Dual Microphone Wind Noise Reduction by Exploiting the Complex Coherence

Christoph Matthias Nelke and Peter Vary

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

Abstract

This contribution presents methods for detecting, estimating and reducing the effect of wind noise picked up by a dual microphone communication device. To discriminate the sound induced by the wind stream the complex coherence properties of the two microphone signals are exploited. Therefore, a coherence analysis of recorded signals is given. Based on the phase variance of the complex coherence function a wind noise detection mechanism is presented. Furthermore, an approach for the enhancement of distorted speech is developed, which uses the phase variance and the magnitude of the complex cross power spectral density (PSD). This stage is realized by a noise PSD estimation and a dedicated spectral subtraction weighting. The reduction of wind noise by means of spectral weighting often leads to severe degradations to the speech signal. However, the evaluation of the proposed algorithm demonstrates a better performance compared to related approaches.

1 Introduction

Nowadays,the hands-free mode is used in many in communication devices, e.g., mobile phones or video conference equipment. While the hands-free scenario has many advantages for the near-end speaker, the far-end participant often experiences a lower audio quality because of the higher levels of background noise sources. This becomes even more severe in the case of wind noise, mainly because of two reasons: wind noise is generated by turbulences in the boundary layer around the used device and thus might be inaudible for the near-end speaker. Furthermore, the wind stream might be much stronger compared to a more shielded hand-held position. The wind stream generates a low frequency instationary distortion which calls for an enhancement of the recorded signals. Because of the non-stationary characteristics, state-of-the-art noise estimation algorithms as [1] fail to produce a sufficiently precise noise PSD estimate. Several single microphone methods exist, which are especially designed for the estimation of wind noise PSD (e.g., [2] and references therein). In this contribution we propose a dual microphone setup for the detection and the estimation of wind noise in a speech signal. Here, the different complex coherence properties of speech and noise are exploited which leads to a sufficient detection rate and good estimation results of the wind noise PSD.

2 Dual Microphone Arrangement

This section presents the coherence properties of the input signals (Sec. 2.1) and the noise reduction system (Sec. 2.2) which can be applied to several hands-free scenarios.

2.1 Signal Statistics

Some noise reduction algorithms exploit the magnitude coherence properties of the desired speech and noise signals for the enhancement of the desired signal. In many acoustic environments the background noise can be assumed to be diffuse, leading to a certain coherence model. The coherence is commonly determined by the ratio of the auto- and cross-PSDs $\Phi_{x_1x_2}(\lambda,\mu)$ and $\Phi_{x_1x_1}(\lambda,\mu)$, $\Phi_{x_2x_2}(\lambda,\mu)$ as normalized cross PSD

$$\Gamma(\lambda,\mu) = \frac{\Phi_{x_1x_2}(\lambda,\mu)}{\sqrt{\Phi_{x_2x_2}(\lambda,\mu)\Phi_{x_2x_2}(\lambda,\mu)}}. \quad (1)$$

A diffuse noise field is often observed by noise sources in the far-field, leading to a frequency dependent coherence function of the receiving microphone signals. In contrast to that wind noise is directly produced by turbulences in a boundary layer close to the microphones. It was shown that the noise field of a wind stream can be created by an arrangement of elementary emitters with monopole, dipole and quadrupole characters [3]. Since these emitters are independent from each other, the noise signals at the two microphone positions are also uncorrelated. In [4] a model is given which predicts low coherence over the whole frequency range. For a coherence analysis measurements where carried out using mock-up phones equipped with two microphones with a distance of 2 or 10 cm. The results are given in Fig. 1 in terms of the magnitude squared coherence (MSC). In the magnified section it can be seen that the coherence of wind is close to zero over the whole frequency range. Additionally, the MSC of corresponding speech signals played by an artificial head (HEAD acoustics HMS II.3) in the hand-free position [5] are depicted by the dashed lines which shows values close to one. A clear distinction is possible from this investigations. In the following we assume a microphone distance of 10 cm because this is commonly used microphone configuration in mobile phones. Nevertheless, all considerations are also applicable to smaller microphone distances.

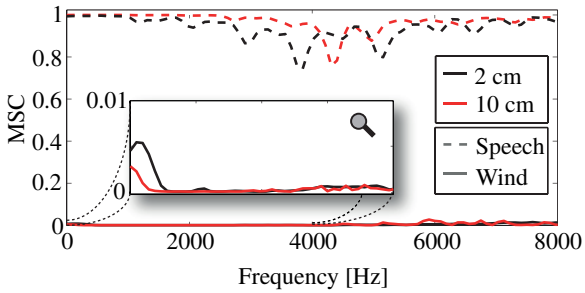

Figure 1: Coherence of Speech and Wind Noise

2.2 Noise Reduction System

The dual microphone system is shown in Fig. 2. The noisy input signals $x_{1|2}(k)$ sampled at a rate of 16 kHz are assumed to be a superposition of the clean speech signals and the wind noise signals. The delay compensation of the

speech signal is not scope of this paper, here we assume aligned speech signals as input for the investigated system. Both input signals are segmented into 20 ms frames with 10 ms overlap using a Hann window and transformed into the frequency domain with a FFT size of 512 (including zero-padding). The spectra $X_{1|2}(\lambda,\mu)$ are used for the detection of speech and wind noise and the noise PSD estimation (see Sec. 3.1 and 3.2). Based on the wind noise indicator $\overline{\sigma}_\varphi(\lambda)$ and the noise PSD estimate $\widehat{\Phi}_N(\lambda,\mu)$ a spectral gain $G(\lambda,\mu)$ is computed (see Sec. 4) and applied to the noisy input. Finally, the time domain output $\widehat{s}(k)$ is synthesized of the enhanced signal $\widehat{S}(\lambda,\mu)$ via IFFT and overlap-add.

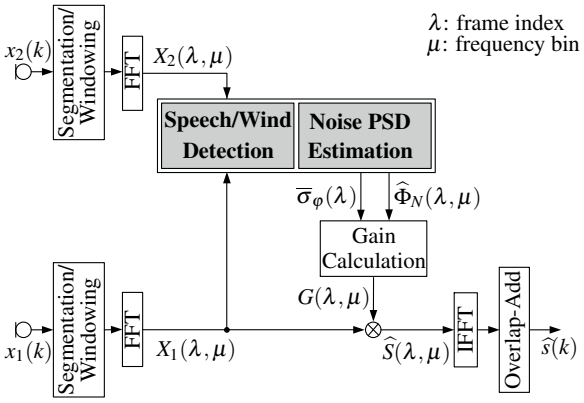

Figure 2: Speech enhancement system

3 Wind Noise Estimation

For the illustration in Fig. 1 the PSDs in (1) are estimated by averaging the auto and cross periodograms of 20 ms frames from 10 seconds audio recordings. For real-time applications the required PSDs for the MSC determination are normally computed via smoothed periodograms

$$\Phi_{x_i x_j}(\lambda,\mu) = \alpha_S \Phi_{x_i x_j}(\lambda-1,\mu) + (1-\alpha_S) X_i(\lambda,\mu) X_j^*(\lambda,\mu), \quad (2)$$

where $\{\}^*$ denotes the complex conjugate. Here the smoothing constant α_S presents a conflict between a sufficient estimation accuracy and a fast adaption to changing coherence properties. Especially, for wind noise the latter aspect is of great importance because of the non-stationary signal characteristics. This requires a small value of α_S. Contrary, a value for α_S closer to zero results in many estimation errors (MSC = 1 in extreme case of $\alpha_S = 0$). In general, coherence given in (1) is complex-valued with the magnitude as shown in Fig. 1 and a phase. It is obvious that even in the case of $\alpha_S = 0$ the phase

$$\varphi(\lambda,\mu) = \angle\{\Phi_{x_1 x_2}(\lambda,\mu)\} = \angle\{X_1(\lambda,\mu)\} - \angle\{X_2(\lambda,\mu)\} \quad (3)$$

provides information about the input signals. For the detection and estimation of wind noise both the magnitude and the phase of the complex coherence are exploited in the following.

3.1 Speech & Wind Detection

As mentioned above, the method for detecting wind noise and the desired speech signal must be capable to track the fast changes of the wind noise. Fig. 3 illustrates the complex coherence of a noisy dual channel signal. Here the smoothing constant was chosen to $\alpha_S = 0.5$. The previously stated effect to the magnitude of the coherence is

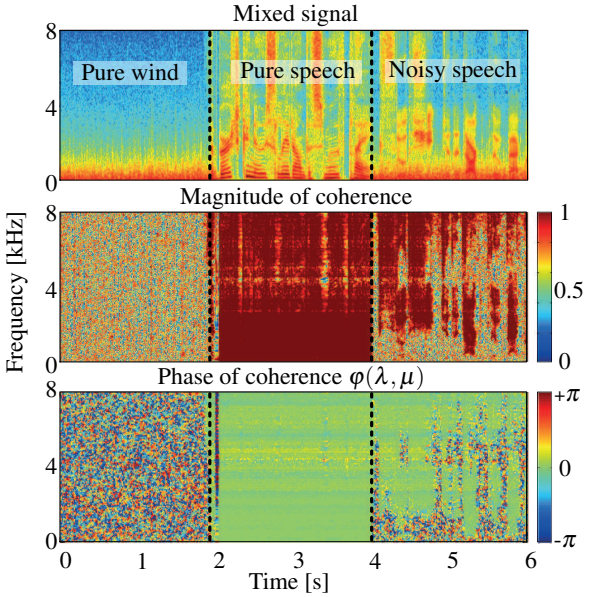

Figure 3: Noisy speech signal and complex coherence

clearly visible in the middle plot. Also in segments where only wind is active ($t = 0...2$ s), high coherent parts appear depicted by several red spots. The bottom plot illustrates that in segments where speech is active ($t = 2...4$ s) the phase $\varphi(\lambda,\mu)$ takes a constant value close to 0. This is because of the constraint of aligned input signals. For not aligned signals, the phase would show a linear decreasing or increasing behavior with the frequency. A quite different behavior can be observed in segments where wind noise is active and the phase results in a randomly distributed noise between $-\pi$ and π. In the last segment ($t = 4...6$ s) the degree of disturbance to the speech signal is visible. The phase as given in (3) of a superposition of a speech spectrum S and a noise spectrum N can be expressed as

$$\varphi = \arctan\Big(\frac{|S||N|(\sin(\varphi_{s1}-\varphi_{n2}) + \sin(\varphi_{n1}-\varphi_{s2}))}{|S|^2 + |S||N|(\cos(\varphi_{s1}-\varphi_{n2}) + \cos(\varphi_{n1}-\varphi_{s2}))} \quad \ldots \quad \frac{+|N|^2 \sin(\varphi_{n1}-\varphi_{n2})}{+|N|^2 \cos(\varphi_{n1}-\varphi_{n2})}\Big). \quad (4)$$

For the sake of brevity the frequency and time indices are omitted in this equation. A direct relation between the SNR and the phase is not possible since the phases of speech signals $\varphi_{s1|2}$ and noise signals $\varphi_{n1|2}$ are randomly distributed. However, it can be seen that in the case of pure wind noise ($S = 0$) φ takes the value of the phase difference of noise signals

$$\varphi_{\text{wind}} = \varphi_{n1} - \varphi_{n2} \quad (5)$$

and in the case of clean speech ($N = 0$)

$$\varphi_{\text{speech}} = 0. \quad (6)$$

A single phase value of the coherence is not meaningful for the degree of distortion, therefore the distribution of the phase within a time period (e.g., one frame) is investigated in the following. The distributions of the phase measured over a segment of 2 seconds is depicted in Fig. 4 derived from clean speech and wind noise signals. It is obvious that the phase of speech results in peak at $\varphi(\lambda,\mu) = 0°$ whereas the phase of wind signal is nearly uniformly distributed in the interval $-\pi...\pi$. A measurable quantity for

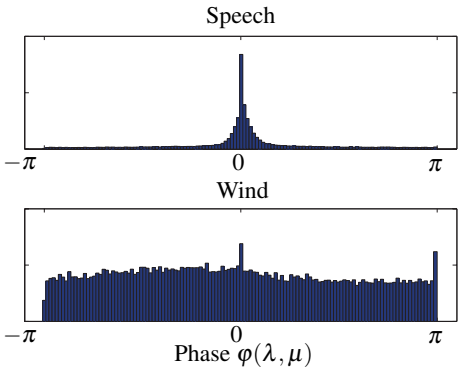

Figure 4: Phase distribution of complex coherence

the distribution of a signal is the variance. Thus we introduce the normalized phase variance given by

$$\overline{\sigma}_{\varphi}(\lambda) = \frac{3}{\pi^2} \sum_{\mu=\mu_1}^{\mu_2} \frac{\varphi(\lambda,\mu)^2}{\mu_2 - \mu_1}, \tag{7}$$

with μ_1 and μ_2 determining the frequency range for the variance computation. The variance of a constant value (as the phase of the speech signals) becomes 0. For an uniform distribution between $\pm\pi$ (as in the case of wind noise) it takes the value $\pi^2/3$. Thus (7) is normalized by this factor leading to values between 0 and 1. In the spectrogram in the top plot in Fig. 3, it is evident that the wind noise merely disturbs the lower frequency range. To provide an indicator for speech and wind noise the phase variance in two frequency ranges is determined. $\overline{\sigma}_{\varphi,l}(\lambda)$ is defined for wind noise detection and $\overline{\sigma}_{\varphi,h}(\lambda)$ to capture speech activity. The ranges in (7) are set to $\mu_1 = 1|129$ and $\mu_2 = 128|257$ for wind detection and speech detection, respectively.

3.2 Noise PSD Estimation

For the estimation of the noise PSD the principle presented in [6] is modified to work with the special constraints of wind noise signals. It is assumed that the speech signal is coherent in both channels and the wind noise is uncorrelated. Furthermore speech and noise are expected to be uncorrelated and their auto PSDs are similar in each microphone channel, which is fulfilled for the measurements carried out in hands-free position. Thus, the noise estimate can be calculated based on the auto- and cross-PSDs of the input signals [6]:

$$\hat{\Phi}_{N,\text{theor}}(\lambda,\mu) = \sqrt{\Phi_{x_1x_1}(\lambda,\mu)\cdot\Phi_{x_2x_2}(\lambda,\mu)} - |\Phi_{x_1x_2}(\lambda,\mu)|. \tag{8}$$

In the case for noise level differences between the two signals the auto PSD of the signal used for the speech enhancement ($\Phi_{x_1x_1}$ for the presented system) should be applied instead of the geometric mean in (8). The general drawback of this noise estimate is again the dependency on the PSD computation as stated in Sec. 3. The value of α_S must not be too close to 1 in order to ensure a sufficient adaption to the wind noise. However this leads to an underestimation of the noise PSD and thus residual noise remains in the enhanced signal. The remaining noise might partially be masked during speech activity but is very annoying in speech pauses where it is audible as a low frequency rumbling noise. Therefore, we propose a modification for the noise estimate update based on the wind indicator $\overline{\sigma}_{\varphi,l}(\lambda)$. As shown in Sec. 3.1, $\overline{\sigma}_{\varphi,l}(\lambda)$ produces values close to 1 in periods with wind noise activity. Thus, it is applied as a cross-fade factor between the theoretical noise estimate from (8) and the input signal of the first channel:

$$\hat{\Phi}_N(\lambda,\mu) = \overline{\sigma}_{\varphi,l}(\lambda)\hat{\Phi}_{N,\text{theor}}(\lambda,\mu) + (1-\overline{\sigma}_{\varphi,l}(\lambda))|X_1(\lambda,\mu)|^2. \tag{9}$$

Besides the speech indicator is compared to a threshold ($\overline{\sigma}_{\varphi,h}(\lambda) < 0.2$) to ensure that speech segments are not taken into account for the noise estimate.

4 Speech Enhancement

The speech enhancement is realized by a spectral weighting as illustrated in Fig. 2. In [2] it turned out that a spectral subtraction rule [7] is a good choice for the reduction of wind noise. For the proposed system this rule is adapted by using side information given by the wind noise detection stage (Sec. 3.1). Commonly used methods for modifying the gains are a gain limitation or the application of a noise overestimation factor $\gamma(\lambda)$ to control the amount of speech distortion and noise reduction:

$$G(\lambda,\mu) = \max\left(\left(1 - \left(\frac{\hat{\Phi}_N(\lambda,\mu)}{\gamma(\lambda)\cdot|X_1(\lambda,\mu)|^2}\right)^2\right)^{1/2}, G_{\min}(\lambda)\right). \tag{10}$$

In periods where speech is active, a moderate noise reduction is beneficial in order to avoid speech distortion. Besides, residual noise is partially masked by the speech signal. During speech pauses a more aggressive noise reduction is desirable which removes a great amount of the distortion. Here we adapt the gain limit $G_{\min}(\lambda)$ and the overestimation factor $\gamma(\lambda)$ according to the speech indicator $\overline{\sigma}_{\varphi,h}(\lambda)$. Both parameters are linearly scaled by the speech indicator $\overline{\sigma}_{\varphi,h}(\lambda)$ between the limits G_{noise} and G_{speech} for the noise only and clean speech cases, respectively:

$$\begin{aligned} G_{\min}(\lambda) = \frac{1}{\sigma_{\varphi,\max} - \sigma_{\varphi,\min}} [&G_{\text{noise}}(\overline{\sigma}_{\varphi,h}(\lambda) - \sigma_{\varphi,\min}) \\ &+ G_{\text{speech}}(\sigma_{\varphi,\max} - \overline{\sigma}_{\varphi,h}(\lambda))] \end{aligned} \tag{11}$$

$\gamma(\lambda)$ is computed in the same way between the limits γ_{speech} and γ_{noise} (by exchanging G with γ in (11)). The parameters $\sigma_{\varphi,\min}$ and $\sigma_{\varphi,\max}$ are thresholds which determine clean speech and pure wind noise frames.

5 Evaluation Results

An evaluation was carried out with measured dual microphone recordings of wind noise and clean speech which were superposed to simulate different SNR values from -15 to 25 dB. For each SNR scenario 300 seconds wind noise and speech samples from [8] were taken. The hands-free setup was used as described in Sec. 2. The proposed algorithm was compared to four dual microphone algorithms where the second and third approaches are designed for the reduction of wind noise without a noise estimation:

- **CohWNest**: Original noise estimation approach proposed in [6] without exploiting the phase information using spectral subtraction gains,
- **SumDiff**: Wind noise suppression rule using the ratio of the sum and difference of the input signals [9],

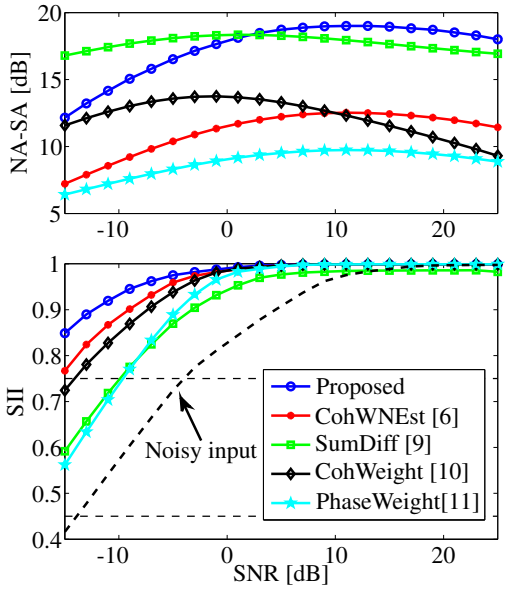

Figure 5: Evaluation results for different input SNR values

- **CohWeight**: Wind noise suppression rule using the MSC as a Wiener filter [10],
- **PhaseWeight**: Direct calculation of the spectral gains using the phase of the coherence function [11].

The PhaseWeight method was proposed for the attenuation of directional noise sources with a fixed phase characteristic without estimating a noise PSD but can also applied for the reduction of wind noise. The framework was used as presented in Sec. 2.2. The smoothing factor α_S was chosen to 0.85 and gains of the four above mentioned algorithms were limited to -30 dB attenuation which makes a good compromise between speech attenuation and noise reduction. For the proposed gain calculation the parameters were set to: $G_{\text{speech}}\widehat{=}$-20 dB, $G_{\text{noise}}\widehat{=}$ -60 dB, γ_{speech}=1, γ_{noise}=5, $\sigma_{\varphi,\text{min}}$=0.2, $\sigma_{\varphi,\text{max}}$=0.6. For the noise estimate using Eqn. (8) α_S was as well 0.85, but for the determination of the phase (Eqn. (3)) a smaller smoothing constant α_φ=0.5 was chosen. The noise reduction performance is determined by means of the noise attenuation minus speech attenuation (NA-SA) measure (e.g., [12]), where an improvement results in higher values. In addition the Speech Intelligibility Index (SII) [13] is applied as measure. 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. As seen in Fig.5, for both measures the proposed method yields to the highest improvements in most of the cases. Only for low SNR conditions (below -5 dB) the SumDiff method achieves a higher noise reduction in terms of the NA-SA value. Informal listening tests confirms this results but it turned out that the high performance of SumDiff is comprised by strong speech attenuation resulting in an audible highpass effect in the output signal. It is also clearly visible that the proposed modifications to [6] by exploiting the phase information provides a significant performance gain. Fig. 6 shows the spectrograms of the noisy input and enhanced signal of the proposed method. For reasons of clarity only the 0-4 kHz range is depicted. It can be seen that the wind noise is reduced by a great amount in speech pauses as well as during speech activity.

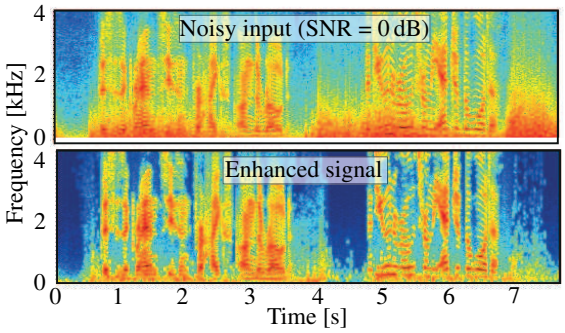

Figure 6: Spectrograms of input and output signals

6 Conclusions

In this contribution a dual microphone approach is presented to reduce wind noise in speech signal recorded by a hands-free device. For the detection, estimation and reduction of the wind noise properties of the complex coherence were exploited. In particular, the phase variance of the coherence provides a sufficient speech and wind noise detection which is required to handle the instationary nature of wind noise. An evaluation with measured signals showed that the proposed method outperforms other dual microphone approaches.

References

[1] 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.

[2] C. 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.

[3] G. Müller and M. Möser, *Handbook of engineering acoustics*. Springer, 2009.

[4] G. M. Corcos, "The structure of the turbulent pressure field in boundary-layer flows," *Journal of Fluid Mechanics*, vol. 18, pp. 353–378, 2 1964.

[5] ETSI EG 201 377-2, "Speech processing, transmission and quality aspects (STQ); Specification and measurement of speech transmission quality; Part 2: Mouth-to-ear speech transmission quality including terminals," April 2004.

[6] M. Dörbecker and S. Ernst, "Combination of two-channel spectral subtraction and adaptive Wiener post-filtering for noise reduction and dereverberation," in *Proc. of European Signal Processing Conf. (EUSIPCO)*, (Trieste, Italy), 1996.

[7] 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.

[8] P. Kabal, "TSP speech database," tech. rep., McGill University, Montreal, Canada, 2002.

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

[10] 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.

[11] P. Aarabi and G. Shi, "Phase-based dual-microphone robust speech enhancement," *IEEE Trans. Syst., Man, Cybern. B*, vol. 34, no. 4, pp. 1763 –1773, 2004.

[12] Quackenbush and Barnwell, *Objective Measures of Speech Quality*. Prentice-Hall, Inc., 1988.

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