Proceedings of the 2003 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Las Vegas, Nevada · October 2003

Applying Scattering Theory to Robot Audition System: Robust Sound Source Localization and Extraction

Kazuhiro Nakadai†, Daisuke Matsuura‡ Hiroshi G. Okuno∗, and Hiroaki Kitano§

†*Honda Research Insutitute Japan, Co., Ltd.*
8-1 Honcho, Wako-shi, Saitama, 351-0114, Japan
‡ *Graduate School of Science and Engineering, Tokyo Institute of Technology,*
∗ *Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University*
§ *Kitano Symbiotic Systems Project, ERATO, Japan Science and Technology Corp.*

Abstract

Robot audition by its own ears (microphones) is essential for natural human-robot communication and interface. Since a microphone is embedded in the head of a robot, the head-related transfer function (HRTF) plays an important role in sound source localization and extraction. Usually, from binaural input, the interaural phase difference (IPD) and interaural intensity difference (IID) are calculated, and then the direction is determined by using IPD and IID with HRTF. The problem of HRTF-based sound source localization is that a HRTF should be measured for each robot in an anechoic chamber, because it depends on the shape of robot's head; HRTF should be interpolated to manipulate a moving talker, because it is available only for discrete azimuth and elevation. To cope with these problems of HRTF, we proposed the *auditory epipolar geometry* as a continuous function of IPD and IID to dispense with HRTF and have developed a real-time multiple-talker tracking system. This auditory epipolar geometry, however, does not give a good approximation to IID of all range and IPD of peripheral areas. In this paper, the *scattering theory* in physics is employed to take into consideration the diffraction of sounds around robot's head for better approximation of IID and IPD. The resulting system shows that it is efficient for localization and extraction of sound at higher frequency and from side directions.

1 Introduction

A robot that operates in daily environment should detect various sound events and pay attention to them to attain robust recognition and interact with people. So, audition is one of the most astects for both robot and human. To realize such robot audition under the real world, at least, the robot should have capability to localize sound sources and extract one of them, because a general sound that a robot hear consists of not a single sound source but multiple sounds such as voices, music, acoustic signals from TV and radio and noises by an air-conditioner and cars running outside.

Some robots developed so far had a capability of sound source localization. Huang *et al*[5] developed a robot that had three microphones. These 3 microphones were installed vertically on the top of the robot, composing a triangle. Comparing the input power of microphones, two microphones that have more power than the other are selected and the sound source direction is calculated. By selecting two microphones from three, they solved the problem that two microphones cannot determine the place of sound source in front or backward. By identifying the direction of sound source from a mixture of an original sound and its echoes, the robot turns the body toward the sound source. The Cog humanoid of MIT has a pair of omni-directional microphones embedded in simplified pinnae [2, 6]. In the Cog, auditory localization is trained by visual information. This approach does not use HRTF, but assumes a single sound source. Humanoids of Waseda University can localize a sound source by using two microphones [9, 16]. These humanoids localize a sound source by calculating *interaural intensity difference* (IID) or *interaural phase difference* (IPD) obtained by a *head-related transfer function* (HRTF).

Such sound source localization based on IPD and IID is expected to be effective because it is based on the human sound localization model proposed as the Jeffress's cross-correlator [7, 10]. On the other hand, HRTF has some difficulty in real-world proccessing. HRTF, which is often used for sound source localization in binaural research, is obtained by measurement of a lot of impulse responses. Because HRTF is usually measured in an anechoic room, sound source localization in an ordinary echoic room needs HRTF including room acoustic, that is, the measurement has to be repeated if the system is installed at a different room. However, deployment to the real world means that the acoustic features of the environment are not known in advance. It is infeasible for any practical system to require such extensive measurement

0-7803-7860-1/03/$17.00 © 2003 IEEE

of the operating space. Thus, a robot audition system without or at least less dependent on HRTF is essential for practical systems.

We proposed *auditory epipolar geometry* which can extract directional information of sound sources without using HRTF [13]. The auditory epipolar geometry is an extension of epipolar geometry in vision[4] (hereafter, *visual epipolar geometry*) to audition. Since the auditory epipolar geometry extracts directional information geometrically, it can dispense with HRTF. We applied the auditory epipolar geometry to a robot audition system based on active audition which improves auditory scene analysis by integration with audition and motion[13]. We attained real-time human tracking [11], sound source separation[12] and simultaneous speech recognition[15] by the robot audition system implemented on the humanoid *SIG* [1].

However, the auditory epipolar geometry judges only three directions – left, right or center – for IID, while it gives continuous estimatation of IPD against a direction of a sound source. In addition, it does not take the effect of scattered sound along the back of the head into account. Because IPD and IID are effective in lower and higher frequency, respectively[10], auditory processing at higher frequency is deteriorated. When a sound source is located at side direction, the scattered sound along the back of the head strongly affects the sound captured by the microphone on the opposite of the sound source. This means that the auditory epipolar geometry is inaccurate for sound from side direction.

In this paper, we use *Scattering Theory*[8] instead of the auditory epipolar geometry. The scattering theory is concerned with the effect obstacles or inhomogeneities have on incident waves. It provides a method to estimate the scattered field from the knowledge of the incident field and the scattering obstacle, that is, this gives accurate estimation of IPD and IID without acoustic measurement by calculating the difference between the total fields at the left and right ears.

The paper is organized as follows: Section 2 presents how to model humanoid head acoustically by scattering theory. Section 3 compares the scattering theory with the auditory epipolar geometry and HRTF. Section 4 proposes new sound source localization and extraction as applications of the scattering theory based IPD and IID estimation. Section 5 evaluates the system, and the last section gives the conclusions.

2 Modeling of Humanoid Head by Scattering Theory

Humanoid head is assumed to be a sphere with radius a. The spherical polar coordinates (r,θ,φ) are used. The ranges of the values are $0 \leq r < \infty$, $0 \leq \theta \leq \pi$ and $0 \leq \varphi < 2\pi$. A sound source at $\boldsymbol{r_0} = (r_0, 0, 0)$ is defined by

$$V^i = \frac{v}{2\pi R f} e^{i\frac{2\pi R f}{v}}, \tag{1}$$

where f and v are the frequency and the velocity of sound, and R is a distance between the source $\boldsymbol{r_0}$ and an observation point $\boldsymbol{r}$.

On the surface $r = a$, the total field of incident and scattered velocity potential is defined by

$$\begin{aligned} S(\theta, f) &= V^i + V^s \\ &= -\left(\frac{v}{2\pi a f}\right)^2 \sum_{n=0}^{\infty} (2n+1) P_n(\cos\theta) \frac{h_n^{(1)}(\frac{2\pi r_0}{v} f)}{h_n^{(1)'}(\frac{2\pi a}{v} f)}, \end{aligned} \tag{2}$$

where P_n and $h_n^{(1)}$ are the first kind Legendre function and the first spherical Hankel function, respectively [1].

The left and right microphones locate at $M_l = (a, \frac{\pi}{2}, 0)$ and $M_r = (a, -\frac{\pi}{2}, 0)$, respectively.

When a sound source is located at $r_0 = (r_0, \theta, 0)$, The total velocity potential at the left and right microphones is defined by $S_l(\theta, f) = S(f, \frac{\pi}{2} - \theta)$ and $S_r(\theta, f) = S(f, -\frac{\pi}{2} - \theta)$. Thus, the IPD $\Delta\varphi_s$ and IID $\Delta\rho_s$ are calculated by

$$\begin{aligned} \Delta\varphi_s(\theta, f) &= \arg\left(S_l(\theta, f)\right) - \arg\left(S_r(\theta, f)\right), \\ \Delta\rho_s(\theta, f) &= 20\log_{10}\frac{|S_l(\theta, f)|}{|S_r(\theta, f)|}. \end{aligned} \tag{3}$$

3 Estimation of IPD and IID

In this section, to make sure the accuracy of the IPD and IID estimation method base on the scattering theory, it is compared with HRTF and estimation based on the auditory epipolar geometry.

3.1 Measurement of HRTF

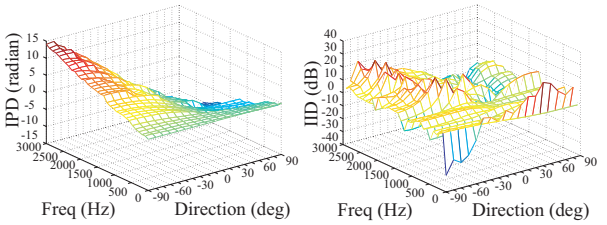
To obtain HRTF of *SIG*, impulse responses are measured in an anechoic room from -90° to 90° at the interval of 10° when the center direction of the robot is 0°. The room has hardly reverberation at the frequency of more than 125 Hz. The robot is installed a pair of microphones in the ear positions, and a loudspeaker is used for sound source. The distance between the humanoid and the loudspeaker is 1 m.

Let $Sp_l(f)$ and $Sp_r(f)$ be the left and right channel spectra of frequency f obtained by FFT from the captured impulse response. Then, the IPD $\Delta\varphi$ and IID $\Delta\rho$ are calculated as follows:

$$\begin{aligned} \Delta\varphi &= \arg\left(Sp_l(f)\right) - \arg\left(Sp_r(f)\right) \\ \Delta\rho &= 20\log_{10}\frac{|Sp_l(f)|}{|Sp_r(f)|}. \end{aligned} \tag{4}$$

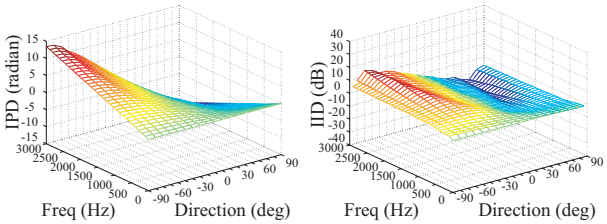
The IPD and IID obtained from HRTF are shown in Figures 1a) and b), respectively.

[1] Our research and publications on the humanoid *SIG* are described in http://winnie.kuis.kyoto-u.ac.jp/SIG/



a) IPD b) IID

Figure 1: *HRTF Measurement in Anechoic Room*



a) IPD b) IID

Figure 2: *Estimation by Scattering Theory*

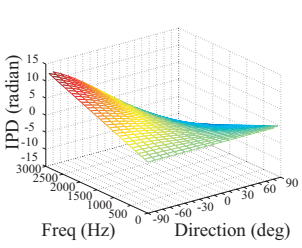


Figure 3: *IPD estimation by auditory epipolar geometry*

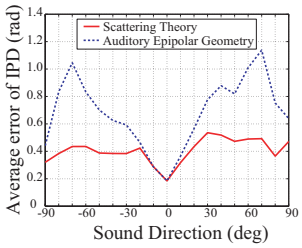


Figure 4: *Error of IPD*

3.2 Auditory Epipolar Geometry

The auditory epipolar geometry is an extension of epipolar geometry in vision (hereafter, *visual epipolar geometry*)[4] to audition. Since the auditory epipolar geometry extracts directional information by using the geometrical relation, it can dispense with HRTF. When the distance between a sound source and *SIG* is more than 50 cm, the influence of the distance can be ignored [12]. Then, when the influence by a head shape is considered, the auditory epipolar geometry is defined by

$$\Delta\varphi_e = \frac{2\pi f}{v} \times r\left(\theta + \sin\theta\right) \tag{5}$$

where f, v, r and θ are the frequency of sound, the velocity of sound, radius of a humanoid's head and the sound direction, respectively. $\Delta\varphi$ is an estimated IPD corresponding to θ.

It is easy to estimate IPD, but it does not have any mathematical model on IID. So, from IID, only three directions such as the left, the center and the right of the robot can be judged. The IPD estimated by the auditory epipolar geometry are shown in Figure 3.

3.3 Comparison of IPD and IID Estimation Methods

From Figures 1a) and b), the head of the robot has acoustic effects as follows:

1. the IID is increasing from 0° to near 60°, has the maximum values at near 60°, and is decreasing to 90°,
2. the IPD is increasing from 0° to 90°, and
3. This tendency is true of every frequency.

Generally, the IID is expected to have the maximum values at 90° as well as the IPD because the distance difference between left and right paths from the sound source to the ears is the maximum at 90°. However, the IID has the maximum values at near 60°. This is caused by energy of scattering signals along the back of the head.

The IPD and IID estimated by the scattering theory, that is, Eq.(3) are shown in Figures 2a) and b), respectively. These figures show that the scattering theory works well in estimation of IPD and IID. Especially, the tendency of the estimated IID matches with that of Figure 1b) well, while the auditory epipolar geometry judges left, center or right.

Figure 4 shows IPD errors in the scattering theory and the auditory epipolar geometry against HRTF. Although, in Figure 3, the auditory epipolar geometry seems to estimate IPD well in comparison with Figure 2a), the error of the IPD estimation by the auditory epipolar geometry is larger than that of the scattering theory. Especially, more than 30°, the error of the auditory epipolar geometry is bigger than that of the scattering theory.

The auditory epipolar geometry does not consider scattering signal which travels along the back of the head. Therefore, the estimation of IPD is getting worse as sound direction goes away from the center direction of the robot. However the scattering theory based method is accurate even when the sound source goes away from the center of the robot. In addition, because it can estimate IID mathematically, it is more efficient in higher frequency.

4 Sound Source Localization and Extraction

To make sure how efficient in real world processing the accurate IPD and IID estimation by the scattering theory is, it is applied to sound source localization and extraction. The basic ideas of the sound source localization and extraction have been described in [14, 12].

The sound source localization system integrates IID and IPD of harmonic components to attain robust localization in the real world[14]. The sound source extraction system uses an active direction-pass filter(ADPF) which extracts sound originating from the specified direction [12]. However, both systems use the auditory epipolar geometry to estimate IPD. As for IID, a simple assumption to judge left, center or right is used. Therefore, the sound source localization and extraction are getting worse in higher frequencies and in the directions away from the center of the robot.

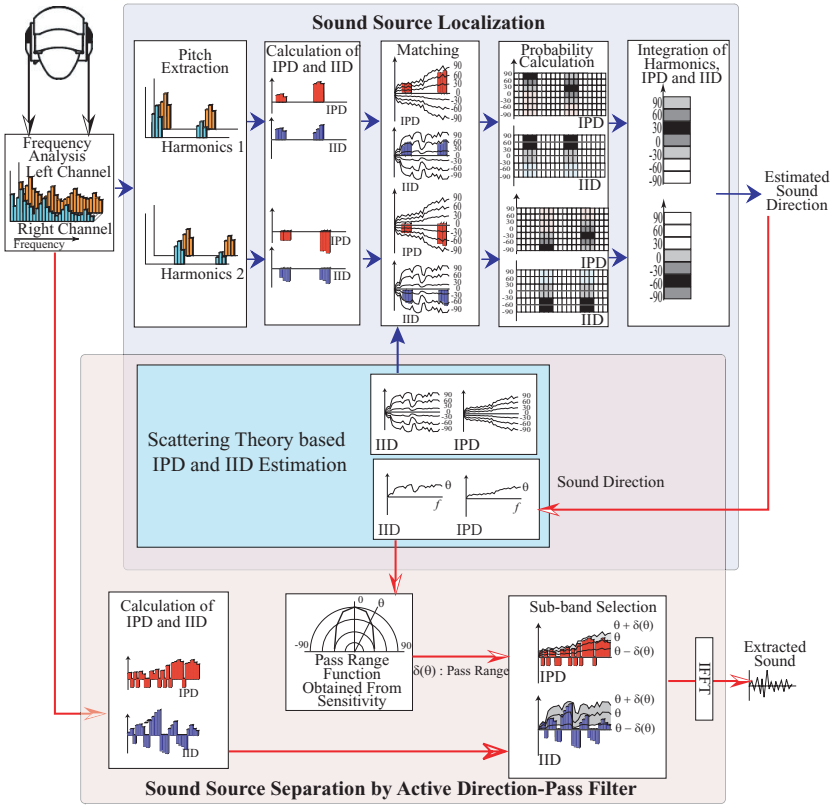


Figure 5: *Sound localization and extraction system using scattering theory*

Then, the IPD and IID estimation method based on the auditory epipolar geometry is replaced with that based on the scattering theory. The new sound localization and extraction are shown in Figure 5. The following sections describe the system in detail.

4.1 Sound Source Localization

The sound source localization system is shown in the upper dark area in Fig. 5. The sound source direction θ is assumed within $\pm 90^\circ$ by 10° from the median plane of the robot. At first, the sound source localization system extracts local peaks from a power spectrum, and they are clustered to be a harmonic sound according to harmonic relationships. The IPD and IID of the peaks included in the extracted harmonic sound is used for sound source localization. For IPD, the system creates hypotheses of IPD by the scattering theory for each θ. Then the distance between each hypothesis and the input harmonic sound is calculated at frequencies lower than f_{th}. The f_{th} is set to 1500 Hz, because the IPD is not efficient in localization for harmonics with more than 1500Hz, and the IID has the feature that it is emphasized at high frequency by the head because of the short wavelength. The cost function is as follows:

$$d_{\mathrm{IPD}}(\theta) = \frac{1}{n_{f \le f_{th}}} \sum_{f=\mathrm{F0}}^{f_{th}} \left(\varphi_h(\theta, f) - \varphi_s(f)\right)^2 / f) \quad (6)$$

where φ_h and φ_s are IPDs of a hypothesis and the corresponding harmonic component, respectively. $n_{f \le f_{th}}$ is number of harmonics lower than f_{th}.

In case of IID, the distance between each hypothesis and the input harmonic sound is calculated at frequencies higher than f_{th} . The cost function for IID is as follows:

$$d_{\mathrm{IID}}(\theta) = \frac{1}{n_{f > f_{th}}} \sum_{f=f_{th}}^{f_{max}} \left(\rho_h(\theta, f) - \rho_s(f)\right)^2 \cdot f) \quad (7)$$

where ρ_h and ρ_s are IID of a hypothesis and the corresponding harmonic component, respectively. $n_{f > f_{th}}$ is the number of harmonics higher than f_{th}.

The distance d_{IPD} and d_{IID} are transferred to belief factors of IPD P_{IPD} and IID P_{IID} using the probability density function. The IID and IPD are integrated based on the Dempster-Shafer theory[3] to get robust sound localization in the real world.

$$\begin{aligned} P_{\mathrm{IPD+IID}}(\theta) &= P_{\mathrm{IPD}}(\theta) P_{\mathrm{IID}}(\theta) + (1 - P_{\mathrm{IPD}}(\theta)) P_{\mathrm{IID}}(\theta) \\ &\quad + P_{\mathrm{IPD}}(\theta)(1 - P_{\mathrm{IID}}(\theta)) \end{aligned} \quad (8)$$

As a result of the integration, θ with maximum $P_{\mathrm{IPD+IID}}$ is regarded as a sound source direction θ_s.

4.2 Sound Source Extraction by ADPF

The architecture of the active direction-pass filter (ADPF) is shown in lower dark area in Figure 5. The ADPF extracts sound sources from a spectrum of input sound, IPD

and IID of the input sound, and sound source direction. The detailed algorithm of the ADPF as follows:

1. The pass range $\delta(\theta_s)$ of the ADPF is selected according to pass range function. The pass range function δ has a minimum value in the *SIG* front direction, because it has maximum sensitivity. δ has a larger value at the peripheral because of lower sensitivity. Let $\theta_l = \theta_s - \delta(\theta_s)$ and $\theta_h = \theta_s + \delta(\theta_s)$.
2. From a stream direction, the IPD $\Delta\varphi_E(\theta)$ and IID $\Delta\rho_E(\theta)$ are estimated for each sub-band by auditory epipolar geometry. Likewise, the IPD $\Delta\varphi_H(\theta)$ and IID $\Delta\rho_H(\theta)$ are obtained from HRTFs.
3. The sub-bands are collected if the IPD and IID satisfy the specified condition.
$$\begin{aligned} f \le f_{th} &: \Delta\varphi_s(\theta_l, f) \le \Delta\varphi(f) \le \Delta\varphi_s(\theta_h, f), \text{and} \\ f > f_{th} &: \Delta\rho_s(\theta_l, f) \le \Delta\rho(f) \le \Delta\rho_s(\theta_h, f). \end{aligned} \quad (9)$$
4. A wave consisting of collected sub-bands is constructed.

5 Evaluation

The sound source localization and extraction are evaluated against a harmonic sound which includes 30 frequency components from 100 Hz to 3000 Hz. a loudspeaker is used for the sound source. The distance between the loudspeaker and the robot is 1 m. A room of $3\,m \times 3\,m$ with 0.2 – 0.3 sec of reverberation time.

For sound source localization, the direction of the sound source varies from 0° to 90° at intervals of 10°. The reported localization system based on the auditory epipolar geometry is also used to compare with the proposed localization system base on the scattering theory. The reported system uses Eq.(5) to estimate IPD for lower frequency and judgment of left, center or right for higher frequency according to the IID.

For sound source extraction, the directions of the sound source are 0°, 30°, 60° and 90°. The pass range of the ADPF varies from $\pm 5°$ to $\pm 90°$ at intervals of $5°$. The HRTF based system and the auditory epipolar geometry based system are used to check the proposed sound extraction system. The HRTF based extraction system uses Eq. (10) for conditions in sub-band selection instead of Eq. (9).

$$\begin{aligned} f \le f_{th} &: \Delta\varphi_H(\theta_l) \le \Delta\varphi \le \Delta\varphi_H(\theta_h), \text{and} \\ f > f_{th} &: \Delta\rho_H(\theta_l) \le \Delta\rho \le \Delta\varphi_H(\theta_h) \end{aligned} \quad (10)$$

The auditory epipolar geometry based extraction system uses Eq. (11) instead of Eq. (9).

$$f \le f_{th} : \Delta\varphi_e(\theta_l) \le \Delta\varphi \le \Delta\varphi_e(\theta_h). \quad (11)$$

In every case, the extraction ratio R defined by Eq.(12) is measured.

$$R = 10\log_{10}\frac{\sum_{j=1}^{n}\sum_{i=1}^{m}(|sp(i,j)| - \beta|sp_o(i,j)|)^2}{\sum_{j=1}^{n}\sum_{i=1}^{m}(|sp(i,j)| - \beta|sp_s(i,j)|)^2}. \quad (12)$$

where $sp(i,j)$, $sp_o(i,j)$, and $sp_s(i,j)$ are the spectra of the original signal, the signal observed by robot microphones and the signal separated by the ADPF, respectively. m and n is the number of sub-bands and samples, respectively. β is the attenuation ratio of amplitude between original and observed signals.

The result of localization is shown in Figure 6. The extraction results of 0°, 30°, 60° and 90° are shown in Figures 7a), b), c) and d), respectively.

In Figure 6, the scattering theory based localization is more accurate than the auditory epipolar geometry. This means that the scattering theory based localization makes the best use of IID in higher frequency. The localization is getting worse as the sound direction goes away from the center direction of the robot in both systems. This is caused by the auditory fovea which means the resolution of localization is much higher in the center direction than in the periphery. However, the error of localization is smaller in the scattering theory based localization, especially in case of more than 50° of the input sound source direction. This is consistent with the result that the error of IPD is bigger in side directions in Figure 4. In

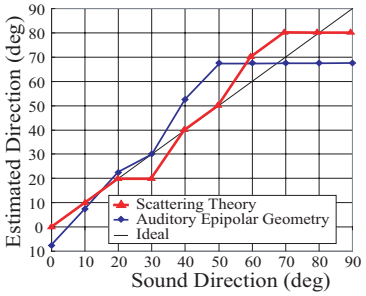
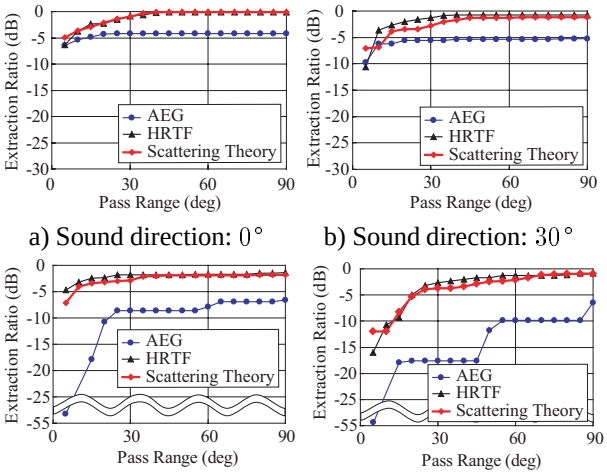


Figure 6: *Localization of Harmonic Sound of 100 Hz*

Figure 7a) – d), The performance of sound extraction by using the scattering theory and the HRTF is similar. This means that the scattering theory works well for sound extraction by estimating accurate IPD and IID. The scattering theory based extraction has an advantage in a sense that it does not require measurement in advance while the HRTF is obtained by a lot of measurement. By the scattering theory based extraction, the pass rage of the ADPF can be narrowed because it has higher extraction ratio at the same pass range than the auditory epipolar geometry based extraction. The narrower pass range are useful to reduce background noise. This is another advantage of the proposed extraction.

The performance of the auditory epipolar geometry based sound extraction is worse than that of the scattering theory because it is not taken the case of $f > f_{th}$ in Eq.(11) into account. The scattering theory based extraction improves by using IID information for sub-band selection in higher frequency of more than f_{th}. In Figures 7c) and d), the performance of the auditory epipolar geometry based extraction is getting much worse. This means that be-

cause IPD estimation is not accurate in side directions, the performance of sound extraction in every frequency is getting worse. The scattering theory based sound extraction is more accurate as shown in Figure 4. Therefore, the performance of sound extraction by the scattering theory is equal to that of the HRTF.



a) Sound direction: 0° b) Sound direction: 30°

c) Sound direction: 60° d) Sound direction: 90°

Figure 7: *Extraction of Harmonic Sound of 100 Hz (AEG: Auditory Epipolar Geometry)*

6 Conclusion

We improved sound source localization and separation for a robot by applying the scattering theory. In comparison with the auditory epipolar geometry, the scattering theory estimates IPD and IID more accurately. It has an advantage of mathematical estimation while the HRTF requires a lot of measurement in advance. This paper shows that sound localization and extraction of single sound source. However, in real-world processing, a mixture of sounds should be coped with, and robot's microphones and sound sources should be taken into account. We have already reported on a robot audition system with such sound localization and separation. The proposed localization and extraction methods are extensions of the reported ones. The proposed methods improve the accuracy in comparison with the reported auditory epipolar geometry based methods. To improve the robot audition system by applying the scattering theory based methods is a future work.

Acknowledgments

This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Scientific Research (A), 15200015 and Scientific Research on Priority Areas, 15017251. We thank Mr. K. Takashima of Nittobo Acoustic Engineering Co.,Ltd. for his valuable advice. We thank our ex-members in Kitano Symbiotic Systems Project: Dr. T. Lourens, Honda Research Institute Japan Co., Ltd. and Mr. K. Hidai, Sony Corp. for their discussions on visual processing. We also thank C. Okuno and Y. Tanemura for their help.

References

[1] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi. *Electromagnetic and Acoustic Scattering by Simple Shapes*. Hemisphere Publishing Co., 1987.

[2] R. Brooks, C. Breazeal, M. Marjanovie, B. Scassellati, and M. Williamson. The cog project: Building a humanoid robot. In C.L. Nehaniv, editor, *Computation for metaphors, analogy, and agents*, pages 52–87. Springer-Verlag, 1999.

[3] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[4] O. D. Faugeras. *Three Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, MA., 1993.

[5] J. Huang, N. Ohnishi, and N. Sugie. Building ears for robots: sound localization and separation. *Artificial Life and Robotics*, 1(4):157–163, 1997.

[6] R. E. Irie. Multimodal sensory integration for localization in a humanoid robot. *Proc. of the Second IJCAI Workshop on Computational Auditory Scene Analysis (CASA'97)*, pages 54–58. IJCAI, 1997.

[7] L.A. Jeffress. A place theory of sound localization. *Journal of Comparative Physiology, Psycology*, 41:35–39, 1948.

[8] P. Lax and R. Phillips. *Scattering Theory*. Academic Press, NY., 1989.

[9] Y. Matsusaka *et al*. Multi-person conversation via multi-modal interface — a robot who communicates with multi-user. In *EUROSPEECH-99*, pages 1723–1726. ESCA, 1999.

[10] B. C. J. Moore. *An Introduction to the Psychology of Hearing*. ACADEMIC PRESS, 1989.

[11] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, and H. Kitano. Real-time auditory and visual multiple-object tracking for robots. In *IJCAI-01*, pages 1424–1432. MIT Press, 2001.

[12] K. Nakadai, K. Hidai, H. G. Okuno, and H. Kitano. Real-time speaker localization and speech separation by audio-visual integration. In *ICRA 2002*, pages 1043–1049. IEEE, 2002.

[13] K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano. Active audition for humanoid. In *AAAI-2000*, pages 832–839. AAAI, 2000.

[14] K. Nakadai, H. G. Okuno, and H. Kitano. Epipolar geometry based sound localization and extraction for humanoid audition. In *IROS-2001*. IEEE, 2001.

[15] K. Nakadai, H. G. Okuno, and H. Kitano. Auditory fovea based speech separation and its application to dialog system. In *IROS-2002*, pages 1314–1319. IEEE, 2002.

[16] A. Takanishi, S. Masukawa, Y. Mori, and T. Ogawa. Development of an anthropomorphic auditory robot that localizes a sound direction (*in japanese*). *Bulletin of the Centre for Informatics*, 20:24–32, 1995.