Speech Intelligibility Enhancement in Hearing Aids based on optimized Spatial Beam-Forming and Computational Auditory Scene Analysis

Anton Schlesinger and Marinus M. Boone

Delft University of Technology, P.O.B 5046, 2600 GA Delft, Netherlands, Email: a.schlesinger@tudelft.nl

Introduction

A combined processing scheme for the enhancement of speech intelligibility in hearing aids is presented. The approach utilises an optimized beam-forming method in connection with a biologically inspired processing-model of modulation perception and binaural interaction.

Departing with consensus about the relation of speech intelligibility (SI) and the signal-to-noise ratio (SNR), the necessary improvement of SI by a hearing aid complies with the improvement in SNR by at least 5 dB to be beneficial to the majority of hearing-impaired people [5]. In reaching this goal, many systems have been developed. The challenges that these approaches have to face are real-world situations, such as situations in which many talkers speak at the same time and in which fluctuating noise and reverberation deteriorate the input-signal. The objective function is to suppress interference while retaining a high degree of target speech quality.

The processing scheme that we present here is based on two approaches that proved to solve this task, respectively. The optimal beam-forming is based on endfire-arrays of microphones in the arms of a pair of spectacles. Sound, emanating in the line of vision is enhanced, while sound from all other directions is attenuated. The second approach is borrowed from the working principle of the human auditory system. It is an implementation of psychological and psychoacoustical models of how the acoustic organ forms an inner representation of the perceived sound-scape to grab only the object of interest while inhibiting interfering sources. The algorithm is attached to Computational Auditory Scene Analysis (CASA), a new field in acoustics that started out to mimic the human auditory processing performance.

It is a long formulated expectation that a combination of these methods of noise-reduction might yield a gain in SI [3]. While the beamforming enhances the sound in the line of vision without a sound-object related discrimination, the CASA-algorithm accomplishes this task and is vice versa supported by a better SNR at its input. This contribution gives a short review of the underlying technologies and concludes with a measurement of the achieved SNR of the combined processing scheme.

Review of underlying methods

For the improvement of SI in hearing aids, array-technology proved to be an excellent possibility [1]. A long succession of works at our institute dealt with the application of array-technology in hearing aids. This led to the introduction of the hearing glasses to the market a few years ago. The achieved improvement in SNR fulfils with a weighted 8 dB the above mentioned requirement. For ensuring the natural binaural listening, two end-fire arrays are integrated in the temples of a pair of spectacles. In order to obtain a robust hearing aid that is small in size and highly directive, an optimal beam-forming method was applied. This method utilizes microphone filters that realize a maximum directivity for a discrete array. The optimization procedure can be formulated as the minimization of the signal output while adhering the constraint that the amplification of the signal in the target direction is unity [1]. Using efficient digital signal processors, the end-fire arrays consist of 4 omnidirectional microphones along a line of only 72 mm length.

The CASA-approach that is applied here refers to a publication of Kollmeier and Koch [3]. This publication describes an algorithm of noise reduction in speech signals that was developed in accordance with psychological and psychoacoustical models of modulation perception and binaural interaction. The underlying principle follows the observation that the human auditory processing uses an orthogonal display of independent cues that are related to a sound-object (e.g., a talker). Formulated in a simplified way, these cues are time-information (exploitation of modulation, - a *periodotopical* organization), frequency-information (a *tonotopical* organization) and binaural disparity (interaural time and level differences, which form a *topological* organization). By this means sound-objeccts are resolved and an inner representation of the sound-scape is gathered. By intentional selection of the cerebrum, sound-objects of interest are sent to the speech area for further analysis.

The algorithm in [3] was modelled, analogously. After a peripheral frequency-analysis of the binaural input, the envelope is used to transform time-information into a modulation-display. The evaluation of the signal envelopes eases the decomposition of the sound-scape. Speech signals expose slowly changing modulation-frequencies that are roughly constant along their energy bearing frequencies-bands. It is consequently possible to disentangle a scene of sound objects that share the same carrier-frequencies. Another feature of the envelope is its robustness towards reverberation and noise-intrusions, which is one reason of the stability of the proposed algorithm. After imaging the input in a carrier-frequency and modulation-frequency representation for the left and the right channel, interaural time and level differences are assessed to establish a localization of sound-objects. By a weighting-function only objects in the target-direction (line of vision) are maintained, while objects that are beyond a threshold of level- and time-differences are suppressed. Eventually, an optimized representation of the

envelope is used to modulate the underlying band-filtered time-signal, which is subsequently resynthesised and sent to the listener.
The implementation in this work is related to the published characteristics of the Kollmeier and Koch algorithm. No profound optimization of the algorithm was performed. Informal listening tests and an analysis with the modulation transfer function (MTF) were used to achieve a stable operation in enhancing SI in different acoustical situations.

Tested results of the combined processing

In order to assess the processing scheme in real-world scenarios, we opted to implement the processing of the hearing glasses in the binaural IEM-room-simulator [4]. This software-based simulator enables an allocation of sound-objects in a user-defined acoustic environment. The incorporation of the processing of the hearing glasses was accomplished by recording the transfer-functions of the hearing glasses at playback positions of 5th order ambisonics. The measurement was performed in an anechoic room at the TU-Delft. The hearing glasses were worn by a manikin to account for the transfer-characteristics of head and torso. The CASA-processing was subsequently computed on the binaural output of the hearing glasses in an arranged virtual sound-scape. To provide an objective comparison with the original situation, for either methods, separately, and the combined processing scheme, MTF's based on a modified envelope regression method were calculated in octave-bands and used to determine the apparent SNR [2]. The test-situations were similar to the situations that were arranged in [3]. The metrics were calculated on two minutes of phonetically balanced speech material. Below, three test-situation are exemplarily analysed.
The first test-situation is anechoic and comprises a male target-speaker in the line of vision and a noise-signal of speech-shaped noise (by the target speaker) at 90° (A1N). The SNR was -10 dB, whereas all SNR-values were measured in an rms-window of 10 s and all sources were placed at 2 m distance from the receiver. For the second test, the setup remained, but a reverberation of 1.33 s was added and the SNR was set to -2 dB (R1N). In the third example, four speakers were arranged at 60° (female voice), 120° (male v.), 240° (m. v.) and 300° (f. v.). No reverberation was added and the SNR was -6 dB (A4S). Table 1 gives the results of the test.

Conclusions and outlook

The calculation of the apparent SNR shows an improvement of the SI for each approach, separately, and their combination comparing the original condition. While the hearing glasses enhance the SNR especially good at constant intrusions, the CASA-algorithm exposes its performance in babble-noise situations, where several talkers compete with each other. Considering tests A1N and R1N, the SNR erodes towards high frequencies at the left ear after processing. In that case, the hypothetical trade-off between noise reduction and signal distortion is not met [2]. Altogether, we observe a stable and significant improvement of SNR in different poor acoustical situations. By adjusting the parameters of the presented CASA-algorithm we are confident to tap more of the potential that the combined processing scheme might yield.

Table 1: Results mark SNR in dB for left ear and right ear, K denotes the manikin, KC the manikin with CASA-processing, HG are the hearing glasses and HGC are hearing glasses with CASA-processing, see text for conditions A1N, R1N and A4S.

A1N	center-frequencies octave-bands [Hz]				
	250	500	1000	2000	4000
K	-11/-16	-10/-16	-10/-17	-5/-14	0/-17
KC	-11/-16	-9/-14	-8/-15	-6/-15	-2/14
HG	-8/-5	-6/-5	-6/-6	-3/-3	-6/-9
HGC	-6/-3	-5/-3	-5/-5	-1/0	-4/-6
R1N					
K	-4/-8	-3/-8	-3/-9	2/-7	5/-10
KC	-3/-7	0/-6	0/-5	-1/-5	2/-5
HG	-1/0	0/1	0/0	3/3	1/-2
HGC	0/2	2/4	2/2	-2/0	-3/-2
A4S					
K	-9/-11	-9/-11	-10/-10	-11/-7	-5/-9
KC	-7/-8	-6/-7	-7/-5	-8/-8	-4/-10
HG	-5/-7	-4/-8	-5/-8	-7/-8	-2/-1
HGC	-1/-1	1/-2	0/-3	-5/-5	-3/-1

References

[1] Boone, M. M.: Directivity measurement of a highly directive hearing aid: the hearing glasses., AES 124th Convention Paper, 2006.

[2] Goldsworthy, R. L. and Greenberg, J. E.: Analysis of speech-based speech transmission index methods with implications for nonlinear operations., J. Acoust. Soc. Am., 2004.

[3] Kollmeier, B. and Koch, R.: Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction., J. Acoust. Soc. Am., 1994.

[4] Noisternig, M. et al.: A 3d ambisonic based binaural sound reproduction system., AES Conf. Paper, 2003.

[5] Duquesnoy, A. J. and Plomp, R.: The effect of hearing impairment on the speech-reception threshold of hearing-impaired listeners in quite and noise., J. Acoust. Soc. Am., 1983.