Conceiving, compiling, publishing and exploiting the "Icelandic 16-electrode EHG database"

Ásgeir Alexandersson

Thesis of 60 ECTS credits
Master of Science in Biomedical Engineering

January 2015

Tilurð, útgáfa og hagnýting íslenska legrafritsgagnabankans

Ásgeir Alexandersson

60 ECTS eininga ritgerð til
meistaraprófs (MSc) í heilbrigðisverkfræði

Janúar 2015

Conceiving, compiling, publishing and exploiting the "Icelandic 16-electrode EHG database"

Ásgeir Alexandersson

Thesis of 60 ECTS credits submitted to the School of Science and Engineering
at Reykjavík University in partial fulfillment
of the requirements for the degree of
Master of Science in Biomedical Engineering

January 2015

Supervisor:

Brynjar Karlsson,
Professor, Reykjavík University, Iceland

Examiner:

Magnús Örn Úlfarsson
Professor, University of Iceland

Tilurð, útgáfa og hagnýting íslenska legrafritsgagnabankans

Ásgeir Alexandersson

60 ECTS eininga ritgerð lögð fram við tækni- og verkfræðideild
Háskólans í Reykjavík til
meistaraprófs (MSc) í heilbrigðisverkfræði

Janúar 2015

Leiðbeinandi:

Brynjar Karlsson,
Prófessor, Háskólanum í Reykjavík

Prófdómari:

Magnús Örn Úlfarsson
Prófessor, Háskóla Íslands

Copyright

Abstract

Biomedical engineering is the application of engineering principles to biology and medicine, with the goal of improving the quality and effectiveness of patient care. Better understanding of the mechanisms underlying labor can contribute to preventing preterm birth, which is the main cause of mortality and morbidity in newborns. Studies on the electrohysterogram (EHG) have shown promising results regarding preterm labour prediction as well as other uses in obstetric care. A 16 electrode recording system in a 4-by-4 configuration on the abdomen can provide information on the propagation of uterine electrical activity. One of the main problems for researchers studying new technology is a shortage of data to enable analysis and hypothesis testing. In this respect, public access databanks are becoming increasingly important. A new category of scientific article, the Data Descriptor, has been designed to provide detailed descriptions of databases to maximize interpretation, search and reuse of the data. The main aim of this MSc project was to implement public access to 16-electrode EHG data recorded in Iceland. The second aim was to demonstrate the type of research that can be done using the database by studying the relationship between maternal clinical factors and the frequency components of contractions. The thesis describes the process of collecting, converting, annotating and de-identifying the 122 EHG recordings performed by the author between 2008 and 2010. The now publicly available Icelandic 16-electrode EHG database, found at PhysioNet, is then presented along with its Data Descriptor. The study on clinical factors and EHG data from the database suggests that a pregnant woman's body mass index, age and obstetric history may influence frequency components of contractions. This could have implications for the development of a medical device for predicting preterm labour and at the same time demonstrates the use that can be made of the database.

Keywords: *electrohysterogram, preterm birth, database, public access, frequency analysis*

Útdráttur

Í heilbrigðisverkfræði er verkfræðilegum aðferðum í læknisfræði beitt til að þróa ný tæki og aðferðir við greiningu og meðferð sjúkdóma. Fyrirburafæðingar eru aðal orsök dauða og veikinda nýbura en með því að auka skilninginn á mekaníski virkni legsins myndast möguleikar á að spá fyrir um og koma í veg fyrir fæðingu fyrir tímann. Rannsóknir á legrafritum benda til að þá tækni mætti nota við greiningu á fyrirburafæðingum. Með því að nota 16 rafskaut í 4-sinnum-4 uppsetningu er hægt að greina útbreiðslu á rafvirkni legsins. Fyrir þá vísindamenn sem eru að rannsaka nýja tækni er algengt vandamál að það vanti gögn til að vinna með. Vegna þessa hafa opnir gagnabankar orðið mikilvægir í vísindaþróun. Ný gerð vísindagreina (*e*. Data Descriptor) hefur verið búin til sem lýsir gagnabönkum vel og eykur möguleika á túlkun, leit og nýtingu gagnanna. Megin tilgangur þessa meistaraverkefnis var að veita opinn aðgang að 16-rafskauta legrafritsmælingum frá Íslandi. Annað markmið var að sýna dæmi um hagnýtingu gagnanna með því að rannsaka samband klínískra þátta hjá þunguðum konum og tíðnisviðs samdrátta. Þessi ritgerð lýsir aðferðunum sem notaðar voru við að safna saman, breyta, merkja og gera órekjanlegar 122 legrafritsmælingar sem framkvæmdar voru af höfundi á árunum 2008 til 2010. Síðan er gagnabankanum, sem hefur verið birtur hjá PhysioNet undir nafninu „Icelandic 16-electrode EHG database“, lýst ásamt samsvarandi Data Descriptor. Rannsóknin á klínísku þáttunum og legrafritunum frá gagnabankanum benda til þess að þyngdarstuðull, aldur og fyrri fæðingasaga þungaðra kvenna hafi áhrif á tíðnisvið samdrátta. Þetta gæti skipt máli varðandi þróun á lækningatækjum sem spá eiga fyrir um fyrirburafæðingar og sýnir um leið hvernig gagnabankinn getur nýst.

Lykilorð: *legrafrit, fyrirburafæðing, gagnabanki, opinn aðgangur, tíðnigreining*

Conceiving, compiling, publishing and exploiting the "Icelandic 16-electrode EHG database"

Ásgeir Alexandersson

60 ECTS thesis submitted to the School of Science and Engineering
at Reykjavík University in partial fulfillment
of the requirements for the degree of
Master of Science in Biomedical Engineering

January 2015

Student:

_____

Ásgeir Alexandersson

Supervisor:

_____

Brynjar Karlsson

Examiner:

_____

Magnús Örn Úlfarsson

Tilurð, útgáfa og hagnýting íslenska legrafritsgagnabankans

Ásgeir Alexandersson

60 ECTS eininga ritgerð lögð fram við tækni- og verkfræðideild
Háskólans í Reykjavík til
meistaraprófs (MSc) í heilbrigðisverkfræði

January 2015

Nemandi:

_____

Ásgeir Alexandersson

Leiðbeinandi:

_____

Brynjar Karlsson

Prófdómari:

_____

Magnús Örn Úlfarsson

Acknowledgements

I would like to thank Ahmad Diab for his help with compiling the data, Bjarni Vilhjálmur Halldórsson for his advice regarding the statistical analysis, Dima Alamedine, Héloïse Brihaye, Mathieu Laplanche, Daniel Suárez Escudero and Mélissa Moulart for their work on segmenting the contractions and Shirin Najdi and Arnaldo Batista for modifying the rdann function from the WFDB toolbox for MATLAB, which enabled the reading of annotation information from the database.

Many thanks to Alda, Guðrún, Alexander and the rest of my family for their inspiration and support.

Finally, a special thanks to Brynjar Karlsson, who I will always be grateful to for giving me unwavering support and guidance in my venture into uncharted waters.

Contents

List of tables

List of figures

1 Introduction

The field of medicine has a long history, but the systematic use of technology to provide a wide range of effective diagnostic and therapeutic instruments is a relatively new phenomenon. Ever since technology started having the dramatic impact on medical care seen in the last few decades, engineering professionals have become intimately involved in many medical ventures. As a result, the discipline of Biomedical Engineering has emerged as an integrating medium for the two dynamic professions of medicine and engineering and has assisted in the battle against illness and disease by providing tools such as biosensors, biomaterials, imaging techniques and artificial intelligence that can be used for research, diagnosis, and treatment. Biomedical engineers seek new solutions to the difficult health care problems confronting modern society[1]. One of the biggest global biomedical challenges is the prevention of preterm birth which is the main cause of mortality and morbidity in newborns[2]. Despite advances in the field, no good objective method currently exists to evaluate the stepwise progression of pregnancy through to labour, neither at term nor preterm[3–6]. One emerging technology that has shown promising results in predicting preterm labour is the external monitoring of uterine electrical activity using an electrohysterogram (EHG)[4–8]. One of the main problems for EHG researchers, as with other research teams standing on the frontline of scientific discovery, is a shortage of data to enable analysis and hypothesis testing. In this respect, public access databanks[9] are becoming increasingly important to engineers and scientists in biomedicine as well as in other fields. A new category of scientific article, the Data Descriptor, has been designed to provide detailed descriptions of databases to maximize the interpretation, search and reuse of the data[10]. The aim of this MSc project was to implement public access to EHG data recorded in Iceland. The data was to be converted into appropriate formats whilst making sure that all participant data was de-identified.

EHG signals are low pass filtered by the conductive properties of tissues lying between the uterus and the electrodes[11]. Multiparous women (history of previous deliveries) have more effective uterine activity during labour[12]. The second aim of this project was to test the hypothesis that the body mass index (BMI) and obstetric history of a pregnant woman have an effect on the frequency components of contractions. This hypothesis testing demonstrates how the clinical variables and EHG data from the database can be exploited retrospectively to generate interesting scientific results. Anyone can do studies such as this with the now publicly available Icelandic 16-electrode EHG database.

This thesis first describes the enormous scope of preterm birth as a public health problem, the current practices in pregnancy monitoring and the state of the art in using EHG recordings. The process of collecting, converting, annotating and de-identifying the EHG data recorded by the author between 2008 and 2010 in Iceland is explained in materials and methods. The now publicly available Icelandic 16-electrode EHG database is then described in detail in the results and the corresponding Data Descriptor is presented. The methodology of frequency analysis of contractions and linear regression analyses of frequency parameters vs clinical variables is also explained in materials and methods. The results suggest that a pregnant woman's body mass index, age and obstetric history may influence the frequency components. This could have implications for the development of a medical device for predicting preterm labour and at the same time demonstrates the use that can be made of the database.

2 Background

2.1 Pregnancy monitoring

The uterus is a hollow muscular organ located in the female pelvis[13]. It is continuous inferiorly with the cervix, which acts as its neck and communication with the vagina (see Figure 2.1). The superior part is the fundus and on either side of this the uterus communicates with the fallopian tubes at the cornu. The wall is made of smooth muscle that encloses the uterine cavity[12].

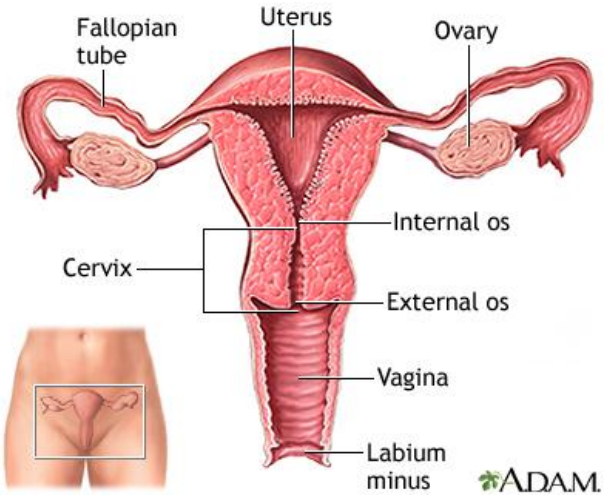

Figure 2.1: The uterus[13]

The uterus is normally palpable abdominally at 12-14 weeks gestation and by 20 weeks the fundus is usually at the level of the navel. After 24 weeks, the pubic symphysis-fundal height (in centimetres) approximately corresponds to the gestation, +/- 2 cm[12]. The uterus nourishes, protects, and ultimately expels the foetus by contraction of the smooth muscle. Ineffective contractions, however, also occur throughout the third trimester (last three months of pregnancy), which do not lead to delivery[12].

In clinical practice, contractions are currently assessed non-invasively with a tocodynamometer (TOCO). The TOCO is a strain gauge positioned over the pregnant uterus which responds to changes in uterine tension transmitted to the abdomen. The device identifies the frequency of contractions but not their intensity[14]. The TOCO records variations in local pressures thus making it highly dependent on the position of the TOCO. It is also influenced by the movements of the foetus within the intrauterine cavity, by maternal parameters (such as the amount of fat tissue), artefacts (respiratory or maternal movements) and by the elastic strap tension used to fasten the TOCO. The TOCO therefore must be constantly recalibrated and re-positioned to work correctly[15,16]. The TOCO is usually part of a device called a cardiotocograph, which is used to try to evaluate the well-being of a foetus. The cardiotocograph combines a

TOCO with a Doppler ultrasound device which is used to measure the foetal heart rate[12]. A cardiotocograph attached to a pregnant woman can be seen in Figure 2.2.

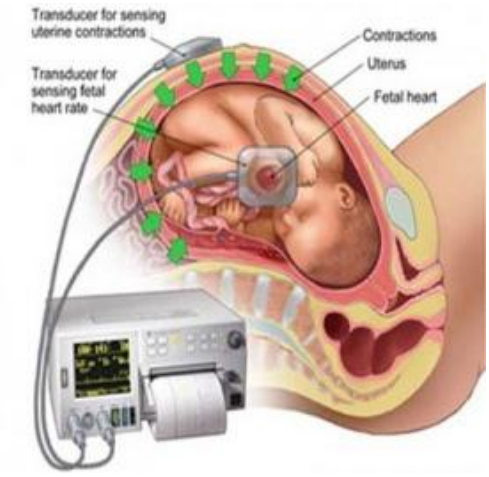

Figure 2.2: A cardiotocograph device attached to a pregnant woman[17]

A cardiotocography (CTG) trace can be seen in Figure 2.3. The upper trace shows the foetal heart rate and the lower trace shows the TOCO output. Each small square represents 30 seconds. The numbers on the foetal heart rate trace represent beats per minute and the numbers on the TOCO trace are supposed to represent the pressure in mmHg but they are, due to the factors mentioned above, in fact arbitrary. Two contractions are labelled in Figure 2.3 and small spikes due to foetal movement can also be seen.

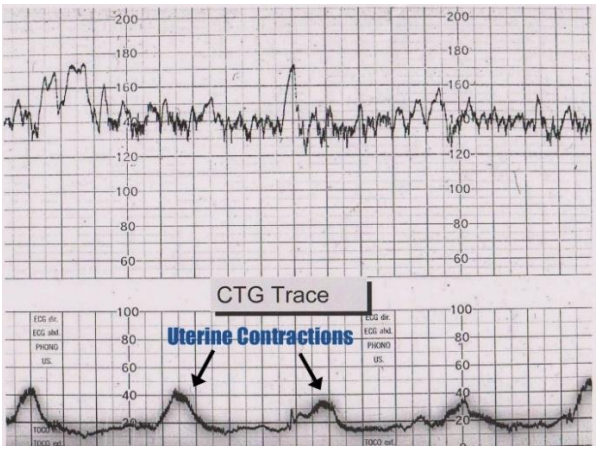

Figure 2.3: A cardiotocography (CTG) trace. The upper trace shows the foetal heart rate and the lower trace shows the tocodynamometer output. Two contractions are labelled[18].

2.2 Preterm birth

Labour is the process whereby the foetus and placenta are expelled from the uterus, normally occurring between 37 and 42 weeks gestation. The diagnosis is made when painful uterine contractions accompany dilatation and effacement of the cervix[12]. The steps leading to labour are not fully understood, but prostaglandin production has a crucial role both in reducing cervical resistance and increasing release of the oxytocin hormone from the posterior pituitary gland which aids stimulation of contractions[12]. Once labour is established, the uterus contracts for 45-60 seconds about every 2-3 minutes. The first stage of labour lasts until the cervix is dilated by 10 cm (fully dilated). The cervix often dilates slowly for the first 3cm; this is the "latent phase of labour". The second stage lasts from full dilatation of the cervix to delivery. The third stage is the time from delivery of the foetus to delivery of the placenta[12]. The rate of cervical dilatation and delivery time is on average slower in nulliparous women (no previous pregnancies beyond 20 weeks) compared to multiparous women. In fact, poor uterine activity and consequent slow labour is a common feature of the nulliparous woman and in induced labour, but is rare in multiparous women. Strengthening this activity artificially is called augmentation. Initially this is performed with amniotomy (artificial rupture of the membranes); if this fails, synthetic oxytocin is administered as a dilute solution[12]. The urge to push that is characteristic of the active second stage can be diminished after epidural analgesia, resulting in a prolonged second stage[12].

Preterm birth is defined as birth before 37 completed weeks of gestation. Around 15 million babies are born preterm every year. On average, 12% of babies are born preterm in lower-income countries and around 9% in higher-income countries. Preterm birth rates are increasing in almost all countries[2]. Preterm birth is the leading cause of new-born deaths (babies in the first four weeks of life) and the second leading cause of death in children under five years (after pneumonia). Over 1 million children die each year due to complications of preterm birth. Babies who are born preterm often require special care and face greater risks of serious health problems, including cerebral palsy, intellectual impairment, chronic lung disease, and vision and hearing loss. Most preterm births happen spontaneously. Common causes include multiple pregnancies, infections and maternal chronic conditions such as diabetes and high blood pressure; however, often no cause is identified[2,19]. If a woman is likely to be in preterm labour, between 24 and 34 weeks gestation, she is given corticosteroids. These reduce perinatal morbidity and mortality by promoting pulmonary maturity. As they take 24 hours to act, delivery is often artificially delayed using tocolysis (medication to suppress labour)[12].

Currently, there is no effective way of preventing preterm birth. The main reason for this is that there is no objective manner to evaluate the stepwise progression of pregnancy through to labour, either at term or preterm. Women may experience painful contractions without being in true labour. Various techniques have been adopted to monitor and/or diagnose labour, but they are either subjective or indirect and do not provide an accurate prediction of when labour will take place[3–6].

2.3 The electrohysterogram

Uterine contractility is a direct consequence of the underlying electrical activity in smooth muscle cells. Spontaneous electrical activities in the muscle are composed of intermittent bursts of spike action-potentials. Single spikes are able to initiate contractions, but multiple, higher-frequency, coordinated spikes are needed for forceful and maintained contractions[3]. The action potentials in the myometrial cells result from voltage- and time-dependent changes in membrane ionic permeabilities. The depolarizing phase of the spike is due to an inward current carried by Ca^{2+} ions and Na^{+} ions. The primary mode for elevation of Ca^{2+} is voltage-gated L-type channels that are regulated by agonists. The increase in Ca^{2+} produces Ca-calmodulin and activates myosin light chain kinase, which phosphorylates myosin and a contraction occurs[3]. The frequency, amplitude, and duration of contractions are mainly determined by the frequency of occurrence of uterine electrical bursts, the total number of cells that are simultaneously active during bursts, and the duration of uterine electrical bursts, respectively[3]. Each burst stops before the uterus has completely relaxed. Myometrial cells are coupled together electrically by gap junctions composed of connexin proteins which provide channels of low electrical resistance between cells, and thereby furnishes pathways for the efficient conduction of action potentials. Throughout most of pregnancy, these cell-to-cell channels or contacts are low, with poor coupling and decreased electrical conductance, which favours quiescence of the muscle and the maintenance of pregnancy. At term, however, the cell junctions increase due to changing oestrogen and progesterone levels and form an electrical syncytium required for coordination of myometrial cells for effective contractions[3].

The electrohysterogram (EHG) is the electrical signal recorded on the abdominal surface of a pregnant woman. Studies have shown that external monitoring of uterine electrical activity using EHG is representative of uterine contractility[11] and is able to detect a higher number of uterine contractions than TOCO[16].

A distinction between two main frequency areas can be made regarding EHG recordings: A low frequency area represents the periodic occurrence of contractions (at most around 0.005 Hz) whilst a fast wave frequency band, directly related to cellular activity, is representative of uterine activity. The fast waves contain two specific domains: a low frequency band (0.2 – 0.45 Hz) and a high frequency band (0.8 - 3 Hz)[11]. The fast wave low (FWL) frequency component may be related to propagation of EHG and the fast wave high (FWH) component may be related to the excitability of the uterus[20].

Various research groups have been investigating EHG and numerous methods have already shown promising results in predicting preterm labour[4–8]. The obtained results have, however, not shown sufficient reliability to provide an accurate tool for the detection of preterm labour.

Several types of parameters have been extracted from the EHG. They can be divided into three classes: linear parameters, non-linear parameters and parameters related to EHG propagation[21]. Linear methods in both time and frequency domains were among the first to be used to extract features from the EHG. There are four well known frequency-related parameters: mean frequency, peak frequency, median frequency and the limit frequency that contains up to 95% of the energy in the power spectral density (PSD). Of these, median frequency tends to be the best parameter that can be used on bipolar signals (difference between adjacent monopolar electrodes) for distinguishing between pregnancy and labour contractions[22]. It has, however, also been shown that peak frequency, mean frequency and 95%-limit frequency increases slightly throughout pregnancy[22]. Maternal age does not seem to affect peak frequency for women at term and in labour and the same can be said for parity (number of previous deliveries)[3]. Combined propagation velocity and power spectrum peak frequency appear to predict preterm delivery within 7 days better than Bishop-score (based on an assessment of the cervix and station of the foetal head in the pelvis by vaginal examination), contractions on TOCO, and cervical length measured by transvaginal ultrasound[6]. Compared to linear methods, non-linear methods may provide a superior way to differentiate between pregnancy and labour contractions[8,23,24] and multichannel recordings seem to improve this classification rate[25].

The abdominal muscles, subcutaneous fat and skin between the uterus and the abdominal electrodes form a conducting medium (volume conductor), through which the EHG signal propagates up to the maternal abdominal surface. EHG signals are low pass filtered by the conductive properties of these tissues[11,26]. The attenuation of high frequencies depends on the distance between the muscle and the recording site, on the conductive properties of the

tissues underlying the electrodes and on the skin impedance. The skin impedance can be considerably reduced by rubbing the skin and cleansing with acetone[11]. The fat layer forms a dielectric and a capacitance is developed as the high skin conductance and muscle conductance is separated by the high resistance in the subcutaneous fat. This creates a resistor capacitor (RC) time constant and a resulting low pass filter, with the cut-off frequency decreasing with increasing time constant[27]. Conductivity and permittivity are themselves frequency-dependant in biological tissues[28]. The low-pass filtering effect has also been seen in studies on striated muscle electromyograms where there is a loss of power in the high frequency region in subjects with a thicker skinfold[29,30].

2.3.1 16-electrode electrohysterograms

Most of the early studies on EHG used two to five electrodes[3,11] and therefore concentrated on the local excitability of the uterus. In 2007, a collaborative group from France and Iceland, involving biomedical researchers, engineers and medical doctors, started using electrodes in a 4-by-4 configuration on the abdomen[20]. This was mainly to better observe and analyse the spatial characteristics and propagation of the electrical activity during contractions rather than just the activity on a single location of the uterus as had been done before. The results from a preliminary study showed a very acceptable SNR (signal to noise ratio) on bipolar signals[31]. The use of the monopolar signals singly was however considered problematic, even with adaptive filtering methods due to external common mode noise (maternal ECG, respiration movements etc.). Efforts to use recent techniques such as empirical mode decomposition (EMD) and canonical component analysis (CCA) to clean up the signal have since met with some success[32]. The preliminary study data was also used to present a moving picture of the electrical activity during contractions. The activity was clearly observed and correlated well with the simultaneous tocograph trace[20]. In the preliminary study, the electrodes were placed one at a time for the 4-by-4 grid, which was a time-consuming task and achieving the desired inter-electrode distance required great operator precision. To address these issues, a placement guide system was specifically developed by the author in 2008[33]. The system has a standardized setup ensuring a consistent distance between electrodes. The system has a shorter electrode attachment time than for the one-by-one electrode attachment method and the placement of the electrodes into the placement guide can be done before a recording, shortening the setup time in an actual recording to around five minutes. This enables recordings to be performed when there is little time and reduces any inconveniences for participants and health professionals.

The 16-electrode placement guide system was used by the author to perform pregnancy and labour EHG recordings in Iceland between 2008 and 2010. Parts of the data have already been used for developing and analysing various signal processing methods and have led to several publications[8,21–25,34–42]. In particular the work has concentrated on efforts to accurately distinguish true labour contractions from normal pregnancy contractions with some success[8,21–25,34–38,41,42]. Parts of the data were included in the BioModUE_PTL project which led to the development and validation of a biophysics based multiscale model of the EHG, going from the cell to the electrical signal measured on the abdomen[43]. None of these past publications, however, have described all of the Icelandic EHG recordings in detail and studies have so far only used parts of the data.

2.4 Public access databases

Free public access to data enables the international scientific community to generate greater academic understanding and technological development. The data is made available without restrictions which encourages open-minded exploitation of the data.

PhysioNet offers free web access to large collections of recorded physiologic signals (found in the PhysioBank)[9]. Each month, about 45,000 visitors from around the world use PhysioNet, retrieving about 4 terabytes of data[44]. PhysioNet was established in 1999 by a diverse group of computer scientists, physicists, mathematicians, biomedical researchers, clinicians, and educators at Boston's Beth Israel Deaconess Medical Center/Harvard Medical School, Boston University, McGill University and the Massachusetts Institute of Technology (MIT) [45]. The PhysioNet web site is a public service of the PhysioNet Resource funded by the National Institutes of Health's National Institute of Biomedical Imaging and Bioengineering (NIBIB) and National Institute of General Medical Sciences (NIGMS)[45]. PhysioNet does not accept data sets containing protected health information listed in the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and reviews projects before making them open access[46].

The peer reviewed journal Scientific Data[10], published by Nature publishing group, is an open access, online-only journal for descriptions of scientifically valuable, public datasets and is open to submissions from a broad range of natural science disciplines[47].

> "Scientific Data primarily publishes Data Descriptors, a new type of scientific publication designed to promote an in-depth understanding of research datasets. Data Descriptors […] are devised to maximize data reuse and enable searching, linking and

data-mining. […] Data Descriptors include detailed descriptions of the methods used to collect the data and technical analyses supporting the quality of the measurements. Data Descriptors focus on helping others reuse data, rather than testing hypotheses, or presenting new interpretations, methods or in-depth analyses." [47]

3 Material and Methods

3.1 EHG recordings

3.1.1 16-electrode EHG recording protocol

To enable a standardized electrode setup for the 16-electrode EHG recordings, an alignment frame, a double sided hypoallergenic adhesive sheet and a silicone backing were designed and manufactured in 2008[33]. This was done as a part of the author's 3rd year research project (18 ECTS) at the Faculty of Medicine at the University of Iceland. The dimensions of the double sided adhesive sheet and silicone backing, with a 17.5 mm distance between adjacent electrode centres, can be seen in Figure 3.1. Reusable Ag/AgCl electrodes with a 13.0 mm outer diameter and an 8.0 mm inner diameter are used for recordings.

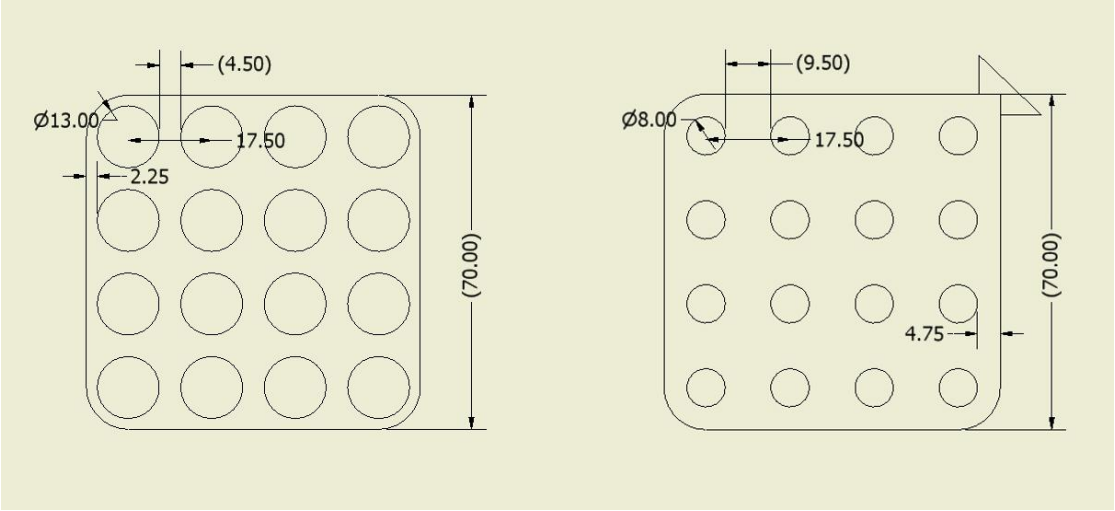

Figure 3.1: Dimensions (mm) of the silicone backing (left) and double sided adhesive sheet (right). The thickness of the silicone backing is 1.5mm.

The following recording protocol was developed that uses the placement guide system:

- The participant first off receives an explanation of what will be done and why. The participant is reminded that she can ask for the recording to be stopped at any time and she has no obligation to answer all questions. Before recordings can start, a signed consent from the participant has to be acquired.
- The alignment frame is used to align and attach an uncovered side of the double sided adhesive sheet to the silicone backing. The electrodes are then placed into the holes in the silicone backing and attached to the adhesive sheet. This can be done before an actual recording. The electrode numbering scheme, as seen when looking at the abdomen of the participant, can be seen in Figure 3.2.

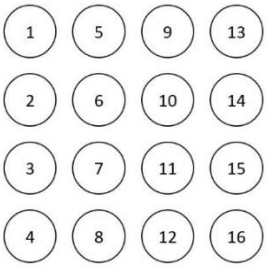

Figure 3.2: The electrode numbering scheme, as seen when looking at the abdomen of the participant.

- The abdominal skin is carefully prepared using an abrasive paste and alcohol solution. After filling the electrode holes with electrode gel, wiping the excess gel away with a straight edged card and uncovering the other side of the double sided adhesive sheet, the adhesive-silicone-electrode matrix is attached to the abdomen.
- The desired position on the abdomen is with the third vertical line of electrodes (electrodes 9 to 12) placed on the median axis of the uterus and the 10th – 11th pair of electrodes half way between the fundus and pubic symphysis. The navel is avoided by displacing the matrix up or down whilst staying as close as possible to the desired position. The skin over the iliac crests on both sides is prepared in the same way as the abdomen and reference electrodes (patient ground and reference) with electrode gel are then attached on each side using adhesive washers, with inner diameters corresponding to the inner diameter of the electrodes. The electrode positions can be seen in Figure 3.3.

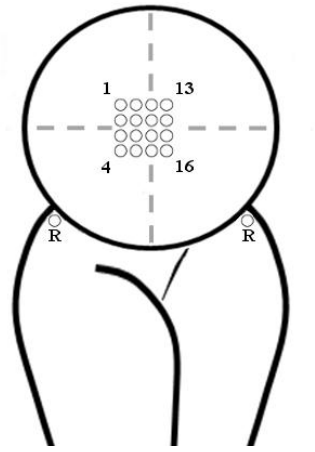

Figure 3.3: Ideal position of the 4-by-4 electrode grid. The third vertical line of electrodes (electrodes 9 to 12) is placed on the median axis of the uterus and the 10th – 11th pair of electrodes half way between the fundus and pubic symphysis. The corner electrodes of the 4-by-4 grid are labelled along with the reference electrodes, R (patient ground and reference).

- A TOCO is also attached to the abdomen during recordings. When possible, a support, such as a pillow, is positioned under the right side of the woman to prevent potential aortocaval

compression syndrome (compression of the abdominal aorta and inferior vena cava by the gravid uterus when the woman is flat on her back. Symptoms can be fatigue, yawning, light-headedness, dizziness, nausea, sweating, pallor, shortness of breath, general feeling of being unwell). A photo of the setup during a recording can be seen in Figure 3.4.

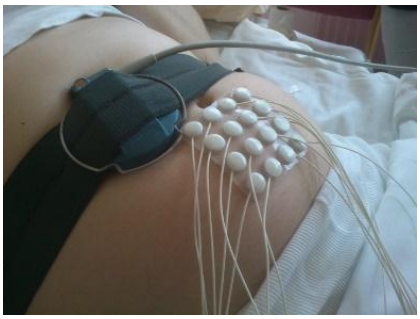

Figure 3.4: The recording setup. The abdominal electrodes and the tocodynamometer can be seen.

- The intended duration of a pregnancy recording is one hour and the intended duration of a labour recording is at least half an hour but the participant is allowed to stop the measurement at any time.
- Each participant is assigned an ID number and for each recording, information on participant age, BMI, gestational age, placental position, gravidity (number of times participant has been pregnant, including current pregnancy), parity (number of previous deliveries) and history of caesarean is recorded in an excel file. Eventual outcomes of the pregnancies, i.e. mode of delivery, gestational age at delivery, use of synthetic oxytocin and/or epidural during labour and the time of these with regard to the birth are recorded after the event, through contact with the participant or through the participants obstetric record (which the participant has agreed to at the time of the recordings). During a recording, the researcher records participant movements, equipment manipulation, participant-reported possible contractions and foetal movements and any other unusual events. The participant and researcher can converse freely during the recordings and no restrictions are placed on the participant in terms of changing position if needed.
- An electrode that is unconnected or floating will give a signal which very quickly goes to saturation and is therefore easily recognized. If during a recording, an electrode gives a trace that is visually very unlike the traces of other electrodes or displays values notably different from other values, it is pressed more firmly onto the abdomen to ensure connection to the skin. If all electrode traces seem suboptimal (e.g. very noisy visually), the reference

electrode connections are tested. When suboptimal traces do not improve, the electrodes are reconnected and the recording then resumed. If the readout of the tocograph goes down to zero, the TOCO is recalibrated to 20.

Using this protocol, recordings are made in a very similar way to the preliminary study, with the same electrode configuration, the same electrodes and same recording device but with a slightly smaller inter-electrode distance (17.5mm vs 21mm).

3.1.2 16-electrode EHG recordings in Iceland

The 16-electrode placement guide system was used to perform pregnancy and labour EHG recordings in Iceland at Landspitali University Hospital, Akureyri Hospital and the Akureyri Primary Health Care Centre between 2008 and 2010. All recordings were performed by the author (ÁA) using the protocol. The researcher (ÁA) stayed with the participants during recordings and recorded events first hand and monitored the equipment and electrode readouts continually.

Participants were invited to take part in the recordings during antenatal care visits or at the labour wards and had normal singleton pregnancies and no known risks for preterm labour. Informed consent was obtained from every participant and the protocol was approved by the National Bioethics Committee in Iceland (VSN 02-0006-V3). After each pregnancy recording, the participant was invited to take part in another recording one to two weeks later. The aim of these multiple, or longitudinal, measurements was to observe the evolution of contractions during pregnancy and into labour.

The measurements were performed using a sixteen channel multi-purpose physiological signal recorder (Embla A10), most commonly used for investigating sleep disorders. An anti-aliasing filter with a high cut-off frequency of 100 Hz was used but no high pass filter was used. The signal sampling rate was 200 Hz and the signal was digitized to 16 bits. The sixteen monopolar electrode signals were originally stored in the EDF (European Data Format) format by the Somnologica software used to control the Embla A10. Paper TOCO traces were scanned to JPEG files.

3.1.3 The Icelandic 16-electrode EHG database

All EDF files, Excel files containing clinical recording data and events and scans of tocographs were gathered together. In each recording year between 2008 and 2010, a different ID naming

system had been used. A new ID system was therefore created, where each of the old ID names received a new ID of the form ice# (# a number between 001 and 045). The numbering in the new ID system is not sequential between years so the ascending numbers do not represent a chronological recording order. As before, the principal investigator of the Icelandic EHG study has the only copy of the document connecting the old ID names with actual participant personal information.
Recordings were classified into two types:

- Pregnancy recordings - defined as recordings performed at antenatal care clinics on participants in the third trimester and not suspected to be in labour. They were performed at Akureyri Primary Health Care Centre and Landspitali University Hospital.
- Labour recordings - defined as measurements performed on participants suspected to be in labour, present in the labour ward and who delivered within 24 hours. They were performed at Landspitali University Hospital and Akureyri Hospital.

Cygwin, a large collection of GNU and Open Source tools which provide functionality similar to a Linux distribution on Windows[48], was used to install the GNU General Public License led WFDB (WaveForm DataBase) Software Package[49] and to convert the EDF files obtained during the recordings into WFDB compatible files. Binary sixteen-bit amplitude signal files (.dat)[50] and header files (.hea) were generated using the edf2mit WFDB application[51]. For each database record, a header file specifies the names of the associated signal files and their attributes[52]. Header files contain a record line, which specifies the record name, the number of signals (16 for these recordings) and the sampling frequency (200 Hz for these recordings). Header files also contain signal specification lines for each electrode signal. They contain the following (with the corresponding information from the Icelandic EHG data in parentheses):

- File name - The name of the file in which the signal data is kept.
- Format - An integer that specifies the storage format of the signal (format 16).
- ADC gain (ADC units per physical unit) - A floating-point number that specifies the difference in sample values that would be observed if a step of one physical unit occurred in the original analogue signal (131.068/mV where mV is the specified unit).
- ADC resolution (bits) - Specifies the resolution of the analogue-to-digital converter used to digitize the signal (15).

- ADC zero - An integer that represents the amplitude (sample value) that would be observed if the analogue signal present at the ADC inputs had a level that fell exactly in the middle of the input range of the ADC. Together with the ADC resolution, the contents of this field can be used to determine the range of possible sample values (0).
- Initial value - Specifies the value of sample 0 in the signal (0).
- Checksum - A 16-bit signed checksum of all samples in the signal. If the entire record is read without skipping samples, and the header's record line specifies the correct number of samples per signal, this field is compared against a computed checksum to verify that the signal file has not been corrupted.
- Block size - This field is an integer and is usually zero (0).
- Description - Any text between the block size field and the end of the line is taken to be a description of the signal.

Signal names identify the digitized signals in PhysioBank data collections and are placed at the end of signal specification lines (description). New signal names under the EMG class were created for the Icelandic 16-electrode EHG database. The signal names are of the form "EHGn" (where "n" is a number between 1 and 16) and represent electrodes 1 to 16 of the EHG electrode matrix[53].

During conversion, recording dates and times from the EDF files were exported to the record lines of the header files. These dates were deleted manually. No other information regarding the original recording files (such as dates) could be found in the signal files, header files or in the corresponding file properties after conversion.

Clinical recording information from the Excel files was added manually to the end of each header file as info strings commencing with '#'. The information added was:

-Participant ID (of the form ice#)

-Record number (showing what number the recording is of all recordings for that participant)

-Record type (labour, pregnancy)

-Age of the participant (years)

-BMI (body mass index) of participant just before pregnancy

-BMI of participant at time of recording

-Gravidity (previous pregnancies including current pregnancy)

-Parity (previous pregnancies beyond 20 weeks)

-Previous caesarean (Yes, No)

-Placental position

-Gestational age at recording (weeks/days), according to first trimester ultrasound

-Gestational age at delivery (weeks/days)

-Mode of delivery (Vaginal, Vaginal/Induction, Elective caesarean, Emergency caesarean due to slow progress, Emergency caesarean due to other than slow progress). Vaginal delivery indicates spontaneous onset unless appended with /Induction.

-Synthetic oxytocin use in labour (Yes, No)

-Epidural during labour (Yes, No)

-Comments for recording (such as explanations of equipment manipulation)

-Comments for delivery (such as timings of synthetic oxytocin use, epidural use or labour recordings with regard to time of birth)

WAVE, a component of the WFDB Software Package, is an X Window System client application and can be used to visual and annotate WFDB-compatible signal files[54]. All of the converted signals were viewed in WAVE and signal values in a random place in the recording were compared to values at the same time in the EDF files, visualised with an open source EDF viewer, EDFbrowser[55]. A screen shot of an EHG signal in WAVE and EDFbrowser can be seen in Figure 3.5 and Figure 3.6 respectively. The figures show the same signal values between viewers at a specific recording time.

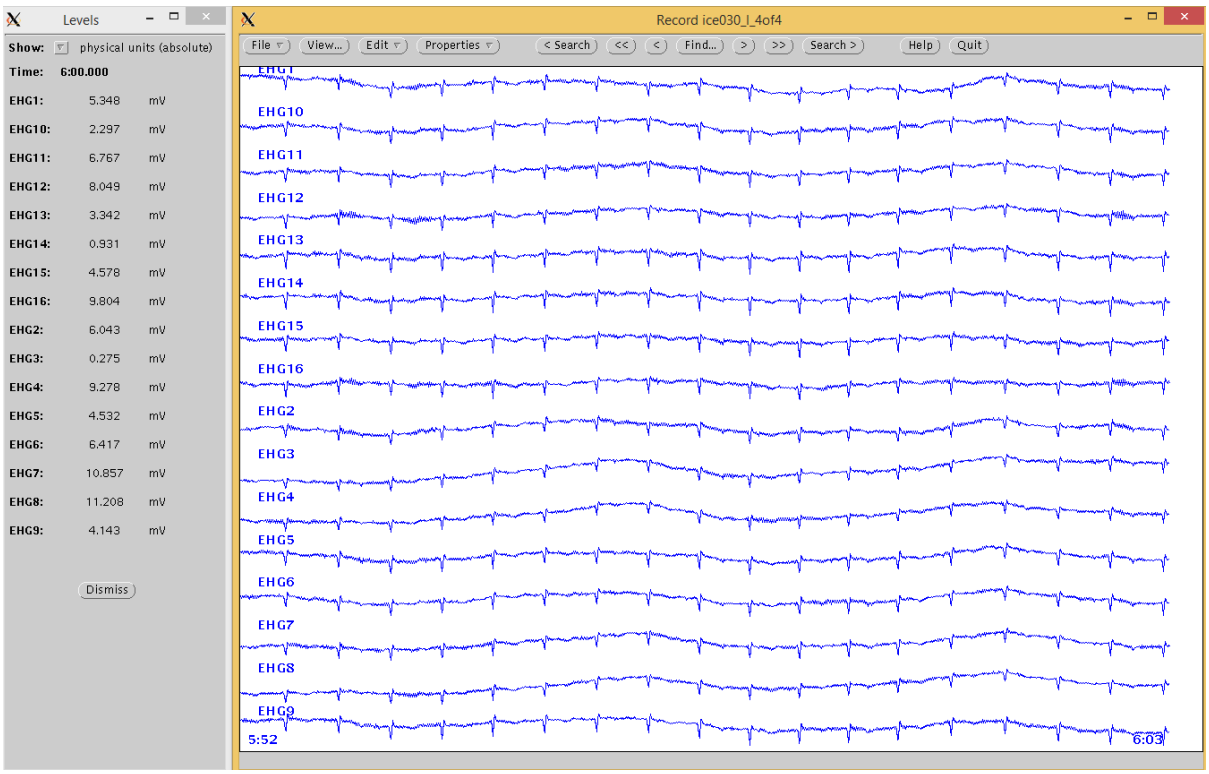

Figure 3.5: An EHG recording visualised in WAVE. The signal values at the recording time of 06:00:00 can be seen on the left side.

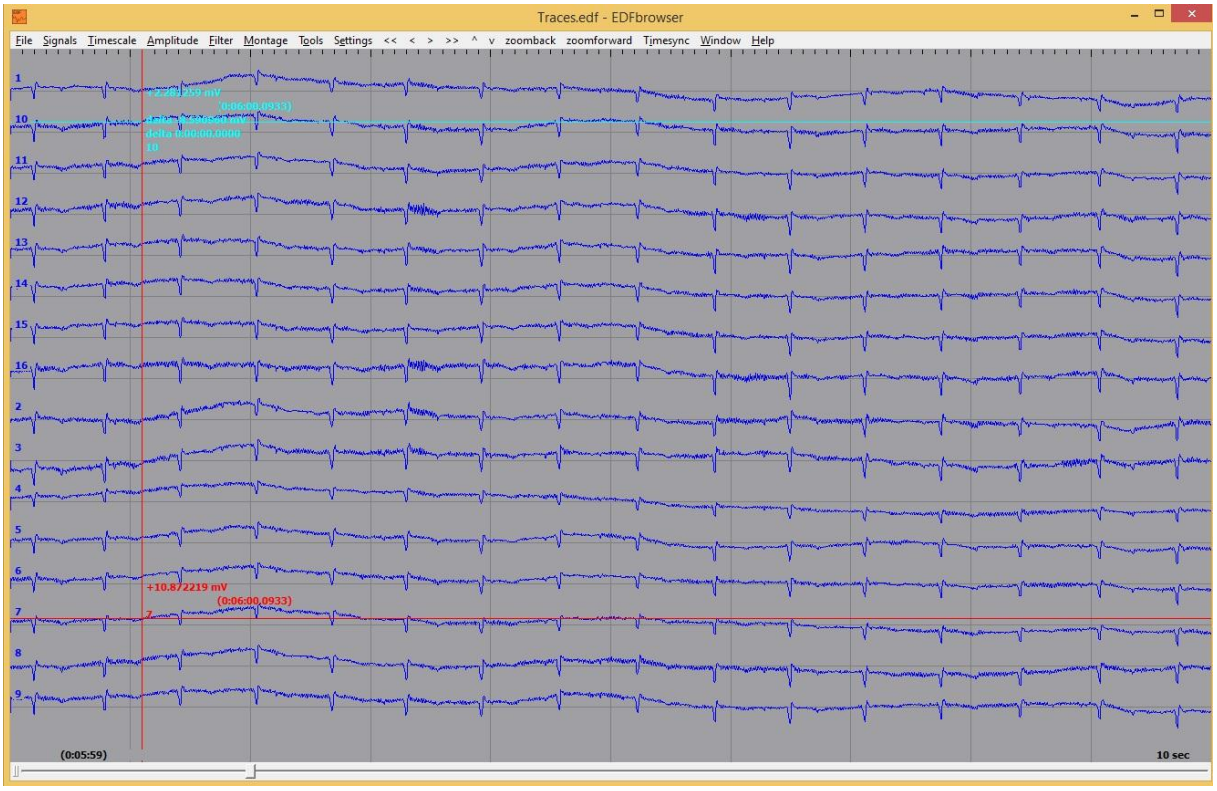

Figure 3.6: The same EHG recording as in Figure 3.5 visualised in EDFbrowser. The signal values for electrodes 10 and 7 at the recording time of 06:00:00 can be seen.

Event information was taken from the Excel files and categorised into the following types:

- C - Contraction. Used when the participant feels a contraction (the TOCO may not always have caught it) or there is a very likely contraction on the tocograph (not always used when there is an obvious contraction on the tocograph).
- (c) - Possible contraction. Used when there is not a very likely contraction but the participant has pressure sensation or a contraction is suspected on the tocograph.
- pm - Participant movement.
- pos - Participant change of position.
- fm - Foetal movement. Used when the participant feels foetal movement.
- em - Equipment manipulation. Used when electrodes are pressed more firmly onto the abdomen if otherwise not explained in comments.

A text file with a table of these annotation definitions was created and appended to Wave.Anntab in .Xdefaults in the home directory[56]. The WFDB-compatible signal files were opened using WAVE and annotation files containing the events were created by manually adding to the correct positions in the signal files.

The wfdb2mat WFDB application was used to convert the WFDB-compatible signal files to MATLAB .mat files[57]. The rdann function from the WFDB toolbox for MATLAB[58] creates .mat files containing annotation information. It had to be modified as the function did not accept windows like paths (such as c:\ehg\...) or annotation strings with more than 2 characters.

CTG machines, corresponding to the types used for the recordings, were studied with regard to recording delays when turned on, to best enable synchronization of the EHG signals with the tocographs. The JPEG tocograph scans were imported to a graphics editor (GIMP[59]) and a time axis corresponding to the recording was inserted onto the scanned images. All information that could possibly lead to the identification of the participant, such as personal data and dates, were manually removed from these images. The images were then exported to new JPEG image files, ensuring that the image file properties did not contain any dates corresponding to recording dates.

At the time of data conversion, PhysioNet was not on the list of Scientific Data's recommended data repositories[60]. The author contacted Scientific Data, suggesting PhysioNet as a recommended repository. Scientific Data now includes PhysioNet as a recommended repository.

A supplement to the existing research approval (VSN 02-0006-V3) was written and submitted to the National Bioethics Committee of Iceland regarding the open access publication of the non-identifiable data. A similar letter was sent to the Data Protection Authority. These applications can be seen in the appendix. At the request of the National Bioethics Committee, the heads of department from the hospitals and health care centre from the study were also notified of the open access publication.

To start off with, a private project was created on a PhysioNetWorks workspace. Samples of the data were added to the project. After approval from a member of the PhysioNet team, the project became an active project, meaning that the rest of the database could be uploaded and other users could download the data after invitation from the owner. The database was then submitted for reviewing with regard to open access publication.

3.2 Frequency analysis of contractions

The research teams from Iceland and France have gone through the EHG recordings performed with the placement guide system and segmented contraction bursts[21,22]. During segmentation,

the monopolar signals are first filtered with a 0.1 Hz high-pass filter and a 35 Hz low-pass filter. Bipolar signals are calculated from adjacent vertical monopolar signals and the bipolar and monopolar signals are then plotted parallel to the tocograph. The signal files are segmented according to the contractions seen on the tocograph and each contraction signal can then be treated with the CCA-EMD technique[32].

The CCA-EMD contraction signals were analysed in MATLAB. The power spectral density (PSD) is a function of frequency and represents the relative contribution of each frequency to a signal[22]. The PSD of each monopolar electrode signal from each contraction was estimated using Welch's power spectral density estimate (pwelch function). The PSD was then used to calculate the following frequency parameters:

Mean power frequency

Mean power frequency (MPF) is calculated as the sum of the product of the power spectrum intensity and the frequency, divided by the total sum of the intensity of the PSD[22].

$$MPF = \frac{\sum_i^n I_i f_i}{\sum_i^n I_i}$$

Where n denotes the number of frequency components in the power spectrum, f_i the value of the frequency and I_i the intensity of the power spectrum corresponding to f_i.

Peak frequency

Peak frequency (PEAK) is the frequency that has the largest corresponding amplitude peak in the PSD.

Median frequency

Median frequency (F50) is the frequency where the sums of the parts above and below in the PSD are the same.

95%-limit frequency

95%-limit frequency (F95) is the frequency below which the frequency-power spectrum contains 95% of the total energy of the signal.

The psd2.m function used for the calculations, created by Marcos Duarte, can be found in the appendix.

3.2.1 Statistical analysis

The relative standard deviation (RSD) is the absolute value of the coefficient of variation (CV). It is defined as the ratio of the standard deviation, σ, to the mean, μ and can be expressed as a percentage (%RSD)

$$\%RSD = \frac{\sigma}{\mu} * 100$$

A high percentage indicates high variability in the dataset.

Multiple linear regression is used for modelling the relationship between a dependant variable and one or more explanatory variables. The model can be described as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

Where y is the predicted independent variable, x_i are the explanatory variables, β_i are the regression coefficients and n is the number of explanatory variables. The p-value for each regression coefficient tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value indicates that the null hypothesis can be rejected. The coefficient of determination, R^2, shows how well data fits a model and indicates what proportion of variations of outcomes the model explains. In multiple regression analysis, the adjusted R^2 is adjusted for the number of variables included in the regression equation.

The Pearson product-moment correlation coefficient is a measure of the strength of the linear relationship between two variables.

The function *colldiag* from the perturb R package[61] investigates collinearity, where two or more explanatory variables are highly correlated. The function is an implementation of regression collinearity diagnostic procedures and computes condition indexes of the matrix of explanatory variables. It also computes variance decomposition proportions associated with each condition index[62].

Excel was used to calculate the %RSD of each frequency parameter from all 16 electrodes of each contraction. From the results of this analysis, it was decided to use the averages of the frequency parameters from each of the 16 monopolar electrodes from a contraction for the multiple linear regression analyses.

The R software environment for statistical computing was used for multiple linear regression analysis. The following parameters, taken from the Icelandic 16-electrode EHG database, were classified as explanatory variables and analysed along with the frequency parameters from the PSD (classified as dependant variables):

- Recording type (made into a dummy variable – 1 for labour and 0 for pregnancy)
- Age of participant [years]
- BMI before pregnancy [kg/m2]
- BMI at recording [kg/m2]
- Gravidity
- Parity
- History of caesarean (made into a dummy variable – 1 for yes and 0 for no)

- Gestational age at recording (weeks presented in decimal notation, e.g. 31.14 is 31 weeks and 1 day)
- Days before labour (calculated after labour – a recording two weeks before labour would be 14 days, a labour recording would be 0 days)

Analysis was performed on two sets of data. The first set was all of the segmented contractions. The second set was all segmented contractions recorded from participants with pregnancies ending with labour of spontaneous onset, this excluded contractions from participants that were induced or had a caesarean when they were not already in spontaneous labour.

A correlation coefficient matrix was calculated for all variables (explanatory and dependant). Simple linear regression (linear regression with one explanatory variable) was performed for the different dependant variables with each explanatory variable. Multiple linear regressions were applied with all explanatory variables in relation to each of the frequency parameters. Statistical significance of a regression coefficient was defined as p-value < 0.05. Variables with correlation coefficients above 0.7 were not included together in the same regression analysis[63] and regressions were performed twice for each such pair, each analysis with one of the variables. The significant variables from the all-inclusive analyses were then used in a final multiple linear regression. In the event of a correlated pair giving significant regression coefficients, the explanatory variable having a larger effect on the adjusted R^2 was chosen. For each dependant variable, the model giving the highest adjusted R^2 was noted.

All models were investigated for collinearity and if the largest condition index was 30 or higher (indicating unstable parameter estimates – interception included)[62] then the parameter/s with variance decomposition proportions higher than 50% were to be labelled and either one removed as these variables may cause collinearity problems[62].

For the analysis of all of the segmented contractions, the explanatory variables from the final multiple linear regression analyses were plotted in a one to one relationship with the dependant variables to visualize the relationships (created in MATLAB).

The commands used in MATLAB and R can be found in the appendix.

4 Results

4.1 The Icelandic 16-electrode EHG database

The National Bioethics Committee approved the supplement (VSN 02-0006-V4) and the Data Protection Authority did not need to act as no identifiable data was in the database. PhysioNet reviewed the database and accepted it for public access release. The database can be found in the public access PhysioBank under the name "The Icelandic 16-electrode EHG database"[64]. The web page presents a short description of EHG recordings and the 16 electrode placement guide protocol along with a description of the data files and some usage notes regarding the data. All recordings in the database were performed by the author (ÁA). Data from the preliminary study by Karlsson et al.[20] is not included in the database.

The database consists of 122 recordings performed on 45 participants. Of these 45 participants, 32 were measured more than once during the same pregnancy and the highest number of recordings for a participant was seven recordings. Ten recordings were performed during labour and five participants took part in measurements during both pregnancy and labour. The lowest gestational age was 29 weeks and 5 days (29w5d - pregnancy recording) and the highest gestational age was 41 weeks and 5 days (41w5d - labour recording). The average recording duration for pregnancy recordings was 61 minutes (range 19-86 minutes) and the average recording duration for labour recordings was 36 minutes (range 8-64 minutes).

File names in the database are of the form ice###_*type*_*record number*, where ice### is the ID of the participant (e.g. ice001), *type* refers to the type of recording: p (pregnancy) or l (labour), and *record number* is the number of recording for that particular participant (e.g. 1of3). Each recording has three associated files:

-A binary signal file containing the data from the 16 monopolar electrodes (.dat file)

-A scanned tocograph with a manually inserted recording time axis (.jgp file). Each small square represents 30 seconds.

-A header file (.hea file). Lines 2-17 of the header files are signal specification lines and the strings at the end of these lines correspond to the signal labels. The signals are labelled with "EHG*n*" where *n* refers to the relevant electrode number (electrode numbers are not in ascending order). Information from each recording is in info strings (commencing with #) at the end of the header files.

For 111 of the recordings, there is also an annotation file containing the type and timing of events (.atr file).

The database also includes:

-The zip file icelandic16ehgmat.zip that includes MATLAB versions of all the signal files along with header files (file names of the form ice###_*type*_*record number*m) and annotations (created using rdann.m, file names of the form ice###_*type*_*record number*m_ann).This is provided for the convenience of users that want to analyse the data in MATLAB.

-RECORDS.txt containing a list of the records by record name, with one record name per line.

The table in the appendix contains the clinical information from each recording (information from the header files excluding comments) along with recording durations and whether or not the recording has a corresponding annotation file. This information can also be found in info.txt in the database.

The annotations and the tocograph complement each other. Some contractions that are present in the annotations are not obvious on the tocograph and some obvious contractions on the tocograph are not in the annotations. This explains in part why some recordings are without annotations. Noted foetal movement could last for differing amounts of time and participants did not always notify if they felt foetal movement. Sometimes foetal movement can be seen on the tocograph. There can be differences in when participants start to feel a contraction or foetal movements and so differences in when participants notify about events. Events were therefore occasionally approximated to the nearest whole minute. Despite trying to take TOCO delays into account, there can be misalignment in the synchronization of EHG signals with the tocograph. Due to factors such as these, the inserted recording times on the tocographs and the annotation times may be up to ± 30 seconds from the actual recording or event times.

Sample numbers start at 1 in the MATLAB files but at 0 in the WFDB signal files and so there is a sample number discrepancy of 1 between files. The MATLAB files contain the absolute raw units. Division by 131.068 gives the physical units in mV.

Even though some pregnancies ended in caesarean, the participant was on occasion already in spontaneous labour. These incidences are explained in the comments sections of the header files. If a recording was close to birth, then the timing of the recording with regard to the birth is in the comments of the header file.

Users can view the data and annotations through two web interfaces: LightWave, a lightweight JavaScript viewer[65], and ATM (Automated Teller Machine)[66]. LightWave viewer enables visualization of the signals and annotations with an autoplay option and an annotation searcher. It also provides a summary of the signal specification lines, annotations and clinical information in the Tables tab (see Figure 4.1).

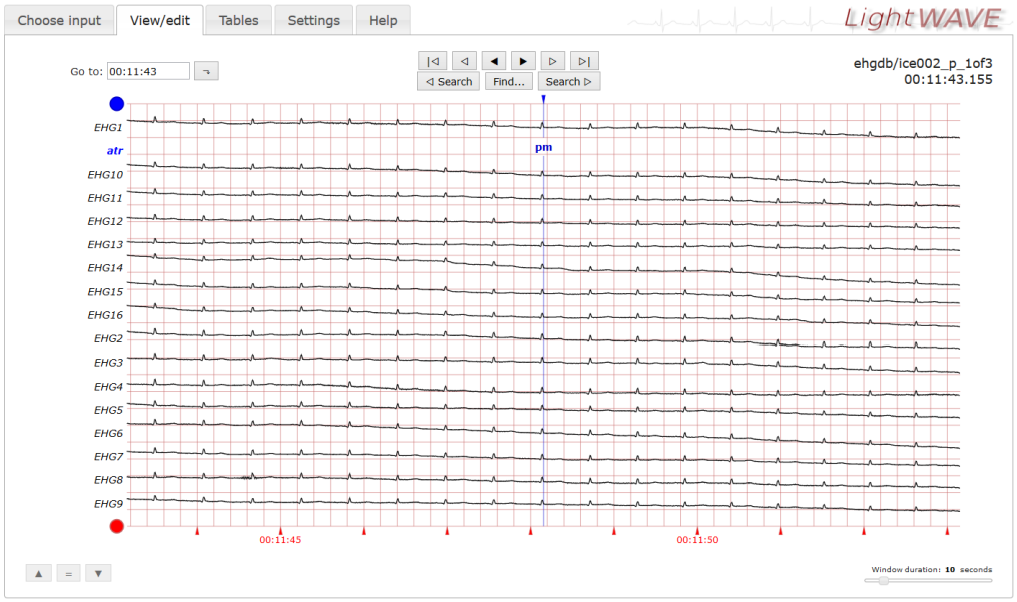

Figure 4.1: The LightWave online viewer at PhysioNet. A section of the ice002_p_1of3 recording can be seen along with an annotation for participant movement (pm).

ATM enables visualization of the annotated signals (see Figure 4.2) and has a toolbox with software that can convert WFDB signal files to text, CSV, EDF or .mat files. It can also show signals and annotations as text and can make a tarball or zip file of the record for ease of downloading.

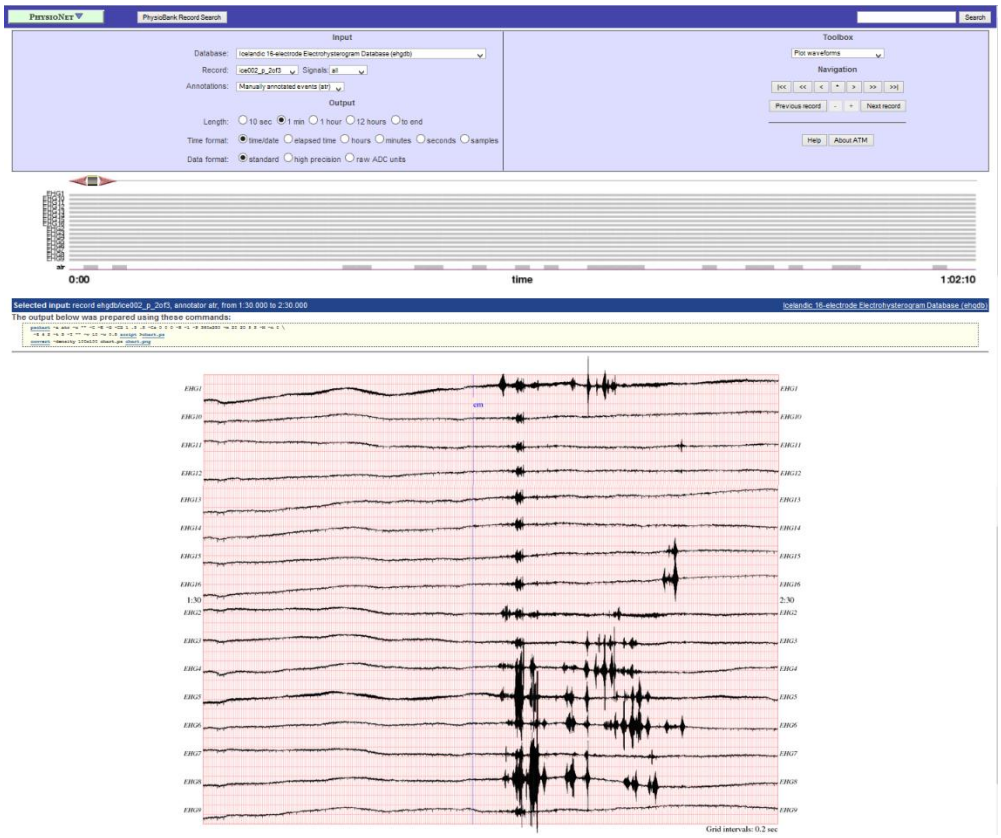

Figure 4.2: The ATM online viewer at PhysioNet. A section of the ice002_p_2of3 recording can be seen along with an annotation for equipment manipulation (em).

WAVE can also be used for viewing and analysing the signals[54]. Annotations can be opened by using the –a option with the wave WFDB application[67].

The WFDB (WaveForm DataBase) Software Package can be used to work with the recordings[49]. For example, the command rdann -r *record name* -a ann -f 0 can be used to read the annotation files[68].

The WFDB toolbox for MATLAB can be used to work with the MATLAB files[58]. The new version of the toolbox (version 0.9.9) will include the modified version of the rdann function.

Scientific Data will add PhysioNet to their list of recommended data repositories and a Data Descriptor describing the Icelandic 16-electrode EHG Database is in review.

Research teams from three countries are already using the database for EHG analysis and collaboration with the teams is ongoing. These teams are from Universidade Nova de Lisboa, Lisbon, Portugal (Arnaldo Guimarães Batista agb@fct.unl.pt), Faculty of Engineering & Information Technology, University of Technology, Sydney, Australia (Ganesh Naik Ganesh.Naik@uts.edu.au) and ETH, Zurich, Switzerland (Donnacha Daly ddaly@ethz.ch).

4.2 Frequency parameters vs clinical variables

4.2.1 All contractions

Frequency analysis was performed on all 16 electrodes for 318 contractions from 31 participants during 68 recordings (9 labour recordings). A total of 57 contractions were labour contractions and 261 were pregnancy contractions. The results from the relative standard deviation (%RSD) calculations on all 16 electrodes for each contraction can be seen in Table 1.

Table 1: Relative standard deviation (%RSD) analysis of 318 segmented contractions from the Icelandic 16-electrode EHG database. The mean, median, maximum and minimum values of %RSD from the 16 electrodes of all the contractions are given. MPF: Mean power frequency [Hz], PEAK: Peak frequency [Hz], F50: Median frequency [Hz], F95: 95%-limit frequency [Hz].

Frequency parameter	**Mean %RSD**	**Median %RSD**	**Maximum %RSD**	**Minimum %RSD**
MPF	19.2	16.3	66.7	3.9
PEAK	22.1	19.6	67.8	4.6
F50	14.4	13.1	80.7	1.4
F95	35.5	27.6	135.5	3.3

Due to the high variability in the %RSD analysis, the averages of the frequency parameters from each of the 16 monopolar electrodes from a contraction were used for the linear regression analyses.

The Pearson correlation coefficient matrix for all explanatory and dependant variables from all contractions can be seen in Table 2. Correlation coefficients higher than 0.7 are in bold font. This occurs, as expected, for BMI before pregnancy vs BMI at recording, Gravidity vs Parity, Gestational age vs Days before labour, MPF vs F95 and F50 vs PEAK.

Table 2: Pearson correlation coefficient matrix for 318 segmented contractions from the Icelandic 16-electrode EHG database. Correlation coefficients higher than 0.7 are in bold font. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], MPF: Mean power frequency [Hz], PEAK: Peak frequency [Hz], F50: Median frequency [Hz], F95: 95%-limit frequency [Hz].

	Labour	Age	BMIb	BMIr	Gravidity	Parity	Caes	GA	Days	MPF	PEAK	F50	F95
Labour	**1.000**	0.075	0.227	0.295	0.080	0.003	0.074	0.489	-0.494	0.337	0.150	0.160	0.287
Age	0.075	**1.000**	0.372	0.341	0.256	0.508	0.253	0.158	-0.286	-0.163	-0.056	-0.207	-0.097
BMIb	0.227	0.372	**1.000**	**0.927**	0.126	0.063	-0.219	0.154	-0.084	0.082	-0.156	-0.217	0.155
BMIr	0.295	0.341	**0.927**	**1.000**	0.279	0.189	-0.123	0.272	-0.223	0.095	-0.101	-0.160	0.147
Gravidity	0.080	0.256	0.126	0.279	**1.000**	**0.777**	-0.115	0.198	-0.328	0.013	0.085	0.053	0.031
Parity	0.003	0.508	0.063	0.189	**0.777**	**1.000**	0.084	0.121	-0.287	-0.067	0.154	0.120	-0.056
Caes	0.074	0.253	-0.219	-0.123	-0.115	0.084	**1.000**	-0.090	-0.141	0.014	-0.004	-0.022	-0.008
GA	0.489	0.158	0.154	0.272	0.198	0.121	-0.090	**1.000**	**-0.883**	0.124	0.025	-0.023	0.124
Days	-0.494	-0.286	-0.084	-0.223	-0.328	-0.287	-0.141	**-0.883**	**1.000**	-0.225	-0.011	-0.012	-0.219
MPF	0.337	-0.163	0.082	0.095	0.013	-0.067	0.014	0.124	-0.225	**1.000**	0.088	0.500	**0.964**
PEAK	0.150	-0.056	-0.156	-0.101	0.085	0.154	-0.004	0.025	-0.011	0.088	**1.000**	**0.767**	-0.070
F50	0.160	-0.207	-0.217	-0.160	0.053	0.120	-0.022	-0.023	-0.012	0.500	**0.767**	**1.000**	0.310
F95	0.287	-0.097	0.155	0.147	0.031	-0.056	-0.008	0.124	-0.219	**0.964**	-0.070	0.310	**1.000**

4.2.1.1 Mean Power Frequency

Results from the linear regression analyses with Mean Power Frequency as the dependant variable can be seen in Table 3. Regression coefficients were statistically significant for four explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), Age of participant, Gestational age at recording and Days before labour.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.166) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy and Days before labour were always statistically significant in the multiple regression analyses of all explanatory variables. The model containing only these variables had all regression coefficients statistically significant and an adjusted R^2 of 0.165. The plots of MPF vs these parameters can be seen in Figures 4.3 – 4.6. The average MPF from all contractions was 0.346 Hz. According to the model, MPF is 0.089 Hz higher in labour compared to pregnancy, MPF decreases by 0.007 Hz for each yearly increase in age (age range of participants 19-39 years), MPF increases by 0.004 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2) and MPF decreases by 0.001 Hz for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 3: Results from the linear regression analyses with Mean Power Frequency as the dependant variable for 318 segmented contractions from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. The light dark grey row represents the model with the highest adjusted coefficient of determination when each regression coefficient is statistically significant. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.121	**7.E-10**																	0.114
		-0.004	**0.004**															0.026
				0.003	0.143													0.007
						0.004	0.092											0.009
								0.001	0.819									2.E-04
										-0.011	0.234							0.004
												0.008	0.800					2.E-04
														0.006	**0.027**			0.015
																-0.002	**5.E-05**	0.051
0.117	**3.E-07**	-0.007	**4.E-05**	0.004	**0.048**			0.004	0.343			0.050	0.137	-0.001	0.829			0.147
0.085	**2.E-04**	-0.008	**3.E-06**	0.005	**0.019**			0.001	0.890			0.046	0.156			-0.001	**0.010**	0.165
0.118	**3.E-07**	-0.008	**7.E-05**	0.005	**0.036**					0.010	0.295	0.048	0.148	-5.E-04	0.871			0.148
0.086	**1.E-04**	-0.008	**9.E-06**	0.005	**0.017**					0.005	0.614	0.047	0.141			-0.001	**0.010**	0.166
0.121	**1.E-07**	-0.006	**1.E-04**			0.003	0.186	0.003	0.522			0.034	0.288	-0.001	0.672			0.141
0.091	**6.E-05**	-0.007	**2.E-05**			0.003	0.160	-0.001	0.900			0.028	0.368			-0.001	**0.020**	0.156
0.122	**1.E-07**	-0.006	**2.E-04**			0.003	0.153			0.006	0.512	0.032	0.313	-0.001	0.691			0.141
0.092	**5.E-05**	-0.007	**6.E-05**			0.003	0.161			0.001	0.921	0.029	0.347			-0.001	**0.021**	0.155
0.089	**7.E-05**	-0.007	**5.E-06**	0.004	**0.049**											-0.001	**0.006**	0.165

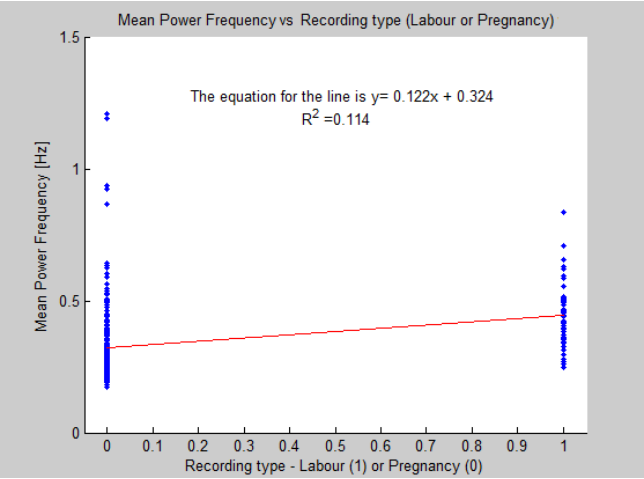

Figure 4.3: A plot of Mean Power Frequency vs Recording type (Labour (1) or Pregnancy (0)) for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 7E-10 (significant).

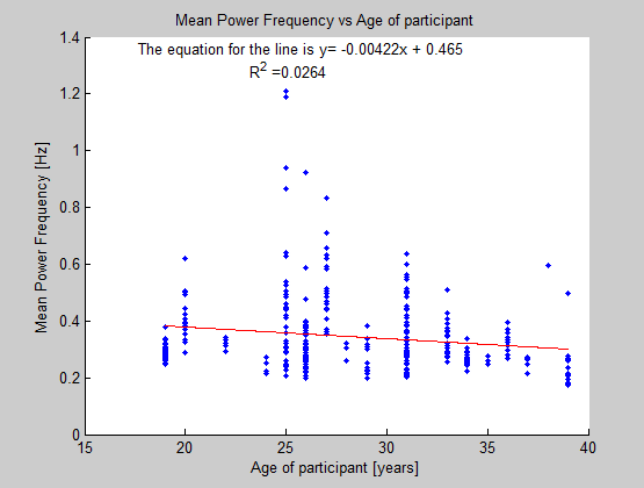

Figure 4.4: A plot of Mean Power Frequency vs Age of participant for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.004 (significant).

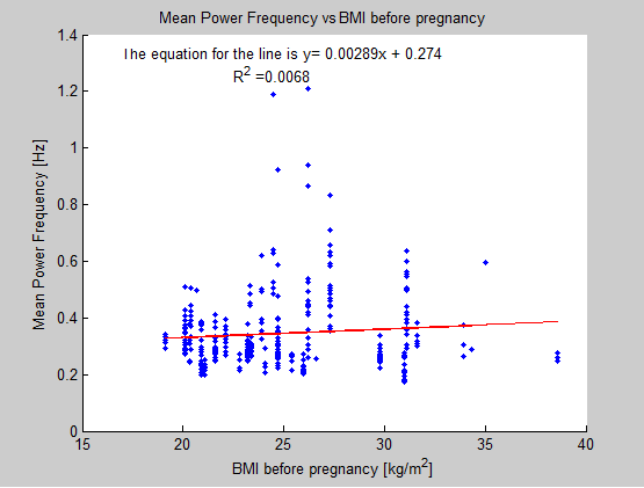

Figure 4.5: A plot of Mean Power Frequency vs BMI before pregnancy for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.143 (not significant).

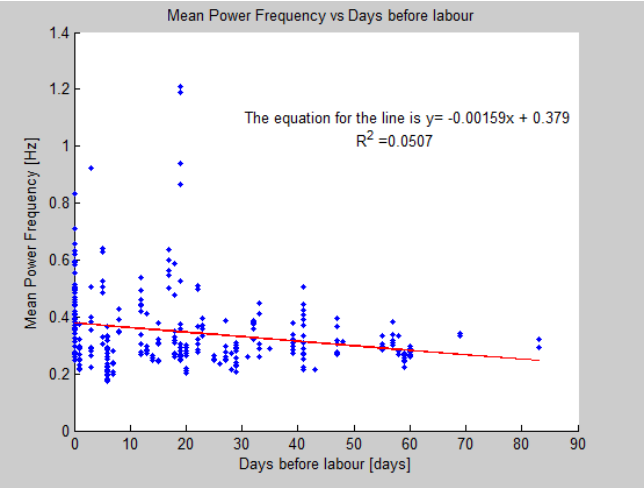

Figure 4.6: A plot of Mean Power Frequency vs Days before labour for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 5E-05 (significant).

4.2.1.2 Peak frequency

Results from the linear regression analyses with Peak frequency (PEAK) as the dependant variable can be seen in Table 4. Regression coefficients were statistically significant for three explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), BMI before pregnancy and Parity.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.099) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), BMI before pregnancy, BMI at recording, Gravidity, Parity and Days before labour were most frequently statistically significant in the multiple regression analyses of all explanatory variables. The model containing Recording type (Labour or Pregnancy), BMI before pregnancy, Parity and Days before labour had all regression coefficients statistically significant and an adjusted R^2 of 0.097. The plots of PEAK vs these parameters can be seen in Figures 4.7 – 4.10. The average PEAK from all contractions was 0.159 Hz. According to the model, PEAK is 0.029 Hz higher in labour compared to pregnancy, PEAK decreases by 0.002 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2), PEAK increases by 0.010 Hz for every unit increase in parity (range 0-3) and PEAK increases by 0.0003 Hz for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 4: Results from the linear regression analyses with Peak frequency as the dependant variable for 318 segmented contractions from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. The light dark grey row represents the model with the highest adjusted coefficient of determination when each regression coefficient is statistically significant. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.015	**0.007**																	0.023
		-4.E-04	0.323															0.003
				-0.002	**0.005**													0.024
						-0.001	0.073											0.010
								0.002	0.129									0.007
										0.007	**0.006**							0.024
												-0.001	0.939					2.E-05
														3.E-04	0.656			0.001
																-2.E-05	0.840	1.E-04
0.024	**2.E-04**	2.E-04	0.757	-0.002	**4.E-04**			0.002	0.090			-0.012	0.233	-0.001	0.181			0.060
0.027	**5.E-05**	3.E-04	0.568	-0.002	**2.E-04**			0.003	**0.035**			-0.008	0.394			3.E-04	**0.034**	0.069
0.025	**1.E-04**	-0.001	0.263	-0.002	**0.003**					0.010	**5.E-04**	-0.011	0.245	-0.001	0.175			0.088
0.028	**2.E-05**	-5.E-04	0.352	-0.002	**0.001**					0.011	**1.E-04**	-0.008	0.383			3.E-04	**0.019**	0.099
0.023	**6.E-04**	-2.E-04	0.652			-0.002	**0.005**	0.003	**0.024**			-0.004	0.663	-0.001	0.352			0.046
0.026	**1.E-04**	-1.E-04	0.814			-0.002	**0.003**	0.004	**0.010**			-0.001	0.902			2.E-04	0.085	0.052
0.024	**2.E-04**	-0.001	0.056			-0.002	**0.009**			0.012	**4.E-05**	-0.006	0.536	-0.001	0.331			0.082
0.028	**3.E-05**	-0.001	0.076			-0.002	**0.007**			0.013	**1.E-05**	-0.003	0.736			3.E-04	**0.047**	0.090
0.029	**1.E-05**			-0.002	**8.E-05**					0.010	**2.E-04**					3.E-04	**0.008**	0.097

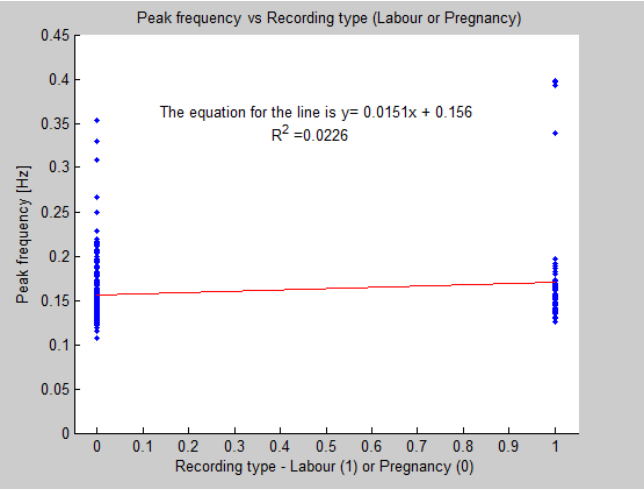

Figure 4.7: A plot of Peak frequency vs Recording type (Labour (1) or Pregnancy (0)) for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.007 (significant).

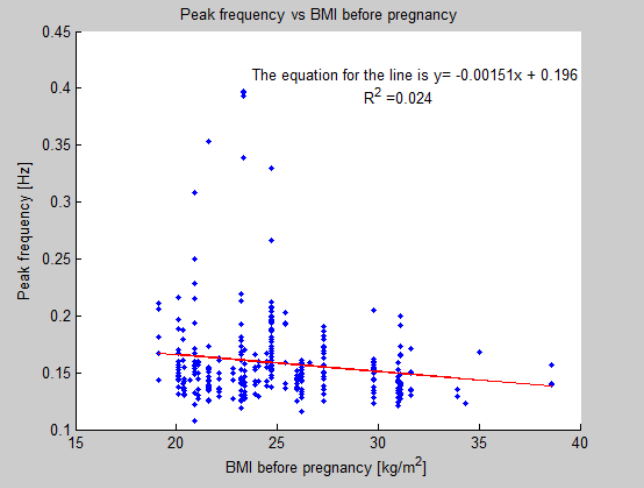

Figure 4.8: A plot of Peak frequency vs BMI before pregnancy for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.005 (significant).

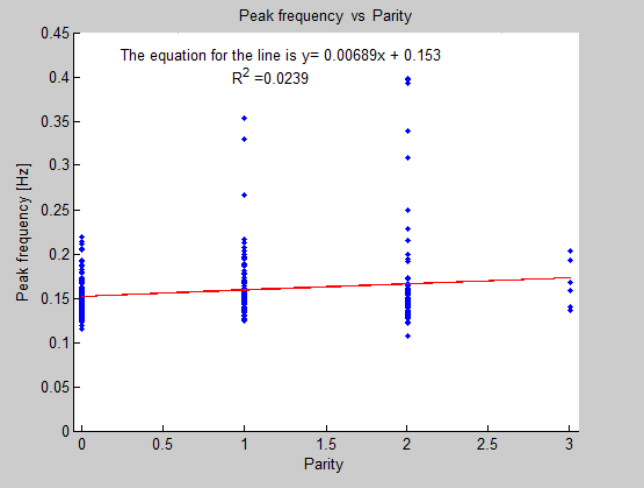

Figure 4.9: A plot of Peak frequency vs Parity for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.006 (significant).

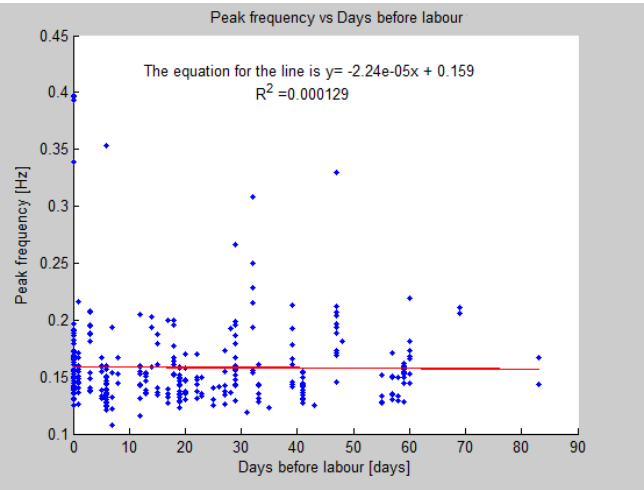

Figure 4.10: A plot of Peak frequency vs Days before labour for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.840 (not significant).

4.2.1.3 Median frequency

Results from the linear regression analyses with Median frequency (F50) as the dependant variable can be seen in Table 5. Regression coefficients were statistically significant for five explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, BMI at recording and Parity.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.167) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity, History of caesarean and Gestational age at recording.

Regression coefficients of Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, BMI at recording, Gravidity, Parity and Gestational age at recording were most frequently statistically significant in the multiple regression analyses of all explanatory variables. The model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity and Gestational age at recording had all regression coefficients statistically significant and the highest adjusted coefficient of determination of (R^2=0.168) of all models. The plots of F50 vs these parameters can be seen in Figures 4.11 – 4.15. The average F50 from all contractions was 0.209 Hz. According to the model, F50 is 0.034 Hz higher in labour compared to pregnancy, F50 decreases by 0.003 Hz for each yearly increase in age (age range of participants 19-39 years), F50 decreases by 0.002 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2), F50 increases by 0.016 Hz for every unit increase in parity (range 0-3) and F50 decreases by 0.002 Hz for every weekly increase in gestation age (range 29.71-41.29 weeks).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 5: Results from the linear regression analyses with Median frequency as the dependant variable for 318 segmented contractions from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.020	**0.004**																	0.025
		-0.002	**2.E-04**															0.043
				-0.003	**1.E-04**													0.047
						-0.002	**0.004**											0.026
								0.001	0.350									0.003
										0.006	**0.033**							0.014
												-0.004	0.691					0.001
														-4.E-04	0.684			0.001
																-3.E-05	0.828	1.E-04
0.035	**1.E-05**	-0.001	**0.042**	-0.003	**0.000**			0.003	**0.045**			-0.012	0.317	-0.002	**0.037**			0.118
0.032	**5.E-05**	-0.001	**0.044**	-0.003	**0.000**			0.003	**0.036**			-0.006	0.579			2.E-04	0.155	0.112
0.036	**3.E-06**	-0.002	**2.E-04**	-0.002	**0.002**					0.016	**3.E-06**	-0.010	0.361	-0.002	**0.029**			0.167
0.035	**8.E-06**	-0.002	**2.E-04**	-0.002	**0.002**					0.017	**2.E-06**	-0.005	0.647			3.E-04	0.066	0.163
0.033	**3.E-05**	-0.002	**0.004**			-0.002	**0.002**	0.004	**0.009**			-0.002	0.831	-0.002	0.099			0.105
0.031	**1.E-04**	-0.002	**0.004**			-0.003	**0.001**	0.004	**0.009**			0.002	0.890			2.E-04	0.300	0.100
0.035	**4.E-06**	-0.003	**3.E-06**			-0.002	**0.004**			0.018	**1.E-07**	-0.004	0.685	-0.002	0.076			0.164
0.034	**9.E-06**	-0.003	**3.E-06**			-0.002	**0.003**			0.018	**7.E-08**	-2.E-04	0.984			2.E-04	0.136	0.162
0.034	**5.E-06**	-0.003	**8.E-06**	-0.002	**0.003**					0.016	**2.E-06**			-0.002	**0.041**			0.168

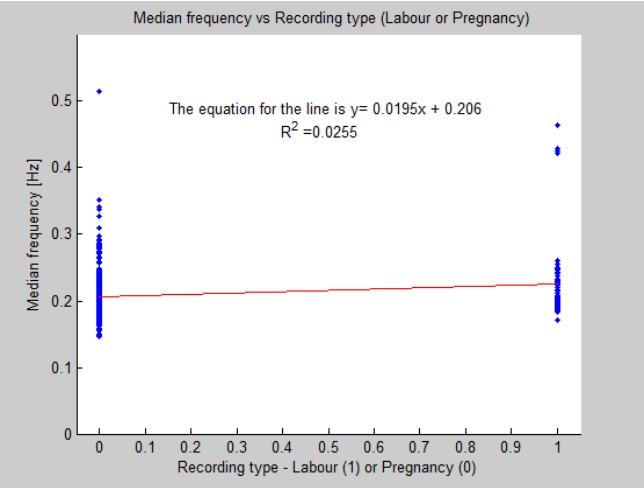

Figure 4.11: A plot of Median frequency vs Recording type (Labour (1) or Pregnancy (0)) for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.004 (significant).

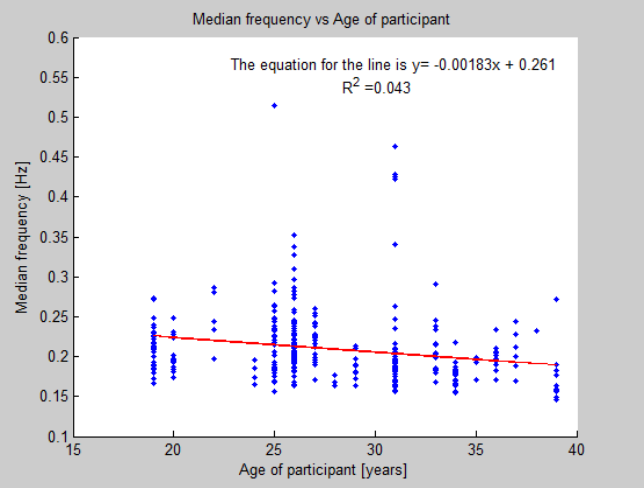

Figure 4.12: A plot of Median frequency vs Age of participant for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 2E-04 (significant).

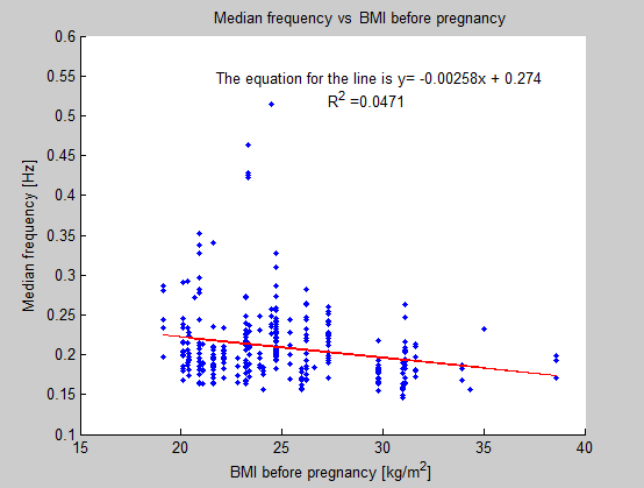

Figure 4.13: A plot of Median frequency vs BMI before pregnancy for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 1E-04 (significant).

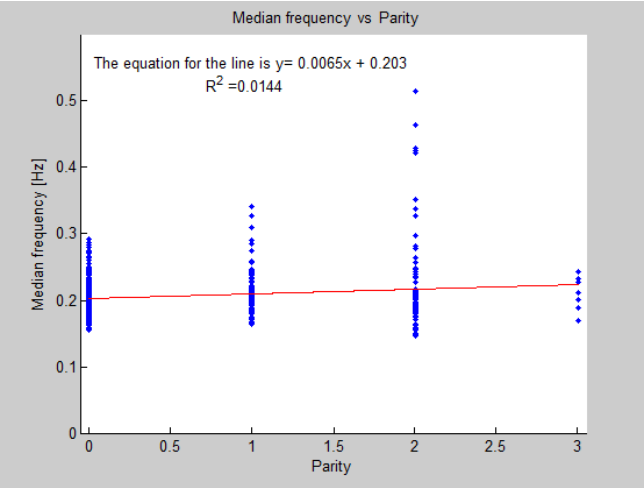

Figure 4.14: A plot of Median frequency vs Parity for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.033 (significant).

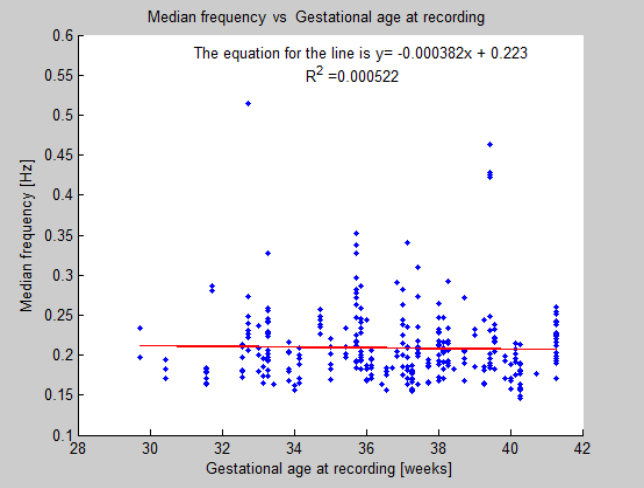

Figure 4.15: A plot of Median frequency vs Gestational age at recording for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.684 (not significant).

4.2.1.4 95%-limit frequency

Results from the linear regression analyses with 95%-limit frequency (F95) as the dependant variable can be seen in Table 6. Regression coefficients were statistically significant for five explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), BMI before pregnancy, BMI at recording, Gestational age at recording and Days before labour.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.128) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Gravidity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, BMI at recording and Days before labour were always statistically significant in the multiple regression analyses of all explanatory variables. The model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy and Days before labour had all regression coefficients statistically significant and the highest adjusted coefficient of determination (R^2=0.130) of all models. The plots of F95 vs these explanatory variables can be seen in Figures 4.16 – 4.19. The average F95 from all contractions was 1.050 Hz. According to the model, F95 is 0.351 Hz higher in labour compared to pregnancy, F95 decreases by 0.035 Hz for each yearly increase in age (age range of participants 19-39 years), F95 increases by 0.037 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2), and F95 decreases by 0.008 for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 6: Results from the linear regression analyses with 95%-limit frequency as the dependant variable for 318 segmented contractions from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.592	**2.E-07**																	0.083
		-0.014	0.084															0.009
				0.031	**0.006**													0.024
						0.032	**0.009**											0.022
								0.015	0.583									0.001
										-0.051	0.321							0.003
												-0.025	0.889					6.E-05
														0.035	**0.027**			0.015
																-0.009	**8.E-05**	0.048
0.511	**1.E-04**	-0.033	**0.001**	0.038	**0.003**			0.024	0.367			0.247	0.207	0.002	0.932			0.107
0.333	**0.011**	-0.039	**9.E-05**	0.043	**0.001**			0.002	0.938			0.214	0.262			-0.007	**0.006**	0.128
0.515	**1.E-04**	-0.034	**0.002**	0.039	**0.003**					0.036	0.528	0.229	0.239	0.003	0.874			0.106
0.333	**0.011**	-0.039	**4.E-04**	0.043	**0.001**					0.003	0.954	0.212	0.259			-0.007	**0.005**	0.128
0.546	**4.E-05**	-0.026	**0.005**			0.028	**0.038**	0.012	0.667			0.108	0.562	-0.004	0.815			0.093
0.379	**0.004**	-0.030	**0.001**			0.029	**0.030**	-0.009	0.745			0.069	0.707			-0.006	**0.018**	0.110
0.544	**5.E-05**	-0.025	**0.014**			0.029	**0.031**			0.003	0.960	0.098	0.597	-0.003	0.850			0.093
0.374	**0.005**	-0.028	**0.005**			0.028	**0.031**			-0.030	0.608	0.074	0.682			-0.006	**0.015**	0.110
0.351	**0.007**	-0.035	**1.E-04**	0.037	**0.002**											-0.008	**0.004**	0.130

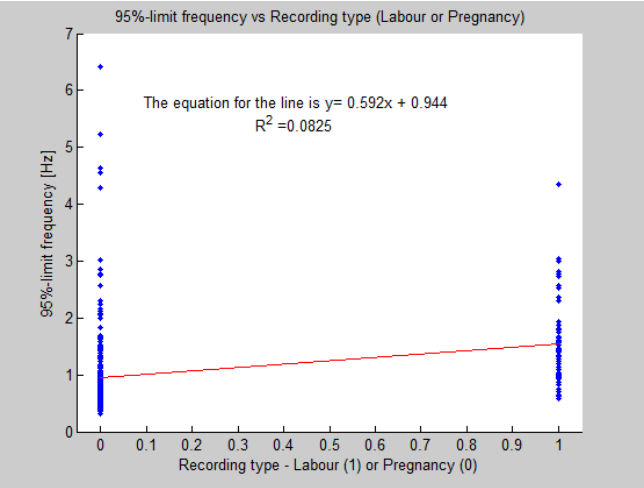

Figure 4.16: A plot of 95%-limit frequency vs Recording type (Labour (1) or Pregnancy (0)) for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 2E-07 (significant).

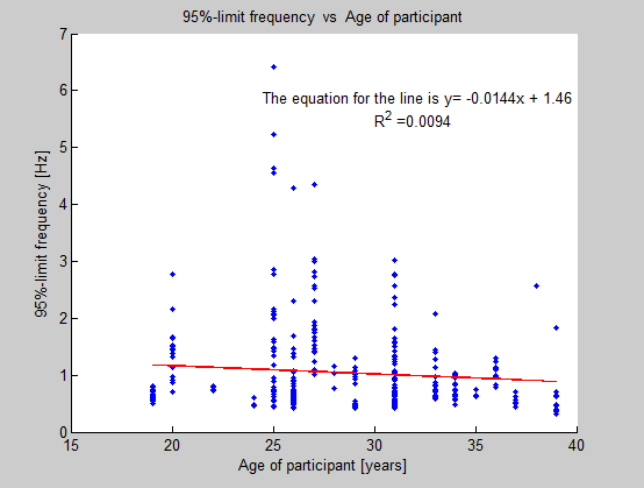

Figure 4.17: A plot of 95%-limit frequency vs Age of participant for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.084 (not significant).

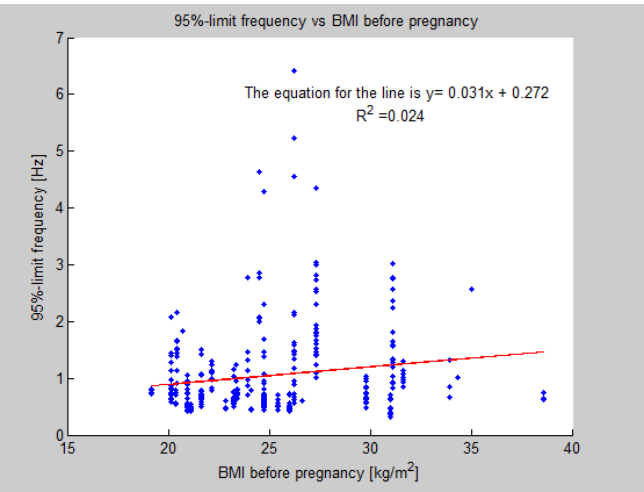

Figure 4.18: A plot of 95%-limit frequency vs BMI before pregnancy for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 0.006 (significant).

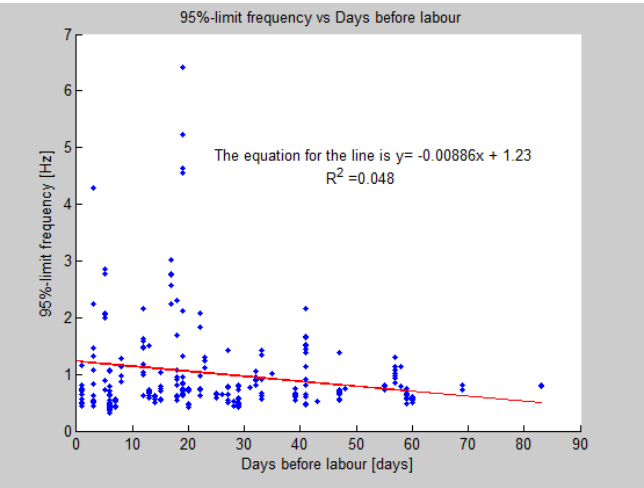

Figure 4.19: A plot of 95%-limit frequency vs Days before labour for 318 contractions from 31 participants during 68 16-electrode EHG recordings (9 labour recordings). The equation for the line and the coefficient of determination (R^2) is shown. The p-value for the simple regression analysis was 8E-05 (significant).

4.2.2 Contractions from all pregnancies ending with labour of spontaneous onset

For all pregnancies ending with labour of spontaneous onset, frequency analysis was performed on all 16 electrodes for 292 contractions from 25 participants during 59 recordings (8 labour recordings). A total of 53 contractions were labour contractions and 239 were pregnancy contractions. The averages of the frequency parameters from each of the 16 monopolar electrodes from a contraction were used for the linear regression analyses. The Pearson correlation coefficient matrix for all explanatory and dependant variables from these contractions can be seen in Table 7. Correlation coefficients higher than 0.7 are in bold font. This occurs, as expected, for BMI before pregnancy vs BMI at recording, Gravidity vs Parity, Gestational age vs Days before labour, MPF vs F95 and F50 vs PEAK.

Table 7: Pearson correlation coefficient matrix for 292 segmented contractions from pregnancies ending with labour of spontaneous onset from the Icelandic 16-electrode EHG database. Correlation coefficients higher than 0.7 are in bold font. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], MPF: Mean power frequency [Hz], PEAK: Peak frequency [Hz], F50: Median frequency [Hz], F95: 95%-limit frequency [Hz].

	Labour	Age	BMIb	BMIr	Gravidity	Parity	Caes	GA	Days	MPF	PEAK	F50	F95
Labour	**1.000**	0.049	0.270	0.347	0.079	-0.013	0.083	0.484	-0.493	0.409	0.149	0.194	0.340
Age	0.049	**1.000**	0.459	0.430	0.270	0.505	0.271	0.192	-0.331	-0.167	-0.037	-0.221	-0.097
BMIb	0.270	0.459	**1.000**	**0.924**	0.129	0.090	-0.270	0.167	-0.090	0.110	-0.156	-0.222	0.185
BMIr	0.347	0.430	**0.924**	**1.000**	0.291	0.242	-0.165	0.286	-0.233	0.114	-0.102	-0.170	0.169
Gravidity	0.079	0.270	0.129	0.291	**1.000**	**0.817**	-0.120	0.209	-0.337	0.010	0.084	0.054	0.029
Parity	-0.013	0.505	0.090	0.242	**0.817**	**1.000**	0.091	0.193	-0.303	-0.145	0.191	0.103	-0.131
Caes	0.083	0.271	-0.270	-0.165	-0.120	0.091	**1.000**	-0.095	-0.160	0.031	-0.001	-0.006	0.002
GA	0.484	0.192	0.167	0.286	0.209	0.193	-0.095	**1.000**	**-0.903**	0.184	0.004	-0.004	0.178
Days	-0.493	-0.331	-0.090	-0.233	-0.337	-0.303	-0.160	**-0.903**	**1.000**	-0.223	-0.001	0.013	-0.212
MPF	0.409	-0.167	0.110	0.114	0.010	-0.145	0.031	0.184	-0.223	**1.000**	0.103	0.409	**0.964**
PEAK	0.149	-0.037	-0.156	-0.102	0.084	0.191	-0.001	0.004	-0.001	0.103	**1.000**	**0.845**	-0.068
F50	0.194	-0.221	-0.222	-0.170	0.054	0.103	-0.006	-0.004	0.013	0.409	**0.845**	**1.000**	0.217
F95	0.340	-0.097	0.185	0.169	0.029	-0.131	0.002	0.178	-0.212	**0.964**	-0.068	0.217	**1.000**

4.2.2.1 Mean Power Frequency

Results from the linear regression analyses with Mean Power Frequency as the dependant variable can be seen in Table 8. Regression coefficients were statistically significant for five explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), Age of participant, Parity, Gestational age at recording and Days before labour.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.230) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, History of caesarean and Days before labour were frequently statistically significant in the multiple regression analyses of all explanatory variables. The model containing only these variables had all regression coefficients statistically significant and the highest adjusted coefficient of determination of (R^2=0.231) of all models. The average MPF from the analysed contractions was 0.341 Hz. According to the model, MPF is 0.089 Hz higher in labour compared to pregnancy, MPF decreases by 0.010 Hz for each yearly increase in age (age range of participants 19-39 years), MPF increases by 0.008 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2), MPF is 0.080 Hz higher if there is a history of caesarean and MPF decreases by 0.001 Hz for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 8: Results from the linear regression analyses with Mean Power Frequency as the dependant variable for 292 segmented contractions from pregnancies ending with labour of spontaneous onset from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.139	**3.E-13**																	0.167
		-0.004	**0.004**															0.028
				0.004	0.062													0.012
						0.004	0.053											0.013
								0.001	0.869									9.E-05
										-0.022	**0.013**							0.021
												0.016	0.593					0.001
														0.009	**0.002**			0.034
																-0.001	**1.E-04**	0.050
0.107	**3.E-06**	-0.009	**2.E-06**	0.007	**0.003**			0.005	0.281			0.091	**0.009**	0.003	0.278			0.215
0.089	**9.E-05**	-0.010	**2.E-07**	0.008	**0.001**			0.002	0.626			0.084	**0.013**			-0.001	**0.014**	0.229
0.109	**2.E-06**	-0.008	**1.E-04**	0.007	**0.007**					-0.004	0.658	0.080	**0.019**	0.004	0.223			0.213
0.088	**1.E-04**	-0.009	**1.E-05**	0.008	**0.002**					-0.008	0.434	0.076	**0.021**			-0.001	**0.006**	0.230
0.121	**1.E-07**	-0.007	**7.E-05**			0.004	0.129	0.002	0.613			0.051	0.105	0.002	0.553			0.197
0.106	**4.E-06**	-0.007	**2.E-05**			0.004	0.084	-9.E-05	0.985			0.044	0.154			-0.001	0.063	0.206
0.118	**3.E-07**	-0.006	**0.001**			0.004	0.102			-0.011	0.268	0.049	0.117	0.002	0.417			0.200
0.101	**1.E-05**	-0.006	**4.E-04**			0.004	0.067			-0.014	0.137	0.043	0.162			-0.001	**0.025**	0.212
0.089	**1.E-04**	-0.010	**2.E-07**	0.008	**0.001**							0.080	**0.014**			-0.001	**0.007**	0.231

4.2.2.2 Peak frequency

Results from the linear regression analyses with Peak frequency (PEAK) as the dependant variable can be seen in Table 9. Regression coefficients were statistically significant for three explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), BMI before pregnancy and Parity.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.136) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), BMI before pregnancy, BMI at recording, Gravidity, Parity, Gestational age at recording and Days before labour were frequently statistically significant in the multiple regression analyses of all explanatory variables. The model containing Recording type (Labour or Pregnancy), BMI before pregnancy, Parity and Days before labour had all regression coefficients statistically significant and an adjusted R^2 of 0.131. The average PEAK from the analysed contractions was 0.160 Hz. According to the model, PEAK is 0.034 Hz higher in labour compared to pregnancy, PEAK decreases by 0.003 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2), PEAK increases by 0.013 Hz for every unit increase in parity (range 0-3) and PEAK increases by 0.0005 Hz for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 9: Results from the linear regression analyses with Peak frequency as the dependant variable for 292 segmented contractions from pregnancies ending with labour of spontaneous onset from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. The light dark grey row represents the model with the highest adjusted coefficient of determination when each regression coefficient is statistically significant. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.015	**0.011**																	0.022
		-0.0003	0.531															0.001
				-0.002	**0.008**													0.024
						-0.001	0.081											0.010
								0.002	0.153									0.007
										0.009	**0.001**							0.037
												-2.E-04	0.983					2.E-06
														6.E-05	0.942			2.E-05
																-2.E-06	0.986	1.E-06
0.033	**1.E-05**	0.0012	**0.042**	-0.004	**1.E-05**			0.002	0.244			-0.027	**0.019**	-0.002	**0.025**			0.082
0.036	**2.E-06**	0.0015	**0.017**	-0.004	**4.E-06**			0.002	0.084			-0.021	0.052			4.E-04	**0.004**	0.093
0.034	**4.E-06**	2.E-04	0.742	-0.003	**2.E-04**					0.012	**1.E-04**	-0.024	**0.027**	-0.002	**0.010**			0.124
0.037	**6.E-07**	5.E-04	0.464	-0.003	**6.E-05**					0.013	**5.E-05**	-0.019	0.068			5.E-04	**0.001**	0.136
0.029	**1.E-04**	5.E-04	0.404			-0.003	**0.001**	0.003	**0.036**			-0.012	0.252	-0.002	0.120			0.055
0.032	**3.E-05**	6.E-04	0.255			-0.003	**5.E-04**	0.004	**0.011**			-0.007	0.468			3.E-04	**0.025**	0.064
0.032	**9.E-06**	-4.E-04	0.442			-0.003	**0.001**			0.015	**2.E-06**	-0.015	0.144	-0.002	**0.039**			0.115
0.035	**2.E-06**	-3.E-04	0.638			-0.003	**4.E-04**			0.016	**5.E-07**	-0.010	0.313			4.E-04	**0.007**	0.125
0.034	**1.E-06**			-0.003	**2.E-05**					0.013	**2.E-06**					5.E-04	**0.001**	0.131

4.2.2.3 Median frequency

Results from the linear regression analyses with Median frequency (F50) as the dependant variable can be seen in Table 10. Regression coefficients were statistically significant for four explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy and BMI at recording.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.191) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI at recording, Parity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, BMI at recording, Gravidity, Parity, Gestational age at recording and Days before labour were frequently statistically significant in the multiple regression analyses of all explanatory variables. The model containing Recording type (Labour or Pregnancy), Age of participant, BMI at recording, Parity and Days before labour had all regression coefficients statistically significant and the highest adjusted coefficient of determination of (R^2=0.193) of all models. The average F50 from all contractions was 0.209 Hz. According to the model, F50 is 0.043 Hz higher in labour compared to pregnancy, F50 decreases by 0.002 Hz for each yearly increase in age (age range of participants 19-39 years), F50 decreases by 0.003 Hz for each unit increase in BMI at recording (BMI range 22.9-42.2 kg/m^2), F50 increases by 0.018 Hz for every unit increase in parity (range 0-3) and F50 increases by 0.0004 Hz for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 10: Results from the linear regression analyses with Median frequency as the dependant variable for 292 segmented contractions from pregnancies ending with labour of spontaneous onset from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.022	**8.E-04**																	0.038
		-0.002	**1.E-04**															0.049
				-0.003	**1.E-04**													0.049
						-0.002	**0.003**											0.029
								0.001	0.357									0.003
										0.005	0.078							0.011
												-0.001	0.914					4.E-05
														-7.E-05	0.941			2.E-05
																3.E-05	0.826	2.E-04
0.040	**1E-06**	-7.E-04	0.309	-0.003	**1.E-04**			0.003	0.079			-0.017	0.184	-0.002	**0.040**			0.139
0.042	**3E-07**	-5.E-04	0.494	-0.004	**5.E-05**			0.003	**0.027**			-0.011	0.363			4.E-04	**0.012**	0.145
0.041	**2E-07**	-0.002	**0.010**	-0.003	**0.001**					0.014	**2.E-05**	-0.015	0.219	-0.002	**0.017**			0.183
0.044	**7E-08**	-0.002	**0.026**	-0.003	**0.001**					0.015	**9.E-06**	-0.009	0.416			4.E-04	**0.005**	0.189
0.037	**4E-06**	-0.001	**0.030**			-0.003	**0.001**	0.004	**0.009**			-0.004	0.737	-0.002	0.141			0.128
0.040	**1E-06**	-0.001	0.062			-0.003	**4.E-04**	0.005	**0.003**			0.001	0.943			3.E-04	**0.045**	0.133
0.041	**2E-07**	-0.002	**2.E-04**			-0.003	**0.001**			0.017	**3.E-07**	-0.007	0.488	-0.002	**0.048**			0.186
0.043	**7E-08**	-0.002	**6.E-04**			-0.003	**5.E-04**			0.018	**1.E-07**	-0.003	0.807			4.E-04	**0.017**	0.191
0.043	**5E-08**	-0.002	**2.E-04**			-0.003	**2.E-04**			0.018	**9.E-08**					4.E-04	**0.017**	0.193

4.2.2.4 95%-limit frequency

Results from the linear regression analyses with 95%-limit frequency (F95) as the dependant variable can be seen in Table 11. Regression coefficients were statistically significant for six explanatory variables in simple regression analysis: Recording type (Labour or Pregnancy), BMI before pregnancy, BMI at recording, Parity, Gestational age at recording and Days before labour.

In the multiple regression analyses of all explanatory variables, the highest adjusted coefficient of determination (R^2=0.1794) was from the model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, Parity, History of caesarean and Days before labour.

Regression coefficients of Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, BMI at recording, History of caesarean and Days before labour were frequently statistically significant in the multiple regression analyses of all explanatory variables. The model containing Recording type (Labour or Pregnancy), Age of participant, BMI before pregnancy, History of caesarean and Days before labour had all regression coefficients statistically significant and an adjusted R^2 of 0.1785. The average F95 from all contractions was 1.021 Hz. According to the model, F95 is 0.315 Hz higher in labour compared to pregnancy, F95 decreases by 0.053 Hz for each yearly increase in age (age range of participants 19-39 years), F95 increases by 0.065 Hz for each unit increase in BMI before pregnancy (BMI range 19.1-38.6 kg/m^2), F95 is 0.439 Hz higher if there is a history of caesarean and F95 decreases by 0.008 for every day increase in days before labour (range 0-83 days).

The condition indexes of the matrix of explanatory variables never went above 30 during regression collinearity diagnostic procedures.

Table 11: Results from the linear regression analyses with 95%-limit frequency as the dependant variable for 292 segmented contractions from pregnancies ending with labour of spontaneous onset from the Icelandic 16-electrode EHG database. Each row represents a linear regression analysis for the corresponding explanatory variables. Regression coefficients that are statistically significant (p-value <0.05) are in bold font. The dark grey row represents the model with the highest adjusted coefficient of determination. The light dark grey row represents the model with the highest adjusted coefficient of determination when each regression coefficient is statistically significant. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], RC: Regression coefficient, *p*: p-value, R^2: Coefficient of determination.

Labour		Age		BMIb		BMIr		Gravidity		Parity		Caes		GA		Days		
RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	RC	*p*	R^2
0.681	**2E-09**																	0.116
		-0.014	0.0989															0.009
				0.036	**0.001**													0.034
						0.036	**0.004**											0.029
								0.013	0.618									0.001
										-0.119	**0.025**							0.017
												0.007	0.968					5.E-06
														0.049	**0.002**			0.032
																-0.008	**3.E-04**	0.045
0.423	**0.002**	-0.048	**3E-05**	0.059	**8.E-05**			0.028	0.277			0.521	**0.015**	0.025	0.163			0.160
0.318	**0.022**	-0.054	**3E-06**	0.065	**1.E-05**			0.012	0.652			0.461	**0.024**			-0.008	**0.007**	0.176
0.432	**0.002**	-0.039	**0.002**	0.054	**3.E-04**					-0.046	0.432	0.447	**0.033**	0.028	0.118			0.159
0.307	**0.026**	-0.047	**2.E-04**	0.061	**5.E-05**					-0.067	0.252	0.404	**0.043**			-0.008	**0.002**	0.179
0.527	**1.E-04**	-0.031	**0.003**			0.034	**0.023**	0.008	0.771			0.219	0.261	0.013	0.458			0.129
0.436	**0.002**	-0.035	**0.001**			0.038	**0.012**	-0.007	0.808			0.168	0.378			-0.005	**0.049**	0.139
0.500	**3.E-04**	-0.023	**0.034**			0.036	**0.016**			-0.098	0.093	0.210	0.274	0.019	0.306			0.137
0.396	**0.005**	-0.028	**0.013**			0.039	**0.009**			-0.122	**0.039**	0.160	0.389			-0.006	**0.016**	0.152
0.315	**0.022**	-0.053	**3E-06**	0.065	**2.E-05**							0.439	**0.026**			-0.008	**0.003**	0.179

5 Discussion

The public access implementation of the database has been a success. The newly initiated collaboration with the research teams from three countries shows that data of this kind is in demand in the international scientific community. These three teams independently and spontaneously solicited the data prior to it being published on PhysioNet which further accentuates the need for data such as this to be available. The publication of the Data Descriptor (at Scientific Data or elsewhere) will present the database to an even wider audience. This could lead to novel EHG analysis and development towards new life-saving technology.

The WFDB MATLAB toolbox and online toolbox of the ATM viewer enable a wide range of analysis of the signals. The Icelandic 16-electrode EHG database therefore not only gives public access to the signal data but PhysioNet also provides tools to work with the data.

The aim was to publish the database as is, without giving the user any detailed directions and thus encouraging open-minded exploitation of the data. Making raw data available is the best way to do this. There are however some pointers about how to make sense of the data which is communicated to the user in the Data Descriptor:

- The data is sampled at 200 Hz. The EHG signals are generally assumed to be of very low frequency, from almost DC up to 3 Hz. Decimating the raw signal (after low pass filtering) is advised, before or after creating a bipolar/multipolar signal or other de-noising. This creates more manageable files and it is often better to work with signals were the difference in the frequency of the signal to be observed is not as far away from the sampling frequency as in this case. This will also cut down calculation time in any complex analysis. Results have shown that decimation of this signal has little or even a positive effect on the performance of methods[41].
- A user doing intensive work using these signals will have to develop an effective methodology to keep track of the signals, the way they have been pre-treated and the associated clinical parameters. Inspiration on how to organize such work can be sought in the structure of the SQL database framework developed by the Icelandic/French research team[69].
- Due to the constant monitoring of the researcher during recording, it is assumed that although there may be parts of the data where the contact is faulty between the skin and the electrode, they are few and moreover they are easily recognizable.

- The individual monopolar signals contain the measurement of the electric potential at each site in the matrix as referenced to an electrode far away over the iliac crest were little electrical activity is suspected. The raw signals therefore contain everything that creates a potential difference between the monopolar electrode and the reference. This includes the maternal electrocardiogram and some electromyogram signals from striated muscles. The signal also contains artefacts related to movements of the participant, foetal movement, foetal heart activity and even foetal respiration has been observed. Foetal hiccups can give mysterious periodic spikes that are clearly visible.

In this respect, the allowing of free conversation between participant and researcher in the protocol could have affected the signal quality. Breathing movement has been shown to move the uterus underneath the abdominal skin[70] and so conversing during the recording may have caused even more artefacts. The frequency of normal breathing is also at a similar frequency to contraction frequencies and so maternal breathing is a factor that should be taken into account in EHG studies.

There is another EHG database that can be found on PhsyioNet called "The Term-Preterm EHG Database"[71]. This contains EHG recordings performed in Slovenia with four electrodes. The position of the four electrodes are quite similar to the corner electrodes in the 16-electrode matrix and the Slovenian group also used AgCl electrodes and so comparison of recordings could be a possibility. On the whole, EHG research groups have not been using the same types of electrodes. This could be a problem as these different electrodes could have varying sensitivity to the low frequency signals. This question will be addressed in a BSc project in the spring of 2015 at the Faculty of Medicine at the University of Iceland. The author will be a co-supervisor for this project and will teach the required recording techniques to the student.

The frequency analysis showed some interesting results. In general, contractions are infrequent during pregnancy, and so there was no guarantee of measuring contractions during pregnancy recordings. Recording durations of one hour increased the chances of this but numerous recordings ended without a contraction.

There was high variation between electrodes from the same contraction. It is impossible to treat the skin equally under each electrode and this could in part explain this variability[11]. The tissue under the electrodes is also not homogenous. The electrodes in the matrix vary in distance from the median axis and this has been shown to influence the median and high parts of the EHG spectrum[26]. However, the variation could in part also be due to inherent differences in the

uterine muscle tissue. Regardless of the exact reason for this variation, it seemed advisable to use the average over all electrodes for the frequency analysis rather than using the spectrum from only one of the electrodes.

The condition index was higher than 30 in preliminary tests, with effect over 50% for GA and the intercept. Collinearity with the intercept can however be caused by near-constancy of other variables. In fact, by shifting GA by 28 (GA was then GA from 28 weeks) the condition index decreased below 30 without affecting the results of the multiple linear regression (as was expected). After this, the condition index never went above 30, probably due to the fact that highly correlated parameters were not included together in the same regression analysis.

Preliminary tests on placental position, comparing anterior position to other positions, showed no effect. However, positions such as lateral or anterior/high may or may not have been close to the electrodes. In future studies, the position of the electrodes with regard to placental position, could be recorded more accurately as studies have shown that placental position can have an effect on frequency parameters[26].

Table 12 shows the regression coefficients of the explanatory variables in the final multiple linear regression analyses for the different dependant variables of all segmented contractions and contractions from pregnancies ending with labour of spontaneous onset. All regression coefficients were statistically significant and the final models had the highest (or very close to) adjusted coefficients of determination.

Table 12: The explanatory variables in the final multiple linear regression models for the different dependant variables of all segmented contractions and contractions from pregnancies ending with labour of spontaneous onset. The values shown in the table are the regression coefficients. Labour: Recording type (Labour (1) or Pregnancy (0)), Age: Age of participant [years], BMIb: BMI before pregnancy [kg/m^2], BMIr: BMI at recording [kg/m^2], Caes: History of caesarean, GA: Gestational age at recording [weeks], Days: Days before labour [days], MPF: Mean power frequency [Hz], PEAK: Peak frequency [Hz], F50: Median frequency [Hz], F95: 95%-limit frequency [Hz].

	Regression coefficients - All segmented contractions				Regression coefficients - Contractions from pregnancies ending with labour of spontaneous onset			
	MPF	PEAK	F50	F95	MPF	PEAK	F50	F95
Labour	0.089	0.029	0.034	0.351	0.089	0.034	0.043	0.315
Age	-0.007		-0.003	-0.035	-0.010		-0.002	-0.053
BMIb	0.004	-0.002	-0.002	0.037	0.008	-0.003		0.065
BMIr							-0.003	
Gravidity								
Parity		0.010	0.016			0.013	0.018	
Caes					0.080			0.439
GA			-0.002					
Days	-0.001	0.0003		-0.008	-0.001	0.0005	0.0004	-0.008

The highest R^2 (0.231) was found in the model for MPF for contractions from pregnancies ending with labour of spontaneous onset, meaning that the model accounts for 23.1% of the variance in MPF. The plots of the frequency parameters vs explanatory variables show visually the high variability in the data, but even noisy, high-variability data can have a significant trend.

All of the frequency parameters were higher in labour compared to pregnancy. This has been shown before[22] and was not a surprising result. The differences in parameters between pregnancy and labour contractions were, however, quite small. There are many changes going on during labour, from the latent phase of labour to the active pushing phase. This leads to a high variability of any characteristic of labour contractions which makes it hard to individually distinguish them from pregnancy contractions. The partogram, used in the obstetric wards during labour, show the progression of cervical dilatation. Correlating labour contractions with partograms could show the short term changes in the contractility of the uterus during labour. This analysis could also enable a prediction of when a woman may be in need of augmentation. Another observation is that on the same tocograph, some contractions seem to peak more quickly and have a shorter duration than other contractions. Whether these are two fundamentally different types of contractions (or more) is not known. If they are, then this may need to be taken into account when creating a model to predict labour.

BMI before pregnancy (BMIb) was statistically significant in all models for the different frequency parameters and BMI at recording (BMIr) for all but MPF. The correlation coefficients were slightly higher for BMIb and the corresponding models had a higher adjusted coefficient of determination compared to BMIr (except for F50 in the contractions from pregnancies ending with labour of spontaneous onset). The larger distance between electrode and source as well as the filtering effects of fat could explain the effects of BMI. BMI before pregnancy would be expected to give a better indication of tissue thickness rather than BMI at recording. This is due to varying amounts of oedema, foetus size and amniotic fluid in pregnancy, which can affect BMI without changes in tissue thickness. BMI before pregnancy does however only give an indication of abdominal tissue thickness and has high individual variability. To study the possible effects of tissue thickness further, the next set of EHG recordings could involve an ultrasound of the tissue wall and in this way the tissue thickness and fat composition could be recorded directly instead of indirectly with the BMI. PEAK and F50 seemed to decrease with increasing BMI, but MPF and F95 seemed to increase. This seems to be contradictory and could mean that the effect of BMI is due to hidden correlative factors that have not been identified. However, there could be physiological explanations for this.

Power in higher frequencies have greater weight when using the MPF and thus it can be more sensitive to changes occurring in that portion of the spectrum. This could explain apparent opposite effects on MPF and median frequency[29]. Similar effects can occur on parameters evaluating the higher end of the spectrum such as F95, as there is a relative decrease in the higher frequencies due to the low pass filtering. Further study is needed to answer this question. Another explanation for these results could be that the frequency changes are due to an inherent difference in muscle activity between women with low vs high BMI. Studies have shown that women with higher BMI have a prolonged first stage of labour[72]. The results suggest that this should be studied further in future EHG research. The results also indicate that previous studies that did not take BMI into account may have over- or underestimated relationships between parameters and clinical aspects.

The linear regressions for the contractions from pregnancies ending with labour of spontaneous onset had higher coefficients of determination compared to all contractions. By only including contractions from the spontaneous labour group, the days before labour had a more accurate representation and this could explain this finding. Models containing Days before labour had a higher adjusted coefficient of determination compared to Gestational age at recording (except for F50 for all contractions). This was to be expected as Days gives a better representation of how close the uterus is to delivery. However, Days before labour is a variable that is calculated after delivery and therefore Gestational age must be used as a parameter in labour prediction. A decrease in any frequency parameter would be expected for increasing Days from labour[22]. This is however not the case for PEAK and F50 and could be explained by the fact that the changes are not linear, that there is a hidden correlative factor or the greater sensitivity of MPF and F95 to changes occurring in the high end of the spectrum[29]. A residual plot could help with assessing these possibilities[73].

Increasing age of participants seemed to have a decreasing effect on MPF, F50 and F95. There was no relationship to PEAK as was expected[3]. Age had no major correlations with parameters in the Pearson analysis, but did have a positive correlation of between 0.25 – 0.51 with all of the BMI variables and variables on obstetric history (Parity, Gravidity, and History of Caesarean). This could explain the findings in part but age could be an independent predictor. Tissues calcify and lose elasticity with age but it is not known whether this explains these effects or not. Despite the fact that the effect could be explained by a correlative factor, the results suggest that age should be studied more intensely in the future and might need to be considered when creating models to predict labour.

Increasing parity seems to increase PEAK and F50. Parity has previously been studied with regard to PEAK, but there was no observed effect[3]. Gravidity also seems to have an effect on PEAK and F50 but with an apparent lower effect on the adjusted coefficient of determination. It is well known that nulliparous women have slower labour progression than multiparous women[12]. One can imagine that after the first labour, the uterus has "gained experience", and is quicker to adapt in the following pregnancies. Gravidity could have the same effect, but with shorter pregnancies not having the same impact. Parity could possibly be used in a model for predicting labour. Gravidity could be used for the nulliparous woman, or both parameters could be used together in a model. The condition index was not over 30 in some models with both parity and gravidity but the high correlation between parameters calls for cautious interpretation.

History of caesarean was statistically significant in the models for MPF and F95 for contractions from pregnancies ending with labour of spontaneous onset. A scar in the muscle tissue could possibly have an effect on contractions, but another feasible explanation for the results could be that of the 20 contractions from participants with a history of caesarean, 18 were from the same woman. The analyses did not account fully for inter-participant differences and diminish the strength of the result.

All of the analysis for this thesis was done on monopolar signals as created by the CCA-EMD method. Most of the recent work on EHG has used bipolar signals and it would be interesting to repeat the analysis with the bipolar signals. The analysis in this thesis studied the relationship between clinical variables and frequency components of contractions. The next steps would involve "reversing" the models to see if frequency parameters along with clinical variables can predict imminent labour.

6 Conclusion

The Icelandic 16-electrode EHG database has been released and is available for public access at PhysioNet. A Data Descriptor paper is in review at the peer reviewed journal Scientific Data and will be published there or elsewhere in the coming months. Three research teams have already started analysing the database, which will hopefully pave the way towards life-saving medical device development. The main aim of this project was to implement public access to EHG data recorded in Iceland and this has been completed.

A study on the relationship between clinical variables and frequency components of contractions using data from the database shows that BMI, age and obstetric history (gravidity and parity) may have an effect on the frequency components. In particular, the median frequency seems to decrease with increasing age and BMI, but increase with increasing parity. The second aim of this project was to test the hypothesis that the BMI and obstetric history of a pregnant woman have an effect on the frequency components of contractions. The results suggest that this may be the case but further analysis is needed, in particular using other manners of de-noising the signals. This has not been shown before to the author's knowledge and demonstrates the use that can be made of the database. These findings could have an impact on the way EHG signals will be analysed in the future and on the development of medical devices for predicting preterm labour.

7 References

1. *Introduction to Biomedical Engineering*. (eds Enderle, J.D., Bronzino, J.D.) (Elsevier/Academic Press, 2012).
2. March of Dimes, PMNCH, Save the Children & WHO. *Born Too Soon: The Global Action Report on Preterm Birth*. (eds Howson, C.P., Kinney, M.V., Lawn, J.E.) (World Health Organization Geneva, 2012).
3. Garfield, R. E. & Maner, W. L. Physiology and electrical activity of uterine contractions. *Semin. Cell Dev. Biol.* **18,** 289–295 (2007).
4. Schlembach, D., Maner, W. L., Garfield, R. E. & Maul, H. Monitoring the progress of pregnancy and labor using electromyography. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **144 Suppl 1,** S33–39 (2009).
5. Lucovnik, M. *et al.* Use of uterine electromyography to diagnose term and preterm labor. *Acta Obstet. Gynecol. Scand.* **90,** 150–157 (2011).
6. Lucovnik, M. *et al.* Noninvasive uterine electromyography for prediction of preterm delivery. *Am. J. Obstet. Gynecol.* **204,** 228.e1–228.10 (2011).
7. Vinken, M. P. G. C., Rabotti, C., Mischi, M. & Oei, S. G. Accuracy of frequency-related parameters of the electrohysterogram for predicting preterm delivery: a review of the literature. *Obstet. Gynecol. Surv.* **64,** 529–541 (2009).
8. Hassan, M. *et al.* Better pregnancy monitoring using nonlinear correlation analysis of external uterine electromyography. *IEEE Trans. Biomed. Eng.* **60,** 1160–1166 (2013).
9. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101,** e215–e220 (2000).
10. Home : Scientific Data. Available at: http://www.nature.com/sdata/ [Accessed 6.1.15]
11. Devedeux, D., Marque, C., Mansour, S., Germain, G. & Duchêne, J. Uterine electromyography: a critical review. *Am. J. Obstet. Gynecol.* **169,** 1636–1653 (1993).

12. Impey, L. *Obstetrics and gynaecology*. (Blackwell Science, 1999).

13. Uterus: MedlinePlus Medical Encyclopedia Image. Available at: http://www.nlm.nih.gov/medlineplus/ency/imagepages/19263.htm [Accessed 6.1.15]

14. EULIANO, T. Y. *et al.* Monitoring uterine activity during labor: a comparison of three methods. *Am. J. Obstet. Gynecol.* **208,** 66.e1–66.e6 (2013).

15. Euliano, T. Y., Nguyen, M. T., Marossero, D. & Edwards, R. K. Monitoring contractions in obese parturients: electrohysterography compared with traditional monitoring. *Obstet. Gynecol.* **109,** 1136–1140 (2007).

16. Alberola-Rubio, J. *et al.* Comparison of non-invasive electrohysterographic recording techniques for monitoring uterine dynamics. *Med. Eng. Phys.* **35,** 1736–1743 (2013).

17. Monitoraggio Cardiotocografico CTG - GINECHO. Available at: http://www.ginecho.it/monitoraggio-cardiotocografico/ [Accessed 6.1.15]

18. How to read a CTG - Geeky Medics. Available at: http://geekymedics.com/2011/05/29/how-to-read-a-ctg/ [Accessed 6.1.15]

19. WHO | Preterm birth. *WHO*. Available at: http://www.who.int/mediacentre/factsheets/fs363/en/ [Accessed 6.1.15]

20. Karlsson, B., Terrien, J., Gudmundsson, V., Steingrimsdottir, T. & Marque, C. Abdominal EHG on a 4 by 4 grid: mapping and presenting the propagation of uterine contractions. in *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007* (eds. Jarm, T., Kramar, P. & Zupanic, A.) 139–143 (Springer Berlin Heidelberg, 2007).

21. Alamedine, D. *et al.* Selection algorithm for parameters to characterize uterine EHG signals for the detection of preterm labor. *Signal Image Video Process.* **8,** 1169–1178 (2014).

22. Moslem, B., Karlsson, B., Diab, M. O., Khalil, M. & Marque, C. Classification performance of the frequency-related parameters derived from uterine EMG signals. *Conf.*

Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. **2011,** 3371–3374 (2011).

23. Hassan, M., Terrien, J., Alexandersson, A., Marque, C. & Karlsson, B. Nonlinearity of EHG signals used to distinguish active labor from normal pregnancy contractions. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2010,** 2387–2390 (2010).

24. Diab, A., Hassan, M., Marque, C. & Karlsson, B. Performance analysis of four nonlinearity analysis methods using a model with variable complexity and application to uterine EMG signals. *Med. Eng. Phys.* **36,** 761–767 (2014).

25. Hassan, M., Terrien, J., Alexandersson, A., Marque, C. & Karlsson, B. Improving the classification rate of labor vs. normal pregnancy contractions by using EHG multichannel recordings. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2010,** 4642–4645 (2010).

26. Marque, C. K., Terrien, J., Rihana, S. & Germain, G. Preterm labour detection by use of a biophysical marker: the uterine electrical activity. *BMC Pregnancy Childbirth* **7,** S5 (2007).

27. Petrofsky, J. The effect of the subcutaneous fat on the transfer of current through skin and into muscle. *Med. Eng. Phys.* **30,** 1168–1176 (2008).

28. Stoykov, N. S., Lowery, M. M., Taflove, A. & Kuiken, T. A. Frequency- and time-domain FEM models of EMG: capacitive effects and aspects of dispersion. *IEEE Trans. Biomed. Eng.* **49,** 763–772 (2002).

29. Bilodeau, M. *et al.* Changes in the electromyographic spectrum power distribution caused by a progressive increase in the force level. *Eur. J. Appl. Physiol.* **71,** 113–123 (1995).

30. Baars, H., Jöllenbeck, T., Humburg, H. & Schröder, J. Surface-electromyography: skin and subcutaneous fat tissue attenuate amplitude and frequency parameters. *ISBS - Conf. Proc. Arch.* **1,** (2007).

31. Terrien, J., Marque, C., Steingrimsdottir, T. & Karlsson, B. Evaluation of adaptive filtering methods on a 16 electrode electrohysterogram recorded externally in labor. in *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007* (eds. Jarm, T., Kramar, P. & Zupanic, A.) 135–138 (Springer Berlin Heidelberg, 2007).

32. Hassan, M., Boudaoud, S., Terrien, J., Karlsson, B. & Marque, C. Combination of canonical correlation analysis and empirical mode decomposition applied to denoising the labor electrohysterogram. *IEEE Trans. Biomed. Eng.* **58,** 2441–2447 (2011).

33. Alexandersson, A., Terrien, J. & Karlsson, B. Multi-Electrode Placement Guide for Uterine Electromyography. Research project for 3rd year medical students. (University of Iceland, 2008).

34. Terrien, J., Hassan, M., Alexandersson, A., Marque, C. & Karlsson, B. Evolution of phase synchronization of the two frequency components of the electrohysterogram (EHG): application to the detection of human labor. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2010,** 17–20 (2010).

35. Hassan, M., Alexandersson, Á., Terrien, J., Karlsson, B. & Marque, C. Wavelet phase synchronization between EHGs at different uterine sites: comparison of pregnancy and labor contractions. in *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010* (eds. Bamidis, P. D. & Pallikarakis, N.) 21–24 (Springer Berlin Heidelberg, 2010).

36. Hassan, M., Terrien, J., Marque, C. & Karlsson, B. Comparison between approximate entropy, correntropy and time reversibility: application to uterine electromyogram signals. *Med. Eng. Phys.* **33,** 980–986 (2011).

37. Diab, A., Hassan, M., Marque, C. & Karlsson, B. Quantitative performance analysis of four methods of evaluating signal nonlinearity: application to uterine EMG signals. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2012,** 1045–1048 (2012).

38. Karlsson, B., Hassan, M. & Marque, C. Windowed multivariate autoregressive model improving classification of labor vs. pregnancy contractions. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2013,** 7444–7447 (2013).

39. Diab, A., Hassan, M., Boudaoud, S., Marque, C. & Karlsson, B. Nonlinear estimation of coupling and directionality between signals: Application to uterine EMG propagation. *Conf. Proc.* **2013,** 4366–4369 (2013).

40. Diab, A., Marque, C., Diab, A., Karlsson, B. & Hassan, M. Comparison of methods for evaluating signal synchronization and direction: Application to uterine EMG signals. in *2013 2nd International Conference on Advances in Biomedical Engineering (ICABME)* 14–17 (2013). doi:10.1109/ICABME.2013.6648835

41. Diab, A., Hassan, M., Karlsson, B. & Marque, C. Effect of decimation on the classification rate of non-linear analysis methods applied to uterine EMG signals. *IRBM* **34,** 326–329 (2013).

42. Diab, A., Hassan, M., Laforêt, J., Karlsson, B. & Marque, C. Estimation of coupling and directionality between signals applied to physiological uterine EMG model and real EHG signals. in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013* (ed. Romero, L. M. R.) 718–721 (Springer International Publishing, 2014).

43. Marque, C. *et al.* A multiscale model of the electrohysterogram the BioModUE_PTL project. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2013,** 7448–7451 (2013).

44. PhysioNet. Available at: http://physionet.org/ [Accessed 6.1.15]

45. About PhysioNet. Available at: http://physionet.org/resource.shtml [Accessed 6.1.15]

46. Data Sharing Services on PhysioNet. Available at: https://physionet.org/users/help/data-sharing/ [Accessed 6.1.15]

47. About : Scientific Data. Available at: http://www.nature.com/sdata/about [Accessed 6.1.15]

48. Cygwin. Available at: https://cygwin.com/ [Accessed 6.1.15]

49. The WFDB Software Package. Available at: http://physionet.org/physiotools/wfdb.shtml [Accessed 6.1.15]

50. SIGNAL(5). Available at: http://www.physionet.org/physiotools/wag/signal-5.htm [Accessed 6.1.15]

51. EDF2MIT(1). Available at: http://www.physionet.org/physiotools/wag/edf2mi-1.htm [Accessed 6.1.15]

52. HEADER(5). Available at: http://www.physionet.org/physiotools/wag/header-5.htm [Accessed 6.1.15]

53. PhysioBank Signals. Available at: http://www.physionet.org/physiobank/signals.shtml [Accessed 6.1.15]

54. The WFDB Software Package. Available at: http://physionet.org/physiotools/wfdb.shtml#WAVE [Accessed 6.1.15]

55. EDFbrowser. Available at: http://www.teuniz.net/edfbrowser/ [Accessed 6.1.15]

56. X11 resources for WAVE. Available at: http://www.physionet.org/physiotools/wug/node61.htm [Accessed 6.1.15]

57. WFDB2MAT(1). Available at: http://www.physionet.org/physiotools/wag/wfdb2m-1.htm [Accessed 6.1.15]

58. WFDB Toolbox for MATLAB and Octave. Available at: http://physionet.org/physiotools/matlab/wfdb-app-matlab/ [Accessed 6.1.15]

59. GIMP - The GNU Image Manipulation Program. Available at: http://www.gimp.org/ [Accessed 6.1.15]

60. Recommended Repositories : Scientific Data. Available at: http://www.nature.com/sdata/data-policies/repositories [Accessed 6.1.15]

61. CRAN - Package perturb. Available at: http://cran.r-project.org/web/packages/perturb/index.html [Accessed 6.1.15]

62. Belsley, D. A. *Regression diagnostics: identifying influential data and sources of collinearity*. (Wiley-Interscience, 2004).

63. Tabachnick, B. G. *Using multivariate statistics*. (HarperCollins College Publishers, 1996).

64. Icelandic 16-electrode EHG database. Available at: http://physionet.org/physiobank/database/ehgdb/ [Accessed 6.1.15]

65. LightWAVE 0.63. Available at: http://physionet.org/lightwave/ [Accessed 6.1.15]

66. PhysioBank ATM. Available at: http://www.physionet.org/cgi-bin/atm/ATM [Accessed 6.1.15]

67. WAVE(1). Available at: http://www.physionet.org/physiotools/wag/wave-1.htm [Accessed 6.1.15]

68. RDANN(1). Available at: http://www.physionet.org/physiotools/wag/rdann-1.htm [Accessed 6.1.15]

69. Terrien, J., Marque, C., Gondry, J., Steingrimsdottir, T. & Karlsson, B. Uterine electromyogram database and processing function interface: An open standard analysis platform for electrohysterogram signals. *Comput. Biol. Med.* **40,** 223–230 (2010).

70. De Lau, H., Rabotti, C., Haazen, N., Oei, S. G. & Mischi, M. Towards improving uterine electrical activity modeling and electrohysterography: ultrasonic quantification of uterine movements during labor. *Acta Obstet. Gynecol. Scand.* **92,** 1323–1326 (2013).

71. The Term-Preterm EHG Database (TPEHG DB). Available at: http://physionet.org/physiobank/database/tpehgdb/ [Accessed 6.1.15]

72. Bogaerts, A., Witters, I., Van den Bergh, B. R. H., Jans, G. & Devlieger, R. Obesity in pregnancy: Altered onset and progression of labour. *Midwifery* **29,** 1303–1313 (2013).

73. Bowerman, B. L. *Business statistics in practice*. (McGraw-Hill/Irwin, 2014).

8 Appendix

8.1 Letter to the Data Protection Authority

Persónuvernd
Rauðarárstíg 10
105 Reykjavík

Reykjavík, 26. ágúst 2014

Erindi: Tilkynning no S 696 (sjá annað viðhengi)

Meginatriði ofangreindrar rannsóknar er þróun vöðvarafritstækni (EMG) til að nota við hríðamælingar en í klínisku starfi eru þrýstingsmælar nú notaðir. Niðurstöður úr mælingum rannsóknarinnar og frá öðrum rannsóknarhópum úti í heimi gefa vísbendingar um að hægt verði að greina á milli hríða og tálhríða (þ.e. samdrátta sem ekki leiða til fæðingar) og þar með að spá fyrir um fyrirburafæðingar með frekari framþróun á þessari aðferð.

Á árunum 2008-2010 voru gerðar alls 122 mælingar á 45 þunguðum konum á Sjúkrahúsinu á Akureyri, Landspítala og Heilsugæslustöðunni á Akureyri. 112 mælingar voru gerðar á seinni þriðjungi meðgöngu og 10 mælingar voru gerðar í fæðingu.

Nú er ósk okkar að fá að birta þessi gögn í alþjóðlegum opnum gagnabanka fyrir lífeðlisfræðileg merki, gagnabanka PhysioNet: www.physionet.org . Þannig gæfust betri tækifæri til alþjóðlegs vísindalegs samstarfs og einnig möguleikar á mikilvægum vísindalegum uppgötvunum. Fyrirhugað nafn á gagnasafninu er „Icelandic 16-electrode EHG database“.

Gögn sem birtast á physionet.org fylgja reglum um HIPAA (*Health Insurance Portability and Accountability Act)* Privacy Rule og birtir gagnabankinn einungis óauðkennanleg gögn (de-identified data) sem ekki eru takmörkuð varðandi gagnadreifingu (samkvæmt HIPAA privacy rule). PhysioNet heimasíðan er í umsjá PhysioNet Resource, stofnað af vísindamönnum frá Beth Israel Deaconess Medical Center/Harvard Medical School, Boston Háskóla, McGill Háskóla og MIT háskóla. Gagnabankinn er rekinn af fjármagni frá National Institute of Biomedical Imaging and Bioengineering og National Institute of General Medical Sciences sem eru undir National Institutes of Health (http://physionet.org/resource.shtml). Aðalþjónn (server) fyrir PhysioNet er í MIT háskólanum.

Þau gögn sem fyrirhugað er að birta í gagnabankanum eru eftirfarandi:

- Mæligögn frá rafskautunum á .dat formi og. mat formi
- Merkingar (annotations) innan mælingar um hvenær samdrættir áttu sér stað, konan hreyfði sig, konan fann fyrir fósturhreyfingum og átt var við mælibúnað á .ann formi
- Skann af hríðariti (TOCO) frá mælingu á .jpg formi. Dagsetningar og önnur auðkennanleg gögn sjást ekki.
- Upplýsingar um konuna og mælinguna á .hea formi (textaskjal). Fyrirhugað er að birta:
    - Dulnefni konunar (á forminu ICE*** þar sem *** er tala)
    - Tegund mælingar (meðganga eða fæðing)
    - Aldur konunar við mælingu
    - BMI konunar fyrir meðgöngu
    - BMI konunar við mælingu
    - Fjölda fyrri meðganga og fæðinga
    - Hvort saga sé um fyrri keisara

- Staðsetningu fylgju
- Meðgöngulengd við mælingu
- Meðgöngulengd við fæðingu
- Fæðingarmáta
- Hvort Syntocinon, epidural eða belgjarof hafi verið notað í fæðingu og tímasetningu þessa miðað við fæðingu barns

Einnig mun koma fram að mælingarnar voru framkvæmdar á árunum 2008-2010 á Sjúkrahúsinu á Akureyri, Landspítala og Heilsugæslustöðunni á Akureyri.

Sem fyrr mun ábyrgðarmaðurinn einn hafa aðgang að listanum sem tengir dulnefni við kennitölur.

Rannsakendur:

Ásgeir Alexandersson, læknir og meistaranemi í heilbrigðisverkfræði við HR.

Brynjar Karlsson, prófessor, tækni- og verkfræðideild HR.

Þóra Steingrímsdóttir, prófessor í fæðinga- og kvensjúkdómalækningum, læknadeild HÍ.

Í hnotskurn:

Við förum fram á við Persónuvernd að fá að birta óauðkennanleg gögn frá rannsókninni „Legrafrit I - Þróun mæliaðferða“ í alþjóðlega opna gagnabankanum hjá PhysioNet.

Sambærilegt bréf er sent Vísindasiðanefnd í dag.

Fyrir hönd annarra rannsakenda,

með vinsemd og virðingu,

_____

Þóra Steingrímsdóttir, prófessor í fæðinga- og kvensjúkdómalækningum, læknadeild HÍ,

ábyrgðarmaður rannsóknar

thoraste@landspitali.is

s. 8633781

8.2 Letter to the National Bioethics Committee of Iceland

Vísindasiðanefnd
Hafnarhúsinu
Tryggvagötu 17
101 Reykjavík

Reykjavík, 26. ágúst 2014

Erindi: VSN nr. 02-006-V3: Legrafrit I. Þróun mæliaðferða.

Meginatriði ofangreindrar rannsóknar er þróun vöðvarafritstækni (EMG) til að nota við hríðamælingar en í klínisku starfi eru þrýstingsmælar nú notaðir. Niðurstöður úr mælingum rannsóknarinnar og frá öðrum rannsóknarhópum úti í heimi gefa vísbendingar um að hægt verði að greina á milli hríða og tálhríða (þ.e. samdrátta sem ekki leiða til fæðingar) og þar með að spá fyrir um fyrirburafæðingar með frekari framþróun á þessari aðferð.

Á árunum 2008-2010 voru gerðar alls 122 mælingar á 45 þunguðum konum á Sjúkrahúsinu á Akureyri, Landspítala og Heilsugæslustöðunni á Akureyri. 112 mælingar voru gerðar á seinni þriðjungi meðgöngu og 10 mælingar voru gerðar í fæðingu.

Nú er ósk okkar að fá að birta þessi gögn í alþjóðlegum opnum gagnabanka fyrir lífeðlisfræðileg merki, gagnabanka PhysioNet: www.physionet.org . Þannig gæfust betri tækifæri til alþjóðlegs vísindalegs samstarfs og einnig möguleikar á mikilvægum vísindalegum uppgötvunum. Fyrirhugað nafn á gagnasafninu er „Icelandic 16-electrode EHG database“.

Gögn sem birtast á physionet.org fylgja reglum um HIPAA (*Health Insurance Portability and Accountability Act)* Privacy Rule og birtir gagnabankinn einungis óauðkennanleg gögn (de-identified data) sem ekki eru takmörkuð varðandi gagnadreifingu (samkvæmt HIPAA privacy rule). PhysioNet heimasíðan er í umsjá PhysioNet Resource, stofnað af vísindamönnum frá Beth Israel Deaconess Medical Center/Harvard Medical School, Boston Háskóla, McGill Háskóla og MIT háskóla. Gagnabankinn er rekinn af fjármagni frá National Institute of Biomedical Imaging and Bioengineering og National Institute of General Medical Sciences sem eru undir National Institutes of Health (http://physionet.org/resource.shtml). Aðalþjónn (server) fyrir PhysioNet er í MIT háskólanum.

Þau gögn sem fyrirhugað er að birta í gagnabankanum eru eftirfarandi:

- Mæligögn frá rafskautunum á .dat formi og. mat formi
- Merkingar (annotations) innan mælingar um hvenær samdrættir áttu sér stað, konan hreyfði sig, konan fann fyrir fósturhreyfingum og átt var við mælibúnað á .ann formi
- Skann af hríðariti (TOCO) frá mælingu á .jpg formi. Dagsetningar og önnur auðkennanleg gögn sjást ekki.
- Upplýsingar um konuna og mælinguna á .hea formi (textaskjal). Fyrirhugað er að birta:
    - Dulnefni konunar (á forminu ICE*** þar sem *** er tala)
    - Tegund mælingar (meðganga eða fæðing)
    - Aldur konunar við mælingu
    - BMI konunar fyrir meðgöngu
    - BMI konunar við mælingu
    - Fjölda fyrri meðganga og fæðinga

- Hvort saga sé um fyrri keisara
- Staðsetningu fylgju
- Meðgöngulengd við mælingu
- Meðgöngulengd við fæðingu
- Fæðingarmáta
- Hvort Syntocinon, epidural eða belgjarof hafi verið notað í fæðingu og tímasetningu þessa miðað við fæðingu barns

Einnig mun koma fram að mælingarnar voru framkvæmdar á árunum 2008-2010 á Sjúkrahúsinu á Akureyri, Landspítala og Heilsugæslustöðunni á Akureyri.

Sem fyrr mun ábyrgðarmaðurinn einn hafa aðgang að listanum sem tengir dulnefni við kennitölur.

Rannsakendur:

Ásgeir Alexandersson, læknir og meistaranemi í heilbrigðisverkfræði við HR.

Brynjar Karlsson, prófessor, tækni- og verkfræðideild HR.

Þóra Steingrímsdóttir, prófessor í fæðinga- og kvensjúkdómalækningum, læknadeild HÍ.

Í hnotskurn:

Við förum fram á við Vísindasiðanefnd að fá að birta óauðkennanleg gögn frá rannsókninni „Legrafrit I - Þróun mæliaðferða“ í alþjóðlega opna gagnabankanum hjá PhysioNet.

Sambærilegt bréf er sent Persónuvernd í dag.

Fyrir hönd annarra rannsakenda,

með vinsemd og virðingu,

_____

Þóra Steingrímsdóttir, prófessor í fæðinga- og kvensjúkdómalækningum, læknadeild HÍ,

ábyrgðarmaður rannsóknar

thoraste@landspitali.is

s. 8633781

8.3 Codes for MATLAB and R

Frequency parameter calculation in MATLAB

```
function varargout = psd2(varargin)
%PSD2 Power Spectral Density and frequency estimates.
%  PSD2 estimates the power spectral density (PSD), mean power frequency (MPF),
%  peak frequency (PEAK), and limit frequency (F95) that contains up to 95%
%  of the PSD using Welch's averaged periodogram method (PWELCH Matlab function).
%  [MPF,PEAK,F50,F95,F,P]=psd2(X,Fs,NFFT,WINDOW,NOVERLAP)
%  [MPF,PEAK,F50,F95,F,P]=psd2(X,Fs) uses default values
%  Inputs:
%    X: data vector
%    Fs: sampling frequency
%    NFFT,WINDOW,NOVERLAP: see PWELCH matlab function for help
%      default values: NFFT=WINDOW=half the data length, NOVERLAP=50%
%  Outputs:
%    MPF: mean power frequency
%    PEAK: peak frequency (mode)
%    F50: median frequency of the power spectral density
%    F95: frequency limit that contains up to 95% of the power spectral density
%    F: frequency vector
%    P: PSD estimated vector
%  Example:
%
[MPF,PEAK,F50,F95,F,P]=psd2(randn(1,2000)+sin(2*pi*10*linspace(0,20,2000)),100);
%
%  Marcos Duarte  mduarte@usp.br 11oct1998
if nargin == 2
    x=varargin{1};
    fs=varargin{2};
    window = ceil(length(x)/2);
    nfft = window;
    noverlap = ceil(window/2);
elseif nargin==5
    x=varargin{1};
    fs=varargin{2};
    nfft=varargin{3};
    window=varargin{4};
    noverlap=varargin{5};
else
    error('Incorrect number of inputs')
end
%power spectral density:
[p,f] = pwelch(x,window,noverlap,nfft,fs);
%mpf:
mpf = trapz(f,f.*p)/trapz(f,p);
%peak:
[m,peak] = max(p);
peak = f(peak);
%50 and 95% of PSD:
area = cumtrapz(f,p);
f50 = find(area >= .50*area(end));
if ~isempty(f50)
    f50 = f(f50(1));
else
    f50 = 0;
end
f95 = find(area >= .95*area(end));
if ~isempty(f95)
    f95 = f(f95(1));
else
    f95 = 0;
end
```

```
varargout{1} = mpf;
varargout{2} = peak;
varargout{3} = f50;
varargout{4} = f95;
varargout{5} = f;
varargout{6} = p;
```

Pearson correlation analysis, multiple linear regression and regression collinearity diagnostic analysis in R

con=read.table("*filename*.txt",header=TRUE) %Read the data

cor(con, use="complete.obs", method="pearson") %Pearson correlation analysis

results=lm(*dependant variable*~*explanatory variable 1* + *explanatory variable 2* + *⋯*,data=con) %Multiple linear regression

summary(results) %Summary of the multiple linear regression

library(perturb) %Load perturb library

colldiag(results) % Regression collinearity diagnostic procedure

Simple linear regression and plotting of data in MATLAB

```
a=*explanatory variable* %x axis
b=*dependant variable* %y axis

figure(1);
hold on;
title('*explanatory variable* vs *dependant variable*');
xlabel('*explanatory variable* [units]'); %including units
ylabel('*dependant variable* [Hz]'); %including units


[p,s] = polyfit(a,b,1);    % p output from polyfit returns 2 coefficients
                             fitting r = a_1 * x + a_2
r = p(1) .* a + p(2); % compute a new vector r, used for the plotted line

plot(a,b,'.');% data points
plot(a, r,'r');% calculated line
r2=1 - s.normr^2 / norm(b-mean(b))^2 %calculation of r2
gtext(['The equation for the line is y= ',num2str(p(1),3),'x +
',num2str(p(2),3),10,'R^2 =',num2str(r2,3)])
```

8.4 Clinical information from each recording (information from the header files excluding comments) in the Icelandic 16-electrode EHG database along with recording durations and whether or not the recording has a corresponding annotation file. Table continues on next two pages.

Filename	ID	Record type	Record number	Age (years)	BMI before pregnancy	BMI at recording	Gravidity	Parity	Previous caesarean	Placental position	Gestational age at recording (w/d)	Gestational age at delivery (w/d)	Mode of delivery	Synthetic oxytocin use in labour	Epidural during labour	Duration of recording	Annotation file
ice001_l_1of1	ice001	labour	1	31	23.3	27.6	3	2	No	Fundus	39/3	39/3	Vaginal	No	No	00:08:20	No
ice002_p_1of3	ice002	pregnancy	1	38	20.7	25.9	4	1	No	Posterior	38/1	40/4	Vaginal	No	No	01:08:05	Yes
ice002_p_2of3	ice002	pregnancy	2	38	20.7	25.9	4	1	No	Posterior	39/1	40/4	Vaginal	No	No	01:02:10	Yes
ice002_p_3of3	ice002	pregnancy	3	38	20.7	26.4	4	1	No	Posterior	40/1	40/4	Vaginal	No	No	01:00:35	Yes
ice003_p_1of2	ice003	pregnancy	1	23	28.4	34.1	1	0	No	Anterior	37/6	40/0	Vaginal	No	No	01:01:10	Yes
ice003_p_2of2	ice003	pregnancy	2	23	28.4	34.6	1	0	No	Anterior	38/6	40/0	Vaginal	No	No	01:00:45	Yes
ice004_p_1of1	ice004	pregnancy	1	30	31.0	35.8	4	2	No	Posterior	36/5	41/3	Vaginal/Induction	No	No	01:06:10	Yes
ice005_p_1of3	ice005	pregnancy	1	28	29.0	31.2	2	0	No	Posterior	32/4	40/3	Vaginal	Yes	No	01:01:55	Yes
ice005_p_2of3	ice005	pregnancy	2	28	29.0	31.2	2	0	No	Posterior	34/6	40/3	Vaginal	Yes	No	00:59:20	Yes
ice005_p_3of3	ice005	pregnancy	3	28	29.0	31.6	2	0	No	Posterior	36/6	40/3	Vaginal	Yes	No	01:05:30	Yes
ice006_p_1of2	ice006	pregnancy	1	30	31.6	37.4	2	1	No	Posterior	36/1	37/6	Vaginal/Induction	Yes	No	01:02:10	Yes
ice006_p_2of2	ice006	pregnancy	2	30	31.6	37.6	2	1	No	Posterior	37/4	37/6	Vaginal/Induction	Yes	No	01:02:45	Yes
ice007_p_1of3	ice007	pregnancy	1	19	23.2	27.8	1	0	No	Posterior	32/5	41/2	Vaginal	No	No	01:06:40	Yes
ice007_p_2of3	ice007	pregnancy	2	19	23.2	28.1	1	0	No	Posterior	35/5	41/2	Vaginal	No	No	01:13:55	Yes
ice007_p_3of3	ice007	pregnancy	3	19	23.2	29.4	1	0	No	Posterior	37/5	41/2	Vaginal	No	No	01:01:50	Yes
ice008_p_1of4	ice008	pregnancy	1	26	20.9	25.3	5	2	No	Anterior	35/5	40/2	Vaginal	No	No	01:02:00	Yes
ice008_p_2of4	ice008	pregnancy	2	26	20.9	25.3	5	2	No	Anterior	37/6	40/2	Vaginal	No	No	01:00:55	Yes
ice008_p_3of4	ice008	pregnancy	3	26	20.9	25.0	5	2	No	Anterior	39/2	40/2	Vaginal	No	No	01:06:30	Yes
ice008_p_4of4	ice008	pregnancy	4	26	20.9	25.3	5	2	No	Anterior	40/1	40/2	Vaginal	No	No	01:05:30	Yes
ice009_p_1of2	ice009	pregnancy	1	24	22.8	25.3	2	1	No	Posterior	33/1	39/0	Vaginal	No	No	01:25:30	Yes
ice009_p_2of2	ice009	pregnancy	2	24	22.8	26.1	2	1	No	Posterior	38/1	39/0	Vaginal	No	No	00:49:35	Yes
ice010_p_1of3	ice010	pregnancy	1	22	19.1	23.0	1	0	No	Posterior	29/5	41/4	Vaginal	Yes	Yes	01:02:40	Yes
ice010_p_2of3	ice010	pregnancy	2	22	19.1	23.0	1	0	No	Posterior	31/5	41/4	Vaginal	Yes	Yes	01:02:00	Yes
ice010_p_3of3	ice010	pregnancy	3	22	19.1	23.8	1	0	No	Posterior	34/5	41/4	Vaginal	Yes	Yes	01:04:15	Yes
ice011_p_1of3	ice011	pregnancy	1	37	25.4	30.9	4	3	No	Anterior	36/0	40/0	Vaginal	Yes	No	01:06:10	Yes
ice011_p_2of3	ice011	pregnancy	2	37	25.4	31.1	4	3	No	Anterior	38/0	40/0	Vaginal	Yes	No	01:00:50	Yes
ice011_l_3of3	ice011	labour	3	37	25.4	31.8	4	3	No	Anterior	40/0	40/0	Vaginal	Yes	No	00:25:10	No
ice012_p_1of4	ice012	pregnancy	1	19	23.4	26.8	2	0	No	Posterior	33/0	40/6	Vaginal	No	Yes	01:01:05	Yes
ice012_p_2of4	ice012	pregnancy	2	19	23.4	26.8	2	0	No	Posterior	35/0	40/6	Vaginal	No	Yes	01:00:50	Yes
ice012_p_3of4	ice012	pregnancy	3	19	23.4	27.4	2	0	No	Posterior	37/0	40/6	Vaginal	No	Yes	01:00:55	Yes
ice012_p_4of4	ice012	pregnancy	4	19	23.4	27.8	2	0	No	Posterior	38/0	40/6	Vaginal	No	Yes	00:59:10	Yes
ice013_p_1of3	ice013	pregnancy	1	28	30.5	35.4	2	1	No	Posterior	34/4	40/4	Vaginal	No	No	01:02:20	Yes
ice013_p_2of3	ice013	pregnancy	2	28	30.5	36.0	2	1	No	Posterior	38/0	40/4	Vaginal	No	No	00:52:15	Yes
ice013_p_3of3	ice013	pregnancy	3	28	30.5	35.9	2	1	No	Posterior	39/0	40/4	Vaginal	No	No	01:03:40	Yes
ice014_p_1of3	ice014	pregnancy	1	33	28.1	32.2	3	2	No	Posterior	36/2	40/1	Vaginal	No	No	01:00:45	Yes
ice014_p_2of3	ice014	pregnancy	2	33	28.1	32.7	3	2	No	Posterior	38/2	40/1	Vaginal	No	No	01:02:35	Yes
ice014_p_3of3	ice014	pregnancy	3	33	28.1	32.8	3	2	No	Posterior	39/2	40/1	Vaginal	No	No	00:51:10	Yes
ice015_p_1of1	ice015	pregnancy	1	23	25.1	30.7	1	0	No	Anterior	36/4	38/3	Emergency caesarean due to other than slow progr	No	No	01:14:55	Yes
ice016_l_1of1	ice016	labour	1	20	23.9	30.5	3	0	No	Anterior	39/3	39/3	Vaginal	Yes	Yes	00:50:55	Yes
ice017_p_1of3	ice017	pregnancy	1	31	19.5	24.4	2	1	No	Anterior	36/5	42/1	Vaginal/Induction	Yes	No	00:46:50	Yes
ice017_p_2of3	ice017	pregnancy	2	31	19.5	25.2	2	1	No	Anterior	39/3	42/1	Vaginal/Induction	Yes	No	01:00:40	Yes
ice017_p_3of3	ice017	pregnancy	3	31	19.5	25.3	2	1	No	Anterior	40/6	42/1	Vaginal/Induction	Yes	No	00:19:10	Yes
ice018_p_1of1	ice018	pregnancy	1	37	24.1	28.9	4	2	Yes	Posterior	38/1	38/4	Elective caesarean	No	No	01:01:10	Yes
ice019_p_1of1	ice019	pregnancy	1	29	21.1	25.2	4	2	No	Fundus	40/2	41/1	Vaginal	No	No	00:55:15	No
ice020_l_1of1	ice020	labour	1	24	20.9	27.2	1	0	No	Posterior/Lateral left	41/5	41/5	Vaginal	Yes	Yes	01:04:10	No
ice021_p_1of3	ice021	pregnancy	1	27	25.4	30.4	1	0	No	Anterior	35/3	40/2	Vaginal	Yes	Yes	01:03:40	Yes
ice021_p_2of3	ice021	pregnancy	2	27	25.4	30.4	1	0	No	Anterior	36/6	40/2	Vaginal	Yes	Yes	01:03:00	Yes
ice021_p_3of3	ice021	pregnancy	3	27	25.4	30.4	1	0	No	Anterior	37/6	40/2	Vaginal	Yes	Yes	01:02:35	Yes

Filename	ID	Record type	Record number	Age (years)	BMI before pregnancy	BMI at recording	Gravidity	Parity	Previous caesarean	Placental position	Gestational age at recording (w/d)	Gestational age at delivery (w/d)	Mode of delivery	Synthetic oxytocin use in labour	Epidural during labour	Duration of recording	Annotation file
ice022_p_1of1	ice022	pregnancy	1	21	19.6	24.4	1	0	No	Anterior	34/2	38/0	Vaginal	No	No	01:08:10	Yes
ice023_p_1of1	ice023	pregnancy	1	29	31.6	33.5	2	1	No	Posterior	32/4	40/5	Vaginal	Yes	Yes	01:04:50	Yes
ice024_l_1of1	ice024	labour	1	35	38.6	42.2	2	1	No	Posterior	39/6	39/6	Vaginal	No	No	00:25:20	Yes
ice025_l_1of1	ice025	labour	1	27	27.3	30.8	2	0	No	Anterior	41/2	41/2	Vaginal	Yes	Yes	00:51:40	Yes
ice026_p_1of4	ice026	pregnancy	1	25	26.2	28.5	1	0	No	Anterior	35/0	39/5	Vaginal	Yes	No	00:52:45	Yes
ice026_p_2of4	ice026	pregnancy	2	25	26.2	28.6	1	0	No	Anterior	37/0	39/5	Vaginal	Yes	No	01:00:55	Yes
ice026_p_3of4	ice026	pregnancy	3	25	26.2	28.9	1	0	No	Anterior	38/0	39/5	Vaginal	Yes	No	01:00:50	Yes
ice026_p_4of4	ice026	pregnancy	4	25	26.2	29.0	1	0	No	Anterior	39/0	39/5	Vaginal	Yes	No	01:05:15	Yes
ice027_p_1of7	ice027	pregnancy	1	31	31.1	34.5	7	2	No	Anterior	29/5	38/1	Vaginal	Yes	No	00:56:45	No
ice027_p_2of7	ice027	pregnancy	2	31	31.1	35.0	7	2	No	Anterior	33/3	38/1	Vaginal	Yes	No	01:01:15	Yes
ice027_p_3of7	ice027	pregnancy	3	31	31.1	35.2	7	2	No	Anterior	34/5	38/1	Vaginal	Yes	No	01:00:25	No
ice027_p_4of7	ice027	pregnancy	4	31	31.1	35.5	7	2	No	Anterior	35/5	38/1	Vaginal	Yes	No	01:00:10	Yes
ice027_p_5of7	ice027	pregnancy	5	31	31.1	35.2	7	2	No	Anterior	36/5	38/1	Vaginal	Yes	No	01:00:25	Yes
ice027_p_6of7	ice027	pregnancy	6	31	31.1	35.5	7	2	No	Anterior	37/5	38/1	Vaginal	Yes	No	01:01:40	Yes
ice027_l_7of7	ice027	labour	7	31	31.1	35.5	7	2	No	Anterior	38/1	38/1	Vaginal	Yes	No	00:32:00	No
ice028_p_1of3	ice028	pregnancy	1	31	21.6	24.9	4	1	No	Posterior	34/1	38/0	Vaginal	No	No	01:03:30	Yes
ice028_p_2of3	ice028	pregnancy	2	31	21.6	25.3	4	1	No	Posterior	36/1	38/0	Vaginal	No	No	01:00:30	Yes
ice028_p_3of3	ice028	pregnancy	3	31	21.6	25.7	4	1	No	Posterior	37/1	38/0	Vaginal	No	No	01:06:20	Yes
ice029_p_1of2	ice029	pregnancy	1	25	24.5	28.3	3	2	No	Anterior/Superior	32/5	35/3	Vaginal/Induction	Yes	Yes	01:00:20	Yes
ice029_p_2of2	ice029	pregnancy	2	25	24.5	28.8	3	2	No	Anterior/Superior	34/5	35/3	Vaginal/Induction	Yes	Yes	01:00:40	Yes
ice030_p_1of4	ice030	pregnancy	1	33	20.1	25.6	2	1	Yes	Posterior/Superior	33/6	37/0	Vaginal	Yes	Yes	01:01:25	No
ice030_p_2of4	ice030	pregnancy	2	33	20.1	26.0	2	1	Yes	Posterior/Superior	35/6	37/0	Vaginal	Yes	Yes	01:04:25	Yes
ice030_p_3of4	ice030	pregnancy	3	33	20.1	26.4	2	1	Yes	Posterior/Superior	36/6	37/0	Vaginal	Yes	Yes	01:01:00	Yes
ice030_l_4of4	ice030	labour	4	33	20.1	26.4	2	1	Yes	Posterior/Superior	37/0	37/0	Vaginal	Yes	Yes	00:31:30	No
ice031_p_1of3	ice031	pregnancy	1	38	35.0	35.9	6	3	Yes	Anterior	36/1	39/0	Vaginal	Yes	No	01:00:15	Yes
ice031_p_2of3	ice031	pregnancy	2	38	35.0	35.6	6	3	Yes	Anterior	38/3	39/0	Vaginal	Yes	No	01:00:35	Yes
ice031_l_3of3	ice031	labour	3	38	35.0	35.8	6	3	Yes	Anterior	39/0	39/0	Vaginal	Yes	No	00:37:55	Yes
ice032_p_1of5	ice032	pregnancy	1	31	26.0	28.7	1	0	No	Posterior	31/6	40/1	Vaginal	No	No	00:34:10	Yes
ice032_p_2of5	ice032	pregnancy	2	31	26.0	28.7	1	0	No	Posterior	34/0	40/1	Vaginal	No	No	01:01:15	Yes
ice032_p_3of5	ice032	pregnancy	3	31	26.0	29.4	1	0	No	Posterior	36/0	40/1	Vaginal	No	No	01:07:00	Yes
ice032_p_4of5	ice032	pregnancy	4	31	26.0	29.8	1	0	No	Posterior	37/2	40/1	Vaginal	No	No	01:08:25	Yes
ice032_p_5of5	ice032	pregnancy	5	31	26.0	30.1	1	0	No	Posterior	40/0	40/1	Vaginal	No	No	01:00:35	Yes
ice033_p_1of1	ice033	pregnancy	1	20	20.4	24.1	3	0	No	Anterior	33/2	39/1	Vaginal	Yes	Yes	01:00:45	Yes
ice034_p_1of5	ice034	pregnancy	1	36	22.1	24.9	3	2	No	Anterior/High	30/3	38/5	Vaginal/Induction	Yes	Yes	01:01:45	Yes
ice034_p_2of5	ice034	pregnancy	2	36	22.1	25.2	3	2	No	Anterior/High	32/3	38/5	Vaginal/Induction	Yes	Yes	01:04:35	Yes
ice034_p_3of5	ice034	pregnancy	3	36	22.1	25.7	3	2	No	Anterior/High	35/3	38/5	Vaginal/Induction	Yes	Yes	01:03:20	Yes
ice034_p_4of5	ice034	pregnancy	4	36	22.1	25.8	3	2	No	Anterior/High	37/5	38/5	Vaginal/Induction	Yes	Yes	00:50:00	Yes
ice034_l_5of5	ice034	labour	5	36	22.1	25.8	3	2	No	Anterior/High	38/5	38/5	Vaginal/Induction	Yes	Yes	00:36:00	No
ice035_p_1of3	ice035	pregnancy	1	34	34.3	36.3	3	2	Yes	Anterior/High	32/0	39/0	Elective caesarean	No	No	01:00:25	Yes
ice035_p_2of3	ice035	pregnancy	2	34	34.3	36.5	3	2	Yes	Anterior/High	34/0	39/0	Elective caesarean	No	No	01:00:25	Yes
ice035_p_3of3	ice035	pregnancy	3	34	34.3	37.0	3	2	Yes	Anterior/High	36/0	39/0	Elective caesarean	No	No	01:00:15	Yes
ice036_p_1of6	ice036	pregnancy	1	39	31.0	35.5	5	2	No	Anterior/High	31/6	41/1	Emergency caesarean due to other than slow progr	No	No	01:00:30	Yes
ice036_p_2of6	ice036	pregnancy	2	39	31.0	36.2	5	2	No	Anterior/High	34/2	41/1	Emergency caesarean due to other than slow progr	No	No	00:58:40	Yes
ice036_p_3of6	ice036	pregnancy	3	39	31.0	35.5	5	2	No	Anterior/High	37/3	41/1	Emergency caesarean due to other than slow progr	No	No	01:04:55	Yes
ice036_p_4of6	ice036	pregnancy	4	39	31.0	35.8	5	2	No	Anterior/High	38/3	41/1	Emergency caesarean due to other than slow progr	No	No	01:03:15	Yes
ice036_p_5of6	ice036	pregnancy	5	39	31.0	35.8	5	2	No	Anterior/High	39/3	41/1	Emergency caesarean due to other than slow progr	No	No	00:43:50	Yes
ice036_p_6of6	ice036	pregnancy	6	39	31.0	35.8	5	2	No	Anterior/High	40/2	41/1	Emergency caesarean due to other than slow progr	No	No	01:02:10	Yes
ice037_p_1of2	ice037	pregnancy	1	33	26.6	31.9	2	1	Yes	Posterior/High	33/5	38/5	Emergency caesarean due to other than slow progr	No	No	01:03:35	Yes

Filename	ID	Record type	Record number	Age (years)	BMI before pregnancy	BMI at recording	Gravidity	Parity	Previous caesarean	Placental position	Gestational age at recording (w/d)	Gestational age at delivery (w/d)	Mode of delivery	Synthetic oxytocin use in labour	Epidural during labour	Duration of recording	Annotation file
ice037_p_2of2	ice037	pregnancy	2	33	26.6	32.6	2	1	Yes	Posterior/High	36/5	38/5	Emergency caesarean due to other than slow progr	No	No	00:59:20	Yes
ice038_p_1of1	ice038	pregnancy	1	29	28.4	32.2	4	1	Yes	Anterior/High	33/3	37/2	Elective caesarean	No	No	01:00:40	Yes
ice039_p_1of6	ice039	pregnancy	1	34	29.8	29.9	1	0	No	Anterior/High	31/4	40/0	Vaginal	No	No	01:02:20	Yes
ice039_p_2of6	ice039	pregnancy	2	34	29.8	29.6	1	0	No	Anterior/High	33/4	40/0	Vaginal	No	No	00:58:45	Yes
ice039_p_3of6	ice039	pregnancy	3	34	29.8	30.0	1	0	No	Anterior/High	35/5	40/0	Vaginal	No	No	01:01:50	Yes
ice039_p_4of6	ice039	pregnancy	4	34	29.8	30.1	1	0	No	Anterior/High	37/2	40/0	Vaginal	No	No	01:02:10	Yes
ice039_p_5of6	ice039	pregnancy	5	34	29.8	30.5	1	0	No	Anterior/High	38/2	40/0	Vaginal	No	No	01:02:30	Yes
ice039_p_6of6	ice039	pregnancy	6	34	29.8	30.3	1	0	No	Anterior/High	39/2	40/0	Vaginal	No	No	00:56:20	Yes
ice040_p_1of5	ice040	pregnancy	1	39	20.7	28.2	2	0	No	Posterior/High	36/5	41/6	Vaginal/Induction	Yes	No	01:03:00	Yes
ice040_p_2of5	ice040	pregnancy	2	39	20.7	28.9	2	0	No	Posterior/High	37/5	41/6	Vaginal/Induction	Yes	No	00:52:20	Yes
ice040_p_3of5	ice040	pregnancy	3	39	20.7	29.4	2	0	No	Posterior/High	38/5	41/6	Vaginal/Induction	Yes	No	01:00:35	Yes
ice040_p_4of5	ice040	pregnancy	4	39	20.7	29.7	2	0	No	Posterior/High	39/5	41/6	Vaginal/Induction	Yes	No	01:01:55	Yes
ice040_p_5of5	ice040	pregnancy	5	39	20.7	29.7	2	0	No	Posterior/High	40/6	41/6	Vaginal/Induction	Yes	No	01:01:20	Yes
ice041_p_1of2	ice041	pregnancy	1	26	33.9	37.6	3	1	No	Anterior/Right	36/0	38/5	Vaginal/Induction	No	No	01:02:30	Yes
ice041_p_2of2	ice041	pregnancy	2	26	33.9	37.9	3	1	No	Anterior/Right	36/5	38/5	Vaginal/Induction	No	No	01:03:05	Yes
ice042_p_1of2	ice042	pregnancy	1	25	20.3	22.6	2	0	No	Anterior/High	34/4	40/3	Vaginal	Yes	Yes	01:00:40	Yes
ice042_p_2of2	ice042	pregnancy	2	25	20.3	22.9	2	0	No	Anterior/High	38/2	40/3	Vaginal	Yes	Yes	00:46:55	Yes
ice043_p_1of2	ice043	pregnancy	1	25	24.1	26.2	2	0	No	Anterior	33/2	40/5	Emergency caesarean due to other than slow progr	No	No	01:06:15	No
ice043_p_2of2	ice043	pregnancy	2	25	24.1	26.9	2	0	No	Anterior	36/4	40/5	Emergency caesarean due to other than slow progr	No	No	01:01:15	Yes
ice044_p_1of3	ice044	pregnancy	1	28	23.2	28.4	1	0	No	Anterior/High	36/3	40/6	Vaginal	Yes	No	01:01:55	Yes
ice044_p_2of3	ice044	pregnancy	2	28	23.2	28.7	1	0	No	Anterior/High	38/5	40/6	Vaginal	Yes	No	01:04:10	Yes
ice044_p_3of3	ice044	pregnancy	3	28	23.2	29.1	1	0	No	Anterior/High	40/5	40/6	Vaginal	Yes	No	00:58:30	Yes
ice045_p_1of4	ice045	pregnancy	1	26	24.7	29.4	4	1	No	Anterior/High	33/2	40/0	Vaginal	No	No	01:06:45	Yes
ice045_p_2of4	ice045	pregnancy	2	26	24.7	29.8	4	1	No	Anterior/High	35/6	40/0	Vaginal	No	No	01:05:55	Yes
ice045_p_3of4	ice045	pregnancy	3	26	24.7	30.1	4	1	No	Anterior/High	37/3	40/0	Vaginal	No	No	01:07:15	Yes
ice045_p_4of4	ice045	pregnancy	4	26	24.7	30.5	4	1	No	Anterior/High	39/4	40/0	Vaginal	No	No	01:11:00	Yes

8.5 Submitted Data Descriptor to Scientific Data

Title

The Icelandic 16-electrode EHG database

Authors

Asgeir Alexandersson[1], Thora Steingrimsdottir[2], Jeremy Terrien[3], Catherine Marque[3], Brynjar Karlsson[1]

Affiliations

1. Reykjavik University, School of Science and Engineering, Reykjavik, Iceland
2. Landspitali University Hospital, Ob-Gyn department, Reykjavik, Iceland
3. Université de Technologie de Compiègne, Biomécanique et Bio-ingénierie, Compiègne, France

Corresponding author(s): Asgeir Alexandersson (alexandersson@gmail.com)

Abstract

External recordings of the electrohysterogram (EHG) can provide new knowledge on uterine electrical activity associated with contractions. Better understanding of the mechanisms underlying labor can contribute to preventing preterm birth which is the main cause of mortality and morbidity in newborns. Promising results using the EHG for labor prediction and other uses in obstetric care are the drivers of this work. This paper presents a database of 122 4-by-4 electrode EHG recordings performed on 45 pregnant women using a standardized recording protocol and a placement guide system. The recordings were performed in Iceland between 2008 and 2010. Of the 45 participants, 32 were measured repeatedly during the same pregnancy and participated in two to seven recordings. Recordings were performed in the third trimester (112 recordings) and during labor (10 recordings). The database includes simultaneously recorded tocographs, annotations of events and obstetric information on participants. The publication of this database enables independent and novel analysis of multi-electrode EHG by the researchers in the field and hopefully development towards new life-saving technology.

Background & Summary

Preterm birth is defined as birth before 37 completed weeks of gestation. On average, 12% of babies are born preterm in lower-income countries and 9% in higher-income countries. Babies who are born preterm often require special care and face greater risks of serious health problems, including cerebral palsy, intellectual impairment, chronic lung disease, and vision and hearing loss. Preterm birth is the leading cause of newborn deaths (babies in the first four weeks of life) and the second leading cause of death in children under five years (after pneumonia). Preterm birth rates are increasing in almost all countries[1].

Currently, there is no effective way of preventing preterm birth. The main reason is that no good objective method currently exists to evaluate the stepwise progression of pregnancy through to labor, neither at term nor preterm. Various techniques have been adopted to monitor and/or diagnose labor, but they are either subjective or indirect and do not provide an accurate prediction of when labor will take place[2–5]. Studies have shown that external monitoring of uterine electrical activity using an electrohysterogram (EHG) is representative of uterine contractility[6] and show promising results in predicting preterm labor[3–5,7,8]. Most of the early studies on EHG used two to five electrodes[2,6] and therefore concentrated on the activity of the uterus near the location of the electrodes. In 2007, a collaborative group from France and Iceland, involving biomedical researchers, engineers and medical doctors, started using monopolar electrodes in a 4-by-4 configuration on the abdomen aimed at

providing information on uterine electrical activity propagation[9]. Guided by this preliminary work, the 16 electrode system was used to perform pregnancy and labor recordings in Iceland at Landspitali University Hospital, Akureyri Hospital and the Akureyri Primary Health Care Centre. In total, a database of 122 EHG recordings was created between 2008 and 2010. The majority of participants were measured multiple times during the same pregnancy and took part in two to seven recordings. These multiple, or longitudinal, measurements were aimed at observing the evolution of contractions during pregnancy and towards labor.

Parts of the data have already been used for developing and analyzing various signal processing methods and have led to several publications[8,10–23]. In particular the work has concentrated on efforts to accurately distinguish true labor contractions from normal pregnancy contractions, with some success[8,10–17,20–23]. Compared to linear methods, non-linear methods may provide a superior way to differentiate between pregnancy and labor contractions[8,11,22] and multichannel recordings seem to improve this classification rate[10]. Efforts have also been made to exploit the information on the propagation and/or synchronization of different locations of the uterus which the extended geometry of the measurement system provides. The results indicate that propagation parameters can be important in accurately recognizing true labor[24,25]. Parts of the data were included in the BioModUE_PTL project which led to the development and validation of a biophysics based multiscale model of the EHG, going from the cell to the electrical signal measured on the abdomen[26]. None of these past publications, however, have described the database in detail and studies have so far only used parts of the data presented.

We provide open access to the database so that the international scientific community can freely generate greater understanding of the mechanics of the uterus and develop applications that improve obstetric care and hopefully accurately predict preterm labor. This paper describes the recording methods used and gives a detailed description of the Icelandic 16-electrode EHG database.

Methods

Data collection

The recordings were performed between 2008 and 2010 in Iceland. Pregnancy recordings, defined as recordings performed at antenatal care clinics on participants in the third trimester and not suspected to be in labor, were performed at Akureyri Primary Health Care Centre and Landspitali University Hospital. Labor recordings, defined as measurements performed on participants suspected to be in labor, present in the labor wards and who delivered within 24 hours, were performed at Landspitali University Hospital and Akureyri Hospital. Participants were invited to take part in the recordings during antenatal care visits or at the labor wards and had normal singleton pregnancies and no known risk factors for preterm birth. Informed consent was obtained from every participant and the protocol was approved by the National Bioethics Committee in Iceland (VSN 02-006-V4). After each pregnancy recording, the participant was invited to take part in another recording one to two weeks later. All data can be found at Data Citation 1: PhysioBank www.physionet.org/physiobank/database/ehgdb.

Recording protocol

Reusable Ag/AgCl electrodes with a 13.0 mm outer diameter and an 8.0 mm inner diameter were used for the recordings. An alignment frame, a double sided hypoallergenic adhesive sheet and a silicone backing were designed and manufactured to enable a standardized electrode setup with a 17.5 mm distance between adjacent electrode centers. The dimensions of the double sided adhesive sheet and silicone backing can be seen in Fig.1.

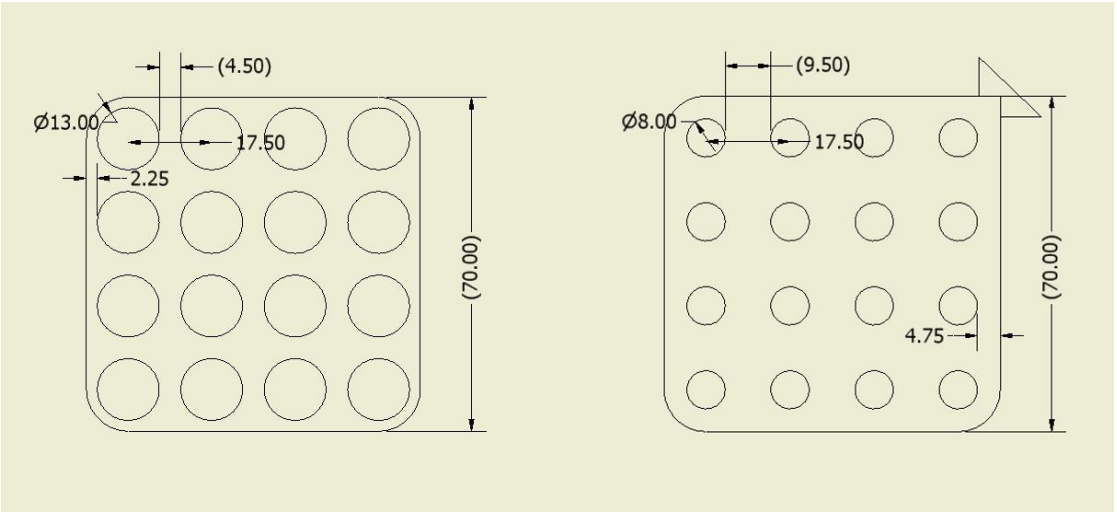

Figure 1: Dimensions (mm) of the silicone backing (left) and double sided adhesive sheet (right). The thickness of the silicone backing is 1.5mm.

The alignment frame was used to align and attach an uncovered side of the double sided adhesive sheet to the silicone backing. The electrodes were then placed into the holes in the silicone backing and attached to the adhesive sheet. The electrode numbering scheme, as seen when looking at the abdomen of the participant, can be seen in Fig.2.

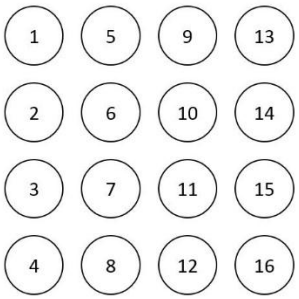

Figure 2: The electrode numbering scheme, as seen when looking at the abdomen of the participant.

The abdominal skin was carefully prepared using an abrasive paste and alcohol solution. After filling the electrode holes with electrode gel, wiping the excess gel away with a straight edged card and uncovering the other side of the double sided adhesive sheet, the electrode-adhesive-silicone matrix was attached to the abdomen.

The desired position on the abdomen was with the third vertical line of electrodes (electrodes 9 to 12) placed on the median axis of the uterus and the 10th – 11th pair of electrodes half way between the uterine fundus and pubic symphysis. The navel was avoided by displacing the matrix up or down whilst staying as close as possible to the desired position. The skin over the iliac crests on both sides was prepared in the same way as the abdomen and reference electrodes (patient ground and reference) with electrode gel were then attached on each side using adhesive washers, with inner diameters corresponding to the inner diameter of the electrodes. The electrode positions can be seen in Fig.3.

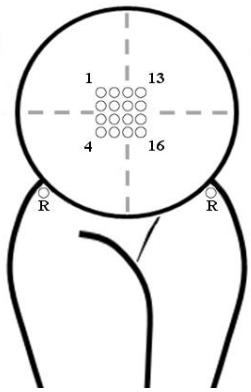

Figure 3: Ideal position of the 4-by-4 electrode grid. The third vertical line of electrodes (electrodes 9 to 12) is placed on the median axis of the uterus and the 10th – 11th pair of electrodes half way between the uterine fundus and pubic symphysis. The corner electrodes of the 4-by-4 grid are labelled along with the reference electrodes, (R).

A tocodynamometer was also attached to the abdomen during recordings. When possible, a support, such as a pillow, was positioned under the right side of the participant to prevent potential aortocaval compression syndrome. A photo of the setup during a recording can be seen in Fig. 4.

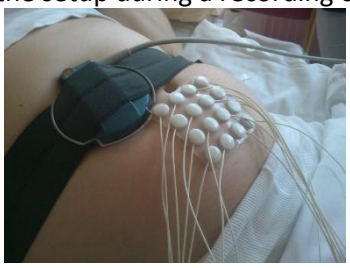

Figure 4: The recording setup. The abdominal electrodes and the tocodynamometer can be seen.

The intended duration of a pregnancy recording was one hour and the intended duration of a labor recording was at least half an hour but the participant could stop the measurement at any time.

The measurements were performed using a sixteen channel multi-purpose physiological signal recorder (Embla A10), most commonly used for investigating sleep disorders. An anti-aliasing filter with a high cut-off frequency of 100 Hz was used but no high pass filter was used. The signal sampling rate was 200 Hz and the signal was digitized to 16 bits. The sixteen monopolar electrode signals were originally stored in the EDF (European Data Format) format by the Somnologica software used to control the Embla A10.

All the recordings were performed by the same person (AA). Each participant was assigned an ID number and for each recording, information on participant age, BMI, gestational age, placental position, gravidity, parity, history of cesarean section, eventual mode of delivery and gestational age at delivery was noted. During a recording, the researcher recorded participant movements, equipment manipulation, participant-reported possible contractions and fetal movements, and any other unusual events. The participant and researcher conversed freely during the recordings and no restrictions were placed on the participant in terms of changing position if needed.

Data processing

No processing of data was performed beyond converting between file formats. The aim is that those that use the data can do so with a fresh start. Future additions to this database may include pre-treated signals and segmented contractions.

The EDF files obtained during the recordings were converted into WFDB (WaveForm DataBase - www.physionet.org/physiotools/wfdb.shtml) compatible files using Cygwin software (cygwin.com). Signal (.dat) and header (.hea) files were generated using the edf2mit WFDB application (www.physionet.org/physiotools/wag/edf2mi-1.htm). Annotation files containing events during recordings were created by manually adding to the WFDB-compatible signal files using WAVE, an X Window System client application (www.physionet.org/physiotools/wfdb.shtml#WAVE). The wfdb2mat WFDB application was used to convert the WFDB-compatible signal files to MATLAB .mat files (www.physionet.org/physiotools/wag/wfdb2m-1.htm). The rdann function from the WFDB toolbox for MATLAB (www.physionet.org/physiotools/matlab/wfdb-app-matlab) was used to create .mat files containing annotation information. The tocograph paper traces were scanned to JPEG images and a time axis corresponding to the recording was inserted onto the scanned images. All information that could possibly lead to the identification of the participant, such as personal data and dates, was manually removed from these images using a graphics editor.

Data Records

The data records in the Icelandic 16-electrode EHG database are stored in a PhysioBank database at the PhysioNet[27] web site (see Data Citation 1 - PhysioBank www.physionet.org/physiobank/database/ehgdb).

A total of 122 recordings were performed on 45 participants. Of the 45 participants, 32 were measured more than once during the same pregnancy and the highest number of recordings for a participant was seven recordings. Ten recordings were performed during labor and five participants took part in measurements during both pregnancy and labor. The lowest gestational age was twenty nine weeks and 5 days (29w5d - pregnancy recording) and the highest gestational age was forty one weeks and five days (41w5d - labor recording). The average recording duration for pregnancy recordings was 61 minutes (range 19-86 minutes) and the average recording duration for labor recordings was 36 minutes (range 8-64 minutes).

File names in the database are of the form ice###_*type*_*record number*, where ice### is the ID of the participant (e.g. ice001), *type* refers to the type of recording: - p (pregnancy) or l (labor), and *record number* is the number of recording for that particular participant (e.g. 1of3).

Each recording has three associated files:

-A binary signal file containing the data from the 16 monopolar electrodes (.dat file)

-A scanned tocograph with a manually inserted recording time axis (.jgp file). Each small square represents 30 seconds.

-A header file (.hea file):

The fields in the header files (.hea) are according to the WFDB convention and are listed in detail on the PhysioNet website (www.physionet.org/physiotools/wag/header-5.htm). Lines 2-17 of the header files are signal specification lines and the strings at the end of these lines correspond to the signal labels. The signals are labelled with "EHGn" where "n" refers to the relevant electrode number (electrode numbers are not in ascending order).

Information from each recording is at the end of the header files and includes:

-Participant ID
-Record number
-Record type (labor, pregnancy)

-Age of the participant (years)
-BMI (body mass index) of participant before pregnancy
-BMI of participant at time of recording
-Gravidity (number of times participant has been pregnant, including current pregnancy)
-Parity (number of previous deliveries)
-Previous cesarean (Yes, No)
-Placental position
-Gestational age at recording (weeks/days), according to a first trimester ultrasound
-Gestational age at delivery (weeks/days),
-Mode of delivery (Vaginal, Vaginal/Induction, Elective cesarean, Emergency cesarean due to slow progress, Emergency cesarean due to other than slow progress). Vaginal delivery indicates spontaneous onset unless appended with /Induction.
-Synthetic oxytocin use in labor (Yes, No)
-Epidural during labor (Yes, No)
-Comments for recording
-Comments for delivery

For 111 of the recordings, there is also an annotation file containing the type and timing of events (.atr file).

The types of event are:
- C - Contraction. Used when the participant feels a contraction or there is a very likely contraction on the tocograph (not always used when there is an obvious contraction on the tocograph).
- (c) - Possible contraction. Used when there is not a very likely contraction but the participant has pressure sensation or a contraction is suspected on the tocograph.
- pm - Participant movement.
- pos - Participant change of position.
- fm - Fetal movement. Used when the participant feels fetal movement.
- em - Equipment manipulation. Used when electrodes are pressed more firmly onto the abdomen if otherwise not explained in the comments.

The database also includes:
-The zip file icelandic16ehgmat.zip that includes MATLAB versions of all the signal files along with header files (file names of the form ice###_*type*_*record number*m) and annotations (file names of the form ice###_*type*_*record number*m_ann). This is provided for the convenience of users that want to analyze the data in MATLAB.
-RECORDS.txt containing a list of the recordings by record name, with one record name per line.

Table 13 (see 8.4 Clinical information from each recording) contains the clinical information from each recording (information from the header files excluding comments) along with recording duration and whether or not the recording has a corresponding annotation file. This information can also be found in info.txt in the database.

Table 13: The clinical information from each of the 122 recordings in the Icelandic 16-electrode EHG database along with recording duration and whether or not the recording has a corresponding annotation file.

Technical Validation

The EHG signal has been shown to be representative of uterine contractions[6] and EHG is, in general, a well proven technique[5]. The proof of concept and technical validation for the recording method used for the database was made in a preliminary study in 2007[9].

The preliminary study applied recognized EHG techniques to a new recording setup involving a 4-by-4 grid. This was mainly to better observe and analyze the spatial characteristics and propagation of the electrical activity during contractions rather than just the activity in a single location of the uterus as had been done before[2,6]. The results from the preliminary study showed a very acceptable SNR (signal to noise ratio) of bipolar signals (difference between adjacent monopolar electrodes)[28]. The use of the monopolar signals singly was however considered problematic, even with adaptive filtering methods due to external common mode noise (maternal ECG, respiration movements etc.). Efforts to use recent techniques such as empirical mode decomposition (EMD) and canonical component analysis (CCA) to clean up the signal have since met with some success[29]. The preliminary study data was also used to present a moving picture of the electrical activity during contractions. The activity was clearly observed and correlated well with the simultaneous tocograph trace[9].

In the preliminary study, electrodes were placed one at a time for the 4-by-4 grid, which was a time-consuming task and achieving the desired inter-electrode distance required great operator precision. To address these issues, a placement guide system was specifically developed (described in Methods). The system has a standardized setup ensuring a consistent distance between electrodes. The database therefore contains recordings made in a very similar way to the preliminary study, with the same electrode configuration, the same electrodes and same recording device but with a slightly smaller inter-electrode distance (17.5mm vs 21mm). The system has a shorter electrode attachment time than for the one-by-one electrode attachment method and the placement of the electrodes into the silicone backing can be done before a recording, shortening the setup time for an actual recording to around five minutes. This enables recordings to be performed when there is little time and reduces any inconveniences for participants and health professionals. The data from the preliminary study is not included in the database presented here.

All recordings were performed by the same researcher (AA) using the same protocol. The researcher (AA) stayed with the participants during recordings, recorded events first hand and monitored the equipment and electrode readouts continually. The tocodynamometer was recalibrated to 20 if readouts were zero. In the Embla A10 machine, an electrode that is unconnected or floating will give a signal which very quickly goes to saturation and is therefore easily recognized. If during a recording, an electrode gave a trace that was visually very unlike the traces of other electrodes or displayed values notably different from other values, it was pressed more firmly onto the abdomen to ensure connection to the skin. If all electrode traces seemed suboptimal (e.g. very noisy visually), the reference electrode connections were tested. In a few cases when suboptimal traces did not improve, the electrodes were reconnected and the recording then resumed. We therefore assume that although there may be parts of the data where the contact is faulty between the skin and the electrode, they are few and moreover they are easily recognizable.

Parts of the data have already been used for developing and analyzing various signal processing methods and have led to several publications[8,10–23]. The technical quality of the data has been thoroughly checked throughout this work and has never been found to be lacking.

Usage Notes

Although EHG has become, in general, a well proven technique, the EHG signals are known to be problematic in the sense that they are very low frequency and very low amplitude.

The individual monopolar signals contain the measurement of the electric potential at each site in the matrix as referenced to an electrode far away over the iliac crest were little electrical activity is suspected. The raw signals therefore contain everything that creates a potential difference between the monopolar electrode and the reference. This includes the maternal ECG and some EMG from striated muscles. The signal also contains artifacts related to movements of the participant, fetal movement, fetal heart activity and even fetal respiration has been observed. Fetal hiccups can give mysterious periodic spikes that are clearly visible.

The annotations and the tocograph complement each other. Some contractions that are present in the annotations are not obvious on the tocograph and some obvious contractions on the tocograph are not in the annotations. This explains in part why some recordings are without annotations. Noted fetal movement could last for differing amounts of time and participants did not always notify if they felt fetal movement. Sometimes fetal movement can be seen on the tocograph.

The first sample of a signal file is indexed with 1 in the .mat files but with 0 in the .dat files and so there is an index number discrepancy of one between the two file formats (.mat and .dat).

The MATLAB files contain the absolute raw units. Division by 131.068 gives the physical units in mV.

Even though some pregnancies ended in cesarean, the participant was on occasion already in spontaneous labor. These incidences are explained in the comments sections of the header files.

If a recording is close to birth, then the timing of the recording with regard to the birth is in the comments of the header file.

Our aim is to publish this database as is, without giving the user any detailed directions and thus encouraging open-minded exploitation of the data. Making raw data available is the best way to do this. There are however some pointers about how to make sense of the data which we would like to communicate to the users.

1. The data is sampled at 200 Hz. The EHG signals are generally assumed to be of very low frequency, from almost dc up to 3 Hz maximum. Decimating the raw signal (after low pass filtering) is advised, before or after creating a bipolar/multipolar signal or other de-noising. This will create more manageable files and it is often better to work with signals were the difference in the frequency of the signal to be observed is not as far away from the sampling frequency as in this case. This will also cut down calculation time in any complex analysis. Results have shown that decimation of this signal has little or even a positive effect on the performance of methods[20].
2. Please be aware that there are inherent imprecisions in the synchronization between the tocograph and the recordings and annotations. There can be differences in when participants start to feel a contraction or fetal movements and so differences in when participants notify about events. Events were therefore occasionally approximated to the nearest whole minute. There are also delays internal to the tocodynamometers. Due to factors such as these, the inserted recording times on the tocographs and the annotation times may be up to ± 30 seconds from the actual recording or event times.
3. A user doing intensive work using these signals will have to develop an effective methodology to keep track of the signals, the way they have been pre-treated and the associated clinical

parameters. Inspiration on how to organize such work can be sought in the structure of the SQL database framework which we developed[30].

Users can view the data and annotations through two web interfaces: LightWave (a JavaScript viewer – www.physionet.org/lightwave) and ATM (Automated Teller Machine – www.physionet.org/cgi-bin/atm/ATM). ATM´s toolbox includes software that can convert WFDB signal files to text, CSV, EDF or .mat files and can show samples and annotations as text.

WAVE can also be used for viewing and analyzing the signals (www.physionet.org/physiotools/wfdb.shtml#WAVE). Annotations can be opened by using the "–a" option with the wave WFDB application (www.physionet.org/physiotools/wag/wave-1.htm).

The WFDB Software Package can be used to work with the recordings (www.physionet.org/physiotools/wfdb.shtml). For example, the command "rdann -r *record name* -a ann -f 0" can be used to read the annotation files (www.physionet.org/physiotools/wag/rdann-1.htm). The WFDB toolbox for MATLAB/Octave can also be used to work with the files (www.physionet.org/physiotools/matlab/wfdb-app-matlab).

Acknowledgements

The Icelandic Centre for Research (RANNIS) supported this work. Óskar Þór Lárusson and Snorri Már Snorrason assisted with the development of the placement guide system. Shirin Najdi & Arnaldo Batista modified the rdann function from the WFDB toolbox for MATLAB, which enabled the reading of annotation information from this database. The authors of this manuscript had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Author Contributions

AA wrote most of the text in the Data Descriptor, was the main developer of the placement guide system, participated in designing the recording protocol and performed the EHG measurements.
TS was responsible for applications to the National Bioethics Committee and the Data Protection Authority in Iceland, evaluated and recruited possible participants and was responsible for the clinical aspects of the recordings.
JT assisted in the development of the placement guide system, participated in designing the recording protocol and taught AA how to perform EHG measurements.
CM heads the French group and participated in designing the recording protocol.
BK heads the Icelandic group, assisted in the development of the placement guide system and participated in designing the recording protocol.
Both CM and BK lead the original development of 4-by-4 electrode EHG recordings, managed the associated grants and supervised the students working on the project.

Competing Financial Interests

The authors declare no competing financial interests.

References

1. March of Dimes, PMNCH, Save the Children & WHO. *Born Too Soon: The Global Action Report on Preterm Birth*. (eds Howson, C.P., Kinney, M.V., Lawn, J.E.) (World Health Organization Geneva, 2012).
2. Garfield, R. E. & Maner, W. L. Physiology and electrical activity of uterine contractions. *Semin. Cell Dev. Biol.* **18,** 289–295 (2007).
3. Schlembach, D., Maner, W. L., Garfield, R. E. & Maul, H. Monitoring the progress of pregnancy and labor using electromyography. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **144 Suppl 1,** S33–39 (2009).
4. Lucovnik, M. *et al.* Use of uterine electromyography to diagnose term and preterm labor. *Acta Obstet. Gynecol. Scand.* **90,** 150–157 (2011).
5. Lucovnik, M. *et al.* Noninvasive uterine electromyography for prediction of preterm delivery. *Am. J. Obstet. Gynecol.* **204,** 228.e1–228.10 (2011).
6. Devedeux, D., Marque, C., Mansour, S., Germain, G. & Duchêne, J. Uterine electromyography: a critical review. *Am. J. Obstet. Gynecol.* **169,** 1636–1653 (1993).
7. Vinken, M. P. G. C., Rabotti, C., Mischi, M. & Oei, S. G. Accuracy of frequency-related parameters of the electrohysterogram for predicting preterm delivery: a review of the literature. *Obstet. Gynecol. Surv.* **64,** 529–541 (2009).
8. Hassan, M. *et al.* Better pregnancy monitoring using nonlinear correlation analysis of external uterine electromyography. *IEEE Trans. Biomed. Eng.* **60,** 1160–1166 (2013).
9. Karlsson, B., Terrien, J., Gudmundsson, V., Steingrimsdottir, T. & Marque, C. Abdominal EHG on a 4 by 4 grid: mapping and presenting the propagation of uterine contractions. in *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007* (eds. Jarm, T., Kramar, P. & Zupanic, A.) 139–143 (Springer Berlin Heidelberg, 2007).

10. Hassan, M., Terrien, J., Alexandersson, A., Marque, C. & Karlsson, B. Improving the classification rate of labor vs. normal pregnancy contractions by using EHG multichannel recordings. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2010,** 4642–4645 (2010).
11. Hassan, M., Terrien, J., Alexandersson, A., Marque, C. & Karlsson, B. Nonlinearity of EHG signals used to distinguish active labor from normal pregnancy contractions. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2010,** 2387–2390 (2010).
12. Terrien, J., Hassan, M., Alexandersson, A., Marque, C. & Karlsson, B. Evolution of phase synchronization of the two frequency components of the electrohysterogram (EHG): application to the detection of human labor. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2010,** 17–20 (2010).
13. Hassan, M., Alexandersson, Á., Terrien, J., Karlsson, B. & Marque, C. Wavelet phase synchronization between EHGs at different uterine sites: comparison of pregnancy and labor contractions. in *XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010* (eds. Bamidis, P. D. & Pallikarakis, N.) 21–24 (Springer Berlin Heidelberg, 2010).
14. Moslem, B., Karlsson, B., Diab, M. O., Khalil, M. & Marque, C. Classification performance of the frequency-related parameters derived from uterine EMG signals. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2011,** 3371–3374 (2011).
15. Hassan, M., Terrien, J., Marque, C. & Karlsson, B. Comparison between approximate entropy, correntropy and time reversibility: application to uterine electromyogram signals. *Med. Eng. Phys.* **33,** 980–986 (2011).
16. Diab, A., Hassan, M., Marque, C. & Karlsson, B. Quantitative performance analysis of four methods of evaluating signal nonlinearity: application to uterine EMG signals. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2012,** 1045–1048 (2012).

17. Karlsson, B., Hassan, M. & Marque, C. Windowed multivariate autoregressive model improving classification of labor vs. pregnancy contractions. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2013,** 7444–7447 (2013).
18. Diab, A., Hassan, M., Boudaoud, S., Marque, C. & Karlsson, B. Nonlinear estimation of coupling and directionality between signals: Application to uterine EMG propagation. *Conf. Proc.* **2013,** 4366–4369 (2013).
19. Diab, A., Marque, C., Diab, A., Karlsson, B. & Hassan, M. Comparison of methods for evaluating signal synchronization and direction: Application to uterine EMG signals. in *2013 2nd International Conference on Advances in Biomedical Engineering (ICABME)* 14–17 (2013). doi:10.1109/ICABME.2013.6648835
20. Diab, A., Hassan, M., Karlsson, B. & Marque, C. Effect of decimation on the classification rate of non-linear analysis methods applied to uterine EMG signals. *IRBM* **34,** 326–329 (2013).
21. Diab, A., Hassan, M., Laforêt, J., Karlsson, B. & Marque, C. Estimation of coupling and directionality between signals applied to physiological uterine EMG model and real EHG signals. in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013* (ed. Romero, L. M. R.) 718–721 (Springer International Publishing, 2014).
22. Diab, A., Hassan, M., Marque, C. & Karlsson, B. Performance analysis of four nonlinearity analysis methods using a model with variable complexity and application to uterine EMG signals. *Med. Eng. Phys.* **36,** 761–767 (2014).
23. Alamedine, D. *et al.* Selection algorithm for parameters to characterize uterine EHG signals for the detection of preterm labor. *Signal Image Video Process.* **8,** 1169–1178 (2014).
24. Hassan, M., Terrien, J., Karlsson, B. & Marque, C. Interactions between uterine EMG at different sites investigated using wavelet analysis: comparison of pregnancy and labor contractions. *EURASIP J. Adv. Signal Process.* **2010,** 918012 (2010).

25. Terrien, J., Steingrimsdottir, T., Marque, C. & Karlsson, B. Synchronization between EMG at different uterine locations investigated using time-frequency ridge reconstruction: comparison of pregnancy and labor contractions. *EURASIP J. Adv. Signal Process.* **2010,** 242493 (2010).
26. Marque, C. *et al.* A multiscale model of the electrohysterogram the BioModUE_PTL project. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2013,** 7448–7451 (2013).
27. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101,** e215–e220 (2000).
28. Terrien, J., Marque, C., Steingrimsdottir, T. & Karlsson, B. Evaluation of adaptive filtering methods on a 16 electrode electrohysterogram recorded externally in labor. in *11th Mediterranean Conference on Medical and Biomedical Engineering and Computing 2007* (eds. Jarm, T., Kramar, P. & Zupanic, A.) 135–138 (Springer Berlin Heidelberg, 2007).
29. Hassan, M., Boudaoud, S., Terrien, J., Karlsson, B. & Marque, C. Combination of canonical correlation analysis and empirical mode decomposition applied to denoising the labor electrohysterogram. *IEEE Trans. Biomed. Eng.* **58,** 2441–2447 (2011).
30. Terrien, J., Marque, C., Gondry, J., Steingrimsdottir, T. & Karlsson, B. Uterine electromyogram database and processing function interface: An open standard analysis platform for electrohysterogram signals. *Comput. Biol. Med.* **40,** 223–230 (2010).

Data Citations

1. Alexandersson, A., Steingrimsdottir, T., Terrien, J., Marque, C. & Karlsson, B. PhysioBank www.physionet.org/physiobank/database/ehgdb (2014)