Estimating Direct-to-reverberant Energy Ratio Using D/R Spatial Correlation Matrix Model

Yusuke Hioka, *Member, IEEE,* Kenta Niwa, *Member, IEEE,* Sumitaka Sakauchi, *Nonmember,* Ken'ichi Furuya, *Senior Member, IEEE,* and Yoichi Haneda, *Senior Member, IEEE*

***Abstract*—We present a method for estimating the direct-to-reverberant energy ratio (DRR) that uses a direct and reverberant sound spatial correlation matrix model (Hereafter referred to as the spatial correlation model). This model expresses the spatial correlation matrix of an array input signal as two spatial correlation matrices, one for direct sound and one for reverberation. The direct sound propagates from the direction of the sound source but the reverberation arrives from every direction uniformly. The DRR is calculated from the power spectra of the direct sound and reverberation that are estimated from the spatial correlation matrix of the measured signal using the spatial correlation model. The results of experiment and simulation confirm that the proposed method gives mostly correct DRR estimates unless the sound source is far from the microphone array, in which circumstance the direct sound picked up by the microphone array is very small. The method was also evaluated using various scales in simulated and actual acoustical environments, and its limitations revealed. We estimated the sound source distance using a small microphone array, which is an example of application of the proposed DRR estimation method.**

***Index Terms*—Direct-to-reverberation energy ratio, D/R spatial correlation matrix model, microphone array, sound source distance.**

I. Introduction

In analyzing characteristics of a reverberant environment, estimating the direct-to-reverberant energy ratio (DRR) is quite helpful because various acoustic parameters, such as reverberation time, diffuseness, etc., can be calculated from it [2]. Several methods are available for estimating DRR. The most primitive way is to calculate DRR directly from the impulse response. However, this requires measurement of the room impulse response. Larsen *et al.* proposed a method for estimating DRR from simply the short beginning part of the impulse response [3], but it still necessitates prior processing to identify the initial part of the impulse response. Falk *et al.* [4] proposed a method that focuses on the long-term temporal dynamics of speech signals. This method performs very well, but it requires an *a priori* calculation of the relation between DRR and the overall reverberation-to-speech modulation energy ratio (ORSMR), which was proposed by the authors in the paper. Obtaining this prior information might be laborious because it changes depending on the acoustic environment.

DRR also has an important aspect relating to human hearing. In the last few decades, many researchers have studied the human auditory perception of the distance to a sound source [5]–[9]. A recent summary paper on human hearing has concluded that DRR may provide absolute distance information especially in reverberant environments, whereas the direct sound can only provide relative distance information [10]. Conventional instruments for measuring the sound source distance include ultrasonic sensors and microphone arrays [11]. Ultrasonic sensors are often used for measuring the distance to a target; however, the reflected ultrasonic waves used in the measurement can be scattered or reflected away from the receiver's (or ultrasonic sensor's) direction if the target has a round or uneven surface [12]. On the other hand, although the use of a microphone array is not influenced by the shape of the surface of the target, conventional methods [13], [14] fail to correctly estimate the distance when the environment has strong reverberation. This is because the methods assume a model of input signal that do not take into account the effect of reverberation explicitly.

As in the case of human auditory perception, the DRR may be a cue to estimate the absolute sound source distance. In fact, a few attempts exploiting the DRR to estimate sound source distance have been reported. Lu *et al.* recently proposed a procedure to derive DRR [15] in order to estimate the sound source distance. They first estimated the energy of the reverberant component by eliminating the direct component through an equalization-cancellation (EC) technique. To eliminate the direct sound, the EC technique exploits the fact that a large difference between the direct sound and reverberation exists in the inter-channel (or spatial) correlation of the binary input signal. Vesa proposed another method [16], [17] that utilises the magnitude-squared coherence (MSC) instead of DRR. Although these methods do not require one to perform an impulse response measurement, they are only appropriate for binaural input signals, which are strongly affected by the human head related impulse response (HRIR). That is, they have limited applicability to microphone array signals, which consist from more than two microphones and have no effect of HRIR. To the best of our knowledge, no DRR estimation method using a microphone array has been reported.

The novelty of this paper is the proposal of DRR estimation method that is the first attempts to use signals measured by

 The authors are with the NTT Cyber Space Laboratories, NTT Corporation, Musashino, Tokyo, 180-8585 Japan.(e-mail: hioka.yusuke@lab.ntt.co.jp, niwa.kenta@lab.ntt.co.jp, sakauchi.sumitaka@lab.ntt.co.jp, furuya.kenichi@lab.ntt.co.jp, haneda.yoichi@lab.ntt.co.jp)

Part of this work has appeared in Proceedings of ICASSP2010 [1].

Manuscript received October 21, 2010; revised January 11, 2011; accepted March 8, 2011.

a general microphone array. It uses the direct and reverberant sound (D/R) spatial correlation matrix model (called the spatial correlation model hereafter). The spatial correlation model assumes that the direct sound propagates from the direction of the sound source but that the reverberation arrives from all directions. Then, we calculate the DRR from the power spectra of both components. These components are estimated from the correlation matrix of the measured signals using the spatial correlation model. We can utilise the estimated DRR to calculate the sound source distance because DRR keeps its one-to-one relation for a range of distances where the effect of model error in the spatial correlation model is small.

This paper is organised as follows. In Sec. II, we introduce the spatial correlation model and present a method for estimating DRR based on the model. In Sec. III, we evaluate the performance of the proposed method by measuring the accuracy of the estimated DRR. In Sec. IV, we investigate the influence of the physical environment on the performance of the proposed method. Finally, in Sec. V, we show the results of estimating the sound source distance by using the estimated DRR. Comments on the outcome of this study and future work conclude this paper.

II. DRR ESTIMATION BASED ON SPATIAL CORRELATION MODEL

A. Direct and reverberant sound spatial correlation matrix model

First, we decompose the impulse response between the sound source and a microphone $H(\omega)$ into two components: the direct component $H_{\mathrm{D}}(\omega)$ and reverberant component $H_{\mathrm{R}}(\omega)$, as illustrated in Fig. 1. For the sake of simplicity, we describe the impulse responses in the form of frequency transfer functions where ω denotes the frequency. Note that the early reflections of the impulse response are also included in $H_{\mathrm{R}}(\omega)$. If we have an M-sensor microphone array, the input signal of the m-th microphone expressed in the time-frequency domain is given by

$$X^{(m)}(\omega,t) = \left(H_{\mathrm{D}}^{(m)}(\omega) + H_{\mathrm{R}}^{(m)}(\omega)\right) S(\omega,t), \tag{1}$$

where t denotes the temporal frame index and $S(\omega,t)$ is the short-time Fourier transform of the sound source. According to this expression, the cross correlation between the p-th and q-th microphones can be derived as

$$\begin{aligned} &E[X^{(p)}(\omega,t)X^{(q)*}(\omega,t)] \\ &= E\left[|S(\omega,t)|^2 \left\{ H_{\mathrm{D}}^{(p)}(\omega)H_{\mathrm{D}}^{(q)*}(\omega) + H_{\mathrm{R}}^{(p)}(\omega)H_{\mathrm{R}}^{(q)*}(\omega) \right.\right. \\ &\left.\left. + H_{\mathrm{D}}^{(p)}(\omega)H_{\mathrm{R}}^{(q)*}(\omega) + H_{\mathrm{R}}^{(p)}(\omega)H_{\mathrm{D}}^{(q)*}(\omega) \right\}\right], \end{aligned} \tag{2}$$

where $E[\cdot]$ and $*$ denote the expectation and complex conjugate, respectively. Now we put the following assumptions on the input signal: the aperture size of the microphone array is sufficiently small for recognizing the direct component as a plane wave; the reverberant component is diffuse; and the cross correlation between the direct and reverberant components (the third and fourth terms on the right side of Eq. (2)) is sufficiently small. Under these assumptions, the spatial correlation matrix [18] of the microphone array $\mathbf{R}(\omega)$ can be approximated by two matrices, given by

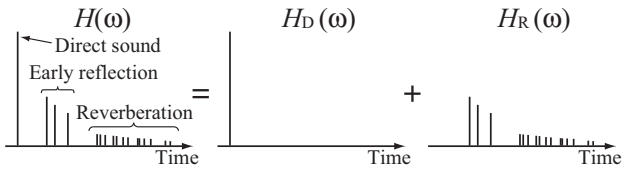


Fig. 1. Decomposition of impulse response.

$$\begin{aligned} \mathbf{R}(\omega) &= E[\mathbf{X}(\omega,t)\mathbf{X}^H(\omega,t)] \\ &\simeq P_{\mathrm{D}}(\omega) \begin{bmatrix} 1 & d_{12}^{\mathrm{P}} & \cdots & d_{1M}^{\mathrm{P}} \\ d_{21}^{\mathrm{P}} & 1 & \cdots & d_{2M}^{\mathrm{P}} \\ \vdots & \vdots & \ddots & \vdots \\ d_{M1}^{\mathrm{P}} & d_{M2}^{\mathrm{P}} & \cdots & 1 \end{bmatrix} \\ &\quad + P_{\mathrm{R}}(\omega) \begin{bmatrix} 1 & r_{12} & \cdots & r_{1M} \\ r_{21} & 1 & \cdots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{M1} & r_{M2} & \cdots & 1 \end{bmatrix}, \end{aligned} \tag{3}$$

where

$$\mathbf{X}(\omega,t) = \left[X^{(1)}(\omega,t)\ X^{(2)}(\omega,t)\ \cdots\ X^{(M)}(\omega,t)\right]^T \tag{4}$$

$$d_{pq}^{\mathrm{P}} = \exp\left(j\mathbf{k}^T \cdot (\mathbf{r}_p - \mathbf{r}_q)\right), \tag{5}$$

$$r_{pq} = \mathrm{sinc}\left(\omega \frac{||\mathbf{r}_p - \mathbf{r}_q||}{c}\right), \tag{6}$$

$\mathbf{r}_m$, c, and ^{H} are the coordinates of the m-th microphone, the speed of sound, and the Hermitian transform, respectively, and $||\cdot||$ is Euclidean distance. Furthermore, $\mathbf{k}$ in Eq. (5) denotes the wave number vector [19] for the sound source in the direction of (θ,ϕ) defined in Fig. 2, i.e. $\mathbf{k} = \frac{\omega}{c}[\sin\theta\cos\phi, \cos\theta\cos\phi, \sin\phi]^T$.

The first term on the right side of Eq. (3) expresses the spatial correlation of the direct component. As there exists a time difference of arrival between microphones in the cross correlation of direct sound, the spatial correlation is expressed by a simple phase difference. In the modelling of the second term, we utilised the feature that the spatial correlation of diffuse sound can be expressed by a sinc function [20]. In Eq. (3), $P_{\mathrm{D}}(\omega)$ and $P_{\mathrm{R}}(\omega)$ are defined by

$$\begin{aligned} P_{\mathrm{D}}(\omega) &= E[|S(\omega,t)|^2|H_{\mathrm{D}}(\omega)|^2], \\ P_{\mathrm{R}}(\omega) &= E[|S(\omega,t)|^2|H_{\mathrm{R}}(\omega)|^2]. \end{aligned}$$

Because of the plane wave assumption, the magnitudes of the transfer function for each microphone can be considered to be identical, i.e., $|H_{\mathrm{D}}^{(p)}(\omega)||H_{\mathrm{D}}^{(q)}(\omega)| = |H_{\mathrm{D}}(\omega)|^2$ and $|H_{\mathrm{R}}^{(p)}(\omega)||H_{\mathrm{R}}^{(q)}(\omega)| = |H_{\mathrm{R}}(\omega)|^2$. In the rest of this paper, the modelling of the spatial correlation matrix given in Eq. (3) will be called the "spatial correlation model".

B. DRR estimation using power spectra of direct and reverberant components

Since the microphone array configuration is initially known and the direction of the sound source can be estimated by using one of the conventional methods [11], we can calculate d_{pq}^{P}

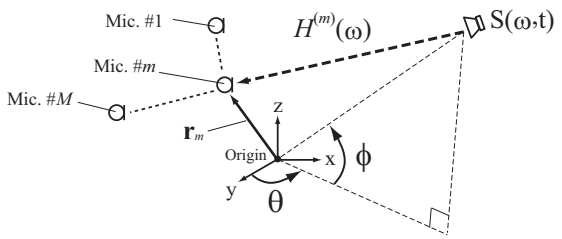


Fig. 2. Modelling of measured signal.

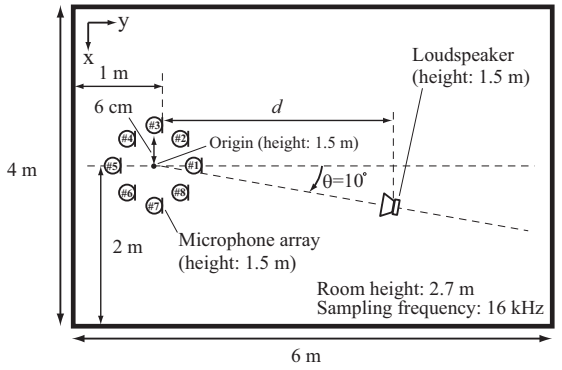


Fig. 3. Positions of microphone array and speaker in simulated reverberant room.

and r_{pq} in Eq. (3). We can then estimate the unknown power spectra of the direct and reverberant components, $P_{\mathrm{D}}(\omega)$ and $P_{\mathrm{R}}(\omega)$, by solving the simultaneous equations in Eq. (7), which were derived by reformulating Eq. (3).

$$\underbrace{\begin{bmatrix} 1 & 1 \\ d_{12}^{\mathrm{P}} & r_{12} \\ \vdots & \vdots \\ d_{1M}^{\mathrm{P}} & r_{1M} \\ d_{21}^{\mathrm{P}} & r_{21} \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}}_{\mathbf{F}(\omega)} \underbrace{\begin{bmatrix} P_{\mathrm{D}}(\omega) \\ P_{\mathrm{R}}(\omega) \end{bmatrix}}_{\mathbf{P}(\omega)} = \underbrace{\begin{bmatrix} R_{11}(\omega) \\ R_{12}(\omega) \\ \vdots \\ R_{1M}(\omega) \\ R_{21}(\omega) \\ R_{22}(\omega) \\ \vdots \\ R_{MM}(\omega) \end{bmatrix}}_{\widetilde{\mathbf{R}}(\omega)} \tag{7}$$

Here, $R_{pq}(\omega)$ in $\widetilde{\mathbf{R}}(\omega)$ denotes the p-th row and q-th column components of $\mathbf{R}(\omega)$, which can be calculated from the observed measured signals. The estimated power spectra of the direct and reverberant components are obtained by solving Eq. (7) using the least-squares method:

$$\widehat{\mathbf{P}}(\omega) = \mathbf{F}^{+}(\omega)\widetilde{\mathbf{R}}(\omega), \tag{8}$$

where $^{+}$ and $\widehat{\ }$ are the Moore-Penrose pseudo inverse and estimated value, respectively.

Finally, the estimated DRR is calculated from the estimated power spectra $\widehat{P_{\mathrm{D}}}(\omega)$ and $\widehat{P_{\mathrm{R}}}(\omega)$:

$$\mathrm{DRR}_{\mathrm{estimate}} = 10\log_{10}\left(\frac{\sum_{\omega}\widehat{P_{\mathrm{D}}}(\omega)}{\sum_{\omega}\widehat{P_{\mathrm{R}}}(\omega)}\right), \tag{9}$$

where

$$\widehat{\mathbf{P}}(\omega) = \begin{bmatrix} \widehat{P_{\mathrm{D}}}(\omega) \\ \widehat{P_{\mathrm{R}}}(\omega) \end{bmatrix}. \tag{10}$$

TABLE I
DEFAULT CONDITIONS AND PARAMETERS IN COMPUTER SIMULATION.

Fs: Sampling frequency [Hz]	16,000
M: Number of microphones	8
Microphone arrangement	circular
Diameter of array [cm]	12
α: absorption coefficient	0.15
Corresponding reverberation time [ms]	0.55
Frame length [samples]	512
Frame shift [samples]	256
Window	Hamming
SNR [dB]	∞

Note that the estimation of the power spectra (Eq. (7)) uses a lot of redundant equations. One possible solution to temper the redundancy is to reduce the number of microphones, but this affects the estimation accuracy, as discussed later in Sec. III-C. Another possible way is to use only the upper or lower triangle of the spatial correlation matrix because the triangle consists of the conjugate values of the other. However, it is not appropriate because using only the triangle results in non-vanishing imaginary components which is not a realistic estimate for a power spectrum.

III. Evaluation of Basic Performance

A. Simulation settings and evaluation criteria

To evaluate the proposed DRR estimation, we performed simulations of reverberant environments. Table I and Fig. 3 show the default parameters and physical positions of the microphone array and loudspeaker used in the simulation. The sound source was 3-s long Gaussian white noise, unless otherwise stated, and the input signals of the microphone array were prepared by convolving the simulated impulse response generated by the image method [21]. The input signals in the time-frequency domain $X^{(m)}(\omega, t)$ were calculated by performing windowed short-time Fourier transform. The DRR was estimated from a 3-s long multiple channel input signal. Simulations were performed for 100 different Gaussian white noise signals. ∞ for the SNR in Tab. I means the input signal was noise-free; i.e., no noise signal was added to the input signal, and all surfaces of the room had the same absorption coefficient α. The corresponding reverberation time was calculated from the reverberation curve [22] of the given impulse response.

The DRR itself is a ratio between two values; so it is reasonable to evaluate the proportion of the estimate to the correct value. Due to DRR is defined in units of decibels in Eq. (9), the proportion between the estimated DRR and correct DRR corresponds to the difference between $\mathrm{DRR}_{\mathrm{estimate}}$ and $\mathrm{DRR}_{\mathrm{actual}}$. Thus, as an evaluation criterion, we calculated the DRR difference defined by

$$\epsilon_{\mathrm{DRR}} = |\mathrm{DRR}_{\mathrm{estimate}} - \mathrm{DRR}_{\mathrm{actual}}|, \tag{11}$$

where the actual DRR is directly calculated from the impulse response defined by

$$\mathrm{DRR}_{\mathrm{actual}} = 10\log_{10}\left(\frac{\sum_{\omega}|H_{\mathrm{D}}(\omega)|^2}{\sum_{\omega}|H_{\mathrm{R}}(\omega)|^2}\right). \tag{12}$$

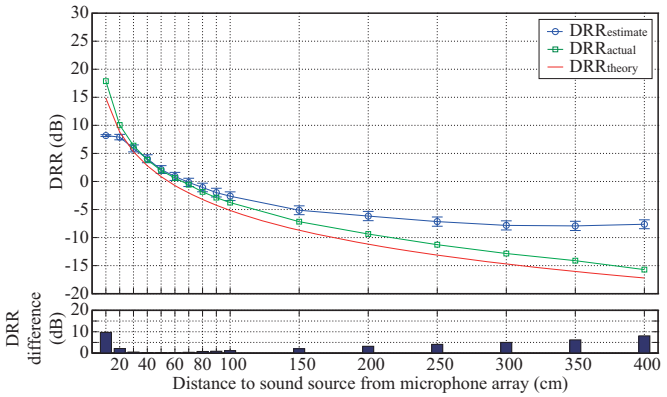


Fig. 4. Comparison of estimated and actual DRR in simulated room. Lines with circles and squares in the upper graph show $\mathrm{DRR}_{\mathrm{estimate}}$ and $\mathrm{DRR}_{\mathrm{actual}}$, respectively. The remaining line shows the theoretical DRR in a diffuse sound field, in order to show the difference between it and the actual reverberations. The log DRR difference between $\mathrm{DRR}_{\mathrm{estimate}}$ and $\mathrm{DRR}_{\mathrm{actual}}$ is shown in the lower graph.

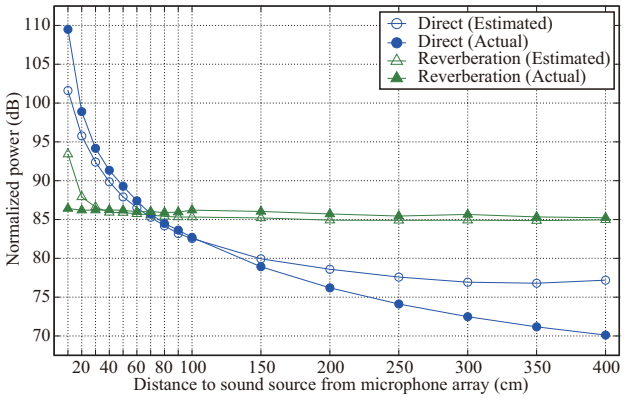


Fig. 5. Comparison of estimated (circles) and actual (triangles) power spectra of direct sound and reverberation.

When the $\mathrm{DRR}_{\mathrm{estimate}}$ is identical to $\mathrm{DRR}_{\mathrm{actual}}$, the proposed method is considered to be completely successful in estimating DRR, and the DRR difference ϵ_{DRR} should be 0 dB. Furthermore, we also calculated the DRR according to established diffuse sound field theory ($\mathrm{DRR}_{\mathrm{theory}}$) [20], as

$$\mathrm{DRR}_{\mathrm{theory}} = 10\log_{10}\left(\frac{A\bar{\alpha}}{16\pi d^2}\right). \tag{13}$$

This value is used to show how the actual reverberation differs from a completely diffuse field. Here, A and $\bar{\alpha}$, are respectively the surface area of the walls and the average absorption coefficient.

B. DRR estimation with default parameter settings

Figure 4 shows the results of DRR estimation performed with the default parameter settings. The upper graph shows the average of the estimated DRRs, while the lower graph shows the DRR difference. The error bars in the upper graph show the standard deviations of the DRR difference over 100 trials. The error increases when the source is very near or far from the microphone array. This trend is in line with the regression analysis of the DRR differences at short (10 – 30 cm) and long (100 – 400 cm) distances shown in Tab. II and Tab. III, respectively. Note that the correlation in the left side table is described in its absolute value. The small p-values imply that the data is statistically significant. The slope is negative ($= -0.45$) at short distances, which means the DRR error decreases as the distance increases. On the other hand, the slope is positive ($= 0.02$) at long distances, which means the DRR error increases with distance.

Figure 5 shows the actual and estimated power of the direct sound and reverberation used to calculate the DRRs in Fig. 4. These results show that the estimation error at short distances is caused by the estimation error of the direct sound and reverberation power, whereas the error at long distances is mainly caused by the error in the estimated power of the direct sound.

The estimation error at long distances is model error. That is, the model ignores the correlation between the direct and reverberant components, i.e., the third and fourth terms of Eq. (2). The maximum error resulting from this omission should be $2|H_{\mathrm{D}}(\omega)||H_{\mathrm{R}}(\omega)|$. As the distance to the source from the microphone array increases, the direct component decreases but the reverberation stays the same. Therefore, the ratio of the error to the power of the direct component $|H_{\mathrm{D}}|^2$, which is $\frac{2|H_{\mathrm{R}}|}{|H_{\mathrm{D}}|}$, becomes inversely proportional to the DRR; i.e., the error increases as the DRR decreases. Thus, the model error more prominently affects the estimation of direct component at long distances where DRR decreases.

The error at shorter distances, on the other hand, could be due to the discrepancy between the modelled and the actual spatial correlation of the direct component. The plane wave assumption of the received sound is only valid for sound sources located in the far-field defined by $d > \frac{D^2}{\lambda}$ [23], where D is the array aperture size. For the octagonal microphone array used in this simulation ($D = 12$ cm), the boundary distance between the far-field and near-field was approximately 33 cm at 8 kHz. This boundary is near the point at which the estimation error started to rapidly increase (d less than 30 cm). It seems that the proposed spectrum estimation (Eq. (8)) tried to compensate the modelling error of the direct component by fitting the reverberation component because it works on a least square error basis.

As support for the above hypothesis, Fig. 6 is the DRR difference of the estimated DRR when we used a spherical wave model for the direct sound component. In other words, we used d_{pq}^{S} defined by Eq. (14) instead of d_{pq}^{P} in Eq. (5) for the matrix $\mathbf{F}(\omega)$ in Eq.(7).

$$d_{pq}^{\mathrm{S}} = \frac{||\mathbf{r}_S - \mathbf{r}_o||^2}{||\mathbf{r}_S - \mathbf{r}_p||||\mathbf{r}_S - \mathbf{r}_q||} \cdot \exp\left(j\frac{\omega}{c}\left(||\mathbf{r}_S - \mathbf{r}_p|| - ||\mathbf{r}_S - \mathbf{r}_q||\right)\right), \tag{14}$$

Here, $\mathbf{r}_S$ denotes the coordinates of the sound source's position, which is initially known, and $\mathbf{r}_o$ is the coordinates of reference microphone. Derivation of d_{pq}^{S} is explained in the Appendix. Each line shows the results for a different sound source position. Notice that the DRR difference for very short distances decreases when the sound source position is set

TABLE II
RESULTS OF REGRESSION ANALYSIS OF DRR DIFFERENCES AT SHORT DISTANCES (10 – 30 CM).

Regression statistics	
Correlation	0.93
Determination	0.86

	Coefficient	Standard error	t-statistic	p-value	Lower 95%	Upper 95%
Intercept	13.21	0.23	58.51	< 0.01	12.76	13.65
Distance to sound source	-4.51×10^{-1}	1.04×10^{-2}	-43.15	< 0.01	-4.71×10^{-1}	-4.30×10^{-1}

TABLE III
RESULTS OF REGRESSION ANALYSIS OF DRR DIFFERENCES AT LONG DISTANCES (100 – 400 CM).

Regression statistics	
Correlation	0.92
Determination	0.84

	Coefficient	Standard error	t-statistic	p-value	Lower 95%	Upper 95%
Intercept	-1.60	0.11	-13.65	< 0.01	-1.84	-1.37
Distance to sound source	2.29×10^{-2}	0.04×10^{-2}	56.20	< 0.01	2.21×10^{-2}	2.37×10^{-2}

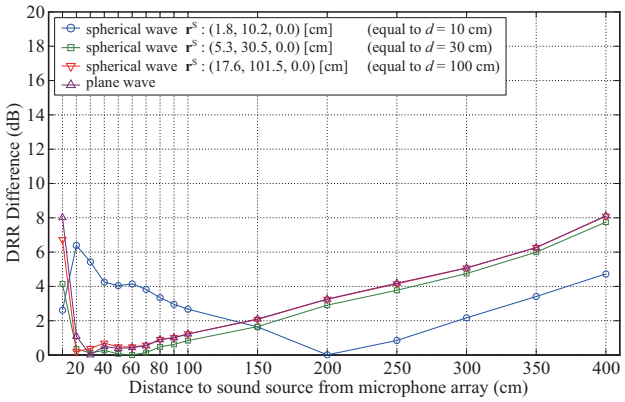


Fig. 6. Log DRR difference when spatial correlation of direct sound is modelled under a spherical wave assumption. Each line shows the measured DRR difference for various sound source positions. For comparison, the line with upward-pointing triangles shows the DRR difference of the DRR estimated with a plane wave assumption.

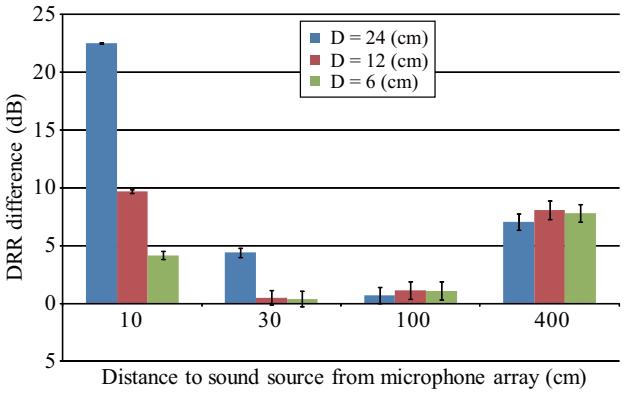


Fig. 7. Average DRR differences and their standard deviation calculated for different aperture sizes. The bar graph shows the average of DRR differences and the error bar is the standard deviation calculated over 100 trials using different input signals.

appropriately, e.g. $\mathbf{r}^{\mathrm{S}} = (1.8, 10.2)$ [cm] for $d = 10$ cm. This proves the above hypothesis: the plane wave assumption for the direct sound caused the estimation error at short distances. Despite the good performance at very short distances, the spherical wave assumption does not show a big improvement from the result for plane wave assumption in the middle and long distances. As proof, the difference of d^{S}_{pq} from d^{P}_{pq} gets very small when $\mathbf{r}^{\mathrm{S}}$ is far from the microphone array. Therefore, we can conclude that the modelling of the direct sound based on the plane wave assumption is effective for most ranges.

C. Evaluation of different sizes of microphone arrays

We evaluated the proposed method while varying the aperture size and the number of microphones. Figure 7 shows the average and standard deviation of DRR differences for different aperture sizes. In this experiment, we only changed the aperture size of the octagonal microphone array from 6 cm to 24 cm. From the results, we can see that the large aperture size ($D = 24$) suffered from estimation error even at 30 cm while the smaller aperture sizes gave a much better estimation. As we mentioned in the previous section, the aperture size of the microphone array determines the range of distances where the plane wave assumption does not hold. On the other hand, the difference was small at longer distances. These results mean that the microphone array does not need large aperture size.

Figure 8 shows the effect of varying the number of microphones. The microphones were equally arranged on a circle with a diameter of 12 cm. More microphones increases spatial resolution, and 16 microphones gave the best estimation accuracy at every distance. However, the difference was slight at the middle distances (30 cm or 100 cm) where the proposed method works well; thus, it seems that a smaller number of microphones could be an adequate compromise between the estimation accuracy and calculation load.

D. Influence of errors in the given direction of the target source

Although we assumed that the direction of sound source θ was known in advance, as there are various conventional methods for estimating the direction of arrival (DOA) [11], in practice, we need to estimate the DOA beforehand. In fact, DOA estimates sometimes show small errors especially when the room is highly reverberant. To know the influence of such errors in the DOA information, we observed the DRR differences while adding to the DOA an error value $d\theta$

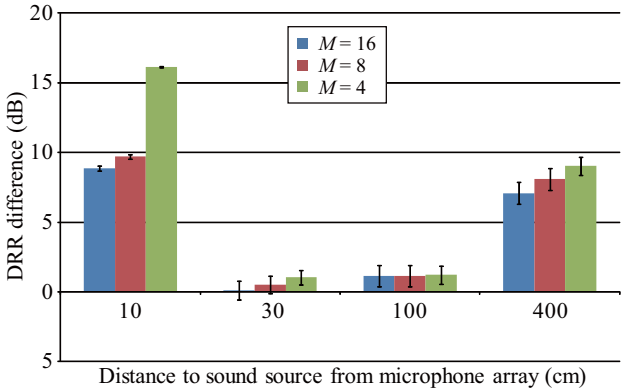


Fig. 8. Average DRR differences and their standard deviation calculated for different numbers of microphones. The bar graph shows the average of DRR differences and the error bar is the standard deviation calculated over 100 trials using different input signals.

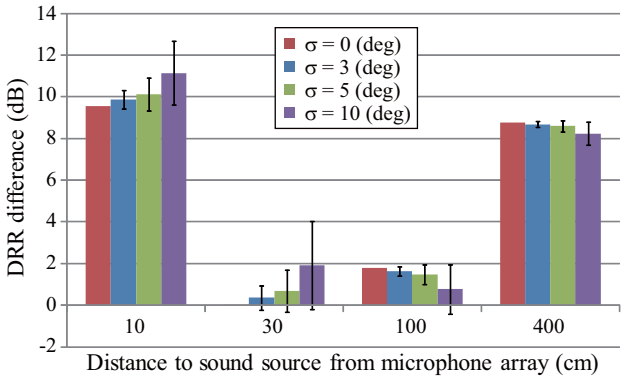


Fig. 9. Average DRR differences and their standard deviation calculated when the given direction to the source has an error.

that followed a zero mean normal distribution with standard deviation σ. $d\theta$ was varied in 100 trials. The average DRR difference and standard deviation are shown in Fig. 9. The average DRR difference varies slightly with σ, but these variations are very small, often smaller than the deviation. On the other hand, a larger σ has more of an effect to the deviation especially at short distances. Because the errors in the given DOA information only influence the direct component of the spatial correlation model, the DRR estimate is more affected at short distances where more of the direct component exists.

E. Performance of method in real acoustic environment

To confirm the effectiveness of the proposed method in an actual acoustic environment, we performed an experiment in a reverberant chamber. The room size and position of the microphone array used in this experiment were the same as those in Fig. 3, except that the loudspeaker was located in the direction of $\theta = 30°$. Every wall and the ceiling of the room were reflecting planes. The flooring was covered by carpet, and thus it might have absorbed more sound energy than the walls and ceiling. We used omni-directional condenser microphones (SONY ECM-C10) and a monitor

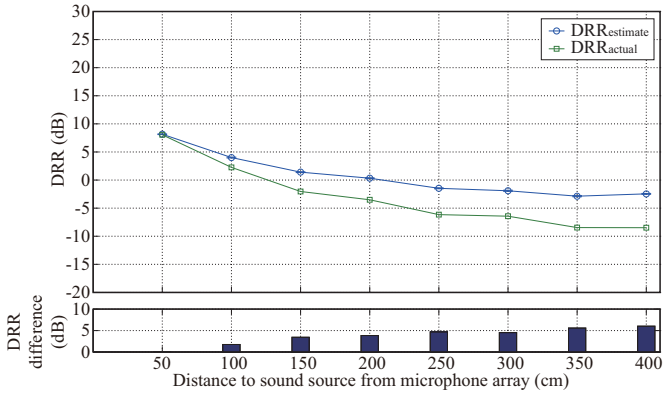


Fig. 10. DRR estimation in actual reverberant room ($T_{60} \approx 1.8$ s). DRR is overestimated at longer distances.

speaker (AURATONE C5 Sound Cube). The reverberation time of the room was approximately 1.8 s. Figure 10 shows the estimated DRR and DRR $_{\mathrm{actual}}$, and Fig. 11 shows the estimated power of each direct sound and reverberation. Note that we used signals prepared by convolving the sound source signal and impulse response preliminarily measured in the real acoustic environment in order to have the correct DRR, i.e., $\mathrm{DRR}_{\mathrm{actual}}$.

The DRR curve in Fig. 10 indicates that the proposed method overestimates DRR for longer distances. We can see that this trend is significant from the regression analysis of the DRR error in Tab. IV. The slope of the regression line is positive with significance, which means the overestimation error increases when the distance to the sound source increases. Although this is similar to the trend found in the results for the simulated environment in Fig. 4, the error in the actual acoustic environment is mainly caused by underestimation of reverberation, not by the overestimation of the direct sound (see Fig. 11). One of the reasons is that the environment was not completely diffuse because the flooring of the room was carpet, which absorbs the sound more than the other reflective walls and ceiling. In such case, the sinc function for the spatial correlation does not hold because the reverberation no longer arrives truly from all directions, thus the method will underestimate the amount of reverberation. On the other hand, the error in the direct component will be less if the reverberant component is smaller, as mentioned in Sec. III-B.

Finally, Fig. 12 shows the estimated DRR in less reverberant environment (the approximate reverberation time was 0.7 s). We can still find the similar trend (the method overestimates DRR for longer distance) in this result.

IV. Investigating Influence of Environment Acoustics

To see how the physical characteristics of the environment affect the proposed method, we performed simulations of different environmental conditions. We investigated the influence of reverberation time, early reflections, and room size. We examined not only the trend of the DRR difference depending on these characteristics but also the results of the analysis of variance (ANOVA) [5] that quantitatively tests for significant differences among multiple parameters.

TABLE IV
RESULTS OF REGRESSION ANALYSIS OF DRR DIFFERENCES IN THE EXPERIMENTAL RESULTS OF FIG. 10.

Regression statistics	
Correlation	0.95
Determination	0.91

	Coefficient	Standard error	t-statistic	p-value	Lower 95%	Upper 95%
Intercept	0.29	0.04	6.78	< 0.01	0.21	0.38
Distance to sound source	1.52×10^{-2}	0.02×10^{-2}	88.69	< 0.01	1.49×10^{-2}	1.55×10^{-2}

TABLE V
RESULTS OF TWO-WAY ANOVA TESTING FOR SIGNIFICANT DIFFERENCES IN THE DATA FOR DIFFERENT REVERBERATIONS AND DISTANCES TO SOUND SOURCE.

Source of variation	SS	df	MS	F	p-value
Absorption coefficient	979.76	5	195.95	661.67	< 0.01
Distance to sound source	51056.17	4	12764.04	43100.00	< 0.01
Interaction	1270662	20	635.33	2145.29	< 0.01
Within	879.56	2970	0.30		

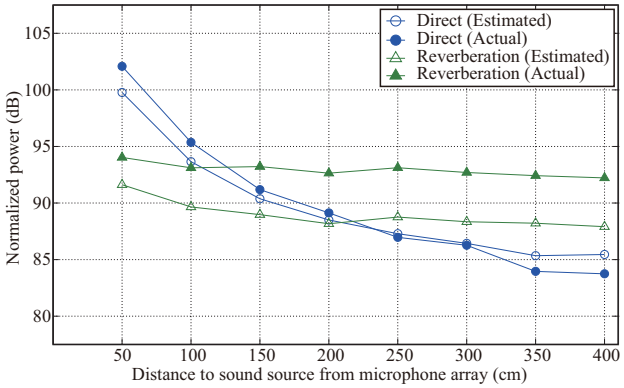


Fig. 11. Comparison of estimated and actual power spectra of direct sound and reverberation in actual acoustic environment.

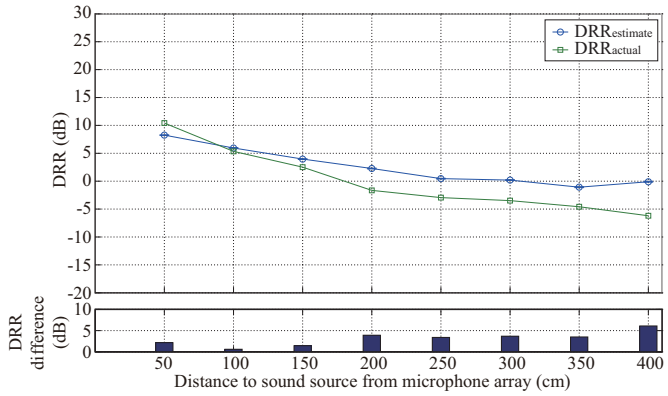


Fig. 12. DRR estimation in actual reverberant room ($T_{60} \approx 0.7$ s).

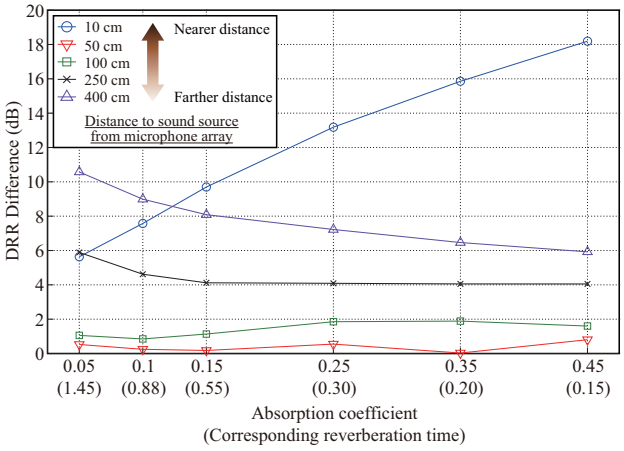


Fig. 13. Log DRR difference measured for different reverberation times. Each line shows DRR difference measured at a particular distance to the sound source from the microphone array. Values in parentheses on the x-axis show approximate reverberation times corresponding to each absorption coefficient.

A. Influence of reverberation time

Figure 13 shows the DRR differences measured in environments with different absorption coefficients (or reverberation time). Note that the environment gets less reverberant as the absorption coefficient increases. The approximate reverberation time measured from a given impulse response is also indicated in parentheses.

There are different trends in the error curve depending on the distance to the sound source. For longer distances, the error increases as the room gets more reverberant. When the room is unreverberant, it would be easy to measure the direct sound because the amount of reverberation, which obscures the direct sound and causes the error, as stated in Sec. III-B, will be small. For medium and shorter distances, the performance was mostly independent of the reverberation except at very short distances, i.e., 10 cm, where more reverberation resulted in higher performance. In this case, the less reverberant situation gave the worst result because the reverberant component will be incorrectly estimated if the diffuseness assumption does not hold.

Table V shows the results of a two-way ANOVA where the factors are the absorption coefficients and the distance to the sound source from the microphone array. The p-values are smaller than 0.01, which means the data has significant differences among the reverberation times and among the distances to the sound source.

B. Influence of early reflections

As we stated using Fig. 1, the impulse response between the sound source and microphone can be classified into three components: direct sound, early reflections, and reverberation. However, the spatial correlation model accounts for the direct

TABLE VI
RESULTS OF TWO-WAY ANOVA TESTING FOR SIGNIFICANT DIFFERENCES IN DATA MEASURED AT DIFFERENT MICROPHONE ARRAY POSITIONS AND DIFFERENT ABSORPTION COEFFICIENTS.

Source of variation	SS	df	MS	F	p-value
Position of microphone array	58.40	1	58.40	3.24	0.07
Absorption coefficient	1730.88	5	346.18	19.23	< 0.01
Interaction	13.84	5	2.77	0.15	0.98
Within	107798.93	5988	18.00		

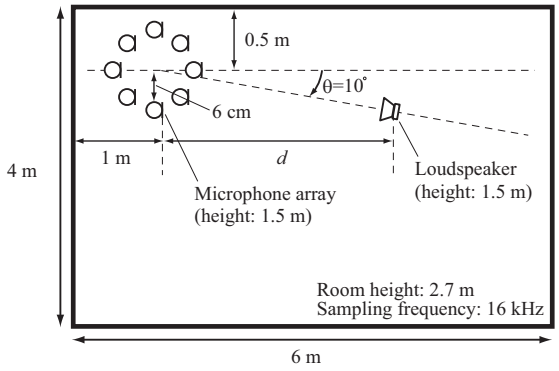


Fig. 14. Position of microphone array and sound source at edge of room in simulation to evaluate the effect of early reflections. The microphone array and speaker are positioned parallel from the default positions shown in Fig. 3.

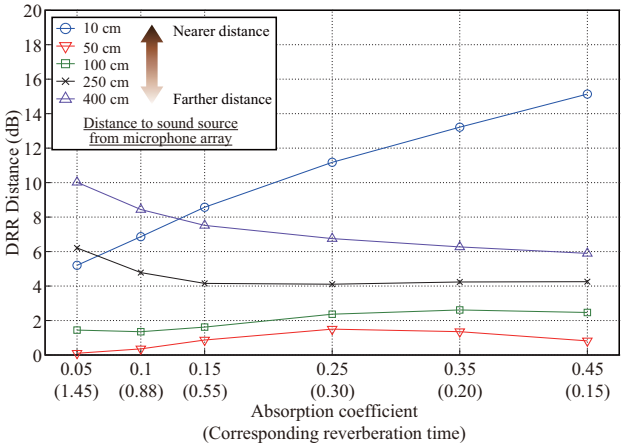


Fig. 15. Log DRR difference measured near a wall of the room. Each line shows DRR difference measured at a particular distance to the sound source from the microphone array. Values in parentheses on the x-axis show approximate reverberation times corresponding to each absorption coefficient.

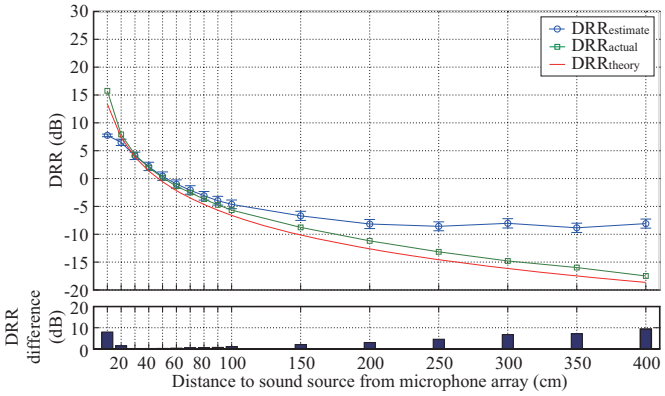


Fig. 16. Comparison of estimated and actual DRR for small room.

sound and the reverberation, but not the early reflections. As the early reflections are mainly sound reflected by the ceiling, floor, and walls, their amount is larger near the walls than at the centre of a room. Thus, we evaluated the influence of early reflections by comparing the DRRs estimated at the edge and centre of a room. Figure 14 shows the locations of the microphone array and sound source for the simulation.

Figure 15 shows the DRR differences when the microphone array was located near a wall. The DRR differences were not much different from those measured at the room centre (Fig. 13). We can be convinced about this fact by looking at Tab. VI, which shows the results of two-way ANOVA whose factors are the microphone array position and the absorption coefficient. From this result, we cannot find a significant difference for different positions of the microphone array because the p-value is larger than 0.01. Thus, we can say that the early reflections have less influence on the DRR estimation accuracy compared with other factors such as ignoring the correlation between the direct and reverberant components, which we discussed in Sec. III-B.

C. Influence of room size

We investigated the influence of the room size because the volume of a room is an important factor that determines the room's acoustics. We simulated a larger room, $3 \times 5 \times 2.7$ (D $\times$ W $\times$ H [m]). and a smaller room, $6 \times 9 \times 2.7$, than the default room. The microphone array was located at the centre between the top and bottom walls, and 1 m away from the left wall as in the default settings described in Fig. 3. The other parameters were set to the default values.

Figure 16 and Fig. 17 show the estimated DRRs for the small and large rooms. Basically, the same trend as we saw in Fig. 4 appears in these results as well; however, the estimates for the small room were a bit more accurate than those for the large room at middle distances (30 – 150 cm). The results of the ANOVA (Tab. VII) validate this trend; the p-value is smaller than 0.01. A cause of this trend would be that the diffusion assumption of the spatial correlation model is less valid for a larger room; the environment becomes close to being anechoic when the size of room increases to that of a free field.

V. USING DRR IN SOUND SOURCE DISTANCE ESTIMATION

Finally, as an example of application of the DRR calculated by the proposed method, we show the results of estimating

TABLE VII
RESULTS FOR DIFFERENT ROOM SIZES. SHOWN ARE THE RELATIONS BETWEEN THREE DIFFERENT ROOM SIZES. ALL 16 DISTANCES TO THE SOUND SOURCE WERE MEASURED.

Source of variation	SS	df	MS	F	p-value
Between groups	131.16	2	65.58	7.09	0.001
Within groups	44374.35	4797	9.25		

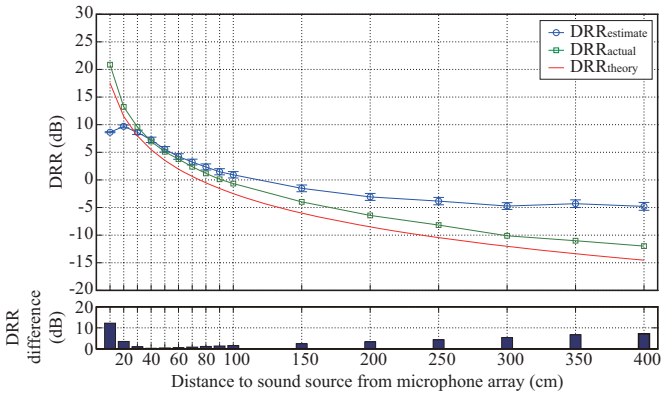


Fig. 17. Comparison of estimated and actual DRR for large room.

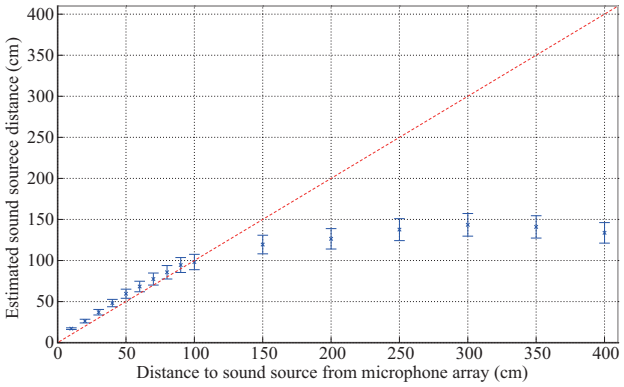


Fig. 18. Results of estimating distance of white noise sound source from microphone array. Each "x" shows the average of the estimated distances with the standard deviation denoted by error bars, and the dashed line shows the correct value.

sound source distances. Note that the method stated below is only an example; the user is free to employ other DRR-based distance estimation methods.

The distance is directly calculated from the estimated DRR by using the following equation. Equation (15) is derived from the theoretical relation between distance and DRR, i.e., Eq. (13). Note that the surface area of the walls A and the average absorption coefficient $\bar{\alpha}$ have to be known preliminarily.

$$d_{\text{estimate}} = \sqrt{\frac{A\bar{\alpha}}{16\pi}} 10^{-(\text{DRR}_{\text{estimate}}/20)} \tag{15}$$

In the simulation, we used the default settings (as in Sec. III-B) except for the absorption coefficient, which was set as $\bar{\alpha} = 0.05$ (or $T_{60} \approx 1.45$ sec). Because the distance estimation would normally be performed for more realistic signals than white noise, we also did simulations with speech signals (five male and five female voices, stating both English and Japanese sentences). Due to the nonstationarity of the speech signal, the frames without speech could degrade the DRR estimation accuracy. Therefore, we applied voice activity detection (VAD) in order to omit the frames without speech before the calculation of the covariance matrix. For the VAD, we simply determined the frames that satisfied the following condition as speech frames. The threshold ν was set as 0.1.

$$X_P^{(m)}(t) > \nu \cdot \max_{t \in \mathcal{T}} \{X_P^{(m)}(t)\} \tag{16}$$

subject to

$$X_P^{(m)}(t) := \sqrt{\sum_{\omega} |X^{(m)}(\omega, t)|^2} \tag{17}$$

where $\mathcal{T}$ is the set of all frames.

Figures 18 shows the results of the distance estimation for white noise. We had mostly correct results up to about 100 cm, but the errors suddenly increased beyond 150 cm. This trend can also be seen in the regression analyses shown in Tab. VIII. We see the slope of the regression line keeps 1.00 with large values for the coefficient of determination up to 80 cm; nonetheless, it includes a small bias given by the non-zero intercept. This implies the proposed DRR estimation cannot be used if the source is far away. A similar trend can also be seen in the results for the speech signals (Fig. 19).

In conclusion, the proposed DRR estimation method can be utilised to measure the sound source distances for a restricted range of distances. This restriction depends on the distances at which the proposed method correctly estimates DRR. It is interesting that the human auditory system shows a very similar trend regarding distance perception; there is a maximum distance at which a human can distinguish the source distance, called the "auditory horizon effect" [6], [7]. Investigating this analogy to the auditory distance perception may lead to a way to improve the accuracy of DRR estimation at longer distances. There are some recent works for sound source localisation that may outperform the proposed methodology. However, it is still interesting that the source distance can be calculated through DRR, which is estimated by using microphone array that has relatively short aperture size.

VI. CONCLUDING REMARKS

We have presented a new DRR estimation method using a microphone array. The method is based on a "spatial correlation" model wherein the direct sound is a plane wave arriving from the direction of the sound source while reverberation arrives from every direction uniformly, i.e. diffuse sound. The proposed method is able to estimate the energy of the direct sound and reverberation directly from the microphone

TABLE VIII
REGRESSION ANALYSES OF DIFFERENT DISTANCE RANGES.

Range of distances [cm]	10–30	10–80	10–100	10–150	10–200
Correlation	0.96	0.97	0.97	0.95	0.92
Determination	0.92	0.95	0.94	0.91	0.85
Intercept	6.79	7.37	9.64	17.15	26.17
Slope	1.00	1.00	0.94	0.78	0.60

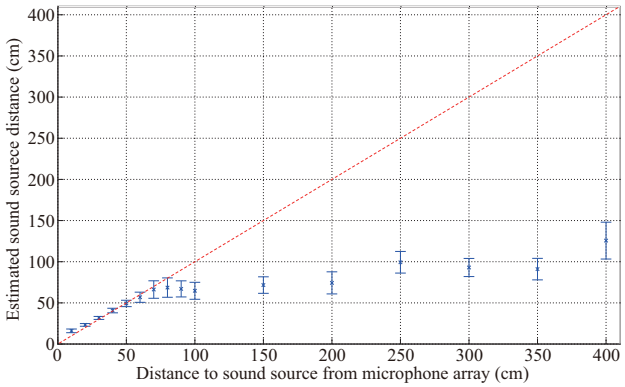


Fig. 19. Results of estimating the distance of ten different speech signals from microphone array. Each "x" is the average estimated distance with the standard deviation denoted by error bars, and the dashed line shows the correct value.

array's measurements, and it does not require preliminary measurement of the impulse response.

The simulation results confirmed that the proposed method is effective especially in reverberant environments. At the same time, we found that the distance range over which the proposed method is able to estimate DRR accurately is restricted. The lower bound of this range is determined by the aperture size of the microphone array because the modelling of the spatial correlation that is based on the plane wave assumption does not hold for sound sources located at short distances from the array. On the other hand, the upper bound of the range is determined by the ratio of the direct component to the reverberant components, which is DRR itself. We investigated the performance of the proposed method in various physical environments and estimated sound source distances based on the estimated DRR.

In the future, we will try to extend the range of distances over which the proposed method correctly estimates DRR and devise measures against various interfering noises. This includes further investigation into the cause of DRR estimation errors that were seen at the long distances. Moreover, it is interesting that the human auditory system shows a similar trend in its distance perception. Investigating the analogy to auditory distance perception may be a way to improve the accuracy of the distance estimation using DRR.

APPENDIX
DERIVATION OF d_{pq}^{S}

The derivation of components in the spatial correlation matrix under the spherical wave model d_{pq}^{S}, which was introduced in Eq. (14), is explained. Let us first define the direct component of p-th microphone input signal under the spherical wave model [24]

$$X^{(p)}(\omega,l) = S(\omega,l)\cdot \frac{||\mathbf{r}_S-\mathbf{r}_o||}{||\mathbf{r}_S-\mathbf{r}_p||}\exp\left\{\frac{j\omega}{c}(||\mathbf{r}_S-\mathbf{r}_p||-||\mathbf{r}_S-\mathbf{r}_o||)\right\}, \quad (18)$$

where $\mathbf{r}_S$, $\mathbf{r}_p$, and $\mathbf{r}_o$ denote the coordinates of sound source, p-th microphone, and reference microphone, respectively. Then the spatial correlation between microphone p and q is derived by

$$E[X^{(p)}(\omega,l)X^{(q)*}(\omega,l)] = P_{\mathrm{D}(\omega)}\cdot \frac{||\mathbf{r}_S-\mathbf{r}_o||^2}{||\mathbf{r}_S-\mathbf{r}_p||||\mathbf{r}_S-\mathbf{r}_q||}\exp\left\{j\frac{\omega}{c}(||\mathbf{r}_S-\mathbf{r}_p||-||\mathbf{r}_S-\mathbf{r}_q||)\right\}. \quad (19)$$

Thus d_{pq}^{S} is given by

$$d_{pq}^{\mathrm{S}} = \frac{||\mathbf{r}_S-\mathbf{r}_o||^2}{||\mathbf{r}_S-\mathbf{r}_p||||\mathbf{r}_S-\mathbf{r}_q||}\cdot\exp\left\{j\frac{\omega}{c}(||\mathbf{r}_S-\mathbf{r}_p||-||\mathbf{r}_S-\mathbf{r}_q||)\right\}. \quad (20)$$

ACKNOWLEDGMENTS

The authors wish to thank the editor and the anonymous reviewers for their valuable comments and suggestions. We would also thank Mr. Shinya Kokubo of NTT Advanced Technology Corporation for his support in collecting the data in the reverberant chamber.

REFERENCES

[1] 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 *Proceedings of ICASSP 2010*, 2010, pp. 149–152.

[2] J. Jo and M. Koyasu, "Measurement of reverberation time based on the direct-reverberant sound energy ratio in steady state," *Proceedings of Inter-noise 75*, pp. 579–582, 1975.

[3] E. Larsen, C. Schmitz, C. Lansing, W. O'Brien Jr., B. Wheeler, and A. Feng, "Acoustic scene analysis using estimated impulse responses," in *Conference Record of the Thirty-seventh Asilomar Conference on Signals, Systems & Computers*, vol. 1, 2003, pp. 725–729.

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

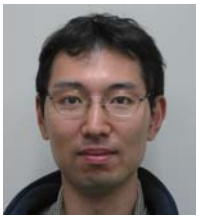
[5] S. H. Nielsen, "Auditory distance perception in different rooms," *J. Audio Eng. Soc*, vol. 41, no. 10, pp. 755–770, 1993. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=6982

[6] A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, pp. 517–520, 1999.

[7] A. W. Bronkhorst, "Modeling auditory distance perception in rooms," in *Proc. EAA Forum Acusticum Sevilla*, 2002.

[8] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *Journal of Acoustic Society of America*, vol. 112, no. 5, pp. 2110–2117, 2002.

[9] ——, "Assessing auditory distance perception using virtual acoustics," *Journal of Acoustic Society of America*, vol. 111, no. 4, pp. 1832–1846, 2002.
[10] P. Zahorik, D. S. Brungrat, 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.
[11] M. Brandstein and D. Ward, Eds., *Microphone Arrays*. Springer-Verlag, 2001.
[12] P. Holmberg, "Ultrasonic sensor array for position and rotation estimates of planar surface," *Sensors and Actuators A*, vol. 44, pp. 37–43, 1994.
[13] M. Omologo and P. Svaizer, "Use of crosspower-spectrum phase in acoustic event location," *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 3, pp. 288–292, 1997.
[14] F. Asano, H. Asoh, and T. Matsui, "Sound source localization and separation in near field," *IEICE Transactions on Fundamentals*, vol. E83-A, no. 11, pp. 2286–2294, 2000.
[15] 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.
[16] S. Vesa, "Sound source distance learning based on binaural signals," in *Proceedings of 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007.
[17] ——, "Binaural sound source distance learning in rooms," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1498–1507, 2009.
[18] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993, ch. 4.4.2.
[19] ——, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.
[20] M. Tohyama, *The Nature and Technology of Acoustic Space*. Academic Press, 1995.
[21] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
[22] H. Kuttruff, *Room Acoustics*. Applied Science Publishers LTD, 1973, ch. III.5.
[23] S. Doclo and M. Moonen, "Design of far-field and near-field broadband beamformers using eigenfilters," *Signal Processing*, vol. 83, pp. 2641–2673, 2003.
[24] M. Brandstein and D. Ward, Eds., *Microphone Arrays*. Springer-Verlag, 2001, ch. 2.3, pp. 24–25.



Yusuke Hioka (S'04–M'05) received B.E., M.E., and Ph.D. degrees in electrical engineering in 2000, 2002, and 2005 from Keio University, Yokohama, Japan. In 2005, he joined Nippon Telegraph and Telephone Corporation (NTT), and currently he is a Research Engineer at NTT Cyber Space Laboratories. From June to December 2010, he was also a Visiting Research Fellow at Massey University, Wellington, New Zealand, and since January 2011, he has been a visiting researcher at Victoria University of Wellington, New Zealand. His research interests include microphone array signal processing and room acoustics. Dr. Hioka is a member of IEICE and the Acoustical Society of Japan.

Kenta Niwa received B.E. and M.E. degrees from Nagoya University in 2006 and 2008, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 2008, he has been engaged in research on microphone arrays. Currently, he is a researcher at NTT Cyber Space Laboratories. He was awarded the Awaya Prize by the Acoustical Society of Japan (ASJ) in 2010. He is a member of the Acoustical Society of Japan.

Sumitaka Sakauchi received the B.S. degree from Yamagata University in 1993, the M.S. degree from Tohoku University in 1995, and the Ph.D. degree from Tsukuba University in 2005. In 1995, he joined Nippon Telegraph and Telephone (NTT). Since then, he has been engaged in research on acoustic echo cancellers and noise reduction. He is now a Senior Research Engineer in the Acoustic Information Group of NTT Cyber Space Laboratories. He received the Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2001 and the Young Engineer Award from the Acoustical Society of Japan (ASJ) in 2003. He is a member of IEICE and the Acoustical Society of Japan.

Ken'ichi Furuya received the B.E. and M.E. degrees in acoustic design from Kyushu Institute of Design, Fukuoka, Japan, in 1985 and 1987, respectively, and the Ph.D. degree from Kyushu University, Japan, in 2005. He joined Nippon Telegraph and Telephone Corporation (NTT) in 1987. He is now a Senior Research Engineer at NTT Cyber Space Laboratories, Tokyo, Japan. His current research interests include signal processing in acoustic engineering. Dr. Furuya was awarded the Sato Prize by the Acoustical Society of Japan (ASJ) in 1991. He is a member of the Acoustical Society of America, and IEICE.

Yoichi Haneda received the B.S., M.S., and Ph.D. degrees from Tohoku University, Sendai, in 1987, 1989, and 1999, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1989, he has been investigating the modelling of acoustic transfer functions, acoustic signal processing, and acoustic echo cancellers. He is now a Senior Research Engineer, Group Leader at NTT Cyber Space Laboratories. He received the paper awards from the Acoustical Society of Japan and from the Institute of Electronics, Information, and Communication Engineers of Japan in 2002. Dr. Haneda is a member of the IEICE and the Acoustical Society of Japan.