SPEECH ENHANCEMENT USING ICA WITH BESSEL FEATURES

Balakrishna L, Ashok Kumar P.V,
Department of Electrical Engineering
Blekinge Institute of Technology
Karlskorna, Sweden
bkla08@student.bth.se, vapo08@student.bth.se,

Chetana Prakash, Surayakanth V. Gangashetty
Speech and Vision Laboratory
International Institute of Information Technology,
Hyderabad, India
chetana@research.iiit.ac.in, svg@iiit.ac.in

***Abstract*—The Independent Component Analysis with Reference (ICA - R) also called as constrained ICA (cICA) extracts only the desired source signals from the mixture of source signals by incorporating some prior information into the separation process. To overcome the problem of designing the reference signal when there is no prior information about the desired signal in the cICA, an improved method is proposed to use a different speech signal generated by the same physical source. The cICA is extended to use Bessel coefficients of the observed signals and the reference signal for processing as they converge faster than the other transformations. Since the Bessel functions provide the desired properties, efficient in representing speech signals, less memory storage they have been exploited in speech processing [1]. The results demonstrate the efficiency of the proposed method.**

***Keywords:* Bessel functions, Independent component analysis, Speaker recognition, Empirical Mode Decomposition, Speech enhancement.**

I. INTRODUCTION

ICA belongs to a class of Blind Signal Separation (BSS) methods for separating data into underlying informational components, where such data can take the form of images, sounds etc... As the name suggests, ICA separates a set of signal mixtures into a corresponding set of statistically independent component signals or source signals. Here the property of being unrelated is of fundamental importance, because it can be used separate mixture of source signals. The defining feature of the extracted signals is that each extracted signal is statistically independent of all other extracted signals i.e. each extracted signal will be generated by a different physical process. [2]. ICA has been widely used in various fields, e.g., speech signal processing, image processing etc. Compared to the conventional ICA algorithm, the technique of the cICA provides a general framework to incorporate the additional requirements or prior information, e.g., statistical properties or rough templates of the sources. With the prior information, the cICA algorithm usually has a better performance. ICA-R is very useful in many applications such as automatic speaker verification, speaker identification and so on [8]. In the algorithm, the norm between the references and the estimated signals is used as the inequality constraints in the contrast function to extract the desired signals. The ICA-R extracts only the desire source signal which is the closest one, based on prior knowledge of the reference signal. Here the reference signal plays a major role in extraction of the desired signal. Therefore by providing or constructing the proper reference signal helps in better extraction of the desired signal. The crucial problem is the design of reference signal in advance which is close to the desired signal when the desired source signal is very week in mixed signals and also when there is no prior information about the desired source signal.

In this paper the ICA-R algorithm is extended to use Bessel coefficients of the observed signals and the reference signal for processing as they are more efficient in representing speech-like waveform. To overcome the problem of designing the reference signal when there is no prior information available about the desired signal here we proposed to use a different speech signal generated by the same physical source as the reference signal for the algorithm. The adaptive solutions using the Newton-Like learning are proposed in the algorithm to solve the constrained optimization problems.

The paper is organized as follows: In Section 2, we describe the Bessel expansion of signals. Section 3, describes the various types of ICA along with the ambiguities of ICA-R. In Section 4, we discuss one of the existing approaches for ICA-R with EMD based reference. The proposed approach for ICA-R based on Bessel features is described in Section 5. The study on ICA-R is discussed in Section 6 by taking suitable examples. In this section we also study the performance of the proposed approach.

II. BESSEL EXPANSION

The Bessel expansion is suitable for non-stationary signals representation [4]. The wave equation inside cylindrical structures (tubes) includes the first kind of Bessel function. The vocal tract can be modeled as organ pipe like cylindrical tubes with a sound source at one end (the larynx or voice box) and open at other ends (the lips or nose). It is a good reason to choose the first kind Bessel functions in terms of naturalness for representing the sounds produced in the vocal tract it could be approximated as acoustic tubes for short-time intervals analysis [5].

A. Coefficients Computation using Bessel Expansion

The zero-order Bessel series expansion of the signal $x(t)$ considered over some arbitrary interval (0, a) is expressed as [6]:

$$x(t) = \sum_{m=1}^{Q} C_m J_0\left(\frac{\lambda_m}{a}t\right) \qquad (1)$$

Where $\{\lambda_m$, m=1, 2, 3…$\}$ are the ascending-order positive roots of $J_0(\lambda) = 0$, and $J_0\left(\frac{\lambda_m}{a}t\right)$ are the Zero-Order Bessel functions. The sequence of Bessel functions forms an orthogonal set on the interval $0 \leq t \leq a$ with respect to the weight t, that is:

$$\int_0^a t J_0\left(\frac{\lambda_m}{a}t\right) J_0\left(\frac{\lambda_n}{a}t\right) dt = 0, \quad \text{for } m \neq n \tag{2}$$

Using the orthogonality of the set $\{J_0\left(\frac{\lambda_m}{a}t\right)\}$, the Bessel coefficients C_m are computed by using the following equation:

$$C_m = \frac{2\int_0^a t x(t) J_0\left(\frac{\lambda_m}{a}t\right) dt}{a^2[J_1(\lambda_m)]^2} \tag{3}$$

With$1 \leq m \leq Q$, where Q is the order of the Bessel expansion and $J_1(\lambda_m)$ is the first–order Bessel functions. The Bessel expansion order Q must be know a priori. The interval between successive zero-crossing of the Bessel functions $J_0(\lambda)$ increases slowly with time and approaches π in the limit. If order Q is unknown, then in order to cover full signal bandwidth, the half of the sampling frequency, Q must be equal to length of the signal.

The Bessel series coefficients C_m are unique for a given signal; similar as the Fourier series coefficients are unique for a given signal [7]. However, unlike the sinusoidal basis functions in the Fourier series, the Bessel functions decay over time. This feature of the Bessel functions makes the Bessel series expansion suitable for non-stationary signals [6-8]. The Bessel features can be used in many applications such as speech segmentation, speaker verification, speaker identification, and language identification and so on.

III. Independent Component Analysis

The ICA recovers a set of unknown mutually independent sources signals from their observed linear mixtures. Suppose *M* independent source signals $s(t)$exist, and *N* observed mixtures of the source signals *x* (t) then $s(t)$ and $x(t)$ are given by:

$$s(t) = [s_1(t) \ldots \ldots s_M(t)]^T \tag{5}$$

$$x(t) = [x_1(t) \ldots \ldots x_N(t)]^T \tag{6}$$

The linear ICA assumes that these mixtures are linear, instantaneous, and noiseless.

$$x(t) = As(t) \tag{7}$$

A is a $N \times M$ mixing matrix that contains the mixing coefficients. The goal of ICA is to find a $M \times N$ de-mixing matrix *W;* such that *M* output signals were extracted.

$$y(t) = Wx(t) = WAs(t) = PDs(t) \tag{8}$$

$D \in R^{M \times M}$is a permutation matrix and $D \in R^{M \times M}$ is diagonal scaling matrix. The ICA is related to blind source separation (BSS), without knowing information about the original signal we are separating the mixed signal [9].

A. ICA with Reference

The ICA – R also called as constrained ICA (cICA) is good for extracting several source signals from a large number of observed mixtures by providing a reference signal which should be closely related to the desired source signal. The adaptive solutions using the Newton-like learning rule is used to solve the constrained optimization problem.

Instead of separating all *M* number of independent sources from *N* mixed signals , ICA-R extracts *L (L<M)* number of desired sources from *N* mixed signals by incorporating some a priori information into the ICA learning algorithm as reference signals [10]. These reference signals are denoted by $r(t) = [r_1(t), \ldots, r_L(t)]^T$, carry some information of the desire sources but not identical to corresponding desired signals. We briefly describe the one-unit ICA-R. It finds one weight vector *W*, so that the output signal, $y(t) = w^T x(t)$ recovers desired source $s^*(t)$ by using $r(t)$ as the reference signal [9, 10].

$$J(y) = \rho[E\{G(y)\} - E\{G(v)\}]^2 \tag{9}$$

Here ρ is a positive constant, ν is a Gaussian variable having zero mean and unit variance, $G(.)$ can be any non-quadratic function. The closeness between the ICA-R output *y* and the reference signal *r* is measured by $\varepsilon(y, r)$, which has a minimal value when $y = PDs^*$. A threshold ξ is used to distinguish the desired source s^* from other source signals such that $g(w) = \varepsilon(w^T x, r) - \xi \leq 0$ is satisfied only when y=PDs^* among all source signals, g(w) as feasible constraint to the contrast function, the problem of one unit ICA-R can be modeled in the framework of constrained independent component analysis [11] .

$$\text{Maximize} \quad J(y) = \rho[E\{G(y)\} - E\{G(v)\}]^2 \tag{10}$$

$$\text{Subject to} \quad g(w) \leq 0,\ h(w) = E\{y^2\} - 1 = 0 \tag{11}$$

Where $h(w)$is the equality constraint used to ensure the contrast function $J(y)$ and the weight vector *w* are bounded. In a Newton-like leaning algorithm is derived by finding the maximum of an augmented Lagrangian function corresponding [10, 11]. The updated weights are given by:

$$w_{k+1} = w_k - \eta \frac{R_{xx}^{-1} L'_{w_k}}{\delta(w_k)} \tag{12}$$

Where k is iteration index, η is the learning rate, R_{xx} is the covariance matrix of the input mixtures x, also

$$L'_{w_k} = \rho E\{xG'_y(y)\} - 0.5\mu E\{xg'_y(w_k)\} - \lambda E\{xy\} \tag{13}$$

$$(w_k) = \rho E\{xG''_{y^2}(y)\} - 0.5\mu E\{xg''_{y^2}(w_k)\} - \lambda \tag{14}$$

Where $G'_y(y)$ and $G''_{y^2}(y)$ are the first and second derivatives of $G(y)$ with respect to *Y*, and $g'_y(w_k)$ and $g''_{y^2}(w_k)$ are those of$g(w_k)$. The optimum multipliers μ and λ are found by iteratively updating the gradient-ascent method. Here γ is the scalar penalty parameter. Further,

$$\mu_{k+1} = \max\{0. \mu_k + \gamma g(w_k)\} \tag{15}$$

$$\lambda_{k+1} = \lambda_k + \gamma\, h(w_k). \tag{16}$$

B. Ambiguities of ICA-R

There are several important issues that should be noticed [13]. The major is choosing the value for the threshold, ξ. When the threshold value is low the ICA-R algorithm will not converge and if the value is too large, the ICA-R may converge to other source signals because of some source signals whose closeness measure may be less than the threshold. So the threshold value should be carefully selected. Here the value is influenced by some factors. One is the choice of the closeness *ε(y, r).* A common choice is the mean square error (MSE) given by *ε(y, r)* = *E* $\{(y - r)^2\}$ and another is the correlation ε(*y, r*) $=-E\{yr\}$. The other factor affecting the threshold value is the choice of the reference signal$r(t)$. If the reference signal is very similar to the desired source signal, then the threshold value should be very small so that the algorithm can globally converge. Otherwise, the value should be large. The other issue is the design or choosing the reference signal. Until now, in most literature the main method is constructing a simple impulse signal by observing the waveform of sensor signals and/or by exploiting strong a priori information about the desired source signal [13]. In this method either the desired source signal is strong enough to be observed in a sensor signal or one should have sufficient priori information of the source signal. Thus, the method sometimes becomes more trivial and results in increased computational load and time. In the next session we discuss on the existing approach for ICA-R with EMB based reference.

IV. ICA-R WITH EMD BASED REFERENCE

In this approach the ICA-R algorithm extracts the target speech signal from the mixture of different speech signals by constructing the reference signal using Empirical Mode Decomposition (EMD) [3]. The EMD is a general nonlinear and non-stationary signal processing method. It decomposes a signal into a finite and often small number of intrinsic mode functions (IMF's), and can be used as a filter by reconstructing the original signal with partial IMFs [7]. Here the EMD is used to obtain the approximate envelope of the power spectrum of the desired speech. In this approach EMD is used as low-pass filter to obtain the envelope of the speech power spectrum. Here the speech signal is significantly enhanced by including prior information, i.e. the knowledge of the speech power spectra into the ICA separation process. This method greatly facilitates many applications such as automatic speech recognition and speaker identification. In the next section we had proposed a different approach of extracting the reference signal using Bessel features. Here the desired signal and the reference signal are two different speech utterances of a same speaker. Thus a suitable threshold value is easy to set, regardless of which closeness measure is adopted.

V. APPROACH FOR ICA-R USING BESSEL FEATURES

In this section we proposed an approach using ICA-R algorithm and Bessel features to extract the desired speech signal from the mixture of speech signals in scenarios when there is no prior information available about the desired signal. Here the reference speech signal and desired speech signal to be extracted are two different speech utterances generated from a same physical source. Our idea is to perform the processing of the ICA-R algorithm using the Bessel coefficients of the reference signal and the set of observed mixed signals to extract the desired speech signal. In this approach we calculated the Bessel coefficients of the reference and source signals by framing with 50% overlap. The calculated Bessel coefficients of the reference signal and set of mixed speech signals are passed as the inputs to the ICA-R algorithm for processing. The ICA-R algorithm extracts the Bessel coefficients of the desired signal from the set of coefficients of the mixed signals which are close to the coefficients of the reference signal. Here the threshold ξ is very critical to the convergence of the algorithm. Initially it was initialized to a very small value to avoid the algorithm going to a local optimum, and then is gradually increased to converge at the global maximum [11]. By performing the ICA-R algorithm using Bessel features on the mixed speech signal, the desired Bessel coefficients of the target speech signal are extracted. Using the extracted Bessel coefficients we re-synthesize the desired speech signal by performing inverse Bessel transformation and overlap-add method.

VI. RESULTS AND DISCUSSIONS

We have considered four speech signals each of different lengths with zero mean as source signals, as shown in Fig.1. Zeros were appended to the four signals to equal to max of the length among the four signals. Now each source signal has 88320 samples and has sampling frequency of 16000 Hz. The four equal length signals were mixed randomly using a mixing matrix A to obtain a set of mixed speech signals as shown in Fig.2. Our goal was to extract the target speech signal s3 as shown in Fig.3(a) using the ICA-R algorithm by the proposed approach by providing the reference signal r, as shown in Fig. 3(b).

As mentioned above the Bessel basis functions were calculated for the set of mixed signals and the reference signal by framing with 20 m.sec (320 samples) as frame size and 50% overlap and were passed as inputs to the ICA-R algorithm for processing.

In the ICA-R algorithm we use$G(y) = \log(\cosh(y))$, which is good general purpose function [12]. The closeness between the ICA-R output signal y and the corresponding reference signal r is defined as the mean square error and is given by:

$$\varepsilon(y,r) = E\{(y-r)^2\} \tag{17}$$

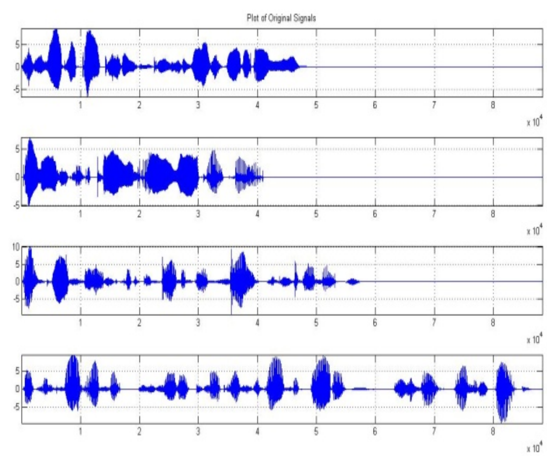

Figure 1: Four source signals (a) s1 (b) s2 (c) s3 (d) s4. The x-axis is time in m.sec and y-axis is amplitude

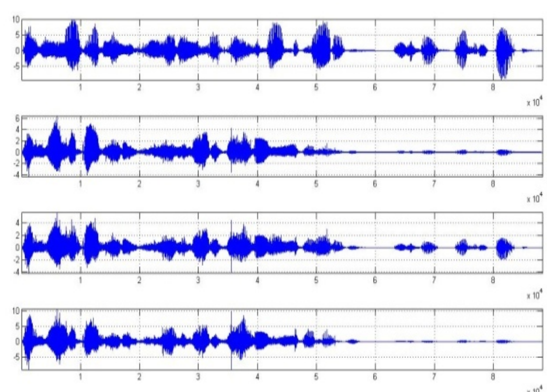

Figure 2: Randomly mixed source signals (a) microphone1 (B) microphone2 (c) microphone3 (d) microphone4. The x-axis is time in m.sec and y-axis is amplitude.

In ICA-R algorithm μ,γ,λ are the critical to the convergence of the algorithm of ICA –R [9, 10, 12 and 13]. It may be initialized with a small value to avoid the algorithm going to local optimum, and then is gradually increased to converge at the global maximum [11]. By performing the ICA-R with Bessel features on the four mixed signals the Bessel basis functions for the target speech signal were extracted. Performing inverse Bessel and overlap-add method the desired speech signal is re-synthesized and was shown in Fig.3(c). Observing Fig.3(a) and Fig.3(c) it can see that extracted speech signal is of good quality. The accuracy of the recovered signal y compared to the desired speech signal $s3$ is measured by the mean square error and Signal to Noise Ratio (SNR) in db:

$$SNR(dB) = 10\log_{10}\left(\frac{\sigma^2}{\mathrm{mse}}\right) \quad (18)$$

The computational results of the mean square error and the SNR for the two different methods are shown in Table 1. The computed results show that the ICA-R using Bessel features is better in extraction of the desired signal from mixture of different speech signals with less mean square error and with very high SNR index.

Table 1: Comparison of ICA-R with EMD and ICA-R with Bessel Features using MSE and SNR as measures

Performance Measures	**Methods**	
	EMD based Reference	***Bessel Features***
MSE	1.3400e-004	8.3400e-008
SNR(dB)	38.729	70.7883

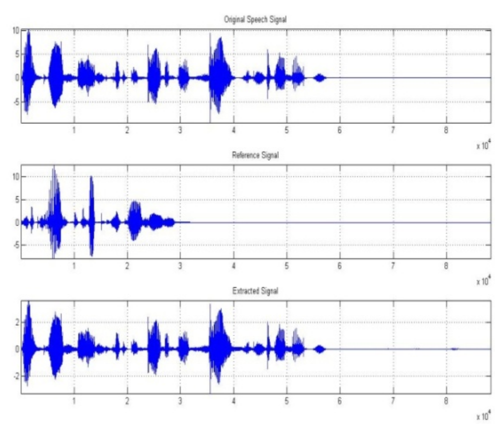

Figure 3: (a) Original/Desired speech signal. (b) Reference signal (different speech utterance of the target speech signal). (c) Extracted speech signal. The x-axis is time in m.sec and y-axis is amplitude

VII. Summary and Conclusions

In this paper, we propose an approach for extracting a desired speech signal from a mixed source signal using the ICA-R algorithm and Bessel features. Here the desired speech signal and the reference signal are two different speech utterances of a same speaker. In the current existing literature most of the methods deal with designing the reference signal with prior information of the target speech signal. In the proposed method we do not require any prior information about the desired speech signal that has to be extracted. This is very useful for many applications such as speaker verification, speaker identification and so on even when the desired signal is very week in mixed signals [8]. The performance analysis shows that the proposed method is effective. This shows that the computation done at the feature level i.e. the Bessel coefficients of the signals yields better results than on the sample values. An example of real world application where the proposed method is useful is the voice recognition and speaker verification login even in the presence of external disturbances and also in voice security applications.

References

[1] C. S. Chen, K. Gopalan, P. Mitra, "Speech signal analysis and synthesis via Fourier-Bessel representation", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 497-500, Tampa, Florida, March, 1958.

[2] P. Comon, "Independent component analysis: a new concept," *Signal Processing*, vol. 36, no. 3, pp. 287-314, 1994.

[3] Yongrui Zheng, Qiuhua Lin, Fuliang Yin, and Hualou Liang, "Speech enhancement using ICA with EMD -based reference," *ICA 2006*, LNCS 3889, pp. 739-246, 2006.

[4] J. Schroeder, "Signal processing via Fourier – Bessel series expansion," *Digital Signal Processing*, vol. 3, pp. 112-124, 1993.

[5] F. S. Gurgen, C. S. Chen, "Speech enhancement by Fourier-Bessel coefficients of speech and noise," *IEEE Proceedings*, vol. 137, pt. I, no. 5, October 1990.

[6] R .B. Pachori, P. Sircar, "Analysis of multicomponent AM-FM signals using FB-DESA method," *Digital Signal Processing,* vol. 20, pp 42-62. 2010.

[7] R. B. Pachori , D. J. Hewson, H. Snoussi, and J. Duchene, "Analysis of center of pressure signals using empirical mode decomposition and Fourier-Bessel expansion," in *Proceedings of the IEEE Tencon Conference*, Hyderabad, India, November 2008.

[8] Aapo Hyvärinen, Erkki Oja, "Independent component analysis: Algorithms and applications," Neural *Networks*, vol.13, no. 4-5, pp. 411-430, 2000.

[9] Lu. W. Rajapakse, J. C, "Constrained independent component analysis in advance in neural information processing systems," MIT press, pp 570- 576, 200.

[10] Lu. W. Rajapakse, J. C, "ICA with reference," proc. Third Int .Conf on ICA and Blind source separation (ICA2001), pp. 120-12, 2001.

[11] A. Hyvarinen, "Survey on independent component analysis," *Neural computation. Surveys,* vol. 2, pp. 94-128, 1999.

[12] Zhi-Lin Zhang, "Morphological constrained ICA for extracting weak temporally correlated signals", Neuro Computing 71, pp. 1669-1679, 2008.