

AN AUDIO SEPARATION SYSTEM BASED ON THE NEURAL ICA METHOD

MICHAL BRÁT, MIROSLAV ŠNOREK

Czech Technical University in Prague

Faculty of Electrical Engineering

Department of Computer Science and Engineering

Karlovo náměstí 13, 121 35 Praha 2

Email: bratm@fel.cvut.cz, snorek@cslab.felk.cvut.cz

KEYWORDS

Data Mining, Signal Mining, Blind Signal Separation - BSS, Independent Component Analysis - ICA, Fast Fourier Transformation - FFT, Principal Component Analysis - PCA, Self-Organizing Map - SOM, Learning Vector Quantization – LVQ.

ABSTRACT

This contribution deals with the problems based on data mining, especially signal mining. The main representative of signal mining is Blind Signal Separation. This group of problems can be solved by traditional (mathematical) methods or also untraditional techniques that utilize artificial intelligence such as neural networks. They are not possible to use alone, therefore this contribution focuses on pre-processing of input signals too. In conclusion we show our developed system based on self-organizing neural network and several experiments with it.

1. INTRODUCTION

At this time the amount of data in electronic format in many academic and other disciplines is increased. Otherwise one from many problems of huge data is in their incomprehensiveness and blind information about them. The group of data problems comes under a discipline, which is called data mining. One part of data mining, that is concentrated only on the signal problems, is well known as data stream mining or also signal mining. We

may solve these problems by traditional methods such as mathematical algorithms, especially statistical algorithms or also other untraditional techniques based on artificial intelligence like neural networks.

2. PROBLEM DEFINITION: BLIND SIGNAL SEPARATION

Imagine a group of people who are sitting in a room and speaking simultaneously (see Figure 1). We are member of speaking group and we want to obtain speech from only a person who is speaking important information for us. We must quite concentrate on this person. Human ability of speech recognition can exactly focus on speech from one person and other noise is eliminated. We want to implement the same recognition abilities in a computer science.

This problem based on separation of a signal is well known as “cocktail party problem”. It is one problem of Blind Signal Separation (BSS). The separation is called **blind** because we do hardly know quite anything about an environment in which mixing of signals takes place. It is special section of signal mining, which focuses on signal separation with minimal information about input signals. They are just hard problems from data mining.

The BSS problem covers in as well other signal process. This is economic data stream mining, which wants to obtain knowledge about data stream. Other process is based on separation of damaged medical signals such as EEG or MEG. All these problems are almost solved by traditional techniques. The main representative of these techniques is Independent Component Analysis (ICA) [1]. It

could be used other techniques based on adaptive filters, decision rules and others.

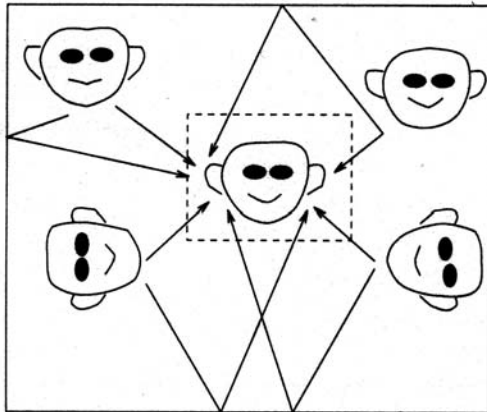


Figure 1: A typical situation in a cocktail party problem

3. STANDARD TECHNIQUES BASED ON MATHEMATICAL ALGORITHMS

The traditional methods for solving of problems, which come out the BSS problem, are almost based on complex mathematical algorithms. The main representative of these techniques is the ICA method. The basic idea of it comes out non-linear transformation of signals in co-ordinates system. The new one represents turning of co-ordinates to direction for better view of signals. Firstly, co-ordinates are turned to direction of maximal variance (this is second statistical moment, in fact it is only linear transformation). Then it is used non-linear transformation of signals. Co-ordinates are turned in direction of maximal kurtosis (it is third statistical moment). More details about it are in [3].

The ICA method is very useful but its computation by mathematical algorithms is quite complex. It can be implemented by easier techniques - using artificial intelligent especially neural networks. Neural networks can be usable for many applications and solutions of hard and non-algorithm problems. The basic idea of the neural ICA method comes out mathematical solution, but implementation is completely different.

4. IMPLEMENTATION OF THE ICA METHOD BASED ON NEURAL NETWORKS

First idea about neural solution of the ICA method has been inspired by article *Non-*

linear Blind Source Separation by Self-Organizing Maps [4]. This meaning was not quite perfect because author has entirely used SOM without using other methods for modification of input signals. Therefore we have prepared first version of a system, which is improvement of the idea came out promising article.

This system (ExNeurICA_PS) is based on neural networks with pre-processing of input signals. A structure of this system is shown in Figure 2. The basic idea of the system consists of pre-processing of input signals and a core of system using neural networks. Pre-processing of input signals is done by the PCA method. It is in fact the same pre-processing such as for mathematical solution of the ICA method. Co-ordinates are turned to direction of maximal variance.

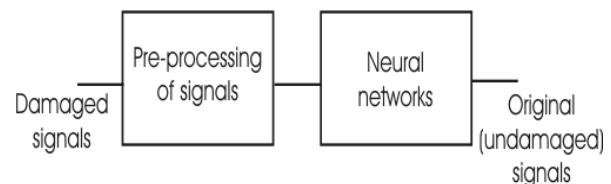


Figure 2: The structure of our systems

The second part of this system utilizes neural network, especially Kohonen's self-organizing map SOM. This neural network seems to be also used for non-linear transformation because of its architecture.

We have prepared several experiments with this system [3]. These results seemed to be not perfect therefore we have prepared new system. More details about previous system are in [2].

5. METHOD IMPROVEMENT: FREQUENCY DOMAIN APPROACH

The structure of new system is the same as previous system, but meaning is completely different. Audio signal in time domain is not quite applicable because it is dependent of quality and level of signal. Therefore almost all audio signals are processed in frequency domain because of easier elaboration. Generally, the signals in frequency domain keep better features.

The same idea about signals in frequency domain is usable for implementation of the ICA method. A developed system is just based on frequency pre-processing and clustering

according to self-organizing neural networks. Transformation from time to frequency domain has been performed by fast Fourier transformation (FFT).

Now we can define variables for computing of this system. The input signals are $x(t)$. They are in fact the damaged (or also mixed) signals, which are separated. The separated signals are marked as $s'(t)$. The original signals $s(t)$ mean etalon for test of quality results. In fact we have not these signals in real application. In addition to they are the basic variable and the inside (only in system) variable is Fourier's image $X(k)$.

5.1. Fast Fourier Transformation - FFT

This transformation has been known a long time but in era without computers it was disapproved and not much used. At this time this transformation is quite used, mainly in discipline, which deals with an audio process. The signal is transformed by the equation

$$X(k) = \sum_{i=0}^{N-1} x(i) e^{-\frac{j2\pi i k}{N}}, \quad k = 0, 1, 2, \dots, N-1$$

where $x(i)$ represents the mixed signal (in the time domain) and $X(k)$ is Fourier's image of the mixed signal (in the frequency domain). It is FFT, but we need also inversion of FFT (iFFT). It is defined by the equation

$$s'(i) = \sum_{k=0}^{N-1} S'(k) e^{\frac{j2\pi i k}{N}}, \quad i = 0, 1, 2, \dots, N-1$$

where $S'(k)$ represents Fourier's image of the estimated signal (in the frequency domain) and $s'(i)$ is the estimated signal (in the time domain).

5.2. Neural Networks SOM and LVQ

We use the same neural networks such as was used in first system. First used neural network SOM has been used as classifier [5], because of its non-linear ability of transformation. The basic idea of it is based on "change a position of neurons" (in fact it is only change the weight of neurons). These neurons are attracted to clustering. The basic idea of using SOM in develop system is shown in Figure 3. The spectral lines, which are very close among them, are clustered. Each cluster means an audio signal in frequency domain. After iFFT, these signals are separated to time

domain. SOM is unsupervised neural networks therefore we do not exactly set a number of clusters. This is very important, because a number of clusters must be the same as a number of signals. This condition cannot be followed.

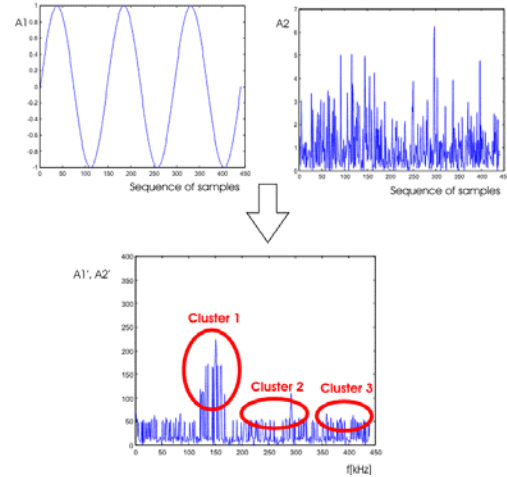


Figure 3: The basic idea based on SOM (This is not possible to set a number of clusters therefore there are different a number of clusters than signals.)

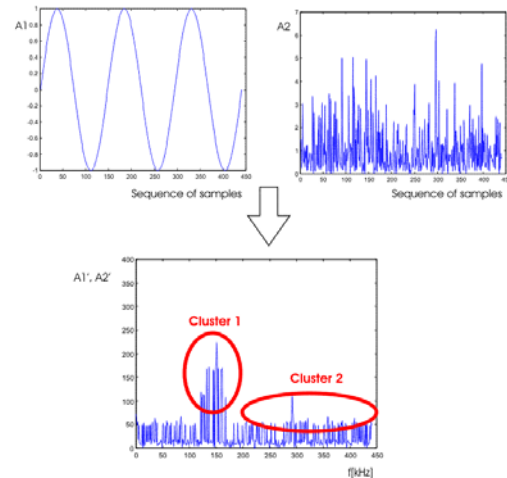


Figure 4: The basic idea based on LVQ (There is exactly to set a number of clusters. The number has to be the same as a number of signals.)

Accordingly we have used LVQ because of its similarity with SOM. This neural network is simply put "SOM with supervised learning" [5]. The idea of this system is the same as with SOM, but we can set exact number of clusters. The basic idea using LVQ is shown in Figure 4.

After both clustering (by SOM or LVQ) we transform signals in time domain from

frequency clusters. For example in Figure 4, the Cluster 1 is first separated signal and Cluster 2 is second separated signal. We describe only situation with two signals, but this idea is used for more signals. For easier explanation we show only this approach.

This system was programmed in Java programming language. It follows that it is independent of operation system. This system will be located on web page <http://cs.felk.cvut.cz/~bratm>.

6. EXPERIMENTS

We would like to describe several experiments with a developed system. Some experiments utilize simple audio signals (e.g. mixture of audio signals with exactly fixed frequency) and also songs or human speech. The experiments have been performed on a PC with the Intel 600 MHz processor, with 256 MB operation memory. The operation system has been Windows 2000. We have prepared more experiments, but now we show only experiments with audio signals.

The input (mixed) signals are shown in Figure 5 a) and b). There are two mixed signals, in fact damaged signals, which have to be repaired. This is simulation of cocktail party.

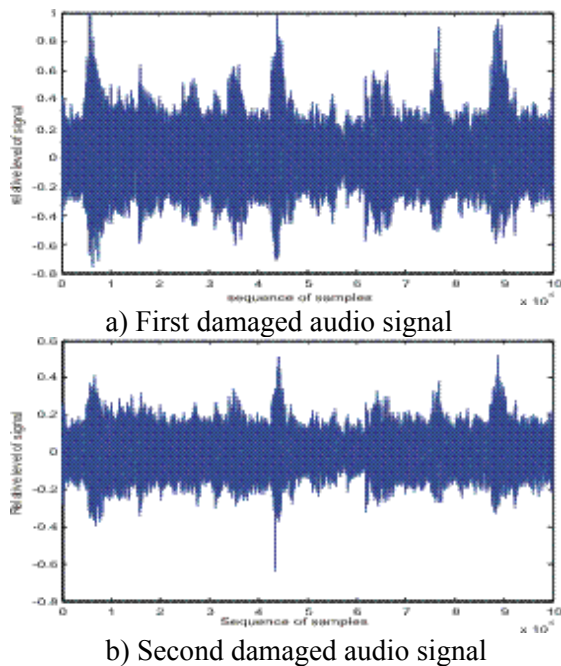


Figure 5: The input mixed signals (speech /one, two, .../ & song simultaneously)

The mixed signals have been pre-processing by FFT. After that we have only used SOM, because the results are quite good. The results can be shown in Figure 6 a) and b).

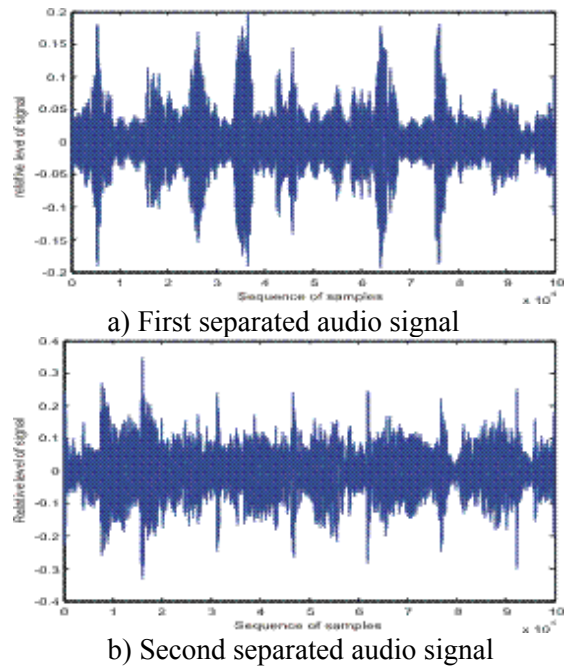


Figure 6: The separated signals

We would like to compare quality of separated signals. Quality can be obtained from joint density. This graph of joint density must be square, but it can be also turned. Figure 7 shows joint density of mixed signals. It is non-orthogonal (non-squared). Figure 8 shows joint density after separation. It is much better then mixed signals. If we look at audio signals or we are listening songs, it is quite good. In conclusion we resume our results.

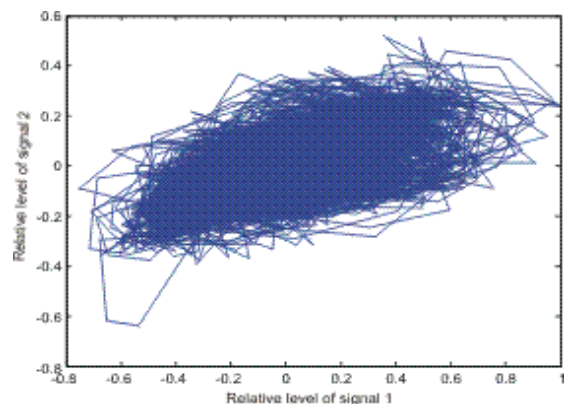


Figure 7: Joint density of mixed signals

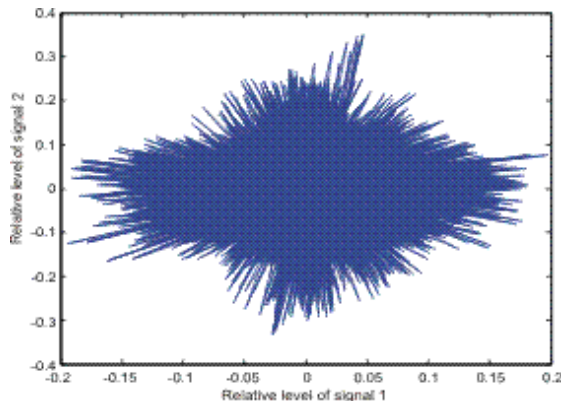


Figure 8: Joint density of separated signals

7. CONCLUSION

This system based on FFT and SOM seems to be very usable for audio separation problems. We can resume that this idea can be used for solving of the BSS problem especially cocktail party problem. We also know that this system is not completely perfect. Firstly we thought that the system based on clustering using SOM is inapplicable. But during developed experiments we ascertain that this system is quite good, maybe better than a system utilizes an idea of clustering by LVQ. We are increasing this system based on LVQ neural network and after that we compare the results. We would like to present our results on next conference.

REFERENCES

- [1] Hyvärinen, A., Karhunen, J., Oja, E. 2001. "Independent Component Analysis". Canada. ISBN 0-471-40540-X
- [2] Brát, M., Šnorek, M. 2002. "Extended Neural ICA for Blind Signal Separation". pages 125-132. MOSIS. ISBN 80-85988-71-2.
- [3] Brát, M. 2003. "Blind Signal Separation – Data Streams Mining Using Neural Network". Postgraduate Study Report DC-PSR-2002-11. CTU.
- [4] Pajunen, P., Hyvärinen, A., Karhunen, J. 2000. "Non-linear Blind Source Separation by Self-Organizing Maps". Helsinki University of Technology. Espoo.
- [5] Šíma, J., Neruda, R. 1996. "Teoretické otázky neuronových sítí". MATFYZPRESS. ISBN 80-85863-18-9.

AUTHORS BIBLIOGRAPHY

MICHAL BRÁT was born in south-bohemia in Počátky, Czech Republic, in 1977. He studied Computer Science and Engineering at Czech Technical University. At this time he is Ph.D. student



at the Department of Computer Science and Engineering of Faculty of the same University (CTU).

He is interested on a processing of signals, audio and video process and artificial intelligent, especially neural networks.

MIROSLAV ŠNOREK was born in south bohemian town Písek, CZ, in 1947. He studied Technical Cybernetics at Czech technical University Prague and he graduated in 1970. He is currently



Associated Professor at the Department of Computer Science and Engineering of Electrical Faculty of the same university (CTU).

He is the head of Neural Network Group. His research interests include unsupervised clustering, GMDH algorithm and neural network applications in modelling and interfacing computers to the real world.