Geometrical Interpretation of the PCA Subspace Approach for Overdetermined Blind Source Separation

S. Winter,[1, 2] H. Sawada,[1] and S. Makino[1]

[1] *NTT Communication Science Laboratories, NTT Corporation, 2-4 Hikaridai Seika-cho, Soraku-gun, Kyoto 619-0237, Japan*
[2] *Department of Multimedia Communication and Signal Processing, University of Erlangen-Nuremberg, 91058 Erlangen, Germany*

We discuss approaches for blind source separation where we can use more sensors than sources to obtain a better performance. The discussion focuses mainly on reducing the dimensions 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 in accordance with 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], the use of 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 output signal space with reduced dimensions.

In this paper we present results for 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, which depends on the frequency and sensor spacing as described in [6].

We compared the two methods by performing experiments with real world data in a reverberant environment. We found that for low frequencies the PCA-based method behaves similarly to the geometry-based method, and support this result analytically. For high frequencies the former method 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. This paper generalizes the results in [7] to include arbitrary arrangements of arbitrary numbers of sensors.

2. BSS USING MORE SENSORS THAN SOURCES

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

We consider a convolutive BSS model with N sources $s_i(t)$ $(i = 1, \ldots, N)$ at positions $\mathbf{q}_i$ and M sensors $(N < M)$ that give mixed signals $x_j(t)$ $(j = 1, \ldots, M)$ at positions $\mathbf{r}_j$ 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 temporally and spatially uncorrelated with unit variance and Gaussian distribution. With $E\{\cdot\}$ denoting the expectation value and the superscript H the hermitian operator, the spatial correlation
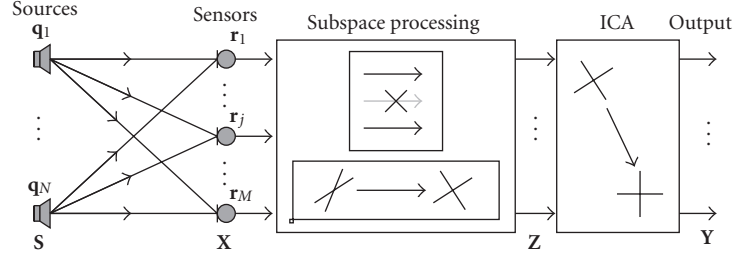
FIGURE 1: General framework of overdetermined BSS.

matrix is therefore given by

$$E\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2\mathbf{I}, \qquad (2)$$

where $\mathbf{n} = [n_1, \ldots, n_M]^T$.

We employed a narrowband frequency-domain approach to solve the convolutive BSS problem including the subspace processing [8]. 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 time-frequency domain signals $\mathbf{X}(f, m) = [X_1(f, m), \ldots, X_M(f, m)]^T$ by an L-point sliding window discrete Fourier transform (DFT), where $f = 0, f_s/L, \ldots, f_s(L-1)/L$ (f_s is sampling frequency, m is time index). After the subspace processing of $\mathbf{X}(f, m)$, we obtain 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

$$\mathbf{Y}(f, m) = \mathbf{W}(f)\mathbf{Z}(f, m), \qquad (3)$$

where $\mathbf{Y}(f, m) = [Y_1(f, m), \ldots, Y_N(f, m)]^T$ and $\mathbf{W}(f)$ is an $N \times N$ matrix whose elements are $W_{ki}(f)$. We call the conjugately transposed row vectors of $\mathbf{W}(f)$ separation vectors $\mathbf{w}_k(f) = [W_{k1}, \ldots, W_{kM}]^H \cdot Y_k(f, m)$ is a frequency-domain representation of the output $y_k(t)$. The output signals $\mathbf{Y}(f, m)$ are made mutually independent.

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

We applied the complex version of FastICA proposed in [9] to $\mathbf{Z}$ to obtain the separation matrix $\mathbf{W} \cdot \mathbf{Z}$ is assumed to have a zero mean and unit variance. By using negentropy maximization as a basis, the separation vector $\mathbf{w}_k$ for each signal is gradually improved by

$$\mathbf{w}_k \longleftarrow E\left\{\mathbf{Z}(\mathbf{w}_k^H\mathbf{Z})^* g\left(|\mathbf{w}_k^H\mathbf{Z}|^2\right)\right\}$$
$$- E\left\{g\left(|\mathbf{w}_k^H\mathbf{Z}|^2\right) + |\mathbf{w}_k^H\mathbf{Z}|^2 g'\left(|\mathbf{w}_k^H\mathbf{Z}|^2\right)\right\}\mathbf{w}_k \qquad (4)$$

until the difference between consecutive separation 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 after each step to already existing separation vectors.

3. SUBSPACE SELECTION

3.1. Relative order of subspace selection and signal separation

The use of more sensors than sources usually improves the separation result. We can exploit the performance improvements due to beamforming principles. For the signal separation, we have to employ some form of dimension reduction in order to map the number of mixed signals to the number of output signals. It appears to be preferable to reduce the dimensions before rather than after ICA as explained in the following.

If we assume virtual sources composed, for example, of noise we could separate as many sources as sensors. Then we could select the desired 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 [10, 11]. 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. Normally more information is available before we use ICA to select an appropriate subspace (e.g., sensor spacing and eigenvalues that give the covariance) than afterwards (eigenvalues are distorted due to scaling ambiguity). In addition, reducing dimensions before ICA reduces the risk of overlearning of the ICA algorithm caused by the virtual sources [12]. 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 to select 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 (5). The principal components are given by the eigenvectors of $\mathbf{R}_{\mathbf{xx}}$ onto which the mixed signals are projected,

$$\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{X}\mathbf{X}^H\}. \qquad (5)$$

In a practical sense Asano et al. [5] consider room reflections to be uncorrelated noise from the direct source signals $s_i(t)$ on condition that the time shift between direct and
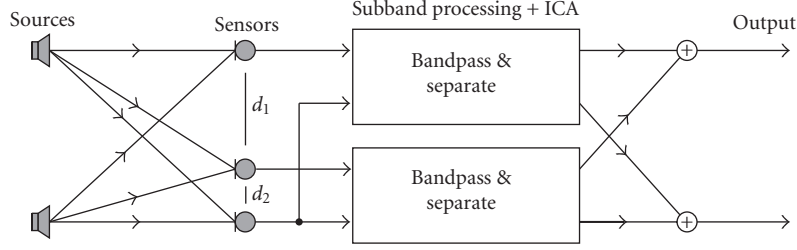
FIGURE 2: Geometry-based subspace selection.

reflected signals is greater than the window length used for the 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. Let $\mathbf{E}$ denote a matrix with the first N principal components and $\mathbf{D}$ a diagonal matrix with the corresponding eigenvalues. Then the sphering matrix $\mathbf{V}$ that is used to project mixed signals $\mathbf{X}$ to $\mathbf{Z}$ is given by

$$\mathbf{V} = \mathbf{D}^{-1/2} \cdot \mathbf{E}^H. \tag{6}$$

Thus by selecting the subspace that is spanned by the first N principal components, dimensions are effectively reduced by removing noise while retaining the signal of interest [13].

Since PCA linearly combines the mixed signals, the noise reduction can be backed up by an increase in the signal-to-noise ratio (SNR) known from array processing [14]. In an ideal case of coherently adding together the signals of several sensors, which are disturbed by spatially and temporally uncorrelated noise, the increased SNR_{new} is given by

$$\text{SNR}_{\text{new}} = 10\log_{10}(M) + \text{SNR}_{\text{single}}. \tag{7}$$

M denotes the number of sensors and $\text{SNR}_{\text{single}}$ the SNR at a single sensor.

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

3.3. Subspace selection based on geometrical knowledge

A method for blind source separation has been proposed using several separating subsystems whose sensor spacings 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 $d_1 > d_2$ using one sensor as a common sensor as shown in Figure 2.

The frequency range of the mixed signals was divided into lower and higher frequency ranges. According to [8], for
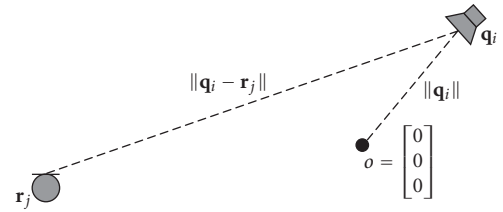

FIGURE 3: Near-field model.

a frequency to be adequate for a specific sensor and source arrangement, the condition in (8) should be fulfilled:

$$f < \left| \frac{\alpha c}{2 \cdot (\|\mathbf{q}_i - \mathbf{r}_j\| - \|\mathbf{q}_i\|)} \right|. \tag{8}$$

Here α is a parameter that governs the degree to which the phase difference exceeds π, c the sound velocity, $\mathbf{r}_j$ the position of the jth sensor, and $\mathbf{q}_i$ the position of the ith source as shown in the general near-field model in Figure 3.

The appropriate sensor pairs were chosen for each frequency range and individually 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 place 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. GEOMETRICAL UNDERSTANDING OF PCA- BASED APPROACH

4.1. Experimental results

We examined the behavior of 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.

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

FIGURE 4: Far-field model.

We assumed the frequency-dependent mixing matrix **H** to be

$$\mathbf{H}(f) = \left(e^{J(2\pi f/c)(\|\mathbf{q}_i - \mathbf{r}_j\| - \|\mathbf{q}_i\|)} \right)_{ji}, \tag{9}$$

where $1 \le i \le N$, $1 \le j \le M$, and c denotes the sound velocity. **H** can be derived assuming a far-field model for the attenuation (Figure 4). The far-field assumption results in a specific but constant attenuation at every sensor for each source signal. Therefore we assume without loss of generality that the attenuation is included in the signal amplitude and omit it from the mixing matrix. For simplicity the phase is based on a more general near-field model (Figure 3) and depends on the differences between the exact distances from the source to the sensor $\|\mathbf{q}_i - \mathbf{r}_j\|$ and to the origin $\|\mathbf{q}_i - [0, \ldots, 0]^T\|$.

The amplitudes of the sensor weights (sensor gains) for a specific output signal of an equispaced linear sensor array (i.e., $\mathbf{r}_2 - \mathbf{r}_1 = \mathbf{r}_3 - \mathbf{r}_2$) for $M = 3$, $N = 2$, are shown in Figure 5 for a specific output signal. They depend on the frequency bin and sensor position. Since they look similar for all output signals, the sensor gains are only shown for one output signal. The unnormalized sensor gains are given by the corresponding row of **WV**. For better comparison the sensor gains are normalized by the respective maximum in each frequency bin. We used the experimental conditions given in the first two lines of Table 2. We can see that the PCA-based method also emphasizes 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 the more the higher the frequency becomes. In Figure 6 the normalized sensor gain is given for $M = 7$ linearly arranged sensors with equal spacing and reveals very similar behavior, particularly for low frequencies. The outer microphones are preferred, which confirms the idea of the geometry-based approach.

Figure 7 was generated under the same conditions as Figure 5 except that we used real speech signals and impulse responses instead of artificial signals and a mixing matrix. Although not as smooth as Figure 5, it still illustrates the principle that outer sensors are preferred for low frequencies and justifies the assumptions made for the ideal case.

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

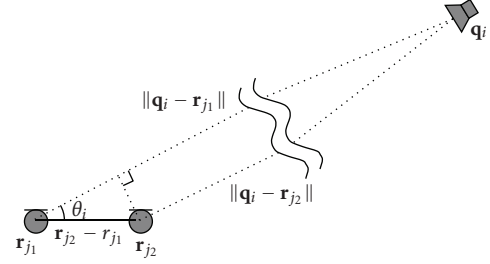and second principal components represented by the eigenvectors with the largest and second largest eigenvalues, respectively, is shown in Figures 8 and 9 for each frequency bin. The figures were generated under the same conditions as Figure 5.

4.2. Interpretation of experimental results

Based on our analytical results as detailed in Appendix A, in this section we provide an explanation for the sensor weighting for low frequencies that is illustrated for the first and second principal components in Figures 8 and 9, respectively. We will then show how the eigenvalues of the correlation matrix $\mathbf{R_{xx}}$ influence the combination of the principal components and contribute to the overall sensor weighting as observed in Figure 5.

For low frequencies the first principal component in Figure 8 weights every sensor approximately equally. This experimental result can be backed up analytically for arbitrary sensor arrangements. Based on the mixing model in (9), the mixed signals **X** are given by

$$\mathbf{X}(f, m)$$
$$= \mathbf{H}(f)\mathbf{S}(f, m) = \left(\sum_{i=1}^{N} S_i(f, m) e^{J(2\pi f/c)(\|\mathbf{q}_i - \mathbf{r}_j\| - \|\mathbf{q}_i\|)} \right)_j. \tag{10}$$

$\mathbf{S}(f, m) = [S_1(f, m), \ldots, S_N(f, m)]$ denotes the time-frequency domain representation of the source signals **s** according to Section 2. Due to the far-field assumption the attenuation from the ith source to an arbitrary sensor is independent of the selected sensor. Therefore, without loss of generality we assume that the attenuation is included in the signal amplitude S_i.

For low frequencies the phase difference between two sensors for a signal from source i,

$$\Delta\varphi_i = \frac{2\pi f}{c} (\|\mathbf{q}_i - \mathbf{r}_{j_1}\| - \|\mathbf{q}_i - \mathbf{r}_{j_2}\|), \tag{11}$$

becomes very small. Therefore we can approximate the phase $\varphi_{ji} := (2\pi f/c)\|\mathbf{q}_i - \mathbf{r}_j\|$ by the least square error (LSE) solution

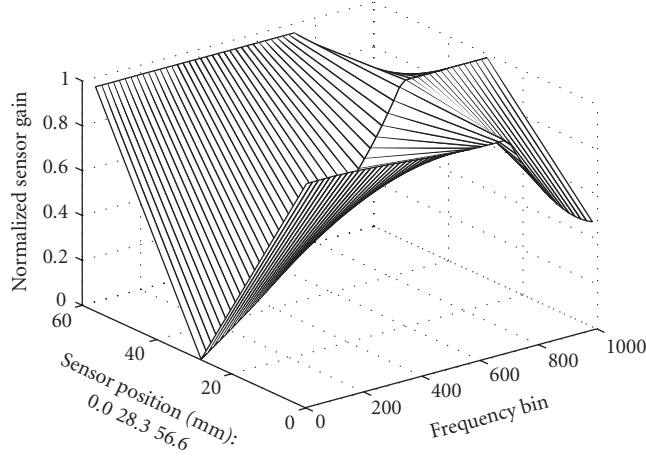$$\overline{\varphi}_i := \frac{2\pi f}{c} \|\mathbf{q}_i - \mathbf{r}\|. \tag{12}$$

FIGURE 5: Normalized sensor gain with PCA-based subspace selection for 3 sensors (artificial signals).

TABLE 2: Experimental conditions.

Source direction	50° and 150°
Sensor distance	$d_1 = d_2 = 28.3$ mm
Source signal duration	7.4 s
Reverberation time	$T_{60} = 200$ ms
Sampling rate	8 kHz
DFT length	250 ms
Window type	von Hann
Filter length	2048 points
Shifting interval	512 points
Frequency range parameter	$\alpha = 1.2$
Threshold for FastICA	10^{-3}
Added sensor noise	≈ -14 dB

$\overline{\varphi}_i$ is independent of the sensor j and turns out to be the solution of

$$\sum_{j=1}^{M} \sin(\varphi_{ji} - \overline{\varphi}_i) = A \sin(\varphi_{i\Sigma} - \overline{\varphi}_i) = 0. \qquad (13)$$

It is given by

$$\overline{\varphi}_i = \varphi_{i\Sigma} + k\pi, \quad k \in \mathbb{N}. \qquad (14)$$

A and $\varphi_{i\Sigma}$ are the parameters of the single sinewave that results from the summation of the sinewaves on the left-hand side of (13). The parameters can be determined by a vector diagram as shown in Figure 10. The definition of $\overline{\varphi}_i$ in (12) is based on $\mathbf{r}$, which can be interpreted as the position of a virtual sensor. Its signal can be approximated by using only the first principal component.

If a virtual sensor coincides with an actual sensor, then the first principal component is sufficient to describe its signal. No higher order principal component is necessary. The further away an actual sensor is from the virtual sensor(s), the more correction is required by higher order principal components to describe the mixed signal at this sensor. This
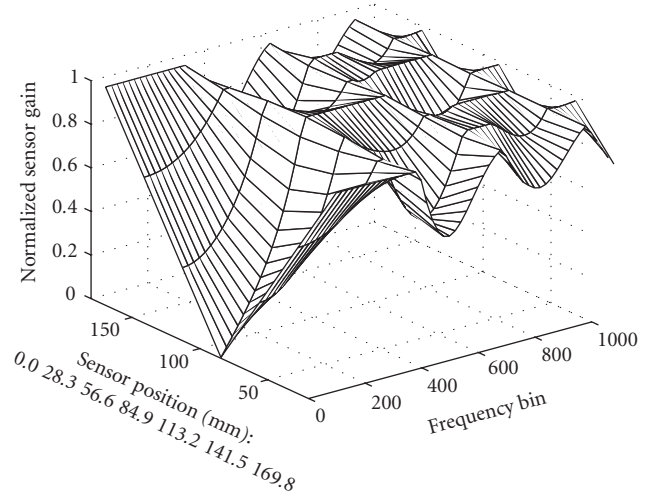

FIGURE 6: Normalized sensor gain with PCA-based subspace selection for 7 sensors (artificial signals).

is important when it comes to the final sensor selection as described later. With an equally spaced linear sensor array the average position of all sensors becomes a possible solution for $\mathbf{r}$ (cf. (A.28)):

$$\mathbf{r} = \frac{1}{M} \sum_{j=1}^{M} \mathbf{r}_j. \qquad (15)$$

If in addition there is an odd number of sensors as in Figures 5 and 6, the central sensor's signal is completely described by the first principal component. However, as we will see later, the first principal component contributes almost nothing to the final result. This explains why the signal of the central sensor is hardly present in the final result (Figure 5).

With the approximation (12) the first principal component $\mathbf{p} = [p_1, \ldots, p_M]$ can now be determined. Following the definition of the (first) principal component we maximize the power $E\{(\mathbf{p}^H \mathbf{X})^* (\mathbf{p}^H \mathbf{X})\}$ with the constrain $\|\mathbf{p}\| = 1$.
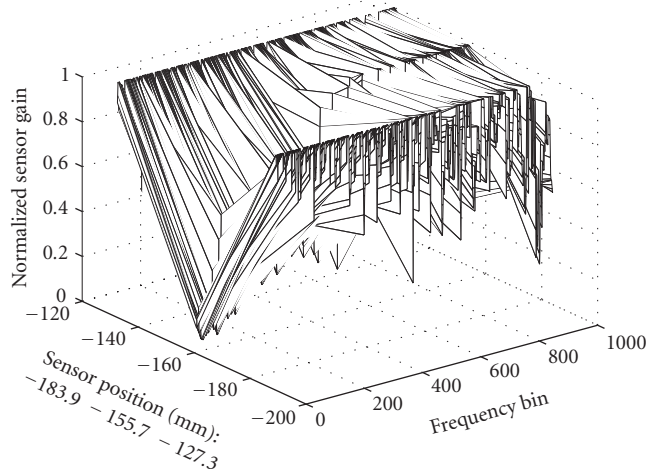
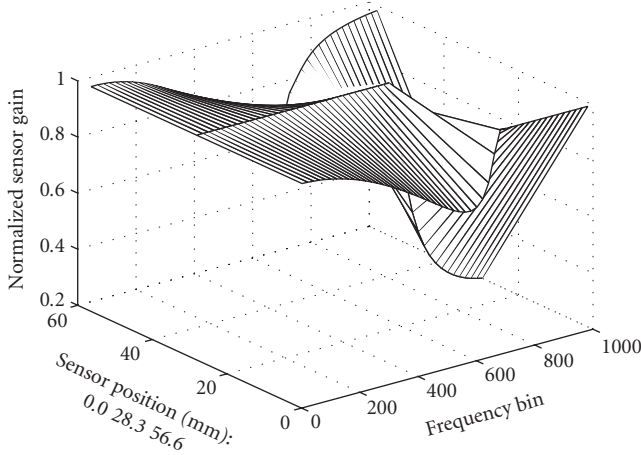FIGURE 7: Normalized sensor gain with PCA-based subspace selection for 3 sensors (real speech signals).

FIGURE 8: Normalized first principal component for 3 sensors.

Without any further assumptions the elements p_j result in a constant

$$p_j = p = \text{const} \quad \forall j. \tag{16}$$

This means that for the first principal component all sensors contribute approximately with the same gain for low frequencies (Figure 8).

Since the first principal component describes the signal of the virtual sensor at **r** almost completely and principal components are orthogonal to each other, this signal will not be included in higher order principal components. Instead, higher order principal components will describe the signals at different positions. This explains why in Figure 9 the central sensor has nearly zero gain and the outer sensors are emphasized for low frequencies.

Now we take a look at the corresponding eigenvalues of $\mathbf{R_{xx}}$. According to (6) the square roots of their inverses determine the weight of each principal component. By definition, as the order of a principal component increases, its eigenvalue decreases. Typical eigenvalues depending on the frequency are shown in Figure 11. For low frequencies the eigenvalue corresponding to the first principal component is very large compared with the eigenvalue corresponding to the second principal component. This in turn means that the first and second principal components are attenuated and amplified, respectively, by their inverses. Thus the second and higher order principal components have a dominant influence when they are combined with the first principal component by the subsequent ICA stage. Therefore, the closer a sensor is to the virtual sensor position **r** of the first principal component, the less it contributes to the final result.

Different settings, such as the unequally spaced sensors used in our additional experiments, also exhibit basically the same behavior, particularly for low frequencies.

5. COMPARISON OF 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 [15]. The frequency ranges were calculated based on the criteria discussed in Section 3.3.

We measured the performance in terms of the signal-to-noise plus interference ratio (SNIR) in dB. It is given for output k by

$$\text{SNIR}_k = 10 \log \frac{\sum_t y_k^s(t)^2}{\sum_t y_k^c(t)^2}, \tag{17}$$

where $y_k^s(t)$ is the portion of $y_k(t)$ that comes from a source signal $s_k(t)$ and $y_k^c(t) = y_k(t) - y_k^s(t)$.
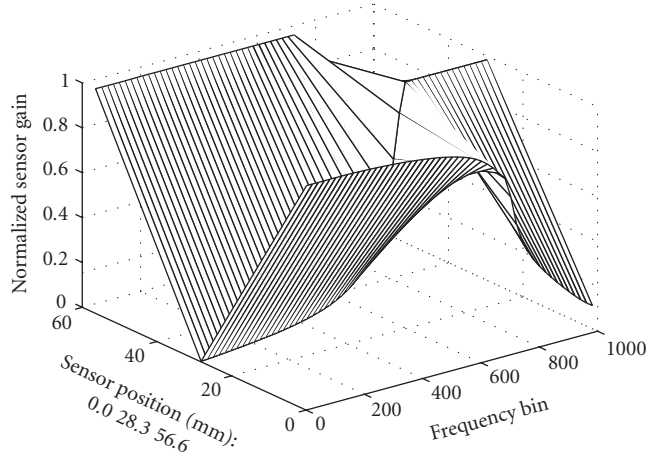
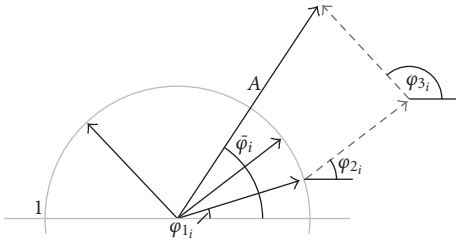FIGURE 9: Normalized second principal component for 3 sensors.

FIGURE 10: Vector diagram for determining $\overline{\varphi}_i$ ($M = 3$).
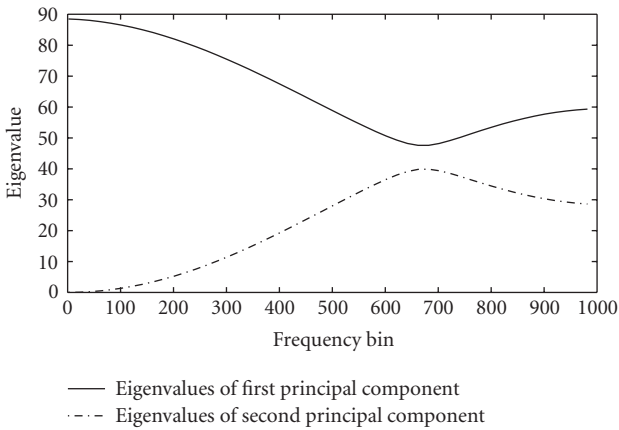

FIGURE 11: Typical absolute eigenvalues of first and second principal components for each frequency bin.

To avoid the permutation problem influencing the result we selected the permutation that resulted in the highest SNIR in each frequency bin. This SNIR was calculated in a similar way to that described above. The solution is identical to that obtained when the permutation problem is perfectly solved. The experimental conditions are given in Table 2.

Figures 12 and 13 show the results with both methods for 12 pairs of speech signals. Figure 12 reveals that both subspace methods exhibit 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 (Figure 13). We confirmed the superior performance with additional experiments using different sensor spacings.

5.2. Interpretation of experimental results

To interpret the experimental results described in Section 5.1 we distinguish between noiseless and noisy cases.

As we have seen in Section 4.1 the PCA-based method also emphasizes the outer microphones for low frequencies. 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 Section 3.3.

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

As stated in Section 3.2 temporally and spatially uncorrelated noise is normally reduced if we coherently combine 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, by definition, always uses only two sensors, and so cannot exploit the noise reduction as much as the PCA-based approach.

In the noiseless case the noise suppression advantage provided by the PCA-based method has no effect and therefore does not improve the result.
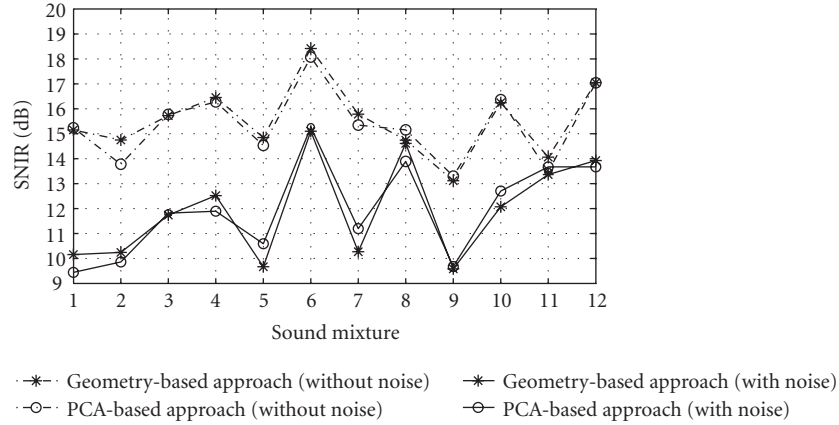
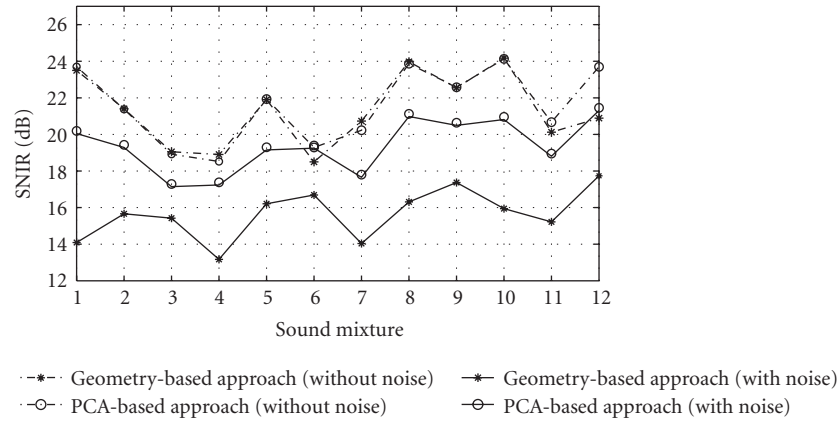FIGURE 12: Comparison of PCA- and geometry-based subspace selection for low frequency range (0–2355 Hz).

FIGURE 13: Comparison of PCA- and geometry-based subspace selection for high frequency range (2356–4000 Hz).

6. CONCLUSION

We have compared two subspace methods for use as preprocessing steps in overdetermined BSS. We found experimentally and analytically 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 the appropriate phase difference means it can utilize all pairs of sensors to suppress the noise. This deepens the geometrical understanding of the PCA-based method.

APPENDIX

A. DERIVATION OF SENSOR SELECTION BY PCA FOR LOW FREQUENCIES

Experimental results have shown that the first principal component weights all sensors equally for low frequencies (Section 4.1). As a result, the central sensors contribute far less than the outer sensors to the final sensor selection. In this section we analytically derive the equal weighting of the first

principal component and determine the position of the virtual sensor whose signal is completely represented by the first principal component. After initial definitions and approximations obtained using virtual sensor in Section A.1 we derive the first principal component in Section A.2. Finally, we determine the position of the virtual sensor in Section A.3 as a least square error (LSE) solution. An outline of this derivation can be found in Section 4.2.

A.1. Definitions and assumptions

We assume a mixing system with N sources and M sensors under far-field conditions. The frequency-domain time series of the source signals $\mathbf{s} = [s_1, \ldots, s_N]^T$ according to Section 2 are given by

$$\mathbf{S}(f, m) := [S_1(f, m), \ldots, S_N(f, m)]^T. \qquad (A.1)$$

According to Section 4.1, the frequency-dependent mixing matrix can be written as

$$\mathbf{H}(f) = \left(e^{J(2\pi f/c)(\|\mathbf{q}_i - \mathbf{r}_j\| - \|\mathbf{q}_i\|)} \right)_{ji}, \qquad (A.2)$$

where c denotes the sound velocity, $\mathbf{q}_i$ the ith source location, and $\mathbf{r}_j$ the jth sensor location. The far-field assumption means that the attenuation from the ith source to a sensor is independent of the selected sensor. Therefore, without loss of generality, we assume that the attenuation is included in the signal amplitude S_i. Then we obtain the mixed signal vector $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{HS} = \left(\sum_{i=1}^{N} S_i e^{J(2\pi f/c)(\|\mathbf{q}_i - \mathbf{r}_j\| - \|\mathbf{q}_i\|)} \right)_j. \quad (A.3)$$

We define an arbitrary eigenvector of the covariance matrix $\mathbf{R_{xx}}$ which corresponds to a principal component by

$$\mathbf{p} := [p_1, \ldots, p_N]^T, \quad \|\mathbf{p}\| = 1. \quad (A.4)$$

The scalar product of $\mathbf{p}$ and the mixed signals $\mathbf{X}$ yield

$$\begin{aligned}
\mathbf{p}^H \mathbf{X} &= \mathbf{p}^H \mathbf{HS} \\
&= \sum_{j=1}^{M} p_j^* \sum_{i=1}^{N} S_i e^{J(2\pi f/c)(\|\mathbf{q}_i - \mathbf{r}_j\| - \|\mathbf{q}_i\|)} \\
&= \sum_{i=1}^{N} S_i e^{-J(2\pi f/c)\|\mathbf{q}_i\|} \sum_{j=1}^{M} p_j^* e^{J(2\pi f/c)\|\mathbf{q}_i - \mathbf{r}_j\|}.
\end{aligned} \quad (A.5)$$

For low frequencies the phase difference $(2\pi f/c)(\|\mathbf{q}_i - \mathbf{r}_{j_1}\| - \|\mathbf{q}_i - \mathbf{r}_{j_2}\|)$ between two sensors becomes very small, and we can approximate the phase $\varphi_{ji} := (2\pi f/c)\|\mathbf{q}_i - \mathbf{r}_j\|$ in (A.5) by the LSE solution $\overline{\varphi}_i$ of

$$\arg\min_{\overline{\varphi}_i} \left\| \sum_{j=1}^{M} p_j^* e^{J\varphi_{ji}} - e^{J\overline{\varphi}_i} \sum_{j=1}^{M} p_j^* \right\|^2. \quad (A.6)$$

Then we can approximate $\mathbf{p}^H \mathbf{X}$ by

$$\mathbf{p}^H \mathbf{X} \approx \left(\sum_{i=1}^{N} S_i e^{-J(2\pi f/c)\|\mathbf{q}_i\|} e^{J\overline{\varphi}_i} \right) \left(\sum_{j=1}^{M} p_j^* \right). \quad (A.7)$$

A.2. Derivation of first principal component

The first principal component is found by maximizing the power

$$E\left\{ (\mathbf{p}^H \mathbf{X})(\mathbf{p}^H \mathbf{X})^* \right\} \quad (A.8)$$

with the constraint $\|\mathbf{p}\|^2 = 1$. This leads to a constrained problem

$$\max_{\mathbf{p}} E\left\{ (\mathbf{p}^H \mathbf{X})(\mathbf{p}^H \mathbf{X})^* \right\}, \quad \|\mathbf{p}\|^2 = 1. \quad (A.9)$$

By defining

$$\begin{aligned}
\overline{x^2} := E\Bigg\{ &\left(\sum_{i=1}^{N} S_i e^{-J(2\pi f/c)\|\mathbf{q}_i\|} e^{J\overline{\varphi}_i} \right) \\
&\cdot \left(\sum_{i=1}^{N} S_i e^{-J(2\pi f/c)\|\mathbf{q}_i\|} e^{J\overline{\varphi}_i} \right)^* \Bigg\}
\end{aligned} \quad (A.10)$$

and using (A.7), we can approximate (A.8) by

$$\begin{aligned}
&E\left\{ (\mathbf{p}^H \mathbf{X})(\mathbf{p}^H \mathbf{X})^* \right\} \\
&\approx E\Bigg\{ \left(\sum_{i=1}^{N} S_i e^{-J(2\pi f/c)\|\mathbf{q}_i\|} e^{J\overline{\varphi}_i} \right) \left(\sum_{j=1}^{M} p_j^* \right) \\
&\qquad \cdot \left(\sum_{i=1}^{N} S_i e^{-J(2\pi f/c)\|\mathbf{q}_i\|} e^{J\overline{\varphi}_i} \right)^* \left(\sum_{j=1}^{M} p_j \right) \Bigg\} \\
&= \left(\underbrace{\sum_{j=1}^{M} p_j p_j^*}_{=\|\mathbf{p}\|=1} + \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j \right) \cdot \overline{x^2} \\
&= \left(1 + \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j \right) \cdot \overline{x^2}.
\end{aligned} \quad (A.11)$$

Since $\overline{x^2}$ does not depend on $\mathbf{p}$ we only have to maximize the first part of (A.11). Therefore (A.9) becomes

$$\max_{\mathbf{p}} \left(1 + \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j \right), \quad \|\mathbf{p}\|^2 = 1. \quad (A.12)$$

Using the Lagrange multipliers approach [16] with δ being the Lagrange multiplier we obtain the following problem:

$$\begin{aligned}
&\nabla_{\mathbf{p}} \left(\left(1 + \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j \right) + \delta(\|\mathbf{p}\|^2 - 1) \right) \\
&= \underbrace{\nabla_{\mathbf{p}}(1 - \delta)}_{=0} + \nabla \left(\sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j + \delta\|\mathbf{p}\|^2 \right) \\
&= \left(\frac{\partial}{\partial p_i} \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j + \delta \sum_{i=1}^{M} p_i^* p_i \right)_i \\
&= 2 \left(\sum_{j \neq i} p_j + \delta p_i \right)_i = \mathbf{0}.
\end{aligned} \quad (A.13)$$

This linear equation can be written as

$$2 \begin{bmatrix} \delta & & 1 \\ & \ddots & \\ 1 & & \delta \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix} = \mathbf{0}. \quad (A.14)$$

We obtain a nontrivial solution if and only if

$$\det \begin{bmatrix} \delta & & 1 \\ & \ddots & \\ 1 & & \delta \end{bmatrix} = (\delta - 1)^{M-1}(\delta + (M-1)) = 0, \quad (A.15)$$

that is, for $\delta_1 = 1$ or $\delta_2 = 1 - M$.

Solution for $\delta = \delta_1 = 1$

The solution is as follows:

$$p_i = -\sum_{j \neq i} p_j. \quad (A.16)$$

Using (A.16) in (A.11) yields

$$
\left(1 + \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j\right) \cdot \overline{x^2} = \left(1 + \sum_{i=1}^{M} p_i^* \sum_{j \neq i} p_j\right) \cdot \overline{x^2}
$$
$$
= \left(1 + \sum_{i=1}^{M} p_i^* \left(-p_i\right)\right) \cdot \overline{x^2}
$$
$$
= (1 - 1) \cdot \overline{x^2} = 0.
$$
(A.17)

Solution for $\delta = \delta_2 = 1 - M$

The solution is as follows:

$$
p_i = p \quad \forall i.
$$
(A.18)

With $\|\mathbf{p}\|^2 = M|p|^2$ it follows:

$$
|p| = \frac{1}{\sqrt{M}}.
$$
(A.19)

Applying this result to (A.11) yields

$$
\left(1 + \sum_{i=1}^{M} \sum_{j \neq i} p_i^* p_j\right) \cdot \overline{x^2} = (1 + M(M-1)pp^*) \cdot \overline{x^2}
$$
$$
= M \cdot \overline{x^2} \geq 0.
$$
(A.20)

The resulting power is for δ_2 larger than for δ_1. Therefore the maximum is obtained for $\delta = \delta_2 = 1 - M$ and $p_i = p \; \forall i$.

A.3. Approximation of phase

With $p_i = p$ and (A.6) we can now approximate the phase. The minimum of (A.6) can be found by

$$
\nabla_{\mathbf{r}} \left\| \sum_{j=1}^{M} p_j^* e^{J(2\pi f/c)\|\mathbf{q}_i - \mathbf{r}_j\|} - e^{J(2\pi f/c)\|\mathbf{q}_i - \mathbf{r}\|} \sum_{j=1}^{M} p_j^* \right\|^2
$$
$$
= \nabla_{\mathbf{r}} \left\| p^* \left(\sum_{j=1}^{M} e^{J\varphi_{ji}} - M e^{J\overline{\varphi}_i} \right) \right\|^2
$$
$$
= \nabla_{\mathbf{r}} \underbrace{p p^*}_{=1} \underbrace{\left(\sum_{j=1}^{M} e^{J\varphi_{ji}} \sum_{k=1}^{M} e^{-J\varphi_{ki}} \right)}_{=0}
$$
$$
+ \nabla_{\mathbf{r}} \underbrace{p p^*}_{=1} \underbrace{\left(M^2 e^{J\overline{\varphi}_i} e^{-J\overline{\varphi}_i} \right)}_{=0}
$$
$$
- \nabla_{\mathbf{r}} \left(\sum_{j=1}^{M} e^{J\varphi_{ji}} e^{-J\overline{\varphi}_i} + \sum_{j=1}^{M} e^{-J\varphi_{ji}} e^{J\overline{\varphi}_i} \right)
$$
$$
= J \frac{2\pi f}{c} \sum_{j=1}^{M} \left(e^{J(\varphi_{ji} - \overline{\varphi}_i)} - e^{-J(\varphi_{ji} - \overline{\varphi}_i)} \right) \nabla_{\mathbf{r}} \|\mathbf{q}_i - \mathbf{r}\|
$$
$$
= -\frac{4\pi f}{c} \sum_{j=1}^{M} \sin\left(\varphi_{ji} - \overline{\varphi}_i\right) \frac{\mathbf{q}_i - \mathbf{r}}{\sqrt{\|\mathbf{q}_i - \mathbf{r}\|}}
$$
$$
= 0.
$$
(A.21)

Since this equation must be true for all i, $\mathbf{p} = \mathbf{q}_i$ is not a solution. Thus we have to solve

$$
\sum_{j=1}^{M} \sin\left(\varphi_{ji} - \overline{\varphi}_i\right) = A \sin\left(\varphi_{i\Sigma} - \overline{\varphi}_i\right) = 0 \quad \forall i.
$$
(A.22)

This equation is a sum of sine functions with identical frequencies and can therefore be expressed as one sine function with amplitude A and phase $\varphi_{i\Sigma}$. The parameters A and $\varphi_{i\Sigma}$ can be determined by a vector diagram as shown in Figure 10. Each sine wave on the left hand of (A.22) is represented by a vector with amplitude 1 and angle φ_{ji}. The amplitude and angle of the sum of these vectors give A and $\varphi_{i\Sigma}$, respectively. Then $\overline{\varphi}_i$ is given by

$$
\overline{\varphi}_i = \varphi_{i\Sigma} + k\pi, \quad k \in \mathbb{N},
$$
(A.23)

which implies

$$
\|\mathbf{q}_i - \mathbf{r}\| = \frac{c}{f} \left(\frac{1}{2\pi f} \varphi_{i\Sigma} + k \right) \quad \forall i,
$$
(A.24)

and can be interpreted as spheres with the source locations $\mathbf{q}_i$ as their centers. The intersection of all the spheres is the solution of $\mathbf{r}$.

As a special case let us consider a linear array with equally spaced sensors, that is,

$$
\mathbf{u} := \mathbf{r}_j - \mathbf{r}_{j+1}, \quad \forall j \in [1; M-1] \subset \mathbb{N}.
$$
(A.25)

With the far-field assumption and $u := (2\pi f/c)\|\mathbf{u}\| \cos \theta_i$ we obtain

$$
\varphi_{Mi} = \frac{2\pi f}{c} \|\mathbf{q}_i - \mathbf{r}_M\|,
$$
$$
\varphi_{ji} = \frac{2\pi f}{c} \|\mathbf{q}_i - \mathbf{r}_j\|
$$
$$
= \frac{2\pi f}{c} \left(\|\mathbf{q}_i - \mathbf{r}_M\| + (M-j) \cdot \|\mathbf{u}\| \cos \theta_i \right)
$$
$$
= \varphi_{Mi} + (M-j) \cdot u.
$$
(A.26)

Then $\overline{\varphi}_i$ turns out to be

$$
\overline{\varphi}_i = \frac{1}{M} \sum_{j=1}^{M} \varphi_{ji} + k\pi, \quad k \in \mathbb{N}.
$$
(A.27)

If we limit the possible solutions to the line spanned by the linear array, (A.27) goes along with

$$
\mathbf{r} = \frac{1}{M} \sum_{j=1}^{M} \mathbf{r}_j,
$$
(A.28)

which is the center of the sensor array.

REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.

[2] M. Joho, H. Mathis, and R. H. Lambert, "Overdetermined blind source separation: using more sensors than source signals in a noisy mixture," in *Proceedings of 2nd International Conference on Independent Component Analysis and Blind*

Signal Separation (ICA '00), pp. 81–86, Helsinki, Finland, June 2000.

[3] A. Westner and V. M. Bove Jr., "Blind separation of real world audio signals using overdetermined mixtures," in *Proceedinds of 1st International Conference on Independent Component Analysis and Blind Signal Separation (ICA '99)*, Aussois, France, January 1999.

[4] A. Koutras, E. Dermatas, and G. Kokkinakis, "Improving simultaneous speech recognition in real room environments using overdetermined blind source separation," in *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech '01)*, pp. 1009–1012, Aalborg, Denmark, September 2001.

[5] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 3, pp. 204–215, 2003.

[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 *Proceedings of 12th IEEE International Workshop on Neural Networks for Signal Processing (NNSP '02)*, pp. 465–474, Martigny, Switzerland, September 2002.

[7] S. Winter, H. Sawada, and S. Makino, "Geometrical understanding of the PCA subspace method for overdetermined blind source separation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03)*, vol. 2, pp. 769–772, Hong Kong, April 2003.

[8] 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., pp. 299–327, Springer, Berlin, Germany, 2005.

[9] 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, 2000.

[10] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proceedinds of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 2, pp. 1041–1044, Istanbul, Turkey, June 2000.

[11] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, 2004.

[12] A. Hyvärinen, J. Särelä, and R. Vigário, "Bumps and spikes: artifacts generated by independent component analysis with insufficient sample size," in *Proceedings of 1st International Workshop on Independent Component Analysis and Blind Signal Separation (ICA '99)*, pp. 425–429, Aussois, France, January 1999.

[13] 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.

[14] S. U. Pillai, *Array Signal Processing*, Springer, New York, NY, USA, 1989.

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

[16] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig, *Taschenbuch der Mathematik*, Harri Deutsch, Frankfurt am Main, Germany, 3rd edition, 1997.

S. Winter received the Dipl.-Ing. degree in electrical engineering from the University of Erlangen-Nuremberg, Germany, in 2002. In 2001, he was an Intern at Siemens Medical Solutions, Danvers, Mass, where he worked in the Algorithm Development Division. In 2002 he researched for his Dipl.-Ing. thesis at the Communication Science Laboratories, Research and Development Division of Nippon Telegraph and Telephone Corporation (NTT), Kyoto, Japan. His topic included subspace techniques for overdetermined blind source separation of audio signals. He continued researching there in 2003 while being on leave from the Department of Multimedia Communication and Signal Processing, University of Erlangen-Nuremberg. His current research interests include multichannel adaptive algorithms and their application to underdetermined blind source separation of speech signals.

H. Sawada received the B.E., M.E., and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993, and 2001, respectively. In 1993, he joined NTT Communication Science Laboratories. From 1993 to 2000, he was engaged in research on the computer-aided design of digital systems, logic synthesis, and computer architecture. Since 2000, he has been engaged in research on array signal processing, blind source separation for convolutive mixtures, and speech enhancement. Currently, he is a Senior Research Scientist at the NTT Communication Science Laboratories. He received the Best Paper Award of the IEEE Circuit and System Society in 2000. He is a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE), and a Member of the Institute of Electronics, Information and Communication Engineers (IEICE), and the Acoustical Society of Japan (ASJ).

S. Makino received the B.E., M.E., and Ph.D. degrees from Tohoku University, Japan, in 1979, 1981, and 1993, respectively. He joined NTT in 1981. He is now an Executive Manager at the NTT Communication Science Laboratories. He is also a Guest Professor at the Hokkaido University. His research interests include blind source separation of convolutive mixtures of speech, adaptive filtering technologies, and realization of acoustic echo cancellation. He is the author or coauthor of more than 200 articles in journals and conference proceedings and has been responsible for more than 150 patents. He is a member of the Conference Board of the IEEE SP Society and an Associate Editor of the IEEE Transactions on Speech and Audio Processing. He is also an Associate Editor of the EURASIP Journal on Applied Signal Processing. He is a member of the Technical Committee on Audio and Electroacoustics of the IEEE SP Society as well as the Technical Committee on Blind Signal Processing of the IEEE CAS Society. He is also the Organizing Chair of the ICA2003 in Nara, and the General Chair of the IWAENC2003 in Kyoto. He is an IEEE Fellow, a Council Member of the ASJ, and a Member of the IEICE.