A practical guide to non-invasive foetal electrocardiogram extraction and analysis

Joachim Behar[1], Fernando Andreotti[2], Sebastian Zaunseder[2], Julien Oster[3] and Gari D Clifford[4]

[1] Biomedical Engineering Faculty, Technion-IIT, Haifa, Israel
[2] Institute of Biomedical Engineering, Faculty of Electrical and Computer Engineering, TU Dresden, Helmholtzstr. 10, 01069 Dresden, Germany
[3] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford OX1 3PJ, UK
[4] Departments of Biomedical Informatics & Biomedical Engineering, Emory University & Georgia Institute of Technology, Atlanta, GA, USA

E-mail: jbehar@tx.technion.ac.il

Abstract

Non-Invasive foetal electrocardiography (NI-FECG) represents an alternative foetal monitoring technique to traditional Doppler ultrasound approaches, that is non-invasive and has the potential to provide additional clinical information. However, despite the significant advances in the field of adult ECG signal processing over the past decades, the analysis of NI-FECG remains challenging and largely unexplored. This is mainly due to the relatively low signal-to-noise ratio of the FECG compared to the maternal ECG, which overlaps in both time and frequency.

This article is intended to be used by researchers as a practical guide to NI-FECG signal processing, in the context of the above issues. It reviews recent advances in NI-FECG research including: publicly available databases, NI-FECG extraction techniques for foetal heart rate evaluation and morphological analysis, NI-FECG simulators and the methodology and statistics for assessing the performance of the extraction algorithms. Reference to the most recent work is given, recent findings are highlighted in the form of intermediate summaries, references to open source code and publicly available databases are provided and promising directions for future research are motivated. In particular we emphasise the need and specifications for building a new open reference database of NI-FECG signals, and the need for new algorithms to be benchmarked on the same database, employing the same evaluation statistics. Finally we motivate the need for research in

NI-FECG to address morphological analysis, since this represent one of the most promising avenues for this foetal monitoring modality.

Keywords: noninvasive foetal ECG, foetal monitoring, source separation

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

1. Introduction

Worldwide, an estimated 2.65 million stillbirths occur yearly, of which 98% occur in countries of low and middle income and with more than 45% in the intrapartum period (Bhutta *et al* 2011) (i.e. occurring during childbirth). Thus, there is a need for effective monitoring techniques that can provide information on the foetal health during the pregnancy and at delivery. In addition, heart defects are the leading cause of birth defect-related death (California Department of Public Health 2014). As an example it is estimated that 9 in 1000 babies born in the UK have congenital related heart disease at birth (NHS 2014). Nowadays, more and more cases of congenital heart disease are diagnosed prepartum during routine imaging via an ultrasound scan. However, there are some cardiac defects that cannot be identified with this modality and the use of this technology requires extensive training and relatively high costs. There is therefore an interest in antepartum technology that could provide additional information about the cardiac health of the foetus at a relatively lower cost, in a more automated manner.

Intermittent observations of foetal heart sounds through auscultation became a standard clinical practice by the middle of the 20th century. Foetal heart rate (FHR) monitors became widely available by mid-1970s (Kennedy 1998) and enabled continuous FHR monitoring, mainly via one dimensional ultrasound or 'cardiotocography' (CTG)[5]. CTG monitors were expected to drastically reduce foetal hypoxia. However, rapidly conflicting studies were published and challenged the efficacy and reliability of the FHR monitors. The introduction of these monitors in clinical practice increased the rate of a painful and expensive cesarean section, higher prevalence of postnatal depression (Philip and Angela 1992), and postoperative pain negatively affecting breastfeeding and infant care (Karlström *et al* 2007) while its benefits on foetal health were not clearly demonstrated. Although intrapartum abnormalities detected through electronic foetal monitoring correlate with umbilical cord base excess and its use is associated with decreased neonatal seizures, it has no effect on perinatal mortality or paediatric neurological morbidity (Graham *et al* 2006).

The first documented attempt to record a non-invasive foetal electrocardiogram (NI-FECG) from the maternal abdomen was by Cremer (1906). It has been shown that during the first stage of labour, FHR tracings obtained by Doppler CTG and NI-FECG measurements correlate well with each other (Reinhard *et al* 2010). Similar results were also obtained for acceleration and deceleration counts as well as other characteristic CTG patterns. However, because the NI-FECG does not need to average over multiple beats, it has the potential to provide a more accurate estimation of the FHR. Moreover, due to the sensor modality, NI-FECG can also identify key morphological information related to the electrical activity of the foetal heart. However, the FECG is difficult to extract from the recorded abdominal signal mixture (see figure 1) and this has restricted its use to-date. Throughout the last few decades, NI-FECG analysis has been the focus of many research studies, mainly due to improvements in quality

[5] Technically, CTG includes recording (or displaying) the foetal heartbeat ('cardio') as well as the uterine contractions ('toco'). The latter are not directly addressed in this review, except in its relation to foetal heart rate changes.

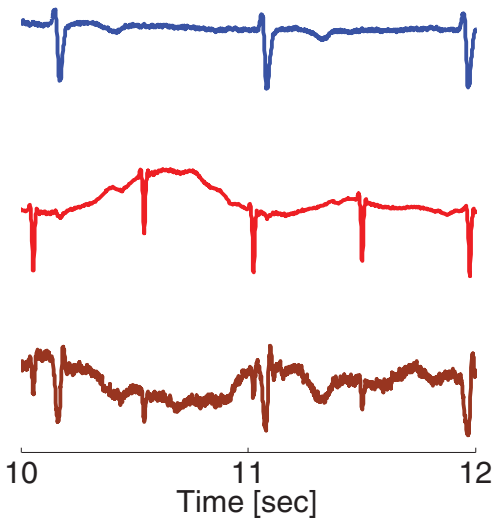

Figure 1. From top to bottom: chest ECG (largely dominated the maternal ECG), SECG i.e. ECG of the foetus recorded by screwing a small electrode on the foetal head (or sometimes foot) during delivery, and abdominal ECG (mixture of maternal and foetal ECG). The purpose of NI-FECG analysis is to separate between the maternal and foetal components of the signal mixture recorded on the maternal abdomen. Reproduced from Clifford *et al* (2014d).

of sensors, speed of processors, advances in signal processing and the decrease in price of recording devices (Sameni and Clifford 2010).

2. Electronic foetal monitoring

Electronic foetal monitoring (EFM) techniques can be invasive or non-invasive with intermittent or continuous assessment. Current techniques include foetal phonocardiography (PCG), Doppler ultrasound, cardiotocography (CTG), foetal magnetocardiography (FMCG) and foetal electrocardiography. Although electronic foetal PCG has been reported in the literature, it is highly prone to noise and can only be used late in pregnancy, and therefore is almost unused in clinical practice (Kovács *et al* 2011). From 20 weeks onwards the heart can be monitored using Doppler ultrasound, the FECG, and the FMCG (Peters *et al* 2001). Around the 28th–32nd week of gestation, a thin layer called the vernix caseosa is formed and dissolves in the 37th–38th week in normal pregnancies (Stinstra 2001). This layer is highly non-conductive and results in a limited NI-FECG amplitude on the maternal abdomen (Oldenburg and Macklin 1977, Oostendorp *et al* 1989). A number of studies (Pieri *et al* 2001, Graatsma *et al* 2009) have observed the adverse effect of this layer on the performance of the NI-FECG extraction but the limitations have been inconsistently reported.

Doppler ultrasound is routinely used for FHR monitoring during pregnancy and delivery. However, despite being a non-invasive modality, the ultrasound is not passive and it has not been conclusively demonstrated that ultrasound radio frequency exposure is completely safe for the foetus (Barnett and Maulik 2001). In addition, while using Doppler ultrasound, the maternal HR (MHR) can be picked up instead of the FHR because of the misorientation of the ultrasound transducer (which can therefore be transducing the maternal arteries). Neilson *et al* (2008) reported five

Table 1. Continuous foetal monitoring modalities (Clifford *et al* 2014d, Peters *et al* 2001). SNR: signal to noise ratio.

Method	System	Gestational age	Comments
PCG	Phonocardiography: acoustic recording from abdomen	⩾28–30 weeks	- Manual auscultation employs a Pinard stethoscope but electronic stethoscopes with amplification are sometimes used - Lowest SNR of all methods employed - Requires expert to locate foetus - Prone to endogenous and exogenous acoustic noise sources (e.g. gastrointensinal activity and speech) - Not routinely employed for foetal monitoring or foetal health assessment
CTG	Cardiotography; ultrasound transducer and uterine contraction pressure-sensitive transducer	⩾20 weeks	- Contraction monitoring with a pressure transducer - Smoothed HR time series - No beat-to-beat data and cardiac function descriptor limited to HR - Not passive; ultrasound irradiation - Prone to maternal/foetal HR confusion - Prone to dropout due to foetal or maternal movement
FMCG	Foetal magnetocardiogram. Detection of the foetal heart's magnetic field through SQUID sensors positioned near the maternal abdomen	⩾20 weeks	- Expensive - Requires skilled personnel - Morphological analysis of the FMCG easier than FECG because of higher SNR - No long term monitoring possible to date because of apparatus size and cost
NI-FECG	Non-invasive foetal ECG. Standard ECG electrodes with varying skin preparation methods	⩾20 weeks	- Low cost - Does not required skilled user - Continuous monitoring possible - FHR and possibly morphological analysis - Low SNR, with potential issues from 28th to 37th week from vernix
SECG	Scalp ECG. A single scalp electrode is placed on the foetal scalp	At delivery	- Accurate FHR - Morphological analysis possible - High SNR - Invasive, carrying small additional risk, not routinely recommended - Can only be used at delivery - Single channel only

cases out of 10 000 monitored pregnancies that were fatal for the foetus where the MHR rather than the FHR was recorded by the CTG. FMCG, while providing higher quality FHR is too cumbersome and expensive to be widely used clinically. The scalp ECG (SECG), while providing accurate FHR time series, is an invasive technique and monitoring can only be performed at delivery. Table 1 lists the EFM techniques and presents their advantages and drawbacks.

Since the introduction of the EFM in the mid 1970s and the uncertainty around the benefit of this method, the medical community has been very cautious about the adoption of any alternative method or adjunct to the CTG (despite the lack of evidence for the utility of the CTG itself). In practice, this has resulted in large clinical trials targeted at assessing the benefits of novel monitoring methods. For example, foetal pulse oximetry (FPO) was recently introduced (with the first FDA approved device in 2000 (Bloom *et al* 2006)) and clinical studies were run to assess its clinical benefit. Studies on FPO showed that using this monitoring method in addition to the CTG was not associated with a significant difference in neonatal outcomes (Bloom *et al* 2006, East *et al* 2006) and conclusions on whether it reduced caesarean frequency could not be drawn.

In clinical practice there are two main widespread methods currently in use for continuous foetal monitoring: the CTG, which is the most widespread, and the SECG. However, neither method is ideal; the CTG only provides an estimate of the FHR while the SECG is invasive. Due to those limitations, the NI-FECG has been suggested as an alternative monitoring method that can combine the advantages of both the CTG and SECG, i.e. be non-invasive while providing an accurate FHR, as well as the additional information on the electrical activity of the heart, contained in the ECG morphology and inter-beat timing. For this reason, the NI-FECG is the subject of considerable research.

3. A meta-review

Since the early 2000s, several reviews have been published on the topic of electronic foetal monitoring. A detailed book on foetal electrocardiography by Symonds *et al* (2001) provides an in depth history of the development of FECG extraction and analysis, including invasive SECG and NI-FECG. In particular, the book includes a substantial chapter that reviews the publications on morphological analysis of the FECG and on the interpretation of the FHR. However, the book is focused on the clinical interpretation of the NI-FECG and not on the NI-FECG signal processing techniques. Moreover, there has been a number of novel contributions to the field since its publication over 14 years ago. Hasan *et al* (2009) published a rather general overview on NI-FECG, reviewing the physiological background, hardware, extraction algorithms and description of the different sources of noise. While lacking detail, it provides a good entry point to the field of NI-FECG. Almost concurrently, Sameni and Clifford (2010) published an extensive review on NI-FECG covering the physiological background of the foetal heart, modelling of NI-FECG using the dipole model, existing publicly available databases and a detailed overview of FECG extraction algorithms with relative merits of each. Although a number of signal processing algorithms were qualitatively reviewed in these two papers, the absence of a 'large' public database and of a defined evaluation methodology made it impossible to objectively compare the relative performance of extraction algorithms proposed to-date. To address these issues the authors of the latter article and colleagues initiated an international competition: 'the PhysioNet/Computing in Cardiology Challenge 2013' (hereafter referred as the 'Challenge') Silva *et al* (2013) and a follow-up special journal issue (Clifford *et al* 2014d). This editorial provided an extensive review of the Challenge and the publications from the 29 teams that competed. The editorial review highlighted the strengths and weaknesses of the 17 open source algorithms that competed for the Challenge. The main strength of the editorial review is that it allowed for the objective ranking of the algorithms' performances since they were evaluated on the same database and using the same

methodology. A public training set was provided, and a hidden test set was used to score submitted algorithms in an independent manner, using the same statistics. The editorial review, however, did not discuss the strengths and weaknesses of the evaluation statistics in detail. Nor did it address the topic of morphological analysis of the NI-FECG. Most recently, Agostinelli *et al* (2015) provided a good general overview of the field including the challenge of electrode placement on the maternal abdomen and potential usefulness of some morphological measures. However the topics addressed in this review were not treated in depth, with a relatively trivial notion of treating the issue as a linear additive noise and signal problem, ignoring movement, nonstationarities, and the complex near-field effects observed in FECG extraction. Several inaccuracies are present in their review such as the description of electrode locations, and that the use of blind source separation techniques requires too many electrodes to be practical (existing commercial monitors contradict this conjecture). Moreover, the review failed to reference the most recent publications in the field (including the wealth of articles generated from the Challenge), possibly because their methodology relied entirely on a PubMed search, although this does not explain all of the major omissions. Perhaps most importantly, the review did not provide any detail on the relative approaches and merits of each FECG extraction technique, nor any numerical results of experiments performed with these techniques. Consequently, no discussion of where the field should move is given.

The current review presented here addresses the recent advances made in digital signal processing for extracting and exploiting the NI-FECG information. In particular we aim at current and prospective research for creating NI-FECG algorithms targeted at morphological analysis and the use of standard statistics for scoring the extraction algorithms (and thus being able to benchmark between algorithms using the same statistics). Reference to the most recent work is given, important findings are highlighted in short 'intermediate conclusions' summarising important design consideration for the practitioner. Finally, directions for future research are motivated. The physiological background of NI-FECG is not reviewed in this article and the reader is referred to Sameni and Clifford (2010) and Symonds *et al* (2001) for a good overview on the topic.

4. Research foci and challenges

The research in NI-FECG is motivated by the possibility of acquiring the following physiological information through this modality: (1) rhythm information using the extracted FHR. The NI-FECG can provide a beat to beat estimation of the FHR, which is superior to what is provided by Doppler ultrasound and enable antenatal monitoring in contrast to SECG, which can only be used at delivery; (2) FECG morphological information, such as PR, QT intervals and ST level. Studies of these parameters have proven to be clinically relevant in adults and animal models, and thus have the potential to add information to the monitoring of the foetal heart; (3) Contraction monitoring by sensing the electrical impulses generated by the myometrial activity (Hayes-Gill *et al* 2012). Knowledge of contraction occurrence is believed to aid the interpretation of rapid changes in the FHR during delivery (although there is little evidence for this, possibly because the lack of precise definitions around contractions); (4) foetal respiration, foetal movement (as suggested in Sameni (2008) and Vullings *et al* (2013)) and foetal orientation (Taylor *et al* 2003).

The challenges for extracting the information listed above from the NI-FECG are: (1) Relatively low signal-to-noise ratio of the FECG compared to the maternal ECG (MECG, see illustration in figure 1). This is due to the size of the foetal source which is much smaller than its maternal counterpart and the different dielectric media in between the foetal source and the sensors; (2) foetal non-stationarities, such as foetal movement, which are frequent at the earlier stages of pregnancy or foetal respiratory effort (see example in figure 2); (3) need for algorithms that either do not require the intervention, or require minimal intervention from a

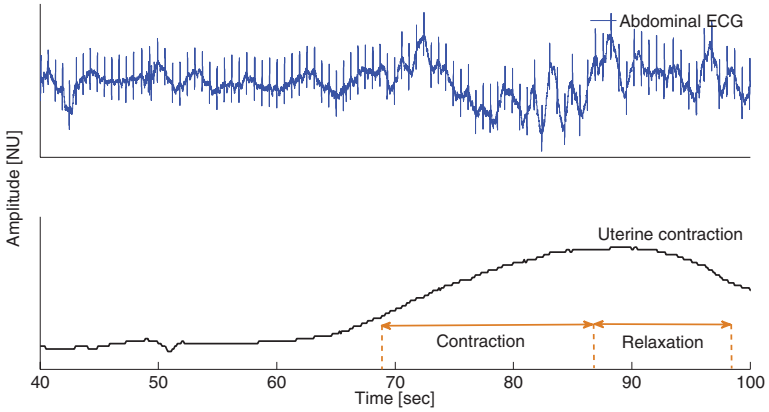

Figure 2. Example of a uterine contraction and its effect on the quality of the abdominal ECG signal. Note the degradation in the abdominal ECG quality after 70 s, which corresponds to the onset of the contraction recorded by an intrauterine pressure transducer.

medical professional; (4) limited feasibility in recording the NI-FECG after the formation of the vernix caseosa layer and until it dissolves (Oostendorp 1989); (5) statistics that provide a confidence measure on to the extraction performances and avoid medical errors such as confounding the FHR and the maternal HR (MHR); (6) algorithms that enable the reconstruction of the FECG waveform (as opposed to being restricted to foetal R-peak detection) from the NI-FECG mixture and enable morphological analysis of the FECG.

5. Commercial NI-FECG monitors

In recent years, a handful of NI-FECG commercial monitors have appeared in the market and more competing solutions are in development Particularly interesting are two monitors that have obtained FDA clearance: the Monica AN24 monitor (Monica Healthcare, Nottingham, UK) and the Meridian M100/M1000 monitors from MindChild Medical (North Andover, MA)[6]. Both monitors have proved to be accurate in detecting FHR (Sameni *et al* 2009, Reinhard *et al* 2012), and some seminal work on extracting morphological information has shown promise for the latter monitor (Clifford *et al* 2011, Behar 2014, Behar *et al* 2014e). However, the studies associated with these monitors were limited in patient numbers and population size and further research is needed to address a broad range of populations and conditions.

Commercial applications for NI-FECG monitoring are in their early days, and there is a growing interest in improving their performance in order to reach the point where they will provide actionable information to the clinician. As such, both because it is a relatively low cost non-invasive technique, and because of the information the NI-FECG has the potential to provide, there is a strong motivation for pushing forward research in this area.

6. Electrode placement

When attempting to collect NI-FECG data, careful thoughts about the electrode configuration are necessary. Similarly to adult electrocardiography, the morphology of the cardiac signal strongly

[6] Other monitors planned for release include Nemo Healthcare (De Lismortel, Netherlands) and PregSense (Nuvo, Israel).

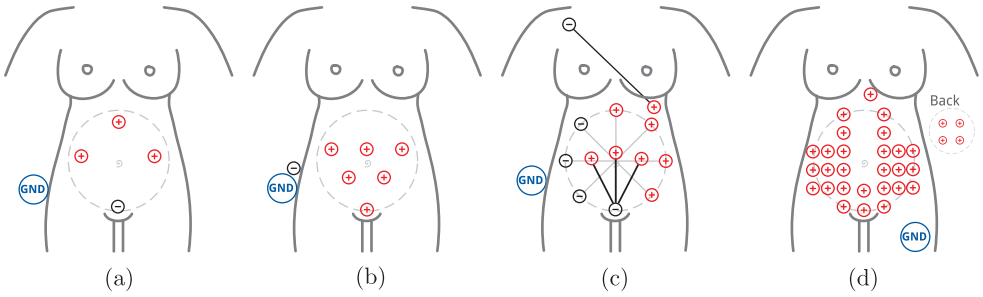

Figure 3. Different electrode configurations used in the literature by (a) Monica Healthcare Ltd AN24 (Nottingham, UK) (Pieri *et al* 2001, Reinhard *et al* 2010); (b) (Rabotti *et al* 2008, Rooijakkers *et al* 2014); (c) (Andreotti *et al* 2014); (d) MindChild Medical Inc. Meridian Monitor (North Andover, USA) (Clifford *et al* 2011, Wolfberg *et al* 2011). 'GND' represents the common ground (inactive) electrode, while ⊕ is the active and ⊖ the reference electrode. This distinction between active and reference electrodes serves to specify which differential leads were considered. (a) 5 electrodes GND. (b) 8 electrodes GND. (c) 14 electrodes Back GND. (d) 32 electrodes.

depends on the lead configuration employed. The foetal position cannot be easily identified prior to electrode placement because it changes across patients particularly at low gestational weeks. Therefore a general-purpose 'optimal' electrode placement is impracticable (Agostinelli *et al* 2015). Nonetheless, numerous configurations have been proposed in the literature ((Andreotti 2011, Agostinelli *et al* 2015)—see figure 3), in an attempt to standardise the recording procedure. Several authors rely on the most common foetal presentation (i.e. vertex, breech or shoulder), to reduce the number of leads used by aiming at usual positions for foetal head/thorax, consequently minimising the application's complexity. Meanwhile, other authors attempt to cover most of the abdomen to maximise the chances of obtaining FECG signal (Agostinelli *et al* 2015). Despite the variety of electrode configurations proposed, to date, very few studies have compared the different options that exist. Rooijakkers *et al* (2014) is one exception, in which the effect of increasing the distance between leads was analysed. The authors suggested two bipolar lead placement schemes, one with 5 electrodes for intrapartum recordings (inter-electrode distance of 16 cm—similar to figure 3(a) with an additional central electrode) and a 6 electrodes scheme for preterm recordings (distance 20 cm—similar to figure 3(b)). Such configurations should optimally record the MECG, FECG and uterine activity. However, this clinical study was limited to 5 patients, each recording 20 min in duration at high gestational weeks (>39 weeks). Another study, which considered signal quality throughout different channels was performed by Clifford *et al* (2012) and Clifford *et al* (2011). The authors described an 'over-complete' set of electrodes and noted that some channels, at particular time instants, contributed most with FECG signals. However, they also noted that the signal quality varies across patients (and time), thus an automatic selection of channels based on signal quality measures is necessary. Lastly, the optimal number of electrodes may depend on the extraction method used. For instance, extraction routines may required one or several abdominal leads, meanwhile other methods need one (or more) MECG reference leads. Further details about these signal processing requirements are discussed in section 8.3.1.

7. Public databases

7.1. Existing databases

To date there are five publicly available databases that can be used to evaluate NI-FECG extraction algorithms' performance. These are summarised below and in table 2:

Table 2. Summary table of the existing open databases.

Database	NR	NF	f_s (Hz)	RA	Duration
DDB	1	1	250	No	10 s
NIFECGDB	55	1	1000	No	Varying
ADFECGDB	5	5	1000	Yes	5 min
PCDB	447	>10[a]	1000	Yes	1 min
FECGSYNDB	1750	9[b]	1000	Yes	5 min

NR: number of records
NF: number of foetuses
f_s sampling frequency
RA: reference annotations available (yes/no).
[a] An unknown number of additional subjects were used in the hidden test dataset.
[b] Only nine different VCG are used in the *fecgsyn* simulator, but an unlimited number can be defined and generated.

- The Daisy database (DDB) (De Moor *et al* 1997) is not so much a database, as a single unrepresentative snippit of data. It consists of 8 ECG channels (5 abdominal and 3 thoracic) from a single foetus, lasting for 10 s, using a sampling frequency $f_s = 250$ Hz, without reference annotations. Although this recording is often used, it is unrepresentative as it is a particularly clean and short set of recordings.
- The non-invasive foetal electrocardiogram database (NIFECGDB), available on PhysioNet Goldberger *et al* (2000), $f_s = 1$ kHz, includes 55 multichannel abdominal ECG (AECG) recordings taken from a single subject (21–40 weeks of gestation) and recorded using a g.BSamp Biosignal Amplifier (GTech GMBH, Austria). No reference annotations are available. Each record has 2 thoracic and 3 or 4 abdominal signals.
- The abdominal and direct foetal electrocardiogram database (ADFECGDB), available on PhysioNet (Goldberger *et al* 2000) $f_s = 1$ kHz with 5 min of recordings (4 abdominal channels and the SECG) from 5 women in labour (38–41 weeks of gestation) and with reference FQRS annotation derived from the SECG. Data were recorded using the KOMPOREL foetal monitoring system (ITAM, Zabrze, Poland).
- The 2013 PhysioNet/Computing in Cardiology Challenge Database (PCDB) (Silva *et al* 2013) consists of 447 min of data from five different databases (including the NIFECGDB, ADFECG and simulated data using the *fecgsyn*). FQRS reference was provided to a subset of the CDB, the remaining reference was kept as hidden test set. Data were resampled at $f_s = 1$ kHz. This database represents the largest publicly available dataset to date. Each record has four abdominal channels and no maternal reference. The open training set has reference FQRS available while the closed validation and test sets had restricted access to the reference annotations to facilitate independent testing. In this respect it is a unique database, which prevents over tuning of parameters. Data and references for set-a (training set) are still available and the platform is still open for scoring annotations for the recordings of set-b (validation set.)
- The Foetal ECG Synthetic Database (FECGSYNDB) (Andreotti *et al* 2016) includes 1750 five-minute realistic simulations of abdominal mixtures using 34 channels (32 abdominal and 2 MECG channels), totalising 145.8 h of data and 1.1 million foetal peaks. The FECSYNDB comprised seven cases of physiological events, for 10 different maternal-foetal heart dipoles' arrangements, at five noise levels. Each combination was simulated five times for statistical purposes, hence combined into a total of $7 \times 10 \times 5 \times 5 = 1750$ simulations. The simulated data utilised the *fecgsyn* simulator by Behar *et al* (2014a).

Together with providing the first significant open NI-FECG database (the PCDB), the 2013 Physionet/Computing in Cardiology Challenge highlighted the need for a larger and more complete dataset. This database should ideally have a number of characteristics which are summarised within this section. The description below intends to define the specification for building such database so that any existing and novel NI-FECG extraction and analysis algorithm can be evaluated on it.

Part of the database should have a reference SECG and at least one chest MECG channel. The scalp recordings can be used as the *silver* standard for FHR and ECG morphological analysis. Chest recordings can be used to ensure accurate MQRS detection and be used as reference channel for some source extraction techniques that require a lead free of any FECG contribution. Moreover, it can be used to avoid confusing the FHR and MHR, a mistake which sometimes occurs during Doppler ultrasound recordings and can be life threatening for the foetus (Neilson *et al* 2008).

The database should also include at least eight abdominal channels (see section 6.5 of Sameni (2008)) and potentially more (Clifford *et al* 2011) to capture the multidimensional nature of both the foetal and maternal ECG. This would allow better performance of blind source separation (BSS) based methodologies and maximise the chances of obtaining foetal signals, since the optimal positioning of the electrodes with respect to the foetal heart position is *a priori* unknown (i.e. channel selection might be required). The sampling frequency and amplitude resolution of $f_s = 1$ kHz, 16-bits usually are sufficient. In particular, low frequency resolution can cause misalignment problems in removing the MECG when using template subtraction methods (see figure 3 in Behar *et al* (2014d)). A sampling frequency that is too low can also affect the accuracy of morphological measurements such as the QT interval (Baumert *et al* 2016).

The database should be 'large enough' both in term of the number of recordings (from independent pregnancies), recording's length and heterogeneity of conditions. In particular, the records should be long enough in order to assess the adaptability of the approaches with respect to the non-stationary nature of the FECG (intrinsic to the ECG signal but also due to foetal motion, resulting in a change of the foetal heart orientation). For example, figure 2 shows a case where an uterine contraction is affecting the quality of the abdominal ECG signal thus making the extraction of the NI-FECG more challenging. Mild contractions are generally 15–20 min apart and last between 60–90 s (Cleveland Clinic 2015) and thus a minimal length of 30 min recording at delivery time should be considered to ensure (statistically) capturing one such event per record. Monitoring the FHR during these events is crucial and none of the published paper to date assessed the capability of the algorithms to cope with this type of physiological events.

Recordings should be performed at different stages of pregnancy, in order to assess when and in what proportion NI-FECG extraction is feasible. If possible, data on foetal activity or 'sleep' should be captured. Meanwhile it would enable the study of heart rate variability (HRV) and ECG morphological analysis as a predictor of cardiovascular diseases in infants. When assessing the performance of the NI-FECG it is important to compare the success of the evaluated FHR against an alternative technique such as ultrasound or SECG, rather than manually annotating foetal R-peaks on the abdominal recordings. Indeed, manually annotating the foetal R-peaks (in order to use them as the reference annotations) implies that these peaks must be visible to the human eyes in the first place on at least one abdominal ECG channel. However, there are instances where the NI-FECG can also be extracted from abdominal recordings where no FECG can be identified by human inspection of the surface electrode

recordings (see figure 7 in Behar *et al* (2014d)). These records should not be discarded from the analysis, since this creates a bias of any system towards analysing easier data.

A range of clinical information such as outcome of delivery, maternal history (e.g. IUGR, foetal anaemia and premature rupture of the membrane) should be reported. Information on the delivery outcome can include presence of metabolic acidosis, neonatal encephalopathy, Apgar score, need for neonatal intubation This allows the assessment of whether or not the information extracted from the NI-FECG is predictive of delivery outcomes. It is also important to obtain rhythm annotations as well as information on early and late deceleration from experts since these events are typically the ones where the NI-FECG monitor should not fail to extract the FECG. Current parameters for foetal FHR analysis should be considered and annotated by medical experts. These parameters were standardised by the FIGO Guidelines (Rooth *et al* 1987), which include definitions for physiological and pathological accelerations, decelerations, tachychardias and bradychardias. Of note, there is still some controversy about the definition around some of these parameters (Bernardes *et al* 1997, Ayres-de Campos and Bernardes 2010).

7.3. Simulations

One of the main limitations of the current NI-FECG databases is the absence of pathological cases, pregnancy with adverse outcome or rare events such as similar FHR and MHR at times of abrupt heart rate changes. Another limitation of existent databases is that they do not enable the performance assessment of the separation algorithms as a function of the signal to noise ratio (SNR) between the ECG mixture and noise, the SNR between the maternal and foetal components. These considerations have motivated the development of an artificial NI-FECG model to create realistic abdominal recordings that model specific physiological phenomena. The original simulator was developed by Sameni *et al* (2007b) and considered a single dipole per cardiac source linearly related to the body surface potentials by a projection matrix that takes into account the temporal movements and rotations of the cardiac dipole. This model build upon the adult ECG model from McSharry *et al* (2003). The most recent update to the NI-FECG model called *fecgsyn* was contributed in Behar *et al* (2014a) and open sourced on Physionet[7] with an easy to use user interface (Alvi *et al* 2014). This recent update took into account the non-stationary mixing due to respiration and foetal movement and realistic heart rate changes (both normal and pathological). Figure 4 illustrates the volume conductor representation used in *fecgsyn*. Further works using the *fecgsyn* resulted in the development of the FECGSYNDB (Andreotti *et al* 2016), which enables the benchmarking of FQRS extraction and FQRS detection algorithms. The *fecgsyn* and FECGSYNDB are particularly useful in providing a first analysis of morphological extraction techniques, since the FECG morphology is known before being added to the abdominal signal (see example on figure 5). The simulator was also used in the recent work of Liu and Luan (2015). The FECGSYNDB should be understood as an exemplary usage of the simulator but the simulator itself provides the possibility for researchers to simulate specific events which can be adapted to suit one's experimental needs.

One of the main limitation of the model in its current state is that the vectocardiograms used to generate the FECG signals were acquired from adult ECG signals which will produce non-physiological foetal intervals. This is because the adult ECG waveform is basically linearly re-scaled to match the FHR but this shrinkage should not be linear in order to produce foetal intervals falling in their physiological range (see table 4). In addition, variation of intervals such as FQT (and other morphological parameters) as a function of the FHR is not

[7] Available at: www.physionet.org/physiotools/ipmcode/fecgsyn/

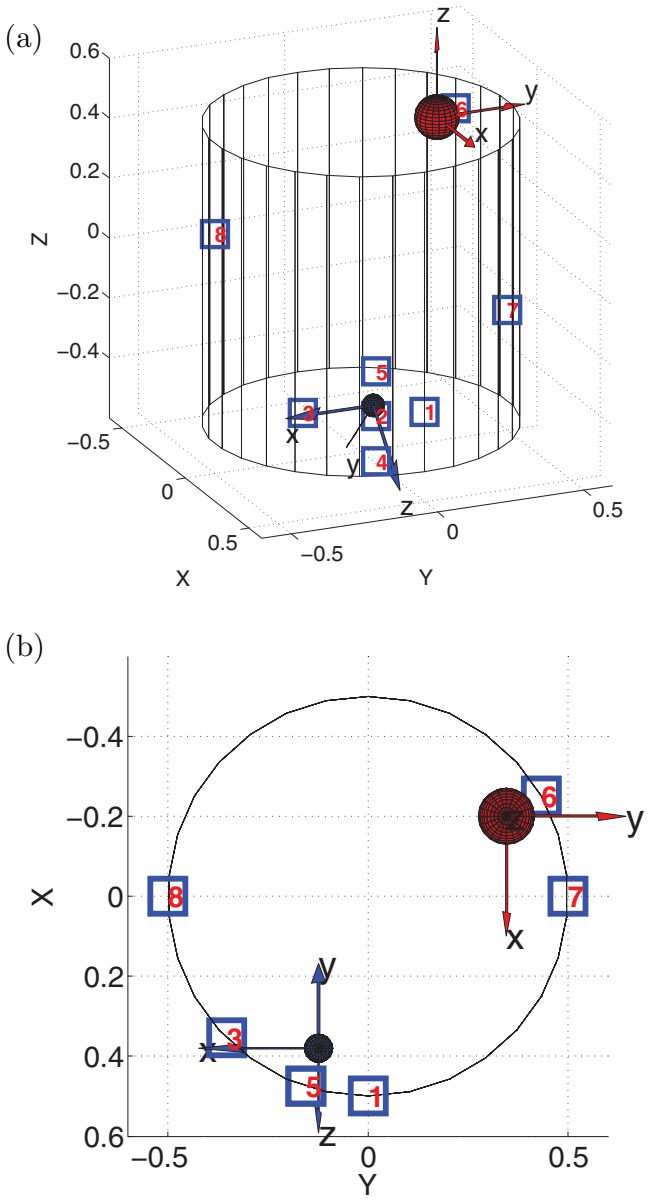

Figure 4. (a) Volume conductor representation. The two spheres represent maternal (large sphere) and the foetal (smaller sphere) hearts. The blue numbered squares show the location of the electrodes on the volume conductor, while the arrows represent the orientation of the foetal and maternal body axis which are defined by the heart orientation; (b) Top view of the volume conductor. Note that in (b) electrodes 2 and 4 are not represented because the top view is overlapping. Figure reproduced from Behar *et al* (2014a).

modelled (other than a linear re-scaling). For this last point a corrected FQT/FHR relationship (equivalent of QT_c) should be derived in the first place from real data measurements. As a final improvement, a more realistic volume conductor geometry could be considered.

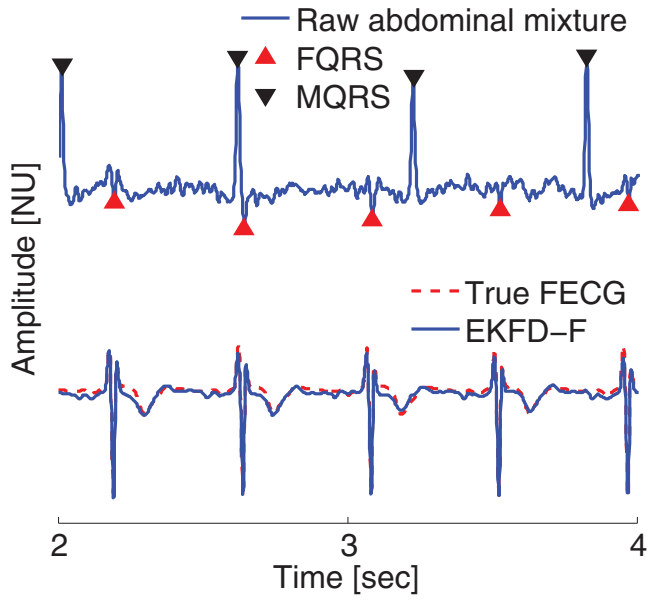

Figure 5. From top to bottom: abdominal mixture, extracted NI-FECG by the EKFD on top of the true FECG. Note that the FECG morphology is reconstructed well using the method by Behar *et al* (2013). Figure reproduced from Behar *et al* (2013). The *fecgsyn* simulator was used to generate the abdominal mixture signal.

7.4. Intermediate conclusions

- There are currently five publicly available real NI-FECG databases (three of which available on Physionet (Goldberger *et al* 2000)). We recommend to use the most recent and complete database of real signals i.e. the PCDB (Silva *et al* 2013). It is important to stress that new algorithms must be evaluated on set-a (training set) and set-b (validation set). Evaluation on the set-a only can solely be a 'proof of concept'. When using any of the NI-FECG databases, recordings with inaccurate reference annotations should be excluded.
- The NI-FECG is usually recorded at high sampling frequency (typically between 500 Hz and 1 kHz) (Taylor *et al* 2005, Widrow *et al* 1975) and AD converter resolution of 16 bit similar to what is traditionally used for adult clinical ECG.
- We motivate the critical need for creating of a novel database of NI-FECG signals with the following important characteristics: (1) rhythm annotations from experts; (2) a minimum of 30 min recording length for each record; (3) sampling frequency of 1 kHz and resolution of 16 bits; (4) recording of SECG, a minimum of one chest channel and a minimum of eight abdominal channels; (5) outcomes at delivery.
- Because there is currently no such 'complete' database, an artificial NI-FECG model such as *fecgsyn* (Behar *et al* 2014a) (see section 7.3), may be used until it is created, or in addition to such database as a way to test/produce of proof of concept for any extraction algorithm.

8. Estimation of the foetal heart rate

The range of normal foetal heart rate varies with gestational age. The heart starts beating at the end of the first month of the pregnancy, when a single hollow ventricle is created (and a

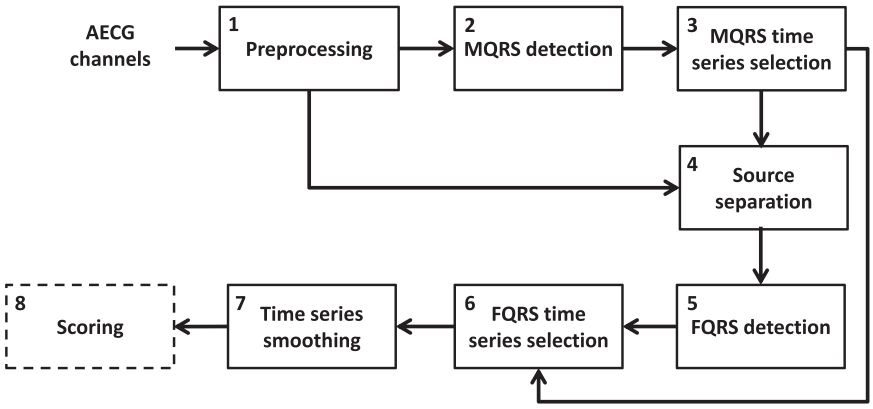

Figure 6. NI-FECG and FHR extraction block diagram. (1) The abdominal ECG (AECG) channels are preprocessed (see section 8.1); (2) MQRS detection is performed (see section 8.2); (3) at least one of the MQRS time-series is selected (see section 8.4); (4) source separation to extract the FECG from the abdominal mixture (see section 8.3.1); (5) FQRS detection (see section 8.2); (6) one of the FQRS time-series is selected (see section 8.4); (7) the resulting time-series is smoothed; (8) finally the post-processed signal is scored against the reference FQRS annotations or FHR trace (see section 10). Picture reproduced from Behar *et al* (2014d).

ventricular rhythm is established at 60–80 bpm (Symonds *et al* 2001)). By the ninth week, the sino-auricular node has developed and the baseline FHR becomes 175 bpm. In the following weeks FHR declines and reaches between 110–160 bpm at delivery (von Steinburg *et al* 2012). There are a large number of FHR patterns that have been studied and it is beyond the scope of this article to review this extensive literature on FHR interpretation. For more information on such patterns the reader is referred to Sweha and Hacker (1999). However, the large inter- and intra-expert variance in interpretation of clinical guidelines means that there is little evidence for the efficacy of the use of such patterns (Downs and Zlomke 2007). This, together with the lack of clear, non self-referential definitions of FHR patterns is another reason that open clinical databases and standardised open source algorithms are needed.

Figure 6 shows a generic block diagram which represents the typical steps undertaken in processing the AECG to extract the NI-FECG and FHR: (1) the AECG channels are preprocessed (see section 8.1); (2) MQRS detection is performed (see section 8.2); (3) at least one of the MQRS time-series is selected (see section 8.4); (4) source separation to extract the FECG from the abdominal mixture (see section 8.3.1); (5) FQRS detection (see section 8.4); (6) one of the FQRS time-series is selected (see section 8.4); (7) the resulting time-series is smoothed; (8) finally the post-processed signal is scored against the reference FQRS annotations or FHR trace (see section 10). Note that step (8) is not part of 'routine' assessment but only used for the development and benchmarking of algorithms. Some of these key steps are discussed below.

8.1. Preprocessing

The prefiltering step is crucial for most biosignal processing application. In Behar *et al* (2014d) 18 source separation algorithms for NI-FECG extraction were evaluated with different baseline wander cut-off frequency (f_b). For all approaches but three a higher f_b ($f_b = 10$ Hz) performed better (up to 3% increase in the F_1 accuracy measure) than a lower f_b ($f_b = 2$ Hz). Similar findings were obtained in Behar *et al* (2014b) where a grid search was performed over the prefiltering parameters (f_b and the higher cut-off frequency, f_h). In Andreotti *et al* (2016) a

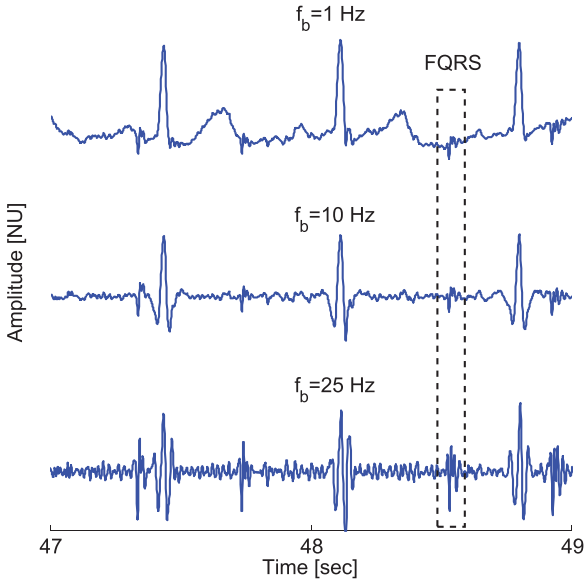

Figure 7. Effect of varying the low cut-off frequency f_b for the preprocessing step. Note that at $f_b = 10$ Hz most of the frequency content of the P and the T-wave of the MECG have been filtered out, leaving only the FQRS and the MQRS. Picture reproduced from Behar *et al* (2014b).

cut-off of 3 Hz was used. In the above mentioned publications the value of f_b were much higher than the ones traditionally used for adult ECG prefiltering (in the range 0.05 Hz–0.7 Hz). Figure 7 illustrates why better outcomes in MECG cancellation are expected with a higher f_b than what is traditionally used for adult ECG prefiltering: a high f_b suppresses the P and T waves, leaving only the MQRS and FQRS (and some band-limited noise) for separation in the abdominal mixture.

8.2. QRS detection

Many QRS detectors were published over the past 50 years. For adult QRS detection (thus MQRS) some very accurate detectors were open sourced and evaluated (see Johnson *et al* (2016) for a benchmark of the most well-known open sourced QRS detectors). For FQRS detection many FQRS detection algorithms were open sourced during the Challenge 2013. These FQRS detectors have often been adapted from traditional adult QRS detectors with some parameters, such as the refractory period, adjusted for the higher metabolic rate. However, so far, it has been impossible to compare their relative performance, because the outcome of the detection were the combined result of many cascaded steps (figure 6), that are necessary to separate the FECG signal from the abdominal mixture. To the authors' best knowledge, no published work to date has focused on benchmarking FQRS detectors on the extracted NI-FECG for a given source separation algorithm.

8.3. Source separation

8.3.1. Before the 2013 PhysioNet/Computing in Cardiology Challenge. The FQRS location is the primary feature that any algorithm must extract from the AECG signal mixture. This peak

Table 3. Foetal separation algorithms for NI-FECG extraction.

Category	Examples
BSS	BSS_{ica} (De Lathauwer *et al* 2000, Zarzoso and Nandi 2001), $BSS_{\pi CA}$ (Sameni *et al* 2008), GP (Niknazar *et al* 2013a), TD (Niknazar *et al* 2012)
AM	AM_{lms} (Widrow *et al* 1975), AM_{rls} and AM_{esn} (Behar *et al* 2014b)
TS	TS_c (Cerutti *et al* 1986), TS_m (Martens *et al* 2007), TS_{pca} (Kanjilal *et al* 1997), TS_{lp} (Ungureanu *et al* 2007), TS_{ekf} (Sameni 2008, Zaunseder *et al* 2012)

BSS: blind source separation.
AM: adaptive filtering methods.
TS: template subtraction.
GP: Gaussian processes.
TD: tensor decomposition.

detection is used for computing the FHR, detecting rhythm abnormalities, or is further used as an anchor point for extracting features from the FECG waveform. Ascertaining the location of the FQRS is simplified by first separating the FECG from the AECG, and several approaches have been suggested in the literature. These include: principal component analysis (Kanjilal *et al* 1997), independent component analysis (ICA) (De Lathauwer *et al* 2000, Zarzoso and Nandi 2001), or periodic component analysis (πCA) (Sameni *et al* 2008) which makes use of the ECG's periodicity. In essence, these approaches are a form of blind source separation (or in the case of πCA semi-blind source), which aim to separate the underlying statistically independent sources into three categories: MECG, FECG and noise. Other techniques, which operate in lower dimensions (i.e. using a lower number of, or single, abdominal channel) include adaptive filtering (Widrow *et al* 1975, Camps *et al* 2001, Behar *et al* 2014b), template subtraction (Cerutti *et al* 1986, Kanjilal *et al* 1997, Martens *et al* 2007, Ungureanu *et al* 2007), Kalman filtering (KF) (Sameni 2008, Zaunseder *et al* 2012, Andreotti *et al* 2014), Gaussian processes (Niknazar *et al* 2013a), tensor decomposition (Niknazar *et al* 2012) and the combination of multiple methods (Behar *et al* 2014d). Refer to chapter 3 of the thesis from Behar (2014) for a detailed presentation of some of the main algorithms. Table 3 classifies these algorithms into three categories: blind source separation (BSS), adaptive filtering methods (AM) and template subtraction methods (TS).

8.3.2. The 2013 PhysioNet/Computing in Cardiology Challenge. The 2013 PhysioNet/Computing in Cardiology Challenge (Silva *et al* 2013) addressed the topic of NI-FECG extraction. Over 50 teams entered the competitions and benchmarked their algorithms on the same dataset and using the same statistics. A variety of separation algorithms were used such as: template subtraction (Andreotti *et al* 2013, Christov *et al* 2013, Dessì *et al* 2013, Di Maria *et al* 2013, Kropf *et al* 2013, Kuzilek and Lhotska 2013, Lipponen and Tarvainen 2013, Liu and Li 2013, Llamedo *et al* 2013, Lukoševičius and Marozas 2013, Maier and Dickhaus 2013, Marco *et al* 2013, Starc 2013), Kalman filtering (Andreotti *et al* 2013, Haghpanahi and Borkholder 2013), tensor decomposition (Akhbari *et al* 2013, Niknazar *et al* 2013b), deflation (Fatemi *et al* 2013), adaptive filtering (Rodrigues 2013) as well as a fusion of different algorithms (Behar *et al* 2013b). Separation steps used in the best entries of the Challenge and follow-up special issue (Clifford *et al* 2014d) included: fusion of multiple source separation algorithms (Behar *et al* 2014d), template subtraction (Andreotti *et al* 2014), combination of template subtraction and ICA (Varanini *et al* 2014) and adaptive filtering (Rodrigues 2013).

8.3.3. After the Challenge 2013. Recent work from Ghaffari *et al* (2015) directed at FQRS detection without having to separate the FECG from the abdominal mixture. The authors

performed MQRS detection and looked for high energy zones out of MQRS domain. Although the authors report a high performance for their algorithm they do not separate their data into train-validation-test sets and they did not evaluate the algorithm on set-b of the PCDB which would have provided an objective performance measure for comparison against other algorithms. Instead, the authors presents a table summarising the performance of their algorithm against other published ones, but with an evaluation performed on different databases, and using differently defined 'accuracy' measures for some of the table entries. Finally, the reported accuracy figures sometime come from the evaluation of an algorithm on a test set of data and sometimes on a training set which makes unfair and irrelevant comparisons. Such publications are a classic example of the non-rigorous approach often seen in the signal processing literature, where the authors' results are often inflated due to over fitting, and hand selection of non-comparable metrics. The work of Liu and Luan (2015) used an adaptive algorithm based on a combination of ICA, ensemble empirical mode decomposition (EEMD), and wavelet shrinkage (WS) denoising. In Ghazdali *et al* (2015) the authors developed a new BSS approach for NI-FECG extraction. In Martinek *et al* (2015) the authors suggested a multichannel adaptive neuro-fuzzy interference system to extract the NI-FECG. Although some of these works seem promising, their evaluations suffer from similar limitations as in Ghaffari *et al* (2015), in that it is not possible to draw substantial conclusions with respect to the relative performance of these algorithms and with respect to existing ones. In Noorzadeh *et al* (2015) the authors used two phonocardiogram (PCG) reference channels (one for the foetus and one for the mother) and focused on extracting the FECG and MECG morphology using a Gaussian process based algorithm. Although the approach is of interest from a machine learning perspective, its clinical practicality is questionable given that two additional PCGs sensors have to be correctly positioned and can be prone to noise. Moreover, the data used in the study was recorded from a single pregnant woman in the 8th month of pregnancy. In Andreotti *et al* (2016), the authors benchmarked eight extraction techniques in terms of FQRS detection accuracy, using a large database of simulated signals (45.8 h of data with almost 1.1 million foetal R-peaks) containing different types of events (further described in section 7.3). This contribution provides a large collection of simulations, extraction algorithms and performance analysis, all made freely available on www.Physionet.org (Goldberger *et al* 2000). The authors' results have shown that all of the three categories of techniques (presented in table 3) are able to extract FQRS complexes with some relative differences in the extraction outcome depending on the types of non-stationarities present in the abdominal mixture. Within the TS category TS_{pca} (Kanjilal *et al* 1997) performed the best, within the AM category AM_{esn} (Behar *et al* 2014b) performed the best and within the BSS category BSS_{ica} performed the best. Finally, Yaping *et al* (Ma *et al* 2015) proposed a multichannel nonlinear adaptive noise canceller based on the generalized functional link artificial neural network (GFLANN). They compared their algorithm against alternative AMs and made use of a training and test set, although the volume of their database remained small (14, 1 min segments for the training set and 8, 1 min segments for the test set). They showed that their algorithm was outperforming all the benchmark techniques with exception of the ESN (Behar *et al* 2014b). However, the computational time of the GFLANN was 3–6 times lower than for the ESN.

There have been a number of contributions published since the 2013 Challenge. However, many of those did not take into account the critical points stressed during the Challenge: the necessary usage of a separate training and test set and the requirement for reporting a novel algorithm and benchmark performances using the same database and same evaluation statistics. Given the existence of the Challenge there is little reason not to include these data in any future publications. We therefore further stress the need for standardised evaluation in the comparison of new NI-FECG algorithms.

8.4. Channel selection

After the source separation step is performed the algorithm has to deal with a number of ECG residuals (in the case of the TS or AM algorithms) which contain the FECG and noise components (in the case of blind source separation). One of the channel RR time series need to be selected or the RR-intervals time series from all (or a subset) of the extracted channels need to be combined. Multiple approaches were suggested in the context of the Challenge 2013. These were usually based on the regularity measure of the RR time series (Andreotti *et al* 2014, Behar *et al* 2014d, Liu *et al* 2014, Varanini *et al* 2014). However, such approaches are likely limited in the instances of abnormal or non-stationary rhythms (acceleration/deceleration). This needs to be evaluated.

Additional ideas emerged with the Challenge 2014 (Moody *et al* 2014) that addressed the topic of *Robust Detection of Heart Beats in Multimodal Data*. Participants were asked to robustly estimate the R-peak location from multiple channels of biosignals including (adult) ECG. In particular one of the best approaches in this competition made use of signal quality indices to assess what channel to trust (Johnson *et al* 2015). Signal quality was successfully used in the context of NI-FECG in Liu *et al* (2014), in order to exclude bad quality abdominal channels and in (Behar *et al* 2014b, Behar *et al* 2013c) to remove bad quality maternal chest ECG leads that were used as a reference in an AM framework. However, signal quality was not used on the residual signals extracted to select what channel was the most trustworthy. The beat comparison measure (BCM) (Behar *et al* 2014d) was also used as a first check to exclude QRS time series detected on the residual signal that matched the MQRS time series. This is important in cases where a TS or AM technique would fail and leave a high amplitude MECG contribution in the residual signal. These signal quality indices (SQIs) allow more accurate MQRS extraction, to some extend discard some very noisy segment and avoid confusing the MQRS with the FQRS time series. However, no SQI specifically designed for abdominal FECG quality evaluation has been proven to improve NI-FECG extraction.

8.5. Intermediate conclusions

- Using a high baseline wander cut-off frequency f_b to prefilter the signal has improved FQRS detection. It naturally removes the maternal (and foetal) P and T-waves, thus facilitating the source separation (Behar *et al* 2014b, 2014d, Andreotti *et al* 2016).
- For MQRS detection we recommend the use of classical open sourced adult QRS detectors such as the ones benchmarked in Johnson *et al* (2016)[8] or in Oster *et al* (2013).
- There are different families of source separation algorithms which can be categorised as Behar *et al* (2014d): (1) blind/semi-blind (BSS), (2) adaptive filtering methods (AM) and (3) template subtraction (TS). In addition, some work combined multiple separation algorithms.
- Among the template subtraction techniques TS_{pca} gave the best results over alternative template subtraction (TS) techniques and was successfully used by some of the top Challenge participants (Behar *et al* 2014d, Lipponen and Tarvainen 2014, Varanini *et al* 2014) and in a recent exhaustive algorithms benchmark on simulated data (Andreotti *et al* 2016). Among the adaptive techniques the AM_{esn} has proved to be the best in the work of Andreotti *et al* (2016), Behar *et al* (2014b) and Ma *et al* (2015).
- Adding an ICA step after performing source separation using one of the temporal techniques showed to improve the FQRS detection performances (Behar *et al* 2014d, Varanini *et al* 2014).

[8] Available at: https://github.com/alistairewj/peak-detector

- No single extraction method showed to be systematically the best when considering different types of non-stationarities in simulated data (Andreotti *et al* 2016). Accordingly, Behar *et al* (2014d) showed that combining a series of NI-FECG extraction algorithm was performing better than any individual algorithm on the Physionet Challenge database.
- When using ICA careful thought should be given when setting the number of inputs and outputs (expected sources). Andreotti *et al* (2016) highlighted that an inappropriate number of sources could result in sub-optimal source separation performance of the NI-FECG. In their work the authors advised to remove the low energy components (low eigenvalue) obtained using a PCA dimension reduction step.
- Most source separation algorithms have parameters that need to be tuned. We recommend using grid search or random search (Behar *et al* 2013a, 2014d)[9] on a separate training dataset in order to evaluate what parameters has the highest impact and find optimal values rather than hand-tuning them (which is sometimes performed on the same set).
- Open-source code for NI-FECG extraction techniques are available through the OSET toolbox[10], on Physionet under the Challenge 2013[11] as well as on the *fecgsyn* Physionet[12].

9. Morphological analysis

The ECG allows for a more in-depth interpretation of the electrical activity of the heart than only HR and HRV. However, morphological analysis of the FECG waveform is currently not performed in clinical practice with the exception of the STAN monitor (Neoventa Medical, Mölndal, Sweden), which uses an invasive scalp ECG (SECG) electrode.

Several ECG morphological analysis features have been studied in a foetal monitoring context (see chapter 6 of Symonds' book for a good summary—Symonds *et al* 2001). These include: width and shape of the QRS complex, R/S ratio (for foetal vector cardiography), P wave morphology (inversion, notching and disappearance), PR interval, QT interval and ST-segment. A summary of the temporal and morphological intervals in the FECG is given in table 4. Note that the length of these intervals highly depends on the gestational age.

In the next paragraphs particular focus is given to the QT measurements, the ST segments and the T/QRS ratio, which have been proposed in literature and STAN monitoring, showing promising results:

QT-segment: In adults, changes in the T-wave occur during myocardial ischemia. It has been shown that acute myocardial ischemia will modify the duration of the QT interval and increase repolarisation heterogeneity (Candil and Luengo 2008). The FQT interval has thus been of much interest in the monitoring of foetal hypoxia. In a study by Oudijk *et al* (2004), a significant shortening of the QT interval has been shown to be associated with intrapartum hypoxia resulting in metabolic acidosis, whereas in normal labour none of such changes do occur. In Behar *et al* (2014e) the authors showed the possibility to recover the foetal QT from the NI-FECG. Three clinicians manually annotated the foetal QT on the NI-FECG and FSE from 22 labouring women. The annotations were fused and the errors found between NI-FECG and FSE QT were in the range of QT annotations performed on adult ECGs.

[9] Available at: www.physionet.org/physiotools/random-search/
[10] Available at: http://spc.shirazu.ac.ir/products/Featured-Products/oset/
[11] Available at: http://physionet.org/challenge/2013/sources/
[12] Available at: http://physionet.org/physiotools/ipmcode/fecgsyn/

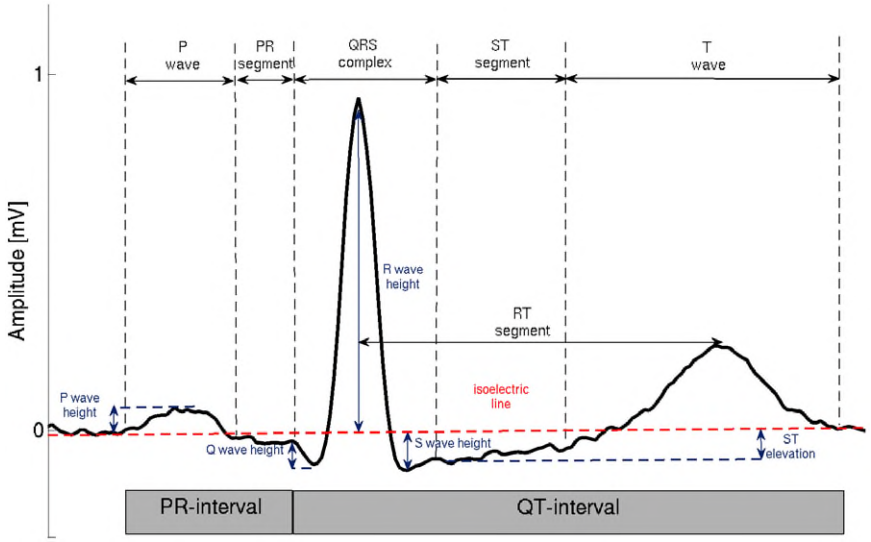

Figure 8. ECG waveform with characteristic waves. Note that the PR segment is sometimes called the PQ segment.

ST-segment: Hypoxia typically causes ST segment and T wave elevations in the FECG. This is due to a catecholamine surge, h-adrenoceptor activation and myocardial glycogenolysis (Hökegård *et al* 1981, Rosen *et al* 1984, Widmark *et al* 1991). In Clifford *et al* (2011), the authors recorded the NI-FECG on 32 term labouring women who had a FSE placed after clinical indication. They evaluated the accuracy of the ST segment extracted on the NI-FECG (by an automated algorithm) against the reference ST segment extracted from the FSE (using the same automated algorithm.) The root mean square error between the ST change calculated by both modalities averaged over all processed segments was 3.2% thus showing that accurate extraction of the ST segment from the NI-FECG is feasible. However, one of the main limitations of the study was the usage of the same algorithm for measuring the ST segments on the FSE and NI-FECG. Ideally, the reference FSE should be annotated manually.

T/QRS-ratio: The T/QRS ratio is defined as the height of the T-peak divided by the QRS amplitude. The interest in using the T/QRS ratio as a proxy for the ST segment originates from an animal experiment by Greene *et al* (1982), where the authors examined 10 chronically instrumented foetal lambs, 115 days to term. The study showed that the normal T/QRS ratio was lower than 0.30, whereas it was in the range of 0.17 to 0.59 for eight of the lambs after inducing hypoxia and reverted to normal with normoxia. The SECG-based STAN monitor uses the T/QRS ratio as a proxy for the ST segment deviation. The use of the STAN analyser together with competency-based training on foetal monitoring has recently shown a significant decrease in the number of cesarean sections at St George's Maternity Unit (St George's hospital, London, UK). In the mean time hypoxic ischaemic encephalopathy and early neonatal death has slightly decreased (Chandraharan *et al* 2013). However, a recent Cochrane study (Neilson *et al* 2006) reviewed six trials that compared the effect of analysing SECG waveforms during labour with alternative foetal monitoring methods, and showed that no significant difference in primary outcomes were achieved using the STAN ST proxy (based on five trials using different versions of the

Table 4. Definition of temporal and morphological intervals in the FECG.

Time intervals		
Parameters name	Definition	Interval
P wave duration	Duration between the onset and end of the P-wave	22–83 ms Stinstra *et al* (2002) ($n = 400$, wg: 16–42)[a]
PR segment	Duration between P and R peaks	66–166 ms Stinstra *et al* (2002) ($n = 534$, wg: 16–42)[a] 86–141 ms Taylor *et al* (2005) ($n = 12$, wg: 24–41)
RT segment	Duration between R and T peaks	—
QRS complex	Duration between the onset of the Q wave and end of the S wave	18–75 ms Stinstra *et al* (2002) ($n = 579$, wg: 16–42)[a] 26–61 ms Brambati and Pardi (1980) ($n = 421$, wg:17–41)[b] 47–85 ms Taylor *et al* (2005) ($n = 12$, wg: 24–41)
T wave	Duration between onset and end of the T wave	85–180 ms Stinstra *et al* (2002) ($n = 412$, wg: 16–42)[a]
QT segment	Duration between Q onset and end of the T wave	149–339 ms Stinstra *et al* (2002) ($n = 412$, wg: 16–42)[a], 207–338 ms Abboud *et al* (1990) ($n = 21$, wg: 32–41), 233–329 ms Taylor *et al* (2005) ($n = 11$, wg: 24–41)

Amplitude and area measures	
Parameter name	Definition
P wave height	Amplitude from isoelectric line to peak of R-wave
P wave area	Area between the P wave and the isoelectric line
PR segment elevation	Amplitude difference between the end of the P wave and the start of the Q wave
Q wave height	Amplitude from isoelectric line to peak of Q-wave
R wave height	Amplitude from isoelectric line to peak of R-wave
S wave height	Amplitude from isoelectric line to peak of S-wave
ST segment elevation	Amplitude between the end of the S wave at the J point and the begining of the T-wave
T wave height	Amplitude from isoelectric line to peak of T-wave
T wave area	Area between T wave and isoelectric line
T/QRS-ratio	Ratio of height of T wave to amplitude of the QRS complex

Note: (Table adapted from Symonds *et al* (2001) p 66). See figure 8 for a pictorial representation of a subset of these wave delineations.
n: number of foetuses
wg: number of weeks of gestation.
[a] Measurement on magnetocardiography.
[b] By observation of the authors' figure 2.

STAN monitor with a total of 15,338 women). More recently, another study including 11,108 women has revealed that ST-segment analysis, as an adjunct to conventional intrapartum electronic foetal heart-rate monitoring, did not improve perinatal outcomes and did not decrease operative-delivery rates either (Belfort *et al* 2015). This suggests that the STAN proxy for the ST level is either not accurate enough, or that it does not provide meaningful information for foetal monitoring.

The assumption of the above mentioned studies, is that heart abnormalities or foetal hypoxic suffering will manifest in the FECG waveform. This is supported by observations in adults and animal models. However a limited number of studies have managed to show a

significant improvement in foetal outcomes or decrease the proportion of cesarean deliveries, because of the difficulty in performing accurate morphological analysis on the FECG (either through SECG or AECG). Symonds *et al* concluded in 2001: 'The issue of the value of current use of the FECG morphological characteristics and time intervals for the prediction of foetal compromise remains promising but unresolved' (Symonds *et al* 2001). It is fair to say that in 2016, 15 years later, the problem remains unresolved.

9.1. Preprocessing

Preliminary processing steps are crucial for the outcome of morphological analysis. For instance, the effect of the baseline frequency, f_b, on the FQT extraction is yet to be determined. It is known that a high f_b will distort the adult T-wave morphology, negatively affecting the QT estimation. However, the baseline wander is rather substantial on the abdominal recordings and needs to be significantly attenuated. The trade-off between less QT distortion and good baseline removal needs to be studied. In adult ECG a cut-off around 0.5 Hz is traditionally used as a upper bound to preserve most of the T-wave (Kligfield *et al* 2007). Several other preprocessing steps play an important role for the reliability of morphological analysis, one can cite for example the accuracy of the detection of the FQRS, the generation of an average template (in case of averaging of several beats), and the reliability of the automatic segmentation techniques. Moreover, the influence of alternative extraction methods on morphological analysis was evaluated (Andreotti *et al* 2016). It was shown by using simulated data that the transformation from the observation to source domain by using a BSS technique, can significantly change the estimates of FQT intervals. Morphological analysis should therefore only be performed in the original observation domain.

9.2. Algorithms

In Sameni *et al* (2007b) a Bayesian filtering framework was introduced for the purpose of extracting the FECG. This original algorithm (denoted EKFS) used prior information on the MECG cycle morphology in order to suppress it from the abdominal mixture. The FECG could then be estimated by subtracting the MECG from the AECG thus leaving the FECG plus noise in the residual. Niknazar *et al* (2013a) extended the EKFS so that both priors on the MECG and FECG cycles morphology were modelled (algorithm denoted EKFN). At each sample of the AECG signal the contribution of the FECG and MECG were estimated. As such the algorithm allowed to estimate the beat by beat FECG morphology while filtering out the noise contaminants. The algorithm was essentially evaluated on artificial data generated using the dipole model of Sameni *et al* (2007a). The equations were recently reformulated in Behar *et al* (2014c) (denoted EKFD) to account for some drift in the filtering process as well as to allow the prior on the FECG cycle morphology to evolve so to account for the non-stationarity of the FECG. The EKFD showed better performance in estimating the foetal QT from simulated data generated using the *fecgsyn* simulator (Behar *et al* 2014a) than EKFS or EKFN. Figure 5 shows an example of an artificial AECG mixture and the estimated FECG using the EKFD algorithm on top of the 'true' FECG (i.e. FECG generated by the simulator and before being mixed on the abdomen with the MECG and noise.) More work for the validation of this NI-FECG morphological extraction approach is needed, particularly by evaluating the algorithms on real data and for a variety of morphological measures (table 4).

Andreotti *et al* (2016) compared the ability of retrieving the FQT and T/QRS ratio from the extracted FECG using a variety of separation algorithms. They showed on simulated

data that the simple TS class of methods was performing better than the alternative methods such as ICA and the echo state neural network. This was interpreted as being due to the fact that more adaptive extraction algorithms are more likely to cause foetal T-wave distortion and in the case of ICA the T-wave measured in the source domain will also be different than in the observation domain. Moreover, all techniques used in Andreotti *et al* (2016) have been previously optimised for accurate FQRS detection, therefore, those methods' parameters might be suboptimal for morphology analysis. This original experiment needs to be reproduced on real data as its conclusions will direct the research in NI-FECG morphological extraction by identifying what class of methods (BSS, AM, TS) is the most suited to this purpose.

9.3. Intermediate conclusions

- Careful consideration should be given to the prefiltering cut-off frequencies when considering morphological analysis. A trade-off between baseline wander removal and distortions of the FECG morphology should be further investigated.
- Performing morphological analysis after a BSS step (i.e. in the source domain) can lead to inaccurate estimation of the morphological parameters (Andreotti *et al* 2016).
- We motivate pushing forward morphological analysis of the NI-FECG and in particular the study of the FQT segment, whose usefulness is supported by studies in adult ECG and animal models.
- The Dual Kalman filter built upon the Bayesian filtering framework introduced in Sameni *et al* (2007b) represents a good candidate algorithm for NI-FECG morphological analysis (Behar *et al* 2013).
- Preprocessing, template generation and segmentation are crucial aspects of morphological analyses and need further investigation/development.
- There is, to date, no open-source segmentation algorithm specifically designed for FECG morphological analysis (whether for the SECG or the NI-FECG). This is lacking in the field.

10. Performance measures

10.1. Obtaining reference annotations

The process of obtaining annotations from experts should be considered carefully, since it is a time consuming process which suffers from inter-rater variability. While aiming at a consensus annotation, it is useful to consider crowd sourcing annotations from a number of individuals such as in the recent work of Zhu *et al* (2014, 2015, 2013). Zhu *et al* combined adult QT annotations from different individuals and automated algorithms in order to improve the estimation of the reference annotations. This algorithm has recently been used in Behar *et al* (2014e) to estimate foetal QT (FQT) by merging the annotations from three medical professionals (see figure 9). We recommend using a consensus of at least three human annotations for obtaining morphological references such as FQT interval particularly given that no open SECG segmentation and signal quality algorithm exists to date. (Although signal quality research on adult data might be sifficient and recent research indicates that automated algorithms can provide accurate ST estimates (Clifford *et al* 2011).) Regarding FQRS references, we recommend to record and use the SECG, then employ an accurate FQRS detector (used in the Challenge) in order to identify the location of reference QRS complexes. If the SECG is not available, multiple expert annotation of raw ultrasound data may prove acceptable.

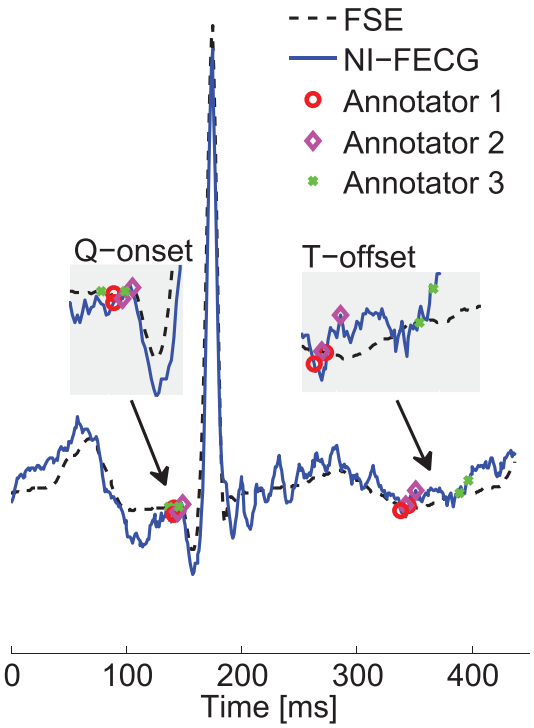

Figure 9. Comparison of annotations performed on an average FSE and NI-FECG ECG beat by three experts. This example illustrates the importance of having multiple experts labelling the data particularly in the case of the NI-FECG which is noisier and thus more challenging to annotate. For this example each annotator recorded the Q-onset and T-end for the FSE and NI-FECG separately. Thus each annotator produced two annotations points for the Q-onset (one on the FSE and one on the NI-FECG) and two for the T-end (one on the FSE and one on the NI-FECG). Picture adapted from Behar (2014).

10.2. Foetal heart rate

It is possible to assess the error for FQRS detection using a window based metric (e.g. FQRS within $\pm$XXX ms window) or using a distance based measures (e.g. absolute error or root mean square error between the detected R-peak and the true R-peak locations—see table 5). Following the ANSI/AAMI guidelines (ANSI/AAMI/ISO EC57 1998/(R)2008) the classical statistics for evaluating QRS detectors are: sensitivity (Se) and positive predictive value (PPV). A detected QRS is considered a true positive if it is within 150 ms from the reference annotation for adults. In the context of NI-FECG extraction, a window of 50 ms is classically used to take into account the higher foetal HR than HR in adults. The *bxb()* function from the WFDB Toolbox (Goldberger *et al* 2000, Silva and Moody 2014) can be used to compute the Se and PPV statistics. Defining a matching window is necessary because one cannot expect that every detectors will detect a QRS complex with a single sample accuracy.

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \qquad \mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{1}$$

Table 5. Summary of the statistics used for assessing the performances of FQRS detection and FECG morphological analysis (on the example of FQT and T-wave height).

Foetal heart rate statistics	
Statistics	Definition and comments
$\text{Se} = \dfrac{\text{TP}}{\text{TP} + \text{FN}}$	The proportion of true FQRS that have been detected out of the total number of true FQRS. This statistic tells how good an algorithms is at finding the true FQRS. The ANSI/AAMI guidelines (ANSI/AAMI/ISO EC57 1998/(R)2008) recommend to report Se when evaluating a QRS detector.
$\text{PPV} = \dfrac{\text{TP}}{\text{TP} + \text{FP}}$	The proportion of the true FQRS that have been detected out of all the detected FQRS. This statistic tells how good the algorithm is at identifying true FQRS out of all the detections it makes. The ANSI/AAMI guidelines (ANSI/AAMI/ISO EC57 1998/(R)2008) recommend to report the PPV when evaluating a QRS detector.
$F_1 = 2 \cdot \dfrac{\text{PPV} \cdot \text{Se}}{\text{PPV} + \text{Se}}$	The F_1 statistic is an average of the Se and PPV thus providing a good summary metric in the absence of TN. It is a harmonic mean and is suited for situations when the average of rates is desired (Sasaki 2007). It can be used as an accuracy measure when training an algorithm and to summarise the final overall performance of it in accurately detecting the R-peak locations. This statistic was introduced in the context of NI-FECG in Behar *et al* (2014b).
$E1^k/E4^k = \dfrac{1}{12}\sum_{i=1}^{12} \Delta h_i^2$	Mean square error computed from the differences between matched reference and test FHR measurements at 12 instances for each one minute segment (i.e. one approximately every 5 s). $\Delta h_i = h_i^t - h_i^r$, with h_i^t the test FHR on the ith segment and h_i^r the reference FHR for the ith segment. The purpose of this statistic is to assess the accuracy of an algorithm in evaluating the HR from the detected R-peaks. This statistic was introduced in Silva *et al* (2013).
$E2^k/E5^k = \sqrt{\dfrac{1}{M}\sum_{i=1}^{N} \Delta RR_i^2}$	Root mean square error computed from the differences between matched reference and test RR intervals. $\Delta RR_i = RR_i^t - RR_i^r$ with RR_i^t the reference RR, N is the number of test measurements and M is the number of reference measurements. The purpose of this statistic is to assess an algorithm in its performance to accurately evaluate the RR interval. This statistic was introduced in Silva *et al* (2013).
$HR_m = N/M$	Percentage of the HR within ± 5 bpm of the reference HR. M number of reference points, N number of test points that are within ± 5 bpm of their corresponding reference measurement. The purpose of this statistic is to evaluate the accuracy of an algorithm in extracting the heart rate with a tolerance of ± 5 bpm as used in industrial practice (ANSI/AAMI 2002). Thus in this case the statistic is window based and not RMS based. This measure was suggested in Behar *et al* (2014b).

(*Continued*)

Table 5. (*Continued*)

Morphological analysis statistics	
Statistics	Definition and comments
$\mathrm{SNR} = 20 \cdot \log(\sqrt{\sum_{i=1}^{K} r_i^2 / \sum_{i=1}^{K} (f_i - r_i)^2})$	The SNR, in decibels (dB), between a reference r and an extracted signal f with K samples. This measure can only be used with simulated data where the ground truth NI-FECG signal is known on each abdominal channel (i.e. before being mixed with the maternal and foetal signals.) It as been used in Behar *et al* (2013).
$\widetilde{\mathrm{FQT}} = \lvert \mathrm{FQT}_a - \mathrm{FQT}_r \rvert$	Absolute error between the measured abdominal FQT (FQT_a) and a reference FQT interval (in seconds). This reference can be obtained by annotating or automatically detecting FQT in the SECG recordings or using simulated data (in which case the reference FECG is known). The statistic was used in Andreotti *et al* (2016). It can be used for any of the time interval measures listed in table 4.
$\widetilde{\mathrm{TQRS}} = \lvert \mathrm{TQRS}_a - \mathrm{TQRS}_r \rvert$	The foetal T/QRS error ($\widetilde{\mathrm{FQRS}}$) is evaluated by taking the absolute error T/QRS ratios (see table 4) from the extracted NI-FECG and the reference ratio that can be obtained from the FSE or known when using simulated data (Andreotti *et al* 2016).

where TP, FP and FN are true positive (correctly identified FQRS), false positive (wrongly detected FQRS) and false negative (missed FQRS) detections respectively. For algorithm parameter optimisation, the F_1 statistic can be used as the accuracy measure. In the context of binary classification, F_1 is defined as:

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

The F_1 measure is an harmonic mean. It is suited to situation when the average of rates is desired (Sasaki 2007). In order to get an idea of why the F_1 measure is preferable over the arithmetic mean, consider the following example: $\mathrm{Se} = 1$ and $\mathrm{PPV} \approx 0$ (which could correspond to a QRS detector annotating every single sample of the ECG as a R-peak) then the arithmetic mean would be $\approx$0.5. Meanwhile, we would have $F_1 \approx 0$ which is certainly more representative of the outcome of the QRS detection algorithms. The F_1 measure was introduced in the context of QRS detection in Behar *et al* (2014b) and well adopted by many participants for the follow-up special issue to the Physionet Challenge 2013 (Andreotti *et al* 2014, Behar *et al* 2014d, Di Maria *et al* 2014, Liu *et al* 2014) as well as used for the Challenge 2014 (Johnson *et al* 2015).

Using a window-based approach also has limitations: let assume that two sets of annotations are provided, a reference set with stable fiducial markers (i.e. every beat is marked at the same location on the QRS complex, say the R-wave peak), and another with highly variable markers (e.g. because its creator attempted unsuccessfully to mark QRS onsets, with an unreliable method). Assume, however, that the second set does contain an annotation for each beat, within *bxb*'s window of tolerance. Using *bxb*, both sets would be in perfect agreement, but no information on the precision of the FQRS locations (and consequently the estimated RR intervals) would be given. It is therefore important to compare also the 'interval measurements' and not only the accuracy of the QRS detection. This is the idea behind the use of *mxm* in the E2/E5 Challenge scores (refer to table 5).

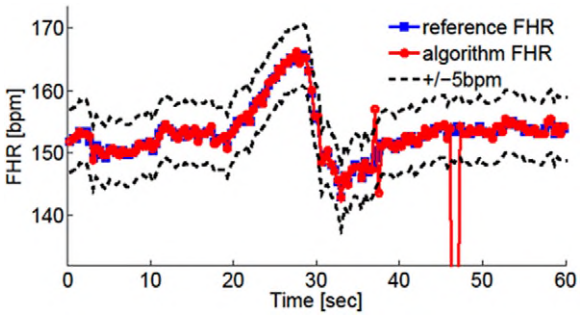

Figure 10. FHR trace. The plot displays the reference and extracted FHR for the first 60 sec of channel 4, record 172a from the NIFECGDB. Black lines indicate the delineation at ±5 bmp with respect to the reference FHR trace.

The motivation for performing FQRS detection can be two-folds: (i) obtaining an accurate estimate of the heart rate; (ii) localising precisely the absolute position of the R-peaks in order to use them as anchor points for further processing such as reconstructing the morphology of the NI-FECG and automated QT and ST measurements. Some error measurements better serve the purpose of (i) whereas other will better serve (ii). Ultimately the choice of the FQRS detection precision measurement should be conditioned by the clinical problem i.e. it should reflect how well the algorithm is performing in identifying a pathology or an abnormal event, for example. In order to clarify the purpose and advantages or limitations of the main statistics, a discussion together with a summary table (table 5) are included below.

Using the RMS error as a distance measure between two RR time-series (event E1/E4 of the Challenge) might be sub-optimal in many cases. Figure 10 illustrates this statement by showing a real example where the reference and algorithms estimated FHR are mostly equivalent apart from a small interval at 47 s. Because of this local error the usage of the RMS measure will drive the error for this record to a high value. The problem will only be present when the FHR time-series outputted by the algorithm is not smoothed. However rigorous smoothing will also 'remove' the high/low variations in HR due to the presence of ectopic beats for example or might lead to an inaccurate outcome in the case of high variability/inadequate calibration of the smoothing function parameters.

10.3. Morphological analysis

A number of performance statistics have been used for NI-FECG morphological analysis. Some SNR based measures have been used for evaluating the extraction performance using the simulator model presented in section 7.3 because the 'true' NI-FECG (i.e. the ground truth) is known with the simulator. The SNR, in decibels (dB), between a reference r and an extracted signal f with K samples is defined as in Behar *et al* (2013):

$$\mathrm{SNR} = 20 \cdot \log\left(\sqrt{\sum_{i=1}^{K} r_i^2 / \sum_{i=1}^{K} (f_i - r_i)^2}\right)$$

However, ultimately the ability of an algorithm to extract an accurate FECG morphology should be assessed in terms of clinically significant parameters, such as the QT segment length or ST elevation. This is because computing an SNR based similarity measure between a reference and extracted ECG gives an overall picture of the extraction, but provides no insights into whether it is possible to recover clinically important parameters from the extracted NI-FECG

(Behar *et al* 2013). One of the main challenges in defining these measures with real datasets is the lack of reference (ground truth). By opposition to adult ECG, where the intervals of interest are manually annotated on the ECG and used as the reference, it is not possible to ensure that manual annotations performed on the extracted NI-FECG are accurate since it is not known whether or not the NI-FECG has been correctly separated from the signal mixture in the first place. For that purpose the recording and annotation of the scalp ECG is required in order to be used as the ground truth. This means that such data and analysis can only be recorded during labour.

Time intervals: for time intervals (see table 4), the absolute error between the measured and reference intervals (in seconds) can be computed. The reference interval can be obtained by manual or automated detection of the segment on the SECG or using simulated data (in which case the reference FECG is known). For example for QT analysis the statistic is expressed below in mathematical form:

$$\widetilde{\mathrm{FQT}} = |\mathrm{FQT}_a - \mathrm{FQT}_r|$$

where FQT_a corresponds to the segment length detected/measured on the abdominal ECG, FQT_r to the segment length of the reference annotation (detected or manually measured on the SECG) and $\widetilde{\mathrm{FQT}}$ to the absolute error in estimating the FQT. The measure is given on the example of the evaluation of the FQT accuracy but it can be used for any time interval listed in table 4.

T/QRS ratio: The foetal T/QRS error ($\widetilde{\mathrm{FQRS}}$) is evaluated by taking the absolute error T/QRS ratios (see table 4) from the extracted NI-FECG (TQRS_a) and the reference ratio (TQRS_r) that can be obtained from the FSE (or known when using simulated data—Andreotti *et al* 2016).

$$\widetilde{\mathrm{TQRS}} = |\mathrm{TQRS}_a - \mathrm{TQRS}_r|$$

10.4. Intermediate conclusion

The choice of statistics very much depends on the intended use of the algorithm:

- The Se, PPV, F_1 (ANSI/AAMI/ISO EC57 1998/(R)2008, Behar *et al* 2014b) statistics allow for the assessment of the presence of R-peaks within a tolerance window (typically 50 ms for this application (Behar *et al* 2014b)). The length of this window is shorter for FECG than for adult ECG (standard of 150 ms) due to the higher FHR.
- The statistic for *E*2/*E*5 using the Physionet *mxm* function allows to assess whether the RR intervals are accurately evaluated. It assesses whether a QRS detector is consistent in positioning the fiducial point on the QRS waveform (not necessarily the R-peak).
- The statistic for *E*1/*E*4 allows to assess whether the detection can result in accurate FHR evaluation. It is however prone to outlier (RMS based statistic) and we recommend using the HR_m (Behar *et al* 2014b) (see table 5).
- Ultimately the ability of an algorithm to extract an accurate FECG morphology should be assessed in terms of clinically significant parameters. Example of such measures were suggested in Andreotti *et al* (2016) and are listed in table 5 for the example of the evaluation of FQT extraction.
- When evaluating a novel NI-FECG algorithms and presenting results for benchmark methods the results should be presented for the evaluation of the novel and benchmark algorithms on the same database and using the same evaluation statistics. Given the recent

number of open source NI-FECG extraction algorithms (that can be used as benchmarks) and of the PCDB this should be a requirement for a novel scientific contribution to be accepted for publication.

11. Discussion

The field of NI-FECG is promising and the apparition of the first generation of commercial monitors confirms the general interest in using this monitoring technique. However, there are many important aspects that need to be researched in order to allow NI-FECG to become a standard for continuous foetal monitoring.

The first critical step is to build an open access reference database of NI-FECG signals. The recent introduction of the PCDB helped to define what the requirements of such a database are: it should encompass enough recordings, contains physiological events such as contraction and HR acceleration and and contains medical annotations. Such a database will allow benchmarking the variety of NI-FECG extraction algorithms that were introduced in the past decade and any novel contribution. Indeed, although the number of algorithms published for this application keeps flourishing, it is very challenging to make an objective statement on their relative performance. In the meantime the *fecgsyn* simulator presented in section 7.3 can be used for proof of concept.

A number of steps are critical in extracting the FHR from abdominal recordings. These were divided as being: (1) preprocessing, (2) MQRS detection, (3) MQRS time series selection, (4) source separation, (5) FQRS detection, (6) FQRS time series selection and (7) time series smoothing to obtain the FHR. Algorithms for NI-FECG extracting were classified in three categories: (1) blind/semi-blind (BSS), (2) adaptive filtering (AM) and (3) template subtraction (TS). The TS_{pca}, AM_{esn} and BSS_{ica} techniques have demonstrated high performance within their category.

The recent work published in Andreotti *et al* (2016) and Behar *et al* (2013) are paving the way to addressing the problematic of NI-FECG morphological analysis in a quantitative manner. Although accurate FHR extraction is needed for foetal monitoring, we believe that the main space for innovation in the field of NI-FECG is in designing methods for accurate morphological analysis of the NI-FECG. Being able to accurately evaluate the various morphological parameters (see table 4) will open the path to many research studies aiming at defining clinical markers of foetal hypoxia, follow-up of foetal growth and identification of cardiac defects.

A large number of NI-FECG algorithms have recently been open-sourced and the location of the source code were referenced in this publication. Any new publication on the topic should consider benchmarking a subset of the most promising ones (e.g. TS_{pca}, AM_{esn} and BSS_{ica}) and report an appropriate subset if not all the statistics reviewed in table 5 on set-a (training set) and set-b (validation set) of the PCDB or private databases. In any instance, any novel algorithm should be benchmarked against the best performing existing algorithms on the same database while reporting the same evaluation statistics.

12. Conclusion

This contribution focused on the processing of the ECG recorded from a set of abdominal sensors recorded on pregnant women, and the extraction of clinically relevant information from this signal. The NI-FECG offers many advantages over alternative foetal monitoring techniques, the most important one being the opportunity to enable morphological analysis of

the FECG, which is important for determining whether an observed FHR event is physiological or pathological during delivery. Commercial applications for NI-FECG monitoring are in their infancy, but there is a growing interest in improving their performance and extending the types of information they provide (such as morphology), with a view to providing more clinically useful decision support. Therefore there is a strong motivation for pushing forward the research in this field.

Acknowledgments

JB is supported in part at the Technion by the Aly Kaufman Fellowship. FA is financially supported by the Conselho Nacional de Desenvolvimento Tecnologico (CNPq—Brazil).

References

Abboud S, Barkai G, Mashiach S and Sadeh D 1990 Quantification of the fetal electrocardiogram using averaging technique *Comput. Biol. Med.* **20** 147–55

Agostinelli A, Grillo M, Biagini A, Giuliani C, Burattini L, Fioretti S, Di Nardo F, Giannubilo S R, Ciavattini A and Burattini L 2015 Noninvasive fetal electrocardiography: an overview of the signal electrophysiological meaning, recording procedures, and processing techniques *Ann. Noninvasive Electrocardiol.* **20** 303–13

Akhbari M, Niknazar M, Jutten C, Shamsollahi M B and Rivet B 2013 Fetal electrocardiogram R-peak detection using robust tensor decomposition and extended Kalman filtering *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 189–92

Alvi M, Andreotti F, Zaunseder S, Oster J, Clifford G D and Behar J 2014 fecgsyn: a graphical user interface for the simulation of maternal-foetal activity mixtures on abdominal electrocardiogram recordings *Proc. of the IEEE Computing in Cardiology Conf. (Cambridge, MA, USA)* vol 41

Andreotti F 2011 Extraction of the fetal ECG from electrocardiographic long-term recordings *BSc Thesis* Technische Universität Dresden

Andreotti F, Behar J, Zaunseder S, Oster J and Clifford G D 2016 An open-source framework for stress testing NI-FECG extraction algorithms *Physiol. Meas.* **37** 627

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

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 Conf. (Zaragoza, Spain)* vol 40 pp 193–96

ANSI/AAMI/ISO EC57 1998/(R)2008 Testing and reporting performance results of cardiac rhythm and ST-segment measurement algorithms

ANSI/AAMI 2002 Cardiac monitors, heart rate meters, and alarms (www.pauljbennett.com/pbennett/work/ec13/ec13.pdf)

Ayres-de Campos D and Bernardes J 2010 Twenty-five years after the figo guidelines for the use of fetal monitoring: time for a simplified approach? *Int. J. Gynecol. Obstet.* **110** 1–6

Barnett S B and Maulik D 2001 Guidelines and recommendations for safe use of Doppler ultrasound in perinatal applications *J. Maternal-Fetal Neonatal Med.* **10** 75–84

Baumert M, Schmidt M, Zaunseder S and Porta A 2016 Effects of ecg sampling rate on qt interval variability measurement *Biomed. Signal Process. Control* **25** 159–64

Behar J 2014 Extraction of clinical information from the non-invasive fetal electrocardiogram *PhD Thesis* University of Oxford

Behar J, Andreotti F, Zaunseder S, Li Q, Oster J and Clifford G D 2014a An ECG model for simulating maternal-foetal activity mixtures on abdominal ECG recordings *Physiol. Meas.* **35** 1537–49

Behar J, Johnson A E W, Clifford G D and Oster J 2014b A comparison of single channel fetal ECG extraction methods *Ann. Biomed. Eng.* **42** 1340–53

Behar J, Johnson A E W, Oster J and Clifford G D 2013a An echo state neural network for foetal electrocardiogram extraction optimised by random search *NIPS (Lake Tahoe, Nevada, USA)*

Behar J, Oster J and Clifford G D 2013b Non invasive FECG extraction from a set of abdominal channels *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 297–300
Behar J, Andreotti F, Oster J and Clifford G D 2013 A Bayesian Filtering Framework for accurate extracting of the non-invasive FECG morphology *Proc. of the Computing in Cardiology Conference (CinC) (Cambridge, MA)* vol 41 pp 53–6
Behar J, Oster J and Clifford G D 2014d Combining and benchmarking methods of foetal ECG extraction without maternal or scalp electrode data *Physiol. Meas.* **35** 1569–89
Behar J, Oster J, Li Q and Clifford G D 2013c ECG signal quality during arrhythmia and its application to false alarm reduction *Trans. Biomed. Eng.* **60** 1660–6
Behar J *et al* 2014e Evaluation of the fetal QT interval using non-invasive foetal ECG technology *Proc. of the Society for Maternal-Fetal Medicine—34th Annual Meeting—The Pregnancy Meeting (New Orleans, LA, 8 February 2014)* vol 210 pp S283–4
Belfort M A *et al* 2015 A randomized trial of intrapartum fetal ecg st-segment analysis *New Engl. J. Med.* **373** 632–41
Bernardes J, Costa-Pereira A, Ayres-de Campos D, Van Geijn H and Pereira-Leite L 1997 Evaluation of interobserver agreement of cardiotocograms *Int. J. Gynecol. Obstet.* **57** 33–7
Bhutta Z A, Yakoob M Y, Lawn J E, Rizvi A, Friberg I K, Weissman E, Buchmann E and Goldenberg R L 2011 Stillbirths: What difference can we make and at what cost? *Lancet* **377** 1523–38
Bloom S L *et al* 2006 Fetal pulse oximetry and cesarean delivery *New Engl. J. Med.* **355** 2195–202
Brambati B and Pardi G 1980 The intraventricular conduction time of fetal heart in uncomplicated pregnancies *Br. J. Obstet. Gynaecol.* **87** 941–8
California Department of Public Health 2014 Types of birth defects (www.ehib.org/page.jsp?page_key=459)
Camps G, Martinez M and Soria E 2001 Fetal ecg extraction using an fir neural network *Proc. of the Conf. on Computers in Cardiology IEEE* pp 249–52
Candil J J and Luengo C M 2008 QT interval and acute myocardial ischemia: past promises, new evidences *Rev. Esp. Cardiol.* **61** 561–63
Cerutti S, Baselli G, Civardi S, Ferrazzi E, Marconi A M, Pagani M and Pardi G 1986 Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings *J. Perinat. Med.* **14** 445–52
Chandraharan 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** 428–29
Christov I, Simova I and Abacherli R 2013 Cancellation of the maternal and extraction of the fetal ECG in noninvasive recordings *Proc. of Computers in Cardiology (Zaragoza, Spain)* vol 40
Cleveland Clinic 2015 Labor & Delivery (https://my.clevelandclinic.org)
Clifford G D, Behar J, Li Q and Rezek I 2012 Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms *Physiol. Meas.* **33** 1419–33
Clifford G D, Silva I, Behar J and Moody G B 2014d. Editorial: Non-invasive fetal ECG analysis *Physiol. Meas.* **35** 1521–36
Clifford G, 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** 47
Cremer M 1906 Über die Direkte Ableitung der Aktionstrome des Menschlichen Herzens vom Oesophagus und Über das Elektrokardiogramm des Fetus *Münchener Med. Wochenschrift* **53** 811–13
De Lathauwer L, De Moor B and Vandewalle J 2000 Fetal electrocardiogram extraction by blind source subspace separation *IEEE Trans. Biom. Eng.* **47** 567–72
De Moor B, De Gersem P, De Schutter B and Favoreel W 1997 Daisy: a database for identification of systems *Journal* A **38** 4–5
Dessì A, Pani D and Raffo L 2013 Identification of fetal QRS complexes in low density non-invasive biopotential recordings *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 321–24
Di Maria C, Duan W, Bojarnejad M, Pan F, King S, Zheng D, Murray A and Langley P 2013 An algorithm for the analysis of foetal ECG from 4-channel non-invasive abdominal recordings *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain)* vol 40
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** 1637–64
Downs T and Zlomke E 2007 Fetal heart rate pattern notification guidelines and suggested management algorithm for intrapartum electronic fetal heart rate monitoring *Permanente J.* **11** 22–8

East C E, Brennecke S P, King J F, Chan F Y and Colditz P B 2006 The effect of intrapartum fetal pulse oximetry, in the presence of a nonreassuring fetal heart rate pattern, on operative delivery rates: a multicenter, randomized, controlled trial (the foremost trial) *Am. J. Obstet. Gynecol.* **194** 606.e1

Fatemi M, Niknazar M and Sameni R 2013 A robust framework for noninvasive extraction of fetal electrocardiogram signals *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 201–4

Ghaffari A, Mollakazemi M J, Atyabi S A and Niknazar M 2015 Robust fetal qrs detection from noninvasive abdominal electrocardiogram based on channel selection and simultaneous multichannel processing *Australasian Physical & Engineering Sciences in Medicine* pp 1–12

Ghazdali A, Hakim A, Laghrib A, Mamouni N and Raghay S 2015 A new method for the extraction of fetal ecg from the dependent abdominal signals using blind source separation and adaptive noise cancellation techniques *Theor. Biol. Med. Modelling* **12** 1–20

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

Graatsma E M, Jacod B C, Van Egmond L A J, Mulder E J H and Visser G H A 2009 Fetal electrocardiography: feasibility of long-term fetal heart rate recordings *BJOG: Int. J. Obstet. Gynaecol.* **116** 334–8

Graham E M, Petersen S M, Christo D K and Fox H E 2006 Intrapartum electronic fetal heart rate monitoring and the prevention of perinatal brain injury *Obstet. Gynecol.* **108** 656–66

Greene K R, Dawes G S, Lilja H and Rosen K G 1982 Changes in the st waveform of the fetal lamb electrocardiogram with hypoxemia *Am. J. Obstet. Gynecol.* **144** 950–58

Haghpanahi M and Borkholder D A 2013 Fetal ECG extraction from abdominal recordings using array signal processing *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain)* vol 40

Hasan M A *et al* 2009 Detection and processing techniques of FECG signal for fetal monitoring *Biol. Proced. online* **11** 263–95

Hayes-Gill B, Hassan S, Mirza F G, Ommani S, Himsworth J, Solomon M, Brown R, Schifrin 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

Hökegård K H, Eriksson B, Kjellmer I, Magno R and Rosen K 1981 Myocardial metabolism in relation to electrocardiographic changes and cardiac function during graded hypoxia in the fetal lamb *Acta Physiol. Scand.* **113** 1–7

Johnson A E W, Behar J, Andreotti F, Clifford G D and Oster J 2015 Multimodal heart beat detection using signal quality indices *Physiol. Meas.* **36** 1665

Johnson A E W, Behar J, Clifford G D and Oster J 2016 R-peak estimation using multimodal lead switching *Physiol. Meas.* in preparation

Kanjilal P P, Palit S and Saha G 1997 Fetal ECG extraction from single-channel maternal ECG using singular value decomposition *Trans. Biomed. Eng.* **44** 51–9

Karlström A, Engström-Olofsson R, Norbergh K G, Sjöling M and Hildingsson I 2007 Postoperative pain after cesarean birth affects breastfeeding and infant care *J. Obstet. Gynecol. Neonatal Nursing* **36** 430–40

Kennedy R G 1998 Electronic fetal heart rate monitoring: retrospective reflections on a twentieth-century technology *J. R. Soc. Med.* **91** 244

Kligfield P *et al* 2007 Recommendations for the standardization and interpretation of the electrocardiogram. Part I: the electrocardiogram and its technology a scientific statement from the american heart association electrocardiography and arrhythmias committee, council on clinic *J. Am. Colloid Cardiol.* **49** 1109–27

Kovács F, Horváth C, Balogh Á T and Hosszú G 2011 Fetal phonocardiography—past and future possibilities *Comput. Methods Programs Biomed.* **104** 19–25

Kropf M, Schreier G, Modre-Osprian R and Hayn D 2013 A robust algorithm for fetal QRS detection using the non-invasive maternal abdomen ECG *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 313–16

Kuzilek J and Lhotska L 2013 Advanced signal processing techniques for fetal ECG analysis *Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 177–80

Lipponen J A and Tarvainen M P 2013 Advanced maternal ECG removal and noise reduction for application of fetal QRS detection *Proc. of the Computing in Cardiology Conf.* vol 40 pp 161–64

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
Liu C and Li P 2013 Systematic methods for fetal electrocardiographic analysis: determing the fetal heart rate, RR interval and QT interval *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain)* pp 309–12
Liu G and Luan Y 2015 An adaptive integrated algorithm for noninvasive fetal ecg separation and noise reduction based on ica-eemd-ws *Med. Biol. Eng. Comput.* **53** 1113–27
Llamedo M, Martín-Yebra A, Laguna P and Martínez J P 2013 Noninvasive fetal ECG estimation based on linear transformations *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain*, *22–25 September 2013)* vol 40 pp 285–8
Lukoševičius M and Marozas V 2013 A robust framework for noninvasive extraction of fetal electrocardiogram signal *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain)* vol 40
Maier C and Dickhaus H 2013 Fetal QRS detection and RR interval measurement in noninvasively registered abdominal ECGs *Proc. of the Computing in Cardiology Conf.* (*Zaragoza, Spain*, *22–25 September 2013*) vol 40 pp 165–8
Marco L Y D, Marzo A and Frangi A 2013 Multichannel foetal heartbeat detection by combining source cancellation with expectation-weighted estimation of fiducial points *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain*, *22–25 September 2013)* vol 40 pp 329–32
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
Martinek R, Kelnar M, Koudelka P, Vanus J, Bilik P, Janku P, Nazeran H and Zidek J 2015 Enhanced processing and analysis of multi-channel non-invasive abdominal foetal ECG signals during labor and delivery *Electron. Lett.* **51** 1744–6
Ma Y, Xiao Y, Wei G and Sun J 2015 A multichannel nonlinear adaptive noise canceller based on generalized flann for fetal ecg extraction *Meas. Sci. Technol.* **27** 015703
McSharry P E, Clifford G D, Tarassenko L and Smith L A 2003 A dynamical model for generating synthetic electrocardiogram signals *Trans. Biomed. Eng.* **50** 289–94
Moody G B, Moody B and Silva I 2014 Robust detection of heart beats in multimodal data: the PhysioNet/ Computing in Cardiology Challenge 2014 *Proc. of the Computing in Cardiology Conf.* vol 41
Neilson D R, Freeman R K and Mangan S 2008 Signal ambiguity resulting in unexpected outcome with external fetal heart rate monitoring *Am. J. Obstet. Gynecol.* **198** 717–24
Neilson J P *et al* 2006 Fetal electrocardiogram (ECG) for fetal monitoring during labour *Cochrane Database Syst. Rev.* **2006** CD000116
NHS 2014 Congenital heart disease (www.nhs.uk/conditions/Congenital-heart-disease/Pages/Introduction.aspx)
Niknazar M, Rivet B and Jutten C 2012 Fetal ecg extraction from a single sensor by a non-parametric modeling *2012 Proc. of the 20th European IEEE Signal Processing Conf.* pp 949–53
Niknazar M, Rivet B and Jutten C 2013a Fetal ECG extraction by extended state Kalman filtering based on single-channel recordings *Trans. Biomed. Eng.* **60** 1345–52
Niknazar M, Rivet B and Jutten C 2013b Fetal QRS complex detection based on three-way tensor decomposition *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain*, *22–25 September 2013)* vol 40 pp 185–88
Noorzadeh S, Rivet B and Guméry P Y 2015 A multi-modal approach using a non-parametric model to extract fetal ECG *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing* pp 832–6
Oldenburg J T and Macklin M 1977 Changes in the conduction of the fetal electrocardiogram to the maternal abdominal surface during gestation *Am. J. Obstet. Gynecol.* **129** 425–33
Oostendorp T, Van Oosterom A and Jongsma H 1989 Electrical properties of tissues involved in the conduction of foetal ecg *Med. Biol. Eng. Comput.* **27** 322–4
Oostendorp T 1989 Modelling the fetal ECG *PhD Thesis* Katholieke Universiteit te Nijmegen (https://books.google.co.uk/books?id=BPt5PAAACAAJ)
Oster J, Behar J, Colloca R, Li Q, Li Q and Clifford G D 2013 Open source Java-based ECG analysis software and Android app for atrial fibrillation screening *Proc. of the Computing in Cardiology Conf. IEEE (Zaragoza, Spain)* vol 40 pp 731–4
Oudijk M A, Kwee A, Visser G H, Blad S, Meijboom E J and Karl R G 2004 The effects of intrapartum hypoxia on the fetal QT interval *Br. J. Obstet. Gynaecol.* **111** 656–60
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
Philip M B and Angela L T 1992 Increased risk of postnatal depression after emergency *Med. J. Aust.* **157** 172–4

Pieri J F, Crowe J A, Hayes-Gill B R, Spencer C J, Bhogal K and James D K 2001 Compact long-term recorder for the transabdominal foetal and maternal electrocardiogram *Med. Biol. Eng. Comput.* **39** 118–25

Rabotti C, Mischi M, van Laar J O E H, Oei G S and Bergmans J W M 2008 Estimation of internal uterine pressure by joint amplitude and frequency analysis of electrohysterographic signals *Physiol. Meas.* **29** 829–41

Reinhard J, Hayes-Gill B R, Schiermeier S, Hatzmann W, Herrmann E, Heinrich T M and Louwen F 2012 Intrapartum signal quality with external fetal heart rate monitoring: a two way trial of external Doppler CTG ultrasound and the abdominal fetal electrocardiogram. *Arch. Gynecol. Obstet.* **286** 1103–7

Reinhard J, Hayes-Gill B R, Yi Q, Hatzmann H and Schiermeier S 2010 Comparison of non-invasive fetal electrocardiogram to Doppler cardiotocogram during the 1st stage of labor *J. Perinat. Med.* **38** 179–85

Rodrigues R 2013 Fetal ECG detection in abdominal recordings: a method for QRS Location *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain)* vol 40

Rooijakkers M, Song S and Rabotti C 2014 Influence of electrode placement on signal quality for ambulatory pregnancy monitoring *Comput. Math. Methods Med.* **2014** 960980

Rooth G, Huch A and Huch R 1987 Figo news: guidelines for the use of fetal monitoring *Int. J. Gynecol. Obstet.* **25** 159–67

Rosen K, Dagbjartsson A, Henriksson B, Lagercrantz H and Kjellmer I 1984 The relationship between circulating catecholamines and st waveform in the fetal lamb electrocardiogram during hypoxia *Am. J. Obstet. Gynecol.* **149** 190–5

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 Electrophysiol. Ther. J.* **3** 4–20

Sameni R, Clifford G D, Jutten C and Shamsollahi M B 2007a Multi-channel ECG and noise modeling: application to maternal and fetal ECG signals *EURASIP J. Adv. Signal Process.* **2007** 94

Sameni R, Clifford G, Ward J, Robertson J, Pettigrew C and Wolfberg A 2009 Accuracy of fetal heart rate acquired from sensors on the maternal abdomen compared to a fetal scalp electrode *Am. J. Obstet. Gynecol.* **201** S241

Sameni R, Jutten C and Shamsollahi M B 2008 Multichannel electrocardiogram decomposition using periodic component analysis *Trans. Biomed. Eng.* **55** 1935–40

Sameni R, Shamsollahi M B, Jutten C and Clifford G D 2007b A nonlinear Bayesian filtering framework for ECG denoising *Trans. Biomed. Eng.* **54** 2172–85

Sasaki Y 2007 The truth of the f-measure Teaching, Tutorial materials, Version: 26th October

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 *Proc. of the Computing in Cardiology Conf. IEEE (Zaragoza, Spain)* vol 40 pp 149–52

Silva I and Moody G 2014 An open-source toolbox for analysing and processing physionet databases in matlab and octave *J. Open Res. Softw.* **2** e27

Starc V 2013 Noninvasive fetal ECG detection from maternal abdominal recording: the Physionet/CINC Challenge *Proc. of the Computing in Cardiology Conf. (Zaragoza, Spain, 22–25 September 2013)* vol 40 pp 317–20

Stinstra J *et al* 2002 Multicentre study of fetal cardiac time intervals using magnetocardiography *Br. J. Obstet. Gynaecol.* **109** 1235–43

Stinstra J G 2001 The reliability of the fetal magnetocardiogram *PhD Thesis* Universiteit Twente (http://doc.utwente.nl/35964/1/t000001d.pdf)

Sweha A and Hacker T W 1999 Interpretation of the electronic fetal heart rate during labor *Am. Family Phys.* **59** 2487–500

Symonds E M, Sahota D and Chang A 2001 *Fetal Electrocardiography* (Singapore: World Scientific)

Taylor M J, Smith M J, Thomas M, Green A R, Cheng F, Oseku-Afful S, Wee L Y, Fisk N M and Gardiner H M 2003 Non-invasive fetal electrocardiography in singleton and multiple pregnancies *BJOG: Int. J. Obstet. Gynaecol.* **110** 668–78

Taylor M J, Thomas M J, Smith M J, Oseku-Afful S, Fisk N M, Green A R, Paterson-Brown S and Gardiner H M 2005 Non-invasive intrapartum fetal ECG: preliminary report *Br. J. Obstet. Gynaecol.* **112** 1016–21

Ungureanu M, Bergmans J W M, Oei S G and Strungaru R 2007 Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique *Biomed. Tech.* **52** 56–60
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
von Steinburg S P, Boulesteix A, Lederer C, Grunow S, Schiermeier S, Hatzmann W, Schneider K M and Daumer M 2012 What is the 'normal' fetal heart rate? *PeerJ* **1** e82
Vullings R, Mischi M, Oei S G and Bergmans J W 2013 Novel Bayesian vectorcardiographic loop alignment for improved monitoring of ECG and fetal movement *IEEE Trans. Biomed. Eng.* **60** 1580–8
Widmark C, Jansson T, Lindecrantz K and Rosén K 1991 ECG waveform, short term heart rate variability and plasma catecholamine concentrations in response to hypoxia in intrauterine growth retarded guinea-pig fetuses *J. Dev. Physiol.* **15** 161–8
Widrow B, Glover J R, McCool J M, Kaunitz J, Williams C S, Hearn R H, Zeidler J R, Dong E and Goodlin R C 1975 Adaptive noise cancelling: principles and applications *Proc. IEEE* **63** 1692–716
Wolfberg A J, Clifford G D and Ward J 2011 Fetal ECG monitoring *US Patent* 2012/0083676 Al
Zarzoso V and Nandi A K 2001 Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation *Trans. Biomed. Eng.* **48** 12–8
Zaunseder S, Andreotti F, Cruz M, Stepan H, Schmieder C, Malberg H, Jank A and Wessel N 2012 Fetal qrs detection by means of Kalman filtering and using the event synchronous canceller *Int. J. Bioelectromagnetism.* 83–9
Zhu T, Behar J, Papastylianou T and Clifford G D 2014 CrowdLabel: a platform for crowd-sourcing accurate foetal QT labels *Proc. of the Computing in Cardiology Conf. IEEE (Cambridge, MA, USA)* vol 41
Zhu T, Dunkley N, Behar J, Clifton D and Clifford G 2015 Fusing continuous-valued medical labels using a Bayesian model *Ann. Biomed. Eng.* **43** 2892–902
Zhu T, Johnson A E W, Behar J and Clifford G D 2013 Crowd-sourced annotation of ECG signals using contextual information *Ann. Biomed. Eng.* **42** 871–84