# On the Application of Auditory Scene Analysis in Hearing Aids

Anton Schlesinger[1], Marinus M. Boone[2]

[1] *University of Technology Delft, The Netherlands, Email: a.schlesinger@tudelft.nl*

[2] *University of Technology Delft, The Netherlands, Email: m.m.boone@tudelft.nl*

## Introduction

The individual onset of hearing impairment is often experienced as an increased listening effort in noisy situations. The problem is for the most part attributed to a sensorineural hearing loss, by which the majority of hearing-impaired people is affected. Unfortunately, the restoration by a technical solution amounts to a difficult engineering task.

Of great importance for the formulation of this task were two findings. At first, Plomp and colleagues [4, 29] revealed that hearing aids which primarily compensate for the raised threshold even increase the unintelligibility of speech in adverse acoustical situations. The consequence is frequently observed when hearing aid wearers switch off their devices the moment they enter noisy environments. The second finding refers to the necessary improvement in SNR, to be helpful for the majority of hearing-impaired people. Put in figures, the decisive discovery was that a noise reduction of 4 to 5 dB would minimize auditory handicapped of all degree by 50 % [4, 29].

Much work has been devoted to the task of recovering a desired speech-signal in noise (for a review see for example [1, 23]). By inspection of numerous technical approaches, two guiding principles are evident. First, approaches that replace auditory dysfunctions tend to realize an audiological success where approaches that mainly compensate for dysfunctions fail [5, 28]. Second, the utilization of a binaural processing and the preservation of binaural cues at the input of the auditory system is at an advantage as compared to a mere monaural processing and supply [1, 24].

There are two robust and successful approaches that both replace the affected auditory function of noise reduction and accomplish a binaural location-based grouping. These are (A) beamforming and (B) Computational Auditory Scene Analysis (CASA).

(A) Beamforming is a well established method in today's hearing aids. The directional characteristic of an array of microphones is utilized to enhance the reception of a target sound. The method of beamforming is generally categorized into fixed and adaptive beamformers. Fixed beamformers employ filters which are time-invariant, i.e., which have a fixed directivity-pattern. Adaptive beamformers change their filter continuously to steer a null of their beampattern in the direction of noise-intrusions. Adaptive beamformers show a remarkable potential in laboratory and favorable conditions. Once the acoustic conditions become more complex, the performance of adaptive algorithms tends to decline [16]. In spite of this fact, current developments demonstrate an ongoing improvement in that branch (see e.g. [17, 9, 7, 12]).

Fixed beamforming comprises mainly three techniques. These are the delay-and-sum, the gradient and the optimal beamforming method. By delay-and-sum beamforming the output in the target direction is maximized. Delay-and-sum beamforming is conceptually suitable to achieve a high gain in hearing aids [31]. However, the method suffers from the requirement that the employed array has to be large in comparison with wavelengths to establish sufficient directivity. The gradient-method allows for a frequency-independent gain in directivity and a much smaller spacing among the transducers. This method is often applied in behind-the-ear (BTE) hearing aids. Although it establishes a theoretical gain of up to 6 dB (when using 2 microphones), the method troubles with a high noise-sensitivity at low frequencies, which worsens with higher-order gradients and thus makes such implementations impractical [25]. As a matter of the small distance between transducers and the shielding effect of the pinna, common methods of beamforming in BTE hearing aids seldom exceed a gain in SNR of 3 dB.

Highly efficient is the optimal beamforming method (also known as Minimum Variance Distortionless Response (MVDR) beamformers) that generates a superdirective beampattern. The method was first developed in seismology [8] and has found much interest in hearing aid research in the recent decade in fixed and adaptive designs [3, 11, 12, 24, 25].

(B) The mammalian ability to concentrate on a single sound source within a complex auditory scene is subsumed under the physiological term of auditory scene analysis. Humans are capable to localize and to separate up to six concurrent talkers [1]. This remarkable ability is known as the "cocktail party effect". The marvelous performance of the auditory system stimulated its imitation by a bionic approach and established the research field CASA. Auditory scene analysis is generally understood as a twofold process. An analysis of the acoustic input begins with its decomposition into a place-code of audible frequencies in the cochlea. As the stimulus travels further along the nuclei of the auditory path to the cortex, a manifold feature-space is erected. The entire extend of this feature-space is not clarified [26]. Most prominent, along with the tonotopic organization, are neural responses to features like on- and offset, periodicity as well as location-based (binaural) cues. The issue of extracting a single sound source from the whole is assumed to be a preattentive binding of features to sound objects. Physiological evidence suggests the binding as an oscillatory correlation process, where the temporal structure in neural responses serves as the indication of features that belong to a particular sound object [1, 27]. The majority of CASA processors that mimic the "cocktail party effect" comply with the basic structure of the psychoacoustic model of binaural interaction. Anatomically, the principal binaural processing is located at the brainstem and midbrain levels. In these areas, sensitivities to interaural time- and level differences (ITD and ILD) are established which are maintained throughout higher neural stages up to the primary auditory cortex. The model of binaural interaction is thus a central part in the auditory system to create an inner representation of the soundscape. The basic structure of the corresponding CASA processor spans from the peripheral filterbank to a cross-correlation stage in each frequency-band. Many binaural CASA processors have proven to be successful in enhancing speech in noise [1, 5, 10, 18, 22]. However, apart from experimental implementations, the application of binaural processors in hearing aids is hindered by their high computational demands.

The subject of this review is a brief review of successful speech processors based on CASA regarding their application in hearing aids. This is followed by an example of a CASA processor in conjunction with a MVDR beamformer. Advantages of this combined processing scheme and current hurdles are outlined and finally, their overcoming is discussed in the conclusion of this contribution.

# CASA, Three Speech Processors

Three binaural CASA algorithms are briefly introduced that each represent the auditory scene analysis at the lower neural stages by means of different concepts. The summary of the algorithms is ordered by their complexity, starting with an algorithm by Kollmeier and Peissig (P-K) that is based on a model of binaural interaction [14, 5, 28]. Thereafter, the algorithm of Kollmeier and Koch (K-K) and the algorithm of Albani (A) are considered. The K-K model incorporates next to the model of binaural interaction a model of modulation perception [22] and the A model features additionally to the model of binaural interaction a model of across frequency interaction [15]. In view of an application in hearing aids, the following characteristics need to be considered:

(1) The P-K speech processor performs a frequency analysis of the binaural signal and, subsequently, analyzes the ILD and the ITD from the envelope in each frequency band. Based on this information, time frequency bins that deviate from a defined listening direction are attenuated. Moreover, the P-K processor uses a measure of coherence between the ears and is thus able to distinguish between direct and diffuse contributions of the sound field. Altogether, the P-K algorithm is able to attenuate lateral noise sources and reverberation. Regarding the improvement of speech intelligibility, the algorithm showed high efficiency in moderate SNR-situations [18]. Moreover, the P-K algorithm features a low systems delay (e.g., 7 ms in [28, 30]) and is clear in design.

(2) After the peripheral decomposition in frequency bands, the K-K processor calculates ILD and ITD in the binaural modulation spectra and suppresses centre-frequency and modulation-frequency bins that outside a predefined preferential direction. Concurrent talkers become easily separable by different fundamental modulation frequencies and the interfering talker is efficiently suppressed. The K-K processor preserves the continuity of the signal and introduces only little distortion. In the range from highly adverse to moderate SNR situations, the K-K processor yields a small (as compared to the P-K algorithm) but stable audiological gain [22, 18]. Also referring to the P-K algorithm, the transfomation into the modulation-domain asks for a higher system delay and a higher computational effort. The analysis of limiting values has yet to be done [22].

(3) Different from the processors above, the A processor incorporates a localization model that resolves spatial ambiguities as it occurs for centre-frequencies in the cone of confusion. The localization is based on an across-frequency interaction process that is essentially the summation of band-wise neural activity response probabilities of ILDs and ITDs. The processor is able to localize up to four sources in reverberation [15, 18] and is subsequently advised to track a certain source as well as to attenuate all other sources. A stabilization of the localization process and the speech enhancement is achieved by a feedback mechanism that applies parts the precedence effect and neural fascilitation. It is not known to the authors if the A processor has been evaluated regarding its performance to improve speech intelligibility.

# Example of Application

The afore summarized binaural CASA processors were developed in the early 1990s. Their application in hearing aids was not only hindered by the high computational effort. Also the mere date transfer among the hearing aids constitutes a challenge. Today these issues undergo a change, digital signal processors are shrinking along with their power consumption and efficient wireless transmission protocols might soon enable the exchange between hearing aids.

In a study we currently analyze CASA processors and their application in a combination with beamformers. Such a fusion provides a series of advantages. By means of the array processing, a stable, source-independent improvement of the SNR in the target-direction is realized. Using this gain in SNR, the binaural processor is put in its optimal working-range and its symmetrical beam-pattern is confined to the preferential direction by the superposition of the array's beam-pattern. The binaural processor further enhances speech intelligibility and quality by interpretation of source-related, i.e., spatial and prosodic information. Both methods therefore complement one another and realize an effective speech enhancement. The strength of the particular combined

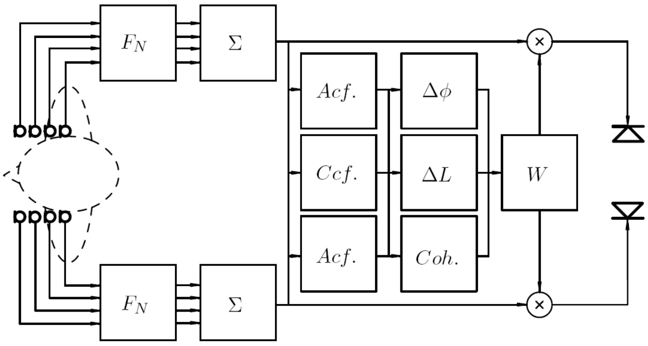


Figure 1: Hearing aid prototype, a combined processing scheme composed of a fixed binaural MVDR beamformer (implemented in the arms of hearing glasses, each beamformer is built up by the boxes $F_N$, which is the multiplication stage with a complex filter and thereafter the summation stage $\sum$) and a binaural (K-P) CASA processor. The boxes across the channels form the CASA processor (*Acf.*: is a autocorrelation function, *Ccf.*: a cross-correlation function, $\Delta\phi$: denotes the calculation stage of interaural phase difference, $\Delta L$: denotes the calculation of interaural level difference, *Coh.*: represents the calculation of the normalized coherence between the ears and $W$: is a weighting multiplier).

processing scheme was foreseen [12, 20, 22]. Here we give an example of the hearing glasses conjuncted with the P-K processor as sketched in Fig. 1. From a conceptional point of view, the presented combined processing scheme was already implemented by Lockwood et. al [12]. The authors combined the P-K-algorithm with two cardioid-receivers of BTE-hearing aids. By calculating the SNR, an improvement of 1-2 dB, as compared to the P-K processor with omnidirectional receivers, was realized in different acoustical situations. As we combine the fixed MVDR beamforming method (as implemented in the hearing glasses, see Fig. 1) with the P-K-algorithm and deliver a dichotic output, a higher audiological gain is expected. The hearing glasses use arrays of four microphones in each temple and realize under realistic conditions a gain of 7.2 dB in the line of sight. In anticipation of a fast advance in the development of digital signal processors and in consideration of the spectacles as a suitable carrier for an across-channel processor, the combined processing scheme is considered as a thinkable evolution of a highly efficient speech processor in future hearing aids.
To have maximum flexibility over the prototypical processing chain, the spatial beam-pattern of hearing glasses was rendered in virtual acoustics through a software-package [13]. Therefore, the transferfunctions of the hearing glasses were recorded and integrated in the sound reproduction system. The binaural recordings of an alterable acoustical situations are subsequently applied to an offline implementation of the binaural processor. While the beamforming of the hearing glasses has already been optimized to yield a compromise between a maximum improvement in speech intelligibility and listening ease [3], the binaural processor had to be adapted to the specific binaural output of the hearing glasses. In order to omit the laborious task of adapting the algorithmic parameters by hand to achieve an optimal performance of the combined processing scheme, a genetic algorithm [6] was applied to search for an optimal set of parameters. As an objective function, an envelope regression method to calculate the speech transmission index (STI) was employed. This method is considered to be capable to assess non-linear systems [19].
In the analyzed acoustic scene, the objective STI showed a high improvement for the combined processing scheme with respect to the mere enhancement through the hearing glasses. In a subjective listening test with 16 subjects of normal hearing, no significant improvement for the same relationship was found. The discrepancy between the objective (STI) and subjective result is attributed to the objective function (STI) by which the CASA processor was optimized. The finding asks for a replacement of the monaural and broad band STI, by a CASA based assessment stage that accounts for the individual hearing and a particular spatial sound scene (i.e., to be in compliance with the binaural masking level difference (BMLD) [26]). The full report on this study has been published in [30].

# Conclusion

This contribution introduced the problem speech intelligibility for hearing-impaired people in noise. Two methods, beamforming and CASA, were mentioned to enhance speech in noise. CASA, a computational approach to achieve the superior human speech processing, was exemplified on the basis of three speech processors. We then briefly reviewed a conjunction of a CASA processor (P-K algorithm) with a binaural beamformer (MVDR beamformer). A genetic optimization of the complex CASA processor suffered from a weak objective function, the monaural STI. Therewith, the improvement in STI was not approved by an audiological test of the combined processing scheme.
Several objective models of the auditory scene analysis developed to assess speech intelligibility in spatial conditions have been developed [2, 21, 32]. The adaption of these models to optimize and to assess CASA speech processors for an individual hearing results in an integral bionic cycle and should yield highly efficient hearing aids.

# References

[1] *Computational Auditory Scene Analysis.* A John Wiley & Sohns, Inc., Publication, 2006.

[2] R. Beutelmann and T. Brand. Prediction of speech intelligibility in spatial noise and reverberartion for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.*, 2006.

[3] M. M. Boone. Directivity measurements on a highly directive hearing aid: the hearing glasses. *AES 120th Convention*, 2006.

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

[5] B. Kollmeier et al. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *J. of Rehabilitation Research and Development*, 1993.

[6] C. Houck et al. A genetic algorithm for function optimization: a matlab implementation. *North Carolina State University, Raleigh, NC, Technical Report*, 1995.

[7] C. Liu et al. A two-microphone dual delay-line approach for extraction of a speech sound in the presence of multiple interferers. *J. Acoust. Soc. Am.*, 2008.

[8] Capon et al. Multidimensional maximum likelihood processing of a large aperture seismic array. *Proceedings of the IEEE*, 1967.

[9] D. Welker et al. Microphone-array hearing aids with binaural output–part ii: A two-microphone adaptive system. *IEEE Transactions on Speech and Audio Processing*, 1997.

[10] J. B. Allen et al. Multimicrophone signal-processing technique to remove room reverberation from speech signals. *J. Acoust. Soc. Am.*, 1977.

[11] J. G. Desloge et al. Microphone-array hearing aids with binaural output–part i: Fixed-processing systems. *IEEE Transactions on Speech and Audio Processing*, 1997.

[12] M. E. Lockwood et al. Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms. *J. Acoust. Soc. Am.*, 2004.

[13] M. Noistering et al. A 3d ambisonics based binaural sound reproduction system. *in Proc. AES 24th Int. Conf. on Multichannel Audio, Banff*, 2003.

[14] N. I. Durlach et al. Note on the equalization and cancellation theory of binaural masking level difference. *J. Acoust. Soc. Am.*, 1960.

[15] S. Albani et al. *Model of binaural localization resolving multiple sources and spatial ambiguities. In Psychoacoustics, speech and hearing aids.* World Scientific, 1996.

[16] S. Mohan et al. Localization of multiple acoustic sources with small arrays using a coherence test. *J. Acoust. Soc. Am.*, 2008.

[17] T. Rohdenburg et al. Objective assessment of self-steering binaural hearing aid microphone arrays. *33rd International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008.

[18] T. Wittkop et al. Speech processing for hearing aids: Noise reduction motivated by models of binaural interaction. *ACTA ACUSTICA united with ACUSTICA*, 1997.

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

[20] V. Hohmann. Modellbasierte signalverarbeitung in hörgeräten. *Fortschritte der Akustik - DAGA*, 2008.

[21] I. Holube and B. Kollmeier. Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated prediction model. *J. Acoust. Soc. Am.*, 1996.

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

[23] H. Levitt. Noise reduction in hearing aids: An overview. *Journal of Rehabilitation Research and Development*, 2001.

[24] T. Lotter and P. Vary. Dual-channel speech enhancement by superdirective beamforming. *EURASIP Journal on Applied Signal Processing*, 2006.

[25] Ivo Merks. *Binaural Application of Microphone Arrays for Improvement Speech Intelligibility in a noisy Environment.* PhD thesis, Technische Universiteit Delft, 2000.

[26] B.C.J. Moore. *An Introduction to the Psychology of Hearing.* Academic Press, 2003.

[27] Kalle J. Palomäki, Guy J. Brown, and DeLiang Wang. A binaural processor for missing data speech recognition in the presence of noise and samll-room reverberation. *Speech Communication*, 2004.

[28] J. Peissig. *Binaurale Hörgerätestrategien in komplexen Störschallsituationen.* PhD thesis, Georg-August-Universtät zu Göttingen, 1992.

[29] Reinier Plomp. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J. Acoust. Soc. Am.*, 1978.

[30] A. Schlesinger and M. M. Boone. Evolutionary optimization for hearing aids of computational auditory scene analysis. *AES 126th Convention, München*, 2009.

[31] Wim Soede. *Improvement of Speech Intelligibility in Noise.* PhD thesis, Technische Universiteit Delft, 1990.

[32] S.J. van Wijngaarden and R. Drullman. Binaural intelligiblity prediction based on the speech transmission index. *J. Acoust. Soc. Am.*, 2008.