# Spatio–Temporal FastICA Algorithms for the Blind Separation of Convolutive Mixtures

Scott C. Douglas, *Senior Member, IEEE*, Malay Gupta, *Member, IEEE*, Hiroshi Sawada, *Senior Member, IEEE*, and Shoji Makino, *Fellow, IEEE*

***Abstract*—This paper derives two spatio–temporal extensions of the well-known FastICA algorithm of Hyvärinen and Oja that are applicable to the convolutive blind source separation task. Our time–domain algorithms combine multichannel spatio–temporal prewhitening via multistage least-squares linear prediction with novel adaptive procedures that impose paraunitary constraints on the multichannel separation filter. The techniques converge quickly to a separation solution without any step size selection or divergence difficulties, and unlike other methods, ours do not require special coefficient initialization procedures to obtain good separation performance. They also allow for the efficient reconstruction of individual signals as observed in the sensor measurements directly from the system parameters for single-input multiple-output blind source separation tasks. An analysis of one of the adaptive constraint procedures shows its fast convergence to a paraunitary filter bank solution. Numerical evaluations of the proposed algorithms and comparisons with several existing convolutive blind source separation techniques indicate the excellent relative performance of the proposed methods.**

***Index Terms*—Blind source separation (BSS), fixed-point algorithm, independent component analysis, speech enhancement.**

## I. Introduction

Blind source separation (BSS) refers to two classes of multichannel signal processing tasks in which the goal is to extract multiple useful signals from multiple linear mixtures of these signals without specific knowledge of the source properties or the mixing characteristics. The first task, known as instantaneous BSS, assumes no multipath in the mixing system, and thus an instantaneous demixing process is adequate. Instantaneous BSS is often associated with independent component analysis (ICA), a class of methods for decomposing multichannel data based on information-theoretic criteria. The second task, known as convolutive BSS, assumes a general multipath channel, thus requiring multichannel filtering to achieve separation. The convolutive BSS task fits several important signal processing problems, such as multitalker speech separation from multimicrophone audio recordings and cochannel interference mitigation in wideband multi-input multi-output wireless communications systems. Although speech separation and enhancement is the focus of this paper, the algorithms developed here are applicable to other convolutive BSS tasks as well as to multichannel blind deconvolution tasks. Algorithmic solutions to the instantaneous BSS task are generally easier to develop than those for the convolutive BSS task due to the single-matrix parametrization of the former task. Numerous algorithms for instantaneous BSS have been developed; see the texts [1], [2] for examples.

Several researchers have explored convolutive extensions of instantaneous BSS procedures. The simplest of these extensions treat the separation task in the (discrete) Fourier domain and apply existing spatial-only complex-valued ICA and BSS methods within each frequency bin. To obtain the best performance, these methods must address the permutation, amplitude, and scaling inconsistencies across different frequency bins of the separation system at convergence to reconstruct the separated output signals. Early solutions to the above problems were proposed in [3] and [4]. Recent advances in permutation/scaling ambiguity resolution can be found in [5]–[9].

A potentially more elegant solution is to develop convolutive BSS algorithms using a time–domain separation criterion. An example of this approach is the information-theoretic natural gradient convolutive BSS and multichannel blind deconvolution algorithms developed in [10] and [11]. While this procedure can be successful, the source distributions must be approximately known, and the number of sources must be exactly known as they are simultaneously extracted. Also, due to the highly nonlinear and unconstrained nature of the coefficient updates, it can be challenging to select appropriate step sizes of the algorithm to obtain fast convergence in a wide range of data scenarios.

The FastICA algorithm of Hyvärinen and Oja [2], [12] is one of the most well-known and popular procedures for both ICA and instantaneous BSS. For an $m$-element linear non-Gaussian signal mixture, the procedure consists of a signal prewhitening stage followed by a set of $m$ fixed-point iterative procedures that extract independent components using a non-Gaussianity signal measure. Coefficient vector orthogonality is used to guarantee uniqueness of the extracted components. The algorithm enjoys a number of useful properties, including fast convergence, guaranteed global convergence for certain mixing conditions and contrasts, and robust behavior when noise is present. Recently, an extension of the FastICA algorithm to convolutive mixtures has been presented [13]. This particular extension does not exploit the convolutive nature of the mixing system, and the initialization scheme it uses is computationally complex. For

Manuscript received July 1, 2006; revised April 4, 2007. An early and shorter version of this paper was published in the Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005), Philadelphia, PA, May 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mari Ostendorf.

S. C. Douglas and M. Gupta are with the Department of Electrical Engineering, Southern Methodist University, Dallas, TX 75275 USA.

H. Sawada and S. Makino are with the NTT Communication Science Laboratories, NTT Corporation, Kyoto 619-0237, Japan.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2007.899176

an $m$-source separation task with $L'$ separation coefficients, a prewhitening matrix of size $(mL') \times (mL')$ must be computed, and $m$ systems of $(mL')$ linear equations must be solved using this prewhitening matrix in order to initialize the coefficients of the separation system. Unlike the symmetric coefficient orthogonality conditions that are used in the original FastICA procedure, the algorithm in [13] uses signal deflation, which leads to error accumulation in the separated outputs at each separation stage.

In this paper, we present two novel spatio–temporal extensions of Hyvärinen and Oja's FastICA algorithm to both convolutive BSS and multichannel blind deconvolution tasks. These algorithms combine multichannel whitening via multistage least-squares linear prediction with fixed-point iterations that use new adaptive techniques for imposing paraunitary constraints on the multichannel separation filter. A unique feature of our approaches are their ability to easily and individually reconstruct the sources as they appear in the observed signal mixtures, thereby generating $m$ output signals for each single source, a technique originally proposed in [14] and later renamed the single-input multiple-output (SIMO) BSS separation task [15]. Unlike other approaches, the proposed methods do not diverge and do not require special step sizes to be selected (e.g., natural gradient-based methods [10], [11] with fixed step-size parameters applied to signal mixtures of varying power levels) and do not require special initialization strategies or permutation solvers to function (e.g., frequency–domain methods described in [3]–[9]). Performance comparisons using data collected from real rooms show that our methods outperform other existing approaches when no knowledge of the sensor configuration and source positions are available, and their performance for three-source, three-microphone mixtures is significantly better than existing methods.

The organization of the paper is as follows. In Section II, we describe the convolutive BSS and reconstruction-focused SIMO BSS tasks and give a concise derivation of our proposed spatio–temporal extension of the FastICA algorithm. As our proposed convolutive BSS algorithms employ paraunitary filter constraints, two methods for adaptively imposing paraunitary constraints on a multichannel system structure are given, and their behaviors are elucidated in Section III. Section IV describes the prewhitening strategy used in our spatio–temporal FastICA algorithms and provides a block diagram for computing the SIMO BSS solution using our procedures. Section V gives the results of numerical experiments on multichannel room recordings comparing our proposed algorithms with several existing methods in the signal processsing literature. Section VI contains our conclusions.

## II. Problem Formulation and Algorithm Derivation

In this section, we describe the signal measurement model and statistical framework on which our separation methods are based and give a general form for our proposed spatio–temporal extensions of the FastICA algorithm.

Let $s_i(k), 1 \leq i \leq m$ denote $m$ spatially independent source signals, such that $s_i(k)$ is statistically independent of $s_j(l)$ for $i \neq j$. These sources are measured in an $n$-dimensional signal mixture with $n \geq m$ as

$$x_j(k) = \nu_j(k) + \sum_{i=1}^{m} \sum_{p=-\infty}^{\infty} a_{jip} s_i(k-p) \tag{1}$$

for $1 \leq j \leq n$, where $\{a_{ijp}\}$ are the coefficients of the multichannel mixing system, and $\nu_j(k)$ is uncorrelated Gaussian sensor noise. The goal of convolutive BSS is to compute a set of separated signals $\{y_i(k)\}$ for $1 \leq i \leq m$ as

$$y_i(k) = \sum_{j=1}^{n} \sum_{p=0}^{L'-1} b_{ijp} x_j(k-p) \tag{2}$$

where $L'$ is a filter length parameter, such that each $y_i(k)$ is nearly a filtered version of one $s_{i'}(k)$ for some unique set of assignments $i' \rightarrow i$, $1 \leq \{i', i\} \leq m$. If this is the case, all sources in $\{s_{i'}(k)\}$ are uniquely represented in the output signals $\{y_i(k)\}$ up to ordering and filtering ambiguities. The separation task then becomes one of adapting the $mnL'$ parameters $\{b_{ijp}\}$ to approximately achieve a separated result.

In the SIMO BSS task [14], [15], the goal is to extract estimates of the sources as they appear in the signal mixtures, ideally given by the $mn$ signal set

$$x_{ij}(k) = \sum_{p=-\infty}^{\infty} a_{jip} s_i(k-p) \tag{3}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. In practice, each $\widehat{x}_{ij}(k)$ is estimated from the separated signals $y_i(k)$ as

$$\widehat{x}_{ij}(k) = \sum_{p=-\infty}^{\infty} g_{jip} y_i(k-p) \tag{4}$$

where the coefficients $\{g_{jip}\}$ must be estimated or calculated from the separation system, the extracted signals $\{y_i(k)\}$, and/or the original input signal mixtures $\{x_j(k)\}$. Thus, both the convolutive BSS and SIMO BSS tasks require separated signals $\{y_i(k)\}$ to be estimated.

Criteria for adapting the parameters $\{b_{ijp}\}$ can be loosely classified into three types: those based on non-Gaussianity of the sources, those based on nonstationarity of the sources, and those based on the correlation properties of the sources [16]. As our approaches are extensions of the FastICA algorithm, they exploit the non-Gaussianity of the sources, a reasonable assumption for speech signals. Our approaches are based on the following assumption of the sources $s_i(k)$ themselves, an assumption that is also used in [17], [18] to develop criteria for convolutive BSS.

*Assumption:* Each source signal $s_i(k)$ is a linearly filtered version of an underlying zero-mean unit-variance non-Gaussian random process $\zeta_i(k)$ as

$$s_i(k) = \sum_{l=0}^{\tilde{L}-1} d_{il} \zeta_i(k-l) \tag{5}$$

where $\tilde{L}$ is the assumed model-order, and each sequence $\zeta_i(k)$ is statistically independent in time and space, and $\{d_{il}\}$ are the coefficients defining the correlation properties of the sources $s_i(k)$.

The above assumption allows one to effectively consider the convolutive BSS task as a multichannel blind deconvolution task, whereby the goal is to extract all of the innovation processes $\{\zeta_i(k)\}$ in $\{y_i(k)\}$ by enforcing their spatio–temporal independence properties. A solution to this task can be obtained by formulating an adaptive procedure in the combined coefficients $c_{ijl}$, where

$$c_{ijl} = \sum_{t=1}^{n} \sum_{p=0}^{L'-1} \sum_{q=-\infty}^{\infty} b_{itp} a_{tj(q-p)} d_{j(l-q)} \tag{6}$$

and

$$y_i(k) = \sum_{l=-\infty}^{\infty} \sum_{j=1}^{m} c_{ijl} \zeta_j(k-l). \tag{7}$$

In the sequel, the adaptive procedure we develop will be translated to an adaptive procedure on the observed system parameters as opposed to the combined system parameters $\{c_{ijl}\}$. The interaction of coefficients $\{a_{ijp}\}$ and $\{d_{il}\}$ is largely responsible for the initial conditions on $\{c_{ijl}\}$. With proper initialization, some algorithms tend to demonstrate a faster convergence.

We have the freedom to define additional constraints on the output signals $\{y_i(k)\}$ to make our adaptive algorithm design easier. It is advantageous to enforce the same second-order uncorrelated properties possessed by $\{\zeta_i(k)\}$ on $\{y_i(k)\}$, as then the combined system represented by $\{c_{ijl}\}$ satisfies the *paraunitary filter constraints*

$$\sum_{p=1}^{m} \sum_{q=-\infty}^{\infty} c_{ipq} c_{jp(l+q)} = \begin{cases} 1, & \text{if } i = j \text{ and } l = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Paraunitary filters are the multichannel extension of allpass filters and have numerous applications in coding and system modeling; for details, see [19]. The constraints in (8) are the spatio–temporal extension of orthonormality constraints on the rows (or columns) of a square matrix used by the FastICA procedure for instantaneous BSS.

Consider the goal of extracting a single independent source in $y_i(k)$ by adjusting the parameters $\{c_{ijl}\}$, $1 \leq j \leq m$, and $-\infty < l < \infty$ according to an appropriate non-Gaussianity measure or contrast. The time index of the source we extract in this formulation is unimportant, because we place no constraints on the overall delay of the system and because all samples of a given source $\{s_j(k)\}$ for all $k$ can be obtained by a filtering operation using the converged parameters. This problem is essentially what the single-unit FastICA procedure already solves in the instantaneous BSS case; the only difference is the doubly infinite nature of the filtering model which will have to be truncated to finite length for implementation purposes. By enforcing the $\{y_i(k)\}$ signals to each have unit variance and collectively be spatio–temporally uncorrelated; however, we can use the constraints in (8) to restrict the filter coefficients $\{c_{ijl}\}$ to the space of paraunitary filters, where paraunitariness can be imposed jointly across all filter channels or preferentially to certain filter channels according to a prescribed channel ordering. The use of paraunitary constraints in the prewhitened convolutive BSS problem is a key concept in our algorithm development.

To make paraunitary constraints practical, we need to translate the constraints to a set of filter coefficients that are easily adjusted. Suppose $\{h_{ijp}\}$ describes a multichannel prewhitening filter in the principal signal subspace, such that the $m$ signals

$$v_i(k) = \sum_{j=1}^{n} \sum_{p=-\infty}^{\infty} h_{ijp} x_j(k-p) \tag{9}$$

are uncorrelated in space and in time with unit variance. Define

$$\mathbf{v}(k) = [\mathbf{v}_1^T(k)\ \mathbf{v}_2^T(k)\ \cdots\ \mathbf{v}_m^T(k)]^T \tag{10}$$

$$\mathbf{v}_j(k) = [v_j(k)\ v_j(k-1)\ \cdots\ v_j(k-L+1)]^T. \tag{11}$$

Furthermore, define

$$\mathbf{w}_i = [\mathbf{w}_{i1}^T\ \mathbf{w}_{i2}^T\ \cdots\ \mathbf{w}_{im}^T]^T \tag{12}$$

$$\mathbf{w}_{ij} = [w_{ij0}\ w_{ij1}\ \cdots\ w_{ij(L-1)}]^T \tag{13}$$

as the separation system coefficient vector for the $i$th system output. Then, we compute the $i$th separated signal sequence as

$$y_i(k) = \sum_{j=1}^{m} \sum_{p=0}^{L-1} w_{ijp} v_j(k-p) = \mathbf{w}_i^T \mathbf{v}(k) \tag{14}$$

for $1 \leq k \leq N$, assuming a data record length of $N$ samples, where $\{w_{ijp}\}$ are the adjustable system parameters. Due to the statistical orthogonality of the prewhitened signals $\{v_j(k)\}$ in space and in time, we can constrain $\{w_{ijp}\}$ to be jointly paraunitary, i.e.,

$$\sum_{p=1}^{m} \sum_{q=-L}^{L} w_{ipq} w_{jp(l+q)} = \begin{cases} 1, & \text{if } i = j \text{ and } l = 0 \\ 0, & \text{for } 0 < |l| \leq Q \end{cases} \tag{15}$$

where $Q$ is an integer less than $L$ due to the finite length of the separation system, and $w_{ipq}$ is assumed to be zero outside of the range $0 \leq q \leq L-1$. Thus, our single-unit FastICA procedure in this scenario has the following structure.

Step 1) Compute $y_i(k)$ in (14) for $1 \leq k \leq N$.

Step 2) Update the coefficient vector as

$$\mathbf{w}_i \longleftarrow \frac{1}{N} \sum_{k=1}^{N} f(y_i(k))\mathbf{v}(k) - f'(y_i(k))\mathbf{w}_i \tag{16}$$

where $f(y)$ is the FastICA algorithm nonlinearity and $f'(y)$ is its derivative [2].

Step 3) Impose all or a subset of the paraunitary constraints defined by (15) on the $\{w_{ijl}\}$ coefficients within $\mathbf{w}_i$ depending on the type of extraction method (e.g., sequential or parallel).

Note that the choice of FastICA algorithm nonlinearity $f(y)$ in (16) is governed by the same rules and considerations as in the instantaneous BSS case. As has been observed, many choices of nonlinearity are possible, with the most popular choices being $f(y) = |y|^{q-1}y$ for $q \geq 1$ and $f(y) = \tanh(\alpha y)$, $\alpha > 0$.

## III. Adaptive Paraunitary Constraints

The spatio–temporal FastICA procedure outlined in the previous section requires a technique to impose paraunitary filter constraints on a multichannel linear filter. Due to the FastICA coefficient updates, we are motivated to develop procedures that impose these constraints on multichannel FIR filters. While a direct projection method could be developed, such a technique would likely involve a significant computational overhead if conventional methods ( e.g., Gram–Schmidt) were employed. In this section, we describe two simple adaptive procedures for imposing paraunitary filter constraints on the separation system coefficients $\{w_{ijl}\}$. Since these procedures are adaptive, several iterations of each procedure are needed after each coefficient update in (16) to maintain system paraunitariness.

### A. Sequential Orthogonalization

Our first procedure assumes a standard sequential implementation of the FastICA procedure, in which sources are extracted one by one from the signal mixtures. The procedure assumes that $\mathbf{w}_j$ for $1 \leq j \leq i-1$ have converged to separating solutions. The procedure is given as follows.

1) Normalize the length of the coefficient vector as

$$\mathbf{w}_i \longleftarrow \frac{\mathbf{w}_i}{\sqrt{\mathbf{w}_i^T \mathbf{w}_i}}. \tag{17}$$

2) While $\mathbf{w}_i$ is not paraunitary with $\mathbf{w}_1\ \mathbf{w}_2, \ldots, \mathbf{w}_{i-1}$,

$$\mathbf{w}_i \longleftarrow \frac{3}{2}\mathbf{w}_i - \frac{1}{2}\mathbf{g}(\mathbf{w}_i, \mathbf{w}_i) - \sum_{j=1}^{i-1} \mathbf{g}(\mathbf{w}_i, \mathbf{w}_j) \tag{18}$$

where

$$\mathbf{g}(\mathbf{w}_i, \mathbf{w}_j) = [\mathbf{g}_{ij1}^T\ \mathbf{g}_{ij2}^T\ \cdots\ \mathbf{g}_{ijm}^T]^T \tag{19}$$
$$\mathbf{g}_{ijk} = \mathbf{C}_{ij}\mathbf{w}_{jk} \tag{20}$$

and the $(p,q)$th element of $\mathbf{C}_{ij}$ is given by

$$[\mathbf{C}_{ij}]_{pq} = \begin{cases} \sum\limits_{k=1}^{m}\sum\limits_{l=0}^{L-1} w_{jkl} w_{ik(l+p-q)}, & \text{if } |p\text{-}q| < \frac{L\text{-}1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

To understand the above method, define the $(n \times 1)$-dimensional $i$th system vector polynomial as

$$\underline{\mathcal{W}}_i(z) = \sum_{l=0}^{L-1} [w_{i1l}\ w_{i2l}\ \cdots\ w_{iml}]^T z^{-l}. \tag{22}$$

Then, (18)–(21) can be expressed using polynomials as

$$\underline{\mathcal{W}}_i(z) \longleftarrow \frac{3}{2}\underline{\mathcal{W}}_i(z) - \Bigg[\frac{1}{2}\left[\underline{\mathcal{W}}_i^T(z^{-1})\underline{\mathcal{W}}_i(z)\right]_{-(L-1)/2}^{(L-1)/2}\underline{\mathcal{W}}_i(z) + \sum_{j=1}^{i-1}\left[\underline{\mathcal{W}}_j^T(z^{-1})\underline{\mathcal{W}}_i(z)\right]_{-(L-1)/2}^{(L-1)/2}\underline{\mathcal{W}}_j(z)\Bigg]_0^{L-1} \tag{23}$$

where $[\cdot]_J^K$ denotes truncating the polynomials of its argument to order $-J$ through $-K$. Extensive simulations of this iterative subprocedure indicate that (23) causes

$$\left[\underline{\mathcal{W}}_i^T(z^{-1})\underline{\mathcal{W}}_i(z)\right]_{-(L-1)/2}^{(L-1)/2} \longrightarrow 1 \tag{24}$$
$$\left[\underline{\mathcal{W}}_j^T(z^{-1})\underline{\mathcal{W}}_i(z)\right]_{-(L-1)/2}^{(L-1)/2} \longrightarrow 0 \quad \text{for } 1 \leq j < i. \tag{25}$$

The above constraints are a spatio–temporal extension of the orthonormality constraints imposed on $\mathbf{w}_i$ in the original FastICA procedure and imply that the separation system is paraunitary.

As further justification of the iterative procedure for enforcing paraunitary constraints, let $L \to \infty$, and define

$$\underline{\mathbf{w}}_i = \underline{\mathcal{W}}_i(z)\big|_{z=e^{j\omega}} \tag{26}$$
$$\underline{\mathbf{W}}_i = \left[\underline{\mathcal{W}}_1(z)\ \underline{\mathcal{W}}_2(z)\ \cdots\ \underline{\mathcal{W}}_{i-1}(z)\right]\big|_{z=e^{j\omega}}. \tag{27}$$

Then, (23) can be rewritten as

$$\underline{\mathbf{w}}_{i,\text{new}} = \frac{3}{2}\underline{\mathbf{w}}_i - \frac{1}{2}||\underline{\mathbf{w}}_i||^2\underline{\mathbf{w}}_i - \underline{\mathbf{W}}_i\underline{\mathbf{W}}_i^H\underline{\mathbf{w}}_i \tag{28}$$

where $\cdot^H$ denotes Hermitian transpose. Define the variables

$$a_{0i} = ||\underline{\mathbf{w}}_i||^2 - 1 \tag{29}$$
$$a_{1i} = ||\underline{\mathbf{W}}_i^H\underline{\mathbf{w}}_i||^2 \tag{30}$$

where $||\underline{\mathbf{w}}_i||^2 = \underline{\mathbf{w}}_i^H\underline{\mathbf{w}}_i$. The condition $|a_{0i}| = a_{1i} = 0$ implies that $[\underline{\mathcal{W}}_1(z)\cdots\underline{\mathcal{W}}_i(z)]$ form an $m$-dimensional paraunitary sequence if $[\underline{\mathcal{W}}_1(z)\cdots\underline{\mathcal{W}}_{i-1}(z)]$ is already paraunitary. Then, (28) implies that these state variables evolve as

$$a_{0i,\text{new}} = \frac{1}{4}a_{0i}^2(a_{0i}-3) + (a_{0i}-1)a_{1i} \tag{31}$$
$$a_{1i,\text{new}} = \frac{1}{4}a_{0i}^2 a_{1i}. \tag{32}$$

This pair of nonlinear coupled scalar equations can be easily simulated for different initial conditions, and such simulation studies show that $a_{0i}$ and $a_{1i}$ converge to zero for a wide range of initial value pairs. Empirically, we have observed convergence of this system if

$$\begin{matrix} -1 < a_{0i} < 2 \\ a_{1i} < a_{0i} + 1 \end{matrix} \quad \text{or} \quad \begin{matrix} 0 < ||\underline{\mathbf{w}}_i||^2 < 3 \\ ||\underline{\mathbf{W}}_i^H\underline{\mathbf{w}}_i||^2 < ||\underline{\mathbf{w}}_i||^2 \end{matrix} \tag{33}$$

which are typically satisfied in practice.

We now investigate the numerical performance of the iterative paraunitary constraint scheme in (18)–(21). For these evaluations, we have chosen $n = 10$ and $L = 51$. For each simulation run, $\mathbf{w}_i$ was initialized to a 510-element vector containing zero-mean uncorrelated Gaussian noise of variance $10^{-4}$ summed with a single nonzero unity-valued "center" tap at position $k = L(i-1) + (L+1)/2$. Shown in Fig. 1 for $1 \leq i \leq 5$ are the average evolutions of $E\{\widehat{a_{0i}^2}\}$ and $E\{\widehat{a_{1i}}\}$, computed from the elements of $\mathbf{C}_{ij}$ for $1 \leq j \leq i$, as averaged over 100 different simulation runs. As can be seen, convergence to a paraunitary condition given by $E\{\widehat{a_{0i}^2}\} \approx 0$ and $E\{\widehat{a_{1i}}\} \approx 0$ is fast, approaching the machine precision of MATLAB in about ten iterations. We observed similar convergence speeds of this procedure in practical speech separation experiments as well.

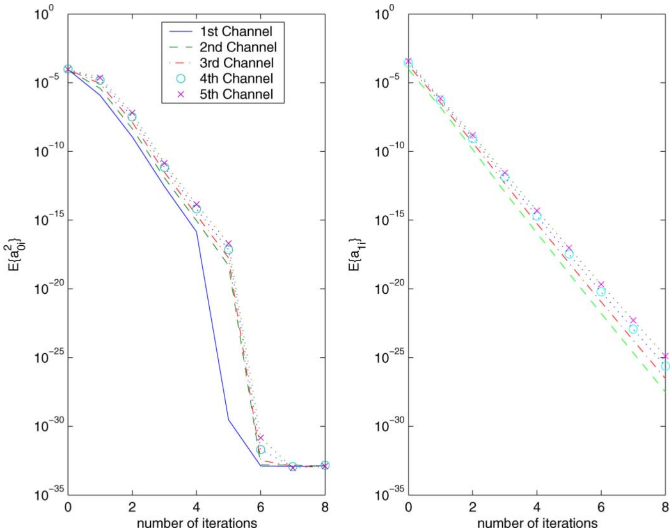



Fig. 1. Evolution of $E\{\widehat{a_{0i}^2}\}$ and $E\{\widehat{a_{1i}}\}$ for (18)–(21).

*B. Symmetric Orthogonalization*

The above adaptive subprocedure imposes paraunitariness in a sequential fashion to the system vectors $\{\mathbf{w}_i\}$. In other words, the value of $\mathbf{w}_1$ affects the solutions of $\mathbf{w}_i, 2 \leq i \leq m$, the value of $\mathbf{w}_2$ affects the solutions of $\mathbf{w}_i$, $3 \leq i \leq m$, and so on. This effect is similar to that observed in the original FastICA algorithm applied in a sequential fashion with Gram–Schmidt coefficient orthogonalization. While this type of constraint is robust in terms of source acquisition, it causes error accumulation in the separated system outputs, such that sources extracted "later" in the separation process have worse signal-to-interference ratios.

We now describe a symmetric method for imposing adaptive paraunitary constraints jointly on all $\mathbf{w}_i$, $1 \leq i \leq m$. Complete details regarding this procedure, including a proof of convergence and an analysis of its convergence speed, can be found in [20]. The procedure is given as follows.

1) Normalize all vectors $\{\mathbf{w}_i\}$, $1 \leq i \leq m$, as

$$\mathbf{w}_i \longleftarrow \frac{\mathbf{w}_i}{\sqrt{\mathbf{w}_i^T \mathbf{w}_i}}. \tag{34}$$

2) While the $\{\mathbf{w}_i\}$ are not jointly paraunitary, do for all $1 \leq i \leq m$

$$\mathbf{w}_i \longleftarrow \frac{3}{2}\mathbf{w}_i - \frac{1}{2}\sum_{j=1}^{m} \mathbf{g}(\mathbf{w}_i, \mathbf{w}_j) \tag{35}$$

where $g(\mathbf{w}_i, \mathbf{w}_j)$ is as defined in (19).

In order to better see the structure of this algorithm, define the $z$-transform matrix

$$\underline{\mathbf{W}}(z) = \sum_{l=0}^{L-1} \begin{bmatrix} w_{11l} & \cdots & w_{1ml} \\ \vdots & & \vdots \\ w_{m1l} & \cdots & w_{mml} \end{bmatrix} z^{-l}. \tag{36}$$

Then, this algorithm can be written as

$$\underline{\mathbf{W}}(z) \longleftarrow \frac{3}{2}\underline{\mathbf{W}}(z) - \frac{1}{2}\left[\left[\underline{\mathbf{W}}(z)\underline{\mathbf{W}}^T(z^{-1})\right]_{-(L-1)/2}^{(L-1)/2} \underline{\mathbf{W}}(z)\right]_0^{L-1}. \tag{37}$$

For $L = 1$, this procedure is a Newton-based adaptive orthogonalization scheme; see [21] for more details as well as a convergence analysis of the procedure. Typically, between 10 and 20 iterations of this procedure are needed at each FastICA update to obtain a system that is sufficiently close to paraunitariness for speech separation applications.

## IV. System Architecture and Prewhitening Method

We now describe the system architecture for implementing our spatio–temporal FastICA procedures. With careful design, it is possible to both achieve good separation quality and perform signal reconstruction efficiently for SIMO BSS tasks.

Previous work on SIMO BSS has yielded two strategies for finding $\{\widehat{x}_{ij}(k)\}$ in (4) after the separated sources $\{y_i(k)\}$ have been found. The first strategy uses traditional linear estimation to calculate the $\{g_{ijp}\}$ coefficients, where the $\{x_{ij}(k)\}$ are the desired signals, and the $\{y_j(k)\}$ are the reference signals [22]. This approach is complicated if the $\{y_i(k)\}$ are not uncorrelated in time, however, as it involves $m$ disjoint $n(M+1)$-dimensional estimation tasks. Moreover, it requires signal averaging between the $\{y_j(k)\}$ and the $\{x_i(k)\}$. The second calculates the inverse of the separation system for the $\{g_{ijp}\}$[14]. This procedure is challenging due to the difficulty of calculating a multichannel system inverse that does not exploit a specific system structure. These procedures generally require filter lengths $M$ that are longer than that of either the separation system or the original mixing channel.

We propose a different strategy. Consider Fig. 2, which shows a signal processing architecture containing a prewhitening stage, a separation stage, and a signal reconstruction stage.

The goal of the prewhitening stage is to decorrelate the original signal mixtures in both space and time. We propose a multistep prewhitening structure using pairs of multichannel linear systems with transfer function matrices given by

$$\mathbf{P}_i(z) = \mathbf{D}_i^{(P)} \begin{bmatrix} P_{11i}(z) & P_{12i}(z) & \cdots & P_{1mi}(z) \\ 0 & P_{22i}(z) & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & P_{mmi}(z) \end{bmatrix} \tag{38}$$

$$\mathbf{Q}_i(z) = \mathbf{D}_i^{(Q)} \begin{bmatrix} Q_{11i}(z) & 0 & \cdots & 0 \\ Q_{21i}(z) & Q_{22i}(z) & & \vdots \\ \vdots & & \ddots & 0 \\ Q_{m1i}(z) & \cdots & & Q_{mmi}(z) \end{bmatrix} \tag{39}$$

where the causal $\{\mathbf{P}_i(z)\}$ and $\{\mathbf{Q}_i(z)\}$ are multichannel FIR filters of length $K$, the causal filters $\{P_{jji}(z)\}$ and $\{Q_{jji}(z)\}$ have unity zero-lag coefficient values, and $\mathbf{D}_i^{(P)}$ and $\mathbf{D}_i^{(Q)}$ are $(m \times m)$ diagonal scaling matrices. The coefficients for the $j$th row of the $\{\mathbf{P}_i(z)\}$ transfer function matrices are calculated by solving a least-squares multichannel forward linear prediction task, e.g., by minimizing the output power of the $j$th output signal. The diagonal entries of $\mathbf{D}_i^{(P)}$ are then calculated so that the scaled forward error residuals have unity variances. These scaled error residuals are used as inputs to the $\{\mathbf{Q}_i(z)\}$ multichannel system, in which the $j$th row of the $\{\mathbf{Q}_i(z)\}$ transfer function matrices are calculated by solving a second least-squares multichannel

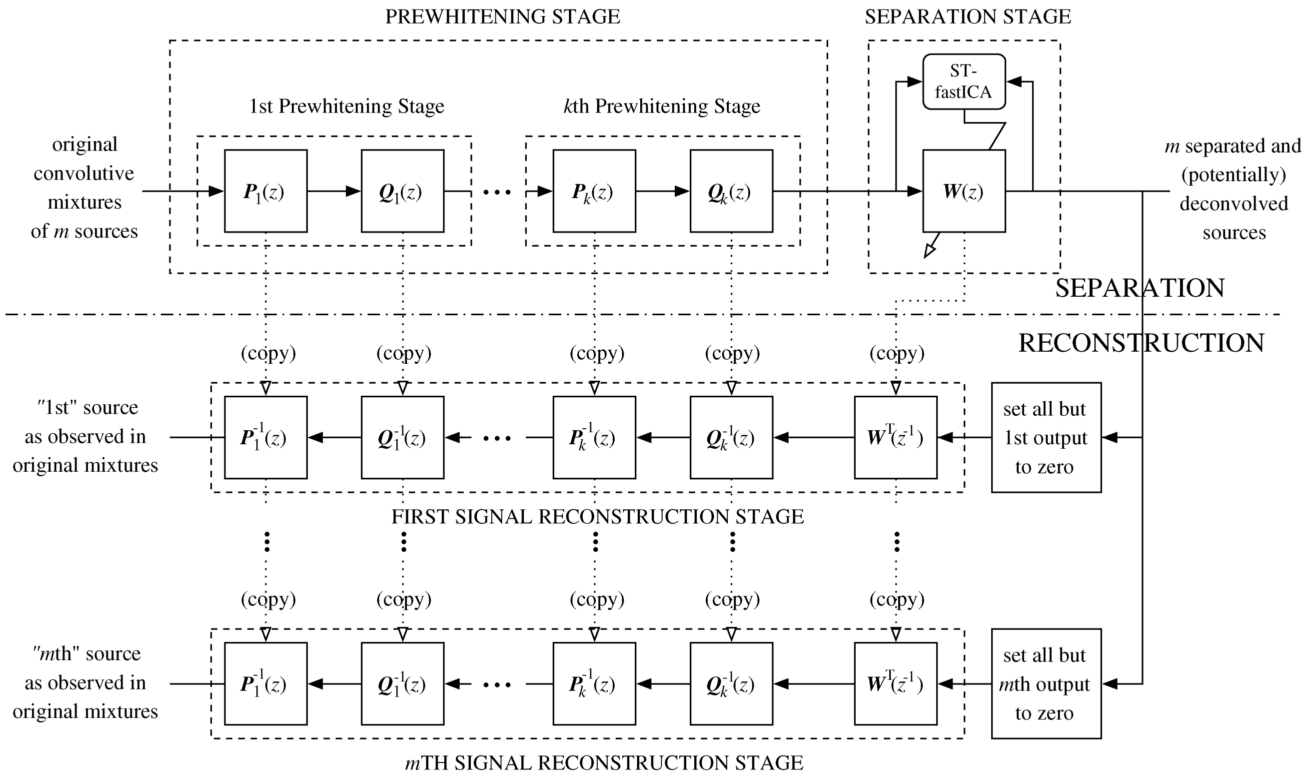



Fig. 2. Block diagram of the combined separation and signal reconstruction system.

forward linear prediction task. The diagonal entries of $\mathbf{D}_i^{(Q)}$ are subsequently calculated so that these error residuals have unity variances. Note that this proposed method is a block-based procedure.

Several stages of this processing strategy are usually required because the estimation of $\mathbf{P}_i(z)$ and $\mathbf{Q}_i(z)$ is performed in a disjoint and sequential fashion. The exact number of stages $k$ can be made adaptive, with a stopping criterion that depends on how much $\mathbf{P}_k(z)$ and $\mathbf{Q}_k(z)$ differ from identity.

The above prewhitening strategy has an important advantage: both $\mathbf{P}_i(z)$ and $\mathbf{Q}_i(z)$ can be easily inverted without calculating any new filter coefficients, using the linear system equivalent of backsubstitution. Thus, so long as the linear system within the separation stage can be easily inverted, creating the inverse of the entire prewhitening-separation system is straightforward.

The goal of the second stage is to perform separation of the prewhitened signal mixtures based on non-Gaussianity. One can use a sequential extraction strategy, where the sources are extracted one by one using the adaptive paraunitary constraint procedure in (18)–(21). This first algorithm is called the *spatio–temporal FastICA 1* (STFICA1) algorithm in the Simulations section. Alternatively, one can use a parallel extraction strategy, where $m$ separation units are updated at each iteration, and the adaptive paraunitary constraint procedure in (34)–(35) is used. This algorithm is called the *spatio–temporal FastICA 2* (STFICA2) algorithm in the Simulations section. Table I provides MATLAB code for implementing the STFICA1 algorithm complete with appropriately chosen stopping criteria, where `v` is the $(m \times N)$ prewhitened signal matrix, `L` is the separation system filter length, and `numiter` is the maximum number of iterations of the FastICA routine.

The goal of the last set of parallel stages, shown at the bottom of Fig. 1, is to reconstruct the individual sources as they appear in the original mixtures. The reconstruction of the $i$th separated signal involves setting all but the $i$th signal $y_i(k)$ to zero and then passing this signal through the inverse of the prewhitening and separation systems. For this calculation, note that

$$\begin{aligned}[\mathbf{W}(z)\mathbf{Q}_k(z)\mathbf{P}_k(z)\cdots\mathbf{Q}_1(z)\mathbf{P}_1(z)]^{-1} \\ = \mathbf{P}_1^{-1}(z)\mathbf{Q}_1^{-1}(z)\cdots\mathbf{P}_k^{-1}(z)\mathbf{Q}_k^{-1}(z)\mathbf{W}^T(z^{-1}) \quad (40)\end{aligned}$$

due to the paraunitariness of $\mathbf{W}(z)$ as constructed by the separation stage. Because of the triangular structures of the $\mathbf{P}_i(z)$ and $\mathbf{Q}_i(z)$ systems, they can be easily inverted.

The method we have described has a number of advantages over competing approaches.

1) No step size needs to be selected.
2) Knowledge of the source distributions is not needed, so long as their statistics imply a nonzero contrast value.
3) For the STFICA1 algorithm, the number of non-Gaussian sources within the mixture need not be known *a priori*.
4) For SIMO BSS, the system inverse used for signal reconstruction is computed directly and exactly.
5) Convergence of the STFICA1 and STFICA2 procedures appears to be as fast as its spatial-only counterparts. The single-unit STFICA1 procedure usually requires fewer than ten iterations per unit when the sources are i.i.d.

## V. Numerical Evaluations

We now present numerical evaluations comparing the performance of the proposed convolutive BSS methods with several existing techniques in the signal processing literature.

TABLE I
MATLAB IMPLEMENTATION OF THE STFICA1 ALGORITHM UPDATE

```
-------------------------------------------------------------+------------------------------------------
function [y,W] = stfica1(v,L,numiter);                       | function [Wi] = orthW(W,m,L,numorth);
                                                             |
L = L + rem((L+1),2);                                        | i = size(W,2);
[m,N] = size(v);                                             | Wi = W(:,i)/norm(W(:,i));
W = kron(eye(n),[zeros((L-1)/2,1);1;zeros((L-1)/2,1)]);      | for k=1:numorth
V = zeros(m*L,N);                                            |   Wt = zeros(m*L,1);
for i=1:m                                                    |   for j=1:i-1
  V((i-1)*L+1:i*L,:) = toeplitz([v(i,1);zeros(L-1,1)],v(i,:)); |     Wt = Wt + gfun(Wi,W(:,j),m,L);
end                                                          |   end
y = zeros(m,N);                                              |   Wi = 3/2*Wi - 1/2*gfun(Wi,Wi,m,L) - Wt;
for i=1:m                                                    | end
  Wold = zeros(m*L,1);                                       +------------------------------------------
  k = 0;                                                     | function [G,C] = gfun(U,V,m,L);
  y(i,:) = W(:,i)'*V;                                        |
  crit = 1;                                                  | Wi = zeros(L,m);    Wi(:) = U;
  while (crit*(k<numiter))                                   | Wj = zeros(L,m);    Wj(:) = V;
    Wold = W(:,i);                                           | Ct = zeros((3*L-1)/2,1);
    k = k+1;                                                 | Z = zeros((L-1)/2,1);
    f = y(i,:)'.^3;          % f = tanh(20*y(i,:)');         | ll = (L+1)/2:(3*L-1)/2;    llr = L:-1:1;
    fp = 3*sum(y(i,:).^2);  % fp = 20*sum(sech(20*y(i,:)).^2); | for i=1:m
    W(:,i) = V*f - fp*W(:,i);                                |   Ct = Ct + filter(Wi(llr,i),1,[Wj(:,i);Z]);
    W(:,i) = orthW(W(:,1:i),m,L,10);                         | end
    y(i,:) = W(:,i)'*V;                                      | C = Ct(ll);
    crit = (abs(abs(W(:,i)'*Wold)-1)>0.0001);                | Gt = filter(C(llr),1,[Wj;zeros((L-1)/2,m)]);
  end                                                        | Gt = Gt(ll,:);
end                                                          | G = Gt(:);
-------------------------------------------------------------+------------------------------------------
```

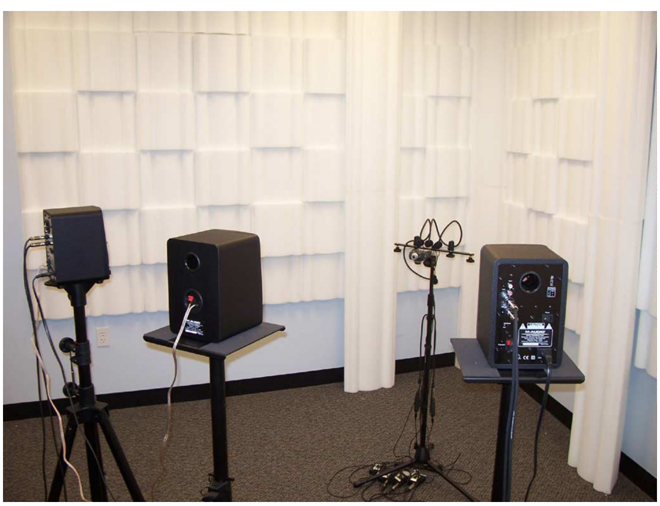


Fig. 3. Laboratory measurement environment used for numerical evaluations.

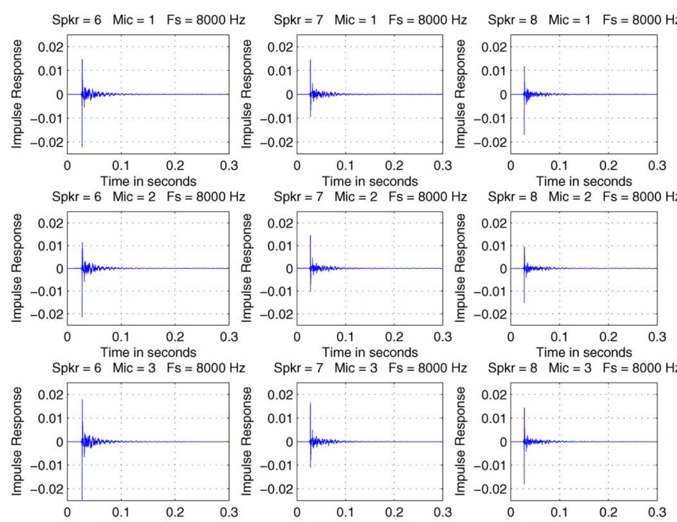



Fig. 4. Impulse responses measured from the uniform linear array, $RT = 300$ ms condition.

Data for these evaluations was generated in an acoustically isolated laboratory environment with up to three loudspeakers playing recordings of talkers (one female and two male) as the sources. The sources were located 127 cm away from three omnidirectional microphones and were spaced at angles of $-30°$, $0°$, and $27.5°$ from the angle of incidence of the microphone array. Two arrangements of microphones were used: 1) a nearly uniform linear array with 4-cm spacing and 2) an equilateral triangular array arranged in a vertical plane with 8-cm spacing. Two room reverberation conditions were created in the room corresponding to reverberation times of 300 and 425 ms, respectively. Fig. 3 shows a photograph of the laboratory setup for the 300-ms reverberation experiment with three sources and a uniform linear array. All measurements were made using 7 s of data per channel and a 48-kHz sampling rate and were downsampled to an 8-kHz sampling rate for processing. Figs. 4 and 5 show the impulse responses of the loudspeaker/microphone paths for reverberation times of 300 and 425 ms, respectively, as calculated using pseudorandom noise sequences. For purposes of correlating these plots with the photograph in Fig. 3, the microphones are labeled as 1, 2, and 3 from right to left in the photograph, and the loudspeakers are labeled as 6, 7, and 8 from right to left in the photograph with directions of arrival of $27.5°$, $0°$, and $-30°$, respectively.

The two proposed STFICA algorithms were run on this data with one stage (for the $RT = 300$ ms conditions) or two stages

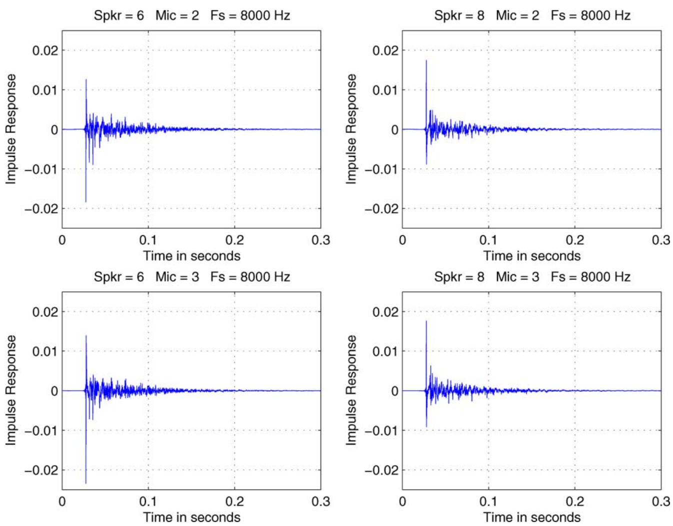



Fig. 5. Impulse responses measured from the uniform linear array, $RT = 425$ ms condition.

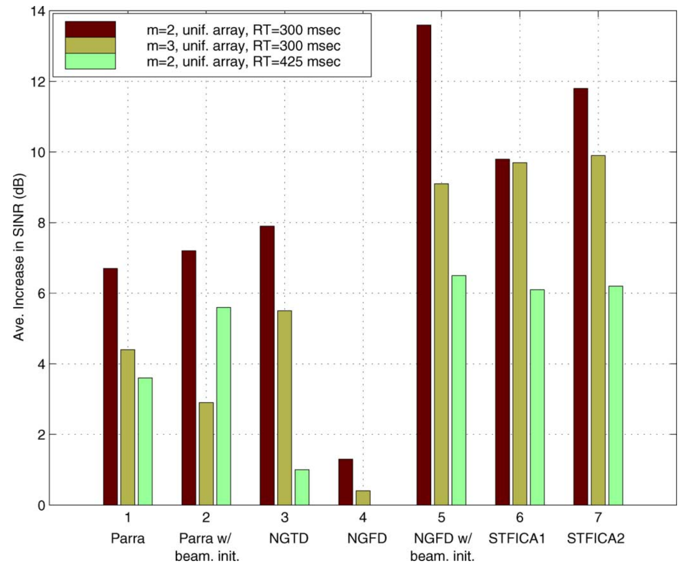



Fig. 6. Performance of various separation methods under two reverberation conditions ($RT = 300$ ms and 425 ms) for a uniform linear microphone array, $m = 2$ and $m = 3$ source cases.

(for the $RT = 425$ ms conditions) of least squares prewhitening using both $\mathbf{P}_i(z)\}$ and $\{\mathbf{Q}_i(z)\}$ filters with lengths of $M = 400$ taps per filter. Each separation system used $L = 300$ taps per input–output filter channel. In addition to these two STFICA algorithms, five additional algorithms were evaluated using this data:

- Parra's decorrelation-based method with the following choices: $\{$number of diagonalized matrices$\} = 5$, $\{$number of data blocks averaged$\} = 10$, $\{$FFT size $= 1024\}$, $L' = 400$-tap time-domain filters per input-output channel, and a 1000-iteration limit [6];
- Parra's decorrelation-based method with beamforming initialization using the above parameter settings [7],
- bin-wise natural gradient frequency-domain (NGFD) method with "center-spike" initialization, $\mu = 0.09$, $L' = 1024$, with 100 iterations;
- bin-wise natural gradient frequency-domain (NGFD) method with beamforming initialization using the above parameter settings;
- Natural gradient time domain (NGTD) method using causal FIR filters and "center-spike" initialization, a step size schedule of $\mu = 0.0005$ for 200 data passes followed by $\mu = 0.0001$ for a single data pass followed by $\mu = 0.00001$ for a single data pass, $L' = 512$ [11].

These algorithms were selected due to their widespread use as tools within modified approaches [5], [8], [9]. After separation is performed, least-squares methods are used to estimate the contributions of the source recordings to each of the recorded mixtures as well as the separated results from each algorithm. By calculating power ratios, we compute the average improvement in signal-to-interference-plus-noise ratio (SINR) for each algorithm in each data case.

Fig. 6 shows the performance of the various algorithms on data collected using two- and three-microphone uniform arrays in terms of average improvement in SINR. We consider each data case separately.

*Case 1: $m = 2$, Uniform Linear Array, $RT = 300$ ms:* The best performing algorithm in this case is NGFD algorithm with beamforming initialization, in which the average SINR improvement is 13.8 dB across the two outputs. The STFICA1 and STFICA2 algorithms achieve a 11.8- and 9.8-dB average SINR improvement. These latter two algorithms, however, do not require knowledge of the directions of arrival of the sources for their initialization, whereas the NGFD algorithm without this information only achieves an average improvement of 1.3 dB.

*Case 2: $m = 3$, Uniform Linear Array, $RT = 300$ ms:* In the three-source cases, the proposed STFICA algorithms perform the best, achieving 9.7 and 9.9 dB of average SINR improvement. The best-performing algorithms in the comparison employed geometric knowledge of the source and sensor positions and obtained a 9.1-dB average SINR improvement in the best case (NGFD).

*Case 3: $m = 2$, Uniform Linear Array, $RT = 425$ ms:* In this more-reverberant environment, the best performing algorithm was the NGFD algorithm with beamforming knowledge, in which a 6.5-dB improvement in SINR was obtained. The proposed STFICA algorithms gave nearly the same performace at 6.1 and 6.2 dB, respectively, and the latter methods do not require knowledge of sensor positions to work well.

In general, the best versions of the methods based on non-Gaussian statistics generally out-perform the methods based on second-order decorrelation. Also, the STFICA2 algorithm using symmetric orthogonalization usually outperforms the STFICA1 algorithm using sequential orthogonalization. It should be noted that both STFICA procedures were easy to evaluate in these experiments as they need only to be run once. The other procedures are gradient-based methods that require significant hand-tuning of step size parameters, careful selection of the number of data passes, and a careful initialization strategy in order to obtain good separation performance and to avoid divergence.

Fig. 7 shows the performance of three algorithms in a three-source mixture case for a triangular microphone array. In this

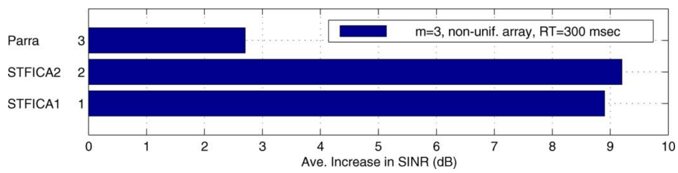



Fig. 7. Performance of various separation methods under an $RT = 300$ ms reverberation condition for a nonuniform linear microphone array, $m = 3$ source case.

case, direction of arrivals are not as easily calculated, so only algorithms with generic (e.g., "center-spike") initializations have been evaluated. As can be seen, the proposed STFICA algorithms perform well in this situation despite having an unknown array geometry, unknown source positions, and generic initialization strategies.

The run-time of the proposed algorithms was found at par with the other algorithms. In the two source scenario, the sequential version of the algorithm took 1.85 s for a single iteration of the algorithm per source, where as the symmetric version of the algorithm took 2.75 s on a Pentium 4 (3.6-GHz) processor with 2 GB of RAM. For the three-source case, the sequential version of the algorithm took 4.00 s, where as the symmetric version of the algorithm took 6.50 s. These times also included the time taken for two-stages of least squares prewhitening of the input data.

## VI. Conclusion

This paper presents two spatio–temporal extensions of the well-known FastICA procedure that are useful for convolutive blind source separation tasks such as speech separation of multichannel microphone recordings in room environments. Our methods employ least-squares prewhitening along with novel iterative schemes for maintaining paraunitary constraints on the separation system. The procedures can be used to reconstruct the contributions of each of the sources in each of the sensor signals without estimating or calculating additional impulse responses or signal properties. When applied to multichannel recordings of two- and three-source speech mixtures, the algorithms are found to perform well and in a robust manner as compared to other existing approaches to convolutive BSS tasks. A chief advantage of the proposed methods is their simple setup; the algorithms do not require significant parameter tuning in order to obtain good separation performance.

## References

[1] A. Cichocki and S.-I. Amari*, Adaptive Blind Signal and Image Processing*. New York: Wiley, 2002.

[2] A. Hyvärinen, J. Karhunen, and E. Oja*, Independent Component Analysis*. New York: Wiley, 2001.

[3] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, Nov. 1998.

[4] S. Ikeda and N. Murata, "A method of blind separation based on temporal structure of signals," in *Proc. 5th Int. Conf. Neural Inf. Process*, Kitakyushu, Japan, Oct. 1998, pp. 737–742.

[5] N. Mitianoudis and M. E. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 489–497, Sep. 2003.

[6] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.

[7] L. Parra and C. Alvino, "Geometric source separation: Merging convolutve source separation with geometric beamforming," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 6, pp. 352–362, Sep. 2002.

[8] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[9] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-converging algorithm combining ICA and beamforming," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 2, pp. 666–678, Mar. 2006.

[10] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. 1st IEEE Workshop Signal Process. Adv. Wireless Commun.*, Paris, France, Apr. 1997, pp. 101–104.

[11] S. C. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," *IEEE Trans. Speech Audio Process.*, vol. 13, pp. 92–104, Jan. 2005.

[12] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computat.*, vol. 9, no. 7, pp. 1483–1492, Oct. 1997.

[13] J. Thomas, Y. Deville, and S. Hosseini, "Time domain fast fixed-point algorithms for convolutive ICA," *IEEE Signal Process. Lett.*, vol. 13, no. 4, pp. 228–230, Apr. 2006.

[14] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. Int. Workshop Ind. Compon. Anal. Blind Signal Separation*, Aussois, France, Jan. 1999, pp. 365–371.

[15] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, "SIMO-model-based independent component analysis for high-fidelity blind separation of acoustic signals," in *Proc. 4th Int. Symp. Ind. Compon. Anal. Signal Separation*, Nara, Japan, Apr. 2003, pp. 993–998.

[16] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," in *Conf. Rec. IEEE Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Kyoto, Japan, Sep. 2003, pp. 275–278.

[17] P. Comon, E. Moreau, and L. Rota, "Blind separation of convolutive mixtures: A contrast-based joint diagonalization approach," in *Proc. 3rd. Int. Conf. Ind. Compon. Anal. Blind Signal Separation*, San Diego, CA, Dec. 2001, pp. 686–691.

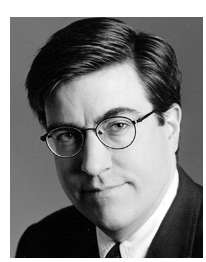
[18] D.-T. Pham, "Contrast functions for blind separation and deconvolution of sources," in *Proc. 3rd. Int. Conf. Ind. Compon. Anal. Blind Signal Separation*, San Diego, CA, Dec. 2001, pp. 37–42.

[19] P. P. Vaidyanathan*, Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

[20] S. C. Douglas, "An adaptive constraint method for paraunitary filter banks with applications to spatio–temporal subspace tracking," *EURASIP J. Appl. Signal Process.*, vol. 2007, 2007, article ID 80301.

[21] S. C. Douglas, "The singular value manifold and numerical stabilization of algorithms with orthogonality constraints," in *Proc. IEEE Workshop Sensor Array Multichannel Processing*, Boston, MA, Jul. 2006.

[22] Y. Inouye and K. Tanabe, "Super-exponential algorithms for multichannel blind deconvolution," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 881–888, Mar. 2000.



**Scott C. Douglas** (SM'88–M'92–SM'98) received the B.S. (with distinction), M.S., and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1988, 1989, and 1992, respectively.

He is currently an Associate Professor in the Department of Electrical Engineering, Southern Methodist University, Dallas, TX. His research activities include source separation, speech enhancement, adaptive computational imaging devices, and hardware implementation of DSP systems. He is the author or coauthor of two books, six book chapters, and more than 150 articles in journals and conference proceedings. He frequently consults with industry and offers short courses and tutorials in the areas of signal processing and source separation. He has served as an editor or coeditor of several books and conference proceedings, including the *Digital Signal Processing Handbook*, the *International Symposium on Active Control of Sound and Vibration*, and the *IEEE Workshops on Machine Learning for Signal Processing* conference records.

Dr. Douglas is the recipient of the 2002 IEEE Signal Processing Society Best Paper Award in Audio and Electroacoustics and the 2003 Gerald J. Ford Research Fellowship. He has served on several technical committees of the IEEE Signal Processing Society and is a Past Chair of the Neural Networks for Signal Processing Technical Committee. He has helped organize several meetings of the IEEE, including the ICASSP and ISCAS conference series, the *Digital Signal Processing and Signal Processing Education Workshop* series, and the *Machine Learning for Signal Processing Workshop* series. He is the General Chair of the International Conference on Acoustics, Speech, and Signal Processing 2010 in Dallas. He is also one of the key authors and developers of curriculum and technology for The Infinity Project, a joint effort between educational, civic, and government organizations to implement engineering curricula at the precollege level. He is a member of Phi Beta Kappa and Tau Beta Pi.

**Malay Gupta** (M'03) was born in Kanpur, India. He received the B.E. degree (with distinction) in electrical engineering from the University of Roorkee (now Indian Institute of Technology), Roorkee, India, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of New Mexico, Albuquerque, in 2002 and 2006, respectively.

In December 2005, he joined Southern Methodist University, Dallas, TX, where he is currently working as a Postdoctoral Researcher. From August 1997 to June 2000, he was with the control and automation division of Larsen and Toubro, Ltd., Mumbai, India. His research interests include multiuser communication systems, blind source separation, independent component analysis, and speech enhancement.

**Hiroshi Sawada** (M'02–SM'04) received the B.E., M.E., and Ph.D. degrees in information science from Kyoto University, Kyoto, Japan, in 1991, 1993 and 2001, respectively.

He joined NTT Corporation, Kyoto, in 1993. He is now a Senior Research Scientist at the NTT Communication Science Laboratories. From 1993 to 2000, he was engaged in research on the computer-aided design of digital systems, logic synthesis, and computer architecture. In 2000, he was with the Computation Structures Group, Massachusetts Institute of technology, Cambridge, for six months. From 2002 to 2005, he lectured on computer architecture at Doshisha University, Kyoto. Since 2000, he has been engaged in research on signal processing, microphone array, and blind source separation (BSS). More specifically, he is working on the frequency–domain BSS for acoustic convolutive mixtures using independent component analysis (ICA). He is the author or coauthor of three book chapters, more than 20 journal articles, and more than 80 conference papers.

Dr. Sawada received the Ninth TELECOM System Technology Award for Student from the Telecommunications Advancement Foundation in 1994 and the Best Paper Award from the IEEE Circuits and Systems Society in 2000. He is an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and a member of the Audio and Electroacoustics Technical Committee of the IEEE SP Society. He was a tutorial speaker at ICASSP 2007. He serves as the publications chairs of the WASPAA 2007 at Mohonk Mountain House, New Paltz, NY, and served as an organizing committee member for ICA 2003 in Nara and the communications chair for IWAENC 2003 in Kyoto. He is a member of the IEICE and the ASJ.

**Shoji Makino** (A'89–M'90–SM'99–F'04) received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively.

He is now an Executive Manager at the NTT Communication Science Laboratories, Kyoto, Japan. He is also a Guest Professor at the Hokkaido University, Sapporo, Japan. His research interests include adaptive filtering technologies and realization of acoustic echo cancellation and blind source separation of convolutive mixtures of speech.

Dr. Makino received the ICA Unsupervised Learning Pioneer Award in 2006. He is the author or coauthor of more than 200 articles in journals and conference proceedings and is responsible for more than 150 patents. He is a Tutorial speaker at ICASSP2007. He is a member of both the Awards Board and the Conference Board of the IEEE Signal Processing Society. He was an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING. He is a member of the Technical Committee on Audio and Electroacoustics of the IEEE Signal Processing Society and the Chair-Elect of the Technical Committee on Blind Signal Processing of the IEEE Circuits and Systems Society. He is a member of the International IWAENC Standing committee and a member of the International ICA Steering Committee. He is the General Chair of the WASPAA2007 at Mohonk, New Paltz, NY, the General Chair of the IWAENC2003 in Kyoto, and the Organizing Chair of the ICA2003, Nara, Japan.