Subjective and objective evaluation methods of complex hearing aids

Holube, I., Fredelake, S., Hansen, M.

Institute of Hearing Technology and Audiology, University of Applied Sciences, Oldenburg/Ostfriesland/Wilhelmshaven, Oldenburg, Germany

Keywords: Hearing aids, multi-channel dynamic compression, noise reduction algorithm, speech intelligibility, paired comparisons

Introduction

Hearing aids and their components are evaluated with different methods at several stages during their development. Electroacoustical measurements based on current standards do not reflect the abilities of complex hearing aids. Therefore, new objective measures of their performance should be applied as evaluation methods for today's signal processing hearing aids. Due to the complex nature of their modern algorithms, the technical descriptors should try to also reflect as much as possible the subjective judgements as obtained from hearing aid users.

Subjective evaluation methods

Subjective evaluation methods include speech intelligibility tests, subjective ratings on different quality and performance scales, and questionnaires. Questionnaires will not be discussed in this paper. When using speech intelligibility tests, mostly word or sentence scoring is applied while presenting the speech material in quiet or in noise at different levels or signal-to-noise ratios (SNR). Speech intelligibility tests in quiet normally lack the ability to show an advantage of dynamic compression algorithms at medium speech level. Improvements are mainly restricted to low levels where dynamic compression applies more gain than linear hearing aids (see e.g., Moore et al., 1992). Speech intelligibility tests in noise on the other hand focus on the SNR for 50 % speech intelligibility, i.e. the speech reception threshold (SRT). Unfortunately, the SRT of the majority of speech audiometry material is located at low, negative SNRs, at least for normal hearing listeners and those with a mild hearing loss. This is considered as one reason why noise reduction algorithms have shown mostly none or only a small advantage in these tests so far (see e.g., Chalupper, 2006).

Recently, the Acceptable Noise Level (ANL, Nabelek et al., 1991) has again come into focus as a measure for the evaluation of noise reduction algorithms (Mueller et al. 2006). For this measure, the most comfortable level of speech in quiet (MCL) and the tolerable level of a background noise while speech is present (BNL) are subjectively adjusted by the listeners similar to the procedure used in a just follow conversation (JFC) test. The term ANL was introduced as the difference (MCL-BNL). The results of individual listeners show large deviations of up to 30 dB, but ANLs are mostly larger than 0 dB, thereby giving noise reduction algorithms a chance for improvements. Schlüter et al. (2007) applied this measurement method to several noise reduction schemes and found a measurable, but smaller than expected difference in the ANL when noise reduction was turned on and off.

Another approach was proposed by Gabrielsson and Sjogren (1979). They introduced several sound quality scales for absolute judgements. Nevertheless, when evaluating small differences, paired comparisons are preferred because of their accuracy. Subjective judgements of paired comparison testing can be transferred to a preference rank scale using e.g., Thurstone (1927). The results derived for this article were obtained using this procedure. For a detailed description see Fredelake (2006).

Objective evaluation methods

Objective measures are based on a comparison of the output of the hearing aid to its input. They determine primarily the effect of compression and noise reduction algorithms. To test different objective methods, a simulated hearing aid was implemented on a PC, comprising the digital signal processing of a multi-channel compression and a noise reduction system, and different evaluation tools were applied.

The modulation spectrum of the output signal can be altered compared to the modulation spectrum of the input signal in dependence on the non-linear signal processing of the hearing aid. These changes can be described by the modulation transfer function, as proposed, e.g. by Festen and Van Dijkhuizen (1999), when using speech as a natural input signal. By averaging across the modulation frequencies and weighted averaging across the peripheral frequencies, the modulation transfer function can be merged to one single value, MTF. Details on the calculation procedure can be found in Holube et al. (2005). This MTF measure was also applied to a number of commercial hearing aids. The dynamic compression algorithm reduced the modulation depth of the input signal by an amount depending on the parameter settings of the compressor and on the modulation frequency. Some noise reduction algorithm, on the other hand, seemed to aim at increasing the MTF by reducing the noise in the soft parts of the input signal.

While the traditional standards focus on the long-term average transfer function of the hearing aid, which is a useful concept for linear hearing aids, the performance of a non-linear hearing aid for speech, or other dynamic signals, can more adequately be described by the effect of compression on the short-level distribution of the input and output signals, on a short time frame by frame basis (Elberling and Naylor, 1996). The short-term levels were plotted into an input-output diagram and a first-order linear regression line was fitted to the data. The inverse of its slope was interpreted as the effective compression ratio CR_{eff}. When comparing 1/MTF to CR_{eff}, both values matched each other closely. Both values are smaller than the corresponding static compression ratio and are decreasing even more with increasing attack times.

Another objective measure is the coherence based quality measure developed by Anderson et al. (2006). The model used for this method includes the coherence between the input and the output signal in combination with a simplified auditory model (Speech Intelligibility Index, SII). From the coherence, the signal-to-distortion ratio (SDR) is computed and used as a substitution for the SNR in the SII calculation. Also this measure shows similar results when comparing to the MTF for a dynamic compression algorithm.

The forth measure which is comparable to the MTF for noise reduction algorithms is the ΔS/N proposed by Hagerman and Olofsson (2004). For deriving the ΔS/N, speech is added to noise at a specific SNR and processed by the algorithm. The same processing is done for the same speech added to the same but inverted noise. From the two output signals, two related signals, corresponding to the speech and the noise alone, can be calculated. Their SNR is subtracted from the SNR at the input of the processing scheme, thus yielding the ΔS/N.

Results

The different objective and subjective evaluation methods show a high correlation between most of them. The highest correlation was found for the simulated dynamic compression algorithm. Fig. 1 shows the relation between the subjective rank scales resulting from paired comparison testing and the MTF for normal hearing (left) and for hearing impaired subjects (right). The different symbols represent different parameter settings of the dynamic compression algorithm (number of channels, compression ratio, and release time). Normal hearing as well as hearing-impaired listeners rated the unprocessed signal or dynamic compression with a small number of channels, low compression ratios and long release times the highest. A difference in the two subject groups is prominent in the different ranges of the preference rank scale. This difference is due to the reduced ability of hearing-impaired subjects to perceive dissimilarities between the presented intervals.

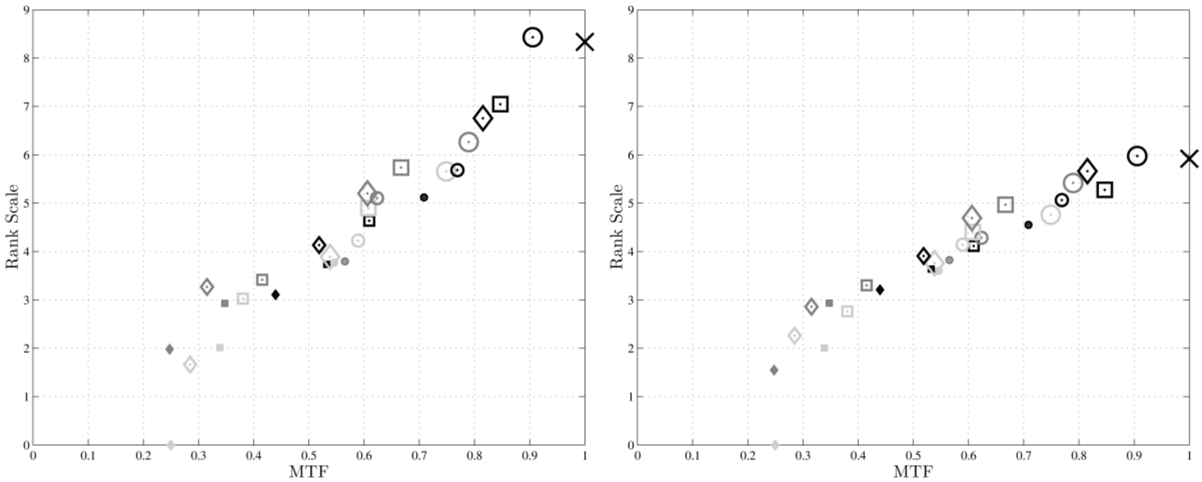

Fig. 1: Rank scale of overall quality from paired comparison testing vs. MTF for normal hearing (left) and for hearing-impaired listeners (right) for different parameter settings of the dynamic compression algorithm. The crosses represent linear processing (taken from Fredelake, 2006).

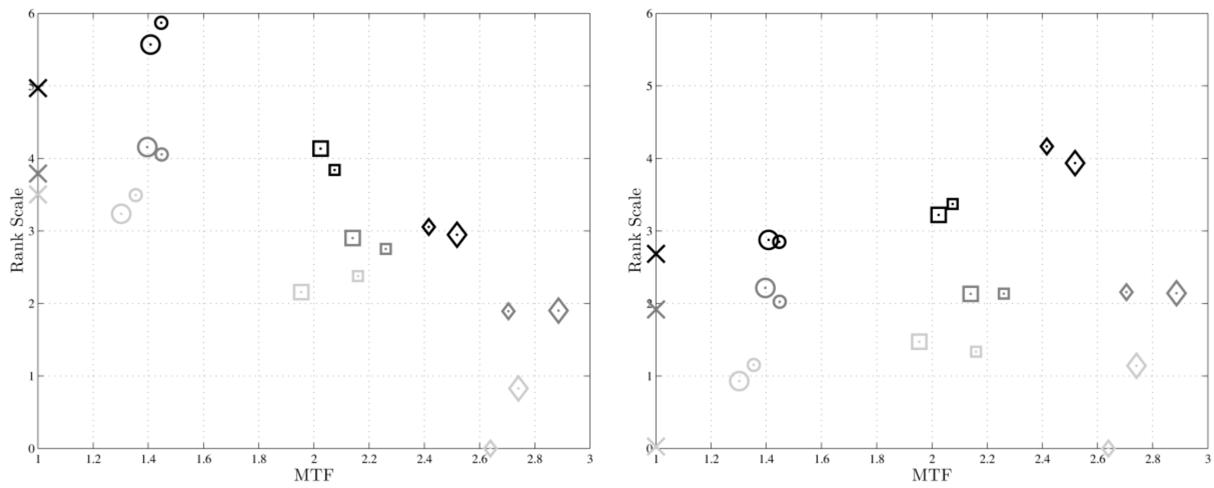

Fig. 2: Rank scale from paired comparison testing focussing on the naturalness of the sounds vs. MTF or normal hearing (left) and for hearing-impaired listeners (right) for different parameter settings of the noise reduction algorithm. The crosses represent linear processing (taken from Fredelake, 2006).

Subjective and objective assessments of noise reduction algorithms vary and depend strongly on the focus of the subjective rating, e.g. speech quality vs. noisiness. Fig. 2 shows, as an example, the results of the paired comparison testing focussing on the naturalness of the sound samples vs. the MTF. The noise reduction algorithm was based on Ephraim and Malah (1984). Available parameters were the amount of noise reduction, a smoothing factor, and the signal-to-noise ratio of the input signal. Naturally, better SNRs were rated higher. In general, a different pattern can be observed for hearing-impaired subjects when being compared to normal hearing subjects. Normal hearing subjects reported a reduced naturalness when noise reduction is increasing. On the other hand, hearing-impaired subjects did not show this effect and even reported an increase in naturalness from the noise reduction algorithms at the better SNRs. They seemed to be less disturbed by artefacts introduced by the algorithm.

Conclusions and Outlook

Several reasonable approaches for the analysis of complex hearing aids have been proposed by different research groups. When comparing the objective measures to each other and to subjective judgements in paired comparison tests, high correlations were observed for different compression systems. The relationship between subjective and objective results for a noise reduction algorithm was less satisfying. This might be due to subjective criteria and the influence of the hearing loss which is not yet included in the objective measures. Still under discussion is the transfer of the results in the laboratory to real world environments.

Acknowledgement

This study has been funded by AGIP, the Lower Saxony Department of Science and Culture, Hanover, the European regional funding EFRE, HörTech centre of excellence, Oldenburg, and Kind Hörgeräte, Burgwedel

Literature

Chalupper, J. (2006) “Are there perceptual benefits of digital noise reduction?”, AudioInfos, 61,14-22.

Elberling, C. and Naylor, G. (1996). “Evaluation of non-linear signal-processing hearing aids using speech and noise signals,” In: Issues in Advanced Hearing Aid Research Conference, May 1996, Lake Arrowhead, CA, USA, 24.

Ephraim, Y. and 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, 1109-1121.

Festen, J. M. and Van Dijkhuizen, J. N. (1999). “Modelling the speech-reception threshold for amplitude-compressed speech,” In: Dau, T., Hohmann, V., Kollmeier, B. (eds) Psychophysics, physiology and models of hearing, World Scientific, Singapore, 249-254.

Fredelake, S. (2006). “Untersuchung der perzeptiven Relevanz der Modulationstransferfunktion von Hörgerätealgorithmen,“ Master thesis, University of Oldenburg.

Gabrielsson, A. and Sjorgren, H. (1979). “Perceived sound quality of hearing aids,” Scandinavian Audiology 8, 159-169.

Hagerman, B. and Olofsson, A. (2004). “A Method to Measure the Effect of Noise Reduction Algorithms Using Simultaneous Speech and Noise,” Acta Acustica united with Acustica 90, 356-361.

Holube, I., Hansen, M. Schultz-Amling, R. and Fredelake, S. (2005). “The use of the modulation transfer function to describe the performance of hearing aids,” In: Hearing Aid Fitting by Rasmussen, A., Poulsen, T., Andersen, T., Simosen,

J. B. and Larsen, C. B. (Eds), 21st Danavox Symposium, Aug 31 – Sept 2, Kolding, Denmark.

Anderson, M. C., Arehart, K. H., Kates, J. and Harvey, L. O. (2006). "Preditions of speech quality under conditions of noise and distortion," International Hearing Aid Conference, Lake Tahoe, CA, USA.

Moore, B. C. J., Seloover Johnson, S., Clark, T. M. and Pluvinage, V. (1992). "Evaluation of a Dual-Channel Full Dynamic Range Compression System for People with Sensorineural Hearing Loss," Ear and Hearing 13(5), 349-370.

Mueller, G., Weber, J., Hornsby, B. W. Y. (2006). „The Effects of Digital Noise Reduction on the Acceptance of Background Noise," Trends in Amplification, 10(2), 83-94.

Nabelek, A. K., Tucker, F. M. and Letowski, T. R. (1991). "Toleration of background noises: relationship with patterns of hearing aid use by elderly persons", Journal of Speech and Hearing Research 34, 679-685.

Schlüter, A., Holube, I., Bitzer, J., Simmer, U. und Brand, T. (2007). „Evaluation of Noise Reduction Algorithms in Hearing Aids with the Acceptable Noise Level Test (ANLT)," this conference proceedings.

Thurstone, L. L. (1927). "A Law of Comparative Judgement," Psychological Review 34, 273-286.