MULTI-CHANNEL WIND NOISE REDUCTION USING THE CORCOS MODEL

Daniele Mirabilii *Emanuël A.P. Habets*

International Audio Laboratories Erlangen*, Am Wolfsmantel 33, 91058 Erlangen, Germany
{daniele.mirabilii,emanuel.habets}@audiolabs-erlangen.de

ABSTRACT

Outdoor recordings of speech are often corrupted by wind noise, which is difficult to reduce due to its high non-stationarity. In this work, a multi-channel wind noise reduction method is presented, based on a joint estimation of the speech and wind noise power spectral densities. In contrast to existing methods that assume uncorrelated wind noise, the estimation phase is performed exploiting the spatial characteristics of wind noise measured by closely-spaced microphones. Here the characteristics are approximated by a fluid-dynamics model, termed the Corcos model. An additional contribution is the employment of a frequency dependent trade-off parameter in the reduction phase, which depends on the ratio of the difference-signal power to the sum-signal power of a sub-set of two microphones. In particular, the trade-off parameter of the parametric multi-channel Wiener filter is used to control the trade-off between noise reduction and speech distortion. The proposed method can be used also to reduce spatially uncorrelated wind noise. An evaluation in terms of speech quality and signal-to-noise ratio improvements is carried out to compare the proposed method to an existing multi-channel wind noise reduction method, using recorded and simulated wind noise samples.

***Index Terms*—** Noise reduction, wind noise, multi-channel, Corcos model

1. INTRODUCTION

Wind noise can result in an adverse acoustic condition in outdoor recordings due to the interference generated by the interaction of the air flow with the microphone membrane. This results in highly non-stationary low frequency distortions which can severely degrade the quality of desired signals, e.g., speech. Wind noise reduction is therefore crucial to enhance the distorted signal. Existing background noise reduction approaches typically assume slowly time-varying noise statistics, and are not able to deal with the rapid fluctuations due to the wind. Moreover, hardware solutions like wind screens cannot always be used, especially for small commercial devices like smartphones or cameras.

Existing wind noise-reduction approaches can be classified based on the number of microphones exploited in the processing, i.e., single-channel [1–4] and multi-channel [5–9]. In this respect, a well-established online scheme is given by a system composed of (a) a detection phase in which the wind noise presence is assessed, (b) an estimation phase in which the speech and wind noise power spectral denisties (PSD) are obtained, and (c) a reduction phase, using one or more microphone. Single-channel reduction is generally performed by means of spectral weighting, e.g., by the Wiener filter. Multi-channel reduction approaches commonly assume that the wind noise is spatially uncorrelated, and perform an additional spatial enhancement, e.g., using a multi-channel Wiener filter (MWF). The authors in [6] proposed to reduce wind noise using an MWF, exploiting a closed-form maximum-likelihood (ML) estimator of the speech and wind noise PSDs. Since spatially uncorrelated wind noise contributions and equal noise power among the microphone signals are assumed, the noise covariance matrix is given by a scaled identity matrix.

In [10], we showed how the spatial coherence of wind noise measured with closely-spaced microphones is non-zero and can be approximated by the Corcos model [11] as an exponential decay which depends on the microphone distance, the air stream direction and velocity. In this contribution, we propose a parametric multi-channel Wiener filter (PMWF) for wind noise reduction, designed for devices with closely-spaced microphones. In contrast to earlier works, we assume that the wind noise signals are spatially correlated. More specifically, we model the spatial coherence of wind noise contributions following the Corcos model. The proposed algorithm exploits (a) a joint estimation of the speech and wind noise PSDs by means of the minimization of the Frobenius norm of a composite error matrix, originally presented in [12] for de-reverberation and (b) a speech distortion/noise reduction trade-off parameter given by the so-called power ratio, proposed in [9] and extended in our previous work [13]. This reduction approach can be applied to largely-spaced microphones systems, by replacing the spatial coherence matrix based on the Corcos model with an identity matrix. In particular, the proposed estimator reduces to the ML estimator in [6] when we assume uncorrelated wind-noise in the estimation procedure.

The paper is structured as follows. In Section 2, we formulate the problem and introduce the notation. In Section 3, we describe the joint PSDs estimation. In Section 4, we describe the trade-off parameter of the PMWF. In Section 5, we assess the performance of the proposed approach in terms of speech quality and signal-to-noise ratio, and we make a comparison with the method in [6]. In Section 6, we draw conclusions and summarize this work.

2. PROBLEM FORMULATION

Without loss of generality, we assume an uniform linear array (ULA) of N microphones with an inter-microphone distance d. The i-th microphone signal, where $i \in \{1, 2, \ldots, N\}$, can be expressed in the short-time Fourier transform (STFT) domain as

$$Y_i(l,k) = G_i(k)S(l,k) + V_i(l,k), \tag{1}$$

where l and k denote the time frame and the frequency bin indices respectively, $S(l,k)$ denotes the speech signal, $V_i(l,k)$ denotes the i-

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

th wind noise contribution and $G_i(k)$ denotes the time-invariant relative transfer function from the speech source as received by the first microphone to the i-th microphone. For outdoor recordings, early reflections and late reverberation can be neglected, so that $G_i(k)$ is the relative direct-path transfer function, given by

$$G_i(k) = \exp\left(\frac{-\iota\,\omega_k d_{i1}\cos(\theta_\mathrm{s})}{c}\right), \tag{2}$$

where $\iota = \sqrt{-1}$, $\omega_k = 2\pi k F_s/K$ denotes the discrete angular frequency, K denotes the length of the discrete Fourier transform and F_s denotes the sampling frequency, $d_{i1} = d|i-1|$ denotes the relative distance between the i-th microphone and the first microphone that is used as a reference, θ_s denotes the speech direction of arrival (DOA) in terms of azimuthal angle and c denotes the speed of propagation of radiating acoustic sources. In the following, we assume that $S(l,k)$ and $V_i(l,k)$ are uncorrelated for all $i \in \{1,2,\ldots,N\}$. The wind noise contributions $V_{i,j}(l,k)$ exhibit a spatial coherence approximated by the Corcos model [10], i.e.,

$$\gamma_{ij}(k) = \exp\left(\frac{\omega_k d_{ij}[-\alpha(\theta_\mathrm{w}) + \iota\cos(\theta_\mathrm{w})]}{U_c}\right), \tag{3}$$

with $\gamma_{ij}(k) = \gamma^*_{ji}(k)$, where $d_{ij} = d|i-j|$, θ_w denotes the DOA of the wind stream with respect to the microphone axis, $\alpha(\theta_\mathrm{w})$ denotes a DOA-dependent decay rate parameter defined as

$$\alpha(\theta_\mathrm{w}) = \alpha_1|\cos(\theta_\mathrm{w})| + \alpha_2|\sin(\theta_\mathrm{w})|, \tag{4}$$

where α_1 and α_2 are, respectively, the longitudinal and the lateral coherence decay rates, experimentally determined in [14]. Finally, U_c is the convective turbulence speed in a boundary layer, where $U_c \approx 0.8U$, with U denoting the free-field wind stream velocity. For an air stream with constant DOA and speed, (3) is assumed to be time-invariant. The spatial coherence matrix of the wind noise contribution given by

$$\boldsymbol{\Gamma}(k) = \begin{bmatrix} \gamma_{11}(k) & \cdots & \gamma_{1N}(k) \\ \vdots & \ddots & \vdots \\ \gamma_{N1}(k) & \cdots & \gamma_{NN}(k) \end{bmatrix} \tag{5}$$

is therefore Hermitian, positive-definite and time-invariant. For a sufficiently large inter-microphone distance d, (5) can be approximated by the identity matrix, since (3) asymptotically converge to zero when d increases to infinity, so that the wind noise contributions can be assumed uncorrelated. The speech and free-field air stream DOAs (θ_s, θ_w) as well as the convective turbulence speed U_c are assumed to be known in the following. The microphone signals can be expressed in a vector notation as

$$\mathbf{y}(l,k) = \mathbf{g}(k)S(l,k) + \mathbf{v}(l,k), \tag{6}$$

where

$$\begin{aligned} \mathbf{y}(l,k) &= [Y_1(l,k), ..., Y_N(l,k)]^\mathrm{T} \\ \mathbf{g}(k) &= [G_1(k), ..., G_N(k)]^\mathrm{T} \\ \mathbf{v}(l,k) &= [V_1(l,k), ..., V_N(l,k)]^\mathrm{T}. \end{aligned}$$

The speech and the wind noise contributions are modelled as statistically independent zero-mean complex-Gaussian random processes, i.e.,

$$\begin{aligned} S(l,k) &\sim \mathcal{N}_c(0, \Phi_{ss}(l,k)) \\ \mathbf{v}(l,k) &\sim \mathcal{N}_c(0, \Phi_{vv}(l,k)\boldsymbol{\Gamma}(k)), \end{aligned}$$

where $\Phi_{ss}(l,k)$ denotes the speech PSD and $\Phi_{vv}(l,k)$ denotes the wind noise PSD. Furthermore, we assume equal noise power at each microphone, i.e.,

$$\Phi_{v_iv_i}(l,k) = \Phi_{v_jv_j}(l,k) = \Phi_{vv}(l,k).$$

The main goal is to obtain an estimate of the speech component $S(l,k)$, given the noisy observations vector $\mathbf{y}(l,k)$, through the complex weighting

$$\widehat{S}(l,k) = \mathbf{h}^\mathrm{H}(l,k)\mathbf{y}(l,k), \tag{7}$$

where $\mathbf{h}(l,k)$ is the PMWF that can be decomposed into [15]

$$\mathbf{h}(l,k) = \underbrace{\frac{\mathbf{g}^\mathrm{H}(k)\widehat{\boldsymbol{\Phi}}_v^{-1}(l,k)}{\mathbf{g}^\mathrm{H}(k)\widehat{\boldsymbol{\Phi}}_v^{-1}(l,k)\mathbf{g}(k)}}_{\mathbf{h}_\mathrm{MVDR}(l,k)} \underbrace{\frac{\widehat{\gamma}(l,k)}{\widehat{\gamma}(l,k)+\lambda}}_{H(l,k)}, \tag{8}$$

where $\mathbf{h}_\mathrm{MVDR}(l,k)$ denotes the minimum-variance distortionless-response (MVDR) beamformer and $H(l,k)$ denotes the parametric single-channel Wiener filter. The parameter λ allows us to control the trade-off between speech distortion and noise reduction. In (8), $\widehat{\boldsymbol{\Phi}}_v(l,k) = \widehat{\Phi}_{vv}(l,k)\boldsymbol{\Gamma}(k)$ denotes the short-term estimate of the noise covariance matrix and $\widehat{\gamma}(l,k) = \widehat{\Phi}_{ss}(l,k)/\widehat{\Phi}_{\tilde{v}\tilde{v}}(l,k)$ denotes the estimated a-priori signal-to-noise ratio (SNR) at the output of the beamforming, where $\widehat{\Phi}_{\tilde{v}\tilde{v}}(l,k) = [\mathbf{g}^\mathrm{H}(k)\widehat{\boldsymbol{\Phi}}_v^{-1}(l,k)\mathbf{g}(k)]^{-1}$. Here, $\widehat{\Phi}_{ss}(l,k)$ and $\widehat{\Phi}_{vv}(l,k)$ are the estimates of the speech and wind noise PSDs respectively, whose computation is the objective of the estimation phase. If $\lambda = 1$, (8) provides a minimum mean square error (MMSE) estimate of the speech, i.e., the MWF. To decrease the amount of residual noise, $\lambda > 1$ can be chosen. In this respect, the noise reduction is increased at the cost of possibly introducing speech distortion. In Section 4, we propose an approach to reduce wind noise while limiting speech distortions by making λ dependent on the microphone signals.

3. POWER SPECTRAL DENSITY ESTIMATION

The estimation approach described in this section was initially presented in [12], where the authors obtained a closed-form solution for the estimates of speech and late reverberant PSDs. Analogously in this work, the optimal set of speech and wind noise PSDs $\widehat{\Phi}(l,k) = [\widehat{\Phi}_{ss}(l,k)\ \widehat{\Phi}_{vv}(l,k)]^\mathrm{T}$ is obtained by minimizing the distance between the covariance matrix of the noisy observations

$$\widehat{\boldsymbol{\Phi}}_y(l,k) = \mathbb{E}\{\mathbf{y}(l,k)\mathbf{y}^\mathrm{H}(l,k)\}, \tag{9}$$

and the analytical model

$$\boldsymbol{\Phi}_y(l,k) = \Phi_{ss}(l,k)\mathbf{g}(k)\mathbf{g}^\mathrm{H}(k) + \Phi_{vv}(l,k)\boldsymbol{\Gamma}(k), \tag{10}$$

where $\mathbb{E}\{.\}$ denotes the expected value and $\boldsymbol{\Gamma}(k)$ given by (5). Here, $\widehat{\boldsymbol{\Phi}}_y(l,k)$ is recursively estimated, i.e.,

$$\widehat{\boldsymbol{\Phi}}_y(l,k) = \alpha\widehat{\boldsymbol{\Phi}}_y(l-1,k) + (1-\alpha)\mathbf{y}(l,k)\mathbf{y}^\mathrm{H}(l,k), \tag{11}$$

with $\alpha \in [0,1)$. We define the composite error matrix between (11) and (10) as the difference

$$\boldsymbol{\Phi}_e(l,k) = \widehat{\boldsymbol{\Phi}}_y(l,k) - \boldsymbol{\Phi}_y(l,k), \tag{12}$$

and we compute the optimal set $\widehat{\Phi}(l,k)$ as the minimizer of the squared Frobenius norm of (12), i.e.,

$$\widehat{\Phi}(l,k) = \underset{\Phi(l,k)}{\mathrm{argmin}}\|\boldsymbol{\Phi}_e(l,k)\|^2_\mathrm{F}. \tag{13}$$

The closed-form solution of (13) is given by [12]

$$\widehat{\Phi}_{ss}(l,k) = \frac{A_{22}(k)\mathrm{b}_1(l,k) - A_{12}(k)\mathrm{b}_2(l,k)}{A_{11}(k)A_{22}(k) - A_{12}^2(k)} \tag{14}$$

$$\widehat{\Phi}_{vv}(l,k) = \frac{A_{11}(k)\mathrm{b}_1(l,k) - A_{21}(k)\mathrm{b}_2(l,k)}{A_{11}(k)A_{22}(k) - A_{12}^2(k)}, \tag{15}$$

where $A_{i,j}(k)$ denote the entries of $\mathbf{A}(k) \in \mathbb{R}^{2\times 2}$ symmetric time-invariant matrix defined by [12]

$$\mathbf{A}(k) \equiv \begin{bmatrix} [\mathbf{g}^{\mathrm{H}}(k)\mathbf{g}(k)]^2 & \mathbf{g}^{\mathrm{H}}(k)\mathbf{\Gamma}(k)\mathbf{g}(k) \\ \mathbf{g}^{\mathrm{H}}(k)\mathbf{\Gamma}(k)\mathbf{g}(k) & \mathrm{Tr}[\mathbf{\Gamma}^{\mathrm{H}}(k)\mathbf{\Gamma}(k)] \end{bmatrix}, \tag{16}$$

and $b_i(l,k)$ denote the elements of the short-term vector $\mathbf{b}(l,k) \in \mathbb{R}^{2\times 1}$ defined by [12]

$$\mathbf{b}(l,k) \equiv \begin{bmatrix} \mathbf{g}^{\mathrm{H}}(k)\widehat{\mathbf{\Phi}}_y(l,k)\mathbf{g}(k) \\ \mathrm{Tr}[\widehat{\mathbf{\Phi}}_y(l,k)\mathbf{\Gamma}^{\mathrm{H}}(k)] \end{bmatrix}. \tag{17}$$

Assuming a sufficiently large microphone distance d, such that the spatial coherence in (3) can be assumed to be zero-valued for $i \neq j$, the spatial coherence matrix is defined by $\mathbf{\Gamma}(k) = \mathbf{I} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$, where $\mathbf{I}$ denotes the identity matrix. Under this assumption, it can be shown that (14) and (15) are equal to the ML estimators in [6] and therefore represent a more general solution.

4. SPEECH DISTORTION VS NOISE REDUCTION CONTROL

In our previous work [13], we derived the difference-signal power to the sum-signal power ratio of a mixture of speech and wind noise signals for a dual-microphone system, referred to as power ratio in the following. The obtained expression takes into account the spatial characteristics of wind noise measured using closely-spaced microphones, which are described by the Corcos model. In particular, the power ratio associated to speech signals yields values close to zero for a large portion of the frequency range, while the power ratio associated to wind noise tends towards one. Therefore, it serves as a reliable feature to discern between clean speech and wind noise. Given a sub-set of two microphone signals $Y_i(l,k)$, $Y_j(l,k)$, we define the difference signal and the sum signals by

$$Y_{\mathrm{diff}}(l,k) = Y_i(l,k) - Y_j(l,k),$$
$$Y_{\mathrm{sum}}(l,k) = Y_i(l,k) + Y_j(l,k),$$

respectively, and their PSDs by

$$\Phi_{\mathrm{diff}}(l,k) = \mathbb{E}\{|Y_{\mathrm{diff}}(l,k)|^2\},$$
$$\Phi_{\mathrm{sum}}(l,k) = \mathbb{E}\{|Y_{\mathrm{sum}}(l,k)|^2\}.$$

The power ratio is defined by

$$\mathrm{PR}(l,k) = \Phi_{\mathrm{diff}}(l,k)/\Phi_{\mathrm{sum}}(l,k). \tag{18}$$

Using (1) and (2), (18) can be written as

$$\mathrm{PR} = \frac{4\Phi_{ss}\sin^2\left(\frac{\omega_k d_{ij}\cos(\theta_{\mathrm{s}})}{2c}\right) + 2\Phi_{vv}\left[1 - \mathrm{Re}\{\gamma_{ij}\}\right]}{4\Phi_{ss}\cos^2\left(\frac{\omega_k d_{ij}\cos(\theta_{\mathrm{s}})}{2c}\right) + 2\Phi_{vv}\left[1 + \mathrm{Re}\{\gamma_{ij}\}\right]}, \tag{19}$$

where the time-frequency dependency is omitted for brevity, $\mathrm{Re}\{.\}$ is the real part operator and $\gamma_{ij}(k)$ is given by (3). It is possible to define the clean speech power ratio if $\Phi_{vv}(l,k) = 0$, obtaining

$$\mathrm{PR}_{\mathrm{s}}(l,k) = \tan^2\left(\frac{\omega_k d_{ij}\cos(\theta_{\mathrm{s}})}{2c}\right), \tag{20}$$

Algorithm: Multi-channel wind noise reduction

Compute $\mathbf{A}$ using (16);
for *all time frames* **do**
  for *all frequency bins* **do**
    Compute $\widehat{\mathrm{PR}}$ using (18);
    Compute $\widehat{\mathbf{\Phi}}_y$ using (11) and λ using (22) ;
    Compute $\mathbf{b}$ using (17);
    Compute $[\widehat{\Phi}_{ss}\ \widehat{\Phi}_{vv}]^{\mathrm{T}}$ using (14) and (15);
    Compute $\mathbf{h}$ using (8);
    Compute $\widehat{S} = \mathbf{h}^{\mathrm{H}}\mathbf{y}$;

which presents a periodic behaviour throughout the frequency spectrum if $\theta_{\mathrm{s}} \neq 90°$. However, if d_{ij} is sufficiently small, the power ratio of clean speech takes values close to zero for low frequencies, where most of the wind noise energy is preponderant. The pure wind noise power ratio is given by

$$\mathrm{PR}_{\mathrm{w}}(l,k) = \frac{1 - \exp\left(\frac{-\alpha(\theta_{\mathrm{w}})\omega_k d_{ij}}{U_c}\right)\cos\left(\frac{\omega_k d_{ij}\cos(\theta_{\mathrm{w}})}{U_c}\right)}{1 + \exp\left(\frac{-\alpha(\theta_{\mathrm{w}})\omega_k d_{ij}}{U_c}\right)\cos\left(\frac{\omega_k d_{ij}\cos(\theta_{\mathrm{w}})}{U_c}\right)}, \tag{21}$$

where, for increasing d_{ij}, it can be approximated by $\mathrm{PR}_{\mathrm{w}}(l,k) \approx 1$.

We propose to control the trade-off between speech distortion and noise reduction by adjusting $\lambda(l,k)$ in (8) using the power ratio as follows

$$\lambda(l,k) = 1 + \rho \cdot \widehat{\mathrm{PR}}(l,k) \tag{22}$$

where $\widehat{\mathrm{PR}}(l,k) = \widehat{\Phi}_{\mathrm{diff}}(l,k)/\widehat{\Phi}_{\mathrm{sum}}(l,k)$ is the power ratio computed by recursively estimating the power of the sum and difference-signal, i.e.,

$$\widehat{\Phi}_{\mathrm{diff}}(l,k) = \alpha\widehat{\Phi}_{\mathrm{diff}}(l-1,k) + (1-\alpha)|Y_{\mathrm{diff}}(l,k)|^2 \tag{23}$$

$$\widehat{\Phi}_{\mathrm{sum}}(l,k) = \alpha\widehat{\Phi}_{\mathrm{sum}}(l-1,k) + (1-\alpha)|Y_{\mathrm{sum}}(l,k)|^2. \tag{24}$$

This way, time-frequency instances distorted mainly by wind noise, for which $\widehat{\mathrm{PR}} \approx 1$, are more strongly attenuated by the PMWF than the MWF. For low-distorted or clean speech instances, for which $\widehat{\mathrm{PR}} \approx 0$, the noise attenuation of the PMWF is approximately equal to that of the MWF.

5. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed reduction approach in terms of objective speech quality and SNR improvement using PESQ [16] and frequency weighted segmental SNR (fwSNR) scores, respectively. We compared the obtained performance with the MWF proposed in [6] from two different experiments. In both experiments, ten different speakers (five male and five females) were randomly selected from the Libri speech Corpus [17] and each speaker was mixed with wind noise samples characterised by five different stream directions $\theta_{\mathrm{w}} = [0, 30°, 45°, 60°, 90°]$, with a defined input SNR level. The results were first averaged over the stream directions for each speaker and then over all the speakers.

The number of microphone used in every processing scheme was $N = 4$. The sampling frequency was 16 kHz, the frame length was 32 ms with 75% of overlap between consecutive frames and the smoothing parameter in (11), (23) and (24) was set to $\alpha = 0.6$. The

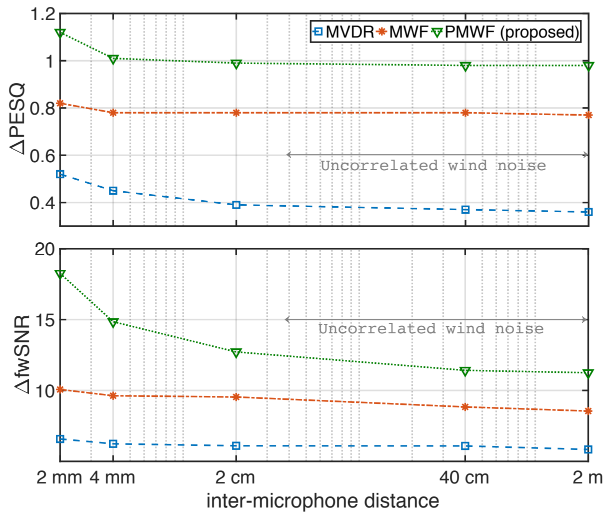

Fig. 1: PESQ and fwSNR improvements from for three processing scheme against increasing microphone distance (iSNR = 0 dB).

estimated a-priori signal-to-noise ratio $\widehat{\gamma}(l, k)$ in (8) was computed using the decision-directed approach, as explained in [12], with an averaging factor $\beta = 0.8$. The parameter ρ in (22) was set to 4.

5.1. Microphone distance dependency

We simulated wind noise using the method presented in [10] with an increasing inter-microphone distance d, air stream velocity U_c = 1.8 m/s and five different directions θ_{w}. In the range 2-40 mm the spatial coherence can be modelled by (3), while above approximately 40 mm we can assume spatially uncorrelated wind noise contributions. We convolved the speech signal with the direct path impulse responses and we mixed the signals with the generated wind noise to obtain a 0 dB input SNR. The speech was kept in the broad-side position ($\theta_{\mathrm{s}} = 90°$).

We compared the performance of three processing scheme, namely a) MVDR beamforming, b) MWF, and c) PMWF, i.e., our proposed method. The estimation phase was performed assuming the spatial coherence matrix as in (5) for every processing scheme. However, as described in Section 3, the speech and wind noise PSDs estimates given by (14) and (15) are equal to the estimates given by the method in [6] for sufficiently large microphone distances, i.e., above 40 mm the MWF approach is equal to the baseline [6] and the obtained performance improvements of the proposed PMWF are due to the trade-off parameter in (22) and not due to a more accurate PSD estimates.

In Fig. 1, the results from the described experiment are shown: the dashed green line depicts the performance of the proposed method which clearly outperforms the MVDR and the MWF for every microphone distance. In particular, analyzing the range 40 mm-2 m, the obtained improvements show that the proposed approach can be also used to reduce spatially uncorrelated wind noise.

5.2. SNR and model dependencies

We fixed the inter-microphone distance at $d = 4$ mm and mixed the convolved speech signals with wind noise samples at -10, 0, 10 dB. The parameters of the simulated wind noise were the same of the previous experiment. In addition, we included measured wind noise from the experiments described in [10], whose reduction was informed with the best fitting parameter of the Corcos model on the measured spatial coherence. The speech was kept in the end-fire position ($\theta_{\mathrm{s}} = 0°$). For each scheme, we computed the performance using two different spatial coherence matrices: the one defined in (5) and the identity matrix (uncorrelated wind noise). Our proposed method consists of the PMWF where the spatial coherence matrix is given by the Corcos model, while the baseline method [6] consists of the MWF where the spatial coherence is given by an identity matrix.

Table 1: PESQ and fwSNR improvements for the three processing scheme, two coherence matrix assumptions and three iSNR levels. Corcos = (5), **I** = identity matrix. Dark grey = proposed approach, light grey = baseline [6].

		-10 dB		0 dB		10 dB	
	Algorithm \ Coh. Matrix	Corcos	**I**	Corcos	**I**	Corcos	**I**
Δ_{PESQ}	MVDR	0.47	0.45	0.41	0.39	0.38	0.36
	MWF	0.78	0.62	0.76	0.59	0.75	0.56
	PMWF	**1.04**	0.83	**1.02**	0.78	**0.93**	0.74
Δ_{fwSNR}	MVDR	4.16	2.01	7.59	3.01	6.24	2.67
	MWF	12.18	5.17	12.34	6.47	8.84	5.68
	PMWF	**15.27**	8.34	**16.23**	9.92	**11.10**	8.69

In Table 1, the results from the second experiment are shown: in the case of closely-spaced microphones, the proposed method given by the combination of the PMWF and the Corcos model-coherence matrix presents the highest PESQ and fwSNR improvements for every level of the iSNR. Moreover, the PMWF with the trade-off parameter (22) outperforms the baseline method [6] also when spatially uncorrelated wind noise is assumed.

Although not shown here due to space constraints, it is interesting to note that the best performance was obtained for female speakers. This could be explained by the fact that female speakers commonly have a higher pitch than male speakers, and therefore the speech and wind noise are better separated in the STFT domain.

Audio examples and spectrograms can be found at
https://www.audiolabs-erlangen.de/resources/2018-ICASSP-WNR.

6. CONCLUSION

We presented a novel multi-channel approach aiming towards the reduction of wind noise in speech recorded with closely-spaced microphones. The noise reduction was performed using a PMWF. In contrast to the established assumption of spatially uncorrelated wind noise contributions, we exploited the spatial properties of wind noise approximated by a fluid dynamics model, namely the Corcos model. In particular, the Corcos model was used to jointly estimate the speech and wind noise PSDs, exploiting a recently developed closed-form solution. We then introduced an approach to control the trade-off between speech distortion and noise reduction. The experimental results showed that the proposed method outperforms a baseline method in terms of PESQ and signal-to-noise ratio improvements in different setups and under different assumptions. Moreover, we showed the robustness of the proposed approach for increasing microphone distance and hence spatially uncorrelated wind noise contributions.

7. REFERENCES

[1] C. Hofmann, T. Wolff, M. Buck, T. Haulick, and W. Kellermann, “A morphological approach to single-channel wind-noise suppression,” in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, 2012.

[2] E. Nemer, W. LeBlanc, M. Zad-Issa, and J. Thyssen, “Single microphone wind noise suppression,” Aug. 20 2013. US Patent 8,515,097.

[3] C. M. Nelke, N. Chatlani, C. Beaugeant, and P. Vary, “Single microphone wind noise psd estimation using signal centroids,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[4] C. M. Nelke and P. Vary, “Wind noise short term power spectrum estimation using pitch adaptive inverse binary masks,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[5] C. M. Nelke and P. Vary, “Dual microphone wind noise reduction by exploiting the complex coherence,” in *Proc. of the ITG Conference on Speech Communication*, 2014.

[6] P. Thuene and G. Enzner, “Maximum-likelihood approach to adaptive multichannel-Wiener postfiltering for wind-noise reduction,” in *Proc. of the ITG Conference on Speech Communication*, 2016.

[7] J. Park, J. Park, S. Lee, J. Kim, and M. Hahn, “Coherence-based dual microphone wind noise reduction by Wiener filtering,” in *Proc. of the 8th International Conference on Signal Processing Systems*, pp. 170–172, ACM, 2016.

[8] S. Franz and J. Bitzer, “Multi-channel algorithms for wind noise reduction and signal compensation in binaural hearing aids,” in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, 2010.

[9] G. W. Elko, J. M. Meyer, and T. F. Gaensler, “Noise-reducing directional microphone array,” Mar. 18 2016. US Patent App. 15/073,754.

[10] D. Mirabilii and E. A. P. Habets, “Simulating multi-channel wind noise based on Corcos model,” in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, 2018.

[11] G. Corcos, “The structure of the turbulent pressure field in boundary-layer flows,” *Journal of Fluid Mechanics*, vol. 18, no. 3, pp. 353–378, 1964.

[12] O. Schwartz, S. Gannot, and E. A. Habets, “Joint estimation of late reverberant and speech power spectral densities in noisy environments using frobenius norm,” in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2016.

[13] D. Mirabilii and E. A. P. Habets, “On the difference-to-sum power ratio of speech and wind noise based on the Corcos model,” in *Proc. IEEE Intl. Conf. on the Science of Electrical Engineering (ICSEE)*, 2018.

[14] R. H. Mellen, “On modeling convective turbulence,” *The Journal of the Acoustical Society of America*, vol. 88, no. 6, pp. 2891–2893, 1990.

[15] J. Benesty, J. Chen, and Y. Huang, “Noise reduction with multiple microphones: a unified treatment,” in *Microphone Array Signal Processing* (J. Benesty and Y. Huang, eds.), Springer Topics in Signal Processing, ch. 5, pp. 85–114, Berlin, Heidelberg: Springer-Verlag, 2008.

[16] ITU-T, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” Feb. 2001.

[17] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, “Librispeech: an asr corpus based on public domain audio books,” in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.