Microphone Diversity based Wind Noise Reduction in a Car Environment using MEMS Arrays

Simon Grimm[1], Jürgen Freudenberger[1], Harald Schnepp[2]
[1] Institute for System Dynamics, HTWG Konstanz, Germany, Email: {sgrimm, jfreuden}@htwg-konstanz.de
[2] Daimler AG Ulm, Enabling Technologies for Communication, Germany, Email: harald.schnepp@daimler.com

Abstract

This work studies wind noise reduction methods using a microphone array in a car environment. In particular, an endfire array with two Micro Electro Mechanical Systems (MEMS) microphones is considered as a substitute for an ordinary microphone capsule of the same size. Even for small distances of the microphones, the different signal properties at the microphones can be exploited to achieve a significant reduction of wind noise. Based on these signal properties, a combination of a beamformer and a single channel post filter is calculated to obtain an improved speech reference. The reference is used to calculate adaptive frequency domain least-mean-square (FLMS) filters which are applied to the microphone signals. The proposed approach results in an improved speech signal regarding the signal-to-noise-ratio (SNR) while keeping the linear speech distortion low.

Introduction

Hands-free communication applications in a car environment always face the problem of unwanted noise components in the microphone signals. The noise consists of car noise as well as wind noise components, caused by open windows, fans or open convertible hoods that create airflow turbulence over the microphone membranes and result in low frequency signal components of high amplitude [1]. Commonly single channel noise suppression algorithms like the Wiener filter and spectral subtraction are used for noise suppression [2]. Multichannel approaches improve the speech quality [3]. Yet noise reduction algorithms in car environments are typically based on the assumption that the noise is stationary or varies only slowly. In [4], Wilson et. al. demonstrated that wind noise consists of local short time disturbances which are highly non-stationary. This makes the reduction of wind noise a challenging task.

The suppression of wind noise is mostly covered in the context of digital hearing aids or mobile devices in the literature [5, 6, 7]. For single channel wind noise reduction, often the different PSD properties of speech and wind noise are exploited [5, 6, 8]. Several other methods exist that aim to reduce wind noise for a single microphone [9, 10, 11, 12, 13]. Using more than one microphone, the diversity of the sound field can be taken into account to indicate wind noise and reduce it successful. In [6], a spectral weighting filter based on the coherence between two microphones is proposed. The coherence is also used in [14], where additionally to the magnitude squared coherence (MSC) the information that relies on the phase component is applied to synthesize a spectral filter function. In [15], the decomposition of the multichannel Wiener filter (MWF) into a beamformer and a single channel post filter for an arbitrary microphone arrangement is presented. The approach is based on the assumption, that the wind noise is decorrelated at the microphones, while having equal power spectral densities (PSDs).

We propose a two-channel approach that exploits the diversity of two closely located MEMS microphones. The microphones are positioned as an endfire array in a car environment. Even for distances of a few centimeters, the variation in the microphone signals can be used to reduce wind noise significantly. The coherence properties of speech and wind noise signals are exploited to obtain an improved speech reference compared with a single microphone capsule. This reference signal is obtained by using a frequency bin selection approach in low frequencies, that is combined with a delay-and-sum beamformer for higher frequencies to reduce the speech distortion. Additionally, a wind noise reference is created by a delay-and-subtract combination of the signals that is able to block the speech energy almost completely in the frequency range where wind noise is present.

Based on these references, a single channel post filter for the speech reference is created to obtain an improved signal regarding the SNR. The acquired signal is used for an FLMS algorithm to calculate time-variant filters that estimate the coherent signal components between the speech reference and the microphone signals. By applying the obtained filters to the microphone input signals, the simulations show that the proposed approach is capable of improving the signal-to-noise-ratio while keeping the linear distortion of the speech signal low or even improve it compared with a single microphone capsule.

Signal Model and Notation

In the following, the signal model and the notation is briefly explained. The acoustic in the car environment is considered as linear and time-invariant. 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 to the i^{th} microphone signal plus an additional noise term $n_i(k)$. This noise term consist of wind noise as well as car noise.

The resulting signal can be written in the short time frequency domain as

$$Y_i(\kappa,\nu) = H_i(\nu)X(\kappa,\nu) + N_i(\kappa,\nu) \quad (1)$$

where $Y_i(\kappa,\nu)$, $X(\kappa,\nu)$ and $N_i(\kappa,\nu)$ correspond to the short time spectra of the signals, $H_i(\nu)$ denotes the acoustic transfer function, $H_i(\nu)X(\kappa,\nu)$ is the spectrum of the speech component and $N_i(\kappa,\nu)$ the spectrum of the noise in the i^{th} microphone. κ and ν denote the subsampled time index and the frequency bin index, respectively. In the following these indices are often omitted when possible.

Wind Noise Reduction Algorithm

In this section, the proposed noise reduction algorithm is presented. At first the beamformer is derived based on assumptions for the speech and wind noise signal properties. Then the wind noise reference is obtained, which is used by the post filter for a further SNR improvement of the speech reference. Finally, the FLMS filter calculation using the speech reference is shown.

Since the MEMS microphones are configured as an endfire array, the signals have to be aligned first to compensate the different times of arrival for the speech signal. This is achieved by delaying the front microphone with a suitable sample delay τ to be in phase with the rear microphone, e.g.

$$\hat{Y}_1(\kappa,\nu) = Y_1(\kappa,\nu) \cdot \begin{cases} e^{-j2\pi\frac{\nu}{N}\tau} & \text{for } \nu \in 0,\ldots,\frac{N}{2}-1 \\ e^{j2\pi\frac{\nu}{N}\tau} & \text{for } \nu \in \frac{N}{2},\ldots,N \end{cases} \quad (2)$$

where N denotes the block length of the short time Fourier transform

Speech Reference

We assume, the wind is decorrelated at the microphones and the short time wind noise PSDs are unequal over the frequency bins that contain wind noise. In contrast, we assume the speech signal to be highly correlated as well as the speech PSDs to be equal for the aligned signals in low frequencies due to the small distance of the MEMS microphones. Based on these assumptions, a combined approach for the speech reference is presented, that uses the selection of wind noise free frequency bins up to 1kHz and a delay-and-sum beamformer (DSB) for the higher frequencies.

Frequency bin selection (FBS)

The short time power spectral densities of the aligned microphone signals are calculated and compared frequency bin wise. The signal with the least energy in the dedicated frequency bin is assumed to be free of wind noise in that bin, since the energy of the speech signal should be the same in both microphone signals after alignment, whereas the PSDs of the wind noise are assumed to be different for each microphone. A dedicated binary map based on the PSD comparison in each frequency bin is calculated as

$$m_1(\kappa,\nu) = \begin{cases} 1, & \text{for } \Phi_{y1}(\kappa,\nu) < \Phi_{y2}(\kappa,\nu) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where Φ_{y1} and Φ_{y2} denote the PSDs of the front and the rear microphone signals, respectively. Therefore it follows for m_2

$$m_2(\kappa,\nu) = 1 - m_1(\kappa,\nu) \quad (4)$$

The resulting signal $\tilde{X}_{FBS}$ can be calculated based on these maps as

$$\tilde{X}_{FBS} = \left(m_1 \cdot \hat{Y}_1 + m_2 \cdot Y_2\right).$$

Delay-and-sum combining

The DSB is capable to offer an improved speech reference compared with a single microphone signal, because it combines the direct path of the desired signal coherently. Incoherent noise signals, spectral coloration, and the direct-to-reverberant ratio are reduced. Using Eq.(2), the DSB output signal $\tilde{X}_{DS}$ is computed as

$$\tilde{X}_{DS} = 0.5 \cdot (\hat{Y}_1 + Y_2). \quad (5)$$

Hybrid combining

Since wind noise is mostly present in low frequencies, the frequency bin selection approach is combined with the delay-and-sum beamformer, whereas the selection is used up to a defined crossover frequency $f_c = 1kHz$ and the DSB is used for frequencies higher than f_c to augment the speech signal.

$$\tilde{X}_{ref}(\kappa,\nu) = \begin{cases} \tilde{X}_{FBS}(\kappa,\nu), & \text{for } f_c \leq 1kHz \\ \tilde{X}_{DS}(\kappa,\nu), & \text{for } f_c > 1kHz \end{cases} \quad (6)$$

Noise Reference

Additionally to the speech reference, a wind noise reference is required to process the microphone signals. A reference for the wind noise can be obtained exploiting the fact that the wind noise components in the two microphones are incoherent. To block the coherent speech, a noise reference is obtained by a delay-and-subtract approach, again using Eq.(2)

$$\tilde{N} = \hat{Y}_1 - Y_2. \quad (7)$$

The signal $\tilde{N}$ denotes the estimated wind noise. Since wind noise is mostly present in low frequencies, the noise reference is filtered with a low pass filter with cutoff frequency $f_g = 1kHz$ to ensure that only incoherent wind noise occurs in the reference and no leakage of speech in higher frequencies affects the estimation.

Single channel Wiener post filter

The Wiener filter (WF) is a prevailing single channel noise reduction technique [2]. Wiener filtering can be applied to suppress the noise in the estimated speech reference with a SNR depending filter. The SNR estimate is calculated using the PSDs for the estimated speech and noise references $\Phi_{\tilde{X}_{ref}}$ and $\Phi_{\tilde{N}}$ as

$$SNR(\kappa, \nu) = \frac{\Phi_{\tilde{X}_{ref}}(\kappa, \nu)}{\Phi_{\tilde{N}}(\kappa, \nu)} \quad (8)$$

Using Eq.(8), the weighting function according to the Wiener filter approach can be calculated as

$$G_{WF} = \frac{SNR}{SNR + \gamma} \quad (9)$$

where γ is a noise overestimation parameter. The SNR-weighted speech reference is obtained by

$$\tilde{X}_{WF} = \tilde{X}_{ref} \cdot G_{WF} \quad (10)$$

FLMS filter

Since wind noise is assumed to be highly instationary, the signal $\tilde{X}_{WF}$ can contain musical noise due to the wiener filtering which assumes that the noise term PSDs are only varying slowly in time. Therefore, similar to [16], we use $\tilde{X}_{WF}$ as a speech reference for the estimation of coherent signals between this reference and one of the microphones. The obtained filters are applied to the input microphone signals to suppress incoherent wind noise components and are estimated as adaptive and time variant filters G_i that are calculated based on the error signals

$$E_i(\kappa, \nu) = \tilde{X}_{WF}(\kappa, \nu) - Y_i(\kappa, \nu) \cdot G_i(\kappa, \nu) \quad (11)$$

which leads to the update of the filter coefficients for the i^{th} microphone

$$G_i(\kappa + 1, \nu) = G_i(\kappa, \nu) + \mu \cdot \frac{Y_i(\kappa, \nu)^* \cdot E_i(\kappa, \nu)}{\Phi_{yi}(\kappa, \nu) + \epsilon} \quad (12)$$

where μ is the adaptation step size and ϵ a regularization parameter to prevent division by zero. * denotes the conjugate complex value. Finally, the resulting output signal $\tilde{X}$ for the microphone array is calculated as

$$\tilde{X} = \begin{cases} 0.5 \cdot (Y_1 \cdot G_1 + Y_2 \cdot G_2), & \text{for } f_c \leq 1kHz \\ \tilde{X}_{DS}, & \text{for } f_c > 1kHz. \end{cases} \quad (13)$$

Simulation Results

In the following, the results of the simulations for the proposed wind noise reduction algorithm are presented. Therefore three wind noise scenarios in a car environment are used, that are recorded with a MEMS array in endfire configuration. Scenario A is a recording where the window at the drivers site is half open, whereas in scenario B the same window is completely open. Scenario C denotes a recording where both windows (at the drivers and the co-drivers site) are completely open. The driving speed was $100km/h$ for all scenarios. The signals for testing the algorithms are ITU speech signals convolved with impulse responses, which where measured with an artificial head at the drivers position and the MEMS array. The array is mounted above the sun visor at the driver seat position. The spacing between the front and the rear microphone of the MEMS array equals $21.4mm$. For the simulations a sampling rate $fs = 16kHz$ and a FFT size of 512 samples is used. The FFT shift is 128 samples and each block is windowed before it is transformed in the frequency domain. The noise recordings and the speech recordings were done separately an mixed in the simulation.

The proposed wind noise reduction approach is compared to a single microphone regarding the segmental signal-to-noise ratio (SSNR) and the LSD. The adaptation step size for the FLMS filter estimation was set to $\mu = 0.4$ and the overestimation parameter for the Wiener post filter was set to $\gamma = 25$. Table 1 shows the results regarding the SSNR. As can be seen, the wind noise is reduced in all scenarios.

Signal	Wind Noise Scenario		
	A	B	C
Y_1	2.9 dB	4.1 dB	1.1 dB
wind noise reduced signal	8.2 dB	9.5 dB	8.2 dB

Table 1: SSNR with wind noise reduction

Table 2 shows the effect of the noise reduction algorithm on speech distortion compared with a single microphone. Due to the coherent combining of the desired signal, the speech distortion is not increased or even improved in scenario C.

Signal	Wind Noise Scenario		
	A	B	C
Y_1	3.5 dB	3.5 dB	3.5 dB
wind noise reduced signal	3.1 dB	3.3 dB	2.5 dB

Table 2: LSD with wind noise reduction

Figure 1 shows the spectogramm of scenario C for the front microphone in comparison to the output of wind noise reduction algorithm. It can be observed that the wind noise terms in the low frequencies are successfully

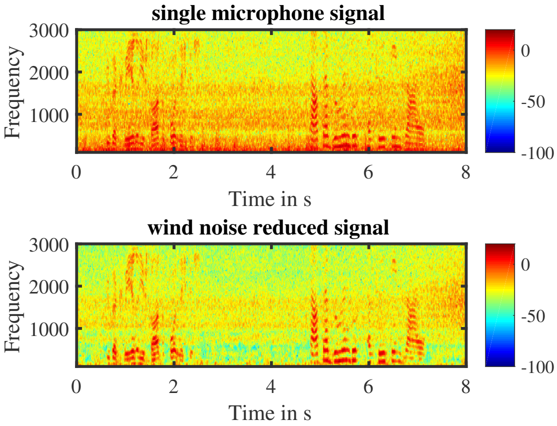

Figure 1: Spectogramm for single microphone (top) and output signal of the proposed wind noise reduction algorithm (bottom)

suppressed, whereas the DSB obtains further signal augmentation for higher frequencies.

Conclusions

In this paper, a wind noise reduction technique using a MEMS endfire array in a car environment was examined. Using two closely spaced microphones allows to exploit the different signal properties of speech and wind noise to acquire suitable reference signals. Based on these references, a single channel post filter was derived to improve the speech reference further. The improved signal was used for an FLMS algorithm to estimate time variant filters based on coherent properties between the reference and the microphone. The filters where applied to the microphone signals to suppress incoherent wind noise signals. The simulation results demonstrate that a significant reduction of wind noise is possible while keeping the speech distortion low.

References

[1] Stuart Bradley, Tao Wu, Sabine von Hünerbein, and Juha Backman, "The mechanisms creating wind noise in microphones," in *Audio Engineering Society Convention 114*, 3 2003.

[2] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, Wiley & Sons, 2006.

[3] 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, 2005.

[4] D. Keith Wilson and Michael J. White, "Discrimination of wind noise and sound waves by their contrasting spatial and temporal properties," *Acta Acustica United with Acustica*, vol. 96, no. 96, pp. 991–1002, 2010.

[5] James W. Kates, *Digital Hearing Aids*, Plural Publishing, 2008.

[6] Sven Franz and Joerg Blitzer, "Multi-channel algorithms for wind noise reduction and signal compensation in binaural hearing aids," in *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 8 2010.

[7] Christoph Matthias Nelke and Peter Vary, "Measurement, analysis and simulation of wind noise signals for mobile communication devices," in *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 9 2014.

[8] Christoph Matthias Nelke, Navin Chatlani, Christophe Beaugeant, and Peter Vary, "Single microphone wind noise PSD estimation using signal centroids," in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 5 2014.

[9] Shingo Kuroiwa, Youji Mori, Satoru Tsuge, Masashi Takashina, and Fuji Ren, "Wind noise reduction method for speech recording using multiple noise templates and observed spectrum fine structure," in *International Conference on Communication Technology*, 11 2006.

[10] Brian King and Les Atlas, "Coherent modulation comb filtering for enhancing speech in wind noise," in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 9 2008.

[11] Elias Nemer and Wilf Leblanc, "Single-microphone wind noise reduction by adaptive postfiltering," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 10 2009.

[12] Christian Hofman, Tobias Wolff, Markus Buck, Tim Haulik, and Walter Kellermann, "A morphological approach to single-channel wind-noise suppression," in *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 9 2012.

[13] Christoph Matthias Nelke, Niklas Nawroth, Marco Jeub, Christophe Beaugeant, and Peter Vary, "Single microphone wind noise reduction using techniques of artificial bandwidth extension," in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 8 2012.

[14] Christoph Matthias Nelke and Peter Vary, "Dual microphone wind noise reduction by exploiting the complex coherence," in *Proceedings of Speech Communications - 11. ITG Symposium*, 9 2014.

[15] Philipp Thüne and Gerald Enzner, "Maximum-likelihood approach to adaptive multichannel-Wiener postfiltering for wind-noise reduction," in *ITG Conference on Speech Communication*, 10 2016.

[16] J. Freudenberger, S. Stenzel, and B. Venditti, "Microphone diversity combining for in-car applications," *EURASIP Journal on Advances in Sig. Proc.*, 2010, Article ID 509541, 13 pages.