GEOMETRICAL INTERPRETATION 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 is based on geometric considerations and selects a subset of sensors according to 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. These results provide 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].

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

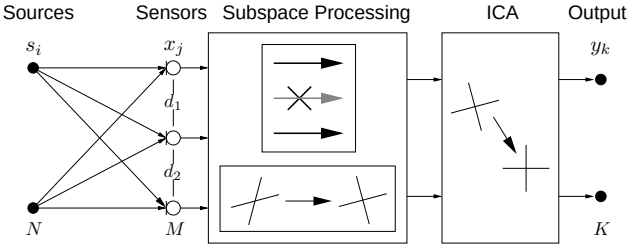

Fig. 1. General framework of overdetermined BSS

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. These results provide a better understanding of the PCA-based approach.

2. BSS USING MORE SENSORS THAN THE NUMBER OF SOURCES

The general framework of overdetermined BSS is shown in Fig. 1. After the mixing process there is a subspace processing stage followed by the actual ICA stage. The reasons for this position of the subspace processing stage will be explained in Sec. 3.1. The subspace processing stage can be subdivided into a sphering stage and a dimension reduction stage.

We consider a convolutive BSS model with N sources $s_i(t)$ $(i = 1, \ldots, N)$ and M sensors ($N < M$) 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 noise is considered to be uncorrelated with unit variance. With $E\{\cdot\}$ denoting the expectation value and $\cdot^H$ the hermitian operator, the spatial correlation matrix

is therefore given by

$$E\left\{\mathbf{n}\mathbf{n}^H\right\} = \sigma_n^2\mathbf{I} \tag{2}$$

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.

Thus time-domain signals $\mathbf{x}(t) = [x_1(t), \ldots, x_M(t)]^T$ are converted into frequency-domain time-series signals $\mathbf{X}(f,m) = [X_1(f,m), \ldots X_M(f,m)]^T$ 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 of $\mathbf{X}(f,m)$, we get 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 $W_{ki}(f)$ $(i,k = 1,\ldots,N)$ of the separating system, we solve an ICA problem $\mathbf{Y}(f,m) = \mathbf{W}(f)\ \mathbf{Z}(f,m)$, where $\mathbf{Y}(f,m) = [\ Y_1(f,m),\ \ldots,\ Y_N(f,m)]^T$ and $\mathbf{W}(f)$ is a $N \times N$ matrix whose elements are $W_{ki}(f)$. We call the row vectors of $\mathbf{W}(f)$ unmixing vectors $\mathbf{w}_k^H(f)$. $Y_k(f,m)$ is a frequency-domain representation of the output $y_k(t)$ and they are made so as to be mutually independent.

Then we obtain time-domain filters by applying an inverse discrete Fourier transform (DFT) to $\mathbf{W}(f)$. This has an advantage in that subspace processing and ICA is employed for instantaneous mixtures, which are easier to solve than convolutive ones.

We applied the complex version of FastICA proposed in [7] to $\mathbf{Z}$ to obtain the unmixing matrix $\mathbf{W}$. $\mathbf{Z}$ is assumed to have a zero mean and unit variance. By using negentropy maximization as a basis, the unmixing vector $\mathbf{w}_k$ for each signal is gradually improved by

$$\begin{aligned}\mathbf{w}_k \leftarrow\ & E\left\{\mathbf{Z}(\mathbf{w}_k^H\mathbf{Z})^* g(|\mathbf{w}_k^H\mathbf{Z}|^2)\right\} - \\ & E\left\{g(|\mathbf{w}_k^H\mathbf{Z}|^2) + |\mathbf{w}_k^H\mathbf{Z}|^2 g'(|\mathbf{w}_k^H\mathbf{Z}|^2)\right\}\mathbf{w}_k\end{aligned} \tag{3}$$

until the difference between consecutive unmixing vectors falls below a certain threshold. $\cdot^*$ denotes the complex conjugate. $g(\cdot)$ denotes the derivative of a nonlinear function $G(\cdot)$, which was here chosen as $G(x) = \log(a+x)$ with $a = 0.1$. $\mathbf{w}_k$ is orthonormalized with respect to already existing unmixing vectors after each step.

3. SUBSPACE SELECTION

3.1. Where to place the subspace selection stage

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 as explained in the following.

If we assume virtual sources composed e.g. of noise we could separate as many sources as the number of sensors. Then we could select the wanted sources and therefore the subspace after ICA. But we would face a similar problem to the one that arises when solving the permutation problem, which appears when we apply ICA to convolutive mixtures in the frequency-domain [8, 9]. The more signals we have, the more difficult it is to characterize the components of each frequency bin uniquely and relate them to the components of adjacent frequency bins or distinguish virtual and real sources. We usually have more information before using ICA to select an appropriate subspace (such as sensor spacing and eigenvalues, that give the covariance) than after using ICA (eigenvalues are distorted due to scaling ambiguity). In addition, reducing dimensions before ICA reduces the risk of overlearning of the ICA algorithm due to the virtual sources [10]. In summary it is better to reduce the dimensions before employing ICA.

3.2. 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 $\mathbf{R}_{\mathbf{xx}}$ as given in (4). The principal components are given by the eigenvectors of $\mathbf{R}_{\mathbf{xx}}$ onto which the mixed signals are projected.

$$\mathbf{R}_{\mathbf{xx}} = E\left\{\mathbf{X}\mathbf{X}^H\right\} \tag{4}$$

Asano et al. 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 [11].

Since PCA linearly combines the mixed signals the noise reduction can be backed up by the increase of signal-to-noise ratio (SNR) known from array processing [12]. In the ideal case of coherently adding up several sensors the increased SNR_{new} is given by

$$\text{SNR}_{new} = \log(M) \cdot \text{SNR}_{old} \tag{5}$$

where M denotes the number of sensors and SNR_{old} the SNR at a single sensor.

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.

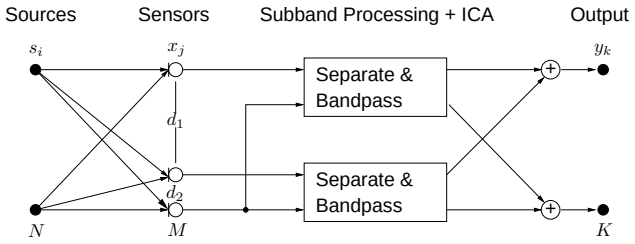

Fig. 2. Geometry-based subspace selection

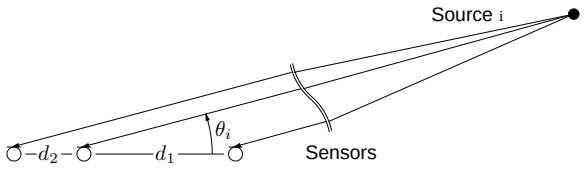

Fig. 3. Definition of source direction

3.3. 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.

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 (6) should be fulfilled.

$$f \leq \frac{\alpha c}{2d\left(\cos(\theta_1) - \cos(\theta_2)\right)} \tag{6}$$

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 as shown in Fig. 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.

Table 1. Summarized comparison

PCA-based selection	**Geometry-based selection**
Statistical consideration	Geometrical considerations
Different subspace for each frequency range	Few different subspaces depending on number of sensors
First sphering, then dimension reduction	First dimension reduction, then sphering

4. GEOMETRICAL UNDERSTANDING OF THE PCA-BASED APPROACH

4.1. Experimental results

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.

We assumed the mixing matrix $\mathbf{H}$ for 3 sensors and 2 sources depending on the frequency to be

$$\mathbf{H} = \begin{bmatrix} b_1 & b_2 \\ b_1 e^{-j\beta_1} & b_2 e^{-j\beta_2} \\ b_1 e^{-2j\beta_1} & b_2 e^{-2j\beta_2} \end{bmatrix} \tag{7}$$

where $\beta_i = 2\pi f d_i \cos\theta_i / c$ and c denotes the sound velocity. It can be derived by using Fig. 3 and assuming that each source signal has a specific but constant attenuation b_i at every sensor and $d_1 = d_2$.

The normalized sensor gain depending on the frequency bin and sensor position for $b_1 = b_2$ is shown in Fig. 4. 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. In Fig. 5 the normalized sensor gain is given for 7 sensors and reveals in particular for low frequencies a very similar behaviour. The outer microphones are preferred and therefore confirmes the idea of the geometry based approach.

To investigate the effect of PCA in even more detail we analyzed the eigenvectors and eigenvalues of the correlation matrix $\mathbf{R_{xx}}$ of the mixed signals. A typical result for the first and second principal components represented by the eigenvectors with the largest and second largest eigenvalues, respectively, is shown in Figs. 6 and 7 for each frequency bin. The figures were generated with the same conditions as Fig. 4

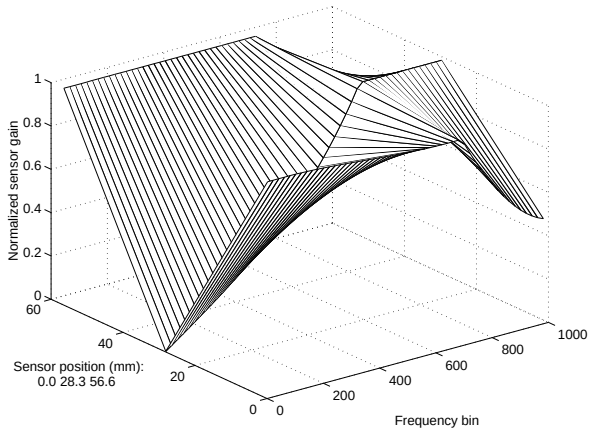

Fig. 4. Normalized sensor gain with PCA-based subspace selection for 3 sensors

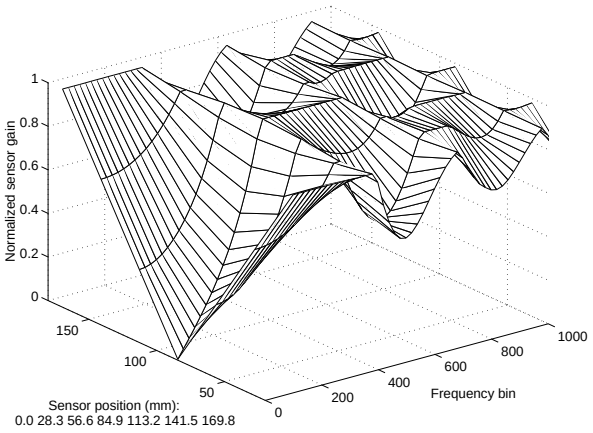

Fig. 5. Normalized sensor gain with PCA-based subspace selection for 7 sensors

4.2. Interpretation of experimental results

For the first principal component all sensors contribute approximately with the same gain for low frequencies. The second principal component in Fig. 7 already shows the emphasis of the outer sensors for lower frequencies. This behavior can be backed up by the following considerations.

Using the mixing matrix in (7) we obtain the mixed signals $\mathbf{X}$ in the frequency domain as

$$\mathbf{X}=\mathbf{HS}=\mathbf{H}\begin{bmatrix} S_1 \\ S_2 \end{bmatrix}=\begin{bmatrix} b_1S_1+b_2S_2 \\ b_1e^{-j\beta_1}S_1+b_2e^{-j\beta_2}S_2 \\ b_1e^{-2j\beta_1}S_1+b_2e^{-2j\beta_2}S_2 \end{bmatrix} \quad (8)$$

where $\mathbf{S}$ denotes the frequency-domain time-series of the source signals $\mathbf{s}$ according to Sec. 2. The scalar product of the mixed signals $\mathbf{X}$ and the principal component $\mathbf{p} = [p_1\ p_2\ p_3]^T$ yields

$$\mathbf{p}^H\mathbf{X}=\mathbf{p}^H\mathbf{HS}=b_1S_1(p_1^*+p_2^*e^{-j\beta_1}+p_3^*e^{-2j\beta_1})+ b_2S_2(p_1^*+p_2^*e^{-j\beta_2}+p_3^*e^{-2j\beta_2}) \quad (9)$$

For low frequencies the phase difference β_i becomes very small and we can approximate it by the least square error (LSE) solution $\bar{\beta}_i$ of

$$\min_{\bar{\beta}_i}||(p_1^*+p_2^*e^{-j\beta_i}+p_3^*e^{-2j\beta_i})- (p_1^*e^{-\bar{\beta}_i}+p_2^*e^{-\bar{\beta}_i}+p_3^*e^{-\bar{\beta}_i})|| \quad (10)$$

Thus we get

$$\mathbf{p}^H\mathbf{X}\approx(b_1S_1e^{-j\bar{\beta}_1}+b_2S_2e^{-j\bar{\beta}_2})(p_1^*+p_2^*+p_3^*) \quad (11)$$

The first principal component in Fig. 6 is found by maximizing the power $E\{(\mathbf{p}^H\mathbf{x})(\mathbf{p}^H\mathbf{x})^H\}$ with the constraint $||\mathbf{p}||=1$. By the Lagrange multiplier approach

$$\nabla\left(E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\}+\gamma(||\mathbf{p}||-1)\right)=\mathbf{0} \quad (12)$$

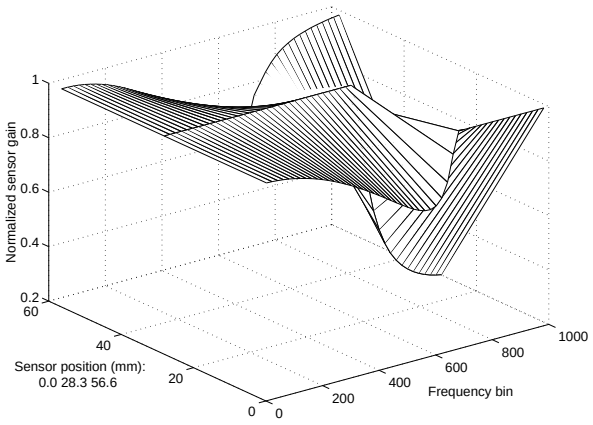

Fig. 6. Normalized first principal component for 3 sensors

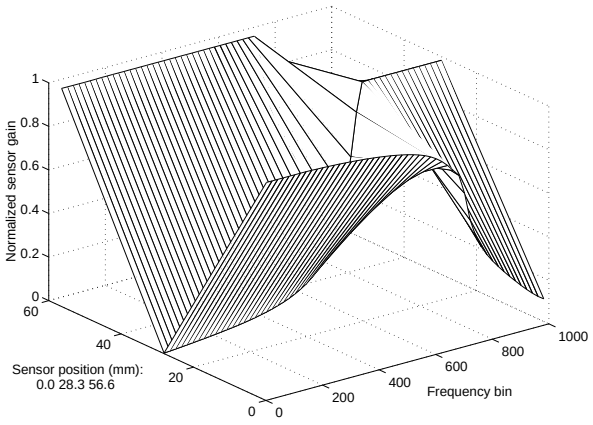

Fig. 7. Normalized second principal component for 3 sensors

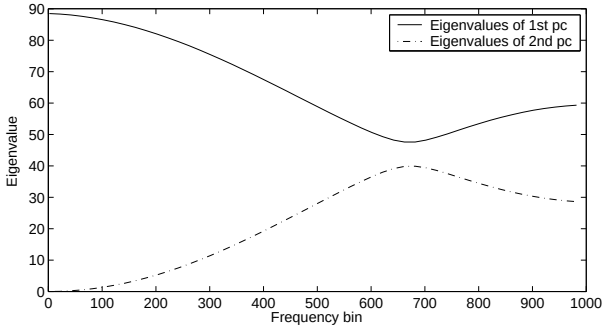

Fig. 8. Typical absolute eigenvalues of first and second component for each frequency bin

where ∇ is the Nabla operator and γ the Lagrange multiplier.

We can show that with the approximation in (11) the maximum is obtained if $p_1 = p_2 = p_3$. This means that all sensors have approximatly equal influence. In this case the LSE solution for $\bar{\beta}_i$ equals β_i.

To explain the emphasis of the outer two sensors with the second principal component we show that the second sensor is completely contained in the first principal component. The projection of the mixed signal on the first principal component yields

$$\frac{\mathbf{p}^H\mathbf{X}}{\mathbf{p}^H\mathbf{p}}\mathbf{p} \overset{\substack{||\mathbf{p}||=1 \\ p_1=p_2=p_3}}{\approx} \begin{bmatrix} b_1 e^{-j\beta_1} S_1 + b_2 e^{-j\beta_2} S_2 \\ b_1 e^{-j\beta_1} S_1 + b_2 e^{-j\beta_2} S_2 \\ b_1 e^{-j\beta_1} S_1 + b_2 e^{-j\beta_2} S_2 \end{bmatrix} \quad (13)$$

From comparing this result with (8) it follows that the middle sensor is nearly exactly represented by the first principal component. In contrast it does not exactly represent the outer two sensors. Thus, to be able to represent them by the principal components they must be considered again in the second principal component. This results in the emphasis of the outer sensors by the second principal component.

We will now explain why the second principal component is dominant even after employing ICA. After projecting the mixed signals onto the principal components they are normalized by the inverse square root of the respective eigenvalue. A typical frequency dependent eigenvalue distribution is shown in Fig. 8. For low frequencies the eigenvalue of the first component is very large compared to the eigenvalue of the second component. This means in turn that the first component is attenuated and the second component is amplified. Thus the second component has a dominant influence when it is combined with the first principal component by the subsequent ICA stage.

Different settings used in additional experiments also show basically the same behavior, particularly for low frequencies.

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}

5. COMPARISON OF THE PCA- AND GEOMETRY-BASED APPROACHES

5.1. Experimental results

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 [13]. The source directions θ_i were estimated by the MUSIC algorithm [14]. The frequency ranges were calculated based on the criteria discussed in Sec. 3.3.

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 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.

Figures 9 and 10 show the results for both methods for 12 pairs of speech signals. Figure 9 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. 10). We confirmed the superior performance with additional experiments using different sensor spacings.

5.2. Interpretation of experimental results

To interprete the experimental results of Sec. 5.1 we distinguish between the noiseless and noisy case.

As we have seen in Sec. 4.1 the PCA based method also emphasizes the outer microphones for low frequencies.

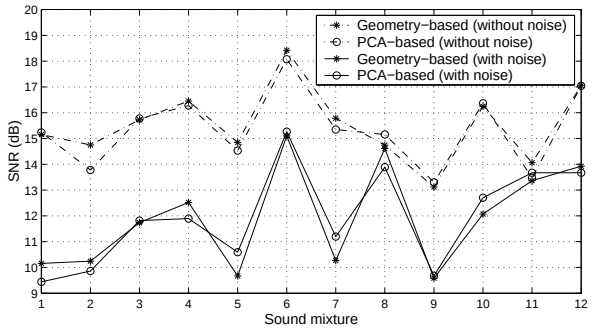

Fig. 9. Comparison of PCA- and geometry-based subspace selection for low frequency range ($0 - 2355$ Hz)

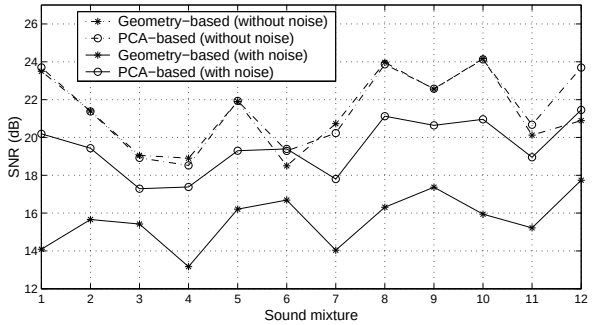

Fig. 10. Comparison of PCA- and geometry-based subspace selection for high frequency range ($2356 - 4000$ Hz)

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. 3.3.

Therefore the contribution of the middle sensor is very small for low frequencies. In addition the PCA based method might have problems in finding appropriate principal components due to low phase differences which are disturbed by noise. Thus the PCA based approach can not make great use of the remaining sensor for low frequencies either and therefore does not improve the performance.

As stated in Sec. 3.2 uncorrelated noise is normally reduced if we coherently add up the mixtures received at several sensors. The PCA based method can utilize all available sensors for high frequencies, since then the smaller sensor distance is appropriate. In contrast the geometry based approach uses by definition always only two sensors. Thus the latter cannot exploit the noise reduction as much as the PCA based approach can.

In the noiseless case the advantage of noise suppression of the PCA based method has no effect and therefore does not improve the result.

6. 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.

7. REFERENCES

[1] A. Hyvärinen, 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] E. Bingham and A. Hyvärinen, "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.

[8] 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.

[9] 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.

[10] A. Hyvärinen, J. Särelä, and R. Vigrio, "Bumps and spikes: Artifacts generated by independent component analysis with insufficient sample size," in *Proc. ICA99*, 1999, pp. 425–429.

[11] 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.

[12] S.U. Pillai, *Array signal processing*, Springer Verlag, 1989.

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

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