ADAPTIVE NOISE CANCELLATION AND BLIND SOURCE SEPARATION

M. G. Jafari, and J. A. Chambers

Centre for Digital Signal Processing Research, King's College London, London WC2R 2LS, U.K.
E-mail: maria.jafari@kcl.ac.uk

ABSTRACT

The use of adaptive noise cancellers (ANCs) to reduce the noise level prior to source separation is investigated in this paper. The foetal electrocardiogram (ECG) extraction problem in particular is addressed, which as well as with noise, is compounded by the non-stationary nature of the measurements. Consequently, computer simulations show that the combined Kalman filter and natural gradient algorithm [1], cascaded with a parallel ANC network, leads to a technique that can significantly improve separation performance. Moreover, it is shown that in some cases, the performance of the method is better than that of the JADE algorithm [2].

1. INTRODUCTION

Blind source separation (BSS) is concerned with the recovery of the original independent sources whose mixtures, generated by travel through an unknown medium, are observed. Conventional BSS techniques are based on the assumption that the mixing occurs instantaneously and in the absence of any additive noise. However, this is rarely true in practice, and when the instantaneous model does not hold, separation performance deteriorates. The analysis of biomedical measurements, for instance, is often hindered by the presence of noise and interfering signals. Foetal electrocardiogram (FECG) extraction, in particular, requires that mains noise, as well as maternal ECG (MECG) components, breathing artefacts and so on, be suppressed in order to obtain an ECG clear enough to observe the foetus' heart rate. It was first addressed around 1965 by Widrow and his colleagues at Stanford University [3], who utilised adaptive noise cancellers to remove periodic sinusoidal interference [4]. In this paper, we investigate the use of adaptive noise cancellers as a pre-processing stage to the natural gradient algorithm (NGA) [5], to address the problem of source separation from instantaneous noisy mixtures. In particular, we focus on the foetal ECG extraction problem, which is compounded by the fact that ECGs are generally non-stationary. However, it has been shown in [6, 1] that the combined

Kalman filter and NGA method (KF-NGA) performs significantly better than NGA when the original sources are mixed by time-varying environments. Thus, a parallel ANC network is cascaded with the KF-NGA approach, and simulation results show that it leads to a structure that can offer better separation performance than the block-based JADE algorithm.

Hence, we begin with stating the BSS problem in Section 2, followed by a brief description of the KF-NGA approach in Section 3. The ANC method is summarised in Section 4, while the performance of the various techniques is investigated in Section 5. Conclusions are drawn in Section 6.

2. PROBLEM STATEMENT

The m observed signals generated when n sources are mixed by a time-invariant instantaneous channel, and no noise is present, are given by [7]

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) \qquad (1)$$

where $\mathbf{x}(k) \in \mathbb{R}^m$ is the vector of observed signals, and $\mathbf{s}(k) \in \mathbb{R}^n$ is the vector of source signals, assumed to be zero-mean and mutually independent. $\mathbf{A} \in \mathbb{R}^{m \times n}$ is an unknown, full column rank mixing matrix, and typically it is assumed that there are at least as many sensors as sources, that is $m \geq n$. The sources are recovered using the following linear separating system

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k) \qquad (2)$$

where $\mathbf{y}(k) \in \mathbb{R}^n$ is an estimate of $\mathbf{s}(k)$, and $\mathbf{W}(k) \in \mathbb{R}^{n \times m}$ is the separating matrix. Conventional BSS assumes that at most one source has Gaussian distribution because, for Gaussian random variables, uncorrelatedness corresponds to independence. The sources can only be recovered up to a multiplicative constant, and their order cannot be predetermined, so that perfect separation is achieved when the global mixing-separating matrix, defined as

$$\mathbf{P}(k) = \mathbf{W}(k)\mathbf{A} \qquad (3)$$

tends toward a matrix with only one non-zero term in each row and column [7], and is given by

$$\mathbf{P}(k) = \mathbf{J}\mathbf{D} \qquad (4)$$

where $\mathbf{J} \in \mathbb{R}^{n \times n}$ is a permutation matrix modeling the ordering ambiguity, and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix accounting for the scaling indeterminacy. The natural gradient algorithm is a BSS method whose update equation is given by the following expression [5, 7]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta(k)[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)]\mathbf{W}(k) \quad (5)$$

where $(\cdot)^T$ denotes the transpose operator, $\mathbf{I}$ is the identity matrix, $\mathbf{f}(\mathbf{y}(k))$ is an odd non-linear function of the output $\mathbf{y}(k)$, called the activation function, and $\eta(k)$ is a positive learning parameter. Usually the learning rate is assumed to be a very small positive constant which is either fixed or decreases exponentially to zero [8, 7].

3. COMBINED KF-NGA APPROACH

The KF-NGA approach uses NGA as the basic BSS block, which updates adaptively the separating matrix, thus estimating the source signals. Algorithm tracking ability is provided by the KF technique, which uses the recovered sources and the observed signals, to estimate the mixing matrix, and is described by the following expressions [9]

$$\mathbf{h}_K^p(k) = \mathbf{T}\mathbf{h}_K^c(k-1) \quad (6)$$

$$\mathbf{M}(k) = \mathbf{T}\mathbf{M}(k-1)\mathbf{T}^T + \mathbf{Q} \quad (7)$$

$$\mathbf{K}(k) = \mathbf{M}(k)\mathbf{S}^T(k)(\mathbf{C}(k) + \mathbf{S}(k)\mathbf{M}(k)\mathbf{S}^T(k))^{-1} \quad (8)$$

$$\mathbf{h}_K^c(k) = \mathbf{h}_K^p(k) + \mathbf{K}(k)(\mathbf{x}(k) - \mathbf{S}(k)\mathbf{h}_K^p(k)) \quad (9)$$

$$\mathbf{M}(k) = (\mathbf{I} - \mathbf{K}(k)\mathbf{S}(k))\mathbf{M}(k) \quad (10)$$

where $\mathbf{h}_K^p(k)$ and $\mathbf{h}_K^c(k)$ denote respectively the predicted and corrected estimate of the vector $\mathbf{h}(k)$, and the vector of sensor measurements $\mathbf{x}(k)$ is taken as the desired response of the Kalman filter. $\mathbf{S}(k)$ and $\mathbf{T}$ are, respectively, the known observation matrix and state transition matrix, $\mathbf{Q}$ and $\mathbf{C}(k)$ are respectively the covariance matrices of the process noise, and of the measurement noise. The Kalman gain is the matrix $\mathbf{K}(k)$, $\mathbf{M}(k)$ represents the parameter error covariance matrix, and $\mathbf{I}$ is the identity matrix. For our implementation, we re-arrange the mixing matrix into an mn-dimensional column vector, defined as

$$\mathbf{h}(k) = vec(\mathbf{A}^T(k)) \quad (11)$$

Unlike the method in [1], no assumptions are made about the distribution of the source signals, so that the observation matrix is obtained by normalising, rather than quantising, the source estimates generated by NGA. Normalisation of the outputs simply ensures that they satisfy the assumption that the sources have unit variance, and is found to improve significantly the performance of the Kalman filter. Thus, the observation matrix $\mathbf{S}(k)$ in (6)-(10) is replaced by the mxmn matrix $\hat{\mathbf{S}}(k)$, defined as

$$\hat{\mathbf{S}}(k) = \mathbf{I}_m \otimes \hat{\mathbf{s}}^T(k) \quad (12)$$

where $\mathbf{I}_m$ is the m-dimensional identity matrix, and $\otimes$ denotes the Kronecker product. The elements of the vector $\hat{\mathbf{s}}(k)$ are given by

$$\hat{s}_i(k) = g_i\left(y_i(k), \sqrt{c_i(k)}\right) \quad (13)$$

where $c_i(k)$ is the estimated variance output at the current iteration, and is given by

$$\bar{m}_i(k+1) = (1-\lambda)\bar{m}_i(k) + \lambda\left[y_i(k+1) - \bar{m}_i(k)\right] \quad (14)$$

$$c_i(k+1) = (1-\lambda)c_i(k) + \lambda\left[y_i(k+1) - \bar{m}_i(k+1)\right]^2 \quad (15)$$

where $y_i(k)$ is the i-th source signal estimated by NGA, and λ, $0 < \lambda < 1$ controls the memory. The mixing coefficients vector estimated by KF is then re-arranged into the mxn matrix $\mathbf{A}_K(k)$, and its pseudo-inverse generates an additional separating matrix, $\mathbf{W}_K(k)$, that updates periodically, every T_p samples, the NGA estimate. Hence, the combined approach can be formulated as

- estimate the separating matrix $\mathbf{W}(k)$ with the NGA update equation (5)

- if $k \, mod \, T_p = 0$, $\mathbf{W}(k+1) = \mathbf{W}_K(k)$;

- else $\mathbf{W}(k+1)$ is updated by (5).

The operation of the Kalman filter is based on a feedback mechanism. At time $k-1$, equations (6) and (7) are used to predict the value of the variable to be estimated at time k. This is then adjusted using the available measurements by (8)-(10), to give the corrected estimate at that time, and the process is repeated at the next iteration. The recursive nature of the filter generally leads to very accurate estimates, fast convergence speed, and good tracking behaviour; however it also increases the computational load of the adaptive process.

The KF-NGA method exploits the advantages of KF by using the estimated sources to obtain the corrected estimate for the mixing matrix at the current iteration. Conversely, the disadvantageous increase in computational complexity is partly mitigated by the parallel implementation of NGA and KF.

4. ADAPTIVE NOISE CANCELLER

The basic ANC system and the problem it addresses are illustrated in Figure 1. The primary input is the received signal $x_p(k)$, representing the desired data $s(k)$ corrupted by additive noise $n_0(k)$, or
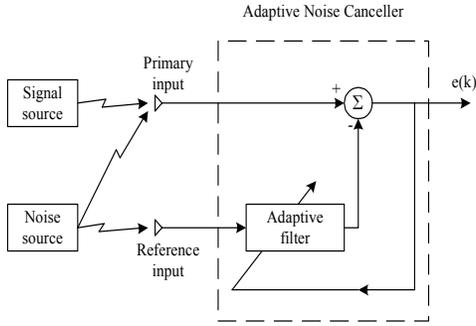
$$x_p(k) = s(k) + n_0(k) \quad (16)$$

Fig. 1. Diagram of the adaptive noise canceller and the problem it addresses, from [10].

Fig. 2. Adaptive noise canceller and NGA method.

where $n_0(k)$ is uncorrelated with the source signal $s(k)$. A second sensor provides the reference input, which is a noise source $n_1(k)$ uncorrelated with the desired source signal, but correlated in some unknown way with $n_0(k)$ [10]. Thus, the function of the adaptive filter in the ANC system is to produce an output that estimates $n_0(k)$, and that is given by [3]

$$y(k) = \sum_{i=0}^{M-1} w_i(k) n_1(k-i) \qquad (17)$$

where $w_i(k)$ are the adaptive filter weights which, like $s(k)$, $n_0(k)$, and $n_1(k)$, are real valued. The output of the system is given by the error signal

$$e(k) = x_p(k) - y(k) \qquad (18)$$

also used by the adaptive filter to adjust the weights automatically, so that its attempts to minimise the mean square error. Thus, the filter coefficients update equation is given by

$$w_i(k+1) = w_i(k) + \eta n_1(k) e(k), \ i = 0, \dots, M-1 \qquad (19)$$

where η is a step-size parameter, and M is the filter length. When several noise sources or interfering signals are present, the adaptive noise canceller can be replaced by a multi-reference system, which utilises a number of reference inputs. Each reference signal is filtered, and a linear combination of the filter outputs is subtracted from the primary input, to give the error signal.

5. ADAPTIVE NOISE CANCELLER, NATURAL GRADIENT ALGORITHM AND KALMAN FILTERING

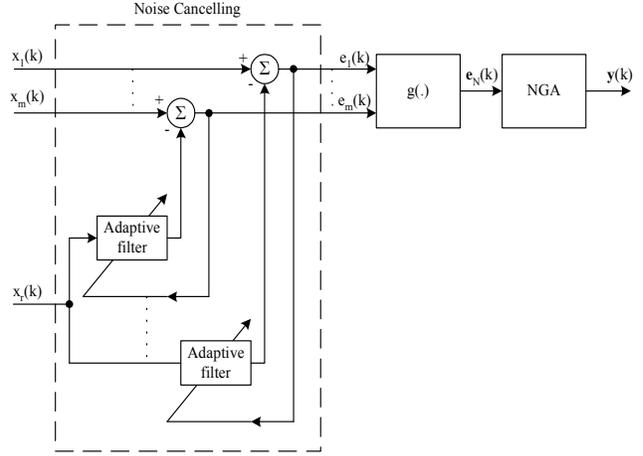In biomedical signal processing, where several primary signals are often available, a more appropriate approach to noise cancellation is to obtain several noisy recordings from the same source signal, and use one ANC for each primary input, working in parallel with the remaining adaptive noise cancellers. Since the system has a multi-channel output, a single estimate is usually selected according to some pre-defined criterion [11]. Here, however, the foetal ECG extraction problem is addressed using adaptive noise cancelling as a pre-processing operation to the natural gradient algorithm, and therefore the use of a parallel network of ANCs resulting in a vector of outputs is convenient in this case. Collectively, the set of primary inputs can be regarded as a vector of the form

$$\mathbf{x}_p(k) = \mathbf{d}(k) s(k) + \mathbf{n}_0(k) \qquad (20)$$

where $\mathbf{x}_p(k) = [x_{p_1}(k), \dots, x_{p_m}(k)]^T$, and $\mathbf{n}_0(k) = [n_{0_1}(k), \dots, n_{0_m}(k)]^T$ represents the noise vector. In addition, the vector $\mathbf{d}(k) = [d_1(k), \dots, d_m(k)]^T$ acts upon a single source signal $s(k)$ in some unknown manner, such that several observations of the same source are obtained [11]. In our implementation, one reference input $n_1(k)$ is present, which is uncorrelated with the source signal, but correlated with the noises in the primary sensors, and the operation of the adaptive filter in every ANC is memoryless, such that only one filter coefficient is adaptively changed. Thus, the signal at the output of the q-th adaptive filter is given by

$$y_q(k) = w_q(k) n_1(k) \qquad (21)$$

where $w_q(k)$ is the q-th adaptive filter weight. The expression in equation (21) effectively means that the filter output is a scaled version of the reference noise signal, which in turn implies that the underlying relationship between the reference noise and the noise in the primary input is instantaneous, an assumption that is consistent with the instantaneous mixing model used used in Section 2. Considering
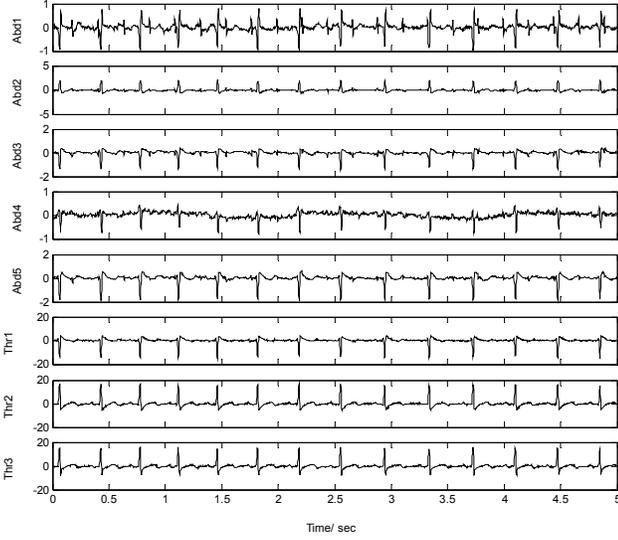
Fig. 3. An 8-channel cutaneous potential recording from a pregnant woman. The signals denoted Abd1-5 were recorded from the abdominal region, while Thr1-3 were obtained from the thoracic area.

Fig. 4. ECG components separated by the NGA algorithm when the mixtures are Abd1 - Abd3 (upper-most plots), Ab3-Ab5 (middle plots), and Abd5, Thr1 and Thr2 (lower-most plots).

the overall parallel network, an adaptive filter output vector can be defined as

$$\mathbf{y}(k) = \mathbf{w}(k) n_1(k) \tag{22}$$

where $\mathbf{y}(k) = [y_1(k), \ldots, y_m(k)]^T$, and $\mathbf{w}(k) = [w_1(k), \ldots, w_m(k)]^T$ are respectively the output and filter weight vectors. The resulting error signal vector is

$$\mathbf{e}(k) = \mathbf{x}_p(k) - \mathbf{y}(k) \tag{23}$$

and its q-th element is used to update the q-th adaptive filter weight according to

$$w_q(k+1) = w_q(k) + \eta n_1(k) e_q(k), \quad q = 0, \ldots, m \tag{24}$$

The parallel adaptive noise cancelling set-up is depicted in Figure 2. Its output, the error signal vector, is normalised such that it has unit energy on the average, and is fed to NGA instead of the mixture signals. Normalisation of the q-th element of the error signal vector is carried out according to

$$e_{q_N}(k) = g_q\left(e_q(k), \sqrt{E\left\{e_q^2(k)\right\}}\right) \tag{25}$$

where $E\left\{e_q^2(k)\right\}$ is the variance of $e_q(k)$ at the current iteration. Hence, the noise cancellers perform second-order decorrelation prior to blind source separation, which is expected to result in improved separation quality. Finally, the ANC-NGA combination replaces the natural gradient algorithm in the KF-NGA structure. We shall refer to this method as the modified KF-NGA approach.
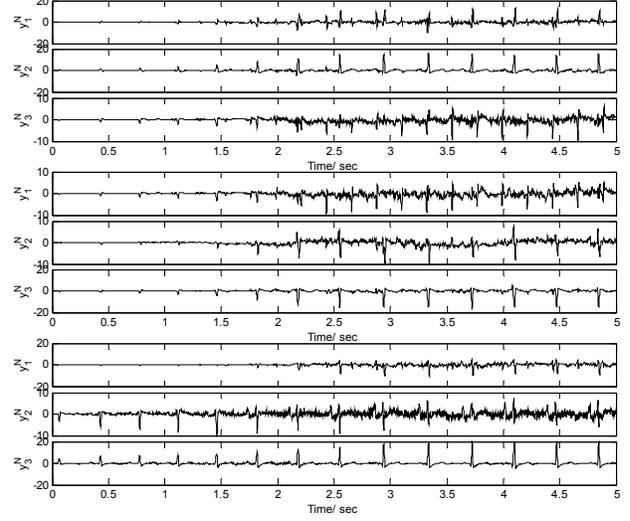
The performance of the modified KF-NGA technique is assessed by extracting the original sources from the ECG recordings Abd1 - Abd3 and Abd3 - Abd5, while for the case of several thoracic measurements and only one abdominal recording, the mixtures used are Abd5, Thr1 and Thr2 (see Figure 3). For comparison purposes, separation is firstly carried out with the conventional NGA and with the JADE algorithms, resulting in the recovered signals shown respectively in Figures 4 and 5. The results show that, although the natural gradient algorithm displays slow convergence speed, following convergence its performance is similar to that of JADE. Nevertheless, for the first set of mixtures, neither algorithm completely separates the maternal and foetal ECG components, as is evident from the output y_1^N in the three upper graphs of Figure 4, and can be seen by closely observing the JADE output y_2^J. Moreover, inspection of the three middle plots of both figures shows that y_2^J and y_2^N contain mixtures of maternal ECG and respiratory motion, an issue which in practice is irrelevant to the solution of the foetal ECG extraction problem, since y_1^J and y_1^N are good estimates of the foetal electrocardiogram, but is a poor result in the context of blind source separation for noisy mixtures. Also, neither NGA nor JADE can extract the foetal ECG when two thoracic and one abdominal measurements are available.

The elements of the normalised error vectors obtained at the output of the parallel ANC networks for the three sets of mixtures are shown in Figure 6, where the thoracic recording Thr3 is selected as the reference signal, and the step-
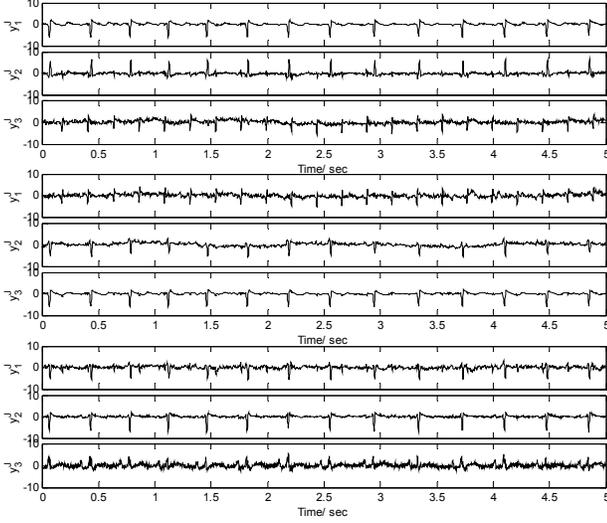
Fig. 5. Components separated with the JADE algorithm from the mixtures Abd1 - Abd3 (upper-most plots), Abd3 - Abd5 (middle plots) and Abd5, Thr1 and Thr2 (lower-most plots).

Fig. 6. Outputs of the parallel ANC network, obtained for the primary inputs Abd1 - Abd3 (upper plots), Abd3 - Abd5 (middle plots), and Abd5, Thr1 and Thr2 (lower plots). The reference signal is Thr3 in all cases.

size parameter η in (24) and for all ANCs is equal to 0.01. The results reveal that ANCs work best on the signals Abd1 - Abd3, which is not surprising since the foetal contributions can be seen clearly in the first ECG recording. Conversely, no obvious improvements in the waveform shape is obtained following de-noising of the signals Abd5, Thr1 and Thr2. In general, it is apparent that the selection of appropriate primary inputs is crucial for correct noise cancellation, and therefore good ECG extraction, to take place. More importantly, the waveforms in Figure 6 contain much larger foetal contributions than the corresponding noisy measurements in Figure 3, and therefore some improvement in the separation quality is expected to be attained when NGA operates on the error vector.

The plots in Figure 7 show the outputs of the ANC-NGA method for the three set of mixtures, and support this observation. Specifically, the maternal contributions in the y_1^{AN} and y_3^{AN} outputs in the upper sets of graphs are significantly reduced, as are the foetal contributions in y_2^{AN}, which can be seen in the corresponding NGA and JADE outputs y_2^{N} and y_2^{J}. This result is not particularly extraordinary, because the foetal ECG extracted by the noise canceller from Abd1 - Abd3 is of good quality. Conversely, when the mixtures Abd3 - Abd5 are fed to the ANCs, the foetal ECG is not immediately extracted. In fact, the improvement in separation quality provided by the ANC-NGA approach in this case is quite remarkable, since it separates the maternal ECG and the respiratory motion, a task that is not accomplished ev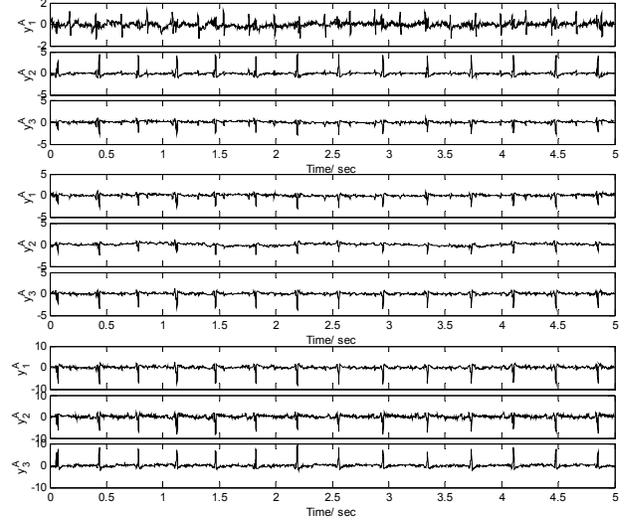en by the JADE algorithm. For the special case of one abdominal and two thoracic mixtures, the ANC-NGA combination succeeds in extracting the foetal ECG component.

Finally, the results obtained with the modified KF-NGA technique, with $T_p = 10$, if $k \leq 100$, while $T_p = 100$, if $k > 100$ for the first two sets of mixtures, and $T_p = 10$, if $k \leq 100$, while $T_p = 50$, if $k > 100$ for the last set, are illustrated in Figure 8. It is implemented because the main disadvantage of the ANC-NGA set-up is that it suffers from slow speed of convergence, which is approximately equal to that of NGA, while it has been shown in [1] that the KF-NGA approach improves considerably the convergence speed of NGA. The outputs in Figure 8 indicate that the modified KF-NGA attains a much faster speed of convergence than NGA, and it successfully extracts foetal and maternal ECGs from the mixtures Abd1 - Abd3 and FECG, MECG and respiratory motion from the mixtures Abd3 - Abd5. In addition, it succeeds in extracting the foetal ECG component from the last set of mixtures.

6. CONCLUSIONS

The foetal ECG problem has been addressed by cascading a parallel ANC network with the KF-NGA method, leading to a structure that, following convergence, can perform better than the block-based JADE algorithm. In particular, the modified KF-NGA approach offers faster speed of convergence, and improved separation quality following initial convergence.
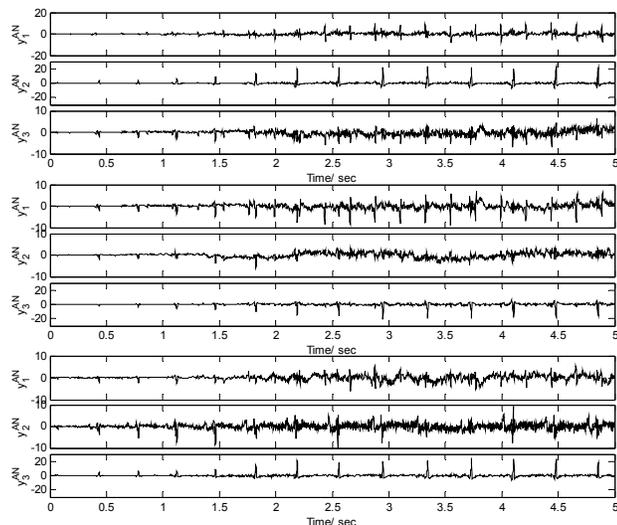
Fig. 7. Components separated with the ANC-NGA method from the mixtures Abd1 - Abd3 (upper plots), Abd3 - Abd5 (middle plots) and Abd5, Thr1 and Thr2 (lower plots).
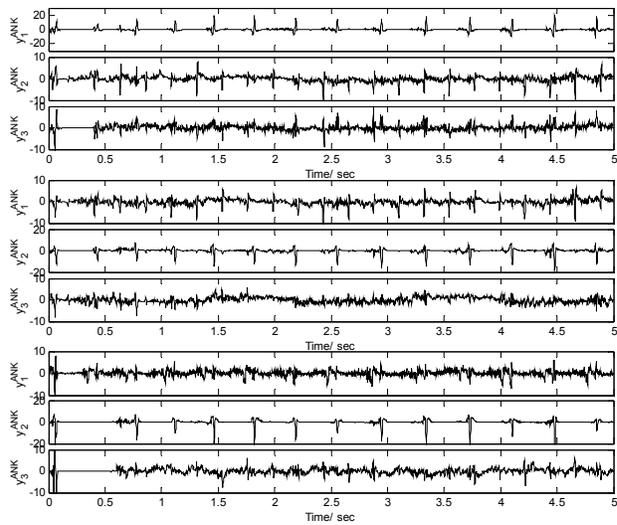

Fig. 8. Source signals recovered with the modified KF-NGA approach, for the observed signals Abd1 - Abd3 (upper-most plots), Abd3 - Abd5 (middle plots), and Abd5, Thr1 and Thr2 (lower-most plots).

7. REFERENCES

[1] M. G. Jafari, H. Seah, and J. A. Chambers, "A combined Kalman filter and natural gradient algorithm approach for blind separation of binary distributed sources in time-varying channels," in *Proc. of the IEEE Int. Conf. on Acoustic, Speech, and Signal Processing*, vol. 5, pp. 2769–2772, 2001.

[2] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings-F*, vol. 140, pp. 362–3370, 1993.

[3] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 3rd ed., 1996.

[4] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: principles and applications," *Proceedings of IEEE*, vol. 63, pp. 1692–1716, 1975.

[5] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 752–763, 1996.

[6] M. G. Jafari and J. A. Chambers, "A combined Kalman filter and natural gradient approach to blind source separation," in *Fifth IMA Int. Conf. on Mathematics in Signal Processing*, 2000.

[7] S. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches," *Proceedings of the IEEE*, vol. 86, pp. 2026–2048, 1998.

[8] A. Cichocki, B. Orsier, A. Back, and S. Amari, "On-line adaptive algorithms in non-stationary environments using a modified comjugate gradient approach," in *Proc. of the IEEE Workshop on Neural Networks for Signal Processing*, pp. 316–325, 1997.

[9] S. Kay, *Fundamentals of statistical signal processing estimation theory*. Prentice Hall, 1993.

[10] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, 1985.

[11] S. A. Vorobyov, A. Cichocki, and Y. V. Bodyanskiy, "Adaptive noise cancellation for multi-sensory signals," *Journal of Fluctuation and Noise Letters*, vol. 1, pp. 12–24, 2001.