University Erlangen-Nuremberg
Telecommunications Laboratory
Multimedia Communications and Signal Processing

Diploma Thesis

Blind Source Separation for Convolutive Mixtures of Speech Signals

Stefan Winter

Advisors: Prof. Dr.-Ing. Walter Kellermann
Dr. Hiroshi Sawada, NTT Corporation
Dr. Shoji Makino, NTT Corporation

Begin: 17.06.2002
End: 17.12.2002

LEHRSTUHL FÜR MULTIMEDIAKOMMUNIKATION UND SIGNALVERARBEITUNG

UNIVERSITÄT ERLANGEN-NÜRNBERG

Professor Dr.-Ing. W. Kellermann

Diploma Thesis

for

Mr. cand.ing. Stefan Winter

Blind Source Separation for Convolutive Mixtures of Speech Signals

Blind source separation (BSS) is a technique to estimate original source signals using only sensor observations that are mixtures of the original signals. If the source signals are mutually independent and non-Gaussian (or non-stationary), techniques of independent component analysis (ICA) can be applied to solve a BSS problem.

In reverberant acoustic environments the BSS solution forms spatial nulls to the directions of jammer signals. In such a situation, a low frequency generally prefers long spacing and a high frequency prefers short spacing of the sensors. Based on this observation, a BSS system with three sensors for two sources seems attractive. In the system, the greater spacing formed by the outer two sensors is used for low frequencies, and a pair including the inner sensor is used for high frequencies. This leads to better results than the conventional system with two sensors.

However, better result might be obtained if three sensors are used for each frequency range. It corresponds to a 3-input 2-output ICA model, which can be solved by a non-holonomic or deflation algorithm.

The main goal of this thesis project is to construct the BSS system using a 3-input 2-output ICA model, and evaluate it. If the 3-2 ICA algorithm is intelligent enough, it might automatically select the outer two sensors for low frequencies to produce a good result. If not, we need an additional control or criteria to lead the algorithm to a good solution. Special attention must be paid to a clear and concise description of the developed algorithms and to its evaluation.

Begin : 17.06.2002

End : 17.12.2002

Advisors:
Prof. Dr.-Ing. W. Kellermann
Dr. Hiroshi Sawada, Nippon Telephone & Telegraph

(Prof. Dr.-Ing. W. Kellermann)

Erklärung:

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Kyoto, Japan, den 17.12.2002

Stefan Winter
Eisfelder Str. 72
96465 Neustadt
Germany

Acknowledgments

"So to the King [Jesus Christ] - eternal, imperishable and invisible, the only God there is - let there be the honor and glory for ever and ever! Amen." (The Bible, 1 Timothy 1:17)

I am very grateful for the unique opportunity that was given to me by Dr. Shoji Makino to write this thesis at NTT Corporation in Japan. I am truly indebted to him and Dr. Hiroshi Sawada, my supervisors, who guided me vigilantly and took care of my progress.

I want to especially thank Prof. Dr. Kellermann, my supervisior at the University in Erlangen, who first suggested writing this thesis in Japan and supported generously this idea.

I express my gratitude to my other colleagues at this excellent research facility for their contribution to the completion of this thesis and to a highly pleasant stay in Kyoto. In particular I want to give thanks to Ms. Shoko Araki, Mr. Ryo Mukai and Dr. Masato Miyoshi for their support. I do not want to forget Mrs. Tomomi Shima and Ms. Sachiko Tsumori who kindly helped me to overcome the obstacles of the Japanese language.

I appreciate the helpful discussions with Mr. Mirko Knaak during his stay at NTT as an exchange researcher from the Technische Univerität Berlin, even when we were on our trips through Japan.

Contents

Zusammenfassung

Blinde Quellentrennung beschäftigt sich mit der Rekonstruktion von Signalen, von denen nur Signalgemische an Sensoren beobachtet werden koennen, ohne dass weitere Informationen über den Mischprozess bekannt sind. Die vorliegende Arbeit behandelt insbesondere die Trennung von gefalteten Sprachsignalgemischen mit Hilfe von Independent Component Analysis (ICA).

Im ersten Teil der Arbeit wird beschrieben, wie ungefaltete Signalgemische getrennt werden können, wenn die ursprünglichen Signale statistisch unabhängig und nicht normalverteilt sind. Mit dem FastICA Algorithmus wird dabei ein praktisches Verfahren erläutert. Anschliessend wird aufgezeigt, wie das Problem der blinden Quellentrennung von gefalteten Signalgemischen beim Wechsel vom Zeit- in den Frequenzbereich auf ungefaltete Gemische zurueckgeführt werden kann.

Der Schwerpunkt der Arbeit liegt auf überbestimmter blinder Quellentrennung, bei der die Anzahl der Sensoren die Anzahl der Signalquellen uebersteigt. Es zeigt sich, dass dieses Problem mit Hilfe von kritisch bestimmter blinder Quellentrennung gelöst werden kann, bei der die Anzahl von Signalquellen und Sensoren übereinstimmt, wenn ein Unterraum des Signalvektorraums betrachtet wird. Für den verwendeten ICA Algorithmus erwies es sich als vorteilhaft, wenn der Unterraum vor der eigentlichen Signaltrennung ausgewählt wird. Zwei Ansätze zur Bestimmung des Unterraums, basierend auf Principal Component Analysis bzw. geometrischen Überlegungen, wurden eingehend miteinander verglichen.

Durch experimentelle und analytische Untersuchungen konnte die Erkenntnis gewonnen werden, dass bei tiefen Frequenzen beide Ansätze ein ähnliches Verhalten aufweisen. Dies lässt sich darauf zurückführen, dass die PCA-basierte Methode in diesem Fall automatisch die Sensoren mit dem grössten Abstand auswählt, wie es auch aufgrund der geometrischen Überlegungen sinnvoll erscheint. Bei hohen Frequenzen ist der PCA-basierte Ansatz im Vorteil, weil er aufgrund der passenden Phasendifferenz zwischen den Sensoren alle Sensorpaare einsetzen kann, um Rauschen zu unterdrücken. Ohne zusätzliches Rauschen zeigen beide Ansätze auch bei höheren Frequenzen eine vergleichbare Trennleistung, da der Vorteil der Rauschunterdrückung nicht zum Tragen kommt.

Die Ergebnisse vertiefen das Verständnis des PCA-basierten Ansatzes aus einer geometrischen Perspektive.

Abstract

Blind source separation (BSS) addresses the problem of estimating original source signals using only sensor observations that are mixtures of the original signals. This thesis deals in particular with the problem of BSS for convolutive mixtures of speech signals by applying techniques of independent component analysis (ICA).

Initially, a description of the method of estimating the source signals from their instantaneous mixtures is given. Here, we assume that the original source signals are mutually independent and have a non-Gaussian probablility density function. In particular we consider a fixed point ICA algorithm that is based on higher order statistics and maximizes non-Gaussianity. Following this, by switching from the the time-domain to the frequency-domain, we show that the problem of BSS of convolutive mixtures can be reduced to that of BSS of instantaneous mixtures.

The major contribution of this thesis is to the problem of overdetermined BSS where the number of sensors exceeds the number of source signals. As it turns out, by employing a subspace processing stage, this problem can be narrowed down to critically-determined BSS where the number of sensors and sources are equal. We found that for the utilized FastICA algorithm it is most advantageous if we employ the subspace processing before the separation of the mixtures. We present experimental and analytical results from the comparison of two previously proposed subspace approaches. One is based on principal component analysis (PCA), the other one is based on geometrical considerations.

Our results show that, for low frequencies, the PCA-based method exhibits a similar behavior to that of the geometry-based method. This is a result of the PCA-based approach automatically emphasising the outer sensors with larger spacing as suggested by the geometry-based approach. For high frequencies, the PCA-based approach performs better when exposed to noisy speech mixtures. This is because in contrast to the geometry-based approach it can utilize all pairs of sensors for high frequencies to suppress the noise. Without the addition of noise both approaches perform similarly as the noise suppression advantage of the PCA based method has no effect.

This thesis deepens our understanding of the PCA-based method from a gemetrical point of view.

Abbreviations

ASJ	Acoustical Society of Japan
BSS	Blind Source Separation
dB	decibel
DOA	Direction of Arrival
ICA	Independent Component Analysis
kurt	kurtosis
LSE	Least Square Error
MUSIC	Multiple Signal Classification
PCA	Principal Component Analysis
RWCP	Real World Computing Partnership
SNIR	Signal-to-Noise plus Interference Ratio
SNR	Signal-to-Noise Ratio
STDFT	Short-Time Discrete Fourier Transform
STFT	Short-Time Fourier Transform

Symbols

$*$	Convolution
$\cdot^*$	Complex conjugate
$\cdot^H$	Conjugate transpose
$\lvert \cdot \rvert$	Absolute value
$\lvert\lvert \cdot \rvert\rvert$	Euclidean norm
Exp1 := Exp2	Exp1 is defined by Exp2
α	Weighting factor
β_{ji}	Phase factors
γ	Lagrange multiplier
δ	Modified Lagrange multiplier
ϵ	Threshold for FastICA
Λ, $\tilde{\Lambda}$	Diagonal matrix with eigenvalues
λ_i, $\tilde{\lambda}_i$	Eigenvalues
∇	Nabla operator
θ_{ji}	Direction of arrival of source i with regard to sensor j
θ_i	Approximated direction of arrival of source i
μ	Discrete frequency parameter
ρ_{jj}, $\tilde{\rho}_{jj}$	Scaling factors
σ_n	Noise variance
ω	Phase factors
a_j	Real part of p_j
b_j	Complex part of p_j
$\mathbb{C}$	Complex numbers
c	Sound velocity
c_{ji}	Absolute value $H_{ji}(f)$
c_i	Approximated value of c_{ji}
D	Block overlap for STDFT
d_i	Distance between sensor number $i-1$ and i ($d_0 = 0$)
$E\{\cdot\}$	Expectation value
$\mathbf{E}$	Matrix with eigenvectors
e_{li}	Energy contribution of source number i to output number l
f	Continuous frequency parameter
$f(\cdot)$	Probability density function
$G(\cdot)$	Nonlinear function

$g(\cdot)$	Derivative of $G(\cdot)$
$\mathbf{H}$	Instantaneous mixing matrix
$\mathbf{H}(f)$	Mixing matrix in frequency domain
$\mathbf{H}^+$	Pseudoinverse of mixing matrix $\mathbf{H}$
$H_{ji}(f)$	Element of $\mathbf{H}(f)$
$\mathbf{h}j$	Columns of $\mathbf{H}\mathbf{W}_{permu}^{-1}$
h_{ji}	Attenuation from source i to sensor j
$h_{ji}(t)$	Room impulse response from source i to sensor j
$\mathbf{I}$	Identity matrix
i	Index
j	Index, imaginary unit
J_G	Cost function (measure for non-Gaussianity)
K	Relevant window length of windowing function
L	Framesize of STDFT
l	Index (summing up distances)
M	Number of sensors / BSS inputs
m	Discrete time parameter of STDFT
N	Number of sources / BSS outputs
$\mathbf{n}(t), \tilde{\mathbf{n}}$	Noise vectors
$\mathbf{p}$	Eigenvector corresponding to a principal component
p_j	j-th element of $\mathbf{p}$
$\mathbf{R}_{\mathbf{XX}}$	Covariance matrix of mixed signals (in general with noise)
$\mathbf{R}_{\tilde{\mathbf{X}}\tilde{\mathbf{X}}}$	Covariance matrix of mixed signals (without noise)
$\mathbf{R}_{\mathbf{SS}}$	Covariance matrix of source signals
$\mathbf{R}_{\mathbf{NN}}$	Covariance matrix of noise
r	Parameter of $G(\cdot)$
S_i	i-th source signal in frequency domain
$\mathbf{S}$	Vector with source signals S_i in frequency domain
s_i	i-th source signal in time domain
$\mathbf{s}$	Vector with source signals s_i in time domain
T	Number of samples
t, τ	Continuous time parameters
$\mathbf{W}$	Unmixing matrix
$\mathbf{W}_{ICA}$	Separation matrix
$\mathbf{W}_{PCA}$	$\mathbf{W}_{PCA} := \mathbf{W}_{white}\mathbf{W}_{sub}^{PCA}$
$\mathbf{W}_{permu}$	Permutation matrix
$\mathbf{W}_{rescale}$	Rescaling matrix

$\mathbf{W}_{scale}$	Scaling matrix
$\mathbf{W}_{sep}$	Separation matrix
$\mathbf{W}_{sub}$	Subspace processing matrix
$\mathbf{W}_{sub}^{geo}$	Geometry-based subspace processing matrix
$\mathbf{W}_{sub}^{PCA}$	PCA-based subspace processing matrix
$\mathbf{W}_{white}$	Whitening matrix
$\mathbf{w}_{+,i}$	Row vector of separation matrix $\mathbf{W}_{ICA}$
$w(\tau)$, $w[k]$	Windowing function
X_j	j-th mixed signal in frequency domain
$\mathbf{X}$	Vector with mixed signals X_j in frequency domain
x_j	j-th mixed signal in time domain
$\mathbf{x}$	Vector with mixed signals x_j in time domain
y_i	i-th output signal in time domain
y_{li}	Part of output l that comes from source i
$\mathbf{y}$	Vector with output signals y_i in time domain
z_j	j-th whitened signal in time domain
$\mathbf{z}$	Vector with whitened signals z_j in time domain

Chapter 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. The term 'blind' refers to the fact that there is no other information available besides the observations itself. In particular, the sources and the mixing system is unknown. If source signals are mutually independent and non-Gaussian (or non-stationary), we can employ independent component analysis (ICA) to solve a BSS problem.

BSS is a versatile tool in a variety of situations for the reason that only few but nethertheless realistic assumptions are made. The applications reach from the separation of biomedical signals over financial market analysis and image processing to telecommunication. They all have in common that the original sources reveal more information than mixtures of original sources for example about a patient's condition or the contents of an image. However, often only mixtures of hidden sources can be obtained, for example electroencephalograms (EEG) or mixtures of communication channels [10].

Another huge appliation field for BSS is audio signal processing. We encounter the problem of source separation in our daily life where we have to focus our attention to a particular speaker while there is a lot of interference from surrounding people and noise. This is often referred to as the cocktail party problem. In this thesis we concentrate on the separation of mixed speech signals that are recorded in a reverberant acoustic environment.

In a reverberant environment we have to account for the convolutive nature of the mixtures. This can be done by switching to the frequency domain where the convolutive BSS problem reduces to an instantaneous BSS problem. Thus, in the first section

of chapter 2 we explain the basic principles of BSS for instantaneous mixtures based on ICA. We describe in particular the FastICA algorithm. In Sec. 2.2 we address the issues which must be taken in to account if switch to the frequency domain.

The emphasis of this thesis is placed on overdetermined BSS, where the number of sensors exceeds the number of sources. Although in many cases equal numbers of source signals and sensors are assumed [10], using an overdetermined systems often yields better results [14, 19, 35]. In chapter 3 we extend the basic framework from chapter 2 to overdetermined BSS. In particular we describe two previously proposed subspace approaches that reduce the problem of overdetermined BSS to the problem of critically-determined BSS where the number of sources and sensors is equal. The subspace methods are based on principal component analysis (PCA) and geometric considerations, respectively [5, 29].

In chapter 4 we compare both subspace methods more thoroughly in particular for 2 sources and 3 sensors. The presented experimental and analytical results explain why the PCA-based approach automatically selects the outer two sensors for low frequencies. This leads to a similar behavior of the PCA- and geometry-based approaches and provides a better understanding of the PCA-based approach..

Finally, in chapter 5, we present the results from overdetermined BSS based on the two subspace approaches, that compare the separation performance using real world data in a reverberant environment. We explain why the PCA-based approach yields better results than the geometry-based approach if exposed to noisy speech mixtures.

Chapter 2

General framework of BSS

In this chapter we explain the basic principles of BSS based on ICA. Though our goal is to separate convolutive speech mixtures, we begin with the separation of instantaneous, complex mixtures and explain it in detail in Sec. 2.1. As we show in Sec. 2.2, the separation of instantaneous mixtures gives the basic module for separating convolutive mixtures if we approach the problem in the frequency domain.

2.1 BSS for instantaneous mixtures

Instantaneous mixtures represent the most simple case in BSS. After describing the mixing model and the respective assumptions, we explain the process of separating the mixtures, paying special attention to the FastICA algorithm proposed by Hyvärinen et al. [10].

2.1.1 Underlying model

We consider a linear mixing model with N discrete source signals $s_i(t) \in \mathbb{C},\ (1 \leq i \leq N)$ represented in the N-dimensional complex vector space by

$$\mathbf{s}(t) := \begin{bmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix} \in \mathbb{C}^N \tag{2.1}$$

where t denotes the continuous time parameter. In the context of ICA we consider each source signal $s_i(t)$ as a continuous random variable s_i. Depending on the particular ICA algorithm, different assumptions with respect to their stochastic properties

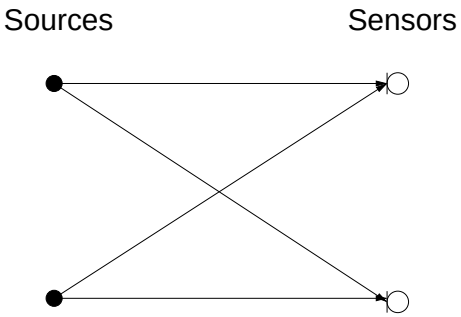

Figure 2.1: Basic mixing model for instantaneous mixtures

are made. Inherent in all algorithms is the assumption that the source signals are statistically independent, i.e.

$$f(s_1, \ldots, s_i, \ldots, s_N) = f(s_1) \cdot \ldots \cdot f(s_i) \cdot \ldots \cdot f(s_N) \tag{2.2}$$

where $f(s_i)$ denotes the probability density function (pdf) of signal s_i [25]. Since source signals with a Gaussian distribution are statistically independent, even if they are mixed, at least one more assumption is made. Most ICA algorithms either assume non-Gaussian pdfs or non-stationarity. The algorithm we employed is based on the assumption of non-Gaussian pdfs. Since we deal with audio signals, we can narrow them down to super-Gaussian pdfs [20]. This means that the kurtosis $\mathrm{kurt}(s_i)$ defined for the zero-mean random variable s_i by

$$\mathrm{kurt}(s_i) = E\{s_i^4\} - 3\left(E\{s_i^2\}\right)^2 \tag{2.3}$$

is positive. $E\{\cdot\}$ denotes the expectation value.

Besides the source signals $\mathbf{s}(t)$ we consider M linearly aligned sensors, at which we can observe M mixed signals $x_j(t) \in \mathbb{C},\ (1 \leq j \leq M)$ represented in the M-dimensional complex vector space by

$$\mathbf{x}(t) := \begin{bmatrix} x_1(t) \\ \vdots \\ x_M(t) \end{bmatrix} \in \mathbb{C}^M \tag{2.4}$$

We can now describe the mixing process by a linear mapping

$$\mathbf{H} : \mathbb{C}^N \to \mathbb{C}^M \tag{2.5}$$

where $\mathbf{H}$ is given by a complex valued $M \times N$ mixing matrix

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1N} \\ \vdots & \ddots & \vdots \\ h_{M1} & \cdots & h_{MN} \end{bmatrix} \in \mathbb{C}^{M \times N} \tag{2.6}$$

This yields

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) \tag{2.7}$$

We assume that $\mathbf{H}$ has full rank. The case that $\mathbf{H}$ does not have full rank can be dealt with by underdetermined BSS. Each element $h_{ji} \in \mathbb{C}$ of the mixing matrix $\mathbf{H}$ can be considered as attenuation factor of the signal $s_i(t)$ when it is recorded at sensor number j.

Depending on the relationship between the number of sensors M and the number of sources N, we can distinguish three different types of BSS systems. Following the notation in [21] we have

- $M < N$: underdetermined BSS
  This type is not be considered in the present thesis. More information on this topic can be found e.g. in [34].
- $M = N$: critically-determined BSS
- $M > N$: overdetermined BSS; sometimes, if coming from a mathematical point of view, it is referred to as the undercomplete basis problem [14]

2.1.2 Separating mixed signals

To obtain the original source signals $\mathbf{s}(t)$ from the observed signals $\mathbf{x}(t)$ we want to find an inverse linear mapping

$$\mathbf{W} : \mathbb{C}^M \rightarrow \mathbb{C}^N \tag{2.8}$$

that inverts the mixing process and separates the mixed signals $\mathbf{x}(t)$. $\mathbf{W}$ can be described by a $N \times M$ unmixing matrix which yields

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{2.9}$$

where

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_N(t) \end{bmatrix} \in \mathbb{C}^N \tag{2.10}$$

denotes the separated output signals. Ideally the unmixing matrix $\mathbf{W}$ would fulfill

$$\mathbf{W} \cdot \mathbf{H} = \mathbf{I} \tag{2.11}$$

where $\mathbf{I}$ denotes the unity matrix. Due to ambiguities that are addressed in more detail in Sec. 2.2.3, we allow for scaled versions of the original signals with a different order. We note that even in the ideal case the unmixing matrix is not necessarily the pseudoinverse. By definition, we do not know the mixing matrix $\mathbf{H}$, i.e. we cannot invert the mixing matrix directly. Therefore we need other approaches that exploit specific characteristics of the signals to separate them.

One of these approaches is ICA, which is based on the assumption made in Sec. 2.1.1 that the original sources are statistically independent. Mixing several independent, non-Gaussian source signals $\mathbf{s}(t)$ in a way described by Eq. (2.7) means generating mixed signals $\mathbf{x}(t)$ that are statistically dependent. ICA builds on this fact by unmixing the observed signals $\mathbf{x}(t)$ in a way that gives statistically independent output signals $\mathbf{y}(t)$. It can be shown that this reconstructs the original source signals [10].

In the case of statistically independent source signals with Gaussian pdfs, even mixtures of these signals would be statistically independent [25]. Thus we would not be able to separate them only based on statistical independence. This leads to the assumption that our source signals $\mathbf{s}(t)$ have non-Gaussian pdfs, which is true for speech signals. According to the central limit theorem, the sum of independent, equally distributed random variables converges to a Gaussian distribution even if the initial pdf is non-Gaussian. This implies that the pdf of a mixture of statistical independent source signals is closer to a Gaussian distribution than the original pdfs. Under the given conditions, non-Gaussianity and statistical independence are to a certain extent interchangeable. This fact can be exploited by searching for an unmixing matrix $\mathbf{W}$ that maximizes non-Gaussianity of the separated signals to find the independent components. The maximization must be done with the constraint that the variance of the separated signal remains constant. The algorithm we implemented is built on this principle and described in the following.

As we show in chapter 3 we can reduce overdetermined BSS to critically-determined BSS. Thus we restrict ourselves in the remainder of the present chapter to critically-determined BSS with $M = N$. Since we assumed that the mixing matrix $\mathbf{H}$ has full rank it is non-singular and therefore invertible. The case of $\mathbf{H}$ being singular can be dealt with by underdetermined BSS.

Whitening

The ICA algorithm we employed is derived using the assumption that the mixed signals have unit variance and are uncorrelated [6]. If the mixed signals do not comply with this assumption we can employ a preprocessing step that normalizes and uncorrelates them. This is often called whitening or sphering [10]:140. It is a linear operation and can be described by a whitening matrix $\mathbf{W}_{white}$. To distinguish the unprocessed mixed signals $\mathbf{x}(t)$ observed at the sensors from the whitened ones, we denote the latter by $\mathbf{z}(t)$. Then we can write

$$\mathbf{z}(t) = \mathbf{W}_{white}\mathbf{x}(t) \tag{2.12}$$

with

$$E\{\mathbf{z}\mathbf{z}^H\} = \mathbf{I} \tag{2.13}$$

due to the normalized and uncorrelated nature of $\mathbf{z}$. $\cdot^H$ denotes the hermitian operator. The problem of decorrelation can be addressed by e.g. principal component analysis (PCA). It is closely related to ICA, since making signals independent includes making them uncorrelated. Following the idea of PCA, we determine the symmetric correlation matrix $\mathbf{R}_{\mathbf{xx}}$ of the mixed signals $\mathbf{x}$ and do the eigenvalue decomposition as given in Eq. (2.14).

$$\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{E}\Lambda\mathbf{E}^H \tag{2.14}$$

$\mathbf{E}$ denotes the unitary matrix of the eigenvectors of $\mathbf{R}_{\mathbf{xx}}$ and Λ the diagonal matrix of the corresponding eigenvalues. We can then write the whitening matrix $\mathbf{W}_{white}$ as

$$\mathbf{W}_{white} = \Lambda^{-\frac{1}{2}}\mathbf{E}^H \tag{2.15}$$

which can be easily verified by

$$E\{\mathbf{z}\mathbf{z}^H\} = E\{\Lambda^{-\frac{1}{2}}\mathbf{E}^H\mathbf{x}\mathbf{x}^H\mathbf{E}^H\Lambda^{-\frac{1}{2}}\} = \Lambda^{-\frac{1}{2}}\mathbf{E}^H\mathbf{R}_{\mathbf{xx}}\mathbf{E}\Lambda^{-\frac{1}{2}} = \mathbf{I} \tag{2.16}$$

ICA

After we have uncorrelated signals $\mathbf{z}$ with unit variance, ICA only has to provide an $N \times N$ unitary matrix

$$\mathbf{W}_{ICA} := \begin{bmatrix} \mathbf{w}_{+,1}^H \\ \vdots \\ \mathbf{w}_{+,N}^H \end{bmatrix} \tag{2.17}$$

for rotating the random variables and thereby making them independent. We call the vectors $\mathbf{w}_{+,i}$ unmixing vectors. Unitarity implies that the row or column vectors, respectively, are mutually orthogonal and that

$$\mathbf{W}_{ICA}\mathbf{W}_{ICA}^{H} = \mathbf{I} \tag{2.18}$$

We employed FastICA, as presented in [10], to obtain independent and, therefore, separated signals. Since we deal with complex numbers, we use the complex valued version of FastICA, which is derived in details in [6].

FastICA, in general, is a fixed-point algorithm[1] that is based on maximizing non-Gaussianity, as mentioned earlier. It can also be classified as a deflationary algorithm. This means that it yields the unmixing vectors $\mathbf{w}_{+,i}$ one after another instead of computing them in parallel by determining the unmixing matrix $\mathbf{W}_{ICA}$ as a whole.

As measure for non-Gaussianity to be maximized, the contrast function

$$J_G(\mathbf{w}_{+,i}) = E\left\{G(|\mathbf{w}_{+,i}^{H}\mathbf{z}|^2)\right\} \tag{2.19}$$

is used where $G : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ denotes a smooth even function for real, nonnegative numbers. Together with the constraint

$$E\{|\mathbf{w}_{+,i}^{H}\mathbf{z}|\} = ||\mathbf{w}_{+,i}||^2 = 1 \tag{2.20}$$

where $||\cdot||$ denotes the Euclidean norm, the basic algorithm in Eq. (2.21) can be derived, by which the unmixing vector $\mathbf{w}_{+,i}$ for a single output signal y_i can be gradually improved until the difference between consecutive unmixing vectors falls below a certain threshold ϵ.

$$\mathbf{w}_{+,i} \leftarrow E\left\{\mathbf{z}(\mathbf{w}_{+,i}^{H}\mathbf{z})^* g(|\mathbf{w}_{+,i}^{H}\mathbf{z}|^2\right\} - E\left\{g(|\mathbf{w}_{+,i}^{H}\mathbf{z}|^2) + |\mathbf{w}_{+,i}^{H}\mathbf{z}|^2 g'(|\mathbf{w}_{+,i}^{H}\mathbf{z}|^2)\right\}\mathbf{w}_{+,i} \tag{2.21}$$

$\cdot^*$ denotes the complex conjugate and $g(\cdot)$ the derivative of the nonlinear function $G(\cdot)$, which was here chosen as

$$G(x) = \log(r + x) \tag{2.22}$$

where is an arbitrary parameter that we set to $r = 0.1$ according to [6]. To account for Eq. (2.20) we must normalize the unmixing vector $\mathbf{w}_{+,i}$ each time it gets improved.

$$\mathbf{w}_{+,i} \leftarrow \frac{\mathbf{w}_{+,i}}{||\mathbf{w}_{+,i}||} \tag{2.23}$$

[1]Iterative algorithm of the form $x_{n+1} = f(x_n)$, $(n = 0, 1, 2, ...;\ x_0$ given) for solving a fixed-point equation of the form $x = f(x)$ [8]

The initial value for $\mathbf{w}_{+,i}$ from the point at which the iteration starts can be chosen arbitrarily. When $\mathbf{w}_{+,i}$ has converged, it yields one unmixing vector, corresponding to one independent component. To obtain several unmixing vectors, we repeat this procedure as many times as there are still signals to be separated. Thereby, we must ensure that the unmixing vectors do not converge to the same solution. We know that an unmixing vector must be orthogonal to all other unmixing vectors. Thus, before we normalize an unmixing vector, we first orthogonalize it by the Gram-Schmidt algorithm with respect to already existing unmixing vectors.

$$\mathbf{w}_{+,i+1} \leftarrow \mathbf{w}_{+,i+1} - \sum_{l=1}^{i} \mathbf{w}_{+,l}\mathbf{w}_{+,i+1}^{H}\mathbf{w}_{+,l} \tag{2.24}$$

Its basic idea is to remove all non-orthogonal components with the result that only the orthogonal component remains. After we obtained all desired unmixing vectors $\mathbf{w}_i$ we can summarize them in the separation matrix $\mathbf{W}_{ICA}$.

In conclusion, we arrive at the unmixing matrix $\mathbf{W}$ in a 2-stage process. First we calculate a whitening matrix $\mathbf{W}_{white}$ that yields uncorrelated and normalized signals. Based on these signals, we can compute the actual separation matrix $\mathbf{W}_{ICA}$, which gives

$$\mathbf{W} = \mathbf{W}_{ICA}\mathbf{W}_{white} \tag{2.25}$$

Then we obtain the separated signals $\mathbf{y}(t)$ by applying the unmixing matrix $\mathbf{W}$ to the mixed signals $\mathbf{x}(t)$.

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{2.26}$$

The block diagram of the basic ICA module is given in Fig. 2.2

2.1.3 Influence of noise

Let us consider an instantaneous mixing model where we have uncorrelated, Gaussian noise with equal variance added to the sensors.

$$\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{n}(t), \qquad E\{\mathbf{n}\mathbf{n}^{H}\} = \mathbf{I} \tag{2.27}$$

We assume that the noise is uncorrelated to the source signals $\mathbf{s}(t)$. We can rewrite this noisy mixing model as [10]:294

$$\mathbf{x}(t) = \mathbf{H}\left(\mathbf{s}(t) + \tilde{\mathbf{n}}(t)\right) \tag{2.28}$$

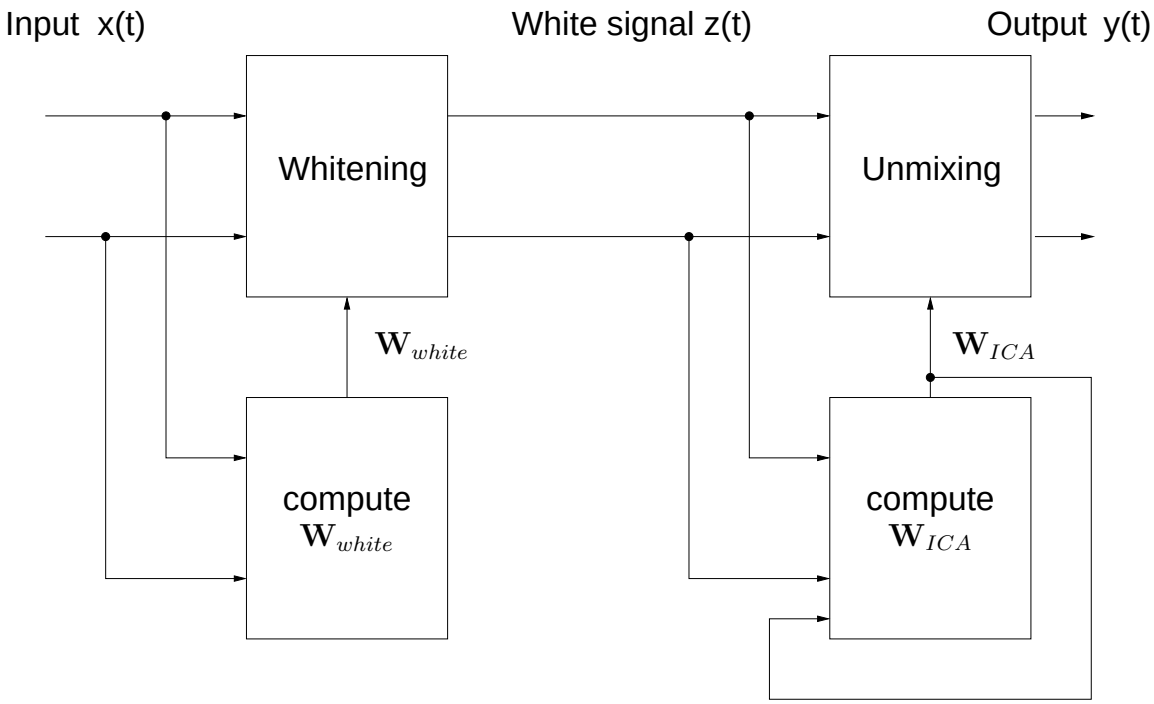

Figure 2.2: Basic module for instantaneous source separation

by defining virtual noise

$$\tilde{\mathbf{n}}(t) := \mathbf{H}^{+}\mathbf{n}(t) \tag{2.29}$$

where $\mathbf{H}^{+}$ denotes the pseudoinverse of $\mathbf{H}$. Now we can further define virtual noisy source signals by

$$\tilde{\mathbf{s}}(t) := \mathbf{s}(t) + \tilde{\mathbf{n}}(t) \tag{2.30}$$

which yields

$$\mathbf{x}(t) = \mathbf{H}\tilde{\mathbf{s}}(t) \tag{2.31}$$

Obviously the virtual source signals $\tilde{\mathbf{s}}(t)$ are still independent and have a non-Gaussian pdf, i.e. they comply with the assumptions made in Sec. 2.1.2 for ICA. The mixing matrix is not changed by the noise. Thus we can still employ ICA to separate the mixed signals $\mathbf{x}(t)$. The mixing matrix and, therefore, the unmixing matrix, are not affected by noise at all. However, while the independent components contain the separated source signals, they are distorted by noise, which becomes clear when we apply the ideal unmixing matrix to the mixed signals.

$$\mathbf{W}\mathbf{x} = \mathbf{W}(\mathbf{H}\mathbf{s}(t) + \mathbf{n}) = \mathbf{s}(t) + \mathbf{W}\mathbf{n} \tag{2.32}$$

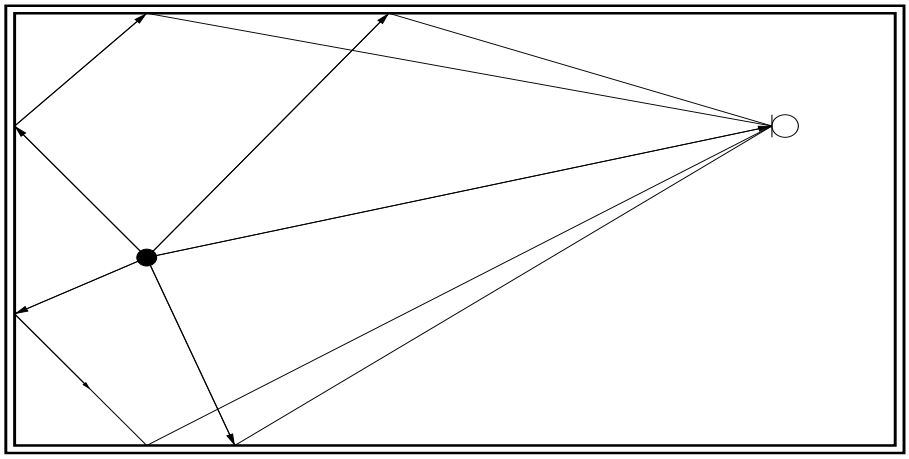

Figure 2.3: Convolutive mixing model

Therefore ICA can be employed to separate noisy mixtures but additional measures are necessary to remove the noise. As we point out in Sec. 3.2, overdetermined BSS combined with an appropriate preprocessing step yields a possibility to achive this noise reduction.

2.2 BSS for convolutive mixtures

In section 2.1 we explained basic principles of BSS based on ICA for instantaneous mixtures. In the present section we enhance the idea of BSS to convolutive mixtures. This type of mixture is common in real world speech applications where reverberation and time-delays cannot be neglected. After describing the underlying model, we show how the separation of convolutive mixtures can be reduced to the separation of instantaneous mixtures. Then we discuss issues that always arise from indeterminancies of the separation process, to which special attention must be paid when considering convolutive mixtures.

2.2.1 Convolutive mixing model

Dealing with speech signals, we consider a reverberative room with time-independent properties, where delayed and attenuated versions of the original source signals arrive at the sensors as indicated in Fig. 2.3. We account for this situation by time-independent room impulse responses that reflect the characteristics of the environment and depend on the position of source and sensor, respectively. Thus, before a

signal from source number j arrives at sensor number i, it is convoluted by an impulse response $h_{ij}(t)$. Using the same notation for source signals $\mathbf{s}(t)$ and mixed signals $\mathbf{x}(t)$ as in Sec. 2.1.1, the output $x_j(t)$ at sensor j can be computed by

$$x_j(t) = \sum_{i=1}^{N} \int_{\tau=-\infty}^{\infty} s_i(\tau) \cdot h_{ji}(t-\tau) = \sum_{i=1}^{N} s_i(t) * h_{ji}(t) \tag{2.33}$$

where $*$ denotes the convolution. This means that the scalar elements of $\mathbf{H}$ in Eq.(2.7) become filters.

If we assume time-limited signals we can switch to the frequency domain by the Fourier transform. Then the convolution turns into a multiplication and we obtain a time-independent mixing matrix $\mathbf{H}(f)$ in the frequency domain, where f denotes the frequency parameter. We further assume a far-field situation where the distance between sources and sensors is very large, compared to the distance between the sensors, which holds for many applications with sensor arrays [7]:3. Then we can approximate the wave fronts arriving at the sensors by plain wave fronts (Fig. 2.4). This leads to a common model in array processing and we can express the frequency-

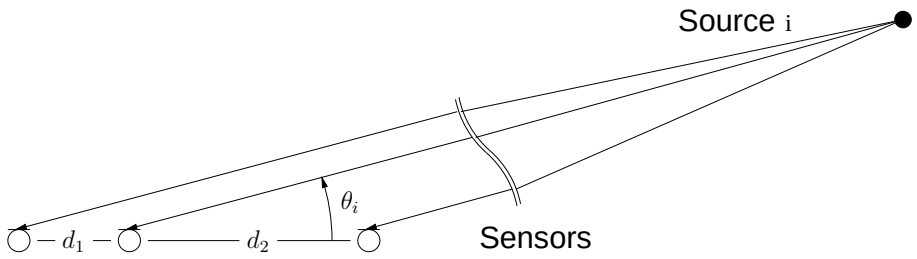

Figure 2.4: Definition of source direction

dependent mixing matrix $\mathbf{H}(f)$ by

$$\mathbf{H}(f) = \begin{bmatrix} H_{11}(f) & \dots & H_{1N}(f) \\ \vdots & \ddots & \vdots \\ H_{M1}(f) & \dots & H_{MN}(f) \end{bmatrix} = \begin{bmatrix} c_{11}e^{j\beta_{11}} & \dots & c_{1N}e^{j\beta_{1N}} \\ \vdots & \ddots & \vdots \\ c_{M1}e^{j\beta_{M1}} & \dots & c_{MN}e^{j\beta_{MN}} \end{bmatrix} \tag{2.34}$$

where c_{ji} accounts for the attenuation and $e^{\beta_{ji}}$ for the phase delay. β_{ji} is given by

$$\beta_{ji} = \frac{2\pi f \cos(\theta_{ji}) \sum_{l=0}^{i-1} d_l}{c} \tag{2.35}$$

where θ_{ji} denotes the direction of arrival (DOA) of source i with regard to sensor j and c the sound velocity. d_l stands for the distance between the linearly aligned and consecutively numbered sensors number $l-1$ and l. d_0 is set to zero, since we

assume that the first sensor serves as a reference point. Due to reverberation, θ_{ji} does not necessarily coincide with the angle that is related to the physical position of the source and might vary depending on the frequency in practice.

2.2.2 Separating signals

For separating convolutively-mixed signals, we can distinguish between two fundamentally different kinds of approaches. One possibility is applying ICA in the time domain [1]. The other possibility, which we employed in this thesis, applies ICA in the frequency domain [31]. It is advantageous in that it allows frequency dependent and, therefore, narrowband algorithms. Therefore, as explained in the following, we can easily employ the same principles as we did for instantaneous mixtures. This is computationally less expensive than time-domain approaches.

Applying the Fourier transform to the observed signals does not change the mixing matrix [10]:365. However, switching to the frequency domain reduces the problem of finding filters with many parameters to finding simple unmixing matrices with scalar elements depending on the frequency. This is similar to solving the ICA problem for instantaneous mixtures, except for the fact that an additional scaling and permutation problem occurs.

The short-time Fourier transform (STFT) is a capable means to obtain time dependent frequency-domain representations $X_j(f,m)$ of the time-domain mixtures $\mathbf{x}(t)$ [16, 24]:

$$X_j(f,t) = \mathrm{STFT}(x(t)) := \int_{\tau=-\infty}^{\infty} x(t+\tau) \cdot w(\tau) e^{-j(2\pi f)\tau} d\tau \tag{2.36}$$

where $w(t)$ represents a windowing function and τ denotes the time shift. The STFT accounts for the short time stationarity of speech signals and provides the necessary time resolution in each frequency bin to estimate the expectation values [18]. In the frequency domain we can replace the convolution in Eq. (2.33) by a multiplication and we obtain

$$X_j(f,t) = \sum_{i=1}^{N} H_{ji}(f) \cdot S_i(f,t) \tag{2.37}$$

where $S_i(f,t)$, similar to $X_j(f,t)$, denotes the frequency-domain time-series representation of the time-domain source signal $s_i(t)$. Summarizing the sources and mixtures in the same way as in Sec. 2.1.1, we obtain

$$\mathbf{X}(f,t) = \mathbf{H}(f)\mathbf{S}(f,t) \tag{2.38}$$

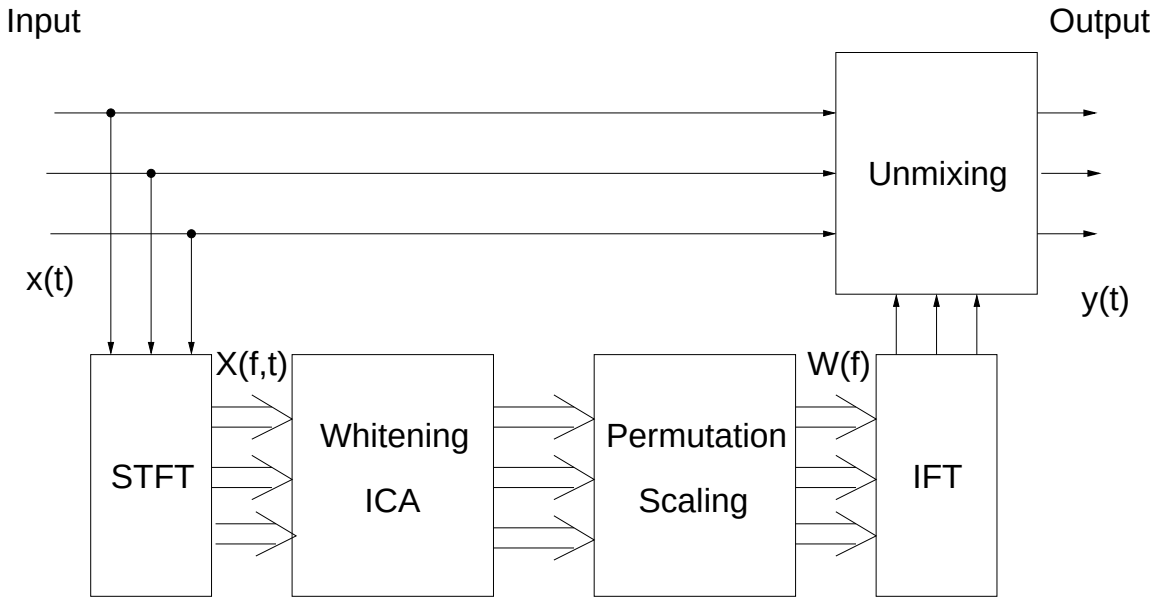

Figure 2.5: Frequency-domain BSS

where $\mathbf{H}(f)$ is the frequency dependent $M \times N$ mixing matrix from Eq. (2.34) composed by the $H_{ji}(f)$. $\mathbf{X}(f,t)$ and $\mathbf{S}(f,t)$ are the STFTs of $\mathbf{x}(t)$ and $\mathbf{s}(t)$ respectively.

Equation (2.38) has basically the same structure as Eq. (2.7), i.e. the problem of convolutive mixtures is hereby transformed to instantaneous mixtures that are frequency dependent. Since the mixing and therefore the unmixing matrix at a specific frequency instance do not depend on the mixing matrix at other frequency instances we can apply the method described in section 2.1 to obtain an unmixing matrix $\mathbf{W}(f)$ for each frequency instance. When we have obtained the frequency dependent unmixing matrix, we must solve the permutation and scaling problem. They are addressed below in more detail. To obtain unmixed time-domain signals, we can calculate impulse responses by the inverse Fourier transform (IFT) from $\mathbf{W}(f)$ and convolute them with the mixed signals. In contrast to separating the signals in the frequency domain this avoids the overhead and difficulties which would arise in practice from inverting the STFT due to the windowing.

The diagram explaining the data flow is shown in Fig. 2.5. In summary, we see that obtaining the unmixing system can be achieved by applying well known techniques of instantaneous BSS in the frequency domain. Thus, for our further considerations, we assume that we are in the frequency domain and apply the described algorithms and ideas in the frequency domain.

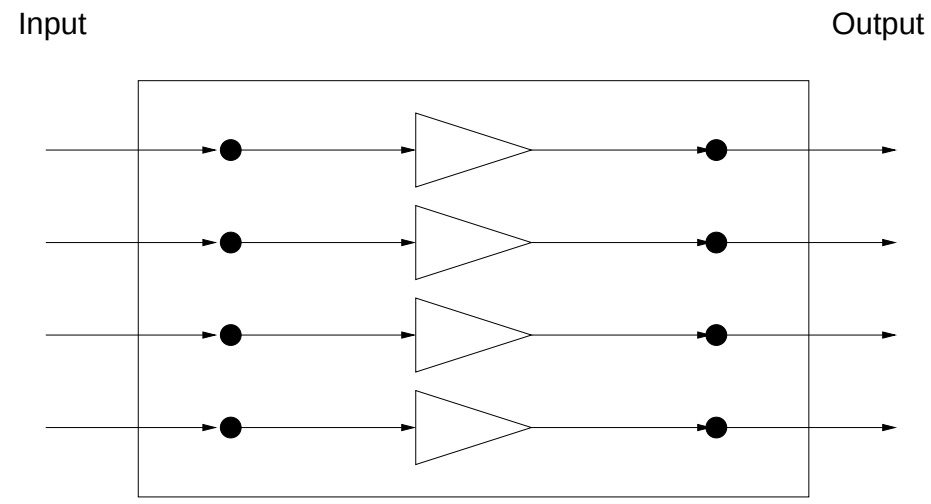

Figure 2.6: Rescaling

2.2.3 Scaling and permutation problem

Two important issues that must be addressed with frequency-domain ICA are the scaling and permutation problem [4, 13, 30, 32]. They arise from the fact that the independence of the separated signals is not affected by a constant, complex valued factor or the order of the signals. Due to these ambiguities in ICA, the order and power of the separated signals might be different from the order and power of the source signals. These problems are also present with instantaneous mixtures, but often do not play an important role. However, they can have serious effects on the performance of separating convolutive mixtures, particularly with frequency-domain ICA. The reasons and possible solutions are discussed in the following. For simplicity's sake we consider the mixing and unmixing matrix in the present section at discrete frequency instances and relate to them by the term 'frequency bins'.

Scaling Problem

In frequency-domain ICA, we compute an unmixing system for every frequency bin. Since each frequency bin can have its own, independent scaling factors, they form the frequency response of a filter. If this ambiguity is not resolved, it results in filtered versions of the original signals, which does not influence their statistical independence and complies with the fact that filtering the signals does not influence the mixing matrix [10]:264. Thus, we must rescale the signals that are separated by ICA (Fig. 2.6)

The method we use to solve the scaling problem is based on the idea of Ikeda et al. [12] and is also utilized for dewhitening in time-domain ICA [1]. We define $\mathbf{W}_{scale}$ as a diagonal matrix with the scaling factors s_{jj} on its diagonal and $\mathbf{W}_{permu}$ as a permutation matrix that has only a single 1 in each column and row. Then we can write the unmixing matrix $\mathbf{W}$ with $\mathbf{W}_{sep} := \mathbf{H}^{-1}$as[2]

$$\mathbf{W} = \mathbf{W}_{scale}\mathbf{W}_{permu}\mathbf{W}_{sep} = \begin{bmatrix} \rho_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \rho_{MM} \end{bmatrix} \mathbf{W}_{permu}\mathbf{H}^{-1} \tag{2.39}$$

Following Ikeda et al. [12] we put the separated signals back to the sensor input with the inverse $\mathbf{W}^{-1}$ of the unmixing matrix. Denoting the columns of $\mathbf{H}\mathbf{W}_{permu}^{-1}$ by $\mathbf{h}_j$ we get

$$\begin{aligned} \mathbf{W}^{-1} &= (\mathbf{W}_{scale}\mathbf{W}_{permu}\mathbf{H}^{-1})^{-1} && (2.40)\\ &= \mathbf{H}\mathbf{W}_{permu}^{-1}\mathbf{W}_{scale}^{-1} && (2.41)\\ &= \mathbf{H}\mathbf{W}_{permu}^{-1}\begin{bmatrix} \rho_{11}^{-1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \rho_{MM}^{-1} \end{bmatrix} && (2.42)\\ &= \begin{bmatrix} \rho_{11}^{-1}\mathbf{h}_1 & \cdots & \rho_{MM}^{-1}\mathbf{h}_M \end{bmatrix} && (2.43) \end{aligned}$$

Defining the columns of the unmixing matrix $\mathbf{W}^{-1}$ as $\mathbf{w}_j^{-1}$ we can write

$$\mathbf{w}_j^{-1} = \rho_{jj}^{-1}\mathbf{h}_j \tag{2.44}$$

Now we define $\tilde{\rho}_{jj}$ as the weighted sum of the elements w_{kj}^{-1} of the column vector $\mathbf{w}_j^{-1}$ and obtain:

$$\tilde{\rho}_{jj} = \sum_{k=1}^{M}\alpha_k w_{kj}^{-1} = \sum_{k=1}^{M}\alpha_k\rho_{jj}^{-1}h_{kj} = \rho_{jj}^{-1}\cdot\underbrace{\sum_{k=1}^{M}\alpha_k h_{kj}}_{:=\bar{h_j}} = \rho_{jj}^{-1}\cdot\bar{h_j} \tag{2.45}$$

Thus, we can multiply the unmixing matrix $\mathbf{W}$ with a diagonal matrix $\mathbf{W}_{rescale}$ given by

$$\mathbf{W}_{rescale} := \begin{bmatrix} \tilde{\rho}_{11}^{1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \tilde{\rho}_{MM}^{1} \end{bmatrix} \tag{2.46}$$

[2]We can always define the scaling matrix $\mathbf{W}_{scale}$ in a way so that the order of $\mathbf{W}_{scale}$ and $\mathbf{W}_{permu}$ as given in Eq. (2.39) is possible

This yields

$$\mathbf{W}_{rescale}\mathbf{W} = \mathbf{W}_{rescale}\mathbf{W}_{scale}\mathbf{W}_{permu}\mathbf{H}^{-1} \quad (2.47)$$

$$= \begin{bmatrix} \bar{h_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \bar{h_M} \end{bmatrix} \mathbf{W}_{permu}\mathbf{H}^{-1} \quad (2.48)$$

As we can see, this removes the arbitrary scaling factor s_{ii}. The disadvantage is that the output is now scaled by $\bar{h_j}$. We can influence this scaling by choosing the weighting factors α_k. For dewhitening used in time-domain ICA, where permutation is not a problem, $\alpha_k = 1$ for $k = j$ and $\alpha_k = 0$ for $k \neq j$ was found to give the best performance. In the case of accounting for the permutation problem, $\alpha_k = \frac{1}{M}$ for $k = 1, \ldots, M$ has been proposed [1]. This choice gives a more balanced and, therefore, robust solution, and was used in our implementation.

The scaling factors are complex. While we do not know the absolute values of the gain factors of the mixing matrix, we can estimate the phase if we know the DOA. By this we are at least able to adjust the phase more correctly. A common method to estimate the DOA is the MUSIC algorithm [33]. If we know the DOA we can estimate the frequency dependent mixing matrix in Eq. (2.34) assuming that $c_i = 1$. We now consider the complete transfer function from the sources to the output given by $\mathbf{WH}$, where $\mathbf{H}$ is the estimated mixing matrix. In the ideal case the phase should be zero for all output signals[3]. If the phase is not zero, we can align it and thereby correct it.

Permutation problem

In the case of instantaneous mixtures the permutation problem is not as important as in the case of convolutive mixtures because the order of the signals is not known anyway. But once we start applying ICA in the frequency domain in each frequency bin, the order of the separated signals plays a significant role. If we do not have the same order in each frequency bin, we do not know which element belongs to which frequency response. Normally ICA itself does not give the necessary information to decide the correct order. Thus, other features of the signals must be exploited to bring the signals into a defined order (Fig. 2.7). However, there exist some approaches of so-called constrained BSS which make it possible to avoid the permutation problem [26]. Other approaches avoid the permutation and scaling problem by employing

[3]This ideal case might lead to non-causal frequency responses $\mathbf{W}$

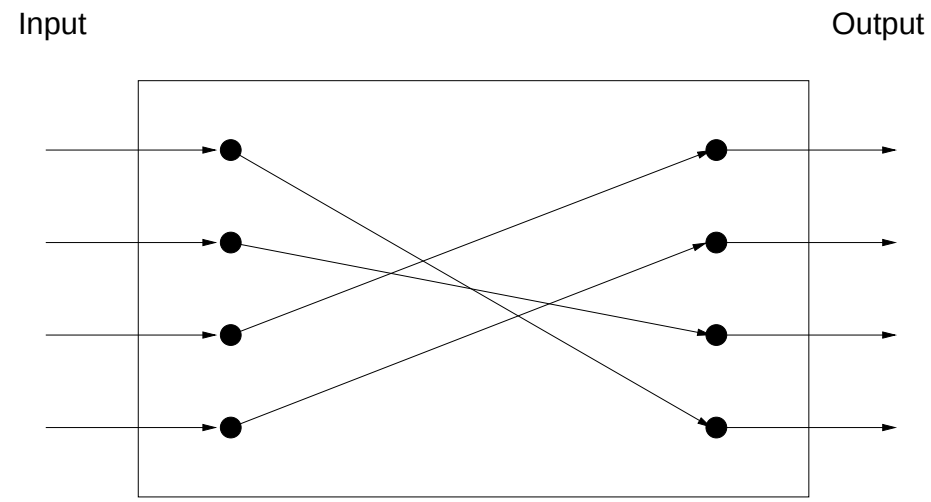

Figure 2.7: Solving permutation

only part of the separation process in the frequency domain [10]:366. These are not covered by this thesis.

Several approaches for solving the permutation problem exist. These have in common that they exploit features that are specific to each original signal and help to relate the unmixing vectors of each frequency bin to a particular signal. These features can be e.g. the direction of arrival (DOA), statistical correlation or the null direction if BSS is considered as a beamformer. While the performance of each single method might not be satisfying and very dependent on the particular situation, it can be improved by combining different approaches to exploit the advantage of each [17]. However, the more signals must be brought into the correct order, the more difficult the permutation problem becomes. A larger number of signals means that the difference between the features becomes smaller, which makes it more difficult to distinguish them. This is very obvious e.g. with the DOA. The more sources we have, the closer their location is and the less the difference of the DOA becomes.

Theoretically optimal solution

For the reason of evalution, we avoided any negative influence of the permutation problem by selecting the optimal permutation and thus aligning the frequency responses as best as possible. We calculated the frequency dependent energy $e_{li}(f)$ in dB of the contribution of each source i to each output l by

$$e_{li}(f) = 10 \log_{10} \left(\int Y_{li}^2(f,t) dt \right) \tag{2.49}$$

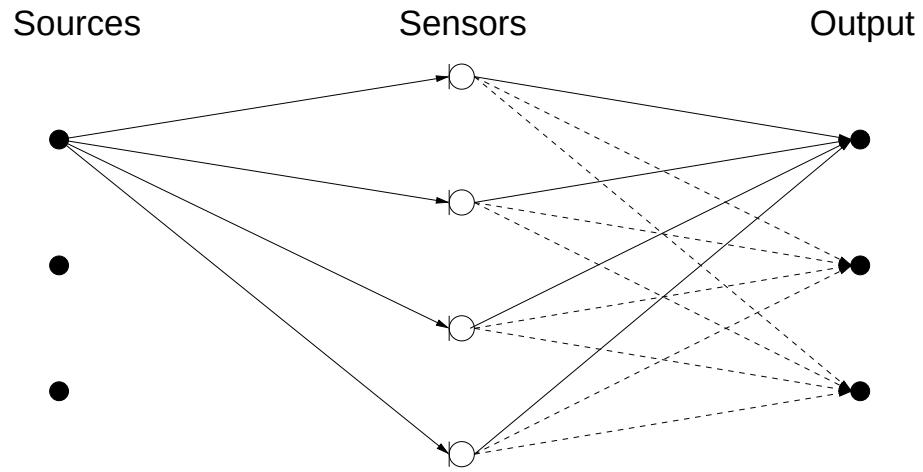

Figure 2.8: Optimal permutation

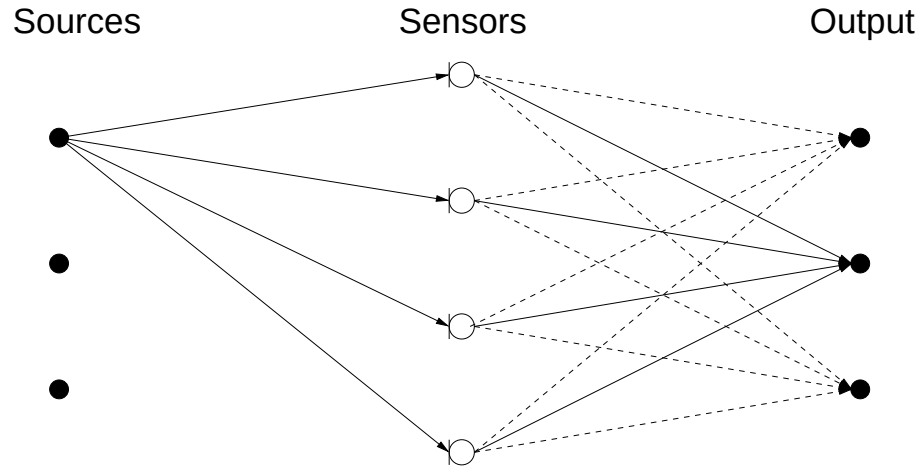

Figure 2.9: Wrong permutation

Y_{li} denotes the part of the output Y_l, $(1 \leq l \leq N)$, that comes from source number i. Assuming that the output should have the order of the source signals, i.e. $Y_i = S_i$, after separating the mixed signals $\mathbf{X}$, most of the energy of S_i should go to the output Y_i (Fig. 2.8). Due to small separation errors, we also find some contribution of S_i in other output signals $Y_l, (l \neq i)$. However, its contribution, measured by the energy is smaller than the contribution of S_i to its correct output signal Y_i. Thus, if the order of the unmixing system is not correct, we have the highest contribution of S_i to an output $Y_l, (l \neq i)$ (Fig. 2.9). Then we can rearrange the order of the unmixing system so that Y_i gets the highest contribution of S_i.

However, to implement this optimal solution the source signals and the mixing matrix must be available. Thus this method is only suitable for evaluation and not for practical application.

Chapter 3

Overdetermined BSS

The use of more sensors than the number of sources usually improves the separation result. In particular, we can improve performance by means of noise reduction, a technique known from beamforming theory [27]. In this chapter we extend the critically-determined BSS approach from chapter 2 to overdetermined BSS with more sensors than sources, i.e. we allow for $M > N$. We show that overdetermined BSS can actually be reduced to critically-determined BSS. As seen in Sec. 2.2 we can handle convolutive mixtures in the time domain by instantaneous mixtures in the frequency domain. As we continue considering convolutive mixtures we assume that our time-domain signals have already been transformed to the frequency domain. Therefore, the algorithms and ideas presented in this chapter are all applied in the frequency domain based on the model in Sec. 2.2.1.

3.1 General subspace selection

Let us consider the sensor signals $\mathbf{S}$ and separated output signals $\mathbf{Y}$ as elements of the N-dimensional complex vector space, and the mixed signals $\mathbf{X}$ as elements of the M-dimensional complex vector space.

$$\mathbf{S},\ \mathbf{Y} \in \mathbb{C}^N,\ \mathbf{X} \in \mathbb{C}^M, \qquad N < M \tag{3.1}$$

Then the mixing process can be described by a linear mapping $\mathbf{H}$

$$\mathbf{H} : \mathbb{C}^N \longrightarrow \mathbb{C}^M \tag{3.2}$$

and overdetermined BSS by a linear mapping $\mathbf{W}$

$$\mathbf{W} : \mathbb{C}^M \longrightarrow \mathbb{C}^N \tag{3.3}$$

where $\mathbb{C}^N$ is a subspace of $\mathbb{C}^M$. $\mathbf{H}$ is given by an $M \times N$ mixing matrix. Since we assume that it has full rank and $N < M$, its rank is given by the number of sources N. This relates to the assumption that two or more sources are not at the same position. It is obvious that mapping the mixed signals $\mathbf{X}$ from the higher dimensional sensor space $\mathbb{C}^M$ to the lower dimensional output space $\mathbb{C}^N$ includes selecting a subspace or in other words reducing the dimensions. An important question is where in the overdetermined BSS process we should optimally select the subspace. Basically we can reduce the dimensions before or after finding independent components by ICA. We can also use deflationary ICA itself to achieve dimension reduction. In the following we first discuss the best position for the subspace selection with respect to the implemented ICA algorithm. It appears to be more advantageous to reduce the dimensions before rather than after ICA. Then we describe two previously proposed methods for subspace selection. We assume that we know the number of sources N. If N is not known we can choose among several algorithms that were designed to estimate N [33].

3.1.1 Subspace selection before or after ICA

Let us consider the virtual sources depicted in Fig. 3.1 that consist, for example, of noise at the sensors. Then we might use a critically-determined BSS approach for M

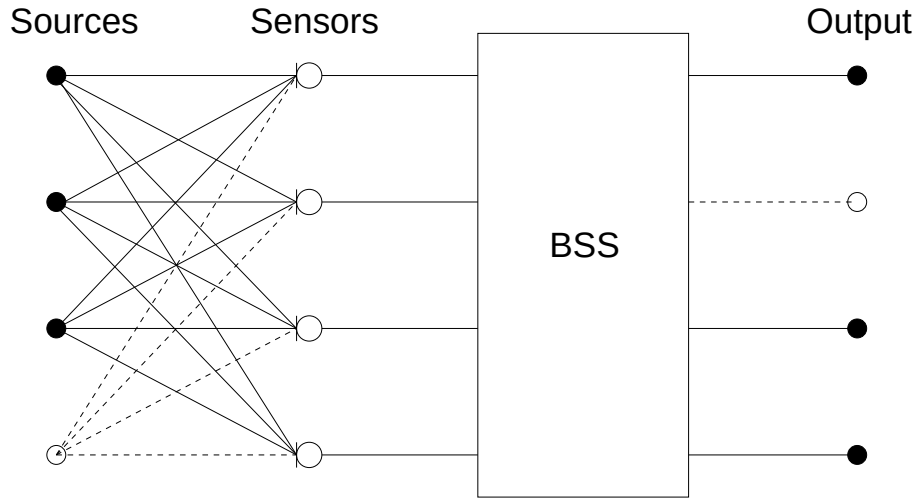

Figure 3.1: Virtual source

mixed signals to separate them. Separating them would provide us with the desired source signals but also with the virtual sources that we actually do not want. If we now want to sort out the virtual sources we 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 [13, 30]. The more signals we have, the more difficult it is to characterize the unmixing vectors of each frequency bin uniquely and relate them to the unmixing vectors of adjacent frequency bins or distinguish virtual and real sources.

We usually have more information before using ICA to select an appropriate subspace than after using ICA. As explained in more detail in Secs. 3.2 and 3.3 it can be, for example, the sensor spacing or the eigenvalues of the spatial correlation matrix $\mathbf{R_{XX}}$. The eigenvalues give the covariances of the mixed signals and are closely related to the power of both real and virtual source signals. However, since the mixed signals become unmixed the sensor spacing information is lost during the separation process. Similarly, due to the necessary step of normalizing the mixed signals $\mathbf{X}$ before using FastICA (cf. Sec. 2.1.2) the covariances become distorted and are no longer available. The latter is closely related to the problem that the power and therefore the covariance of the virtual sources is very small. By normalizing with the inverse square root of the covariance we therefore multiply with a large number and emphasize the noise [10]:129. Most ICA algorithms besides FastICA need normalization and therefore they are also affected by this problem. A promising approach to avoid the normalization problem is given by non-holonomic algorithms as proposed in [2]. However, the convergence speed might not be fast enough and we do not consider them in more detail in this thesis.

In addition, reducing the dimensions before ICA reduces the risk of the ICA algorithm overlearning due to the virtual sources [11]

Another advantage of dimension reduction before ICA becomes clear when we take the computational workload into account. The subspace selection stage always has to deal with as many signals as there are sensors whether it is employed before or after ICA. In contrast, the number of signals that the ICA stage must process depends on the position of the subspace selection stage. If it is employed before ICA, only the reduced number of signals must be separated. Otherwise ICA must also process as many signals as there are sensors. Thus we can save computational power and time if we first select a subspace before using ICA. In other words, dimension reduction before ICA constrains the search for independent components to a less costly subspace.

In conclusion we can state that for most ICA algorithms, including FastICA, it is more advantageous to reduce the dimensions before rather than after employing ICA. We can describe the subspace processing by an $N \times M$ matrix $\mathbf{W}_{sub}$. Since we employ it

before ICA $\mathbf{W}_{ICA}$ becomes a square $N \times N$ matrix.

3.1.2 Subspace selection by ICA

As mentioned earlier FastICA is a deflationary algorithm, i.e. it is in principle able to separate one signal after another. This would make it possible to stop separation after the desired number of signals is obtained and thereby reduce the dimensions. This case resembles that where we assume virtual sources to enable us to apply ICA to all sensor signals. Due to the permutation problem we cannot be sure that the first N signals that we separate by FastICA belong to the original sources and not to the virtual sources. Thus stopping separation when the desired number of signals has been obtained may result in one or more noisy virtual components instead of the correct components. Even worse, this bad separation result cannot be corrected afterwards by solving the permutation problem since not all correct components were separated. This problem might again be avoided if deflationary ICA algorithms were used that allow constrained BSS. They can be initialized so that they separate the same specified signal in every frequency bin and therefore avoid the permutation problem.

3.2 PCA-based subspace selection

PCA can be used for whitening the mixed signals as explained in Sec. 2.1.2 and also offers a very powerful mechanism for reducing the dimensions and so is widely used with various applications [15]. As we have already shown, PCA in general gives principal components that are by definition uncorrelated. It has already been successfully employed for example in [4, 5, 14] for dimension reduction in overdetermined BSS. While in [14] the effects of PCA were only investigated for instantaneous mixtures, in [5] and [4] PCA was applied to convolutive mixtures. However, the relationship between sensor distance, frequency bin and subspace selection was not assessed.

In our context PCA is based on the spatial correlation matrix $\mathbf{R_{XX}}$ of the mixed signals $\mathbf{X}$, which is similar to that in Eq. (2.14). Let us define the spatial correlation matrix $\mathbf{R_{SS}}$ of the source signals as

$$\mathbf{R_{SS}} := E\left\{\mathbf{S}\mathbf{S}^H\right\} \tag{3.4}$$

and likewise the noise correlation matrix $\mathbf{R_{NN}}$ as

$$\mathbf{R_{NN}} := E\left\{\mathbf{N}\mathbf{N}^H\right\} = \sigma_n^2\mathbf{I} \tag{3.5}$$

where σ_n^2 denotes the noise power. Both matrices are diagonal since the source signals and the noise are statistically independent. Together with the $M \times N$ mixing matrix $\mathbf{H}$ we can then write [4, 9]

$$\begin{aligned}
\mathbf{R_{XX}} &= E\left\{\mathbf{X}\mathbf{X}^H\right\} = E\left\{(\mathbf{H}\mathbf{S}+\mathbf{N})(\mathbf{H}\mathbf{S}+\mathbf{N})^H\right\} \\
&= E\left\{(\mathbf{H}\mathbf{S}+\mathbf{N})(\mathbf{S}^H\mathbf{H}^H+\mathbf{N}^H)\right\} \\
&= E\left\{\mathbf{H}\mathbf{S}\mathbf{S}^H\mathbf{H}^H\right\} + \underbrace{E\left\{\mathbf{H}\mathbf{S}\mathbf{N}^H\right\}}_{=0} + \underbrace{E\left\{\mathbf{N}\mathbf{S}^H\mathbf{H}^H\right\}}_{=0} + E\left\{\mathbf{N}\mathbf{N}^H\right\} \\
&= \mathbf{H}\mathbf{R_{SS}}\mathbf{H}^H + \sigma_n^2\mathbf{I}
\end{aligned} \tag{3.6}$$

We can decompose $\mathbf{R_{\tilde{X}\tilde{X}}} := \mathbf{H}\mathbf{R_{SS}}\mathbf{H}^H$ by the eigenvalue decomposition (EVD) into

$$\mathbf{R_{\tilde{X}\tilde{X}}} = \mathbf{E}\tilde{\Lambda}\mathbf{E}^H \tag{3.7}$$

where

$$\tilde{\Lambda} = \begin{bmatrix} \tilde{\lambda}_1 & & \mathbf{0} \\ & \ddots & \\ & & \tilde{\lambda}_N \\ \mathbf{0} & & \mathbf{0} \end{bmatrix} \tag{3.8}$$

denotes a diagonal matrix with the eigenvalues $\tilde{\lambda}_i$ of $\mathbf{R_{\tilde{X}\tilde{X}}}$ and $\mathbf{E}$ stands for the respective eigenvectors. Without loss of generality, we can arrange the eigenvalues in decreasing order with respect to their absolute value

$$|\tilde{\lambda}_i| > |\tilde{\lambda}_{i+1}| \tag{3.9}$$

$\mathbf{R_{\tilde{X}\tilde{X}}}$ has $M - N$ eigenvalues equal to 0 since $\mathbf{H}$ and $\mathbf{R_{SS}}$ are of rank N [22, 27]. Together with Eq. (3.6) it follows from here that the EVD of $\mathbf{R_{XX}}$ is given by [9]

$$\mathbf{R_{XX}} = \mathbf{E}\Lambda\mathbf{E}^H \tag{3.10}$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_M \end{bmatrix} = \begin{bmatrix} \tilde{\lambda}_1 + \sigma_n^2 & & & & & \\ & \ddots & & & \mathbf{0} & \\ & & \tilde{\lambda}_N + \sigma_n^2 & & & \\ & & & \sigma_n^2 & & \\ & \mathbf{0} & & & \ddots & \\ & & & & & \sigma_n^2 \end{bmatrix} \tag{3.11}$$

i.e. the power of the N source signals $\mathbf{S}$ is concentrated in the first N eigenvalues, whereas the noise power is uniformly distributed throughout the eigenvalues [5].

We obtain the principal components by projecting the mixed signals onto the eigenvectors of $\mathbf{R_{XX}}$. Then the first N principal components corresponding to the largest eigenvalues contain a mixture of direct source signals and noise. By contrast the remaining principal components consist solely of noise. Thus by selecting the subspace that is spanned by the first N eigenvectors, dimensions are effectively reduced by removing noise while keeping the signals of interest [23].

Since PCA linearly combines the mixed signals, the noise reduction can be backed up by the increase in the signal-to-noise ratio (SNR) known from array processing [27]:13. In the ideal case of coherently adding together several sensors the increased SNR_{new} is given by

$$\mathrm{SNR}_{new} = \log(M) \cdot \mathrm{SNR}_{old} \tag{3.12}$$

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

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\text{signal power}}{\text{noise power}} \right) \tag{3.13}$$

The fact that PCA-based subspace selection effectively removes noise is also closely related to the optimal behavior of PCA from the standpoint of information theory. That is, subspace reduction by PCA preserves information as well as possible in the least-squares sense [10]:267.

Due to the nature of the PCA-based subspace approach, whitening is already included, i.e. the whitened signals are calculated implicitly. However, if we consider whitening and subspace selection as two different steps we first have to whiten the signals before we can select the subspace. Thus the unmixing matrix W is given by

$$\mathbf{W} = \mathbf{W}_{ICA} \underbrace{\mathbf{W}_{sub}^{PCA} \mathbf{W}_{white}}_{:=W_{PCA}} \tag{3.14}$$

where $\mathbf{W}_{sub}^{PCA}$ denotes the actual subspace selection matrix based on PCA. This leads to the structure depicted in Fig. 3.2. We summarize $\mathbf{W}_{sub}^{PCA} \mathbf{W}_{white}$ by

$$\mathbf{W}_{PCA} := \mathbf{W}_{sub}^{PCA} \mathbf{W}_{white} \tag{3.15}$$

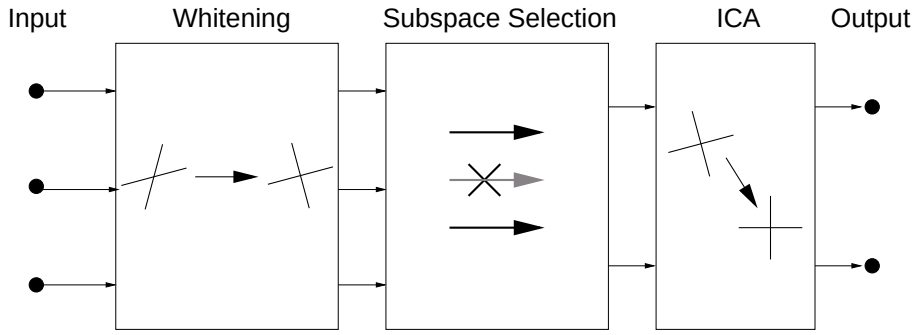

Figure 3.2: Overdetermined BSS with PCA-based subspace selection

3.3 Geometry-based subspace processing

Sawada et al. proposed a method for blind source separation using several separating subsystems whose sensor spacings could be configured individually [29]. The idea is based on the observation that BSS behaves like a beamformer in that it forms spatial nulls in the jammer directions [3], i.e. the jammer is suppressed. From this it follows that the optimal sensor spacing depends on the frequency. If the distance between the sensors is larger than half the wave length spatial aliasing occurs. This means that nulls are not only formed toward the jammer but also in other directions and this downgrades the separation performance. The performance can also suffer from a sensor spacing that is too small. In this case the resulting phase difference, which plays a key role in separating signals, is too small, too. In other words low frequencies prefer a wide sensor spacing whereas high frequencies prefer a narrow sensor spacing. 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. 3.3.

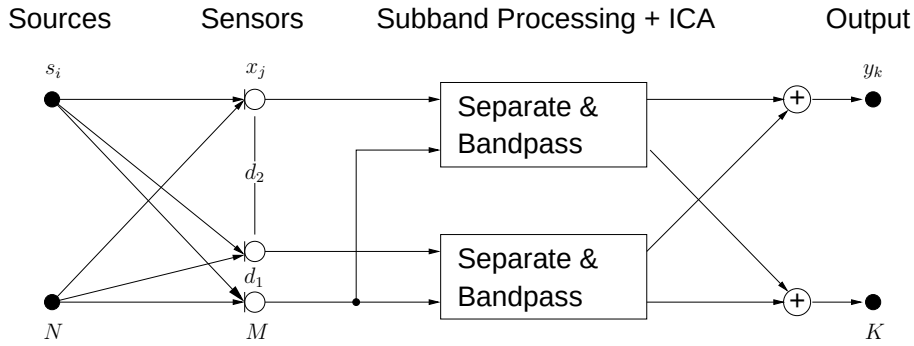

Figure 3.3: Geometry-based subspace selection [29]

The frequency range of the mixed signals was divided into lower and higher frequency

ranges. According to [29], for a frequency to be adequate for a given sensor spacing d the condition in (3.16) should be fulfilled.

$$f \leq \frac{\alpha c}{2d\left(\cos(\theta_1) - \cos(\theta_2)\right)} \tag{3.16}$$

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. 2.4. This result can be derived using the mixing model described in Sec. 2.2.1. The DOA can be estimated by the MUSIC algorithm [33]. Appropriate sensor pairs were chosen for each frequency range and used separately for separation in each frequency range. The mixed signals were whitened before ICA was applied to each chosen pair. This means that subspace selection and whitening are two separate steps, and their order is different from that in PCA-based subspace processing. Thus the unmixing matrix $\mathbf{W}$ can be decomposed in the frequency domain into

$$\mathbf{W} = \mathbf{W}_{ICA}\mathbf{W}_{white}\mathbf{W}_{sub}^{geo} \tag{3.17}$$

where $\mathbf{W}_{sub}^{geo}$ denotes the geometry-based subspace selection. This leads to the structure depicted in Fig. 3.4.

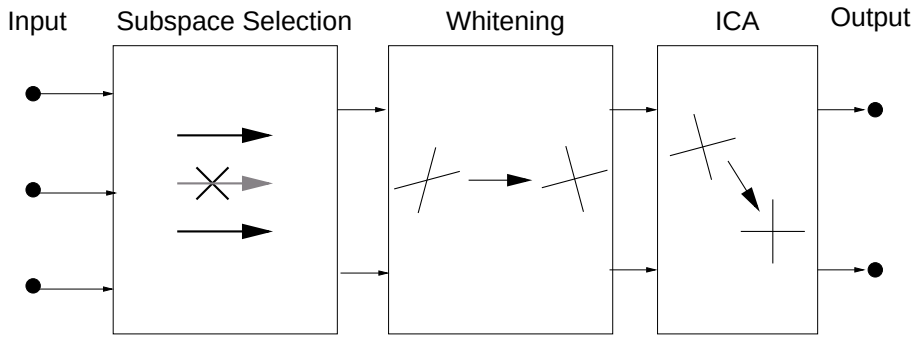

Figure 3.4: Overdetermined BSS with geometry-based subspace selection

The similarities and differences between the two subspace selection methods are summarized in Table 3.1.

Figure 3.5 shows the complete framework for 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 whitening stage and a dimension reduction stage. The order is different in the two methods described here.

Table 3.1: Summarized comparison

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

3.4 Implemetation details

We used MATLAB® to implement discrete versions of the above mentioned algorithms for BSS. Thus the signals, the mixing and unmixing system were sampled in time and frequency domain. Thereby k denotes the discrete time and μ the discrete frequency. Instead of the STFT we employed the short-time discrete Fourier transform (STDFT)

$$X_j[\mu, m] = \mathrm{STDFT}(x[k]) := \sum_{k=0}^{L-1} x[m(L-D)+k] \cdot w[k] e^{-j\frac{2\pi\mu}{L}k} \tag{3.18}$$

$$\tag{3.19}$$

where m denotes the discrete time parameter, L the framesize or filter length, respectively, and D the block overlap. The possible values for μ are given by

$$0 \leq \mu \leq L-1 \tag{3.20}$$

For $w[k]$ we chose the Hann window

$$w[k] = \begin{cases} 0.5 - 0.5\cos\left(\frac{2\pi k}{K}\right), & 0 \leq k \leq K-1 \leq L-1 \\ 0, & \text{else} \end{cases} \tag{3.21}$$

where K defines the relevant window length. The time-domain unmixing filters were obtained by the inverse discrete Fourier transform.

Since we do not know the exact pdfs or expectation values we approximated the frequency dependent expectation values by using time averages given exemplarily for $\mathbf{X}$ by

$$\mathbf{E}\{\mathbf{X}\} \approx \frac{1}{T} \sum_{m=0}^{T-1} \mathbf{X}[\mu, m] \tag{3.22}$$

where T denotes the number of available samples with regard to m. We assume finite impulse response (FIR) filters.

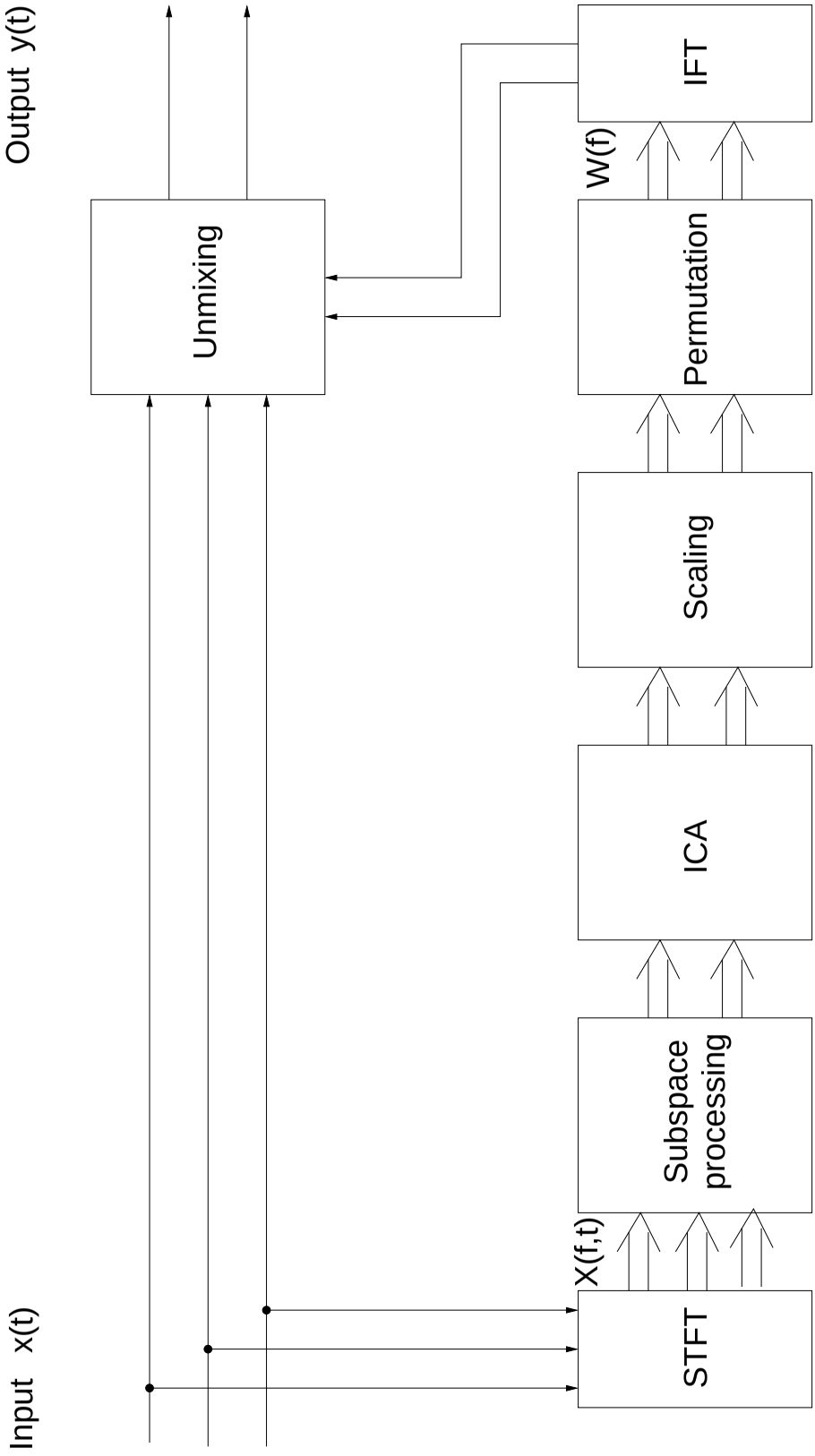

Figure 3.5: General framework of overdetermined BSS

Chapter 4

Geometric interpretation of the PCA-based subspace approach

In this chapter we investigate the behavior of the PCA-based subspace approach more closely and relate it to the geometry-based method. In particular we consider a mixing model with $N = 2$ sources and $M = 3$ equispaced sensors.

4.1 Experimental results

The geometry-based subspace approach selects sensors according to the sensor position, DOA and frequency. In order to investigate the relationship between the geometry- and PCA-based approaches we experimentally examined the behavior of the PCA-based subspace method with regard to the resulting sensor selection. We considered real and artificially generated source signals for the following reason.

Speech signals do not always comply with assumptions like uncorrelatedness and independence, which are made when applying PCA and ICA to them. Therefore, so that we could also assess the ideal behavior, we used artificial source signals with the desired properties produced by a random generator in the frequency domain instead of real speech signals. The super-Gaussian pdf $f(S_i)$ of the frequency-domain source signals S_i is given by

$$f(S_i) = \frac{1}{\sqrt{2\pi}|3S_i^{\frac{2}{3}}|} e^{-\frac{S_i^{\frac{2}{3}}}{2}} \tag{4.1}$$

It is depicted in Fig. 4.1 together with the Gaussian pdf for zero mean and unit variance. In order to obtain mixtures from the artificial source signals we assumed

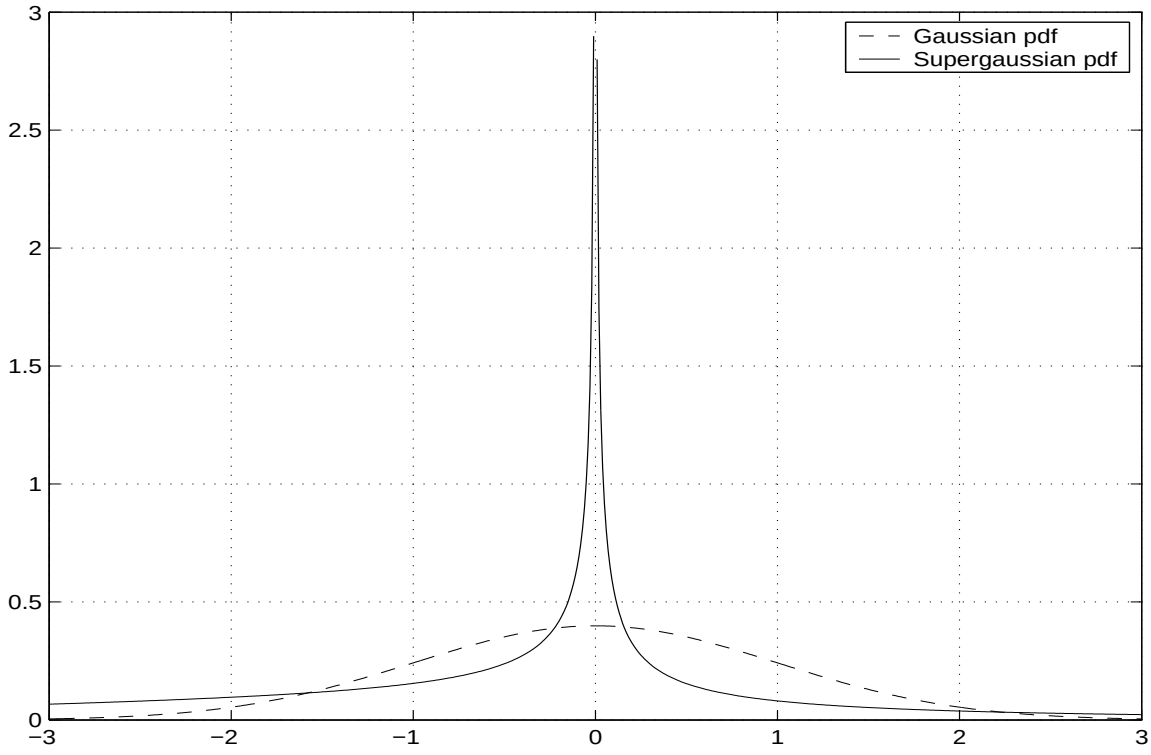

Figure 4.1: Probability density functions

a frequency dependent mixing matrix $\mathbf{H}(f)$ according to Sec. 2.2.1. As we already mentioned, the DOA θ_{ji} does not necessarily coincide with the angle that is related to the physical position of the source and might vary depending on the frequency. However, for theoretical analysis we can approximate a fixed DOA for each source by [29]

$$\theta_{ji} \approx \theta_i \tag{4.2}$$

Since we assume a far-field situation, we can also neglect the difference in attenuation between different sensors for a particular source signal. Thus, we assume that each source signal has a specific, but constant, attenuation at each sensor given by

$$c_{ji} \approx c_i \tag{4.3}$$

This yields a simplified mixing matrix

$$\mathbf{H}(f) = \begin{bmatrix} c_1 e^{j\beta_{11}} & \dots & c_N e^{j\beta_{1N}} \\ \vdots & \ddots & \vdots \\ c_1 e^{j\beta_{M1}} & \dots & c_N e^{j\beta_{MN}} \end{bmatrix} \tag{4.4}$$

where β_{ji} is now given by

$$\beta_{ji} = \frac{2\pi f \cos(\theta_i) \sum_{l=0}^{i-1} d_l}{c} \tag{4.5}$$

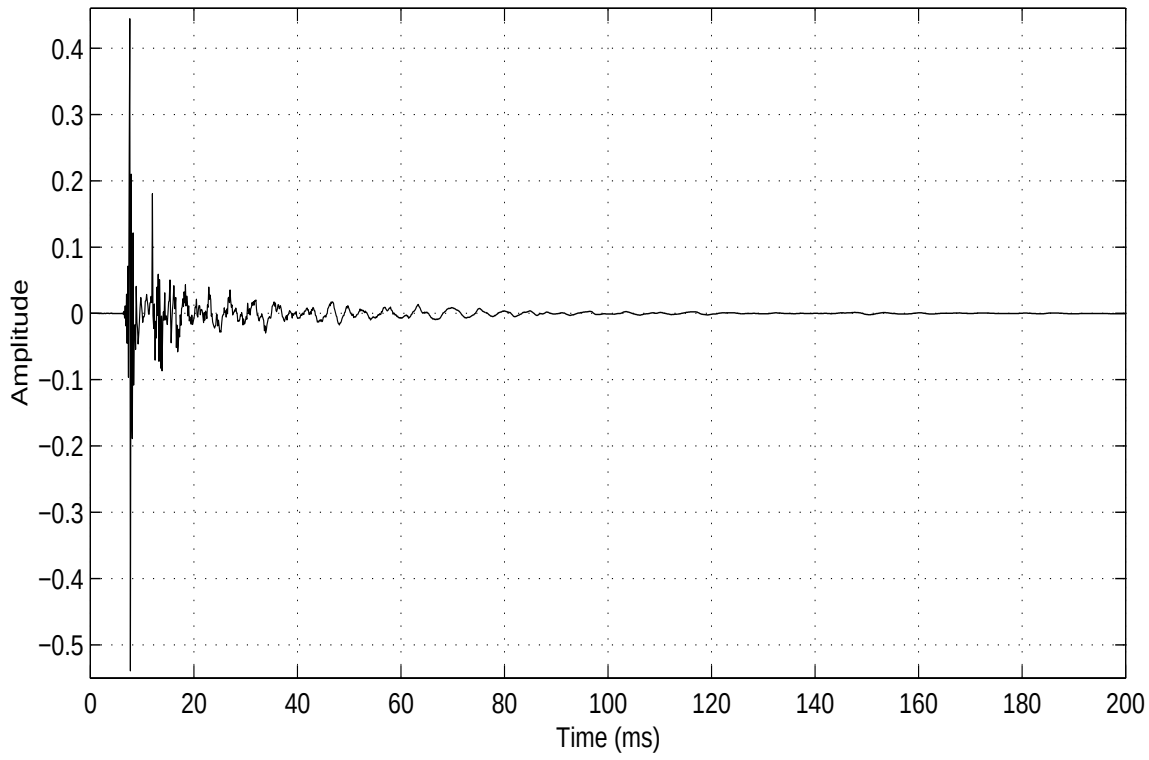

Figure 4.2: Example impulse response

For the real signals we employed 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 [28]. An example of the impulse responses is shown in Fig. 4.2. The typical reverberation time is about $190ms$. The source directions θ_i were estimated by the MUSIC algorithm [33].

In Figs. 4.3-4.7 the normalized[1] sensor gain of the unmixing system for each frequency bin and sensor is shown for different experimental conditions. Denoting the row vectors of the unmixing matrix $\mathbf{W}$ by $\mathbf{w}_i$ and the elements of $\mathbf{w}_i$ by w_{ij} the normalized and frequency dependent sensor gains are given by

$$\frac{|\mathbf{w}_i|}{\max(|w_{i1}|, \ldots, |w_{iN}|)} \tag{4.6}$$

The figures were generated by depicting the normalized, absolute values of the unmixing vectors that were obtained by applying the PCA-based subspace approach and afterwards FastICA to real and artificially generated source signals, respectively. The experimental conditions for each figure are summarized in Table 4.1. Figures 4.3 and 4.4 show the ideal case with artificial signals for 3 uniformly spaced sensors for the

[1] each frequency bin was normalized to its maximum gain

Table 4.1: Experimental conditions

Figure	4.5	4.3, 4.4	4.6	4.7
Source signals	recorded speech	artificially generated		
Length of source signals	7.4 seconds	1000 samples		
Sampling rate	8 kHz	N/A		
Number of sources	2			
Direction of arrival	$\theta_1 = 50°$, $\theta_2 = 150°$			
Number of sensors	3			7
Distance of sensors (mm)	$d_1 = 28.3$ $d_2 = 28.3$		$d_1 = 28.3$ $d_2 = 56.6$	$d_i = 28.3$ $(i = 1, \ldots, 7)$
Mixing process	recorded impulse responses	artificial mixing matrix		
Windowing function	Hann	N/A		
Filter length	2048 points			
Shifting interval	512 points			
Threshold ϵ for FastICA	10^{-3}			

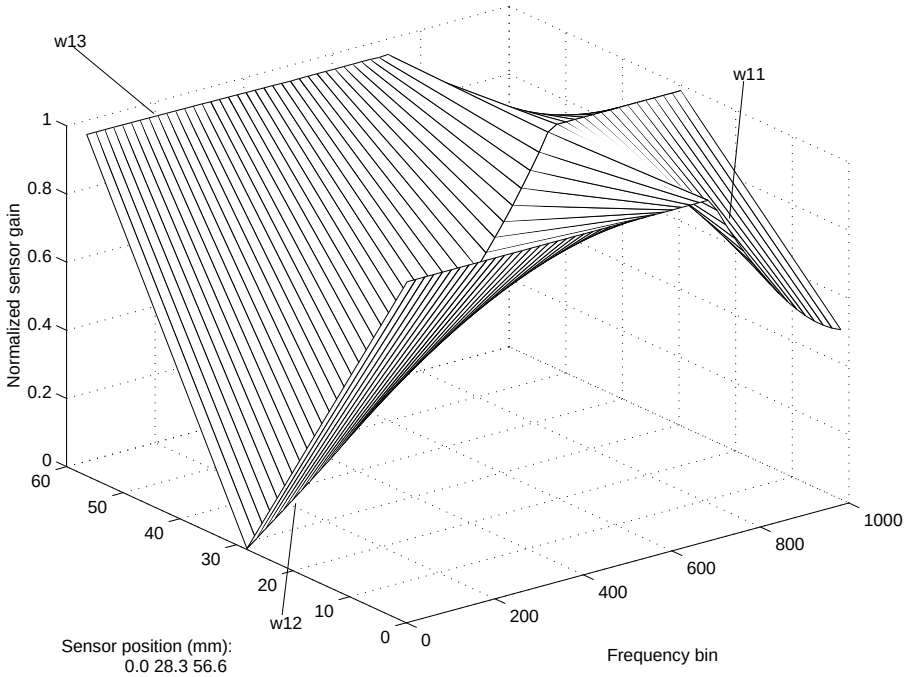

Figure 4.3: Normalized sensor gain with PCA-based subspace selection for 3 uniformly spaced sensors and first artificial source signal

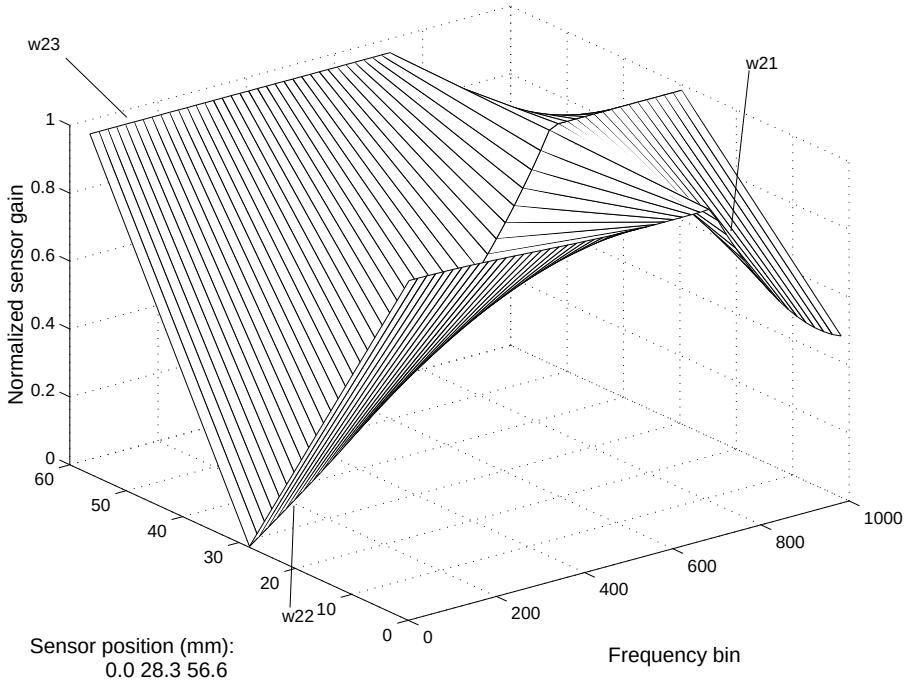

Figure 4.4: Normalized sensor gain with PCA-based subspace selection for 3 uniformly spaced sensors and second artificial source signal

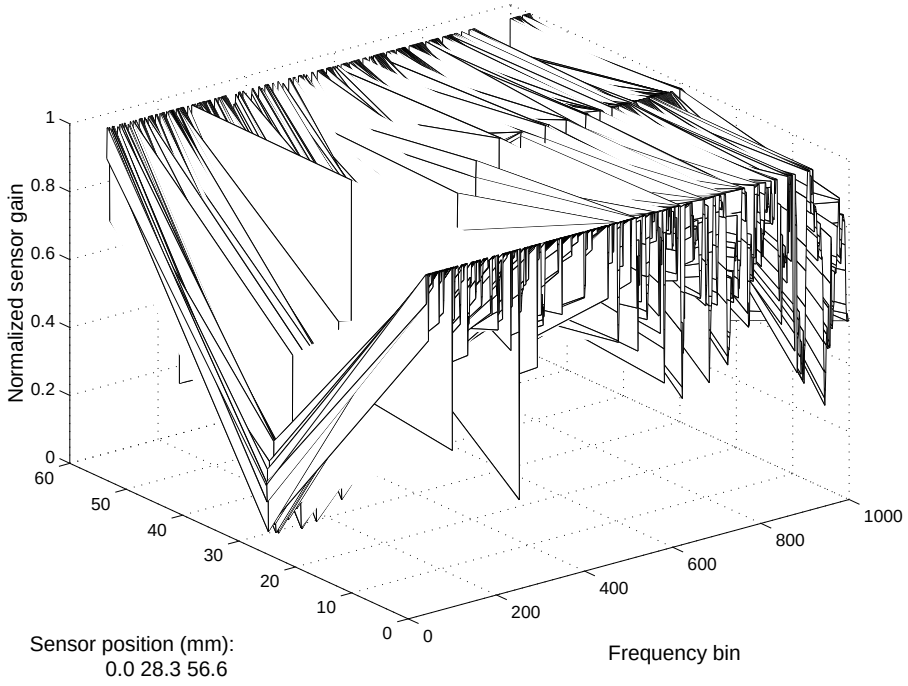

Figure 4.5: Normalized sensor gain with PCA-based subspace selection for 3 uniformly spaced sensors and real speech signals

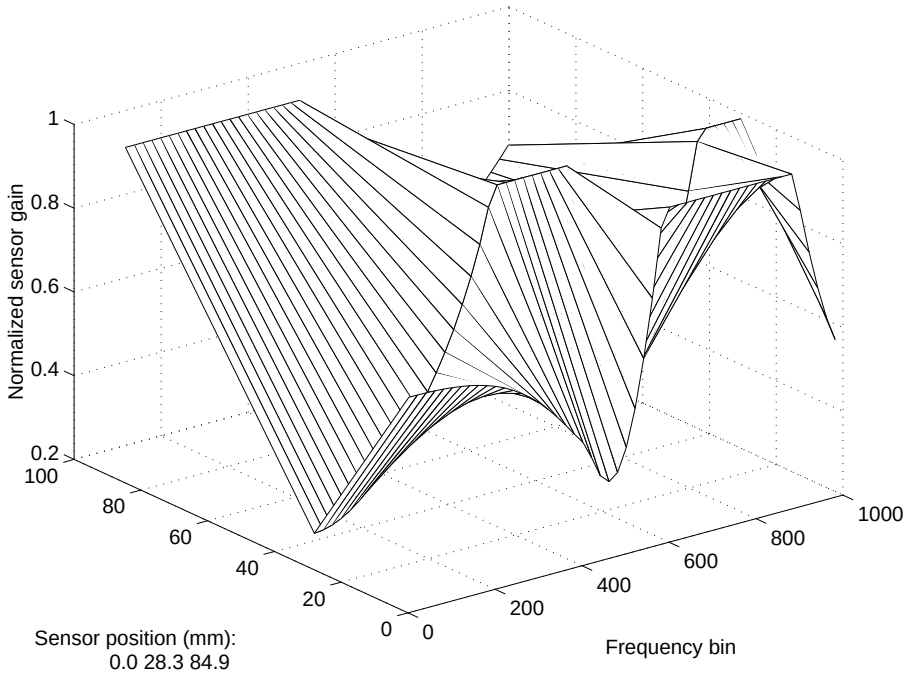

Figure 4.6: Normalized sensor gain with PCA-based subspace selection for 3 unequally spaced sensors and artificial signals

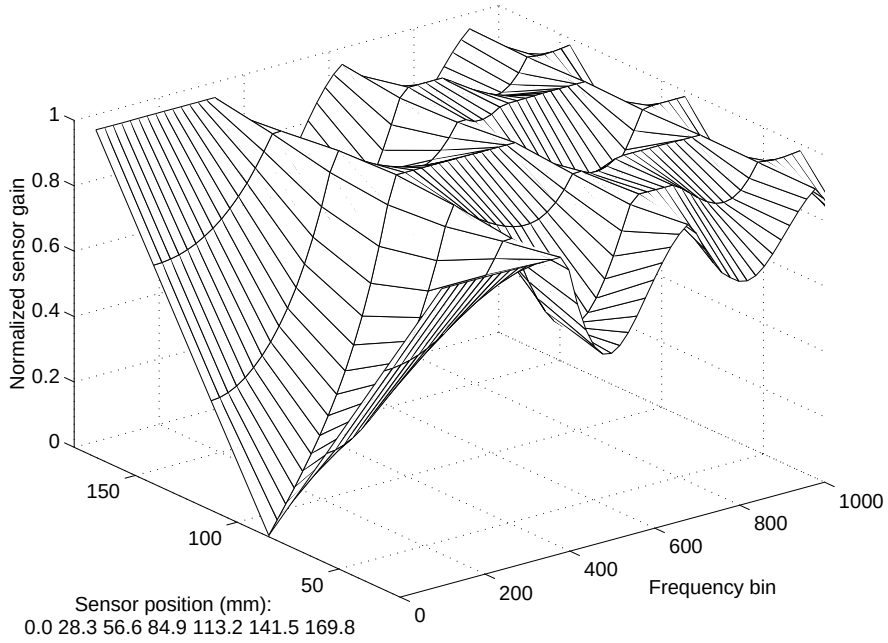

Figure 4.7: Normalized sensor gain with PCA-based subspace selection for 7 uniformly spaced sensors and artificial signals

first and the second source signal, respectively. They reveal most clearly the principle that the PCA-based method also emphasizes the outer sensors with a wide spacing for low frequencies as the geometrical considerations in [29] suggest. However, the remaining sensor in the center is not excluded as with the geometry-based approach but contributes more the higher the frequency becomes. At the highest frequencies it contributes about twice as much as the outer sensors. The similarity between the landscapes of both signals can be explained by the fact that we only account for the absolute value and not the phase. Since this is also true for the following figures, we omitted the landscape of the second signal there. We used the same conditions as for Fig. 4.3 to generate Fig. 4.5 except that we employed real speech signals and impulse responses instead of artificial signals and a mixing matrix. While it is not as clear as Fig. 4.3 it still illustrates the principle that outer sensors are preferred for low frequencies and justifies the assumptions made for the ideal case. This principle is also apparent if we choose different conditions such as an unequal sensor spacing or more than 3 sensors as was done in Figs. 4.6 and 4.7, though the sensor gain landscape becomes more complicated, particularly for high frequencies.

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 according to the approach described in Sec. 3.2. 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 Fig. 4.8 and 4.9 for each frequency bin. The result is normalized and was generated using the same conditions as for Fig. 4.3. The elements w_{ij} now belong to $\mathbf{W}_{PCA}$. For the first principal component in Fig. 4.8 all sensors contribute with approximately the same gain for low frequencies. In contrast, the outer sensors are emphasized for high frequencies. The sensor gain of the second principal component in Fig. 4.9 already shows the emphasis of the outer sensors for lower frequencies as well as the emphasis of the middle sensor for higher frequencies, which is dominant in the resulting sensor gain landscape in Figs. 4.3-4.7.

4.2 Interpretation of experimental results

4.2.1 Low frequencies

In this section we provide an explanation for the experimental result whereby the PCA-based subspace approach automatically selects the outer two sensors for low

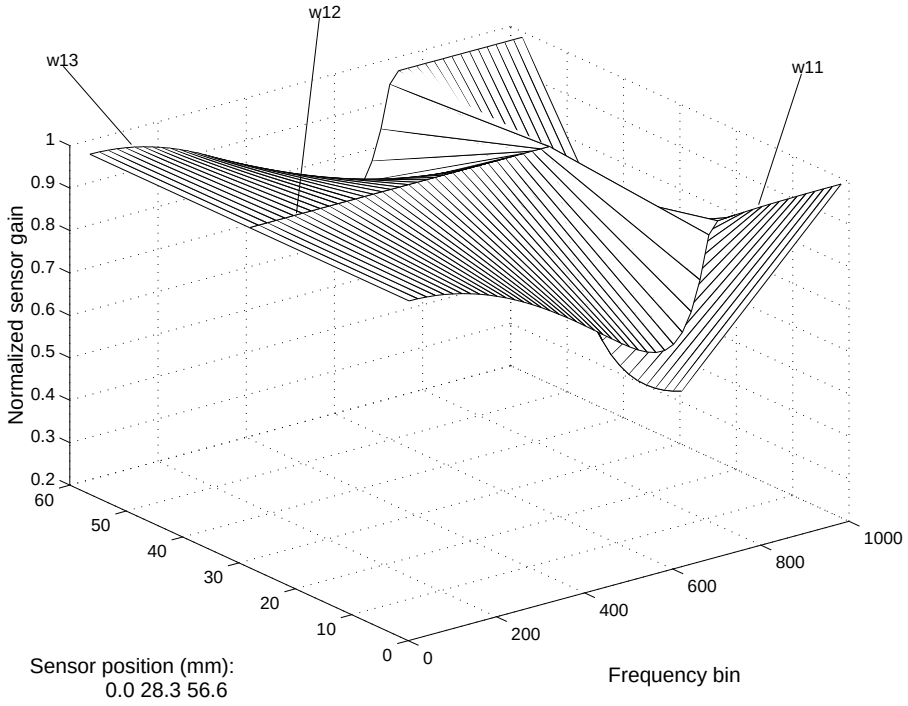

Figure 4.8: Normalized sensor gain corresponding to the first principal component of Fig. 4.3

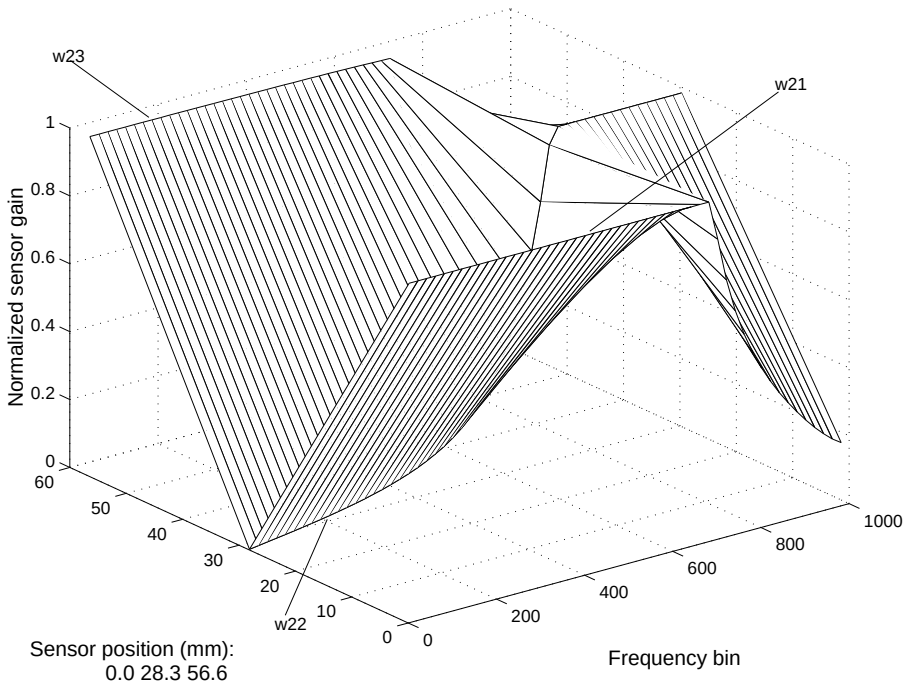

Figure 4.9: Normalized sensor gain corresponding to the second principal component of Fig. 4.3

frequencies to produce a good result. First we show analytically why the second principal component already exhibits this behavior. Then we explain its dominance in the final sensor selection landscape.

In order to explain the landscape of the eigenvectors that correspond to the first and second principal components in Figs. 4.8 and 4.9 we assume a 3×2 mixing matrix $\mathbf{H}$ as described in Sec. 2.2.1 and uniformly spaced sensors ($d_1 = d_2 := d$). This yields

$$\mathbf{H} = \begin{bmatrix} c_1 & c_2 \\ c_1 e^{j\omega_1} & c_2 e^{j\omega_2} \\ c_1 e^{j2\omega_1} & c_2 e^{j2\omega_2} \end{bmatrix} \tag{4.7}$$

where

$$\omega_i = \frac{2\pi f d \cos\theta_i}{c} \tag{4.8}$$

Then we obtain the mixed signals $\mathbf{x}$ as

$$\mathbf{X} = \mathbf{HS} = \mathbf{H} \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} c_1 S_1 + c_2 S_2 \\ c_1 e^{j\omega_1} S_1 + c_2 e^{j\omega_2} S_2 \\ c_1 e^{2j\omega_1} S_1 + c_2 e^{2j\omega_2} S_2 \end{bmatrix} \tag{4.9}$$

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

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \tag{4.10}$$

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

$$\begin{aligned} \mathbf{p}^H \mathbf{X} = \mathbf{p}^H \mathbf{HS} = c_1 S_1 (p_1^* + p_2^* e^{j\omega_1} + p_3^* e^{2j\omega_1}) + \\ c_2 S_2 (p_1^* + p_2^* e^{j\omega_2} + p_3^* e^{2j\omega_2}) \end{aligned} \tag{4.11}$$

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

$$\bar{\omega}_i = \min_{\bar{\omega}_i} ||(p_1^* + p_2^* e^{j\omega_i} + p_3^* e^{2j\omega_i}) - (p_1^* e^{\bar{\omega}_i} + p_2^* e^{\bar{\omega}_i} + p_3^* e^{\bar{\omega}_i})|| \tag{4.12}$$

Thus we obtain

$$\mathbf{p}^H \mathbf{X} \approx (c_1 S_1 e^{j\bar{\omega}_1} + c_2 S_2 e^{j\bar{\omega}_2})(p_1^* + p_2^* + p_3^*) \tag{4.13}$$

The first principal component corresponding to Fig. 4.8 is found by maximizing the power $E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\}$ with the constraint $||\mathbf{p}|| = 1$. By the Lagrange multiplier approach

$$\nabla \left(E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\} + \gamma(||\mathbf{p}|| - 1) \right) = \mathbf{0} \tag{4.14}$$

where ∇ is the Nabla operator and γ the Lagrange multiplier, we can show that with the approximation in Eq. (4.13) the maximum is obtained if $p_1 = p_2 = p_3$, which means that all sensors have approximately equal influence. In this case the LSE solution for $\bar{\omega}_i$ equals ω_i. A more detailed derivation is given in appendix A.

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

$$\frac{\mathbf{p}^H\mathbf{X}}{\mathbf{p}^H\mathbf{p}}\mathbf{p} \overset{\substack{||\mathbf{p}||=1 \\ p_1=p_2=p_3}}{\approx} \begin{bmatrix} c_1 e^{j\omega_1} S_1 + c_2 e^{j\omega_2} S_2 \\ c_1 e^{j\omega_1} S_1 + c_2 e^{j\omega_2} S_2 \\ c_1 e^{j\omega_1} S_1 + c_2 e^{j\omega_2} S_2 \end{bmatrix} \tag{4.15}$$

From comparing this result with Eq. (4.9) 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.

To explain the dominance of the second principal component even after employing ICA we have to examine the process of normalizing the mixed signals after projecting them onto the principal components. According to Sec. 2.1.2 the normalizing is performed by the inverse square root of the respective eigenvalue. A typical frequency dependent eigenvalue distribution is shown in Fig. 4.10. For low frequencies the eigenvalue of the first principal component is very large compared with the eigenvalue of the second principal component. This in turn means that the first principal component is attenuated and the second principal component is amplified. Thus the second principal component has a dominant influence when combined with the first principal component by the subsequent ICA stage.

4.2.2 High frequencies

The sensor gain landscape for high frequencies in Fig. 4.3 also resembles the result of the geometry-based subspace approach in [29]. The sensor distance at the highest

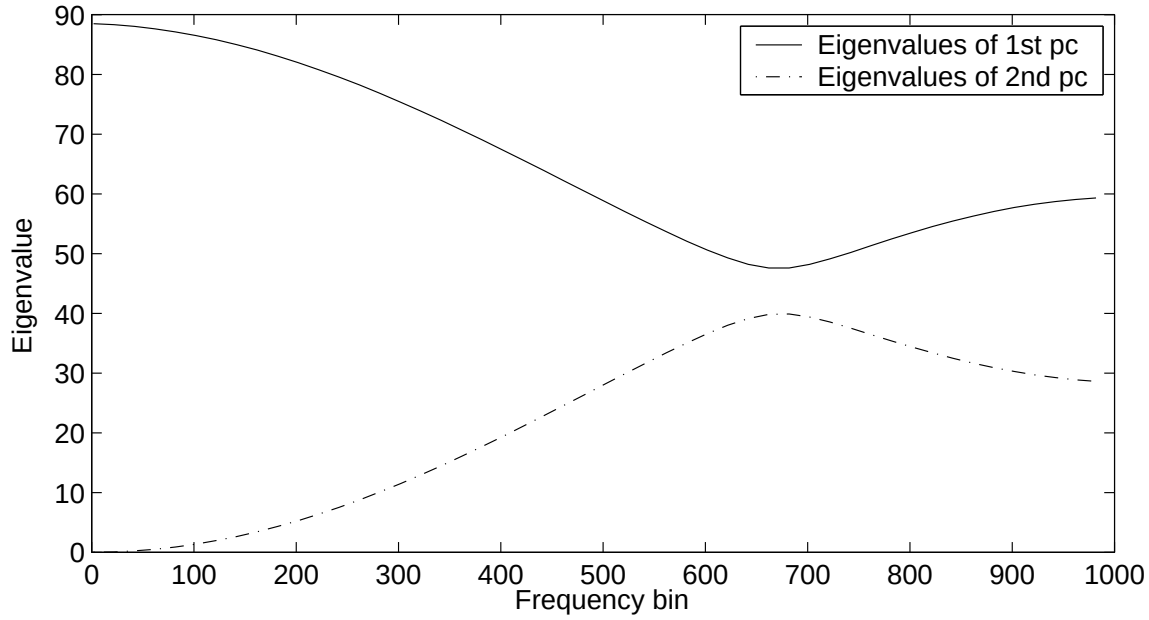

Figure 4.10: Typical absolute eigenvalues corresponding to the first and second principal component for each frequency bin

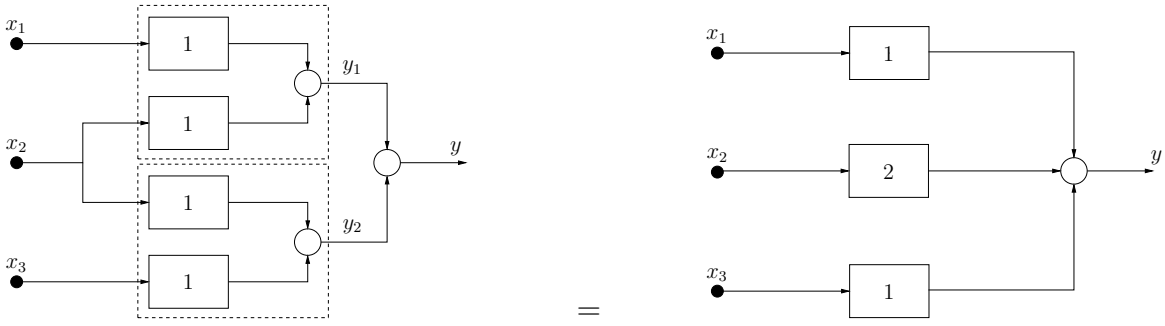

Figure 4.11: Combined processing for high frequencies (normalized, phase omitted)

frequency equals approximately half the wavelength. This is the maximum possible distance while still avoiding spatial aliasing and should be the optimum distance. Thus two adjacent sensors should be selected to obtain the best result. As mentioned in Sec. 4.1 the contribution of the center sensor is roughly twice that of the outer ones. This is equal to processing the mixed signals for each adjacent pair of sensors separately and adding the results later as shown in Fig. 4.11. There is a continuous transition between the lowest and highest frequency with respect to choosing the relative sensor gains.

Chapter 5

Comparison of the PCA- and geometry-based approaches

In this chapter we present experimental results to compare the performance of the two previously described subspace methods. We provide an explanation based on the results reported in chapter 4.

5.1 Experimental results

To compare the PCA- and geometry-based methods, we separated real speech mixtures that we obtained by convolving impulse responses $h_{ji}[k]$ and pairs of speech signals $s_i[k]$, and optionally adding artificial Gaussian noise $n_j[k]$ with zero mean and variance $\sigma_n^2 = 10^{-4}$. We employed speech signals and impulse responses from the same databases as in Sec. 4.1 and used the same conditions as given in Table 4.1 for Fig. 4.3. In addition we conducted experiments for filter lengths of 1024 and 4096 points with a shifting interval of 265 and 1024 points, respectively. The frequency ranges for the geometry-based method were calculated based on the criteria discussed in Sec. 3.3 where $\alpha = 1.2$. We measured the performance in terms of the signal-to-noise plus interference ratio (SNIR) in dB at output number i in terms of

$$\mathrm{SNIR}_i = 10 \log \left(\frac{\sum_k y_i^s[k]^2}{\sum_k y_i^{cn}[k]^2} \right) \tag{5.1}$$

where $y_i^s[k]$ is the portion of a output $y_i[k]$ that comes only from a source signal $s_i[k]$ and $y_i^{cn}[k]$ is the portion of $y_i[k]$ that comes from the remaining signals (interference)

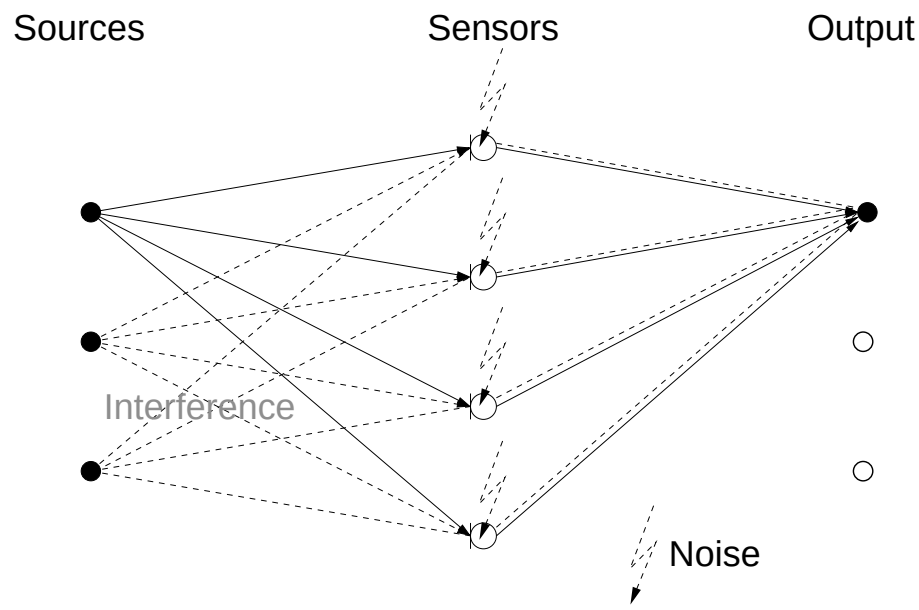

Figure 5.1: Calculating SNIR

and optional noise as indicated in (Fig 5.1).

$$y_i^{cn}[k] = y_i[k] - y_i^s[k] \tag{5.2}$$

Figures 5.2-5.7 show the results for both subspace approaches. In order to account for all outputs $y_i[k],\ (1 \leq i \leq N)$ we calculated an averaged SNIR given by

$$\text{SNIR} = \frac{1}{N}\sum_{i=1}^{N}\text{SNIR}_i \tag{5.3}$$

Each figure depicts the results for 12 different pairs of speech signals obtained for a specific filter length. Additionally we distinguished between the low and the high frequency ranges that were used by the geometry-based approach. Figures 5.2, 5.4 and 5.6 reveal that, regardless of filter length, both subspace methods perform similarly for low frequencies with and without 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, Figs. 5.3, 5.5 and 5.7 suggest, again regardless of filter length, that the performance differs significantly if we consider the high frequency range. 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.

In Figures 5.8-5.11 show the performance for part of the low frequency range broken down into single frequency bins. As an example we chose source pair number 3.

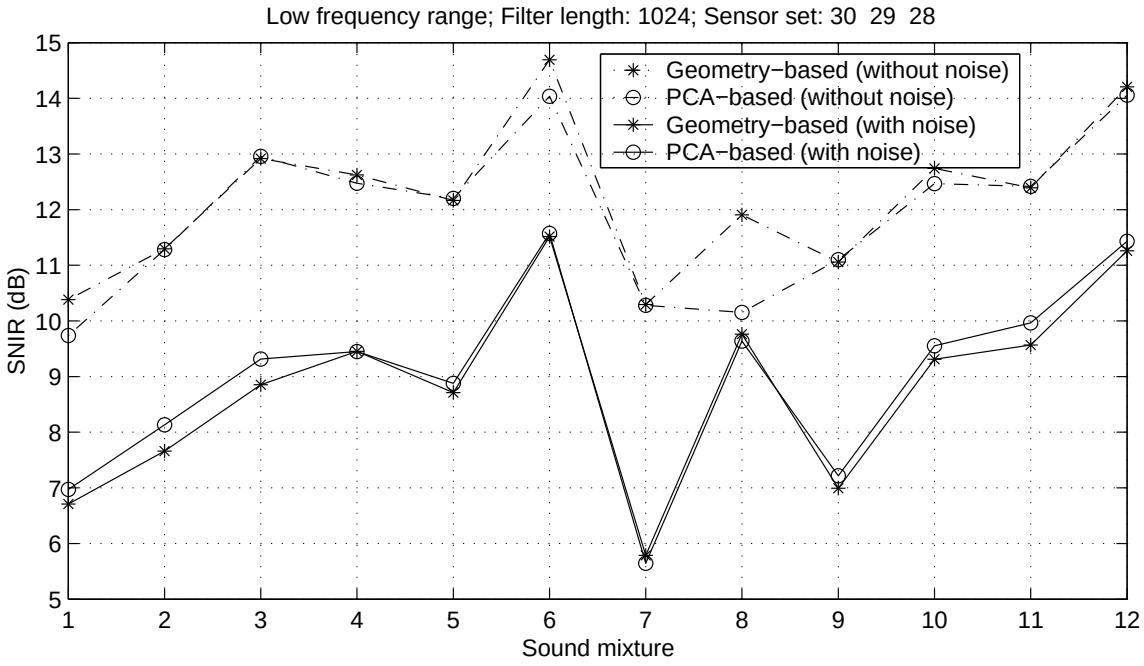

Figure 5.2: Comparison of PCA- and geometry-based subspace selection

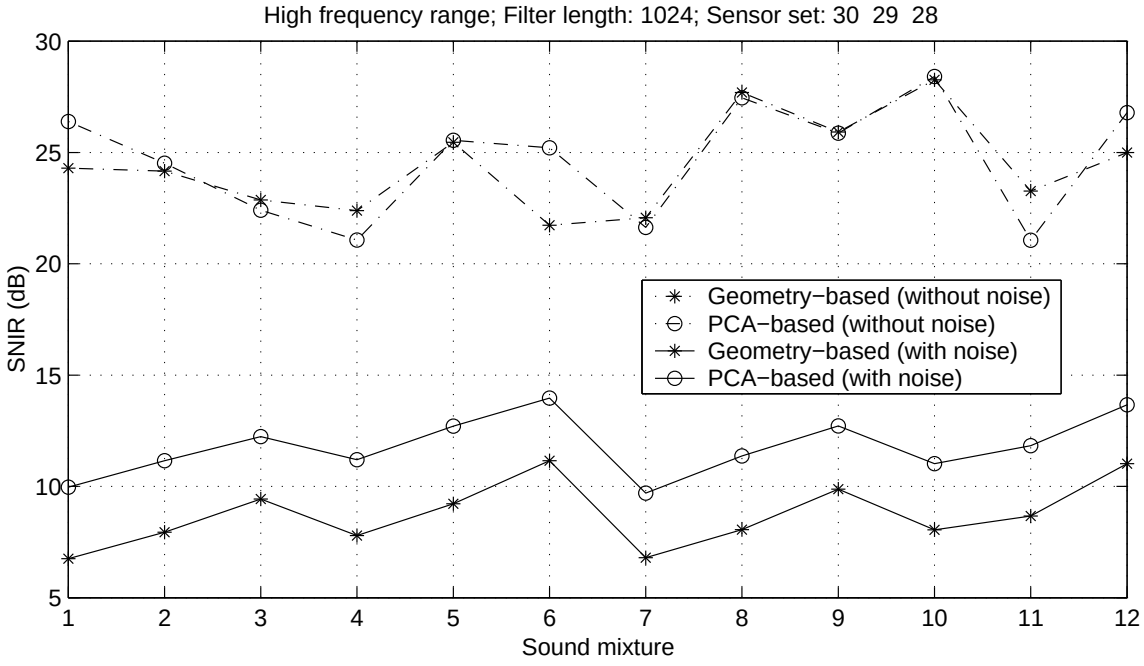

Figure 5.3: Comparison of PCA- and geometry-based subspace selection

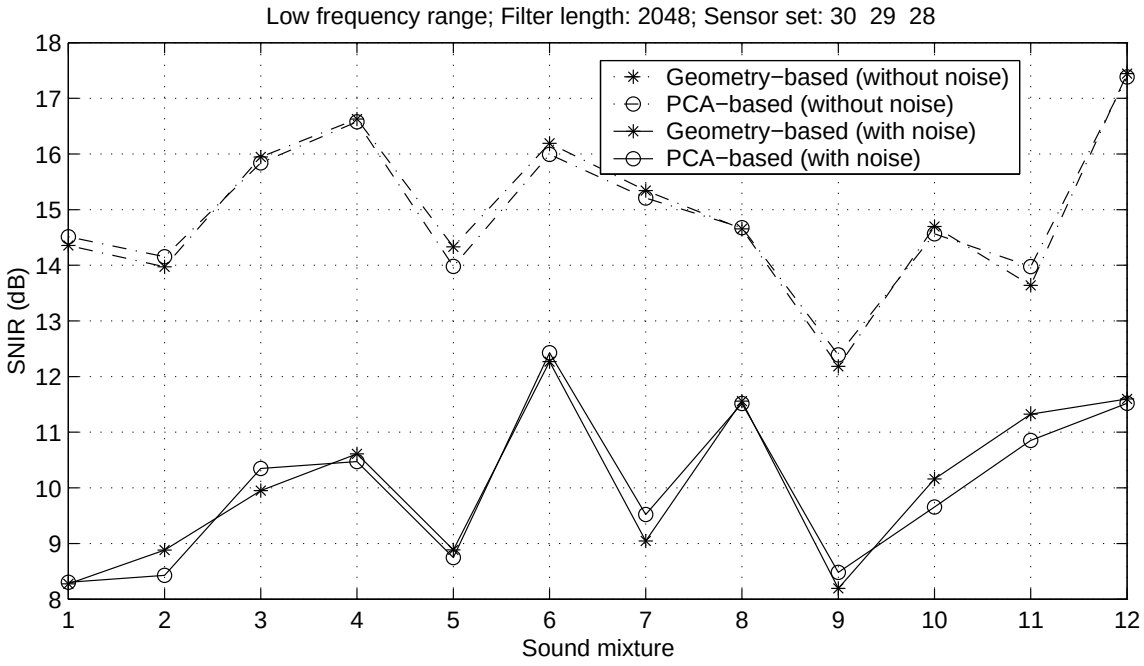

Figure 5.4: Comparison of PCA- and geometry-based subspace selection

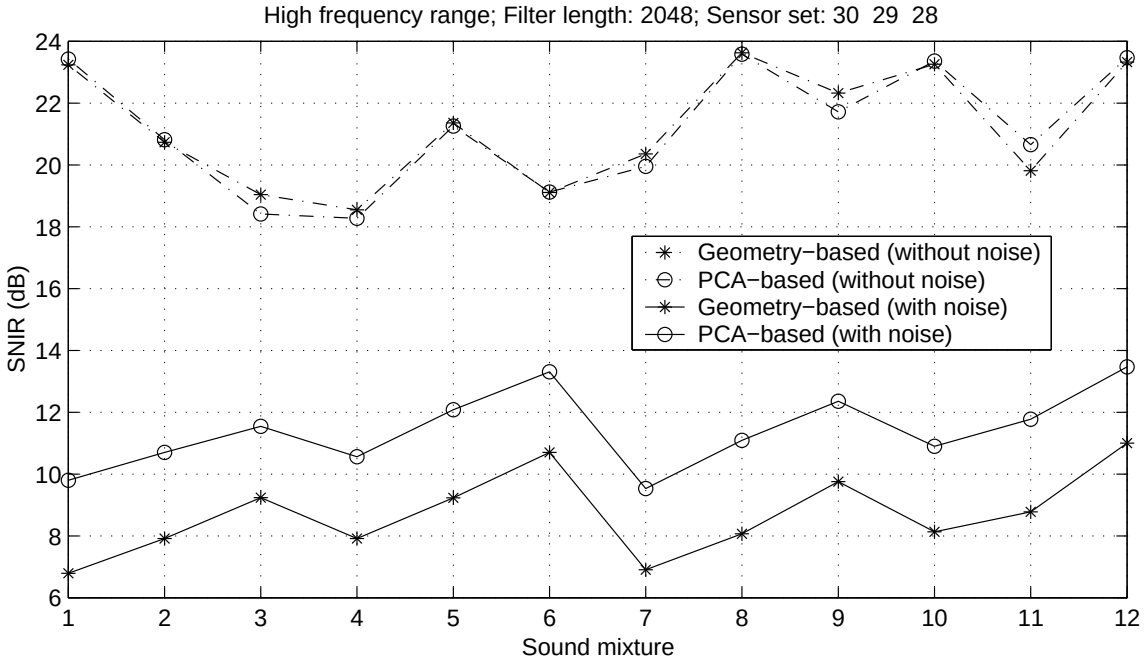

Figure 5.5: Comparison of PCA- and geometry-based subspace selection

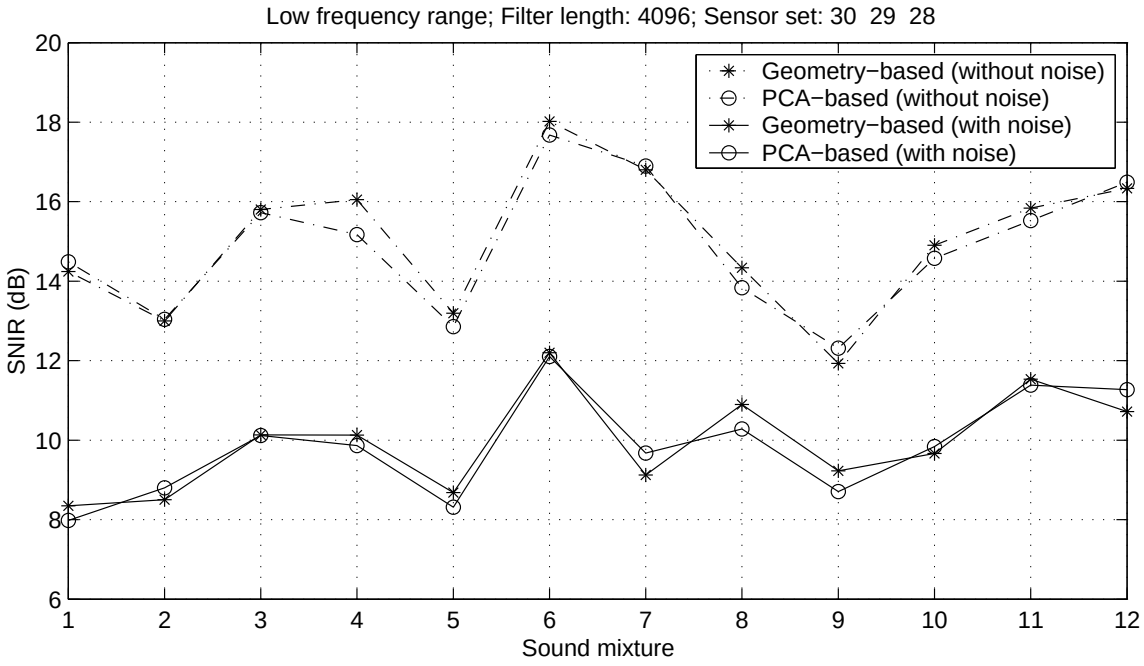

Figure 5.6: Comparison of PCA- and geometry-based subspace selection

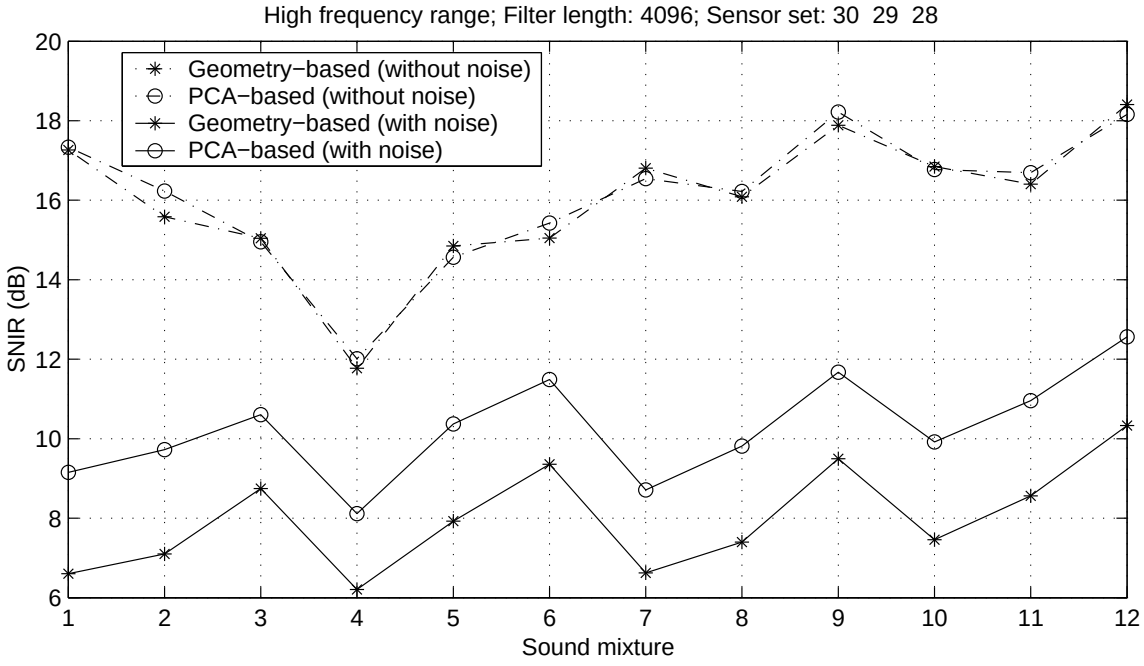

Figure 5.7: Comparison of PCA- and geometry-based subspace selection

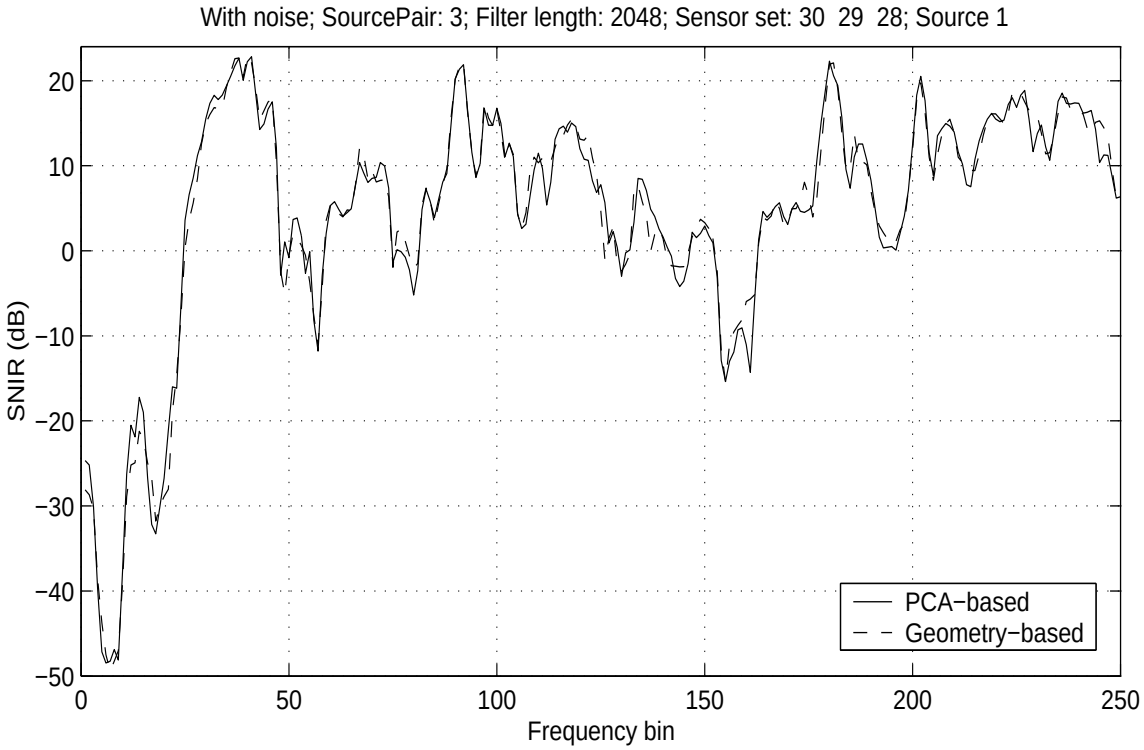

Figure 5.8: Frequency dependent SNIR with added noise

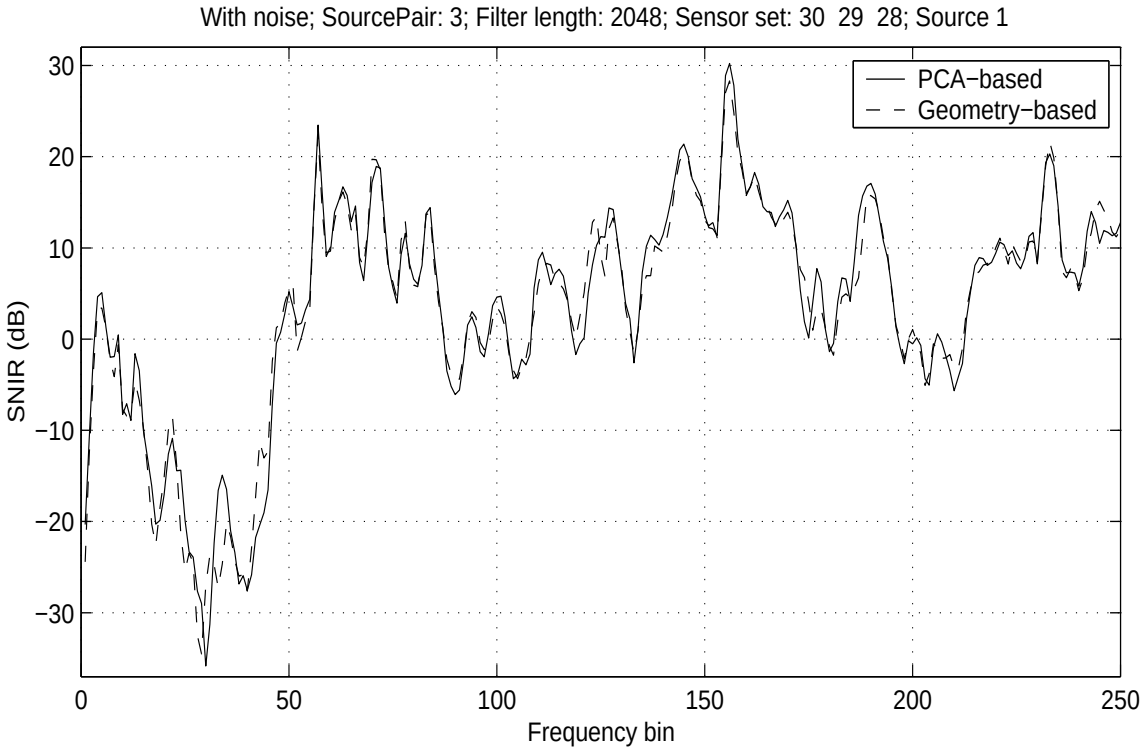

Figure 5.9: Frequency dependent SNIR with added noise

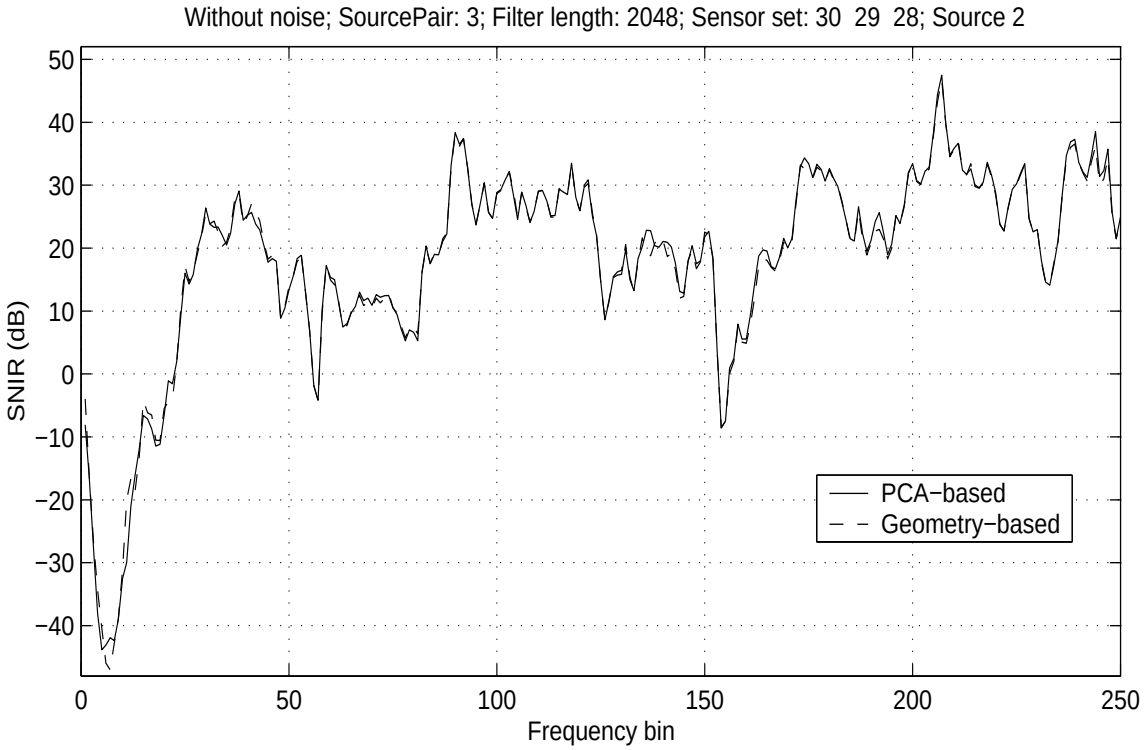

Figure 5.10: Frequency dependent SNIR without additional noise

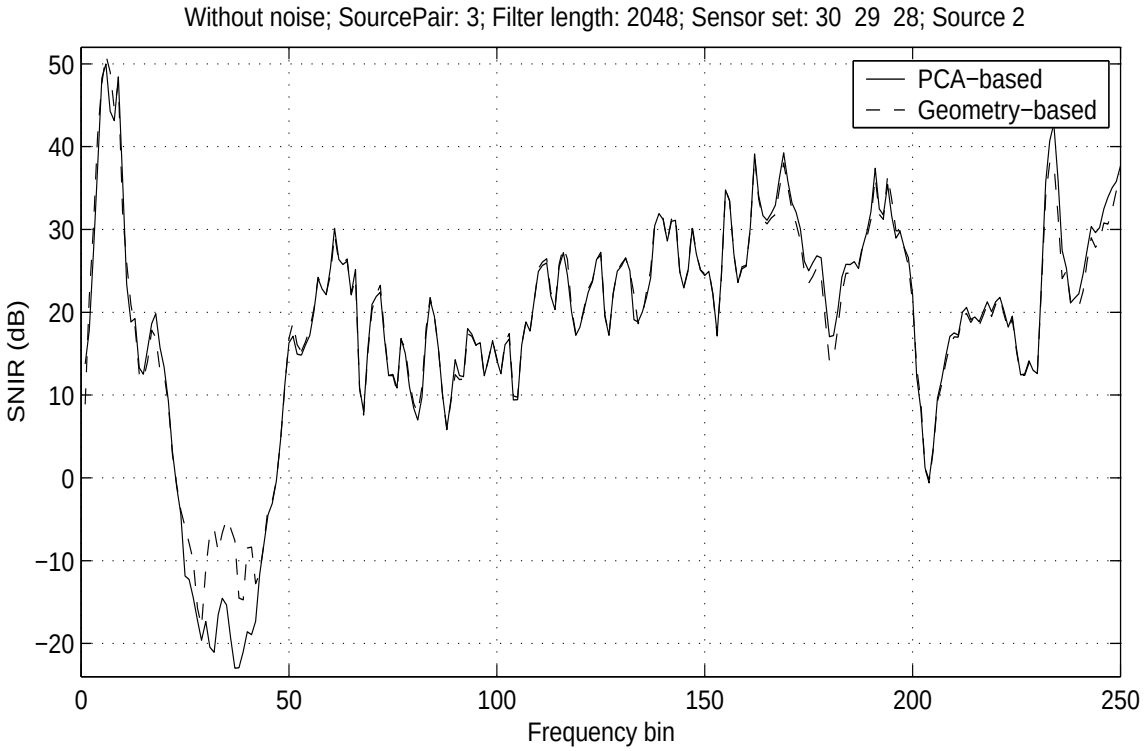

Figure 5.11: Frequency dependent SNIR without additional noise

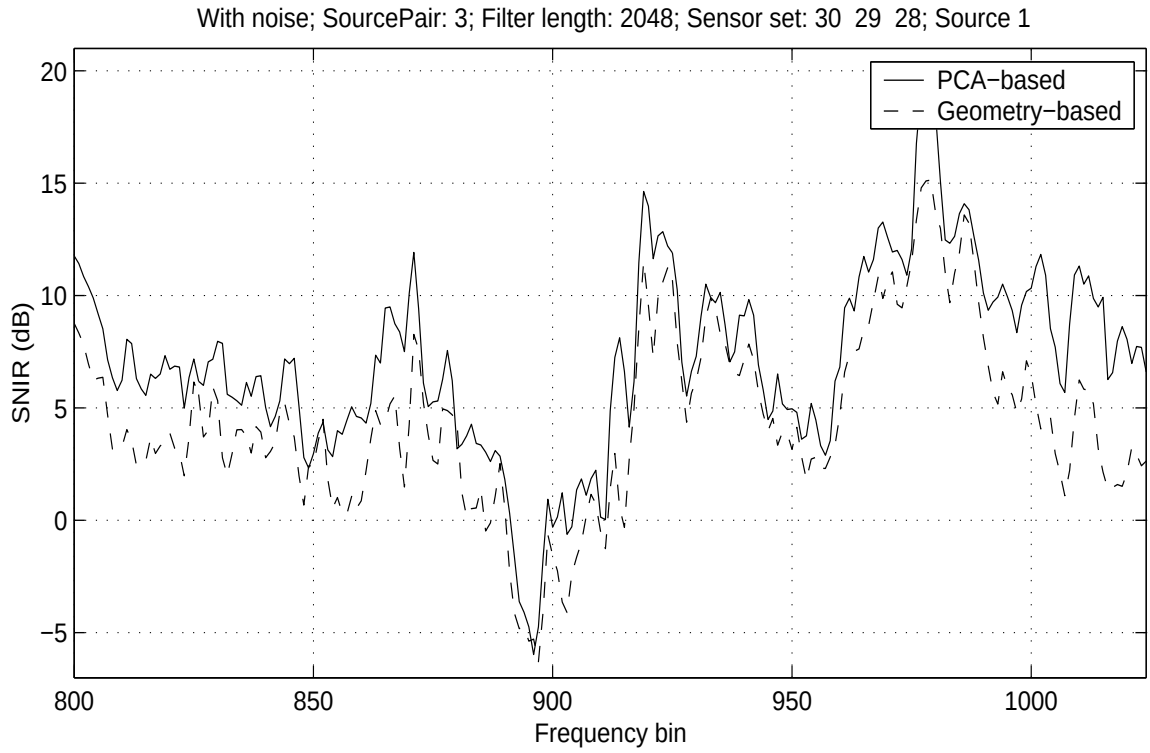

Figure 5.12: Frequency dependent SNIR with added noise

These figures again evidence the similar behavior of the PCA- and geometry-based approaches for low frequencies. Their behavior is most similar over a wide range if we account only for reverberation. However, in Fig. 5.11 we can see a great failure in the frequency bins number 25 to 40. If noise is added both methods still perform very similarly for low frequencies.

In Figs. 5.12-5.15 the performance of the same signals as in Figs. 5.8-5.11 is depicted for the high frequency range. We can clearly see that the PCA-based approach is advantageous in the high frequency range if noise is added. But again we encounter a great failure in Fig. 5.15 in the frequency bins above 1005.

5.2 Interpretation of experimental results

To interprete the experimental results of Sec. 5.1 we distinguish between noiseless and noisy cases as well as between frequency ranges.

As stated in Sec. 3.2 uncorrelated noise is normally reduced if we coherently add up the mixtures received at several sensors. While the PCA-based method is in

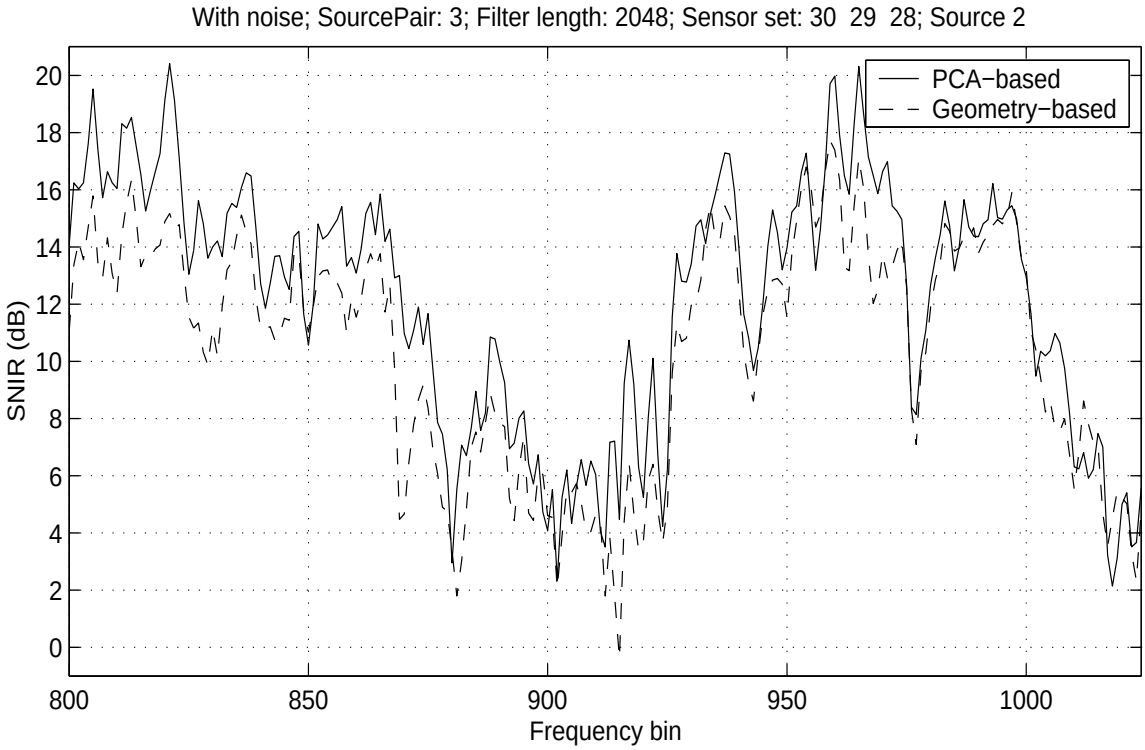

Figure 5.13: Frequency dependent SNIR with added noise

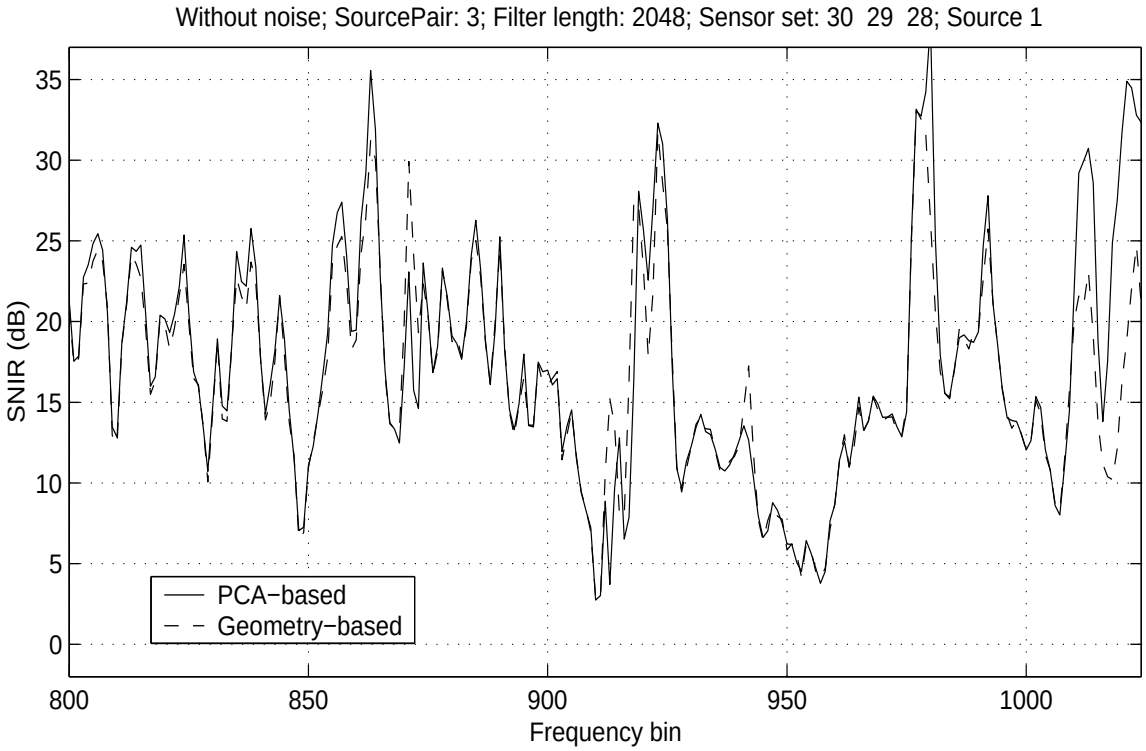

Figure 5.14: Frequency dependent SNIR without additional noise

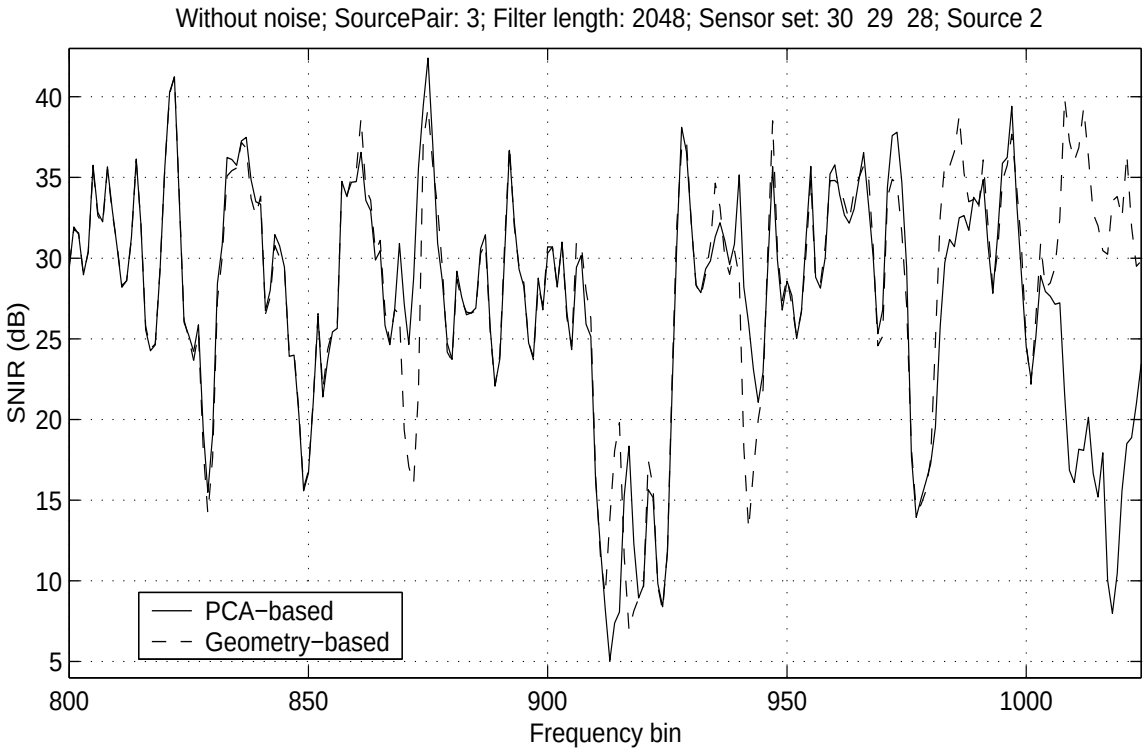

Figure 5.15: Frequency dependent SNIR without additional noise

general capable of utilizing all available sensors, the geometry-based approach by definition always uses only two sensors. Thus the latter cannot exploit the noise reduction to the same degree as the PCA-based approach. However, as seen in Sec. 4.1 the PCA based method also emphasizes the outer sensors for low frequencies. This normally provides the highest possible phase difference for low frequencies, which is important for correctly separating the mixed signals with 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 encounter problems when trying to find appropriate appropriate principal components due to low phase differences which are disturbed by noise. Thus the PCA-based approach cannot make great use of the remaining sensor to reduce noise either and therefore does not improve the performance for low frequencies.

In contrast, for high frequencies a smaller sensor distance is appropriate. Section 4.2.2 showed that the behavior of the PCA-based approach also resembles the behavior of the geometry-based approach for high frequencies by automatically selecting the small sensor spacing. However, the PCA-based approach can utilize all available sensor pairs with small a spacing whereas the geometry-based method still allows only the contribution of one sensor pair. Therefore, while the PCA-based approach

follows geometric considerations it is also capable of effectively reducing noise.

In the noiseless case the fact that the PCA-based subspace approach follows geometric considerations both in the low and high frequency range remains unchanged. Since the PCA-based approach cannot reduce the noise for the low frequency range it does not matter whether there is noise or not in the low frequency range with regard to the choice of the subspace method. In contrast, the advantage of the noise suppression provided by the PCA-based method for the high frequency range has no effect in the noiseless case and therefore does not improve the result.

The exact reasons for the failures in Figs. 5.11 and 5.15 remains unclear and needs further investigation. One possible reason could be that the geometry-based approach is more robust against outliers than the PCA-based approach. The sensor distance is not influenced by the characteristics of the signals and therefore gives a clear basis for the geometry-based approach to select the subspace. In contrast, the PCA-based approach relies on statistical characteristics of the signals. They can change and are not as certain as the sensor spacing.

Chapter 6

Summary and conclusion

In this thesis we investigated the problem of BSS for convolutive mixtures of speech signals. We first explained how we can estimate source signals from their instantaneous mixtures if the original source signals are mutually independent and non-Gaussian. In particular we considered the FastICA algorithm proposed in [6]. We then pointed out how we can reduce the problem of BSS for convolutive mixtures to instantaneous mixtures if we switch from the time domain to the frequency domain. We then specifically addressed the problem of overdetermined BSS and showed how it can be narrowed down to critically-determined BSS if we employ a subspace preprocessing stage. We found that for FastICA it is most advantageous if we undertake the subspace processing before we separate the mixtures.

We have compared two subspace approaches both experimentally and analytically. We found that for low frequencies the PCA-based method exhibits a similar performance to the geometry-based method because it automatically also emphasizes the outer sensors with a larger spacing. For high frequencies the PCA-based approach performs better when exposed to noisy speech mixtures because due to an appropriate phase difference it can utilize all pairs of sensors to suppress the noise. These results deepen our understanding of the PCA-based method from a geometrical point of view.

Further investigations should include the comparison of the algorithms described in this thesis with non-holonomic algorithms.

Appendix A

Derivation of sensor selection by PCA for low frequencies

A.1 Definitions and assumptions

We assume a mixing system with 2 sources and 3 equispaced sensors with distance d. The sensors are supposed to be linearly aligned and consecutively numbered. The source signals are given by

$$\mathbf{S} := \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} \tag{A.1}$$

According to Sec. 4.2, if the first sensor serves as reference point, the frequency dependent mixing matrix can be written as

$$\mathbf{H}(f) = \begin{bmatrix} c_1 & c_2 \\ c_1 e^{j\omega_1} & c_2 e^{j\omega_2} \\ c_1 e^{j2\omega_1} & c_2 e^{j2\omega_2} \end{bmatrix} \tag{A.2}$$

where

$$\omega_i = \frac{2\pi f d \cos\theta_i}{c} \tag{A.3}$$

θ_i denotes the DOA of source number i and c the sound velocity. Then we obtain the mixed signals $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{HS} = \mathbf{H} \begin{bmatrix} S_1 \\ S_2 \end{bmatrix} = \begin{bmatrix} c_1 S_1 + c_2 S_2 \\ c_1 e^{j\omega_1} S_1 + c_2 e^{j\omega_2} S_2 \\ c_1 e^{j2\omega_1} S_1 + c_2 e^{j2\omega_2} S_2 \end{bmatrix} \tag{A.4}$$

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

$$\mathbf{p} := \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} a_1 + jb_1 \\ a_2 + jb_2 \\ a_3 + jb_3 \end{bmatrix} \tag{A.5}$$

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

$$\begin{aligned} \mathbf{p}^H\mathbf{X} = \mathbf{p}^H\mathbf{HS} = c_1S_1(p_1^* + p_2^*e^{j\omega_1} + p_3^*e^{j2\omega_1}) + \\ c_2S_2(p_1^* + p_2^*e^{j\omega_2} + p_3^*\ e^{j2\omega_2}) \end{aligned} \tag{A.6}$$

For low frequencies the phase difference ω_i becomes very small and we can approximate the phase $j\omega_i,\ (0 \leq j \leq M-1)$ by the least square error (LSE) solution $\bar{\omega}_i$ of

$$\begin{aligned} \min_{\bar{\omega}_i} ||(p_1^* + p_2^*e^{j\omega_i} + p_3^*e^{j2\omega_i}) - \\ (p_1^*e^{j\bar{\omega}_i} + p_2^*e^{j\bar{\omega}_i} + p_3^*e^{j\bar{\omega}_i})|| = \\ = \min_{\bar{\omega}_i} ||((p_1^* + p_2^*e^{j\omega_i} + p_3^*e^{2j\omega_i}) - (p_1^*e^{j\bar{\omega}_i} + p_2^*e^{j\bar{\omega}_i} + p_3^*e^{j\bar{\omega}_i}))|| \\ = \min_{\bar{\omega}_i} ||((p_1^* + p_2^*e^{j\omega_i} + p_3^*e^{2j\omega_i}) - e^{j\bar{\omega}_i}(p_1^* + p_2^* + p_3^*))|| \end{aligned} \tag{A.7}$$

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

$$\mathbf{p}^H\mathbf{X} \approx (c_1e^{j\bar{\omega}_1}S_1 + c_2e^{j\bar{\omega}_2}S_2)(p_1^* + p_2^* + p_3^*) \tag{A.8}$$

A.2 Derivation of first principal component

The first principal component is found by maximizing the power

$$E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\} \tag{A.9}$$

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

$$\max_{\mathbf{p}} E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\}, \qquad ||\mathbf{p}|| = 1 \tag{A.10}$$

We define:

$$\begin{aligned} \bar{x^2} &:= E\{(c_1e^{j\bar{\omega}_1}S_1 + c_2e^{j\bar{\omega}_2}S_2) \\ &\quad (c_1e^{j\bar{\omega}_1}S_1 + c_2e^{j\bar{\omega}_2}S_2)^*\} \qquad \text{(A.11)} \\ \gamma &:= \delta \cdot \bar{x^2}, \qquad \delta \text{ variable (modified Lagrange multiplier)} \qquad \text{(A.12)} \end{aligned}$$

and utilize the Lagrange multipliers approach [8]:364:

$$\nabla \left(E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\} + \gamma(||\mathbf{p}|| - 1) \right) = \mathbf{0} \tag{A.13}$$

$$\begin{aligned} E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\} &\approx E\{((c_1 e^{j\bar{\omega}_1} S_1 + c_2 e^{j\bar{\omega}_2} S_2)(p_1^* + p_2^* + p_3^*) \\ &\quad (c_1 e^{j\bar{\omega}_1} S_1 + c_2 e^{j\bar{\omega}_2} S_2)^*(p_1 + p_2 + p_3)\} \\ &= (\underbrace{p_1 p_1^* + p_2 p_2^* + p_3 p_3^*}_{=||\mathbf{p}||=1} + p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2)) \cdot \bar{x^2} \\ &= (1 + p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2)) \cdot \bar{x^2} \end{aligned} \tag{A.14}$$

$\Rightarrow$

$$\begin{aligned} &\nabla \left(E\{(\mathbf{p}^H\mathbf{X})(\mathbf{p}^H\mathbf{X})^*\} + \gamma(||\mathbf{p}|| - 1) \right) = \\ &= \nabla (1 + p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2)) \cdot \bar{x^2} + \delta \cdot \bar{x^2}(||\mathbf{p}|| - 1) \\ &= \underbrace{\nabla(\bar{x^2} - \delta \cdot \bar{x^2})}_{=\mathbf{0}} + \bar{x^2} \nabla(p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2) + \delta||\mathbf{p}||) = \mathbf{0} \end{aligned}$$

$\Rightarrow$

$$\nabla(p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2) + \delta||\mathbf{p}||) = \mathbf{0} \tag{A.15}$$

If we use $p_i = a_i + jb_i$ we obtain

$$\begin{aligned} &\nabla(p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2) + \delta||\mathbf{p}||) = \\ &= \nabla((a_1 - jb_1)(a_2 + jb_2 + a_3 + jb_3) + (a_2 - jb_2)(a_1 + jb_1 + a_3 + jb_3) + \\ &\quad (a_3 - jb_3)(a_1 + jb_1 + a_2 + jb_2) + \delta(a_1^2 + b_1^2 + a_2^2 + b_2^2 + a_3^2 + b_3^2)) \\ &= \begin{bmatrix} 2(\delta a_1 + a_2 + a_3) \\ 2(a_1 + \delta a_2 + a_3) \\ 2(a_1 + a_2 + \delta a_3) \\ 2(\delta b_1 + b_2 + b_3) \\ 2(b_1 + \delta b_2 + b_3) \\ 2(b_1 + b_2 + \delta b_3) \end{bmatrix} \\ &= 2 \begin{bmatrix} \delta & 1 & 1 & & & \\ 1 & \delta & 1 & & \mathbf{0} & \\ 1 & 1 & \delta & & & \\ & & & \delta & 1 & 1 \\ & \mathbf{0} & & 1 & \delta & 1 \\ & & & 1 & 1 & \delta \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \mathbf{0} \end{aligned} \tag{A.16}$$

$\Rightarrow$ Basically we have to solve two independent subsystems:

$$\begin{array}{ccc|c}
\delta & 1 & 1 & 0 \\
1 & \delta & 1 & 0 \\
1 & 1 & \delta & 0 \\
\hline
\delta & 1 & 1 & 0 \\
0 & \delta - \frac{1}{\delta} & 1 - \frac{1}{\delta} & 0 \\
0 & 1 - \frac{1}{\delta} & \delta - \frac{1}{\delta} & 0 \\
\hline
\delta & 1 & 1 & 0 \\
0 & \delta - \frac{1}{\delta} & 1 - \frac{1}{\delta} & 0 \\
0 & 0 & \frac{\delta^2+\delta-2}{\delta+1} & 0
\end{array} \tag{A.17}$$

We get a non-trivial solution if and only if

$$\frac{\delta^2 + \delta - 2}{\delta + 1} = 0 \quad \Rightarrow \quad \delta_1 = -2,\ \delta_2 = 1 \tag{A.18}$$

Solution for $\delta = \delta_1 = 1$:

$$a_1 = -(a_2 + a_3); \qquad a_2, a_3 \text{ variable} \tag{A.19}$$

$$b_1 = -(b_2 + b_3); \qquad b_2, b_3 \text{ variable} \tag{A.20}$$

$$\Rightarrow p_1 = -(p_2 + p_3); \qquad p_2, p_3 \text{ variable} \tag{A.21}$$

Inserting in Eq. (A.14) yields:

$$\begin{aligned}
&(1 + p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2)) \cdot \bar{x^2} = \\
&= (1 - p_1^* p_1 + p_2^*(-p_2 - p_3 + p_3) + p_3^*(-p_2 - p_3 + p_2)) \cdot \bar{x^2} \\
&= (1 - \underbrace{(p_1^* p_1 + p_2^* p_2 + p_3^* p_3)}_{=||\mathbf{p}||=1}) \cdot \bar{x^2} \\
&= 0
\end{aligned} \tag{A.22}$$

Solution for $\delta = \delta_2 = -2$:

$$a_1 = a_2 = a_3 = a \tag{A.23}$$

$$b_1 = b_2 = b_3 = b \tag{A.24}$$

$$p_1 = p_2 = p_3 = p = a + jb \tag{A.25}$$

Let

$$a + jb =: |p| e^{j\phi} \tag{A.26}$$

Then it follows with $||\mathbf{p}|| = 3|p|^2 = 1$

$$|p| = \pm\frac{1}{\sqrt{3}} \tag{A.27}$$

ϕ can be chosen arbitrarily

Inserting in Eq. (A.14) yields:

$$\begin{aligned}
&(1 + p_1^*(p_2 + p_3) + p_2^*(p_1 + p_3) + p_3^*(p_1 + p_2)) \cdot \bar{x^2} = \\
&= (1 + p^*(p + p) + p^*(p + p) + p^*(p + p)) \cdot \bar{x^2} \\
&= (1 + 3p^*(p + p)) = (1 + 6p^*p) \cdot \bar{x^2} = (1 + 6|p|^2) \cdot \bar{x^2} \\
&= 3 \cdot \bar{x^2} \geq 0
\end{aligned} \tag{A.28}$$

$\Rightarrow$ We obtain the maximum for $\delta = \delta_2 = -2$ and $p_1 = p_2 = p_3 = p$

A.3 Approximation of phase difference

With $p_1 = p_2 = p_3 = p$ and Eq. (A.7) we can now determine the approximation of the the phase. The minimum of Eq. (A.7) can be found by

$$\frac{d}{d\bar{\omega_i}}||(p_1^* + p_2^* e^{j\omega_i} + p_3^* e^{2j\omega_i}) - e^{j\bar{\omega}_i}(p_1^* + p_2^* + p_3^*)|| = 0 \tag{A.29}$$

We obtain for the first derivative:

$$\begin{aligned}
&\frac{d}{d\bar{\omega_i}}||(p_1^* + p_2^* e^{j\omega_i} + p_3^* e^{2j\omega_i}) - e^{j\bar{\omega}_i}(p_1^* + p_2^* + p_3^*)|| = \\
&= \frac{d}{d\bar{\omega_i}}||p^*((1 + e^{j\omega_i} + e^{2j\omega_i}) - 3e^{j\bar{\omega}_i})|| = \\
&= \frac{d}{d\bar{\omega_i}}(pp^*((1 + e^{j\omega_i} + e^{2j\omega_i}) - 3e^{j\bar{\omega}_i}) \cdot \\
&\quad ((1 + e^{-j\omega_i} + e^{-2j\omega_i}) - 3e^{-j\bar{\omega}_i})) \\
&= \frac{1}{3}(\underbrace{\frac{d}{d\bar{\omega_i}}((1 + e^{-j\omega_i} + e^{-2j\omega_i})(1 + e^{j\omega_i} + e^{2j\omega_i})}_{=0} + \\
&\quad \underbrace{\frac{d}{d\bar{\omega_i}}(3e^{-j\bar{\omega}_i} \cdot 3e^{j\bar{\omega}_i})}_{=0} - \\
&\quad \frac{d}{d\bar{\omega_i}}(3e^{-j\bar{\omega}_i}(1 + e^{j\omega_i} + e^{2j\omega_i}) + 3e^{j\bar{\omega}_i}(1 + e^{-j\omega_i} + e^{-2j\omega_i})) \\
&= je^{-j\bar{\omega}_i}(1 + e^{j\omega_i} + e^{2j\omega_i}) - je^{j\bar{\omega}_i}(1 + e^{-j\omega_i} + e^{-2j\omega_i})
\end{aligned} \tag{A.30}$$

The second derivative is given by

$$\begin{aligned}
&\frac{d}{d^2\bar{\omega_i}}||p^*(1+e^{j\omega_i}+e^{2j\omega_i})-3e^{j\bar{\omega}_i}|| = \\
&= \frac{d}{d\bar{\omega_i}}(je^{-j\bar{\omega}_i}(1+e^{j\omega_i}+e^{2j\omega_i})-je^{j\bar{\omega}_i}(1+e^{-j\omega_i}+e^{-2j\omega_i})) \\
&= e^{-j\bar{\omega}_i}(1+e^{j\omega_i}+e^{2j\omega_i})+e^{j\bar{\omega}_i}(1+e^{-j\omega_i}+e^{-2j\omega_i}) \qquad \text{(A.31)}
\end{aligned}$$

For $\bar{\omega}_i = \omega_i$ we get

$$\begin{aligned}
&\frac{d}{d\bar{\omega_i}}||p^*(1+e^{j\omega_i}+e^{2j\omega_i})-3e^{j\bar{\omega}_i}|| = \\
&= je^{-j\bar{\omega}_i}(1+e^{j\omega_i}+e^{2j\omega_i})-je^{j\bar{\omega}_i}(1+e^{-j\omega_i}+e^{-2j\omega_i}) \\
&= j(e^{-j\omega_i}+1+e^{j\omega_i})-j(e^{j\omega_i}+1+e^{-j\omega_i}) \\
&= 0 \qquad \text{(A.32)}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{d}{d^2\bar{\omega_i}}||p^*(1+e^{j\omega_i}+e^{2j\omega_i})-3e^{j\bar{\omega}_i}|| = \\
&= e^{-j\bar{\omega}_i}(1+e^{j\omega_i}+e^{2j\omega_i})+e^{j\bar{\omega}_i}(1+e^{-j\omega_i}+e^{-2j\omega_i}) \\
&= (e^{-j\omega_i}+1+e^{j\omega_i})+(e^{j\omega_i}+1+e^{-j\omega_i}) \\
&= 2+2\cos(\omega_i)+2\cos(-\omega_i) \\
&= 2+4\cos(\omega_i) \\
&\overset{|\omega_i|\ll 1}{>} 0 \qquad \text{(A.33)}
\end{aligned}$$

Thus $\bar{\omega}_i = \omega_i$ is a LSE solution of Eq. (A.7) for $p_1 = p_2 = p_3 = p$.

Bibliography

[1] R. Aichner. Time domain blind source separation of non-stationary convolved signals with utilization of geometric beamforming. Diploma thesis, 2002.

[2] S. Amari, T.P. Chen, and A. Cichocki. Nonholonomic orthogonal learning algorithm for blind source separation. *Neural Computation*, 12(6):1463–1484, 2000.

[3] S. Araki, S. Makino, R. Mukai, and H. Saruwatari. Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers. In *Proc. Eurospeech2001*, pages 2595–2598, Sept. 2001.

[4] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki. A combined approach of array processing and independent component analysis for blind separation of acoustic signals. In *Proc. ICASSP2001*, May 2001.

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

[6] E. Bingham and A. Hyvärinen. A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, 10(1):1–8, Feb. 2000.

[7] M.S. Brandstein and D.B. Ward, editors. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Verlag, 2001.

[8] Bronstein, Semendjajew, Musiol, and Muehlig. *Taschenbuch der Mathematik*. Harri Deutsch, Frankfurt am Main, 3rd edition, 1997.

[9] D.A. Harville. *Matrix Algebra from a statistician's perspective*. Springer Verlag, 2000.

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

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

[12] S. Ikeda and N. Murata. A method of ICA in time-frequency domain. In *Proc. ICA99*, pages 365–371, Jan. 1999.

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

[14] 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*, pages 81–86, June 2000.

[15] I.T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 2nd edition, Oct. 2002.

[16] W. Kellermann. Fundamentals of digital signal processing I+II. Lecture notes, 2000.

[17] D. Kolossa, B.-U. Koehler, M. Conrath, and R. Orglmeister. Optimal permutation correction by multiobjective genetic algorithms. In *Proc. ICA2001*, 2001.

[18] I. Kopriva, Z. Devcic, and H.H. Szu. An adaptive short-time frequency domain algorithm for blind separation of non-stationary convolved mixtures. In *Proc. of INNS-IEEE Joint Conf. on Neural Networks*, pages 424–429, July 2001.

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

[20] J. Leblanc and P. De Leon. Speech separation by kurtosis maximization. In *Proc. ICASSP*, volume 2, pages 1029–1032, 1998.

[21] P. De Leon and Y. Ma. Blind separation of l sources from m mixtures of speech signals. In *140th Meeting of the Acoustical Society of America*, 2000.

[22] C.D. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial & Applied Mathematics, 2000.

[23] J.-P. Nadal, E. Korutcheva, and F. Aires. Blind source separation in the presence of weak sources. *Neural Networks*, 13(6):589–596, 2000.

[24] A.V. Oppenheim and R.W. Schafer with J.R. Buck. *Discrete-time signal processing*. Prentice Hall, 2nd edition, 1998.

[25] A. Papoulis and S.U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, 4th edition, 2002.

[26] L. Parra and C. Alvino. Geometric source separation: Merging convolutive source separation with geometric beamforming. *IEEE Transaction on Speech and Audio Processing*, 10(6):352–362, Sept. 2002.

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

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

[29] 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)*, pages 465–474, Sept. 2002.

[30] 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.

[31] P. Smaragdis. Blind separation of convolved mixtures in the frequency domain. *International Workshop on Independence & Artificial Neural Networks*, Feb. 1998.

[32] L. Tong, R.-W. Liu, V.C. Soon, and Y.-F. Huang. Indeterminacy and identifiability of blind identification. *IEEE Trans. Circuits Syst.*, 38:499–509, May 1991.

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

[34] L. Vielva, D. Erdogmus, C. Pantaleon, I. Santamaria, J. Pereda, and J.C. Principe. Underdetermined blind source separation in a time-varying environment. In *Proc. ICASSP02*, volume 3, pages 3049–3052, May 2002.

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