ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 3, September 2013

Bio-inspired Sound Source Localization for Compensating Sound Diffraction Using Binaural Human Head and Torso

Ryuichi Shimoyama
Shimoyama.ryuichi@nihon-u.ac.jp, College of Industrial Technology, Nihon University, Japan

***Abstract*—Both humans and animals have two ears for localizing the sound source.The outer ears of most mammals are relatively large, transformable, and complicated in shape. In humans, the ears are located on opposite sides of the head. Although sound is diffracted by the outer ear, head, and body, sound diffraction has not been considered in previous biological neural approaches to sound source localization. In the present paper, the effect of sound diffraction on sound source localization is evaluated and a new method for compensating for sound diffraction in sound source localization is proposed. A similar model to an owl's auditory system is applied to a binaural human head and torso.**

***Index Terms*—sound source localization, sound diffraction, compensation, owl, head and torso.**

I. INTRODUCTION

Both humans and animals have two ears for localizing the sound source. Neural mechanism in binaural auditory systems has been investigated psychophysically and biologically. In humans, most researchers agree that human utilizes two acoustical cues for localizing the horizontal azimuth of the sound source by the Duplex Theory: interaural time differences (ITDs) at low frequencies and interaural intensity differences (IIDs) at high frequencies [1][2]. The Duplex Theory was defined for pure-tone stimuli through psychophysical measurements. Jeffress proposed that ITDs provide the primary cue for sound source localization at low frequencies [3]. The mid-brain includes many coincident detectors, which fire when nerve spikes transmitted along nerve fibers that are acting as delay lines from the cochleae arrive simultaneously. This model explains how ITDs are transformed into the distribution of fired detectors as encoding of ITDs. For both birds and mammals, the neural system in the mid-brain has been investigated biologically with respect to sound source localization [4]-[14]. Konishi et al. reported that ITDs were also a primary cue even at high frequencies in the mid-brain of barn owls [4] [5]. At higher frequencies, ITDs may be encoded by the firing of multiple detectors due to phase ambiguity. Konishi et al. showed how phase ambiguity confusion was resolved in the owl auditory system. Shimoyama proposed a frequency domain model (computational acoustic vision by solving phase ambiguity confusion: CAVSPAC) which was similar to Jeffress-type models [15]. A distant single sound source could be localized precisely by using CAVSPAC. In this algorithm, interaural phase differences (IPDs) which were transformed from ITDs corresponded to horizontal azimuths of sound sources. Whether ITDs coding in the mammalian brain is similar to that in the avian brain has been a point of contention [6]–[12]. CAVSPAC can explain how ITDs relate geometrically to source azimuths independent of the neural coding strategy of the ITDs. Mechanical Learning through training with vision is another solution for sound source localization. ITDs can be related directly to horizontal azimuth of sound source even under diffraction [16]-[18]. In contrast to birds, which have no outer ear, most mammals have outer ears that are relatively large, transformable, and complicated in shape. In humans, the ears are located on opposite sides of the head. Although sound is diffracted by the outer ear, head, and body, sound diffraction has not previously been considered in Jeffress-type models [2]–[5].

In the present paper, the effect of sound diffraction on sound source localization is evaluated and a new method for compensating for sound diffraction in sound source localization is proposed. The proposed method uses CAVSPAC, one Jeffress-type model, applied to a human head and torso. Sensor interval is treated as a variable and is set wider than the actual interval during the calculating process for compensating the effect of sound diffraction on source localization. By making slight adjustments to sensor interval, a sensor interval-independent azimuth is identified. In Section II, the mathematical model and the proposed algorism are described. In Section III, experimental resuls under rebeverative condition are presented, comparing conventional CAVSPAC with the poposed method.

II. METHOD

Binaural microphones are mounted in the ear holes of a head and torso simulator. The effect of sound diffraction strictly depends on the sound wavelength and the three-dimensional shape and size of the outer ear, head, and body. In this section, a new method for compensating for sound diffraction in sound source localization is described.

A. CAVSPAC

Fig. 1 shows the geometry of sensors and a sound source. The azimuth θ for a sound source is identified at the origin, which is located at the midpoint between a pair of sensors. The sensor interval is S. The path difference Δl is expressed in terms of the phase difference $\Delta\varphi$ by

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 3, September 2013

$$\Delta l = \frac{c\Delta\varphi}{2\pi f}, \quad (1)$$

where c is the speed of sound and f is frequency. Due to "phase ambiguity", the following form is used rather than (1):

$$\Delta l = \frac{c\left(\Delta\varphi + 2\pi m\right)}{2\pi f}, \quad (2)$$

where m is an integer. The relationship between the path difference and the azimuth is also obtained from plane geometry:

$$\Delta l' = \sqrt{\left(L\cos\theta\right)^2 + \left(L\sin\theta + S/2\right)^2} - \sqrt{\left(L\cos\theta\right)^2 + \left(L\sin\theta - S/2\right)^2}. \quad (3)$$

Equations (2) and (3) are solved numerically under the following constraint:

$$\varepsilon = \left|\Delta l - \Delta l'\right| < \varepsilon_0, \quad (4)$$

where ε_0 is a given threshold value. Multiple azimuths are calculated corresponding to different integer values of m. The true source azimuth does not depend on frequency. This true azimuth is extracted as a frequency-independent value from the multi-azimuths identified.

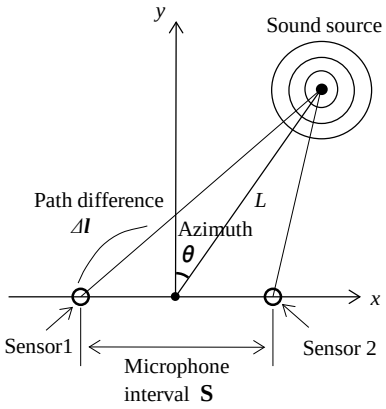

Fig. 1:Geometry of a sound source and a pair of sensors

B. Proposed method

The existence of a frequency-independent azimuth as described above is not always true when sound is radiated once over a brief period under reverberate conditions. However, if a sound continues long enough to be measured several times, several identified azimuths can be averaged over time to reduce random variation. In the present paper, a new method is proposed for reducing random variation of sound when sound is radiated once over a brief period. For identifying the source azimuth, two physical parameters are used, sensor interval S in (3) and speed of sound c in (2). These cannot be changed physically and humans cannot consciously recognize these values. Grothe et al. reported that "the binaural auditory system in mammals is less "hardwired" than has been imagined and appears instead to be highly dynamic, able to adjust rapidly its tuning properties to take account of the context in which sounds are heard"[11]. It seems reasonable to treat these parameters as variable in the calculating process for identifying the source azimuth. In the present paper, only sensor interval S in (3) is treated as a variable. The azimuth θ is a function of sensor interval S. The azimuth may be expected to change continuously and regularly when the sensor interval changes slightly. Then, the following additional constraint is introduced to CAVSPAC:

$$\varepsilon_{\theta S} = \frac{\partial\theta}{\partial S} << 1, \quad (5)$$

where $\varepsilon_{\theta S}$ is constant as small as possible. By adding (5), true azimuth can be identified from even a brief sound which is radiated once under reverberate conditions.

C. Algorithm

The proposed algorithm is described as follows.

- The pair of sound signals detected with sensors are $x(t)$ and $y(t)$. The cross-power spectra G_{12} is expressed as

$$G_{12} = X(f)Y^*(f), \quad (6)$$

where $X(f)$ and $Y(f)$ are the respective spectra after DFT processing and * indicates the complex conjugate.

The phase of G_{12} is express as

$$\angle G_{12} = \angle\left(X(f)Y^*(f)\right). \quad (7)$$

- The phase value given by (7) is substituted into $\Delta\varphi$ in (2) at each frequency. The azimuths θ_{mi} are identified, corresponding to integer m at frequency f_i from

$$\frac{c\left(\Delta\varphi_i + 2\pi m\right)}{2\pi f_i} \cong \sqrt{\left(L\cos\theta_{mi}\right)^2 + \left(L\sin\theta_{mi} + S/2\right)^2} - \sqrt{\left(L\cos\theta_{mi}\right)^2 + \left(L\sin\theta_{mi} - S/2\right)^2}. \quad (8)$$

- True azimuth is extracted as a frequency-independent value from the multiple azimuths identified by using the evaluation function $K(\theta_j)$. θ_{mi} is replaced with $\theta_{ij,}$ as m depends on frequency. evaluation function $K(\theta_j)$ is defined as

$$K(\theta_j) = \sum_i \theta_{ij}. \quad (9)$$

The frequency-independent azimuth θ_j is obtained as the maximizer of $K(\theta_j)$.

- A new process is added to CAVSPAC. The similar calculation is repeated for different sensor intervals. K azimuths θ_{kl} are selected for evaluation from the maximizer of function $K(\theta_l)$ as K candidates for each sensor interval S_l. another evaluation function $H(\theta_k)$ is defined as

$$H(\theta_k) = \sum_l \theta_{kl}. \quad (10)$$

The sensor-interval-independent azimuth θ_k is obtained as a maximizer of $H(\theta_k)$.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 3, September 2013

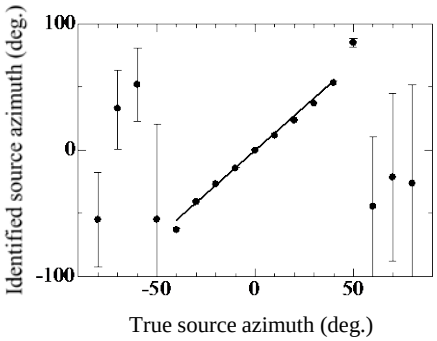

Fig. 2:This figure shows a comparison of the identified azimuths and true source azimuths. (S = 0.13 m)

III. EXPERIMENTAL RESULTS

The azimuth for a golf ball-sized loudspeaker at a distance 1 m (*L* in Fig. 1) from a human head and torso was identified. Broadband noise was continuously radiated from the loudspeaker, in which reverberate sound existed continuously. This method may be appropriate for different sound qualities, since phase does not depend on amplitude. The head and torso simulator (Type 4100-D, Bruel&Kjaer) in which the microphones were mounted was rotated on a turntable. Noise level was approximately 61 dB at the opposite outer ear when the loudspeaker was located at 90 degrees. The microphone interval was approximately 0.13 m. A pair of acoustical signals were measured and processed on a workstation (GX280, Dell) equipped with an A/D converter board (24-bit resolution, PCI-4474, N.I.). Sampling frequency was 24 kHz using a Hanning window for DFT. One frame included 4800 data for each channel. Measuring time was 0.2 s. The threshold value ε_0 was 0.005 m in (4).

Fig. 2 shows a comparison of the true azimuths and the identified azimuths by CAVSPAC. The value of the microphone interval was set to 0.13 m which was equal to the actual interval (*S* in (3)). Calculating process included (6) through (9) except (10). The identified azimuths were proportional to the true azimuths over a range of −40 degrees to +40 degrees. The proportionality coefficient was approximately 1.47. The source azimuth could not be identified for azimuths exceeding 50 degrees, as shown in Fig. 2. The frequency characteristics of the identified source azimuth and the corresponding distribution of function $K(\theta_j)$ are shown in Figs. 3(a) through (c), for the true source azimuths +30 degrees, +40 degrees, and +70 degrees, respectively. There was no frequency-independent azimuth in the case of Fig. 3(c), and so the source azimuth could not be identified when the loudspeaker was located at an azimuth of +70 degrees. The maximum peak of function $K(\theta_j)$ in (9) did not always show the true source azimuth for a wide range of azimuths. This implied that sound diffraction narrowed the identified range of possible source azimuths. The path difference Δl in Fig. 1 is calculated by using straight paths.

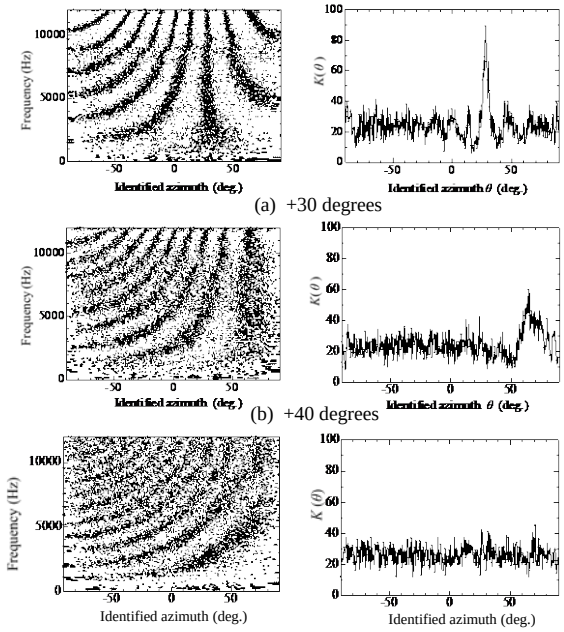

Fig. 3: These figures show frequency characteristics of identified azimuth and distribution of function $K(\theta_j)$ at different true source azimuths: (a) +30 degrees, (b)+40 degrees and (c)+70 degrees. (S = 0.13 m)

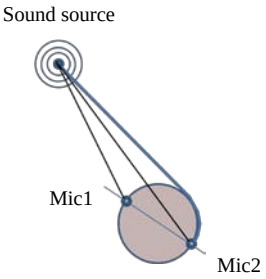

Fig. 4:This figure show two sound paths and a geometrical model at large source azimuth. It seems that the path difference Δl may be wider than microphone interval *S*. Geometrical model is not suitable for large source azimuth due to sound diffraction.

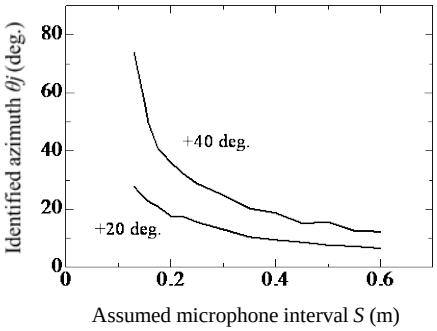

Fig. 5: This figure shows the identified azimuths for different assumed microphone interval at two source azimuths: +20 degrees and +40 degrees. Both azimuths converged to constant values for wider interval.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 3, September 2013

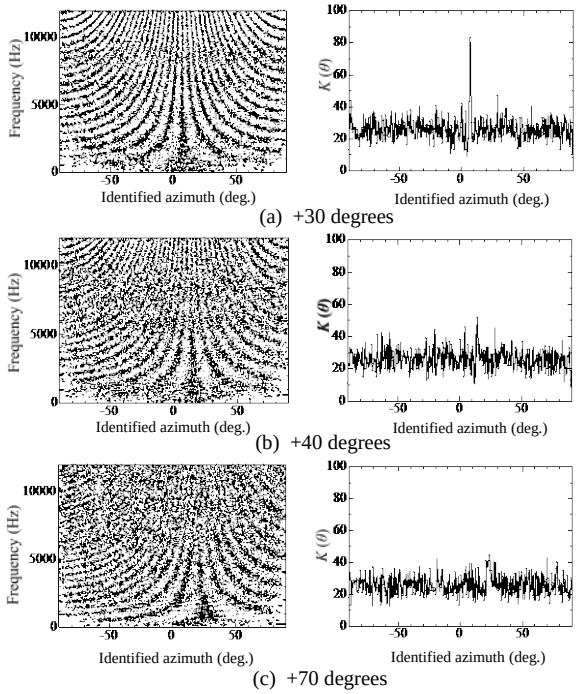

Fig. 6: These figures show frequency characteristics of identified azimuth and corresponding distribution of function $K(\theta_j)$ at different true source azimuths: (a) +30 degrees, (b)+40 degrees and (c)+70 degrees. The microphone interval was set wider (S=0.5 m) for calculation than physical interval. Maximum peaks of function $K(\theta_j)$ did not correspond to true azimuths. Function $K(\theta_j)$ included noise.

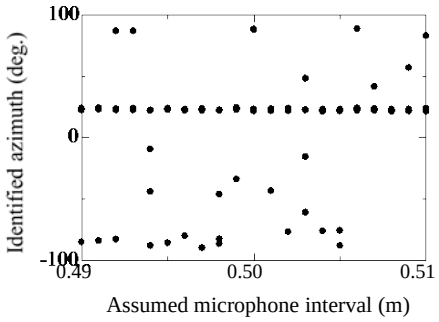

Fig. 7:This figure shows five identified azimuths corresponding to five peaks of function $K(\theta_j)$ at different microphone intervals when true azimuth is set +70 degrees. Azimuth +23 degrees was identified constantly even for different microphone intervals.

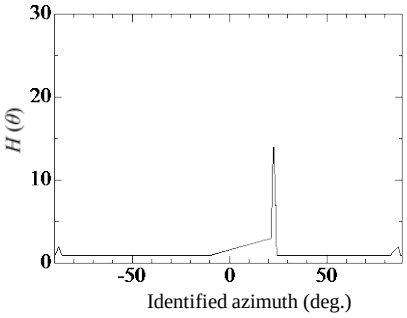

Fig. 8 shows distribution of function $H(\theta_k)$. Maximum peak was less than the true azimuth.

Sound actually diffracts around the far side of the head. When the true source azimuth exceeds 50 degrees, the path difference Δl might be estimated as wider than the maximum: microphone interval, S, as shown in Fig. 4. A geometrical model is not appropriate in this case. To avoid this situation, one solution is to set the microphone interval S in (3) as wider than the actual interval in order to satisfy $\Delta l < S$.

Let us check the relation between the identified azimuth and the microphone interval. Fig. 5 shows the identified azimuths for different assumed microphone intervals for two cases: +20 degrees and +40 degrees. The microphone interval S in (3) was taken to be from 0.13 m to 0.6 m for the same acoustical signals. Both identified azimuths decreased and converged to constant values for wider microphone interval. The identified azimuth shifted slightly when the microphone interval was assumed to be wider than 0.4 m. When the microphone interval was assumed to set 0.5 m, the frequency characteristics of the identified source azimuth and the corresponding distribution of function $K(\theta_j)$ are shown in Figs.6(a) through (c). There was a frequency-independent azimuth approximately at 9.9 degrees in Fig.6(a). Wide sensor interval expanded the identified range of possible source azimuth. However, function $K(\theta_j)$ included noise. Random noise can be usually cancelled by averaging over time when a sound continues long enough to be measured several times. It is necessary to distinguish signal from noise using a brief sound which is measured once.

Next, five peaks of function $K(\theta_j)$ were selected as probable candidates for evaluation, when the loudspeaker was located at an azimuth of +70 degrees. It is expected that the true identified azimuth may shift regularly and continuously when the microphone interval changes slightly. Fig. 7 shows five candidate azimuths corresponding to five peaks of function $K(\theta_j)$ in (9) for different microphone intervals. The microphone interval was varied over a small neighborhood of 0.5 m for the same azimuth, +70 degrees. The source azimuth was consistently identified as +23 degrees for different microphone intervals. Fig.8 shows the distribution of function $H(\theta_k)$ in (10), corresponding to Fig.7. The maximizer of $H(\theta_k)$ was +23 degrees clearly, comparing with Fig.6(c).The value of the identified azimuth was less than the true azimuths. So, a proportional coefficient of 2.98 was introduced to compensate the azimuth for sound diffraction. Fig.9 shows a comparison of the true azimuth and the compensated azimuth. The source azimuths could be identified even for azimuths exceeding 50 degrees. The compensated azimuths had good agreement with the true source azimuths.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 3, September 2013

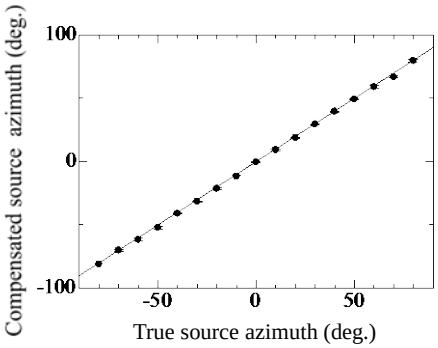

Fig. 9:This figure shows a comparison of compensated source azimuth and true azimuth. Coefficient value 2.98 is constant for different source azimuths. The compensated azimuths had good agreement with true azimuths in which standard deviation was low.

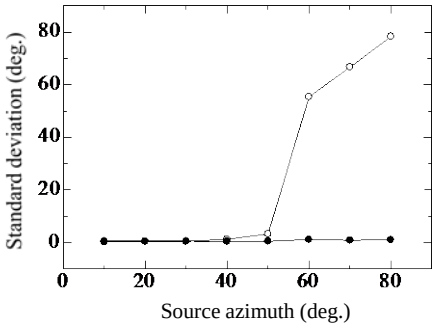

Fig. 10: This figure shows standard deviation values for different source azimuths by two methods. White circle indicates result under *S*=0.13 m. Black circle indicates result by proposed method(*S*=0.5 m, Maximizer of $H(\theta_k)$). Standard deviation values were low after compensation in the proposed

Figure 10 shows standard deviation values for different source azimuths by two methods. White circle indicates result under S=0.13 m. Black circle indicates result by proposed method (S=0.5 m, Maximizer of $H(\theta_k)$).

Standard deviation values were relatively low after compensation by the proposed method. The standard deviation value of the identified azimuth was 1.1 degrees when the true source azimuth was +80 degrees and five measurements were conducted under the same conditions. The relations between number of evaluated data, K and standard deviation values when measurements are repeated five times under same condition are shown in Table 1. Much evaluated data than two were appropriate for compensating sound diffraction.

Table 1: This table shows relations between number of evaluated data, *K* and standard deviation values when measurements are repeated five times under same condition. Much evaluated data than two were appropriate.

Number of evaluated data	Standard deviation (deg)
1	25
2	0.5
3	0.4
4	0.2
5	0.2
6	0.0

IV. CONCLUSION

The results are summarized as follows. Sound diffraction narrowed the identified range of possible source azimuths. Thus, sound diffraction is not negligible in a Jeffress-type model. The compensated azimuths agreed with the true azimuths and the standard deviation of the identified azimuth was relatively low.

This method is appropriate for a brief sound which is radiated only once. It should also be appropriate for different sound qualities, since phase does not depend on amplitude. This method will be applied to robot audition in future work.

ACKNOWLEDGMENT

This research was supported in part by a grant from Nihon University in 2009.

REFERENCES

[1] F. L. Wightman and D. J. Kistler,"Sound localization" in Human Psychophysics, W. A. Yost, A. N. Popper and R. R. Fay, Eds. (Springer-Verlag, New York), Chap. 5, pp. 155-192, 1993.

[2] E. A. Macpherson and J. C. Middlebrooks,"The duplex theory of sound localization revisited," J. Acoust. Soc. Am. Vol.11, No.5, Pt.1, pp.2219-2235, 2002.

[3] L. A. Jeffress, "A Place theory of sound localization,"J..Comp. Physiol. Vol.41, pp. 35–39, 1947

[4] C. E. Car and M. Konishi, "A circuit for detection of interaural time difference in the brain stem of the barn owl," J. Neurosci., vol.10, pp.3227-3246, 1990.

[5] M. Konishi, "Study of sound localization by owls and its relevance to humans," J. Comp. Biochem. Physiol., A, vol.126, pp. 459-469, 2000.

[6] D. McAlpine, D. Jiang, and R. Palmer, "Interaural delay sensitivity and the classification of low best-frequency biaural responses in the inferior colliculus of the guinea pig," Hearing Research Vol.97, pp.136-152, 1996.

[7] L. O. Trussell, "Synaptic mechanisms for coding timing in auditory neurons," Annu. Rev. Physiol., vol.61, pp.477-496, 1999.

[8] B. Grothe, "New roles for synaptic inhibition in sound localization," Nature rev. Neurosci., vol.4, pp.1-11, doi:10.1038/nrn1136, 2003.

[9] P. Joris and T. C.T. Yin,"A matter of time: internal delays in binaural processing," Trends in Neurosci., vol.30, 2, pp.70-78, 2006.

[10] N. A. Lesica, A. Lingner and B. Grothe,"Population coding of interaural time differences in Gerbils and barn owls," J.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 3, September 2013

Neurosci., vol.30, 35, pp.11696-11702, 2010.

[11] B. Grothe, M. Pecka and D. Mcalpine,"Mechanismsof sound localization in mammals," Physiol. Rev., vol.90, pp.983-1012, 2010.

[12] G. Ashida and C. E. Car,"Sound localization: Jeffress and beyond," Current opinion in Neurobiol., vol.21, pp.745-751, 2011.

[13] S. Karino, P. H. Smith, T. C. T. Yin and P. X. Joris, "Axonal branching patterns as sources of delay in the mammalian auditory brainstem: a re-examination," J. Neurosci., vol.31, 8, pp.3016-3031, 2011.

[14] N. H. Salminen, H. Tiitinen and P. J. C. May,"Auditory spatial processing in the human cortex," The Neuroscientist, Vol.XX, No.X, pp.1-11, 2012.

[15] R. Shimoyama and K. Yamazaki, "Computational acoustic vision by solving phase ambiguity confusion," Acoust. Sci. & Tech., vol.30, 3, pp. 199–208, 2009.

[16] F. Palmieri, M. Datum and A. Shah, "Learning binaural sound localization through a neural network," Bioengi. Conf., Proc. 1991 IEEE Annual Northeast, pp. 13-14, 1991.

[17] J. Backman and M. Kajalainen, "Modelling of human directional and spatial hearing using neural networks," Acoust. Speech, Signal Process. Vol.1, pp. 125-128, 1993.

[18] P. Smaragdis and P. Boufounos, "Position and trajectory learning for microphone arrays," IEEE Trans. Audio, Speech, Lang. Process. Vol.15, No.1, pp.358-368, 2007.

AUTHOR'SPROFILE

He is Professor of Nihon University, Japan. Doctor of Engineering in System Science at Okayama University, in 1994. 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. His research interest includes Artificial Intelligence on auditory system and Robotics. IEEE, IEICE, ASJ,SICE.