# Artificial Intelligence for Noninvasive Fetal Electrocardiogram Analysis

# Artificial Intelligence for Noninvasive Fetal Electrocardiogram Analysis

Eleni Fotiadou

geboren te Drama, Griekenland

ISBN: 978-90-386-5281-8

Print & Design: Ridderprint | www.ridderprint.nl

# Artificial Intelligence for Noninvasive Fetal Electrocardiogram Analysis

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens,

voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op maandag 14 juni 2021 om 16:00 uur

door

Eleni Fotiadou

geboren te Drama, Griekenland

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

| | |
|---|---|
| **Voorzitter:** | prof.dr.ir. P.G.M.Baltus |
| **1e promotor:** | prof.dr.ir. M. Mischi |
| **Copromotoren:** | dr.ir. R. Vullings<br>dr. J.O.E.H. van Laar |
| **Leden:** | prof.dr. S.G. Oei<br>prof.dr.ir. E. Cantatore<br>prof.dr. T.M. Heskes (Radboud University)<br>prof.dr.ir. P.M. Djuric (Stony Brook University)<br>prof.dr. J. Hesser (Heidelberg University) |

Het onderzoek of ontwerp dat in dit thesis wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

## SUMMARY

### Artificial intelligence for noninvasive fetal electrocardiogram analysis

Most women experience uncomplicated pregnancies. However, some women have high-risk pregnancies, where the fetus or the newborn has an increased risk of experiencing an adverse health condition. These health conditions include delayed fetal growth, preterm birth, birth injuries and can even lead to infant mortality. Many of these conditions are preventable and can be managed with timely clinical intervention. For early detection of pregnancy- and labor-related complications, there is a need for effective fetal monitoring techniques to provide information about the fetal condition and contribute to an improved perinatal outcome.

The most common approach for fetal monitoring in clinical practice is cardiotocography (CTG), which monitors changes in the fetal heart rate (HR) in response to uterine contractions. CTG monitoring of the fetal HR is commonly based on either a Doppler-based ultrasound transducer or an invasive fetal scalp electrode. The Doppler-ultrasound technique, which can be performed both during pregnancy and labor, is unobtrusive but it frequently suffers from signal loss. The scalp electrode provides more reliable measurements, but its applicability is limited to the intrapartum period after the completion of 32 weeks of gestation. In general, the specificity of CTG is low, meaning that an abnormal CTG may also be encountered for healthy fetuses, leading to unnecessary interventions. Moreover, the perinatal outcome has not been improved since CTG was introduced in clinical practice. Techniques like fetal blood sampling and ST waveform analysis (STAN) used in conjunction with CTG did not manage to decrease neonatal morbidity and mortality either. Hence, additional information about the condition of the fetus is required.

Noninvasive fetal electrocardiography, based on multiple electrodes positioned on the maternal abdomen, may be a valuable alternative for fetal monitoring, allowing for the extraction of the fetal HR and uterine contractions similar to conventional CTG. However, unlike Doppler ultrasound, fetal electrocardiography can provide accurate, beat-to-beat, HR information, which is necessary for reliable analysis of the fetal HR variability, an important indicator of fetal distress. Moreover, other than Doppler ultrasound, noninvasive fetal electrocardiography offers the possibility to perform morphological analysis of the fetal electrocardiogram (ECG). Most

cardiac defects have some manifestation in the morphology of the ECG signals, while ECG waveform changes may indicate fetal hypoxia. Morphological analysis of the ECG waveform is currently performed through the invasive STAN method. However, making use of a scalp electrode, this method is merely employed during labor. Moreover, the use of a single lead for the ECG acquisition is being criticized. In contrast, abdominal measurements of the fetal ECG can be performed during pregnancy employing multiple electrodes, overcoming the main shortcomings of the STAN method.

Despite the numerous advantages of the noninvasive fetal electrocardiography, its use in clinical practice is limited and yet does not incorporate morphological ECG analysis. The reason is the low signal-to-noise ratio (SNR) of the obtained fetal ECG signal, which is obscured by a mixture of maternal ECG and various other noises and interferences. Even after accurate maternal ECG removal, the residual noise in the fetal ECG may hamper the reliable extraction of the ECG morphological characteristics and, in some cases, even the QRS complex locations, necessary for the extraction of the fetal HR. In this thesis, we aim to advance the field of noninvasive fetal electrocardiography by improving the fetal HR extraction process to achieve more reliable CTG monitoring and by enhancing the quality of the extracted fetal ECG signals to enable morphological ECG analysis.

First, a method is proposed for reliable extraction of the fetal HR. Fetal R-peak detection in noninvasive fetal ECG recordings is challenging due to the low SNR and the non-stationarity of the fetal ECG signals. A deep learning model was developed that directly estimates the fetal HR from the multichannel fetal ECG, without the need for R-peak detection. Deep learning models implicitly perform extraction of features, which have often proven to be more informative than expert features. The proposed model combines convolutional neural networks with long short-term memory networks to capture both short-term as well as long-term temporal HR patterns. The robustness of the algorithm is increased by a classifier that estimates the reliability of the HR estimation. The proposed method, which is invariant both to changes in the fetal orientation and to electrode position, outperforms the best methods proposed in the literature.

In the second part of the thesis, signal processing methods for enhancing the quality of the fetal ECG signals are proposed. Several complex physiological and instrumental noises with spectral overlap with the fetal ECG remain even after the

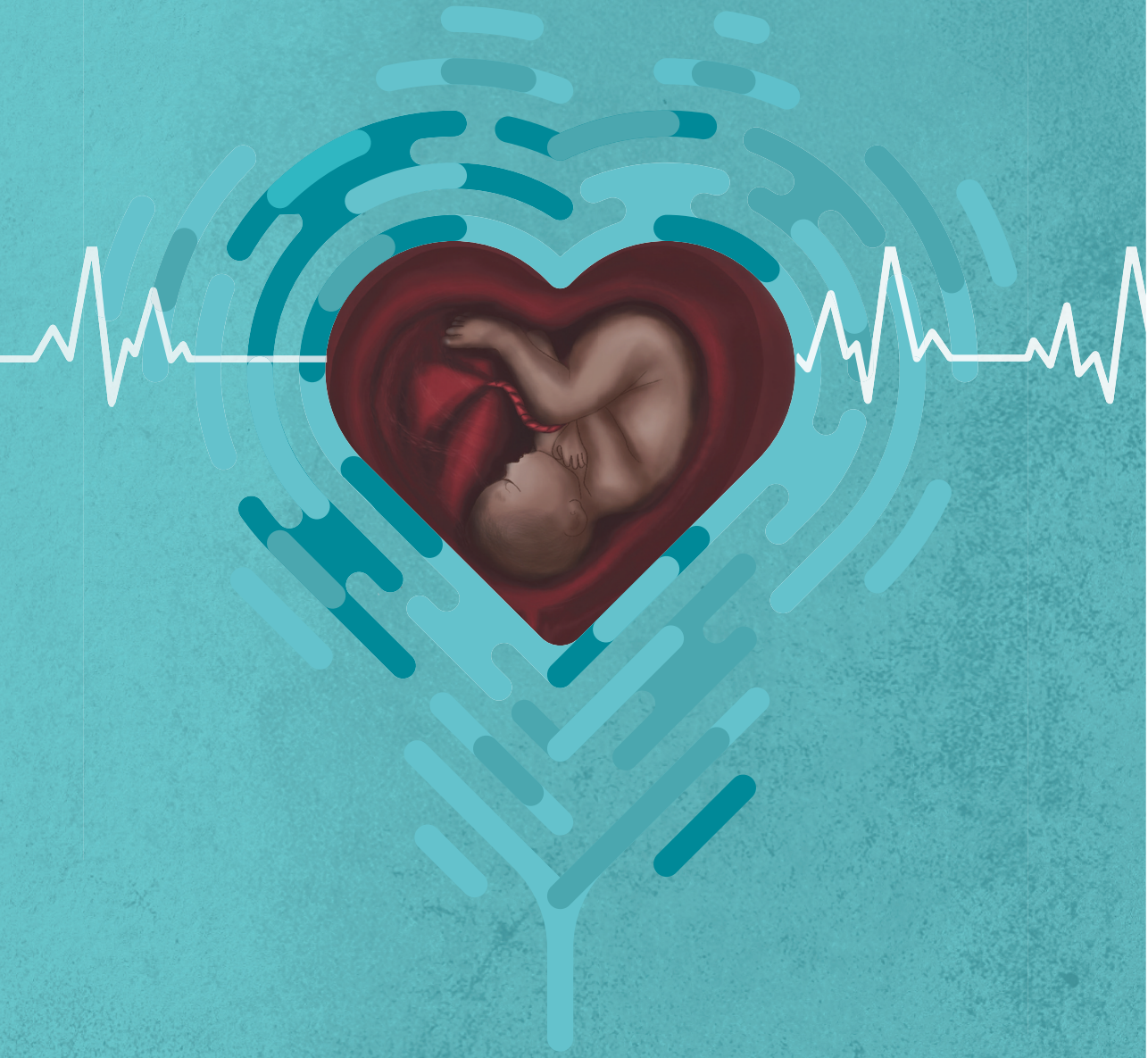
application of powerful fetal ECG extraction algorithms. The low SNR of the fetal ECG hinders the morphological analysis of its waveform in clinical practice. To address this issue, initially the use of a time-sequenced adaptive filter is proposed that significantly improves the quality of the multichannel fetal ECG recordings, revealing the morphology of the signals. To overcome the limitations of this filter, such as requirement for accurate estimates of R-peak locations prior to use and inability of handling abrupt changes in ECG morphology, a second alternative denoising method is developed that is based on deep learning. An encoder-decoder deep convolutional network is trained to remove the residual noise by capturing the structure of the fetal ECG in the encoder of the network and recovering the signal details through the decoder. The method is tuned and assessed both for single channel and multichannel fetal ECG signals. In both cases, experiments on both synthetic and real data show that the network can achieve a substantial quality improvement of the noisy signals. Clearly, the multichannel network delivers cleaner and more reliable fetal ECG signals.

In combination with current fetal monitoring techniques, noninvasive fetal electrocardiography could support clinical decisions. The work presented in this thesis may open the way for morphological analysis of the fetal ECG signals and contribute to the spread of this technique in clinical practice.

## Contents

## ABBREVIATIONS

| | |
|---|---|
| aNLMS | Augmented normalized least mean squares |
| aTSAF | Augmented time sequenced adaptive filter |
| AUC | Area under the ROC curve |
| AV | Atrioventricular (node) |
| CHD | Congenital heart defect |
| CTG | Cardiotocography |
| CNN | Convolutional neural network |
| DAE | Denoising autoencoder |
| DNN | Deep neural network |
| ECG | Electrocardiogram |
| EHG | Electrohysterogram |
| EMG | Electromyogram |
| FBS | Fetal blood sampling |
| FECGSYNDB | Fetal electrocardiogram synthetic database |
| HR | Heart rate |
| ICA | Independent component analysis |
| LMS | Least mean squares |
| LSTM | Long short-term memory |
| MAE | Mean absolute error |
| MCG | Magnetocardiogram |
| MSE | Mean squared error |
| NLMS | Normalized least mean squares |
| NMSE | Normalized mean squared error |
| PCG | Phonocardiography |
| PLI | Powerline interference |
| PReLU | Parametric rectified linear unit |
| ReLU | Rectified linear unit |
| RMSE | Root mean squared error |
| RNN | Recurrent neural network |
| SA | Sinoatrial (node) |
| SNR | Signal to noise ratio |
| STAN | ST waveform analysis |
| TSAF | Time sequenced adaptive filter |
| UC | Uterine contraction |
| VCG | Vectorcardiogram |



# CHAPTER 1

# INTRODUCTION

## 1.1 Motivation

Annually, an estimated 2.5 million neonatal deaths and 2.6 million stillbirths occur worldwide [1]. Half of all stillbirths occur intrapartum, while most of them result from preventable conditions [2]. One of the leading causes for obstetric complications is oxygen deficiency during labor. When oxygen deficiency lasts long, it can cause irreversible damage to the central organs of the fetus, such as the brain, a condition known as perinatal asphyxia [3]. Fetal surveillance is crucial to allow for timely intervention and prevent the damage of the fetal vital organs. However, even though obstetrical monitoring has advanced during the last years, perinatal morbidity and mortality rates still remain high. Thus, there is a need for improved fetal monitoring techniques that can provide information of the fetal health status during both pregnancy and labor. In addition, congenital heart defects (CHDs) are the most common types of congenital malformations and a leading cause of death due to fetal defects [4]. CHDs affect approximately 1 in 100 live births [5] and cause death to almost 300.000 babies annually during pregnancy or infancy [6]. Timely detection of CHD can lead to reduced mortality and morbidity but is hampered by the low detection rates of the current screening technology [6]. Therefore, there is an urgent need for additional technologies that could provide supplementary information about the cardiac health status of the fetus and improve the prenatal CHD detection rate.

Electronic fetal monitoring was introduced in clinical practice in the 1960s through cardiotocography (CTG), which is a simultaneous recording of the fetal heart rate (HR) and uterine contractions (UCs). CTG was developed as a way to assess fetal distress since it enables the evaluation of the fetal response to stress caused by uterine contractions [7]. An example of CTG recording is shown in Fig. 1.1. With the introduction of the CTG, perinatal asphyxia was expected to be successfully detected. As perinatal asphyxia is considered the major cause of cerebral palsy, the underlying expectation was to achieve a reduction in neonatal mortality and morbidity. Unfortunately, not only the incidence of cerebral palsy has not decreased but also there has been a significant increase in the incidence of unnecessary operative deliveries [8]. Although CTG is the current standard for fetal monitoring, its diagnostic value is limited [8]. One reason for this is that CTG patterns are interpreted visually, while at the same time the inter- and intra-observer variability is high. The fourth annual Confidential Enquiry into Stillbirths and Deaths in Infancy (CESDI) reported that the CTG interpretation was questioned in more than half of all perinatal deaths [9].

Moreover, the CTG has good sensitivity but very poor specificity, meaning that from the fetuses with abnormal CTG only 40-60% present intrapartum hypoxia, i.e. oxygen deficiency that affects the peripheral tissues [10].

In view of the poor specificity of the CTG, additional diagnostic tools are necessary to assess the fetal well-being. The most important tools complementary to CTG are fetal blood sampling (FBS) [11] and analysis of the fetal electrocardiogram (ECG) waveform [12]. During FBS, a small amount of blood is taken from the fetus and its pH is identified. However, there are risks associated with FBS such as infection [14]. Moreover, since FBS provides instantaneous information only, the procedure needs several repetitions in case the CTG remains abnormal. Lastly, systematic reviews have demonstrated no benefit of using fetal blood sampling for reducing unnecessary interventions or influencing long-term outcomes [15].

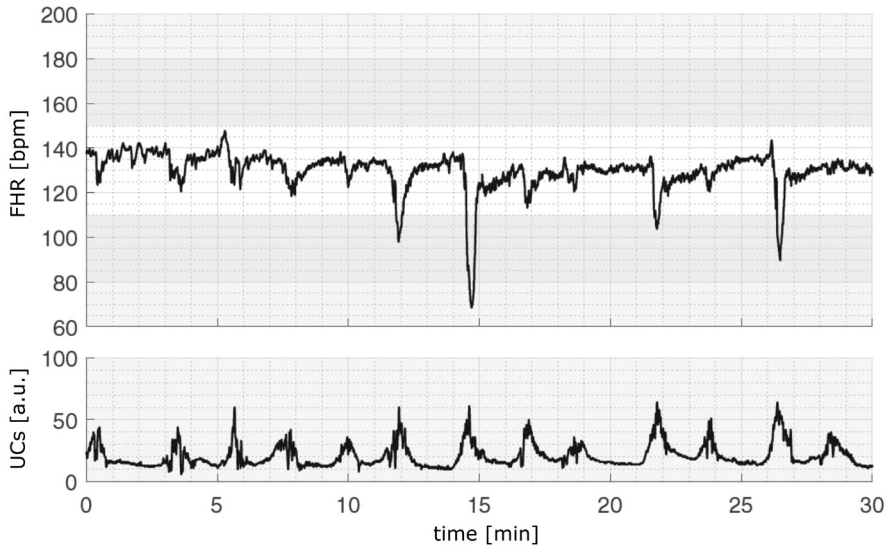



Fig. 1.1 Example of a CTG, a simultaneous recording of the fetal heart rate (FHR) and the uterine contractions (UCs) [13].

Another method used in conjunction with CTG is ST waveform analysis (STAN). STAN analyzes changes in the ST segment and the T/QRS ratio of the fetal ECG that occur secondary to fetal hypoxia. STAN recordings are performed through an invasive electrode attached to the fetal scalp (Fig. 1.2a). An example of a STAN recording is illustrated in Fig. 1.2b. Even though initial studies showed promising results in improving perinatal outcome with the use of STAN [16], [17], subsequent studies could not reproduce the original findings [18], [19], [20], [21]. Moreover, the

use of STAN is limited in several ways. First, since STAN uses a scalp electrode, it can only be used during labor after rupture of the membranes and sufficient cervical dilation. Second, STAN alarms should be ignored when CTG is reassuring [22]. Because of the high intra-observer variability of the CTG interpretation, this limits the success of the STAN method. Finally, ST analysis is performed on a single-lead scalp ECG that might not provide the optimal ECG derivation, possibly diminishing the performance of the waveform analysis [23], [24], [25]. From the above it is clear that both FBS and STAN have poor diagnostic value and several limitations. Therefore, there is a need for a non-invasive, safe, and long-term monitoring method that provides more reliable information about the fetal condition and that can be used during both pregnancy and labor, as well as in patients with elevated risk of preterm labor.

Besides the poor specificity of CTG, its use in clinical practice is also limited for another reason. The fetal HR is the main source of information from which the physiological condition of the fetus can be assessed. Most CTG monitors use Doppler ultrasound technology to determine the fetal HR [27]. This technology is noninvasive, but it uses autocorrelation techniques to determine the fetal HR, leading to an HR that is averaged over several heartbeats [28]. This limits the computation of the HR variability that ideally requires beat-to-beat HR information. Because the autonomic nervous system regulates the HR, the presence of HR variations indicates autonomic regulation and can provide important information about fetal distress [29], [30], [31]. Recently, there has been strong interest in computerized HR variability analysis that can quantify beat-to-beat HR variations [29], [30], [32], [33]. There are several other technical limitations associated with the use of Doppler CTG, one of them being the frequent incidence of periods of signal loss [34]. Moreover, Doppler CTG measurements are especially inaccurate in cases of obese patients, premature deliveries and during second stage of labor. Another disadvantage of the Doppler CTG is in that it is sensitive to maternal and fetal movements, requiring frequent repositioning of the ultrasound probe. Last but not least, although long-term fetal HR monitoring is often recommended, there is no strong proof that exposure to long-term ultrasound radiation is completely harmless for the fetus [35].

An alternative way to determine the fetal HR is through measuring the electrical activity of the heart (i.e. the fetal ECG). The electrical activity of the heart can be assessed either by electrodes placed on the maternal abdomen or the fetal scalp. As mentioned before, fetal scalp measurements are invasive and can only be performed during labor after the membranes have ruptured. On the contrary, recording the

fetal ECG through electrodes on the maternal abdomen (Fig. 1.3) is an unobtrusive technique and can therefore be applied in all stages of pregnancy. Fetal ECG recordings offer the possibility not only to extract beat-to-beat HR information, overcoming the main limitation of Doppler ultrasound, but also to assess the fetal ECG signal morphology. The depolarization and repolarization properties of the heart are reflected in the fetal ECG morphology. It has been reported in the literature that fetal hypoxia is reflected by changes in the fetal ECG waveform [36], [37]. In fact, this was the reason why the STAN technology emerged. In contrast to STAN, noninvasive fetal ECG can be measured also during pregnancy. Moreover, as noninvasive fetal ECG provides multilead measurements, it may overcome some of the current shortcomings in ST waveform analysis [38], [39]. In addition, most cardiac defects have some manifestation in the morphology of the ECG signals [4], [40], [41].

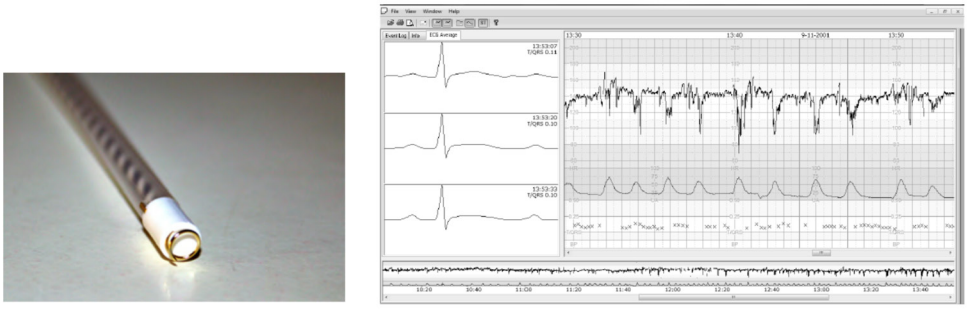
Through abdominal recordings, apart from measuring the fetal HR, it is also possible to retrieve the maternal HR and the uterine activity [42]. Therefore, noninvasive fetal ECG recordings could be used as an alternative or complementary solution to current fetal monitoring. Unfortunately, the noninvasiveness of the fetal ECG comes at the cost of low signal-to-noise ratio (SNR). The fetal ECG is mixed with various noise sources that complicate its extraction from the signal mixture. The difficulty to extract a noise-free fetal ECG signal from the abdominal measurements explains why the application of abdominal fetal ECG recordings in clinical practice is still limited.

## 1.2 Goals of this study

From the previous section, it is clear that noninvasive fetal ECG analysis can provide additional information regarding the fetal condition. The main shortcoming of this technique and the reason of its limited application in clinical practice is the low SNR of the recorded fetal ECG. This is due to the fact that the fetal ECG signal is generated from a small source (fetal heart) and has to propagate through several attenuating media to reach the maternal abdominal surface; moreover, various interferences and noises are additionally recorded by the abdominal electrodes. This results in recorded fetal ECG signals with an amplitude of about 10μV [27] that largely overlap in frequency and time domain with the maternal ECG and other interferences and noises. Even after accurate maternal ECG removal and the application of advanced signal processing techniques for fetal ECG extraction [4], [43], the residual noise

in the fetal ECG may impede reliable ECG waveform analysis. In some cases, the amount of noise is such that even the extraction of the QRS complex locations, which is required for the computation of the fetal HR and fetal HR variability, becomes complicated.

This thesis has two objectives: i) to improve the fetal HR extraction process from noninvasive fetal ECG recordings and achieve more reliable beat-to-beat HR information and ii) to enhance the SNR of the fetal ECG signals that are extracted from the mixture of abdominal signals such as to enable, in the future, the morphological analysis of the fetal ECG waveform. Special attention is dedicated to investigating different postprocessing methods of the extracted fetal ECG signals for improving their quality, to advance the field of noninvasive fetal electrocardiography and stimulate research in this area, with the end goal of enabling fetal ECG waveform analysis in clinical practice.



(a) (b)

Fig. 1.2 (a) Scalp electrode [26]. (b) Example of a STAN recording. In the right panel, the fetal HR is shown as the top graph, the uterine activity as the second graph from above and the results of the ST analysis are presented with the crosses in the bottom graph. In the left panel, three fetal ECG complexes, each one corresponding to a single cross are shown [23].

## 1.3 Thesis outline

Chapter 2 gives an overview of the physiological and technical background that is relevant for this thesis. The background provides a description of the fetal heart anatomy and development and of the fetal ECG and its clinical relevance. Moreover, several fetal monitoring techniques are explained with a focus on noninvasive

fetal electrocardiography. After the background chapter, the first part of the thesis describes a technique developed for reliable fetal HR detection from noninvasive fetal ECG recordings. The low SNR and the non-stationarity of the noninvasive fetal ECG signals make QRS detection, which is typically required for HR extraction, a very demanding task. In Chapter 3, a new method is proposed that instead of performing QRS detection, directly determines the fetal HR from the extracted fetal ECG signals. To this end, a deep learning technique is employed that combines convolutional neural networks with recurrent neural networks.


Fig. 1.3 Noninvasive fetal ECG recording (Photo by Bart van Overbeeke) [23]. Electrodes on the maternal abdomen measure a signal mixture, from which the fetal ECG should be extracted.

The noninvasive fetal ECG is often severely contaminated by a considerable amount of noise sources. Even after the application of advanced signal processing techniques for fetal ECG extraction, the signal quality remains usually very low. The second

part of this thesis focuses on the development of postprocessing methods for the enhancement of the fetal ECG obtained from noninvasive abdominal recordings. Chapter 4 proposes an improvement of the time-sequenced adaptive filter for enhancing the quality of multichannel fetal ECG. The improved filter is demonstrated in this chapter to be effective in reducing major noise components and revealing the ECG waveform. However, accurate location of fetal ECG complexes is required to synchronize the filter, and the method cannot handle abrupt changes in fetal ECG morphology, e.g. in cases of arrhythmia. To address these shortcomings, a novel fetal ECG denoising method, based on deep learning, is proposed in Chapter 5. The method uses a deep fully convolutional encoder-decoder network that learns an end-to-end mapping from noise-contaminated single channel fetal ECG signals to clean ones. An adaptation of the method to exploit the spatiotemporal patterns existing in multichannel signals is outlined in Chapter 6. This method can achieve a substantial improvement in the quality of the noisy signals, while being able to preserve beat-to-beat morphological variations without requiring any prior knowledge of the locations of the ECG complexes and the noise spectra.

In the last chapter of this thesis (Chapter 7), the main findings of this work are summarized and directions for future research are discussed. Chapters 3-6 are studies that have been published in international peer-reviewed journals. Each of these chapters is written to be self-contained, causing some overlapping among them.

## 1.4 List of publications

### Journal Papers

**JP-**1 Fotiadou E., van Laar J.O.E.H, Oei S.G., Vullings R., Enhancement of low-quality fetal electrocardiogram based on time-sequenced adaptive filtering. *Med. Biol. Eng. Comput.*, 2018, 56, 2313-2323. doi:10.1007/s11517-018-1862-8

**JP-**2 Fotiadou E, Konopczyński T, Hesser J, Vullings R., End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising. *Physiol. Meas.*,2020, 41:15005. doi: 10.1088/1361-6579/ab69b9

**JP-**3 Fotiadou E., Vullings R., Multi-Channel Fetal ECG Denoising with Deep Convolutional Neural Networks. *Front. Pediatr.*, 2020, 8, 508. doi:10.3389/fped.2020.00508

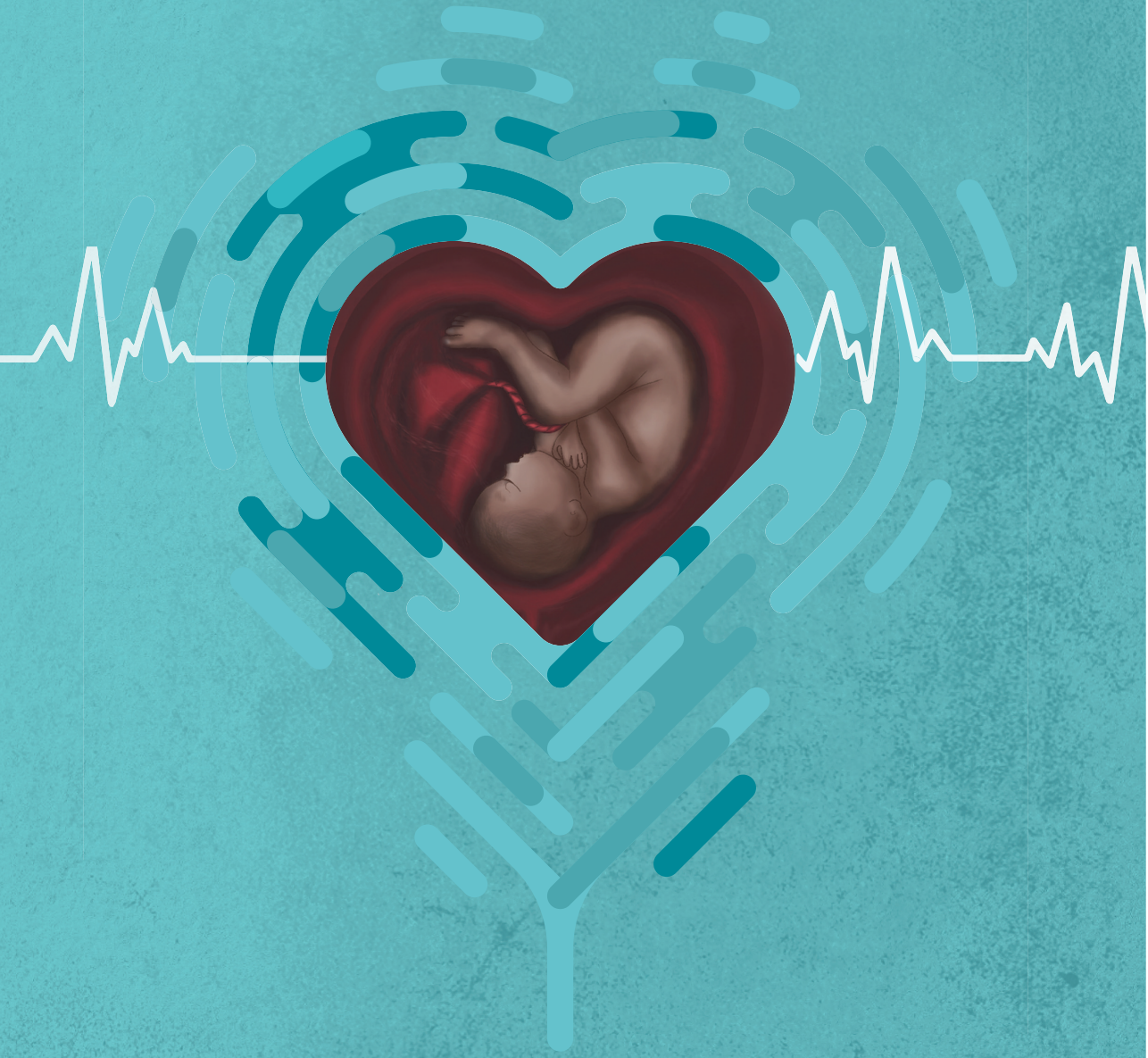
**JP-**4 Fotiadou E., van Sloun R.J.G., van Laar J.O.E.H, Vullings R., A Dilated Inception CNN-LSTM Network for Fetal Heart Rate Estimation. *Physiol. Meas.,* 2021, doi: 10.1088/1361-6579/abf7db



## Conference Proceedings

**CP-**1 Fotiadou E., Xu M., van Erp B., van Sloun R. J. G., Vullings R., Deep Convolutional Long Short-Term Memory Network for Fetal Heart Rate Extraction,2020 42*nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, QC, Canada, 2020, pp. 1-4, doi: 10.1109/EMBC44109.2020.9175442.

**CP-**2 Fotiadou E., Konopczyński T., Hesser J., Vullings R., Deep Convolutional Encoder-Decoder Framework for Fetal ECG Signal Denoising, 2019 *Computing in Cardiology (CinC)*, Singapore, Singapore, 2019, pp. 1-4, doi: 10.23919/CinC49843.2019.9005722.



# CHAPTER 2

# BACKGROUND

## 2.1 Physiology of the fetal heart

### 2.1.1 Fetal cardiac development

The human heart is one of the first organs to form and function in the fetus [44]. However, several weeks are needed until it takes the shape of a four-chambered heart (Fig. 2.1). Around 18 to 19 days after fertilization, the heart develops from a tissue known as mesoderm, near the head of the embryo, in a region called cardiogenic area [45]. Initially, the cardiogenic area has the appearance of two endocardial tubes, which are then fused to form a primitive heart tube with five distinct regions (Fig. 2.1). The primitive heart tube develops and reshapes to a structure with four champers at day 28. The heart valves are then formed between week five and eight. The fetal heart starts to beat around day 22, only 3 weeks after fertilization [44].

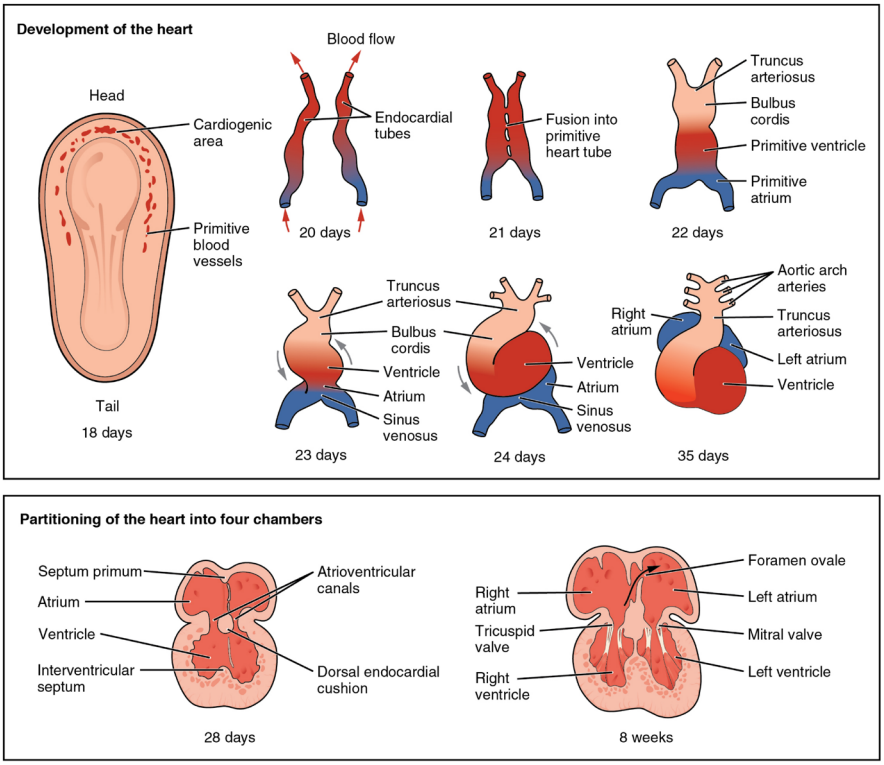



Fig. 2.1 Fetal heart development stages [46]. In the top subfigure the development of the fetal heart during the first eight weeks is displayed and in the bottom subfigure the formation of the heart chambers. Blue and red color represent blood inflow and outflow.

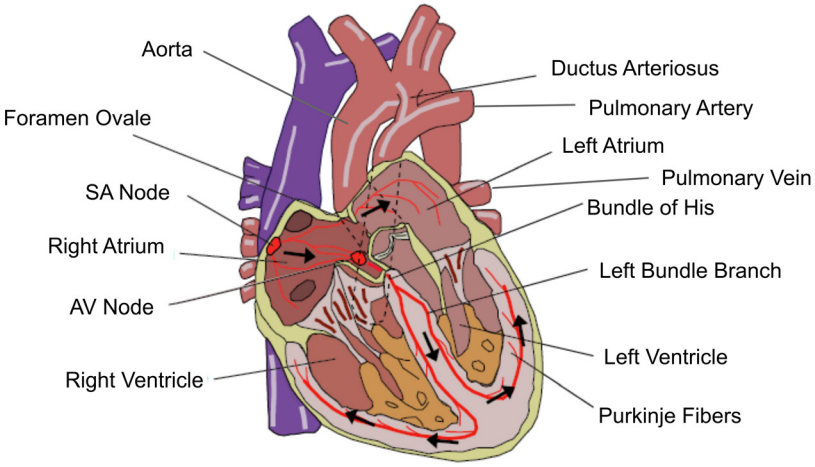



Fig. 2.2 Anatomy of the fetal heart. Image adapted from [47].

### 2.1.2 Fetal heart anatomy

The heart is a muscular organ whose primary role is to pump blood and distribute oxygen and nutrients through the whole body. The heart has four champers: the left and right atria at the top, and the left and right ventricles at the bottom (Fig. 2.2). A wall, called septum, divides the heart into right and left sides. In adults, the right side pumps deoxygenated blood to the lungs to receive oxygen, while the left side pumps oxygenated blood to the rest of the body through the main cardiac artery called aorta [8]. In the adult circulation, deoxygenated blood reaches the heart through the right atrium, it fills the right ventricle through the tricuspid valve, and reaches the lungs through the pulmonary artery. In the lungs, gas exchange occurs, and the deoxygenated blood is replenished with oxygen. Afterwards, the oxygenated blood enters the heart through the pulmonary veins in the left atrium and passes through the mitral valve to the left ventricle, which will pump it again to the aorta and the systemic circulation reaching the whole body. The blood delivers oxygen and nutrients to the organs and tissues and then returns to the heart for the next round of the adult circulation.

The anatomy of the fetal heart is illustrated in Fig. 2.2. The fetal heart exhibits some functional differences with respect to the adult heart that originate from the fact that the fetal circulation differs from the adult one [48], [49], [50]. Fetuses receive oxygen and nutrients from the mother via the placenta and the umbilical cord [50]. The waste

products produced by the fetuses are carried back to the placenta and end up in the maternal circulation to be cleared. Since the lungs of the fetus are not yet functional, blood is pumped into them much less compared to adults. Instead, both ventricles pump together most of the blood through the systemic circulation [50]. Two shunts in the fetal heart play an important role with shunting blood away from the lungs: the foramen ovale and the ductus arteriosus. The foramen ovale allows blood from the right atrium to enter the left atrium while the ductus arteriosus allows blood from the pulmonary artery to enter the aorta. There is also a third shunt, the ductus venosus, which shunts a portion of umbilical vein blood flow directly to the inferior vena cava. Thus, it allows oxygenated blood from the placenta to bypass the liver.

At birth the umbilical cord is clamped and snipped, and the baby does no longer receive oxygen and nutrients from the mother. As soon as babies are born, they need to use their lungs to breathe and receive oxygen from the air. Soon after birth, the shunts close since they are no longer needed (in healthy babies).

A network of specialized muscle cells is found in the cardiac walls that transmits signals to the rest of the heart causing it to contract. This group of cells is called the cardiac conduction system. The main parts of this system are the sinoatrial node (SA node), atrioventricular node (AV node), bundle of His, bundle branches and Purkinje fibers (Fig. 2.2). The SA node, which is situated in the walls of the right atrium, is known as the heart's pacemaker because it generates impulses (action potentials) and initiates the contraction of the cardiac muscles, thereby producing a heartbeat. The SA node contains the highest number of cells with the ability to produce a cardiac impulse by auto-depolarization. These cells also have the fastest depolarization rate than all the cardiac cells; therefore, due to the presence of a refractory period, the SA node controls the rhythm of the heart [51]. Other heart cells have also the ability to act as pacemakers but, normally, because the cells of the SA node produce impulses at a faster rate, these other cells are dominated by the SA node.

After the SA node initiates an electrical impulse, the impulse travels throughout the left and right atria and depolarizes the myocardial cells, causing the contraction of the atria. From the atria the electrical impulse travels to the AV node and to the ventricles via the bundle of His. The bundle of His splits into two pathways, known as right and left bundle branch, to stimulate the right and left ventricles. However, before the impulse moves to the bundle of His, it is delayed strategically to allow the atria to empty their contents into the ventricles before starting the ventricular

contraction. The Purkinje fibers, which are situated on the ventricular walls, allow the ventricles to contract simultaneously, and are, therefore, necessary for maintaining an effective cardiac contraction. Every cycle of atrial and ventricular contractions represents one heartbeat.

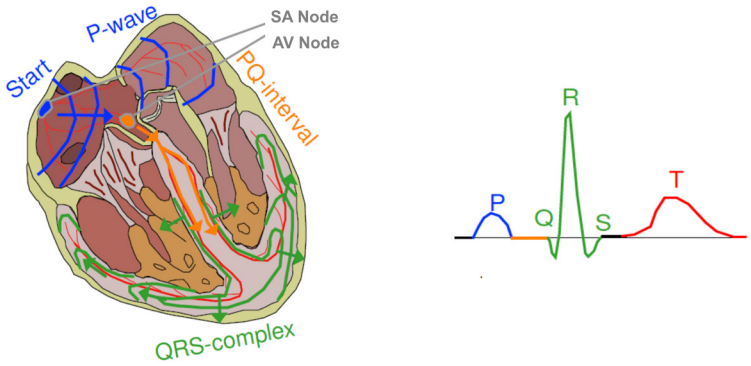



Fig. 2.3 Representation of the cardiac electrical activation sequence, resulting in the typical ECG waveform. Image adapted from [47].

### 2.1.3 The fetal electrocardiogram

At rest, the cells of the cardiac muscle are in polarized state, meaning that the outer surface of the cell membrane has a positive charge, while the inner surface is negatively charged [51], [52]. When the cells are stimulated externally, the polarity reverses and the intracellular potential becomes positive. This is referred to as depolarization. Once one cell is depolarized, it causes its neighboring cells to depolarize, producing a self-propagating wave through cells of the cardiac muscle [52]. After the depolarization, a slower recovery phase follows, called repolarization, until the cell reaches again the polarized rest state. The process of depolarization and repolarization is called action potential.

An ECG recording is the sum of all action potentials generated by all heart cells as a function of time. More correctly, in general the ECG measures the potential differences between two electrodes at the skin level. Fig. 2.3 illustrates how the sequence of the cardiac electrical activation translates into the characteristic waves that we encounter in a typical ECG, while some useful nomenclature about the ECG is shown in Fig. 2.4. The ECG has a pattern, the so-called PQRST complex, which was named by Einthoven in 1895 [53]. The first wave encountered in the ECG is the

P-wave, which is associated with the depolarization (contraction) of the atria. In the next 50 ms approximately, only very weak signals are recorded as it takes some time for the depolarization wave to travel through the AV node while the number of heart cells that take part in the atrioventricular conduction is very small [54]. After that, the QRS complex follows, which is associated with the ventricular depolarization. The amplitude of the QRS complex is significantly larger than the P-wave amplitude because the amount of muscle fibers that are located in the ventricular walls is much higher than the fibers found in the atrial walls. This is explained by the fact that the ventricles need to pump the blood through most of the body while the atria only need to pump the blood to the ventricles. While the ventricles are depolarized, the atrial repolarization (relaxation) occurs, which is masked by the large ventricular activity. The recovery of the cells in the ventricles produces the T-wave.

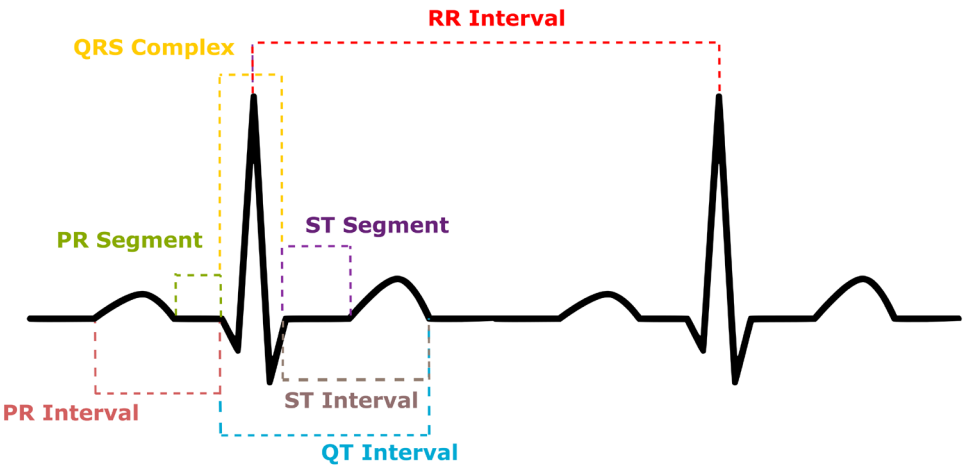



Fig. 2.4 Nomenclature of the ECG.

The electrical cardiac activity can be modeled using a dipole model. As the depolarization front travels through the cardiac muscle, it can be modeled by a rotating electrical dipole. The travelling dipole generates an electrical field in the heart that can be approximated by a single vector, called the electrical heart vector [56], [57]. The path that this vector follows over time in the three-dimensional space is known as the fetal vectorcardiogram (VCG) (Fig. 2.5). In this context, an ECG can be regarded as the projection of the electrical field produced by the rotating dipole onto the lead vector of two electrodes [55].

The fetal and adult heart are rather similar; however, some differences exist in the morphology of their ECG and in their HR [58]. The amplitude of the fetal ECG is lower, while the fetal heart beats two to three times faster than the adult heart.

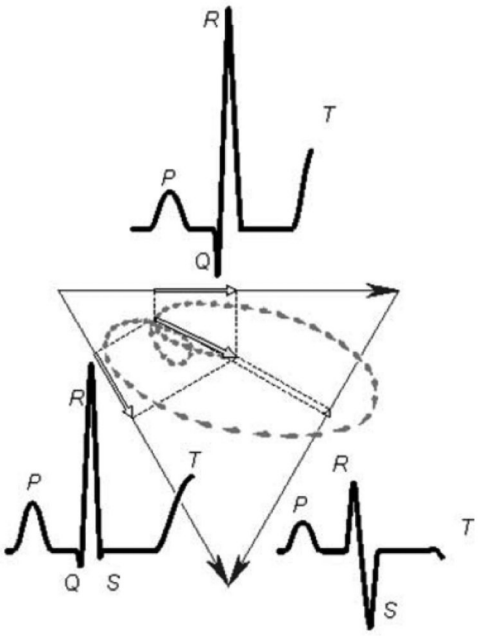



Fig. 2.5 2-D illustration of the fetal vectorcardiogram (gray) and the projection of the dipole vector onto the three leads of the Einthoven triangle (black) [55], [23].

The fetal and adult heart are rather similar; however, some differences exist in the morphology of their ECG and in their HR [58]. The amplitude of the fetal ECG is lower, while the fetal heart beats two to three times faster than the adult heart. Moreover, the HR variability of fetuses is lower [59] and the normal ranges for the HR and the HR variability vary throughout the pregnancy [60]. Another significant difference between the fetal and adult hearts is the electrical heart axis, i.e. the major direction of the overall electrical activity of the heart. For adults, the electrical axis points towards the left ventricle, while for fetuses it points towards the right ventricle due to its larger relative mass [61]. Thus, the ECG representation for the fetus differs from the same ECG representation for the adult, i.e., if it were possible to place electrodes on the fetal body in the same configuration as placed on the adult body, the fetal ECG would still look different due to these physiological differences.

## 2.2 Fetal Monitoring

Several methods exist for monitoring the fetal well-being so that obstetricians and midwifes can better assess the fetal health status. Ultrasound echocardiography is a widely adopted non-invasive technology that produces images of the fetal heart's interior and is mainly used for diagnosis of congenital heart defects, typically during the second trimester of pregnancy [62]. The disadvantages of this technique are in the cost of high-end equipment and the need for highly trained operators to acquire high-quality images and perform accurate assessment [63]. Moreover, with fetal echocardiography it is difficult to detect specific abnormalities, like coarctation of the aorta or ventricular septal defects [64].

A popular and relatively simpler method is cardiotocography (CTG), which is routinely used both during pregnancy and labor to monitor the fetal HR together with the uterine contractions [65]. The focus of CTG monitoring is to identify fetal HR patterns associated with inadequate oxygen supply to the fetus. CTG monitoring can be performed by four methods: Doppler ultrasound, phonocardiography (PCG), magnetocardiography (MCG) and electrocardiography (ECG) [27]. Currently, the most widely adopted method in clinical practice is Doppler CTG, which uses ultrasound technology and the Doppler principle to interpret changes in frequency of sound waves that are reflected from pulsating tissue in the fetal heart. As highlighted in the Introduction section, the Doppler ultrasound method has many limitations, including its high sensitivity to motion, frequent signal loss, and limited applicability (i.e. it cannot be applied 24/7).

An alternative approach to fetal HR extraction is fetal PCG, which is a recording of the sounds and murmurs made by the heart. These sounds accompany the mechanical activity of the heart and are related to changes in the blood flow and the opening and closing of the cardiac valves, and thus could provide diagnostic information [66]. Typically, the heart sounds are picked up by sound transducers (microphones) placed on the maternal abdomen [67]. Fetal PCG is an inexpensive and passive methodology, since no energy is transmitted to the fetus, enabling long-term fetal monitoring. However, fetal HR extraction from PCG recordings is very challenging because of the large amount of noise that is present, impeding the adoption of the method in clinical practice. Noise is produced by several sources including maternal heart, respiratory and digestive sounds and fetal movements resulting in PCG signals of very low signal-to-noise ratio (SNR) [68].

Another method used for fetal HR monitoring is the fetal MCG, which is a noninvasive recording of the magnetic fields generated by the electrical activity of the fetal heart [69], [27]. It is a passive and safe means of fetal monitoring that uses superconducting quantum interference device (SQUID) sensors placed on the maternal abdomen. MCG signals have high quality allowing for beat-to-beat HR variability analysis and waveform information extraction. Nevertheless, fetal MCG is a very expensive technique that requires trained staff. Moreover, it requires a dedicated shielded environment to avoid electromagnetic interference. As a result, this method is not used in clinical practice.



CTG measurements can also be performed with electrocardiography, both invasively and non-invasively. Invasive recordings are performed through an electrode placed on the fetal scalp, but they are limited to the intrapartum period and only after the completion of 32 weeks of gestation. Non-invasive measurements are done through electrodes placed on the maternal abdomen. Therefore, the fetal signal has much lower intensity than the maternal signal and is additionally contaminated by various other noises and interferences resulting in low-quality fetal ECG signals. However, the passive nature of the ECG-based CTG monitoring enables long measurements that can be utilized for studying beat-to-beat HR variability. Moreover, these fetal ECG measurements are much cheaper than fetal MCG. Finally, ECG recordings, similar to MCG, offer the possibility to study the fetal ECG signal morphology.

In general, CTG measurements often result in unnecessary interventions and significantly increase the number of Caesarean deliveries [70]. Despite the fact that CTG is preferred by most healthcare providers, this type of monitoring has not demonstrated consistent improvements in perinatal outcomes [71]. To avoid unnecessary operative deliveries and prevent fetal hypoxia, adjunctive techniques have been developed to further assess fetal oxygenation.

One of these techniques is fetal scalp blood sampling. During this procedure, a sample of blood is taken from the fetus to evaluate whether the pH is low, which indicates the presence of acidosis. However, systematic reviews have demonstrated no benefit of using fetal blood sampling in reducing unnecessary interventions and influencing long-term outcomes [15]. Fetal pulse oximetry [72] is used to measure oxygen saturation of fetal blood as an adjunct to CTG monitoring, similar to fetal blood sampling. It utilizes a probe that is placed on the fetal head through the dilated cervix during labor. According to studies, the addition of fetal pulse oximetry does not reduce overall Caesarean section rates and better methods are required

to evaluate the fetal well-being [72], [73]. ST analysis (STAN, Neoventa Medical Ab Mölndal) was introduced in the 1990s to detect metabolic acidosis and improve perinatal outcome [17]. The STAN method analyses the ST waveform of the fetal ECG, acquired through an invasive scalp electrode during labor. ST analysis together with CTG monitoring showed promising results in reducing the rate of operative deliveries and metabolic acidosis [16], [17]. However, subsequent randomized trials did not reproduce these original findings [19], [18], [20], [21].

## 2.3 Clinical significance of the fetal electrocardiogram

### 2.3.1 Fetal heart rate and fetal heart rate variability

The most frequently used feature of the fetal ECG is the QRS complex, which is employed for the extraction of the fetal HR through calculation of the RR intervals. The range of normal fetal HR varies significantly with gestational age. At 20 weeks of gestation the baseline fetal HR ranges from 110 to 160 bpm [4]. However, the HR varies continuously under normal physiological conditions. This beat-to-beat variation in the RR time interval is referred to as fetal HR variability (HRV).

The clinical importance of the fetal HRV was first recognized in 1965 [74], when fetal distress was found to be preceded by variations in interbeat intervals before noticing remarkable changes in the fetal HR. In 1978, it was found in adults that lower HRV is associated with a higher risk of mortality after myocardial infarction [75]. Later, HRV analysis was used apart from the cardiology domain in other applications such as prediction of diabetic neuropathy [76].

According to the opinion of clinicians, monitoring the fetal HR is of utmost importance both during pregnancy and labor. During pregnancy, analysis of fetal HRV and the patterns of the fetal HR can be used to estimate the fetal brain development [77] and predict developmental outcomes in early childhood [78]. During labor, according to several studies, fetal distress can be detected by using HRV features and examining abnormal fetal HR patterns [31], [30], [29], [79]. Fetal distress typically occurs when the fetus does not receive enough oxygen and, if not managed in time, this can lead to complications, such as hypoxic-ischemic encephalopathy (HIE), cerebral palsy (CP) and even fetal mortality [79]. Analysis of the HRV is currently limited

by the use of Doppler ultrasound CTG, which does not allow to study beat-to-beat variability. Unlike Doppler ultrasound CTG, fetal electrocardiography does not have this limitation.

### 2.3.2 Congenital heart defects



Congenital heart defect (CHD) is defined as a clinically significant structural heart defect present at birth and is a major cause of serious morbidity and mortality [80]. Timely detection of CHD during pregnancy has several advantages including prompt treatment after birth, preparation of the parents for the arrival of their sick child and the option to terminate the pregnancy in case of severe defects. Moreover, it was indicated that prenatal CHD diagnosis increases survival rates and reduces long-term morbidity [81], [82]. Currently, detection of CHD is done in mid-pregnancy via ultrasound examination. However, approximately half of the severe cases are still missed [83], [84].

Fetal electrocardiography could be an additional tool for fetal heart assessment. It was demonstrated that the waveform of the fetal ECG changes in the presence of CHD [40], [41]. It was recently shown that it is possible to standardize the fetal ECG for the fetal orientation in mid-pregnancy for a healthy fetus and that in fetuses with specific CHD the standardized ECG clearly differs [85]. Moreover, a deep neural network was able to detect CHD from a fetal vectorcardiogram at 20 weeks of gestation with accuracy of 76%, higher than fetal echocardiography [6]. In that case noninvasive fetal ECG measurements were initially processed to yield a 3-dimensional vectorcardiogram.

### 2.3.3 ST interval

The ability of the fetal heart to pump blood depends on a balance between energy-producing and energy-consuming processes. Normally, the available oxygen exceeds its request and the fetal heart utilizes oxygen-dependent aerobic metabolism to produce energy. In such a case, the energy balance is positive, and the ST waveform is normal [86]. If the amount of oxygen happens to be insufficient, the energy balance becomes negative and this results in hypoxia. Hypoxia produces changes in the morphology of the ST interval, which is normally characterized by a horizontal or upward sloping ST segment and a T-wave with stable amplitude. During hypoxia,

the ST segment becomes depressed, exhibiting a downward slope [86]. The fetus responds to the negative energy balance with an increase in adrenaline production, that further decreases the energy balance and initiates the use of stored glycogen for energy production [87]. Therefore, the aerobic metabolism of the fetus is supported by the anaerobic metabolism to restore the energy balance to equilibrium [87]. The breakdown of glycogen causes an increase in the height of the T wave. Changes in ST interval basically reflect the fetal defense mechanisms against hypoxia, and monitoring of these changes in clinical practice adds value to the current CTG monitoring approach.

### 2.3.4 QT interval

The QT interval is determined from the beginning of the QRS complex until the end of the T wave, and thus includes both the ventricular depolarization and repolarization. In a number of studies, changes in the duration of the QT interval were associated with an increased risk of death due to myocardial ischemia [88]. In fetuses, the QT interval has been of interest for monitoring fetal hypoxia. During labor, alterations in the length of the QT interval were found to be associated with hypoxia resulting in metabolic acidosis [89].

### 2.3.5 Fetal movement

Fetal movement is a valuable indicator of fetal health, as a marked movement decrease can be a precursor to fetal death [90]. The oldest and most commonly used method to assess fetal movement is maternal counting of fetal activity based on her perception [91]. However, it is possible to perform automatic and long-term detection of fetal movement based on the non-invasive fetal ECG.

In abdominal recordings, the fetal ECG is the projection of the VCG on the electrode lead vectors on the maternal abdomen. In the absence of maternal abdominal movement, the lead vectors are stationary and changes in the morphology of the fetal ECG can be associated with fetal movement [23], [90], [92]. Undoubtedly, also changes in the physiological condition of the fetus result in changes in the ECG morphology. However, normally these changes happen in larger time scales comparing to fetal movement, which is expected to take place in small time scales. Therefore, the time scale of the changes can be used to discriminate the causes of the

morphological ECG changes. Moreover, movement causes coherent changes in the ECG of the various electrodes. Different type of coherence among the different ECG channels exist when the changes in ECG morphology originate from e.g. hypoxia. Therefore, spatial effects can also be explored to distinguish movement from physiological changes. It is important to point out that movement of the fetal limbs does not alter the morphology of the abdominal fetal ECG, so the application of fetal ECG is limited to detection of thoracic movement.



## 2.4 Noninvasive fetal ECG extraction

### 2.4.1 Noise in the fetal ECG

Besides the fetal ECG signal, several other physiological and non-physiological interferences and noises are recorded by the abdominal electrodes [23]. These include, but are not limited to, the powerline grid, maternal ECG, abdominal muscle activity (electromyogram, EMG) and electrical activity from the uterus (electrohysterogram, EHG). An example of an abdominal recording is shown in Fig. 2.6, where for the sake of clarity the powerline interference has been removed.

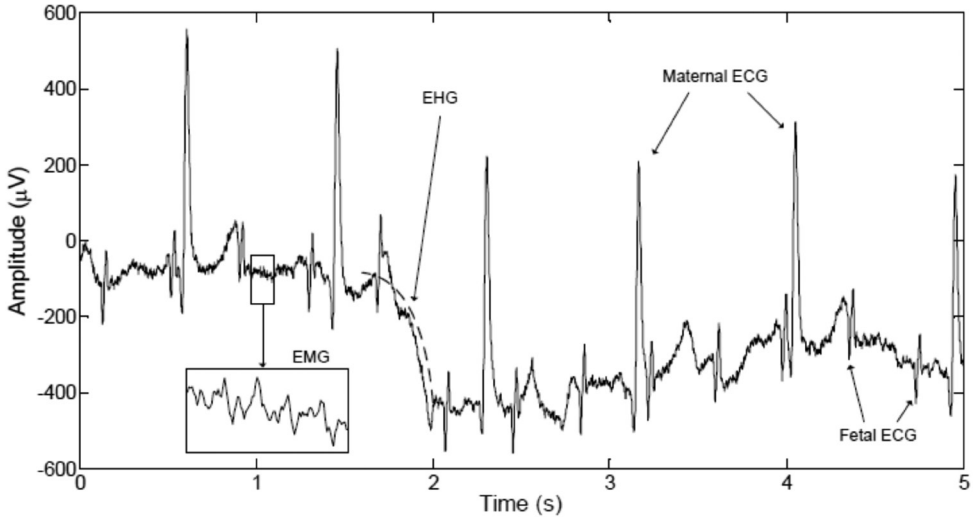



Fig. 2.6 Example of an abdominal recording. Apart from the fetal ECG, several interferences and noises are also recorded and some of them are indicated in the figure [23].

The ECG of the mother has about ten times larger amplitude than the fetal ECG, and the frequency content ranges approximately from 0 Hz to 80 Hz, similar to the fetal ECG [93]. The EHG includes frequencies from 0 Hz to roughly 3 Hz, while the EMG from 0 Hz to 200 Hz [23]. Lastly, the powerline interference has a frequency component of either 50 Hz or 60 Hz, along with its harmonics. The fact that these interferences and noises overlap in frequency with the fetal ECG makes their suppression a rather challenging task.

### 2.4.2 Signal preprocessing

Before extracting and analyzing the fetal ECG, the first steps in the signal processing pipeline usually involve the removal or suppression of interferences and noises such as the powerline interference (PLI) and baseline wander. There are several works in the literature for their removal or suppression, with the main challenge being to preserve the fetal ECG components.

The PLI is a common interference in biomedical signals and is determined by the power supply network. It occurs due to differences in electrode impedances and parasitic currents through the patient body and electrode cables. PLI is a problem in fetal ECG analysis as it can reach amplitudes much greater than the abdominal fetal ECG signal [94]. Solutions exist for reducing the PLI, such as cable shielding and proper skin preparation, but these are often insufficient to suppress it. The review of Țarălungă et al. [94] describes the main methods for removing the PLI, which are notch filters, Kalman filters, wavelets, adaptive filters, neural networks and blind source separation.

Baseline wander is a strong noise component, which can be caused by factors such as respiration, changes in electrode impedance and motion. Baseline wander can mask important information from the fetal ECG and thus its removal is an essential preprocessing step. The simplest method for baseline wander suppression is the use of a high-pass filter that blocks the low-frequency components characteristic of the noise, while allowing the main fetal ECG components to pass through the filter. The frequency of the high-pass filter is usually set to roughly 0.5 Hz, slightly affecting the low-frequency fetal ECG components such as the ST segment [95]. Other methods were also investigated in the literature such as techniques based on average and median filtering [58], wavelet-based and spline approximation algorithms [95]. However, for medical applications, high-pass filtering is typically the best choice because of its low computational complexity and relatively high accuracy [95].

### 2.4.3 Fetal ECG extraction

After preprocessing the abdominal recordings, the main signal processing step follows, which involves the removal of the maternal ECG and the extraction of the fetal ECG signal. Several methods have been proposed in the literature for extraction of the fetal ECG signal [43]. One popular category of methods is template subtraction, which refers to approaches where the maternal component is (dynamically) estimated (template) and then subtracted from the signal mixture [96], [97], [98], [99]. Kalman filtering and extended Kalman filtering have also been used to filter the maternal ECG component [100], [101]. Wavelet transform is another widely used method for fetal ECG extraction that, because of the spectral overlap of the signal and the several noises, is usually used in combination with other techniques such as adaptive filtering [102] or blind source separation [103]. When using blind source separation methods for fetal ECG estimation, the latent sources of the signal mixture, one of them being the fetal ECG, are estimated assuming that they are statistically independent [104], [105], [106], [107]. Several methods have been also developed that use one or more additional reference inputs from electrodes placed on the maternal chest. These inputs are assumed to contain the maternal ECG and different techniques have been proposed for estimating and subtracting it from the mixture [108], [109], [110]. In general, each method has strengths and weaknesses and combination of different techniques can potentially improve the fetal ECG extraction performance [111].

### 2.4.4 Signal postprocessing

Even after the maternal ECG removal step, the extracted fetal ECG is still substantially contaminated by residual noises that are often non-stationary, complex, and have spectral overlap with the fetal ECG. Despite the vast literature on the separation of maternal and fetal ECG, very few works focused on the postprocessing step for enhancing the quality of the fetal ECG signals so as to enable accurate analysis of the fetal ECG morphology. The oldest, simplest and more extensively used method consists of averaging of consecutive ECG complexes, at the expense of losing individual variations in pulse shape [112], [113]. Different wavelet denoising techniques have also been investigated for improving the SNR of the fetal ECG signals [114], [115]. In this thesis, Chapters 4, 5 and 6 are dedicated to developing postprocessing techniques for fetal ECG enhancement that aim at improving the performance of the existing methods tackling their main shortcomings.

### 2.4.5 Complications in fetal ECG analysis

Apart from the various noises and interferences that corrupt the noninvasive fetal ECG, the analysis of the fetal ECG signals is further complicated due to changes in the volume conductor between the fetal heart and the abdominal electrodes. The volume conductor consists of several anatomical layers with different electrical conductivities that surround the fetus and through which the cardiac signals must pass to reach the maternal body surface. These layers include the amniotic fluid, vernix caseosa, skin of the mother and fat [4]. The volume conductor is not constant and its changes can distort, attenuate or amplify the fetal ECG [116], [117], [47], [58]. These changes mainly originate from the development of vernix caseosa layer [23]. Between the $28^{th}$ and $32^{nd}$ week of gestation, the vernix caseosa, which is a thin protective layer, is formed [58]. This layer with very low conductivity introduces an electrical shield around the fetus, making the recording of the fetal ECG through abdominal electrodes very difficult. When the pregnancy is normal, the vernix caseosa slowly dissolves around the $37^{th}$ and $38^{th}$ week of gestation. However, since the $32^{nd}$ week the vernix caseosa starts breaking down and some holes are formed which restore the possibility to perform abdominal recordings [23]. In general, limitations in fetal ECG analysis caused by this layer are mostly expected between $28^{th}$ week and $37^{th}$ week of pregnancy.

### 2.4.6 Devices for fetal ECG monitoring

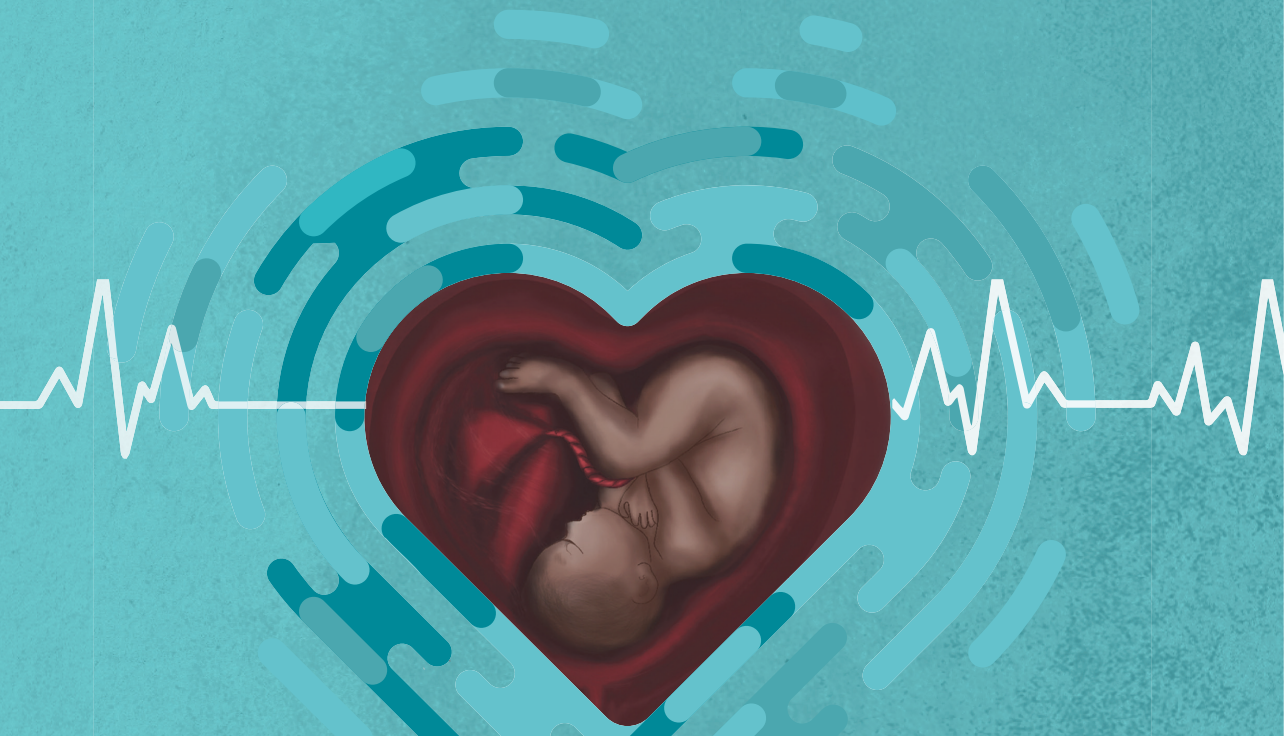
(a) (b) (c) (d)

Fig. 2.7 Commercial fetal monitors based on noninvasive fetal ECG. (a) Monica AN24, (b) Monica Novii Wireless Patch System, (c) MERIDIAN M110 Fetal Monitoring System and (d) Nemo Fetal Monitoring System [43].

The first commercial fetal HR monitor, which was approved in the United States by the Food and Drug Administration (FDA) was the Monica AN24 monitor (Monica Healthcare, Nottingham, UK) in 2011. After Monica AN24 monitor, more commercial

fetal monitors appeared such as Monica Novii Wireless Patch System in 2014, MERIDIAN M110 Fetal Monitoring System in 2017 (MindChild Medical, Inc., North Andover, MA, USA) and Nemo Fetal Monitoring System in 2018 (Nemo Healthcare, Veldhoven, The Netherlands). These devices can measure uterine activity and fetal HR. Each device uses a different number of electrodes and different ways of applying them on the maternal abdomen as demonstrated in Fig. 2.7.

# PART I Fetal heart rate extraction



# CHAPTER 3

# A DILATED INCEPTION CNN-LSTM NETWORK FOR FETAL HEART RATE ESTIMATION

Based on [JP-4]: Fotiadou E., van Sloun R.J.G., van Laar J.O.E.H, Vullings R., A Dilated Inception CNN-LSTM Network for Fetal Heart Rate Estimation. Physiol. Meas., 2021, doi: 10.1088/1361-6579/abf7db

**Abstract.** *Objective:* Fetal heart rate monitoring is routinely used during pregnancy and labor to assess fetal well-being. The non-invasive fetal electrocardiogram (ECG), obtained by electrodes on the maternal abdomen, is a promising alternative to standard fetal monitoring. Subtraction of the maternal ECG from the abdominal measurements results in fetal ECG signals, in which the fetal heart rate can be determined typically through R-peak detection. However, the low signal-to-noise ratio and the nonstationary nature of the fetal ECG make R-peak detection a challenging task. *Approach:* We propose an alternative approach that instead of performing R-peak detection employs deep learning to directly determine the fetal heart rate from the extracted fetal ECG signals. We introduce a combination of dilated inception convolutional neural networks (CNN) with long short-term memory networks to capture both short-term and long-term temporal dynamics of the fetal heart rate. The robustness of the method is reinforced by a separate CNN-based classifier that estimates the reliability of the outcome. *Main results:* Our method achieved a positive percent agreement (within 10% of the actual fetal heart rate value) of 97.3% on a dataset recorded during labor and 99.6% on set-A of the 2013 Physionet/Computing in Cardiology Challenge exceeding top-performing state-of-the-art algorithms from the literature. S*ignificance:* The proposed method can potentially improve the accuracy and robustness of fetal heart rate extraction in clinical practice.

## 3.1 Introduction

Hypoxia that occurs when the brain does not receive adequate oxygen poses a significant risk for fetuses. The failure of oxygen delivery to the fetus can cause permanent brain damage, developmental delays, or even death in severe cases. Although oxygen deficiency can happen at any stage during pregnancy most injuries typically occur during labor. The heart rate (HR) pattern of the fetus changes as a response to reduced oxygenation [118], [119]. Therefore, the fetal HR must be monitored during labor but also pregnancy, especially in high-risk pregnancies.



Many hospitals routinely use continuous electronic fetal HR monitoring during pregnancy and labor. Electronic fetal HR monitoring measures fetal HR in response to the contractions of the uterus and can be performed internally or externally. External fetal HR monitoring is the most common method and can be performed by placing an ultrasound transducer on the maternal abdomen. However, this method often provides inaccurate results, as it is affected by the movement of the mother and the fetus and suffers from signal loss in case of obese patients [43]. Internal monitoring is carried out by placing an electrode on the fetal scalp [118]. It provides a more accurate and consistent HR because factors such as movement have a smaller impact on the quality of the measurement. However, scalp fetal HR measurements can only take place during labor after the membranes have ruptured and there is sufficient dilation of the cervix. Additionally, the scalp electrode may cause injury to the fetus.

An alternative way of measuring the fetal HR externally is by measuring the noninvasive fetal electrocardiogram (ECG) by placing electrodes on the maternal abdomen. This method has the potential to provide accurate measurements and can be performed both during pregnancy and labor. As opposed to ultrasound measurements, the fetal ECG recordings enable beat-to-beat HR extraction, necessary for reliable analysis of the HR variability [120]. To extract the fetal HR from the fetal ECG recordings, one needs to detect the fetal R-peaks since the HR is related to the distance between two successive peaks. However, the noninvasiveness of the fetal ECG comes at the cost of a reduction in signal-to-noise ratio (SNR) [121]. The noninvasive abdominal recordings are severely contaminated by electrical interferences and noises such as the maternal ECG, powerline interference, muscle noise, equipment noise, etc., thus complicating the detection of the fetal R-peaks.

Several methods have been proposed in the literature in the area of fetal ECG and fetal HR extraction from noninvasive abdominal recordings. The 2013 Physionet/Computing in Cardiology Challenge aimed to encourage fetal HR estimation from abdominal recordings [122]. Along with the challenge, a database became publicly available to allow for the comparison of different algorithms. The methods presented in the challenge were unique but, as pointed out in the review of Clifford et al. [123], most followed a five-step approach. This includes preprocessing of the abdominal signals, estimation, and suppression of the maternal ECG signal, R-peak detection, and postprocessing of the fetal HRs. A variety of algorithms were proposed for maternal ECG subtraction, such as adaptive filtering [108], [109], [58], template subtraction [98], [99], blind source separation [105], [124], [106], [107] or a combination of different algorithms [125], [102], [103]. For an extensive review of fetal ECG extraction methods see [121] or [126]. However, even after the maternal ECG is removed, the SNR of the extracted fetal ECG signals is usually low, and this can result in faulty R-peak detections.

The winner of the challenge, Varanini et al. [104], used two QRS detectors in forward and backward directions after enhancing the extracted multichannel fetal ECG signal with independent component analysis. Afterwards, the fetal ECG channel with the best R-R series was selected for HR estimation. However, since the orientation of the fetus can change, causing variations to the SNR of each channel, using multichannel information can be more robust. Warmerdam et al. [127] suggested an adaptive multichannel R-peak detection method that combines HR information and ECG waveform. However, a linear autoregressive model was used as an HR model that is not able to describe complex accelerations and decelerations that occur during labor. When tested on set-A of the 2013 Physionet/Computing in Cardiology Challenge [122], Warmerdam et al. achieved 99.6% accuracy in R-peak detection while Varanini et al. 98.6%. Since the challenge, more research has been performed in the area of fetal HR extraction with promising results [128], [129].

Deep learning methods have achieved remarkable success in tasks such as image classification [130] and speech recognition [131] and the expectations on how this technology could help to improve health care are high [132]. Several works have been reported in the area of ECG processing such as adult ECG denoising [133], [134], adult arrhythmia detection [135], prediction of fetal acidemia [136], fetal ECG denoising [137], [138] and fetal ECG signal reconstruction [139]. Present signal processing algorithms for fetal HR extraction have limited performance because

there are no accurate models for the noise that remains after the maternal ECG is suppressed. However, complex deep learning models might be more efficient in HR estimation especially in low-SNR signals. Two works in the literature already attempted fetal QRS detection with deep learning methods. Zhong et al. [140] proposed a convolutional neural network (CNN) model for QRS complex detection from noninvasive abdominal recordings without canceling the maternal ECG. Subsequently, Lee et al. [141] advanced this approach by using multichannel signals, deeper architecture, and postprocessing, leading to an improved positive predictive value of 92.77%. However, both methods did not manage to reach the performance of conventional signal processing algorithms. Moreover, there is a limitation in the validation of both approaches to only 7 subjects of the set-A of the 2013 Physionet / Computing in Cardiology Challenge dataset [122].



Recently, the authors proposed a deep learning model that directly estimates the fetal HR from the extracted fetal ECG signals [142]. The model, which combines CNNs with long short-term memory (LSTM) networks, achieved comparable performance with a top-performing state-of-the-art HR extraction algorithm. A limitation of the method was that it failed to correctly estimate the HR in cases of extremely low-quality signals obtained during the second stage of labor, especially when decelerations happened. In clinical practice, absence of information is unfavorable but wrong information is worse as it might lead to misdiagnosis.

In this chapter, we further improve and extend our previous work aiming at more robust and efficient HR estimation. Inspired by [143], we propose to use a dilated inception CNN encoder in our network to achieve multiscale feature extraction and a variety of receptive fields. An LSTM decoder network is chained to the encoder to learn long-term temporal feature relations and extract the fetal HR. Our main contributions are outlined as follows:

- A deep hybrid dilated inception CNN-LSTM (DICNN-LSTM) encoder-decoder network that extracts the fetal HR from noninvasive abdominal recordings. To the best of our knowledge, we are the first that employed deep learning to estimate the fetal heart HR without explicitly detecting the QRS complexes.

- The reliability of our method is reinforced by a classifier, which uses a CNN network to identify the periods of time that the extracted fetal HR is inaccurate, increasing the suitability of our method for clinical application.

- Experimental results demonstrate the advantage of our method over top-performing state-of-the-art algorithms.

The rest of the chapter is organized as follows. Section 3.2 presents the proposed fetal HR extraction method, the HR reliability classifier and the data used. Experimental results are provided in Section 3.3. Finally, the results are discussed in Section 3.4 and conclusions are drawn in Section 3.5.

## 3.2 Materials and methods

### 3.2.1 Data

Two different datasets were used in this study. The first one is a private dataset obtained in a collaboration between the Eindhoven University of Technology and the Máxima Medical Center, Veldhoven, The Netherlands. This dataset is a part of the study described in [144]. It contains 28 abdominal recordings measured by 4 electrodes during labor at a sampling frequency of 500 Hz. The data were collected from 28 women with a gestational age between 36 and 42 weeks with a total duration of roughly 91h. Each recording was partially obtained during the first and second stage of labor. Simultaneous scalp HR recordings were performed and stored at 4 Hz. In nearly all recordings, clock drift was present between the scalp fetal HR and the fetal ECG, leading to desynchronization after a certain amount of time. To limit this desynchronization, the fetal HR was resampled using a manually determined resample factor. The fetal ECG signals were extracted from the abdominal measurements by the methods described in [98], [145]. According to [145], a fixed-lag Kalman smoother with adaptive noise estimation was used to filter the powerline interference. After this step, the maternal ECG was suppressed by a template subtraction technique, known as weighted averaging of maternal ECG segments (WAMES) [98]. WAMES method dynamically segments the maternal ECG complex in separate parts and generates an individual template for each part. Each template is determined by linearly combining time-shifted, offset-compensated, and scaled corresponding parts in previous complexes. The individual templates are then combined to yield a maternal ECG template, which is subsequently subtracted from the recorded, preprocessed data. At the moment of conducting the study, the authors

were provided merely with the extracted fetal ECG data, while access to the raw data was not available. Fig. 3.1 illustrates two examples of fetal ECG signals contained in this dataset. Fig. 3.1(a) presents a signal recorded during the first stage of labor, while Fig. 3.1(b) a signal obtained during second stage of labor. Note that the vertical axis limits for the signals in Fig. 3.1(a) and (b) differ for better visualization. Notably the signal in Fig. 3.1(a) has significantly higher quality than the signal in Fig. 3.1(b). All the signals of this dataset obtained during second stage of labor have high amounts of noise, making it virtually impossible to distinguish the fetal R-peaks.



The second dataset is the set-A of the 2013 Physionet/Computing in Cardiology Challenge [122]. It consists of 75 1-minute noninvasive abdominal signals sampled at 1000 Hz. The data were obtained from multiple sources, using a variety of instrumentation with different frequency response, resolution and configuration, although in all cases they are presented at 1000 samples per second. Reference annotations of the fetal QRS complexes were made available as well which enabled us to determine and evaluate the fetal HR. The reference annotations were produced, usually with reference to a direct scalp fetal ECG signal. Following the suggestion of Behar et al. [146], organizer of the challenge, seven recordings (a33, a38, a47, a52, a54, a71, and a74) were discarded due to inaccurate annotations. The algorithm of Varanini et al. [104] was used to extract the fetal ECG signals from the abdominal recordings. According to the method of Varanini et al., first the baseline wander and the powerline interference were removed. Afterwards, the maternal ECG was estimated through independent component analysis (ICA) and singular value decomposition and subsequently subtracted from the signals. Finally, a second ICA was employed to enhance the fetal ECG signal. We need to note that since our fetal HR extraction framework was developed for signals sampled on 500 Hz, before applying our algorithm on the Physionet dataset, but after applying Varanini's method, the signals were resampled to 500 Hz.

16 (54h) out of the 28 recordings of our private dataset were randomly selected and used to train the fetal HR extraction network. 6 recordings of 22.5h were kept as validation set to tune the parameters of the network. The remaining 6 (14.5h) recordings of our private set, together with the set-A of the Physionet database (68 min) [122], were kept as a test set to evaluate the performance of the network. The scalp fetal HR was used as the desired output of the network (labels).

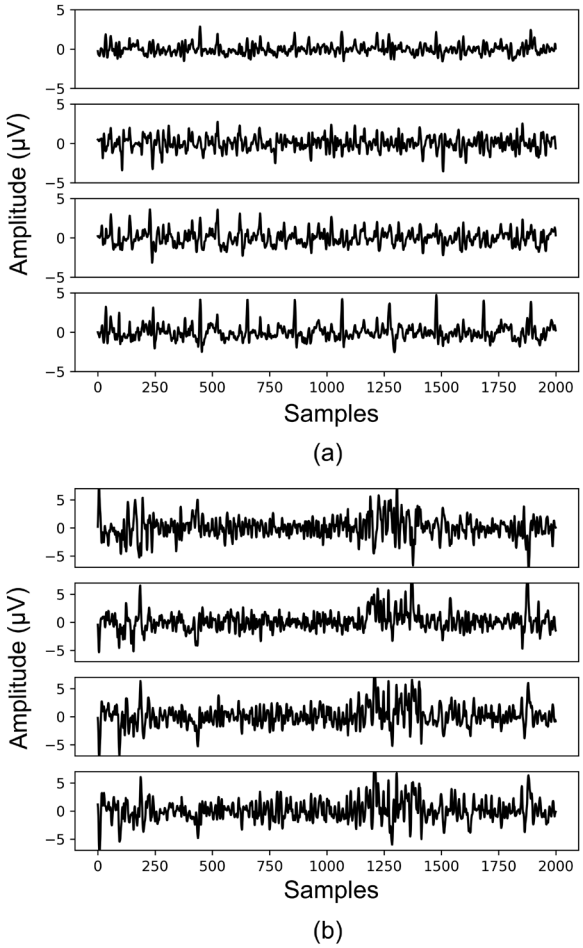



Fig. 3.1. Two example 4-channel fetal ECG signals from our private dataset [144], where the signal presented in (a) was obtained during first stage of labor and in (b) during second stage of labor.

### 3.2.2 Fetal heart rate extraction

In this section, we present our DICNN-LSTM model for fetal HR extraction from noninvasive fetal ECG signals. The model, which is illustrated in Fig. 3.2, is comprised of two main blocks. The DICNN network block (encoder) consists of six stacked dilated convolution inception modules,which are depicted in detail in Fig. 3.3. The DICNN network is used as a feature extractor and the extracted features are fed to the LSTM network block (decoder) that is responsible for estimating the fetal HR.

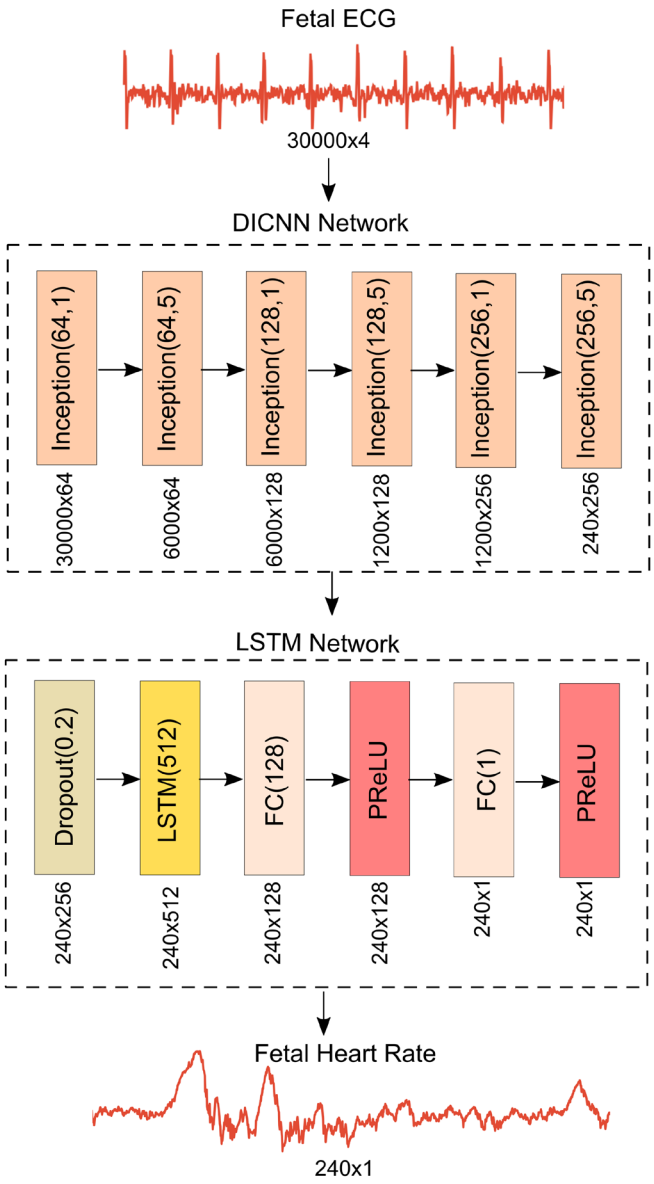



Fig. 3.2. Overview of the proposed DICNN-LSTM network. A dilated inception CNN network extracts features from the input 4-channel fetal ECG signal (1-minute duration), which are then fed to an LSTM network for the estimation of the fetal heart rate (4values/s). Next to each layer the output layer size is provided. The Inception(i,j) block is given in more detail in Fig. 3.3, while i indicates the number of filters of the convolutional layers and j the size of the max pooling window. FC stands for a fully connected layer.

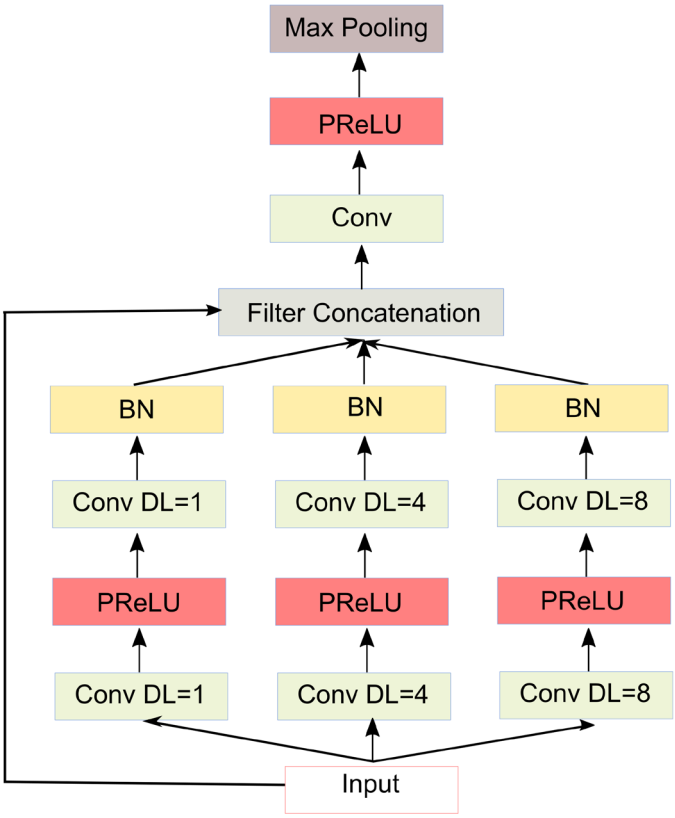



Fig. 3.3. The dilated inception block used in the DICNN-LSTM network of Fig. 3.2. Conv stands for a convolutional layer, DL for dilation rate, and BN for a batch normalization layer.

#### *3.2.2.1 Network description*

##### 3.2.2.1.1 Dilated convolution

Dilated convolutions were originally developed for wavelet transforms [147] and later proposed for multiscale context aggregation in [148]. Given a 1-D signal $u$ and a kernel $h$, the output $v$ of a dilated convolution is defined as:

$$v[n] = \sum_{k=0}^{K-1} h[k]u[n - dk], \tag{3.1}$$

where $K$ and $d$ are the kernel size and the dilation factor respectively. Notice than when $d$ is 1, then the dilated convolution is the same as the conventional convolution.

Fig. 3.4 illustrates the 1-D dilated convolution operation with dilation rates 1, 2, and 4 and kernel size 3. As shown in the figure, d-1 signal samples are skipped during the dilated convolution.

Dilated convolutions are used to increase the receptive field of the network, i.e. the region in the input space that a CNN feature is affected by. When using conventional convolutions, the receptive field is linearly related to the depth of the layer. Dilated convolutions can achieve a receptive field that is exponentially related to the layer depth when exponentially increasing dilation rates (d=1,2,4...) are used.



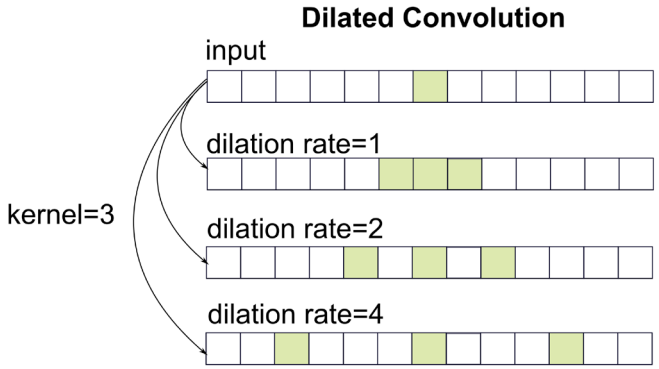



Fig. 3.4. 1-D dilated convolution with a kernel size of 3 and dilation rates 1, 2, and 4. The green block in the input signal (first row) indicates the unit of interest. In the output signals (second, third, and fourth row) the green blocks show the receptive field for each different dilation rate.

#### 3.2.2.1.2 Encoder network

Szegedy et al. [149] proposed an inception model for image classification, intending to capture multiscale information in the input images. Salient signal parts can have large variations in size and thus choosing the right kernel size is not easy. Larger kernel sizes are preferred for more globally spread information, while smaller sizes are preferred for local information. Moreover, very deep networks are susceptible to overfitting and are computationally expensive. To address these issues, the original inception module was developed that uses a wider instead of deeper network by allowing filters of different kernel sizes to operate on the same level. More recently, Shi et al. [143] presented a dilated convolution inception model for single image super-resolution. In their model, instead of using convolutions with different kernel sizes, they employed convolutions of different dilation rates to learn multiscale information.

Inspired by both works we propose a stacked dilated inception convolutional encoder network to extract useful features for fetal HR extraction from noninvasive fetal ECG signals. We found that for our problem varying the dilation rate rather than the kernel size leads to better performance (in the validation dataset). As shown in Fig. 3.3, each inception module of the encoder consists of three parallel branches of two convolutional layers of different dilation rates (1, 4, and 8). Since our data are temporal, we use 1-D convolutions. Batch normalization layers are used after the convolutions to speed up the training process by normalizing the intermediate outputs. The outputs of these layers are subsequently concatenated together with the input of the module through a residual connection. The role of the residual connection is twofold. First, it allows the input features to be reused, leading to better performance. Second, it allows the gradients to propagate easier across our deep network preventing the vanishing gradient problem. A convolutional layer is applied after the concatenation of the features to select the most meaningful features and reduce the dimension of the feature vector. At the end of the module, a max-pooling layer is used to reduce the temporal dimension of the signal.

The parametric rectified linear unit (PReLU) [150] is used as an activation function between the layers. In contrast to the rectified linear unit (ReLU), in which the negative part is completely dropped, PReLU assigns a nonzero slope to it. The output of PReLU, *f(z)*, for an arbitrary input z, is defined as:

$$f(z) = \begin{cases} z, & for\ z > 0 \\ bz, & for\ z \leq 0 \end{cases}, \tag{3.2}$$

where the slopes *b* are optimized during the training of the network. The motivation to use PReLU is that it solves the "dying ReLU" problem when ReLU neurons become inactive and output zero for any input. Moreover, when tested in the validation set, we noticed a slight increase in our HR estimation accuracy when using PReLU instead of ReLU.

Six of these dilated convolution-based inception modules are stacked together to form the encoder of our model. We tried also different numbers of inception modules but six gave the best performance based on tests on our validation dataset. The encoder is both wide and deep and thus relatively complex, which is necessary

to extract relevant information from the typically very noisy fetal ECG signals. A 240x256 feature vector is extracted from the fetal ECG signals at the output of the encoder.

#### 3.2.2.1.3 Decoder network

The extracted feature vector is sent to the decoder network that is responsible for estimating the fetal HR. The first layer of the decoder is a dropout layer with a dropout rate of 0.2. This layer randomly ignores 20% of the neurons in the corresponding layer during training. This means that the weights for these neurons will not be updated on the backward pass. As a result, the network becomes less sensitive to the specific weights of the neurons and consequently can generalize better and is less likely to overfit.



Then, an LSTM layer with 512 nodes is used to model the temporal dynamics of the extracted features. LSTMs [151] were developed as an improved form of recurrent neural networks (RNNs) to handle the problem of vanishing and exploding gradient. Instead of simple RNN neurons, LSTMs are characterized by more complex memory blocks that consist of several gates to control the information flow in the internal memory cells. They were explicitly designed for learning long-term dependencies in sequential data, meaning that they can remember and associate past with present information. Thus, an LSTM layer is stacked after the feature extraction step and dropout to learn dependencies in the feature sequence. The LSTM output is fed to two fully connected layers with 128 and 1 neurons respectively that output the fetal HR. PReLU activation functions are applied also to the decoder of our network. The parameters of both the encoder and decoder of our network were determined such that the fetal HR estimation performance was maximized on our validation dataset. Due to the large number of parameters cross validation to tune them was prohibitive.

#### 3.2.2.1.4 Input and output signals

The input signal of the network is a 4-channel ECG signal of one-minute duration. Since the average normal fetal HR is 100 to 160 bpm, the input signal typically contains 100 to 160 heartbeats. Because our reference HR from the scalp electrode is 4 HR values/s (Section 3.2.1), the output of the network is set to 240 fetal HR values (60s x 4values/s).

To speed up the training process, the input signal X is normalized over the four channels separately by subtracting the mean and dividing by the standard deviation of the signals through:

$$X_{norm} = \frac{X - E[X]}{\sqrt{Var[X] + \epsilon}}, \tag{3.3}$$

where *E[X]* and *Var[X]* are the mean and the variance of the input signal respectively. A small constant $\epsilon$, chosen as 0.001, is added to the denominator to prevent division by zero. The output signals (labels) are not normalized.

#### 3.2.2.1.5 Loss function

Mean square error (MSE) is the most commonly used loss function in regression problems. As the name suggests, MSE measures the square difference between the HR estimations and the actual observations. Due to the squaring, estimations that are far away from their actual values are penalized more heavily. However, when estimating the fetal HR less importance should be given to clear outliers because these will be ignored by the clinicians in their visual inspection of the fetal HR traces. For this reason, the mean absolute error (MAE) was selected as a loss function for our problem. MAE is defined as

$$MAE = \frac{1}{Q}\sum_{i=1}^{Q}\left|FHR_{target_i} - FHR_{predicted_i}\right|, \tag{3.4}$$

where Q is the length of the output sequence (Q=240), $FHR_{predicted}$ is the estimated fetal HR and $FHR_{target}$ is the fetal HR measured by the scalp electrode.

#### 3.2.2.1.6 Training

The Adam algorithm [152] was selected as an optimization algorithm while the learning rate was set to 0.0001. The batch size was set to 8. At each training iteration, 8 random one-minute segments were chosen from the 54h training sequences. The scalp fetal HR was used as the desired output of the network. The network was trained for 1000 epochs and the model that minimized the loss on the validation set was finally chosen.

#### *3.2.2.2 Performance Evaluation*

##### 3.2.2.2.1 Evaluation metrics

The accuracy of the proposed method was assessed by measuring the mean absolute error (Equation 3.4) and the mean squared error defined by:

$$MSE = \frac{1}{Q}\sum_{i=1}^{Q}(FHR_{target_i} - FHR_{predicted_i})^2. \quad (3.5)$$

We determined the reliability of our method in terms of positive percent agreement (PPA), which is the percentage of fetal HR outputs that were within 10% of the actual HR value [153].We also present results where we used a smaller tolerance of 5% for the calculation of PPA (we will call this PPA_5).In addition, we used a second reliability metric, the coverage, that corresponds to the time that the method outputted a nonzero fetal HR value.



##### 3.2.2.2.2 Reference methods

The performance of our method was compared to the performance of the algorithm of Warmerdam et al. [127], that outperformed the algorithms of Varanini et al. [104] and Behar et al. [125], when tested in set-A of the 2013 Physionet /Computing in Cardiology Challenge dataset [122]. Warmerdam proposed a multichannel hierarchical probabilistic framework for detecting the fetal R-peaks. Two models are incorporated in this framework, a gaussian QRS model and an autoregressive HR model. Their framework consists of three inference levels for inferring each next R-peak location: state estimation, QRS and HR model estimation, and noise estimation. Initially, the QRS model is used to determine the next R-peak location. A sanity check is performed to evaluate if this is indeed an R-peak location and if not, the location is extrapolated based on the HR model. In addition to R-peak detection, the algorithm of Warmerdam et al. identifies the periods of time that the extracted fetal HR is unreliable and does not output an HR value for these. The code of the algorithm was provided by the authors for our experiments.

We performed an additional comparison of our method with the algorithms of Varanini et al. [104] and Behar et al. [125] in the Physionet dataset. Both [104] and [125] are fetal ECG extraction approaches that also calculate the fetal HR. We did not compare on our private dataset because we have access only to the extracted fetal ECG signals and not the raw data. For both algorithms we used online implementations provided by the authors in the Physionet website.

After the fetal ECG extraction process (section 3.2.1), the method of Varanini performs R-peak detection on individual channels using a two-step approach. First, a derivative filter is used for R-peak detection. Second, a forward and backward autoregressive model is trained from the obtained RR-series. The autoregressive model is subsequently used in combination with the derivative signal to detect the fetal R-peaks. In the end, the channel with the best RR-series is chosen based on some statistical features of the RR-series.

Behar uses a combination of various source separation techniques to cancel the maternal ECG such as template subtraction and blind source separation methods. Fetal R-peak extraction is then performed using a Pan and Tompkins R-peak detector on each separate fetal ECG channel. Finally, the channel with the smoothest RR-series is selected.

### 3.2.3 Fetal heart rate reliability estimation

When the fetal ECG signals have extremely low quality it is difficult to accurately estimate the fetal HR. To increase the robustness of our network, we propose to use a simple classification framework that determines whether the extracted fetal HR is reliable or not. In fact, the classifier determines if a fetal ECG segment could yield a reliable fetal HR depending upon its signal quality. Fig. 3.5 illustrates the proposed classifier for fetal HR reliability estimation. The input 4-channel fetal ECG signal is initially segmented in parts of 3000 samples that correspond to 6s. Each segment passes through a CNN network that produces a value between 0 (unreliable) and 1 (reliable) indicating the reliability of the fetal HR. For each 6-second segment, one output value is obtained. However, since the DICNN-LSTM network outputs 4 fetal HR values/s, we copy the same value 24 times to obtain consistent outputs. Afterwards, we concatenate all outputs obtained from the 6-second signal segments. A median filter of size 101 (26s) is subsequently applied to smoothen the reliability result for the whole input signal. Filtering removed some outliers and was found to slightly improve the classification performance on the validation dataset. The reasoning about our choice to filter the output as well as about the relatively large filter size is that it prevents the output from jumping between reliable and unreliable. Towards a clinical application we want to avoid such jumping as we want to prevent frequent raising of alarms. It is preferable to have more sustained alarms that could encourage the hospital staff to take measures to improve signal quality by e.g. asking the patient to move less. After the filtering operation, we set a cut-off threshold to

determine if the final outcome will be 0 (unreliable) or 1 (reliable) for each fetal HR value.

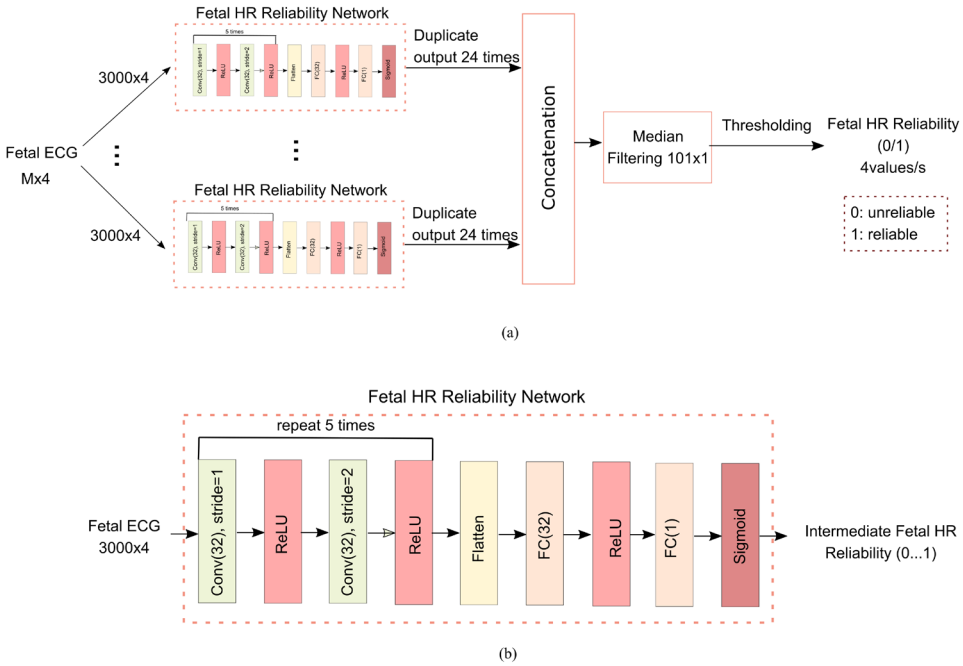



Fig. 3.5. (a) Illustration of the fetal heart rate reliability classifier. A CNN network is used to classify each 3000-sample-long segment of the 4-channel fetal ECG (M samples) and the outputs are subsequently concatenated and filtered to obtain the final result. (b) Description of the CNN network used to classify each fetal ECG segment. Conv stands for convolutional layer and FC for fully connected.

#### *3.2.3.1 Fetal heart rate reliability network*

The fetal HR reliability network, shown in Fig. 3.5(b), consists of 10 1-D convolutional layers followed by 2 fully connected layers. After every two convolutional layers subsampling by two is applied to the input (convolutional layers with stride 2). We empirically determined that employing 32 filters in each convolutional layer achieves a satisfactory result and avoids overfitting. The kernel size was set to 15 for all the convolutional layers. The two fully connected layers have 32 and 1 units respectively. All layers except for the last are followed by a ReLU activation function, while the last is followed by a sigmoid function. The latter produces a score between 0 and 1 indicating if the input will produce an unreliable (value close to 0) or reliable result (value close to 1).

#### 3.2.3.1.1 Training

The same training data that were used to train the DICNN-LSTM model were employed for the training of the fetal HR reliability network, i.e 16 recordings of our private dataset. However, the labels of the data were different because in this case we are not estimating heart rates but reliabilities. To obtain the training labels, the data were initially passed through the DICNN-LSTM network and the heart rates were estimated. Afterwards, the HR estimations were compared with the actual HR values obtained by the scalp electrode. When the estimations were close to the target values ($|FHR_{target} - FHR_{prediction}| < 0.1 * FHR_{target}$), then the data were labeled as reliable with label 1, otherwise unreliable with label 0.

We trained the network by minimizing the binary cross-entropy between the HR estimations and the labels using the Adam optimizer with learning rate 0.0001. The batch size was 32, comprising 16 positive (reliable) and 16 negative (unreliable) stochastically selected examples. After training the network for 1000 epochs, the model that minimized the validation loss was selected.

#### *3.2.3.2 Performance evaluation*

The performance of the HR reliability classification network was assessed in terms of area under the receiver operator characteristic curve (AUC), true negative rate (TNR) or specificity, and true positive rate (TPR) or sensitivity. AUC metric ranges from 0 to 1, with 0 meaning that all estimations are wrong and 1 that all estimations are correct.

## 3.3 Results

In this section, we apply our proposed model for fetal HR extraction and report the experimental results for both our private and the Physionet datasets. The effectiveness of our method is validated both quantitatively as well as qualitatively. The results are presented in three parts. In the first part experiments regarding the DICNN-LSTM are provided. In the second part the fetal HR reliability classifier is evaluated and in the third part the results of the combined system are presented.

### 3.3.1 DICNN-LSTM Network Experiments

#### *3.3.1.1 Kernel size*

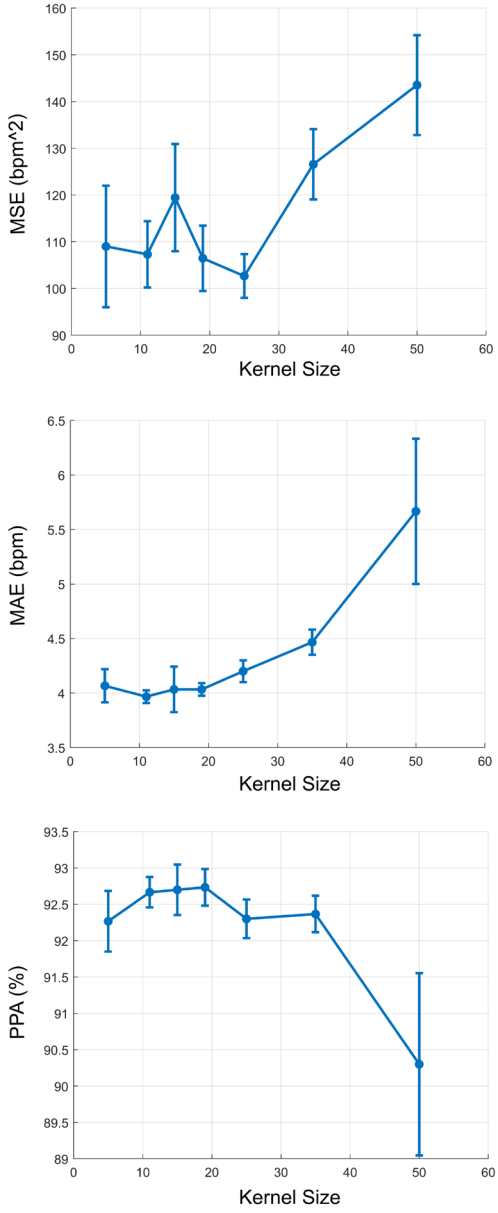



Fig. 3.6. Performance of the DICNN-LSTM network when varying the kernel size of all the convolutional layers on the validation set. At each data point the standard deviation error bars show variations in performance due to differences in the initialization of the weights and random selection of training samples.

By changing the size of the kernels different features could be learned and the complexity of the network changes. We are interested in investigating if the detection accuracy changes when the dimensions of the convolutional window are altered. Thus, we performed experiments where we varied the kernel size of all the convolutional layers in the encoder of our DICNN-LSTM network but kept all the other parameters fixed. The experiments were performed on our validation dataset. The different kernels sizes that we tried are the following: 5, 11, 15, 19, 25, 35, and 50. For each one of them we repeated training the network three different times and then we averaged the performances of these three networks. The reason for this is that there is some randomness in the training process caused by differences in the initialization of the weights and random selection of training samples. Therefore, even when two networks are trained with the same architectural- and hyper-parameters, these trained networks will not be identical.

Fig. 3.6 illustrates the performance of our network in terms of MSE, MAE and PPA when varying the kernel size. We notice that for smaller kernel sizes (up to 19) the performance of the network is relatively stable. This means that the kernel size does not seem to strongly affect the accuracy of the network and using any kernel size up to 19 is a good choice. However, for kernel sizes larger than 19 the performance starts to drop and becomes worse as the kernel size increases. This means that probably with large kernel sizes we miss details of some smaller features that are relevant for heart rate detection. Contrary to that, when using smaller kernel sizes, we detect not only smaller features but also larger ones since we have a relatively deep network with many convolutional layers stacked on top of each other. In addition, with the dilated inception scheme of our network we achieve a variety in the receptive fields of the network.

#### *3.3.1.2 Model ensemble*

Deep neural networks (DNNs) learn via a stochastic training algorithm that makes them sensitive to the training data and may learn a different set of weights each time they are trained that consequently produce different estimations. A way to reduce this high variance of the DNNs is to train multiple models and combine their outputs i.e. to use ensemble learning. Ensemble learning typically results in more stable and improved estimations when compared to a single model. There are many ensemble learning techniques like varying the training data or the model architecture and the combination of the estimations. We decided to choose the combination of the three best models, trained with different kernel sizes, and average their outputs.

Table 3.1 shows the performance of the three best performing models together with the performance of the model ensemble. The best performances are marked in bold. The convolutional layers of these three models have kernel sizes of 5, 15, and 19. The performance of the three models is comparable while the combination of the separate models leads to more accurate results.

Table 3.1 Performance of the best performing DICNN-LSTM models and the ensemble of them on the validation set.

| Model | MSE (bpm^2) | MAE (bpm) | PPA (%) |
|---|---|---|---|
| **Model1_kernel5** | 101.5 | 3.9 | 92.6 |
| **Model2_kernel15** | 107 | 3.8 | 93.1 |
| **Model3_kernel19** | 111.6 | 4 | 93 |
| **Model Ensemble** | **98.7** | **3.7** | **93.4** |



### 3.3.2 Fetal Heart Rate Reliability Classifier Performance

Fig. 3.7 illustrates the performance of the fetal HR reliability estimations for the validation dataset. The classifier achieved AUC of 0.91. We additionally calculated the 95% AUC confidence interval but since our sample size is very large (approximately 300,000) the upper and lower limits differ less than 0.01 from the AUC. By changing the threshold for our classifier, we can achieve the desired sensitivity and specificity for our problem. By choosing a high threshold we achieve higher sensitivity, meaning that most of the reliable data are indeed classified as reliable. Contrary to that, with a lower threshold higher specificity is obtained, leading to correctly identifying most of the unreliable data. We selected a threshold of 0.55 for our classifier and provided the TPR (sensitivity) and false positive rate (FPR) values as well in Fig. 3.7. The specificity can be calculated as 1-FPR. With the selected threshold the classifier can identify 91% of the reliable HRs and 61% of the unreliable ones.

We give slightly more importance to having a higher sensitivity than specificity for two reasons. First, we do not want to miss a lot of correctly estimated information to achieve a high coverage in the detection of the fetal HR values. Second, in clinical practice, even if some unreliable values are displayed on the screens of the clinicians, most of them look more like outliers, unconnected to the correct heart rate trace, and as such not taken into considerations in the clinical decision-making process.

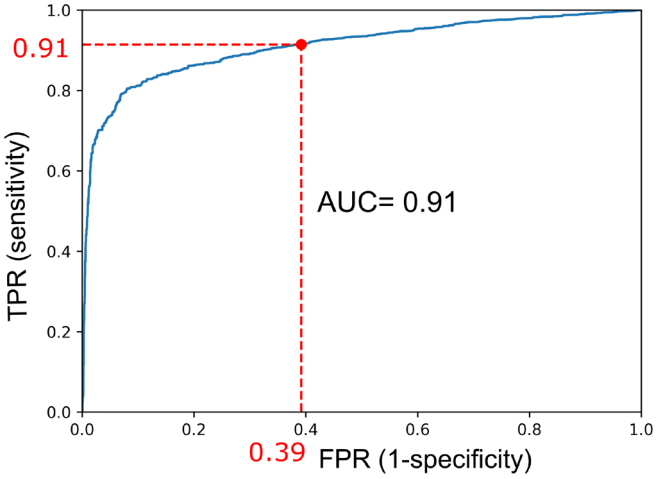



Fig. 3.7. Receiver operator characteristic curve (true positive rate (TPR) plotted against false positive rate (FPR)) of the fetal HR reliability classifier for the validation set. The area under the curve (AUC) is specified in the plot. The 95% confidence margin of error was less than +/- 0.01. The red dot on the plot indicates the operating point when the classifier threshold is set to 0.55. In that case sensitivity and specificity have values 0.91 and 0.61 respectively.

Table 3.2 provides the classification performance for the two test datasets. The classifier achieved AUC of 0.92 for our private test dataset and 0.83 for the Physionet dataset. This means that the classifier can differentiate relatively well between the reliable and unreliable fetal heart rates. The AUC of our private test dataset is similar to the one of the validation set but the AUC of the Physionet dataset is lower. The reason for this is that the Physionet dataset is highly unbalanced, containing 98.6% positive and only 1.4% negative examples. However, an AUC of 0.83 is still quite high meaning that our classifier generalized well to this dataset. The 95% AUC confidence interval for our private test set is very narrow (<+-0.01 difference from AUC) because of the large sample size. For the Physionet dataset, that is relatively small in size, the confidence interval for the AUC is [0.78, 0.86]. The TPR and TNR when the classifier threshold is set to 0.55 are also provided in Table 3.2. According to the evaluation on the test sets the classifier can identify 82-92% of the reliable heart rates and 66-75% of the unreliable ones.

Table 3.2 Performance of the fetal HR rellability classifier (for determining FPR and TPR classification threshold was set to 0.55)

| Private Test Dataset | | | Physionet Dataset | | |
|---|---|---|---|---|---|
| AUC | TPR | TNR | AUC | TPR | TNR |
| 0.92 | 0.92 | 0.66 | 0.83 | 0.82 | 0.75 |

### 3.3.3 Fetal Heart Rate Extraction Performance



Table 3.3 shows the performance of the fetal HR extraction network in comparison with the algorithm of Warmerdam et al. [127] for our private test dataset. In the table, we provided results both with and without incorporating the fetal HR reliability classifier. Our model achieved a PPA of 93.9% (91% for PPA_5), comparable with

Table 3.3 Fetal HR extraction performance on our private test dataset

| Algorithm | MSE (bpm^2) | MAE (bpm) | PPA (%) | PPA_5 (%) | Coverage (%) |
|---|---|---|---|---|---|
| **DICNN-LSTM** | 104.5/59.6* | 3.3/2.4* | 93.9/96* | 91/93.6* | 100 |
| **DICNN-LSTM + Fetal HR Reliability Classifier** | 49.4 | 2 | 97.3 | 95.4 | 87.9 |
| **Warmerdam et al. [127]** | 129.7 | 3.4 | 93.6 | 91.5 | 94.5 |

*Calculated only in the periods that warmerdam et at. [127] outputs a heart rate value.

the method of Warmerdam with a PPA of 93.6% (91.5% for PPA_5). However, the algorithm of Warmerdam identifies the periods of time that the extracted HR is unreliable and does not output an HR value for these. Thus, for this dataset it has coverage of 94.5%. To have a fair comparison we evaluated the performance of our method also for the case where we excluded exactly the same periods. In this case the PPA achieved by our network is 96%, clearly outperforming the algorithm of Warmerdam. In addition, the proposed method achieved significantly lower MAE and MSE, (2.4 vs 3.4 bpm for MAE and 59.6 vs 129.7 bpm$^2$ for MSE). When we discarded the unreliable fetal HR values, estimated by our reliability classifier, our PPA increased to 97.3% and the MAE and MSE decreased to 2 bpm and 49.4 bpm$^2$. However, our coverage fell to 87.9%, lower than Warmerdam's method. We additionally calculated the percentage of the time that the two methods in comparison agree on estimating the reliability of a fetal HR segment and we found that it is 90.1% on our test set.

We should note here that our private dataset is very challenging because it is partially obtained during the second stage of labor. At this stage, the uterine contractions are stronger and more frequent, and the woman is actively pushing leading to additional interferences from the abdominal muscles. As a result, the extracted fetal ECG signals have very low quality as they are strongly contaminated by noises. The significantly low SNR of the signals complicates the heart rate extraction, but by identifying these periods as unreliable, we are more confident that correct information will be presented to the clinicians. As an example, our reliability classifier labelled the fetal HR extracted from the signals of Fig. 3.1(a) as reliable and from the lower quality signals of Fig. 3.1(b) as unreliable.

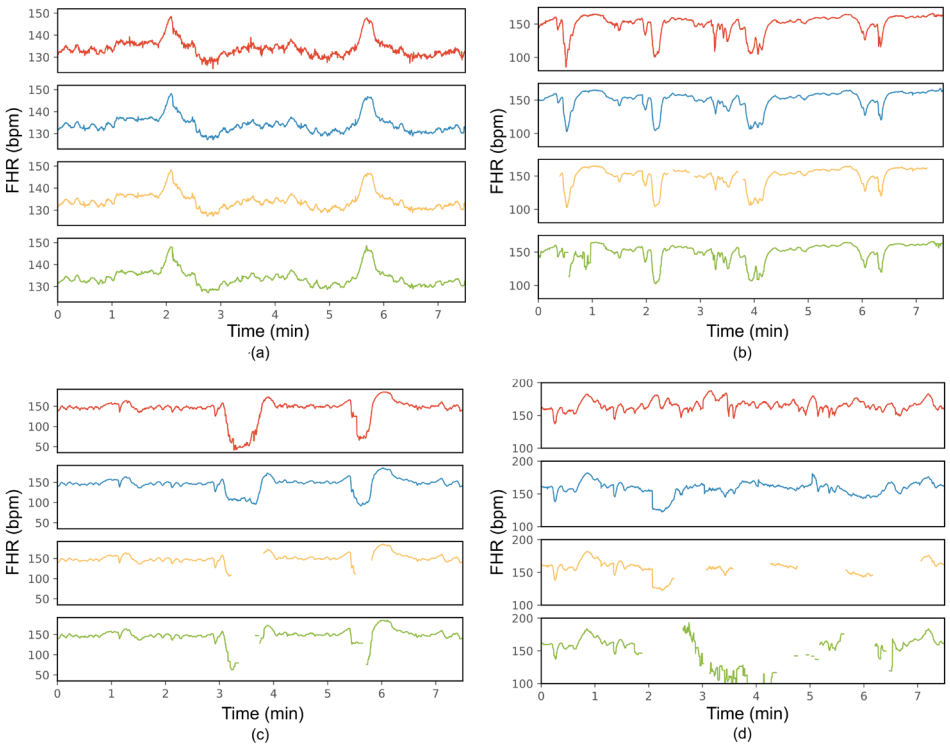



Fig. 3.8. Four fetal heart rate (FHR) segments from our private test set of 7.5 minutes duration each. For each case a-d: the first subplot (red) depicts the scalp FHR, the second (blue) the estimated FHR from the DICNN-LSTM network, the third (yellow) the estimated FHR from the DICNN-LSTM network in combination with the reliability classifier (only reliable FHR values are displayed) and the fourth (green) the FHR obtained from the algorithm of Warmerdam *et al.* [127] (only when an FHR value is outputted).

Fig. 3.8 depicts the fetal HR extraction by our network in comparison to the ground truth HR measured by the scalp electrode and the algorithm of Warmerdam. Four cases (a)-(d) of 7.5 min fetal HR segments are presented in the figure. Results are demonstrated both with and without employing the HR reliability classifier. Note that the vertical axis limits for each case (a)-(d) are not the same for better visualization. As can be seen in Fig. 3.8(a) our method follows very precisely the reference HR obtained by the scalp electrode and so does the method of Warmerdam. The reliability classifier determined successfully that the estimations for this segment were reliable. Fig. 3.8(b) is another example of almost perfect estimation. Here, the reliability classifier wrongly classified merely small signal portions as unreliable. In the case of Fig. 3.8(c) both methods failed to correctly determine the fetal HR deceleration parts. However, they successfully identified that the estimation was wrong. Fig. 3.8(d) depicts a case that was partially wrongly estimated by our network with the reliability classifier being also partially successful in that case.



Table 3.4 demonstrates the performance of the fetal HR extraction on set-A of the Physionet/Computing in cardiology dataset for our network in comparison to the methods of Warmerdam, Varanini and Behar. Our network slightly surpasses the methods of Warmerdam and Varanini for this dataset, achieving a PPA of 98.6% as opposed to 97.9% and 98.4% respectively. The algorithm of Behar achieved lower performance with PPA of 91%. However, when we lowered the tolerance for the calculation of PPA the algorithm of Warmerdam was the one to achieve the best performance.

Table 3.4 Fetal HR extraction performance on set-A of 2013 Physionet/Computing in cardiology challenge.

| Algorithm | MSE (bpm^2) | MAE (bpm) | PPA (%) | PPA_5 (%) | Coverage (%) |
|---|---|---|---|---|---|
| **DICNN-LSTM** | 14.2 | 1.6 | 98.6 | 95.6 | 100 |
| **DICNN-LSTM + Fetal HR Reliability Classifier** | 6.9 | 1.1 | 99.6 | 98.7 | 82 |
| **Warmerdam et al. [127]** | 30.8 | 1.5 | 97.9 | 96.5 | 100 |
| **Varanini et al. [104]** | 23.8 | 1 | 98.4 | 95.4 | 100 |
| **Behar et al. [125]** | 172 | 5.7 | 91 | 78.8 | 100 |

When we incorporated the reliability classifier our PPA increased to 99.6% but our coverage fell to 82% even though this dataset does not contain particularly low-quality fetal ECG signals. Moreover, our method achieved significantly lower MSE than the other methods, indicating that it provides fewer outliers. In terms of MAE the performances of all the methods are similar, apart from the one of Behar that is less accurate.

## 3.4 Discussion

Fetal R-peak detection in noninvasive fetal ECG recordings is demanding due to the low quality and the non-stationarity of the fetal ECG signals. In this work, we employed deep learning for directly determining the fetal HR from fetal ECG signals, without the need of R-peak detection. We proposed a deep hybrid dilated inception CNN-LSTM network that captures both short-term as well as long-term temporal HR patterns. To increase the reliability of our method, a classifier based on a CNN network was developed that estimates the accuracy of the detected fetal HR.

Initially, we performed experiments where we varied the kernel size of the convolutional layers of our DICNN-LSTM network. According to the results, we concluded that smaller kernel sizes are preferred, probably because features on smaller timescales are relevant for fetal HR extraction. Moreover, even when using small kernels, larger features are still exploited by our network due to the large receptive field achieved by using dilations in combination with a relatively deep network. In addition, we found that a model ensemble of several trained models with different kernel sizes leads to better performance than a single model. We decided to use the average of the three best performing models that we trained in our experiments as our fetal HR estimator.

The network achieved accurate estimations of the fetal HR for the most cases tested. However, mainly during the second stage of labor, there were cases that our method failed in the fetal HR extraction. Presenting false information to the clinicians can be very dangerous as this can lead to wrong decision making in critical moments during labor. Therefore, we developed a simple classifier that decides if the output of the DICNN-LSTM network is reliable so that in case of clinical application of the method only reliable information is presented to the clinicians. The developed classifier that is based on a simple deep CNN network managed to differentiate relatively well

between the reliable and unreliable HRs (AUC of 0.92 and 0.83 for our private and the Physionet test dataset respectively). However, since the hyperparameters of the classifier were not fully optimized, we are confident that the classification accuracy can be improved, and an even more robust system can be achieved.

Our DICNN-LSTM network achieved a positive percent agreement (PPA) of 93.9% on a dataset obtained during labor similar to a top-performing state of the art algorithm [127] (PPA of 93.6%) in the area of fetal HR extraction. However, since [127] excludes some unreliable fetal HR signal parts, when we also excluded the same parts our PPA increased to 96%. It is remarkable that Doppler ultrasound, which is widely adopted in clinical practice, according to the literature, achieves much lower PPA. In a study conducted to a population of 75 women in labor [153], Doppler ultrasound obtained a PPA of 73%, while abdominal fetal electrocardiography 81.7%. In a second study [154], which recruited 71 women in labor, PPA for ultrasound was measured as 63%, while for noninvasive fetal ECG it was 84.4%.

When we combined the DICNN-LSTM network with the HR reliability classifier the PPA was raised to 97.3%, while the coverage of our method dropped to 87.9%. We should note here that we selected a threshold of 0.55, however, this might not be the optimal threshold according to a clinician. For clinical applications, the threshold could be adapted so as the desired importance should be given to the sensitivity and specificity of the classifier. Moreover, as an alternative to the binary output of the classifier (i.e. suppress fetal HR output or not), the classifier could also be used to provide reliability of the fetal HR and have clinicians appraise this information themselves to trust or distrust the provided fetal HR. We could also use a different approach to assess the reliability of the fetal HR. The variance of the estimations of the different models used in the ensemble might give some information about the reliability of the outcome. Possibly for that we need a bigger ensemble than the one used in this work.

Next to our private dataset, we tested our network on a public set of Physionet achieving a PPA of 98.6% outperforming the algorithms of Warmerdam, Varanini and Behar that scored 97.9%, 98.4% and 91% respectively. However, the algorithm of Warmerdam achieved the best performance when computing the PPA metric with lower tolerance. As reported in [127], in the same dataset of Physionet, Warmerdam scored 99.6%, Varanini 98.6% and Behar 92.9% in terms of accuracy in R-peak detection. This implies that even though R-peaks can be detected accurately, the

resulting fetal HR that is calculated from the R-peaks can be inaccurate. The main reason for this is the relatively wide range around an annotated R-peak during which a detected R-peak is still considered accurate. In combination with the HR reliability classifier the PPA achieved by our network increased to 99.6% for the Physionet database. However, the coverage of our method fell to 82% because many correctly estimated heart rates were wrongly classified as unreliable.

Comparing the results obtained on the two different datasets we notice that the accuracy of the detection on the Physionet dataset is better. We believe that this is mainly due to the fact that our private dataset was partially obtained during the second stage of labor resulting in fetal ECG signals of lower quality as compared to the Physionet dataset. In addition, the Physionet dataset is significantly smaller in size than our dataset (68 min vs 14.5 h). Lastly, the algorithms used to extract the fetal ECG were different between the two datasets. We do not believe that this is the cause of the difference but can possibly account for a small part of it. Moreover, we notice that our approach clearly outperforms the one of Warmerdam on our private set (when compared in the same fetal HR signal parts) but in the Physionet dataset the performances are comparable. We believe that this is due to the limited ability of the model of Warmerdam to describe complicated accelerations and decelerations during labor. The DICNN-LSTM method, being capable of modeling more complex dynamics, provided consistent results in both datasets. However, we believe that even in cases when the performances are comparable, there is still value in our work because we demonstrate that a neural network can achieve similar performance to complex signal processing methods.

Finally, we have built a deep neural network to estimate fetal HR directly from the extracted fetal ECG signals. An alternative strategy would be to first extract features of the fetal HR manually and then apply a network on them for estimating the fetal HR. Although we believe that the encoder of our network can extract informative HR features, it would be interesting to investigate also this option.

### 3.4.1 Limitations of the study

This study has several limitations. First, the proposed fetal HR extraction network has many parameters such as the number of layers, dilation rates, number of nodes for each layer, type of layers, size of input and output signals etc. To select these parameters, first we divided our private dataset into training, validation and test

sets. Then, we chose the parameter values that led to the best performance on the validation set. An alternative way to choose the parameters is leave-on-out cross validation that is appropriate in cases of small datasets like ours. The reason we did not chose this approach is that it is computationally very expensive. Moreover, due to the large number of parameters we did not perform sensitivity analysis by varying their values (apart from the kernel size), although it could potentially reveal erratic behavior resulting from overfitting to the training set. Our training set is relatively small for training such a complex model. The training recordings are long, leading to many available data (54h), but they come from 16 subjects and a single acquisition system. However, our method was successfully tested not only on a test set from the same database but also to a completely unseen set from Physionet. The Physionet data constitute heterogeneous data obtained from multiple sources with different acquisition systems. It is promising that, even though our network was trained on data obtained from a single system, it was able to generalize to the Physionet dataset obtained by different systems. Nevertheless, in order to confirm our findings, validation on more diverse datasets is needed.



According to our results, as already mentioned, the DICNN-LSTM network sometimes failed to correctly estimate the fetal HR during the second stage of labor. The second stage of labor is the phase of labor that begins after full cervical dilatation and ends with the delivery of the baby. This phase lasts approximately 20 to 60 min. During this stage uterine contractions are stronger and more frequent, and the maternal abdominal muscles are particularly active due to intense pushing. The second stage is very critical for the fetus because exactly at this period it is often subjected to reduced oxygenation due to the increase in the intensity of the contractions. During this stage fetal HR decelerations often happen that are synchronous to the contractions. Our network frequently failed to precisely estimate these decelerations. Fortunately, the HR reliability classifier mostly classified these results as unreliable. However, it has been demonstrated that prolongation of the second stage of labor and the fetal HR decelerations is correlated with perinatal mortality and morbidity [155]. Therefore, the study of the fetal HR during this period is extremely important. Merely excluding these parts as unreliable should be avoided as much as possible. Possibly the capacity of our network is not sufficient to handle these cases of extreme noise and a separate network should be trained particularly for this stage. It might also be that we need more training data obtained during this stage, so our network learns better how to handle those segments. Finally, the error could lie much earlier in our signal processing chain, in the fetal ECG extraction step from the abdominal

recordings. Considering the high amount of muscle noise, we could argue that the maternal-fetal ECG separation might not have been correctly performed. This has a strong impact on subsequent processing.

In general, regarding our fetal HR reliability classifier, we need to stress that it would benefit from further improvement. Our main purpose in this work was to show that it is possible to classify unreliable fetal HRs with a relatively simple network. According to our experiments in the heterogenous Physionet dataset we found that the classifier rejected too much data of good quality. Unlike our classifier, the algorithm of Warmerdam successfully estimated the reliability of the result in this dataset. Thus, our results suggest that more investigation is required for robust fetal HR reliability estimation. We might do need two separate classifiers, one trained for second stage of labor and one for pregnancy and first stage of labor due to the distinct characteristics of the different stages. On the other hand, one could argue that we do not need an HR reliability classifier outside the second stage of labor since the network estimations are already accurate during the first stage of labor. These hypotheses need to be confirmed with additional experiments and more data.

Our method extracts the fetal HR sampled at 4Hz. The non-invasive fetal ECG can potentially provide beat-to-beat fetal HR information, while beat-to-beat HR variability analysis has been reported to provide important information about fetal distress [29]. However, according to a study [156], differences in variability indices between beat-to-beat and 4Hz sampled HR signals were found to not affect physiological changes observed during labor progression, while 4 Hz sampling provided better results in entropy indices. In addition, most central monitoring systems require the fetal HR values to be communicated at a frequency of 4 Hz and thus choosing for this frequency allows our method to be easier implemented in clinical practice.

Moreover, we provided the performance of our method in terms of PPA as in [153], [154] and PPA_5 for assessing the fetal HR extraction reliability with lower error tolerance. For an actual fetal HR of 140 bpm, the PPA metric considers acceptable HR values in the range [126,154], while PPA_5 in the range [133,147]. We need to stress that this error of 7 bpm (accepted by PPA_5) might still be relatively high for reliable calculation of fetal HR variability.

Another important issue to consider when developing a method intended to run online as well as when comparing different algorithms is computational complexity.

Our model ensemble is relatively complex having 137M trainable parameters and our reliability classifier has 237k. The advantage of our method is that it is a neural network and as such its deployment can exploit massive parallelization, which can be exploited via GPU computation. The other algorithms that we compared to might be less complex but do not have this benefit and consequently might be even slower in practice



## 3.5 Conclusion

In this study, we presented a deep dilated inception CNN-LSTM network for fetal heart rate extraction from noninvasively obtained fetal ECG signals. A quality assessment method, based on a CNN network, was additionally developed to exclude signal parts that will yield an unreliable fetal heart rate. The proposed method achieves accurate heart rate detection outperforming top-performing methods proposed in the literature. Our method may be used to achieve more reliable heart rate monitoring and contribute to the spread of noninvasive electrocardiography in clinical practice. Our results indicate that more complex algorithms and more data are needed to also make accurate fetal heart rate estimations during the second stage of labor.

# PART II Signal processing methods for fetal ECG enhancement

# CHAPTER 4

# ENHANCEMENT OF LOW-QUALITY FETAL ELECTROCARDIOGRAM BASED ON TIME-SEQUENCED ADAPTIVE FILTERING

Based on [JP-1]: Fotiadou E., van Laar J.O.E.H, Oei S.G., Vullings R., Enhancement of low-quality fetal electrocardiogram based on time-sequenced adaptive filtering. Med. Biol. Eng. Comput., 2018, 56, 2313-2323. doi:10.1007/s11517-018-1862-8

**Abstract**. Extraction of a clean fetal electrocardiogram (ECG) from noninvasive abdominal recordings is one of the biggest challenges in fetal monitoring. An ECG allows for the interpretation of the electrical heart activity beyond the heart rate and heart rate variability. However, the low signal quality of the fetal ECG hinders the morphological analysis of its waveform in clinical practice. The time-sequenced adaptive filter has been proposed for performing optimal time-varying filtering of non-stationary signals having a recurring statistical character. In our study, the time-sequenced adaptive filter is applied to enhance the quality of multichannel fetal ECG after the maternal ECG is removed. To improve the performance of the filter in cases of low signal-to-noise ratio (SNR), we enhance the ECG reference signals by averaging consecutive ECG complexes. The performance of the proposed augmented time-sequenced adaptive filter is evaluated both in synthetic and real data from PhysioNet. This evaluation shows that the suggested algorithm clearly outperforms other ECG enhancement methods, in terms of uncovering the ECG waveform, even in cases with very low SNR. With the presented method, quality of the fetal ECG morphology can be enhanced to the extent that the ECG might be fit for use in clinical diagnostics.

## 4.1 Introduction

The fetal electrocardiogram (ECG) can be used to monitor the fetal condition from early pregnancy until delivery [157]. Analysis of its waveform provides information that can assist clinicians in making more appropriate and timely decisions during labor [36]. This can be applicable in the case of fetal hypoxia that occurs as a result of oxygen deprivation of the fetus during parturition. The condition is often accompanied by acidosis and is life-threatening unless prompt interventions are undertaken to restore well-oxygenated blood to fetus. Fetal hypoxia is found to be associated with changes in the ECG waveform [158]. The recording of the fetal ECG can be carried out by an invasive electrode or by placing skin electrodes on the maternal abdomen. Unfortunately, in abdominal recordings the obtained signals are substantially contaminated by interferences and noise that vary depending on the gestational age, position of electrodes, skin impedance, etc. [157]. The main sources of interferences and noise include the maternal ECG, maternal and fetal muscle noise, powerline interference, base-line wander, movement artifacts and multiple layers of dielectric biological tissues through which the electrical signals must pass. The signals from some of these interferences and noises overlap both in time and frequency with the fetal ECG, complicating the extraction of the fetal ECG through conventional filtering techniques.

Despite the difficulties in acquiring fetal ECG signals noninvasively, a number of different techniques has been proposed in the literature such as neural networks [110], [159], wavelet based methods [160], [161], singular value decomposition [124], blind source separation [105], [106], [107], [162], adaptive filtering [108], [163], [100], [101], [109], as well as combinations of different algorithms [164], [103], [102]. Clifford et al. [123] reviews the key achievements and the follow up research generated as a result of the PhysioNet/Computing in Cardiology Challenge 2013 [122]. The challenge focused on fetal heart rate estimation and QT measurement in an automated way and managed to accelerate algorithm development in these areas [165], [146], [166], [104]. However, the extracted fetal ECG signal usually has a low signal-to-noise ratio (SNR) and additional processing is required to further enhance its quality. Beat-to-beat averaging [113] can be employed to improve the SNR of the signal. This approach however has the disadvantage that individual variations in pulse shape can be lost. In [101] an adaptive Kalman filter is developed that varies the number of complexes to be averaged according to the characteristics of the ECG signal. The filter is able to infer whether the ECG signal is corrupted by

noise or dynamic variations and adapt the number of averages accordingly, in that way preserving the pulse variations. However, the filter is not extensively evaluated in fetal ECG signals with low SNR.

The proposed method focuses on the postprocessing of the extracted fetal ECG signals, for enhancing their quality, based on adaptive filtering. Adaptive filters [167], [168], [109] have the ability to adjust their parameters autonomously and have been widely used to remove uncorrelated noise components when the noise characteristics are a priori unknown. Since the fetal ECG is a time-varying signal, adaptive filtering seems appropriate for estimating the fetal ECG. However, a Least Mean Squares (LMS) adaptive filter [169] has been proven inadequate to fulfill this role due to the low SNR, complexity, and non-stationarity of the fetal ECG. Despite the substantial background noise removal achieved by the LMS adaptive filter, the signal distortion was too severe, causing relevant ECG morphology to disappear.

The fetal ECG signal is highly non-stationary and an LMS filter is typically unable to track these rapidly varying non-stationarities. In [170], the time-sequenced adaptive filter (TSAF) has been suggested for the estimation of a class of non-stationary signals having a recurring statistical character and has shown good performance for detecting fetal ECG [108], [163]. The TSAF can be conceptualized as a bank of LMS adaptive signal enhancers that can achieve a rapidly varying impulse response. Fetal pulses differ among each other but have similar statistical properties. When the pulses are aligned according to a fiducial point, the statistics can be computed over the ensemble of pulses. The fiducial point at which the statistical properties of the signal renew is called the regeneration time. Each adaptive signal enhancer becomes an expert at filtering a specific signal segment between the regeneration times. The advantage of the method is that it does not require *a priori* knowledge of signal characteristics. However, an external input is needed to determine the regeneration times. In [163] and [108], the R-peak locations were detected to determine the regeneration times; [163] used a matched filter followed by peak detection, whereas [108] used a synchronized Doppler ultrasound signal.

The results of applying the TSAF to fetal ECG enhancement, presented in [163] and [108], show that, in case of poor SNR, the characteristic waves of the ECG cannot be distinguished after filtering. Moreover, as mentioned in [108], when the input SNR is relatively low, the effect of the filter's regeneration time can be seen in some cardiac cycles. This chapter presents an improvement of the TSAF for fetal ECG

enhancement and demonstrates the feasibility of time-sequenced adaptive filtering for enhancing the quality of the fetal ECG to the extent that the morphology of the signal is preserved. This improvement involves the enhancement of the reference inputs prior to their use in the TSAF, yielding an increase in the SNR of the TSAF output, and the elimination of aforementioned effects of the regeneration time via the use of overlapping filters.

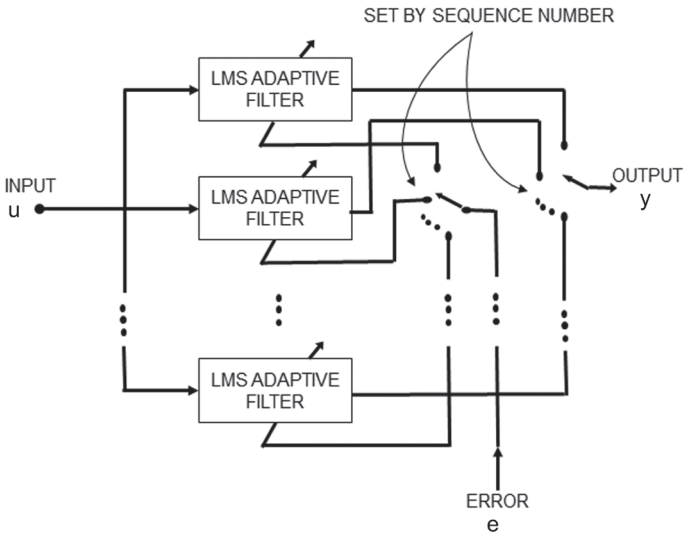



Fig. 4.1 Structure of the time-sequenced adaptive filter. The filter is realized as a bank of LMS adaptive filters. The sequence number determines which LMS filter is used at each point in time.

## 4.2 Materials and methods

### 4.2.1 The time-sequenced adaptive filter

The TSAF provides a separate LMS enhancer for each signal sample in an ECG cycle, as opposed to the original adaptive signal enhancer [167] that consists of a single LMS filter. Thus, there is a number of adaptive filters equal to the length of an ECG complex and each one is updated in every ECG cycle. The structure of the filter is depicted in Fig. 4.1. An external input, known as the sequence number, determines which LMS enhancer to be used at each point in time. The LMS adaptive filter is a

stochastic gradient-based algorithm that utilizes the gradient vector of the filter tap weights to converge to the optimal Wiener solution [109]. Each iteration of the LMS algorithm consists of the following steps:

1. Calculation of the filter output *y(n)*.

$$y(n)=w^T(n)u(n), \tag{4.1}$$

where $u(n)$ is the filter input and $w(n)$ are the filter weights. The size of both $u(n)$ and $w(n)$ is equal to Px1, where P is the filter order.

2. Estimation of the error signal which is needed to update the filter coefficients in the next step.

$$e(n)=d(n)-y(n) \tag{4.2}$$

where $d(n)$ corresponds to the n-th sample of the desired solution (in our case the signal to be enhanced $X_1$, see Fig. 4.2).

3. Update of the filter weights for the next iteration.

$$w(n+1)=w(n)+\mu e(n)u(n) \tag{4.3}$$

where $\mu$ is a step-size parameter.

The weights $w(n)$ are initialized as zero. The step-size parameter controls the rate at which the weights change. The choice of this parameter is critical for balancing the convergence of the filter versus its stability. Selecting a step-size that guarantees stability and ensures fast enough convergence is difficult due to the sensitivity of the LMS algorithm to scaling of its input [167]. To resolve this issue, in this work we use the Normalized Least Mean Squares (NLMS) filter instead of the LMS. NLMS is an extension of the standard LMS algorithm with the difference that it uses a time-varying step-size, yielding a faster convergence as opposed to the LMS algorithm [171]. The step-size used by the NLMS filter is defined as:

$$\mu'(n) = \frac{\bar{\mu}}{u(n)^T u(n)} \tag{4.4}$$

where $\mu$ is a scalar which allows for a change in the adaptation speed. The normalization of the step size with the power of the input signal $u(n)$ makes the algorithm insensitive to scaling.

## 4.2.2 Proposed method

### *4.2.2.1 Enhancement of reference signals*

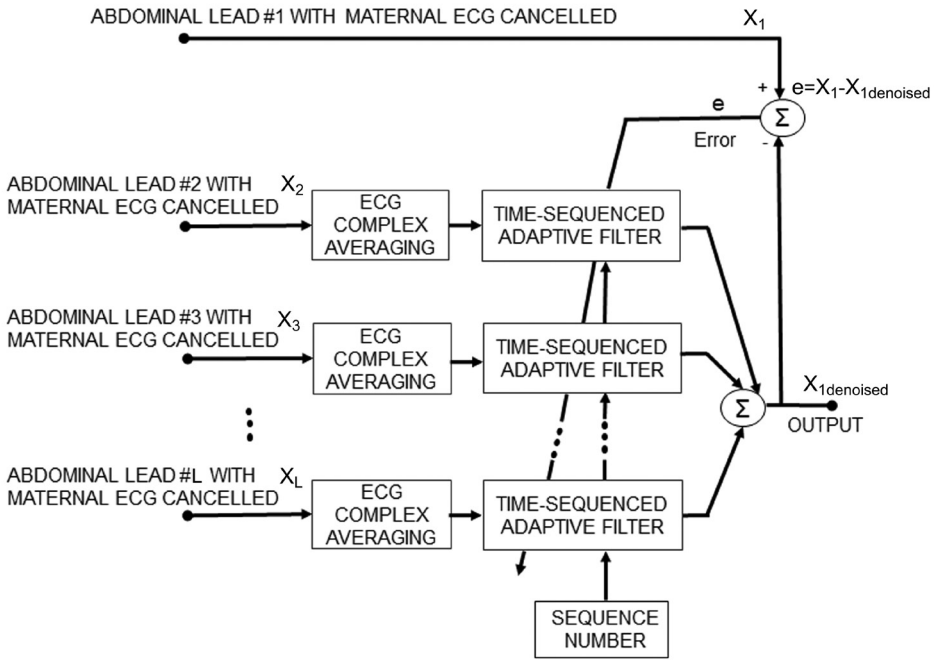



Fig. 4.2 Block diagram of the proposed L-channel time-sequenced fetal ECG enhancement method. The reference channels (#2-#L) are enhanced prior to filtering. The filter output is an estimate of the denoised fetal ECG signal of channel #1.

In this study, we have performed preprocessing and maternal ECG removal following the method developed by Varanini et al. [104]. The residual signals comprise of fetal ECG and a significant amount of remaining noise. Typically, the fetal ECG components among different channels are correlated, whereas some remaining noise components such as muscle noise are mostly uncorrelated. We propose the use of an augmented TSAF (aTSAF) to enhance the quality of multichannel fetal ECG by attenuating the uncorrelated noise (Fig. 4.2). The channel to be denoised is considered as the primary channel and the other channel(s) as the reference channel(s). In Fig. 4.2, the use of channel 1 as primary channel is just an example. Every channel that we wish to denoise can be considered as primary channel. The reference channels should be correlated with the primary one and should have sufficiently high SNR to yield satisfactory performance of the TSAF. To ensure an acceptable SNR of the reference inputs, in this work we enhance the quality of the reference channels

by ensemble averaging, prior to using them in the TSAF. The ensemble averaging includes the detection of the R-peak locations by the method described in [104], the alignment of successive complexes according to these locations, and the ensemble averaging of aligned ECG complexes. In the ensemble averaging, we used the ECG of 30 consecutive heartbeats, to preserve clinically relevant variations in the ECG and at the same time produce substantial enhancement of the ECG [172].

#### *4.2.2.2 Scheme for faster algorithm convergence*

After the reference channels have been enhanced, the TSAF combines them via a dynamic linear combination, where the weights in this linear combination are optimized for minimizing the error between the TSAF output and the primary input. To maintain a high rate of convergence, without risking instability of the process, a scheme is employed that is described in the work of Cano et al. [173]. According to it, the filter weights $w$ are updated not just in the prior ECG cycle but also in the current cycle where they are used for filtering. The scheme makes the assumption that immediate neighboring data samples are highly correlated and can be used to approximate each other; thus, the weights can be updated in the cycle being filtered. It is only employed for a given number of ECG cycles to assist the algorithm to achieve faster convergence.

#### *4.2.2.3 Reduction of the regeneration time effect*

The TSAF requires the knowledge of the location of the fetal ECG fiducial points in order to determine the regeneration times. The sequence number at each regeneration time is (re)set to 1 and increases by one for each data sample. Since the fetal PR interval is approximately 100 ms [174], the regeneration times are chosen to occur 160 ms before the detected R-peaks. In such a way, the start of the sequence occurs before the P-wave starts. The sequence length defines how many adaptive signal enhancers constitute the TSAF. Ideally, we want the sequence length to be equal to the ECG complex length, but this varies from cycle to cycle. Hence, we defined the sequence length to be 110% of the mean interval between the fetal R-peaks. In this way, it is ensured that in most cases the sequence length is bigger than the length of the ECG cycle, allowing all the samples in the cycle to be processed by a separate adaptive filter. This implies, however, that in most cases there are overlapping signal parts that are filtered twice. These parts are smoothed by averaging the contributions of both overlapping complexes. To be more specific, the contribution of the first ECG complex is gradually reduced while the contribution

of the second complex is gradually increased. Inevitably, there are cases where the length of the sequence is smaller than the size of the ECG complex. In this situation, gaps exist between successive complexes. The samples in these gaps are filtered by the same LMS adaptive filter that was used for the last sample in the sequence. The weights of this filter are updated only once in the current cycle. This guarantees that all the filters in a cycle converge with similar speed. By selecting the start and length of the sequence in the way described before, we avoid that the characteristic waves of the ECG complex fall outside the borders of the sequence in cases of ECG complexes longer than the sequence length. This is very important since the waves contain all the useful information about the physiological state of the fetus. Following the described approach, the regeneration time effect noticed in [108] can be significantly reduced.



### 4.2.3 Data description

The fetal ECG signals, even after the maternal ECG has been removed, are still affected by noise. It is hence impossible to have a gold reference (i.e. clean signal) that can be used to quantitatively validate the proposed algorithm. As a surrogate, in this study the proposed method is extensively validated based on simulated signals of the Fetal ECG Synthetic Database (FECGSYNDB) of PhysioNet [175], [176]. To illustrate the potential of our method on real data, we have included results obtained from data in the Abdominal and Direct Fetal Electrocardiogram Database of PhysioNet [177].

The FECGSYNDB consists of 1750 synthetic abdominal signals with 34 channels, sampling frequency of 250 Hz and duration of 5 minutes. The database includes 10 simulated pregnancies with seven different physiological events as shown in Table 4.1. The signal-to-noise ratio (SNR) of the simulated signals varies from 0 to 12 dB in steps of 3dB. In each simulation, signals are generated five times for statistical purposes. In this work, we use 6 channels (i.e. (1, 8, 11, 22, 25 and 32) for evaluation of our algorithm, as suggested by Andreotti et al. [175]. The signals simulating twin pregnancy ($5^{th}$ case, Table 4.1) are excluded from our analysis, since the proposed algorithm is not developed to handle this case.

The Abdominal and Direct Fetal Electrocardiogram Database contains multichannel fetal electrocardiogram recordings obtained from 5 different women in labour, between 38 and 41 weeks of gestation. Each recording contains 4 signals acquired from maternal abdomen and one scalp ECG signal. The recordings have duration of 5 minutes and are sampled at 1000 Hz.

Table 4.1. Description of the seven cases of physiological events of the synthetic signals of the Fetal ECG Synthetic Database

| Case | Description |
|---|---|
| Baseline | Abdominal mixture (no noise or events) |
| Case 0 | Baseline (no events) + noise |
| Case 1 | Fetal movement + noise |
| Case 2 | Acceleration or deceleration of maternal and fetal heart rate + noise |
| Case 3 | Uterine contraction + noise |
| Case 4 | Ectopic beats for both fetus and mother + noise |
| Case 5 | Twin pregnancy + noise |

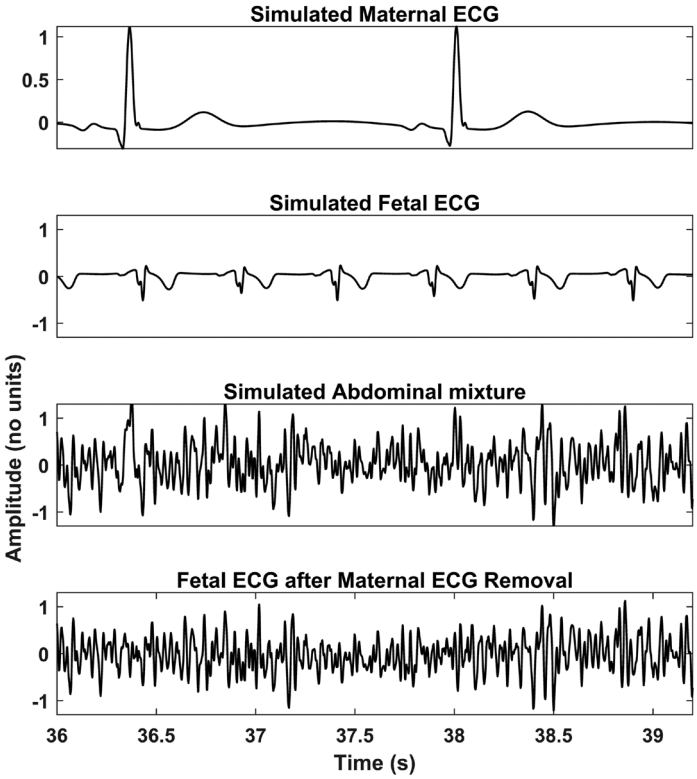



Fig. 4.3 An example of simulated signals of FECGSYNDB. First row shows the simulated maternal ECG, second the simulated fetal ECG, third the abdominal mixture, where noise is also added, and last the extracted fetal ECG after the method of Varanini [104] is applied. The displayed segment corresponds to the 3$^{rd}$ channel (3$^{rd}$ of the 6 channels in use) of the 10$^{th}$ simulated pregnancy, SNR 0, case 0 and 2$^{nd}$ repetition.

The signals of both databases are preprocessed before the proposed method is applied to them. First the signals are resampled to 500Hz to have a common reference. Then, the open-source algorithm of Varanini et al. [104] is applied to the signals. According to this algorithm, first the baseline wander and the powerline interference are removed. After that, the maternal ECG is estimated through Independent Component Analysis (ICA) and singular value decomposition and subsequently subtracted from the signals. Finally, a second ICA is employed to enhance the fetal ECG signal and two QRS detectors are applied in forward and backward directions to obtain the R-peak locations. Fig. 4.3 and Fig. 4.4 show an example of simulated data and real data respectively together with the result of the aforementioned preprocessing. In both cases the extracted fetal ECG after the maternal ECG removal contains a significant amount of remaining noise.



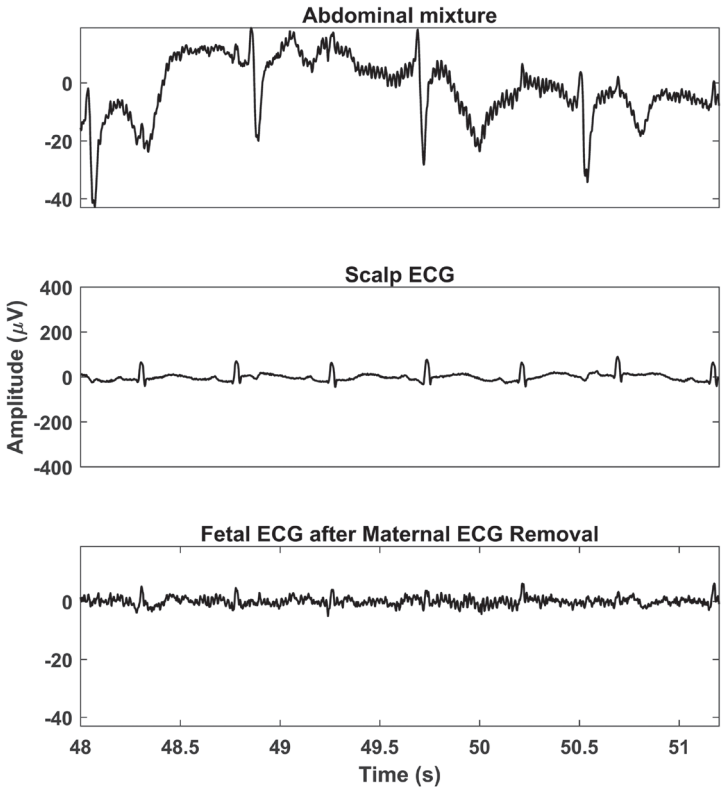



Fig. 4.4 An example of real signals of Abdominal and Direct Fetal Electrocardiogram Database. The first row shows a segment of the abdominal mixture of channel 1 for recording 'r07'. The second displays the simultaneously recorded scalp fetal ECG, while the last the extracted fetal ECG after the method of Varanini [104] is applied.

### 4.2.4 Performance measure

The performance of the method is assessed based on the SNR improvement in the filter output when compared to the input signal. The metric is computed as follows:

$$SNR_{imp}[dB] = 10log_{10}\frac{\sum_{m=1}^{M}\left|X_{noisy_m}-X_{clean_m}\right|^2}{\sum_{m=1}^{M}\left|X_{denoised_m}-X_{clean_m}\right|^2} \quad (4.5)$$

where $X_{clean}$ denotes the original clean fetal ECG signal, $X_{noisy}$ the noisy signal, $X_{denoised}$ the enhanced one and M the length of the signals. The metric is measured for each channel and subsequently summed over all ECG channels and over all signals.

The proposed method is compared to several other ECG enhancement algorithms. The first one is a wavelet denoising method that decomposes the signal, thresholds the detail coefficients and reconstructs the signal to obtain its enhanced version. The order 6-Daubechies wavelet is selected as mother wavelet because of its similarity to an actual ECG. A fixed threshold is used which is estimated by the minimax principle [178]. Secondly, our algorithm is compared to the ensemble averaging over 30 consecutive ECG complexes. The next reference method is the adaptive Kalman filter described in [101], where the number of the ECG complexes to be averaged is adapted according to the signal characteristics. An additional comparison is made with the multichannel NLMS adaptive signal enhancer [167] and with the NLMS adaptive signal enhancer for which the reference channels are enhanced via our suggested averaging method (we will refer to this method as augmented NLMS; aNLMS), as described in subsection 4.2.2.1. Finally, our method is also compared to the time-sequenced adaptive filter (TSAF) without preprocessing of the reference channels but including compensation for regeneration time effects.

## 4.3 Results

### 4.3.1 Parameter optimization

The proposed algorithm has few parameters that must be chosen: the parameter $\overline{\mu}$ that is used in the calculation of the step size of NLMS algorithm and P, the length of the adaptive filters. The value of $\overline{\mu}$ is critical for the algorithm's performance since a wrong choice will restrain the algorithm from convergence. However, once

chosen appropriately, this parameter does not need further readjustments because of the scaling-tolerance of the NLMS algorithm. The choice for P is less critical.

In order to select these two parameters, the simulated signals of the FECGSYNDB are separated in a training and a test set. The first 5 simulated pregnancies are used as training data while the last 5 as test. The training set is used to optimize the parameter values, while the test set is used to evaluate the performance of the algorithm. An iterative procedure is performed over several parameter values and the ones that produce the highest SNR improvement in the training set are finally selected. For $\overline{\mu}$ a search is performed between the values 0.0005 and 0.1 with 8 logarithmic steps. Regarding the filter length the values of 50 and 100 to 800 with steps of 100 are used. For TSAF, NLMS and aNLMS the parameter values are selected similarly. The parameter values that produce the best results for all the algorithms are 0.005 for $\overline{\mu}$ and 100 for P.



Table 4.2. The number of the signals with successful R-peak detection for each different case of FECGSYNDB (125 synthetic signals per case of the test simulated pregnancies, 6-10).

| Case | Number of signals |
|---|---|
| Baseline | 121 |
| Case 0 | 94 |
| Case 1 | 106 |
| Case 2 | 84 |
| Case 3 | 78 |
| Case 4 | 30 |
| Total | 513 |

### 4.3.2 Evaluation on Fetal ECG Synthetic Database

Our method is evaluated on data from the Fetal ECG Synthetic Database (FECGSYNDB) described in Section 4.2.3. Since the proposed algorithm requires the knowledge of the fetal R-peaks, only the signals for which at least 80% of the fetal R-peaks are detected within an error of 50ms from the actual R-peak [179] are included in the evaluation of the algorithm. Table 4.2 shows the number of signals with successful R-peak detection for each different case of the FECGSYNDB. The numbers correspond to the simulated pregnancies 6-10. As expected in the baseline case, where no noise is added to the abdominal mixture, the extracted fetal ECG signals are of relatively good quality and the R-peak detection succeeds in almost all

the cases. Case 4, where ectopic beats are simulated, is the most challenging one and the R-peak detection succeeds only in 30 out of 125 occurrences.

Fig. 4.5 illustrates the performance of the proposed method in comparison to the denoising algorithms mentioned in Section 4.2.4 for the different SNR values of the input signals. The input SNR refers to the SNR of the signals after the maternal ECG has been removed and this SNR is averaged over all channels and measurements. From Fig. 4.5, it can be observed that, for very low SNR values of the input signals,

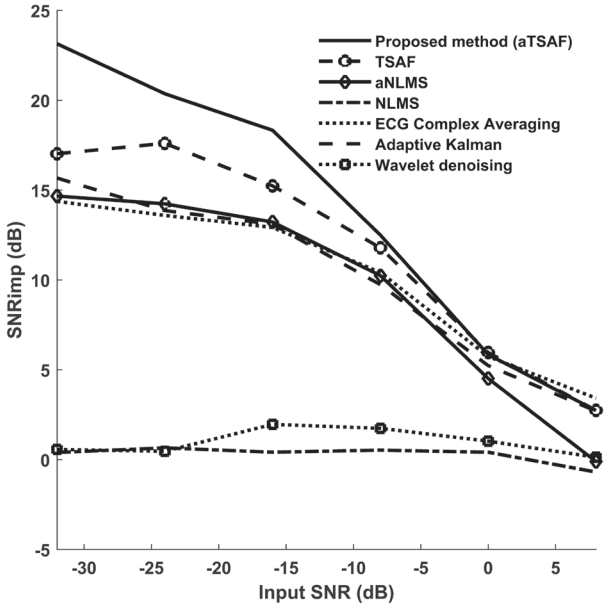



Fig. 4.5 Performance of different fetal ECG enhancement algorithms in comparison with the proposed algorithm by means of improvement in SNR of the output when compared to the SNR of the input signal. The results correspond to the average over all channels and cases for the synthetic signals of FECGSYNDB (simulated pregnancies 6-10).

the proposed algorithm clearly outperforms the other algorithms. For input SNR less than -15 dB the suggested method provides an additional SNR improvement of at least 3 dB over the best performing other method. For high SNR values (more than 0dB) the conventional TSAF algorithm produces a similar result with our algorithm since there is no need to filter the reference channels. It is worth observing the increase in performance of the aNLMS algorithm over the conventional NLMS. Without our proposed preprocessing of reference signals, NLMS performs significantly worse. From Fig. 4.5, it can be seen that the averaging of ECG complexes produces good results when the SNR is higher than -10. However, for lower SNR values of the

input signals, the SNR improvement is significantly lower than that of the proposed algorithm. Apparently, different ECG segments should be considered separately to truly enhance the ECG. Plain averaging works on the entire ECG complex and treats it as a whole. On the other hand, our method uses different filtering schemes for different parts of the ECG.

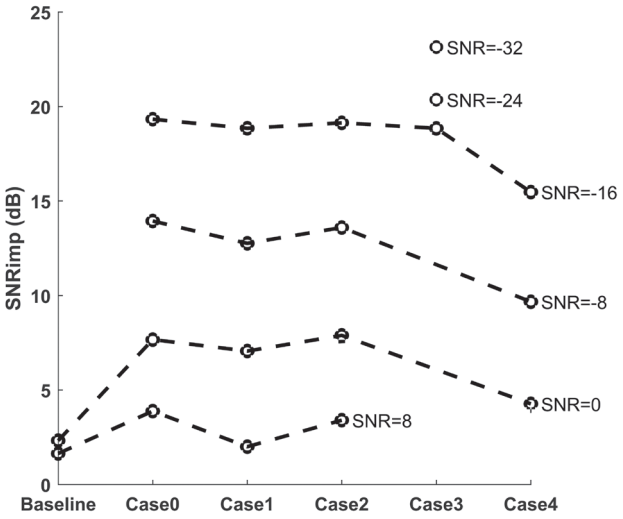



Fig. 4.6 Performance of the proposed aTSAF algorithm for each different case and input SNR of the synthetic signals of FECGSYNDB. The input SNR corresponds to the average of all channels and after the maternal ECG is removed. The circles, 'o', show the performance only for the cases that there are available input signals with the specific SNR (-32dB to 8 dB).

Fig. 4.6 depicts the performance of the suggested method for each different case separately and for all the range of available input SNRs. The input SNR values are clustered in 6 groups of values -32 dB to 8 dB in steps of 8. Input signals with SNR from all these groups are not available for every case, since the SNR is measured after the maternal ECG is removed from the simulated signals. As we can see in Fig. 4.6, the efficiency of aTSAF for each specific input SNR is similar for the different cases apart from case 4 and the baseline case. Case 4, as mentioned before, is a challenging case that simulates the presence of ectopic beats. In this case, abrupt changes in the morphology of the ECG signals occur, for short amount of time, making it relatively impossible for the algorithm to adapt fast enough. In this case, our algorithm adapts to regular heartbeats but is incapable of tracking the sudden changes. Regarding the baseline case, less improvement in SNR is achieved compared to the other cases. The reason is that this case does not contain added noise that can be removed by the aTSAF.

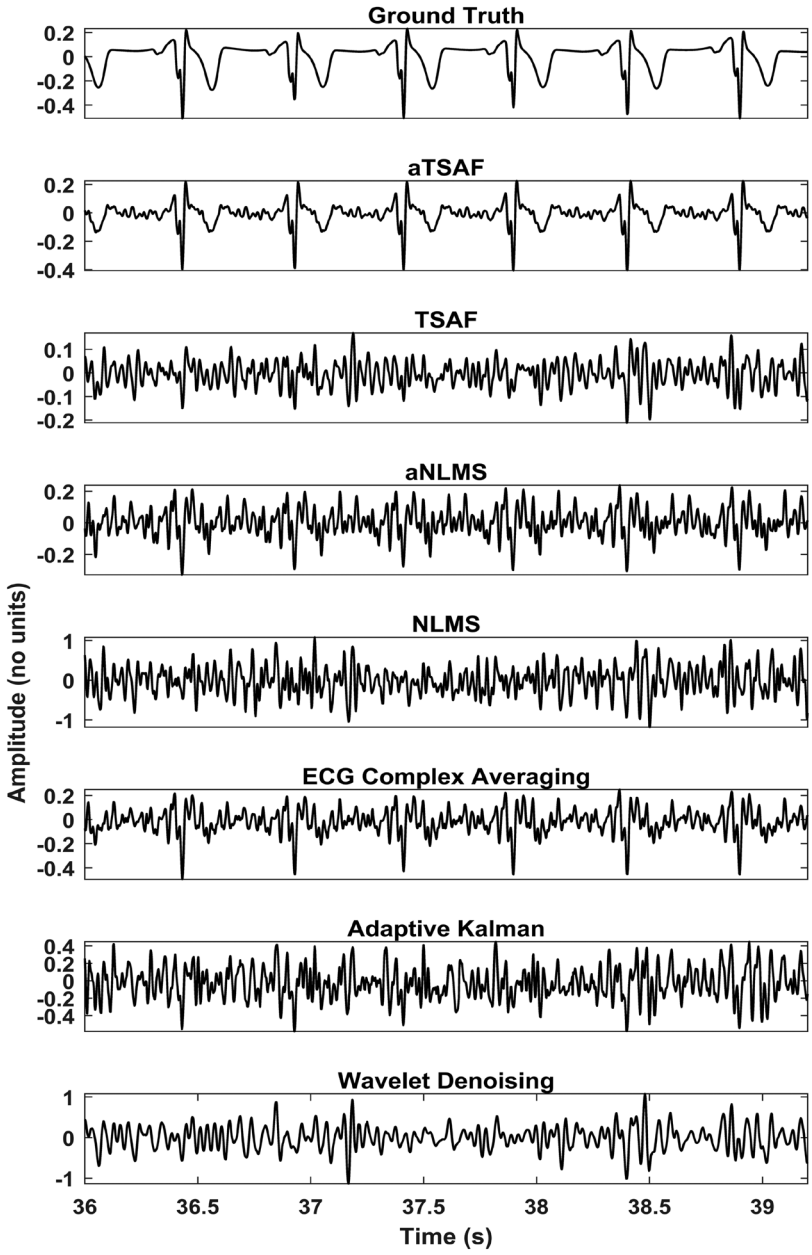



Fig. 4.7 Enhancement result of the different algorithms for the simulated signals described in Fig. 4.3. In the first row the simulated clean fetal ECG signal is displayed for comparison. Only the proposed method (aTSAF) provides high quality result in this challenging case of low-SNR input signals (-18dB on average).

The results of processing the simulated signals, which were described in Fig. 4.3, by our aTSAF method are illustrated in Fig. 4.7. For the depicted recording, the average SNR of the 6 channels after the maternal ECG has been removed is -18dB. The figure demonstrates that the proposed aTSAF method is capable of suppressing the noise to the extent that the individual waves in the ECG complex become distinguishable. Moreover, there is virtually no distortion of the ECG signal. In this particular case, the suggested algorithm clearly outperforms the other methods. The additional step of enhancing the quality of reference channels – which is the main contribution of this study – makes a significant difference in the performance of the TSAF algorithm, especially for cases with low-quality signals. For the other algorithms, the ECG complex averaging and the aNLMS produce the best results but show still a significant amount of noise in the filtered signals. For the other methods, the morphological information of the ECG cannot be seen.



### 4.3.3 Evaluation on the Abdominal and Direct Fetal ECG Database

In this subsection, the performance of the same algorithms evaluated in Section 4.3.2 is evaluated on actual data from the Abdominal and Direct Fetal ECG database. Concerning the algorithm parameters, the same parameters that were optimized in FECGSYNDB are used. In the real data, because of the lack of ground truth data, the performance of the various methods is evaluated qualitatively, as opposed to quantitatively for the simulated data. To demonstrate the potential of abdominal ECG recordings, we have shown the scalp ECG, after high-pass filtering for baseline wander removal, as well in our figures (see e.g. Fig 4.8). It should be noted that the scalp ECG is a different ECG lead than the abdominal leads and that they should not be identical, even in case of perfect enhancement. Nevertheless, the individual ECG segments should coincide between abdominal leads and scalp lead.

Fig. 4.8 illustrates the results of various algorithms for the fetal ECG enhancement of channel 1 of recording 'r07', which was already depicted in Fig. 4.4. As seen in Fig. 4.8, the ECG signal filtered by the proposed aTSAF algorithm is relatively free from noise and the individual waves correspond well to those in the scalp ECG. Preprocessing of the reference channels appears to have a substantial contribution to the performance of both the TSAF and NLMS algorithms, making even the small waves distinguishable. Without our proposed preprocessing these characteristic waves are often not visible, either because they were suppressed by the filter or because the noise was not suppressed enough.

Fig. 4.9 demonstrates the results of the processing of the 4 other recordings of Abdominal and Direct Fetal ECG database with the proposed algorithm. The corresponding scalp ECGs are presented together with the results for channel 1 of each recording. In all cases the suggested method produces a relatively clean ECG signal with morphology that corresponds relatively well to that of the scalp ECG, especially for the P-wave and QRS-complex.

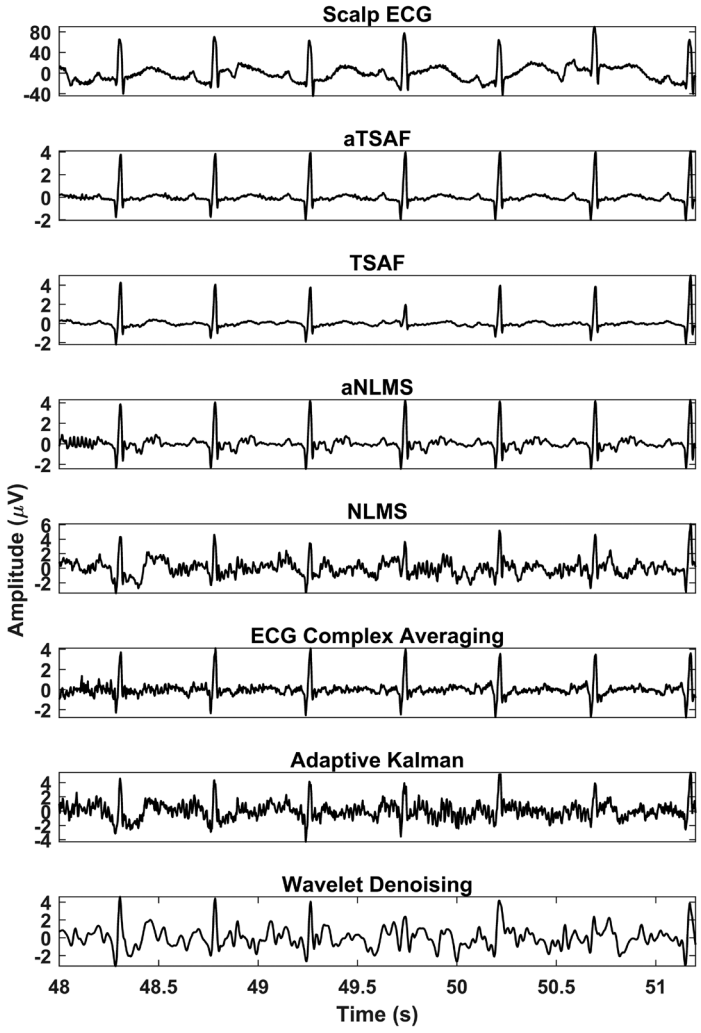



Fig. 4.8 Comparison of the enhancement results of the different algorithms for channel 1 of recording 'r07' of Abdominal and Direct Fetal Electrocardiogram Database. In the first row the simultaneously recorded scalp ECG is presented. The noisy signal is shown in Fig. 4.4. In the output of the proposed method (aTSAF) the characteristic ECG waves are clearly visible and correspond well to those of the scalp ECG.

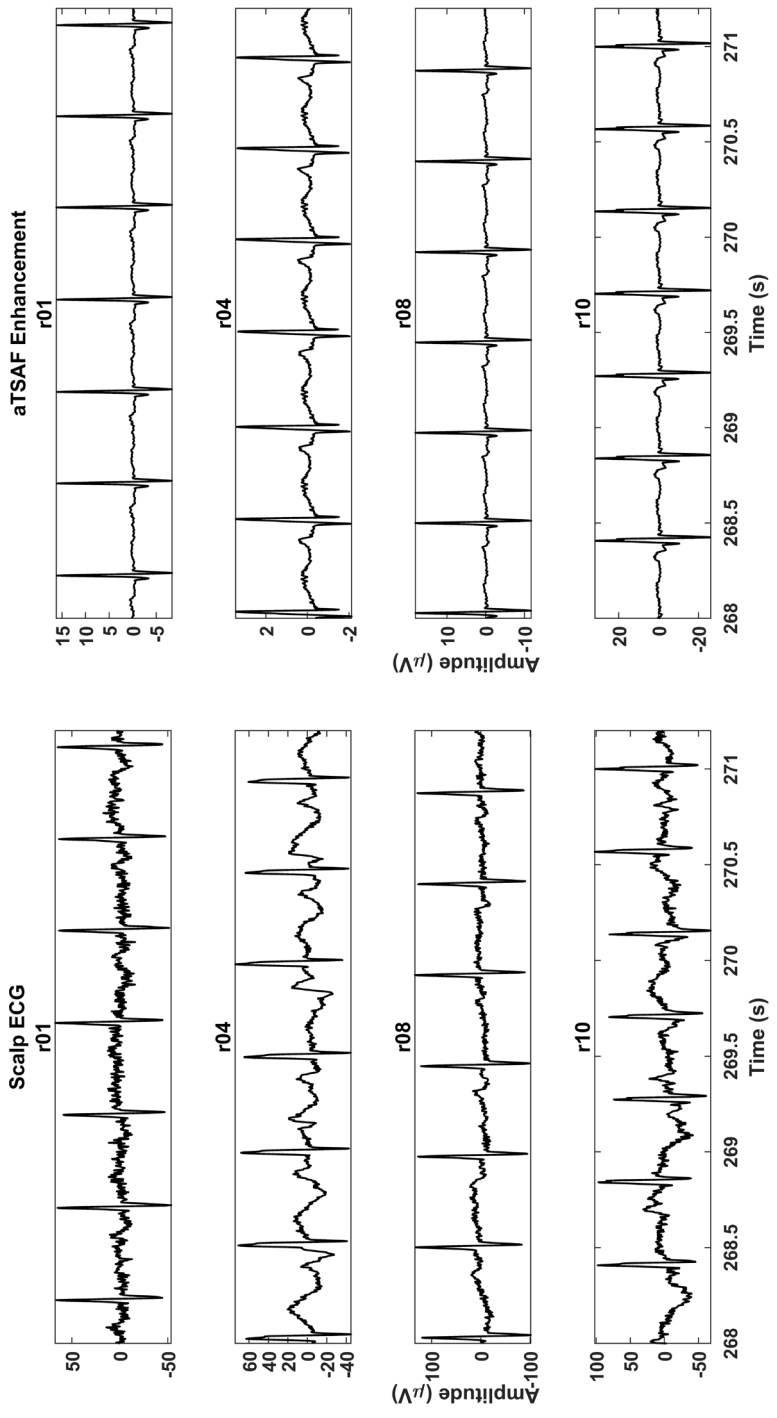



Fig. 4.9 The enhancement result of the proposed method (aTSAF) for the recordings 'r01', 'r04', 'r08' and 'r10' of the Abdominal and Direct Fetal Electrocardiogram Database. In the left column the corresponded scalp ECGs are displayed. In all cases the enhanced signals are clean and the morphological signal characteristics are similar to those of the scalp ECGs.

## 4.4 Discussion

A method is presented here for postprocessing of the extracted fetal ECG after the maternal ECG is removed. In most cases, the extracted fetal ECG still contains a substantial amount of noise that impedes the interpretation of the morphology of the ECG signal by clinicians. Usually, the QRS complex can be detected without further processing of the fetal ECG, due to the high amplitude of the R-peak. However, the smaller waves, like the P and T waves, can often not be readily distinguished. Thus postprocessing of the extracted fetal ECG signal, to enhance its quality, becomes of paramount importance. For this purpose, in the proposed method, the TSAF is improved by increasing the quality of the reference channels and by exploiting overlapping filters to minimize the effect of regeneration times. The improved filter is found to be effective in reducing major components of noise. But more than that, the main contribution of the filter is that after the filtering the signal morphology is retained to the extent that even the small signal waves can be visually distinguished. As a plus, the method is relatively insensitive to the choice of parameters and as such is rather generally applicable.

A limitation of the method is that an estimate of the R-peak locations is required to determine the regeneration times of the filter. However, this does not necessarily impede the use of our filter. First of all, a lot of valuable work has been already done in the area of fetal R-peak estimation [165], [146], [166], [158], [104] with very promising results. Second, in a practical application the user could be informed to, in case of missing R-peaks, distrust the output of the filter. Also, there is the possibility of using other measurement modalities to yield the regeneration times, like synchronous Doppler echocardiography as was proposed by Adam et al. [108].

Another shortcoming of the method is in cases of arrhythmia and ectopic beats. In these cases, sudden and brief changes are happening to ECG signal morphology and the filter is not able to adapt fast enough. As a consequence, it will only adapt to the morphology of regular beats and would be incapable of tracking the abrupt changes. This effect is also caused because of the averaging of ECG complexes for the enhancement of reference channels. Because of the averaging, brief variations in the ECG morphology are filtered out. However, without this enhancement step the TSAF is unable to efficiently remove the noise. Moreover, the target application of this work is to enhance the fetal ECG quality for detecting hypoxia. The STAN method [172] that is used for fetal monitoring with an invasive electrode, averages

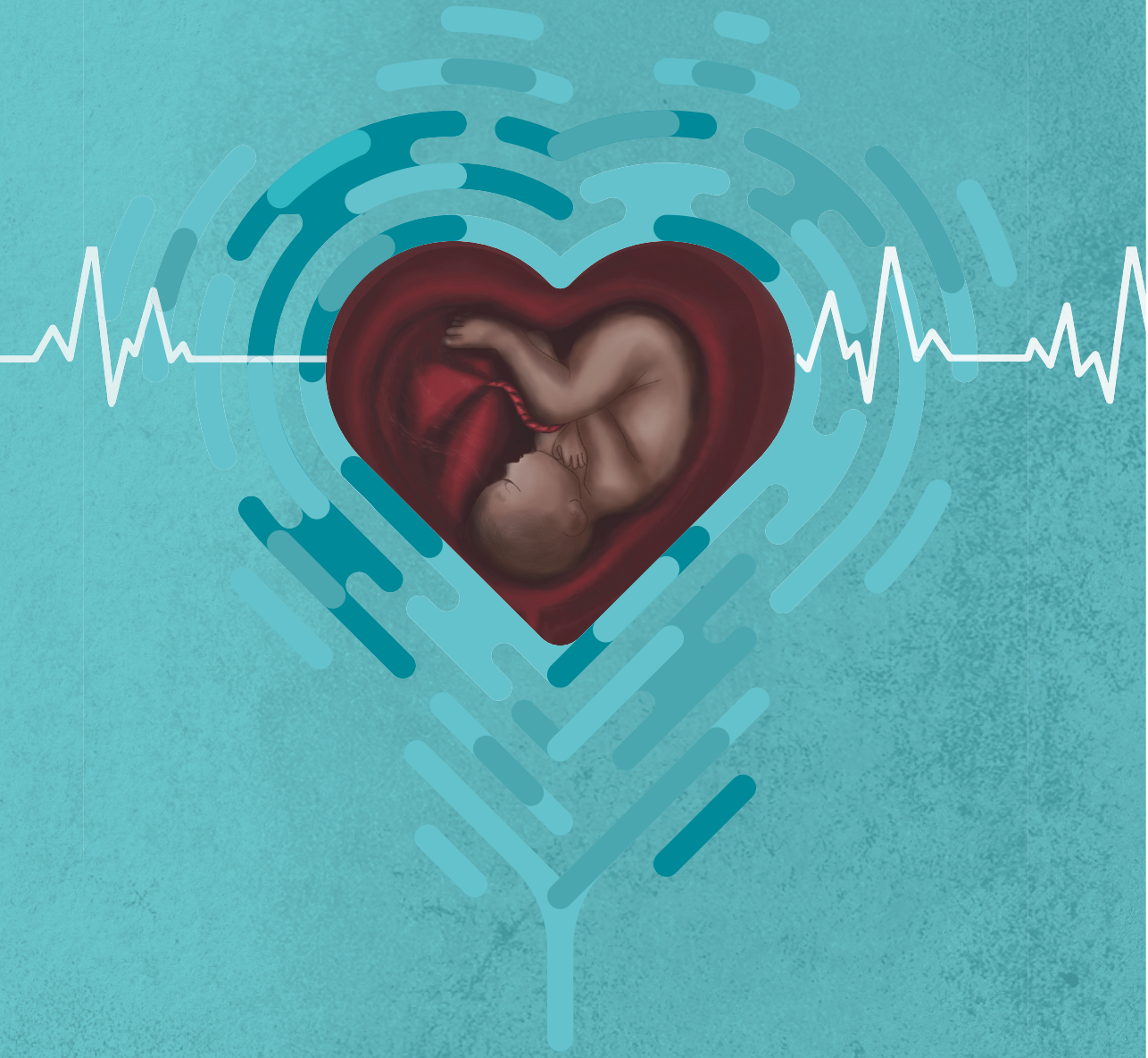
30 consecutive heartbeats for ECG signal enhancement. For a typical fetal heart rate of about 140 bpm, 30 heartbeats correspond to a time interval of 13s. Transient or structural changes in the ECG waveform that have clinical relevance with respect to developing hypoxia are hypothesized in the STAN methodology to occur over longer time scales. Consequently, averaging 30 consecutive ECG complexes preserves clinically relevant variations in the ECG and at the same time yields a substantial enhancement of signal quality. Based on the same reasoning, we also average 30 consecutive ECG complexes to enhance our reference channels, while maintaining relevant information for detecting hypoxia.



The high-quality signals that the proposed method delivers can give the opportunity to clinicians to measure the exact timing of different morphological features of the ECG signal. Besides, it can facilitate and advance the research towards automated detection of fetal ECG intervals and segments. Extracting morphological features from the ECG signal allows for the estimation of the wellbeing of the fetus. Fetal acidosis is known to affect ECG morphology [36], [180], while asphyxia of the fetus is thought to be associated with changes in the P wave, PQ interval and ST segment [181]. Moreover, fetal growth might also influence the timing of ECG waves [182].

## 4.5 Conclusion

This chapter presented a method to improve the performance of the time-sequenced adaptive filter for fetal ECG enhancement. In the proposed method, the quality of the reference channels is enhanced prior to filtering via ensemble averaging of multiple consecutive ECG complexes. The evaluation of our filter, both on simulated and real fetal ECG signals, shows that the proposed algorithm outperforms the conventional time-sequenced and the NLMS adaptive filtering techniques. Our results indicate that preprocessing of the reference channels provides a more accurate estimate of the underlying fetal ECG signal. The proposed algorithm can reveal the characteristic waves of the fetal ECG signal, even in cases with relatively low SNR. A limitation of our method it that, in case of rapidly changing ECG morphology, for instance in presence of ectopic beats, our algorithm is unable to track these changes, yielding a suboptimal performance. Moreover, an estimate of the position of the fetal pulse locations is required to determine the regeneration times of the filter. Future work could focus on performing automated extraction of relevant morphological features such as PR intervals, QT intervals, and ST segments.



# CHAPTER 5

# END-TO-END TRAINED ENCODER-DECODER CONVOLUTIONAL NEURAL NETWORK FOR FETAL ELECTROCARDIOGRAM SIGNAL DENOISING

Based on [JP-2]: Fotiadou E, Konopczyński T, Hesser J, Vullings R., End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising. Physiol. Meas.,2020, 41:15005. doi: 10.1088/1361-6579/ab69b9

**Abstract**. *Objective:* Noninvasive fetal electrocardiography has the potential of providing vital information for evaluating the health status of the fetus. However, the low signal-to-noise ratio of the fetal electrocardiogram (ECG) impedes the applicability of the method in clinical practice. Quality improvement of the fetal ECG is of great importance for providing accurate information, enabling support in medical decision making. In this chapter, we propose the use of artificial intelligence for the task of one-channel fetal ECG enhancement as a postprocessing step after maternal ECG suppression. *Approach:* We propose a deep fully convolutional encoder-decoder framework, learning end-to-end mappings from noise-contaminated fetal ECGs to clean ones. Symmetric skip-layer connections are used between corresponding convolutional and transposed convolutional layers to help recovering the signal details. *Main results:* Experiments on synthetic data show an average signal-to-noise ratio (SNR) improvement of 7.5dB for input SNR in the range of -15 to 15dB. Application of the method on real signals and subsequent ECG interval analysis demonstrates a root mean squared error of 9.9 and 14ms for the PR and QT interval, respectively, when compared with simultaneous scalp measurements. The proposed network can achieve a substantial noise removal both on synthetic and real data. In cases of highly noise-contaminated signals some morphological features might be unreliably reconstructed. *Significance:* The presented method has the advantage of preserving individual variations in pulse shape and beat-to-beat intervals. Moreover, no prior knowledge on the power spectra of the noise or the pulse locations is required.

## 5.1 Introduction

During pregnancy and labor, monitoring of the fetal heart condition is of paramount importance. Fetal monitoring can support medical decision taking, while an early disease diagnosis can increase the effectiveness of the appropriate treatment. Since the fetus is well protected within the woman's womb, it is inaccessible for direct measurements. Nowadays, the gold standard for fetal heart assessment is cardiotocography that provides visual representation of the fetal heart rate together with uterine contractions [183]. However, cardiotocography is prone to signal loss and does not provide the heart rate with beat-to-beat variations. The noninvasive fetal ECG, derived from abdominal electrodes, has the potential to provide beat-to-beat heart rate information [120] with the added possibility of assessing ECG morphology such as the PR and QT intervals or ST segments. Despite the ease of its applicability, fetal ECG signals are substantially contaminated by interferences and noises that vary depending on the gestational age, position of electrodes, skin impedance etc. [157]. Most significantly, the signals are masked by the maternal ECG and background noises caused by the abdominal and uterine electromyogram. Additional noises and interferences, such as electrode movement and powerline interference and the multiple layers of dielectric biological tissues that the signals must pass further lower the signal-to-noise ratio (SNR) of the fetal ECG. The high amount of the interfering sources that are typically non-stationary and overlap with the fetal ECG in time and frequency domain, render the fetal ECG extraction and heart rate detection challenging signal processing tasks and limit the applicability of the method in clinical practice.

Despite the advances in adult ECG signal processing, analysis and interpretation of the fetal ECG is still in its infancy. Various techniques were proposed in the literature for extracting the fetal ECG from noninvasive abdominal recordings. The main methods include adaptive filtering [108], [100], [109], [58], blind source separation [105], [106], [107], [162] and template subtraction [98], [99], [97] . Clifford et al. [123] gave a thorough review of the key achievements and the follow-up research as a result of the PhysioNet/Computing in Cardiology Challenge 2013 [122]. The aim of the challenge was to encourage heart rate estimation and QT interval measurements in an automated manner and succeeded in stimulating research in these areas [165], [146], [104], [166]. In general, a five-step approach is followed by most algorithms that includes preprocessing, maternal ECG estimation and subtraction, fetal heart

rate estimation, and postprocessing of heart rates. Usually, the extracted fetal ECG has still a low SNR and additional processing is required to further enhance its quality. A traditional approach of improving the SNR of the signal is beat-to-beat averaging [113] at the expense of loss in the individual variations in pulse shape. In a previous work [184] the authors used a time-sequenced adaptive least mean squares (LMS) filter, as a postprocessing step, to enhance the quality of the extracted multichannel fetal ECG. Despite the substantial noise removal, the method requires the knowledge of the R-peaks location prior to use and is not designed to handle arrhythmia cases.

Recently, deep neural network models (DNNs), like stacked denoising autoencoders, recurrent neural networks (RNNs) and convolutional neural networks (CNNs) have been widely used in the area of signal and image denoising with great success [185], [186], [187], [188], [189]. Several attempts were also made in the field of ECG signal processing such as adult ECG denoising [134], [133], adult arrhythmia detection [135] and fetal QRS complex detection [190], [141]. Xiong et al. [134] used an improved denoising autoencoder (DAE), reformed by a wavelet transform method to remove baseline wander noise, electrode contact noises, and motion artifacts from adult ECG signals. Antczak in his work [133] trained an RNN model for denoising synthetic ECG data and subsequently used transfer learning to enhance the quality of real data.

To the best of our knowledge, so far there is no work attempting fetal ECG denoising using CNNs. Inspired by the work of Mao et al. [189] on image restoration, we propose the use of a deep convolutional encoder-decoder network with symmetric skip-layer connections for single channel fetal ECG denoising. Residual noise in the fetal ECG, after the maternal ECG has been removed, is often non-stationary, complex and has spectral overlap with the fetal ECG. Our method removes the residual noise by capturing the structure of the fetal ECG by the convolutional layers and recovering the signal details with the help of the transposed convolutional layers.

The rest of the chapter is organized as follows: Section 5.2 presents the proposed fetal ECG enhancement method and the data used. Experimental results are provided in Section 5.3. Finally, the results are discussed in Section 5.4 and conclusions are drawn in Section 5.5.

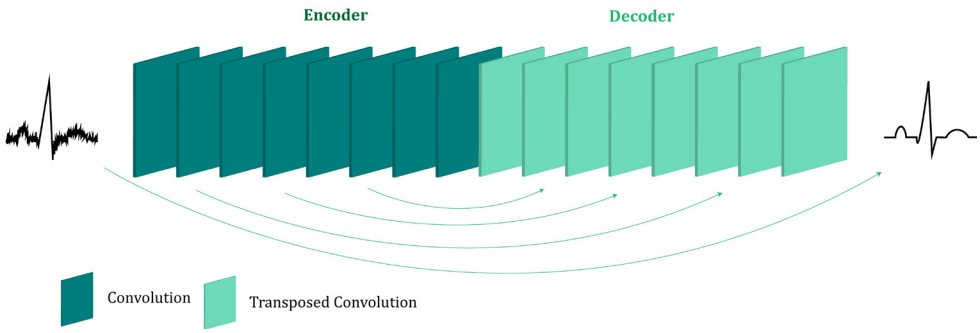



Fig. 5.1. The architecture of the proposed fetal ECG denoising network. The network consists of an encoder of 8 convolutional layers and a decoder of 8 transposed convolutional layers. Symmetric skip connections are added between every two corresponding convolutional and transposed convolutional layers.



## 5.2 Methods and data

### 5.2.1 Network Architecture

In this section, we present the proposed CNN model for removing the residual noise after extracting fetal ECG from antenatal abdominal recordings. We adopt the network architecture of Mao et al. [189], developed for image restoration, and modify it to make it suitable for ECG signal denoising. The proposed method, as shown in Fig. 5.1, aims to learn an end-to-end noise removal function from noise-corrupted fetal ECG to its clean version. The network contains two stages: an encoding and a decoding stage. The encoder acts as a feature extractor that removes the noise while preserving the primary ECG components. The decoder recovers the signal details and delivers a "clean" ECG as an output. Skip-connections are added between every two corresponding convolutional and transposed convolutional layers to help with recovering a clean ECG while also tackling the optimization difficulty caused by gradient vanishing in deep architectures.

### 5.2.2 Network Parameters

The network depth, together with the selection of kernel size and the use or not of subsampling or dilation in the convolutional layers define the receptive field of the network. Increasing the receptive field size can make use of the context information

in a larger signal region. It was indicated that the effective patch size of the denoising methods is highly correlated with the receptive field of the network [187], [191]. A simple strategy to achieve large receptive field is to increase the number of layers in the network. However, this becomes computationally very intensive. Alternatively, subsampling operations can be performed by the network, but this is not generally preferred in denoising tasks for the sake of preserving the signal details [188].

The fetal ECG usually exhibits high levels of noise and thus it requires a large effective patch size to capture more information for efficient denoising. Moreover, in order to exploit the self-similarity of the underlying fetal ECG signal the network should permit correlations to extend to several heartbeats. In the proposed framework we selected to use a relatively deep network of 8 convolutional and 8 mirrored transposed convolutional layers. Both the convolutions and the transposed convolutions are one-dimensional. Furthermore, subsampling by 2 is employed after every convolutional layer apart from the first one. We found that including subsampling operations in the convolutional layers helps rather than harms in terms of denoising performance since the receptive field of the network is drastically increased. Without subsampling a huge number of convolutional layers should be added to obtain the same receptive field. Moreover, in case of a practical implementation of the network, employment of the network should be fast, and the subsampling is also beneficial as it reduces the size of feature maps. It should be noted here that the additional skip-connections account for the lost signal details introduced by the subsampling in a great extent. Regarding the kernel size we use 15 in all layers, which for the case of signals sampled with 500 Hz corresponds to 30ms. With the above choices an effective receptive field of roughly 3.6s is achieved that corresponds to 5 to 10 heartbeats. For the sake of computational efficiency, we selected not to add more layers, even though this would increase even further the receptive field.

The input of the network is fixed to 1920 samples, slightly higher than the effective receptive field (1800 samples), to facilitate the subsampling operations. The number of filters applied to produce the feature maps is [64, 128, 256, 256, 512, 512, 1024, 2048] for the 8 convolutional layers respectively. Consequently, considering that the encoder subsamples the input signal by 2 in 7 layers, the dimension of the bottleneck feature vector is 2048x15. The transposed convolutional layers use mirrored numbered of filters. Leaky rectified linear units (Leaky ReLU) are utilized for non-linearity after each layer. The implementation of the network was done in Keras.

### 5.2.3 Skip Connections

In deep networks, transposed convolution does not work very well in recovering the details of the input data from only the data abstraction, possibly because too much detail is already lost in the convolution [189]. The suggested network is not only deep but, as mentioned above, the encoder heavily downsamples the input signal. To address this problem, skip connections are added from every two convolutional layers to the corresponding mirrored transposed convolutional layers. The feature maps passed by the skip connections carry much signal detail, which helps transposed convolution to recover an improved clean version of the fetal ECG. In addition, skip connections facilitate training of deep networks as they aid in back-propagating the gradient to the bottom layers [192], [193]. The way that the skip connections are used in the network is demonstrated in Fig. 5.2. As can been from the figure, the output of a convolutional layer is added to the output of the corresponding transposed convolutional layer and subsequently an activation function is applied to their sum.



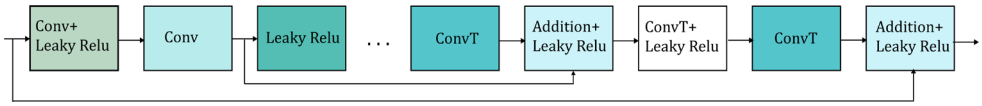



Fig. 5.2. The way that the skip connections are used in the proposed network. Only the first two skip connections are shown for simplicity. Conv stands for convolution and ConvT for transposed convolution.

### 5.2.4 Data

The fetal ECG signals, even after the maternal ECG has been removed, are still affected by noise. It is hence impossible to have real fetal ECG pairs (noisy and clean) to train the network. Thus, we used a rich dataset of simulated fetal ECG for training. The algorithm was then extensively validated in a separate simulated test dataset, while results are additionally provided for real signals and compared with simultaneously performed scalp fetal ECG recordings.

#### *5.2.4.1 Simulated Data*

For the creation of an extensive simulated dataset we used the fecgsyn toolbox [175]. By employing fecgsyn, fetal-maternal mixtures can be created while there is the possibility to model a number of non-stationary events that affect the morphology and dynamics of the abdominal ECG by rotating, translating and modulating

the available sources. The synthetic abdominal signals consist of 34 channels (32 abdominal and 2 maternal ECG reference channels). With the use of the toolbox, we created a dataset where different physiological events are considered, such as ectopic beats, uterine contraction, noise, fetal movement etc., similar to the Fetal ECG Synthetic Database (FECGSYDB) [175]. Unfortunately, the data simulated with the fecgsyn toolbox is based on 9 different vectorcardiograms (VCGs) only. This means that training merely on this dataset has the risk of overfitting to these specific VCGs since there in relatively small variation in the timing of the ECG intervals. One possibility to introduce more variety in the dataset is to model more VCGs but we have chosen a different approach. Our approach includes modifying the fecgsyn toolbox to obtain a big variety of ECG morphologies based on the already available VCGs. The modified fecgsyn toolbox is able to receive a VCG and then apply random modifications of the lengths of the VCG intervals and segments, as well as of the amplitudes of the waves, such that a new VCG is created. This VCG was then used as a base to create the abdominal fetal ECG. The ranges of modifications were selected to cover a wide range of variations of the ECG morphological features, while still ensuring physiologically plausible ECGs.

To further increase the diversity of ECG patterns in our dataset we generated one additional set of data. For this set, adult ECG signals were used to simulate the fetal ECG. Morphologically, adults and fetuses have rather similar ECG patterns; they both comprise a PQRST complex, yet the amplitudes and lengths of adult ECG segments and intervals are large compared to those of the fetal ECG. The adult ECGs were collected from 3 different databases of Physionet [176], the PTB Diagnostic ECG Database [194], the St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database [176] and the QT database [195]. The signals were preprocessed to remove baseline wander and noise and to resemble fetal ECG. A high-pass filter with cut-off frequency at 1 Hz was first applied to the signals, followed by a Savitzky-Golay filtering of order 8 and length 31. Based on the fact that the speed of the fetal heart rate is two to three times faster than the adult heart rate, the adult ECG signals were resampled at half frequency to look similar to fetal ECG signals. Amplitude scaling of the ECG was not necessary, as the data was normalized before entering the network. As a final step, noise was added to the signals. For this purpose, a number of 6-channel abdominal recordings, for which the study protocol is described in [174], was employed. In some recordings, after the maternal ECG suppression, the fetal ECG was impossible to be detected in any electrode or a subset of them. This

was caused either by the shielding of the fetus by the vernix caseosa or because some electrodes were too far from the fetal heart. In these cases, we assumed that the electrodes recorded only noise (apart from the maternal ECG that is subtracted) and this noise was added to the preprocessed adult ECG to simulate noisy fetal ECG. The advantage of using real noise recordings is that the data used for training the network is highly similar to the data we aim to use in testing and ultimately employment of the network. By learning how to remove real noise, the network is likely to perform better in real data.

#### *5.2.4.2 Real Data*

To investigate the impact of training with synthetic data on denoising in a real dataset, we used the Abdominal and Direct Fetal Electrocardiogram Database [177]. This database contains multichannel fetal ECG recordings obtained from 5 different women in labor, between 38 and 41 weeks of gestation. Each recording contains 4 signals acquired from maternal abdomen and one scalp ECG signal. The recordings have duration of 5 minutes and are sampled at 1000 Hz.



#### *5.2.4.3 Data Preprocessing*

The signals of all datasets were preprocessed before entering the network either for training or for testing. Regarding the abdominal recordings, the open-source algorithm of Varanini et al. [104] was initially applied to them. According to this algorithm, first the baseline wander and the powerline interference were removed. After that, the maternal ECG was estimated through Independent Component Analysis (ICA) and singular value decomposition and subsequently subtracted from the signals. Finally, a second ICA was employed to enhance the fetal ECG signal.

The fetal ECG signals of all datasets were resampled to 500Hz to have a common reference. Then, they were divided in segments of 1920 samples and finally normalized to have zero mean and unity standard deviation. The normalization was performed in each segment separately.

### 5.2.5 Network Training

In order to train the weights of the convolutional and the transposed convolutional kernels a loss function was minimized. A standard loss function used in denoising optimization problems is the mean squared error (MSE):

$$\mathcal{L} = \frac{1}{N*M}\sum_{n=1}^{N}\sum_{m=1}^{M}\left(X_{clean_{n,m}} - X_{denoised_{n,m}}\right)^2 \tag{5.1}$$

where N is the number of the training data in a batch, M the length of the signals, $X_{denoised}$ is the denoised fetal ECG and $X_{clean}$ is the clean version of it, used as ground truth. We found, that even when we normalized our data with zero mean and standard deviation one, still there were some scale differences among the data, since we combined different datasets. To avoid favoring the data with the higher scale instead of MSE we adopted as loss function the normalized mean squared error (NMSE):

$$\mathcal{L} = \frac{1}{N*M}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\left(X_{clean_{n,m}} - X_{denoised_{n,m}}\right)^2}{\bar{X}^2_{clean_n}} \tag{5.2}$$

where we normalized the MSE by the mean squared amplitude of the target signal $\bar{X}^2_{clean}$. We found that training with NMSE led to an improved network performance. The Adam algorithm [152] was selected as optimization algorithm.

The simulated dataset was separated in two parts for test and training. In the training set we added the signals generated with the modified fecgsyn toolbox based on the VCGs 1-7, while the data based on VCG 8-9 was assigned to the test set. The data simulated from adult ECG were also carefully separated into training and test set. In addition, since the VCGs from the fecgsyn toolbox came from fitting a model on 9 subjects from the PTB Diagnostic ECG Database, only these subjects were excluded to avoid any mixing of the training and test sets. Finally, the total number of training segments used in our experiments was 840000 and of test segments 200000. The network was trained on data with noise from -15dB to 15dB for 75 epochs until convergence was reached.

### 5.2.6 Performance Measure

The performance of the network was assessed based on the SNR improvement achieved after the denoising comparing to the SNR of corrupted fetal ECG. The metric is computed as follows:

$$\begin{aligned} SNR_{imp}[dB] &= SNR_{post-denoising} - SNR_{pre-denoising} \\ &= 10log_{10}\frac{\sum_{m=1}^{M}\left|X_{noisy_m} - X_{clean_m}\right|^2}{\sum_{m=1}^{M}\left|X_{denoised_m} - X_{clean_m}\right|^2}, \end{aligned} \tag{5.3}$$

where $X_{noisy}$ denotes the corrupted fetal ECG signal.

The proposed method was compared to two other ECG enhancement algorithms. The first one is beat-to-beat averaging, where we averaged 30 successive ECG complexes, similar to STAN method [25]. The R-peaks were detected before the averaging with the Pan Tompkins algorithm [196] in the clean signals. In this way, the averaging performance was not affected by the performance of the R-peak detector. The second method is a wavelet denoising method that decomposes the signal, thresholds the detail coefficients, and reconstructs the signal to obtain its enhanced version. The symlet wavelet with 6 vanishing moments was selected as a mother wavelet because of its similarity to an actual ECG. A fixed threshold was used which was estimated by the minimax principle [178].



## 5.3 Results

### 5.3.1 Evaluation on Simulated Data

The denoising network was evaluated in the simulated test dataset. Fig. 5.3 illustrates the performance of the developed fetal ECG denoising network in comparison to the 30-complex averaging and wavelet denoising methods. The input SNR refers to the SNR of the fetal ECG signals prior to denoising. As we can see in Fig. 5.3,

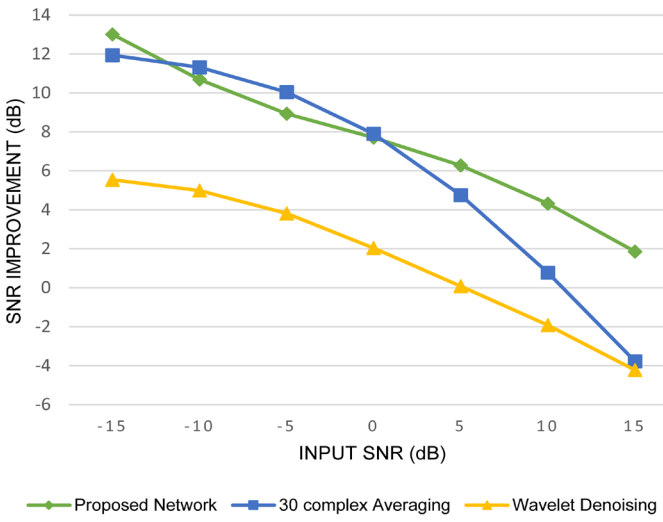



Fig. 5.3. Performance of the proposed fetal ECG denoising network in comparison with other denoising methods in terms of SNR improvement of the denoised signal when compared with the noisy one.

the denoising network provides a considerable amount of SNR improvement throughout the whole range of input SNR. In higher SNR ranges (more than 0dB) our network outperforms wavelet denoising and averaging techniques. This was expected since the averaging method does not preserve individual variations in the ECG complexes while our method is capable of doing so. Moreover, wavelet denoising distorts the signal amplitude, whereas the denoising network does a better job in preserving it, especially in cases of low noise. For lower SNR values averaging performs slightly better comparing to our method. The averaging method uses a larger signal segment for denoising (30 complexes) comparing to the network (5-10 complexes). Apparently, an even larger effective denoising patch is needed for cases of heavy noise. On the other hand, unlike the averaging, the proposed network does not make use of any prior information about the input signal. It should be mentioned here that the averaging method assumes that the R-peaks are known. However, it is not guaranteed that the R-peaks can be accurately estimated in very noisy signals. The strength of our method is that no prior processing of the signals, such as detecting the R-peaks, is required and consequently errors in denoising originating from wrong peak detection are avoided.

Fig. 5.4 depicts some typical results obtained in the test dataset for a wide range of SNR values. For relatively high SNR values (cases (a)-(c)) the network suppresses the noise to an extent that the ECG waves become distinguishable. Moreover, there is no or little distortion in the signal amplitude. For lower SNR values, around -5dB, usually the QRS complexes are successfully recovered as it can be observed in Fig. 4(d) and 4(e). However, there are cases that some P or T waves are created by the network where they do not actually exist. As example, in Fig. 5.4(d) the P-waves in the denoised signal differ from the ground truth ECG both in location and polarity. In other instances, the waves are difficult to recover, or their amplitude is distorted (Fig. 5.4(e)). Finally, in very low SNR (less than -10dB) even the QRS complexes appear challenging to be recovered (Fig. 5.4(f)). Thus, in low SNR ranges of the input signals even though the SNR of the output signal is significantly increased, some morphological features might not always be reliably extracted.

Fig. 5.5 illustrates the potential of our denoising method in a case with arrhythmia. The quality of the output signal has been enhanced to a great extent and nearly all the individual variations among the complexes are preserved. Unlike our algorithm, the averaging of successive complexes fails in maintaining abrupt changes in signal morphology.

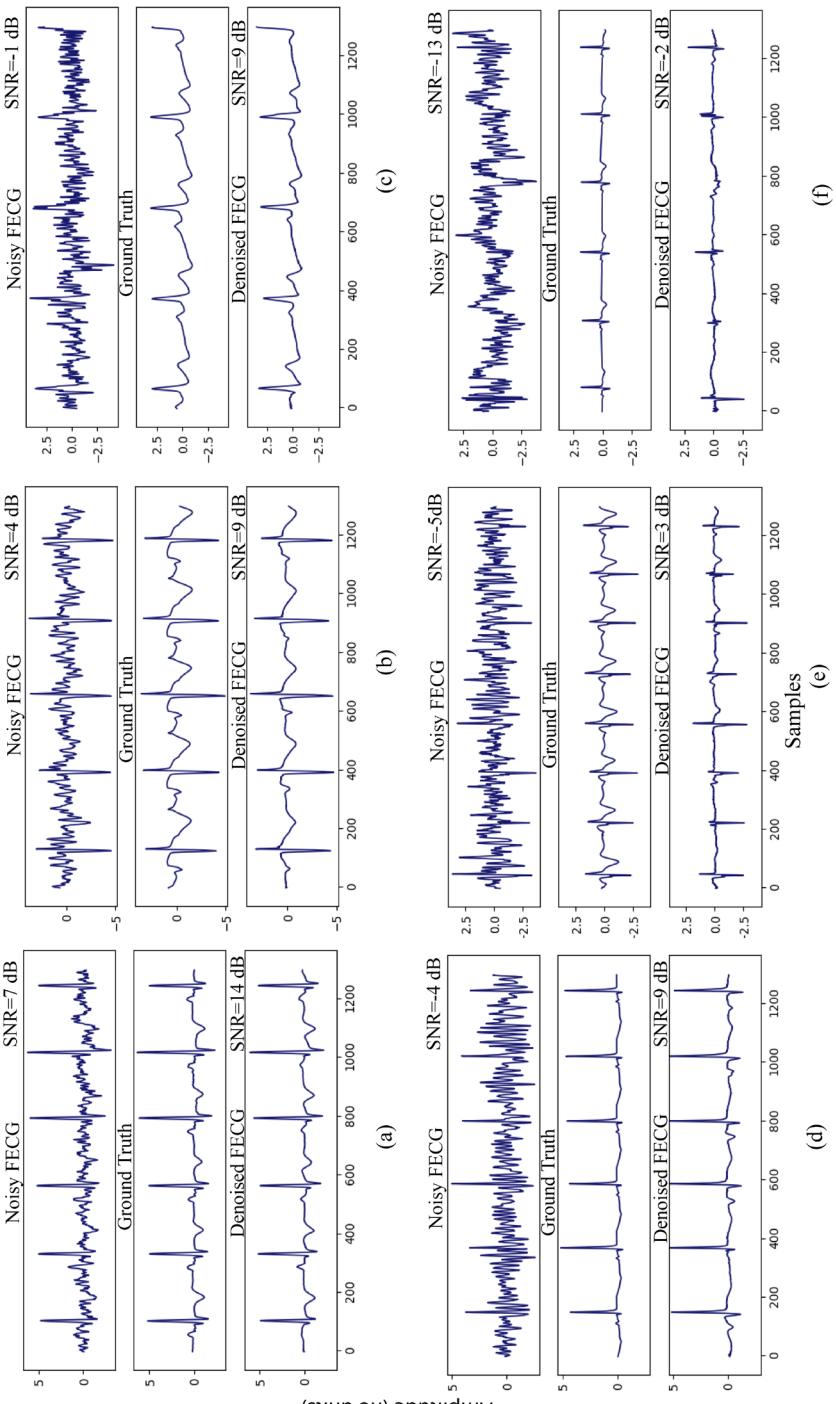



Fig. 5.4. Denoising results for simulated fetal ECG signals in the test dataset with different amount of noise added to them. For relatively high input SNR the network achieves a high-quality result while for lower SNR values some morphological features might be unreliable.

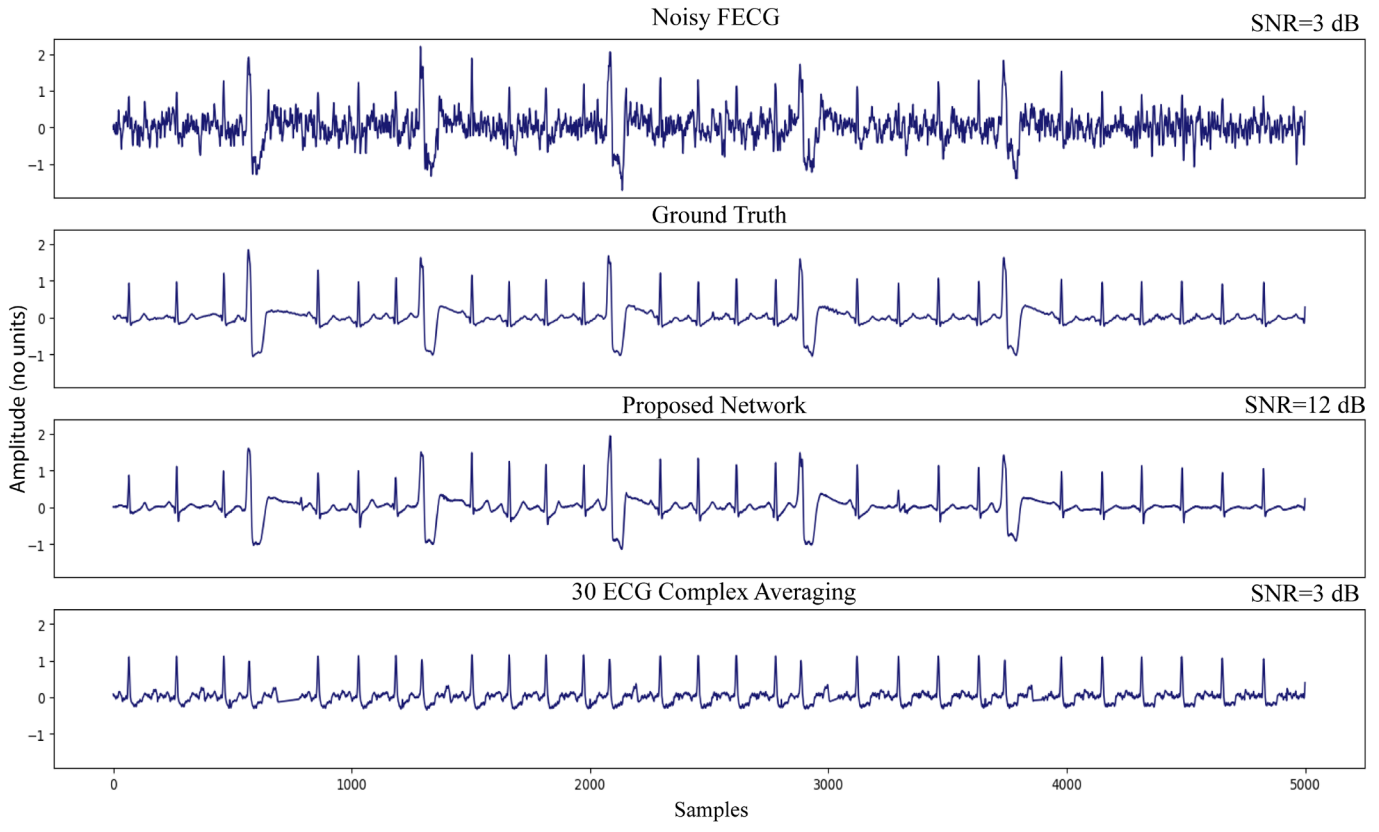



Fig. 5.5. An example of denoising a simulated fetal ECG exhibiting arrhythmia with the proposed deep CNN network in comparison with the 30 ECG complex averaging method. Unlike averaging, the suggested network can preserve diagnostic morphological information of the signal.

### 5.3.2 Performance on Real Fetal ECG Signals

The fetal ECG denoising network was merely trained on simulated data. Therefore, it is interesting to examine how well it performs in terms of enhancing the quality of real fetal ECG. Since ground truth real signals are not available, the performance of the network was tested by comparing the denoised signals of the Abdominal and Direct Fetal Electrocardiogram Database with the simultaneously recorded scalp ECG. It should be noted that the scalp ECG is a different ECG lead than the abdominal leads and that they should not be identical, even in case of perfect denoising. Nevertheless, the individual ECG segments should coincide between the abdominal leads and the scalp lead. Two intervals were computed and compared between the denoised signals and the scalp ECG in our analysis: the QT interval and the PR interval. The QT interval is defined as the interval from the onset of QRS complex to the end of the T wave (offset of ventricular repolarization). The PR interval corresponds to the period that extends from the beginning of the P wave (onset of atrial depolarization) to the beginning of the QRS complex (onset of ventricular depolarization).



The scalp measurements contain considerable amount of noise that made it challenging to determine the intervals. To enable the interval detection, quality enhancement of the scalp ECG was necessary. For this purpose, we performed averaging of 7 successive ECG complexes. To have a fair comparison, the same procedure was applied to the denoised signals. Even after 7 complex averaging there was substantial noise left in some segments of the scalp ECGs and therefore, we had to exclude them from the comparison. However, by averaging more complexes we risked assessing the performance of the averaging method rather than our proposed denoising. Choosing to use 7 complexes was a compromise between sufficient quality enhancement of the scalp ECGs and preservation of variations among consecutive complexes.

The end of the T wave is difficult to define [197] and to calculate it we adopted the tangent technique. According to this technique, a tangent line is determined down the steepest slope of the terminal limb of the T wave. Afterwards, the end of the T wave is defined by the intersection of this line with the baseline. In a similar way the beginning of the P wave was determined. The intervals were calculated first in an automated way and then checked manually by an expert to correct for errors. From each recording, only the channel with the least amount of noise was analyzed. Since the recordings are 5 minutes long this corresponds approximately to 620 intervals

per recording (~3100 in total). The signal parts, where the amount of noise was too high, or the intervals were not recognizable, were excluded from the comparison (both for the scalp and denoised signals). After this exclusion, 2000 intervals were compared.

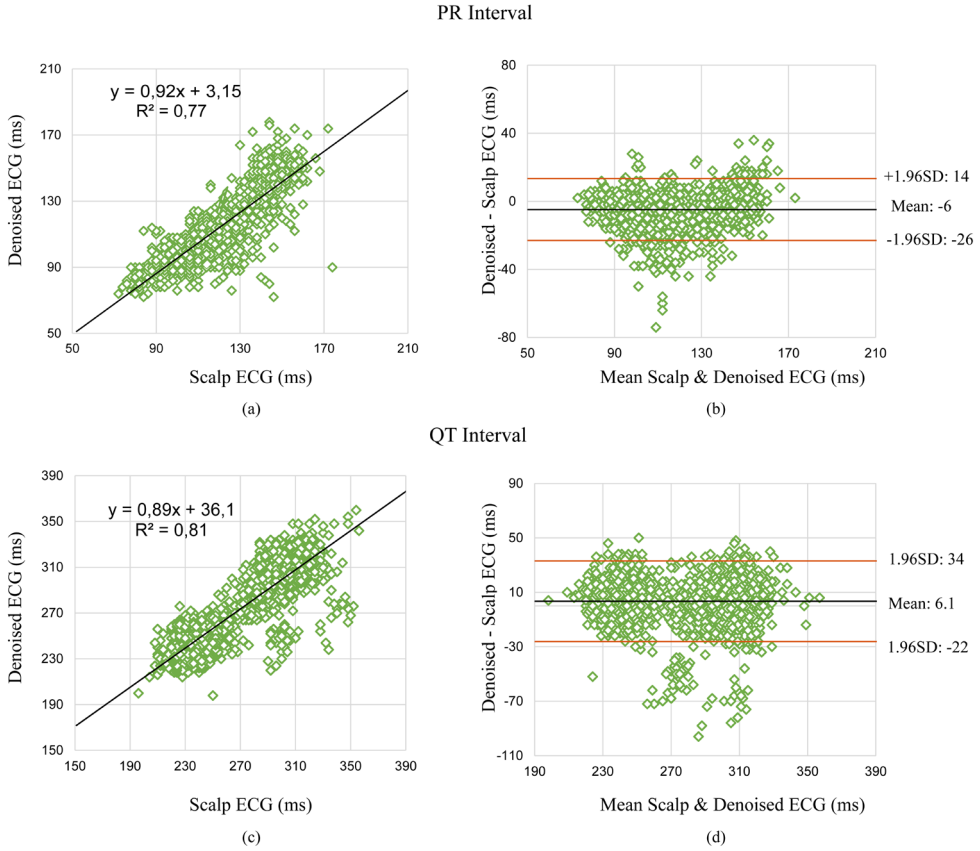



Fig. 5.6. Interval comparison between the denoised noninvasive fetal ECG signals and the simultaneously recorded scalp ECG: (a), (c) correlation plots and (b), (d) Bland Altman plots for the PR and QT interval respectively. In (a) and (c) line fit is given by y = intercept + gradient x and $R^2$ is the corresponding squared Pearson correlation coefficient.

The root mean squared error (RMSE) and the squared Pearson correlation coefficient ($R^2$) were computed to measure how well the intervals match between the scalp and the denoised fetal ECG signals. Moreover, we evaluated the RMSE95 and $R^2$95 where the extreme 5% values were excluded. This was done to ensure that the outliers did not bias the estimation of the metrics. Table 5.1 displays the calculated values for

the two intervals. The RMSE for the PR interval was estimated to be 9.9ms, while after the exclusion of the extreme values 7.4ms. For the QT interval RMSE is equal to 14ms and RMSE95 11ms. We should note here that in the scalp measurements the average value for the PR interval was found 120ms and for the QT interval 270ms. For the PR interval the squared correlation coefficient is 0.77 while for the QT 0.81. The removal of the extreme values raised the value to 0.86 for the PR and 0.87 for the QT interval respectively. We provide additionally the correlation and the Bland Altman plots for the estimated intervals in Fig. 5.6. The Bland Altman plot shows that the PR interval was measured on average 6ms shorter in the denoised signals when compared to the scalp electrode. Regarding the QT interval, it was computed 6.1ms longer in the noninvasive fetal ECG signals. Moreover, we observe that the limits of agreement are not very wide, especially for the QT interval, but there are measurements falling outside them. This means that the denoising does not provide consistently good results but also some outliers.



Table 5.1. Comparison of the PR and QT interval between the denoised fetal ECG signals and the scalp ECG. All the values are expressed in ms.

| Interval | RMSE | $R^2$ | RMSE95 | $R^2$95 |
|---|---|---|---|---|
| PR | 9.9 | 0.77 | 7.4 | 0.86 |
| QT | 14 | 0.81 | 11 | 0.87 |

Finally, a visual result of the denoising is provided in Fig. 5.7 for channel 2 of the recording “r04” of the Abdominal and Direct Fetal Electrocardiogram Database. For better visualization the scalp ECG was high-passed filtered to remove the baseline wander. The denoised ECG signal is free from noise and the individual waves correspond relatively well to those in the scalp ECG.

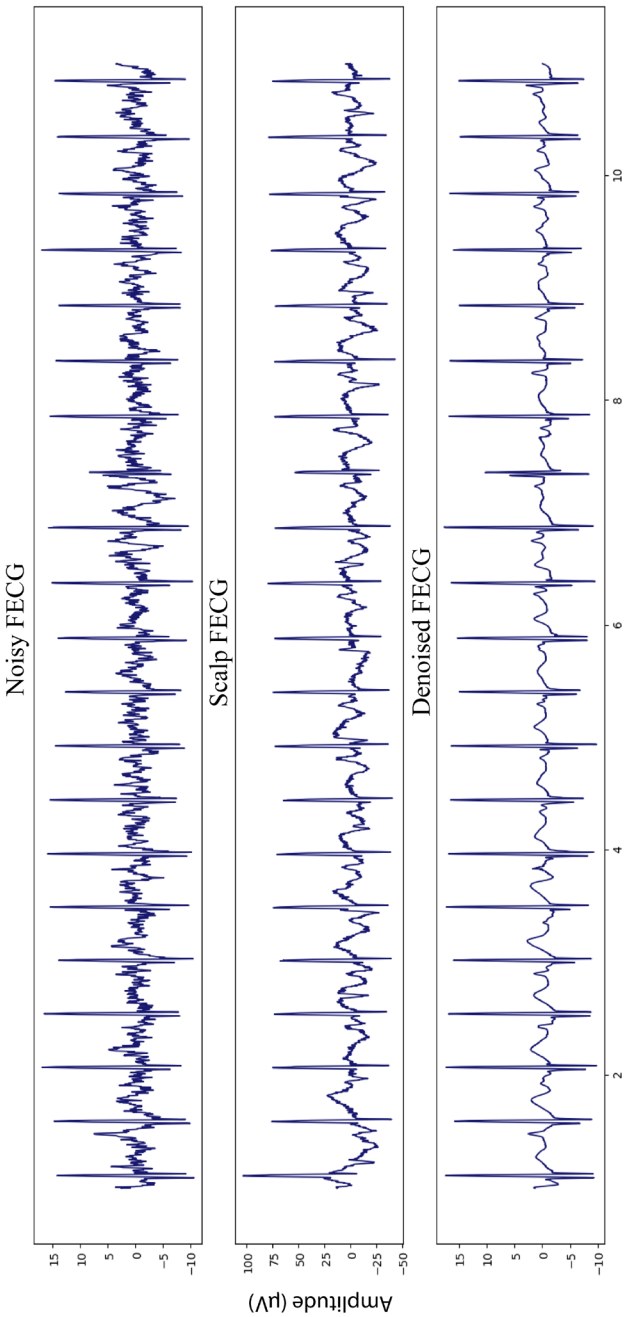



Fig. 5.7 the denoising result of the proposed network for channel 2 of recording "r04" of the Abdominal and Direct Fetal Electrocardiogram Database. The enhanced signal is clean and the morphological signal characteristics are similar to those of the scalp ECG.

## 5.4 Discussion

This chapter presented a deep CNN network for denoising fetal ECG signals. Fetal ECG signals, obtained from noninvasive recordings, contain a substantial amount of noise even after the maternal ECG is suppressed. This renders it difficult for the clinicians to examine and interpret the morphology of the ECG signals. Usually, the QRS complex can be detected without further processing, due to the high amplitude of the R-peak. However, the smaller waves, like the P and T waves, can often not be readily distinguished. The proposed CNN network was developed to postprocess the extracted fetal ECG signals in order to further enhance their quality. The network was trained merely on simulated data and its performance was examined both on synthetic and real signals.



The network was found to be efficient in reducing residual noise in the synthetic ECG dataset. When the SNR of the signals was relatively high the T and the P waves were mostly reconstructed with no or little distortion. In most of these cases the waves in the noisy input signals were impossible or too difficult to detect. However, in cases with low SNR, even if a significant SNR improvement was achieved, some morphological features were distorted, absent or "fake". This means that the network cannot reliably reconstruct the ECG morphology when the signals are severely corrupted by noise. Possibly longer signal segments are needed rather than 4s to capture more information about the underlying signal structure. This finding suggests that if the developed network is intended to be used in practice, then prior quality assessment of the input signal is necessary as a reliability measure of the result. If the quality of the signals is assessed as low, then the network should not be used for denoising. Alternatively, possibly longer signal segments can be considered and more layers should be added to the network to achieve a larger effective denoising patch at the expense of higher computational intensity and possible need of even more data. Moreover, in our analysis we have experimented only with single channel ECG. However, modifying our network to handle multichannel fetal ECG is straightforward by employing 2D convolutions. Allowing the convolutions to extend both in space and time might lead to more accurate results.

In a real dataset we measured how well the PR and QT intervals of the denoised signals correspond to those measured in the scalp ECG. The intervals were extracted in the signals created as a running average of 7 heartbeats. There is a limitation in this comparison, since there is some further denoising performed because of the

averaging. However, the need to remove the noise in the scalp, for obtaining reliable intervals, outweighs the mentioned limitation. We found a RMSE of 9.9ms for the PR interval and 14ms for the QT interval. These values show high similarity between the two measurements especially if we consider errors and high ambiguity in the computation of the intervals. The RMSE value for the QT interval is similar to the one found by Behar et al. [198] (13.6ms) where annotations of several cardiologists were combined and compared to the scalp ECG. However, compared to [198], we used less subjects in our comparison. Furthermore, the running average computation in [46] was different from ours since they selected only the ECG complexes with high similarity for the averaging in 1-minute segments. They found that the QT interval can be reliably extracted only if computed as a running average of several heartbeats. In our case the averaging is less necessary since most of the noise is removed by the denoising network.

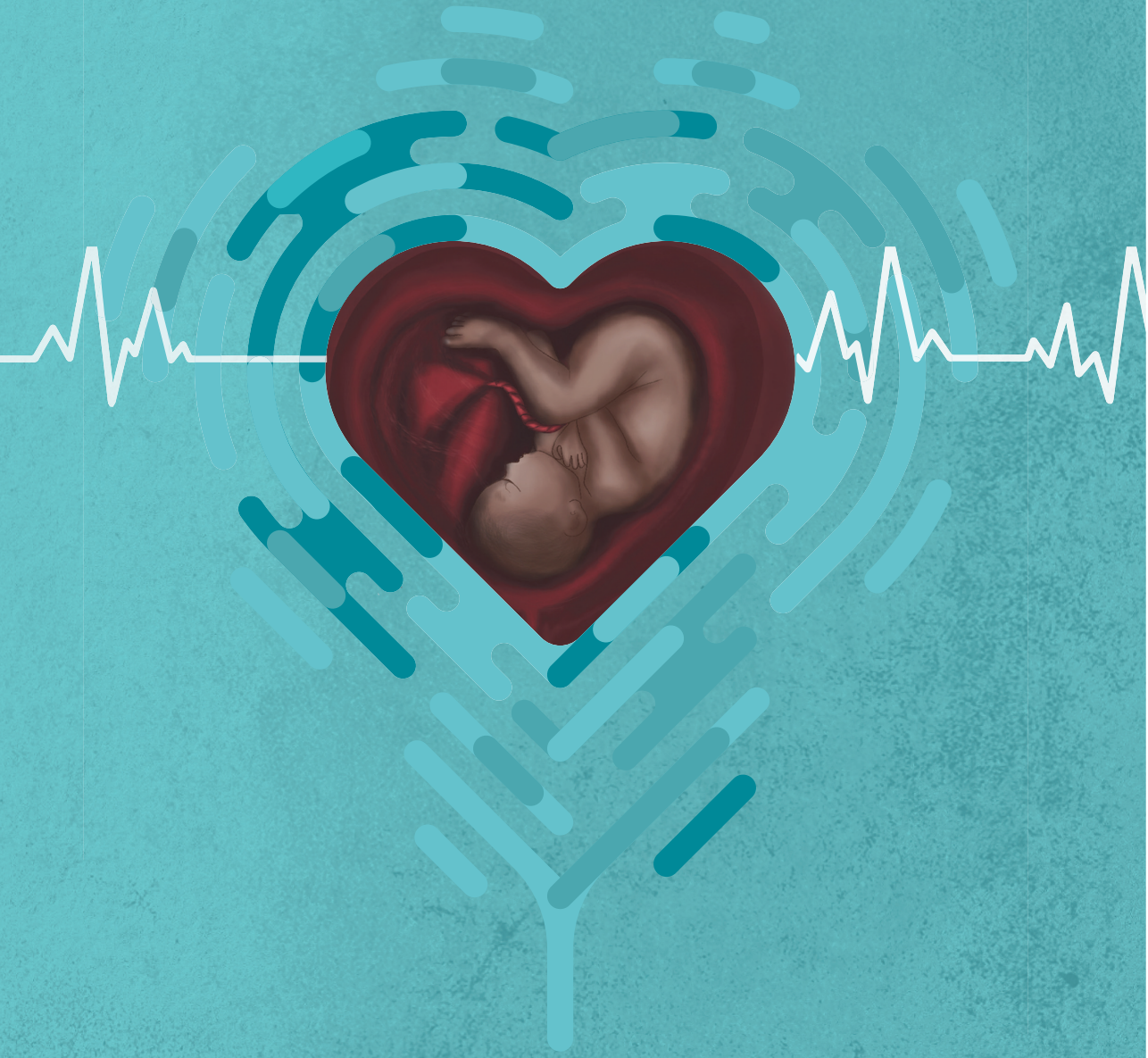
It bears mentioning, however, that validation in a more extensive real dataset is needed to confirm our findings. Moreover, the signals in our study were recorded between 38 and 41 weeks of gestation. This is because it is impossible to have scalp measurements before birth. However, it would be interesting to assess the accuracy of the method in earlier gestational ages also.

Despite the limitations in the validation of our method, this is the first study to show the potential of deep CNNs in efficiently removing noise from noninvasive fetal ECG signals. The principal advantage of the method over the widely used running average method is that no prior R-peak detection is necessary and that individual variations in pulse shape and beat-to-beat interval are retained. This is especially beneficial in arrhythmia cases. Nowadays, the averaging performed in the fetal ECG precludes its use in real time arrhythmia analysis and hence fetal arrhythmia can only be assessed through echocardiography. This work brings us a step closer to broadening our understanding of mechanisms of fetal arrhythmia by examining the fetal ECG. Moreover, the quality of the denoised signals is high enough to allow for measuring the exact timing of different morphological features of the ECG signal by the clinicians. Besides, it can facilitate and advance the research towards automated detection of fetal ECG intervals and segments. Extracting morphological features from the ECG signal allows for the estimation of the well-being of the fetus. Metabolic acidosis was found to be associated with QT length variations [89]. Several studies have also demonstrated a physiologically negative correlation between the PR interval and RR interval which becomes positive with evolving acidosis [199], [200]. However, the

role of the ECG intervals in fetal monitoring is yet to be established, but for this to be achieved the technological limitations related to the noninvasive ECG should be overcome.

## 5.5 Conclusion

In this chapter, we proposed a deep encoder-decoder framework for noninvasive fetal ECG signal denoising. Convolutions and transposed convolutions are combined to remove the noise by extracting primary signal content and recovering details. Experimental results in synthetic signals showed that the network is able to achieve a substantial quality improvement of the noisy signals. However, when the signals are heavily corrupted by noise, some morphological features are unreliable, urging the need for reliability measure of the network's output. Experiments on real signals demonstrated high correlation of the PR and QT interval in the denoised signals when compared with the scalp ECG. The principal advantage of the method is that individual variations among different pulses can be preserved and that, opposed to most other fetal ECG denoising methods, the method does not require knowledge on R-peak locations.



# CHAPTER 6

# MULTICHANNEL FETAL ECG DENOISING WITH DEEP CONVOLUTIONAL NEURAL NETWORKS

Based on [JP-3]: Fotiadou E., Vullings R., Multi-Channel Fetal ECG Denoising with Deep Convolutional Neural Networks. Front. Pediatr., 2020, 8, 508. doi:10.3389/fped.2020.00508

**Abstract**. Noninvasive fetal electrocardiography represents a valuable alternative continuous fetal monitoring method that has recently received considerable attention in assessing fetal health. However, the noninvasive fetal electrocardiogram (ECG) is typically severely contaminated by a considerable amount of various noise sources, rendering fetal ECG denoising a very challenging task. This work employs a deep learning approach for removing the residual noise from multichannel fetal ECG after the maternal ECG has been suppressed. We propose a deep convolutional encoder-decoder network with symmetric skip-layer connections, learning end-to-end mappings from noise-corrupted fetal ECG signals to clean ones. Experiments on simulated data show an average signal-to-noise ratio (SNR) improvement of 9.5 dB for fetal ECG signals with input SNR ranging between -20 and 20 dB. The method is additionally evaluated on a large set of real signals, demonstrating that it can provide significant quality improvement of the noisy fetal ECG signals. We further show that employment of multichannel signal information by the network provides superior and more reliable performance as opposed to its single channel network counterpart. The presented method is able to preserve beat-to-beat morphological variations and does not require any prior information on the power spectra of the noise or the pulse location.

## 6.1 Introduction

The fetal electrocardiogram (ECG) can be used to monitor the condition of the fetal heart from early pregnancy until delivery [157]. Nowadays, fetal monitoring is mainly performed by cardiotocography or by ECG recordings where an electrode is directly placed on the fetal scalp. Cardiotocography records the fetal heart rate together with the uterine contractions. The advantages of the method are that it is performed noninvasively and is safe for the patient. On the other hand, it is prone to signal loss, while recorded changes of the heart rate are not always precise [16]. Scalp ECG recordings are a more reliable means of monitoring the fetal health. However, they are invasive, may pose a health risk to the fetus, and can only be performed during labor, when the membranes have ruptured.

Noninvasive fetal electrocardiography, performed by placing electrodes on the maternal abdomen, is a promising alternative to standard fetal monitoring. In comparison with cardiotocography, it provides more accurate information because it does not need to average over multiple beats for the heart rate extraction. Moreover, it provides the possibility to assess the ECG morphology, related to the electrical activity of the fetal heart. The advantage of the method over the scalp ECG measurements is that it can be performed already during pregnancy, it is safe for the fetus and comfortable for the mother. However, the difficulty to extract a clean fetal ECG from the abdominal mixture is the main reason that the application of the method in clinical practice is still limited. The interferences and noises in the abdominal recordings among others include the maternal ECG, powerline interference, baseline wander, muscle noise from the fetus and mother and movement artifacts. Considering that the signals of some of these interferences and noises overlap both in time and frequency with the fetal ECG, the extracted fetal ECG signals usually have very low signal-to-noise ratio (SNR). Therefore, the noninvasive recordings are in practice merely used for heart rate analysis.



There are typically three main steps in the fetal ECG extraction process: preprocessing, separation and postprocessing [201]. Preprocessing includes removal of unwanted noise and interferences such as powerline interference and baseline wander. In the separation step, the maternal ECG is estimated and then subtracted from the signals to obtain the fetal ECG. Finally postprocessing is employed to enhance the quality of the extracted fetal ECG signals. The work on noninvasive fetal ECG analysis has mainly targeted the first two steps, together with the improvement of the acquisition devices [121], while only few works focused on the postprocessing

of the obtained signals. Beat-to-beat averaging is a traditional method which is often used to improve the SNR of the extracted signals, at the expense of losing individual variations in pulse shape [113]. Different wavelet denoising techniques were additionally proposed in the literature for the postprocessing of the extracted fetal ECG signals [114], [115]. In a previous work [184], the authors employed an augmented time-sequenced adaptive filter to enhance the quality of the extracted fetal ECG. Despite the significant quality improvement that the method achieves, the location of the fetal pulses is required to synchronize the filter and the method cannot handle abrupt changes in fetal ECG morphology, e.g. in cases of arrhythmia.

Recently, deep neural network models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) and stacked denoising autoencoders have been successfully applied for a variety of purposes including signal and image denoising [189], [188], [187], [186], [185]. Moreover, few works reported adult ECG signal denoising [133], [134], fetal QRS detection [140], [141] and fetal ECG signal reconstruction [139]. Zhong et al. [202] presented a deep convolutional encoder-decoder framework for preprocessing abdominal recordings to remove noise. However, they did not extract the fetal ECG from the preprocessed signals to ensure that it is not suppressed by the network. The authors were the first to propose a deep convolutional encoder-decoder network for postprocessing noninvasive single channel fetal ECG [203], [137], achieving a substantial quality improvement of the noisy signals. The method tackled some of the shortcomings of the state-of-the-art noninvasive fetal ECG postprocessing methods, since it can preserve beat-to-beat morphological variations and does not require prior knowledge about the location of the fetal pulses. However, in cases of heavily corrupted signals, the method was unable to reliably reconstruct some relevant morphological features of the ECG, sporadically even causing presence of “fake” waves, i.e. waves in the reconstructed ECG that should not have been there or should have had opposite sign. For a practical application this might be dangerous, leading to wrong diagnosis.

In this work, we are dealing with the aforementioned problem by extending our model to handle multiple fetal ECG channels. Multiple electrodes measure the electrical activity of the heart from different angles. We propose to use a deep convolutional encoder-decoder network with symmetric skip connections that learns how to optimally combine the input channels to deliver a reliable clean, multichannel ECG as output. The method eliminates the residual noise in the fetal ECG by capturing the signal structure in the convolutional layers and recovering the details by the transposed convolutional layers.

## 6.2 Materials and Methods

### 6.2.1 Data

#### *6.2.1.1 Simulated Data*

For the training, but also for the evaluation of the proposed network, we created an extensive simulated fetal ECG dataset that consists of two parts. The first part was built by employing the fecgsyn toolbox developed by Behar et al. [204], [176]. The toolbox enables the creation of abdominal mixtures with adjustable noise sources, heart rate, heart rate variability, fetal movement, ectopic beats and contractions. A Gaussian model is used to simulate the ECG beats, as originally developed by McSharry [205] and further improved by Sameni [206]. Any number of electrodes can be positioned on the maternal abdomen for the simulations. Unfortunately, the simulated fetal ECGs are based merely on 9 available vectorcardiograms (VCGs). Since there is limited variation in the shape and lengths of the individual PQRST waves in these VCGs, there is an increased risk of overfitting the network. This means that the network might learn to reproduce these limited morphologies and enforce resemblance of the denoised signals with the training data. In fact, what happened in our initial experiments is that the P and T waves of the denoised signals were shifted with respect to their ground truth data to match the locations of the training data. For this reason, we built a modified version of the toolbox that creates a variety of new ECG morphologies based on the already available VCGs. The modified toolbox receives a VCG as input, alters the length of the VCG intervals along with the amplitudes of the PQRST waves and subsequently uses it as a base to form the abdominal fetal ECG. Initially, for all 9 VCGs, the points of interest, which are the beginning and end of the P wave, T wave, and QRS complex were annotated and saved to be later available to the simulator. In every iteration of the modified simulator, one of the 9 VCGs is randomly selected and subsequently the start and the end of the waves are randomly shifted in position. Since the shift of the start and shift of the end point of each wave are not identical, also the length of the waves is automatically varied this way. The amplitude of each wave is changed as well by random scaling. The modified VCG is the starting point that the abdominal fetal ECG can be created. With the help of the modified toolbox we created a large dataset of 4-channel abdominal mixtures, where different physiological events were considered, such as heart rate decelerations and accelerations, fetal movement, ectopic beats, uterine contraction etc., similar to the Fetal ECG Synthetic Database [175]. The VCG alterations were chosen so as to include an ample range of variations of the ECG

morphological features, while still ensuring their physiological plausibility. When obtaining the simulated data, we varied the placement of the four electrodes to make the method invariant to variations in the electrode position.

To further enrich the ECG morphologies in our dataset and reduce the risk of overfitting to the training data we generated an additional set of simulated signals based on adult ECG from the PTB Diagnostic ECG Database of Physionet [194]. The database comprises of both normal and pathological signals with 15 leads, sampled at 1000 Hz. 549 records from 290 male and female subjects are available. Adult and fetal ECG have similar morphology, but the adult ECG intervals and amplitudes are larger compared to the fetus. The adult ECG was preprocessed to remove noise and resemble the fetal ECG. First, a high-pass filter with cut-off frequency of 1 Hz was applied followed by Savitzky-Golay filtering of order 8 and length 31. Afterwards, considering that the fetal heart beats two to three times faster than the adult heart, the signals were resampled to half frequency. Adjustment of the signals amplitude was not necessary because they were, in a later data preparation step, anyway normalized before entering the network. As a next step, 4-channel signals were created by making random combinations of four leads, where a maximum of two was chosen out of the six first limb leads. Finally, "real" noise was added to the signals. For the "real" noise we employed a set of 6-channel abdominal recordings of an ongoing study of which the protocol is described in [174]. In a subset of these recordings we found it impossible to detect the fetal ECG, either because of the shielding of the fetus by the vernix caseosa or because the fetal heart was far from some electrodes. We considered that these measurements, after the maternal ECG suppression and powerline interference removal, consist of pure noise and added them to the preprocessed adult ECG to generate our simulated fetal ECG signals.

#### *6.2.1.2 Real Data*

In order to evaluate how well our algorithm performs in real signals we employed two databases. The first one is a private set of noninvasive fetal ECG measurements, obtained in collaboration with the Máxima Medical Center, Veldhoven, the Netherlands [174], [85]. The dataset contains 462 6-channel recordings of different women, at least 18 years old, between 18 and 24 weeks of gestation. The fetal ECG was recorded with adhesive Ag/AgCI electrodes on the abdomen of the pregnant women while they were in semi-upright position. Six electrodes were placed around the navel to produce six channels of electrophysiological measurements, while two additional electrodes, placed close to the navel, served as common reference

and ground. Each recording lasted from 5 up to 50 minutes and was digitized and stored at 500 Hz sampling frequency by a fetal monitoring system (Nemo Healthcare BV, The Netherlands). Since the signals were measured through six electrodes, we selected the first, third, fourth and fifth dimensions to form the 4-channel fetal ECG signal.

The second real dataset is the Abdominal and Direct Fetal Electrocardiogram Database which consists of 4-channel abdominal fetal ECG recordings obtained by five women in labor, between 38 and 41 weeks of gestation [177]. Each recording comprises four different signals acquired from the maternal abdomen together with a reference direct fetal ECG registered from the fetal head. The configuration of the abdominal electrodes consisted of four electrodes placed around the navel, a reference electrode placed above the pubic symphysis and a common reference electrode placed on the left leg. The recordings have duration of five minutes and are sampled at 1000 Hz.



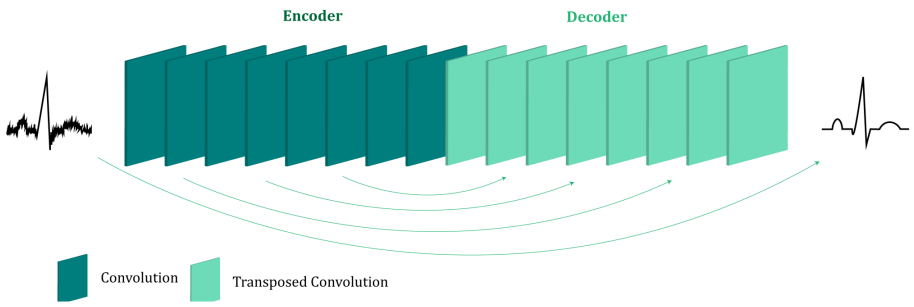



Fig. 6.1. The architecture of the proposed multichannel fetal ECG denoising network. The network consists of an encoder of eight convolutional layers and a decoder of eight transposed convolutional layers, which are linked symmetrically by skip connections.

#### *6.2.1.3 Data preprocessing*

The signals of all the datasets were preprocessed before entering the network. The fetal ECG extraction was performed with the help of the open-source algorithm of Varanini et al. [104] and the signals were resampled to 500 Hz to have a common reference. Finally, the fetal ECG signals were divided in segments of 1920x4 samples and normalized to have zero mean and unity standard deviation. The normalization was performed along each channel separately.

### *6.2.2 Network description*

The proposed fetal ECG denoising CNN network is illustrated in Fig. 6.1. It consists of an encoder of eight convolutional layers and a decoder of eight symmetric transposed convolutional layers. The network receives a noisy fetal ECG signal as input and delivers a denoised one as output. The convolutional layers act as a feature extractor which captures the abstraction of the fetal ECG while eliminating the noise. Subsequently, the transposed convolutional layers decode the fetal ECG abstraction to recover the signal details. The convolutional layers are symmetrically connected with the transposed convolutional ones via skip connections. The role of the skip connections is twofold. First, they help back-propagating the gradients to bottom layers, facilitating the training of our deep network. Second, they pass signal content from the bottom to top layers to aid in recovering the signal details.

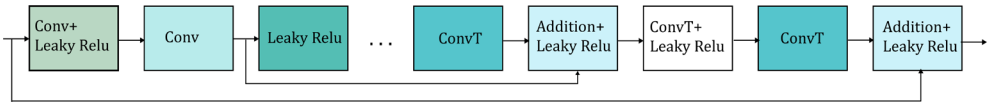



Fig. 6.2. Detailed illustration of the way that the skip connections (represented by the arrows) are applied in the network. Only two skip connections are shown for simplicity. Conv stands for convolution and ConvT for transposed convolution.

The noninvasive fetal ECG typically contains a high amount of noise and thus a large denoising patch can lead to more efficient noise removal by using context information from a larger signal region. It was indicated in the literature that the denoising patch is highly correlated with the receptive field of the network, i.e. the region in the input space that a CNN feature can be affected by [187], [191]. The receptive field of the network is determined by the kernel size, the depth of the network and whether subsampling or dilation is used in the convolution operations. A common approach to increase the receptive field is to increase the number of layers in the network but this is computationally expensive. We chose to use a relatively deep network of eight convolutional and eight transposed convolutional layers. Since our data are temporal, we adopt one-dimensional convolutions and transposed convolutions. In addition, subsampling by two is performed after each convolutional layer, apart from the first, and upsampling by two after the transposed convolutional layers, apart from the last one. Subsampling operations are not generally preferred in denoising tasks in order to preserve the signal details [188]. On the other hand, in our case they lead to a significant increase of the receptive field, necessary for removing the large amount

Table 6.1. Detailed overview of the proposed network architecture.

| | Layer | Output Size | Filter Size | Kernel Size |
|---|---|---|---|---|
| Encoder | Convolution (stride=1) | 1920x64 | 64 | 15 |
| | LeakyRelu(0.2) | 1920x64 | | |
| | Convolution (stride=2) | 960x128 | 128 | 15 |
| | LeakyRelu(0.2) | 960x128 | | |
| | Convolution (stride=2) | 480x256 | 256 | 15 |
| | LeakyRelu(0.2) | 480x256 | | |
| | Convolution (stride=2) | 240x256 | 256 | 15 |
| | LeakyRelu(0.2) | 240x256 | | |
| | Convolution (stride=2) | 120x512 | 512 | 15 |
| | LeakyRelu(0.2) | 120x512 | | |
| | Convolution (stride=2) | 60x512 | 512 | 15 |
| | LeakyRelu(0.2) | 60x512 | | |
| | Convolution (stride=2) | 30x1024 | 1024 | 15 |
| | LeakyRelu(0.2) | 30x1024 | | |
| | Convolution (stride=2) | 15x2048 | 204 | 15 |
| | LeakyRelu(0.2) | 15x2048 | | |
| Decoder | Transposed Convolution(stride=2) | 30x1024 | 1024 | 15 |
| | LeakyRelu(0.2) | 30x1024 | | |
| | Transposed Convolution(stride=2) | 60x512 | 512 | 15 |
| | Addition | 60x512 | | |
| | LeakyRelu(0.2) | 60x512 | | |
| | Transposed Convolution(stride=2) | 120x512 | 512 | 15 |
| | LeakyRelu(0.2) | 120x512 | | |
| | Transposed Convolution(stride=2) | 240x256 | 256 | 15 |
| | Addition | 240x256 | | |
| | LeakyRelu(0.2) | 240x256 | | |
| | Transposed Convolution(stride=2) | 480x256 | 256 | 15 |
| | LeakyRelu(0.2) | 480x256 | | |
| | Transposed Convolution(stride=2) | 960x128 | 128 | 15 |
| | Addition | 960x128 | | |
| | LeakyRelu(0.2) | 960x128 | | |
| | Transposed Convolution(stride=2) | 1920x64 | 64 | 15 |
| | LeakyRelu(0.2) | 1920x64 | | |
| | Transposed Convolution(stride=1) | 1920x4 | 4 | 15 |
| | Addition | 1920x4 | | |
| | Linear Activation | 1920x4 | | |

of noise present in the fetal ECG signals. Moreover, in order to exploit the self-similarity of the ECG signals the network should permit the convolutions to extend to several heartbeats. Regarding the kernel size we empirically determined that 15 achieves satisfactory results by being large enough to include sufficient signal information without excessively increasing the number of network parameters.

The input and output of the network have dimension 1920x4 which corresponds to 4-channel ECG of 3.84 s. For non-linearity after each layer, leaky rectified linear units (LeakyRelu) with a slope of 0.2 are utilized. The aforementioned parameter choices led to a receptive field of roughly 4 sec that corresponds to 5-10 heartbeats. A detailed description of the network architecture and the parameters is given in Table 6.1.

#### *6.2.2.1 Skip connections*

In shallow networks transposed convolutions work well for recovering the signal details but as the network goes deeper, they do not longer work satisfactory [189]. Our network is deep and heavy subsampling is performed for the sake of increasing the receptive field of the network, resulting in significant loss of signal information. To address this issue, skip connections are added between every two convolutional and mirrored transposed convolutional layers as shown by the arrows in Fig. 6.1. The skip connections carry signal information and account to a great extent for the lost signal details introduced by the subsampling. Moreover, these skip connections allow the gradient update rules to back-propagate to the bottom layers directly, dealing with the gradient vanishing problem occurring in deep architectures. The way that the skip connections are used in the network is depicted in Fig. 6.2.

#### *6.2.2.2 Network training*

For training the network the normalized mean squared error loss was minimized, which is defined as:

$$\mathcal{L} = \frac{1}{N*L*M} \sum_{n=1}^{N} \sum_{l=1}^{L} \sum_{m=1}^{M} \frac{\left(X_{clean_{n,l,m}} - X_{denoised_{n,l,m}}\right)^2}{\bar{X}^2_{clean_{n,l}}}, \qquad (6.1)$$

where N is the number of the training data in a batch, L is the number of channels, M is the length of the signals, $X$ represents the fetal ECG and $\bar{X}^2$ is the mean squared amplitude of $X$. In our experiments N=64, L=4 and M=1920. The Adam algorithm was selected [152] as an optimization algorithm while the learning rate was set to 0.00001. The training method that we followed is supervised, meaning that we need clean fetal ECG signals as labels together with the noisy signals. For this reason, the training of the network was performed based only on simulated data. The simulated data were separated in two sets for the training and testing of the method. The training set contains the signals simulated by the modified fecgsyn toolbox based on VCG 1-7 and 449 preprocessed records from 212 subjects of the PTB dataset. The

test set contains the simulated signals based on VCG 8-9 from the modified fecgsyn toolbox, plus 100 preprocessed records of 78 subjects of the PTB dataset. The SNR of the training set ranges from -15 to 15 dB. The network was trained for 21 epochs until convergence was reached.

### 6.2.3 Performance Evaluation

In the simulated dataset, the performance of the method was evaluated based on the SNR improvement of the fetal ECG signals achieved by the network. The metric is estimated for a channel, *l*,of a signal as:

$$SNR_{imp} = 10log_{10}\frac{\sum_{m=1}^{M}\left|X_{noisy_{l,m}} - X_{clean_{l,m}}\right|^2}{\sum_{m=1}^{M}\left|X_{denoised_{l,m}} - X_{clean_{l,m}}\right|^2}. . \quad (6.2)$$

The metric was computed for each channel and subsequently averaged over all the ECG channels and test signals.

For real fetal ECG signals there is no ground truth available, because even after the maternal ECG suppression there is still noise present in the signals. Thus, it is impossible to have a gold reference to quantitatively validate the results. Simultaneous scalp recordings may help but they can be performed only during labor. Unfortunately, since our real private dataset was obtained during the second trimester of pregnancy, it was not possible to measure the scalp ECG to have a clean reference. For this dataset, in order to provide some quantitative results along with the qualitative, we decided to generate a surrogate "clean" ground truth signal by calculating the running median of 100 heartbeats. We then measure how well the quality of the denoised signals was enhanced by computing the improvement in SNR performance defined by Equation (6.2). The metric was calculated for 455 cases, where sufficient QRS complexes were detected for the generation of the "ground truth" signal.

In the Abdominal and Direct Fetal Electrocardiogram Database, since simultaneous scalp measurements are provided together with the noninvasive fetal ECG, the performance of our method was evaluated by comparing with the scalp electrode. The scalp ECG is however a different lead than the abdominal ones and we cannot compare them directly since, even in case of perfect denoising by our method, the morphology of

the ECG will not be the same between different leads. Instead, we estimated a denoised scalp ECG as a linear combination of the four abdominal fetal ECG channels:

$$\hat{X}_{scalp} = a^T X_{denoised},\ a = (X_{denoised} {X_{denoised}}^T)^{-1} X_{denoised} X_{scalp}^T, \quad (6.3)$$

where $X_{scalp}$ is the [1 x 250] scalp ECG and $\hat{X}_{scalp}$ the [1 x 250] estimation of the scalp ECG from the abdominal fetal ECG channels. The coefficients of the [4 x 1] linear combination, *a*, were computed on windows of half a second that corresponds to 250 samples. The dimension of $X_{denoised}$ is 4x250. Because the scalp ECG measurements contain considerable amount of noise and this could affect the comparison, we denoised the scalp ECG by high pass filtering followed by averaging of 30 ECG complexes. Nevertheless, we provided comparative results both when the estimation was done based on the noisy scalp ECG as well as on the denoised scalp ECG.

Four different quantitative measures were employed for the comparison, the Pearson correlation coefficient (*R*), the mean squared error (MSE), the mean absolute error (MAE) and the signal-to-noise ratio (SNR). The metrics are defined by the following equations:

$$R = \frac{cov(\hat{X}_{scalp} X_{scalp})}{\sigma_{\hat{X}_{scalp}} \sigma_{X_{scalp}}}, \quad (6.4)$$

$$MSE = \frac{1}{K} \sum_{i=1}^{K} (X_{scalp_i} - \hat{X}_{scalp_i})^2, \quad (6.5)$$

$$MAE = \frac{1}{K} \sum_{i=1}^{K} |X_{scalp_i} - \hat{X}_{scalp_i}|, \quad (6.6)$$

$$SNR = 10 log_{10} \frac{\sum_{i=1}^{K} \left| X_{scalp_i} \right|^2}{\sum_{i=1}^{K} \left| X_{scalp_i} - \hat{X}_{scalp_i} \right|^2}, \quad (6.7)$$

where *cov* stands for the covariance, $\sigma$ the standard deviation and K the length of the signals. The metrics were computed for the five signals of the database and subsequently averaged to obtain one final value.

### 6.2.4 Reference methods

Our method was evaluated in comparison with 3 other ECG denoising methods. The first method is the single channel CNN denoising network, where each fetal ECG channel is denoised separately [137]. The second method is a wavelet denoising

algorithm that removes the noise by thresholding the detail coefficients after the signal decomposition. The symlet wavelet was selected due to its resemblance with an ECG, while a fixed threshold was used, estimated by the minimax principle [178]. The last method is the widely used beat-to-beat averaging method. We selected to average 30 beats similar to the averaging performed by the STAN method [25]. The QRS complexes were detected by a Pan Tompkins detector in the clean fetal ECG signals and not the noisy ones because we do not intend to assess the performance of the QRS detector but the performance of the averaging method. However, we should note that it is not guaranteed that the QRS complexes can be accurately estimated in the presence of acute noise.

## 6.3 Results



### 6.3.1 Performance on Simulated Signals

The improvement in SNR performance of the proposed network in comparison to the other denoising algorithms, for input SNR from -20 to 20 dB, is illustrated in Fig. 6.3. As demonstrated in this figure, the CNN network provides a considerable amount of SNR improvement throughout the whole range of input SNR. The proposed method outperforms the beat-to-beat averaging and the wavelet denoising methods for all the input SNR values. This was anticipated because the averaging method does not preserve individual variations among complexes, while our method is capable of doing so. Moreover, the wavelet denoising distorts the signal amplitude, whereas the proposed network preserves it better. The multichannel network additionally outperforms the single channel nearly for the whole range of input SNR values. More specifically, for input SNR less than 0 dB the multichannel algorithm provides an SNR improvement of at least 10 dB with respect to the input signal and at least 2 dB further improvement as compared to the single channel method. As the input SNR increases the performances of the two methods become gradually comparable, while for input SNR more than 11 dB the single channel network slightly surpasses the multichannel. This was something to expect because for signals of lower quality, information from multiple channels will be beneficial for recovering the ECG structure. On the other hand, if a fetal ECG channel has sufficiently high quality not only the other channels are unnecessary for denoising it but could also slightly affect the quality of the denoised fetal ECG, especially in case their SNR is low. This can be

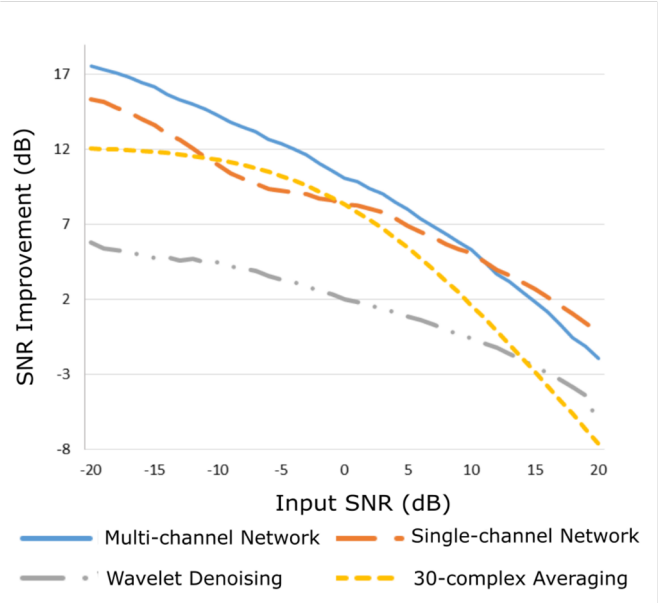



Fig. 6.3. Performance of the proposed multichannel convolutional network in comparison with other denoising methods in terms of improvement in SNR of the denoised fetal ECG signals when compared with the noisy ones.

explained better by the following: By using any set of three linearly independent ECG leads, the VCG can be constructed, which is the three-dimensional representation of the electrical activity of the heart. A VCG can explain about roughly 90% of an ECG signal [57]. This means that when a signal is reconstructed from different channels, 10% of the signal information should be considered as not reconstructable. In case of very high signal quality, the single channel denoiser can perform better than the multichannel since it could theoretically reconstruct 100% of the signal.

By observing Fig. 6.3, we see that, for all methods, there is an input SNR for which the denoisers decrease the SNR. This input SNR value is 9,12, 18 and 20 dB for the wavelet, averaging, multichannel network and single channel network denoising methods respectively. Since it is not common to obtain fetal ECG signals of very high quality (more than 18dB), we do not consider it as a limitation of our method. We additionally noticed that there is a upgoing trend for the SNR improvement metric as the input SNR decreases. However, we did not test for signals of quality even lower than -20 dB because real fetal ECG signals typically do not have quality less than -20dB.

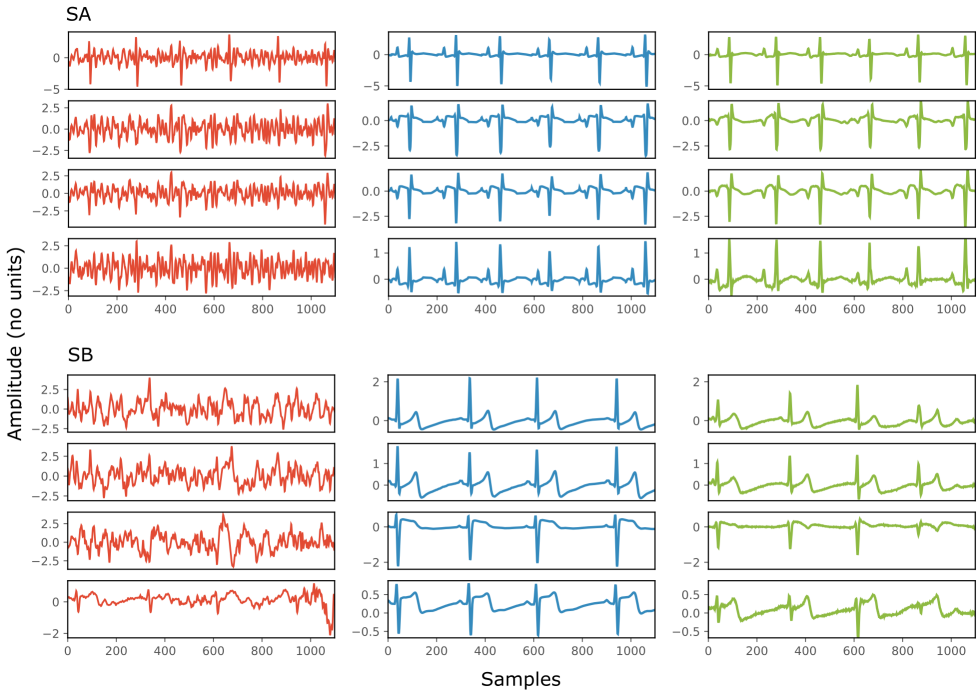



Fig. 6.4. Denoising results by the proposed method for two simulated signals (SA and SB) of the test dataset. For both signals: each panel in the left presents one channel of the noisy 4-channel fetal ECG signal (red), in the middle the corresponding channels of the clean signal (blue) are shown and in the right the denoised fetal ECG signal by our network (green). The horizontal axis depicts the samples at 500 Hz, while the vertical the amplitude of the signals. The SNR values of the noisy and the denoised fetal ECG for both signals are given in Table 6.2 (SNRin and SNRout respectively).

Fig. 6.4 depicts two typical denoising results from our test dataset. The SNR values of the signals before and after denoising are provided in Table 6.2. Note that in Fig. 6.4 the vertical axes limits for the noisy signals differ from those of the ground truth and denoised signals for better visualization. However, the axes limits for the clean and denoised fetal ECG are the same to allow for their comparison. As can be noted, the network suppresses the noise in a great extent for both signals simulated-A (SA) and simulated-B (SB). In the case of signal SA the similarity of the network's output with the clean signals is very high for all channels and all ECG waves are clearly distinguishable.

Table 6.2. The SNR values in dB for the four channels of the simulated signals depicted in Fig. 6.4, before (SNRin) and after (SNRout) denoising.

| Channel | Signal SA | | Signal SB | |
|---|---|---|---|---|
| | SNRin | SNRout | SNRin | SNRout |
| 1 | 1 | 17 | -9 | 5 |
| 2 | -2 | 13 | -9 | 4 |
| 3 | -3 | 14 | -10 | 3 |
| 4 | -12 | 8 | 0 | 3 |

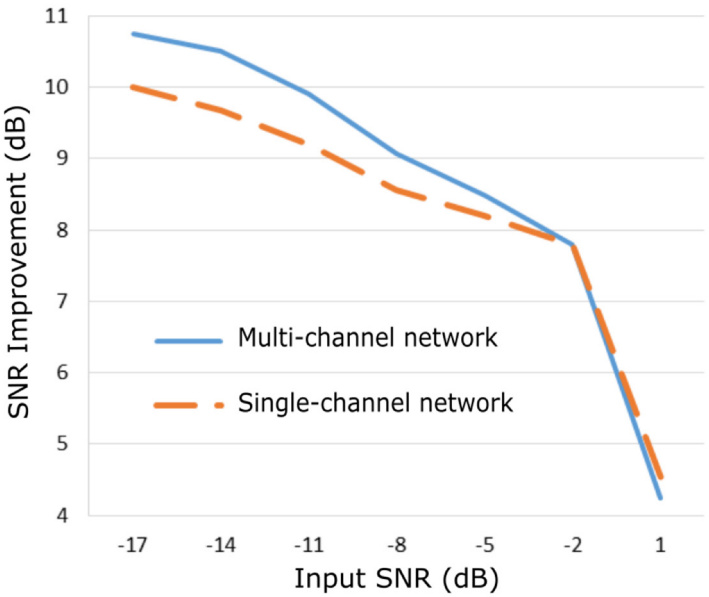



Fig. 6.5. Performance of the proposed fetal ECG denoising method in a large real dataset [174] in terms of improvement in SNR of the denoised signals.

Even for channel 4, with input SNR of -12 dB, the network provides a high-quality result, since it combines all channels to reconstruct it. For signal SB the majority of ECG channels have very low quality (around -9 dB). The SNR after denoising with our network is significantly higher (3.75 dB on average). However, we notice some distortion on the signal amplitude, while particularly the P-waves are suppressed by the network. Moreover, despite channel 4 having the least amount of noise before entering the network, we observe the least improvement after denoising, evidencing that indeed the network's output is obtained through combination of information from all leads.

### 6.3.2 Evaluation on Real Fetal ECG Signals

The proposed method was evaluated on our extensive noninvasive fetal ECG dataset [174] and the results are presented in Fig. 6.5. Fig. 6.5 illustrates the improvement in SNR for input SNR ranging from -17 to 1 dB. The input SNR corresponds to the SNR of the noisy fetal ECG signals when we assume that the ground truth signal is the running median of 100 heartbeats. We need to stress that this is not the actual SNR of the signals but merely an approximation of it. In fact, the more noise is present in the signals or the more physiological variation, the less accurate the constructed "clean" signal is. Examining the Fig. 6.5 and 6.3, where the performance in the simulated dataset is illustrated, we observe an analogy between them. In both graphs the multichannel denoiser surpasses the single channel for lower input SNR while for higher SNR values the two methods perform comparably. The performance improvement as compared to the single channel approach is lower for the real signals than for the simulated ones, but this might be due to the lack of actual ground truth signals for comparison. By all means the evaluation in this dataset is suboptimal but it provides a performance indicator on a large real dataset.



Fig. 6.6 demonstrates the result of denoising two signals of this database, while Table 6.3 provides the corresponding SNR values before and after denoising. Note that the vertical axes limits for the noisy signals differ from the ones of the "clean" and denoised ones for clearer visualization. Both signals in Fig. 6.6, especially signal real-B (RB), have a significant amount of noise before denoising (see Table 6.3). The "clean" reference signals as well contain few noise but in most of them the ECG morphology is relatively clear. On the other hand, all the possible variations among the successive complexes is lost due to the heavy averaging performed. The multichannel network achieved a fairly remarkable result in denoising those signals. Comparing the morphology of the denoised with the "clean" reference signals, the various ECG waves and segments correspond relatively well. In this comparison, we acknowledge that the running median of 100 heartbeats is not the gold standard.

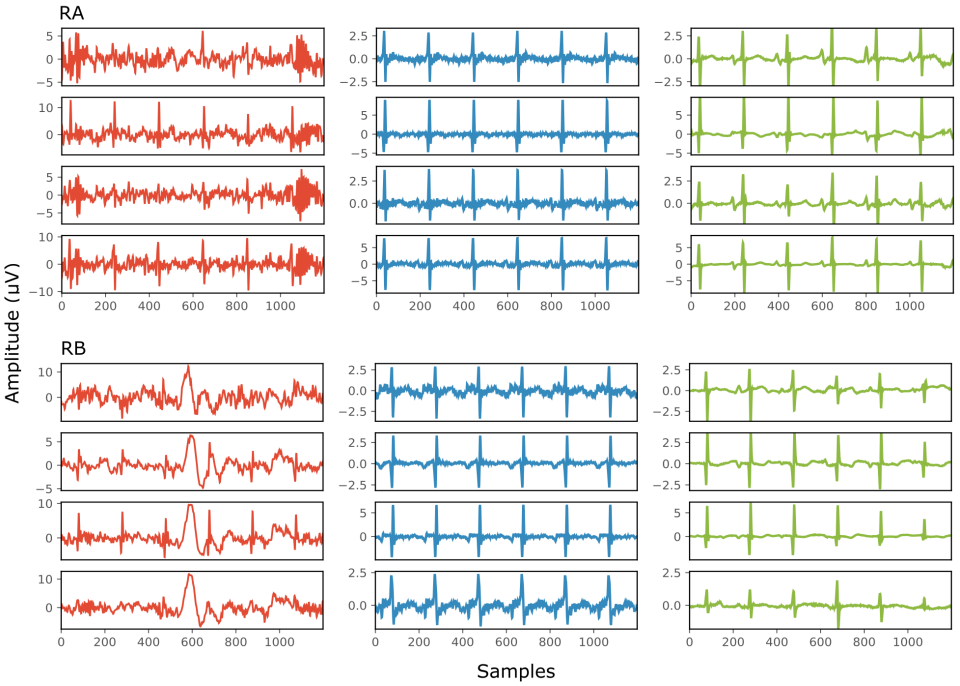



Fig. 6.6. The results of denoising two signals (RA and RB) from our private real fetal ECG dataset [174]. For each signal: the noisy 4-channel fetal ECG signal extracted from the abdominal measurements is presented in the left (red), the running median of 100 heartbeats for each channel in the middle (blue) and the denoised fetal ECG signal by our network in the right (green). The horizontal axis depicts the samples at 500 Hz, while the vertical the amplitude of the signals in μV. The SNR values of the noisy fetal ECG signals (SNRin) together with the values for the denoised ones (SNRout) are given in Table 6.3.

However, taking the median of the heartbeats brings evidence for the location of the ECG waves, especially the P-waves, information that cannot be seen in the noisy signals. It is important to recognize that in the denoised signals by our network, these locations seem to correspond with the locations in the median signals. As a matter of fact, the denoised signals appear to exhibit better quality and clearer morphology than the reference. Some morphological features seem to be distorted, as we can see in Fig. 6.6 for signal RB. However, the overall performance in those low-quality signals is relatively good.

Table 6.3. The SNR values in dB for the four channels of the real signals depicted in Fig. 6.6, before (SNRin) and after (SNRout) denoising.

| Channel | Signal RA | | Signal RB | |
|---|---|---|---|---|
| | SNRin | SNRout | SNRin | SNRout |
| 1 | -8 | 3 | -13 | 3 |
| 2 | -3 | 3 | -8 | 5 |
| 3 | -8 | 2 | -6 | 6 |
| 4 | -1 | 6 | -14 | 1 |

Fig. 6.7 illustrates the performance for a fetal ECG signal of our noninvasive fetal ECG dataset in comparison to the single channel network, 30-complex averaging and wavelet denoising. For simplicity we present only one channel out of the four. As shown in the figure, all methods provide a noise-free result. However, our method retains the individual ECG complex differences as opposed to the averaging method and does not distort the signal amplitude as opposed to wavelet denoising. In addition, the morphology of the denoised ECG is clearer in our case. The single channel network provided a similar result to the multichannel for this signal.

The performance of the network on the Abdominal and Direct Fetal ECG Database is illustrated in Table 6.4. The scalp ECG was compared with the aforementioned linear combination of abdominal signals, as described in Equation (6.3). In Table 6.4 we provide the results of this comparison for 2 cases; when we used the original scalp ECG and when we denoised it. For each performance metric the values before and after denoising with the multichannel and single channel network are presented, while with bold the best performing method is marked.

First, we believe that denoising of the scalp ECG was important to allow for better comparison with the scalp ECG estimation from the denoised abdominal leads. By averaging 30 successive ECG complexes we might have lost some morphological variations among the successive beats of the scalp lead but achieved significant quality improvement. Even the scalp ECG approximated by the noisy fetal ECG signals has better resemblance with the denoised scalp lead, e.g. correlation coefficients of 0.74 vs 0.53. Second, we observe that both the multichannel and single channel networks achieve significant quality improvement of the fetal ECG signals for all the metrics presented in Table 6.4. We should note here once more that by no means the scalp estimation is expected to be the same with the scalp ECG even after perfect denoising,

because the latter is a different lead than the abdominal leads. Last, the multichannel network outperforms the single channel in terms of all computed performance metrics.

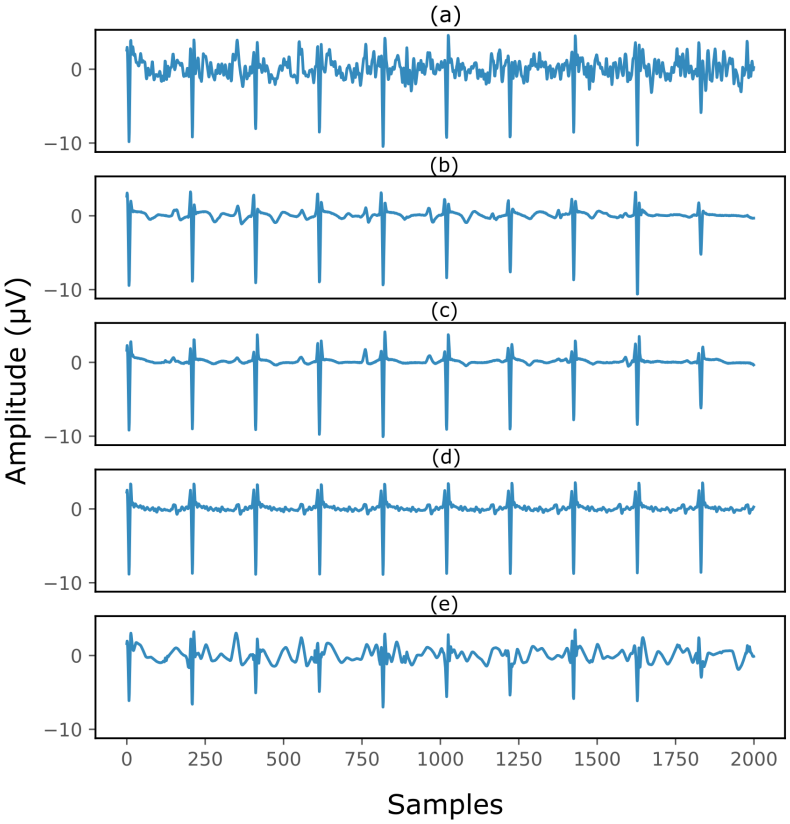



Fig. 6.7. The result of denoising a real fetal ECG signal from our private dataset [174] with different denoising algorithms. For simplicity, only one channel is displayed. The panels show: (a) the noisy extracted fetal ECG, (b) the denoised signal by the proposed method, (c) the denoised signal by the single-channel denoising network [137], (d) the result of 30-complex averaging and (e) the result after wavelet denoising. The horizontal axis depicts the samples at 500 Hz, while the vertical the amplitude of the signals in μV.

However, the differences are relatively small. It might be because the extracted fetal ECG signals already have decent quality and, as we have already found in simulated signals, employing multiple channels is more advantageous in cases of signals exhibiting lower SNR. Larger difference was found regarding the MSE metric (62.1 vs 68.4 $\mu V^2$), indicating that the single channel network may provide more outliers, while the multichannel a smoother outcome.

Table 6.4. Performance of the multichannel CNN network vs the single-channel one [137] on the Abdominal and Direct Fetal ECG Database [177] in terms of comparison of the scalp ECG with a scalp estimated from the denoised abdominal fetal ECG.

| Metric | Original scalp ECG | | | Denoised scalp ECG | | |
|---|---|---|---|---|---|---|
| | Noisy Input | Multichannel Output | Single-channel Output | Noisy Input | Multichannel Output | Single-channel Output |
| **R** | 0.53 | **0.66** | 0.65 | 0.74 | **0.87** | 0.85 |
| **MSE ($\mu V^2$)** | 555.8 | **440.7** | 449.3 | 116 | **62.1** | 68.4 |
| **MAE ($\mu V$)** | 15 | **12.9** | 13 | 7.3 | **5.4** | 5.5 |
| **SNR (dB)** | 1.5 | **2.7** | 2.5 | 3.7 | **6.4** | 6.1 |

Fig. 6.8 provides two qualitative results of the scalp estimation, when fetal ECG denoising is performed with the proposed multichannel method. In both cases, the scalp estimated by the denoised fetal ECG is free from noise and the individual waves and intervals correspond relatively well to those of the scalp ECG. We do not expect absolute correspondence, not only because the scalp ECG is a different lead, but also because it was averaged over 30 complexes.

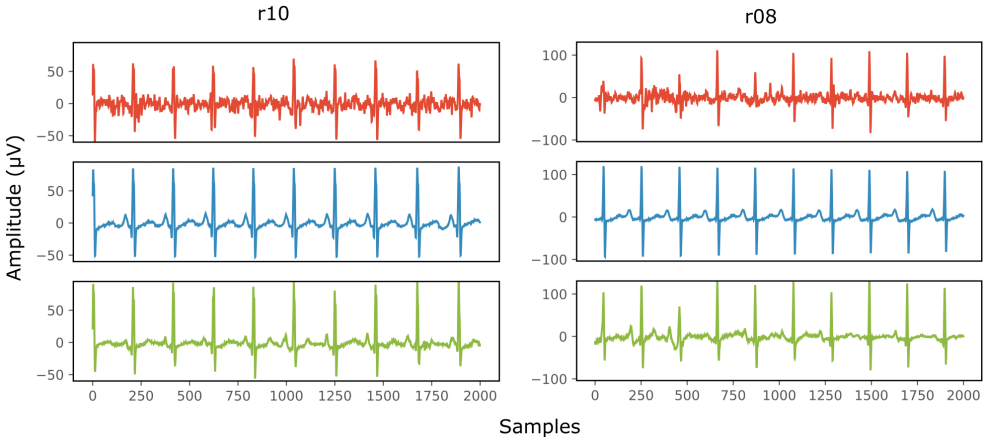



Fig. 6.8. Comparison of the scalp ECG with an estimation of it as a linear combination of the abdominal fetal ECG for 2 records (r10 and r08) of the Abdominal and Direct Fetal ECG Database [177]. In the first row (red) the scalp ECG estimated from the noisy fetal ECG is presented, in the second (blue) the ECG as measured by the scalp electrode (running average of 30 complexes) and in the last one (green) the scalp ECG estimated from the denoised fetal ECG by the proposed network. The horizontal axis depicts the samples at 500 Hz, while the vertical the amplitude of the signals in μV.

## 6.4 Discussion

We have proposed a CNN network for postprocessing noninvasively extracted multichannel fetal ECG signals to improve their quality. The noninvasive fetal ECG is substantially contaminated by various noises, even after the application of various signal processing tools proposed in literature, such as maternal ECG suppression. The low quality of the fetal ECG is the principal reason that the applicability of noninvasive fetal electrocardiography in clinical practice is limited.The suggested multichannel network was trained on a wide dataset of simulated 4-channel ECG signals, with SNR ranging from -15 to 15 dB, while it was extensively validated both on simulated as well as on real datasets.

Experiments on simulated data showed a significant improvement in the quality of the noise-corrupted fetal ECG signals. The network combined information from all the channels to efficiently remove the noise and uncover the ECG signal morphology even in the presence of acute noise. However, the network suppressed some morphological characteristics in cases there was not sufficient content for denoising i.e. when most signal channels were severely corrupted. The multichannel network outperformed the single channel [137] in cases of low SNR of the input signals, while for SNR more than 11 dB the single channel network exhibited slightly better performance. This behavior could be anticipated. A multi-lead signal configuration captures the spatiotemporal nature of the cardiac electrical activity. For low quality signals this is beneficial as more signal information can be exploited to better reconstruct each channel. However, if we wish to denoise a channel that already has high quality, using spatiotemporal content may be not always the best choice. Nevertheless, it is very uncommon in practice to obtain fetal ECG of such high quality. Yet, in case this would happen, the output of the multichannel network would still be of such quality that it could be used for further clinical interpretations.

The evaluation of our network on a large real fetal ECG dataset showed an analogous behavior to that on the simulated data; for low quality fetal ECG the multichannel network outperformed the single channel, while for higher SNR the performances of the networks were comparable. We cannot make a direct comparison because the evaluation method for the real signals was suboptimal. We are aware that the approximation of the ground truth signals with the running median of 100 heartbeats was not very accurate. However, it gave us an indication that the method is efficient in real data too. We additionally presented some qualitative denoising results for two

signals of this database in Fig. 6.6 to support our claim. The network outputted clean denoised signals with good correspondence of the individual ECG waves between the reference and denoised signals. A few recordings in our dataset had input SNR that was even lower than -17 dB. Based on visual analysis of the output of our proposed denoiser, we could argue that the performance of the denoiser breaks down at these very low signal quality levels and the network is no longer capable of reconstructing a reliable fetal ECG. This limitation probably comes from the fact that the network was trained for input SNR range of -15 to 15 dB. Thus, the network did not learn to remove efficiently the noise when the quality of the signals is even lower. This indicates that we might need to perform experiments for a wider SNR input range. However, the capacity of the network might no longer be sufficient for handling such an ample range of signal qualities and further research is needed to evaluate this.

The CNN network was additionally evaluated in the Abdominal and Direct Fetal ECG Database. Simultaneously recorded scalp ECGs were compared to an estimated scalp ECG from the denoised abdominal channels, also here demonstrating that the method can provide significant quality improvement of the noisy fetal ECG signals. Comparison of the performances of the multichannel and single-channel networks for this database, revealed that they achieve comparable results, probably because the input signals were of relatively good quality. It is difficult to compare the performances between the two real datasets for several reasons. Most importantly, the sizes of the two datasets differ a lot (455 vs 5) and so do the gestational ages of the subjects (18-24 vs 38-41 weeks).



As mentioned in the introduction and also in [137], the shortcoming of denoising single channel fetal ECG with a convolutional network is that the network can output signals that look as if they were ideally denoised, but that can have “fake” waves that can differ both in location and polarity when compared to the actual ECG waves. This happens mostly when the quality of the input signals is relatively low and the network, not having enough signal information, reconstructs a clean signal from unreliable information in the encoded latent space. We demonstrated that by employing multichannel signals this problem is eliminated to a large extent. When the quality of the signals is very low, the amplitude of the small signal waves, like the P-wave and T-wave, and less often of the R-peaks in the denoised signals can be distorted rather than “fake”. This means that some waves may be virtually absent, or the output does not even resemble an ECG anymore. This makes the method safer to use in clinical practice, because clinicians will typically discard a distorted signal

but a signal that looks like a high-quality ECG but in fact contains "fake" information might lead to erroneous decision-making.

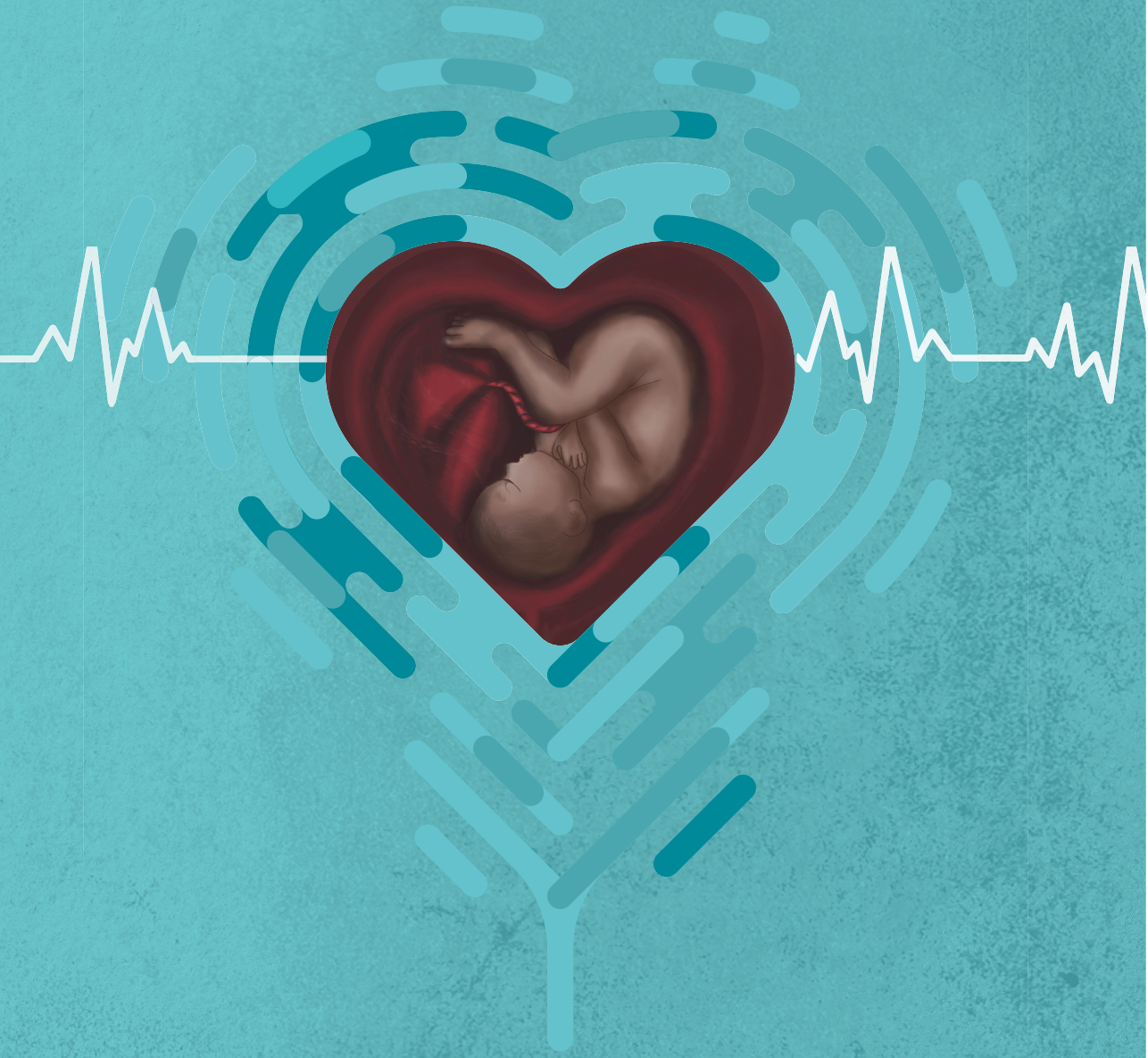
To summarize, we have shown the potential of deep CNNs for removing noise from noninvasive multi-lead fetal ECG. We validated the method on a wide dataset of simulated but also real recordings with both early as well as late gestational ages (18 to 24 and 38 to 41 weeks). Primarily, we demonstrated that employing multichannel information for denoising does not only lead to more clean signals but also to more reliable results, when compared to single-channel information. The main advantage of the method is that, as opposed to the widely used averaging method, no prior processing of the signal is needed to extract the locations of the R-peaks and variations in ECG morphology among consecutive heartbeats are preserved. This is especially important in case that arrhythmias are present. Up to now, arrhythmia is assessed through echocardiography because the averaging that was performed to enhance the quality of the fetal ECG hinders its application for arrhythmia analysis. Moreover, the quality of the denoised signals is high enough to allow for measuring the timing of intervals, like the PR and QT interval. However, in order to confirm this, we need to perform a thorough comparison of the ECG intervals between the denoised and the clean signals. If we wish to obtain reliable results, a large annotated dataset is necessary, but this requires time and experts to perform these annotations.

Certainly, there is room for improvement of our method. Most importantly, the capacity of the network could be further increased to handle even more noisy signals. Moreover, we can explore denoising directly the raw abdominal signals, without cancelling the maternal ECG. Most probably a more complex network architecture is needed for such a task and appropriate data for training.

## 6.5 Conclusion

An end-to-end trained deep CNN network was presented for denoising of fetal ECG signals. Convolutions and transposed convolutions were combined in the network, modeling the denoising problem as an encoding of primary signal content and subsequent decoding to recover details. Essentially, we proposed to employ spatiotemporal information in the ECG signal by using multiple ECG leads simultaneously as input to the network. The network then learned how to combine the input channels and deliver a reliable clean ECG as output. Our experiments

showed an SNR improvement of 9.5 dB for simulated fetal ECG signals with input SNR in the range [−20 ,20] dB and 8.7 dB for real fetal ECG signals with input SNR in the range [-17,1] dB. The proposed network can achieve a substantial quality improvement of the noisy signals and outperform a single-channel alternative.



# CHAPTER 7

# DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

## 7.1 Discussion

Monitoring the condition of the fetus is of vital importance for timely detection of fetal distress and appropriate treatment or intervention. Currently, fetal monitoring is performed through cardiotocography (CTG), the simultaneous registration of fetal heart rate (HR), averaged over a number of cardiac cycles, and uterine contractions. However, while CTG is associated with an increase in interventions, it has shown no remarkable improvement in perinatal death and cerebral palsy rates. Moreover, Doppler ultrasound, which is mainly used in CTG monitors, suffers from signal loss due to fetal motion and in preterm pregnancies. This is especially evident in obese women because of the higher attenuation of ultrasound. It has been demonstrated that CTG in combination with fetal electrocardiogram (ECG) significantly reduces the rates of metabolic acidosis and operative deliveries [207]. Unlike Doppler ultrasound, more reliable beat-to-beat fetal HR can be obtained by the fetal ECG. Next to this, analysis of the fetal ECG waveform provides additional diagnostic information for assessing the fetal condition [12], [36], [16]. However, usually in clinical practice the fetal ECG is obtained invasively and therefore can only be used during labor, after sufficient cervical dilatation and after the membranes have ruptured.

The fetal ECG can be measured noninvasively by placing electrodes on the maternal abdomen. In this way, fetal monitoring with fetal electrocardiography can be performed during both pregnancy and labor and may represent a valid an alternative to conventional monitoring methods. Currently, the use of noninvasive fetal electrocardiography in clinical practice is limited by the fact that the abdominal recordings are severely corrupted by many electrical interferences and noises resulting in fetal ECG signals of low signal-to-noise ratio (SNR). Extensive research focused on removing the maternal ECG, which is the dominant interference [43]. However, even after the maternal ECG has been removed, the extracted fetal ECG signals typically have low quality. The low signal quality hampers the analysis of the fetal ECG waveform and in many cases even the reliable extraction of the fetal HR, despite the highest amplitude of the fetal R-peaks. In the first part of this thesis we focused on improving the reliability of fetal HR detection. In the second part we focused on postprocessing of the extracted fetal ECG signals for enhancing their quality.

In many practical situations, maternal ECG signal removal is not sufficient to ensure reliable extraction of the fetal HR from noninvasive fetal ECG recordings. As an

example, during labor the maternal abdominal muscles are particularly active, causing interferences that exceed in amplitude the fetal ECG while overlapping with it in the frequency domain. Fetal HR extraction typically requires fetal R-peak detection, which is very challenging especially in these cases of low SNR. Moreover, since there are no accurate models to characterize the residual noise in the extracted fetal ECG signals, traditional signal processing methods show relatively limited performance in determining the fetal HR. In Chapter 3, an inception network was presented that directly estimates the fetal HR from the extracted fetal ECG signals without the intermediate step of fetal R-peak detection. The network combines convolutional neural networks (CNNs) with long short-term memory networks (LSTMs) to capture both short-term and long-term temporal relations of the fetal HR. Despite the good performance of this network in the first stage of labor, when the fetal ECG signals have extremely low quality, e.g. during second stage of labor, it is very challenging to reliably estimate the fetal HR. During the second stage of labor uterine contractions are stronger and more frequent, and the woman is actively pushing, contaminating even further the abdominal recordings with intense muscle noise. To increase the reliability of the method and its suitability for clinical application, a CNN-based classifier was developed that estimates the accuracy of the detected fetal HR. The proposed method outperformed other state-of-the-art methods proposed in the literature in terms of fetal HR detection accuracy.



To widely adopt noninvasive fetal electrocardiography in clinical practice, quality improvement of the extracted fetal ECG signals is of key relevance. Even after the application of powerful fetal ECG extraction algorithms, several instrumental and physiological noise sources are still present in the extracted fetal ECG signals obscuring the morphology of the signals. Adaptive filters have been extensively used to remove uncorrelated noise components in cases that the noise characteristics are a priori unknown. Typically, in noninvasive fetal ECG recordings the fetal ECG components measured by different electrodes are correlated, while the noise components, such as muscle noise, are mostly uncorrelated. However, since the fetal ECG signals are highly non-stationary a least mean squares (LMS) adaptive filter cannot track these rapidly varying non-stationarities, resulting in severe signal distortion when applying LMS adaptive filters. In Chapter 4, an augmented time-sequenced adaptive filter (aTSAF) was presented to address the limitations of the LMS filter on enhancing the quality of the extracted multichannel fetal ECG signals. Fetal ECG complexes that originate from consecutive heart beats have similar statistical properties. The aTSAF exploits this property and uses a separate

adaptive filter for each separate segment of a fetal ECG complex. Since we need to align the ECG complexes in time, the R-peak locations need to be a priori known to apply this method. The aTSAF enhances the quality of each separate fetal ECG channel by using the other channels as reference, meaning that the denoising performance highly depends on the quality of the reference channels. Based on this reasoning we enhanced the quality of the reference channels by ensemble ECG complex averaging, prior to the filter application. According to experiments on both simulated and real signals the aTSAF method was found to be efficient in reducing major noise components and revealing the ECG signal morphology. Moreover, aTSAF outperforms other fetal ECG enhancement methods, especially in cases of signals with low SNR.

The aTSAF method requires the R-peaks locations to synchronize the filter but as mentioned above it is not always possible to obtain an accurate estimation of them. Moreover, the filter needs time to adapt its coefficients, making it inefficient in adapting to abrupt changes of the fetal ECG morphology that may occur e.g. in the presence of arrhythmia or ectopic beats. Besides that, short-term variations in the ECG waveform are lost because of the averaging filters applied to the reference channels. However, aTSAF can be used when we are interested in detecting fetal hypoxia, as changes in the fetal ECG waveform that are caused from developing hypoxia happen at longer times scales. Next to that, structural heart malformations affect similarly each fetal ECG complex and thus screening of congenital heart defects (CHDs) should be also possible.

Yet, to address the issue related to the need for a priori knowledge of fetal R-peak locations by the aTSAF, in Chapter 5, another fetal ECG denoising method was proposed that is based on deep learning. The presented method does not need any prior knowledge on the location of the fetal ECG complexes and is able to preserve beat-to-beat variations in the ECG signal morphology. A deep fully convolutional encoder-decoder network was developed that learns end-to-end mappings from noise-contaminated single-channel fetal ECG signals to clean ones. The encoder of the network captures the ECG signal abstraction while eliminating the corruptions and the decoder upsamples the feature maps and recovers the signal details. Since our network is deep and difficult to train, we added skip connections between the corresponding layers of the encoder and decoder to achieve faster convergence and better denoising results. Our method is supervised, meaning that we need noisy fetal ECG data together with the corresponding clean data for training the network.

However, the fetal ECG signals even after maternal ECG suppression are still affected by noise and, as such, we do not have clear reference data. To address this issue, we decided to train the network on simulated data and evaluate its performance both on real and simulated data. To minimize risks of overfitting, we created a large simulated fetal ECG dataset for training where we incorporated many variations in the fetal ECG morphology and added noise. The large capacity of our network allowed us to cope with different levels of noise in a single model. Experiments on simulated and real data showed that the network can achieve substantial quality improvement of the noisy fetal ECG signals. However, in cases of very low SNR, the method failed to reliably reconstruct the fetal ECG signal morphology, occasionally causing artificial (i.e. fake) ECG waves.

To address the latter issue, in Chapter 6 our network was adapted to deal with multiple fetal ECG channels. Multiple electrodes measure the electrical activity of the heart from different angles, leading to distinct fetal ECG signals yet with corresponding locations of ECG segments and waves. By combining information from different channels, convolutions can extend also in space rather than in time only, exploiting the correlations among the different channels and, therefore, leading to more reliable signal reconstruction. Indeed, we demonstrated that by employing multichannel signals the limitation of the single channel denoising network was eliminated to a large extent. Our experiments showed a significant quality improvement of the noise-corrupted signals both in simulated and real data. In cases when the majority of the channels were severely corrupted the network distorted some morphological characteristics but did not create any fake information. The proposed method outperformed not only its single-channel network counterpart but also other widely used fetal ECG denoising algorithms such as ECG complex averaging and wavelet denoising. The main advantage of our method is that it retains individual variations among different ECG complexes that is necessary for the study and detection of fetal arrhythmias and that is does not require a priori knowledge on fetal R-peak locations.



## 7.2 Future directions

Fetal electrocardiography exists for more than one century as the first recordings were made by Cremer in the early 1900s [208]. However, difficulties in acquiring and processing the electrophysiological signals but also lack of availability of gold standard databases limited the development of this technique. Nowadays in clinical

practice the fetal ECG is mainly monitored using an invasive scalp electrode and used to extract the fetal HR and perform ST segment analysis. To overcome the limitations of the invasive measurements, over the last years the field of noninvasive fetal electrocardiography has regained some interest as an alternative to electronic fetal monitoring. Few solutions have already been introduced in clinical practice for CTG acquisition based on noninvasive electrophysiological measurements. However, the performance of these devices can be rather poor in some cases, e.g. during the second stage of labor. Moreover, there are no available solutions for unobtrusive extraction and analysis of ECG morphological parameters. Thus, there is still ample room for improvement in the area of noninvasive fetal electrocardiography and significant advances should be made before its clinical uptake as standard monitoring technique. The methods presented in this thesis are intended to make a step in this direction. However, several improvements need to be made before the proposed methods can be used in clinical practice. The most critical improvement points for each method are discussed below.

The fetal HR extraction method, presented in Chapter 3, sometimes fails to make correct estimations during the second stage of labor. At this stage the fetal ECG signals have extremely low quality, making the fetal HR extraction very challenging, especially during HR decelerations. This was the reason why we additionally developed a fetal HR reliability classifier to identify the periods with correct estimations. However, our classifier was found to reject too many data of good quality (i.e. it has low sensitivity), especially when tested on a public dataset of Physionet. This Physionet dataset is rather small, comprising 68 1-minute recordings and containing only relatively good quality data that lead to correct fetal HR estimations. Since this set is small and contains mostly good-quality examples, it is difficult to make conclusions about the generalizability of our classifier. Thus, we do need a larger and more diverse test dataset, also including data from second stage of labor, to evaluate our classifier. Having more data available, the parameters of the classifier network can also be optimized, and the architecture can be refined to achieve improved classification performance. Nevertheless, we should improve not only the classification of the accuracy of fetal HR estimations, but also the accuracy of the estimations themselves. Accurate fetal HR estimations during the second stage of labor are very important because during this period the fetus is often subjected to reduced oxygenation. With a larger dataset, it might be worth training a separate network that will exclusively target the second stage of labor and perhaps

achieve more accurate estimations during the deceleration periods. Moreover, it is worth investigating the performance of a hybrid approach combining our method with a model-based approach, such as the method of Warmerdam et al. [127]. Such an approach could benefit from the strengths of both approaches; for instance, the method of Warmerdam has a robust mechanism for detecting unreliable fetal HR outcomes, while our method has high accuracy. In any case, we should ensure the feasibility of implementing the final system in a medical device.

The method for postprocessing the extracted fetal ECG signals, presented in Chapter 4, can significantly increase the signal quality even for low SNR and is relatively insensitive to the filter parameters. However, the filter is dependent on accurate estimation of the R-peak locations which, as already mentioned, is not always feasible. Thus, for application of the filter in clinical practice the user needs to be informed that the filter output should not be trusted in the absence of detected R-peaks. Moreover, the filter cannot be applied in cases when there is interest in abrupt or short-term changes in the fetal ECG morphology, such as in cases of arrhythmia, because of the averaging performed on the reference channels and the fact that the filter needs time to adapt its coefficients. Perhaps the averaging of the reference channels could be replaced with another enhancement method that retains the variations among different complexes. To this end, wavelet denoising could be considered, although it still shows limitations to deal with strong non-stationarity. However, the proposed filter may be suitable for applications such as detecting hypoxia were ST changes in the ECG waveform happen occur at longer time scales. Nevertheless, a large patient study should be performed to assess whether ST segment analysis can be performed on our filtered data from noninvasive ECG recordings as accurately as on invasive scalp ECG recordings. In that case, as shown before [38], the fetal orientation should be taken into account. Fetuses in different positions will have different ECG waveforms and thus the amount of ST elevation will differ depending on the position of the fetus. Therefore, normalization of the fetal orientation is crucial before performing ST analysis; this could be done with the help of ultrasound imaging as suggested in [85].

The denoising methods presented in Chapter 5 and 6 do not need any prior information on the locations of the ECG complexes and most importantly are able to preserve morphological variations among them. Although we proved that significant improvement in the quality of the extracted ECG signals can be achieved, the validation of our denoising network in real signals was suboptimal. Before

using the method in clinical practice, a thorough validation in a large real dataset is needed. For this we need noninvasive fetal ECG data together with simultaneous scalp recordings annotated by experts. In this way it will be possible to compare the ECG intervals. If the intervals are proven to have a good agreement with the scalp recordings, the method will be suited for use in clinical practice. In addition, we will need to establish the minimum signal quality leading to reliable reconstruction of the clean signals. It would also be interesting to train a similar network to perform the fetal ECG extraction directly from the abdominal signals. However, this problem is extremely complex and requires large datasets for training a deep network with sufficient capacity.

Since the denoising network retains variations among successive ECG complexes, it is worth performing a study to assess if fetal arrhythmias can be detected from the noninvasive fetal ECG signals. For such a study, additional data obtained from fetal echocardiography could be used as ground truth. Moreover, the denoised ECG data obtained from the denoising methods presented in this thesis could be used to define normal ranges for the fetal ECG in different stages of pregnancy. Having established the normal ranges, deviations from them could indicate the presence of abnormalities, such as CHD. When abnormal ECGs are recognized, referral should be given to more specialized centers, hopefully increasing the current detection rate of CHD, which is rather low (~60% [83]). Also, for this study, a large patient population is needed that will possibly include fetuses with different types of CHD.

Over the last years, noninvasive fetal electrocardiography has shown remarkable advances and its potential has been demonstrated to complement the fetal monitoring techniques that are currently used in clinical practice. Developments both in signal processing and in hardware design resulted in few commercial solutions for CTG acquisition based on noninvasive abdominal recordings. Yet, these applications are still in their infancy. We believe that the methods proposed in this thesis can help to further advance and spread this technology in clinical practice. Most importantly, we hope to achieve more reliable signal analysis and also morphological analysis of the fetal electrocardiogram, which today is only possible using the invasive scalp electrocardiography method.

This morphological ECG analysis could improve the perinatal outcome in several ways. For example, it could be implemented in a portable CTG device with several additional functionalities. First, an automated algorithm could indicate the presence

of fetal arrhythmias. This could serve like a screening method referring patient suspected of having arrhythmias to a perinatal cardiologist for further diagnosis. Severe cases of arrhythmia can lead to fetal death but if detected during pregnancy can be monitored closely and in some cases treated in utero. A second algorithm could be incorporated in the device that detects presence of CHD. This could also serve as a screening method and referral to a specialized center should be then given, leading to an increase in the detection rate of CHD. Prenatal CHD diagnosis significantly increases survival rates and decreases long-term morbidity. Moreover, the device could measure beat-to-beat HR variability and/or perform ST segment analysis in order to detect fetal hypoxia, helping obstetricians eventually to perform proper interventions.

# *Bibliography*

[1] W. H. Organization, The WHO application of ICD-10 to deaths during the perinatal period: ICD-PM, 2016.

[2] J. E. Lawn, A. Manandhar, R. A. Haws RA and G. L. Darmstadt, "Reducing one million child deaths from birth asphyxia--a survey of health systems gaps and priorities," Health Res Policy Syst, vol. 5, no. 4, 2007.

[3] H. M. Aslam, S. Saleem, R. Afzal et al., "Risk factors of birth asphyxia," Ital. J. Pediatr, vol. 40, no. 94, 2014.

[4] R. Sameni and G. D. Clifford, "A Review of Fetal ECG Signal Processing; Issues and Promising Directions," Open Pacing Electrophysiol Ther J., p. 4–20, 2010.

[5] D. Vervoort and M. Cardarelli, "The global unmet need of congenital cardiac care: a quantitative analysis of the global burden of disease," Cardiology in the Young, pp. 1-6, 2020.

[6] R. Vullings, "Fetal Electrocardiography and Deep Learning for Prenatal Detection of Congenital Heart Disease," in 2019, Singapore, Computing in Cardiology.

[7] R. K. Freeman, T. J. Garite, M. P. Nageotte and L. A. Miller, Fetal heart rate monitoring, Lippincott Williams & Wilkins, 2012.

[8] Z. Alfirevic, G. M. L. Gyte, A. Cuthbert and D. Devane, "Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour," Cochrane Database Syst Rev., vol. 2, no. 2, 2017.

[9] C. Pehrson, J. L. Sorensen and I. Amer-Wåhlin, "Evaluation and impact of cardiotocography training programmes: a systematic review," BJOG, vol. 118, no. 8, pp. 926-35, 2011.

[10] E. Chandraharan, "STAN: An Introduction to Its Use, Limitations and Caveats," St. George's Healthcare NHS Trust, London, 2010.

[11] D. C. Young, J. H. Gray, E. R. Luther and L. J. Peddle, "Fetal scalp blood pH sampling: Its value in an active obstetric unit," Am J Obstet Gynecol, vol. 136, no. 3, pp. 276-281, 1980.

[12] K. G. Rosén and I. Kjellmer., "Changes in the fetal heart rate and ECG during hypoxia," Acta Physiologica Scandinavica, vol. 93, no. 1, p. 59–66, 1975.

[13] V. Chudácek, J. Spilka, M. Burša et al., "Open access intrapartum CTG database," BMC Pregnancy Childbirth, vol. 14, no. 16, 2014.

[14] Society for Maternal-Fetal Medicine (SMFM), S. M. Berry, J. Stone et al., "Fetal blood sampling," Am J Obstet Gynecol, vol. 209, pp. 170-180, 2013.

[15] E. Chandraharan, "Fetal scalp blood sampling during labour: is it a useful diagnostic test or a historical test that no longer has a place in modern clinical obstetrics?," Bjog-Int J Obstet Gy, vol. 121, no. 9, pp. 1056-1062, 2014.

[16] I. Amer-Wåhlin, C. Hellsten, H. Norén, H. Hagberg, A. Herbst and I. Kjellmer, "Cardiotocography only versus cardiotocography plus ST analysis of fetal

electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial," Lancet, vol. 358, no. 9281, pp. 534-8, Aug. 2001.

[17] J. Westgate, M. Harris, J. S. Curnow and K. R. Greene, "Plymouth randomized trial of cardiotocogram only versus ST waveform plus cardiotocogram for intrapartum monitoring in 2400 cases," Am. J. Obstet. Gynecol., vol. 169, no. 5, pp. 1151-1160, 1993.

[18] K. Ojala , M. Vaarasmaki, K. Makikallio, M. Valkama and A. Tekay, "A comparison of intrapartum automated fetal electrocardiography and conventional cardiotocography—a randomised controlled study," BJOG, vol. 113, p. 419–423, 2006.

[19] M. E. Westerhuis, G. H. Visser, K. G. Moons KG and N. Zuithoff, "Cardiotocography plus ST analysis of fetal electrocardiogram compared with cardiotocography only for intrapartum monitoring: a randomized controlled trial," Obstet Gynecol., vol. 117, no. 2, pp. 406-407, 2011.

[20] C. Vayssiere C, E. David, N. Meyer, R. Haberstich, V. Sebahoun, E. Roth, R. Favre, I. Nisand and B. Langer, "A French randomized controlled trial of ST-segment analysis in a population with abnormal cardiotocograms during labor," Am J Obstet Gynecol, vol. 197, no. 3, p. 299.e1–299.e6, 2007.

[21] E. Schuit, I. Amer-Wahlin , K. Ojala, C. Vayssiere et al., "Effectiveness of electronic fetal monitoring with additional ST analysis in vertex singleton pregnancies at >36 weeks of gestation: an individual participant data metaanalysis," Am J Obstet Gynecol, vol. 208, no. 3, pp. 187.e1-187.e13, 2013.

[22] I. Amer-Wåhlin, S. Arulkumaran, H. Hagberg, K. Maršál and G. H. A. Visser, "Fetal electrocardiogram: ST waveform analysis in intrapartum surveillance," BJOG, vol. 114, no. 10, p. 1191–1193, 2007.

[23] R. Vullings, Non-invasive fetal electrocardiogram : analysis and interpretation, Eindhoven: Technische Universiteit Eindhoven, 2010.

[24] A. D. J. Hulsenboom, G. J. J. Warmerdam, J. Weijers et al., "Head orientation and electrode placement potentially influence fetal scalp ECG waveform," PLoS ONE, vol. 14, no. 10, p. e0223282, 2019.

[25] A. D. J. Hulsenboom, K. M. Verdurmen, R. Vullings, M. B. van der Hout-van der Jagt, A. Kwee and J. O. E. H. van Laar, "Relative versus absolute rises in T/QRS ratio by ST analysis of fetal electrocardiograms in labour: A case-control pilot study," PLoS ONE, vol. 14, no. 3, 2019.

[26] G. J. J. Warmerdam, "Analysis of fetal heart rate variability from non-invasive electrocardiography recordings," Technische Universiteit Eindhoven, 2018.

[27] M. Peters, J. Crowe, J. Piéri, H. Quartero, B. Hayes-Gill, D. James, J. Stinstra and S. Shakespeare, "Monitoring the fetal heart non-invasively: a review of methods," J. Perinat. Med, vol. 29, pp. 408-416, 2001.

[28] G. W. Lawson, R. Belcher, G. S. Dawes and C. W. Redman, "A comparison of ultrasound (with autocorrelation) and direct electrocardiogram fetal heart rate detector systems," Am J Obstet Gynecol, vol. 147, no. 6, pp. 721-2, 1983.

[29] J. O. E. H. van Laar, C. H. L. Peters, R. Vullings, S. Houterman, J. W. M. Bergmans and S. G. Oei, "Fetal autonomic response to severe acidaemia during labor," BJOG, vol. 117, no. 4, p. 429–437, 2010.

[30] J. O. E. H. van Laar, C. H. L. Peters, S. Houterman, P. F. F. Wijn, A. Kwee and S. G. Oei, "Normalized spectral power of fetal heart rate variability is associated with fetal scalp blood pH," Early Hum. Dev., vol. 87, no. 4, pp. 259-263, 2011.

[31] G. J. J. Warmerdam, R. Vullings, J. O. E. H. van Laar, M. B. van der Hout-van der Jagt, J. W. M. Bergmans, L. Schmitt and S. G. Oei, "Detection rate of fetal distress using contraction-dependent fetal heart rate variability analysis," Phys Meas, vol. 39, no. 2, p. 025008, 2018.

[32] A. Georgieva, S. J. Payne, M. Moulden and C. W. G. Redman , "Artificial neural networks applied to fetal monitoring in labour," Cont. Dev. of Neural Compt. & Appln., vol. 22, p. pages85–93, 2013.

[33] G. Magenes, M. G. Signorini and D. Arduini, "Classification of cardiotocographic records by neural networks," in Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, Como, Italy, 2000.

[34] P. Hamelmann, R. Vullings, A. F. Kolen, J. W. M. Bergmans, J. O. E. H. van Laar, P. Tortoli and M. Mischi, "Doppler ultrasound technology for fetal heart rate monitoring: a review,," IEEE T Ultrason Ferr, vol. 67, no. 2, pp. 226-238, 2020.

[35] H. Kieler, S. Cnattingius, B. Haglund, J. Palmgren and O. Axelsson, "Ultrasound and adverse effects," Ultrasound Obstet Gynecol, vol. 20, no. 1, pp. 102-103, 2002.

[36] Greene K. R. et al., "Changes in the ST waveform of the fetal lamb electrocardiogram with hypoxemia," Am. J. Obstet. Gynecol., vol. 144, pp. 950-8, 1982.

[37] K. G. Rosén, K. H. Hökegård and I. Kjellmer, "A Study of the Relationship between the Electrocardiogram and Hemodynamics in the Fetal Lamb during Asphyxia," Acta Physiologica Scandinavica, vol. 98, no. 3, pp. 275-284, 1976.

[38] R. Vullings, K. M. J. Verdurmen, A. D. J. Hulsenboom et al., "The electrical heart axis and ST events in fetal monitoring: A post-hoc analysis following a multicentre randomised controlled trial," PLoS One, vol. 12, no. 4, p. e0175823, 2017.

[39] P. Andriessen, A. Zwanenburg, J. O. E. H. van Laar et al., "ST waveform analysis for monitoring hypoxic distress in fetal sheep after prolonged umbilical cord occlusion," PLoS One, vol. 13, p. e0195978, 2018.

[40] L. A. Hamilton, E. Fisher, C. Horn, I. DuBrow and D. Vidyasagar, "A new prenatal cardiac diagnostic device for congenital heart disease," Obstet. Gynecol., vol. 50, no. 4, p. 491–4, 1977.

[41] C. Velayo et al., "Understanding congenital heart defects through abdominal fetal electrocardiography: Case reports and clinical implications," J. Obstet. Gynaecol. Res., vol. 37, no. 5, p. 428–435, 2011.

[42] C. Lempersz, L. Noben, B. de Vries et al., “The Noninvasive Fetal Electrocardiogram During Labor: A Review of the Literature,” Obstet Gynecol Surv, vol. 75, no. 6, pp. 369-380, 2020.

[43] R. Kahankova et al., “A Review of Signal Processing Techniques for Non-Invasive Fetal Electrocardiography,” IEEE Rev Biomed Eng, vol. 13, pp. 51-73, 2020.

[44] J. Gordon Betts, Anatomy & physiology, Houston, Texas: OpenStax College, Rice University, 2013, p. 787–846.

[45] A. Moorman, S. Webb, N. A. Brown, W. Lamers and R. H. Anderson, “Development of the heart: (1) formation of the cardiac chambers and arterial trunks,” Heart, vol. 89, no. 7, p. 806–14, 2003.

[46] “OpenStax College / CC BY (https://creativecommons.org/licenses/by/3.0)”.

[47] J. Stinstra, Reliability of the fetal magnetocardiogram, University of Twente, Enschede, The Netherlands: Ph.D. dissertation, 2001.

[48] G. Mielke and N. Benda, “Cardiac output and central distribution of blood flow in the human fetus,” Circulation, vol. 103, no. 12, p. 1662–1668, 2001.

[49] N. P. DePasquale and G. E. Burch, “The electrocardiogram, ventricular gradient and spatial vectorcardiogram during the first week of life,” Am J Cardiol, vol. 12, p. 482–93, 1963.

[50] S. Yagel, N. H. Silverman and U. Gemburch, Fetal cardiology: embryology, genetics, physiology, echocardiographic evaluation, diagnosis and perinatal management of cardiac diseases. 2nd ed, New York, USA: Informa Healthcare Inc, 2009.

[51] A. Guyton and J. Hall, Textbook of medical physiology, 11th, Elsevier Inc, 2006.

[52] T. G. Laske, M. Shrivastav and P. A. Iaizzo, Handbook of Cardiac Anatomy, Physiology, and Devices: The cardiac conduction system, Humana Press, 2005.

[53] W. Einthoven, “Über die Form des menschlichen Electrocardiogramms,” Pflugers Arch Gesamte Physiol Menschen Tiere, vol. 60, no. 3, pp. 101-123, 1895.

[54] Z. Ihara, A. van Oosterom and R. Hoekema, “Atrial repolarization as observable during the pq interval,” J Electrocardiol, vol. 39, p. 290–297, 2006.

[55] E. Frank, “General theory of heat-vector projection,” Circulation Research, vol. 2, no. 3, p. 258–270, 1954.

[56] H. C. Burger and J. B. Van Milaan, “Heart-vector and leads,” Br Heart J, vol. 8, no. 3, p. 157–161, 1946.

[57] D. B. Geselowitz, “On the theory of the electrocardiogram,” Proc of IEEE, vol. 77, no. 6, p. 857–876, 1989.

[58] R. Sameni, Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings, Ph.D. thesis, Sharif University of Technology – Institut National Polytechnique de Grenoble, 2008.

[59] S. P. von Steinburg, A.-L. Boulesteix, C. Lederer et al., “What is the `normal’ fetal heart rate,” PeerJ, vol. 1, no. 12, p. e82, 2013.

[60] S. P. von Steinburg, A. Boulesteix, C. Lederer, S. Grunow, S. Schiermeier, W. Hatzmann, K. M. Schneider and M. Daumer, "What is the `normal' fetal heart rate?," PeerJ, vol. 1, no. e82, 2013.

[61] K. M. J. Verdurmen, A. D. J. Hulsenboom, J. O. E. H. van Laar, P. F. F. Wijn, R. Vullings and S. G. Oei, "Orientation of the electrical heart axis in mid-term pregnancy," Eur J Obstet Gynecol Reprod Biol, vol. 207, pp. 243-246, 2016.

[62] A. C. Caliskan , "Fetal Echocardiography," in Echocardiography - New Techniques, 2012.

[63] L. Yeo and R. Romero, "Intelligent navigation to improve obstetrical sonography," Ultrasound Obstet Gynecol, vol. 47, p. 403–409, 2016.

[64] R. Chaoui, R. Bollmann and H. Hoffmann, "Fetal echocardiography. I: Methods, limitations and indications," Zentralblatt fur Gynakologie, vol. 112, no. 19, pp. 1197-1208, 1990.

[65] R. M. Grivell, Z. Alfirevic, G. M. L. Gyte and D. Devan, "Antenatal Cardiotocography for Fetal Assessment," Cochrane Database Syst Rev, vol. 9, 2015.

[66] H. Tang, T. Li, T. Qiu and Y. Park, "Fetal heart rate monitoring from Phonocardiograph Signal Using Repetition Frequency of Heart Sounds," J. Electr. and Comput. Eng., 2016.

[67] F. Kovács, C. Horváth, Á. T. Balogh and G. Hosszú, "Fetal phonocardiography—past and future possibilities," Comput Methods Programs Biomed, vol. 104, no. 1, p. 19–25, 2011.

[68] M. Cesarelli, M. Ruffo, M. Romano and P. Bifulco, "Simulation of foetal phonocardiographic recordings fortesting of FHR extraction algorithms," Comput Methods Programs Biomed, vol. 107, no. 3, pp. 513-523, 2012.

[69] J. F. Strasburger, B. Cheulkar and R. T. Wakai, "Magnetocardiography for Fetal Arrhythmias," Heart Rhythm, vol. 5, no. 7, p. 1073–1076, 2008.

[70] J. Hornbuckle, A. Vail, K. R. Abrams and J. G. Thornton, "Bayesian interpretation of trials: the example of intrapartum electronic fetal heart rate monitoring," Br J Obstet Gynaecol, vol. 107, pp. 3-10, 2000.

[71] R. J. Knupp, W. W. Andrews and A. T. Tita, "The future of electronic fetal monitoring," Best Pract Res Clin Obstet Gynaecol, vol. 67, pp. 44-52, 2020.

[72] S. L. Bloom et al., "Fetal Pulse Oximetry and Cesarean Delivery," N Engl J Med, p. 2195–2202, 2006.

[73] C. E. East, L. Begg, P. B. Colditz and R. Lau, "Fetal pulse oximetry for fetal assessment in labour," Cochrane Database Syst. Rev., vol. 10, 2014.

[74] E. H. Hon and S. T. Lee, "Electronic evaluations of the fetal heart rate patterns preceding fetal death, further observations," Obstet Gyne, vol. 87, p. 814–26, 1965.

[75] M. M. Wolf, G. A. Varigos, D. Hunt and J. G. Sloman, "Sinus arrhythmia in acute myocardial infarction."," The Medical Journal of Australia 2.2, p. 52–53, 1978.

[76] A. Stys and T. Stys, "Current Clinical Applications of Heart Rate Variability," Clin Cardiol, vol. 21, no. 10, pp. 719-724, 1998.

[77] D. Hoyer, E.-M. Kowalski, A. Schmidt et al., “Fetal autonomic brain age scores, segmented heart rate variability analysis, and traditional short term variability,” Front Hum Neurosci, vol. 8, 2014.

[78] J. A. DiPietro, M. H. Bornstein, C.-S. Hahn, K. Costigan and A. Achy-Brou, “Fetal Heart Rate and Variability: Stability and Prediction to Developmental Outcomes in Early Childhood,” Child Dev., vol. 78, no. 6, p. 1788–1798, 2007.

[79] C. A. Lear, J. A. Westgate, A. Ugwumadu et al., “Understanding Fetal Heart Rate Patterns That May Predict Antenatal and Intrapartum Neural Injury,” Seminars in Pediatric Neurology, vol. 28, pp. 3-16, 2018.

[80] S. C. Mitchell, S. B. Korones and H. W. Berendes, “Congenital heart disease in 56,109 births. Incidence and natural history.,” Circulation, vol. 43, no. 3, pp. 323-32, 1971.

[81] O. Franklin, “Prenatal diagnosis of coarctation of the aorta improves survival and reduces morbidity,” Heart, vol. 87, no. 1, p. 67–69, 2002.

[82] C. L. van Velzen et al., “Prenatal detection of transposition of the great arteries reduces mortality and morbidity,” Ultrasound Obstet. Gynecol, vol. 45, no. 3, p. 320–5, 2015.

[83] C. L. Van Velzen et al., “Prenatal detection of congenital heart disease - Results of a national screening programme,” BJOG An Int. J. Obstet. Gynaecol, vol. 123, no. 3, p. 400–407, 2016.

[84] I. Stümpflen , A. Stümpflen, M. Wimmer and G. Bernaschek, “Effect of detailed fetal echocardiography as part of routine prenatal ultrasonographic screening on detection of congenital heart disease,” Lancet, vol. 348, no. 9031, p. 854–857, 1996.

[85] C. Lempersz et al., “The standardized 12-lead fetal electrocardiogram of the healthy fetus in mid-pregnancy: A cross-sectional study,” PLOS ONE, vol. 15, pp. 1-12, 2020.

[86] I. Amer-Wahlin, B. Yli and S. Arulkumaran, “Fetal heart rate patterns and ECG ST segment changes preceding metabolic acidaemia at birth,” BJOG, vol. 112, no. 2, pp. 160-165, 2005.

[87] I. Amer-Wåhlin, “Fetal ECG waveform analysis for intrapartum monitoring,” Phd thesis, Lund University, Lund, Sweden, 2003.

[88] J. J. Candil and C. M. Luengo, “QT Interval and Acute Myocardial Ischemia: Past Promises, New Evidences,” vol. 61, no. 6, pp. 561-563, 2008.

[89] M. A. Oudijk, A. Kwee, G. H. Visser, S. Blad, E. J. Meijboom and K. G. Rosen, “The effects of intrapartum hypoxia on the fetal QT interval,” BJOG., vol. 111, no. 7, p. 656–60, 2004.

[90] M. J. Rooijakkers, H. de Lau, C. Rabotti, S. G. Oei and M. Mischi, “Fetal movement detection based on QRS amplitude variations in abdominal ECG recordings,” in Annu Int Conf IEEE Eng Med Biol Soc, 2014.

[91] E. Sadovsky and W. Z. Polishuk, “Fetal movements in utero nature, assessment, prognostic value, timing of delivery,” Obstet Ginecol, vol. 50, p. 49–55, 1977.

[92] M. J. Rooijakkers, C. Rabotti, H. de Lau, S. G. Oei, J. W. Bergmans and M. Mischi, "Feasibility study of a new method for low-complexity fetal movement detection from abdominal ECG recordings," IEEE J Biomed Health Inform, vol. 20, no. 5, pp. 1361-1368, 2015.

[93] S. Abboud and D. Sadeh, "Spectral analysis of the fetal electrocardiogram," Comput. Biol. Med., vol. 19, no. 6, pp. 409-415, 1989.

[94] D. D. Țarălungă, G. M. Ungureanu, I. Gussi, R. Strungaru and W. Wolf, "Fetal ECG Extraction from Abdominal Signals: A Review on Suppression of Fundamental Power Line Interference Component and Its Harmonics," Comput Math Methods Med, vol. 2014, 2014.

[95] J. Cardoso, G. Lenis, N. Pilia, A. Loewe, W. H. W. Schulze and O. Dössel, "Comparison of Baseline Wander Removal Techniques considering the Preservation of ST Changes in the Ischemic ECG: A Simulation Study," Comput Math Method M, vol. 2017, 2017.

[96] S. M. Martens, C. Rabotti, M. Mischi and R. J. Sluijter, "A robust fetal ECG detection method for abdominal recordings," Phys Meas, vol. 28, no. 4, pp. 373-388, 2007.

[97] S. Cerutti , G. Baselli , S. Civardi, E. Ferrazzi, A. M. Marconi, M. Pagani and G. Pardi, "Variability analysis of fetal heart rate signals as obtained from abdominal electrocardiographic recordings," J Perinat Med., vol. 14, no. 6, p. 445–52, 1986.

[98] R. Vullings, C. H. L. Peters, R. J. Sluijter, M. Mischi, S. G. Oei and J. W. M. Bergmans, "Dynamic segmentation and linear prediction for maternal ECG removal in antenatal abdominal recordings," Physiol. Meas., vol. 30, no. 3, pp. 291-307, 2009.

[99] M. Ungureanu, J. W. M. Bergmans, S. G. Oei and R. Strungaru, "Fetal ECG extraction during labor using an adaptive maternal beat subtraction technique," Biomed. Tech. (Berl), vol. 52, no. 1, p. 56–60, 2007.

[100] M. Niknazar, B. Rivet and C. Jutten, "Fetal ECG extraction by extended state kalman filtering based on single-channel recordings," IEEE Trans. Biomed. Eng., vol. 60, no. 5, pp. 1345-52, 2013.

[101] Vullings R et al., "An Adaptive Kalman Filter for ECG Signal Enhancement," IEEE Trans. Biomed. Eng., vol. 58, no. 4, pp. 1094-103, 2011.

[102] S. Wu, Y. Shen et al., "Research of fetal ECG extraction using wavelet analysis and adaptive filtering," Comput. Biol. Med., vol. 43, no. 10, pp. 1622-7, 2013.

[103] V. Vigneron, A. Paraschiv-Ionescu et al., "Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising," in Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings., Paris, France, 2003.

[104] M. Varanini , G. Tartarisco, L. Billeci, A. Macerata, G. Pioggia and R. Balocchi, "A multi-step approach for non-invasive fetal ECG analysis," in Computing in Cardiology 2013, Zaragosa, 2013.

[105] J. L. Camargo-Olivares, R. Martín-Clemente, S. Hornillo-Mellado, M. M. Elena and I. Roman, "The maternal abdominal ECG as input to MICA in the fetal ECG extraction problem," IEEE Signal Process. Lett., vol. 18, no. 3, pp. 161-4, 2011.

[106] R. Martín-Clemente, J. L. Camargo-Olivares, S. Hornillo-Mellado, M. Elena and I. Roman, "Fast technique for noninvasive fetal ECG extraction," IEEE Trans. Biomed. Eng., vol. 58, no. 2, pp. 227-30, 2002.

[107] Y. Ye, P. C. Y. Sheu, J. Zeng, G. Wang and K. Lu, "An efficient semi-blind source extraction algorithm and its applications to biomedical signal extraction," Science in China Series F-Information Sciences, vol. 52, no. 10, pp. 1863-1874, 2009.

[108] D. Adam and D. Shavit, "Complete foetal ECG morphology recording by synchronized adaptive filtration," Med. Biol. Eng. Comput., vol. 28, pp. 287-92, 1990.

[109] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong and R. C. Goodlin, "Adaptive noise cancelling: principles and applications," Proc. IEEE, vol. 63, no. 12, pp. 1692-716, 1975.

[110] Camps G. et al., "Fetal ECG extraction using an FIR neural network," Comput. Cardiol., vol. 28, pp. 249-52, 2001.

[111] R. Jaros, R. Martinek and R. Kahankova , "Non-Adaptive Methods for Fetal ECG Signal Processing: A Review and Appraisal," Sensors, vol. 18, no. 11, 2018.

[112] E. H. Hon, "Noise Reduction In Fetal Electrocardiography," Med. Arts Sci., vol. 18, p. 63–66, 1964.

[113] Lindecrantz K., "Processing of the fetal ECG; an implementation of a dedicated real time microprocessor system," Chalmers University of Technology, Gothenburg, 1983.

[114] M. Ahmadi, M. Ayat, K. Assaleh and H. Al-Nashash, "Fetal ECG Signal Enhancement using Polynomial Classifiers and Wavelet Denoising," in 2008 Cairo International Biomedical Engineering Conference, Cairo, 2008.

[115] G. Baldazzi, E. Sulas, E. Brungiu, M. Urru, R. Tumbarello, L. Raffo and D. Pani, "Wavelet-based Post-processing Methods for the Enhancement of Non-invasive Fetal ECG," in 2019 Computing in Cardiology , Singapore, 2019.

[116] T. F. Oostendorp, A. van Oosterom and H. W. Jongsma, "The effect of changes in the conductive medium on the fetal ECG throughout gestation," Clin Phys Physiol Meas, vol. 10, no. 4B, p. 11–20, 1989.

[117] T. F. Oostendorp and A. van Oosterom, "Electrical properties of tissues involved in the conduction of foetal ECG," Med Biol Eng Comput, vol. 27, p. 322–4, 1989.

[118] A. K. Sundström et al, "Fetal Surveillance," Neoventa Medical AB, Gothenburg, Sweden, 2000.

[119] C. J. Shaw, C. C. Lees and D. A. Giussani, "Variations on fetal heart rate variability," J. Physiol., p. 1279–1280, Mar. 2016.

[120] J. Jezewski, J. Wrobel and K. Horoba, "Comparison of doppler ultrasound and direct electrocardiography acquisition techniques for quantification of fetal heart rate variability," IEEE Trans. Biomed. Eng., vol. 53, no. 5, pp. 855-864, 2006.

[121] R. Sameni and G. D. Clifford, "A Review of Fetal ECG Signal Processing; Issues and Promising Directions," Open Pacing Electrophysiol Ther J., vol. 3, pp. 4-20, Jan. 2010.

[122] I. Silva, J. Behar, R. Sameni, T. Zhu, J. Oster, G. D. Clifford and G. B. Moody, "Noninvasive Fetal ECG: the PhysioNet/Computing in Cardiology Challenge 2013," Comput Cardiol, vol. 40, pp. 149-152, 2013.

[123] G. D. Clifford, I. Silva, J. Behar and G. B. Moody, "Non-invasive fetal ECG analysis," Physiological Measurement, vol. 35, pp. 1521-36, 2014.

[124] P. P. Kanjilal, S. Palit and G. Saha, "Fetal ECG extraction from single-channel maternal ECG using singular value decomposition," IEEE Trans. Biomed. Eng., vol. 44, no. 1, pp. 51-9, 1997.

[125] J. Behar et al., "Combining and benchmarking methods of foetal ECG extraction," Physiol. Meas, vol. 35, no. 8, p. 1569–1589, 2014.

[126] J. Behar, F. Andreotti, S. Zaunseder, J. Oster and G. D. Clifford, "A practical guide to non-invasive foetal electrocardiogram extraction and analysis," Physiol Meas, vol. 37, no. 5, pp. R1-R35, 2016.

[127] G. J. J. Warmerdam, R. Vullings, L. Schmitt, J. O. E. H. Van Laar and J. W. M. Bergmans , "Hierarchical Probabilistic Framework for Fetal R-Peak Detection, Using ECG Waveform and Heart Rate Information," IEEE Trans. Signal Process., vol. 66, no. 16, pp. 4388 - 4397, Aug. 2018.

[128] F. Jamshidian-Tehrani and R. Sameni, "Fetal ECG extraction from time-varying and low-rank noninvasive maternal abdominal recordings," Phys Meas, vol. 39, no. 12, p. 125008, 2018.

[129] N. Zhang, J. Zhang, H. Li, O. Mumini, O. Samuel, K. Ivanov and L. Wang, "A Novel Technique for Fetal ECG Extraction Using Single-Channel Abdominal Recording," Sensors, vol. 17, no. 3, p. 457, 2017.

[130] R. Waseem and W. Zenghui, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," Neural Comput., vol. 29, no. 9, pp. 1-98, 2017.

[131] A. B. Nassif et al., "Speech Recognition Using Deep Neural Networks: A Systematic Review," IEEE Access, vol. 7, pp. 19143-19165, 2019.

[132] A. Esteva et al., "A guide to deep learning in healthcare," Nat. Med., vol. 25, p. 24–29, 2019.

[133] K. Antczak, "Deep Recurrent Neural Networks for ECG Signal Denoising," arXiv:1807.11551, 2018.

[134] P. Xiong, H. Wang, M. Liu, S. Zhou, Z. Hou and X. Liu, "ECG signal enhancement based on improved denoising auto-encoder," Eng. Appl. Artif. Intell., vol. 52, pp. 194-202, 2016.

[135] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," Procedia Comput. Sci., vol. 120, pp. 268-275, 2017.

[136] Z. Zhao et al., "DeepFHR: intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network," BMC Med Inform Decis Mak, vol. 19, p. 286, 2019.

[137] E. Fotiadou, T. Konopczyński, J. Hesser and R. Vullings, "End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising," Physiol. Meas., vol. 41, no. 1, p. 015005, 2020.

[138] E. Fotiadou and R. Vullings, "Multi-Channel Fetal ECG Denoising With Deep Convolutional Neural Networks," Front Pediatr, vol. 8, p. 508, 2020.

[139] P. R. Muduli, R. R. Gunukula and A. Mukherjee, "A deep learning approach to fetal-ECG signal reconstruction," in 2016 Twenty Second National Conference on Communication (NCC), Guwahati, 2016.

[140] W. Zhong , L. Liao , X. Guo and G. Wang, "A deep learning approach for fetal QRS complex detection," Physiol. Meas, vol. 39, no. 4, p. 045004, 2018.

[141] J. S. Lee, M. Seo, S. W. Kim and M. Choi, "Fetal QRS Detection Based on Convolutional Neural Networks in Noninvasive Fetal Electrocardiogram," in 4th Int. Conf.on Frontiers of Signal Processing, Poitiers, 2018.

[142] E. Fotiadou, M. Xu et al., "Deep Convolutional Long Short-Term Memory Network for Fetal Heart Rate Extraction," in 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020.

[143] W. Shi et al., "Single image super-resolution with dilated convolution based multi-scale information learning inception module," in IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017.

[144] C. Lempersz, L. Noben, G. van Osta et al., "Intrapartum non-invasive electrophysiological monitoring: A prospective observational study," Acta Obstet. Gynecol. Scand., vol. 99, no. 10, pp. 1387-1395, 2020.

[145] G. J. J. Warmerdam, R. Vullings, L. Schmitt, J. O. E. H. V. Laar and J. W. M. Bergmans, "A fixed-lag Kalman smoother to filter power line interference in electrocardiogram recordings," IEEE Trans. Biomed. Eng., vol. 64, no. 8, pp. 1852-61, 2016.

[146] J. Behar, J. Oster and G. D. Clifford, "Non-invasive FECG extraction from a set of abdominal sensors," in Computing in Cardiology 2013, Zaragosa, 2013.

[147] M. Holschneider et al., "A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform," in Wavelets, Berlin, Heidelberg, Springer, 1990, pp. 286-297.

[148] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in ICLR, 2016.

[149] C. Szegedy et al., "Going Deeper with Convolutions," arXiv: 1409.4842, Sep. 2014.

[150] K. He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in IEEE International Conference on Computer Vision (ICCV), Santiago, 2015.

[151] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Physica D, vol. 404, p. 132306, 2020.

[152] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Int. Conf. for Learning Representations, San Diego, 2015.

[153] W. R. Cohen et al, “Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes: Maternal surface electrode fetal monitoring,” Acta Obstet Gynecol Scand., vol. 91, no. 11, p. 1306–13, 2012.

[154] T. Y. Euliano, S. Darmanjian, M. T. Nguyen, J. D. Busowski, N. Euliano and A. R. Gregg, “Monitoring fetal heart rate during labor: A comparison of three methods,” J Pregnancy, 2017.

[155] A. L. Tranquilli, “Fetal heart rate in the second stage of labor: recording, reading, interpreting and acting,” J Matern Fetal Neonatal Med, vol. 25, no. 12, pp. 2551-2554, 2012.

[156] H. Goncalves , A. Costa, D. Ayres-de Campos, C. Costa-Santos, A. P. Rocha and J. Bernardes, “Comparison of real beat-to-beat signals with commercially available 4 Hz sampling on the evaluation of foetal heart rate variability,” Med Biol Eng Comput, vol. 51, p. 665–76, 2013.

[157] J. O. van Laar, G. J. Warmerdam, K. M. Verdumen, R. Vullings, C. H. Peters, S. Houterman, P. F. Wijn, P. Andriessen, C. van Pul and S. G. Oei, “Fetal heart rate variability during pregnancy, obtained from non-invasive electrocardiogram recordings,” Acta Obstet. Gynecol. Scand., vol. 93, no. 1, pp. 93-101, 2014.

[158] Rooijakkers J. M. et al., “Low-complexity R-peak detection for ambulatory fetal monitoring,” Physiol. Meas., vol. 33, no. 7, pp. 1135-50, 2012.

[159] Jia W. et al., “Fetal ECG extraction based on adaptive linear neural network,” in 2010 3rd International Conference on Biomedical Engineering and Informatics, Yantai, China, 2010.

[160] P. A. Hassanpour H, “Fetal ECG extraction using wavelet transform,” in 2006 International Conference on Computational Inteligence for Modelling Control and Automation and International Conference on Intelligent Agents Web Technologies and International Commerce (CIMCA’06), Sydney, 2006.

[161] N. S. Khamene A, “A new method for the extraction of fetal ECG from the composite abdominal signal,” IEEE Trans. Biomed. Eng., vol. 47, no. 4, pp. 507-16, 2000.

[162] H. Zhang, Z. Shi, C. Guo and E. Feng, “Semi-blind source extraction algorithm for fetal electrocardiogram based on generalized autocorrelations and reference signals,” J. Comput. Appl. Math., vol. 223, no. 1, pp. 409-20, 2009.

[163] Ferrara E. R. and Widrow B., “Fetal Electrocardiogram Enhancement by Time-Sequenced Adaptive Filtering,” IEEE Trans. Biomed. Eng., Vols. BME-29, no. 6, pp. 458-60, 1982.

[164] Gao P. et al., “Blind separation of fetal ECG from single mixture using SVD and ICA,” in Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, Singapore, 2003.

[165] F. Andreotti, M. Riedl, T. Himmelsbach et al., “Maternal signal estimation by Kalman filtering and Template Adaptation for fetal heart rate extraction,” in Computing in Cardiology 2013, Zaragosa, 2013.

[166] P. Podziemski and J. Gieraltowski, “Fetal heart rate discovery: Algorithm for detection of fetal heart rate from noisy, noninvasive fetal ECG recordings,” in Computing in Cardiology 2013, Zaragosa, 2013.

[167] Ferrara E. R. and Widrow B. “Multichannel adaptive filtering for signal enhancement,” IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 3, pp. 766 - 70, 1981.

[168] Ramli R. M. et al., “A review of adaptive line enhancers for noise cancellation,” Aust. J. Basic & Appl. Sci., vol. 6, no. 6, pp. 337-52, 2012.

[169] Ferrara E. R., “The time-sequenced adaptive filter,” Ph.D. dissertation, Stanford University, Stanford, CA, 1977.

[170] Ferrara E. R. and Widrow B., “The time-sequenced adaptive filter,” IEEE Trans. Acoust., Speech, Signal Process., vol. 29, no. 3, pp. 679 - 83, 1981.

[171] Douglas S. C. and T. H.-. Meng, “Normalized Data Nonlinearities for LMS Adaptation,” IEEE Trans. on Signal Process., vol. 42, no. 6, pp. 1352-65, 1994.

[172] Rosén K. G. and Lindecrantz K., “STAN-the Gothenburg model for fetal surveillance during labour by ST analysis of the fetal electrocardiogram,” Clin. Phys. Physiol. Meas., vol. 10, no. 4B, pp. 51-6, 1989.

[173] Cano G. G. et al., “Enhancement of low-level ECG components in noise with time-sequenced adaptive filtering,” J. Electrocardiol., vol. 23, pp. 176-83, 1990.

[174] K. M. J. Verdurmen, C. Lempersz, R. Vullings and C. Schroer, “Normal ranges for fetal electrocardiogram values for the healthy fetus of 18–24 weeks of gestation: a prospective cohort study,” BMC Pregnancy Childbirth, vol. 16, no. 1, p. 227, 2016.

[175] F. Andreotti , J. Behar, S. Zaunseder, J. Oster and G. D. Clifford, “An open-source framework for stress-testing non-invasive foetal ECG extraction algorithms,” Physiol. Meas., vol. 37, no. 5, pp. 627-48, 2016.

[176] A. L. Goldberger , L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mar, J. E. Mietus, G. B. Moody and C. -K. Peng, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” Circulation, vol. 101, no. 23, pp. e215-20, 2000.

[177] J. Jezewski , A. Matonia , T. Kupka, D. Roj and R. Czabanski, “Determination of fetal heart rate from abdominal signals: Evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram,” Biomedical Engineering, vol. 57, no. 5, pp. 383-94, 2012.

[178] Sardy S., “Minimax threshold for denoising complex signals with Waveshrink,” IEEE Trans. Signal Process., vol. 48, no. 4, pp. 1023-8, 2000.

[179] Andreotti F. et al., “Robust fetal ECG extraction and detection from abdominal leads,” Physiol Meas., vol. 35, no. 8, pp. 1551-67, 2014.

[180] Jenkins H. M. et al., “Can fetal electrocardiography improve the prediction of intrapartum fetal acidosis?,” Br. J. Obstet. Gynaecol, vol. 93, pp. 6-12, 1986.

[181] Pardi G. et al., “Fetal electrocardiogram changes in relation to fetal heart rate patterns during labor,” American Journal of Obstetrics and Gynecology, vol. 118, no. 2, pp. 243-250, 1974.

[182] Pardi G. et al., “The clinical relevance of the abdominal fetal electrocardiogram,” J. Perinat. Med., vol. 14, p. 371–377, 1986.

[183] M. G. Signorini, G. Magenes, S. Cerutti and D. Arduini, “Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings,” IEEE Trans Biomed Eng., vol. 50, no. 3, p. 365–374, 2003.

[184] E. Fotiadou, J. O. E. H. van Laar, S. G. Oei and R. Vullings, “Enhancement of low-quality fetal electrocardiogram based on time-sequenced adaptive filtering,” Med. Biol. Eng. Comput., vol. 56, no. 12, p. 2313–2323, 2018.

[185] X. Lu, Y. Tsao, S. Matsuda and C. Hor, Speech Enhancement Based on Deep Denoising Autoencoder, INTERSPEECH, 2013.

[186] A. Maas, Q. V. Le, T. M. O’Neil, O. Vinyals, P. Nguyen and A. Y. Ng, Recurrent Neural Networks for Noise Reduction in Robust ASR, INTERSPEECH, 2012.

[187] V. Jain and S. Seung, “Natural Image Denoising with Convolutional Networks,” in Proc. Advances in Neural Inf. Process. Syst., Vancouver, 2008.

[188] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, “Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising,” Trans. Image Process., vol. 26, no. 7, pp. 3142-3155, 2017.

[189] X. Mao, C. Shen and Y. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” in NIPS, Barcelona, 2016.

[190] W. Zhong , L. Liao , X. Guo and G. Wang, “A deep learning approach for fetal QRS complex,” Physiol. Meas, vol. 39, no. 4, 2018.

[191] H. C. Burger, C. J. Schuler and S. Harmeling, “Image denoising: Can plain neural networks compete with BM3D?,” in IEEE Conf. Comput. Vision and Pattern Recognition (CVPR), Providence, RI, 2012.

[192] R. K. Srivastava, K. Greff and J. Schmidhuber, “Training very deep networks,” in Adv. Neural. Inf. Process. Syst., Montreal, 2015.

[193] K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comp. Vis. Patt. Recogn., Las Vegas, 2016.

[194] R. Bousseljot , D. Kreiseler and A. Schnabel, “Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet,” Biomedizinische Technik, vol. 40, no. s1, 1995.

[195] P. Laguna, R. G. Mark, A. L. Goldberger and G. B. Moody, “A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG,” in Computers in Cardiology.

[196] J. Pan and W. J. Tompkins, “A Real-Time QRS Detection Algorithm,” IEEE Trans. Biomed. Eng., Vols. BME-32, no. 3, pp. 230 - 236, 1985.

[197] P. M. Rautaharju, B. Surawicz and L. S. Gettes, “AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram,” J. Am. Coll. Cardiol., vol. 53, no. 11, pp. 982-991, 2009.

[198] J. Behar et al., “Evaluation of the fetal QT interval using noninvasive fetal ECG technology,” Physiol. Meas., vol. 37, no. 9, p. 1392–1403, 2016.

[199] H. G. Murray, The fetal electrocardiogram: current clinical developments in Nottingham, Kooperation de Gruyter, 1986.

[200] W. J. van Wijngaarden, H. H. de Haan, J. Sahota and H. Symonds, “Changes in the PR interval - fetal heart rate relationship of the electrocardiogram during fetal compromise in chronically instrumented sheep,” Am J Obstet Gynecol., vol. 175, no. 3, Part 1, p. 548–54, 1996.

[201] J. Behar, F. Andreotti, S. Zaunseder, J. Oster and G. D. Clifford, “A practical guide to non-invasive foetal electrocardiogram extraction and analysis,” Physiol. Meas., vol. 37, no. 5, pp. R1-R35, 2016.

[202] W. Zhong, X. Guo and G. Wang, “Non-invasive Fetal Electrocardiography Denoising Using Deep Convolutional Encoder-Decoder Networks,” in Proceedings of 2019 Chinese Intelligent Systems Conference, Singapore, 2019.

[203] E. Fotiadou, T. Konopczyński, J. Hesser and R. Vullings, “Deep Convolutional Encoder-Decoder Framework for Fetal ECG Signal Denoising,” in 2019 Computing in Cardiology (CinC), Singapore, 2019.

[204] J. Behar, F. Andreotti, S. Zaunseder, Q. Li, J. Oster and G. D. Clifford, “An ECG simulator for generating maternal-foetal activity mixtures on abdominal ECG recordings,” Physiol. Meas., vol. 35, no. 8, pp. 1537-1550, 2014.

[205] P. E. McSharry, G. D. Clifford, L. Tarassenko and L. Smith, “A dynamical model for generating synthetic electrocardiogram signals,” IEEE Trans Biomed Eng, vol. 50, no. 3, pp. 289-294, 2003.

[206] R. Sameni, G. D. Clifford, C. Jutten and M. B. Shamsollahi , “Multichannel ECG and noise modeling: application to maternal and fetal ECG signals,” EURASIP J. Adv. Sig. Pr., vol. 2007, no. 1, p. 43407, 2007.

[207] I. Amer-Wåhlin, A. Ugwumadu, B. M. Yli et al., “Fetal electrocardiography ST-segment analysis for intrapartum monitoring: a critical appraisal of conflicting evidence and a way forward,” Am. J. Obstet. Gynecol., vol. 221, no. 6, pp. 577 - 601.e11, 2019.

[208] M. Cremer, “Über die direkte ableitung der aktionsströme des menschlichen herzens vom oesophagus und über das elektrokardogramm des fötus,” Münchener Medizinische Wochenschrift, vol. 53, pp. 811-813, 1906.

# Acknowledgements

As I could have not achieved any of this without the support and guidance that I received from many people, I would like to dedicate this last part of my thesis to thank them.

Firstly, I would like to express my gratitude to my first promotor prof. dr. Massimo Mischi. Thank you for giving me the opportunity to work in the BM/d group, in a nice and friendly environment. I always admire your passion and enthusiasm for research, and I feel honored to work in your group.

I would like to say thank you to my copromotors dr. Rik Vullings and dr. Judith van Laar for giving me the opportunity to participate in this research project. Rik, I am very grateful for the guidance, support and inspiration during these four years. Literally I could not have asked for a better supervisor and I owe you a lot. You have such a deep and admirable knowledge in the area of pregnancy monitoring, and I learned a lot from you. You also helped me to feel more confident and believe more in myself. I don´t know if I could finish this PhD without your encouragement, valuable insights and advice. What I will never forget is that you showed a lot of understanding and you supported me when I had some difficult moments. Judith, thank you for your clinical expertise that has been very valuable to me. I learned a lot from you in the clinical area and with your critical feedback you helped me to improve my work.

I would like to thank the committee members prof. dr. G. Oei, prof. dr. E. Cantatore, prof. T. Heskes, prof. dr. J. Hesser and prof. dr. P. M. Djurić for having accepted to be part of my committee and for their time and effort to assess this thesis.

My special thanks to prof. dr. J. Hesser that gave me the opportunity to spend some time working in Mannheim. Thank you for the inspiration, for discussing ideas with me and contributing to this work. I am very grateful to Tomasz Konopczyński for introducing me to deep learning and helping me so much to understand all the concepts. I really appreciate your help, Tomasz, and I enjoyed spending time with you.

I would like to say a big thank for the people in Nemo Healthcare. Thank you for inviting me in meetings and symposiums about fetal monitoring as I learned a lot from them. You also made me feel welcome in Nemo healthcare and you were always friendly and nice with me.

I am very grateful to prof. dr. Jan Bergmans that gave me the opportunity to pursue a PhD in the SPS group. It has really been a pleasure for me to work in the nice environment of this group. I really appreciated that Jan was always very easy to approach whenever I needed his help or advice. A big thank you goes also to the secretaries of the SPS group, Carla, Marieke and Sabine, and also Judith who were always there to help with all the administrative tasks during my PhD.

Next, I would like to thank the people in the BM/d group. Thank you, Ruud, about sharing ideas with me and contributing also to this work. Linda, I will never forget our trip to Vietnam. It was such an amazing trip with beautiful memories. Paul and Chiara thank you for our enjoyable trip to Sardinia, we had really a great time. Thank you Marina, Gabriele, Lieke, Maretha, Elisabetta, Guy, Anouk, Bas, Nish, Tom, Jelte, Laura, Beatrijs, Federica, Simona, Iris, Xufei, Anastasia, Nienke, Celine, Luuk, Rogier, Yizhou, Peiran, Chuan, Ben.

My years at TU/e have been very special also thanks to my "Greek lunch" friends. Special thanks to Chara, we didn't only drunk countless coffees together but also shared the same house during the first lockdown, traveled together, worked out together and shared many beautiful moments. Francesca, I was so happy to meet you the first day of my PhD. We immediately became friends and I enjoyed your stories, our walks, our trips, our brunches. Nailia you are the master of coffee breaks and thank you for sharing some of them with me. Thank you Panos, Roger, Arash, Cristina, Amir, Eleonora, Antoine and Christos for the nice lunches and other beautiful moments we spent together. I must admit it would be difficult to finish my PhD without the trips to the Greek islands and thank you my friends that we had such an amazing time there.

Furthermore, I would like to express my gratitude to my friends outside the university. Thank you, Louisa, Ellidora and Pelagia for being the best housemates I could have. We shared amazing moments during the last years and you are always next to me. I feel very lucky to meet you and for the trips, parties, dinners and all the funny moments we had together. Thank you Thomi, for always being close to me. Thank you, Giannis and Panos for our "secret" meetings. Thank you Panos, as without you I wouldn't be able to stay fit during the lockdown (a healthy mind in a healthy body!!!). Thank you, my new Turkish friends. I was lucky to meet you during the lockdown because you made it more enjoyable and happier. Of course, I cannot forget my best friend Kiveli. Without you, I wouldn't even be in the Netherlands

now. You were the reason that I came to Eindhoven, maybe the best decision I took in my life. I owe you the biggest thank you in the world that you are my friend, that you helped me overcome difficulties and although we are far, we are always next to each other. Also, I would like to thank Miguel. I am so happy I met you and we spent my PhD years together. You always knew how to make me release my PhD stress, believe in myself and feel more confident. Definitely, I owe part of my PhD success to you.

Lastly, I would like to thank my family for their love and their support. Even if you don't always agree with my choices, especially because they bring me far from you, you are always there to help me, willing to do everything for me. I owe you a lot because without your support, I wouldn't be here now, writing this thesis.

## Curriculum Vite

Eleni Fotiadou was born in Drama, Greece in 1984. In 2009, she received her M.Sc. degree in Electrical and Computer Engineering from Aristotle University of Thessaloniki, Greece. After her graduation she worked for 2 years in CERTH-ITI, Thessaloniki, Greece as a research assistant in computer vision. Starting in 2011, she conducted a PDEng traineeship at the Eindhoven University of Technology (TU/e) within the Video Coding and Architectures (VCA) group of the Signal Processing Systems (SPS) group. After that she worked as a research scientist in the department of Smart Personal Spaces at Philips Research in Eindhoven. In February 2017, she started a PhD project on noninvasive fetal electrocardiography analysis in the SPS group (BM/d Team) at TU/e of which the results are presented in this thesis. The project was a collaboration project between TU/e, Máxima Medical Center Veldhoven and Nemo Healthcare.