Electronic Research Archive of Blekinge Institute of Technology
<http://www.bth.se/fou/>

This is an author produced version of a paper published in IEEE Transactions on Audio, Speech, and Language Processing. This paper has been peer-reviewed but may not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:
Sällberg, Benny; Grbic, Nedelko; Claesson, Ingvar
Complex-Valued Independent Component Analysis for Online Blind
Speech Extraction
IEEE Transactions on Audio, Speech, and Language Processing
2008, Vol. 16(8) pp. 1624-1632

DOI: 10.1109/TASL.2008.2002058

Access to the published version may
require subscription.

Published with permission from:
IEEE

Complex-Valued Independent Component Analysis for Online Blind Speech Extraction

Benny Sällberg, *Member, IEEE*, Nedelko Grbić, *Member, IEEE*, and Ingvar Claesson, *Member, IEEE*

Abstract—This paper presents a theoretical analysis of a certain criterion for complex-valued independent component analysis (ICA) with a focus on blind speech extraction (BSE) of a spatio-temporally nonstationary speech source. In the paper, the proposed criteria denoted KSICA is related to the well-known FastICA method with the Kurtosis contrast function. The proposed method is shown to share the important fixed-point feature with the FastICA method, although an improvement with the proposed method is that it does not exhibit the divergent behavior for a mixture of Gaussian-only sources that the FastICA method tends to do, and it shows better performance in online implementations. Compared to the FastICA, the KSICA method provides a 10 dB higher source extraction performance and a 10 dB lower standard deviation in a data batch approach when the data batch size is less than 100 samples. For larger batch sizes, the KSICA method performs equally well. In an online application with spatially stationary sources the KSICA method provides around 10 dB higher interference suppression, and 1 MOS-unit lower speech distortion compared to the FastICA for 0.15 s time constant in the algorithm update parameter. Thus, the FastICA performance matches the KSICA performance for a time constant above 1 s. Finally, in an online application with a moving speech source, the KSICA method provides 10 dB higher interference suppression, compared to the FastICA for the same algorithm settings. All in all, the proposed KSICA method is shown to be a viable alternative for online BSE of complex-valued signal mixtures.

Index Terms—Array signal processing, higher order statistics, speech enhancement.

I. INTRODUCTION

BLIND extraction of signals is a reoccurring problem in a variety of signal processing applications including speech enhancement, extraction of biomedical signals, etc. The notation “blind” implies that the signal extraction is relying only on certain assumptions regarding the statistical properties of the source signals, such as an assumption regarding the independence of the source signals. Some of these blind approaches can be sorted under the field of independent component analysis (ICA) [1], [2]. This paper focuses on a new method for performing blind speech extraction (BSE) using an ICA approach. This approach is explored through a real-time speech enhancement application. It should be noted that in BSE, it is desirable to extract a dominant speech source (or a group of dominant sources) from an observed mixture of many sources [3]–[7].

Manuscript received December 09, 2007; revised May 16, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Susanto Rahardja.

The authors are with the Department of Signal Processing, Blekinge Institute of Technology, Ronneby SE-371 79, Sweden (e-mail: bsa@bth.se; ngr@bth.se; icl@bth.se).

Digital Object Identifier 10.1109/TASL.2008.2002058

To extract speech, a convolution model is adopted where a set of source signals are emitted in a room and received by an array of microphones. Such a convolution model in the time domain corresponds to a multiplicative model in the frequency domain [8], [9]. It is in many cases desirable to perform the ICA in the frequency (subband) domain [10]–[12], which in general yields a faster convergence rate and a lower computational load as opposed to a corresponding full-band time domain method. The BSE discussed in this paper was furthermore performed on complex-valued data generated with the help of a specific Fourier-transform-based time–frequency subband transformation. One popular method used to perform BSE is the FastICA method [2], [13]. The FastICA method has been reported to be a fast and efficient method for blind extraction of signals. Bingham *et al.* [13] derived the FastICA method for complex-valued signal compositions with a focus on sources having circular distributions in order to simplify their derivations. In addition to this, Douglas [14] presented an alternative version of the FastICA method for separation of complex-valued signal mixtures, without relying upon circularity assumptions of the source signals.

Essentially, the FastICA method is intended for a batch processing approach where signal statistics are estimated for a certain period of time, described as the data batch duration, whereafter the ICA problem is solved in an iterative manner until a prespecified stopping criterion is met. Mukai *et al.* [11] show that a batch-based ICA method, based on a natural gradient approach, achieves better performance for fixed sources than an online method where the continuously received data is used to update statistical estimates. However, in an applied BSE application the risk is that the sources are spatio-temporally nonstationary, e.g., the spatial activity pattern of speech sources is typically nonstationary as the speakers could move around, and pauses and bursts in the speech make a speech signal temporally nonstationary. Because of this, the batch processing approach is unsuitable in such nonstationary environments due to its inherent estimation delay. Quite simply, this delay limits the method’s ability to track sources in a nonstationary environment. For this reason, the use of batch-based ICA methods (e.g., FastICA) cannot be recommended in such nonstationary environments. To address this problem, this paper focuses on a new ICA method which is related to the FastICA method. However, unlike FastICA, the new method is intended for performing online estimations of the source statistics used during BSE.

The use of ICA methods described in this paper are based on a fourth-order cumulant [15], i.e., the Kurtosis measure. The use of cumulants (higher order statistics) in order to find approximate and simple features distinguishing desired sources from undesired sources dates back to the early pioneering work of

ICA (see, for instance, [16]–[19] and references therein). However, the use of higher order cumulants for performing Kurtosis maximization has met with critique because of its sensitivity to data outliers. When an outlying data sample enters the Kurtosis algorithm, it may result in divergence in the algorithm. One practical solution to this problem is to make use of some sort of signal-conditioning before the data is sampled by a device, e.g., by using a compressor before the sampler. The rationale for using such measures as the Kurtosis, despite their sensitivity to outliers, is that they will yield a polynomial structure in their adaptive weight update equations, and this is identified in this communication as an important feature that may generate a feasible real-time digital signal processor (DSP)-based implementation.

In addition to the pure ICA task, many approaches also incorporate temporal (and/or spatial) postprocessors in order to improve performance further [20], [21]. The blind extraction of source signals discussed in this paper was performed purely without any postprocessors. It is noted, however, that additional postprocessors may be added in the future to improve the performance of the proposed method.

The outline of this paper is as follows: the adopted ICA data model is presented in Section II. The FastICA method for complex-valued data is briefly repeated, from [13], in Section III. The proposed new ICA method is presented in Section IV. A brief discussion on batch processing for ICA is given in Section V, and applied ICA is discussed in Section VI. The FastICA and the proposed method are evaluated and compared in Section VII. A summary with conclusions is given in Section VIII.

II. ICA DATA MODEL

The model assumes an array of M microphones where each received real-valued time signal is denoted $x_m[t]$ for $m \in (1, 2, \dots, M)$ and where t denotes continuous time. Each received time signal is sampled and decomposed into a time–frequency representation using a filter bank with K subbands, and where each subband signal is denoted $X_{m,k}[n]$ with $k \in (1, 2, \dots, K)$ and n is a sample index in the subband domain. The subband decomposition is carried out in the evaluation part by using a discrete Fourier transform (DFT) modulated uniform analysis filter bank and an efficient polyphase realization (see, for instance, [22] for details regarding the filter bank). Since the analysis in the next section is general and identical for all subbands, the notation will intentionally omit the subband index k . For the sake of simplicity, the focus in the presentation is only on one of the K subbands. Each subband signal is then compactly denoted $X_m[n] \equiv X_{m,k}[n]$. The received subband signals are represented using a signal vector $\mathbf{X}[n] = (X_1[n], X_2[n], \dots, X_M[n])^T$ of size $M \times 1$, where $(\cdot)^T$ denotes the vector transpose. In this paper, the commonly used noise-free ICA mixing model for complex-valued data is adopted as

$$\mathbf{X}[n] = \mathbf{A}[n]\mathbf{S}[n]. \quad (1)$$

Here, $\mathbf{A}(n)$ is a time-varying source mixing matrix of size $M \times L$, while the L original, independent sources

$\mathbf{S}[n] = (S_1[n], S_2[n], \dots, S_L[n])^T$ are assumed to obey $E\{\mathbf{S}[n]\mathbf{S}[n]^H\} = \mathbf{I}_L$. In this formula, $(\cdot)^H$ denotes the complex conjugate transpose, while $E\{\cdot\}$ represents the expectation operator, and $\mathbf{I}_L$ is an identity matrix of the size $L \times L$. According to [13] and [15], the Kurtosis value of a complex-valued signal $X[n]$ that has a circular distribution can be defined as

$$K\{X[n]\} = E\{|X[n]|^4\} - 2E\{|X[n]|^2\}^2. \quad (2)$$

The Kurtosis value of each original source signal is, due to the assumption $E\{|S_l[n]|^2\} = 1$, equal to $K\{S_l[n]\} = E\{|S_l[n]|^4\} - 2$, for $l \in (1, 2, \dots, L)$, and it is assumed that the sources are ordered so that the dominant source with the highest absolute Kurtosis value, is $S_1[n]$, i.e., $|K\{S_1[n]\}| > |K\{S_l[n]\}|$ for $l \in (2, 3, \dots, L)$. This assumption implies that a blind extraction method would extract the dominant source $S_1[n]$. The subband output signal $Y[n]$ is a linear combination of the observed input signals $\mathbf{X}[n]$ weighted by the filter vector $\mathbf{W}[n]$ of size $M \times 1$ according to

$$Y[n] = \mathbf{W}[n]^H \mathbf{X}[n] = \mathbf{W}[n]^H \mathbf{A}[n]\mathbf{S}[n]. \quad (3)$$

To continue, the adopted signal model uses one filter vector tap per subband. This model captures signal dynamics up to the frame length used in the filter bank. While it is possible to capture longer time scales by using several filter-taps per subband, i.e., FIR filtering, this method is not considered in this paper since the theoretical analysis is greatly simplified if using only one filter-tap per subband. It is desirable that the BSE method finds a $\mathbf{W}[n]$ so that $Y[n] = S_1[n]$, i.e., $\mathbf{W}[n]^H \mathbf{A}[n] = (1, 0, \dots, 0)$. The time-domain output signal $y[t]$ is then computed from the subband output signals $Y[n]$ (i.e., $Y_k[n]$ with the subband notation) by a DFT modulated synthesis filter bank matched to the analysis filter bank [22]. We will, for the sake of simplicity in the presentation, henceforth drop the time index n and adopt the following notation $\tilde{\mathbf{W}} \equiv \mathbf{W}[n]$ and $\mathbf{W} \equiv \mathbf{W}[n+1]$, this in order to discriminate the current filter vector $\tilde{\mathbf{W}}$ from the future filter vector $\mathbf{W}$. The tilde-sign (“~”) will henceforth be used to denote that *a priori* knowledge is being used.

III. FastICA METHOD REVISITED

The FastICA [13] contrast function used here is the Kurtosis contrast function $G(u) = (1/2)u^2$, i.e.,

$$J_G(\mathbf{W}) = E\left\{G\left(|\mathbf{W}^H \mathbf{X}|^2\right)\right\} = \frac{1}{2}E\{|Y|^4\}. \quad (4)$$

The rationale of focusing on the Kurtosis contrast function is that it will yield a polynomial structure in its update equation. A polynomial structure can be realized in hardware using a pre-determined series of multiplication operations. This is often beneficial in a real-time implementation where a low complexity is preferable. According to [13], the optima of $J_G(\mathbf{W})$ conditioned on $E\{|\mathbf{W}^H \mathbf{X}|^2\} = 1$ follows the Kuhn–Tucker conditions, where the cost function $J_C(\mathbf{W}) = J_G(\mathbf{W}) - \beta E\{|Y|^2\}$ is being used, and β is a real-valued parameter.

The filter vector update equation at a point $\tilde{\mathbf{W}}$ according to an approximative Newton’s method [13], [23], with an additional

unity norm constraint applied, is

$$\mathbf{W}^+ = -\tilde{\mathbf{W}} + \frac{E\{\mathbf{X}\mathbf{X}^H\}^{-1}E\{\tilde{\mathbf{Y}}^*\tilde{\mathbf{Y}}|\mathbf{X}\}}{2E\{|\tilde{\mathbf{Y}}|^2\}} \quad (5)$$

$$\mathbf{W} = \frac{\mathbf{W}^+}{\|\mathbf{W}^+\|} \quad (6)$$

where $\mathbf{W}^+$ is a temporary variable, and $\tilde{\mathbf{Y}} = \tilde{\mathbf{W}}^H\mathbf{X}$. The normalization approach in (6) is used to avoid the trivial solution $\mathbf{W} = \mathbf{0}_M$, where $\mathbf{0}_M$ denotes a null-vector of size $M \times 1$. This normalization approach will preserve the power of the source signal [2]. The normalization is henceforth assumed to be performed after each update of the temporary variable $\mathbf{W}^+$.

A. Fixed-Point Behavior of the FastICA Method

It is customary to perform a preprocessing whitening of the input data using, for instance, principal component analysis (PCA) (e.g., [1]) in order to speed up the convergence of the FastICA method. The PCA decorrelation has the same impact as if the mixing matrix $\mathbf{A}$ would possess a unitary property, i.e., $E\{\mathbf{X}\mathbf{X}^H\} = \mathbf{A}E\{\mathbf{S}\mathbf{S}^H\}\mathbf{A}^H = \mathbf{A}\mathbf{A}^H = \mathbf{I}_M$.

The output signal of the FastICA method is denoted $Y = \mathbf{W}^H\mathbf{X} = \mathbf{W}^H\mathbf{A}\mathbf{S}$, according to (3). It is convenient to define a vector $\mathbf{Q}$ as $\mathbf{Q} = \mathbf{A}^H\mathbf{W}$ (or $\mathbf{W} = \mathbf{A}\mathbf{Q}$), which yields that $Y = \mathbf{Q}^H\mathbf{S}$. The unity norm constraint yields that $\|\mathbf{W}\|_2 = 1$ and $\|\mathbf{Q}\|_2 = 1$.

To prove the fixed-point behavior of FastICA, let the filter vector at the previous iteration equal the optimal solution that extracts the dominant source S_1

$$\tilde{\mathbf{Q}} = \mathbf{Q}_{\text{opt}} = (1, 0, \dots, 0)^T. \quad (7)$$

In other words, the optimal solution $\mathbf{W}_{\text{opt}}^H = \mathbf{Q}_{\text{opt}}^H\mathbf{A}^H$ corresponds to the first row of the inverse (or if $L \neq M$, the Moore–Penrose pseudo inverse) $\mathbf{A}^{-1} = \mathbf{A}^H$ of the matrix $\mathbf{A}$. The definition of $\mathbf{Q}_{\text{opt}}$ in (7) yields that

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{Q}}^H\mathbf{S} = \mathbf{Q}_{\text{opt}}^H\mathbf{S} = S_1. \quad (8)$$

In this way, the behavior of the FastICA method at an optimal solution is

$$\mathbf{Q}^+ = \frac{1}{2}K\{S_1\}\mathbf{Q}_{\text{opt}}. \quad (9)$$

If the dominant source S_1 possesses a nonzero Kurtosis value, i.e., $K\{S_1\} \neq 0$, then the updated filter vector provided by FastICA, $\mathbf{W} = \mathbf{W}^+/\|\mathbf{W}^+\|_2 = \pm\mathbf{W}_{\text{opt}}$, is a stable fixed-point optimal solution, and the sign $\pm$ is determined by the sign of the Kurtosis of the dominant source. The FastICA method is, by virtue of this property, denoted a fixed-point method. However, the FastICA is undefined for Gaussian-only mixtures, since $K\{S_1\} = 0$ yields a division-by-zero in the filter vector normalization stage in (6). These conclusions are already established properties of the FastICA method (see for instance [2], [13]).

B. Local Consistency of the FastICA Method

This paper follows the analysis of the local consistency of FastICA conducted in [13]. This analysis was conducted

around the optimal point $\mathbf{W}_{\text{opt}}$ which extracts the dominant source. The analysis is conducted by evaluating a second-order Taylor expansion (any term with an order higher than two is omitted) of $J_G(\tilde{\mathbf{W}})$ at the point $\tilde{\mathbf{W}} = \mathbf{W}_{\text{opt}} + \mathbf{P}$, where $\mathbf{P}$ is a small perturbation vector to the optimal solution (here, the term “small” implies that $\|\mathbf{P}\|_2 \ll 1$). The unity norm constraint (6) yields that the perturbation vector must satisfy $\|\mathbf{W}_{\text{opt}} + \mathbf{P}\|_2^2 = \|\mathbf{W}_{\text{opt}}\|_2^2 = 1$, and thus the perturbed optimal solution is evaluated at a hyper-sphere. The second-order Taylor series expansion of $J_G(\mathbf{W})$ around the point $\mathbf{W}_{\text{opt}}$ is

$$J_G(\mathbf{W}_{\text{opt}} + \mathbf{P}) = J_G(\mathbf{W}_{\text{opt}}) + 2E\{|S_1|^4\} \text{Re}\{\mathbf{P}^H\mathbf{W}_{\text{opt}}\} + 2\mathbf{P}^H\mathbf{P}. \quad (10)$$

As long as the optimal solution $\mathbf{W}_{\text{opt}}$ as well as the perturbed optimal solution $\mathbf{W}_{\text{opt}} + \mathbf{P}$ obey the unity norm constraint, the following relationship holds (from [13]):

$$2\text{Re}\{\mathbf{P}^H\mathbf{W}_{\text{opt}}\} = -\mathbf{P}^H\mathbf{P}. \quad (11)$$

This relationship yields

$$J_G(\mathbf{W}_{\text{opt}} + \mathbf{P}) = J_G(\mathbf{W}_{\text{opt}}) - K\{S_1\}\mathbf{P}^H\mathbf{P}. \quad (12)$$

The term $\mathbf{P}^H\mathbf{P}$ is always greater than or equal to 0, and the type (local maximum or minimum) of the optimal solution $\mathbf{W}_{\text{opt}}$ is therefore dependent on the sign of $K\{S_1\}$, i.e., the sign of the Kurtosis of the dominant source. If $K\{S_1\} > 0$, the optimal solution $\mathbf{W}_{\text{opt}}$ is a local maximum. The optimum is a local minimum if $K\{S_1\} < 0$, and it is a saddle point if $K\{S_1\} = 0$, i.e., if S_1 is Gaussian distributed. This result is identical to that in [13].

C. Implications of Performing Applied BSE Using FastICA

When the FastICA is used for blind signal extraction, it is required that at least the dominant source is non-Gaussian, i.e., has a nonzero Kurtosis value. This has implications in a real-time realization of the FastICA method, where the true source statistics are estimated using sample-based estimators. Such a sample-based estimator should have a finite memory, or at least a rather short integration time, in order to be able to track changes in the environment and to restrain the memory requirement. Hence, the FastICA method risks divergence in a scenario where a spatio-temporally nonstationary source is mixed with one or several stationary Gaussian sources, and the non-Gaussian source becomes inactive for the duration of the sample-based estimator's memory length. This behavior is emphasized in the evaluation, Section VII, where a recursive sample-based estimator is used to estimate source signal statistics. The remedy to this behavior is increasing the memory length of the sample-based estimator. However, in that case the method fails in tracking sources in a rapidly changing environment. In addition to this, the memory requirement in the realization can, in some cases, be increased. These aforementioned issues regarding applied BSE, and more specifically those related to the shortcomings of the FastICA method, are the main reasons for proposing the new ICA method.

IV. PROPOSED ICA METHOD

The problem that needs to be resolved is that the FastICA method is divergent for Gaussian-only source mixtures. The idea of this paper is to propose an alternative Kurtosis measure that weighs the fourth-order term $E\{|Y|^4\}$ towards the square second-order term $E\{|Y|^2\}^2$ in the Kurtosis measure so as to avoid the divergent behavior. An alternative method for BSE is therefore proposed that shares important properties with the FastICA method, such as the fixed-point behavior, but which circumvents the divergent behavior of FastICA for a mixture of Gaussian-only sources. A general cost function for a weighted Kurtosis measure that encapsulates both the FastICA method and the proposed method is described by the cost function $C(\mathbf{W}, \alpha)$

$$C(\mathbf{W}, \alpha) = \frac{1}{2} K\{Y\} + \alpha E\{|Y|^2\}^2 \\ = \frac{1}{2} E\{|Y|^4\} - (1 - \alpha) E\{|Y|^2\}^2 \quad (13)$$

where α is a real-valued parameter. If, for instance, $\alpha = 1$, then $C(\mathbf{W}, 1) = (1/2)E\{|Y|^4\}$ equals the Kurtosis contrast function used in the FastICA method, i.e., (4). Various α -values yield BSE algorithms with different properties. This paper focuses on a certain α -value: $\alpha = -1$. This value yields a non-divergent BSE algorithm that can be used to construct a new method, henceforth denoted Kurtosis maximization in the Sub-band domain ICA (KSICA). With the help of this method, the corresponding contrast function can be defined as $J(\mathbf{W}) = C(\mathbf{W}, -1)$

$$J(\mathbf{W}) = \frac{1}{2} E\{|Y|^4\} - 2E\{|Y|^2\}^2. \quad (14)$$

A Newton-based method is used here to optimize the contrast function $J(\mathbf{W})$ of the KSICA method, where the gradient and the Hessian matrix of $J(\mathbf{W})$ are evaluated as

$$\nabla_{\mathbf{W}} J(\mathbf{W}) = E\{Y^* |Y|^2 \mathbf{X}\} - 4E\{|Y|^2\} E\{Y^* \mathbf{X}\} \quad (15)$$

$$\nabla_{\mathbf{W}}^2 J(\mathbf{W}) = 2E\{|Y|^2 \mathbf{X} \mathbf{X}^H\} \\ - 4(E\{Y^* \mathbf{X}\} E\{Y \mathbf{X}^H\} \\ + E\{|Y|^2\} E\{\mathbf{X} \mathbf{X}^H\}). \quad (16)$$

Following the approximations outlined for the FastICA method, it is assumed that $E\{|Y|^2 \mathbf{X} \mathbf{X}^H\} \approx E\{|Y|^2\} E\{\mathbf{X} \mathbf{X}^H\}$ and that $E\{Y^* \mathbf{X}\} E\{Y \mathbf{X}^H\} \approx E\{|Y|^2\} E\{\mathbf{X} \mathbf{X}^H\}$, which yields

$$\nabla_{\mathbf{W}}^2 J(\mathbf{W}) \approx -6E\{|Y|^2\} E\{\mathbf{X} \mathbf{X}^H\}. \quad (17)$$

The filter vector update for KSICA according to Newton's method at a point $\tilde{\mathbf{W}}$, and under a unity norm constraint, is

$$\mathbf{W}^+ = \tilde{\mathbf{W}} + \frac{E\{\mathbf{X} \mathbf{X}^H\}^{-1} E\{\tilde{Y}^* |\tilde{Y}|^2 \mathbf{X}\}}{2E\{|\tilde{Y}|^2\}} \quad (18)$$

$$\mathbf{W} = \frac{\mathbf{W}^+}{\|\mathbf{W}^+\|}. \quad (19)$$

The update equation of KSICA in (18) is identical to the update equation of FastICA in (5), except for the point that FastICA has a negative $-\tilde{\mathbf{W}}$ term in its update.

A. Fixed-Point Behavior of the KSICA Method

Following the assumptions in Section III-A, where a pre-processing whitening of the input data is performed so that $E\{\mathbf{X} \mathbf{X}^H\} = \mathbf{I}_M$, and the analysis is conducted around the optimal point $\mathbf{Q}_{\text{opt}} = \mathbf{A} \mathbf{W}_{\text{opt}}$, it may be concluded that the behavior of the KSICA method at an optimal solution follows

$$\mathbf{Q}^+ = \frac{1}{2} (K\{S_1\} + 4) \mathbf{Q}_{\text{opt}} \quad (20)$$

$$\mathbf{Q} = \frac{\mathbf{Q}^+}{\|\mathbf{Q}^+\|} = \mathbf{Q}_{\text{opt}}. \quad (21)$$

Consequently, the KSICA is a fixed-point method like the FastICA method, i.e., once the KSICA has found an optimal solution, it stays at that optimal solution as the iterations proceed. Furthermore, since the term $(1/2)(K\{S_1\} + 4)$ in (20) is always positive and nonzero, the KSICA is always stable and it is not inhibited by any probability distribution assumptions regarding the source signals. This is contrary to the FastICA method, in which at least the dominant source must have a nonzero Kurtosis value in order to avoid divergence. This property of the KSICA makes it tractable in a real-time application where the non-Gaussianity assumption cannot always be guaranteed. It must be stressed that while the KSICA and FastICA share the fundamental assumptions imposed by the theory of ICA, which disallows separation of Gaussian-only sources, the KSICA does not diverge in the case of Gaussian-only sources. It is this difference from FastICA which makes the KSICA a superior candidate for performing BSE in an online setting where constant activity of non-Gaussian sources may not be guaranteed.

B. Local Consistency of KSICA

An analysis of the local consistency of the KSICA method follows the local consistency-analysis of the FastICA method in Section III-B. Therefore, a preprocessing whitening stage of the data is also included here. The second-order Taylor expansion of $J(\mathbf{W})$ around the point $\mathbf{W}_{\text{opt}}$ is

$$J(\mathbf{W}_{\text{opt}} + \mathbf{P}) = J(\mathbf{W}_{\text{opt}}) + 2(K\{S_1\} - 2) \text{Re}\{\mathbf{P}^H \mathbf{W}_{\text{opt}}\} \\ - 6\mathbf{P}^H \mathbf{P}. \quad (22)$$

Using the relationship in (11) yields

$$J(\mathbf{W}_{\text{opt}} + \mathbf{P}) = J(\mathbf{W}_{\text{opt}}) - (K\{S_1\} + 4) \mathbf{P}^H \mathbf{P}. \quad (23)$$

Consequently, the term $-(K\{S_1\} + 4) \mathbf{P}^H \mathbf{P}$ is always less than or equal to 0, and the optimal solution $\mathbf{W}_{\text{opt}}$ is thereby always a local maxima to $J(\mathbf{W})$, independent of the distribution of the sources. This result further implies that the optimal solution $\mathbf{W}_{\text{opt}}$ related to the dominant source S_1 is in fact always a global maximum solution.

V. BATCH PROCESSING FOR ICA

When performing BSE using a batch processing approach, input data is collected for a certain amount of time corre-

sponding to the batch duration. The set of recorded data in a batch indexed b is denoted $\mathbf{X}_b$ of the size $N \times M$, where N is the number of data samples in the batch and M is the number of microphones, as before. The statistical measures are then estimated for all data available in the data batch. The BSE method then follows a prespecified iterative schedule for each data batch b in order to numerically find an optimal solution.

- 1) Compute an *a priori* output signal vector $\tilde{\mathbf{Y}}_b = \mathbf{X}_b \tilde{\mathbf{W}}_b^* = \mathbf{X}_b \mathbf{W}_{b,i-1}^*$ at each iteration $i = 1, 2, \dots$.
- 2) Estimate source signal statistics in the temporary variables $\hat{p}_{\tilde{\mathbf{Y}}}$, $\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}$, and $\hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}$

$$\begin{aligned}\hat{p}_{\tilde{\mathbf{Y}}} &= \frac{1}{N} \tilde{\mathbf{Y}}_b^H \tilde{\mathbf{Y}}_b \\ \hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}} &= \frac{1}{N} \mathbf{X}_b^T \mathbf{X}_b \\ \hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}} &= \frac{1}{N} \mathbf{X}_b^T \text{diag} \left\{ \tilde{\mathbf{Y}}_b^* \right\}^2 \tilde{\mathbf{Y}}_b\end{aligned}$$

where $\text{diag} \{ \cdot \}$ produces a diagonal matrix.

- 3) Compute a temporary filter weight vector $\mathbf{W}^+$

$$\begin{aligned}\text{KSICA} : \mathbf{W}^+ &= \mathbf{W}_{b,i-1} + \frac{\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}^{-1} \hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}}{2\hat{p}_{\tilde{\mathbf{Y}}}} \\ \text{FastICA} : \mathbf{W}^+ &= -\mathbf{W}_{b,i-1} + \frac{\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}^{-1} \hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}}{2\hat{p}_{\tilde{\mathbf{Y}}}}.\end{aligned}$$

- 4) Update the normalized filter weight vector

$$\mathbf{W}_{b,i} = \frac{\mathbf{W}^+}{\|\mathbf{W}^+\|}.$$

- 5) When a stopping criterion is met, or if a specific number of iterations has passed, let $\mathbf{W}_{b,\text{opt}} = \mathbf{W}_{b,i}$ and stop the iterations for this batch. Otherwise, go to 1).

One may initialize the starting point $\mathbf{W}_{b,0}$ as a random vector with unit norm.

VI. ON APPLIED BSE USING ICA

In an applied real-time application of ICA, it is necessary to use sample-based estimators to estimate the statistics of the unknown sources since their true statistics are generally not directly available. Expectations are replaced by their sample-based averages, where auto regressive (AR) averages are convenient while their “memory length” or integration time can easily be changed by adjusting a small set of parameters [1]. In this paper, we have made use of a first-order AR averaging technique, where an approximation $\hat{m}_X(n)$ of the expectation operator for the signal $X(n)$ is defined as

$$\hat{m}_X(n) = \lambda \hat{m}_X(n-1) + (1-\lambda)X(n). \quad (24)$$

The parameter $\lambda \in [0, 1]$ is a constant associated with the integration time of the AR-average.

A. Online Parameter Estimation

The KSICA and the FastICA share the expectations $E\{|\tilde{Y}|^2\}$, $E\{\mathbf{X}\mathbf{X}^H\}$, and $E\{\tilde{Y}^*|\tilde{Y}|^2\mathbf{X}\}$. These expectations are herein approximated using the following first-order AR averages:

$$\hat{p}_{\tilde{\mathbf{Y}}}(n) = \lambda \hat{p}_{\tilde{\mathbf{Y}}}(n-1) + (1-\lambda) |\tilde{Y}(n)|^2 \quad (25)$$

$$\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n) = \lambda \hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n-1) + (1-\lambda) \mathbf{X}(n) \mathbf{X}(n)^H \quad (26)$$

$$\hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}(n) = \lambda \hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}(n-1) + (1-\lambda) \tilde{Y}(n)^* |\tilde{Y}(n)|^2 \mathbf{X}(n). \quad (27)$$

The inverse of the estimator in (26) is found through the matrix inversion lemma [23], as

$$\begin{aligned}\mathbf{T} &= \hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n-1)^{-1} \mathbf{X}(n) \\ \hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n)^{-1} &= \lambda^{-1} \hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n-1)^{-1} \\ &\quad - \frac{\mathbf{T} \mathbf{X}(n)^H \hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n-1)^{-1}}{\lambda^2(1-\lambda)^{-1} + \lambda \mathbf{X}(n)^H \mathbf{T}}\end{aligned} \quad (28)$$

where the temporary vector $\mathbf{T}$ has been incorporated for the sake of clarity in the presentation. The online KSICA update equation is

$$\mathbf{W}^+ = \mathbf{W}(n) + \frac{\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n)^{-1} \hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}(n)}{2\hat{p}_{\tilde{\mathbf{Y}}}(n)} \quad (29)$$

$$\mathbf{W}(n+1) = \frac{\mathbf{W}^+}{\|\mathbf{W}^+\|}. \quad (30)$$

The online FastICA update equation looks similar to this

$$\mathbf{W}^+ = -\mathbf{W}(n) + \frac{\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}(n)^{-1} \hat{\mathbf{r}}_{\mathbf{X}\tilde{\mathbf{Y}}}(n)}{2\hat{p}_{\tilde{\mathbf{Y}}}(n)} \quad (31)$$

$$\mathbf{W}(n+1) = \frac{\mathbf{W}^+}{\|\mathbf{W}^+\|}. \quad (32)$$

VII. EVALUATION OF KSICA AND FastICA

Two different evaluations are performed to get an overall picture of the proposed method. First, the KSICA and FastICA methods are evaluated in a batch processing approach according to Section V. The second part of the evaluation deals with an analysis of the methods’ capabilities to extract human speech from an observed mixture of speech and interfering noise. In the second part, online estimates are used to measure the source statistics, and the methods are operating in an online mode according to Section VI.

A. Evaluation in a Batch Processing Mode

Two sources and two sensors are used in this part of the evaluation, and the elements of the mixing matrix $\mathbf{A}$ from (1) are randomized and computed so that the unitary property of $\mathbf{A}$ is preserved.

Two cases are evaluated; first, the sources are assumed to have circular probability distribution functions where the source signals are $S_1(n) = A_1(n)e^{j\phi_1(n)}$ and $S_2(n) = A_2(n)e^{j\phi_2(n)}$, where $A_1(n)$, $A_2(n)$, $\phi_1(n)$, and $\phi_2(n)$ are independent real-

Fig. 1. Mean (solid) and standard deviation (dashed) of performance index (33) in dB in a batch processing setup where KSICA “○” and FastICA “×.” The source signals have circular distributions (upper), and the source signals do not have circular distributions (lower). The mean and standard deviation of the performance index should ideally be $-\infty$ dB.

valued random processes. Here, $A_1(n)$ follows a Laplacian distribution which is common in speech models [24], $A_2(n)$ follows a Gaussian distribution, while the random phase signals $\phi_1(n)$ and $\phi_2(n)$ follow a uniform distribution in the interval $[-\pi, +\pi]$. In the second case, the source signals are not circular and the real and imaginary parts of $S_1(n)$ also follow a Laplacian distribution, whereas the real and imaginary parts of $S_2(n)$ follow a Gaussian distribution. The variances of the two sources are identical and set to unity. Furthermore, the two sources are independent. To evaluate the performance of the batch approach, one can analyze the behavior of the vector $\mathbf{Q}_i = \mathbf{W}_i^H \mathbf{A}$ at iteration index i . When the extraction method converges, the absolute value of the elements in $\mathbf{Q}_i$ should have the values one and zero, i.e., $|Q_i(1)| = 1$ and $|Q_i(2)| = 0$ since that implies that $Y_i = S_1$. Our performance index $p(i)$ measures the deviation of the solution from this desired behavior according to

$$p(i) = \frac{|Q_i(2)|^2}{|Q_i(1)|^2}. \quad (33)$$

In other words, if $p(i) = 0$ which corresponds to $-\infty$ dB, this implies that the source signal is fully extracted. The outcome of 1000 realizations is averaged, and the resulting mean performance index, together with the standard deviation for various batch sizes, is provided in Fig. 1. The analysis shows that the performance of the KSICA method exceeds that of the FastICA when the batch size is lowered. This is true both in the case with circular sources and in the case with noncircular sources. Furthermore, the standard deviation of the performance index is significantly lower in the proposed KSICA method when compared to the FastICA method. This result is further validation that the KSICA method is more appropriate to use in an online setting. It can be added in this discussion that the FastICA approach generally converges within three to five iterations and

Fig. 2. Source setup in an online mode: The two microphones are situated in p_1 and p_2 , the interfering noise source is situated in p_3 . For a stationary evaluation, the speech source is situated in p_4 , and for a nonstationary evaluation, the speech source is moving along p_7 starting in p_5 and ending in p_6 .

that the KSICA converges to the same steady state value within about seven more iterations.

B. Evaluation in an Online Mode

This evaluation concerns the performance of the KSICA and FastICA methods in an online setting in which two microphones are used. Two different source configurations are evaluated: one configuration uses spatially stationary sources and the second configuration uses a spatially nonstationary moving speech source. The configurations are outlined according to Fig. 2 where two microphones are situated in p_1 and p_2 , and the noise source is situated in p_3 . For the first configuration, the speech source is spatially stationary and situated in p_4 . In the second configuration the speech source is moving along the path p_7 from p_5 to p_6 and is in a constant motion.

1) *Evaluation With a Spatially Stationary Speech Source:* The evaluation assesses improvements in signal-to-interference ratio (SIR) and degradation of perceptual speech quality according to the ITU-T standard p.862, Perceptual Evaluation of Speech Quality (PESQ) [25]. A key factor in an online BSE method is the learning rate λ (see Section VI). The performance measures will be assessed for a variety of λ -values in order to provide a complete picture of the methods' performances.

The interfering noise source is spatially and temporally stationary, and it is situated in p_3 (see Fig. 2) while the speech source is spatially stationary and temporally nonstationary, and situated in p_4 (see Fig. 2). The source signals are prerecorded with a sampling frequency of 8 kHz, 25 s long, and subject to free-field propagation [26]. The speech source is active 50% of the time. The filter bank uses 128 subbands and a two times oversampling. The prototype filter is a 192-tap-long Hamming window.

A measure of the SIR improvement and the PESQ measure is used to evaluate the BSE methods in a spatially stationary online configuration. The filter weights at each iteration are stored and used for filtering the original convolved, but unmixed, source signals. This enables direct access to the evaluation measures. The SIR improvement performance measure, denoted P_{SIR} , is defined as

$$P_{\text{SIR}} = \frac{\text{Var}\{y_s(t)\} \text{Var}\{x_{1,w}(t)\}}{\text{Var}\{x_{1,s}(t)\} \text{Var}\{y_v(t)\}} \quad (34)$$

where $\text{Var}\{\cdot\}$ denotes an estimator of variance, $y_s(t)$ and $y_v(t)$ represent the speech and the interfering noise components of the

Fig. 3. Performance measures SIR improvement (left) and PESQ (right). Input SIR is +10 dB “x” KSICA and “*” FastICA. Input SIR is -10 dB “v” KSICA and “o” FastICA.

enhanced output signal, and, similarly, the signals $x_{1;s}(t)$ and $x_{1;v}(t)$ represent the speech and interfering noise components of the first simulated microphone signal. The first microphone is used as a reference in the analysis. The PESQ is an automated method for the objective assessment of perceptual sound quality, and it uses a perceptual model of how sound quality is perceived by humans. The PESQ computes a perceptual model for a clean reference speech signal $x_{1;s}(t)$ and a perceptual model for the processed output speech component $y_s(t)$. The perceptual difference between the clean speech signal and the processed speech signal is mapped on the mean opinion score (MOS) [27], yielding a value between one and five, where one indicates a poor perceptual speech quality and five indicates an excellent perceptual speech quality.

The evaluated performances of the two BSE methods are presented in Fig. 3. This figure shows that the KSICA method provides a stable and good performance for a time constant around 0.15 s, which corresponds to a fast converging method. FastICA provides a performance that is only similar to the KSICA for a time constant above 1 s, which corresponds to a comparatively slower converging method. It may also be noted that the FastICA continuously provides a lower speech quality as opposed to the proposed KSICA method, where the MOS-distance is exceeding one MOS-unit for a time constant around 0.15 s in the 10-dB SIR case and exceeding 0.5 MOS-unit in the -10-dB SIR case. The difference in speech quality decreases between the two methods as the time constant increases.

2) *Evaluation With a Moving Speech Source:* The interfering noise source is spatially and temporally stationary, and it is situated in p_3 (see Fig. 2) while the speech source is spatially and temporally nonstationary, moving along the path p_7 from p_5 to p_6 (see Fig. 2). The source signals, the propagation model, and the system parameters are the same as in the previous online evaluation (see Section VII-B1).

In order to capture the performance in this dynamic configuration, the SIR improvement of the adaptive filter weights is assessed for each frame of the input data (the frame length is 64 samples), and in each subband. It is assumed that the source is stationary during the length of a frame (the radial velocity of the

Fig. 4. SIR improvement for KSICA (black, solid) and FastICA (black, dashed) where the time constant of the λ -parameter is 0.15 s. KSICA (grey, solid) and FastICA (grey, dashed) where the time constant of the λ -parameter is 1 s. The input SIR is 0 dB.

speech source is $0.0288^\circ/\text{frame}$). The SIR improvement is computed as the mean array gain in the speech source direction over the mean array gain in the direction of the interfering source.

The evaluated performances of the two BSE methods are presented in Fig. 4. This figure shows that the KSICA method provides 10 dB higher SIR improvement in relation to the FastICA method when the time constant of the λ -parameter corresponds to 0.15 s. The FastICA does perform as well as the KSICA if the time constant is set to 1 s. However, long time constants are obviously undesirable in a nonstationary application, as can be seen in Fig. 4 where the 1-s setting approaches the 0.15-s setting first after 20 s of adaptation.

VIII. SUMMARY AND CONCLUSION

This paper presents a new method, denoted KSICA, for online BSE of complex-valued signal mixtures. The proposed method uses a Kurtosis contrast function that is a modification of the FastICA Kurtosis contrast function, and this modification is introduced in order to improve certain aspects of the FastICA method in an applied setting. The KSICA method and the FastICA method are encapsulated by a uniform cost function (13) in which a scalar constant weighs the fourth-order term to the square second-order term in the Kurtosis measure. The weighting introduced by FastICA (based on the Kurtosis contrast function) yields a sensitivity to Gaussian-only signal mixtures, while the FastICA method diverges for such signal mixtures. The proposed method uses a specific choice of weighting parameter in order to circumvent this divergent behavior of the FastICA.

An analysis of the proposed method is derived through comparison with the FastICA method. Evaluation of the method in a batch processing configuration shows that FastICA is sensitive to low batch sizes, whereas the proposed KSICA method is considerably less sensitive. For instance, the performance of the KSICA method is improved by 10 dB (also with 10 dB lower standard deviation) with respect to the FastICA method if the data batch size is less than 100 samples. The two methods both derive from the assumption that the sources have circular distributions. In the batch approach, it is furthermore shown that the KSICA is insensitive to a violation of this circularity

assumption. In other words, the KSICA yields good results also when the sources are noncircular. The performance of the two methods equates as the batch size is considerably increased, above 100 data samples. The insensitivity of the proposed KSICA method compared to the FastICA method in a batch setup is seen as an indication that the KSICA method is preferable in an online setup where the requirement of non-Gaussianity for at least the dominant source cannot be constantly guaranteed.

The two methods are further analyzed in an online approach, with two active sources in a free-field propagation model. The interfering noise source, a Gaussian source, is both spatially and temporally stationary, while the target speech source, with a higher Kurtosis value, is spatially stationary and temporally nonstationary. The statistical measures are estimated in the online setup using received real data only and the KSICA and FastICA update schemes are performed on a sample-by-sample basis. Also, the KSICA method shows significant performance improvements over the FastICA method in the online setup, both in terms of SIR improvement and in preservation of perceptual speech quality. For example, when performed at a time constant of 0.15 s in the algorithm update equation, the KSICA provides a 10 dB higher SIR improvement and more than one MOS-unit better perceptual speech quality in a 10-dB input SIR scenario. In order for the FastICA to reach the same performance, the time constant needs to be increased above 1 s. Furthermore, in an online approach, where the speech source is moving, the KSICA provides about 10 dB higher interference suppression compared to the FastICA at a time constant of 0.15 s. If the time constant is 1 s, the FastICA approaches the performance of KSICA after 20 s.

Further research may include an analysis of various values for the parameter α in the general cost function (13). This paper only evaluates two values for α , where $\alpha = -1$ yields the proposed KSICA method while $\alpha = 1$ yields the FastICA method (with a Kurtosis contrast function). Future research should also investigate the algorithm's performance under the influence of observation noise. Future analysis should extend the presented evaluation by evaluating real measured data, under various operating conditions, e.g., the number of microphones, the number of subbands, the use of more taps in the filter-and-sum subband beamformer's FIR filters, and changing the oversampling ratio. Also, evaluation of the algorithm performance under double-talker situations is also important to undertake.

REFERENCES

- [1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing—Learning Algorithms and Applications*. New York: Wiley, 2003.
- [2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [3] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind extraction of a dominant source signal from mixtures of many sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, vol. 3, pp. 61–64.
- [4] B. Sällberg, M. Swartling, N. Grbić, and I. Claesson, "Real time implementation of a blind beamformer for subband speech enhancement using kurtosis maximization," in *Proc. Int. Workshop Acoust., Echo, Noise Control*, 2006, pp. 485–489.

- [5] B. Sällberg, N. Grbić, and I. Claesson, "Online maximization of subband Kurtosis for blind adaptive beamforming in realtime speech extraction," in *Proc. IEEE 15th Int. Conf. Digital Signal Process.*, 2007, pp. 603–606.
- [6] B. Sällberg, N. Grbić, and I. Claesson, "Online blind speech extraction based on a locally quadratic Kurtosis criteria and a preprocessing automatic gain controller," in *Proc. IEEE 49th Int. Symp. ELMAR*, 2007, pp. 139–142.
- [7] B. Sällberg, N. Grbić, and I. Claesson, "An adaptive blind beamformer with an integrated single-channel noise reduction method for robust realtime blind speech extraction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2008, pp. 309–312.
- [8] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Elsevier Neurocomputing*, vol. 22, no. 1–3, pp. 21–34, 1998.
- [9] N. Grbić, X.-J. Tao, S. E. Nordholm, and I. Claesson, "Blind signal separation using overcomplete subband representation," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 524–533, Sep. 2001.
- [10] B. W. Gillespie, H. S. Malvar, and D. A. F. Florêncio, "Speech dereverberation via maximum-kurtosis subband adaptive filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2001, vol. 6, pp. 3701–3704.
- [11] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Robust real-time blind source separation for moving speakers in a room," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 5, pp. 469–472.
- [12] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Blind source separation of many signals in the frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. 5, pp. 969–972.
- [13] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, 2000.
- [14] S. C. Douglas, "Fixed-point FastICA algorithms for the blind separation of complex-valued signal mixtures," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, 2005, pp. 1320–1325.
- [15] C. Nikias and A. Petropulu, *Higher-Order Spectral Analysis—A Non-linear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [16] J.-F. Cardoso, "Source separation using higher order moments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1989, vol. 4, pp. 2109–2112.
- [17] A. Cichocki, R. Thawonmas, and S. Amari, "Sequential blind signal extraction in order specified by stochastic properties," *Electron. Lett.*, vol. 33, no. 1, pp. 64–65, Jan. 1997.
- [18] J. P. LeBlanc and P. L. De Le' on, "Speech separation by kurtosis maximization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1998, vol. 2, pp. 1029–1032.
- [19] F. J. Theis and Y. Inouye, "On the use of joint diagonalization in blind signal processing," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2006, pp. 3586–3589.
- [20] S.-Y. Low, S. Nordholm, and R. Togneri, "Convolutional blind signal separation with post-processing," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 529–548, Sep. 2004.
- [21] R. Aichner, M. Zourub, H. Buchner, and W. Kellermann, "Post-processing for convolutional blind source separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. 5, pp. 37–40.
- [22] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [23] S. Haykin, *Adaptive Filter Theory*. New York: Wiley, 2002.
- [24] W. Zhang and S. Gazor, "Statistical modelling of speech signals," in *Proc. IEEE Int. Conf. Signal Process.*, 2002, vol. 1, pp. 480–483.
- [25] *Perceptual Evaluation of Speech Quality (PESQ)*, ITU-T p.862.
- [26] D. Johnson and D. Dudgeon, *Array Signal Processing—Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [27] *Methods for Subjective Determination of Transmission Quality*, ITU-T p.800, Annex B, 1996.

Benny Sällberg (M'04) was born in Sweden in 1979. He received the M.Sc. degree and Lic. of Technology in telecommunications degree from the Blekinge Institute of Technology, Ronneby, Sweden, in 2003 and 2006, respectively. He is currently pursuing the Ph.D. degree in telecommunications.

His research interests include speech enhancement, adaptive beamforming, blind speech extraction, and he has a pronounced focus on robust methods for real-time speech enhancement.

Nedelko Grbić (M'97) was born in Sweden in 1971. He received the B.S. degree from the University/College of Falun, Borlänge, Sweden, in 1993, the M.Sc. and Ph.D. degrees from the Blekinge Institute of Technology, Ronneby, Sweden, in 1997 and 2001, respectively.

He was appointed as an Associate Professor in 2006 at the Blekinge Institute of Technology. His research interests include array techniques in the field of speech enhancement, adaptive beamforming, blind equalization, blind signal separation and blind speech extraction in various applications such as binaural hearing aids, hands-free speech communication, conference telephony, and underwater acoustics.

Ingvar Claesson (M'91) was born in Broby, Sweden, in 1957. He received the Dipl.Eng. and Ph.D. degrees from Lund University, Lund, Sweden, in 1980 and 1986, respectively.

He was appointed Senior Lecturer in Telecommunication Theory at Lund University in 1986, and was made Associate Professor in 1992. Since May 1998, he has held the chair of Signal Processing at the Blekinge Institute of Technology, Ronneby, Sweden. In 1990, he was one of the founders of the Department of Signal Processing, Blekinge Institute of Technology, and is currently Dean at Blekinge Institute of Technology, Head of Research, and Principal Supervisor in Signal Processing there. His current research interests are in adaptive signal processing, blind equalization, adaptive beamforming, speech enhancement, blind signal separation, active noise control, filter design, and antenna arrays.