WIND NOISE SHORT TERM POWER SPECTRUM ESTIMATION USING PITCH ADAPTIVE INVERSE BINARY MASKS

Christoph M. Nelke, and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{nelke,vary}@ind.rwth-aachen.de

ABSTRACT

This paper presents a method to enhance a speech signal disturbed by wind noise. The wind noise is generated by turbulences in an air stream close to the microphone which picks up the desired speech signal. As the majority of speech enhancement algorithms works in the frequency domain, the short term power spectrum (STPS) of the unwanted noise must be estimated to reduce the wind noise. Conventional algorithms for background noise estimation fail in the case of wind noise due to its non-stationary characteristics. Hence, it is necessary to use special methods for the estimation and reduction of wind noise. The proposed system exploits the spectral characteristics of speech and noise to estimate the wind noise STPS. The spectral power distribution of wind noise and the pitch frequency of speech are used to generate a binary mask for the noise STPS estimation. This method is dependent on a precise pitch estimation. To reduce estimation errors a robust pitch estimation method using knowledge from prior estimates is presented. An evaluation and comparison with other wind noise reduction techniques shows improved speech enhancement of the proposed method.

***Index Terms*—** wind noise reduction, single microphone, pitch adaptive filtering, binary masks, speech enhancement

1. INTRODUCTION

Wind noise can severely degrade the speech quality and intelligibility as shown, e.g., in [1]. Since it is desired to use communication devices nearly everywhere, also outdoors, this might extremely disturb a conversation. As the design of mobile devices, such as mobile phones or hearing aids, is getting smaller and more space saving, the application of wind shields to prevent the noise generation acoustically is not feasible. This makes it necessary to remove wind noise in the recorded speech signal by means of signal processing. Most noise reduction systems work with a spectral weighting in the short-term DFT domain. For these systems the estimation of the noise is the most crucial part. In conventional systems, stationary or quasi-stationary noise signals are assumed and they aim to estimate the power spectral density (PSD) of the noise which is an averaged or smoothed version of the noise power in each frame. These algorithms are based on the assumption that stationary noise and speech can be separated by their temporal statistics (e.g. [2],[3],[4]). Thus they fail for the estimation of wind noise because of its fast changing signal level. The proposed method exploits the harmonic structure in terms of the pitch frequency as side information for the estimation of the wind noise power. Furthermore the characteristic spectral shape of wind noise is taken into account. In Sec. 2 the general structure of the speech enhancement system is presented. The proposed algorithm for the estimation of the wind noise STPS is explained in Sec. 3, where a modification for the enhanced pitch estimation (Sec. 3.4) and the concept of inverse binary masks (Sec. 3.1) are given. The performance of the proposed method is compared with other wind noise estimation methods using real recordings in Sec. 4.

1.1. Relation to prior works

For the estimation of the PSD of background noise a variety of algorithms exists. State-of-the-art conventional estimation methods using a single microphone are [2],[3],[4]. These algorithms all assume a rather stationary or at least slowly changing noise signal. Approaches especially designed for wind noise are given in [5],[6] and [7]. In [5] morphological operations are carried out to find connected areas in the time-frequency plane of the noisy signal to estimate the wind noise. The method proposed in [6] estimates the wind noise STPS with the combination of stored wind noise templates and information from the noisy signal. Our previously presented method ([7]) computes so-called signal centroids which reflects the spectral center-of-gravity to detect wind noise. An algorithm which also exploits the harmonic structure of speech to estimate non-stationary noise signals can be found in [8]. Methods for the reduction of wind noise using two or more microphones can be found in [9], [10] or [11]. These methods all take into account the correlation between the microphone signals.

2. SYSTEM OVERVIEW

The considered noise reduction system is depicted in Fig. 1. The noisy input signal $x(k)$ which is assumed to be a superposition of the clean speech $s(k)$ and wind noise $n(k)$ is first segmented and windowed using a square-root Hann window. Using the FFT, the short-term frequency representation $X(\lambda, \mu)$ with the frame index λ and the frequency bin μ is given. A spectral gain $G(\lambda, \mu)$ is applied to reduce the unwanted noise. As explained in Sec. 1, usually the PSD of the noise signal is estimated. In the case of wind noise any smoothing or averaging of the noise power is disadvantageous because of the fast changing noise signal. Therefore an estimate of the wind noise short term power spectrum (STPS) $\hat{\Phi}_N(\lambda, \mu)$ is used for the computation of the gains $G(\lambda, \mu)$. The estimation of $\hat{\Phi}_N(\lambda, \mu)$ from the noisy speech signal is the most crucial part of this system. Further details on the noise estimation and the determination of the required pitch frequency f_0 are given in Sec. 3.4 and Sec. 3.2. Finally, the enhanced signal is transformed back into the time domain and reconstructed via overlap-add using again a square-root Hann window yielding $\hat{s}(k)$.

978-1-4673-6997-8/15/$31.00 ©2015 IEEE

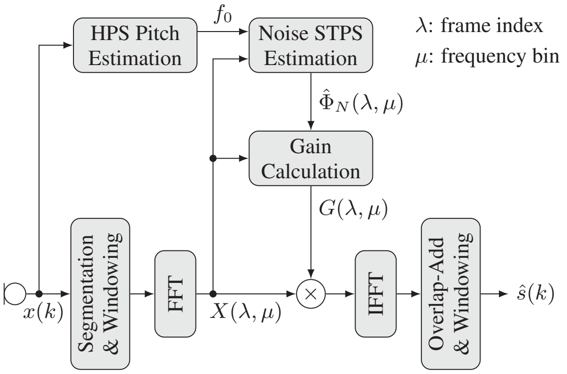

Fig. 1. Speech enhancement system

3. WIND NOISE ESTIMATION

Wind noise greatly differs from other typical noise signals (car, babble, traffic) in terms of its temporal and spectral characteristics. The acoustic signal of wind noise is generated by turbulences in an air stream close to the microphone and results in a non-stationary low-frequency noise signal [12]. For the computation of the spectral gains $G(\lambda,\mu)$ an estimate of the current noise STPS $\hat{\Phi}_n(\lambda,\mu)$ in each frame is required. Many conventional algorithms exploit that the noise signal statistics are changing slower than the desired speech signal. However, this is often not fulfilled for wind noise. As shown in [1], wind noise shows similar temporal statistics as speech signals. Therefore, wind noise reduction methods aim to exploit the spectral characteristics which greatly differ from speech signals. Due to the low frequency behaviour of wind noise, the main spectral overlap and thus the main degradation is given during voiced speech segments. In contrast to that, unvoiced speech has the main energy at higher frequencies ($>$ 2000 Hz) where wind noise has only a marginal influence. Thus the main task is to enhance voiced speech whereas for unvoiced speech a noise suppression with a simple high pass filter is sufficient as shown in [5]. It is also auxiliary that high-pass filtered wind noise and unvoiced speech have a similar sound. In this contribution we use the harmonic structure of voiced segments for the estimation of the noise STPS. In Sec. 3.1 the concept of inverse binary masks (IBM) is explained. In Sec. 3.2 and 3.3 the noise STPS estimation is shown and the necessary pitch estimation is presented in 3.4.

3.1. Inverse Binary Masks

Binary masks are usually used to separate speech and noise by applying a spectral gain

$$G_{\text{BM}}(\lambda,\mu)=\begin{cases}1, \text{if } |S(\lambda,\mu)|^2 > \text{LC}(\mu),\\ 0, \text{otherwise}\end{cases} \tag{1}$$

to the noisy spectrum $X(\lambda,\mu)$. The resulting output signal only contains parts where the speech power $|S(\lambda,\mu)|^2$ is higher than a certain local criterion LC(μ). This criterion is usually a certain threshold which might depend on the local SNR. Applying an ideal binary mask can improve the intelligibility or the performance of an automatic speech recognition system (e.g., [13] and references therein).

Normally, binary masks completely cancel out parts of the undesired noise signal. This leads to a sufficient but also aggressive noise suppression which may introduce unwanted artefacts to the output signal. Furthermore, due to the binary gain of the mask based processing follows, that the noise is not reduced in time-frequency units where both speech and noise are active. This residual noise also results in annoying effects in the output signal.

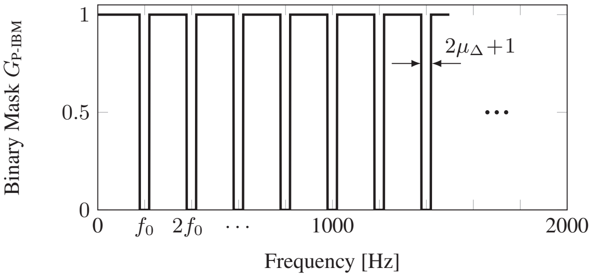

Fig. 2. Pitch adaptive inverse binary mask

The task of binary masks in this paper is different to this common application. As aforementioned, the main objective is to enhance voiced speech segments. Therefore, we introduce the pitch adaptive inverse binary mask (P-IBM) as shown in Fig. 2. The aim is to cancel out the voiced speech segments in the time-frequency plane by applying the P-IBM to the noisy signal. This means that the binary mask is defined as follows

$$G_{\text{P-IBM}}(\lambda,\mu)=\begin{cases}0, \text{if } \mu\in \text{M}\\ 1, \text{otherwise.}\end{cases} \tag{2}$$

with

$$\text{M}=\{\cup_{\kappa\in\mathbb{N}}[\kappa\cdot\mu_0-\mu_\Delta,\ldots,\kappa\cdot\mu_0,\ldots\kappa\cdot\mu_0+\mu_\Delta]\} \tag{3}$$

and μ_0 depicts the discrete frequency bin corresponding to the pitch frequency f_0. μ_Δ determines a frequency range around the pitch bin to ensure the cancellation of the speech signal by the P-IBM. An estimated speech-free spectrum is then given by

$$\tilde{X}(\lambda,\mu)=G_{\text{P-IBM}}(\lambda,\mu)\cdot X(\lambda,\mu) \tag{4}$$

in which the speech components are set to zero and is used in the following for the wind noise STPS estimation.

3.2. Noise Estimation

For the required noise STPS of the wind signal, the speech-free spectrum $\tilde{X}(\lambda,\mu)$ as computed in Eq. 4 is considered. In the frequency bins which are not set to zero, the noisy speech signal reveals directly the wind noise spectrum between the multiples of its pitch frequency. Now the remaining parts $\kappa\cdot\mu_0-\mu_\Delta\ldots\kappa\cdot\mu_0+\mu_\Delta$ of $\tilde{X}(\lambda,\mu)$ which were set to zero by the binary mask are linearly interpolated according to

$$|\hat{N}(\lambda,\mu)|^2=\begin{cases}|\tilde{X}_{\text{inter}}(\lambda,\mu)|^2 & \text{, if } \mu\in \text{M}\\ |\tilde{X}(\lambda,\mu)|^2 & \text{, otherwise}\end{cases} \tag{5}$$

with linear interpolation $\tilde{X}_{\text{inter}}(\lambda,\mu)$ between the adjacent spectral bins $\tilde{X}(\lambda,\kappa\cdot\mu_0-\mu_\Delta-1)$ and $\tilde{X}(\lambda,\kappa\cdot\mu_0+\mu_\Delta+1)$. Investigations with other interpolation types showed no improvements, therefore the linear interpolation is used in the following. The P-IBM can only be applied in frames with voiced speech activity where the pitch structure defines the binary mask. In unvoiced segments or segments with no speech activity the wind noise can be directly estimated from the lower frequency range of the input signal. A detection of wind and voiced speech segments can be realized by the so-called signal centroids (SC) measured in Hz

$$\text{SC}(\lambda) = \frac{f_s}{N} \frac{\sum_{\mu=1}^{L} \mu \cdot |X(\lambda,\mu)|^2}{\sum_{\mu=1}^{L} |X(\lambda,\mu)|^2}, \tag{6}$$

where L depicts the frequency range for the computation, f_s the sampling frequency and N the FFT size. The frequency range for the SC computation was set to 0. . . 2000 Hz, in which voiced speech is assumed to be active. The SCs are the *centres-of-gravity* in the spectrum and thus they depict the spectral power distribution. In [7] it was shown that the SCs of voiced speech are between 300-800 Hz and for wind noise clearly below 50 Hz and the SCs were exploited to estimate the SNR in the current frame. The SC of voiced speech is dependent on the pitch and the spectral envelope which determines the spectral power distribution. Here we apply a threshold th_{SC} = 85 Hz to determine segments of pure wind ($\text{SC}(\lambda) < th_{\text{SC}}$). In these frames the noise estimate is set to the input signal $|\hat{N}(\lambda,\mu)|^2 = |X(\lambda,\mu)|^2$. This also removes the wind noise in unvoiced segments where no speech power is given in the lower frequency range. The reliability check proposed in Sec. 3.3 ensures that the higher frequencies are protected. The rather low value of the threshold th_{SC} guarantees a low misdetection rate of voiced speech segments in order to protect the desired speech signal. Many approaches for noise estimation apply recursive smoothing of the STPS to compute the required noise PSD estimate (e.g., [4]). Due to the non-stationary characteristics of wind noise a subsequent smoothing would lower the adaptation speed of the estimator. Therefore, the STPS estimate is directly set to $\hat{\Phi}_N(\lambda,\mu) = |\hat{N}(\lambda,\mu)|^2$.

3.3. Reliability Check

As shown in many publications the wind noise signal is characterized by a low frequency energy distribution (e.g., [12] or references therein). In order to prevent an overestimation of the wind noise for higher frequencies a reliability check is performed. In [7] it was shown that the spectrum of wind noise can be approximated by an $1/f$ slope over the frequencies f. Therefore, the noise STPS estimate is limited at higher frequencies ($\mu > \mu_{\text{rel}}$) by an $1/f^2$ slope starting from the averaged power $\sigma^2_{N,\text{low}}(\lambda)$ in the lower band ($\mu < \mu_{\text{rel}}$) of the noise estimate from Eq. 5

$$\hat{\Phi}_N(\lambda,\mu) = \min\left\{\hat{\Phi}_N(\lambda,\mu), \sigma^2_{N,\text{low}}(\lambda)/\mu^2\right\} \text{ for } \mu > \mu_{\text{rel}}. \tag{7}$$

In addition to this upper limit of the noise estimate, the STPS $\hat{\Phi}_N(\lambda,\mu)$ is set to zero in frames where the SC indicates no wind activity ($\text{SC}(\lambda) > 800\,\text{Hz}$).

3.4. Improved Harmonic Product Pitch Estimation

For the determination of $G_{\text{P-IBM}}(\lambda,\mu)$ the current pitch frequency is required. Therefore the pitch frequency is estimated in each frame. In [14] an evaluation of several pitch estimation algorithms in terms of their robustness to wind noise was carried out. It turned out that methods working in the cepstral or frequency domain achieve the best results. For the proposed system the Harmonic Product Spectrum (HPS) was chosen as pitch estimator ([15]):

$$\tilde{\mu}_0(\lambda) = \arg\max_{\mu} \left\{ \frac{\prod_{l=1}^{M_H} |X(\lambda, l\cdot\mu)|}{\prod_{l=1}^{M_H} |X(\lambda, l\cdot\mu + [\mu/2])|} \right\}, \tag{8}$$

where $[l]$ denotes the closest natural number to l and M_H is the number of considered harmonics. In [16] Eq. 8 was used to compute the pitch frequency of band-limited speech, where the frequencies below 300 Hz are completely missing. It turned out that in the case of wind noise, where mainly the lower frequencies are corrupted, the HPS also gives quite good results for the pitch estimation.

However, in frames where strong wind noise is active the pitch estimation might fail. Because the noise STPS estimation from Sec. 3.2 is sensitive to pitch estimation errors, a post-processing as presented in the following lowers the number of estimation errors. The pitch of human speech is speaker dependent and shows a strong temporal correlation of adjacent frames (see, e.g., [17]). It is favourable to use this characteristic for a post-processing of the individual pitch estimates. A simple approach would be a smoothing of the estimates to lower the variance. The special temporal behaviour of wind noise noise leads to a specific kind of errors. During a sudden rise of the wind signal power, the pitch estimation mainly fails in single or only a few consecutive frames. The post-processing step proposed for our system is implemented as a buffer $\mathbf{B}_{\mu_0}(\lambda)$ storing the last K pitch estimates

$$\mathbf{B}_{\mu_0}(\lambda) = [\hat{\mu}_0(\lambda-1), \hat{\mu}_0(\lambda-2), \ldots, \hat{\mu}_0(\lambda-K)]. \tag{9}$$

A low variance of the stored estimates in $\mathbf{B}_{\mu_0}(\lambda)$ is given for the aforementioned strong temporal correlation of the pitch frequency and thus indicates correct estimates. For the given system a reliable pitch buffer is assumed if the standard deviation (STD) within the buffer is smaller than 50% of mean pitch $\bar{\mu}_0$ of the buffer which leads to

$$\frac{\text{STD}\left\{\mathbf{B}_{\mu_0}(\lambda)\right\}}{\bar{\mu}_0} < 0.5. \tag{10}$$

In this case the final pitch estimate is given by comparing the estimate from the current frame $\tilde{\mu}_0(\lambda)$ with the stored values. Here, also a deviation of 50% to the mean value is tolerated

$$\hat{\mu}_0(\lambda) = \begin{cases} \tilde{\mu}_0(\lambda) & \text{, if } \bar{\mu}_0 \cdot 0.5 < \tilde{\mu}_0(\lambda) < \bar{\mu}_0/0.5 \\ \hat{\mu}_0(\lambda-1) & \text{, otherwise,} \end{cases} \tag{11}$$

otherwise the last reliable pitch estimate is taken. The buffer processing could be tuned to be more aggressive by applying a lower threshold than 50%, but it turned out that a great amount of pitch errors can be corrected with this setting. The pitch buffer is updated only in segments with speech activity and if the first condition of Eq. 11 is fulfilled. This post-processing also reduces doubling errors which commonly appear for all pitch estimation methods. Voiced speech segments in the noisy input signal are given for higher values of the signal centroid ($\text{SC}(\lambda) > th_{\text{SC}}$) as explained in Sec. 3.2. In segments with no voiced speech and no reliable pitch estimate (second case of Eq. 11) the buffer is emptied again to prevent an erroneous correction after speech pauses where changes of the pitch frequency are common. In addition the post-processing is only applied if the buffer is at least 50% filled.

Due to the limited frequency resolution in the DFT domain the pitch bin μ_0 can only take integer values. Therefore, it is recommended to check for every multiple of the pitch frequency $\kappa \cdot \mu_0$

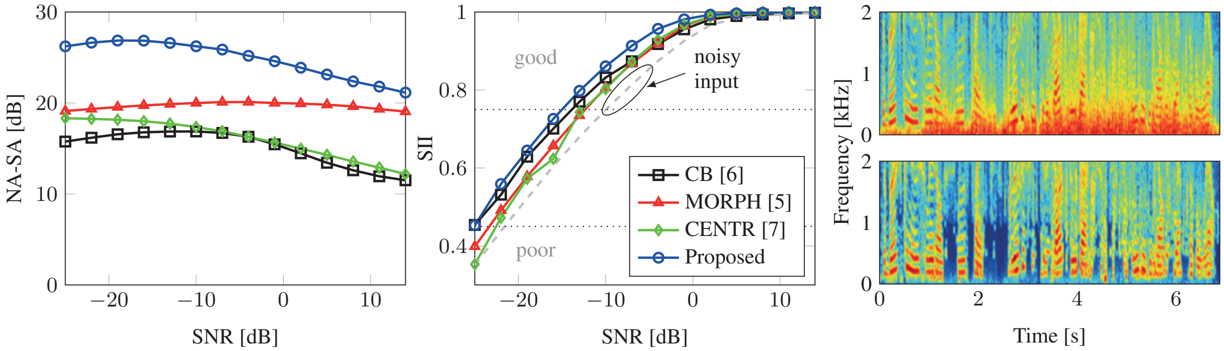

Fig. 3. Results: *left*: segmental noise attenuation - speech attenuation (NA-SA); *middle*: Speech Intelligibility Index (SII); *right*: noisy (SNR = -5 dB, top) and enhanced speech (bottom) by proposed system

in Eq. 3 if it corresponds to a local maximum by comparing it to its adjacent bins and correct the pitch bin if necessary.

4. EVALUATION

4.1. Experimental Setup

Table 1. Simulation parameters

sampling frequency	f_s = 16 kHz
frame-length	40 ms
window	$\sqrt{\text{Hann}}$
overlap	50%
FFT size	N = 1024 (incl. zero-padding)
number of pitch harm.	M_H = 8
pitch search range	80 - 450 Hz
pitch buffer size	K = 10 ($\hat{=}$ 200 ms)
pitch width	$\mu_\Delta \hat{=}$ 30 Hz
centroid threshold	th_{SC} = 85 Hz
reliabilty check threshold	$\mu_{\text{rel}} \hat{=}$ 500 Hz

The performance of the proposed system was evaluated with wind noise recordings from [1]. These signals were recorded outdoors and reflect realistic settings for, e.g., a phone-call scenario. From both the the *strong wind* and the *normal wind* example 240 seconds were taken[1] and mixed with random sentences from the TSP database [18] w.r.t. to realistic SNR scenarios for wind noise. The proposed system was compared to three other approaches especially designed for the estimation of wind noise: the codebook based method (CB) from [6], the morphological approach (MORPH) from [5] and the simpler centroid based wind estimation (CENTR) we proposed in [7]. The parameters of the whole noise reduction framework are given in Tab. 1. For the final reduction fast adapting spectral gains $G(\lambda, \mu)$ are required which precludes any kind of smoothing of the gains. For wind noise reduction the spectral subtraction gain rule [19] showed the best performance and was used for all simulations. The estimation of the pitch frequency requires a larger frame-size than 40 ms for sufficient results (see, e.g., [20]). Therefore frames of 90 ms are considered for the used HPS method using the same frame shift as for the noise reduction (20 ms).

4.2. Results

All algorithms are evaluated in terms of their speech enhancement. The noise reduction performance is determined by means of the segmental noise attenuation (NA) minus segmental speech attenuation (SA) measure (e.g., [21]), where an improvement results in higher values (Fig. 3 left plot). In addition the Speech Intelligibility Index (SII) [22] is applied as measure (Fig. 3 middle plot). The SII provides a value between 0 and 1 where a SII higher than 0.75 indicates a good communication system and values below 0.45 correspond to a poor system. The results of the simulations are shown in Fig. 3 where the gray dashed line depicts the SII of the noisy unprocessed signal. It can be seen that over the whole considered SNR range (-25. . . 15 dB) the proposed system shows the highest performance for both measures. The utmost right plots of Fig. 3 exemplifies the performance of the proposed method in the shown spectrograms of a noisy signal (SNR = -5 dB) and the enhanced output signal of the proposed system. Informal listening test confirmed the results whereas all algorithms can produce some high-pass effects to the speech, in particular in segments with high wind energy. Due to the pitch adaptive processing, the presented method is capable to preserve more of the desired speech signal. Investigations with conventional noise reduction system including, e.g., [4] showed no or only marginal improvements in terms of wind noise reduction.

5. CONCLUSIONS

In this contribution a single microphone method for the estimation of wind noise STPS is presented. Because of the non-stationary temporal behaviour of wind noise, the spectral characteristics of speech and wind were exploited. Applying an inverse binary masked controlled by the pitch frequency of the speech signal leads to a sufficient noise STPS estimate. In this context a post-processing of the pitch estimation is given which lowers the number of estimation errors. Evaluation with real wind noise recordings shows that the proposed system can efficiently remove the noise with a higher performance than other wind noise reduction schemes.

[1] Link for downloading corresponding noise signals provided in [1]

6. REFERENCES

[1] C.M. Nelke and P. Vary, "Measurement, analysis and simulation of wind noise signals for mobile communication devices," in *Proc. of Intern. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sophia-Antipolis, France, September 2014, Proc. of Intern. Workshop on Acoustic Echo and Noise Control (IWAENC).

[2] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.

[3] R.C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Dallas, Texas, USA, 2010.

[4] T. Gerkmann and R. Hendriks, "Noise power estimation based on the probability of speech presence," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 2011.

[5] C. Hofmann, T. Wolff, M. Buck, T. Haulick, and W. Kellermann, "A morphological approach to single-channel wind-noise suppression," in *Proc. of Intern. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aachen, Germany, Sept. 2012.

[6] S. Kuroiwa, Y. Mori, S. Tsuge, M. Takashina, and F. Ren, "Wind noise reduction method for speech recording using multiple noise templates and observed spectrum fine structure," in *Intern. Conf. on Communication Technology*, Guilin, China, 2006.

[7] C.M. Nelke, N. Chatlani, C. Beaugeant, and P. Vary, "Single microphone wind noise PSD estimation using signal centroids," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Florence, Italy, May 2014.

[8] D. Ealey, H. Kelleher, and D. Pearce, "Harmonic tunnelling: tracking non-stationary noises during speech," in *INTERSPEECH*, Aalborg, Denmark, September 2001, pp. 437–440.

[9] G.W. Elko, "Reducing noise in audio systems," Patent US7171008, 2007.

[10] C.M. Nelke and P. Vary, "Dual microphone wind noise reduction by exploiting the complex coherence," in *ITG-Fachtagung Sprachkommunikation*, Erlangen, Germany, September 2014.

[11] S. Franz and J. Bitzer, "Multi-channel algorithms for wind noise reduction and signal compensation in binaural hearing aids," in *Proc. of Intern. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, 2010.

[12] S. Bradley, T. Wu, S.v. Hünerbein, and J. Backman, "The mechanisms creating wind noise in microphones," in *Audio Engineering Society, 114th Convention*, 2003.

[13] S. Gonzalez and M. Brookes, "Mask-based enhancement for very low quality speech," in *Proc. of IEEE Intern. Conf. on Acoustics, Speech, and Signal Process. (ICASSP)*, Florence, Italy, May 2014.

[14] C.M. Nelke, N. Nawroth, M. Jeub, C. Beaugeant, and P. Vary, "Single microphone wind noise reduction using techniques of artificial bandwidth extension," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Bucharest, Romania, August 2012.

[15] A. Noll, "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum and a maximum likelihood estimate," *Proc. of the Symposium on Computer Processing in Communications*, vol. 14, pp. 779–797, 1970.

[16] Bernd Iser, Gerhard Schmidt, and Wolfgang Minker, *Bandwidth extension of speech signals*, vol. 13, Springer, 2008.

[17] P. Vary and R. Martin, *Digital Speech Transmission. Enhancement, Coding and Error Concealment*, Wiley-VCH Verlag, 2006.

[18] P. Kabal, "TSP speech database," Tech. Rep., McGill University, Montreal, Canada, 2002.

[19] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113 – 120, 1979.

[20] S. Gonzalez and M. Brookes, "A pitch estimation filter robust to high levels of noise (PEFAC)," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, Barcelona, Spain, 2011.

[21] S. Gustafsson, R. Martin, P. Jax, and P. Vary, "A psychoacoustic approach to combined acoustic echo cancellation and noise reduction," *IEEE Trans. Speech and Audio Process.*, vol. 10, no. 5, pp. 245–256, July 2002.

[22] ANSI S3.5-1997, "Methods for the calculation of the speech intelligibility index," 1997.