Experimental Upper Bound for the Performance of Convolutive Source Separation Methods

Kenneth E. Hild II, Deniz Erdogmus*, Jose C. Principe**

k.hild@ieee.org, derdogmus@ieee.org, principe@cnel.ufl.edu

Biomagnetic Imaging Laboratory, Rm. C-324B

The University of California at San Francisco, San Francisco, CA 94122

*Departments of Computer Science and Engineering, and Biomedical Engineering

OGI School of Science and Engineering, Oregon Health & Science Univ., Beaverton, OR 97006

**Computational NeuroEngineering Laboratory, Rm. NEB 451

The University of Florida, Gainesville, FL 32611

Contact Author: Jose Principe, 352-392-2662 (office), 352-392-0044 (fax)

EDICS: SSP-PERF, SSP-SSEP, SPC-INTF, MAL-ICAN

Abstract—An important but seldom-discussed problem in the field of blind source separation of real convolutive mixtures is the determination of the role of the demixing filter structure and the criterion/optimization method in limiting separation performance. This issue requires the knowledge of the optimal performance for a given structure, which is unknown for real mixtures. Herein the authors introduce an experimental upper bound on the separation performance for a class of convolutive blind source separation structures, which can be used to approximate the optimal performance. As opposed to a theoretical upper bound the experimental upper bound produces an estimate of the optimal separating parameters for each dataset in addition to specifying an upper bound on separation performance. Estimation of the upper bound involves the application of a supervised learning method to the set of observations found by recording the sources one at a time. Using the upper bound it is demonstrated that structures other than the FIR should be considered for real (convolutive) mixtures, there is still much

room for improvement in current convolutive BSS algorithms, and the separation performance does not appear to be limited by local minima.

Index Terms—Convolutive source separation, BSS, ICA, upper bound, speech enhancement.

I. INTRODUCTION

There are three components of an adaptive filter each of which can limit separation performance of convolutive blind source separation (BSS) methods [1]. First, the demixing filter topology may be insufficient for the given mixture. Second, the criterion used may not be capable of finding the best solution given the demixing filter. Third, the optimization method also affects the solution since it can, among other items, cause convergence to a local minimum. Using the term ‘algorithm’ to denote the combination of the criterion and the optimization method, one can ask to what degree (1) the filter topology or (2) the algorithm limits separation performance. Knowing the answer to these two questions would be very helpful in, *e.g.*, prioritizing research efforts. One way to address these questions requires the estimation of the optimal separation performance for a given filter structure. Notice that BSS criteria cannot be used for this endeavor since they necessarily employ a proxy for separation performance such as correlation [2], [3], kurtosis [4], or mutual information [5].

Using the approach presented here, however, it is possible to approximate the optimal separation performance for a class of demixing filters. This approach is appropriate for any application that involves a convolutive mixture of sources as long as the designer has at least a limited access, as explained in a later section, to the inputs of the mixing channel in question. The list of possible applications includes speech recognition devices, audio enhancement for hearing aids, and multi-user digital and analog communication systems. The ability to approximate the optimal performance allows one to make direct comparisons of the separation performance of different *filter structures* without the added uncertainty of the performance of the BSS *algorithm*. In addition it provides an absolute level of separation performance against which any algorithm can be judged. Relative algorithm performance is easily obtained by comparing different algorithms using the same structure. Using the proposed approach one can go a step

further and determine how closely a given algorithm approaches the optimum performance level for the filter structure under consideration. Unlike theoretical bounds, the experimental upper bound also produces the set of parameters needed to approximate the best performance.

Application of the experimental upper bound to real, convolutive data yields several discoveries, three of which are listed next. First, for convolutive BSS one should consider replacing the ubiquitously used FIR structure with either a Gamma or a Laguerre filter. These two alternative structures, which are briefly reviewed in Section II.B, require fewer adaptable parameters. Consequently, the expected level of misadjustment is reduced, which may, in turn, permit an improvement in separation performance relative to that of the FIR. Second, the separation performance of two recent convolutive BSS algorithms is well below the level of performance that is achievable (for the data under consideration). Third, the separation performance of these algorithms does not appear to be limited by the presence of local minima in the performance surface.

II. CONVOLUTIVE BLIND SOURCE SEPARATION

In the convolutive mixing paradigm a (N_s x 1) vector of unknown sources at time n, $\boldsymbol{s}(n)$, are collected at N_s sensors after having undergone an unknown convolutive mixing represented by $\boldsymbol{H}(z)$, which is a (N_s x N_s) matrix of Z-transforms of the individual mixing filters. The receiver only has access to the (N_s x 1) observation vector, $\boldsymbol{x}(n)$, the individual constituents of which are given by,

$$x_i(n) = \sum_{j=1}^{N_s} h_{ij}(n) * s_j(n) \tag{1}$$

for i = 1, 2, …, N_s, where "*" represents convolution and h_{ij}(n) represents the length-L_h impulse response associated with the ij^{th} entry of $\boldsymbol{H}(z)$.

Demixing can be performed in the frequency-domain [12], [13] or in the time-domain using either the feedforward (FF) [3], [5], [9] or feedback (FB) [2], [14], [15], [16] architecture. The time-domain FF demixing architecture is given by,

$$y_i(n) = \sum_{j=1}^{N_s} w_{ij}(n) * x_j(n) \tag{2}$$

where $y_i(n)$ is the i^{th} output at time n and $w_{ij}(n)$ are the parameters of the demixing filter from the j^{th} sensor to the i^{th} output having L_w adjustable parameters and an effective length of L_{w*} samples (for an FIR filter $L_w = L_{w*}$). Under the linear model framework the individual demixing filters belong either to the autoregressive (AR) [6], moving average (MA, also known as FIR) [6], or autoregressive-moving average (ARMA) [6] models, or to generalized feedforward structures such as the Laguerre and Gamma [7]. Notice that the acronyms FF and FB are used here to refer to the multiple-input, multiple-output (MIMO) global connectivity of the demixing filters irrespective of their local structure, which can be (locally) feedforward or feedback. For the purposes here the most important difference between these two types of systems is that overall stability can be guaranteed for FF systems whenever the local filters are stable, whereas local stability is insufficient to guarantee overall stability for FB systems. In simple terms, the goal of BSS is to adjust the parameters of $w_{ij}(n)$ such that each output, $y_i(n)$, approximates a single source, $s_j(n)$.

Suppose there are $N_s = 2$ sources and observations and that the demixing structure is, without loss of generality, restricted to have z^{-L} terms on the main diagonal. In this case, the two separating solutions are,

$$\boldsymbol{W}(z) = \begin{bmatrix} z^{-L} & \dfrac{-H_{12}(z)z^{-L}}{H_{22}(z)} \\ \dfrac{-H_{21}(z)z^{-L}}{H_{11}(z)} & z^{-L} \end{bmatrix} \quad \textit{and} \quad \boldsymbol{W}(z) = \begin{bmatrix} z^{-L} & \dfrac{-H_{11}(z)z^{-L}}{H_{21}(z)} \\ \dfrac{-H_{22}(z)z^{-L}}{H_{12}(z)} & z^{-L} \end{bmatrix} \tag{3}$$

for the non-permuted and permuted solutions, respectively, where $L \geq 0$ is a user-defined parameter that is needed to implement acausal solutions. Since $\boldsymbol{Y}(z) = \boldsymbol{W}(z)\boldsymbol{X}(z) = \boldsymbol{W}(z)\boldsymbol{H}(z)\boldsymbol{S}(z)$, Equation (3) may be verified by noticing that $\boldsymbol{G}(z) = \boldsymbol{W}(z)\boldsymbol{H}(z)$ at the separating solution must be either a diagonal or anti-diagonal matrix.

The upper bound is used below to compare the class of FF structures that incorporate generalized feedforward filters, which include the FIR, Gamma, and Laguerre filter. As a brief reminder the Gamma and Laguerre filters are IIR and have a memory depth, L_{w*}, that is a function of the single feedback parameter, μ. Despite being IIR both are considered to be generalized feedforward filters since the adaptable coefficients (assuming μ is fixed) determine the locations of only the zeros of the transfer

function. The μ parameter can take any value between 0 and 2 and, for either filter, $\mu = 1$ corresponds to an FIR filter. To increase the memory depth the designer can either increase the filter order, L_w, or he/she can make an appropriate change in the value of μ. Changing μ governs the trade-off between memory depth and resolution, the latter of which is expressed in units of taps per sample. Table I shows how the memory depth and the resolution are affected by L_w and μ for each of these filters. The memory depth for the Gamma filter is from Principe *et al.* [7] and the value for the Laguerre filter is an approximation of the minimum length needed to capture 90% of the total energy contained in the impulse response. Notice that the memory depth of the FIR filter, unlike the other two, is directly coupled to the number of adjustable parameters.

Several characteristics of the Gamma and Laguerre are important for this discussion. The output of each, $v(n)$, is a weighted summation of the signals produced by passing the input, $u(n)$, through $L_w - 1$ generalized delay functions,

$$v(n) = \sum_{i=1}^{L_w} a_i u_i(n)$$

$$\begin{aligned} &\textit{for the Gamma}: && u_1(n) = u(n) \\ & && u_i(n) = (1-\mu)u_i(n-1) + \mu\, u_{i-1}(n-1) \quad (i = 2, \cdots, L_w) \end{aligned} \tag{4}$$

$$\begin{aligned} &\textit{for the Laguerre}: && u_1(n) = u(n) \\ & && u_2(n) = (1-\mu)u_2(n-1) + \mu\, u_1(n-1) \\ & && u_i(n) = (1-\mu)u_i(n-1) + (\mu-1)u_{i-1}(n) + u_{i-1}(n-1) \quad (i = 3, \cdots, L_w) \end{aligned}$$

The Laguerre uses an identical all pass section for each generalized delay function except for the very first one, which is a low pass filter. The Gamma also uses identical generalized delay functions. The difference is that the Gamma delays are all either low pass filters (for $0 < \mu < 1$) or high pass filters (for $1 < \mu < 2$). This is critical since repeatedly applying either a low pass or a high pass filter causes the input correlation matrix to become ill conditioned as i becomes large, such as occurs for large L_w. The value of L_w for which this becomes problematic depends on μ as shown in Figure 1. This figure shows a contour plot of c, the log of the condition number of $\boldsymbol{R}_{\boldsymbol{u}}$, where $\boldsymbol{u}(n) = [u_1(n)\ u_2(n)\ \ldots\ u_{Lw}(n)]^{\mathrm{T}}$ and $\boldsymbol{R}_{\boldsymbol{u}}$ =

$E[\boldsymbol{u}(n)\boldsymbol{u}(n)^T]$ is the correlation matrix of $\boldsymbol{u}(n)$ (the data used to generate this plot are the same data used in the first experiment described in Section V). Notice that this definition of $\boldsymbol{R}_{\boldsymbol{u}}$ becomes the usual auto-correlation matrix only when $\mu = 1$, in which case $\boldsymbol{u}(n) = [u(n)\ u(n\text{-}1)\ \ldots\ u(n\text{-}L_w\text{+}1)]^T$. Values of $c \leq$ -16 (the particular value depends on the machine precision) indicate that $\boldsymbol{R}_{\boldsymbol{u}}$ is nearly singular. The condition of $\boldsymbol{R}_{\boldsymbol{u}}$ is important for the proposed upper bound since it requires estimation of $\boldsymbol{R}_{\boldsymbol{u}}^{-1}$; therefore, some form of matrix conditioning should be used whenever $c \leq$ -16. The condition of $\boldsymbol{R}_{\boldsymbol{u}}$ is also relevant for BSS algorithms that do not require $\boldsymbol{R}_{\boldsymbol{u}}^{-1}$ since ill conditioning is indicative of a large eigenvalue spread, which causes unacceptably slow convergence of the adaptive parameters [6]. Notice that for $\mu = 1$, which corresponds to an FIR, $\boldsymbol{R}_{\boldsymbol{u}}$ is well-conditioned for all values of L_w shown. However, for the Gamma filter values of $\mu \leq 0.5$ require that L_w be constrained to be less than or equal to 25. The Laguerre filter does not suffer from this matrix-conditioning problem since it uses an all pass filter and it is the preferred topology when longer filters are needed. The only papers known to the authors that consider a generalized feedforward filter for convolutive blind source separation are a paper by the authors [5] and one written by Stanacevic *et al*. [11].

III. FIGURES OF MERIT FOR SEPARATION

The proposed upper bound requires a measurable quantity of separation performance of real convolutive mixtures. Numerous methods have been used in the published literature to indicate separation performance. The list includes MSE [20], [21] bit/symbol error rate [22], [23], Frobenius distance [24], multi-channel row and multi-channel column ISI [25], plot of global mixing filter responses [26], [27], [28], SNR [29], [30], [31], ISR [19], SIR [32], [33], [34], [35], [36], [37], [38], one-at-a-time SIR [8], ISI [39], [40], [41], bias and standard deviation of filter coefficients [42], [43], plot of estimated sources in the time or frequency domain [44], [45], [46], hand-segmented SIR [47], automatic speech recognition rate [48], [33], [34], and the mean opinion score [18]. Several of these are not ideal for comparisons because they are either subjective, such as the plots and the mean opinion score, or require knowledge of the mixing filters, which makes them inapplicable for real mixtures. Several others are not desirable for speech because they give preference to BSS methods that temporally whiten the estimated sources, such

as the ISI-based measures, or are more suitable for digital communications. The automatic speech recognition is an excellent measure, but it requires that the sources are speech signals and it requires a large amount of training data and time. For real recordings two of the more interesting methods are the one-at-a-time SIR (signal-to-interference ratio) and the hand-segmented SIR. The latter method can be used when all the sources are recorded at the same time, but it requires that the sources must not overlap significantly in the time domain. The one-at-a-time SIR has no restrictions on the temporal overlap of the sources, but it requires that each source be recorded separately. The drawback for this approach is that each source will have a different background noise if recordings are not made in a silent environment such as a soundproof room.

Schobben *et al*. [8] introduced the one-at-a-time SIR performance metric and suggested its use for real recordings. Since sound waves are additive, the separately recorded sources can be summed together to produce proper mixtures which may then be presented to any BSS algorithm. For example, if there are N_s = 2 sources and sensors then recording each source separately produces two sets of two-sensor observations. To "create" the mixtures, the two signals occurring at the first sensor, one due to the first source and one due to the second, are summed together. This should also be done for the two signals occurring at the second sensor. The SIR is easily estimated by passing the set of observations due to a single source, one at a time, through the previously optimized demixing filters and measuring the power of that source in each output. More formally, the one-at-a-time SIR, hereafter referred to as simply SIR, is defined as,

$$SIR = \arg\max_{\boldsymbol{k}} \frac{1}{N_s} \sum_{i=1}^{N_s} 10 \log_{10} \left(\frac{P_{y(i)|s(k_i)}}{P_{y(i)} - P_{y(i)|s(k_i)}} \right) \quad \text{(dB)} \qquad (5)$$

where k_i, for i = 1, 2, .., N_s, is an element of {1, 2, ..., N_s}, k_i not equal to k_j for i not equal to j, $P_{y(i)|s(k_i)}$ is the power of output i due only to source k_i, (*e.g.*, the power of $(h_{11}(n\text{-}L) + h_{21}(n)*w_{12}(n))*s_1(n)$ when i = 1, k_i = 1 and N_s = 2), and $P_{y(i)}$ is the total power of output i. The N_s-factorial set of k_i terms, $\boldsymbol{k}$, are determined by assuming a particular permutation of the output signals with respect to the original sources. Since the

order of the outputs is immaterial for separation quality (and represents an indeterminacy for BSS), the $\boldsymbol{k}$ that maximizes Equation (5) produces the SIR of interest. Values of SIR above 15 dB for convolutive mixtures are indicative of fairly good separation. While there is no single perfect measure the one-at-a-time SIR is one of the most promising for objective general-purpose comparisons. Moreover, recording sources in this fashion allows the approximation of the upper bound of separation performance for a given mixture, as discussed next.

IV. EXPERIMENTAL (APPROXIMATE) UPPER BOUND

For $N_s = 2$ the signals that are produced when recording one source at a time are $h_{11}(n)*s_1(n)$, $h_{21}(n)*s_1(n)$, $h_{12}(n)*s_2(n)$, and $h_{22}(n)*s_2(n)$. Suppose that we have an FIR adaptive filter and that $-h_{22}(h)*s_2(n)$ is used as the input and $h_{12}(n)*s_2(n)$ is delayed by L and used as the desired signal. This is identical to a system identification problem in adaptive filter theory where the desired transfer function is $H_{12}(z)z^{-L}/H_{22}(z)$. Notice that the desired transfer function equals the optimal (non-permuted) solution for $w_{12}(n)$ as previously specified in Equation (3). Figure 2 shows how to use supervised training to estimate both filters of $\boldsymbol{W}(z)$ for either permutation, where the desired permutation is the one that results in a larger value of SIR. The parameters are found using the mean square error criterion, which is guaranteed to yield a performance surface free of local minima when used with an FIR structure [1]. Furthermore, while this supervised learning method does not maximize SIR, it does minimize the power of the interfering signal, which is the denominator of the SIR. Verification of this claim is easily obtained by noting that the error signal of the system identification is precisely the appropriate interfering signal of the BSS system (and recalling that the performance surface is free of local minima). For example, when estimating $w_{12}(n)$ for the case of no permutation the error signal of the system identification is $(w_{12}(n)*h_{22}(n) + h_{12}(n-L))*s_2(n)$, which equals $g_{12}(n)*s_2(n)$, where $g_{ij}(n)$ is the impulse response associated with the ij^{th} entry of $\boldsymbol{G}(z)$.

This approach is easily extended for $N_s > 2$, although a numerical solution is required since the filters in each row of $\boldsymbol{W}(z)$ must be determined simultaneously. Furthermore, this approach can be used to estimate the demixing parameters for the FF/FIR for any combination of acausality parameter, L, and filter length,

L_w, and it can be used with structures such as the FF/Gamma and the FF/Laguerre. While this approach is not applicable for global FB structures, it may be used with any type of FF structure. The Gamma and the Laguerre have the advantage that, similar to the FF/FIR, the performance surface of both is guaranteed to be free of local minima (when the feedback parameter is fixed) and it is trivial to guarantee stable operation (unlike AR filters).

The experimental upper bound does not replace the need for BSS. Instead, it is meant to be used as a tool during the design process. In order for it to be viable the application in question must allow the possibility to record one signal at a time through the channel. Note that this recording need only be done once and is done off-line so that continual access to the individual inputs of the channel is not necessary. Since the experimental upper bound produces results that are specific to a given dataset the set of one-at-a-time recordings should include all combinations of signals and environments that are expected to occur frequently. Expected background noise sources should also be recorded one at a time as they may be treated as additional sources for the purposes of BSS. It is not possible to treat certain types of noise separately, such as sensor noise and model error. When these are present and non-negligible Equation (1) should be changed as outlined in the appendix. The main shortcoming of the upper bound is that it is pessimistic when the source spectra do not significantly overlap. This occurs since only N_s-1 sources are present in each system identification task; hence, it is not able to take into account spectral diversity among all sources.

Before proceeding the validity of the proposed upper bound is checked using a synthetic mixture for two values of L_h. For these examples an FF/FIR structure is used for both the mixing and demixing with $N_s = 2$, $H_{ii}(z) = 1$, and $L_w = L_h$. For this particular choice the optimal solution has an SIR of infinity. The SIR that results after optimizing the demixing parameters for $L_h = 5$ is, in fact, infinite. This indicates that the magnitude of the error is smaller than the precision of the machine. Likewise, the result for $L_h = 4000$ is 51.6 dB. The input SIR, measured before the demixing is applied, is 1.5 dB for $L_h = 5$ and 0.7 dB for $L_h = 4000$. Typical results obtained using a BSS algorithm for synthetic mixtures are on the order of 20-40 dB for $L_h = 5$ (depending on the BSS algorithm), whereas results for $L_h = 4000$ are on the order of 0-5 dB.

V. RESULTS

Binaural recordings were collected in a 2.5 m by 3.2 m by 2.4 m double-walled soundproof room. This room has a 0.1 s reverberation time, defined as the time for the mean-squared sound pressure of a 1 kHz signal to drop 60 dB. The sources (pre-recorded speech from a male and a female speaker) were played over a single loudspeaker, one at a time, and collected at a distance of 2.3 m by a pair of Audio Technica AT853 miniature condenser microphones having a cardiod pickup pattern. The microphones were placed inside synthetic ears made by Knowles Electronics, which were placed on either side of a dummy head of width 0.14 m. The data were recorded at f_s = 44.1 kHz for N = 120k samples. By rotating the stand on which the dummy head was attached, recordings were collected at 0^o to 360^o azimuth in increments of 10^o. Recordings were also taken at 45^o, 135^o, 225^o, and 315^o. The entire dataset is available on the Internet at http://www.cnel.ufl.edu/itl.html by selecting the "Data" link. For the following results the data are downsampled to 11.025 kHz after using an appropriate anti-aliasing filter.

Results are shown for the experimental upper bound and two different BSS algorithms, *i.e.*, JBD and MRMI-SIG. These two were chosen because they use a FF/FIR structure and because they outperformed numerous convolutive BSS methods in a previous comparison, where MRMI-SIG performed the best for noiseless synthetic mixtures and JBD performed the best for real mixtures [10]. MRMI-SIG has the additional feature that it is easily modified to incorporate other FF structures, whereas JBD is suited only for the FF/FIR since the parameters of each filter are found by truncating the associated impulse response.

Parra *et al.* introduced JBD [9], which uses a frequency-domain criterion that minimizes the cross power spectra. This method uses second-order statistics and requires that the sources are non-stationary. Using a frequency-domain criterion for convolutive BSS is advantageous in that convolution in the time-domain becomes multiplication in the frequency-domain. Consequently, the problem reduces to numerous instantaneous demixing problems (one for each frequency bin), each of which is very simple. The disadvantage is that each frequency bin has an unknown permutation (this may also be true for time-domain methods [32]). Perfect separation in each frequency bin equates to perfect separation only if the permutation ambiguity is resolved. JBD computes the demixing filter in the frequency-domain, converts

the filter to the time-domain using the inverse transform, and then truncates the resulting time-domain filter that is then used to produce the final source estimates. The truncation forces the solutions to be smooth in the frequency domain, thus coupling the information between the different frequency bins. Since arbitrary permutations require filters that are longer than what is available due to the truncation, the convergence of the algorithm tends to avoid the permutation ambiguity [9].

The authors introduced MRMI-SIG [5], which minimizes an approximation of Renyi's mutual information between length-L_t segments of the estimated sources. The mutual information is approximated by summing the N_s marginal Renyi entropies and subtracting the (single) joint Renyi entropy. For convolutive mixtures MRMI-SIG requires that each of the sources has a super-Gaussian distribution, which is normally the case for speech signals. Experimental evidence indicates that the set of parameters that minimize/maximize Renyi's entropy are very nearly the same as those that minimize/maximize Shannon's entropy when the sources are super-Gaussian. To the extent that this approximation holds MRMI-SIG minimizes the (Shannon) mutual information [50] between segments of the estimated sources, which can be shown to be a contrast as long as the user-defined segment-length is chosen sufficiently large [49]. It is not known, however, whether Shannon's mutual information is discriminating (there may be local minima). MRMI-SIG is an information-theoretic approach that operates entirely in the time-domain. This makes it more computationally intensive than JBD, especially as L_w becomes large or μ approaches 0.

The solutions for the experimental upper bound are found using the well-known Wiener-Hopf solution, $\boldsymbol{R}_{\boldsymbol{u}}^{-1}\boldsymbol{p}_{\boldsymbol{u}d}$ [1], where $\boldsymbol{R}_{\boldsymbol{u}}$ is the correlation matrix of $\boldsymbol{u}(n)$ (as defined previously), $\boldsymbol{p}_{\boldsymbol{u}d} = \mathrm{E}[\boldsymbol{u}(n)d(n)]$ is the cross-correlation vector between the input vector, $\boldsymbol{u}(n)$, and the desired signal, $d(n)$, and both $u(n)$ and $d(n)$ are as shown in Figure 2. The temporal correlation matrix, $\boldsymbol{R}_{\boldsymbol{u}}$, should not be confused with the spatial auto-correlation matrix that is commonly used to sphere the observations in BSS. For the JBD algorithm the length of the FFT is varied from 256 to 16384. The values for the other user-defined parameters are kept as close to the default values as

possible, although they had to be reduced for large FFT-lengths. The results shown are for the FFT-length that produced the highest SIR. The user-defined parameters for MRMI-SIG are $L_t = 2L_{w^*} - 1$ and $\sigma = 1$, where σ is the kernel size. The acausality parameter in all cases is fixed at $L = 0$. There are three important parameters for each filter topology, L_w, L_{w^*}, and μ, but because of the relationship between them previously given in Table I only two of these need to be given to completely specify the parameters of the FF/Gamma and FF/Laguerre and only one is needed to uniquely specify the FF/FIR.

For the first experiment the loudspeaker is oriented at 45^o azimuth when rendering the female speech, $s_2(n)$, and at 0^o azimuth when rendering the male speech, $s_1(n)$, where a value of 0^o corresponds to the location directly in front of the dummy head. Figures 3 and 4 show how the performance of the upper bound for the FF/Laguerre varies with L_w and μ. For these two figures the input SIR is 4.5 dB. Notice in Figure 3 that the FF/Laguerre (for this data) always outperforms the FF/FIR as a function of the number of adaptable parameters. The fact that the SIR improves for the FF/Laguerre as μ is decreased from 1 to 0.25 indicates that increasing the memory depth for this dataset is more important to performance than maintaining a high resolution.

Figure 4 shows how the FF/Gamma compares to the FF/Laguerre and FF/FIR as a function of memory depth. The FF/Gamma uses matrix regularization whenever its inclusion improves the SIR. In this figure $\mu = 0.5$ for the FF/Laguerre and μ for the FF/Gamma is chosen so that the filter length of these two topologies is equal, whereas the filter length of the FF/FIR is roughly three times that of the other two. Notice that the FF/Laguerre performance nearly equals that of the FF/FIR despite having only 1/3 the number of adaptable parameters. A paper by Stanacevic *et al.* also makes this claim [11]. The measure of performance used in Stanacevic's paper was the fit of the impulse response of the FIR and Laguerre filters to an estimated room impulse response. The results here substantiate this claim using a much more rigorous basis, namely by using a direct measure of separation

performance on real convolutive mixtures. Figure 4 also shows that the separation performance of all three topologies is comparable for small memory depths and that the eigenvalue spread limits the performance of the FF/Gamma to approximately 12.5 dB.

Besides determining the preferred filter topology in an algorithmic-independent fashion, the experimental upper bound is also useful for providing an absolute measure of algorithm performance. For example, comparing the JBD performance in Figure 4 with the upper bound that uses the same topology indicates that JBD is performing well for small memory depths. However, the vertical line in this figure indicates how much room there is for improvement as L_{w*} increases. A similar judgment can be made for MRMI-SIG using the results shown in Table II (due to computational complexity the results for MRMI-SIG are restricted to a subset of L_{w*}, also, due to differing filter lengths care should be taken when using this table to compare topologies). Similar to JBD the performance of MRMI-SIG for all three topologies is near the upper bound when the memory depth is 25, but then falls off rapidly as the memory depth is increased. For this figure the SIR prior to demixing is the same as before, *i.e.*, 4.5 dB.

Results of the second experiment are shown in Figures 5 and 6 as the location of $s_2(n)$ is changed to -90^o, -45^o, 0^o, 45^o, and 90^o azimuth, while $s_1(n)$ remains fixed at 0^o azimuth. A plot of the input SIR is shown since each data point in these two figures represents a different mixture. In fact, when the second source is located at 0^o azimuth the two sources are collocated so that the mixing matrix is not full rank. However, it is still possible to produce some degree of separation due to spectral differences of the speech signals. It is expected that the result at 0^o would noticeably drop if the sources perfectly overlapped in the frequency domain. The values of $L_{w*} = 250$ for Figure 5 and $L_{w*} = 25$ for Figure 6 are used since these values represent the best performing memory depth for the JBD and MRMI-SIG algorithms, respectively. Also, as was done in Table II, FF/Gamma uses $\mu = 0.3$ and FF/Laguerre uses $\mu = 0.5$. As such Figures 5 and 6 are more appropriate for determination of absolute algorithm performance than for filter topology comparison. Even though the memory depth in both figures is optimized for the respective BSS algorithm, there is a noticeable difference between how well the two BSS algorithms performed

compared to how well they could perform. Keep in mind that this difference would be much larger if the memory depth were optimized for the upper bounds.

An additional set of experiments was conducted in order to assess bias in the expected SIR performance. In the previous two experiments the demixing parameters of JBD and MRMI-SIG were initialized randomly. To estimate performance bias both algorithms were also adapted after initializing the demixing parameters with the solution found using the experimental upper bound. Estimating the performance bias in this way is valid to the extent that the solution using the upper bound is near the optimal solution. The results are not shown, but the difference in SIR of the solutions found with this initialization and with random initialization was negligible in almost every case. Hence, the results reported here are indicative of the bias introduced by each algorithm and are not simply due to limitations imposed by local minima of the respective performance surface.

VI. CONCLUSIONS

The experimental upper bound reveals how much the structure and the algorithm limit separation performance. For the data used in these experiments the limitation of the FF/Gamma (with μ = 0.3) relative to the FF/FIR filter structure, *e.g.*, is roughly 1.5 dB for L_{w*} = 100. Likewise, the MRMI-SIG algorithm in this case further limits the performance by approximately 11.9 – 5.3 = 6.6 dB. Figure 6 indicates that MRMI-SIG with the FF/Gamma structure produces a result that is only slightly above the input SIR for three of the five mixtures and is very nearly equal to the experimental upper bound for the remaining two. This type of information is invaluable for finding and eliminating weaknesses of any given BSS algorithm. Without the experimental upper bound one only has the knowledge of relative separation performance, which can lead to inaccurate conclusions, *e.g.*, when an algorithm performs better for dataset A than for dataset B although the performance for A is well below the upper bound and the performance for B nearly equals the upper bound. Many other inferences are also possible using the proposed experimental upper bound. As a final example, it could be argued from Figure 4 that the FF/FIR and the FF/Laguerre structures produce asymptotically identical results so that there is no need to consider any structure other than the FF/FIR, which is used almost exclusively for time-domain demixing.

However, Figure 4 and Table II also indicate that the two BSS algorithms suffer from misadjustment. The consequence for practical BSS algorithms is that L_w should be kept reasonably small. Based on the data used here small values of L_w tend to favor the use of the FF/Gamma (for $L_{w^*} < 100$) or the FF/Laguerre structure over the FF/FIR.

In conclusion, the one-at-a-time SIR allows for an objective and unambiguous comparison of any linear demixing BSS method, including FB structures. Moreover, the data collection procedure that it requires makes possible the proposed approach of estimating the experimental upper bound to separation performance of FF structures. This approach is novel not because of an advancement of signal processing technique, but rather due to the recognition that recording the sources one at a time and applying the system identification method known from adaptive filter theory causes the minimization of the power of the interfering signal in the BSS paradigm. The result is a tool that allows for a quick estimation of the optimal separation coefficients and the associated upper bound of the SIR for any given set of (real or synthetic) mixtures. The approximate upper bound is not subject to local minima or local permutations, it uses supervised learning, it minimizes the power of the interfering signal in each output, it provides an estimate of performance bias, it can be used to evaluate the difficulty of a given real (convolutive) data set, it illustrates the advantage of using a Gamma or Laguerre filter topology, it allows quick estimation of the optimal values for the feedback parameter, filter length, and acausality parameter for one or more representative sets of data, and it exposes just how poorly two recent BSS algorithms (including the authors') perform for real mixtures.

APPENDIX

For sake of explication the previous discussion of the experimental upper bound ignores the presence of noise and model error. This is justifiable in certain conditions, *e.g.*, for audio applications as long as the recordings are made in a soundproof room, the transfer functions of the microphones and the amplifiers are nearly linear, and the signal power as measured at each transceiver is much greater than the power of the corresponding sensor noise. To the extent possible background noises should be treated as additional

sources which are to be recorded separately. To account for all other types of non-negligible perturbations Equation (1) should be modified to the following,

$$x_i(n) = \sum_{j=1}^{N_s} h_{ij}(n) * s_j(n) + e_i(n) \tag{6}$$

where $e_i(n)$ represents noise and model error at the i^{th} sensor. The separating solutions are still given by Equation (3) since the concern in BSS is to separate the sources from each other without regard to contamination by $e_i(n)$. However, when $e_i(n)$ is not zero for all i and all n, the Wiener-Hopf solution using the available signals can no longer be guaranteed to converge asymptotically to the separating solution. Suppose once again that there are $N_s = 2$ sources and observations. Recording the sources one at a time produces the signals $h_{11}(n)*s_1(n) + e_{11}(n)$, $h_{21}(n)*s_1(n) + e_{21}(n)$, $h_{12}(n)*s_2(n) + e_{12}(n)$, and $h_{22}(n)*s_2(n) + e_{22}(n)$, where two subscripts are used for each $e(n)$ since the recordings of $s_1(n)$ and $s_2(n)$ are made at different times. For consistency with Equation (6) it is necessary that $e_i(n) = e_{ii}(n) + e_{ij}(n)$, where i does not equal j. These four signals are used as $u(n)$ and $d(n)$ as previously shown in Figure 2. When it is reasonable to assume that $e_{ij}(n)$ is not correlated with $s_i(n)$ for all combinations of i, j, the $e_{ij}(n)$ term occurring in $d(n)$ has no effect on determining the Wiener-Hopf solution of the demixing filter [1]. The $e_{ij}(n)$ term occurring in $u(n)$, on the other hand, always biases the solution. In order to mitigate the bias the Wiener-Hopf solution should be replaced with the Error Whitening Criterion (EWC) [17]. Equation (5), which in the present case refers to the signal-to-interference/noise ratio (SINR), remains the same with the understanding that a portion of the output power, $P_{y(i)}$, is due to $e(k_i)$.

ACKNOWLEDGEMENTS

Work partially supported by NSF ECS #0300340.

REFERENCES

[1] M.G. Larimore, C.R. Johnson, and J.R. Treichler, *Theory and Design of Adaptive Filters*, Pearson Education, Harlow, UK, 2001.

[2] S. Van Gerven and D. Van Compernolle, “Signal separation by symmetric adaptive decorrelation: stability, convergence, and uniqueness,” *IEEE Trans. Signal Proc.*, Vol. 43, pp. 1602-1612, July 1995.

[3] L. Parra, C. Spence, and B. De Vries, “Convolutive blind source separation based on multiple decorrelation,” *Neural Networks for Signal Proc.*, Cambridge, pp. 23-32, Aug. 1998.

[4] C. Simon, Ph. Loubaton, C. Vignat, C. Jutten, and G. d'Urso, “Blind source separation of convolutive mixtures by maximization of fourth-order cumulants: The non-IID case,” *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, Vol. 2, pp. 1584-1588, Nov. 1998.

[5] K.E. Hild II, D. Pinto, D. Erdogmus, and J.C. Principe, “Convolutive blind source separation by minimizing mutual information between segments of signals,” *Submitted to IEEE Trans. Circuits and Systems I*, June 2004.

[6] S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, Englewood Cliffs, NJ, 2001.

[7] J. Principe, N. Euliano, and W. Lefabvre, *Neural and Adaptive Systems*, John Wiley & Sons, NYC, 1999.

[8] D. Schobben, K. Torkkola and P. Smaragdis, “Evaluation of blind signal separation methods,” *Intl. Workshop on Independent Component Analysis and Signal Separation*, Aussois, pp. 261-266, Jan. 1999.

[9] L. Parra, C. Spence, “Convolutive blind source separation of non-stationary sources,” *IEEE Trans. Speech and Audio Processing*, pp.320-327, May 2000.

[10] K.E. Hild II, “Blind Separation of Convolutive Mixtures Using Renyi's Divergence,” Ph.D. Dissertation, The University of Florida, Nov. 2003.

[11] M. Stanacevic, M. Cohen, and G. Cauwenberghs, “Blind separation of linear convolutive mixtures using orthogonal filter banks,” *Intl. Workshop on Independent Component Analysis*, San Diego, pp. 260-265, Dec. 2001.

[12] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, Vol. 22, No. 1-3, pp. 21-34, Nov. 1998.

[13] C. Mejuto, A. Dapena, and L. Castedo, “Frequency-domain Infomax for blind separation of convolutive mixtures,” *Intl. Workshop on Independent Component Analysis*, Helsinki, pp. 315-320, June 2000.

[14] N.A. Kanlis, J.Z. Simon, and S.A. Shamma, “Complete training analysis of feedback architecture networks that perform blind source separation and deconvolution,” *Intl. Workshop on Independent Component Analysis*, Helsinki, pp. 139-144, June 2000.

[15] N. Charkani and Y. Deville, “Self-adaptive separation of convolutively mixed signals with a recursive structure. Part I: Stability analysis and optimization of asymptotic behaviour,” *Signal Processing*, Vol. 73, pp. 225-254, Jan. 1999.

[16] N. Charkani and Y. Deville, “Self-adaptive separation of convolutively mixed signals with a recursive structure. Part II: Theoretical extensions and application to synthetic and real signals,” *Signal Processing*, Vol. 75, pp. 117-140, June 1999.

[17] J.C. Principe, Y.N. Rao, D. Erdogmus, “Error Whitening Wiener Filters: Theory and Algorithms,” *Least-Mean-Square Adaptive Filters*, S. Haykin and B. Widrow (eds.), Sept. 2003.

[18] J.I. Sohn and M. Lee, “Selective Noise Cancellation Using Independent Component Analysis,” *Intl. Conf. on Artificial Neural Networks*, Istanbul, pp. 530-537, June 2003.

[19] V. Zarzoso and A.K. Nandi, “Closed-form semi-blind separation of three sources from three real-valued instantaneous linear mixtures via quaternions,” *Intl. Symposium on Signal Processing and its Applications*, Kuala Lumpur, Vol. I, pp. 1-4, Aug. 2001.

[20] D. Yellin and E. Weinstein, “Criteria for multichannel signal separation,” *IEEE Trans. on Signal Proc.*, Vol. 42, pp. 2158-2167, Aug. 1994.

[21] L. De Lathauwer, B. De Moor, and J. Vandewalle, “An algebraic approach to blind MIMO identification,” *Intl. Workshop on Independent Component Analysis*, Helsinki, pp. 211-214, June 2000.

[22] S. Icart and R. Gautier, “Blind separation of convolutive mixtures using second and fourth order moments,” *Intl. Conf. Acoustics, Speech, and Signal Proc.*, Atlanta, Vol. 5, pp. 3018-3021, May 1996.

[23] D. Gesbert, P. Duhamel, and S. Mayrargue, “On-line blind multichannel equalization based on mutually referenced filters,” *IEEE Trans. on Signal Proc.*, Vol. 45, pp. 2307-2317, Sept. 1997.

[24] P. Comon, E. Moreau, and L. Rota, “Blind separation of convolutive mixtures, a contrast-based joint diagonalization approach,” *Intl. Workshop on Independent Component Analysis*, San Diego, pp. 686-691, Dec. 2001.

[25] R.H. Lambert, “Difficulty measures and figures of merit for source separation,” *Intl. Workshop on Independent Component Analysis*, Aussois, pp. 133-138, Jan. 1999.

[26] M. Kawamoto, A.K. Barros, K. Matsuoka, and N. Ohnishi, “A method of real-world separation implemented in frequency domain,” *Intl. Workshop on Independent Component Analysis*, Helsinki, pp. 267-272, June 2000.

[27] A. Mansour, C. Jutten, and P. Loubaton, “Adaptive subspace algorithm for blind separation of independent sources in convolutive mixture,” *IEEE Trans. Signal Proc.*, Vol. 48, pp. 583-586, Feb. 2000.

[28] I. Fijalkow and P. Gaussier, “Self-organizing blind MIMO deconvolution using lateral-inhibition,” *Intl. Workshop on Independent Component Analysis*, Aussois, pp. 221-226, Jan. 1999.

[29] M. Girolami, “Symmetric adaptive maximum likelihood estimation for noise cancellation and signal separation,” *Electronic Letters*, Vol. 33, pp. 1437-1438, Aug. 14, 1997.

[30] H.C. Wu and J.C. Principe, “A unifying criterion for blind source separation and decorrelation: simultaneous diagonalization of correlation matrices,” *Neural Networks for Signal Proc.*, Amelia Island, pp. 496-505, Sept. 1997.

[31] W. Baumann, B.U. Kohler, D. Kolossa, and R. Orglmeister, “Real time separation of convolutive mixtures,” *Intl. Workshop on Independent Component Analysis*, San Diego, pp. 65-69, Dec. 2001.

[32] L.C. Parra and C.V. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” *IEEE Trans. Speech and Audio Proc.*, Vol. 10, pp. 352-362, Sept. 2002.

[33] A. Koutras, E. Dermatas, and G. Kokkinakis, “Blind signal separation and speech recognition in the frequency domain,” *IEEE Intl. Conf. Electronics, Circuits and Systems*, Pafos, Vol. 1, pp. 427-430, Sept. 1999.

[34] A. Koutras, E. Dermatas, and G. Kokkinakis, “Continuous speech recognition in a multi-simultaneous-speaker environment using decorrelation filtering in the frequency domain,” *Intl. Work. Speech and Computer*, St. Petersburg, pp. 253-256, Oct. 1998.

[35] K. Rahbar and J.P. Reilly, “Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices,” *Intl. Conf. Acoustics, Speech, and Signal Processing*, Salt Lake City, Vol. 5, pp. 2745-2748, May 2001.

[36] R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, “Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming,” *Neural Networks for Signal Proc.*, Martigny, pp. 445-454, Sept. 2002.

[37] L. Parra and C. Spence, “On-line convolutive blind source separation of non-stationary signals,” *Journal of VLSI Signal Proc. Systems for Signal Image and Video Tech.*, Vol. 26, pp. 39-46, Aug. 2000.

[38] K. Rahbar and J.P. Reilly, “Blind source separation algorithm for MIMO convolutive mixtures,” *Intl. Workshop on Independent Component Analysis*, San Diego, pp. 242-247, Dec. 2001.

[39] S.C. Douglas and S.Y. Kung, “Kuicnet algorithms for blind deconvolution,” *Neural Networks for Signal Proc.*, Cambridge, pp. 3-12, Aug. 1998.

[40] O. Shalvi and E. Weinstein, “New criteria for blind deconvolution of nonminimum phase systems,” *IEEE Trans. on Information Theory*, Vol. 36, pp. 312-321, March 1990.

[41] M. Kawamoto, Y. Inouye, A. Mansour, and R.W. Liu, “Blind deconvolution algorithms for MIMO-FIR systems driven by fourth-order colored signals,” *Intl. Workshop on Independent Component Analysis*, San Diego, pp. 692-697, Dec. 2001.

[42] D. Yellin and E. Weinstein, “Multichannel signal separation: Methods and analysis,” *IEEE Trans. on Signal Proc.*, Vol. 44, pp. 106-118, Jan. 1996.

[43] U.A. Lindgren, H. Broman, “Source separation using a criterion based on second-order statistics,” *IEEE Trans. on Signal Proc.*, Vol. 46, pp. 1837-1850, July 1998.

[44] S. Ikeda, N. Murata, “A method of ICA in time-frequency domain,” *Intl. Workshop on Independent Component Analysis*, Aussois, pp. 365-370, Jan. 1999.

[45] N. Mitianoudis and M. Davies, “New fixed-point algorithms for convolved mixtures,” *Intl. Workshop on Independent Component Analysis*, San Diego, pp. 633-638, Dec. 2001.

[46] M.S. Brandstein, “On the use of explicit speech modeling in microphone array applications,” *Intl. Conf. Acoustics, Speech, and Signal Processing*, Seattle, Vol. 6, pp. 3613-1616, May 1998.

[47] C.L. Fancourt and L. Parra, “The coherence function in blind source separation of convolutive mixtures of non-stationary signals,” *Neural Networks for Signal Proc.*, Falmouth, pp. 303-312, Sept. 2001.

[48] F. Ehlers, and H.G. Schuster, “Blind separation of convolutive mixtures and an application in automatic speech recognition in a noisy environment,” *IEEE Trans. Signal Proc.*, Vol. 45, pp. 2608-2612, Oct. 1997.

[49] D.T. Pham, “Mutual information approach to blind separation of stationary sources,” *IEEE Trans. on Information Theory*, Vol. 48, pp. 1935-1946, July 2002.

[50] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., NYC, 1991.

Figure 1. Contour of c, the log condition number of the correlation matrix for the Gamma filter as a function of feedback parameter, μ, and filter length, L_w.

Figure 2. System identification formulations for the permuted and non-permuted solutions of the approximate upper bound.

Figure 3. SIR of the experimental upper bound for the FF/Laguerre topology as a function of the filter length, L_w, for different μ.

Figure 4. SIR of the experimental upper bound for the FF/Laguerre topology as a function of the feedback parameter, μ, for different L_w.

Figure 5. SIR as a function of memory depth, L_{w^*}, for the upper bound and for JBD. For the upper bound (FF/Laguerre) $\mu = 0.5$ and for the upper bound (FF/Gamma) μ is such that the filter length of the Gamma and Laguerre is equal (resulting in $0.32 \leq \mu \leq 0.4$).

Figure 6. SIR as a function of azimuth using $L_{w^*} = 250$.

Figure 7. SIR as a function of azimuth using $L_{w^*} = 25$.

TABLE I: Memory depth and resolution as a function of L_w and μ.

TABLE II: SIR values (in dB) for the MRMI-SIG algorithm. The value of the associated upper bound is given in parentheses.

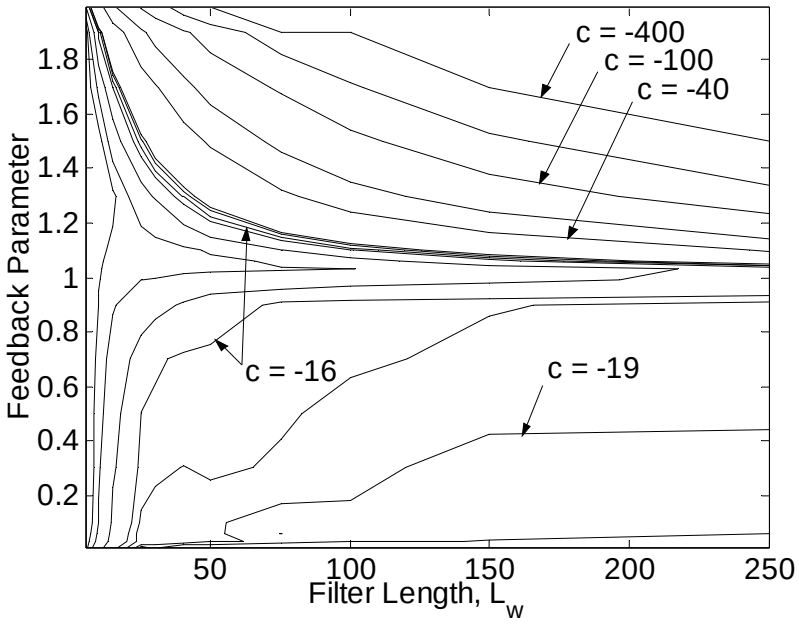
c = -400
c = -100
c = -40
c = -16
c = -19
Feedback Parameter
Filter Length, L_w
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
50
100
150
200
250

Fig. 1

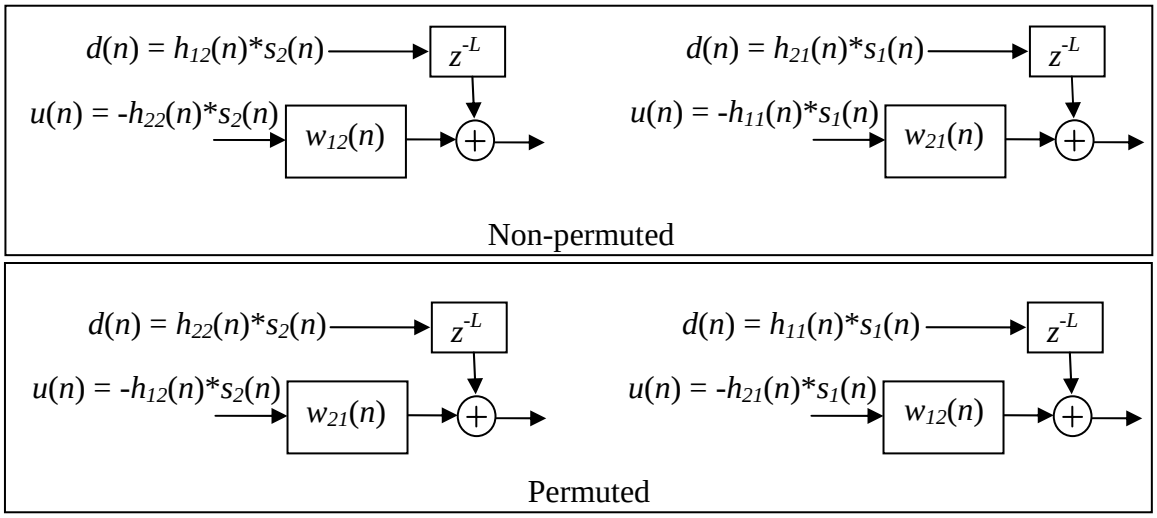
d(n) = h12(n)*s2(n)
z-L
u(n) = -h22(n)*s2(n)
w12(n)
d(n) = h21(n)*s1(n)
z-L
u(n) = -h11(n)*s1(n)
w21(n)
Non-permuted
d(n) = h22(n)*s2(n)
z-L
u(n) = -h12(n)*s2(n)
w21(n)
d(n) = h11(n)*s1(n)
z-L
u(n) = -h21(n)*s1(n)
w12(n)
Permuted

Fig. 2

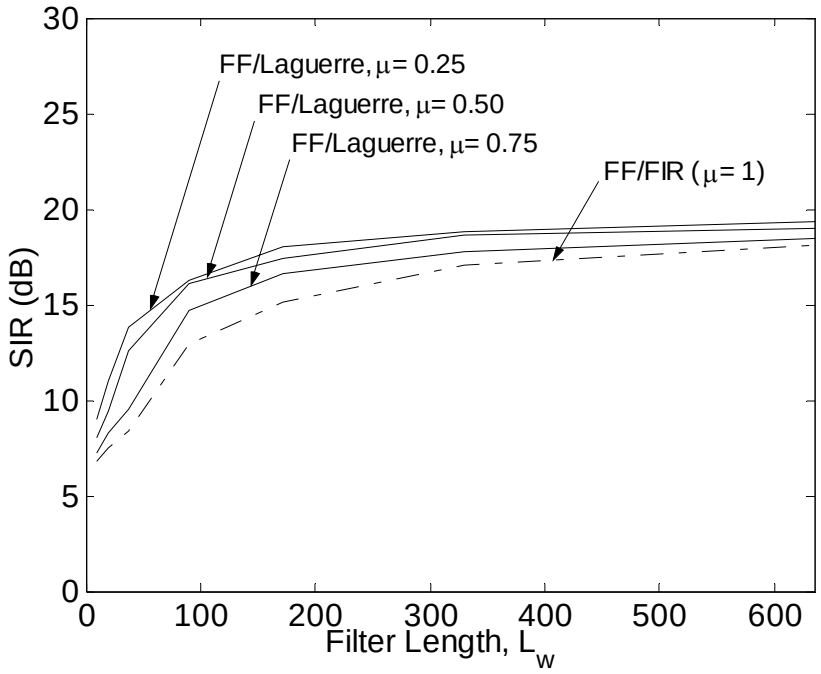
30
25
20
15
10
5
0
SIR (dB)
0
100
200
300
400
500
600
Filter Length, L_w
FF/Laguerre, μ= 0.25
FF/Laguerre, μ= 0.50
FF/Laguerre, μ= 0.75
FF/FIR (μ= 1)

Fig. 3

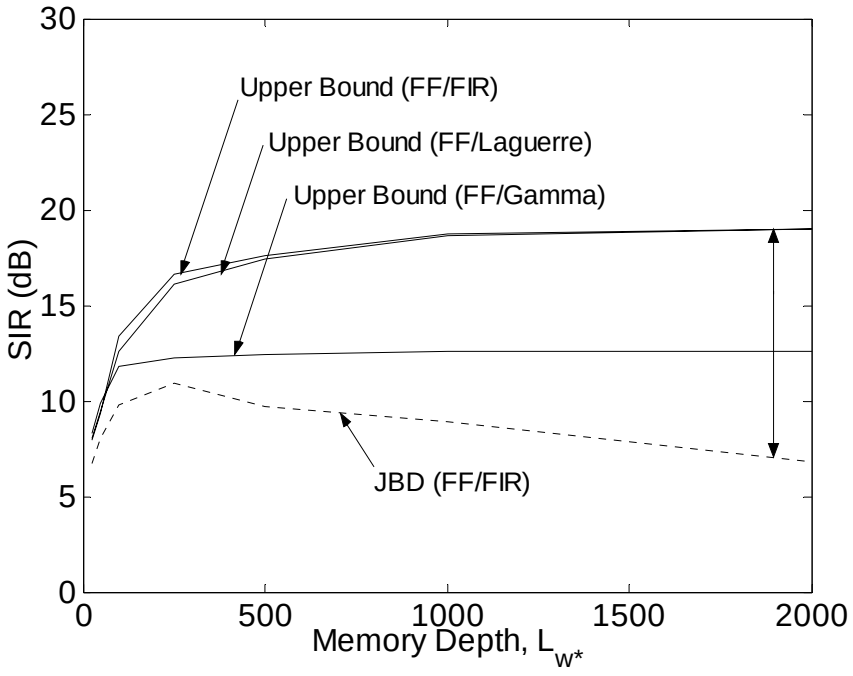
30
25
20
15
10
5
0
SIR (dB)
0
500
1000
1500
2000
Memory Depth, L_{w^*}
Upper Bound (FF/FIR)
Upper Bound (FF/Laguerre)
Upper Bound (FF/Gamma)
JBD (FF/FIR)

Fig. 4

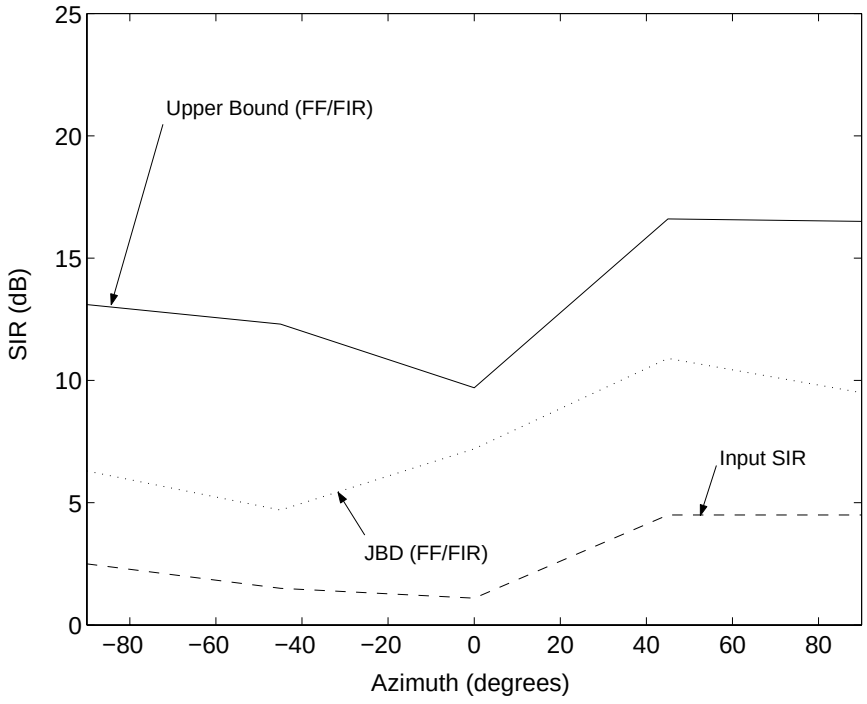
Upper Bound (FF/FIR)
Input SIR
JBD (FF/FIR)
SIR (dB)
Azimuth (degrees)
25
20
15
10
5
0
−80
−60
−40
−20
0
20
40
60
80

Fig. 5

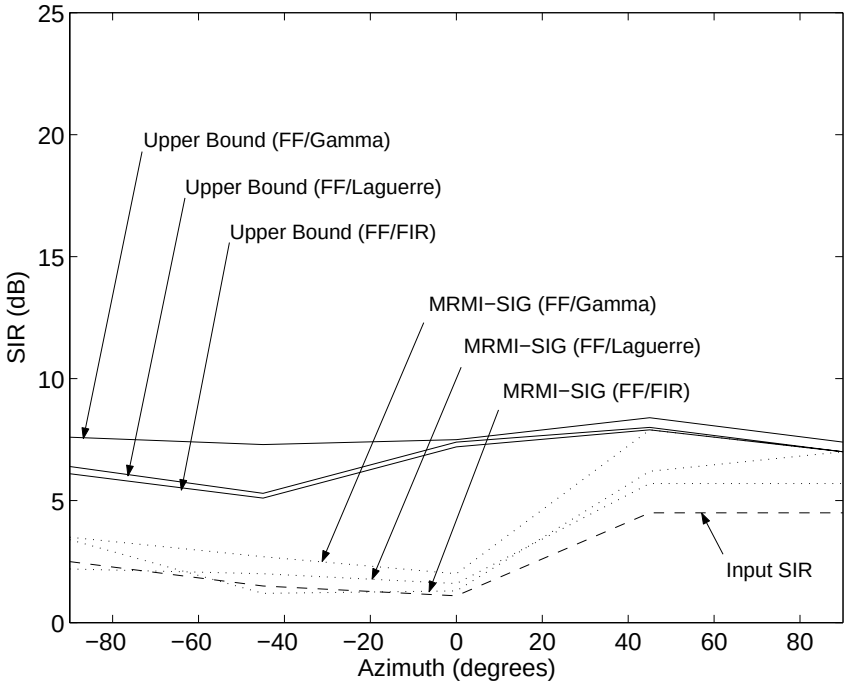
Upper Bound (FF/Gamma)
Upper Bound (FF/Laguerre)
Upper Bound (FF/FIR)
MRMI−SIG (FF/Gamma)
MRMI−SIG (FF/Laguerre)
MRMI−SIG (FF/FIR)
Input SIR
SIR (dB)
Azimuth (degrees)
25
20
15
10
5
0
−80
−60
−40
−20
0
20
40
60
80

Fig. 6

Structure	Memory Depth, L_{w^*}	Resolution
FIR	L_w	1
Gamma	L_w/μ	μ
Laguerre	$(1 + 0.4\|\mu\text{-}1\|\log_{10}L_w)\ L_w/\mu$	$\mu / (1 + 0.4\|\mu\text{-}1\|\log_{10}L_w)$

Table 1.

Memory Depth, L_{w^*}	MRMI-SIG (FF/FIR) $\mu = 1$	MRMI-SIG (FF/Laguerre) $\mu = 0.5$	MRMI-SIG (FF/Gamma) $\mu = 0.3$
25	5.7 dB (7.9)	6.2 dB (8.0)	7.9 dB (8.4)
50	4.3 (9.3)	6.0 (9.4)	5.9 (9.8)
100	4.8 (13.4)	5.7 (12.6)	5.3 (11.9)

Table 2.