International Journal of Humanoid Robotics

Detecting and Segmenting Sound Sources by using Microphone Array

Satoshi Kagami[1,2,3], Yuuki Tamai[1,3], Hiroshi Mizoguchi[1,3], Koichi Nishiwaki[1,2], Hirochika Inoue[1]

1: Digital Human Research Center,
National Institute of Advanced Industrial Science and Technology
2-41-6, Aomi, Koto-ku, Tokyo, 135-0064, JAPAN.
s.kagami@aist.go.jp

2: Japan Science and Technology Agency
3: Tokyo University of Science

This paper describes microphone array that can be used for sound localization and sound capture. The system has omni-directional sensitivity, and can be as a tele-microphone for apropriate direction. Sound capture by microphone array is achieved by Sum and Delay Beam Former (SDBF). Simulation of sound pressure distribution of 32 & 64ch circular microphone array are shown. According to simulation results, dedicated Firewire (IEEE1394) 32-channel board are developed with maximum sampling rate of 11kHz sample. The 32ch circular microphone array system is evaluated by using frequency components of the sound.

Keywords: Microphone Array; Sound Localization; Sound Segmentation.

1. Introduction

Sound source localization and capture are one of fundamental function for a robot that behaves in a human world. Not only voice capture by omitting background noise for a voice recognition, but also to notice human position or to catch a cue from human being will be important for a human-robot interaction. To achieve those two purpose, omni-directional (or wide range of) sensor coverage and tele-microphone function are simultaneously required.

Microphone arrays are often used for separating sound signals which are from a specific position and sound source localization[2,3,4]. The method of sound signal separation by a microphone array is already used for a voice recognition-specific microphone.

In this paper, authors developed a 32ch circular microphone array and a 128ch square microphone array for a robotics application. A circular microphone array can detect and capture arbitrary sound directions. Capturing sound by microphone array is achieved by Sum and Delay Beam Former. In order to build a microphone array, it is necessary to sample every channels simultaneously at high speeds. Dedicated Firewire (IEEE1394) 32-channel board is developed with maximum sampling

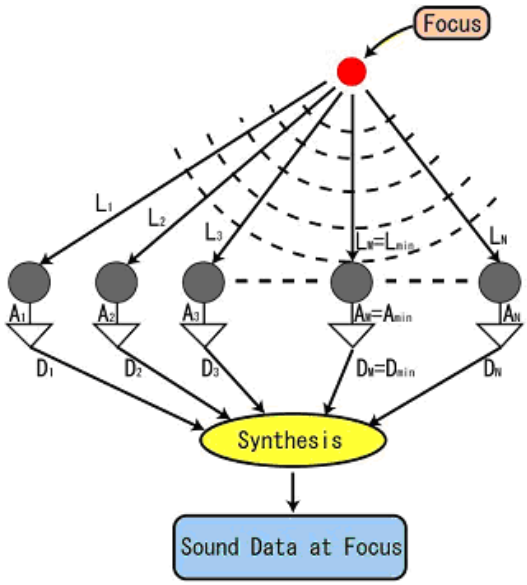

Fig. 1. Sound Capture by Microphone Array

rate of 11.025kHz sample.

2. Sound Capture Algorithm

Assume that focusing point C is an ideal omnidirectional sound source (Fig.1 and there are N microphones. Let $L_i (i < N)$ be distance from C and L_{min} be minimum distance among all L_i. Amplitude A_i and phase shift time D_i for each microphone related to the minimum distance microphone are calculated as follows.

$$A_i = \frac{L_{min}}{L_i} \tag{1}$$

$$D_i = \frac{L_{min} - L_i}{V_s} \tag{2}$$

Here V_s is a sound speed. Amplitude A_i was set as each microphone receives the same loudness from the point C. Consider sin curve $\sin(2\pi Ft)$ with frequency F is captured at i^{th} microphone, then following wave was arrived.

$$N_i(t) = A_i \sin(2\pi F(t + D_i)) \tag{3}$$

Where decreasing ratio with the minimum distance microphone B_i and time shift E_i at point C are described as follows.

$$B_i = \frac{L_i}{L_{min}} \tag{4}$$

$$E_i = \frac{L_i}{V} \tag{5}$$

Therefore, from the point C, following wave is obtained.

$$O_i(t) = \sin(2\pi F(t + \frac{L_{min}}{V})) \tag{6}$$

Every microphone input wave is synchronized and accumulated at the point C, so that total wave is as follows:

$$Q(t) = \sum_{i=1}^{N} O_i(t) = \sum_{i=1}^{N} \sin(2\pi F(t + \frac{L_{min}}{V})) \tag{7}$$

3. Simulation

In this chapter, simulations by using Sum and Delay Beam Former (SDBF) of 32 & 64ch circular microphone array.

3.1. *Sound Pressure Equations*

Simulation can be done by using the following parameters: microphone arrangement, number of microphone, frequency F, location $P(x, y)$. Let R_{iP} is the distance from i^{th} microphone and point P. Then total wave at point P can be described at follows:

$$Q_P(t) = \sum_{i=1}^{N} \frac{L_i}{R_{iP}} \sin(\omega t + \frac{L_{min} + R_{iP} - L_i}{V}) \tag{8}$$

where $\omega = 2\pi F$.
Set $\Theta_{iP} = \frac{L_{min}+R_{ic}-L_i}{v}$ and $F_{iP} = \frac{L_i}{R_{iP}}$. Then

$$\begin{aligned} Q_P(t) &= \sum_{i=1}^{N} F_{iP} \sin(\omega t + \Theta_{iP}) \\ &= A_p \sin(\omega t + \alpha) \end{aligned} \tag{9}$$

where,

$$\begin{aligned} A_p &= \sqrt{(\sum_{i=1}^{N} F_{iP} \cos(\Theta_{iP}))^2 + (\sum_{i=1}^{N} F_{iP} \sin(\Theta_{iP}))^2} \\ \alpha &= \arctan \frac{\sum_{i=1}^{N} F_{iP} \sin(\Theta_{iP})}{\sum_{i=1}^{N} F_{iP} \cos(\Theta_{iP})} \end{aligned} \tag{10}$$

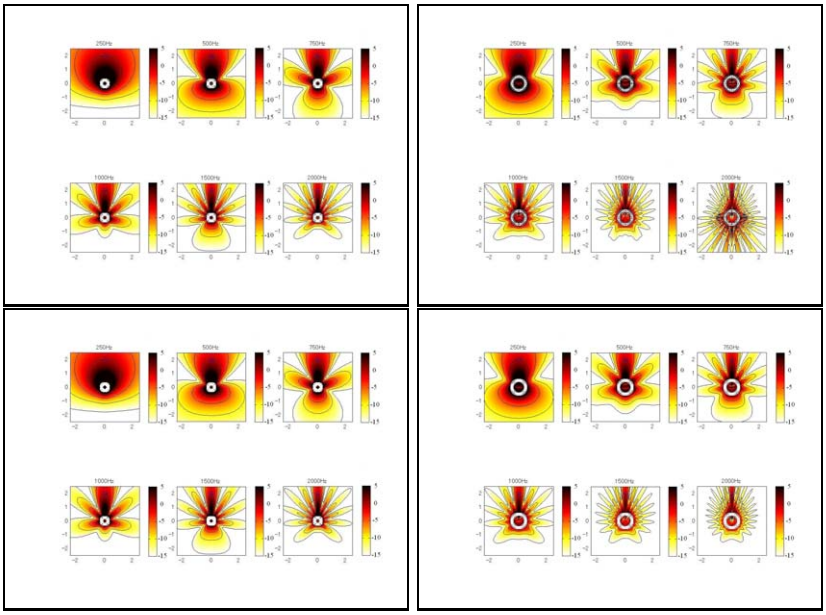

Fig. 2. Simulation result for 32ch, 64ch(ϕ 50[cm]) and ϕ 100[cm]) circle microphone

A_p is a total amplitude of the point w$P(x, y)$, and α is a phase shift.

Sound pressure level S_{PL}(dB) is calculated by using sound pressure P_1(Pa:N/m^2) and base sound pressure (P_0)($2 \times 10^{-5} N/m^2$) as follows:

$$S_{PL} = 20 \log_{10} \frac{P_1}{P_0} \tag{11}$$

In this simulation, sound pressure level of the point $P(x, y)$ is

$$S_{PL_P} = 20 \log_{10} \frac{A_p}{Q} \tag{12}$$

3.2. *Simulation Results*

In simulation, sound pressure sensitivity distribution map is created. Frequency ingredient of human voice is said to be 100-3000Hz, so, Sin wave of those bands are used in simulation.

Fig.2(top) shows simulation results of 32ch circular microphone array with diameter 50 and 100[cm]. Fig.2(bottom) shows simulation results of 64ch circular microphone array in the same condition.

In 32ch circular microphone array, above 1[kHz] is focused in 50cm diamiter caseFig.2(top left). At low frequency, high sound pressure area spreads out in wide range. So if sound source only generates low frequency sound, sound capture and sound source localization by SDBF will have limited performance.

Table 1. IEEE1394 32ch Microphone board specs.

	32ch Microphone
Protocol	IEEE1394
IF	TI TSB12LV32+TSB41AB1
FPGA	ALTERA EP1C6F256
AD	AD7490 x 4 (32ch)
Res.	12bits
Ratio	x800-5000 (hard wired)
Input	ECM
Transfer(Typ.)	Isochronous 15[us]

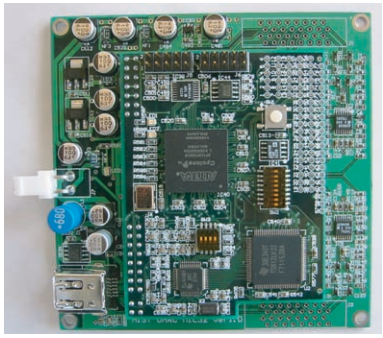

Fig. 3. IEEE1394 32ch Mic board

In 32ch circular microphone array with 100cm diamiterFig.2(top left), low frequency (especially at 500 and 750[hz]. However at 2[khz] there are many spiked high gain area which is caused by less resolution of microphone.

In 64ch circular microphone array with 50cm diamiterFig.2(bottom left), result doesn't change a lot from 32ch 50cm diamiter case. However, for 100cm diameter caseFig.2(bottom right), 2[khz] are improved from the sufficient resolution of microphone.

4. System Design and Implementation

A system that all components are driven by software, can have wide flexibility for developing the application. Especially focusing algorithm and digital filtering technique are important. Therefore, in this paper, real-time OS based online sound focus & capture system was developed. Fig.4 shows 32ch circular microphone array on top of Nomad XR4000.

Fig. 4. 32ch circular microphone array

4.1. *32ch Firewire Microphone board*

We developed a compact IEEE1394 microphone board for mobile robot purpose. (Fig.3(right)). The board transmits data by using isochronous transmission to host PC, so that no realtime OS is required. Table 1(right) shows specifications of the board.

5. Experimental Results

The performance of sound capture are evaluated by FFT results, and of sound source separation are done by sound pressure distribution.

5.1. *Sound Capture Experiments*

In 32ch circular microphone array, origin is set to its center. Two sound sources are placed at point A (1.0, 1.0) [m] and B (-1.0, 1.0) [m] and each sound sources generates 1.0[kHz] and 1.6[kHz] Sin wave simultaneously.

Fig.5(left) shows the experimental results of sound capture by 32ch circular microphone array. The system focuses on point A and B, and two sound data are captured. They are covered over FFT. At point A, the power spectrum of 1.0[kHz] is about 8[dB] larger than one of 1.6[kHz]. At point B, one of 1.6[kHz] is about 8[dB] larger than one of 1.0[kHz]. The performance of sound capture by the microphone array is a little weak, but the sound can be selectively captured.

5.2. *Sound Localization Experiments*

In 32ch circular microphone array, we put one to four speakers around the system. Fig.6(top) shows those conditions and detected sound peak. Normal human conversation are played from those speakers. Fig.6(bottom) shows the accuracy of sound

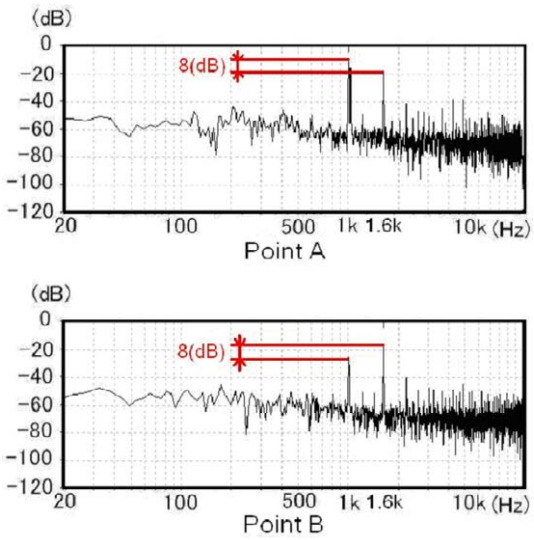

Fig. 5. Sound Capture Experiment by using 32ch Circle Microphone

source direction when only one speaker is put around the system. Maximum error is about 3[deg].

6. Voice Recognition

Julious/Julian is adopted for voice recognition software in Japanese. It is originally developed as a head-set based voice recognition system, but we adopt the system for robot usage. 100 words dictionary that includes noum, adjactive, verb with simple syntax are implemented. There are no conjunctive and no context in a given dictionary.

Recognition is done by placing two speakers, where each speakers are placed 200[cm] away from center of microphone array and 135[deg] separation. In case of two voices are played, recognition ratio is 71.4ratio is 78.6microphone input is less than 5

7. Concluding Remarks

In this paper, we described about online sound source localization and capture by using SDBF with number of microphones. Phase shifting parameter is calculated from the distance in between destination point and each microphone. More than one spot can be captured simultaneously, so that omni-directional sensitivity and tele-microphone effect can be obtained.

In simulation, we evaluate 32 & 64 circular microphone array, and decided to

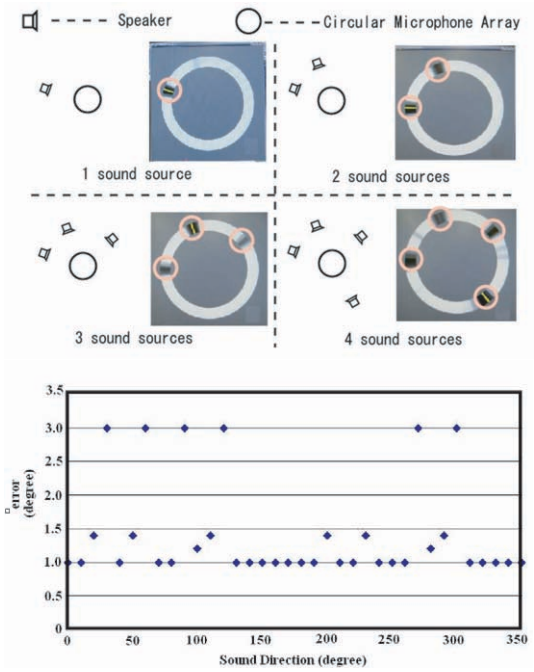

Fig. 6. Sound Localization Experiment and Detection Error by using 32ch Circle Microphone Array

implement 50cm diamiter 32ch circular microphone. We also developed 32ch simultaneous AD board with Firewire (IEEE1394) interface.

The performance of sound capture by 32ch circular microphone array is better than commercially available tele-microphone, and furthermore the system has omnidirectional sensitivity for locate sound sources. We also shows a brief experimental results of voice recognition in Japanese. In the next, we will put the system on humanoid H7 to achieve vision & sound combined human interaction functions.

References

1. Y. Tamai S. Kagami H. Mizoguchi, K. Shinoda and K. Nagashima. Invisible messenger: Visually steerable sound beam forming system based on face tracking and speaker array. In *Proceedings of the SICE Annual Conference 2003 (SICE2003)*.
2. H. Nomura, Y. Kaneda, and J. Kozima. Microphone array for near sound field. *Journal of Acoustic Society of Japan*, Vol. 53, No. 2, pp. 110–116, 1997.
3. Futoshi Asano, Shiro Ikeda, Michiaki Ogawa, Hideki Asoh and Nobuhiko Kitawaki. A Combined Approach of Array Processing and Independent Component Analysis for Blind Separation of Acoustic Signals.
4. Don H. Johnson and Dan E. Dundgeon. *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993. ISBN:0-13-048513-6.