# Estimating the number of sources using independent component analysis

Hiroshi Sawada*, Ryo Mukai, Shoko Araki and Shoji Makino

*NTT Communication Science Laboratories, NTT Corporation*

## 1.  Introduction

Estimating the number of sources provides useful information for signal processing applications, such as blind source separation (BSS) in the frequency domain [1]. It is well known that the number of dominant eigenvalues of the spatial correlation matrix corresponds to the number of sources [2,3]. However, it is difficult to distinguish dominant eigenvalues from the other eigenvalues in a reverberant case as shown in Sect. 3. This difficulty has already been pointed out in [3], where they propose the use of support vector machines (SVM) to classify eigenvalue distributions and determine the number of sources. However, the SVM needs to be trained beforehand and experimental results were provided only for 1- or 2-source cases.

This letter first discusses the problem of the conventional eigenvalue-based method when applied to a reverberant condition. Then we propose a new approach for estimating the number of sources that employs independent component analysis (ICA). Experimental results show the characteristics of the proposed approach compared with the conventional eigenvalue-based method.

## 2.  Convolutive mixtures and frequency-domain approximation

Suppose that $N$ source signals $s_1(t), \ldots, s_N(t)$ are convolutively mixed and observed at $M$ sensors

$$x_j(t) = \sum_{k=1}^{N} \sum_{l=0}^{P-1} h_{jk}(l) s_k(t-l) + n_j(t), \qquad (1)$$

where $h_{jk}(l)$ represents the impulse response from source $k$ to sensor $j$, $P$ represents the duration of the impulse responses, and $n_j(t)$ is an additive Gaussian noise for each sensor. The goal is to estimate the number of sources $N$ only from the observations $x_1(t), \ldots, x_M(t)$. We assume that the number of sensors $M$ is larger than or equal to the number of sources $N$, i.e. $N \leq M$.

It is effective to convert the time-domain convolutive mixtures $x_j(t)$ into frequency-domain time-series signals $x_j(f, \tau)$ with a short-time Fourier transform (STFT). This is because convolutive mixtures (1) can be approximated as multiple instantaneous mixtures in the frequency domain:

$$x_j(f, \tau) = \sum_{k=1}^{N} h_{jk}(f) s_k(f, \tau) + n_j(f, \tau), \qquad (2)$$

where $h_{jk}(f)$ is the frequency response from source $k$ to sensor $j$, and $s_k(f, \tau)$ and $n_j(f, \tau)$ are frequency-domain time-series signals of $s_k(t)$ and $n_j(t)$ obtained similarly with STFT, respectively. The vector notation of Eq. (2) is

$$\boldsymbol{x}(f, \tau) = \sum_{k=1}^{N} \boldsymbol{h}_k(f) s_k(f, \tau) + \boldsymbol{n}(f, \tau), \qquad (3)$$

where $\boldsymbol{x} = [x_1, \ldots, x_M]^T$, $\boldsymbol{h}_k = [h_{1k}, \ldots, h_{Mk}]^T$ and $\boldsymbol{n} = [n_1, \ldots, n_M]^T$.

## 3.  Conventional eigenvalue-based method

The conventional eigenvalue-based method estimates the number of sources in the following way [2,3]. It performs eigenvalue decomposition for the spatial correlation matrix $\boldsymbol{R}(f) = \langle \boldsymbol{x}(f, \tau) \boldsymbol{x}(f, \tau)^H \rangle_\tau$, where $\langle \cdot \rangle_\tau$ denotes the averaging operator and $\cdot^H$ denotes the conjugate transpose. Let $\lambda_1 \geq \cdots \geq \lambda_N \geq \cdots \geq \lambda_M$ be the sorted eigenvalues of $\boldsymbol{R}$. If the approximation by Eq. (2) holds sufficiently well, the number of dominant eigenvalues is equal to the number of sources $N$, and the remaining $M - N$ smallest eigenvalues are the same as the noise power: $\lambda_{N+1} = \cdots = \lambda_M = \sigma_n^2$. However, there are two problems in a real reverberant condition.

**Reverberation** If the reverberation of a mixing system is long and strong, the approximation provided by Eq. (2) does not hold sufficiently well and the number of dominant eigenvalues might be more than the number of sources.

**Unrecovered power** If some of the mixing vectors $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N$ are similar, the number of dominant eigenvalues might be less than the number of sources. In this case, the first few eigenvalues represent almost all powers. A typical situation can be seen in low frequencies, where the phase differences among sensors are very small.

Because of these two problems, the eigenvalue-based method does not work well in a real reverberant condition. Figure 1 shows normalized component powers estimated by the eigenvalue-based method in an environment whose conditions are summarized in Fig. 2. The left hand plot shows a one-source case. Because of reverberations, the normalized powers of the second principal components were around $-20\,\text{dB}$. To distinguish between the source and the noises (including reverberations), a threshold of around $-15\,\text{dB}$ is suitable for the one-source case. However, if such a threshold is used for the three-source case shown in the right hand plot,
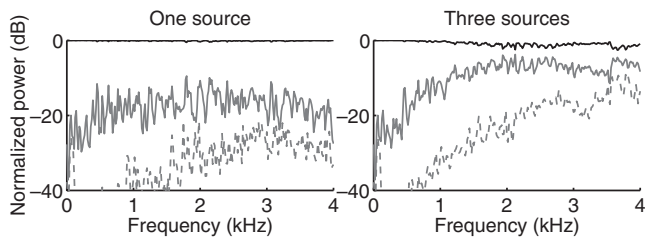
*e-mail: sawada@cslab.kecl.ntt.co.jp

**Fig. 1** Normalized component powers estimated by the eigenvalue-based method (The powers of the first, second and third principal components are represented by solid, gray, and dashed-gray lines, respectively).
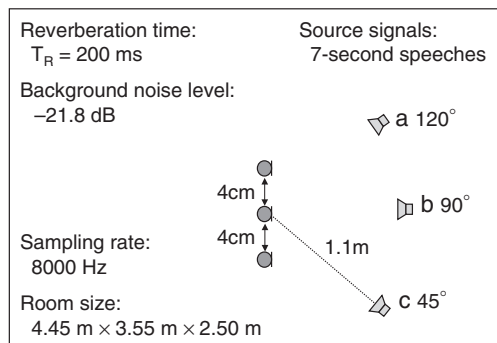


**Fig. 2** Experimental conditions.

the number of sources is estimated at two in most frequency bins. Therefore, it is hard to find a threshold that works well for both cases.

## 4. Proposed ICA-based method

In this section, we propose a new method for estimating the number of sources that solves the two problems mentioned above.

To solve the problem of **unrecovered power**, the proposed method recovers the power of each signal measured at the sensors by using ICA [4] and a scaling technique. It first applies ICA for $x(f, t)$ to obtain $M$ separated signals:

$$y(f, \tau) = W(f)x(f, \tau), \tag{4}$$

where $W$ is an $M \times M$ separation matrix and $y = [y_1, \ldots, y_M]^{\mathrm{T}}$ is the vector of the separated signals. $N$ elements of $y$ correspond to sources and the remaining $M - N$ elements correspond to noise components (including reverberations). The scaling ambiguity of ICA means that the power of each element of $y$ might be different from the power of each source or noise, and the $M - N$ noise components are generally enhanced. To recover the power of each source or noise measured at the sensors, we determine the scaling ambiguity by

$$W \leftarrow \Lambda W, \quad \Lambda = \mathrm{sqrt}(\mathrm{diag}[(WW^{\mathrm{H}})^{-1}]), \tag{5}$$

where $\mathrm{diag}(\cdot)$ retains only the diagonal elements and makes the non-diagonal elements zero, and $\mathrm{sqrt}(\cdot)$ calculates the square root of each element. We call Eq. (5) *power-recovery scaling* since the power of each source measured at the

sensors is recovered at each output:

$$|y_i(f, \tau)|^2 = \sum_{j=1}^{M} |h_{jk}(f)s_k(f, \tau)|^2 \tag{6}$$

if ICA is properly solved and $y_i$'s are made mutually independent. This solves the problem of unrecovered power.

Then we consider the problem of **reverberation**. Assume that a separated signal $y_k(f, \tau)$ is a source component and another separated signal $y_i(f, \tau)$ is not a source component but includes the reverberation of the source corresponding to $y_k(f, \tau)$. Then, the reverberant component $y_i(f, \tau)$ should have a similar time structure to $y_k(f, \tau)$. The similarity can be calculated by the correlation

$$\frac{\langle v_i(\tau) \cdot v_k(\tau - \Delta\tau) \rangle_\tau}{\sqrt{\langle v_i^2(\tau) \rangle_\tau} \cdot \sqrt{\langle v_k^2(\tau - \Delta\tau) \rangle_\tau}} \tag{7}$$

of envelopes $v_i(\tau) = |y_i(f, \tau)| - \langle |y_i(f, \tau)| \rangle_\tau$ with an appropriate time delay $-\Delta\tau$. If the correlation is large enough, we consider $y_i(f, \tau)$ to be a reverberant component.

The overall procedure of the method is as follows.

1. Obtain independent components $y_i(f, \tau)$ by using ICA (4) and the scaling formula (5), and then calculate the normalized power $NP_i = \langle |y_i(\tau)|^2 \rangle_\tau / \sum_{k=1}^{M} \langle |y_k(\tau)|^2 \rangle_\tau$ for $i = 1, \ldots, M$.
2. If the normalized power $NP_i$ is smaller than a certain threshold, e.g. 0.01 ($-20\,\mathrm{dB}$), we consider $y_i$ to be a noise component.
3. If the normalized power $NP_i$ is smaller than a certain threshold, e.g. 0.2, and one of the correlations (7) among other components is larger than a certain threshold, e.g. 0.5, we consider $y_i$ to be a reverberant component.
4. Otherwise, we consider the $i$-th component $y_i$ to be a signal.

These thresholds can be determined beforehand by the power levels of the background noise and reverberations.

## 5. Experimental results

We performed experiments to estimate the number of sources. Sensor measurements were made under the conditions summarized in Fig. 2. We tested cases of one, two and three sources, while the number of sensors was three for all cases. Figure 3 shows the numbers of sources estimated by using the conventional eigenvalue-based method and the
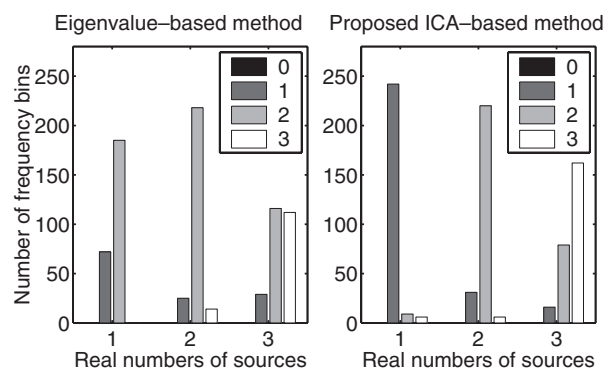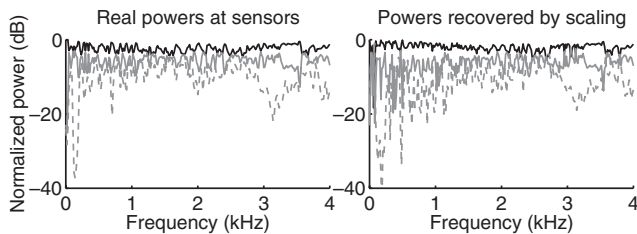


**Fig. 3** Estimated numbers of sources.

**Fig. 4** Power recovery by scaling formula (5) when there were three sources.



**Fig. 5** Identifying reverberant components when there was one source.

proposed ICA-based method. The vertical axis shows the number of frequency bins for each estimated number of sources. The STFT frame size was 512 for a sampling rate of 8,000 Hz, and thus the total number of frequency bins to cover 0–4,000 Hz was 257. By taking the maximum vote, the ICA-based method successfully estimated the number of sources in all cases, whereas the eigenvalue-based method estimated the number of sources to be two in all cases.

Figure 4 shows how well the powers of the sources were recovered by using ICA and the proposed scaling technique. The left hand plot shows the normalized powers of the three sources measured at the sensors, and the right hand plot shows those estimated with ICA and the scaling formula (5). The powers were sufficiently well recovered to estimate the number of sources. Compared with the result obtained with the eigenvalue-based method (the right hand plot in Fig. 1), the advantage of the proposed method becomes clear.

Figure 5 shows how the proposed method copes with the reverberation problem. The left hand plot shows the normalized power of the first and second largest components of the scaled ICA outputs when there was one source. It was hard to decide solely from these normalized powers whether the second component was a signal or a noise because the powers of the second components were not sufficiently small in many frequency bins. However, by calculating the correlation (7) of the envelopes between the first and second components, it
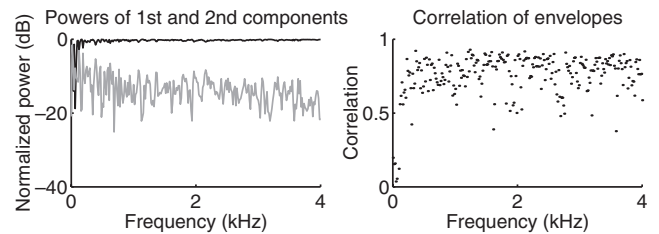
became clear that the second component was a noise component including reverberations. The right hand plot shows the correlations, which were sufficiently large (around 0.8) in many frequency bins.

## 6. Conclusion

We have proposed a new method for estimating the number of sources. This method recovers the power of each source measured at the sensors by using ICA and power-recovery scaling. The method can also identify a noise component that includes reverberations by calculating the correlation of the envelopes.

## References

[1] H. Sawada, R. Mukai, S. Araki and S. Makino, "Frequency-domain blind source separation," in *Speech Enhancement*, J. Benesty, S. Makino and J. Chen, Eds. (Springer, Berlin, 2005), pp. 299–327.

[2] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech Signal Process.*, **33**, 387–392 (1985).

[3] K. Yamamoto, F. Asano, W. van Rooijen, E. Ling, T. Yamada and N. Kitawaki, "Estimation of the number of sound sources using support vector machines and its application to sound source separation," *Proc. ICASSP 2003*, pp. 485–488 (2003).

[4] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis* (John Wiley & Sons, New York, 2001).