

# BLIND SOURCE SEPARATION OF TIME-VARYING, INSTANTANEOUS MIXTURES USING AN ON-LINE ALGORITHM

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

Computational NeuroEngineering Laboratory (www.cnel.ufl.edu)  
The University of Florida, Gainesville, Florida, 32611, USA  
Email: k.hild@ieee.org, deniz@cnel.ufl.edu, principe@cnel.ufl.edu

## ABSTRACT

A blind source separation algorithm is presented that performs on-line separation of an unknown, time-varying, instantaneous mixture of independent sources. This procedure utilizes the information-theoretic MeRMald-SIG criterion and an on-line PCA algorithm, referred to as SIPEX-G, that has been recently submitted for publication. Results indicate that the combination of the two on-line criteria is able to track a rapidly changing mixing matrix. A performance comparison of separation criteria is also made using the same, aforementioned on-line PCA algorithm. Results show the superior performance of the proposed method.

## 1. INTRODUCTION

Much of the research in the field of blind source separation has been focused on the separation of instantaneous (noiseless, over-determined, linear, time-invariant) mixtures. This is not, however, a very realistic model for many real-world mixtures. There are several changes to the (assumed) mixing model that can be made to remedy this problem. Examples include the addition of noise to the model, the possibility that the mixture is under-determined (fewer observations than sources), the use of convolutive and/or non-linear demixing architectures, and finally, the case that the mixture is time-varying. The last item is the one that is addressed herein.

Time-varying mixtures are an especially important topic in the demixing of speech signals for the application of hearing aids. This is due to the fact that the mixing matrix can change very rapidly [1]. The more rapidly the mixing matrix changes, for a fixed sample frequency, the fewer the number of data samples are available to estimate the demixing coefficients. To wit, it is not wise to use data samples at time  $k - N$  (for  $N$  large) for estimating the update to the demixing coefficients at time  $k$  since the statistics of the observations may have changed drastically in the last  $N$  samples. The MeRMald and MeRMald-SIG algorithms have already been shown to be very data efficient [2], [3]; therefore, they appear to be a good candidate for tracking a rapidly changing mixture. Since, in addition, it is desired to produce an on-line method, only the MeRMald-SIG criterion is considered presently.

A previous study, which focused on demonstrating the viability of an on-line information-theoretic criterion, briefly discussed time-varying, instantaneous demixing [3]. In [3], only the rotation of the mixture was varied as a function of time. This greatly simplified the experimental procedure, but is not very realistic since

the sphering of the mixture was always constant. Furthermore, this approach prevented the direct comparison of the MeRMald-SIG algorithm with other methods (that do not constrain the demixing matrix to be a pure rotation). This research aims to extend the results of the previous paper by adding to the MeRMald-SIG algorithm, an on-line procedure for sphering the data. The combination of the two, therefore, is a true on-line procedure and it allows for the comparison of several separation criteria.

## 2. SYSTEM DESCRIPTION

A simple instantaneous mixing model for hearing aid applications is to assume that the sound pressure varies inversely with the distance of the speaker to each sensor (i.e. to each ear). This ignores the fact that the signals at the sensors that are farther from a particular source than a reference sensor are delayed with respect to the reference signal. It also ignores the effect of the head-related transfer function. It does, however, provide a simple instantaneous model that can be easily simulated.

Given this model, the mixing matrix,  $\mathbf{H}_k$ , is found at every time sample (determined by the sampling frequency) for the desired speaker geometry. In addition, the sphered observations,  $\mathbf{x}_k$ , are related to the (zero-mean) sources,  $\mathbf{s}_k$ , by the following,

$$\mathbf{x}_k = \mathbf{W}_k^T \mathbf{H}_k^T \mathbf{s}_k \quad (1)$$

where  $\mathbf{x}_k$  and  $\mathbf{s}_k$  are  $(N \times 1)$  vectors,  $\mathbf{H}_k$  and  $\mathbf{W}_k$  are the  $(N \times N)$  mixing and sphering matrices, respectively, at time  $k$ , and  $T$  is the matrix transpose. For now, it is assumed that there are only two sources (speakers) and two sensors, and that each sensor is located in one of the ears of the observer. As an example, consider the case where the distance between the ears is 1/6 meter and the speakers are located 1 m directly in front of and 2 m directly to the left of the observer. The resulting  $\mathbf{H}_k$  matrix is then,

$$\mathbf{H}_k^T = \begin{bmatrix} \sqrt{144/145} & 12/23 \\ \sqrt{144/145} & 12/25 \end{bmatrix} \quad (2)$$

The number 12/23 is found by taking the inverse of the distance between the second speaker and the left ear of the observer,  $1/(2 -$

### 3. ALGORITHM DESCRIPTION

There are two algorithms that constitute the proposed method, the SIPEX-G algorithm [4] is an on-line criterion for updating the sphering matrix and the MerMaId-SIG algorithm is an on-line method for updating the demixing matrix. Each of these will be described briefly before proceeding.

The SIPEX-G algorithm is introduced in [4]. One of the motivations for this algorithm is that the sphering matrix is a product of an eigenvector matrix, which is orthonormal, and the inverse square root of the corresponding eigenvalue matrix, which is diagonal. Ignoring the diagonal matrix for a moment, it is easy to see that the observations can be made orthogonal to each other using only the orthonormal matrix, which can be implemented using only Givens rotations [5]. In addition, since the sphering matrix is determined using only second-order techniques, the resulting cost function is simply,

$$J = \sum_{i=1}^{N-1} \gamma_i \text{Var}(y_i) \quad (4)$$

where  $\text{Var}(y_i)$  is the variance of the  $i^{\text{th}}$  output and  $\gamma_i$  is a corresponding scalar gain chosen such that  $\gamma_1 > \gamma_2 > \dots > \gamma_{N-1} > 0$ . In the following section, this gain (there is only one gain for  $N = 2$ ) is set to two.

Simulations have shown that this method converges more quickly (convergence is simultaneous for all principle components) than several other methods. For example, in [4] it is shown that the convergence is roughly 5 times as fast as Sanger's rule [6]. One other adjustment must be made to this criteria in order to sphere the data, that is to scale each output to have unit variance. The scale factors, which are the inverse square root of the eigenvalues of the unsphered observation covariance matrix, are already computed by the SIPEX-G algorithm and therefore require no additional computation.

The MerMaId-SIG algorithm is described completely in [3]. It is a non-parametric, information-theoretic criterion that aims to minimize the Renyi's quadratic mutual information between the outputs. The cost function is given by,

$$J = \sum_{i=1}^N H_{R_2}(y_i) \quad (5)$$

where  $H_{R_2}(y_i)$  is the Renyi's (quadratic) marginal entropy [7] of the  $i^{\text{th}}$  output which is given by,

$$H_{R_2}(y_i) = -\log \int_{-\infty}^{\infty} f_{Y_i}(y_i)^2 dy \quad (6)$$

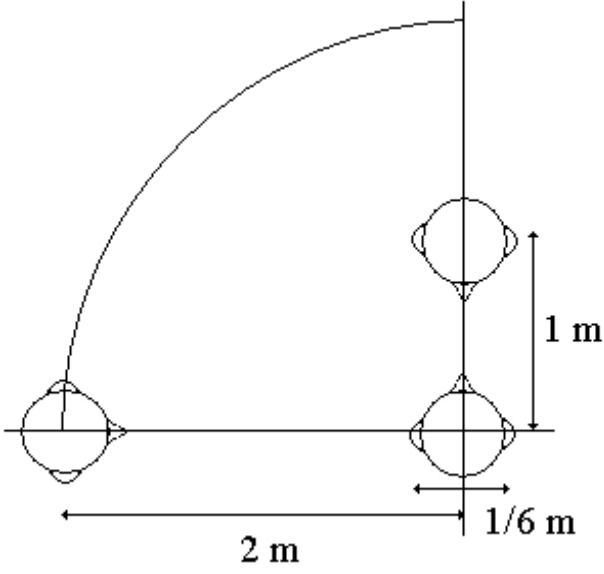


Fig. 1. Top view of two speakers and a single observer (lower right person).

1/12), where 1/12 is half of the head-width. See Figure 1 to help visualize the geometry for this example.

Using equation (1), the estimated sources,  $\mathbf{y}_k$ , are then determined to be,

$$\mathbf{y}_k = \mathbf{R}_k^T \mathbf{x}_k \quad (3)$$

where  $\mathbf{y}_k$  and  $\mathbf{R}_k$  are the  $(N \times 1)$  output and  $(N \times N)$  demixing matrices, respectively. For MerMaId-SIG, the  $\mathbf{R}_k$  matrix is constrained to be a pure rotation. Furthermore, the  $\mathbf{W}_k$  sphering matrix is ideally  $\Phi \Lambda^{-1/2}$ , where  $\Phi$  is the matrix of eigenvectors of

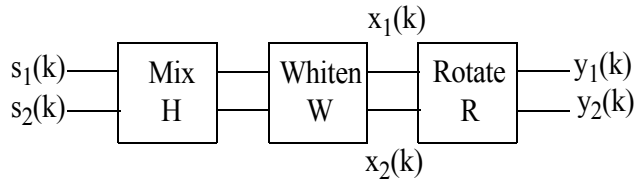


Fig. 2. Block diagram for BSS of  $N = 2$  sources and observations.

the autocorrelation of  $\mathbf{H}^T \mathbf{s}$  and  $\Lambda$  is the corresponding eigenvalue matrix. Notice, however, that the ideal value of  $\mathbf{W}_k$  changes as the mixing matrix changes, therefore it must be continually adapted as will the demixing matrix. See Figure 2 for a block diagram of the mixing and demixing process.

where  $f_Y(y_i)$  is the marginal pdf of the  $i^{\text{th}}$  output. When Parzen windows is used with Gaussian kernels, the resulting expression for the Renyi's marginal entropy is,

$$H_{R_2}(y_i) = -\log \frac{1}{L^2} \sum_{j=1}^L \sum_{k=1}^L G(y_i(j) - y_i(k), 2\sigma^2) \quad (7)$$

where  $G(x, \sigma^2)$  is a Gaussian pdf evaluated at  $x$ , having zero-mean and a variance of  $\sigma^2$ , and  $L$  is the blocksize. The above expression is for the original MeRMald algorithm, the only change required to produce the MeRMald-SIG algorithm is to drop one of the summations in equation (7) and sum over only consecutively occurring samples. This produces the desired result,

$$H_{R_2}(y_i) = -\log \frac{1}{L} \sum_{j=1}^L G(y_i(j) - y_i(j-1), 2\sigma^2) \quad (8)$$

Notice that the blocksize,  $L$ , in this equation does not prevent on-line adaptation. It merely represents the number of samples before the accumulated result is applied to the tap weights.

#### 4. RESULTS

The Signal-to-Interference Ratio, SIR, will be used to measure the performance of the separation techniques. The definition of the SIR is given as,

$$SIR = \frac{1}{N} \sum_{i=1}^N 10 \log_{10} \left( \frac{\max(q_i)^2}{q_i^T q_i - \max(q_i)^2} \right) \quad (\text{dB}) \quad (9)$$

where  $\mathbf{q} = \mathbf{HWR}$ ,  $\mathbf{q}_i$  is the  $i^{\text{th}}$  column of  $\mathbf{q}$ , and  $\max(\mathbf{q}_i)$  is the maximum element of  $\mathbf{q}_i$ . This criterion effectively measures the distance of the overall mixing matrix,  $\mathbf{q}$ , from the product of a permutation matrix and a diagonal matrix.

An instantaneous mixture of two sources, speech from a male and female speaker, is used to compare the MeRMald-SIG algorithm with Bell and Sejnowski's version of Infomax [8] and Yang's MMI (Minimum Mutual Information) method [9]. The time-varying mixing matrix was determined as stated above for the case that one of the sources is 1 m directly in front of the observer and the second speaker is initially 2 m directly in front of the observer, but is moving in a circle (counter-clockwise) of radius 2 m around the observer. This direction of motion was chosen because it produces a faster change in the mixing matrix coefficients than most other directions and therefore represents a worst-case type scenario. Note that this case, when the speaker walks in a circle around the observer, does not correspond to a change in only the rotation matrix. Also, as before, the head-width is 1/6 m.

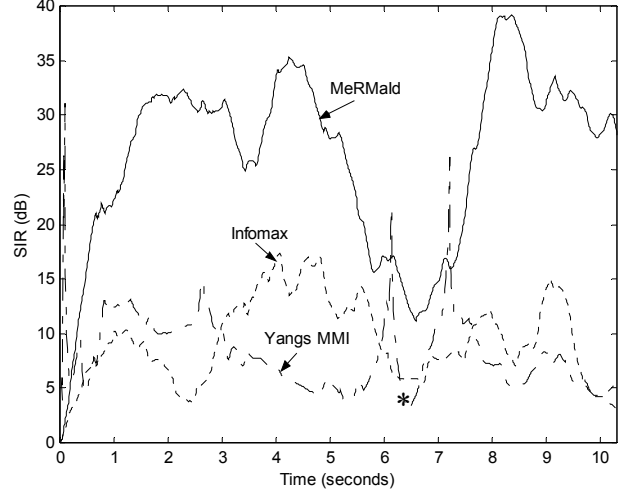


Fig. 3. SIR versus time for three different on-line separation methods for counter-clockwise motion of 1 m/s.

All three methods utilize the same spatial pre-whitening provided by the SIPEX-G algorithm, so that the separation techniques can be directly compared. In addition, the Infomax and Yang's MMI methods use Amari's natural gradient [10], and Yang's method uses the adaptive scheme for calculating cumulants. A blocksize of 200 was used and the step sizes for each method was chosen to optimize the SIR.

Figure 3 shows the first set of results, with some smoothing applied. In this case the speaker that is moving is doing so at a constant rate of 1 m/s. At this rate the speaker will be directly behind the observer at a time of about 6.3 s (marked by the asterisk), and will end up traversing about  $3\pi/2$  radians of a circle. Notice that when the speaker in motion passes either directly in front of or directly behind the observer, the mixing matrix becomes ill-conditioned (due to the location of the other speaker), and we can expect that during those times, the performance will suffer. As can be seen from the figure, this is in fact the case. Also notice that outside of those regions, the MeRMald-SIG algorithm quickly adapts to a solution with an SIR of over 20 dB, which is the value at which almost no audible interference can be heard. The other methods, on the other hand, never reach 20 dB SIR. In fact, Yang's method performs as well as it does only because the step size required for stability was so small that there was virtually no adaptation. The change in the performance seen for Yang's method is, consequently, due entirely to the sphering routine.

A second plot, shown in Figure 4, is used to display the results in a geometric manner. It is desired to show at each unit of time, given the coefficients of the demixing matrix, the point in two-dimensional space where the algorithm believes the two sources are located. This is not possible because of the indeterminacies of the demixing process; however, this can be mostly resolved if the demixer is allowed to know the true power of the sources. The vertical line in the plot shows the convergence of the estimate of the location of the stationary speaker due to the MeRMald-SIG algorithm, which has an ideal value of (0, 2). The circle (for  $3\pi/2$  radians) is the ideal path of the second speaker. The estimate in

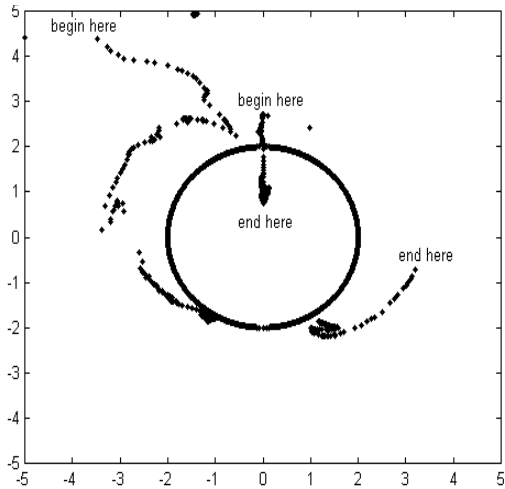


Fig. 4. Two-dimensional geometric estimates of the locations of the two sources.

this case starts out up and to the left, but then begins to trace out the circle. Notice, however, that it always overestimates the distance of the second speaker.

A second example was used to test the algorithm for the case that the speaker is moving much faster than in the previous example. In this case, shown in Figure 5, the speaker is now moving at a very brisk 5 m/s. At this rate, the speaker is able to go around the observer just over 4 times during the 10.7 seconds of data. Therefore we can expect 8 points in time where the mixing matrix is ill-conditioned (indicated by the asterisks). As can be seen, the SIR falls in each of these locations, but is quickly restored to a value of over 20 dB between each of these areas. Due to the poor performance on the easier data set, the other two separation methods were not tested here.

## 5. CONCLUSIONS

The speed of convergence of the separation and PCA algorithms is absolutely critical for real-world applications. If either one were very slow, it wouldn't matter how fast the other one was. Due to the extreme data efficiency of the two algorithms presented herein, namely MeRMaId-SIG and SIPEX-G, it has been shown that time-varying demixing (of instantaneous mixtures) is, in fact, possible.

## 6. ACKNOWLEDGEMENTS

This work was partially supported by NSF ECF #9900394.

## REFERENCES

- [1] Kari Torkkola, "Blind separation for audio signals - are we there yet?," ICA '99, pp. 239-244, Jan. 1999.
- [2] Kenneth E. Hild II, Deniz Erdogmus, and Jose C. Principe, "Blind source separation using Renyi's Mutual Information,"

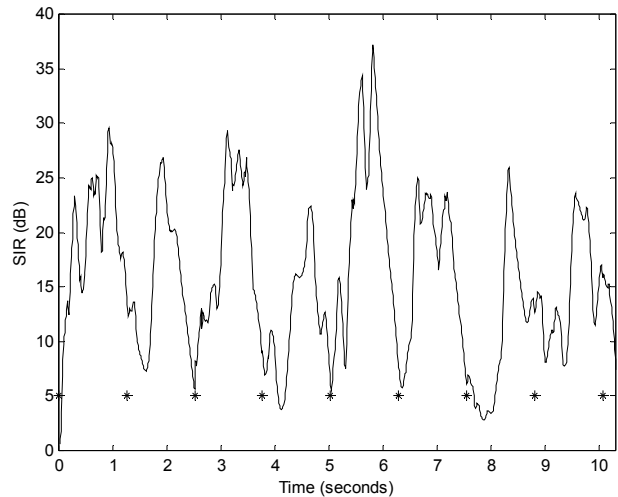


Fig. 5. SIR versus time for the MeRMaId-SIG algorithm for counter-clockwise motion of 5 m/s.

IEEE Signal Proc. Letters, Vol. 8, No. 6, pp. 174-176, June 2001.

- [3] Kenneth E. Hild II, Deniz Erdogmus, and Jose C. Principe, "On-line Minimum Mutual Information Method For Time-varying Blind Source Separation," accepted to Third Intl. Workshop on Independent Component Analysis and Signal Separation, ICA '01, San Diego, CA, Dec. 2001.
- [4] Deniz Erdogmus, Yadunandana N. Rao, Jose C. Principe, Jing Zhao, and Kenneth E. Hild II, "Simultaneous extraction of principal components using Givens rotations and output variances," submitted to the IEEE Intl. Conf. on Acoustics, Speech and Signal Proc., ICASSP '02, Orlando, FL, May 2002.
- [5] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, 2nd ed., John Hopkins University Press, Baltimore, Maryland, 1989.
- [6] T. D. Sanger, "Optimal unsupervised learning in a single layer linear feedforward neural network," *Neural Networks*, Vol. 2, No. 6, pp. 459-473, 1989.
- [7] A. Renyi, *Probability Theory*, North-Holland Publishing Company, Amsterdam, Netherlands, 1970.
- [8] Anthony J. Bell and Terrence J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, Vol. 7, No. 6, pp. 1129-1159, Nov. 1995.
- [9] Howard Hua Yang and Shun-ichi Amari, "Adaptive online learning algorithms for blind separation: maximum entropy and minimum mutual information," *Neural Computation*, Vol. 9, No. 7, pp. 1457-1482, Oct. 1997.
- [10] Shun-ichi Amari, "Neural learning in structured parameter spaces - natural Riemannian gradient," *Proc. of 1996 Advances in Neural Information Proc. Systems, NIPS '96*, pp. 127-133, MIT Press, 1997.