Innovative Processing Algorithms for Fetal Magnetoencephalographic Data

Dissertation der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard-Karls-Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Katrin Sippel
aus Ulm

Tübingen
2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	10.07.2020
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Wolfgang Rosenstiel
2. Berichterstatter:	Prof. Dr. Hubert Preissl

Acknowledgment

A thousand thanks to all those who supported me physically, mentally or materially during this time, to all those who explained and taught me so many things or simply listened to me and especially to all those who read and corrected my manuscripts.

Materially, this work was partially supported by the Luminous project (EU-H2020-FETOPEN GA No.686764), the Helmholtz Alliance ICEMED-Imaging and Curing Environmental Metabolic Diseases and German Research Foundation (TR-SFB 654 Plasticity and Sleep) and a grant from German Research Foundation (DFG; SP 1533/2-1).

First of all I would like to thank Prof. Rosenstiel and Prof. Preissl very much for giving me the opportunity to write my doctoral thesis in this interdisciplinary field of computer based methods in fetal magnetencephalography.

I thank my colleagues from Sand, Armin, Carina, Christian, Dirk , Jörg, Markus, Martin, Sebastian, Tanja and Thomas for the numerous methodical discussions, their support and the incredibly funny lunch breaks at the Mensa Morgenstelle.

Martin once again deserves a very special thank you. His priceless scientific advices supported me unbelievably during the whole doctoral thesis and taught me so much.

Furthermore, I would like to thank my colleagues from the fMEG Franziska, Julia H., Haliza, Magdalene, Julia M., Ilena and Lorenzo for the nice working atmosphere and the constructive coffee breaks.

Franziska has supported me in all aspects of the AURORA study and always found time for my countless questions and ideas, especially at the beginning of my PhD. With her help, I was able to develop a good sense for the fMEG data.

I thank Julia M. for every single exchange of ideas. The conversations with her were so productive and often very enlightening. Without her many of my thoughts would still be in my head and not published in structured written form.

Thank you Julia H. for the common suffering in the final writing phase of our PhDs.

I thank Ilena for always motivating me and listening to me when I couldn't get any further, politely and patiently, and then giving me energy again with a smile and an optimistic advice.

Grazie mille a Lorenzo, per la miglior prova di lettura di sempre e per avermi restituito la fede nella magia e nell'impossibile.

A special thanks also goes to Magdalene, who supported me with so much during my PhD as well as during my pregnancy and who also evaluated data by hand for hours for my AURORA study.

In particular, I want to thank Hubert for leaving me a completely free hand in choosing my final thesis topic and thereby educating me to scientific independence. Whenever I had a question, he had not only an open door and an open ear, but also a lot of HARIBO.

Vor allem aber möchte ich meiner ganzen großen Familie danken, in der mich jeder Einzelne auf irgendeine Art und Weise unterstützt hat. Euch zu haben ist so ein Gewinn!

Ein ganz großes Dankeschön geht an meine Eltern und Schwiegereltern die sich in den ganzen letzten Jahren so regelmäßig, zuverlässig und liebevoll um Liam und Madita und noch so viel mehr gekümmert haben. Ohne euch wäre diese Arbeit nicht möglich gewesen! Aber der allergrößte Dank geht an die drei wichtigsten Menschen im meinem Leben: Jens, Liam und Madita.

Danke Jens, dass du so oft deine Schulstunden bis weit nach Mitternacht vorbereitet hast, damit du dich mittags um die Kids kümmern konntest, während ich gearbeitet habe.

Danke Liam, dass du mich mit deinem "Und Mama, wie wars heute an der Uni? Bist du mit deiner Doktorarbeit fertig geworden?" immer wieder ermahnt und gepusht hast.

Danke Madita, dass du mit mir an zahlreichen fMEG Studien mitgemacht hast, ohne dich hätte es die AURORA Studie vielleicht nie gegeben.

List of Abbrevations

AURORA	**au**tonomous and central nervous **r**eaction t**o** mate**r**n**a**l voice
FAIRY	**f**ully **a**utomated process**i**ng fo**r** fetal magnetoencephalograph**y**
FAUNA	**f**ully **a**utomated s**u**btractio**n** of heart **a**ctivity
FLORA	**f**u**l**ly aut**o**mated **R**-peak detection **a**lgorithm
fMEG	**f**etal **m**agneto**e**ncephalo**g**raphy
GA	**g**estational **a**ge
HR	**h**eart **r**ate
HRV	**h**eart **r**ate **v**ariability
HTA	**h**ilbert **t**ransformation **a**lgorithm
MEG	**m**agneto**e**ncephalo**g**raphy
OP	**o**rthogonal **p**rojection
SARA	**S**QUID **a**rray for **r**eproductive **a**ssessment
SATM	**s**emi **a**utomated **t**emplate **m**atching
SQUID	**s**uperconducting **q**uantum **i**nterference **d**evice

Abstract

Fetal magnetoencephalography (fMEG) facilitates the investigation of both the nature and development of the fetal central and autonomic nervous system, starting at 20 weeks of gestational age. Like magnetoencephalography in children and adults, fetal magnetoencephalography is a noninvasive method and therefore completely harmless for both the mother and the child. Magnetic sensors in fMEG devices are arranged around the abdomen of the pregnant woman. The good spatial and temporal resolution allows to measure maternal and fetal magnetocardiograms simultaneously with fetal brain activity. The fMEG signals are mainly used to measure the auditory and visual event-related brain responses or the spontaneous brain activity. Important questions concerning the developmental process of the fetal brain, as well as the maternal influence on the metabolic and cognitive state of the newborn, can be clarified by the analysis of fMEG signals.
The evaluation of the fetal brain activity poses some challenges, as the signals of fetal and maternal heart activity are 10-1000 times stronger than the fetal brain signal. Therefore, it is mandatory to detect and remove the heart activity of both the mother and the fetus before analyzing the fetal brain activity. The currently used methods for this detection and removal work well for most datasets, but the processing includes numerous manual steps and is therefore very time consuming. Furthermore, signal redistribution is a problem with the current methods, which makes later detection of the fetal brain activity challenging.

The aim of this work was to make the evaluation of fMEG data faster, better and nevertheless, easy to use. In this thesis two new fully-automated procedures for the detection and removal of the heart activity are presented. The fully automated R-peak detection algorithm (FLORA) improves R-peak detection by combining and extending the advantages of the previously used methods. The algorithm for the fully automated subtraction of heart activity (FAUNA) improves the signal quality and facilitates detection of brain activity without the problem of redistribution. Furthermore these methods lead to a higher reliability of the data analysis since no manual interventions are necessary. Combining both methods in a tool for fully automated processing for fetal magnetoencephalography (FAIRY) makes data evaluation now easy and fast. This prepares the processing of fMEG data for the era of "Big Data" and "Automated Science".

Additionally a study about the fetal and neonatal autonomous and central nervous response to maternal voice (AURORA) was performed using the new data processing methods. The results showed a reduced movement of fetuses between 26 and 32 weeks of pregnancy and a lower heart rate during the fist 20 seconds of stimulation in the last weeks of pregnancy as a reaction to maternal voice. We additionally found a higher amplitude of the brain response to voice onset of a stranger female voice in newborns.

Zusammenfassung

Fetale Magnetenzepahalographie (fMEG) ermöglicht die Untersuchung der Entwicklung des zentralen und des autonomen Nervensystems bei Feten ab der 20. Schwangerschaftswoche. Wie normale Magnetenzephalographie bei Erwachsenen und Kindern ist auch fMEG eine nicht-invasive Methode und in der Anwendung vollkommen harmlos für Mutter und Kind. Die magnetischen Sensoren sind hierbei um das Abdomen der schwangeren Frau angeordnet. Die gute räumliche und zeitliche Auflösung erlaubt es, mütterliche und fetale Magnetokardiogramme gleichzeitig mit der fetalen Hirnaktivität zu messen. Die Signale der fetalen Magnetoenzephalographie werden vor allem zur Messung von auditiven und visuellen ereignisbezogenen Hirnreaktionen oder der spontanen Hirnaktivität verwendet. Wichtige Fragen zum Entwicklungsprozess des fetalen Gehirns und des autonomen Nervensystems sowie der mütterliche Einfluss auf den metabolischen und kognitiven Zustand des Neugeborenen können durch die Analyse der fetalen Magnetoenzephalographie-Signale geklärt werden.
Die Auswertung der fetalen Hirnaktivität birgt einige Herausforderungen, da die Signale der fetalen und mütterlichen Herzaktivität etwa 10-1000 mal stärker sind als das fetale Hirnsignal. Daher ist es zwingend erforderlich, die Herzaktivität der Mutter und des Fetus zu erkennen und zu entfernen, bevor die fetale Hirnaktivität analysiert wird.
Die derzeit verwendeten Methoden für die Erkennung und Entfernung der Herzaktivität funktionieren für die meisten Datensätze zuverlässig, die Verarbeitung enthält jedoch einige manuelle Schritte, was das Ganze sehr zeitaufwändig macht. Darüber hinaus ist die Signal Redistribution beim Entfernen der Herzaktivität ein bekanntes Problem, welches es schwierig macht, die Hirnaktivität später zu identifizieren. Das Ziel dieser Arbeit war es, die Auswertung der fMEG Daten schneller, besser und trotzdem leicht handhabbar zu machen.

In dieser Arbeit werden zwei neue vollautomatisierte Methoden zur Erkennung und Entfernung der Herzaktivität vorgestellt. Der vollautomatisierte R-Peak Erkennungsalgorithmus (FLORA) verbessert die R-Peak Erkennung, indem er die Vorteile der zuvor verwendeten Methoden kombiniert und erweitert. Der Algorithmus zur vollautomatisierten Subtraktion der Herzaktivität (FAUNA) verbessert die Signalqualität und vereinfacht die Erkennung der Hirnaktivität, ohne Redistribution. Die Zuverlässigkeit der Daten wird dadurch erhöht, da keine manuelle Auswahl getroffen werden muss. Die Kombination beider Methoden in einem Programm zur vollautomatisierten Verarbeitung für die fetale Magnetoenzephalographie (FAIRY) macht die Datenauswertung nun einfach und schnell. Damit wird die fMEG Datenverarbeitung auf die "Big Data"- und "Automated Science"-Ära vorbereitet.

Des Weiteren wurde eine Studie über die autonome und zentralnervöse Reaktion von Feten und Neugeborenen auf die mütterliche Stimme (AURORA) mit den neuen Datenverarbeitungsmethoden durchgeführt. Die Ergebnisse zeigten eine reduzierte Bewegung der Feten zwischen der 26. und 32. Schwangerschaftswoche und eine niedrigere Herzfrequenz während der ersten 20 Sekunden der Stimulation in den letzten Schwangerschaftswochen, als Reaktion auf die mütterliche Stimme. Zusätzlich fanden wir eine höhere Amplitude der Gehirnreaktion als Reaktion auf eine fremde Frauenstimme bei Neugeborenen.

Contents

1. Introduction

Once upon a time scientists developed a fascinating procedure called fetal magnetoencephalography, which will now be taken one step closer to the era of automated science.

Fetal magnetoencephalography offers unique possibilities for the examination of the fetal central and autonomic nervous system development [1, 2] but in the whole world only few devices exist. Like magnetoencephalography for children and adults [3], fetal magnetoencephalography is a noninvasive method and therefore completely harmless for both the mother and the fetus.
It is primarily used for measuring the auditory or visually event related brain responses of the fetus [4, 5, 6, 7, 8, 9, 10] as well as spontaneous brain activity [11, 12]. The analysis of fetal brain signals allows to address important questions regarding the developmental process happening in the fetal brain and how they are influenced by the maternal metabolism. With its good spatio-temporal resolution, it also allows to monitor the maternal and fetal magnetocardiograms simultaneously with the recording of fetal brain activity. This enables the evaluation of maternal and fetal heart rate, different parameters of the heart rate variability as a proxy for autonomous nervous system functioning [13, 14] and fetal behavioral states starting at 20 weeks [15, 16]. The event-related brain responses can be detected starting at 20 -24 weeks of gestational age.
The theory that the fetal development is very sensitive to environmental or metabolical influences of the mother (like smoking, pollution, malnutrition, chronic metabolic diseases, depression and anxiety) has led to an increasing interest in this research topic, called "fetal programming", over the recent years [17, 18]. Since the origin of many diseases is attributed to factors during fetal development it is very important to investigate how the fetal central and autonomous nervous systems develop normally, because only then deviations from this normal trajectory can be detected early and treated accordingly.

But fetal magnetoencephalography also faces some difficulties. Both the mother and the fetus undergo a huge development between the 20th and the 40th week of gestation. This leads to a huge change in the composition and strength of the recorded signal. As the fetus grows, the magnitude of both the heart and the brain signal get higher and visible on more sensors, but also the movement of the fetus can cause artifacts of a higher amplitude. Furthermore, the magnetical sensors are very susceptible to external influences (e.g. a lawnmower or a helicopter near the building).
Methods used for the processing and artifact removal of these data thus has to be very adaptive and flexible.

The methods that are currently used for the removal of artifacts work reliably for the maternal and fetal heart, but they reach their limits in the evaluation of the fetal brain or in very noisy datasets. In those cases a lot of data have to be evaluated through a great deal of manual work and therefore also a lot of time.

The present work shows the creation of new methods for the evaluation of fetal magnetoencephalographic data that compensate the weak points of the previous evaluation procedures, automate processes and minimize the required manual evaluation time.
Since each manual intervention could influence the evaluation result, an automated data processing improves the reproducibility of the evaluations. Included in an optimal tailored graphical user environment, the selection of datasets for evaluation is transparent and fast. The time saved by automated processes can be used for additional explorative evaluation approaches to generate new hypotheses for future studies and to exploit the high potential of the fetal magnetoencephalographic data.

2. Background

2.1. Recording Techniques

2.1.1. Magnetoencephalography

Magnetoencephalography (MEG, see Fig. 2.1) is a non-invasive method that measures brain activity, with good temporal and spatial resolution, first used by David Cohen in 1968 [19]. The electrical currents of active nerve cells produce a magnetic field. This field is not distorted by passing through the layers of biological tissue and can be detected by a superconducting quantum interference device (SQUID) sensor. The SQUID sensors have to be cooled down to -269°C by liquid helium to reach superconductivity. They are able to detect magnetic fields with tangential but not radial directions towards their superconducting loop. To avoid the influence of environmental noise, like low frequency waves or other electromagnetic radiations, the system is operated in a multilayer magnetically-shielded room of μ-metal and aluminum. Inside this room, external magnetic fields are highly attenuated with an attenuation factor around 1000 for frequencies above 1 Hz. For measuring brain activity in adults, 200 – 300 SQUID magnetometers are arranged in a hemispherical shape around the head (see Fig. 2.1). The magnitude of adult brain signals measured by MEG is $\sim$ 1 pico Tesla (pT, $= 10^{-12}$T) for spontaneous activity, and $\sim$ 100 femto Tesla (fT, $= 10^{-15}$T) for event related brain responses [20].

Pictures courtesy of the University Hospital Tübingen.

Figure 2.1.: An adult woman prepared for a MEG measurement. Her head is inside the MEG device where 275 SQUID sensors surround the surface of her head.

2.1.2. Fetal Magnetoencephalography (fMEG)

Fetal magnetoencephalography (fMEG) is a specific application of MEG. Due to the tissue-passing property of magnetic fields, this method can be used to non-invasively asses the brain activity of a fetus inside the maternal abdomen. Besides the magnetic field of the fetal brain, fetal and maternal heart activity are also recorded [1, 21].
The Helmholtz Center Munich at the University of Tuebingen operates a SQIUD array for reproductive assessment (SARA), built by VSM MedTech Ltd., Port Coquitlam, Canada. The 156 SQUID sensors of the SARA are spread over a shell-shaped array around the maternal body from the throat to the tailbone (see Fig. 2.2c).

(a) The SARA is placed inside a magnetically shielded room.*

(b) Sensors (grey) record magnetical activity of the maternal heart, the fetal heart and the fetal brain.

(c) During the measurements the pregnant woman is sitting on the device in a comfortable position.*

(d) The SARA can also be used for heart and brain activity measurements in newborns.*

* Pictures courtesy of the University Hospital Tübingen.

Figure 2.2.: Fetal Magnetoencephalography is a method to asses heart and brain activity in neonates and in fetuses alongside with the maternal heart activity.

It is also possible to measure the brain activity of neonates with the same device. For this purpose a cradle is attached to the system and the neonate is then bedded inside with his/her head inside the shell-shaped array (see Fig. 2.2d).

Data on the SARA device in Tuebingen are usually recorded with a sampling rate of 1220.7 Hz for measurements of spontaneous activity, and 610.35 Hz for measurements during visual or auditory stimulations. The resulting datasets include a matrix of the measured magnetical activity over all 156 sensors and all sampling points, along with the 3D coordinates of all sensors and all kind of stimulation triggers. Additionally the dataset include the 3D coordinates of 4 magnetic coils. Three of them are placed on the left, back and right side of the participants waist, and the forth coil is placed on the surface of the maternal abdomen as near as possible to the fetal head. These 4 coils can be used to determine the position of mother and fetal head in relation to the SQUID sensors.

The magnitude of the magnetic activity of the fetal brain is ≈ 100 fT for spontaneous brain activity and $\approx 10 - 50$ fT for event-related responses. The magnitude of the fetal heart activity is $\approx 1 - 10$ pT while the magnitude of the maternal heart activity is $\approx 10 - 50$ pT (see Fig. 2.3). These differences in magnitude lead to several problems in signal processing and evaluation of fetal brain activity which are described in the following chapters.

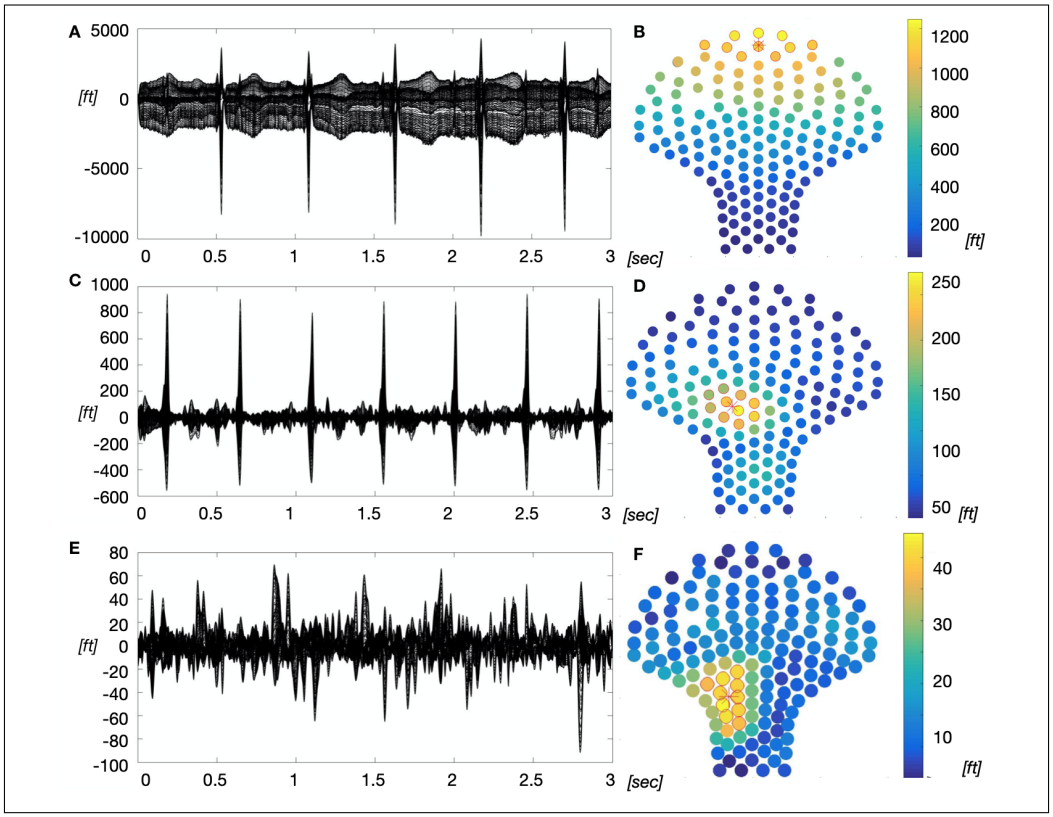

Figure 2.3.: First row: raw signal with prominent maternal heart component (**A**) and the root mean square representation on all SQUID sensors (**B**). Second row: signal after removal of the maternal heart activity with prominent fetal heart component (**C**) and the root mean square representation on the sensors (**D**). Third row: remaining activity after removal of maternal and fetal heart activity (**E**), and the root mean square representation of the data on the sensors with prominent cluster of activity (**F**).

Fetal magnetoencephalography is a unique method that can simultaneously record maternal and fetal heart activity as well as fetal brain signals. However, the high noise level and the difference in signal strength between the heart and brain components generate some challenges for signal processing and evaluation.

2.1.3. Other Fetal Recording Techniques

Sonography is the most common technique for investigation of fetal health and well-being. It uses ultrasound to generate a two-dimensional black and white live image of the fetus inside the maternal abdomen. With modern sonographs it is possible to see the embryo / amniotic sac from 5 weeks of gestational age. In Germany, sonography is an integral part of medical checkups during pregnancy. The Doppler Sonography is using the Doppler effect to generate information about the blood flow. Color coded doppler sonography is even able to color the blood flow direction in doppler images using two different colors (normally blue and red) to describe if the blood is flowing away from or towards the transducer.

Cardiotocography is a method that also uses Doppler sonography to asses fetal heartbeat. This method is not used before the end of the second trimester because until 25 weeks of gestation the fetus is too small and the heart signal gets lost if the fetus changes its position inside the abdomen. Cardiotocography is also used to asses uterus contractions [22].

Electrocardiography is manly used during labour to asses fetal heart beat and heart rate variability of the fetus to detect a possible hypoxia of the fetus as early as possible. The electrode is attached either to the maternal abdomen or, in an advanced state of birth, directly to the scalp of the child. Since the benefit of electrocardiography is small although it is more invasive, it is recommended to use electrocardiography only if the cardiotocography shows alarming features. [23].

Magnetic Resonance Imaging in fetuses, besides anatomical information, enables the detection of changes in brain activity (blood oxygenation level in functional magnetic resonance imaging). Although Kok et al. [24] already showed in 2004 that investigations in a 1.5T magnetic resonance imaging scanner during the 3rd trimester of pregnancy caused no harmful effects, it is only slowly becoming established in fetal research.

There are several techniques used for fetal assessment but none of them is able to investigate fetal brain activity in such a high temporal resolution like fetal magnetoencephalography.

2.2. Biological and Medical Background

This section provides information about the biological processes that are used in the following chapters.

2.2.1. Heart Activity

The human heart starts his first rhythmical contractions around 21 days after conception, long before the brain of the fetus is formed. It is driven by electrical impulses of the sinoatrial node and consists mainly of muscle cells. Therefore, heart activity also generates an electric and a magnetic field. Measuring this field with electro- or magnetometers in healthy adults shows a typical waveform also called electro- or magnetocardiogram (see Fig. 2.4). The duration of such a wave (also called PQRST-complex) is between 0.3 and 1.0 s in adults. The waves and peaks of the PQRST-complex are the result of the activity of the different parts of the heart muscles. The highest peak of the PQRST-complex is the R-peak resulting from the muscular activity of the heart chambers. Given its amplitude it is used to measure the heart rate based on the actual difference between two successive R-peaks [25]. The distance between two R-peaks of consecutive PQRST-complexes is called RR-interval.

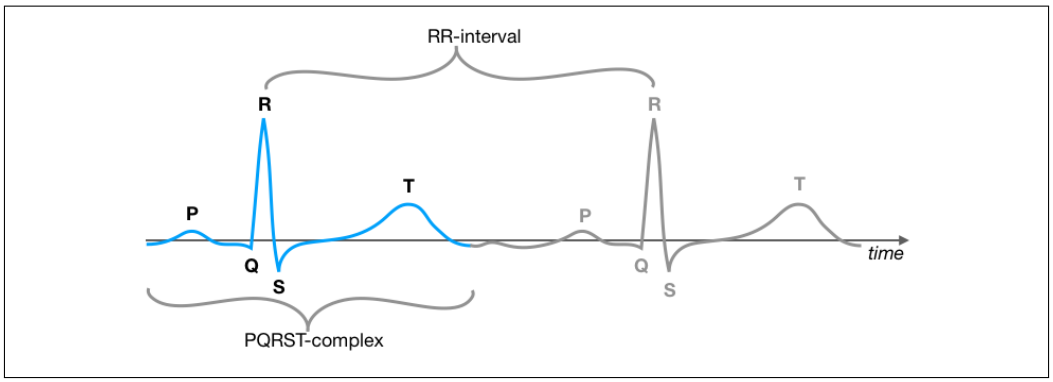

Figure 2.4.: Schematic description of an adult electrocardiogram PQRST-complex including P-wave, Q-peak, R-peak, S-peak and T-wave.

2.2.2. Heart Rate Variability

The human heart rate is driven by the autonomic nervous system consisting of the sympathetic and the parasympathetic nerves. While the sympathetic nerves are responsible for heart rate up-regulation in stressful situations, the parasympathetic system is responsible for relaxation and down-regulation. The RR-interval in healthy adults underlies a lot of variation. The heart rate variability is an indicator of the sympathovagal balance [26]. The parameters of the heart rate variability are influenced by age, gender and body mass index of a person [27]. Disturbance of this balance has been associated with heart failure and cardiovascular diseases [28]. It has also been shown that the heart rate variability is positively

correlated with speed and accuracy of cognitive processes. [29]. Heart rate variability of mothers and fetuses has also been examined in several fMEG publications over the last years [30, 16, 13, 31, 32]. The heart rate variability parameters that can be assessed are separated in time and frequency domain measures [33].

Time Domain Measures The heart rate variability parameters of the time domain are normally evaluated for a time interval of at least 3 min to make sure that a sufficient number of R-peaks are used.

- **Mean RR-interval**, the average distance between two successive R-peaks.
- **Mean heart rate**, inverse proportional to the mean RR-interval.

$$HR = \frac{60}{mean(RR-interval)} \tag{2.1}$$

- **Standard deviation from normal-to-normal (SDNN)**, the standard deviation between two successive normal heart beats. To remove non-normal RR-intervals that are caused by additional (ectopic) or missing heart beats, a data preprocessing step is necessary. SDNN is associated with both, the sympathetic and parasympathetic activity and it has been show that it is also a good indicator for the arousal in fetuses [16].
- **Root mean square of successive differences (RMSSD)**, associated with the recovery and short-term adaption of the heart rate, calculated by

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1}-RR_i)^2} \tag{2.2}$$

  where N is the number of RR-intervals.
- **Normal to normal 50/10**, in adults the percentage of normal RR-intervals that differ more than 50 ms from the previous RR-interval. In fetuses the percentage of normal RR-intervals that differ more than 10 ms from the previous RR-interval.

Frequency Domain Measures To calculate the measures of the frequency domain the heart rate is resampled with 4Hz and a spectral analysis of the heart rate signal is performed

- **Low frequency (LF)** the range of low frequency activity in adults is between 0.04 and 0.15 Hz, and between 0.08 and 0.20 Hz in the fetus. It is associated with both, the sympathetic and parasympathetic activity.
- **High frequency (HF)** the range of high frequency activity in adults is between 0.15 and 0.4 Hz, and between 0.4 to 1.7 Hz in the fetus. It is associated primarily with the parasympathetic activity.
- **Low frequency / high frequency ratio** the proportion between the low and the high frequency. It is an indicator for the sympathovagal balance.

2.2.3. Development of the Fetal Auditory System

The fetal cochlea develops during the first trimester of gestation and matures during the second trimester. At the beginning of the second trimester, the cochlear hair cells are built and the cochlear fluid is separated into scala vestibuli, scala tympany and scala media. The cochlear nerve is myelinized, the basic requirements for fetal hearing [34].
Fetal auditory perception starts around 19 weeks of gestational age for frequencies between 250 and 500 HZ. Between the 20th and 22nd and until the 27th week of gestation the cochlea starts growing in size and at the end of the 27th week nearly all fetuses react to stimulations with 100 Hz. From 31 - 33 weeks of gestational age on also higher frequencies of 1 to 3 kHz can be heard [35].
During the second trimester the brain forms a prominent subplate below the cortical plate. Neurons from the thalamus migrate through this subplate to the cortical plate. The synapses of these neurons reach the cortical plate around 26 weeks of gestational age [36, 37]. Due to these thalamo-cortical connections the fetal brain from then on also has the ability of higher auditory processing [38]. Several fMEG studies showed auditory event-related responses starting between 26-29 weeks of gestational age [39, 40, 41, 4, 42, 5].

Fetal auditory perception starts around 19 weeks of gestational age. However, auditory processing is only possible after 26 weeks of gestational age when the thalamo-cortical connections are built.

2.2.4. Auditory Event-Related Responses

The human brain shows a specific response to external auditory stimuli like tones, words or speech that can be measured by electroencephalography or MEG. This event-related response occurs additive to the ongoing brain activity [43].
In adults, the auditory event related response consists of several positive and negative components (e.g. P100, N200 ect.) named after their polarity and latency after the event onset. In children similar components are visible, but the latency is longer compared to adults [44].
In healthy fetuses the auditory event-related response normally consists of a single component around 200-500 ms at the end of the second trimester, and around 100-300 ms at the end of the third trimester. These latency can even be longer in fetuses with intra-uterine growth-restriction or fetuses of mothers with gestational diabetes 60 min after glucose ingestion, compared to healthy fetuses of the same gestational age [42, 9]. In healthy neonates, the latency of this first component is also around 100-300 ms [41].

Auditory event-related responses are changes in brain activity as a result of specific auditory events (eg. tones or words).

2.3. State of the Art - Maternal Voice

2.3.1. Fetal Reaction to Maternal Voice

The differentiation of voices is a complex skill in which several components are involved. Some studies described the sounds inside the maternal abdomen very loud, reaching levels as high as 90 dB [45] , while others as barely audible. Most of these sound perception studies used an underwater microphone positioned inside the cervix or inside the uterine body after rupture of the amniotic sac. However, they agree that the mother's voice is perceived very well and phonemes, prosody, and unique characteristics of the maternal voice are well preserved [46, 47, 48]. These studies describe that the live spoken maternal voice is perceived multimodally (auditory, vestibular and cutaneous) [49, 50], since high intensity and low frequency components of the maternal voice are perceived as vibrations [51]. External Stimuli like maternal recorded voice or stranger voices can also be perceived in utero but they appear around 15dB lower [45] and low pass filtered since there is a higher loss for higher frequencies passing through the maternal abdomen [50, 52]. An ultrasound study could show that fetuses from the 25th week react opening the mouth to specific maternal live sounds like "LA" [53]. Nevertheless, behavioral ultrasound studies showed that fetal movement reactivity to live spoken maternal voice is not as high as to maternal touch of the abdomen [54] and that fetal reactions to maternal recorded voice presented to the abdomen is even lower [55].

Two studies using functional magnetic resonance imaging found that in fetuses of 33-34 weeks of gestational age voice is processed in the left hemisphere and the activation in the left temporal lobe is higher during stimulation with maternal than stranger voice [56, 57].

Assessing the heart rate reaction to the onset of maternal voice in term fetuses some cardiotocography studies showed heart rate decelerations to live spoken [58] and to recorded maternal voice [59, 60, 61]. Some other studies showed accelerating heart rate responses in fetuses at term [62, 63], while they also described the response for fetuses between 34-37 weeks GA as first decelerating then accelerating, in fetuses of 31-33 weeks GA as decelerating and for fetuses younger than 31 weeks GA as not visible [64].

The fact that the results of these studies are inconsistent could be explained by the fact that both behavioral state and the vagal tone influence the reaction of the fetus [58]. In particular responses of the fetal heart rate are higher during active sleep compared to quiet sleep [65], and differences in heart rate reaction to maternal and stranger voice are only visible in fetuses with high vagal tone [61]. Beside that, the volume of the stimulation also plays a role with increasing the intensity of the stimulation the heart rate reaction to maternal voice seems to change from deceleration to acceleration [66].

There are several studies about the fetal perception and reaction to maternal voice, but they do not give a uniform picture. When the fetus starts recognizing the maternal voice and is able to discriminate it from other female voices still unclear.

2.3.2. Neonatal Reaction to Maternal Voice

The neonates ability to recognize the voice of their mother is seen as an important part of the mother-child bonding. Since the 1980ies it is known that human newborns prefer to hear their mothers voice already a few hours after birth [67]. In the first 24 hours of life the newborn respiration rate drops after hearing maternal voice [60]. Later in life this effect is not visible anymore. Presentation of the maternal voice also leads to increased orienting movements towards the direction of maternal voice in the first 2 hours after birth on the one hand [68], but reduced body movements in the first three days of life on the other hand [55].
A study in neonates 2-7 days after birth showed that hearing maternal voice leads to an increase in the delta wave amplitude in the frontal and parietal brain areas [69]. Another study with a single child of 24/75 days after birth showed similar results: high delta activation to auditory presentation of familiar voices and familiar text passages, theta activation to familiar voice and unfamiliar presented text passages [70]. EEG studies in the last years showed that in newborn the auditory event related-responses to maternal voice were significantly higher than the responses to stranger voices [71, 72].

It hast been shown that newborns react different to the voice of their mother compared to other voices. Differences in the reaction to maternal voice are also visible in the brain activity of newborns.

2.4. Mathematical Background

This section provides information about the mathematical procedures that are used in the method sections in the next chapters.

2.4.1. Correlation

The correlation is a measure of the linear dependence of two vectors. If two vectors are not correlated, it does not necessarily mean that they are independent since they could also have a nonlinear relationship.

Covariance The covariance describes a linear relationship of two vectors (a,b) with equal length dependent on their variance.

$$cov(a,b) = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \overline{a}) * (b_i - \overline{b}) \tag{2.3}$$

where $\overline{a}$, $\overline{b}$ are the means of a, b respectively.

Correlation Coefficient The correlation coefficient ρ is the normalized version of the covariance. The magnitude of the correlation coefficient shows the strength of the linear correlation. The correlation coefficient between two vectors a and b is calculated by:

$$\rho(a,b) = \frac{cov(a,b)}{\sigma_a \sigma_b} \tag{2.4}$$

where σ_a is the standard deviation of a and σ_b is the standard deviation of b. Calculating the correlation coefficients of n vectors results in a $n \times n$ correlation coefficient matrix C. The entries range between -1 (perfect inverse correlated) over 0 (not correlated) to 1 (perfect directly correlated). The diagonal entries of C are always 1, because each vector is perfectly directly correlated with itself [73, 74].

$$C = \begin{pmatrix} 1 & \rho(a,b) \\ \rho(b,a) & 1 \end{pmatrix} \tag{2.5}$$

Cross-Correlation The cross-correlation describes the correlation of two vectors of time series for different time shifts τ as a function of the shift. Given a vector a with n elements and a vector b with m elements. The output vector is of length $c = n + m - 1$.

2.4.2. Regression

Regression is used to estimate the relationship between an outcome variable y from one (simple regression) or more (multiple regression) predictor variable(s) X by building a model of X and minimizing the error in

$$y_i = model(X) + error_i \tag{2.6}$$

Ridge regression, also known as Tikhonov regularization, is a method of regression analysis [75]. It is able to estimates the predictor X and outcome y of a linear model also in cases of multicollinearity (that means that two more predictor variables correlate highly to each other). Since the matrix $(X^T X)^{-1}$ is close to singular in cases of multicollinearity the ridge regression extends the least square estimate

$$\beta = ((X^T X)^{-1} X^T y) \tag{2.7}$$

by the ridge parameter k and the identity matrix I

$$\beta = ((X^T X + kI)^{-1} X^T y) \tag{2.8}$$

Choosing a small positive number for k reduces the variance of the estimates. Using that estimate can minimize the n-dimensional weight vector β to generate the best model.

$$\min_{\beta}((y - X\beta)^T (y - X\beta) + k\beta^T \beta) \tag{2.9}$$

The outcome of a new datapoint x with n dimensions than can be predicted by

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \tag{2.10}$$

2.4.3. Principal Component Analysis

The principal component analysis is an exploratory statistical approach. It reduces complexity and dimensionality of data by searching for linear combinations of variables that explain the main variance of the data (main components).
Given a data matrix X including p vectors of n dimensions

$$\sum_{i=1} pa_i x_i = Xa \tag{2.11}$$

where a is a p-dimensional vector of constants. The variance of such a linear combination is given by:

$$v_{Xa} = a^T cov(X) a \tag{2.12}$$

To find the linear combination with the maximum variance, $a^T cov(X)a$ needs to be maximized.

After this combination is found, the component is projected out and the procedure is repeated for the left data. This finally results in a set of not more than p principal components sorted by their influence on the variance of the data [76, 77, 78].

2.4.4. Independent Component Analysis

The independent component analysis tries to decompose a signal vector x_w into multiple independent components c by using linear transformation.
Before the independent component analysis can be performed, the data have to be centered and whitened.
Centering means that the mean of x is subtracted.

$$x_c = x - \overline{x_c} \tag{2.13}$$

Whitening sets the variance of the centered data to 1 by multiplying the centered data to the eigenvector matrix E and a diagnoal matrix D.

$$x_w = DEx_c \tag{2.14}$$

D contains reciprocals of the square root of the eigenvalues.

$$D = \begin{pmatrix} e_1^{-\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & e_n^{-\frac{1}{2}} \end{pmatrix} \tag{2.15}$$

It is assumed, that the signal vector x_w consists of $n!$ independent and non-gaussian components and could also be written as:

$$x_w = a_1 * c_2 + a_2 * c_2 + \cdots n_1 * c_n \tag{2.16}$$

$$c = Ax \tag{2.17}$$

In this case A is a quadratic mixing matrix and c has the same dimensions as x. The task of the independent component analysis is to estimate A and x_c. Since it is assumed that the single components are non-gaussian the independent component analysis tries to adaptively calculate a vector m by using a contrast function to maximize the non-gaussianity of the data.

$$x_c i = m_i^T c \tag{2.18}$$

Doing that in an iterative process minimizes the difference between the actual and the estimated data.
By inverting the mixing matrix A, the components c can be retransformed to the original data $x_w = Mc$ (inverse independent component analysis). $M = A^{-1}$ therefore is called the unmixing matrix. [79, 80].

fastICA is a very fast and efficient way to compute an independent component analysis. It uses Newton's method as contrast function for the approximation of the non-gaussianity [81].

2.4.5. Fourier Transformation

The discrete Fourier transformation is a method to compute the discrete Fourier transform of a signal [82, 83]. It is based on the assumption that a time signal consists of infinite different sine oscillations of different frequencies and phases. The discrete Fourier transform is able to transform a signal $t_n = t_0, t_1, \ldots, t_{N-1}$ from time to frequency domain $F_k = F_0, F_1, \ldots, F_{N-1}$. This transformation is defined by:

$$F_k = \sum_{n=0}^{N-1} t_n \cdot \left[\cos\left(\frac{2\pi}{N} kn\right) - i \cdot \sin\left(\frac{2\pi}{N} kn\right)\right] \tag{2.19}$$

Fast Fourier Transform is used for an algorithm that is able to compute a discrete Fourier transform fastly and efficiently [84].

Welch's Method for power spectral density estimation is another method to perform a fast Fourier transform. For this case the data are separated to smaller overlapping windows where the discrete Fourier transform is calculated individually. After that, the result is squared and averaged to reduce the variance of the individual measurements [85].

Hilbert Transform is also an algorithm using the fast Fourier transform, consisting of the following four steps:

1. Calculating a fast Fourier transform of a signal x.
2. Generating a vector h with:
   $h(i) = 0$ for $i = \left(\frac{n}{2}\right) + 2, \cdots, n$
   $h(i) = 1$ for $i = 1, \left(\frac{n}{2}\right) + 1$
   $h(i) = 2$ for $i = 2, 3, \cdots, \left(\frac{n}{2}\right)$
3. Calculating the scalar product $x \cdot h$.
4. Calculating the inverse fast Fourier transform of $x \cdot h$.

The result H consists of a real and an imaginary part so that $H = H_{real} + kH_{imag}$. The imaginary part is a product of the real part, but with a phase shift of $\frac{\pi}{2}$ (90°) [86].

2.4.6. Permutation Analysis

The permutation analysis is a statistical test using randomized samples. It assumes that n subjects were tested under both the condition A and B and that there is a difference in the result of the conditions. To see if this difference is significant p random permutations (usually 10,000) of the labels ("A" and "B") of each subject are performed and the difference between the randomly generated conditions is calculated. This way a distribution is generated based on the data, unlikley other statistical procedures that compare the result to a standard normal distribution. After the permutation, the 0.05 confidence interval is calculated for the randomly generated differences. Where the original difference between A and B exceeds this confidence interval the difference is considered significant [87].

2.4.7. Wilcoxon Tests

Wilcoxon Signed Rank Test The Wilcoxon signed rank test is the non-parametric equivalent of the paired samples t-test. It is used for one continuous outcome variable using one categorical predictor with two conditions. For using this test, both groups have to contain data form the same entities or participants.
First, the difference of the two conditions is calculated for each of the n subjects individually. The sign ($+$ or $-$) of each difference is noted and then the absolute value of the differences is ranked starting with rank 1 for the smallest, but non-zero, absolute difference. Differences with value zero are excluded from this step. Then, the rank-sum over all ranks of positive differences is built (positive rank-sum) and the same is done for the ranks of all negative differences (negative rank-sum) for both conditions respectively. After that, either a positive or a negative rank-sum is used as test statistics R. The z-score of this test statistics is calculated by

$$z = \frac{R - \overline{R}}{\sigma_{\overline{R}}} \tag{2.20}$$

where $\overline{R}$ is the mean rank and $\sigma_{\overline{R}}$ is the standard error of this statistic, calculated by

$$\overline{R} = \frac{n(n+1)}{4} \tag{2.21}$$

$$\sigma_{\overline{R}} = \sqrt{\frac{n(n+1)(2n+1)}{24}} \tag{2.22}$$

An absolute value of the z-score > 1.96 results in a p-value < 0.05 [88, 74, 89, 90].

Wilcoxon Rank-Sum Test The Wilcoxon rank-sum test, also known as Mann-Whitney test, is in contrast to the the Wilcoxon signed rank test. It is used for two conditions with different subjects in each condition (n_1 subjects in condition 1 and n_2 subjects in condition 2).
To calculate if there are significant differences between the two conditions, first all values of both conditions are sorted in ascending order and ranked from 1 (lowest value) to $n = n_1 + n_2$ for the highes value. After that, the rank-sum is calculated for each group

individually. One of that rank sums is then used as test statistics R. The z-score of this test statistics is again calculated by

$$z = \frac{R - \overline{R}}{\sigma_{\overline{R}}} \tag{2.23}$$

but, due to the possibly unequal number of entities in each condition in case of the Wilcoxon rank-sum test, the mean rank $\overline{R}$ is calculated by

$$\overline{R} = \frac{n_1(n_1 + n_2 + 1)}{2} \tag{2.24}$$

And the standard error $\sigma_{\overline{R}}$ is calculated by

$$\sigma_{\overline{R}} = \frac{n_1 n_2(n_1 + n_2 + 1)}{12} \tag{2.25}$$

Again, an absolute value of the z-score > 1.96 results in a p-value < 0.05 [91, 74, 89, 90].

2.4.8. Forward Modeling of Magnetical Activity

To evaluate the developed analysis methods it is necessary to generate a model of magnetic field distribution. This can be generated by so called forward modeling.

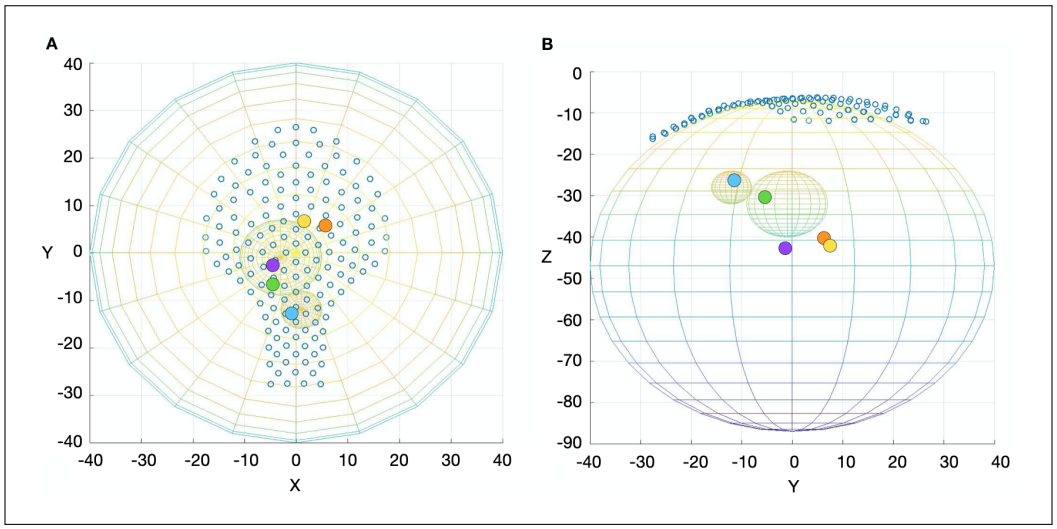

Figure 2.5.: The forward model in **A** front view and **B** 90°rotation. The big sphere describes the maternal body/abdomen, the small spheres describe the fetal body and fetal head. The small colored points describe the point sources for the simulated activity of fetal brain (light blue), fetal heart (light green) and maternal heart activity (purple, yellow and orange). The dark blue circles represent the position of the 156 SQUID sensors of the SARA.

The model we used consisted of three spherical conductors describing the maternal body, the fetal body and the fetal head (see Fig. 2.5). The two spheres describing the fetus were placed inside the sphere describing the maternal body. The fetal brain activity was generated from a point source inside the fetal head sphere, the fetal heart activity from a point source inside the fetal body sphere. The maternal heart activity was generated by three point sources inside the maternal body sphere, but outside the spheres describing the fetus [92, 93].

The magnetic induction B of a source current J^i in a conductor G can be modeled by using the quasistatic approximation of the Maxwell's equation as it is described by Sarvas [94]. The representation of the modeled signal is different for each sensor s_j depending on the distance and the orientation of the source.

2.5. Data Processing Methods for Fetal Magnetoencephalography

Fetal magnetoencephalographic data contains information about the maternal heart activity, the fetal heart activity and the fetal brain activity. To assess this information several data processing steps are necessary. The two most important ones are R-peak detection - for the analysis of the heart rate and heart rate variability - and heart activity removal - which is used to first uncover fetal heart activity and later fetal brain activity.

2.5.1. R-peak Detection

2.5.1.1. Semi Automated Template Matching (SATM)

Data Editor is the in-house software of the SQIUD array for reproductive assessment (SARA) and it provides an R-peak detection based on semi-automated template matching (SATM). Prior to R-peak detection the data are detrended based on the whole trial and filtered with a low-pass of 60Hz and a high-pass filter of 0.5 Hz. Then, several miliseconds around the R-peak of a heartbeat are manually marked and selected as a template. This template is then used to locate and mark all similar heartbeats in the dataset automatically by computing the correlation (see chapter 2.4.1) between the template and the signal. This is done for each of the 156 sensors individually.
The semi-automated template matching runs for maternal and fetal heart separately.

2.5.1.2. Hilbert Transformation Algorithm (HTA)

This automated approach for maternal and fetal R-peak detection (see Chapter 2.2.1) is a MATLAB Script using the built-in *hilbert* function to compute the Hilbert transformation of the fMEG signal [95]. Prior to the Hilbert tranformation the dataset is filtered with a low-pass of 60 Hz and a high-pass filter of 1 Hz using a 4th order butterworth filter with zero phase distortion. Subsequently, the power spectral density for each sensor is calculated and the 10 sensors with the highest mean value are selected. On the data of

these 10 selected sensors the Hilbert transformation and the rate of change of the Hilbert amplitude are calculated (see Chapter 2.4.5). The rate of change of the Hilbert amplitude for each sensor m and data point n is defined by

$$R_{m,n} = \sqrt{(x_{m,n+1} - x_{m,n})^2 + (H_{m,n+1} - h_{m,n})^2} \tag{2.26}$$

Afterwards the cumulative Hilbert amplitude $S(n)$ is calculated.

$$S(n) = sum_m R_{m,n} \tag{2.27}$$

Due to the characteristics of the Hilbert transform, the $S(n)$ is always positive.
Thereafter multiple thresholds are defined, all local maxima points of $S(n)$ above those thresholds are identified as R-peaks and the RR-intervals (see Chapter 2.2.1) of these R-peaks are calculated. Then it is calculated how many RR-intervals lie outside the normal range of RR-intervals. This normal range is defined as 0.29-0.55 s (which corresponds to a heart rate of 110-210 bpm) for fetal and 0.55-1.5 s (which corresponds to a heart rate of 40-110 bpm) for maternal heart activity. The threshold with the lowest number of outliers is then used as fixed threshold [95, 96].
The Hilbert Transformation algorithm also provides the option to compare the results to a prior performed R-peak detection of the Data Editor Software.

2.5.2. Heart Activity Removal

2.5.2.1. Orthogonal Projection (OP)

A widely used method for heart activity removal in fMEG data is the Orthogonal Projection (OP) [97, 98]. Orthogonal projection separates heart activity in a matrix from other activity by attenuation of the estimated signal space of the heart signal. Prior to the orthogonal projection, a R-peak detection of the heart activity has to be performed. Over all detected R-peaks an average PQRST-complex is built resulting in a matrix X. X is of dimension $n \times t$ where n is the number of sensors and t the number of time points of the average PQRST-complex. A signal space vector $v = v_1, \cdots, v_n$ is then selected at the time point t with the largest amplitude of the PQRST-complex. This vector is then projected out. Thereafter, the next vector is selected by using Gramm-Schmidt orthogonalization [99] and projected out. This procedure is repeated until the remaining signal stops exceeding a chosen threshold (see Fig 2.6) [98]. This procedure effectively removes heart activity, but it also redistributes the remaining signal over the sensors. Assuming the brain signal is located on the i_{th} sensor. Each time a signal space vector $v = v_1, \cdots, v_n$ is out projected by matrix multiplication, the components v_i (if $v_i \neq 0$) redistributes the original signal of the i_{th} sensor to the other sensors [100].

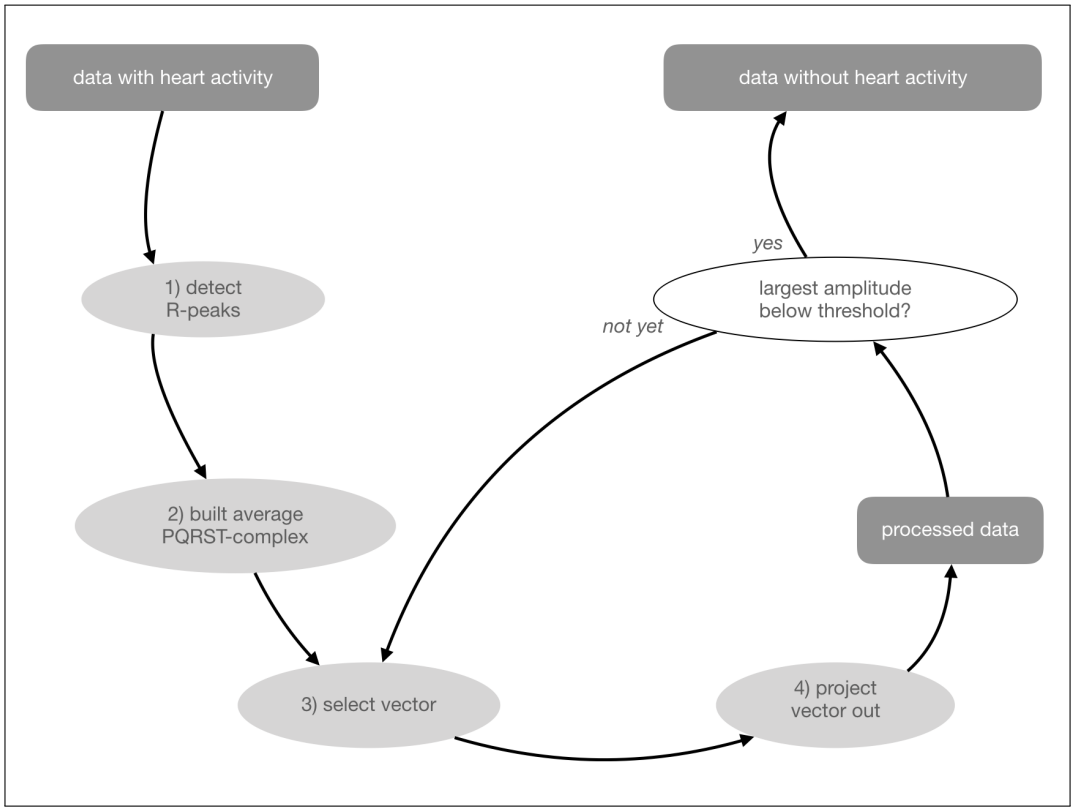

Figure 2.6.: Scheme of the orthogonal projection loop. Steps 3) and 4) are repeated on the already processed data until the amplitude of the remaining signal does not exceed the threshold value anymore.

3. Aims of this work

Improvement of the R-peak detection

The identification of the R-peaks of the heart activity forms the basis of the heart rate and the heart rate variability analysis. The challenge of a good R-peak detection algorithm is to detect R-peaks as specific as possible, ignoring data peaks originating from noise such as fetal or maternal movement. At the same time as sensitive as possible, detecting even weak or covered peaks. For critical and noisy datasets, the previously used methods tend to be either sensitive or specific but not both at the same time and furthermore these methods are very time consuming.

Aim of this work is to improve the quality of R-peak detection by developing an adaptive method that combines the advantages of both the previously used standard methods, and detects maternal and fetal heart rate even in noisy datasets with high specificity and high sensitivity.

Improvement of the heart activity removal

Due to the large differences in the signal strength it is mandatory to remove maternal and fetal heart activity before evaluating the fetal brain activity. The previously used method is effective in removing heart activity but often causes signal redistribution. This leads to a complex and also time-consuming manual selection of brain signal sensors.

Aim of this work is to develop a method that automatically and effectively removes maternal and fetal heart activity in a way that avoids redistribution of the brain activity, and paves the way for further automated brain evaluation.

Improvement of the user convenience

The processing of fetal magnetoencephalography data is performed by researchers with different levels of prior knowledge in data processing. Therefore, data processing should be easy to learn and fast to perform, also in the context of student theses.

Aim of this work is to develop a user-friendly, well-structured graphical user interface for the the processing of fetal magnetoencephalography data with an infrastructure that makes the evaluation of the datasets clear, fast and easy.

4. FLORA

This chapter describes a **f**u**l**ly aut**o**mated **R**-peak detection **a**lgorithm (FLORA). Parts of this chapter has already been published in the *Computer methods and programs in biomedicine* journal, volume 173, pages 35 to 41, in 2019 [101].

The identification of the R-peaks of the heart activity forms the basis of the heart rate and the heart rate variability analysis. The challenge of a good R-peak detection algorithm is to detect R-peaks as sensitive and specific as possible. It should detect as many R-peaks as possible (sensitivity) while ignoring data peaks originating from fetal or maternal movement (specificity). The combination of high specificity and high sensitivity should result in a number of peaks per minute in the range of the natural heart rate, and small and well-distributed distances between consecutive R-peaks (RR-interval). While high specificity and low sensitivity lead to a too low heart rate and huge RR-intervals, a low specificity and high sensitivity cause a too high heart rate estimation, too small RR-intervals and to a lower signal-to-noise ratio.

A combination of the semi-automated evaluation by experts using template matching (SATM, see chapter 2.5.1.1) and the automated Hilbert transformation approach (HTA, see chapter 2.5.1.2) [96] is the standard procedure for both maternal and fetal R-peak detection. Both of these commonly used methods function for most maternal heart evaluations since the representation of the maternal heart signal is quite strong and virtually stationary.
In fetal heart rate, however, the evaluation is somewhat more difficult. Since the mMCG signal is between 10 and 100 times stronger than the fMCG signal [1], the maternal heart signal must first be removed from the data before the fetal heart evaluation can be carried out. In addition, the strength of the detected heart activity is highly dependent on the quality of the recording, on the gestational age and on the position of the fetus. It is even more difficult to detect the fMCG signal when it is superimposed by muscle artifacts or if the fetus moves during the recording.

For both heart rate and heart rate variability analysis, the detection of the maternal/fetal heart activity must be as specific and sensitive as possible. For critical and noisy datasets, the previously used methods tend to be either sensitive or specific but not both.

4.1. Methods

4.1.1. Data Sample

In total, 55 datasets of singleton pregnant women were evaluated. The gestational age of the fetuses ranged from 26 to 39 weeks (mean 30.89 $\pm$ 3.31). The length of the recordings varied between 6 and 27 minutes (mean 12.96 $\pm$ 5.73) and included both spontaneous data and different stimulation paradigms. Data were recorded by a 156 sensors SARA (SQUID Array for Reproductive Assessment, VSM MedTech Ltd., Port Coquitlam, Canada) system at the University of Tuebingen with a sampling rate of 610.35 Hz for measurements with stimulation and 1220.7 Hz for spontaneous measurements.

4.1.2. Fully Automated R-peak Detection Algorithm (FLORA)

FLORA can be divided into four main steps. First, a noise analysis is performed. Second, the heart frequency in the raw dataset is analyzed to identify the individual characteristics of each recording. Third, R-peaks are detected by combining three different approaches. Fourth, contingently existing gaps are filled. This procedure is the same for both the maternal and the fetal R-peak detection. Each of these steps will be explained in more detail in the following section (the workflow of FLORA is depicted in Figure 4.1).

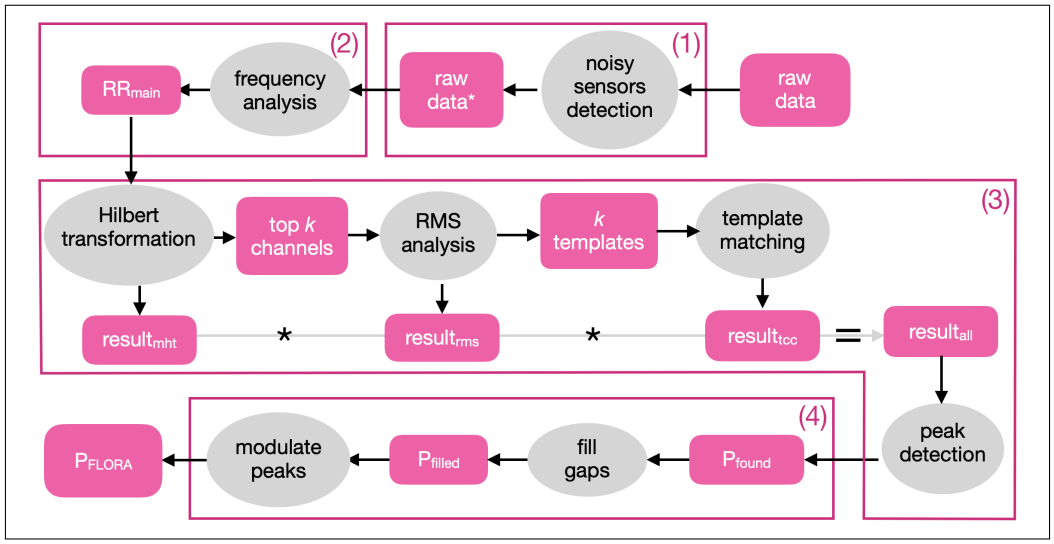

Figure 4.1.: Sketch of the FLORA algorithmic procedure including noisy sensors detection (1), frequency analysis (2), peak detection (3) and interpolation of missing peaks (4).

4.1.2.1. Noisy Sensors Detection

It is possible that some of the 156 sensors show some unspecific noise caused by a defect in the sensor, or by a retainer or a tattoo on the pregnant woman. Such sensors impede the evaluation of the dataset and should be removed before the data processing.

Prior to processing, the raw dataset was filtered with a 4th order Butterworth band pass filter from 1 to 35 Hz. This is necessary to minimize the amount of muscular artifacts of the mother. As a first step of the sensor detection, the data is normalized and variance v_i for each sensor s_i is calculated.
Then, for each sensor s_i, the distance d of its location to every other sensor s_j is calculated by:

$$d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{4.1}$$

where $x_{i,j}$,$y_{i,j}$ and $z_{i,j}$ are the coordinates of $s_{i,j}$ in 3 dimensional space. If $d < 5$ sensors s_i and s_j are considered as 'neighbor sensors'. For each sensor s_i, the mean correlation coefficient $\overline{\rho_{S_i}}$ and according Bonferroni-corrected p-value $\overline{p^*_{S_i}}$ between s_i and all n 'neighbor sensors' s_k are calculated.

$$\overline{\rho_{S_i}} = \frac{1}{n} \sum_{k=1}^{n} \rho(S_i, S_k) \tag{4.2}$$

$$\overline{p^*_{S_i}} = \frac{1}{n^2} \sum_{k=1}^{n} p(\rho(S_i, S_k)) \tag{4.3}$$

Since sensors with high unspecific noise show high signal variance and/or very low or not significant correlation to the signal of their neighbor sensors they can be easily identified and their signal can be set to zero. Threshold values were determined by explorative analysis. The sensors s_i of the noise corrected dataset were defined by

$$s_i = \begin{cases} s_i & \text{if } v_i < 0.2 \;\&\; \overline{\rho_{S_i}} > 0.15 \;\&\; \overline{p^*_{S_i}} < 0.05 \\ 0 & \text{else} \end{cases} \tag{4.4}$$

4.1.2.2. Frequency Analysis

The aim of the frequency analysis is to identify the individual heart rate characteristic of each dataset. Therefore, a power spectral density analysis (see Chapter 2.4.5) is performed on the whole dataset. The result of the spectral analysis is smoothed using a moving average window of 20 data points. The frequency at the maximum of this curve represents the main natural frequency of the heart rate f_{nat} (see Fig. 4.2). The range for maternal frequency analysis is 0.8-2.2 Hz (corresponds to a heart rate of 48 - 132 bpm) and 1.5-3 Hz (corresponding to a heart rate of 90 - 180 bpm) for fetal. The resulting main RR-interval $RR_{main} = \frac{1}{f_{nat}}$ is used as parameters for further analysis (see Fig. 4.1).

4.1.2.3. Peak Detection

Hilbert transformation The magnitude of the Hilbert transformation is calculated for all sensors and for all sampling points (t), resulting in a vector ($result_{mht}$) with dimension $1 \times t$. Since this single vector contains information about the signal of all sensors at one it is robust to changes of fetal heart signal location originating from fetal movements. R-peaks are identified by a top-down peak search that commences with the highest maximum point in the transformed dataset and continues with the next maximum with the minimal distance

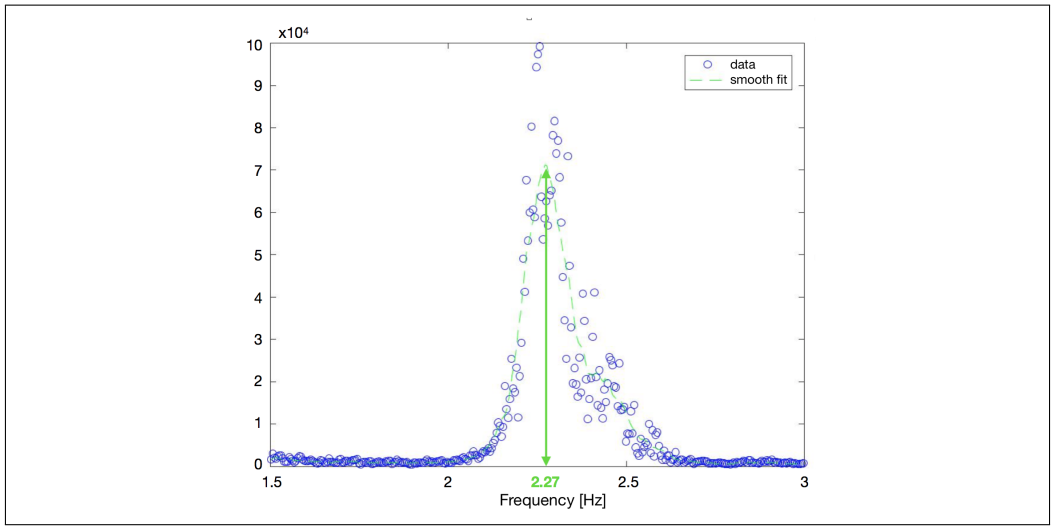

Figure 4.2.: Frequencies between 1.5 and 3 Hz of one dataset, after removing the maternal heart activity. In this case $f_{nat} = 2.27$ which matches $RR_{main} = 0.44$ and a mean heart rate of about 136 bpm.

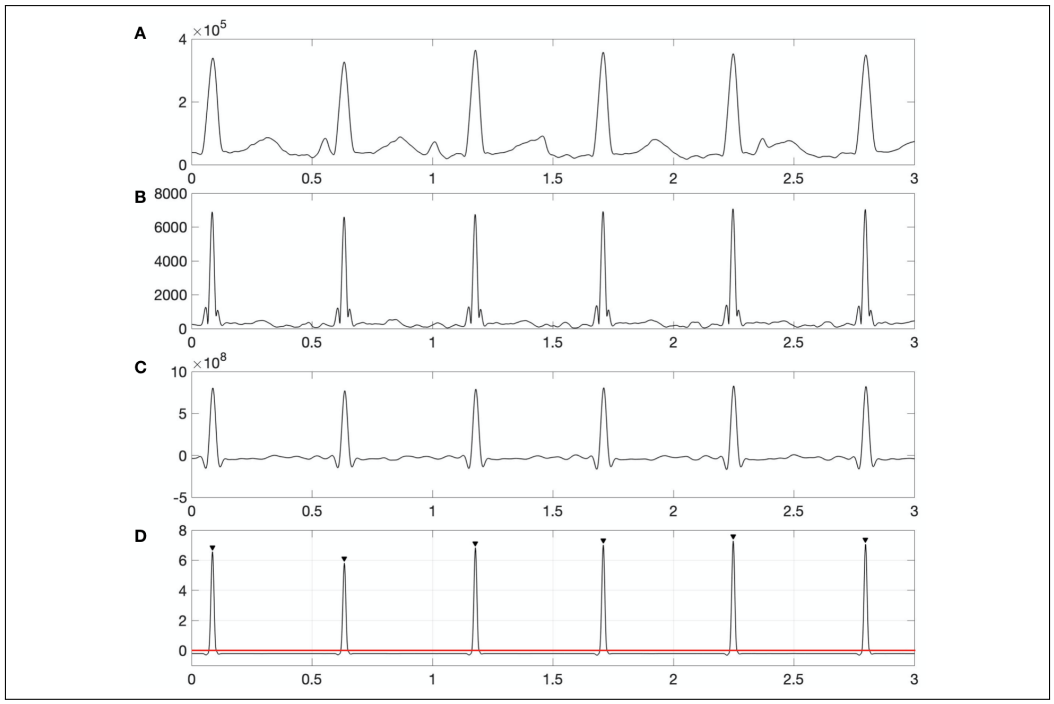

Figure 4.3.: A three-second excerpt from the result of **A** the hilbert transformation ($result_{all}$), **B** the result of the root mean square analysis ($result_{rms}$), **C** the result of the template cross-correlation ($result_{tcc}$) and **D** the z-scored combined result ($result_{all}$). Peaks of the combined signal have to exceed the threshold of zero (red line).

(d_{min}) to the previous one. This minimum distance is defined by the result of the frequency analysis: $d_{min} = RR_{main} * 0.7$. Factor 0.7 is chosen since it is assumed that the heart rate does not deviate more than 30% from the main RR-interval RR_{main}. Subsequently, the average over the whole dataset at the time points of all detected heartbeats (P_h) is calculated and k sensors with the highest absolute R-peak value are selected ($k = 20$ for maternal, $k = 5$ for fetal MCG).

Root mean sqare analysis and template generation after the signal of the selected k sensors is taken, (a matrix with dimension $k \times t$), the root mean square over the k sensors at each sampling point t is calculated resulting in a vector ($result_{rms}$) with dimension $1 \times t$. Peak detection is performed with the same parameters as in the previous step (Hilbert transformation). The result is averaged over all peak times (P_v) to generate one template for each of the k sensors. Each template has the length of the average RR-interval, with 40% of the time before and 60% after the R-peak. This procedure is adopted to ensure that the characteristics of the P, Q, S and T wave are also visible (see Fig. 4.4).

Template Matching After generating the template with the root mean square analysis, the cross-correlation (see Chapter 2.4.1) between the template and the signal is calculated for each of the k sensors individually. These k cross correlations are summed up to a general template cross-correlations vector ($result_{tcc}$) with dimension $1 \times t$.

Combination of results To combine the results of the previous three analysis steps, the product of their resulting curves is calculated:

$$result_{all} = result_{mht} * result_{rms} * result_{tcc} \tag{4.5}$$

A final peak search is performed on the z-scored $result_{all}$ vector, using the minimum peak distance d_{min} and a minimum peak height of 0, what means that a peak has to exceed the x-axis to be detected (see Fig. 4.3). The identified R-peaks (P_{found}) are then used for further analysis.

4.1.2.3.1. Interpolation of Missing Peaks By dividing all the resulting RR-intervals by RR_{main}, the gaps in the R-peak detection can be easily identified. The number of missing beats is subsequently calculated by dividing the size of the gap by the mean heart rate before and after the gap. The gaps are initially filled by dividing the length of the gap by the number of missing beats (m) + 1 and create m artificial peak time points. Since such an artificial filling would have a profound effect on measures like the heart rate variability, the algorithm searches for additional peaks of $result_{all}$ in the time window of $\pm$ 20 ms around the artificial time point. These artificial time points are then replaced by the highest peak in this time window. The final number of detected R-peaks in FLORA (P_{FLORA}) is the sum of detected peaks (P_{found}) and filled/modulated peaks.

4.1.3. Comparison to Standard Methods

To evaluate the performance of FLORA, the algorithm was compared to both the previously used methods Semi Automated Template Matching and Hilbert Transformation Approach, as well as to the combination of both methods.

Semi Automated Template Matching (SATM) as implemented in the Data Editor Software which is the standard software provided by the fMEG hardware manufacturer (VSM MedTech Ltd.) (see Chapter 2.5.1.1). The user is requested to mark a regular heart beat manually which is subsequently used as a template for template-matching R-peak detection. The resulting R-peaks (P_{SATM}) are exported to a marker-file.

Hilbert Transformation Approach (HTA) This approach by Wilson et al. [96] uses Hilbert transformation (see Chapter 2.5.1.2) on data of the 10 sensors with the highest mean power spectral density (see Chapter 2.4.5). On the resulting signal a threshold based peak search is performed to detect the R-peaks.

Combination of SATM and HTA (COMB) For a standard analysis of our fMEG studies, both methods are always applied one after the other to get the best R-peak detection for the current dataset. The result in which more R-peaks are detected is usually chosen for further evaluation. Using these standard methods, an additional visual check is always necessary to ensure that a higher number of peaks is not caused by arbitrarily added peaks.

For our sample of 55 selected datasets, the evaluation by an expert found SATM to be more precise for 21 datasets and HTA better for the other 34 datasets. This combination is denoted as COMB below.

4.1.4. Evaluation Metrics

Four metrics were used to evaluate the detection methods which are explained in more detail in the following sections.

Number of Detected Peaks per Minute (NP)

The number of R-peaks found in the dataset is the first indicator for the quality of the algorithm. Although in general, less peaks are found in critical datasets, particularly when using the HTA it could also be that fetal movements are misinterpreted as R-peak which in turn would lead to higher number of peaks and lower signal to noise ratio. Nevertheless, the number of peaks for fetal analysis should produce results in the normal range of fetal heart rate (120 - 160 bpm).

Difference between RR Measures (RR-DIFF)

A mean heart rate of healthy fetuses between 120 and 160 bpm, and results in a normal RR-interval between 0.375 and 0.5 seconds. The histogram of a heart rate recording should therefore resemble a Gaussian distribution of the RR-interval with a peak within that range. The more gaps there are in the analysis, the higher the mean RR-interval will be, while the more additional peaks the algorithm finds, the lower the mean RR-interval will be. Due to the fact that the mean RR-distance is dependent on the arousal and gestational age of the fetus, the main natural RR-interval (RR_{main}) is calculated by extracting the main frequency f_{nat} using a frequency analysis of the data (see Fig. 4.2). The more reliable the R-peak detection works, the lower the difference between RR_{main} and the mean RR-interval RR_{est} of the estimated heart rate will be.

$$RR_{main} = \frac{1}{f_{nat}} \tag{4.6}$$

$$RR_{diff} = |RR_{main} - RR_{est}| \tag{4.7}$$

Percentage of Normal to Normal Intervals (PNN)

The percentage of normal-to-normal is the amount of successive normal RR-intervals. 'Normal' in this case is defined as all RR_{est} values between $RR_{main}/2$ and $2 * RR_{main}$. Since the normal-to-normal and other parameters of the heart rate are generally used for heart rate variability analysis, a high percentage of normal to normal intervals is important for a valid heart rate variability result. While in the RR_{diff} value many smaller gaps do not weigh as heavily as one very big gap, the opposite applies with regard to the percentage of normal-to-normal.

Signal to Noise Ratio (SNR)

Although the above-mentioned measurements give us an indication about the reliability of the algorithms, they cannot guarantee that the R-peaks added by FLORA correspond to real R-peaks or that they are just arbitrary added points instead. Therefore, the signal to noise ratio of the averaged RR-peaks (see Fig. 4.4) of the raw data is calculated, where amp_{peak} describes the amplitude of the data at the R-peak, which is then divided by the median of the amplitude over the whole average sample. Precisely located R-peaks would result in a high signal to noise ratio, whereas arbitrary added R-peaks would result in a lower amplitude at the time of the R-peaks amp_{peak}, in more noise over the whole averaged sample (higher $median(amp)$) and therefore in a lower signal to noise ratio.

$$SNR = \frac{amp_{peak}}{median(amp)} \tag{4.8}$$

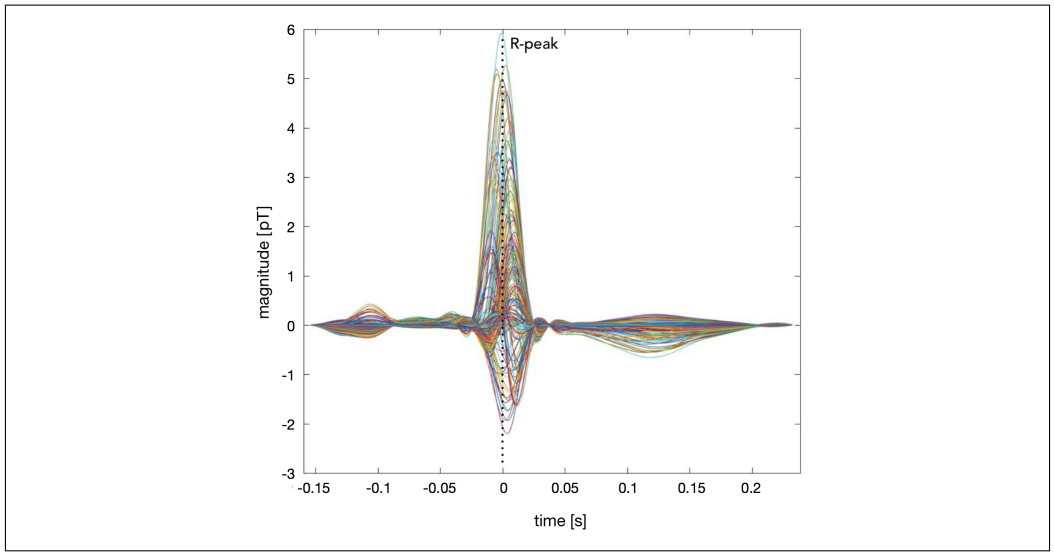

Figure 4.4.: Shape of averaged fetal heart beats over one dataset. Time point 0 refers to the detected R-peaks (the moment of highest magnetic activity during one heartbeat).

4.1.4.1. Statistics

The results of SATM, HTA, COMB and FLORA were tested for normal distribution using a One-sample Kolmogorov-Smirnov test. Since they were not normally distributed, SATM, HTA and COMB were compared with FLORA in all four metrics (number of peaks, RR-difference, percentage of normal-to-normal and signal to noise ratio) by using a Wilcoxon Signed Rank Test (see Chapter 2.4.7). To decide whether a result is significant, the significance levels were adjusted by Bonferroni correction to $p = \frac{0.05}{3}$. Due to extreme outliers in the results, it was decided to report [median $\pm$ standard deviation] instead of the mean.

4.2. Results

FLORA algorithm was analyzed for maternal and fetal R-peak detection. Maternal R-peak detection is easily performed and works very well for all three algorithms. Hence, there were no obvious differences and the following results relate to fetal R-peak detection only.

4.2.1. Number of Detected Peaks per Minute

A comparison of the number of detected peaks per minute as obtained from the four different methods showed a significant difference between SATM [132.67 $\pm$ 36.86] and FLORA [139.17 $\pm$ 23.15] as well as between HTA [137.40 $\pm$ 41.05] and FLORA. No significant difference is observed between COMB [138.35 $\pm$ 29.24] and FLORA (see Fig. 4.5).

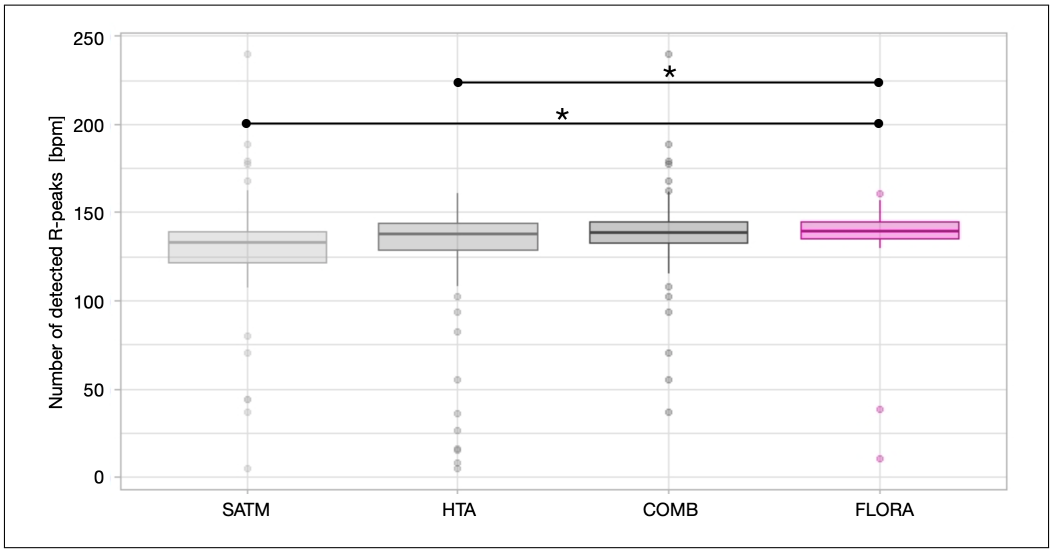

Figure 4.5.: Number of R-peaks, identified per minute.
* denotes significant differences in the Wilcoxon signed rank test ($p < 0.05$), Bonferrroni corrected for multiple comparison.

4.2.2. Difference between RR Measures

The difference between RR_{main} and the mean RR-interval RR_{est} of the estimated HR also showed a significant difference between SATM [0.01 $\pm$ 1.57] and FLORA [0.001 $\pm$ 0.72] as well as HTA [0.005 $\pm$ 1.96] and FLORA. Again, no significant difference is found between COMB [0.004 $\pm$ 0.20] and FLORA (see Fig. 4.6).

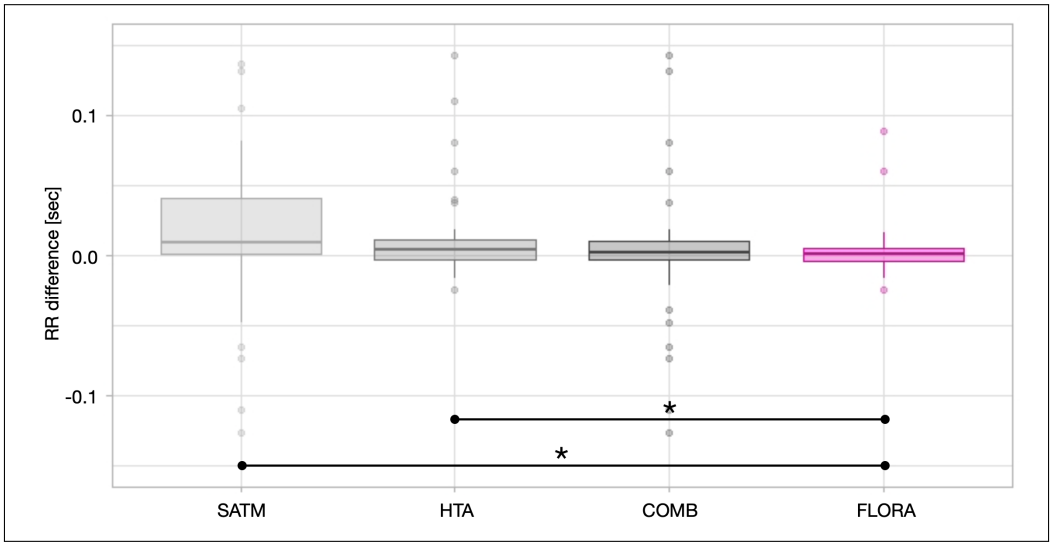

Figure 4.6.: The difference between RR_{nat} and the averaged RR-interval RR_{est} as estimated by the different R-peak detection methods.
* denotes significant differences in the Wilcoxon signed rank test ($p < 0.05$), Bonferrroni corrected for multiple comparison.

4.2.3. Percentage of Normal to Normal Intervals

The percentage of normal to normal intervals showed significant differences between FLORA [100 ± 5.81] and all other methods, SATM [98.48 ± 19.26], HTA [99.97 ± 23.16] and COMB [99.85 ± 14.57] (see Fig. 4.7).

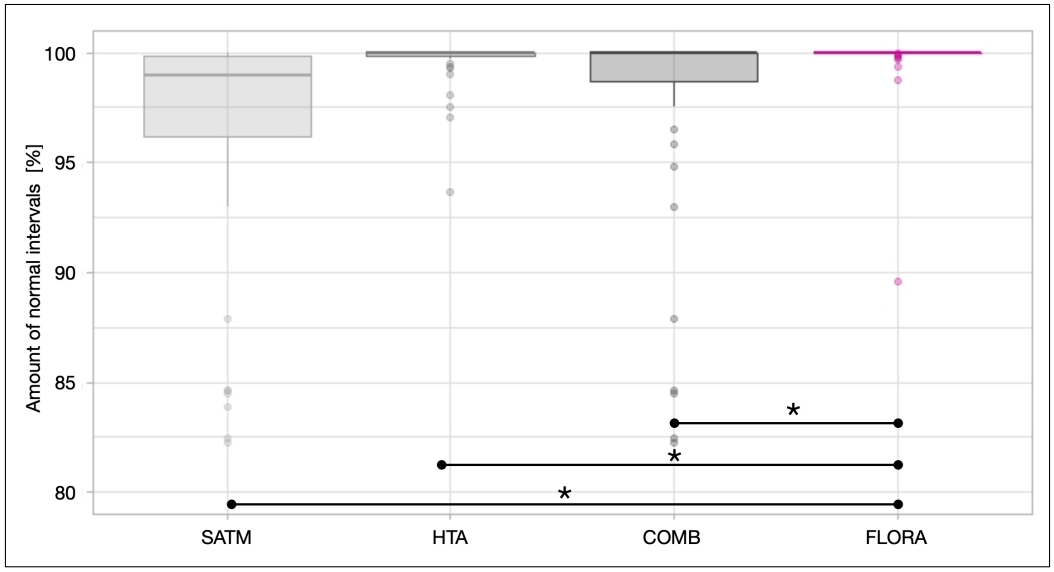

Figure 4.7.: The percentage of normal RR-intervals.
* denotes significant differences in the Wilcoxon signed rank test ($p < 0.05$), Bonferrroni corrected for multiple comparison.

In Figure 4.8 the percentage of normal to normal intervals with regard to gestational age is displayed. HTA shows a significant positive correlation with the gestational age [R = 0.35, p = 0.01]. Correlation values for FLORA and SATM are smaller and not significant. This shows that HTA results in a lower percentage of normal to normal intervals, for fetuses with lower gestational age.

Figure 4.9 displays the percentage of normal to normal intervals in relation to the measurement duration. HTA [R = 0.42, p = 0.002] and SATM [R = 0.28, p = 0.04] both show a significant positive correlation with the measurement duration. Correlation value for FLORA is smaller and not significant. This shows that, both HTA and SATM results in a lower percentage of normal to normal intervals in shorter measurements.

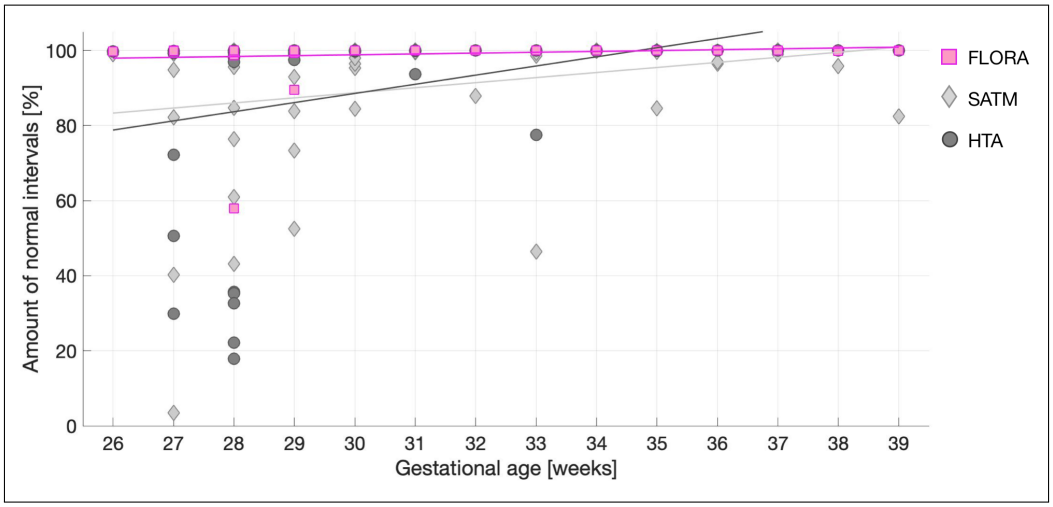

Figure 4.8.: The relation of the gestational age of the fetus and the amount of normal RR-intervals in percent (PNN). Each dot represents the PNN of one dataset. The colored lines emphasize the correlation between the PNN and the gestational age of the fetus.

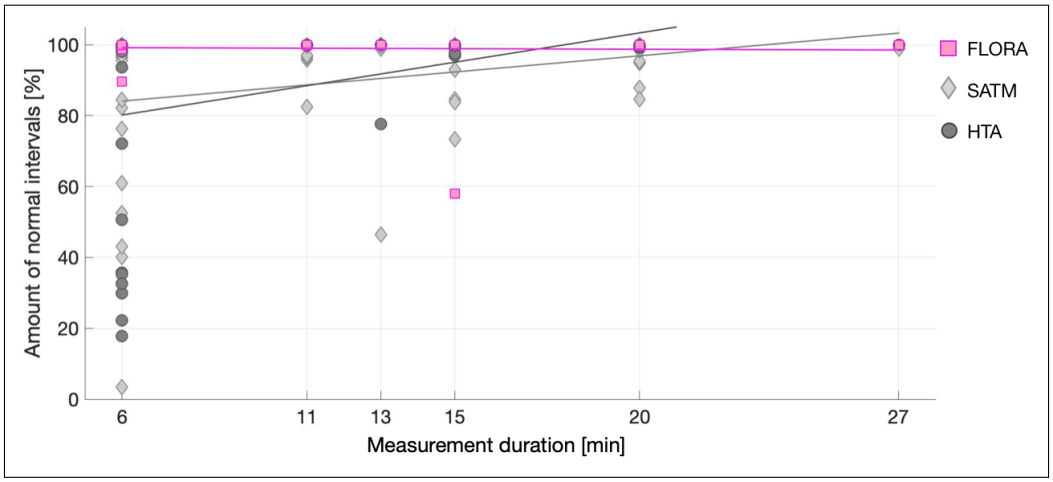

Figure 4.9.: The relation of the measurement duration and the amount of normal RR-intervals in percent (PNN). Each dot represents the PNN of one dataset. The colored lines emphasize the correlation between the PNN and the duration of the measurement.

4.2.4. Signal to Noise Ratio

The signal to noise ratio of FLORA [20.38 $\pm$ 4.50] is significantly different to signal to noise ratio of HTA [17.83 $\pm$ 6.48]. No significant difference is shown between FLORA and SATM [20.03 $\pm$ 6.62], or COMB [19.34 $\pm$ 5.43] (see Fig. 4.10).

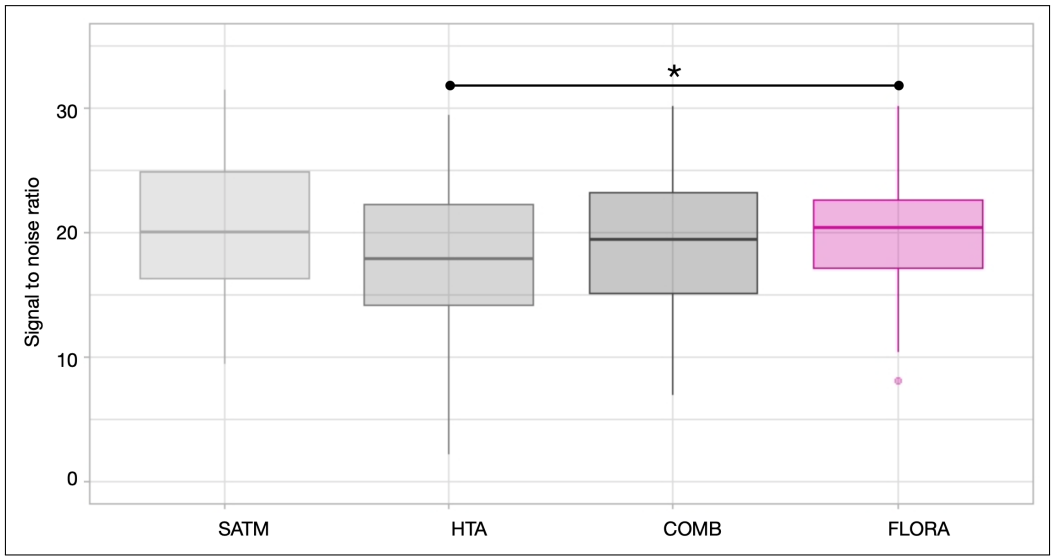

Figure 4.10.: The signal to noise ratio shows how clearly the R-peak contrasts from the rest of the averaged heart signal.
* denotes significant differences in the Wilcoxon signed rank test ($p < 0.05$), Bonferrroni corrected for multiple comparison.

4.3. Discussion and Conclusion

Since FLORA was developed to improve and simplify R-peak detection in fetal magnetoencephalography, its performance was tested on 55 different real datasets and compared with the currently established procedures. Four different evaluation metrics were used to compare performance between the different approaches.

Our results show that FLORA not only performs just as well as the established methods (SATM and HTA and their combination COMB) but also that it has some advantages over these. FLORA is significantly more accurate in number of detected peaks than SATM and HTA and the difference between RR measures of FLORA is also lower than both of the previously used methods. Even if there is no significant difference in these measures between FLORA and COMB, in percentage of normal to normal intervals, the results of FLORA are significantly different from those of the standard methods SATM and HTA, as well as of their combination COMB. Based on the results on percentage of normal to normal intervals we suggest that FLORA works reliably over the whole range of gestational ages and it is independent from the measurement duration.
FLORA resulted in a significantly higher signal to noise ratio than HTA indicating that, despite the higher number of detected peaks, the peaks derive from heart beats that were actually detected and not from randomly added points.

In sum, FLORA works user-independently which is very important for reproducibility. By removing noisy sensors, estimating the physiological heart rate of the subject, optimizing the peak search, filling gaps and modulating the interpolated peaks, FLORA generates high quality heart rate datasets that can be easily post-processed. Finally, due to its automatization, FLORA can also be used for batch-processing of a large amount of datasets.

5. FAUNA

This chapter describes the **f**ully **a**utomated s**u**btractio**n** of heart **a**ctivity (FAUNA). Parts of this chapter has already been published in the proceedings of the *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5685-5689 [102].

The evaluation of fetal brain activity is even more challenging since the fetal brain signal is superimposed by fetal and maternal heart activity which have a signal strength 10-1000 times larger. Thus, it is mandatory to remove this heart activity before the analysis of fetal brain activity. The detailed characterization of the interfering sources is only possible with multisensor systems covering a large part of the maternal abdomen. A widely used method for heart activity removal in fMEG data is the Orthogonal Projection (OP) [97, 96].
OP separates heart activity from other activity in a dataset by attenuation of the estimated signal space of the heart signal. This process is fairly effective in removing the maternal and, in most cases, also the fetal heart signals but it has its limitations. One drawback of the method is a possible redistribution of the signal [100], which can lead to inconsistent localization and involves the risk that some brain activity is removed together with the heart signals. Since the fMEG signal is only present in a small number of sensors, the identification of these sensors has to work reliably to make further automated evaluation steps possible.

5.1. Methods

First the proposed algorithm is explained step by step. Subsequently it is compared to the current standard method orthogonal projection (see Chapter 2.5.2.1) and verified in its functionality.

5.1.1. FAUNA

FAUNA is an algorithm for a **f**ully **a**utomated s**u**btractio**n** of heart **a**ctivity without the issue of redistribution. It is inspired by the frequency-dependent subtraction method [103] but without the step of individual sensor selection. The idea behind both, the frequency-dependent subtraction and FAUNA is to estimate the pure heart activity and subtract it from the original dataset.
The procedure of heart removal is done twice (for the fetal datasets) to remove both the maternal and fetal heart activity. In neonatal datasets a single cycle is sufficient since there is no maternal heart activity included.

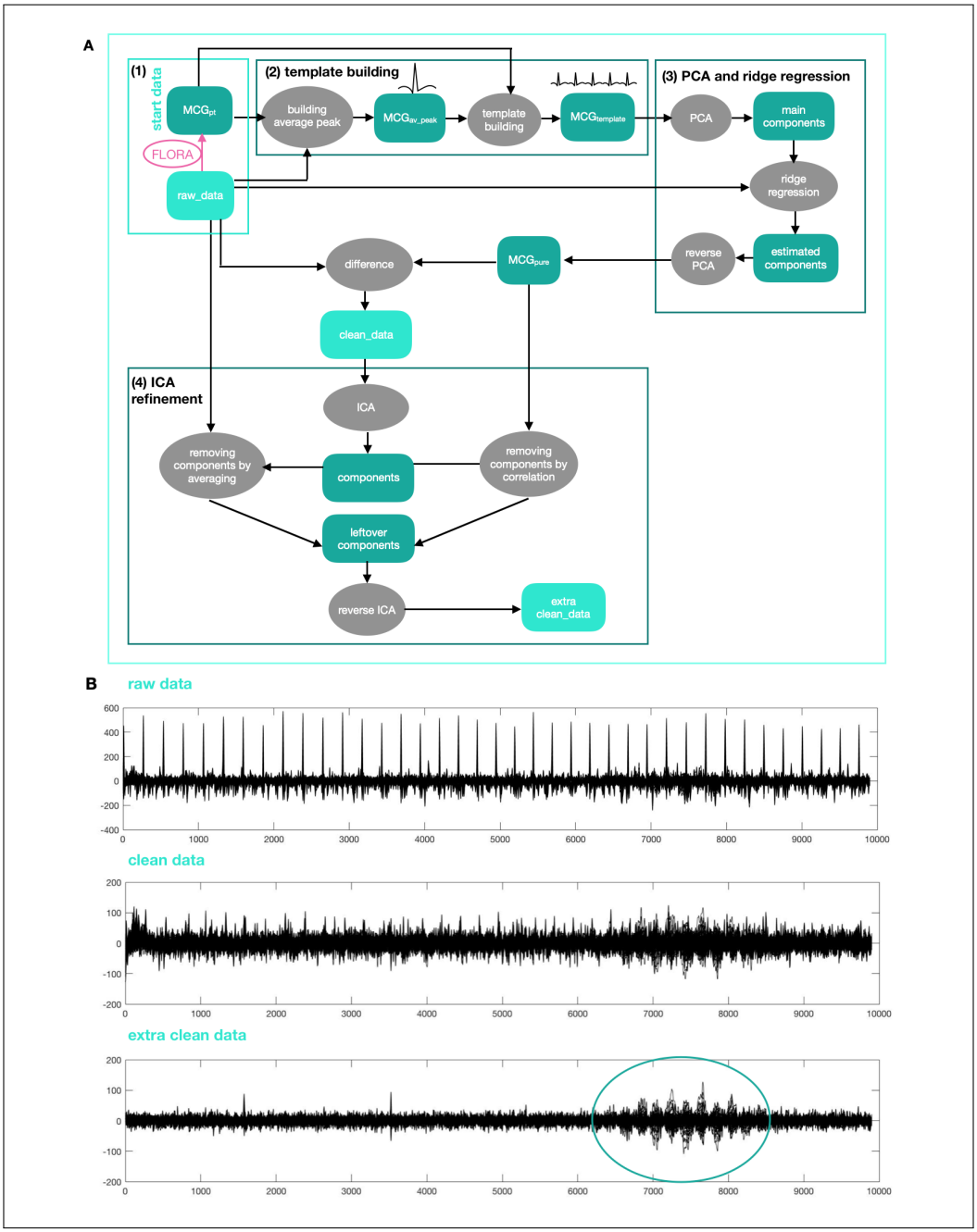

Figure 5.1.: Sketch of the FAUNA algorithmic procedure **A** including the preprocessing with FLORA (1), the building a heart beat template (2), the PCA and Ridge regression (3) and the ICA refinement (4). Subfigure **B** shows the same dataset as raw data at the beginning of step (1), as clean data after step (3) and as extra clean data after step (4). The circle in the extra clean data highlights a brain activity burst which is only visible after performing all steps of FAUNA. ©2019 IEEE

Before performing the heart activity removal, a detection of the R-peaks is necessary. FAUNA uses the result of an external automated R-peak detection algorithm (FLORA, see Chapter 4) that computes the maternal and fetal R-peaks [101]. In the first step of the FAUNA procedure, the time points of these R-peaks are used to generate a template of the heart activity. By using the average heart beats for the template the risk of removing anything else than the heart activity is minimized. In the second step several components of the heart activity are selected and estimated from the original dataset and removed afterwards. The last step divides the remaining signal into independent components and additionally removes the components that correlate with the heart activity.
An overview of the different steps of FAUNA is shown in Figure 5.1. The steps are described in the following paragraphs.

Building a Heart Beat Template

Firstly, the raw dataset is filtered using a second order butterworth filter from 1-35 Hz. Thereafter, an average heartbeat is built using the average over all R-peaks (MCGpt) for each sensor individually. By concatenating this average heart beat template with the distance of the original RR-intervals, an artificial pure heart signal is built, again for each sensor individually (see Fig. 5.1 A(2).

Principle Component Analysis and Ridge Regression

To uncover the signal characteristics of the heart rate, a Principal Component Analysis is performed on the artificial heart signal and the 4 main components are selected (for fetal heart the number of main components is reduced to 3).To build a model that considers signal variations, a ridge regression is trained to estimate these 4 components based on the original dataset. For each component, a separate unscaled ridge regression (k = 0.01) is calculated.

$$\hat{\beta}_i = ((X^T X + kI)^{-1} X^T y_i) \tag{5.1}$$

X is the original data, y_i the component and $\hat{\beta}_i$ is the estimated weight vector (see Chapter 2.4.2). The resulting ridge regression models serve as a spatial filter [104], which is a linear combination of all input sensors, with the aim of extracting the heart component as close to the template as possible while reducing the noise and all other activity that is not related to that component. The spatial filter allows to extract the heart components from the MEG signal with a good signal-to-noise ratio. In contrast to a template-based approach, extracting the heart components by a spatial filter has the benefit that dynamic variations of the heart components are accounted for. As the heart components are also modeled for the time points where the R-peak detection failed to detect a peak, those peaks can be visible in the spatially filtered signal and can be used to fill up missing R-peaks. Afterwards, the estimated heart components are used for a reverse Principal Component Analysis to transform them back to a dynamic estimation of the heart signal for all sensors (see Fig. 5.1 A(3)), which then is subtracted from the original dataset.

Refinement with Independent Component Analysis

Since the former procedure removes a large amount but not all heart signal components, in the next step an additional Independent Component Analysis (see Chapter 2.4.4) is performed on the resulting dataset. First, the correlation (see Chapter 2.4.1) of each component and the pure heart signal is calculated and second, the average of each component at the R-peak time points is generated. 40% of the components with the highest correlation and average components that reach a threshold of 1 were also removed from the remaining dataset. By reversing the Independent Component Analysis with the leftover components, the dataset without interfering heart activity is generated (see Fig.5.1 C).

5.1.2. Data Generation Model

A model with real and simulated data was used. Real data were collected by a 156 sensor system (SARA, SQUID Array for Reproductive Assessment, VSM MedTech Ltd., Port Coquitlam, Canada) at the University of Tübingen with a sampling rate of 610 Hz. The model data were adapted to this system and to this sampling rate.

As the real fetal brain activity is unknown in real datasets and the aim is to extract the fetal brain activity as well as possible, we do not have a ground-truth of the fetal brain activity in a real dataset. Therefore, we combined artificially generated fetal brain activity with real fMEG data. We generated a dataset consisting of real fMEG background activity and heart signals, generated artificial fMEG data based on a forward model [92, 94] and superimposed that artificial fMEG data with real maternal magneto-cardiography (mMCG) signals, real fetal magneto-cardiography (fMCG) signals and real background noise. To extract these real mMCG and fMCG signals for the model, a dataset was selected where an Independent Component Analysis could separate multiple mMCG and fMCG components. A reverse Independent Component Analysis was performed on six mMCG components and on two fMCG components to generate the maternal (Fig. 5.2 A,E) and fetal heart signal (Fig. 5.2 B,F) included in the model. The artificial fetal brain signal was put into the time trace at specific trigger time points (see Fig. 5.2 C,G). Triggers were set with a random distance of 10-15 sec. An empty fMEG measurement was performed to generate the background noise (Fig. 5.2 D,H).

After generating an artificial dataset with this model, the resulting dataset was processed by Orthogonal Projection (OP) and FAUNA.

5.1.3. Evaluation

The evaluation of fMEG signals is usually made over the whole time course of a recording or on data averaged over a specific trigger. To cover both options we first analyzed some of our evaluation parameters over the whole recording time, and secondly over data averaged over all fMEG simulation triggers. Evaluation was made after heart activity removal on the model dataset with OP and FAUNA respectively. To show the activity distribution over

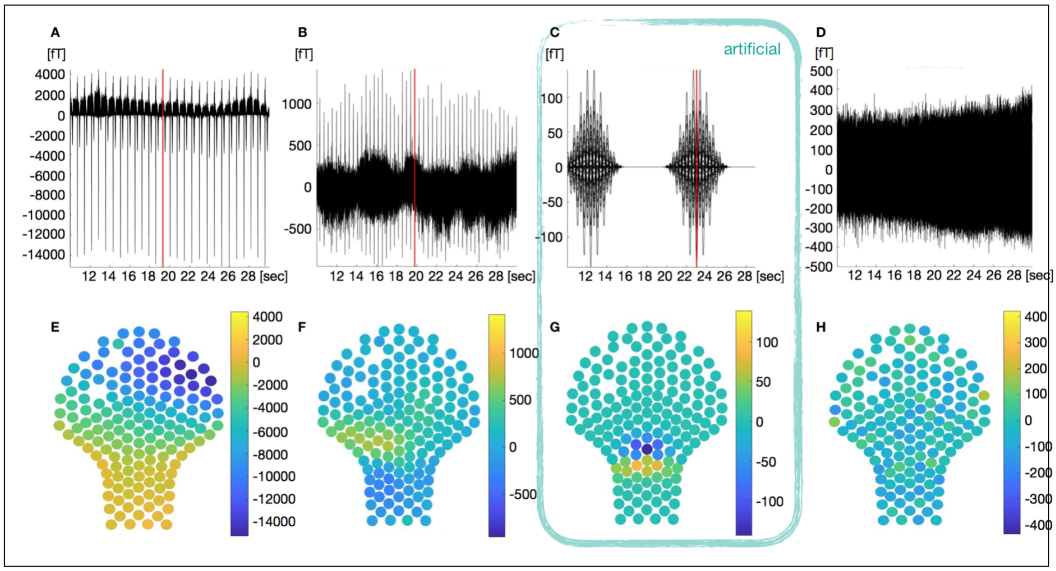

Figure 5.2.: The first row shows the magnetic activity, averaged over all sensors, in a segment of 20 sec duration for A) maternal heart activity, B) fetal heart activity, C) the artificial fetal brain activity and D) noise. The second row represents the magnitude of all sensors at the time point of the red line for E) maternal heart activity, F) fetal heart activity, G) the artificial fetal brain activity and H) noise. ©2019 IEEE

the sensors, the root mean square (RMS) activity was calculated for each sensor s.

$$RMS(s) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}|x_n|^2} \tag{5.2}$$

N is the total number of samples over the whole recording time and x_n is the value of the activity at time point n.

5.1.3.1. Root Mean Square Difference over Sensors

The difference of the simulated fetal brain activity $fMEG_{sim}$ and the leftover activity $fMEG_{left}$ after processing with each method was calculated for each data sample. Then the root mean square difference (RMSD) was calculated for each sensor s individually.

$$RMSD(s) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}|x_n - y_n|^2} \tag{5.3}$$

N in this case is the number of samples over the whole recording time, x is the value of the simulated fetal brain activity, and y the leftover activity after processing with OP or FAUNA. We compared the root mean square difference to the simulated signal for both, all the sensors and a selection of the 10 sensors where the magnitude of the simulated brain signal was the highest.

5.1.3.2. Correlation Analysis

To evaluate how much of the simulated brain signal is left in the data after the removal of maternal and fetal heart signals, the correlation coefficient between the simulated brain signal and the remaining data is calculated. This calculation is done on the whole time course and on the data averaged over all the fMEG simulation triggers.

5.1.3.3. Signal to Noise Ratio

The signal to noise ratio was calculated by dividing the root mean square of the simulated brain signal $fMEG_{sim}$ by the root mean square difference between the simulated brain signal $fMEG_{sim}$ and the remaining signal $fMEG_{left}$ after removing the heart activity. This was done for each of both methods (OP and FAUNA) once for the whole time course and once for the data averaged over all fMEG simulation triggers.

$$SNR = \frac{RMS(fMEG_{sim})}{RMS(fMEG_{left} - fMEG_{sim})} \tag{5.4}$$

5.1.3.4. Statistics

Results from all the above mentioned metrics were compared for OP and FAUNA by using a Wilcoxon Signed Rank Test (see Chapter 2.4.7) since the results were not normally distributed. This comparison was done for spontaneous fMEG activity and fMEG activity averaged over all the fMEG simulation triggers. Three sensors were excluded during the preprocessing with FLORA [101] and thus were excluded for statistical comparison in all methods to have an equivalent number of sensors. Results are reported in [mean $\pm$ standard deviation].

5.2. Results

5.2.1. Root Mean Square Difference over Sensors

Comparing the root mean square difference for all sensors resulted in significant ($p < 0.001$) lower values for FAUNA [1.26$\pm$0.23] compared to OP [1.38$\pm$0.13] (see Fig. 5.3 A-D).
Comparing the root mean square difference only for the 10 sensors with the highest activity in the simulated fMEG signal also showed a significant difference (p=0.019531) between OP [1.00$\pm$0.16] and FAUNA [0.67$\pm$0.10] (see Fig. 5.3 A-C red circles and E). The root mean square difference here is also significantly lower for FAUNA.

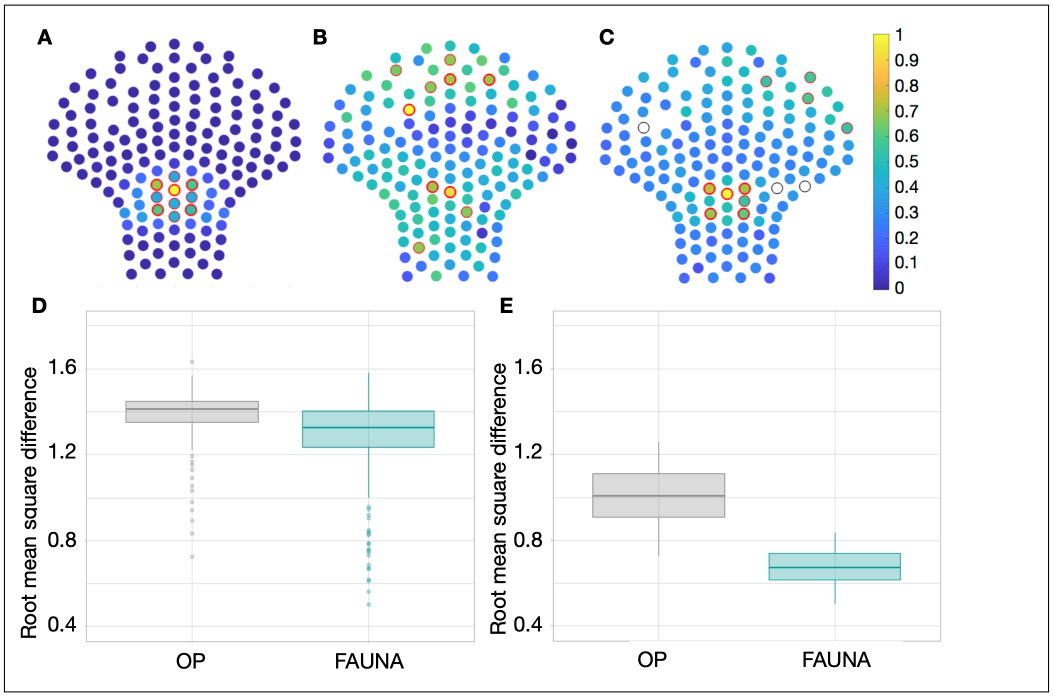

Figure 5.3.: First row: The normalized root mean square over the whole recording time for all sensors for A) the simulated signal, B) the signal after processing the heart activity removal with OP and C) the signal after processing the heart activity removal with FAUNA. The sensors with the 10 highest root mean square values are marked with a red circle, top 5 with a bold red circle. Second row: Boxplot of the root mean square between the actual fetal brain activity and the reconstructed signal by either OP or FAUNA for D) all sensors E) 10 sensors with highest activity of the simulated fMEG signal. ©2019 IEEE

5.2.2. Correlation Analysis

A comparison of the correlation over the whole time course showed a significant difference ($p < 0.001$) between OP [0.11±0.13] and FAUNA [0.21±0.22] (see Fig. 5.4 A-C,G).
Comparing the correlation for the averaged data also showed a significant difference ($p < 0.001$) between OP [0.50±0.30] and FAUNA [0.69±0.28] (see Fig. 5.4 D-F, H). Correlation is significantly higher for FAUNA in both cases.

5.2.3. Signal to Noise Ratio

A comparison of signal to noise ratio over the whole time course for all sensors showed a significant difference ($p < 0.001$) between OP [0.09±0.21] and FAUNA [0.25±0.47] (see Fig. 5.5 A).
A comparison of signal to noise ratio on the averaged dataset for all sensors showed a significant difference ($p < 0.001$) between OP [0.50±1.10] and FAUNA [0.78±1.20] (see Fig. 5.5 B).
Signal to noise ratio is significantly higher for FAUNA in both cases.

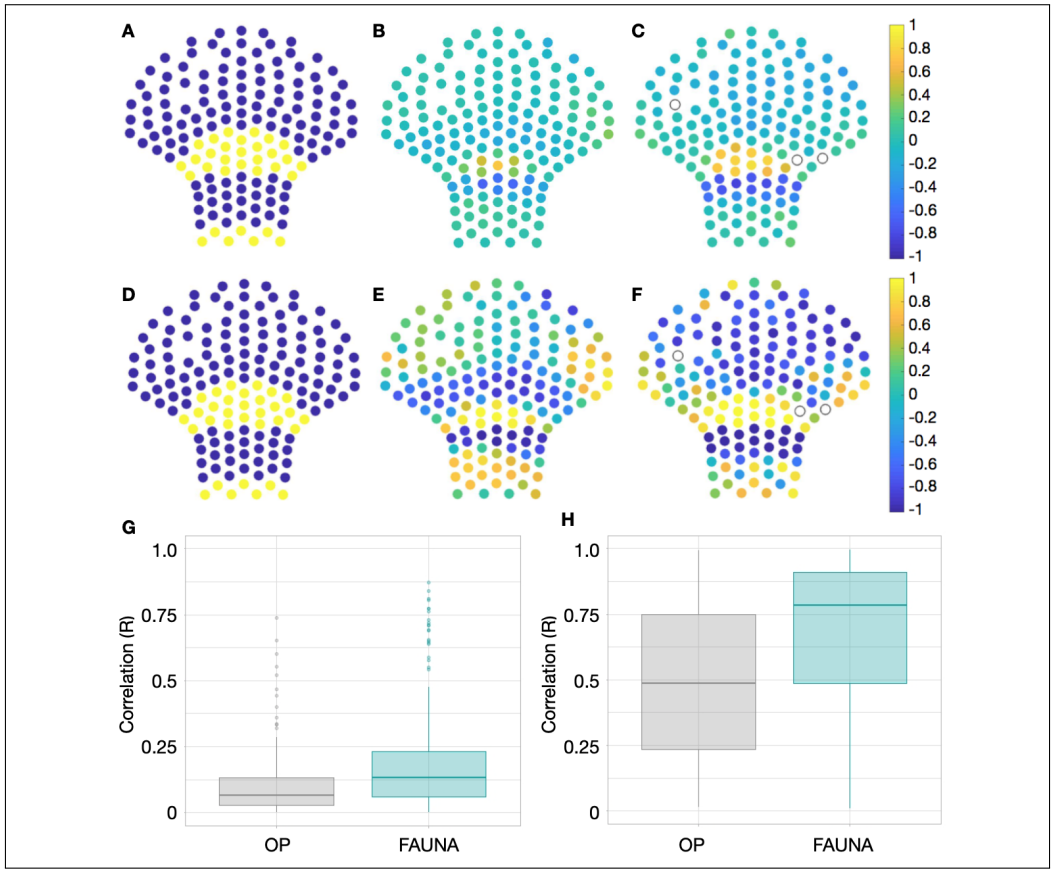

Figure 5.4.: First row: correlation of the original fMEG signal over the whole recording time for A) simulated signal B) OP and C) FAUNA. Second row: correlation between the original fMEG signal averaged over all fMEG simulation triggers for D) simulated signal E) OP and F) FAUNA. Third row: boxplot of the correlation with the original fMEG signal for all sensors over G) the whole time of the recording and H) data averaged over all fMEG simulation triggers. ©2019 IEEE

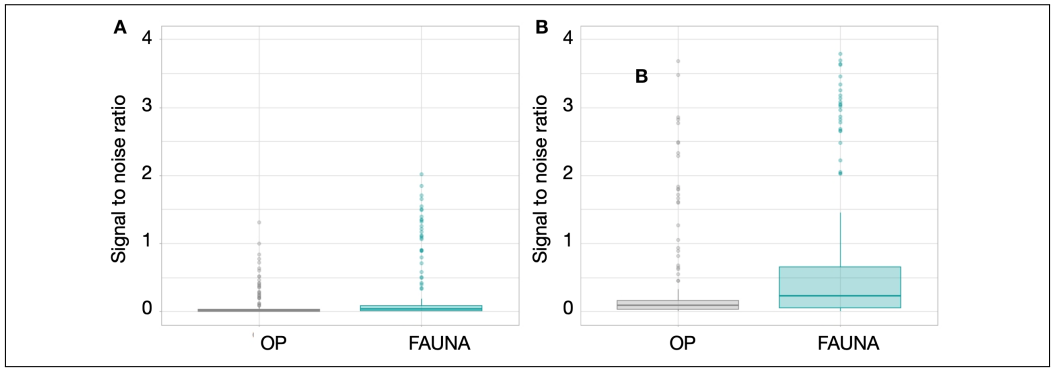

Figure 5.5.: Boxplots of signal to noise ratio for all sensors over A) the whole time measurement and B) data averaged over all fMEG simulation triggers. ©2019 IEEE

5.3. Conclusion

Until now, orthogonal projection (OP) was a well established method for heart activity removal in fMEG data. In this work we introduced our new algorithm for Fully AUtomated subtractioN of heart Activity (FAUNA) and compared both methods by using a model of real heart and simulated brain activity to evaluate both methods on completeness and redistribution. Therefore, we used the root mean square difference, the correlation coefficient and the signal to noise ratio as criteria.
The significantly higher correlation values between FAUNA and the simulated signal showed that more brain signal is left after removing heart activity with FAUNA than with OP.
This indicates that by using methods based on orthogonal projection some brain signal is removed along with the heart signal. The impact of the redistribution is visible when looking at the root mean square difference for all sensors. The remaining fMEG signal after heart activity removal with FAUNA shows a cluster of high root mean square values in the same region where the original simulated fetal brain signal was located. Also five of the ten sensors with the highest root mean square values are the same. In the remaining fMEG signal, after heart activity removal with OP, there is no visible cluster and only two out of the five sensors with highest root mean square values are congruent with the simulated fetal brain signal (see Figure 5.3 A-C). The findings from the root mean square difference support both these observations, showing a significantly lower difference for FAUNA. The signal to noise ratio over all sensors is significantly higher for FAUNA, compared to OP, for both the spontaneous and the averaged brain activity. This is in accordance with the findings from the correlation analysis, which also showed that after processing with FAUNA there is generally more brain signal left.
In conclusion, we developed and successfully tested a new algorithm for the removal of heart activity from fMEG data. Compared to a well established method, FAUNA provides better results in terms of signal to noise ratio and does not have the drawback of a redistributed signal. As FAUNA is also fully automated, it is a superior alternative to the currently used methods for the removal of heart signals. Using FAUNA for heart activity removal, the remaining signal containing fetal brain activity should form clusters of high root mean square values, even for spontaneous brain activity where the data contains a lot more noise than in averaged brain activity. Such clusters of high amplitude activity can easily be identified by automated procedures.

6. AURORA Study

Inside the maternal womb the fetus experienced a relatively uniform environment, constantly surrounded by amniotic fluid and sounds of the maternal intestines and the maternal voice. After birth, the world for the newborn is unsteady and the neonate has to adapt to his new environment. Recognizing the mothers voice after birth is an important factor for mother-child bonding. The maternal voice can have a calming effect on the newborn and inspires confidence, what is important for the newborn child to adapt to the environment outside the maternal womb. Studies showed that newborns even in the first hours of life react different to the voice of their mother than to other female or male voices (see Chapter 2.3.1) and also studies in fetuses showed that there are also differences in reaction to the maternal voice compared to the fathers or other female voices. That leads to the assumption that the ability of voice discrimination is already learned in utero.

6.1. Study Design

6.1.1. Goal of the Study

In this study we wanted to investigate the fetal and neonatal **au**tonomous and central nervous **r**esponse t**o** mate**rn**a**l** voice (AURORA). We wanted to assess the fetal/neonatal heart rate, actocardiogram and heart rate variability during stimulation with maternal voice compared to a strange female voice or silence. Furthermore we want to complete this with the analysis of the power spectral density of the brain signal during these periods as well as event-related brain activity to the onset of maternal/stranger voice.

Hypothesis about the Heart Rate

Based on the studies of Kisilevsky et al. [62, 63, 64], we expected no difference in the heart rate between the stimulation with maternal and stranger voice before 31 weeks gestational age. From 32 weeks of gestational on we expected an initial decrease of the heart rate in the first 30 seconds after voice onset for both stimulations and subsequent increase of the fetal heart rate especially during a stimulation with maternal voice.

Hypothesis about the Power Spectral Density

We expected generally higher values during stimulation compared to silence, especially in the alpha and theta frequency bands, since theta is correlated with novelty [105] and alpha with attention [106], what was in former studies suggested to be the reason for the initial heart rate drop. Since in studies with newborn children a familiar voice was associated

with a change in delta frequency band [69, 70], we expected differences in the same band while fetuses listen to mother or stranger voice. Moreover those differences should start around 31 weeks gestational age and increase over gestation.

Hypothesis about the Auditory Responses

We assumed that the fetus shows an auditory event-related response to both, the maternal and the stranger voice starting at 26 weeks of gestation. We also assumed that the fetus at this stage of development is not yet able to distinguish the voice of his mother from other female voices and there is no difference between auditory response to maternal and stranger voice before 31 weeks gestational age. Finally, we expected an increasing amplitude of the response to the mother's voice according to evidences in newborn [71, 72].

6.1.2. Study Population

For this study we examined 41 singleton pregnant women older than 18 years, with German as mother language. Exclusion criteria for the fetal measurements were: smoking, drug or alcohol abuse of the mother, high-risk pregnancy or abnormal fetal development. Additional exclusion criteria for the neonatal measurements were: complications during birth and bad condition of the newborn.

6.1.3. Protocol

6.1.3.1. Audio Recording

The voice of each participating woman was recorded while reading Grimms "Dornröschen" (see Appendix B). This recording took part in a sound proof room to ensure a good sound quality. Recordings were made by using a H4next handy recorder attached to a condenser microphone. The condenser microphone was placed in a 45°angle with $\approx$ 20 cm distance to the mouth of the women (see Fig. 6.1). The women were instructed to read with normal reading speed as reading to a child but without adjusting voices when reading direct speech. Before and after turning over the pages they should make a little break, same when they wanted to drink something. If they made a reading-mistake they were instructed to say "nochmal" ("again") and repeat the whole sentence. Thereafter, the audio files were edited using Audacity software version 2.0.5.0. First, the recording was converted from stereo to mono sound. Then, eventual interferences and disturbances (drinking, page turns, coughing, reading errors etc.) were cutted out. Thereafter 4 cutouts (Text 1-4) of exactly 3 min length were prepared as stimuli for the fMEG sessions. The cutouts for all women started at the same text passages:

1. Vorzeiten lebten ein König und eine Königin, ...
2. Der König, der sein liebes Kind vor dem Unglück gern bewahren wollte, $\cdots$
3. In dem Augenblick aber, wo sie den Stich empfand, $\cdots$
4. Es ging aber die Sage in dem Land von dem schönen schlafenden Dornröschen $\cdots$

but due to individual reading speed the text segments varied in the number of words between women. Each 3 min segment was normalized to a peak amplitude of -1.0 dB after removing the DC offset. On the last 3 seconds of each 3 min segment a "fade out" - effect was applied to avoid a reaction to interrupted speech.

Pictures courtesy of the University Hospital Tübingen.

Figure 6.1.: Setting of the audio recording in the sound proof room. The text is placed in front of the woman on the table, the condenser microphone is placed beside the table in a 45°angle with $\approx$ 20 cm distance to the mouth.

6.1.3.2. Fetal Measurements

To see the changes over the course of pregnancy, we scheduled four age groups for fetal measurements, according to the gestational age (26th - 31st; 32nd - 34th; 35th - 37th and 38th - 42nd week). For each participant one measurement in each of these four age groups was planned (see Fig. 6.2).

6.1.3.3. Neonatal Measurement

The final measurement took part between the second and the eighth week of life. The newborn was intended to be in a good mood and sated, in order to avoid too much movement during the measurement, because movement would negatively influence the quality of the recorded data.

6.1.4. Recording Paradigm

The duration of one recording is 27 minutes, consisting of 9 consecutive 3 min sections (5 sections of silence and 4 sections of auditory stimulation - 2 sections maternal and 2 sections stranger voice, see Fig. 6.3). The voices are always presented in alternating manner during one measurement. In each measurement two different text segments are

presented each in maternal and stranger voice. We generated a randomized list of protocols regarding the following:

- measurements start alternating with maternal or stranger voice
- text segment 1 is presented in each measurement
- in 3 consecutive measurements Text 2-4 are presented once, but in randomized order
- if the texts are presented in adjacent or alliterated order is chosen randomly
- which text is presented first is also chosen randomly

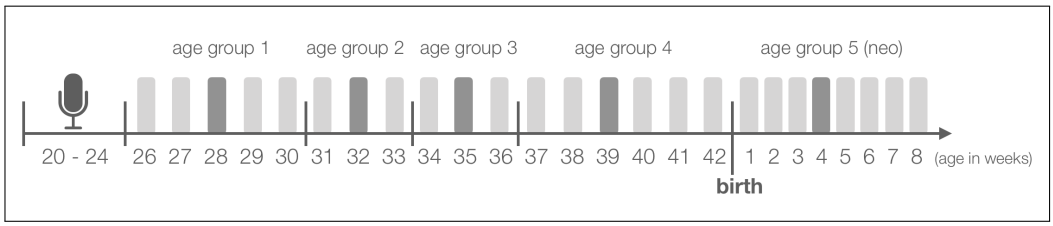

Figure 6.2.: Schematic measurements time course of the AURORA Study. One measurement per age group was planned for each woman. Additional measurements for longitudinal data collection were voluntary.

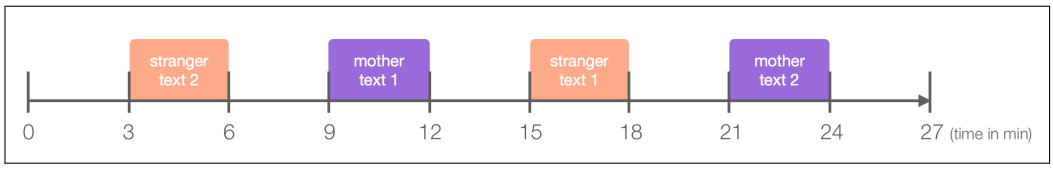

Figure 6.3.: Example Paradigm for Maternal Voice measurement. Audio recordings of two different text segments read by maternal (purple) or stranger voice (orange) were presented in random order. Duration of the recordings as well as the duration of the silent phases before/after the stimulations were three minutes.

The speech sound is produced by an audiobox outside the shielded room and then conducted through tubes into the shielded room. In the case of fetal measurement, the tubes end in an air-filled balloon placed on the mother's abdominal surface (see Fig. 6.4). In the case of measurement in newborns, the tubes end in a sound-shell specially designed for infants and placed on the left ear of the newborn (see Fig. 6.5).
The audio files used for stimulation were presented at a peak volume of 65 dB for newborns (measured at the ear) and a peak volume of 95 dB for fetuses (measured at the mother's abdominal surface). The maximum volume heard by the fetuses will therefore also be around 65 dB, since the volume is reduced by about 30 dB by the abdominal wall of the mother [46]. Thus the intensity of the stimulation is comparable with intrauterine noise [107].

6.1.5. Session

Each maternal voice session takes about 2 hours of time and usually takes place in the morning.

Fetal Session After arriving at the fMEG Center the pregnant woman is asked to change her clothes, because for an fMEG measurement a metal-free clothing is required. Thereafter she fills out the Profile of mood states (POMS) questionnaire and, in the first session also an anamnesis questionnaire together with our midwife. Then, a short ultrasound measurement is performed to assess the fetal position, the distance between the ear and the abdominal wall and the distance between the heart and the abdominal wall. Subsequently the woman is asked to take place on the SARA (SQUID Array for Reproductive Assessment, see Chapter 2.1.2).

Pictures courtesy of the University Hospital Tübingen.

Figure 6.4.: Pregnant woman on the fMEG device with auditory stimulation (green sound-balloon)

A small magnetic coil (head coil) is placed on the maternal abdomen near the fetal head. A strap with three more coils is placed around the belly of the woman to mark the left, right and back position of the abdomen. These coils can be used to locate the mother's position and the position of the fetal head in a 3 dimensional space.

The sound-balloon is placed between the maternal abdomen and the fMEG device to provide a good and stable sound quality to the fetus during the measurement. Another strap around the thighs and buttocks prevent the woman to slide off the SARA during the measurement. Pillows are placed around the woman until her seat is comfortable. Earplugs are placed in the ears of the woman to avoid that she hears the voices in the sound-ballon during the measurement. If the woman wants to hear music during the measurement this is also possible with these earplugs. If everything is ok, the midwife and the assistant leave the magnetically shielded room, close the door and start the recording. The pregnant woman is under video and audio surveillance during the whole measurement that communication is assured all the time. After 27 minutes of recording the midwife opens and enters the room, the straps and coils are removed and the woman can slowly leave the device. Subsequently, a second ultrasound is performed to recheck the position of the fetus and to estimate if the fetus moved during the measurement. Thereafter, the woman is asked to fill a questionnaire about the measurement where, among other questions, the movement of the fetus and the comfort of the woman are asked. Finally, before leaving, the pregnant woman can change her clothes again.

Pictures courtesy of the University Hospital Tübingen.

Figure 6.5.: Neonate with auditory stimulation. The sound is provided through a tube ending in an ear-shell which is placed on the left ear of the newborn.

Neonatal Session For the newborn measurements the procedure is a little bit different. After arriving at the fMEG center, the parents are asked to change the clothes of the baby if they are not metal-free. Then the parents have time to feed and calm the baby, because

for the success of the measurement it is best when the baby is sated and asleep. When this is the case, the baby is placed in a newborn cradle on the fMEG device, lying on the right side. On the left ear a sound-shell is then sticked where the stimuli during the recording are presented (see Fig. 6.5). During the neonatal recording one parent is allowed to stay inside the room to calm the baby if necessary. The parent is instructed to move slowly and as little as possible to not interfere with the recorded data. After the recording the clothes of the baby are again changed.

6.2. Methods

6.2.1. Data Sample

For the AURORA study a total of 231 measurements in 41 women were planned. For 39 women we planned 5 appointments during pregnancy, 2 women we measured longitudinally once a week. Additional appointments were made for newborns if, during the first appointment the newborn was to restless, so that measurement was not possible and the parents then wanted to come for another newborn measurement. In 23 of the planned appointments the mother was sick, already gave birth in the meanwhile or simply didn't show up. Of all 208 started measurements, 16 were not completed because in 8 case the mother was sick or dizzy and in the 8 newborn measurements the neonate was restless, so that the recording had to be interrupted. Finally 192 measurements were completed. In addition 6 datasets had to be excluded afterwards: all 5 measurements of 1 subject, because the child was diagnosed with spina bifida, and 1 dataset was corrupted. So in total 186 datasets from 40 women were left and used for further evaluations. All newborns had an 10 min APGAR score of 9 or 10 (mean=9.9, std=0.18) and no hearing disabilities were reported.

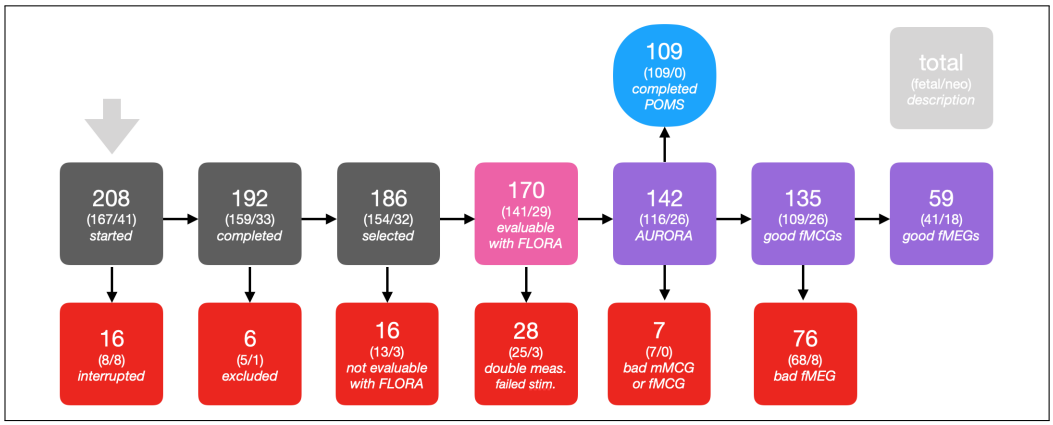

Figure 6.6.: Schematic representation of all data selection steps. Excluded datasets were represented in red. The numbers in brackets describe the amount of fetal/neonatal datasets. The 170 datasets that were evaluable with our methods are marked in pink. The final datasets that were used for the analysis of the AURORA study, fetal heart, fetal brain analysis were marked in purple. The amount of completed POMS questionnaires is represented in blue.

As shown in Figure 6.6, 135 dataset with good fMCG quality were preselected with the FAIRY tool. For the same 135 datasets the 10 channels with the highest RMS value were selected as brain activity. These channels were visually examined in FAIRY and 76 subjects were excluded because the selected channels showed pattern of muscular activity or maternal movement artifacts in the power spectral density plots.

Table 6.1.: Number of fMCG and fMEG datasets in each age-group

	age-group 1 26-31 weeks	**age-group 2** 32-34 weeks	**age-group 3** 35-37 weeks	**age-group 4** 38-42 weeks	**age-group 5** newborn
fMCG	n = 27	n = 29	n = 27	n = 26	n = 26
fMEG	n = 6	n = 11	n = 12	n = 12	n = 18

6.2.2. Fetal Heart Activity

6.2.2.1. Heart Rate

The timepoints of the R-peaks were transferred into the current heart rate and the mean heart rate for each second was calculated. Then, we selected the first 60 seconds after and 20 seconds before stimulus onset of each 4 stimulation intervals for each measurement. Since we were interested in change of heart rate after onset of maternal or stranger voice stimulation we defined the mean heart rate in the 20 seconds before stimulus onset as baseline and subtracted it. Then we averaged the two selections for maternal and for stranger stimulus onset respectively resulting in a 2 x 80 dimensional matrix for each measurement.

6.2.2.2. Actogram

Continuous actogram calculation is also based on the timepoints of the R-peaks detected by the FLORA Algorithm. For each R-peak the center of gravity was calculated by using the signal distribution over the sensors at the time-point of the R-peak. The magnitude of the signal was then multiplied with the X and Y coordinates of the sensors and resulted in an average X and Y 3D coordinate for the center of gravity. The actogram was calculated as the distance of the center of gravity between every two subsequent R-peaks and finally averaged for each second, resulting in the continuous actogram. Then, we selected the first 60 seconds after and 20 seconds before stimulus onset of the continuous actogram for each 4 stimulation intervals in each measurement. We defined the mean continuous actogram in the 20 seconds before stimulus onset as baseline and subtracted it. Finally, we averaged the two selections for maternal and for stranger stimulus onset respectively.

6.2.2.3. Heart Rate Variability

The total recording time was, according to the stimulation paradigm, separated into 9 different 3 minutes segments of which 5 were silent, 2 during stimulation with maternal voice and 2 during stimulation with stranger voice. The heart rate variability was calculated for each 3 minute segment individually as described in MatHusin et al. 2020 [32] using MATLAB 2016b. We performed our evaluation on the standard deviation between the normal RR-intervals (SDNN), because this is a well described parameter, representing the overall variability and used in many studies [108, 109, 16]. For each subject the standard deviation between the normal RR-intervals was averaged over all segments with same stimulation type (mother, stranger or silence) resulting in a 3 values for each measurement.

6.2.3. Fetal Brain Activity

6.2.3.1. Power Spectral Density

For each dataset the average power spectral density was calculated on the 10 preselected brain activity channels. Power spectral density was calculated for each of the 9 different 3 minute sections using the pwelch function in MATLAB 2016. Again for each type of stimulus ('mother', 'stranger', 'silence') one average power spectral density was built per dataset. Thereafter the mean activity of the delta (< 4Hz), theta ($4-8$Hz), alpha ($8-13$Hz), and beta ($13-30$Hz), frequency band was calculated for each dataset and stimulus type. Mean activities were transferred to logarithmic scale to ensure normal distribution.

6.2.3.2. Voice Onset Response

The current findings about the fetal ability to recognize the maternal voice are not consistent and the possible mechanisms underlying this ability are still unclear. Since we wanted to include multiple features of the maternal and stranger voice we calculated the voice onset response as a special form of auditory event-related responses. To calculate the voice onset response the selected brain activity was filtered with 1-10 Hz for fetal and 1-15 Hz for neonatal measurements. Since we had no single syllables but ongoing speech we averaged the fetal brain activity of each subject for every voice onset time-point after at least 500 ms of silence during both maternal and both stranger stimulation sequences, respectively. Averaging started 200 ms before voice onset time-point and ended 1s after voice onset time-point. The mean brain activity of the 200 ms period before voice onset time-point was regarded as baseline and subtracted in order to calculate the voice onset response.

6.2.4. Profile of Mood States Questionnaire (POMS)

The profile of mood states (POMS) was performed to preclude that a negative mood (like depression, anxiety or hostility) of the mother influences the reaction of the fetus. The questionnaire consists of 35 items naming different feelings and mood states that should be rated on a seven point scale from 0 (not at all / überhaupt nicht) to 6 (extremely / sehr stark) according to the current condition of the pregnant woman. We used the German version of the POMS (see Appendix C). There are four subscales in the German POMS:

- Niedergeschlagenheit / depression and anxiety (negative subscale), 14 items
- Tatendrang / vigor (positive subscale), 7 items
- Müdigkeit / fatigue (negative subscale), 7 items
- Missmut / hostility (negative subscale), 7 items

The scales were calculated as the sum of the item values, as proposed by Albani et. al ([110]).
Since questionnaires that were not completely filled were excluded and we did not provide the POMS questionnaire in the newborn measurements, a total of 107 questionnaires were left for evaluation.

6.2.5. Statistical Analysis

6.2.5.1. Permutation analysis

The significance of the continuous parameters were calculated by permutation analysis for each of the 5 age groups separately. We used a 10000 fold permutation analysis for continuous heart rate and continuous actogram and a 1000 fold permutation analysis for voice onset response, since the sample size was quite small for the brain parameters in each age group. Permutations were only made between different stimulation types (stranger and mother) for the same subject, not between subjects. In each permutation analysis the difference between mother and stranger was calculated. This difference was seen as significant when exceeding the 95% quantile and highly significant for exceeding the 99%quantile of the permutations.

6.2.5.2. Mixed Models and ANOVA

For power spectral density a linear mixed-effects model was generated in R. We investigated the main effect of stimulation type (mother/stranger/silence), age-group (1-5) and gender (m/f) and age-group*gender interaction. ANOVA was performed using Satterthwaite's method. For significant ANOVA results a post-hoc T-test was performed using the false discovery rate (FDR) to correct for multiple testing.

6.3. Results

6.3.1. Heart Rate

We could evaluate the heart rate, providing stable outlier free results, even for the first age group (see Fig. 6.7 (a)). We found that in age group 4 (term fetuses >37 weeks GA) there is a significantly higher heart rate during stimulation with stranger voice, compared to stimulation with maternal voice between 7.5-9.5 seconds and at 16.5 seconds after stimulation onset (see Fig. 6.7 (b)). The heart rate for the other 4 age groups showed no significant differences between mother and stranger stimulation.

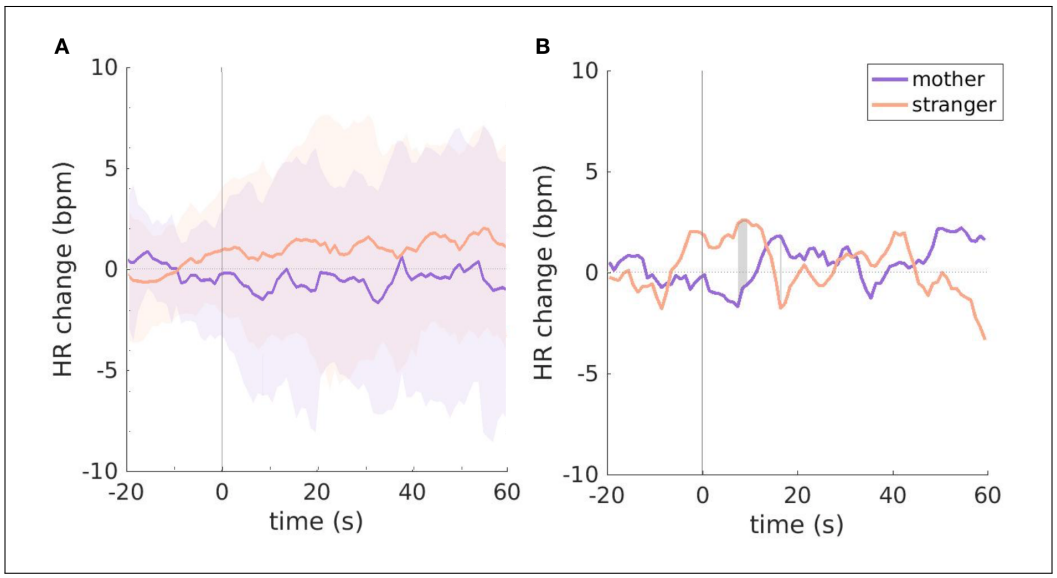

Figure 6.7.: Heart rate for **A** age group 1 (<32 weeks GA), mean responses with standard deviation (shaded areas) and **B** age group 4 (>37 weeks GA), mean responses with significant areas (grey).

6.3.2. Actogram

For the actogram we found a reduction in movement during stimulation with maternal voice in the first age group, getting significant between 8.5-11.5 seconds, 27.5-33.5 seconds, 38.5-40.5 seconds and 47.5-49.5 seconds after stimulus onset (see Fig. 6.8 (a)). Age group 3 showed higher movement activity at 8.5 sec, between 10.5-13.5 sec, between 34.5-35.5 sec and between 45.5-49.5 sec, getting highly significant between 46.5-48.5 sec after stimulus onset. In age group 5 (neonates) we also found significantly lower movement to stimulation with maternal voice compared to stranger voice between 5.5-9.5 sec and 25.5-28.5 sec after stimulus onset (see Fig. 6.8 (b)). Age-groups 2 and 4 did not show any significant differences in actogram between stimulation with maternal and stranger voice.

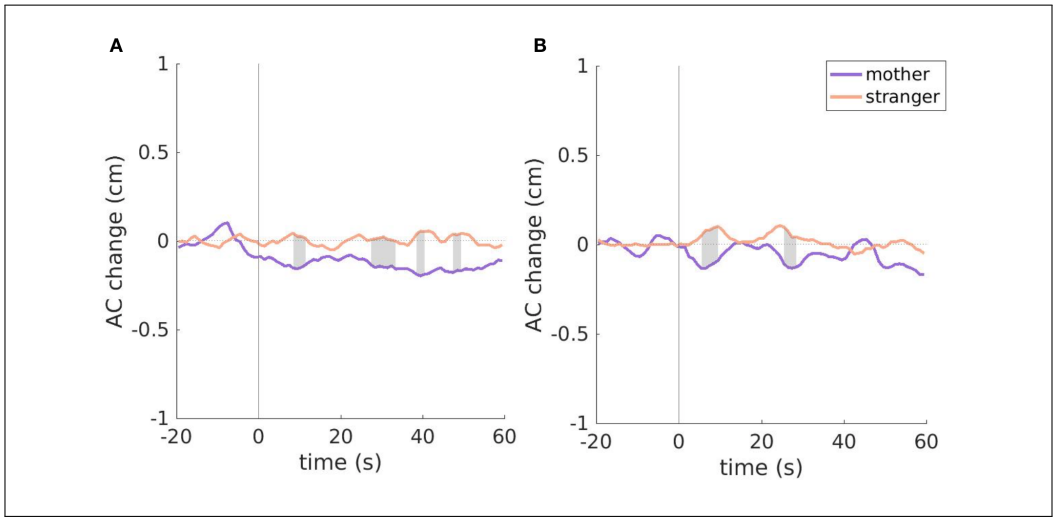

Figure 6.8.: Actogram with mean movement for **A** age group 1 (<32 weeks GA) **A** neonates. Significant differences are marked in grey.

6.3.3. Heart Rate Variability

We found no main effect for stimulation type or sex in the standard deviation between the normal RR-intervals (SDNN) which we used as measure for the heart rate variability. But we found a main effect of age-group (see Tab. 6.2 and Fig. 6.9). The SDNN in age-group 1 [mean = 8.97 , std = 2.35] is significantly lower than in all other age groups. The SDNN in age-group 5 (newborns) [mean = 16.78 , std = 4.84] is significantly higher than in all other age groups. Between the heart rate variability of age-group 2 [mean = 11.49 , std = 3.03] , age-group 3 [mean = 11.78 , std = 3.19] and age-group 4 [mean = 11.81 , std = 3.63] there is no significant difference (see Tab. 6.3).

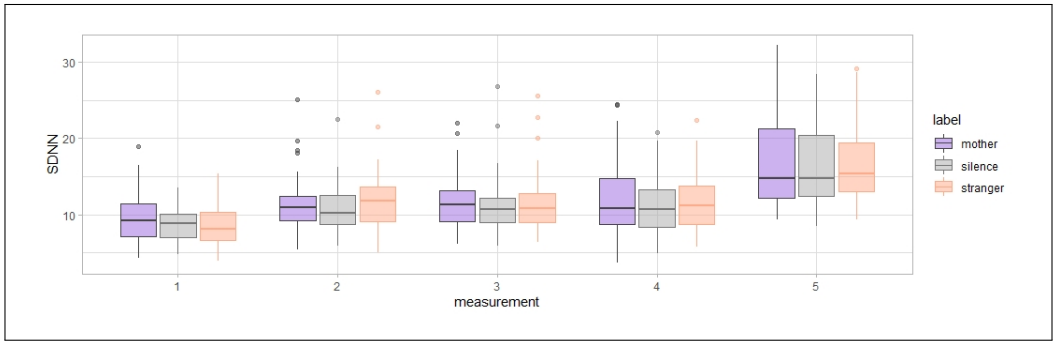

Figure 6.9.: Standard deviation between normal RR-intervals (SDNN) for all five age-groups.

Table 6.2.: ANOVA results for standard deviation between normal RR-intervals (SDNN).

	Sum Sq	Mean Sq	DF	F-value	P-value
label	23.00	11.50	2 / 338.69	1.3232	0.26766
sex	6.82	6.82	1 / 35.41	0.7851	0.38157
age-group	2448.71	612.18	4 / 346.65	70.4476	< 2e-16 ***
sex:age-group	73.77	18.44	4 / 346.65	2.1224	0.07762
label:sex	3.11	1.56	2 / 338.69	0.1791	0.83610
label:age-group	21.67	2.71	8 / 338.69	0.3117	0.96143
label:sex:age-group	24.99	3.12	8 / 338.69	0.3594	0.94116

Significances: ***<0.001, **<0.01, *<0.05.

Table 6.3.: Pairwise differences in standard deviation between normal RR-intervals (SDNN).

Contrast	Estimate	SE	DF	T-ratio	P-value
1 - 2	-1.8270	0.480	346	-3.803	0.0002 ***
1 - 3	-2.2379	0.500	348	-4.477	<.0001***
1 - 4	-2.1630	0.506	346	-4.278	<.0001***
1 - 5	-7.8900	0.498	348	-15.857	<.0001***
2 - 3	-0.4110	0.481	345	-0.854	0.4921
2 - 4	-0.3361	0.498	346	-0.674	0.5561
2 - 5	-6.0630	0.492	349	-12.334	<.0001***
3 - 4	0.0749	0.503	343	0.149	0.8817
3 - 5	-5.6520	0.502	348	-11.255	<.0001***
4 - 5	-5.7269	0.512	347	-11.175	<.0001***

Significances: ***<0.001, **<0.01, *<0.05.

6.3.4. Power Spectral Density

We found no main effect for stimulation type or sex in any of the four frequency bands but we found a main effect of age-group and a age-group*sex interaction for all of the four frequency bands (see Tab. 6.4 and Fig. 6.10). Pos hoc T-tests revealed that the differences were mainly due to the generally increased PSD in newborns but also showed differences between female and male fetuses in age-groups 1 and 3 (see Tab. 6.5 - 6.8 and Fig. 6.11).).

Table 6.4.: ANOVA results for power spectral density analysis

	Sum Sq	Mean Sq	DF	F-value	P-value
delta (1-4 Hz)					
label	0.00309	0.00155	2 / 137.211	0.0641	0.9380
sex	0.00755	0.00755	1 / 25.674	0.3128	0.5808
age-group	2.21677	0.55419	4 / 149.799	22.9576	8.263e-15 ***
sex:age-group	1.87429	0.46857	4 / 152.624	19.4108	6.266e-13 ***
theta (4-8 Hz)					
label	0.00070	0.00035	2 / 137.198	0.0220	0.9782
sex	0.00339	0.00339	1 / 25.599	0.2143	0.6474
age-group	1.60341	0.40085	4 / 150.650	25.3584	4.517e-16 ***
sex:age-group	1.16744	0.29186	4 / 153.674	18.4635	2.081e-12 ***
alpha (8-13 Hz)					
label	0.00038	0.000189	2 / 137.09	0.0162	0.9839
sex	0.00273	0.002730	1 / 25.50	0.2343	0.6325
age-group	1.09283	0.273207	4 / 150.51	23.4466	4.421e-15 ***
sex:age-group	0.81133	0.202833	4 / 153.53	17.4071	8.491e-12 ***
beta (13-30 Hz)					
label	0.00032	0.00016	2 / 136.938	0.0150	0.9851
sex	0.00152	0.00152	1 / 25.341	0.1421	0.7093
age-group	1.35558	0.33890	4 / 150.724	31.6012	< 2.2e-16 ***
sex:age-group	0.66806	0.16701	4 / 153.827	15.5737	1.017e-10 ***

Significances: ***<0.001, **<0.01, *<0.05.

Table 6.5.: T-test results for power spectral density analysis of **delta (1-4 Hz)**

Contrast	Estimate	SE	DF	T-ratio	P-value
1 - 4	0.12226	0.0540	144	2.266	0.0499 *
1 - 5	-0.20022	0.0629	156	-3.185	0.0044 **
2 - 5	-0.29780	0.0507	161	-5.876	<.0001 ***
3 - 5	-0.32042	0.0503	161	-6.369	<.0001 ***
4 - 5	-0.32248	0.0437	153	-7.377	<.0001 ***
f,1 - m,1	-0.6432	0.1536	72.6	-4.187	0.0004 ***
f,1 - f,5	-0.6042	0.0604	141.3	-10.009	<.0001 ***
f,1 - f,4	-0.2816	0.0623	140.5	-4.520	0.0001 ***
f,4 - f,5	-0.3226	0.0522	144.4	-6.180	<.0001 ***
m,1 - m,4	0.5261	0.0881	146.2	5.970	<.0001 ***
m,4 - m,5	-0.3224	0.0701	156.9	-4.597	0.0001 ***

Significances: ***<0.001, **<0.01, *<0.05.

Table 6.6.: T-test results for power spectral density analysis of **theta (4-8 Hz)**

Contrast	Estimate	SE	DF	T-ratio	P-value
1 - 5	-0.2192	0.0507	157	-4.325	0.0001 ***
2 - 5	-0.1897	0.0408	162	-4.650	<.0001 ***
3 - 5	-0.2571	0.0405	162	-6.349	<.0001 ***
4 - 5	-0.2948	0.0353	154	-8.361	<.0001 ***
f,1 - m,1	-0.38337	0.1209	76.2	-3.171	0.0062 **
f,3 - m,3	-0.30276	0.1028	46.5	-2.946	0.0125 *
f,1 - f,5	-0.47327	0.0488	141.7	-9.696	<.0001 ***
f,1 - f,4	-0.21019	0.0504	140.8	-4.171	0.0003 ***
f,4 - f,5	-0.26308	0.0422	145.1	-6.237	<.0001 ***
m,1 - m,4	0.36142	0.0712	147.0	5.076	<.0001***
m,4 - m,5	-0.32655	0.0565	158.2	-5.778	<.0001 ***

Significances: ***<0.001, **<0.01, *<0.05.

Table 6.7.: T-test results for power spectral density analysis of **alpha (8-13 Hz)**

Contrast	Estimate	SE	DF	T-ratio	P-value
1 - 5	-0.1922	0.0435	157	-4.415	0.0001 ***
2 - 4	0.1194	0.0363	158	3.285	0.0025 **
2 - 5	-0.1348	0.0350	162	-3.846	0.0004 ***
3 - 5	-0.2033	0.0348	162	-5.844	<.0001 ***
4 - 5	-0.2542	0.0303	154	-8.392	<.0001 ***
f,1 - m,1	-0.299351	0.1041	75.6	-2.877	0.0141 *
f,3 - m,3	-0.251792	0.0885	46.2	-2.844	0.0144 *
f,5 - m,5	0.074241	0.0806	34.0	0.921	0.4815
f,1 - f,5	-0.378956	0.0419	141.5	-9.042	<.0001 ***
f,1 - f,4	-0.178279	0.0433	140.7	-4.121	0.0003 ***
f,4 - f,5	-0.200677	0.0362	144.9	-5.540	<.0001 ***
m,1 - m,4	0.302276	0.0611	146.8	4.944	<.0001 ***
m,4 - m,5	-0.307640	0.0485	158.1	-6.337	<.0001 ***

Significances: ***<0.001, **<0.01, *<0.05.

Table 6.8.: T-test results for power spectral density analysis of **beta (13-30 Hz)**

Contrast	Estimate	SE	DF	T-ratio	P-value
1 - 5	-0.22851	0.0417	157	-5.480	<.0001 ***
2 - 3	0.07681	0.0311	146	2.472	0.0243 *
2 - 4	0.12945	0.0348	158	3.719	0.0006 ***
2 - 5	-0.15494	0.0336	162	-4.618	<.0001 ***
3 - 5	-0.23176	0.0333	162	-6.959	<.0001 ***
4 - 5	-0.28439	0.0290	154	-9.800	<.0001 ***
f,1 - m,1	-0.26087	0.0989	76.6	-2.638	0.0239 *
f,3 - m,3	-0.22244	0.0839	46.6	-2.652	0.0245 *
f,1 - f,5	-0.38654	0.0402	141.6	-9.617	<.0001 ***
f,1 - f,4	-0.16395	0.0415	140.7	-3.951	0.0005 ***
f,4 - f,5	-0.22259	0.0347	145.1	-6.409	<.0001 ***
m,1 - m,4	0.27571	0.0586	147.0	4.704	<.0001 ***
m,4 - m,5	-0.34619	0.0465	158.5	-7.445	<.0001 ***

Significances: ***<0.001, **<0.01, *<0.05.

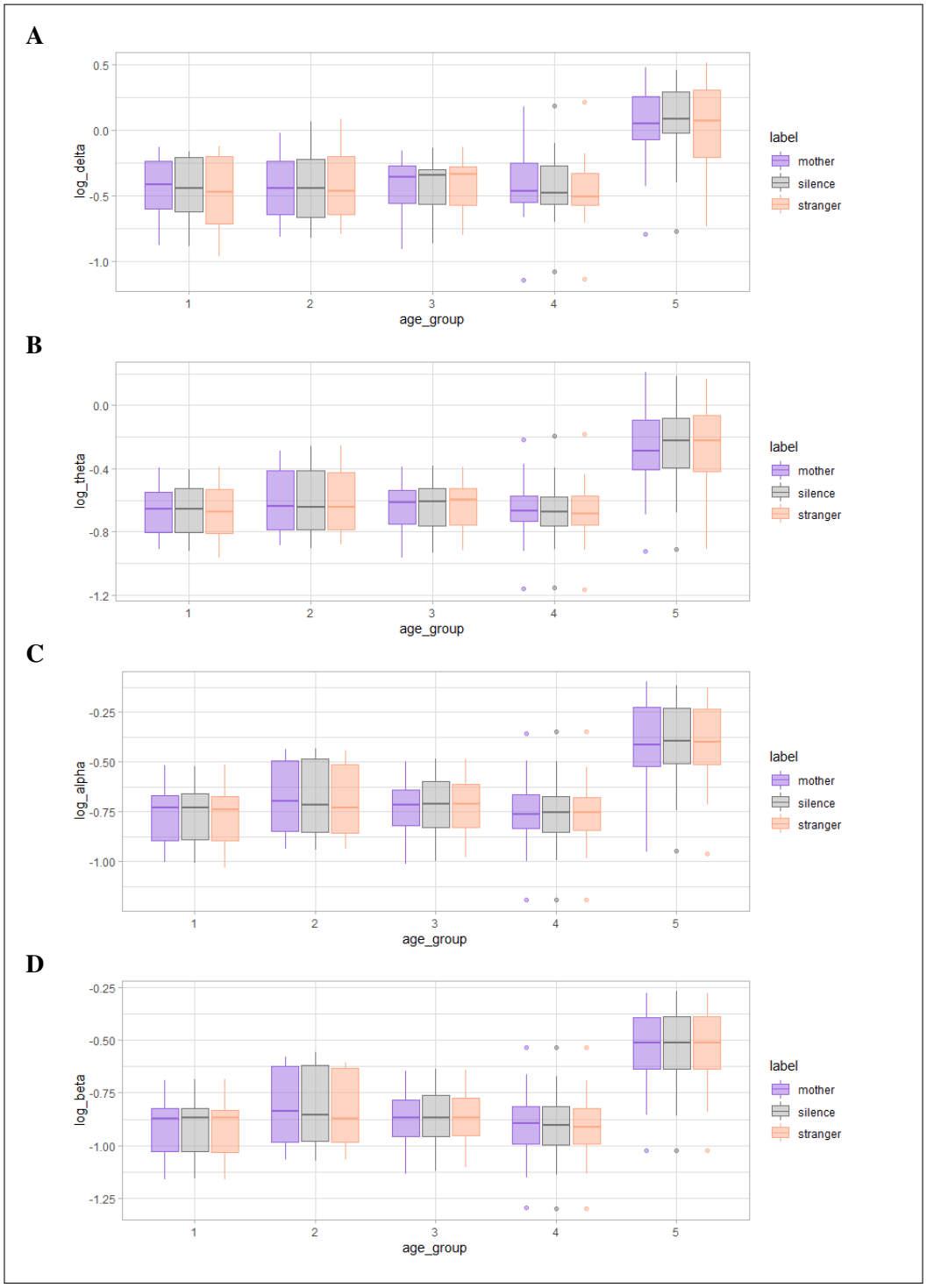

Figure 6.10.: Mean log10 PSD in different bandwidths for all three stimulation types and all 5 age groups for **A** delta band (< 4Hz), **B** theta band ($4 - 8$Hz), **C** alpha band ($8 - 13$Hz) and **D** beta band ($13 - 30$Hz) during silence (grey), stimulation with maternal (purple) or stranger voice (orange).

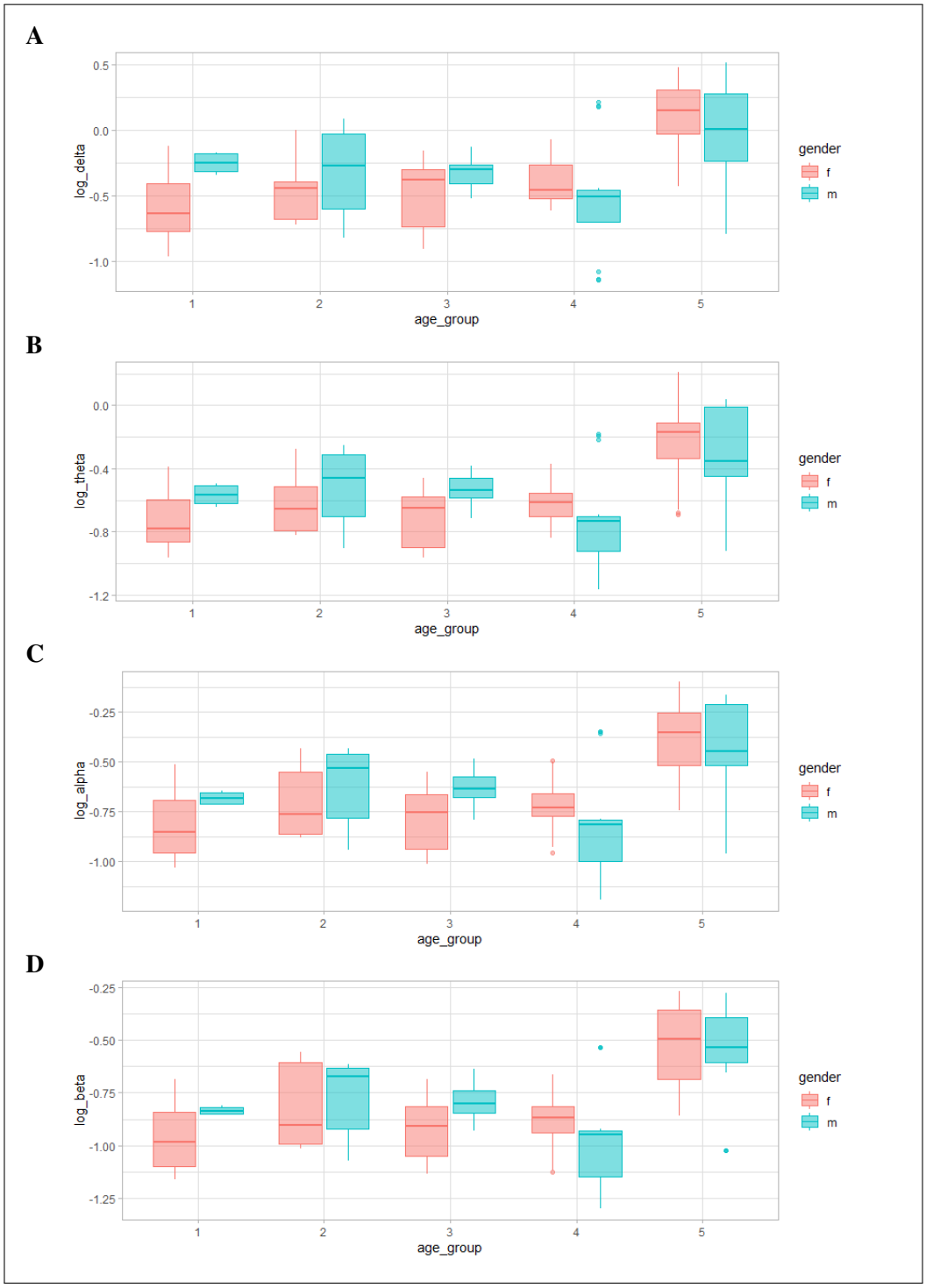

Figure 6.11.: Mean log10 PSD in different bandwidths for male and female in all 5 age groups for **A** delta band (< 4Hz), **B** theta band ($4 - 8$Hz), **C** alpha band ($8 - 13$Hz) and **D** beta band ($13 - 30$Hz) for male (blue) or female (pink) fetuses / neonates.

6.3.5. Voice Onset Response

In age-group 1, the reaction to maternal voice was significantly higher between 629 - 634 ms after voice onset (see Fig. 6.12 (a)). Age group 2 showed a higher reaction to stranger voice between 360 - 362 ms after voice onset. Age-group 3 did not show any significant differences between maternal and stranger voice. In age group 4 reaction to stranger voice was significantly higher between 11-19 ms, 821-847 ms and 863-872 ms. In age-group 5 (neonates) we found a significantly higher amplitude in auditory event-related response to stimulation with stranger voice compared to maternal voice between 87-204 ms, 590-606 ms and 641-683 ms after stimulus onset. Differences were highly significant($p<0.01$) between 100-140 ms and 649-652 ms (see Fig. 6.12 (b)).

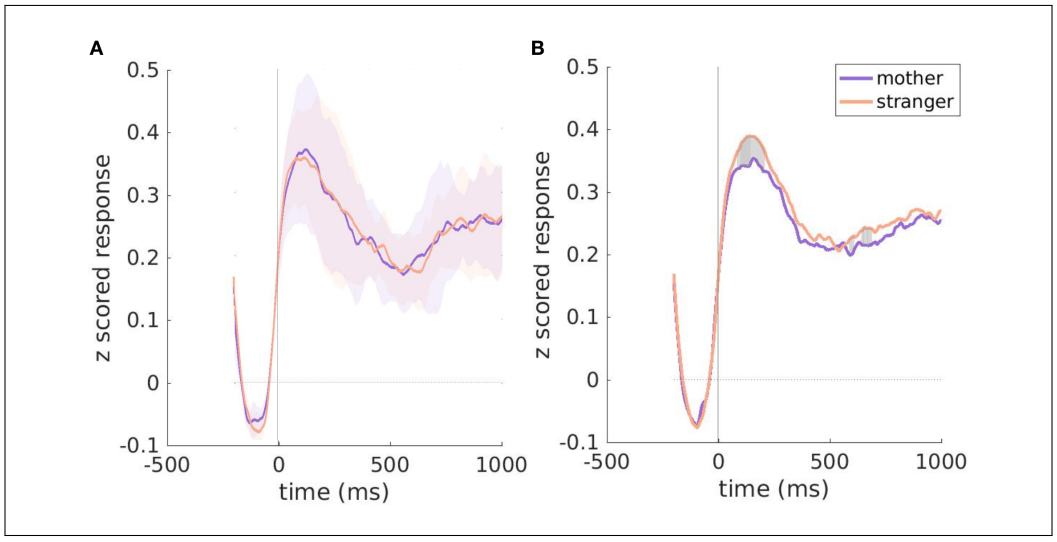

Figure 6.12.: Voice onset response for **A** age group 1 (<32 weeks GA), mean responses with standard deviation (shaded areas) and **B** neonates, mean responses with significant areas (grey).

6.3.6. Profile of Mood States Questionnaire (POMS)

Since the means on the negative sub-scales for depression/anxiety and hostility were much lower than the means in the factor structure of the POMS (depression/anxiety m=9.51, hostility m=6.35) [110] we did not perform any further calculations or correlations with the POMS values, so far.

Table 6.9.: Results of the POMS questionnaire

Item	mean	std	min	max	max achievable
Niedergeschlagenheit (depression/anxiety)	3.38	5.0	0	27	84
Tatendrang (vigor)	16.2	8.6	0	32	42
Müdigkeit (fatigue)	10.2	7.1	0	27	42
Missmut (hostility)	1.9	3.3	0	16	42

6.4. Discussion

We could show significant differences in the heart rate, actogram and voice onset response in reaction to maternal compared to stranger female voice. The differences in the actogram were visible in neonates and in fetuses in the first age group (26-31 weeks); the differences in heart rate were significant in age-group 4 (38-42 weeks); the differences in voice onset response were significant in age-group 5 (newborn).
For heart rate variability we could not find any difference in the standard deviation between normal RR-intervals (SDNN) according to the stimulation type. In the power spectral density we also could not observe any difference according to the stimulation type. However, in our explorative analysis, we found a significant interaction between the age group and the sex of the fetus/neonate.

Our hypothesis about the fetal heart rate was not confirmed. The heart rate decelerations to maternal voice that we found in term fetuses go in line with the findings of some previous studies [59, 60, 61] but contradict the results of the more recent studies of Kisilevsky et al.[62, 63], which found accelerating heart rate responses. These differences could be associated with differences in stimulation intensity, since it was shown that fetuses react with a deceleration in heart rate to an auditory stimulation of intensity of 80 dB, but with an acceleration to an auditory stimulation of intensity of 90 dB [66]. Since our stimulation had only peak values of 90 dB the average stimulation intensity could be around 80 dB. Additionally, it was shown that the differences in fetal response to a familiar or unfamiliar voice is only present in fetuses with high vagal tone [61].
Since we have already found differences in the actogram and the voice onset response in the first age-group, it could be that the fetus concentrates on distinguishing the maternal voice from other female voices already before the 32nd week of gestation. Other studies on fetal movement in response to maternal voice did not find any significant differences [55, 62], but these investigations were made in term fetuses, where we also did not find differences. The reduced body movement to maternal voice stimulation in newborns seems to be in line with other studies in newborn [55].
Since we found only the normal expected increase in SDNN over gestation, as it was shown in former fetal magnetoencephalography (fMEG) studies [16], the stimulation with maternal voice seems not to influence the autonomic nervous system.
Our hypothesis about the changes in the theta and delta frequency band of the fetal brain activity was not confirmed. However, with fMEG we are only able to measure the power

spectral density of the whole brain and not of specific brain areas. Therefore, we assume that the areas where this change in theta and delta frequencies was shown in newborn studies [69, 70] are too small to generate a change in frequency bands over the whole brain. The differences in power spectral density of all four frequency bands between male and female fetuses are very interesting. The power of the spectral density of the brain in delta, theta alpha and beta frequency band is known to change over life. Also, sex differences in the brain power spectral density are found over the whole lifespan. Based on our findings we assume that these differences already start before birth.

Our hypothesis about the auditory event-related response to the maternal voice was also not confirmed. Since EEG studies in newborns showed a higher auditory event-related response to stimulation with maternal voice [71, 72] we were surprised to find a higher response to stimulation with a stranger female voice in our newborn group. A possible explanation for this difference might be the behavioral state of the newborns. In our fMEG measurements the newborns usually sleep during the stimulation while in the other studies the newborns were awake.

With 135 high quality datasets of fetal heart activity, with 26 - 29 datasets for each age group, we could perform a solid analysis of the fetal reaction to maternal and stranger female voice concerning the autonomic nervous system. Since we had only 59 high quality datasets for the fetal brain activity with only 6 datasets for the first age-group, further data sets have to be evaluated in order to make a statistically more valid statement.

The general hypothesis of an emerging difference in the response to mother and stranger voice that changes over time could not be confirmed. Also, the significant results in the heart rate, movement and voice onset response do not give a uniform picture. The mechanisms behind fetal perception of maternal voice and the ability to discriminate it from other female voices seem to be very complex. Further research therefore should include the fetal heart rate variability parameters to examine if there are differences in fetal heart rate reaction depending on the fetal vagal tone. Additionally, the fetal behavioral states should be investigated since voice perception may differ if the children are sleeping or awake. Further research should also focus on the investigation of sex differences in the fetal central and autonomic nervous system.

The fetal perception, and differentiation, of maternal voice seems to be a very complex cognitive process depending on many variables. To finally answer the question of when and how the fetus is starting to recognize the maternal voice, and of when and how the fetus is able to discriminate the maternal voice from other female voices, a more detailed study has to be made, including much more participants and taking a lot of additional parameters into account.

7. FAIRY TOOL

FAIRY is a matlab tool for the **f**ully **a**utomated **i**nterference **r**emoval and evaluation for fetal magnetoencephalograph**y**. This tool was developed to make the algorithms for fully automated R-peak detection (FLORA) and fully automated subtraction of heart activity (FAUNA) easy usable for all people working with fMEG data (including students, physicians, psychologists, nutritionists, midwifes and research assistants), without programming skills. FAIRY has a well structured graphical user interface (GUI) that easily leads the user through the processing and evaluation process, so that the usage can be learned within hours and no huge training period is necessary (see Appendix A).

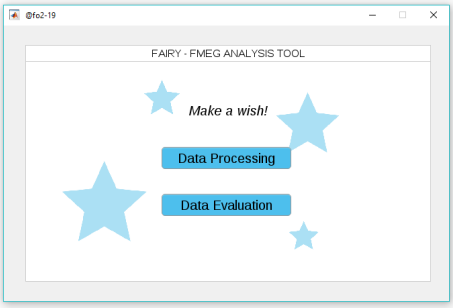

Figure 7.1.: Startscreen of the FAIRY tool.

That FLORA and FAUNA show advantages compared to the previously used methods was already shown in Chapters 4 and 5. The aim of this section is to show the influence of the methods on the evaluation of a real study, especially on aspects of time effort, number of evaluable datasets, fMCG and fMEG quality.

On the one hand all 186 AURORA datasets were manually evaluated by an expert using both semi automated template matching (SATM) and the hilbert transform algorithm (HTA) method for R-peak detection and orthogonal projection (OP) for heart activity removal. On the other hand they were automatically processed using the FAIRY tool which includes FLORA, as R-peak detection algorithm, and FAUNA to remove the heart activity.

7.1. Time Expenses of Processing

Here we are particularly concerned with the human working time involved in the processing of data, as this has been very high with the previous method of processing fMEG data. Since the manual working time is individually different, depending on the user's experience but also on the size and quality of the dataset, we measured the processing time of some datasets and calculated an average value.

The manual evaluation of a dataset with the conventional methods (see Chapter 2.5) usually takes 5 to 10 minutes. Datasets with a high noise level can take up to 20 minutes to process. With conventional methods, this also includes the entire computing time, which cannot be considered separately, since it is in constant interaction with the user and the user must

therefore always be attentive. With FAIRY the manual work and the computing time are separated by the possibility of the batch processing, which offers the possibility to carry out the calculations overnight. The initialization of the batch processing with the FAIRY processing tool takes about 5 to 10 minutes depending on the number of data records to be processed. After the computing time, which does not require the user's attention, the FAIRY evaluation tool checks the data sets in about 30 - 60 seconds per dataset.
If the processing effort is now calculated using the 188 data records of the AURORA study as an example, the processing time with conventional methods is over 31 hours, while processing with FAIRY should take a maximum of 3 hours and 20 minutes. FAIRY has thus reduced the manual workload of the study in this case by approx. 90%.

7.2. Number of Fetal Heart Rate Datasets

Of the initially 186 datasets the fetal heart rate detection could be done by all of the three methods in 147 datasets. 14 datasets could be processed only by SATM and HTA but not by FLORA, 10 datasets provided results for HTA and FLORA but not for SATM, for another 12 datasets we had only results for FLORA and for 3 datasets we had no result at all (see Tab. 7.1).
The 147 datasets, for which all 3 methods worked for the evaluation of the fetal magnetocardiogram (fMCG), showed the same results as the FLORA evaluation in Chapter 4.2. The number of detected R-peaks was significantly higher in FLORA compared to both SATM and HTA, the difference between RR measures significantly lower. The percentage of normal to normal intervals for FLORA was significantly higher compared to SATM, HTA and the combination of both (COMB) and the signal to noise ratio showed no significant differences at all.
14 datasets could be processed only by SATM and HTA but not by FLORA. Most of them looked very noisy after or even before mMCG removal and the sensor plots showed movement artifacts.
Of the 12 datasets for which we had only results for FLORA 4 didn't have a marker by mistake and 8 were marked als 'not evaluable' by the user. For 6 of these 8 marked datasets FLORA showed good fMCG evaluation.

Table 7.1.: The result of the evaluation of 186 datasets with three different R-peak detection methods. The table shows for how many datasets the evaluation has been successful (yes) or not (no).

n	SATM	HTA	FLORA
147	yes	yes	yes
14	yes	yes	no
10	no	yes	yes
12	no	no	yes
3	no	no	no

7.3. Heart Rate Continuity

The quality and continuity of the fMCG is especially important when it comes to HRV evaluation since ectopic or missing beats influence the result of the HRV a lot. According to the guidelines for HRV analysis [111], RR-intervals of ectopic or missing beats are removed prior to HRV calculation. The amount of R-peaks that have to be removed is an indicator for the quality of the dataset. We compared the percentage of ectopic beats in the fetal heart activity in 168 datasets processed by the FLORA algorithm to 171 datasets processed with the old methods. For the 171 datasets we used either the result of the SATM (41 datasets) or the HTA (130 datasets) according to the expert recommendation. Since the results were not normally distributed and the amount of datasets was not equal we tested for significance by using the Wilcoxon-Mann-Whitney ranksum test. Figure 7.2 shows a significant ($p < 0.001$) lower percentage of ectopic beats in datasets after processing the datasets with the FLORA algorithm in the FAIRY tool (0% to 6.61%, mean 0.05%) compared to the combination of SATM and HTA (0% to 21.98%, mean 0.45%).

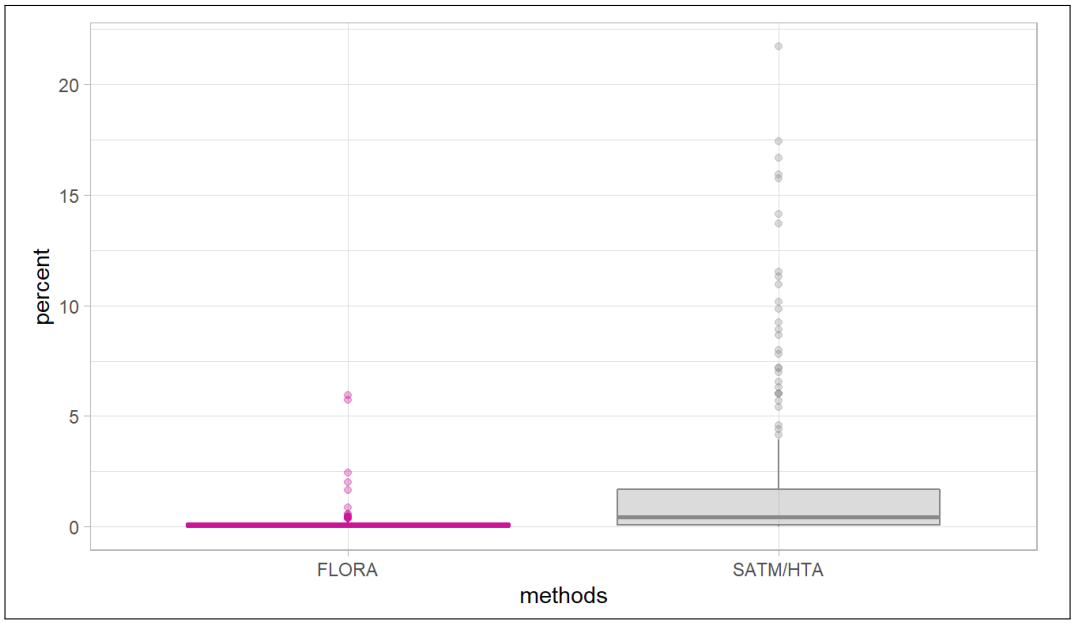

Figure 7.2.: Percentage of RR-intervals that are considered non-normal and therefore had to be removed prior to HRV analysis after processing with FLORA compared to after processing with SATM/HTA for 147 datasets.

7.4. Brain Activity Cluster

One aim of developing the FAUNA algorithm was to detect clusters of brain activity after mMCG and fMCG removal, which was not possible until now because of the signal redistribution. To investigate if the sensor with highest activity were cluster-like or randomly distributed we selected the 10 sensors with the highest RMS value after fMCG removal and calculated the mean of the inner distance between these 10 sensors. We compared all 147 datasets for which fMCG detection worked for FLORA and the expert recommended method SATM or HTA. The means of the inner distances were as well calculated for the 10 sensors with highest RMS value after heart activity removal by FAUNA as for the 10 sensors with highest root mean square value after the removal using OP.
To know the normal range for the result of the mean of the inner distance values, the distance of each sensor to his nearest neighbor and to the most distant sensor was calculated. The minimum distances ranged from 2.12 cm to 3.53 cm (mean 2.78 cm) the maximum distances from 28.31 cm to 54.34 cm (mean 40.01 cm). Additionally, the mean of the inner distance of each sensor and his 9 closest neighbor sensors was calculated. The resulting values are in a range between 4.26 cm to 6.51 cm with a mean of 5.21 cm. Since the results were not normally distributed but calculated for the same datasets, we tested for significance by using the Wilcoxon signrank test. Figure 7.3) shows a significant ($p < 0.001$) lower mean of the inner distance in the 10 sensors with highest root mean square value after processing the datasets with the FAUNA algorithm in the FAIRY tool (4.65 to 29.88, mean 9.55) compared to OP (5.58 to 31.80, mean 19.16).

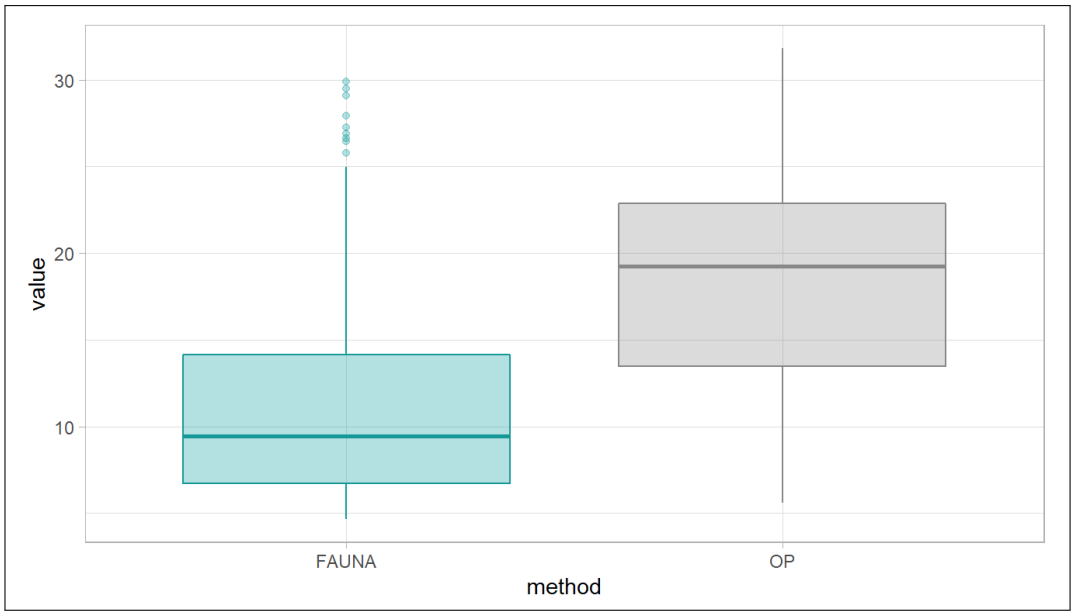

Figure 7.3.: Mean inner distance of 10 sensors with highest RMS after processing with FAUNA compared to after processing with OP for 147 datasets.

7.5. Visual Evaluation with FAIRY

After visual evaluation of the 142 preselected AURORA dataset using the FAIRY evaluation tool 135 datasets were selected data with a good and continuous fetal heartrate. Only 7 datasets were excluded because they had outliers that could influence variance and mean of following heart rate or heart rate variability analyses.
Visual evaluation of fMEG data resulted in 59 really good and usable automatically fMEG selection. As selection criteria we used a cluster shaped distribution of the 10 sensors with highest root mean square and a logarithmical shape of the power spectral density of these channels. 6 datasets had visible leftovers of the maternal head region, in 31 datasets the cluster with the highest root mean square sensors was obviously muscle activity from the maternal tights and in another 39 datasets the quality of the selected channels did not match the selection criteria. For datasets where the fetal heart activity was excluded, the fetal brain activity was automatically excluded too.

8. Discussion and Outlook

We could show that with FLORA, FAUNA and the FAIRY TOOL we created an environment for the processing and evaluation of fetal magnetoencephalographic data that compensates the weak points of the previously used methods, minimizes the required evaluation time by automating processes and, nevertheless, it is easy to use.

With the AURORA study we had a good showcase to compare the performance of the new algorithms to the performance of the previous processing methods on real data. We could show that the introduced methods are more sensitive and more specific than the old ones.

FLORA significantly improves the quality of the heart data, especially for younger age groups where there were often large gaps in the heart rate. These gaps would have falsified the data or would not have provided enough data points. Thanks to FLORA, the heart data is now so consistent that only a few data sets have to be excluded due to outliers, and the results can show a time-trace free of outliers. The fully automated method is comparable to the combination of automated Hilbert transformation algorithm (HTA) and semi-automated template matching (SATM) [95, 96], but requires much less time. The proportion of RR intervals that can be used is nearly 100% for almost all data sets. The number of non-normal RR-intervals that has to be removed for the calculation of heart rate variability is on average 1 out of 200 intervals according to FLORA whereas, using the old methods, approximately 1 out of 20 had to be removed. In addition, by reducing manual intervention by automating processes the reproducibility of the R-peak detection is ensured.

The removal of heart activity until now has been done by orthogonal projection (OP) [98]. This method worked reliable but had the problem of redistributing the signal [100]. We could show that FAUNA improves the removal of heart activity while neither redistributing the signal nor removing too much of the brain activity. This also paves the way for an evaluation of spontaneous brain activity which would have been very challenging with the previous methods, as good signal to noise ratio was only achieved by averaging the data on certain events. The distribution of the remaining magnetic activity now also shows a much clearer cluster structure. The mean inner distance within the 10 channels with highest root mean square activity calculated using OP is, on average, twice as high than the one calculated using FAUNA.

FAIRY provides a good framework that easily combines the two algorithms and is capable to process a large number of raw fMEG data sets at once. This prepares the processing of fMEG data for the era of "Big Data" and "Automated Science". Furthermore, it

facilitates the process of data evaluation. Even new employees and students can quickly and easily familiarize with the processing of the data without having to go through a long learning phase. In addition, the user interface avoids mistakes by the user, as tables are created automatically and the overview of all important data at a glance makes it much easier to estimate whether the quality of a data set is sufficient for further processing or not.

Out of the 142 datasets used for the AURORA study we had 135 ($\approx 95\%$) high quality fetal heart datasets and 59 datasets ($\approx 42\%$) high quality fetal brain datasets, what is really good for fMEG data. With these high quality datasets we could evaluate the fetal heart rate, actocardiogram, heart rate variability, the power spectral density of the fetal brain activity and auditory event-related responses.
Until now, actocardiograms have only been used in paper form to provide information on fetal behavioral states for certain studies. Now, the actogram can also be used to monitor changes in fetal movement behavior over time and under different conditions.
Auditory event-related responses have been detected a lot in previous studies but, for the first time, we measured voice onset responses after at least 500 ms breaks during continuous voice stimulation. Thanks to FAUNA, the quality of the data is so good that even in the first age-group (26 -31 weeks) a clear auditory response is visible, with very little variance between the different fetuses of this age group. We had a high drop out rate for the fetal brain activity but the remaining data were of high quality. So the event-related responses we found in all age-groups are of comparable quality than a response in newborns.
We have found that the analysis of the power spectral density in the fetal brain offers an interesting option but is probably not suitable for auditory stimulation. Since speech is processed in just small areas of the brain, frequency changes in these regions are probably too weak to be detected as we only have a total brain signal. However, analysis of power spectral density could be used to further investigate gender differences or sleep and behavioral states of the fetus, as those have an influence on the whole brain signal and not only on single areas.

One limitation of FLORA, FAUNA and the FAIRY TOOL is that they are specifically tailored to fMEG data. However since FLORA and FAUNA are written in Matlab, they also can be used in other scientific areas such as electrocardiography.
Another limitation is that the number of drop outs in fetal brain activity still could be reduced. Especially in early weeks of pregnancy, the signal of fetal brain activity is covered by signals of maternal thigh muscle activity or maternal head movements and therefore difficult to detect. For this reason, one of the next steps is to develop a method to detect muscle activity and to mute sensors or components with muscle activity so that the signal of fetal brain activity becomes recognizable again.
Since, according to some earlier studies [16], fetal behavior has a major influence on how and whether heart and brain parameters change under external influences, a second further step is the development of an automatic detection of fetal behavioral states based on the new high-quality datasets we now have.

FLORA improves the quality of the heart data by better detecting R-peaks and intelligently filling gaps, enabling automated heart rate evaluation.
FAUNA separates the heart data cleanly from the brain activity without redistributing the signal and paves the way for automated brain evaluation.
FAIRY offers a user-friendly, well-structured interface, making the evaluation of fMEG data fast and magically easy.

Bibliography

[1] Hubert Preissl, Curtis L Lowery, and Hari Eswaran. Fetal magnetoencephalography: current progress and trends. *Experimental neurology*, 190:28–36, 2004.

[2] Hubert Preissl, Curtis L Lowery, and Hari Eswaran. Fetal magnetoencephalography: viewing the developing brain in utero. *International review of neurobiology*, 68:1–23, 2005.

[3] Riitta Hari, Sylvain Baillet, Gareth Barnes, Richard Burgess, Nina Forss, Joachim Gross, Matti Hämäläinen, Ole Jensen, Ryusuke Kakigi, François Mauguière, et al. Ifcn-endorsed practical guidelines for clinical magnetoencephalography (meg). *Clinical Neurophysiology*, 2018.

[4] Rossitza Draganova, Hari Eswaran, Pamela Murphy, Curtis Lowery, and Hubert Preissl. Serial magnetoencephalographic study of fetal and newborn auditory discriminative evoked responses. *Early human development*, 83(3):199–207, 2007.

[5] Liviu Moraru, Reza Sameni, Uwe Schneider, Jens Haueisen, Ekkehard Schleußner, and Dirk Hoyer. Validation of fetal auditory evoked cortical responses to enhance the assessment of early brain development using fetal meg measurements. *Physiological measurement*, 32(11):1847, 2011.

[6] Liviu Moraru, Reza Sameni, Uwe Schneider, Jens Haueisen, Ekkehard Schleußner, and Dirk Hoyer. Validation of fetal auditory evoked cortical responses to enhance the assessment of early brain development using fetal meg measurements. *Physiological measurement*, 32(11):1847, 2011.

[7] Jana Muenssinger, Tamara Matuz, Franziska Schleger, Rossitza Draganova, Magdalene Weiss, Isabelle D Kiefer-Schmidt, Annette Wacker-Gussmann, Rathinaswamy Bhavanandhan Govindan, Curtis L Lowery, Hari Eswaran, et al. Sensitivity to auditory spectral width in the fetus and infant–an fmeg study. *Frontiers in human neuroscience*, 7:917, 2013.

[8] Franziska Schleger, Karin Landerl, Jana Muenssinger, Rossitza Draganova, Maren Reinl, Isabelle Kiefer-Schmidt, Magdalene Weiss, Annette Wacker-Gußmann, Minna Huotilainen, and Hubert Preissl. Magnetoencephalographic Signatures of Numerosity Discrimination in Fetuses and Neonates. *Developmental Neuropsychology*, 39(4):316–329, may 2014.

[9] Katarzyna Linder, Franziska Schleger, Isabelle Kiefer-Schmidt, Louise Fritsche, Stefanie Kümmel, Martin Heni, Magdalene Weiss, Hans-Ulrich Häring, Hubert

Preissl, and Andreas Fritsche. Gestational Diabetes Impairs Human Fetal Postprandial Brain Activity. *The Journal of Clinical Endocrinology & Metabolism*, 100(11):4029–4036, nov 2015.

[10] Rossitza Draganova, Anna Schollbach, Franziska Schleger, Johanna Braendle, Sara Y Brucker, Harald Abele, Karl O Kagan, Diethelm R Wallwiener, Andreas Fritsche, Hari Eswaran, et al. Fetal auditory evoked responses to onset of amplitude modulated sounds. a fetal magnetoencephalography (fmeg) study. *Hearing research*, 363:70–77, 2018.

[11] Naim Haddad, Rathinaswamy B Govindan, Srinivasan Vairavan, Eric Siegel, Jessica Temple, Hubert Preissl, Curtis L Lowery, and Hari Eswaran. Correlation between fetal brain activity patterns and behavioral states: an exploratory fetal magnetoencephalography study. *Experimental neurology*, 228(2):200–205, 2011.

[12] Douglas F Rose and Hari Eswaran. Spontaneous neuronal activity in fetuses and newborns. *Experimental neurology*, 190:37–43, 2004.

[13] Ellen Fehlert, Kathrin Willmann, Louise Fritsche, Katarzyna Linder, Haliza Mat-Husin, Franziska Schleger, Magdalene Weiss, Isabelle Kiefer-Schmidt, Sara Y Brucker, Hans-Ulrich Häring, et al. Gestational diabetes alters the fetal heart rate variability during an oral glucose tolerance test: a fetal magnetocardiography study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 124(12):1891–1898, 2017.

[14] Diana Escalona-Vargas, Jessica L Coker, Shona Ray-Griffith, Eric R Siegel, Curtis L Lowery, Zachary N Stowe, and Hari Eswaran. Fetal assessment in buprenorphine-maintained women using fetal magnetoencephalography: a pilot study. *Addiction*, 113(10):1895–1904, 2018.

[15] J G Nijhuis, H F Prechtl, C B Martin, and R S Bots. Are there behavioural states in the human fetus? *Early human development*, 6(2):177–95, apr 1982.

[16] Johanna Brändle, Hubert Preissl, Rossitza Draganova, Erick Ortiz, Karl O Kagan, Harald Abele, Sara Y Brucker, and Isabelle Kiefer-Schmidt. Heart rate variability parameters and fetal movement complement fetal behavioral states detection via magnetography to monitor neurovegetative development. *Frontiers in human neuroscience*, 9:147, 2015.

[17] David Barker. Developmental origins of adult health and disease. *Journal of epidemiology and community health*, 58(2):114, 2004.

[18] Petter D Gluckman and Mark A Hanson. The developmental origins of health and disease. In *Early life origins of health and disease*, pages 1–7. Springer, 2006.

[19] David Cohen. Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786, 1968.

[20] Matti Hämäläinen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413, 1993.

[21] Curtis L Lowery, Hari Eswaran, Pamela Murphy, and Hubert Preissl. Fetal magnetoencephalography. In *Seminars in Fetal and Neonatal Medicine*, volume 11, pages 430–436. Elsevier, 2006.

[22] Diogo Ayres-de Campos, Catherine Y Spong, Edwin Chandraharan, and FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. Figo consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *International Journal of Gynecology & Obstetrics*, 131(1):13–24, 2015.

[23] James P Neilson. Fetal electrocardiogram (ecg) for fetal monitoring during labour. *Cochrane database of systematic reviews*, (12), 2015.

[24] René D Kok, Marieke M de Vries, Arend Heerschap, and Paul P van den Berg. Absence of harmful effects of magnetic resonance exposure at 1.5 t in utero during the third trimester of pregnancy: a follow-up study. *Magnetic resonance imaging*, 22(6):851–854, 2004.

[25] Marc Gertsch. *Das EKG: auf einen Blick und im Detail*. Springer-Verlag, 2008.

[26] MARIANNE Bootsma, CEES A Swenne, HARM H Van Bolhuis, PETER C Chang, V Manger Cats, and AV Bruschke. Heart rate and heart rate variability as indexes of sympathovagal balance. *American Journal of Physiology-Heart and Circulatory Physiology*, 266(4):H1565–H1571, 1994.

[27] Ivana Antelmi, Rogério Silva De Paula, Alexandre R Shinzato, Clóvis Araújo Peres, Alfredo José Mansur, and Cesar José Grupi. Influence of age, gender, body mass index, and functional capacity on heart rate variability in a cohort of subjects without heart disease. *The American journal of cardiology*, 93(3):381–385, 2004.

[28] Borejda Xhyheri, Olivia Manfrini, Massimiliano Mazzolini, Carmine Pizzi, and Raffaele Bugiardini. Heart rate variability today. *Progress in cardiovascular diseases*, 55(3):321–331, 2012.

[29] Caroline Di Bernardi Luft, Emílio Takase, and David Darby. Heart rate variability and cognitive function: Effects of physical effort. *Biological psychology*, 82(2):186–191, 2009.

[30] Isabelle Kiefer-Schmidt, Julia Raufer, Johanna Brändle, Jana Münßinger, Harald Abele, Diethelm Wallwiener, Hari Eswaran, and Hubert Preissl. Is there a relationship between fetal brain function and the fetal behavioral state? a fetal meg-study. *Journal of perinatal medicine*, 41(5):605–612, 2013.

[31] Tabea Y Fehling, Franziska Schleger, Haliza Mat-Huzin, Hubert Preissl, Andreas Fritsche, Jan Pauluschke-Fröhlich, Harald Abele, Johanna Brändle, Sara Y Brucker, and Isabelle Kiefer-Schmidt. Vegetative entwicklungsdiagnostik bei iugr-feten mittels fetaler verhaltensstadien und herzratenvariabilitätsparameter–eine fmkg-studie. *Geburtshilfe und Frauenheilkunde*, 78(10):P–338, 2018.

[32] Haliza Mat Husin, Franziska Schleger, Ilena Bauer, Ellen Fehlert, Isabelle Kiefer-Schmidt, Magdalene Weiss, Karl O. Kagan, Sara Brucker, Jan Pauluschke-Fröhlich, Hari Eswaran, Hans-Ulrich Häring, Andreas Fritsche, and Hubert Preissl. Maternal weight, weight gain, and metabolism are associated with changes in fetal heart rate and variability. *Obesity*, 28(1):114–121, 2020.

[33] Standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Circulation*, 93(5):1043–1065, 1996.

[34] Jean K Moore and Fred H Linthicum Jr. The human auditory system: a timeline of development. *International journal of audiology*, 46(9):460–478, 2007.

[35] Peter G Hepper and B Sara Shahidullah. The development of fetal hearing. *Fetal and Maternal Medicine Review*, 6(3):167–179, 1994.

[36] Hugo Lagercrantz and Jean-Pierre Changeux. The emergence of human consciousness: from fetal to neonatal life. *Pediatric research*, 65(3):255–260, 2009.

[37] Ivica Kostović and Miloš Judaš. The development of the subplate and thalamocortical connections in the human foetal brain. *Acta paediatrica*, 99(8):1119–1127, 2010.

[38] Charles C Lee. Thalamic and cortical pathways supporting auditory processing. *Brain and language*, 126(1):22–28, 2013.

[39] Uwe Schneider, Ekkehard Schleussner, Jens Haueisen, Hannes Nowak, and Hans-Joachim Seewald. Signal analysis of auditory evoked cortical fields in fetal magnetoencephalography. *Brain topography*, 14(1):69–80, 2001.

[40] Hari Eswaran, Hubert Preissl, James D Wilson, Pam Murphy, Stephen E Robinson, Douglas Rose, Jiri Vrba, and Curtis L Lowery. Short-term serial magnetoencephalography recordings of fetal auditory evoked responses. *Neuroscience letters*, 331(2):128–132, 2002.

[41] Manuela Holst, Hari Eswaran, Curtis Lowery, Pamela Murphy, Jonathan Norton, and Hubert Preissl. Development of auditory evoked fields in human fetuses and newborns: a longitudinal meg study. *Clinical Neurophysiology*, 116(8):1949–1955, 2005.

[42] Isabelle Kiefer, Eric Siegel, Hubert Preissl, Maureen Ware, Burkhard Schauf, Curtis Lowery, and Hari Eswaran. Delayed maturation of auditory-evoked responses in

growth-restricted fetuses revealed by magnetoencephalographic recordings. *American journal of obstetrics and gynecology*, 199(5):503–e1, 2008.

[43] Ville Mäkinen, Hannu Tiitinen, and Patrick May. Auditory event-related responses are generated independently of ongoing brain activity. *Neuroimage*, 24(4):961–968, 2005.

[44] Liliane Aparecida Fagundes Silva, Fernanda Cristina Leite Magliaro, ACM Carvalho, and Carla Gentile Matas. Maturation of long latency auditory evoked potentials in hearing children: systematic review. In *CoDAS*, volume 29, page e20160107, 2017.

[45] Kenneth J Gerhardt and Robert M Abrams. Fetal exposures to sound and vibroacoustic stimulation. *Journal of Perinatology*, 20(S1):S21, 2000.

[46] Denis Querleu, X Renard, C Boutteville, and G Crepin. Hearing by the human fetus? In *Seminars in perinatology*, volume 13, pages 409–420, 1989.

[47] Minna Huotilainen. A new dimension on foetal language learning. *Acta Paediatrica*, 102(2):102–103, 2013.

[48] Christine Moon, Hugo Lagercrantz, and Patricia K Kuhl. Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, 102(2):156–160, 2013.

[49] Christine Moon. Prenatal experience with the maternal voice. In *Early Vocal Contact and Preterm Infant Brain Development*, pages 25–37. Springer, 2017.

[50] Denis Querleu, Xavier Renard, Fabienne Versyp, Laurence Paris-Delrue, and Gilles Crèpin. Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 28(3):191–212, 1988.

[51] Christine M Moon and William P Fifer. Evidence of transnatal auditory learning. *Journal of Perinatology*, 20(S1):S37, 2000.

[52] Jean-Pierre Lecanuet and Carolyn Granier-Deferre. Speech stimuli in the fetal environment. In *Developmental neurocognition: Speech and face processing in the first year of life*, pages 237–248. Springer, 1993.

[53] Gabriella A Ferrari, Ylenia Nicolini, Elisa Demuru, Cecilia Tosato, Merhi Hussain, Elena Scesa, Luisa Romei, Maria Boerci, Emanuela Iappini, Guido Dalla Rosa Prati, et al. Ultrasonographic investigation of human fetus responses to maternal communicative and non-communicative stimuli. *Frontiers in psychology*, 7:354, 2016.

[54] Viola Marx and Emese Nagy. Fetal behavioural responses to maternal voice and touch. *PloS one*, 10(6):e0129118, 2015.

[55] Peter G Hepper, D Scott, and Sara Shahidullah. Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, 11(3):147–153, 1993.

[56] Renaud Jardri, Delphine Pins, Véronique Houfflin-Debarge, Caroline Chaffiotte, Nathalie Rocourt, Jean-Pierre Pruvo, Marc Steinling, Pierre Delion, and Pierre Thomas. Fetal cortical activation to sound at 33 weeks of gestation: a functional mri study. *Neuroimage*, 42(1):10–18, 2008.

[57] Renaud Jardri, Véronique Houfflin-Debarge, Pierre Delion, Jean-Pierre Pruvo, Pierre Thomas, and Delphine Pins. Assessing fetal response to maternal speech using a noninvasive functional brain imaging technique. *International Journal of Developmental Neuroscience*, 30(2):159–161, 2012.

[58] Kristin M Voegtline, Kathleen A Costigan, Heather A Pater, and Janet A DiPietro. Near-term fetal response to maternal spoken voice. *Infant Behavior and Development*, 36(4):526–533, 2013.

[59] William P Fifer and Christine M Moon. The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatrica*, 83:86–93, 1994.

[60] Elizabeth M Ockleford, Margaret A Vince, Claire Layton, and Margaret R Reader. Responses of neonates to parents' and others' voices. *Early human development*, 18(1):27–36, 1988.

[61] Laura S Smith, Pawel A Dmochowski, Darwin W Muir, and Barbara S Kisilevsky. Estimated cardiac vagal tone predicts fetal responses to mother's and stranger's voices. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 49(5):543–547, 2007.

[62] Barbara S Kisilevsky, Sylvia MJ Hains, Kang Lee, Xing Xie, Hefeng Huang, Hai Hui Ye, Ke Zhang, and Zengping Wang. Effects of experience on fetal voice recognition. *Psychological science*, 14(3):220–224, 2003.

[63] Barbara S Kisilevsky, Sylvia MJ Hains, Christine Ann Brown, Charlotte T Lee, Bernadine Cowperthwaite, Sherri Schmidt Stutzman, Melissa L Swansburg, Kang Lee, Xing Xie, Hefeng Huang, et al. Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development*, 32(1):59–71, 2009.

[64] Barbara S Kisilevsky and Sylvia MJ Hains. Onset and maturation of fetal heart rate response to the mother's voice over late gestation. *Developmental science*, 14(2):214–223, 2011.

[65] Marie-Claire Busnel, C Granier-Deferre, and JP Lecanuet. Fetal audition. *Annals of the New York Academy of Sciences*, 662(1):118–134, 1992.

[66] Lynn J Groome, Donna M Mooney, Scherri B Holland, Lynn S Bentz, Jana L Atterbury, and Roscoe A Dykman. The heart rate deceleratory response in low-risk

human fetuses: Effect of stimulus intensity on response topography. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 30(2):103–113, 1997.

[67] Anthony J DeCasper and William P Fifer. Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448):1174–1176, 1980.

[68] D Querleu, C Lefebvre, M Titran, X Renard, M Morillion, and G Crepin. Reaction of the newborn infant less than 2 hours after birth to the maternal voice. *Journal de gynécologie, obstétrique et biologie de la reproduction*, 13(2):125–134, 1984.

[69] Mariko O Uchida, Takeshi Arimitsu, Kiyomi Yatabe, Kazushige Ikeda, Takao Takahashi, and Yasuyo Minagawa. Effect of mother's voice on neonatal respiratory activity and eeg delta amplitude. *Developmental psychobiology*, 60(2):140–149, 2018.

[70] Z Radicevic, M Vujovic, Lj Jelicic, and M Sovilj. Comparative findings of voice and speech: language processing at an early ontogenetic age in quantitative eeg mapping. *Experimental brain research*, 184(4):529–532, 2008.

[71] Raye-Ann deRegnier, Charles A Nelson, Kathleen M Thomas, Sandi Wewerka, and Michael K Georgieff. Neurophysiologic evaluation of auditory recognition memory in healthy newborn infants and infants of diabetic mothers. *The Journal of pediatrics*, 137(6):777–784, 2000.

[72] Maude Beauchemin, Berta Gonzalez-Frankenberger, Julie Tremblay, Phetsamone Vannasing, Eduardo Martínez-Montes, Pascal Belin, Renee Beland, Diane Francoeur, Ana-Maria Carceller, Fabrice Wallois, et al. Mother and stranger: an electrophysiological study of voice processing in newborns. *Cerebral cortex*, 21(8):1705–1711, 2010.

[73] Maurice George Kendall. Rank correlation methods. 1948.

[74] Andy Field. *Discovering statistics using IBM SPSS statistics*. sage, 2013.

[75] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[76] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[77] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[78] Ian Jolliffe. *Principal component analysis*. Springer, 2011.

[79] Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

[80] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[81] Aapo Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, 10(3):626–634, 1999.

[82] Matteo Frigo and Steven G. Johnson. FFTW: An adaptive software architecture for the FFT. In *Proc. 1998 IEEE Intl. Conf. Acoustics Speech and Signal Processing*, volume 3, pages 1381–1384. IEEE, 1998.

[83] Matteo Frigo and Steven G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[84] James W Cooley, Peter AW Lewis, and Peter D Welch. The fast fourier transform and its applications. *IEEE Transactions on Education*, 12(1):27–34, 1969.

[85] Peter Welch. The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on audio and electroacoustics*, 15(2):70–73, 1967.

[86] Lawrence Marple. Computing the discrete-time" analytic" signal via fft. *IEEE Transactions on signal processing*, 47(9):2600–2603, 1999.

[87] Phillip I Good. Permutation, parametric and bootstrap tests of hypotheses: a practical guide to resampling methods for testing hypotheses. *Permutation, parametric and bootstrap tests of hypotheses: a practical guide to resampling methods for testing hypotheses*, 100(4), 2005.

[88] Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer, 1992.

[89] Jean Dickinson Gibbons and Subhabrata Chakraborti. *Nonparametric statistical inference*. Springer, 2011.

[90] Myles Hollander, Douglas A Wolfe, and Eric Chicken. *Nonparametric statistical methods*, volume 751. John Wiley & Sons, 2013.

[91] Joseph L Hodges Jr and Erich L Lehmann. Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, pages 598–611, 1963.

[92] John C Mosher, Richard M Leahy, and Paul S Lewis. Matrix kernels for meg and eeg source localization and imaging. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 5, pages 2943–2946. IEEE, 1995.

[93] John C Mosher, Richard M Leahy, and Paul S Lewis. Eeg and meg: forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46(3):245–259, 1999.

[94] Jukka Sarvas. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine & Biology*, 32(1):11, 1987.

[95] Umit D Ulusar, RB Govindan, James D Wilson, Curtis L Lowery, Hubert Preissl, and Hari Eswaran. Adaptive rule based fetal qrs complex detection using hilbert transform. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4666–4669. IEEE, 2009.

[96] James D. Wilson, R. B. Govindan, Jeff O. Hatton, Curtis L. Lowery, and Hubert Preissl. Integrated approach for fetal QRS detection. *IEEE Transactions on Biomedical Engineering*, 55(9):2190–2197, 2008.

[97] Jack McCubbin, Stephen E. Robinson, R. Cropp, A. Moiseev, Jiri Vrba, Pamela Murphy, Hubert Preissl, and Hari Eswaran. Optimal reduction of MCG in fetal MEG recordings. *IEEE Transactions on Biomedical Engineering*, 2006.

[98] Jiri Vrba, Stephen E. Robinson, Jack McCubbin, Pamela Murphy, Hari Eswaran, James D. Wilson, Hubert Preißl, and Curtis L. Lowery. Human fetal brain imaging by magnetoencephalography: Verification of fetal brain signals by comparison with fetal brain models. *NeuroImage*, 2004.

[99] Richard Courant and David Hilbert. *Methods of Mathematical Physics: Partial Differential Equations*. John Wiley & Sons, 2008.

[100] Jiri Vrba, Stephen E Robinson, Jack McCubbin, Curtis L Lowery, Hari Eswaran, James D Wilson, Pamela Murphy, and Hubert Preißl. Fetal meg redistribution by projection operators. *IEEE Transactions on Biomedical Engineering*, 51(7):1207–1218, 2004.

[101] Katrin Sippel, Julia Moser, Franziska Schleger, Hubert Preissl, Wolfgang Rosenstiel, and Martin Spüler. Fully automated r-peak detection algorithm (flora) for fetal magnetoencephalographic data. *Computer Methods and Programs in Biomedicine*, 173:35–41, 2019.

[102] Katrin Sippel, Julia Moser, Franziska Schleger, Hubert Preissl, Wolfgang Rosenstiel, and Martin Spüler. Fully automated subtraction of heart activity for fetal magnetoencephalography data. In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5685–5689, 2019.

[103] Diana Escalona-Vargas, Eric R Siegel, Pamela Murphy, Curtis L Lowery, and Hari Eswaran. Selection of reference channels based on mutual information for frequency-dependent subtraction method applied to fetal biomagnetic signals. *IEEE Transactions on Biomedical Engineering*, 64(5):1115–1122, 2016.

[104] Martin Spüler. Spatial filtering of eeg as a regression problem. In *7th Graz Brain-Computer Interface Conference*, 09 2017.

[105] A Jeewajee, C Lever, S Burton, J O'keefe, and N Burgess. Environmental novelty is signaled by reduction of the hippocampal theta frequency. *Hippocampus*, 18(4):340–348, 2008.

[106] Wolfgang Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*, 16(12):606–617, 2012.

[107] Carl V Smith, Brian Satt, Jeffrey P Phelan, and Richard H Paul. Intrauterine sound levels: intrapartum assessment with an intrauterine microphone. *American journal of perinatology*, 7(04):312–315, 1990.

[108] Dirk Hoyer, Esther Heinicke, Susann Jaekel, Florian Tetschke, Dania Di Pietro Paolo, Jens Haueisen, Ekkehard Schleußner, and Uwe Schneider. Indices of fetal development derived from heart rate patterns. *Early human development*, 85(6):379–386, 2009.

[109] Peter Van Leeuwen, Dirk Cysarz, Friedrich Edelhäuser, and Dietrich Grönemeyer. Heart rate variability in the individual fetus. *Autonomic Neuroscience*, 178(1-2):24–28, 2013.

[110] Cornelia Albani, Gerd Blaser, Michael Geyer, Gabriele Schmutzer, E Brähler, Harald Bailer, and Norbert Grulke. The german short version of" profile of mood states"(poms): psychometric evaluation in a representative sample. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 55(7):324–330, 2005.

[111] Marek Malik, J Thomas Bigger, A John Camm, Robert E Kleiger, Alberto Malliani, Arthur J Moss, and Peter J Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European heart journal*, 17(3):354–381, 1996.

[112] RB Govindan, S Vairavan, UD Ulusar, JD Wilson, SS McKelvey, H Preissl, and H Eswaran. A novel approach to track fetal movement using multi-sensor magnetocardiographic recordings. *Annals of biomedical engineering*, 39(3):964–972, 2011.

Appendices

A. FAIRY Handbook

FAIRY is a matlab tool for the **f**ully **a**utomated process**i**ng fo**r** fetal magnetoencephalogra**phy**. This tool was developed to make the algorithms for fully automated R-peak detection (FLORA) and fully automated subtraction of heart activity (FAUNA) easy usable for all people working with fMEG data (including students, physicians, psychologists, nutritionists, midwifes and research assistants), without programming skills. FAIRY has a well structured graphical user interface (GUI) that easily leads the user through the processing and evaluation process, so that the usage can be learned within hours and no huge training period is necessary. FAIRY is divided into two main segments. First data processing, second evaluation (see Fig. A.1). The data processing is the batch processing of the heart activity detection and removal, for fetal and maternal heart. The evaluation includes a table for the results of the processing and a detailed overview for each individual dataset, to control the quality of the maternal and fetal heart rate and the remaining brain activity and to decide for each dataset which parts will be used for further evaluation.

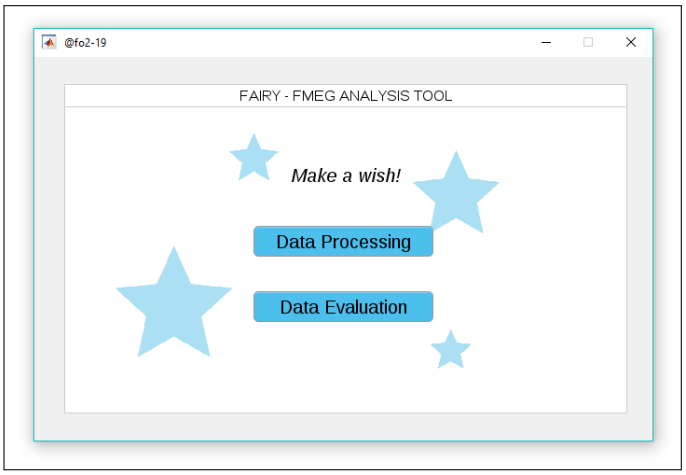

Figure A.1.: Startscreen of the FAIRY tool.

A.1. FAIRY Data Processing

The fairy subtool for data processing offers the possibility to perform heart recognition (FLORA) and heart removal (FAUNA) for large amounts of data at once. For this purpose, the names of the raw data sets are specified in the first mask (see Fig. A.2) and a storage location for the data output is selected. The names of the data records can be exported as a list. This ensures that no data records are forgotten or added when data is processed again or further.

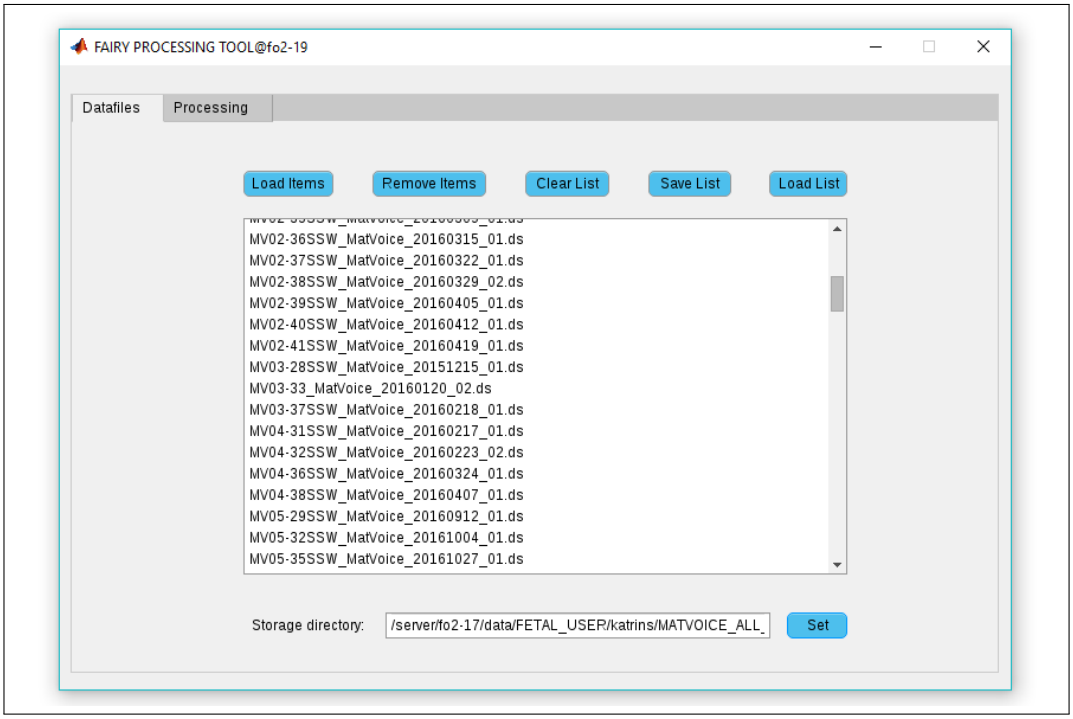

Figure A.2.: FAIRY Processing Tab. Here datasets and storage directory can be selected.

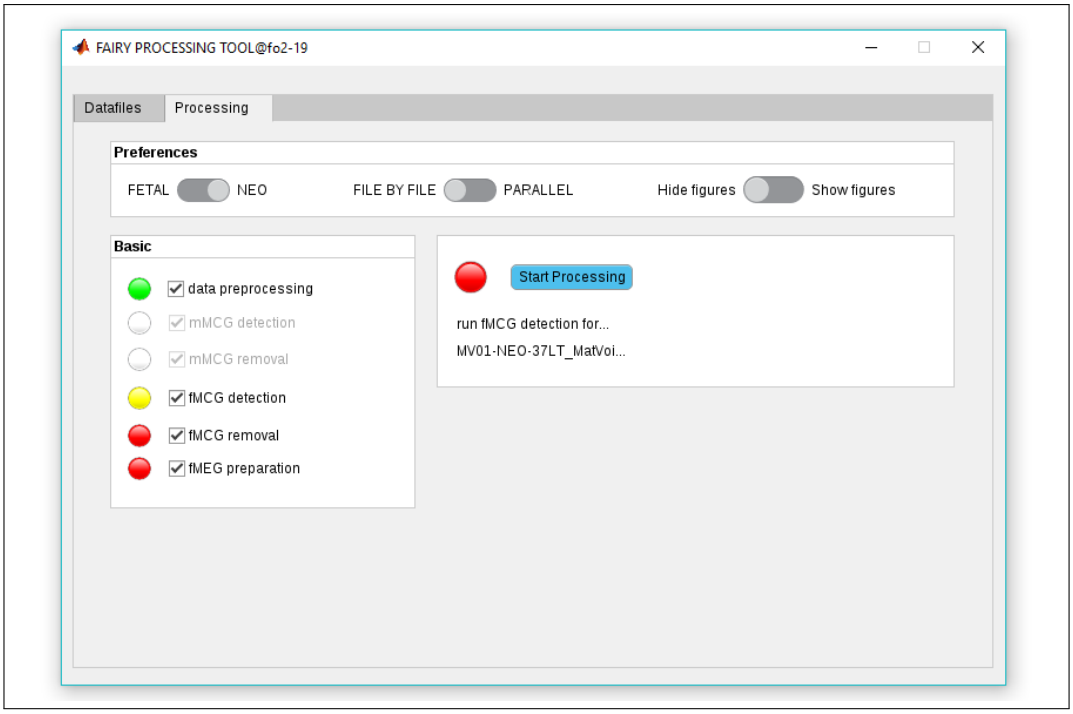

Figure A.3.: Fairy Processing Tab during the evaluation of neonatal datasets.

In the second mask (see Fig. A.3), the individual steps of the processing are listed, which can all be executed together or individually.
1) data preprocessing removes data from squid sensors that have too high a variance or correlate too low with their neighboring sensors (see Chapter 4.1)
2) mMCG detection uses FLORA for the recognition of maternal cardiac activity (see Chapter 4).
3) mMCG removal removes the maternal cardiac activity with the FAUNA method (see Chapter 5).
4) fMCG detection uses FLORA to detect maternal cardiac activity (see Chapter 4). Additionally, the actogram for the fetus is calculated [112].
5) fMCG removal removes fetal heart activity using the FAUNA method (see Chapter 5).
6) fMEG preparation searches for clusters of fetal brain activity and selects the 10 channels with the highest RMS.
It should be noted that the data should contain either only fetal or only neonatal datasets, as the type of dataset must be specified in the second mask. this is important because steps 2 and 3 are omitted for neonatal datasets. The data can be processed one after the other as well as in parallel. Nevertheless, a step is executed first for all data before moving on to the next step. In addition, the mask offers the possibility to either show or hide the individual images created during data processing. Hiding the images has a positive effect on the performance of data processing.

A.2. FAIRY Evaluation

After the data has been processed with the FAIRY processing tool, the quality of the datasets and their processing can be checked in the FAIRY evaluation sub-tool. In the first mask of the evaluation tool, first the names and the location of the desired data are given. This information corresponds exactly with the first mask in the processing tool. A list of datafiles which was saved in the processing step can be reloaded here very easily. In the second mask there is an empty table which can be filled by using the update button. This table contains the most important parameters of the datasets: Name; if it is a neonatal dataset; age (either in pregnancy weeks or in days of life); average heart rate of the mother (only for fetal datasets), average heart rate of the child; length of the measurement in minutes; sampling frequency of the measurement (see Fig. A.4). By selecting any field, the dataset in this row can either be deactivated/activated with one click or or viewed in detail. Clicking on the the 'edit file'-button a separate window opens for the evaluation of this dataset.

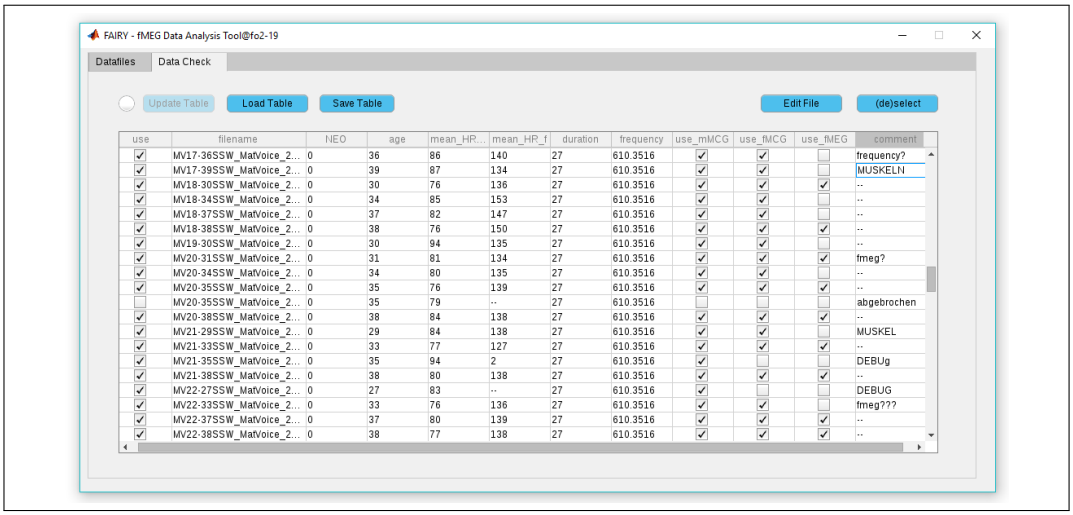

Figure A.4.: FAIRY Evaluation Table. This table gives an overview of the results of the preprocessing and the decisions in the evaluation.

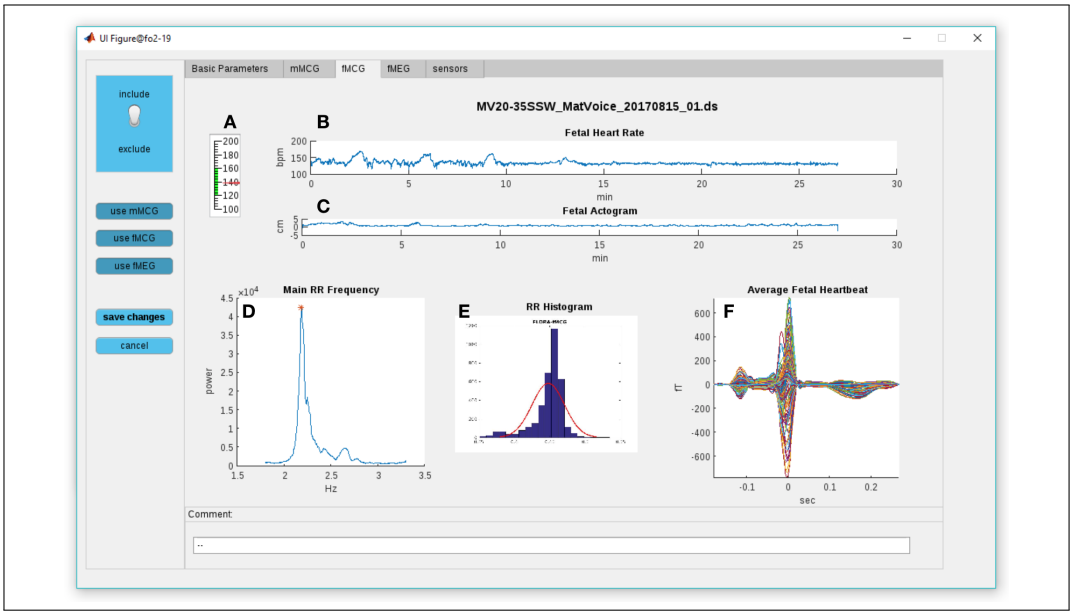

Figure A.5.: FAIRY Evaluation Window - fetal heart activity Tab. Displaying **A** the mean fetal heart rate, **B** the fetal heart rate over time, **C** the fetal actogram, **D** the power spectral density of the data before removal of the fetal heart activity with selected RR-main, **E** the RR histogram, and **F** the averaged fetal heartbeat.

A.3. Evaluation Window

This evaluation window provides an overview over some figures generated during data processing with the processing tool (see Fig. A.5). On the left side of the window there are selection buttons, in the lower part a comment line and in the right part several tabs. In the first tab you can see an overview of the squid channels removed during preprocessing as well as two fields in which the age or the sampling rate can be entered manually, if they could not be recorded automatically during evaluation. In the second tab there is an overview of various parameters of maternal heart activity: mean heart rate, heart rate over the course of the measurement, frequency of raw data in the range of maternal heart rate, histogram of the calculated RR intervals and an averaged heart beat over all R-peaks.

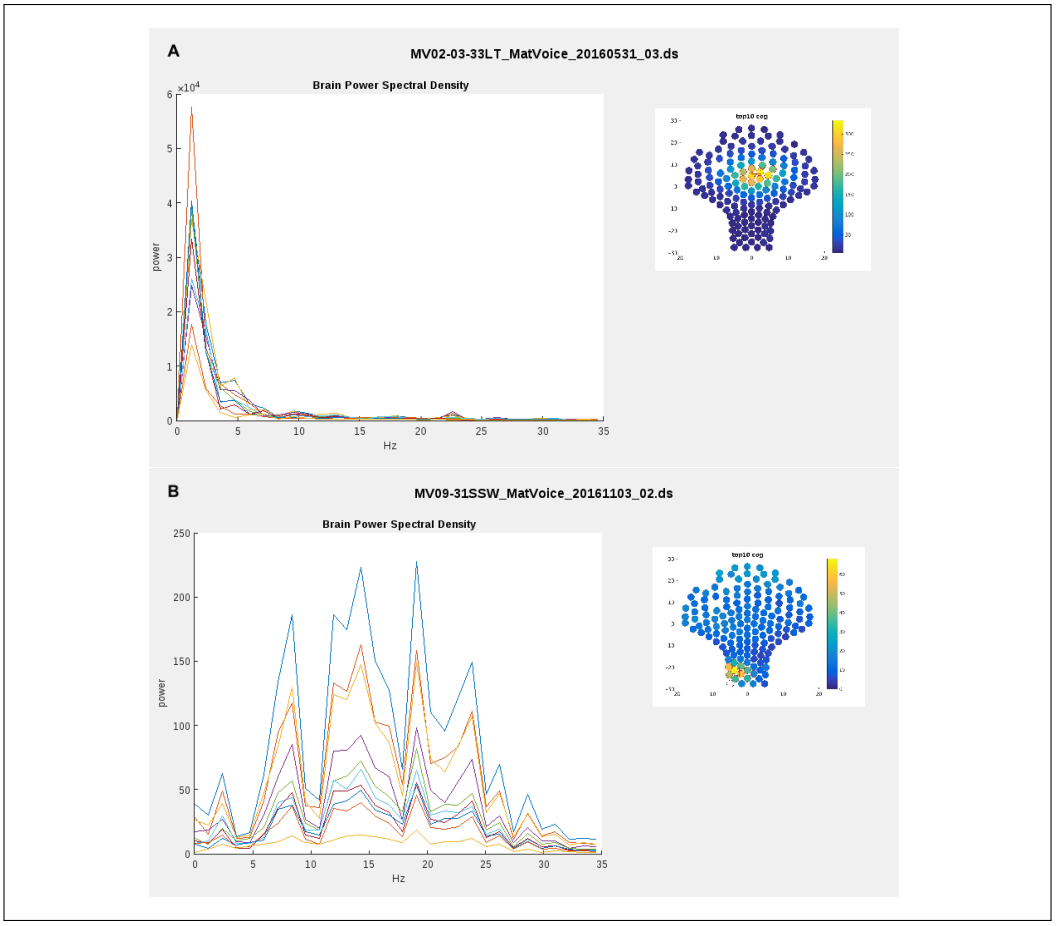

Figure A.6 Two fMEG PSD examples with **A** good PSD as it should look like for brain activity and **B** typical PSD for tight muscle activity of the mother.

The same parameters are available for the fetal heart activity in the next tab, and additionally the fetal actogram (see Fig. A.5) The fourth tab contains the power spectral density analysis of the ten selected fMEG channels, as well as a RMS plot of all SQUID sensors with the ten selected fMEG channels marked (see Fig. A.6).

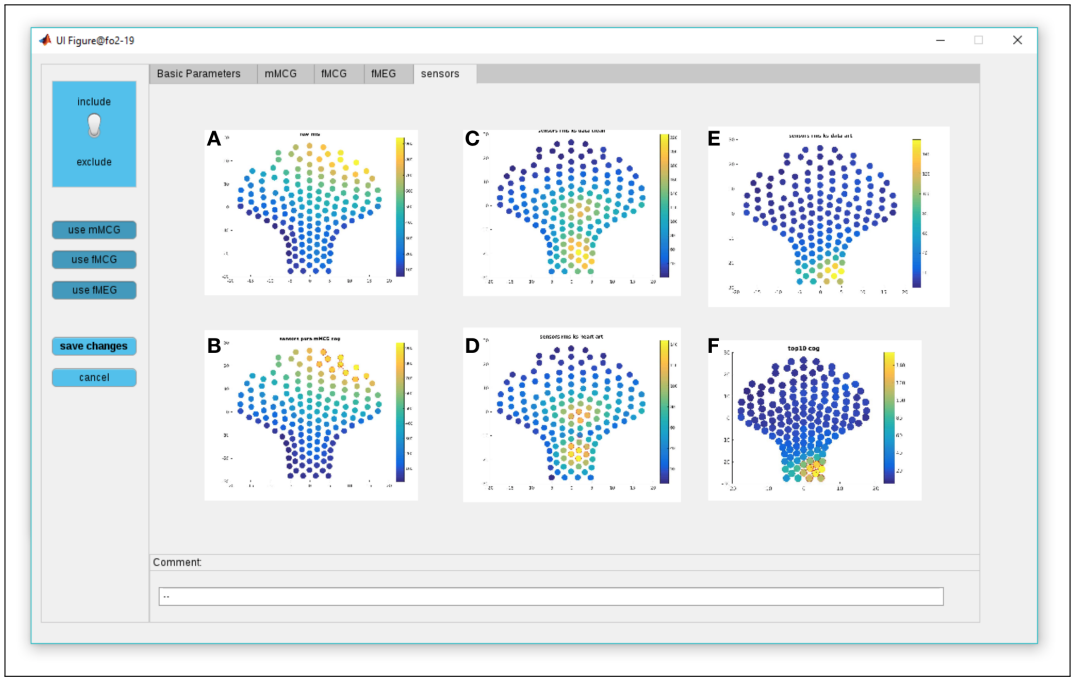

Figure A.7 FAIRY Evaluation Window - Sensor Tab. Displaying sensor RMS for raw data **A**, pure maternal heart **B**, data after mMCG removal **C**, pure fetal heart **D**, data after fMCG removal **E**, and selected fMEG **F**.

In the last tab there are another 6 RMS plots of all SQUID sensors for each processing step (see Fig. A.7). While viewing the individual tabs, the buttons on the left can be used to select whether the quality of mMCG fMCG and fMEG is good enough to be included in further evaluation steps. This is then transferred to the table after closing the evaluation window. Special features of the dataset can be noted in the comment field. These comments are also included in the table.

After all datafiles in the table have been checked, the table can be exported from the GUI to an excel file.

B. Dornröschen

Vorzeiten lebten ein König und eine Königin, die sprachen jeden Tag: "Ach wenn wir doch ein Kind hätten!" und kriegten immer keins. Da geschah es, als die Königin einmal im Bade saß, dass ein Frosch aus dem Wasser ans Land kroch und zu ihr sprach: "Dein Wunsch wird erfüllt werden. Ehe ein Jahr vergeht, wirst du eine Tochter haben." Was der Frosch gesagt hatte, das geschah, und die Königin bekam ein Mädchen, das war so schön, dass der König vor Freude sich nicht zu fassen wusste und ein großes Fest gab.

Er lud nicht bloß seine Verwandten, Freunde und Bekannten, sondern auch die weisen Frauen dazu ein, damit sie dem Kind hold und gewogen wären. Es waren ihrer dreizehn in seinem Reiche. Weil er aber nur zwölf goldene Teller hatte, von denen sie essen sollten, musste eine daheim bleiben. Das Fest wurde mit aller Pracht gefeiert und als es zu Ende war, beschenkten die weisen Frauen das Kind mit ihren Wundergaben, die eine mit Tugend, die andere mit Schönheit, die dritte mit Reichtum, und so mit allem, was auf der Welt zu wünschen ist.

Als elf ihre Sprüche eben getan hatten, trat plötzlich die dreizehnte herein. Sie wollte sich dafür rächen, dass sie nicht eingeladen war, und ohne jemand zu grüßen oder nur anzusehen, sprach sie: "Die Königstochter soll sich in ihrem fünfzehnten Jahr an einer Spindel stechen und tot hinfallen." Und ohne ein Wort weiterzusprechen, kehrte sie sich um und verließ den Saal.

Alle waren erschrocken. Da trat die zwölfte hervor, die ihren Wunsch noch übrig hatte, und weil sie den bösen Spruch nicht aufheben, sondern ihn nur mildern konnte, so sagte sie: "Es soll aber kein Tod sein, sondern ein hundertjähriger tiefer Schlaf, in den die Königstochter fällt."

Der König, der sein liebes Kind vor dem Unglück gern bewahren wollte, ließ den Befehl ausgeben, dass alle Spindeln im ganzen Reich verbrannt werden sollten. An dem Mädchen aber wurden die Gaben der weisen Frauen erfüllt. Denn es war so schön und sittsam, freundlich und verständig, das jedermann, der es ansah, es lieb haben musste.

An dem Tag, an dem es grade fünfzehn Jahre alt wurde, waren der König und die Königin nicht zu Hause, und das Mädchen blieb ganz allein im Schloss zurück. Da ging es allerorten herum, besah Stuben und Kammern, wie es Lust hatte, und kam endlich auch an einen alten Turm. Es stieg die enge Wendeltreppe hinauf und gelangte zu einer kleinen Tür. In dem Schloss steckte ein verrosteter Schlüssel. Als es ihn umdrehte, sprang die Tür auf, und da saß in einem kleinen Stübchen eine alte Frau mit einer Spindel und spann emsig ihren Flachs.

"Guten Tag, du altes Mütterchen", sprach die Königstochter, "was machst du da?" "Ich spinne", sagt die Alte und nickte mit dem Kopf. "Was ist das für ein Ding, das so lustig herumspringt?" fragte das Mädchen, nahm die Spindel und wollte auch spinnen. Kaum

hatte sie aber die Spindel angerührt, so ging der Zauberspruch in Erfüllung, und sie stach sich damit in den Finger.

In dem Augenblick aber, wo sie den Stich empfand, fiel sie auf das Bett nieder, das dort stand, und lag in einem tiefen Schlaf. Und dieser Schlaf verbreitete sich über das ganze Schloss. Der König und die Königin, die eben heimgekommen und in den Saal getreten waren, fingen an einzuschlafen und der ganze Hofstaat mit ihnen. Da schliefen auch die Pferde im Stall, die Hunde im Hof, die Tauben auf dem Dach, die Fliegen an der Wand, ja, das Feuer, das auf dem Herd flackerte, wurde still und schlief ein. Der Braten hörte auf zu brutzeln, der Koch, der dem Küchenjungen, weil er etwas versehen hatte, eine Ohrfeige geben wollte, ließ es sein und schlief. Und der Wind legte sich, und auf den Bäumen vor dem Schloss regte sich kein Blättchen mehr.

Rings um das Schloss aber begann eine Dornenhecke zu wachsen, die jedes Jahr höher wurde und endlich das ganze Schloss umzog und darüber hinauswuchs, dass gar nichts mehr davon zu sehen war, nicht einmal die Fahne auf dem Dach.

Es ging aber die Sage in dem Land von dem schönen schlafenden Dornröschen, denn so wurde die Königstochter genannt, sodass von Zeit zu Zeit Königssöhne kamen und durch die Hecke in das Schloss dringen wollten. Es war ihnen aber nicht möglich, denn die Dornen hielten fest zusammen, als hätten sie Hände, und die Jünglinge blieben darin hängen, konnten sich nicht wieder losmachen und starben eines jämmerlichen Todes.

Nach langen, langen Jahren kam wieder einmal ein Königssohn in das Land und hörte, wie ein alter Mann von der Dornenhecke erzählte. Es solle ein Schloss dahinter stehen, indem eine wunderschöne Königstochter, Dornröschen genannt, schon seit hundert Jahren schlafe. Und mit ihr schliefen der König und die Königin und der ganze Hofstaat.

Der Alte wusste auch von seinem Großvater, dass schon viele Königssöhne gekommen wären und versucht hätten, durch die Dornenhecke zu dringen, aber sie wären darin hängengeblieben und eines traurigen Todes gestorben. Da sprach der Jüngling: "Ich fürchte mich nicht, ich will hinaus und das schöne Dornröschen sehen." Der gute Alte mochte ihm abraten, wie er wollte, er hörte nicht auf seine Worte.

Nun waren aber gerade die hundert Jahre verflossen, und der Tag war gekommen, wo Dornröschen wieder erwachen sollte. Als der Königssohn sich der Dornenhecke näherte, waren es lauter schöne, große Blumen. Die taten sich von selbst auseinander und ließen ihn unbeschädigt hindurch, und hinter ihm taten sie sich wieder als eine Hecke zusammen.

Im Schlosshof sah er die Pferde und die scheckigen Jagdhunde liegen und schlafen. Auf dem Dach saßen die Tauben und hatten das Köpfchen unter den Flügel gesteckt. Und als er ins Haus kam, schliefen die Fliegen an der Wand, der Koch in der Küche hielt noch die Hand so, als wollte er dem Jungen eine Ohrfeige geben, und die Magd saß vor dem schwarzen Huhn, das gerupft werden sollte.

Da ging er weiter und sah im Saal den ganzen Hofstaat liegen und schlafen, und oben auf dem Thron schlummerten der König und die Königin. Da ging er noch weiter. Alles war so still, dass einer seinen Atem hören konnte. Und endlich kam er zu dem Turm und öffnete die Tür zu der kleinen Stube, in der Dornröschen schlief.

Da lag es und war so schön, dass er die Augen nicht abwenden konnte, und er bückte sich und gab ihm einen Kuss. Wie er es mit dem Mund berührt hatte, schlug Dornröschen die Augen auf, erwachte und blickte ihn ganz freundlich an.
Da gingen sie zusammen hinab, und der König erwachte und die Königin und der ganze Hofstaat, und sie sahen einander mit großen Augen an. Und die Pferde im Hof standen auf und schüttelten sich, die Jagdhunde sprangen und wedelten, die Tauben auf dem Dach zogen das Köpfchen unterm Flügel hervor, sahen umher und flogen ins Feld. Die Fliegen an den Wänden krochen weiter, das Feuer in der Küche erhob sich, flackerte und kochte das Essen, der Braten fing wieder and zu brutzeln, und der Koch gab dem Jungen eine Ohrfeige, dass er schrie, und die Magd rupfte das Huhn fertig.

Und da wurde die Hochzeit des Königssohnes mit dem Dornröschen in aller Pracht gefeiert, und sie lebten vergnügt bis an ihr Ende.

C. Profile Of Mood States Questionnaire (POMS)

Datum:_____ Code-Nr.:_____

POMS

Anleitung: (bitte genau durchlesen)

Sie finden nachstehend eine Liste mit Wörtern, die verschiedenartige Gefühle oder Gefühlszustände beschreiben. Bitte lesen Sie sorgfältig jedes einzelne Wort und setzten Sie dann in der Spalte ein Kreuz ein, die am besten Ihre Gefühlzustände beschreibt.
Bitte lassen Sie keine Zeile aus!

		überhaupt nicht	sehr schwach	schwach	etwas	ziemlich	stark	sehr stark
1	zornig	0	1	2	3	4	5	6
2	abgeschlafft	0	1	2	3	4	5	6
3	unglücklich	0	1	2	3	4	5	6
4	lebhaft	0	1	2	3	4	5	6
5	unsicher	0	1	2	3	4	5	6
6	lustlos	0	1	2	3	4	5	6
7	traurig	0	1	2	3	4	5	6
8	aktiv	0	1	2	3	4	5	6
9	gereizt	0	1	2	3	4	5	6
10	verdrießlich	0	1	2	3	4	5	6
11	betrübt	0	1	2	3	4	5	6
12	energisch	0	1	2	3	4	5	6
13	ängstlich	0	1	2	3	4	5	6
14	hoffnungslos	0	1	2	3	4	5	6
15	überreizt	0	1	2	3	4	5	6
16	müde	0	1	2	3	4	5	6
17	verärgert	0	1	2	3	4	5	6
18	entmutigt	0	1	2	3	4	5	6
19	neidig	0	1	2	3	4	5	6
20	fröhlich	0	1	2	3	4	5	6
21	verbittert	0	1	2	3	4	5	6
22	erschöpft	0	1	2	3	4	5	6
23	schwermütig	0	1	2	3	4	5	6
24	verzweifelt	0	1	2	3	4	5	6
25	träge	0	1	2	3	4	5	6
26	hilflos	0	1	2	3	4	5	6
27	ermattet	0	1	2	3	4	5	6
28	munter	0	1	2	3	4	5	6
29	wütend	0	1	2	3	4	5	6
30	schwungvoll	0	1	2	3	4	5	6
31	schlecht gelaunt	0	1	2	3	4	5	6
32	minderwertig	0	1	2	3	4	5	6
33	erschreckt	0	1	2	3	4	5	6
34	tatkräftig	0	1	2	3	4	5	6
35	entkräftet	0	1	2	3	4	5	6

Bitte prüfen Sie, ob Sie alle Feststellungen zutreffend beantwortet haben!