

RESEARCH

Open Access



# Wind noise reduction for a closely spaced microphone array in a car environment

Simon Grimm\* and Jürgen Freudenberger

## Abstract

This work studies a wind noise reduction approach for communication applications in a car environment. An endfire array consisting of two microphones is considered as a substitute for an ordinary cardioid microphone capsule of the same size. Using the decomposition of the multichannel Wiener filter (MWF), a suitable beamformer and a single-channel post filter are derived. Due to the known array geometry and the location of the speech source, assumptions about the signal properties can be made to simplify the MWF beamformer and to estimate the speech and noise power spectral densities required for the post filter. Even for closely spaced microphones, the different signal properties at the microphones can be exploited to achieve a significant reduction of wind noise. The proposed beamformer approach results in an improved speech signal regarding the signal-to-noise-ratio and keeps the linear speech distortion low. The derived post filter shows equal performance compared to known approaches but reduces the effort for noise estimation.

**Keywords:** MVDR beamforming, Multichannel Wiener filter, Wind noise reduction, Speech signal processing, MEMS microphones

## 1 Introduction

Hands-free communication applications in a car environment always face the problem of unwanted noise components in the microphone signals. Commonly, single-channel algorithms like the Wiener filter and spectral subtraction are used for noise suppression [1, 2]. Multichannel approaches are able to improve the speech quality further [3–6]. Considering more than one microphone, closely spaced microphones are often used in communication systems for signal augmentation by forming a differential microphone array [7–11]. This allows to create a directivity-dependent beam pattern to augment a desired signal direction, while suppressing noise coming from other incident angles.

The use of micro-electro-mechanical system (MEMS) microphones as a replacement for ordinary microphone capsules has gained interest in [12–14], especially for the application of directive beamforming [15, 16] due to its reduced size and cost compared with an ordinary microphone capsule. However, differential microphone arrays

are not ideal in the presence of wind noise. The directional beam pattern may lead to a significant amplification of the wind noise due to the correlation properties of the noise terms [17]. The required first-order low-pass filter for the equalization regarding the speech signal makes this behavior even worse. One proposed solution for a differential microphone array is to switch to a single microphone with an omnidirectional response if wind noise is detected [17].

Besides car noise, wind noise components often occur in hands-free communication applications in a car environment, 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 [18].

Noise reduction algorithms in car environments are typically based on the assumption that the noise is stationary or varies only slowly in time. In [19], 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 [17, 20, 21]. For single-channel wind noise reduction,

\*Correspondence: [sgrimm@htwg-konstanz.de](mailto:sgrimm@htwg-konstanz.de)  
Institute for System Dynamics, HTWG Konstanz, University of Applied Sciences,  
Alfred-Wachtel-Strasse 8, 78462 Konstanz, Germany

often the different power spectral density (PSD) properties of speech and wind noise are exploited [17, 20, 22]. Several other methods exist that aim to reduce wind noise for a single microphone [23–27].

The utilization of more than one microphone allows to take the diversity of the sound field into account to indicate wind noise and reduce it successfully. In [20], a spectral weighting filter based on the coherence between two microphones is proposed. The coherence is also used in [28], where in addition to the magnitude squared coherence (MSC) the information that relies on the phase component is applied to synthesize a spectral filter function.

In [29], the decomposition of the multichannel Wiener filter into a minimum variance distortionless response (MVDR) beamformer and a single-channel Wiener post filter for an arbitrary microphone arrangement is presented. The approach is based on the assumption that the wind noise is uncorrelated at the microphones, while having equal noise power spectral densities, but arbitrary acoustic transfer functions (ATFs). From these assumptions follows for closely spaced microphones that a simple delay-and-sum (DS) beamformer achieves maximum signal-to-noise-ratio (SNR) beamforming, because equal ATFs from the speech source to the microphones can be assumed for low frequencies.

In this work, we propose a wind noise reduction approach for a closely spaced microphone array consisting of two MEMS microphones, which is considered as a substitute for an ordinary cardioid microphone capsule. The decomposition of the MWF in a beamformer and a single-channel post filter is used similar to [29] as well as the assumption that the wind noise is uncorrelated at the microphones. But in contrast to [29], we assume that the noise powers at the microphones may differ. Since the geometry of the microphone array and the location of the desired speech source are known, additional assumptions about the speech and noise signal properties can be made to design a low-complexity wind noise reduction algorithm. Even for distances of only 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 form a beamformer, as well as to obtain estimates of the speech and noise PSDs for the post filter. Simulations with recorded wind noise show that the proposed approach improves the signal-to-noise-ratio, while keeping the linear distortion of the speech signal low.

The remainder of this paper is structured as follows. The signal model and the notation are briefly introduced in Section 2. In Section 3, the proposed wind noise reduction approach is presented. Simulation results are discussed in Section 4, followed by a conclusion in Section 5.

## 2 Signal model and notation

In the following, the signal model and the notation is briefly explained. We consider a linear MEMS microphone array, which is mounted in a car in front of the speaker's seat in an endfire configuration. The acoustics in the car environment are considered as linear and time invariant. Using the sub-sampled time index  $\kappa$  and the frequency bin index  $\nu$ , the spectrum  $Y_i(\kappa, \nu)$  of the  $i$ th microphone 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  $X(\kappa, \nu)$  corresponds to the short-time spectrum of the speech signal.  $H_i(\nu)$  denotes the acoustic transfer function,  $S_i(\kappa, \nu) = H_i(\nu)X(\kappa, \nu)$  is the spectrum of the speech component, and  $N_i(\kappa, \nu)$  is the spectrum of the noise at the  $i$ th microphone. For two microphones, the signals can be written as vectors

$$\mathbf{S}(\kappa, \nu) = [S_1(\kappa, \nu), S_2(\kappa, \nu)]^T \quad (2)$$

$$\mathbf{N}(\kappa, \nu) = [N_1(\kappa, \nu), N_2(\kappa, \nu)]^T \quad (3)$$

$$\mathbf{H}(\nu) = [H_1(\nu), H_2(\nu)]^T \quad (4)$$

$$\mathbf{Y}(\kappa, \nu) = \mathbf{S}(\kappa, \nu) + \mathbf{N}(\kappa, \nu). \quad (5)$$

Vectors and matrices are written in bold, and scalars are normal letters.  $^T$  denotes the transpose of a vector,  $*$  denotes the complex conjugate, and  $^\dagger$  denotes the conjugate transpose.

We assume that the speech and noise signals are zero-mean random processes with the short-time power spectral densities  $\Phi_{N_i}^2(\kappa, \nu)$  and  $\Phi_{S_i}^2(\kappa, \nu)$  at the  $i$ th microphone. It is assumed that the speech and noise terms are uncorrelated. The noise correlation matrix can be expressed as

$$\mathbf{R}_N(\kappa, \nu) = \mathbb{E} \left\{ \mathbf{N}(\kappa, \nu) \mathbf{N}(\kappa, \nu)^\dagger \right\} \quad (6)$$

and similar the speech correlation matrix as

$$\mathbf{R}_S(\kappa, \nu) = \mathbb{E} \left\{ \mathbf{S}(\kappa, \nu) \mathbf{S}(\kappa, \nu)^\dagger \right\} = \Phi_X^2(\kappa, \nu) \mathbf{H} \mathbf{H}^\dagger, \quad (7)$$

where  $\mathbb{E}$  denotes the mathematical expectation and  $\Phi_X^2(\kappa, \nu)$  the PSD of the clean speech signal. Due to the short-time PSD fluctuations, the PSDs are time and frequency dependent. However, for brevity, the indices  $(\kappa, \nu)$  are often omitted in the following.

## 3 Wind noise reduction algorithm

In this section, the proposed noise reduction algorithm is derived. The filtering is only applied in the low frequency range which is affected by wind noise. It should be noted that the noise signal consists of wind as well as car noise components. However, in the presence of wind noise, the wind noise components are dominant at low frequencies. In the following, we consider only the non-stationary

wind noise components at low frequencies and neglect the slowly varying driving noise. Such stationary noise components can be estimated and reduced by state-of-the-art noise reduction approaches.

The proposed wind noise reduction approach is derived from the commonly used speech distortion weighted multichannel Wiener filter [3], which is defined as

$$\mathbf{G}^{\text{MWF}} = (\mathbf{R}_S + \mu \mathbf{R}_N)^{-1} \Phi_X^2 \mathbf{H} \tilde{H}^* \quad (8)$$

where  $\tilde{H}$  is the acoustic transfer function of an arbitrary chosen microphone channel.  $\mu$  is a noise overestimation parameter which allows a trade-off between noise reduction and speech distortion. The output signal  $Z_{\text{MWF}}$  of the Wiener filter is obtained by

$$Z_{\text{MWF}} = \mathbf{Y} \cdot \mathbf{G}^{\text{MWF}\dagger}. \quad (9)$$

In [30, 31], it is shown that  $\mathbf{G}^{\text{MWF}}$  can be decomposed into an MVDR beamformer

$$\mathbf{G}^{\text{MVDR}} = \frac{\mathbf{R}_N^{-1} \mathbf{H}}{\mathbf{H}^\dagger \mathbf{R}_N^{-1} \mathbf{H}} \quad (10)$$

and a single-channel Wiener post filter

$$G^{\text{WF}} = \frac{\gamma^{\text{out}}}{\gamma^{\text{out}} + \mu} \quad (11)$$

as

$$\mathbf{G}^{\text{MWF}} = \mathbf{G}^{\text{MVDR}} \cdot G^{\text{WF}} \cdot \tilde{H}^*. \quad (12)$$

The term  $\gamma^{\text{out}}$  is the narrow-band SNR at the beamformer output which is defined as

$$\gamma^{\text{out}} = \text{tr}(\mathbf{R}_S \mathbf{R}_N^{-1}), \quad (13)$$

where  $\text{tr}(\cdot)$  denotes the trace operator. We exploit this decomposition for the proposed wind noise reduction. Firstly, we derive a beamformer for the considered microphone setup.

### 3.1 Beamformer

In the following, we consider time-aligned signals where the alignment compensates the different times of arrival for the speech signal. This is achieved by delaying the front microphone with a suitable sample delay  $\tau$  to be in phase with the rear microphone,

$$\hat{Y}_1(v) = Y_1(v) \cdot \begin{cases} e^{-j2\pi \frac{v}{L} \tau} & \text{for } v \in 0, \dots, \frac{L}{2} - 1 \\ e^{j2\pi \frac{v}{L} \tau} & \text{for } v \in \frac{L}{2}, \dots, L - 1 \end{cases} \quad (14)$$

where  $L$  denotes the block length of the short-time Fourier transform. After this alignment, we assume that the ATFs in  $\mathbf{H}$  are identical, because the low frequency speech components have a large wavelength compared with the microphone distance.

$$H = \hat{H}_1 = H_2 \quad (15)$$

$$\mathbf{H} = H \cdot [1, 1]^T \quad (16)$$

which leads to the speech correlation matrix depending only on the PSD of the speech signal at one of the microphones

$$\mathbf{R}_S = \Phi_X^2 |H|^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \Phi_S^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (17)$$

Furthermore, it can be assumed that the wind noise terms for both microphone signals are uncorrelated even for small distances of the microphones [28, 32]. This simplifies the noise correlation matrix as well as its inverse since the cross-terms can be neglected

$$\mathbf{R}_N^{-1} = \begin{pmatrix} \frac{1}{\Phi_{N_1}^2} & 0 \\ 0 & \frac{1}{\Phi_{N_2}^2} \end{pmatrix}. \quad (18)$$

The numerator term of the  $\mathbf{G}^{\text{MVDR}}$  in (10) can be written as

$$\mathbf{R}_N^{-1} \mathbf{H} = H \cdot \begin{pmatrix} \frac{1}{\Phi_{N_1}^2} \\ \frac{1}{\Phi_{N_2}^2} \end{pmatrix} \quad (19)$$

and the denominator as

$$\mathbf{H}^\dagger \mathbf{R}_N^{-1} \mathbf{H} = |H|^2 \cdot \left( \frac{1}{\Phi_{N_1}^2} + \frac{1}{\Phi_{N_2}^2} \right). \quad (20)$$

Since  $H$  is not known, it is set to  $H = 1$ . This results in the minimum variance (MV) beamformer coefficients

$$G_i^{\text{MV}} = \frac{\frac{1}{\Phi_{N_i}^2}}{\frac{1}{\Phi_{N_1}^2} + \frac{1}{\Phi_{N_2}^2}}, \quad (21)$$

which can be interpreted as a noise-dependent weighting of the input signals. Note that the MV beamformer achieves the same narrow-band output SNR as the MVDR beamformer but no distortion-free response [5]. Finally, the output of the beamformer can be written as

$$Y_{\text{MV}} = (\hat{Y}_1 \cdot G_1^{\text{MV}} + Y_2 \cdot G_2^{\text{MV}}). \quad (22)$$

Using (17) and (18), we are able to calculate the narrow-band output SNR of the beamformer as

$$\gamma^{\text{out}} = \Phi_S^2 \cdot \left( \frac{1}{\Phi_{N_1}^2} + \frac{1}{\Phi_{N_2}^2} \right) = \frac{\Phi_S^2}{\Phi_{N_{\text{beam}}}^2}, \quad (23)$$

where  $\Phi_{N_{\text{beam}}}^2$  denotes the noise PSD at the beamformer output. This PSD can be calculated as

$$\Phi_{N_{\text{beam}}}^2 = \frac{\Phi_{N_1}^2 \cdot \Phi_{N_2}^2}{\Phi_{N_1}^2 + \Phi_{N_2}^2}. \quad (24)$$

### 3.2 Special cases

In the following, we consider some special cases for the beamformer derived in (22). Assuming  $\Phi_{N_1}^2 = \Phi_{N_2}^2$  and uncorrelated noise terms as in [29], then  $G_i^{\text{MV}}$  reduces to

the simple weighting of a delay-and-sum beamformer (a simple summing of the aligned signals)

$$G_i^{DS} = \frac{\frac{1}{\Phi_{N_1}^2}}{\frac{1}{\Phi_{N_1}^2} + \frac{1}{\Phi_{N_2}^2}} = \frac{1}{2}, \quad (25)$$

which results in the output signal

$$Y_{DS} = \frac{1}{2} (\hat{Y}_1 + Y_2). \quad (26)$$

A delay-and-sum beamformer is also proposed in [17] for closely spaced microphones with wind noise.

We keep the condition of uncorrelated noise terms and assume a special case where the short-time noise PSDs are varying over time and frequency. This is motivated by the highly non-stationary local short-time wind noise disturbances [19] and implies that only one microphone is affected by wind noise at any given time and frequency index  $\kappa$  and  $\nu$

$$\Phi_{N_1}^2(\kappa, \nu) \ll \Phi_{N_2}^2(\kappa, \nu) \quad (27)$$

or

$$\Phi_{N_1}^2(\kappa, \nu) \gg \Phi_{N_2}^2(\kappa, \nu). \quad (28)$$

Then, the noise PSD-dependent weighting in (21) reduces to a selection approach of the dedicated frequency bins by comparing the short-time PSDs of the microphone signals  $\Phi_{Y_i}^2$ , because the speech signal PSDs  $\Phi_{S_i}^2$  are assumed to be identical for both microphones. Therefore, the resulting output signal  $Y_{FBS}$  can be written as

$$Y_{FBS}(\kappa, \nu) = \begin{cases} Y_1(\kappa, \nu), & \Phi_{Y_1}^2(\kappa, \nu) < \Phi_{Y_2}^2(\kappa, \nu) \\ Y_2(\kappa, \nu), & \Phi_{Y_1}^2(\kappa, \nu) > \Phi_{Y_2}^2(\kappa, \nu) \end{cases} \quad (29)$$

### 3.3 PSD estimation

Next, we derive estimates for the speech and noise PSDs which are required for the beamformer and post filter. As mentioned in [29], most single-channel noise estimation procedures (i.e., [33–35]) rely on the assumption that the noise signal PSDs are varying more slowly in time than the speech signal PSD. This is not the case for wind noise. The fast varying short-time PSDs make noise estimation a challenging task for a single microphone. However, using more than one microphone, the different correlation properties for speech and wind noise can be used for the estimation.

A reference for the wind noise can be obtained by exploiting the fact that the wind noise components in the two microphones are incoherent while the speech components are coherent. To block the speech signal, a delay-and-subtract approach is used to obtain a noise reference

$$N = \frac{\hat{Y}_1 - Y_2}{2}, \quad (30)$$

which depends only on incoherent wind noise terms. The PSD of this noise reference is

$$\Phi_N^2 = \mathbb{E} \{ NN^* \} \quad (31)$$

$$= \mathbb{E} \left\{ \left( \frac{\hat{Y}_1 - Y_2}{2} \right) \left( \frac{\hat{Y}_1 - Y_2}{2} \right)^* \right\} \quad (32)$$

$$= \frac{1}{4} \left( \mathbb{E} \{ \hat{Y}_1 \hat{Y}_1^* \} - \mathbb{E} \{ \hat{Y}_1 Y_2^* \} - \mathbb{E} \{ Y_2 \hat{Y}_1^* \} + \mathbb{E} \{ Y_2 Y_2^* \} \right) \quad (33)$$

$$= \frac{1}{4} \left( \mathbb{E} \{ \hat{N}_1 \hat{N}_1^* \} - \mathbb{E} \{ \hat{N}_1 N_2^* \} - \mathbb{E} \{ N_2 \hat{N}_1^* \} + \mathbb{E} \{ N_2 N_2^* \} \right). \quad (34)$$

The cross-terms vanish, because the wind noise terms are uncorrelated. Hence, we obtain

$$\Phi_N^2 = \frac{\Phi_{N_1}^2}{4} + \frac{\Phi_{N_2}^2}{4}. \quad (35)$$

Note that the delay-and-subtract signal in (30) is used in other applications as the output of a differential microphone array [17]. Obviously, this is not suitable for microphone positions that are sensitive to wind noise, because the noise terms are heavily amplified.

By summing the aligned signals according to (26), we augment coherent signal components. The combined signal  $Y_{DS}$  has the PSD

$$\Phi_{Y_{DS}}^2 = \mathbb{E} \{ Y_{DS} Y_{DS}^* \} \quad (36)$$

$$= \mathbb{E} \left\{ \left( \frac{\hat{Y}_1 + Y_2}{2} \right) \left( \frac{\hat{Y}_1 + Y_2}{2} \right)^* \right\} \quad (37)$$

$$= \frac{1}{4} \left( \mathbb{E} \{ \hat{Y}_1 \hat{Y}_1^* \} + \mathbb{E} \{ \hat{Y}_1 Y_2^* \} + \mathbb{E} \{ Y_2 \hat{Y}_1^* \} + \mathbb{E} \{ Y_2 Y_2^* \} \right) \quad (38)$$

$$= \mathbb{E} \{ SS^* \} + \frac{1}{4} \left( \mathbb{E} \{ \hat{N}_1 \hat{N}_1^* \} + \mathbb{E} \{ \hat{N}_1 N_2^* \} + \mathbb{E} \{ N_2 \hat{N}_1^* \} + \mathbb{E} \{ N_2 N_2^* \} \right). \quad (39)$$

Again, the noise cross-terms vanish and we obtain

$$\Phi_{Y_{DS}}^2 = \Phi_S^2 + \frac{\Phi_{N_1}^2}{4} + \frac{\Phi_{N_2}^2}{4}. \quad (40)$$

Combining (35) and (40) yields the PSD of the clean speech signal

$$\Phi_S^2 = \Phi_{Y_{DS}}^2 - \Phi_N^2 \quad (41)$$

and the noise PSD at the  $i$ th microphone

$$\Phi_{N_i}^2 = \Phi_{Y_i}^2 - \Phi_S^2. \quad (42)$$

Note that this derivation only holds for uncorrelated noise terms.  $\Phi_S^2$  may still contain correlated noise. However, we neglect the correlated driving noise as stated at the beginning of this section. In contrast to Zelinskis post

filter [36], which also assumes zero correlation between the microphone signals, we assume the short-time noise PSDs to be different ( $\Phi_{N_1}^2 \neq \Phi_{N_2}^2$ ).

### 3.4 Post filter

As described in (12), the beamformer is followed by a single-channel Wiener post filter to achieve additional noise suppression. We use the post filter

$$G^{WF} = \frac{\gamma}{\gamma + \mu}. \quad (43)$$

with the SNR estimate

$$\gamma = \frac{\Phi_S^2}{\Phi_N^2}. \quad (44)$$

That is, the noise PSD is estimated according to (35) instead of (23), because this estimate showed a better performance in the simulations regarding SNR and speech distortion. Note that  $\Phi_N^2 \geq \Phi_{N_{beam}}^2$  holds, with equality if  $\Phi_{N_1}^2 = \Phi_{N_2}^2$ . Hence, the noise estimation in (44) results in an overestimation of the noise power if the short-time PSDs at the microphones vary. This is similar to using an overestimation parameter  $\mu > 1$ .

Finally, the output of the complete wind noise reduction algorithm is

$$Z = (\hat{Y}_1 \cdot G_1^{MV} + Y_2 \cdot G_2^{MV}) \cdot G^{WF} \quad (45)$$

$$= Y_{MV} \cdot G^{WF}. \quad (46)$$

This wind noise reduction algorithm is only applied for frequencies below a cutoff frequency  $f_c$ , because wind noise mostly contains low frequency components and the assumptions about the signal properties are only valid for low frequencies. Figure 1 shows the block diagram of the signal processing structure.

## 4 Simulation results

In the following, simulation results for the algorithm proposed in Section 3 are presented for wind noise in a car. For the signal measurements, a linear MEMS microphone

array in an endfire configuration was mounted above the sun visor at the driver seat position. To investigate varying microphone distances, an array with four sensors was used. The microphone distances were 7.1, 14.3, and 21.4 mm.

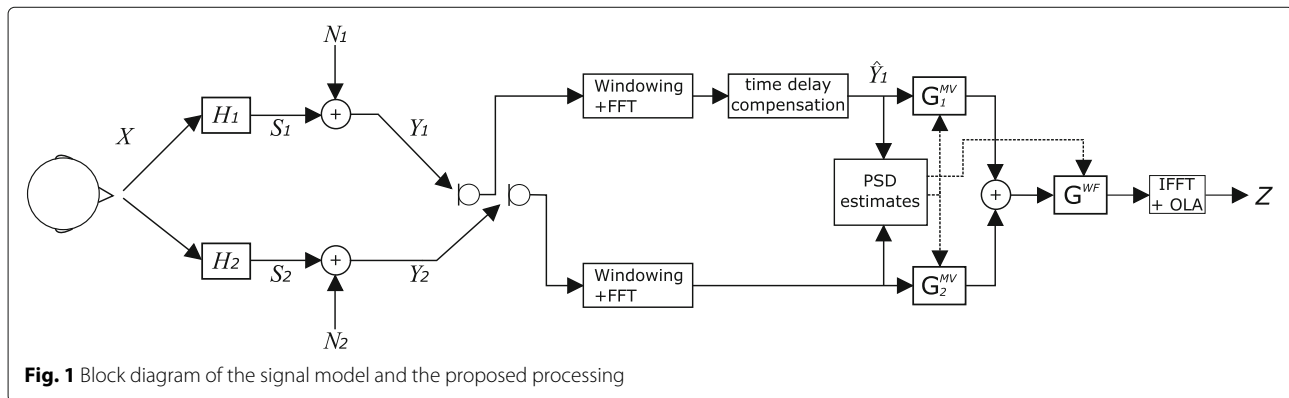
The noise recordings and the speech recordings were done separately and mixed in the simulation. For the noise recordings, the driving speed was 100 km/h and both front windows at the driver side as well as the co-driver side were completely open to allow a turbulence airflow over the MEMS array. The speech signals for testing were four ITU speech signals convolved with the impulse responses, which were measured from the mouth reference point of an artificial head (HMS II.5 from HEAD acoustics) at the driver's position to the MEMS array microphones.

For the simulations, a sampling rate  $f_s = 16$  kHz and an fast Fourier transform (FFT) size of 512 samples was used. The FFT shift was 128 samples, and each block was windowed before it was transformed into the frequency domain. The cutoff frequency  $f_c$  was set to 1 kHz.

As quality measures, we consider the segmental signal-to-noise ratio (SSNR), the log spectral distance (LSD), as well as short-time objective intelligibility measure (STOI) as described in [37]. The STOI is a metric for speech intelligibility.

It should be noted that the SSNR and LSD measures are calculated for the frequency region below the cutoff frequency  $f_c$  since the frequency region above  $f_c$  is not affected by the proposed wind noise reduction approach. Therefore, the signals are transformed back into the time domain and are low-pass filtered to calculate the SSNR and LSD values. The STOI is calculated over the complete frequency range with 15 third-octave bands.

The LSD measures the linear speech distortion and is calculated as the average logarithmic spectral distance of two PSDs. These are the signals under test, i.e., the speech component of the filtered output signal and the clean speech reference  $X$ . The PSDs are calculated over all speech active blocks using an ideal voice activity detector.





For further details regarding the LSD calculation, we refer to [38].

The SSNR is calculated based on [39]. However, we calculate the SSNR by the ratio of the signal energy of the speech and the noise components in speech active frames as

$$SSNR = \frac{1}{K} \sum_{l=0}^{K-1} \left[ 10 \log_{10} \left( \frac{\sum_{k=RI}^{RI+M-1} |\tilde{s}(k)|^2}{\sum_{k=RI}^{RI+M-1} |\tilde{n}(k)|^2} \right) \right]_{-10}^{35}. \quad (47)$$

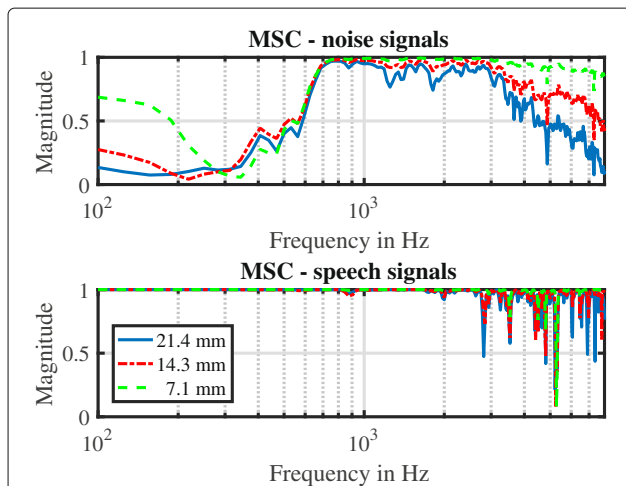
$\tilde{s}(k)$  and  $\tilde{n}(k)$  are the speech and noise components at the output of the dedicated noise reduction approach in the time domain.  $k$  is the time index,  $M$  is the frame length,  $R$  is the frame shift, and  $K$  is the total number of considered frames. The frame length was 512 samples, and the frame shift was 256 samples. The SSNR values are limited between  $-10$  and  $35$  dB.

Car noise, which is also present in the microphone signals, is not considered in our algorithm. Thus, the SSNR improvements in absolute value can be lower compared with measured noise signals which contain wind noise only.

#### 4.1 Coherence properties

Figure 2 shows the results of the magnitude squared coherence calculation of speech and noise for varying microphone distances. The magnitude squared coherence for two signals  $u_1(k)$  and  $u_2(k)$  is calculated as

$$MSC = \left| \frac{\mathbb{E}\{u_1 u_2^*\}}{\sqrt{\mathbb{E}\{u_1 u_1^*\} \cdot \mathbb{E}\{u_2 u_2^*\}}} \right|^2, \quad (48)$$



**Fig. 2** Magnitude squared coherence for the noise signals  $N_1$  and  $N_2$  (top) as well as the speech signals  $S_1$  and  $S_2$  (bottom) with different microphone distances

where  $U_1$  and  $U_2$  denote the corresponding short-time spectra. The mathematical expectation values of the input signals are estimated by the Welch periodogram using recursive smoothing. A very high smoothing factor of 0.9995 was chosen to average over many signal frames. An MSC value close to one means the signals are highly correlated, whereas a value close to zero indicates that the signals are uncorrelated.

As can be observed, the assumption that noise is uncorrelated while speech is highly correlated is fulfilled for frequencies below 600 Hz for all microphone distances, which justifies the assumptions made in Section 3.

#### 4.2 Beamformer output

In Table 1, the SSNR gain of the beamformer output is compared with a single microphone. This comparison is considered, because the approach in [17] suggest to switch from a differential microphone array to a single omnidirectional microphone if wind noise is detected. The SSNR of the single microphone is 2.14 dB. For further comparison, the results of the delay-and-sum beamformer  $Y_{DS}$  are shown, which is the summing of the aligned signals as described in (26) (and also proposed in [17] for combining of wind noise-affected signals). Also, the output of a frequency bin selection ( $Y_{FBS}$ ) approach as stated in (29) is examined. The noise estimates in (42), as derived in Section 3.3, are used for the beamformer. Moreover, the ideal noise PSDs are used to get a benchmark. Since the noise signals were recorded separately for the simulations, the ideal noise PSDs are obtained by using the noise only signals for the PSD calculation. The PSDs are calculated by the Welch periodogram using recursive smoothing. However, the short-time recursive PSD smoothing was omitted, because this achieved the best results due to the high non-stationarity of the wind noise.

As can be observed, all beamformer approaches are able to improve the SSNR in the considered frequency region compared with a single microphone, where all SNR gains are getting larger as the distance between the microphones is increased. It is interesting to see that the delay-and-sum approach  $Y_{DS}$  has the worst performance for all microphone distances, whereas the frequency bin selection approach shows results similar to

**Table 1** SSNR gain compared with single microphone for different beamformer outputs

Signal	Microphone distance		
	7.1 mm	14.3 mm	21.4 mm
$Y_{DS}$ (Eq. (26))	0.96 dB	1.22 dB	1.58 dB
$Y_{FBS}$ (Eq. (29))	1.80 dB	2.25 dB	2.47 dB
$Y_{MV}$ (Eq. (22)) (noise estimate)	1.77 dB	2.34 dB	2.80 dB
$Y_{MV}$ (Eq. (22)) (noise benchmark)	1.80 dB	2.51 dB	3.04 dB

the MV beamformer. This indicates that the short-time PSDs at the microphones vary heavily. Comparing the performance with estimated noise PSDs with that of the beamformer with the actual noise PSDs, we observe that the results regarding the SSNR are similar, i.e., the PSD estimates are sufficiently accurate.

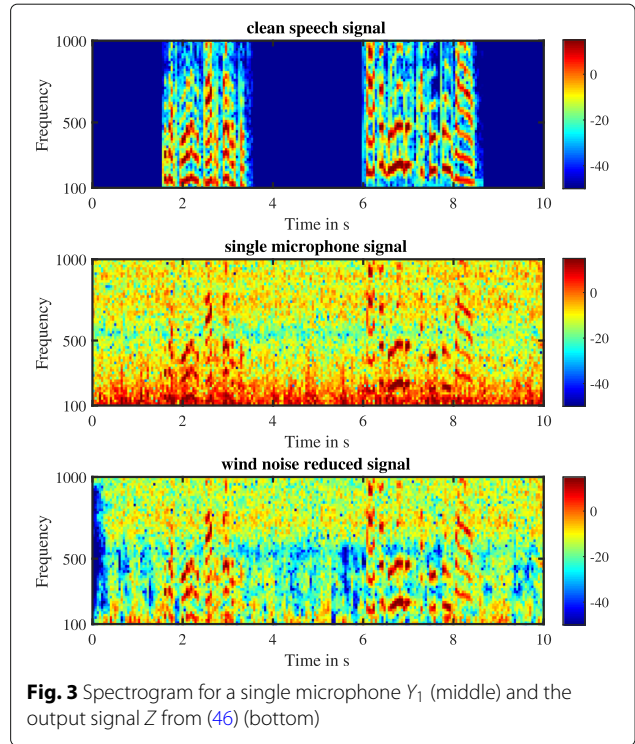
#### 4.3 Post filter output

Now, the SSNR as well as the LSD for the complete MWF including the post filter (as derived in (46)) are examined. To compare the post filter of (43) with other approaches, a wind noise reduction filter by Franz et al. [20] that defines a filter function based on the magnitude squared coherence is used as a reference. The proposed post filter in (43) as well as the post filter derived in [20] are applied to the beamformer output  $Y_{MV}$  which uses the noise estimates. As can be seen in Table 2, the SSNR can be further improved while keeping the speech distortion below 1 dB compared with the single microphone signal  $Y_1$ .

For the post filter comparison, the noise overestimation parameter  $\mu$  was set to achieve a similar LSD value as the post filter in [20]. The short-time PSDs used for the post filter, as well as the calculated MSC needed for the filter design in [20], were recursively smoothed by the same factor of 0.85 to make a fair comparison. As can be seen, both post filters are able to achieve the same noise reduction.

Table 2 also contains values for the STOI. The STOI is closely related to the percentage of correctly understood words averaged across a group of users. The maximum STOI value is one and larger values indicate better speech intelligibility. The noisy speech signals are compared with the time domain signal of the clean speech  $X$ . It can be seen in Table 2 that the STOI is increased for the beamformer output  $Y_{MV}$  compared with the single microphone  $Y_1$ . The results indicate that additional post filtering improves the STOI, where the post filters obtain similar STOI values.

Figure 3 shows the spectrogram for the omnidirectional reference microphone, as well as the output  $Z$  of our proposed wind noise reduction algorithm with a microphone distance of 21.4 mm. It can be observed that the high energetic noise terms in the low frequencies are successfully suppressed. Above 600 Hz the noise reduction is not as strong, i.e., the assumptions for the wind noise signal



**Fig. 3** Spectrogram for a single microphone  $Y_1$  (middle) and the output signal  $Z$  from (46) (bottom)

properties with this noise recording are only valid for frequencies below 600 Hz (cf. Fig. 2).

#### 4.4 Wind noise only scenario

Finally, the wind noise reduction is considered in a scenario containing only wind noise and no driving noise. The SSNR of the single microphone  $Y_1$  is 4.86 dB in this scenario. The results can be seen in Table 3. Again, the beamformer output  $Y_{MV}$  with noise estimation is used with both post filter approaches as in Section 4.3. All parameters except for the overestimation parameter are the same. The table contains results for two different values of the overestimation parameter for the Wiener post filter in order to demonstrate the trade-off between speech distortion and noise reduction. With  $\mu = 8$ , the Wiener filter and the filter from [20] obtain similar performance values. Reducing the overestimation parameter to  $\mu = 1$  also reduces the SNR gain, but results in better LSD and STOI values. Comparing the results with the gains in

**Table 2** Results for the post filter output for wind noise and driving noise

Signal	SSNR gain	LSD	STOI
$Y_1$	–	2.3 dB	0.799
$Y_{MV}$ (Eq. (22))	2.80 dB	2.2 dB	0.821
$Y_{MV}$ (Eq. (22)) + post filter after [20]	5.43 dB	3.3 dB	0.825
$Z$ (Eq. (46)), $\mu = 6$	5.44 dB	3.3 dB	0.826

**Table 3** Results for the post filter output in a scenario containing only wind noise

Signal	SSNR gain	LSD	STOI
$Y_1$	–	2.3 dB	0.896
$Y_{MV}$ (Eq. (22))	3.42 dB	2.3 dB	0.917
$Y_{MV}$ (Eq. (22)) + post filter after [20]	9.18 dB	3.1 dB	0.900
$Z$ (Eq. (46)), $\mu = 8$	9.34 dB	3.3 dB	0.906
$Z$ (Eq. (46)), $\mu = 1$	5.69 dB	2.4 dB	0.923

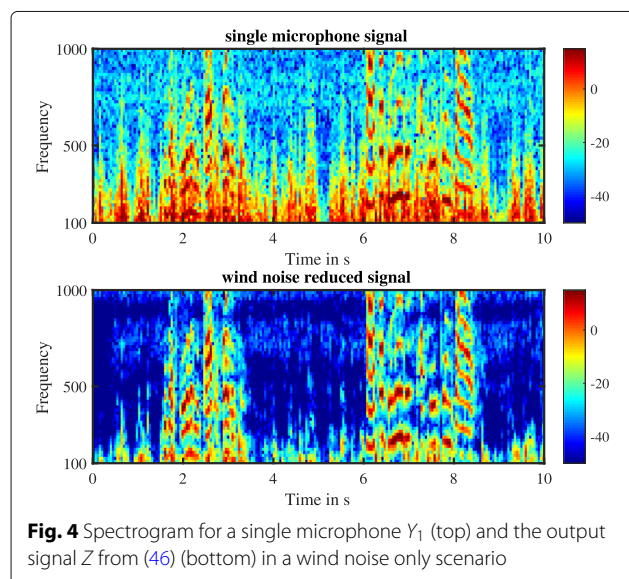
Table 2, the achieved SSNR values are higher due to the absence of the driving noise.

Figure 4 shows the spectrogram of the output  $Z$  for the wind noise only scenario. The noise is significantly reduced over a wide frequency range. Since the coherent driving noise terms are not present in this scenario, noise reduction can also be observed for frequencies above 600 Hz.

## 5 Conclusions

In this paper, a wind noise reduction approach for a compact endfire array was examined. Based on the decomposition of the MWF, a beamformer and a post filter were derived. Due to the known geometry of the MEMS microphone array in endfire configuration and knowledge about the position of the speech source, assumptions about the signal properties of the speech and wind noise components were made. The acquired estimates of the PSDs for the wind noise as well as the speech signals are used to design a beamformer as well as a post filter for wind noise reduction. The simulations based on noise recordings in a car environment show that a significant wind noise reduction is possible while keeping the speech distortion low.

Further investigations should be made to combine the proposed wind noise reduction approach with the reduction of car noise. The driving noise is neglected in our study. The compact microphone array can be part of an array of more widely spaced microphones, where the spatial diversity of the sound field can be exploited for further noise reduction. Since the non-stationary noise terms are mostly reduced with the proposed approach, state-of-the-art noise estimation procedures can be chosen that rely on the assumption that the driving noise is only slowly varying.



Wind noise-induced disruptions are a commonly known problem with differential beamforming, e.g., with the closely spaced microphone arrangements in hearing aids [17]. Hence, the proposed noise reduction approach may also be applicable for hearing aids.

## Abbreviations

ATF: Acoustic transfer function; DS: Delay-and-sum; FFT: Fast Fourier transform; LSD: Log spectral distance; MEMS: Micro-electro-mechanical system; MSC: Magnitude squared coherence; MV: Minimum variance; MVDR: Minimum variance distortionless response; MWF: Multichannel Wiener filter; PSD: Power spectral density; SNR: Signal-to-noise-ratio; SSNR: Segmental signal-to-noise-ratio; STOI: Short-time objective intelligibility measure

## Acknowledgements

We thank the Daimler AG, Department Enabling Technologies for Communication, Ulm, for providing the measurement data.

## Availability of data and materials

The measurement data was used by courtesy of Daimler AG. It is not available for public access.

## Authors' contributions

Both authors developed the idea of the proposed algorithm. JF initiated the theoretical description, while SG implemented the algorithm and refined it in the simulations. The simulations and a majority of the manuscript writing were done by SG, while JF supervised the simulations and helped in improving the text. Both authors read and approved the final manuscript.

## Authors' information

Simon Grimm (SG) is a member of the signal processing group at the Institute for System Dynamics at the HTWG Konstanz since 2014. His work is primarily concerned with the development of signal processing algorithms for multichannel noise reduction approaches in noisy acoustic environments. He received his B. Eng. in 2012 and his M. Eng. in 2014. Dr. Jürgen Freudenberger (JF) is a professor at the HTWG Konstanz since 2006, where he is the head of the signal processing group at the Institute for System Dynamics. His work is primarily concerned with the development of algorithms in the field of signal processing and coding for reliable data transmission as well as efficient algorithm implementation for hardware and software.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 12 September 2017 Accepted: 9 July 2018

Published online: 27 July 2018

## References

1. P Vary, R Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*. (Wiley, Chichester, 2006)
2. E Hänsler, G Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. (Wiley, New Jersey, 2004)
3. S Doclo, A Spriet, M Moonen, J Wouters, in *Speech Enhancement*. Speech distortion weighted multichannel Wiener filtering techniques for noise reduction (Springer, Berlin, 2005). Chap. 9. [https://doi.org/10.1007/3-540-27489-8\\_9](https://doi.org/10.1007/3-540-27489-8_9)
4. S Doclo, A Spriet, J Wouters, M Moonen, Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction. *Speech Comm.* **49**(7-8), 636–656 (2007). <https://doi.org/10.1016/j.specom.2007.02.001>
5. S Stenzel, J Freudenberger, Blind matched filtering for speech enhancement with distributed microphones. *J. Electr. Comput. Eng.* **2012**, 636 (2012). Article ID 169853



6. T Matheja, M Buck, T Fingscheidt, A dynamic multi-channel speech enhancement system for distributed microphones in a car environment. *EURASIP J. Adv. Signal Proc.* **2013** (2013)
7. J Benesty, C Jingdong, *Study and Design of Differential Microphone Arrays*. (Springer, Berlin, 2013)
8. GW Elko, *Differential microphone arrays*. In: Y Huang, J Benesty. (eds) *Audio Signal Processing for Next-Generation Multimedia Communication System*. (Springer, Boston, 2004)
9. H Teutsch, GW Elko, in *International Workshop on Acoustic Signal Enhancement*. First- and second-order adaptive differential microphone arrays, (2001), pp. 35–38
10. J Benesty, M Souden, Y Huang, A perspective on differential microphone arrays in the context of noise reduction. *IEEE Trans. Audio, Speech, Lang. Process.* **20**(2), 699–704 (2012). <https://doi.org/10.1109/TASL.2011.2163396>
11. GW Elko, Microphone array systems for hands-free telecommunication. *Speech Commun.* **20**(3), 229–240 (1996). [https://doi.org/10.1016/S0167-6393\(96\)00057-X](https://doi.org/10.1016/S0167-6393(96)00057-X). Acoustic Echo Control and Speech Enhancement Techniques
12. M Turqueti, J Saniie, E Oruklu, in *2010 53rd IEEE International Midwest Symposium on Circuits and Systems*. MEMS acoustic array embedded in an FPGA based data acquisition and signal processing system, (2010), pp. 1161–1164. <https://doi.org/10.1109/MWSCAS.2010.5548866>
13. I Hafizovic, C-IC Nilsen, M Kjølervbakken, V Jahr, Design and implementation of a MEMS microphone array system for real-time speech acquisition. *Appl. Acoust.* **73**(2), 132–143 (2012). <https://doi.org/10.1016/j.apacoust.2011.07.009>
14. J Tietze, F Domínguez, Bd Silva, L Segers, K Steenhaut, A Touhafi, Soundcompass: a distributed mems microphone array-based sensor for sound source localization. *Sensors*. **14**(2), 1918–1949 (2014). <https://doi.org/10.3390/s140201918>
15. G Elko, Small directional microelectromechanical systems (MEMS) microphone arrays. *Proc. Meet. Acoust.* **19**(1), 030033 (2013). <https://doi.org/10.1121/1.4799608>. <http://asa.scitation.org/doi/pdf/10.1121/1.4799608>
16. A Palla, L Fanucci, R Sannino, M Settin, in *2015 10th International Conference on Design Technology of Integrated Systems in Nanoscale Era (DTIS)*. Wearable speech enhancement system based on MEMS microphone array for disabled people, (2015), pp. 1–5. <https://doi.org/10.1109/DTIS.2015.7127384>
17. JW Kates, *Digital Hearing Aids*. (Plural Publishing, San Diego, 2008)
18. S Bradley, T Wu, S von Hünenbein, J Backman, in *Audio Engineering Society Convention 114*. The mechanisms creating wind noise in microphones, (2003)
19. DK Wilson, MJ White, Discrimination of wind noise and sound waves by their contrasting spatial and temporal properties. *Acta Acustica United Acustica*. **96**(96), 991–1002 (2010)
20. S Franz, J Blitzer, in *International Workshop on Acoustic Signal Enhancement (IWAENC)*. Multi-channel algorithms for wind noise reduction and signal compensation in binaural hearing aids, (2010)
21. CM Nelke, P Vary, in *International Workshop on Acoustic Signal Enhancement (IWAENC)*. Measurement, analysis and simulation of wind noise signals for mobile communication devices, (2014)
22. CM Nelke, N Chatlani, C Beaugeant, P Vary, in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*. Single microphone wind noise PSD estimation using signal centroids, (2014)
23. S Kuroiwa, Y Mori, S Tsuchi, M Takashina, F Ren, in *International Conference on Communication Technology*. Wind noise reduction method for speech recording using multiple noise templates and observed spectrum fine structure, (2006)
24. B King, L Atlas, in *Proceedings of International Workshop on Acoustic Signal Enhancement (IWAENC)*. Coherent modulation comb filtering for enhancing speech in wind noise, (2008)
25. E Nemer, W Leblanc, in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. Single-microphone wind noise reduction by adaptive postfiltering, (2009)
26. C Hofman, T Wolff, M Buck, T Haulik, W Kellermann, in *Proceedings of International Workshop on Acoustic Signal Enhancement (IWAENC)*. A morphological approach to single-channel wind-noise suppression, (2012)
27. CM Nelke, N Nawroth, M Jeub, C Beaugeant, P Vary, in *Proceedings of European Signal Processing Conference (EUSIPCO)*. Single microphone wind noise reduction using techniques of artificial bandwidth extension, (2012)
28. CM Nelke, P Vary, in *Proceedings of Speech Communications - 11. ITG Symposium*. Dual microphone wind noise reduction by exploiting the complex coherence, (2014)
29. P Thüne, G Enzner, in *ITG Conference on Speech Communication*. Maximum-likelihood approach to adaptive multichannel-Wiener postfiltering for wind-noise reduction, (2016)
30. KU Simmer, J Bitzer, C Marro, in *Microphone Arrays: Signal Processing Techniques and Applications*, ed. by MS Brandstein. Post-filtering techniques (Springer, Berlin Heidelberg, 2001), pp. 39–60
31. KU Simmer, J Bitzer, in *Jahrestagung für Akustik (DAGA), Aachen*. Multi-microphone noise reduction — theoretical optimum and practical realization, (2003)
32. GM Corcos, The structure of the turbulent pressure field in boundary-layer flows. *J. Fluid Mech.* **18**(3), 353–378 (1964). <https://doi.org/10.1017/S002211206400026X>
33. R Martin, Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Trans. Speech Audio Process.* **9**, 504–512 (2001)
34. J Freudenberger, S Stenzel, B Venditti, in *Proc. European Signal Processing Conference (EUSIPCO), Glasgow*. Spectral combining for microphone diversity systems, (2009), pp. 854–858
35. J Freudenberger, S Stenzel, in *IEEE Workshop on Statistical Sig. Proc. (SSP)*. Time-frequency dependent voice activity detection based on a simple threshold test (IEEE, Nice, 2011)
36. R Zelinski, in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms, (1988), pp. 2578–25815. <https://doi.org/10.1109/ICASSP.1988.197172>
37. CH Taal, RC Hendriks, R Heusdens, J Jensen, An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Trans. Audio Speech Lang. Process.* **19**(7), 2125–2136 (2011). <https://doi.org/10.1109/TASL.2011.2114881>
38. PA Naylor, ND Gaubitch, *Speech Dereverberation*, 1st edn. (Springer, London, 2010)
39. K Kondo, *Subjective Quality Measurement of Speech*. (Springer, Berlin, 2012)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)