A Directivity Based Reference for the Multichannel Wiener Filter

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 noise reduction technique. Recently, it was shown that the broadband output signal-to-noise ratio (SNR) can be improved with the generalized MWF (G-MWF) which uses a linear combination of the microphone channels as a speech reference. In this work, directivity based references for the G-MWF are introduced which are suitable for closely spaced microphone arrangements, e.g. for hearing aids. The proposed references are based on differential microphone processing approaches using dipole, cardioid, and hypercardioid beam patterns. Simulation results are presented for a monaural hearing aid in a cylindrical isotropic noise field and a cafeteria environment. These results show an improved noise suppression performance compared with the standard MWF.

Introduction

Research on speech enhancement using multiple microphones has gained significant interest in the past few years, since more than one microphone allows to acquire information about the spatial sound field [1, 2, 3, 4, 5]. One of those signal enhancement methods is the speech distortion weighted - multichannel Wiener filter (SDW-MWF), which produces a minimum-mean-squared error (MMSE) estimate of an unknown desired signal [6, 7, 8]. The desired signal of the SDW-MWF is usually the speech component in one of the microphone signals, referred to as the reference microphone signal.

However, for spatially distributed microphones the selection of the reference microphone may have a large influence on the performance of the MWF since it depends on the positions of the speech and noise sources as well as the microphones [9, 2, 4, 3]. The generalized MWF (G-MWF) was proposed in order to improve the broadband output SNR [4]. With the G-MWF, the speech reference is not a single microphone channel as in the standard multichannel Wiener filter. Instead a linear combination of the channels is used to design the overall transfer function. This reference choice is still optimal regarding the narrow-band SNR [1], but can potentially improve the broad-band output SNR [9].

New G-MWF reference designs were proposed recently. In [2], the overall transfer function was designed as the envelope of the individual transfer functions to achieve a partial equalization of the acoustic system. This approach has the advantage requires only on the second order statistics of the speech and noise terms in the microphone signals. In [10, 11], the time-difference-of-arrival (TDOA) was taken into account using a delay-and-sum beamformer reference. It was shown for spatially distributed microphone setups that this reference design results in an improved SNR as well as a reduced reverberation compared with the MWF with a signal microphone reference.

For closely spaced microphones, the spatial diversity cannot be exploited due to the small microphone distances. However, differential beamforming is often used for this microphone arrangements, because it achieves a spatial filtering [12]. In this work, directivity based references for the G-MWF are examined regarding their noise reduction capabilities. These references are based on differential beamforming. Besides the second order statistics, only knowledge of the TDOA between the microphones is required for these reference designs.

In the following sections, the signal model and the notation is described. Furthermore, the speech distortion weighted - multichannel Wiener filter and its generalization is introduced. The new reference designs based on differential beamforming are proposed afterwards. This is followed by the simulation results.

Signal Model and Notation

In this section, the signal model for the multichannel Wiener filter is briefly introduced. The i^{th} microphone signal $y_i(k)$ can be described by the convolution of the speech signal $x(k)$ with the acoustic impulse response $h_i(k)$ from the mouth-reference-point (MRP) to the i^{th} microphone plus an additional noise term $n_i(k)$. The acoustic is considered as linear and time-invariant. The resulting signal can be written in the short time frequency domain as

$$Y_i(\eta,\nu) = S_i(\eta,\nu) + N_i(\eta,\nu) \tag{1}$$
$$S_i(\eta,\nu) = X(\eta,\nu)H_i(\nu) \tag{2}$$

where $Y_i(\eta,\nu)$, $S_i(\eta,\nu)$ and $N_i(\eta,\nu)$ correspond to the short time spectra of the time domain signals and $H_i(\nu)$ denotes the acoustic transfer function. η and ν denote the subsampled time index and the frequency bin index, respectively. In the following these indices are often omitted for brevity. The signals as well as the acoustic transfer functions can be written as M-dimensional vectors

$$\mathbf{Y} = [Y_1, Y_2, \dots, Y_M]^T \tag{3}$$
$$\mathbf{S} = [S_1, S_2, \dots, S_M]^T \tag{4}$$
$$\mathbf{N} = [N_1, N_1, \dots, N_M]^T \tag{5}$$
$$\mathbf{H} = [H_1, H_2, \dots, H_M]^T \tag{6}$$
$$\mathbf{Y} = \mathbf{S} + \mathbf{N} \tag{7}$$

^{T} denotes the transpose of a vector, * the complex conjugate and † the conjugate transpose. Vectors and matrices are written in bold and scalars are normal letters.

The speech and noise signals are assumed to be zero-mean random processes with the power spectral densities (PSDs) $\Phi_{S_i}^2$ and $\Phi_{N_i}^2$. Assuming a single speech source, the speech correlation matrix $\boldsymbol{R}_S$ has rank one and can be expressed as

$$\boldsymbol{R}_S = \mathbb{E}\left\{\mathbf{S}\mathbf{S}^\dagger\right\} = \Phi_X^2 \mathbf{H}\mathbf{H}^\dagger \tag{8}$$

where $\mathbb{E}\{\}$ denotes the mathematical expectation and Φ_X^2 the PSD of the clean speech signal at the MRP. Similarly, $\boldsymbol{R}_N = \mathbb{E}\left\{\mathbf{N}\mathbf{N}^\dagger\right\}$ denotes the noise correlation matrix. Furthermore, it is assumed that the speech and noise terms are uncorrelated.

The Generalization of the MWF

The SDW-MWF aims to estimate the speech signal of a chosen microphone channel. It is commonly implemented as

$$\mathbf{G}^{\mathbf{MWF}} = (\boldsymbol{R}_S + \mu \boldsymbol{R}_N)^{-1} \boldsymbol{R}_S \mathbf{u}\,, \tag{9}$$

where μ is an overestimation parameter that sets a trade-off between noise reduction and speech distortion and $\mathbf{u}$ is a reference vector. The filtered output signal Z is obtained by

$$Z = \mathbf{G}^{\mathbf{MWF}^\dagger} \mathbf{Y} \tag{10}$$

In case of the SDW-MWF, $\mathbf{u}$ is a vector that selects the reference microphone, i.e., the vector $\mathbf{u}$ contains a single one and all other elements are zero

$$\mathbf{u} = [0, \ldots, 1, \ldots, 0]^T\,. \tag{11}$$

The MWF realization using (11) is referred to as the S-MWF in the following. For the S-MWF, the overall transfer function $\tilde{H}_d$ is equal to the ATF of a reference microphone, i.e. $\tilde{H}_d = H_{ref}$. Since $\boldsymbol{R}_S$ is a rank one matrix, it should be noted that any non-zero vector $\mathbf{u}$ achieves the same (optimal) narrow-band output SNR [1]. This implies that the vector $\mathbf{u}$ can be chosen in an arbitrary way to influence the overall transfer function $\tilde{H}_d$. In [4], the generalized MWF was presented, which allows to define a speech reference for the MWF by the elements u_i of the vector $\mathbf{u}$. The vector $\mathbf{u}$ can be used to define the desired overall transfer function $\tilde{H}_d$ as

$$\tilde{H}_d = \mathbf{u}^\dagger \mathbf{H} = \sum_i {u_i}^* \cdot H_i \text{ for } u_i \in \mathbb{C} \tag{12}$$

By using the decomposition of the MWF [13], it can be seen that $\tilde{H}_d$ is the resulting overall transfer function

$$\mathbf{G}^{\mathbf{MWF}} = \frac{\Phi_X^2}{\Phi_X^2 + \mu(\mathbf{H}^\dagger \boldsymbol{R}_N^{-1}\mathbf{H})^{-1}} \frac{\boldsymbol{R}_N^{-1}\mathbf{H}}{\mathbf{H}^\dagger \boldsymbol{R}_N^{-1}\mathbf{H}} \tilde{H}_d^* \tag{13}$$

$$= G^{WF}\, \mathbf{G}^{\mathbf{MVDR}}\, \tilde{H}_d^*\,, \tag{14}$$

for $\mu = 0$, because $\mathbf{G}^{\mathbf{MVDR}}$ has a unity gain transfer characteristic. In this case, the speech output signal Z_S of the processing algorithm can be written as

$$Z_S = \tilde{H}_d \cdot X\,. \tag{15}$$

Directivity Based References for the G-MWF

In the following, the new directional references are presented. We consider closely spaced microphones in an endfire array configuration pointing towards the desired speech source. The references only differ in a delay τ to form the required beam pattern. In this work, an endfire array consisting of two microphones is considered for the differential beamforming references. The corresponding reference vector $\mathbf{u}$ can be written as

$$\mathbf{u} = G^{EQ^*}[1, -e^{j2\pi\nu\tau}]^T \tag{16}$$

where G^{EQ} denotes a compensation filter for the first order high pass behavior created by the differential output. The directional references create a spatial null for signals arriving from an incident angle $\theta \in [90°, 180°]$, where the incident angle from the front of the endfire array is $\theta = 0°$. The spatial null is dependent on a delay τ, which can vary within $0 \leq \tau \leq \tau_0$. τ_0 is the time required for the speech signal to travel between the microphones. The beam patterns corresponding to a given delay time are shown in Table 1.

Table 1: Selected directional beam patterns for τ

Beam pattern	Delay τ
Dipole	0
Cardioid	τ_0
Hypercardioid	$\frac{1}{3}\tau_0$

The equalization filter G^{EQ} in (16) is required to compensate the high pass behavior of the differential microphone array output. In this work, it is derived in anechoic and noise free conditions in the MMSE sense by minimizing the error between the uncompensated output of the differential array and a selected reference speech signal

$$G^{EQ} = \underset{G^{EQ}}{\operatorname{argmin}}\ \mathbb{E}\left\{|([1, -e^{-j2\pi\nu\tau}]\mathbf{S})G^{EQ} - S_{ref}|^2\right\}\,. \tag{17}$$

Simulation Results

In order to investigate the proposed differential beamforming references for the G-MWF, a simulation scenario with a monaural hearing aid consisting of two closely spaced microphones mounted on an artificial head is examined. An anechoic environment with an cylindrically isotropic noise field as well as a more realistic situation in a cafeteria are considered for the simulations. The datasets to create the simulations are obtained from the database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses in [14]. The sampling rate for all simulations is 48 kHz and an FFT length of $L = 512$ was used with a blockshift of 128 samples. For the overlapp-add implementation, a Hamming window was used. The spacing between the microphones of the hearing aid is 7.6 mm. For the simulation scenarios, the power spectral densities, the cross power spectral densities as well as the TDOA for the microphones are assumed to be known. The values of $\boldsymbol{R}_S$ and $\boldsymbol{R}_N$ are calculated for the whole dataset and the G-MWF is applied as a batch job. The overestimation parameter is set to $\mu = 1$ for all simulations.

Segmental SNR

In the following, the segmental SNR (SSNR) of the proposed references is compared with the results for the S-MWF. The SSNR is calculated for frequencies between 400 Hz and 4kHz by band-limiting the signals. The obtained results for the cylindrically isotropic noise field in the anechoic environment, as well as the cafeteria environment with real background noise are presented in Table 2. It should be noted that the SSNR gain is quite small in general due to the batch job processing, i.e. the filters are stationary.

Table 2: Segmental SNR comparison for different G-MWF references (speech source angle: $\theta = 0°$)

Reference	Anechoic environment	Cafeteria environment
Y_1	5.6 dB	4.1 dB
S-MWF	7.7 dB	5.3 dB
Dipole reference	8.5 dB	6.5 dB
Cardioid reference	8.6 dB	6.6 dB
Hypercardioid reference	8.8 dB	7.0 dB

As can be observed, the S-MWF which uses a reference channel as the overall transfer function is capable to improve the SSNR by over 2 dB for the anechoic environment compared with a single microphone of the hearing aid (Y_1). The differential references show a significant SSNR improvement for both simulation scenarios compared with the S-MWF, where the hypercardiod reference selection achieves the best performance regarding the SSNR.

Incident Angle Dependent SNR

Now the suppression of directional noise is examined in the anechoic simulation environment. Therefore the direction of a single noise source of white Gaussian noise is varied for incident angles θ from 0° to $-180°$. The speech source stays at a fixed position with $\theta = 0°$. The performance of the differential beamforming references is compared with the S-MWF that uses a single reference channel. The results are depicted in Figure 1.

Compared with the S-MWF that uses the reference microphone, the dipole, the cardioid as well as the hypercardioid show an improved segmental SNR for nearly all incident angles expect for incident angles between $-150°$ and $-180°$, where the dipole reference shows a slightly inferior performance. The SSNR has its maximum value close to the incident angles where the differential beamformer is supposed to suppress the noise source best. These angles are $-90°$ for the dipole, $-180°$ for the cardioid, and $-110°$ for the hypercardioid beam pattern. As can be observed, these theoretical values agree with the simulation results.

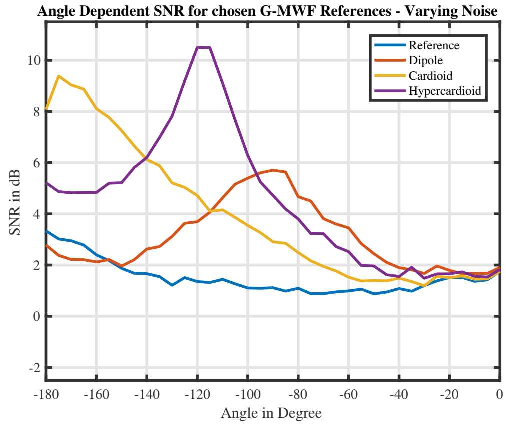

Figure 1: Segmental SNR for an angle varying noise source and a fixed speech source (speech source angle: $\theta = 0°$)

Log Spectral Distance

The log spectral distance measures the linear distortion in comparison with the clean speech signal at the MRP. In Table 3 the results regarding the log spectral distance are presented. As can be observed, the values are slightly improved for the differential beamforming references in the anechoic scenario compared with the single microphone Y_1 as well as the S-MWF. In the cafeteria environment, the LSD for all references is similar to that of the S-MWF. This shows that the spectrum of the speech signal for the differential beamforming references is not distorted in comparison to the S-MWF despite the superior noise reduction.

Table 3: Log spectral distance comparison for different G-MWF references (speech source angle: $\theta = 0°$)

Reference	Anechoic environment	Cafeteria environment
Y_1	3.6 dB	3.9 dB
S-MWF	2.6 dB	4.5 dB
Dipole reference	2.0 dB	4.5 dB
Cardioid reference	2.2 dB	4.4 dB
Hypercardioid reference	2.1 dB	4.4 dB

Conclusions

In this paper, new directivity based references for the generalized multichannel Wiener filter were introduced. With the proposed differential beamforming references, the directional response of the G-MWF is implicitly designed. Simulations were performed with a monaural hearing aid consisting of two closely spaced microphones.

The results for an anechoic environment and a cafeteria show that the directional beam pattern references are superior to the S-MWF in terms of SNR improvement. The highest SSNR gains are obtained by the hypercardioid reference. The log-spectral distance is not increased for all differential beamforming references compared with the S-MWF despite the improved noise reduction capability.

The ability of the differential beamforming references to suppress directional noise is examined by varying the incident angle of a directional noise source. It was shown that the proposed references show a superior performance regarding the directional noise reduction compared with the S-MWF. The best SSNR values are achieved at noise incident angles where the differential beamforming references are supposed to create a spatial null for the dedicated directional beam pattern. These incident angles are close to the theoretical values.

References

[1] S. Stenzel and J. Freudenberger, "Blind matched filtering for speech enhancement with distributed microphones," *Journal of Electrical and Computer Engineering*, Aug 2012, Article ID 169853, 15 pages.

[2] S. Stenzel, T.C. Lawin-Ore, J. Freudenberger, and S. Doclo, "A multichannel Wiener filter with partial equalization for distributed microphones," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY*, 2013.

[3] T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Alternative formulation and robustness analysis of the multichannel Wiener filter for spatially distributed microphones," in *Proceedings of the International Workshop on Acoustic Signal Enhancement (IWAENC), Antibes, France*, Sep 2014, pp. 208–212.

[4] T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Generalized multichannel Wiener filter for spatially distributed microphones," in *Proc. ITG Conference on Speech Communication, Erlangen, Germany*, Sep 2014, pp. 1–4.

[5] T. C. Lawin-Ore and S. Doclo, "Analysis of the average performance of the multichannel Wiener filter based noise reduction using statistical room acoustics," *Signal Processing*, vol. 107, pp. 96–108, Feb 2015.

[6] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids," in *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2003, pp. 147–150.

[7] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multichannel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, 2004.

[8] S. Doclo, A. Spriet, M. Moonen, and J. Wouters, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*, chapter 9. Springer, Berlin, Heidelberg, New York, 2005.

[9] T. C. Lawin-Ore and S. Doclo, "Reference microphone selection for MWF-based noise reduction using distributed microphone arrays," in *Proceedings of 10. ITG Symposium on Speech Communication*, Sept. 2012, pp. 31–34.

[10] S. Grimm and J. Freudenberger, "A phase reference for a multichannel Wiener filter by a delay and sum beamformer," in *Jahrestagung für Akustik (DAGA), Nürnberg*, Mar 2015, pp. 208–212.

[11] S. Grimm, J. Freudenberger, T. C. Lawin-Ore, and S. Doclo, "Phase reference for the generalized multichannel Wiener filter," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1 – 17, Jul 2016.

[12] Jacob Benesty and Chen Jingdong, *Study and Design of Differential Microphone Arrays*, Springer, Berlin, Heidelberg, 2013.

[13] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., pp. 945–978. Springer Berlin Heidelberg, 2008.

[14] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "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, no. 1, pp. 298605, Jul 2009.