Friedrich-Alexander-Universität Erlangen-Nürnberg

Lehrstuhl für Multimediakommunikation und Signalverarbeitung

Prof. Dr.-Ing. Walter Kellermann

Research Internship

Evaluation of Circular Convolution Effects for Independent Vector Analysis

Huajian Fang

September 2018

Betreuer: M.Sc. Thomas Haubner

Erklärung

Ich versichere, dass ich die vorliegende 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.

Ort, Datum

Unterschrift

Contents

Abstract

The blind source separation (BSS) problem refers to the restoration of the original sources from the observed signals on microphones without knowing the mixing process. A statistical signal processing algorithm, i.e. independent component analysis (ICA) and its modified versions were given by some researchers and scholars to separate the mixed signals. After ICA, independent vector analysis (IVA) was proposed to give realatively better performances. It preserves the inherent frequency dependency within each source, while removing the dependency between different sources. The interfrequency dependency in data is captured by defining the source prior as a dependent multivariate super-Gaussian distribution, which means the permutation problem is completely avoided. The performances of the algorithm are influenced by many factors, such as reverberant conditions or a microphone array with a tricky geometric configuration. Since the drawbacks such as computational load and slow convergence for the time domain implementation, the frequency domain approach is efficiently used instead. Observed signals are convolutive mixtrues including delays and convolutions. Since the permutation problem can be sloved by using IVA, another problem, i.e. the ciucular convolution effects, is evaluated here.

Abbreviations

ICA	Independent component analysis
IVA	Independent vector analysis
DFT	Discrete Fourier Transfrom

Symbols

$+$	Addition
$-$	Subtraction
$\cdot$	Multiplication
$/$	Division
$\sum$	Summation
$\circledast$	Circular convolution
$\sqrt{h}$	Square root of h
$\int$	Integration
$\|\|h\|\|$	Absolute value
$h^{\dagger}$	Hermitian conjugate of h

Chapter 1

Introduction

Blind source separation (BSS) aims to separate original source signals from mixed observations at sensors without knowing prior information. The blind separation is an attractive topic and has drawn the attention of many scholars over the last decades. It has wide applications, such as speech enhancement, speech recognition, and crosstalk separation in telecommunication system etc. The reasons for making this problem difficult are manifold, including reverberant conditions, with a large number of microphones, and ill-posed problem. For example, the impulse response of the propagation path for each source is similar. The fundamental assumption is that source signals are statistically independent.

Before the IVA algorithm was proposed, a statistical method, i.e. ICA, had been used. The simplest form of ICA, the linear instantaneous model for the mixing process, is not valid in practice. The modified model, i.e. convolutive mixtures, is used instead because of the reflections from walls and floor. The source signals are mixed with time delays and convolutions before recorded by microphones. The implementation in the time domain brings the problems like computational load and slow convergence because the ICA method for convolutive mixtures is much more complicated than that for instantaneous mixtures. However, in the frequency domain, ICA for instantaneous mixtures can be applied in each frequency bin and simplicity is preserved. The permutation ambiguity over frequency bins inherent in the ICA algorithm becomes a serious problem at the same time.

Another problem, circularity problem, leads to distortion for sources and performance degradation. The detailed explanation of this problem is given in chapter 3. This problem is influenced by factors, such as the length of the room impulse response and the length of the applied window.

A novel approach has been proposed, i.e. IVA, based on ICA, by exploiting higher order frequency dependencies [KALL07]. The fact that the better performances are obtained with the IVA algorithm is claimed even in ill-posed conditions[KALL07]. There are different types of IVA algorithms currently. The auxiliary function based IVA which is utilized for the evaluation here gives more effective convergence and better results than the natural gradient convergence method [Ono11]. Interfrequency dependencies in source signals are modeled by the new source priors instead of independent priors at each frequency bin. A dependent multivariate super-Gaussian is defined for each source prior. Therefore, the dependency of frequencies for each source is preserved, which means that the permutation problem inherent in ICA is theoretically solved. The aforementioned circular convolution of IVA is primarily evaluated here by changing different parameters, such as the reverberation time, the length of the applied window, and the number of iteration times and so on. The target is to evaluate the aliasing effects of the IVA method due to the circular convolution model.

The rest of this report is organized as follows: the next chapter is to explain the source model exploited, and the implementation in the time domain is followed after that. Then the several methods are utilized to evaluate the performance in different scenarios. At last, the discussion and summary of the evaluation are given.

Chapter 2

Source Model

The assumption of independent Laplacian distribution does not specify any directionality, nevertheless, the dependent super-Gaussian distribution gives high dependence among frequency bins. For the purpose of evaluation, the ideal source is created and fed into the system. The dependent super-Gaussian is created by the combination of some other distributions. A source variable of K-dimension in the frequency domain is specified as:

$$s_i = \sqrt{v} \cdot \mathbf{z}_i + \mu_i$$

where the symbol v defines a scalar random variable, $\mathbf{z}_i$ denotes a K-dimensional random variable, and μ is a K-dimensional deterministic variable which is neglected in the generation of the ideal source here for the sake of brevity[KALL07]. The captial letter K here stands for the number of frequency bins. The random variable $\mathbf{z}$ is Gaussian distributed with zero mean and $\sum_i$ covariance matrix

$$p(\mathbf{z}_i) = \alpha_z \exp\left(-\frac{\mathbf{z}_i^{\dagger} \sum_i^{-1} \mathbf{z}_i}{2}\right)$$

The random variable v is supposed to be a kind of Gamma distribution:

$$p(v) = \alpha_v v^{\frac{(K-1)}{2}} \exp\left(-\frac{v}{2}\right)$$

where α_v denotes the normalization factor. With the help of the model above, the distribution of the source variable can be described as :

$$\begin{aligned} p(\mathrm{s}_i) &= \int_0^\infty p(\mathbf{s}_i|v)p(v)dv \\ &= \alpha \exp\left(-\sqrt{(s_i-\mu_i)^\dagger \sum_i^{-1}(s_i-\mu_i)}\right). \end{aligned}$$

As described aforementioned, most signals' frequency dependencies between frequency bins are captured. Ideal sources are created in the time-frequency domain, then the parameters, such as the window's name and length and the hopsize, are the same as the ones used for real speech signals, such that the difference between those two could be observed. The parameters of a Gamma distribution can be defined according to the number of frequency bins and the defined Gamma distribution formula given above. For the generation of the ideal source, the correlated covariace matrix is utilized instead of the identity matrix introduced in IVA. The random correlated covariance specified is supposed to be positive definite. In the time-frequency domain, the covariance matrix defined over K dimentions is different for each time block. The signal in the time-frequency domain then can be converted into the time domain and fed into the system. The way of how to obtain the plots shown below will be given in the Chapter 3. In the figure 1, the demixng filter in the time domain is shown when the ideal source of ten seconds long is fed into the system. The layout for the microphones and loudspeakers is shown in figure 3. The x-axis is the time axis but converted into a sample index axis with a sampling frequency of 16 kHz. The y-axis shows the amplitude of the demixing filter. The reverberation of the room is defined as $0.4s$. The parameter called the number of iteration is used here to stop the alogrithm when the number of the iteration is reached. The algorithm is also stopped when the cost decrement is lower than cost decrement threshold. The number of iteration times is specified with 30 in the scenario of figure 1. The length of the room impulse response is longer than that of the demixing filter, and they are specified with 6400 and 4096 samples repectively. In the figure 2,

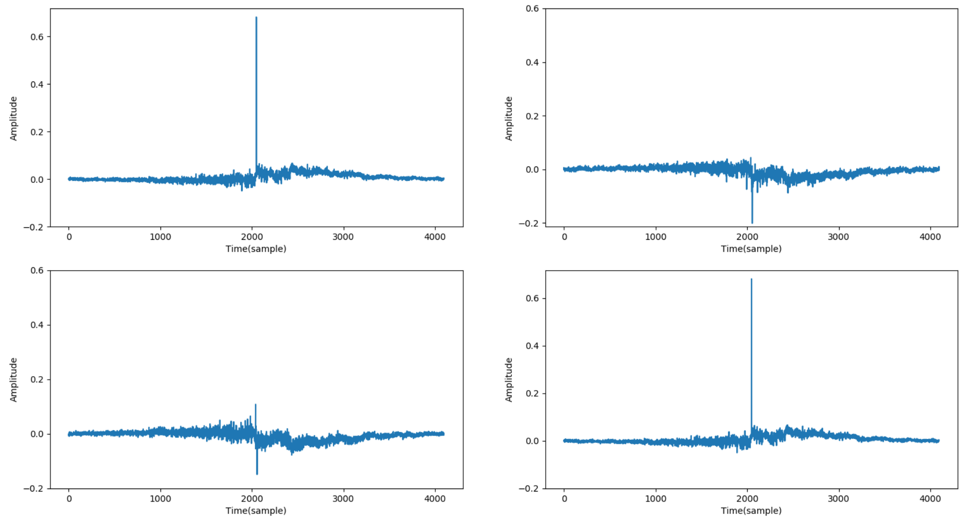

Figure.1. Demixing filter of the ideal source by the finite-length filter with the room impulse response of length 6400 and demixing filter of length 4096.

four overall impulse responses can be seen. The x-axis is also the sample index axis, and y-axis also indicates the amplitude value. Two subplots on the diagonal are desired and the other two plots on the anti-diagonal correspond to interference terms.The way how to interpret these four subplots is given in the next chapter. The overall impulse response plot of the ideal sources obtained by the finite-length filter shows no spikes around the ends.

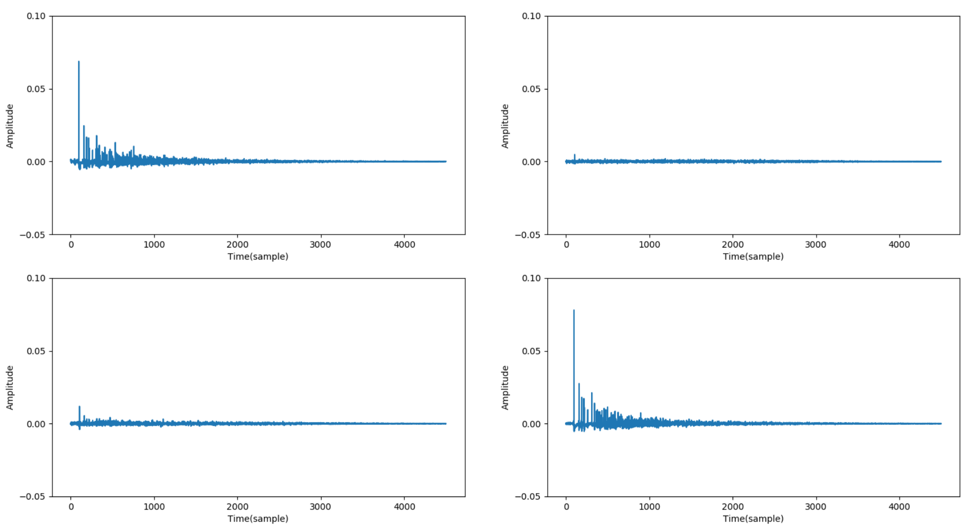

Figure.2. Overall impulse responses of the ideal source obtained by the finite-length filter with the room impulse response of the length 6400 and the demixing filter of the length 4096.

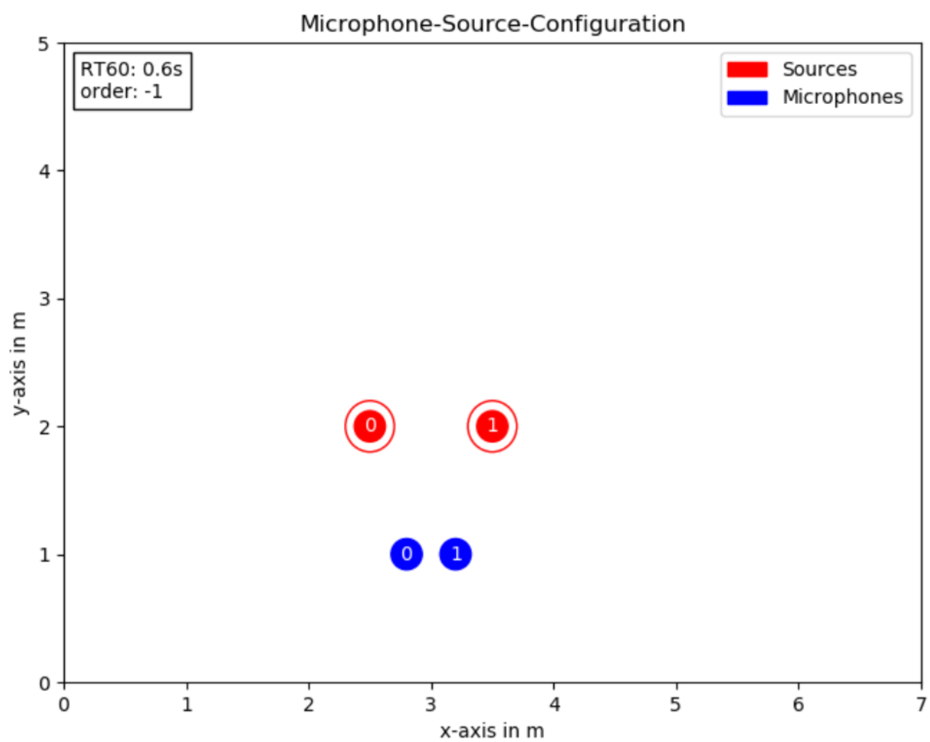

Figure.3. The layout of the microphones and sources.

Chapter 3

Demixing Model

The separation implemented in the time domain is much more numerically complex and slower. This problem can be overcomed in the frequency domain, in which the multiplicaiton in the frequency domian corresponds to the circular convolution in the time domain. The circular covolution differs from the linear one. It is known that the circular convolution can be approximated as the linear convolution with aliasing. IVA models the demixing system as multiplication in the frequency domain.

If two sequences $x_1[n]$ and $x_2[n]$ of length L and P are assumed respectively, the maximum length of the resultant sequence of the linear convolution is $L + P - 1$. The length of the DFT coefficients and its relation with the sum of lengths of the finite-length sequences decide whether a circular convolution is the same as the linear convolution. If the length of the DFT coefficients N is greater than $L + P - 1$, the circular convolution which corresponds to the multiplication of the DFT coefficients with length N is approximated exactly by the linear convolution. The aliasing effects can be avoided by padding sufficient zeros at the end of the time-domain filter.

The procudures for the separation model are that sources convolve with the room impulse responses and excute short-time Fourier transform, then the frequency domain singals are obtained. If the length of the window applied in short-time Fourier transform is sufficiently longer than that of the mixing filter, the convolution in the time domain can be approximated by the multiplication in the frequency domain, which is described by

$$x_i^{(k)}[n] \approx \sum_{j=1}^{L} h_{ij}^{(k)} s_j^{(k)}[n]$$

where index $k = 1, 2, ..., K$ denotes the kth frequency bin. The demixing step is finally given

$$\hat{s}_j^{(k)}[n] = \sum_{i=1}^{M} g_{ji}^{(k)} x_i^{(k)}[n] \approx s_j^{(k)}[n]$$

where $g_{ji}^{(k)}$ is the separation filter at the kth frequency bin, and M is the number of observed signals.

The circularity problem in the IVA algorithm refers to the fact that the interval at which the DTFT is sampled is the reciprocal of the duration of the input sequence [SMK+03]. Ideally, the demixig filter is infinitely periodic while converting the filter from the frequency domain into the time domain. However, the one-period realization is utilized for the realistic usage. The overall impulse response from the source s_j to the estimated source $\hat{s}_{j'}$ is given by

$$u_{j'j}(l) = \sum_{j=1}^{L} \sum_{\tau=0}^{T-1} g_{j'i}(\tau) h_{ij}(l - \tau)$$

where $g_{j'i}$ denotes the demixing filter [SMK+03]. In figure 4, it shows an example for four overall impulse reponses with two sources and two mircrophones, which is obtained by convolving the room impulse responses in this scenario with the demixing filters obtained from IVA. The layout for the microphones and loudspeakers is the same as shown in figure 3. The reverberation time here is set as 0.4s. The number of iteration is specified with 30. The length of the room impulse response is 6400 and the length of the demixing filter is 4096. The extraction of the source s_1 is expressed as

$$\hat{s}_1 = (g_{11}h_{11} + g_{12}h_{21})s_1 + (g_{11}h_{12} + g_{12}h_{22})s_2.$$

Convolution is done between the filters. The ideal case is to minimize the interference from the source s_2, which means to set ideally the part $(g_{11}h_{12} + g_{12}h_{22})$ to 0. The top left subplot $u_{00}(l)$ and right bottom subplot $u_{11}(l)$, which are $(g_{11}h_{11} + g_{12}h_{21})$ and $(g_{21}h_{12} + g_{22}h_{22})$ repectively, correspond to the extraction of the desired source signals, and the other two, which are $(g_{11}h_{12} + g_{12}h_{22})$ and $(g_{21}h_{11} + g_{22}h_{21})$, correspond to the suppression of the interference signals. Figure 5 gives the overall impulse responses

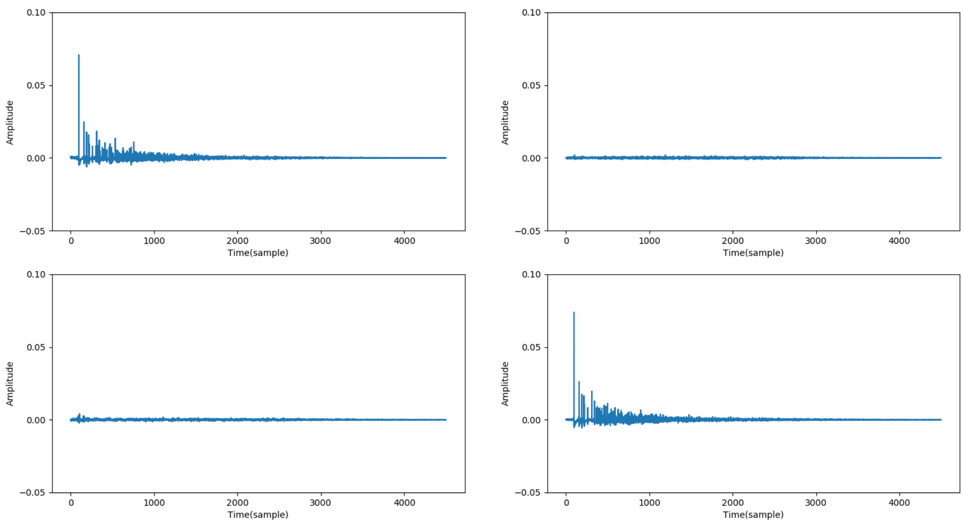

Figure.4. Overall impulse reponses obtained by the finite-length filter with the room impulse response of the length 6400 and the demixing filter of the length 4096.

obtained by the periodical filter for the real input souce signals. The periodical filter does not create spikes. However, there might be two spikes around the two ends of the one-period filter in the case of the one-period realization as shown in the ICA paper [SMK+03]. The spikes have an adverse influence on the estimated source signals. The reasons include that the frequency responses are undersampled and the corresponding time-domain filter overlaps with the adjacent period, which means its one-period realization is the result of the overlap. Therefore, the frequency responses should be controlled, such that the corresponding time-domain filter contians enough zeros at the end. The other reason is that adjacent periods work together to perform some filtering. From the subplots in figure 4, there are no spikes around the two ends, which roughly means there is no circularity effects in the IVA algorithm under this parameters combination.

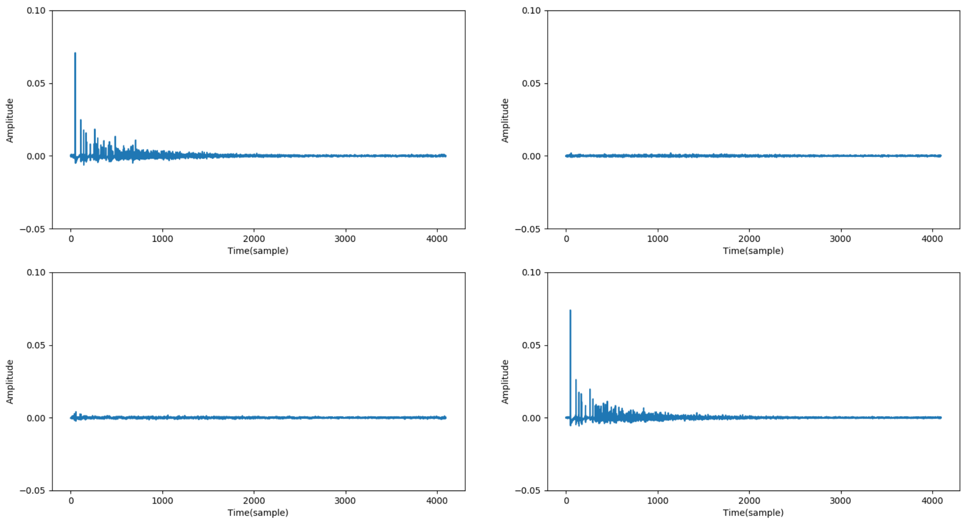

Figure.5. Overall impulse responses obtained by the periodical filter with the room impulse response of the length 6400 and the demixing filter of the length 4096.

Chapter 4

Performace Measures

There are various ways to evaluate the performance of the algorithm. In this chapter, some of them are chosen to evaluate the quality of separated source signals. The evaluation of the performance is either with the help of both real data and estimated data or working with mixing and demixing filter. For example, the BSS Eval method proposed by Emmanuel Vincent [VGF06] only exploits the source and the estimated signals without knowing the mixing or demixing models in advance. Depending on the purpose of the evaluation, the method used by Hiroshi Sawada [SMK$^+$03] can be evolved into two types, signal-dependent method or signal-independent one. They will all be introduced afterward. The criteria such as signal-to-interference ratio(SIR) and signal-to-distortion ratio(SDR) are primarily used here. The higher SIR or SDR value is, the less interference or overall distortion there is in the separated signal. In the rest of this chapter, some measurement methods will be introduced.

4.1 Performance Eval

In a paper proposed by Hiroshi Sawada and other scholars, they exploit the impulse responses of the mixing process and the separated filters to evaluate the separation. As introduced in the previous section, the overall impulse responses are given by the convolution of the mixing filter and the estimated separation filter shown as $u_{j'j}(l) =$

$\sum_{j=1}^{L}\sum_{\tau=0}^{T-1} g_{j'i}(\tau)h_{ij}(l-\tau)$. SIR is calculated by the ratio of the energy of the target component $u_{rr}(l)$ and the power of the interference components $\sum_{j'\neq j} u_{j'j}(l)$ for each source, which is described by

$$SIR = \frac{\mathbf{E}\left(u_{rr}(l)\right)}{\mathbf{E}(\sum_{j'\neq j} u_{j'j}(l))}.$$

In order to compute SDR, the desired overall impulse response is decomposed into the sum of a scalar version of the target impulse response and a error distortion component, which can be written as

$$u_{j'j}(l) = \alpha \cdot h_{j'j}(l) + e_{j'}(l).$$

The real-value scalar α is computed to minimize the distortion component and maximum the SDR value. SDR is computed by the ratio of the power of $\alpha \cdot h_{j'j}(l)$ and $e_{j'}(l)$. It is written as

$$SDR = \frac{\mathbf{E}\left(\alpha \cdot h_{j'j}(l)\right)}{\mathbf{E}\left(e_{j'}(l)\right)}.$$

From the matrix of overall impulse response, it is not difficult to calculate SIR. However, in order to find the scalar value which minimizes the distortion, the problem that $u_{j'j}(l)$ and $h_{j'j}(l)$ have different lengths should be solved. Since the permutation indeterminacy problem is threoretically avoided in the IVA method, the reference impulse response can be obtianed easily. As introduced in the paper, reference impulse response $h_{j'j}(l)$ is chosen following the minimum distortion principle [SMK+03]. In this way, SDR calculation is implemented without trying all possibilites and selecting the best result.

It is known that the overall impulse reponse obtained after applying the minimum distortion principle is to approximate the room impulse response sequence. In order to align the reference impulse response with the longer overall impulse reponse, cross-correlation is exploited and truncation is done after the alignment. In signal processing, cross-correlation is a measure of similarity of two sequences as a function of the lag of one relative to the other [Wiki]. The corss-correlation is computed between $u_{j'j}(l)$ and

$h_{j'j}(l)$ for each pair. The overall impulse response $u_{j'j}(l)$ is truncated with the same legnth as $h_{j'j}(l)$ according to the position of maximum correlation coefficient. Then the α value is calculated by

$$\begin{aligned}
\min_{\alpha} ||e_{j'}|| &= \min_{\alpha} ||u_{j'j} - \alpha h_{j'j}|| \\
&= \min_{\alpha} (u_{j'j} - \alpha h_{j'j})(u_{j'j} - \alpha h_{j'j})^{\dagger} \\
&= \min_{\alpha} \left(u^{\dagger}_{j'j} u_{j'j} - \alpha u^{\dagger}_{j'j} h_{j'j} - \alpha h^{\dagger}_{j'j} u_{j'j} + \alpha^2 h^{\dagger}_{j'j} h_{j'j} \right).
\end{aligned}$$

Finally, the α value can be obtained by differentiating the fucniton above with respective to the α and setting the partial differentiate to zero, which gives the result as follows:

$$\alpha = \frac{u^{\dagger}_{j'j} h_{j'j}}{h^{\dagger}_{j'j} h_{j'j}}.$$

The signal-independent SIR and SDR can be calculated as the method introduced above. The signal-dependent ones which takes the influence of the source signals can also be evolved based on these. The only difference between the signal-dependent one and the signal-independet one is to convolve the overall transfer funcitons with source signals. Other steps are the same as the signal-independent one.

4.2 BSS Eval

Another performance measurement method used here is the subspace projection approach proposed by Emmanuel Vincent [VGF06]. The subspace projection method is to decompose the estimated signal into four components as follows

$$\hat{s}_j = s_{target} + e_{interf} + e_{noise} + e_{artif}$$

where s_{target} is a modified version of the true source signal with tolerant distortion which is only considered as time-invariant gains here, e_{interf} is the interference from other unwanted sources, and e_{noise} and e_{artif} are noise an artifacts terms respectively. It is known that the advantage of the approach is that the perfomance measurement can be done with only the estimated source and the true source signal. The mixing system and the demixing technique is not necessarily known in advance. In the proposed paper, several critiria are emplyed such as SDR, SIR, SNR and SAR. Only two of them are considered here and energy ratios are expressed as follows:

$$SDR = 10\log_{10}\frac{||s_{target}||^2}{||e_{interf} + e_{noise} + e_{artif}||^2}$$

$$SIR = 10\log_{10}\frac{||s_{target}||^2}{||e_{interf}||^2}$$

The noiseless situation is assumed here, the e_{noise} term can be neglected. The decomposition is implemented based on orthogonal projections, whose orthogonal projector denotes as $\prod\{y_1, \cdots, y_k\}$. The decomposed components are written as

$$\begin{aligned} s_{target} &= \prod\{s_j\} \cdot \hat{s}_j \\ e_{interf} &= \prod\{(s_{j'})_{1\leq j'\leq n}\} \cdot \hat{s}_j - s_{target} \\ e_{artif} &= \hat{s}_j - s_{target} - e_{interf} \end{aligned}$$

Based on the values of SDR and SIR, the performance is known. SDR could be viewed as the ability of the algorithm to reject the overall distortion and SIR shows the ability to reject interference from unwanted sources. The subspace projection method is relatively easy to understand from the perspective of the geometry. The implementation is equivalent to calculate least squares problem, which is a approach to address the over-determined problem. All vecotrs in the subspace spanned by the desired source or all input source signals are used to approximate the estimated desired signal as close as possible. The projector is a matrix to linearly combine all vectors in the subspace. In the Python implementation, a stragety which approximates the subspace with few delay versions of base vectors is used to simplify calculation. If all delay versions of the source are considered as the subspace vectors, the computation complexity is large and the calculation speed is slow. The calculation is done efficiently in the frequency domain. One should notice that the source images on the microphones should be used instead of the real source signals.

4.3 PESQ

There are different types of methods to assess speech quality. Two main categories, subjective listening test, and objective quality measures, can be distinguished. Both have their own pros and cons. For example, compared to subjective ones which are valuable but time-consuming and expensive, objective methods need to measure the distance between the original and estimated signals and combine carefully with the auditory model. Subjective listening tests include Mean Opinion Score (MOS) method which is the arithmetic mean over all the scores obtained from all listeners, Multi-Stimulus with Hidden Reference and Anchor (MUSHRA) which is another method to evaluate the subjective audio quality etc.. Objective quality measures include many perceptually related quality methods. Peceptual speech quality measure (PESQ) which belongs to objective methods is mainly used here. PESQ is an intrusive measure, which means a source signal is input as a reference. The possible sample rates used by PESQ are only $8kHz$ or $16kHz$. The range of the PESQ score is ranging from -0.5 to 4.5.

Chapter 5

Evaluation

In chapter 3, the fact that there are no spikes at two ends means no circularity problem. The assumpution in IVA is that it models the demixing system as multiplication in the frequency domain with the sufficiently long length of the demixing filter. Compared to figure 4, if the length of the demixing filter is fairly short, whcih is not enough to approximate the room impulse response, the interference peaks can be observed and this case is showed in figure 6. The reverberation time is defined as 0.4s. The number of iteration times is 30. After trying with the demixing filter of different lengths(512, 1024, 2048, 4096, 8192, 16384, 32768 samples), the suppression filters are acceptable as long as the length of the demixing filter is greater than or equal to 4096(with sampling frequency 16 kHz here). This also applies to feeding an ideal signal source into the system. Except the circularity problem, several methods are used to measure SIR and SDR values and they are recorded in the output files. The plot of evaluation results from the BSS Eval measure method is created by fixing the length of the room impulse response and varying the length of the demixing filter, which can be seen in figure 7. The number of iterations is specified with 30. The length of the room impulse response which is called nsample shown in the tile of the plot is defined as 6400 samples. The reverberation time is 0.4s. If the length of the room impulse response is fairly shorter than that of the demixing filter, performance degradation is caused, which means low SIR and SDR values. The signal independent method can also give roughly similar

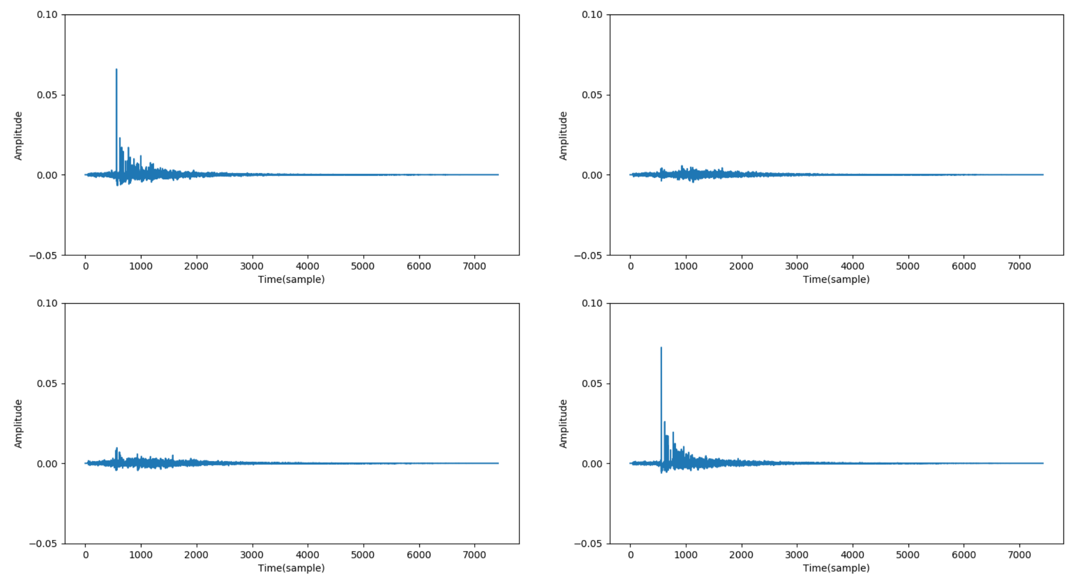

Figure.6. Overall impulse reponses of the speech signal by the finite-length filter with room impulse response of 6400 length and demixing filter of length 1024.

values for the SIR and SDR plots, but slightly different depending on the reverberation time. It is also compared with the case of 50 iteration times, and there is no much difference for the envelope of the plot between those two cases.

The reverberation time also plays an important role in the performance evaluation. From the plots, the longer the reverberation time is, the lower SIR or SDR value is. The reverberation is stronger, the late parts of the room impulse response have more impacts. For example, the maximum values of SIR and SDR reach the maximum both at the demixing filter of length 16384 in the scenario with 0.8s reverberation time in figure 8. For the case of $0.2s$ reverberation time as shown in figure 9, the SIR and SDR values reach the maximum at the demixing filter of length 8192 and 4096 respectively. The assumption becomes more critical.

The fact that the envelope of the plot does not change much can be observed while increasing the length of the mixing filter from 6400 samples to 8400 or 12000 samples under the specific reverberation time, which means the majority of the room impulse

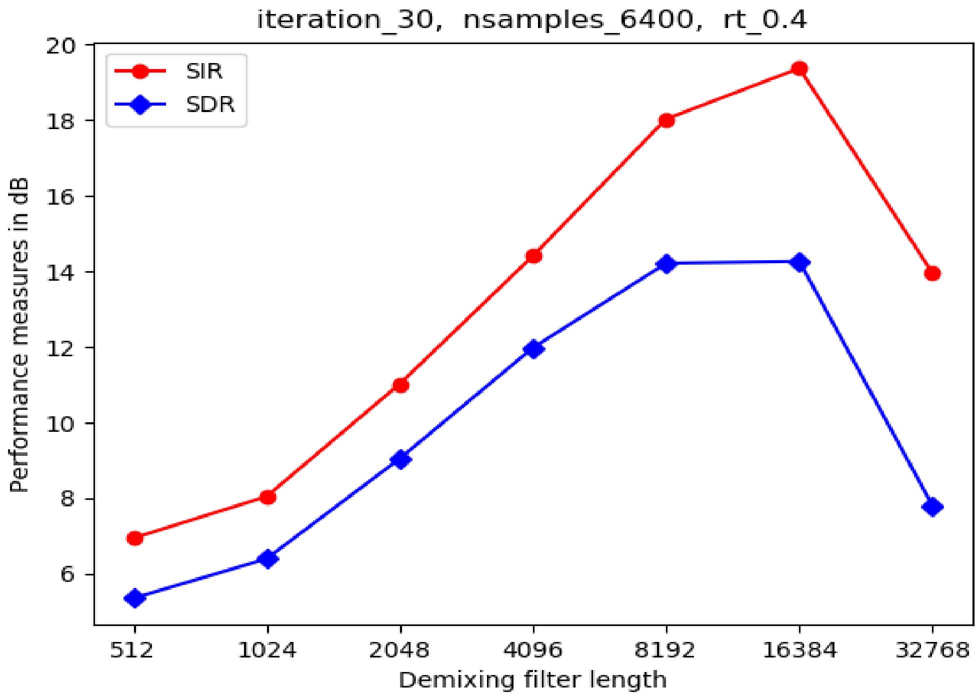

Figure.7. The influence of the length of the demixing filter with the fixed length of room impulse response with 0.4s reverberation time.

response can be represented by the first part. Only one source-microphone configuration is used as for all cases above which is shown in figure 3. The influence of the configuration of the microphones is also looked roughly. For good source microphone configuration, the length of the mixing filter does not need to be really long and acceptable suppression filters are obtained. However, the length needs to be long enough for difficult configurations. For example, for two microphones that are aligned with each other and perpendicular to the source, the length must be greater than 8192. If the positions of microphones and sources are randomized, the influence of the length of the demixing filter is characterized by the plots in figure 10 and 11. They are created by averaging over 30 random positions.

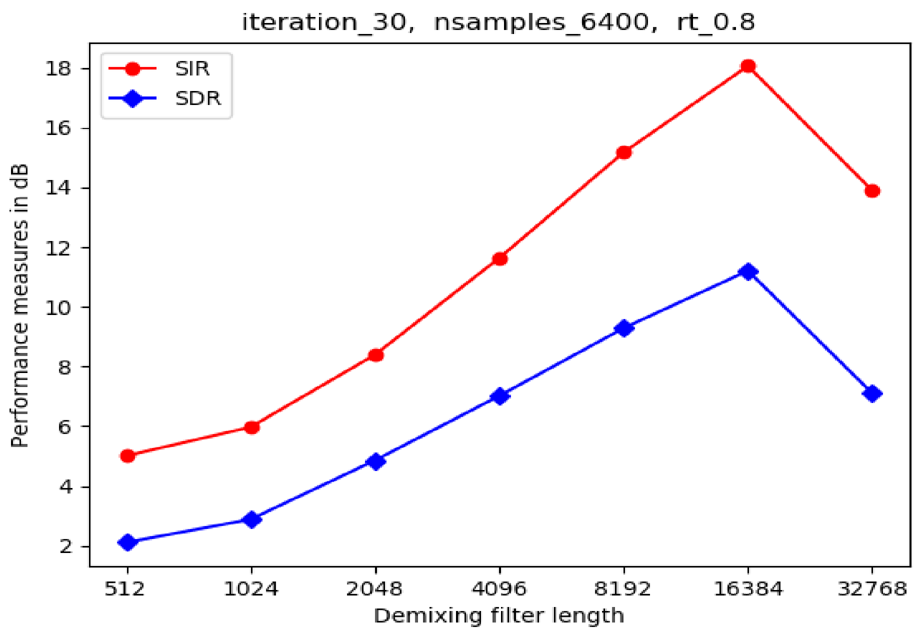

Figure.8. The influence of the length of the demixing filter with the fixed length of room impulse response with 0.8s reverberation time.

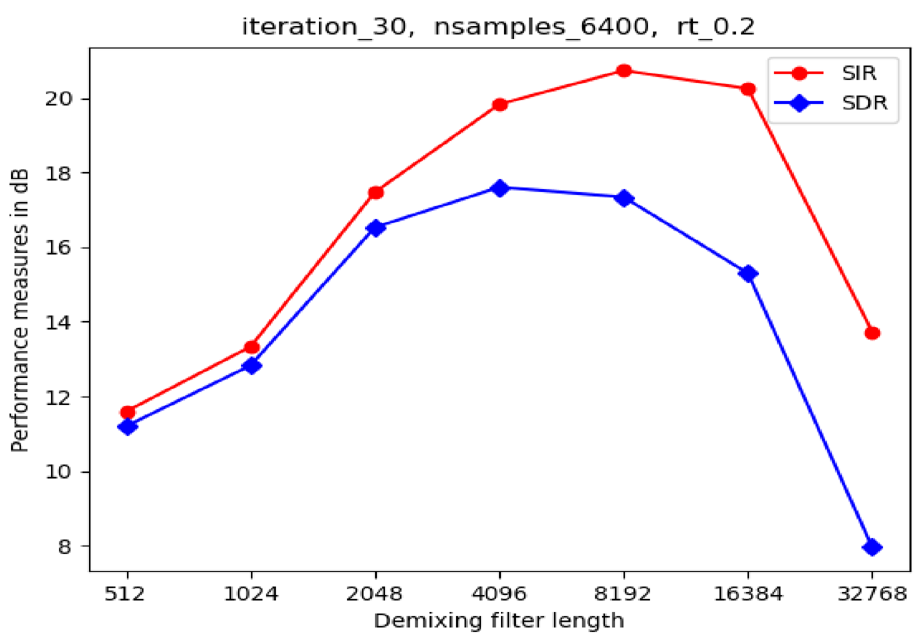

Figure.9. The influence of the length of the demixing filter with the fixed length of room impulse response with 0.2s reverberation time.

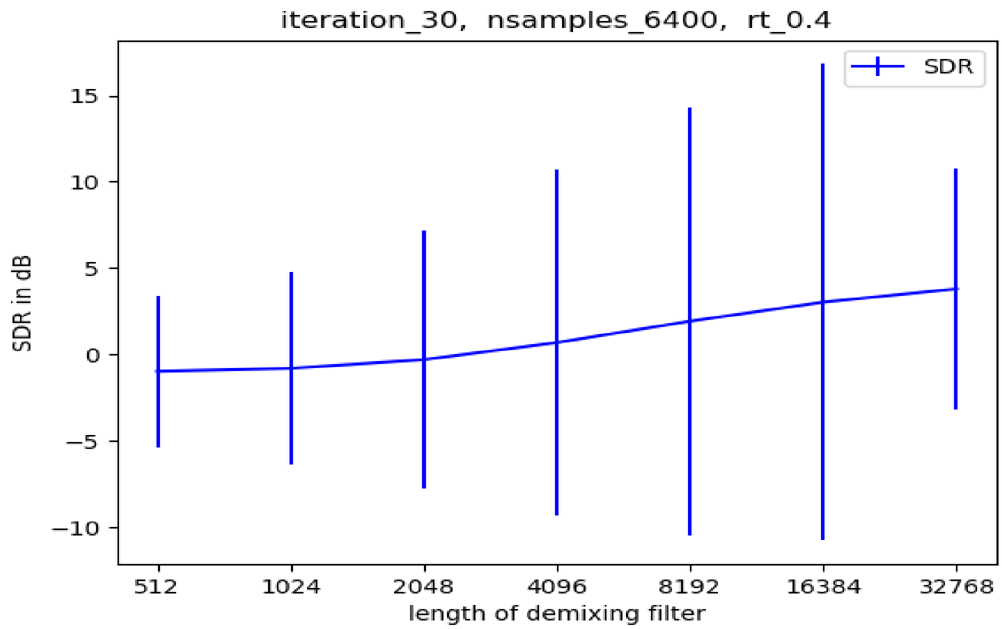

Figure.10. The influence of the length of the demixing filter with random positions for sources and microphones.

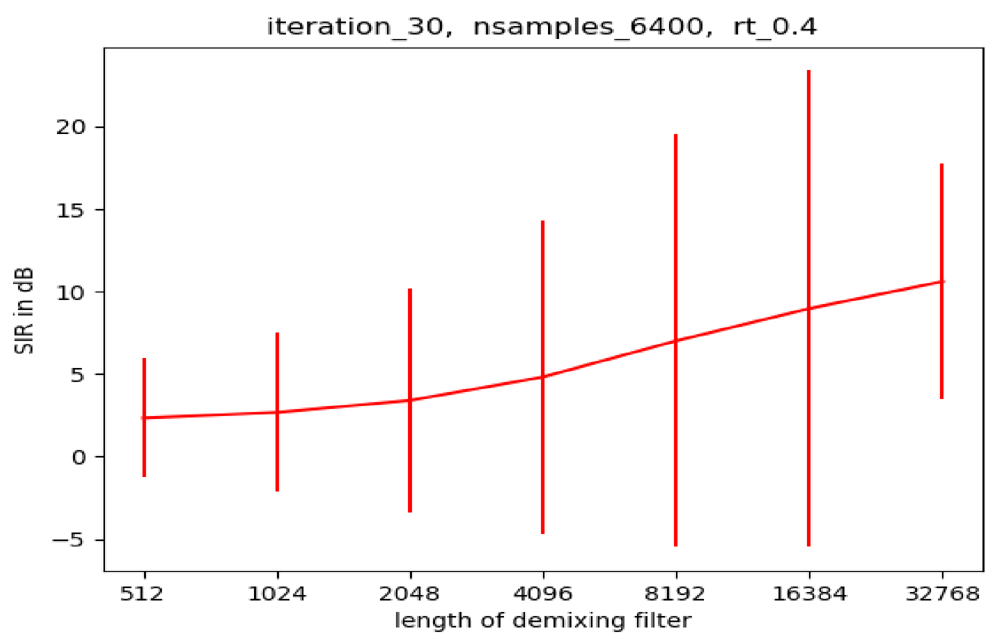

Figure.11. The influence of the length of the demixing filter with random positions for sources and microphones.

Chapter 6

Summary

Ideal sources are created according to the model introduced in [KALL07], but the correlated covariace matrix is utilized instead of the identity matrix. The computation of time-domain demixing filters of IVA is done. No circular convolutive spikes shown in the ICA paper are observed. The assumption that IVA models the demixing system as multiplication in the frequency domain with the sufficiently long length of the demixing filter is highly important, however it depends on the reverberation time. Several evaluation methods are utilized to show the performance measures.

Bibliography

[KALL07] Kim, Taesu ; Attias, Hagai T. ; Lee, Soo-Young ; Lee, Te-Won: Blind source separation exploiting higher-order frequency dependencies. In: *IEEE transactions on audio, speech, and language processing* 15 (2007), Nr. 1, S. 70–79

[Ono11] Ono, Nobutaka: Stable and fast update rules for independent vector analysis based on auxiliary function technique. In: *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on* IEEE, 2011, S. 189–192

[SMK+03] Sawada, Hiroshi ; Mukai, Ryo ; Kethulle, Sébastien ; Shoko, Ryhove ; Makino, Araki S.: Spectral smoothing for frequency-domain blind source separation. In: *in Proc. IWAENC 2003* Citeseer, 2003

[VGF06] Vincent, Emmanuel ; Gribonval, Rémi ; Févotte, Cédric: Performance measurement in blind audio source separation. In: *IEEE transactions on audio, speech, and language processing* 14 (2006), Nr. 4, S. 1462–1469