A Phase Reference for a Multichannel Wiener Filter by a Delay and Sum Beamformer

Simon Grimm, Jürgen Freudenberger
Institute for System Dynamics, HTWG Konstanz, Germany, Email: {sgrimm, jfreuden}@htwg-konstanz.de

Abstract

The multichannel Wiener filter (MWF) is a well-established multichannel noise reduction technique. Most commonly, the MWF estimates the speech component in one of the microphone signals, which is called the reference microphone. The choice of the speech reference determines the broadband output signal-to-noise ratio (SNR). Recently, MWF approaches were proposed that combine the microphone signals to form a better speech reference, in order to improve the broadband output SNR. These techniques allow an arbitrary phase reference, because the phase of the estimated signal does not influence the output SNR. However, the phase of the reference affects the linear distortion of the speech signal. Besides noise, reverberation may degrade the speech quality. In order to improve the direct-to-reverberant-ratio (DRR), we propose a phase reference for the MWF, which is the phase of a delay-and-sum beamformer. This approach requires a time-difference-of-arrival (TDOA) estimate to align the signals properly. The phase of the aligned signals is used only as a phase reference for the MWF. This approach does not influence the output SNR, but reduces reverberation.

Introduction

In applications where hands-free communication systems are used, the environmental conditions have a significant impact on the speech quality. One reason is the linear distortion of the speech signal and reverberation, caused by room acoustics, which reduces the DRR. Due to background noise the SNR is decreased and as a result, the signal quality is further degraded. To achieve an improvement in speech communication, the influence of the acoustic transfer functions (ATF) from a speaker source to the microphones and also background noise should be taken into account.

The use of multiple microphones allows to apply beamforming techniques for speech enhancement in reverberant and noisy environments. Several approaches that use a reference channel were presented, which contain the relative transfer function - generalized sidelobe canceler (TF-GSC) [4], the minimum variance distortion-less response (MVDR) beamformer [5] and the speech distortion weighted - multichannel Wiener filter (SDW-MWF) [6]. Due to the selection of an arbitrary reference channel, the ATF from a speech source to the chosen reference microphone remains as the overall transfer function. This has an impact on the broadband output SNR of the beamformer [7]. In [1] the envelope of the individual transfer functions with an arbitrary phase reference was chosen as the overall transfer function to achieve a partial equalization of the acoustic system, which results in an improvement on the broadband output SNR. However, the reverberation caused by the acoustic environment is not reduced with this approach. In [9] it is shown, that the reverberation relies on the all-pass component of an ATF. This implies, that a delay-and-sum beamformer, as proposed in [8], can lead to a better DRR.

In this paper, it is shown that by choosing a suitable phase reference for the overall transfer function, the DRR can be improved. Therefore the phase of a delay-and-sum beamformer is used as a phase reference instead of the phase of a reference channel. The aim of this approach is a reduction of the reverberation, which leads to an improvement in speech quality.

Signal Model and Problem formulation

In this section, the signal model and the corresponding notation is presented. We consider the acoustic system as time-invariant and linear. The beamformer array consists of M microphones. The i^{th} microphone signal $y_i(k)$ can be expressed as the convolution of the source speech signal $x(k)$ with the acoustic impulse response $h_i(k)$ from the speech source to the i^{th} microphone plus an additive noise term $n_i(k)$. The resulting microphone signals can be written as following in the short time frequency domain

$$Y_i(\kappa,\nu) = H_i(\nu)X(\kappa,\nu) + N_i(\kappa,\nu) \tag{1}$$

where $Y_i(\kappa,\nu)$, $X(\kappa,\nu)$ and $N_i(\kappa,\nu)$ correspond to the short time spectra of the signals, $H_i(\nu)$ represents the ATF of the acoustic impulse response and $S_i(\kappa,\nu) = H_i(\nu)X(\kappa,\nu)$ the speech signal component at the i^{th} microphone. κ and ν correspond to the subsampled time index and the frequency bin index respectively. In the following these are omitted. The speech signals and the ATFs can be written as M-dimensional vectors:

$$\boldsymbol{S} = [S_1\ S_2 \cdots S_M]^T \tag{2}$$
$$\boldsymbol{N} = [N_1\ N_2 \cdots N_M]^T \tag{3}$$
$$\boldsymbol{H} = [H_1\ H_2 \cdots H_M]^T \tag{4}$$
$$\boldsymbol{Y} = [Y_1\ Y_2 \cdots Y_M]^T \tag{5}$$
$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{N} \tag{6}$$

^{T} denotes the transpose of the vector, * the complex conjugate and † corresponds to the conjugate transpose. It is further assumed that the speech source and the noise

signals are zero-mean random processes with the variances σ_{Ni}^2 and σ_X^2 and therefore the correlation matrix of the speech signal component can be expressed as:

$$\boldsymbol{R}_S = \mathbb{E}\{\boldsymbol{S}\boldsymbol{S}^\dagger\} = \sigma_X^2 \boldsymbol{H}\boldsymbol{H}^\dagger \tag{7}$$

According to [1] the MWF can be decomposed using the matrix inversion lemma:

$$G^{MWF} = \frac{\sigma_X^2}{\sigma_X^2 + \mu_W (\boldsymbol{H}^\dagger \boldsymbol{R_N^{-1}} \boldsymbol{H})^{-1}} \frac{\boldsymbol{R_N^{-1}} \boldsymbol{H}}{\boldsymbol{H}^\dagger \boldsymbol{R_N^{-1}} \boldsymbol{H}} \tilde{H}^* \tag{8}$$

$$= G^{WF}\ G^{MVDR}\ \tilde{H}^* \tag{9}$$

μ_W is a tradeoff parameter, R_N^{-1} the inverse of the noise correlation matrix and $\tilde{H}^*$ the complex conjugate of the overall transfer function from the speech source to the output of the beamformer. Apparently, the parametric MWF can be decomposed into an MVDR beamformer G^{MVDR}, a filter that is equal to the overall transfer function $\tilde{H}$, and a single-channel Wiener post filter G^{WF}. In [1] $\tilde{H}$ is chosen as:

$$\tilde{H} = \sqrt{\boldsymbol{H}^\dagger \boldsymbol{H}}\ e^{j\phi_{ref}} \tag{10}$$

The phase reference ϕ_{ref} could be selected arbitrary. For the aim of SRR improvement, the phase-term of a delay-and-sum beamformer ϕ_{DS} must be procured. This beamformer design requires a time-difference-of-arrival (TDOA) estimate, which is a challenging task in noisy and reverberant environments and could be achieved by the generalized cross-correlation method [2] or several other methods as proposed in [3]. The phase of the delay-and-sum beamformer is acquired as shown in Figure 1. The TDOA estimate allows the computation of a linear all-pass phase-term H_{align}, which is multiplied to one of the microphone signals in the frequency domain to achieve a coherent signal aligning of the direct paths. After the summation of the aligned signals, the phase reference ϕ_{DS} is obtained.

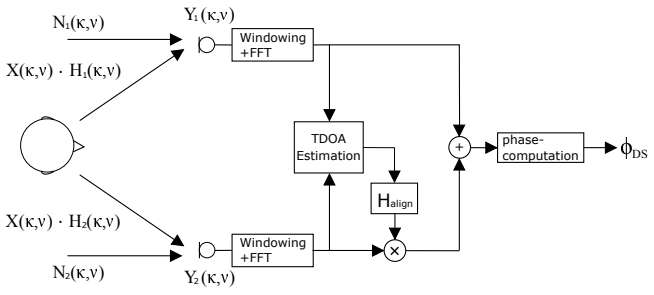

Figure 1: Structure of the delay-and-sum beamforming and the resulting phase-acquirement

If the phase reference in Eq.(10) is set to

$$\phi_{ref} = \phi_{DS} \tag{11}$$

a coherent phase combining of the direct path signals is achieved, which results in less reverberation, because the direct path is enhanced and incoherent reflections are attenuated.

Simulation Results

In order to verify the proposed approach, we consider a system without the noise reduction Wiener post filter, i.e. we select $\mu_W = 0$ in Eq.(8). Consequently, the overall transfer function (from the speakers mouth to the output of the system) is determined by the choice of the reference channel, because G^{MVDR} has a unity gain transfer function. Note that for $\mu_W > 0$ the Wiener post filter alters the overall transfer function. However, these changes are independent of the phase reference.

Firstly, we compare the Schroeder plots [10] corresponding to the resulting ATFs of an exemplary in-car scenario. For this scenario we consider two cardioid microphones that were mounted close to the rear-view mirror in a car (8 cm microphone distance). Impulse responses were measured with an artificial head.

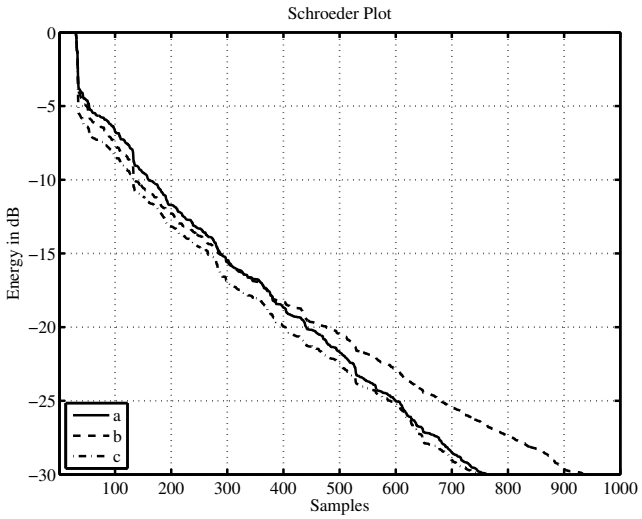

Figure 2: Schroeder plot of the resulting acoustic transfer functions of the in-car environment: (a) AIR from the speech signal source to microphone 1, (b) overall transfer function as chosen in Eq.(10) with phase reference of microphone 1, (c) overall transfer function as chosen in Eq.(10) but with the phase reference ϕ_{DS} of the delay-and-sum beamformer

The Schroeder plots of the resulting overall transfer functions are shown in Figure 2. Curve (a) depicts the Schroeder plot of the ATF from the speaker source to the first microphone. Curve (b) shows the resulting Schroeder plot for the overall transfer function of Eq.(10) with the phase reference of microphone 1. It could be observed, that the decay time is increased, as a result of the decoupling between the phase and magnitude of the first microphone channel. However, the energy of the first reflections is reduced due to the partial equalization of the acoustic channel as could be seen from the first 300 samples of the Schroeder plot. Curve (c) shows the Schroeder plot for the same overall transfer functions as seen in (b), but with the phase ϕ_{DS} as the selected phase reference.

Compared with (b), a reduced decay time is observed due to the coherent combining of the phase terms. The overall reverberation time is even shorter, compared with the decay time of the ATF of (a). As a result, the direct components of the ATFs are enhanced, which leads to an improvement in speech quality of the overall system.

In Figure 3 the Schroeder plot of the ATFs of a classroom scenario as measured in [11] are shown. The impulse responses were recorded with omnidirectional microphones at two different spatial locations with a microphone distance of 0.5m. The reverberation time RT_{60} has a value between 1.5 and 1.8 seconds over all frequencies. Due to the longer reverberation time, as compared with the in-car scenario, the resulting Schroeder plots show a different behaviour. Curve (d) shows the resulting transfer function when one of the microphone channels is selected as $\tilde{H}$. Curve (e) and (f) show the overall transfer function of Eq.(10) where (e) has the phase reference of microphone 1 and (f) has the phase reference of a delay-and-sum beamformer. It can be observed in (f), that the direct signal component for the first few samples is augmented, due to the coherent aligning of the signals in the phase reference. But compared to (d) the decay time is increased. While (f) still shows a better performance than (e), because of the delay-and-sum phase reference, the decay time is still slightly increased compared to (d). This is again caused by the phase and magnitude decoupling of the filter design.

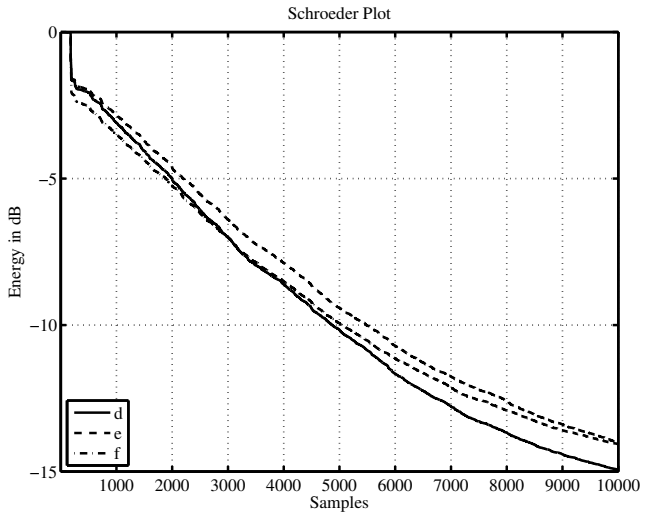

Figure 3: Schroeder plot of the resulting acoustic transfer functions of the classroom environment: (d) AIR from the speech signal source to microphone 1, (e) overall transfer function as chosen in Eq.(10) with phase reference of microphone 1, (f) overall transfer function as chosen in Eq.(10) but with the phase reference ϕ_{DS} of the delay-and-sum beamformer

In Table 1 the direct-to-reverberant ratio calculations after [12] for the two acoustic scenarios are presented. For the direct path, the first eight milliseconds of the ATFs are taken into account. For the in-car scenario it could be observed that the overall transfer function, as proposed in Eq.(10), increases the DRR. By selecting ϕ_{DS} as the phase reference, the DRR is increased further. For the classroom scenario however, the DRR is decreased compared to microphone channel 1 as the overall transfer function selection, when the envelope of the ATFs and the phase reference of microphone 1 is chosen. This could be improved by taking the phase reference of the delay-and-sum beamformer. As a result, the DRR is slightly higher than the DRR of the ATF of microphone channel 1.

Table 1: DRR of the overall transfer function for choosing a different phase and magnitude difference

ATFs	selected reference $\tilde{H}^*$		
	H_1	$\sqrt{\boldsymbol{H}^\dagger \boldsymbol{H}} e^{j\phi_{H_1}}$	$\sqrt{\boldsymbol{H}^\dagger \boldsymbol{H}} e^{j\phi_{DS}}$
in-car scenario	9.52 dB	10.51 dB	11.18 dB
classroom scenario	-2.45 dB	-2.71 dB	-1.33 dB

Conclusions

In this paper, we proposed a new approach for the phase reference selection of the multichannel Wiener filter with partial equalization. By comparing the Schroeder plots of the proposed overall transfer function with different phase references, it could be seen that the decay time is shortened if the phase of a delay-and-sum beamformer is used. Also the DRR is increased. In the case of the in-car scenario, the decay time of the overall transfer function is even shorter as the ATF from the signal source to one of the microphones. For the classroom scenario the decay time is increased due to the decoupling of the phase and magnitude components, if the phase reference of one microphone channel is used. Hence, the proposed delay-and-sum phase reference approach reduces the reverberation compared with the phase reference from a microphone. The effect of $\mu_W \neq 0$ in Eq.(8) on the overall transfer function is neglected in this paper and therefore a topic for further investigations.

References

[1] S. Stenzel, T. C. Lawin-Ore, J. Freudenberger, S. Doclo: A Multichannel Wiener Filter With Partial Equalization For Distributed Microphones. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2013), 1-4

[2] C. H. Knapp and G. C. Carter: The Generalized Correlation Method for Estimation of Time Delay. IEEE Transactions On Acoustics, Speech And Signal Processing (August 1976), 320-327

[3] J. Chen, J. Benesty and Y. Huang: Time Delay Estimation in Room Acoustic Environments: An Overview. EURASIP Journal on Applied Signal Processing (2006), 1-19

[4] S. Gannot, D. Burshtein and E. Weinstein: Signal enhancement using beamforming and nonstationarity with applications to speech. IEEE Transactions On Signal Processing (2001), 1614-1626

[5] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot and J. Dmochowski: New Insights Into The MVDR Beamformer in Room Acoustics. IEEE Transactions On Acoustics, Speech And Language Processing (2010), 158-170

[6] S. Doclo, A. Spriet, J. Wouters and M. Moonen: Frequency-Domain Criterion For The Speech Distortion Weighted Multichannel Wiener Filter For Robust Noise Reduction. Speech Communication (July 2007), 636-656

[7] T. C. Lawin-Ore and S. Doclo: Reference Microphone Selection for MWF-based Noise Reduction Using Distributed Microphone Arrays. Proceedings of 10. ITG Symposium on Speech Communication (September 2012), 31-34

[8] J. B. Allen, D. A. Berkley and J. Blauert : Multimicrophone signal processing technique to remove room reverberation from speech signal. J. Acoust. Soc. Am. (1977), 912-915

[9] Q.- G. Liu, B. Champagne and P. Kabal: Room speech dereverberation via minimum-phase and all-pass component processing of multi-microphone Signals. IEEE Pacific Rim Conference on Communications, Computers and signal Processing (May 1995), 571-574

[10] M. R. Schroeder: New Method of Measuring Reverberation Time. J. Acoust. Soc. Am. 37 (1965), 409-412

[11] R. Stewart and M. Sandler: Database of Omnidirectional and B-Format Impulse Responses. IEEE International Conference on Acoustics, Speech and Signal Processing (March 2010), 165-168

[12] Patrick A. Naylor and Nikolay D. Gaubitch: Speech Dereverberation. Springer Verlag, London, 2010