UNBIASED COHERENT-TO-DIFFUSE RATIO ESTIMATION FOR DEREVERBERATION

Andreas Schwarz, Walter Kellermann

Multimedia Communications and Signal Processing
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Cauerstr. 7, 91058 Erlangen, Germany
{schwarz, wk}@lnt.de

ABSTRACT

We investigate the estimation of the time- and frequency-dependent coherent-to-diffuse ratio (CDR) from the measured spatial coherence between two omnidirectional microphones. We illustrate the relationship between several known CDR estimators using a geometric interpretation in the complex plane, discuss the problem of estimator bias, and propose unbiased versions of the estimators. Furthermore, we show that knowledge of either the direction of arrival (DOA) of the target source or the coherence of the noise field is sufficient for an unbiased CDR estimation. Finally, we apply the CDR estimators to the problem of dereverberation, using automatic speech recognition word error rate as objective performance measure.

Index Terms— Dereverberation, Reverberation Suppression, Spatial Coherence, Diffuse Noise Suppression

1. INTRODUCTION

The idea of using short-time spatial coherence estimates for signal enhancement in the short-time Fourier transform (STFT) domain dates back to 1977, when Allen et al. [1] proposed to essentially use the magnitude of the spatial coherence in each time- and frequency bin as a gain for reverberation suppression. Other heuristic methods for noise reduction and dereverberation using coherence estimates have since been proposed, e.g., in [2], and most recently in [3], where a soft threshold function is used to compute a gain from the coherence magnitude, and the parameters of the threshold function are adapted depending on the histogram of the coherence magnitude.

Short-time coherence estimates have also been investigated for noise suppression by postfilters as part of beamformers, and solutions which are optimal under certain conditions have been derived for the suppression of uncorrelated [4, 5] and diffuse [6] noise. Compared to the heuristic methods, an important result from these postfilters is that optimum diffuse noise suppression is not possible based on only the magnitude of the spatial coherence.

More recently, explicit estimators for the coherent-to-diffuse ratio (CDR), i.e., the ratio between direct and diffuse signal components, have been formulated [7, 8], building on the earlier optimum postfilter derivations. These results have since been generalized from omnidirectional microphones to other microphone directivities [9, 10], spherical microphone arrays [11], and applied to dereverberation with different noise coherence functions [12].

In this paper, we first describe the signal model for the acquisition of a noisy or reverberated signal by two omnidirectional microphones, and define the CDR. Then, we investigate and visualize several known CDR estimators, and propose improved unbiased variants. Using a geometric interpretation in the complex plane, we show that knowledge of either the target signal direction or the noise coherence is sufficient for an unbiased CDR estimation, and derive estimators for the cases of unknown target signal direction and unknown noise coherence. Finally, we apply the CDR estimators to dereverberation, processing reverberated speech and evaluating the recognition accuracy achieved by an automatic speech recognizer.

2. SIGNAL MODEL

We consider the acquisition of a reverberated or noisy speech signal by two omnidirectional microphones with spacing d. The auto- and cross-power spectra of the microphone signals x_i are $\Phi_{x_i x_j}(k,f)$, $i,j = 1,2$, with the frame index k and frequency f. Assuming that microphones are identical and closely spaced, $\Phi_{x_1 x_1} = \Phi_{x_2 x_2} = \Phi_x$. The complex spatial coherence function is then defined as

$$\Gamma_x(k,f) = \frac{\Phi_{x_1 x_2}(k,f)}{\Phi_x(k,f)}. \tag{1}$$

Furthermore, it is assumed that the direct signal and noise or reverberation components, with power spectra Φ_s and Φ_n, respectively, are orthogonal, so that $\Phi_x = \Phi_s + \Phi_n$. The direct sound is modeled as a plane wave with the direction of arrival (DOA) θ with respect to the microphone axis, where $\theta = 0\,^\circ$ corresponds to broadside direction. The noise or reverberation component is modeled as a diffuse sound field; for (late) reverberation, this assumption can be made since observation window lengths used in practice are much shorter than room impulse responses [13]. The corresponding spatial coherence functions for the direct and diffuse sound components are then given by

$$\Gamma_s(f) = e^{j2\pi f \Delta t}, \tag{2}$$

$$\Gamma_n(f) = \Gamma_{\text{diff}}(f) = \operatorname{sinc}(2\pi f \frac{d}{c}), \tag{3}$$

respectively, with the time difference of arrival (TDOA) $\Delta t = d\sin(\theta)/c$. The coherence of the mixture of the direct-path signal and diffuse noise can be written as a function of the coherent-to-diffuse ratio $CDR(k,f) = \Phi_s(k,f)/\Phi_n(k,f)$:

$$\Gamma_x(k,f) = \frac{CDR(k,f)\Gamma_s(f) + \Gamma_n(f)}{CDR(k,f)+1}. \quad (4)$$

This can be rewritten as a parametric line equation in the complex plane, highlighting that Γ_x lies on a straight line connecting Γ_n and Γ_s:

$$\Gamma_x(k,f) = \Gamma_s(f) + \frac{1}{CDR(k,f)+1}(\Gamma_n(f) - \Gamma_s(f)). \quad (5)$$

Note that the line parameter $[CDR(k,f)+1]^{-1}$ is equivalent to the *diffuseness* defined in [14].

3. COHERENT-TO-DIFFUSE RATIO ESTIMATION

Solving (4) for the CDR yields (we omit the time- and frequency-dependency of the coherence in the following):

$$CDR(k,f) = \frac{\Gamma_n - \Gamma_x}{\Gamma_x - \Gamma_s}. \quad (6)$$

In theory, although the coherence values may be complex, the CDR is real-valued; however, when inserting a coherence estimate $\hat{\Gamma}_x$ (e.g., computed from recursively estimated spectra), the resulting CDR value will in general be complex-valued, due to mismatch between the coherence models and room acoustics and the variance of the spectrum estimates. A number of different practical estimator realizations have therefore been proposed, which implicitly account for these errors, and which we will compare in the following. In order to illustrate their behavior, we visualize the output of different estimators over the complex plane of possible coherence values $\hat{\Gamma}_x$ in Fig. 1. Results for a direct path TDOA $\Delta t = 0$ (broadside) are shown in the first row, while in the second row, results are shown for $\Delta t = \frac{1}{5f}$. The $\circ$ marks the coherence of a fully coherent signal with the respective TDOA, while the $\times$ marks the coherence of an ideal diffuse signal. The straight white line between these points marks the coherence values which would occur in theory under ideal conditions for different CDR values, according to (5). We define the *bias* of a CDR estimator as the deviation from (6) for coherence values along this line; i.e., an *unbiased* estimator should exactly match (6) for these values, as can be verified by inserting Γ_x according to (4) into the estimator equation. Furthermore, due to various effects mentioned before, the coherence estimates $\hat{\Gamma}_x$ which are observed in practice will not lie exactly on the line, therefore a good estimator should also be *robust* in the sense that some deviations of the coherence estimate from the assumed model, e.g., caused by an inexact DOA estimate, do not lead to large deviations of the estimate.

Using the assumptions described in Sect. 2, McCowan et al. [6] derived an estimate for the target signal under presence of diffuse noise. Jeub et al. [15] evaluated the same method specifically for the suppression of reverberation, and, instead of directly deriving an estimate for the target signal, formulated a CDR estimator [7]. Both McCowan and Jeub rely on the assumption that the direct path is perfectly time-aligned in both microphones, which can be achieved by obtaining a TDOA estimate $\widehat{\Delta t}$ and applying a corresponding delay to one of the channels [15]. For ease of analysis and comparison, we here represent this delay directly in the CDR estimator as a phase shift applied to the coherence estimate:

$$\widehat{CDR}_{\text{Jeub}}(k,f) = \frac{\Gamma_n - \mathrm{Re}\{e^{-j2\pi f\widehat{\Delta t}}\hat{\Gamma}_x\}}{\mathrm{Re}\{e^{-j2\pi f\widehat{\Delta t}}\hat{\Gamma}_x\} - 1}. \quad (7)$$

Delaying one of the channels to achieve time alignment of the direct path however also affects the phase of the coherence of the diffuse signal component. Since this is not accounted for in this estimator, the estimate is biased for non-zero TDOAs, as will be shown later.

Thiergart et al. [8, 10] proposed another estimator where the direct path coherence Γ_s is computed from a TDOA estimate $\widehat{\Delta t}$,

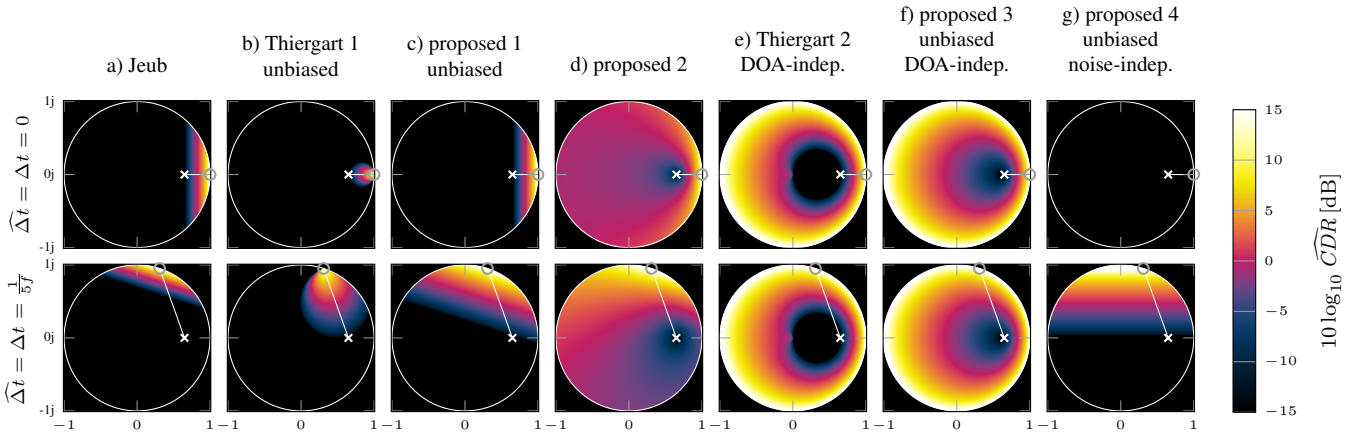

Fig. 1. Coherent-to-diffuse ratio estimates as a function of complex spatial coherence $\hat{\Gamma}_x$, for $d = 8\,\mathrm{cm}$, $f = 1\,\mathrm{kHz}$. Different estimators (columns) and TDOA values (rows). Coherence of fully diffuse (Γ_n) and fully coherent (Γ_s) signals are highlighted.

and the real part of (6) is taken:

$$\widehat{CDR}_{\text{Thiergart1}}(k,f) = \operatorname{Re}\left\{\frac{\Gamma_n - \hat{\Gamma}_x}{\hat{\Gamma}_x - e^{j2\pi f\widehat{\Delta t}}}\right\}. \tag{8}$$

While unbiased, this estimator is not robust towards phase deviations of the coherence estimate [10], since, for a measured coherence with a magnitude close to one, even small phase differences between $\hat{\Gamma}_x$ and Γ_s can have a large effect on the CDR estimate. This can be seen in Fig. 1b, where, unlike in Fig. 1a, the CDR for coherence values along the unit circle sharply drops to zero.

We propose a new estimator based on (7), where we correct the diffuse coherence model by multiplying Γ_n with the phase term $e^{-j2\pi f\widehat{\Delta t}}$, thereby removing the bias of the estimator caused by the time alignment, while preserving the robustness towards phase deviation (see Fig. 1c):

$$\widehat{CDR}_{\text{prop1}}(k,f) = \frac{\operatorname{Re}\{e^{-j2\pi f\widehat{\Delta t}}\Gamma_n - e^{-j2\pi f\widehat{\Delta t}}\hat{\Gamma}_x\}}{\operatorname{Re}\{e^{-j2\pi f\widehat{\Delta t}}\hat{\Gamma}_x\} - 1}. \tag{9}$$

In a second, heuristically motivated variant, which is illustrated in Fig. 1d, we use the magnitude instead of the real part as in (9). We found that this increases robustness towards model errors and leads to increased dereverberation performance [16]:

$$\widehat{CDR}_{\text{prop2}}(k,f) = \left|\frac{e^{-j2\pi f\widehat{\Delta t}}\Gamma_n - e^{-j2\pi f\widehat{\Delta t}}\hat{\Gamma}_x}{\operatorname{Re}\{e^{-j2\pi f\widehat{\Delta t}}\hat{\Gamma}_x - 1\}}\right|. \tag{10}$$

Note that this estimator has a small bias for non-zero TDOAs; compensation of the bias however only has a negligible effect on practical performance and is therefore omitted here.

Thiergart et al. [8, 10] alternatively proposed to use the instantaneous phase of the cross-power spectrum (which is the same as the phase of the estimated coherence $\arg\hat{\Gamma}_x$) as a phase estimate for the direct path model, which has the advantage of not requiring an explicit TDOA estimate:

$$\widehat{CDR}_{\text{Thiergart2,DOA-indep.}}(k,f) = \operatorname{Re}\left\{\frac{\Gamma_n - \hat{\Gamma}_x}{\hat{\Gamma}_x - e^{j\arg\hat{\Gamma}_x}}\right\}. \tag{11}$$

However, the instantaneous phase of the mixture is not an unbiased estimate for the phase of the direct path, since, for low CDR, the coherence of the mixture is dominated by the coherence of the diffuse signal [10]. For $\theta \neq 0\,^\circ$, this leads to a bias in the CDR estimate, as can be observed in Fig. 1e, second row.

It is in fact possible to derive an unbiased estimator for the CDR which does not require knowledge of the TDOA, since the assumption that $|\Gamma_s| = 1$, i.e., that the direct signal is fully coherent, is sufficient to solve (6). This can be visualized using the observation that, according to (5), Γ_x, Γ_s and Γ_n all lie on a straight line in the complex plane, and it is furthermore known that Γ_s lies on the unit circle and Γ_n on the real axis. The resulting estimator equation is given in (12) and illustrated in Fig. 1f.

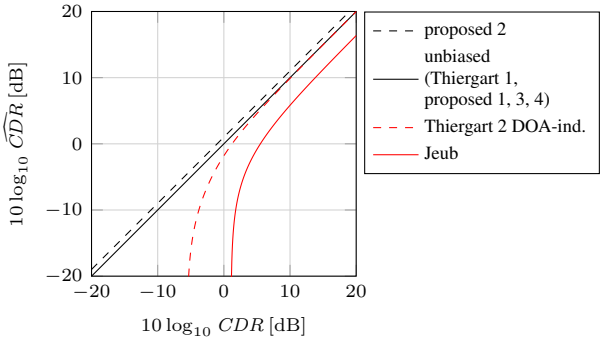

Fig. 2. Comparison of true CDR and estimated CDR for $d = 8\,\text{cm}$, $f = 1\,\text{kHz}$, $\widehat{\Delta t} = \frac{1}{5f}$.

Analogously, we can conclude that unbiased coherent-to-noise ratio estimation is theoretically possible in noise fields with unknown coherence Γ_n, exploiting only the knowledge that the noise coherence is real. The corresponding estimator, illustrated in Fig. 1g, is given by:

$$\widehat{CDR}_{\text{prop4}}(k,f) = \begin{cases} \frac{1}{\frac{\operatorname{Im}\Gamma_s}{\operatorname{Im}\hat{\Gamma}_x} - 1}, & \text{for } \frac{\operatorname{Im}\Gamma_s}{\operatorname{Im}\hat{\Gamma}_x} \geq 1 \\ \infty, & \text{for } 0 < \frac{\operatorname{Im}\Gamma_s}{\operatorname{Im}\hat{\Gamma}_x} < 1 \\ 0, & \text{for } \frac{\operatorname{Im}\Gamma_s}{\operatorname{Im}\hat{\Gamma}_x} \leq 0, \end{cases} \tag{13}$$

where the case differentiation accounts for cases where $\operatorname{Im}\hat{\Gamma}_x$ has values outside of the expected range, i.e., a larger magnitude than $\operatorname{Im}\Gamma_s$, or a different sign. An important constraint that limits practical applicability of this estimator is that $\Delta t \neq 0$, since otherwise the imaginary parts disappear. Note that in [17] a noise estimate was derived in a similar way from the imaginary part of a cross spectrum estimate.

Fig. 2 compares the true CDR value and the different estimates for mixtures of coherent and ideally diffuse signals (corresponding to the values along the white line in Fig. 1, second row). The proposed estimators (9), (12) and (13) are unbiased, as is the DOA-dependent estimator proposed by Thiergart et al. (8). Our heuristically motivated estimator (10) has a small, constant bias that is dependent on Δt and f. The estimators by Jeub et al. (7) and the DOA-independent estimator by Thiergart et al. (11) show a significant bias, here in the form of an underestimation; for other values of Δt and f, overestimation can also occur.

$$\widehat{CDR}_{\text{prop3}}(k,f) = \frac{\Gamma_n \operatorname{Re}\{\hat{\Gamma}_x\} - |\hat{\Gamma}_x|^2 - \sqrt{\Gamma_n^2 \operatorname{Re}\{\hat{\Gamma}_x\}^2 - \Gamma_n^2 |\hat{\Gamma}_x|^2 + \Gamma_n^2 - 2\Gamma_n \operatorname{Re}\{\hat{\Gamma}_x\} + |\hat{\Gamma}_x|^2}}{|\hat{\Gamma}_x|^2 - 1} \tag{12}$$

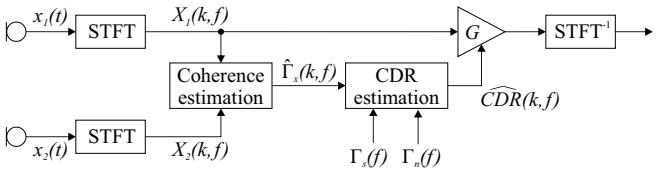

Fig. 3. Coherence-based dereverberation system.

4. CDR-BASED DEREVERBERATION

Fig. 3 shows the structure of a reverberation or diffuse noise suppression system using a short-time CDR estimate [15]. First, microphone signals are transformed into the STFT domain, and short-time estimates $\hat{\Gamma}_x(k,f)$ of the spatial coherence are obtained according to (1) from power spectra estimated by recursive averaging. Then, the CDR is estimated, using models for the direct path and reverberation coherence, and a suppression gain is computed from the CDR, here, using spectral magnitude subtraction [18]:

$$G(k,f) = \max\left\{G_{\min}, \sqrt{\frac{\mu}{\widehat{CDR}(k,f)+1}}\right\}, \qquad (14)$$

with the oversubtraction factor μ and the gain floor $G_{\min}$. Since the focus of our evaluation is the CDR estimation, we apply the suppression simply to the first microphone signal. Alternatives would be to average the power spectra between microphones [4], or to use beamforming to combine the signals, which would however change the CDR and therefore require consideration in the suppression filter computation [19].

We use sets of measured impulse responses from three rooms: Room A ($6\,\mathrm{m} \times 6\,\mathrm{m} \times 3\,\mathrm{m}$, $T_{60} \approx 0.4\,\mathrm{s}$), Room B (lecture hall, $7\,\mathrm{m} \times 11\,\mathrm{m} \times 3\,\mathrm{m}$, $T_{60} \approx 1\,\mathrm{s}$), and Room C (foyer, $54\,\mathrm{m} \times 7\,\mathrm{m} \times 3\,\mathrm{m}$, $T_{60} \approx 3.5\,\mathrm{s}$). In each room, impulse responses were measured for 40-70 different source positions in $l = 1$, 2 and $4\,\mathrm{m}$ distance from the microphones, in the angular range $\theta = -90\ldots90\,^\circ$. All processing takes place at a sampling rate of 16 kHz using a DFT-based filter bank [20] with window length 1024, DFT length 512, and downsampling factor 128. The coherence estimates are obtained by recursive averaging with the forgetting factor $\lambda = 0.68$.

We use the word error rate (WER) of an automatic speech recognizer for an objective evaluation of the overall dereverberation performance. 500 utterances from the GRID corpus [21] are reverberated by convolution with each of the measured 2-channel impulse responses, and processed by the dereverberation methods. We compare all CDR estimators discussed in this paper, except the proposed variant 4, which does not work for $\Delta t = 0$ and is therefore only of limited use in this scenario. In addition to the CDR-based methods, we evaluate a version of [1] (where we directly use the magnitude of the coherence as the gain, and apply the enhancement to only one microphone), and the coherence-to-gain-mapping proposed by Westermann et al. [3] (using offline-estimated coherence statistics for each room and position, and the parameter $k_p = 0.25$, which was found to yield good results across all rooms). We also evaluate the exponential decay model by Lebart et al. [22] (assuming perfect knowledge of the reverberation time). For the method of Lebart and the CDR-based methods, spectral magnitude subtraction (14) is applied to the first microphone, with $G_{\min} = 0.1$; to ensure a fair comparison, μ is optimized in the range $0.7\ldots2.5$ for maximum recognition performance individually for each room and dereverberation method. Ideal TDOA knowledge is used for the CDR estimators which require a TDOA estimate $\widehat{\Delta t}$. The employed ASR engine is PocketSphinx [23] using MFCC+Δ+ΔΔ features, cepstral mean normalization, and a speaker-independent acoustic model trained on clean speech.

Table 1. Average ASR Word Error Rate.

		Room A	Room B	Room C	mean
	clean	7.9	7.9	7.9	7.9
	reverberated	13.0	50.8	63.6	42.5
	Lebart	10.3	26.5	45.2	27.3
Coherence-based	Allen	11.2	37.5	52.6	33.8
Coherence-based	Westermann	10.3	34.2	49.9	31.5
CDR-based	Jeub	10.6	25.3	35.0	23.6
CDR-based	† Thiergart 1	13.5	36.7	47.6	32.6
CDR-based	† proposed 1	10.6	24.8	33.4	22.9
CDR-based	proposed 2	**10.0**	**23.3**	**32.6**	**22.0**
CDR-based	* Thiergart 2	10.3	28.4	45.1	27.9
CDR-based	*† proposed 3	10.2	27.1	42.4	26.6

* DOA-independent

† unbiased

Table 1 shows the resulting WER for the letter and number in the utterance, averaged over all source positions. The magnitude-based methods by Allen and Westermann have a relatively weak dereverberating effect, and, unlike the methods based on spectral subtraction, offer no direct way of tuning the amount of suppression, therefore WER improvements are generally lower than with the CDR-based methods. An exception is room A, where we found a significant mismatch between the measured late reverberation coherence and the diffuse model, which benefits Westermann's adaptation to the coherence statistics. Overall, dereverberation using the proposed estimators leads to a reduced WER, both for the case of known and unknown DOA. It is worth noting that the DOA-independent CDR estimator allows effective dereverberation without requiring any prior information or long-term estimation of signal characteristics, relying only on the short-time coherence estimate.

5. CONCLUSION

We have investigated and illustrated the behavior of several known CDR estimators. Based on the observation that known methods either yield a biased estimate for non-zero TDOAs, or, in one case, are not robust enough for application to signal enhancement, we have then proposed unbiased estimators, both for the case where prior knowledge on the DOA is available, and where either the DOA or the noise coherence are unknown. We have shown that applying the proposed methods in a CDR-based dereverberation system leads to a consistent improvement in ASR accuracy.

6. REFERENCES

[1] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *The Journal of the Acoustical Society of America*, vol. 62, no. 4, pp. 912–915, 1977.

[2] R. Le Bouquin-Jeannes, A. A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 5, pp. 484–487, Sept. 1997.

[3] A. Westermann, J. M. Buchholz, and T. Dau, "Binaural dereverberation based on interaural coherence histograms," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2767–2777, 2013.

[4] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. ICASSP*, 1988.

[5] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, May 1998.

[6] I. A. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 6, pp. 709–716, 2003.

[7] M. Jeub, C. M. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in *Proc. EUSIPCO*, 2011.

[8] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones," in *Proc. ICASSP*, 2012.

[9] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Diffuseness estimation with high temporal resolution via spatial coherence between virtual first-order microphones," in *Proc. WASPAA*, 2011.

[10] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," *The Journal of the Acoustical Society of America*, vol. 132, pp. 2337, 2012.

[11] D. P. Jarrett, O. Thiergart, E. A. P. Habets, and P. A. Naylor, "Coherence-based diffuseness estimation in the spherical harmonic domain," in *Proc. 27th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, 2012.

[12] M. Jeub and P. Vary, "Binaural dereverberation based on a dual-channel Wiener filter with optimized noise field coherence," in *Proc. ICASSP*, 2010.

[13] F. Jacobsen and T. Roisin, "The coherence of reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 204–210, 2000.

[14] G. Del Galdo, M. Taseska, O. Thiergart, J. Ahonen, and V. Pulkki, "The diffuse sound field in energetic analysis," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2141–2151, Mar. 2012.

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

[16] A. Schwarz, A. Brendel, and W. Kellermann, "Coherence-based dereverberation for automatic speech recognition," in *Proc. DAGA*, 2014.

[17] N. Ito, N. Ono, E. Vincent, and S. Sagayama, "Designing the Wiener post-filter for diffuse noise suppression using imaginary parts of inter-channel cross-spectra," in *Proc. ICASSP*, 2010.

[18] E. Haensler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley-Interscience, 2004.

[19] S. Lefkimmiatis and P. Maragos, "A generalized estimation approach for linear and nonlinear microphone array post-filters," *Speech Communication*, vol. 49, no. 7–8, pp. 657–666, July 2007.

[20] M. Harteneck, S. Weiss, and R.W. Stewart, "Design of near perfect reconstruction oversampled filter banks for subband adaptive filters," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 8, pp. 1081–1085, Aug. 1999.

[21] M. Cooke, J. Barker, S. Cunningham, and Xu Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2421–2424, 2006.

[22] K. Lebart, J.-M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 359–366, 2001.

[23] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishankar, and A. I. Rudnicky, "PocketSphinx: a free, real-time continuous speech recognition system for hand-held devices," in *Proc. ICASSP*, 2006.