BLIND ALIGNMENT OF ASYNCHRONOUSLY RECORDED SIGNALS FOR DISTRIBUTED MICROPHONE ARRAY

Nobutaka Ono, Hitoshi Kohno, Nobutaka Ito, and Shigeki Sagayama

Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
{onono, h-kohno, ito, sagayama}@hil.t.u-tokyo.ac.jp

ABSTRACT

In this paper, aiming to utilize independent recording devices as a distributed microphone array, we present a novel method for alignment of recorded signals with localizing microphones and sources. Unlike conventional microphone array, signals recorded by independent devices have different origins of time, and microphone positions are generally unknown. In order to estimate both of them from only recorded signals, time differences between channels for each source are detected, which still include the differences of time origins, and an objective function defined by their square errors is minimized. For that, simple iterative update rules are derived through auxiliary function approach. The validity of our approach is evaluated by simulative experiment.

Index Terms— blind alignment, source localization, time delay, cross correlation

1. INTRODUCTION

Microphone array techniques are much useful for localizing sound sources and separating mixture of sounds and they have greatly developed in several decades. One of the significant factors for the performance is the number of microphones. Even if using a simple delay and sum beamforming, many microphones with large aperture yield acute directivity. In applying independent component analysis, more microphones than sources makes the problem overdetermined, which is easily solved rather than an underdetermined case.

However, in realistic applications, the use of many microphones are not always feasible. In conventional array signal processing, it is supposed that received signals have the same time origins because time delays between channels are significant cues for localizing and separating sound sources. For satisfying the condition, in conventional microphone array system, microphones have to be connected to multiple A/D converters controlled by a common temporal clock.

Our aim is to organize independent audio recording devices as a wireless and distributed microphone array. The concept of distributed microphone array has been recently discussed in several literatures [1, 2, 3, 4, 5]. In utilizing independent devices, even if mismatch of their sampling frequencies is negligible, the origins of time are generally much different. Therefore, synchronization among them is a significant issue and it is one of the general problems in sensor networks [6]. In this paper, the problem of estimating the time origins in a blind manner is discussed, which means the estimation is performed by only recorded signals and any other kinds of temporal information such as RF (radio frequency) channels are not used. Our method is based on the consistency between unknown time origins, positions of microphones and sources, and observed time differences among channels for each source. The problem can be considered as an extension of TDOA (Time Difference of Arrival)-based source localization [7], and simultaneous localization of microphones and sources [8]. An objective function defined by square errors of time differences is considered and unknown parameters are estimated by minimizing it. Through auxiliary function approach, simple iterative update rules are derived. We show the feasibility by simulative experiment.

2. FORMULATION

2.1. Fundamental Equations

Suppose that K sound sources are observed by L microphones. Let $\boldsymbol{s}_i = (x_i\ y_i\ z_i)^t$ $(1 \le i \le K)$ and $\boldsymbol{r}_n = (u_n\ v_n\ w_n)^t$ $(1 \le n \le L)$ be positions of sound sources and microphones, respectively, where ^{t} represents transpose. Let t_n be the time origin of the observed signal recorded by the nth microphone. As the first step of the formulation of the blind alignment problem, we assume that clocks on microphones are accurate in this paper and the mismatch of sampling frequencies will be discussed in the next work.

In localization of sources and microphones, one of the most significant cues is the time difference between observed signals. When the microphone m and n receive only ith source in a certain time interval, the time delay of the mth observed signal to the nth observed signal can be represented by

$$\begin{aligned}\tau_{imn} &= \left(\frac{|\boldsymbol{s}_i - \boldsymbol{r}_m|}{c} - t_m\right) - \left(\frac{|\boldsymbol{s}_i - \boldsymbol{r}_n|}{c} - t_n\right) \\ &= \frac{|\boldsymbol{s}_i - \boldsymbol{r}_m| - |\boldsymbol{s}_i - \boldsymbol{r}_n|}{c} - (t_m - t_n), \qquad (1)\end{aligned}$$

where the fist term of eq. (1) represents the difference of arrival, and the second term represents the difference of the time origin.

2.2. Necessary Observations

In eq. (1), τ_{imn} are obtained from any pairs of observed signals and $L-1$ of them are ideally independent variables for each source. While, unknown variables are t_n, $\boldsymbol{s}_i = (x_i\ y_i\ z_i)^t$, and $\boldsymbol{r}_n = (u_n\ v_n\ w_n)^t$ for $1 \le i \le K$ and $1 \le n \le L$. Note that all of unknown variables are relative. It means that an absolute time origin and six degrees of freedom of position variables caused by transformation (three degrees) and rotation (three degrees) are not determined in this framework. Then, to determine unknown variables,

$$K(L-1) \ge (L-1) + 3(K+L) - 6 \qquad (2)$$

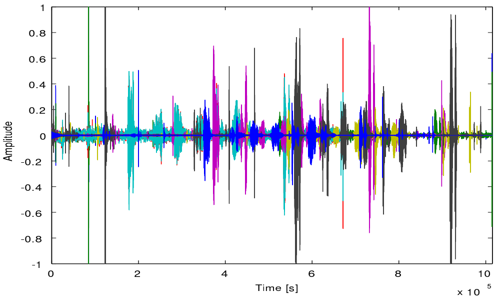

Figure 1: Asynchronously recorded signals

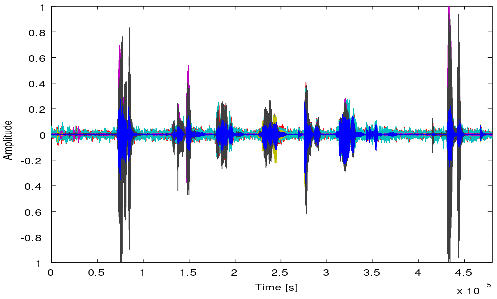

Figure 2: Coarsely-synchronized signals

should be satisfied at least. Rearranging eq. (10), we obtain

$$(K-4)(L-4) \quad \geq \quad 9, \tag{3}$$

as a necessary condition. Note that the number of sources K are easily increased when a moving sound source is observed because it has different positions in different time frames.

2.3. Coarse-fine Alignment Scheme

On asynchronous recorded signals, acquiring the information of τ_{imn} is not a direct problem. Even if source signals are sparse and they are not overlapped, it is much difficult to find correspondences of acoustic events generated by sources on each channel shown in Fig. 1. Furthermore, real observations include silent intervals and overlaps of source signals. Therefore, autonomously detecting single-source frames is essential. For that, we propose coarse-fine alignment scheme in the following.

1. Select one channel as a reference.
2. Calculate the cross correlation between each channel and the reference using the whole signal length and obtain the time difference from the maximum.
3. Align each channel using the time difference (as shown in Fig. 2). Note that it is just coarse alignment and fine mismatches of time origins between channels remains.
4. The frame analysis is applied and the normalized cross correlation is calculated frame by frame.
5. Only frames with a larger maximum than a threshold in the normalized cross correlation are selected as single-source frames and time differences between any channels are obtained from the frames. Each single-source frame is handled as different source frame.

3. DERIVATION OF UPDATE RULES

3.1. Objective Function

In order to find unknown parameters: $\mathbf{\Theta} = \{\boldsymbol{s}_i, \boldsymbol{r}_n, t_n | \ 1 \leq i \leq K, 1 \leq n \leq L\}$, the square errors of eq. (1):

$$\begin{aligned} J(\mathbf{\Theta}) &= \frac{1}{c^2KL^2}\sum_{i=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}\varepsilon_{imn}^2 && (4)\\ \varepsilon_{imn} &= |\boldsymbol{s}_i-\boldsymbol{r}_m|-|\boldsymbol{s}_i-\boldsymbol{r}_n|-c(\tau_{imn}+t_m-t_n)\\ &= \sqrt{(x_i-u_m)^2+(y_i-v_m)^2+(z_i-w_m)^2}\\ &\quad -\sqrt{(x_i-u_n)^2+(y_i-v_n)^2+(z_i-w_n)^2}\\ &\quad -c(\tau_{imn}+t_m-t_n) && (5)\end{aligned}$$

is considered as an objective function to be minimized.

3.2. Auxiliary Function Approach

The objective function $J(\mathbf{\Theta})$ includes the square root of x_i, y_i, z_i, etc, or, the absolute function of the vectors $\boldsymbol{s}_i$, $\boldsymbol{r}_m$, etc. Minimizing it is a kind of nonlinear optimization problem and we could apply a general optimization method such as the gradient descent method or the quasi-Newton method. But here, for deriving simple and efficient iterative solution, applying auxiliary function approach is considered. The auxiliary function approach is an extension of well-known EM algorithm and has been recently applied to solve several kind of nonlinear optimization problems in the signal processing field [9, 10, 11].

In order to find parameters $\boldsymbol{\theta}$ to minimize an objective function $J(\boldsymbol{\theta})$, an auxiliary function $Q(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$ to satisfy

$$J(\boldsymbol{\theta}) = \min_{\bar{\boldsymbol{\theta}}} Q(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \tag{6}$$

is utilized, where $\bar{\boldsymbol{\theta}}$ is called auxiliary variables. The principle of the auxiliary function method is based on the fact that $J(\boldsymbol{\theta})$ is non-increasing under the updates:

$$\begin{aligned} \bar{\boldsymbol{\theta}}^{(l+1)} &= \text{argmin}_{\bar{\boldsymbol{\theta}}}\ Q(\boldsymbol{\theta}^{(l)}, \bar{\boldsymbol{\theta}}), && (7)\\ \boldsymbol{\theta}^{(l+1)} &= \text{argmin}_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}^{(l+1)}), && (8)\end{aligned}$$

where l is the index of iterations. The brief proof is given by the following.

1. $Q(\boldsymbol{\theta}^{(l)}, \bar{\boldsymbol{\theta}}^{(l+1)}) = J(\boldsymbol{\theta}^{(l)})$ from eq. (6) and eq. (7),
2. $Q(\boldsymbol{\theta}^{(l+1)}, \bar{\boldsymbol{\theta}}^{(l+1)}) \leq Q(\boldsymbol{\theta}^{(l)}, \bar{\boldsymbol{\theta}}^{(l+1)})$ from eq. (8),
3. $J(\boldsymbol{\theta}^{(l+1)}) \leq Q(\boldsymbol{\theta}^{(l+1)}, \bar{\boldsymbol{\theta}}^{(l+1)})$ from eq. (6),

then,

$$J(\boldsymbol{\theta}^{(l+1)}) \leq J(\boldsymbol{\theta}^{(l)}), \tag{9}$$

which guarantees non-increasing of objective functions. For efficient updates, an auxiliary function $Q(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$ such that eq. (7) and eq. (8) are given by closed forms is desired.

3.3. Design of Two-step Auxiliary Functions

We derive auxiliary functions of eq. (4) in two steps. One of the difficulties to find a minimum of eq. (4) lies on the fact that ε_{imn} consists of two terms with different indexes of m and n. Due to it, $\partial J/\partial \boldsymbol{r}_m = 0$ includes not only the variable of $\boldsymbol{r}_m$ but $\boldsymbol{r}_n (1 \leq n \leq L)$, which leads to simultaneous equations. For separating variables, we consider the following lemma.

Lemma 1 *For any A_1, A_2 and B, an inequality:*

$$(A_1 + A_2 - B)^2 \leq 2(A_1 - a_1)^2 + 2(A_2 - a_2)^2 \qquad (10)$$

is hold under $a_1 + a_2 = B$. The equality is satisfied if and only if

$$\begin{aligned} a_1 &= A_1 - \frac{1}{2}(A_1 + A_2 - B), & (11)\\ a_2 &= A_2 - \frac{1}{2}(A_1 + A_2 - B). & (12) \end{aligned}$$

Proof: Let $f(a_1, a_2)$ be

$$\begin{aligned} f(a_1, a_2) &= 2(A_1 - a_1)^2 + 2(A_2 - a_2)^2 - (A_1 + A_2 - B)^2 \\ &\quad + 4\lambda(a_1 + a_2 - B), \qquad (13) \end{aligned}$$

where λ represents a Lagrange multiplier for the constraint $a_1 + a_2 = B$. Differentiating $f(a_1, a_2)$ by a_1, a_2, and λ, and letting them be zero yield

$$\begin{aligned} A_1 - a_1 - \lambda &= 0, & (14)\\ A_2 - a_2 - \lambda &= 0, & (15)\\ a_1 + a_2 - B &= 0. & (16) \end{aligned}$$

By solving them, it is clear that $f(a_1, a_2)$ takes the minimum value in eq. (11) and eq. (12), and it is equal to zero. ■

By letting

$$\begin{aligned} A_1 &= |\boldsymbol{s}_i - \boldsymbol{r}_m|, & (17)\\ A_2 &= -|\boldsymbol{s}_i - \boldsymbol{r}_n|, & (18)\\ B &= c(\tau_{imn} + t_m - t_n), & (19) \end{aligned}$$

we obtain an auxiliary function:

$$\begin{aligned} J_1(\boldsymbol{\Theta}, \boldsymbol{\mu}) &= \frac{2}{c^2KL^2}\sum_{i=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}\{(|\boldsymbol{s}_i - \boldsymbol{r}_m| - \mu_{imn}^{m})^2 \\ &\quad + (|\boldsymbol{s}_i - \boldsymbol{r}_n| - \mu_{imn}^{n})^2\} \qquad (20) \end{aligned}$$

satisfying $J(\boldsymbol{\Theta}) = \min_{\boldsymbol{\mu}} J_1(\boldsymbol{\Theta}, \boldsymbol{\mu})$ where $\boldsymbol{\mu} = \{\mu_{imn}^{m}, \mu_{imn}^{n}\}$ are auxiliary variables. $J(\boldsymbol{\Theta}) = J_1(\boldsymbol{\Theta}, \boldsymbol{\mu})$ if and only if

$$\begin{aligned} \mu_{imn}^{m} &= |\boldsymbol{s}_i - \boldsymbol{r}_m| - \frac{1}{2}\varepsilon_{imn}, & (21)\\ \mu_{imn}^{n} &= |\boldsymbol{s}_i - \boldsymbol{r}_n| + \frac{1}{2}\varepsilon_{imn}. & (22) \end{aligned}$$

Although the microphone position vectors $\boldsymbol{r}_m$ and $\boldsymbol{r}_n$ $(1 \leq m, n \leq L)$ are separated into independent terms in the auxiliary function J_1, it still includes the absolute function of the vectors $\boldsymbol{s}_i - \boldsymbol{r}_m$ and $\boldsymbol{s}_i - \boldsymbol{r}_m$, which makes it difficult to solve $\partial J_1/\partial \boldsymbol{s} = 0$ and $\partial J_1/\partial \boldsymbol{r}_m = 0$. So, we consider to apply the auxiliary function approach, again. Note that the following lemma.

Lemma 2 *For any vector $\boldsymbol{x}$, any unit vector $\boldsymbol{e}$, and positive scalar a,*

$$(|\boldsymbol{x}| - a)^2 \leq |\boldsymbol{x} - a\boldsymbol{e}|^2 \qquad (23)$$

is hold. The equality is satisfied if and only if $\boldsymbol{e} = \boldsymbol{x}/|\boldsymbol{x}|$.

Proof:

$$\begin{aligned} & |\boldsymbol{x} - a\boldsymbol{e}|^2 - (|\boldsymbol{x}| - a)^2 \\ =\; & |\boldsymbol{x}|^2 - 2a\boldsymbol{x}\cdot\boldsymbol{e} + a^2|\boldsymbol{e}|^2 - |\boldsymbol{x}|^2 + 2a|\boldsymbol{x}| - a^2 \\ =\; & 2a(|\boldsymbol{x}| - \boldsymbol{x}\cdot\boldsymbol{e}) \\ =\; & 2a|\boldsymbol{x}|(1 - \cos\theta) \\ \geq\; & 0, \qquad (24) \end{aligned}$$

where θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{e}$. ■

By letting

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{s}_i - \boldsymbol{r}_m, & (25)\\ a &= \mu_{imn}^{m}, & (26) \end{aligned}$$

we have the second auxiliary function:

$$\begin{aligned} J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e}) &= \frac{2}{c^2KL^2}\sum_{i=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}\{(\boldsymbol{s}_i - \boldsymbol{r}_m - \boldsymbol{e}_{im}\mu_{imn}^{m})^2 \\ &\quad + (\boldsymbol{s}_i - \boldsymbol{r}_n - \boldsymbol{e}_{in}\mu_{imn}^{n})^2\} \qquad (27) \end{aligned}$$

satisfying $J_1(\boldsymbol{\Theta}, \mu) = \min_{\boldsymbol{e}} J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e})$ where $\boldsymbol{e} = \{\boldsymbol{e}_{im}\}$. $J_1(\boldsymbol{\Theta}, \mu) = J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e})$ if and only if

$$\begin{aligned} \boldsymbol{e}_{im} &= (\boldsymbol{s}_i - \boldsymbol{r}_m)/|\boldsymbol{s}_i - \boldsymbol{r}_m|, & (28)\\ \boldsymbol{e}_{in} &= (\boldsymbol{s}_i - \boldsymbol{r}_n)/|\boldsymbol{s}_i - \boldsymbol{r}_n|. & (29) \end{aligned}$$

Note that $\boldsymbol{r}_m$ minimizing J_2 is obtained in a closed form unlike J_1.

3.4. Derivation of Update Rules

Exploiting the two kinds of auxiliary functions, the objective function is monotonically decreased in the following iterative way.

Step 1: Update $\boldsymbol{\mu}$ to minimize $J_1(\boldsymbol{\Theta}, \boldsymbol{\mu})$. Then, $J_1(\boldsymbol{\Theta}, \boldsymbol{\mu})$ is equal to $J(\Theta)$.

Step 2: Update $\boldsymbol{e}$ to minimize $J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e})$. Then,

$$J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e}) = J_1(\boldsymbol{\Theta}, \boldsymbol{\mu}) = J(\boldsymbol{\Theta}) \qquad (30)$$

is satisfied.

Step 3: Update $\boldsymbol{\Theta}$ to minimize $J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e})$. Then, $J(\Theta)$ also decreases since generally

$$J_2(\boldsymbol{\Theta}, \boldsymbol{\mu}, \boldsymbol{e}) \geq J_1(\boldsymbol{\Theta}, \boldsymbol{\mu}) \geq J(\boldsymbol{\Theta}). \qquad (31)$$

Step 4: Return to Step 1.

The update rules for Θ is derived from solving $\partial J_2/\partial\Theta = 0$. The whole update rules are summarized as follows.

$$\begin{aligned} \varepsilon_{imn} &\leftarrow |\boldsymbol{s}_i - \boldsymbol{r}_m| - |\boldsymbol{s}_i - \boldsymbol{r}_n| - c(\tau_{imn} + t_m - t_n) & (32)\\ \mu_{imn}^{m} &\leftarrow |\boldsymbol{s}_i - \boldsymbol{r}_m| - \frac{1}{2}\varepsilon_{imn} & (33)\\ \mu_{imn}^{n} &\leftarrow |\boldsymbol{s}_i - \boldsymbol{r}_n| + \frac{1}{2}\varepsilon_{imn} & (34)\\ \boldsymbol{e}_{in} &\leftarrow (\boldsymbol{s}_i - \boldsymbol{r}_n)/|\boldsymbol{s}_i - \boldsymbol{r}_n| & (35)\\ \boldsymbol{s}_i &\leftarrow \frac{1}{L^2}\sum_{m=1}^{L}\left(L\boldsymbol{r}_m + \boldsymbol{e}_{im}\sum_{n=1}^{L}\mu_{imn}^{m}\right) & (36)\\ \boldsymbol{r}_n &\leftarrow \frac{1}{KL}\sum_{i=1}^{K}\left(L\boldsymbol{s}_i - \boldsymbol{e}_{in}\sum_{m=1}^{L}\mu_{inm}^{n}\right) & (37)\\ t_n &\leftarrow t_n + \frac{1}{cKL}\sum_{i=1}^{K}\left(L|\boldsymbol{s}_i - \boldsymbol{r}_n| - \sum_{m=1}^{L}\mu_{inm}^{n}\right) & (38) \end{aligned}$$

4. EXPERIMENTAL EVALUATION

In order to confirm that the feasibility of simultaneously estimating positions of microphones, sources, and time origins by minimizing eq. (4), a simulative experiment was performed. A room with $10 \times 10 \times 10\mathrm{m}^3$ was supposed and ideal spherical-wave propagation of acoustic wave was simulated. For satisfying the necessary condition of eq. (3), the number of microphones and sources were set to 9 and 8, respectively. Their positions were determined randomly. As source signals, real-recorded hand claps were used and each source was not overlapped each other. The sampling frequency was 44100Hz and the signal length was 5.0s. The differences of the time origins less than 1.0s were randomly given to observed signals. In this setup, the degree of freedom of unknown parameters is $(8-1)+3(9+8)-6=52$.

As discussed in section 2.3, the coarse alignments were first performed, and then, the frame analysis was applied, single-source frames were detected, and time differences for each source were obtained, where the frame length was 100ms. Finally, initial values were given randomly, and update rules were iteratively applied to them. The number of iterations were 60000. Fig. 3 shows that true, initial and finally-estimated positions of microphones and sources projected into the xy plane. Even though the initial positions were much different from the true positions and observed time differences included the differences of unknown time origins, it is confirmed that the estimation was well performed.

5. CONCLUSION AND FUTURE WORK

In this paper, a blind method for alignment of signals recorded by independent devices with estimation of microphones and sources are presented. Although the derived update rules guarantees monotonically decrease of the objective function, the local minimum problem still theoretically remains. Exploiting amplitude ratios, not only time differences between channels is a possible way to give a better initial positions of microphones and sources, which will also facilitate fast convergence. Applying the proposed framework to real-recorded signals is on-going.

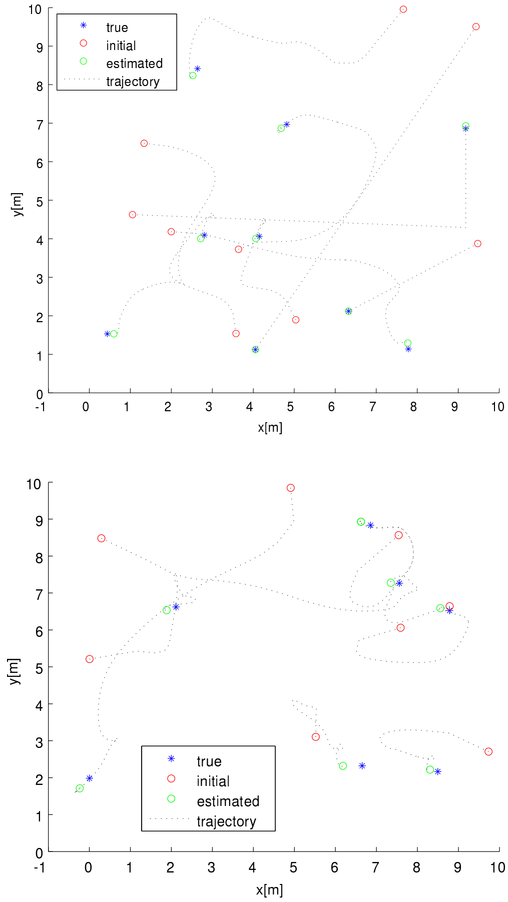

Figure 3: The estimation results of microphones (upper) and sources (lower) positions projected into the xy plane

6. REFERENCES

[1] R. Lienhart, I. Kozintsev, S. Wehr, and M. Yeung, "On the importance of exact synchronization for distributed audio processing," Proc. ICASSP, pp. 840-843, 2003.

[2] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," EURASIP Journal of Applied Signal Processing, vol. 2003, no. 4, pp. 338-347, 2003.

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

[4] Z. Liu, Z. Zhang, L. He, and P. Chou, "Energy-based sound source localization and gain normalization for ad hoc microphone arrays," Proc. ICASSP, pp. 761-764, 2007.

[5] E. Robledo-Arnuncio, T. S. Wada, and B. H. Juang, "On Dealing with Sampling Rate Mismatches in Blind Source Separation and Acoustic Echo Cancellation," Proc. WASPAA, pp. 34-37, 2007.

[6] S. Ando and N. Ono, "A Bayesian theory of cooperative calibration and synchronization in sensor networks," Trans. Soc. Instru. Control Engineers (SICE), vol. E-S-1, pp.21-26, 2005.

[7] P. Stoica and J. Li, "Source localization from range-difference measurements," IEEE. Signal Processing Mag., vol. 23, no. 69, pp. 63-65, Nov. 2006.

[8] K. Kobayashi, K. Furuya, A. Kataoka, "A blind source localization by using freely positioned microphones," Trans. IEICE, vol. J86-A, No.6, pp. 619-627, 2003. (in Japanese)

[9] D. D. Lee and H. S. Seung, "Algorithms for Non-Negative Matrix Factorization," Proc. NIPS, pp. 556–562, 2000.

[10] J. Le Roux, H. Kameoka, N. Ono, A. de Cheveigne, and S. Sagayama, "Single and Multiple F0 Contour Estimation Through Parametric Spectrogram Modeling of Speech in Noisy Environments," IEEE Trans. ASLP, vol. 15, no. 4, pp.1135-1145, May., 2007.

[11] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, "Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram," Proc. EUSIPCO, Aug., 2008.