2011 IEEE/RSJ International Conference on
Intelligent Robots and Systems
September 25-30, 2011. San Francisco, CA, USA

Improvement of Speaker Localization by Considering Multipath Interference of Sound Wave for Binaural Robot Audition

Ui-Hyun Kim, Takeshi Mizumoto, Tetsuya Ogata, and Hiroshi G. Okuno

***Abstract*—This paper presents an improved speaker localization method based on the generalized cross-correlation (GCC) method weighted by the phase transform (PHAT) for binaural robot audition. The problem with the conventional direction-of-arrival (DOA) estimation based on the GCC-PHAT method is a multipath interference whereby a sound wave travels to microphones via the front-head path and the back-head path in binaural robot audition. This paper describes a new time delay factor for the GCC-PHAT method to compensate multipath interference on the assumption of spherical robot head. In addition, the restriction of the time difference of arrival (TDOA) estimation by the sampling frequency is also solved by applying the maximum likelihood (ML) estimation in frequency domain. Experiments conducted in the SIG-2 humanoid robot show that the proposed method reduces localization errors by 17.8 degrees on average and by over 35 degrees in side directions comparing to the conventional DOA estimation.**

I. Introduction

SPEAKER localization is one of the most important techniques to achieve more natural and intelligent human-robot interaction (HRI). This is because robots need to be able to identify the direction of a sound source to judge the acoustic situation and watch the position of a talker to notify him/her that they are now ready to receive an order or express their interest in the conversation.

Recently, many researchers and engineers have conventionally used lots of microphones for robots to reinforce their speaker localization performance. However, using numerous microphones causes some problems: rising maintenance costs for microphones and computational power, and losing the general-purpose interface due to the different morphology of the microphone array for each robot. “Binaural” literally means having or relating to two ears. For robots, this means only two microphones located in the left and right sides of the robot head like human ears. The market for binaural audition hardware is growing because the cost of binaural audition hardware is much cheaper and uses much less computational power than multi-channel audition devices. Moreover, binaural audition hardware and its applications can be easily embedded on PCs, TVs, and other information and communication technology (ICT) devices. For these reasons, the binaural speaker localization system is necessary for robots.

Speaker localization usually deals with two techniques: voice activity detection (VAD) and sound source localization (SSL). VAD is used to provide delimiters for the beginning and end of a speech segment as exactly as possible from background noise such as music or other non-speech signals. It first extracts some features or quantities from the input signal and compares these measured values with a threshold. Many VAD algorithms have been based on zero-crossing rate, periodicity estimation, and signal energy level detection. The most well-known algorithm of this kind is the G.729B VAD [1].

SSL is defined as the identification of the location or origin of a detected sound in terms of direction and distance. It has been extensively studied by a number of researchers. As a result, the primary clues for sound localization have been discovered, including the inter-aural level difference (ILD), the inter-aural time difference (ITD), and the spectral modification due to the pinna, head, shoulder and torso. These clues are contained in the head related transfer function (HRTF) [2].The ITD, more commonly referred to as the time difference of arrival (TDOA), plays an important role in sound localization; the sound signals arrive at each microphone at different times due to the finite speed of sound and different positions of microphones from the sound source. One of the most popular algorithms using this ITD clue for SSL is the generalized cross-correlation method (GCC) and its phase transform (PHAT) weighting. GCC-PHAT is known as one of the most successful formulations of GCC and performs very well in noisy environments [3].

Many robot audition systems have been developed using the GCC-PHAT method, and their performance has gradually improved. However, most of these robot audition systems utilize their microphone array, which consists of lots of microphones, to protect the localization performance from various technical problems [4]. Since the binaural robot audition system consists of only two microphones embedded in the robot head, there are difficulties in obtaining a performance as good as that when using the microphone array.

In this paper, we describe two problems in binaural robot audition with the direction-of-arrival (DOA) estimation based on the GCC-PHAT method:

1) Restriction of the TDOA estimation by the sampling frequency in time domain: the cross-correlation function is calculated as the inverse Fourier transform of the cross-power spectrum in the GCC-PHAT method.

Ui-Hyun Kim, Takeshi Mizumoto, Tetsuya Ogata, and Hiroshi G. Okuno are with the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan (e-mail:{euihyun, mizumoto, ogata, okuno} @kuis.kyoto-u.ac.jp).

2) Multipath interference due to the diffraction of the sound wave along the shape of the robot head: sound wave easily bends around the robot head, and thus different TDOAs occur via the front-head path and the back-head path.

These two problems badly affect the DOA estimation, especially in the lateral direction of a sound source coming from around ±90 degrees. For accurate speaker localization, we solve these two problems by doing the following:

1) Applying the DOA estimation that can be calculated in the frequency domain by the maximum likelihood (ML).

2) Applying the new time delay factor considering multipath interference to the GCC-PHAT method once the robot head is assumed to be spherical.

These two solutions are implemented and evaluated experimentally in our binaural speaker localization system of a SIG-2 humanoid robot.

The outline of the paper is as follows: Section II summarizes the conventional DOA estimation based on the GCC-PHAT method and defines its two problems in binaural robot audition. Section III addresses their solutions. Sections IV and V outline our binaural speaker localization system and experimental results with discussions. Finally, Section VII concludes this paper.

II. Conventional Direction-Of-Arrival Estimation

This paper uses a time-frequency domain approach with a *T*-point short-time Fourier transform (STFT). The received signal from a *m*-th microphone can be mathematically modeled as

$$X_m[f,n] = H_m[f]S[f,n] + N_m[f,n], \tag{1}$$

where $X_m[f,n]$, $S[f,n]$, and $N_m[f,n]$ are *f*-th elements of the STFT of the measured signal from the *m*-th microphone, a sound source, and uncorrelated additive noise, respectively, on the *n*-th time-frame index. $H_m[f]$ is the transfer function between a sound source and the *m*-th microphone, $f \in \{0, fs / T, \ldots, fs (T-1) / T\}$ is a frequency, *fs* is a sampling frequency, and *T* is a frame size for the STFT.

Assuming that the distance between a sound source and a pair of microphones is significantly greater than half the distance between the pair of microphones, we can consider a sound as a plan wave and sound incidence reaching each microphone as parallel incidence [5]. Since sound sources usually occur far from measured microphones in the localization situation, this far-field assumption has been generally used as a simple formula for the DOA estimation. In an ideal scenario, since the transfer function between a far-field sound source and the microphones includes a propagation delay and a scaling factor, the received signal from the *m*-th microphone can be represented from (1) as follows:

$$X_m[f,n] = \alpha_m S[f,n] e^{-j2\pi f \tau_m} + N_m[f,n], \tag{2}$$

where α_m and τ_m are an attenuation factor and the time delay from the position of the sound source to the *m*-th microphone, respectively. TDOA τ_{ij} between two microphones *i* and *j* is defined by the relationship in (2) assuming microphone *i* as a reference:

$$\tau_{ij} = \tau_j - \tau_i = \frac{d_{ij}}{c} \sin(\frac{\theta}{180}\pi), \tag{3}$$

where d_{ij} is the distance between two microphones, $\theta \in \{-90, \ldots, +90\}$ is an angle of sound incidence, and *c* is the speed sound (340.5 m/s, at 15 °C, in air).

A. Generalized Cross-Correlation Method with the Phase Transform Weighting

The DOA estimation of a sound source can be obtained by estimating TDOA. One of the most common algorithms to estimate TDOA τ_{ij} from unknown parameters τ_i and τ_j is the GCC-PHAT method [6], which is defined as

$$\hat{R}_{x_i x_j}[n] = \sum_{f=0}^{fs(T-1)/T} G^{PHAT} X_i[f,n] X_j^*[f,n] e^{j2\pi f \tau_{ij}}, \tag{4}$$

$$G^{PHAT} = \frac{1}{\left| X_i[f,n] X_j^*[f,n] \right|}, \tag{5}$$

where $\hat{R}x_ix_j$ is the estimate of the cross-correlation function, * is the complex conjugate, and G^{PHAT} is a normalization factor to preserve only the phase information.

The cross-correlation function can be calculated as the inverse Fourier transform of the cross-power spectrum with the PHAT weighting for its computational efficiency [7] as follows:

$$\begin{aligned} csp_{ij}[t,n] &= ISTFT(G^{PHAT} X_i[f,n] X_j^*[f,n]) \\ &= ISTFT(\frac{X_i[f,n] X_j^*[f,n]}{\left| X_i[f,n] X_j^*[f,n] \right|}), \end{aligned} \tag{6}$$

where csp_{ij} is the coefficient of the cross-power spectrum phase (CSP) analysis, *t* is the time index, and *ISTFT* is the inverse short-time Fourier transform. As the coefficient of the CSP analysis presents a delta pulse centered on the delay, TDOA τ_{ij} is estimated as

$$\hat{\tau}_{ij}[n] = \arg\max_t (csp_{ij}[t,n]) \frac{c}{fs}. \tag{7}$$

After TDOA τ_{ij} is obtained, DOA θ of a sound source can be estimated by (3), which is rewritten as follows:

$$\hat{\theta}[n] = \sin^{-1}(\frac{\hat{\tau}_{ij}[n]}{d_{ij}}) \frac{180}{\pi}. \tag{8}$$

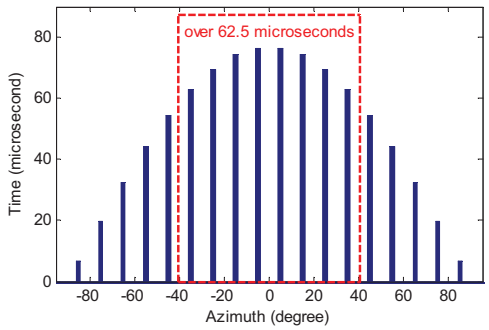

Fig. 1. Differences between TDOAs with 10-degree unit intervals when two microphones are 15 cm apart.

B. Two Problems in Binaural Robot Audition

Two problems in binaural robot audition with the GCC-PHAT method are explained in detail.

1) Restriction of the TDOA Estimation by the Sampling Frequency in Time Domain: Basically, the conventional DOA estimation comes from the TDOA estimation by the GCC-PHAT method, which is finally calculated in the time domain. In the conventional DOA method, estimated TDOA τ_{ij} must be restricted by the sampling frequency because the maximum value in (7) exists in the time domain through the inverse Fourier transform in (6).

This restriction causes TDOA estimation to be impossible in some cases. For instance, if a sampling frequency is 16 kHz, the minimum TDOA that can be estimated in the time domain is limited to 62.5 μsec (1 sec / 16 kHz). In other words, since the difference between TDOAs coming from -90 degrees and -80 degrees is less than 62.5 μsec when two microphones are at least 35 cm apart, we cannot distinguish TDOAs coming from -90 degrees and -80 degrees. Figure 1 shows a chart of these differences between TDOAs with 10-degree unit intervals in the azimuth from -90 degrees to +90 degrees. This chart was simulated with two microphones installed 15 cm apart in the binaural audition system of the SIG-2 humanoid robot, where the far-field assumption holds for any sound source at the distance of greater than 1.0 m. As Fig. 1 shows, the reliable TDOAs for distinguishing DOAs are only from -40 degrees to +40 degrees.

Simple solutions to this problem are to widen the distance between the two microphones, to increase the sampling frequency, or to use a microphone array by adding extra microphones, but these solutions also have their limitations, as a matter of course.

2) Multipath interference due to the diffraction of the sound wave along the shape of the robot head: In the conventional DOA estimation, TDOAs are estimated under the assumption that the microphones are located in free space. However, this assumption cannot be applied to the TDOA estimation using two microphones installed in the robot head. This is because the sound wave easily bends and spreads

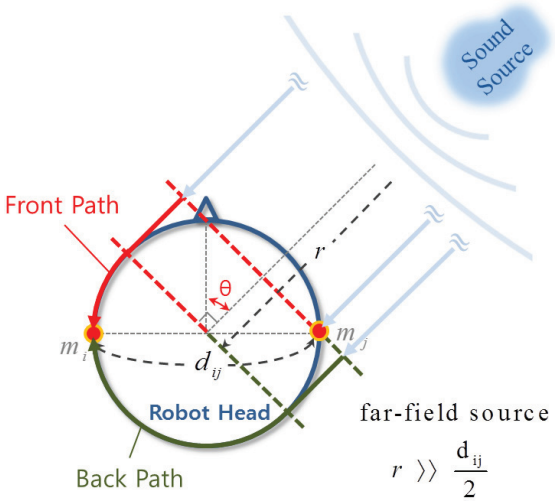

Fig. 2. Multipath interference due to diffraction of sound wave with spherical-head assumption.

along the shape of the robot head, and these attributes of sound cause different TDOAs from a sound source along the front-head path and the back-head path with multipath interference. Figure 2 illustrates these two paths of the sound wave with the assumptions that the robot head is spherical and the sound wave reaches each microphone as parallel incidence. Therefore, to more accurately estimate DOAs in binaural robot audition, the conventional DOA estimation has to be considered with these two paths of a sound wave and multipath interference in binaural robot audition.

III. Performance Improvement By Solving Problems

This section gives our solutions to the two problems mentioned in Section II-B through the analysis of causes. In binaural speaker localization, the two problems cause unreliable DOA estimations, especially around the lateral directions of a sound source coming from around ±90 degrees. This is because estimation errors increase as the sound incidence goes to -90 degrees or +90 degrees. For accurate binaural speaker localization, the two problems must be solved.

A. ML-based DOA Estimation in frequency domain

To solve the restriction of the TDOA estimation by the sampling frequency in the time domain, we applied the ML-based DOA estimation that is calculated in the frequency domain. This ML-based DOA estimation can be derived from (3)-(8) as follows:

$$\hat{\theta}[n] = \arg\max_{\theta} \sum_{f=0}^{fs(T-1)/T} \frac{X_j[f,n]X_i^*[f,n]}{\left|X_j[f,n]X_i^*[f,n]\right|} e^{j2\pi\frac{f}{T}fs\frac{d_{ij}}{c}\sin(\frac{\theta}{180}\pi)}, \quad (9)$$

where the estimated DOA θ of a sound source can be obtained by finding a degree θ that maximizes the sum of the

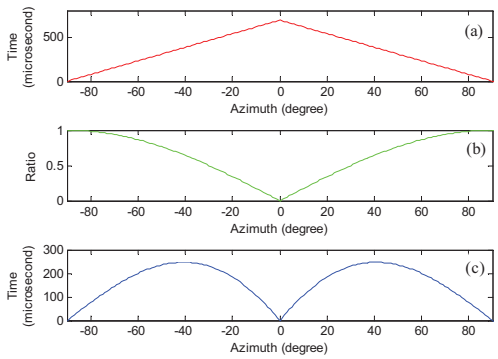

Fig. 3. Deriving compensation factor for multipath interference. (a) Absolute values of different time delays between two paths along front and back heads. (b) Absolute values of ILD ratios complied with sine function. (c) Distances calculated from (a) multiplied by (b) to compensate multipath interference.

cross-power spectrum with the PHAT weighting in frequency domain. This ML-based DOA estimator is able to get round the restriction by the sampling frequency in the time domain and has an advantage of 1 degree resolution estimation.

B. New Time Delay Factor

To solve multipath interference due to the two paths caused by the diffraction of the sound wave along the shape of the robot head, we first apply the simplified formula to these two paths after assuming that the robot head is a spherical:

$$Path_{front} = \frac{d_{ij}}{2c}\{\frac{\theta}{180}\pi + \sin(\frac{\theta}{180}\pi)\}, \quad (10)$$

$$Path_{back} = \frac{d_{ij}}{2c}\{\mathrm{sgn}(\theta)\pi - \frac{\theta}{180}\pi + \sin(\frac{\theta}{180}\pi)\}, \quad (11)$$

where $Path_{front}$ and $Path_{back}$ are the time delays along the front-head path and the back-head path, respectively, *sgn* is the signum function that extracts the sign of θ, i.e, if θ has the negative sign, then *sgn(θ)* will be -1. After deriving the formulas for two paths, the difference between the time delays of two paths in each sound direction can be obtained by

$$\begin{aligned} Diff_{front-back} &= Path_{back} - Path_{front} \\ &= \frac{d_{ij}}{2c}\{\mathrm{sgn}(\theta)\pi - \frac{2\theta}{180}\pi\}, \end{aligned} \quad (12)$$

where $Diff_{front-back}$ becomes 0 when θ is -90 or +90 degrees. Suppose that the intensity of the multipath interference from $Path_{back}$ for each sound direction complies with that of the ILD ratios between two microphones located in the robot head and this intensity of ILD ratios shows signs of the absolute values of the sine function in the ideal condition. We can consider $Diff_{front-back}$ multiplied by the absolute of the sine function as the factor to compensate for multipath interference:

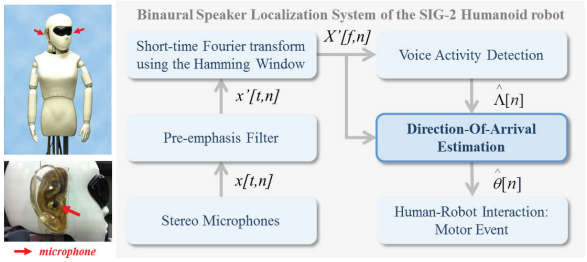

Fig. 4. System flow chart for binaural speaker localization.

$$Multi_{front-back} = \frac{d_{ij}}{2c}\{\mathrm{sgn}(\theta)\pi - \frac{2\theta}{180}\pi\}\left|\sin(\frac{\theta}{180}\pi)\right|, \quad (12)$$

where $Multi_{front-back}$ is the compensation factor for multipath interference in the binaural robot audition. This derived $Multi_{front-back}$ is shown in Fig. 3, which was simulated with a pair of microphone 15 cm apart. The final time delay factor for the DOA estimation can be derived with $Path_{front}$ and $Multi_{front-back}$ as follows:

$$\begin{aligned} \tau_{multi} &= Path_{front} + Multi_{front-back} \\ &= \frac{d_{ij}}{2c}\{\frac{\theta}{180}\pi + \sin(\frac{\theta}{180}\pi)\} \\ &\quad + \frac{d_{ij}}{2c}\{\mathrm{sgn}(\theta)\pi - \frac{2\theta}{180}\pi\}\left|\sin(\frac{\theta}{180}\pi)\right|. \end{aligned} \quad (13)$$

This new time delay factor τ_{multi} is utilized with the ML-based DOA estimation (9) as follows:

$$\hat{\theta}[n] = \arg\max_{\theta} \sum_{f=0}^{fs(T-1)/T} \frac{X_j[f,n]X_i^*[f,n]}{\left|X_j[f,n]X_i^*[f,n]\right|} e^{j2\pi\frac{f}{T}fs\tau_{multi}}. \quad (14)$$

IV. System Implementation

The ML-based DOA estimation with the new time delay factor τ_{multi} has been implemented in our binaural speaker localization system of a SIG-2 humanoid robot. Figure 4 shows the flow of this implemented speaker localization system. For the body of the system, we also used the pre-emphasis filter and the VAD algorithm as significant building blocks for speaker localization.

A. Pre-emphasis Filter

The goal of the pre-emphasis filter is to enhance the high frequencies of the speech spectrum, which are generally reduced by the speech production process [8]. Therefore, applying this filter to the acoustic signals measured from microphones attributes generally improves speaker localization performance.

The pre-emphasis filter is as follows:

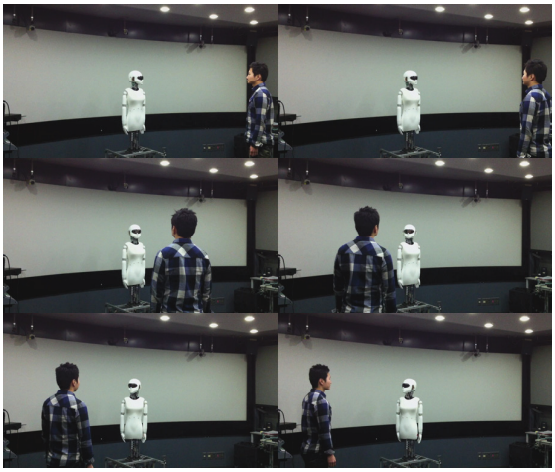

Fig. 5. Experimental setup.

$$x'[t,n] = x[t,n] - \omega x[t-1,n], \tag{15}$$

where $x[t,n]$ is the measured signal and $x'[t,n]$ is the pre-emphasized signal in time domain. Values of ω are typically in the range of 0.95 to 0.98.

B. A Statistical Model-Based VAD Algorithm

We used a statistical model-based VAD algorithm proposed by Sohn et al. This statistical model-based VAD algorithm requires fewer parameters for optimization than the G.729B VAD and indicates the presence or absence of speech with high accuracy [9].

The speech absent frame and the speech present frame are determined by the ML-based decision rule:

$$\begin{aligned} &\text{if} \quad \hat{\Lambda}[n] > \eta \quad \text{then} \quad n = \text{speech present} \\ &\text{else} \qquad\qquad\qquad\quad n = \text{speech absent} \end{aligned} \tag{16}$$

where

$$\hat{\Lambda}[n] = \frac{1}{T} \sum_{f=0}^{f_s(T-1)/T} \{\gamma[f,n] - \log \gamma[f,n] - 1\}, \tag{17}$$

$\gamma[f,n]=|X[f,n]|^2/\lambda_N[f,n]$ is the posteriori SNRs, $\lambda_N[f,n]$ is an estimated variances of $N[f,n]$, and η is a threshold, respectively.

For estimating the variance of noise $N[f,n]$, we recorded ambient noise for one second before operating the SIG-2 humanoid robot and utilized it as a priori noise.

V. Experiments and Results

We evaluate our ML-based DOA estimation in various ways to verify that it can make fewer localization errors than the conventional DOA estimation in binaural robot audition. For this purpose, the SIG-2 humanoid robot is utilized.

A. Experimental Setup

The experiment room contained the noise of air conditioners and personal computers. To create a noisy environment, background music that had lyrics was played as additive noise. The average dB of the music was about 61.2. The SIG-2 humanoid robot using the Sennheiser ME 104 omnidirectional microphone was placed in the center of the room and tested at its 1.5 m radius. A male speaker was placed at each locus of a 10-degrees-unit azimuth from -90 degrees to 90 degrees as shown in Fig. 5. The speech sources occurred 10 times at each locus in four different ways:

1) With the conventional DOA estimation using (6)-(8) mentioned in Section II.

2) With our ML-based DOA estimation using (9) by solving the problem mentioned in Section II-B1 and III-A.

3) With our ML-based DOA estimation using $Path_{front}$ in (10) to identify multipath interference from the back-head path.

4) With our ML-based DOA estimation using (14) considered multipath interference mentioned in Section II-B1 and III-B.

B. Results in Speaker Localization

Figure 6 shows the root mean square error (RMSE) of the experimental results on 190 occasions (azimuth change: 19 times × speech: 10 times) for each experimental method. According to Fig. 6, our ML-based DOA estimation had fewer localization errors than the conventional DOA estimation. Especially, our ML-based DOA estimation using (14) could reduce the average RMSE of the conventional DOA estimation by 17.8 degrees and the RMSEs of side directions by over 35 degrees.

C. Discussion

Through analyzing experimental results, we could verify that applying the ML-based DOA estimation that is calculated in the frequency domain all-round improves the localization performance (see cyan bar in Fig. 6). In addition to that, the ML-based DOA estimation using the new time delay factor τ_{multi} improved the localization performance effectively, especially around the side directions (See dark red bar in Fig. 6). This means that localization performance can be significantly improved by considering the two paths and its multipath interference of a sound wave along the shape of the robot head. The effect of multipath interference in localization can be identified by observing parabolic increases in RMSEs when we had tested the ML-based DOA estimation with $Path_{front}$ only (See orange bar in Fig. 6). These parabola increases were almost the same of those of the compensation factor we derived in Section III-B (See Fig. 3-c). As a result, despite our speaker localization system having only two microphones located in the robot head, our system showed the overall performance which is as good as other systems utilizing a microphone array.

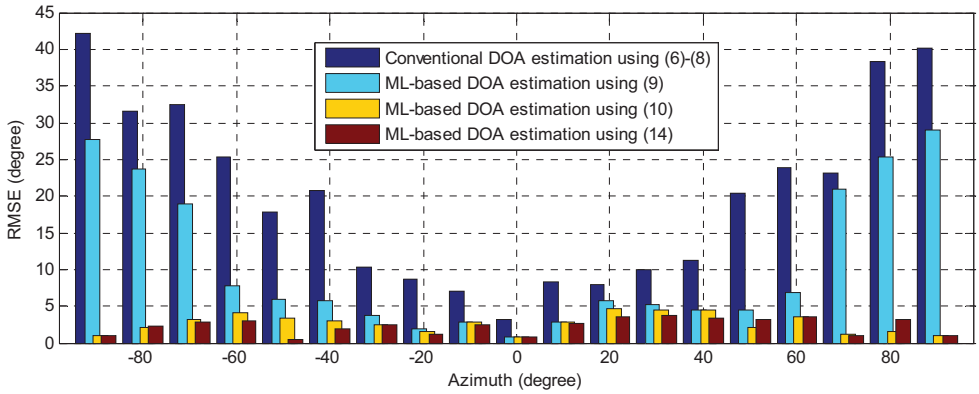

Fig. 6. RMSE of experimental results in speaker localization.

VI. Conclusion and Future Work

In this paper, we solved two problems with the conventional DOA estimation based on the GCC-PHAT method in binaural robot audition and their solutions. For the multipath interference due to the diffraction of the sound wave around the robot head, we applied the new time delay factor considering multipath interference to the GCC-PHAT method after the robot head is spherical. For the restriction of the TDOA estimation by the sampling frequency, the ML-based DOA estimation that is calculated in the frequency domain is applied to the GCC-PHAT method instead of the CSP analysis. Experimental results demonstrated that considering the front-head and back-head paths with multipath interference can be an important solution for improving localization performance in binaural robot audition.

Nakadai et al. exploits the scattering theory for a spherical head to estimate ILDs for lateral directions to improve the performance of localization [10]. The comparison of our multipath approach and the scattering theory may be interesting future work. Another important future work includes localization in whole horizontal space. This paper focuses on the azimuth localization from -90 to +90 degrees using our ML-based DOA estimation because of the front-back confusion problem in the binaural audition system [11]. To disambiguate the front-back confusion problem, some researchers adopt an approach called “active audition”, which means that the robot moves its microphones embedded on its body or head [12], [13]. We are currently developing a head that can move only its ears to disambiguate the front-back confusion problem.

Acknowledgment

This research was partially supported by JSPS Grant-in-Aid for Scientific Research (S), JST Japan-France Cooperative Research Project BINAAHR, and the Global COE Program of Graduate School of Informatics, Kyoto University.

References

[1] A. Benyassine, E. Shlomot, H.Y. Su, D. Massaloux, C. Lamblin, and J.-P. Petit, “ITU-T Recommendation G.729 Annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications,” IEEE Communications Magazine, vol. 35, pp. 64–73, Sept. 1997.

[2] C. I. Cheng and G. H. Wakefield, “Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space,” J. Audio Eng. Soc., vol. 49, pp. 231–249, Apr. 2001.

[3] V. M. Trifa, A. Koene, J. Moren, and G. Cheng, “Real-time Acoustic Source Localization in Noisy Environments for Human-robot Multimodal Interaction,” IEEE International Symposium on Robot and Human Interactive Communication, pp. 393-398, Jeju, Republic of Korea, Aug. 2007.

[4] U. H. Kim, J. Kim, D. Kim, H. Kim, and B. J. You, "Speaker Localization Using the TDOA-based Feature Matrix for a Humanoid Robot", IEEE International Symposium on Robot and Human Interactive Communication, pp. 610-615, Munich, Germany, Aug. 2008.

[5] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition)*. Cambridge, MA: MIT Press, 1997.

[6] C. H. Knapp and G. C. Carter, “The generalized correlation method for estimation of time delay”, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 24, no. 4, pp. 320-327, 1976.

[7] M. Matassoni and P. Svaizer, “Efficient time delay estimation based on cross-power spectrum phase,” European Signal Processing Conference (EUSIPCO), Florence, Italy, September. 2006.

[8] F. Bimbot, J. F. Bonastre, C. Fredouille, G. Gravier, I. Margrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. A. Reynolds, “A Tutorial on Text-Independent Speaker Verification,” EURASIP Journal on Applied Signal Processing, vol. 4, pp. 430-451, 2004.

[9] J. Sohn, N. S. Kim, and W. Sung, “A Statistical model-based voice activity detection,” IEEE Signal Processing Letters, vol. 6, no. 1, pp. 1-3, 1999.

[10] K. Nakadai, D. Matsuura, H. G. Okuno, and H. Tsujino, “Improvement of Recognition of Simultaneous Speech Signals Using AV Integration and Scattering Theory for Humanoid Robots”, Speech Communication, Vol.44, Issues 1-4 (Oct. 2004) 97--112, Elsevier.

[11] H. D. Kim, K. Komatani, T. Ogata, and H. G. Okuno, “Binaural Active Audition for Humanoid Robots to Localize Speech over Entire Azimuth Range, Applied Bionics and Biomechanics, Special Issue on "Humanoid Robots”, Vol.6, Issue 3 & 4(Sep. 2009), pp.355-368, Taylor & Francis 2009.

[12] K. Nakadai, D. Matsuura, H. G. Okuno, and H. Kitano, “Robot recognizes three simultaneous speech by active audition”, International Conference on Robotics and Automation, pp.398-405, 2003.

[13] E. Berglund, and J. Sitte, “Sound source localisation through active audition”, IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.653-658, 2005.