Performance comparison of binaural beamforming and MWF-based noise reduction algorithms for hearing aids

Daniel Marquardt[1], Volker Hohmann[2], Simon Doclo[1]

[1] *Signalverarbeitung - Universität Oldenburg*, [2] *Medizinische Physik - Universität Oldenburg*

Email: {daniel.marquardt, volker.hohmann, simon.doclo}@uni-oldenburg.de

Introduction

Noise reduction algorithms in hearing aids are crucial to improve speech understanding in background noise for hearing-impaired persons. For binaural hearing aids multi-microphone algorithms, exploiting signals from both the left and the right hearing aid, are considered to be promising techniques for noise reduction, because in addition to spectral information spatial sound information can be exploited. The performance of these algorithms however highly depends on prior assumptions about the acoustic environment and/or estimates of the signal statistics. In this paper the performance of 3 algorithms for a simulated everyday situation has been investigated using objective and subjective performance measures.

Configuration and Notation

Considering a microphone array consisting of $2N$ microphones, the frequency-domain representation of the n-th microphone signal Y_n is:

$$Y_n(\omega) = X_n(\omega) + V_n(\omega), \quad n = 1 \dots 2N, \tag{1}$$

with $X_n(\omega)$ representing the speech component and $V_n(\omega)$ representing the noise component. For conciseness the frequency variable ω will be ommitted from now on. Each hearing aid processes the $2N$-dimensional signal vector consisting of all microphone signals $\mathbf{Y} = \mathbf{X} + \mathbf{V}$. The correlation matrices are defined as:

$$\mathbf{R}_y = \mathcal{E}\left\{\mathbf{Y}\mathbf{Y}^H\right\}, \mathbf{R}_v = \mathcal{E}\left\{\mathbf{V}\mathbf{V}^H\right\}, \mathbf{R}_x = \mathbf{R}_y - \mathbf{R}_v. \tag{2}$$

In addition, we define the $2N$-dimensional vector $\mathbf{e}$ with one element equal to one and all other elements equal to zero, identifying the reference speech signal of the hearing aid (e.g. first microphone). In binaural algorithms two ouput signals are produced i.e. the objective is to design $2N$-dimensional filters $\mathbf{W}$ for the left and the right hearing aid.

Noise Reduction Algorithms

In the following section a brief overview of the 3 considered binaural noise reduction algorithms is given.

Beamformer with Binaural Postfilter (BPF) [1]
The monaural output of an MVDR (minimum variance distortionless response) beamformer $Y' = \mathbf{W}_b^H\mathbf{Y}$ is used to generate a real-valued postfilter gain G applied to the left and right hearing aid reference signal Y_l and Y_r i.e.

$$\mathbf{W}_b = \frac{\mathbf{\Gamma}^{-1}\mathbf{d}}{\mathbf{d}^H\mathbf{\Gamma}^{-1}\mathbf{d}}, \qquad G = \frac{\left(|d_l|^2 + |d_r|^2\right)|Y'|^2}{|Y_l|^2 + |Y_r|^2}, \tag{3}$$

where $\mathbf{\Gamma}$ represents the noise coherence matrix and $\mathbf{d}$ the propagation vector between the speech source S and the microphones.

Speech Distortion Weighted Multichannel Wiener Filter (MWF) [2]
The MWF minimizes the mean square error between an unknown reference (e.g. speech component in microphone signal) and a filtered version of the microphone signal, resulting in

$$\mathbf{W}_{mwf} = \left(\mathbf{R}_x + \mu\mathbf{R}_v\right)^{-1}\mathbf{R}_x\mathbf{e}, \tag{4}$$

where the parameter μ allows a trade-off between speech distortion and noise reduction.

Spatial Prediction (SP)
Based on an approach in [3], a frequency-domain approach exploiting spatial correlation of the speech signals has been presented in [4]. By first estimating the spatial prediction vector $\mathbf{h}$, minimizing the residual noise energy and imposing speech distortion to be zero, the spatial prediction filter is equal to

$$\mathbf{W}_{sp} = \frac{\mathbf{R}_v^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{R}_v^{-1}\mathbf{h}}, \qquad \mathbf{h} = \frac{\mathbf{R}_x\mathbf{e}}{\mathbf{e}^T\mathbf{R}_x\mathbf{e}}. \tag{5}$$

Test Procedure

For evaluating the performance of the algorithms, Behind-The-Ear Head-Related Impulse Responses (BTE-IR) from [5] have been used to generate speech signals from a fixed postition (placed -25° from receiver look direction) mixed with a continuous babble noise stream, with two seconds of noise between subsequent speech segments. The files were processed using an overlap-add framework at $f_s = 16\,\text{kHz}$, using 16 ms block size and 75% overlap between successive blocks.

In **MWF** and **SP** a voice activity detector (VAD) is required, classifying frames as noise or speech + noise. We assumed a perfect VAD and implemented a *batch* and *adaptive* strategy for estimating smoothed versions of the required correlation matrices.

1. **batch**: The correlation matrices are estimated offline using all available noisy speech vectors $\mathbf{Y}$ and corresponding noise vectors $\mathbf{V}$, which is a unrealistic assumption, however promising good results.

2. **adaptive**: The correlation matrices are adaptively smoothed based on the VAD classification. The filter $\mathbf{W}$ is adapted during noise-only periods, which is a realistic processing strategy provided that a perfect VAD is available.

BPF relies on prior assumptions about the acoustic environment. Two methods have been used:

1. $\mathbf{BPF_{opt}}$: The propagation vector $\mathbf{d}$ is computed using measured BTE-IRs, and anechoic BTE-IRs measured using the same hearing aid are used to compute $\mathbf{\Gamma}$.
2. $\mathbf{BPF_{sim}}$: A head model [6] is used for computing $\mathbf{d}$ and $\mathbf{\Gamma}$, assuming the position of the speech source is exactly known.

Results

The performance evaluation was based on two objective measures and one subjective hearing test.

Objective: For evaluating the noise reduction performance, the intelligibility weighted SNR improvement (ΔiSNR) for different input SNRs (measured at the left reference microphone) was calculated. In addition Perceptual Similarity Measure (PSM) from PEMO-Q [7] was used to determine the signal distortion introduced by the algorithms. For the reference signal in PSM noise was added to the clean speech signal with the same SNR as the filtered output signal. As shown in Figure 1, the **batch** processing leads to the best ΔiSNR, where **SP** outperforms **MWF**. The differences in ΔiSNR between **batch** and **adaptive** processing reduce with decreasing SNR. The beamformer shows the lowest ΔiSNR, especially for $\mathbf{BPF_{sim}}$ due to erroneous estimation of $\mathbf{d}$ and $\mathbf{\Gamma}$. In terms of PSM the beamformer with binaural postfilter shows less signal distortion compared to **SP** and **MWF**, especially at low SNR. This might be explained by lower noise reduction in $\mathbf{BPF_{opt}}$ and $\mathbf{BPF_{sim}}$.

Subjective: For measuring the improvement in Speech Reception Threshold (ΔSRT) an adaptive test, namely the “Oldenburg Sentence Test,” has been conducted with 10 normal hearing listeners. As shown in Figure 2, similarly to the objective measures the **batch** processing shows the best results (up to 5.7 dB improvement), whereas **MWF** outperforms **SP** in **batch** and especially **adaptive** processing. This might be explained by annoying spatial artefacts in the **SP** output reported by some listeners, which are introduced by erroneous estimation of the spatial prediction vector. As expected from the objective measures **BPF** shows less SRT improvement than **MWF** and **SP** but surprisingly $\mathbf{BPF_{sim}}$ does not perform significantly worse than $\mathbf{BPF_{opt}}$.

Conclusion

We have shown that an adaptive version of the **MWF** and **SP** binaural noise reduction algorithms (assuming a perfect VAD) results in a significant SRT improvement even in a difficult listening situation. In comparison to the batch algorithms the adaptive procedure performs worse due to erroneous signal estimates. It has also been shown that a fixed MVDR-beamformer with binaural postfilter indeed enables improvement in speech intelligibility but suffers from limited noise reduction performance compared to the **MWF** and **SP** algorithms.

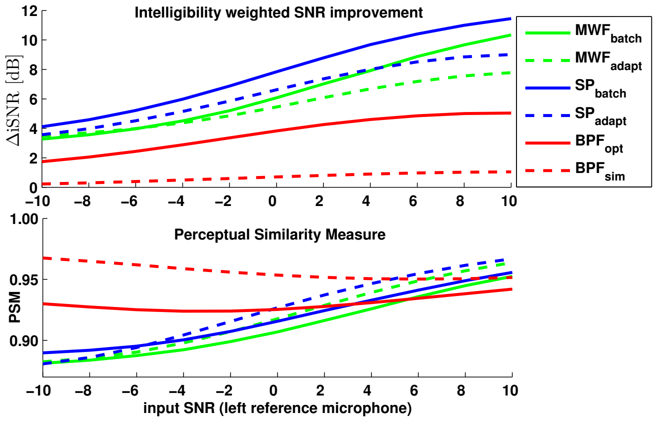

Figure 1: iSNR improvement and PSM for different input SNRs

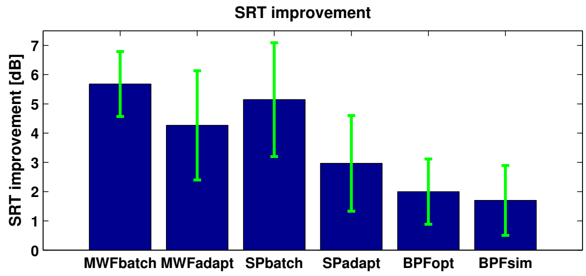

Figure 2: SRT improvement of the different algorithms

References

[1] Thomas Rohdenburg, “Development and objective perceptual quality assessment of monaural and binaural noise reduction schemes for hearing aids,” PhD thesis, Carl von Ossietzky Universität Oldenburg, Physics, Oldenburg, 2008.

[2] S. Doclo, A. Spriet, J. Wouters, M. Moonen, “Frequency-Domain criterion for the Speech Distortion Weighted Multichannel Wiener Filter for robust noise reduction,” Speech Communication, special issue on Speech Enhancement, vol. 49, no. 7-8, pp. 636-656, Jul.-Aug. 2007.

[3] J. Chen, J. Benesty, and Y. Huang, “A minimum distortion noise reduction algorithm with multiple microphones,” IEEE Trans. Audio, Speech, and Language Processing, vol. 16, no. 3, pp. 481–493, March 2008.

[4] B. Cornelis, M. Moonen, J. Wouters, “Comparison of frequency-domain noise reduction strategies based on Multichannel Wiener Filtering and Spatial Prediction,” in Proc. ICASSP, Taipei, Taiwan, Apr. 2009, pp. 129-132.

[5] H. Kayser *et al.*, “Database of multichannel In-Ear and Behind-The-Ear Head-Related and Binaural Room Impulse Responses,” Eurasip Journal on Advances in Signal Processing, vol. 2009, pp. 10 pages, 2009.

[6] P. C. Brown and R. O. Duda, “A structural model for binaural sound synthesis,” IEEE Trans. on Speech and Audio Processing, vol. 6, no. 5, pp. 476-488, Sep. 1998.

[7] R. Huber, B. Kollmeier, “PEMO-Q a new method for obejctive audio assessment using a model of auditory perception,” IEEE Trans. Audio, Speech and Language Processing, vol. 14, pp. 1902-1911, 2006.