

INDEPENDENT COMPONENTS ANALYSIS FOR FETAL ELECTROCARDIOGRAM EXTRACTION: A CASE FOR THE DATA EFFICIENT MERMAID ALGORITHM

Dorothee E. Marossero¹, Deniz Erdogmus²,
Neil Euliano¹, Jose C. Principe², Kenneth E. Hild II²

¹NeuroDimension Inc., 1800 N Main St, Suite D4, Gainesville, FL 32608

²CNEL, University of Florida, Gainesville, FL 32611

Abstract. Fetal heart rate (FHR) monitoring is currently the primary methodology for antenatal determination of fetal well-being. Currently, the FHR can be detected with ultrasonography, but the additional information from fetal electrocardiogram (FECG) is only available via an invasive scalp electrode. A cost effective noninvasive monitoring through standard ECG electrodes could be used on nearly every patient in lieu of the ultrasound monitors. In this method, a number of electrodes are positioned on the abdomen of the mother to collect, simultaneously, various combinations of the signals including the heartbeats of the mother and the fetus. For accurate fetal heart-rate estimation, a clean FECG must be extracted from the collected mixtures. It is well known that this can be achieved using blind source separation (BSS) techniques. In this paper, the performance of the Mermaid algorithm, which is based on minimizing Renyi's mutual information, is evaluated on this problem of great practical importance. The effectiveness and data efficiency of Mermaid and its superiority over alternative information theoretic BSS algorithms are illustrated using artificially mixed ECG signals as well as fetal heart rate estimates in real ECG mixtures.

INTRODUCTION

Fetal heart rate (FHR) monitoring is currently the primary methodology for antenatal determination of fetal well-being. Currently, the FHR can be detected with ultrasonography, but the additional information from fetal electrocardiogram (FECG) is only available via an invasive scalp electrode. A cost effective noninvasive monitoring through standard ECG electrodes could be used on nearly every patient in lieu of the ultrasound monitors. Data collection using ECG is performed by electrodes located at various points on the mother's skin, especially at the abdomen. The resulting signals are mixtures of maternal electrocardiogram (MECG), fetal electrocardiogram (FECG), and interfering noise such as base-line wander, power line interference, maternal electromyogram, thermal noise, and other noise from electronic equipment [1]. Hence, in order to be monitored, the FECG has to be separated from these noisy elements. Numerous methods have been used to solve this extraction problem [2] besides the approaches that assume an instantaneous mixture of independent source signals and employ the well-established blind source separation techniques [3].

Under the instantaneous linear mixture assumption, which is used for the blind extraction of FECG, the problem is formulated as follows. Suppose,

there are m independent source signals $\{s_1, \dots, s_m\}$ that are linearly mixed to form the n mixtures (recordings of each electrode in the current setting) denoted by $\{x_1, \dots, x_n\}$. In vector-matrix form, this equation is expressed as $\mathbf{x} = \mathbf{H}\mathbf{s}$, where the vector \mathbf{x} consists of the mixtures, the vector \mathbf{s} consists of the source signals, and the matrix \mathbf{H} defines the linear mixture. The observations will include measurement noise in reality, as well. However, for the theoretical development of BSS algorithms, this is usually omitted. Especially in information theoretic approaches, where the solution can equivalently be expressed as maximally distant from a Gaussian distribution, any Gaussian noise will not, on average, affect the solution. In this formulation, neither the source signals nor the mixing matrix are known, therefore the problem of determining the source signals, using only the observations, is done *blindly*. Of course, in practice, only a finite number of samples of the observation vector \mathbf{x} are available and the only basic assumption about the source signals is their independence.

In this paper, we will demonstrate the performance of an information theoretic independent components analysis (ICA) algorithm named MeRMaId (for Minimum Renyi's Mutual Information) [4]. Performance comparisons of the batch and sample-by-sample versions of Mermaid are presented with FastICA [5] and InfoMax [6] on the separation of artificially mixed ECG data and real mixtures. In the artificial mixtures, the performance evaluation is based on the signal-to-distortion ratio (SDR) computed using the actual mixing matrix. For the real mixtures, an end-to-end system criterion using the Pan-Tompkins QRS detection algorithm [7] is developed.

MERMAID ALGORITHM FOR ICA

In the following, we assume that the mixture is square. In practice, when the number of mixtures is greater than the number of sources (assuming that the latter is known), PCA is applied to reduce the dimensionality of the mixture vector preserving as much variance as possible, which also helps improve the signal-to-noise ratio (SNR) of the measurements. In BSS, spatial pre-whitening is usually applied to the observations in order to improve the convergence speed of the algorithms. In Mermaid, the whitening is an integral part of the algorithm. It is well known that after the spatial whitening of the mixtures, the separation can be achieved using only a rotation. In Mermaid, Renyi's mutual information is minimized to determine the optimal values for the rotation angles, whereas the preliminary whitening is based on only the second order statistics of the observed data. Given the covariance matrix for the observed vector \mathbf{x} , it is easy to determine the eigenvectors of this matrix and the corresponding eigenvalues. The whitening matrix \mathbf{W} is then given by $\mathbf{W} = \mathbf{\Lambda}^{-1/2} \mathbf{\Phi}^T$, where $\mathbf{\Phi}$ is the orthonormal eigenvector matrix corresponding to the diagonal eigenvalue matrix $\mathbf{\Lambda}$. The whitened mixture \mathbf{x}^w is then separated using a rotation matrix parameterized in terms of Givens rotations with $\mathbf{Y} = \mathbf{R}(\mathbf{\Theta}) \mathbf{x}^w$, where $\mathbf{\Theta}$ denotes the vector of Givens rotation angles.

In [4], it has been shown that the optimal separation matrix parameters $\mathbf{\Theta}_{\text{opt}}$ can be obtained through the minimization of the following simplified cost function

$$J = \sum_{o=1}^n H_{\alpha}(y^o) \quad (1)$$

where $H_\alpha(y^o)$ is the marginal order- α Renyi's entropy for the random variable y^o of the output channel o , and n is the dimensionality of the output vector \mathbf{y} . This cost function, together with the assumed separation topology, exploits effectively the property of invariance-under-rotations of Renyi's joint entropy. Besides coinciding with the minimization of Renyi's mutual information between the separated output signals, the minimization of (1) can also be interpreted as the minimization of the average Gaussianity of the separated output signals, since the whitening-rotation scheme guarantees that each output is of unit-variance and the Gaussian distribution has the maximum Shannon entropy among fixed-variance density functions.

The rotation matrix $\mathbf{R}(\Theta)$, is given in terms of the Givens rotation angles θ_{ij} by the product of in-plane rotation matrices, $\mathbf{R}_{ij}(\theta_{ij})$

$$\mathbf{R}(\Theta) = \prod_{i=1}^{n-1} \prod_{j=i+1}^n \mathbf{R}_{ij}(\theta_{ij}) \quad (2)$$

Each of these simple rotation matrices represents a rotation in the i - j plane of the n -dimensional output space. These in-plane rotation matrices are simply identity matrices whose elements at the intersection of i^{th} and j^{th} rows and columns are modified as follows: $(i,i)^{\text{th}}$, $(i,j)^{\text{th}}$, $(j,i)^{\text{th}}$, and $(j,j)^{\text{th}}$ entries are modified to read $\cos\theta_{ij}$, $-\sin\theta_{ij}$, $\sin\theta_{ij}$, and $\cos\theta_{ij}$, respectively. There are a total of $n(n-1)/2$ such in-plane rotations that define an overall rotation in n dimensions. All matrix products in (1) can be performed from the left or from the right, as long as the convention is carried out to the computation of the gradient for the angle updates.

Given the output samples $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ corresponding to a specific choice of the Givens angles, one can utilize the following nonparametric estimator for Renyi's α -entropy [8], which is based on the Parzen window estimate of the underlying probability distribution functions (pdf) [9].

$$\hat{H}_\alpha(y^o) = \frac{1}{1-\alpha} \log \left[\frac{1}{N^\alpha} \sum_{j=1}^N \left(\sum_{i=1}^N \kappa_\sigma(y_j^o - y_i^o) \right)^{\alpha-1} \right] \quad (3)$$

In (3), $\kappa_\sigma(\cdot)$ is the kernel function, which is usually selected to be a symmetric differentiable pdf (for example the zero-mean Gaussian pdf with standard deviation σ). The parameter σ controls the kernel width, and is named the kernel size. Usually, larger kernel sizes lead to high bias, low variance density estimates, whereas smaller kernel sizes lead to low bias, high variance density estimates.

In order to minimize the cost function in (1) estimated using the nonparametric formula in (3), one could update the Givens rotation angles using a batch gradient [8,10] in off-line adaptation or a stochastic gradient in on-line adaptation [11,12]. For convenience, these two gradient expressions are provided below. The batch gradient is obtained by direct derivation of (1) with respect to the Givens angles.

$$\frac{\partial J}{\partial \theta_{kl}} = - \sum_{o=1}^n \frac{\sum_{j=1}^N \left(\sum_{i=1}^N \kappa_\sigma(y_j^o - y_i^o) \right)^{\alpha-2} \left(\sum_{i=1}^N \kappa'_\sigma(y_j^o - y_i^o) \left(\frac{\partial y_j^o}{\partial \theta_{kl}} - \frac{\partial y_i^o}{\partial \theta_{kl}} \right) \right)}{\sum_{j=1}^N \left(\sum_{i=1}^N \kappa_\sigma(y_j^o - y_i^o) \right)^{\alpha-1}} \quad (4)$$

where the output sensitivities $\partial y_j^o / \partial \theta_{kl}$ are computed using the fact that $y_j^o = \mathbf{R}^o \mathbf{x}_j^w$, where \mathbf{R}^o is the o^{th} row of the rotation matrix and j is the sample index.

$$\frac{\partial y_j^o}{\partial \theta_{kl}} = \frac{\partial \mathbf{R}^o}{\partial \theta_{kl}} \mathbf{x}_j^w = \left(\frac{\partial \mathbf{R}}{\partial \theta_{kl}} \right)^o \mathbf{x}_j^w \quad (5)$$

Similarly, the stochastic gradient, called Mermaid-SIG, which uses a window of L samples (these samples could be selected as the most recent L samples to minimize the memory requirement or a random set of L previously collected samples to minimize any possible correlation over time between the samples, thus make them closer to being independent), at time step k is

$$\frac{\partial J_k}{\partial \theta_{kl}} \approx - \frac{\sum_{j=k-L+1}^k \kappa_{\sigma}^{\alpha-2} (y_j^o - y_{j-1}^o) \cdot \kappa'_{\sigma} (y_j^o - y_{j-1}^o) \left(\frac{\partial y_j^o}{\partial \theta_{kl}} - \frac{\partial y_{j-1}^o}{\partial \theta_{kl}} \right)}{\sum_{j=k-L+1}^k \kappa_{\sigma}^{\alpha-1} (y_j^o - y_{j-1}^o)} \quad (6)$$

In [11], the superior tracking capability of this algorithm in nonstationary mixture cases is demonstrated in even fast time-varying mixtures.

To summarize, the Mermaid algorithm determines the separated sources using the following steps:

1. Initialize Givens angles (to all zeros or randomly).
2. Compute the whitening matrix as prescribed using all samples in off-line separation or update the whitening matrix using an adaptive PCA algorithm in on-line.
3. In off-line separation, use the batch gradient in (4), which is computed using all available samples. In on-line separation, use the stochastic gradient in (6), which uses only L samples, including the most recent sample at time k .
4. Update the Givens rotation angles using steepest descent

$$\Theta = \Theta - \eta \frac{\partial J}{\partial \Theta} \quad (7)$$

Previous studies on audio signal separation have demonstrated that the batch Mermaid algorithm is data efficient, whereas the stochastic Mermaid algorithm is fast in convergence and accurate in tracking time-varying mixtures, which might as well be interpreted as data efficiency.

In the following sections, we seek to test these hypotheses by comparing Mermaid to two benchmark ICA algorithms: Hyvarinen's FastICA [5] and Bell & Sejnowski's InfoMax [6] (with Amari's natural gradient [13]).

MEASURES OF PERFORMANCE

We include three basic aspects to compare the different algorithms: signal separation performance, algorithmic complexity, and robustness. Due to the unsupervised nature of blind source separation algorithms, we propose to first evaluate the algorithms using artificially mixed signals. The artificial mixtures are designed to serve as controlled experimental setups to determine the capabilities of different separation algorithms, as well as to make an objective and unbiased comparison of their performance. Since in this situation the actual mixing matrix is known, SDR can be utilized. The

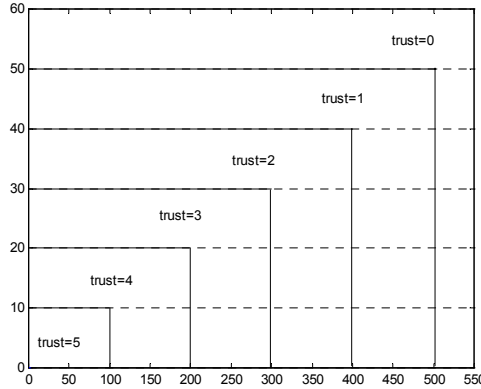


Figure 1. Discretization regions for the trust factor where the horizontal and vertical axes are FHR variance and FN+FP, respectively.

SDR, defined in (8), is an index that measures the closeness of the product of the mixing and separation matrix to a perfect separation solution (with permutation and scaling ambiguities permitted). Letting $\mathbf{O}=\mathbf{R}\mathbf{W}\mathbf{H}$ be the overall matrix defining the relationship between the separated sources and the actual sources, the SDR is given by

$$SDR = \frac{1}{n} \sum_{i=1}^n 10 \log_{10} \left(\frac{(\max_j O_{ij})^2}{\sum_{j=1}^n O_{ij}^2 - (\max_j O_{ij})^2} \right) \quad (8)$$

Essentially, a larger SDR signifies a better separation of the original sources, thus means a better performance of the separation algorithm used.

In real mixtures (recorded from real patients), the true mixing matrix is unknown. The measured signals are a combination of many sources, therefore, the algorithmic performance must be measured in terms of the end-goal of the designed system. In the fetal ECG extraction case, at the output of the Mermaid algorithm, the Pan Tompkins online QRS detection algorithm [7] is applied to the separated fetal ECG signal. First, a band-pass filter is used to remove interfering noise. The outputs of the filter are differentiated and squared. This is followed by an integration stage realized using a moving-average filter. Finally, thresholds are set to find the locations of the peaks, which gives us the beat-to-beat heart rate and its variance. The orders and bandwidths of the different filters and the thresholds used are different when applied to the Maternal ECGs and the fetal ECGs, since these signals have different power spectra and amplitudes. An estimation of the numbers of peaks that we found when none was present (false positive) and of peaks that we missed (false negative) are also computed on line. A false positive (FP) peak is estimated when the RR interval between the previous peak and the present one is less than 70% of the 5 previous RR intervals' average. A false negative (FN) is estimated when the RR interval between the previous peak and the present one is greater than 130% of the 5 previous RR intervals' average.

Based on these, a parameter called the "Trust factor" (TF) is developed. This parameter is calculated from the most important aspects that lead us to trust or distrust the fetal heart rate calculation done in real-time: the numbers of FN and FP added together versus the variance of the fetal heart rate. A good separation level corresponds to small values for the FP, FN, and variance, simultaneously. We created 6 regions for the TF (see Fig. 1). The

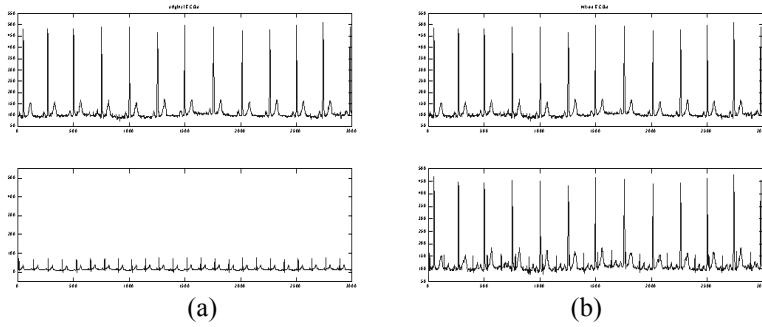


Figure 2. Artificial ECG mixture simulation a) source signals b) mixtures.

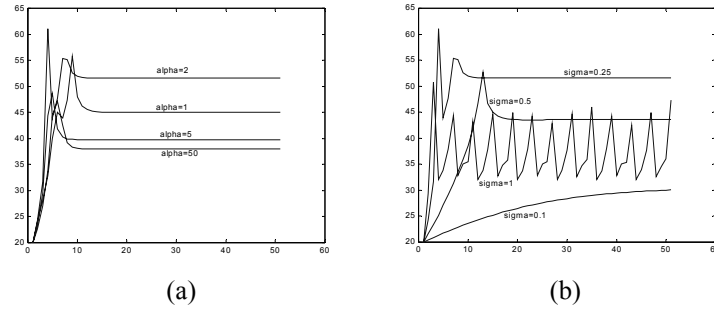


Figure 3. The effect of a) entropy order and b) kernel size on the performance of Mermaid in separating the two artificially mixed ECG signals shown in terms of SDR (dB) versus iterations.

variable TF ranges from 0 to 5, 0 being the worst case where the heart-rate estimate is not reliable, and 5 being the best case where the fetal heart rate is obtained with highest certainty.

COMPARISON OF PERFORMANCE ON ARTIFICIAL MIXTURES

In these simulations, an artificial mixture of two clean ECG signals with an original sampling frequency of 500 Hz was used. One of the source signals was downsampled by a factor of 2 and its amplitude was reduced by 3, in order to make it appear as a fetal ECG. The realistic *unknown* mixing matrix was selected to be

$$H = \begin{bmatrix} 1 & 0.05 \\ 0.9 & 1 \end{bmatrix} \quad (9)$$

in order to describe accurately the environment where the MECG influence in the mixtures is significant. The original source signals and the mixed signals are shown in Fig. 2.

First, the batch algorithm that uses (4) is considered. The effect of entropy order and kernel size on the performance of Mermaid is investigated. The ECG signals are super-Gaussian, therefore, in accordance with the suggestion in [10] we expect better performance using an entropy order that is greater than or equal to 2. We observe in Fig. 3a that the quadratic entropy yields the best separation (using a Gaussian kernel function with a size of $\sigma = 0.25$). We also notice in Fig. 3b that extremely large and small kernel sizes result in degradation of performance. Notice that in all cases, the SDR of the final solution is much greater than 20 dB, which corresponds to an average signal power to distortion power ratio of 100 over the two output

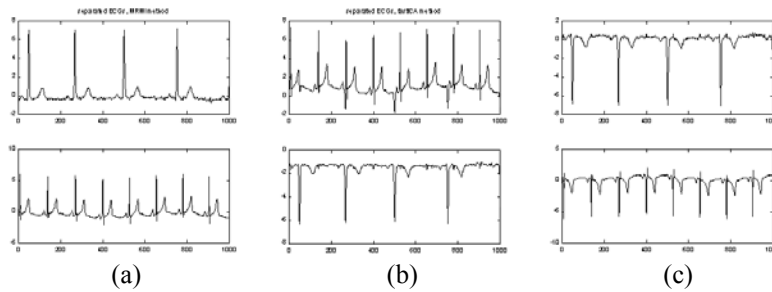


Figure 4. Separated signals using a) Mermaid trained with 100 samples b) FastICA trained with the same 100 samples c) FastICA trained with 1000 samples.

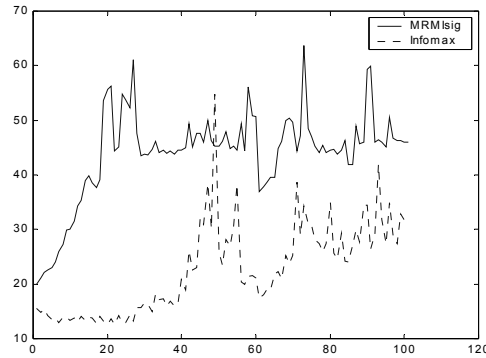


Figure 5. SDR (dB) of Mermaid-SIG (solid) and InfoMax (dotted) versus the number of iterations in on-line separation of artificially mixed ECG signals.

channels. Usually, we consider an SDR of 20dB or higher an acceptable solution. Of course, as the number of channels (sources) increase, the final SDR obtained will decrease due to the increasing number of interfering signals versus a single desired source signal at each channel. Therefore, a margin over 20 dB in a two-dimensional separation problem is necessary for acceptable generalization of performance to higher order problems.

In this controlled off-line separation setting, we compare the performance of Mermaid (code available at www.cnel.ufl.edu) with that of FastICA [5] (code available at www.cis.hut.fi/projects/ica/fastica). Fig. 4 shows the separated signals using Mermaid and FastICA. Both algorithms converge to their final solutions in 10 to 12 iterations and their final separation performances in terms of SDR are approximately 38dB, 25dB, and 29dB, respectively. These results conform to the well-known fact that FastICA requires a large number of samples to achieve high separation performance as it relies on the evaluation of the kurtosis of the output signals. Mermaid, on the other hand, achieves an SDR level much higher than FastICA with only a few samples drawn from the mixture. This observation is in accordance with the data efficiency of Mermaid as discussed in [HildSPL]. As for the computation time required to achieve these final solutions, the total CPU time required for convergence on a standard Pentium2 400 MHz desktop PC is compared: Mermaid(100) requires 0.82sec, whereas FastICA(100) takes 0.80sec, and FastICA(1000) takes 1.37sec to complete the computations.

In the same setting, the on-line separation capabilities of Mermaid-SIG, which uses (6), and InfoMax with natural gradient are compared. In these simulations, Mermaid-SIG uses a step size of 0.05 and a window length of 100 samples. Quadratic entropy is used and the $\sigma = 0.25$ as before. On the

other side, InfoMax uses a step size of 0.001 and the updates are accumulated over windows of 100 samples to be applied all at once. For a fair comparison, both algorithms utilize the same whitening matrix to pre-whiten the mixtures. Furthermore, the step size values are set to yield approximately the same level of variance in SDR after convergence (by trial and error). The result presented in Fig. 5 illustrates the fast convergence and data efficiency properties of Mermaid-SIG. Besides, it achieves an average SDR of more than 45dB after convergence, whereas InfoMax finds a solution with a 30dB separation. For a brief complexity comparison, we note that the total CPU time required to perform this simulation, which used a total of 10000 samples in 100 updates was 0.20sec for Mermaid-SIG and 0.22sec for InfoMax on the same desktop mentioned before.

Thus, we conclude that Mermaid-SIG is an effective, fast converging stochastic BSS algorithm that is a strong alternative to the batch Mermaid; the latter has a complexity that increases as $O(N^2)$ with the number of samples, whereas the complexity of the former is $O(L)$ with the window length. The strength of Mermaid-SIG, compared with Mermaid, lies in the fact that with the stochastic updates it is possible to make use of more samples in a shorter time, thus extract more information about the mixture, while keeping the computational complexity down at acceptable levels.

Table I. Results for non-clean artificial data.

Algorithm	# samples	SDR max (dB)	# of iterations
Noisy Artificial Data			
Mermaid-SIG	100	50	20
Infomax	100	25	80
Fast ICA	500	20	batch
Varying Mixing Matrix			
Mermaid-SIG	100	30	50
Infomax	100	7	80
Fast ICA	500	22	batch

Additional experiments were performed on noisy artificial mixtures of ECG signals, in scenarios where the mixing matrix is stationary and time varying (as could be the situation in real measurements). Table I summarizes the results. In all cases, Mermaid-SIG provides higher SDRs and converges faster using less data samples and/or less iterations. Thus, in the artificial mixture experiments, we conclude that the Mermaid algorithm outperforms the other two BSS algorithms.

APPLICATION OF MERMAID TO REAL ECG MIXTURES

In this section, the separation results on several 5-minute data sets from 8-lead ECG abdominal recordings, sampled at 200Hz are presented. The mixtures are subjected to a DC-removal preprocessing stage, which consisted of simple order-6 FIR high-pass filters. A sample from the 8 mixtures after DC removal is shown in Fig. 6. Filtering the mixtures commutes with the mixing matrix since both are linear operations on the signals, thus this procedure effectively filters the original source signals.

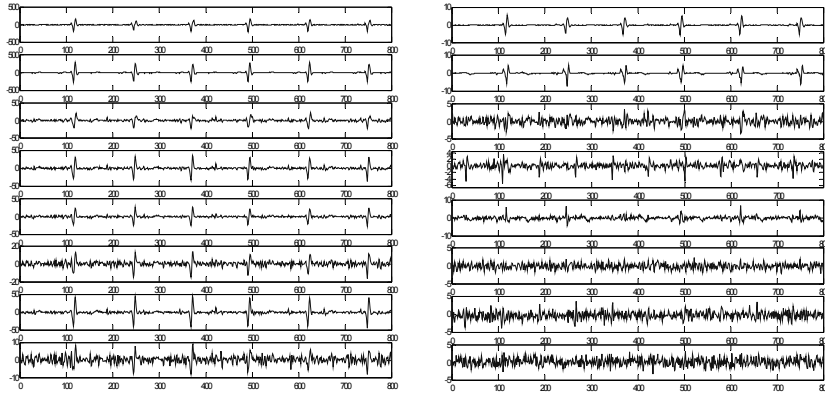


Figure 6. Measured real ECG signals after the DC-removal pre-processing stage (left) and at the separated outputs from Mermaid-SIG (right).

After DC removal and pre-whitening, FastICA (with the symmetric approach and the Gaussian nonlinearity), InfoMax (with natural gradient), and Mermaid-SIG (with Gaussian kernels and quadratic entropy) are applied to the 8-channel mixture to extract 8 source estimates. The separated signals obtained with Mermaid-SIG are shown in Fig. 5. The data set shown here is from “Datafeb06”.

Table II. Average Trust Factor for the 3 Algorithms on Different Datasets

<i>Data Set</i>	<i>Mermaid-SIG</i>	<i>Infomax</i>	<i>FastICA</i>
Average	4.4	2.8	3.3
Datamay2	5.0 (100,0.1)	3.0 (100,0.0001)	5.0 (100,0.1)
Datafeb06	4.5 (100,0.2)	4.0 (100,0.0001)	2.0 (100,0.1)
Datafeb6	3.3 (100,0.1)	2.6 (1000,0.0001)	3.7 (500,0.01)
Datajune24	3.0 (100,0.1)	2.5 (100,0.0001)	3.0 (100,0.01)
Datajuly05_3	5.0 (100,0.1)	2.0 (1000,0.0001)	1.5 (100,0.01)
Datajuly17	5.0 (100,0.1)	2.0 (1000,0.0001)	5.0 (500,0.1)
Datajuly17_2	3.7 (100,0.1)	3.2 (1000,0.0001)	3.7 (100,0.01)

All separated source estimates are then subjected to the Pan-Tompkins QRS detection algorithm [7] that detects the location of the beats in the fetal ECG, which gives us the heart rate based on the R-R interval estimates. The performance of the three separation algorithms are compared in terms of the average Trust Factor they yield. Table II summarizes the performance of the three algorithms. For each data set, TF is calculated every 4 seconds and the bold numbers in the table correspond to the average TF for each data set over time. The average TF for each algorithm is then also averaged over data sets. For all algorithms a search for the best possible results has been performed by adjusting the window and step sizes. The best performing values are given in parantheses next to the average TF levels. From Table II, we observe that Mermaid-SIG clearly outperforms Infomax and also does better than FastICA on average.

CONCLUSIONS

Fetal ECG extraction is an important practical problem that can be solved using blind source separation techniques. In this paper, a performance comparison of three information theoretically motivated ICA algorithms is presented: Mermaid, Infomax, and FastICA. The experimental results in artificially mixed ECG signals and fetal heart rate detection from real ECG measurements showed that Mermaid outperforms the other two algorithms. Mermaid is shown to be more data efficient both in batch and on-line operation modes, which is an important feature for real-time implementation, since the mixture model could be time-varying in real mixtures.

Acknowledgments: This work is supported by NSF grant DMI-0239060. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would also like to thank Dr. T. Euliano for collecting the data and V. Meka for designing and building the data collection amplifier.

REFERENCES

- [1] J. R. Mazzeo, "Non-invasive fetal electrocardiography", Med Prog Technol. 1994; pp. 20-75.
- [2] V. Zarzoso, A. K. Nandi, "Comparison between blind separation and adaptive noise cancellation techniques for fetal electrocardiogram extraction", In Proc. IEE Colloquium on Medical Applications of Signal Processing, number 107, London, October 6 1999; pp. 1/1-1/6
- [3] E. Bacharakis, A.K. Nandi, V. Zarzoso, "Fetal ECG extraction using blind source separation methods," Proceedings of EUSIPCO'96, pp. 395-398, Trieste, Italy, 1996.
- [4] K.E. Hild II, D. Erdogmus, J.C. Principe, "Blind source separation using Renyi's mutual information," IEEE Signal Processing Letters, vol. 8, pp. 174-176, 2001.
- [5] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," IEEE Transactions on Neural Networks, pp. 626-634, 1999.
- [6] A. Bell, T. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," Neural Computation, vol. 7, pp. 1120-1159, 1995.
- [7] J. Pan, W.J. Tompkins, "A real time QRS detection algorithm," IEEE Transactions on Biomedical Engineering, vol. 32, no. 3, pp. 837-843, 1985.
- [8] D. Erdogmus, K.E. Hild II, J.C. Principe, "Blind source separation using Renyi's α -marginal entropies," Neurocomputing, vol. 49, no. 1, pp. 25-38, 2002.
- [9] E. Parzen, "On estimation of a probability density function and mode," Time Series Analysis Papers, Holden-Day, Inc., California, 1967.
- [10] K.E. Hild II, D. Erdogmus, J.C. Principe, "Blind source separation of time-varying, instantaneous mixtures using an on-line algorithm," Proceedings of ICASSP'02, vol. 1, pp. 993-996, Orlando, Florida, May 2002.
- [11] S. Amari, "Neural learning in structured parameter spaces: Natural Riemannian gradient," Proceedings of NIPS'96, pp. 127-133, 1996.