2D POSITION IDENTIFICATION OF A SOUND SOURCE USING PHASE DIFFERENCE SPECTRUM IMAGES

Ryuichi Shimoyama

Nihon University, College of Industrial Technology
1-2-1 Izumicho, Narashino,Chiba, 275-8575,Japan

Shimoyama.ryuichi@nihon-u.ac.jp

Abstract

The visually handicapped person can localize the position of the sound source only by hearing. The auditory space might be constructed on the auditory system in the brain. Such an auditory mechanism remains unknown. A human-like robot that can construct an auditory space with two microphones might reveal an alternative solution to clarify such a mechanism. The present paper describes a new computational algorithm for constructing a 2D auditory space with two microphones. The cross-power spectrum phase of the sound pressure transmitted from a broadband sound source in a reverberation room is discontinuous with respect to the frequency. The discontinuous phase differences spectra (images) are used to identify both the distance to the sound source and the azimuth of the sound source. A head and torso simulator with a built-in microphone in each ear was fixed at the center of the reverberation room. The 2D positions were identified for the source placed at various positions. The results showed that the source positions could be identified over a range of 3 meters and an approximate range of azimuths of from -50° to +50°.

Keywords: 2D, Position identification, Sound source, Phase difference, Images, Head and torso.

Presenting Author's biography

Ryuichi Shimoyama. He received Doctor of Engineering in System Science at Okayama University. He was an invited researcher at Center for Intelligence Systems of Vanderbilt University, USA and Polo Sant'Anna Valdera of Pisa University, Italy, from 2005 to 2006. He is Professor of Nihon University, Japan. His current research interest includes artificial intelligence on auditory system and robotics. He is a member of IEEE, IEICE, IPSJ, ASJ and JSST.

1   Introduction

The visually handicapped person can localize the position of a sound source only by hearing. The auditory space might be constructed on the auditory system in the brain. Such an auditory mechanism remains unknown. The 3D auditory space consists of the horizontal azimuth, the vertical azimuth of the sound source, and the distance to the source. The vertical azimuth of the sound source is beyond the scope of the present paper.

As for the horizontal azimuth, several algorithms for sound source localization with two microphones using the interaural time difference (ITD) have been proposed [1,2]. A sound source can be localized with an average accuracy of ±1.5° for robotic acoustic tracking [3]. The present author proposed a computational algorithm for sound source localization using the frequency spectra of interaural phase difference (IPD) [4,5,6], because ITD sometimes provides an incorrect azimuth under the reverberative conditions. Nakashima *et al.* [7,8] proposed binaural models based on the ITD and the interaural level difference (ILD). However, the ILD is not useful for the far field under the reverberative condition due to its low signal level.

Human characteristics of auditory distance perception were examined through psychophysical experiments. Gardner discussed human distance perception in an anechoic space [9]. Bronkhorst *et al.* reported that the energy ratio of the direct and reflected sound is important with respect to distance estimation in human hearing [10]. The results of Mershon and King indicated that the state of reverberation can serve as absolute cues for auditory distance perception [11]. Takahashi *et al.* investigated the dependence of the phase response of a transfer function on the distance to the sound source in the coherent field [12] by measuring sound pressure with a single microphone. The phase of sound pressure inevitably depends on the type of sound.

The present paper describes a new algorithm for computationally identifying both the distance and the azimuth, which may be used to construct the 2D auditory space. The frequency spectra of interaural phase difference (IPD) were used to achieve independence from the type of sound and for adaptability to the far field in the reverberative condition. The cross-power spectrum phase of the sound pressure was measured to identify both the distance to the sound source and the azimuth of the sound source. A head and torso simulator equipped with a microphone in each ear was fixed at the center of the reverberation room. The 2D position of the sound source was identified for various source positions.
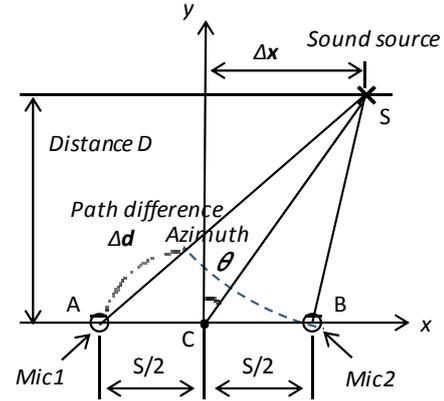

Fig. 1   Geometry of the sound source and the sensors (plane figure)

2   Method and procedures

2.1   Method

Generally, the cross-power spectrum phase of the sound pressure, which is measured using a pair of microphones, is discontinuous over the range of measured frequency when broadband noise is transmitted from a loudspeaker in a reverberation room. The microphone detects the sound, synthesizing directly propagated sound and the sound reflected from the floor, walls, and ceiling. The reflected sound might affect this discontinuity. Such discontinuity in the frequency spectrum depends on the distance between the sound source and the microphones. In the present study, the discontinuous phase difference spectrum (images) is used to identify both the azimuth of the sound source and the distance to the source.

The CAVSPAC method [6] is applied to identify the azimuth θ of the source. This method is briefly described as follows. A simple mathematical model is considered, as shown in Fig. 1. The sound propagates directly from the sound source to two sensors without reflection or diffraction. The sound source is set at a distant D in front of a pair of sensors, Mic1 and Mic2, separated from each other by a distance *S*. The phase difference $\Delta\varphi$ of the sound pressure between the two sensors is expressed as follows:

$$\Delta\varphi = \frac{360 f \Delta d}{c} \quad \text{(degrees)} \qquad (1)$$

where Δd is the path difference between paths AS and BS, *f* is the frequency, and *c* is the speed of sound. The phase is defined ambiguously, as follows:

$$\Delta\varphi \rightarrow \Delta\varphi + 360n \quad \text{(degrees)} \qquad (2)$$

where *n* is an integer. The value of $\Delta\varphi$ equals to that of $\Delta\varphi + 360n$. The path difference is then defined

as follows:

$$\Delta d = \frac{c(\Delta\varphi + 360n)}{360f} \quad \text{(degrees)}. \qquad (3)$$

The azimuth of the sound source is calculated based on the plane geometry by solving Eqs. (3), (4), and (5) numerically:

$$\Delta d = \sqrt{D^2 - (S/2 + \Delta x)^2} - \sqrt{D^2 - (S/2 - \Delta x)^2} \quad \text{(m)}, \qquad (4)$$

$$\theta = \frac{360}{2\pi}\arctan\left(\frac{\Delta x}{D}\right) \quad \text{(degrees)}. \qquad (5)$$

The azimuth θ depends only slightly on the distance to the sound source D under the condition in which D/S is greater than 3. An arbitrary value of the distance D, which is three times greater than S, should be substituted in Eqs. (4) and (5) in order to find the azimuth. Multiple azimuths of the sound source are then defined at higher frequencies where the wavelength is shorter than the distance between sensors, corresponding to integer n. The frequency convergence of the phase difference is treated as additional information for finding the true azimuth, considering the broadband sound signal.
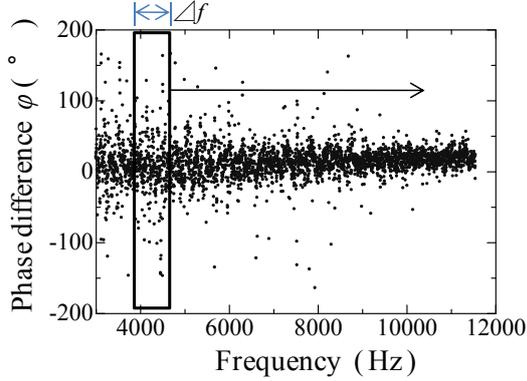

Fig. 2　Method of calculating the frequency spectrum of the standard deviation from the phase difference of the sound pressure.
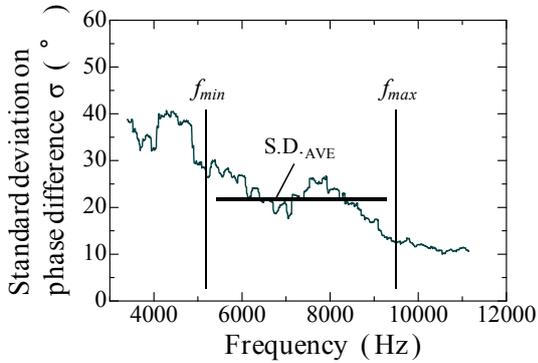

Fig. 3　Frequency spectrum of the standard deviation and its average value from f_{min} to f_{max}.

Next, a new method is proposed for determining the distance D to the source. Figure 2 shows a sample of the frequency spectrum of the phase difference of the measured sound pressure. The discontinuity of the phase difference spectrum is quantified as follows using the standard deviation. If 2m+1 data of the discrete phase difference are included within the frequency width Δf, where m is an integer, the standard deviation σ_i at the center frequency f_i is given by the following equations:

$$\sigma_i = \sqrt{\frac{1}{2m+1}\sum_{j=i-m}^{i+m}(\varphi_j - \overline{\varphi_i})}, \qquad (6)$$

$$\overline{\varphi_i} = \frac{\sum_{j=i-m}^{i+m}\varphi_j}{2m+1}, \qquad (7)$$

where φ_j is the phase difference value at frequency f_j, and $\overline{\varphi_i}$ is the average value at frequency f_i. The frequency spectrum of the standard deviation is obtained when the center frequency f_i is swept over the measured frequency range. The frequency spectrum of the standard deviation is shown in Fig. 3. The average value S.D.$_{AVE}$ over the range of frequency from f_{min} to f_{max} depends on the distance to the source. However, S.D.$_{AVE}$ also depends on the azimuth of the sound source. Therefore, the relationship between S.D.$_{AVE}$ and the distance and S.D.$_{AVE}$ and the azimuth are measured experimentally and are stored in the database in advance. The distance is identified by referring to this database from the identified azimuth.

2.2　Procedures

The proposed algorithm is shown in Fig. 4. A pair of acoustical signals are measured simultaneously. The cross-power spectrum phase is calculated after the DFT procedure. Multiple azimuths are obtained for a target source at relatively high frequencies, because the phase difference spectrum leads to multiple azimuths due to the phase ambiguity. The frequency-independent azimuth is extracted from these multiple azimuths using the azimuth convergence over the frequency. The average standard deviation of the distance is calculated from the same phase difference spectrum. The increase in S.D.$_{AVE}$ is corrected using the identified azimuth value. Finally, the distance D is identified by interpolating the data in the database.

3　Applications

Figure 5 shows the configuration of the head and torso simulator with two built-in microphones and a loudspeaker. The measurement and the data processing of the acoustical signals were conducted using a computer (Dell, GX280) equipped with a
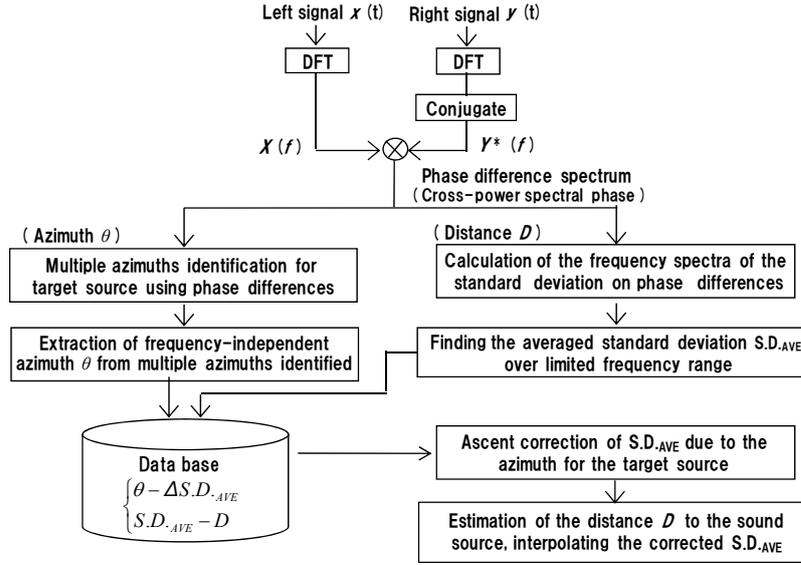
Fig. 4 Algorithm for 2D position identification of a sound source.

24-bit A/D converter (N.I., PCI-4474). The head and torso simulator (Bruel & Kjaer, type 4128-C) was used for sound detection, as shown in Fig. 6. The distance between the two microphones was approximately 13 cm. The head and torso simulator was placed on a motor-controlled turntable. The sound source was a golf-ball-sized loudspeaker that had relatively broad directivity and a narrow source area. All measurements were performed at the center of the room. Broadband noise was transmitted from the loudspeaker. The parameters were set as $\Delta f = 800$ Hz, m = 160, f_{min} = 4 kHz, and f_{max} = 11 kHz.

3.1 Azimuth identification for the source

A loudspeaker was placed at a distance of 3 m (D in Fig. 1) from the head and torso at a height of 1.2 m. Figures 7, 8, and 9 show the experimental results for the case in which the azimuth was set to 0°. The phase difference spectrum of the sound pressure in Fig. 7 was discontinuous over the measured frequency. This discontinuous distribution (image) depends not only on the azimuth of the source but also on the distance to the source. The frequency spectra of the multiple azimuths for the source are shown in Fig. 8, which were calculated using Eqs. (3), (4), and (5). The true azimuth was obtained at approximately zero degrees at the peak in Fig. 9, at which the source azimuth converged over the range of frequency. The identified azimuths also approximately matched the true azimuths for distances of D = 1 and 2 m. The sound pressure was then measured when the azimuth θ of the head and torso simulator in Fig. 5 (D = 1 m) was changed by rotating the turntable. The relationship between the true source azimuth θ and the identified source azimuth is shown in Fig. 10 for different
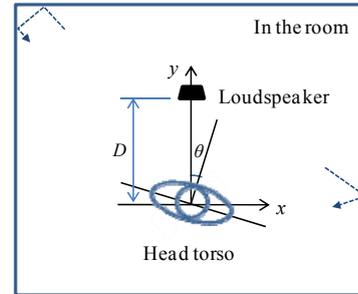

Fig. 5 Geometry of the head and torso simulator and the loudspeaker (plane figure).

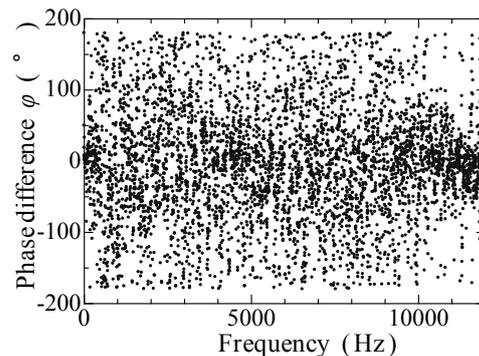Fig. 6 Head and torso simulator with built-in microphones.

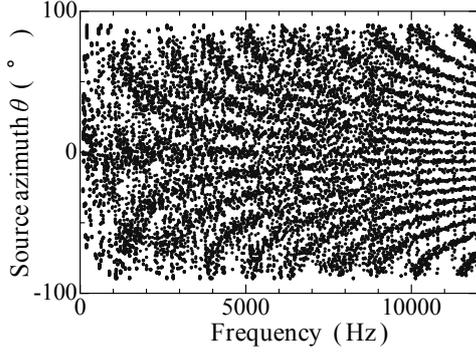Fig. 7 Phase difference spectrum (D = 3 m, θ = 0°).

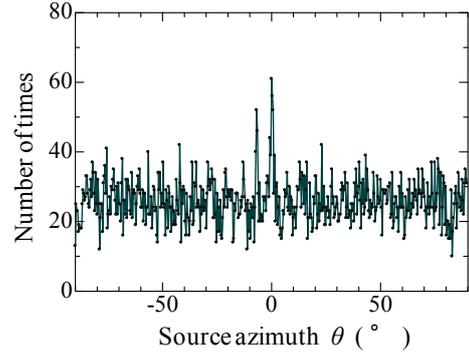Fig. 8 Frequency spectra of multiple identified azimuths $(D = 3$ m, $\theta = 0°)$.

Fig. 9 Source azimuth convergence over frequency $(D = 3$ m, $\theta = 0°)$.
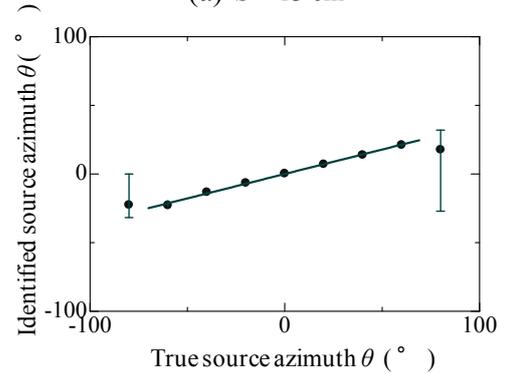
microphone intervals. The identified azimuths were proportional to the true azimuths from -50° to +50°, in Fig. 10(a). For the azimuths exceeding 50°, the azimuths were obtained with a large dispersion. This limitation in the azimuth detection restricts the area for source localization. Figure 10(b) shows the case in which the microphone interval S was set to 50 cm, rather than 13 cm (see Fig. 1). This means that a different value of S is substituted into Eq. (4) for azimuth identification, while the microphone interval is physically maintained at 13 cm. The dispersion decreased over a wider range of azimuth with a longer interval. The interval S and the speed of sound c in Eqs. (3) and (4) are physical constants. The values of these parameters cannot be known in advance for the human auditory system. Nakashima *et al.* suggested that the physical constants could be estimated using a machine learning algorithm, such as a neural network [8]. The values of variables c and S should be adjusted to find the precise source azimuth θ using a machine learning algorithm.

3.2 Distance identification to the source

The loudspeaker was placed in front of the head and torso simulator. When the distance was changed from 1 to 3 m, the relationship between the distance and the average standard deviation value was approximately linear, as shown in Fig. 11. The average value increased with distance. For a source placed directly in front of the head and torso, the distance can be calculated from the obtained average standard deviation value by interpolating the relation in Fig. 11. Figure 12 shows the ascent of the standard deviation value corresponding to the azimuth for the source. The average value S.D.$_{AVE}$ increased nonlinearly with increasing azimuth. This nonlinear relation varied greatly with minute variations in distance. The average was used for distance identification as the real line indicated in Fig.12. If the azimuth of the source is identified, then the average standard deviation value could be corrected using the ascent value interpolated in Fig. 12. Then, the distance can be identified from the corrected average value in Fig. 11.
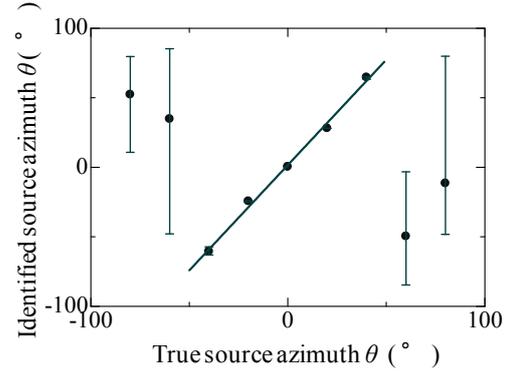
(a) $S = 13$ cm

(b) $S = 50$ cm

Fig. 10 The identified azimuths corresponding to the true when Head and torso was rotated from -80°to +80° with different values of microphone interval S ($D = 1$ m).

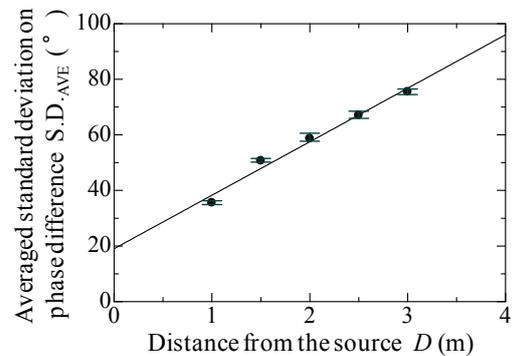
Fig. 11 S.D.$_{AVE}$ as a function of the distance between the loudspeaker and Head and torso simulator $(\theta = 0°)$.

3.3 2D position identification of the source

The 2D positions of the loudspeaker were identified using the fixed head and torso with the loudspeaker placed at various positions. Fig. 13 shows a comparison the true position with the identified position of the loudspeaker. The 2D positions of the loudspeaker could be approximately traced. The four positions at a constant angle of +30° were identified with an error of 10 cm. The maximum error was approximately 45 cm at three positions at a constant distance of 2 m. The cross-frontal source position could be identified with a significant error in distance, because the distance identification process was complicated due to the dependence on the azimuth.

4 Conclusion

The 2D positions of the sound source were identified using the fixed head and torso when the loudspeaker was placed at various positions. Broadband noise was transmitted from the loudspeaker in the reverberation room. The discontinuous phase difference spectra (images) on the sound pressure were used to identify the azimuth of the sound source and the distance to the source. As a result, (1) the 2D positions of the loudspeaker were approximately traced, and (2) the source positions could be identified over a range of 3 meters and an approximate range of azimuths of from -50° to +50°.

As reported by Bronkhorst, Merson, and King, the rate of reverberation might serve as an absolute cue even for computational distance identification. This means that the identified distance inevitably changes depending of the reverberative condition of the room. Real-time relative distance detection will be investigated in the future in an attempt to detect movement by sound alone, such as approaching or going away.

ACKNOWLEDGEMENT

5 References

[1] R. M. Stern and H. S. Colburn. Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective internal position. J. Acoust. Soc. Am., 64 : 127-140, 1978.

[2] W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. J. Acoust. Soc. Am., 80 : 1608-1621, 1986.

[3] J. Murray, H. Erwin, S. Wermter. Robotic sound-source localization and tracking using interaural time difference and cross-correlation. Center for hybrid intelligence systems, University of Sunderland, Sunderland .
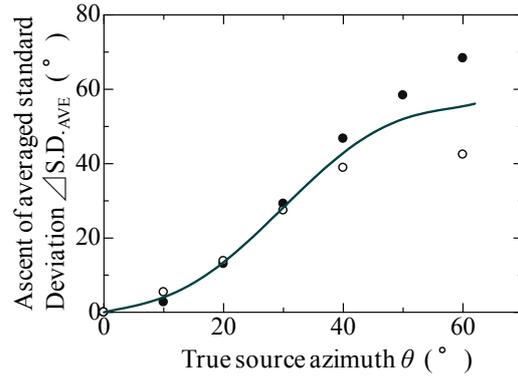

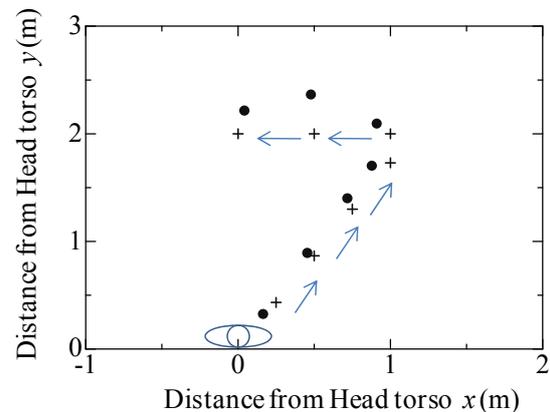Fig. 12   S.D.$_{AVE}$ as a function of the source azimuth ($D = 1$ m ●, $D = 2$ m ○).

Fig. 13 2D positions of the loudspeaker and the identified (True positions: ●, Identified: ○).

[4] R. Shimoyama and K. Yamazaki. Acoustic source localization using phase difference spectrum images. Acoust. Sci. & Tech., 24:161-171, 2003 .

[5] R. Shimoyama and K. Yamazaki. Multiple acoustic source localization using ambiguous phase differences under reverberative conditions. Acoust. Sci. & Tech., 25:446-456, 2004 .

[6] R. Shimoyama and K. Yamazaki. Computasional acoustic vision by solving phase ambiguity confusion. Acoustical science and technology. 30 (3): 199-208, 2009

[7] H. Nakashima, Y. Chisaki, T. Usagawa and M. Ebata, Frequency domain binaural model based on interaural phase and level differences. Acoustical science and technology. 24 (4): 172-178, 2003.

[8] H. Nakashima, M. Kawamoto, M. Ito and T. Mukai, A two-dimensional position estimation of two sound sources using two microphones with reflectors. Journal of the society of instrument and control engineering, 664-670, 2009(in Japanese).

[9] M.B.Gardner, Distance estimation of 0°or apparent 0°-oriented speech signals in anechoic space. The journal of the Acoustical Society of America, 48 (45)1,47-53,1969.

[10] A. W. Bronkhorst and T. Houtgast. Auditory distance perception in rooms. Nature, 397(6719),

517-520, 1999.

[11] D.H.Mershon and L.E.King, Intensity and reverberation as factors in the auditory perception of egocentric distance, Perception & Psychophysics, 18(6), 409-415, 1975.

[12] Y. Takahashi, M. Tohyama and Y. Yamazaki. Phase response of transfer functions and coherent field in a reverberant room. IEICE, J89-A (4): 291-297, 2006 (in Japanese).