TELESCOPIC MICROPHONE ARRAY USING REFLECTOR FOR SEGREGATING TARGET SOURCE FROM NOISES IN SAME DIRECTION

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

NTT Cyber Space Laboratories, NTT Corporation, Tokyo, Japan

ABSTRACT

A spatial sensitivity control method for segregating the sound sources in the same direction by using an acoustic reflector is proposed. Our goal is to clearly pick up the target source at an arbitrary position using a microphone array. Though many methods have been studied for spatial sensitivity control, it is difficult to robustly suppress the power of noise sources in the same direction of the target source in a room. To overcome this problem, we attach a reflector to a microphone array to capture the reflected sounds whose characteristics vary depending on the distance from the array to the source. Assuming that the acoustical properties of the reflector are known e.g., measuring the transfer functions, those reflected sounds can be used as effective clues for segregating the sound sources in the same direction. With the proposed method, a filter for minimizing the output noise power is derived by taking into consideration of the acoustic properties of the reflector. Experiments were conducted in an actual room by using 96 microphones and a large reflector. We confirmed that the spatial sensitivities for segregating the target source at an arbitrary position from noise sources can be achieved by using the proposed method.

Index Terms— Microphone array, Spatial sensitivity control, Acoustical reflector, Transfer function

1. INTRODUCTION

The recent progress in video screen technology has resulted in several new styles of video image expression. For instance, 3D TV or free viewpoint TV[1, 2] can zoom in on an object like we do with a zoom lens to look at the details of an object. Similar to a zoom lens, our goal in this study is to clearly pick up sound of a target source. As a zoom lens narrows the view angle and varies the focal point distance to focus on the target object, we control the spatial sensitivity of a microphone array to emphasize a target sound source located at an arbitrary position while suppressing surrounding noise sources. Hereafter, we call our new array "telescopic microphone array". Applying such a telescopic array in conjunction with 3D TV will make new video screen technologies more attractive.

Various methods for controlling the spatial sensitivity of a microphone array [3] have been studied. Directivity control based on the beamforming method [4] is a popular strategy for emphasizing sound sources arriving from an arbitrary direction. With this method, the steering vector [3] is used in designing filters. However, it cannot be used to suppress noise sources located in the same direction of the target source. On the other hand, the active noise control (ANC) technique can be used to separate the target source located in front of/behind noise sources [5], but it usually requires the room transfer functions of all noise source positions. Since ANC is very sensitive to changes in room transfer functions, it is difficult to achieve spatial sensitivity control that is robust against variation in the room environment. Adaptive beamforming [6] such as with the minimum variance distortionless response (MVDR) method [7], is another option that may solve this problem [8]. However it is also sensitive to changes in room transfer functions required to constrain the beamformer response to the target source.

Our telescopic microphone array enables robust control of spatial sensitivity to overcome this problem. Several studies on microphone arrays that exploit reflection or reverberation to estimate the source distance have been reported [9, 10]. These studies are based on the fact that reverberation is an effective clue for estimating sound source distance for human auditory perception [11]. On the other hand, reflection may also be a useful distance cue because the propagation paths of the reflections vary depending on the distance to the source. To exploit this feature, we artificially generate reflections by introducing reflectors that are attached to the microphone array. Because the reflectors are located very close to the microphone array, the amplitude level of the reflections will be large; thus the change in a room's acoustic characteristics, which affects the room transfer function between the source and microphone array, should be relatively small and can be ignored. In other words, a room transfer function generated by simulation or measurement in an anechoic chamber can be used as prior information to design a beamforming filter. In this paper, we discuss how the reflectors affect the modeling of room transfer functions. Experimental results from a real acoustic environment proved that the proposed telescopic microphone array segregates sound sources located in the same direction but at different distances.

This paper is organized as follows. The basic framework of the microphone array input and brief explanation about the conventional MVDR method are described in Section 2. In Section 3, a spatial sensitivity control method that uses a reflector is proposed. Experimental results and discussions are discussed in Section 4, followed by concluding remarks in Section 5.

2. CONVENTIONAL METHOD

2.1. Problem formulation

Let us consider that M (>1) microphones receive a target and $K-1$ noise sources, as shown in Fig. 1. Our goal is to segregate the target source positioned at an arbitrary position without distortion, even if K is a large number. The room transfer functions from the target and the k-th noise source to the m-th microphone are denoted as $a_{\mathrm{S},m}(t)$ and $a_{\mathrm{N}_k,m}(t)$, respectively, where t denotes the time sample index. When the target source signal and the k-th noise are respectively denoted as $s(t)$ and $n_k(t)$, the observed signal received by the m-th microphone is given by

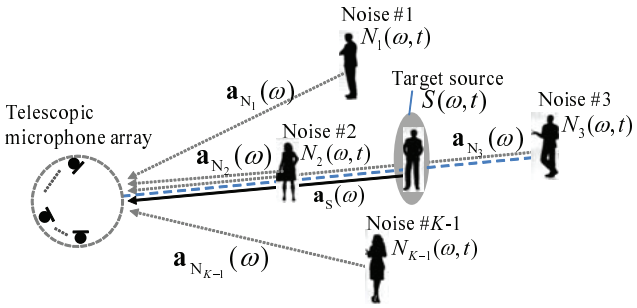

Fig. 1. Definition of symbols

$$x_m(t) = \sum_l a_{S,m}(l)s(t-l) + \sum_{k=1}^{K-1}\sum_l a_{N_k,m}(l)n_k(t-l). \quad (1)$$

By convolving filters $w_m(l)$ of length L, the output signal $y(t)$ is given by

$$y(t) = \sum_{m=1}^{M}\sum_{l=0}^{L-1} w_m(l)x_m(t-l). \quad (2)$$

Our goal is to retrieve $s(t)$ by allowing delay to maintain causality, as expressed in Eq. (3), even if noise sources are positioned in the same direction of the target source.

$$y(t) = s(t - \frac{L}{2}). \quad (3)$$

2.2. Microphone array input model in frequency-domain

To conduct convolution by multiplication, Eq. (1) is reformulated in the frequency domain given by

$$\mathbf{x}(\omega,\tau) = \mathbf{a}_S(\omega)S(\omega,\tau) + \sum_{k=1}^{K-1}\mathbf{a}_{N_k}(\omega)N_k(\omega,\tau), \quad (4)$$

where $\mathbf{x}(\omega,\tau) = [X_1(\omega,\tau),\ldots,X_M(\omega,\tau)]^T$ is the vector of microphone observation and $S(\omega,\tau)$, $N_k(\omega,\tau)$, and $X_m(\omega,\tau)$ are the short-time Fourier transform (STFT) of $s(t)$, $n_k(t)$, and $x_m(t)$, respectively. Note that τ and ^{T} denotes the frame-time and the transposition, respectively. The vectors

$$\mathbf{a}_S(\omega) = [A_{S,1}(\omega),\ldots,A_{S,M}(\omega)]^T,$$
$$\mathbf{a}_{N_k}(\omega) = [A_{N_k,1}(\omega),\ldots,A_{N_k,M}(\omega)]^T,$$

consist of the room transfer functions from the target and the k-th noise sources to the m-th microphone denoted as $A_{S,m}(\omega)$ and $A_{N_k,m}(\omega)$, respectively. Thus, the output signal $Y(\omega,\tau)$ is calculated by

$$Y(\omega,\tau) = \mathbf{w}^H(\omega)\mathbf{x}(\omega,\tau), \quad (5)$$

where $\mathbf{w}(\omega) = [W_1(\omega),\ldots,W_M(\omega)]^T$ and ^{H} denote the filter in frequency domain and the Hermitian conjugate, respectively. The time domain output signal in Eq. (3) is obtained since $y(t)$ is the inverse short-time Fourier transform (ISTFT) of $Y(\omega,\tau)$.

Hereafter, $S(\omega,\tau)$ and $N_k(\omega,\tau)$ are assumed to be uncorrelated with each other; therefore, Eq. (6) is expressed as

$$\mathrm{E}\{\mathbf{s}(\omega,\tau)\mathbf{s}(\omega,\tau)^H\} = \boldsymbol{I}_K, \quad (6)$$
$$\mathbf{s}(\omega,\tau) = [S(\omega,\tau), N_1(\omega,\tau), ..., N_{K-1}(\omega,\tau)]^T,$$

where $\mathrm{E}\{\cdot\}$ is the expectation operator.

2.3. Filter design of MVDR method

In the MVDR method [6], the filter $\mathbf{w}(\omega)$ is optimized by minimizing the output noise power $P_{\text{out}}(\omega)$ while the response gains to U different positions are constrained,

$$\mathbf{w}_{\text{opt}}(\omega) = \arg\min_{\mathbf{w}}\{P_{\text{out}}(\omega)\}, \quad (7)$$
$$\text{subject to} \quad \mathbf{w}^H(\omega)\mathbf{c}(\omega) = \mathbf{g}(\omega), \quad (8)$$

where

$$P_{\text{out}}(\omega) = \mathrm{E}\{|Y(\omega,\tau)|^2\} = \mathbf{w}^H(\omega)\mathbf{R}(\omega)\mathbf{w}(\omega),$$
$$\mathbf{R}(\omega) = \mathrm{E}\{\mathbf{x}(\omega,t)\mathbf{x}^H(\omega,t)\},$$
$$\mathbf{C}(\omega) = [\mathbf{h}_S(\omega), \mathbf{h}_{N_{\sigma(1)}}(\omega),\ldots,\mathbf{h}_{N_{\sigma(U-1)}}(\omega)],$$
$$\mathbf{g}(\omega) = [1,0,\ldots,0]^T,$$
$$\mathbf{h}_S(\omega) = [H_{S,1}(\omega),\ldots,H_{S,M}(\omega)]^T,$$
$$\mathbf{h}_{N_{\sigma(u)}}(\omega) = [H_{N_{\sigma(u)},1}(\omega),\ldots,H_{N_{\sigma(u)},M}(\omega)]^T.$$

Here, $\mathbf{R}(\omega)$ and $\sigma(u)$ denote the spatial correlation matrix [3] and the index of the noise source to be drastically suppressed, respectively, and $\mathbf{h}_S(\omega)$ and $\mathbf{h}_{N_{\sigma(u)}}(\omega)$ are the propagation vectors of the target and the $\sigma(u)$-th noise source, respectively. The optimum filter $\mathbf{w}_{\text{opt}}(\omega)$ is calculated by

$$\mathbf{w}_{\text{opt}}(\omega) = \mathbf{R}^{-1}(\omega)\mathbf{C}(\omega)\Big(\mathbf{C}^H(\omega)\mathbf{R}^{-1}(\omega)\mathbf{C}(\omega)\Big)^{-1}\mathbf{g}(\omega) \quad (9)$$

By using the assumption of Eq. (6), we note that $\mathbf{R}(\omega)$ is calculated from the propagation vectors of virtual sources defined by

$$\mathbf{R}(\omega) = \frac{1}{K}\left[\mathbf{h}_S(\omega)\mathbf{h}_S^H(\omega) + \sum_{k=1}^{K-1}\mathbf{h}_{N_k}(\omega)\mathbf{h}_{N_k}^H(\omega)\right]. \quad (10)$$

We specifically discuss the MVDR method that uses $\mathbf{R}(\omega)$ calculated by Eq. (10).

In the original MVDR method [6], the components in the propagation vectors $\mathbf{h}_S(\omega)$ and $\mathbf{h}_{N_{\sigma(u)}}(\omega)$ are substituted by the steering vectors [3] given by

$$H_{S,m}(\omega) = \exp\Big(-j\omega\frac{\mathbf{q}_S^T\mathbf{p}_m^{(0)}}{\nu}\Big), \quad (11)$$
$$H_{N_{\sigma(u)},m}(\omega) = \exp\Big(-j\omega\frac{\mathbf{q}_{N_{\sigma(u)}}^T\mathbf{p}_m^{(0)}}{\nu}\Big), \quad (12)$$

since the sound sources are assumed to arrive as plane waves. Here $\mathbf{p}_m^{(0)}$, $\mathbf{q}_S$, $\mathbf{q}_{N_{\sigma(u)}}$, and ν are the position vectors of the m-th microphone, that of the target, that of the $\sigma(u)$-th noise source, and sound velocity, respectively. Because $\mathbf{h}_S(\omega)$ and $\mathbf{h}_{N_{\sigma(u)}}(\omega)$ include only the directional but not the distance information of the source positions, the noise sources located in the same direction as that of the target source cannot be suppressed. On the other hand, applying room transfer functions to the components of the propagation vectors would enable us to distinguish the sources aligned in the same direction.

$$H_{S,m}(\omega) = A_{S,m}(\omega), \quad (13)$$
$$H_{N_{\sigma(u)},m}(\omega) = A_{N_{\sigma(u)},m}(\omega). \quad (14)$$

However, the microphone array would be very sensitive to changes in room transfer function, i.e., the room's acoustic characteristics.

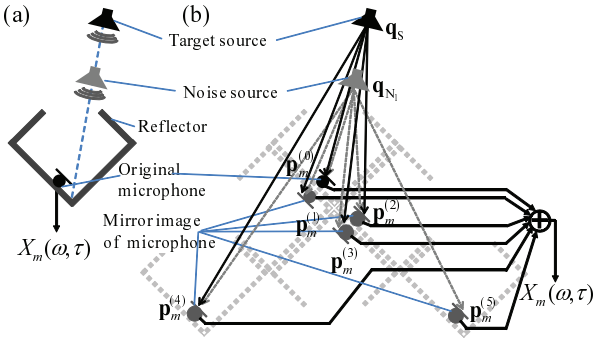

Fig. 2. Generation of reflected sounds by using positioning reflector

3. PROPOSED METHOD

The basic idea that underlies our approach is that we modify the modeling of the propagation vector $\mathbf{h}_S(\omega)$ and $\mathbf{h}_{N_{\sigma(u)}}(\omega)$ to include distance information of sound sources while being independent from the room's acoustic characteristics. To accomplish this, we introduce a reflector attached to the microphone array that produces strong reflections included in the array input signal. Normally a reflection is produced by any solid object that reflects sound waves. In an indoor environment, such objects could be the walls, ceiling or floor of the room or any other large flat items. When we have two sound sources aligned in the same direction from a microphone as described in Fig. 2(a), the propagation paths of reflections vary depending on the source position whereas the direct sounds propagate the same path. This implies that the active observation of reflections enables us to distinguish the source distance although the sources are located in the same direction. The most straightforward approach for exploiting the reflections would be using reflections generated by such objects, e.g., walls. However, this will exhibit the same problem as we saw in the application of a room transfer function to the propagation vector because the paths of reflections vary depending on the room's acoustic characteristics.

Assume that a large reflector is attached to the microphone array, as described in Fig. 2(a) which generates strong reflections. Because the length of the propagation path is much shorter, the reflections generated by the reflector will be dominant in the observed signal compared to the reflections generated by the walls; thus, the reflector will enable us to ignore the effect of reflections generated by other items.

Assuming that reflections are modeled using the image method [12], each reflection is regarded as a signal generated from an image source. Analogously, we could also say that the input signal consists of a source received by real and mirror image microphones as described in Fig. 2 (b). In other words, we can define the propagation vectors as composed of the direct sound and reflections up to the D-th order, as expressed in Eqs.(15) and (16),

$$H_{S,m}(\omega)=\sum_{d=0}^{D}\frac{\kappa^{(d)}(\omega)}{||\mathbf{p}_m^{(d)}-\mathbf{q}_S||}\exp\left(-j\omega\frac{||\mathbf{p}_m^{(d)}-\mathbf{q}_S||}{c}\right), \quad (15)$$

$$H_{N_{\sigma(u)},m}(\omega)=\sum_{d=0}^{D}\frac{\kappa^{(d)}(\omega)}{||\mathbf{p}_m^{(d)}-\mathbf{q}_{N_{\sigma(u)}}||}\exp\left(-j\omega\frac{||\mathbf{p}_m^{(d)}-\mathbf{q}_{N_{\sigma(u)}}||}{c}\right), \quad (16)$$

where $\mathbf{p}_m^{(d)}$ and $\kappa^{(d)}(\omega)$ denote the coordinates of a mirror microphone of the d-th image source and reflection coefficient of the d-th image source, respectively. Note that the 0-th reflection corresponds to the direct sound; thus $\kappa^{(0)}(\omega)=1$. By substituting $H_{S,m}(\omega)$ and $H_{N_{\sigma(u)},m}(\omega)$ into $\mathbf{h}_S(\omega)$ and $\mathbf{h}_{N_{\sigma(u)}}(\omega)$ by Eqs. (15) and (16), solving Eq. (9) will derive a filter $\mathbf{w}(\omega)$ that distinguishes source distances and is independent from the room's acoustic characteristics. It is preferable to design a reflector whose shape generates higher order reflections. Heuristically, a rectangular shape performs well, as described in the following section; however, this solution should be studied for future work.

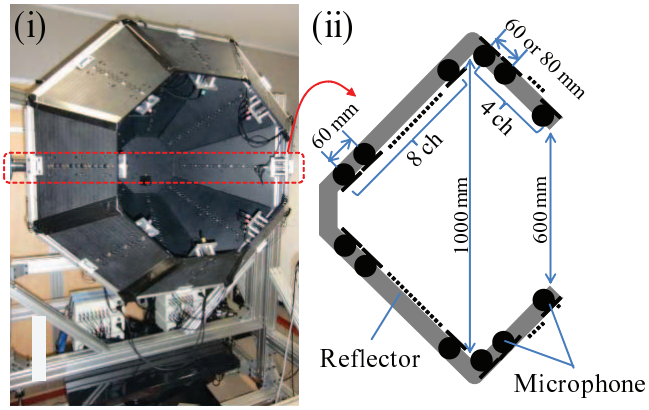

Fig. 3. (i) Telescopic microphone array using large reflector, (ii) is cross-section of (i)

Table 1. Experimental parameters

Sampling frequency	48 kHz
Filter length	2048 taps
Number of microphones, M	96 ch
Number of measurement positions, (= Number of sound sources, K)	238 (= 17 directions × 14 distances)
Angle interval for measurement	2.0°
Distance interval for measurement	0.5 m
Number of gain constraints, U	2
Reverberation time of room, T_{60}	180 msec
Room size	11.8 m (W) × 6.3 m (D) × 3.0 m (H)

4. EXPERIMENTS

4.1. Experimental conditions

Experiments were conducted to evaluate the effectiveness of the proposed method. Figure 3 (i) shows the telescopic microphone array with a large reflector, to which 96 omni-directional microphones are attached. Figure 3 (ii) shows a cross-sectional schematic in the horizontal plane of the telescopic microphone array. The shape of the reflector is based on a chamfered rectangle. The reflector was made from several plates of acrylonitrile butadiene styrene (ABS) with a total thickness of 10 mm. The length of the largest part of the reflector is 1.0 m. To reduce the level of observed noise arriving from the back of the microphone array, all microphones were built inside the reflector's enclosure.

For calculating $\mathbf{w}(\omega)$, the transfer functions from multiple virtual source positions to the telescopic microphone array were measured in a real acoustic environment whose reverberation time (T_{60}) was 180 msec. 238 different virtual source positions, i.e., combinations of 17 directions and 14 distances (Fig. 4), were selected. Other experimental parameters are listed in Table 1.

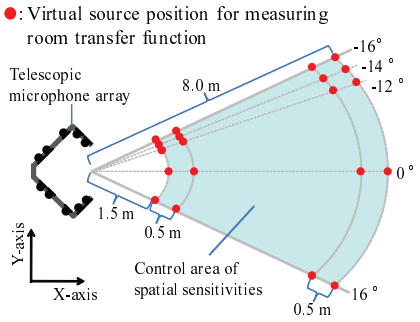

Fig. 4. Measurement points of transfer functions

4.2. Experimental results

Figure 5 shows the spatial sensitivities of different frequencies when the target source was positioned 5.5 m from the center of the array in the direction of 0°. If the spatial sensitivity is low at every position except for the target position, the telescopic microphone array effectively segregates the target source from noise sources. From the results, the spatial sensitivity was small even in the same direction of the target source as long as the distance from the microphone array was different. On the other hand, Fig. 6 shows the spatial sensitivities when the direction of the target source was changed to 10°. As in the case where the target source was positioned in 0°, a distinctively high spatial sensitivity could be seen at the position of the target source. Finally, Fig. 7 shows the spatial sensitivities when the distance between the microphone array and target source was changed to 7.5 m. Again, the proposed method effectively emphasized the target source even though the target was located at a relatively distant position. From these experimental results, it is confirmed that the proposed telescopic microphone array distinctively segregates a target source located at a particular position.

5. CONCLUSION

We proposed a spatial sensitivity control method for our telescopic microphone array, which uses a reflector to segregate a target source from noise sources. By placing the reflector close to the microphone array, we artificially generate reflections to segregate sound sources located in the same direction but at different distances from the microphone array. By taking into consideration the acoustical properties of the reflector, the filters for forming an optimal spatial beam were calculated. Experiments using the telescopic microphone array composed of 96 microphones and a large reflector were conducted in real acoustical environment. After investigating the spatial sensitivities of the proposed microphone array, it was confirmed that the target source could be segregated even if there were noise sources in the same direction of the target.

There are many other issues that need further study, e.g., investigating the performance of the proposed microphone array in more reverberant environments, and reducing the number of transfer functions to be measured preliminarily. Furthermore, we believe more theoretical design of the reflector is also needed.

6. REFERENCES

[1] T. Fujii and M. Tanimoto, "Free-viewpoint TV system based on the ray-space representation," *SPIE ITCom*, vol. 4864-22, pp. 175–189, 2002.

[2] K. Niwa, T. Nishino, and K. Takeda, "Encoding large array signals into a 3D sound field representation for selective listening point audio based on blind source separation," *ICASSP 2008*, pp. 181–184, 2008.

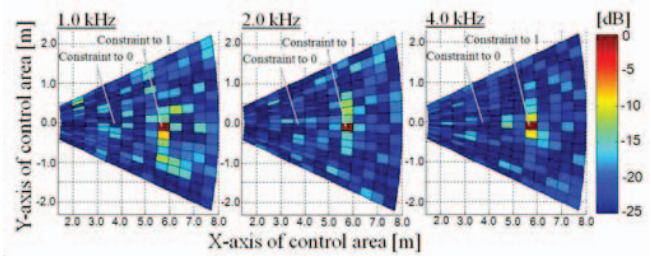

Fig. 5. Spatial sensitivities when target source is positioned in 0° direction and 5.5 m from center of microphone array

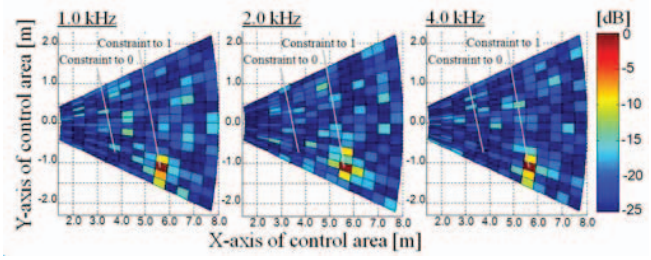

Fig. 6. Spatial sensitivities when target source is positioned in 10° direction and 5.5 m from center of microphone array

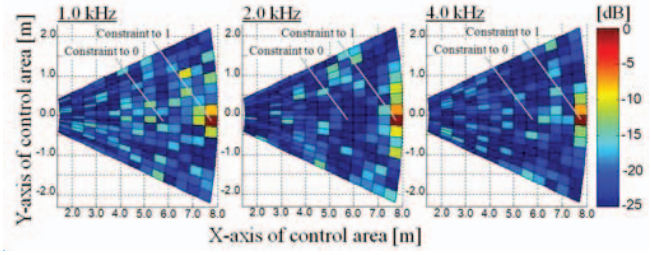

Fig. 7. Spatial sensitivities when target source is positioned in 0° direction and 7.5 m from center of microphone array

[3] D. H. Johnson and D. E. Dudgeon, *Array processing: concepts and techniques*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[4] J. L. Flanagan and D. A. Berkley, G. W. Elko, J. E. West, M. M. Sondhi, "Autodirective microphone systems," *Acoustica*, vol. 73, no. 2, pp. 58–71, 1991.

[5] S. M. Kuo, D. R. Morgan, "Active noise control: a tutorial review," *Proc. of IEEE*, vol. 87 (6), pp. 943–973, 1999.

[6] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60 (8), pp. 926–935, 1972.

[7] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, 1969.

[8] R Mukai, H Sawada, S Araki, S Makino, "Frequency-domain blind source separation of many speech signals using near-field and far-field models," *EURASIP J. Appl. Signal Proc.*, p. 13, Article ID 83683, 2006.

[9] 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," *ICASSP 2010*, pp. 149–152, 2010.

[10] 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 Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1793–1805, Sep. 2010.

[11] 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.

[12] J. B. Allen and D. A. Berkley., "Image method for efficiently simulating small-room acoustics," *J.Acoust.Soc.Am.*, vol. 65, no. 4, pp. 943–950, 1979.