# Multiple Sources' Direction Finding by using Reliable Component on Phase Difference Manifold and Kernel Density Estimator

K. Fujimoto†, N. Ding†, and N. Hamada††
†Signal processing Lab, School of Integrated Design Engineering, Keio University
Hiyoshi 3-14-1, Yokohama 223-8522 Japan
†fujimoto@hamada.sd.keio.ac.jp, †ding@hamada.sd.keio.ac.jp, ††hamada@sd.keio.ac.jp

***Abstract*—This paper proposes a novel direction-of-arrival estimation method in a general 3-dimensional array configuration for multiple speech signals uttered simultaneously. The method is based on sparseness in the time-frequency representation of speech signal and is applicable to an underdetermined case where the sources outnumber sensors. At first, we introduce a parameterized closed surface to which we refer the phase difference manifold. This is defined in the space of phase difference vectors between sensors in order to provide the one-to-one correspondence between the induced phase difference on this sphere and a propagating direction vector of the source. Instead using the conventional pseudo-inverse mapping algorithm, the selection of phase difference vectors located or closely located on the phase difference manifold as a set of reliable observations. Finally, the author's method utilizing kernel density algorithm is generalized for arbitrary array sensors case. The conducted experiments demonstrate that the method utilizing the reliable cell selection and the kernel density estimator with appropriate bandwidth determination performed effectively.**

## I. INTRODUCTION

The localization of sound sources is essential in the study of human-machine communication systems and it is widely used for a variety of applications. Numerous methods for estimating the direction-of-arrival (DOA) of sources using a microphone array have been studied extensively [1]. While different techniques such as MUSIC exist [2], most typical DOA estimation systems utilize the time-delay of arrival (TDOA) between different microphones. The generalized cross correlation phase transform method and its variance are well-known [3]. For multiple simultaneously-uttered sources even for the underdetermined case where sources outnumber sensors, the effectiveness of source separation schemes has been demonstrated. The separation-based methods are a variety of histogram techniques, clustering approaches such as k-means algorithm, and ICA-based approaches. The underlying DOA estimation problems addressed in this paper are summarized as follows. a) Array configuration of multiple sensors in 3-dimensional arrangement is arbitrary, b) Multiple speech sources uttered simultaneously, and c) Sources outnumber sensors.

This paper presents a new approach for estimating multiple speaker's DOAs based on the time-frequency property known as W-disjoint orthogonality and its variances [4]-[8]. The existing methods, such as [4],[7]-[9], cope with the problem under the same conditions a)-c) mentioned above. Huang et al.[4] introduced the method accumulating delays estimated at individual Short Time Fourier Transform (STFT) components, and searching the peak positions of combined delay histograms which are generated by a set of microphone pairs. The method described in [8] applies their DOA finding algorithm at individual components of the STFT domain to estimate a set of DOA vectors. It finally employs a fusion process to detect DOAs of speakers. Their closed form solution provided by the Moore-Penrose type pseudo-inverse of the over-determined linear relationship between the propagation direction vector and the delay vector is effectively utilized for multiple-sensor scenario. Araki, et al.[7] also employ the source sparseness assumption in the context of blind source separation, and apply $k$-means algorithm to cluster the normalized vectors in the STFT domain. They assume the centroid of each cluster provides the DOA of the source corresponding to the cluster and apply the pseudo-inverse operation as in [8]. In contrast to these studies including the method described in this paper, Nesta et al. [10] proposed a method utilizing the blind separation based on the independent component analysis. They use the ratio of the elements of the de-mixing matrices obtained through the source separation and employ these to accurate estimation.

The proposed DOA estimation method in this paper is based on the sparseness of speech signal representation of STFT for coping with multiple sources even in underdetermined condition. At the first part of this paper we introduce a parameterized sphere, which is referred to the phase difference manifold, defined in the space of phase difference vectors between sensors. The proposed manifold provides a one-to-one correspondence between the induced phase difference vector on this manifold and the propagation direction vector of the source. With the use of this unique relationship, we need not to use the conventional pseudo-inverse mapping. The process of the proposed DOA estimation algorithm is summarized by the following three steps. Step 1) From given phase difference observations, we select subset of observations which are closely located on the phase difference manifold. This is because these are considered to be reliable for DOA estimation. Step 2) Apply the established mapping from phase manifold to the propagation direction vector for the selected

phase difference vectors. Step 3) Generalized author's DOA estimation method based on kernel density estimation [11] is applied, and finally, the peaks of the estimated probability density function of DOA yield the consequence direction angles of the sources.

## II. ARRAY SYSTEM

Consider an array of $M$ sensors with a given geometry. The sensors are omni-directional and observe acoustic signals generated by far-field speech sources. Let $\mathbf{r}_m = [x_m, y_m, z_m]^T$ $(m = 1, \cdots, M)$ denote the location of the $m$-th sensor in 3-D space, and we assume the first sensor is located at the origin $(\mathbf{r}_1 = \mathbf{0})$ without loss of generality. The source direction vector referred to as the propagation direction vector is defined by

$$\mathbf{a}(\phi, \theta) = [sin\theta cos\phi, sin\theta sin\phi, cos\theta]^T \tag{1}$$

where, $\phi$ $(-\pi \leq \phi \leq \pi)$ and $\theta$ $(0 \leq \theta \leq \pi)$ denote the azimuth and elevation angles of the source, respectively. The source's propagating wave with traveling speed $c$ induces the TDOAs $\delta_m$ $(m = 1, \cdots, M)$ between $m$-th sensor and the reference. They are integrated in the following ($M$-1) dimensional vector.

$$\boldsymbol{\delta} := [\delta_2, \cdots, \delta_M]^T = -\frac{\mathbf{R}\mathbf{a}(\phi, \theta)}{c} \tag{2}$$

where,

$$\boldsymbol{\delta_m} = -\frac{\mathbf{r}_m^T \mathbf{a}(\phi, \theta)}{c}, \quad \mathbf{R} := [\mathbf{r}_2, \cdots, \mathbf{r}_M]^T$$

The TDOAs are estimated by using the phase difference or cross-phase between the discrete Fourier transform (DFT) of microphone signals.

Let $X_m(l)$ be the $L$-point DFT of $m$-th microphone signal and $l$ (0,$\cdots$,$L$/2) is the frequency bin index, then define the following phase difference($PD$) vector as a function of frequency index $l$.

$$\boldsymbol{\varphi}(l) = [\varphi_2(l), \cdots, \varphi_M(l)]^T \tag{3}$$

where

$$\varphi_m(l) = \angle X_m(l) - \angle X_1(l)$$

The TDOA $\delta_m$ can be estimated by the phase difference of observations as follow.

$$\hat{\boldsymbol{\delta}} := [\hat{\delta_2}, \cdots, \hat{\delta_M}]^T$$

$$\hat{\delta_m} = -\frac{1}{\Delta\omega l}\varphi_m(l) \tag{4}$$

where $\Delta\omega = \frac{2\pi f_s}{L}$($f_s$:sampling frequency) is the frequency interval between adjacent frequency points in the DFT domain.

As a consequence from Eqs.(2)-(4) for a single source case, the approximated expression $\hat{\delta} \approx \delta$ can be written in terms of $PD$ vector as follow.

$$\boldsymbol{\varphi}(l) \approx \kappa(l)\mathbf{R}\mathbf{a}(\phi, \theta) \tag{5}$$

where $\kappa(l)$=$\Delta\omega l$ is the angular wave number at $l$. The problem addressed in this paper is to approximate the left-hand phase difference estimation by assigning $\mathbf{a}(\phi, \theta)$ in the right-hand side of eq(5). Generally, our interest is in this case of $M$ >3 because an over-determined relationship has to be solved for the case.

In contrast to our approach for this problem discussed in III, the references [7] and [8] use the Moore-Penrose pseudo-inverse from $\boldsymbol{\varphi}(l)$ to $\mathbf{a}(\phi, \theta)$, then the nonlinear optimal computation from $\mathbf{a}(\phi, \theta)$ to $(\phi, \theta)$ is applied in [8].

It is noted that the problem for 2-dimensional case where all sensors as well as sources are on the same plane, for instance $z_m$=0 for all $m$ and $\theta$=$\pi/2$, is much simpler than the general discussions above. This restricted cases have been discussed partly by authors in [13].

## III. METHODS

### A. Phase difference manifold

Consider the case at a frequency bin $l$. At first, we refer to the right hand side of Eq.(5) as a $PD$ manifold. It is a parametric closed surface in the ($M$-1)-dimensional $PD$ space associated with the parameters $(\phi, \theta)$ varying within $-\pi \leq \phi \leq \pi$, $0 \leq \theta \leq \pi$, respectively. We denote the $PD$ manifold as follows:

$$\boldsymbol{\xi}(l; \phi, \theta) = \kappa(l)\mathbf{R}\mathbf{a}(\phi, \theta) \tag{6}$$

where $\mathbf{a}(\phi, \theta)$ is a unit sphere, since it satisfies $\| \mathbf{a}(\phi, \theta) \|$=1 in 3-dimensional space. Therefore, the Eq.(6) provides one-to-one correspondence between each point on the unit sphere $\mathbf{a}(\phi, \theta)$ and the corresponding $PD$ vectors on the $PD$ manifold $\boldsymbol{\xi}(l; \phi, \theta)$. The $PD$ manifold comprises all PD vectors each of which is induced by the source with a specific $\mathbf{a}(\phi, \theta)$.

The $PD$ manifold can be expressed by

$$\begin{aligned} \boldsymbol{\xi}(l; \phi, \theta) &= \mathbf{R}\mathbf{a}(\phi, \theta)\kappa(l) \\ &= (sin\theta cos\phi \mathbf{r}_x + sin\theta sin\phi \mathbf{r}_y \\ &\quad + cos\theta \mathbf{r}_z)\kappa(l) \end{aligned} \tag{7}$$

where

$$\mathbf{r}_p = [p_2, \cdots, p_M], \quad for\ p = x, y, z$$

From above, it is obvious that the $PD$ manifold is a closed surface in the 3-dimensional subspace denoted by $\Re$ spanned by $\mathbf{r}_x$, $\mathbf{r}_y$ and $\mathbf{r}_z$. As a consequence, we can define the unique inverse mapping from a point on the $PD$ manifold to the angles $(\phi, \theta)$ such that,

$$(\phi, \theta) = \Xi^{-1}\big[\boldsymbol{\xi}(l; \phi, \theta)\big]$$

### B. Inverse mapping $\Xi^{-1}$

The relationship Eq.(7) derives an inverse mapping $\Xi^{-1}$ from the $PD$ manifold $\boldsymbol{\xi}(l; \phi, \theta)$ to the direction angles $(\phi, \theta)$ by the following process $B$-1) -$B$-3). These consecutive procedures will be restrictively applied for the selected phase difference observations located on the $PD$ manifold.

$B$-1) Orthonomal basis in $\Re$

Since $\Re$ is spanned by the basis system $\{\mathbf{r}_x, \mathbf{r}_y, \mathbf{r}_z\}$, we may generate an orthonormal basis system $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ by applying the Gram-Schmidt orthogonalization.

$B$-2) Representation of $\boldsymbol{\xi}(l;\phi,\theta)$
In the new basis system, $\boldsymbol{\xi}(l;\phi,\theta)$ can be represented by

$$\boldsymbol{\xi}(\hat{\phi},\theta) = c_1(\phi,\theta)\mathbf{u}_1 + c_2(\phi,\theta)\mathbf{u}_2 + c_3(\phi,\theta)\mathbf{u}_3 \quad (8)$$

where

$$c_i(\phi,\theta) =< \boldsymbol{\xi}(\hat{\phi},\theta), \mathbf{u}_i > \quad , i = 1,2,3 \quad (9)$$

$<,>$ denotes the inner product.
In specific, we have

$$\begin{bmatrix} c_1(\phi,\theta) \\ c_2(\phi,\theta) \\ c_3(\phi,\theta) \end{bmatrix} = \begin{bmatrix} cos\phi sin\theta \|\mathbf{r}_x\| + sin\phi sin\theta d_1 + cos\theta d_2 \\ sin\phi sin\theta \|\hat{\mathbf{u}}_2\| + sin\theta d_3 \\ cos\theta \|\hat{\mathbf{u}}_3\| \end{bmatrix} \quad (10)$$

where $d_1$=$< \mathbf{u}_1, \mathbf{r}_y >$, $d_2$=$< \mathbf{u}_1, \mathbf{r}_z >$, $d_3$=$< \mathbf{u}_2, \mathbf{r}_z >$.

$B$-3) Transformation to $\mathbf{a}(\phi,\theta)$
Let define the matrix

$$\mathbf{T} = \begin{bmatrix} \frac{1}{\|\mathbf{r}_x\|} & -\frac{d_1}{\|\hat{\mathbf{u}}_2\|\|\mathbf{r}_x\|} & \frac{d_3 d_1 - \|\hat{\mathbf{u}}_2\| d_2}{\|\mathbf{r}_x\|\|\hat{\mathbf{u}}_2\|\|\hat{\mathbf{u}}_3\|} \\ 0 & \frac{1}{\|\hat{\mathbf{u}}_2\|} & -\frac{d_3}{\|\hat{\mathbf{u}}_2\|\|\hat{\mathbf{u}}_3\|} \\ 0 & 0 & \frac{1}{\|\hat{\mathbf{u}}_3\|} \end{bmatrix}$$

, then we can prove the following equation.

$$\mathbf{a}(\phi,\theta) = \mathbf{T} \begin{bmatrix} c_1(\phi,\theta) \\ c_2(\phi,\theta) \\ c_3(\phi,\theta) \end{bmatrix} \quad (11)$$

Finally, angles $(\phi,\theta)$ are uniquely obtained from $\mathbf{a}(\phi,\theta)$ on unit sphere that is, $\| \mathbf{a}(\phi,\theta) \|$=1.

### C. Reliable $T$-$F$ cell selection

For underdetermined multiple sources case, the sparseness of the time-frequency ($T$-$F$) or STFT domain representation of speech signals is essential. Therefore, we start from the $PD$ vectors at individual $T$-$F$ points which are denoted by $\varphi(k,l)$ where $k, l$ is time frame and frequency bin indices respectively. One of the basic ideas in this paper is to select a set of $T$-$F$ cells at which the $PD$ vectors are closely located to the $PD$ manifold $\boldsymbol{\xi}(l;\phi,\theta)$. This is because the phase difference estimation is thought to be reliable. For a given $PD$ vector at a $T$-$F$ cell $\varphi(k,l)$, we evaluate the Euclid distance between a vector $\varphi(k,l)$ and the $PD$ manifold in $\boldsymbol{\varphi}$-space by

$$d\{(\varphi(k,l), \boldsymbol{\xi}(l;\phi,\theta)\} := \min_{\phi,\theta} \| \varphi(k,l) - \boldsymbol{\xi}(l;\phi,\theta) \| \quad (12)$$

We select all $T$-$F$ cells of ($k$,$l$) at which $\varphi(k,l)$s satisfy $d\{(\varphi(k,l), \boldsymbol{\xi}(l;\phi,\theta)\} < th$ , since the estimated $PD$ on or closed to the $PD$ manifold would be a reliable one. The threshold value th is sufficiently small and determined empirically. The concrete procedure is omitted for the sake of space.
In addition to above $T$-$F$ cell selection, the conventional cell selection algorithm is also applied. The method is based on the DOA feature consistency with surrounding $T$-$F$ cells. [13]

### D. Error distribution of $(\phi,\theta)$

Here, the proposed estimation is established by generalizing previously proposed our DOA estimation scheme with a pair of sensor to the algorithm for arbitrary multi-sensor case. The selected $T$-$F$ cells consist of unknown multiple clusters of cells, each of which associated with one of the sources, for instance, $n$-th source with DOA parameter $(\phi_n,\theta_n)$. Thus reliable $PD$ vectors $\varphi(k,l)$ belonging this cluster locate around a point $\boldsymbol{\xi}(l;\phi_n,\theta_n)$. We assume that the error of the observed $\varphi(k,l)$ is independent identical omni-directional Gaussian distribution with zero mean, and the variance $\sigma$.
Let define the tangent plane of $\boldsymbol{\xi}(l;\phi,\theta)$ at $(\phi,\theta)$=$(\phi_n,\theta_n)$, and assume that the observed $PD$ vector $\varphi(k,l)$ locates on the tangent plane with 2-$D$ circular Gaussian with its mean $\boldsymbol{\xi}(l;\phi_n,\theta_n)$. According to the one-to-one correspondence between $\boldsymbol{\xi}(l;\phi,\theta)$ and $(\phi,\theta)$ in III, the standard deviations of Gaussian distribution with respect to $\phi$ and $\theta$ around $(\phi_n,\theta_n)$ are given by the following equations.

$$\sigma_x = \frac{\sigma}{\kappa(l) \parallel \frac{d\boldsymbol{\xi}}{dx} \parallel_{(\phi,\theta)=(\phi_n,\theta_n)}}, \quad for\ x = \phi,\ \theta \quad (13)$$

### E. Kernel density estimator

Finally the kernel density estimation algorithm[13] is applied to estimate the probability density function $p(\phi,\theta)$ for whole phase difference data $\varphi(k,l)$ at the selected $T$-$F$ cells. At first, observed $\varphi(k,l)$ on the $PD$ manifold $\boldsymbol{\xi}(l;\phi,\theta)$ gives $(\hat{\phi},\hat{\theta})$ by using the inverse mapping

$$(\hat{\phi},\hat{\theta}) = \Xi^{-1}(\varphi(k,l)) \quad (14)$$

As the consequence of this process for the selected $PD$ data $\varphi(k,l_i)$, let denote the estimated DOA angles for individual data by

$$(\hat{\phi}_i^{[l_i]}, \hat{\theta}_i^{[l_i]}) \quad i = 1, \cdots, I \quad (15)$$

Then, the kernel density estimator applied to above data gives

$$\hat{p}(\phi,\theta) = \frac{1}{I}\sum_{i=1}^{I} \frac{1}{\epsilon(l_i)\delta(l_i)} K\left(\frac{\phi - \hat{\phi}_i^{[l_i]}}{\epsilon(l_i)}, \frac{\theta - \hat{\theta}_i^{[l_i]}}{\delta(l_i)}\right) \quad (16)$$

where $\epsilon(l_i)$ and $\delta(l_i)$ are the bandwidths of a 2-$D$ kernel function $K(\phi,\theta)$ with respect to $\phi$ and $\theta$, and are respectively written by

$$\epsilon(l_i) = \sigma_\phi|_{l=l_i}\hbar, \quad \delta(l_i) = \sigma_\theta|_{l=l_i}\hbar \quad (17)$$

where $\hbar$ is a control parameter of these bandwidths.

## IV. EXPERIMENTS

**Experiment1**: Experiments are conducted in a conference room (Width=18m, Depth=15m, Height=8m) using a regular tetrahedron microphone array with 4cm on each side. Other experimental parameters are listed as follows: Sampling frequency=8kHz, Sound speed c=340m, STFT Frame Length=1024points, Window=Hamming, Frame overlap=512points. For five sources case, the result of the proposed $T$-$F$ cell selection followed after the inverse mapping of phase difference vectors is shown in Figure1.

Three axes in Fig.1 (a) represent three phase differences $\boldsymbol{\varphi_2}$, $\boldsymbol{\varphi_3}$, and$\boldsymbol{\varphi_4}$, and a set of normalized phase difference data plot at whole T-F cells is shown. On the other hand, Fig.1(b) shows the plot of the inversely-mapped and selected phase difference data is shown in the transformed/normalized space. Figure demonstrates the effectively distributed results.

**Experiment2**: We conducted experiments using image method to confirm performance of proposed method under the reverberation condition is $T_{60}$=1200ms. The real angles of five speakers are; $(\phi_1, \theta_1)$, $(\phi_2, \theta_2)$, $(\phi_3, \theta_3)$, $(\phi_4, \theta_4)$, $(\phi_5, \theta_5)$=(30,45),(60,60),(90,150),(120,120),(240,135) degrees. Fig.2 illustrates the estimated probability density of DOA derived from the proposed method and the histogram obtained by the conventional method [8]. In the figure, vertical lines show the real directions of speakers. The proposed method estimates accurate results such as, (33, 44), (61,58), (93,151), (121,122), (241,135)degrees.

## V. CONCLUSION

In this paper, we presented a novel method for estimating DOAs of multiple sources by using the phase difference of time-frequency components. The method is applicable for arbitrary array configuration in 3-dimesnsional space. A unique relationship between phase difference vector and the direction angles are established by introducing the phase different manifold, and this relationship enables us to apply a previously proposed kernel density algorithm utilizing a statistics of phase difference error.

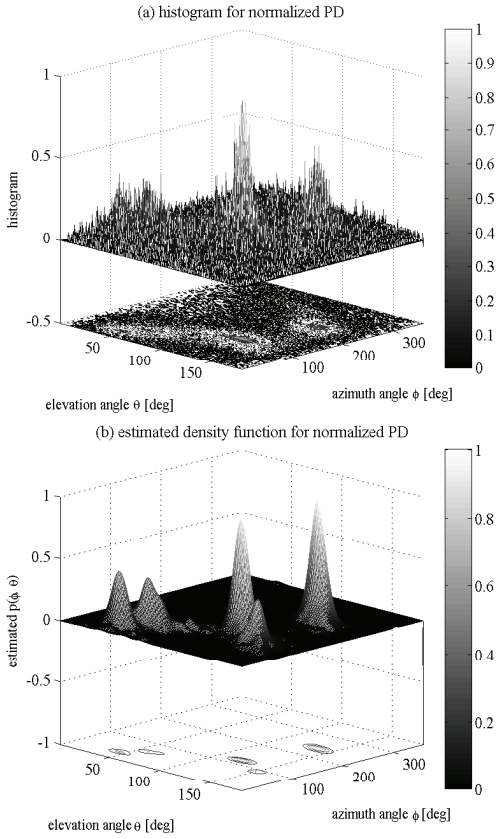


Fig. 2. Histogram of normalized $PD$ (a) and Estimated density function (b)

## REFERENCES


[1] E.D.D.Claudio and R. Parisi, Microphone Arrays. Springer-Verlag, 2001, ch. Multi-Source Localization Strategies, pp. 181-201.

[2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. on Antennas and Propagation, vol. 34, pp.276-280, 1986.

[3] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delays," IEEE Trans. on Acoust. Speech Signal Process., vol. ASSP-24, pp. 320-327, 1976.

[4] J. Huang, N. Ohnishi, and N. Sugie, "A biomimetic system for localization and separation of multiple sound sources," IEEE Trans.on Instrumentation and Measurement, vol. 44, pp. 733-738, 1995.

[5] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," IEEE Transactions on Signal Processing, vol. 52, no. 7, pp. 1830-1847, 2004.

[6] S. Arberet, R. Gribonval, and F. Bimbot, "A robust method to count and locate audio sources in a multichannel underdetermined mixture," IEEE Transactions on Signal Processing, vol. 58, no. 1, pp. 121-133, 2010.


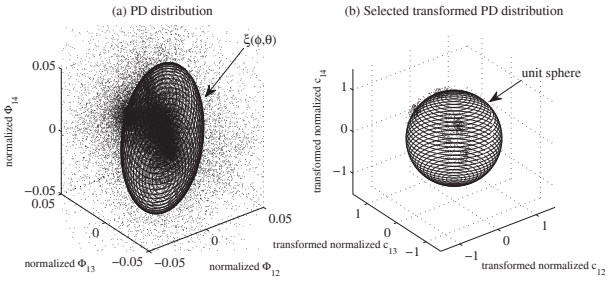


Fig. 1. $PD$ transform and $T$-$F$ cell distribution in $\boldsymbol{\varphi}$-$l$ space


[7] S. Araki, H. Sawada, R. Murai, and S. Makino, "Doa estimation for multiple sparce sources with arbitrarily arranged multiple sensors," Journal of Signal Processing Systems, 2009.

[8] B. Berdugo, J. Rosenhouse, and H. Azhari, "Speakers direction finding using estimated time delays in the frequency domain," Signal Processing, vol. 82, pp. 19-30, 2002.

[9] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 5, pp. 1592-1604, 2007.

[10] F. Nesta, P. Svaizer, and M. Omologo, "Cumulative state coherence transform for a robust two-channel multiple source localization," Proc. ICA, pp. 290-297, 2009.

[11] N. Ding and N. Hamada, "DOA estimation of multiple speech sources from a stereophonic mixture in underdetermined case", Trans. on Fundamentals, IEICE accepted for publication

[12] N. Ding, K. Fujimoto, and N. Hamada, "Kernel density estimator approach for solving underdetermined source localization problem in arbitrary microphone configuration", 26th Signal Processing Symposium, Sapporo, Nov. 2011.

[13] F. Abrard and Y. Deville, " A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources, " Signal Processing, vol. 85, pp. 1389-1403, 2005.

[14] C. M. Bishop, Pattern recognition and machine learning. Springer, 2006.