Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2010, Article ID 450487, 7 pages
doi:10.1155/2010/450487

Research Article

Underdetermined DOA Estimation via Independent Component Analysis and Time-Frequency Masking

Peter Jančovič, Xin Zou, and Münevver Köküer

School of Electronic, Electrical & Computer Engineering, University of Birmingham, Birmingham, B15 2TT, UK

Correspondence should be addressed to Peter Jančovič, p.jancovic@bham.ac.uk

Received 22 March 2010; Revised 22 June 2010; Accepted 20 July 2010

Academic Editor: Ioan Tabus

This paper presents an algorithm for the estimation of the direction of arrival (DOA) in underdetermined situations, that is, there is more sources than sensors. The algorithm performs the estimation in an iterative manner, each iteration consists of two-steps: first estimation of the DOA of a dominant source via the Independent Component Analysis and then removal of the detected source from the mixture via time-frequency masking. Experiments, performed using speech signals mixed in real environment when only two microphones are used but three and four sources are present, demonstrate that the proposed algorithm can estimate the DOAs more accurately than two previously used underdetermined DOA algorithms.

1. Introduction

The estimation of the direction of arrivals (DOAs) of sources using a microphone array is an important problem in multichannel speech signal processing with many applications, for instance, in teleconference systems. The DOA estimation is usually performed in the time-frequency domain. When the number of sources is less than the number of sensors, the DOA can be estimated by employing the subspace approach. An example of this is the MUSIC (MUltiple SIgnal Classification) algorithm [1]. Algorithms in the subspace approach identify the noise subspace by using eigenvector decomposition and then estimates the DOA based on searching the vectors orthogonal to the noise subspace.

Recently, a DOA estimation algorithm employing the Independent Component Analysis (ICA) was proposed in [2]. In this algorithm the DOA is estimated from the ICA matrices. Unlike the MUSIC algorithm, which exploits only the second-order statistics, the ICA-based DOA estimation exploits also higher-order statistics. This algorithm can be employed when the number of sources is less than or equal to the number of sensors. However, in practical applications it is often the case that the number of sources is greater than the number of sensors due to the presence of various ambient sounds and limited number of sensors on a device. This is referred to as the underdetermined situation.

Algorithms for DOA estimation proposed for the underdetermined situations are typically based on the W-disjoint orthogonality (W-DO) assumption [3] of speech signals in time-frequency domain, that is, only one source is active at a time-frequency point. Based on this assumption, the DOA can be estimated from a collection of time-frequency signal features that appear to belong to each source. There have been several algorithms proposed and these mainly differ in two aspects: the type of signal features used and the way to find the clusters of these features. The normalised time-frequency samples were used as features in [4], their amplitudes and phases were used in [5–8], and Hermitian angle in [9]. To find the cluster centres, histogram method was used in [5–7], the k-means clustering in [4, 8, 9] and the Gaussian mixture model in [8]. A combination of the use of time-frequency masking and ICA was proposed in [10] for underdetermined blind source separation. This algorithm first aims to convert the situation to determined by removing a number of sources from the mixture signals, which is performed by employing the clustering-based DOA estimation and time-frequency masking, and then applies the ICA to separate the remaining sources.

In this paper, we present an underdetermined DOA estimation algorithm employing the ICA and time-frequency masking. The proposed algorithm performs the DOA estimation in an iterative manner; each iteration consists of the estimation of the DOA of a dominant source via the ICA and of the removal of the dominant source from the signal mixtures via time-frequency masking. As a result, the DOAs of the sources contained in the mixture will be estimated one by one. Experiments are performed using speech signals mixed in real environment. Evaluations are presented for various frame-sizes and lengths of the signal. Experimental results demonstrate that the proposed algorithm can estimate the DOAs for three and four sources using only two microphones with a better accuracy than the above mentioned clustering-based and combined time-frequency masking & ICA algorithms.

2. The Mixing Model and Independent Component Analysis

2.1. Mixing Model. In real-world environments, the microphones capture not only the original sources but also their delays due to the reverberation. Considering there are N sources and M microphones, this scenario can be written using the convolutive mixing model as

$$x_m(t) = \sum_{n=1}^{N} \sum_{l=0}^{L} h_{mn}(l) s_n(t-l), \tag{1}$$

where $m = 1, \ldots, M$, s_n is the signal from the source n, x_m is the mixture signal captured by the microphone m, and h_{mn} is the impulse response from the source n to the microphone m with L being the maximum time delay due to the reverberation.

The DOA estimation is usually performed in the time-frequency domain, which is obtained by splitting the entire signal into short segments (frames) and taking the short-time Fourier transform of each frame. In this domain, the convolutive mixtures can be approximated as instantaneous mixtures and the mixing model can then be written in matrix notation as

$$\mathbf{X}(f,l) = \mathbf{H}(f)\mathbf{S}(f,l), \tag{2}$$

where f is the frequency index and l is the frame-time index. $\mathbf{H}(f)$ is M by N mixing matrix corresponding to the impulse response h_{mn}, and $\mathbf{X}(f,l) = [X_1(f,l), \ldots, X_M(f,l)]^T$ and $\mathbf{S}(f,l) = [S_1(f,l), \ldots, S_N(f,l)]^T$ denote the vectors of the short-time Fourier transform of the mixture signals at all the microphones and of the source signals, respectively, at frequency f and frame-time l.

2.2. Independent Component Analysis. The Independent Component Analysis (ICA) can be used to separate the observed mixture signals and as such also to estimate the DOAs of the sources when the number of sources N is less or equal to the number of microphones M. As such, we employ ICA by considering that the number of independent source signals is equal to M. The ICA technique is used for each frequency bin f to separate the mixture signals $\mathbf{X}(f,l)$ by means of

$$\mathbf{Y}(f,l) = \mathbf{W}(f)\mathbf{X}(f,l), \tag{3}$$

where the index l goes over all frames, $\mathbf{Y}(f,l) = [Y_1(f,l), \ldots, Y_M(f,l)]^T$ is the vector of the estimated separated signals. The $\mathbf{W}(f)$ is the $M \times M$ ICA unmixing matrix which is obtained by applying an ICA algorithm on a collection (over all frames l) of $\mathbf{X}(f,l)$. There are several ICA algorithms, each derived based on a different criteria. One of the most widely used is the negentropy-based FastICA [11], which we also employed in this paper.

3. The Proposed Underdetermined DOA Estimation via ICA and Time-Frequency Masking

The proposed algorithm aims to estimate the DOA of N source signals which are observed by M microphones in underdetermined situations, that is, $M < N$. The ICA itself cannot be used to estimate the DOA of the sources directly in underdetermined situations. However, since the ICA aims to estimate independent source components, the estimated signals $Y_1(f,l), \ldots, Y_{M-1}(f,l)$ should correspond to any $(M-1)$ out of N of the source signals and the $Y_M(f,l)$ will be then a mixture of the remaining $N-(M-1)$ sources. Based on this, it should be possible to use the ICA for estimation of the DOA of up to $M-1$ sources from M mixture signals. This principle was exploited in [12] where it was employed for separation of dominant sources in underdetermined situations, but the number of dominant sources did not exceed the number of microphones. Here we employ this principle for the DOA estimation of all the sources, number of which is greater than the number of microphones.

Considering the above principle of the ICA and the sparseness of speech in the time-frequency domain, we propose a scheme in which the ICA is employed iteratively to estimate the DOAs of the sources. Each iteration of the proposed algorithm consists of two steps: (i) the ICA-based estimation of the DOA of a dominant source and (ii) the removal of the current dominant source via time-frequency masking. By iteratively applying the above two steps, one can obtain the DOA estimates of all sources. The iterative process can be stopped based on the number of sources being reached, when this information is available, or otherwise based on the energy of the remaining mixture spectrum falling below a given threshold value. These two steps of the iterative process are described in details in the following. Note that for clarity the algorithm is demonstrated considering two microphones, however, the same procedure applies in a case of more microphones.

3.1. The ICA-Based DOA Estimation of a Dominant Source. The application of an ICA algorithm as described in Section 2.2 provides the ICA unmixing matrix $\mathbf{W}(f)$ for each frequency bin f. The DOA can be estimated by using

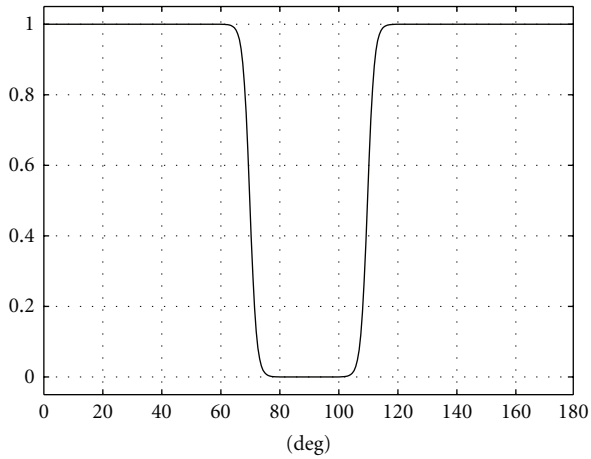

Figure 1: The masking function employed in the proposed algorithm.

the inverse of $\mathbf{W}(f)$ or by searching for a minimum in the directivity pattern [2]. The former method was employed here. Denoting the inverse of $\mathbf{W}(f)$ by $\mathbf{A}(f)$, the DOA can be calculated using the elements of each column n of the matrix $\mathbf{A}(f)$ as

$$\theta_n(f) = \cos^{-1}\left(\frac{\text{angle}(A_{2n}(f)/A_{1n}(f))}{2\pi f c^{-1} d}\right), \tag{4}$$

where d denotes the distance of the microphones and c is the speed of sound. The term $\text{angle}(A_{2n}(f)/A_{1n}(f))$ denotes the phase difference between the two elements of the nth column vector of $\mathbf{A}(f)$. The obtained DOAs $\theta_n(f)$ for $n = 1$ and $n = 2$ are both used.

After the above ICA-based DOA estimation procedure is performed for each frequency bin f, a histogram of the obtained DOAs collected over all frequency bins (and each column n) is constructed and the DOA of the dominant source in the current iteration, denoted by θ^*, is obtained by finding the highest peak of the histogram. Note that if the current DOA estimate is close to a DOA estimate obtained in previous iterations, they are considered belonging to the same source and thus the source DOA is updated by their average.

3.2. Removing of the Source by Time-Frequency Masking. In this step, the current dominant source is removed from the mixture spectra. Such modified mixture spectra are then used in the following iteration to estimate the DOA of another source.

As demonstrated in [3], a large percentage of speech energy is carried by a small percentage of time-frequency samples, and as such the W-DO assumption is approximately satisfied. Therefore, the current dominant source can be removed from the mixture by employing time-frequency masking. This is performed by multiplying the mixture spectrum at the microphone m with a masking function $\mathcal{F}$ at each time-frequency point as

$$X_m(f,l) = \mathcal{F}(\theta(f,l),\theta^*)X_m(f,l), \tag{5}$$

where $\theta(f,l)$ is the DOA of a source at the time-frequency point (f,l), which is calculated based on the phase difference of the time-frequency points of microphone mixture signals in $\mathbf{X}(f,l)$, that is, $\theta(f,l)$ is calculated using (4) but replacing the term $\text{angle}(A_{2n}(f)/A_{1n}(f))$ by the term $\text{angle}(X_2(f,l)/X_1(f,l))$. The θ^* in (5) denotes the DOA, from the set of DOAs estimated so far in Section 3.1. The masking function $\mathcal{F}$ is designed to deweight the signal from the dominant source DOA θ^* in the mixture, that is, it will have a value close to zero when $\theta(f,l)$ is close to θ^*. The $\mathcal{F}$ is defined as

$$\mathcal{F}(\theta(f,l),\theta^*) = \frac{1}{1+e^{-\beta(|\theta(f,l)-\theta^*|-\Delta)}}, \tag{6}$$

where parameters β and Δ determine the slope and the centre-offset of the curve, respectively. In our experiments, setting $\beta = 50$ and $\Delta = 20^\circ$ provided good performance. The masking function with these parameters when $\theta^* = 90^\circ$ is depicted in Figure 1.

4. Experimental Results

Experiments were performed using speech signals mixed in real environment, which are available from [13]. These are English speech utterances of about 7.3 s duration, sampled at 8 kHz. The sources were 1.2 m away from a linear microphone array containing four microphones at 4 cm of each other—note that only two adjacent microphones were used at a time for experiments. The DOAs of the four sources are 30°, 70°, 110°, and 150° in front of the microphone array. The configuration of the sources and microphones is shown in Figure 2. In the experiments, the entire mixture signals of 7.3 s duration were split into 10 segments of a specified length, each segment shifted by the corresponding number of samples in order to cover the entire 7.3 s signal. A given segment of the mixture signals was split into frames of a specified length, with 128 samples shift between frames, and DFT was applied. Experiments were performed using various lengths of the segment and frame. The negentropy-based complex FastICA algorithm [11] was employed for estimation of the ICA unmixing matrix. Experimental results are presented in terms of the mean absolute error of the estimated DOAs to the true DOAs. The results for each source are obtained as the average over the DOA errors from the 10 experimental runs (corresponding to the signal segments) for a given microphone combination and over all the two adjacent microphone combinations. The overall DOA error is obtained as the average over the DOA errors of all sources.

Since the proposed algorithm employs ICA to provide the DOA estimate of a dominant source in the mixture, we first analyse how much/often a source in the mixture is dominant. As the ICA exploits the second- and higher-order statistics, this analysis is performed in terms of the variances and kurtoses of the sources for each frequency bin. Note that the variances and kurtoses are normalised such as to sum to 1 over all the sources. As we are using two microphones, we assess the difference between the variance/kurtosis of the dominant (i.e., strongest) and the third strongest source

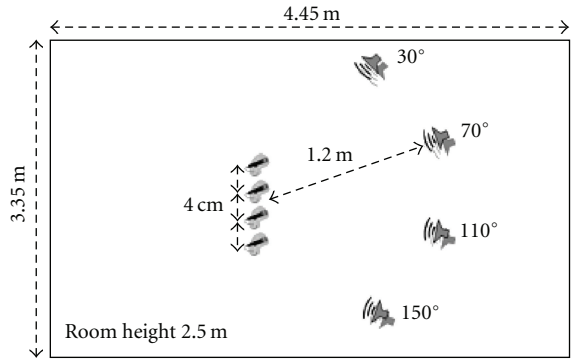

Figure 2: Configuration of the sources and microphones (reproduced based on [13]).

(in both cases of three and four sources) in the mixture. If the difference is large, the DOA estimate of the dominant source will be affected little by the third strongest source (and therefore also by the fourth in the case of four sources). The histograms of the differences in variances and kurtoses in the case of three and four source signals in the mixture are depicted in Figure 3. It can be seen that the means of the histograms are located at around 0.4-0.5 in the case of three sources and at around 0.2-0.3 in the case of four sources. The histograms indicate that in the case of three sources, the difference of the normalised variance/kurtosis between the dominant and the third strongest source signal was larger than 0.2 in 91%/86% of the frequency bins. The corresponding results for the case of four sources were 75%/67%. These results indicate that there is a large amount of frequency bins when one source has considerably larger value of variance and kurtosis. Corresponding to these results, it was found that the estimated DOA of a source was within 10° range of the true DOA in 86% and 81% of the frequency bins in the case of three and four sources, respectively.

Next, we present the evaluation of the proposed algorithm with regard to the automatic estimation of the number of sources in the mixture. The iterative process of the proposed algorithm was stopped automatically based on the remaining energy in the signal and the difference between the remaining energies in the current and previous iterations. There were in total 210 experimental runs of the algorithm, using three and four sources and various segment lengths of the signal. Out of these 210 runs, the number of sources was estimated incorrectly in 9 cases, which gives approximately 4.3% error in the estimation of the number of sources. In Figure 4 are shown examples of histograms of the DOAs estimated over all frequency bins obtained from our experiments when three sources at 30°, 70°, and 110° were present. It can be seen that the DOA of the source at 110° was estimated first, then the source at 70° and finally source at 30°. After these three iterations of the algorithm, over 95% of the signal energy was removed and the algorithm was considered as converged. All the experimental results of the proposed algorithm presented below were obtained by using the automatic estimation of the number of sources.

Here, we present evaluations of the proposed algorithm in order to find the effect of the frame length (used to split the signal) on the DOA estimation accuracy. Experiments were performed with the frame length set to 256, 512, 1024, and 2048 samples, respectively. The presented results are obtained as the average over ten runs with signal segments of 4 s length. The obtained overall DOA errors are presented in Figure 5. It can be seen that the frame length of 1024 samples, which corresponds to 128 ms, provided the best performance and this is used in all of the following experiments. These results show that the proposed ICA-based DOA estimation achieves better accuracy with longer frame length than the 20–30 ms which is typically used in speech processing. The obtained results reflect the use of similar long frame length by others, for example, [12].

Next, we performed experiments to demonstrate the performance as a function of the length of the signal used for the DOA estimation. The segment length was set to a value from 2 s to 7 s, in 1 s steps. The obtained results are presented in Figure 6. It can be seen that as the length of the signal decreases, the DOA error increases, however, the increase is only mild. This indicates that even a short length of the signal, such as 2 s, can provide acceptable DOA accuracy, which may be important in situations with fast moving sources.

Finally, we present experimental results obtained by the proposed algorithm and compare them to two previously used algorithms. These experiments were performed for the cases of three and four sources when two microphones are used, using the signal segment length of 7 s. The first algorithm included for comparison, denoted as "Algorithm 1", was a conventional algorithm presented in [4], which performs k-means clustering of the normalised time-frequency points. The second algorithm included for comparison, denoted as "Algorithm 2", was based on the work presented in [10]. Although the task in [10] was the source separation and only for the case of three sources and two microphones, we implemented the part of the algorithm corresponding to the DOA estimation as follows. First, the conventional clustering-based DOA estimation, as described above, was performed to provide three and four DOA estimates for the case of three and four sources, respectively. Then, in the case of three sources, the source corresponding to the estimated DOA with the largest number of assigned time-frequency points was removed from the mixture signals. Correspondingly, in the case of four sources, two sources were removed from the mixture signals. This led to a determined situation and as such the ICA was employed to estimate the DOAs of the remaining two sources. Both of the algorithms included for comparison considered the number of sources being known and the frame length was set to 512 samples as this obtained the best results. The results are presented in Table 1. Comparing the two previous algorithms, it can be seen that the "Algorithm 2" obtained better results than the "Algorithm 1", which can be attributed to the use of higher-order statistics by employing the ICA for estimation of the DOAs of the two remaining sources. It can be seen that the proposed algorithm obtained significantly lower DOA estimation error than both of the

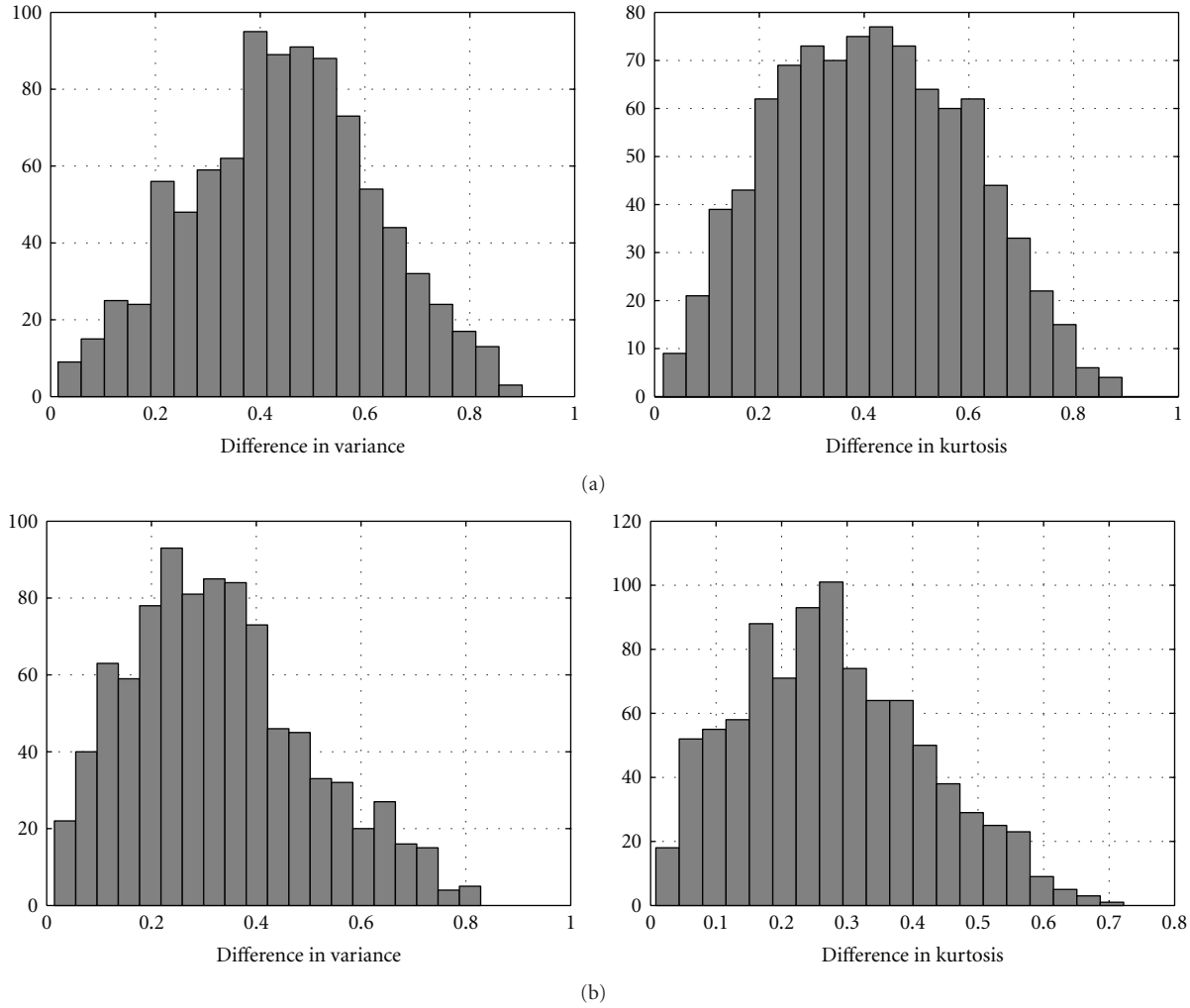

Figure 3: Histogram of the differences between the variances (left) and kurtoses (right) of the dominant and the weakest source (for the case of three sources) (a) and of the dominant and the second weakest source (for the case of four sources) (b).

Table 1: Mean absolute error in estimated DOAs (in degrees) obtained by the proposed algorithm and the previous algorithms when two microphones are used and three/four sources are present.

Algorithm	Mean absolute error in estimated DOAs (°)						
	3 sources			4 sources			
	$s_1(30°)$	$s_2(70°)$	$s_3(110°)$	$s_1(30°)$	$s_2(70°)$	$s_3(110°)$	$s_4(150°)$
Algorithm 1	9.6	11.6	11.8	20.4	16.4	11.6	15.9
Algorithm 2	8.4	6.5	8.5	11.5	13.0	12.3	13.3
Proposed	6.1	4.2	3.6	8.7	5.3	4.4	6.5

previous algorithms for both cases of three and four sources present.

5. Conclusion

In this paper, we presented an underdetermined DOA estimation algorithm based on the combination of the ICA and time-frequency masking. The algorithm performed the estimation in an iterative manner, each iteration consists of the estimation of the DOA of a dominant source via ICA and then the removal of this source via time-frequency masking. Experiments were performed using speech signals mixed in real environment. Evaluations of the proposed algorithm were presented for various lengths of the frame and various lengths of the signal. Comparisons were provided with two previously used underdetermined DOA estimation algorithms. Experimental results demonstrated that the proposed algorithm can provide more accurate DOAs than the previous algorithms when only two microphones are used and three or four sources are present.

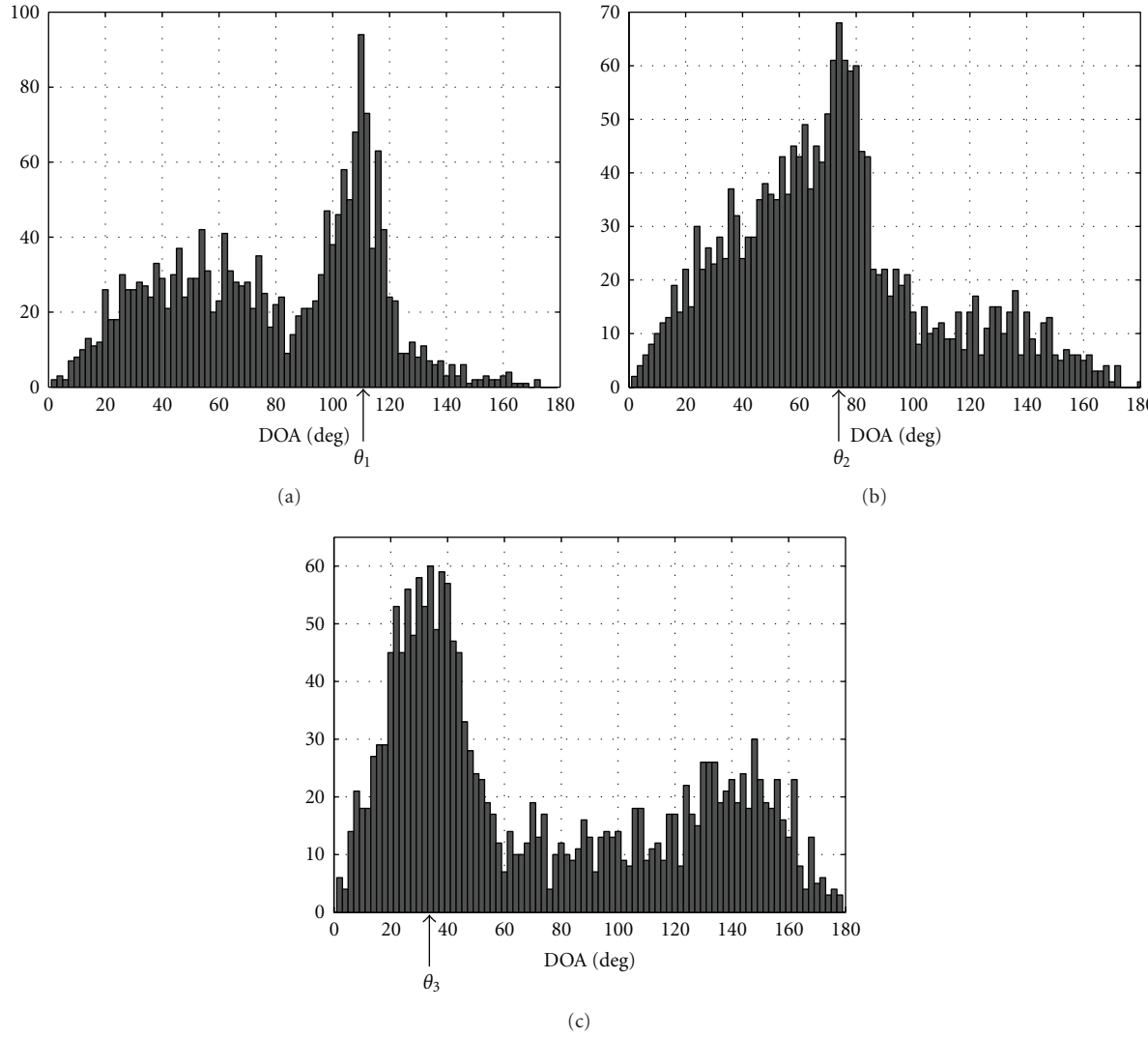

Figure 4: Histogram of the estimated DOAs via ICA from the original mixtures (a), from mixtures after the source from θ_1 was masked (b), and from mixtures after the sources from both θ_1 and θ_2 were masked (c).

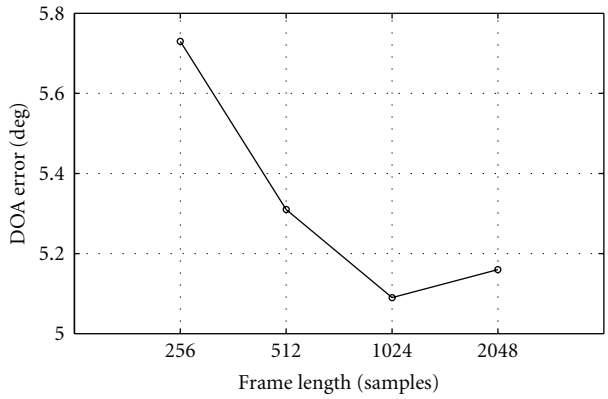

Figure 5: Overall mean absolute error in estimated DOAs (in degrees) obtained by the proposed algorithm as a function of the frame length when two microphones are used and three sources are present. The signal length of 4 s was used.

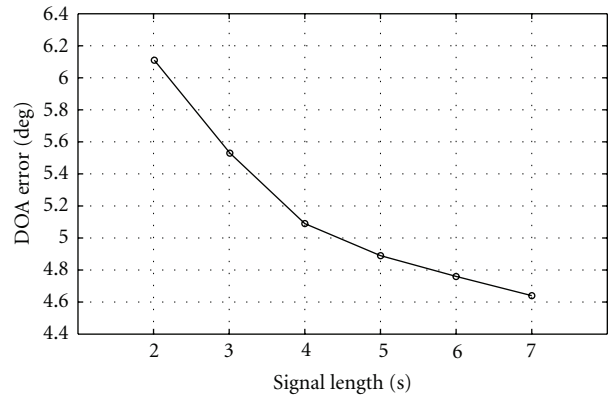

Figure 6: Overall mean absolute error in estimated DOAs (in degrees) obtained by the proposed algorithm as a function of the length of the signal when two microphones are used and three sources are present.

Acknowledgment

This paper was supported by UK EPSRC Grant EP/F036132/1.

References

[1] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[2] H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *Proceedings of the International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 411–414, 2003.

[3] Ö. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1846, 2004.

[4] S. Araki, H. Sawada, R. Mukai, and S. Makino, "DOA estimation for multiple sparse sources with normalized observation vector clustering," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '06)*, vol. 5, pp. 33–36, Toulouse, France, 2006.

[5] S. Rickard and F. Dietrich, "DOA estimation of many W-disjoint orthogonal sources from two mixtures using duet," *IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 311–314, 2000.

[6] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[7] M. Matsuo, Y. Hioka, and N. Hamada, "Estimating DOA of multiple speech signals by improved histogram mapping method," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '05)*, pp. 129–132, 2005.

[8] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, 2007.

[9] V. G. Reju, S. N. Koh, and I. Y. Soon, "Underdetermined convolutive blind source separation via time-frequency masking," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 1, Article ID 5061881, pp. 101–116, 2010.

[10] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Blind separation of more speech than sensors with less distortion by combining sparseness and ICA," in *Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC '05)*, pp. 271–274, Kyoto, Japan, September 2003.

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

[12] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of dominant target sources using ICA and time-frequency masking," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 6, Article ID 1709904, pp. 2165–2173, 2006.

[13] http://www.kecl.ntt.co.jp/icl/signal/sawada/demo/bss2to4/index.html.