International Workshop on Acoustic Echo and Noise Control (IWAENC2003), Sept. 2003, Kyoto, Japan

GEOMETRICAL UNDERSTANDING OF THE PCA SUBSPACE METHOD FOR OVERDETERMINED BLIND SOURCE SEPARATION

*Stefan Winter** *Hiroshi Sawada* *Shoji Makino*

NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{wifan,sawada,maki}@cslab.kecl.ntt.co.jp

ABSTRACT

In this paper, we discuss approaches for blind source separation where we can use more sensors than the number of sources for a better performance. The discussion focuses mainly on reducing the dimension of mixed signals before applying independent component analysis. We compare two previously proposed methods. The first is based on principal component analysis, where noise reduction is achieved. The second involves selecting a subset of sensors based on the fact that a low frequency prefers a wide spacing and a high frequency prefers a narrow spacing. We found that the PCA-based method behaves similarly to the geometry-based method for low frequencies in the way that it emphasizes the outer sensors and yields superior results for high frequencies, which provides a better understanding of the former method.

1. INTRODUCTION

Blind source separation (BSS) is a technique for estimating original source signals using only sensor observations that are mixtures of the original signals. If source signals are mutually independent and non-Gaussian, we can employ independent component analysis (ICA) to solve a BSS problem. Although in many cases equal numbers of source signals and sensors are assumed [1], using more sensors than source signals (overdetermined systems) often yields better results [2–4]. Different techniques are employed to map the mixture signal space to the reduced dimensional output signal space.

In this paper we present the results of overdetermined BSS based on two different methods of subspace selection. Each provides better separation results than when the number of sensors and sources is the same. The first method utilizes the principal components obtained by principal component analysis (PCA) as described in [5]. The second method is based on geometrical selection that depends on the frequency and sensor spacing as described in [6].

We compared the two methods by undertaking experiments using real world data in a reverberant environment. We found that for low frequencies the PCA-based method behaves similarly to the geometry-based method, while for high frequencies the former yields better results, since it normally removes the noise subspace more efficiently than the geometry-based method. This provides a better understanding of the PCA-based approach.

*The author is on leave from the Chair of Multimedia Communication and Signal Processing, University Erlangen-Nuremberg.

2. BSS USING MORE SENSORS THAN THE NUMBER OF SOURCES

The general framework of overdetermined BSS is shown in Fig. 1.

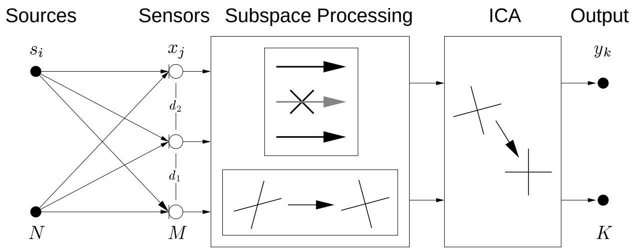

Fig. 1. General framework of overdetermined BSS

After the mixing process there is a subspace processing stage followed by the actual ICA stage. The subspace processing stage can be subdivided into a sphering stage and a dimension reduction stage. Their order is different in the two methods described here.

We consider a convolutive BSS model with N sources $s_i(t)$ $(i = 1, \ldots, N)$ and $N < M$ sensors that give mixed signals $x_j(t)$ $(j = 1, \ldots, M)$ with added noise $n_j(t)$. The mixing process can be described by

$$x_j(t) = \sum_{i=1}^{N} \sum_{l=0}^{\infty} h_{ji}(l) s_i(t-l) + n_j(t) \qquad (1)$$

where $h_{ji}(t)$ stands for the impulse response from source i to sensor j.

The use of more sensors than the number of sources usually improves the separation result. We can exploit the performance improvement technique known from beamforming theory. When achieving the separation, we have to apply some dimension reduction in order to map the number of mixed signals to the number of output signals. It appears to be more advantageous to reduce the dimensions before rather than after ICA.

We employed a frequency-domain approach to solve the convolutive BSS problem including the subspace processing. First, we calculate the frequency responses of the separating system, then we obtain time-domain representations by applying an inverse discrete Fourier transform (DFT). This has an advantage in that subspace processing and ICA is used for instantaneous mixtures, which are easier to solve than convolutive ones. Time-domain signals $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]$ are converted into frequency-domain time-series signals $\mathbf{X}(f, m) = [X_1(f, m), \ldots X_M(f, m)]$ by an L-point short time DFT, where $f = 0,\ f_s/L,\ \ldots,\ f_s(L-1)/L$ (f_s: sampling frequency; m: time dependence). After subspace processing, we have uncorrelated signals $\mathbf{Z}(f, m) = [Z_1(f, m),\ \ldots,\ Z_N(f, m)]^T$ reduced to the dimension N. To obtain the frequency responses $W_{kj}(f)$ of the separating system, we solve an ICA problem $\mathbf{Y}(f, m) = \mathbf{W}(f)\mathbf{Z}(f, m)$, where $\mathbf{Y}(f, m) = [Y_1(f, m), \ldots, Y_K(f, m)]^T$ and $\mathbf{W}(f)$ is a $N \times N$ matrix whose elements are $\mathbf{w}_k^H(f)$. $Y_k(f, m)$ is a frequency-domain representation of $y_k(t)$ and they are made so as to be mutually independent.

The application of ICA to convolutive mixtures in the frequency-domain implies that we have to solve the permutation problem afterwards [7, 8]. The more signals we have to permute, the more difficult it is to characterize the components of each frequency bin uniquely and relate them to the components of adjacent frequency bins. Thus we want to have as few components as possible, i.e., the least possible number of dimensions.

If we reduce the dimensions after ICA, we face a similar problem to the one that arises when solving the permutation problem, i.e., the question of how to discriminate between the components. We usually have more information before using ICA to select an appropriate subspace (such as eigenvalues = covariance and sensor spacing) than after using ICA (eigenvalues are distorted due to scaling ambiguity). Therefore it is better to reduce the dimensions before employing ICA.

We applied the complex version of FastICA proposed in [9] to $\mathbf{Z}(f)$ to obtain separated signals. By using negentropy maximization as a basis, the unmixing vector for each signal is gradually improved by

$$\begin{aligned} \mathbf{w} \leftarrow\ & E\left\{\mathbf{Z}(\mathbf{w}^H\mathbf{Z})^* g(|\mathbf{w}^H\mathbf{Z}|^2)\right\} - \\ & E\left\{g(|\mathbf{w}^H\mathbf{Z}|^2) + |\mathbf{w}^H\mathbf{Z}|^2 g'(|\mathbf{w}^H\mathbf{Z}|^2)\right\}\mathbf{w} \end{aligned} \tag{2}$$

until the difference between consecutive unmixing vectors falls below a certain threshold. $E\{\cdot\}$ denotes the expectation value, $\cdot^H$ the hermitian operator and $\cdot^*$ the complex conjugation. $\mathbf{Z}$ is assumed to have a zero mean and unit variance. $g(\cdot)$ denotes the derivative of a nonlinear function $G(\cdot)$, which was here chosen as $G(x) = \log(a + x)$ with $a = 0.1$. $\mathbf{w}$ is orthonormalized with respect to already existing unmixing vectors after each step.

3. SUBSPACE SELECTION

3.1. Subspace selection based on statistical properties

Asano et al. proposed a BSS system that utilizes PCA for selecting a subspace [5]. PCA in general gives principal components that are by definition uncorrelated and is suited to dimension reduction [1, 2]. Here PCA is based on the spatial correlation matrix $\mathbf{R}_{\mathbf{xx}} = E\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\}$.

They consider room reflections to be uncorrelated noise from the direct source signals $s_i(t)$ in a practical sense on condition that the time shift between direct and reflected signal is larger than the window length used for the short time DFT. By assuming uncorrelatedness, it follows that the first N principal components with the largest eigenvalues contain a mixture of direct source signals and noise. N denotes the number of sources. By contrast the remaining principal components consist solely of noise.

Thus by selecting the subspace that is spanned by the first N principal components, dimensions are effectively reduced by removing noise while keeping the signal of interest [10].

Here it is important to note that sphering takes places before dimension reduction, which is based on the principal components found by sphering and is applied in the sphered signal space.

3.2. Subspace selection based on geometrical knowledge

A method for blind source separation has been proposed using several separating subsystems whose sensor spacing could be configured individually [6]. The idea is based on the fact that low frequencies prefer a wide sensor spacing whereas high frequencies prefer a narrow sensor spacing. This is due to the resulting phase difference, which plays a key role in separating signals. Therefore three sensors were arranged in a way that gave two different sensor spacings using one sensor as a common sensor as shown in Fig. 2.

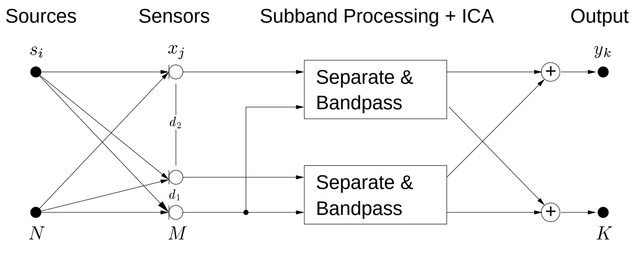

Fig. 2. Geometry-based subspace selection

The frequency range of the mixed signals was divided into lower and higher frequency ranges. According to [6] for a frequency to be adequate for a given spacing d the condition in (3) should be fulfilled. Here α is a parameter that governs the degree to which the phase difference exceeds π, c denotes the sound velocity and θ_i stands for the i-th source's direction.

Table 1. Summarized comparison

PCA-based selection	**Geometry-based selection**
Statistical consideration	Geometrical considerations
Different subspace for each frequency range	Two different subspaces
First sphering, then dimension reduction	First dimension reduction, then sphering

Table 2. Experimental conditions

Direction of sources	50° and 150°
Distance of sensors	$d_1 = d_2 = 28.3$ mm
Length of source signals	7.4 seconds
Sampling rate	8 kHz
Window type	Hanning
Filter length	2048 points
Shifting interval	512 points
Frequency range parameter	$\alpha = 1.2$
Threshold for FastICA	10^{-3}

$$f \leq \frac{\alpha c}{2d\left(\cos(\theta_1) - \cos(\theta_2)\right)} \tag{3}$$

The appropriate sensor pairs were chosen for each frequency range and separately used for separation in each frequency range. Before ICA was applied to each chosen pair, the mixed signals were sphered. It is important to note that sphering takes places after dimension reduction, which is based on geometrical considerations and is applied in the mixed signal space.

The similarities and differences between the two subspace selection methods are summarized in Table 1.

4. EXPERIMENTAL RESULTS

4.1. Sensor selection of the PCA-based approach

We examined the behavior of the PCA-based subspace selection with regard to the resulting sensor selection. Speech signals do not always comply with the assumptions of uncorrelatedness and independence, which are made when applying PCA and ICA to them. Therefore, to assess the ideal behavior, we used artificial signals produced by a random generator in the frequency-domain with the desired properties instead of real speech signals. The normalized sensor gain depending on the frequency bin and sensor position is shown in Fig. 3. We used the experimental conditions given in the first two lines of Table 2.

We can see that the PCA-based method also emphasizes the outer sensors with a wide spacing for low frequencies as the geometrical considerations in [6] suggest. However, the remaining sensor is not excluded but contributes more the higher the frequency becomes. Different settings used in additional experiments revealed similar behavior, particularly for low frequencies.

4.2. Comparison of the PCA- and geometry based approaches

To compare the PCA- and geometry-based methods, we separated mixtures that we obtained by convolving impulse responses $h_{ji}(t)$ and pairs of speech signals $s_i(t)$, and optionally adding artificial noise $n_j(t)$. We used speech signals from the Acoustical Society of Japan (ASJ) continuous speech corpus and impulse responses in the Real World Computing Partnership (RWCP) sound scene database from real acoustic environments [11]. The source directions θ_i were estimated by the MUSIC algorithm (the direction orthogonal to the linearly mounted sensor array is 90°) [12]. The frequency ranges were calculated based on the criteria discussed in Sec 3.2.

We calculated the SNR at output k as $10 \log \left(\sum_t y_k^s(t)^2 / \sum_t y_k^c(t)^2\right)$, where $y_k^s(t)$ is a portion of $y_k(t)$ that comes from a source signal $s_k(t)$ and $y_k^c(t) = y_k(t) - y_k^s(t)$.

To avoid any influence of the permutation problem on the result we selected the best permutation by calculating the SNR in each frequency bin in a similar way to that described above. The solution is ideal under the condition that the permutation problem is perfectly solved.

The experimental conditions are given in Table 2.

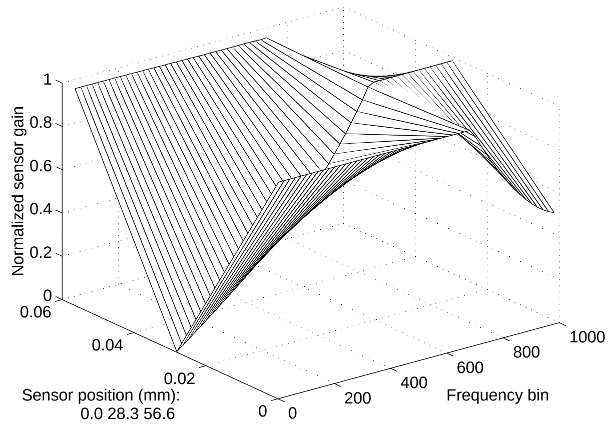

Fig. 3. Normalized sensor gain with PCA-based subspace selection

Figures 4 and 5 show the results for both methods for 12 pairs of speech signals. Figure 4 reveals that both subspace methods show a similar behavior for low frequencies independent of added noise. This confirms that the PCA-based approach also emphasizes the wider sensor spacing in the same way as the geometry-based method.

However, for high frequencies, while both approaches still perform similarly if we only account for reverberation, the PCA-based approach works better than the geometry-based approach if noise is added (Fig. 5). We confirmed the superior performance with additional experiments using different sensor spacings.

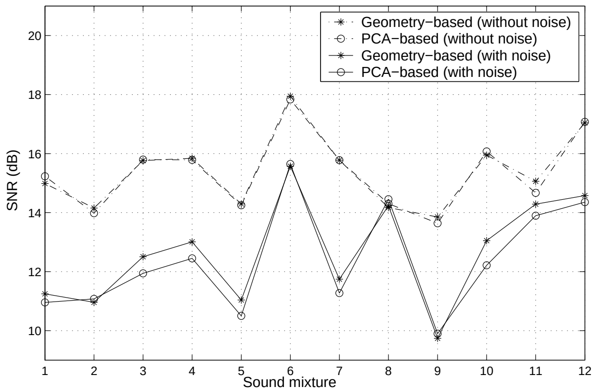

Fig. 4. Comparison of PCA- and geometry-based subspace selection for low frequency range

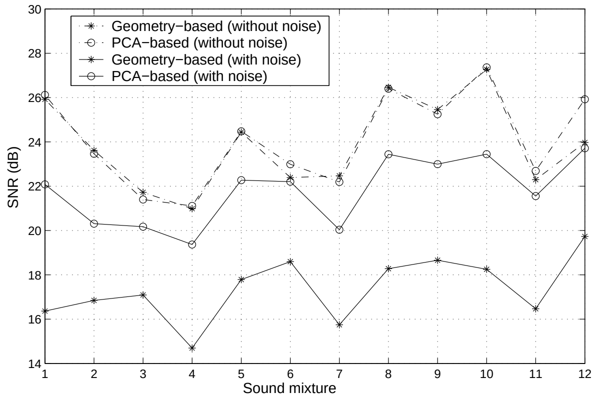

Fig. 5. Comparison of PCA- and geometry-based subspace selection for high frequency range

4.3. Interpreting experimental results

We can explain the similar performance of the PCA- and geometry-based methods for low frequencies by the fact that the PCA-based method also emphasizes the outer sensors as shown in Sec. 4.1. This normally provides the highest possible phase difference for low frequencies, which is important for correctly separating the mixed signals by the subsequent ICA stage as mentioned in Sec. 3.2.

For high frequencies the PCA-based method yields better results when noise is added because it can utilize all the sensors (Fig. 3) whereas the geometry-based method still utilizes only two sensors. By using all the sensors the PCA-based approach can effectively suppress the uncorrelated noise.

5. CONCLUSION

We have compared two subspace methods for use as preprocessing steps in overdetermined BSS. We found that for low frequencies the PCA-based method exhibits a similar performance to the geometry-based method because it also emphasizes the outer sensors. For high frequencies the PCA-based approach performs better when exposed to noisy speech mixtures because due to appropriate phase difference it can utilize all pairs of sensors to suppress the noise. This deepens the geometrical understanding of the PCA-based method.

6. REFERENCES

[1] A. Hyvaerinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2000.

[2] M. Joho, H. Mathis, and R.H. Lambert, “Overdetermined blind source separation: using more sensors than source signals in a noisy mixture,” in *Proc. ICA2000*, June 2000, pp. 81–86.

[3] A. Westner and J. Bove, “Blind separation of real world audio signals using overdetermined mixtures,” in *Proc. ICA99*, Jan. 1999.

[4] A. Koutras, E. Dermatas, and G. Kokkinakis, “Improving simultaneous speech recognition in real room environments using overdetermined blind source separation,” in *Proc. Eurospeech2001*, Sept. 2001, pp. 1009–1012.

[5] F. Asano, Y. Motomura, H. Asoh, and T. Matsui, “Effect of PCA filter in blind source separation,” in *Proc. ICA2000*, June 2000, pp. 57–62.

[6] H. Sawada, S. Araki, R. Mukai, and S. Makino, “Blind source separation with different sensor spacing and filter length for each frequency range,” in *IEEE International Workshop on Neural Networks for Signal Processing (NNSP2002)*, Sept. 2002, pp. 465–474.

[7] M. Z. Ikram and D. R. Morgan, “Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment,” in *Proc. ICASSP 2000*, June 2000, pp. 1041–1044.

[8] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust approach to the permutation problem of frequency-domain blind source separation,” in *Proc. ICASSP2003*, 2003, submitted.

[9] E. Bingham and A. Hyvaerinen, “A fast fixed-point algorithm for independent component analysis of complex valued signals,” *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, Feb. 2000.

[10] J.-P. Nadal, E. Korutcheva, and F. Aires, “Blind source separation in the presence of weak sources,” *Neural Networks*, vol. 13, no. 6, pp. 589–596, 2000.

[11] Real World Computing Partnership, “RWCP sound scene database in real acoustic environments,” `http://tosa.mri.co.jp/sounddb/indexe.htm`.

[12] H.L. Van Trees, *Optimum array processing*, Wiley, New York, 2002.