A New Non-Maximally Decimated UEPS for Blind Source Separation

Salah Al-Din Badran, Abdel-Hamid Soliman,Ismail Shahin

Abstract—The problem of blind source separation (BSS) refers to recovering original audio signals, called source signals, from the mixed signals, called convolutive mixtures or observation signals, in a reverberant environment. The mixture is a function of a sequence of original speech signals mixed in a reverberant room or ambient. The objective is to separate mixed signals to obtain the original signals without degradation and without prior information of the features of the sources, such as the locations, the spectral nature or the mixing method. The strategy used to achieve this objective is to use multiband schemes. Multiband schemes work at a lower sampling rate, have less computational cost and a quicker convergence than the full-band approach. For more challenging scenarios, an unequal-passbands scheme (UEPS) with oversampled decimation is proposed for blind source separation application; the greater number of bands, the more efficient the separation. The results are compared to the currently best performing method and three standard methods and show a significant improvement in the computational saving and convergence speed.

Keywords—Convolutive, mixture, speakers, separation, NMD_UEPS.

I. Introduction

THE speech signals of multiple speakers in a room are mixed with other audio sources, such as music [1], and noises [2]. Different methods have proposed to separate the mixed signals for different applications, such as noise cancellation and speech processing for people with hearing difficulties [3]. Many researchers have suggested different statistical solutions for blind separation applications in audio signal processing and cognitive psychology [4].

Nowadays, speech recognition technology is desirable in speaker identification applications [1]. Though, the recognition worsens rapidly when many people talk at the same time or background noise is added to the speech signals (e.g., air-conditioning, engine noises of machinery, noise, etc.). This challenge motivates researchers to find a way to create new devices and methods that extract the original speech signals from the undesired signals. This scenario which seeks to detect an individual speech among mixtures captured by microphones [5],.

The motivation for this study are good performing results, which showed that the UEPS has a superior performance over the EPS [5] for applications that need a huge number of parameters, such as system identification and convolutive mixtures BSS.

The principle of BSS can be stated as follows: it is required to reconstruct the N sources signals that are received by M mixtures. The processing is then blind, i.e., it has no prior knowledge of the mixing system and the sources are unobservable. It is necessary to have an additional assumption; otherwise, the blind source separation appears as an unsolvable problem. This is why most of the BSS techniques assume that the sources are independent. Initially, it was modelled based on a single product, i.e., the sensors receive at each instant a linear combination of the source signals. Later, modelling close to reality was introduced. One of these models considers transmission channel as a system based on a filtering operation. In other words, the captured signals are linearly dependent on both the source and their delayed versions. The mixing system is said to be a linear convolutive system.

Different approaches have offered different solutions to the BSS problem, some of them were stated above. One approach is to use the Independent Component Analysis (ICA) technique, formalized by Comon in [7].

Our proposal is a blind source separation approach for reverberant rooms (convolutive mixtures) that uses NMD_UEPS filter bank scheme with real-coefficients, to extract the original speech with a better convergence speed and lower computational complexity. To achieve this goal, this proposed structure uses a novel normalisation approach for the adaptation algorithm. Multiple bands are used in the separation system. The rationale for using multiple bands is to enhance the convergence speed and decrease the computational cost as it is compared to the full-band algorithm.

II. Multiband Blind Source Separation in Time Domain

We first propose a further reduction to equation [8]:

$$R_{nn}(l) \approx diag\{R_{nn}(l)\} \tag{1}$$

of the normalisation factor to a scalar. In this case,

$$R_{nn}(l) \approx (y_n(l))^T y_n(l) I \tag{2}$$

with $y_n^i(l)$ is given by equation (3):

$$y_n(l) = \sum_{m=1}^{M} \tilde{X}_{K_m}^{\ T}(l)\tilde{\omega}_{mn} = [y_n(lU) \ \ \cdots \ \ y_n(lU+K-1)]^T \tag{3}$$

S. B. is with Sohar University, P.O. Box: 44, P. Code 311, Sohar Sohar, Oman (sbadran@soharuni.edu.om)

A. S. is with the School of Engineering, Staffordshire University, Mellor Building, College Rd, Stoke-on-Trent, ST4 2DE, UK (e-mail: a.soliman@staffs.ac.uk).

I. S. is with the Electrical and Computer Engineering Department, University of Sharjah, Sharjah, UAE.

This reduction will significantly decrease the computational cost. For higher order separation filters, equation (4):

$$N_{blockFB} = 2M^2KUD + M^2(K + DK) + M^2\left(\left(\sum_{c=1}^{U} c\right) + U\right) \quad (4)$$

can be simplified by considering only the dominant terms, resulting in

$$N_{blockFB} \approx 4M^2U^3 \quad (5)$$

The expression above was obtained by considering the normalisation factor of the simplified equation (2), $D = U$ and $K = 2D$.

Our proposal is to use a NMD_UEPS filter bank with octave bands. The proposed blind source separation system is with two inputs and two outputs considering an UEPS filter bank with L bands. This is a modified version of the proposed work presented in [8]. The signals at the separation filters inputs of each band $\tilde{\omega}_{mn}^{i}(k)$ are decimated by half of the maximal decimation factor to minimise the alaising throughout the adaptation process of the coefficients. The output signals for the separation filter of each band are decimated by a factor of 2 to restore the maximal sampling rate of the scheme before the reconstruction stage of the output signal.

For PR in a two-channel cosine-modulated filter bank with EP, we can write

$$p^{\ell}(k) = 2p_m(k)\cos\left[\frac{\pi}{2}(\ell + 0.5)\left(k - \frac{K_M - 1}{2}\right) + \varphi_{\ell}\right] \quad (5)$$

$$q^{\ell}(k) = 2p_m(k)\cos\left[\frac{\pi}{2}(\ell + 0.5)\left(k - \frac{K_M - 1}{2}\right) - \varphi_{\ell}\right] \quad (6)$$

where $\varphi_{\ell} = (-1)^{\ell}\frac{\pi}{4}$ for $\ell = 0.1$ and $0 \leq k \leq K_M - 1$. For a two-channel octave-bands filter bank, (equations (5) and (6)), are equivalent to the EP with L-band tree filter bank. The number of coefficients of each separation filter in i^{th} band $\omega_{mn}^{i}(k)$ must be at least [8]

$$U_i = 2\left[\frac{U - 1 + K_{Q_i}}{\mathcal{F}_i}\right] \quad (7)$$

for the i^{th} synthesis filter bank, the order is K_{Q_i}.

Every separation filter can be regulated individually by its own coefficients. The full-band approach in equation:

$$\widetilde{W}(l + 1) = \widetilde{W}(l) - \gamma\nabla_{\widetilde{W}}^{NG}\varsigma(l) \quad (8)$$

is applied to the multiple bands, and the update equation of the filter parameters of the i^{th} band is expressed as follows:

$$\widetilde{W}^{i}(r) = \widetilde{W}^{i}(r-1) - \frac{2}{b_i}\left\{\sum_{l=1}^{b_i}\begin{bmatrix}\widetilde{W}_{12}^{i}R_{21}^{i}R_{11}^{i-1} & \widetilde{W}_{11}^{i}R_{12}^{i}R_{22}^{i-1}\\ \widetilde{W}_{22}^{i}R_{21}^{i}R_{11}^{i-1} & \widetilde{W}_{21}^{i}R_{12}^{i}R_{22}^{i-1}\end{bmatrix}\right\}\begin{bmatrix}\gamma_1^{i}I & 0\\ 0 & \gamma_2^{i}I\end{bmatrix} \quad (9)$$

where

$$R_{mn}^{i}(l) = \left[Y_m^{i}(l)\right]^{H} Y_n^{i}(l) \quad (10)$$

and

$$Y_n^{i}(l) = \begin{bmatrix} y_n^{i}(lU_i) & \cdots & y_n^{i}(lU_i - D_i + 1)\\ y_n^{i}(lU_i + 1) & \ddots & y_n^{i}(lU_i - D_i + 2)\\ \vdots & \ddots & \vdots\\ y_n^{i}(lU_i + K_i - 1) & \cdots & y_n^{i}(lU_i - D_i + K_i)\end{bmatrix} \quad (11)$$

The above matrices have dimensions D_i x D_i (with $1 \leq D_i \leq U_i$) and $K_i \times D_i$ (with $K_i \geq D_i$), respectively, b_i is the number of blocks, K_i is the size of each block, and γ_n^{i} is the n^{th} adaptation step of the i^{th} band, and r is the number of iterations for all bands.

For the purpose of lower computational cost, the normalisation factor $R_{nn}^{i-1}(m)$ can be simplified to a scalar (see equation (2)). In this case,

$$R_{nn}^{i}(l) \approx \left[y_n^{i}(l)\right]^{T} y_n^{i}(l)I \quad (12)$$

with $\boldsymbol{y}_n^{i}(l)$ corresponding with the first column of the matrix of delay-time block (equation (11)).

The implementation of the proposed multiband version is more flexible than the full-band version. For example, it can work with separation filters of different lengths and independent learning rates of various bands.

III. Results

In our experiment, discarding the use of a filter bank with complex coefficients, we contrast only the full-band and multiband schemes, both with approximate normalisation methods according to equations (2() and (11). The UEPS is tested with octave-band filter banks and perfect reconstruction for $L = 4$ bands Fig. 1 demonstrates $P_{i,j}(z)$ for of the multiband UEPS. Fig. 1 also $P_{i,j}(z)$ for the multiband UEPS.

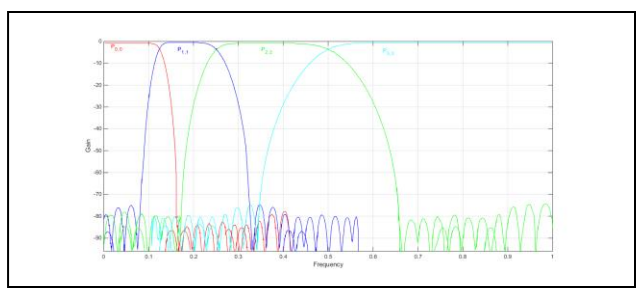


Fig. 1: $P_{i,i}(z)$ for $L = 4$

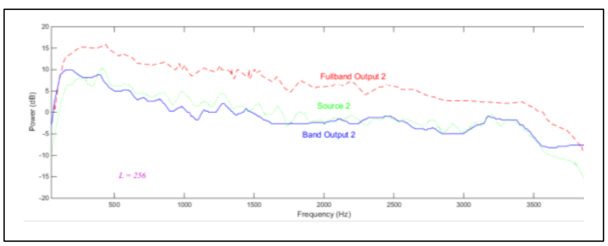
Our proposed algorithm is validated with some standards and another best performing work using PESQ metrics as a benchmark for testing multiband scheme for the FB scheme.

The performance of the BSS in $i = 3$ is always inferior to the other bands, because of employing a reduced filter's length, leading to a minimised signal-to-interference-ratio.

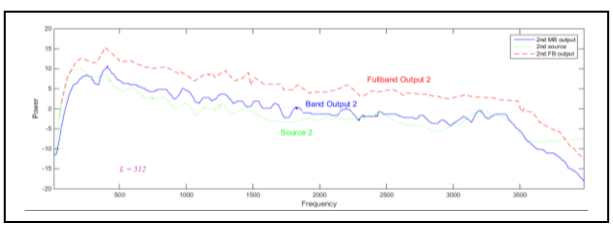
The perceptual evolutions are carried out by one female and one male English speakers using a PESQ (Perceptual Evaluation Signal Quality) tool [9] and the results are compared with three standards, that are used as a benchmark, and one best performing work [10], see Table 1. Fig. 2 depicts the estimates of the original sources of full-band and multiband for following lengths of mixtures: L = 256, 512, and 1024, and their spectral power. Fig. 3 provides visualization for the spectrum of the first and second original sources compared to the spectrum of the estimated signals at the outputs of the first and second band, respectively. The obtained outcomes demonstrate how resilient the proposed scheme is to the scaling of the output signals and the whitening of the sources.

TABLE 1: THE PESQ OF THE PROPOSED METHOD IS COMPARED TO THREE STANDARDS AND BEST PERFORMING WORK [10].

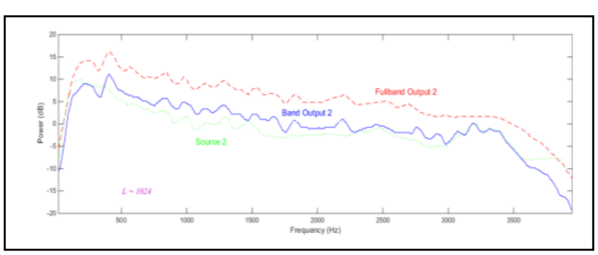
	FastICA	JADE	SOBI	[10]	Proposed Method
PESQ of the Female Speaker	3.25	3.29	2.58	3.29	3.31
PESQ of the Male Speaker	4.27	4.14	3.45	4.38	4.43



(a) Mixing filter length 256



(b) Mixing filter length 512



(c) Mixing filter length 1024

Fig. 1: FB and MBs spectrum with different filter lengths

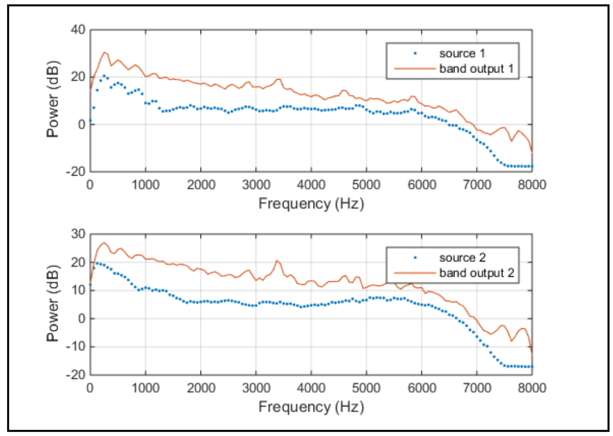


Fig. 3 The spectra of the original sources 1 and 2, and the separated signals at bands y1 & y2.

IV. CONCLUSION

Simulations are performed on speech signals. These simulations demonstrate the superior performance of the proposed multiband scheme over the full-band scheme. The results show a significant improvement in the proposed algorithm in comparison to three standards and a new best performing algorithm.

Simulations are performed on speech signals. These simulations demonstrate the superior performance of the proposed multiband scheme over the full-band scheme.

[1] Salah Al-Din Ibrahim Badran, Samad Ahmadi, Pooneh Bagheri Zadeh, and Ismail Shahin, "A Novel Tap Selection Design for Filters in Unequal-Passbands

Scheme," in The Eleventh International Conference on Digital Telecommunications (ICDT 2016), Lisbon, 2016, pp. 22-26.

[2] Salah Al-Din Badran, Samad Ahmadi, and Dylan Menzies, "A Novel Technique to Solve the Convergence Speed in the Post Nonlinear Models," International Research on Technology, Engineering and Science (IRTES), vol. 5, no. 1, pp. 106- 110, August 2015.

[3] Salah Al-Din Badran, Samad Ahmedi, Dylan Menzies, and Ismail Shahin, "Efficient Separation for Convolutive Mixtures," International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 8, no. 5, pp. 718 - 722, May 2014.

[4] Salah Al-Din Badran, Samad Ahmadi, and Dylan Menzies, "A New Approach for the Blind Separation of Sources in Reverberating Environments," International Research on Technology, Engineering and Science (IRTES), vol. 3, no. 1, pp. 86 - 90, April 2013.

[5] Simon Haykin and Zhe Chen, "The Cocktail Party Problem," *Neural Computation*, vol. 17, no. 9, pp. 1875-1902, September 2005.

[6] Simon Haykin, *Adaptive Filter Theory*, 4th ed. New Jersey, United States: Prentice Hall, 2002.

[7] Pierre Comon, "Independent Component Analysis, a New Concept," *Signal Processing, Elsevier*, vol. 36, no. 3, pp. 287 - 314, March 1994.

[8] Salah Al-Din Ibrahim Badran, Samad Ahmadi, Pooneh Bagheri Zadeh, and Ismail Shahin, "A Novel Tap Selection Design for Filters in Unequal-Passbands Scheme," in *The Eleventh International Conference on Digital Telecommunications (ICDT 2016)*, Lisbon, 2016, pp. 22-26.

[9] International Telecommunications Union (ITU). (2005, November) Recommendation P.862. [Online]. http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en.

[10] Ibrahim Missaoui and Zied Lachiri, "Blind speech separation based on undecimated wavelet packet perceptual," *International Journal of Computer Science Issues*, vol. 8, no. 3, pp. 1694-0814, May 2011.