Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Evaluating Estimation of Direct-to-Reverberation Energy Ratio using D/R Spatial Correlation Matrix Model

Yusuke Hioka, Kenta Niwa, Sumitaka Sakauchi, Ken'ichi Furuya and Yoichi Haneda

NTT Cyber Space Laboratories, NTT Corporation, 3-9-11 Midori-cho, Musashino, Tokyo 180-8585, Japan

PACS: 43.55.Br,43.55.Mc,43.60.Fg

ABSTRACT

We evaluate the accuracy of direct-to-reverberation energy ratio (DRR) estimation that uses the direct sound to reverberation spatial correlation matrix model (DRSC model). The DRSC model, which expresses the spatial correlation matrix with two different matrices of direct sound and reverberation, assumes that the direct sound propagates only from the direction of the sound source but that the reverberation arrives from every direction uniformly. The DRR is calculated from the power spectra of both the direct sound and reverberation that are estimated from the spatial correlation matrix of the observed signal. The method was evaluated using various scales in both simulated and actual acoustical environments. The evaluation results confirmed the effectiveness of DRR estimation using the DRSC model and also revealed its limitations.

INTRODUCTION

Estimating the direct-to-reverberant energy ratio (DRR) is helpful for determining the features of a reverberant environment because various acoustic parameters, such as reverberation time, diffuseness, etc., can be calculated from DRR (Jo and Koyasu 1975). There is also another important aspect in DRR relating to human hearing. Recent research on human hearing has concluded that DRR may provide absolute distance information, especially in reverberant environments (Zahorik et al. 2005). There are several conventional instruments to measure the distance to the sound source. For example, an ultrasonic sensor is a well-known tool for distance measurement, but it costs more than passive methods that use only microphones, not transmitters. There are also some works on estimating source distance by using a microphone array(Asano and H. Asoh 2000, Yu and Silverman 2004). Although the microphone array techniques require only microphones, they fail to correctly estimate the distance when the environment is highly reverberant. This is because the time and sound level differences of the arrival of direct sound between microphones, which are exploited as keys to determining the sound source positions, becomes ambiguous due to the existence of reverberation. Even in such an environment, we can still estimate distance from the DRR because DRR keeps its one-to-one relation with the source distance in a reverberant environment.

Several methods are available for estimating DRR. The most primitive way is to calculate DRR directly from the impulse response. However, this is a complicated process because measurement of the impulse response is required. Larsen *et al*. proposed a method for estimating DRR from simply the short beginning part of the impulse response (Larsen et al. 2003), but it still necessitates prior processing to identify that part. Lu *et al*. also proposed a procedure to estimate DRR (Lu and Cooke 2008). They utilized a binaural input signal and estimated the energy of the reverberant component by eliminating the direct component using an equalization-cancellation (EC) technique. To eliminate direct sound, the EC technique exploits the large difference between the direct sound and reverberation that exists in the inter-channel (or spatial) correlation of the binary input signal. However, the EC technique loses its DRR estimation accuracy in highly reverberant environments because it is based on a model in which no reverberation component propagates from the same direction as the sound source.

We have proposed a DRR estimation method using the direct sound to reverberation (D/R) spatial correlation matrix model (called "DRSC model" hereafter), which consists of the spatial correlation matrices of direct sound and reverberation (Hioka et al. 2010). The DRSC model assumes that the direct sound propagates only from the direction of the sound source but that the reverberation arrives from every direction uniformly. Then, we calculate DRR from the power spectra of both components, which are estimated from the correlation matrix of the observed signals. In the previous report, the discussion on the effectiveness of the method was limited to a particular condition, so its limitations and advantages in comparison to the conventional method were not investigated. Therefore, in this contribution, we evaluate the performance of the proposed method under various conditions and then reveal its limitations. Furthermore, we compare the results of DRR estimation with those of the conventional method based on the EC technique(Lu and Cooke 2008).

This paper is organized as follows. We first introduce the DRSC model and then propose a method for estimating DRR based on the model. Then, we evaluate the DRSC model by measuring the accuracy of the estimated DRR from various viewpoints. We also discuss the influence of parameters set by users to investigate the appropriate values for each parameter. Finally, the evaluation results that were obtained in actual environments are shown, and then we conclude this paper with some comments.

DRR ESTIMATION BASED ON DRSC MODEL

Modelling of spatial correlation matrix

First, we decompose the transfer function $H(\omega)$ between a sound source and a microphone into two components, the direct component $H_{\mathrm{D}}(\omega)$ and reverberant component $H_{\mathrm{R}}(\omega)$, as described in Fig. 1. Note that the early reflection of the im-

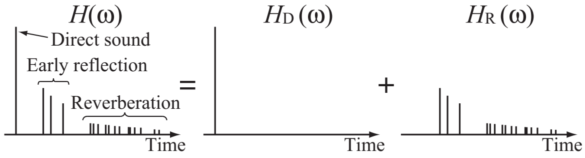

Figure 1: Decomposition of transfer function.

pulse response is also included in $H_{\mathrm{R}}(\omega)$. When we have an M-sensors 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. By this expression, the cross correlation between the p-th and q-th microphones is 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, under the assumption that 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, the spatial correlation matrix of the microphone array $\mathbf{R}(\omega)$ can be approximated by two matrices given by

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

where

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

$$d_{pq} = \exp\left(j\omega \frac{(\mathbf{r}_p - \mathbf{r}_q)\cdot \mathbf{a}(\theta)}{c}\right), \tag{5}$$

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

and $\mathbf{r}_m$, c, and ^{H} are the coordinates of the m-th microphone, sound speed, and Hermitian transformation, respectively, and $||\cdot||$ is the Euclidean distance. Furthermore, $\mathbf{a}(\theta) = [\sin\theta, \cos\theta]^T$ is the look-direction unit vector of sound that propagates from θ when the y-axis is set to 0 deg as described in Fig. 2.

On the right side of Eq. (3), the first term expresses the spatial correlation of the direct component. As there is a time difference of arrival between microphones in the cross-correlation of the direct sound, the spatial correlation is expressed by simple phase difference. In the modelling of the second term, we utilized the feature that the spatial correlation of diffuse sound can be expressed by a sinc function (Tohyama 1995). In Eq. (3), $P_{\mathrm{D}}(\omega)$ and $P_{\mathrm{R}}(\omega)$ are defined by

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

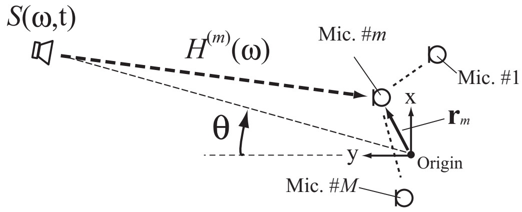

Figure 2: Modelling of observed signal.

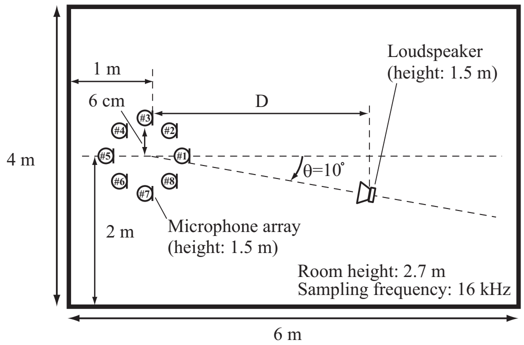

Figure 3: Positions of microphone array and speaker in simulated reverberant room.

Note that in the derivation of Eq. (3), the aperture size of the microphone array is assumed to be sufficiently small. This means that the array recognizes the received sound as a plain wave and that the magnitude of the transfer function for each microphone can be considered as identical, i.e.,

$$|H_{\mathrm{D}}^{(p)}(\omega)||H_{\mathrm{D}}^{(q)}(\omega)| = |H_{\mathrm{D}}(\omega)|^2$$

and

$$|H_{\mathrm{R}}^{(p)}(\omega)||H_{\mathrm{R}}^{(q)}(\omega)| = |H_{\mathrm{R}}(\omega)|^2.$$

However, if we assume the received sound is a spherical wave, the spatial correlation between microphones, which is modelled as

$$d_{pq} = \frac{||\mathbf{r}_S - \mathbf{r}_q||}{||\mathbf{r}_S - \mathbf{r}_p||} \exp\left(j\frac{\omega}{c}\left(||\mathbf{r}_S - \mathbf{r}_p|| - ||\mathbf{r}_S - \mathbf{r}_q||\right)\right), \tag{7}$$

is used instead of Eq. (5). Here, $\mathbf{r}_S$ denotes the coordinates of a sound source position preliminarily assumed (called the "focusing position" hereafter).

DRR estimation using power spectra of direct and reverberant components

As the microphone array configuration is generally known *a priori*, and the direction of the sound source can be estimated by various conventional methods (Brandstein and Ward 2001), d_{pq} and r_{pq} in Eq. (3) can be specified. Thus, we estimate the unknown power spectra of both direct and reverberant components, $P_{\mathrm{D}}(\omega)$ and $P_{\mathrm{R}}(\omega)$, by solving the simultaneous equation given by Eq. (8), which is derived by reformulating Eq. (3).

$$\underbrace{\begin{bmatrix} 1 & 1 \\ d_{12} & r_{12} \\ \vdots & \vdots \\ d_{1M} & r_{1M} \\ d_{21} & r_{12} \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix}}_{\mathbf{F}(\omega)} \underbrace{\begin{bmatrix} P_{\mathrm{D}}(\omega) \\ P_{\mathrm{R}}(\omega) \end{bmatrix}}_{\mathbf{P}(\omega)} = \underbrace{\begin{bmatrix} R_{11}(\omega) \\ R_{12}(\omega) \\ \vdots \\ R_{1M}(\omega) \\ R_{21}(\omega) \\ R_{22}(\omega) \\ \vdots \\ R_{MM}(\omega) \end{bmatrix}}_{\widetilde{\mathbf{R}}(\omega)} \tag{8}$$

Table 1: Basic settings of conditions and parameters set in computer simulation.

Fs: Sampling frequency [Hz]	16,000
M: Number of microphones	8
Microphone arrangement	circular
Radial size of array [cm]	6
$\bar{\alpha}$: absorption coefficient	0.15
Frame length [samples]	512
SNR [dB]	∞

Table 2: Relation between average absorption coefficient and reverberation time.

$\bar{\alpha}$	0.05	0.10	0.15	0.25	0.35	0.45
T_{60}[sec]	1.45	0.88	0.55	0.30	0.20	0.15

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 observed signals. The estimated power spectra of direct and reverberant components are given by solving Eq. (8) using the least-square method given by

$$\widehat{\mathbf{P}}(\omega) = \mathbf{F}^{+}(\omega)\widetilde{\mathbf{R}}(\omega), \qquad (9)$$

where $^{+}$ and $\widehat{\ }$ are the Moore-Penrose pseudo inverse and estimated value, respectively.
Finally, the estimated DRR is given by using the estimated power spectra $\widehat{P_{\mathrm{D}}}(\omega)$ and $\widehat{P_{\mathrm{R}}}(\omega)$ in the following Eq. (10).

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

EVALUATION BY COMPUTER SIMULATION

Simulation settings and evaluation criteria

To evaluate the performance of the proposed DRR estimation using the DRSC model, we performed experiments in simulated reverberant environments. The basic settings of the assumed conditions and parameters used in the simulation are described in Tab. 1 and Fig. 3. In Tab. 1, ∞ in SNR means the input signal is noise-free, i.e., no noise signal is added to the input signal. The sound source was 3-s long Gaussian white noise unless otherwise stated, and the input signals of a microphone array were prepared by convolving the simulated impulse response generated by the image method (Allen and Berkley 1979). For each condition, we applied the DRR estimation for 100 trials.

As an evaluation criterion, we calculated the log DRR difference in the estimated DRR, defined by

$$\varepsilon_{\mathrm{DRR}} = |\mathrm{DRR}_{\mathrm{estimate}} - \mathrm{DRR}_{\mathrm{actual}}|. \qquad (11)$$

The actual DRR was directly calculated from the impulse response, defined by

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

As DRR varies depending on the distance from the microphones, evaluating the ratio of the estimated DRR to the actual DRR is more reasonable than evaluating the difference of DRRs. 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 log DRR difference should be 0 dB. Furthermore, for several results shown below, we also calculated principle-based

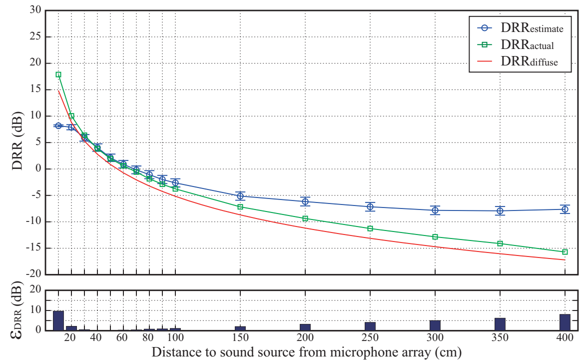

Figure 4: Comparison between estimated and actual DRR in simulated room. In upper graph, lines with circles and squares show $\mathrm{DRR}_{\mathrm{estimate}}$ and $\mathrm{DRR}_{\mathrm{actual}}$, respectively. Other line shows principle-based DRR in diffuse sound field, which helps to determine distance of actual reverberation from completely diffuse field. Log DRR difference between $\mathrm{DRR}_{\mathrm{estimate}}$ and $\mathrm{DRR}_{\mathrm{actual}}$ is shown in lower graph.

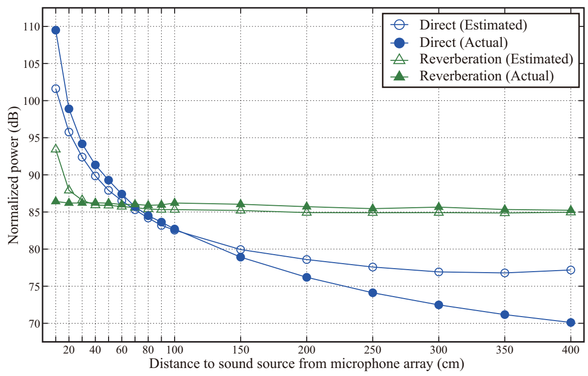

Figure 5: Comparison between estimated and actual power spectra of direct sound and reverberation, which are denoted by circular and triangular markers, respectively.

DRR in a diffuse sound field ($\mathrm{DRR}_{\mathrm{diffuse}}$)(Tohyama 1995), given by

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

This value helps to determine the distance of the actual reverberation from the completely diffuse field. Here, S and $\bar{\alpha}$ are the surface area of the walls and the average absorption coefficient, respectively. As a reference for how reverberant the environment is, the approximate values of reverberation time, which are calculated from a given impulse response, for each $\bar{\alpha}$ are shown in Tab. 2.

DRR estimation with basic parameter settings

Figure 4 shows the results of DRR estimation performed with the basic condition and parameter settings. The upper graph shows the average of estimated DRRs with their standard deviations, while the lower graph shows the log DRR difference. These results reveal a trend of the error increasing in both very near and far distances. To discover the cause of these errors, both the actual and estimated power of the direct sound and reverberation were calculated (Fig. 5). The results show that the estimation error in nearer distances is caused by the estimation error of both the direct sound and reverberation power, but that the error in farther distances is mainly caused by the error in the estimated power of the direct sound.

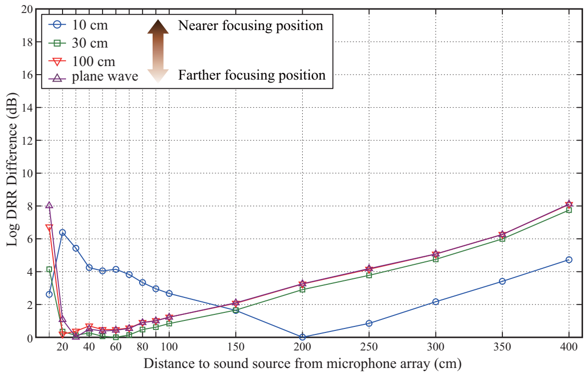

Figure 6: Log DRR difference calculated when spatial correlation of direct sound is modelled by assuming spherical wave. Each line shows log DRR difference measured for changed focusing positions. For comparison, line with upward triangles shows log DRR difference of estimated DRR using plane wave assumption.

For nearer distances, the discrepancy between the modelled and actual spatial correlation of the direct component could be conceived as the cause of the error. This is because the plane wave assumption of the received sound is only valid for the sound sources located in the far-field defined by $d > \frac{D^2}{\lambda}$ (Doclo and Moonen 2003), where D is the array aperture size. For the octagonal microphone array used in this simulation ($D = 12$ cm), the boundary distance between far-field and near-field was approximately 33 cm at 8 kHz. This is supported by the results obtained where the estimation error started to rapidly increase when d became smaller than 30 cm. To prove this assumption, we performed a simulation by assuming the received sound was a spherical wave using Eq. (7) for d_{pq}. Figure 6 shows the log DRR difference of the estimated DRR when the focusing position was changed. The log DRR difference of very near distances, i.e., $d = 10$ cm, decreased when the focusing position was set appropriately. In other words, the proposed method was able to correctly estimate DRR in nearer distances if it applied the assumption of a spherical wave in its modelling of direct sound. However, its good performance was achieved only in a specific range of distances. As proof, the difference was very small at other focusing positions, compared to the result when the plane wave assumption was applied. Therefore, we can conclude that the modelling of direct sound based on plane wave assumption is reasonable for most cases.

The estimation error found in the farther distances is conceived as the result of the power of the direct component being too small to be accurately detected from the observed signal, which mostly consists of reverberant components. This consideration can be proved from the fact that the estimation accuracy was improved by increasing the number of microphones, which is stated later. Thus, there is an upper limit of the distance for which the method is able to estimate DRR correctly using this particular microphone array configuration.

Evaluation for influence of reverberation time

To determine the influence of room reverberation, the performance of the proposed method in different reverberation conditions was evaluated. Figure 7 shows the log DRR differences measured in an environment with different absorption coefficients. From the results, we can see a trend of the error in farther distances increasing as the room becomes more reverberant, while the error in nearer distances increases as the room becomes less reverberant. Such a trend can also be explained by the effects of modelling errors in the spatial correlation of direct sound for nearer distances, and of low power direct sound for farther distances.

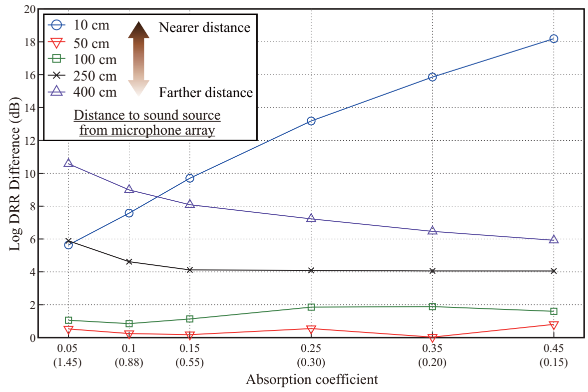

Figure 7: Log DRR difference measured for different reverberation times. Each line shows log DRR difference measured at particular distance to sound source from microphone array. Parenthetic values in x-axis show approximate reverberation time corresponding to each absorption coefficient.

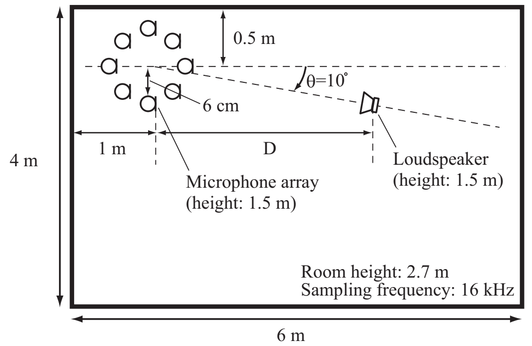

Figure 8: Position of microphone and sound source applied to evaluate influence of early reflection. Both microphone array and speaker are parallel shifted from basic position shown in Fig. 3.

When the room is unreverberant, it is easier to measure the direct sound because the amount of reverberation, which obscures the direct sound, is reduced. This means the direct sound power spectrum will be measured more correctly. Thus the results for farther distances are better when the room is unreverberant. For the nearer distances, except for the specific nearer distance of 10 cm, the estimation accuracy is a little improved as the condition gets closer to the diffuse sound field. This is natural because the DRSC model assumes a fully diffuse sound for the reverberation.

Evaluation for influence of early reflection

As we stated using Fig. 1, the impulse response between the sound source and microphone can be classified into three components: direct sound, early reflection, and reverberation. However, in the definition of the DRSC model mentioned above, we took account of the direct sound and the reverberation, but not the early reflection. As the early reflection is mainly developed by the sound reflection from the ceiling, floor, and walls, the amount of early reflection is larger at the edge than the centre of a room. Thus, here we evaluate the influence of early reflection by comparing the DRRs estimated at the edge and centre of a room. In the simulation, the microphone array and sound source were located as described in Fig. 8.

Figure 9 shows the log DRR differences when the microphone array was located at the edge of a room. In comparison with the

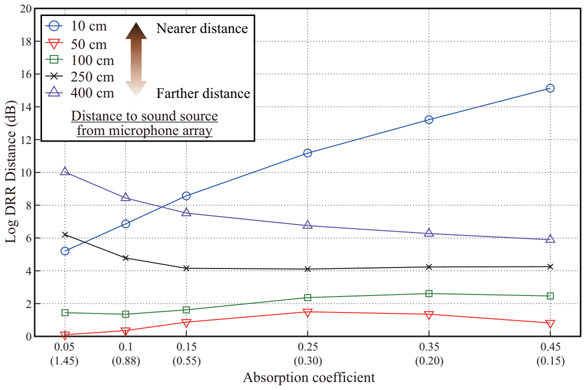

Figure 9: Log DRR difference measured at edge of room. Each line shows log DRR difference measured at particular distance to sound source from microphone array. Parenthetic values in x-axis show approximate reverberation time corresponding to each absorption coefficient.

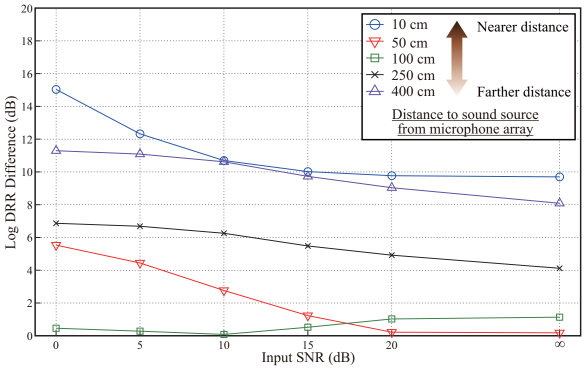

Figure 10: Influence of spatially uncorrelated noise for estimated DRR. Each line shows log DRR difference measured at particular distance to sound source from microphone array.

results measured at the centre of a room shown in Fig. 7, the log DRR difference increased at the distances of 50 and 100 cm. As these distances are in the range at which the proposed method works best, the early reflection may adversely affect the estimation accuracy of the DRR. In contrast, the log DRR differences at nearer and farther distances did not differ much from the result of the log DRR difference measured at the room centre. Thus, we suppose that other factors that cause the log DRR difference growth, such as modelling errors and low power direct sound, affect the result more than early reflection does.

Evaluation for influence of spatially uncorrelated noise

In a practical environment, generally some amount of electric noise is added to the sound received by each microphone. Basically, as such noise is dependent on the property of each microphone, there is no correlation between the noise of each microphone. To investigate how such spatially uncorrelated noise affects the performance of DRR estimation, we examined the log DRR difference for different input SNR. As Fig. 10 shows, the method basically works better as the input SNR increases, but the SNR has a larger impact on the errors of nearer distances than on the errors of farther distances. As conceived above, we considered that the estimation accuracy in farther distances declined because the direct sound component was completely obscured by the reverberation. In the low SNR condition, the situation is quite similar to the conditions for direct sound from farther distances. Therefore, the noise did not have large impact on the estimation accuracy at farther distances.

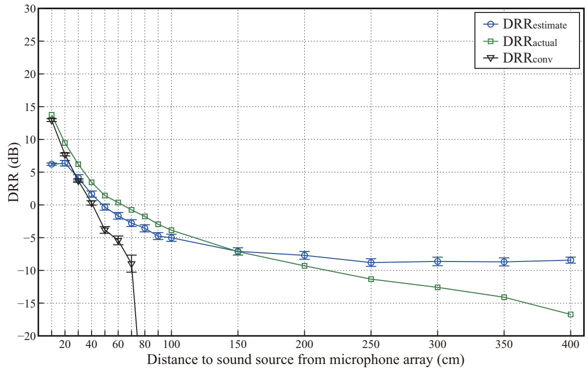

Figure 11: Evaluation of DRR estimation compared to conventional method (highly reverberant condition).

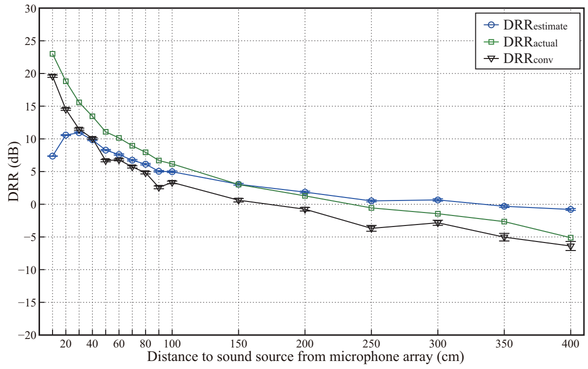

Figure 12: Evaluation of DRR estimation compared to conventional method (little reverberation condition).

Comparison to conventional DRR estimation method

Finally, we compare the accuracy of the DRR estimation to that of the conventional method(Lu and Cooke 2008). Figures 11 and 12 show the DRRs estimated at the basic parameter settings except for the absorption coefficients set at $\bar{\alpha} = 0.15$ and $\bar{\alpha} = 0.45$ respectively. As the conventional method works under the binaural signal, we used only a pair of microphones in the octagonal microphone array (microphones #3 and #7 in Fig. 3), separated by 12 cm.

From Fig. 11, we can see that the conventional method almost failed to estimate DRR correctly in highly reverberant conditions except at very near distances. This is because the method is based on a model in which no reverberation propagates from the same direction as the sound source. As this assumption does not hold in highly reverberant conditions, the method did not work correctly. Furthermore, the effect of this modelling error can also be seen in the unreverberant case in Fig. 12. In the results, there is a difference between the actual DRR and the DRR estimated by the conventional method, which is almost constant at every distance. On the other hand, the proposed method worked well, especially in highly reverberant conditions, because it is based on the DRSC model, which is a more realistic model than that assumed in the conventional method.

DISCUSSION ON INFLUENCE OF PARAMETER SETTINGS

To discover the appropriate values for each of the parameters that are basically set by the users, we investigated the relation between the parameters and DRR estimation accuracy. Here, we discuss the impact of two major parameters: the number of microphones and the frame length.

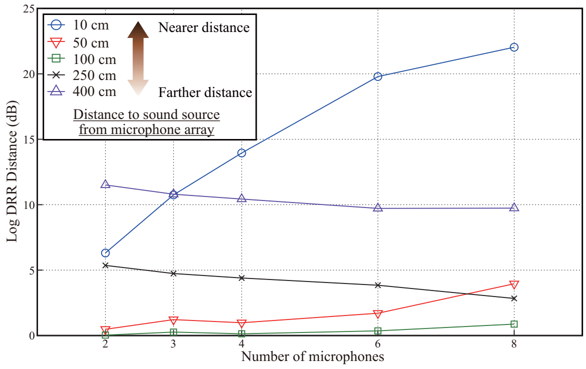

Figure 13: Errors in estimated DRR for different numbers of microphones. Each line shows log DRR difference measured at particular distance to sound source from microphone array.

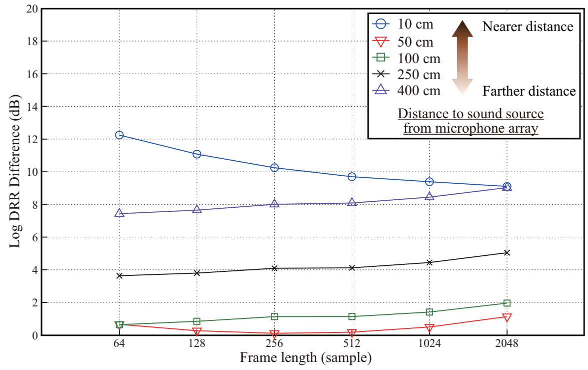

Figure 14: Effect of frame length on calculation for estimated DRR. Each line shows log DRR difference measured at particular distance to sound source from microphone array.

Evaluation for influence of number of microphones

Generally, the number of microphones impinges on the spatial resolution of the microphone array. Here, the influence of that number was investigated by applying the proposed method for a linear equi-spaced microphone array with different numbers of microphones. The inter-microphone distance was set at 4 cm, which does not cause spatial aliasing in the whole frequency band, and we changed the number of microphones from between two to eight.

The results of the examinations are shown in Fig. 13. The log DRR difference increased as more microphones were used in nearer distances, but it decreased in farther distances. As we stated above, the region where the plane wave assumption holds enlarges depending on the array aperture size D. Hence, in the nearer distances, the log DRR difference increased with the number of microphones. For example, when M is 2, the boundary of regions that we assume are far-field and near-field is approximately 4 cm. Therefore, the proposed method mostly estimated DRR correctly even though the sound source was located at 10 cm. On the other hand, for a larger number of microphones, more accurate estimation of DRR was achieved in the farther distances. Generally, the more microphones we use, the better the performance of the array. In the same manner, it is conceived that the method could extract the direct sound obscured by the reverberation more accurately as the number of microphones increased.

Evaluation for influence of frame length

The impact of frame length, which determines the time-frequency resolution of short-time Fourier transform (STFT) calculated in Eq. (2), was investigated. When the frame length is very short, the latter part of reverberation is not included in the same frame. Therefore, the estimation accuracy could be degraded as the frame length becomes much shorter than the length of the impulse response between the sound source and microphone array. When the frame length is longer, on the other hand, the amount of reverberation included in each frame increases. In other words, again the reverberation could obscure the direct sound component, and this may result in increase of the log DRR difference.

Figure 14 shows the results of experiments where frame length was changed. For every sound source distance except for 10 cm, a longer frame length caused a small increase in the log DRR difference, but the effect on the estimation accuracy was small. When the sound source is located very near to the microphone, the reverberation arrives much later than the direct sound reaches the microphone. In contrast, the difference in the arrival time of direct sound and reverberation becomes small

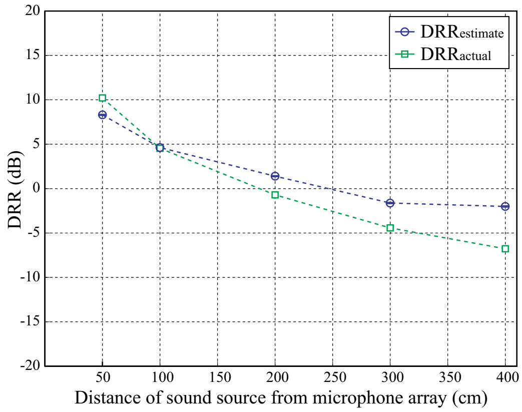

Figure 15: Results of DRR estimation in actual reverberant room.

when the sound source is located farther away. Thus, the shorter frame size affected the estimation accuracy in only the 10-cm case.

RESULTS OF EXPERIMENTS IN REAL ACOUSTIC ENVIRONMENT

To confirm the effectiveness of our method in an actual environment, we also performed an experiment in a real reverberant room. The room size and position of the microphone array used in this experiment were the same as those given in Fig. 3, except that the loudspeaker was located in the direction of $\theta = 0°$. The reverberation time of the room was approximately 400 ms. The estimated DRR and $\mathrm{DRR}_{\mathrm{actual}}$ calculated from the measured impulse response are shown in Fig. 15. The results prove that the proposed method is still effective in actual reverberant environments.

CONCLUDING REMARKS

In this paper, we have evaluated from various viewpoints a method for estimating DRR based on our DRSC model. The proposed method is able to estimate DRR directly from the received sound, and it does not require preliminary measurement of the impulse response.

From the simulation results, we found that a range of distance where the proposed method is able to estimate DRR correctly exists. The lower bound of this range is determined by the aperture size of the microphone array because the modelling of spatial correlation that is based on the plane wave assumption does not hold for the sound sources located at nearer distances. The upper bound of the range is determined by the number of microphones because the estimation errors in farther distances are caused by the power of the direct sound being too small to be detected. This is natural, as the gain of a microphone array is improved by increasing the number of microphones.

We also evaluated the proposed method in various simulation conditions. Basically, the obtained results were logical, and we could explain most of them by the existence of a range where the method correctly estimates DRR, as mentioned above. Furthermore, in a comparison with the conventional DRR estimation method, the proposed method showed much better performance in DRR estimation accuracy.

Improvement of the method for various types of noise, including not only spatial uncorrelated noise but also other interfering sounds that exist in most general environments, is an important future subject of investigation.

REFERENCES

J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65:943–950, 1979.

F. Asano and T. Matsui H. Asoh. Sound source localization and separation in near field. *IEICE Transactions on Fundamentals*, pages 2286–2294, 2000.

M. Brandstein and D. Ward. Springer-Verlag, 2001.

S. Doclo and M. Moonen. Design of far-field and near-field broadband beamformers using eigenfilters. *Signal Processing*, 83:2641–2673, 2003.

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. *Proceedings of 2010 IEEE International Conference of Acousics, Speech and Signal Processing (ICASSP2010)*, pages 149–152, 2010.

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*, pages 579–582, 1975.

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. *Conference Record of the Thirty-seventh Asilomar Conference on Signals, Systems & Computers*, 1:725–729, 2003.

Y .C. Lu and M. Cooke. Binaural distance perception based on direct-to-reverberant energy ratio. *Proceedings of the 11th International Workshop on Acoustic Echo and Noise Control (IWAENC2008)*, pages 409–420, 2008.

M. Tohyama. *The Nature and Technology of Acoustic Space*. Academic Press, 1995.

Y. Yu and H. F. Silverman. An improved tdoa-based location estimation algorithm for large aperture microphone arrays. *Proceedings of 2004 IEEE International Conference of Acousics, Speech and Signal Processing (ICASSP2004)*, pages 77–80, 2004.

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*, 91(3): 409–420, 2005.