Effects of Noise Reduction on Speech Intelligibility, Perceived Listening Effort, and Personal Preference in Hearing-Impaired Listeners

Inge Brons¹, Rolph Houben¹, and Wouter A. Dreschler¹

Abstract
This study evaluates the perceptual effects of single-microphone noise reduction in hearing aids. Twenty subjects with moderate sensorineural hearing loss listened to speech in babble noise processed via noise reduction from three different linearly fitted hearing aids. Subjects performed (a) speech-intelligibility tests, (b) listening-effort ratings, and (c) paired-comparison ratings on noise annoyance, speech naturalness, and overall preference. The perceptual effects of noise reduction differ between hearing aids. The results agree well with those of normal-hearing listeners in a previous study. None of the noise-reduction algorithms improved speech intelligibility, but all reduced the annoyance of noise. The noise reduction that scored best with respect to noise annoyance and preference had the worst intelligibility scores. The trade-off between intelligibility and listening comfort shows that preference measurements might be useful in addition to intelligibility measurements in the selection of noise reduction. Additionally, this trade-off should be taken into consideration to create realistic expectations in hearing-aid users.

Keywords
noise reduction, hearing aids, intelligibility, perceived listening effort, preference

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 aid’s gain in specific frequency bands to suppress unwanted background noise. Generally, hearing-aid noise reduction is presented as a black box, as no details on the processing are given in the specifications of the hearing aid. It is therefore often unclear how the hearing-aid gain changes if the noise reduction is switched on in a specific hearing aid and whether the effects differ between hearing aids. This lack of information on the effects of noise reduction complicates the selection of noise reduction for an individual user.

In a recent study, we compared noise reduction from different hearing aids to gain some insight in the effects of noise reduction (the black box) on the speech signal (Brons, Houben, & Dreschler, 2013). In short, acoustical analyses showed that noise-reduction implementations differ among hearing aids, and perceptual measurements showed that these differences are perceptually relevant for normal-hearing listeners. Noise-reduction implementations differed perceptually from each other in the degree to which they influenced the noise annoyance and speech naturalness perceived by normal-hearing listeners, resulting in differences in preference. Finally, small differences in speech intelligibility and listening effort were found among noise-reduction systems but not between noise reduction on and off.

In this follow-up study, we investigated whether these findings also hold true for hearing-impaired listeners. It

¹Clinical and Experimental Audiology, Academic Medical Center, Amsterdam, The Netherlands

Corresponding author:
Inge Brons, Clinical and Experimental Audiology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.
Email: I.Brons@amc.uva.nl
might be that hearing-impaired listeners are less sensitive to differences between processing conditions because of suprathreshold deficits such as reduced frequency selectivity or impaired modulation detection (Marzinzik, 2000). On the other hand, because of their decreased ability to understand speech in noise, it might be more important for hearing-impaired listeners to avoid distortions of the speech signal. In this phase, we evaluated noise reduction in isolation. In a later stage, possible interactions between noise reduction and compression should be addressed. The goal of the current study was to answer the following research question: Does hearing-aid noise reduction influence speech intelligibility, listening effort, noise annoyance, speech naturalness, and preference for listeners with moderate sensorineural hearing loss, compared with (a) no noise reduction and (b) noise reduction from other linearly fitted hearing aids?

**Methods**

The methods for hearing aid recording, perceptual measurements, and statistical analyses were identical to those described by Brons et al. (2013). Approval for this experiment was obtained from the Medical Ethical Committee of the Academic Medical Center on November 29, 2011 (MEC2011_310).

**Hearing-Aid Recordings**

We recorded hearing-aid output of three linearly fitted hearing aids from different manufacturers (Phonak Exelia M, ReSound Azure AZ80-DVI, and Widex Mind 440) using the method described by Houben, Brons, and Dreschler (2011). Analyses of recordings from these hearing aids fitted for different hearing losses revealed that noise-reduction processing in this selection of hearing aids was independent of hearing loss (i.e., the patterns of gain reduction remained the same for the same input signals when the hearing aids were fitted for another hearing loss), so that it was not necessary for the current purpose to fit the hearing aids to other targets than in Houben et al. We therefore took the same hearing aids and settings as in that study. In short, all hearing aids were linearly fitted with all signal-processing features turned off for the unprocessed condition, and with the strongest available noise-reduction setting on for the noise-reduction conditions. Compensation for individual hearing loss was done after recording and filtering. Compensation was done according to the linear National Acoustic Laboratories’ prescription for hearing-aid gain and frequency response—Revised, Profound (NAL-RP) (Byrne & Dillon, 1986). Recordings of the three hearing aids with noise reduction activated were randomly coded as conditions NR1, NR2, and NR3. This coding is the same as in Brons et al. (2013).11

All recordings were filtered with an inverse filter (Houben et al., 2011) that corrected for differences in frequency response between hearing aids. Thus, if noise reduction was inactive, all recordings had the same output spectrum as the input signal. There were no perceptual differences between recordings from different hearing aids when noise reduction was inactive, as was verified by Houben et al. (2011). We therefore used recordings of one hearing aid that formed the unprocessed condition, representing all hearing aids with noise reduction inactivated. Recordings with noise reduction activated were filtered with the same filter. Differences among noise-reduction conditions from different hearing aids were thus caused by the noise reduction only and not by other hearing aid characteristics.

Stimuli consisted of Dutch sentences (Versfeld, Daalder, Festen, et al., 2000) in babble noise (20 talkers reading simultaneously different passages; Auditec, St. Louis, MO), recorded from the hearing aids with a hearing-aid input noise level of 65 dB(A). The recorded stimuli were presented monaurally to the subjects with Sennheiser HDA200 headphones. During the listening experiments, the noise level was 65 dB(A) for all the stimuli in the unprocessed condition, and the average speech levels ranged from 61 to 75 dB(A) to obtain all input SNRs that were required for the stimuli (−4 to +10 dB SNR). Additional amplification was applied according to the linear NAL-RP prescription (Byrne & Dillon, 1986) to compensate for listeners’ individual hearing loss.

Acoustical analyses of the noise-reduction processing of these hearing aids are given in Brons et al. (2013) and summarized here in Figure 1, where the long-term average gain reduction for the three noise-reduction conditions is plotted for the six different SNRs (−1 through +4 dB) that were used in this study. A more negative gain value in Figure 1 indicates stronger noise reduction. All noise-reduction algorithms apply more gain reduction at lower SNRs, except for NR3 at frequencies between 1 and 2 kHz. NR3 reduced gain only for frequencies below 1 kHz and increased gain slightly between 1 and 2 kHz. NR1 and NR2 applied gain reduction over a broader range of frequencies, but NR1 was more cautious around 1 kHz. The analyses in Brons et al. showed that NR1 and NR2 change gain rather quickly to amplify speech when present and to attenuate the noise when speech is not present (e.g., in between speech segments), whereas NR3 applies its gain reduction more gradually.
Subjects

Twenty hearing-impaired subjects between 48 and 69 years of age (average = 61.3 years) participated in this study. We used the results obtained in Brons et al. (2013) from normal-hearing subjects for a power calculation. The power calculation revealed that a number of 10 subjects would be sufficient to detect a difference of about 13% in intelligibility score (which is equal to about 1 dB change in SRT50, i.e., the speech reception threshold, the SNR at which the subject can correctly reproduce 50% of the sentences). Also, 10 subjects would be sufficient to detect a difference of 1 rating point in perceived listening effort. Because we expected that the variation between subjects would be higher for hearing-impaired subjects than for normal-hearing subjects, we decided to include 20 subjects in the current 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, Vlaming, and Dahlquist (2010). Figure 2 shows the hearing thresholds for the ears included (one per subject) averaged over all subjects, and the corresponding standard deviation. Figure 2 also shows the standard audiogram N3 on which the selection of the subjects was based.

The two intelligibility outcome measures were (a) the subjects’ individual SRT50 and (b) the percentage correct words at a fixed SNR of +4 dB. The outcomes for listening effort and preference were measured at both subjects’ individual SRT50 (averaged over the four conditions) and at a fixed SNR of +4 dB. The +4 dB SNR was previously used in Brons et al. (2013) for measurements with normal-hearing subjects and roughly corresponds to daily listening situations for the hearing impaired (Smeds, Wolters, & Rung, 2012).

Intelligibility

Following the adaptive procedure described by Plomp and Mimpen (1979), we measured the SRT50. At the fixed SNR of +4 dB, we measured the percentage of words correctly repeated. This is similar to Brons et al. (2013) but at a higher SNR. Both measurements started with 13 training sentences followed by one list of 13 sentences per processing condition. The order of the lists
and noise-reduction conditions as well as the combinations of list and condition were balanced across subjects to minimize possible effects of differences between lists or training effects.

**Listening Effort**

The subjects rated the listening effort on a 9-point rating scale that ranged from no effort to extremely high effort. The five labels on the scale were based on the ITU-T P.800 standard (ITU-T: Recommendation: P.800, 1996). The subjects gave ratings for the four processing conditions at the SRT50 level and at −4, +4 and +10 dB SNR, thus for 16 different conditions. Each subject started with a practice run of 16 sentences, which was followed by three additional runs that were used for analysis.

**Paired-Comparison Rating**

We used paired-comparison ratings to measure noise annoyance, speech naturalness, and overall preference. All processing conditions were compared with each other, by presenting the same sentence in two different conditions. For each combination of processing conditions, subjects indicated which was best on noise annoyance, which on speech naturalness, and which they would prefer for prolonged listening. For each of these three criteria, there were seven possible answers, ranging from A is much less annoying/much more natural/much better to B is much less annoying/much more natural/much better, with the option to indicate no difference between A and B. These seven response categories correspond with those of the comparison category rating method described in ITU-T P.800 (ITU-T: Recommendation: P.800, 1996). All six combinations of conditions were measured three times, both at individual SRT50 level and at +4 dB SNR.

**Results**

**Intelligibility**

Figure 3 shows the group results for speech intelligibility. We used paired t tests to determine whether the noise-reduction algorithms changed intelligibility compared with unprocessed. SRT50 results showed no difference between noise reduction and unprocessed, but analysis of the rationalized arcsine units-transformed (Studebaker, 1985) percentage correct scores showed that scores for NR2 at +4 dB input SNR were significantly lower than for unprocessed \( p = .005; \text{Bonferroni-corrected threshold } = .05/3 = 0.0167 \).

The other outcome measures described in this article were measured at the individual SRT50, averaged over noise reductions and rounded to whole decibels, ranging from −1 to +4 dB.

**Listening Effort**

Figure 4 shows the group-average listening-effort ratings relative to that for unprocessed at SRT50 level and

---

**Figure 3.** Mean and 95% confidence interval of the SRT50 (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, and NR3 are the hearing-aid noise reductions. The horizontal line indicates which processing conditions differ significantly from each other \( ^* p < .05/6 = .0083; \text{Bonferroni-corrected thresholds for six comparisons} \).
averaged over the three fixed levels (upper panel), and the average absolute ratings for the three fixed SNRs separately (lower panel).

Because the rating scale has an upper and a lower bound, the variance will be lower in the scores near the bounds of the scale than in the middle. We therefore applied an arcsine transformation to the listening effort scores to satisfy the criteria for an analysis of variance (ANOVA; Fink, 2009). A repeated measures ANOVA on the arcsine-transformed absolute data on SRT level showed a significant effect of processing condition, $F(3, 57) = 2.9, p = .043$, but pairwise comparisons were not significant after Bonferroni correction. The arcsine-transformed absolute data for the fixed SNRs showed

![Figure 4](image-url)

**Figure 4.** Mean and 95% confidence interval of the listening effort ratings assigned by the 20 subjects relative to unprocessed (Δ listening effort, upper panel) at SRT level (left) and averaged over the three fixed SNRs (right), and absolute ratings at −4, +4, and +10 dB SNR (lower panel). Horizontal lines indicate which processing conditions differ significantly from each other (*p < .05/6 = .0083; Bonferroni-corrected threshold for six comparisons).
significant effects of SNR, \(F(2, 38) = 124.3, p < .001\), and processing condition, \(F(3, 57) = 2.8, p = .047\). Bonferroni-corrected pairwise comparisons of processing conditions showed that more effort was required for NR3 than for both unprocessed \((p = .0022, \text{threshold } = .0083)\) and NR1 \((p = .0080, \text{threshold } = .0083)\).

**Paired Comparison Rating**

Figure 5 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, 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. In each comparison, both conditions receive a (sign-reversed) rating so that the mean of all ratings is zero. For statistical analysis, we modeled the data using a log-linear modeling approach for ordinal paired comparisons (Dittrich, Hatzinger, & Katzenbeisser, 2004). A repeated measures ANOVA on the results showed a significant effect of processing condition for all criteria, \(F(3, 57) > 4.3, p < .008\), except for speech naturalness at SRT level, \(F(3, 57) = 1.1, p = .173\). Horizontal lines in Figure 5 indicate which processing conditions differed significantly from each other after Bonferroni correction.

**Discussion**

**Intelligibility**

Word scores for NR2 were significantly lower than those for unprocessed. In the results for normal-hearing listeners (Brons et al., 2013), NR2 also had the lowest word score although not significantly lower than unprocessed. Most studies have found no effect of noise reduction on speech intelligibility (Loizou & Kim, 2011; Nordrum, Erlar, Garstecki, & Dhar, 2006). Results of Hu and Loizou (2007) suggest that noise reduction reduces intelligibility more at lower SNRs. The more the noise is dominating the input signal, the more difficult it is for the noise-reduction algorithm to recognize the speech and to correctly separate the speech and noise. This might introduce more classification errors, resulting in speech distortions. Our intelligibility results did not confirm this larger decline due to noise reduction at lower SNRs; whereas the intelligibility was reduced somewhat by noise reduction at +4 dB SNR, it was not at the SRT level, which was measured at lower SNRs for most subjects. A possible explanation can be found in the dynamic behavior of the noise reduction. At higher SNR, the noise reduction shows larger and quicker changes in gain to separate between speech and noise, whereas at lower SNRs, gain is more gradually reduced because speech is not well recognized (Brons et al., 2013).

Speech distortions by quick gain transitions may be more detrimental to speech intelligibility than a more gradual suppression of the speech signal. Hilkhuyzen, Gaubitch, Brookes, & Huckvale (2012) found no interaction between noise reduction and SNR. However, they did not take measurements at positive SNRs.

**Listening Effort**

Subjects rated effort at the fixed SNRs significantly higher for NR3 than for unprocessed and NR1. This finding agrees with that for normal-hearing listeners, who also rated the highest listening effort for NR3 (Brons et al., 2013). Whereas Bentler, Wu, Kettel, & Hurtig (2008) found a reduction of perceived listening effort due to hearing-aid noise reduction, most other studies using a rating scale for determining listening effort did not (Alcântara, Moore, Brian, Kühl, & Launer 2003; Brons et al., 2013; Desjardins & Doherty, 2014). Desjardins and Doherty (2014) measured listening effort both using a dual task and a rating scale and found a reduction in listening effort with the dual task due to hearing-aid noise reduction in the same conditions where ratings of listening effort showed no difference between noise reduction on and off. This implies that a method more sensitive than a rating scale is effective. However, the positive effect of noise reduction on listening effort in Desjardins and Doherty was only found at SRT<sub>50</sub> levels and not for higher, arguably more relevant, SNRs. Recently, cognitive factors such as listening effort have enjoyed increased attention in the evaluation of hearing-aid functions. Apart from factors such as noise type and SNR, the cognitive capacity of the listener may also be important in determining which noise reduction processing should be applied in a specific situation (Lunner, Rudner, & Rönnberg, 2009; Rudner, Lunner, Behrens, Thorén, & Rönnberg, 2012).

The data in this study reveal that the absolute effort ratings by the hearing-impaired subjects for the -4 and +4 SNR conditions were higher than those reported for normal-hearing subjects in Brons et al. (2013). This difference between subject groups was also found by Luts, Eneman, Wouters, et al. (2010) and reflects the fact that hearing-impaired listeners have more difficulty with listening to speech in noise, which is also reflected in the intelligibility results.

**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. Except for the higher speech-
Figure 5. Mean rating scores derived from the paired-comparison data for the three judgment 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 lines indicate which processing conditions differ significantly from each other (*p < .05/6 = .0083; **p < .001/6 = .00017; Bonferroni-corrected threshold for six comparisons).
naturalness rating for NR2 by the hearing impaired, the results at +4 dB SNR agree well with those of Brons et al. (2013) for normal-hearing subjects. Although the strength of preference cannot directly be compared because the ratings by the normal-hearing subjects were based on comparisons of five conditions instead of four, the ranking of these four conditions is the same in the normal-hearing listeners and the hearing-impaired listeners.

NR2 reduced noise annoyance more than the other conditions and had higher speech naturalness than the other conditions. The combination of reduced noise annoyance with a high speech naturalness is remarkable because in general stronger reduction of noise is accompanied by more speech distortion (Houben, Dijkstra, & Dreschler, 2012; Loizou, 2007). Perhaps this uncharacteristically high rating of speech naturalness was based on different aspects. At the SNR under consideration (SNR = +4 dB), normal-hearing listeners rated no difference in speech naturalness among conditions (Brons et al., 2013). This suggests that at +4 dB SNR, the noise reduction introduced no audible distortions because that would have reduced the perceived speech naturalness for normal-hearing listeners. In the absence of audible distortions, the hearing-impaired subjects might have used other cues to rate naturalness, for instance, the absence of noise (Marzinzik, 2000). At −4 dB SNR, normal-hearing subjects rated speech naturalness lower with noise reduction, indicating that the distortions due to noise reduction increase with decreasing SNR. To determine whether this effect was also present in the data for hearing-impaired subjects, we repeated the analysis with the subjects divided in two groups based on their SRT (12 subjects with SRT −1, 0, or +1 dB and 8 subjects with SRT +2, +3, or +4 dB). For the first group, the ratings at SRT were based on relatively unfavorable SNRs, and here subjects rated the speech naturalness highest for the unprocessed condition and significantly lower for NR3. For the latter group, the ratings were based on higher SNRs, and subjects from this group rated the naturalness lowest for the unprocessed condition and significantly higher for NR2. This finding confirms that noise reduction appears to affect speech naturalness more at lower SNRs as was previously found for normal-hearing subjects. Neher, Grimm, and Hohmann (2014) also found that preference for noise reduction over no noise reduction was stronger at higher than at lower input SNRs. Boymans and Dreschler (2000) and Ricketts and Hornsby (2005) also found preference for noise reduction on over off at positive SNRs, in contrast to Alcántara et al. (2003), who measured mainly at negative SNRs.

The condition that was most preferred by the subjects (NR2) also produced the lowest intelligibility scores. Such a trade-off between preference and intelligibility was previously found to be inherent to noise reduction in several studies (Brons, Houben, & Dreschler, 2012; Neher et al., 2014; Wang, 2008). Apparently, due to the reduced noise level, it is more comfortable to listen to speech and noise that were processed with noise reduction, even though the reduction in speech level or distortions of the speech signal may cause lower intelligibility scores.

**Signal-to-Noise Ratios for Evaluation of Noise Reduction**

Noise-reduction processing depends on the input SNR (Hoetink, Körössy, & Dreschler, 2009). We therefore measured not only at an individual SNR (SRT90) for each subject (to ensure an equal performance level for all subjects) but also at a fixed SNR to ensure equal noise-reduction processing for all subjects. Group results were similar between the two independent datasets obtained at SRT 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 listener’s 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.

**Limitations**

The results of this study were measured in a laboratory setting and cannot easily be generalized beyond the limited number of conditions included. In addition, noise reduction was studied in isolation, whereas in practice it will often be used in combination with other hearing-aid features such as directional microphones and other signal processing algorithms. The most important signal processing that should be investigated in combination with noise reduction is dynamic-range compression, the effects of which may be opposite to that of noise reduction (Anderson, Arehart, & Kates, 2009; Chung, 2007).

**Implications for Hearing Aid Fitting**

Because of the mentioned limitations, the hearing aid specific results cannot be used to conclude which of the hearing aids tested is best. Nevertheless, this study indicates that differences in the types and implementations of hearing-aid noise reduction are perceptually relevant and this may have consequences for the selection and fitting
of hearing aids. At least for listeners with specific complaints in noisy environments, it might be worthwhile to perform technical and perceptual comparisons to select the best noise-reduction system for the individual listener. Additionally, it is important to raise realistic expectations on the effects of noise reduction. Listeners should be aware that no improvement in intelligibility scores in noise should be expected, but that single-microphone noise reduction might improve listening comfort and reduce noise annoyance.

Conclusions

Noise reduction from 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.

Overall, the results of hearing-impaired subjects agree well with those obtained for normal-hearing listeners in a previous study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the Heinsius-Houbolt Fund.

Note

1. The fourth noise-reduction condition in that study, NR4, appeared to be inactive for positive signal-to-noise ratios (SNRs) and was therefore left out of the current study.

References

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

Anderson, M. C., Arehart, K. H., & Kates, J. M. (2009). The acoustic and perceptual effects of series and parallel processing. *EURASIP Journal on Advances in Signal Processing*, 1–20. doi:10.1155/2009/619805.

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

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

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

Brons, I., Houben, R., & Dreschler, W. A. (2012). Perceptual effects of noise reduction by time-frequency masking of noisy speech. *The Journal of the Acoustical Society of America*, 132, 2690–2699.

Brons, I., Houben, R., & Dreschler, W. A. (2013). Perceptual effects of noise reduction with respect to personal preference, speech intelligibility, and listening effort. *Ear and Hearing*, 34, 29–41.

Byrne, D., & Dillon, H. (1986). The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, 7, 257–265.

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

Desjardins, J. L., & Doherty, K. A. (2014). The effect of hearing aid noise reduction on listening effort in hearing-impaired adults. *Ear and Hearing*. Retrieved from http://ovidsp.tx.ovid.com/sp-3.13.0b/ovidweb.cgi?AS= GECBFDPDNBDNMKNCBLKPEOBDOHMAA00&Link+Set=S.sh.22.23.27.31%7c45%7csl_10.

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.

Fink, E. L. (2009). The FAQs on data transformation. *Communication Monographs*, 76, 379–397.

Hilkhuyzen, G., Gaubitch, N., Brooke, 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, 531–539.

Hoetink, A. E., Körösséy, 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, 444–455.

Houben, R., Brons, I., & Dreschler, W. A. (2011). A method to remove differences in frequency response between commercial hearing aids to allow direct comparison of the sound quality of hearing-aid features. *Trends in Amplification*, 15, 77–83.

Houben, R., Dijkstra, T. M. H., & Dreschler, W. A. (2012). Analysis of individual preferences for tuning of noise-reduction algorithms. *Journal of Asian Earth Sciences*, 60, 1024–1037.

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

ITU-T: Recommendation: P.800. (1996). Methods for subjective determination of transmission quality. In *International telecommunications union, telecommunications standardization sector* (pp. 1–37). Geneva, Switzerland: International Telecommunications Union.
Loizou, P. C. (2007). *Speech enhancement: Theory and practice*. Boca Raton, FL: 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, 47–56.

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

Luts, H., Eneman, K., & Wouters, J., et al. (2010). Multicenter evaluation of signal enhancement algorithms for hearing aids. *Journal of the Acoustical Society of America*, 127, 1491–1505.

Marzinzik, M. (2000). *Noise reduction schemes for digital hearing aids and their use for the hearing impaired*. Aachen, Germany: Shaker Verlag.

Neher, T., Grimm, G., & Hohmann, V. (2014, September/October). Perceptual consequences of different signal changes due to binaural noise reduction: Do hearing loss and working memory capacity play a role? *Ear & Hearing*, 35, e213–e227. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=ovfto&AN=00003446-201409000-00017.

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, 81–91.

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

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, 270–277.

Rudner, M., Lunner, T., Behrens, T., Thorén, E. S., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal of the American Academy of Audiology*, 23, 577–589.

Smeds, K., Wolters, F., & Rung, M. (2012). Estimation of realistic signal-to-noise ratios. Poster Presented at the International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA.

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

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

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