ESTIMATION OF THE NUMBER OF SOURCES AND THEIR LOCATIONS IN COLORED NOISE USING REVERSIBLE JUMP MCMC

Futoshi Asano[1,2]*, Hideki Asoh*[1] *and Kazuhiro Nakandai*[2]

National Institute of Advanced Industrial Science and Technology (AIST)[1],
Honda Research Institute Japan Co., Ltd (HRI-JP)[2]

ABSTRACT

The authors proposed a method of estimating the source locations in colored noise using a hierarchical model in Bayesian estimation framework. In this paper, the reversible jump MCMC method is introduced into this method to jointly estimate the number of sources. By introducing this, the cases in which the number of sources is unknown or dynamically changes can be handled. The results of the experiments show an improvement over the conventional methods in terms of the source localization performance in a room reverberation.

Index Terms— source localization, Bayesian estimation, hierarchical model, reversible jump MCMC

1. INTRODUCTION

For source localization in spatially colored noise such as reverberation of rooms, the performance of the conventional estimators such as the maximum likelihood (ML) method or the MUSIC method is often reduced [1]. The authors proposed a method of joint estimation of the noise covariance and the source location using a hierarchical model in a Bayesian estimation framework [2]. By using the hierarchical model, the common structure of the covariance can be extracted, and therefore, the stable estimate of the covariance can be obtained from a smaller amount of data. A problem in this method is that the order of the model, i.e., the number of sources, must be known in advance. In real applications, the number of sources is often unknown. Moreover, even if the number of *physical* sources is known, the number of *active* sources may dynamically change for a source signal such as speech due to its sparseness in the frequency domain. Therefore, the number of sources must be jointly estimated.

In the Bayesian framework, a method for jointly estimating the model order, the reversible jump MCMC (Markov chain Monte Carlo) method, was proposed by Green [3] and was applied to the estimation of the frequency of sinusoids by Andrieu *et al.* [4]; this is essentially the same as the source localization problem. In this paper, the authors introduce this method into their hierarchical model and examine its performance using the simulated and actually recorded data.

2. JOINT ESTIMATION USING HIERARCHICAL MODEL

In this section, the joint estimation framework proposed by the authors [2] is briefly reviewed to facilitate an understanding of the following sections.

2.1. Model of signal/noise

The observation vector is assumed to be modeled as

$$\boldsymbol{z}_{j,k} = \boldsymbol{A}_j(\boldsymbol{\theta}_j)\boldsymbol{s}_{j,k} + \boldsymbol{v}_{j,k} \tag{1}$$

where the mth element of $\boldsymbol{z}_{j,k}$ denotes the short-time Fourier transform (STFT) of the mth sensor input at the time frame index k. The symbol j denotes the index for the time block that consists of K observations as $\boldsymbol{Z}_j = [\boldsymbol{z}_{j,1}, \cdots, \boldsymbol{z}_{j,K}]$. The symbol $\boldsymbol{A}_j(\boldsymbol{\theta}_j)$ denotes the array manifold matrix. The source direction $\boldsymbol{\theta}_j = [\theta_{j,1}, \cdots, \theta_{j,N_j}]^T$ within the block is assumed to be invariant. The symbols $\boldsymbol{s}_{j,k}$ and $\boldsymbol{v}_{j,k}$ are the source vector and noise vector, respectively. The covariance matrix is assumed to be modeled as

$$\boldsymbol{R}_j = E[\boldsymbol{z}_{j,k}\boldsymbol{z}_{j,k}^H] = \boldsymbol{A}_j\boldsymbol{\Gamma}_j\boldsymbol{A}_j^H + \boldsymbol{K}_j \tag{2}$$

where $\boldsymbol{\Gamma}_j = E[\boldsymbol{s}_{j,k}\boldsymbol{s}_{j,k}^H]$ and $\boldsymbol{K}_j = E[\boldsymbol{v}_{j,k}\boldsymbol{v}_{j,k}^H]$. The symbols M and N_j denote the number of sensors and sources, respectively.

2.2. Joint estimation framework

The parameters to be estimated are $\boldsymbol{\Theta}_j = \{\boldsymbol{\theta}_j, \boldsymbol{K}_j, \boldsymbol{S}_j, N_j\}$, for $j = 1, \cdots, J$ where $\boldsymbol{S}_j = [\boldsymbol{s}_{j,1}, \cdots, \boldsymbol{s}_{j,K}]$. The estimation of N_j is introduced in Section 3, and thus is omitted in this section. The parameters $\{\boldsymbol{\theta}_j, \boldsymbol{K}_j, \boldsymbol{S}_j\}$ are estimated using the combination of Gibbs sampling and the Metropolis algorithm [5] in each time block. Then, the prior of $\{\boldsymbol{K}_j; j = 1, \cdots, J\}$ is estimated using the hierarchical model. The assumed sampling model is:

$$\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J \sim \text{i.i.d. inverse-Wishart}(\nu_0, (\nu_0\boldsymbol{K}_0)^{-1}) \tag{3}$$

where ν_0 and $\boldsymbol{K}_0$ are the parameters of the inverse-Wishart distribution to be estimated. The procedure for the joint estimation is summarized as follows:

© EURASIP, 2012 - ISSN 2076-1465

1. Set $\{\boldsymbol{K}_j^{(1)}\}$, $\{\boldsymbol{\theta}_j^{(1)}\}$, $\boldsymbol{K}_0^{(1)}$, $\nu_0^{(1)}$ and $p = 1$.
2. Sample $\boldsymbol{s}_{j,k}^{(p+1)} \sim p(\boldsymbol{s}_{j,k}|\boldsymbol{Z}_j, \boldsymbol{\theta}_j^{(p)}, \boldsymbol{K}_j^{(p)}) \quad \forall j, k$
3. Sample $\boldsymbol{K}_j^{(p+1)} \sim p(\boldsymbol{K}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{\theta}_j^{(p)}) \quad \forall j$
4. Sample $\boldsymbol{\theta}_j^{(p+1)}$ as: $\quad \boldsymbol{\theta}_j^* \sim J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)}) \quad \forall j$
$$\boldsymbol{\theta}_j^{(p+1)} = \begin{cases} \boldsymbol{\theta}_j^* & r > r_{thr} \\ \boldsymbol{\theta}^{(p)} & \text{otherwise} \end{cases}$$
5. Sample $\boldsymbol{K}_0$ as:
$$\boldsymbol{K}_0^{(p+1)} \sim p(\boldsymbol{K}_0|\boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)}, \nu_0^{(p)})$$
6. Sample ν_0 as:
$$\nu_0^{(p+1)} \sim p(\nu_0|\boldsymbol{K}_0^{(p+1)}, \boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)})$$
7. Go back to Step 2 with $p \leftarrow p + 1$.

Here, $\cdot^{(p)}$ denotes the index for the iteration. The symbol $J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)})$ denotes the proposal distribution of $\boldsymbol{\theta}_j$ given $\boldsymbol{\theta}_j^{(p)}$. The symbol r is the acceptance ratio defined as:

$$r := \frac{p(\boldsymbol{\theta}_j^*|\boldsymbol{Z}_j, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{K}_j^{(p+1)})}{p(\boldsymbol{\theta}_j^{(p)}|\boldsymbol{Z}_j, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{K}_j^{(p+1)})} \tag{4}$$

The symbol r_{thr} is an appropriate threshold. For the concrete distribution for the sampling and other details, see [2].

3. ESTIMATION OF NUMBER OF SOURCES USING REVERSIBLE JUMP MCMC

3.1. Reversible jump MCMC

In this section, the reversible jump MCMC method[3] is introduced in the joint estimation frame work described in the previous sections. In this paper, the following three moves [4] were employed in Step 4 of the joint estimation procedure described in Section 2.2:

- **Birth**:
  1. Increase the number of sources: $N_j^* = N_j^{(p)} + 1$
  2. Propose a new source with the location θ_{j,N_j^*} randomly selected from the possible locations.
  3. Evaluate the acceptance ratio r described in Section 3.2
  4. If $r > r_{thr}$, $N_j^{(p+1)} = N_j^*$ and $\boldsymbol{\theta}_j^{(p)} = \boldsymbol{\theta}_j^* (= [\boldsymbol{\theta}_j^{(p)},\ \theta_{j,N^*}])$.
- **Death**:
  1. Decrease the number of sources: $N_j^* = N_j^{(p)} - 1$.
  2. Eliminate one of the sources randomly from $\boldsymbol{\theta}_j^{(p)}$ to yield $\boldsymbol{\theta}_j^*$.
  3. Evaluate the acceptance ratio r.
  4. If $r > r_{thr}$, $N_j^{(p+1)} = N_j^*$ and $\boldsymbol{\theta}_j^{(p+1)} = \boldsymbol{\theta}_j^*$.
- **Update**:
  1. Conduct Step 4 in Section 2.2

One of these three moves is selected randomly during the iteration. The threshold r_{thr} can be obtained as $r_{thr} \sim \mathcal{U}(0, 1)$, where $\mathcal{U}()$ denotes the uniform distribution.

3.2. Acceptance ratio

The acceptance ratio in reversible jump MCMC is defined as [3, 4]

$$r = \text{posterior ratio} \times \text{proposal ratio} \tag{5}$$

In this paper, the proposal ratio is assumed to be unity for the sake of simplicity. Thus, the acceptance ratio is given by the same expression as (4). However, $\boldsymbol{S}_j^{(p+1)}$ cannot be used in (4) since the dimension of $\boldsymbol{S}_j^{(p+1)}$ is changed by the move. Therefore, $\boldsymbol{S}_j^{(p+1)}$ must be eliminated from (4) using integration.

Assuming that $\boldsymbol{\theta}_j$ has uniform prior distribution and is independent of $\boldsymbol{S}_j$ and $\boldsymbol{K}_j$, the full conditional distribution of $\boldsymbol{\theta}_j$ is given by

$$\begin{aligned} p(\boldsymbol{\theta}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) &\propto p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) p(\boldsymbol{\theta}_j|\boldsymbol{S}_j, \boldsymbol{K}_j) \\ &\propto p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) \end{aligned} \tag{6}$$

Assuming that the noise $\boldsymbol{v}_{j,k}$ has a complex Gaussian distribution, the likelihood $p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j)$ is given by

$$\begin{aligned} p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) &= \pi^{-MK}|\boldsymbol{K}_j|^{-K} \times \qquad (7) \\ \exp&\left\{ -\sum_{k=1}^{K} (\boldsymbol{z}_{j,k} - \boldsymbol{A}_j \boldsymbol{s}_{j,k})^H \boldsymbol{K}_j^{-1} (\boldsymbol{z}_{j,k} - \boldsymbol{A}_j \boldsymbol{s}_{j,k}) \right\} \end{aligned}$$

From the integration of $\boldsymbol{S}_j$ and omitting the unnecessary terms, (7) becomes

$$\begin{aligned} p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{K}_j) &= \int p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) d\boldsymbol{S}_j \\ &\propto |\boldsymbol{K}_j|^{-K} |\boldsymbol{\Sigma}_j|^K \exp\{-\mathrm{tr}\,[\boldsymbol{C}_j \boldsymbol{P}_j]\} (8) \end{aligned}$$

where

$$\boldsymbol{C}_j := \sum_{k=1}^{K} \boldsymbol{z}_{j,k} \boldsymbol{z}_{j,k}^H \tag{9}$$

$$\boldsymbol{\Sigma}_j := \left(\boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{A}_j \right)^{-1} \tag{10}$$

$$\boldsymbol{P}_j := \boldsymbol{K}_j^{-1} - \boldsymbol{K}_j^{-1} \boldsymbol{A}_j \boldsymbol{\Sigma}_j \boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \tag{11}$$

Table 1. Parameters for analysis.

Parameter	Value
Sampling frequency	16 kHz
Frame length (STFT lengthj	512 points@
Frame shift	128 points
Block length	2.0 s (Exp. I) / 0.2 s (Exp. II and III)
Number of iterations	1000
Frequency	1500 Hz (Exps. I and II) / 1400-2100Hz (Exp. III)

From these, the logarithm of the acceptance ratio becomes

$$\begin{aligned}\log r &= \log p(\boldsymbol{\theta}_j^*|\boldsymbol{Z}_j, \boldsymbol{K}_j) - \log p(\boldsymbol{\theta}_j^{(p)}|\boldsymbol{Z}_j, \boldsymbol{K}_j) \\ &\propto K\left(\log|\boldsymbol{\Sigma}_j^*| - \log|\boldsymbol{\Sigma}_j^{(p)}|\right) \\ &\quad - \left(\mathrm{tr}\left[\boldsymbol{C}_j\boldsymbol{P}_j^*\right] - \mathrm{tr}\left[\boldsymbol{C}_j\boldsymbol{P}_j^{(p)}\right]\right) \qquad (12)\end{aligned}$$

As shown in Section 4, the dependency of $|\boldsymbol{\Sigma}_j^*|$ on $\boldsymbol{\theta}_j^*$ is low compared to that of $\mathrm{tr}\left[\boldsymbol{C}_j\boldsymbol{P}_j^*\right]$ [1]. Thus, when the number of sources is unchanged by the move, $\log|\boldsymbol{\Sigma}_j^*| \simeq \log|\boldsymbol{\Sigma}_j^{(p)}|$. In this case, the acceptance ratio $\log r$ is mainly determined by the second term, $-\left(\mathrm{tr}\left[\boldsymbol{C}_j\boldsymbol{P}_j^*\right] - \mathrm{tr}\left[\boldsymbol{C}_j\boldsymbol{P}_j^{(p)}\right]\right)$.

When $N_j^* > N_j^{(p)}$ (birth move), this second term tends to increase. This can be understood by the fact that the likelihood, in general, increases as the degree of freedom in the model increases. Therefore, when the model order is determined based on the likelihood, a "penalty" term is usually introduced as in AIC/MDL. In (12), the first term, $K\left(\log|\boldsymbol{\Sigma}_j^*| - \log|\boldsymbol{\Sigma}_j^{(p)}|\right)$ functions as a penalty. The determinant of $\boldsymbol{\Sigma}_j^*$ can be decomposed using its eigenvalues as:

$$\log|\boldsymbol{\Sigma}_j^*| = \sum_{i=1}^{N_j^*} \log\lambda_i \qquad (13)$$

where $\{\lambda_i\}$ denotes the eigenvalues. By appropriate scaling of data $\boldsymbol{Z}_j$, the eigenvalues can also be scaled as $\lambda_i < 1, \forall i$. Thus, when $N_j^* > N_j^{(p)}$, the first term in (12) decreases. Similarly, in the case of the death move, the first term increases.

In the first term of (12), K can be viewed as the factor that controls the magnitude of the penalty, and an appropriate value should be selected. However, K is the number of frames in a block and is usually determined by the application. Therefore, in the proposed method, instead of using the actual K, it is replaced by the arbitral constant κ in (12). The value of κ is optimized experimentally in Section 4.

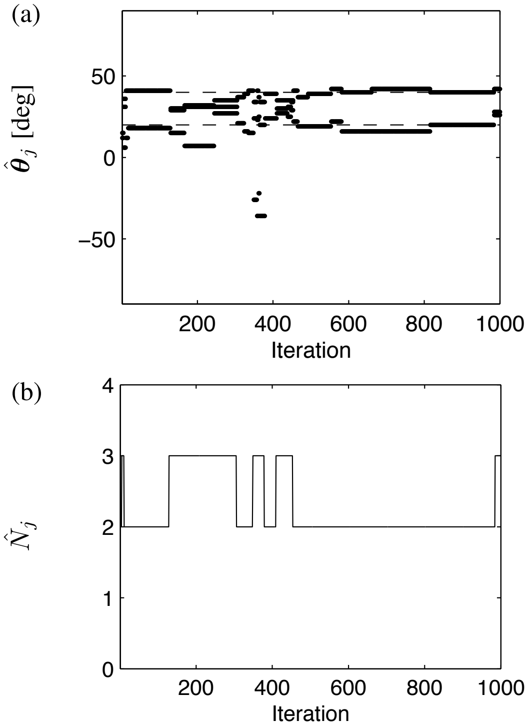


Fig. 1. Variation of $\boldsymbol{\theta}_j^{(p)}$ and $\hat{N}_j$. The dotted line in (a) shows the true $\boldsymbol{\theta}_j$.

4. EXPERIMENT

4.1. Exp. I - base line

The introduction of the reversible jump MCMC method into the joint estimation framework is examined first in an ideal condition with the simulated data. A microphone-array input is generated by convolving the room impulse responses measured in a meeting room (room size:8 m × 9 m × 3 m, reverberation time $\simeq$ 0.5 s) with the source signal (Gaussian noise). The true number of sources N_j is selected from $\{1, 2, 3\}$. The angular distance between the sources is 20° and the distance between the sources and the microphone array is 1.5 m. An 8-element microphone array mounted on the head of a robot is used. To focus on the performance of the joint estimation of $\boldsymbol{\Theta}_j$ including N_j, a longer block length, 2.0 s, is chosen. In this case, the hierarchical modeling of $\boldsymbol{K}_j$ is not necessary and is not employed in Exp. I since sufficient amount of data is available. Thus the number of blocks is set as $J = 1$. The parameters for signal analysis are summarized in Table 1. The estimate $N_j^{(p)}$ is also selected from $\{1, 2, 3\}$.

Fig. 1 shows an example of the variations in $\boldsymbol{\theta}_j^{(p)}$ and $N_j^{(p)}$ during the iterations. As the final estimate $\hat{N}_j$, $N_j^{(p)}$ with the highest frequency is employed. Then, $\{\boldsymbol{\theta}_j^{(p)}\}$ with $N_j^{(p)} = \hat{N}_j$ is averaged.

[1] The exception is the case where two or more of the sources are identical and the column of $\boldsymbol{\Sigma}_j^*$ is linearly dependent. This case is eliminated in the iteration.

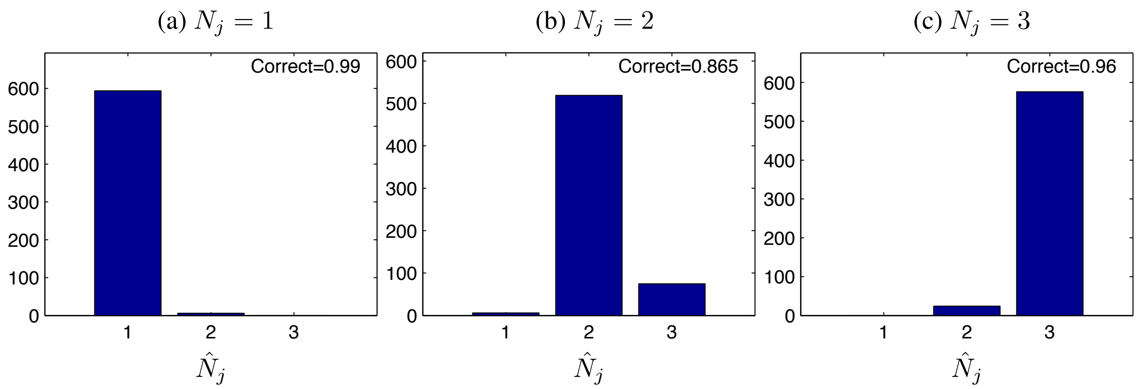


Fig. 2. Histogram of $\hat{N}_j$ for different true N_j in Exp. I.

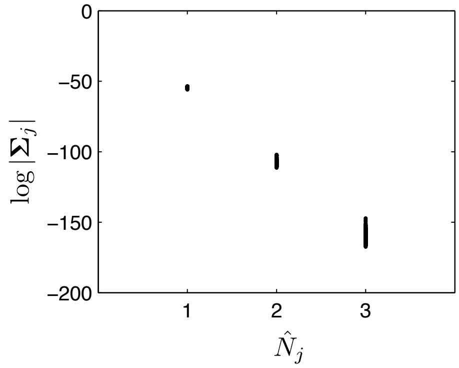


Fig. 3. Value of $\log|\mathbf{\Sigma}_j|$ for different $\hat{N}_j$.

Table 2. MAE for Exp. I.

	$N_j = 1$	$N_j = 2$	$N_j = 3$
MAE	13.01	2.89	7.54
C4	0.76	0.80	0.39
C8	0.77	0.95	0.70

Fig. 2 shows the histogram of $\hat{N}_j$ for 600 trials. From this, it can be seen that the correct $\hat{N}_j$ is estimated with high probability.

Table 2 shows the MAE for different true N_j. MAE is defined as $(1/N_{trial}) \sum_t |\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j|$, where t and N_{trial} indicate the index and the total number of trials, respectively. The values of C4 and C8 indicate the probabilities of $\text{MAE} \leq 4°$ and $\text{MAE} \leq 8°$, respectively. In this table, the error for $N_j = 1$ is relatively high. By increasing κ, the error for $N_j = 1$ can be reduced, resulting in an increase in the MAE for $N_j = 3$ instead.

Fig. 3 shows the value of $\log|\mathbf{\Sigma}_j|$ for different $\hat{N}_j$. From this, it can be seen that the value decreases as $\hat{N}_j$ increases while the variation in the same $\hat{N}_j$ is relatively small. The optimum value of κ is determined so that the sum of mean absolute error (MAE) for all true $N_j = \{1, 2, 3\}$ is minimized.

4.2. Exp. II - hierarchical model

In this subsection, the performance when including the hierarchical model of $\boldsymbol{K}_j$ is evaluated using the simulated data. The block length is reduced to 0.2 s. The number of blocks is $J = 20$. In a single trial (20 blocks), true N_j is invariant while the location of sources is randomly selected (angular distance between the sources is 20°). 30 trails were conducted so that the number of final estimates is the same as that in Exp. I.

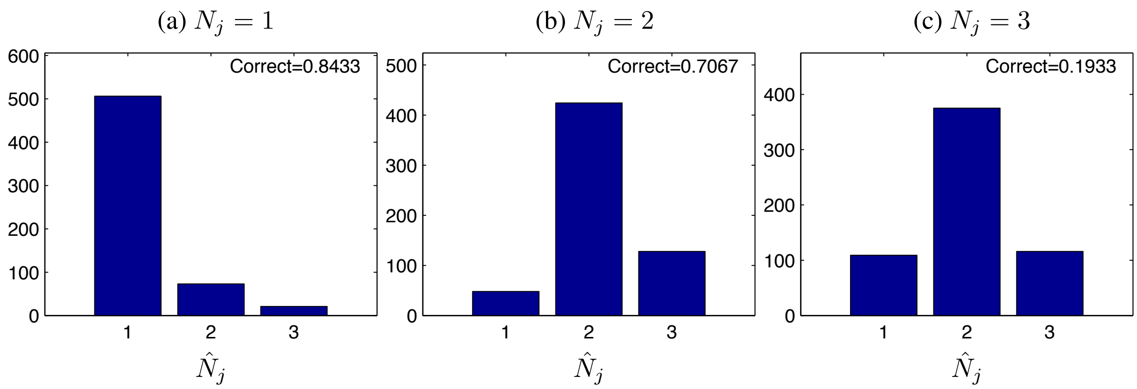
Fig. 4 shows the histogram of $\hat{N}_j$. As compared to Fig. 2, the error for $N_j = 3$ has increased. Table 3 shows the MAE defined as $(1/(N_{trial} \times J)) \sum_t \sum_j |\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j|$. With regard to the estimation of $\boldsymbol{\theta}_j$, the error is comparable to that of Exp. I. Since the amount of data in a single block is reduced to 1/10th of that in Exp. I, the effect of the hierarchical model is confirmed.

Table 3. MAE for Exp. II.

	$N_j = 1$	$N_j = 2$	$N_j = 3$
MAE	1.29	4.44	9.69
C4	0.99	0.62	0.42
C8	0.99	0.92	0.80

4.3. Exp. III - real data

In this experiment, the proposed method is applied to a more realistic dynamic environment. Two human speakers are walking around the robot (HRI-JP, Hearbo) at a distance of 1.5 m. Speech signals from the human speakers are recorded using an 8-element microphone array (approximately circular configuration) mounted on the head of the robot. The reverberation time is approximately 0.3 s. The interval between the human speakers is approximately 30°. The data in the 20 consecutive time blocks in the middle frequency range (1400-2100 Hz, 22 bins) are analyzed. The locations of humans are obtained using ultrasonic transmitters and are used as true $\boldsymbol{\theta}_j$.



(a) $N_j = 1$

(b) $N_j = 2$

(c) $N_j = 3$

Fig. 4. Histogram of $\hat{N}_j$ for different true N_j in Exp. II.

While the number of physical sources is two, the number of active sources in each combination of the time block and the frequency bin is unknown and time-varying among $\{0, 1, 2\}$ due to the sparseness of the speech source signal. Thus, the block-frequency combination with low power is first eliminated from the estimation and $\hat{N}_j$ is selected from $\{1, 2\}$ for the rest of the block-frequency combinations. With regard to the evaluation of $\{\hat{\boldsymbol{\theta}}_j\}$, the estimates are classified according to $\hat{N}_j$ and the values of MAE, C4 and C8 were calculated for each case. For the sake of comparison, the MUSIC and ML estimator with the assumption of $\hat{N}_j = 2$ are also evaluated.

Fig. 5 shows the histogram of $\hat{N}_j$. Table 4 summarizes the MAE, C4 and C8 values. From these, it can be seen that the results of the proposed method for $\hat{N}_j = 2$ are similar to those of MUSIC and ML. On the other hand, the results for the proposed method for $\hat{N}_j = 1$ are improved.

Table 4. MAE for Exp. III.

	Proposed			
	$\hat{N}_j = 1$	$\hat{N}_j = 2$	ML	MUSIC
MAE	**19.36**	47.14	47.50	48.90
C4	**0.35**	0.20	0.24	0.23
C8	**0.52**	0.36	0.37	0.33

4.4. Conclusion

In this paper, a method of joint estimation for the number of sources and their locations was proposed and was examined. From the results of the simulation (Exps. I and II), the potential of the joint estimation was shown. For the real recorded data (Exp. III), the effect of the proposed method was recognized but was limited to the case of $\hat{N}_j = 1$. Since the reverberation was generated by the actual room impulse responses in both the simulated and recorded data, the major difference between them is considered to be the source signal. A speech signal is sparse in the frequency domain and the power balance of the sources changes dynamically even when both sound sources are active. This sometimes results in the signal-to-noise ratio being very low and makes the detection of sources difficult. For the future, an additional framework such as sequential Monte Carlo (SMC) in the time domain and the averaging in the frequency domain can be introduced to obtain the final stable trajectory of sources.

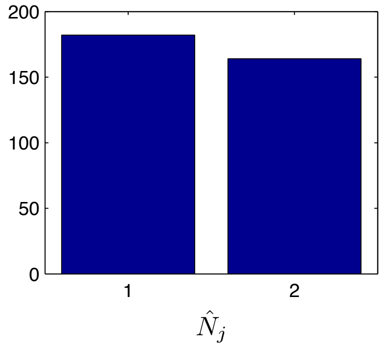


Fig. 5. Histogram of $\hat{N}_j$ in Exp. III.

5. REFERENCES

[1] F. Asano and H.Asoh, “Joint estimation of sound source location and noise covariance in spatially colored noise,” in *Proc. Eusipco 2011*, 2011.

[2] F. Asano, H.Asoh, and K. Nakadai, “Sound source localization in spatially colored noise using a hierarchical bayesian model,” in *Proc. ICASSP 2012*, 2012.

[3] P. J. Green, “Reversible jump MCMC computation and bayesian model determination,” *Biometrika*, vol. 82, pp. 711–732, 1995.

[4] C. Andrieu and A. Doucet, “Joint Bayesian model selection and estimation of noisysinusoids via reversible jump MCMC,” *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2667–2676, 1999.

[5] P. D. Hoff, *A first course in Bayesian statistical methods*, Springer, 2009.