Localization of multiple environmental sound sources by MUSIC method with weighted histogram

Mari YAMAMOTO[1]; Yosuke TATEKURA[2]

[1] Shizuoka University, Japan [2] Shizuoka University, Japan

ABSTRACT

This paper presents an estimation method of environmental sound sources localization based on MUSIC method by small-sized microphone array. MUSIC spectrum depends on array manifold vector calculated from microphone array configuration and eigenvectors provided an observed signal. By considering the relation between a microphone array and MUSIC spectrum, we examine the arrangement of microphones for high accurate localization with fewer number of elements. And to estimate the environmental sound sources, which have various frequency properties, we propose a weighted histogram method for direction of arrival (DOA) estimation considering frequency properties of the observed signal. We introduce histogram obtained by detected peaks by MUSIC spectrum at each frequency bin. In addition, the sound source positions are estimated on intersection of the estimated directions which are obtained by a pair of microphone arrays. In the numerical simulation, the localization accuracy can be improved with reduced aliasing using a microphone array of irregular interspace. And, the proposed method can be improved the estimation accuracy for the electronic sound, which has sparse frequency property.

Keywords: DOA estimation, Environmental sound sources, Microphone array, MUSIC method, Weight histogram   I-INCE Classification of Subjects Number(s): 01.4, 74.7

1. INTRODUCTION

The needs of monitoring technique in daily living space are increasing for detecting users or conditions e.g. to watch elderly person of single life. An image sensor-based system, which is used many generally, can easily get conditions if an image includes dynamic changes such as people moving. However, it is difficult for the image sensor to detecting the abnormal noise of some kind of apparatuses or outbreak sounds from the blind spot of the camera. Thus, it is effective to utilize sound information positively to process more diversified information in monitoring. The purpose of this study is to estimate sound source localization using only sound information by small-sized microphone array.

Many of localization technique have been developed into a sound source signal using a speech signal for the purpose of a high-performance meeting system or a speech recognition system(1, 2). However, our living is surrounded by various environmental sounds including a machine drive sound, an electronic sound and so on. Therefore, it is necessary to perform these detection. An estimation accuracy depend on each sound source, because frequency properties of the environmental sound are different every signal unlike a speech signal.

In direction of arrival (DOA) estimation of multiple sound sources, multiple signal classification (MUSIC) method(3, 4) is well-known providing high spatial resolution. To estimate DOA with high accuracy, any studies have been proposed a method by large-sized device using a large number of microphones(5, 6). In this paper, we estimate DOAs and positions of sound sources using small-sized microphone array to apply MUSIC method. We focus on MUSIC method depending on an array manifold vector (AMV) provided from a microphone arrangement and eigenvector of the correlation matrix of the observed signal. From AMV properties, we consider on the effect to DOA estimation accuracy by the difference in the arrangement of microphone arrays. And, we examine the arrangement of microphones for high accuracy with fewer number of elements. In addition, we propose a DOA estimation method with a weighted histogram considering frequency properties of the observed signal.

[1] m-yamamo@spalab.eng.shizuoka.ac.jp
[2] tatekura@spalab.eng.shizuoka.ac.jp

In numerical simulation, we perform DOA estimation of multiple environmental sound sources and consider the relation between MUSIC spectrum and AMV properties. And, by the proposed method, we can expect that the detection rate is improved for sound sources which have sparse frequency property. Additionally, the sound source positions are estimated on the intersection of the estimated directions, which are obtained by a pair of microphone arrays. By distributing them, we can expect that the rate of position estimation improves.

2. METHOD OF LOCALIZATION

2.1 DOA Estimation by MUSIC Method

MUSIC method is a technique of DOA estimation based on subspace method. Assume that a set of M microphones observe N sources. Eigenvectors $e_m (m = 1,...,M)$ which can be obtained by eigenvalue decomposition of the correlation matrix of the observed signal are classified into signal subspace $E_s = [e_1 ... e_N]$ and noise subspace $E_n = [e_{N+1} ... e_M]$ by scale of eigenvalues. Signal subspace indicates the sound source direction and noise subspace is orthogonal to it. Using the orthogonality between eigenvectors e_i $(i = N+1,...,M)$ of noise subspace and AMV $a(\theta, f)$, MUSIC spectrum in the direction θ at arbitrary frequency f is given as follows:

$$P(\theta, f) = \frac{\|a(\theta, f)\|^2}{\sum_{i=N+1}^{M} |a^H(\theta, f) e_i(f)|^2}. \tag{1}$$

MUSIC spectrum shows a sharp peak when θ corresponds with true sound source direction. Here, we define a time delay between microphones τ_m and interval between 1st microphone and m-th microphone, d_m, AMV $a(\theta, f)$ with spatial information of signal sources is written as follows:

$$\begin{aligned} a(\theta, f) &= \begin{pmatrix} a_1 \\ \vdots \\ a_M \end{pmatrix} \\ &= \begin{pmatrix} e^{-j2\pi f \tau_1} \\ \vdots \\ e^{-j2\pi f \tau_M} \end{pmatrix}, &(2) \\ \tau_m &= d_m \sin(\theta_s)/\mathrm{c}. &(3) \end{aligned}$$

The estimation accuracy of MUSIC method depends on the arrangement of microphone array and the observed signal.

Since Eq. (1) is spectrum of narrow-band signal, wide-band MUSIC spectrum is averaged in all frequencies.

2.2 Expansion to Histogram

In wide-band MUSIC spectrum, the peak value strongly depends on quantity of frequency component of each sound source. Therefore, we adopt not the value of narrow-band MUSIC spectrum but directions of their peaks, and calculate histogram of the estimated direction.

We detect L peaks from narrow-band MUSIC spectrum at each frequency bin. Assume that the indicated directions of detected peaks are $\theta_l (l = 1,...,L)$. In k-th frequency bin, a peak determination is expressed as follow:

$$h_k(\theta) = \begin{cases} 1 & (\theta = \theta_l) \\ 0 & (otherwise). \end{cases} \tag{4}$$

Thus, the sum of detected peaks in the arbitrary direction θ is given as follows:

$$H(\theta) = \sum_{k=1}^{K} h_k(\theta). \tag{5}$$

DOA of the sound sources is estimated from direction indicated by peak of the histogram. Hence, the estimation not to depend on the value of peaks of narrow-band MUSIC spectrum is possible and it is expected that the sound source dependency of estimation accuracy is reduced.

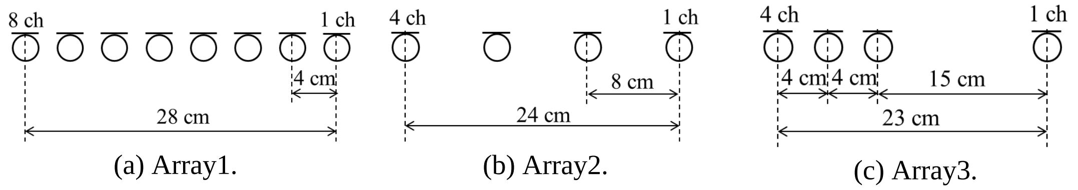


Figure 1 – The arrangement of microphone arrays.

We propose weighted histogram method using weight β_k prepared by k-th frequency bin. In the arbitrary direction θ, the weighted sum of detected peaks is given as follows:

$$H_{\mathrm{W}}(\theta) = \sum_{k=1}^{K} \{\beta_k h_k(\theta)\}. \tag{6}$$

An improvement of the detection rate is expected by the weight introducing from the observed signal.

2.3 Position Estimate Method

It is difficult to localize the sound sources in the distance or side positions by a small-sized microphone array because the aperture length of the microphone array is too short. Therefore, the localization accuracy can be improved by distributedly arrangement of a pair of microphone arrays. The sound source positions can be estimated by finding the intersection of the DOAs from each microphone array.

3. MICROPHONE ARRANGEMENT AND PROPERTY OF AMV

In this section, we evaluate the AMV properties of three kinds of microphone arrays and investigate the estimation accuracy by the MUSIC spectrum that is expected from their properties. Fig. 1 shows the arrangement of three microphone arrays. (a) and (b) are regular interspace arrays, which have difference number and interspace of elements each other. (c) is irregular interspace array, which is 15 cm between 1 ch and 2 ch, and 4 cm from 2 ch to 4 ch. The arrangement (c) is the same as Kinect sensor(7), which is a well-known small-sized device. The number of elements of (b) and (c) are equal, and the array aperture are also approximately equal.

3.1 Angular Resolution

In the DOA estimation, it is considered that the angular resolution of MUSIC spectrum depends on AMV. If variation of AMV by one degree is large, the angular resolution becomes higher.

The differences of AMV A_{sub} between direction θ ($-90^\circ \leq \theta \leq 90^\circ$) and direction next to each other in arbitrary frequency f are given as follows:

$$\begin{aligned} A_{\mathrm{sub}}(\theta, f) &= \frac{1}{2}\{a(\theta, f) - a(\theta-1, f) + a(\theta, f) - a(\theta+1, f)\} \\ &= \frac{1}{2M}\{\sum_{m=1}^{M} |(a_m(\theta, f) - a_m(\theta-1, f))| + \sum_{m=1}^{M} |(a_m(\theta, f) - a_m(\theta+1, f))|\}. \end{aligned} \tag{7}$$

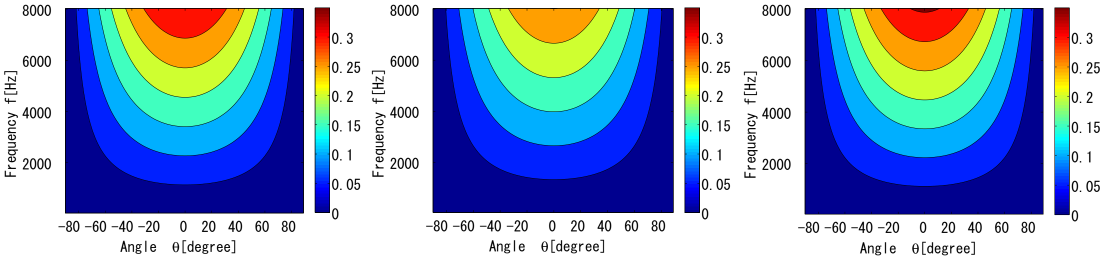
Fig. 2 shows the A_{sub} calculated from three microphone arrays. It indicates the angular resolution becomes higher in the central direction and high frequency in all cases. In comparison with Fig. 2 (a) and (b), the shorter the interspace between microphones or the more number of elements, the higher the angular resolution. However, in Fig. 2 (c), it is suppose that the microphone array with high angular resolution can be achieved by arranging microphone at irregular interspace even if the microphone elements are few.

Additionally, in all microphone arrays, A_{sub} is approximately zero in the band less than about 1000 Hz. Thus, it is suppose that to shape right peaks of MUSIC spectrum is difficult in this frequency band.

3.2 Correlation Coefficient

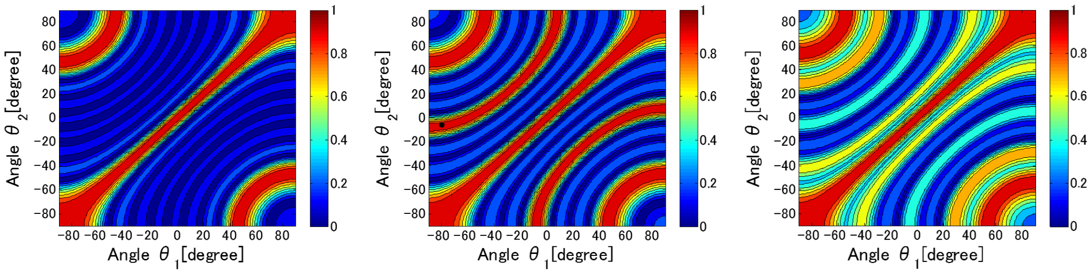
The peaks of MUSIC spectrum occur by orthogonality of AMV and noise subspace of the correlation matrix of the observed signal. So, it assumes that false estimated peaks are caused by high similarity for AMV that directions indicate right sound sources.

Correlation coefficient of AMV A_{cor} between direction θ_1 ($-90^\circ \leq \theta_1 \leq 90^\circ$) and direction θ_2 ($-90^\circ \leq$



(a) Array1. (b) Array2. (c) Array3.

Figure 2 – Variation of AMV with the directions of arrival.



(a) Array1. (b) Array2. (c) Array3.

Figure 3 – A correlation coefficient A_{cor} of AMV at 5000 Hz.

$\theta_2 \leq 90°$) in arbitrary frequency f is written as follows:

$$A_{\mathrm{cor}}(f) = \frac{a(\theta_1, f)^H \cdot a(\theta_2, f)}{\sqrt{\sum_{m=1}^{M} a_m(\theta_1, f)^2}\sqrt{\sum_{m=1}^{M} a_m(\theta_2, f)^2}}, \tag{8}$$

where $\cdot^H$ is complex conjugate transposition. Fig. 3 shows a correlation coefficient A_{cor} at 5000 Hz of three microphones, where diagonal from the upper right to the lower left in these figures mean autocorrelation coefficient. In Fig. 3 (a) and (b), which are microphone arrangement with regular interspace, tendency of the correlation value is bipolarized and extremely high correlation value occurred on the position of spatial aliasing. Therefore, it is supposed that the false peaks are concentrated at outbreak positions of the spatial aliasing in regular interspace array.

On the other hand, Fig. 3 (c) shows dispersion of the correlation value. It means that althogh the influence of the spatial aliasing is decrease, the influence of the side lobe is increase. So, directions of the false estimations are also dispersion in the microphone array with irregular interspace.

4. NUMERICAL EVALUATION

4.1 Evaluation of AMV Properties

In the numerical simulation, we performed the experimentation of sound sources DOAs estimation applied MUSIC method with three microphone arrays as shown in Fig 1.

4.1.1 Experimental Conditions

Fig. 4 shows experimental arrangement. And simulation conditions are listed as follows; the sampling frequency, the impulse response length, the flame length, the flame shift length and the evaluation band range are 16000 Hz, 2048 points, 512 points, 32 points and 1000–8000 Hz, respectively. The room reverberation is considered only primary reflection.

The sound sources are used five kinds from RWCP Sound Scene Database in Real Acoustical Environments(8); toy, pan, buzzer, pipong and phone. These frequency properties are different from each other, for example, toy has large power in wide-band while phone has few frequency components.

The number of sound sources assumes two or three, and all sources are arranged to the distance of 150 cm from center of Microphone Array 1 in Fig. 4. Sound sources arrangements include 23 patterns by two sources or 25 patterns by three sources, and we performed with all combination by five kinds of sounds. The volume

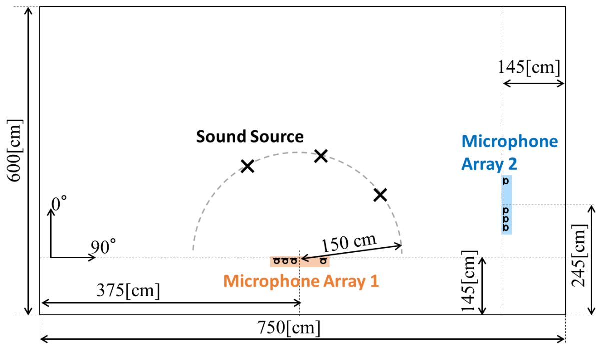
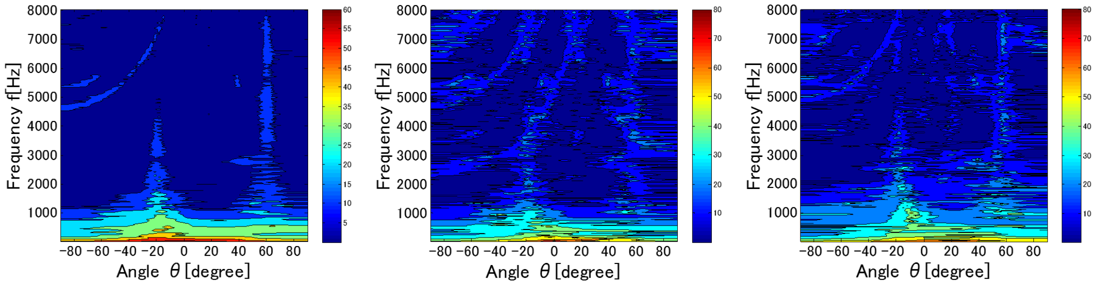


Figure 4 – Experimental arrangement.



(a) Array1. (b) Array2. (c) Array3.

Figure 5 – The power distribution maps of MUSIC spectrum at every frequency bin of three sources, where directions of arrival are $60°, 40°$ and $-20°$, respectively.

of all sounds are equalized in 4 ch of Array1 or in 2 ch of Array2 and Array3.

The estimated DOAs are calculated from peaks of wide-band MUSIC spectrum which are additionally averaged of narrow-band MUSIC spectrum in the evaluation band range. In this paper, the angular resolution of narrow-band MUSIC spectrum is set to 3°, so in evaluation of the accuracy, we assume that an estimation succeeded if an error was less than 3°.

4.1.2 Results of DOA Estimation

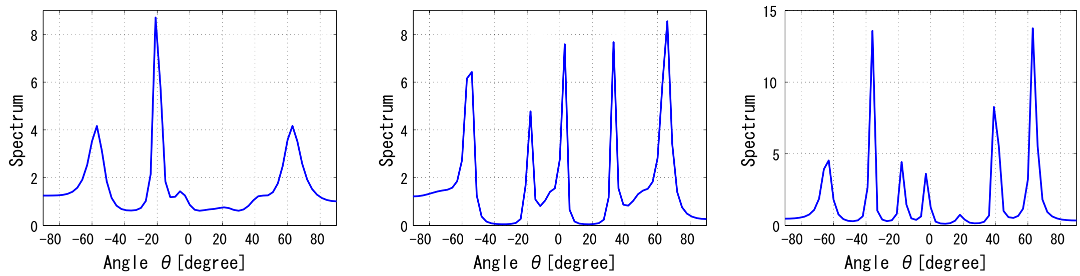
We evaluate the accuracy of DOA estimation from results using Microphone Array 1 of Fig. 4.

Fig. 5 and Fig. 6 shows power distribution maps of MUSIC spectrum at every frequency bin by three sources and MUSIC spectra of three sources at 5000 Hz, where the sound sources are buzzer, phone and pan, and each directions is $60°, 40°$ and $-20°$, respectively. In any microphone arrays, it indicates that right DOA can not be estimated less than about 1000 Hz from the result of Fig. 5. Using the microphone array with regular interspace, the false estimation peaks, which indicated high value, are occurred on the spatial aliasing positions. On the other hand, using Array3, they are dispersion by frequency bin and do not have regularity. In comparison with Fig. 6 and Fig. 3, the false estimation peaks occur on the directions which have high correlation of AMV with right direction. From these results, we confirm that MUSIC spectrum strongly depends on AMV.

Fig. 7 shows the averaged accuracy rate of direction estimation value at each sound source, where (a) and (b) are estimation results of two sources and three sources, respectively. A tendency of dependence to sound sources is similarity in all arrays, the accuracy rate becomes low when the frequency components of the sound source are sparse. In almost sound sources, the accuracy rate of Array3 is higher than that of Array2. This reason is considered that the right peaks can detect easily for dispersion of the false peak directions. Thus, irregular interspace arrangement can improve the DOA estimation accuracy.

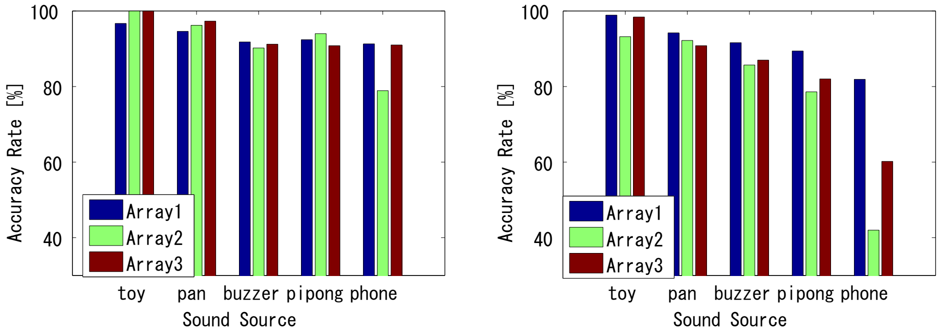
4.2 Evaluation by Weighted Histogram Method

In the numerical simulation, we perform the experimentation of sound sources DOA estimation and positions estimation using three methods; MUSIC method, the conventional histogram method and the proposed weighted histogram method.



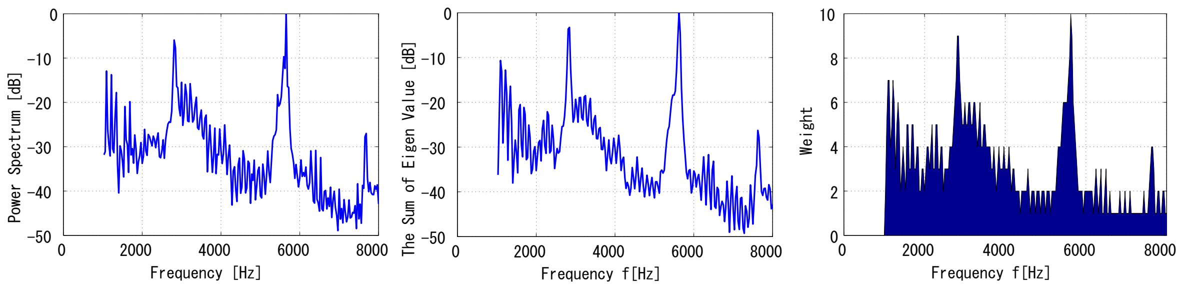
(a) Array1. (b) Array2. (c) Array3.

Figure 6 – MUSIC spectra of three sources at 5000 Hz, where directions of arrival are $60^{\circ}, 40^{\circ}$ and -20°, respectively.



(a) 2 sources. (b) 3 sources.

Figure 7 – The averaged accuracy rate of direction estimation value at each sound source.



(a) Power Spectrum. (b) The Sum of Eigen Value. (c) Weight.

Figure 8 – Acquiring the weight using observed signal, in which the sound sources are buzzer and phone.

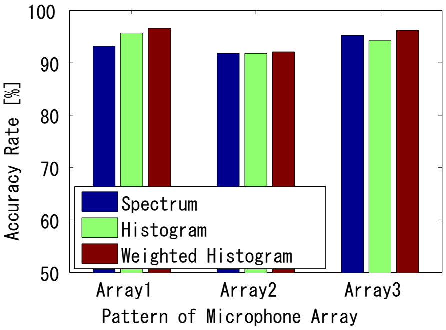
4.2.1 Experimental Conditions

The experimental conditions and arrangement are the same as subsection 4.1.1. In the position estimation, a distance error of the intersection which is closest to the right position in multiple points calculated using estimated DOA is evaluated. We assume that an estimation succeeded when the distance error was less than 15 cm.

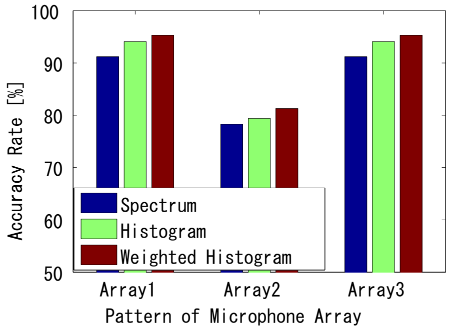
In this study, the weight is obtained from eigenvalue distribution of the correlation matrix by the observed signal. The sum of eigenvalue is divided into 10 steps between maximum value and minimum value at each frequency bin, and convert the 10 steps to the weight from $+10$ to $+1$. Fig 8 shows power spectrum of the observed signal, the sum of eigenvalue and the calculated weight, respectively, in which buzzer and phone is used as sound sources. It is found that the weight depends on frequency properties of the observed signal.

4.2.2 Results of DOA Estimation

Fig 9 and Fig 10 show the averaged accuracy of direction estimation value using three microphone arrays at each method and using proposed method at sound sources, respectively, where (a) and (b) are estimation

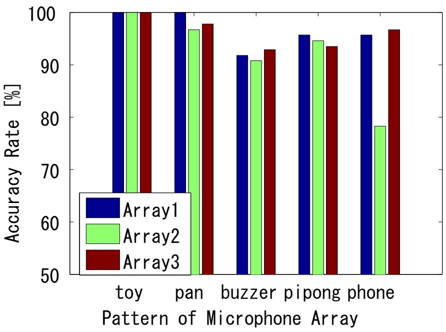


(a) 2 sources.

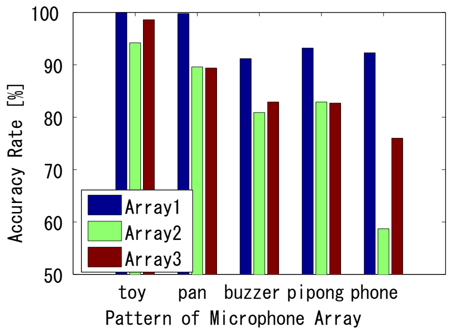


(b) 3 sources.

Figure 9 – The averaged accuracy rate of the direction estimation value at each method.



(a) 2 sources.



(b) 3 sources.

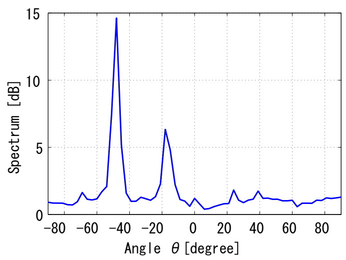
Figure 10 – The averaged accuracy rate of the direction estimation value using weighted histogram method at each sound sources.

results of two sources and three sources, respectively. From the result of Fig 9, the effectivity of the proposed method is proved because the most highest DOA accuracy is indicated in all array cases. Although the number of elements of Array3 is fewer than that of Array1, the estimation accuracies are almost equal. Additionally, in Fig 10, though the accuracy of Array2 for phone is greatly decrease, it was improved in Array3 case. In these results, the microphone with irregular interspace arrangement can decrease the influence of the sound source dependence in all methods.

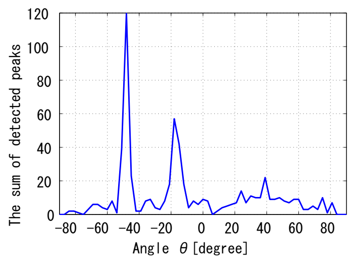
Fig. 11 shows examples of DOA estimation using Array3, where the sound source arrangement is phone, buzzer and toy at $40^\circ, -20^\circ$ and -50°, respectively. (a), (b) and (c) are shown MUSIC spectrum, the conventional histogram and the proposed weighted histogram, respectively. Although MUSIC spectrum cannot detect a peak of phone because of the false estimation peak, the detection becomes possible by using the histogram. In addition, in comparison with (b) and (c), the peak using the proposed method becomes higher relatively. Therefore, by using the proposed weight, the peaks of sound sources which have few frequency components such as phone can be detected easily. Fig. 12 shows the averaged accuracy rate of direction estimation value using Array3 at each sound source, where (a) and (b) are estimation results of two sources and three sources, respectively. In arrangement of three sources, the accuracy rate of phone is 60.2 % at MUSIC spectrum and 63.2 % at the conventional histogram method but the proposed method is 76.0 % and it is improved over 10 %. However, in some sound sources such as pan and buzzer, the accuracy rates of the conventional histogram method are higher than those of the proposed method. This reason is considered that the effect does not provided for sound sources which have power in low frequency, because band of less than 1000 Hz is excluded from the evaluation in this experiment.

4.2.3 Results of Position Estimation

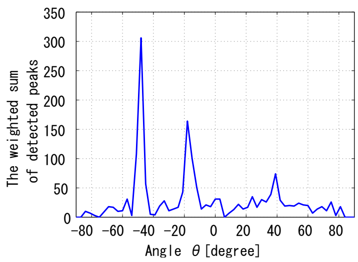
Fig 13 shows the position estimation accuracy rate at each method, where (a) and (b) are estimation results of two sources and three sources, respectively. In comparison with accuracy, Array1 is the highest and Array2 is the lowest. There are the same as results of DOA estimation in all methods. And, the proposed method is the most highest accuracy rate in all microphone array. From these results, the position estimation accuracy depends on the DOA estimation accuracy. However, the position estimation rate is significantly reduced in



(a) MUSIC Spectrum.

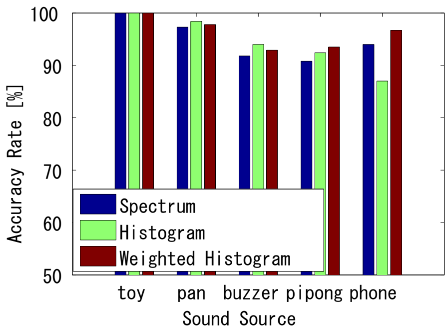


(b) Conventional Histogram.

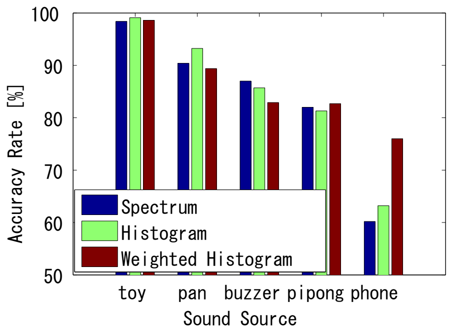


(c) Weighted Histogram.

Figure 11 – Result of the direction estimate using Array3, where directions of arrival are 40°, −20° and −50°, respectively.

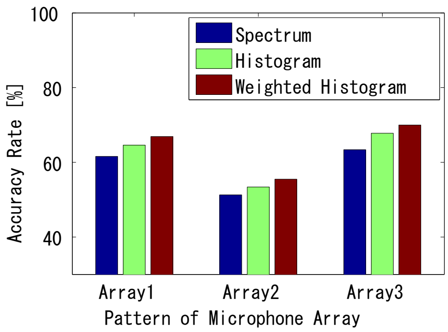


(a) 2 sources.

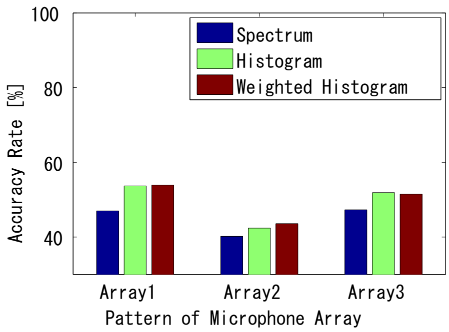


(b) 3 sources.

Figure 12 – The averaged accuracy rate of the direction estimation value using Array3 at each sound source.



(a) 2 sources.



(b) 3 sources.

Figure 13 – The localization accuracy rate at each method.

comparison with the rate of DOA estimation. In this paper, the arrangement of multiple sound sources are set from Microphone Array 1 of Fig. 4. So, there is too near angle of the arrangement each source from Microphone Array 2. This is the reason of the accuracy degradation. Thus, to improve the position estimation accuracy, it is necessary to improve angular resolution of DOA estimation.

In addition, the experiments are evaluated of the accommodation limit which select the most nearest point to right position in multiple intersections. Therefore, it is necessary that a new method achieves to select right intersections automatically by clustering the DOA estimations of the same sound source.

5. CONCLUSIONS

In this paper, for the sake of localization of multiple environmental sound sources, we focus on the MUSIC spectrum depending on AMV and eigenvector of the correlation matrix of the observed signal. We investigate AMV properties of the microphone arrays, which have different arrangements, and examine about the arrangement of microphones for high quality localization. We have confirmed that the angular resolution of

AMV is higher in the central direction and high frequency, and irregular arrangement of microphones can be prevent from degradation of the spatial resolution by the fewer elements and suppress the spatial aliasing. In addition, we propose the DOA estimation method by the weighted histogram which is considered frequency properties of the observed signal.

In the numerical simulation, we have confirmed that MUSIC spectrum are reflected the AMV properties. Thus, it is considered that the estimation accuracy can be improved by using microphone array by weighting from the AMV informations. The improvement of the estimation accuracy of an electrical sound can be possible using the weighted histogram method. Additionally, we have confirmed that irregular interspace arrangement of microphones can be decrease sound source dependency. However, it is necessary to improve angular resolution of DOA estimation because the localization rate is significantly degraded. And to localize multiple sound sources at the same time, it is necessary to distinguish the kind of the sound source between two microphone arrays.

REFERENCES

1. Maeda S, Someya T, Suyama K. Multiple Sound Source Localization Using Reliability Weighted Distribution Bsed on Sparseness of Speech Signals. IEICE Trans Fundam (J). 2012;J95-A(3):247–260.

2. Nagata Y, Iwasaki S, Hariyama T, Horiguchi H, Fujioka T, Abe M. Step by Step DOA Estimation if Multiple Sound Sources Based on MUSIC with Two-Channel Input. IEICE Trans Fundam (J). 2009;J92-A(11):864–873.

3. Schmidt RO. Multiple Emitter Location and Signal Parameter Estimation. IEEE Transactions on Antennas and Propagation. 1986;34(3):276–280.

4. Benesty J, Chen J, Huang Y. Microphone Array Signal Processing. Springer-Verlag;2008.

5. Oakamoto T, Nishimura R, Iwaya Y. Estimation of sound source positions using a surrounding microphone array. J Acoust Soc Jpn. 2007;28(3):181–189.

6. Nishiura T, Yamada T, Nakamura S, Shikano K. Localization of Multiple Sound Sources Based on CSP Analysis with a Microphone Array. IEICE Trans Inf & Syst (J). 2000;J83-D(8):1713–1721.

7. http://www.xbox.com/ja-JP/

8. http://research.nii.ac.jp/src/en/RWCP-SSD.html