SUMMARY In this paper, we propose a DOA (Direction Of Arrival) estimation method of speech signal using three microphones. The angular resolution of the method is almost uniform with respect to DOA. Our previous DOA estimation method [1] using the frequency-domain array data for a pair of microphones achieves high precision estimation. However, its resolution degrades as the propagating direction being apart from the array broadside. In the method presented here, we utilize three microphones located at vertices of equilateral triangle and integrate the frequency-domain array data for three pairs of microphones. For the estimation scheme, the subspace analysis for the integrated frequency array data is proposed. Through both computer simulations and experiments in a real acoustical environment, we show the efficiency of the proposed method.

key words: microphone array, DOA estimation, microphone pair, eigenspace analysis, harmonic structure, spatial resolution

1. Introduction

Human-machine interfaces based on speech have attracted much concern recently, because they provide a human-friendly access to a machine. As a core technology in speech-based human-machine interfaces, high quality speech signal acquisition is required for high-performance speech recognition. Microphone array [2],[3] is a well-known effective technique to improve the quality of received speech signal. Most of speech enhancement techniques using microphone array [2],[3] require the direction of speaker in advance.

Among the conventional DOA estimation methods [4], the MUSIC (MUltiple Signal Classification) [5] is well-known for providing high spatial resolution. The MUSIC, as its original form, is applicable for narrowband input signal. For broadband signals, in which speech signal is included, many DOA estimation methods have been reported as well [6]–[8]. Wang et al. proposed the Coherent Signal-Subspace(CSS) [7] that enables to apply MUSIC to broadband signals. In the method, the components in several frequency bands are gathered together (this process is called focusing). The focusing requires rough DOA pre-estimation in advance, and the pre-estimation error highly affects the final estimation result in practice.

The physical scale of array is another subject to be considered from a practical point of view. Generally, the performance of an array signal processing for estimating DOA, as well as that for rejecting interferences, is improved by increasing both the number of sensors and the array aperture size. However, they are often restricted due to the limited physical size of the apparatus on which the array sensors are installed. For this subject, some studies of DOA estimation using a few microphones have been reported [1],[9],[10]. The propagation time delay and the sound pressure difference between sensors are used in [9] and in [10], respectively.

Previously, we proposed a DOA estimation method for speech using two microphones [1]. We used the virtually generated multichannel array data, or simply called “frequency array data,” given by a pair of microphones. However, there are two drawbacks in this method. One is that the estimation resolution degrades as the propagating direction apart from the array broadside. Another one is that the method cannot discriminate whether the speaker is in front or behind. These properties are inherent in linear array, and the only arrangement of a two-sensor microphone array is linear.

In this study, we propose a new speech DOA estimation method using three microphones. This method can estimate omni-directional DOA with realizing spatially uniform resolution.

The main proposals in this research are summarized as follows.

- Array of three microphones located at vertices of equilateral triangle (we call it “equilateral-triangular microphone array” in the below).
- A new DOA estimation scheme by integrating the frequency array data for three pairs of microphones

The former aims to realize uniform resolution with respect to DOA. In the equilateral-triangular sensor arrangement, we can take three pairs of microphones and each of them is a linear 2-sensors array. Since the broadside of each microphone pair faces at a different angle individually, we can expect uniform resolution by integrating the frequency array data of these three pairs. The second idea is the integrated use of the three frequency array data. At the estimation stage in our proposed method, we apply the subspace analysis to the integrated frequency array data. The process does not require any a priori DOA knowledge.

The additional advantage of the proposed method is its robustness to reverberation. That is because the array data integration gives a spatial averaging effect.
This paper is organized as follows. In the following Sect. 2, we mention the problem settings, and a brief review of our previous method [1] is explained in Sect. 3. Then the details of the proposed method and its analyses are described in Sect. 4. Simulation and experimental results are shown in Sect. 5 to confirm the efficiency of the proposed method, and some concluding remarks are stated in Sect. 6.

2. Problem Settings

We use an equilateral-triangular microphone array as shown in Fig. 1. The three microphones are located at the vertices of an equilateral triangle. A speaker in the direction \( \theta \) utters a voiced speech signal \( s(n) \). The microphones receive the signal \( x(n) \), \( y(n) \) and \( z(n) \) respectively, with additive sensor noise signals \( n_x(n) \), \( n_y(n) \) and \( n_z(n) \) that can be modeled as spatially uncorrelated. From such configuration, we can take three pairs of microphones that have equal distance \( \sqrt{3R} \) between microphones and each pair faces to different direction of every \( \frac{\pi}{3} \) [rad]. For a linear array, including a microphone pair as the simplest case, it has highest spatial resolution to its facing (broadside) direction. The aim of the proposed method is to realize uniform resolution by integrating these three pairs.

Here we assume for the input signal without loss of generality.

a) One voiced speech signal is received.

A speech signal mainly contains voiced sound localizing at some time segments [11] and usually it is possible to extract a single speaker time segments even for a double speak case.

b) The location of the speaker is restricted on the array plane.

Later, the effects of circumstances violating this assumption are considered in Sect. 5.3.

3. Speech DOA Estimation Using a Pair of Microphones

We briefly review the speech DOA estimation method using a pair of microphones proposed in [1]. Let us consider the two channel signals \( [x_1(n), x_2(n)] \) in Fig. 2, obtained by a pair of microphones, represented by

\[
\begin{align*}
x_1(n) &= s(n) + n_1(n) \\
x_2(n) &= s(n - \tau) + n_2(n),
\end{align*}
\]

where \( s(n) \) is a voiced speech signal, \( \tau \) is the time delay between two microphones which is a function of the source signal’s DOA \( \theta \), and \( n_1(n) \) and \( n_2(n) \) are mutually uncorrelated white noise signals. Thus, the Fourier transforms of \( x_1(n) \) and \( x_2(n) \), and their cross spectrum, are represented by

\[
\begin{align*}
X_1(\omega) &= S(\omega) + N_1(\omega) \\
X_2(\omega) &= S(\omega) e^{-j \omega \tau} + N_2(\omega)
\end{align*}
\]

and

\[
G_{12}(\omega) = E[X_1^*(\omega)X_2(\omega)] = P_S(\omega) e^{-j \omega \tau}
\]

respectively, where \( P_S(\omega) \) and the expectation \( E[\cdot] \) denote the power spectral density of \( s(n) \) and the average of DFT at several frames respectively, and \( \ast \) means the complex conjugate. When we set \( \omega = \omega_m \), where \( \omega_m \) is the \( m \)-th higher harmonics of the fundamental frequency \( \omega_0 \) of \( s(n) \), i.e.

\[
\omega_m = m \omega_0,
\]

the phase term in \( G_{12}(\omega_m) \) is replaced by \( e^{-j m \omega_0 \tau} \). This phase term is interpreted as a time delay, \( m \) times \( \tau \), of a narrow-band signal whose central frequency is \( \omega_0 \). This interpretation leads us to the idea that the \( G_{12}(\omega_m) \) might be the virtual multichannel array signals, which are narrow-band signals acquired by an equally spaced linear multiple sensors. For determining the frequency \( \omega_0 \) and its harmonics, we use the harmonic structure of voiced sound \( s(n) \). That is, we set \( \omega_0 \) as the fundamental frequency of voiced sound in speech. Because the power of a voiced sound is localized in its harmonic frequencies, the SNRs at these frequencies are rather high, and as a result, harmonic elements contribute to improving the estimation accuracy. Thus, we define the following frequency array data \( G(\omega_m) \) for a pair of microphone signals.

\[
G(\omega_m) = \begin{bmatrix} G_{12}(\omega_m) \\ \bar{G}_{12}(\omega_m) \\ \bar{G}_{12}(\omega_m) \end{bmatrix}
\]

Fig. 1 Model of input signal to the equilateral-triangular microphone array.

Fig. 2 Microphone pair model to derive a frequency array data.
\( (a, b, \cdots \in \mathbf{m}) \)

Since the power spectrum distribution depends on speaker and phoneme, here we select the \( M \) harmonics that contains the voiced speech components in higher SNR condition. In Eq. (7), \( \mathbf{m} \) is a set of the \( M \) harmonics order selected by thresholding the magnitude-squared coherence function [13], as given by Eq. (8)–Eq. (10)

\[
m = (m \mid \eta_{xy}(\omega_0)) \geq T, \: (m = 1, 2, \cdots, M) \tag{8}
\]

\[
\eta_{xy}(\omega_0) = 10 \log_{10} \frac{|y_{xy}(\omega_0)|^2}{1 - |y_{xy}(\omega_0)|^2} \tag{9}
\]

\[
|x_{xy}(\omega_0)|^2 = \frac{E[X[Y(\omega_0)]]^2}{E[X(\omega_0)]^2} \tag{10}
\]

The \( M \) is the highest order of the candidate harmonics determined by the criterion stated in [1]. The fundamental frequency \( \omega_0 \) is estimated by evaluating logarithmic harmonic product spectrum [14]. Here we note that the magnitude of each component in \( \mathbf{G}(\omega_0) \) are normalized as shown in Eq. (7). Finally, the DOA estimation is performed by applying the MUSIC [5] to the frequency array data \( \mathbf{G}(\omega_0) \).

4. Proposed Method

4.1 Model of Input Signal

The short-time Fourier transforms of each microphone input signals \( x(n), y(n) \) and \( z(n) \) in Fig. 1 are given by

\[
\begin{align*}
X(\omega) &= S(\omega)e^{-j\omega\tau_x} + N_x(\omega) \\
Y(\omega) &= S(\omega)e^{-j\omega\tau_y} + N_y(\omega) \\
Z(\omega) &= S(\omega)e^{-j\omega\tau_z} + N_z(\omega)
\end{align*}
\]

where \( \tau_x(x, y, z) \) denotes the signal arrival delay at microphone \( x \) with respect to the reference point located at the array origin \( o \). Here we can define the cross spectra of three microphone as shown in Eq. (12).

\[
\begin{align*}
G_{xy}(\omega) &= E[X^*(\omega)Y(\omega)] = P_x(\omega)e^{-j\omega\tau_{xy}} \\
G_{yx}(\omega) &= E[Y^*(\omega)X(\omega)] = P_y(\omega)e^{-j\omega\tau_{yx}} \\
G_{xz}(\omega) &= E[Z^*(\omega)X(\omega)] = P_z(\omega)e^{-j\omega\tau_{xz}}
\end{align*}
\]

The delay variables in Eq. (12) are the function of DOA \( \theta \) given by

\[
\begin{align*}
t_{xy}(\theta) &= \sqrt{3}R \sin(\theta + \frac{\pi}{3})/c \\
t_{yx}(\theta) &= \sqrt{3}R \sin(\theta)/c \\
t_{xz}(\theta) &= \sqrt{3}R \sin(\theta - \frac{\pi}{3})/c
\end{align*}
\]

where \( c \) denotes the sound velocity. Then, we form the frequency array data \( \mathbf{G}_{xy}(\omega_0, \theta), \mathbf{G}_{yx}(\omega_0, \theta) \) and \( \mathbf{G}_{xz}(\omega_0, \theta) \) by extracting the \( M \) harmonic components as shown in Sect. 3. For simplicity, we omit the \( \omega_0 \) in the following part of this paper.

4.2 Integration of Three Frequency Array Data

Now let us consider the difference of delay term (which determines the phase value) between two frequency array data for a signal propagating from direction \( \phi \).

\[
\begin{align*}
t_{xy}(\phi) &= t_{xy}(\theta) - t_{xy}(\phi) = 3R \sin(\phi - \frac{\pi}{3})/c \tag{16} \\
t_{yx}(\phi) &= t_{yx}(\theta) - t_{yx}(\phi) = 3R \sin(\phi + \frac{\pi}{3})/c \tag{17}
\end{align*}
\]

Then, we define the following \( M \times M \) diagonal matrices called rotation matrices that consist of the phase compensating components with respect to the signal from direction \( \phi \).

\[
\mathbf{G}_{xy}(\phi) = \text{diag} \left[ e^{-j\phi t_{xy}(\phi)} e^{-j\phi t_{yx}(\phi)} \cdots \right] \tag{18}
\]

\[
\mathbf{G}_{yx}(\phi) = \text{diag} \left[ e^{-j\phi t_{yx}(\phi)} e^{-j\phi t_{xy}(\phi)} \cdots \right] \tag{19}
\]

Using these rotation matrices, we define the following data called integrated array data.

\[
\mathbf{G}_m(\phi, \theta) = \left[ \mathbf{G}_{xy}(\phi)\mathbf{G}_{xy}(\theta) + \mathbf{G}_{yx}(\theta) \right] + \left[ \mathbf{G}_{xy}(\phi)\mathbf{G}_{xz}(\theta) \right]/3 \tag{20}
\]

It is noted that the phases of each terms in the right side of Eq. (20) are equal if and only if \( \phi = \theta \).

4.3 Subspace Analysis of Integrated Array Data Matrix

Here let us note the delay term of each rotated frequency array data in Eq. (20).

\[
\begin{align*}
\mathbf{G}_{xy}(\phi)\mathbf{G}_{xy}(\theta) &\rightarrow t_{xy}(\phi) + t_{xy}(\theta) = t_{xy}(\phi, \theta) \\
\mathbf{G}_{yx}(\theta) &\rightarrow t_{yx}(\theta) = t_{yx}(\phi, \theta) \tag{21} \\
\mathbf{G}_{xy}(\phi)\mathbf{G}_{xz}(\theta) &\rightarrow t_{xy}(\phi) + t_{xz}(\theta) = t_{xz}(\phi, \theta)
\end{align*}
\]

The sign “\( \rightarrow \)” above denotes to extract the delay term of a frequency array data. Now the following lemma is satisfied.

[Lemma]

The equation \( t_{xy}(\phi, \theta) = t_{yx}(\phi, \theta) = t_{xz}(\phi, \theta) \) is satisfied if and only if \( \phi = \theta \). (The proof is described in the Appendix A).

From this lemma, we can replace our DOA estimation problem by searching rotation matrices, which equalize the delay terms of all three frequency array data. To solve this subject, we can find the following theorem (see Appendix B for its proof).

[Theorem]

The integrated frequency array data \( \mathbf{G}_m(\phi, \theta) \) is equal to a steering vector \( \mathbf{s}(\phi) \) defined by

\[
\mathbf{s}(\phi) = \left[ e^{-j\phi t_{xy}(\phi)} e^{-j\phi t_{yx}(\phi)} \cdots \right]^T \tag{22}
\]

for \( \phi \) that satisfies \( t_{xy}(\phi, \theta) = t_{yx}(\phi, \theta) = t_{xz}(\phi, \theta) \), and vice versa. That is

\[
\mathbf{s}(\phi) = \mathbf{s}(\theta) \iff \mathbf{G}_m(\phi, \theta) = \mathbf{s}(\phi) \tag{23}
\]

This theorem leads the DOA estimation problem to search the parameter \( \phi \) satisfying the equality \( \mathbf{G}_m(\phi, \theta) = \mathbf{s}(\phi) \). In order to determine \( \phi \), we use the subspace structure
of the following covariance matrix \( \mathbf{R}_m(\phi) \) for \( \mathbf{G}_m(\phi, \theta) \).

\[
\mathbf{R}_m(\phi) = \mathbf{G}_m \mathbf{G}_m^H.
\]

Because \( \mathbf{R}_m(\phi) \) is an Hermitian matrix, each eigenvector \( \mathbf{v}_i \) of \( \mathbf{R}_m(\phi) \) is mutually orthogonal. Namely,

\[
\mathbf{v}_i^H \mathbf{v}_j = \delta_{ij},
\]

where \( \delta_{ij} \) is the Kronecker delta given by

\[
\delta_{ij} = \begin{cases} 
1 & i = j \\
0 & i \neq j 
\end{cases}.
\]

We also have

\[
\mathbf{R}_m(\phi) = \sum_{i=1}^{\tilde{M}} \lambda_i \mathbf{v}_i(\phi) \mathbf{v}_i^H(\phi).
\]

From the well-known theorem on array covariance matrix [12], the eigenvector \( \mathbf{v}_1 \) corresponding the largest eigenvalue \( \lambda_1 \) is equal to the vector \( \mathbf{G}_m \) in the case of rank-1 model. The estimated value \( \hat{\theta} \) is given by the following null search strategy.

\[
\hat{\theta} = \arg \max_{\theta} |P(\theta)|,
\]

where,

\[
P(\phi) = \frac{1}{\sum_{i=2}^{\tilde{M}} s^H(\phi) \mathbf{v}_i(\phi) \mathbf{v}_i^H(\phi) s(\phi)}.
\]

Figure 3 shows the flow diagram of the proposed method.

4.4 Uniform Spatial Resolution

Let us recall the array of two microphones in Fig. 2. The received signal defined by Eq. (3) and Eq. (4) are rewritten in matrix form as

\[
\mathbf{X}(\omega) = \mathbf{S}(\omega) \left[ \begin{array}{c} 1 \\ e^{-j\omega \tau} \end{array} \right] + \left[ \begin{array}{c} N_1(\omega) \\ N_2(\omega) \end{array} \right] = \mathbf{S}(\omega) \mathbf{s} + \mathbf{N}(\omega).
\]

The DOA information of the signal \( S(\omega) \) is contained in the steering vector \( \mathbf{s} \) that involves the phase difference \( \omega \tau \) between the two sensors. While the frequency \( \omega \) is determined, the phase term is a function of \( \theta \), therefore, the following relation is derived.

Due to the noise in practice, the phase value of each sensor signal deviates randomly. Eventually, \( \hat{\tau} \) in Eq. (31) is devi-
ated and this causes the DOA estimation error. For measuring the sensitivity of DOA estimation with respect to noise, we introduce noise robustness factor (NRF) \( I(\theta) \), which is defined by

\[
I(\theta) \equiv \frac{|d\theta|^{-1}}{|\cos \theta|}.
\]

We can find from \( I(\theta) \) that the estimation accuracy degrades as \( \theta \) being apart from the broadside direction.

On the other hand, the three pairs of microphones in the proposed method face at different directions of every \( \frac{\pi}{3} \) [rad] and they are integrated by summation. So averaging NRFs of three frequency array data as follows derives the NRF of the equilateral-triangular microphone array.

\[
I_{tr}(\theta) = \frac{1}{3} \left( I(\theta - \frac{\pi}{3}) + I(\theta) + I(\theta + \frac{\pi}{3}) \right).
\]

In Fig. 4, we show the NRF \( I(\theta) \) and \( I_{tr}(\theta) \). The \( I_{tr}(\theta) \) can keep its value nearly constant to omni-direction.

4.5 Discriminability for Opposite-Direction

Another advantage of the proposed method is the discriminability for a direction and its opposite. Now, let us consider a signal arriving from the back of the linear array, i.e. \( \pi > |\theta| > \frac{\pi}{2} \). Because \( \sin \theta = \sin(\pi - \theta) \), the linear array cannot discriminate whether the signal arrives from its front or the back. In contrast, the proposed method can discriminate the signal from omni-direction because the preceding theorem and lemma in Sect. 4.3 hold for \( \theta = [-\pi, \pi] \).

4.6 Suppressing Effects to the Reverberation

The integrated use of plural frequency array data in our

\[\text{IEICE TRANS. FUNDAMENTALS, VOL.E87–A, NO.3 MARCH 2004}\]
method is expected to be effective for suppressing the influence of reverberation. Usually, reverberation is known to be spatially diffuse due to the multiple reflection paths. On the other hand, each frequency array data is generated by the data acquired at different spatial position. From these facts, the reverberation components in each term of Eq. (20) are mutually uncorrelated. Thus, the integrated array data can be expected to possess an anti-reverberation effect.

5. Simulation and Experiment Results

5.1 Evaluation with Computer Simulation

For the computer simulation, we use the real 5 phoneme data (/a/, /e/, /i/, /o/, /u/) uttered by 10 subjects (5 each for male and female) as a source signal and take 5 trials for every data. As the conventional methods for comparison, we adopt our previous method [1] with linearly located 2, 3 and 4 microphones. In the case of 3 and 4 microphones, we use the average of multiple frequency array data before the covariance matrix derivation. Furthermore, we also compare with the MUSIC [5] with CSS [7] on the harmonics [15]. In the method, the pre-estimation is obtained by the beamformer method [16]. The parameters shown in Table 1 are adopted to every method. For the conventional methods, we use the same harmonics selected in the proposed method.

The spatial resolution is evaluated by the deviation of estimation error (DEE), which is given by

\[ DEE = \sqrt{\overline{[\hat{\theta}_i - \theta_T]^2}}, \]  

(35)

where \(\hat{\theta}_i\) and \(\theta_T\) are the estimated and true DOA respectively, and \(\overline{\cdot}\) means average for \(i\).

5.1.1 Evaluation for the Anechoic Case

We perform a numerical simulation using ideally generated microphone array input signals without reverberation. The microphone array input signal is virtually generated by delaying the signal with an appropriate samples according to \(\theta\), and sum up with additive white noise as a sensor noise.

Figure 5 shows the \(P(\phi)\) (called “spectrum”) of the proposed method and the spectra given by the conventional methods. The spectrum given by the proposed method shows the most prominent peak characteristic at the estimated angle. Figure 6 shows the DEEs for each method. In this simulation, we also compare with the MUSIC-CSS whose pre-estimated DOA involves random errors having Gaussian probability density function. From this result, we can recognize that the proposed method keeps its nearly constant spatial resolution to every direction, and its accuracy is better than that of the MUSIC-CSS with precise pre-estimated angle. Later in Fig. 12 (see the case \(\psi = 0\)), a nearly constant spatial resolution for omni-direction is shown for the proposed method.

5.1.2 Evaluation for a Simulated Reverberant Condition

For evaluating reverberation suppressing effect in the proposed method, we perform computer simulation using the room impulse responses simulated by the image method [17]. The room model for the simulated reverberant condition is summarized in Fig. 7, and for the reflection coefficients \(\beta\) in [17], we take them the values as shown in Table 2

![Fig. 5 Estimated spectra. (female /a/, \(\theta = 30^\circ\)](image)

![Fig. 6 Deviation of estimation error at ideally anechoic case. (solid line: Proposed, broken line: Conventional, dash dotted line: MUSIC-CSS)](image)

<table>
<thead>
<tr>
<th>Table 1 Parameters for simulation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input SNR</td>
</tr>
<tr>
<td>Sampling Frequency</td>
</tr>
<tr>
<td>Wave Velocity (c)</td>
</tr>
<tr>
<td>Array Aperture (R)</td>
</tr>
<tr>
<td>Threshold (T)</td>
</tr>
<tr>
<td>Window</td>
</tr>
<tr>
<td>FFT point</td>
</tr>
<tr>
<td>Frame Length</td>
</tr>
<tr>
<td>Frame Overlap</td>
</tr>
<tr>
<td>Data Length</td>
</tr>
</tbody>
</table>

\(\dagger\) The prominent peak of the spectrum results in the definiteness of the estimated DOA.
except for the values relating to ceiling and floor which are fixed at 0.5 in every case. We also denote the approximate reverberation time $T_R$ calculated using the given impulse response [18] in Table 2. The other parameters are the same as given in Table 1 except that the threshold $T$ is 10 dB. In Fig. 8, the DEEs for each method are shown. From these results, the proposed method keeps its accuracy and uniformity even for reverberation existence.

5.2 Experiments at Real Acoustic Environment

To verify that the proposed method is effective even at real acoustic environment, we performed some experiments at a conference room whose physical sizes are shown in Fig. 9. The speech data and parameters are the same as in the preceding computer simulation except for the input SNR lying around 18 dB and the threshold $T$ is 12 dB, and here we also made 5 trials for each data. We regard the mean value of the estimation results $\theta_{\text{MEAN}}$ shown in Table 3 as the true direction $\theta_T$ and evaluate the DEE around this value. Figure 10 shows the results of the experiment. This result shows that the proposed method provides the best resolution. In contrast, the MUSIC-CSS is crucially degraded by the existing pre-estimation error. It is noted that the resolution of the proposed method is nearly uniform for omni-direction (see the case $\psi = 0$ [deg] in Fig. 13).

5.3 Influence by the Elevation Angle Error

Although we assume that the speaker is located on the array plane, it may deviate from this plane in practice. Here we measure the influence occurred by the deviation of the elevation angle $\psi$ between the speaker’s location and array plane. Figure 11 and Fig. 12 show the influence of the elevation angle to the spectra and DEE, respectively. From these results, the elevation angle error below 10 degree degrades the estimation merely by a few degrees. Furthermore, this
In this contribution, we have proposed a DOA estimation method of a speech signal using microphones located at vertices of equilateral triangle. The proposed method uses the subspace analysis of the integrated frequency array data, and it achieves a uniform estimation resolution. Through both computer simulation and experiment in a real acoustic environment, we have confirmed that the satisfactory enhancement has been achieved by the proposed method. For a future subject, the estimation error and possibility for the elevation direction estimation should be considered. Another future study is to estimate directions of more than one simultaneous speaker.

Acknowledgement

This work is supported in part by a Grant in Aid for the 21st century Center Of Excellence for Optical and Electronic Device Technology for Access Network from the Ministry of Education, Culture, Sport, Science, and Technology in Japan.

References

Appendix A: Proof of Lemma

- $\mathfrak{r}_{xy}(\phi, \theta) = \mathfrak{r}_{yx}(\theta) \Rightarrow \phi = \theta$ and $\phi = -\frac{3}{2}\pi - \theta$
- $\mathfrak{r}_{zz}(-\frac{3}{2}\pi - \theta, \theta) \neq \mathfrak{r}_{xy}(-\frac{3}{2}\pi - \theta, \theta) = \mathfrak{r}_{yx}(\theta)$
- $\mathfrak{r}_{xy}(\theta) = \mathfrak{r}_{yz}(\phi, \theta) \Rightarrow \phi = \theta$ and $\phi = \frac{3}{2}\pi - \theta$
- $\mathfrak{r}_{xy}(\frac{3}{2}\pi - \theta, \theta) \neq \mathfrak{r}_{yx}(\frac{3}{2}\pi - \theta, \theta)$
- $\mathfrak{r}_{xz}(\phi, \theta) = \mathfrak{r}_{yz}(\phi, \theta) \Rightarrow \phi = \theta$ and $\phi = -\theta$
- $\mathfrak{r}_{yz}(\theta) \neq \mathfrak{r}_{xz}(-\theta, \theta) = \mathfrak{r}_{xy}(-\theta, \theta)$

Appendix B: Proof of Theorem

The magnitude of $k$-th element in $G_m$ is less than 1 not as far as all the interpolated delay terms are equal.

$$\|G_m\|_k = \left| e^{-jk\omega\tau_{xy}(\phi, \theta)} + e^{-jk\omega\tau_{yx}(\theta)} + e^{-jk\omega\tau_{yz}(\phi, \theta)} \right| / 3$$

$$\leq \left| e^{-jk\omega\tau_{xy}(\phi, \theta)} \right| + \left| e^{-jk\omega\tau_{yx}(\theta)} \right| + \left| e^{-jk\omega\tau_{yz}(\phi, \theta)} \right| / 3$$

$$= 1$$

The equality is satisfied only if the three complex values are equal.

Appendix C: Derivation of Eq. (33)

$$I(\theta) \equiv \left| \frac{d\theta}{d\theta} \right|^{-1}$$

$$= \left| \pm 1 \right|^{-1}$$

$$= \left| \frac{1}{\sqrt{1 - \theta^2}} \right|$$

$$= \left| \sqrt{1 - \sin^2 \theta} \right|$$

$$= |\cos \theta|$$