right frequencies to reduce noise, but leave the speech intact. The next sections explain the basic approaches and main properties of both parts of a noise-reduction algorithm.

1.2 Classification of the environment

The upper panels of Figure 1.1 show typical time-amplitude waveforms of speech, stationary noise and a mixture of speech and stationary noise. The figure shows that speech has a very characteristic temporal pattern, with high variation in amplitude over time corresponding to the opening and closing of the vocal tract. Noise-reduction algorithms make use of this characteristic pattern of speech for the classification of the environment. The algorithms often use the envelope of the signal for this purpose, which is plotted for the same signals in the lower panels of Figure 1.1. The envelope of speech shows large level fluctuations (lower left panel), whereas the envelope of stationary noise shows very little fluctuations (lower middle panel). Adding noise to the speech reduces the depth of the envelope fluctuations (lower right panel) compared to those of speech.

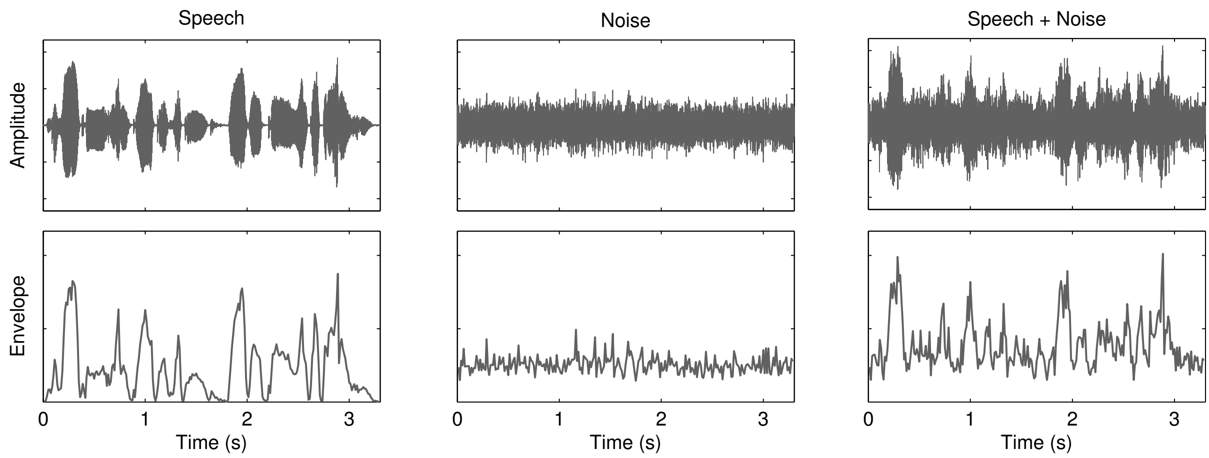


Figure 1.1: Upper part: time-amplitude waveforms of speech, stationary noise and the mixture of these speech and noise signals. Lower part: the envelope of the same signals.

Most real-life noises have envelope characteristics that lie in between that of speech and stationary noise, with more fluctuations than stationary noise, but with lower amplitude and at other modulation frequencies than for speech (see next paragraphs).

Classification based on modulation depth

One property of the envelope signal that a noise-reduction algorithm can use for identification of speech and noise is the modulation depth (Schaub 2008). The modulation depth is the difference between the peaks and valleys of the signal envelope. As visible in Figure 1.1, modulation depth is higher for speech than for noise. Adding noise to the speech reduces the modulation depth (lower right panel in Figure 1.1) compared to that of speech alone. Therefore, the modulation depth can be used to estimate the SNR at the input of the hearing aid (input SNR). Frequency channels with high modulation depth are likely to be dominated by speech and thus to have high input SNR, whereas frequency channels with low modulation depth are dominated by noise and have low input SNR.

Classification based on modulation spectrum

Another property of the envelope signal that can be used to identify speech and noise is the modulation spectrum (Dillon 2001; Chung 2004). The modulation spectrum provides information on how fast the signal envelope changes in level. Figure 1.2 shows the modulation spectrum for the three envelope signals in the lower part of Figure 1.1. The envelope of speech is typically dominated by modulations with frequencies below 10 Hz. In general, modulation frequencies between 3 and 6 Hz are dominant, corresponding to the rate at which syllables are produced. The envelope of stationary noise shows the same amount of modulations at all frequencies. Thus, if the modulation spectrum for a specific frequency channel is dominated by frequencies between 3 and 6 Hz it is likely that speech is present.

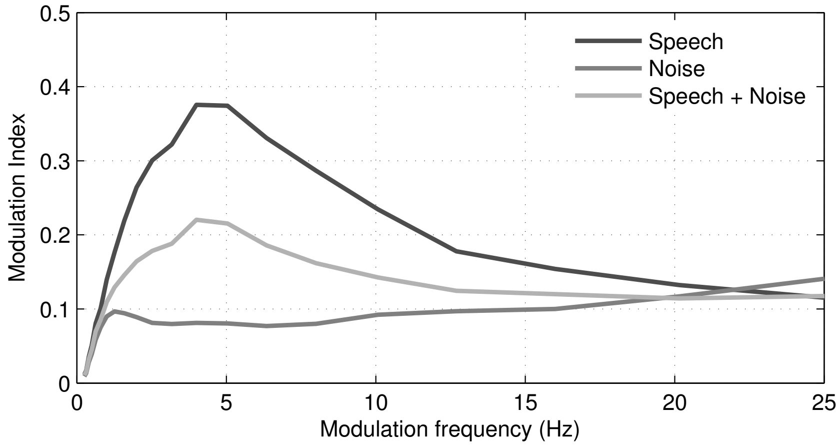


Figure 1.2: Modulation spectra of the envelope of speech, stationary noise and the mixture of speech and noise. A higher modulation index means a higher modulation depth at that frequency.

Classification based on synchrony detection

An additional clue for speech presence is the synchrony between signal envelopes of different frequency bands (Elberling 2002; Dillon 2001). For instance, modulations at the pitch of the voice (100-400 Hz) are synchronized across frequency channels. This is illustrated in Figure 1.3, which shows the waveform of speech during a vowel in four different frequency channels. The speech envelope in each frequency channel has the same periodicity. Thus, if such synchronization across frequencies is detected, speech is likely to be present.

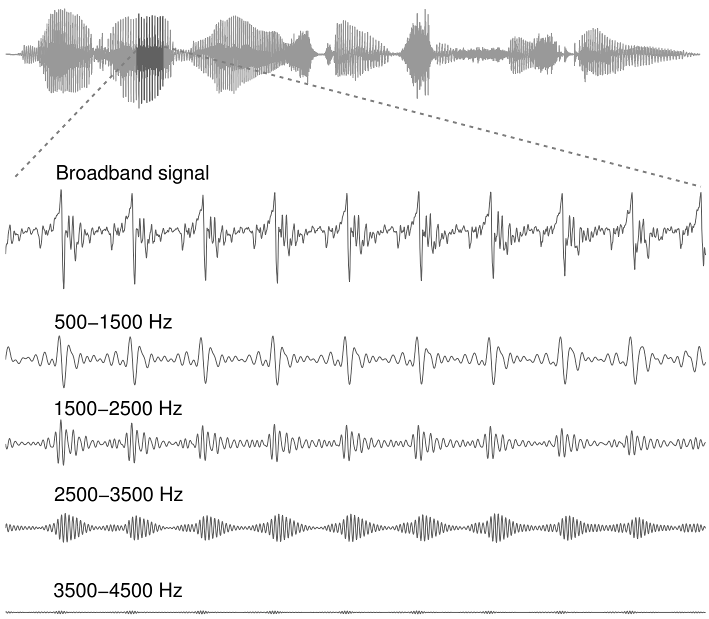


Figure 1.3: Synchronization across frequency channels during speech. The upper signal shows in light gray the time-amplitude signal of one sentence, with in dark gray the fragment (part of a vowel) that is shown in the lower five signals. The 'Broadband signal' shows the fragment in its original form; the lower four signals show the fragment in four different frequency channels (each 1000 Hz wide).

Classification based on speech pause detection

Another way to make use of the dynamical characteristics of speech is to use the pauses in speech to estimate the noise signal and to update this estimate over time (Loizou 2007). The most straight-forward way to do this is the use of an absolute level threshold. Signal parts where the level does not exceed this threshold are assumed to contain no speech and can therefore be considered as an update of the noise estimate, which in turn can be used to estimate the SNR. Tracking the minimum level of the incoming signal as an estimate for the noise forms the basic principle for many more sophisticated noise-estimation algorithms (Loizou 2007).

1.2.1 Reduction of the hearing-aid gain

If the properties of the incoming signal indicate the presence of noise, the noise-reduction algorithm should decide if and how it will adjust the hearing-aid gain in order to reduce the noise. The main challenge for the noise reduction is to reduce background noise while retaining the level and quality of target speech. To what extent the noise reduction succeeds in this task depends not only on the accuracy of the classification of speech and noise but also on how this information is translated into changes in hearing aid gain. The most important variables for the gain reduction will be discussed below.

Amount of gain reduction

The amount of gain reduction in a frequency channel is generally chosen to be proportional to the estimated SNR for that channel (Dillon 2001; Chung 2004). If the signal is clearly dominated by speech, thus at high input SNR, the gain should not be reduced in that channel, to preserve speech information. If the input SNR of a channel is very low, i.e. when noise is present but no speech is recognized in that channel, the noise-reduction algorithm will reduce gain in that channel maximally. In most hearing aids the maximum amount of gain reduction is limited, so that below a certain SNR the gain is not further reduced if the SNR decreases further (Chung 2004; Bentler and Chiou 2006). If the maximum amount of gain reduction is low, the residual background noise level will be relatively high. If the maximum amount of gain reduction is high, the residual noise will be less but that may be at the expense of speech level and sound quality. Reduction of speech level and quality can for instance occur when errors are made in the classification of the environment (speech-dominated signal parts are wrongly considered as noise-dominated and are therefore reduced in gain) (Loizou and Kim 2011). Sound quality can also be affected when the gain differs largely between neighbouring time windows or frequency channels (see next sections).

Time constants for gain reduction

The effectiveness of noise reduction is also determined by how fast the noise reduction reacts to changes in the environment (Bentler and Chiou 2006). Noise-reduction algorithms generally have several time constants to react properly to different changes in the input signal, so that for instance the decrease of gain as a reaction to noise can be slower than the reaction to a sudden appearance of speech. The time that an algorithm needs to reach its maximum gain reduction after noise has started varies between algorithms from several seconds to more than 30 seconds. In contrast, most algorithms react within several milliseconds to the appearance of speech by increasing the gain, in order not to loose any speech information (Chung 2004). Algorithms that change the gain very quickly can be able to reduce the noise even in the short pauses that occur during speech. However, quick transitions in gain may be experienced as restless and can cause distortions to the speech signal, which may affect speech intelligibility. On the other hand, algorithms that adjust the gain only slowly will result in more residual noise during speech and after changes in the environment.

Number of frequency channels for gain reduction

The potential benefit of noise reduction in improving the SNR is mainly due to the possibility to apply noise reduction separately for different frequency channels (Chung 2004). Because noise reduction only adjusts the overall gain for a frequency channel, the SNR within a frequency channel will not change during one time frame. Only the combination of reduced gain in some frequency channels (the noise-dominated ones) and preserved gain in others (the speech-dominated ones) makes that the instantaneous SNR of the overall signal (all frequency channels together) can improve due to noise reduction. This argues in favour of a high number of frequency channels. Within hearing aids, however, the number of channels is limited because increasing the number of channels also increases processing delay. Many state-of-the-art noise-reduction algorithms that are not (yet) implemented in hearing aids apply noise reduction in high numbers of frequency bins separately (Loizou 2007). However, applying different amounts of gain to neighbouring frequency bins introduces unwanted distortions, often referred to as musical noise (Berouti et al. 1979). It is therefore generally more favourable for the sound quality to limit the number of frequency channels, although this may not result in the most optimal improvement of SNR.

Frequency weighting of the gain reduction

Dividing the signal in different frequency channels allows the algorithm to assign different weights to separate frequency channels (Kuk and Paludan-Müller 2006). One can for instance decide to allow more gain reduction for the low frequencies than for the higher frequencies. Many environmental noises, for instance traffic noise, contain energy mainly in the low frequencies. For such situations listening comfort can easily be improved by reducing the gain more for lower frequencies than for higher frequencies ("high-pass filter"). This method also removes the low-frequency content of the speech signal, but if this part of the speech would otherwise be inaudible due to the noise, the filter will not reduce intelligibility. Reduction of the low-frequency noise also prevents the noise from masking high-frequency information in the speech ("upward spread of masking") (Levitt 2001).

Several algorithms base their maximum gain reduction per frequency band on the articulation index, which describes how much the audibility of speech in each frequency band contributes to the intelligibility of the speech (Kryter 1962). Frequency bands which are known to be important for speech intelligibility will receive less gain reduction than frequency bands with lower perceptual importance. Additionally, in order to maintain audibility of the input signal, listeners' hearing loss at different frequencies, as well as the level of the input signal may play a role in determining the noise-reduction strength for separate frequency bands (Kuk and Paludan-Müller 2006).

1.3 Implementation in hearing aids

It may be clear from the description of the variables that there is not one optimal value for each of them. In fact, all variables involve the trade-off between reducing noise and retaining the quality of speech. Generally, most variables are preset by the manufacturer and cannot be changed by the clinician who adjusts the hearing aid settings to its user. Some hearing aids provide the possibility to adjuste the maximum allowed amount of gain reduction. For the other variables, the manufacturers have the difficult task to make a well-considered choice for their settings. Because each setting has its advantages and disadvantages, and because they may interact, these choices may differ between manufacturers. Unfortunately, noise reduction in hearing aids is generally presented as a "black box" so that no information is available on the actual implementation and its underlying rationales. In order to nonetheless obtain some insight in noise-reduction implementations in hearing aids, we have made recordings of the output of some hearing aids and analyzed how the hearing aid gain differed for the same input signals if noise reduction was activated or de-activated.

Figure 1.4 shows for four different hearing aids (HA1 to HA4) how noise reduction influenced the hearing aid gain for the same input signal. As a reference condition, one "conceptual" noise reduction is added to Figure 1.4. This noise reduction received speech and noise separately and thus had knowledge on the real input SNR (which is often referred to as "ideal" noise reduction). This algorithm can obviously not be used in hearing aids, but for research purposes it provides a useful tool to investigate effects of noise-reduction settings. In Figure 1.4, this conceptual condition is added for comparison because it provides information on the real input SNR of the different time-frequency units.

The time signal at the top of the Figure 1.4 shows the input signals, consisting of stationary noise which is combined with speech after 2 seconds. Note that speech and noise are plotted separately here for clarity, but only the conceptual algorithm had the separate signals as input. The hearing aid noise-reduction systems only had the sum of both signals as input and had to estimate whether this was speech, noise, or both.

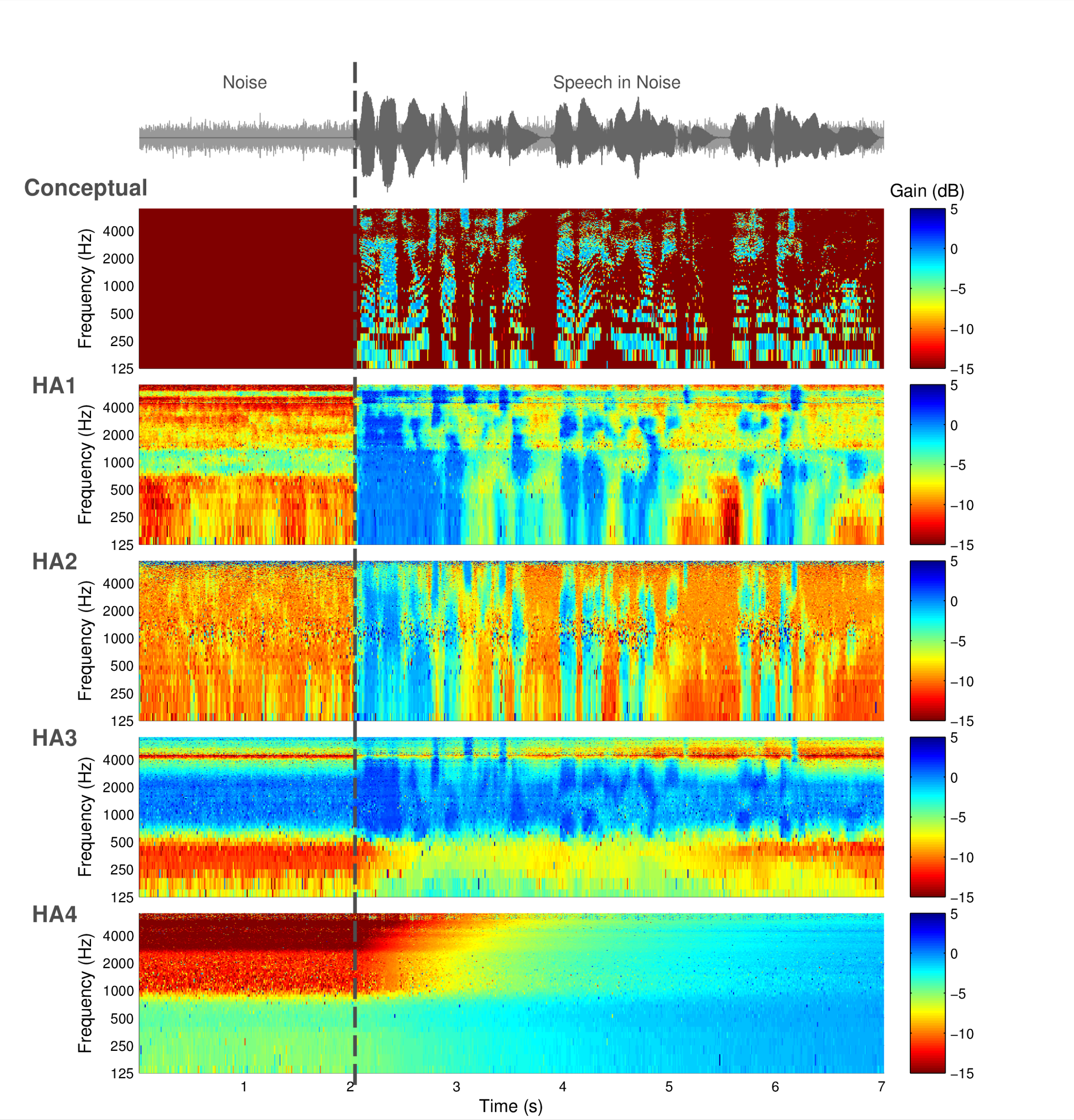


Figure 1.4: The effect of noise reduction on the hearing aid gain in four different hearing aids (HA1 to HA4). One conceptual algorithm (which received speech and noise separately and thus had knowledge on the real input SNR) was added for comparison. The time signals plotted at the top show the input noise (light gray) and speech (dark gray) signals, which were mixed before they entered the hearing aids.

The colours in the lower five panels show how the noise reduction influenced the gain as a function of time (horizontal axis) and frequency (vertical axis). Blue areas indicate no difference in gain between noise reduction *on* and *off*, thus no activity of the noise reduction. Red areas represent strong reduction in gain due to noise reduction. Thus, as would be expected, noise reduction was generally most active if the input signal consisted of noise only (the first two seconds) and less when the input signal also contained speech (starting after 2 seconds). Next sections will discuss what information can be deduced from Figure 1.4 on the variables explained before.

Amount of gain reduction

The maximum amount of gain reduction for the conceptual algorithm was set at 15 dB. All time-frequency bins where the SNR was below 0 dB, thus where noise was dominating, received maximum gain reduction. For positive SNRs, the amount of gain reduction decreased as a function of SNR, so that gain was not reduced for time-frequency units with high input SNR. The colour plot for the conceptual algorithm thus provides direct information on the real input SNR so that we can compare how well the other algorithms separated between speech and noise.

Noise reduction in the four hearing aids was set at the strongest available settings. During the first two seconds, where only noise was present and thus the algorithms should reduce gain maximally this resulted in maximum gain reduction of about 10 to 13 dB (red areas) for hearing aids HA1 to HA3, and more than 15 dB (dark red areas) for HA4 in the high frequencies.

Time constants for gain reduction

At the start of Figure 1.4, noise was already present for a longer time, so that the figure does not show how quick the algorithms adjusted the gain if the situation turned from silence to noise. However, we see that as soon as speech started (vertical line at t = 2 s), all algorithms reacted quickly with reducing noise reduction. Such a quick reaction on speech presence is important in order to retain speech information.

While speech was present, the algorithms in hearing aid HA1 and HA2 adjusted the hearing aid gain from time to time in order to reduce noise during the short pauses in the speech. Comparison with the gain-reduction pattern of the conceptual noise reduction shows that speech was most of the time recognized well by HA1 and HA2, but clearly not all speech-dominated parts were preserved. The reaction of HA3 differed between the low- and high-frequency regions. For frequencies up to 500 Hz, the gain changed only slowly. Gain in the higher frequencies reacted more quickly on changes in speech, but not by reducing gain during noise but by increasing gain somewhat during speech. HA4 adjusted its gain only slowly. After the speech started it took some seconds before gain was restored, but then gain was completely preserved as long as speech was present. So it seems that for this hearing aid preserving audibility of speech was considered more important than reducing the noise during speech.

Number of channels for gain reduction

The number of frequency channels within the four hearing aids can not be deduced from Figure 1.4. According to the technical specification of the hearing aids, the number of frequency channels used for noise-reduction processing varied from 8 to 20. The conceptual algorithm calculated gain for each frequency bin separately. As explained before, this had the advantage that speech and noise could be separated more accurately, but this resulted in high gain contrast between neighbouring time-frequency bins, which might be disadvantageous for the sound quality.

Frequency weighting of the gain reduction

Figure 1.4 reveals that the noise-reduction algorithms from the four different hearing aids have different strategies for frequency weighting. The conceptual algorithm had no rules for frequency weighting implemented. Thus if the input signal consisted of noise only, this algorithm applied the same amount of gain reduction (15 dB) over all frequencies. This was also the case for HA2. Noise reduction in HA1 was more cautious for frequencies around 1000 Hz, where it only reduced gain up to about 5 dB (green horizontal line) instead of 12 dB (red) for other frequencies. HA3 was even more careful for frequencies around 1000 and 2000 Hz. This is probably because these frequencies generally contain important speech information. In that light the behaviour of HA4 is rather unexpected: the gain during noise was preserved in the low frequencies but strongly reduced in the high-frequency regions. The rationale behind this weighting is unknown, but from recordings with other input noises we know that the frequency weighting for this hearing aid differs between noises (see for instance Chapter 3 for the gain-reduction pattern for babble noise).

Summarizing, Figure 1.4 reveals that noise reduction from different hearing aids have diverging effects on the hearing aid gain for the same speech in noise input signal. In that light, it is puzzling why there is so little public knowledge on noise-reduction implementations in hearing aids and on the motivation of manufacturers to choose for specific noise-reduction settings in their products.

1.4 Perceptual effects of noise reduction

Noise-reduction algorithms were at first developed with the aim of improving speech intelligibility in noisy environments (Loizou 2007). However, studies that investigated the effect of noise reduction on speech intelligibility revealed no improvement in intelligibility due to noise reduction (see for instance Nordrum et al. 2006). In fact, in some cases noise reduction even tended to reduce speech intelligibility (Hu and Loizou 2007a). The reason for this could be that, as mentioned before, noise reduction may affect the speech signal and speech level due to errors in the classification and due to processing artefacts.

Even if noise reduction is able to reduce the noise level without seriously affecting the speech level and quality, this does not necessarily result in an improvement in speech intelligibility in noise. In that case, the noise that is removed by the noise reduction was not the noise that masked the speech signal, so that otherwise the auditory and cognitive system of the listener would have been able to neglect these noise parts that were now reduced by the noise reduction. Although this does not result in an objective improvement in speech intelligibility, the listener can experience more listening comfort if parts of the noise are already removed by the noise reduction. It is imaginable that this lightens the cognitive load or listening effort needed to understand speech, which in turn can lead to less fatigue due to listening in noisy environments (Sarampalis et al. 2009). Unlike speech intelligibility, which is a common objective measure for hearing aid benefit, objective measurements for cognitive load or listening effort are scarce and still in an initial phase of development and validation.

Studies evaluating hearing aid noise reduction generally use subjective measures to gain insight in the effects of noise reduction on perception (Bentler 2005). For instance, subjects can be asked to rate the sound quality within a specific condition, or to chose from different conditions the most comfortable one. In the same way, listening effort can be subjectively measured. The results of previous studies on subjective effects of hearing aid noise reduction show however no uniform results. For instance, some studies showed that listeners preferred noise reduction over no noise reduction (Boymans and Dreschler 2000; Ricketts and Hornsby 2005), whereas others found no difference in listening comfort or sound quality due to noise reduction (Alcàntara et al. 2003; Bentler et al. 2008). Also results on listening effort were inconsistent (Alcàntara et al. 2003; Bentler et al. 2008).

The studies mentioned here compared noise reduction with no noise reduction within a hearing aid, but there is little or no knowledge on how noise reduction was implemented in the specific hearing aids. As we have just seen in Figure 1.4 for a selection of four hearing aids, noise-reduction implementations can differ widely between hearing aids. It is imaginable that these differences between hearing aids will also result in different results for perceptual outcomes like listening comfort and listening effort. In this thesis, the perceptual consequences of the different noise-reduction implementations in Figure 1.4 will be investigated more systematically.

1.5 Outline of this thesis

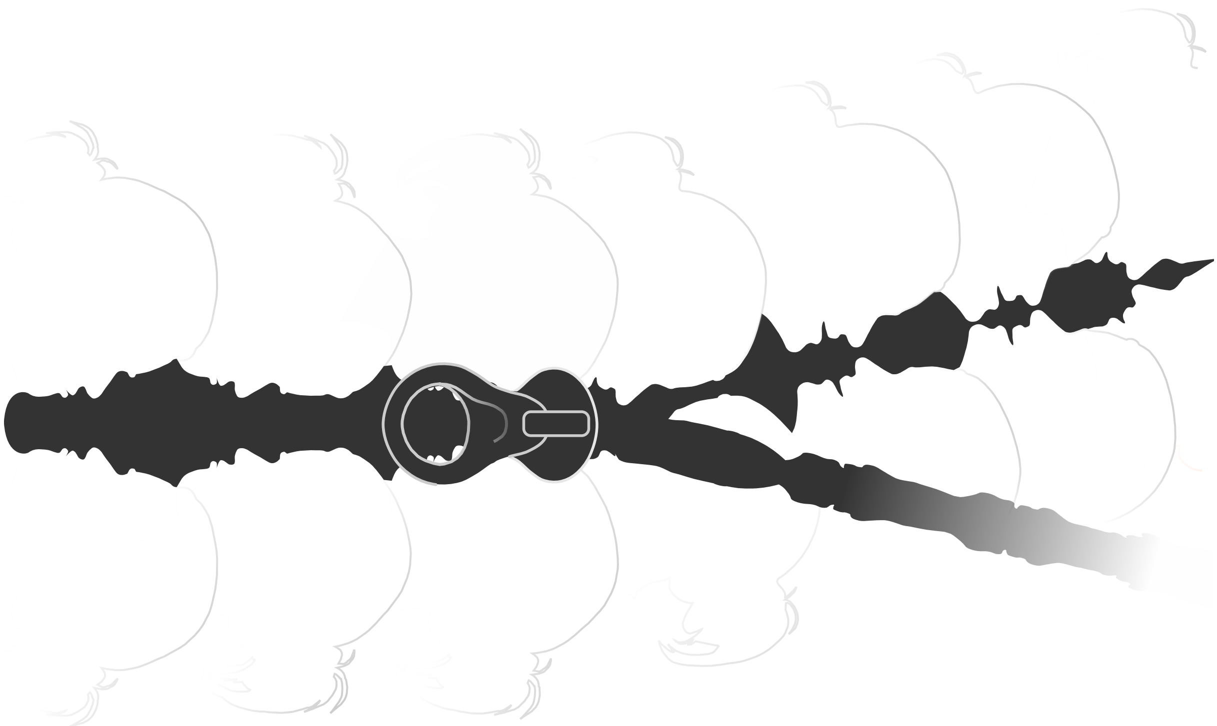
This thesis studies the effects of single-microphone noise-reduction algorithms on perceptual outcomes (noise annoyance, speech naturalness, personal preference, speech intelligibility and listening effort).

The first part of the thesis (Chapters 2 to 5) describes the evaluation of noise reduction as implemented in different commercial hearing aids. **Chapter 2** describes how we recorded the hearing aid output and removed differences in frequency response between hearing aids to allow direct comparison of noise-reduction algorithms between hearing aids. In **Chapter 3** we used this method to perform perceptual measurements with normal-hearing subjects to evaluate the effects of noise reduction in linear hearing aids in terms of noise annoyance, speech naturalness, personal preference, speech intelligibility, and listening effort. **Chapter 4** extends Chapter 3 in that it describes the results of comparable measurements for listeners with sensorineural hearing loss. Thus far, we evaluated noise reduction in a linear setting. However, for listeners with sensorineural hearing loss, hearing aids usually apply dynamic-range compression, which may interact with noise reduction. In **Chapter 5** we therefore explored the combination of noise reduction and compression in the same commercial hearing aids as in Chapter 4. This chapter describes both acoustical measurements and perceptual measurements with hearing-impaired subjects.

The second part of the thesis (Chapters 6 and 7) describes studies using noise-reduction algorithms that were not implemented in hearing aids. Because the source code of these algorithms was available we could adjust parameters to investigate effects of different settings. In **Chapter 6** we evaluated the effect of maximum attenuation strength on perceptual outcomes (noise annoyance, speech naturalness, personal preference, speech intelligibility, and listening effort) for ideal noise reduction (comparable with the “conceptual” condition in Figure 1.4 of the current chapter). Additionally, we evaluated the effect of two different (non-ideal) noise-estimation methods on the same perceptual outcomes to obtain insight in the effect of classification errors. In **Chapter 7** we determined how much attenuation could be applied by a noise-reduction algorithm until subjects detected distortions caused by the noise reduction. We investigated whether this detection threshold for noise-reduction distortion differed between normal-hearing and hearing-impaired subjects and whether it was related to their individual preference for noise-reduction strength.

Chapter 8 summarizes the results of this thesis and provides some recommendations for clinical practice and for future research.

2 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.

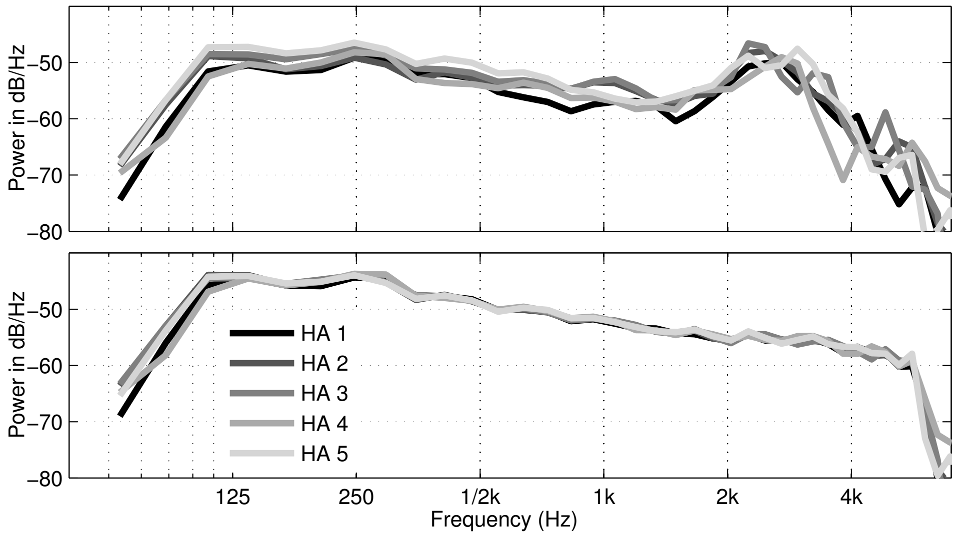


Figure 2.1: Narrowband analyses of the hearing aid output for an input of pink noise at 70 dB SPL. The top panel shows the spectra of the raw recordings for the five hearing aids, the bottom panel shows the spectra for the recordings that were filtered with the inverse filter and bandwidth limited.

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 x 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).

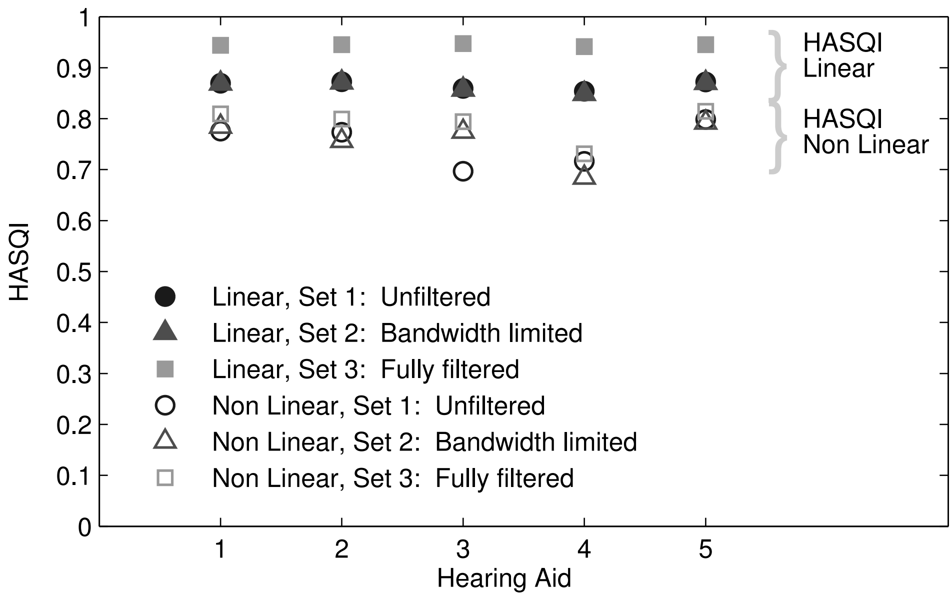


Figure 2.2: Results of the HASQI objective quality model for the three stimulus sets.

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).

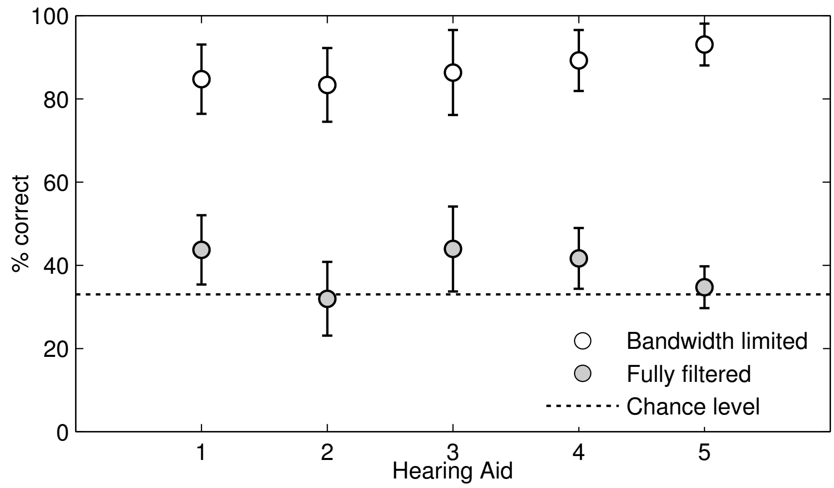


Figure 2.3: Percentage of times the subjects selected the correct stimulus as deviant from the other two. Signals were only compared to others of the group they belonged to, i.e. bandwidth filtered only (open circles), or fully filtered (bandwidth limited and inversely filtered, filled circles). Chance 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 undistinguishable 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 implemented 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.

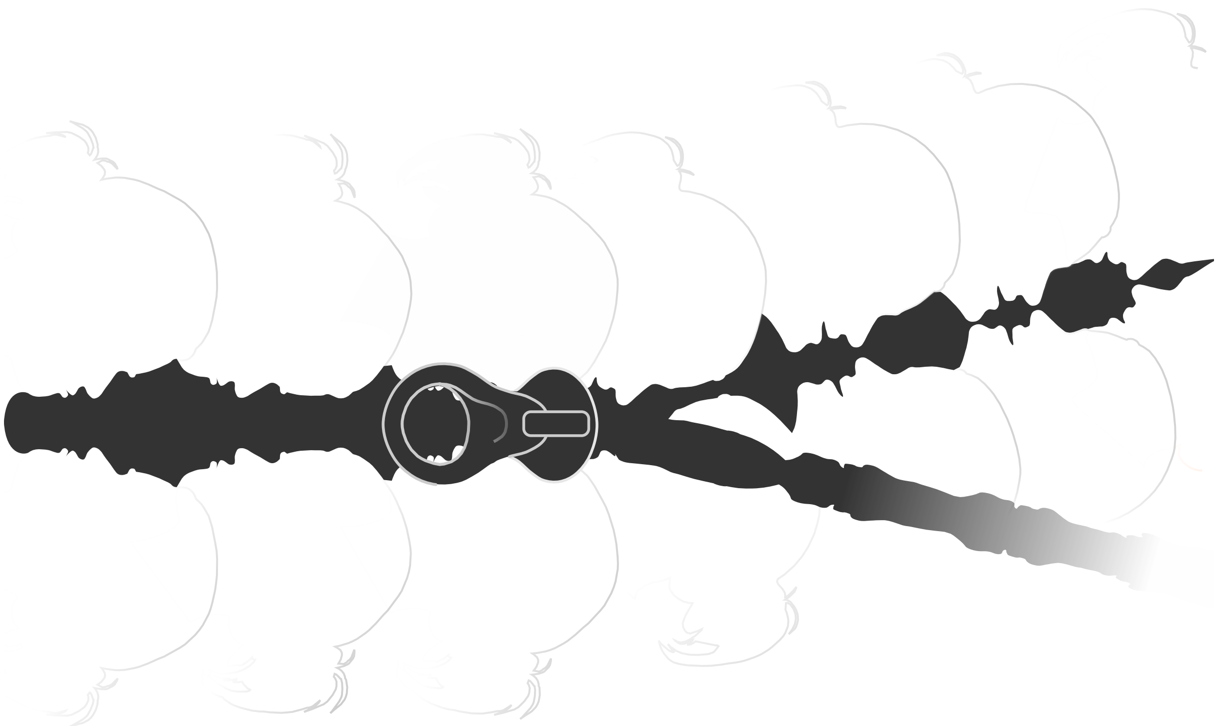
The normal-hearing subjects preferred the recordings with a flattened frequency response. Perhaps this result carries over to listeners with hearing loss, especially for subjects with mild conductive loss. If this would also hold for other hearing losses one might contemplate to add a simplified version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

3 Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort



Inge Brons, Rolph Houben, Wouter A. Dreschler

Adapted from *Ear and Hearing, 2013, 34(1), pp.29-41.*

3.1 Introduction

One of the main reasons for hearing aid dissatisfaction is the difficulty of listening to speech in noisy environments. Consequently, most currently marketed hearing aids use a single-microphone noise-reduction system to make listening in noisy environments more comfortable. The noise-reduction algorithm continuously analyzes the input signal to estimate the ratio between speech and noise, and reduces the gain for a frequency band if the band is dominated by noise (Chung 2004). Unfortunately, details regarding the properties of noise reduction (e.g., gain-reduction strength, signal-to-noise ratio (SNR) dependency, time constants) in hearing aids are rarely provided by manufacturers. Furthermore, it is unknown whether the perceptual effects of noise reduction (e.g., intelligibility, listening effort, and preference) differ among hearing aids or even among listeners. Consequently, clinicians have no guidelines for selecting the best noise-reduction system and settings. However, if more information were available regarding noise reduction and its effect on the perception of the user, clinicians could actively select the best individual noise-reduction system and settings, thereby increasing hearing aid satisfaction.

Numerous studies have examined the effects of noise reduction on speech intelligibility in noisy environments. These studies reveal that single-microphone noise-reduction system does not improve speech intelligibility (Nordrum et al. 2006; Loizou and Kim 2011).

Despite its inability to increase speech intelligibility, noise reduction has reportedly benefited several listeners using hearing aids (Bentler 2005). Therefore, researchers are increasingly seeking explanations for this perceived benefit. Apparently, benefit from noise reduction can be expected in terms of listening comfort, effort, and personal preference. However, conclusions from studies that have evaluated noise reduction on these outcomes differ. For instance, some studies found that listeners preferred to have noise reduction *on* compared with noise reduction *off* (Boymans and Dreschler 2000; Ricketts and Hornsby 2005). Although this preference for noise reduction suggests that noise reduction increases listening comfort and sound quality, other studies based on rating scales for listening comfort and sound quality found no difference between noise reduction *on* and noise reduction *off* (Alcàntara et al. 2003; Bentler et al. 2008). These two studies also included a rating scale for listening effort. Alcàntara et al. (2003) found equal listening-effort ratings when noise reduction was *on* and *off*, whereas Bentler et al. (2008) found a reduction in listening effort when noise reduction was *on*.

We identified two factors that might have contributed to the apparently conflicting results regarding the perceptual effects of noise reduction: (1) differences between the noise-reduction systems used and (2) individually different weightings of factors underlying the overall preference. We designed a perceptual experiment to further investigate these factors.

3.1.1 Differences between noise-reduction systems

One possible explanation for the diverging results regarding the perceptual effects of noise reduction in hearing aids may be that different hearing aids, and thus different noise-reduction systems, were used in the studies. Each study compared noise reduction *on* with noise reduction *off* within one type of hearing aid. However, technical measurements have shown that noise-reduction systems from different hearing aids can differ substantially in the amount of gain reduction used (Bentler and Chiou 2006; Hoetink et al. 2009). Currently, there are no perceptual data from comparisons of noise reduction from different hearing aids, but we hypothesize that the differences in gain reduction will also have perceptual consequences.

The first goal of this study was to test whether noise-reduction systems on the market differ perceptually. Because noise reduction aims to reduce background noise while retaining speech quality and intelligibility, we determined whether subjects attributed differences among noise-reduction systems in terms of noise annoyance and speech naturalness. This led to the first research question:

Q1. Are there differences regarding noise annoyance or speech naturalness (a) between noise reduction *on* and noise reduction *off* within the same hearing aid and (b) among noise-reduction systems from different hearing aids?

Because we did not know whether differences between noise-reduction systems are perceptually relevant, we used normal-hearing subjects in this initial study. Normal-hearing subjects form a more homogeneous group of listeners than hearing-impaired listeners do. For instance, because of suprathreshold deficits such as reduced frequency resolution and impaired modulation detection, hearing-impaired listeners could differ from one another in which noise-reduction effects are perceptible, complicating the interpretation of the results. In addition, the use of normal-hearing subjects allowed us to compare noise-reduction systems without compensation for an individual hearing loss. Thus, we did not have to use frequency-dependent linear gain or dynamic range compression. Of course, if the present study shows that there are perceptually relevant differences among noise-reduction systems for normal-hearing subjects, the next step should focus on effects for hearing-impaired subjects, taking all these complicating factors into account.

Aside from noise annoyance and speech naturalness, we also determined the overall preference, intelligibility scores, and perceived listening effort of the subjects. Thus, our second research question was as follows:

Q2. Does noise reduction influence the preference, intelligibility, or perceived listening effort of normal-hearing subjects compared with (a) no noise reduction and (b) noise reduction from other hearing aids?

3.1.2 Factors underlying individual preferences

A second possible reason for the diverging results regarding the perceptual effects of noise reduction in hearing aids is that most studies concentrate on group results. However, recent work in our laboratory provided evidence that even normal-hearing listeners differ significantly in their preference for noise-reduction strength (Houben et al. 2012). This study used paired comparisons to determine the preferred setting for noise-reduction strength in an algorithm that was designed for hearing aids. Five of the 10 normal-hearing subjects had an optimized noise-reduction strength that differed significantly from that of the averaged group data. Although the study provided no decisive answer concerning the factors underlying the differences in preference, the results suggest that an individual approach is required in the investigation of the perceptual effects of noise reduction.

The second goal of our study was to determine which factors might influence the preference of the subject for a specific type of noise reduction or for no noise reduction. We looked for correlations between the overall preference and noise annoyance, speech naturalness, intelligibility, and listening effort. Aside from group results, we also determined whether individuals differed in their preference, and we examined the factors related to this preference. Our third research question therefore was:

Q3. Is the overall preference of normal-hearing subjects related to the intelligibility scores, perceived listening effort, noise annoyance, or speech naturalness obtained for the same noise-reduction conditions?

It was also useful to include normal-hearing subjects for this objective. If we were to find substantial differences among subjects even in this homogeneous group of listeners, these differences would be caused by individual differences because there would be no differences in hearing ability to confound the results.

3.2 Methods

3.2.1 Method for the comparison of noise reduction from different hearing aids

In Chapter 2 we developed and evaluated a method that allows for direct comparison of noise-reduction systems of different hearing aids, without the confounding effects of other hearing aid characteristics. Briefly, we made recordings from linearly fitted hearing aids with all the processing features deactivated. On the basis of the difference between the input and output, we designed an equalization filter for each individual hearing aid. This filter was intended to remove perceptual differences among recordings from different hearing aids with signal processing turned *off*.

Once such an inverse filter is available for a specific linearly fitted hearing aid, it can also be applied to recordings from the same hearing aid with noise reduction turned on. The only difference among hearing aids is then caused by the noise reduction because all hearing aids are perceptually equal when noise reduction is turned *off*.

In Chapter 2 several tests were described to verify whether our methods for hearing aid fitting, recording, and filtering indeed removed all the perceptual differences among recordings from different hearing aids with all the processing features deactivated. First, we verified the linearity of the hearing aid gain by electroacoustical measurements. All the hearing aids had a linear response for input levels between 50 and 95 dB SPL for frequencies up to 6 kHz. Second, we calculated the objective hearing aid speech quality index (HASQI; Kates and Arehart 2010) for their hearing aid recordings of speech-in-noise signals. The equalization filter improved the HASQI score compared with band-pass filtering (the mean linear HASQI index was 0.863 for the band-pass-filtered signals and 0.945 for the equalized signals) and reduced the differences in the HASQI score among hearing aids (the maximum HASQI index difference between two hearing aids was 0.02 after band-pass filtering only, and 0.006 after equalization). These HASQI results show that the equalization filter minimized differences among hearing aids. Third, we performed listening experiments to determine whether the recordings from different hearing aids were perceptually similar after filtering. After filtering, all six normal-hearing subjects were unable to detect perceptual differences among recordings from different hearing aids. Thus, both the objective and subjective evaluations showed that the equalization filter removed perceptual relevant differences among the recordings from different hearing aids if the noise-reduction feature was turned off, without affecting the sound quality.

3.2.2 Hearing aid fitting and recordings

The hearing aids tested in this study were of four different brands of frequently used behind-the-ear hearing aids (Phonak Exélia M, ReSound Azure AZ80-DVI, Starkey Destiny 1200, and Widex Mind 440). This selection was a representative sample of the commercial hearing aids available at the time of the study. The hearing aid numbers used in this study were randomly assigned to the test hearing aids and are different from the numbers used in Chapter 2.

We applied the same methods of hearing aid fitting and recording as described in Chapter 2. In fact, we took the hearing aids, hearing aid settings, and equalization filters from the test set in Chapter 2, so that we were sure that our verification of the method was also applicable to the new recordings. We turned *off* all the signal-processing features in the hearing aids (directionality, feedback control, noise reduction, compression, frequency transposition, etc.) and carefully adjusted their gains to obtain the same insertion gain for all hearing aids. The target insertion gain was based on the NAL-RP prescription (Dillon 2001) for a conductive hearing loss of 30 dB at 500 Hz and 15 dB at 2 kHz. This resulted in an insertion gain of approximately 10 dB in the low and mid frequencies (between 125 Hz and 2 kHz), decreasing to 0 dB for the higher frequencies.

We recorded the hearing aid output with the use of a B&K Head and Torso Simulator (HATS Type 4128C), which was fitted with a custom-made tight-fitting earmold without venting. The recordings were made in a sound-treated double-walled booth (2.20 2.53 2.00 m). The speaker was placed 62 cm in front of the recording microphone (on axis) to minimize the influence of room reflections. All the free-field hearing aid input signals were corrected for the speaker response.

We designed an inverse filter for each hearing aid to remove any differences in frequency response that remained among hearing aids despite the careful adjustment of the hearing aid gain. To obtain the required filter response, we compared the hearing aid output to the output of a reference microphone. We used the Matlab function “fir2” to calculate the filter coefficients (500 taps) based on the required response. In addition, the frequency response was limited to 100 Hz to 5.8 kHz with elliptical filters of the seventh-order.

As described earlier, when a filter has been designed for a specific linearly fitted hearing aid, it can also be applied to recordings from the same hearing aid with noise reduction turned *on* to examine the isolated effect of noise reduction. We selected in each hearing aid the strongest available noise-reduction setting. For the purpose of exploring possible perceptual effects of noise reduction, investigators often use the maximum setting (Ricketts and Hornsby 2005; Mueller et al. 2006; Nordrum et al. 2006; Palmer et al. 2006). In some hearing aids, the maximum noise-reduction setting was not the setting recommended by the manufacturer for an initial fit. However, if there are no perceptual differences between setting the noise reduction on and off or among different noise-reduction systems even if maximum noise reduction is applied, it is reasonable to hypothesize that there will also be no perceptual differences for lower settings.

3.2.3 Stimuli

We made hearing aid recordings of Dutch female speech (Versfeld et al. 2000) in a multitalker babble noise (Luts et al. 2010). The speech material consists of unrelated, low-context sentences, containing five to nine words per sentence. The signals were presented to each hearing aid with a noise level of 70 dB(A) and four different speech levels (63, 66, 70, and 74 dB(A)) to form stimuli at SNRs of -7, -4, 0, and +4 dB. These levels were well within the linear range of the hearing aids (50-95 dB SPL). The negative SNRs were chosen to prevent ceiling effects for intelligibility. We rated the listening effort and paired comparisons at -4 dB SNR as well so that we had one common SNR across all outcomes. Because hearing-impaired listeners have more difficulty with low SNRs, they are less likely to listen in a setting with -4 dB SNR during their daily activities. Therefore, we additionally measured listening effort at 0 and +4 dB SNR, and noise annoyance, speech naturalness and overall preference at +4 dB SNR. These SNRs are more relevant for hearing-impaired listeners and thus more relevant in evaluating hearing aid processing. The noise was continuous while the speech was paused for approximately 1 sec between sentences. One list (36 sec) preceded the stimulus lists in each condition to allow the hearing aid to adapt to the input signals.

We applied the inverse filters to all the hearing aid recordings. This step resulted in five different conditions. One condition represented the situation in which the noise reduction was turned *off* (hereafter the "unprocessed" condition). The four additional conditions represented the situations in which the noise reduction (NR) was turned *on* for each hearing aid (randomly coded NR1-NR4).

Stimuli for all the measurements consisted of single sentences with 0.5 sec of noise before and after the sentence. The stimuli were presented diotically with Sennheiser HDA200 headphones, which had been calibrated with a B&K Artificial Ear Type 4153. The noise level was 70 dB(A) for all the stimuli in the unprocessed condition.

3.2.4 Noise-reduction processing

Although we had no details on the implementation of the different noise-reduction systems, acoustical analyses gave insight into how they reacted to our input signals. We plotted the time signals and difference spectrograms to study the dynamic characteristics of the noise-reduction conditions. Figure 3.1 shows the results of this study. For each noise-reduction system, the upper plot shows the time signal of the hearing aid output with the noise reduction *off* (dark gray) compared with the output with the noise reduction *on* (light gray). In addition, the spectrogram-like color plots show the difference between noise reduction *on* and *off* (i.e., the gain reduction caused by the noise reduction) as a function of time and frequency. Note that each plot starts and ends with 0.5 sec noise only, with the sentence in between. The more negative the gain value, the stronger the noise reduction. Thus, red areas correspond to the maximum noise reduction (>10 dB reduction of the gain), whereas blue areas correspond to an inactive noise reduction. Figure 3.2 shows the long-term average gain reduction (averaged over 13 sentences) due to the four noise-reduction conditions as a function of frequency. Table 3.1 summarizes these plots by giving the median and the 5th and 95th percentile of the gain values (the difference between noise reduction *on* and *off*, in dB) across all time-frequency bins. The 5th percentile provides an estimate of the maximum gain reduction applied by the noise reduction. Again, a more negative value indicates stronger suppression. Similarly, the 95th percentile is an estimate of the minimum gain reduction.

3.2.5 Subjects

The number of subjects chosen for this study was based on a power calculation for speech intelligibility and listening effort. We used a within-subject standard deviation of approximately 13.9% taken from Bosman (1989). The slope of the psychometric function at the speech-reception threshold (SRT_{50}; the SNR at which 50% of the sentences are correctly repeated by the subject) was 16% per dB. To be able to detect a difference of 16% (thus 1 dB change in SRT_{50}) nine subjects should be included for α = 0.05 and 1-β = 0.8. For a power calculation for listening-effort rating, we used the within-subject standard deviation found by Luts et al. (2010), which averaged 0.88 points on the 7-point listening effort scale. With α = 0.05 and 1-β = 0.8, nine subjects are sufficient to detect a 1-point difference on the listening-effort rating scale. For the outcomes measured with paired comparisons, we had no appropriate data available for a power calculation. However, for these data we were especially interested in individual differences, so that the number of subjects was of less importance for these outcome measures. On the basis of the power calculations, we decided to include 10 normal-hearing subjects.

Ten normal-hearing subjects between 19 and 23 years of age (average = 20.8 years) participated in this study. Their hearing thresholds were 15 dB HL or better at 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz.

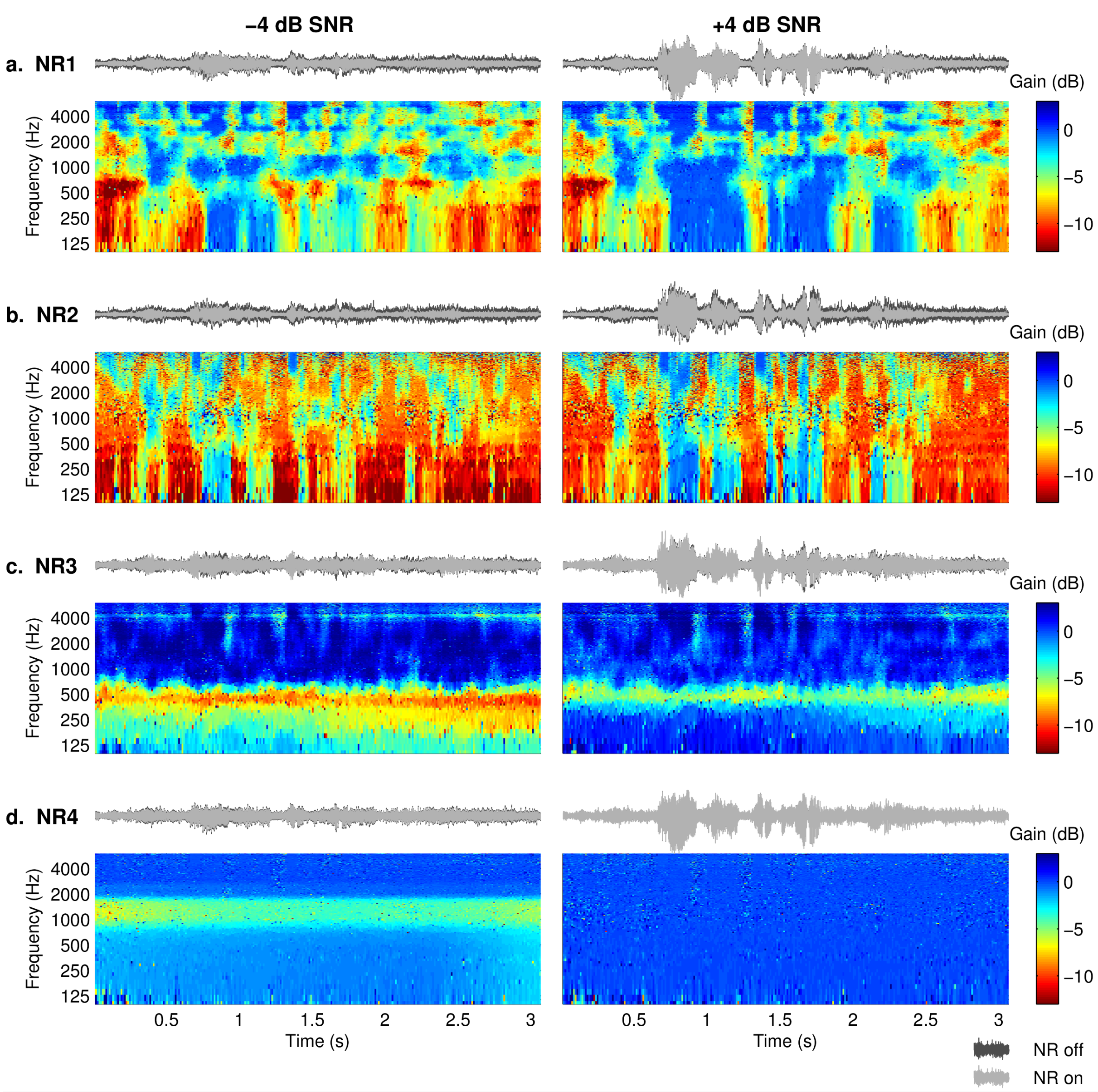


Figure 3.1: Acoustical effects of the four noise-reduction systems on speech in babble noise at -4 dB SNR (left column) and +4 dB SNR (right column). The time signal of the hearing-aid output with noise reduction *off* (dark background signal) and *on* (light foreground signal) for each processing condition is shown. The changes in gain caused by noise reduction (the difference between noise reduction *on* and *off*) as a function of time and frequency are also shown.

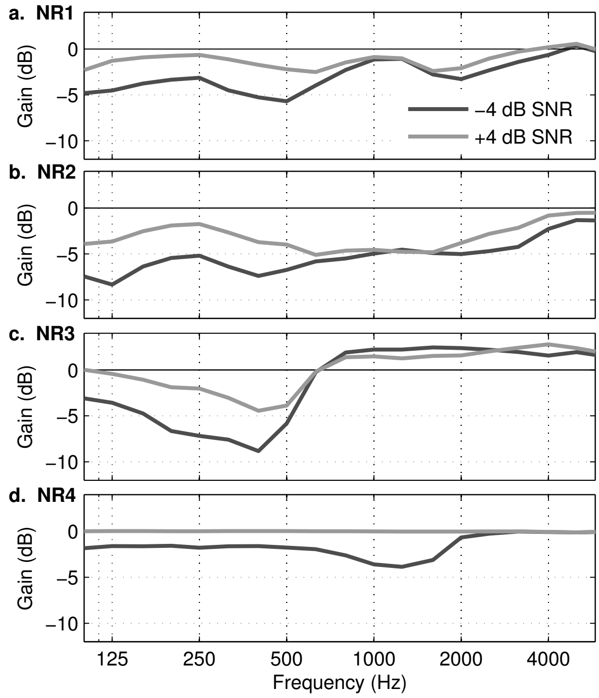


Figure 3.2: Gain-reduction spectra for the four noise-reduction conditions. The difference between noise reduction *on* and *off* as a function of frequency is shown, averaged over 13 sentences in noise at two different input SNRs.

Table 3.1: Median and 5th and 95th percentile values of the gain difference between noise reduction and unprocessed conditions for the time-frequency ins that are presented in Figure 3.1.

	-4 dB Signal-to-noise ratio			+4 dB Signal-to-noise ratio		
	median	5th percentile	95th percentile	median	5th percentile	95th percentile
NR1	-4.1	-11.5*	-0.4	-3.4	-10.0*	+0.7
NR2	-7.0	-12.7*	-1.2	-6.8	-11.1*	+0.3
NR3	+0.8	-9.3*	+2.9	+0.6	-6.1*	+3.0
NR4	-0.4	-4.5	+0.5	-0.0	-0.9	+0.9

All values are in dB.

* Calculated over time-frequency bins between 100 and 750 Hz. Taking the whole frequency range into account resulted in an underestimation of the maximum gain reduction, because maximum gain reduction was acquired in the lower frequencies ($\leq$ 750 Hz), where the number of time-frequency bins was much lower than in the higher frequencies.

3.2.6 Paired-comparison rating

We used paired-comparison rating (a two-interval, seven-alternative forced choice paradigm) to measure noise annoyance, speech naturalness, and overall preference. This method was based on a standard of the International Telecommunication Union (ITU-T P.835, ITU-T 2003), according to which subjects must give separate ratings for the speech signal, background noise, and overall quality. The ITU standard uses a rating scale to measure quality. We chose to use paired comparisons instead because these are more sensitive to subtle differences in conditions (Böckenholt 2001).

For each pair of stimuli (the same sentences in two different processing conditions), the subjects answered three questions. The first time they listened to the two fragments A and B, the subjects were asked to concentrate on the speech and to rate in which of the two fragments the speech was more natural and to indicate the strength of the difference. After they made a choice, they listened to the same fragments again, now concentrating on the annoyance of the noise and selecting the least annoying fragment. The subjects could listen to both fragments again before they answered the third question, but this was not required. For the third question, the subjects were asked which fragment that they would prefer for prolonged listening. For each question, there were seven possible answers, ranging from "A is much more natural/much less annoying/much better" to "B is much more natural/much less annoying/much better." The seven choice categories were derived from the comparison category rating method described in ITU-T P.800 (ITU-T 1996). The subjects were able to indicate no difference between A and B. They were allowed to listen to the fragments as often as they preferred before they answered each question.

All five conditions were paired with each of the other conditions, which resulted in 10 different stimulus pairs. Three runs of 10 comparisons were performed at both -4 and +4 dB SNR, which resulted in a total of 60 comparisons per subject (10 Pairs x 3 Runs x 2 SNRs). The choice for three runs of comparisons was based on previous studies that used paired comparisons to determine preference for noise-reduction algorithms or settings (Boymans and Dreschler 2000; Ricketts and Hornsby 2005; Luts et al. 2010). All the subjects started with four training pairs. Subsequently, five subjects started with all the comparisons at -4 dB SNR, and the other five subjects started at +4 dB SNR.

3.2.7 Intelligibility

We measured intelligibility as the percentage of words that the subjects repeated correctly at -4 and -7 dB SNR. Each subject started with 13 training sentences containing all five processing conditions, starting at +4 dB SNR. After every three sentences, the SNR decreased one step (4 dB for the first two steps and 3 dB for the last step), terminating with an SNR of -7 dB for the last four sentences. After this training list, we used one list per processing condition per SNR to determine the intelligibility scores. We balanced the order of conditions across all the subjects to minimize the possible effects of training on the group data. We also balanced the measurement lists across the conditions, to minimize possible effects of lists. Every new combination of processing condition and SNR started with 3 training sentences, followed by 10 sentences that were used to calculate the percentage of correct words. All words, not only key words, were included in this calculation.

3.2.8 Listening-effort rating

The subjects rated the listening effort on a 9-point rating scale that ranged from "no effort" to "extremely high effort." This test is similar to the test used by Luts et al. (2010) but differed in that our scale used five labeled buttons instead of seven. The labels were based on ITU-T P.800 methodology (ITU-T 1996). The subjects gave ratings for the five processing conditions at three SNRs (-4, 0, and +4 dB), thus for 15 different conditions. Each subject started with a practice run of 15 conditions. This practicing run was followed by three additional runs that we used for analysis.

3.3 Results

3.3.1 Paired-comparison rating

Figure 3.3 shows the average rating score for each processing condition. We assigned scores from -3 to 3 for each condition, according to the ITU-T recommendation P.800 (ITU-T 1996). For instance, if the subject rated condition A slightly better than condition B, we assigned a score of 1 to condition A and a score of -1 to condition B. The scale for the noise annoyance is inverted in Figure 3.3. For each outcome, a symbol plotted above the zero line means a better performance on that judgment criterion.

Because we could not expect the scorings to represent a linear interval scale, we used the log-linear modeling approach for ordinal paired comparisons described by Dittrich et al. (2004) for the statistical analysis of the paired-comparison rating data. The model is a log-linear representation of the Bradley-Terry model (Bradley and Terry 1952) and is extended for paired-comparison data with multiple response categories, including a "no difference" option. By fitting this model to the paired-comparison data, we obtained estimates of the "worth" parameters, which describe the location of the five processing conditions on the subjects preference scale. This scale can be interpreted similarly to a ratio scale, thus providing not only the ranking of preference for the five conditions but also information regarding the strength of preference.

We estimated the worth parameters separately for noise annoyance, speech naturalness, and overall preference. We fitted a model for each individual run of 10 comparisons, which resulted in three models per subject per SNR per judgment criterion. We tested the goodness-of-fit for all of the models by comparing the obtained model with a saturated model (a model reproducing the data perfectly). All the p values were above 0.95, indicating a high agreement with the saturated model, and thus, all models could be accepted.

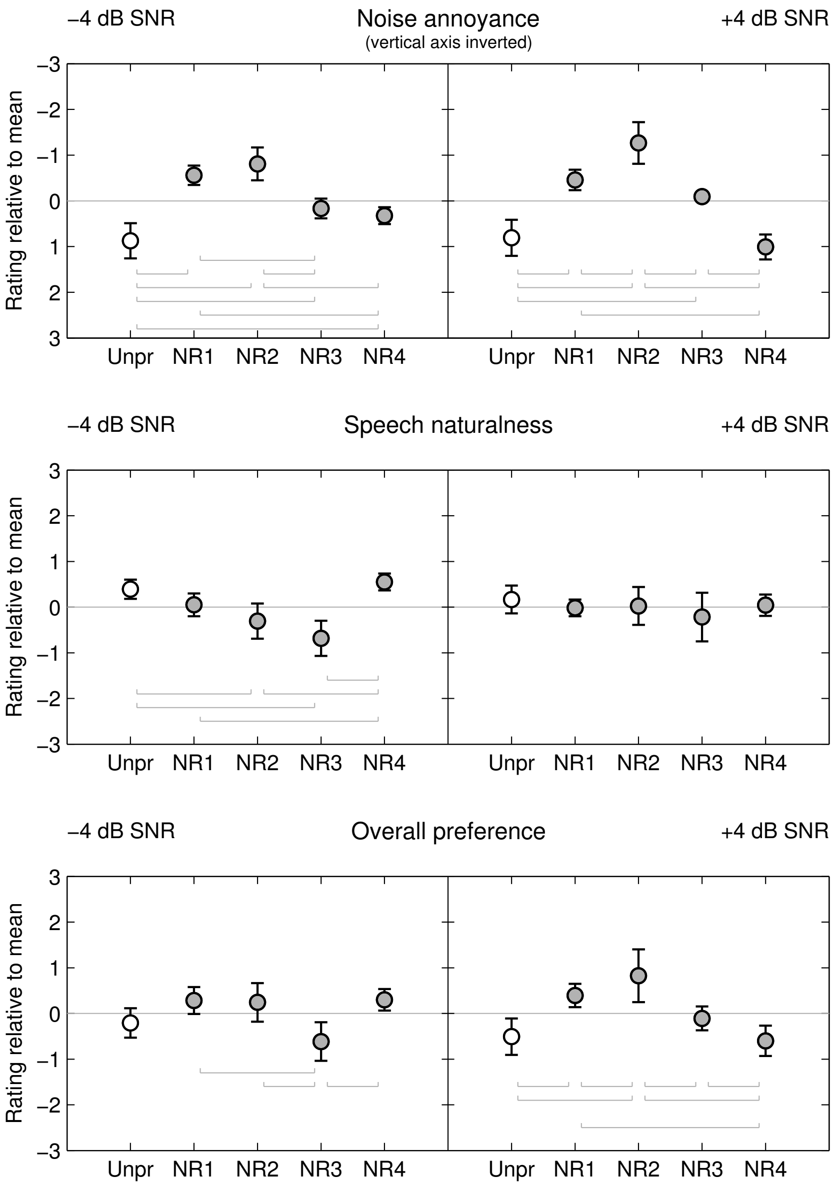


Figure 3.3: Mean rating scores derived from the paired-comparison data for the three judgment criteria and two SNRs. Scores from -3 to +3 were assigned with 0 indicating no difference; -1 and +1 indicating a minor difference; -2 and +2 indicating a moderate difference and -3 and +3 indicating a major difference. Error bars show the 95% confidence interval among subjects (without Bonferroni correction). Higher values indicate better performance. Horizontal bars indicate which processing conditions differ significantly from each other after Bonferroni correction for 10 comparisons.

We performed repeated-measures analysis of variance (ANOVA) on the estimated worth parameters for each judgment criterion separately (noise annoyance, speech naturalness, and overall preference) with SNR and processing condition as fixed effects and subject as a random effect. The resulting F statistics and p values are presented in Table 3.2. We found a significant effect of processing condition for each of the three judgment criteria. In addition, we found significant interactions between processing condition and SNR for all three criteria. Because of the significant interaction between processing condition and SNR, we performed a subsequent repeated-measures ANOVA for each SNR separately, with processing condition as a fixed effect and subject as random effect. The resulting values for F and p are also given in Table 3.2. The effect of processing condition was significant for each judgment criterion at both SNRs, except for the speech naturalness at +4 dB SNR. The horizontal lines in Figure 3.3 indicate which conditions differed significantly from each other after Bonferroni correction.

Table 3.2: Main analysis of variance outcomes for the paired-comparison results.

		Noise annoyance		Speech naturalness		Overall preference	
		Both signal-to-noise ratios					
Effect	df	F	*p*	F	*p*	F	*p*
Processing condition	4	24.57	< 0.001	3.21	0.023	5.61	0.001
SNR	1	0.02	0.892	2.68	0.136	1.34	0.276
Processing condition x SNR	4	8.23	< 0.001	4.42	0.002	8.79	< 0.001
		-4 dB signal-to-noise ratio					
Processing condition	4	18.26	< 0.001	8.21	< 0.001	3.65	0.014
		+4 dB signal-to-noise ratio					
Processing condition	4	24.29	< 0.001	0.33	0.854	8.04	< 0.001

3.3.2 Intelligibility

The left panel in Figure 3.4 shows the percentage of words correctly repeated averaged over all 10 subjects. For statistical analysis, we transformed these percentages to rationalized arcsine units (Studebaker 1985) and subsequently performed a repeated-measures ANOVA on the transformed data with SNR and processing condition as fixed effects and subject as a random effect. We found significant effects of SNR ($F[1, 9] = 39.0$, $p < 0.001$) and processing condition ($F[4, 36] = 4.4$, $p = 0.005$). Post hoc Bonferroni-corrected pairwise comparisons showed that the scores for NR2 were significantly worse than those for NR4 (uncorrected $p = 0.0045$). The differences between the conditions are shown in the right panel of Figure 3.4, in which the scores of each noise-reduction condition are plotted relative to the scores of the unprocessed condition.

3.3.3 Listening-effort rating

The left panel in Figure 3.5 shows the mean listening-effort ratings assigned by the 10 subjects. Note that a higher value means that the listening effort was lower. To satisfy the ANOVA criteria, we transformed the listening-effort ratings with an arcsine transformation.

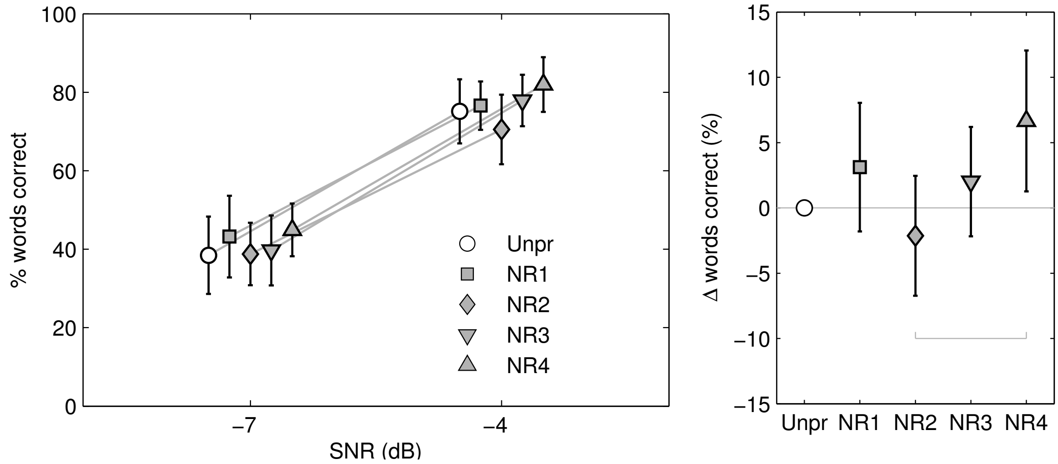


Figure 3.4: Left panel: Mean and 95% confidence interval of the percentage of words correctly repeated by the 10 subjects at -7 and -4 dB SNR. Right panel: Mean results from both SNRs relative to the unprocessed condition. Horizontal bars indicate which processing conditions differ significantly from each other after Bonferroni correction for 10 comparisons.

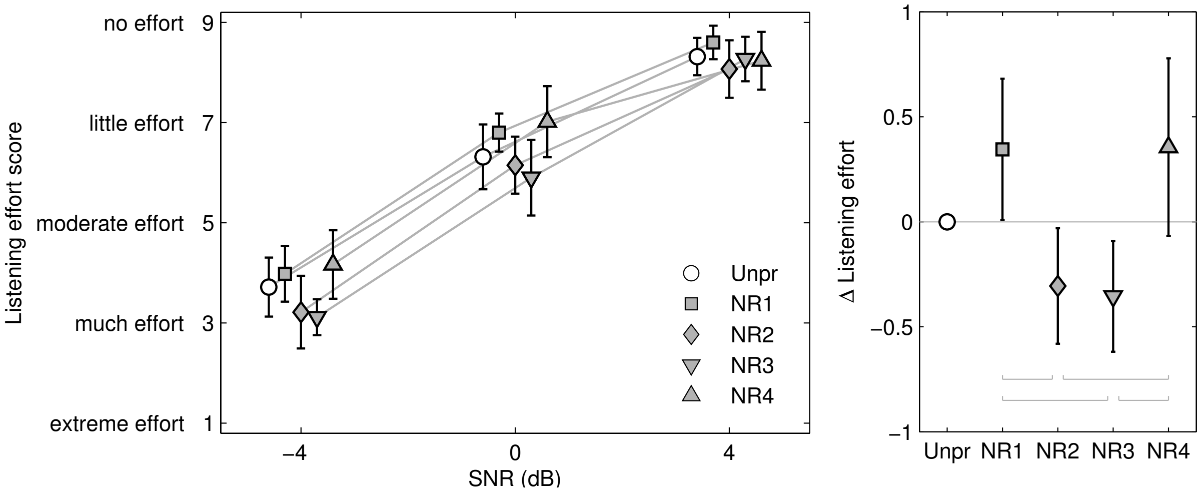


Figure 3.5: Left panel: mean and 95% confidence intervals of the listening effort ratings assigned by the 10 subjects at -4, 0 and +4 dB SNRs. Right panel: mean results from all 3 SNRs relative to the unprocessed condition. Horizontal bars indicate which processing conditions differ significantly from each other after Bonferroni correction for 10 comparisons.

We performed a repeated-measures ANOVA with SNR and processing condition as fixed effects and subject as a random effect. We found significant effects of SNR ($F[2, 18] = 155.4$, $p < 0.001$) and processing condition ($F[4, 36] = 6.0$, $p < 0.001$). Post hoc pairwise comparisons with Bonferroni correction showed that NR1 and NR4 involved significantly lesser effort than NR2 and NR3 did (uncorrected $p < 0.05$). The right panel of Figure 3.5 shows the differences between each processing condition and the unprocessed condition averaged across all three SNRs.

3.3.4 Relations between outcome measures

Because all the outcomes were measured at -4 dB SNR, we used the (transformed) data from this SNR to determine whether the overall preference was related to the other outcome measures. For noise annoyance, speech naturalness, overall preference (worth estimates), and listening effort, we calculated for each subject the average of the three repeats. Thus, we obtained one value per condition per subject for each outcome measure (and thus 5 processing conditions x 10 subjects = 50 values per outcome measure). We standardized the data by subtracting the mean and dividing by the standard deviation so that all the outcome measures had zero mean and a standard deviation of 1. Figure 3.6 shows the standardized results for all the outcome measures at -4 dB SNR.

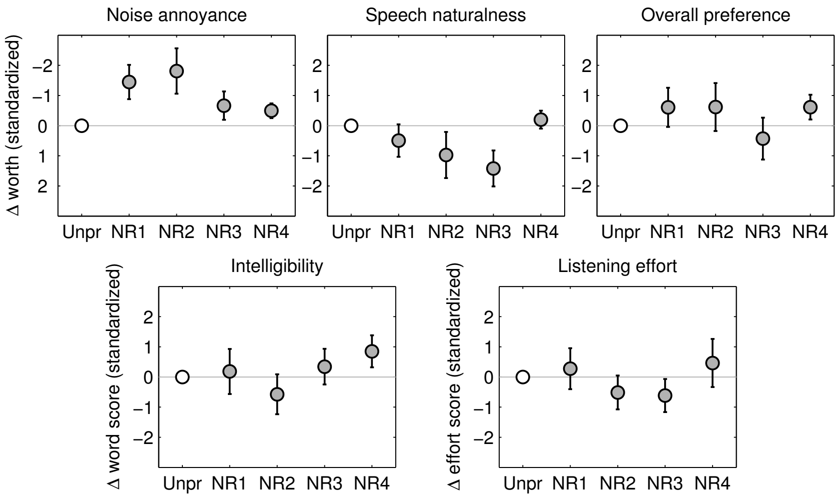


Figure 3.6: Results for all outcomes at -4 dB SNR (means and 95% confidence intervals over 10 subjects). For each outcome measure, data were transformed, standardized, and plotted relative to the unprocessed condition.

We calculated Pearsons correlation coefficients between the overall preference and each outcome. After Bonferroni correction, we found significant correlations between the overall preference and the noise annoyance ($r = 0.48$, $n = 50$, $p < 0.001$) and between the overall preference and the speech naturalness ($r = 0.50$, $n = 50$, $p < 0.001$). Noise annoyance and speech naturalness were not significantly correlated to each other ($r = -0.04$, $n = 50$, $p = 0.76$). Overall preference was not significantly correlated to intelligibility scores ($r = 0.01$, $n = 50$, $p = 0.94$) or listening effort ($r = 0.26$, $n = 50$, $p = 0.07$).

To determine whether overall preference could be predicted by noise annoyance and speech naturalness, we applied linear regression analysis to the worth estimates. We performed this analysis both on the group and individual levels. We used the data from each individual run of 10 comparisons, which resulted in 15 values (5 processing conditions x 3 runs) for noise annoyance, speech naturalness, and overall preference per subject per SNR. For the group analysis, this process resulted in 150 values per judgment criterion (10 subjects x 15 worth values). We used backward stepwise regression with thresholds of 0.05 for entering or removing terms. The dependent variable was the overall preference, and the independent variables were the noise annoyance and speech naturalness. In this way, we could determine whether the overall preference could be predicted by either noise annoyance or speech naturalness, by these factors together, or by none of these factors.

The left panel in Figure 3.7 shows the standardized regression coefficients (β) for noise annoyance and speech naturalness for each individual and for the entire group at -4 dB SNR. Because the data were standardized, a higher coefficient indicates a greater effect of that variable on the overall preference. Thus, for all of the subjects together ("group"), both noise annoyance and speech naturalness contributed equally to the overall preference (β noise annoyance = 0.54 and β speech naturalness = 0.57), together explaining 56% of the variance in overall preference (R^2 = 0.56). For subjects 1, 2, 4, and 8 noise annoyance and speech naturalness were also both included in the best model (nonzero β coefficients, see Figure 3.7). The R^2 values for subjects 1, 2, 4, and 8 were, 0.79, 0.88, 0.91, and 0.76, respectively. In contrast, for subject 3, neither noise annoyance nor speech naturalness was included. For the remaining five subjects, only one of the variables remained in the best regression model: speech naturalness for subjects 5, 6, and 10 (with R^2 0.83, 0.60, and 0.44, respectively) and noise annoyance for subjects 7 and 9 (with R^2 0.41 and 0.49, respectively).

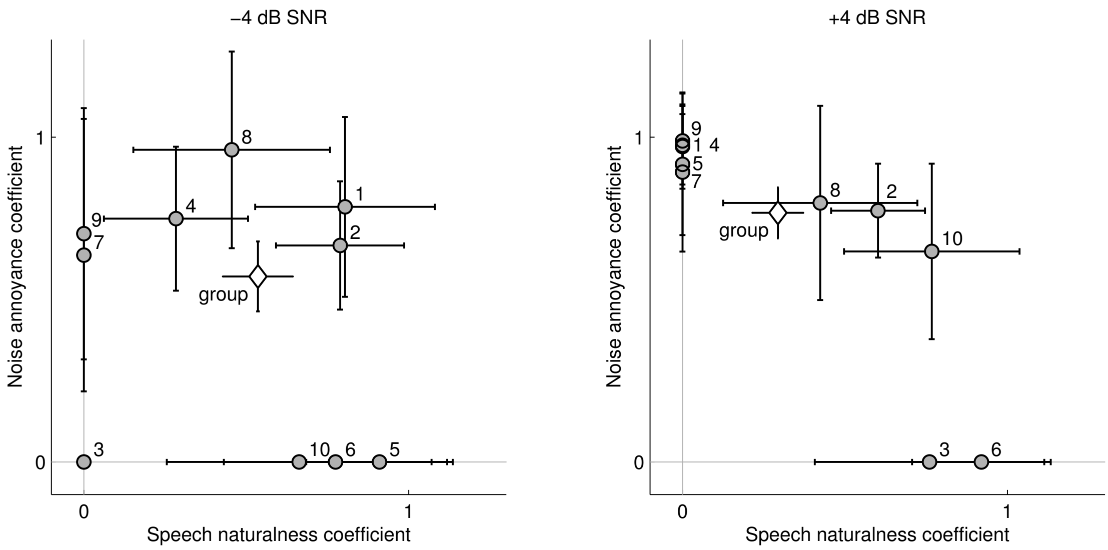


Figure 3.7: Standardized regression coefficients (β) for noise annoyance (vertical axis) and speech naturalness (horizontal axis) at -4 dB SNR (left panel) and +4 dB SNR (right panel). Mean and 95% confidence intervals were given for each individual subject (circles) and for the group results (diamonds).

Because worth estimates for noise annoyance, speech naturalness, and overall preference were also measured at +4 dB SNR, we performed the same regression analysis at this SNR. This process yielded the coefficients represented in the right panel of Figure 3.7. For the group, the noise annoyance was weighted more than the speech naturalness (β noise annoyance = 0.77 and β speech naturalness = 0.29), together explaining 78% of the variance in overall preference (R^2 = 0.78). For 5 of the 10 subjects, noise annoyance was the only explaining variable, and for 2 subjects, speech naturalness was the only explaining variable. For the remaining three subjects, both factors were included in the best model. R^2 values for the individual regression models at +4 dB SNR ranged between 0.58 and 0.98 and were all higher than at -4 dB SNR (except for subject 8, for whom R^2 was 0.04 lower at +4 dB SNR than at -4 dB SNR).

3.4 Discussion

With respect to our research questions, we can summarize our findings as follows:

Q1. Hearing aid noise reduction is able to reduce the annoyance of babble noise for normal-hearing listeners at -4 dB SNR (all four hearing aids) and at +4 dB SNR (three of the four hearing aids). At -4 dB SNR, however, two of the four noise-reduction systems also reduce speech naturalness. The noise-reduction systems differ from one another in how strongly they reduce noise annoyance and preserve speech naturalness.

Q2. Normal-hearing listeners prefer noise reduction over no noise reduction within two of the four hearing aids at +4 dB SNR. For our selection of hearing aids, noise reduction provides no statistically significant benefit in terms of intelligibility or listening effort compared with no noise reduction. Compared with each other, however, noise-reduction systems differ mutually in terms of all the outcome measures.

Q3. The overall preference of normal-hearing listeners correlates to noise annoyance and speech naturalness, but not to intelligibility or listening effort. There are differences among individual listeners in whether they place more weight on noise annoyance or on speech naturalness in determining their overall preference.

3.4.1 Noise annoyance, speech naturalness, and overall preference

This study is, to our knowledge, the first in which the perceptual effects of noise-reduction systems from different hearing aids are directly compared with each other. The recording and filtering technique developed in Chapter 2 allowed us to compare all combinations of noise-reduction systems directly with each other in a paired-comparison design. Even within the small sample of hearing aids included, the measurements show that noise-reduction systems differ perceptually, as was previously suggested based on technical differences (Bentler and Chiou 2006; Hoetink et al. 2009).

The paired-comparison data (Figure 3.3) show that noise-reduction systems were able to reduce the noise annoyance at +4 dB SNR without affecting the speech naturalness. In the more difficult listening situation of -4 dB SNR, however, the results reveal a trade-off between noise reduction and speech distortion. Apparently, at this lower SNR, it was harder for the noise-reduction system to differentiate between speech and noise, so the reduction of noise was accompanied by distortion of the speech signal. Indeed, Figure 3.1 shows that the gain applied by the four noise-reduction systems differs between the two input SNRs (-4 and +4 dB). For instance, in the right column of Figure 3.1 we see less gain reduction (blue areas) in the spectrograms of NR1 and NR2 during the moments that speech is present (speech presence is visible from the time-signal plots shown above the spectrogram), indicating that the noise reduction does not reduce the gain during speech presence. In the left column, these blue areas for NR1 and NR2 are smaller than in the right column, indicating that at -4 dB SNR the noise reduction had more difficulty recognizing the speech in the speech-plus-noise mixture. This implies that the main cause for the reduced speech naturalness at -4 dB SNR is the unwanted suppression of the speech signal. In addition, the quick changes in gain introduced by NR1 and NR2 may have caused distortions to the speech signal. The underlying mechanisms for NR3 and NR4 seem to differ strongly from those of NR1 and NR2. NR3 reduces the gain for frequencies of up to 500 Hz, which seems to be independent of the presence of speech. This low-frequency gain reduction by NR3 is much stronger at -4 dB SNR than at +4 dB SNR (Table 3.1, Figure 3.1 and Figure 3.2: at -4 dB SNR the gain is reduced up to 9 dB, whereas at +4 dB SNR the maximum lies at approximately 6 dB). This low-frequency gain reduction is likely the cause of the reduced speech naturalness for NR3 at -4 dB SNR. NR4 only reduces the gain between 1 kHz and 2 kHz at -4 dB SNR and does not reduce the gain at +4 dB SNR. The small amount of gain reduction by NR4 at -4 dB SNR seemed to be able to reduce noise annoyance, although to a lesser extent than that by NR1 and NR2. The advantage of NR4 at -4 dB SNR was that the speech naturalness was kept intact, so that it was not less preferred than NR1 and NR2. However, at +4 dB SNR, NR4 lacks the advantage of reduced noise annoyance, which was found most clearly for NR1 and NR2.

Our results may help to understand the diverging results from previous studies. First, we found that subjects preferred noise reduction *on* over noise reduction *off* in only two of the four hearing aids. The four studies mentioned in the Introduction used another type of hearing aid, which may have contributed to their apparently conflicting results. Second, our results showed significant preferences for noise reduction *on* over noise reduction *off* only at +4 dB SNR and not at -4 dB SNR. Indeed, the two studies that found a significant preference for noise reduction *on* over noise reduction *off* both measured preference at positive SNRs (Boymans and Dreschler 2000, at +5 dB SNR; Ricketts and Hornsby 2005, at +1 and +6 dB SNR), as opposed to Alcàntara et al. (2003), who measured mainly at negative SNRs and did not find changes in quality or comfort because of noise reduction. In the study by Bentler et al. (2008), the measurement conditions differed per listener so that no systematic effect of the SNR could be derived. Although numerous other factors (for instance, the type of noise, other hearing aid characteristics, and hearing ability), play a role, we conclude that differences among noise-reduction systems and their dependency on SNR should be considered in the interpretation of noise-reduction studies.

3.4.2 Intelligibility

Our finding that none of the noise-reduction systems changed intelligibility scores compared with unprocessed corresponds to previous findings (Bentler 2005; Nordrum et al. 2006). We also found that intelligibility differed slightly between two noise-reduction systems. Scores for NR2 and NR4 differed on average 8.8%, which was just significant ($p = 0.0045$, with Bonferroni-corrected $\alpha = 0.005$). One should keep in mind that this difference was measured in a laboratory situation with only one noise type and at the most sensitive point of the psychometric function (50%). Thus, it remains to be seen whether this improvement leads to an actual benefit in real-life situations.

Nordrum et al. (2006) also compared the effect of noise-reduction systems from different hearing aids on intelligibility but did not find significant differences among noise-reduction systems. The authors did not equalize other hearing aid characteristics, so results from different hearing aids could not directly be compared with each other. Furthermore, all previous intelligibility measurements with hearing aid noise reduction were performed with hearing-impaired listeners. The difference among listeners complicates comparison with our results because we measured at lower SNRs.

Luts et al. (2010) compared single-microphone noise-reduction systems that were developed and optimized for use in hearing aids. For normal-hearing subjects and the same sentence material and background noise as we used, the authors found an SRT_{50} (the SNR at which 50% of the sentences were correctly repeated) of -5.2 dB SNR in the unprocessed condition for normal-hearing listeners. The two single-microphone noise-reduction systems they evaluated did not change this value. Although the measurement methods differed (i.e., sentence scoring in an adaptive procedure versus word scoring at a fixed SNR), the results of Luts et al. were in agreement with ours.

3.4.3 Listening effort

Noise reduction did not change listening effort compared with the unprocessed condition. If noise reduction was *on*, two of the noise-reduction systems required slightly more effort than the other two systems. In the Introduction, we mentioned two studies that determined listening effort for listening with noise reduction *on* and noise reduction *off* within a single hearing aid. One study found a reduction in listening effort because of noise reduction (Bentler et al. 2008), and the other study found no effect (Alcàntara et al. 2003). On the basis of our results, it seems reasonable that, other than factors like noise type and SNR, the differences among noise-reduction systems contributed to differences among the results from different studies.

Although there is a common assumption that noise reduction may reduce listening effort compared with no noise reduction, there is little evidence confirming this assumption, and our data did not confirm it. The main difficulty in studying listening effort is the lack of a proper method to measure listening effort (Bentler 2005; Edwards 2007; Lunner et al. 2009). Such a test would ideally be able to evaluate noise reduction in situations relevant for the user, thus at SNRs where the speech is highly intelligible. In these situations speech-intelligibility tests suffer from ceiling effects, but the effort required to obtain the same intelligibility score may differ among situations. For instance, Sarampalis et al. (2009) used a dual-task paradigm to measure listening effort and found decreasing response time (indicating decreasing listening effort) with increasing SNR, even at SNRs where speech was highly intelligible. Although the dual-task paradigms such as in the example of Sarampalis et al. seem promising, the use of a secondary task in the nonauditory domain complicates the interpretation of the results and often requires a specialized test equipment. The disadvantage of our method was that we found a ceiling effect for listening effort at the SNR where speech was highly intelligible (+4 dB). Thus, a more appropriate method is required to measure the potential effects of noise reduction on listening effort.

3.4.4 Relations between all outcome measures

NR4 performed better than the other noise-reduction conditions in terms of intelligibility and listening effort (Figure 3.6), although it had the weakest gain reduction (Table 3.1). In contrast, NR2 applied the strongest gain reduction (Table 3.1) and reduced noise annoyance more than any other condition but did not perform well with respect to intelligibility and listening effort (Figure 3.6). This difference may be explained by the fact that the stronger reduction in noise by NR2 also affected the speech more severely. Considering these results, it is surprising that NR3, which reduced speech naturalness, preference, and listening effort the most, did not have the worst intelligibility score. However, from Figure 3.1c and Figure 3.2c, it is clear that NR3 reduces gain for lower frequencies ($\leq$ 500 Hz), while increasing the gain for higher frequencies ($>$ 500 Hz), with no clear relation to the presence or absence of speech. Thus, the reduction in speech naturalness for NR3 was most likely caused by the low-frequency gain reduction, in contrast to the reduced speech naturalness for NR2, which was caused by the suppression of different speech fragments and quick changes in gain. Although our normal-hearing subjects did not prefer the spectral shaping caused by NR3 and they did not benefit from it either in terms of speech intelligibility, this might be different for hearing-impaired listeners. For the normal-hearing subjects, the audibility of all stimuli was maintained. However, hearing-impaired listeners could benefit from the small increase of gain that NR3 applies for frequencies above 1 kHz. These frequencies have been shown to be important for speech perception in noise (Smoorenburg 1992), so maintaining or increasing the audibility of this part of the signal could be beneficial for hearing-impaired subjects.

3.4.5 Relations among preference, noise annoyance, and speech naturalness

We found that the overall preference of our subjects was related to noise annoyance and speech naturalness. This finding corresponds with results of other studies. For instance, Hu and Loizou (2007b) evaluated different speech-enhancement algorithms using the ITU-T P.835 methodology, according to which subjects give ratings for the background intrusiveness, the signal distortion, and the overall quality of sound samples. The authors also performed linear regression on the ratings, which resulted in a noise coefficient of 0.37 and a speech coefficient of 0.57. The authors concluded that listeners integrate the effects of both signal and background distortions when assigning ratings for overall quality, but that they seem to place more emphasis on the speech distortion than on the background noise. Marzinzik (2000) used paired comparisons with the same three judgment criteria and also concluded that speech distortions counterbalanced the reduction of noise in the overall preference judgments. However, he performed no further analysis to determine the weighting between both parameters.

To date, studies have only looked at group results and provided no insight into possible individual differences. However, Houben et al. (2012) performed paired comparisons on different settings for the strength of noise-reduction in an algorithm designed for hearing aids and found significant interindividual differences in preferences among normal-hearing subjects. Because stronger noise reduction introduces more speech distortion, the authors hypothesized that listeners differ in their preference for a trade-off between noise annoyance and speech distortion. Our regression results support this hypothesis. Although the preference of some subjects seems to be a balanced weighting between noise annoyance and speech naturalness, for other subjects one of these factors was clearly more decisive than the other. Thus, we hypothesize that whereas some subjects accepted a degradation of speech quality to reach a less-noisy situation, others rejected noise reduction as soon as the reduction was at the cost of speech naturalness. The finding of individual weighting of background noise and speech quality agrees with findings of Versfeld et al. (1999), where individual subjects differed in which factor was most decisive for their overall preference for different types of signal processing and distortion: intelligibility was balanced against clarity, and distortion of the signal was balanced against the amount of added background noise.

The differences between the left and right panels of Figure 3.7 show that several subjects seem to be inconsistent in whether they place more weight on speech naturalness or noise annoyance. The majority of these subjects (subjects 1, 4, 5, and 10) placed more weight on noise annoyance at +4 dB SNR than they did at -4 dB SNR. As shown in Figure 3.3, differences in noise annoyance were larger at +4 dB SNR than at -4 dB SNR, whereas differences in speech naturalness were smaller at +4 dB than at -4 dB SNR. If subjects do not perceive a reduction in speech naturalness, this aspect will no longer play a role in their choice. Therefore, it is not surprising that noise annoyance gets more weighting at +4 dB SNR. However, for half the subjects speech naturalness still played a role in their preference at +4 dB SNR. Thus, it seems that speech naturalness is not completely unaffected at +4 dB SNR. For several subjects (subjects 2, 3, 6, 8, and 10), the small changes in speech naturalness were still too disturbing to fully benefit from the reduction in noise annoyance.

With respect to the effect of SNR, it is remarkable that Hu and Loizou (2007b) found that their normal-hearing listeners placed more emphasis on speech distortion rather than on the noise when judging the overall quality of stimuli at +5 and +10 dB SNR. These authors evaluated different state-of-the-art noise-reduction algorithms, which are not yet implemented in hearing aids. The implementations used in that study were given in Loizou (2007). We applied these algorithms to our noisy speech and listened to these stimuli in comparison to our own hearing aid recordings. It seemed that Loizou's algorithms removed more noise and that they affected the speech quality more than our selection of hearing aid noise-reduction algorithms. It seems that our hearing aid noise reduction was fine-tuned to preserve speech quality, whereas Loizou's algorithms were primarily aimed at reducing the background noise.

Because noise annoyance and speech naturalness were measured using the same procedure as was used to measure overall preference, they were more likely to be correlated to preference than intelligibility and listening effort. For each pair of processing conditions, subjects answered the three questions about noise annoyance, speech naturalness, and overall preference successively, so that the judgment for overall preference could correlate strongly with their previous answers on the speech and noise criteria. Furthermore, all three answers were with regard to the same sentence, whereas noise reduction might act differently on other sentences. To test whether the correlations we found were caused by the experimental design, we shuffled the answers across all the repeats. Thus, we considered whether the values for noise annoyance from the first run combined with that for speech naturalness from the second run could predict the overall preference from the third run, and so on. At the group level, the correlations between overall preference and both other judgment criteria were reduced but still significant. At -4 dB SNR, the group regression coefficient for speech naturalness was slightly higher than the coefficient for noise annoyance, whereas at +4 dB SNR, the noise annoyance was clearly dominating. At an individual level, there were more subjects for whom none of the factors were included in the model (six subjects at -4 dB SNR and two subjects at +4 dB SNR). For the remaining subjects, the effect of SNR was still visible, with more emphasis on speech naturalness at -4 dB SNR and on noise annoyance at +4 dB SNR. From these results, we conclude that the succession of the three questions indeed enhanced the relationship between the overall preference and the noise and speech criteria. However, the effect of SNR on the weighting of the judgment criteria and the differences among subjects in their preferred weighting remained consistent when the successive answers were separated.

3.4.6 Limitations

This study was the first exploration into a largely uninvestigated area, and the conclusions only apply for the limited conditions that we measured.

First, our study population differs from the target population of hearing aid users. The results obtained with our normal-hearing subjects might be representative of listeners with a conductive hearing loss. For listeners with a sensorineural hearing loss, the evaluation becomes more complicated because hearing aids for this type of hearing loss apply dynamic-range compression. The interactions of noise reduction with compression have been studied only occasionally thus far (Chung 2007; Anderson et al. 2009). The results obtained emphasize that the configuration of noise reduction and compression strongly influences the processing. The interactions between noise reduction and compression demand more exhaustive investigations, especially into the complex systems that are currently available for hearing aid users.

In addition, we evaluated just one type of noise and speech combined at a limited number of SNRs. To obtain a more complete impression of the effect of noise reduction, one should investigate a much broader range of speech, noise types, and SNRs (Houben 2011).

Last, we made measurements in a laboratory setting and presented the stimuli to the listeners via headphones. Field studies should reveal whether our results hold in real-life listening situations.

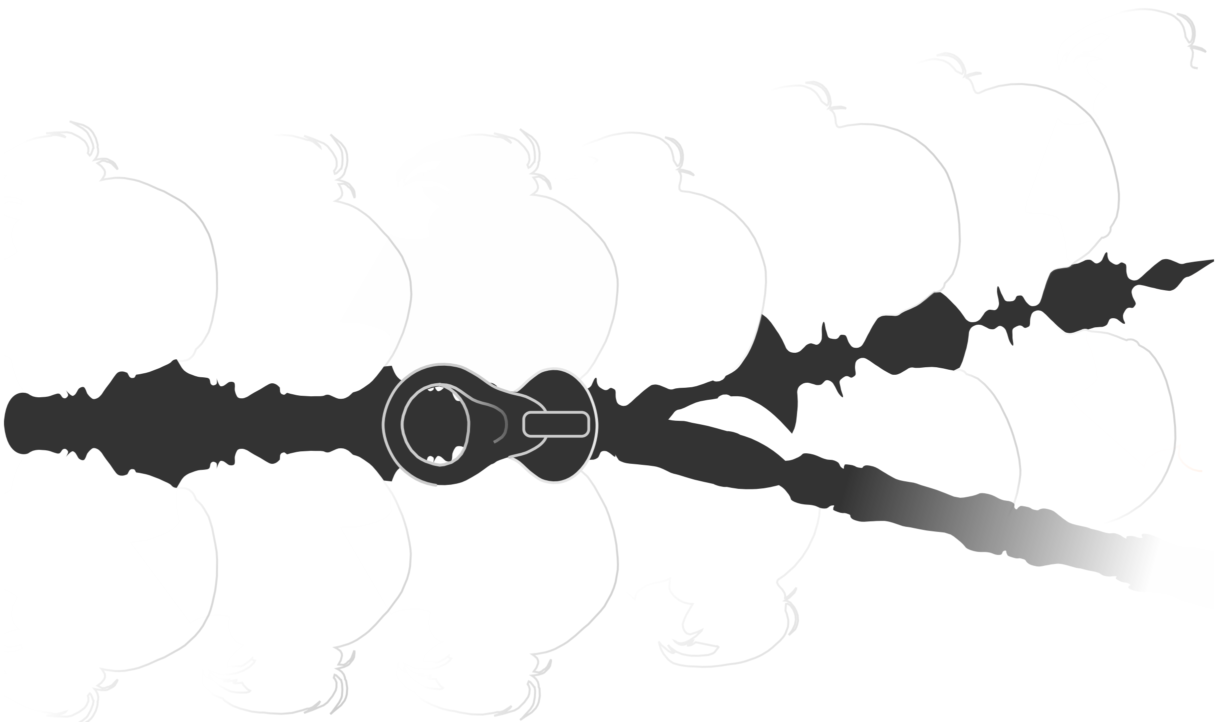
3.5 Conclusions

We conclude that noise reduction differs among hearing aids in the degree that they reduce the noise annoyance and the speech naturalness perceived by normal-hearing listeners. These differences among noise-reduction systems may explain the divergent results of previous studies on the effects of noise reduction on preference and listening effort. Differences in intelligibility were small, as shown in previous noise-reduction studies. Our results imply that it may be useful to give hearing aid users the possibility to compare different noise-reduction systems.

In addition, we conclude that individuals differ in their preferred weighting of noise annoyance and speech naturalness. This finding suggests that listeners may benefit from individualization of noise reduction in hearing aids, and supports the earlier statement that hearing aid users should have the possibility of comparing noise-reduction systems.

Clearly, the next step toward developing guidelines for clinicians to fit noise reduction in hearing aids should include listeners with sensorineural hearing loss, and a set-up to determine the effects of the interaction between noise reduction and compression. If such research would finally lead to fitting rules that help clinicians to actively select the best noise-reduction system and settings for individual listeners, hearing aid satisfaction may increase.

4 Effects of noise reduction on speech intelligibility, listening effort, and personal preference in hearing-impaired listeners



Inge Brons, Rolph Houben, Wouter A. Dreschler

Submitted for publication

4.1 Introduction

Single-microphone noise reduction is a common feature in modern hearing aids that should determine whether the input signal is contaminated with noise and then adjust the hearing aids gain in specific frequency bands to suppress unwanted background noise. In Chapter 3, we directly compared noise reduction from different hearing aids perceptually with each other. For normal-hearing subjects, noise-reduction algorithms appeared to differ perceptually between hearing aids. In this follow-up study, we investigated whether these findings also hold true for hearing-impaired listeners:

Q1. Does hearing aid noise reduction influence speech intelligibility, listening effort, noise annoyance, speech naturalness, and preference for listeners with a moderate sensorineural hearing loss, compared with (a) no noise reduction and (b) noise reduction from other linearly fitted hearing aids?

In this phase, we evaluated noise reduction in isolation without the influence of dynamic-range compression.

4.2 Methods

The methods for hearing aid recording, perceptual measurements, and statistical analyses were identical to those described in Chapter 3.

4.2.1 Hearing aid recordings

We recorded hearing aid output of three linearly-fitted hearing aids from different brands (Phonak Exélia M, ReSound Azure AZ80-DVI, and Widex Mind 440) using the method described in Chapter 2. Acoustical analyses of the noise-reduction processing of these hearing aids are given in Chapter 3. Recordings of the three hearing aids with noise reduction activated were randomly coded as conditions NR1, NR2, and NR3. For one hearing aid we also recorded the output when noise reduction was inactive, resulting in an "unprocessed" condition, representing all hearing aids with noise reduction inactivated.

Stimuli consisted of Dutch sentences in babble noise, recorded with an input noise level of 65 dB(A). Stimuli were presented monaurally to the subjects with Sennheiser HDA200 headphones. The noise level was 65 dB(A) for all the stimuli in the unprocessed condition. Additional amplification was applied according to the linear NAL-RP prescription (Byrne et al. 1991) to compensate for listeners individual hearing loss.

4.2.2 Subjects

Twenty hearing-impaired subjects between 48 and 69 years of age (average = 61.3 years) participated in this study. The subjects audiograms were similar (i.e., no more than 10 dB difference at octave frequencies) to audiogram type N3 (moderate hearing loss with moderate slope) in the set of standard audiograms proposed by Bisgaard et al. (2010). All outcomes were measured at both subjects' individual average SRT_{50} and at a fixed SNR of +4 dB.

4.2.3 Intelligibility

Following the adaptive procedure described by Plomp and Mimpen (1979), we measured Speech Reception Thresholds in noise (SRT_{50}). At the fixed SNR of +4 dB, we measured the percentage of words correctly repeated, similar to the procedure used in Chapter 3, but at higher SNR. Both measurements started with 13 training sentences followed by one list of 13 sentences per processing condition.

4.2.4 Listening effort

The subjects rated the listening effort on a nine-point rating scale that ranged from "no effort" to "extremely high effort". The subjects gave ratings for the four processing conditions at the SRT_{50} level and at -4, +4, and +10 dB SNR.

4.2.5 Paired-comparison rating

We used paired-comparison rating to measure noise annoyance, speech naturalness, and overall preference. For each combination of processing conditions, subjects indicated which was best on each of the three criteria. All six combinations of conditions were measured three times, both at individual SRT_{50} level and at +4 dB SNR.

4.3 Results

4.3.1 Intelligibility

Figure 4.1 shows the group results for speech intelligibility. A repeated-measures ANOVA on the SRT_{50} results (left panel) revealed no significant effect of processing condition ($F[3,57] = 1.0$, $p = 0.39$). The other outcomes were measured at the individually averaged SRT_{50} rounded to whole decibels, ranging from -1 to +4 dB. A repeated measures ANOVA on the rau-transformed percentages (right panel) revealed no significant effect of processing condition ($F[3,57] = 2.3$, $p = 0.085$). However, pairwise comparisons with Bonferroni correction for six comparisons showed a significant difference between unprocessed and NR2 (uncorrected $p = 0.005$).

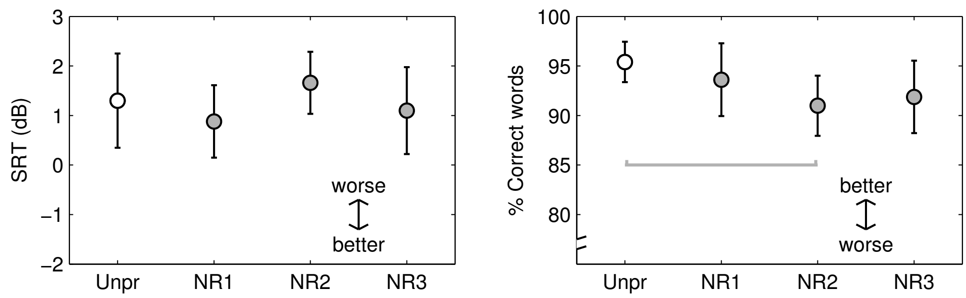


Figure 4.1: Mean and 95% confidence interval of the SRT_{50} (left panel) and of the percentage of words correctly repeated by the subjects at +4 dB SNR (right panel). "Unpr" is the unprocessed reference condition and NR1, NR2, NR3 are the hearing aid noise reductions. Horizontal bars indicate which processing conditions differ significantly from each other after Bonferroni correction for six comparisons.

4.3.2 Listening effort

Figure 4.2 shows the group-average listening-effort ratings relative to that for unprocessed at SRT_{50} level and averaged over the three fixed levels (right panel) and the average absolute ratings for the three fixed SNRs separately (left panel). A repeated-measures ANOVA on the arcsine-transformed data on SRT_{50} level showed a significant effect of processing condition ($F[3,57] = 2.9$, $p = 0.043$), but pairwise comparisons were not significant after Bonferroni correction. The data for the fixed SNRs showed significant effects of SNR ($F[2,38] = 124.3$, $p < 0.001$) and processing condition ($F[3,57] = 2.8$, $p = 0.047$), but not in Bonferroni-corrected pairwise comparisons of processing conditions.

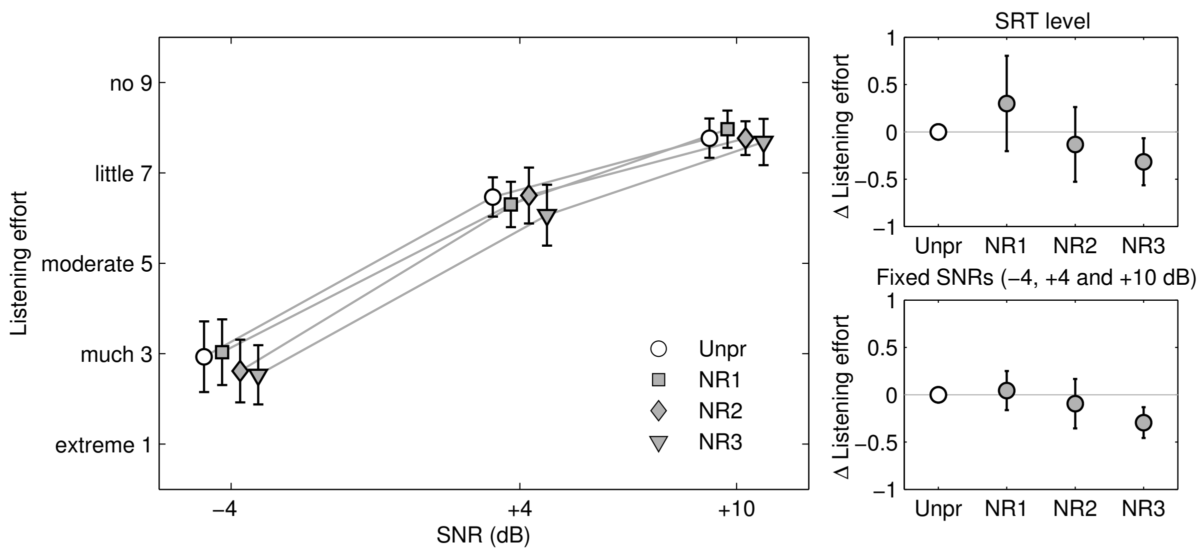


Figure 4.2: Mean and 95% confidence interval of the listening effort ratings assigned by the 20 subjects relative to unprocessed (Δ listening effort, right panels) at SRT_{50} level (upper right panel) and averaged over the three fixed SNRs (lower right panel), and absolute ratings at -4, +4, and +10 dB SNR (left panel).

4.3.3 Paired-comparison rating

Figure 4.3 shows the average rating scores for each processing condition for the three judgment criteria. Scores from -3 to 3 represent the seven categories in the paired-comparison scale. For statistical analysis, we modeled the data using a log-linear modeling approach for ordinal paired comparisons (Dittrich et al. 2004). A repeated-measures ANOVA on the results showed a significant effect of processing condition for all criteria except for speech naturalness at SRT_{50} level. Horizontal lines in Figure 4.3 indicate which processing conditions differed significantly from each other after Bonferroni correction.

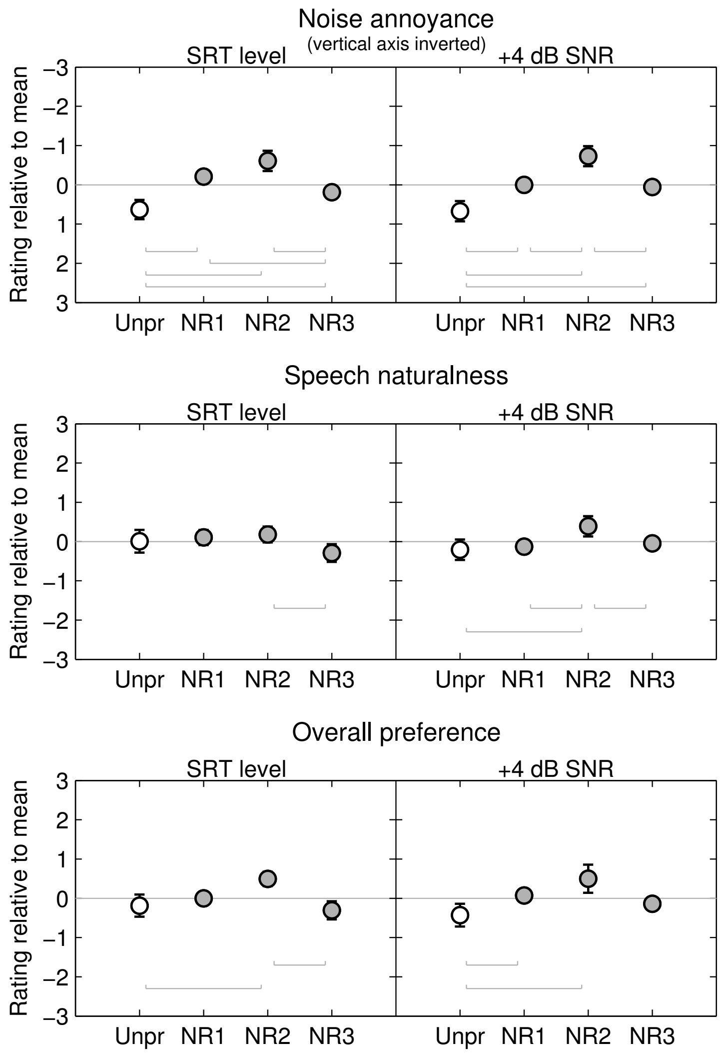


Figure 4.3: Mean rating scores derived from the paired-comparison data for the three judgement criteria and two SNRs. Scores from -3 to +3 were assigned as 0, indicating no difference; -1 and +1 indicating a minor difference; -2 and +2 indicating a moderate difference; and -3 and +3 indicating a major difference. Error bars show the 95% confidence interval among subjects. Horizontal bars indicate which processing conditions differ significantly from each other ($p < 0.05/6 = 0.0083$; Bonferroni-corrected threshold for 6 comparisons).

4.4 Discussion

4.4.1 Intelligibility

Word scores for NR2 were lower than those for unprocessed. In the results of Chapter 3, NR2 also had the lowest word score, although it was not significantly lower than unprocessed. Most studies found no effect of noise reduction on speech intelligibility (Nordrum et al. 2006). Results of Hu and Loizou (2007a) suggest that noise reduction reduces intelligibility more at lower SNRs. Our results do not support this; we measured no effect of noise reduction at lower SNR (SRT_{50}) and a small negative effect at positive SNR. Hilkhuysen et al. (2012) found no interaction between noise reduction and SNR. However, they did not take measurements at positive SNRs.

4.4.2 Listening effort

We found no significant effect of noise reduction on listening effort. The ranking of conditions was equal for the SRT_{50} level and fixed SNR (see Figure 4.2). The same ranking was obtained previously for normal-hearing listeners (Chapter 3). Although not statistically significant, this finding suggests that the results for hearing-impaired listeners agree with those of normal-hearing subjects.

4.4.3 Noise annoyance, speech naturalness, and overall preference

Hearing-impaired listeners indicated differences in noise annoyance, speech naturalness, and overall preference between the conditions of noise reduction on and off and between noise-reduction algorithms of different linear hearing aids. Results at +4 dB SNR agreed well with previous results of normal-hearing subjects (Chapter 3), except for speech naturalness. Hearing-impaired subjects rated speech naturalness higher with noise reduction (NR1 and NR2), indicating that they might use other cues to rate naturalness, for instance the absence of noise (Marzinzik 2000).

We repeated the analysis with the subjects divided in two groups based on their SRT_{50} (12 subjects with SRT_{50} -1, 0, or 1 dB and eight subjects with SRT_{50} +2, +3, or +4 dB), and we found that subjects with a low SRT_{50} rated naturalness at SRT_{50} level lower after noise reduction than subjects with a high SRT_{50}. This finding confirms that noise reduction affects speech naturalness more at lower SNRs, as was previously found for normal-hearing subjects.

The condition that was most preferred by the subjects (NR2) also caused the lowest intelligibility scores. This trade-off between quality and intelligibility is inherent to noise reduction (Wang 2008; Chapter 6).

4.4.4 Signal-to-noise ratios for evaluation of noise reduction

Noise-reduction processing depends on the input SNR (Hoetink et al. 2009). We therefore measured not only at an individual SNR for each subject (to ensure an equal performance level for all subjects) but also at a fixed SNR to ensure equal noise-reduction processing over all subjects. Group results were similar between the two independent datasets obtained at SRT_{50} level and at fixed SNR. This implies that, for the (small) range of hearing losses included, the approach of a fixed and individual SNR did not influence the results. This conclusion might not hold for a broader range of hearing losses. In that case, the approach of evaluating both from a listeners perspective (individually adjusted SNR) and a processing perspective (fixed SNR) might be considered, although results from fixed SNRs are easier to interpret because the effects of noise reduction and hearing ability are easier to separate.

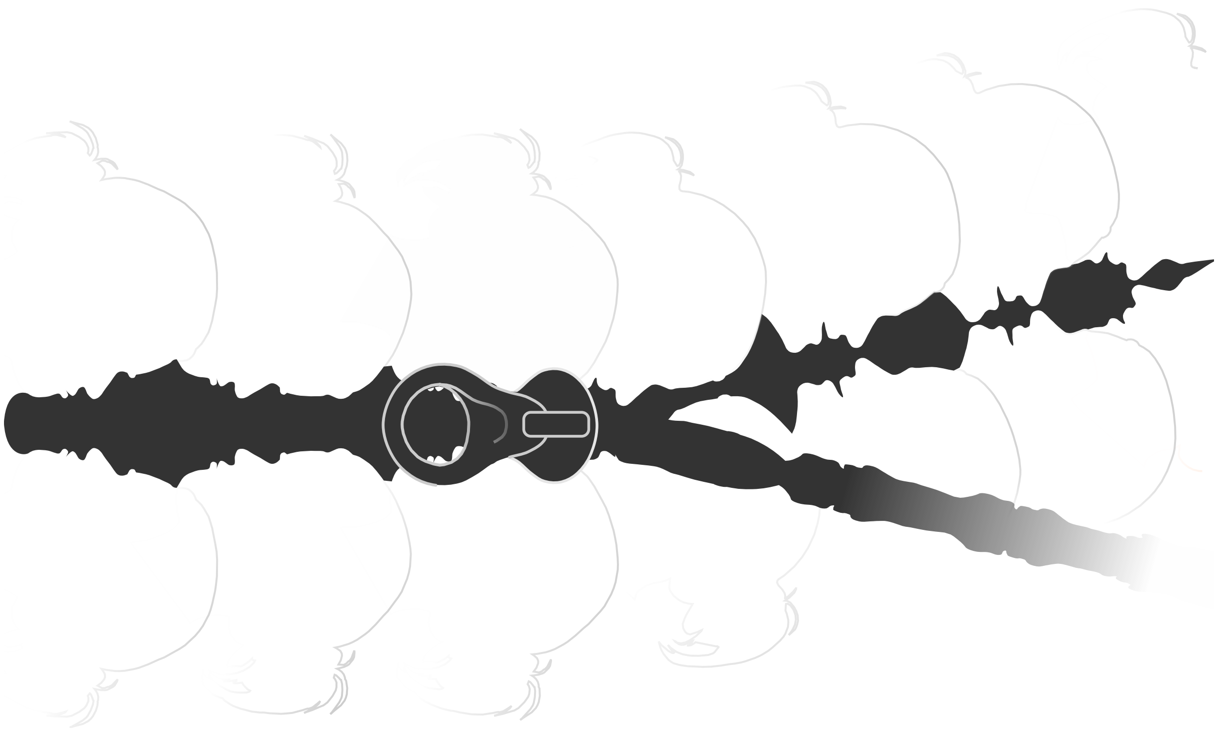
4.4.5 Limitations

The results of this study were measured in a laboratory setting and cannot be generalized beyond the limited number of conditions included. An important factor not included in this study is the fact that most hearing aids apply dynamic-range compression, which might interact with the effects of noise reduction (Chung 2007; Anderson et al. 2009).

4.5 Conclusions

Noise reduction from all three hearing aids tested was able to reduce the annoyance of babble noise perceived by listeners with moderate sensorineural hearing loss. The noise reduction that reduced noise annoyance the most and that was most preferred caused poorer intelligibility scores, confirming a trade-off between listening comfort and intelligibility. The results of hearing-impaired subjects agree well with those of normal-hearing listeners in a previous study. Subsequent experiments should reveal how dynamic-range compression influences the effects of noise reduction.

5 Acoustical and perceptual comparison of noise reduction and compression in hearing aids



Inge Brons, Rolph Houben, Wouter A. Dreschler

To be submitted for publication

5.1 Introduction

Noise reduction and dynamic-range compression are common features in modern hearing aids. The role of noise reduction is to reduce hearing aid gain for background noises and to preserve the gain for incoming speech signals (Bentler and Chiou 2006). The role of compression is to adjust the hearing aid gain based on the input level to fit all incoming signals into the restricted dynamic range of the hearing aid user (Dillon 2001). Compression results in more amplification of low input levels and less amplification of high input levels. Studies investigating the perceptual effects of noise reduction and compression have provided inconsistent results. In general, both the features do not appear to or only slightly influence speech intelligibility in noise; however, both noise reduction and compression can provide benefit in terms of listening comfort and effort. The optimal settings for intelligibility may however differ from the optimal settings for sound quality, both for noise reduction and compression (Dillon 2001; Anderson et al. 2009; Chapters 4 and 6).

Compared with compression, only limited literature is available on noise reduction in hearing aids. This may be because noise reduction in hearing aids is commonly presented as a "black box" i.e., there is no information on the details of signal processing. While compression ratios of a hearing aid can be manually adjusted as a function of frequency, noise reduction can only be influenced by switching an unknown algorithm *on* or *off*, with in some hearing aids a few options for processing strength (e.g., "mild", "moderate", or "strong"). Because of the broad application of noise reduction in hearing aids, these relatively large uncertainties about its implementation are unexpected and undesired.

Although noise reduction and compression are generally applied together, literature on possible interactions between them is scarce (Chung et al. 2007; Anderson et al. 2009). Because noise reduction should reduce the noise level and compression should increase low-level sounds, it is possible that both the features counteract each other, based on their configuration and on the level of background noise. Chung (2007) found that in some hearing aids, the combination of noise reduction and compression caused lesser reduction of noise level than noise reduction alone (i.e., in a linear setting). However, in other hearing aids, compression did not or even positively influence the effect of noise reduction on noise level. Therefore, implementation of noise reduction and compression differ largely between hearing aids, making it difficult to investigate their effects systematically and to draw conclusions that can be generalized across hearing aids.

In the current study, we evaluated the combination of noise reduction and compression in a small set of hearing aids in three subsequent steps. In Experiment 1, we performed acoustical analyses to measure how compression and noise reduction influenced the gain in four different hearing aids. In Experiment 2, we determined whether changes in gain caused by noise reduction and compression were audible in hearing-impaired listeners using three of the above hearing aids. We looked at differences between conditions *within* hearing aids as well as at differences in identical conditions *between* hearing aids. Comparisons between hearing aids were possible because we corrected the differences in frequency response between hearing aids by using inverse filters (see Chapter 2). In Experiment 3, we determined whether combined processing of noise reduction and compression influenced intelligibility and preference compared with (a) no processing (i.e., linear gain only) and (b) combined processing in other hearing aids.

5.2 Experiment 1: Acoustical evaluation

Because most manufacturers do not provide details on the implementation of compression and noise reduction, we can use acoustical analyses to gain an insight into how these features react to speech in noise signals. Therefore, we analyzed recordings of four different hearing aids to answer the following question:

Q1 : How do noise reduction and compression, separately as well as in combination and in different hearing aids, react to speech in babble noise at an input SNR of +4 dB in terms of (a) dynamic gain patterns and (b) change in the overall speech and noise levels?

5.2.1 Methods

Hearing aid recordings

The study included four brands of frequently used behind-the-ear hearing aids (Phonak Exélia M, ReSound Azure AZ80-DVI, Starkey Destiny 1200, and Widex Mind 440) that were randomly assigned a number (from HA1 to HA4). This designation was same as that employed in Chapters 3 and 4, but differed from that employed in Chapter 2.

For linear conditions, we applied the same methods of hearing aid fitting and recording as those described in Chapter 2. We took the hearing aids, hearing aid settings, and equalization filters from the test set in Chapter 2 so that we were sure that the verification of the method was also applicable to the new recordings. All the signal-processing features in the hearing aids (directionality, feedback control, noise reduction, compression, frequency transposition, etc.) were turned *off*, and the frequency-gain patterns of the hearing aids were carefully adjusted to obtain a linear gain that was the same for all hearing aids.

For compression conditions, we fitted the same hearing aids so that their insertion gain, as determined on a B&K Head and Torso Simulator (HATS Type 4128C), was in accordance with the NAL-NL1 prescription for the audiogram coded as N3 from the set of standard audiograms proposed by Bisgaard et al. (2010).

We recorded all sound signals in four different conditions per hearing aid: linear with noise reduction turned *off* ("Unprocessed" condition), linear with noise reduction turned *on* (NR), compressive with noise reduction turned *off* (C), and compressive with noise reduction turned *on* (CNR).

We designed inverse filters for each hearing aid to remove differences in frequency response that remained among the hearing aids (see Chapter 2). Two filters were designed for each hearing aid: one to correct the differences in gain in the linear setting and another to correct the differences in gain in the compressive setting. After filtering, the spectra of recordings of stationary noise with an input level of 65 dB SPL were equal for conditions with noise reduction turned *off*, both for the linear and compressive hearing aids. For linear hearing aids, the inverse filter also corrected the differences in gain between hearing aids at other input levels. However, for compressive hearing aids, only the gain for an input level of 65 dB SPL was equalized. Because of differences in compression ratios as a function of frequency, the gain of compressive aids for other input levels may have slightly differed between hearing aids. These remaining differences in gain between compressive hearing aids are shown in Figure 5.1. This figure shows an effective gain (i.e., the difference between the original input signal and the filtered recording) for stationary noise with input levels of 55 dB SPL (upper curves), 65 dB SPL (middle curves) and 80 dB SPL (lower curves) for hearing aids HA1-HA4. The target that is plotted is the NAL-NL1 prescription for the N3 audiogram. While the gain curves for 65 dB SPL input level agree well between hearing aids caused by the inverse filter, the gain for other input levels differs due to limitations in hearing aid fine-tuning to compensate for detailed differences in amplitude compression. For instance, hearing aid HA3 applied compression for frequencies below 500 Hz although linear gain was prescribed for those frequencies. HA3 did not allow a change in low-frequency compression without affecting the compression at higher frequencies.

In addition to the inverse filter, we bandpass filtered the recordings to limit the response to frequencies ranging from 100 Hz to 5800 Hz.

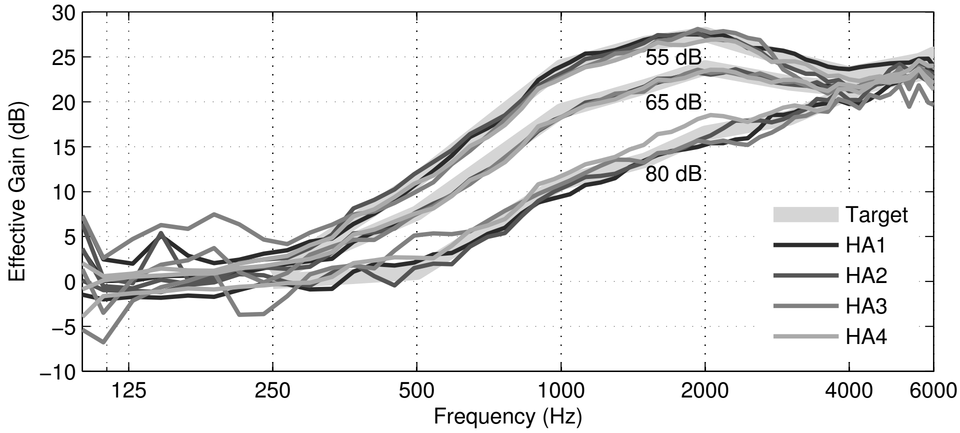


Figure 5.1: Hearing aid gain after filtering for each hearing aid (from HA1 to HA4) for input levels of 55 dB SPL (upper curves), 65 dB SPL (middle curves) and 80 dB SPL (lower curves). The thick lines show the NAL-NL1 prescription for a moderate sensorineural hearing loss according to the N3 audiogram as defined by Bisgaard et al. (2010).

Sound signals

We made hearing aid recordings of female speech from the Dutch VU98 speech material (Versfeld et al. 2000) in a multitalker babble noise. The signals were presented to each hearing aid with an average noise level of 65 dB (A) and an SNR of +4 dB. The babble noise was played continuously while the speech was paused for approximately 1 s between sentences. One list (approximately 36 s) preceded the list that was used for analysis in each condition to allow the noise reduction algorithm in each hearing aid to adapt to the input signals.

To obtain additional acoustical information on the effect of signal processing, we applied the method described by Hagerman and Olofsson (2004). This method allows the calculation of speech and noise levels separately after hearing aid processing by making an additional recording with the noise inverted before it was combined with the speech and presented to the hearing aid. The speech and noise levels can then be separated by taking the sum or the difference of both the recordings.

5.2.2 Results

Figure 5.2 shows the effects of different hearing aid settings on the signal for a recorded sentence. The hearing aid output is plotted as a time signal in gray with the background in dark gray indicating the unprocessed condition (Unpr) and the foreground in light gray indicating the processing condition NR, C, or CNR. The spectrogram-like color plots show the difference between the processing condition (NR, C, or CNR) and the unprocessed condition (Unpr) as a function of time and frequency. The more negative the gain value, the stronger the gain reduction induced by the processing. Thus,

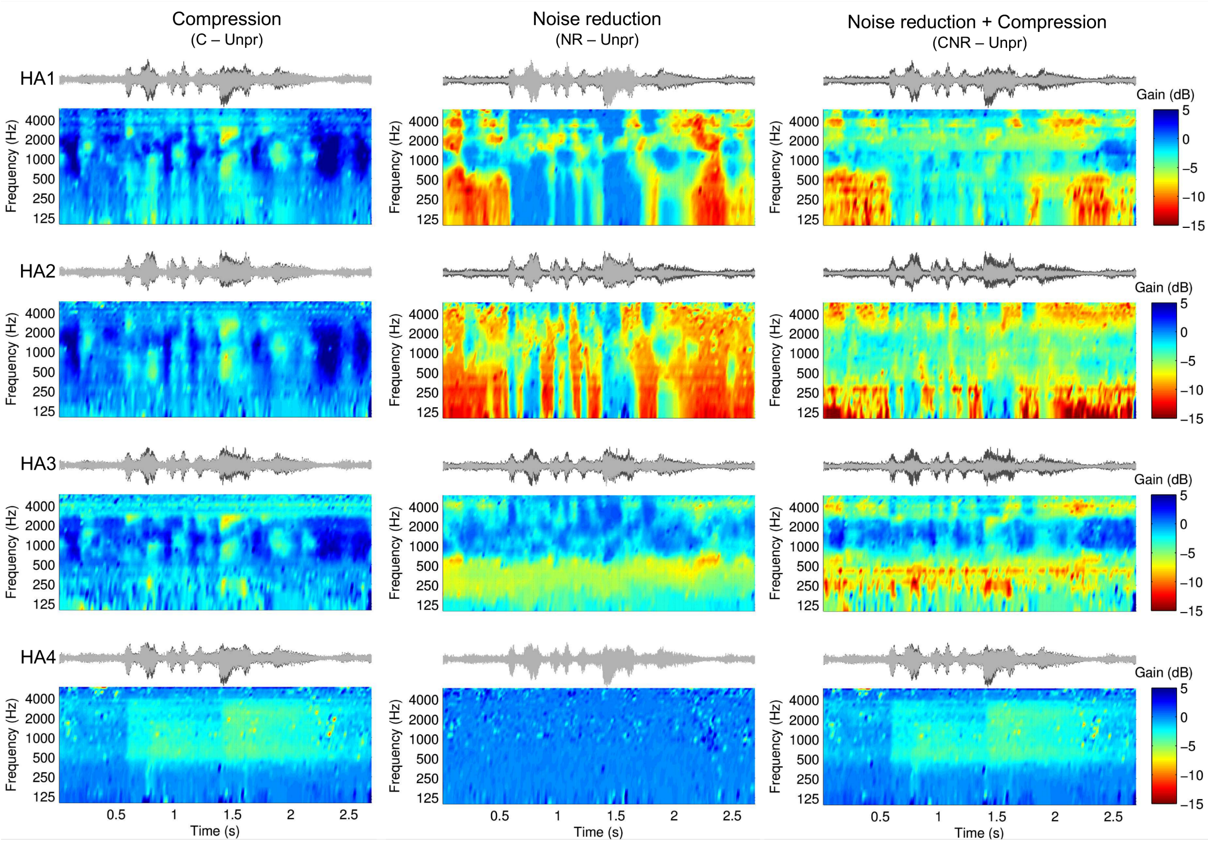


Figure 5.2: Acoustical effects of noise reduction and compression in the four hearing aids tested for a sentence in babble noise with an input SNR of +4 dB. The left column shows the effect of compression (C-Unpr), the middle column the effect of noise reduction (NR-Unpr) and the right column the effect of combined processing of noise reduction and compression (CNR-Unpr). For each processing condition, the time signal of the hearing aid output is shown for the unprocessed condition (dark gray background signal) and for the condition with processing on (C, NR, and CNR respectively for the three columns, the light gray foreground signal). The spectrogram-like color plots show the difference between processed and unprocessed conditions as a function of time and frequency.

red areas correspond to more gain reduction due to processing, whereas blue areas represent no processing or even an increase in gain (dark blue).

Figure 5.3 shows the results obtained with the inversion method of Hagerman and Olofsson (2004). For each processing condition, we extracted the output speech and output noise by respectively adding and subtracting the two recordings with inverse noise from each other. We compared the resulting speech and noise levels after processing with those obtained in the same way for the unprocessed (Unpr) condition. Figure 5.3 shows level reductions in speech and noise signals separately for each processing condition, averaged over 13 sentences.

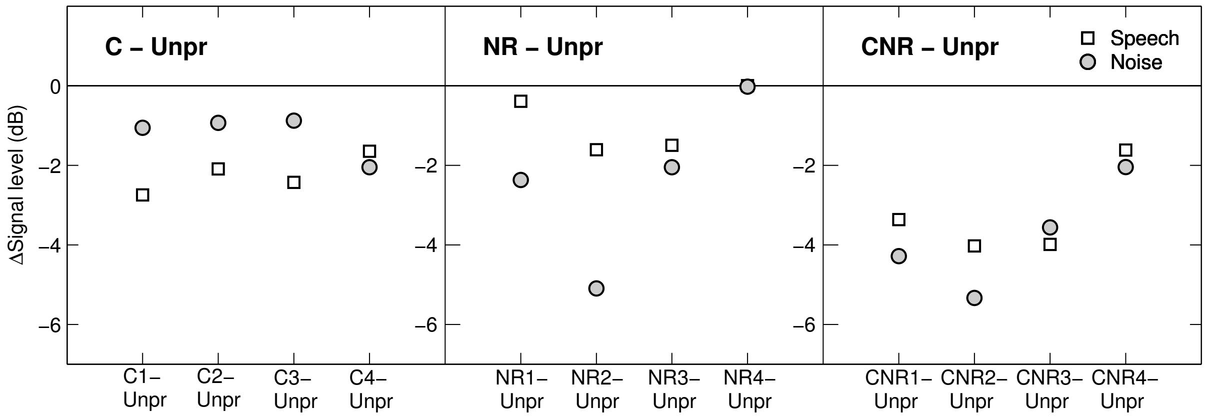


Figure 5.3: Average reduction in speech and noise levels due to the processing of the speech in babble noise at an input SNR of +4 dB The left panel shows the effect of compression (C), the middle panel shows the effect of noise reduction in a linear setting (NR), and the right panel shows the effect of the combination of noise reduction and compression (CNR), all relative to the unprocessed condition (Unpr).

5.2.3 Discussion

The processing of noise reduction (middle column in Figure 5.2; NR-Unpr) differed between the hearing aids in gain depth and dynamics, which was previously shown in Chapter 3 (Figure 3.1). NR4 did not change the gain at this input SNR. All other noise reduction algorithms reduced the noise level more than the speech level (Figure 5.3). Although this implies that noise reduction improves the SNR, it does not necessarily mean that noise reduction also improves intelligibility (Chung 2007). For instance, the sections of the noise that were removed were not necessarily the sections that masked the speech. In addition, the audibility or quality of speech may have been affected, as indicated by the reduced overall speech level. However, Wu and Stangl (2013) found that the changes in SNR as calculated with the method of Hagerman and Olofsson (2004) were related to the changes in acceptable noise level (ANL) of subjects. Thus, although changes in the estimated output SNR cannot directly be translated to an improvement in intelligibility, they may point at other perceived benefits of processing (see Experiment 3).

The changes in gain pattern caused by compression (C-Unpr; left column in Figure 5.2) were similar between hearing aids, except C4, which appeared to have much longer release times than other compression conditions (i.e., the compression reacted relatively slowly to a decrease in input level). Figure 5.3 shows that compression in hearing aids HA1-HA3 reduced speech level more than noise level, which is common for positive input SNR because the speech level was higher than the noise level; thus, speech was less amplified than noise (Hagerman and Olofsson 2004; Rhebergen et al. 2009; Wu and Stangl 2013).

The combined processing of noise reduction and compression (CNR) resulted in the strongest reduction in signal levels, except in hearing aid HA4. Because noise reduction reduced gain in noise and compression during speech, their combined effect on speech and noise showed a relatively constant reduction in gain (Figure 5.2, particularly for CNR1 and CNR2). Although it is imaginable that noise reduction and compression could cancel each other's effect on the gain (e.g., when compression increased the noise level decreased by noise reduction (Chung 2007; Anderson et al. 2009)) it was not the case in the hearing aids tested.

Wu and Stangl (2013) also used the Hagerman and Olofsson (2004) method to evaluate signal processing in hearing aids. They used a different hearing aid; however, within that hearing aid, they evaluated the same conditions as we did (unprocessed, NR, C, and CNR). With the hearing aid and speech in speech-shaped stationary noise at an SNR of +5 dB, they found the same ranking of conditions in their output SNR (note that the change in output SNR in our hearing aids can be derived from the difference between speech and noise levels in Figure 5.3). Activating noise reduction in the linear setting improved the SNR, whereas compression reduced SNR compared with linear gain. When noise reduction was activated in the compressive setting, the negative effect of compression on SNR was offset by noise reduction; however, the output SNR did not reach the same level as that in the linear noise-reduction setting (Wu and Stangl 2013).

5.3 Experiment 2: Detectability of differences

Acoustical measurements showed different strategies for noise reduction and compression between hearing aids. However, it is not known whether these differences are perceptually relevant. As a first step to investigate this, we determined whether the differences could be detected by hearing-impaired listeners for different combinations of conditions *within* as well as *between* hearing aids. To limit the number of conditions, we left out hearing aid HA4 because the spectrogram did not show any changes in gain due to noise reduction in this hearing aids at an SNR of +4 dB.

We designed a listening experiment to determine the percentage of correct detection of differences between conditions to answer the following questions:

Q2. Can hearing-impaired listeners distinguish between combined processing of noise reduction and compression and (a) no processing (CNR compared with unprocessed *within* each hearing aid) and (b) combined processing of other hearing aids (CNR compared *between* hearing aids)?

Q3. Can hearing-impaired listeners detect the effect of noise reduction equally well within a linearly fitted hearing aid (NR compared with unprocessed *within* each hearing aid) as that within a hearing aid fitted with compression (CNR compared with C *within* each hearing aid)?

Q4. Are there any audible differences between hearing aids that are fitted according to the same compressive fitting rule (NAL-NL1) after correction for differences in their frequency response at an input level of 65 dB SPL (C compared *between* hearing aids)?

Questions Q2 and Q3 were meant to provide insight on the effects of noise-reduction processing in compressive hearing aids. Question Q4 was meant to investigate whether careful hearing aid fitting and inverse filtering could remove audible differences between compressions in the tested hearing aids.

5.3.1 Methods

Hearing aid recordings

We used the recordings of speech in babble noise with an input SNR of +4 dB from hearing aids HA1-HA3, which were also used in Experiment 1.

Subjects

Twenty listeners with a moderate sensorineural hearing loss who participated in the study in Chapter 4 were invited for a second visit. Of these, 16 participated in this study. The four remaining subjects did not participate for personal reasons. We used the original audiograms of the subjects because their hearing loss had not deteriorated by $\geq$10 dB in the intervening period (approximately 10 months), determined during the second visit by measuring their air-conduction hearing thresholds of 0.5, 1, 2, and 4 kHz for the ear included. To use the same compression ratios for all participants, only subjects with an audiogram resembling that of the standard audiogram N3 (Bisgaard et al. 2010) were included. Figure 5.4 shows the hearing thresholds for the ears included in the experiment averaged over all the subjects, the corresponding standard deviations, and the standard audiogram N3.

Amplification

The NAL-NL1 prescription rule yields slight differences in the prescribed gain and compression characteristics for individual subjects. The differences between prescribed compression ratios in individual subjects appeared to be small because subjects with audiograms very close resembling the N3 audiogram were selected (Bisgaard et al. 2010; also see the group-average audiogram in Figure 5.4). Therefore, we could use the recordings made for Experiment 1.

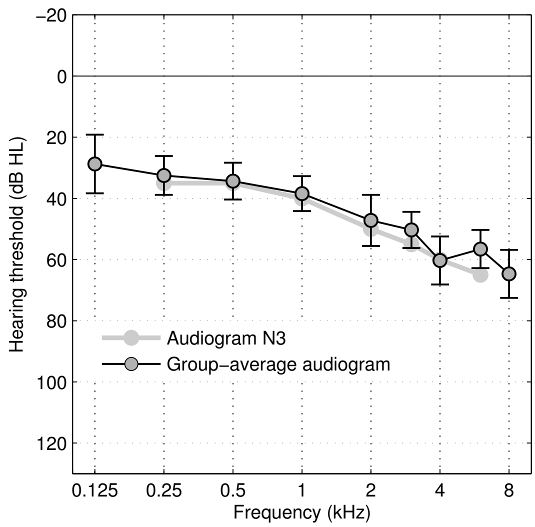


Figure 5.4: The average hearing thresholds of the 16 subjects (one ear per subject) Error bars show the standard deviation. In gray the standard audiogram N3 is shown.

The differences in prescribed gain for an input level of 65 dB SPL differed more between individual subjects than relative differences in gain between input levels caused by different compression ratios (i.e., the difference in gain between input levels of 65 dB SPL and 80 and 55 dB SPL). Therefore, we applied individual inverse filters for each subject. In terms of the gain spectra of Figure 5.1, this means that the amplification profile for the input level of 65 dB SPL was individualized for each subject and that the amplification profiles for 55 and 80 dB SPL were derived from that at 65 dB SPL, with the standard compression ratio from profile N3.

Stimuli

Stimuli were taken from the recorded VU98 lists and included single sentences with 0.5 s of noise before and after the sentences. The stimuli were presented monaurally with Sennheiser HDA200 headphones, which had been calibrated with a B&K Artificial Ear Type 4153. The noise level was 65 dB (A) (and the average speech level 69 dB(A) to realize an SNR of +4 dB) for stimuli in the unprocessed condition if no NAL-NL1 amplification was applied.

Measurement procedure

We used a similar detection task as that used in Chapter 2. In brief, all combinations of conditions required to answer our questions (Q2-Q4) were used in an oddball paradigm where the subjects' task was to select from three stimuli which one sounded differently from the other two. Participants were allowed to listen to the fragments as often as they preferred before they responded. Directly after their responses, participants received a feedback on which stimulus was different from the other two.

In all, 15 distinct stimulus pairs were included (2 x 3 to answer Q2, 2 x 3 to answer Q3, and 3 to answer Q4), and each stimulus pair was tested thrice in AAB and thrice in BBA configuration to provide 90 trials per subject. This was divided in three separate blocks of 30 trials. The blocks alternated with two blocks of intelligibility measurement for Experiment 3.

5.3.2 Results

Figure 5.5 shows the group results divided over four panels for the separate (sub)questions of this experiment. Asterisks in the lower part of each panel indicate the detection rates that differed from chance level (33%) according to one-sided t-tests with Bonferroni correction for three comparisons.

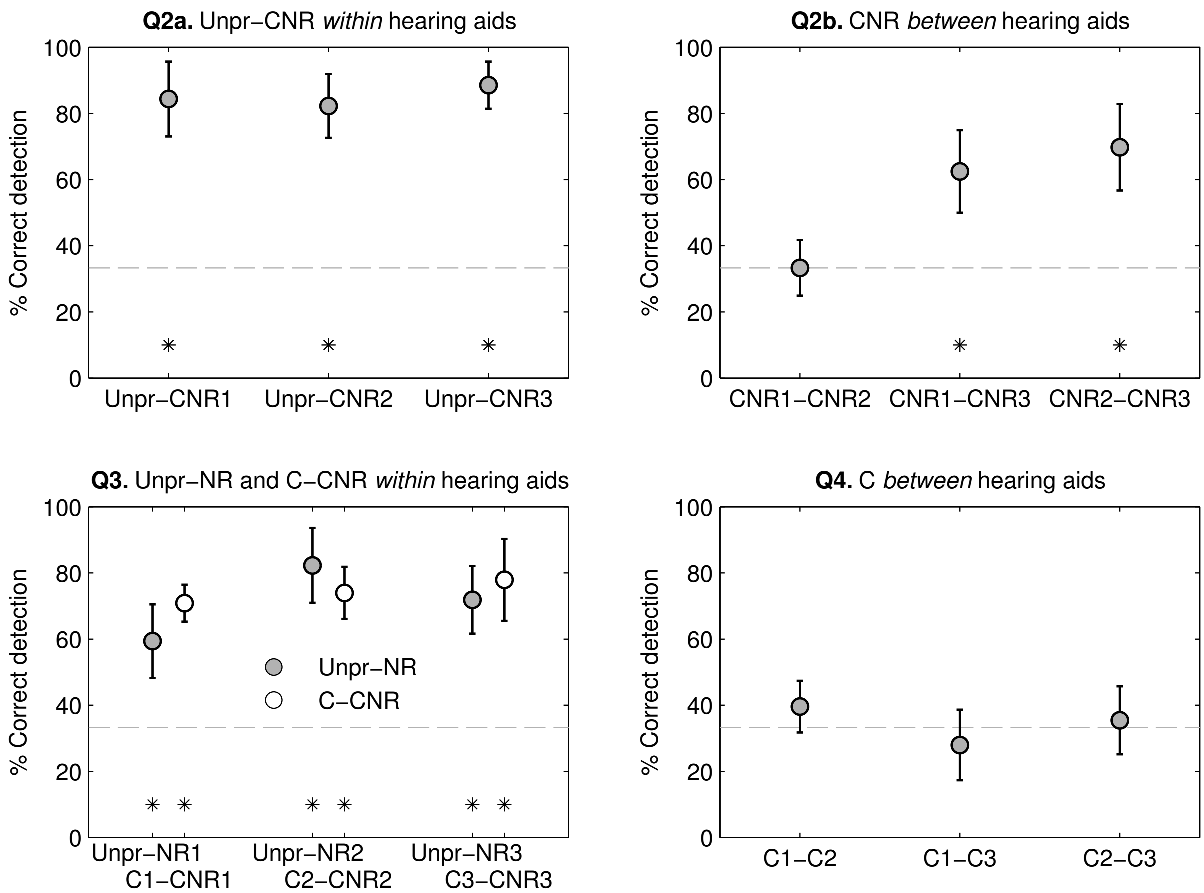


Figure 5.5: Percentage of trials in which the odd stimulus was correctly identified for each combination of conditions and averaged for the 16 subjects. Each panel corresponds to one research (sub)question for this experiment. Error bars show 95% confidence intervals among subjects. Dashed horizontal lines show the chance level (33%). Asterisks indicate the combinations for which the detection rate was significantly higher than the chance level. Unpr: unprocessed, NR: noise reduction in a linear setting; C: compression; CNR: noise reduction in a compressive setting.

The upper panels show that the detection rate for combined processing compared with unprocessed within all three hearing aids was significantly higher than the chance rate (Q2a). Between hearing aids, the detection of differences in combined processing was above chance level for both combinations with CNR3, but not for the combination of CNR1 and CNR2 (Q2b).

To answer Q3, we compared the detection rate for NR with linear amplification (i.e., the difference between Unpr and NR; Figure 5.5, bottom left graph) with that for NR with compression (i.e., the difference between C and CNR; Figure 5.5, bottom left graph) for each hearing aid. This was done by using two-sided t-tests with Bonferroni correction for three comparisons. The detection percentages for the effect of noise reduction within a linear setting did not differ from those for noise reduction in a compressive setting within each hearing aid (uncorrected p-values: $p = 0.04$ for HA1, $p = 0.20$ for HA2, and $p = 0.15$ for HA3).

The bottom right panel shows that the participants could not differentiate between compression in the different hearing aids (all detection rates were at chance level).

5.3.3 Discussion

Combined processing of noise reduction and compression (Q2)

Q2a: Hearing-impaired listeners were able to distinguish combined noise reduction and compression processing (CNR) from unprocessed (i.e., linear amplification) in all the three hearing aids (see upper left panel in Figure 5.5). Thus, combined processing of noise reduction and compression changed the gain sufficiently to be audible. Detection percentages for the combined effect of noise reduction and compression were higher than those for the effect of noise reduction alone (Figure 5.5; percentages for Unpr-CNR in the top left panel were higher than those for Unpr-NR in the bottom left panel). These results agree well with the acoustical results from Experiment 1. Figure 5.3 shows that the CNR conditions changed speech and noise levels more than the NR conditions did. Note that higher detection percentages do not necessarily imply more benefit of the processing conditions. This will be investigated in Experiment 3.

Q2b: The results show that hearing-impaired listeners were able to distinguish between CNR conditions in different hearing aids. The upper right panel of Figure 5.5 indicates that the combined processing of noise reduction and compression (CNR) in HA3 could be distinguished from that in HA1 and HA2, which were not discernible from each other. The acoustical results (Figure 5.2) showed that CNR3 had a higher contrast between the lower (250-500 Hz) and higher (1000-2000 Hz) frequencies and that this did not change in the presence of speech. In contrast, CNR1 and CNR2 resulted in a higher reduction in gain during the absence of speech. This may explain the higher detection rate for CNR3.

Noise-reduction effect in linear and compressive aids (Q3)
Hearing-impaired listeners could detect the effect of noise reduction equally well in linearly fitted hearing aids as in hearing aids fitted with compression (Figure 5.5, bottom left panel). Although this finding does not prove that the perceptual effect of noise reduction was the same in linear and compressive settings, it is plausible that compression did not influence noise reduction to a great extent, at least not in the hearing aids and conditions tested in this study.

Detectability of differences in compression between the hearing aids (Q4)
The subjects could not differentiate between compression processing in the three different hearing aids for the speech in noise signals that were used. Thus, for the current set of hearing aids, careful hearing aid fitting and inverse filtering achieved the removal of audible differences between compression in the hearing aids tested. This does not guarantee that it will work for all compressive hearing aids (see Chapter 2). The equalization filter cannot correct the differences in attack and release times, which may cause audible differences between hearing aids even if they are fitted with equal compression ratios. For instance, this could be the case for HA4 that was not included in the perceptual measurements but showed different compression characteristics compared with the other three hearing aids in Experiment 1 (Figure 5.2).

In the current study, the fact that the subjects could not detect differences between the three C conditions means that any perceptual differences that were found between the CNR conditions (see Figure 5.5 upper right panel and preference results in Experiment 3) can be attributed to the differences in either noise-reduction processing or interaction between noise reduction and compression.

5.4 Experiment 3: Perceptual effects

The final step in this study was to evaluate the combined effects of noise reduction and compression on speech intelligibility, noise annoyance, speech naturalness, and personal preference. Similar measurements were previously performed in Chapter 4 for the same hearing aids but with linear amplification (i.e., the three NR conditions were compared with each other and with the unprocessed condition). Here we evaluated the three CNR conditions, that were more representative of the application of noise reduction in the hearing aids. We compared the three CNR conditions with each other and with their joint reference condition “Unprocessed,” which was equal for all the three hearing aids. This is the same reference condition as that used in Chapter 4.

The research questions of Experiment 3 were as follows:

Q5. Does the combined processing of noise reduction and compression influence speech intelligibility in babble noise compared with (a) no processing (CNR compared with unprocessed *within* hearing aids) or compared with (b) the combined processing in other hearing aids (CNR compared *between* hearing aids)?

Q6. Does the combined processing of compression and noise reduction influence listeners' preference (noise annoyance, speech naturalness, or overall preference) compared with (a) no processing (CNR compared with unprocessed *within* hearing aids) or compared with (b) the combined processing in other hearing aids (CNR compared *between* hearing aids)?

5.4.1 Methods

Measurements were done during the same visit as those for Experiment 2. Hearing aid recordings, subjects, stimuli, and amplification were the same as described before. We used recordings from four different hearing aid conditions: CNR conditions in HA1-HA3 and unprocessed condition in HA1, which did not perceptually differ from the unprocessed condition in the other two hearing aids (see Chapter 2).

Intelligibility

We measured the percentage of words correctly repeated by the subjects at a fixed input SNR of +4 dB for each condition. Each subject started with one list of 13 training sentences, containing all processing conditions. After this training list, we used two lists (test and retest) of 13 sentences per processing condition. Test and retest were separated by a block of 30 trials for the detection task (Experiment 2). We balanced the order of conditions across subjects to minimize the effects of training and fatigue on group data. We balanced the lists across conditions to minimize the effects of differences between lists. We considered the first 3 sentences of each condition as training sentences and used the last 10 sentences to calculate the percentage of correct words.

Paired comparisons

We used paired-comparison rating (a two-interval, seven-alternative forced choice paradigm) to measure noise annoyance, speech naturalness, and overall preference (see Chapter 3 for details). In brief, subjects listened to all the possible combinations of conditions and rated on a seven-point rating scale in which of the two conditions they found the speech to be more natural, in which of the two conditions they found the noise to be less annoying, and which of the two conditions they would prefer for prolonged listening. All subjects performed three runs of 6 comparisons, resulting in 18 comparisons per subject. All the subjects started with two training pairs.

5.4.2 Results

Intelligibility (Q5)

Figure 5.6 shows the percentage of words correctly repeated averaged over all the 16 subjects. For statistical analysis, we transformed these percentages to rationalized arcsine units (Studebaker 1985) and subsequently performed a repeated-measures analysis of variance (ANOVA) on the transformed data, with processing condition as the fixed effect. The effect of processing was not significant (p = 0.09).

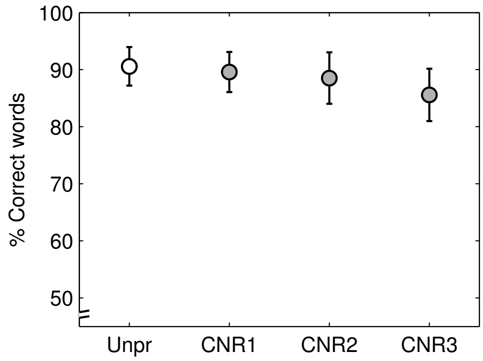


Figure 5.6: Mean and 95% confidence interval of the percentage of words correctly repeated by the subjects at an SNR of +4 dB. The scores for the four different conditions did not differ significantly from each other. Unpr is the unprocessed reference condition; CNR1, CNR2, and CNR3 are the conditions with noise reduction and compression in the three hearing aids tested.

Paired Comparisons (Q6)

Figure 5.7 shows the average rating scores for each processing condition for the three criteria. We assigned scores from -3 to 3 for each condition, according to the ITU-T recommendation P.800 (1996). The scale for the noise annoyance is inverted in Figure 5.7 so that for each outcome a symbol plotted above the zero line means a better than average performance on that judgment criterion. For statistical analysis we used a log-linear modeling approach for ordinal paired comparisons described by Dittrich et al. (2004) in the same way as that described in Chapter 3. We performed a repeated-measures ANOVA on the worth parameters estimated with that model for each judgment criterion separately (noise annoyance, speech naturalness, and overall preference), with processing condition as the fixed effect. The effect of processing was significant for all the three judgment criteria ($p < 0.001$ for noise annoyance, p = 0.005 for speech naturalness, and p = 0.001 for the overall preference). Horizontal lines in Figure 5.7 indicate the processing conditions that differed significantly from each other in a pairwise comparison with Bonferroni correction for six comparisons.

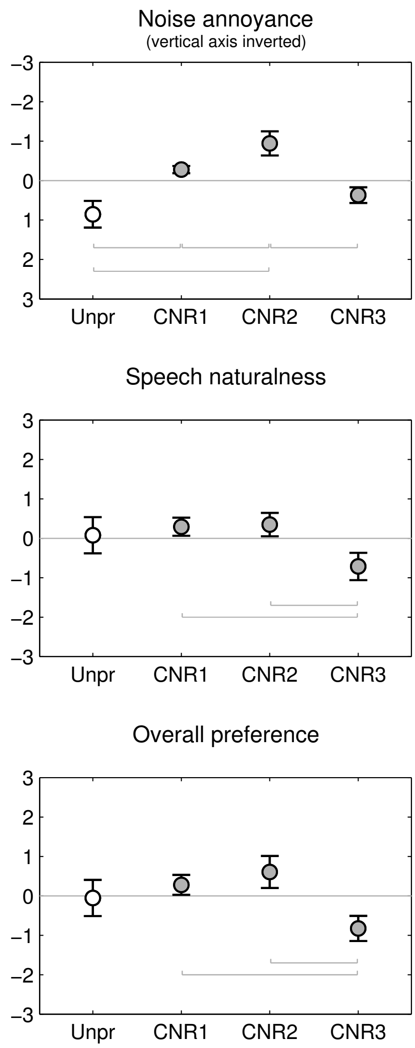


Figure 5.7: Mean rating scores derived from the paired-comparison data for the three judgement criteria. Scores from -3 to +3 were assigned with 0 indicating no difference; -1 and +1 indicating a minor difference; -2 and +2 indicating a moderate difference; and -3 and +3 indicating a major difference. Error bars show the 95% confidence interval among subjects. Horizontal bars indicate which processing conditions differ significantly from each other ($p < 0.05$ after Bonferroni correction for 6 comparisons). Unpr: unprocessed reference condition; CNR: combined processing of noise reduction and compression.

5.4.3 Discussion

Intelligibility (Q5)

The combined processing of noise reduction and compression did not influence speech intelligibility at an SNR of +4 dB compared with no processing or combined processing in other hearing aids. The finding of no effect of noise reduction on intelligibility is common (Nordrum et al. 2006). Most other studies have compared noise reduction *on* and *off* within a compressive hearing aid (i.e., CNR vs. C), whereas we compared noise reduction and compression with linear amplification (CNR vs. Unprocessed). Chung (2007) made the same comparison as we did in two different hearing aids and also found no difference in speech intelligibility scores between linear setting and combined processing of noise reduction and compression. However, one of those hearing aids slightly reduced intelligibility when compression was activated compared with linear processing (both without noise reduction). In that hearing aid, switching the noise reduction *on* in the compressive setting offset the negative effect of compression and brought intelligibility scores back to those of the linear condition. This demonstrates the difficulty in studying the interaction between compression and noise reduction. If Chung had only measured the effect of noise reduction in a compressive hearing aid, she would have found an improvement due to noise reduction. In fact, the underlying cause may just as well be that noise reduction undid the unfavorable gain alteration caused by compression.

In Chapter 4 intelligibility measurements were described for the unprocessed and NR conditions in the same hearing aids. We found in Chapter 4 that intelligibility slightly reduced for hearing aid HA2 due to noise reduction. In the current experiment, the CNR condition of the same hearing aid did not affect intelligibility. Thus, in combination with compression, noise reduction of HA2 seems less deteriorating for speech intelligibility than in combination with linear amplification. This is remarkable because although CNR2 had higher speech intelligibility, it reduced the speech level more than NR2 (Figure 5.3). This might be explained by Figure 5.2, which shows that the dynamic behavior also differed between CNR2 and NR2. Although CNR2 caused a relatively constant reduction in gain during speech (because noise reduction reduced gain mainly during noise and compression during speech for our stimuli), NR2 caused the gain to change quickly to reduce the noise during speech pauses while retaining the speech signal. Thus, in this particular hearing aid, it seems that the quick and rather large changes in gain caused by noise reduction during speech were more detrimental for speech intelligibility than a continuous moderate reduction in gain caused by noise reduction and compression combined.

Noise annoyance, speech naturalness, and overall preference (Q6)

Paired-comparison data showed that the combined processing of noise reduction and compression in HA1 and HA2 (CNR1 and CNR2) reduced noise annoyance compared with unprocessed and CNR3. CNR2 resulted in the strongest reduction in noise annoyance. Speech naturalness was lower for CNR3 than for the three other conditions, which was also observed for the overall preference.

The pattern for noise annoyance agreed well with that of the objectively determined reduction in noise level in Figure 5.3 (right panel, from CNR1 to CNR3). The results for noise annoyance were also comparable with those obtained in Chapter 4 for noise reduction in linear hearing aids (NR conditions of the same three hearing aids), except for HA3 in which noise reduction without compression (NR3) previously reduced noise annoyance but noise reduction with compression (CNR3) in the current experiment did not. This seemed to be inconsistent with the result from Experiment 1, where CNR3 reduced noise level more than NR3 (Figure 5.3). However, the negative judgments for speech naturalness and overall preference for CNR3 (Figure 5.7) show that its gain reduction (which was concentrated in the low frequencies; Figure 5.2) was perceived as unnatural. This unnaturalness may also have caused the subjects to judge the noise in this condition as more annoying.

Although the speech level was reduced in CNR1 and CNR2 (Figure 5.3), the subjects rated speech naturalness for these conditions not differently from unprocessed. However, the individual preference data suggest that subjects differed from each other in their opinion on speech naturalness. Some subjects rated speech naturalness for CNR1 and CNR2 higher than that for Unprocessed, whereas others did the opposite. Although there is no sufficient data for exhaustive analyses by dividing our subjects in groups based on their opinion on naturalness, these observations may form a basis for further investigations into individual differences in preference. In Chapter 3 we found that normal-hearing subjects differed from each other in whether they based their preference mainly on speech naturalness or on noise annoyance. The large individual differences in preference for noise-reduction strength between normal-hearing listeners found by Houben et al. (2012) support this hypothesis. The current findings also suggest the differences between hearing-impaired individuals, in whether they perceive processed speech as more natural (less noise) or less natural (more processing artifacts) than unprocessed speech. Although CNR1 and CNR2 decreased noise annoyance compared with unprocessed, they were not significantly preferred over unprocessed. Within the same hearing aids, significant preferences for NR1 and NR2 over unprocessed were found in Chapters 3 and 4, both for normally hearing and hearing-impaired listeners. Anderson et al. (2009) determined the preference scores for different configurations of compression and noise reduction and found that their NR condition was the most preferred. Their CNR conditions were even less preferred than unprocessed. However, compared with compression alone (C), the CNR conditions were more preferred. Comparable results were recently obtained by Wu and Stangl (2013) who determined the acceptable noise level (ANL) in different processing conditions. Subjects accepted less background noise in the compression condition (C) than unprocessed, but the activation of noise reduction offset this effect. Thus, the results on perceptual measurements with CNR highly depend on whether the reference condition is unprocessed or compressed. The positive effect of CNR is expected to be higher compared with compression than compared with unprocessed.

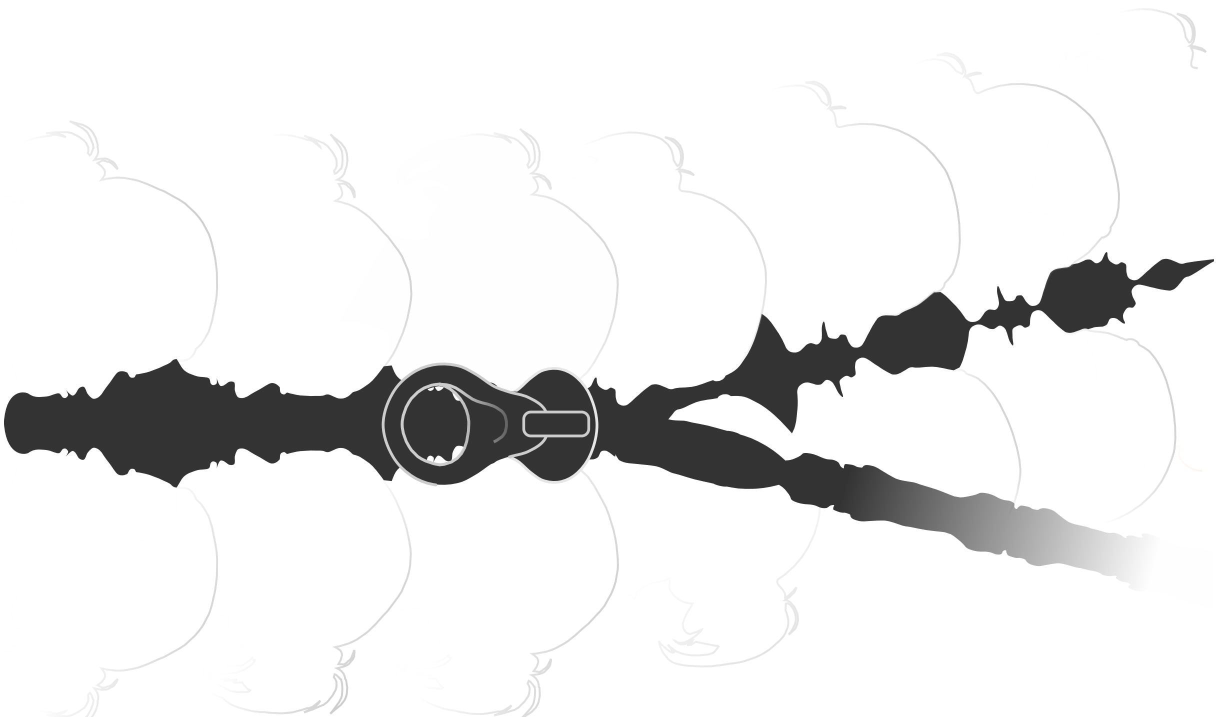
5.5 Conclusions

Acoustical analyses (Experiment 1) of hearing aid recordings of speech in babble noise with an input SNR of +4 dB showed that there are differences between hearing aids in terms of gain changes due to noise reduction, compression, and their combined processing. Combined processing of noise reduction and compression in the four hearing aids tested reduced both the speech and noise levels. This reduction due to combined processing was stronger than that for noise reduction or compression separately, indicating that both features do not cancel each other if combined for the hearing aids tested.

Experiment 2 showed that differences in processing within hearing aids were detectable for hearing-impaired listeners. Between hearing aids, the listeners could not detect differences in compression (C) condition; however, combined with noise reduction (CNR), the processing in one hearing aid was discernable from the other two.

The combined effect of noise reduction and compression did not influence speech intelligibility (Experiment 3). However, the combined processing reduced noise annoyance, which agreed with the reduction in noise levels found in Experiment 1. The reduction in speech levels found in Experiment 1 resulted in a reduction in speech naturalness only for one hearing aid (CNR3). That processing condition (CNR3) was less preferred than the other two and less preferred than unprocessed. Preference for CNR conditions relative to unprocessed seemed to be lower than previously found for the NR conditions, where processing of HA1 and HA2 were preferred over unprocessed. This indicates that the influence of compression should be considered for the development and evaluation of noise reduction algorithms for hearing aid application.

6 Perceptual effects of noise reduction by time-frequency masking of noisy speech



Inge Brons, Rolph Houben, Wouter A. Dreschler

Adapted from *The Journal of the Acoustical Society of America, 2012, 132(4), pp.2690-2699.*

6.1 Introduction

Understanding speech in the presence of background noise is a demanding task. Consequently, there has been extensive research in the development of noise-reduction algorithms. These algorithms should improve user satisfaction with modern communication devices (e.g., hearing aids) in noisy environments. Noise-reduction algorithms should detect and remove the background noise, without affecting the target speech. Many noise-reduction techniques are based on the time-frequency representation of the signal, thus analyzing the content of each frequency channel during each time window. For all of these individual time-frequency units, the signal-to-noise ratio (SNR) is calculated or estimated and is used to determine whether the time-frequency unit will be retained (when speech is dominant) or attenuated (when noise is dominant). The resulting time-frequency attenuation pattern is often referred to as a time-frequency mask.

Numerous studies have examined the effects of time-frequency masking on speech intelligibility in noise. Time-frequency masking has been shown to cause large improvements in intelligibility (Brungart et al. 2006; Wang et al. 2008). This improvement, however, reflects the non-realistic situation that speech and noise are given separately as input (i.e., ideal noise-tracking). In a realistic situation in which the speech and noise are not individually known, the algorithm must estimate the noise without prior knowledge of the input signal. Time-frequency masking then provides no benefit in terms of intelligibility, mainly because of errors in the noise estimation (Loizou and Kim 2011). However, intelligibility is not the only outcome that is relevant for user satisfaction. Therefore, in this study, we evaluated different types of time-frequency masking not only on speech intelligibility but also on listening effort, speech naturalness, noise annoyance, and overall preference. We determined how these outcomes were influenced by two main aspects of noise reduction: (1) the strength of attenuation and (2) the method for noise tracking. We used four different types of time-frequency masking. With the first two conditions, we compared strong (infinite) attenuation with limited attenuation. With the second, third, and fourth conditions, we compared ideal noise tracking and two types of non-ideal noise-tracking.

Our first condition is known as the ideal binary mask (IBM) (Wang 2005). This noise reduction receives speech and noise separately as input so that it does not need to estimate the noise from the mixed signal. The IBM applies a binary pattern of attenuation to the noisy signal. All of the time-frequency units that have a signal-to-noise ratio (SNR) above a specified threshold are preserved, while all of the units with a lower SNR are eliminated. Usually, the threshold is 0 dB SNR. Researchers often use the IBM to investigate the effects of different noise-reduction parameters on intelligibility, independent of noise-estimation errors. For example, the IBM has been used to evaluate the influence of the time- and frequency resolution of noise reduction (Anzalone et al. 2006; Li and Loizou 2008a; Wang et al. 2008), the frequency range on which noise reduction is active (Anzalone et al. 2006; Li and Loizou, 2008a; Wang et al., 2009), the signal-to-noise ratio threshold below which noise reduction is applied (Li and Loizou 2008b; Kjems et al. 2009), and the type of background noise (Li and Loizou 2008b; Kjems et al. 2009).

The IBM has proven to be able to provide approximately 13 dB SNR improvement in speech intelligibility in the presence of noise (Wang et al. 2009). However, the rapid binary attenuation transitions of the IBM can introduce musical noise (Wang 2008). Musical noise has a tonal character, and it occurs because of small isolated peaks that remain in the spectrum after the signal is removed in other time-frequency units (Berouti et al. 1979). In some cases, musical noise can be more disturbing to the listener than the original distortions caused by interfering noise (Loizou 2007).

A method for reducing musical noise is to limit the attenuation so that the noise-dominated time-frequency units will be attenuated but not eliminated (Anzalone et al. 2006). Although limiting the attenuation could improve the sound quality, it could also reduce the potential intelligibility benefit. This hypothesized trade-off between the subjective perception of the sound quality and the objective benefit in terms of speech intelligibility should receive more attention in the evaluation of noise reduction algorithms (Wang 2008). Therefore, our second noise-reduction condition was an ideal mask as well, but with a tempered attenuation function (ideal tempered mask, ITM).

Tempering the attenuation function is especially useful in combination with non-ideal noise estimators, which we used in our third and fourth conditions. Because noise estimators introduce errors in the SNR estimation, applying a binary attenuation function that either retains or completely removes time-frequency units would cause not only musical noise but also additional distortions from estimation errors. Estimation errors can be classified into two types: type I errors occur when time-frequency units are wrongly classified as being speech-dominated, and type II errors occur when units are wrongly classified as being noise-dominated (Li and Loizou 2008b). We used two noise-estimation algorithms for our third and fourth time-frequency masks: for one algorithm, the type I errors dominated (Hu and Loizou 2008), and for the other algorithm, the type II errors dominated. Both of the algorithms were combined with the tempered attenuation function.

To summarize, whereas most of the studies with the IBM concentrate on its effect on intelligibility only, we evaluated this algorithm also on other perceptual aspects (listening effort, noise annoyance, speech naturalness, and overall preference). Additionally, we limited the maximum attenuation to determine whether the expected disadvantages of the IBM can be reduced and to what degree this limited attenuation reduces the advantages in terms of speech intelligibility. Finally, we replaced the ideal noise classifier by real noise estimators to determine whether this realistic noise reduction has perceptual advantages in spite of the expected lack of intelligibility improvement. These three steps resulted in the following research questions:

Q1. How does the IBM influence perception in terms of speech intelligibility, listening effort, noise annoyance, speech naturalness, and overall preference?

Q2. How do the perceptual effects differ between the IBM and an ideal mask with non-binary attenuation limited to a maximum of 10 dB ("ideal tempered mask")?

Q3. How do the perceptual effects of noise reduction differ between noise reduction with and without prior knowledge of the speech and noise signals? ("non-ideal" versus "ideal" masking).

6.2 Methods

This study was performed in parallel with our study to investigate the perceptual effects of noise-reduction algorithms that are implemented in hearing aids. As such, the subjects, the measurement procedures and the statistical methods are identical to those described in Chapter 3.

6.2.1 Subjects

Ten normal-hearing subjects (who were all university students) between 19 and 23 years of age (average = 0.8 years) participated in this study. Their hearing thresholds were 15 dB Hearing Level or better at 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz.

6.2.2 Signal processing

Table 6.1 provides an overview of the five processing conditions. Stimuli for the unprocessed condition were passed through the IBM algorithm with the attenuation set to 0 dB (i.e., no attenuation) for all of the time-frequency units, which corresponded to a linear mask with all ones and no zeros. Thus, the complete speech-in-noise signal was retained in this condition.

Table 6.1: Overview of the differences between the five processing conditions.

Condition	Noise tracker	Attenuation function	Attenuation (dB)
Unprocessed	-	-	-
IBM	Ideal	Binary	0 or ∞ dB
ITM	Ideal	Gradual	0 to 10 dB
MartinTM	Martin	Gradual	0 to 10 dB
MCRA2TM	MCRA2	Gradual	0 to 10 dB

The MATLAB implementation of the IBM algorithm used in this study was provided by Loizou and was previously used in Li and Loizou (2008b) and Hu and Loizou (2008). Briefly, time-frequency units were calculated by fast Fourier transformation on 20-ms Hamming-windowed segments, with a 50% overlap between the segments. All sound files had a sample rate of 44.1 kHz, leading to an FFT size of 882 samples per frame. The SNR of each time-frequency unit was compared against a threshold of 0 dB SNR to determine whether it was retained (at a positive SNR) or eliminated (at a negative SNR). We coded speech and noise that was processed with this original IBM algorithm as condition IBM.

For the next conditions, we introduced a new attenuation function to temper the IBM. Time-frequency units with (estimated) SNR below 0 dB were attenuated by 10 dB. For higher SNRs, the attenuation decreased logarithmically as a function of SNR (see Figure 6.1), which is comparable to a Wiener filter (Loizou 2007; see also Chapter 7). This condition was coded as ITM. The choice of 10 dB attenuation was based on the maximum attenuation provided by noise-reduction algorithms in hearing aids (Hoetink et al. 2009).

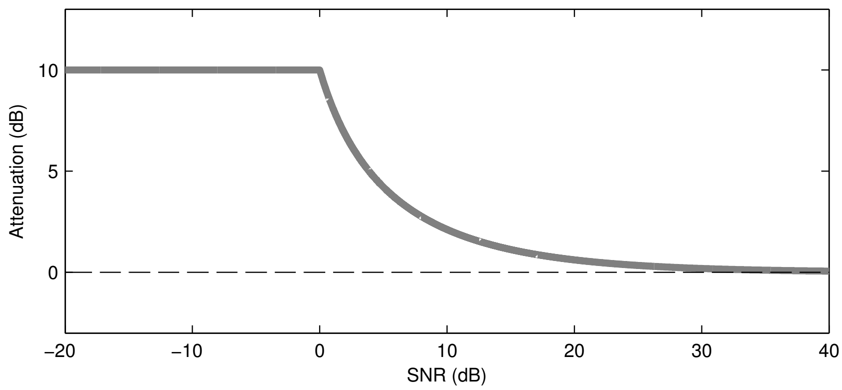


Figure 6.1: Attenuation applied by the tempered mask as a function of (estimated) SNR.

We also used the tempered attenuation function for the last two conditions, but it was preceded by real noise estimation instead of ideal noise tracking. The Martin noise estimator, used in the condition coded as MartinTM, was implemented by Loizou (2007) according to the description in Martin (2001). The Martin algorithm is a minimum tracking algorithm, which means that it makes a rough estimate of the noise level in each frequency band by tracking the minimum of the input power in that band. The second noise estimator we used is a minimum controlled recursive average algorithm (called MCRA2) and was also implemented by Loizou (2007), according to the description in Rangachari and Loizou (2006). This algorithm updates the noise estimate in each frame using a time-frequency-dependent smoothing factor that varies with the probability that speech is present.

For both noise trackers, the SNR of each time-frequency unit was estimated from the estimated noise spectrum using the decision-directed approach (Ephraim and Malah 1984). This method estimates the a priori SNR, $\hat{\varepsilon}_k$ during each time frame (m) and in each frequency band (k), as a weighted average of the estimated SNRs of the previous and current frame:

$$\hat{\varepsilon}_k(m) = \alpha \frac{(G_k(m-1)Y_k(m-1))^2}{\hat{D}_k^2(m-1)} + (1-\alpha)max(\frac{Y_k^2(m)}{\hat{D}_k^2(m)} - 1, 0) \tag{6.1}$$

The weighting factor α was set to 0.98. Y_k(m - 1) is the input spectrum of the previous frame, and G_k(m - 1) expresses attenuation of the previous frame according to the tempered attenuation function based on its estimated SNR. Thus, the numerator of the first fraction represents the output of the previous frame. Dividing by the estimated noise power of that frame ($\hat{D}_k^2(m-1)$, the output of the noise estimator), we obtain the a posteriori SNR. The second fraction represents the estimated SNR of the current frame, with the input in the numerator and the estimated noise in the denominator. Thus, the SNR estimate is largely determined by the estimated SNR of the previous frame. This approach causes the SNR estimate to change gradually, so that the attenuation will not change radically from frame to frame, which reduces musical noise. It is also important to note that, in this approach, the SNR estimate depends on the attenuation function. The method assumes that the output of the previous frame (the numerator of the first fraction) is an estimate of the speech signal in that frame. However, even if the noise is correctly detected, our attenuation function only attenuates it by 10 dB. Thus, the output signal will still contain noise while the approach incorrectly assumes that it does not. This scenario could result in an overestimation of the SNR, leading to less attenuation than desired.

Figure 6.2 shows the four different time-frequency masks for the same sentence. For the IBM, black pixels indicate those noise-dominated time-frequency units that were removed (infinite attenuation), and white pixels indicate the speech-dominated units that were retained. For the other conditions, the attenuation is shown in gray, ranging from white (0 dB attenuation) to gray (10 dB attenuation).

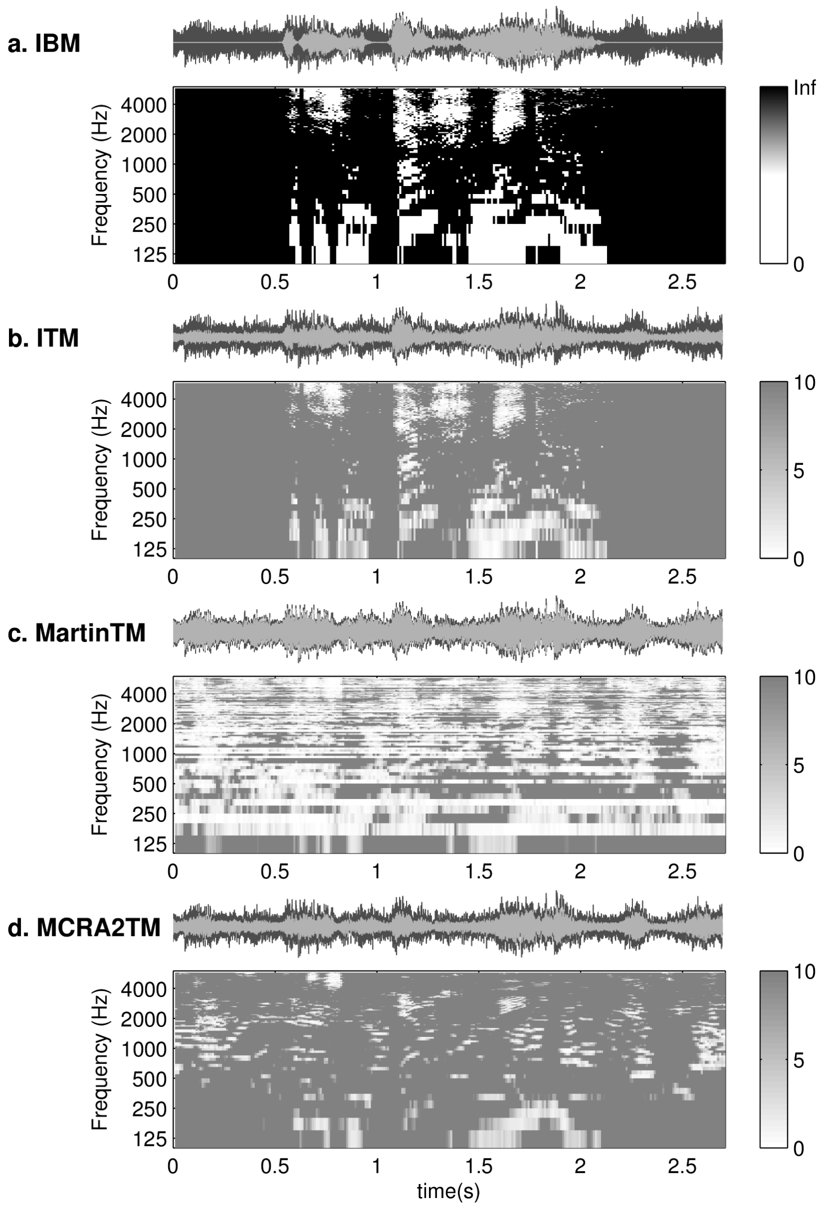


Figure 6.2: Attenuation by the four noise-reduction conditions for the same input sentence. For each processing condition, the time signals show the unprocessed signal (dark background signal) and the processed signal (light foreground signal). The spectrogram-like plots show the attenuation pattern as a function of time and frequency. For the IBM (a), black pixels indicate noise-dominated time-frequency units that were removed (infinite attenuation), and white pixels are the speech-dominated units that were retained. For the other conditions (b-d), the attenuation is color coded between white (0 dB) and gray (10 dB).

Because noise-reduction algorithms are mainly targeted at hearing aids and other mobile devices, we limited the bandwidth of the stimuli after processing to 100-5800 Hz with elliptical filters of the seventh order.

6.2.3 Stimuli

The input signals for the noise-reduction algorithms consisted of 260 unique concatenated Dutch sentences, produced by a female speaker (Versfeld et al. 2000) in multitalker babble noise (Luts et al. 2010). We combined speech and noise at SNRs of -22, -19, -16, -13, -10, -7, -4, 0 and +4 dB, based on the A-weighted representation of the signals. Thirteen sentences (36 s) preceded the stimulus sentences to allow the noise estimators to adapt to the input signals. The noise was continuous, while the speech paused one second between successive sentences.

Stimuli for the perceptual measurements consisted of single sentences cut from the processed signals, with 0.5 s of noise before and after the sentence. The stimuli were presented diotically with Sennheiser HDA200 headphones. The noise level was 70 dB(A) for all stimuli in the unprocessed condition.

6.2.4 Intelligibility

We measured speech intelligibility as the percentage of words that the subjects repeated correctly at fixed SNRs. Each subject started with 13 training sentences containing all five processing conditions, starting at +4 dB SNR. After every three sentences, the SNR decreased one step (4 dB for the first two steps and 3 dB for the last step), terminating with an SNR of -7 dB for the last four sentences. After this training, we used one list of 13 sentences per processing condition per SNR to determine the intelligibility scores. Every new combination of algorithm and SNR started with three training sentences, followed by ten sentences that were used to calculate the percentage correct. Stimuli from all of the five processing conditions were presented at -10, -7 and -4 dB SNR. Additional measurements were performed for the IBM at -22, -19 and -16 dB SNR and for the ITM at -13 dB SNR. We balanced the order of the conditions over all of the subjects to minimize the possible effects of training on the group data. We also balanced the sentence lists over the conditions to minimize the possible effects of differences between lists.

6.2.5 Listening effort rating

The subjects rated their perceived listening effort on a nine-point rating scale that ranged from “no effort” to “extremely high effort.” This rating scale is similar to the test used in Luts et al. (2010) but differs in that our scale used five labeled buttons instead of seven. The five labels are based on ITU-T P. 800 methodology (ITU-T 1996). Subjects gave ratings for all five processing conditions at three SNRs (-4, 0, and +4 dB). We considered the first run of 15 ratings to be practice, and we used the subsequent three runs for analysis.

6.2.6 Paired comparison rating

We used paired-comparison rating (a two-interval, seven-alternative forced choice paradigm) to measure speech naturalness, noise annoyance, and overall preference, successively. This method was based on the ITU-T P.835 method (ITU-T 2003) in which subjects must give separate ratings for the speech signal, the background noise, and the overall quality. The ITU standard uses a rating scale to measure the quality. We chose to use paired comparisons instead because these are more sensitive to subtle differences between conditions (Böckenholt 2001).

For each pair of stimuli, the subjects answered three questions. The first time that they listened to the two fragments A and B, subjects were asked to concentrate on the speech and to rate in which of the two fragments the speech was more natural and to indicate the strength of the difference. After they made a choice, they listened to the same fragments again, now concentrating on the annoyance of the noise and selecting the least annoying fragment. The subjects could listen to both fragments again before they answered the third question. For the third question, the subjects were asked which fragment they would prefer for prolonged listening. For each of the three questions, there were seven possible answers, ranging from “A is much more natural/much less annoying/much better” to “B is much more natural/much less annoying/much better.” The seven choice categories were derived from the Comparison Category Rating method described in ITU-T P. 800 (ITU-T 1996).

All five conditions were paired with each of the other conditions, which resulted in ten different stimulus pairs. Three runs of ten comparisons were performed at both -4 and +4 dB SNR, which resulted in a total of 60 comparisons per subject (10 pairs x 3 runs x 2 SNRs). All of the subjects started with four training pairs. Subsequently, five subjects started with all of the comparisons at -4 dB SNR, and the other five subjects started at +4 dB SNR.

6.3 Results

6.3.1 Intelligibility

Figure 6.3 shows the percentage of words correctly repeated averaged over all ten subjects. For statistical analysis, we transformed the percentages of correct words to rationalized arcsine units (rau) (Studebaker 1985) and subsequently performed a repeated-measures analysis of variance (ANOVA) on the transformed data for -10, -7, and -4 dB SNR, with SNR and processing condition as fixed effects. We found significant effects of SNR ($F[2,18] = 631.7$, $p < 0.001$) and processing condition ($F[4,36] = 332.8$, $p < 0.001$), and a significant interaction between processing condition and SNR ($F[8,72] = 21.0$, $p < 0.001$). Post hoc Bonferroni-corrected pairwise comparisons showed that the scores for IBM and ITM were higher than for the other three conditions at all SNRs. At -4 dB SNR, IBM and ITM were not significantly different from each other, but at -10 and -7 dB SNR, scores for IBM were significantly higher than for ITM (Bonferroni-corrected p-values were < 0.001 for all of the differences mentioned). Intelligibility scores for the other conditions (unprocessed, MartinTM, and MCRA2TM) did not differ significantly from each other.

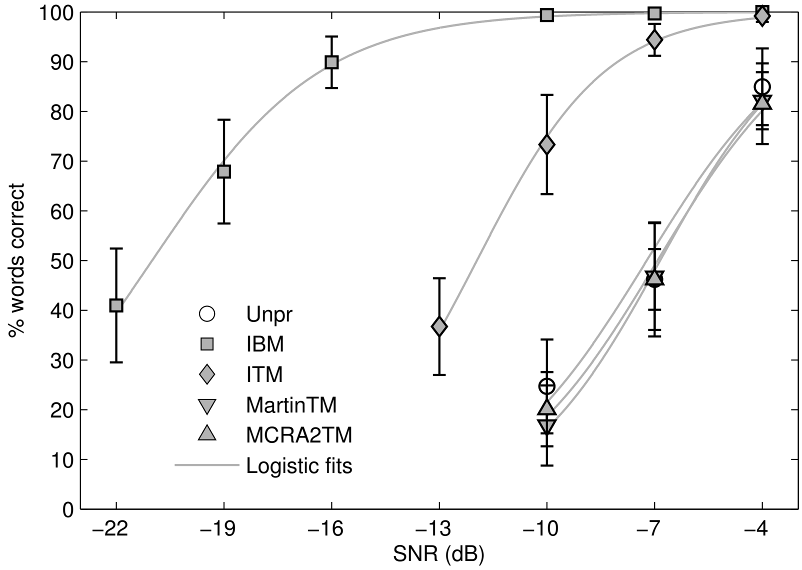


Figure 6.3: Mean percentage of words correctly repeated by the 10 subjects at the different SNRs. Error bars show the 95% confidence interval between subjects (without Bonferroni correction).

6.3.2 Listening effort rating

Figure 6.4 shows the mean listening-effort ratings assigned by the 10 subjects. Note that a higher value means that the listening effort was lower. To satisfy the ANOVA criteria, we transformed the listening effort ratings with an arcsine transformation. We performed a repeated measures ANOVA with SNR and processing condition as fixed effects. We found significant effects of SNR ($F[2,18] = 227.4$, $p < 0.001$) and processing condition ($F[4,36] = 35.5$, $p < 0.001$) and a significant interaction between processing condition and SNR ($F[8,372] = 15.73$, $p < 0.001$). The horizontal lines in Figure 6.4 indicate which conditions differed significantly from each other after Bonferroni correction.

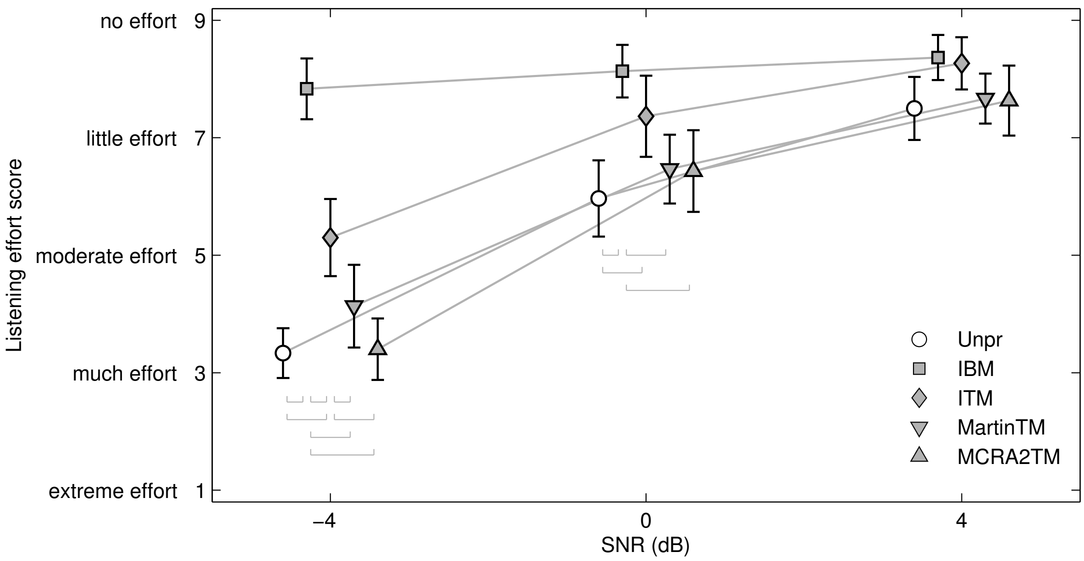


Figure 6.4: Mean and 95% confidence intervals of the listening effort ratings assigned by the 10 subjects at -4, 0 and +4 dB SNR. The horizontal bars indicate which processing conditions differ significantly from each other after Bonferroni correction for 10 comparisons.

6.3.3 Paired comparison rating

Figure 6.5 shows the average rating score for each processing condition. We assigned scores from -3 to 3 for each condition, according to the ITU-T recommendation P. 800 (ITU-T 1996). If the subject rated condition A slightly better than condition B, then we assigned a score of 1 to condition A and a score of -1 to condition B. Similarly, scores of -2 and +2 indicate a moderate difference, and scores of -3 and +3 indicate a major difference, and a score of 0 indicates no difference. The scale for the noise annoyance is inverted in Figure 6.5. As a result, for each outcome, a positive value means a better performance on that judgment criterion. Error bars show a 95% confidence interval between the subjects.

Because, in general, scorings do not represent a linear interval scale, we used the log-linear modeling approach for ordinal paired-comparisons described by Dittrich et al. (2004) for the statistical analysis of the paired-comparison rating data. The model is a log-linear representation of the Bradley-Terry model (Bradley and Terry 1952) and is extended for paired-comparison data with multiple response categories, including a “no difference” option. By fitting this model to the paired-comparison data, we obtained estimates of the so-called “worth” parameters, which describe the location of the five processing conditions on the subject’s preference scale. This scale can be interpreted similarly to a ratio scale, thus providing not only the ranking of preferences for the five conditions but also information regarding the strengths of the preferences.

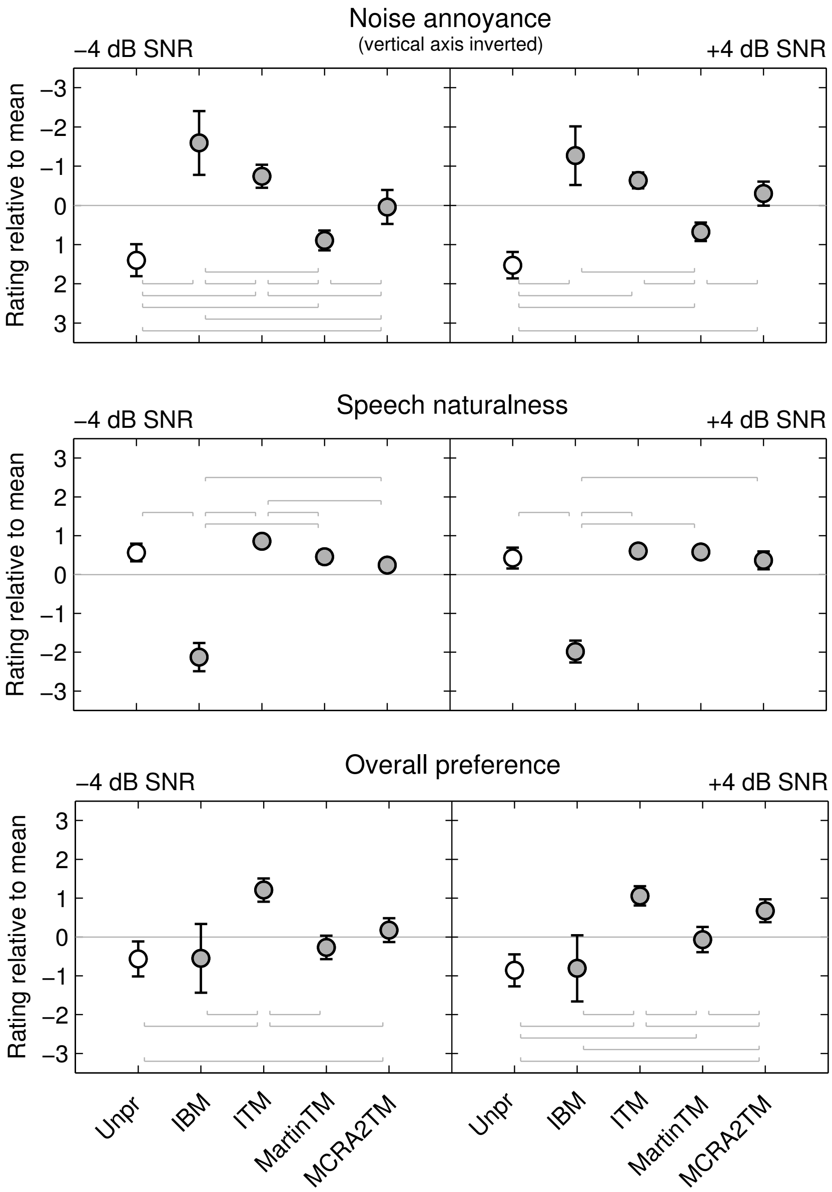


Figure 6.5: Mean and 95% confidence intervals of the rating scores derived from the paired-comparison data for the three judgment criteria and two SNRs. We assigned values from -3 to +3 to the answers, in accordance with ITU P. 800 (ITU-T 1996). The horizontal bars indicate which processing conditions differ significantly from each other after Bonferroni correction for 10 comparisons.

We estimated the worth parameters separately for the noise annoyance, speech naturalness, and overall preference. We fitted a model for each individual run of ten comparisons, which resulted in three models per subject per SNR per judgment criterion. We tested the goodness-of-fit for all of the models by comparing the obtained model with a saturated model (a model reproducing the data perfectly). All of the p-values were >0.95, indicating a high agreement with the saturated model; thus all of the models could be accepted.

We did repeated-measures ANOVAs on the estimated worth parameters for each judgment criterion separately (noise annoyance, speech naturalness, and overall preference) with SNR and processing condition as fixed effects. The resulting F-statistics and p-values are presented in Table 6.2. We found a significant effect for processing condition for each of the three judgment criteria. The effect of SNR (+4 or -4 dB SNR) was only significant for the overall preference. The interaction between processing condition and SNR was significant for all three criteria. Because of the significant interaction between processing condition and SNR, we redid the repeated-measures ANOVA but treated each SNR separately, with processing condition as a fixed effect. The resulting values for F and p are also given in Table 6.2. In both analyses (-4 and +4 dB SNR), the effect of processing condition was significant for each judgment criterion. The horizontal lines in Figure 6.5 indicate which conditions differed significantly from each other after Bonferroni correction.

Table 6.2: Main analysis of variance outcomes for the paired-comparison results.

		Noise annoyance		Speech naturalness		Overall preference	
		Both signal-to-noise ratios					
Effect	df	F	p	F	p	F	p
Processing condition	4	6.87	< 0.001	47.88	< 0.001	7.90	< 0.001
SNR	1	2.00	0.19	2.02	0.19	7.84	< 0.05
Processing condition x SNR	4	8.35	< 0.001	3.23	< 0.05	5.97	< 0.001
		-4 dB signal-to-noise ratio					
Processing condition	4	6.81	< 0.001	43.36	< 0.001	6.21	< 0.001
		+4 dB signal-to-noise ratio					
Processing condition	4	39.97	< 0.001	6.40	< 0.001	9.87	< 0.001

6.4 Discussion

With respect to our research questions, we can summarize our findings as follows.

Q1. Our results show that the IBM reduces noise annoyance, causing a strong increase in intelligibility and a decrease in listening effort. However, the IBM also strongly reduces speech naturalness and is therefore not preferred over no processing at the SNRs tested (-4 and +4 dB SNR).

Q2. Tempering the IBM by limiting its maximum attenuation to 10 dB with a gradual instead of a binary attenuation function (ITM) removes the disadvantage of reduced speech naturalness. The tempered attenuation causes the ITM to be preferred over the IBM and over all of the other conditions. Compared to the unprocessed condition, the ITM still increases intelligibility and reduces the listening effort and noise annoyance, but to a lesser extent than the IBM.

Q3. Replacing the ideal noise tracker with real noise estimators removes the benefit of increased intelligibility, as expected. The MCRA2 noise estimator reduced the noise annoyance compared with no noise reduction and was also preferred over no noise reduction. The Martin noise estimator caused much smaller reductions in the noise annoyance and was only slightly preferred over no noise reduction at +4 dB SNR.

6.4.1 Ideal binary mask

As expected, the IBM strongly improved intelligibility. Li and Loizou (2008b) described intelligibility measurements with the same IBM implementation. In their first experiment, they investigated the effect of the threshold for retaining or removing time-frequency units on intelligibility. Because changes in this threshold have the same effect as changes in the input SNR (Brungart et al., 2006), we can interpret the results for the different thresholds as results for different input SNRs. In spite of the differences in speech material, the obtained word scores for speech in 20-talker babble agree well between Li and Loizou (2008b) and our study (we estimate from their Figure 2 that the SRT_{50} was at -19 dB SNR, compared to -21 dB SNR in our study). Additionally, Wang et al. (2009) found a 13.4 dB improvement from IBM relative to the unprocessed condition in stationary speech-shaped noise (corrected for their threshold of -6 dB), compared to 13.8 in our study. However, Brungart et al. (2006) found smaller improvements in intelligibility for speech in speech-shaped noise (we estimate from their Figure 5 (left) an SRT_{50} improvement of 5 dB). This difference can most likely be attributed partly to the differences in speech material. Brungart et al. (2006) used words embedded in a fixed carrier sentence as speech material, whereas the other studies used sentence materials.

The IBM reduced the noise annoyance more than all of the other conditions, but also reduced speech naturalness the most. Subjects described the IBM condition as “a computer-like voice” or “watery sounding”. This unnaturalness is probably the reason that the IBM was not preferred over unprocessed or other noise reduction conditions at -4 and +4 dB SNRs. However, the large error bars for the overall preference of the IBM indicate that individual subjects differ in whether they prefer IBM processing or not. Whereas the preference for the other conditions can be mainly based on the sound quality, the IBM requires a choice between low quality and high intelligibility. Which of these criteria will obtain the higher weight could differ between listening situations, and several subjects indicated that they had difficulty with the choice.

It can be expected that at +4 dB SNR intelligibility scores will be 100% for all of the conditions. However, the listening effort scores still show differences between the ideal and non-ideal conditions at this SNR. Thus, even if speech is fully intelligible, there could be a benefit of noise reduction in terms of reduced listening effort. If the SNR decreases, the perceived effort increases for all of the conditions except for the IBM because it still removes all of the noise and provides a large intelligibility improvement. It seems that the unnaturalness of speech does not increase the perceived listening effort at SNRs high above the SRT_{50}.

The IBM seems optimal for improving speech intelligibility (Loizou and Kim 2011), but it can only be used in the exceptional situation in which both speech and noise are given separately as input signals. A few attempts have been made to estimate the binary mask from a single input signal in which speech and noise were mixed. For example, Kim et al. (2009) trained binary SNR classifiers to estimate the binary mask. Their approach do not require an accurate SNR estimate, but only an accurate classification of SNR $<$ 0 dB and SNR $>$ 0 dB. Their classifier estimated the IBM accurately enough to result in large intelligibility improvements for normal-hearing listeners. It remains to be seen whether these results can be generalized, as Brookes and Huckvale (2011) were not able to reproduce these intelligibility improvements. Estimation of the binary mask thus requires more investigation.

6.4.2 Ideal tempered mask

Tempering the IBM (i.e., limiting the attenuation to 10 dB and making the attenuation function gradual instead of binary) reduced the intelligibility benefit compared with IBM, but there was still an improvement in SRT_{50} of 4.8 dB due to ITM compared with the unprocessed condition. Anzalone et al. (2006) also limited the attenuation of their IBM. They did not use a gradual attenuation function, and their binary attenuation was 0 or 14 dB. For speech in speech-shaped noise, their IBM improved the SRT_{50} more than 7 dB for normal-hearing listeners and approximately 9 dB for hearing-impaired listeners. These values are somewhat higher than those found in our results, but multiple differences in processing strategy make it difficult to compare the studies. For example, we used a gradual instead of a binary attenuation and based the attenuation on the SNR, whereas they used the speech energy for the binary decision.

Loizou and Kim (2011) have shown in their Appendix that the IBM is optimal in that it maximizes a simplified form of the articulation index, a measure known to correlate highly with speech intelligibility. As a result, the binary infinite attenuation function is theoretically optimal for improving intelligibility. From these derivations, it follows that tempering the attenuation, as in our ITM condition, leads to sub-optimal attenuation functions for intelligibility. Our intelligibility results confirm these theoretical hypotheses. Both versions of the ideal mask (IBM and ITM) cause intelligibility scores of (almost) 100% at -4 dB SNR. However, the perceived listening effort is higher for the ITM than for the IBM. In contrast to the IBM, the amount of residual noise, and thus the listening effort for the ITM, increases with a decreasing SNR. Although more effortful than the IBM, the ITM is less effortful than the unprocessed condition at all SNRs.

The paired-comparison data show that the ITM reduces noise annoyance, but in contrast to the IBM, not at the cost of speech naturalness. Taken altogether, it is not surprising that ITM is clearly preferred over all of the other conditions. The ITM combines improved speech intelligibility with less degraded sound quality.

Cao et al. (2011) investigated the effect of adding stationary noise after IBM processing and found that, for certain noise levels, the addition of noise after IBM processing can further improve the intelligibility. They hypothesized that filling the sudden silences caused by IBM processing enhances the perceived continuity of the speech, leading to an intelligibility improvement. Instead of adding noise to the processed signal, limiting the attenuation could lead to a similar effect because the limitation leads to more residual noise in the processed signal. In our study, we limited the ITM attenuation to 10 dB. The output SNR can thus be at best 10 dB higher than the input SNR. For the ITM condition, our input SNRs thus lead to output SNRs of at best -3 to +6 dB (for input SNRs of -13 to -4 dB). Cao et al. (2011) found significant intelligibility improvements for output SNRs of 8 dB and higher. Thus, for our ITM condition it is not expected that the residual noise improves speech intelligibility. To investigate possible effects of residual noise on intelligibility one would need to use the input SNRs where the IBM did not give 100% speech intelligibility (-22 to -16 dB) combined with an ITM limitation of 30 dB. This would theoretically result in output SNRs of +8 to +14 dB, the region where Cao et al. (2011) found an improvement caused by the noise. Although it could be worthwhile to determine if the ITM can be optimized, combining the required 30 dB attenuation with non-ideal noise tracking could lead to large distortions due to noise-estimation errors.

6.4.3 Ideal noise tracking compared to noise estimation

The intelligibility results from MartinTM and MCRA2TM confirm the hypothesis that the positive effects of IBM will be negated if the prior knowledge of the noise and speech signal (ideal noise tracking) cannot be used (real noise estimator). Although the attenuation function we used has proven to be able to improve the SRT_{50} by roughly 5 dB (see ITM), no improvement remained in the conditions in which noise and speech had to be estimated from the mixed signal. In terms of the perceived listening effort, these realistic noise-reduction conditions also provided no benefit over the unprocessed condition. However, both MartinTM and MCRA2TM significantly reduced the perceived noise annoyance compared to the unprocessed condition, without affecting the speech naturalness. This reduction in noise annoyance was higher for MCRA2TM than for MartinTM. MCRA2TM was, at both SNRs, preferred over the unprocessed condition, whereas MartinTM was slightly preferred only at +4 dB SNR.

The fact that MCRA2TM was preferred over unprocessed at -4 dB SNR whereas the IBM was not, is noteworthy given the fact that the intelligibility scores for MCRA2TM were approximately 19% lower than for the IBM, and the perceived listening effort was lower than for the IBM. Thus, it seems that the sound quality was more decisive for the subjects' preference than the benefit in terms of intelligibility or listening effort. The objective intelligibility can thus be balanced against the subjectively perceived quality by modifying the attenuation function. The optimum balance depends on both the situation and the individual listeners. For non-ideal masks, it is important to take into account that modifying the attenuation function could lead to a decrease in intelligibility.

6.4.4 Comparison of different noise estimators

The paired-comparison results show a difference between the two noise-estimation conditions in the degree to which they reduce the noise annoyance. This difference is not surprising given the attenuation patterns in Figures 6.2c and 6.2d, which are very different from each other. In the MartinTM condition, the majority of noise-dominated time-frequency units were not attenuated or were only slightly attenuated. This lack of attenuation resulted from the fact that the minimum tracking method often causes an underestimation of the true noise level (Loizou 2007; Chen and Loizou 2012). Additionally, the noise-spectrum estimate was updated slowly by the Martin algorithm, resulting in the stripes in Figure 6.2c (Loizou 2007). The small number of attenuated time-frequency units was still sufficient to slightly reduce the noise annoyance and did not affect speech naturalness. At +4 dB SNR, this reduced noise annoyance caused the MartinTM to be slightly preferred over the unprocessed condition. Thus, although the changes are small, this noise reduction can improve the sound quality at higher SNRs.

In contrast, in the MCRA2TM condition, the majority of noise-dominated units were attenuated (see Figure 6.2d), resulting in a reduced perceived noise annoyance. Although MCRA2TM also attenuated many of the speech-dominated units, this attenuation was apparently not perceived as reduced speech naturalness. This condition probably sounds like an overall attenuation of the unprocessed signal. Although MCRA2TM was not able to improve intelligibility or listening effort, the subjects preferred this condition at +4 dB SNR over all of the other conditions except for the ITM, which was based on ideal noise tracking.

Li and Loizou (2008b) defined two types of noise-estimation error: type I error, which occurs when a noise-dominated time-frequency unit is retained, and type II error, which occurs when a speech-dominated unit is removed. We estimated these error percentages for the Martin and MCRA2 noise estimators with our speech and noise stimuli as input signals. For this purpose we removed the noise-only timeframes between the sentences. Similar to the IBM, we compared the estimated SNR of each time-frequency unit to the threshold of 0 dB SNR to obtain a binary mask. We compared these masks to the IBM to calculate the percentages of misclassified units (Li and Loizou, 2008b, Appendix). For an input signal of -4 dB SNR, the Martin algorithm resulted in an average of 77% type I errors (i.e., 77% of all of the noise-dominated units were classified as speech-dominated) against 16% for the MCRA2 algorithm. In contrast, the percentage of type II errors (speech-dominated units classified as noise-dominated) was 19% for the Martin algorithm and 75% for the MCRA2 algorithm.

Li and Loizou (2008b) concluded that the binary mask did not improve intelligibility if the error percentage exceeded 85% for either error type in isolation. If both error types are present, this percentage is expected to be lower. Our noise estimators both approached this percentage already for one type of error but had additional errors of the other type. This result confirms that, for our speech and noise material, no intelligibility improvement can be expected from noise reduction based on the Martin or MCRA2 noise-estimation algorithms.

Li and Loizou (2008b) also concluded that type I and type II errors have different effects on intelligibility. Whereas type I errors affected intelligibility even at low percentages, type II errors can occur in up to 60% of the speech-dominated units before they cause a substantial decrease in intelligibility. Our results also seem to indicate that type II errors, which dominated in our MCRA2TM condition, are less detrimental for the overall preference. This result suggests that, if the same attenuation function is used, a noise reduction with a noise estimator that tends to underestimate the SNR would outperform a noise estimator that tends to overestimate the SNR. This statement is further supported by Chen and Loizou (2012). They systematically introduced different degrees of SNR overestimation and underestimation in a Wiener-filter based noise-reduction algorithm. The resulting intelligibility scores confirmed that SNR overestimation for time-frequency units with negative SNR was much more harmful to speech intelligibility than SNR underestimation.

6.4.5 Limitations

Paired-comparison ratings are especially useful if differences between the stimuli are small. However, our IBM condition was clearly different from all of the other conditions. The high contrasts with the IBM could have led to smaller perceived differences between the other processing conditions.

For the non-ideal conditions, both the noise-estimation algorithm and the decision-directed approach for SNR estimation determine the final attenuation and estimation errors. Our results do not allow us to distinguish between these two approaches. We can, however, calculate the influence of the decision-directed approach on the error rates. We repeated the error calculations with the weighting factor α set to 0 (see Equation 6.1), so that the SNR estimation was based only on the input signal and noise estimate of the current frame. Compared with an α of 0.98, the percentage of Type I errors increased for both estimators (from 77% to 85% for Martin and from 16% to 39% for the MCRA2). The Type II errors decreased (from 19% to 18% for Martin and from 75% to 57% for MCRA2). Thus, the decision-directed approach tends to enhance SNR overestimation. This bias is introduced by the clipping function (the max operator, see Equation (6.1)) implemented in the decision-directed approach (Chen and Loizou, 2012). Additionally, as discussed in section 6.2.2, the combination of the decision-directed approach and a limited attenuation function also leads to SNR overestimation. As a consequence, without the decision-directed approach, fewer time-frequency units will be attenuated. This strategy will retain more noise, which would probably result in more noise annoyance and less preference than with the decision-directed approach.

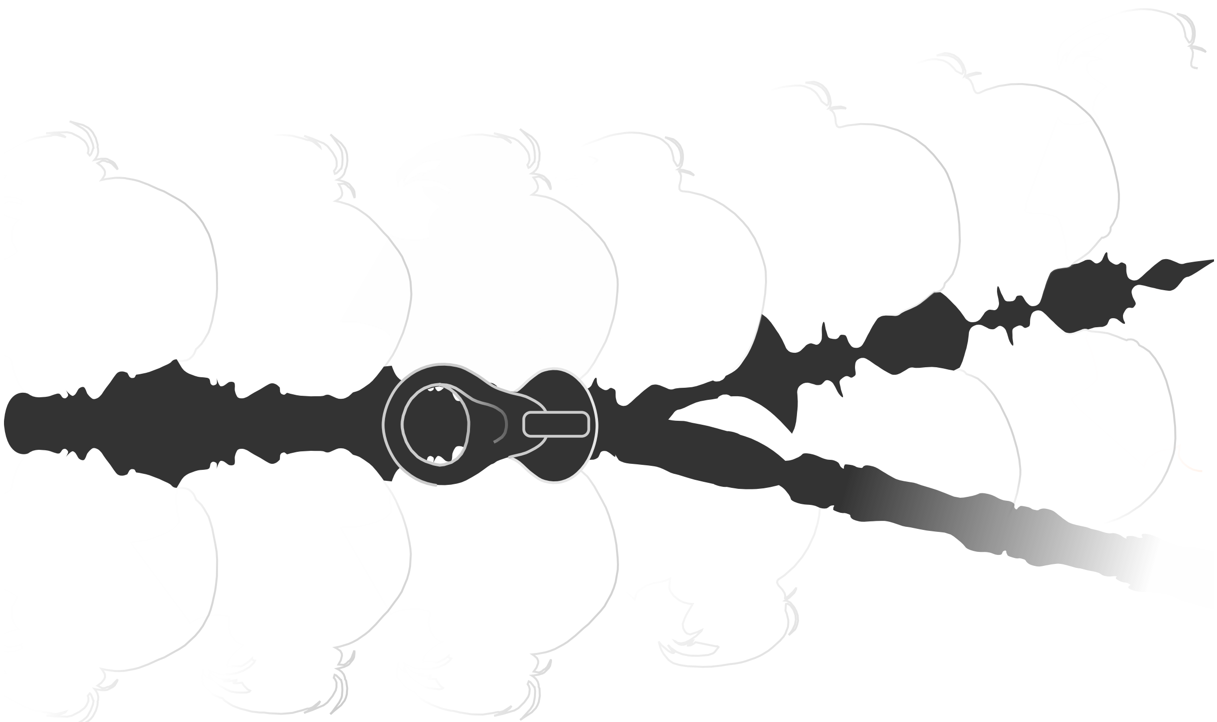
We used a threshold of 0 dB SNR in the attenuation function for all of the conditions, which is a common choice for the IBM (Wang 2008). Additionally, the use of the same attenuation function for the three tempered conditions allowed us to compare the effect of the different noise-tracking methods. However, one must be aware of the possible effects of the attenuation function on the results. As noted in section 6.4.1, the threshold for retaining or removing time-frequency units for the IBM is directly related to the SNR of the input signal (Kjems et al. 2009). For example, a threshold of -8 dB SNR instead of 0 dB SNR would lead to the same results for input signals at -4 dB SNR as would a 0-dB threshold for input signals at +4 dB SNR. For the non-ideal conditions, however, the SNR estimate is not independent of the attenuation function (see section 6.2.2). The attenuation function that we used was not the function for which the noise estimators were developed and tested. Thus, the optimal choice for the attenuation threshold and the maximum attenuation depends on the noise-estimation algorithm, and this choice needs further investigation.

An important application of noise reduction is in hearing aids, to make listening in noisy environments easier. For this application, our normal-hearing study population differs from the hearing-impaired target population. Subsequent investigations are required to determine the effects for listeners with sensorineural hearing losses. Previous studies have shown that the IBM also improves intelligibility for hearing-impaired listeners (Anzalone et al. 2006; Wang et al. 2009). The improvement in intelligibility might be even higher for hearing-impaired listeners than for normal-hearing listeners because the performance in noise is worse for hearing-impaired listeners, whereas performance after IBM processing is comparable between normal-hearing and hearing-impaired listeners (Wang et al. 2009). Anzalone et al. (2006) observed a difference between normal-hearing and hearing-impaired listeners in their informal comments about the speech quality. The hearing-impaired listeners appeared to be less sensitive to the reduced quality. However, as mentioned before, the binary attenuation in that study was 0 or 14 dB, so that the noise was not completely removed but instead was only attenuated by 14 dB. This change reduces the speech distortions compared to the IBM, especially for the higher input SNRs presented to the hearing-impaired listeners. Thus, it is not certain whether hearing-impaired listeners indeed had lower sensitivity to distortions or whether their stimuli contained less distortion. On the one hand, it appears reasonable that hearing-impaired listeners have a higher detection threshold for speech distortions, so that they will accept a stronger attenuation. On the other hand, avoiding speech distortions could be more important for hearing-impaired listeners because of their dependence on clean speech signals. Additionally, there are also individual differences between hearing-impaired listeners in how well they accept background noise (Mueller et al. 2006) and in the maximum attenuation strength that they prefer for noise reduction (Houben et al. 2011). More investigation is required to determine to what extent a preference for noise reduction is determined by a hearing loss and to what extent by other individual differences.

6.5 Conclusions

We conclude that, although the IBM improves speech intelligibility in noise, listeners do not prefer it over the unprocessed condition at SNRs of -4 and +4 dB because it sounds unnatural. Tempering the IBM (limiting the attenuation to a maximum of 10 dB and smoothing the attenuation function) overcomes this drawback of the IBM while maintaining an intelligibility improvement, although to a lesser extent. Other values for the maximum attenuation should be evaluated to find an optimum for the trade-off between intelligibility and sound quality. With a real noise-estimation algorithm as the basis for noise reduction, estimation errors negate the potential intelligibility benefit of noise reduction. However, such realistic noise reduction can reduce the noise annoyance that is perceived by listeners so that they prefer it over the unprocessed signal. Although noise reduction based on noise-estimation algorithms does not yet provide an objective benefit in terms of intelligibility, possible subjective benefits should receive more attention in the development and evaluation of noise-reduction algorithms.

7 Detection threshold for sound distortion because of noise reduction in normally hearing and hearing-impaired listeners



Inge Brons, Wouter A. Dreschler, Rolph Houben

Submitted for publication

7.1 Introduction

Most modern hearing aids have a single-microphone noise-reduction algorithm that is designed to make listening in noisy environments more comfortable and less effortful. The algorithm should reduce background noise without affecting the target speech signal. For this purpose, the noise-reduction algorithm continuously analyses the input signal to differentiate between speech and background noise. Based on the estimated ratio between speech and noise, the algorithm can adjust the hearing-aid gain during each timeframe for noise-dominated frequency channels (Chung 2004). However, because speech and noise enter the hearing aid as a mixed signal, it is difficult for the noise-reduction algorithm to separate them completely. Therefore, reduction of noise is usually accompanied by signal distortions. An increase in noise-reduction strength will decrease the amount of residual noise (desired effect) while adding distortions (other than noise-level differences) to the speech and remaining noise (unwanted effect). Because noise-reduction algorithms offer no way of controlling noise reduction and signal distortion independently (Loizou 2007), the ideal situation of total noise removal without distortion can never be achieved in practice (unless there is a-priori knowledge available on the original speech and noise signals). Thus, a challenge in the design and fitting of a noise-reduction algorithm is to find the optimal trade-off between residual noise and distortion.

Current single-microphone noise-reduction algorithms are not able to improve speech intelligibility in noise, but the reduction of noise can make listening more comfortable and less effortful (Boymans and Dreschler 2000; Bentler et al. 2008), despite speech distortion. Studies that tried to demonstrate such perceptual effects of noise reduction gave inconsistent results (Alcàntara et al. 2003; Bentler et al. 2008; Boymans and Dreschler 2000; Ricketts and Hornsby 2005). These inconsistencies were likely caused by differences between noise-reduction systems and differences between listeners. For instance, noise-reduction systems from different hearing aids vary in the degree they reduce noise annoyance and affect speech naturalness (see Chapters 3 to 5). The reason for differences in preference may be that individual listeners differ in the way they balance for instance noise annoyance and speech naturalness, in their personal preference for a specific noise-reduction system (Houben et al. 2012).

It is not known to what extent hearing loss influences the preferred balance between residual noise and distortion. There are indications that hearing-impaired listeners are less sensitive to speech distortions than normal-hearing listeners. Marzinzik (2000) found that hearing-impaired listeners judged speech to sound more natural with noise reduction than that for unprocessed. They hypothesized that hearing-impaired listeners were less sensitive to speech distortion and therefore looked for other cues while judging speech naturalness, for instance the presence of background noise. If hearing loss is indeed accompanied by a reduced ability to hear signal distortions, hearing-impaired listeners may tolerate more speech distortion and thus stronger noise reduction than normal-hearing listeners. Studies that investigated distortion caused by other types of signal processing also indicated that hearing-impaired listeners are less sensitive to signal distortions (Lawson and Chial 1982; Stelmachowicz et al. 1999). However, there are also results to indicate that listeners with hearing loss are similar to listeners with normal hearing in their ability to perceive distortion changes (Arehart et al. 2010). Currently it is not known whether hearing loss influences the just noticeable differences (jnd) of noise-reduction processing. This knowledge could improve both the customization of noise reduction to the specific hearing loss of the user and the development of new noise-reduction algorithms that are targeted at hearing-impaired users. Measurements of just noticeable differences for distortion caused by noise reduction algorithms are complicated by the fact that noise reduction processes speech and noise simultaneously. The resulting stimuli thus contain different cues, namely the reduction in noise and the introduction of distortions. Subjects can use these different cues for what they perceive as distortion, and the results are difficult to interpret. To overcome this problem we used a noise-reduction system in which we had access to calculated gain reductions. This system allows us to separate the distortions in speech and noise from the reductions in noise level, which, in turn, allows us to measure the detection threshold for distortion (the negative effect of noise reduction) without the confounding effect of differences in noise levels (the positive effect of noise reduction). The main research question that we wanted to answer is:

Q1. Does the detection threshold for noise-reduction induced signal distortions (other than level changes) differ between normal-hearing and hearing-impaired subjects?

Signal distortion refers here to all changes to the speech and noise signal that remain after we corrected separately for changes in overall level of speech and noise. Besides speech distortion, this can also include distortion of the remaining noise. A well-known example of processing artifacts caused by noise reduction is musical noise, a tonal character that was not present in the original signal (Berouti et al. 1979). Musical noise occurs because of small isolated peaks that remain in the signal spectrum, and can even be more disturbing to the listener than the original interfering noise (Loizou 2007). To verify whether noise reduction caused distortions both in the speech and the noise segments of our stimuli, we also measured detection thresholds for distortion of speech and distortion of noise separately:

Q2. Are the thresholds for noise-reduction induced signal distortions related to thresholds for distortions in the speech signal or to the distortions in the noise signal?

In all three conditions the strength and temporal pattern of the attenuations were the same, but the conditions differ in whether this attenuation was applied to speech, noise, or both. For clarity, we will use the terms overall distortion, speech distortion, and noise distortion to distinguish between the three separate conditions, irrespective the subjective impression of whether the distortion is ascribed to the speech or to the noise. To measure the threshold for overall distortion, we presented the participant with signals that consist of speech in noise that was processed by noise reduction. To measure speech distortion the participants listened to processed speech only (no noise present) and to measure noise distortion they listened to processed noise (processed speech present which is equal for target and reference condition, see methods).

The detection thresholds for distortion provide information on how much noise reduction can be applied before a listener experiences audible distortions. For practical use it is important to know the level at which distortions can be detected by hearing impaired listeners (Q1), because we expect that the noise reduction can be set to at least this strength. In addition to this theoretical base value of acceptable noise reduction strength, we also wished to determine the amount of audible distortions that is actually accepted by the hearing-impaired listener in exchange for a reduced level of background noise. This level is expected to be higher than the detection threshold. To investigate these issues, we determined the preference of the subjects for noise reduction strength, and determined whether there was a difference between normal-hearing subjects and hearing-impaired subjects. In addition, we related the optimal values for the trade-off (thus the preferred noise-reduction strengths) to the individual detection thresholds.

Q3. Is there a difference between normal-hearing subjects and hearing-impaired subjects in their preferred noise-reduction strength, both absolute and relative to their detection threshold?

To determine the preference for noise-reduction strength subjects listened to the speech in noise that was processed by noise reduction. Thus, both the negative (distortion) and positive (reduced noise level) effects of noise reduction were included. Participants could vary the noise-reduction strength to choose the setting that they would prefer for prolonged listening. We compared individual preferences of the subjects to their individual detection thresholds for overall distortion.

7.2 Methods

7.2.1 Subjects

We estimated the required sample size using a power calculation based on data from a pilot experiment with three normal-hearing subjects. Their between-subjects standard deviation for the threshold for overall distortion was 1.18 dB. To be able to detect a difference of 2 dB (the stepsize of the adaptive procedure for threshold measurement) between the thresholds of the two subject groups (normal hearing and hearing impaired), seven subjects should be included per group for $\alpha = 0.05$ and $1\text{-}\beta = 0.8$. Because between-subject differences are expected to be somewhat higher for the group of hearing-impaired subjects, 12 subjects were included per group.

Twelve normal-hearing subjects aged between 19 and 50 years (average 26.4 years) and 12 subjects with sensorineural hearing loss aged between 53 and 69 years (average 61.1 years) participated in this laboratory study. Figure 7.1 shows the hearing thresholds and the corresponding standard deviations for the test ears averaged over all subjects.

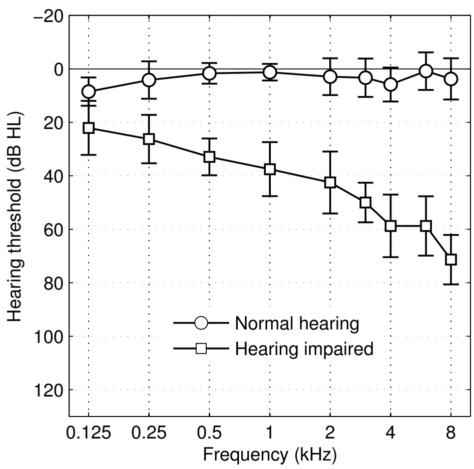


Figure 7.1: Average hearing thresholds for the normal-hearing and hearing-impaired group (one ear per subject). The error bars show the standard deviation between subjects.

7.2.2 Stimuli

The input signals for the noise-reduction algorithm consisted of concatenated Dutch sentences (female speaker) in a stationary background noise with the same spectrum as the speech (Versfeld et al. 2000). We combined speech and noise at an SNR of +5 dB based on the A-weighted sound level of the sentences. Thirteen sentences (approximately 35 s) preceded the sentences actually used, to allow the noise-reduction algorithm to adapt to the input signals. The noise was continuous, but the speech paused one second between successive sentences.

Stimuli for the threshold measurements consisted of five unique sentences, with 0.2s of noise before and after the sentence. All stimuli were presented monotically with Sennheiser HDA200 headphones. The overall level of the original speech in noise at +5 dB SNR was 70 dB(A). For hearing-impaired subjects, the stimuli were amplified according to the individual linear NAL-RP prescription rule (Byrne et al. 1991).

7.2.3 Signal processing

We used a state-of-the-art noise-reduction algorithm from literature (Rangachari and Loizou 2006). This algorithm in its original form causes musical noise. From recordings of noise reduction in commercially available hearing aids, it appeared that no musical noise was audible in a selection of hearing aids (see Chapters 3 to 5). To ensure that our results are representative of hearing aid noise reduction, we made two adjustments to the noise-reduction algorithm to minimize musical noise. First, we limited the maximum gain reduction that can be applied by the noise reduction from infinity (original Wiener filter) to a fixed value between 0 and 30 dB. This procedure has been used previously to adapt an ideal binary mask to resemble hearing aid processing (Anzalone et al. 2006; Chapter 6). The variable used to adjust noise-reduction strength and thereby distortion was the maximum attenuation, as was carried out previously by Houben et al. (2012) with another noise-reduction algorithm to determine preference for noise-reduction strength. Second, as is common in hearing aids, we limited the number of frequency channels for the gain signal to fifteen channels, instead of using all available frequency bins separately.

Figure 7.2 schematically summarizes the steps that were used to process the stimuli, which will be explained in detail in the next paragraphs.

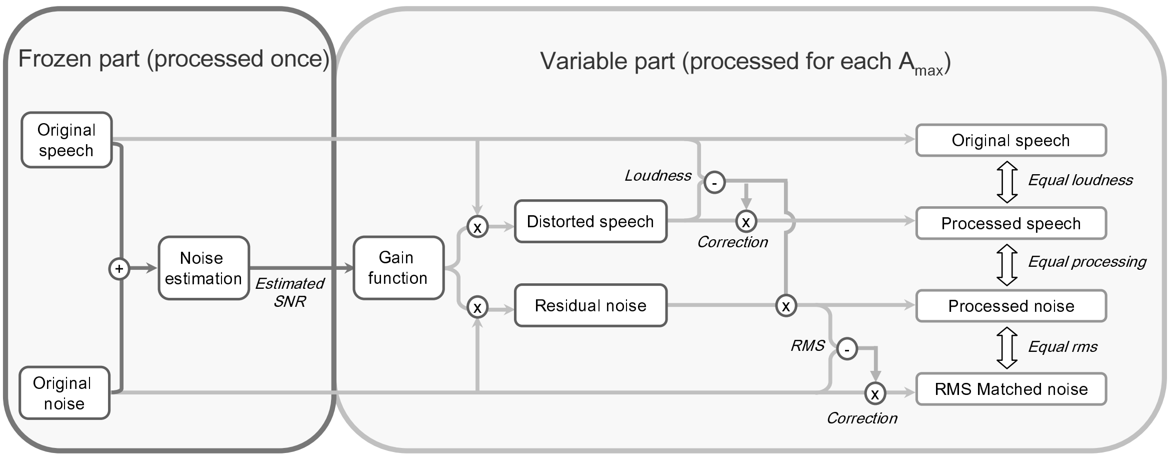


Figure 7.2: Processing scheme for the stimuli. The four resulting signals, shown on the right of the figure, are the basis for the stimuli that were used during the measurements.

Estimation of the signal-to-noise ratio (SNR)

Speech and noise signals were combined before entering the noise-reduction algorithm so that the noise-reduction algorithm had to estimate which part of the signal was speech and which part noise. For this purpose we used a minimum controlled recursive average algorithm (called MCRA2) that estimated the noise spectrum. This algorithm was implemented by Loizou (2007) according to the description by Rangachari and Loizou (2006). The algorithm updates the noise estimate in each time frame using a time-frequency-dependent smoothing factor that varies with the probability that speech is present. Based on the noise estimate made by the MCRA2 algorithm, the signal-to-noise ratio (SNR) of the input signal was estimated using the decision-directed approach (Ephraim and Malah 1984). This method estimates the SNR in each time-frequency unit as a weighted average of the estimated SNRs of the previous and current frames.

Frequency channels

We smoothed the SNR estimate across frequencies to reduce high contrasts in gain. The original algorithm calculated time-frequency units by fast Fourier transformation (FFT) on 20-ms Hamming-windowed segments, with a 50% overlap between the segments. All sound files had a sample rate of 44.1 kHz, leading to an FFT size of 882 samples per frame. We averaged the noise-estimates over several frequency-bins to obtain 15 different frequency channels, logarithmically divided over frequencies between 50 and 8000 Hz.

Gain function

Based on the estimated SNR, the gain function determines the amount that each time-frequency unit will be attenuated. Our gain function consisted of two parts: (1) for SNRs lower than 0 dB, a fixed attenuation was applied (A_{max}) and (2) for SNRs higher than 0 dB, the linear attenuation G_k in each frequency band (k) was determined from the estimated signal-to-noise ratio ($\hat{\varepsilon}_k$, also linear) by a parametric Wiener filter (Lim and Oppenheim, 1979):

$$G_k = (\frac{\hat{\varepsilon}_k}{1+\hat{\varepsilon}_k})^{\beta} \tag{7.1}$$

In order to let the two parts of the gain function correspond to each other at 0 dB SNR, parameter β was related to A_{max} as follows (with A_{max} in dB):

$$\beta = \frac{-A_{max}}{20\,log10(0.5)} \tag{7.2}$$

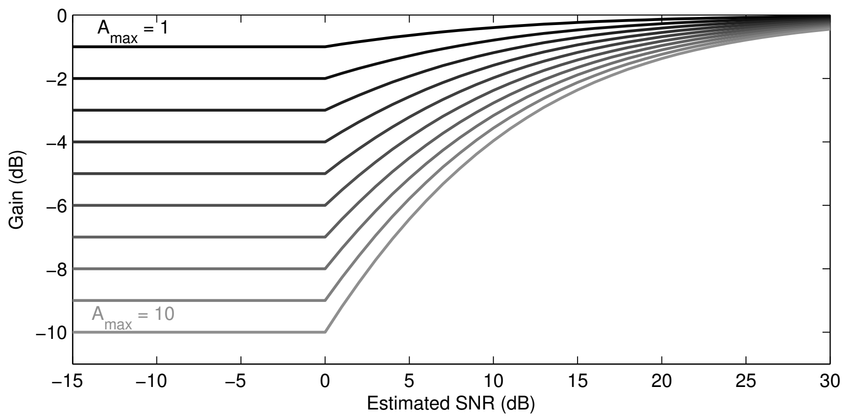


Figure 7.3: Noise-reduction gain as a function of the estimated signal-to-noise ratio (SNR) for different values of A_{max} (the variable determining the noise-reduction strength). Time-frequency units dominated by noise (SNR $<$ 0 dB) were attenuated with the maximum strength (A_{max}), whereas the attenuation for speech-dominated time-frequency units (SNR $>$ 0 dB) was determined by the parametric Wiener filter.

Figure 7.3 shows the resulting attenuation as a function of estimated SNR for different values of A_{max} (all in dB).

The decision-directed approach allows the gain function to influence the SNR estimate so that changing the noise-reduction strength also changes the estimated SNR. For our experiment, this was undesirable because the same SNR estimate for each value of A_{max} was required, to ensure that the attenuation pattern remains fixed between conditions and only the depth of attenuation differs. We therefore estimated the SNR only once (see Figure 7.2) and froze the result. This frozen SNR estimate was then used as input for the gain function for each value of A_{max}. This frozen SNR estimation was obtained with a value of β = 0.5 in the gain function (i.e. the square-root Wiener filter (Loizou 2007)).

Figure 7.4 shows the effect of applying the attenuation on the complete speech-in-noise time signal and the corresponding time-frequency attenuation pattern for one of the stimuli with different values of A_{max} (and corresponding values of β).

Normally, the speech and noise are mixed before they enter the noise reduction and it is not possible to obtain separate estimates of the speech and noise components after processing. However, by applying the recorded attenuation (Figure 7.4) separately to the speech and the noise signals, we could artificially obtain the combinations that are required to separate the perceptual effects of speech distortion and residual noise.

Level corrections

Because noise reduction alters the sound signal, the loudness of signals after processing can differ from that of the input signals. To reduce loudness as a possible cue for

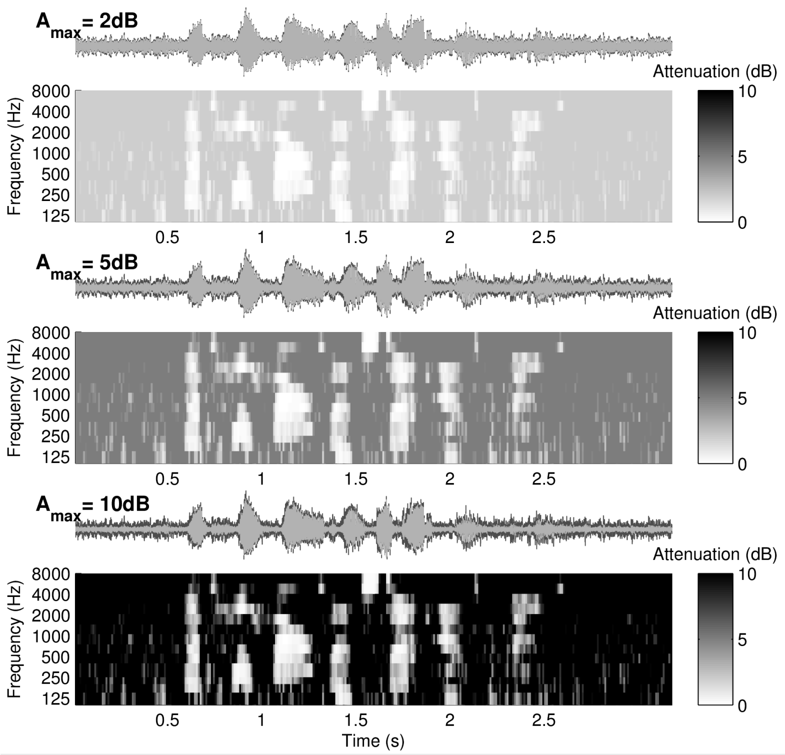


Figure 7.4: Effect of noise reduction for one sentence in stationary noise at an SNR of +5 dB for different values of the maximum attenuation A_{max}. The time signals in dark gray show the input speech and noise signal and light gray curves show the output of the noise-reduction algorithm. The spectrogram-like plots show the attenuation pattern determined by the noise-reduction algorithm with the time-frequency units that were not attenuated in white, the units that were attenuated with 10 dB in black and the units with an attenuation between 0 and 10 dB in gray.

detection in the listening experiment, we removed level differences with the loudness model (Glasberg and Moore, 2002). First, the loudness of the unprocessed speech was estimated using the loudness model. We then amplified the processed speech signal until it had the same calculated loudness as the input signal (difference $<$ 0.1 phons). In the scheme in Figure 7.2 the resulting signal is denoted “processed speech”. The same amplification was then used for the noise after noise reduction, so that the output SNR of the noise reduction was preserved (“processed noise”, Figure 7.2)

Because we only wanted to study the negative effects of noise reduction (i.e., distortion) and not the positive effects (i.e., reduction of noise level), we changed the level of the reference noise (“rms matched noise”, Figure 7.2), to match the level of the noise after noise reduction (“processed noise”).

To further avoid effects of potential cues from remaining small differences in loudness between stimuli, we applied level roving (see description of the procedure for threshold measurements).

7.2.4 Processing conditions for threshold measurements

We determined the detection thresholds for noise-reduction distortion in the complete speech-and-noise signal, and in the speech signal and noise signal separately. For each threshold that was determined, a reference stimulus and a target stimulus were required to perform the detection task (see description of the procedure for threshold measurements). Table 7.1 summarizes the formation of these reference and target stimuli by combinations of the four conditions that resulted from processing (see the right part of Figure 7.2: original speech, processed speech, processed noise and rms matched noise).

Table 7.1: Overview of the combinations of different speech and noise signals used to obtain the target and reference stimuli for the measurement of three different types of distortion.

Threshold to determine	Reference stimulus	Target stimulus
Detection of overall distortion	Original speech	Processed speech
	rms Matched noise	Processed noise
Detection of speech distortion	Original speech	Processed speech
	No noise	No noise
Detection of noise distortion	Processed speech	Processed speech
	rms Matched noise	Processed noise

To determine the detection threshold for overall distortion, the target condition was the combination of processed speech and processed noise, with A_{max} as the adaptive variable. The reference stimulus was the combination of original speech and rms matched noise. Thus, the target and reference had equal speech and noise levels (apart from level roving) but differed in the amount of distortion in both speech and noise.

To determine the threshold for speech distortion, we omitted the noise and compared only the original speech (reference) with the processed speech (target). Again, the adaptive variable was A_{max}, which determined the amount of distortion of the processed speech.

To determine the threshold for noise distortion, the target contained processed noise and the reference unprocessed noise that had the same rms level (rms matched noise). In both the reference and the target, we added processed speech to the noise (see Table 7.1). The speech processing was equal for target and reference (with the same A_{max} as the noise in the target signal), so that the only difference between target and reference was a difference in noise processing. The reason for adding the speech to the noise in this condition was to prevent the participant from listening to the "ghost image" of the removed speech in the processed noise. Such a "ghost image" of speech arises because the noise-reduction algorithm sets the attenuation to zero dB in time-frequency units where speech is present. This lack of attenuation during speech is automatically applied to the noise. By adding the speech, we were able to mask the ghost image and determine the detection threshold for distortions from changes in the noise signal only.

7.2.5 Procedure for threshold measurements

The detection threshold was determined with an adaptive up-down procedure that estimates the 79.4% correct response level (Levitt 1992). For each presented distortion level in this procedure, we used a three-interval two-alternative forced-choice paradigm to determine whether the participant could detect the distortion. Subjects were asked to choose which of the last two sound samples was equal to the first sample. They were instructed to concentrate on differences in distortion rather than differences in loudness. Each run of the adaptive procedure started with a relatively high value for A_{max} (17 dB). First, A_{max} was decreased by 3 dB after each correct response and increased by 3 dB after each incorrect response (one-up one-down procedure). After three turn points, the one-up one down-procedure was changed to one-up three-down: A_{max} was decreased by 2 dB after three successive correct responses and increased by 2 dB after each incorrect response. This procedure continued until six turn points had occurred in the one-up three-down procedure. The detection threshold was defined as the mean of the values of A_{max} at the last six turn points, corresponding to 79.4% correct response (Levitt 1992). If the standard deviation of the values at the six turn points of a run exceeded twice the stepsize (2 x 2 dB), we considered that particular measurement to be unreliable and discarded the threshold from the dataset.

Detection thresholds were measured twice (test and retest) for each of the three measurement conditions (overall distortion, speech distortion, and noise distortion). The first run for each condition (test) was preceded by 12 adaptively presented training sentences. For the second run (retest) only three training sentences were used. The maximum value of A_{max} was 30 dB. For small values of A_{max} the stepsize was reduced. When the adaptive procedure required A_{max} values below 2 dB, the stepsize was decreased to $A_{max}/2$, with a minimum stepsize of 0.125 dB.

To reduce the effect of possible remaining differences in loudness between stimuli, we applied level roving. The reference stimulus was at a random amplification or attenuation of 0 dB, 1 dB, or 2 dB.

7.2.6 Measurement of preferred noise-reduction strength

We determined the individual preference of subjects for noise-reduction strength using a procedure in which the participants could manually adjust the strength for noise reduction. Whereas during the detection task the subjects were presented with the negative effects of noise reduction (i.e. distortion) only, they could now hear both the negative (distortion) and positive (reduction of noise level) effects. To familiarize the participants with both effects of noise reduction, they had to listen to a training sentence in noise twice: first without processing and then with the maximum noise-reduction processing (A_{max} = 30 dB). At this strength both the reduction in noise level and the distortion of speech were clearly audible for all subjects. After this example, the same sentence in noise was presented with a lower level of noise-reduction strength. Each sentence started at a different level of A_{max}, with levels set to 3, 6, 9, 12 and 15 dB. Subjects could now adjust the noise-reduction strength to the level that they preferred. This was achieved with two buttons: one for "more noise reduction" and one for "less noise reduction". Each time they pressed a button the level of A_{max} was increased or decreased with 3 dB and the sentence was presented again at the new noise-reduction strength. Subjects were asked to search for the noise-reduction strength that they would prefer for prolonged listening. This procedure was repeated for each of the five unique sentences that were used during the detection task.

All stimuli consisted of the processed speech embedded in the processed noise (see Figure 7.2). Note that the processing resulted in equal loudness of speech for all stimuli (see Figure 7.2), but that the level of the residual noise was lower for higher values of A_{max}.

7.3 Results

7.3.1 Q1. Detection threshold for overall distortion and hearing impairment

Three measurements (of a total of 48 measurements) for the detection threshold for overall distortion were discarded because the standard deviation of the turn points exceeded the reliability criterion of 4 dB. For each of these measurements the corresponding test or retest measurement was available and thus there was at least one reliable data point for each combination of condition and subject.

Figure 7.5 (left) shows the average detection thresholds for both groups for overall distortion and the corresponding 95% confidence intervals among subjects. The average detection threshold for overall distortion was 2.3 dB higher for the hearing-impaired subjects (6.7 dB) than for the normal-hearing subjects (4.4 dB). This difference was statistically significant (t-test, $p = 0.04$).

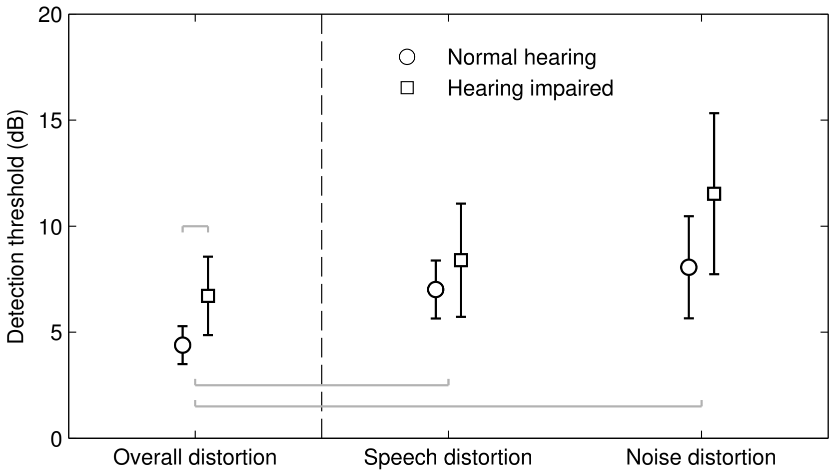


Figure 7.5: Detection thresholds for noise-reduction induced signal distortions averaged over all subjects per group. Error bars show the 95% confidence interval among subjects.

There was no significant correlation between the detection threshold for overall distortion and the hearing threshold of the individual, summarized as the pure tone average (PTA) of 0.5, 1, 2 and 4 kHz (Pearson's correlation coefficient $r = 0.40$, $p = 0.052$).

7.3.2 Q2. Detection thresholds for speech and noise separately

Seven measurements for the threshold for speech distortion and nine measurements for the threshold of noise distortion were discarded because the standard deviation of the turn points exceeded the 4 dB reliability criterion, and one measurement for the noise-distortion threshold was discarded because the adaptive procedure reached the ceiling level for A_{max} of 30 B. For each of these measurements the corresponding test or retest measurement was available and thus there was at least one reliable data point for each combination of condition and subject.

The group-averages for the detection thresholds for speech distortion and noise distortion are shown in the right hand side of Figure 7.5. Average thresholds for speech distortion and noise distortion were higher than the threshold for overall distortion according to paired t-tests after Bonferroni correction for 3 comparisons (uncorrected $p = 0.01$ for the difference between speech distortion and overall distortion, and uncorrected $p < 0.01$ for the difference between noise distortion and overall distortion). The thresholds for speech and noise distortion did not differ from one another (uncorrected $p = 0.092$).

Correlation analysis revealed that thresholds for speech distortion and noise distortion were both correlated with those for overall distortion (Pearson's correlation coefficient $r = 0.68$, $p < 0.001$ for the speech distortion and overall distortion, and $r = 0.46$, $p =$

0.023 for noise distortion and overall distortion), but that thresholds for speech distortion and noise distortion were not significantly correlated to each other (Pearson's correlation coefficient $r = 0.31$, $p = 0.143$).

Further analysis revealed that the differences between normal-hearing and hearing-impaired listeners were not significant for speech distortion or for noise distortion thresholds (t-test, $p = 0.38$ for speech distortion and $p = 0.15$ for noise distortion).

7.3.3 Q3. Preferred noise-reduction strength

Figure 7.6 shows the individually preferred A_{max} values plotted against the individually measured detection thresholds for overall distortion. All data points fell above the diagonal, showing that for each subject the preferred noise-reduction strength was above the detection threshold for distortion. Detection threshold and overall preference were not significantly correlated (Pearson's correlation coefficient $r = 0.54$, $p = 0.07$ for normal-hearing subjects and $r = 0.53$, $p = 0.08$ for hearing-impaired subjects).

Figure 7.7 shows the group-average preferred A_{max} (left panel, "preferred A_{max}") and the difference between the preferred A_{max} and the individual detection threshold for overall distortion (right panel, "relative preferred A_{max}"). The absolute preferred noise-reduction strength did not significantly differ between normal-hearing and hearing-impaired subjects (t-test: $p = 0.25$). The average relative preference was significantly higher for normal-hearing listeners than for hearing-impaired listeners (t-test, $p = 0.02$).

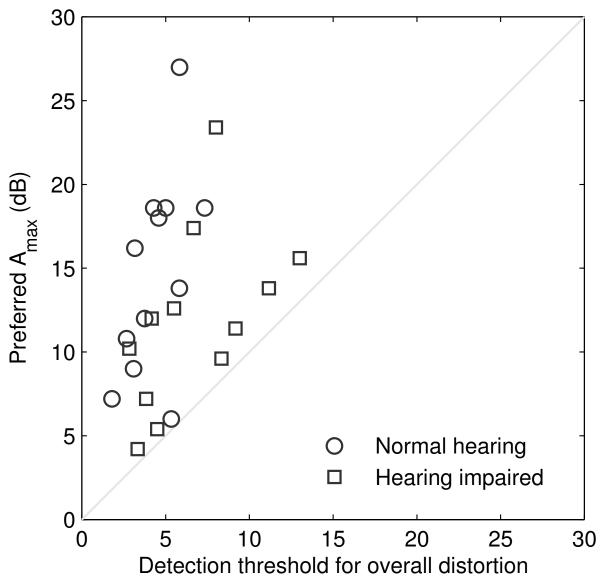


Figure 7.6: Individually preferred noise-reduction strength plotted against the individual threshold for the detection of overall distortion.

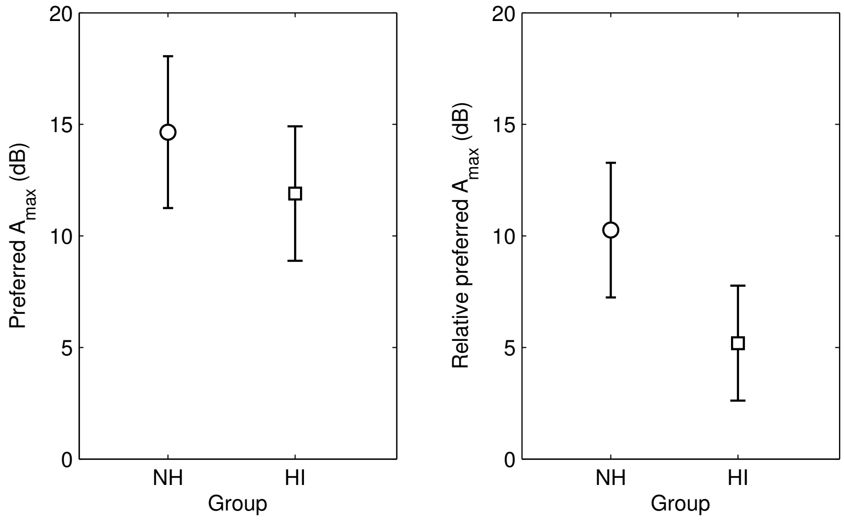


Figure 7.7: Average preferred value for A_{max} for both subject groups ("preferred A_{max}", left panel) and the difference between that preferred value and the detection threshold ("relative preferred A_{max}", right panel). Error bars show the 95% confidence interval among subjects.

7.4 Discussion

7.4.1 Q1. Detection threshold and hearing loss

The detection threshold for distortion caused by noise reduction was higher for hearing-impaired subjects than for normal-hearing subjects, which implies that hearing-impaired listeners are less sensitive to distortions caused by noise reduction than normal-hearing listeners.

Marzinzik (2000) found differences between normal-hearing and hearing-impaired subjects in their judgments on speech naturalness after noise-reduction processing and suggested that this was because hearing-impaired subjects were less able to notice the speech distortions due to noise reduction. This was, however, based on results from quality judgments only and Marzinzik did not measure objectively whether subjects could hear distortions. Our results show that hearing-impaired subjects indeed had a higher threshold for distortion, but it is not unlikely that the hearing impaired in the study of Marzinzik would have been able to detect distortions. In the subjective experiment other cues could have been dominating judgment of the hearing-impaired subjects regarding naturalness.

On average, hearing-impaired subjects could detect distortion for an A_{max} of 6.7 dB. Given that the threshold measured in the present study represents a just noticeable difference that is obtained by direct comparison with an undistorted reference, it is expected that the distortion at threshold are not annoying for the listener. Noise reduction in hearing aids is generally limited to a maximum attenuation in the order of a magnitude of 10 to 20 dB for their strongest settings (Chung 2004). Several algorithms also provide the possibility to set a lower value for the maximum gain reduction. Thus, if noise reduction is set to its maximum strength in hearing aids this may cause audible distortions to the speech signal. Of course this is also dependent on other settings of the noise reduction, such as the time constants (see Chapter 1). Additionally, the amount of distortion depends on the type of input noise and the input SNR (Houben et al. 2011; Chapter 3). For lower SNRs, the noise-reduction algorithm is likely to make more errors in estimating the SNR, therefore causing more distortions to the signal. Thus, for lower input SNRs the detection threshold for distortion is expected to be higher. Indeed, preference results of normal-hearing listeners for hearing aid noise reduction showed that degradation in speech naturalness was a more determining factor for the preference at low input SNR (-4 dB) than at higher input SNR (+4 dB).

The weak relationship between detection thresholds and pure tone average thresholds implies that hearing threshold is not the main factor in determining the detection threshold for speech distortion. This is not surprising, because the NAL-RP amplification compensated for hearing threshold, so that audibility is not expected to be the main factor determining the detection threshold. A more plausible cause is a difference in suprathreshold processing, for example temporal and frequency resolution and modulation detection.

7.4.2 Q2. Detection threshold for speech and noise separately

The detection thresholds for the noise-reduction attenuation pattern applied to speech (speech distortion) and to noise (noise distortion) did not differ from one another but were both higher than the detection threshold for overall distortion. This implies that the effects of noise reduction on speech and on the remaining noise both contribute to the perception of overall distortion: if the detection of overall distortion was mainly determined by one type of distortion, the threshold for that type would not differ from that for overall distortion.

Thresholds for speech distortion and for noise distortion were both significantly correlated to the threshold for overall distortion, but not to each other. This indicates that both speech distortion and noise distortion correspond to different aspects of overall distortion, with speech distortion as the main factor determining the overall distortion, but noise distortion also contributing.

Inter-subject variability was higher for the detection of speech and noise distortion than for the detection of overall distortion. This caused an insignificant difference between the normal hearing and hearing impaired groups. However, the trend of higher thresholds for hearing-impaired listeners for overall distortion was also present in speech distortion and noise distortion, albeit not significantly (Figure 7.5). It remains possible that a large number of subjects would also reveal a significant effect of hearing loss on detection thresholds for speech distortion and noise distortion.

7.4.3 Q3. Preference for noise-reduction strength and detection threshold

As expected, the preferred noise-reduction strength for each individual was higher than or equal to the individual's detection threshold for distortion (see Figure 7.6). As long as the distortion is not audible the effect of noise reduction is a decrease in noise level, whereas the speech remains undistorted. Thus, although we can not directly use the detection threshold for predicting the preferred strength for noise reduction (as correlations were low), it provides important information on the amount of noise reduction that can be applied without being disadvantageous for an individual. Given that we did not measure the sensitivity of subjects to a reduction in noise level in this study, we cannot discriminate whether noise reduction was indeed positive or just neutral. Here it is important to note that we measured preference after the processed signal was amplified to correct for possible reduction in speech level due to noise reduction. If such a correction was not applied, as is often the case in hearing aids (see the acoustical analyses in Chapters 3 and 5), the preferred noise-reduction strength may be lower because noise reduction also affects the speech level.

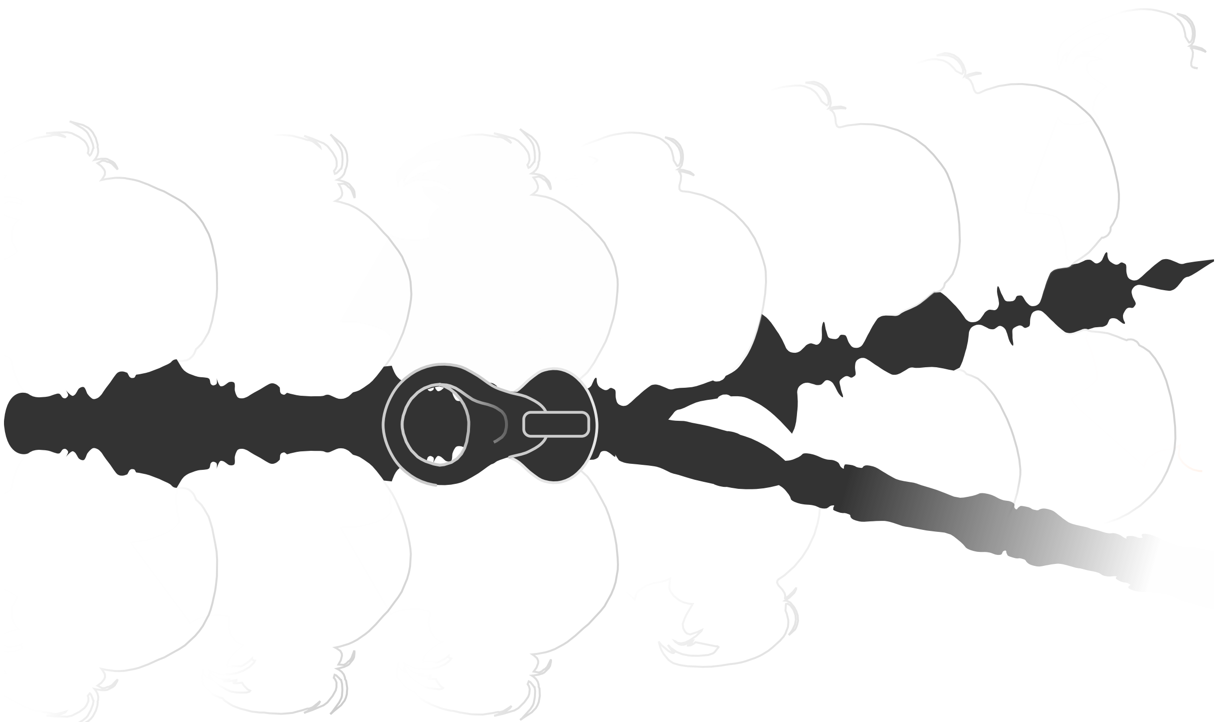
The group-average preferred noise-reduction strength did not differ between normal-hearing and hearing-impaired subjects (Figure 7.7, left panel), which is also in line with previous findings (Marzinzik 2000; Anderson et al. 2009; Luts et al. 2010; Houben et al. 2011). The higher detection thresholds for hearing-impaired subjects did not result in higher preferred noise-reduction strengths. In fact, the preferred noise-reduction strength seemed even lower for hearing-impaired listeners, although this was not significant. A lower mean preferred noise-reduction strength in hearing-impaired listeners combined with a higher threshold, lead to a significant difference in the relative preference between normal-hearing and hearing-impaired subjects. Thus preferred strength of hearing-impaired subjects was closer to their detection threshold than that of normal-hearing subjects (Figure 7.7, right panel). Our interpretation of these findings is that, although stronger noise reduction can be applied for hearing-impaired listeners before distortions are audible, they seem to tolerate fewer distortions than normal-hearing listeners, once distortions are audible. This implies that avoiding speech distortion is more important for hearing-impaired subjects than for normal-hearing subjects, probably due to their impaired ability to understand speech.

Inter-individual differences in preferred noise-reduction strength were high within both subject groups (see Figure 7.6; individual preferred values for A_{max} ranged from 4.2 to 27.0 dB), which is in line with previous findings (Houben et al. 2011). This broad range of preferred settings for A_{max} implies that it would be useful to have the possibility to adjust the noise-reduction strength in hearing aids over a wide range of settings. The individually preferred trade-off between speech distortion and residual noise seems to depend not only on audibility of distortions, but also on how well an individual tolerates the presence of noise. Research into the acceptable noise level (ANL) has shown that noise tolerance is not related to hearing loss, indicating that the acceptance of noise is an individual characteristic (Mueller et al. 2006; Nabelek et al. 2006).

7.5 Conclusions

In this study a method was developed and applied to determine the detection threshold for the negative effects of noise reduction (i.e., distortion of the signal) independently from the positive effects (i.e., reduction of noise level). The results show that detection thresholds for distortion were higher for hearing-impaired subjects than for normal-hearing subjects. This suggests that stronger noise reduction can be applied for hearing-impaired listeners without negative effects on the perceived sound quality. However, the preferred noise reduction strength of hearing-impaired listeners was closer to their individual detection threshold for distortion than that of normal-hearing listeners, implying that hearing-impaired listeners accept fewer audible distortions than normal-hearing listeners do. The overall effect was that the preferred noise-reduction strength did not differ between the groups, but the weighting of underlying causes (audibility and acceptability of distortions) did differ between the groups.

8 Summary and general discussion



Difficulty to understand speech in noisy situations is the number-one complaint of hearing aid users (Kochkin 2002). Hearing aid manufacturers take measures against the problem of speech in noise by implementing signal-processing algorithms that should reduce background noise. The most widely applied measure against noise is single-microphone noise reduction, which has the task to estimate from a single input signal whether it does contain noise or not and to reduce the hearing aid gain accordingly, without affecting speech if present.

Given that there are about 650 000 hearing aids in use in the Netherlands of which the vast majority contains single-microphone noise reduction, it is remarkable that there is only little knowledge on the implementation and effects of noise reduction. Clinicians who are responsible for prescribing and fitting the hearing aids have no insight in the consequences of activating noise reduction in a hearing aid. Thus, for thousands of hearing aid users it is not sure whether their hearing aid is chosen and fitted optimally to compensate for the wide-spread problem of reduced speech perception in noise.

This thesis describes several studies that were conducted to learn more about the effects of different single-microphone noise-reduction algorithms on perceptual outcomes, such as speech intelligibility, listening effort, and personal preference. The major part of the thesis (Chapters 2 to 5) comprises studies investigating noise-reduction implementations from commercial hearing aids. Chapters 6 and 7 describe additional studies that used state-of-the-art noise-reduction algorithms from literature to deepen our knowledge on some specific aspects of noise-reduction.

Chapter 1: Introduction

Chapter 1 explains the basic principles of noise reduction. Noise-reduction processing usually consists of two parts. First, the environment should be classified in being noise, speech or speech in noise. This step generally results in an estimation of the actual signal-to-noise ratio (SNR) per frequency channel. The second step is to adjust (or retain) the hearing-aid gain based on the estimated SNR, where the main challenge is to find an optimal trade-off between reducing noise and retaining speech quality. Many different approaches are possible for both steps, so that there is not one representative noise-reduction algorithm. Conclusions from investigations with one algorithm thus can not be easily generalized to other algorithms. For that reason we used in this thesis different noise-reduction algorithms so that we could also compare between *algorithms* instead of only evaluating the effect of switching one algorithm *on* and *off*.

Chapter 2: Method to compare noise reduction in hearing aids

Chapter 2 describes how we developed and validated a method to allow comparison of noise-reduction algorithms from different hearing aids, without confounding effects due to other differences between hearing aids. An inverse filter was made for each hearing aid to correct for differences in frequency response that remained despite careful adjustment of the hearing aid gain. When that filter was applied on all recordings of that hearing aid, normal-hearing subjects could no longer differentiate recordings from different hearing aids from one another if all hearing aids were linearly fitted with signal processing turned *off*. Once a filter is made for a specific hearing aid, the filter can also be applied on recordings of that hearing aid with noise reduction switched *on*. This allows listening to the effect of noise reduction in isolated form, without confounding effects of other processing. The method for hearing-aid recording and inverse filtering forms the basis of the next three chapters, where noise-reduction implementations from different hearing aids were perceptually compared.

Chapter 3: Noise reduction in linear hearing aids (normal-hearing listeners)

Chapter 3 describes the perceptual comparison of noise reduction from four different linearly fitted hearing aids. Noise reduction appeared to differ between hearing aids in the degree they reduce the noise annoyance and speech naturalness perceived by normal-hearing listeners. Speech naturalness was only degraded at lower SNR (-4 dB), where speech was more difficult to detect for the noise reduction. Intelligibility scores and listening-effort ratings differed not between noise reduction *on* and *off* and only slightly among noise-reduction algorithms. Preference differed both between noise reduction *on* and *off* and between algorithms. The preference results confirm that results on noise-reduction effects of one algorithm can not be generalized to other algorithms. Aside from differences between noise-reduction algorithms, the results reveal differences between individual subjects in their preferred weighting of noise annoyance and speech naturalness.

Chapter 4: Noise reduction in linear hearing aids (hearing-impaired listeners)

Chapter 4 describes the perceptual comparison of noise reduction as in Chapter 3, but now with hearing-impaired subjects and with three instead of four hearing aids. The results of hearing-impaired subjects agreed well with those of normal-hearing subjects in Chapter 3. Intelligibility was slightly reduced by the noise reduction that had most positive ratings for noise annoyance, speech naturalness, and overall preference. These findings confirm the commonly assumed trade-off between intelligibility and listening comfort. Analysis on the individually preferred trade-off between noise annoyance and speech naturalness as was done in Chapter 3 was not possible with the data in this chapter, because stimuli for the hearing-impaired subjects were at higher SNRs where the speech naturalness was not perceptually degraded by the noise-reduction algorithms evaluated.

Chapter 5: Noise reduction in compressive hearing aids

Chapter 5 describes measurements with the same three hearing aids as in Chapter 4, but now we studied noise reduction in a compressive setting, which is more often applied in hearing aids than linear gain. Acoustical analyses showed that noise reduction reduced gain during noise and compression during peaks in the speech signal, so that the combined processing of noise reduction and compression reduced both the speech and the noise level for our input stimuli with an SNR of +4 dB. Thus, compression tended to have a negative effect on the SNR for our input stimuli, which was compensated for by noise reduction. Perceptual measurements with hearing-impaired subjects revealed that the combination of noise reduction and compression did not influence intelligibility for the hearing aids tested. The combined processing of noise reduction and compression was for none of the hearing aids significantly preferred over no processing, in contrast to noise reduction without compression (see Chapter 4, where noise reduction in linear aids was significantly preferred over unprocessed for two hearing aids from the same test set). This shows that noise reduction for application in hearing aids should be developed and tested not only in isolation, but also in combination with compression.

Chapter 6: Ideal and non-ideal noise reduction

Chapter 6 describes perceptual measurements with four different variations of a noise-reduction algorithm. Two conditions had an ideal noise-tracker, which means that they had access to the separate speech and noise input signal. One of these conditions completely removed noise-dominated time-frequency units (“ideal binary mask”), whereas we limited attenuation in the other to a maximum of 10 dB (“tempered mask”) which is more comparable with the amount of gain reduction that is applied in hearing aid noise reduction algorithms. The other two conditions consisted of different real noise estimators, combined with the same “tempered” attenuation function. The ideal conditions revealed a trade-off between quality and intelligibility in relation to the attenuation strength: infinite attenuation improved intelligibility the most, but reduced sound quality due to the large fluctuation in gain that was perceived as sound distortions. Non-ideal noise reduction showed errors in the estimation of speech and noise and therefore did not improve intelligibility. However, it reduced the annoyance caused by the noise and was therefore preferred by the subjects over no processing. Thus, improving the subjective benefit while preserving speech intelligibility might be a more realistic goal in the development of noise reduction than trying to improve intelligibility at the cost of the subjectively perceived benefits.

Chapter 7: Detection thresholds for distortion from noise reduction

Chapter 7 describes a study in which we further explored the trade-off between reduction of noise and distortion of the signal, which is inherent to noise reduction. In this study we artificially separated both effects to measure the detection threshold for distortions, without possible confounding effects of reductions in noise level. The detection threshold was higher for hearing-impaired subjects than for normal-hearing subjects. This implies that for hearing-impaired listeners stronger noise reduction can be applied without affecting the perceived sound quality than for normal-hearing listeners. However, the preferred noise-reduction strengths of hearing-impaired listeners were closer to their individual detection thresholds for distortion than for normal-hearing listeners. This shows that, once distortions are audible, hearing-impaired listeners tolerate fewer distortions than normal-hearing listeners. On average, normal-hearing listeners and hearing-impaired listeners preferred the same noise-reduction strength, but individual differences were large within both subject groups.

Considerations for implementation and presentation of noise reduction in hearing aids

Given our results that noise-reduction implementations differ perceptually between hearing aids (Chapters 3 to 5), from a clinical point of view it is undesired that noise-reduction implementations are commonly presented as a "black box". Clinicians need at least information on the implemented maximum amount of gain reduction, time constants, and frequency weighting of the gain (see Chapter 1) as well as the philosophy behind the chosen parameter settings. Additionally, it is desirable to have more options to adjust noise-reduction parameters within a selected hearing aid for an individual listener. At least the maximum amount of gain reduction should be manually adjustable over a wide range of settings, because individual listeners appeared to differ largely in their most preferred noise-reduction strength (Houben et al. 2012 and Chapters 3 and 7 of this thesis).

The perceived benefit of noise reduction may be influenced by the expectations of the listener. For both the manufacturers who market the hearing aids as well as for the clinicians who inform the hearing aid user, it is important to raise realistic expectations. Especially hearing-impaired listeners getting their first hearing aid should be aware that amplification will also increase audibility of annoying sounds and that noise reduction will not completely remove these sounds (Palmer et al. 2006).

Considerations for selection and fitting of noise reduction in hearing aids

The hearing aid specific results in Chapters 3 to 5 should not be used to draw conclusions on which of the hearing aids tested is best. First, the number of conditions studied was limited, and noise-reduction effects are known to differ between noise types and SNRs. Second, all measurements were done in a laboratory setting and listeners' judgements were based on short sound fragments instead of prolonged listening. Third, we studied noise reduction in isolation or in combination with compression, thereby ignoring all other hearing aid processing features that may play a role in daily use. Fourth, even in the same situation a listener may differ from other listeners in which type of processing he/she prefers.

Nevertheless, the results of Chapters 3 to 5 indicate consistently that noise reduction differs perceptually between hearing aids and that preference for noise reduction differs between listeners. This implies that more attention should be paid to tune the noise reduction to the individual hearing aid user. Because the current noise-reduction implementations barely provide the possibility to adjust settings to fine-tune noise reduction to an individual user within a hearing aid, the choice between hearing aids for an individual becomes relatively critical with respect to the type of noise reduction implemented. However, the lack of knowledge on the implementation causes that noise reduction rarely plays a role in the selection of a hearing aid. But at least for listeners with specific complaints in noisy environments, it is worthwhile to perform technical and perceptual comparisons in order to select the best noise-reduction system for the individual listener and to fine-tune the noise-reduction parameters in a systematic way.

Technical comparison of noise reduction in hearing aids

Technical measurements are required to obtain information on the implementation of noise-reduction parameters if those are not provided in the specifications. The recording method that was introduced in Chapter 2 and applied in subsequent chapters provides a useful tool to learn about noise-reduction characteristics (see Chapters 1, 3 and 5) and to listen to its effects (Chapters 3 to 5). Such measurements can be the basis for perceptual comparisons, e.g. in a master hearing aid (Levitt et al. 1986; Grimm et al. 2006), see below.

A less time-consuming and more easily available method for clinical application is to perform probe-microphone measurements with broad-band stimuli either in a 2cc coupler or in the listener's ear (McCreery et al. 2010). The differences in hearing aid output spectrum between noise reduction *on* and noise reduction *off* show the noise-reduction activity of the hearing aid. When the input signal is stationary noise, the spectral differences show the maximum gain reduction by noise reduction, its frequency-dependency, and the time constant for entering a noise-only situation. Additionally, an input signal of speech (ISTS-signal, Holube et al. 2010) or simulated speech (ICRA signal, Dreschler et al. 2001) in noise reveals if noise reduction remains active during the presence of speech. If noise reduction still reduces gain when speech is present, one should be aware that the audibility of speech may be reduced. Although not exhaustive, such verification provides a quick and useful glance in the black box of hearing aid noise reduction. Ideally, the verification should be done after fitting the hearing aid gain and compression according to the users' individual hearing loss, for compression can influence the efficiency of noise reduction and vice versa (see Chapter 5).

Perceptual comparison of noise reduction in hearing aids

Because not only the noise-reduction implementations differ but also the preference of individual listeners, several noise-reduction systems and settings should be presented to the hearing aid user to gain insight in his personal preferences. Direct comparison of noise reduction *on* and *off*, and if available of different settings for noise reduction within a hearing aid can be done during the hearing aid fitting. However, it is not practical to switch between hearing aids for direct comparison between different noise-reduction systems. Perceptual comparison between noise-reduction algorithms thus requires another approach, for instance a master hearing aid in which different characteristic algorithms are implemented, or in which one algorithm is implemented with the possibility to change a range of parameter settings. Paired comparisons of settings then should reveal for instance whether the listener prefers weak or strong noise reduction, quick or slow gain adjustments and frequency-specific or overall gain reductions. This information can be combined with that obtained with technical measurements as described above, in order to find the best noise-reduction system and settings for an individual listener.

Considerations for future research on noise reduction in hearing aids

The results of the studies described in this thesis imply that there is room for improvement in the use of single-microphone noise reduction to optimize the subjective perception of speech in noise for individual listeners. Additional research is required to clarify the many uncertainties that remain about noise reduction. It is more and more recognized that single-microphone noise reduction does not improve intelligibility but that it might provide benefit in terms of listening effort and listening comfort. It is therefore recommended for future research to evaluate more thoroughly how existing noise-reduction algorithms influence the sound quality in different situations rather than adding new variations to the large amount of existing algorithms in attempts to improve intelligibility. A related research goal is the development of test methods that can objectify perceptual benefit of noise-reduction algorithms.

Evaluation of existing noise reduction

Investigating all available hearing aid noise-reduction implementations extensively is impossible, thus there is a need for a method to somehow generalize among algorithms (Hoetink et al. 2009). The results described in this thesis (especially in Chapters 2 and 6) as well as previous results (Houben et al. 2012) suggest that the trade-off between noise reduction and speech distortion might be a promising approach to characterize noise-reduction algorithms as well as listeners. In Chapter 7 we made a start in investigating this trade-off. The preference task that we used appeared to be a quick and easy way to characterize an individual listener as a “noise hater” (preferring strong reduction of noise, irrespective the consequences for speech quality), as a “distortion hater” (preferring original speech quality, irrespective the presence of background noise), or as someone weighting both aspects more equally, or someone with no explicit preference. To characterize noise-reduction algorithms in terms of the trade-off between noise reduction and speech distortion, the separate effects of noise reduction and distortion should be analyzed. If the implementation of the noise-reduction algorithm allows the recording of the gain, this can be done in the same way as we did in Chapter 7, by applying the gain to the speech and noise signals separately. For the black-box algorithms in hearing aids, the Hagerman and Olofssen (2004) method that we applied in Chapter 5 can be used to separate between the effects on speech and noise.

A systematic evaluation of the effect of specific noise-reduction parameter settings is worthwhile, but impossible within hearing aids since the actual noise-reduction implementations do not allow changing parameter settings in the extent that is required. Noise-reduction algorithms that are described in detail in literature (see for instance Loizou 2007 for several noise-reduction implementations) can be used for the purpose of evaluating parameter settings. Although implementation in hearing aids is often mentioned as a potential application of such algorithms, the state-of-the-art noise-reduction algorithms from literature generally differ largely from the noise reduction implementations in hearing aids. To make the results of such an evaluation more applicable for hearing aids, the maximum amount of gain should be limited and the number of frequency channels should be reduced (see Chapters 6 and 7). Additionally, the possible influence of compression on the noise-reduction effect should be considered to increase applicability of results for realistic use in hearing aids (Chapter 5).

To assist in the evaluation and development of noise-reduction algorithms, one would like to be able to predict the effect of noise reduction on sound quality and intelligibility without the need for time-consuming listening experiments. To achieve this goal, several objective measures were developed that estimate changes in quality or intelligibility by comparing the processed signal with the original speech signal. Examples of such measures are the Extended Speech Intelligibility Index (ESII, Rhebergen et al. 2005) and the Coherence Speech Intelligibility Index (CSII, Kates and Arehart 2005) for predicting speech intelligibility. Examples of models for predicting sound quality are the Hearing Aid Speech Quality Index (HASQI, Kates and Arehart, 2010; see also Chapter 2), the Perceptual Evaluation of Speech Quality (PESQ, Rix et al. 2001), and PEMO-Q (Huber and Kollmeier 2006). Current modeling attempts, however, were not developed and evaluated specifically for the goal of evaluating noise-reduction processing. Some studies explored the performance of a number of these models on predicting noise-reduction effects, with diverging results (Marzinzik 2000; Loizou 2007; Taal et al. 2009). As a result, it is currently not possible to reliably predict the perceptual effects of noise reduction, and listening experiments remain required. Further development and evaluation of objective models is required to make them more appropriate for evaluation of noise-reduction effect, taking into account both the effects on speech quality and on noise level (Marzinzik 2000). The listening experiments described in this thesis provide useful data for the development and evaluation of objective models.

Development of methods to objectify noise-reduction effects

The focus of hearing aid noise-reduction evaluation shifted during the last years towards cognitive measures, especially to listening effort. The underlying theory is that, although noise reduction does not increase intelligibility scores, it might be that the cognitive load required to obtain the same intelligibility scores is decreased due to the reduction of background noise. Several attempts were made to objectively quantify the cognitive load. Examples of these attempts are the use of dual tasks, where the performance of a second task is assumed to reveal how much cognitive load is used by the first (listening) task (Sarampalis et al. 2009), and the use of reaction times, where the time in between the presentation of the speech stimuli and the listener's response to a task (simply repeating the stimulus or, more difficult, performing an arithmetic task for digit stimuli) is assumed to be related to the cognitive load (Houben e.a. under revision). Objective measures for listening effort may also be useful for determination of the optimal trade-off between noise reduction and speech distortion for an individual listener: whereas decreasing noise level decreases the cognitive load, increasing speech distortions are expected to increase cognitive load. However, until objective cognitive measures can be used for these purposes there is still a long way to go in optimization, verification, and validation of the measures.

The results in this thesis imply that noise reduction can indeed positively contribute to the compensation of the number-one complaint of reduced speech perception in noise. However, an important requirement is that the hearing aids are chosen and fitted well. Unfortunately, the latter is not guaranteed in current clinical practice due to lack of knowledge in this field. There is room for improvement if we increase our knowledge along the lines initiated in this thesis. This knowledge can be translated into clinical procedures for selection and fine tuning of noise reduction. Given the large-scale application of noise reduction in the vast majority of modern hearing aids and the relevance of compensating the complaints about noisy situations, it is worthwhile to continue research on the perceptual effects of noise reduction.

Bibliography

Alcàntara, J., Moore, B., Kühnel, V., & Launer, S. (2003). Evaluation of the noise reduction system in a commercial digital hearing aid. *International Journal of Audiology*, *42*(1), 34–42.

Anderson, M. C., Arehart, K. H., & Kates, J. M. (2009). The acoustic and peceptual effects of series and parallel processing. *EURASIP Journal on Advances in Signal Processing*, *2009*, 1–20.

ANSI (2004). *Specification for audiometers*. American National Standards Institute, S3.6-2004.

Anzalone, M. C., Calandruccio, L., Doherty, K. A., & Carney, L. H. (2006). Determination of the potential benefit of time-frequency gain manipulation. *Ear and hearing*, *27*(5), 480–492.

Arehart, K. H., Kates, J. M., & Anderson, M. C. (2010). Effects of noise, nonlinear processing, and linear filtering on perceived speech quality. *Ear and hearing*, *31*(3), 420–436.

Bendat, J. S., & Piersol, A. G. (2011). *Random Data: Analysis and Measurement Procedures*. John Wiley & Sons.

Bentler, R., & Chiou, L.-K. (2006). Digital noise reduction: An overview. *Trends in Amplification*, *10*(2), 67–82.

Bentler, R., Wu, Y.-H., Kettel, J., & Hurtig, R. (2008). Digital noise reduction: Outcomes from laboratory and field studies. *International Journal of Audiology*, *47*(8), 447–460.

Bentler, R. A. (2005). Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence. *Journal of the American Academy of Audiology*, *16*(7), 473–484.

Berouti, M., Schwartz, R., & Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79.*, vol. 4, (pp. 208–211).

Bisgaard, N., Vlaming, M., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification*, *14*(2), 113–120.

Böckenholt, U. (2001). Hierarchical modeling of paired comparison data. *Psychological Methods*, *6*(1), 49–66.

Bosman, A. (1989). *Speech perception by the hearing impaired*. Ph.D. thesis, Rijksuniversiteit Utrecht, Utrecht.

Boymans, M., & Dreschler, W. A. (2000). Field trials using a digital hearing aid with active noise reduction and dual-microphone directionality. *Audiology*, *39*(5), 260–268.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, *39*(3/4), 324–345.

Brookes, M., & Huckvale, M. (2011). Can binary masks improve intelligibility? In *3th International Workshop on Speech in Noise: Intelligibility and Quality, Lyon, F*.

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, *120*(6), 4007–4018.

Byrne, D., Parkinson, A., & Newall, P. (1991). Modified hearing aid selection procedures for severe/profound hearing losses. *The Vanderbilt hearing aid report II*, (pp. 295–300).

Cao, S., Li, L., & Wu, X. (2011). Improvement of intelligibility of ideal binary-masked noisy speech by adding background noise. *The Journal of the Acoustical Society of America*, *129*(4), 2227–2236.

Chen, F., & Loizou, P. (2012). Impact of SNR and gain-function over-and underestimation on speech intelligibility. *Speech Communication*, *54*(2), 272–281.

Chung, K. (2004). Challenges and recent developments in hearing aids: Part i. speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends in Amplification*, *8*(3), 83–124.

Chung, K. (2007). Effective compression and noise reduction configurations for hearing protectors. *The Journal of the Acoustical Society of America*, *121*(2), 1090–1101.

Davis, L. A., & Davidson, S. A. (1996). Preference for and performance with damped and undamped hearing aids by listeners with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *39*(3), 483–493.

Dillon, H. (2001). *Hearing Aids*. Thieme.

Dittrich, R., Hatzinger, R., & Katzenbeisser, W. (2004). A log-linear approach for modelling ordinal paired comparison data on motives to start a PhD programme. *Statistical Modelling*, *4*(3), 1–13.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., & Westermann, S. (2001). ICRA noises: artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. international collegium for rehabilitative audiology. *Audiology: official organ of the International Society of Audiology*, *40*(3), 148–157.

Edwards, B. (2007). The future of hearing aid technology. *Trends in Amplification*, *11*(1), 31–46.

Elberling, C. (2002). About the VoiceFinder. Tech. rep., Oticon.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *32*(6), 1109–1121.

Gabrielsson, A., Schenkman, B. N., & Hagerman, B. (1988). The effects of different frequency responses on sound quality judgments and speech intelligibility. *Journal of Speech and Hearing Research*, *31*(2), 166–177.

Gabrielsson, A., & Sjögren, H. (1979). Perceived sound quality of sound-reproducing systems. *The Journal of the Acoustical Society of America*, *65*(4), 1019–1033.

Glasberg, B. R., & Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society*, *50*(5), 331–342.

Grant, K. W., Summers, V., & Leek, M. R. (1998). Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *104*(2), 1051–1060.

Grimm, G., Herzke, T., Berg, D., & Hohmann, V. (2006). The master hearing aid: A PC-Based platform for algorithm development and evaluation. *Acta Acustica united with Acustica*, *92*(4), 618–628.

Hagerman, B., & Olofsson, A. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acustica united with Acustica*, *90*(2), 356–361.

Hilkhuysen, G., Gaubitch, N., Brookes, M., & Huckvale, M. (2012). Effects of noise suppression on intelligibility: Dependency on signal-to-noise ratios. *The Journal of the Acoustical Society of America*, *131*(1), 531–539.

Hoetink, A. E., Körössy, L., & Dreschler, W. A. (2009). Classification of steady state gain reduction produced by amplitude modulation based noise reduction in digital hearing aids. *International Journal of Audiology*, *48*(7), 444–455.

Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an international speech test signal (ISTS). *International Journal of Audiology*, *49*(12), 891–903.

Houben, R., Dijkstra, T., & Dreschler, W. A. (2011). The influence of noise type on the preferred setting of a noise reduction algorithm. In *Speech Perception and Auditory Disorders*, (pp. 465–472). Nyborg, DK: The Danavox Jubilee Foundation 2012.

Houben, R., Dijkstra, T. M. H., & Dreschler, W. A. (2012). Analysis of individual preferences for tuning of noise-reduction algorithms. *Journal of the Audio Engineering Society*, *60*(12), 1024–1037.

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology*, *52*(11), 753–761.

Hu, Y., & Loizou, P. C. (2007a). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America*, *122*(3), 1777–1786.

Hu, Y., & Loizou, P. C. (2007b). Subjective comparison and evaluation of speech enhancement algorithms. *Speech communication*, *49*(7), 588–601.

Hu, Y., & Loizou, P. C. (2008). Techniques for estimating the ideal binary mask. In *Proc. 11th Int. Workshop Acoust. Echo Noise Control*.

Huber, R., & Kollmeier, B. (2006). PEMO-Q - a new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(6), 1902–1911.

ITU-T (1996). *Methods for subjective determination of transmission quality*. International Telecommunication Union, Recommendation P.800.

ITU-T (2003). *Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm*. International Telecommunication Union, Recommendation P.835.

Kates, J., & Arehart, K. (2010). The hearing-aid speech quality index (HASQI). *J Audio Eng Soc*, *58*(5), 363–381.

Kates, J. M., & Arehart, K. H. (2005). Coherence and the speech intelligibility index. *The Journal of the Acoustical Society of America*, *117*(4), 2224–2237.

Kim, G., Lu, Y., Hu, Y., & Loizou, P. C. (2009). An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *126*(3), 1486–1494.

Kjems, U., Boldt, J. B., Pedersen, M. S., Lunner, T., & Wang, D. (2009). Role of mask pattern in intelligibility of ideal binary-masked noisy speech. *The Journal of the Acoustical Society of America*, *126*(3), 1415–1426.

Kochkin, S. (2002). Consumers rate improvements sought in hearing instruments. *Hearing Review*, *9*(11), 18–22.

Kryter, K. D. (1962). Methods for the calculation and use of the articulation index. *The Journal of the Acoustical Society of America*, *34*(11), 1689–1697.

Kuk, F. K., & Paludan-Müller, C. (2006). Noise-management algorithm may improve speech intelligibility in noise. *The Hearing Journal*, *59*(4), 62–71.

Lawson, G. D., & Chial, M. R. (1982). Magnitude estimation of degraded speech quality by normal- and impaired-hearing listeners. *Journal of the Acoustical Society of America*, *72*(6), 1781–1787.

Legarth, S., Simonsen, C., Bramsløw, L., Le Ray, G., & Zacharov, N. (2010). Evaluation of hearing aid performance based on normal-hearing listeners. *Proceedings of the 3rd International Workshop on Perceptual Quality of Systems, Dresden, Germany*, (pp. 149–154).

Levitt, H. (1992). Adaptive procedures for hearing aid prescription and other audiologic applications. *Journal of the American Academy of Audiology*, *3*(2), 119–131.

Levitt, H. (2001). Noise reduction in hearing aids: An overview. *Journal of rehabilitation research and development*, *38*(1), 111–121.

Levitt, H., Neuman, A., Mills, R., & Schwander, T. (1986). A digital master hearing aid. *Journal of rehabilitation research and development*, *23*(1), 79–87.

Li, N., & Loizou, P. C. (2008a). Effect of spectral resolution on the intelligibility of ideal binary masked speech. *JASA Express Letters*, *123*(4), EL59–EL64.

Li, N., & Loizou, P. C. (2008b). Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. *The Journal of the Acoustical Society of America*, *123*(3), 1673–1682.

Lim, J., & Oppenheim, A. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, *67*(12), 1586–1604.

Loizou, P. C. (2007). *Speech enhancement: theory and practice*. Boca Raton, Florida: CRC Press.

Loizou, P. C., & Kim, G. (2011). Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(1), 47–56.

Lunner, T., Rudner, M., & Rönnberg, J. (2009). Cognition and hearing aids. *Scandinavian journal of psychology*, *50*(5), 395–403.

Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., & Spriet, A. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *The Journal of the Acoustical Society of America*, *127*(3), 1491–1505.

Martin, R. (2001). Noise power spectral density estimation based on optimal smoothing and minimum statistics. *Speech and Audio Processing, IEEE Transactions on*, *9*(5), 504–512.

Marzinzik, M. (2000). *Noise Reduction Schemes for Digital Hearing Aids and their Use for the Hearing Impaired*. Ph.D. thesis, Universität Oldenburg, Oldenburg.

McCreery, R. W., Gustafson, S., & Stelmachowicz, P. G. (2010). Should digital noise reduction be activated in pediatric hearing aid fittings? *A Sound Foundation Through Early Amplification*.

Moore, B. C. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear and Hearing*, *17*(2), 133–161.

Mueller, H. G., Bentler, R. A., & Wu, Y.-H. (2008). Prescribing maximum hearing aid output: Differences among manufacturers found. *The Hearing Journal*, *61*(3), 30–36.

Mueller, H. G., Weber, J., & Hornsby, B. W. (2006). The effects of digital noise reduction on the acceptance of background noise. *Trends in Amplification*, *10*(2), 83–93.

Nabelek, A. K., Freyaldenhoven, M. C., Tampas, J. W., Burchfield, S. B., & Muenchen, R. A. (2006). Acceptable noise level as a predictor of hearing aid use. *Journal of the American Academy of Audiology*, *17*(9), 626–639.

Nordrum, S., Erler, S., Garstecki, D., & Dhar, S. (2006). Comparison of performance on the hearing in noise test using directional microphones and digital noise reduction algorithms. *American Journal of Audiology*, *15*(1), 81–91.

Palmer, C. V., Bentler, R., & Mueller, H. G. (2006). Amplification with digital noise reduction and the perception of annoying and aversive sounds. *Trends in Amplification*, *10*(2), 95–104.

Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research*, *29*(2), 146–154.

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology*, *18*(1), 43–52.

Rangachari, S., & Loizou, P. C. (2006). A noise-estimation algorithm for highly non-stationary environments. *Speech Communication*, *48*(2), 220–231.

Rhebergen, K. S., & Versfeld, N. J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2181–2192.

Rhebergen, K. S., Versfeld, N. J., & Dreschler, W. A. (2009). The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise. *The Journal of the Acoustical Society of America*, *126*(6), 3236–3245.

Ricketts, T. A., & Hornsby, B. W. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing" digital noise reduction". *Journal of the American Academy of Audiology*, *16*(5), 270–277.

Rix, A., Beerends, J., Hollier, M., & Hekstra, A. (2001). Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*, vol. 2, (pp. 749–752).

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *Journal of Speech, Language, and Hearing Research*, *52*(5), 1230–1240.

Schaub, A. (2008). *Digital Hearing Aids*. Thieme.

Smith, S. W. (1997). *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing.

Smoorenburg, G. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *The Journal of the Acoustical Society of America*, *91*(1), 421–437.

Stelmachowicz, P., Lewis, D., Hoover, B., & Keefe, D. (1999). Subjective effects of peak clipping and compression limiting in normal and hearing-impaired children and adults. *The Journal of the Acoustical Society of America*, *105*(1), 412–422.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*(3), 455–462.

Taal, C. H., Hendriks, R. C., Heusdens, R., Jensen, J., & Kjems, U. (2009). An evaluation of objective quality measures for speech intelligibility prediction. In *Tenth Annual Conference of the International Speech Communication Association*, (pp. 1947–1950).

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, *107*(3), 1671–1684.

Versfeld, N. J., Festen, J. M., & Houtgast, T. (1999). Preference judgments of artificial processed and hearing-aid transduced speech. *The Journal of the Acoustical Society of America*, *106*(3), 1566–1578.

Wang, D. (2005). On ideal binary mask as the computational goal of auditory scene analysis. In P. Divenyi (Ed.) *Speech Separation by Humans and Machines*, (pp. 181–197). Boston: Kluwer Academic Publishers.

Wang, D. (2008). Time-frequency masking for speech separation and its potential for hearing aid design. *Trends in Amplification*, *12*(4), 332–353.

Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., & Lunner, T. (2008). Speech perception of noise with binary gains. *The Journal of the Acoustical Society of America*, *124*(4), 2303–2307.

Wang, D., Kjems, U., Pedersen, M. S., Boldt, J. B., & Lunner, T. (2009). Speech intelligibility in background noise with ideal binary time-frequency masking. *The Journal of the Acoustical Society of America*, *125*(4), 2336–2347.

Warner, R. L., & Bentler, R. A. (2002). Thresholds of discomfort for complex stimuli: Acoustic and sound-quality predictors. *Journal of Speech, Language, and Hearing Research*, *45*(5), 1016–1026.

Wu, Y.-H., & Stangl, E. (2013). The effect of hearing aid signal-processing schemes on acceptable noise levels: Perception and prediction. *Ear and hearing*, *34*(3), 333–341.

Zakis, J. A., Hau, J., & Blamey, P. J. (2009). Environmental noise reduction configuration: Effects on preferences, satisfaction, and speech understanding. *International Journal of Audiology*, *48*(12), 853–867.

Samenvatting

De meest gehoorde klacht van hoortoestelgebruikers is dat het erg moeilijk is om spraak te verstaan in rumoerige omgevingen. Hoortoestelfabrikanten proberen de moeilijkheden met spraak in rumoer ("ruis") te verminderen door hoortoestellen te voorzien van signaalbewerkingsalgoritmes die de achtergrondruis moeten verminderen. De meest voorkomende variant is een ruisonderdrukkingsalgoritme, dat op basis van één ingangssignaal inschat of er ruis aanwezig is of niet. Als het algoritme de aanwezigheid van ruis detecteert, wordt de hoortoestelversterking aangepast om de ruis te verminderen. Daarbij moet eventueel aanwezige spraak onaangetast blijven.

In Nederland zijn er ongeveer 650 000 hoortoestellen in gebruik, waarvan de overgrote meerderheid ruisonderdrukking bevat. Het is daarom opmerkelijk dat er slechts weinig bekend is over de implementatie en perceptieve effecten van ruisonderdrukking. Voor audiologen en audiciens die hoortoestellen voorschrijven en instellen is niet bekend wat er precies verandert aan het signaal wanneer ze de ruisonderdrukking activeren in een specifiek toestel, en of dit verschilt tussen toestellen. Dat betekent dat het voor tienduizenden hoortoestelgebruikers niet zeker is of hun hoortoestel optimaal is gekozen en ingesteld om te compenseren voor het probleem van verminderd spraakverstaan in ruis.

Dit proefschrift beschrijft verschillende onderzoeken naar de effecten van ruisonderdrukking op perceptieve uitkomstmaten, zoals spraakverstaan, luisterinspanning, en persoonlijke voorkeur. Het grootste deel van het proefschrift (Hoofdstukken 2 tot en met 5) beschrijft onderzoeken naar ruisonderdrukking in commerciële hoortoestellen. Hoofdstukken 6 en 7 beschrijven onderzoeken waarin gebruik wordt gemaakt van state-of-the-art ruisonderdrukkingsalgoritmes uit de literatuur om de kennis over enkele aspecten van ruisonderdrukking te verdiepen.

Hoofdstuk 1: Introductie

In Hoofdstuk 1 worden de basisprincipes van ruisonderdrukking uitgelegd. Over het algemeen bestaat ruisonderdrukking uit twee delen. Ten eerste moet worden geclassificeerd of de omgeving bestaat uit ruis, spraak, of spraak in ruis. Deze stap resulteert meestal in een schatting van de momentane signaal-ruisverhouding per frequentiekanaal. De tweede stap is het aanpassen (of gelijk houden) van de hoortoestelversterking op basis van de geschatte signaal-ruisverhouding. Daarbij is de belangrijkste uitdaging het vinden van een optimum in de trade-off tussen het verminderen van de ruis en het behouden van de spraakkwaliteit. Omdat voor beide stappen veel verschillende benaderingen mogelijk zijn, is het niet zo dat één algoritme als representatief kan worden gezien voor alle ruisonderdrukkers. Conclusies van onderzoeken met het ene algoritme kunnen daarom niet zo maar worden gegeneraliseerd naar andere algoritmes. Om die reden zijn in dit proefschrift verschillende ruisonderdrukkingsalgoritmes gebruikt, zodat we ook tussen algoritmes konden vergelijken in plaats van alleen het effect te meten van het aan- of uitzetten van één algoritme.

Hoofdstuk 2: Methode om ruisonderdrukking in hoortoestellen te vergelijken

In Hoofdstuk 2 wordt beschreven hoe we een methode hebben ontwikkeld en gevalideerd om ruisonderdrukkers uit verschillende hoortoestellen met elkaar te kunnen vergelijken, zonder dat andere verschillen tussen hoortoestellen deze vergelijking beïnvloeden. Voor ieder hoortoestel werd een inverse filter gemaakt dat corrigeerde voor verschillen tussen toestellen in de frequentierespons die er ondanks zorgvuldige aanpassing van de hoortoestelversterking nog waren. Nadat het filter op alle opnames van een hoortoestel was toegepast, konden normaalhorende proefpersonen geen verschillen meer horen tussen vijf verschillende toestellen met lineaire versterking zonder verdere signaalbewerkingen. Wanneer eenmaal een filter gemaakt is voor een specifiek toestel, kan dat filter ook worden toegepast op opnames van dat toestel waarbij de ruisonderdrukking is ingeschakeld. Hierdoor wordt het mogelijk om alleen naar het effect van ruisonderdrukking te luisteren, zonder de invloed van andere hoortoesteleigenschappen. Deze methode voor het maken en filteren van hoortoestelopnames vormt de basis van de volgende drie hoofdstukken, waar implementaties van ruisonderdrukking in verschillende hoortoestellen onderling worden vergeleken.

Hoofdstuk 3: Ruisonderdrukking in lineaire hoortoestellen (normaalhorenden)

In Hoofdstuk 3 worden ruisonderdrukkers uit vier verschillende hoortoestellen perceptief met elkaar vergeleken. De ruisonderdrukkers bleken onderling te verschillen in de mate waarin ze de ruis verminderden volgens normaalhorende proefpersonen. Bij lage signaal-ruisverhouding (-4 dB), waar het moeilijker is voor de ruisonderdrukking om spraak en ruis te onderscheiden, werd de spraak minder natuurlijk door ruisonderdrukking. Scores voor spraakverstaanbaarheid en luisterinspanning verschilden niet tussen ruisonderdrukking *aan* en *uit*, en slechts weinig tussen verschillende ruisonderdrukkers. Wel waren er verschillen in voorkeur voor de verschillende ruisonderdrukkers. De voorkeursresultaten bevestigen dat resultaten van één ruisonderdrukker niet kunnen worden gegeneraliseerd naar andere ruisonderdrukkers. De resultaten lieten verder zien dat niet alleen ruisonderdrukkers van elkaar verschillen, maar dat ook luisteraars van elkaar verschillen in de mate waarin hinderlijkheid van ruis en natuurlijkheid van spraak meewegen in de persoonlijke voorkeur.

Hoofdstuk 4: Ruisonderdrukking in lineaire hoortoestellen (slechthorenden)

In Hoofdstuk 4 worden dezelfde perceptieve metingen beschreven als in Hoofdstuk 3, maar nu met slechthorende proefpersonen en met drie in plaats van vier hoortoestellen. De resultaten van slechthorende proefpersonen kwamen grotendeels overeen met die van normaalhorende proefpersonen in Hoofdstuk 3. Er was nu echter een kleine achteruitgang in spraakverstaanbaarheid door de ruisonderdrukker die de meest positieve scores behaalde op hinderlijkheid van de ruis, natuurlijkheid van de spraak en persoonlijke voorkeur. Dit ondersteunt de algemene aanname dat ruisonderdrukking gepaard gaat met een trade-off tussen spraakverstaanbaarheid en luistercomfort. Omdat bij de gebruikte signaal-ruisverhouding (+4 dB) de spraakkwaliteit niet werd aangetast, kon nu geen analyse worden gedaan van de individuele voorkeur voor de weging tussen hinderlijkheid van ruis en natuurlijkheid van spraak zoals in Hoofdstuk 3.

Hoofdstuk 5: Ruisonderdrukking in compressie-hoortoestellen

In Hoofdstuk 5 worden metingen beschreven met dezelfde drie hoortoestellen als in Hoofdstuk 4, maar nu werd ruisonderdrukking gecombineerd met compressie, wat in de praktijk vaker voorkomt in hoortoestellen dan lineaire versterking. Uit akoestische analyses bleek dat compressie voor onze ingangssignalen met een signaal-ruisverhouding van +4 dB een negatief effect had op de signaal-ruisverhouding. De ruisonderdrukking corrigeerde voor dit negatieve effect. Uit perceptieve metingen met slechthorende proefpersonen bleek dat de combinatie van ruisonderdrukking en compressie geen invloed had op de spraakverstaanbaarheid. Ook kreeg de combinatie van ruisonderdrukking en compressie in geen van de hoortoestellen significant hogere voorkeur dan lineaire versterking zonder ruisonderdrukking. Dit in tegenstelling tot ruisonderdrukking in lineaire toestellen (zie Hoofdstuk 4, de voorkeur voor ruisonderdrukking met lineaire versterking was significant hoger dan voor lineaire versterking zonder ruisonderdrukking). Hieruit blijkt dat het van belang is om bij het ontwerpen en testen van ruisonderdrukking voor toepassing in hoortoestellen rekening te houden met compressie.

Hoofdstuk 6: Ideale en niet-ideale ruisonderdrukking

In Hoofdstuk 6 worden perceptieve metingen beschreven met vier verschillende variaties op een ruisonderdrukkingsalgoritme. Twee condities hadden een "ideale" ruisschatter, wat betekent dat ze de spraak en ruis van het ingangssignaal afzonderlijk ontvingen en daarmee de beschikking hadden over de werkelijke signaal-ruisverhouding. Eén van deze condities verwijderde per tijdseenheid het signaal uit alle frequentiebanden waarin ruis overheerste ("ideal binary mask"), terwijl bij de andere conditie de verzwakking beperkt was tot een maximum van 10 dB ("tempered mask"), wat meer overeenkomt met de hoeveelheid verzwakking die bij ruisonderdrukking in hoortoestellen wordt toegepast. De resultaten van deze "ideale" condities lieten een trade-off zien tussen kwaliteit en spraakverstaanbaarheid, die gerelateerd was aan de hoeveelheid verzwakking: oneindige verzwakking gaf de sterkste verbetering in spraakverstaanbaarheid, maar verminderde de geluidskwaliteit door de grote en snelle veranderingen in versterking. De andere twee condities bestonden uit realistische ruisschatters, gecombineerd met de verzwakking volgens de "tempered mask". Doordat deze condities fouten maakten in de schatting van spraak en ruis, verbeterden ze de spraakverstaanbaarheid niet. Wel werd de ruis minder hinderlijk, waardoor de proefpersonen voorkeur hadden voor ruisonderdrukking boven geen ruisonderdrukking. Het lijkt daarom zinvoller om in de ontwikkeling van ruisonderdrukking in te zetten op het verbeteren van de subjectieve voordelen van ruisonderdrukking (waarbij in de gaten moet worden gehouden dat de spraakverstaanbaarheid niet achteruit gaat) dan te proberen de spraakverstaanbaarheid te verbeteren ten koste van het subjectieve voordeel.

Hoofdstuk 7: Detectiedrempels voor vervorming door ruisonderdrukking

In Hoofdstuk 7 wordt een studie beschreven waarin we de trade-off tussen vermindering van ruis en vervorming van het signaal verder bestudeerden. Deze beide effecten werden kunstmatig van elkaar gescheiden, zodat we de detectiedrempel voor vervorming konden meten zonder dat veranderingen in ruisniveau daar invloed op hadden. De detectiedrempel was hoger voor slechthorenden dan voor normaalhorenden. Dit suggereert dat ruisonderdrukking voor slechthorenden sterker kan worden ingesteld dan voor normaalhorenden zonder de subjectieve geluidskwaliteit te verminderen. Echter, ten opzichte van hun detectiedrempel hadden slechthorenden voorkeur voor een minder sterke ruisonderdrukking dan normaalhorenden. Dit suggereert dat wanneer er eenmaal hoorbare vervorming aanwezig is, slechthorenden minder vervorming tolereren dan normaalhorenden. Het eindresultaat was dat normaalhorenden en slechthorenden gemiddeld gezien ongeveer eenzelfde sterkte prefereerden, maar dat er grote individuele verschillen waren binnen beide groepen.

Hoofdstuk 8: Samenvatting en algemene discussie

In Hoofdstuk 8 worden de voorgaande hoofdstukken samengevat en wordt besproken wat de resultaten betekenen voor de ontwikkeling en toepassing van ruisonderdrukking en voor verder onderzoek naar de effecten van ruisonderdrukking. Omdat de effecten van ruisonderdrukking bleken te verschillen tussen hoortoestellen is het wenselijk dat hoortoestelfabrikanten meer informatie geven over de implementatie van de belangrijkste parameters van de ruisonderdrukker. Daarnaast zouden er binnen een hoortoestel meer mogelijkheden moeten zijn om ruisonderdrukking aan te passen aan de individuele voorkeur van een gebruiker. Verder is het van belang dat de gebruiker realistische verwachtingen heeft van de mogelijke voordelen en beperkingen van ruisonderdrukking. Op basis van de resultaten uit Hoofdstukken 3 tot en met 5 kunnen geen conclusies worden getrokken over welk hoortoestel het beste is, maar de resultaten laten wel zien dat er meer aandacht zou moeten zijn voor het aanpassen van ruisonderdrukking aan de wensen van de individuele luisteraar. Zolang de informatie over de implementatie van ruisonderdrukking beperkt is, zal met technische metingen moeten worden uitgezocht hoe de ruisonderdrukker reageert in een specifieke omgeving, en met voorkeursmetingen moeten worden bepaald welke vorm van ruisonderdrukking het beste past bij een individuele hoortoestelgebruiker. Verder onderzoek zal nodig zijn om te karakteriseren welke verschillende implementaties van ruisonderdrukking er zijn en hoe die gekoppeld kunnen worden aan de verschillende wensen van de individuele hoortoesteldrager. De trade-off tussen vermindering van ruis en vervorming van spraak lijkt hiervoor een goed uitgangspunt te zijn. Een ander belangrijk aandachtspunt voor vervolgonderzoek is het ontwikkelen, evalueren en toepassen van objectieve maten om het mogelijke effect van ruisonderdrukking op luisterinspanning te meten.

De resultaten van dit proefschrift laten zien dat ruisonderdrukking geen directe verbetering biedt voor het spraakverstaan in lawaai, maar desalniettemin positief kan bijdragen aan het luistercomfort. Er is echter nog veel ruimte voor verbetering in het selecteren en instellen van de ruisonderdrukking voor individuele hoortoestelgebruikers. Gezien de grootschalige toepassing van ruisonderdrukking in hoortoestellen is het de moeite waard om door te gaan met onderzoek om de kennis omtrent de perceptieve effecten van ruisonderdrukking te vergroten.

Dankwoord

Hoewel de kaft van dit boekje slechts één naam vermeldt, had de inhoud ervan er nooit kunnen zijn als er niet nog veel meer mensen bij betrokken waren geweest. Daarom maak ik graag van deze meest gelezen pagina's gebruik om nog vele andere mensen te noemen die ik erg dankbaar ben voor het feit dat het boekje er nu toch is.

De titelpagina's van de afzonderlijke hoofdstukken geven al een realistischer beeld dan de kaft, doordat ze ook de namen Wouter A. Dreschler en Rolph Houben vermelden. Wouter, aan jouw aanstekelijke enthousiasme over het noise-reduction project heeft dit boekje zijn bestaansrecht te danken. Bedankt voor dat enthousiasme, voor je vertrouwen in de komst van het welbekende 'kaftje eromheen', en voor alle begeleiding die nodig was om daadwerkelijk van enthousiasme tot kaftje te komen. Ik ben er trots op het laatste deel van je mooie eindejaars-serie dinsdagse promoties in te mogen vullen. Rolph, als ik het heb over bestaansrecht van dit boekje, dan kan jouw naam zeker niet achterblijven. Jij wist je rol als dagelijks begeleider behoorlijk letterlijk in te vullen door altijd bereikbaar te zijn voor vragen of overleg. Ik heb veel geleerd door jouw onderzoeks-ervaring, al zie je dat misschien niet terug aan dit hoofdstuk zonder onderzoeksvraag en t-test.

De overige leden van de promotiecommissie, prof.dr. D. Başkent, prof.dr. W.J. Fokkens, prof.dr.ir. C.A. Grimbergen, dr. N.J. Versfeld en prof.dr.ir. B. de Vries wil ik bedanken voor het lezen en beoordelen van de eerste versie van dit boekje en voor hun constructieve opmerkingen daarover.

De inhoud van dit boekje is grotendeels te danken aan de medewerking van proefpersonen. Ik heb altijd erg veel bewondering gehad voor het geduld waarmee mensen de soms best moeilijke of saaie metingen uitvoerden, en ben hen daar erg dankbaar voor.

Dat ik de afgelopen jaren met veel plezier gewerkt heb kwam niet alleen door leuk onderzoek en goede begeleiding, maar ook doordat ik te midden van leuke collega's mocht werken. Femke, jij hebt het het langst met mij op de kamer weten vol te houden. Bedankt voor die tijd en voor je blijvende belangstelling en betrokkenheid vanaf de andere kant van de gang. Heel fijn dat je je verjaardag dit jaar als paranimf wilt vieren! Thijs, Jelmer en Marjolijn, bedankt voor de leuke tijd die we hadden als (buur-)roomies. Waren er in die tijd nog flinke potten koffie nodig waren om iedereen in 'het lab' van

belangrijke onderzoeks-input te voorzien, nu is het juist de thee die hier nog voor wat aanloop zorgt. Maj, Marya, en Leonie, dankzij jullie bleef het hier ook gezellig zonder directe roomies. Dat geldt ook voor de andere buren rondom het bootje: Monique Boymans, Koen en László. Fijn dat ik altijd ergens binnen kon lopen voor allerlei soorten praktische hulp of meedenk-hulp. Naast het binnen lopen was ook het buiten lopen altijd erg fijn. Alle lunchwandelaars bedankt voor de gezelligheid, de goede gesprekken, en natuurlijk het ontwijken van de gebieden met een hoog gehalte aan bepaalde beestjes.

De donderdagochtend-meetings met het researchteam heb ik altijd erg leuk en waardevol gevonden. Ondanks, of juist ook dankzij, het feit dat we ons allemaal met uiteenlopende projecten bezig houden heb ik veel kunnen leren van de overleggen, de praatjes, de vragen, en de feedback. Hiske en Patrick, jullie klinische blik hielp daarbij altijd om verder te kijken dan de ivoren toren van het onderzoek. Maaike, Maaike en Maj, bedankt voor de leuke samenwerking in jullie onderzoeks-tijd. Monique Leensen, pas-gepromoveerd voorbeeld, en Thamar en Bastiaan, bijna-gepromoveerde 'lotgenoten', bedankt voor alle praktische adviezen en tips in de laatste fase van het promotietraject.

Ook alle overige AC-medewerkers wil ik bedanken voor hun behulpzaamheid bij het leren audiometreren en bij allerlei praktische vragen, en voor de gezellige gesprekjes onderweg naar het AMC, tijdens de lunchwandeling, of in de wandelgangen.

Tenslotte wil ik alle mensen bedanken die maken dat ook het leven buiten het AMC erg goed is. Alle vrienden, kringgenoten en gemeenteleden die er voor hebben gezorgd dat ik me de afgelopen jaren thuis ben gaan voelen in Hilversum, en alle waardevolle contacten rondom Hilversum (rondom betekent hier ongeveer van Enschede tot Haarlem en van Groningen tot Den Haag): goede vrienden, lieve zussen, en betrokken familieleden. Grote zus, Jorien, fijn dat je paranimf wilt zijn. Papa en mama, de fysieke afstand is behoorlijk toegenomen in de loop van mijn promotietraject (jullie vallen niet meer binnen 'rondom Hilversum'), maar jullie betrokkenheid is er bepaald niet minder op geworden. Toda raba!

Curicullum Vitae

19 October 1985	Born in Vlissingen, the Netherlands	
1997-2003	**Secondary school** Greijdanus College, Zwolle	
2003-2006	**Bachelor Technical Medicine** University of Twente, Enschede	
	BSc Thesis:	Clinical Neurophysiology *Medisch Spectrum Twente, Enschede*
2006-2009	**Master Technical Medicine** University of Twente, Enschede	
	Internships:	Anaesthesiology *Canisius Wilhelmina Ziekenhuis, Nijmegen*
		Respiratory Physiology *Medisch Spectrum Twente, Enschede*
		Clinical Neurophysiology *Medisch Spectrum Twente, Enschede*
		Audiology *UMC St Radboud, Nijmegen*
	MScThesis:	Clinical Neurophysiology *University Medical Center Groningen, Groningen*
2009-2013	**PhD student** Clinical and Experimental Audiology Academic Medical Center, Amsterdam	

PhD Portfolio

Summary of PhD training and teaching activities

Name PhD student:	I. Brons
PhD Period:	2009-2013
Name PhD supervisor:	Prof.dr.ir. W.A.Dreschler

1. PhD training	Year	Workload (ECTS)
General courses		
- BROK (Basiscursus Regelgeving & Organisatie voor Klinisch onderzoekers)	2011	0.9
- Scientific writing in English for publication	2011	1.5
- Oral presentation in English	2011	0.8
- Practical biostatistics	2011	1.1
Specific courses		
- Training for clinical audiometry (internal)	2011	3.0
- Digitale signaalbewerking in hoortoestellen (NAN)	2012	0.5
- Akoestiek voor audiologie	2012	1.0
Seminars, workshops and master classes		
- Weekly department seminars	2009-2013	5.0
- Meetings Werkgroep Auditief Systeem (WAS)	2009-2013	1.0
- Meetings Nederlandse Vereniging voor Audiologie (NVA)	2009-2013	1.0

	Year	Workload (ECTS)
Presentations		
- Perceptual effects of noise reduction *Poster on ARCHES meeting*	2009	0.5
- Perceptual effects of noise reduction *Poster on Workshop on Speech in Noise*	2010	0.5
- Single-channel noise reduction in hearing aids - Recordings for perceptual evaluation *Presentation on Nederlandse Vereniging voor Audiologie*	2010	0.5
- Ruisonderdrukking in hoortoestellen - Zijn er verschillen tussen algoritmes en hoe luisteren wij daarnaar? *Presentation on StAr Seminar*	2010	0.5
- Can we compare the sound quality of noise reduction between commercial hearing aids? A method to level the ground between devices *Presentation on Workshop on Speech in Noise*	2011	0.5
- Perceptual effects of noise reduction in hearing aids *Presentation on WAS-dag*	2011	0.5
- Perceptual comparison of noise reduction in hearing aids *Poster on ISAAR*	2011	0.5
- Perceptual effects of noise reduction in hearing aids *Presentation on ARCHES meeting*	2011	0.5
- Perceptual effects of noise reduction by time-frequency masking of noisy speech *Poster on Workshop on Speech in Noise*	2012	0.5
- Ruisonderdrukking in hoortoestellen *Presentation on KNO wetenschapsdag*	2012	0.5
- Acoustical and perceptual comparison of noise reduction in hearing aids *Poster on Workshop on Speech in Noise*	2013	0.5
- Acoustical and perceptual evaluation of noise reduction in hearing adis *Poster on ISAAR*	2013	0.5

	Year	Workload (ECTS)
(Inter)national conferences		
- ARCHES meeting *November 2009, Nottingham, UK*	2009	0.5
- Workshop on Speech in Noise *January 2010, Amsterdam, the Netherlands*	2010	0.5
- IEEE Benelux Signal Processing Symposium *April 2010, Delft, the Netherlands*	2010	0.2
- Workshop on Speech in Noise *January 2011, Lyon, France*	2011	0.5
- International Symposium on Auditory and Audiological Research (ISAAR) *August 2011, Nyborg, Denmark*	2011	0.8
- ARCHES meeting *November 2011, Leuven, Belgium*	2011	0.5
- Workshop on Speech in Noise *January 2012, Cardiff, UK*	2012	0.5
- Workshop on Speech in Noise *January 2013, Vitoria, Spain*	2013	0.5
- International Symposium on Auditory and Audiological Research (ISAAR) *August 2013, Nyborg, Denmark*	2013	0.8
Other		
- Journal clubs	2009-2013	1.0
- Conferences Nederlandse Vereniging voor Technische Geneeskunde	2010-2013	1.0
- Scientific meetings Nederlandse Vereniging voor Technische Geneeskunde	2011-2013	0.2
- KNO wetenschapsdag	2012	0.2

2. Teaching

	Year	Workload (ECTS)
Lecturing		
- StAr bijscholingsdagen	2012	1.0
- Course Digitale signaalbewerking in hoortoestellen (NAN)	2012	0.5