Estimating the Direct-to-Reverberant Energy Ratio Using a Spherical Harmonics Based Spatial Correlation Model

Prasanga N. Samarasinghe*, *Member, IEEE,* Thushara D. Abhayapala, *Senior Member, IEEE,* and Hanchi Chen, *Student Member, IEEE*

***Abstract*—The direct-to-reverberant ratio (DRR), which describes the energy ratio between the direct and reverberant component of a soundfield, is an important parameter in many audio applications. In this paper, we present a multi-channel algorithm, which utilizes the blind recordings of a spherical microphone array to estimate the DRR of interest. The algorithm is developed based on a spatial correlation model formulated in the spherical harmonics domain. This model expresses the cross correlation matrix of the recorded soundfield coefficients in terms of two spatial correlation matrices, one for direct sound and the other for reverberation. While the direct path arrives from the source, the reverberant path is considered to be a non-diffuse soundfield with varying directional gains. The direct and reverberant sound energies are estimated from the aforementioned spatial correlation model, which then leads to the DRR estimation. The practical feasibility of the proposed algorithm was evaluated using the speech corpus of the ACE (Acoustic Characterization of Environments) Challenge. The experimental results revealed that the proposed method was able to effectively estimate the DRR of a large collection of reverberant speech recordings including various environmental noise types, room types and speakers.**

I. Introduction

The direct-to-reverberant energy ratio (DRR) is one of the most important parameters when it comes to the analysis of room acoustics. It not only determines the acoustic quality of a room but also serves as an integral element in many audio applications, such as speech enhancement and dereverberation [1]–[3], source localization [4], parametric spatial audio coding [5], performance evaluation of beamforming [6] and psychoacoustics, where it is believed that the DRR helps humans to determine the distance to a sound source [4], [7], [8]. The knowledge of DRR also helps the derivation of various other acoustic parameters such as reverberation time (T_{60}), and diffuseness [9], [10].

Due to the broad usefulness of the DRR, its estimation accuracy is considered to be vital. The most primitive method to calculate the DRR is to use the room impulse response (RIR) measured by an omnidirectional microphone. Even though the DRR can be estimated using only the beginning part of RIR [11], reasons such as, the need to use intrusive signals to reliably obtain RIR measurements, the requirement to repeat RIR measurements with moving source and receiver positions and the necessity of prior processing to identify the initial part of the RIR, makes this estimation process less practical. Falk *et al.* [12] proposed an alternative method that utilizes the long-term temporal dynamics of recorded speech signals, however this method requires an *a priori* calculation of a pre-defined energy ratio, which varies with changing acoustic environments.

In order to avoid pre-processing, several authors have more recently investigated the use of so-called blind methods that do not require special measurement signals or a priori information, but instead utilizes the microphone recordings of the reverberant soundfield to directly estimate the DRR. There exists two main approaches to blind DRR estimation, namely *power based estimation* and *coherence based estimation*. In order to simplify the problem formulation, both of these approaches usually assume the reverberant soundfield to be diffuse.

Power based estimators utilize the power spectral density (PSD) of two or more beamformer signals to derive the direct and reverberant signal PSDs. The authors in [13] proposed a PSD estimator using a microphone array and two identical beamformers while the authors of [14] proposed a similar method using multiple directional microphones. Hioka *et al.* [15] recently proposed an improved PSD estimator that uses a microphone array and multiple arbitrary beamformers.

In contrast, coherence based methods use the cross-correlation between two or more microphone signals and appropriate signal processing algorithms to estimate the DRR. They mostly utilizes the differences in propagation properties of direct and reverberant sound to estimate the direct path power and reverberant/diffuse field power separately, which then leads to the estimated DRR. The authors of [4] proposed a binaural system where, through cross-correlation, the direct sound is removed from one channel by subtracting a filtered version of the second channel. The drawback of this approach is that not only the direct path energy but also part of the reverberant path energy flowing from the source direction also gets counted towards the estimated direct sound energy. A multi-channel coherence based method is presented in [16] where the spatial correlation matrix of the microphone array is expressed in terms of two spatial correlation matrices, one for direct sound and one for reverberation. The least squares method is used to derive the direct and reverberant sound energies. Instead of estimating the direct and reverberant soundfield energies separately, some recent coherence based methods use analytically derived relationships between the DRR and the magnitude squared coherence (MSC) function of the microphone array [7], [8] or the MSC between the coincident pressure and particle velocity [17]–[19]. The use of

these analytical relationships avoids the need of least squares based numerical solving. However in general, all coherence based approaches suffer from high estimation variance at low frequencies since the omnidirectional microphone signals are strongly correlated even if the soundfield is diffuse. Addressing this problem, later work proposed the use of directional microphones as in [20] or the use of coherence between eigenbeams/spherical harmonic coefficients instead of direct microphone signals as in [18], [19], [21].

In this paper, we propose a multi-channel coherence estimator based on a spatial correlation model formulated in the spherical harmonics domain. The spatial correlation model is similar to that of [16], but now the correlation matrix is between the spherical harmonic coefficients/eigenbeams derived from a spherical microphone array[1]. Apart from that, the current method assumes a more realistic reverberant soundfield that is non-diffuse such that different reflective surfaces may have different directional gains. This can be considered as the most novel concept compared to all of the aforementioned DRR estimators. We evaluate the performance of this method on the speech corpus of the ACE (Acoustic Characterization of Environments) Challenge) [24], [25]. To investigate the advantages of considering a non-diffuse reverberant field, we carried out a performance comparison with the method in [16]. Furthermore, since the problem specified by the ACE challenge was a fully blind problem (i.e., no prior information about the source direction is provided), the methods were tested after being combined with a conventional direction-of-arrival (DOA) estimation method.

This paper is organized as follows. In Section II, we formulate the DRR estimation problem. In Section III, we derive a spatial correlation matrix in the spherical harmonics domain. In Section IV, we derive the estimated DRR from the spatial correlation matrix derived earlier. Finally, in Section V, we present the results obtained from the proposed estimation algorithm. Comments on the outcome of this study and future work conclude this paper.

II. Problem Formulation

We first consider a spherical array of Q omnidirectional microphones recording the incident soundfield caused by a single source inside the room enclosure of interest. The observed soundfield at the q^{th} $(q = 1, 2, \cdots, Q)$ microphone can be expressed in the time-frequency domain as

$$P(\boldsymbol{x}_q, k, t) = S(k, t)H(\boldsymbol{x}_q, \boldsymbol{y}_o, k) \tag{1}$$

where $k = 2\pi f/c$ is the wavenumber with f and c representing the frequency in Hz and speed of sound in ms^{-1} respectively, $S(k, t)$ is the Short Time Fourier transform of the source signal, t is the temporal frame index, and $H(\boldsymbol{x}_q, \boldsymbol{y}_o, k)$ is the room transfer function (RTF) between the source location $\boldsymbol{y}_o = (r_0, \theta_o, \phi_0)$ and the receiver location $\boldsymbol{x}_q = (r, \theta_q, \phi_q)$. Note that from now on, we omit the time dependency (t) for notational convenience.

In a reverberant enclosure, $P(\boldsymbol{x}_q, k)$ would comprise of the direct path from the source as well as the reverberant path caused by room reflections. This decomposition can be reflected in the RTF as

$$H(\boldsymbol{x}_q, \boldsymbol{y}_o, k) = H_{\text{dir}}(\boldsymbol{x}_q, \boldsymbol{y}_o, k) + H_{\text{rvb}}(\boldsymbol{x}_q, \boldsymbol{y}_o, k) \tag{2}$$

where $H_{\text{dir}}(\cdot)$ and $H_{\text{rvb}}(\cdot)$ represent the direct and reverberant components of the room impulse response, respectively.

Assuming the aperture size of the microphone array is sufficiently small compared to the distance to the source[2], the direct path of (2) can be considered to be a single plane wave of the form

$$H_{\text{dir}}(\boldsymbol{x}_q, \boldsymbol{y}_o, k) = H_D(k)e^{ik\hat{\boldsymbol{y}}_o \cdot \boldsymbol{x}_q} \tag{3}$$

where $H_D(k)$ denotes the direct path gain and $\hat{\boldsymbol{y}}_o$ denotes the unit vector along the incoming direction. Although a common practice in DRR estimation algorithms is to assume the corresponding reverberant path to be a diffuse field, it is not a realistic model in practical acoustic environments due to the non-isotropic gain distribution among reflective surfaces. Therefore, we consider a more generalized model for the reverberant path of (2) in terms of

$$H_{\text{rvb}}(\boldsymbol{x}_q, \boldsymbol{y}_o, k) = \int_{\hat{\boldsymbol{y}}} H_R(k, \hat{\boldsymbol{y}})e^{ik\hat{\boldsymbol{y}} \cdot \boldsymbol{x}_q} d\hat{\boldsymbol{y}} \tag{4}$$

where $H_R(k, \hat{\boldsymbol{y}})$ is the gain of the reflected plane wave arriving from the direction $\hat{\boldsymbol{y}} = (1, \theta, \phi)$ for $\theta \in [0 : \pi]$ and $\phi \in [0 : 2\pi)$ and $\int_{\hat{\boldsymbol{y}}} d\hat{\boldsymbol{y}} = \int_0^{2\pi} \int_0^{\pi} \sin\theta d\theta d\phi$. From (1), (2), (3) and (4), we re-write $P(\boldsymbol{x}_q, k)$ as

$$P(\boldsymbol{x}_q, k) = S(k)\Big(H_D(k)e^{ik\hat{\boldsymbol{y}}_o \cdot \boldsymbol{x}_q} + \int_{\hat{\boldsymbol{y}}} H_R(k, \hat{\boldsymbol{y}})e^{ik\hat{\boldsymbol{y}} \cdot \boldsymbol{x}_q} d\hat{\boldsymbol{y}}\Big). \tag{5}$$

By observing the above derivation, the ratio between the direct path energy P_D and the reverberant path energy P_R can be expressed as

$$\text{DRR} = \frac{P_D}{P_R} = \frac{E\Big\{ \|S(k)\|^2 \|H_D(k)\|^2 \Big\}}{\int_{\hat{\boldsymbol{y}}} E\Big\{ \|S(k)\|^2 \|H_R(k, \hat{\boldsymbol{y}})\|^2 \Big\} d\hat{\boldsymbol{y}}} \tag{6}$$

where $E\{\cdot\}$ denotes the expectation operator and $\|\cdot\|$ is the $2-$norm. Our goal is to estimate the above ratio utilizing the spherical microphone array recordings. Since spherical array based signal processing techniques are well established in the spherical harmonics domain, we first utilize the spherical harmonic decomposition of the reduced wave equation (Helmholtz wave equation) to derive a set of coefficients defining the spatial soundfield enclosed by the array of interest. We then use these coefficients to construct a spatial correlation matrix model, which provides a pathway to estimate P_D, P_R and eventually the desired DRR.

[1] Or other alternative structures as given in [22], [23]

[2] When the aperture size is much smaller than the radiating wavelength, it is said to be a *point source*, which radiates power equally in all directions with a spherical radiation pattern. At great distances with respect to wavelength from the source, the spherically spreading waves can be regarded as plane waves forming a far-field. A common rule of thumb is far-field sources are located at a distance of $r > 2L^2/\lambda$ where L is the aperture radius and λ is the operating wavelength.

III. Spherical Harmonics based spatial correlation

In this section, we use the spherical harmonic decomposition of soundfields to derive a similar relationship to (5), and utilize it to formulate a closed form expression for spatial correlation of reverberant soundfield recordings.

A. Spherical harmonic decomposition of wavefields

Spherical harmonics are a set of orthonormal spatial basis functions, which can be used to represent functions defined over a sphere. Thus, any spherical function $f(\theta,\phi)$ may be expanded as a linear combination of these basis functions in the form of

$$f(\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}a_{nm}Y_{nm}(\theta,\phi). \tag{7}$$

where $Y_{nm}(\cdot)$ and a_{nm} denote the spherical harmonic functions and the corresponding coefficients, respectively. The spherical harmonics are inherently orthonormal, and hence

$$\int_{\hat{\boldsymbol{y}}}Y_{nm}(\theta,\phi)Y^{*}_{n'm'}(\theta,\phi)d\hat{\boldsymbol{y}}=\delta_{nn'}\delta_{mm'}. \tag{8}$$

where * denotes the complex conjugate operator and δ denotes the dirac delta function.

As mentioned earlier, our intention is to derive a similar relationship to (5) in terms of the spherical harmonic decomposition given in (7) to assist the proposed estimation model. The left-hand-side of (5) is the incident soundfield over a spherical surface outlined by the microphone array, and therefore can be expressed in a similar form to (7) as

$$P(\boldsymbol{x}_q,k)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\underbrace{\alpha_{nm}(k)b_n(kr)}_{a_{nm}(k)}Y_{nm}(\theta_q,\phi_q). \tag{9}$$

where k is introduced to represent the frequency dependance, r is the radius of the spherical microphone array (e.g., Eigenmike), $a_{nm}(k)=\alpha_{nm}(k)b_n(kr)$ is a further simplification of the observed incident soundfield based on the assumption that it is a homogeneous incident soundfield [26], [27] and

$$b_n(kr)=\begin{cases}j_n(kr) & \text{for an open array}\\ j_n(kr)-\frac{j'_n(kr)}{h'_n(kr)}h_n(kr) & \text{for a rigid array}\end{cases} \tag{10}$$

with $j_n(\cdot)$ and $h_n(\cdot)$ denoting the spherical Bessel and Hankel functions of order n respectively. Note that $\alpha_{nm}(k)$, the incident soundfield coefficients, can be derived up to order $N=\lceil kr\rceil$ using the microphone array recordings $P(\boldsymbol{x}_q,k)$ for $q=1,2\cdots Q$ [26], [27].

Similarly, the spherical functions in the right-hand-side of (5) can be decomposed in terms of spherical harmonics. These include the reverberant gain function $H_R(k,\hat{\boldsymbol{y}})$ distributed over all possible look directions, and the plane wave soundfields $e^{ik\hat{\boldsymbol{y}}_o\cdot\boldsymbol{x}_q}$ and $e^{ik\hat{\boldsymbol{y}}\cdot\boldsymbol{x}_q}$, as observed by the spherical microphone array. We write $H_R(k,\hat{\boldsymbol{y}})$ in terms of

$$H_R(k,\hat{\boldsymbol{y}})=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\beta_{nm}(k)Y_{nm}(\theta,\phi). \tag{11}$$

where $\beta_{nm}(k)$ are the respective spherical harmonic coefficients, and $e^{ik\hat{\boldsymbol{y}}\cdot\boldsymbol{x}_q}$ by

$$e^{ik\hat{\boldsymbol{y}}\cdot\boldsymbol{x}_q}=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\underbrace{i^nY^{*}_{nm}(\theta,\phi)j_n(kr)}_{a_{nm}(k)}Y_{nm}(\theta_q,\phi_q). \tag{12}$$

where the spherical harmonic coefficients are known for a given planewave incident direction [26].

By substituting (9), (11) and (12) in (5), we derive a modal domain relationship analogous to (5) as

$$\alpha_{nm}(k)=S(k)i^n\big(H_D(k)Y^{*}_{nm}(\theta_0,\phi_0)+\beta_{nm}(k)\big). \tag{13}$$

where the soundfield coefficients recorded by the microphone array are now represented in terms of their respective direct and reverberant components. Please refer to Appendix A for a detailed derivation of the above result. This relationship serves as the basis for the spatial correlation expression formulated in the following section.

B. Spatial correlation in the spherical harmonic domain

Here, we derive a spatial correlation expression in terms of the spatial soundfield coefficients recorded by the microphone array. Based on (13), the cross correlation between α_{nm} and $\alpha_{n'm'}$ is

$$\begin{aligned}&E\Big\{\alpha_{nm}(k)\alpha^{*}_{n'm'}(k)\Big\}=i^n(-i)^{n'}E\Big\{\left\|S(k)\right\|^2\Big\}\\&\Big(E\Big\{\left\|H_D(k)\right\|^2\Big\}Y^{*}_{nm}(\theta_0,\phi_0)Y_{n'm'}(\theta_0,\phi_0)\\&+E\Big\{H_D(k)\beta^{*}_{n'm'}(k)\Big\}Y^{*}_{nm}(\theta_0,\phi_0)+\\&E\Big\{\beta_{nm}(k)H^{*}_D(k)\Big\}Y_{n'm'}(\theta_0,\phi_0)+E\Big\{\beta_{nm}(k)\beta^{*}_{n'm'}(k)\Big\}\Big).\end{aligned} \tag{14}$$

Due to the autonomous behavior of the reflective surfaces in a room (i.e., the reflection gain from reflective surfaces are independent from the direct path gain), the cross correlation between the direct path gain and reverberant path gain coefficients (second and third components of the RHS of (14)) can be assumed to be negligible. Under this assumption, (14) simplifies into

$$\begin{aligned}&E\Big\{\alpha_{nm}(k)\alpha^{*}_{n'm'}(k)\Big\}=i^n(-i)^{n'}E\Big\{\left\|S(k)\right\|^2\Big\}\\&\Big(E\Big\{\left\|H_D(k)\right\|^2\Big\}Y^{*}_{nm}(\theta_0,\phi_0)Y_{n'm'}(\theta_0,\phi_0)\\&\qquad+E\Big\{\beta_{nm}(k)\beta^{*}_{n'm'}(k)\Big\}\Big).\end{aligned} \tag{15}$$

The term $E\Big\{\beta_{nm}(k)\beta^{*}_{n'm'}(k)\Big\}$ of the above equation is preferred to be further simplified to arrive at a tractable estimation for DRR based on (15). Based on a second assumption that the reflection gains from different incoming directions are uncorrelated, that is

$$E\Big\{H_R(k,\hat{\boldsymbol{y}})H^{*}_R(k,\hat{\boldsymbol{y}}')\Big\}=\begin{cases}E\Big\{\left\|H_R(k,\hat{\boldsymbol{y}})\right\|^2\Big\} & \hat{\boldsymbol{y}}=\hat{\boldsymbol{y}}'\\0 & \hat{\boldsymbol{y}}\neq\hat{\boldsymbol{y}}',\end{cases} \tag{16}$$

and by utilizing a spherical harmonic decomposition for the function $E\Big\{ \|H_R(k,\hat{\boldsymbol{y}})\|^2 \Big\}$ of (16) as

$$E\Big\{ \|H_R(k,\hat{\boldsymbol{y}})\|^2 \Big\} = \sum_{v=0}^{\infty}\sum_{u=-v}^{v} \gamma_{vu}(k) Y_{vu}(\hat{\boldsymbol{y}}), \tag{17}$$

where $\gamma_{vu}(k)$ are the corresponding spherical harmonic coefficients, we derive a closed form expression for the term $E\Big\{\beta_{nm}(k)\beta^*_{n'm'}(k)\Big\}$ of (15) as

$$E\Big\{\beta_{nm}(k)\beta^*_{n'm'}(k)\Big\} = \sum_{v=0}^{\infty}\sum_{u=-v}^{v} \gamma_{vu}(k) \Big(\frac{(2v+1)(2n+1)(2n'+1)}{4\pi}\Big)^{1/2} W_1 W_2 \tag{18}$$

where W_1 and W_2 are Wigner coefficients, representing

$$W_1 = \begin{pmatrix} v & n & n' \\ 0 & 0 & 0 \end{pmatrix} \text{ and} \tag{19}$$

$$W_2 = \begin{pmatrix} v & n & n' \\ u & m & -m' \end{pmatrix}. \tag{20}$$

Please refer to Appendix B for a detailed derivation of the above result (18). Now we substitute (18) in (15) to arrive at

$$E\Big\{\alpha_{nm}(k)\alpha^*_{n'm'}(k)\Big\} = i^n(-i)^{n'} E\{\|S(k)\|^2\} \big(E\{\|H_D(k)\|^2\} Y^*_{nm}(\theta_0,\phi_0) Y_{n'm'}(\theta_0,\phi_0) + \sum_{v,u} \gamma_{vu}(k) \Big(\frac{(2v+1)(2n+1)(2n'+1)}{4\pi}\Big)^{1/2} W_1 W_2\big) \tag{21}$$

The above result provides a comprehensive expression for the spatial correlation between two spherical harmonic coefficients of an enclosed soundfield, in terms of its direct path component $H_D(k)$ and reverberant path component $\gamma_{vu}(k)$. It can be utilized in any room acoustic application that seeks the separation of direct and reverberant soundfields. In the following section, we use the above result to estimate the desired DRR.

C. Spatial Correlation matrix

We define the *modal domain spatial correlation matrix* $\boldsymbol{R}(k)$ by

$$\boldsymbol{R}(k) \equiv= E\big\{\boldsymbol{\alpha}(k)\boldsymbol{\alpha}^H(k)\big\} \tag{22}$$

where $\boldsymbol{\alpha}(k) = \big[\alpha_{00}(k) \quad \alpha_{1-1}(k) \quad .. \quad \alpha_{NN}(k)\big]^T_{1\times(N+1)^2}$. By substituting (21) to (22), we obtain

$$\boldsymbol{R}(k) = P_D \begin{bmatrix} b_{0000} & b_{001-1} & \cdots & b_{00NN} \\ b_{1-100} & b_{1-11-1} & \cdots & b_{1-1NN} \\ \vdots & \vdots & \vdots & \vdots \\ b_{NN00} & b_{NN1-1} & \cdots & b_{NNNN} \end{bmatrix} + E\Big\{ \|S(k)\|^2 \Big\} \begin{bmatrix} \boldsymbol{d}_{0000} & \boldsymbol{d}_{001-1} & .. & \boldsymbol{d}_{00NN} \\ \boldsymbol{d}_{1-100} & \boldsymbol{d}_{1-11-1} & .. & \boldsymbol{d}_{1-1NN} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{d}_{NN00} & \boldsymbol{d}_{NN1-1} & .. & \boldsymbol{d}_{NNNN} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{1-1} \\ \vdots \\ \gamma_{VV} \end{bmatrix} \tag{23}$$

where $b_{nmn'm'} = Y^*_{nm}(\theta_0,\phi_0) Y_{n'm'}(\theta_0,\phi_0)$,

$\boldsymbol{d}_{nmnm'} = [d_{nmnm'00}\ d_{nmnm'1-1}\ d_{nmnm'10}\ d_{nmnm'11} \cdots d_{nmnm'VV}]_{(N+1)^2(V+1)^2\times 1}$, and

$$d_{nmnm'vu} = \Big(\frac{(2v+1)(2n+1)(2n'+1)}{4\pi}\Big)^{1/2} W_1 W_2.$$

IV. DRR estimation using a spatial correlation model

In this section, we show how to estimate the desired DRR from the spatial correlation matrix (23). In order to utilize (23) to find the desired DRR, P_D and P_R needs to be individually estimated (both are defined in (6)). While the term P_D is present in (23), P_R is not. Therefore, we ought to derive an equivalent expression for P_R such that its extractable from (23). To achieve this, we utilize the definition of P_R given in (6), the harmonic decomposition in (17) and the symmetrical property of spherical harmonics to derive

$$\begin{aligned} P_R &= E\Big\{ \|S(k)\|^2 \Big\} \int_{\hat{\boldsymbol{y}}} E\Big\{ \|H_R(k,\hat{\boldsymbol{y}})\|^2 \Big\} d\hat{\boldsymbol{y}} \\ &= E\Big\{ \|S(k)\|^2 \Big\} \sum_{v=0}^{V}\sum_{u=-v}^{v} \gamma_{vu}(k) \int_{\hat{\boldsymbol{y}}} Y_{vu}(\hat{\boldsymbol{y}}) d\hat{\boldsymbol{y}} \\ &= E\Big\{ \|S(k)\|^2 \Big\} \gamma_{00}(k) \end{aligned} \tag{24}$$

Note that the above result can be used to replace the first entry $\gamma_{00}(k)$ of the last matrix in (23) by $P_R / E\Big\{ \|S(k)\|^2 \Big\}$. Therefore, (23) now embodies both P_D and P_R which are the vital components to estimate the desired DRR.

Since the spherical microphone array characteristics are initially known, $\boldsymbol{\alpha}(k)$ in (23) can be calculated following the method given in [27], [28]. If the direction of arrival is known or estimated, then $b_{nmn'm'}$ can also be calculated. $\boldsymbol{d}_{nmnm'}$ are composed of known functions. Thus, we can estimate the unknown power spectra P_D and P_R by solving the following set of equations, which were derived by reformulating (23) as

$$\underbrace{\begin{bmatrix} R_{0000} \\ R_{001-1} \\ \vdots \\ R_{00NN} \\ R_{1-100} \\ \vdots \\ R_{NNNN} \end{bmatrix}}_{\widetilde{\boldsymbol{r}}(k)} \approx \underbrace{\begin{bmatrix} b_{0000} & d_{000000} & \cdots & d_{0000VV} \\ b_{000-1} & d_{000-100} & \cdots & d_{000-1VV} \\ \vdots & \vdots & \vdots & \vdots \\ b_{00NN} & d_{00NN00} & \cdots & d_{00NNVV} \\ b_{1-100} & d_{1-10000} & \cdots & d_{1-100-1VV} \\ \vdots & \vdots & \vdots & \vdots \\ b_{NNNN} & d_{NNNN00} & \cdots & d_{NNNNVV} \end{bmatrix}}_{\boldsymbol{B}(k)} \times \underbrace{\begin{bmatrix} P_D \\ P_R \\ E\Big\{ \|S(k)\|^2 \Big\}\gamma_{1-1} \\ \vdots \\ E\Big\{ \|S(k)\|^2 \Big\}\gamma_{VV} \end{bmatrix}}_{\boldsymbol{p}(k)}.$$

(25)

Here, $R_{nmn'm'}$ in $\widetilde{\boldsymbol{r}}(k)$ denotes the $(n^2+n+m+1)^{th}$ row and $(n'^2+n'+m'+1)^{th}$ column components of $\boldsymbol{R}(k)$, which can be calculated from the spherical microphone measurements. The estimated power spectra of the direct and reverberant components can be derived by solving (25) using the least-squares method

$$\hat{\boldsymbol{p}}(k) = \boldsymbol{B}^{\dagger}(k)\widetilde{\boldsymbol{r}}(k) \tag{26}$$

where $[\cdot]^{\dagger}$ and $\hat{[\cdot]}$ represents the Pseudo-inverse and estimated value, respectively. The first and second elements of $\hat{\boldsymbol{p}}(k)$ then gives the desired DRR estimate

$$\mathrm{DRR}_{\mathrm{est}}(k) = 10\log_{10}\frac{\hat{P}_D(k)}{\hat{P}_R(k)}. \tag{27}$$

The basic formulation of this estimation is similar to that given in [16], however, the performance of the present model is expected to be more robust due to (i) the spatial characteristics of the correlation matrix (23), and (ii) the realistic propagation model used for the reverberated field. Note that the solution to (26) provides higher order γ terms which are redundant in the current application. These terms are essentially the higher order harmonic components of the angular power of reverberation (17), which could be useful to calculate the DRR between a point source and a V^{th} order spatial receiver region.

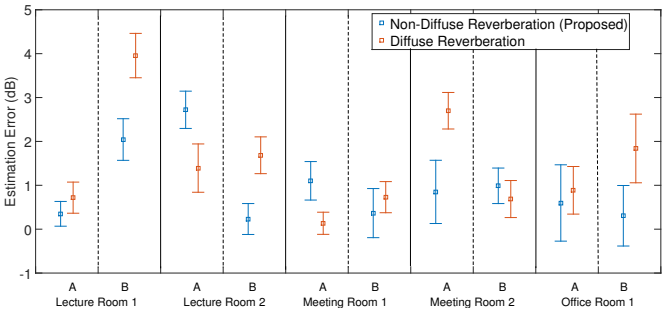


Fig. 1. Mean and standard deviation of the full-band DRR estimation error for 5 rooms and 2 microphone configurations.

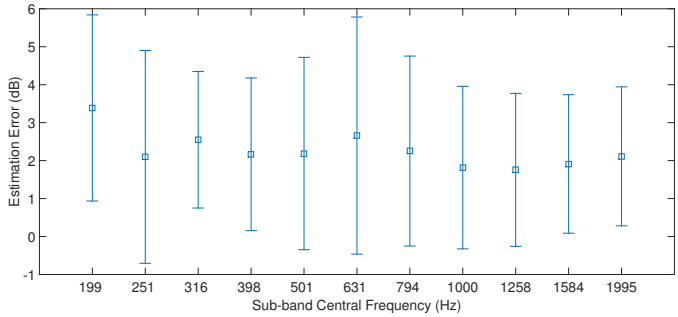


Fig. 2. Mean and standard deviation of the sub-band DRR estimation error averaged over all room configurations.

V. Experiment results using the ACE Challenge database

We evaluate the proposed DRR estimation algorithm with real acoustic data. We used the ACE Challenge database [24], [25] to retrieve a large corpus of multi-channel recordings spanned over a variety of rooms, speakers and environmental conditions.

A. The ACE Challenge database

The ACE Challenge database[3] is a recently developed database to stimulate research in non-intrusive acoustic parameter estimation in realistic environments including noise and reverberation. The database comprises of a variety of anechoic speech measurements, multi-channel room impulse responses (RIR), and multi-channel ambient, fan and live babble noise measurements recorded under the same conditions as the AIRs. A detailed description of the database can be found in [24], [25]. The results and analysis provided in this paper are based on a subset (Eval-EM32) of the above data, which were recorded using an Eigenmike. Included in this subset are; anechoic speech recordings of 5 male talkers and 5 female talkers, each uttering 5 different English phrases, RIR measurements for 5 different rooms at 2 different Eigenmike positions per room (near and far from the source position), and 3 types of noise recordings (babble, fan, ambient) under the same room conditions. The ACE database also supplies a software to construct realistic multi-channel noisy reverberant speech utterances from the above recordings by convolving the various elements of anechoic speech with the RIRs followed by adding the associated noise measurements at 3 different SNR levels (-1 dB, 12 dB and 18 dB). A resulting total of 4500 different speech utterances were utilized to evaluate the proposed DRR estimation algorithm. For performance comparison, the ground truth values for both full-band DRR and sub-band DRR are provided in the ACE database. In [24], the authors mentioned that these ground truth values were determined based on the the method given in [29]. The frequency sub-bands follows the ISO specifications for $1/3-$octave sub-bands, such that the center-frequency (CF) of band 1 is at 25.1189 Hz, CF of band 2 is at 31.6227766 Hz and so on.

[3]Available at www.ace-challenge.org

B. DRR algorithm setup and DOA Estimation

The DRR estimation algorithm involves the implementation of the spatial correlation model given in (25). This requires the formulation of the modal domain correlation matrix $\widetilde{\boldsymbol{r}}(k)$ and the secondary matrix $\boldsymbol{B}(k)$. A detailed description of our approach to this formulation is as follows.

The modal domain coefficient matrix $\widetilde{\boldsymbol{r}}(k)$ gives the correlation between the spherical harmonic coefficients of the spatial region of interest. The available measured data were the 32 channel noisy reverberant Eigenmike recordings from the ACE Challenge database. We first windowed each recording with a window length of 4 ms to balance out the trade-off between spectral leakage and reduced frequency resolution. We next discarded the windows containing speech intervals or non-speech segments to arrive at a better DRR estimation. Later we performed Short-Time Fourier Transform on the signal(s)

TABLE I
ROOM DIMENSIONS (APPROX.), MEAN T_{60}, AND MEAN DRR ACROSS ALL MICROPHONE POSITIONS, CONFIGURATIONS, AND CHANNELS

Name	L (m)	W (m)	H (m)	Vol. (m)	T60 (S)	Pos. 1 DRR min. (dB)	Pos. 1 DRR max. (dB)	Pos. 2 DRR min. (dB)	Pos. 2 DRR max. (dB)
Lecture Room 1	6.93	9.73	3.00	202	0.638	-0.825	14.6	0.872	7.90
Lecture Room 2	13.4	9.29	2.94	365	1.22	-0.370	12.8	-3.75	6.45
Meeting Room 1	6.61	4.73	2.95	92.2	0.437	-1.98	10.8	-3.10	7.57
Meeting Room 2	10.3	9.17	2.63	249	0.371	-2.57	11.2	-1.08	12.5
Office 2	5.10	3.22	2.94	48.3	0.390	-0.444	13.0	-2.28	9.48

along with $1/3$ Octave banding. For each frequency bin belonging to each sub-band we then calculated the spherical harmonic coefficients $\alpha_{nm}(k)$ using the 32-channel Eigenmike data. For this purpose we followed the mode matching approach given in [27], and the spatial soundfield order was determined according to the rule-of-thumb $N = \lceil kr \rceil$ [30]. From $\alpha_{nm}(k)$ we then constructed the spatial correlation matrix $\widetilde{\boldsymbol{r}}(k)$. The frequency variants of $\widetilde{\boldsymbol{r}}(k)$ were appropriately averaged to arrive at individual sub-band results.

We only considered a frequency range between $178 - 2239$ Hz (or CFs $199.52 - 1995.26$ Hz) containing a total of 11 $1/3-$octave bands. The elimination of low frequencies was due to the inherent nature of the human voice spectrum and the difficulty of obtaining sensible DRR estimations in noisy conditions. Due to the inherent properties of speech signals, we assume that there exists sufficient energy within the chosen frequency range. For full-band estimations we averaged the results over the 11 bands of interest. (When comparing the full-band results with the ground truth, we performed a similar averaging for the respective sub-band ground truth values instead of using the direct full-band ground truth provided by the database).

The formulation of matrix $\boldsymbol{B}(k)$ is straightforward as its elements are known functions. The only unknown we needed to estimate was the direction-of-arrival (DOA) which was not a given in the ACE database. To achieve this, we performed frequency smoothed MUSIC DOA estimation in the spherical harmonics domain [31], [32]. When processing the raw data for MUSIC estimation, we discarded all windows except for those containing the beginning of each utterance such that the corresponding impinging signal was almost purely due to the direct path.

C. Full-band DRR estimation results

The overall results for full-band DRR estimation are shown in Fig. 1. This figure shows the mean and standard deviation of the absolute value of the DRR estimation error. Since the DRR itself is a ratio between two values, it's reasonable to evaluate the error as a proportion of the estimate to the ground truth given by the ACE database. As the DRR is defined in units of decibels in the ACE database, the proportion between the estimated DRR and ground truth DRR corresponds to the difference between $\text{DRR}_{\text{estimate}}$ and $\text{DRR}_{\text{ground}}$. Thus as an evaluation criterion, we calculated the DRR estimation error ϵ_{DRR} by

$$\epsilon_{DRR} = |\text{DRR}_{\text{estimate}} - \text{DRR}_{\text{ground}}| \tag{28}$$

The results are computed over all 5 rooms including 2 microphone configurations for each room. The 2 configurations are explained in the ACE database as 'A=long' and 'B=short', interpreting a large and a small distance between the source and Eigenmike, respectively. The actual distance or their consistency across different rooms are however not disclosed.

In order to study the effectiveness of modeling the reverberant soundfields as non-diffuse, we compared the DRR estimation error with that obtained by Hioka et. al. [16], where the reverberant component of (5) was assumed to be a diffuse field. One of the reasons why the environment may not be completely diffuse is because some rooms may have carpet floors and other surfaces, which absorb more than the reflective walls and ceilings. When this happens, the spatial correlation expression formulated in [16] does not hold because not all reverberant waves arrive equally from all directions. While the proposed DRR estimation method avoids the above assumption, it may have other drawbacks due to (i) truncation error (ii) the numerical noise related to matrix inversion. The reflection of the above concerns in Fig. 1 is discussed as follows.

We can observe from Fig. 1 that out of the 10 different room configurations, the modal DRR estimation method yields better results for a majority of 7 cases with the error improvement varying between $0.5 - 2$ dB. This reflects that in general, the effect of using truncated modal expressions for the spatial correlation model and the related numerical noise are less problematic than a diffuse reverberation field assumption. For both methods, the results are weakest in Lecture Room 1, Lecture Room 2 and Meeting Room 2. These rooms are the largest and therefore, the estimation error may be caused by incorrect reverberant field estimation. This is due to the environment becoming close to anechoic with increasing room size.

D. Subband DRR estimation results

Here, we present the sub-band DRR results for $1/3-$octave bands with central frequencies ranging from $199 - 1995$ Hz. Figure 2 shows the mean and standard deviation of the sub-band DRR estimation error (28) for the proposed method. Similar to the full-band results, the results were averaged

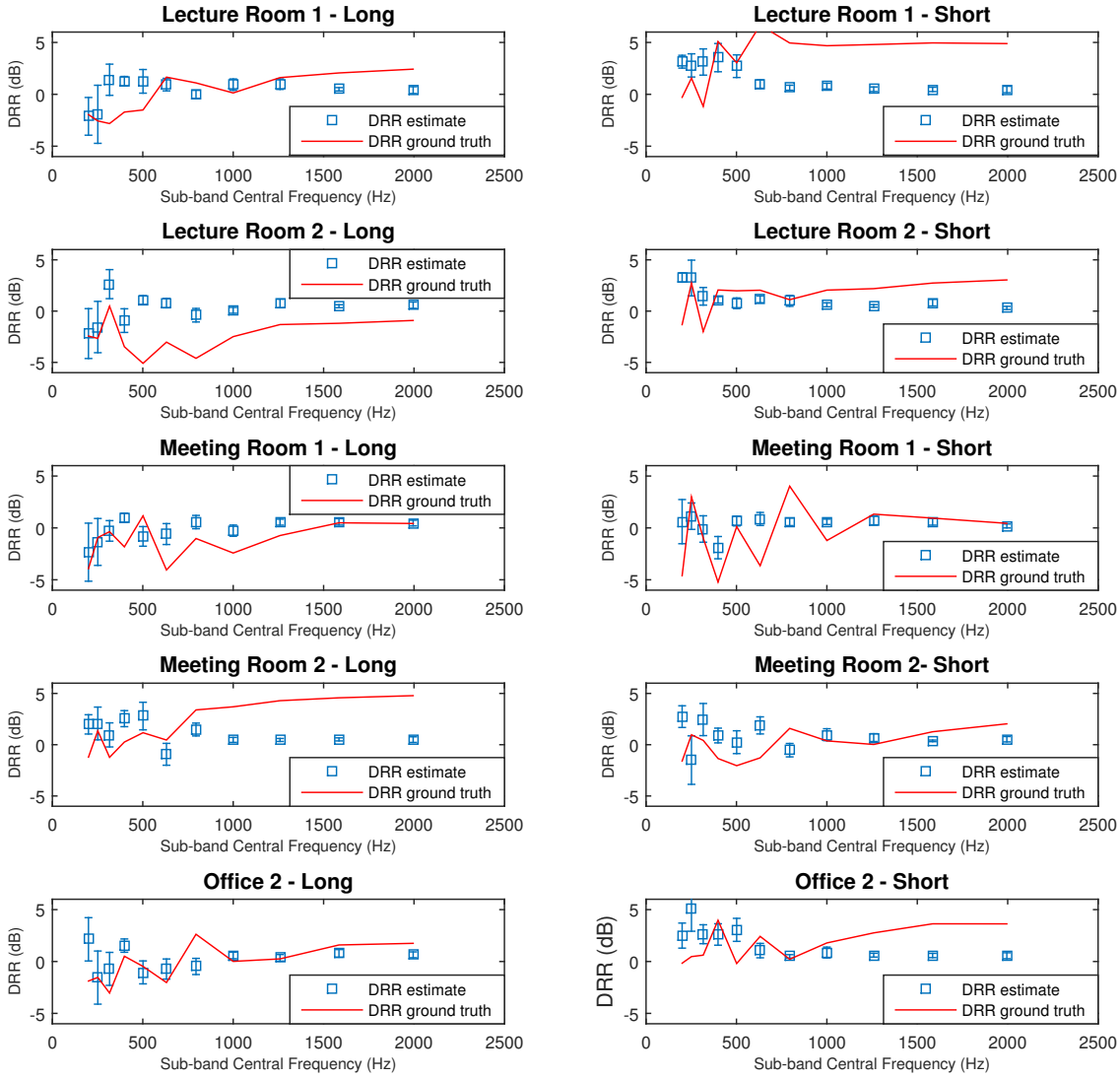


Fig. 3. Mean and standard deviation of the estimated DRR for individual rooms and microphone configurations.

over 4500 speech files spanning 5 rooms, 2 microphone configurations, 10 speakers, 5 utterances, 3 noise types and 3 noise levels.

Figure 2 indicates that the DRR estimation error for the proposed method is is fairly stable around 2dB throughout the frequency range of interest. A slight decrease of mean error is present at high frequencies compared to low frequencies. This is due to a couple of reasons; (i) at each frequency bin, the modal approach truncates the soundfield order at $N = \lceil kr \rceil$ and therefore, the lower frequencies consider less number of soundfield modes compared to higher frequencies; (ii) the frequency resolution at low frequency sub-bands is low, hence the number of frequency bins contributing to the spatial correlation model is limited, which increases the fluctuation of results compared to that of high frequency sub-bands with higher resolution; (iii) the ground truth values are bound to inherent estimation errors from RIR measurement imperfections, numerical noise, limitations of the theories used etc.

E. Influence of room statistics

A more detailed study on the influence of room statistics on the proposed DRR algorithm is given below. We analyze the effects of source-receiver distance, room size, reverberation time, noise levels and noise types.

1) Influence of the source-receiver distance and room size: Figure 3 shows the sub-band DRR estimation and DRR ground truth for the 5 rooms and 2 microphone configurations

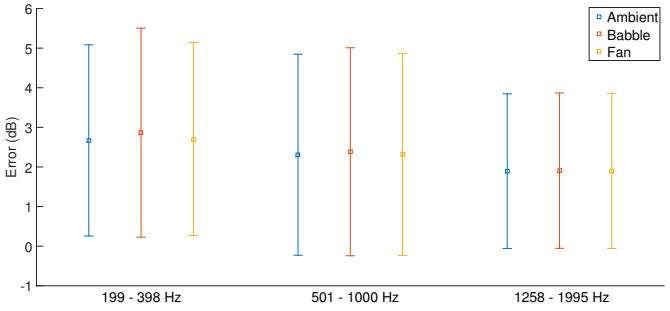


Fig. 4. Mean and standard deviation of DRR estimation error for different noise types at SNR$= -1$ dB.

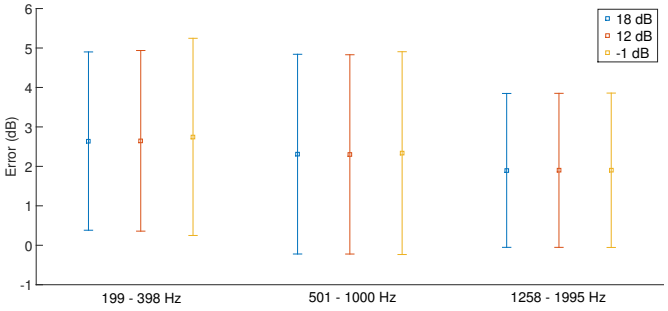


Fig. 5. Mean and standard deviation of DRR estimation error for varying SNR.

as provided by the ACE challenge database. As mentioned previously, the terms 'long' and 'short' express the source-microphone distance. From the 10 cases, the DRR estimation performance is consistently good (within 3 dB error) in 6 scenarios. The remaining 4 cases shows estimation errors up to 5 dB for which the possible reasons are discussed below. From the 'long' configuration, Lecture Room 2 and Meeting Room 2 have the weakest DRR estimation performance. There are two main reasons for this performance degradation. First, it is observed from Table I that these two rooms are the largest. As mentioned in Section V-C, the reverberant signal energy in large rooms are quite low due to the rooms becoming close to anechoic with increasing size. Therefore, the numerical process involved with estimating the reverberant path power will introduce additional noise. Secondly, there exists a formulation error that specifically affects the DRR estimation accuracy when the source-microphone distance is large. This error is stemmed from the assumption made in Section III-A, where the cross-correlation between direct and reverberant path components of the soundfield coefficients $\alpha_{nm}(k)$ was assumed to be zero (14). The maximum error resulting from this at each soundfield mode is $2|H_D(k)Y^*_{nm}(\theta_0,\phi_0)||\beta_{nm}(k)|$. As the receiver is moved further away from the source, the direct component degrades while the reverberant component stays the same. Consequently, the ratio of the error to the power of the direct component $|H_D(k)Y^*_{nm}(\theta_0,\phi_0)|^2$, which is $\frac{2|\beta_{nm}(k)|}{|H_D(k)Y^*_{nm}(\theta_0,\phi_0)|}$, becomes inversely proportional to the true DRR. Hence the formulation error prominently affects the estimation of DRR at long distances where true DRR is low. The results in Lecture Room 2 is a prime example for this.

From the 'short' configuration, Lecture Room 1 and Office 2 display degraded performance. The main reason for this degradation is the plane wave assumption of all incoming sound waves incident at the Eigenmike. This assumption is only valid for sound sources located in the far-field defined by $||\hat{\boldsymbol{y}}_0 - \boldsymbol{x}_q|| > r^2/\lambda$ where λ is the wavelength. Therefore, when the source-microphone distance is less than this limit, the above assumption causes discrepancies in the estimate direct path correlation which then propagates to the estimated DRR.

2) Influence of Reverberation time T_{60}: Rooms with larger T_{60} values experience increased reverberation. From Table I, Lecture Room 2 has the longest T_{60}, which two or more times larger than those of the remaining 4 rooms. When the reverberant power is large, the most vulnerable scenario for error is direct path estimation with a long source-receiver distance. As discussed in section V-E1, the direct path estimation is already difficult due to its degradation with the increased travel distance, and now a large reverberant signal is present to further obscure its detection. For shorter source-receiver distance the performance is expected to improve as observed in Lecture Room 2-short. The DRR estimation of less reverberant rooms can be expected to be least affected by the reverberation time. However, if the reverberant field is incorrectly assumed to be diffuse, when in reality its sparse and weak, the DRR estimation performance is bound to degrade. This result can be observed in Fig. 1, where the diffuse-method show increased estimation errors in most rooms with low-medium T_{60} values.

3) Influence of Noise: In this section, we analyze the influence of different noise types and noise levels on the performance of the proposed DRR estimation algorithm. The ACE challenge database provides microphone recordings for 3 noise types, namely, 'Ambient', 'Fan' and 'Babble'. For each noise type, the SNR is varied over 3 levels at 18 dB, 12 dB and -1 dB. For a more comprehensible presentation, we re-grouped the sub-band results into 'low' ($199-398$ Hz), 'medium' ($501-1000$ Hz), and 'high' ($1258-1995$ Hz) frequency bands as shown in figures 4 and 5.

Figure 4 shows the mean and standard deviation of the sub-band DRR estimation error for the 3 types of noise present. On their own, these noise types have distinct spectral characteristics as described bellow. 'Ambient' noise is prominent in the low frequency range with the energy decreasing gradually with increasing frequency. 'Fan' noise has a very similar spectrum to that of 'Ambient'. In contrast, 'Babble' noise has a wider spectrum due to the fact that it's a collection of superimposed speech signals. Therefore, the energy present in its mid and high frequency bands are higher compared to that of 'Ambient' and 'Fan'. Figure 4 seem to clearly reflect the effects of above spectral characteristics. In the low frequency range, all 3 noise types result in the highest estimation error due to the noise spectral energy dominance in that particular frequency band. There's a gradual decrease of error with increasing frequency following the energy distribution of all 3 noise types. The performance under 'Ambient' and 'Fan' noise are almost identical due to the similarity of their noise spectrum. However, the performance under 'Babble' noise is slightly

higher, specially in the low and mid frequency ranges, due to increased energy in its frequency spectrum.

Figure 5 shows the mean and standard deviation of the subband DRR estimation error for 3 different noise levels with the SNR varying between 18 dB, 12 dB and -1 dB. It is observed that the estimation results are quite similar with 18 dB and 12 dB SNR. At -1 dB SNR, the mean and standard deviation of the estimated error are slightly higher, specially at low frequencies. Apart from the increased interference, this could be partially stemming from the DOA estimation errors.

VI. Conclusion

In this paper, we presented and evaluated a novel approach to estimate the Direct-to-Reverberant Ratio in a noisy field. This method utilizes a spatial soundfield recording at the receiver location of interest and spherical harmonics domain spatial correlation model to reliably estimate the direct path energy, reverberant path energy and the desired DRR. In contrast to existing research, we modeled the reverberant path with a more realistic non-diffuse field. We showed that this approach achieves better DRR estimates compared to that assuming a diffuse model for the reverberant field. The simulation results were obtained using a measurement database made available by the ACE challenge. Further improvement should be sought to make the proposed method more robust to the variation of the acoustical environment, specially within the low frequency range.

Appendix A
Derivation of Equation (13)

By substituting (12) and (11) in (5), we derive

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\alpha_{nm}(k)j_n(kr)Y_{nm}(\theta_q,\phi_q) = S(k)\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n\Big(H_D(k)Y^*_{nm}(\theta_0,\phi_0)+\sum_{n',m'}\beta_{nm}(k) \int_{\hat{\boldsymbol{y}}} Y^*_{nm}(\theta,\phi)Y_{n'm'}(\theta,\phi)d\hat{\boldsymbol{y}}\Big)j_n(kr)Y_{nm}(\theta_q,\phi_q) \quad (29)$$

Based on the orthonormal property of spherical harmonic functions (8), (29) can be simplified into

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\alpha_{nm}(k)j_n(kr)Y_{nm}(\theta_q,\phi_q) = S(k)\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n\Big(H_D(k)Y^*_{nm}(\theta_0,\phi_0)+\beta_{nm}(k)\Big)j_n(kr)Y_{nm}(\theta_q,\phi_q). \quad (30)$$

From (30), the modal coefficients of the spatial soundfield recorded by the microphone array, $\alpha_{nm}(k)$ can be identified as

$$\alpha_{nm}(k) = S(k)i^n\Big(H_D(k)Y^*_{nm}(\theta_0,\phi_0)+\beta_{nm}(k)\Big).$$

This completes the derivation of (13).

Appendix B
Derivation of Equation (18)

In this section, we perform a four step simplification utilizing the inherent properties of spherical harmonics to derive a closed form expression for $E\{\beta_{nm}(k)\beta^*_{n'm'}(k)\}$.

First we use the orthonormal property of spherical harmonics (8) in (11) to derive

$$E\Big\{\beta_{nm}(k)\beta^*_{n'm'}(k)\Big\} = E\Big\{\int_{\hat{\boldsymbol{y}}} H_R(k,\hat{\boldsymbol{y}})Y^*_{nm}(\hat{\boldsymbol{y}})d\hat{\boldsymbol{y}} \int_{\hat{\boldsymbol{y}}'} H^*_R(k,\hat{\boldsymbol{y}}')Y_{n'm'}(\hat{\boldsymbol{y}}')d\hat{\boldsymbol{y}}'\Big\} = \int_{\hat{\boldsymbol{y}}}\int_{\hat{\boldsymbol{y}}'} E\Big\{H_R(k,\hat{\boldsymbol{y}})H^*_R(k,\hat{\boldsymbol{y}}')\Big\}Y^*_{nm}(\hat{\boldsymbol{y}})Y_{n'm'}(\hat{\boldsymbol{y}}')d\hat{\boldsymbol{y}}d\hat{\boldsymbol{y}}' \quad (31)$$

Second, based on the assumption that the reflection gains from different incoming directions are uncorrelated (16), we simplify (31) into

$$E\Big\{\beta_{nm}(k)\beta^*_{n'm'}(k)\Big\} = \int_{\hat{\boldsymbol{y}}} E\Big\{\,\|H_R(k,\hat{\boldsymbol{y}})\|^2\,\Big\}Y^*_{nm}(\hat{\boldsymbol{y}})Y_{n'm'}(\hat{\boldsymbol{y}})d\hat{\boldsymbol{y}}. \quad (32)$$

In order to simplify (32), we next substitute the spherical harmonic decomposition of $E\Big\{\,\|H_R(k,\hat{\boldsymbol{y}})\|^2\,\Big\}$ defined in (17), in (32) and derive

$$E\Big\{\beta_{nm}(k)\beta^*_{n'm'}(k)\Big\} = \sum_{v=0}^{\infty}\sum_{u=-v}^{v}\gamma_{vu}(k)\int_{\hat{\boldsymbol{y}}} Y_{vu}(\hat{\boldsymbol{y}})Y^*_{nm}(\hat{\boldsymbol{y}})Y_{n'm'}(\hat{\boldsymbol{y}})d\hat{\boldsymbol{y}}. \quad (33)$$

Finally, by utilizing one of the integral properties of spherical harmonics given by [33]

$$\sum_{v=0}^{\infty}\sum_{u=-v}^{v}\int_{\hat{\boldsymbol{y}}} Y_{vu}(\hat{\boldsymbol{y}})Y^*_{nm}(\hat{\boldsymbol{y}})Y_{n'm'}(\hat{\boldsymbol{y}})d\hat{\boldsymbol{y}} = \sqrt{(}\frac{(2V+1)(2n+1)(2n'+1)}{4\pi})W_1W_2 \quad (34)$$

where W_1 (19) and W_2 (20) denote Wigner coefficients, in (33), we arrive at a closed form expression for $E\{\beta_{nm}(k)\beta^*_{n'm'}(k)\}$ as

$$E\Big\{\beta_{nm}(k)\beta^*_{n'm'}(k)\Big\} = \sum_{v=0}^{\infty}\sum_{u=-v}^{v}\gamma_{vu}(k)\Big(\frac{(2v+1)(2n+1)(2n'+1)}{4\pi}\Big)^{1/2}W_1W_2.$$

This completes the derivation of equation (18).

References

[1] P. A. Naylor and N. D. Gaubitch, *Speech dereverberation*. Springer Science & Business Media, 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, 1982.

[3] 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.

[4] 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.

[5] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[6] D. P. Jarrett, E. A. Habets, M. R. Thomas, N. D. Gaubitch, and P. A. Naylor, "Dereverberation performance of rigid and open spherical microphone arrays: Theory & simulation," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*. IEEE, 2011, pp. 145–150.

[7] S. Vesa, "Sound source distance learning based on binaural signals," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 271–274.

[8] ——, "Binaural sound source distance learning in rooms," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1498–1507, 2009.

[9] T. Jo and M. Koyasu, "Measurement of reverberation time based on the direct-reverberant sound energy ratio in steady state," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 1975, no. 2. Institute of Noise Control Engineering, 1975, pp. 579–582.

[10] M.-V. Laitinen and V. Pulkki, "Utilizing instantaneous direct-to-reverberant ratio in parametric spatial audio coding," in *Audio Engineering Society Convention 133*. Audio Engineering Society, 2012.

[11] E. Larsen, C. D. Schmitz, C. R. Lansing, W. D. O'Brien Jr, B. C. Wheeler, and A. S. Feng, "Acoustic scene analysis using estimated impulse responses," in *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1. IEEE, 2003, pp. 725–729.

[12] T. H. Falk and W.-Y. Chan, "Temporal dynamics for blind measurement of room acoustical parameters," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 978–989, 2010.

[13] 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 *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 149–152.

[14] O. Thiergart, T. Ascherl, and E. A. Habets, "Power-based signal-to-diffuse ratio estimation using noisy directional microphones," in *IEEE International Conference On Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 7440–7444.

[15] Y. Hioka and K. Niwa, "Psd estimation in beamspace for estimating direct-to-reverberant ratio from a reverberant speech signal," *arXiv preprint arXiv:1510.08963*, 2015.

[16] 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.

[17] M. Kuster, "Estimating the direct-to-reverberant energy ratio from the coherence between coincident pressure and particle velocity," *The Journal of the Acoustical Society of America*, vol. 130, no. 6, pp. 3781–3787, 2011.

[18] H. Chen, P. N. Samarasinghe, T. D. Abhayapala, and W. Zhang, "Estimation of the direct-to-reverberant energy ratio using a spherical microphone array," *arXiv preprint arXiv:1510.08950*, 2015.

[19] H. Chen, T. D. Abhayapala, P. N. Samarasinghe, and W. Zhang, "Direct-to-reverberant energy ratio estimation using a first order microphone," *Transactions on Audio, Speech and Language Processing*, in press 2016.

[20] O. Thiergart, G. Del Galdo, and E. A. 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, no. 4, pp. 2337–2346, 2012.

[21] D. P. Jarrett, O. Thiergart, E. A. Habets, and P. A. Naylor, "Coherence-based diffuseness estimation in the spherical harmonic domain," in *IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*. IEEE, 2012, pp. 1–5.

[22] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *The Journal of the Acoustical Society of America*, vol. 138, no. 5, pp. 3081–3092, 2015.

[23] A. Gupta and T. D. Abhayapala, "Double sided cone array for spherical harmonic analysis of wavefields," in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 77–80.

[24] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "The ace challengeâĂŤcorpus description and performance evaluation," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2015 IEEE Workshop on*. IEEE, 2015, pp. 1–5.

[25] ——, "Estimation of room acoustic parameters: The ace challenge," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1681–1693, Oct 2016.

[26] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography*. London, UK: Academic Press, 1999, pp. 115–125.

[27] T. Abhayapala and D. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. II, 2002, pp. 1949–1952.

[28] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, 2002, pp. 1781–1784.

[29] S. Mosayyebpour, H. Sheikhzadeh, T. A. Gulliver, and M. Esmaeili, "Single-microphone lp residual skewness-based inverse filtering of the room impulse response," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1617–1632, 2012.

[30] D. Ward and T. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

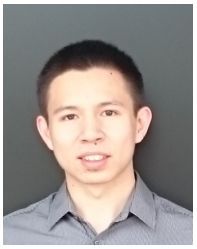
[31] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*. IEEE, 2009, pp. 221–224.

[32] P. T. D. Abhayapala *et al.*, "Modal analysis and synthesis of broadband nearfield beamforming arrays," 1999.

[33] J. Mathews and R. L. Walker, *Mathematical methods of physics*. WA Benjamin New York, 1970, vol. 501.

Prasanga Samarasinghe received her B.E. degree (with first-class honors) in electronic and electrical engineering from the University of Peradeniya, Sri Lanka, in 2009. She completed her Ph.D. degree at the Australian National University (ANU), Canberra in 2014. She is currently working as a research fellow at the Research School of Engineering at ANU. Her research interests include spatial audio, active noise control, and multichannel signal processing.

Thushara Abhayapala received his B.E. degree (with honors) in engineering in 1994 and his Ph.D. degree in telecommunications engineering in 1999, both from the Australian National University (ANU), Canberra. He is currently the deputy dean of the College of Engineering and Computer Science, ANU. He was the the director of the Research School of Engineering at ANU from January 2010 to October 2014 and the leader of the Wireless Signal Processing Program at the National ICT Australia from November 2005 to June 2007. His research interests are in the areas of spatial audio and acoustic signal processing, and multichannel signal processing. He has supervised more than 30 Ph.D. students and coauthored more than 200 peer-reviewed papers. He is an associate editor of IEEE/ACM Transactions on Audio, Speech, and Language Processing. He is a member of the Audio and Acoustic Signal Processing Technical Committee (2011 – 2016) of the IEEE Signal Processing Society. He is a fellow of Engineers Australia.



Hanchi Chen received the B.E. degree (with first class honours) in Electronics and Telecommunication Engineering in 2012, from the Australian National University (ANU), Canberra. He is currently a pursuing a Ph.D. degree at ANU in the field of spatial audio signal processing. His research interests include spatial audio and multi-channel processing techniques, especially active noise control.