SIGNAL-TO-REVERBERANT RATIO ESTIMATION BASED ON THE COMPLEX SPATIAL COHERENCE BETWEEN OMNIDIRECTIONAL MICROPHONES

Oliver Thiergart, Giovanni Del Galdo, and Emanuël A.P. Habets

International Audio Laboratories Erlangen*

ABSTRACT

The signal-to-reverberant ratio (SRR) is an important parameter in several applications such as speech enhancement, dereverberation, and parametric spatial audio coding. In this contribution, an SRR estimator is derived from the direction-of-arrival dependent complex spatial coherence function computed via two omnidirectional microphones. It is shown that by employing a computationally inexpensive DOA estimator, the proposed SRR estimator outperforms existing approaches.

Index Terms— Array signal processing, spatial coherence, signal-to-reverberation ratio

1. INTRODUCTION

The spatial coherence between two microphone signals contains useful information for the statistical characterization of a sound field. The spatial coherence can be used for instance to determine the ratio between the coherent energy and the diffuse energy present in a room. This information, often expressed by the signal-to-reverberant ratio (SRR) [1], is crucial for many applications such as speech enhancement and dereverberation [2] or parametric spatial audio coding [3]. Considering the spatial coherence for estimating the SRR is especially beneficial as only two microphones are required which can be placed nearly arbitrarily. Moreover, the processing can be carried out in the short-time Fourier transform (STFT) domain which is crucial for many real-time applications.

The spatial coherence between two omnidirectional microphones in a purely diffuse sound field is well studied [4]. For mixed sound fields it becomes a complex function depending on the direction-of-arrival (DOA) of the coherent sound. The authors in [5] consider the real part of the complex coherence function to derive an SRR estimator for omnidirectional microphones. In doing so, they assume that the coherent sound arrives from a specific direction, namely the broadside direction, at the two microphones. Considering the squared absolute value of the spatial coherence was proposed in [6]. Since the DOA dependency is not considered, the estimator yields biased results depending on the DOA of the coherent sound. Closely related to considering the spatial coherence for the SRR estimation is the approach in [7]. The authors express the power spectral densities (PSDs) between several omnidirectional microphones as a function of the coherent and diffuse sound energy, and determine both quantities separately. In doing so, this approach requires at least three microphones for estimating the SRR.

In order to exploit all relevant information, we consider the spatial coherence as a complex function for deriving an SRR estimator that requires two omnidirectional microphones. It is shown that considering the DOA dependency is particularly important in strongly coherent fields, i. e., for sound fields where an accurate DOA estimation can usually be carried out. Therefore, as verified by measurements, the SRR computation can be significantly improved even when employing a computationally inexpensive DOA estimator.

The contribution is organized as follows: Section 2 introduces the sound field model. Section 3 reviews the spatial coherence for mixed sound fields. The complex SRR estimator is derived in Sec. 4 and verified in Sec. 5 based on measurements. The conclusions are drawn in Sec. 6.

2. PROBLEM FORMULATION

Consider a sound field where the sound pressure $S(k,t,\boldsymbol{d})$ in an arbitrary point $\boldsymbol{d}$ in the Cartesian coordinate system at time instant t and wavenumber $k = 2\pi f/c$ (frequency f, speed of sound c) is formed by a superposition of *directional sound* (coherent sound) and *diffuse sound*, i. e.,

$$S(k,t,\boldsymbol{d}) = S_{\text{dir}}(k,t,\boldsymbol{d}) + S_{\text{diff}}(k,t,\boldsymbol{d}). \tag{1}$$

The directional sound $S_{\text{dir}}(k,t,\boldsymbol{d})$ is given by a single monochromatic plane wave with DOA expressed by the unit norm vector $\boldsymbol{n}_{\text{dir}}(k)$, i. e.,

$$S_{\text{dir}}(k,t,\boldsymbol{d}) = \sqrt{E_{\text{dir}}(k,t)}\, e^{j\mu(k)+j\xi(k,t)}, \tag{2}$$

where $\mu(k) = -k\,\boldsymbol{n}_{\text{dir}}^{\mathrm{T}}(k)\,\boldsymbol{d}$, $E_{\text{dir}}(k,t)$ is the energy of the wave, and $\xi(k,t)$ is the phase of the wave in the origin. The diffuse sound field $S_{\text{diff}}(k,t,\boldsymbol{d})$ is assumed to be spatially isotropic, meaning that the sound arrives with equal strength from all directions, and spatially homogeneous, meaning that its mean energy $E_{\text{diff}}(k,t) = \mathrm{E}\left\{|S_{\text{diff}}(k,t,\boldsymbol{d})|^2\right\}$ does not

*A joint institution of the University Erlangen-Nuremberg and Fraunhofer IIS, Germany vary with $\boldsymbol{d}$. Such a diffuse field can be modeled by summing an infinite number of monochromatic plane waves with equal magnitudes, random phases, and uniformly distributed propagation directions. In the following, $S_{\text{dir}}(k,t,\boldsymbol{d})$ and $S_{\text{diff}}(k,t,\boldsymbol{d})$ are assumed to be uncorrelated. The energy ratio between the directional sound and diffuse sound is expressed by the SRR $\Gamma(k,t)$, defined as

$$\Gamma(k,t) = \frac{E_{\text{dir}}(k,t)}{E_{\text{diff}}(k,t)}. \tag{3}$$

We aim at estimating the SRR $\Gamma(k,t)$ with $M = 2$ omnidirectional microphones located in $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$. According to the sound field model in (1), the i-th microphone signal with $i \in \{1,2\}$ can be written as

$$P_i(k,t,\boldsymbol{d}_i) = S_{\text{dir}}(k,t,\boldsymbol{d}_i) + S_{\text{diff}}(k,t,\boldsymbol{d}_i) + N_i(k,t), \tag{4}$$

where $N_i(k,t)$ models the microphone self-noise as independent and identically distributed (i. i. d.) zero-mean complex Gaussian noise with energy $E_{\text{noise}}(k,t) = \mathrm{E}\left\{|N_i(k,t)|^2\right\} \forall i$.

3. SPATIAL COHERENCE IN MIXED FIELDS

Let us derive the complex spatial coherence between two displaced omnidirectional microphones with signals $P_1(k,t,\boldsymbol{d}_1)$ and $P_2(k,t,\boldsymbol{d}_2)$ according to the sound field model presented in the previous section. Notice that in the following we omit the dependencies on time t for simplicity. Let $\Phi_{ij}(k)$ denote the PSD computed for the i-th and j-th microphones as

$$\Phi_{ij}(k) = \mathrm{E}\left\{P_i(k,\boldsymbol{d}_i)\, P_j^*(k,\boldsymbol{d}_j)\right\}, \tag{5}$$

representing the auto PSD for $i = j$ and the cross PSD, otherwise. For the signal model in Sec. 2, the PSD $\Phi_{ij}(k)$ equals the sum of the following individual PSDs

$$\Phi_{ij}(k) = \Phi_{\text{dir},ij}(k) + \Phi_{\text{diff},ij}(k) + \Phi_{\text{noise},ij}(k). \tag{6}$$

Since the microphone noise is uncorrelated between the microphones, we have $\Phi_{\text{noise},11}(k) = \Phi_{\text{noise},22}(k) = E_{\text{noise}}(k)$ and $\Phi_{\text{noise},12}(k) = 0$. The PSDs $\Phi_{\text{dir},ij}(k)$ resulting from the directional sound can be written as

$$\Phi_{\text{dir},11}(k) = \Phi_{\text{dir},22}(k) = E_{\text{dir}}(k), \tag{7a}$$

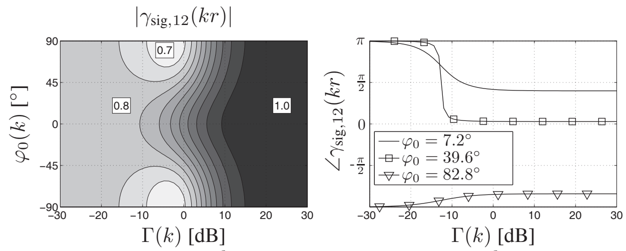
$$\Phi_{\text{dir},12}(k) = E_{\text{dir}}(k)\, e^{j\mu_{\text{dir}}(k)}, \tag{7b}$$

where $\mu_{\text{dir}}(k) = -k\, \boldsymbol{n}_{\text{dir}}^{\mathrm{T}}(k)\, \boldsymbol{r}$ is the phase shift of the directional sound from the first to the second microphone with $\boldsymbol{r} = \boldsymbol{d}_2 - \boldsymbol{d}_1$ being the displacement vector of the two microphones. The diffuse sound PSDs $\Phi_{\text{diff},ij}(k)$ are

$$\Phi_{\text{diff},11}(k) = \Phi_{\text{diff},22}(k) = E_{\text{diff}}(k), \tag{8a}$$

$$\Phi_{\text{diff},12}(k) = \gamma_{\text{diff},12}(kr)\, E_{\text{diff}}(k), \tag{8b}$$

where $r = \|\boldsymbol{r}\|$ is the microphone spacing. The spatial coherence $\gamma_{\text{diff}}(kr)$ for ideal diffuse sound fields assuming no



(a) absolute value, $kr = 1$ (b) phase, $kr = 10$

Fig. 1: Absolute value and phase of $\gamma_{\text{sig},12}(kr)$ for different SRRs and DOAs of the directional sound

microphone noise was derived in [4] for omnidirectional microphones. For spherically isotropic diffuse sound fields, we obtain the well-known result $\gamma_{\text{diff}}(kr) = \frac{\sin(kr)}{kr}$. The noiseless PSDs are

$$\Phi_{\text{sig},ij}(k) = \Phi_{\text{dir},ij}(k) + \Phi_{\text{diff},ij}(k) \tag{9}$$

$$= \Phi_{ij}(k) - \Phi_{\text{noise},ij}(k). \tag{10}$$

This leads to the *noiseless* spatial coherence function

$$\gamma_{\text{sig},12}(kr) = \frac{\Phi_{\text{sig},12}(k)}{\sqrt{\Phi_{\text{sig},11}(k)}\sqrt{\Phi_{\text{sig},22}(k)}} = \frac{\Phi_{\text{sig},12}(k)}{\Phi_{\text{sig}}(k)}, \tag{11}$$

where $\Phi_{\text{sig},11}(k) = \Phi_{\text{sig},22}(k) = \Phi_{\text{sig}}(k)$. In practice, $\Phi_{\text{sig}}(k)$ can be estimated by averaging $\Phi_{\text{sig},11}(k)$ and $\Phi_{\text{sig},22}(k)$ computed directly from the microphone signals after subtraction of the estimated noise floor energy. Inserting (7)–(9) and (3) into (11) yields the noiseless spatial coherence $\gamma_{\text{sig},12}(kr)$ as function of the SRR $\Gamma(k)$, i. e.,

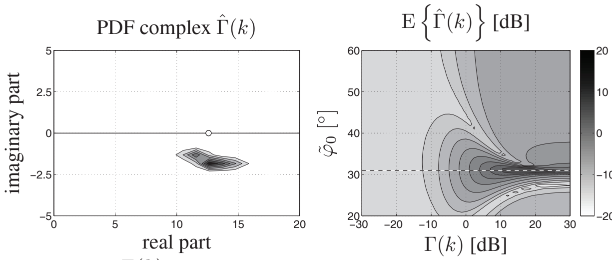
$$\gamma_{\text{sig},12}(kr) = \frac{\Gamma(k)\, e^{j\mu_{\text{dir}}(k)} + \gamma_{\text{diff}}(kr)}{\Gamma(k) + 1}. \tag{12}$$

As expected, we have $\gamma_{\text{sig},12}(kr) = \gamma_{\text{diff}}(kr)$ for $\Gamma(k) = 0$ while $\gamma_{\text{sig},12}(kr) = e^{j\mu_{\text{dir}}(k)}$ for $\Gamma(k) \to \infty$. For $\Gamma(k) > 0$ the coherence is dependent on the DOA of the directional sound and becomes real if the directional sound arrives from broadside direction (i. e., $\boldsymbol{n}_{\text{dir}} \perp \boldsymbol{r}$). Figure 1 shows the absolute value and phase of $\gamma_{\text{sig},12}(kr)$ as function of $\Gamma(k)$ for different kr. The directional sound is propagating in the horizontal plane and arriving from azimuth angle $\varphi_0(k)$ where 0° is the broadside direction. The absolute value in (a) is DOA dependent. It follows from (12) that this dependency vanishes at kr where $\gamma_{\text{diff},12}(kr) \to 0$, since the exponential then disappears when taking the absolute value. The phase $\angle\gamma_{\text{sig},12}(kr)$ depicted in (b) contains relevant information on $\Gamma(k)$ as well. In general, it is DOA dependent for any $kr > 0$.

4. SIGNAL-TO-REVERBERATION ESTIMATION

The SRR is determined by estimating the complex noiseless spatial coherence $\gamma_{\text{sig},12}(kr)$ and solving (12) for $\Gamma(k)$, i. e.,

$$\hat{\Gamma}(k) = \mathrm{Re}\left\{\frac{\gamma_{\text{diff}}(kr) - \hat{\gamma}_{\text{sig},12}(kr)}{\hat{\gamma}_{\text{sig},12}(kr) - e^{j\mu_{\text{dir}}(k)}}\right\}, \tag{13}$$



(a) $\Gamma(k) = 11\,\mathrm{dB}$ (b)

Fig. 2: Estimated SRR at $kr = 5$. In this simulation a plane wave arriving from $\varphi_0 = 31°$ was superimposed to a spherically isotropic diffuse sound field. Microphone noise (21 dB SNR) was present. The noise energy was considered in (9). The expectation in (5) was approximated via averaging over $K = 100$ realizations of the sound field.

where $\hat{\gamma}_{\mathrm{sig},12}(kr)$ is the estimated complex spatial coherence obtained with (5), (10), and (11) and $\hat{\Gamma}(k)$ is the resulting SRR estimate. Clearly, computing $\hat{\Gamma}(k)$ requires information on the DOA of the directional sound, or more precisely, on the spatial frequency $\mu_{\mathrm{dir}}(k)$. A computationally inexpensive estimator for $\mu_{\mathrm{dir}}(k)$ providing relatively accurate results is given by

$$\hat{\mu}_{\mathrm{dir}}(k) = \angle\Phi_{12}(k). \tag{14}$$

In predominantly directional sound fields, the total cross PSD $\Phi_{12}(k)$ approximates (7b) so that its angle provides $\mu_{\mathrm{dir}}(k)$. Figure 2(a) depicts an exemplary probability density function (PDF) of $\hat{\Gamma}(k)$ when omitting the real part operator in (13). The white circle indicates the true $\Gamma(k)$. Microphone noise is present in this simulation and $\mu_{\mathrm{dir}}(k)$ is estimated with (14). It is very likely to obtain complex SRR estimates $\hat{\Gamma}(k)$. The straight-forward way for obtaining a valid (real-valued) SRR value is taking the real part as proposed in (13). When the complex $\hat{\Gamma}(k)$ is almost symmetrically distributed as in the example in Fig. 2(a), then this solution provides the most probable result. Figure 2(b) depicts the real part of $\hat{\Gamma}(k)$ for the same simulation but when considering incorrect *a priori* information on the DOA of the directional sound (indicated by $\tilde{\varphi}_0$) instead of using (14). The dashed line shows the true DOA. Using incorrect $\tilde{\varphi}_0$ for computing $\mu_{\mathrm{dir}}(k)$ leads rapidly to a severe underestimation of the SRR at larger $\Gamma(k)$. Thus, precise DOA information is particularly necessary for non-diffuse sound fields, in which an accurate DOA estimation can usually be carried out.

In contrast to the proposed approach in (13), the authors in [5] assume that the directional sound arrives from the broadside direction. Consequently, they do not require information on the DOA of the directional sound and can consider only the real part of the spatial coherence as the imaginary part becomes zero. However, it follows from Fig. 1(b) that the imaginary part contains relevant information on the SRR, namely when the directional sound does not arrive from broadside direction. Alternatively to [5], one can consider the magnitude squared coherence (MSC) $|\hat{\gamma}_{\mathrm{sig},12}(kr)|^2$ as proposed in [6]. However, this disregards the phase of $\hat{\gamma}_{\mathrm{sig},12}(kr)$ which also contains relevant information on the SRR as depicted in Fig. 1(b).

5. EVALUATION

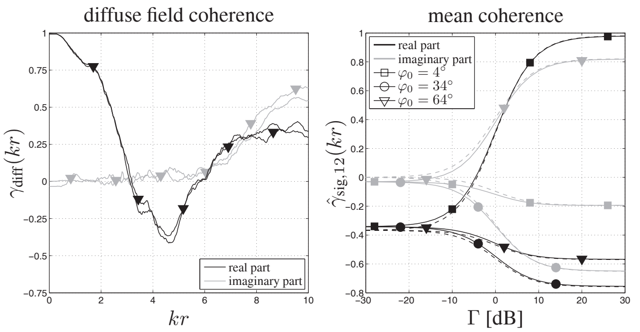
5.1. Measurement Setup

Measurements in an anechoic chamber and a reverberant environment have been carried out to verify the presented approach. The sound was recorded with sampling frequency $f_{\mathrm{s}} = 44.1\,\mathrm{kHz}$ using two omnidirectional microphones in the horizontal plane with spacing $r = 4.4\,\mathrm{cm}$. The directional sound $S_{\mathrm{dir}}(k, t, \boldsymbol{d})$ was generated in the anechoic chamber by reproducing pink noise from direction φ_0 and transforming the recorded signals with a 1024-point STFT (with 50% overlap). Similarly, a semi-spherically isotropic diffuse field $S_{\mathrm{diff}}(k, t, \boldsymbol{d})$ was obtained by reproducing uncorrelated pink noise signals from 25 loudspeakers on a hemisphere in a room with mean reverberation time $\mathrm{RT}_{60} \approx 360\,\mathrm{ms}$. The total pressure signals $P_1(k, t, \boldsymbol{d}_1)$ and $P_2(k, t, \boldsymbol{d}_2)$ in (4) are generated by adding the two recorded sound fields accordingly to the desired SRR $\Gamma(k)$. The PSDs $\Phi_{ij}(k)$ are determined with (5) where the expectation operator is approximated by averaging over K time frames. Finally, the SRR $\hat{\Gamma}(k)$ is determined with (11) and (13) where $\mu_{\mathrm{dir}}(k)$ is estimated with (14). In the following experiment we disregard the microphone noise, i. e., we assume $\Phi_{\mathrm{sig},ij}(k) = \Phi_{ij}(k)$ in (11), since the recordings were performed with relatively high SNR.

5.2. Measurement Results

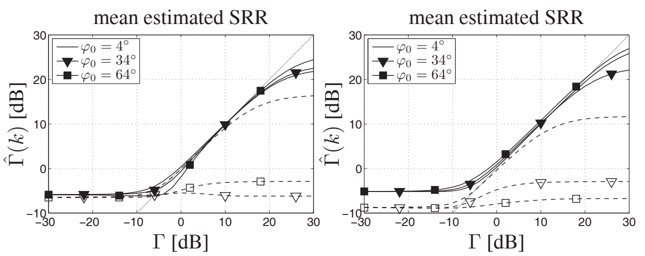
Let us first investigate the diffuse field coherence $\gamma_{\mathrm{diff}}(kr)$, required in (13). It was estimated with (5) and (11) from the diffuse field recording. A relatively long temporal averaging $K = 5500$ ($\approx 60\,\mathrm{s}$) was applied. The lines without markers in Fig. 3(a) depict the real and imaginary part, respectively, of the measured $\gamma_{\mathrm{diff}}(kr)$. The result was verified by rotating the microphone array by $90°$ and repeating the measurement (plots with markers). The results for both measurements are nearly identical showing that the created diffuse field was approximately spatially isotropic as required by the derivation in Sec. 3. The mean of the two estimated coherence functions depicted in Fig. 3(a) is used as $\gamma_{\mathrm{diff}}(kr)$ in the following.

Figure 3(b) illustrates the mean of the estimated spatial coherence $\hat{\gamma}_{\mathrm{sig},12}(kr)$ at $kr = 4.6$ ($f = 5.7\,\mathrm{kHz}$) as function of $\Gamma(k)$ for different DOAs φ_0 of the directional sound ($0°$ is the broadside direction). The black and gray lines indicate the real and imaginary part of $\hat{\gamma}_{\mathrm{sig},12}(kr)$, respectively. The underlying dashed lines show the theoretical coherence computed with (12) using the measured $\gamma_{\mathrm{diff}}(kr)$. The estimation results follow the theoretical functions. Small deviations can be observed for small $\Gamma(k)$ resulting mainly from the relatively short temporal averaging $K = 20$ ($\approx 220\,\mathrm{ms}$) applied here which is more typical in practical applications.



(a) (b) $kr = 4.6$ (≈ 5.7 kHz)

Fig. 3: Plot (a): measured diffuse field coherence $\gamma_{\text{diff}}(kr)$. Plot (b): estimated spatial coherence $\hat{\gamma}_{\text{sig},12}(kr)$ compared to the theoretical function (dashed lines).



(a) $kr = 0.83$ (≈ 1 kHz) (b) $kr = 6.0$ (≈ 7.4 kHz)

Fig. 4: Estimated SRR as function of the true SRR for different DOAs of the directional sound. Solid lines: complex estimator. Dashed lines: considering only the real part of the complex spatial coherence as proposed in [5].

Figure 4 shows the mean of the estimated SRR $\hat{\Gamma}(k)$ as function of the true SRR $\Gamma(k)$ for different wavenumbers, DOAs of the directional sound, and $K = 20$. The black solid lines show the results when using the complex estimator (13)–(14). The dashed lines show the results when considering only the real part of the spatial coherence as proposed in [5]. The gray line depicts the correct results. The proposed complex SRR estimator (black solid lines) provides accurate results particularly for medium $\Gamma(k)$. The overestimation at low $\Gamma(k)$ follows from Fig. 3(b). As the coherence functions are relatively flat at low (also at high) $\Gamma(k)$, the bias in the estimated coherence $\hat{\gamma}_{\text{sig},12}(kr)$ yields a large bias in the estimated SRR $\hat{\Gamma}(k)$. Notice that in a practical application, when considering the theoretical diffuse field coherence $\gamma_{\text{diff}}(kr) = \frac{\sin(kr)}{kr}$ (instead of the measured one), which deviates even more from the true coherence at low $\Gamma(k)$, then the SRR estimation bias is very likely further increased.

Figure 4 illustrates further that the results of the proposed complex estimator are nearly identical for different φ_0 showing the benefit of considering the DOA for the SRR estimation, even though the proposed estimator (14) is relatively simple. In contrast, the results for the estimator proposed in [5] (dashed lines) are strongly DOA dependent. Relatively accurate results are achieved for $\varphi_0 = 4°$. For this DOA, the comparatively large bias at high $\Gamma(k)$ results from the directional sound not arriving exactly at the array broadside (0°).

6. CONCLUSIONS

In this paper we have investigated the complex spatial coherence computed between two omnidirectional microphones in the case of a sound field composed of the superposition of coherent and diffuse sound. The main contribution is an estimator for the signal-to-reverberant ratio (SRR) which was derived from the complex spatial coherence. While traditional methods consider only the real part of the spatial coherence, the proposed estimator also employs its imaginary part, thus avoiding an implicit assumption on the direction-of-arrival (DOA) of the coherent sound. As a result, measurements have shown that the proposed estimator outperforms existing estimators, especially for large values of the SRR.

7. REFERENCES

[1] P. A. Naylor, E. A. P. Habets, J. Y.-C. Wen, and N. D. Gaubitch, "Models, measurement and evaluation," in *Speech Dereverberation*, P. A. Naylor and N. D. Gaubitch, Eds., chapter 2, pp. 21–56. Springer, 2010.

[2] P. J. Bloom, "Evaluation of a dereverberation technique with normal and impaired listeners.," *British Journal of Audiology*, vol. 16, no. 3, pp. 167–176, August 1982.

[3] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc*, vol. 55, no. 6, pp. 503–516, June 2007.

[4] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson Jr., "Measurement of correlation coefficients in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1072–1077, 1955.

[5] M. Jeub, C. M. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in *19th European Signal Processing Conference (EUSIPCO 2011)*, Barcelona, Spain, August 2011.

[6] O. Thiergart, M. Kratschmer, M. Kallinger, and G. D. Galdo, "Parameter estimation in directional audio coding using linear microphone arrays," in *Audio Engineering Society Convention 130*, London UK, May 2011.

[7] Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, and Y. Haneda, "Estimating direct-to-reverberant energy ratio based on spatial correlation model segregating direct sound and reverberation," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, Dallas Texas, USA, March 2010.