SOUND SOURCE LOCALIZATION IN SPATIALLY COLORED NOISE USING A HIERARCHICAL BAYESIAN MODEL

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

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

ABSTRACT

In this paper, source localization in spatially colored noise is addressed. The covariance of colored noise is estimated using a hierarchical model in the joint Bayesian estimation. The results of the experiment show that the spatial resolution was improved compared with the approach without hierarchical modeling.

Index Terms— spatially colored noise, source localization, Bayesian estimation, hierarchical model

1. INTRODUCTION

For source localization in spatially colored noise, a noise whitening approach using the generalized eigenvalue decomposition (GEVD) in the MUSIC/ESPRIT estimator was proposed [1]. In this approach, the noise covariance must be known in advance. In the frequency domain, room reverberation can approximately be viewed as an additive spatially colored noise[2]. However, the noise covariance is unavailable because it is not possible to observe only reverberation. The authors proposed a method of jointly estimating the noise covariance and the source location using the Bayesian framework [2]. In this paper, a hierarchical model [3] of the noise covariance is introduced into this framework so that it can be estimated efficiently with a relatively small quantity of data. This feature is especially desirable for dynamic environments in which moving sources are present.

In room acoustics, it is known that the resonant frequencies are constant across the room and are independent of the observation position. This fact suggests that the covariance for reverberation can be modeled as a combination of a factor that is common to the multiple observations and a factor that is peculiar to a single observation block, as long as the observations are made in the same environment (room). The common factor is estimated using the hierarchical model while the block variability is estimated by joint estimation.

2. MODEL OF SIGNAL AND NOISE

The observation vector consists of the short-time Fourier transform (STFT) of the sensor inputs as $\boldsymbol{z}_{j,k} = [Z_1(\omega, j, k), \cdots, Z_M(\omega, j, k)]^T$, where $Z_m(\omega, j, k)$ denotes STFT of the mth sensor input at the frequency ω and the time frame index k. The symbol j denotes the index for the time block which consists of K observations (frames) as $\boldsymbol{Z}_j = [\boldsymbol{z}_{j,1}, \cdots, \boldsymbol{z}_{j,K}]$. The source direction $\boldsymbol{\theta}_j = [\theta_{j,1}, \cdots, \theta_{j,N}]^T$ within the block is assumed to be invariant. 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 $\boldsymbol{A}_j(\boldsymbol{\theta}_j)$ denotes the array manifold matrix. The symbols $\boldsymbol{s}_{j,k}$ and $\boldsymbol{v}_{j,k}$ are the source vector and noise vector, respectively. The covariance matrix can 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 denote the number of sensors and sources, respectively.

3. JOINT ESTIMATION OF PARAMETERS WITHIN THE BLOCK

In this section, the joint estimation of the parameters $\{\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j\}$ within the block [2] is briefly reviewed to facilitate the understanding of the hierarchical model described in Section 4. The symbol $\boldsymbol{S}_j$ denotes the block source as $\boldsymbol{S}_j = [\boldsymbol{s}_{j,1}, \cdots, \boldsymbol{s}_{j,K}]$.

3.1. Conditional distribution of parameters

3.1.1. Likelihood

Assuming that $\boldsymbol{v}_{j,k}$ has a complex Gaussian distribution, the likelihood for the block observation $\boldsymbol{Z}_j$ is given by

$$p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) \propto |\boldsymbol{K}_j|^{-K} \times \exp\left(-\sum_{k=1}^{K}[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j\boldsymbol{s}_{j,k}]^H \boldsymbol{K}_j^{-1}[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j\boldsymbol{s}_{j,k}]\right) \tag{3}$$

3.1.2. Conditional distribution of $\boldsymbol{s}_{j,k}$

Assuming that the signal $\boldsymbol{s}_{j,k}$ has the Gaussian prior $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_0)$, its full conditional distribution is the following Gaussian distribution:

$$\begin{aligned} p(\boldsymbol{s}_{j,k}|\boldsymbol{Z}_j, \boldsymbol{\theta}_j, \boldsymbol{K}_j) &\propto p(\boldsymbol{s}_{j,k}) p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) \\ &= \mathcal{N}(\boldsymbol{\mu}_{j,k}, \boldsymbol{\Phi}_j) \end{aligned} \tag{4}$$

where

$$\boldsymbol{\Phi}_j = \left(\boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{A}_j + \boldsymbol{\Phi}_0^{-1} \right)^{-1}, \quad \boldsymbol{\mu}_{j,k} = \boldsymbol{\Phi}_j \boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{z}_{j,k} \tag{5}$$

3.1.3. Conditional distribution of $\boldsymbol{K}_j$

It is assumed that the covariance $\boldsymbol{K}_j$ has the complex inverse-Wishart distribution:

$$\begin{aligned} p(\boldsymbol{K}_j) &= \text{inv-Wishart}\left(\nu_0, (\nu_0 \boldsymbol{K}_0)^{-1}\right) \\ &\propto |\boldsymbol{K}_j|^{-(\nu_0 + M)} \exp\left\{-\text{tr}(\nu_0 \boldsymbol{K}_0 \boldsymbol{K}_j^{-1})\right\} \end{aligned} \tag{6}$$

where ν_0 is the virtual sample size. The conditional distribution of $\boldsymbol{K}_j$ is then the following inverse-Wishart distribution:

$$\begin{aligned} p(\boldsymbol{K}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j, \boldsymbol{\theta}_j) &\propto p(\boldsymbol{K}_j) p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) \\ &= \text{inv-Wishart}(\nu_0 + K, [\nu_0 \boldsymbol{K}_0 + \boldsymbol{C}_j]^{-1}) \end{aligned} \tag{7}$$

where

$$\boldsymbol{C}_j = \sum_{k=1}^{K} (\boldsymbol{z}_{j,k} - \boldsymbol{A}_j \boldsymbol{s}_{j,k}) (\boldsymbol{z}_{j,k} - \boldsymbol{A}_j \boldsymbol{s}_{j,k})^H \tag{8}$$

3.1.4. Conditional distribution of $\boldsymbol{\theta}_j$

Regarding $\boldsymbol{\theta}_j$, the Metropolis algorithm [3] is used since $\boldsymbol{A}_j(\boldsymbol{\theta}_j)$ is a nonlinear function of $\boldsymbol{\theta}_j$ and it is difficult to obtain samples from the conditional distribution [4]. The proposal distribution used in the Metropolis algorithm is the following uniform distribution:

$$J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)}) = \mathcal{U}(\boldsymbol{\theta}_j^{(p)} - \boldsymbol{\delta}, \boldsymbol{\theta}_j^{(p)} + \boldsymbol{\delta}) \tag{9}$$

where p is the index for the iteration and $\boldsymbol{\delta}$ is the appropriate constant vector. The new sample $\boldsymbol{\theta}^*$ is accepted when the following acceptance ratio exceeds a certain threshold r_{thr}:

$$r = \frac{p(\boldsymbol{Z}|\boldsymbol{\theta}_j^*, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{K}_j^{(p+1)})}{p(\boldsymbol{Z}|\boldsymbol{\theta}_j^{(p)}, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{K}_j^{(p+1)})} \frac{p(\boldsymbol{\theta}_j^*)}{p(\boldsymbol{\theta}_j^{(p)})} \tag{10}$$

3.2. Joint parameter estimation using the Gibbs sampler

The iterative algorithm for the joint estimation is as follows:

1. Set $\boldsymbol{K}_j^{(1)}$ and $\boldsymbol{\theta}_j^{(1)}$

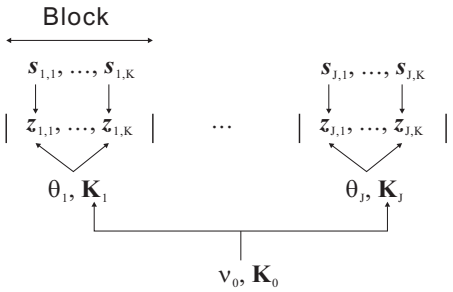

Fig. 1. Hierarchical model of the covariance matrix.

2. Sample $\boldsymbol{s}_{j,k}^{(p+1)} \sim p(\boldsymbol{s}_{j,k}|\boldsymbol{Z}_j, \boldsymbol{\theta}_j^{(p)}, \boldsymbol{K}_j^{(p)}) \quad \forall k$
3. Sample $\boldsymbol{K}_j^{(p+1)} \sim p(\boldsymbol{K}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{\theta}_j^{(p)})$
4. Sample $\boldsymbol{\theta}_j^{(p+1)}$ as: $\boldsymbol{\theta}_j^* \sim J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)})$

$$\boldsymbol{\theta}_j^{(p+1)} = \begin{cases} \boldsymbol{\theta}_j^* & r > r_{thr} \\ \boldsymbol{\theta}^{(p)} & \text{otherwise} \end{cases}$$

5. Go back to Step 2 with $p \leftarrow p + 1$.

4. HIERARCHICAL BAYESIAN MODEL

4.1. Sampling model

Assuming that the noise covariance has a common structure between the block observations $\{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_J\}$, this common structure is estimated using the hierarchical model. We assume the following sampling model:

$$\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J \sim \text{ i.i.d. inv-Wishart}(\nu_0, (\nu_0 \boldsymbol{K}_0)^{-1}) \tag{11}$$

where the parameter set $\{\nu_0, \boldsymbol{K}_0\}$ is common between the J block observations. Fig. 1 depicts this model.

4.2. Conditional distribution of the common paramters

4.2.1. Conditional distribution of $\boldsymbol{K}_0$

According to the sampling model (11), the full conditional distribution of $\boldsymbol{K}_0$ can be decomposed as:

$$p(\boldsymbol{K}_0|\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J, \nu_0) \propto p(\boldsymbol{K}_0) \prod_{j=1}^{J} p(\boldsymbol{K}_j|\boldsymbol{K}_0, \nu_0) \tag{12}$$

Assuming that $\boldsymbol{K}_0$ has the complex Wishart distribution $p(\boldsymbol{K}_0) = \text{Wishart}(\eta, \boldsymbol{\Psi})$ as the prior distribution, the full conditional distribution becomes

$$\begin{aligned} &p(\boldsymbol{K}_0|\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J, \nu_0) \\ &= \text{Wishart}(\boldsymbol{K}_0; \eta, \boldsymbol{\Psi}) \prod_{i=1}^{J} \text{inv-Wishart}(\boldsymbol{K}_j; \nu_0, (\nu_0 \boldsymbol{K}_0)^{-1}) \\ &\propto |\boldsymbol{K}_0|^{\eta + J\nu_0 - M} \exp\left\{-\text{tr}\left(\boldsymbol{K}_0 \boldsymbol{\Lambda}^{-1}\right)\right\} \\ &\propto \text{Wishart}\left(\boldsymbol{K}_0; \eta + J\nu_0, \boldsymbol{\Lambda}\right) \end{aligned} \tag{13}$$

where

$$\boldsymbol{\Lambda} := \left(\boldsymbol{\Psi}^{-1} + \nu_0 \sum_{j=1}^{J} \boldsymbol{K}_j^{-1} \right)^{-1} \tag{14}$$

(14) is analogous to the harmonic mean [3].

4.2.2. *Conditional distribution of* ν_0

Assuming that ν_0 has the prior distribution $p(\nu_0) \propto \exp(-\alpha\nu_0)$ [3], its full conditional distribution becomes

$$\begin{aligned} p(\nu_0|\boldsymbol{K}_0, \boldsymbol{K}_1, \cdots, \boldsymbol{K}_J) &\propto p(\nu_0) \prod_{j=1}^{J} p(\boldsymbol{K}_j|\nu_0, \boldsymbol{K}_0) \\ &\propto \exp(-\alpha\nu_0) \prod_{j=1}^{J} \frac{|\nu_0 \boldsymbol{K}_0|^{\nu_0}}{\Gamma_M(\nu_0)} |\boldsymbol{K}_j|^{-(\nu_0+M)} \times \\ &\quad \exp\left\{-\mathrm{tr}(\nu_0 \boldsymbol{K}_0 \boldsymbol{K}_j^{-1})\right\} \end{aligned} \tag{15}$$

where $\Gamma_M(\nu_0) = \pi^{M(M-1)/2} \prod_{m=1}^{M} \Gamma(\nu_0 - m + 1)$. The symbol $\Gamma(\cdot)$ denotes the Gamma function.

4.3. Iterative algorithm

The procedure for obtaining samples of $\boldsymbol{K}_0$ and ν_0 is as follows:

1. Set $\boldsymbol{K}_0^{(1)}$ and $\nu_0^{(1)}$.
2. Sample $\{\boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)}\}$ using the procedure described in Section 3.2.
3. Sample $\boldsymbol{K}_0$ as:
$$\boldsymbol{K}_0^{(p+1)} \sim p(\boldsymbol{K}_0|\boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)}, \nu_0^{(p)})$$
4. Sample ν_0 as:
$$\nu_0^{(p+1)} \sim p(\nu_0|\boldsymbol{K}_0^{(p+1)}, \boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)})$$
5. Go back to Step 2 with $p \leftarrow p + 1$.

5. EXPERIMENT

5.1. Experiment

5.1.1. *Condition*

Observations were generated by convolving the room impulse responses with the source signal (Gaussian noise). The room used for measuring the impulse response was a middle sized meeting room (8 m × 9 m × 2.5 m) with a reverberation time of 0.5 s. Two sound sources were located on a circle with a radius of 1.5 m. The angular distance between the sources was 20° and the location of the source-pair was randomly selected. Twenty observation blocks with different source-pair locations were used for the estimation ($J = 20$). A microphone array with 8 elements mounted on the head of a robot was placed at the center of the circle. The parameters used for the signal analysis are summarized in Table 1. As the initial value $\boldsymbol{\theta}^{(1)}$, the ML estimate fluctuated by adding a uniform noise was employed.

Table 1. Parameters for analysis.

Parameter	Value
Sampling frequency	16 kHz
Frame length (STFT length)	512 points
Frame shift	128 points
Block length (observation time)	3200 points (0.2 s)
Frequency	1500 Hz
Number of iteration	1000

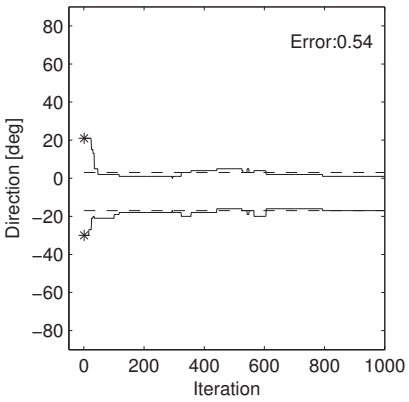

Fig. 2. Variation of sample $\boldsymbol{\theta}^{(p)}$ during the iteration.

5.1.2. *Results*

Fig. 2 shows the variation in $\boldsymbol{\theta}^{(p)}$ during the iteration. It is observed that the samples converged on the true value (the dotted line) with a small number of iterations. The final estimate $\hat{\boldsymbol{\theta}}$ was obtained as the mean of the samples.

Fig. 3 compares the mean absolute error (MAE) of the proposed method with those of the maximum likelihood (ML) and the MUSIC estimators. MAE was calculated as $\mathrm{MAE} = 1/(J \times N_{trial}) \sum |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|$, where the number of trial is $N_{trial} = 30$. The proposed method has a smaller MAE compared with the other two methods.

Fig. 4 indicates MAE for different methods used to obtaining $\boldsymbol{K}_0$. "Wishart" corresponds to the proposed method described in Section 4.2.1. For "Harm", $\boldsymbol{\Lambda}$ in (14), which is the conditional mean of the Wishart distribution, was employed as $\boldsymbol{K}_0$. For "Arith", the arithmetic mean of $\{\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J\}$ was employed. It can be noted that MAE was small for "Wishart" and "Harm". From these re-

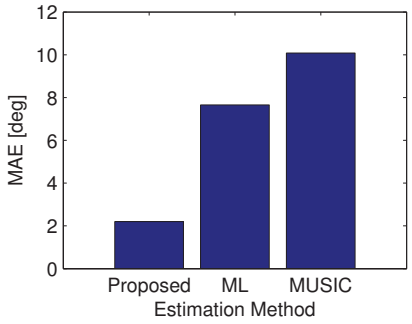

Fig. 3. MAE for different parameter estimation methods.

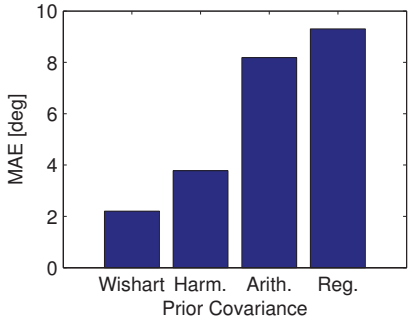

Fig. 4. MAE for different methods used to obtain $\boldsymbol{K}_0$.

sults, it is deduced that the harmonic-mean-like operation in (14) was essential for the hierarchical modeling of $\boldsymbol{K}_0$. For " Reg ", $\sigma\boldsymbol{I}$ was employed as $\boldsymbol{K}_0$. In this case, only the joint estimation without hierarchical modeling was conducted. The role of $\boldsymbol{K}_0 = \sigma\boldsymbol{I}$ is the regularization of $\boldsymbol{C}_j$. The difference between " Wishart " and " Reg " corresponds to the effect of the hierarchical modeling when the quantity of data in a block is small.

Fig. 5 shows MAE when the value of ν_0 is fexed during the iteration. It is observed that MAE is a minimum at around $\nu_0 = 10^2$.

To evaluate the value of $\boldsymbol{K}_0$ that is estimated by the hierarchical model, the obtained sample of $\boldsymbol{K}_0$ was fed to the

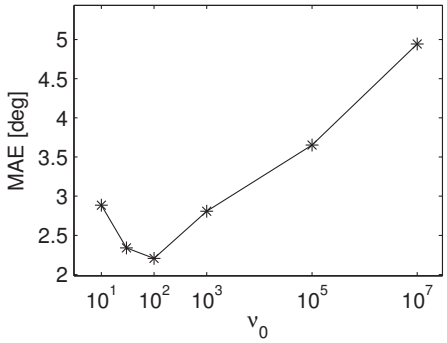

Fig. 5. MAE for different ν_0 values.

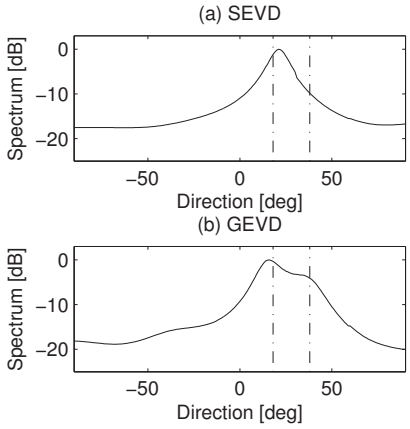

Fig. 6. Evaluation of the $\boldsymbol{K}_0$ value estimated using GEVD-MUSIC spatial spectral estimator.

GEVD-MUSIC. Fig. 6(b) shows the spatial spectrum. Compared with the MUSIC spectrum obtained by using the standard eigenvalue decomposition (SEVD, spatially white assumption) shown in (a), the spatial resolution was found to have improved.

6. CONCLUSION

In this study, we investigated a method of estimating the noise covariance using a hierarchical model. From the results of the experiment, it was shown that the spatial resolution was improved by the hierarchical modeling when the quantity of data in a single observation block is small. In the proposed method, the number of sources are assumed to be known in advance, and was fixed at two in the experiment. In future, we intend to address the estimation of the number of active sources in this joint estimation framework [4].

7. REFERENCES

[1] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

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

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

[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.