BLIND ESTIMATION OF THE COHERENT-TO-DIFFUSE ENERGY RATIO FROM NOISY SPEECH SIGNALS

Marco Jeub, Christoph Nelke, Christophe Beaugeant⋆*, and Peter Vary*

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
⋆Intel Mobile Communications, Sophia-Antipolis, France
{jeub,nelke,vary}@ind.rwth-aachen.de christophe.beaugeant@intel.com

ABSTRACT

A novel dual-channel algorithm is proposed which estimates the coherent-to-diffuse energy ratio (CDR) of background noise in mixed noise fields. The algorithm is based on an estimate of the noise field coherence from a noisy speech signal and a subsequent minima tracking in order to increase the estimation accuracy even in the presence of speech.

The obtained CDR estimate can be used, e.g., for the acoustic environment classification in hearing aids or to control speech enhancement algorithms such as noise reduction or speech dereverberation. Besides, the approach can be used to calculate an estimate of the direct-to-reverberant energy ratio (DRR) blindly from reverberant speech signals.

1. INTRODUCTION

Acoustic environment classification is an important component of modern hearing aids, e.g., [1, 2] and various types of speech communication systems such as mobile phones. Depending on the acoustical environment the device should automatically adjust the operating mode and control all integrated algorithms. When it comes to multichannel speech enhancement algorithms, e.g., for noise reduction or dereverberation, it is of significant interest whether the noise field can be characterized as either coherent or diffuse (non-coherent) or as a mixture. This is important since many algorithms rely, e.g., on a diffuse noise field assumption and hence, do not show sufficient enhancement performance under other noise conditions or can lastly degrade the desired signal.

In this contribution we present a new approach of noise field classification which allows to estimate the ratio between coherent and diffuse noise (CDR) from a dual-channel noisy observation. Besides, the algorithm does not require a voice activity detector and gives reliable estimates even in the presence of speech. Depending on the application, this ratio can be computed either frequency-dependent or frequency-independent.

The remainder is organized as follows. In Section 2, the definition of the noise field coherence as well as the coherence of typical noise fields is introduced briefly. Section 3 discusses the superposition of coherent and diffuse noise fields and in Section 4 the proposed CDR estimation algorithm is discussed including simulation results. Possible applications are given in Section 5 and conclusions are drawn in Section 6.

2. NOISE FIELD COHERENCE

A well-established measure for describing the noise environment is the complex noise field coherence between two signals $x_1(k)$ and $x_2(k)$ with discrete time index k. It is defined in the frequency domain as [3]

$$\Gamma_{x_1x_2}(e^{j\Omega}) = \frac{\Phi_{x_1x_2}(e^{j\Omega})}{\sqrt{\Phi_{x_1x_1}(e^{j\Omega}) \cdot \Phi_{x_2x_2}(e^{j\Omega})}}, \tag{1}$$

where $\Phi_{x_1x_1}(e^{j\Omega})$ and $\Phi_{x_2x_2}(e^{j\Omega})$ represent the auto-power spectral density (PSD) of $x_1(k)$ and $x_2(k)$ in the Fourier domain. $\Phi_{x_1x_2}(e^{j\Omega})$ represents the cross-PSD. The normalized radian frequency is given by $\Omega = 2\pi f/f_s$ with frequency variable f and sampling frequency f_s.

The frequently used term magnitude squared coherence (MSC) is referred to the squared magnitude of Eq. (1) and is given by

$$C_{x_1x_2}(e^{j\Omega}) = \left|\Gamma_{x_1x_2}\left(e^{j\Omega}\right)\right|^2 = \frac{\left|\Phi_{x_1x_2}(e^{j\Omega})\right|^2}{\Phi_{x_1x_1}(e^{j\Omega}) \cdot \Phi_{x_2x_2}(e^{j\Omega})} \tag{2}$$

with the property

$$0 \le C_{x_1x_2}(e^{j\Omega}) \le 1. \tag{3}$$

In a diffuse noise field an infinite number of uncorrelated point sources are propagating in all directions simultaneously with equal energy and low spatial correlation. Assuming two microphones in the far-field and a homogeneous noise field, the spherically isotropic coherence can be calculated by the integration over all possible directions of incident of a directional sound source. The corresponding formula reads [3]

$$\Gamma^{(\mathrm{diff})}_{x_1x_2}(e^{j\Omega}) = \mathrm{sinc}\left(\Omega f_s d_{\mathrm{mic}}/c\right), \tag{4}$$

with distance d_{mic} between two omnidirectional microphones and sound velocity c[1]. The first zero-crossing of the sinc-function reads

$$f_0 = c/(2d_{\mathrm{mic}}), \tag{5}$$

e.g., for a given spacing of 0.15 m this frequency is $f_0 =$ 1.1 kHz. The noise field is highly correlated for frequencies below f_0 while the correlation is low for frequencies above f_0. The corresponding frequency bin is termed μ_0 in the DFT domain.

In this contribution, the coherence function of a coherent noise source, arriving at angle θ at the microphone arrangement which is placed in broadside orientation and assuming a small inter-microphone distance, is given by

$$\Gamma^{(\mathrm{coh})}_{x_1x_2}(e^{j\Omega}) = e^{-j\Omega f_s d_{\mathrm{mic}}\cos(\theta)/c}. \tag{6}$$

This corresponds to a value of one for the MSC

$$C^{(\mathrm{coh})}_{x_1x_2}(e^{j\Omega}) = 1 \;\forall\; \Omega. \tag{7}$$

[1] A sound velocity of $c = 340$m/s is used throughout this paper.

3. COHERENCE OF MIXED COHERENT AND DIFFUSE NOISE

When it comes to mixed noise fields, i.e., a superposition of diffuse (non-coherent) and coherent noise, the cross-PSD in Eq. (1) is given by the sum of the individual cross-PSDs [4]. This summation applies also for the total auto-PSD. It is assumed that all noise sources are uncorrelated with each other. Hence, the generalized complex coherence function reads

$$\Gamma^{(\mathrm{mix})}_{x_1x_2}(e^{j\Omega}) = \frac{\sum\limits_{n=1}^{N} \Phi^{(n)}_{x_1x_2}(e^{j\Omega})}{\sqrt{\sum\limits_{n=1}^{N} \Phi^{(n)}_{x_1x_1}(e^{j\Omega}) \cdot \sum\limits_{n=1}^{N} \Phi^{(n)}_{x_2x_2}(e^{j\Omega})}} \tag{8}$$

with N being the number of superposed noise fields or sources.

For an ideal diffuse noise field, the noise signals arrive with approximately equal power spectral density $\Phi_{\mathrm{d}}(e^{j\Omega})$ at the two microphones and hence,

$$\Phi^{(\mathrm{diff})}_{x_1x_1}(e^{j\Omega}) = \Phi^{(\mathrm{diff})}_{x_2x_2}(e^{j\Omega}) = \Phi_{\mathrm{d}}(e^{j\Omega}). \tag{9}$$

The cross-PSD in this case is given by [4]

$$\Phi^{(\mathrm{diff})}_{x_1x_2}(e^{j\Omega}) = \Phi_{\mathrm{d}}(e^{j\Omega}) \cdot \mathrm{sinc}\,(\Omega f_s d_{\mathrm{mic}}/c)\,. \tag{10}$$

For a coherent noise source which arrives at angle θ with energy $\Phi_{\mathrm{c}}(e^{j\Omega})$, the auto-PSD reads

$$\Phi^{(\mathrm{coh})}_{x_1x_1}(e^{j\Omega}) \approx \Phi^{(\mathrm{coh})}_{x_2x_2}(e^{j\Omega}) = \Phi_{\mathrm{c}}(e^{j\Omega}). \tag{11}$$

The corresponding cross-PSD can be expressed by

$$\Phi^{(\mathrm{coh})}_{x_1x_2}(e^{j\Omega}) = \Phi_{\mathrm{c}}(e^{j\Omega}) \mathrm{e}^{-\mathrm{j}\Omega \mathrm{f_s d_{mic}} \cos(\theta)/\mathrm{c}}. \tag{12}$$

If we assume that the coherent noise signal arrives without a time delay at the two sensors, i.e., $\theta = \pi/2$, the cross-PSD reduces to

$$\Phi^{(\mathrm{coh})}_{x_1x_2}(e^{j\Omega}) = \Phi_{\mathrm{c}}(e^{j\Omega}). \tag{13}$$

To take into account different ratios of coherent and diffuse noise, we introduce the coherent-to-diffuse energy ratio (CDR)

$$\Psi(e^{j\Omega}) = \frac{\Phi_{\mathrm{c}}(e^{j\Omega})}{\Phi_{\mathrm{d}}(e^{j\Omega})}. \tag{14}$$

Having described the auto- and cross-PSD equations for both coherent and diffuse noise, a model for the complex coherence of mixed noise fields (assuming $\theta = \pi/2$ for the coherent noise source) can be expressed by

$$\begin{aligned} \Gamma^{(\mathrm{mix})}_{x_1x_2}(e^{j\Omega}) &= \frac{\Phi_{\mathrm{c}}(e^{j\Omega}) + \Phi_{\mathrm{d}}(e^{j\Omega})\mathrm{sinc}\,(\Omega f_s d_{\mathrm{mic}}/c)}{\Phi_{\mathrm{c}}(e^{j\Omega}) + \Phi_{\mathrm{d}}(e^{j\Omega})} \\ &= \frac{\Psi(e^{j\Omega}) + \mathrm{sinc}\,(\Omega f_s d_{\mathrm{mic}}/c)}{1 + \Psi(e^{j\Omega})}. \end{aligned} \tag{15}$$

For the special case when the CDR $\Psi(e^{j\Omega})$ is equal for all frequencies, the three special cases are included in Eq.(15):

- $\Psi \to \infty \Rightarrow \Gamma^{(\mathrm{mix})}_{x_1x_2}(e^{j\Omega}) \to 1$: coherent noise field,
- $\Psi = 1$ same energy of diffuse and coherent noise,
- $\Psi \to 0 \Rightarrow \Gamma^{(\mathrm{mix})}_{x_1x_2}(e^{j\Omega}) \to \mathrm{sinc}\,(\Omega f_s d_{\mathrm{mic}}/c)$ diffuse noise field.

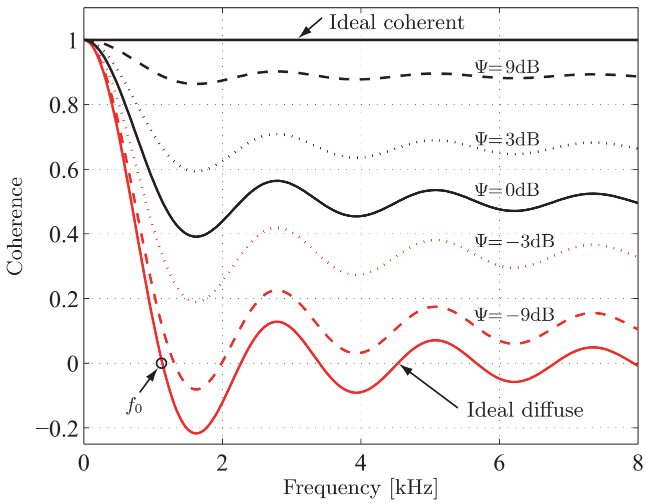

Figure 1: Coherence of mixed coherent and diffuse noise with varying CDR. For the simulations an inter-microphone distance of $d_{\mathrm{mic}} = 0.15\,\mathrm{m}$ is used. Red curves indicate a negative CDR and black curves a positive CDR. Besides, the curves for ideal diffuse and ideal coherent noise fields are plotted. The first zero-crossing of the sinc function is marked by f_0.

In the following, the CDR value will be expressed in dB. Figure 1 shows exemplarily the coherence of a coherent and diffuse noise field for different CDR values. The mixing procedure for the signals is described below. It can be seen that for $\Psi < -9\,\mathrm{dB}$ and for $\Psi > 9\,\mathrm{dB}$ the coherence can be approximated by a diffuse and coherent noise field, respectively. Hence, the CDR estimator should have an operating range of $-9 \leq \Psi \leq 9\,\mathrm{dB}$.

We use the following thresholds for $f > f_0$ to distinguish between purely diffuse and purely coherent noise fields:

- $\Gamma_{x_1x_2}(e^{j\Omega}) < 0.1 \to$ diffuse noise,
- $\Gamma_{x_1x_2}(e^{j\Omega}) > 0.9 \to$ coherent noise.

4. CDR ESTIMATION

4.1 Absence of Speech

In order to estimate the *real-valued* CDR from a given complex coherence function, Eq.(15) can be rearranged to

$$\Psi(e^{j\Omega}) = \frac{\mathrm{sinc}\,(\Omega f_s d_{\mathrm{mic}}/c) - \mathrm{Re}\{\Gamma^{(\mathrm{mix})}_{x_1x_2}(e^{j\Omega})\}}{\mathrm{Re}\{\Gamma^{(\mathrm{mix})}_{x_1x_2}(e^{j\Omega})\} - 1}. \tag{16}$$

where $\mathrm{Re}\{\Gamma_{x_1x_2}\} - 1 > 0$ has to be ensured for the denominator, e.g., by means of an upper threshold of the coherence $\Gamma_{\max} = 0.99$. The function $\mathrm{Re}\{\cdot\}$ returns the real part of its argument.

Thus, having an estimate of the coherence for a given noise signal available, leads directly to the ratio between coherent and diffuse noise. This CDR can also be seen as the ratio between coherent and non-coherent noise or as the ratio between direct speech and reverberant speech (direct-to-reverberant energy ratio (DRR)) in reverberant sound fields. In the following, all estimates are denoted by a hat-operator.

In the practical implementation, the algorithm is realized in the short-term spectral domain. The input signals $x_1(k)$ and $x_2(k)$ are first segmented into overlapping frames of length L. After windowing, these frames are transformed via Fast Fourier Transform (FFT) of length M ($M \geq L$). At discrete frequency bin μ, the signals are denoted by $X_1(\lambda, \mu)$

and $X_2(\lambda,\mu)$. The required coherence in Eq.(16) is calculated by Eq.(1). The auto- and cross-PSD terms in Eq.(1) are estimated recursively by means of the discrete short-time estimates according to

$$\hat{\Phi}_{x_1x_1}(\lambda,\mu) = \alpha\,\hat{\Phi}_{x_1x_1}(\lambda-1,\mu) + (1-\alpha)|X_1(\lambda,\mu)|^2, \tag{17}$$

$$\hat{\Phi}_{x_2x_2}(\lambda,\mu) = \alpha\,\hat{\Phi}_{x_2x_2}(\lambda-1,\mu) + (1-\alpha)|X_2(\lambda,\mu)|^2, \tag{18}$$

$$\hat{\Phi}_{x_1x_2}(\lambda,\mu) = \alpha\,\hat{\Phi}_{12}(\lambda-1,\mu) + (1-\alpha)X_1(\lambda,\mu)\cdot X_2^*(\lambda,\mu), \tag{19}$$

with smoothing factor $0 \le \alpha \le 1$.

4.2 Presence of Speech

The direct computation of Eq.(16) for the classification of background noise is only possible in segments of speech absence or in segments with very low speech energy. This would restrict the application of the CDR estimator to the very limited periods of speech absence and requires a robust voice activity detector. In this subsection it is discussed how to perform a reliable CDR estimation independent of speech activity.

The speech signal is assumed to be the target signal for any subsequent noise reduction or classification system, i.e., a person standing in front of an hearing impaired person, or the speech signal emitted from the mouth to a mobile phone in hand-held or hands-free position. Hence, this speech signal alone would have a coherence close to one if we assume that the person is located within the critical distance (high DRR). Besides, we assume that $\Psi(e^{j\Omega})$ changes slowly over time and sudden changes are not taken into account.

For the upcoming simulations, the noise signals are generated using the approach of [5], where predefined spatial coherence constraints can be employed. In order to demonstrate the principle of the algorithm, white Gaussian noise (WGN) with the same PSD is used for the coherent and diffuse noise which are mixed at a CDR of $\Psi = -9\,\text{dB}$. All signals are uncorrelated with each other. For the noisy signal, speech samples which are coherent among the microphones from the TSP speech database [6] are summed with the mixed noise signals at a signal-to-noise ratio (SNR) [2] of 10 dB. In Fig. 4 the corresponding time-domain waveform of the clean speech signal is shown.

For the simulations, an inter-microphone distance of $d_{\text{mic}} = 0.15\,\text{m}$ is assumed which is a typical distance between two binaural hearing aids. Besides, it is assumed that a possible time-delay of the speech signal among the microphones has been compensated. Such simulation ensures reproducible results and are required in order to evaluate the CDR estimation accuracy which would be inherently difficult with measured data. Further simulation parameters are listed in Table 1.

The influence of an additional speech signal to the noise input signal on the coherence function is illustrated in Fig.2. The estimated coherence for a noise-only and noisy speech frame is given.

In the case of speech absence (dashed black line), the estimated coherence follows the theoretical red solid line for the given CDR (Eq. (15)). When it comes to a noisy speech signal (solid black line), the estimated noise field coherence is biased towards higher values. This can be explained since the coherence of the speech signal alone would be one, as mentioned previously. Besides, the speech signal has the highest

[2] Here, the SNR is defined as the ratio between the desired speech signal to the mixture of diffuse and coherent noise.

Table 1: Main simulation parameters.

Parameter	Settings
Sampling frequency	$f_s = 16\,\text{kHz}$
Frame length	$L = 320$ (20 ms)
FFT length	$M = 512$ (incl.zero-pad.)
Frame overlap	75% (Hann window)
Smoothing factor	$\alpha = 0.8$ (Eqs.(17),(18) and (19))
Smoothing factor	$\alpha_{\text{coh}} = 0.5$ (Eq.(1))
Inter-microphone distance	$d_{\text{mic}} = 0.15\,\text{m}$
SNR	10 dB

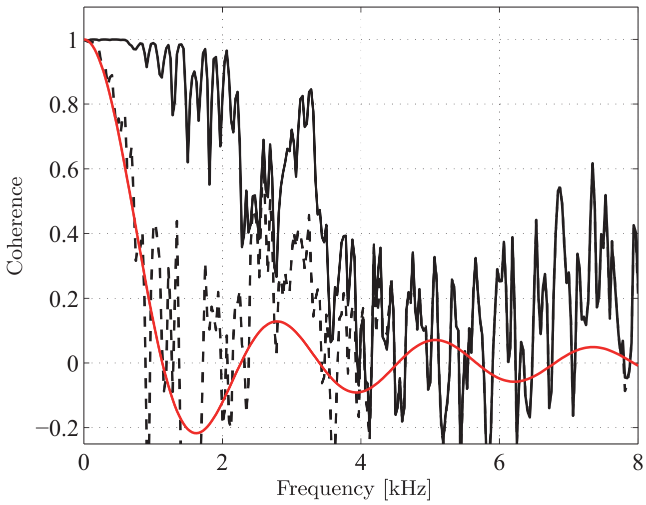

Figure 2: Coherence of noise (dashed) and noisy speech (solid) with frequency-independent CDR of $\Psi = -9\,\text{dB}$ and SNR=10 dB. The ideal diffuse noise field coherence curve is marked by the solid red line.

energy components for a frequency range up to 3 kHz. Hence, the estimated CDR would always return values for the mixing ratio which are higher than the true CDR in segments of speech activity. As mentioned before, for very low SNR conditions (< 0 dB), where the noise signal dominates the overall signal energy, this effect is negligible.

In order to counteract the biased CDR estimate, a minima tracking of the estimated CDR per frequency bin is performed using a large tracking window of 1.5 s. This concept is well-known from noise PSD estimation [7]. Figure 3 shows the estimated CDR over time for one specific frequency bin. In this case the CDR was set fixed to $\Psi = 9\,\text{dB}$ and the SNR to 10 dB. Depicted is the estimated CDR without minima tracking $\widehat{\Psi}(\mu)$ and with minima tracking $\widehat{\Psi}_{(\text{min})}(\mu)$. From this figure, it can be concluded that the speech signal causes a severe bias to the estimated CDR and hence, the performance of any subsequent algorithm which relies on this estimate would be degraded.

The complete CDR estimation algorithm is summarized in Fig.5. The processing is performed for each frame λ and, depending on the application, a frequency-independent or frequency-dependent CDR estimate can be obtained. For complexity reasons, fixed smoothing factors for the recursive estimation of the auto- and cross-PSD and no bias correction as in [7] are employed. The integration can, however, reduce the remaining bias between the true and estimated CDR.

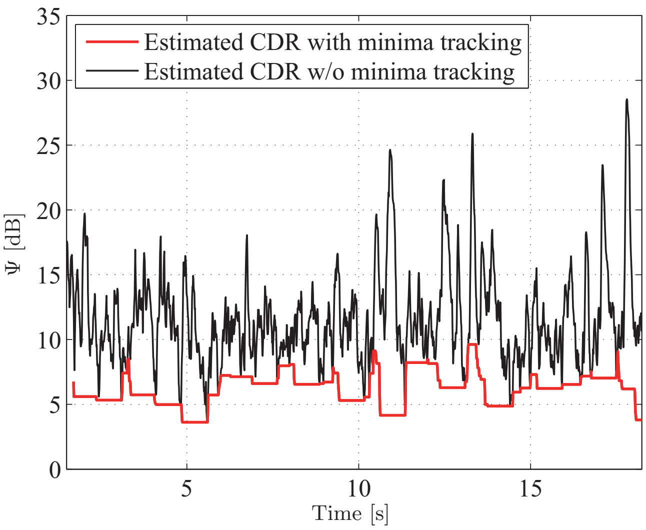

Figure 3: Estimated CDR without minima tracking $\widehat{\Psi}(\mu)$ and with minima tracking $\widehat{\Psi}_{(\min)}(\mu)$ for one frequency bin $\mu = 65$ ($\widehat{=} 2\,\mathrm{kHz}$). The signals are mixed with a fixed CDR $\Psi = 9\,\mathrm{dB}$ and SNR=10 dB.

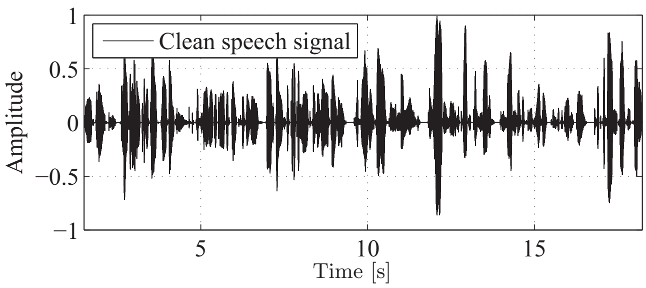

Figure 4: Time-domain waveform of clean speech signal from TSP database.

- transform current frame λ into the frequency domain with frequency bin $\mu = 1, ..., N$
- compute smoothed auto- and cross-PSD $\hat{\Phi}_{x_1x_1}(\lambda,\mu)$, $\hat{\Phi}_{x_2x_2}(\lambda,\mu)$, $\hat{\Phi}_{x_1x_2}(\lambda,\mu)$ by recursive smoothing (α); Eqs.(17),(18) and (19)
- compute coherence $\Gamma_{x_1x_2}(\lambda,\mu)$; Eq.(1) by recursive smoothing (α_{coh})
- compute CDR estimate $\widehat{\Psi}(\lambda,\mu)$; Eq.(16)
- perform minima tracking to obtain $\widehat{\Psi}_{(\min)}(\lambda,\mu)$
- average over frequency

$$\overline{\Psi}(\lambda) = \frac{1}{N-\mu_0}\sum_{\mu=\mu_0}^{N} \widehat{\Psi}_{(\min)}(\lambda,\mu) \tag{20}$$

- process next frame $\lambda + 1$

Figure 5: CDR estimation algorithm. The averaging over frequency is required only for a frequency-independent CDR estimate. Estimating the DRR from a reverberant speech signal does not require the minima tracking.

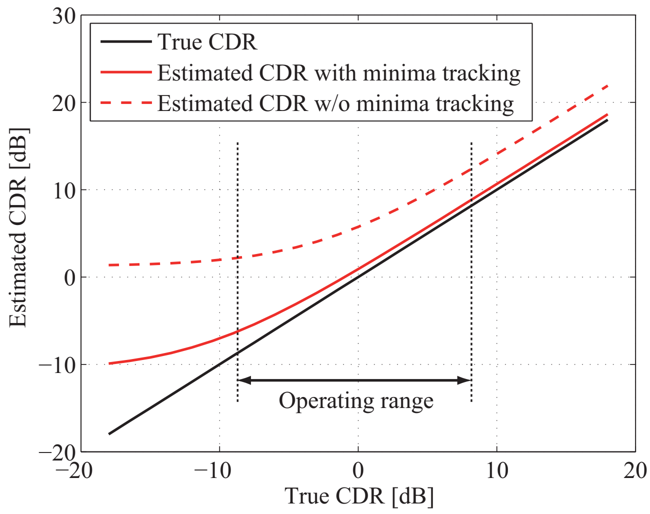

Figure 6: True and estimated CDR for varying CDR values. The results are averaged over all frequency bins for each CDR step and averaged over time.

The overall performance of the proposed CDR estimator with and without the minima tracking is depicted in Fig. 6. The accuracy of the CDR estimate has been increased significantly by applying a minima tracking the direct estimate obtained from Eq.(16). For the desired operating range of $-9 \leq \Psi \leq 9\,\mathrm{dB}$ (see above and Fig.1), the estimator shows sufficient accuracy.

Since the ideal diffuse noise field shows a high coherence for low frequencies (see Fig.1 and Eq.(4)), no clear separation between coherence and diffuse noise is possible. The estimator is only capable for a frequency-dependent CDR estimation above $f_0/4$, which was confirmed heuristically by experiments. In further simulations we have shown that the inter-microphone distance d_{mic} should be equal or greater than 0.02 m.

4.3 Influence of Coherent Noise Direction

It has to be mentioned that the assumption for the angle of the interfering source of $\theta = \pi/2$ in the derivation of the CDR estimator does not have an influence on the estimation performance as shown in Fig. 7. In the simulation, a coherent noise signal from alternative angles was considered. It can be seen that the simplification in the derivation of the CDR estimator does not affect the estimation accuracy. Besides, the main target applications are (binaural) hearing aids and dual-microphone mobile phones where an inter-microphone spacing of 15 cm, 10 cm or 3 cm is assumed and a framewise processing with a typical frame length of 20 ms.

5. APPLICATIONS

The following list gives a few application examples of the proposed CDR estimator.

Speech dereverberation:

- Without minima tracking, the scheme in Fig.5 can give an estimate of the direct-to-reverberant energy ratio (DRR) since the direct speech can be seen as coherent speech and the reverberant speech as diffuse or non-coherent speech. Please note that the use of the heuristically motivated equation

$$\mathrm{DRR}(\mathrm{e}^{\mathrm{j}\Omega}) = \frac{|\mathrm{sinc}\,(\Omega \mathrm{f_s d_{mic}}/\mathrm{c})\,|^2 - |\Gamma^{(\mathrm{mix})}_{\mathrm{x_1x_2}}(\mathrm{e}^{\mathrm{j}\Omega})|^2}{|\Gamma^{(\mathrm{mix})}_{\mathrm{x_1x_2}}(\mathrm{e}^{\mathrm{j}\Omega})|^2 - 1}. \tag{21}$$

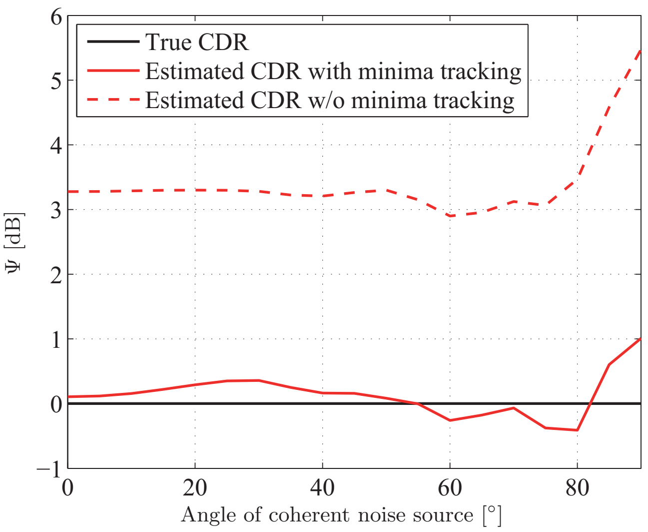

Figure 7: True and estimated CDR for varying angles θ of the coherence noise source. 90° indicates frontal direction.

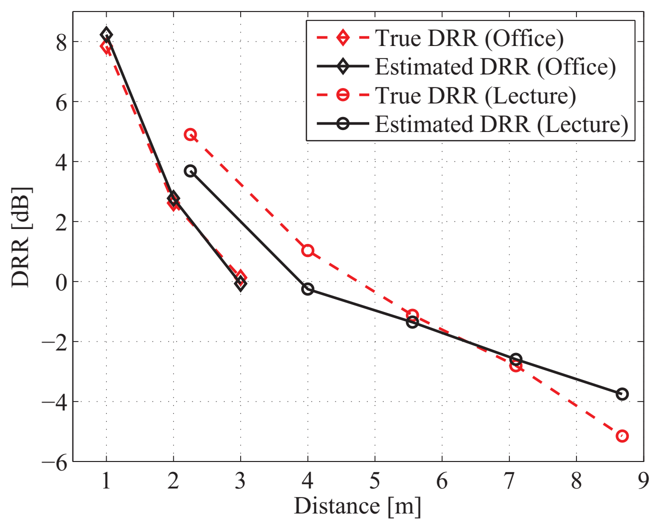

Figure 8: True and estimated DRR over the source-microphone distance calculated from reverberant speech signals using room impulse responses from the AIR database.

instead of Eq.(16), a time-alignment with respect to the direct speech and an additional recursive smoothing of the estimate ($\alpha_{\text{DRR}} = 0.99$) leads to the best performance even without a voice activity detector. See Fig.8 for the DRR estimation performance using binaural room impulse responses of lecture room and office from the AIR database which were convolved with above introduced speech signal[3].

- The CDR can be used to control the two stages of our recently proposed two stage dereverberation algorithm [8] where in a first stage late reverberation is reduced, followed by a subsequent stage where all non-coherent components are attenuated. Hence, knowledge of the CDR can lead to a more suitable combination of the two spectral weighting gains.

Noise PSD estimation:

- The CDR can control a combination of coherence-based noise PSD estimators which work well in diffuse noise fields with an MMSE-based algorithm or Minimum Statistics (MS) which are also capable to track coherent noise (see [9] and the references therein).

Hearing Aids:

- Automatic selection of processing modes.

Mobile phones:

- Interfering talker detection.

6. CONCLUSIONS

In this paper, we present and evaluate a novel approach which estimates the ratio between coherent and diffuse noise in a mixed noise field. It is shown that from the estimated coherence function of a dual-channel signal a reliable estimate of the CDR can be derived. This estimate can either be frequency-dependent or independent. In a second step, the bias which is provoked by speech presence in the signals is removed by a subsequent minima tracking. Experiments show that the estimated and true CDR matches over the normal operating range $-9 \le \Psi \le 9\,\text{dB}$.

Applications are the classification of the acoustical environment for hearing aids and in noise reduction systems for a more efficient gain calculation by adapting the algorithm to different noise field environments.

[3]The true DRR is obtained directly from the impulse response. The direct path is chosen such that a few early reflections are also included, here: 10 ms.

REFERENCES

[1] V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP Journal on Applied Signal Processing*, vol. 18, pp. 2915–2929, 2005.

[2] T. Wittkopp, *Two-channel noise reduction algorithms motivated by models of binaural interaction*, Ph.D. thesis, Universität Oldenburg, Oldenburg, Germany, 2001.

[3] H. Kuttruff, *Room Acoustics*, Spon Press, Oxon, 2009.

[4] A.G. Piersol, "Use of coherence and phase data between two receivers in evaluation of noise environments," *Journal of Sound and Vibration*, vol. 56, no. 2, pp. 215–228, 1978.

[5] E.A.P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, 2008.

[6] P. Kabal, "TSP speech database," Tech. Rep., Department of Electrical & Computer Engineering, McGill University, Montreal, Quebec, Canada, 2002.

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

[8] M. Jeub, M. Schäfer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1732–1745, 2010.

[9] M. Jeub, C. M. Nelke, H. Krüger, C. Beaugeant, and P. Vary, "Robust dual-channel noise power spectral density estimation," in *Proc. European Signal Processing Conference (EUSIPCO)*, Barcelona, Spain, 2011.