Robust fetal ECG extraction and detection from abdominal leads

Fernando Andreotti[1], Maik Riedl[2], Tilo Himmelsbach[1], Daniel Wedekind[1], Niels Wessel[2], Holger Stepan[3], Claudia Schmieder[3], Alexander Jank[3], Hagen Malberg[1] and Sebastian Zaunseder[1]

[1] Institute of Biomedical Engineering, Faculty of Electrical and Computer Engineering, TU Dresden, Germany
[2] Cardiovascular Physics, Department of Physics, Universität zu Berlin, Germany
[3] Department of Obstetrics, University Hospital of Leipzig, Germany

E-mail: fernando.andreotti@mailbox.tu-dresden.de

Abstract

The fetal ECG derived from abdominal leads provides an alternative to standard means of fetal monitoring. Furthermore, it permits long-term and ambulant recordings, which expands the range diagnostic possibilities for evaluating the fetal health state. However, due to the temporal and spectral overlap of maternal and fetal signals, the usage of abdominal leads imposes the need for elaborated signal processing routines.

In this work a modular combination of processing techniques is presented. Its core consists of two maternal ECG estimation techniques, namely the extended Kalman smoother (EKS) and template adaption (TA) in combination with an innovative detection algorithm. Our detection method employs principles of evolutionary computing to detect fetal peaks by considering the periodicity and morphological characteristics of the fetal signal. In a postprocessing phase, single channel detections are combined by means of kernel density estimation and heart rate correction.

The described methodology was presented during the Computing in Cardiology Challenge 2013. The entry was the winner of the closed-source events with average scores for events 4/5 with 15.1/3.32 (TA) and 69.5/4.58 (EKS) on training set-A and 20.4/4.57 (TA) and 219/7.69 (EKS) on test set-B, respectively. Using our own clinical data (24 subjects each 20 min recordings) and statistical measures beyond the Challenge's scoring system, we further validated the proposed method. For our clinical data we obtained an average detection rate of 82.8% (TA) and 83.4% (EKS). The achieved results show that the proposed methods are able produce reliable fetal heart rate estimates from a restricted number of abdominal leads.

Keywords: fetal ECG, abdominal signals, template subtraction, extraction, evolutionary detection

1. Introduction

Congenital heart defects are one of the most common birth deformities (Sameni 2008). Similarly to adult electrocardiography, the morphological analysis of the fetal electrocardiogram (FECG) provides additional diagnostic information, since most cardiac defects have some manifestation on ECG tracings (Sameni and Clifford 2010). However, accurate and reliable detection of fetal peaks is a pre-requirement for further morphological analysis of the FECG.

Since its introduction in the 1960s, electronic fetal monitoring has become part of routine maternal care (Bailey 2009), which includes the analysis of fetal heart rate (FHR) patterns. Standard methods of electronic fetal monitoring are either non-invasive (i.e. cardiotocography) or, more recently, invasive (i.e. fetal scalp electrode). Despite their broad acceptance, these methods are not suited for long-term antepartum monitoring, due to lack of specificity or the exclusive compatibility with intrapartum period. Cardiotocography for instance, aside from usually providing an averaged estimate for FHR, requires a skilled specialist to position the ultrasound probe during the complete measurement. As a consequence, cardiotocography is known for its low specificity (van Laar *et al* 2009) and poor inter and intraobserver reliability (American College of Obstetricians and Gynecologists 2009). Moreover, cardiotocography has been associated with an increase in the number of medical interventions, meanwhile the mortality and morbidity rates have not decreased (Amer-Wåhlin and Maršál 2011, Bailey 2009). On the other hand, a fetal scalp electrode does provide information about the electrical activity of the a fetal heart with high frequency resolution (on a beat-to-beat basis). However, its use is limited to the intrapartum period and since dealing with an invasive method, it may incur the risk of infection.

The FECG derived from surface abdominal leads (abdominal electrocardiogram—abdECG) is regarded as a promising alternative, which allows non-invasive long-term monitoring of the fetal heart (e.g. with a portable abdominal Holter device). Although many studies have contemplated the topic (e.g. Widrow *et al* 1975, Peters *et al* 2001, Ungureanu *et al* 2009), FECG extraction (i.e. maternal interference estimation and cancellation) along with trustworthy fetal QRS detection remains an unsolved problem. The varying (usually low) fetal signal-to-noise ratio (SNR), temporal and spectral overlap of FECG and maternal electrocardiogram (MECG) signals and the morphological similarity between QRS complexes of mother and fetus entail difficulties concerning FECG's clinical usage.

Aiming at accurate detection of fetal QRS (FQRS) complexes and FHR, the Computing in Cardiology (CinC) Challenge 2013 (Silva *et al* 2013, Clifford *et al* 2014) addressed the field of FECG processing and promoted the development of processing methods by providing a large database of abdECG records. This work, which was presented during CinC (Andreotti *et al* 2013), is the further development of Zaunseder *et al* (2013), which compared methods for MECG estimation, namely the extended Kalman smoother (EKS), (Sameni 2008) and the event synchronous canceller (Ungureanu and Wolf 2006). For this contribution, the latter method was further developed to allow a more flexible adaptation of maternal templates, here referred to as template adaptation (TA). EKS and TA are applied in parallel in order to extract the FECG. A newly developed FQRS detector and postprocessing scheme is sequentially used in deriving the FQRS and FHR.

2. Material and methods

2.1. Data

Our analysis made use of the data available within the CinC Challenge for the closed-source events as well as clinical recordings provided by the University Hospital of Leipzig within a collaboration on FECG analysis.

2.1.1. Challenge data. Closed-source data consisted of 175 records drawn from five different sources (including real and simulated data). The records were divided into a training set (set-A) of 75 recordings[4] and a test set (set-B) containing 100[5] recordings. Each recording contained a 1-minute abdECG measurement with four channels at 1000 Hz. For set-A along with the raw signals, reference for FQRS locations was provided. For both sets references were annotated by different specialists using fetal scalp electrode recordings (Silva *et al* 2013, Clifford *et al* 2014).

2.1.2. Own clinical data. Clinical recordings were acquired at the Department of Obstetrics and Gynecology at the University Hospital of Leipzig. The study was approved by the University Hospital of Leipzig's ethics commission record 348-12-24092012 and written informed consent was obtained from each patient. Twenty-four 20-minute abdECG recordings of variable fetal-maternal SNR were selected from a larger collective of measurements. These recordings were taken from ten women (both healthy and pathological patients were present), aged between 21 and 33 years (27.1 ± 4.3 years) and gestational weeks between 20 and 28 weeks (25.0 ± 2.5 weeks). No ectopic beats have been found for either mother or fetus.

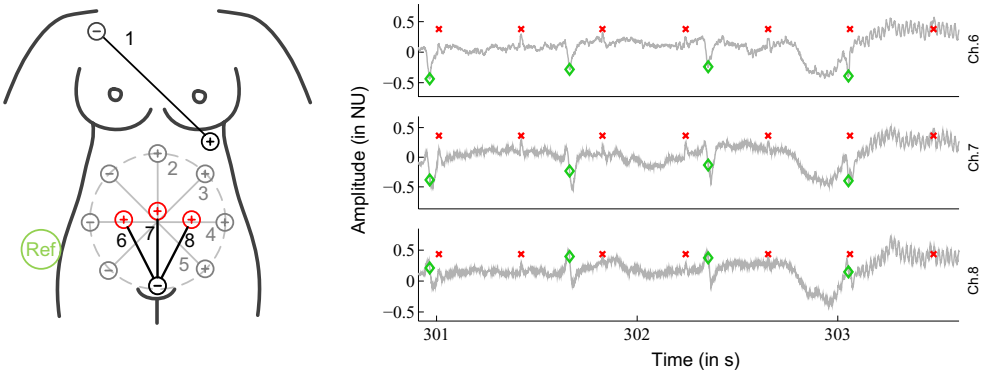
Data were collected by a bipolar lead configuration (see figure 1) at 1000 Hz sampling frequency using the ADInstruments ML138 Octal Bio Amp and ADInstruments PowerLab 16/30 (ADInstruments, Dunedin, New Zealand). The abdominal leads were filtered by a mains filter and a high pass filter (cut-off frequency 1 Hz). The measuring range of the abdominal recordings was 500 μV. To be used as gold standard, each record had its maternal QRS (MQRS) and FQRS annotated by one and corrected by two other trained specialists (see figure 1(*b*)). Reference annotation was performed on channel 1 (for the maternal reference) and the full set of abdominal leads (2–8, in figure 1) for the fetal reference. As preliminary works identified leads 6–8 (black coloured in figure 1(*a*)) as the most suitable ones, only these leads were used during signal processing. Moreover, a low number of leads provides the processing algorithm a more challenging situation which portable devices usually face.

2.2. Performance evaluation

In order to assess the accuracy of our FQRS detection, different measures were used. The first two indexes were provided by the CinC Challenge (function score2013.m) for event 4 and 5. Namely the mean squared error between smoothed and re-sampled versions of the reference FHR (FHR_i) at time i and detected FHR (FHR_i^d) (Score1, in bpm^2) and the root mean square difference of corresponding RR intervals (Score2, in ms), respectively (Silva *et al* 2013). Equations (1) and (2) describe the scoring calculus, with i being the detection index and I the total number of FQRS detections present on the reference

$$\mathrm{Score1} = \frac{12}{I} \cdot \sum_{i=1}^{I/12} (\mathrm{FHR}_i - \mathrm{FHR}_i^d)^2 \tag{1}$$

[4] This work disregards 'a54' since it was discarded by the Challenge's organizers.
[5] The Challenge organizers might not have used all these recordings for providing statistical results.



(a) (b)

Figure 1. (*a*) Electrode configuration for patient recording. Three bipolar leads were used in this study as input to signal processing methods (leads 6–8); (*b*) exemplary raw data (no unit—NU), reference annotations are represented as maternal (◇) and fetal (×).

$$\text{Score2} = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(\text{RR}_i - \text{RR}_i^d)^2}\,. \tag{2}$$

Aside from the Challenge's scoring system, we used sensitivity (SE) and positive predictive value (PPV) as defined in (ANSI/ AAMI EC57 1998):

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

where true positive (TP) denotes correctly detected peaks, false negative (FN) being existing peaks which were not detected, and false positive (FP) nonexistent peaks that were falsely detected. Differing from ANSI/ AAMI EC57 (1998) we defined a ±50 ms acceptance interval between detection and the closest reference annotation to account for the higher expected heart rate (in the norm ±150 ms is suggested).

Finally, equations (5) and (6) describe two measures that were used to summarize the accuracy of FQRS detections (i.e. TP in comparison to FP and FN): accuracy (ACC) as in Karvounis *et al* (2007) and the F_1-measure (Behar *et al* 2013).

$$\text{ACC} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{5}$$

$$F_1 = 2 \cdot \frac{\text{PPV} \cdot \text{Se}}{\text{PPV} + \text{Se}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FN} + \text{FP}}. \tag{6}$$

For our own data, each statistic was calculated for every consecutive 1-minute segment (non-overlapping), to better illustrate how the algorithm's performance may vary along the same recording. Afterwards both intra and interindividual results were evaluated.

2.3. Signal processing

The workflow for signal processing is depicted in figure 2. Its modular design permits the substitution of processing blocks. Each block is composed of several subroutines. In the following subsections, each individual block is thoroughly explained.

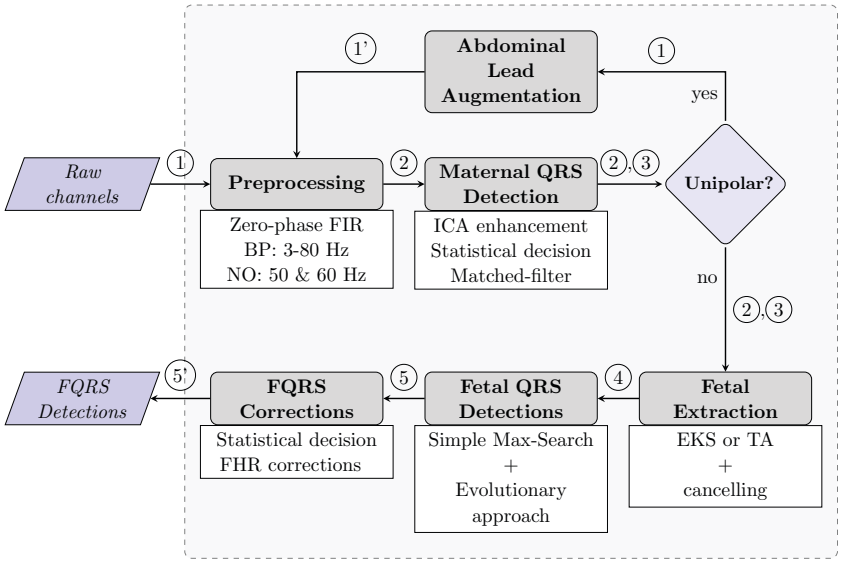


Figure 2. Workflow diagram of signal processing. ① represents the N raw channels; ①' stands for the $M=N\cdot(N-1)/2$ pairwise subtracted ①; ② depicts the N (or $N+M$) preprocessed channels; ③ represents the MQRS detections over ② (one decision); ④ is N (or $N+M$) channels with suppressed maternal signals (extracted FECG); ⑤ is the FQRS detections using evolutionary detector; ⑤' represents the statistically decided and FHR corrected FQRS detections.

2.3.1. Preprocessing. To reduce baseline wander, muscular artifacts and power-line interference we used 1000 taps FIR filters. The first two types of noise were partially suppressed with a bandpass between 3–80 Hz. For power-line removal two notch-filters (at 50 Hz and 60 Hz) were applied, since the origin of the Challenge data was unknown. The three filters were consecutively applied in the forward and backward direction, performing zero-phase filtering. Figure 3 shows the magnitude and phase plots for the resulting filter.

2.3.2. Maternal QRS detection. Considering the non-standardized abdominal derivations, we augmented the maternal signal by using spatial filtering before performing MQRS detection. Independent component analysis (ICA) has been previously proposed for separating FECG and MECG components (De Lathauwer *et al* 2000), thus FastICA algorithm (Hyvärinen 1999) was applied for the purpose of isolating the MECG. The ICA algorithm was used symmetrically, incorporating whitening and making use of a hyperbolic tangent contrast function (as suggested by Hyvärinen (1999)). To account for non-stationarities inside recordings, ICA was applied on non-overlapping sub-segments of 30 s duration.

Permutation indeterminacy denotes a common difficulty which goes along with ICA. The *a priori* unknown positioning of independent components (ICs) inside the ICA's output imposes a challenging task of automatic selection of the IC of interest (i.e. MECG) (Wedekind *et al* 2013). For automated selection, first MQRS detection was performed using a maximal/minimal search algorithm (Sameni 2010) in each of the preprocessed channels (i.e. prior to

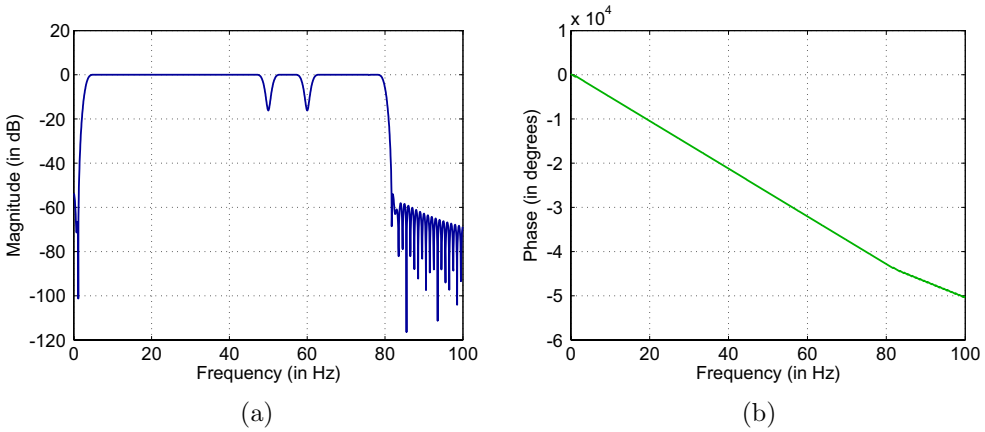


Figure 3. (*a*) Magnitude of frequency response to cascaded filters; (*b*) phase response of the resulting filter.

ICA—primary detection) and in all ICA output channels (secondary QRS detection). Figure 4 depicts the results of using ICA on two of the Challenge's data.

Each channel-based detection was then compared to the others by a voting system, using ACC (from equation (5)—using a $\pm$ 25 ms matching window around reference detection) as measure of correctness. Based on the assumption that MQRS complexes are predominantly detected in both input and output channels, the agreement between primary and secondary QRS detections can be considered as an indicator of the MECG presence on that IC (see figure 3). The detection-channel with highest ACC was then chosen as the best output channel.

To cope with cases when ICA fails to produce reliable ICs, the chosen ICA detections were once more compared to the primary MQRS detections. This was performed based on a statistical decision-making procedure to generate a consensus between primary MQRS detection and best secondary MQRS detection. This consensus took into account the value of FP and FN by using kernel density estimation of the peak distributions in multiple channels as in Silverman (1986). Gaussian kernel functions with 30 ms standard deviation and unitary amplitude were centered at the time instants of each detected peak. The sum of these kernels resulted in a distribution of peaks which ranged from zero to the number of considered channels, so that the height of the local maxima represented the match of detections in multiple channels. Values above a given threshold (defined as half of the number of channels) were accepted as MQRS complexes.

Next, MECG morphology was exploited by making use of a matched filter detection approach. For this purpose, the primary detection channel that showed the strongest agreement with our statistical decision was used. A template (length 120 ms) was obtained by aligning individual beats based on their correlation coefficients, low correlating beats were then excluded and a trustworthier template was generated. The template was cross-correlated with the preprocessed best channel, where high-correlated new peaks were accepted as correct if either (1) they overlapped with previous detections in a given window (three times the size of the template)—these detections were re-aligned to match the R-peak using matched filter's output; or (2) a correlation superior to 80% was achieved outside this window. Previously detected beats were removed, in cases of poor correlation (less than 50%) between template and signal.

Lastly, differences between RR-series and a 5th order smoothed version of the RR-series greater than 150 ms were corrected by adding/removing detections.

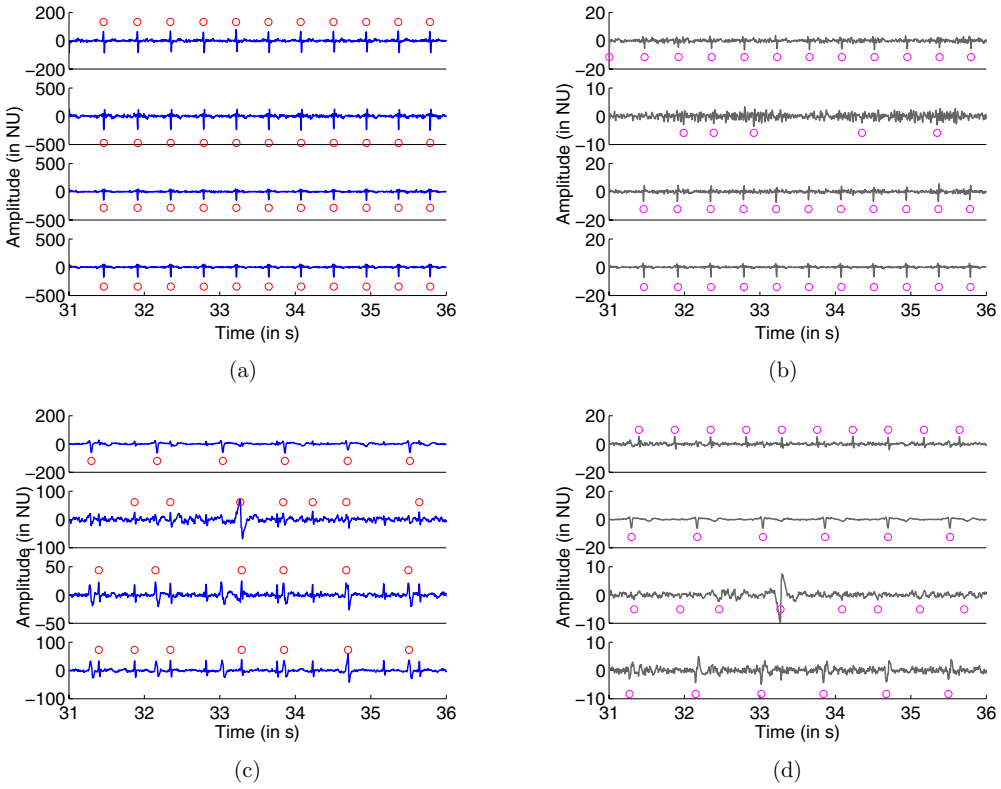


Figure 4. (*a*) Preprocessed signals (record 'a02' from set-A) before ICA; (*b*) output signals after spatial filtering[6] for (*a*); (*c*) preprocessed signals (record 'a08' from set-A) before ICA; (*d*) output signals from ICA for (*c*). MQRS detections are represented by circles. On (*a*), the FECG component has low SNR, while on (*c*), channels 3 and 4, the FQRS have almost the same amplitude as MQRS. For both cases, a maximal/minimal peak detector from Sameni (2010) was used. For (*b*) ICs 1, 3 and 4 show maximum accuracy, on (*d*) IC 2 was selected as the best component.

2.3.3. Abdominal lead augmentation. Before estimating the maternal signals, we exploited the MQRS amplitudes and created a rule for linear combining raw channels. This step is only relevant for Challenge data, where the lead configuration was unknown (see figure 5). A large maternal peak amplitude (i.e. abdominally recorded peak with absolute value greater than 150 mV) was assumed to indicate a sort of unipolar measurement. By means of linear combining the four raw channels (through subtraction) and using the six available combinations of these channels, we augmented the FECG components. The resulting six channels were then preprocessed and, for the further steps, used concurrently with the four initial channels (see figure 2).

2.3.4. Fetal signal extraction. In this work, two temporal methods were used. These methods, often called template subtraction methods, make use of the pseudo-periodicity of MECG and time-decorrelation with the FECG for generating a maternal template beat (Martens 2003). By re-adapting this template onto the preprocessed signals, the maternal component within a single channel can be estimated. Both our methods started from a common ground, which

[6] Figure 4(*b*) shows mainly MECG and noise ICs (i.e. no fetal component), which is because ICA requires at least the same number of leads as sources (Karvounis *et al* 2007).

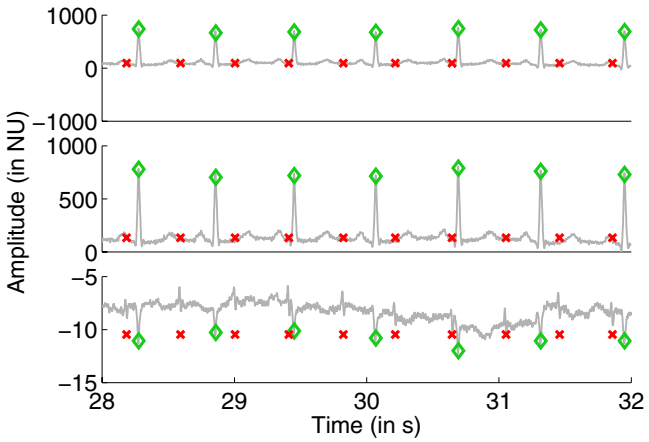


Figure 5. Exemplary CinC recording 'a59' where 2 linear combination of channels is advantageous. The top panel represents raw channel 2; middle shows the raw channel 3; bottom the result of subtracting raw channel 3 from 2 (notice that fetal peaks are visible). Reference maternal and fetal annotations are represented by (◇) and (×), respectively.

was an average maternal template. This template was obtained by wrapping (as in Sameni *et al* (2005) and Sameni (2008)) 30 s segments of abdominal signals and coherent-averaging (Rompelman and Ros 1986) the wrapped beats.

In order to keep the consistency with the previous processing steps, we applied the further mentioned methodology to every 30 s of measurement. Lastly, after obtaining maternal estimates for each channel (using either of the methods), these estimates were subtracted from the respective preprocessed channels, therefore obtaining estimates for FECG plus noise.

2.3.5. Maternal estimation by means of extended Kalman smoother. The Kalman filter is a versatile framework that has been applied for filtering diverse biomedical signals. Its use for FECG extraction was first proposed by Sameni *et al* (2005), a more complete description is present in Sameni (2008) and an implementation is available in Sameni (2010). For our purpose, the Kalman filter was used in estimating the MECG for each channel based on:

(a) average maternal beat, previously described;
(b) a dynamic model comprising a phase information and a mathematical description to the maternal ECG template. The first is derived from MQRS detections, where for each beat a linear interpolation between $[-\pi, \pi]$ is assigned, with zero being the maternal R-peak, therefore resulting in a sawtooth-shaped signal. The latter is based on McSharry *et al*'s (2003) electrocardiogram model, which approximates ECG waveforms by adding N shifted Gaussian kernels;
(c) the observed signals (preprocessed abdominal signals).

Since dealing with a non-linear model, its dynamical model has to be linearized before filtering. This extension is called the extended Kalman filter. During filtering, the Kalman gain provides a weight between observations and inner model, which is based on estimations of noise for observations and model (covariance matrices). After filtering, a backward smoothing stage is applied to provide a more trustworthy estimation of the MECG projection within the abdominal leads, therefore the method is called the extended Kalman smoother (EKS).

The implementation is available in Sameni (2010). The steps for obtaining a maternal estimate can be summarized as follows.

(a) Approximate an averaged beat with N Gaussian kernels, therefore obtaining $3 \cdot N$ parameters (amplitude, width and position).
(b) Filter the signal on sample basis, with the Kalman gain serving as a weight between observation and inner model by means of covariance matrices.
(c) Non-causal backward fixed-interval smoothing.

2.3.6. Maternal estimation by means of template adaptation. Based on the event-synchronous canceller (ESC, Ungureanu and Wolf (2006)) and our previous work (Zaunseder *et al* 2013) we developed the template adaption (TA) method.

Similarly to the ESC, TA makes use of an average maternal beat which is directly adapted back to the original signal without any fixed mathematical modelling (as EKS does). Differing from ESC, TA also allows a width adaption (stretching the template around the QRS-complex) which grants more flexibility. In order to perform this adaption, the template was divided into three patterns (likely to represent the maternal Q, R and S waves). Each sub-pattern was then allowed to vary in width and height, which gave the extraction method some additional degrees of freedom. Furthermore, based on the overall signal trend (height of preceding Q, R or S waves) a template adaption constraint was added to avoid complete cancellation of fetal peaks in cases of complete temporal overlap with maternal QRS. The steps for performing MECG estimation by means of TA are as follows.

(a) Locate sub-patterns (Q, R and S waves) on the previously described template.
(b) Identify the corresponding sub-patterns of each maternal beat on the preprocessed channel (cross-correlating sub-patterns of the template with the preprocessed signal).
(c) Evaluate and constrain the gain for templates whose amplitudes exceed a smoothed version of the beats' amplitude-series (using a 9 beat span).
(d) Temporally re-sample both template and preprocessed channel and selection of 350 feature points (F_p) for width adaptation of the template.
(e) Stretch template's segments for optimal adaption (based on Euclidean distance) to each F_p on the preprocessed channels.

2.3.7. Fetal QRS detection. In order to cope with low fetal SNR due to noise and residuals from MECG extraction, we introduced a novel method for FQRS detection that makes use of basic concepts from evolutionary algorithm (EA). Essentially, based on a set of initial detections, our method corrects these detections following two basic principles: the overall beat-to-beat interval periodicity and the beat morphology on the recording.

EAs have become a popular heuristic near-optimal solution for problems with large search spaces (Whitley 2001). In contrast to traditional algorithms for optimization, EAs are working with probabilistic rules thereby enabling parallel solutions. Due to the lack of boundaries in creating the global cost function, EAs can be applied in an easy and flexible way (Pohlheim 1999). In EAs, artificial individuals (i.e. FQRS detections) are used to populate, compete with one another and discover optimal solutions within this search space (Bäck and Schwefel 1993), i.e. the FECG signal.

Although the EA population is often arbitrarily initialized (Bäck and Schwefel 1993), for the purpose of faster convergence we chose to start our method with initial FQRS detections using a simple maximal search algorithm (Sameni 2010). During initialization, a fitness value (FV_i, depending on periodicity and morphology of FQRS detections) was assigned to every

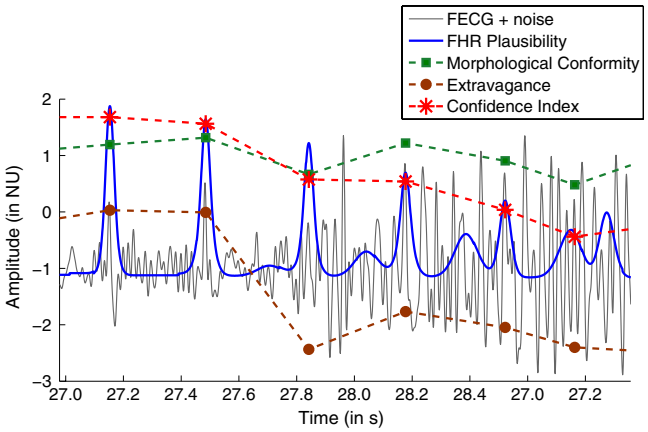


Figure 6. Different parameters used for defining the fitness curve.

individual (i). The value of our global optimal solution is expressed using a fitness function. Possible FQRS were allowed to originate, change location or cease existence on every iteration. Following the ideas of simulated annealing (Kirkpatrick *et al* 1983), the amount of changes (i.e. number of detections to originate/move/vanish) was reduced after every cycle.

Periodicity (or 'FHR plausibility', see figure 6) was determined based on a smoothed RR-interval curve. The algorithm looped through each detection and, based on the local smoothed RR-curve, it placed six Gaussian kernels (three on each side of actual detection) centered on plausible neighbouring peaks. The amplitude of such kernels depended on two factors: (i) local value of FV_i (amplitude increase with rising FV_i) and (ii) distance from actual individual, i.e. $\{1\cdot FV_{i\pm1}, 0.9\cdot FV_{i\pm2}, 0.8\cdot FV_{i\pm3}\}$ (decaying amplitude with increasing distance). The Gaussian standard deviation was defined as 20% of the recording's shortest smoothed RR-interval. The sum of the individually generated curves resulted in the FHR plausibility curve. The neighbouring Gaussian kernels provided a rough measure for certainty about FQRS positions and generated competition between neighbouring individuals. As the number of iterations increased, the kernels were expected to interfere constructively with one another and generate a more skewed FHR plausibility curve (see figure 7).

On the other hand, morphological features can not be described by one single feature signal, particularly due to the presence of noise. In order to assess morphology, we made use of two sub-functions, namely conformity and extravagance. Conformity is a cross-correlation-based measure of likeness between each detection and a generated fetal beat template, while extravagance is a measure of contrast between the detected peak and its surroundings (see figure 6).

The optimization procedure stopped when convergence occurs, i.e. the majority of the detections settled (did not move), or the maximum number of iterations was achieved ($i = 1000$). After providing viable FQRS within each channel, the detector uses the annotation channels which were considered to be most trustworthy (largest overall FV) to generate a set of multichannel detections. Since the optimization was affected by the choice of its process variables, the procedure might get stuck in local optima or not converge producing random differences in its results.

2.3.8. Decision-making and correction. Decision making aims at combining single channel annotations in one consensus annotation. In order to do so, an adaptation of the approach applied for MQRS consensus was used. For creating this consensus, lower local maxima

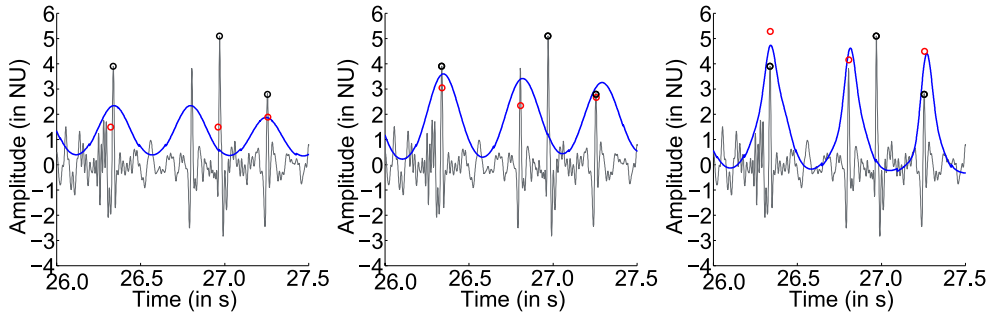


Figure 7. Iterations of the evolutionary algorithm for record 'a01' from the CinC Challenge. The continuous line depicts the 'FHR plausibility' curve (from figure 6). The detection quality indexes (FV) are represented by the height of the actual detection at time i (○). Initial detections (at $i=0$) are depicted by (○) (*a*) detector at iteration step $i=119$, FV = {1.5, 1.5, 1.9}; (*b*) at $i=126$, FV = {3, 2.3, 2.6}; (*c*) at $i=742$, FV = {5.3, 4.2, 4.5}.

between two accepted neighbours were deleted if both peaks had a temporal distance which was smaller than a defined refractory period (300 ms).

Last, heart rate corrections were performed while attempting to obtain a plausible FHR. This correction was done by means of including peaks at the borders of gaps indicated by FHR jumps. Furthermore, FHR 'zigzag' pattern (i.e. a smaller value followed by a higher one and vice versa) were corrected switching the position of the central detected peak to decrease FHR variability. After looping through all fetal RR-intervals, persistent sudden heart FHR greater than 70 ms were removed and existing gaps were filled without regarding any distribution information.

3. Results

The proposed extraction methods were able to separate the MECG and FECG. The evolutionary FQRS detector obtained reliable detections in the presence of noise, even when MECG residuals persisted. Figure 8 shows an example of extraction using EKS and TA.

Table 1 shows the average interindividual performance measures obtained for each dataset. Despite the overall similar results for EKS and TA, both methods behave differently (see figure 8 and table 1). Particularly around maternal QRS complexes, EKS tended to cross-out FQRS complexes in cases where temporal overlap with maternal complexes occurred. Meanwhile TA produces some artefacts on the resulting FECG signal, due to the non-linear adaptation.

EKS produced more variable results which is reflected in the standard deviation of the performance measures. This variance is particularly strong for the value of Score1 (mean squared error between FHR), which may be caused by the removal of overlapping beats and attempt to re-insert these missing beats in a later stage. In average, TA's accuracy outperforms EKS's for CinC data but the opposite happens for our own recordings. The results were obtained from a single run of the proposed work flow for signal processing. Due to the stochasticity of our detector results may slightly differ.

In figure 9 intraindividual results using own clinical data are shown. Record 8, the record exhibiting the most pronounced differences between EKS and TA, is separately shown in figure 10. Recordings 12–14 were included even though no FECG was visible on channels 6–8 during visual inspection (channels 2–5 were used when creating fetal reference annotations).

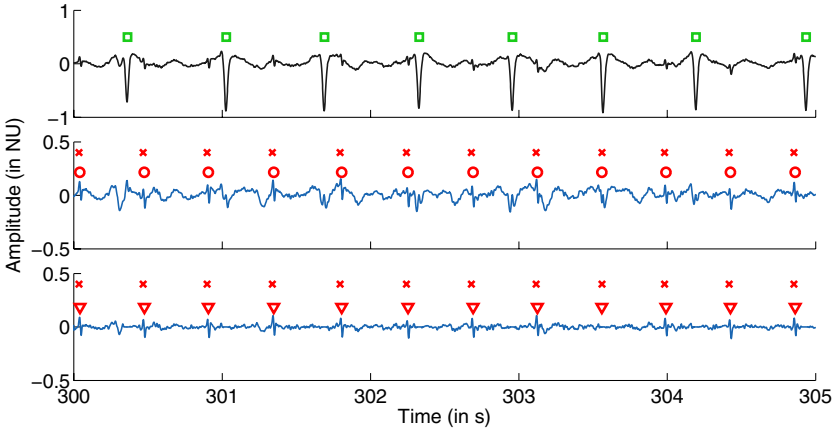


Figure 8. Five seconds segment of our own clinical dataset 2. Maternal QRS detections are marked with (□), fetal reference by (×), FQRS detections over TA output by (○) and FQRS detections over EKS output (▽). The top panel shows preprocessed abdominal channel 6; the middle the extraction results for TA over the channel is displayed; the bottom shows the extraction results for EKS.

4. Discussion

4.1. Relation to other FECG extraction methods

According to Martens (2003), approaches to estimate the fetal signal from abdECG leads can be categorized into spatial, temporal and, the combination of both, spatio-temporal methods. Spatial methods attempt to separate the fetal signal by using information about the spatial distribution of the sources within the abdECG mixture. Methods that appear in the literature include ICA (De Lathauwer *et al* 2000, Zarzoso and Nandi 2001), principal component analysis (PCA) (Martens 2003), singular value decomposition (SVD) (Callaerts *et al* 1990, Kanjilal *et al* 1997) and non-linear state-space projections (NSSP) (Richter *et al* 1998).

Some CinC competitors made use of such methods (alone or combining them with temporal methods) for separating the maternal and fetal ECG (e.g. Behar *et al* 2013, Fatemi *et al* 2013, Lipponen and Tarvainen 2013, Varanini *et al* 2013). Despite its successful usage, even during CinC's Challenge, some of these methods (e.g. ICA) have some drawbacks, e.g. (i) require a large number of channels (Karvounis *et al* 2007), that equals or exceeds the number of sources (Martens *et al* 2007); (ii) assume a stationary mix of the sources, which in cases of varying fetal SNR throughout the channels, fetal movement or appearance of noise is invalid; (iii) they unintentionally separate atrial and ventricular components (Rieta *et al* 2004), which makes further morphological analysis impracticable.

Temporal methods are also widely used in the literature and some competitors made use of them as well (e.g. Behar *et al* 2013, Lukosevicius and Marozas 2013, Kropf *et al* 2013). Our work focuses on such methods because they imply the fewest restrictions on the recording setup (a single channel can be used). Within this contribution two temporal methods (TA and EKS) were exploited and certain limitations were presented for each method. TA is a modification of template subtraction methods which allows not only and amplitude adaptation, as in Lukosevicius and Marozas (2013) and Martens *et al* (2007), but in width which provides the method an extra degree of freedom. EKS, on the other hand, was applied similarly to other competitor's schemes (e.g. Behar *et al* 2013, Akhbari *et al* 2013, Haghpanahi and Borkholder 2013). Results show that both extraction methods can be considered suitable

Table 1. Interindividual performance of FQRS detection depending on extraction method used for CinC data and our own data.

Data	Trial	Se (%)	PPV (%)	ACC (%)	F_1 (%)	Score1	Score2
Set-A	TA	97.4 ± 11.0	97.2 ± 10.7	96.0 ± 13.4	97.3 ± 10.8	15.1 ± 57.5	3.32 ± 14.1
	EKS	93.1 ± 20.3	92.8 ± 20.3	91.2 ± 23.2	93.0 ± 20.3	69.5 ± 245	4.58 ± 14.9
Set-B	TA	—	—	—	—	20.4	4.57
	EKS	—	—	—	—	219	7.69
Own Data	TA	85.8 ± 23.7	85.0 ± 23.3	82.8 ± 26.5	85.4 ± 23.5	102 ± 192	26.5 ± 27.8
	EKS	86.8 ± 23.4	85.9 ± 23.3	83.4 ± 26.9	86.3 ± 23.3	118 ± 201	27.0 ± 29.4

Values represented as (mean ± standard deviation). For hidden set-B, results for Se, PPV, ACC and F_1 are not available.

for FECG extraction. In terms of challenge scores, our scheme outperformed similar schemes, a finding the authors attribute mainly to our novel FQRS detector.

4.2. Limitations and future work

While dealing with clinical data, some limitations on the methodology were noticed. One of them being gestational age, i.e. case 15 (own data) as the earliest recording (20th week of gestation) indicates a low fetal SNR to strongly affect the accuracy of the detections. Another limitation could be associated with premature rupture of membrane and/or breech presentation as in the cases 8 and 12–14, where extraction/detection fails. While comparing EKS and TA, the fact that for most of our own clinical data EKS outperforms TA, indicates that TA may be over-trained for the Challenge's data (although better results were obtained for unknown set-B). These results highlight EKS's best feature: the versatility to adapt to different conditions.

However, EKS still needs improvements. Qualitative analysis reveals that EKS tends to remove fetal QRS complexes in cases of temporal overlap with maternal complexes (see figure 8), which is in accordance with the literature (Sameni 2008). The reason for this phenomenon is the unconstrained Kalman gain. Ideally, the Kalman filter should be able to track amplitude variations, most importantly introduced by the maternal respiration. At the same time, the filter should reject amplitude variations due to fetal maternal overlap. At the moment, EKS works on a sample basis not taking into consideration the amplitude of previous maternal beats. A viable solution would be an approximation of the cases interpolating the amplitude's modulation by means of splines or a Kalman filter (Nemati *et al* 2010). Amplitude ratios of other segments which undergo amplitude modulation due to respiration, e.g. QRS to T-wave amplitude (Langley *et al* 2010), could be exploited in order to improve the behavior of the Kalman gain as well. The usage of sub-segments' amplitude information, supported by TA, is responsible for successful FECG extraction even in cases of complete overlaps. Kalman filter constraints could also be added in order to allow the filter to behave in a similar manner.

Aside from constraining the Kalman gain, the dynamical model from McSharry *et al* (2003) can be improved by considering the non-linear temporal ECG changes (over PR, QRS and QT intervals). By means of dynamic template warping (Niknazar *et al* 2012), a trustworthier phase information (thus, wrapped template) is achieved. Improvement in our model can also be obtained by optimizing Gaussian kernel placement in approximating the maternal template beat and better designing the process and observational covariance matrices.

The evolutionary FQRS detector can be regarded as a sound approach for FQRS detection. Despite its development and adjustment over the CinC data, its applicability in longer and noisier datasets was validated using own clinical data. The observed drop in the results

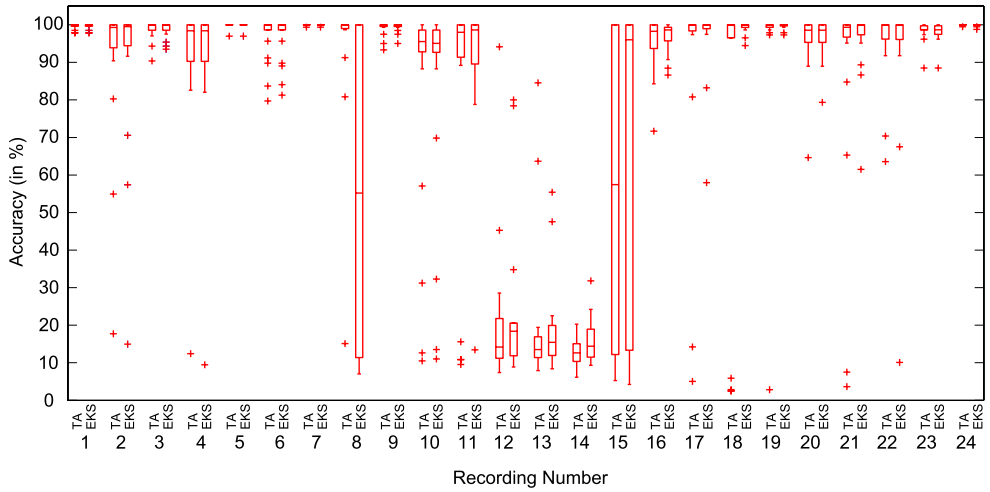


Figure 9. Boxplot showing the accuracy inter and intrapacient variability for our own clinical data using TA and EKS. Outliers are represented with +, the first and third quartiles depict the box's ends, the line within the box represents the median and non-outlier maximum and minimum values are at the ends of the whiskers.

can be attributed to the more challenging data (see figure 10). Leaving Leipzig recordings 12–14 out of the accuracy calculation increases ACC using TA to 91.9 ± 10.1% and for EKS 92.4 ± 12.1%. By weighting between the signal periodicity and morphology a more consistent approach than mere interpolating missing peaks is obtained. Despite its capabilities, the evolutionary approach should be further studied. For example, additional sub-functions to capture morphology and/or periodicity could be easily included.

In general our methodology was able to perform robust FQRS detections for a total of 175 min (Challenge data) and approximately 470 min (own data). However, since neither Challenge nor our own data contained recordings with significant arrhythmia or ectopic beats, a more complete noise stress test is required. For instance, the negative effects of post-processing maternal/fetal QRS detections was not evaluated. Indeed, such a heart rate correction stage should be avoided, for example by better adjusting the EA detector's sub-functions to only allow RR-series smoothing in segments where the signal is not buried into noise. An inference about this work's clinical relevance is not possible due to the difficulty in obtaining FECG recordings with reliable FQRS annotations (for gold standard) and the fact that fetal electrocardiography is still in its early stages. Therefore, only a comparison between methods is possible. For the purpose of benchmarking methods for FECG extraction, the simulator presented in Behar *et al* (2014) and Sameni *et al* (2007) is a better option since it allows a variety of realistic scenarios.

5. Conclusion

In this work, we presented the workflow for our winning closed-source entry on the Computing in Cardiology 2013. The scheme was further evaluated and validated using our own clinical recordings (comprising 7.8 h of multichannel abdominal data). The described methods were able to adapt to the extended clinical test data, performing satisfactory FECG extraction for robust FQRS detection. After our analysis, the most promising direction for future research

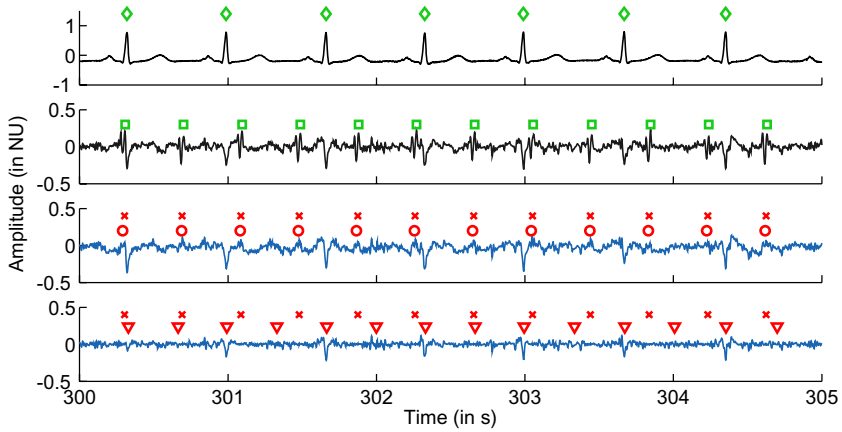


Figure 10. 5 s segment of own clinical dataset 8, where extraction fails due erroneous MQRS detections. Maternal reference is marked by (◇), MQRS detections with (□), fetal reference by (×), FQRS detections over TA output by (○) and FQRS detections over EKS output (▽). The first signal (top) shows channel 1 for comprehension purposes (however, it was not used in this study); the second signal shows abdominal ECG from channel 6; the third signal shows extraction results for TA; lastly the extraction result for EKS is shown.

might be the extension of the proposed Kalman based approach using additional information, e.g. expected amplitudes. By doing so, the mathematical framework of EKS and the flexibility of TA can be combined to generate better maternal estimates.

Acknowledgments

The authors thank S Schröder and S Löther from University Hospital of Leipzig for providing very valuable recordings and helping us to produce reliable annotations. FA thanks the CNPq (Conselho Nacional de Desenvolvimento Tecnológico—Brazil) for its financial support. MR is supported by the DFG (Deutsche Forschungsgemeinschaft—RI2016/2-1).

References

Akhbari M, Niknazar M, Jutten C, Shamsollahi M and Rivet B 2013 Fetal electrocardiogram R-peak detection using robust tensor decomposition and extended Kalman filtering *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 189–92

Amer-Wåhlin I and Maršál K 2011 ST analysis of fetal electrocardiography in labor *Semin. Fetal Neonatal Med.* **16** 29–35

American College of Obstetricians and Gynecologists 2009 ACOG practice bulletin no. 106: intrapartum fetal heart rate monitoring: nomenclature, interpretation, and general management principles *Obstet. Gynecol.* **114** 192–202

Andreotti F, Riedl M, Himmelsbach T, Wedekind D, Zaunseder S, Wessel N, Malberg H and Andreotti F 2013 Maternal signal estimation by Kalman filtering and template adaptation for fetal heart rate extraction *Computing in Cardiology (Zaragoza, Sept. 2013)* pp 193–6

American National Standard 1998 Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms *ANSI/AAMI EC57:1998* Association for the Advancement of Medical Instrumentation (Arlington, VA)

Bäck T and Schwefel H-P 1993 An overview of evolutionary algorithms for parameter optimization *Evol. Comput.* **1** 1–23

Bailey R E 2009 Intrapartum fetal monitoring *Am. Family Phys.* **80** 1388–96

Behar J, Oster J and Clifford G 2013 Non-invasive FECG extraction from a set of abdominal sensors *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 297–300
Behar J, Andreotti F, Zaunseder S, Li Q, Oster J and Clifford G D 2014 An ECG simulator for generating maternal-foetal activity mixtures on abdominal ECG recordings *Physiol. Meas.* **35** 1537–50
Callaerts D, Moor B, Vandewalle J, Sansen W, Vantrappen G and Janssens J 1990 Comparison of SVD methods to extract the foetal electrocardiogram from cutaneous electrode signals *Med. Biol. Eng. Comput.* **28** 217–24
Clifford G, Silva I, Behar J and Moody G 2014 Non-invasive fetal ECG analysis *Physiol. Meas.* **35** 1521–35
De Lathauwer L, De Moor B and Vandewalle J 2000 Fetal electrocardiogram extraction by blind source subspace separation *IEEE Trans. Biomed. Eng.* **47** 567–72
Fatemi M, Niknazar M and Sameni R 2013 A robust framework for noninvasive extraction of fetal electrocardiogram signals *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 201–4
Haghpanahi M and Borkholder D 2013 Fetal ECG extraction from abdominal recordings using array signal processing *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 173–6
Hyvärinen A 1999 Fast and robust fixed-point algorithms for independent component analysis *IEEE Trans. Neural Netw.* **10** 626–34
Kanjilal P, Palit S and Saha G 1997 Fetal ECG extraction from single-channel maternal ECG using singular value decomposition *IEEE Trans. Biomed. Eng.* **44** 51–9
Karvounis E, Tsipouras M, Fotiadis D and Naka K 2007 An automated methodology for fetal heart rate extraction from the abdominal electrocardiogram *IEEE Trans. Inform. Technol. Biomed.* **11** 628–38
Kirkpatrick S, Gelatt C D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80
Kropf M, Modre-Osprian R, Schreier G and Hayn D 2013 A robust algorithm for fetal GRS detection using non-invasive maternal abdomenal ECGs *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 313–6
Langley P, Bowers E and Murray A 2010 Principal component analysis as a tool for analyzing beat-to-beat changes in ecg features: application to ecg-derived respiration *IEEE Trans. Biomed. Eng.* **57** 821–9
Lipponen J and Tarvainen M 2013 Advanced maternal ECG removal and noise reduction for application of fetal QRS detection *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 161–4
Lukosevicius M and Marozas V 2013 Noninvasive fetal QRS detection using echo state network *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 205–8
Martens S 2003 Extraction of the fetal ECG and the uterine contraction EMG from simulated abdominal recordings by principal component analysis *Master Thesis* Technische Universiteit Eindhoven (http://alexandria.tue.nl/extra1/afstversl/E/570506.pdf)
Martens S M M, Rabotti C, Mischi M and Sluijter R J 2007 A robust fetal ECG detection method for abdominal recordings *Physiol. Meas.* **28** 373–88
McSharry P, Clifford G, Tarassenko L and Smith L 2003 A dynamical model for generating synthetic electrocardiogram signals *IEEE Trans. Biomed. Eng.* **50** 289–94
Nemati S, Malhotra A and Clifford G D 2010 Data fusion for improved respiration rate estimation *EURASIP J. Adv. Signal Process.* **2010** 926305
Niknazar M *et al* 2012 Application of dynamic time warping on Kalman filtering framework for abnormal ECG filtering *20th European Symp. on Artificial Neural Networks (ESANN) (Bruges, Apr. 2012)* pp 139–44
Peters M, Crowe J, Piéri J F, Quartero H, Hayes-Gill B, James D, Stinstra J and Shakespeare S 2001 Monitoring the fetal heart non-invasively: a review of methods *J. Perinatal Med.* **29** 408–16
Pohlheim H 1999 *Evolutionäre Algorithmen: Verfahren, Operatoren, Hinweise aus der Praxis* (Berlin: Springer)
Richter M, Schreiber T and Kaplan D 1998 Fetal ecg extraction with nonlinear state-space projections *IEEE Trans. Biomed. Eng.* **45** 133–7
Rieta J, Castells F, Sanchez C, Zarzoso V and Millet J 2004 Atrial activity extraction for atrial fibrillation analysis using blind source separation *IEEE Trans. Biomed. Eng.* **51** 1176–86
Rompelman O and Ros H 1986 Coherent averaging technique: a tutorial review part 1: noise reduction and the equivalent filter *J. Biomed. Eng.* **8** 24–9
Sameni R 2008 Extraction of fetal cardiac signals from an array of maternal abdominal recordings *PhD Thesis* Sharif University of Technology, Institut National Polytechnique de Grenoble (www.sameni.info/Publications/Thesis/PhDThesis.pdf)
Sameni R 2010 *The Open-Source Electrophysiological Toolbox (OSET), Version 2.1* (www.oset.ir)

Sameni R and Clifford G D 2010 A review of fetal ECG signal processing; issues and promising directions *Open Pacing Electrophysiol. Therapy J.* **3** 4–20
Sameni R, Clifford G D, Jutten C and Shamsollahi M B 2007 Multichannel ECG and noise modeling: application to maternal and fetal ECG signals *EURASIP J. Adv. Signal Process.* **2007** 14
Sameni R, Shamsollahi M-B, Jutten C and Babaie-Zadeh M 2005 Filtering noisy ECG signals using the extended Kalman filter based on a modified dynamic ECG model *Computers in Cardiology (Lyon, Sept. 2005)* pp 1017–20
Silva I, Behar J, Sameni R, Zhu T, Oster J, Clifford G and Moody G 2013 Noninvasive fetal ECG: the PhysioNet/Computing in Cardiology Challenge 2013 *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 149–52
Silverman B W 1986 *Density Estimation for Statistics and Data Analysis* vol 26 (London: Chapman and Hall)
Ungureanu G M, Bergmans J W M, Oei S G, Ungureanu A and Wolf W 2009 Comparison and evaluation of existing methods for the extraction of low amplitude electrocardiographic signals: a possible approach to transabdominal fetal ECG *Biomed. Technol.* **54** 66–75
Ungureanu M and Wolf W 2006 Basic aspects concerning the event-synchronous interference canceller *IEEE Trans. Biomed. Eng.* **53** 2240–7
van Laar J, Peters C, Vullings R, Houterman S, Bergmans J and Oei S 2009 Fetal autonomic response to severe acidaemia during labour *Int. J. Obstet. Gynaecol.* **117** 429–37
Varanini M, Tartarisco G, Billeci L, Macerata A, Pioggia G and Balocchi R 2013 A multi-step approach for non-invasive fetal ECG analysis *Computing in Cardiology Conf. (CinC) (Zaragoza, Sept. 2013)* pp 281–4
Wedekind D, Malberg H and Zaunseder S 2013 Cascaded output selection for processing of capacitive electrocardiograms by means of independent component analysis *2013 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF) (Bonn, Oct. 2013)* pp 1–6
Whitley D 2001 An overview of evolutionary algorithms: practical issues and common pitfalls *Inform. Softw. Technol.* **43** 817–31
Widrow B, Glover J R Jr, McCool J, Kaunitz J, Williams C, Hearn R, Zeidler J, Eugene Dong J and Goodlin R 1975 Adaptive noise cancelling: principles and applications *Proc. IEEE* **63** 1692–716
Zarzoso V and Nandi A 2001 Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation *IEEE Trans. Biomed. Eng.* **48** 12–8
Zaunseder S, Andreotti F, Cruz M, Stepan H, Schmieder C, Malberg H, Jank A and Wessel N 2013 Fetal QRS detection by means of Kalman filtering and using the event synchronous canceller *Int. J. Bioelectromagn.* **15** 83–9