International Journal of Systems Signal Control and Engineering Application 9 (5-6): 157-165, 2016
ISSN: 1997-5422

Permutation Correction in Convolutive BSS Using MFCC Coefficients

[1]Mostafa Esmaeilbeig, [2]Hamid Sheikhzadeh and [1]Farbod Razzazi
[1]Department of Electrical and Computer Eng., Science and Research Branch,
Islamic Azad University, Tehran, Iran
[2]Department of Electrical Eng,
Amirkabir University of Tech (Tehran Polytechnic), Hafez Ave., Tehran, Iran

Abstract: In this study, we propose a novel algorithm for solving the permutation ambiguity problem in convolutive blind source separation of speech signals. Transferring convolutive mixtures into time-frequency domain, enables us to separate source signals by employing instantaneous algorithms in each frequency bin. After separation, the main challenge is the scale and permutation ambiguities which can imperil the separation performance. Overcoming this challenge needs the reordering of all separated signals in each frequency bin according to order of source signal. In this study we propose a new algorithm for reordering the separated signals based on statistics of MFCC of speech signals. In each frequency bin, the separated subband signals are transferred back to time-domain and their individual MFCC's are extracted. Then, based on simple statistics of the MFCC's the permutation problem is resolved. The proposed algorithm drastically decreasesthe computational complexity and as a result speeds up the permutation correction process.

Key words: BSS, permutation, MFCC coefficient, convolutive, frequency domain

INTRODUCTION

Blind Source Separation (BSS) is a method for separating mixed signals without any knowledge of the sources and their combination process. In recent decades, the BSS has been the focus intense interest because of its wide range of applications in various fields such as biomedical signal processing (Pendharkar *et al*., 2014) telecommunications, signal processing (Besic *et al*., 2015) seismic and other geological and mechanical systems (Yu *et al*., 2014). However, in this study, we will focus on BSS applied to convolutive mixtures of speech signals. In this scenario, it is assumed that each speaker signal is passed through a linear time-invariant filter and received by sensors (microphones). Accordingly, the separation is much more difficult than the instantaneous case where only the linearly scaled versions of the source signals are received by sensors. Many different approaches have been developed for solving the convolutive BSS problem (Buchner *et al*., 2005). We will provide a very short summary of the convolutive BSS here.

There are two main approaches to convolutive BSS: time-domain and frequency-domain methods. In time-domain approach the sources signal are separated from the mixed signals by optimizing a FIR filter, typically with a large number of taps. This approach in very slow, encounters convergence problems (specially with longer FIR filters) and suffers froma heavy computational cost (Won and Lee, 2008). A particularly more effective time-domain approach is a method based on sparsity of signals. In a sparse signal, most of the samples have a near-zero or insignificant value (which can be practically ignored) and a small subset of samples possesssignificant values. In recent approaches, the Sparse Component Analysis (SCA) is now widely employed for separation of speech signals (Gribonval and Lesage, 2006).

In contrast, frequency-domain approachesare more attractive. Transferring mixed signals from time-domain to frequency-domain, converts the convolutive mixtures to instantaneous ones and as a result, instantaneous separation algorithm such as Independent Component Analysis (ICA) (Nesta *et al*., 2011) can be applied to mixtures (Mukai *et al*., 2006). In another frequency-domain approach, the mixing process is modeled as a beamformer. Then by estimatingthe Direction of Arrival (DOA) in all frequency bins, the separated signals are properly aligned and sources can be separated (Xiong *et al*., 2014; Sawada *et al*., 2011).

Source separation in frequency-domain suffers from two major drawbacks: scale ambiguity and permutation ambiguity. Scale ambiguity is due to the fact that the separated signals in each frequency bin have unknown scale. Hence after the separation, the recovered signals

Corresponding Author: Mostafa Esmaeilbeig, Department of Electrical and Computer Eng., Science and Research Branch, Islamic Azad University, Tehran, Iran

are severely distorted due to random scaling of each subband signal. The permutation ambiguity, on the other hand, stems from the fact that in each frequency bin, the order of the separated signals is random and independent of other frequency bins. This leads to severe distortion and loss of signal continuity in time-domain. Obviously, it is imperative to resolve these two ambiguity problems for the frequency-domain approaches to BSS to be practically useful. In this study, we will deal with the more important problem of the permutation ambiguity.

It is obvious that in order to extract the sources from the separated signals in frequency-domain, one must compare each recovered signal in each frequency bin with the recovered signals in all of the other frequency bins. Assuming L frequency bins and N sources, one needs N! comparisons for each frequency bin and a total of L.N! comparisons for each time-frame of demixed signals. Considering the fact that signal comparisons are time-consuming processes, the computational complexity will be very large for large numbers of frequency bins (L) and sources (N). In reality, the number of frequency bins (L) should be comparable to (and larger than) the length of the actual convolutive mixing filter in time-domain. This constitutes the major motivation of this research to seek alternative and simpler methods for the ambiguity problem in the frequency-domain approaches to the convolutive BSS. To better signify the problem, a review of current state of the research is presented here.

In the last decade, many different approaches have been proposed to overcome the permutation ambiguity. These algorithms can be divided into three main groups. The first group uses the spectrum consistency of source signals the second one exploits the consistency of separation filters and the last one employs signal statistics in subbands (Mazur and Mertins, 2009; Serviere and Pham , 2006; Sawada *et al*., 2004).

In the first category, the correlation between the amplitude of the subband signals in different frequency bins is used for reordering the recovered signals. The variation of amplitude of subband signals of each source has a unique pattern based on the contiguous signal pattern in time-domain. This forms the basis for an approach of clustering the subband separated signals (Anemuller and Kollmeier, 2000). This method suffers from two main drawbacks. First, if the source signal has any discontinuities in time-domain (due to silence time or abrupt change in speech signal) the spectrum consistency will be lost. The second deficiency is the error propagation which means that if in a frequency bin, the signal is wrongly attributed to a source, this error will be repeated in other frequencies and as a result, the error propagation occurs in the rest of frequency bins.

As an example of this group of algorithms, Duran-Daz *et al*. (2012) exploit the amplitude modulation correlation between speech signals in adjacent frequency bins. They use the correlation of logarithmically scaled separated signals in each frequency bins for reordering the recovered signals and overcoming the permutation ambiguity.

The second category uses the continuity of the demixing filter coefficients. As a basic principle, this method exploits the continuity and the smoothness of the frequency response of the separation filters. This method also has two main drawbacks. The first is error propagation similar to the first category. The second one is the inefficiency of separation in highly reverberant environments where the separation filter must include thousands of taps which leads to a high computational cost and increases the processing time.

The third category employs the statistics of subband separated signals in each frequency band. It is shown that these statistics are highly correlated for one source in adjacent frequency bins. Mazur and Mertins (2009) introduce a new method to resolve the permutation ambiguity problem specifically targeted for speech signals. They show that speech signals in each subband has a Generalized Gaussian Distribution (GGD) and the related parameters of the GGD for each source signal demonstrate continuity over adjacent frequency bands to the point that one can exploit this property for resolving the permutation ambiguity.

Also, Sarmiento *et al*. (2015) introduced a new function of generalized divergence, called the AB-divergence $D^{\alpha,\beta}_{AB}(p, q)$. This function represent the dissimilarity between two vectors p and q, therefore the negative of this parameter can be used as a measure of similarity between the two vectors. They employ this parameter for evaluation of similarity of the absolute value of the speech subband signals (represented by STFT coefficients in each frequency band) to resolve the permutation ambiguity.

Our proposed approach is resolve the permutation ambiguity in time-domain. The method is based on the assumed continuity of the separated suuband signals from one single source. Assuming L subbands and N sources, after the separation of mixed signals in time-frequency domain, the L.N separated subband signals are each individually transferred back to time-domain, obtaining L.N time-domain wavelets with N wavelets per subbands which has to be reordered. Each of these N wavelets should be compared to the other wavelets in the adjacent time-frames to recover the correct reordering of the signals. To implement this wavelet

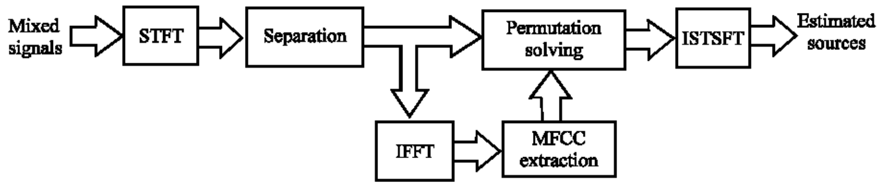

Fig. 1: Diagram of the proposed method

comparison, we must select a suitable feature. We have proposed the use of MFCC of the time-domain wavelets as an efficient feature for this task.

This approach has precedence in the literature. Khanagha and Khanagha (2009) introduced an approach for solving permutation problem in time-domain based on TDOA for localization of sources. They use MFCC and PLPCC feature for clustering and recovery of source signals from mixtures in time-domain. In our proposed method, at first we transfer the mixed signals to time-frequency domain and after separation of the signals in subbands each subband signalsare transferred back to time-domain to obtain a wavelet per source per subband. Next based on the MFCC's of this wavelets, the separated signals are clustered. Then according to this clustering, all time-domain wavelets associated with successive subbands are reordered. Finally, the subband signals are reordered based on wavelet orders in each subband and transferred back to time-domain.

Figure 1 shows the diagram of proposed method. The MFCC has played a significant role in speech detection, speech recognition and also for speaker identificationand verification. We extract the statistics of MFCC of the time-domain wavelets and use them as a criterion for reordering them. We will show that the MFCC has a good performance for solving permutation ambiguity and can be used in time domain instead of frequency domain.

Problem description: To model the convolutive BSS, assume that there are M sensors and N sources signals $s_i(t)$ ($i = 1, ..., N$) which are convolved with environment impulse response and then mixed together to constitute the sensor outputs of $x_j(t)$ ($j = 1, ..., M$). As such, the output signalscan be written as:

$$x_j(t) = \sum_{i=1}^{N} \sum_{k=1}^{\infty} h_{ij}(k)\, s_i(t-k) \quad (1)$$

where, $h_{ij}(k)$ represents the environment impulse response from source i to sensor j. By using Short Time Fourier Transform (STFT), the signals are transferred from time-domain to frequency-domain and the convolution is replaced by multiplication:

$$X(f,\tau) = H(f)S(f,\tau) \quad (2)$$

where, $X(f, \tau) = [X_1(f, \tau)\ X_2(f, \tau), ..., X_M(f, \tau)]^T$ and $S(f, \tau) = [S_1(f, \tau)\ S_2(f, \tau) .... S_N(f, \tau)]^T$ are STFT representations of the mixed and source signals, f stands for frequency and τ stands for time index and H(f) is the channel impulse response Fourier Transform. As a result, it is possible to separate the sources in each frequency band independently. There are many algorithms for instantaneous separation such as ICA and its derivatives. After separation, we face with two problems: scale and permutation ambiguity Eq. 2 can be written as:

$$X(f,\tau) = H(f)S(f,\tau) = \Lambda(f)p(f)S(f,\tau) \quad (3)$$

where, $\Lambda(f)$ is an arbitrary diagonal scaling matrix and p(f) is a frequency-dependent permutation matrix. The scaling matrix models the inevitable amplitude variations in frequency bands. This ambiguity should be resolved to avoid severe distortion due to random scaling of various frequency bands. In Ikeda and Murata (1998) proposed a method to correct this ambiguity by applying a post-filter on the single separated signals which were the inverses of the unmixing filters. Another method, called Minimal Distortion Principle (Matsuoka, 2002) proposes to employ a filter defined as:

$$H'(f) = \mathrm{Diag}(H^{-1}(f)).H(f) \quad (4)$$

where, Diag $(H^{-1}(f))$. H(f) is the inverse of matrix H(f) with all off-diagonal elements set to zero. Most of methods designed to resolve the permutation ambiguity are applied in frequency-domain. Typically, one of methods reviewed in the Introduction section are applied and after solving the permutation ambiguity, the separated signals for each source from all subbands are combined together and transferred back from frequency-domain to time-domain. In contrast to the common methods, instead of correcting the permutation ambiguity in frequency-domain, the separated signals in each subband is separately transferred to time-domain (each time section obtained for a frame in subband is called a 'wavelet' here) and the permutation ambiguity is resolved based on these subband-indexed time-domain wavelets.

MATERIALS AND METHODS

Method for resolving permutation: Historically certain features of speech signal have been employed for speaker identification. One of the most commonly used features is the MFCC which acts as a classifier. Motivated by this, in the proposed method, after separationof all sources in each frequency subband, the subbands signals are individually transferred to time-domain. For each subband and each separated source, we obtain an equivalent time segment, a wavelet which is associated with a source signal. However, the association of obtained wavelets to various source signals will be different in various subbands.

To demonstrate how the MFCC of these wavelets can be helpful, consider the mean of MFCC's of full-band time-domain signal for each source. We demonstrate that beyond a certain frame-length, this mean value: exhibits time-continuity for each source signal and discriminates between various sources. In order to show this we select five speech signals from different speakers form the TIMIT database sampled at 16 kHz and extract MFCC's from each speech signals with different frame sizes and with 50% frame overlap. Next, we calculate the means of these coefficients and compare them for various frame lengths. Figure 2 depicts the results.

As depicted, for larger frame-sizes, the mean values perform better in terms of both time-continuity and discriminability of different sources. To demonstrate the effect of number of MFCC's on the mean value behavior, we show in Fig. 3 that: as the number of MFCC's increases, the mean value decreases but it does not have a considerable effect on the time-continuity. The conclusion reached is that the mean values of MFFC's are effective in source discrimination provided that one considers long-term signals, 64 msec or more (1024 samples or more in our experiment).

In our STFT-based framework of Fig. 1 for resolving the permutation in time-domain for each subband, the MFCC's of the time wavelets are extracted fromall separated signals and their mean values are calculated. Starting from first subband, the separated signals are ordered according to their mean values. For the next subbands the mean values are compared to mean values of previous subband and separated signals are ordered accordingly.

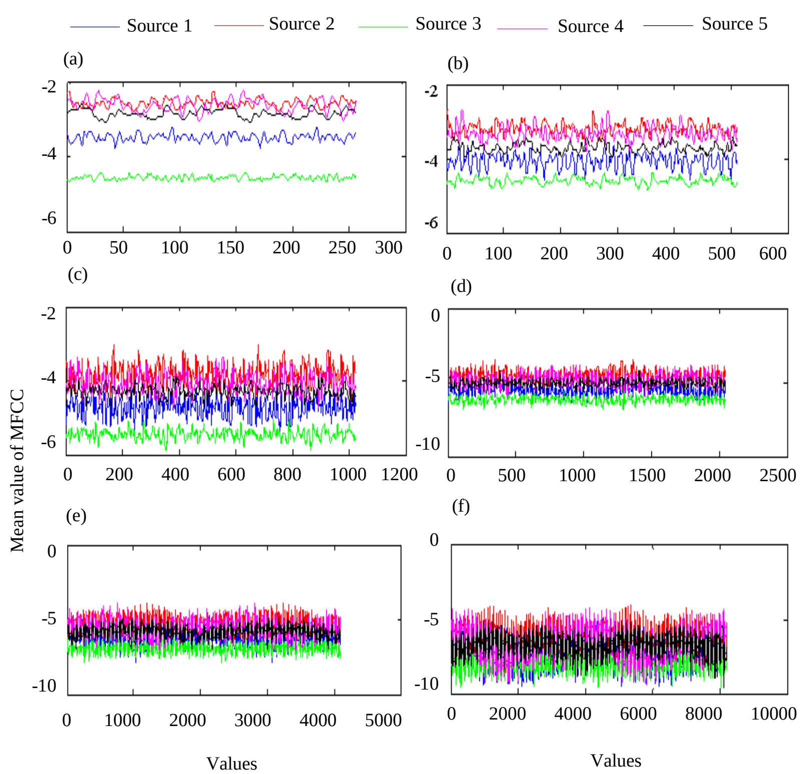

Fig. 2: Mean values of MFCC for different frame sizes of a) 2048; b) 1024; c) 512; d) 256; e) 128; f) 64 samples

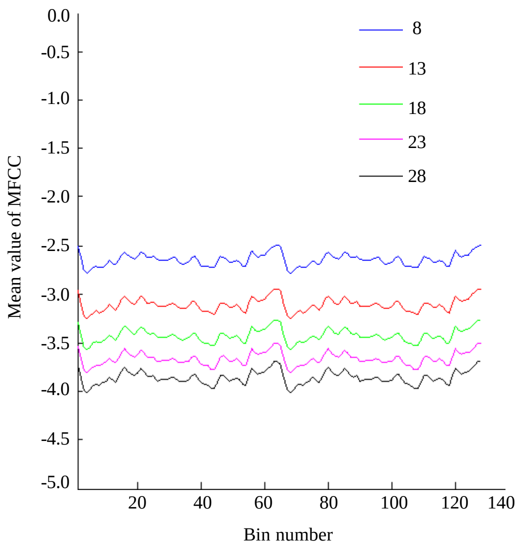

Fig. 3: Effect of number of MFCC's for a frame length of 1024 sample

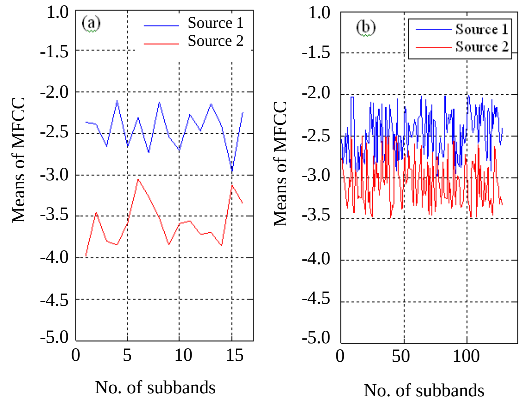

Fig. 4: Effect of number of subbands on MFCC mean value for a 2×2 mixture: a) 16 subbands and b) 128 subbands

In the STFT-based BSS scenario, assumethat there are L subbands and in each subband N sources. The separated sources in each subband l are denoted by $X_i(f_l, \tau)$ where $1 \le i \le N$ and $1 \le l \le L$ and f_l represents the subband frequency in subband l and τ represents the time-frame index. Now for each subband, the separated sources $X_i(f_l, \tau)$ are transferred back to time-domain to obtain the wavelet $x_i(l, t)$ which is the time-domain contribution of the source signal I to subband l. Next, we extract the MFCC's from these wavelets associated with all N sources and calculate the mean valuesof the MFCC's, denoted by m_i (l, τ).

Now for the ith source in a subband k, k>1we calculate the distance between $m_i(k,\tau)$ and the mean value for all other sources: $m_j(k-1, \tau)$ $1 \le j \le N$:

$$q = \underset{j, 1\le j\le N}{\arg\min} \left\| m_i(k,\tau) \right| \left| m_j(k-1,\tau) \right\| \quad (5)$$

Now, it is can be concluded that the wavelets x_i (k,t) and x_q(k-1,t) are both associated with one source signal. The minimization of Eq. (5) is repeated for all neighboring subbands, k = 2, 3, ..., N and all sources $1 \le j \le N$. Now the indices obtained in this process are used in frequency-domain for permutation correction.

There are two potential problems with this simple method: the first is bad clustering due to occasional similarities between the mean values of MFCC's of some sources. As shown in Fig. 2, in some time segments, the mean values for some speech signals have a hard time in discriminating between the sources which leads to errors in ordering. The second problem isthe error propagation in subbands; if an error occursduringclustering in any subband k, this ordering error will propagate to all subbands afterwards.

For first drawback, if some mean values of MFCC in any time segment are the same or close, we mark these time segments and can use other methods for resolving the permutation ambiguity such as frequency-domain spectrum consistency (Fig. 1).

In order to prevent the subband error propagation, instead of comparing the mean value of MFCC of each source signal in a subband k with the previous subbnad k-1, we can compare the mean value of MFCC of present subbandk with the average of mean values of MFCC's, averaged over subbands 1-k-1.

For a 2×2 signal mixture, after separation in subbands with the ICA algorithm applied in 16 subbands, we obtain the mean values of MFCC's $m_i(k, \tau)$ for one frame of the two sources and these values are shown in Fig. 4.

RESULTS AND DUISCUSSION

Simulation . room center with equal separations. We used a Room Impulse Response (RIR) generation method (McGovern, 1994) to produce the channel impulse responses. The room is 2 m wide and 3 m long and 3 m height. The RIR's were generated with 150 msec reverberation time at 16 kHz sampling rate, with a reflection coefficient of 0.5 for floor, roof and walls. For speech signals, we used sentences from the TIMIT database with 16 kHz sampling rate with 30 sec of duration each. The speech signals were convolved with different RIR signals and mixed together. Afterwards by using STFT, the mixture signals were transferred to the

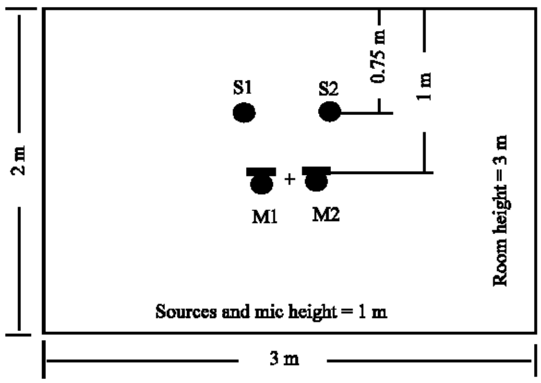

Fig. 5: Simulated test room arrangements and dimensions for 2×2 mixtures

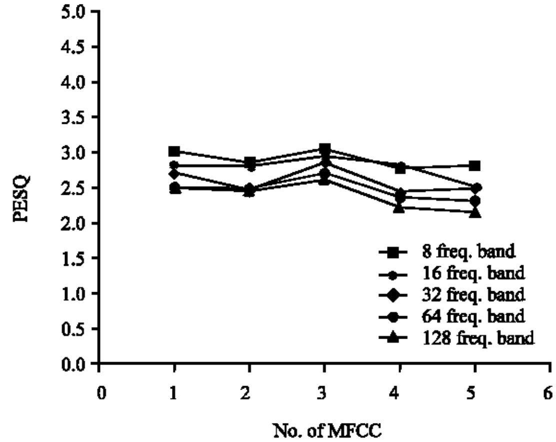

Fig. 6: Effect of number of frequency bands on separation quality

frequency subbands. The number of subbans were changed from 8-128 in various experiments as reported below. The STFT filter bank was a Near Perfect Reconstruction (NPR) one with a Hamming analysis/synthesis window, overlapping by 50% in time.

After separation in all subbands by the fast ICA algorithm, subband signals were transferred to time-domain and the MFCC-based permutation method of study 3 was applied to resolve the permutation problem in subbands as depicted in Fig. 1. Finally, the reordered subband signals were synthesized back in time-domain to obtain the source estimates.

In order to objectively evaluate the separation performance, we used three measures of: Source-to Artifacts-Ratio (SAR), Signal-to-Distortion-Ratio (SDR) and Source-to-Interference-Ratio (SIR) as defined in (Vincent *et al*., 2005). Also the quality of separation was measured by the Perceptual Evaluation of Speech Quality (PESQ) score as described in Nesta *et al*. (2001). As depicted in Fig. 2, as the time-frame length increases, the MFCC-based permutation performs better. Notice that with a given data length,increasing the frame length, inevitably will decrease the number of frequency bands.

Figure 6 shows the effect of number of frequency bands and MFCC's on the PESQ parameter. As depicted, the optimal number of MFCC's is 3 and beyond this value the PESQ score will decrease. Also, it is evident that the quality as measured by PESQ decreases as the number of frequency bands increase beyond 8.

The performance of the proposed method (with 8 subbands) was also compared with those of 4 main competing methods. The first method is the spectrum consistency approach of (Duran-Daz *et al*., 2012) the second one is the DOA approach, the third is the GGD of (Khanagha and Khanagha, 2009) and the forth one is the GD of Duran-Daz. For comparison the SIR, SDR, SAR and PESQ parameters for all these approaches are measured.

As shown in Fig. 7, the performance of proposed method is comparable to those of DOA and Spectral Consistency methods but it inferior to those of GGD and GD.

The main advantage of the proposed method is its simplicity and the resulting speed in resolving the permutation problem. In the spectrum consistency method, the spectra of all separated signals must be calculated and the correlation between all spectra must be calculated. In this case, if we have N k-point subband signals, we must calculate N k-point spectra and also calculate k-point correlations between all possible signal pairs which leads to computational cost increase. In the DOA approach for each signal one must calculate the angle of received signal and compare the angle of all signals in all subbands. In the GGD method at firstthe parameters of the Gaussian distribution must be estimated and compared in each subband. In the GD method, the divergence function must be calculated in each subband.

In our proposed method, we need to calculate the MFCC's for each subband frame, calculate its mean and then only compare this single value in all subbands which drastically decreases the computational complexity. Figure 8 shows the processing time for permutation correction phase for all these methods. As Fig. 8 depicts, the processing time of the proposed method is very small compared to other methods which leads to the increased speed of the overall signal separation.

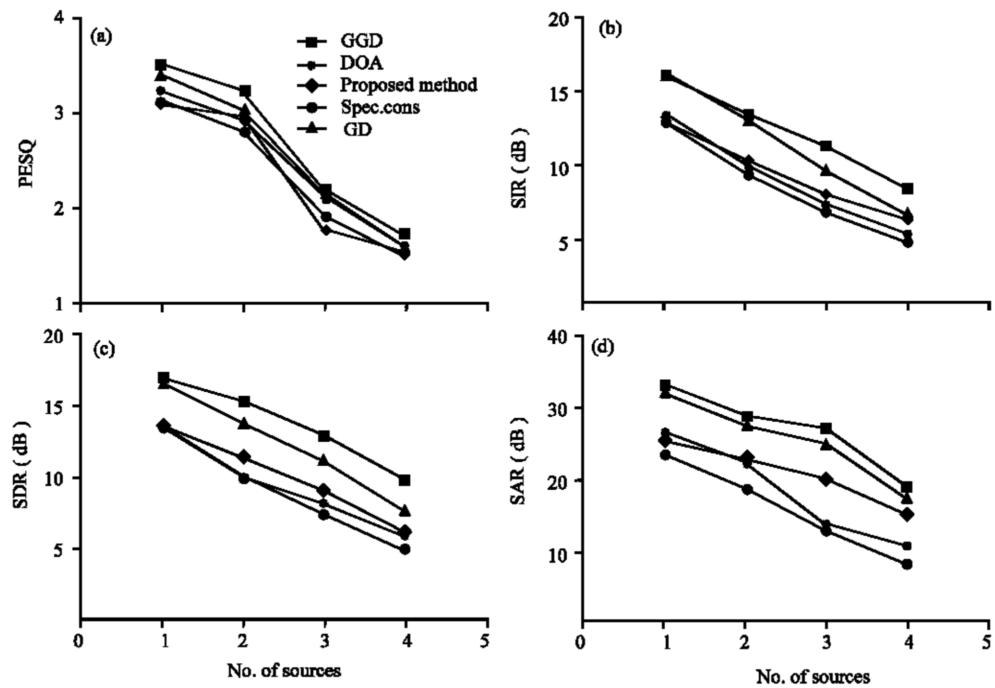

Fig. 7: Performance evaluations of 5 different separation methods verses number of sources: a) PESQ; b) SIR; c) SDR and d) SAR measures

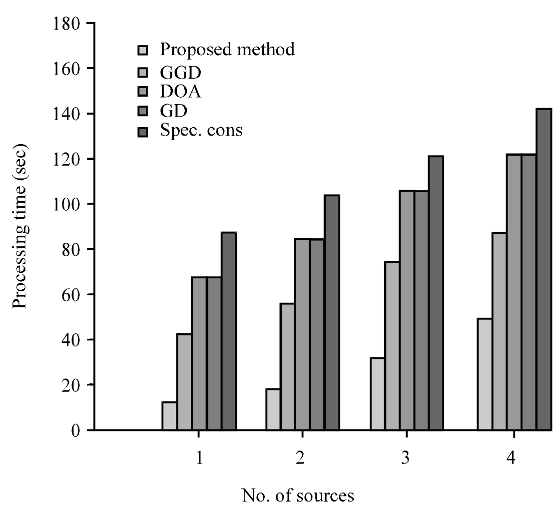

Fig. 8: Processing time of various methods for different number of sources

CONCLUSION

In this research, we presented a new method to resolve the permutation ambiguity problem in the subband-based convolutive BSS. The method is based on comparing the MFCC's of subband speech signals after individually transferring the subband signals to time-domain.

Comparing to four competitive methods, all operating in subband domain, the proposed method is superior to the DOA and spectrum consistency algorithms but inferior to compare to the GGD and GD algorithms, in terms of separation performance as measured by four objective measures of SIR, SAR, SDR and PESQ. In term of computational complexity, the proposed method is superior to all of the four competitive methods.

Currently, there is a trade-off between the performances of the proposed reordering algorithm and the BSS method applied in subb and domain. As the time-frame length increases, the separation performance of typical BSS algorithms such as the ICA degrade (Esmaeilbeig *et al*., 2016). In contrast, the proposed method for permutation correction performs better with longer time-frames. As a future work, we will try to reconcile this trade-off by improving the reordering algorithm to benefit from its computational simplicity while sacrificing less on the quality of the separated signals as measured by the PESQ score.

Finally, so far we have only employed the MFCC means value for permutation resolution. There are other possibilities and scenarios based on the MFCC statistics which can yield better performance.

REFERENCES

Anemuller, J. and B. Kollmeier, 2000. Amplitude modulation decorrelation for convolutive blind source separation. Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Signal Separstion, June 19-22, 2000, Helsinki, Finland, pp: 215-220.

Besic, N., G. Vasile, J. Chanussot and S. Stankovic, 2015. Polarimetric incoherent target decomposition by means of independent component analysis. IEEE. Trans. Geosci. Remote Sens., 53: 1236-1247.

Buchner, H., R. Aichner and W. Kellermann, 2005. A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics. IEEE. Trans. Speech Audio Process., 13: 120-134.

Duran-Daz, I., A. Sarmiento, S. Cruces and P. Aguilera, 2012. Generalized method for solving the permutation problem in frequency-domain blind source separation of convolved speech signals. J. Acoust. Soc. Amer., 131: 139-144.

Esmaeilbeig, M., H. Sheikhzadeh and F. Razzazi, 2016. A novel and fast algorithm for solving permutation in convolutive BSS, based on real and imaginary decomposition. Circuits Syst. Signal Process., 35: 4532-4549.

Gribonval, R. and S. Lesage, 2006. A survey of sparse component analysis for blind source separation: Principles, perspectives and new challenges. Proceedings of the 14th European Symposium on Artificial Neural Networks E SANN'06, April 26-28, 2006, D-Side Publisher, Bruges, Belgium, pp: 323-330.

Ikeda, S. and N. Murata, 1998. An approach to blind source separation of speech signals. Proceedings of International Conference on Artificial Neural Networks (ICANN'98), September 2-4, 1998, ICANN, Sweden, pp: 761-766.

Khanagha, V. and A. Khanagha, 2009. Solving global permutation ambiguity of time domain BSS using speaker specific features of speech signals. Proceedings of the IEEE Symposium on Industrial Electronics and Applications ISIEA09 Vol. 2, October 4-6, 2009, IEEE, Kuala Lumpur, Malaysia, ISBN: 978-1-4244-4681-0, pp: 1007-1011.

Matsuoka, K., 2002. Minimal distortion principle for blind source separation. Proceedings of the 41st Annual Conference on SICE Vol. 4, August 5-7, 2002, IEEE, Osaka, Japan, ISBN: 0-7803-7631-5, pp: 2138-2143.

Mazur, R. and A. Mertins, 2009. An approach for solving the permutation problem of convolutive blind source separation based on statistical signal models. IEEE. Trans. Audio Speech Lang. Process., 17: 117-126.

McGovern, S., 1994. Room impulse response generator. MathWorks Computer Software Company, Natick, Massachusetts, USA. http://www. mathworks.com/matlabcentral/file exchange/5116-room-impulse-response-generator?focused=3772529&tab=function.

Mukai, R., H. Sawada, S. Araki and S. Makino, 2006. Blind source separation of many signals in the frequency domain. Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP Vol. 5, May 14-19, 2006, IEEE, Toulouse, France, ISBN: 1-4244-0469-X, pp: 969-972.

Nesta, F., P. Svaizer and M. Omologo, 2011. Convolutive BSS of short mixtures by ICA recursively regularized across frequencies. IEEE Trans. Audio Speech Lang. Process., 19: 624-639.

Pendharkar, G., G.R. Naik and H.T. Nguyen, 2014. Using blind source separation on accelerometry data to analyze and distinguish the toe walking gait from normal gait in ITW children. Biomed. Signal Process. Control, 13: 41-49.

Sarmiento, A., I. Duran-Diaz, A. Cichocki and S. Cruces, 2015. A contrast function based on generalized divergences for solving the permutation problem in convolved speech mixtures. IEEE. ACM. Trans. Audio Speech Lang. Process., 23: 1713-1726.

Sawada, H., R. Mukai, S. Araki and S. Makino, 2004. A robust and precise method for solving the permutation problem of frequency-domain blind source separation. IEEE Trans. Speech Audio Process, 12: 530-538.

Sawada, H., S. Araki and S. Makino, 2011. Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. IEEE. Trans. Audio Speech Lang. Process., 19: 516-527.

Serviere, C. and D.T. Pham, 2006. Permutation correction in the frequency domain in blind separation of speech mixtures. EURASIP. J. Appl. Signal Process., 2006: 177-177.

Vincent, E., R. Gribonval and C. Fevotte, 2006. Performance measurement in blind audio source separation. IEEE. Trans. Audio Speech Lang. Process., 14: 1462-1469.

Won, Y.G. and S.Y. Lee, 2008. Convolutive blind signal separation by estimating mixing channels in time domain. Electron. Lett., 44: 1277-1279.

Xiong, K., Z. Liu, Z. Liu, D. Feng and W. Jiang, 2014. Underdetermined DOA estimation of multi-path signals based on ICA and sparse reconstruction. Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 4-9, 2014, IEEE, Florence, Italy, ISBN: 978-1-4799-2893-4, pp: 2233-2236.

Yu, K., K. Yang and Y. Bai, 2014. Estimation of modal parameters using the sparse component analysis based underdetermined blind source separation. Mech. Syst. Signal Process., 45: 302-316.