DIRECT-TO-REVERBERANT ENERGY RATIO ESTIMATION BASED ON INTERAURAL COHERENCE AND A JOINT ITD/ILD MODEL

Mehdi Zohourian and Rainer Martin

Ruhr-Universität Bochum, Institute of Communication Acoustics, Bochum, Germany
{mehdi.zohourian, rainer.martin}@rub.de

ABSTRACT

This paper proposes a novel algorithm to estimate the direct-to-reverberant energy ratio (DRR) using hearing aid microphones. The algorithm is based on the interaural magnitude-squared coherence of signals and is able to take both phase and level differences of microphones signals in the binaural configuration into account. We employ a spherical head model to approximate binaural cues. The proposed algorithm uses the common assumption of an ideally diffuse reverberation sound field. We test our approach on signals based on simulated and on measured binaural room impulse responses. Results show improved performance of the proposed algorithm as compared to other coherence-based DRR estimation methods.

Index Terms— Direct-to-reverberant ratio, spatial coherence, binaural hearing aids

1. INTRODUCTION

DRR is an important measure to describe an acoustic transmission path and is beneficial in speech enhancement algorithms, e.g., for dereverberation [1, 2, 3, 4, 5] or for speaker distance estimation [6, 7, 8, 9]. Typically, DRR can be estimated from the measured *room impulse response* (RIR) which practically requires an intrusive measurement within the room or a blind estimation procedure. A common assumption made for DRR estimation is to discard the energy of the early reflections and to consider only the energy of the direct sound and the late reverberation which is modeled as a diffuse sound field. Therefore, several studies have introduced the *coherent-to-diffuse ratio* (CDR) [10, 11] as an estimator of DRR since a fully coherent source implicitly represents the direct component.

Recently, most algorithms have made use of multichannel signals in order to estimate DRR. Broadly speaking, these methods can be classified into two main groups of coherence-based [10, 12, 11] and beamforming-based techniques [7, 8, 13, 14]. Coherence-based algorithms estimate DRR by computing the coherence of microphone signals and modeling the reverberant component as an ideal isotropic noise field. Generally, the coherence (as defined in (13), see below) is a complex-valued quantity and thus requires a proper mapping to a tractable positive real-valued DRR [10, 12]. The authors in [11] introduce and compare a variety of CDR estimators including DOA-dependent/independent and noise coherence aware/unaware techniques. They investigate the performance of these methods on a postfilter used for dereverberation. The binaural extension of this algorithm [11] has been proposed in [15] where the authors employ

This work has received funding from the European Fund for Regional Development, grant no EFRE-0800372 (RaVis-3D).

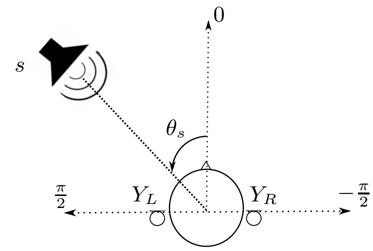


Fig. 1. Binaural microphone array configuration with a source s, and the coordinate system.

a frequency dependent *interaural time difference* [16] for DRR estimation. The performance of various state-of-the-art algorithms for DRR estimation has been investigated for real recordings in [17], yet, the binaural DRR estimation is still less explored.

In this study we propose a binaural DRR estimation approach using hearing aid (HA) microphones. Our algorithm is based on the interaural *magnitude-squared coherence* (MSC) of microphone signals which results in a tractable real-valued measure for DRR estimation. The authors in [18] also propose to use MSC for the estimation of the diffuseness, however, without considering any directional information. They report that the MSC-based diffuseness estimator provides a better performance for $\theta_s = 0$ than for $\theta_s = \frac{\pi}{2}$. The proposed algorithm, however, takes the shadowing effect of the head into consideration by employing joint binaural cues, namely, interaural time/level differences (ITD/ILD). These cues are approximated here using the spherical head model in [19]. The algorithm employs a binaural *direction-of-arrival* (DOA) estimation technique [20] to find appropriate binaural cues in order to align the phase and level of the binaural signals. We also assume an ideal diffuse sound field model for the reverberation.

The remainder of this paper is organized as follows. Section 2 describes the binaural signal model, the narrowband DRR, and the ITD/ILD model used in this paper. Section 3 introduces the proposed approach for DRR estimation using interaural MSC. Section 4 presents the evaluation results for the proposed approach and Section 5 concludes this paper.

2. BINAURAL SIGNAL MODEL

In our scenario depicted in Fig. 1 we consider binaural signals from a single source $s(n)$ received by the microphones of two HAs. Using the convolution operator $*$ the received signal at the left (L) and right (R) microphone is written as

$$y_{L/R}(n) = h_{L/R}(n, \theta_s) * s(n), \tag{1}$$

where $h_{L/R}(n, \theta_s)$ indicates the binaural RIR (BRIR) from source s at azimuth angle θ_s to the left/right microphone and n is the sampling index. Note that we do not consider the effect of ambient noise in our model. Based on the assumption that the effect of early reflections is less prominent, we may decompose the BRIR to $h_{L/R}(n, \theta_s) = h^{(\mathrm{d})}_{L/R}(n, \theta_s) + h^{(\mathrm{lr})}_{L/R}(n)$, where $h^{(\mathrm{d})}$ and $h^{(\mathrm{lr})}$ denote the direct and the late reverberant components, respectively. In a binaural configuration the direct component of the RIR is mainly characterized by the *head-related* RIR (HRIR). Therefore, for the ease of notation we replace $h^{(\mathrm{d})}_{L/R}(n, \theta_s)$ with $d_{L/R}(n, \theta_s)$ in the above equation. We thus write the signal model as

$$y_{L/R}(n) = d_{L/R}(n, \theta_s) * s(n) + r_{L/R}(n), \tag{2}$$

where $r_{L/R}(n) = s(n) * h^{(\mathrm{lr})}_{L/R}(n)$ denotes the reverberant component of the binaural signals at the left/right microphone. To analyze signals in the STFT domain, we take a $K-$point DFT on overlapped and windowed signal frames. Using matrix notation we thus obtain

$$\boldsymbol{Y}(k, \mu) = \boldsymbol{D}(k, \theta_s) S(k, \mu) + \boldsymbol{R}(k, \mu), \tag{3}$$

where k, μ denote the frequency and frame indices, respectively. In this equation, $\boldsymbol{D}(k, \theta_s) = [D_L(k, \theta_s), D_R(k, \theta_s)]^T$ indicates the vector of *head-related transfer functions* (HRTFs) for the source at azimuth position θ_s. Here, $\boldsymbol{Y}(k, \mu) = [Y_L(k, \mu), Y_R(k, \mu)]^T$ and $\boldsymbol{R}(k, \mu) = [R_L(k, \mu), R_R(k, \mu)]^T$ determine the spectral coefficients and the reverberant component of microphone signals, respectively. Furthermore, we introduce the relative HRTF by dividing the HRTF vector $\boldsymbol{D}$ by the reference (left) HRTF D_L as

$$\tilde{\boldsymbol{D}}(k, \theta_s) = [1, G(k, \theta_s) e^{j\Omega_k f_s \Delta\tau(\theta_s)}]^T, \tag{4}$$

where $G(k, \theta_s) = \left| \frac{D_R(k,\theta_s)}{D_L(k,\theta_s)} \right|$ and $\Delta\tau(\theta_s) = \tau_R(\theta_s) - \tau_L(\theta_s)$ indicate the ILD and ITD for the source angle θ_s, respectively. Moreover, $\Omega_k = 2\pi k/K$ denotes the angular frequency normalized to the sampling rate f_s using K DFT bins. We thus write the model (3) as

$$\boldsymbol{Y}(k, \mu) = \tilde{\boldsymbol{D}}(k, \theta_s) X(k, \mu) + \boldsymbol{R}(k, \mu), \tag{5}$$

where $X(k, \mu) = D_L(k, \theta_s) S(k, \mu)$ determines the non-reverberated microphone signal of the reference (left) microphone. Since our implementation works on single time-frames, we shall drop the frame index μ for convenience and reintroduce it only when necessary.
Taking the signal model (5) into account and assuming that the direct and late reverberant components of microphone signals are mutually uncorrelated, we may express the covariance matrix $\boldsymbol{\Phi_{YY}}(k) = \mathrm{E}\{\boldsymbol{Y}(k)\boldsymbol{Y}^H(k)\}$ of microphone signals as

$$\boldsymbol{\Phi_{YY}}(k) = \tilde{\boldsymbol{D}}(k, \theta_s) \tilde{\boldsymbol{D}}^H(k, \theta_s) \Phi_{XX}(k) + \boldsymbol{\Phi_{RR}}(k), \tag{6}$$

where $\Phi_{XX}(k)$ indicates the power of the non-reverberated microphone signal at the reference microphone. Moreover, the covariance matrix of the reverberation tail $\boldsymbol{\Phi_{RR}}(k)$ may be decomposed to

$$\boldsymbol{\Phi_{RR}}(k) = \Phi_{RR}(k) \begin{pmatrix} 1 & \Gamma_{R_L R_R}(k) \\ \Gamma_{R_R R_L}(k) & 1 \end{pmatrix}, \tag{7}$$

where $\Phi_{RR}(k)$ indicates the power of the reverberant components of microphone signals which is assumed to be equivalent across all microphones. Additionally, $\Gamma_{R_R R_L} = \Gamma_{R_L R_R}$ are real-valued quantities indicating the coherence of the reverberant components of microphone signals which may be determined assuming an ideal diffuse sound field as

$$\Gamma_{R_L R_R}(k) = \frac{\sin(2\Omega_k f_s a c^{-1})}{2\Omega_k f_s a c^{-1}}, \tag{8}$$

where a is the head radius and c is the speed of sound.
We thus express (6) as

$$\begin{aligned} \boldsymbol{\Phi_{YY}}(k) &= \\ &\Phi_{XX}(k) \begin{pmatrix} 1 & G(k, \theta_s) e^{j\Omega_k f_s \Delta\tau(\theta_s)} \\ G(k, \theta_s) e^{-j\Omega_k f_s \Delta\tau(\theta_s)} & G^2(k, \theta_s) \end{pmatrix} \\ &+ \Phi_{RR}(k) \begin{pmatrix} 1 & \Gamma_{R_L R_R}(k) \\ \Gamma_{R_L R_R}(k) & 1 \end{pmatrix}. \end{aligned} \tag{9}$$

Therefore, the ratio of the power of the direct and reverberant components of the left and the right microphone signals is given by $\eta_L(k) = \frac{\Phi_{XX}(k)}{\Phi_{RR}(k)}$, and $\eta_R(k) = G^2(k, \theta_s)\eta_L(k)$. We thus obtain the DRR for each frequency bin by averaging $\eta_L(k)$ and $\eta_R(k)$ as

$$\eta(k) = \frac{1 + G^2(k, \theta_s)}{2} \frac{\Phi_{XX}(k)}{\Phi_{RR}(k)}. \tag{10}$$

In order to estimate ITD and ILD cues we employ the spherical head model in [19]. Given the coordinate system in Fig. 1, the ITD and ILD cues are approximated respectively as

$$\widehat{\Delta\tau}(\theta_s) = \frac{a}{c} \left(\theta_s + \sin(\theta_s) \right), \tag{11}$$

$$\hat{G}(k, \theta_s) = \sqrt{\frac{4\omega_0^2 + \delta_R^2(\theta_s)\Omega_k^2 f_s^2}{4\omega_0^2 + \delta_L^2(\theta_s)\Omega_k^2 f_s^2}}\quad, \tag{12}$$

where $\omega_0 = c/a$ and $\delta_{L/R}(\theta_s) = 1.05 + 0.95\cos\left(1.2(\theta_s \pm \pi/2)\right)$.

3. BINAURAL DRR ESTIMATION USING MSC

A useful technique for measuring the spatial characteristics of acoustic environments is the coherence of microphone signals defined as

$$\Gamma_{Y_L Y_R}(k) = \frac{\Phi_{Y_L Y_R}(k)}{\sqrt{\Phi_{Y_L Y_L}(k)\Phi_{Y_R Y_R}(k)}}. \tag{13}$$

Taking the signal model (5) into account the coherence of the binaural signals may be written as

$$\begin{aligned} &\Gamma_{Y_L Y_R}(k) = \\ &\frac{G(k, \theta_s)\Phi_{XX}(k) e^{j\Omega_k f_s \Delta\tau(\theta_s)} + \Phi_{RR}(k)\Gamma_{R_L R_R}(k)}{\sqrt{\left(\Phi_{XX}(k) + \Phi_{RR}(k)\right)\left(G^2(k, \theta_s)\Phi_{XX}(k) + \Phi_{RR}(k)\right)}}, \end{aligned} \tag{14}$$

which is simplified with $\eta_L(k) = \frac{\Phi_{XX}(k)}{\Phi_{RR}(k)}$ to

$$\Gamma_{Y_L Y_R}(k) = \frac{G(k, \theta_s) e^{j\Omega_k f_s \Delta\tau(\theta_s)} \eta_L(k) + \Gamma_{R_L R_R}(k)}{\sqrt{(\eta_L(k) + 1)(G^2(k, \theta_s)\eta_L(k) + 1)}}. \tag{15}$$

In principle, $\Gamma_{Y_L Y_R}(k)$ is complex-valued whereas the DRR is a positive real-valued quantity. Therefore, different heuristically motivated algorithms exist to estimate DRR from the complex-valued coherence. For instance, [10, 12, 11] consider a free-field microphone array which results in $G(k, \theta_s) = 1$ and $\eta_L(k) = \eta_R(k) = \eta(k)$. The authors in [15] extends the method of [11] to a binaural con-

Table 1. A summary of DRR estimation algorithms.

algorithms	DRR cost functions $\eta(k) =$
Jeub [10]	$\frac{\Gamma_{R_L R_R}(k) - \mathrm{Re}\left\{\Gamma_{Y_L Y_R}(k)\exp\{-j\Omega_k f_s \Delta\tau(\theta_s)\}\right\}}{\mathrm{Re}\left\{\Gamma_{Y_L Y_R}(k)\exp\{-j\Omega_k f_s \Delta\tau(\theta_s)\}\right\}-1}$
Thiergart [12]	$\mathrm{Re}\left\{\frac{\Gamma_{R_L R_R}(k)-\Gamma_{Y_L Y_R}(k)}{\Gamma_{Y_L Y_R}(k)-\exp\{j\Omega_k f_s \Delta\tau(\theta_s)\}}\right\}$
Schwartz [11]	$\frac{\mathrm{Re}\left\{\exp\{-j\Omega_k f_s \Delta\tau(\theta_s)\}\left(\Gamma_{R_L R_R}(k)-\Gamma_{Y_L Y_R}(k)\right)\right\}}{\mathrm{Re}\left\{\Gamma_{Y_L Y_R}(k)\exp\{-j\Omega_k f_s \Delta\tau(\theta_s)\}\right\}-1}$
Zheng [15]	[11] using the ITD model in [16]
Proposed	(20)

figuration using only the ITD model devised in [16]. In the present work, however, we propose to estimate DRR by computing the MSC of the microphone signals and taking the effect of joint ITD and ILD into account. A summary of the state-of-the-art DRR estimation algorithms are listed in Table 1.

We propose to solve (15) for $\eta_L(k)$ by computing the MSC of the microphone signals which leads to the following quadratic equation

$$a_1\eta_L^2(k) + a_2\eta_L(k) + a_3 = 0, \tag{16}$$

where

$$a_1 = G^2(k,\theta_s)\left(1 - |\Gamma_{Y_L Y_R}(k)|^2\right), \tag{17}$$

$$\begin{aligned} a_2 = {} & 2G(k,\theta_s)\cos(\Omega_k f_s \Delta\tau(\theta_s))\Gamma_{R_L R_R}(k) \\ & - \left(1 + G^2(k,\theta_s)\right)|\Gamma_{Y_L Y_R}(k)|^2, \end{aligned} \tag{18}$$

$$a_3 = \Gamma^2_{R_L R_R}(k) - |\Gamma_{Y_L Y_R}(k)|^2. \tag{19}$$

The average DRR is thus estimated by solving the quadratic equation in conjunction with (10) which is determined by

$$\hat{\eta}^{\mathrm{binrl}}_{\mathrm{MSC}}(k) = \frac{1 + G^2(k,\theta_s)}{2}\,\frac{-a_2 \mp \sqrt{a_2^2 - 4a_1 a_3}}{2a_1}. \tag{20}$$

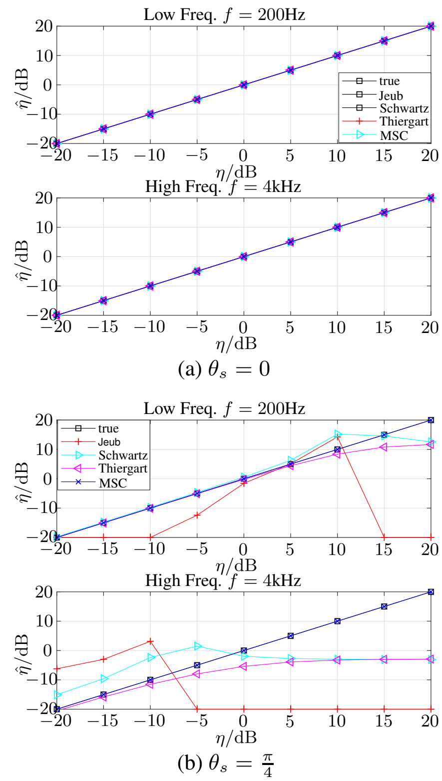
Here, the first solution (-) is always negative and thus not meaningful. Therefore, the second solution to this quadratic equation (+) indicates the estimated DRR. Under the assumption that $\theta_s = 0$, we have $G(k,\theta_s) = 1$ and $\Delta\tau = 0$. In this case, the DRR estimation using the MSC is determined by

$$\hat{\eta}^{\mathrm{FF}}_{\mathrm{MSC}}(k) = \frac{\Gamma_{R_L R_R}(k) - |\Gamma_{Y_L Y_R}(k)|}{|\Gamma_{Y_L Y_R}(k)| - 1}. \tag{21}$$

For the computation of the coherence defined in (13), $\Phi_{Y_m Y_{m'}}(k)$ is estimated using the first order recursive temporal smoothing. In order to obtain the broadband DRR we average the estimated narrowband DRRs across low frequencies $f \leq 1.8$ kHz.

4. EVALUATION RESULTS

In order to motivate the necessity of using a binaural model in DRR estimation algorithm, we evaluate the performance of the proposed binaural DRR estimation algorithm in (20) and algorithms of Jeub [10], Thiergart [12], and Schwartz [11] on simulated data. We simulate the coherence of the mixture of coherent and ideally diffuse signals as in (14) in two frequency bands. For the coherent source we use the ITD and ILD model in (11) and (12) with $a = 8.5$ cm. We compute the coherence of the reverberant signal using (8). We estimate the DRR for true DRRs in the range of $10\log_{10}(\eta) \in [-20 : 5 : 20]$ dB. Figure 2 presents the DRR estimates for one low and one high frequency example and for $\theta_s = 0, \frac{\pi}{4}$. As can be observed when the target is in front of the HA user all estimators achieve unbiased estimation results. However, when the source is not at $\theta_s = 0$ the proposed MSC-based DRR estimation algorithm delivers the most accurate and unbiased DRR estimate as the shadowing of the head is properly taken into consideration.



(a) $\theta_s = 0$

(b) $\theta_s = \frac{\pi}{4}$

Fig. 2. Comparison of coherence-based DRR estimation algorithms including Jeub [10], Thiergart [12], and Schwartz [11], and the proposed MSC-based approach (20) for a simulated data in two frequency bands and for two speaker directions $\theta_s = 0, \frac{\pi}{4}$.

In our second experiment we use synthetic BRIRs. In order to generate synthetic BRIRs we mix real HRIRs [21] and a synthetic late reverberation tail as in (2) with an appropriate DRR. The late reverberation is simulated by applying the exponentially decaying envelope to a white noise sequence $\nu(n)$ as [22, Chapter 13]

$$h^{(\mathrm{lr})}(n) = \exp\left(-\frac{3n\ln(10)}{T_{60} f_s}\right)\nu(n). \tag{22}$$

In our third experiment we use BRIRs recorded in a reverberant room of dimensions 7.5 m × 6.3 m × 3.3 m. We record BRIRs using front microphones of a pair of *behind-the-ear* HAs attached to a *Head and Torso Simulator* (HATS) [23]. Chirp signals were played back from Genelec-2029B loudspeakers which are located at various distances in the range of $d_{\mathrm{true}} \in [50 : 50 : 300]$ cm w.r.t. the HATS. The recordings were made under three reverberation times of $T_{60} = 0.4, 0.7$, and 0.9 s in that room.

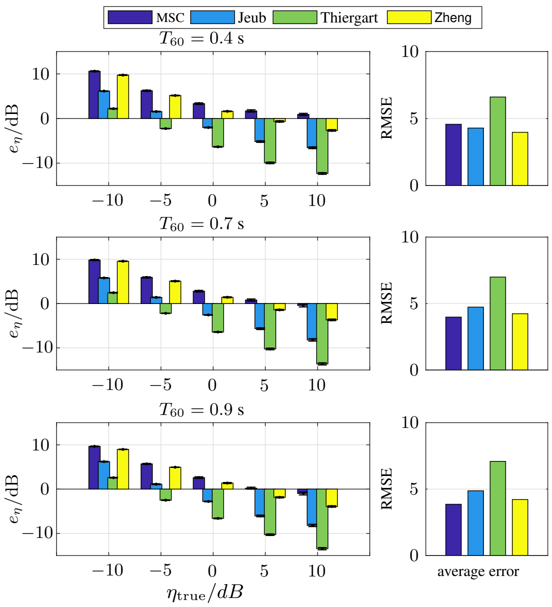


Fig. 3. Performance of DRR estimation algorithms using the proposed binaural MSC (20), Jeub [10], Thiergart [12], and Zheng [15] approaches on synthetic BRIRs for a reverberation time $T_{60} = 0.4$ s (top), $T_{60} = 0.7$ s (middle), $T_{60} = 0.9$ s (bottom) and for $\theta_s = \frac{\pi}{2}$.

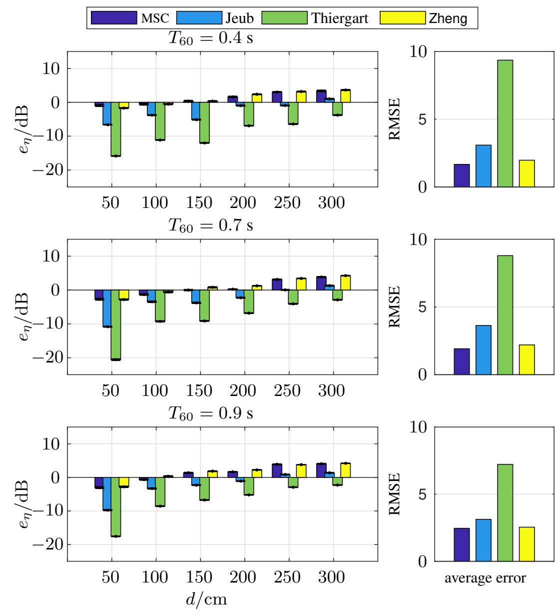


Fig. 4. Performance of DRR estimation algorithms using the proposed binaural MSC (20), Jeub [10], Thiergart [12], and Zheng [15] approaches on measured BRIRs for a reverberation time $T_{60} = 0.4$ s (top), $T_{60} = 0.7$ s (middle), $T_{60} = 0.9$ s (bottom) and for $\theta_s = \frac{\pi}{4}$.

We then convolve BRIRs with a clean speech signal from the TSP database [24] composed of a concatenation of 10 male and female speakers, each of which has a duration of 8 s. The sampling frequency is set to $f_s = 16$ kHz. Signals are segmented using a Hann window of length 32 ms with an overlap of 16 ms between successive DFT frames. The number of DFT bins equals $K = 512$.

In order to evaluate the performance of DRR estimation algorithms we compute the DRR estimation error defined as $e_\eta/\mathrm{dB} = 10\log_{10}\left(\frac{\hat{\eta}}{\eta_{\rm true}}\right)$, which shows the performance of each algorithm in terms of over/under estimation error for each scenario. Moreover, we measure the segmental root mean-square error to evaluate the overall accuracy of each DRR estimation algorithm given by $\mathrm{RMSE}_\eta/\mathrm{dB} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(10\log_{10}\left(\frac{\hat{\eta}_l}{\eta_{\rm true}}\right)\right)^2}$, where $L = 10$ is the number of concatenated speakers as outlined before. Note that, for the computation of the power of the direct path we consider 2 ms in the neighborhood of the direct component and early reflections. In order to approximate binaural cues, we first estimate DOA of the speaker using the target-beamforming technique described in [20].

We compare our results with the method of Jeub [10] and Thiergart [12] which are based on heuristic mapping of complex coherence in conjunction with the free-field propagation model and with the method of Zheng [15] which integrates an ITD model in [16] to the coherence-based DRR estimation of [11] [1].

Figure 3 presents the accuracy of DRR estimation algorithms on synthetic BRIRs for $\theta_s = \frac{\pi}{2}$. In this Figure, each bar illustrates the mean and the standard deviation of the DRR estimation error for that particular DRR. Here, the standard deviation of errors describes the robustness of the algorithm over different speech materials and speakers and is in general very small. According to this figure, the proposed MSC-based DRR estimation algorithm achieves the least RMSE on average, even though, for low DRRs the method of Jeub [10] and Thiergart [12] outperform the proposed algorithm.

Figure 4 shows the result of DRR estimation algorithms on measured BRIRs for $\theta_s = \frac{\pi}{4}$. It is observed that the proposed approach outperforms the methods of Jeub and Thiergart that verifies the utility of using the binaural model in conjunction with the MSC for DRR estimation. Furthermore, our approach achieves relatively similar performance w.r.t. the method of Zheng [15] by constraining the averaging of the estimated narrowband DRRs across low frequency bins. This implies that the closed-form MSC based solution for DRR estimation gives a comparable result w.r.t. a well-explored heuristic mapping of the complex coherence to real-valued DRR. Overall, all algorithms provide small standard deviation which verifies their robustness against different speech signals. Further results (not shown) indicate that the proposed method is robust across various speaker DOAs.

5. CONCLUSION

In this paper we proposed a DRR estimation algorithm using HA microphones. A common assumption in most DRR estimation algorithms is to model the late reverberation as an ideal diffuse sound field. We estimate DRR using interaural magnitude-squared coherence which leads analytically to a real-valued DRR taking the binaural configuration into consideration. In the proposed algorithm the phase and level of binaural signals are aligned with binaural cues approximated using a spherical head model. Results over simulated and real data corroborate the overall robustness of the proposed algorithm w.r.t. other coherence-based DRR estimation methods.

[1] The authors acknowledge Jeub et al. [10] and Zheng et al. [15] who kindly share their source codes.

6. REFERENCES

[1] B. W. Gillespie, H. S. Malvar, and D. AF. Florêncio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2001, pp. 3701–3704.

[2] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," in *Speech dereverberation*, P. A. Naylor and N. D. Gaubitch, Eds. Springer Science & Business Media, 2010.

[3] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B. H. Juang, "Blind speech dereverberation with multi-channel linear prediction based on short time Fourier transform representation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 85–88.

[4] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood PSD estimation for speech enhancement in reverberation and noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1599–1612, Sept 2016.

[5] S. Braun, A. Kuklasiński, O. Schwartz, O. Thiergart, E. A. P. Habets, S. Gannot, S. Doclo, and J. Jensen, "Evaluation and comparison of late reverberation power spectral density estimators," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1056–1071, June 2018.

[6] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *ACTA Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.

[7] Y.c. Lu and M. Cooke, "Binaural estimation of sound source distance via the direct-to-reverberant energy ratio for static and moving sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1793–1805, 2010.

[8] Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, and Y. Haneda, "Estimating direct-to-reverberant energy ratio using D/R spatial correlation matrix model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 8, pp. 2374–2384, 2011.

[9] A. Brendel and W. Kellermann, "Learning-based acoustic source-microphone distance estimation using the coherent-to-diffuse power ratio," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 61–65.

[10] M. Jeub, C. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in *European Signal Processing Conference (EUSIPCO)*, 2011, pp. 1347–1351.

[11] A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 6, pp. 1006–1018, 2015.

[12] 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 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 309–312.

[13] Y. Hioka, K. Furuya, K. Niwa, and Y. Haneda, "Estimation of direct-to-reverberation energy ratio based on isotropic and homogeneous propagation model," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*. VDE, 2012.

[14] J. Eaton, A. H. Moore, P. A. Naylor, and J. Skoglund, "Direct-to-reverberant ratio estimation using a null-steered beamformer," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 46–50.

[15] C. Zheng, A. Schwarz, W. Kellermann, and X. Li, "Binaural coherent-to-diffuse-ratio estimation for dereverberation using an ITD model," in *European Signal Processing Conference (EUSIPCO)*, 2015, pp. 1048–1052.

[16] G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," *The Journal of the Acoustical Society of America*, vol. 62, no. 1, pp. 157–167, 1977.

[17] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of room acoustic parameters: The ACE challenge," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 10, pp. 1681–1693, 2016.

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

[19] C.P. Brown and R.O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, Sep 1998.

[20] M. Zohourian, G. Enzner, and R. Martin, "Binaural speaker localization integrated into an adaptive beamformer for hearing aids," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 515–528, March 2018.

[21] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 6, 2009.

[22] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*, John Wiley & Sons, 2006.

[23] ITU-T Recommendation P.58, *Head and Torso Simulator for telephonometry*, HEAD acoustics, 2018.

[24] P. Kabal, "TSP speech database," *McGill University, Database Version*, 2002.