

A time-frequency approach to blind separation of under-determined mixture of sources

A. MANSOUR
Lab. E³I²,
ENSIETA,
29806 Brest cedex 09, (FRANCE).
mansour@ieee.org

M. KAWAMOTO
Dept. of Electronic and
Control Systems Eng.,
Shimane University,
Shimane 690-8504, (JAPAN)
kawa@ecs.shimane-u.ac.jp

C. PUNTONET
Departamento de Arquitectura y
Tecnología de computadores,
Universidad de Granada,
18071 Granada, (SPAIN).
carlos@atc.ugr.es

ABSTRACT

This paper deals with the problem of blind separation of under-determined or over-complete mixtures (i.e. more sources than sensors). At first a global scheme to separate under-determined mixtures is presented. Then a new approach based on time-frequency representations (TFR) is discussed. Finally, some experiments are conducted and some experimental results are given.

KEY WORDS

ICA, BSS, Time-Frequency domain, over-complete or under-determined mixtures

1. Introduction

Blind separation of sources problem is a recent and an important signal processing problem. This problem involves recovering unknown sources by only observing some mixed signals of them [1]. Generally, researchers assume that the sources are statistically independent from each other and at most one of them can be a Gaussian signal [2]. Other assumptions can be also founded in the literature concerning the nature of the transmission channel (i.e. an instantaneous or a memoryless channel, a convolutive or a memory channel, and a non-linear channel). In addition, a widely used assumption considers that the number of sensors should be equal or greater (for subspace approaches) than the number of sources. These assumptions are fairly satisfied in many divers applications such as robotics, telecommunication, biomedical engineering, radars, etc., see [3].

In recent applications linked to special scenarios in telecommunication (as satellite communication in double-talk mode), robotics (for exemple, robots which imitate human behavior) or radar (in ELeCTronic INTelligence "ELINT" applications), the assumption about the number of sensors can not be satisfied. In fact, in the latter applications the number of sensors is less than the number of sources and often we should deal with a mono-sensor system with two or more sources.

Recently, few authors have considered the under-determined mixtures. Thus by using overcomplete representations, Lewicki and Sejnowski in [4] present an algorithms to learn overcomplete basis. Their algorithm uses a Gaussian approximation of probability density function (PDF) to maximize the probability of the data given the model. Their approach can be considered as a generalization of the Independent Component Analysis (ICA) [2] in the case of instantaneous mixtures. However, in this approach, the sources should be sparse enough to get good experimental results, otherwise the sources are being mapped down to a smaller subspace and there is necessary a loss of information. Using the previous approach, Lee *et al.* [5] separate successfully three speech signals using two microphones. On the other hand, When the sources are sparsely distributed, at any time t , at most one of sources could be significantly different from zero. In this case, estimating the mixing matrix [6, 7, 8] consists of finding the directions of maximum data density by simple clustering algorithm. Using Reimannian metrics and Lie group structures on the manifolds of over-complete mixture matrices, Zhang *et al.* [9] present a theoretical approach and develop an algorithm which can be considered as a generalization of the one presented in [10]. The algorithm of Zhang *et al.* update the weight matrix by minimizing a kullback-Leibler divergence by using natural learning algorithm [11].

In the general case, one can consider that separation of over-complete mixtures still a real challenge for the scientific community. However, some algorithms have been proposed to deal with particular applications. Thus for binary signals used in digital communication, Diamantaras and Chassioti [12, 13] propose an algorithm based on the PDF of the observed mixed signals. The pdf of the observation signals have been modeled by Gaussian pdf and estimated from the histogram of the observed signals. Using differential correlation function, Deville and Savoldelli [14] propose an algorithm to separate two sources from noisy convolutive mixtures. The proposed approach requires the sources to be long-term non-stationary signals and the noise should be long-terme stationary ones. The previous statement means that the sources (resp. noise) should have different (resp. identically) second order statis-

tics at different instances separated by a long period.

2. Channel Model

Hereinafter, we consider that the n sources are non Gaussian signals and statistically independent from each other. In addition, we assume that the noise is an additive white Gaussian noise (AWGN). Let $S(t)$ denote the $n \times 1$ source vector at any time t , $Y(t)$ is $m \times 1$ mixing vector and $B(t)$ is a $m \times 1$ AWGN vector. The channel is represented by a full rank real and constant $m \times n$ matrix \mathbf{H} ($m < n$).

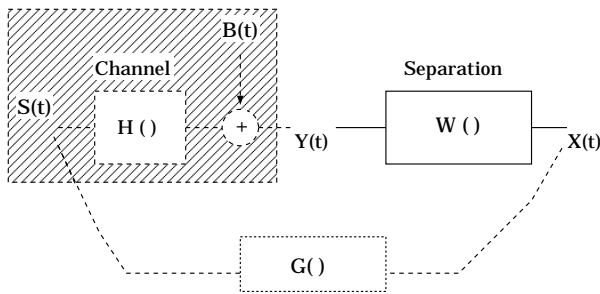


Figure 1. General structure.

The separation is considered achieved when the sources are estimated up to a scale factor and a permutation. That means the global matrix \mathbf{G} can be written as:

$$\mathbf{G} = \mathbf{W}\mathbf{H} = \mathbf{P}\mathbf{\Delta}$$

here, \mathbf{W} is a weight matrix, \mathbf{P} is a permutation matrix and $\mathbf{\Delta}$ is a non-zero diagonal matrix. For a sake of simplicity and without loss of generality, we will consider in the following that:

$$\mathbf{H} = [\mathbf{H}_1 \quad \mathbf{H}_2]$$

Where \mathbf{H}_1 is an invertible matrix and \mathbf{H}_2 is a full rank rectangular matrix.

3. A Separation Scheme

In the case of over-complete mixtures ($m < n$), the invertibility of the mixing matrix \mathbf{H} becomes an ill-conditioned problem. That means the Independent Component Analysis (ICA) will be reduced to extract m independent signals which are not necessarily the origine sources, i.e. the separation can not give a unique solution. Therefore, further assumptions should be considered and in consequence suitable algorithms could be developed. Thus, two strategies can be considered:

- At first one can identify the mixing matrix then using this estimated matrix along with important information about the nature or the distributions of the sources, we should retrieve the original sources.

- In many applications (such as speech signals, telecommunications, *etc*), one can assume the sources have special features (constant modulus, frequency properties, *etc*). Using sources' specifics, the separation becomes possible in the classic manner, i.e. up to permutation and a scale factor.

Beside the algorithms cited and discussed in the introduction of our manuscript, few more algorithms can be founded in the literature. The latter publications are discussed in this section.

3.1 Identification & Separation

One of the first publications on the identification of under-determined mixtures was proposed by Cardoso [15]. In his manuscript, Cardoso proposed an algorithm based only on fourth-order cumulant. In fact, using the symmetries of quadricovariance tensor, an identification method based on the decomposition of the quadricovariance was proposed.

Recently, Comon [16] proved using an algebraic approach, that the identification of static MIMO (Multiple Inputs Multiple Outputs) with fewer outputs than inputs is possible. In other words, he proved that the CANonical Decomposition (CAND) of a fourth-order cross-cumulant tensor can be considered to achieve the identification. In addition, he proved that ICA is a symmetric version of CAND. Using a Sylvester's theorem in multilinear algebra and the fourth order cross cumulant tensor, he proposed an algorithm to identify the mixing matrix in the general case. To recover d-psk sources, comon proposes also a non-linear inversion of \mathbf{H} by adding some non-linear equations and using the fact that the d-psk signals satisfy special polynomial properties (i.e. $s^d(t) = 1$). Later on, Comon and Grellier [17] proposed an extension of the previous algorithm to deal with different communication signals (MSK, QPSK and QAM4). Similar approach was also proposed by De Lathauwer *et al.*, see [18].

Finally, Taleb [19] proposes a blind identification algorithm of M-inputs and 2-outputs channel. He proved that the coefficients of the mixing matrix \mathbf{H} are the roots of a polynomial equations based on the derivative of the second characteristic function of the observed signals. The uniqueness of the solution is proved using Darmois' Theorem [20].

3.2 Direct Separation

Here, we discuss methods to separate special signals. As it is mentioned in the previous subsection that Comon *et al.* [16, 17] proposed an algorithm to separate communication signals. Nakadai *et al.* [21, 22] addressed the problem of a blind separation of three mixed speech signals with the help of two microphones by integrating auditory and visual processing in real world robot audition systems. Their

approach is based on direction pass-filters which are implemented using the interaural phase difference and the interaural intensity difference in each sub-band frequency. Using Dempster-Shafer theory, they determine the direction of each sub-band frequency. Finally, the waveform of one sound can be obtained by a simple inverse FFT applied to the addition of the sub-band frequencies issued by the specific direction of that speaker. Their global system can perform sound source localization, separation and recognition by using audio-visual integration with active movements.

4. Time-Frequency Approach

The algorithm proposed in this section is based on time-frequency distributions of the observed signals. To our knowledge, few time-frequency methods have been devoted to the blind separation of MIMO channel. In fact, for MIMO channel with more sensors than sources, Belouchrani and Moeness [23] proposed a time-frequency separation method exploiting the difference in the time-frequency signatures of the sources which are assumed to be nonstationary multi-variate process. Their idea consists on achieving a joint diagonalization of a combined set of spatial time-frequency distributions which have been defined in their paper.

It is clear from the discussion of the previous sections that the identification of MIMO channel is possible. However, the separation is not evident in the general case. The few published algorithms for the under-determined matter are very linked to signal features of their applications. In our applications, an instantaneous static under-determined mixture of speech signals is considered. This problem can be divided into two steps:

- At first an identification algorithm should be applied. For the moment, we didn't develop a specific identification algorithm. Therefore, any identification algorithms previously mentioned can be used.
- Let us assume that the coefficient of the mixing matrix \mathbf{H} have been estimated. The question becomes How can we recover the sources from fewer sensors? To answer this question, we consider in this section the separation of a few speech signals (for the instance, we are considering just two or at most three sources) using the output of a single microphone (i.e. Multiple Inputs Single output, MISO channel).

Recently, time-frequency representations (TFR) have been developed by many researchers [24] and they are considered as very powerful signal processing tools. In the literature, many different TFR have been developed as Wigner-Ville, Pseudo-Wigner-Ville, Smooth Pseudo-Wigner-Ville, Cho-Willims, Born-Jordan, *etc.* In a previous study [25], we found that for simplicity and performance reasons, the Pseudo-Wigner-Ville can be

considered as a good TFR candidate.

Here we present a new algorithm based on time-frequency representations of the observed signals (TFR) to separate a MISO channel with speech sources. It is known that speech signals are non-stationary signals. However within phonemes (about 80 ms of duration) the statistics of the signal are relatively constant [26]. On the other hand, It is well known that voiced speech are quasi-periodic signals and the non-voiced signals can be considered as white filtered noise [27]. Within a small window corresponding to 51 ms, the pitch can be slightly change. Therefore, one can use this property to pick up the frequency segments of a speaker. The pitch can be estimated using divers techniques [28].

Using the previous facts and Pseudo-Wigner-Ville representations, one can separate up to three speech signals from one observed mixed signal of them. To achieve that goal, we assume that the time-frequency signatures of the sources are disjoints. At first, one should calculate the TFR of the observed signal. Then, in the time-frequency space, we plot a regular gridded. The dimensions of the a small cell of the gridded are evaluated based on the properties of the speech signals and the sampling frequency. Therefore, these dimensions can be considered as 10 to 20 ms in length (i.e. time axis) and 5 to 10% of the sampling frequency value in the vertical axis. Once we plot the gridded, we estimate the energy average in each cell and a threshold is applied to distinguish noisy cells from other. Then the cell with the maximum energy is considered as a potential pitch of one speaker and it is pointed out. After that, we merge in a set of cells, all cells with high level of energy in the neighborhood of the previous cell. At least one harmonic of the pitch should be also selected. The previous steps should repeated as necessary. Finally, the obtained map can be considered as a bi-dimensional time-frequency filters which should be applied on the mixed signal. Using a simple correlation maximization algorithm, one can find the different pieces corresponding to the speech of one speaker.

5. Experimental Results

To demonstrate the validity of the proposed algorithm mentioned in section 4, many computer simulations were conducted. Some results are shown in this section. We considered the following two-input and one-output system.

$$Y(t) = S_1(t) + S_2(t) \quad (1)$$

The sources $S_i(t)$ ($i = 1, 2$) were male and female voices which were recorded by 8 [KHz] sampling frequency. The TFR was calculated by using 128 data of the observed signal $Y(t)$.

Figure 2 shows the results obtained by applying the proposed algorithm (last paragraph in section 4) to the observed signal. From this figure, one might think that the estimated signals are different from the original signals. However, if one hear the estimated signals, one can see that the two original sources $S_1(t)$ and $S_2(t)$ are separated from the observed signal by the proposed algorithm.

6. Conclusion

This paper deals with the problem of blind separation of under-determined (or over-complete) mixtures (i.e. more sources than sensors). At first, a survey on blind separation algorithms for under-determined mixtures is given. A separation scheme based on identification or direct separation is discussed. A new time-frequency algorithm to separate speech signals has been proposed. Finally, some experiments have been conducted and the some experimental results are given. Actually, we are working on a project concern the separation of under-determined mixtures. Further results will be the subject of future communications.

References

- [1] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications.," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E83-A, no. 8, pp. 1498–1512, August 2000.
- [2] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.
- [3] A. Mansour and M. Kawamoto, "Ica papers classified according to their applications & performances.," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86-A, no. 3, pp. 620–633, March 2003.
- [4] M. Lewicki and T. J. Sejnowski, "Learning non-linear overcomplete representations for efficient coding," *Advances in neural Information Processing Systems*, vol. 10, pp. 815–821, 1998.
- [5] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87–90, April 1999.
- [6] P. Bofill and M. Zibulevsky, "Blind separation of more sources than mixtures using sparsity of their short-time fourier transform," in *International Workshop on Independent Component Analysis and blind Signal Separation*, Helsinki, Finland, 19–22 June 2000, pp. 87–92.
- [7] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2363, 2001.
- [8] P. Bofill, "Undetermined blind separation of delayed sound sources in the frequency domain," *NeuroComputing*, p. To appear, 2002.
- [9] L. Q. Zhang, S. I. Amari, and A. Cichocki, "Natural gradient approach to blind separation of over- and under-complete mixtures," in *First International Workshop on Independent Component Analysis and signal Separation (ICA99)*, J. F. Cardoso, Ch. Jutten, and Ph. Ioubaon, Eds., Aussois, France, 11–15 January 1999, pp. 455–460.
- [10] M. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, no. 2, pp. 337–365, 2000.
- [11] S. I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Neural Information Processing System 8*, Eds. D.S. Toureyzky et. al., 1995, pp. 757–763.
- [12] K. Diamantaras and E. Chassiotti, "Blind separation of n binary sources from one observation: A deterministic approach," in *International Workshop on Independent Component Analysis and blind Signal Separation*, Helsinki, Finland, 19–22 June 2000, pp. 93–98.
- [13] K. Diamantaras, "Blind separation of multiple binary sources using a single linear mixture," in *Proceedings of International Conference on Acoustics Speech and Signal Processing 2001, ICASSP 2000*, Istanbul, Turkey, Jun 2000, pp. 2889–2892.
- [14] Y. Deville and S. Savoldelli, "A second order differential approach for underdetermined convolutive source separation," in *Proceedings of International Conference on Acoustics Speech and Signal Processing 2001, ICASSP 2001*, Salt Lake City, Utah, USA, May 7–11 2001.
- [15] J. F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors.," in *Proceedings of International Conference on Speech and Signal Processing 1991, ICASSP'91*, Toronto-Canada, May 1991, pp. 3109–3112.
- [16] P. Comon, "Blind channel identification and extraction of more sources than sensors," in *In SPIE Conference on Advanced Algorithms and Architectures for Signal Processing*, San Diego (CA), USA, July 19–24 1998, pp. 2–13, Keynote address.
- [17] P. Comon and O. Grellier, "Non-linear inversion of underdetermined mixtures," in *First International Workshop on Independent Component Analysis and*

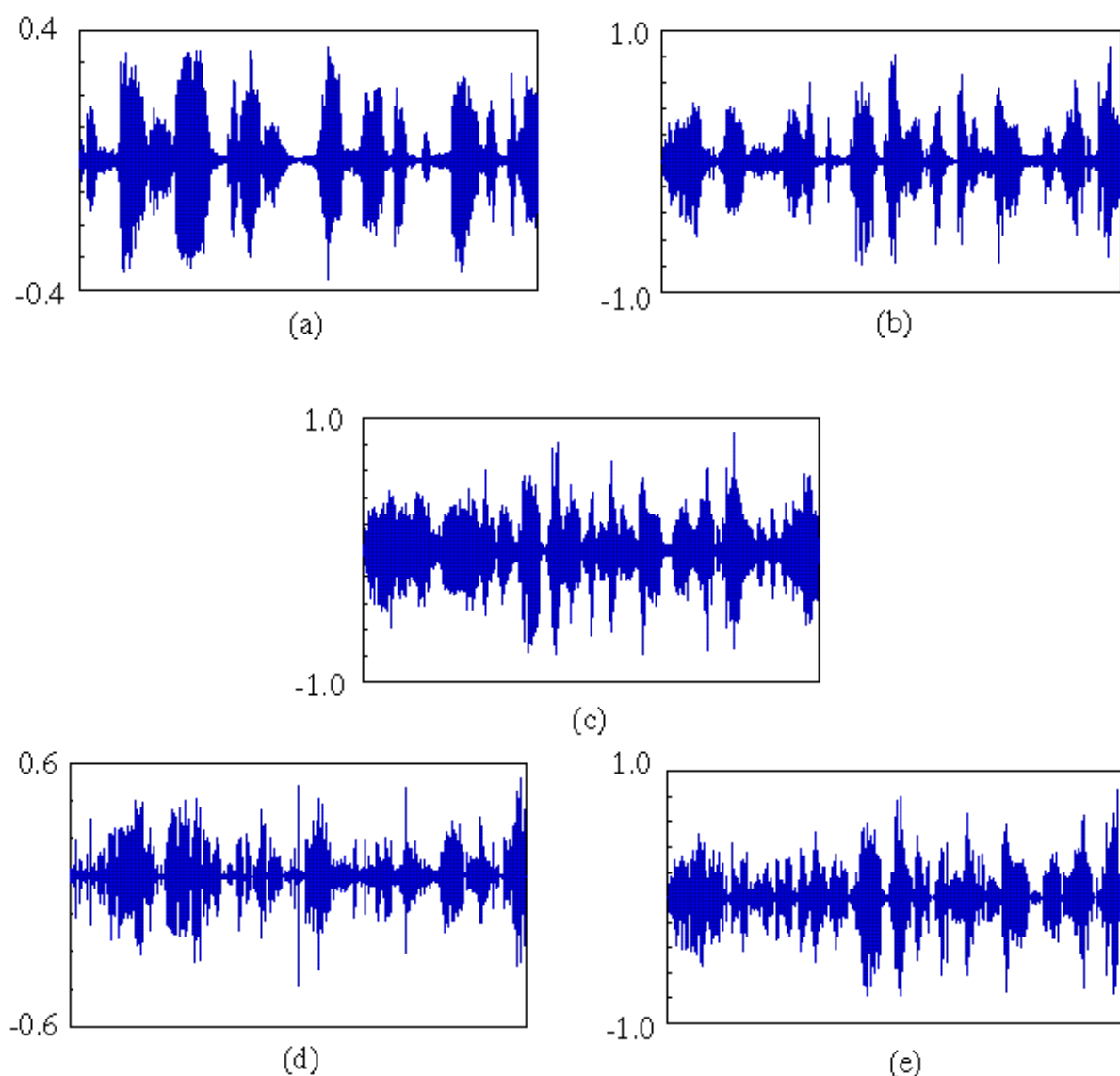


Figure 2. Simulations Results: (a) Source signal $S_1(t)$ (b) Source signal $S_2(t)$ (c) Observed signal $Y_1(t)$ (d) Estimated signal of $S_1(t)$ (e) Estimated signal of $S_2(t)$

signal Separation (ICA99), J. F. Cardoso, Ch. Jutten, and Ph. loubaton, Eds., Aussois, FRANCE, 11-15 January 1999, pp. 461–465.

- [18] L. De Lathauwer, P. Comon, B. De Moor, and J. Vandewalle, “ICA algorithms for 3 sources and 2 sensors,” in *IEEE SP Int Workshop on High Order Statistics, HOS 99*, Caeserea, Israel, 12-14 June 1999, pp. 116–120.
- [19] A. Taleb, “An algorithm for the blind identification of n independent signals with 2 sensors,” in *Sixth International Symposium on Signal Processing and its Applications (ISSPA 2001)*, M. Deriche, Boashash, and W. W. Boles, Eds., Kuala-Lampur, Malaysia, August 13-16 2001.
- [20] G. Darmois, “Analyse générale des liaisons stochastiques,” *Rev. Inst. Intern. Stat.*, vol. 21, pp. 2–8, 1953.
- [21] K. I. Nakadai, K. Hidai, H. G. Okuno, and H. kitano, “Real-time speaker localization and speech separation by audio-visual integration,” in *17th international Joint Conference on Artificial Intelligence (IJCAI-01)*, Seattle, USA, August 2001, pp. 1425–1432.
- [22] H. G. Okuno, K. Nakadai, T. Lourens, and H. kitano, “Separating three simultaneous speeches with two microphones by integrating auditory and visual processing,” in *European Conference on Speech Processing*, Aalborg, Denmark, September 2001, pp. 2643–2646.

- [23] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. on Signal Processing*, vol. 46, no. 11, pp. 2888–2897, 1998.
- [24] P. Flandrin, *Time-Frequency / Time-Scale analysis*, Academic Press, Paris, 1999.
- [25] D. Le Guen and A. Mansour, "Automatic recognition algorithm for digitally modulated signals," in *6th Baiona workshop on signal processing in communications*, Baiona, Spain, 25-28 June 2003, p. To appear.
- [26] J. Thiemann, *Acoustic noise suppression for speech signals using auditory masking effects*, Ph.D. thesis, Department of Electrical & Computer Engineering, McGill University, Canada, July 2001.
- [27] R. Le Bouquin, *Traitement pour la reduction du bruit sur la parole application aux communications radio-mobiles.*, Ph.D. thesis, L'université de Rennes I, July 1991.
- [28] A. Jefremov and B. Kleijn, "Sline-based continuous-time pitch estimation," in *Proceedings of International Conference on Acoustics Speech and Signal Processing 2002, ICASSP 2002*, Orlando, Florida, U.S.A, 13-17 May 2002.