Single Channel Audio Source Separation

BIN. GAO, W.L. WOO, and S.S. DLAY
School of Electrical, Electronic and Computer Engineering
University of Newcastle
Newcastle upon Tyne, NE1 7RU
UNITED KINDOM
Email: bin.gao@ncl.ac.uk, w.l.woo@ncl.ac.uk, s.s.dlay@ncl.ac.uk

Abstract:-Blind source separation is an advanced statistical tool that has found widespread use in many signal processing applications. However, the crux topic based on one channel audio source separation has not fully developed to enable its way to laboratory implementation. The main idea approach to single channel blind source separation is based on exploiting the inherent time structure of sources known as basis filters in time domain that encode the sources in a statistically efficient manner. This paper proposes a technique for separating single channel recording of audio mixture using a hybrid of maximum likelihood and maximum a posteriori estimators. In addition, the algorithm proposes a new approach that accounts for the time structure of the speech signals by encoding them into a set of basis filters that are characteristically the most significant.

Key-Words: - Single Channel Separation, Blind Source Separation, Characteristic Filters, ML, MAP

1 Introduction

Separating mixture of different signals has been focused by many researchers in computational auditory scene analysis (CASA) and Independent Component Analysis (ICA) [1]. Blind source separation (BSS) is a powerful methodology and a family of algorithms has already been developed [2-5]. Extension of BSS to solving nonlinear mixtures has also been introduced [6-11]. Parallel with this, recent advances have been made in blind source separation in the area of single channel signal analysis. The mixture of sources can be separated by given only single channel recording. The main idea for single channel blind source separation (SCBSS) is based on exploiting the inherent time structure of sources to generate basis functions in time domain. The basis functions imply inherent types of non-Gaussian characteristic signals. Maximum likelihood algorithm and Maximum a posteriori estimators are exploited for source separation.

The key point of SCBSS is to exploit a priori knowledge of sources such as the basis functions to generate sparse coding. The training sources are then projected onto a set of basis functions whose coefficients are as sparse as possible. The proposed separation algorithm use hybrid of maximum likelihood and maximum a posteriori estimators [12, 13]. to recover the independent components. If the basis functions are not chosen correctly, this will significantly deteriorate the performance of signal separation.

In this paper, the single channel mixing problem is considered and the objective is to provide optimal estimation of the source signals, the contribution of this paper is to provide a novelty method to extract the most significant characteristic features especially in terms of separating speech mixture source such a crux project based on the general ICA [14] and cross-correlation algorithms. In addition, the main affective factors are discussed and analogized based on separation results. The generalized hybrid of maximum likelihood and maximum a posteriori algorithm is then derived where to estimate original sources. In the proposed method, the real audio sources are exploited to test the performance of the algorithm.

Suppose the observed signal y^t mixed with two independent sources.

$$y^t = \lambda_1 x_1^t + \lambda_2 x_2^t \qquad (1)$$

Here λ_1 and λ_2 are the gain of independent source x_1^t and x_2^t separately. Alternatively, one could use a constrained gain given by $\lambda_1 + \lambda_2 = 1$. The superscripts indicate sample indices of time-varying signals. The gain is affected by many factors, for example attenuation of propagation between sensor and independent source.The independent component can be constructed as the product of basis functions and their coefficients which can be shown as:

$$\mathbf{x}_i^t = \sum_{k=1}^{M} a_{ik} s_{ik}^t = \mathbf{A}_i \mathbf{s}_i^t \qquad (2)$$

Here small length P with $P << T$ from independent source is employed to analysis. The time duration from t to $t+P-1$, an P dimensional column vector $\mathbf{x}_i^t = \left[x_i^t, x_i^{t+1} ...... x_i^{t+P-1}\right]^T$, here the vector symbol T means transpose. $M=P$ means that the delay samples of independent source equal to the number of basis functions. i refers to the source number (there are totally two independent sources). k expresses the type of basis function while basis function's coefficients can be expanded as $\mathbf{s}_i^t = \left[s_{i1}^t, s_{i2}^t .....s_{iM}^t\right]^T$. Since $M=P$, the matrix $\mathbf{A}$ has full rank so that the vector $\mathbf{x}_i^t$ and $\mathbf{s}_i^t$ are reversible in both directions.

$$\mathbf{x_i^t} = \begin{bmatrix} x_i^t \\ x_i^{t+1} \\ \vdots \\ x_i^{t+P-1} \end{bmatrix} = \left[\mathbf{a}_{i1}, \mathbf{a}_{i2}, \ldots \mathbf{a}_{iM}\right] \times \begin{bmatrix} s_{i1}^t \\ s_{i2}^t \\ \vdots \\ s_{iM}^t \end{bmatrix} \quad (3)$$

The inverse of basis matrix $\mathbf{W}_i = \mathbf{A}_i^{-1}$, then the expression 2 is transformed as:

$$\mathbf{s}_i^t = \mathbf{W}_i \mathbf{x}_i^t \quad (4)$$

Hence the generated model satisfies the typical ICA linear combination model. In this paper, algorithms derived from maximum likelihood (ML) and maximum a posteriori (MAP) [2-5] estimators are exploited to solve the single channel blind source separation problem.

2 Separation Algorithms

Firstly, the following figure shows an example of basis functions of real music signal. These basis functions are known *a priori* or obtained as part of the solution to the SCBSS [12, 13]. In this paper, it is assumed that the basis functions have already been obtained using any standard ICA algorithm.

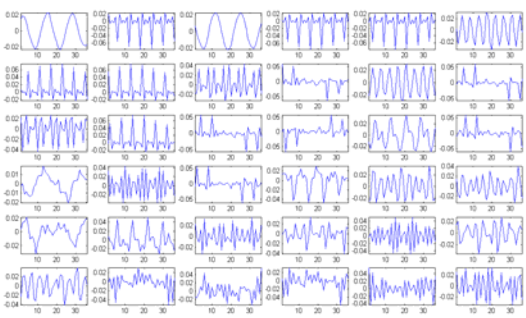


Fig. 1: Real music basis functions **A** (36 basis are shown)

In Figure 1, it is obvious that each of basis functions has non-Gaussian characteristics. This is the stage where we could find out the how of original source signals have been mixed. Since the set of basis functions can be obtained before source separation, the variables $\mathbf{A}_i$ as expression 2 are known. In addition, the basis coefficient s_i^t density can be estimated by exploiting non-Gaussian density function [15]. It is a very natural measure for independence if the joint density $f(x_i^t)$ equal to its marginal densities. The method which decreases the mutual information is by maximizing the marginal densities of the transformed coordinates for the given training data.

$$\Pr(x_i^t \mid \mathbf{W}_i) \cong G(s_i^t \mid \boldsymbol{\theta}_i)\left|\det \mathbf{W}\right| \quad (5)$$

where $\left|\det \mathbf{W}\right|$ is the absolute value of Jacobian of the transformation. In this case the transformation is linear and basis coefficients are assuming independent, therefore the log likelihood can be expressed as:

$$\log p(x_1^{1,\ldots,T} \mid \mathbf{W}) \cong TP \log\left|\det \mathbf{W}\right| + \sum_{t-1}^{TP} \sum_{j=1}^{M} \log p(s_{1j}^t) \quad (6)$$

Our interest is in adapting basis functions of source for $\forall t \in [1, T]$, for convenience $TP = T - P + 1$. The gradient-based learning rules for updating ICA model can be derived by evaluating the appropriate derivatives of log likelihood function. The basis filter gradient $\mathbf{W}$ can be evaluated as follows:

$$\frac{\partial \log p(x_1^{1,\ldots,T} \mid \mathbf{W})}{\partial \mathbf{W}} \propto \left[\mathbf{I} - \varphi(\mathbf{s})\mathbf{s}'\right]\mathbf{W} \quad (7)$$

where the 'nature gradient' in expression 7 is obtained by post-multiplying the gradient by $\mathbf{W}'\mathbf{W}$. $\varphi(\mathbf{s}) = \dfrac{\partial \log p(\mathbf{s})}{\partial \mathbf{s}}$ and $\mathbf{s}'$ denotes the matrix transpose of $\mathbf{s}$. $\mathbf{I}$ is the identity matrix. The $\varphi(\mathbf{s})$ can be obtained by evaluating the derivatives of Generalised Exponential Source (GGD) density function which expressed as follows:

$$G(s; q, \beta, u) = \frac{q\beta^{\frac{1}{q}}}{2\Gamma(\frac{1}{q})} \exp(-\beta\left|s-u\right|^q) \quad (8)$$

The sources densities were assumed to be inverse-cosh densities in traditional ICA formulation [2]. However more current ICA algorithm shows that the inverse-cosh densities can't adequately model sub-Gaussian densities. The more general densities which can model super-Gaussian, Gaussian and sub-Gaussian form is the 'Generalised Exponential (GE)' density [15].where $\Gamma(.)$ is the gamma function. u expresses the mean of coefficient and variance can be determined by β. $\theta = \{u, \beta, q\}$. Here the exponent q denotes the varying of distribution. In the simulation section, this aspect will be discussed in details regarding the affects of the exponent q on the performance of separation.

$$G(s;q,\beta,u)=\left\{\begin{array}{c} sub-Gaussian,\ q>2 \\ Gaussian,\ q=2 \\ \sup er-Gaussian,\ q<2 \\ sparse-coding,\ q<1 \end{array}\right\} \qquad (9)$$

Gradient-based learning rules for updating $\varphi(\mathbf{s})$ then can be expressed as:

$$\varphi(s)=\frac{\partial \log p(s|q,u,\beta)}{\partial s}=sign(s-u)q\beta|s-u|^{q-1} \qquad (10)$$

$\mathbf{W}_i=\mathbf{A}_i^{-1}$ .This expression is exactly from Generalized Gaussian density function. Thus the specific process approach to SCBSS can be concluded as: The observed mixture data y_t and the current estimation individual source $\hat{x}_i^t$ are given. At each time point, firstly, $\mathbf{s}_i^{tn}=\mathbf{W}_i\hat{\mathbf{x}}_i^{tn}$. Current estimated independent source pass through basis filter to generate basis coefficients. Secondly, $s_i^t=\log p(s_i^t)$ each basis coefficients are statistically independent by exploiting Maximum likelihood estimator. Thirdly, Maximum a posterior estimator to obtain likelihood of $\mathbf{x}_i^t$ which basis coefficient' $\mathbf{s}_i^t$ pdf are derived from Generalized Gaussian density function.

$$\Pr(x_i^t\mid \mathbf{W}_i)\cong G(s_i^t\mid \boldsymbol{\theta}_i)|\det \mathbf{W}| \qquad (11)$$

If T samples are independent, then we can decompose the above expression into:

$$\Pr(x_i^{1\ldots T};\mathbf{W}_i)\cong\prod_{t=1}^{T}\Pr(\mathbf{x}_i^t;\mathbf{W}_i) \qquad (12)$$

Therefore the log likelihood of both independent sources as

$$\begin{aligned} L&=\log\left\{\Pr(x_1^{1\ldots T})\Pr(x_2^{1\ldots T})\right\} \\ &\propto\sum_{k=1}^{M}\sum_{t=1}^{T}\log p(s_{1k}^t)+\sum_{k=1}^{M}\sum_{t=1}^{T}\log p(s_{2k}^t) \end{aligned} \qquad (13)$$

The gradients-based learning rule for updating the estimated source can be derived by evaluating the appropriate derivatives of the log likelihood:

$$\frac{\partial L}{\partial z_i^t}\propto\sum_{n=1}^{M}\left[\lambda_2\sum_{k=1}^{M}\left\{\varphi(s_{1k}^{tn})w_{1kn}\right\}-\lambda_1\sum_{k=1}^{M}\left\{\varphi(s_{2k}^{tn})w_{2kn}\right\}\right] \qquad (14)$$

Variables z_1^t and z_2^t are selected to replace $\lambda_1\hat{x}_1^t$ and $\lambda_2\hat{x}_2^t$. Finally, the newest individual gradients are added back to the current estimate source till the separation process converges. For evaluating mixing gain λ_i, it can be estimated by MAP estimator when given the current estimated individual sources.

$$\begin{aligned} &\Pr(\lambda\mid x_1^{1\ldots T},x_2^{1\ldots T})\propto\Pr(x_1^{1\ldots T})\Pr(x_2^{1\ldots T})p_\lambda(\lambda) \\ &\log\Pr(x_1^{1\ldots T})\Pr(x_2^{1\ldots T})p_\lambda(\lambda)\cong L+\log p_\lambda(\lambda) \end{aligned} \qquad (15)$$

The gradient-based rule then can be obtained as:

$$\begin{aligned} &\frac{\partial\log p(s_{ik}^t)}{\partial\lambda_1}=\frac{\partial\log p(s_{ik}^t)}{\partial s_{ik}^t}\frac{\partial s_{ik}^t}{\partial\lambda_1} \\ &\frac{\partial s_{ik}^t}{\partial\lambda_i}=\frac{\partial}{\partial\lambda_i}\left(\frac{\lambda_i s_{ik}^t}{\lambda_i}\right)=-\frac{\lambda_i s_{ik}^t}{\lambda_i^2};\lambda_1=1-\lambda_2 \\ &\Delta\lambda=\varphi(s_{ik}^t)w_{ik}z_i^t\left(-\frac{1}{\lambda_i^2}\right) \end{aligned} \qquad (16)$$

3 Performance Analysis -General basis

Figure 2 shows the flow of the simulations investigated in this paper and the analysis taken place for estimating the specific variables.

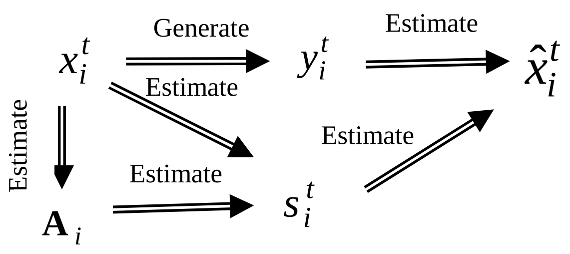
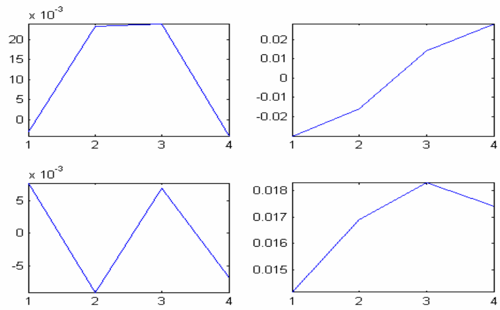


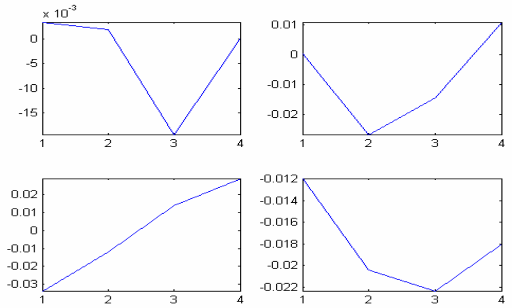
Fig. 2: Simulation structure (independent source $\underline{x}_i(t)$ are generated by Jazz music and speech source.)

In this section, all simulations and analysis are conducted using a PC with Inter Core 2 CPU 6600 @ 2.4GHz 2.4GHz and 2GB RAM. The proposed experiments are devised to test the performance of proposed method on the single channel where the mixtures are exploited by different models (female and male speech, Jazz music and speech sources).

3.1 Separation results by exploiting 4 basis



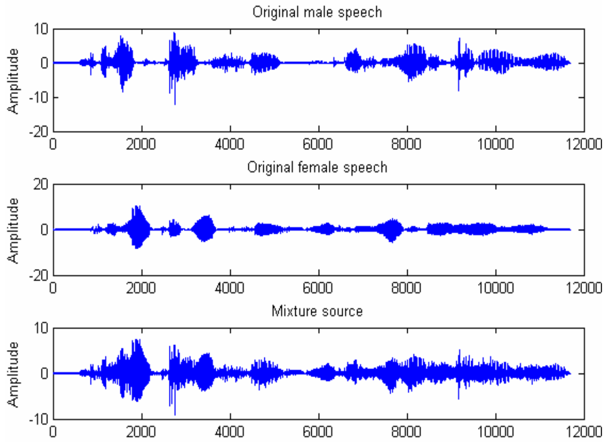
(a)



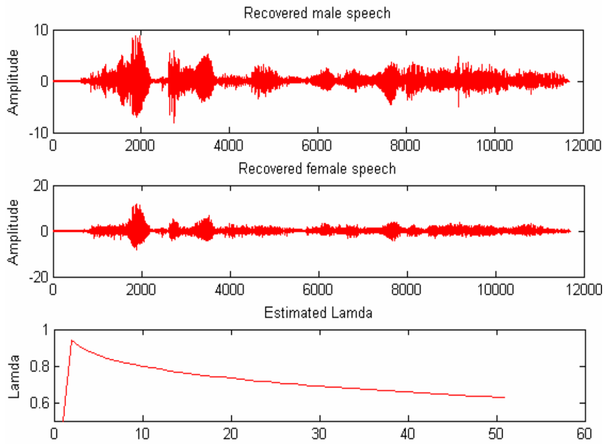
(b)

Fig. 3: Basis Functions. (a) Male basis functions. (b) Female basis functions.

Figure 3 (a) and (b) show the different non-Gaussian sources have specific characteristic of the basis functions.



(a)



(b)

Fig. 4: (a) The original sources and mixture
(b) The estimated sources and λ_i

By contrasting the original and recovered speech sources (the down sampled data were exploited 11697 out of 46785 from male-female speech)., it is computed that the mean square error of estimated two speech signals equal to 0.78064 and 0.62653, respectively.

3.2 Affective factor *q*

Here the *q* factor is derived from expression 8. The question we would like to address in this section is how will the *q* affect separation results? The theorem proves when exponent value decrease, the distribution of basis coefficients become more sparsity [12]. This is the reason why the value of *q* is set to a small number (in Figure 4 *q*=0.2) for separation process. Here is a comparison of separation results which *q*=1 as the same original sources.

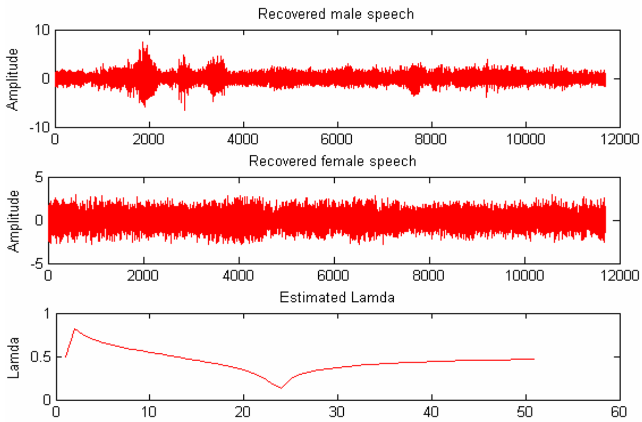


Fig. 5: The estimated sources and λ_i when *q*=1

Table 1: Comparison results of different *q*

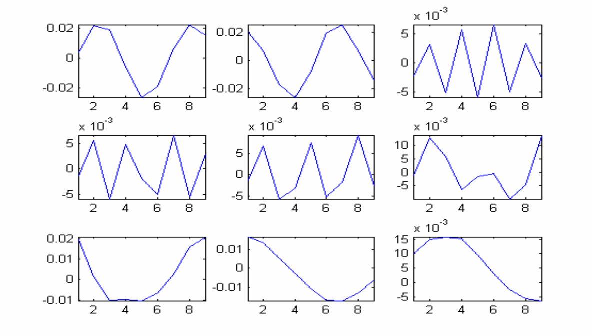
Recovered male speech		
q	MSE	λ
0.2	**0.78064**	0.62653
1	1.4087	0.46589
Recovered female speech		
q	MSE	λ
0.2	**0.6979**	0.37347
1	1.5085	0.53411

MSE: Mean Square Error

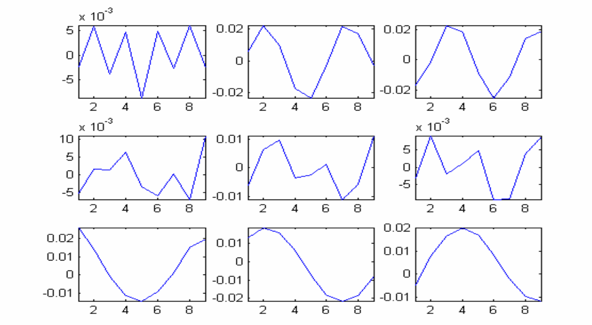
In Table 1, as exponent *q* is getting smaller, the estimation values of individual sources are more close to the original models according to MSE. The problem is the estimation value of λ. It's hard to maintain a correct value when decrease *q*.

3.3 Affective factor: basis functions

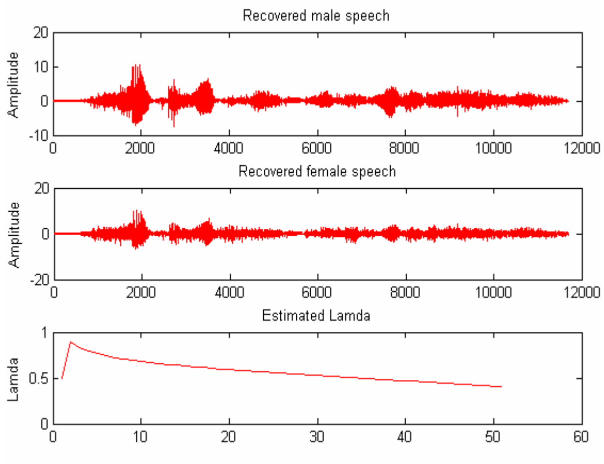
There exists another affective factor in separation process. As previous experiments utilized that the numbers of basis functions are exploited as 4. This experiment will test the results when increase the number of basis functions. The origianl sources are invariant.



(a)



(b)



(c)

Fig. 6: (a) Male basis functions (N=9).
(b) Female basis functions (N=9).
(c) The estimated sources and λ.

Table 2: Comparison results of different basis

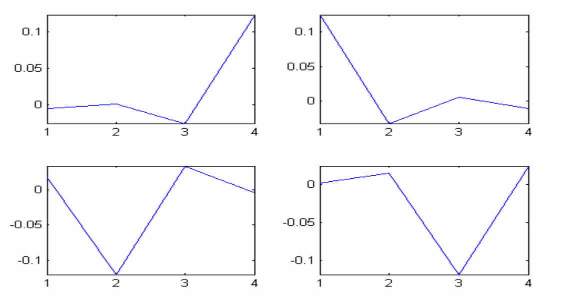
Recovered male speech		
Number of Basis function	MSE	λ
4	**0.78064**	0.62653
9	1.0548	0.40019
Recovered female speech		
Number of Basis function	MSE	λ
4	**0.6979**	0.37347
9	1.088	0.59981

MSE: Mean Square Error

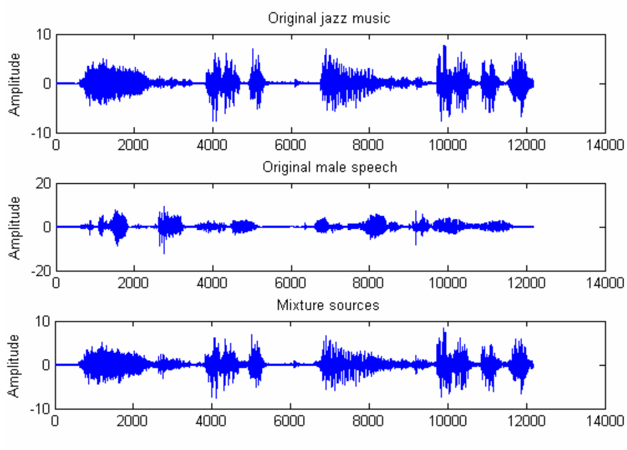
The separation results which number of basis equal to 4 is estimated more accurately than N=9. The problem is also the same as exponent q factor: the estimation value of lamda occur more errors when decrease number of basis functions.

In Figure 4, it is seen that the estimation value of male and female speech may not perform the best separation results with mean square error 0.78064 and 0.6979 respectively. The following experiments will focus on selecting different audio mixture sources to test the performance on the types of mixture which can obtain the best estimation results.

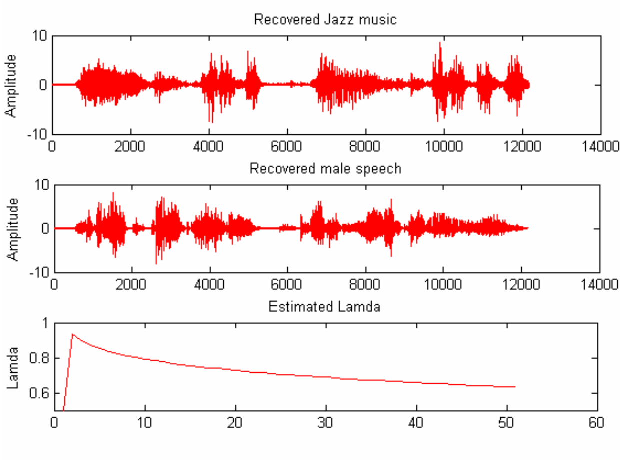
3.4 Different mixture: Jazz and male speech



(a)



(b)



(c)

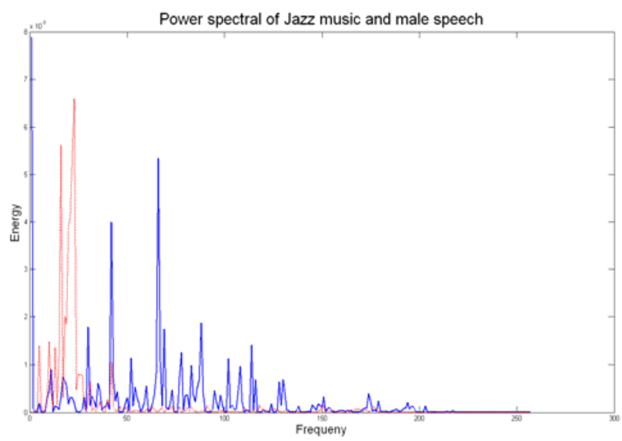
Fig. 7: (a) Jazz music basis functions.
(b) The original sources and mixture
(c) The estimated sources and λ

Table 3: Comparison results of different mixture

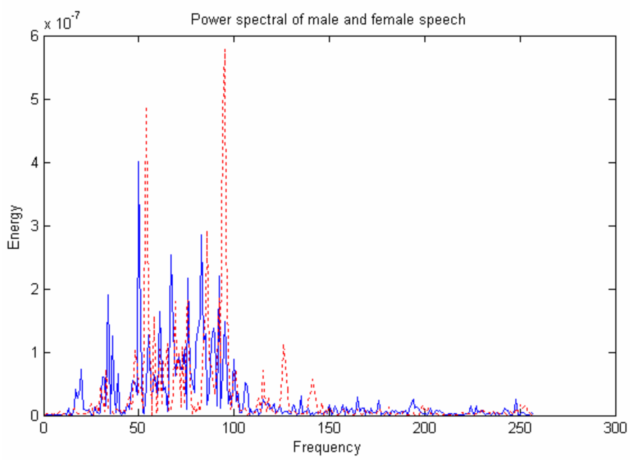
Male and female speech		
Mixture	MSE	λ
Male speech	0.78064	0.62653
Female speech	0.6979	0.37347
Jazz music and male speech		
Mixture	MSE	λ
Male speech	**0.56422**	0.63274
Jazz music	**0.08961**	0.36726

MSE: Mean Square Error

In terms of MSE the mixture which contains music is separated more clearly than male-female mixture. Separation of Jazz music and male speech is the best. The reason why separation results of male-female mixture is worse than Jazz-male mixture can be explained by exploiting their power spectral to analysis. The following two figures show both mixtures' power spectral.



(a)



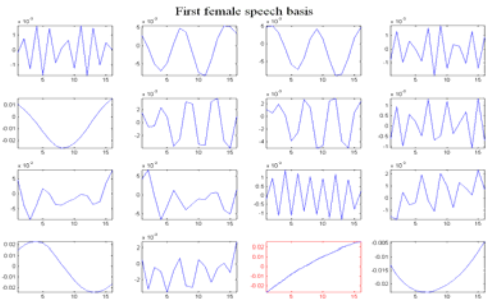
(b)

Fig. 8: (a) Power spectral of Jazz (dot line) and male speech (solid line)
(b) Power spectral of female (solid line) and male speech (dot line)

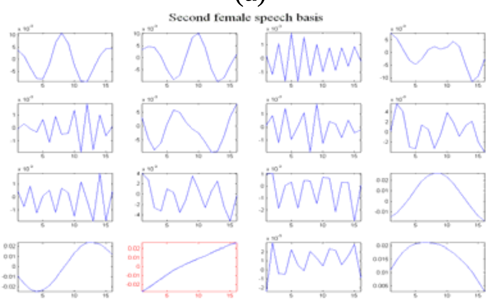
In Figure 8, power spectral density is used to describe the frequency contents of the speech or music signals. The figure shows the frequency components that are present in process and how much power they posses. It is seen that the structures of male-female speech power spectral are very similar. As for the Jazz-male mixture, although there are plenty of overlaps between Jazz music and male speech, the inherent structure of both signals are different due to the basis functions. Thus, the separation results of Jazz-male mixture give the better results.

4 Best Characteristic Speech Basis

The best characteristic basis functions [16] requires that each basis be incomplete in the general space of all possible signals but complete in the subspace of a single class of signal, and that the classes are disjoint in signal space. Based on separation results, there is too much overlap in signal space between two speakers when decrease number of basis functions (e.g. the separation results in table 2 by exploiting 4 basis functions). One way around this obstacle would be to do the separation in some feature space where there is both better class separation, and the possibility of transformation back to signal space. This proposed method based on cross-correlation function to extract most similar features from various male basis functions and female basis functions. Figures 9 and 10 show some feature extraction from general basis functions.

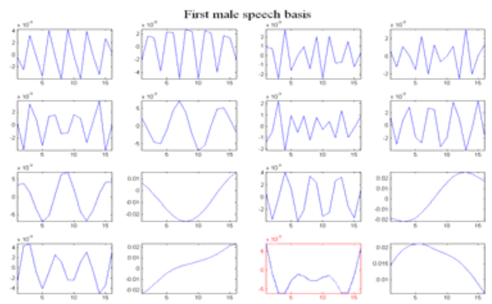


(a)

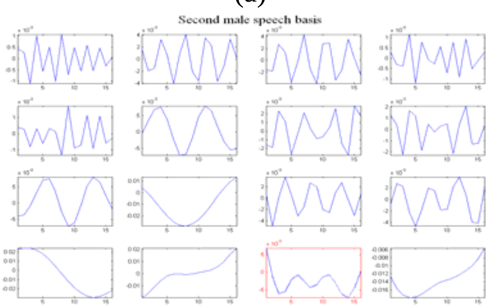


(b)

Fig. 9: (a) First female basis functions.
(b)Second female basis functions.



(a)



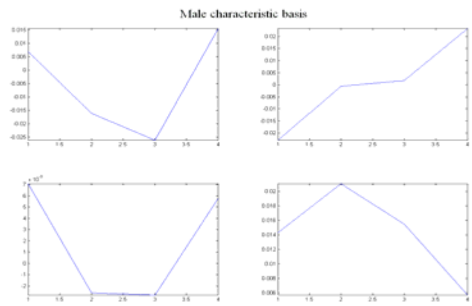
(b)

Fig. 10: (a) First male basis functions.
(b)Second male basis functions.

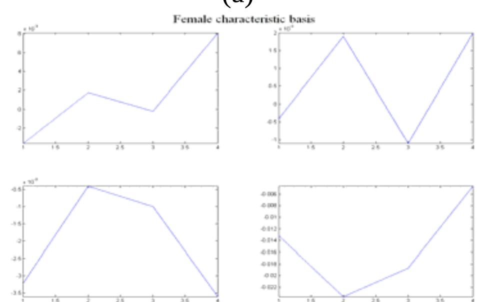
In determining, the best characteristically basis functions for the source signals will first be determined. In this paper, cross-correlation is used to identify the characteristically most similar features inherent in the speech signals. This is carried out by firstly normalizing a set of basis functions (as shown in Fig. 9 and 10) which are obtained using standard signal separation algorithm. The two general basis functions are cross-correlated and largest N values (represent two basis functions are most similar) from the cross-correlation matrix is used to obtain the most characteristically similar features (e.g. the dot line marked basis from both male-female speech basis function). The cross-correlated model can be expressed as:

$$\mathbf{C}_{\mathbf{BA}} = \mathbf{B}^{\mathbf{T}}\mathbf{A} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_N \end{bmatrix} \times \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \ldots & \mathbf{A}_N \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{B}_1\mathbf{A}_1 & \mathbf{B}_1\mathbf{A}_2 & \cdots & \mathbf{B}_1\mathbf{A}_N \\ \mathbf{B}_2\mathbf{A}_1 & \mathbf{B}_2\mathbf{A}_2 & \cdots & \mathbf{B}_2\mathbf{A}_N \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{B}_N\mathbf{A}_1 & \mathbf{B}_N\mathbf{A}_2 & \cdots & \mathbf{B}_N\mathbf{A}_N \end{bmatrix} \tag{9}$$

where B and A denotes the second and first speech basis functions respectively (e.g. B represents the second male basis and A represents the first male basis). The best characteristic features (largest values of) then can be extracted from the cross-correlation matrix and reconstructed as 4 out of 16 characteristic basis functions (less number of basis get better separation results) from male-female speech. In figure 11 shows the extracted basis functions from 10 training male and female speeches respectively which download from TIMIT database.



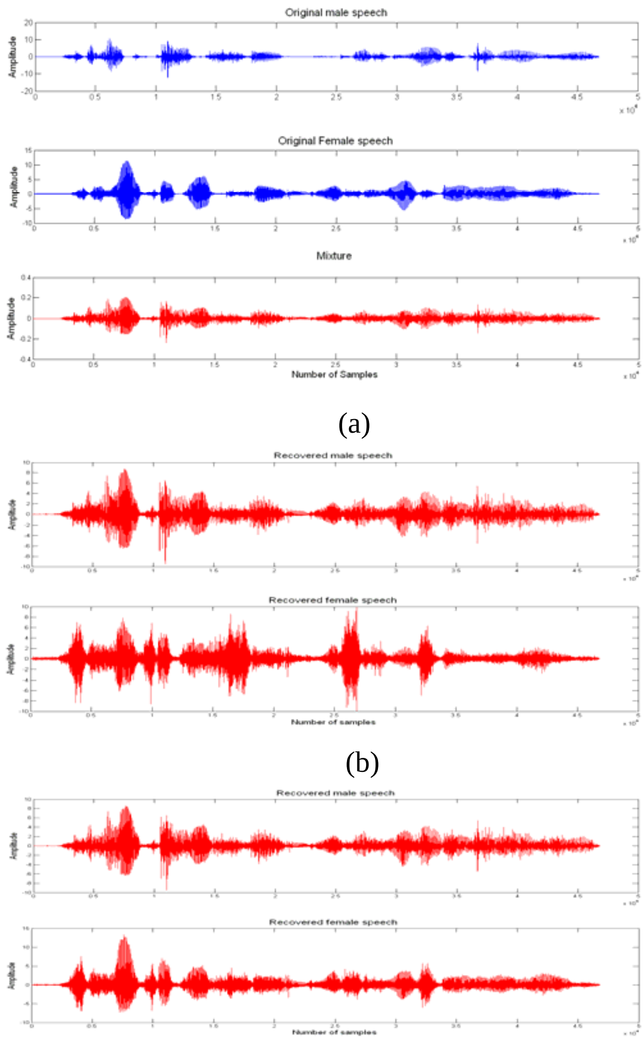
(a)



(b)

Fig. 11: (a) Male characteristic basis.
(b) Female characteristic basis.

We update the new basis filter to obtain the most probably estimate of the source coefficients density. According to Mean Square Error, the separation results especially in terms of recovered female source have been substantially improved by exploiting the most characteristic features. The following figures show the separation results of recovered speech signal by exploiting best characteristic basis.



(a) (b) (c)

Fig. 12: (a) The original sources and mixture
(b) Recovered male-female speech by using the hybrid ML-MAP estimator and general basis functions
(c) : Recovered male-female speech by using the proposed estimator with characteristic features.

In terms of Mean Square Error, the following table concludes how improvements the recovered speech signal are by using characteristic speech basis filters.

Table 4: Performance comparisons

Recovered male-female by general basis	
Mixture	Mean Square Error
Male speech	0.6496
Female speech	1.4241
Recovered speech by characteristically basis	
Mixture	Mean Square Error
Male speech	**0.645**
Jazz music	**0.4669**

In Table 4 and Figure 12 the separation results of both male and female speeches are tabulated which shows an improvement of 0.355% and 50.62% respectively.

5 Comparison to other SCBSS methods

Resent proposed solutions of Single Channel Source Separation (SCSS) [17] problem are categorized into three branches: firstly, model-based SCBSS, secondly, underdetermined blind source separation, and finally, computational auditory scene analysis (CASA). Model-based SCSS [18-20] techniques are similar to model-based single channel speech enhancement techniques. In other words SCSS can be considered as a speech enhancement in which both the target and interference. In underdetermined BSS techniques [21, 22], the sources are projected onto a set of basis functions whose coefficients are as sparse as possible. By using independent component analysis (ICA) nonnegative matrix factorization, or sparse coding, In CASA-based techniques [23, 24], the goal is to replicate the process of human auditory system by exploiting signal processing approaches. The main idea is based on an appropriate transform (such as the short-time Fourier transform (STFT)), the observation signal is segmented into time-frequency cells; then using some criteria to group one source. We now briefly describe and compare two latest novelty techniques approach single channel audio source separation with our proposed method.

First technique SCBSS by using subband filters [25] which exploiting frequency domain to decompose the mixture before separation procedure (the observed mixed signal is converted in to subband with a filter bank so that we can choose a moderate number of subbands and maintain a sufficient number of samples in each subband. In separation process the empirical mode decomposition (EMD) and time domain ICA are exploited to separate each subband mixture. The drawbacks of this method include: firstly the mixture is directly constructed by mixing two original sources but not considering the mixing gain of each source (the gain constants are affected by several factors, such as powers, locations, directions and many other characteristics of the source generators as well as sensitivities of the sensors). Secondly this method exploits spectral techniques. The spectral techniques assume that source signals are disjoint in the spectrogram, which frequently result in audible distortions of the signal in the regions where the assumption mismatches.

The second technique SCBSS by using soft mask filtering [17] make somewhat revise of frequency domain. It exploits the log spectral vector of the mixture which approximated well by the maximum element-wise comparison of the log spectral vector of the sources. In other words at each frequency band the weaker log spectral amplitude is masked by the strong one. The advantage of this method compare with binary mask and wiener filter show that the proposed soft mask filter outperforms both binary and wiener filters.

The advantages of proposed method include: firstly, it avoids strong assumptions by virtue of higher-order statistics. Secondly, there is no longer orthogonality constraint of the subspaces, as the basis functions obtained by the ICA algorithm are not restricted to being orthogonal. Finally the proposed method automatically generates the prior information while other SCBSS methods also require the prior information.

6 Conclusions

A new algorithm based on the hybrid of ML and MAP estimator combined with the encoding of the best characteristic features of the speech has been developed. Real time speech recording has been conducted and the obtained results show significant performance of male-female speech separation using the proposed algorithm. The separation results in both recovered male and female speech improve 0.355% and 50.62% respectively. The separation algorithm of SCBSS was conducted successfully by using to mean square error criterion. Each type of basis function has different characteristics (e.g. speech and music). The relationship between separation results and affective factor q (i.e. the exponent value decrease, the distribution become more super-Gaussian) has been established. The best separation results from different mixtures (separation of Jazz music and male speech is found to be the best. The reason could be explained by exploiting their power spectral density. The relationship between separation results and affective factor number of basis functions (i.e. the number of basis functions decrease, the separation results are getting better) has also been investigated and it is shown that it is not necessary to have very large number of basis functions.

References:

[1] G. J. Brown and M. Cooke, Computational auditory scene analysis, *Computer Speech and Language,* Vol. 8, No. 4, pp. 297–336. 1994

[2] A.Hyvarinen, J.Karhunen, E.Oja, *Independent component analysis and blind source separation*, John Wiley & Sons pp.20-60, 2001.

[3] S.C. Alvarez, A. Cichocki and L.C. Ribas, An Iterative Inversion Approach to Blind Source Separation, *IEEE Trans. on Neural Networks*, Vol. 11, No.6, pp. 1423 – 1437, Nov. 2000.

[4] M. Castella, J.C. Pesquet and A.P. Petropulu, A Family of Frequency-and Time-Domain Contrasts for Blind Separation of Convolutive Mixtures of temporally Dependent Signals, *IEEE Trans. on Signal Processing*, Vol. 53, No.1, Jan. 2005.

[5] H.Attias and C.E. Schreiner, Blind source separation and deconvolution: the dynamic component analysis algorithm, *Neural Computation*, 10 (6): 1373-1424, 1998.

[6] P. Gao, W.L. Woo and S.S. Dlay, Nonlinear Signal Separation for Multi-Nonlinearity Constrained Mixing Model, *IEEE Trans. on Neural Networks*, Vol. 17, No. 3, pp. 796-802, May 2006.

[7] Jingyi Zhang, W.L. Woo, and S.S. Dlay, Blind Source Separation of Post-Nonlinear Convolutive Mixture, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 15, No. 8, pp. 2311-2330, Nov. 2007.

[8] W.L. Woo and S.S. Dlay, Neural Network Approach to Blind Signal Separation of Mono-nonlinearly Mixed Signals, *IEEE Trans. on Circuits and System*, Vol. 52, No. 2, pp. 1236-1247, June 2005.

[9] W.L. Woo and L.C. Khor, Blind Restoration of Nonlinearly Mixed Signals using Multilayer Polynomial Neural Network, *IEE Proceedings on Vision, Image and Signal Processing*, Vol 151, No. 2, pp.51-61, 2002.

[10] W.L. Woo and S. Sali, General Multilayer Perceptron Demixer Scheme for Nonlinear Blind Signal Seperation, *IEE Proc. On Vision, Image and Signal Processing*, Vol. 149, No. 5, pp. 253-262, 2002.

[11] W.L. Woo, and S.S. Dlay, Blind separation of nonlinearly mixed signals using maximum likelihood neural network, *WSEAS Trans. on Systems*, Vol. 2, No. 3, pp. 675-681, 2003.

[12] G. J. Jang and T. W. Lee, A probabilistic approach to single channel source separation, *in Proc. Adv. Neural Inf. Process. Syst.*, pp.1173–1180. 2003.

[13] T.-W. Lee and G.-J. Jang. The statistical structures of male and female speech signals, *in Proc. ICASSP, (Salt Lake City, Utah)*. May 2001.

[14] X.R. Cao and R.W. Liu, General Approach to Blind Source Separation, *IEEE Trans. on Signal Processing*, Vol. 44, No. 3. March, 1996.

[15] R.Everson and S.J.Roberts, Independent component analysis: A flexible nonlinearity and decorrelating manifold approach, *Neural Computation*, Vol.11, No.8.1999

[16] G. J. Jang and T. W. Lee, A maximum likelihood approach to single channel source separation, *Journal of Machine Learning Research*, Vol.4, pp.1365-1392. 2003

[17] M.H. Radfa and R.M. Dansereau, Single-Channel Speech SeparationUsing Soft Mask Filtering, *IEEE Trans. on Audio, Speech and Language Processing*, Vol. 15. 2007.

[18] D. Ellis, Model-based scene analysis, in *Computational Auditory Scene Analysis*: Principles, Algorithms, and Applications, D. Wang and G. Brown, Eds. New York: Wiley/IEEE Press, 2006.

[19] M. J. Reyes-Gomez, D. Ellis, and N. Jojic, Multiband audio modeling for single channel acoustic source separation, in *Proc. ICASSP'04*, Vol. 5, pp. 641–644. May 2004.

[20] T. Kristjansson, H. Attias, and J. Hershey, Single microphone source separation using high resolution signal reconstruction, in *Proc. ICASSP'04*, pp. 817–820. May 2004.

[21] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, and X. Shengli, Underdetermined blind source separation based on sparse representation, *IEEE Trans. Speech Audio Process.*, Vol. 54, No. 2, pp. 423–437, Feb. 2006.

[22] C. Fevotte and S. J. Godsill, A Bayesian approach for blind separation of sparse sources, *IEEE Trans. Audio, Speech, and Lang. Process*., Vol. 14, No. 6, pp. 2174–2188, Nov. 2006.

[23] P. Li, Y. Guan, B. Xu, and W. Liu, Monaural speech separation based on computational auditory scene analysis and objective quality assessment of speech, *IEEE Trans. Audio, Speech, and Lang. Process.*, Vol. 14, No. 6, pp. 2014–2023, Nov. 2006.

[24] G. Hu and D. L. Wang, Monaural speech segregation based on pitch tracking and amplitude modulation, *IEEE Trans. Neural Netw.*, Vol. 15, No. 5, pp. 1135–1150, Sep. 2004.

[25] N.Grbic, X.J.Tao, S.E.Nordholm and I.Claesson, Blind signal separation using overcomplete subband representation, *IEEE Trans. on Audio, Speech and Language Processing*,Vol. 9, No. 5, 524-533. 2005.

[26] K.Kokkinakis and P.C.Loizou, Subband-based blind signal processing for source separation in convolutive mixtures of speech, *IEEE International Conf on Acoustic, Speech and Signal Processing*, Honolulu, HI, pp. 917-920. 2007