2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Using multiple microphone arrays and reflections for 3D localization of sound sources

* Carlos T. ISHI, Jani EVEN, Norihiro HAGITA (Intelligent Robotics and Communication Labs, ATR)

***Abstract*—We proposed a method for estimating sound source locations in a 3D space by integrating sound directions estimated by multiple microphone arrays and taking advantage of reflection information. Two types of sources with different directivity properties (human speech and loudspeaker speech) were evaluated for different positions and orientations. Experimental results showed the effectiveness of using reflection information, depending on the position and orientation of the sound sources relative to the array, walls, and the source type. The use of reflection information increased the source position detection rates by 10% on average and up to 60% for the best case.**

I. INTRODUCTION

IT is well-known that in real environments, the performance of applications having specific sounds as input (such as speech recognition systems) degrades due to changes in the environmental background noises along the time (such as in home, office or shopping malls). In order to deal with such a problem, the authors aim on developing a "sound environment intelligence" system, which has the ability to learn and make use of prior knowledge about the sound environment. For that purpose, the system has to be able to create a "sound environment map" (or simply "sound map") representing "when, where and what type of sound event occurred in the three dimensional space.

The problem of localizing sound sources by microphone array processing has been extensively studied so far [1 ~ 11]. However, only a few works evaluate sound localization in the 3D space, which is important to be considered in real situations where the elevation angles of the target sources relative to the sensors cannot be constrained. Also, due to physical constraints, the estimation of the sound source range (i.e., the distance from the array to the source position) is less robust than the estimation of direction angles by microphone array processing. Thus, in the present work we combine direction information from multiple microphone arrays in order to provide more accurate sound source position estimation in the 3D space.

This work was supported by the Strategic Information and Communication R&D Promotion Programme (SCOPE) under the Ministry of Internal Affairs and Communication, Japan.

C. T. Ishi is with the ATR Intelligent Robotics and Communication Labs., Kyoto, 619-0288 Japan (phone: +81-774-95-2457; fax: +81-774-95-1408; e-mail: carlos@atr.jp).

J. Even is with the ATR Intelligent Robotics and Communication Labs., Kyoto, 619-0288 Japan (phone: +81-774-95-2457; fax: +81-774-95-1408; e-mail: even@atr.jp).

N. Hagita is with the ATR Intelligent Robotics and Communication Labs., Kyoto, 619-0288 Japan (e-mail: hagita@atr.jp).

Further, a well-known problem of microphone array processing is when sound reflections occurring in the environment (such as in the walls, windows, ceiling, and displays) are observed by the microphone array besides the sound direct path. We have made sound recordings by fixing microphone arrays in the ceiling, and have often observed the presence of strong reflections. Although reflections are usually treated as a problem for sound applications, in the present work, we proposed a framework for taking advantage of reflection information in the sound location estimation problem.

Some works extend the problem of sound localization to the sound orientation estimation, by accounting the directivity properties of sound sources and propagation properties in the environment [12 ~ 15]. Although the estimation of sound orientation is also useful for characterizing the sound sources in an environment, in the present work, we focus on the sound location estimation problem. However, we also take into account the effects of the orientation of sound sources with different directivity properties in the evaluation.

It is worth clarifying that the term "localization" is broadly used concerning both "direction estimation" and "position estimation". In the present paper, we will use the term "localization" for indicating "position estimation".

II. THE PROPOSED METHOD

In the proposed method, the directions of multiple sound sources are firstly estimated using multiple arrays, then the directions of reflections are estimated by using spatial information, and all these information are integrated for estimating the location of the sound sources in the 3D space. The following Section II.A introduces the sound direction estimation method, which is based on the method proposed in our past work [10], while Section II.B introduces the proposed method using reflection information.

A. Sound direction estimation based on MUSIC spectrum

In the present work we adopt the MUSIC (Multiple Signal Classification) method for sound direction estimation, due to its high-resolution property [1,3,9,10].

Fig. 1 shows the block diagram of the algorithm for estimation of the direction of arrival (DOA) of sounds, based on the broadband MUSIC spectrum. The algorithm structure is similar to a classical approach of the MUSIC algorithm: 1) getting the Fourier transform (FFT) for computation of the multi-channel spectrum X(k,t); 2) computing the cross-spectrum correlation matrix R_k; 3) making the eigenvalue decomposition of the averaged correlation matrix

over a time block; 4) computing the (narrowband) MUSIC responses for each frequency bin using the eigenvectors corresponding to the noise subspace E_k^n and the steering/position vectors $a_k(\theta,\varphi)$ prepared beforehand for the desired search space; 5) computing the broadband MUSIC response $P(\theta,\varphi)$ by averaging the narrowband responses $P(\theta,\varphi,k)$ over a frequency range; 6) peak picking in the broadband MUSIC response to get the direction of arrival (DOA) of the sound sources.

The broadband MUSIC responses are referred as MUSIC spectrum, while the sequence of the MUSIC spectrums along the time is referred as MUSIC spectrogram.

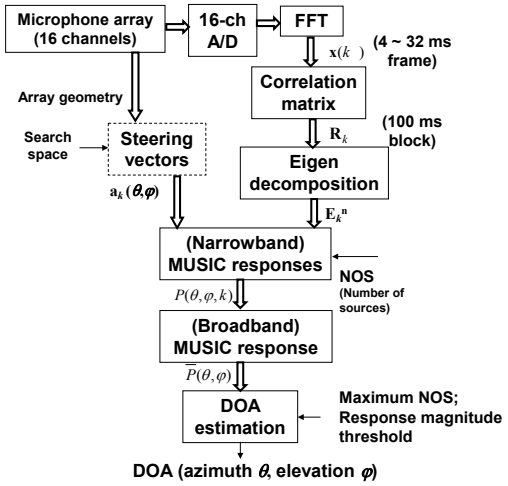

Fig. 1. The MUSIC-based sound direction estimation algorithm, and related parameters.

Although the MUSIC method has advantages of providing high resolution, it has two main issues which influence its performance in real applications: one is that the processing time increases as the number of microphones and the searching space increase, so that the real-time processing could not be achieved for 3D-space search; the second issue is that the number of sources has to be given beforehand for estimating the MUSIC spectrum.

Regarding the first issue, some of the parameters related to the MUSIC response computation were analyzed in [10], in order to obtain real-time processing, while keeping the DOA (direction of arrival) estimation performance. It has been shown that real-time processing is achieved by reducing the FFT frame size to 64 ~ 128 points (equivalent to 4 ~ 8 ms), and a block length of 100 ms, using a Core2Duo/2GHz CPU, for a DOA search space with 5 degree resolution on a spherical mesh, covering about 1200 directions (azimuth angles from 0 ~ 360 degrees, and elevation angles from -30 to 90 degrees) [10].

Regarding the second issue related to the estimation of the narrowband MUSIC spectrum, where the number of active sources at that moment has to be given, we adopted the solution proposed in [10], where the number of sources is fixed (considering that the estimation of the number of sources is not robust), and peaks in the MUSIC spectrum exceeding a threshold are picked. A maximum number of sources was also attributed.

It is worth to remind that although the range estimation is also possible by extending the search space for MUSIC spectrum computation. However, we restrict the MUSIC search space for the direction (azimuth vs. elevation) estimation for allowing real-time processing.

B. Sound localization based on multiple arrays and reflection information

In the present section, we describe the proposed method for sound location estimation based on direction estimation from multiple arrays and reflection information. A block diagram of the proposed approach is shown in Fig. 2.

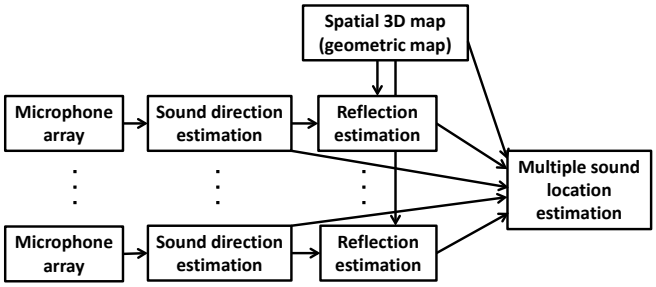

Fig. 2. The proposed sound localization system using multiple arrays and sound reflections.

The proposed method is based on a basic concept that if two or more sound directions detected by multiple arrays cross at certain location, it is probable that there is a sound source at that location. A straight solution is to estimate the sound directions by multiple microphone arrays, and search for the locations where the directions detected by different arrays cross.

We also take into account that the directions of the reflected sounds may also be observed by the array, depending on the position of the array relative to the ceiling, walls, displays, and other materials where sounds tend to reflect. In the proposed method, we consider that if the directions of the direct path and the reflection from ceiling or walls cross at certain location, it is also probable that there is a sound source at that location. Note that by using reflection information, the localization of a sound source may be possible by using a single microphone array, when the directions of both direct path and reflection are observed by the array.

Although reflections are usually treated as problematic sources in sound localization and separation applications using microphone array processing, in the present work we take advantage of reflection information for improving sound localization.

As we don't know beforehand if the detected direction is from a direct path or a reflection, all directions are firstly reversed when they reach a reflecting plane. This is done by mirroring the sensors in the walls and drawing the lines coming from the mirrored sensor positions. Although multiple reflections may occur in the space for the same source, we process only the first reflection, assuming that both power and directivity are weakened from the second reflection.

Sound directions are represented in terms of azimuth and elevation angles, for representation in 3D space.

For the estimated directions, there is an angle uncertainty (AU), so that the position uncertainty (PU) increases according to the distance between the array and the source. From geometrical rules, the position uncertainty can be straightly obtained by the following expression:

$$PU(r) = \pm AU / 360 * 2\pi * r, \quad (1)$$

where r is the source range (i.e. the distance from the array center to the source), AU is the angle resolution (in degrees) in the direction estimation process. For example, for a sound direction estimation with 5 degree resolution on a spherical mesh (AU = 5), and a range of 1 m (r = 1), the estimated position error of the source would be in the interval between ±8.7 cm. For a range of 2 m, the position error would be between ±17.4 cm.

We make use of the above error measurements for judging whether or not two detected directions cross at some location in the space. Firstly, the minimum distance between two directions ("dir_1" and "dir_2") are estimated by drawing lines in these directions and estimating the minimum distance between two lines, from the geometrical formulae

$$\text{dist}(dir_1, dir_2) = \frac{|(\overrightarrow{V_1}\times\overrightarrow{V_2}).\overrightarrow{P_1P_2}|}{|\overrightarrow{V_1}\times\overrightarrow{V_2}|}, \quad (2)$$

where V_i are vectors parallel to the detected directions, and P_i are the positions of the arrays.

The two lines are judged as crossing, if this distance is smaller than the summation of the absolute values of position uncertainty (PU) estimated for each direction, as shown in the following expression:

$$\text{dist}(dir_1, dir_2) < \text{abs}(PU_1(r_1)) + \text{abs}(PU_2(r_2)), \quad (3)$$

where dist() is the minimum distance between the lines drawn over the directions dir_1 and dir_2, r_1 and r_2 are estimations of the source ranges calculated as the distances from each array center to the line with minimum distance between dir_1 and dir_2. We then consider that it is probable that a sound source exists in the position where the two detected directions crossed.

The expressions (1) and (3) are evaluated for each pair of directions, including all direct paths and all possible reflections. When two directions are judged as crossing, the sound source location is determined by a weighted average between the points on the line with minimum distance, as shown in the following expression:

$$pos_{source} = (pos_1 * w_1 + pos_2 * w_2) / (w_1+w_2), \quad (4)$$

where pos_{source} is the estimated position of the sound source, pos_n are the positions (in the Cartesian coordinate) where each direction intersect the line with minimum distance, and w_n are weighting factors. As weighting factors, we decided to use the position uncertainty values $PU_i(r_i)$, or equivalently the distances r_i from the array to the line with minimum distance.

If the estimated source positions by two or more pairs of directions are close, they are treated as a unique sound source, and the estimated position of the pair with the smallest distance between directions ($\text{dist}(dir_1,dir_2)$) is selected.

Finally, pause intervals (blocks where no crosses were detected) smaller than 400 ms (equivalent to 4 blocks) are merged.

III. Analysis Data Description and Evaluation of the Proposed Method

A. Configuration of the microphone array and sound direction estimation parameters

Fig. 3 shows the geometry of the 16-element microphone array used in the present experiment. In order to have direction estimation in the 3D space, i.e. estimation of both azimuth and elevation angles, the microphones were distributed over a half-sphere of 30 cm diameter.

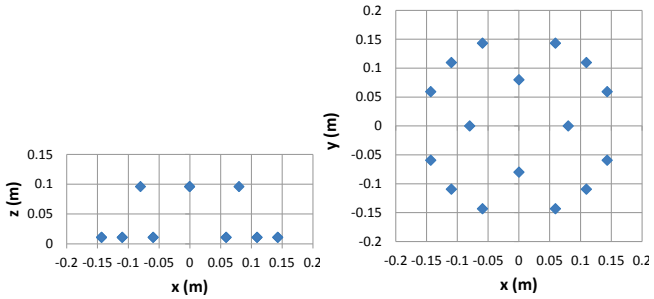

Fig. 3 The geometry of the 16-element microphone array.

For the multi-channel audio capture device, we used the 16-channel A/D converter TD-BD-16ADUSB from the Tokyo Electron Devices Ltd. For the microphones, we used the Sony omnidirectional condenser microphones ECM-C10. Audio was samples at 16 kHz/16 bits resolution for all 16 channels.

For the parameter setting of the sound direction estimation based on the MUSIC spectrum, the number of sources was fixed to 3, the MUSIC power threshold was set to 2.5 dB, and the maximum number of simultaneous sources at the same block was set to 6.

The frequency range of operation for estimating the MUSIC spectrum was set to 1500 ~ 5000 Hz, for avoiding spatial aliasing in high frequencies, and low spatial resolution in low frequencies.

The search space of sound direction estimation was set to 5 degree resolution in a spherical mesh. As the arrays are fixed in the ceiling, a range of 0 ~ 360 degrees is set for azimuth angle, while a range from -10 ~ -80 degrees is set for elevation angle. The elevation range between -80 ~ -90 degrees (right under the array) was removed from the search space since a MUSIC spectral peak appears in that range even when there is no sound source in that direction. This is because of the hardware characteristics of the multi-channel audio capture card used, where noise signals with equal phase are observed in all channels. From the symmetric geometry of the array, these equal-phase signals produce a peak around the line (x;y)=(0;0), which mean elevation angles around ±90 degrees.

B. *Data collection*

For the present experiment, we attached two arrays in the ceiling, as shown in Fig. 4. The locations of the arrays were decided in order to cover the sounds around the desks and the screen where other robot experiments with children have been conducted. An acoustic absorption material was put between the array and the ceiling, so that reflections from the ceiling are avoided. The flooring is constituted by tile carpets where reflections are smaller. Indeed, reflections in the floor could not be observed by the array. Thus, for the present work, we only consider reflections coming from the walls. However, depending on the array position and the reflecting properties of the floor, similar approach can be extended for accounting reflections in the ceiling and floor.

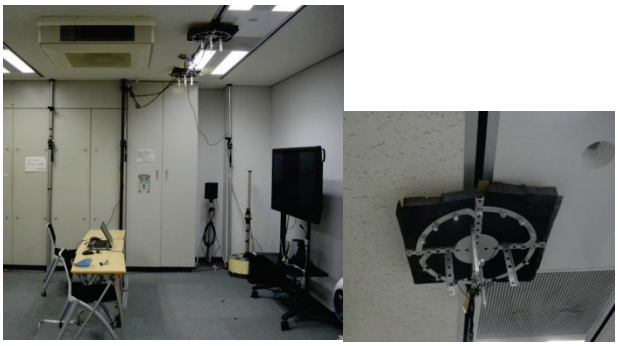

Fig. 4 The microphone arrays attached in the ceiling.

We also consider that the directivity of a sound source depends on its orientation. Therefore, the strength of directivity observed in the array may change according to the orientation of the sound source relative to the array, even if the source position is the same. Further, the directivity of sounds may also change according to the sound source type. In the present work, we used two source types: human speech, and the same speech played in a loudspeaker.

The air-conditioner (shown in the left-top of Fig. 4) was turned on. Since the air conditioner is close to the arrays, it is observed as a strong directional noise source by the arrays.

The target sources were positioned in six locations around the tables in Fig. 4, with four different orientations (front, back, left, right). It is difficult to fix the source position exactly, so that the mouth position was made unchanged as much as possible, when changing the orientations for the same position.

For the loudspeaker, we used the ONKYO GX-77M. The height of the loudspeaker was adjusted to the mouth height of the human speaker sitting in the chair. The human speaker was asked to utter the same sentences in similar speaking styles for all positions and orientations. From the loudspeaker, the same sentences recorded by the human speaker were played back. The volume of the loudspeaker was adjusted to approach the sound pressure level of the human speaker. The duration of the utterances for each trial was between 10 to 15 seconds.

Fig. 5 shows the positions of the microphone arrays (array1, array2), and the positions (1 ~ 6) and orientations (F: frontward, L: leftward, B: backward; R: rightward) of the sources. The height of the arrays is z = 2630 mm, while the height of the sources are z = 1160 mm. The walls are in the planes x=0 and y=0. The other walls are in the planes x=7.4 m and y=-5.6 m. However, reflections in these walls were disregarded from the present experiments since they are more than 4 meters far from the array positions.

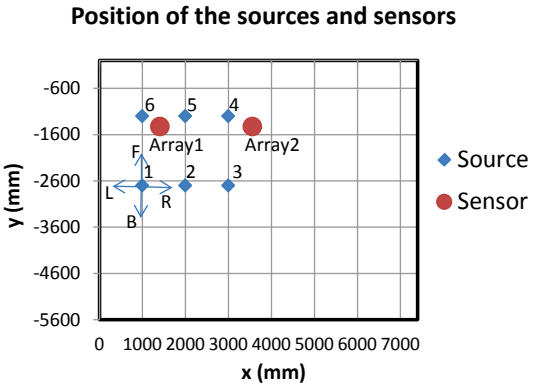

Fig. 5 Position (1 ~ 6) and orientation (F: front, L: left, B: back, R: right) of the target sources and the microphone array sensors (Array1, Array2) in the room. The external thick lines correspond to the walls.

C. *Effects of the source type and the source orientation relative to the array*

In this experiment, we firstly verified how accurate the source positions can be estimated by the measured directions of direct paths and reflections in each array. As evaluation measurements, the distances between the target source position and the lines drawn in the detected directions were computed.

In addition to the systematic errors due to position estimation uncertainty, we also consider the fact that both mouth and loudspeaker are not strict point sources, since the mouth aperture changes according to the phoneme, and sometimes airflow also comes from the nostrils in nasal phonemes, and the loudspeaker has a 9 cm diameter in a box with 13.6 cm width. Further, by also considering that the true positioning of the target sources is not exact, we decided to consider that the direction detected by the array comes from the source, if the distance between the reference and estimated positions is within 40 cm. For computing the position detection rates (in the end of this section), the maximum position estimation error was set to 30 cm.

The detection rates are then computed for each direction type (direct path or reflection) of each array, as the ratio between the number of blocks where the sound directions were correctly estimated and the total number of blocks in the utterance.

Fig. 6 shows the direction detection rates of the two source types ("human" and "loudspeaker"), arranged by each condition (position: "1 ~ 6", and orientation: "F, L, B, R"). The sound directions estimated by each array ("array1, array2") are categorized in direct path ("d"), reflection at the plane y=0 ("ry"), and reflection at plane x=0 ("rx").

From the results in Fig. 6, it can firstly be observed that the detection rates by direct path ("d") and reflections ("ry", "rx") in each array vary according to the both source position and orientation. This reflects the fact that depending on the position and orientation, the array "sees" directly a source or

"sees" the source mirrored in the wall. For example, in the condition "6L" for the "human" source (top of Fig. 6), both direct path ("d") and reflection at x=0 ("rx") are observed with a high rates around 0.8 in Array1. In Array2, the reflection "rx" has detection rate of 0.6, but the direct path ("d") has a very low detection rate around 0.1.

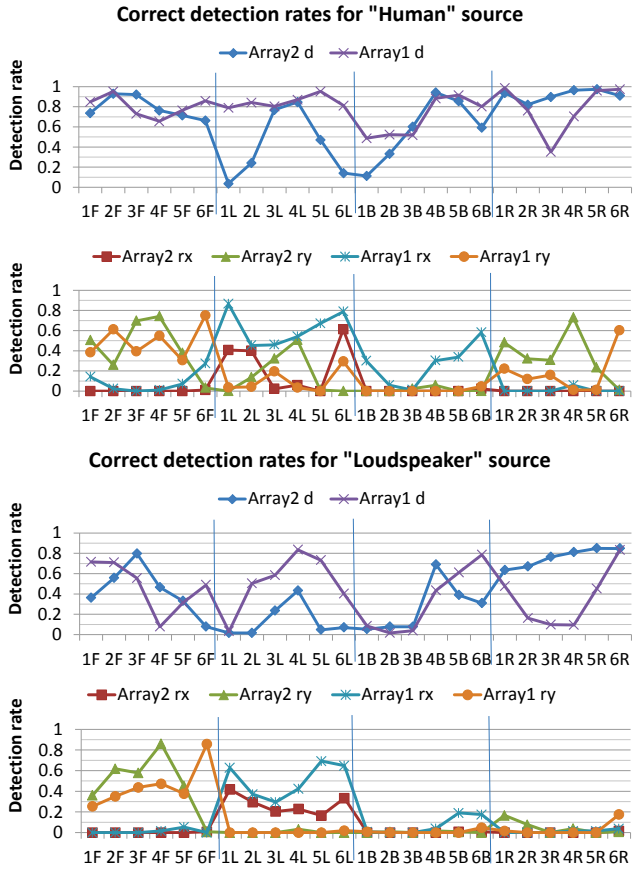

Fig. 6 Source direction detection rates for direct path (d) reflection at plane y=0 (ry) and reflection at plane x=0 (rx) by each array (Array1, Array2), for each position (1 ~ 6) and orientation (F: front, L: left, B: back, R: right) of the target sources ("human" and "loudspeaker").

Comparing the results of "human" and "loudspeaker", higher detection rates can overall be observed in "human". The reason is that directivity is stronger in loudspeakers, so that if the source is not facing the array, the reflections might be observed with comparable strength with the direct path.

As explained in Section II.B, the location where two or more sound direction intersects is probable of being the sound source position. For example, in the condition "6R" for the "human" source, the direct paths of both arrays are detected with rates larger than 0.9. For the "loudspeaker", high detection rates (larger than 0.8) are also observed. This means that the sound activity of sources at position "6" facing to right direction ("R") can be well detected by the directions estimated by the two arrays.

For the "human" source, high detection rates are observed for the direct paths ("d") in both arrays. In some of the cases ({1L, 2L, 5L, 6L} for Array1), the reflection at the plane x=0 ("rx") has detection rates than the direct paths of Array2. These are conditions where the sources are close and facing the wall at the plane x=0.

On the other hand, for the "loudspeaker", the only case where the direct path has the highest rate is the condition {6R}. The cases where the reflections ("rx" or "ry") have the highest rates are {4F, 5F, 6F, 1L, 2L, 5L, 6L}. These are cases where the loudspeaker is not facing to both arrays, so that the reflections on the walls are observed with higher directivity than the direct paths. Among these cases, in {4F, 1L} the rates for direct paths are close to zero. Therefore, the reflection information is shown to be important depending on the orientation of the sources.

In the conditions {1B, 2B, 3B}, the source is not facing to any of the arrays, so that low rates (lower than 0.6) were obtained for the "human" source, for both direct paths and reflections. In the "loudspeaker" source, which has strong directivity, detection rates close to 0 were obtained. In such cases, it would be necessary to increase the number of sensors, in order to cover the whole space.

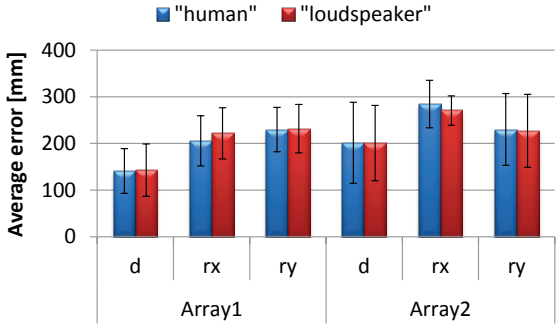

Fig. 7 Average position estimation errors by each direction type ("d", "rx", "ry") in each array, for each target source type ("human" and "loudspeaker").

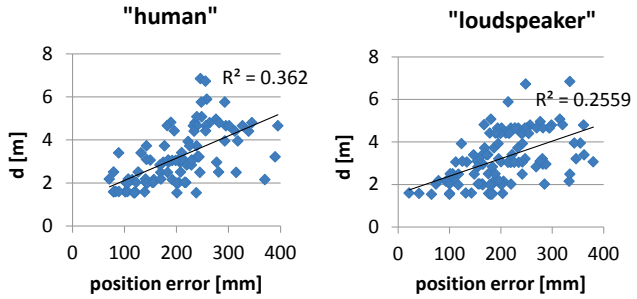

Fig. 8 Position estimation errors for each target source type ("human" and "loudspeaker").

Fig. 7 shows the average position errors, estimated by the distances between the target source position and the line drawn on each direction type (direct path "d", and reflections "rx" and "ry"). It can be observed that the average position errors do not change between the source types, but rather, seem to change according to the direction types in each array. However it is also observed that the average errors differ for the two arrays. These differences are due to the average distances from the array position to the source positions. Note that for reflections, the distance is not the physical distance, but rather the distance from the source position to the positions of the arrays mirrored at the walls. Indeed, correlation between the position estimation errors and the

distances from the array to the source can be observed in the distributions of Fig. 8. Correlation coefficients of 0.6 and 0.5 were observed for "human" and "loudspeaker" sources respectively.

Fig. 9 compares the source position detection rates by integration of direction information, using direct path only ("d") or direct path + reflections ("d,rx,ry"), before and after the merging process (merging pause intervals smaller than 400 ms where no crosses were detected). It is clear that the inclusion of reflection information increases the detection rates for the conditions {1L,2L,6L} for "human", and for the conditions {1F,4F,5F,6F,2L,5L,6L} for "loudspeaker".

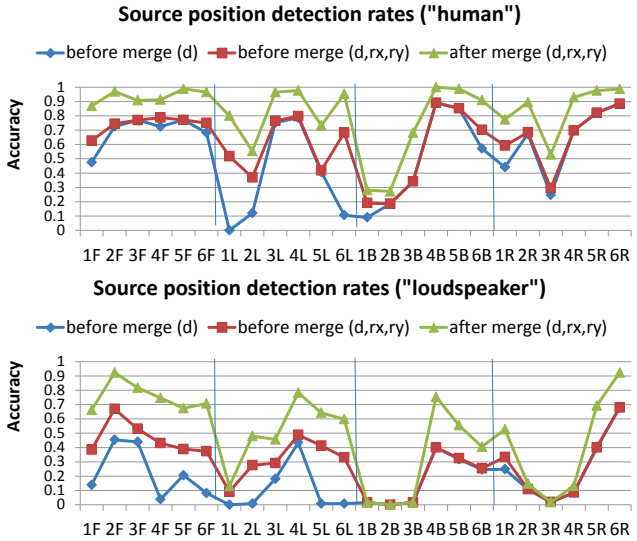

Fig. 9 Source position detection rates by integration of direction information, for each target source type ("human" and "loudspeaker"), using direct path only ("d"), direct path + reflections ("d,rx,ry"), before and after merging.

The inclusion of reflection information improved the position detection rates by about 10% on average, with the best improvement of about 60% for the case "6L" for the "human" source. The merging process increased the detection rates by 15 ~ 20% on average.

IV. Conclusion

In the present work, we conducted sound direction estimation in multiple microphone arrays, and proposed a framework for localizing sound sources in the 3D space, by making use of reflection direction information. It was shown that the performance of sound localization increased by using reflection information.

By comparing the sound direction estimation results for human speech and speech played back by a loudspeaker, it was observed that the directivity patterns of direct path and reflections observed in each microphone array varied according to several factors including the relative position of the array and walls, the source type, and the source position and orientation.

It is worth to emphasize that the results for human and loudspeaker were quite different, because the loudspeaker has stronger directivity properties than humans. It was even observed that if the loudspeaker is not facing to the array, its reflections in the walls are observed with stronger directivity by the arrays, rather than their direct paths. Although it is common to find "laboratory experiments" using loudspeakers in place of real humans in microphone array research, the results in the present work remind us that one should take directivity properties of the different types of sound sources into account.

The experiments of the present work were conducted using fixed arrays in the ceiling. However, similar approaches can be straightly extended for arrays in (moving) robots, if the self-localization of the robot is provided. Furthermore, the fixed array scenario can be used as part of a network robot system, so that robots can be automated to self-positioning in the best place for approaching the target source.

Future topics are integrating the sound localization approach with human tracker to improve speaker activity and embedding to the sound environment intelligence system.

References

[1] F. Asano, M. Goto, K. Itou, and H. Asoh, "Real-time sound source localization and separation system and its application on automatic speech recognition," in *Eurospeech 2001*, 2001, pp. 1013–1016.

[2] K. Nakadai, H. Nakajima, M. Murase, H.G. Okuno, Y. Hasegawa and H. Tsujino, "Real-time tracking of multiple sound sources by integration of in-room and robot-embedded microphone arrays," in *Proc. of IROS 2006*, Beijing, China, 2006, pp. 852–859.

[3] S. Argentieri and P. Danès, "Broadband variations of the MUSIC high-resolution method for sound source localization in Robotics," in *Proc. of IROS 2007*, San Diego, CA, USA, 2007, pp. 2009–2014.

[4] M. Heckmann, T. Rodermann, F. Joublin, C. Goerick, B. Schölling, "Auditory inspired binaural robust sound source localization in echoic and noisy environments," in *Proc. of IROS 2006,* 2006, pp.368–373.

[5] T. Rodemann, M. Heckmann, F. Joublin, C. Goerick, B. Schölling, "Real-time sound localization with a binaural head-system using a biologically-inspired cue-triple mapping," in *Proc. of IROS 2006*, Beijing, China, 2006, pp.860–865.

[6] Y. Sasaki, S. Kagami, H. Mizoguchi, "Multiple sound source mapping for a mobile robot by self-motion triangulation," in *Proc. of IROS 2006*, Beijing, China, 2006, pp. 380–385.

[7] J.-M. Valin, F. Michaud, and J. Rouat, "Robust 3D localization and tracking of sound sources using beamforming and particle filtering," *IEEE ICASSP 2006*, Toulouse, France, pp. IV 841–844.

[8] B. Rudzyn, W. Kadous, C. Sammut, "Real time robot audition system incorporating both 3D sound source localization and voice characterization," Proc. *ICRA 2007*, Roma, Italy, 2007, pp. 4733–4738.

[9] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, "Intelligent sound source localization for dynamic environments," in *Proc. of IROS 2009*, St. Louis, USA, 2009, pp. 664–669.

[10] C. T. Ishi, O. Chatot, H. Ishiguro, N. Hagita, "Evaluation of a MUSIC-based real-time sound localization of multiple sound sources in real noisy environments," in *Proc. IROS 2009*, 2009, pp. 2027–2032.

[11] F. Ribeiro, D. Ba, C. Zhang, and F. D., "Turning enemies into friends: using reflections to improve sound source localization," in Proc. of *ICME2010*, 2010, pp. 731-736.

[12] A. Brutti, M. Omologo, P. Svaizer, "Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays," in *Proc. of Interspeech 2005*, Lisbon, Portugal, 2005, pp. 2337-2340.

[13] H. Nakajima, K. Kikuchi, T. Daigo, Y. Kaneda, K. Nakadai, Y. Hasegawa, "Real-time sound source orientation estimation using a 96 channel microphone array," in *Proc. of IROS 2009*, pp. 676-683.

[14] R. Chakraborty, C. Nadeu, T. Butko, "Detection and positioning of overlapped sounds in a room environment," in *Proc. of Interspeech 2012*, Portland, USA, 2012.

[15] R. Takashima, T. Taniguchi, Y. Ariki, "Estimation of talker's head orientation based on discrimination of the shape of cross-power spectrum phase coefficients," in *Proc. of Interspeech 2012,* 2012.