

Time-Domain Fast Fixed-Point Algorithms for Convolutive ICA

Johan Thomas, Yannick Deville, *Member, IEEE*, and Shahram Hosseini

Abstract—This letter presents new blind separation methods for moving average (MA) convolutive mixtures of independent MA processes. They consist of time-domain extensions of the FastICA algorithms developed by Hyvarinen and Oja for instantaneous mixtures. They perform a convolutive sphering in order to use parameter-free fast fixed-point algorithms associated with kurtotic or negentropic non-Gaussianity criteria for estimating the source innovation processes. We prove the relevance of this approach by mapping the mixtures into linear instantaneous ones. Test results are presented for artificial colored signals and speech signals.

Index Terms—Convolutive mixtures, fixed-point algorithms, independent component analysis (ICA), non-Gaussian signals.

I. INTRODUCTION

BLIND source separation (BSS) consists in estimating a set of N unobserved source signals from P observed mixtures of these sources where the mixture parameters are unknown. Let us denote by $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^T$ the vector of sources and by $\mathbf{x}(n) = [x_1(n), \dots, x_P(n)]^T$ the observations. We consider convolutive mixtures defined by a set of unknown filters with impulse responses $h_{ij}(n)$, where $i = 1, \dots, P$ and $j = 1, \dots, N$. The relations between the sources and the observations may be expressed in the time domain as

$$x_i(n) = \sum_{j=1}^N \sum_{k=-\infty}^{\infty} h_{ij}(k) s_j(n-k), \quad \forall i = 1, \dots, P. \quad (1)$$

The overall relationship then reads in the \mathcal{Z} domain

$$\mathbf{X}(z) = \mathbf{H}(z) \cdot \mathbf{S}(z) \quad (2)$$

where $\mathbf{X}(z)$ and $\mathbf{S}(z)$ are, respectively, the \mathcal{Z} transforms of $\mathbf{x}(n)$ and $\mathbf{s}(n)$, and the $P \times N$ matrix $\mathbf{H}(z)$ consists of the transfer functions $H_{ij}(z) = \mathcal{Z}[h_{ij}(n)]$ of the mixing filters.

In this letter, each source $s_j(n)$ is assumed to be expressed in the \mathcal{Z} domain as

$$S_j(z) = F_j(z) \cdot U_j(z) \quad (3)$$

where $F_j(z)$ corresponds to a filter, and $U_j(z)$ is the \mathcal{Z} transform of a process $u_j(n)$, which is the innovation process of

$s_j(n)$. Denoting $\mathbf{U}(z) = [U_1(z), \dots, U_N(z)]^T$, we can then express the mixing (2) as

$$\mathbf{X}(z) = \mathbf{G}(z) \cdot \mathbf{U}(z) \quad (4)$$

where $\mathbf{G}(z) = \mathbf{H}(z) \cdot \mathbf{F}(z)$, with $\mathbf{F}(z) = \text{diag}(F_1(z), \dots, F_N(z))$.

We make the following assumptions concerning the above mixture model.

- The process $\mathbf{u}(n) = [u_1(n), \dots, u_N(n)]^T$ is real-valued, zero-mean, independent and identically distributed (i.i.d.), and spatially independent, i.e., its components $u_j(n)$ are statistically independent of each other but do not necessarily have the same distribution. We also assume that, at most, one of these components is Gaussian.
- The filter matrices $\mathbf{F}(z)$, $\mathbf{H}(z)$, and thus $\mathbf{G}(z)$ are causal, finite impulse response (FIR), and nonsingular. Note that infinite impulse response (IIR) systems can also be approximated by equivalent (high-order) FIR models.

The goal of convolutive BSS is typically to estimate the contributions of all sources in each observation, i.e., $H_{ij}(z) \cdot S_j(z)$. In deflation-based methods such as [1], this is achieved by using the following procedure.

- 1) Extract the innovation process $u_j(n)$ of a source $s_j(n)$ from the observations.
- 2) Identify P coloring filters and apply them to $u_j(n)$ in order to recover the contributions of $s_j(n)$ in each observation, i.e., $H_{ij}(z) \cdot S_j(z)$.
- 3) Subtract these contributions from all the observations.
- 4) Set $N \leftarrow N - 1$. If $N \neq 1$, go back to step 1) in order to extract another source.

We here consider time-domain BSS methods, which use non-Gaussianity as a criterion to realize the first step of the above procedure and which are therefore based on independent component analysis (ICA) [2]. In the next section, we analyze the principles and limitations of the existing methods, and we propose an approach to extend them so as to obtain the currently missing fast-converging kurtotic and negentropic methods for convolutive mixtures. The experimental performance of the proposed methods is presented in Section III, and conclusions are drawn from this investigation in Section IV.

II. ANALYSIS AND EXTENSION OF BSS METHODS BASED ON NON-GAUSSIANITY

A. Previously Reported Approaches

Delfosse and Loubaton [3] proposed the first deflation-based kurtotic BSS method for *linear instantaneous* mixtures, where the filters $H_{ij}(z)$ are replaced by scalar coefficients. This method

Manuscript received September 15, 2005; revised November 23, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue (Joseph) Wang.

The authors are with the Laboratoire d'Astrophysique de Toulouse-Tarbes—Université Paul Sabatier Toulouse 3, 31400 Toulouse, France (e-mail: jthomas@ast.obs-mip.fr; ydeville@ast.obs-mip.fr; shosseini@ast.obs-mip.fr).

Digital Object Identifier 10.1109/LSP.2005.863638

first consists in deriving a sphered version $\mathbf{v}(n)$ of the observation vector $\mathbf{x}(n)$, i.e., a set of linear combinations of these observations composed of signals that are mutually uncorrelated at time n and that have unit variances. A first output signal is then derived as a linear combination $y(n) = \mathbf{w}^T \mathbf{v}(n)$ of the sphered observations, with a normalized coefficient vector \mathbf{w} selected so as to maximize the square (or the absolute value) of the *nonnormalized* kurtosis of $y(n)$ defined by $kurt(y(n)) = E\{y(n)^4\} - 3(E\{y(n)^2\})^2$ for a zero-mean signal. Delfosse and Loubaton proved in [3] that the local maxima of this criterion correspond to the separation points. They used a gradient-like method to maximize this criterion. This requires one to select an adequate adaptation gain and anyway yields slow convergence. Hyvarinen and Oja solved this problem by introducing a fixed-point algorithm for optimizing the above criterion [2].

A different approach was proposed by Tugnait for *convolutive* mixtures [1]. It directly operates on the observations, i.e., without first sphering them, but then uses the absolute value of the *normalized* kurtosis of the output signal $y(n)$, i.e., $kurt_N(y(n)) = (kurt(y(n)) / (E\{y(n)^2\})^2)$, as the separation criterion. Tugnait proved that the separation points correspond to the local maxima of this criterion when recombining the observations with doubly infinite extraction filters. He proposed to optimize this criterion by using a gradient-based approach, which again yields slow convergence. The tests performed in our team [4] showed that, even when using Newton's optimization scheme instead, convergence remains slow, especially for high-order mixing filters.

This letter therefore aims at filling the gap that results from the above approaches, i.e., at introducing fast-converging kurtotic or negentropic methods for convolutive mixtures. To this end, we first investigate how to extend to convolutive mixtures the approach based on sphering and fixed-point optimization of nonnormalized kurtosis that has been proposed for instantaneous mixtures.

B. New Methods for Extracting an Innovation Process

All above methods require a normalization, as the nonnormalized kurtosis of $y(n)$ tends to infinity when the power of $y(n)$ tends to infinity. In Tugnait's approach, the criterion itself is normalized as the method consists in estimating one of the innovation processes $u_j(n)$ up to a delay and scale factor by maximizing the absolute value of the normalized kurtosis of a convolutive combination of the observations defined as

$$y(n) = \sum_{p=1}^P k_p(n) * x_p(n) = \sum_{p=1}^P \sum_{r=-R}^R k_p(r) x_p(n-r) \quad (5)$$

where $k_p(n)$, $p=1, \dots, P$ are P noncausal FIR filters in practice.

Instead, the two linear instantaneous approaches mentioned in the previous subsection use nonnormalized kurtosis and are based on a normalization of the power of $y(n)$. This results from the sphering stage of these approaches, which yields $E\{y(n)^2\} = \|\mathbf{w}\|^2$, so that selecting \mathbf{w} with $\|\mathbf{w}\|^2 = 1$ guarantees that $E\{y(n)^2\} = 1$. We here extend this method to convolutive mixtures. To this end, the first step of our approach

performs a "convolutive sphering" of the observations, defined as follows. At any time n , we consider the column vector

$$\tilde{\mathbf{x}}(n) = [x_1(n+R), \dots, x_1(n-R), \dots, x_P(n+R), \dots, x_P(n-R)]^T \quad (6)$$

which contains $M = (2R+1)P$ entries. We derive the M -entry column vector $\mathbf{x}'(n) = [x'_1(n), \dots, x'_M(n)]^T$ defined as

$$\mathbf{x}'(n) = \mathbf{B}\tilde{\mathbf{x}}(n) \quad (7)$$

where \mathbf{B} is an $M \times M$ matrix chosen so that

$$E\{x'_i(n)x'_j(n)\} = \delta_{ij}, \forall i, j \in \{1, \dots, M\}. \quad (8)$$

With respect to $\tilde{\mathbf{x}}(n)$, operation (7) may therefore be considered as conventional sphering, which consists of principal component analysis and normalization. Now, with respect to the original observations $x_i(n)$, this may be interpreted differently. Indeed, (6) and (7) show that the signals $x'_i(n)$ are convolutive mixtures of the $x_i(n)$. Equation (8) then means that the signals $x'_i(n)$ are created so as to have unit variances and to be mutually uncorrelated, which may be seen as a spatiotemporal whitening and normalization of the observations $x_i(n)$. Let us denote by $y(n)$ the extracted signal

$$y(n) = \mathbf{w}^T \cdot \mathbf{x}'(n) = \sum_{m=1}^M w_m \cdot x'_m(n) \quad (9)$$

where \mathbf{w} is an M -entry extended column vector of extraction coefficients w_m that, together with (7), yields a convolutive combination $y(n)$ of the observations. The power of $y(n)$ reads $E\{y(n)^2\} = \mathbf{w}^T E\{\mathbf{x}(n)\mathbf{x}(n)^T\} \mathbf{w}$. By constraining $x'(n)$ so as to meet (8), we get $E\{y(n)^2\} = \mathbf{w}^T \mathbf{I} \mathbf{w} = \|\mathbf{w}\|^2$. Therefore

$$E\{y(n)^2\} = 1 \iff \|\mathbf{w}\|^2 = 1. \quad (10)$$

Our method then consists in maximizing the absolute value of the nonnormalized kurtosis of $y(n)$ defined by (9) under the constraint $\|\mathbf{w}\|^2 = \mathbf{w}^T \cdot \mathbf{w} = 1$. The Appendix shows that this criterion lets us extract an estimate $e_l(n)$ of a delayed and scaled source innovation process $\alpha_l u_l(n-r)$, under some conditions.

Moreover, powerful algorithms for performing a constrained optimization of the absolute value of the kurtosis of $y(n)$ may then be straightforwardly derived from those previously reported for linear instantaneous mixtures, because the Appendix shows that the convolutive mixtures $\tilde{\mathbf{x}}(n)$ studied in this letter may be reformulated as instantaneous mixtures in the considered conditions. Especially, we propose as an extension of [2] the following convolutive kurtotic fast fixed-point ICA algorithm based on our modified vector \mathbf{w} .

- Initialize \mathbf{w} to a value \mathbf{w}_0 , e.g., using the approaches proposed below.
- Repeat the following steps 1) and 2) until convergence

$$1) \quad \mathbf{w} = E\left\{\mathbf{x}'(\mathbf{w}^T \mathbf{x}')^3\right\} - 3\mathbf{w} \quad (11)$$

$$2) \quad \mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|}. \quad (12)$$

The above-mentioned initial value \mathbf{w}_0 of \mathbf{w} may be selected at random. An improved approach may be obtained by taking ad-

vantage of the relationship that exists between our vector \mathbf{w} and the coefficients of the FIR filters of Tugnait's approach defined in (5). Indeed, let us consider the columns of our sphering matrix \mathbf{B} and index them as

$$\mathbf{B} = [\mathbf{b}_1^{-R}, \dots, \mathbf{b}_1^{+R}, \dots, \mathbf{b}_P^{-R}, \dots, \mathbf{b}_P^{+R}] \quad (13)$$

based on (6) and (7). Using (7), the signal (9) extracted in our method reads $y(n) = \mathbf{w}^T \mathbf{B} \tilde{\mathbf{x}}(n) = \mathbf{w}^T \sum_{p=1}^P \sum_{r=-R}^R \mathbf{b}_p^r x_p(n-r)$. Identifying this expression with the output signal (5) in Tugnait's approach, we obtain

$$k_p(r) = \mathbf{w}^T \cdot \mathbf{b}_p^r, \quad \forall p \in \{1, \dots, P\}, \quad \forall r \in \{-R, \dots, R\} \quad (14)$$

and therefore

$$\mathbf{K} = \mathbf{w}^T \mathbf{B} \quad (15)$$

where the row vector \mathbf{K} consists of the impulse response coefficients of the filters $k_1(n)$ to $k_P(n)$. This relation lets us initialize our vector \mathbf{w} as in Tugnait's method, i.e., with unit filters $k_p(n) = \delta(n)$, so that $y(n)$ is the sum of all observations $x_p(n)$. That corresponds to $\mathbf{K} = \mathbf{K}_0$ defined as

$$\mathbf{K}_0 = \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{k_1}, \dots, \underbrace{[0, \dots, 0, 1, 0, \dots, 0]}_{k_P}. \quad (16)$$

Equation (15) then yields $\mathbf{w}_0^T = \mathbf{K}_0 \mathbf{B}^{-1}$. This initialization of \mathbf{w} provided better experimental results than a random one and is used in Section III.

For instantaneous mixtures, instead of using the kurtosis, another contrast function based on negentropy was proposed by Hyvarinen to estimate non-Gaussianity [5]. It has been shown to yield better robustness and lower variance than the kurtotic approach. In particular, it is more robust to extreme values than the kurtosis criterion, which involves a fourth-order moment, whose estimation is very sensitive to outliers. Furthermore, a fast and reliable fixed-point algorithm was also developed by Hyvarinen for this type of function. We extend this negentropy algorithm to convolutive mixtures in the same way as the above kurtotic approach, using the following adaptation formula instead of (11) in each iteration:

$$1) \quad \mathbf{w} = E \{ \mathbf{x}' g(\mathbf{w}^T \mathbf{x}') \} - E \{ g'(\mathbf{w}^T \mathbf{x}') \} \mathbf{w} \quad (17)$$

where g and g' are the first and second derivatives of a contrast function G that is used for estimating the negentropy.

The proposed kurtotic algorithm may be optimized in terms of the processing time needed for computing the expectation in (11). Indeed, (11) requires to perform a matrix-vector product $\mathbf{x}'(n) = \mathbf{B} \tilde{\mathbf{x}}(n)$ for each sample index n in order to compute the sum of the resulting terms $\mathbf{x}'(n) (\mathbf{w}^T \mathbf{x}'(n))^3$ for expectation estimation. By computing the vector $\mathbf{w}_1 = \mathbf{w}^T \mathbf{B}$ once for the considered iteration of the algorithm and by replacing \mathbf{x}' by $\mathbf{B} \tilde{\mathbf{x}}$ in the update (11), we obtain

$$1) \quad \mathbf{w} = \mathbf{B} E \left\{ \tilde{\mathbf{x}} (\mathbf{w}_1^T \tilde{\mathbf{x}})^3 \right\} - 3\mathbf{w}. \quad (18)$$

We hence compute only two matrix-vector products to estimate the expectation, i.e., one in $\mathbf{w}_1 = \mathbf{w}^T \mathbf{B}$ and one in (18). The same principle applies to the negentropy algorithm (17).

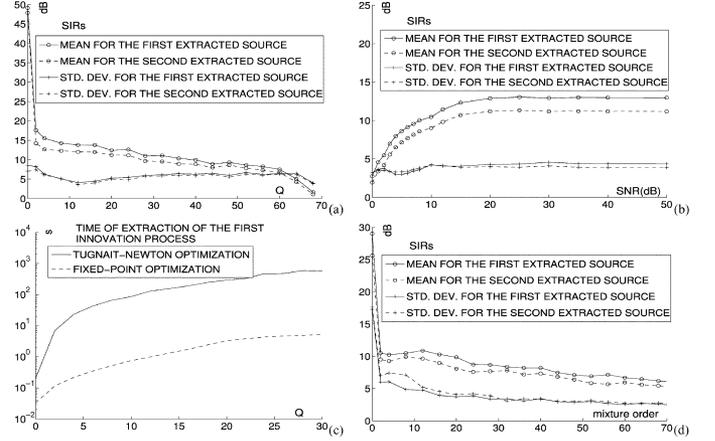


Fig. 1. (a) SIRs of extracted sources depending on Q for artificial colored signals. (b) SIRs of extracted sources of interest depending on the input SNR. (c) Time of extraction of the first innovation process depending on Q for Tugnait-Newton optimization and for our fixed-point optimization. (d) SIRs of extracted sources depending on the mixture order for speech signals.

C. Overall Proposed BSS Methods

The above extraction stage provides an estimate $e_l(n)$ of a source innovation process up to a delay and a scale factor, which we can then color to obtain each contribution of the l th source in the k th observation $x_k(n)$. This can be done by deriving the noncausal coloration filters $C_{kl}(z) = \sum_{r=-R}^R c_{kl}(r)z^{-r}$ that make the signals $c_{kl}(n) * e_l(n)$ be the closest to $x_k(n)$ in the mean-square sense [4]. This is here achieved by noncausal FIR Wiener filters [6], whose impulse response coefficients form vectors \mathbf{c}_{kl} defined by

$$\mathbf{c}_{kl} = \mathbf{R}_{e_l}^{-1} \cdot \mathbf{r}_{e_l x_k} \quad (19)$$

where \mathbf{R}_{e_l} is the autocorrelation matrix of the signal $e_l(n)$, and $\mathbf{r}_{e_l x_k}$ is the cross-correlation vector of the signals $e_l(n)$ and $x_k(n)$. Note that the autocorrelation matrix has a highly regular Toeplitz structure, and there are a number of efficient methods [6] for solving the linear matrix equation (19).

After subtracting the contributions $c_{kl}(n) * e_l(n)$ from all observations, we obtain another mixture configuration with $N-1$ sources. The first step must then be iterated as explained in Section I to extract the innovation process of another source.

III. EXPERIMENTAL RESULTS

In this section, we illustrate the performance of our methods on several examples. These algorithms are first tested for $P=2$ convolutive mixtures of $N=2$ artificial colored signals containing 100 000 samples. The innovation processes $u_j(n)$ have uniform distributions. For this type of signal, the kurtosis turned out to be the best optimization criterion, as compared to negentropy, and is used below. Fig. 1(a) shows the resulting output signal-to-interference ratios (SIRs), depending on the model order Q that we define as the sum of the orders of the mixture and innovation coloration filters $H_{ij}(z)$ and $F_j(z)$. For each extracted source, this SIR is averaged over the two estimated source contributions. For each value of Q , 100 Monte Carlo simulations were made by varying the mixture and innovation coloration filter coefficients with a uniform distribution,

and the resulting mean and standard deviations of the SIRs were computed. The order of R in (6) was set to $R = Q$, since our tests proved that this yields a good trade-off between performance (SIR) and computational cost. Similarly, the order of the noncausal filters $C_{kl}(z)$ used to color the estimated innovation processes was fixed by using $R' = 2Q$. Fig. 1(a) shows that for model orders Q up to 60, the means of SIRs are between 17 and 9 dB for the first source and a little lower for the second source, apart for linear instantaneous mixtures, which yield much higher performance.

In the second series of experiments, we tested our algorithm in the underdetermined case $N = 3$ and $P = 2$, when the two observed signals contain a stationary white Gaussian “noise source” in addition to the above two sources of interest. We fixed the model order to $Q = 20$. By varying the power of the noise source and thus the signal-to-noise ratio (SNR) in the observations, we investigated the robustness of the estimated mixing matrix to noise. To this end, we computed the SIRs of the estimated two sources of interest in which we canceled the contributions of the noise source. Fig. 1(b) shows that our method is fairly robust as SIRs remain higher than 10 dB for SNRs down to 8 dB.

For $N = P = 2$, we also compared the processing time of our method with a modified version of Tugnait’s algorithm introduced in [4], which already achieves higher speed than Tugnait’s approach by using a modified Newton algorithm. We varied the model order Q between 0 and 30 for sources containing $T = 10\,000$ samples ($T > 10\,000$ or $Q > 30$ lead to unacceptably high processing time for 100 runs of Tugnait–Newton’s approach). The results in Fig. 1(c) represent the time of the innovation extraction stage, which is by far the most time consuming. This shows that our method is about 100 times faster than Tugnait–Newton. In addition, it yields slightly higher SIRs, i.e., about 0.5 dB.

The next series of experiments was carried out with two English speech sources sampled at 20 kHz during 5 s. As in the first set of experiments, we varied the order of the mixing filters $H_{ij}(z)$, and we again performed 100 experiments for each filter order. This time, we chose to perform the innovation extraction stage on a 20 000-sample window where the signals are almost stationary and to estimate the SIRs on the overall signals. For these audio signals, the negentropic optimization criterion with $G(x) = e^{-(x^2/2)}$ turned out to yield better performance, probably because of their distributions tails [5]. The results in Fig. 1(d) show that the mean SIRs are between 11 and 7 dB, which is 4 dB lower than with the above 100 000-sample artificial stationary signals. This results from the 20 000-sample window, which yields slightly lower performance for artificial signals, and from the moving average (MA) process model (3), which is only approximately relevant for speech sources. Anyway, this yields significant perceptual quality improvement.

We also tested our negentropic algorithm with real 64th-order mixing filters measured at the ears of a dummy head [7] and with $R = 64$, $R' = 300$. We selected the impulse responses associated with source positions defined by 80- and 120-degree

angles in relation to the dummy head. Using again the above two speech signals, the mean output SIRs were 9.1 dB for the first source and 6.9 dB for the second source. These real filters therefore yield almost the same performance as the 64th-order artificial ones considered in Fig. 1(d).

APPENDIX

RELEVANCE OF CONSIDERED CRITERION

The P considered observations $x_i(n)$ are expressed with respect to the N innovation processes $u_j(n)$ according to (4). They are therefore causal Q th-order FIR mixtures of these processes. Now consider the vector $\tilde{\mathbf{x}}(n)$ defined in (6) and composed of delayed observations. The analysis provided in [8] implies that if

$$PL \geq N(Q + L) \quad (20)$$

where $L = (2R + 1)$ is the number of lags, then $\tilde{\mathbf{x}}(n)$ may also be interpreted as a set of linear instantaneous mixtures of corresponding sources, which are here delayed and scaled versions of the innovation processes $u_j(n)$, with at least as many observations as sources. Therefore, if (20) is met, the investigation for instantaneous mixtures provided in [3] proves rigorously that, by maximizing the absolute value of the nonnormalized kurtosis of the signal $y(n)$ defined in (9) under the constraint (10), we extract a delayed and scaled innovation process $\alpha_l u_l(n-r)$, whose practical estimate is denoted $e_l(n)$ hereafter.

If (20) is not met, the reformulated instantaneous BSS problem is underdetermined, i.e., it involves fewer observations than sources (note that this is especially the case when $P = N$). Some approximations are then necessary. However, when the ratio $PL/N(Q + L)$ associated to (20) tends to 1 (which is the case when $P = N$ and L is large), a delayed and scaled innovation process may still be accurately estimated as a linear combination of the available observations whose absolute nonnormalized kurtosis is maximum under the constraint (10).

REFERENCES

- [1] J. Tugnait, “Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria,” *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 658–672, Mar. 1997.
- [2] A. Hyvarinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Comput.*, vol. 9, pp. 1483–1492, 1997.
- [3] N. Delfosse and P. Loubaton, “Adaptive blind separation of independent sources: A deflation approach,” *Signal Process.*, vol. 45, pp. 59–83, 1995.
- [4] F. Abrard, “Méthodes de séparation aveugle de sources et applications,” Ph.D. dissertation, Univ. Toulouse, France, 2003.
- [5] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [6] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*. Chichester, U.K.: Wiley, 2000.
- [7] B. Gardner and K. Martin. Head Related Transfer Functions of a Dummy Head. [Online] <http://sound.media.mit.edu/ica-bench/>.
- [8] C. Févotte, “Approche temps-fréquence pour la séparation aveugle de sources nonstationnaires,” Ph.D. dissertation, Univ. Nantes, France, 2003.