Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Real-time acoustic blind signal separation system based on the spatio-temporal gradient analysis

Kenbu Teramoto and Md. Tawhidul Islam Khan

Saga University, 1-Honjo, 8408502 Saga, Japan
tera@me.saga-u.ac.jp

This paper presents an autonomous directivity microphone system for the blind source separation based on the newly proposed spatio-temporal gradient algorithm. The proposed blind signal separation algorithm utilizes the linearity among four signals: (1) the sound pressure, (2) $\boldsymbol{x}$, (3) $\boldsymbol{y}$, and (4) $\boldsymbol{z}$-directional particle velocities, all of which are governed by the wave equation and the motion equation. The proposed method, therefore, has an ability to simplify the convoluted blind source separation problems into the instantaneous blind source separation problems over the spatio-temporal gradient space. Several acoustical experiments have been performed with acceptable performance of the proposed method for the real-time acoustic blind source separation.

1 Introduction

Blind source separation (BSS) based on the independent component analysis (ICA) plays an important role in many kinds of areas. The mission of BSS is to reconstruct the transmitted signals from the observed data[]. In acoustical signal processing and related areas, the observed signal consists of an unknown source signal mixed with itself and different time delays. Fourier transform techniques, therefore, are used for replacing convolutions with products in the frequency domain, although, several problems: indeterminacy of permutation and sign exist in ICA[]. In order to avoid the preceding problems, the proposed BSS algorithm utilizes the fact that the observed particle velocity vectors can be denoted as the linear combination of sound pressure which is excited by each sound source. Consequently, the BSS based on the particle velocity measurement makes the convolution problem of blind source separation into the simplest instantaneous mixing problem.

Blind signal separation (BSS) consists of recovering original signals or sources from several observed mixtures. Typically, the observations are obtained at the output of a set of sensors each of which receives a different combination of the source signals. The adjective 'blind' stresses the fact that no priori information about i) source signals and ii) mixture of signals are available. The approach of modeling a sound signal that transfers from the source to the sensor is rather difficult when no priori information about the transfer system is available. The lack of priori knowledge about the mixing process is compensated by a statistically strong but physically often plausible assumption of independency among the source signals. The so-called blindness should not be understood negatively: the weakness of the prior information is precisely the strength of the BSS model, making it a versatile tool for exploiting the spatial diversity provided by sensors. Promising applications can be found in the processing of communication signals, biomedical signals like ECG (electro-cardiogram) and EEG (electro-encephalography) particularly for extracting the signals of fetus heart beat from several acoustic transducers outside the mother's body. Especially, acoustic applications including the cocktail-party problem, sonar problem etc. widen its applications. Equally, the spatio-temporal gradient method that has been applied already in the field of non-destructive evaluation,[] is proposed as one of the acoustical signal processing algorithm. Combining these methods, the spatio-temporal gradient algorithm for blind source separation makes the convolutions of blind source separation into the simplest instantaneous mixture problems.[]

The spatio-temporal gradient algorithm of blind source separation has been described in the present paper analytically as well as experimentally. Two voice signals are utilized in the experiments. The results of acoustic experiments have verified the proposed signal separation method very well. The paper has been organized as follows: the second section presents the problem formulation. The third section explains the experimental analysis. Experimental results as well as related discussions are presented also in this section. Finally, the concluding remarks are mentioned in the section four.

2 Problem Formulation

2.1 Overlaps of Three Plane Waves

Assuming that three band-limited plane waves of acoustic field propagate to three linearly independent directions $\boldsymbol{p}_{i=1,2,3}$ that are shown in Fig.1. The acoustic signals of these plane waves are denoted as follows:

$$\begin{aligned} f_i(x,y,z,t) &= w_i(t - \frac{1}{c}\boldsymbol{r}^T\boldsymbol{p}_i) \\ &= w_i(t + \frac{1}{c}(x\sin\theta_i\cos\phi_i \\ &+ y\sin\theta_i\sin\phi_i + z\cos\theta_i)), \quad (1) \\ & \quad i = 1,2,3, \end{aligned}$$

where, c denotes the phase velocity of the airborne sound and $\boldsymbol{r}$ indicates the position of the observation point ($\boldsymbol{r} = (x \quad y \quad z)^T$). The band-limited

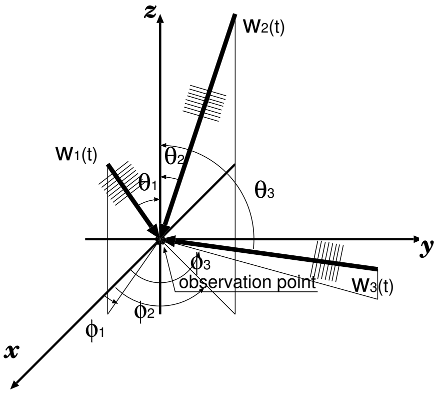

Figure 1: Geometry of the three independent incident plane waves and the observation point

signal $w_i(t)$ denotes the source signal as expressed below:

$$w_i(t) = \int_{-\omega_i}^{\omega_i} a_i(\omega) e^{j\omega t} \mathrm{d}\omega. \tag{2}$$

where ω_i denotes the bandwidth of each source signal. Again, directions of arrival of the acoustic signals can be expressed in the following vectors:

$$\boldsymbol{p}_i = (-\sin\theta_i \cos\phi_i \quad -\sin\theta_i \sin\phi_i \quad -\cos\theta_i)^T \tag{3}$$

Here, angles, θ_i and ϕ_i denote the elevational and azimuthal directions respectively. Furthermore, at the origin, $x = y = z = 0$, the temporal gradient can be obtained as:

$$f_{it}(0,0,0,t) = \frac{\partial}{\partial t} f_i(x,y,z,t)\Big|_{x=y=z=0} = \dot{w}_i(t). \tag{4}$$

Orthogonal triplets of the spatial gradient are derived as follows:

$$\begin{aligned} f_{ix}(0,0,0,t) &= \frac{\partial}{\partial x} f_i(x,y,z,t)\Big|_{x=y=z=0} \\ &= \frac{\sin\theta_i \cos\phi_i}{c} \dot{w}_i(t) \end{aligned} \tag{5}$$

$$\begin{aligned} f_{iy}(0,0,0,t) &= \frac{\partial}{\partial y} f_i(x,y,z,t)\Big|_{x=y=z=0} \\ &= \frac{\sin\theta_i \sin\phi_i}{c} \dot{w}_i(t) \end{aligned} \tag{6}$$

$$\begin{aligned} f_{iz}(0,0,0,t) &= \frac{\partial}{\partial z} f_i(x,y,z,t)\Big|_{x=y=z=0} \\ &= \frac{\cos\theta_i}{c} \dot{w}_i(t) \end{aligned} \tag{7}$$

2.2 Instantaneous Mixture

Based on the above equations (Eqs. (4) to (**??**)), the orthogonal spatial gradients of the observed signals, $f(x,y,z,t)$, at the origin can be denoted as:

$$\nabla f(x,y,z,t)\Big|_{x=y=z=0} = A \begin{pmatrix} \dot{w}_1(t) \\ \dot{w}_2(t) \\ \dot{w}_3(t) \end{pmatrix} \tag{8}$$

where, the mixing matrix A can be defined as:

$$A = (\,\boldsymbol{p}_1 \quad \boldsymbol{p}_2 \quad \boldsymbol{p}_3\,) \tag{9}$$

However, based on the wave equation of airborne sound, particle velocity vector $\boldsymbol{v}(x,y,z,t)$ and the spatial gradients of the sound pressure $f(x,y,z,t)$ satisfy the following equation.

$$\rho \frac{\partial \boldsymbol{v}(x,y,z,t)}{\partial t} = -\nabla f(x,y,z,t) \tag{10}$$

where, ρ is the density of the air. Therefore, after the integration of the above equation with respect to time, Eq. (8) can be denoted as follows:

$$\boldsymbol{v}(0,0,0,t) = -\frac{1}{\rho} A \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \end{pmatrix} \tag{11}$$

2.3 LMS-like BSS-DOA algorithm

For the real time separation, a LMS-like BSS-DOA process is used in the POC experiment.

The mixing process of the ideal side microphone can be denoted as:

$$f_s(t) = (\,\cos\phi_1 \quad \cos\phi_2\,) \begin{pmatrix} w_1(\mathbf{0},t) \\ w_2(\mathbf{0},t) \end{pmatrix}, \tag{12}$$

where $f_s(t)$ denotes the output signal from the side microphone. Therefore the output signal, $f_s(t)$ is proportional to the x directional particle velocity at the observation point. In the same way, the output signal from the mid microphone can be denoted as:

$$f_s(t) = (\,\tfrac{\sin\phi_1 + 1}{2} \quad \tfrac{\sin\phi_2 + 1}{2}\,) \begin{pmatrix} w_1(\mathbf{0},t) \\ w_2(\mathbf{0},t) \end{pmatrix}. \tag{13}$$

Consequently, the output signal show the sum of a signal which is proportional to the y directional particle velocity and a signal which is proportional to the sound pressure at the observation point. Through the above mentioned electrical matrix circuit, the L/R channel signal are obtained as:

$$\begin{pmatrix} f_L(t) \\ f_R(t) \end{pmatrix} = \begin{pmatrix} \beta & 1-\beta \\ \beta & \beta - 1 \end{pmatrix} \begin{pmatrix} f_m(t) \\ f_s(t) \end{pmatrix}, \tag{14}$$

where $\beta(\in (0,1))$ changes the directional pattern of L/R channels. On the basis of the above equation, when β which satisfies

$$\frac{\beta}{1-\beta} = \cos\frac{\pi}{6} \tag{15}$$

is given, the mixing matrix of the MS-microphone can be denoted as:

$$A = \begin{pmatrix} \cos(\phi_1 - \frac{\pi}{6}) + \frac{1}{2} & \cos(\phi_2 - \frac{\pi}{6}) + \frac{1}{2} \\ \cos(\phi_1 - \frac{5\pi}{6}) + \frac{1}{2} & \cos(\phi_2 - \frac{5\pi}{6}) + \frac{1}{2} \end{pmatrix} \tag{16}$$

The online BSS-DOA algorithm is summarized as follows:

at $t = n\Delta t$

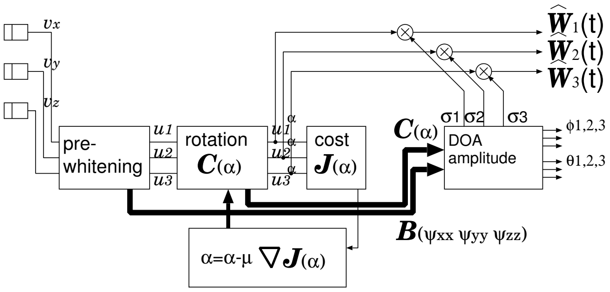

Figure 2: Block diagram of the on-line BSS-DOA processing

1. $f_L(n\Delta t)$ and $f_R(n\Delta t)$ are obtained.

2. **Pre-whitening**

   (a) The mean values of the observed signals, $\overline{f}_L$ and $\overline{f}_R$, are modified as follows:

$$\overline{f}_L^{[n]} \leftarrow \overline{f}_L^{[n-1]} e^{\Delta t/\tau} + f_L(n\Delta t)\Delta t/\tau \quad (17)$$

$$\overline{f}_R^{[n]} \leftarrow \overline{f}_R^{[n-1]} e^{\Delta t/\tau} + f_R(n\Delta t)\Delta t/\tau \quad (18)$$

   (b) The variances and covariance of the observed signals, ψ_{LL} , ψ_{RR} and ψ_{LR} are modified as follows:

$$\begin{aligned}\psi_{LL}^{[n]} &\leftarrow \psi_{LL}^{[n-1]} e^{\Delta t/\tau}\\ &+ (f_L(n\Delta t) - \overline{f}_L^{[n]})^2 \Delta t/\tau\end{aligned} \quad (19)$$

$$\begin{aligned}\psi_{RR}^{[n]} &\leftarrow \psi_{RR}^{[n-1]} e^{\Delta t/\tau}\\ &+ (f_R(n\Delta t) - \overline{f}_R^{[n]})^2 \Delta t/\tau\end{aligned} \quad (20)$$

$$\begin{aligned}\psi_{LR}^{[n]} &\leftarrow \psi_{LR}^{[n-1]} e^{\Delta t/\tau}\\ &+ (f_L(n\Delta t) - \overline{f}_L)\\ &\cdot (f_R(n\Delta t) - \overline{f}_R^{[n]})\Delta t/\tau\end{aligned} \quad (21)$$

   (c) the pre-whitening matrix $\boldsymbol{B}$ is defined:

$$\boldsymbol{B}^{[n]} \leftarrow \begin{pmatrix} \frac{1}{\sqrt{\psi_{LL}^{[n]}}} & -\frac{\psi_{LR}^{[n]}}{\sqrt{\psi_{LL}^{[n]}}\sqrt{\psi_{LL}^{[n]}\psi_{RR}^{[n]}-\psi_{LR}^{[n]\,2}}} \\ 0 & \frac{\sqrt{\psi_{LL}^{[n]}}}{\sqrt{\psi_{LL}^{[n]}\psi_{RR}^{[n]}-\psi_{LR}^{[n]\,2}}} \end{pmatrix} \quad (22)$$

   (d) the normalized orthogonal signal u_1 and u_2 can be obtained as:

$$\begin{pmatrix} u_1^{[n]} \\ u_2^{[n]} \end{pmatrix} \leftarrow \boldsymbol{B}^{[n]} \begin{pmatrix} f_L(n\Delta t) - \overline{f}_L^{[n]} \\ f_R(n\Delta t) - \overline{f}_R^{[n]} \end{pmatrix} \quad (23)$$

3. **Separation**

   (a) Rotation matrix $\boldsymbol{C}$ is defined:

$$\boldsymbol{C}^{[n]} \leftarrow \begin{pmatrix} \cos\alpha^{[n-1]} & -\sin\alpha^{[n-1]} \\ \sin\alpha^{[n-1]} & \cos\alpha^{[n-1]} \end{pmatrix} \quad (24)$$

   (b) independent components $u_1^{[\alpha]}$, $u_2^{[\alpha]}$ are obtained:

$$\begin{pmatrix} u_1^{[\alpha][n]} \\ u_2^{[\alpha][n]} \end{pmatrix} \leftarrow \boldsymbol{C}^{[n]} \begin{pmatrix} u_1^{[n]} \\ u_2^{[n]} \end{pmatrix} \quad (25)$$

   (c) the rotation angle, α, is modified in the LMS-like manner as:

$$\alpha^{[n]} \leftarrow \alpha^{[n-1]} + \mu\Delta J^{[n]}, \quad (26)$$

   where, ΔJ is defined as:

$$\begin{aligned}\Delta J^{[n]} &\leftarrow -u_2^{[n]}((u_1^{[n]})^3 - 3(u_1^{[n]}))\\ &+ u_1^{[n]}((u_2^{[n]})^3 - 3(u_2^{[n]}))\end{aligned}$$

4. **DOA estimation**

   (a) The inverse of the de-mixing matrix:

$$\begin{pmatrix} d_{11}^{[n]} & d_{12}^{[n]} \\ d_{21}^{[n]} & d_{22}^{[n]} \end{pmatrix} = \left(\boldsymbol{C}^{[n]}\boldsymbol{B}^{[n]}]\right)^{-1} \quad (27)$$

   (b) estimation of the standard deviations of the source signals: $\hat{\sigma}_i$ for $i = 1, 2$

$$\hat{\sigma}_i^{[n]} \leftarrow \frac{2((d_{1i}^{[n]})^2 + (d_{2i}^{[n]})^2 + d_{1i}^{[n]} d_{2i}^{[n]})}{2(d_{1i}^{[n]} + d_{2i}^{[n]})} \quad (28)$$

   (c) estimation of DOAs of the source signals: $\hat{\phi}_i$ for $i = 1, 2$

$$\hat{\phi}_i = 2\tan^{-1}\sqrt{3}\frac{d_{1i} + d_{2i}}{d_{1i} - d_{2i}} - \frac{\pi}{2} \quad (29)$$

3 Acoustical Experiments

We have implemented and tested the proposed spatio-temporal blind source separation algorithm with the Proof-Of-Concept (POC) model. For simplicity, 2-dimensional field is assumed and the observation point is set at the origin, ($\boldsymbol{x} = \boldsymbol{0}$).

3.1 Experimental setup

The Proof-Of-Concept (POC) model experiment with two acoustic source signals as shown in Fig. 3. (1) shows the external view of the MS-microphone(SONY ECM999). MS stereo is a two microphone capsule technique using a primary microphone as the mid-signal and a bi-directional microphone for the

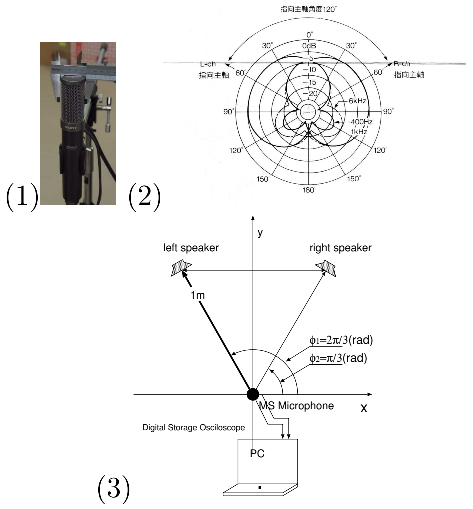

Figure 3: POC-model experimental setup:(1)MS-microphone: SONY ECM-999, (2)the solid curve shows the directional patterns of the L-channel (mid +side) and the dotted curve shows the R-channel one (mid -side)(dotted curve) microphone from the date-sheet of ECM-999, (3)schematic diagram of the experimental setup.

side-signal. A stereo signal is generated combining these two microphones, with different polar responses, in an electrical matrix. Whether decoded during recording or in post production, the amount of stereo spread can be adjusted, a great benefit of MS stereo over other stereo microphone techniques. In (2), the both directional patterns are shown. To construct a stereo (left/right) signal the mid and side signals are combined in an electrical matrix which sums the following signals:

- Mid signal sent to both left and right equally,
- Side signal sent to the left channel,
- Side signal with polarity reversed sent to right channel.

The schematic view of the POC model experiment is shown in (3). The rangial the source signals to the observation point (O) is 1.0m.

Two female voice signals (one is NHK stereo broadcasting opening narration, recorded in 1954 and another is "TOEFL LESSON" narration, recorded in 2001) have been used in the POC model experiment as source signals(cf. Table1). Both sources $w_1(t)$ and $w_2(t)$ were considered statistically independent and each of which has been emitted for 20 seconds.

Table 1: Specifications of source signals

	$w_1(t)$	$w_2(t)$
source	female voice, "NHK Stereo Broadcasting-Opening narration" recoded in 1954	female voice, "TOEFL LESSON narration" recorded in 2001
length	20s	
sampling resolution	16bit, 44.1kHz, normalized by 0000=-1.0, FFFF=1.0	
mean	$m_1 = 0.0$	$m_2 = 0.0$
standard deviation	$\sigma_1 = 0.1171$	$\sigma_2 = 0.0792$
moments		
the 3rd	-2.75×10^{10}	1.10×10^{10}
the 4th	7.28×10^{20}	5.91×10^{20}
the 5th	-7.34×10^{18}	1.64×10^{18}
mutual information	0.043bit	

3.2 Separation of two static source signals

The ripple marks of the two source signals $w_1(t)$, $w_2(t)$ are shown in Fig.4(1) and (2) respectively. The observed Left and Right channel signal (at the observation point) are shown in Fig. 5 respectively. In the figure the mixing characteristics of the two source signals are shown. The characteristics of the separated signals by the POC experimental methods are shown in Fig. 6. The experimental results of the arrival direction of the wave fronts are shown in Fig. 7. Initially, the result shows that the arrival directions settles to the $\phi_1 = \pi/3$ and the $\phi_2 = 2\pi/3$ gradually, which are the directions of the true sound sources.

4 Concluding Remarks

In this paper, the method of the blind source separation by spatio-temporal gradient analysis based on the linear advection equation has been presented. The LMS-like on-line algorithm has been applied successfully for the blind source separation method of two independent sources. The present separation method is considered to be independent of source location which can present simple digital signal processing algorithm characterizing the wave field where incident three plane waves overlap each other. For detecting the directions of arrivals (DOA), the MS-microphone system has been used as the

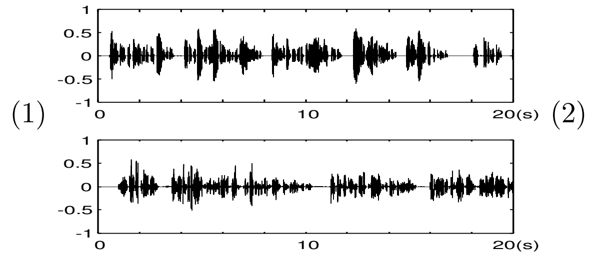

Figure 4: Original sound signals:(1)No.1, (2) No.2

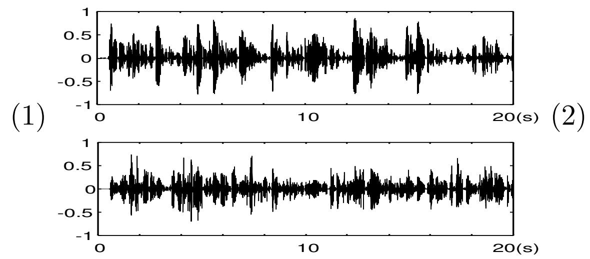

Figure 5: Observed signals of (1)L-ch and (2) R-ch.

analysis shows the linear dependency among source sound pressures. The proposed separation technique has been verified with two dimensional acoustic model experiments and the results have been summarized with the following conclusions and remarks:

1. The spatio-temporal gradient analysis which utilizes the linear advection equation has an ability to simplify the convolution blind source separation problems into the instantaneous blind source separation.

2. The directions of wave propagation can be determined with the proposed technique of blind source separation.

References

[1] J. F. Cardoso, "Source separation using higher order moments," *Proc. IEEE ICASSP*, Vol.4, pp.2109-2112, 1989

[2] S. Amari, S. C. Douglas, et.al, "Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach," *Proc. 11th IFAC Symp. System Identification*, pp.1007-1012, 1997

[3] J.F.Cardoso and B.Laheld,"Equivariant adaptive source separation," *IEEE Trans., Signal Processing*, Vol.44-12, pp.3017-3030, 1997

[4] K. Matsuoka, M. Ohya and M.Kawamoto, "A neural net for blind separation of nonstationary signal", *Neural Networks*, Vol.8, pp.411-419, 1995

[5] N. Murata, S. Ikeda and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals", *Neurocomputing*, Vol.41, pp.1-24, 2001

[6] K.Teramoto, K.Tsuruta, "Blind source separation based on the spatio-temporal gradient analysis," *Proc. of ICSV12*, in CDROM, 2005

[7] A.Hyvaerinen, J.Karhunen, E.Oja, "Independent Component Analysis," J.Wiley and Sons, 2001

[8] H. Ogura and N. Nakasako, "Direct method of blind separation for two mixed signals based on the statistical independency from information theoretic approach," *Proc. of ISNIC-98*, pp.215-220, 1998

[9] H. Sawada, R. Mukai, S. Araki, S. Makino, "A Robust and Precise Method for Solving the Permutation Problem of Frequency-Domain Blind Source Separation," *IEEE Trans. Speech and Audio Processing*, Vol. 12-5, pp.530-538, 2004

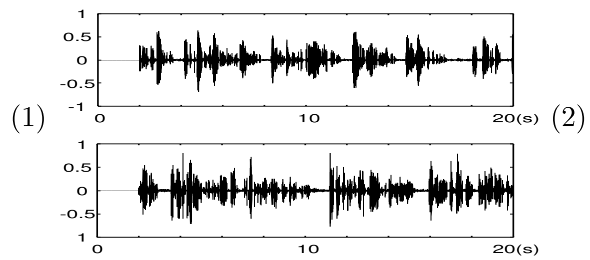

Figure 6: Separated sound signals: (1)No.1, (2)No.2

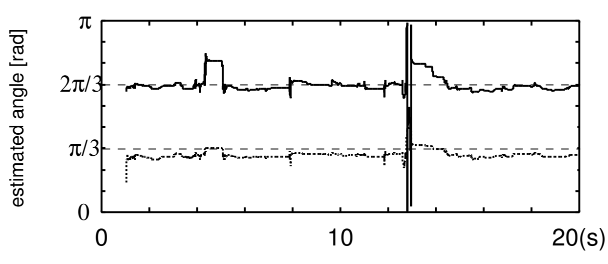

Figure 7: Convergence plots of the estimated directions of the original sound waves: No.1 and No.2.