A TIME-FREQUENCY CORRELATION-BASED BLIND SOURCE SEPARATION METHOD FOR TIME-DELAYED MIXTURES

Matthieu Puigt, Yannick Deville

Laboratoire d'Astrophysique de Toulouse-Tarbes
Observatoire Midi-Pyrénées - Université Paul Sabatier Toulouse 3
14 Av. Edouard Belin, 31400 Toulouse, France
mpuigt@ast.obs-mip.fr, ydeville@ast.obs-mip.fr

ABSTRACT

We propose a time-frequency (TF) blind source separation (BSS) method suited to attenuated and delayed (AD) mixtures, inspired from a method that we previously developed for linear instantaneous mixtures. This approach only requires each of the uncorrelated sources to occur alone in a tiny TF zone, i.e. it sets very limited constraints on the source sparsity and overlap, unlike various previously reported TF-BSS methods. Our approach is based on TIme-Frequency CORRelation (hence its name AD-TIFCORR). It consists in identifying the columns of the (filtered permuted) mixing matrix in TF zones where it detects that a single source occurs. We thus identify columns of scale coefficients and time shifts. This method is especially suited to non-stationary sources.

1. INTRODUCTION

Blind source separation (BSS) consists in estimating a set of N unknown sources from a set of P observations resulting from mixtures of these sources through unknown propagation channels. Most of the approaches that have been developed to this end are based on Independent Component Analysis [1]. More recently, several methods based on time-frequency (TF) analysis have been reported [2, 3, 4, 5, 6]. In particular, some methods based on ratios of TF transforms of the observed signals have been proposed [3, 4]. Some of these methods, i.e. DUET and its modified versions [4], apply to attenuation and delay (AD) channels but require the sources to have no overlap in the TF domain, which is quite restrictive. On the contrary, only slight differences in the TF representations of the sources are requested by the linear instantaneous (LI) TIFROM method that we proposed in [3]. This feature is also provided by the LI-TIFCORR method that we introduced in [5]. We here propose a novel TF-BSS method which is inspired from this LI-TIFCORR method, but suited to AD mixtures. We thus avoid the restriction of the DUET method concerning the sparsity of the sources in the TF domain, while addressing the same class of mixtures.

2. PROBLEM STATEMENT

In this paper, we assume that N unknown sources signals $s_j(n)$ are transferred through AD channels and added, thus providing a set of N mixed observed signals $x_i(n)$. This reads

$$x_i(n) = \sum_{j=1}^{N} a_{ij}\, s_j(n - n_{ij}) \qquad i = 1 \dots N, \tag{1}$$

where a_{ij} are real-valued strictly positive constant scale coefficients and n_{ij} are integer-valued time shifts. We here handle the scale/filter indeterminacies inherent in the BSS problem by extending to AD mixtures the approach that we introduced in [5]. We therefore consider an arbitrary permutation function $\sigma(.)$, applied to the indices j of the source signals, which yields the permuted source signals $s_{\sigma(j)}(n)$. We then introduce scaled and time-shifted versions of the latter signals, equal to their contributions in the first mixed signal, i.e.

$$s'_j(n) = a_{1,\sigma(j)}\, s_{\sigma(j)}\left(n - n_{1,\sigma(j)}\right). \tag{2}$$

The mixing equation (1) may then be rewritten as

$$x_i(n) = \sum_{j=1}^{N} a_{i,\sigma(j)}\, s_{\sigma(j)}\left(n - n_{i,\sigma(j)}\right) = \sum_{j=1}^{N} b_{ij}\, s'_j\left(n - \mu_{ij}\right) \tag{3}$$

with

$$\begin{cases} b_{ij} & = \frac{a_{i,\sigma(j)}}{a_{1,\sigma(j)}} \\ \mu_{ij} & = n_{i,\sigma(j)} - n_{1,\sigma(j)} \end{cases} \tag{4}$$

The Fourier transform of Eq. (3) reads

$$X_i(\omega) = \sum_{j=1}^{N} b_{ij}\, e^{-j\omega\mu_{ij}}\, S'_j(\omega) \qquad i = 1 \dots N. \tag{5}$$

This yields in matrix form

$$\underline{X}(\omega) = B(\omega)\, \underline{S'}(\omega) \tag{6}$$

where $\underline{S'}(\omega) = [S'_1(\omega) \cdots S'_N(\omega)]^T$ and

$$B(\omega) = \left[b_{ij}\, e^{-j\omega\mu_{ij}}\right] \qquad i, j = 1 \dots N. \tag{7}$$

In this paper, we aim at introducing a new method for estimating $B(\omega)$.

3. A NEW TIFCORR METHOD FOR AD MIXTURES

3.1. Time-frequency tool and assumptions

As stated in Section 1, we recently proposed [5] a LI TIme-Frequency CORRelation-based BSS method, that we therefore called "LI-TIFCORR". Starting from this method, we here develop a version which is extended to AD mixtures (hence its name AD-TIFCORR). The TF transform of the signals considered in that approach is the Short-Time Fourier Transform (STFT) defined as:

$$U(n, \omega) = \sum_{n'=-\infty}^{+\infty} u(n') h(n' - n) e^{-j\omega n'} \tag{8}$$

where $h(n'-n)$ is a shifted windowing function, centered on time n. $U(n,\omega)$ is the contribution of the signal $u(n)$ in the TF window corresponding to the short time window centered on n and to the angular frequency ω. Our approach uses the following definitions and assumptions.

Definition 1 *A TF "analysis zone" is a series of adjacent TF windows (n,ω). More precisely, a "constant-frequency (resp. constant-time) analysis zone" is a series of M time-adjacent (resp. M' frequency-adjacent) analysis windows. We denote it (T,ω) (resp. (n,Ω)).*

Definition 2 *A source is said to "occur alone" in an analysis zone if only this source has a TF transform which is not equal to zero everywhere in this analysis zone.*

Definition 3 *A source is said to be "visible" in the TF domain if there exist at least one analysis zone where it occurs alone.*

Assumption 1 *i) Each source is visible in the TF domain and ii) in all considered analysis zones, the variance of at least one source is nonzero.*

Note that Assumption 1-i) is a very limited sparsity constraint. Assumption 1-ii) is only made for the sake of simplicity: it may be removed in practice, thanks to the noise contained by real recordings, for the same reasons as in [3].

Assumption 2 *Over each analysis zone (T,ω), the TF transforms of the sources are uncorrelated.*

3.2. Overall structure of AD-TIFCORR

The AD-TIFCORR method aims at estimating the mixing matrix $B(\omega)$ defined in (7), i.e. the scale coefficients b_{im} and the associated time shifts μ_{im}, with $i=2\dots N$ and $m=1\dots N$ ($i=1$ yields $b_{ij}=1$ and $\mu_{ij}=0$: see Eq. (4)). It is composed of 4 main stages, preceded by a pre-processing stage:

1. The pre-processing stage consists in deriving the STFTs $X_i(n,\omega)$ of the mixed signals, according to Eq. (8).
2. We detect single-source constant-frequency analysis zones and identify the columns of scale coefficients b_{im} in these zones as described in Section 3.3.
3. We detect single-source constant-time analysis zones and then identify the columns of time shifts μ_{im} as explained in Section 3.4.
4. We couple the above-identified scale coefficients and time shifts, using the method proposed in Section 3.5.
5. In the combination stage, we eventually compute the output signals. They may be obtained in the frequency domain by computing
$$\underline{Y}(\omega)=B^{-1}(\omega)\,\underline{X}(\omega) \tag{9}$$
where $\underline{Y}(\omega)=[Y_1(\omega)\cdots Y_N(\omega)]^T$ is the vector of Fourier transforms of the output signals. The time-domain versions of these signals are then obtained by applying an inverse Fourier transform to $\underline{Y}(\omega)$.

3.3. Detection of single-source zones and identification of b_{im}

As stated above, the BSS method that we here introduce first includes a detection stage for finding single-source constant-frequency analysis zones. The frequency-domain mixture equations corresponding to Eq. (1) read

$$X_i(\omega)=\sum_{j=1}^{N} a_{ij}\,e^{-j\omega n_{ij}}\,S_j(\omega) \qquad i=1\dots N. \tag{10}$$

This relationship between the observations and sources remains almost exact when expressed in the TF domain if the time shifts n_{ij} are small enough as compared to the temporal width of the windowing function $h(.)$ used in the STFT transform. In this paper, we assume that this condition is met and therefore that the STFTs of the observations can be expressed with respect to the STFTs of the sources as

$$X_i(n,\omega)=\sum_{j=1}^{N} a_{ij}\,e^{-j\omega n_{ij}}\,S_j(n,\omega) \qquad i=1\dots N. \tag{11}$$

For any couple of signals $v_1(n)$ and $v_2(n)$, we define the cross correlation of the (non-centered version of the) TF transforms of these signals over the considered constant-frequency analysis zone (T,ω) as

$$R_{v_1v_2}(T,\omega)=\frac{1}{M}\sum_{p=1}^{M} V_1(n_p,\omega)\,V_2^*(n_p,\omega), \tag{12}$$

where the superscript * denotes the complex conjugate. The corresponding correlation coefficient reads

$$r_{v_1v_2}(T,\omega)=\frac{R_{v_1v_2}(T,\omega)}{\sqrt{R_{v_1v_1}(T,\omega)R_{v_2v_2}(T,\omega)}}. \tag{13}$$

Applying the general proof of [5] to the mixtures expressed in Eq. (11) directly shows that a necessary and sufficient condition for a source to occur alone in the TF analysis zone (T,ω) is

$$|r_{x_1x_i}(T,\omega)|=1 \qquad \forall i,\,2\le i\le N. \tag{14}$$

In practice, for each analysis zone, we compute the mean $\overline{|r_{x_1x_i}(T,\omega)|}$ of $|r_{x_1x_i}(T,\omega)|$ over all i ($2\le i\le N$). We then order all analysis zones according to decreasing values of $\overline{|r_{x_1x_i}(T,\omega)|}$. The first zones in this ordered list are then considered as the "best" single-source zones.

If a source $S_k(n,\omega)$ occurs alone in the considered TF window (n_p,ω_l) then Eq. (11) and (12) yield

$$I_i(T,\omega_l)=\frac{R_{x_ix_1}(T,\omega_l)}{R_{x_1x_1}(T,\omega_l)}=\frac{a_{ik}}{a_{1k}}e^{-j\omega(n_{ik}-n_{1k})}=b_{im}e^{-j\omega\mu_{im}} \tag{15}$$

with b_{im} and μ_{im} defined by Eq. (4) and $k=\sigma(m)$. Since we assumed all mixing coefficients a_{ik} to be real and positive, all resulting scale coefficients b_{im} are also real and positive. The modulus of the parameter value $I_i(T,\omega_l)$ provided in Eq. (15) is thus equal to b_{im}. The identification stage for the scale coefficients b_{im} therefore consists in successively considering the first analysis zones of the above ordered list. For each such zone, the estimates of b_{im} associated to a column of $B(\omega)$ are set to the values of $|I_i(T,\omega_l)|$. A new column of b_{im} is kept if its distance with respect to each previously found column of b_{im} is above a user-defined threshold ϵ_1. The identification procedure ends when the number of columns of scale coefficients thus kept becomes equal to the specified number N of sources to be separated.

3.4. Detection of single-source zones and identification of μ_{im}

In this section, we describe the detection and identification stages for estimating the time shifts μ_{im}. This approach is based on ratios of mixtures in the TF domain, defined as

$$\alpha_i(n,\omega)=\frac{X_i(n,\omega)}{X_1(n,\omega)}=\frac{\sum_{j=1}^{N} a_{ij}\,e^{-j\omega n_{ij}}\,S_j(n,\omega)}{\sum_{j=1}^{N} a_{1j}\,e^{-j\omega n_{1j}}\,S_j(n,\omega)}. \tag{16}$$

If a source $S_k(n,\omega)$ occurs alone in the considered TF window $(n_{p'},\omega_{l'})$ then, denoting $k=\sigma(m')$, we get

$$\alpha_i(n_{p'},\omega_{l'}) = \frac{a_{ik}}{a_{1k}} e^{-j\omega(n_{ik}-n_{1k})} = b_{im'} e^{-j\omega\mu_{im'}}. \qquad (17)$$

Thanks to expression (17) of the parameters $\alpha_i(n_{p'},\omega_{l'})$ in single-source analysis zones, a natural idea for estimating the time shifts $\mu_{im'}$ consists in taking advantage of the phase of $\alpha_i(n_{p'},\omega_{l'})$[1]. We consider independently each time position $n_{p'}$ associated to TF windows and for each such position, we unwrap the phase of $\alpha_i(n_{p'},\omega_{l'})$ over all associated frequency-adjacent TF windows. If $S_k(n,\omega)$ occurs alone in the analysis zone $(n_{p'},\Omega)$ and we consider the unwrapped phase $\phi_i(n_{p'},\omega_{l'})$ of $\alpha_i(n_{p'},\omega_{l'})$ in this zone, due to (17) we have

$$-\omega_{l'}\mu_{im'} = \phi_i(n_{p'},\omega_{l'}) + 2q_{im'}(n_{p'})\pi, \qquad (18)$$

where $q_{im'}(n_{p'})$ is an unknown integer. Eq. (18) shows that the curve associated to the variations of the phase $\phi_i(n_{p'},\omega_{l'})$ with respect to $\omega_{l'}$ in a single-source analysis zone $(n_{p'},\Omega)$ is a line and that its slope does not depend on the value of $q_{im'}(n_{p'})$ and is equal to $-\mu_{im'}$. This therefore provides not only a means for identifying $\mu_{im'}$, with no phase indeterminacy, but also a way to detect constant-time single-source zones.

The overall detection and identification method that we propose for the parameters $\mu_{im'}$ then operates as follows. We successively consider all constant-time analysis zones $(n_{p'},\Omega)$. In each such zone $(n_{p'},\Omega)$, for each observed signal with index i, we consider the M' points which have two coordinates, resp. defined as the frequencies $\omega_{l'}$ and the corresponding values $\phi_i(n_{p'},\omega_{l'})$ of the unwrapped phase of the parameter $\alpha_i(n_{p'},\omega_{l'})$. We determine the least-mean square regression line and the mean-square error of the points $(\omega_{l'},\phi_i(n_{p'},\omega_{l'}))$ with respect to their associated regression line. The estimates of the tentative parameters $\mu_{im'}$ are set to the integers which are the closest to the opposite of the slopes of the regression lines. The best single-source zones are those which yield the lowest mean-square errors. These zones may be then be used in various ways for eventually identifying the parameters $\mu_{im'}$, e.g. by ordering these zones according to increasing values of their mean-square error or by using clustering techniques. They eventually yield a set of column vectors. Each of these vectors contains the values $\mu_{im'}$, which correspond to all observations with indices i and to the source with index $\sigma(m')$ which occurs in the considered analysis zone. A final stage should therefore be added to our approach, in order to couple each column of parameters $\mu_{im'}$ to the column of parameters b_{im} corresponding to the same source. This stage is described hereafter.

3.5. Coupling the parameters b_{im} and $\mu_{im'}$

3.5.1. Alternative identification method for the parameters b_{im}

In Section 3.4, we introduced a method for detecting constant-time single-source analysis zones and we showed how the phase of the parameters $\alpha_i(n,\omega)$ in such zones may be used to identify the parameters $\mu_{im'}$. We here note that the moduli of these parameters in these zones also make it possible to identify the parameters $b_{im'}$: Eq. (17) shows that, at any frequency $\omega_{l'}$ of such a zone, the modulus of $\alpha_i(n_{p'},\omega_{l'})$ is equal to $b_{im'}$. The latter parameter may therefore be identified as the mean value of the modulus of $\alpha_i(n,\omega)$ over a constant-time single-source analysis zone.

The value thus obtained is denoted $b'_{im'}$ below, in order to distinguish it from the value b_{im} provided by the method that we introduced in Section 3.3. The alternative approach that we propose in this subsection is attractive because each considered analysis zone yields the parameters $b'_{im'}$ and $\mu_{im'}$ corresponding to the *same* source. It therefore inherently provides a solution to the coupling of these types of parameters. However, our experimental tests showed that the parameter value $b'_{im'}$ thus obtained estimate less accurately the actual mixture parameters than the values b_{im} that we obtained in Section 3.3. We therefore introduce a modified approach which takes advantage of both types of parameters hereafter.

3.5.2. Coupling the parameters b_{im} and $(b'_{im'},\mu_{im'})$

Taking advantage of all above-defined principles, we now introduce a method for eventually coupling the parameters b_{im} and $\mu_{im'}$. This method consists in:

1. determining the parameters b_{im} as explained in Section 3.3,
2. independently determining the couples $(b'_{im'},\mu_{im'})$ as explained in Sections 3.4 and 3.5.1,
3. and then mapping the parameters $\mu_{im'}$ towards the parameters b_{im} thanks to the parameters $b'_{im'}$. This is achieved as follows. The above identification of the parameters b_{im} yields N columns of such parameters, each associated with a different source. In the detection of constant-time single-source analysis zones, we keep a number of zones significantly larger than N, by selecting all the zones where the mean-square error with respect to the associated regression line is below a user-defined threshold ϵ_2. For each such zone, we identify the two columns that resp. contain the parameters $b'_{im'}$ and $\mu_{im'}$ corresponding to that zone. We then consider the parameters $b'_{im'}$ and b_{im} resp. as coarse and accurate estimates of the scale parameters associated to the mixing matrix and we map each column of $b'_{im'}$ towards the closest column[2] of b_{im}. Since the parameters $b'_{im'}$ were already coupled with the parameters $\mu_{im'}$, the latter parameters are thus mapped towards the N columns of parameters b_{im}. For each element (associated to the observation index i) in each such column of b_{im}, we should eventually keep only one parameter value $\mu_{im'}$. This is achieved as follows for each such element: among all the values $\mu_{im'}$ which were mapped above towards this element, we keep the value which has the highest number of occurences.

4. EXPERIMENTAL RESULTS

We now present various tests performed with $N=2$ English speech sources sampled at 20 kHz. These signals consist of 2.5 s of continuous speech from different male speakers. Both sources were first centered and scaled so that their highest absolute value is equal to 1. The performance achieved in each test is measured by the overall signal-to-interference-ratio (SIR) Improvement achieved by this system, denoted $SIRI$ below, and defined as the ratio of the output and input SIRs of our BSS system.

All our tests aim at estimating the influence of the time shifts n_{ij} on

[1] One could think of using the phase of $I_i(T,\omega_l)$ instead (see Eq. (15)). However, this approach failed in our first tests, presumably due to the averaging over the temporal zone T used in $I_i(T,\omega_l)$ instead of the single temporal window $n_{p'}$ in $\alpha_i(n_{p'},\omega_{l'})$.

[2] A user-defined threshold ϵ_3 is used to ignore any column of parameters $b'_{im'}$ such that all its distances with respect to the N columns of parameters b_{im} are above that threshold.

η	0	25	100	200
Frobenius norm	2.8e-5	1.6e-2	3.6e-2	6.7e-2

Table 1. Frobenius norm of the difference between the actual matrices of parameters b_{im} and their estimates provided by AD-TIFCORR.

the performance of the proposed method. We therefore use symmetrical mixing matrices defined as

$$A(\omega) = \begin{bmatrix} 1 & 0.9\, e^{-j\omega\eta} \\ 0.9\, e^{-j\omega\eta} & 1 \end{bmatrix}. \qquad (19)$$

The values that we considered for η are $\eta = 0, 25, 100$ and 200. The input SIR of our BSS system is equal to 0.9 dB.

As explained in Section 3.2, the proposed method uses TF representations of the observed signals $x_i(n)$, obtained by computing their STFTs $X_i(n, \omega)$. More precisely, this type of representation is used twice in the AD-TIFCORR approach, i.e. first when considering constant-frequency analysis zones used for estimating the parameters b_{im}, and then when considering constant-time analysis zones used for estimating the parameters $b'_{im'}$ and $\mu_{im'}$. These two types of analysis zones may lead to different optimum values as for the parameters of STFTs and numbers of STFT windows per analysis zone. Therefore, we independently considered two sets of such parameters, resp. associated to the above two types of analysis zones in our approach, i.e.:

- We here denote d (resp. d') the number of samples of observed signals $x_i(n)$ in each time window of the STFTs used in constant-frequency (resp. constant-time) analysis zones.
- As stated above in Definition 1, M (resp. M') is the number of adjacent windows in constant-frequency (resp. constant-time) analysis zones.
- We here denote ρ (resp. ρ') the temporal overlap between the time windows in the STFTs used in constant-frequency (resp. constant-time) analysis zones.

For the sake of simplicity, we fix some parameters in the tests reported here, i.e. $d = 256$, $M = 10$ and $\rho = \rho' = 75\%$. We also fix the user-defined thresholds $\epsilon_i (i = 1, 2, 3)$ to 0.15, 0.1 and 0.1.

The other parameters of our BSS method are varied as follows. The number d' of samples per STFT window is geometrically varied from 512 to 32768 samples. The number M' of windows per analysis zone is set to 16 when $d' = 512$. This value of M' is then increased geometrically with d'. Thus, the absolute width of the frequency bands associated to the frequency domain Ω of the analysis zones $(n_{p'}, \Omega)$ takes the same value whatever d' and is here equal to 625 Hz.

In our approach, the estimates of b_{im} are independent from the parameters d' and M' varied in these tests. Table 1 provides the Frobenius norm of the difference between the estimated and actual matrices of parameters b_{im}. This norm is quite low, showing that our method always succeeds in identifying the parameters b_{im} very accurately.

The overall performance of our approach is shown in Table 2. This table shows that the range of STFT window sizes d' which yields the best performance is $2048 \leq d' \leq 8192$. The $SIRI$s are then above approximately 20 dB even for the highest considered time shift η. Note that the $SIRI$s significantly decrease when η increases. When $d' > 8192$, the method fails finding exactly the actual columns of time shifts in a significant number of tests because the sources tend to become invisible for long STFTs. The method also fails in some cases when $d' \leq 1024$ because the time shifts are non negligible with respect to the STFT window size d'. It succeeds in finding all columns of estimated μ_{im} *exactly* equal to theoretical values in 68% of the cases considered in Table 2. The variations of $SIRI$ with respect to η are also reflected in this success rate for μ_{im}: it decreases from 100% when $\eta = 0$ to 43% when $\eta = 200$.

η	STFT window size d'						
	512	1024	2048	4096	8192	16384	32768
0	76.8	76.8	76.8	76.8	76.8	76.8	76.8
25	32.3	32.3	32.3	32.3	32.3	-13.1	inv.
100	-17.3	-18.3	22.5	22.5	22.5	22.5	inv.
200	-15.7	4.1	19.8	19.8	19.8	5.9	inv.

Table 2. Performance ($SIRI$ in dB) for $\eta = 0$, 25, 100, 200 vs STFT window size d'. "inv." means invisible.

5. CONCLUSION AND EXTENSIONS

In this paper, we proposed a TF BSS method for AD mixtures. It avoids the restrictions of the DUET method, which needs the source to be (approximately) W-disjoint orthogonal. Our approach consists in first finding the TF zones where a source occurs alone and then identifying in these zones the parameters of the (filtered permuted) mixing matrix. Thanks to this principle, this approach applies to non-stationary sources, provided there exists at least a tiny TF zone per source where this source occurs alone. We presented various aspects of the experimental performance of this approach. In our future investigations, we will perform a more detailed characterization of its performance. We will also aim at studying the usefulness of clustering techniques (such as those proposed for LI mixtures in [6]) in this approach and at extending it to general convolutive mixtures.

6. REFERENCES

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, New York, 2001.

[2] A. Belouchrani and M. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 2888–2897, November 1998.

[3] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Processing*, vol. 85, Issue 7, pp. 1389–1403, July 2005.

[4] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures," in *Proceedings of ICASSP 2000*, Istanbul, Turkey, June 5-9 2000, vol. 5, pp. 2985–2988, IEEE Press.

[5] Y. Deville, "Temporal and time-frequency correlation-based blind source separation methods," in *Proceedings of ICA 2003*, Nara, Japan, April 1-4 2003, pp. 1059–1064.

[6] D. Smith, J. Lukasiak, and I. Burnett, "Two channel, block adaptive audio separation using the cross correlation of time frequency information," in *Proceedings of ICA 2004*, Granada, Spain, September 22-24 2004, pp. 889–897.