A Comparison of Single Channel Fetal ECG Extraction Methods

Joachim Behar, Alistair Johnson, Gari D. Clifford, and Julien Oster

Intelligent Patient Monitoring Group, Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

Associate Editor K. A. Athanasiou oversaw the review of this article.

Abstract—The abdominal electrocardiogram (ECG) provides a non-invasive method for monitoring the fetal cardiac activity in pregnant women. However, the temporal and frequency overlap between the fetal ECG (FECG), the maternal ECG (MECG) and noise results in a challenging source separation problem. This work seeks to compare temporal extraction methods for extracting the fetal signal and estimating fetal heart rate. A novel method for MECG cancelation using an echo state neural network (ESN) based filtering approach was compared with the least mean square (LMS), the recursive least square (RLS) adaptive filter and template subtraction (TS) techniques. Analysis was performed using real signals from two databases composing a total of 4 h 22 min of data from nine pregnant women with 37,452 reference fetal beats. The effects of preprocessing the signals was empirically evaluated. The results demonstrate that the ESN based algorithm performs best on the test data with an F1 measure of 90.2% as compared to the LMS (87.9%), RLS (88.2%) and the TS (89.3%) techniques. Results suggest that a higher baseline wander high pass cutoff frequency than traditionally used for FECG analysis significantly increases performance for all evaluated methods. Open source code for the benchmark methods are made available to allow comparison and reproducibility on the public domain data.

Keywords—Fetal ECG, Reservoir computing, Template subtraction, Adaptive noise canceller.

INTRODUCTION

Monitoring the fetal cardiac activity may allow for screening of fetal well-being through analysis of the fetal heart rate (FHR) and morphology of the fetal electrocardiogram (FECG) waveform. For example, shortening of the fetal QT interval has been associated with intrapartum hypoxia resulting in metabolic aci-dosis.[25] Doppler ultrasound is routinely used for measuring the FHR during pregnancy and delivery,[21] even though it has not been demonstrated that ultrasound use is fully safe for the fetus.[3] Moreover, ultrasound is less accurate than FECG for tracking FHR.[12]

The FECG can be recorded in two ways; through an electrode attached to the fetal scalp while the cervix is dilated (during delivery) or non-invasively through electrodes attached to the mother's abdomen. Figure 1 shows an example of a maternal electrocardiogram (MECG), scalp FECG and an abdominal electrocardiogram (AECG) recorded simultaneously. The current primary use of the FECG is for FHR analysis during delivery. Morphological analysis of the FECG waveform is usually not performed with the exception of the STAN monitor (Neoventa Medical, Goteborg, Sweden), which attempts to identify ST segment deviation through a proxy measure (the T/R amplitude ratio). However this technology is invasive, can only be performed during labour and uses a single electrode (on the fetal scalp) which does not cover the three dimensional electrical field emanating from the fetal heart. Conversely, non invasive FECG (NI-FECG) extraction can theoretically be performed at almost any point in the pregnancy with multiple electrodes. This is true as long as the fetal dipole is large enough and that the vernix caseosa layer is not isolating the FECG signal too much during the third trimester. This motivates the research in NI-FECG extraction from the AECG which exhibits a mixture of the MECG, noise and the FECG.

Despite the rich literature on adult ECG, the significant advances in signal processing and the increased computational capabilities of digital processors, clinically useful extraction of the FECG from the mixture of abdominal signals is still a nascent field.[30] This is due to the relatively low signal to noise ratio (SNR) of

Address correspondence to Joachim Behar, Intelligent Patient Monitoring Group, Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK. Electronic mail: joachim.behar@eng.ox.ac.uk
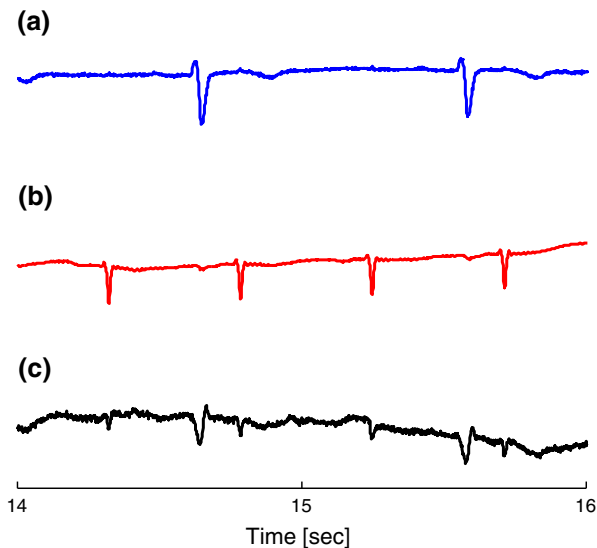
FIGURE 1. Example of (a) maternal chest ECG, (b) fetal scalp ECG and (c) abdominal ECG. Note that the abdominal ECG contains a mixture of both maternal and fetal ECG.

the FECG compared to the MECG, as well as the limited clinical knowledge on how fetal cardiac function and development map to changes in the FECG.

The primary feature that any algorithm needs to extract from the AECG signal mixture is the fetal QRS complex (FQRS) location. This peak detection is used for computing FHR, detecting rhythm abnormalities, or further used as an anchor point for extracting features from the FECG waveform. Ascertaining the location of the FQRS is simplified by first separating the FECG from the AECG, and several approaches have been previously applied. These include: principal component analysis,[19] independent component analysis (ICA),[7,16] or periodic component analysis (πCA)[31] which makes use of the ECG's periodicity. In essence, these approaches are a form of blind source, or semi-blind in the case of πCA, separation, which aim to separate the underlying statistically independent sources into three categories: MECG, FECG and noise. A key assumption of these methods is that of a linear stationary mixing matrix between these sources. Although the stationary aspect of the ICA mixing matrix could be unrealistic over long recordings, in practice the matrix can be regenerated for successive short time periods. The original signals are projected into the "source" domain, where the channels representing the MECG and noise can be canceled. The resulting back-projected signals should primarily consist of FECG components. Other techniques which operate in lower dimensions include adaptive filtering,[36] template subtraction (TS)[8,24,33] and Kalman filtering (KF).[29] See Sameni and Clifford[30] for a good overview of the methods. Despite many interesting

theoretical frameworks the robustness of most of these methods has not been sufficiently quantitatively evaluated. This is mainly due to two factors: (1) the lack of gold standard databases with expert annotations and (2) the methodology for assessing the algorithms is underdeveloped.

The echo state neural network (ESN), which belongs to the family of adaptive filtering techniques,[17] was used for the first time in the context of ECG processing by Petrenas et al.[27] for QRST cancellation during atrial fibrillation. In the present work, an FQRS detection method based on the ESN is introduced and compared to the following baseline techniques: the least mean square (LMS) adaptive filter,[36] the recursive least square (RLS) adaptive filter[36] and TS.[8,24,33] The work presented here is focused on the relative performance comparison of these single channel time-based techniques for extracting FQRS, in contrast to Blind Source Separation (BSS) techniques which can be considered as spatially based. Focusing on single abdominal channel extraction techniques (i.e., techniques that require only one abdominal channel with or without a reference chest channel) would enable the production of low-cost, easy-to-use devices for NI-FECG monitoring. In particular, it is worth noting that the methods are not compared against BSS based approaches which require many abdominal channels to be simultaneously recorded. The details for tuning the algorithms' global parameters and assessing their performance are discussed and a particular focus is given to the effect of the preprocessing step on the AECG mixture, where preprocessing refers to any signal filtering prior to applying a given FECG extraction algorithm.

MATERIALS AND METHODS

Template Subtraction

Four variants of the TS method were implemented and evaluated.[8,19,24,33] They are denoted TS_c,[8] TS_m,[24] TS_{pca},[19] and TS_{lp}[33] with TS_{lp} implemented as described in Vullings et al.[35] A MECG template cycle was built centered on the mother R-peak and considering a duration of 0.20 seconds (s) for the P wave, 0.10 s for the QRS complex and 0.40 s for the T wave. In the case of TS_c the average MECG complex $\underline{t}$ was scaled for each individual MECG cycle with a constant a in order to reduce the mismatch between the template cycle and the MECG complex $\underline{m}$. The scaling constant a was found by searching for the least mean square error (MSE) (e^2) between $\underline{m}$ and $\underline{t}$ (i.e., solving $\mathrm{argmin}\,(e^2) = \mathrm{argmin}(||\underline{t}a - \underline{m}||^2)$.[24] TS_m is similar to TS_c except that three scaling constants are searched for each individual cycle (one for each of the P, QRS and

T waves). TS$_{lp}$ was implemented as described in Vullings *et al.*[35] where the template ECG was built by weighting the seven previous cycles, where the weights selected minimized the MSE. This is in contrast to the other TS methods where the weights of the cycles contributing to the template are equal. In order for the methods to adapt to the non-stationary MECG morphology, the template was updated with incoming cycles.

PCA aims to identify a meaningful orthonormal basis to re-express a given dataset. It can be used for dimensionality reduction, source separation and visualization. In this work TS$_{pca}$, was implemented as described by Kanjilal *et al.*[19] and the first two principal components computed on the aligned ECG cycles were kept. In order for the method to be adaptive, the PCA basis was updated every 10 AECG cycles.

The LMS and RLS Adaptive Filter

The classical method for removing noise from a corrupted signal is to pass it through a filter. The filter can be fixed (i.e., its transfer function is constant) or adaptive. In the context of FECG extraction, adaptive noise cancelation[36] is commonly used to suppress noise from the mixture of signals in the AECG. The AECG $y(n)$ is treated as the sum of the signal of interest (the FECG, $s(n)$) and noise ($\eta(n)$), i.e., $y(n) = s(n) + \eta(n)$. $\eta(n)$ corresponds to the combination of the MECG, other physiological signals such as muscle noise and artifacts such as movement. As the signal recorded on the chest does not have an FECG component due to its location, it serves as an observation of the noise and as a reference for the noise canceling field. The abdominal noise $\eta(n)$ is adaptively removed by a filter whose coefficients $\mathbf{w} = [w_1, \ldots, w_N]$ form a finite impulse response filter with N being the number of coefficients or weights that are recursively updated in order to minimise an error signal $e(n)$. Thus the goal of the LMS, RLS and ESN methods is to learn a model with input $u(n)$ and output $\hat{\eta}(n)$, where $\hat{\eta}(n)$ matches the target signal $y(n)$ as closely as possible in the least MSE sense. Subtracting $\hat{\eta}(n)$ from the abdominal mixture results in the suppression of the most disrupting source of noise; the MECG. This process is represented in Fig. 2.

The LMS Adaptive Filter

The LMS adaptive filter applied to NI-FECG extraction was first published by Widrow *et al.*[36] but only qualitative results were demonstrated. LMS is used to find filter coefficients that minimise the MSE $e^2(n)$ between the filter output $\hat{\eta}(n)$ and the desired response or target $y(n)$. Let $\underline{\mathbf{u}}(n) = [u_1(n - N + 1), \ldots, u_1(n)]^T$, $\forall\, n > N$ be a segment of the input signal (with N being the
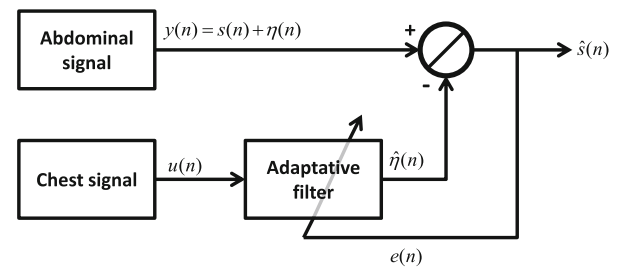

FIGURE 2. Adaptive noise canceling block diagram in the case of one reference input $u(n)$. On the diagram : the FECG $s(n)$, the noise $\eta(n)$, the abdominal ECG $y(n) = s(n) + \eta(n)$, the chest signal $u(n)$, the estimated noise $\hat{\eta}(n)$, the estimation error $e(n)$ and the output signal $\hat{s}(n)$. n corresponds to a time index.

last N input samples), $\underline{\mathbf{w}}(n) = [w_1(n), \ldots, w_N(n)]$ be the filter weights, and $e(n) = y(n) - \mathbf{w}^T\mathbf{u}(n)$ the error rate at each step. The optimal weight vector $\underline{\mathbf{w}}_o$, also called the Wiener weight vector, is given by $\underline{\mathbf{w}}_o = \boldsymbol{R}^{-1}\boldsymbol{P}$ where $\boldsymbol{R}$ is the input correlation matrix and $\boldsymbol{P}$ is the cross correlation between the desired response $y(n)$ and $\mathbf{u}(n)$. LMS algorithms aim to estimate the optimum filter weights that minimise the MSE by utilizing the gradient of the MSE at each step. The weight update equation is given by $\underline{\mathbf{w}}(n + 1) = \underline{\mathbf{w}}(n) - \frac{\mu}{2}\nabla E[e^2(n)]$, where $E[e^2(n)]$ is the expected value of the MSE and μ is the step size that controls the stability and convergence rate. By assuming that the expectation of the MSE ($E[e^2(n)]$) can be adequately approximated by a finite sample of size N (i.e., as $E[e^2(n)] = \sum_{n=1}^{N-1}[y(n) - \mathbf{w}^T\mathbf{u}(n)]^2$), this equation becomes $\underline{\mathbf{w}}(n + 1) = \underline{\mathbf{w}}(n) + \mu e(n)\underline{\mathbf{u}}(n)$. The key LMS adaptive algorithm steps can be summarized as:

$$\hat{\eta}(n) = \underline{\mathbf{w}}^T(n - 1)\underline{\mathbf{u}}(n) \tag{1}$$

$$e(n) = y(n) - \hat{\eta}(n) \tag{2}$$

$$\underline{\mathbf{w}}(n) = \underline{\mathbf{w}}(n - 1) + \mu e(n)\underline{\mathbf{u}}(n) \tag{3}$$

where (1) gives the filter prediction, (2) is used to evaluate the error and (3) is used to update the filter weights at each sample n. Note that there are two parameters to set in order to design the adaptive filter: the filter length N, and the step size μ. Both these parameters must be optimized on a training set. In the context of this work, usage of the LMS technique implicitly assumes that there is a linear relationship between the maternal waveform recorded on the chest and on the abdomen.

The RLS Adaptive Filter

The RLS algorithm minimises the total squared error between the desired signal and the target signal. In contrast to the LMS, which only considers the current error value to adapt its coefficients, the RLS

considers the total error from the beginning of the signal to the incoming data point. The forgetting factor $\lambda \in [0\ 1]$ defines what proportion of past data contribute to the filter coefficient update. In the extreme case of $\lambda = 1$ all past data contribute equally, and as λ approaches zero only the most recent data points play a role. This translates into finding the parameters so that the following "loss-function" $\epsilon(n)$ is minimized:

$$\epsilon(n) = \sum_{i=1}^{n} \beta(n,i)e(i)^2 = \sum_{i=1}^{n} \beta(n,i)[y(i) - \hat{\eta}(n,i)]^2$$

with $\hat{\eta}(n,i) = \mathbf{w}^T(n)\underline{u}(i)$ and where $\beta(n,i) = \lambda^{n-i}$ in the case of the exponentially weighted least squares solution. The RLS algorithm updates the filter coefficients at each iteration[2,34] as follows:

$$e(n) = y(n) - \mathbf{w}^T(n-1)\underline{u}(n)$$
$$\underline{k}(n) = \frac{\mathbf{P(n-1)u(n)}}{\lambda + \underline{u}(n)^T \mathbf{P(n-1)u(n)}}$$
$$\mathbf{P(n)} = (\mathbf{I} - \underline{k}\ \mathbf{u(n)^T})\mathbf{P(n-1)}\frac{1}{\lambda}$$
$$\mathbf{w}(n) = \mathbf{w}(n-1) + \underline{k}(n)e(n),$$

where $\mathbf{P}$ is the covariance and $\mathbf{I}$ is the identity matrix. There are two important parameters in the RLS: the forgetting factor λ and the number of filter coefficients N. RLS tends to converge faster than LMS and is usually more accurate. However, this is at the price of a higher computational complexity. In the context of this work and similar to the LMS approach, RLS assumes that a linear relationship exists between the maternal waveform recorded on the chest and the abdomen.

The ESN

Recurrent neural networks (RNN) are a class of neural networks capable of non-linear modeling of dynamical systems. This is made possible by recurrent connections between neurons, visualized as cycles in the network topology, that allow processing of temporal dependencies.[9] However, parameter estimation of the RNNs has proven to be a difficult task. Indeed, optimization methods that were originally used for training feedforward neural networks, such as the error backpropagation algorithm, do not generally perform as well for training RNNs.[23] The ESN[17] is a recently introduced approach to RNN training, with the RNN (or reservoir) being generated randomly. The reservoir is then fixed and only the weights of the output neurons are learnt and updated using online or offline linear regression. This method outperformed classic fully trained RNNs in many tasks.[23] The ESN is introduced in the context of noise canceling as a nonlinear medium into which the reference signals propagate before the "echo response" given by the network reservoir is weighted by a readout layer.

In the configuration presented here, the MECG recorded on the chest is projected onto a set of non-orthogonal basis functions through the ESN reservoir, comparable to the kernel in kernel learning approaches. The input signal(s) (chest ECG) drive the nonlinear reservoir resulting in a high-dimensional dynamical "echo response."[9] The reservoir also acts as a memory of the input signal thus providing temporal context.[22] In a second step an adaptation algorithm is used to compute the weights of the output neurons. The RLS algorithm was used for this step. This readout layer maps the reservoir states to the output: the observed abdominal AECG.

For K input units, M internal units and L output units: $\underline{u}(n) = [u_1(n), \ldots, u_K(n)]$, $\underline{x}(n) = [x_1(n), \ldots, x_M(n)]$, $\hat{\underline{\eta}}(n) = [\hat{\eta}_1(n), \ldots, \hat{\eta}_L(n)]$, where $\underline{x}(n)$ is the reservoir state vector, $\underline{u}(n)$ is the vector of input signals and $\hat{\eta}(n)$ is the vector of output signals. We define the extended system state as $\underline{z}(n) = [\underline{x}(n)\ \underline{u}(n)]$. The activation of internal units is updated using the following equation:

$$\underline{x}(n+1) = \underline{x}(n) + f(\mathbf{W}\underline{x}(\mathbf{n}) + \mathbf{W_i}\underline{u}(n+1) + \mathbf{W_b}\hat{\underline{\eta}}),$$

where $\mathbf{W} \in \Re^{M \times M}$ is the reservoir weight matrix and $\mathbf{W_i} \in \Re^{M \times K}$ is the input weight matrix (randomly generated and fixed). $\mathbf{W_b} \in \Re^{M \times L}$ is the back projection weight matrix and f is the reservoir neuron activation function, taken to be the hyperbolic tangent. Considering a purely input-driven dynamical pattern recognition task, the system is simplified by setting $\mathbf{W_b} = \mathbf{0}$. The output is computed as:

$$\hat{\underline{\eta}}(n+1) = g(\underline{w}_o(n)\underline{z}(n)),$$

where g is the output neuron activation function (taken to be identity) and $\underline{w}_o$ are the output weights which may be adaptive or fixed. An ESN with a leaky integrator neuron model[18] was used. This satisfies:

$$\underline{x}(n+1) = (1-\alpha)\underline{x}(n) + f(\mathbf{W}\underline{x}(\mathbf{n}) + \mathbf{W_i}\underline{u}(\mathbf{n+1})), \tag{4}$$

with $\alpha \in [0\ 1]$ being the leakage rate or forgetting factor. In the case of $\alpha = 1$, the neurons do not retain any information about their previous state. Initial weights were generated from the uniform distribution on the interval $[-1\ 1]$ for $\mathbf{W_i}$ and $\underline{w}_o$. $\mathbf{W}$ is a random $M \times M$ sparse matrix with approximately $\psi \times M \times M$ uniformly distributed non zero entries ($\psi \in [0\ 1]$ is the sparsity of the reservoir). Figure 3 shows a representation of the ESN based FECG extraction algorithm. In the figure, one chest signals $u(n)$ is used as inputs of the ESN and the abdominal channel is used as the target signal. The reservoir and input weights are randomly initialized once and the generated network is
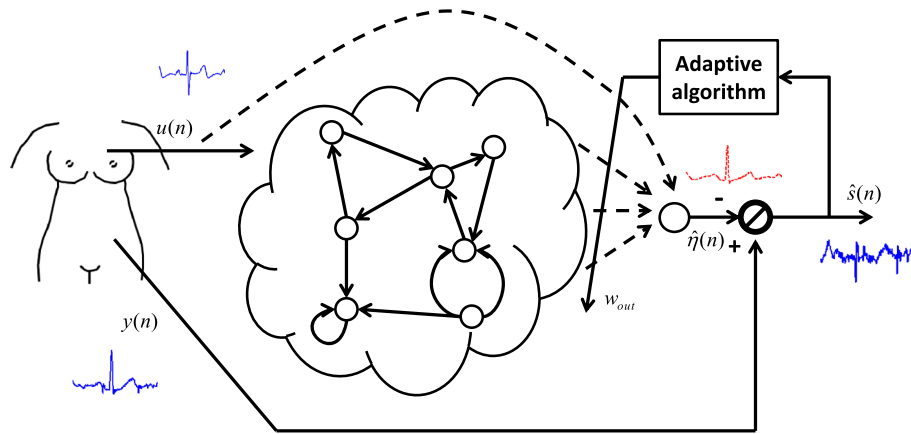
FIGURE 3. The ESN based FECG extraction algorithm showing the relationship between the chest signal $u(n)$, predicted signal $\hat{\eta}(n)$, the jth abdominal ECG signal $y(n)$, and the residual signal $\hat{s}(n)$ containing the FECG. Dashed lines represent adaptive weights. Image of the women is adapted from Zaunseder *et al.*[37].

TABLE 1. Parameters of the ESN, LMS and RLS algorithms.

	Symbol	Definition	Grid search range (step size)
ESN	f_b	High frequency cut-off	[30, 120] (5)
	f_h	Low frequency cut-off	[1, 49] (3)
	M	Size/number of units of the reservoir	[10, 250] (20)
	ψ	Sparsity of the reservoir. Percentage of connections between nodes of the reservoir	20%[†]
	ρ	Spectral radius of W, the reservoir connection matrix	[0, 1] (0.05)
	γ	Input scaling of the input weight matrix W_{in}	1[†]
	α	Leakage rate	[0, 1] (0.1)
	λ	Forgetting factor of the RLS algorithm	0.999
LMS	N	Filter length	[1, 262] (5)
	μ	Step size	[0.01, 0.46] (0.05)
RLS	N	Filter length	[1, 61] (3)
	λ	Forgetting factor	[0.8, 1] (0.0025)

[†]Parameters that were not optimized.

used independently for each abdominal channel available of each individual record. The predicted signal $\hat{\eta}(n)$ is then subtracted from the jth abdominal signal $y_j(n)$ giving the residual signal $\hat{s}(n)$ containing the FECG. Using the ESN for NI-FECG extraction allows for a non-linear relationship between the maternal waveform as recorded on the chest and recorded on the abdomen.

There are a number of global parameters that have an important influence on the algorithm's performance. The main three global parameters of the ESN are[22]: (1) input scaling γ, (2) the spectral radius ρ and (3) the leakage rate. The spectral radius determines how fast the influence of an input disappears in the reservoir with respect to time: choosing a small ρ means that the output is more dependent on recent history.[18] It also affects how stable the reservoir acti-

vations are (see Eq. (4)). $\mathbf{W}$ is first rescaled by $\rho(\mathbf{W})$ (i.e., its dominant eigenvalue) and therefore has unit spectral radius. Next $\mathbf{W}$ is scaled by ρ. Input scaling determines the degree of non-linearity of the reservoir responses. Normalizing the input signal so that it lies in the range $[-1\ 1]$ plays a similar role as scaling $\mathbf{W_i}$ (see Eq. (4)). Table 1 summarises the ESN parameters.

The weights can be allowed to evolve (adaptive filtering) or can be fixed (initialized on a sub-segment of the signal and kept constant). Both options were explored for the ESN approach; these will be denoted ESN_a for the adaptive (i.e., weights are updated online) approach and ESN_{na} for the nonadaptive (i.e., weights are determined on some initial training data and kept constant) approaches. This allows for assessment of whether online updating of the filter coefficients improves the filter performance.

Databases and Evaluation Protocol

Databases

Two databases were used in this study. The first was the Physionet non-invasive FECG database (PNIFECGDB)[14] consisting of 55 multichannel AECG recordings taken from a single subject between 21 and 40 weeks of fetal gestational age. Each record consisted of two chest channels and 3–4 abdominal channels with different electrode configurations (electrode position was varied in order to improve the SNR). All signals were sampled at 1 kHz with 16-bit resolution. A bandpass filter (0.01–100 Hz) and a notch filter (50 Hz) were applied to the data during acquisition. A total of 14 records were manually selected where the FQRS complexes were visible on at least one channel. The gestational age for these records ranged from 22 to 40 weeks. One minute of signal, 30 s after the start of the record, was extracted for all available channels for each record.[1] A total of 2148 FQRS were manually annotated by the first author (with three years of FECG analysis experience) using the channel where the FQRS appeared to be the most visible for each record. These markers were considered to be the reference. For consistency, only the first three abdominal channels were kept so that all records had two chest and three abdominal channels. The three abdominal channels were considered independently for FECG extraction thus providing $3 \times 14 = 42$ min of annotated data (i.e., $3 \times 2148 = 6444$ reference FQRS). Data were of good quality with variable FECG/MECG SNR and some minor artifacts which were not manually discarded. This first database was denoted DB_1.

The second database consisted of a subset of records from a private commercial database.[2] Each record consisted of 28 abdominal channels, one maternal chest channel, and the invasive scalp FECG signal. All signals were sampled at 1 kHz with 16-bit resolution. The chest channel as well as a subset of four abdominal channels where a FECG trace was visible were manually selected for processing. A total of eleven 5 min records from 8 pregnant women were used. An energy QRS detector based upon that of Pan and Tompkins (P&T)[26] was applied to the scalp electrode and the corresponding markers were used as the reference. All records' abdominal channels were considered independently for FECG extraction thus providing $11 \times 5 \times 4 = 3$ h and 40 min of annotated data (overall $4 \times 7752 = 31{,}008$ reference FQRS annotations). Some artifacts were present in a few records and the FECG/MECG SNR was variable. This second database was denoted DB_2.

All ECG signals were downsampled to 250 Hz with an anti-aliasing filter prior to running extraction algorithms and tuning the parameters. In the following work, DB_1 constitutes the training database and DB_2 the test database. Summary statistics for the fetal and maternal HR in DB_2 are reported in Table 2. Note that the data in the training set (DB_1) and the test set (DB_2) were recorded with different hardwares, following a separate protocol and at different stage of pregnancy for different subjects. This allows for reassessment of whether the algorithms are sufficiently flexible to work with data that have similar, but not equivalent, recording configurations. It is particularly pertinent to assess the adaptability of the extraction methods to an unseen set of signals from distinct subjects as there are many free parameters which are tuned on the training set database. Furthermore, the PNIFECGDB was used as the training set because the reference FQRS fiducial markers were not as accurate as for the private database which had a simultaneous fetal scalp signal.

Because only one chest channel was available in DB_2 only one of the two available chest channel was used for DB_1. Each abdominal channel were considered individually. Thus $\mathbf{u}(n) = u(n)$ and $\hat{\boldsymbol{\eta}}(n) = \hat{\eta}(n)$.

Evaluation Protocol

QRS detectors' performances are usually assessed by beat-to-beat comparisons between the detected beats and the reference beats. The classical adult matching window for candidate fiducial points is 150 milliseconds (ms).[1] However, in order to account for the higher FHR, a matching window of 50 ms is commonly employed, as for example in Guerrero-Martinez et al.[15] and Zaunseder et al.[37] In accordance with the ANSI/AAMI guideline[1] the sensitivity (Se) and positive predictive value (PPV) are reported as follows: $Se = TP/(TP + FN)$, $PPV = TP/(TP + FP)$, where TP, FP and FN are true positive, false positive and false negative detections respectively. For algorithm parameter optimization the following performance index (PI), was suggested by Kotas et al.[20] in the context of FQRS detection: $PI = (T - FN - FP)/T = (TP - FP)/(TP + FN)$, where $T = TP + FN$ is the number of annotated FQRS. The F_1-measure could alternatively be used as a measure of an algorithm's accuracy. In the context of binary classification:

$$F_1 = 2 \cdot \frac{PPV \cdot Se}{PPV + Se} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP}. \quad (5)$$

---

[1]The set of the 14 records selected from the PNIFECGDB was: 154, 192, 244, 274, 290, 323, 368, 444, 597, 733, 746, 811, 826, 906.

[2]The study was approved by The Institutional Review Boards at Summa Health System (RP#12018) and Brigham and Women's Hospital (RP#2010-P-002778/1).

TABLE 2. HR reference statistics for DB$_2$ records.

	FHR stats (bpm)				MHR stats (bpm)			
	Mean	SD	Max	Min	Mean	SD	Max	Min
123a	130	5	139	119	66.7	2.74	76.3	61.2
123b	131	6.19	144	117	63.8	2.12	68.5	59.2
125	145	5.44	147	126	80.9	2.77	87.3	74.8
126	159	2.43	163	155	78.9	5.36	96	69.5
172a	154	6.35	159	135	91.5	4.32	102	83.6
172b	142	10.8	153	112	93.6	6.6	109	81.0
210	157	5.51	170	145	90.8	6.77	108	80.2
261	132	3.54	142	122	79.3	2.69	89.7	74.2
265	129	2.57	135	125	64.9	3.42	74	59.3
299a	148	6.92	159	123	83.4	5.65	100	74.8
299b	147	13.1	157	95.5	80.2	5.57	97.2	72.1

All the values are given use trimmed estimates of the FHR distribution, i.e., they exclude the most extreme values (2.5% on either tail). This is due to inherent artifacts in the data causing misleading statistics if not excluded.

From Eq. (5) one can observe that *FN* and *FP* play a symmetric role in penalizing the accuracy measure F_1, which is not the case when considering the *PI* measure. Thus in "Results" section *Se*, *PPV* and F_1 statistics are reported and the F_1 measure was used for parameter optimization. In addition, the performance of the algorithms are evaluated in terms of FHR. FHR can be derived from the RR interval time series and compared with the reference trace HR derived from the reference annotations. Both the FHR and the reference HR were extracted as the reciprocal of the RR interval scaled by a factor of 60. At any given time, the extracted FHR is said to match the reference FHR if it is within ±5 beats per minute (bpm) with respect to the reference trace. The corresponding FHR measure is denoted HR_m.

Preprocessing and General Experimental Set Up

For each of the available abdominal channels, baseline wander was first removed. In the context of NI-FECG extraction, it is common to use a larger than standard low frequency cut-off before performing MECG cancelation (see Martens *et al.*[24] and Zaunseder *et al.*[37] for example), where larger is ~2 Hz. This cut-off is acceptable when the aim of the analysis is not FECG morphology, yet there is no comprehensive study known to the authors that has assessed the prefiltering effect on outcomes for FQRS extraction. As a consequence an exhaustive search was conducted to determine f_b (an "optimal" cut-off frequency to remove the baseline using a high pass filter) and f_h (an "optimal" cut-off frequency to remove the high frequency content using a low pass filter). Two zero phase Butterworth digital filters were cascaded for that purpose: one sixth order high pass filter and one 10th order low pass filter. The reference and input signals were then normalized according to the following procedure: (1) the first 5 s of the signal were used to derive the amplitude range of the ECG signal and the signal was then divided by this value, (2) the mean was computed over the first 5 s and subtracted from the ESN signal, and (3) the resulting signal was transformed using the hyperbolic tangent function. Step (3) was applied in order to avoid outliers which could result in the reservoir state $\underline{\mathbf{x}}(n)$ or LMS/RLS weights $\underline{\mathbf{w}}(n)$ taking unexpected values due to abnormally large values (likely attributable to signal artefact). Not performing this last step could result in loss of useful memory or a highly unpredictive output.[22] Following the preprocessing, each of the algorithms was utilized to filter out the MECG resulting in a residual signal comprising of the FECG and some noise. FQRS detection was performed on the residual signal $\hat{s}$ using a P&T QRS detector with 150 ms refractory period. All the parameter optimization was performed on DB$_1$ while DB$_2$ was used as the independent test set. The *Se*, *PPV*, F_1 and HR_m of each algorithm was evaluated.

TS techniques MQRS detection was run on the chest channel because of its higher SNR and negligible FECG contribution. Each MQRS was then adjusted in order for the R-peaks to be accurately located on each abdominal channel. This was to ensure good construction of the channel-specific MECG template and good alignment between the ECG cycles and the template MECG.

LMS technique An exhaustive search over a range of values of $N \in [1\ 262]$ and $\mu \in [0.01\ 0.46]$ was performed.

RLS technique An exhaustive search over a range of values of $N \in [1\ 61])$ and $\lambda \in [0.8\ 1]$ was performed.

TABLE 3. Overall statistics on DB_1 (training DB) and DB_2 (test DB).

	Se (%)	PPV (%)	F1 (%)	HR_m (%)
TS_c-DB_1	91.2	90.5	90.8	57.9
TS_m-DB_1	90.3	90.0	90.1	56.7
TS_{lp}-DB_1	91.2	92.8	92.0	59.7
TS_{pca}-DB_1	94.7	96.0	95.4	68.7
LMS-DB_1	95.8	95.0	95.4	69.3
RLS-DB_1	96.2	95.6	95.9	70.6
ESN_{na}-DB_1	97.2	97.3	<u>97.2</u>	73.5
ESN_a-DB_1	96.8	97.2	97.0	72.2
TS_c-DB_2	86.8	85.3	86.0	68.1
TS_m-DB_2	86.4	85.2	85.8	67.2
TS_{lp}-DB_2	85.9	85.5	85.7	66.7
TS_{pca}-DB_2	89.9	88.8	89.3	73.9
LMS-DB_2	89.3	86.5	87.9	73.7
RLS-DB_2	89.7	86.8	88.2	74.5
ESN_{na}-DB_2	89.3	86.5	87.9	73.9
ESN_a-DB_2	91.4	88.9	<u>90.2</u>	78.7

Best results in term of F1 and HR_m are underlined.

"a" stands for adaptive filtering and "na" stands for non adaptive filtering. In the case of offline training the weights are determined on the 30 s of signal preceding the studied ECG segment (for both DB_1 and DB_2).

ESN technique The optimization of the ESN parameters was performed using stage wise grid search as follows: (i) a standard prefiltering range using $f_b = 2$ and $f_h = 100$ Hz was used to search for workable ρ and α values, (ii) given ρ and α, a search was performed over a range of preprocessing cut-off frequencies to find optimal f_b and f_h, (iii) using the optimized f_b and f_h, the search for optimal values of ρ and α was repeated, and finally (iv) the number of reservoir neurons (M) was optimized as measured by the F_1 score with the parameters selected in (i–iii). Note that (i–iii) were performed with a high number of reservoir neurons ($M = 250$). In the case of adaptive filtering, an RLS algorithm was used to update the weights of the readout layer which maps the reservoir states to the output of the observed AECG. In the case of non-adaptive filtering, the Wiener–Hopf, direct pseudo inverse and ridge regression were considered to determine the weights on a 30 s signal epoch preceding the studied segments. When varying the number of neurons, M, the experiment was repeated 20 times in order to ensure the variance of the F_1 measure caused by the random initialization of the ESN reservoir connections and weights was negligible.

One chest channel was taken as the reference signal for the LMS, RLS and the ESN. Filter weights were initialized on 30 s preceding the annotated data for each record (i.e., preceding the 1 min of DB_1 and the 5 min of DB_2).

Adding Signal Quality Indices

In Clifford *et al.*[11] different signal quality indices (SQIs) to identify bad quality ECG signals were eval-uated. The most accurate SQI evaluates the agreement between two QRS detectors with different robustness to noise. This metric, termed bSQI $\in$ [0 1] (with 1 representing a good quality signal), was difficult to apply to the abdominal signal as the MQRS and also some FQRS could be detected by the QRS detectors in the case of high amplitude FECG traces. This makes bSQI a weak quality indicator if applied to the abdominal signal. Thus no bSQI was used on this signal but was used on the chest channel (where there is no FECG contribution) with a 10 s window and 9 s overlap. This provided a second by second SQI for the chest channel. In the case where bSQI was inferior to 0.8 the corresponding abdominal segments were removed.

RESULTS

Parameter Optimization

TS

For TS, the TS_{pca} method gave the best results on the training set (see Table 3), thus only results for TS_{pca} were reported in Table 4 for the performance on the individual records of DB_2.

ESN

Figure 4 illustrates the exhaustive search results obtained on the training database DB_1 using the direct pseudoinverse method for computing **w**. Based on this optimization step, the preprocessing and ESN parameters

TABLE 4. Per record statistics on DB$_2$ (test DB).

Rec	TS$_{pca}$		LMS		RLS		ESN$_a$	
	F_1	HR_m	F_1	HR_m	F_1	HR_m	F_1	HR_m
123a	84.1	51.7	83.4	63.8	83.3	51.7	<u>84.3</u>	65.0
123b	95.9	88.4	<u>96.3</u>	92.4	<u>96.3</u>	92.5	95.7	90.2
125	87.2	71.6	78.7	54.5	80.7	59.3	<u>88.2</u>	75.6
126	86.2	69.4	88.4	75.8	88.2	76.1	<u>88.9</u>	76.5
172a	98.1	90.2	99.1	96.4	<u>99.3</u>	97.8	99.2	97.3
172b	94.1	85.4	94.1	88.7	94.5	89.9	<u>94.7</u>	90.7
210	86.2	63.4	84.2	51.2	85.4	62.0	<u>88.9</u>	70.8
261	<u>77.2</u>	58.7	73.6	53.1	73.5	52.9	75.9	56.4
265	77.4	54.5	70.7	42.9	70.5	43.0	<u>78.6</u>	56.7
299a	98.0	89.0	98.9	95.3	<u>99.0</u>	96.8	98.0	93.8
299b	97.4	90.5	<u>98.8</u>	96.7	<u>98.8</u>	97.2	98.0	93.2

Best results in terms of the F_1 measure are underlined.

were determined to be: $f_b = 20$ Hz, $f_h = 95$ Hz, $\alpha = 0.4$, $\rho = 0.4$, $M = 90$. The best results on the training set using the above mentioned parameters were $Se = 97.2\%$, $PPV = 97.3\%$ and $F_1 = 97.2\%$ (see Table 3).

LMS

The parameters μ and N of the adaptive LMS adaptive algorithm were searched (Fig. 4). Selected parameters based on the grid search were $\mu = 0.1$ and $N = 20$ with corresponding performance statistics $Se = 95.8\%$, $PPV = 95.0\%$ and $F_1 = 95.4\%$ (see Table 3).

RLS

The parameters λ and N of the RLS adaptive algorithm were searched (Fig. 4). Parameter values selected based on the grid search were $\lambda = 0.999$ and $N = 20$, with corresponding performance statistics of $Se = 96.2\%$, $PPV = 95.6\%$ and $F_1 = 95.9\%$ (see Table 3).

Qualitative Results

Results in Fig. 5 were produced with $f_b = 20$ Hz, $f_h = 95$ Hz. Figure 5b shows qualitative examples of the ESN algorithms performance on r154 in DB$_1$ with optimal parameters.

Results on Training and Test Sets

Table 3 summarises the results obtained on the training set DB$_1$. It can be seen that ESN based techniques performed better than the TS and adaptive filter techniques. Both ESN techniques gave similar results, with F_1 scores of 97.2% for ESN$_{na}$ and 97% for ESN$_a$. Among the TS techniques, TS$_{pca}$ gave $F_1 = 95.4\%$ and outperformed all other TS techniques. Finally, among

the adaptive filter techniques the LMS (F_1 of 95.4%) was slightly outperformed by the RLS (F_1 of 95.9%).

Table 3 also summarises the results obtained on the test set DB$_2$. The adaptive ESN (ESN$_a$) gave the best F_1 score of 90.2%, improving upon the TS$_{pca}$ technique (89.3%) and the RLS technique (88.2%). Table 4 presents the results for each individual record of DB$_2$ for the optimal ESN, LMS, RLS and TS$_{pca}$ as determined on the training set DB$_1$. Note that in Table 4 the performance of each algorithm was averaged over all individual channels of a given record. The best result in each case is underlined.

Results with Signal Quality Indices

Table 5 presents the results for the test database in terms of F_1; Including the signal quality indices improved the results on DB$_2$ by +1.07 for the TS$_{pca}$ approach, +0.95% for the LMS approach, +1.04% for the RLS approach and +1.11% for the ESN approach while excluding 3.6% of the overall signals length. See Fig. 6 for an example of segment excluded on DB$_2$ using bSQI.

DISCUSSION

The focus of this study was to benchmark the ESN, RLS, LMS and TS techniques in their capacity to accurately extract the FECG from the AECG and facilitate robust FQRS detection. Better results may be achieved with an alternative QRS detector, specifically designed for FQRS detection as in Kotas et al.,[20] but the relative ordering of the algorithms is unlikely to change.

It should be noted that in practice the FHR time series are smoothed prior to being displayed on the clinical monitor. This smoothing operation removes
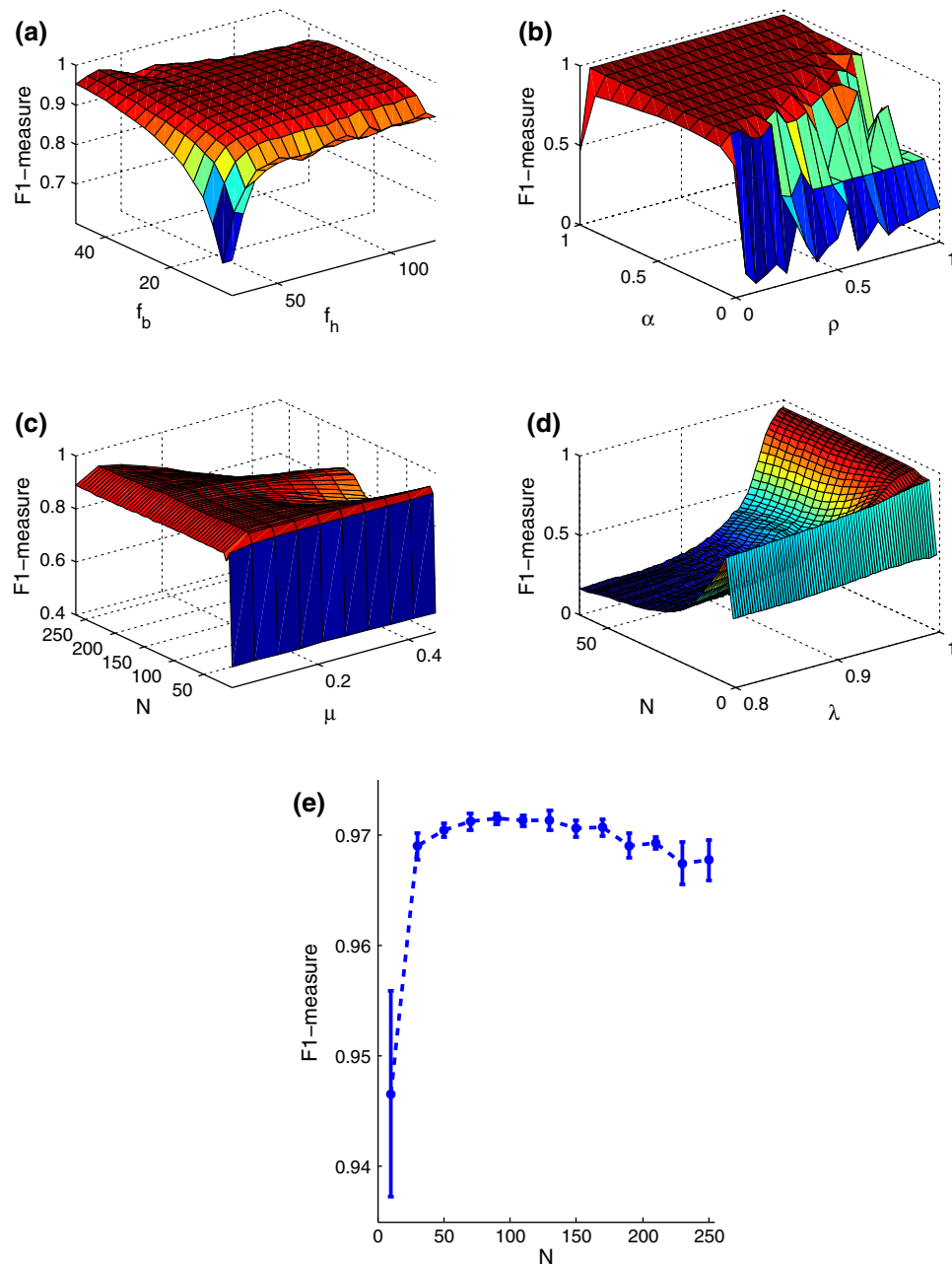
FIGURE 4. Search space for the preprocessing parameters (f_b, f_h) the ESN specific parameters (α, ρ) and the LMS parameters (μ, N) on the training set database DB_1. (a) search for ESN preprocessing parameters ($\alpha = 0.4$, $\rho = 0.4$, $M = 90$), (b) search for ESN parameters α and ρ ($f_b = 20$ Hz, $f_h = 95$ Hz, $M = 90$), (c) search RLS parameters λ and N ($f_b = 20$ Hz, $f_h = 95$ Hz), (d) search LMS parameters μ and N ($f_b = 20$ Hz, $f_h = 95$ Hz), (e) for the number of ESN neurons required ($f_b = 20$ Hz, $f_h = 95$ Hz, $\alpha = 0.4$, $\rho = 0.4$- repeated 20 times for each value of N to look at the variance of the F_1 measure caused by the random initialization of the ESN reservoir connections and weights).

sudden drops or increases of FHR due to missed beat detections. Higher performances would be expected on DB_1 and DB_2 in terms of the FH_m if this smoothing operation was applied.

Studying the performance of the algorithms on individual channels without smoothing the extracted FQRS time series allowed us to draw a direct comparison between the various time based techniques. Blind source separation techniques such as ICA or PCA were not considered in this work as its focus was on time based techniques using a single abdominal channel. However, in a multi abdominal channel system any of the presented methods could be used to remove the MECG contribution on the different
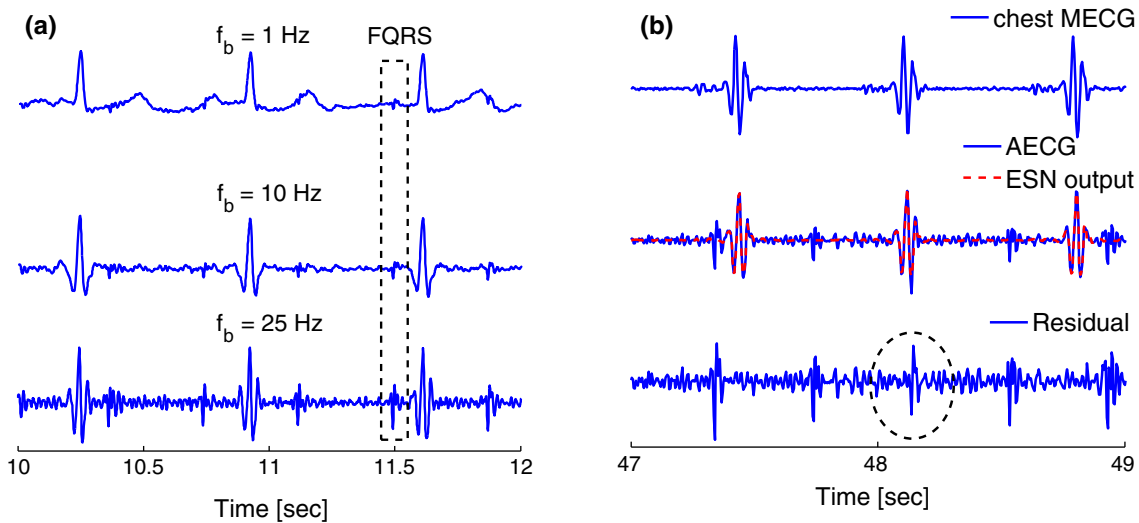
FIGURE 5. (a) Effect of varying f_b (r154, DB$_1$, 3rd AECG, f_h = 110 Hz) for the preprocessing step. Note that at 10 Hz most of the frequency content of the P and the T-wave of the MECG have been filtered out, leaving only the FQRS and the MQRS. (b) Example of the ESN algorithm performance with α = 0.4, ρ = 0.4, M = 90 (optimal parameters). The signal (r154, DB$_1$, third AECG) was prefiltered with f_b = 20 Hz and f_h = 95 Hz. Notice in particular the extraction of the FQRS embedded in the MQRS at t = 48.2 sec (circled by a broken black line).

TABLE 5. Results on DB$_2$ both with and without removing low quality record segments based upon a single SQI.

	F_1 no SQI (%)	F_1 SQI[a] (%)	ΔF_1 (%)
TS$_{pca}$	89.32	90.39	+1.07
LMS	87.88	88.83	+0.95
RLS	88.23	89.27	+1.04
ESN$_a$	90.17	91.28	+1.11

[a] With 3.6% of the overall signals being removed.

channels independently before performing the BSS step, as for example in Behar et al.[5] where higher performance was obtained by adding a BSS step after any of the single channel time based approaches was used. In addition results in Table 4 showed that no algorithm was systematically outperforming the others. This suggests that there is potential to combine different approaches for NI-FECG extraction and further improve the overall results.

The importance of the signal preconditioning and in particular the baseline wander cut-off frequency was studied. Figure 5 shows an example of a segment (r154, DB$_1$, 3rd AECG) which was preprocessed with various different f_b values. Note that at 10 Hz most of the frequency content of the P-wave and the T-wave of the MECG have been filtered out leaving only the FQRS and MQRS. This higher cut-off showed improvement on both training and test sets for all methods. As an example, the TS$_{lp}$ technique gave $F_1 = 83.0\%$ and $F_1 = 80.0\%$ with $f_b = 2$ Hz, $f_h = 95$ Hz on DB$_1$ and DB$_2$ respectively and $F_1 = 92.0\%$ and $F_1 = 85.7\%$ with $f_b = 20$ Hz, $f_h = 95$ Hz on DB$_1$ and DB$_2$ respectively[3]. The cut-off selected in this work was inferred from the grid search in Fig. 4a. Although it was clear that choosing f_b larger than traditional used was beneficial in this case, whether to choose $f_b = 20$ Hz (as in this work) or a lower value inevitably depends on the dataset and filter design considered.

One of the theoretical limitations of such adaptive filtering approaches is that any noise contained in the abdominal, but not the chest ECG signal, will not be removed. Indeed, the assumption behind adaptive noise canceling is that the noise contaminants on the abdominal channels are also present on the chest channels (considered as the noise field). Nevertheless the method is well suited for removing the main noise contaminant from the abdominal signal, namely the MECG, and easing FQRS detection.

The ESN algorithm requires a minimum of one reference and one abdominal channel. This is an important advantage over the blind source separation techniques which while very popular for this application, require a relatively high number of channels (usually between 8 and 16[29]). Furthermore, the ESN does not require any prior information on the MQRS location, unlike the TS or KF approaches. In partic-

---

[3]It is less meaningful to report a direct comparison of prefiltering effects for the ESN and LMS since parameters have been optimized for these two algorithms based on a given prefiltering ($f_b = 20$ Hz, $f_h = 95$ Hz).
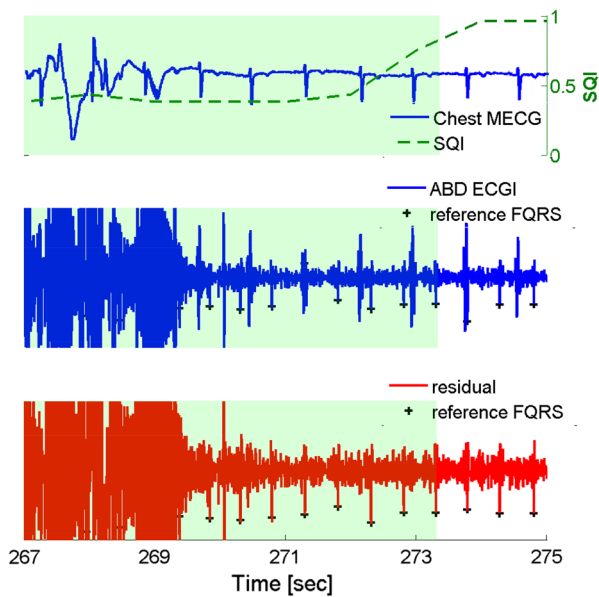
FIGURE 6. Signal quality for identifying bad quality regions (record 123a first abdominal channel). From top to bottom: chest MECG, ABD ECG and residual after performing MECG cancelation using the ESN. In the faded colour area, the part of the signal that would be discarded because of its low quality (SQI ≤ 0.8-window size for assessing quality is 10 s with 9 s overlap).

ular, the KF requires a precise MQRS detection technique, making this technique particularly susceptible to noise in the MECG.

When using the ESN_a with the traditional RLS implementation presented in Sect. 2.2.2 used to train the readout layer, we observed a monotonic increase of the weights over time leading to large weights after a few minutes. When the weights are too large they amplify small differences among the dimensions of $\underline{x}(n)$ which in turn can lead to instability in the presence of a small deviation from the conditions for which the network had been trained.[23] As a consequence we implemented the RLS algorithm introduced in Douglas et al.,[13] which uses RLS in combination with least squares prewhitening. This implementation made the ESN_a stable.

Within the TS techniques, TS_{pca} performed the best on the training and test sets. Within the TS class of techniques used TS_{pca} is certainly the most adaptive. Within the other methods (LMS, RLS, ESN), RLS was found to perform slightly better than LMS and the ESN better than the RLS. This suggests that the MECG cycle was better removed with more adaptive algorithms.

It is important to note that the TS techniques could include discontinuities due to the piece-wise template. Indeed the MECG template cycle was built centered on the mother R-peak and considering a duration of 0.20,

0.10 and 0.40 s for the P, QRS and T waves respectively.[24] The choice of each ECG segments' time interval is realistic although certainly not optimal. However, the discontinuities mentioned are assumed to be minimal, with limited influence on FQRS detection (as opposed to FECG morphological analysis). Performance could be improved by varying the length of each of these intervals with the maternal heart rate rather than being constant.

The main drawback of the TS with respect to the LMS and ESN is that it relies on accurate MQRS detection (the best detectors typically achieve about 99% accuracy over a range of diverse databases). Indeed, a missed MQRS detection will most likely result in FPs (except if located in the FQRS refractory period) and possibly FNs (if the actual FQRS is located within the refractory period of the induced FP). Conversely, the main drawback of the LMS, RLS and ESN algorithms is that they are driven by and subject to poor signal quality in the chest signal. This particularly motivates the use of SQIs. It should also be noted that the number of coefficients for LMS/RLS adaptive filters or neurons in the ESN is a function of the signal sampling frequency and as a consequence parameter optimization should be conducted again if a different sampling frequency is to be considered.

Usage of a SQI on the chest reference channel improved the accuracy measure by ∼1% while suppressing ∼3% of the overall record. It is expected that the added value in using SQI would be higher in the presence of noisier recordings, as the data used in this work were of relatively good quality with some minor local artifacts.

It is to be noted that the number of ESN parameters that need tuning, in addition to the randomness of reservoir initialization and connectivity (that give no insight into the reservoir dynamic), makes the ESN design and implementation difficult. In an attempt to tackle these problems, Rodan et al.[28] compared the performance of the "standard" ESN framework (the one used in this paper) with much simpler networks structures and showed that similar performance to the standard ESN could be achieved using these deterministically constructed reservoirs. Their approach has two main advantages: 1) simplification of the reservoir design and number of parameters to optimise and 2) building a path to theoretical analysis of the ESN. This alternative ESN design will be compared, in the context of our application, to the standard ESN in future work. An alternative is to use random search for finding a set of acceptable parameters as suggested in Behar et al.[4] where we used the approach from Bergstra et al.[6] Random search performed for the ESN and preprocessing parameters showed similar performance as the exhaustive grid search presented in this work but

reducing the number of search iterations. The parameters that were assumed constant in this work in order to restrict the grid search to an acceptable number of free-parameters (such as ϕ and γ) could be searched if random search was employed. In practice it is likely that random search would be used as an initialization method to find which parameters are the most relevant and further parameter tuning would be performed with fine grid.

An important limitation for evaluating NI-FECG extraction algorithms is the absence of large publicly available databases with expert references. This limitation was recently partially addressed with the introduction of the Physionet/Computing in Cardiology 2013 challenge database.[32] However this database is constituted of 4 abdominal channels for each record without any reference chest MECG channel, thus not allowing for evaluation of adaptive noise canceling techniques such as the LMS, RLS and ESN presented in this work.

The ESN performed slightly better than the LMS and RLS on the test database, but the results were not significantly different. One of the main differences between the two approaches is that the ESN allows for a non-linear relationship between the chest and abdominal ECGs whereas the LMS only considers a linear relationship. Thus the results could suggest that the the mapping is mostly linear or that the numerous ESN hyperparameters were over-tuned on the training set database. Recall that the data in the training set (DB_1) and test set (DB_2) were recorded with different hardware, following a separate protocol and at different stage of pregnancy for different subjects. Since the LMS is conceptually simpler and computationally less expensive it is most likely a more appropriate choice for low-cost devices with limited computational power. However in the case of a more advanced hospital-based system, where computing power is generally not a concern, then there is little incentive against using the more computationally complex and accurate solution.

CONCLUSION

This work compared non invasive FECG extraction methods. The methods were qualitatively and quantitatively evaluated. In addition, the preprocessing performed for baseline wander and high frequency removal was studied in some detail. These filters, along with various parameters for the LMS, RLS and ESN, were optimized through exhaustive grid search on a training database. The findings of this research are; (1) The non linear ESN_a method showed slightly superior

performance with respect to the LMS, RLS and TS methods; (2) using a high baseline wander cut-off frequency improved the performance of the extraction algorithms; (3) SQI improved the performance of all methods by excluding bad quality chest ECG which likely led to an unexpected adaptive filters response or false QRS detection. In addition this paper suggested a framework for assessing the algorithms performance in extracting the FQRS and FHR by using the F1 and HR_m measures. Future work requires assessing the algorithms on a larger dataset, benchmarking them against additional methods employed for this task and evaluating the alternative ESN reservoir design as suggested by Rodan *et al.*[28] Open source code for the benchmark methods are made available to allow comparison and reproducibility on the public domain data. The source code is available on Physionet at http://physionet.org/.[10]

REFERENCES

[1]ANSI/AAMI/ISO EC57 (1998/(R)2008) Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms.
[2]Åström, K. J., and B. Wittenmark. *Adaptive Control*, 2nd ed. Reading, MA: Addison Wesley, 1994.
[3] Barnett, S., and D. Maulik. Guidelines and recommendations for safe use of Doppler ultrasound in perinatal applications. *J. Matern-Fetal Neonatal Med.* 10(2):75–84, 2001.
[4]Behar, J., A. Johnson, J. Oster, and G. D. Clifford. *An Echo State Neural Network for Foetal Electrocardiogram Extraction Optimised by Random Search.* Nevada: NIPS Lake Tahoe, 2013a.
[5]Behar, J., J. Oster, and G. D. Clifford. Non invasive FECG extraction from a set of abdominal sensors. In: Computing in Cardiology 2013, Zaragoza, Spain, 2013b.
[6]Bergstra, J., and Y. Bengio. Random search for hyperparameter optimization. *J. Mach. Learn. Res.* 13:281–305, 2012.

[7]Cardoso, J., and A. Souloumiac. Blind beamforming for non Gaussian signals. In: Institute Electrical Engineer Proceedings of Radar Signal Processing, London, vol. 140(6), pp. 362–370, 1993.

[8]Cerutti, S., G. Baselli, S. Civardi, E. Ferrazzi, A. Marconi, M. Pagani, and G. Pardi. Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings. *J. Perinat Med.* 14(6):445–452, 1986.

[9]Chatzis, S., and Y. Demiris. The copula echo state network. *Pattern Recogn.* 45(1):570–577, 2012.

[10]Clifford, G. D., J. Behar, J. Oster, and A. Johnson. IPM Open Source Code. https://physionet.org/users/gari@alum.mit.edu/works/IPMCode/, 2014. Last updated Feb 2014.

[11]Clifford, G. D., J. Behar, Q. Li, and I. Rezek. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiol. Meas.* 33(9):1419–1433, 2012.

[12]Cohen, W. R., S. Ommani, S. Hassan, F. G. Mirza, M. Solomon, R. Brown, B. S. Schifrin, J. M. Himsworth, and B. R. Hayes-Gill. Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes. *Acta Obstet .Gynecol. Scand.* 91(11):1306–1313, 2012.

[13]Douglas, S. Numerically-robust O (N(2)) RLS algorithms using least-squares prewhitening. In: Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing, vol. 1, pp. 412–415, 2000.

[14]Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):215–220, 2000.

[15]Guerrero-Martinez, J., M. Martinez-Sober, M. Bataller-Mompean, and J. Magdalena-Benedito. New algorithm for fetal QRS detection in surface abdominal records. In: Computation in Cardiology, pp. 441–444, 2006.

[16]Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw* 10(3):626–634, 1999.

[17]Jaeger, H. The "echo state" approach to analysing and training recurrent neural networks. Technical Report GMD Report 148. German National Research Center for Information Technology 148, 2001. http://www.faculty.jacobs-university.de/hjaeger/pubs/EchoStatesTechRep.pdf.

[18]Jaeger, H. A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach, 3rd revision. GMD-Forschungszentrum Informationstechnik, 2008. http://www.pdx.edu/sites/www.pdx.edu.sysc/files/Jaeger_TrainingRNNsTutorial.2005.pdf.

[19]Kanjilal, P., S. Palit, and G. Saha. Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. *IEEE Trans. Biomed. Eng.* 44(1):51–59, 1997.

[20]Kotas, M., J. Jezewski, A. Matonia, and T. Kupka. Towards noise immune detection of fetal QRS complexes. *Comput. Methods Programs Biomed.* 97(3):241–256, 2010.

[21]Lewis, M. Review of electromagnetic source investigations of the fetal heart. *Med. Eng. Phys.* 25:801–810, 2003.

[22]Lukoševičius, M. A practical guide to applying echo state networks. *Lect. Notes Comput. Sci.* 7700:659–686, 2012.

[23]Lukoševičius, M., and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* 3(3):127–149, 2009.

[24]Martens, S., C. Rabotti, M. Mischi, and R. Sluijter. A robust fetal ECG detection method for abdominal recordings. *Physiol. Meas.* 28:373–388, 2007.

[25]Oudijk, M., A. Kwee, G. Visser, S. Blad, E. Meijboom, and K. Rosén. The effects of intrapartum hypoxia on the fetal QT interval. *BJOG* 111(7):656–660, 2004.

[26]Pan, J., and W. Tompkins. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* 32(3):230–236, 1985.

[27]Petrenas, A., V. Marozas, L. Sornmo, and A. Lukoševičius. An echo state neural network for QRST cancelation during atrial fibrillation. *IEEE Trans. Biomed. Eng.* 59(10):2950–2957, 2012.

[28]Rodan, A., and P. Tino. Minimum complexity echo state network. *IEEE Trans. Neural Netw.* 22(1):131–144, 2011.

[29]Sameni, R. Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings. Ph.D. Thesis, Sharif University of Technology, Institut National Polytechnique de Grenoble, 2008. http://www.sameni.info/Publications/Thesis/PhDThesis.pdf.

[30]Sameni, R., and G. D. Clifford. A review of fetal ECG signal processing; issues and promising directions. *Open Pacing Electrophysiol. Ther. J.* 3:4–20, 2010.

[31]Sameni, R., C. Jutten, and M. Shamsollahi. Multichannel electrocardiogram decomposition using periodic component analysis. *IEEE Trans. Biomed. Eng.* 55(8):1935–1940, 2008.

[32]Silva, I., J. Behar, G. D. Clifford, and G. B. Moody. Noninvasive fetal ECG: the PhysioNet/Computing in Cardiology Challenge 2013. In: Computing in Cardiology 2013, Zaragoza, Spain, 2013.

[33]Ungureanu, M., J. Bergmans, S. Oei, and R. Strungaru. Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique. *Biomedizinische Technik* 52(1):56–60, 2007.

[34]Vahidi, A., A. Stefanopoulou, and H. Peng. Recursive least squares with forgetting for online estimation of vehicle mass and road grade: theory and experiments. *Vehicle Syst. Dyn.* 43(1):31–55, 2005.

[35]Vullings, R., C. Peters, R. Sluijter, M. Mischi, S. Oei, and J. Bergmans. Dynamic segmentation and linear prediction for maternal ECG removal in antenatal abdominal recordings. *Physiol. Meas.* 30(3):291–307, 2009.

[36]Widrow, B., J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, and R. Goodlin. Adaptive noise cancelling: principles and applications. *Proc. IEEE* 63(12):1692–1716, 1975.

[37]Zaunseder, S., F. Andreotti, M. Cruz, H. Stepan, C. Schmieder, H. Malberg, and A. Jank. (Como, July 2012) Fetal QRS detection by means of Kalman filtering and using the Event Synchronous Canceller. In: 7th International Workshop on Biosignal Interpretation.