Blind Source Separation using Independent Component Analysis

C.K.Kilerci and M.Z.Ahmed

Centre for Security, Communications and Network Research,
Plymouth University, Plymouth, UK
e-mail: info@cscan.org

Abstract

Blind Source Separation (BSS) is a process of estimating a set of signals by using their mixed representations. In this separation process, the user does not have a prior knowledge of the original signals nor their mixing process. This is also called the Cocktail Party Problem. In this paper, a fixed point algorithm of Independent Component Analysis (ICA) was used to separate linearly mixed stationary and non-stationary mixed audio signals. Also, an attempt was made to solve the permutation problem of linearly mixed non-stationary signals by using the sliding window ICA and Minimum Distance Classifier (MDC).

Keywords

Blind Source Separation, Independent Component Analysis, Sliding Window ICA, Cocktail Party Problem, Permutation

1 Introduction

The problem of separating multivariate data without any prior knowledge of both the original sources and the mixing process has attracted a lot of attention during the past decade. This multivariate data can be taken from many different applications such as human brain activity, financial market prediction, geological activity and audio signals. During this project audio signals were selected for testing purposes because of the availability of useful the data.

The goal of this study is to separate linearly mixed stationary and non-stationary audio signals using the ICA technique. The separation was done using a fixed point algorithm using kurtosis, which was proposed by Hyvarinen & Oja (1997). This technique converges very quickly and gives accurate results for linearly mixed audio signals, but for linearly mixed non-stationary signals separation, a modification was required for the algorithm. This modification for the algorithm which is also known as sliding window ICA was implemented. Therefore the way the chosen algorithm works, a permutation effect arises for each separated window depending on the mixing process and frequency characteristics of input signals. In other words, separated windows are not in the same order as the original versions. For this reason, each window was classified using the frequency characteristics of the input signals and the separated windows were placed in the correct order. These improvements make the fixed point algorithm using the kurtosis technique realisable with linearly mixed non-stationary audio signals.

2 Background

2.1 ICA model

Let's assume that, s1 and s2 are the audio signals, where two people speak simultaneously. Then, the speech signals are mixed by an unknown mixing matrix **A**. The aim is to estimate the un-mixing matrix which can be found by ICA under the several assumptions. This un-mixing matrix called **W**. Finally, the output signals can be retrieved from the mixed results, which are the exactly same but opposite representations of the input signals. This process can be also expressed mathematically as shown below;

$$x = A.s \tag{1}$$

$$u = W.x \tag{2}$$

Where **x** is the mixtures and **u** is the estimations of the original signals.

2.2 The fixed point algorithm using kurtosis

Kurtosis is a method of measuring non-Gaussianity. This is an important measurement when determining independent components. Using this method kurtosis can be maximised or minimised when calculating independent components. Kurtosis of a random variable *y* can be calculated as shown below (Hyvarinen *et al.*, 2001);

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{3}$$

From the above formula, the calculated value for *y* can be either positive or negative and this value determines the distribution of the probability density function (pdf). A Gaussian distribution has zero kurtosis while super-Gaussian and sub-Gaussian distributions have positive and negative kurtosis values respectively.

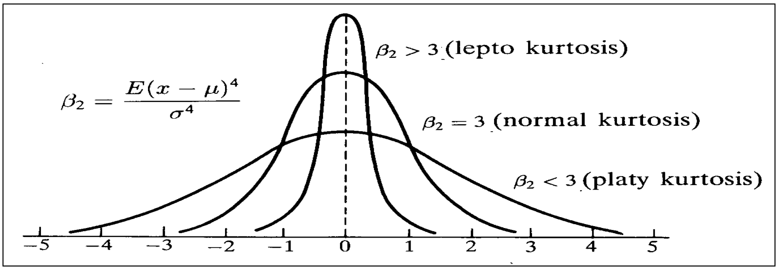

Figure 1: Measures of non-Gaussianity

If a variable has a Gaussian distribution, the kurtosis value cannot be maximised or minimised and for this reason ICA will not work. Another major drawback of the kurtosis can be very sensitive of outliers (Hyvarinen *et al.*, 2001). Hence, observations can be erroneous or irrelevant. To illustrate this drawback, consider the example below;

$$\begin{gathered} Sample\ size = 1000 \\ mean = 0 \\ Variance = 1 \\ Contains\ one\ value = 10 \\ kurtosis = \frac{10^4}{1000} - 3 = 7 \end{gathered}$$

Although, the kurtosis method is not very robust to measure non-Gaussianity, it is an alternative to gradient based algorithms. The reason for this is that a good learning rate must be chosen to obtain a good convergence for gradient based algorithms and failure of wrongly selected learning rate causes slow convergence and failure of separation (Hyvarinen *et al.*, 2001). Additionally, the kurtosis algorithm is one of the versions used in FastICA, (Hyvärinen *et al*., 2005) which is the main reason why this method became popular among the other ICA algorithms.

One-by-one Estimation (steps 3–9)

Fixed-point iteration (steps 4–8)

1. Centering $\mathbf{x} = \tilde{\mathbf{x}} - \mathbf{m}_{\tilde{x}}$
2. Whitening $\mathbf{z} = \mathbf{V}\mathbf{x}, \quad E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$
3. Choose m, No. of ICs to estimate. Set counter $p \leftarrow 1$
4. Choose an initial guess of unit norm for w_p, eg. randomly.
5. Let $\mathbf{w}_p \leftarrow E\{\mathbf{z}[\mathbf{w}_p^T\mathbf{z}]^3\} - 3\mathbf{w}_p\|\mathbf{w}_p\|^2$
6. Do deflation decorrelation
$$\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1}(\mathbf{w}_p^T\mathbf{w}_j)\mathbf{w}_j$$
7. Let $w_p \leftarrow w_p/||w_p||$
8. If w_p has not converged ($|\langle\mathbf{w}_p^{k+1}, \mathbf{w}_p^k\rangle| \neq 1$), go to step 5.
9. Set $p \leftarrow p+1$. If $p \leq m$, go back to step 4.

Figure 2: A fixed point algorithm using kurtosis (Hyvarinen *et al.*, 2001)

From Figure 2, the essential pre-processing steps of centering and whitening, which is also implemented during this study, can be seen. Without the pre-processing steps, some unsuccessful estimations were observed from the experiments. Also the algorithm estimates the signals on a one by one basis and this is also knowm as deflation. When dealing with two input signals, once the algorithm estimates the first signal then whatever is left from the mixture is the estimation for the second signal and this is true for two signals separation (Hyvarinen *et al.*, 2001). For this reason, estimation of the un-mixing matrix is a very crude version of the original mixing matrix.

2.3 Ambiguities of ICA

In the ICA, there are two important ambiguities (Hyvarinen *et al.*, 2001). Firstly, the energies of the independent components cannot be determined because of the two unknowns (**s** and **A**). This means the amplitudes of estimated signals can be different from the original signals. Secondly, the order of the estimated signals cannot be determined. In other words, estimated signals are permutated versions of input signals.

The ambiguities of ICA can have an effect on the estimated output signals and these problems can be solved once the signals are separated successfully by contrasting their amplitudes and changing the orders of the estimated signals.

During this study, the permutation problem of ICA was observed as a result of the separations in the time domain when the statistical properties of the input data were changed such as the change of mixing process. To observe this phenomenon further, non-stationary signals were applied to the ICA algorithm. This is discussed is section 4.2.

2.4 Gender Identification and Minimum Distance Classifier

ICA is a method for performing blind source separation from linear (instantaneous) mixtures. "*The technique assumes a statistical independence between the sources and allows at most one Gaussian component*" (Mitianoudis & Davies, 2002). Separation can be done both the frequency and the time domain. ICA algorithms give successful results in both domains. Although, separation results of both domains are comparable, the computational complexity of the frequency domain approach is much higher (Mitianoudis & Davies, 2002). In addition, Mitianoudis states that "*one major advantage of working in the time domain is that, at least theoretically, the **permutation problem** does not exist*." In other words, the separation results of linearly mixed stationary audio signals will not suffer from the permutation problem. However, the frequency domain approach suffers from the permutation problem (Mitianoudis & Davies, 2001). During this study, all the signals were mixed and unmixed in the time domain to avoid the permutation problem for the linearly mixed stationary signals. On the other hand, separation process of the non-stationary signals in time domain can cause similar problems as found in the frequency domain approach. For this reason, the features of the audio signals play an important role in overcoming this problem.

During this research, female and male audio speech signals were analysed as the input data to the ICA program to increase the convergence speed by increasing independence. Gender difference of the input signals provides different frequency range of the voice. This spectral difference can be used for gender discrimination. (Traunmüller & Eriksson, 1994) states that "*Typical values obtained for fundamental frequencies are 120 Hz for men and 210 Hz for women.*" Additionally, there are many different factors which can affect the final gender decision such as age, language, dialect, accent, health and emotional state (Wu and Childers, 1991).

The gender discrimination technique implemented for this research is known as Minimum Distance Classifier (MDC). MDC is a technique of characterising the information by grouping similarities of data features. MDC requires some knowledge about the data, as with the other classification techniques (Dunham, 2003). For this reason, a training set is used to develop the specific parameters required for the classification purposes and in this case, the required parameters about the data would be the Fast Fourier Transformation (FFT) of the different female and male speeches. In this study, 2 or 3 point FFT of the data can give enough information about the data for classification purposes which is verified experimentally. Also, instead of using the full frequency spectrum, reducing the window size and position in respect to where the most of the spectral power density to locate the spectral band that can uniquely identify the speech characteristics.

2.5 Sliding window ICA for linearly mixed non-stationary audio signals

Linearly mixed stationary signals can be separated very efficiently by ICA algorithms and most of the attempts were made to solve BSS in the literature by assuming signals are linearly mixed and stationary. This assumption simplifies the problem because external factors are constant such as reflections and environmental acoustics. During this research, the same assumptions were made for the initial tests. Afterwards, the mixing process of the signals was changed half way through by generating another random or user defined mixing matrix. Hence, the robustness of the ICA algorithm could be investigated further. For this reason, the algorithm was modified by a technique called sliding window ICA. This technique was proposed for image separation by Hyvarinen *et al.* (2001). Guo & Wu (2010) used sliding window technique with infomax to analyse motor imaginary EEG. In this research, the fixed point algorithm using kurtosis algorithm was adapted to sliding window ICA technique to analyse audio signals

3 Simulation Setup

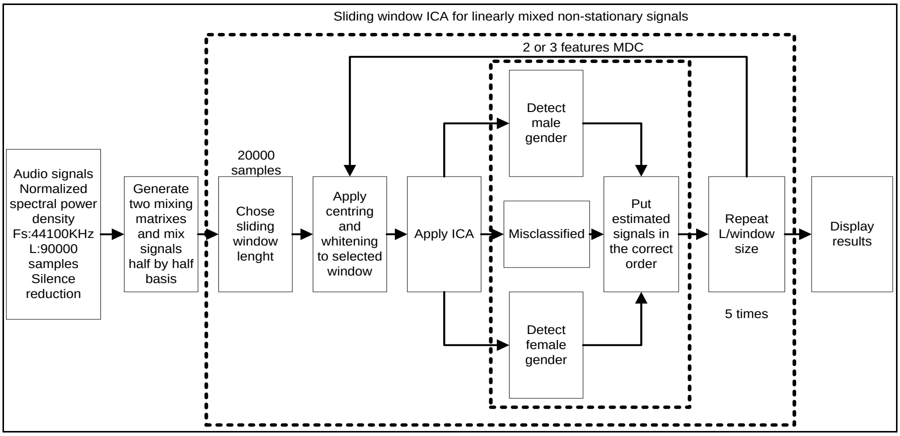

Figure 3: Logic block diagram of the non-stationary signal separation

The simulation was carried out in MATLAB® and the logic block diagram is given in Figure 3. The program takes data from the size of the window which was defined by the user and treats each window as separate signals. In this way, signals would be separated on a window by window basis. Pre-processing was applied to each individual window. The number of repetition was calculated by dividing the length of the signal by the window size. Once the signals were separated successfully, results were passed to MDC and the genders of the signals identified. Using this identification, separated signals were stored into the created male and female variables. This way the effects of the permutation problem was reduced. However, there is, as expected erroneous data appears in the region of the mixing matrix change. There is nothing that can be done to overcome this failure in this method. Therefore, one window of failure is considered as an acceptable failure rate. Although, one window was accepted as a failure, it can be correctly classified and placed inside the correct gender variable depending on how heavily the signals were mixed. Also, size of the window can be reduced to decrease the probability of failure. Conversely, this may cause a discontinuous transition between the windows and can produce audiable high frequency components.

4 Results and Discussion

4.1 MDC simulations

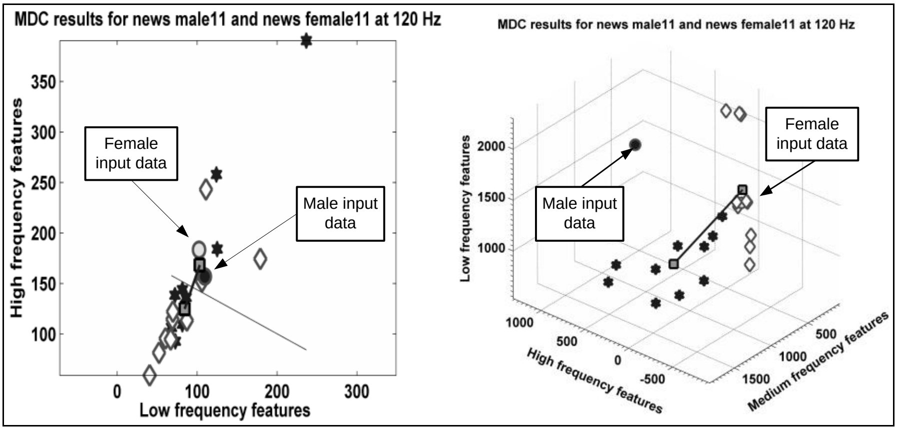

Figure 4: Comparison between two and three features MDC performance

In Figure 4, a two features MDC plot is illustrated on the left hand side. The spectral power density of the data was used at 120 Hz. The training data of the male speech (hexagram) and the female speech (diamond) were not classified perfectly. Centroids (squares) are calculated from the mean values of the training data set and input data (circles) were classified in respect to the distances from the centroids and a decision line (sloped line) was drawn between the two centroids. From the two features MDC plot, it can be clearly seen that the female and male data is misclassified.

On the right hand side plot, a three features MDC is illustrated and this is improvement by carefully selecting of the window's spectral bandwidth to the classification efficiency, where the input data classified correctly.

4.2 Sliding window ICA for linearly mixed non-stationary audio signals simulations

Figure 5, the top two plots represent the original male and female speeches using spectrograms. These audio signals were uploaded to the sliding window ICA program and the statistical properties of the signals were changed. In other words, audio signals appear non-stationary. Hence the permutation problem of ICA was observed. The estimated signals are wrongly ordered as it can be seen from Figure 5. Also, failure is expected at where the mixing process changes, depending on the difference between the mixing matrices. From this point, the three features MDC program used to identify the genders of individual windows so that estimations can be placed to correct gender variable where they are belong.

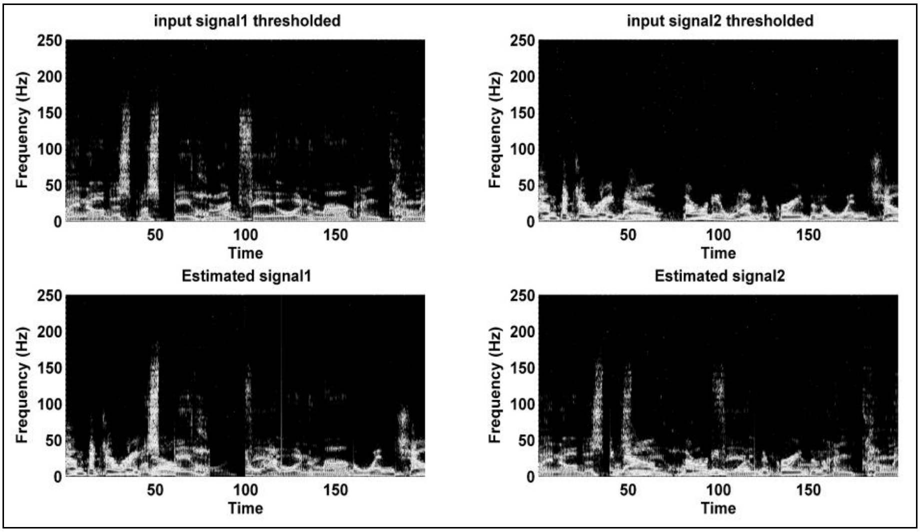

Figure 5: Permutation problem of non-stationary signal separation

Two user defined mixing matrices were generated for the mixing process. Additionally the length of the input data was set to 90000 samples long for the three features MDC. The used window size was 20000 samples long. Hence, input data was separated into five windows. This experiment was repeated for five male and five female audio signals. An audible assessment was done for the evaluation of results and recorded in table 1.

Estimations	M1	F1	M2	F2	M3	F3	M4	F4	M5	F5
Window1	M	F	M	F	M	F	M	F	M	F
Window2	M	F	M	F	M	F	M	F	M	F
Window3	M	F	U	U	M	F	U	U	U	U
Window4	M	F	U	U	M	F	M	F	M	F
Window5	M	F	M	F	M	F	M	F	M	F
M: Male, F: Female, U: Unsuccessful										

Table 1: Three features MDC with sliding window ICA @ 120 Hz

A three feature MDC program gives better classification results for the output signals of the sliding window ICA. Results were verified by listening correctly to the ordered signals. For the classification purposes, the male fundamental frequency spectrum (120 Hz) was used and divided into three equal windows and average of each window taken. Also, as expected some failures were observed at window3. The main reason for this failure is because the ICA cannot estimate two mixing matrices at once. In this case, the random decision was used in window3. To reduce the number of failures, a better classifier can be used such as neural networks.

5 Conclusion

An attempt was made to solve the ICA permutation problem. For linearly mixed non-stationary audio signals, ICA was implemented by dividing data into smaller windows and treating each window as an individual signal. In this research, MDC, the simplest data mining techniques MDC was used to group the male and female speeches by using the frequency characteristics of the data. This technique was implemented by three features from the data. The most spectral power efficient window was used during the classification process. Simulation results have shown that a three feature MDC can successfully classify individual windows. The numbers of failures are reduced and data classified more efficiently. In this way, permutated windows were rearranged respect to gender they are belong to.

6 References

Dunham, M. H. (2003) Data Mining Introductory and Advance Topics. Prentice Hall.

Guo, X. & Wu, X. (2010) 'Motor Imagery EEG Classification Based on Dynamic ICA Mixing Matrix', Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on. 18-20 June 2010. pp. 1-4.

Hyvärinen, A., Särelä, J., Hurri, J. & Gävert, H. (2005) FastICA for Matlab 7.x and 6.x. (Version 2.5) [Computer Program]. Available at: http://www.cis.hut.fi/projects/ica/fastica/

Hyvarinen, A., Karhunen, J. & Oja, E. (2001) Independent Component Analysis. New york: Wiley-Interscience.

Hyvarinen, A. & Oja, E. (1997) 'A fast fixed-point algorithm for independent component analysis'. Neural Comput., 9 (7). pp 1483-1492.

Mitianoudis, N. & Davies, M. (2001) 'A fixed point solution for convolved audio source separation'. IEEE workshop on Applications of Signal Processing on Audio and Acoustics. New York: New Paltz.

Mitianoudis, N. & Davies, M. (2002) 'Audio Source Separation: Solutions and Problems'. International Journal of Adaptive Control and Signal Processing, 18 (3). pp 299-314.

Traunmüller, H. & Eriksson, A. (1994) The frequency range of the voice fundamental in the speech of male and female adults. Stockholm, Sweeden: Department of Linguistics, University of Stockholm. Available.

Wu, K. & Childers, D. G. (1991) 'Gender recognition from speech. Part I: Coarse analysis'. The Journal of the Acoustical Society of America, 90 (4). pp 1828-1840.