Acoustic Multiple Object Positioning System

Viet-Hung Dang
Department of Computer Engineering, KyungHee University, Gyeonggi-do, 446-701, Korea.
dangviethung@oslab.khu.ac.kr

Thuong Le-Tien
HoChiMinh City University of Technology, National University, 268 Ly Thuong Kiet Str., HoChiMinh City, Vietnam
thuongle@hcmut.edu.vn

Young-Koo Lee
Department of Computer Engineering, KyungHee University, Gyeonggi-do, 446-701, Korea.
yklee@oslab.khu.ac.kr

SungYoung Lee
Department of Computer Engineering, KyungHee University, Gyeonggi-do, 446-701, Korea.
sylee@oslab.khu.ac.kr

ABSTRACT

Tracking techniques for unwelcome objects in wireless sensor networks (WSNs) for outdoor setting are based mainly on acoustic signals. However, the research on multi-object acoustic positioning so far has not been much developed in WSNs. The reasons are the communication and computation cost to solve the problems of recorded convolved mixture signals for the time delay difference from each object to each sensor. To overcome those problems, we introduce a new method for acoustic multi-object tracking in which the time delay difference is not paid attention but the relative information of magnitudes recorded at different sensors is. The sensors do pre-processing on the sensed data to get the most important information before compressing and sending it to the base. At the base, the data is uncompressed and analyzed with an independent component analysis (ICA) to get the relative information of magnitudes. Then an optimization method is used to infer the locations of these objects. Analysis and simulation results lead to the conclusion that our method gives good accuracy with a distributed computing manner and does not need much communication.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Applications

General Terms

Design

Keywords

Multiple Object Positioning, ICA, Acoustic, WSNs

1. INTRODUCTION

There are many previous works which use the data collected from arrays of acoustic sensors as input. The methods of this kind are mainly based on finding the relative angles between the sound sources and the receiving sensor arrays, called angles of arrival (AOAs). However, the number of methods for multiple object tracking is limited with this scheme of AOA information, not to mention the one with the scheme of non-array microphone sensors. For techniques that use AOA scheme, most of acoustic approaches so far can only give solutions to one tracked object [7],[8],[11] and just few are for multi-object tracking [5]. For techniques solving multi-object tracking with non-array sensors, the idea of using independent component analysis (ICA) comes naturally. The reasons are: the recorded data is the mixtures of sources and ICA is a powerful algorithm to separate and restore the original source data as long as these sources are statistically independent. Unfortunately, the standard ICA works only when the observations are instantaneous linear mixtures of the sources [2] or no delay time in the mixtures. In practice, an acoustic signal takes different time delays to reach the sensors, generating the observed convolved mixture data. Some methods have been developed to deal with this problem in time domain [1] and in frequency domain [6]. The ones for time domain are limited and the computation cost is too high [11],[9],[4]. Meanwhile, the methods for frequency domain are still complicated where the finite impulse response (FIR) Linear algebra is used and the standard ICA is applied on the complex domain [3]. In addition, all the related techniques so far generally need a centralized computer at the base, to solve the problem. Therefore the communication load is too big that it is hard, if not impossible, to apply to wireless sensor networks (WSNs) when all the recorded data is sent to this base.

In this paper, we describe a method solving the problem of location tracking for multiple objects that emit acoustic signals. The data this approach deals with is the convolved mixture data where sensor arrays are not used or the AOA information is not available. We do not extract the information of time-delay differences out of the mixtures but the information of received signal strength (RSS) ratios of each object to different sensors. We emphasize that there is no previous work solving the problem by extracting the information of RSS ratios for multi-object tracking from the delay mixture data. Our method also has many advantages that make the implementation into WSNs more practical when the whole computation load is shared on the sensors and the communication cost of collecting data for ICA computing is low.

2. PROBLEM STATEMENT

Assume that there are M objects emitting continuous acoustic signals. These signals can be considered as the components $s_j(t), j = 1,..,M$ of the source vector $\mathbf{s}(t)$ and are propagated to N different sensors, $M < N$. Then at each sensor i, the received data set is denoted as a column vector $x_i(t)$, which is also considered as the i-th component of the observed vector $\mathbf{x}(t)$. Assume that the distances from objects to sensors are not very far and sources' velocities are low, then different frequency components have the same attenuation property over distance and Doppler effect can be ignored. We do not need the sophisticated model data having the FIR combinations like in [1],[8],[9],[10]. The data received at each sensor is the actual signal with continuous values of delay, similar to the model in [7]:

$$x_i(t) = \sum_{j=1}^{M} a_{ij} s_j(t - \tau_{ij}),\ i = 1,..,N \tag{1}$$

where a_{ij} is a scalar presenting the magnitude of the signal from source j measured at sensor i and τ_{ij} is the propagation time of this signal. It should be noted that in this paper, any 1-dimensional vector with respect to t or ω is denoted as a column vector. Without any prior knowledge about the sources except for the information that any delayed version of one source is statistically independent of any delayed version of another, we have to detect the locations of all sources.

As mentioned above, previous works for object tracking focus on either the discrepancies between different τ_{ij} or the angles of arrival (AOA). These features are actually the direct measurements for location estimating and generally can give good results. However, τ_{ij} is hard to be extracted and need a big load of communication to send all the data to the base where the algorithm is computed. For AOA information, the deployed sensors must be in arrays, and still, it is hard to indicate the angles when there are multiple objects. In this work, we mainly concern about the relative information between different magnitudes of a source signal at different sensors.

3. PROPOSED METHOD FOR MULTI OBJECT POSITIONING

3.1 Distance information

Note that if there is only one active source and the others are inactive or emitting no sound, applying Short Time Fourier Transformation (STFT) to the signal at this source and to the signal received at each sensor, we get the same magnitude spectrum image. The difference is the scalar coefficient and the phase spectrum image. Obviously, the time-delay τ_{ij} only affects the phase spectrum image:

$$X_i(\omega) = \sum_{j=1}^{M} a_{ij} S_j(\omega) e^{-2\pi\tau_{ij}},\ i = 1,..,N, \tag{2}$$

$$|X_i(\omega)| = \sum_{j=1}^{M} |a_{ij}|\, |S_j(\omega)|,\ i = 1,..,N. \tag{3}$$

where $X_i(\omega)$ and $S_j(\omega)$ are the STFT transformation results of $x_i(t)$ and $s_j(t)$ alternatively. Evidently, the magnitude spectrum data has the form of instantaneous mixtures. In addition, the sources' independence property leads to the fact that the magnitude spectra of different $|S_j(\omega)|$ are statistically independent of each other. In other words, every $|S_j(\omega)|$ is independent of $|S_k(\omega)|$, $j \neq k$. As a result, $|S_j(\omega)|$ in (3) or the image of magnitude spectrum of $s_i(t)$ can be restored by a standard ICA. It should be noted that ICA cannot restore the magnitudes of the original independent components (ICs). Instead of giving exact $|a_{ij}|\,|S_j(\omega)|$, it gives the result $b_j\,|S_j(\omega)|$, $b_j \in R$. However, we can infer the mutual information of energy of each IC observed by the sensors because every $|X_{ij}|$ is composed of independent components which are considered as the orthogonal vectors. Thus the inner product of each IC vector $|b_j|\,|S_j(\omega)|$ and each magnitude spectrum image $|X_i(\omega)|$ contains the information of energy of this IC observed by sensor i.

$$(|b_j|\,|S_j(\omega)|)^T\,|X_i(\omega)| = |b_j|\,|S_j(\omega)|^T \sum_{k=1}^{M} |a_{ik}|\,|S_k(\omega)| \tag{4}$$

or

$$(|b_j|\,|S_j(\omega)|)^T\,|X_i(\omega)| = |b_j a_{ij}|\,|S_j(\omega)|^T\,|S_j(\omega)|. \tag{5}$$

Therefore, for each IC j, for each pair of magnitude spectrum of observed data i and l, the ratio $\left|\frac{a_{ij}}{a_{lj}}\right|$ can be achieved.

$$\left|\frac{a_{ij}}{a_{lj}}\right| = \frac{(|b_j|\,|S_j(\omega)|)^T\,|X_i(\omega)|}{(|b_j|\,|S_j(\omega)|)^T\,|X_l(\omega)|}. \tag{6}$$

Meanwhile,

$$\left|\frac{a_{ij}}{a_{lj}}\right|^2 = \frac{a_{ij}^2\,|s_j(t)|^T\,|s_j(t)|}{a_{lj}^2\,|s_j(t)|^T\,|s_j(t)|} = \frac{E_{ij}}{E_{lj}}, \tag{7}$$

where E_{ij} is the energy sent by source j and received by sensor i. Due to the inverse square law, if sound propagation distance is not very far, then the absorption of gas molecules is insignificant and the energy of sound decreases proportional to the inverse square of distance. In other words,

$$\frac{E_{ij}}{E_{lj}} = \frac{\left(1/d_{ij}\right)^2}{\left(1/d_{lj}\right)^2}. \tag{8}$$

From (6), (7) and (8), we have the relations of all pairs of distances from any tracked object j to sensors:

$$r_{ilj} = \left|\frac{a_{ij}}{a_{lj}}\right| = \frac{d_{lj}}{d_{ij}} = \frac{(|b_j|\,|S_j(\omega)|)^T\,|X_i(\omega)|}{(|b_j|\,|S_j(\omega)|)^T\,|X_l(\omega)|} \tag{9}$$

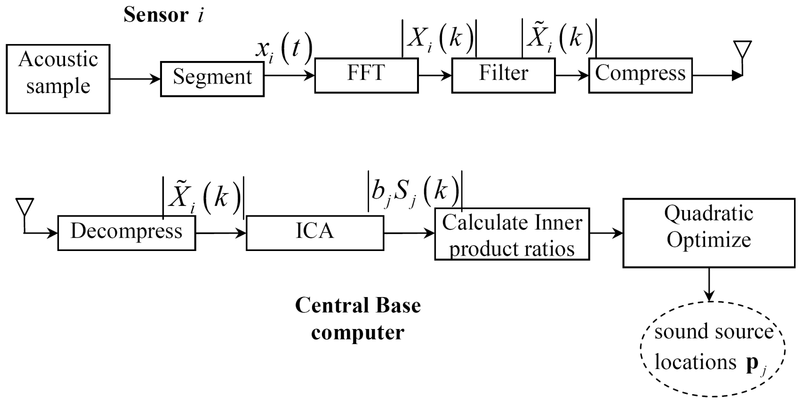

Figure 1: The proposed system with the Sensor's architecture and the Base's architecture.

Based on these relations, the locations of all of the sources will be inferred.

3.2 Multi-object tracking system architecture

Our proposed system for the method is depicted in Figure 1 according to the key idea that has been described in the previous Subsection on distance information. The upper is the Acoustic Sensor' work flow and the lower is the Central Base computer's work flow of the method. The more details of the method are mentioned in followings.

Let us consider a sound signal which is recorded by sensor i in segments of time. Since we can not infer the continuous parametric form for the results of STFT, we use the popular Fast Fourier Transformation (FFT) for these segments. The output of FFT block of Sensor i is the magnitude spectrum image in discrete form $|X_i(k)|$ instead of $|X_i(\omega)|$. The input data from the recorders always includes additional Gaussian noise which has the spectra spreading over the frequency domain. We can detect this level and eliminate out the Gaussian noise. The data of one segment after being filtered is $\left|\tilde{X}_i(\omega)\right|$, which is the magnitude spectrum image with several dominant values presenting the frequency components of the received data. Most of the rest values can be replaced by zeros if they are not greater than the noise level. Note that we may only need one frequency component for positioning a source. Therefore, compressing data is considerably efficient because missing dominant frequency components can be permitted and only few in stead of thousands values are sent to the Base for each time segment. Moreover, only half of the magnitude spectrum image length is needed due to the symmetric property of the magnitude spectrum image (recorded data is real). This is one of the key ideas of the method in order for the communication cost to be significantly reduced. That makes the system practical to be implemented in WSNs.

At the Central Base, the flow is straightforward and consistent with what we analyzed in the previous Subsection. The iteration fastICA that maximizes the nongaussianity [2] is used to obtain all the ICs. After that, at the "Quadratic Optimize" stage, all the ratios of different pairs of distances from a source to different sensors are available. Each ratio defines a constraint or a curve that this source belongs to. Usually, the curves are circles; except for the case $r_{ilj} = 1$, when the curve becomes the line orthogonal to the line segment connecting sensor i and sensor i at the segment's central point. Because of the fact that additional noise always exists in the recorded data and cannot be completely filtered, the errors in the constraints are not avoidable. Moreover, when a source is very close to one sensor and far from the others, its signal can be dominated by noise at the faraway sensors and then ICA would give high errors of separation. Therefore, we need a solution for the location of source j, a vector $\mathbf{p}_j$, $\mathbf{p}_j \in \mathbf{R}^2$, that compromises these constraints.

We propose an object function in quadratic form so that the function is convex and the gradient descent method can be used to guarantee the convergence. Negative gradient technique is used here to achieve the final estimations of source locations via iterations.

$$F_j = \sum_{i}^{N} \sum_{l,l \neq i}^{N-1} (d_{ij} - r_{ilj} d_{lj})^2 \tag{10}$$

and the solution for source j will be

$$\mathbf{p}_j = \arg\min_{\mathbf{p}_j} F_j \tag{11}$$

One remarkable note is that since the method uses the mutual information between every pair of active sensors, when the sensor number is greater than 3, the constraint number becomes much larger than the sensor number. For example, if the sensor number is 5, then the constraint number is 10. The more constraints we get, the more accurate are the results.

4. EXPERIMENT RESULTS

We conduct one main simulation set by building up the deployed setting area within ([0m,6m]x[0m,6m]). Based on characteristics of the real sound sources, we generate four simulated sources that imitate the sounds of vehicles and motors for this simulation set. These continuous sources have equivalent power levels. Four sensors are deployed

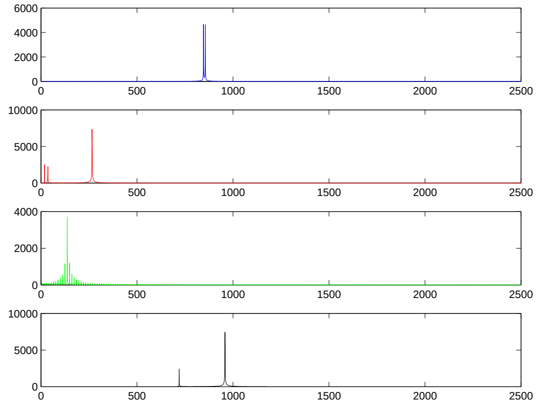

Figure 2: The original spectra images of four sources.

around the corners while the sources are set randomly within the deployed area (see Figure 6). The energies of the sources decrease with the inverse square law to the sensors and with the sound speed of $c = 343m/s$. The sampling frequency is $F_s = 16.384KHz$ and the time segment length is 0.3 seconds. We consider the noise to be caused by the background setting and the characteristics of the microphones. Thus the noise level is the same at all sensors. In order to analyze the system's performance, we increase the magnitude of additional Gaussian noise linearly and choose two parameters of signal over noise ratios for monitoring:

$$SNR_1 = \frac{P_{min}}{P_{noise}} \tag{12}$$

and

$$SNR_2 = \frac{P_{mean}}{P_{noise}}. \tag{13}$$

where P_{min} is the minimum power among the individual source power values received at the sensors and P_{mean} is the mean value of the combination power received at sensors. The reason we use these two values is that ICA algorithm depends on the linear combinations, so it is sensitive to noise which makes the linear combinations become nonlinear combinations. These parameters are the relations between RSS and the noise level and they can emphasize how well the system works with high noise level.

Figure 2 includes the spectrum images of individual sources while Figure 3 shows a typical set of linear combinations with the highest level noise in the simulation set ($SNR_1 = 0.07$ and $SNR_2 = 0.97$). The result of de-noise task on the frequency domain is illustrated in Figure 4. This result is actually obtained when the de-noise task is performed twice. At the sensors, the Gaussian noise level is determined and then is eliminated for the first time. The data then is cut half due to symmetry and compressed. At the Base, after being decompressed, these images are checked again and all the components on the frequency domain that do not appear in all four images are also eliminated for the second time. Figure 5 shows the result where all four sources are separated "successfully". In fact, we can never reconstruct the original spectrum images if the noise exists. Therefore, the separation is considered to be "successful" if an individual spectrum image can hold 75% energy likelihood in compari-

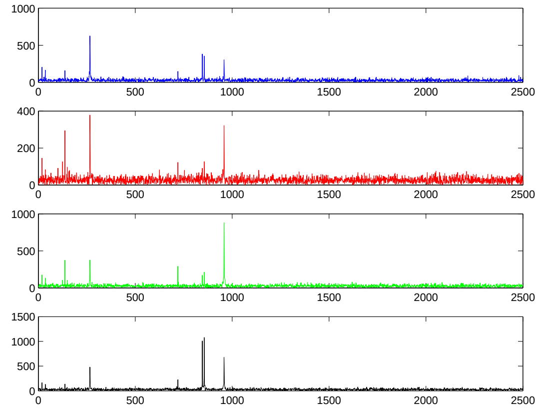

Figure 3: The spectra of received signals at four sensors where there exists the highest noise level in the simulation set, $SNR_1 = 0.07$.

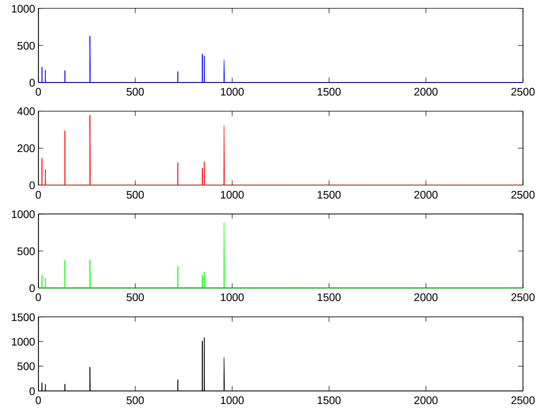

Figure 4: The spectra of received signals at four sensors where noise, with the highest level in the simulation set or $SNR_1 = 0.07$, is eliminated before ICA is used.

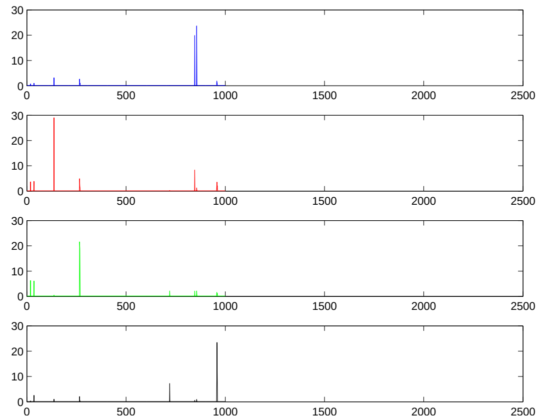

Figure 5: The result after using ICA on the magnitude of the frequency images or the independent components (ICs), in the case $SNR_1 = 0.07$.

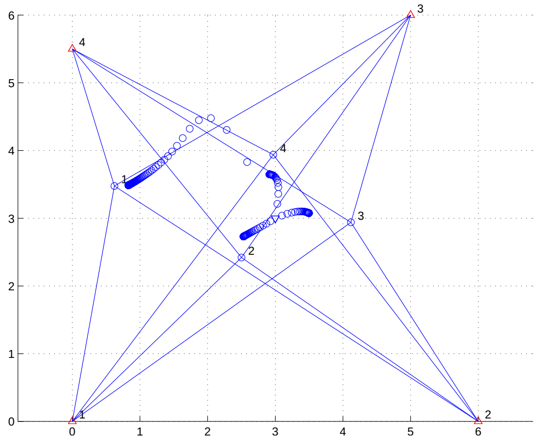

Figure 6: Location estimation results when minimization is used over iterations after a successful separation. The estimated locations reach closer to the actual locations. The level of noise in this case is the highest level used in the simulation, $SNR_1 = 0.07$.

son to the original one. In other words, the maximum value of the cross convolution between normalized original image and normalized separated image is larger than 0.75.

Finally, Figure 6 presents the process of solving the minimization for an example with $SNR_1 = 0.07$, highest noise level of this simulation set. The triangular positions are of the sensors, the circle positions marked with numbers are of the objects, and the strings of blue circle dots represent the updating processes for finding final estimations. The negative gradient algorithm is used to find the best position for each source so that the objective functions F_j in (10) are minimized. The initial guess for each source position's estimation is the same at the point (3m,3m).

The simulation set is done with 2000 trials for each level of noise. For each trial, the source locations are generated uniformly random within the deployed area. Table 1 shows the percentage of successful separation. Meanwhile, Table 2 shows the root mean square (RMS) errors for 4 cases of "successful" separation at each level of noise. With higher noise level, the percentage of "successfully" separation for all four sources decreases, but the percentage for other cases increases especially for the case of separating three sources "successfully". As can be seen, although the desired situation, or when all 4 sources are separated, gives good accuracy, it does not have very high percentage of occurrence. However, we can see that the sum of percentage for successfully separating 3 sources and 4 sources has the percentage from 88% to 91% at most of tested levels of noise. Even when the $SNR_2 < 1$, this sum is 86%. In reality, when the new data comes, the separation is performed again and we can use tracking technique to filter out the wrong detected locations to improve the accuracy. It can be observed that the accuracy decreases slowly with respect to the increasing noise. Obviously, once an IC is restored, if the separation quality is good enough, then the position of this IC can be inferred with high accuracy. The accuracy result ranging from 0.6m to 0.7m for the multiple source location tracking

Table 1: Percentage of successful separation

SNR_1	4.97	1.17	0.44	0.23	0.12	0.07
SNR_2	64.72	15.19	5.73	2.96	1.54	0.97
1 IC	1%	1%	2 %	3%	3%	3%
2 ICs	7%	7%	8%	8%	9%	10%
3 ICs	28%	29%	30%	33%	35%	36%
4 ICs	63%	62%	60%	56%	53%	50%

Table 2: RMS error after separation

SNR_1	4.97	1.17	0.44	0.23	0.12	0.07
SNR_2	64.72	15.19	5.73	2.96	1.54	0.97
1 IC	0.64m	0.66m	0.71m	0.70m	0.69m	0.72m
2 ICs	0.68m	0.71m	0.70m	0.69m	0.69m	0.71m
3 ICs	0.69m	0.71m	0.70m	0.71m	0.70m	0.71m
4 ICs	0.60m	0.65m	0.71m	0.71m	0.72m	0.73m

is considered to be good.

5. CONCLUSIONS

This paper proposed a system design for acoustic source localization in which the information for separation is the ratios of energy values of original sources received at the sensors. In order to obtain these ratios, ICA technique is used to separate the magnitude spectra of received signals. A new quadratic function has been proposed for the task of inferring the source locations. The result at this preliminary research shows that the method gives high accuracy and needs a very low communication cost for big spectrum image data. It is practical for implementation because the whole computation load is shared and calculated in a distributed manner. The sensors for this design must have a powerful ability of computing to perform FFT on a long segment of data in short time. In other words, the proposed system is the acoustic source localization design for the future generation of WSNs. Our research on improvement of the accuracy is being in process whose more analysis details and results will be presented in another paper.

6. ACKNOWLEDGMENTS

This research was supported by the MKE (Ministry of Knowledge and Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2010-(C1090-0902-0002))

7. REFERENCES

[1] R. T. A. Ciaramella. Separation of convolved mixtures in frequency domain ica. *International Mathematical Forum*, 16:769–795, 2006.

[2] E. O. Aapo Hyvarinen. Independent component analysis: a tutorial. *Neural Networks*, 13:4–5, 2000.

[3] E. BINGHAM and A. HYVARINEN. A fast fixed point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, 10:1–8, 2000.

[4] R. C. W. Darren B. Ward. Particle filter beamforming for acoustic source localization in a reverberant environment. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2002.

[5] D. W. Nicoleta Roman. Binaural tracking of multiple moving sources. *IEEE Transaction on Audio, Speech, and Language Processing*, 16:728–739, 2008.

[6] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. pages 22:21–34, 1998.

[7] M. Stanacevic. Micropower gradient flow acoustic localizer. *IEEE Transactions on Circuits and systems*, 52:2148–2157, 2005.

[8] D. K. G. Stanley T.Birchfield. Ieee international conference on acoustics, speech, and signal processing. In *IEEE International Conference*. IEEE Computer Society, 2001.

[9] D. K. G. Stanley T.Birchfield. Fast bayesian acoustic localization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2002.

[10] R. O. T. S. Te-won Lee, Andreas Ziehe. Combining time-delayed decorrelation and ica: Towards solving the cocktail party problem. In *In Proc. ICASSP98*, pages 1249–1252, 1998.

[11] H. M. Yoki Sasaki, Satoshi Kagami. Multiple sound source mapping for a mobile robot by self-motion triangulation. In *International Conference on Intelligent Robots and Systems*, pages 380–385. IEEE, Oct 2006.