

Editorial

Non-invasive fetal ECG analysis

Gari D Clifford^{1,3,4,5}, Ikaro Silva², Joachim Behar¹ and George B Moody²

¹ Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

² Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

³ Department of Biomedical Informatics, Emory University, Atlanta, GA 30322, USA

⁴ Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

E-mail: gari@alum.mit.edu, george@mit.edu, joachim.behar@eng.ox.ac.uk and ikaro@mit.edu

Received 6 July 2014

Accepted for publication 7 July 2014

Published 28 July 2014

Abstract

Despite the important advances achieved in the field of adult electrocardiography signal processing, the analysis of the non-invasive fetal electrocardiogram (NI-FECG) remains a challenge. Currently no gold standard database exists which provides labelled FECG QRS complexes (and other morphological parameters), and publications rely either on proprietary databases or a very limited set of data recorded from few (or more often, just one) individuals.

The PhysioNet/Computing in Cardiology Challenge 2013 enables to tackle some of these limitations by releasing a set of NI-FECG data publicly to the scientific community in order to evaluate signal processing techniques for NI-FECG extraction. The Challenge aim was to encourage development of accurate algorithms for locating QRS complexes and estimating the QT interval in non-invasive FECG signals. Using carefully reviewed reference QRS annotations and QT intervals as a gold standard, based on simultaneous direct FECG when possible, the Challenge was designed to measure and compare the performance of participants' algorithms objectively. Multiple challenge events were designed to test basic FHR estimation accuracy, as well as accuracy in measurement of inter-beat (RR) and QT intervals needed as a basis for derivation of other FECG features.

This editorial reviews the background issues, the design of the Challenge, the key achievements, and the follow-up research generated as a result of the Challenge, published in the concurrent special issue of *Physiological Measurement*.

⁵ Author to whom any correspondence should be addressed.

Keywords: non-invasive foetal ECG, fetal monitoring, challenge, PhysioNet

(Some figures may appear in colour only in the online journal)

1. Introduction

Since the late 19th century, decelerations of fetal heart rate have been known to be associated with fetal distress. Intermittent observations of fetal heart sounds (auscultation) became standard clinical practice by the mid-20th century. The first fetal heart rate (FHR) monitors were developed more than 50 years ago, and became widely available by the mid-1970s. Continuous FHR monitoring was expected to result in dramatic reduction of undiagnosed fetal hypoxia, but disillusionment rapidly set in as studies showed that the outputs of FHR monitors were often unreliable and difficult to interpret, large increase rates of a painful and expensive cesarean section, higher prevalence of postnatal depression (Boyce and Todd 1992) and postoperative pain negatively affecting breastfeeding and infant care (Karlström *et al* 2007). There was little evidence that reductions in adverse outcomes were attributable to the use of FHR monitors.

Improved accuracy in FHR estimation has been achieved through use of more sophisticated signal processing techniques applied to more reliable signals. These improvements, coupled with a better understanding of the limitations of fetal monitoring, have led to wider acceptance. However, there remains a great deal of room for improvement.

Electronic fetal monitoring techniques can be invasive or non-invasive with intermittent or continuous assessment; these techniques include fetal phonocardiography, Doppler ultrasound, cardiotocography (CTG), fetal magnetocardiography (FMCG) and fetal electrocardiography (FECG)—see table 1. At 20 weeks the heart can be heard without amplification (Sameni and Clifford 2010) and monitored by ultrasound (Peters *et al* 2001), the FECG and FMCG can be recorded from 20 weeks onward (Peters *et al* 2001)—figure 1. Doppler ultrasound is routinely used for FHR monitoring during pregnancy and delivery. However, it has not been demonstrated that ultrasound irradiation exposure was completely safe for the foetus (Barnett and Maulik 2001). The FECG can be recorded in two ways; through an electrode attached (screwed) to the fetal scalp while the cervix is dilated (i.e. during delivery) or by non-invasive electrodes placed on the mother's abdomen—non-invasive FECG (NI-FECG).

The ECG allows for interpretation of the electrical activity of the heart far beyond just heart rate and heart rate variability. However morphological analysis of the FECG waveform is usually not performed in clinical practice, with exception of the STAN monitor (Neovinta Medical, Goteborg, Sweden), which uses an invasive scalp electrode. This electrode can only be placed at the very last stage of the pregnancy (antepartum) and has an associated small risk. It is therefore not routinely used. Moreover only one differential electrode is possible, thus the three dimensional electrical field emanating from the fetal heart is unavailable, and only singletons can be monitored. Conversely, the NI-FECG is non-invasive and can theoretically be performed at earlier stages of the pregnancy (although with a weaker field strength). However, the NI-FECG always manifests as a mixture of (significant) noise, fetal activity (from each fetus) and a much larger amplitude of maternal activity (figure 2(a)). The signals overlap in both the time (figure 2(a)) and frequency domains (figure 2(b)) and therefore accurate extraction and analysis of the FECG waveform is challenging.

The FECG was first observed more than a century ago (Cremer 1906). Despite significant advances in adult clinical electrocardiography, signal processing techniques and the potency of digital processors, few significant advances have been made in the extraction and analysis of NI-FECG. This is partly due to the relatively low signal-to-noise ratio (SNR) of the FECG

Table 1. Main methods for non-invasive electronic fetal monitoring. Main reference (Peters *et al* 2001).

Method	System	Gestational age	Comments
CTG	Cardiotography; ultrasound transducer and uterine contraction pressure-sensitive transducer	≥ 20 weeks	<ul style="list-style-type: none"> – contraction through pressure transducer – smoothed HR time series. Rather robust and reliable – no beat to beat data and cardiac function descriptor limited to HR – not passive; ultrasound irradiation
FMCG	Fetal magnetocardiogram. Detection of the fetal heart's magnetic field through SQUID sensors positioned near the maternal abdomen	≥ 20 weeks	<ul style="list-style-type: none"> – expensive – requires skilled personnel – morphological analysis of the FMCG easier than NI-FECG because of higher SNR – no long term monitoring possible to date because of apparatus size/coast etc.
NI-FECG	Standard ECG electrodes with varying skin preparation methods	≥ 20 weeks with dip from 28th to 37th weeks	<ul style="list-style-type: none"> – cheap – easy to handle – continuous monitoring possible – FHR and possibly morphological analysis – low SNR

compared to the maternal ECG (MECG), caused by the various media between the fetal heart and the measuring electrodes, and the fact that the fetal heart is simply smaller. Moreover, there is a less complete clinical knowledge concerning fetal cardiac function and development than for adult cardiology. Another significant barrier to the analysis of NI-FECG is the paucity of (public) gold standard databases with expert annotations and objective signals, such as independent measures of the FECG (through direct scalp electrodes), heart rate, ischemia, rhythm, etc.

The key features in fetal monitoring are FHR rhythm-related, and FECG morphology related (e.g. ST and QT changes). FHR can be used as an indicator of fetal distress (Van Geijn *et al* 1991). In medical practice 1-D Doppler ultrasound is usually used to measure the FHR, but this requires frequent repositioning of the ultrasound transducer and its accuracy is often well below that of the scalp electrode. Moreover, little progress has been made in the use of FHR to provide clinically actionable information. In contrast, some studies have shown that FECG morphology was promising in identifying actionable abnormalities. This includes the QT interval (Oudijk *et al* 2004), QRS morphology, and the ST segment (Clifford *et al* 2011). In particular, it is known that QT interval reacts to situations of stress and exercise. It has been shown that a significant shortening of the QT interval was associated with intrapartum hypoxia (resulting in metabolic acidosis) irrespectively of changes in FHR, whereas in normal labour these changes do not occur (Oudijk *et al* 2004).

The scalp ECG based STAN monitor provides a proxy measure (the T/R amplitude ratio) for the ST segment deviation. Recently the use of the STAN analyser together with competency based training on fetal monitoring showed significant decrease in the number of cesarean at St George's Maternity Unit while hypoxic ischaemic encephalopathy and early

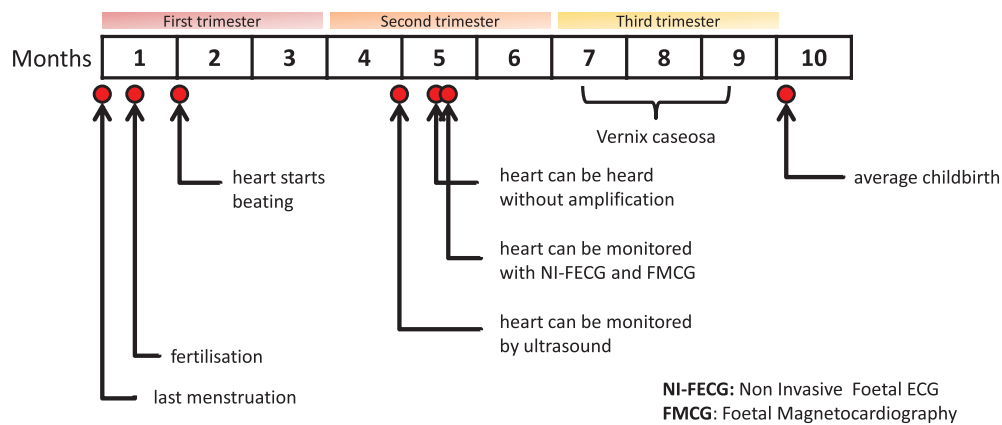


Figure 1. Prenatal development time-line with key landmark with respect to fetal monitoring. At 20 weeks the heart can be heard without amplification (Sameni and Clifford 2010), and monitored using Doppler ultrasound (Peters *et al* 2001), the Non invasive FECG (NI-FECG) and FMCG can be recorded from 20 weeks onward (Peters *et al* 2001) but the vernix caseosa forms around 28th–32nd weeks and dissolves in 37th–38th weeks in normal pregnancies (Stinstra 2001) limiting NI-FECG effectiveness recording during this period.

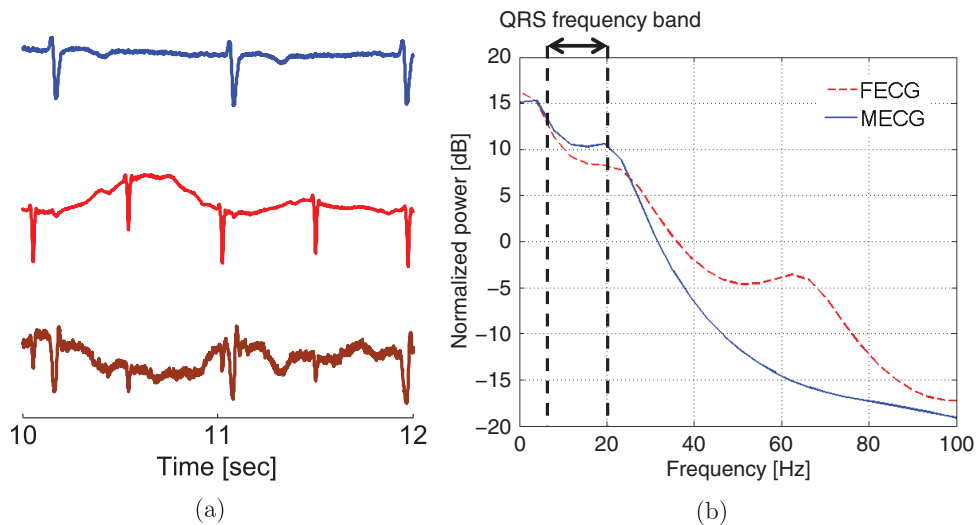


Figure 2. Frequency and temporal overlap of the MECG and FECG signals. (a) From top to bottom: example of maternal chest ECG, fetal scalp ECG and abdominal ECG (AECG). Note that the AECG contains a mixture of both MECG and FECG and that some fetal QRS are overlapping with the maternal QRS—temporal overlap. To produce (a) a notch filter at 60 Hz was used to make the FECG visible on the abdominal channel. (b) Power spectral density distribution (Burg method, order 20) for 5 min of scalp electrode ECG and 5 min of adult ECG. Notice the frequency overlap between the adult and fetal ECG signals particularly in the frequency band of the QRS.

neonatal death decreased slightly (Chandrabaran *et al* 2013). However a recent Cochrane study (Neilson 2006) reviewed six trials aiming at comparing the effect of the analysis of scalp FECG waveforms during labour with alternative methods used for fetal monitoring and showed that no significant difference to primary outcomes were achieved using the STAN ST proxy (evaluated on five trials using different version of the STAN monitor, 15 338 women). This suggests that whether the STAN proxy measure for computing ST is not accurate enough or that ST measure does not provide significant information to improve fetal monitoring.

To date there are only two NI-FECG devices known to the authors that have obtained FDA clearance and regularly published papers on NI-FECG analysis: the Monica AN24 monitor (Monica Healthcare, Nottingham, UK) and the MERIDIAN monitor from MindChild Medical (North Andover, MA). Both monitors have proved to be accurate in detecting the FHR and early work on extracting morphological information has been published (Behar *et al* 2014a, Clifford *et al* 2011). These recent advances in the field are very exciting. However, these studies are still limited in number and population size, and the positive outcomes of these devices on fetal monitoring are yet to be established.

Until the PhysioNet/Computing in Cardiology Challenge 2013 (*the Challenge*) there were three public NI-FECG databases: (i) the Daisy database constituted of 8 channels (4 abdominal and 3 thoracic) and the abdominal ECG (AECEG) lasting for 10 s and using a sampling frequency (fs) of 250 Hz. (ii) The Non-Invasive Fetal Electrocardiogram Database (NIFECGDB), available on PhysioNet (Goldberger *et al* 2000) $fs = 1$ kHz. 55 multichannel abdominal ECG recordings taken from a single subject (21 to 40 weeks of gestation), $fs = 1$ kHz, without reference annotations. (iii) Abdominal and Direct Fetal Electrocardiogram Database (ADFECGDB), available on PhysioNet (Goldberger *et al* 2000) $fs = 250$ Hz with 5 min of recordings (4 abdominal channels) from 5 women in labour (38 to 41 weeks of gestation), $fs = 1$ kHz, scalp ECG available for reference. It is important to note that these three databases are low dimensional (number of recordings, number of abdominal channels available) and few data have any reference annotations, and those that do, only have FQRS complex location from a single annotator.

In summary NI-FECG has the potential to provide:

- Fetal heart rate
- ECG morphological information such as PR, ST, QT intervals
- Contraction monitoring (as in Hayes-Gill (2012))
- Fetal movement (as suggested in Sameni (2008)) & fetal position

The most accurate method for measuring FHR is direct fetal electrocardiographic (FECG) monitoring using a fetal scalp electrode. This is possible only in labour, however, and is not common in current clinical practice, except in deliveries considered to be high risk, because of the associated risks of the scalp electrode usage. Non-invasive FECG monitoring makes use of electrodes placed on the mother's abdomen. This method can be used throughout the second half of pregnancy and is of negligible risk, but it is often difficult to detect the fetal QRS complexes in ECG signals obtained in this way, since the maternal ECG is usually of greater amplitude in them.

2. Overview of the challenge 2013

The key questions of the Challenge were: (1) Can accurate FHR measurements be performed using a set of non-invasive abdominal ECG electrodes? and (2) Can an accurate fetal QT measure be performed in an automated way using the extracted signal? Despite many interesting

Table 2. FECG database reference.

Database name	N records
ADFECGDB (Matonia <i>et al</i> 2006)	25
Simulated FECGs (Behar <i>et al</i> 2014b)	20
NIFECGDB (Goldberger <i>et al</i> 2000)	14
Non-invasive FECG	340
Scalp FECG database	48
Total	447

theoretical frameworks, the robustness of most of the methods for NI-FECG extraction in the literature to date has not been sufficiently quantitatively evaluated. This is due to two main factors : (i) the lack of gold standard databases with expert annotations and (ii) the methodology for assessing the algorithms. The Challenge attempts to address these limitations by making publicly available a set of FECG data to the scientific community for evaluation of signal processing techniques, as well as a scoring system for evaluating the outcomes of these methods.

The data sets used for the Challenge were obtained from five different sources, table 2, yielding a total of 447 records. Two out of the five databases have been previously made public (Goldberger *et al* 2000, Matonia *et al* 2006), and one database was artificially generated using an extended version (Behar *et al* 2014b) of the dipole model described in Sameni *et al* (2007). The other two databases were donated to PhysioNet for this Challenge (the Scalp FECG Database was not made public and used only for scoring open source algorithms in Set C described below). The gold standard used for the initial stage of the Challenge consisted of reference annotations from the data sets (for the non-invasive data sets, the annotations were obtained from FECG QRS estimates derived manually or through additional maternal ECG leads that were not available to competitors). The reference for the second and final stage of the Challenge was obtained by using a Bayesian crowd-sourcing approach (Zhu *et al.*, 2013, 2014) to combine the original reference annotations with the annotations from the all the open-source entries for the first stage. A subset of both the initial and final reference annotations were manually verified by the Challenge organisers, although some minor errors in annotations persisted.

All records were formatted to have a 1 kHz sampling frequency, one minute duration, and four channels of non-invasive abdominal maternal ECG leads. The databases in table 2 were re-arranged into three data sets for the Challenge:

- Set A: 75 records, both records and expert annotations were made public
- Set B: 100 records, only the records were made public
- Set C: 272 records, both records and expert annotations were withheld from the public

The Challenge scores relative to the different events were defined as follows; scores for the FHR based events (E1 and E4) were computed from the differences between matched reference and test FHR measurements at 12 instances (i.e. one every 5 s). Scores for the RR events (E2 and E5) were computed from the differences between matched reference and test RR intervals. The score for the QT measure event (E3) was calculated from the differences between matching reference and test QT intervals. The purpose of the RR events was to assess whether an algorithm was able to extract the absolute FQRS position, i.e. the position of the fetal R-peak on the signal with respect to the reference fiducial markers. The purpose of the FHR events was to assess the performance of an algorithm for providing clinically relevant information, regardless of where the fetal R-peaks were located (so the FQRS time series could be highly smoothed before computing the FHR). As such, the RR and FHR scores represented two distinct events even if they ended highly correlated as the results of the Challenge

Table 3. Scoring methods for the records of the Challenge. E stands for event.

Estimation task	Scoring method	Units	Event
FHR Series	Beat by beat classification error	(beats min ⁻¹) ²	E1, E4
RR Series	Average root square error	milliseconds	E2, E5

showed (Silva *et al* 2013). The Challenge was divided into three phases corresponding to three time periods where participants were allowed to submit a certain number of entries (phase 1: from 25-04-2013 to 01-06-2013, 3 entries; phase 2: from 01-06-2013 to 25-08-2013, five entries; phase 3: from 25-08-2013 to 05-09-2013, 1 entry).

The participants of the Challenge were expected to use set A for the training of their algorithms while sets B and C were used by the Challenge organisers for scoring. It was not possible to score one record in Set A and two in Set B due to some errors in the corresponding reference annotations. The training data set A, and the records for set B, are publicly available at PhysioNet. The Challenge was organised into five events (E): a QT estimation event, and four time series estimation which are the focus of this special issue. The four time series events were defined as presented in table 3: events E1 and E2 were only considered for the open source entries (evaluated on set C); events E4 and E5 were considered for open or closed source entries.

E1 and E2 were scored on a private PhysioNet server running the participant's algorithm on set C. E4 and E5 were automatically scored on PhysioNet's web server by comparing the user's submitted annotation file with the expert annotations. The web based scoring interface in PhysioNet remains open for those wishing to compare their results with those from the official Challenge on events 4 and 5. The scoring methods for the three different FECG estimation tasks are described in table 3. Records that were not annotated by the competitors were given a very high penalty value. The WaveForm DataBase (WFDB) software package version 10.5.19 was used in the scoring of the events related to FHR and RR Series. The final score for a given event was determined by the average score of all the records within the event's data set. The source code used for all scoring remains available at <http://physionet/challenge/2013> and the source code from the open-source competitors can be found at <http://physionet.org/challenge/2013/sources/>.

An open source sample entry was provided to the participants by the organisers of the Challenge. The competitors were welcome to either improve the sample entry or generate their entry following the same interface as of the sample entry. A total of 53 international teams participated in the Challenge yielding 208 sets of annotations and 93 open source entries, with the vast majority outperforming the sample entry (figure 3). The top scores for all the events (E) were: 179.439 (beats min⁻¹)² (E1), 20.793 ms (E2), 18.083 (beats min⁻¹)² (E4), and 4.337 ms (E5). Results presented at the Computing in Cardiology conference 2013 are presented in Table 4. Following the Challenge some participants further refined their algorithms and their updated scores reported in this special issue are summarised in Table 5. Note that only the scores from the participants who submitted a follow-up paper in this special issue are listed in the tables.

3. Review of key algorithms in the challenge

A large number algorithms for FHR and RR series estimation were proposed in the Challenge. The aim of this section is to expose several of the different signal processing techniques that lead to successful fetal ECG estimation. Other algorithms also obtained good scores at the

Table 4. Results presented at the Computing in Cardiology conference 2013 for events 1-5 (E1-E5) for all the papers presented in this special issue. NA: Not Available, because the corresponding participants did not entered the open source events E1-E2. E1 and E4 in bpm² and E2, E5 in ms.

Participants/Events	E1	E2	E4	E5
Andreotti <i>et al</i> (2014)	NA	NA	<u>18.1</u>	<u>4.3</u>
Behar <i>et al</i> (2013) (non-official)	<u>179.4</u>	<u>20.8</u>	29.6	4.7
Haghpanahi and Borkholder (2013)	6298.1	159.9	50.1	9.1
Varanini <i>et al</i> (2013)	<u>187.1</u>	<u>21.0</u>	34.0	5.1
Dessì <i>et al</i> (2013)	684.2	48.0	639.5	23.8
Lipponen and Tarvainen (2013)	NA	NA	28.9	4.8
Di Maria <i>et al</i> (2013)	NA	NA	223.2	19.3
Liu and Li (2013)	2782.3	81.7	264.9	9.0
Lukoševičius and Marozas (2013)	NA	NA	66.3	8.2
Rodrigues (2013)	278.8	28.2	124.8	14.4
Christov <i>et al</i> (2013)	NA	NA	285.1	20.0
Almeida <i>et al</i> (2013)	NA	NA	521.4	33.0

Table 5. Challenge results for the algorithms presented in this special issue i.e. considering further development after the Challenge deadline. NA: Not Available. E1 and E4 in bpm² and E2, E5 in ms.

Participants/Events	E1	E2	E4	E5
Andreotti <i>et al</i> (2013)	NA	NA	<u>15.1</u>	<u>3.3</u>
Behar <i>et al</i> (2014c)	<u>179.4</u>	<u>20.8</u>	29.6	4.7
Haghpanahi and Borkholder (2014)	NA	NA	50.1	9.1
Varanini <i>et al</i> (2014)	187.0	21.0	34.0	5.1
Dessì <i>et al</i> (2014)	281.1	25.93	134.5	12.4
Lipponen and Tarvainen (2014)	NA	NA	28.9	4.8
Di Maria <i>et al</i> (2014)	NA	NA	142.7	19.9
Liu and Li (2014)	NA	NA	47.5	7.6
Lukoševičius and Marozas (2014)	NA	NA	66.3	8.2
Rodrigues (2014)	278.8	28.2	124.8	14.4
Christov <i>et al</i> (2014)	NA	NA	305.7	23.1
Almeida <i>et al</i> (2014)	NA	NA	513.1	35.3

Challenge, however, due to obvious space limitations, it is not practical to mention all of them here. The techniques presented at the Challenge were unique and original, but in general each had a five step approach as follows:

1. Pre-processing
2. Estimation of maternal component
3. Removal of maternal component
4. Estimation of FHR and RR time series
5. Post-processing

The first step generally consists of pre-processing the raw waveforms. In this stage noise, artifacts, baseline wandering (i.e. trends), and power-line interference are removed through the use of filters, averaging, or median filtering. In some cases, an augmented set of channels is also obtained through algebraic manipulation of the existing ones, for instance, by subtracting pairs of signals or inverting individual ones, thus creating a differential signal. At the second stage, an estimate of the maternal signal is obtained by using a form of decomposition, filtering, template generation, or a combination of these three. The two most common forms of subspace decomposition used were Independent Component Analysis (ICA) and Singular Value Decomposition (SVD). For approaches that used a maternal template, the template is

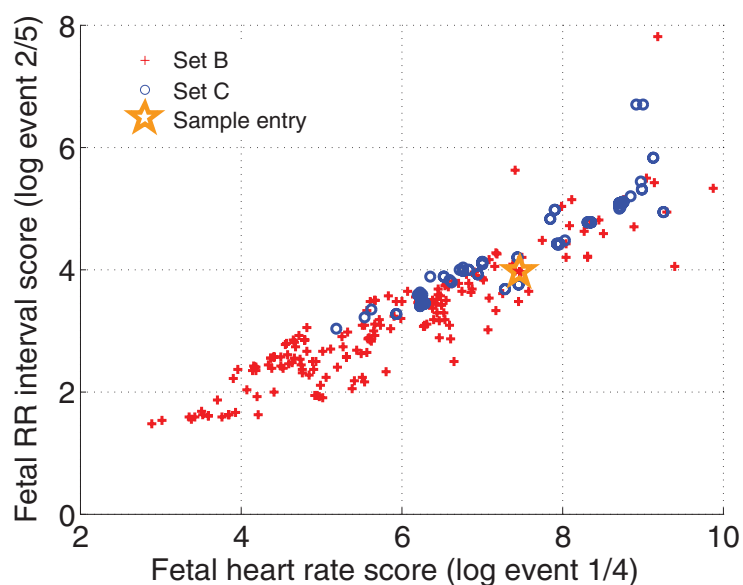


Figure 3. Scatter plot of the scores for the Challenge (best scores are in the lower left corner). Scores for set C and B are marked in blue and red, respectively. The score for the sample entry is highlighted in green.

usually estimated by averaging detected MQRS across space (i.e. channels) and/or time, with the Pam Tompkins algorithm being a popular choice for MQRS detection (Pan and Tompkins 1985). Additionally, an estimated measure of signal quality can be used to weight the channels during the averaging process. In some cases, the temporal template is further decomposed into set of parameters through curve fitting. At the third stage, the maternal component is removed from the waveforms through a combination of one or more of the following techniques: subspace reconstruction, maternal template subtraction (signal cancelling), filtering, and/or asynchronous temporal windowing (temporal gating). The subspace reconstruction (typically done with ICA or SVD) is performed by setting the components of non-fetal subspaces to zero. Signal subtraction using the estimated maternal templates can be performed statically or adaptively. Adaptive methods tend to track changes in curved-fitted parameters or use of adaptive filters such the Kalman and Least Mean Square (LMS) filters. After the maternal component has been attenuated in the third stage, the fourth step is to estimate the FQRS. The FQRS estimation can be performed through RS slope and R amplitude threshold detection, or modification of any of the existing adult QRS detection techniques. In some algorithms, the fourth stage is also accompanied by the merging of QRS annotations from the different channels and/or different QRS detection algorithms (for example, using median voting). The fifth and final step (applied only by some of the competitors) is to constrain the estimated FHR and RR time series through physiological or statistical limits based on heuristics.

As expected, technical challenges and limitations exist in each of the five steps described above. At the pre-processing level, we have the task of designing band-pass and notch filters that will maximally attenuate noise without significantly distorting either the maternal or fetal ECG components. Estimation and removal of the maternal component, (the second and third steps), faces even harder challenges. Algorithms that use a subspace decomposition and reconstruction may make the following implicit assumptions: (a) the number of signal sources

are fixed, discrete, and less than or equal to the number of recorded channels, (b) the subspace representation is stationary, (c) the sources are uncorrelated, (d) the maternal signal has a high signal-to-noise ratio and spans one or more of the dominant spaces (with maternal P, QRS, and T waves possibly projecting into separate spaces). In some cases, the assumption that the data dimension space is larger than the number of independent sources can be resolved by pre-processing the data (via filtering or cancellation, for example). Alternative methods that use maternal template cancellation instead of subspace decomposition/reconstruction also make assumptions. Some of the key assumption of the maternal template cancellation are: (a) the maternal component is uncorrelated with the fetal component, (b) the relationship between the ECG leads are stationary (or short-term stationary) and ergodic, (c) the maternal and fetal wave morphology are either constant, have slow trackable changes, or with no ectopic beats. It was observed that some of these assumptions did not hold for the Challenge data (Di Maria *et al* 2014), for instance, remarks that the MQRS detection is suboptimal if limited to always the first principal component.

The estimation of fetal heart rate and RR series performed by Andreotti *et al* (2013) consisted of five major stages. The first was a pre-processing stage to remove baseline wander, muscle artifact, and power line interference through zero-phase FIR filtering. In the second stage, the MECG was then estimated through a process that begins with Independent Component Analysis (ICA) to generated pseudo-channels. A QRS algorithm was run in all of these channels, and a best channel was selected by comparing the individual channels with a Gaussian kernel based QRS agreement of all channels. The chosen optimal channel was then used to generate a template MQRS for MQRS detection. In some instances, dependent on the MQRS amplitude, the original four channels were subtracted from each other to further generate an augmented set of eight channels. At the third stage, the MQRS was removed from the waveforms using two different approaches: Extended Kalman Filtering based on Sameni *et al* (2005), and maternal template adaptation. The Extended Kalman Filtering approach was based on a non-linear system model of the averaged maternal beat and the inclusion of an innovation process. On the other hand, the MECG template adaptation sought to segment the QRS complex into three distinct sections with whose width and heights were tracked and allowed to vary (the heights of these components were limited to a maximum range in order to avoid interfering with any superimposed FECG). The fourth major stage consisted of FQRS detection. The FQRS detection was treated as an optimisation problem in which fetal beat morphology and beat-to-beat interval consistencies were part of the cost function. Thus, simulated annealing was used as the optimisation tool, with independent FQRS annotations of processed channels as the input. The fifth and final stage of the Andreotti *et al* (2013) algorithm sought to correct, or constraint, the estimated FHR and RR time series. Among the key corrections were the removal of intervals in the estimated time series that were shorter than 300 ms and heart rates changes greater than 70 ms. The authors remark that although the maternal template adaptation yielded better results (less attenuation of the fetal complexes), the Extended Kalman Filtering approach had a significant potential for improvement.

Another proposed algorithm that performed well in the data set was that of Lipponen and Tarvainen (2013). The pre processing was performed with a 6th order Butterworth high-pass filter (cut-off at 2 Hz) and elimination of the 50 Hz spectral Fourier component. An extra set of channels was then obtained through subtraction of the original channels. The MECG was estimated and eliminated from the waveforms in three steps: (a) MQRS were detected in all channels using (Pan and Tompkins 1985), (b) maternal Q, R, S, P and T wavelets for each epoch were then stacked to generate 5 measurement matrix from which eigen-decompositions were obtained, (c) the individual epochs were then filtered by linearly combining the eigenvectors

with the top six eigenvalues for the QRS wavelets zeroed, and the top four eigenvalues for the P and T wavelets zeroed. Thus MECG removal process assumed that the FECG and noise components were not dominant in the space spanned by the principal eigenvectors. The fetal HR and RR time was estimated in four major steps. First, each channel was normalised by their signal quality factor, estimated by passing the channel's envelope through a 100 ms moving average. At the second step, 20 QRS complexes were detected from the largest peaks that were generated by squaring the waveforms and their channel's sum. The second step was followed by obtaining channel specific FQRS templates were from the average of these 20 locations. In fourth and final step, the estimated time FHR and RR series were calculated by summing the correlation of the channels with their respective templates and passing them through a 30 ms moving average filter.

The algorithm proposed by Varanini *et al* (2013) was also among one the top performers. The algorithm had four pre-processing key steps: (1) sample replacement if a signal's sample was higher than a threshold value based on a 60 ms median filter, (2) the channels were then low pass filtered using a first order Butterworth filter with cut off at 3.17 Hz, (3) a detrended signal was then obtained by subtracting the filtered signal of step 2 with the signal from step 1 and passing the difference through a 260 ms median filter (4) finally, a notch filter was applied if a power line interference was detected (the first 3 harmonics were also removed in similar manner). The MECG was estimated through Independent Component Analysis (ICA), with FastICA as the choice software ICA. A QRS detector was then applied on a band pass filtered version of the major independent components. The maternal beats were then gated in time, and a Singular Value Decomposition (SVD) was then used to model the maternal beat from the first 3 largest singular values. The MECG was then removed from the data by subtracting the MECG SVD model from the signals. The final stage of fetal estimation was similar to the estimation process of the maternal signals. The fetal signal was first enhanced through ICA and two QRS detectors were then applied in forward and backward directions. The estimated annotations were then constrained to be smooth by having small mean absolute values in the first derivative, second derivatives, and in the number of fetal QRSs that matched with maternal QRSs.

The algorithm, (Podziemski and Gieratowski 2013), had a unique approach. The first pre processing stage was similar to other approaches (augmenting channels by inverting them, notch filtering, and median filtering). The second stage, however, was unique in that it attempted to estimate the FHR prior to MECG removal. The FHR was done by first detecting FQRS from threshold detection on RS slope and amplitudes. The thresholds were selected heuristically from the training set and allowed to change adaptively so that individual RR intervals were within 75 ms of the mean. An average channel was generated from the two channels that had the best FHR pair-wise agreement. The maternal ECG was only removed in the third stage. In this third stage, the MQRS was detected using the same RS slope and amplitude technique to detect the FQRS. A maternal MQRS template was then estimated from the detected MQRSs and subtracted from the baseline signal from stage 1. A second set of FQRS was then detected from this residue signal. A covariance signal was then estimated from the first set of FQRS detected on stage 2 and the second set of FQRSs detected on the residual signal. This covariance signal was multiplied with the residual signal and used in a final fetal QRS detection attempt (re-using the RS slope and amplitude techniques). The goal of the final fetal QRS detection pass was to find fetal beats that were potentially missed due to the maternal ECG interference. Finally, a pre-processing stage consisted in removal of beats that were too short to be physiologically possible, and rechecking for missed beats with an adaptive threshold where the estimated intervals were too long.

The final example of algorithm discussed in this section is the approach described in Behar *et al* (2013), which obtained top scores in all of the events in the Challenge. The pre-processing

stage of this algorithm consisted of high-pass and notch filtering. The authors note that selection of a high cut-off frequency for the high-pass filtering, such as 10 Hz, led to an improved result in detection, due to the removal of the large maternal P and T wave components. The maternal ECG was estimated by first applying QRS detectors (based modified version of Pan and Tompkins algorithm), followed by the fusion of several different techniques of source separation (including principal component analysis, template subtraction, and ICA). Detection of the FQRS waveforms was then performed by a modified of the Pan and Tompkins algorithm on all the channels. A best FQRS channel was then selected based on the number of occurrences where the instantaneous heart rate variability was greater than 30 bpm. The final stage consisted of post-processing the fetal HR and RR series by smoothing the time series. For cases where the FQRS was considered undetectable, a constant time series at 143 bpm or at its estimated dominant mode, was generated. The authors also used a different scoring function from that of the Challenge for optimising their algorithm. The authors choosed the F_1 statistic (harmonic mean) to evenly balance the performance in terms of positive predictive value and sensitivity of the detectors.

4. Review of articles in the special issue

A total of thirteen articles were reviewed and revised in time to be accepted for this special issue. All authors had originally entered the Challenge, and most submitted updated versions of their algorithms, which should be made available by the authors through their open source licenses. The articles in this issue fall into three general categories based upon their signal processing approaches; temporal, spatial, and frequency (or time-frequency) approaches. We have therefore attempted to group the articles together and present them in this order. However, several articles combine multiple of these approaches to improve the heart rate extraction and do not neatly fall into a single category.

This special issue begins (after this editorial) with Behar *et al's* article describing the NI-FECG simulator that was developed for the Challenge (Behar *et al* 2014b). By constructing a realistic mixing model, with non-stationary effects from breathing and other motion, the training data for the Challenge was enriched with examples that had completely known source signals. By making this code available to the public, it is possible to stress test fetal analysis algorithms in unusual and pathological conditions (such as maternal heart rate dipping below the fetal heart rate) which despite seldom, could have adverse clinical consequences if missed.

Andreotti *et al* (2014) won events 2 and 5. The authors used kernel density estimation for fusing detection algorithms on the different channels for MQRS detection. The use of differential channels to augment the set of the four abdominal channels was also studied. Template adaption and an extended Kalman smoother for removing the maternal contribution were employed. An evolutionary algorithm was used in order to correct for FQRS detection, where weights were chosen between signal periodicity and signal morphology. ICA was only used for MQRS detection but all the processing for extracting of the FECG and detecting the FQRS was performed in the time domain using temporal methods on the available abdominal channels and possible differentials. The authors also used a 470 min private dataset recorded from 10 pregnant women to further evaluate their extraction algorithms. The authors reported that the TS approach performed better than the Kalman filter approach on the Challenge dataset but lower on their additional private dataset.

Behar *et al* (2014c), unofficial entry scored first for event 1 and 4, second for event 3 and third and second for event 2 and 5. Although unofficial, because they also helped create

the competition, the authors actually were blinded to the validation data, and so were at no advantage, apart from having spent more time analysing FECG than most in the competition. Their article presents a comprehensive review of classical methods used in the field for this application (template subtraction, blind source separation and Kalman filter approaches). The key contribution (apart from providing benchmarking algorithms), was the detail in how to train and combine these algorithms in order to achieve better performance.

Haghpanahi and Borkholder (2014) used the deflation approach from Sameni (2008) (iterative subspace decomposition and Kalman filtering) in order to remove the MECG. They also used PCA on the four abdominal channels and selected the best FQRS time series out of the two approaches (deflation/PCA). The authors used kurtosis as a proxy for signal quality and in order to rank the residual signals from the deflation methods and combined a subset of these to infer the FQRS time series.

Varanini *et al* (2014) removed the MECG from the abdominal signal using a PCA-based template subtraction algorithm and then applied ICA on the residuals. They, then selected one of the residuals based on: knowledge of typical FHR, mean of absolute RR first derivative and mean of absolute RR second derivative and the number of detected FQRS. This was a very similar approach to one of the techniques studied in Behar *et al* (2014c), where the authors concluded that using TS_{pca} was better than all other template subtraction techniques and subsequently applying ICA improved the result.

Dessi *et al* (2014) used a template subtraction approach followed by an ICA step, FQRS detection and correction and channel selection. For performing the template subtraction step, the authors noted that performing the operation at a high sampling frequency (they upsampled the data to 8 kHz) was important to enable alignment of each MECG cycle with the template and thus achieving superior cancellation. In order to build the template MECG cycle the authors selected beats based on correlation thresholding to avoid including abnormal beats in template.

Lipponen and Tarvainen (2014) used a PCA-based template subtraction approach in order to remove the MECG. They built the design matrix for the P and QRS and T waves separately and then applied PCA to identify the principal components. The most significant eigenvectors were fitted back to individual wave epochs from the MECG in order to remove them. The approach for suppressing the MECG is similar to Varanini *et al* (2014) and Behar *et al* (2014c) although Lipponen and Tarvainen (2014) separated the MECG cycles into P, QRS and T-waves.

Di Maria *et al* (2014) took a very standard PCA and template subtraction approach. the main focus of the paper was to explore picking the best principal component in order to identify the best MECG channel and the best FECG channel after performing MECG cancellation.

Liu *et al* (2014) performed prefiltering, then MQRS detection, then template subtraction and finally FQRS detection on the residual. It is important to note that they used a quality index (sample entropy) in order to exclude bad quality channels, which is theoretically better than performing FQRS detection on each channel and making the decision based on the regularity of the RR interval (as most entrants did). The authors also showed that by adjusting the MECG template to each cycle (in contrast to performing the simple construction with the template centered on the MQRS location) a performance improvement can be found. This second point was also illustrated in Behar *et al* (2014c), although Liu *et al* (2014) provided an interesting quantification of this phenomenon.

Lukoševičius and Marozas (2014) focused on the application of a QRS detector using an Echo state neural network (ESN), a data-driven statistical machine learning approach. The ESN is trained with the four residual signals (obtained using the MECG cancellation method from Martens *et al* (2007)) as channels of the input stream, and a probability of

QRS detection as the output. It should be noted that the authors did not focus on the extraction algorithms but on the QRS detector using multiple channels.

Rodrigues (2014) employed a Wiener filter which took as the input, the three abdominal channels with a number of coefficients (91) in order to filter out the MQRS from the fourth channel. The authors also used the MIT Abdominal and Direct Fetal Electrocardiogram Database in order to train their algorithm, which may have led to a bias in the results as this database was included in set-a, set-b (and possibly a few records in set-c).

Christov *et al* (2014) described a template subtraction method, with the template length being heart rate dependant, followed by an enhancement method that combined the four abdominal channels. The combined lead was obtained using i) PCA, ii) RMS or iii) Hotelling T-squared. The final combined lead was obtained by taking a mean over these three methods and although FQRS detection was performed on this combined lead

The final article in this collection is by Almeida *et al* (2014), who take a wavelet approach to denoising and extracting the fetal ECG. Although a time-frequency analysis seems very promising, the large cross-over in the spectral domain between the maternal and fetal signals and the noise, means this approach appeared to be limited. However, the authors note that their method is highly dependent on the pre-processing methods employed. This notable remark is true for every method to a greater or lesser degree.

5. Summary and future directions

In summary, the Physionet/Computing in Cardiology Challenge 2013 provided several key additions to the field of non-invasive fetal monitoring. First, a modest (but significant) annotated public database of NI-FECG was created, with a hidden validation set to allow objective future evaluation of algorithms. Second, a range of approaches have been compared, and open source code posted to allow scientific repeatability on the open access database. The existence of multiple independent algorithms allows us to explore the strengths and weaknesses of each approach, and exploit the combination of them to produce robust and accurate 'committees of experts' (e.g. see Behar *et al* (2014c)).

However, several limitations remain. A larger database is needed with more patients, longer recordings, more leads (including maternal ECGs) and abnormalities (such as arrhythmias, inter-uterine growth restriction, fetal acidosis, etc). Moreover, an annotated set of data which includes labels for ST segments and QT intervals under varying normal and abnormal conditions is required. It is hoped that we can produce such a database in the near future to provide the entrants with the opportunity to identify if their algorithms are able to extract such features with no clinically significant distortion. Of course, in order to do so, it will be important to define and identify meaningful analogs of adult measures of abnormality (such as long QT and ST deviations) in the fetal population, adjusted for gestational age. This will become even more important as we attempt to apply NI-FECG extraction algorithms to earlier and earlier stages of pregnancy. The updated open source codes described in this special issue are available on PhysioNet at <http://physionet.org/challenge/2013/sources/>.

Acknowledgments

JB is supported by the UK Engineering and Physical Sciences Research Council, the Balliol French Anderson Scholarship Fund and MindChild Medical Inc. North Andover, MA. This work was funded in part by NIH/NIGMS grant R01GM104987.

References

- Almeida R, Goncalves H, Bernardes J and Rocha A 2014 Fetal QRS detection and heart rate estimation: a wavelet-based approach *Physiol Meas.* **35** 1723–35
- Andreotti F, Riedl M, Himmelsbach T, Wedekind D, Wessel N, Stepan H, Schmieder C, Jank A, Malberg H and Zaunseder S 2014 Robust fetal ECG extraction and heart rate detection from abdominal leads *Physiol Meas.* **35** 1551–67
- Andreotti F, Riedl M, Himmelsbach T, Wedekind D, Zaunseder S, Wessel N and Malberg H 2013 Maternal signal estimation by Kalman filtering and template adaptation for fetal heart rate extraction *Computing in Cardiology (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Barnett S B and Maulik D 2001 Guidelines and recommendations for safe use of Doppler ultrasound in perinatal applications *J. Matern. Fetal Neo. Med.* **10** 75–84
- Behar J, Andreotti F, Zaunseder S, Li Q, Oster J and Clifford G D 2014b An ECG model for simulating maternal-foetal activity mixtures on abdominal ECG recordings *Physiol Meas.* **35** 1537–50
- Behar J, Oster J and Clifford G D 2013 Non invasive FECG extraction from a set of abdominal channels *Computing in Cardiology (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Behar J, Oster J and Clifford G D 2014c Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data *Physiol Meas.* **35** 1569–89
- Behar J *et al* 2014a Evaluation of the fetal QT interval using non-invasive foetal ECG technology SMFM- 34th Annual Meeting: the Pregnancy Meeting (New Orleans, LA, 8 February 2014)
- Boyce P M and Todd A L 1992 Increased risk of postnatal depression after emergency *Med. J. Aust.* **157** 172–4
- Chandrarahan E, Lowe V, Ugwumadu A and Arulkumaran S 2013 Impact of fetal ECG (STAN) and competency based training on intrapartum interventions and perinatal outcomes at a teaching hospital in London: 5 year analysis **120** 428–9
- Christov I, Simova I and Abaecherli R 2014 Extraction of the fetal ECG in noninvasive recordings by signal decompositions *Physiol Meas.* **35** 1713–21
- Clifford G D, Sameni R, Ward J, Robinson J and Wolfberg A J 2011 Clinically accurate fetal ECG parameters acquired from maternal abdominal sensors *Am. J. Obstet. Gynecol.* **205** 47e1–5
- Cremer M 1906 Über die Direkte Ableitung der Aktionströme des Menschlichen Herzens vom Oesophagus und Über das Elektrokardiogramm des Fetus *Münchener Medizinische Wochenschrift* **53** 811–3
- Dessì A, Danilo P and Luigi R 2013 Identification of fetal QRS complexes in low density non-invasive biopotential recordings *Computing in Cardiology (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Dessì A, Pani D and Raffo L 2014 An advanced algorithm for fetal heart rate estimation from non-invasive low electrode density recordings *Physiol Meas.* **35** 1621–36
- Di Maria C, Liu C, Zheng D, Murray A and Langley P 2014 Extracting fetal heart beats from maternal abdominal recordings: selection of the optimal principal components *Physiol Meas.* **35** 1649–64
- Goldberger A L, Amaral L A N, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C-K and Stanley H E 2000 PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals *Circulation* **101** e215–20
- Haghpasahi M and Borkholder D 2014 Fetal QRS extraction from abdominal recordings via model-based signal processing and intelligent signal merging *Physiol Meas.* **35** 1591–605
- Hayes-Gill B, Hassan S, Mirza F G, Ommami S, Himsworth J, Solomon M, Brown R, Schiffrin B S and Cohen W R 2012 Accuracy and reliability of uterine contraction identification using abdominal surface electrodes *Clin. Med. Insights: Women's Health* **5** 65–75
- Karlström A, Engström-Olofsson R, Norbergh K-G, Sjöling M and Hildingsson I 2007 Postoperative pain after cesarean birth affects breastfeeding and infant care *JOGNN* **36** 430–40
- Lipponen J A and Tarvainen M P 2013 Advanced maternal ECG removal and noise reduction for application of fetal QRS detection *Computing in Cardiology (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Lipponen J and Tarvainen M 2014 Principal component model for maternal ECG extraction in fetal QRS detection *Physiol Meas.* **35** 1637–48
- Liu C, Li P, Maria C D, Zhao L, Zhang H and Chen Z 2014 A multi-step method with signal quality assessment and fine-tuning procedure to locate maternal and fetal QRS complexes from abdominal ECG recordings *Physiol Meas.* **35** 1665–83

- Lukoševičius M and Marozas V 2014 Noninvasive fetal QRS detection using echo state network and dynamic programming *Physiol Meas.* **35** 1685–97
- 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** 373–8
- Matonia A, Jezewski J, Kupka T, Horoba K, Wrobel J and Gacek A 2006 The influence of coincidence of fetal and maternal QRS complexes on fetal heart rate reliability *Med. Biol. Eng. Comput.* **44** 393–403
- Neilson J P *et al* 2006 Fetal electrocardiogram (ECG) for fetal monitoring during labour *Cochrane Database Syst. Rev.* **3** CD000116
- Oudijk M A, Kwee A, Visser G H, Blad S, Meijboom E J and Rosén G K 2004 The effects of intrapartum hypoxia on the fetal QT interval *BJOG* **111** 656–60
- Pan J and Tompkins W J 1985 A real-time qrs detection algorithm *IEEE Trans. Biomed. Eng.* **32** 230–6
- 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. Perinat. Med.* **29** 408–16
- Podziemski P and Gieratowski J 2013 Fetal heart rate discovery (FHRD) algorithm for detection of fetal heart rate from noisy, noninvasive fetal ECG recordings *Computing in Cardiology (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Rodrigues R 2014 Fetal beat detection in abdominal ECG recordings: global and time adaptive approaches *Physiol Meas.* **35** 1699–711
- 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 and Clifford G D 2010 A review of fetal ECG signal processing; issues and promising directions *Open Pacing Electrophys. Ther. J.* **3** 4–20
- Sameni R, Clifford G D, Jutten C and Shamsollahi M B 2007 Multi-channel ECG and noise modeling: application to maternal and fetal ECG signals *EURASIP J. Adv. Signal Process.* **2007** 043407
- Sameni R, Shamsollahi M and Jutten C 2005 Filtering electrocardiogram signals using the extended kalman filter *27th Ann. Int. Conf. of the Engineering in Medicine and Biology Society, (Shanghai, China, 1 September 2005)* (New York: IEEE) pp 5639–42
- Silva I, Behar J, Sameni R, Zhu T, Oster J, Clifford G D and Moody G B 2013 Noninvasive fetal ECG: the PhysioNet/Computing in Cardiology Challenge 2013 *Computing in Cardiology (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Stinstra J G 2001 The reliability of the fetal magnetocardiogram *PhD Thesis* Universiteit Twente (<http://doc.utwente.nl/35964/1/t000001d.pdf>)
- Van Geijn H P *et al* 1991 Diagnosis and management of intrapartum fetal distress *Eur. J. Obstet. Gynecol. Reprod. Biol.* **42** S63
- 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 (Zaragoza, Spain, 22–25 September 2013)* vol 40
- Varanini M, Tartarisco G, Billeci L, Macerata A, Pioggia G and Balocchi R 2014 An efficient unsupervised fetal QRS complex detection from abdominal maternal ECG *Physiol Meas.* **35** 1607–19
- Zhu T, Johnson A, Behar J and Clifford G D 2013 Bayesian voting of multiple annotators for improved qt interval estimation *Comput. Cardiol.* **40** 659–62
- Zhu T, Johnson A E, Behar J and Clifford G D 2014 Crowd-sourced annotation of ecg signals using contextual information *Ann. Biomed. Eng.* **42** 871–4