THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

The EEG of the Neonatal Brain – Classification of Background Activity

JOHAN LÖFHEDE

CHALMERS

HÖGSKOLAN I BORÅS
VETENSKAP FÖR PROFESSION

Division of Biomedical Engineering
Department of Signals and Systems
Chalmers University of Technology

Göteborg, Sweden 2009

The EEG of the Neonatal Brain – Classification of Background Activity
JOHAN LÖFHEDE
ISBN 978-91-7385-339-2

Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny Serie nr 3020
ISSN 0346-718X

Division of Biomedical Engineering
Department of Signals and Systems
Chalmers University of Technology
SE – 412 96 Göteborg, Sweden
Telephone + 46 (0)31-772 1000

Skrifter från Högskolan i Borås: 19
ISSN 0280-381X

School of Engineering
University of Borås
SE – 501 90 Borås, Sweden

e-mail: johan@lofhede.se

Cover: A newborn baby at the neonatal intensive care unit, the Queen Silvia children's hospital, Göteborg, Sweden. The time-frequency plot in the background shows EEG power spectrum variations during sleep, and the EEG plot in the foreground shows a burst suppression pattern.
Photo: Yxell, Montage: Erik Lodin.

This research project was supported by a grant from the Margarethahemmet Foundation, the Swedish state under the ALF-agreement and the BIOPATTERN EU Network of Excellence, EU contract 508803.

Printed by Chalmers Reproservice
Göteborg, Sweden 2009

To Hanh and Smulan

Abstract

The brain requires a continuous supply of oxygen and nutrients, and even a short period of reduced oxygen supply can cause severe and lifelong consequences for the affected individual. The unborn baby is fairly robust, but there are of course limits also for these individuals. The most sensitive and most important organ is the brain. When the brain is deprived of oxygen, a process can start that ultimately may lead to the death of brain cells and irreparable brain damage. This process has two phases; one more or less immediate and one delayed. There is a window of time of up to 24 hours where action can be taken to prevent the delayed secondary damage. One recently clinically available technique is to reduce the metabolism and thereby stop the secondary damage in the brain by cooling the baby.

It is important to be able to quickly diagnose hypoxic injuries and to follow the development of the processes in the brain. For this, the electroencephalogram (EEG) is an important tool. The EEG is a voltage signal that originates within the brain and that easily and non-invasively can be recorded at bedside. The signals are, however, highly complex and require special competence to interpret, a competence that typically is not available at the intensive care unit. This thesis addresses the problem of automatic classification of neonatal EEG and proposes methods that would be possible to use in bed-side monitoring equipment for neonatal intensive care units.

The thesis is a compilation of six papers. The first four deal with the segmentation of pathological signals (burst suppression) from post-asphyctic full term newborn babies. These studies investigate the use of various classification techniques, using both supervised and unsupervised learning. In paper V the scope is widened to include both classification of pathological activity versus activity found in healthy babies as well as application of the segmentation methods on the parts of the EEG signal that are found to be of the pathological type. The use of genetic algorithms for feature selection is also investigated. In paper VI the segmentation methods are applied on signals from pre-term babies to investigate the impact of a certain medication on the brain.

The results of this thesis demonstrate ways to improve the monitoring of the brain during intensive care of newborn babies. Hopefully it will someday be implemented in monitoring equipment and help to prevent permanent brain damage in post asphyctic babies.

Keywords: EEG, segmentation, classification, asphyxia, hypoxia, newborn, neonatal, cerebral

List of publications

This thesis is based on the work contained in the following papers:

Paper I **Detection of burst and suppression in the EEG of post asphyctic newborns.**
J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer and K. Lindecrantz. 28th Annual International IEEE EMBS Conference, 2006, 2179-2182.

Paper II **Classifying burst and suppression in the EEG of post asphyctic newborns using a support vector machine.**
J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer and K. Lindecrantz. 3rd International IEEE EMBS Conference on Neural Engineering, 2007, 630-633.

Paper III **Classification of burst and suppression in the neonatal electroencephalogram.**
J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer and K. Lindecrantz. J. Neural Eng. 5, 2008, 402-410.

Paper IV **Comparing a supervised and an unsupervised classification method for burst detection in neonatal EEG.**
J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer and K. Lindecrantz. 30th Annual International IEEE EMBS Conference, 2008, 3836-3839.

Paper V **Automatic classification of background EEG activity in healthy and sick neonates.**
J. Löfhede, M. Thordstein, N. Löfgren, A. Flisberg, M. Rosa-Zurera, I. Kjellmer and K. Lindecrantz. Submitted to the Journal of Neural Engineering.

Paper VI **Does indomethacin for closure of patent ductus arteriosus affect cerebral function?**
A. Flisberg, I. Kjellmer, J. Löfhede, N. Löfgren, M. Rosa-Zurera, K. Lindecrantz and M. Thordstein.

The following papers were generated during the thesis work but are not included in the thesis:

Comparison of three methods for classifying burst and suppression in the EEG of post asphyctic newborns.
J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer and K. Lindecrantz. 29th Annual International IEEE EMBS Conference, 2007, 5136-5139.

Application of a Very Simple Segmentation Algorithm to Burst Suppression Classification in Electroencephalogram Signals.
P. Damaschke and J. Löfhede. Submitted to Pattern Recognition.

Contents

Acknowledgements

The writing of this thesis and the work that it describes would not have been possible without our research group, a collection of people from the University of Borås and the Sahlgrenska University Hospital. First of all, I would like to thank my advisor Prof. Kaj Lindecrantz for his support and never ending enthusiasm. Also a big thank you to my co-supervisors: Dr Nils Löfgren, for his deep knowledge on EEG signal processing and practical advice on how to bend Matlab to ones will, and associate Prof. Magnus Thordstein for being a great golden standard and for coming up with new exciting research ideas. Thanks to Prof. em. Ingemar Kjellmer for the knowledge and inspiration, and Anders Flisberg for good collaboration and for all the interesting signals.

I would like to thank all the people that have been working with me or around me during these five years. The main part of this work has been carried out at the department of Signals and Systems at Chalmers and I would like to thank all the people there for making it such a great place to work, especially the people in the division of biomedical engineering and the crowd at floor 5. Thanks also to former colleague Johan Degerman who came up with the idea behind paper IV, and to Fernando Seoane for all the peptalks and discussions during these years.

For great assistance regarding the non-research related issues, I would like to thank Ann-Christine Lindbom, Agneta Kinnader and Madeleine Persson. Special thanks go my good friend Lars Börjesson for faithfully keeping track of the coffee breaks and for the assistance all the times that I have managed to mess up my computers in various ways.

I would also like to especially acknowledge my good friend Erik Lodin for the montage of the cover illustration.

Finally I would like to thank my fiancé Hanh for her never ending love and support, and for the new adventure that we are about to embark on. Smulan, thanks for being so patient and waiting for daddy to finish his thesis before entering this world.

Notations and abbreviations

σ^2	variance
μ	mean
aEEG	amplitude-integrated EEG
ANN	artificial neural network
AS	active sleep
AUC	area under the curve
AW	active wake
BS	burst-suppression
BSR	burst-suppression ratio
CAR	common average reference
CFM	cerebral function monitoring
CNS	central nervous system
ECG	electrocardiogram
EEG	electroencephalogram
FLD	Fisher's linear discriminant
HMM	hidden Markov model
ICU	intensive care unit
ML	maximum likelihood
NICU	neonatal intensive care unit
PCA	principal component analysis
pdf	probability density function
QS	quiet sleep
QW	quiet wake
ROC	receiver operating characteristic
SEF	spectral edge frequency
SVM	support vector machine
TA	*tracé alternant*
TD	*tracé discontinue*

Part I

Introduction

Chapter 1

Introduction

The brain requires a continuous supply of oxygen and nutrients, and even a short period without them can cause lifelong effects. During delivery there is always a risk of insufficient circulation or blood gas exchange to the baby, something that may lead to asphyxia. This condition includes lack of oxygen, excess of carbon dioxide and a lowered pH value which can lead to permanent brain damage. Babies at risk are kept under close observation during delivery and afterwards at a neonatal intensive care unit (NICU), but it is hard to determine if the babies are recovering and if any brain damage has occurred. Parameters such as the heart rate, blood pressure and oxygen saturation are monitored regularly, but they are only measures of general conditions. If the function of the brain itself is to be monitored, the most direct way is to measure the electrical signals produced by the brain, the electroencephalogram (EEG).

The EEG is a voltage signal that is usually measured using metal electrodes placed on the scalp. It originates from electrical activity of the neuronal cells in the brain, and the EEG signal contains information about

the health status of a patient's brain that can be interpreted by a clinical neurophysiologist. Earlier studies from our group have investigated how certain parameters calculated from the EEG signal can be used for detecting hypoxia (lack of oxygen) in the brain, and even for predicting the outcome after an hypoxic event [1].

In practice, most EEG recordings are evaluated through visual inspection of the unprocessed signal by a clinical neurophysiologist. Obviously, this methodology only allows for intermittent evaluations, and is not suitable for continuous bedside monitoring. Moreover, the expertise needed for this type of evaluation is typically not available at the NICU, and the patient cannot be transferred to a neurophysiologist for diagnosis. Even though methods for remote consultations have been developed [2], this methodology mainly allows for evaluation at distinct time instances and is not suitable for continuous bedside monitoring.

One attempt to simplify long-time monitoring of brain function that can be used for bedside monitoring is the amplitude-integrated EEG (aEEG). This method, in its most commonly used format, displays a filtered version of a two-channel EEG on a compressed time scale. It provides the clinician with a simple way to monitor the brain activity of a patient, and the compressed time scale gives a convenient view of several hours of recorded brain activity. However, this method has some severe limitations. For instance, interference and artifacts have in some cases been demonstrated to be hidden in the compressed signal and mistaken for brain activity, and there are also examples of missed seizure activity [3]. Because of these limitations, neurophysiologists argue that the unprocessed EEG signal has to be taken into consideration when interpreting the aEEG. This means that the staff at the NICU need to be able to interpret the signal at least to the level that they recognize artifacts and can distinguish between these and important brain activity. The staff has a lot of things to keep in mind, and adding complexity to their work would probably be problematic. Figure 1 shows an example of how the equipment surrounding a patient at a NICU may look.

To enable an improved continuous cerebral monitoring, we aim at developing methods for automatic classification and quantification of different types of activity in neonatal EEG. The input of the system should be a number of EEG channels and possibly additional parameters such as blood pressure and electrocardiogram (ECG) which normally are measured during these circumstances. The methods should be of a kind that can be implemented in a compact form suitable for use in a NICU environment, with easily interpreted parameters and alarms for threatening conditions of the brain. These parameters could serve as decision support for the clinician selecting the proper treatment or adjusting medication.

Figure 1: A newborn baby is treated after asphyxia during birth. The bed is cooled with circulating water to reduce further damage to the brain, and various parameters, *e.g.* aEEG (left screen), are monitored (Photo: Yxell).

A functioning system of this kind will enable higher-quality care for high-risk neonates by providing clinicians with the possibility to continuously monitor the function of the brain itself, and not just the underlying support functions. Continuous monitoring will make it possible to follow the development of the status of the child over time, and enables the clinician to modify treatments and to follow the results in a real-time fashion, instead of having to rely on intermittent evaluations made by neurophysiologists.

A system like this could schematically consist of the following parts:

- EEG amplification
- Digitization and storage
- Filtering of the EEG to reduce disturbances from surrounding devices
- Artifact rejection that marks epochs of questionable quality and excludes them from further processing

- Activity classification algorithms that divide the data into different categories, pathological (indicative of disease) or normal. Both pathological and normal EEGs in neonates can be continuous or intermittent (alternating between two types of activity)
- Segment classification that quantifies the proportions of different types of activity in intermittent EEG
- Presentation of easily interpreted information to the attending staff

Figure 2 shows the system as a flowchart, all the way from the raw EEG signal to some easily interpreted parameter which can be displayed to the attending staff. The focus of this thesis is on feature generation, feature selection and classification of some different types of activity that can be found in neonatal EEG. The shaded boxes represent processes that are included in the thesis, but these processes are only active during development and are not to be included in the bedside equipment. Good EEG amplifiers and systems for storage and display of the acquired signal are available for clinical use [4] and are outside the scope of this thesis. The filtering that has been done was limited to simple highpass, lowpass and notchfilters implemented in Matlab, with especially the lowpass filters manually adapted to minimize the 50 Hz components that in some EEG recordings were dominant even after notch filtering.

The types of EEG under consideration consist of four behavioral states that are common in healthy full term neonates, one type that is typical for pre-term neonates (*tracé discontinue*) and burst suppression (BS) which is a type of activity sometimes found in very sick neonates. BS is intermittent activity characterized by a very low signal level (suppression) that is occasionally interrupted by sudden outbursts of higher signal levels (bursts). In the case that BS is detected, the automated analysis takes one step further and segments the pattern into burst and suppression, thus making it possible to calculate parameters such as the burst suppression ratio (BSR) and the suppression lengths which can be useful when arriving at a prognosis for a sick baby [5]. The *tracé discontinue* data was considered in a separate study where we applied the developed methods to the problem of investigating the possible presence of side-effects of a certain drug used to remedy a condition that sometimes can be found in preterm babies.

The methods that have been used include linear classifiers such as Fisher's linear discriminant as well as nonlinear ones such as support vector machines (using a nonlinear RBF kernel) and neural networks. As inputs to the classifiers a number of features of the underlying EEG have been used. These features are parameters that are calculated from sliding windows that are moved along the signal, and enhance different

characteristics that are useful for classification. The classifiers can handle many features in combination, but just using all available features, or adding features at random, would not give the best classification since features that do not add useful information to the classifier will instead add noise. Therefore two methods for feature selection have been used: exhaustive search and restricted search using genetic algorithms. The exhaustive search simply tries all possible combinations of features on the classification problem, and the best combination can be selected. When the number of features grows this method quickly becomes too slow because the number of possible combinations grows very quickly. Restricted search using genetic algorithms on the other hand searches the space of possible combinations using methods inspired by natural selection and will usually find a solution close to the optimal one while trying much fewer combinations than the exhaustive search. However, the number of attempted combinations is still in the range of thousands, and because each attempt involves training and testing the classifier the genetic algorithm is too slow when used with the more advanced classifiers, and Fishers linear discriminator is preferred over them.

The results show that the developed methods can segment all the different intermittent EEG types that have been tested, and that burst suppression activity can be distinguished from normal EEG. If the developed methods were to be implemented as parts of a monitoring system they would provide improved insight into the brain function of babies at the NICU by being able to automatically detect if the baby is having BS and being able to automatically measure *e.g.* the suppression lengths. It can also classify the signal as normal, and indicate if the baby is sleeping quietly by detecting the presence of *tracé alternant*.

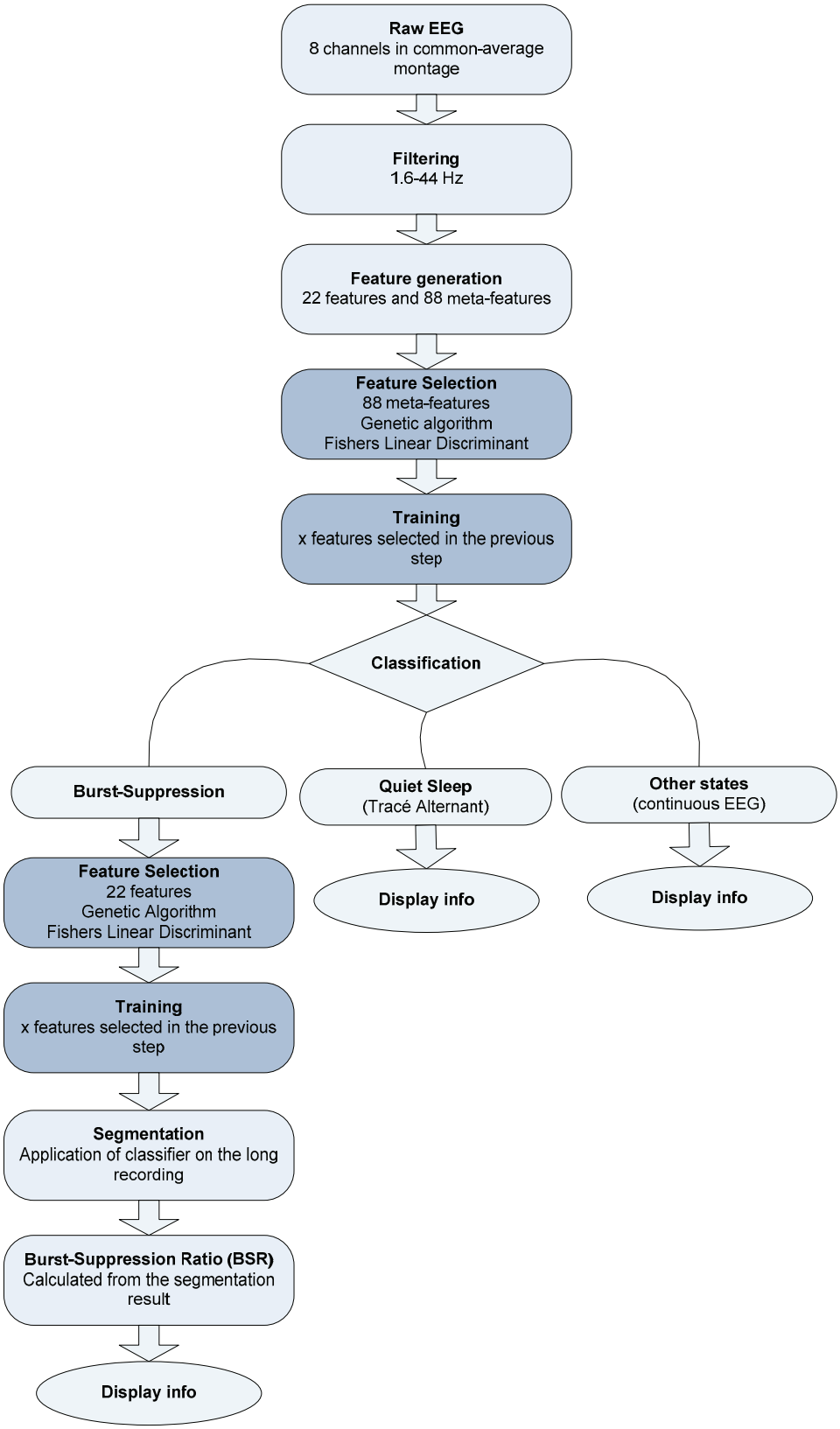


Figure 2: Monitoring system flowchart. The shaded boxes represent steps that are only included during development of the system, and that not need to be implemented in the bed-side equipment.

1.1 Outline of the thesis

Part I contains an introduction to the area of neonatal monitoring and some background to the problems that have been worked on. Part I is divided into the following chapters:

Chapter 2 describes some relevant aspects of brain physiology. Neurons and synapses are described with emphasis on how they give rise to the potentials that are summed into the EEG signal. Then the major parts of the brain such as the cerebrum, cerebellum and brain stem are described to familiarize the reader with the basic structure of the brain.

Chapter 3 describes how the EEG signal is generated and how it can be measured. The electrode placement system and the different montages are described followed by an overview of some types of activity that can be found in the neonatal EEG.

Chapter 4 describes the data that has been available for experiments and used for training of the classifiers.

Chapter 5 describes the various signal processing methods that have been used in this project, with focus on the concept of generating feature signals intended to be used for automatic classification of the underlying signal.

Chapter 6 reviews the techniques for classification of signals that have been used throughout the project.

Chapter 7 describes why feature selection is necessary. Two methods for feature selection, exhaustive search and restricted search using genetic algorithms, are described.

Chapter 8 gives some conclusions and ideas for future work.

Part II contains the papers that are the foundation of the thesis. Summaries of these papers can be found in the beginning of Part II.

Chapter 2

The brain

The chapter starts with a section dealing with the neuron, or nerve cell. This is the smallest unit in the construction of the wiring and logical system of the brain and of the entire nervous system. Then the anatomy of the brain is briefly described at a higher level, focusing on the cerebral cortex which is the main source of the EEG.

Most of the information in this chapter is based on [6].

2.1 Neurons

The neuron is the smallest functional unit in the brain and makes up a large part of the volume of the central nervous system (CNS). The rest of the volume consists of cells that support the neurons in various ways. A neuron has a very simple computational capability in itself, but by working together in vast neural networks they together form the complex system that is the human brain.

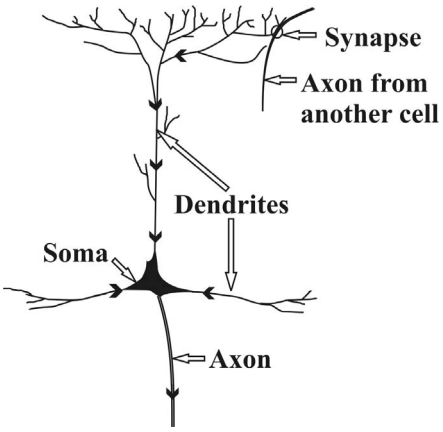


Figure 3: A neuron, with axon, dendrites and the cell body (soma).

The neuron generally has several different incoming connections, called *dendrites*, and one outgoing connection, called an *axon* (Figure 3). The axon of a neuron usually connects to dendrites of other neurons through *synapses*. A neuron works by summing the inputs from all dendrites before initiating an impulse along the axon, towards other neurons. The axons can be more than a meter in length (in the peripheral nervous system), and are often bundled into nerves. In the CNS, these aggregations are often called tracts.

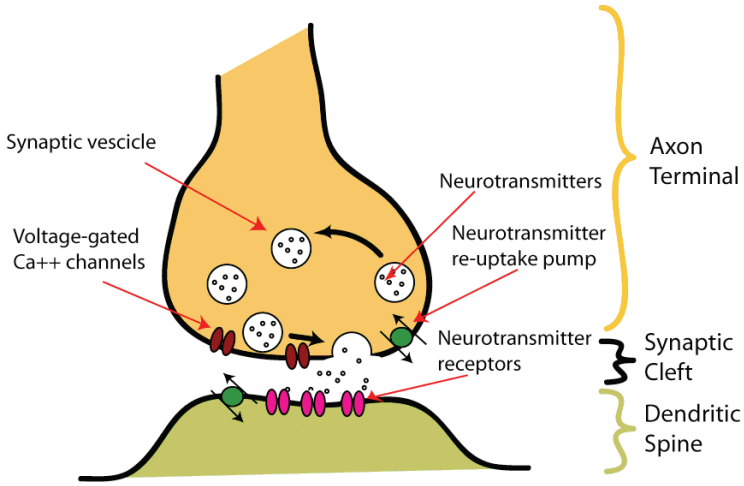


Figure 4: Schematic representation of a synapse. An incoming action potential opens calcium channels in the synapse, and the calcium causes synaptic vesicles to release their contents of neurotransmitter molecules into the synaptic cleft. The molecules diffuse over to the post-synaptic cell, where a fraction of them bind to receptors in the cellular membrane. The receptors trigger in- or out-flux of ions and, when the membrane potential in the post-synaptic neuron reaches a certain threshold, a post-synaptic action potential is triggered. Afterwards, the neurotransmitters are released and may be recycled at the axon terminal.

Information flows through neurons as a depolarization of the cell membrane that causes ion channels in neighboring parts of the membrane to depolarize, and thus spread as a chain reaction that form an action potential that propagate along the axon and dendrites. When an action

potential reaches a synapse (Figure 4), it triggers a number of synaptic vesicles to release its content of chemical neurotransmitters into the synaptic cleft (the space between the pre-synaptic and the post-synaptic cell). The neurotransmitter molecules then attach themselves to receptors located in the cellular membrane on the receiving neuron on the other side of the synapse, triggering an inrush of ions that may start a new action potential and/or other alterations of the cell.

However, not all action potentials that reach a synapse trigger a new action potential in the receiving dendrite. The receiving neuron has thousands (or hundreds of thousands) of synapses. These synapses can be either excitatory, meaning that each synaptic activation increases the probability of the initiation of a post-synaptic action potential, or inhibitory, meaning that the synaptic activation decreases the probability of triggering an action potential. This increased or decreased probability of activation is transient; after a while the neurotransmitter molecules are removed from the synaptic cleft in preparation for receiving the next impulse. The neuron exhibits both temporal and spatial summation. Temporal summation means that many impulses in quick succession are summed and can together trigger an action potential. Spatial summation means that if many synapses are activated simultaneously their sum can trigger an action potential.

The spatial summation corresponds to summing a number of weighted inputs and using a threshold to decide if a binary output signal should be sent. The strength of a synaptic coupling can be enhanced by repeated activation, something that is believed to be the basis of learning. These characteristics are mimicked in the artificial neural network computational model described in section 6.5.

2.2 Brain anatomy

The central nervous system (CNS) consists of the neurons described above, and of cells that support them. Examples of supporting cells are *oligodendrocytes* which wrap nerve fibers in fatty sheets that isolate them from one another and increase the signal propagation speed, or *astrocytes* which help to regulate the composition of the extracellular fluid by, for example, removing excessive neurotransmitter molecules that have leaked from the synapses.

The CNS can be divided into the *brain* and the *spinal cord* (Figure 5). The spinal cord is the information highway which conducts sensory inputs from the body to the brain, and which also conducts commands from the brain to the various parts of the body but primarily to the muscles. The spinal cord does not contribute to the EEG and is therefore outside the scope of this text.

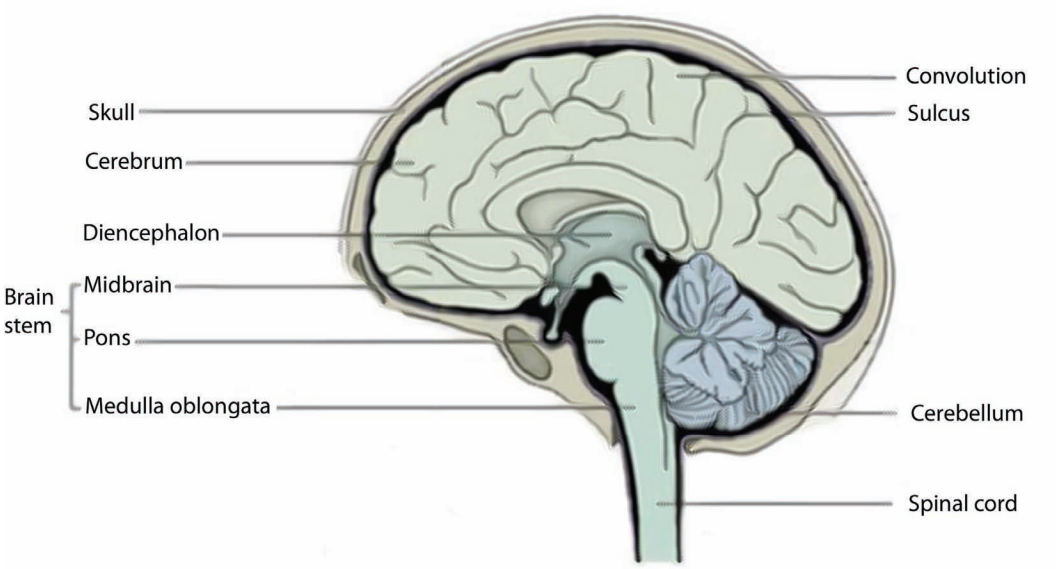


Figure 5: The brain, with the cerebrum, cerebellum and brain stem.

The brain can further be broken down into the *cerebrum*, the *cerebellum* and the *brain stem*. The brain controls the body through electrical impulses running along nerves from the brain via the spinal cord or the cranial nerves to different locations throughout the body, or through chemical messengers (hormones) that are created by various glands and spread to the body via the circulatory system.

2.2.1 The cerebrum and the cerebral cortex

The largest part of the brain is called the cerebrum and it has a layered structure. The outer layer, the *cerebral cortex*, is about 1.5-4.0 mm thick and consists of nerve cells, which are brownish gray in color and therefore called gray matter. The cortex is convoluted, which gives it a large area, and is itself a layered structure with six layers, numbered 1-6 from the surface inward. The part of the cerebrum under the cortex is called white matter. It consists mainly of axons wrapped in *myelin*, a fatty substance that isolates the axons from each other and increases the signaling speed.

The cerebral cortex is involved in most high-level functions such as perception, the generation of voluntary movements, reasoning, learning and memory. Different functions are mapped to different parts of the cortex. It can, for example, be shown that different touch sensors throughout the body are mapped to different parts of the cortex. Other parts of the cortex receive signals from the eyes, or the ears, or are responsible for planning movements. Various cortical areas are highly interconnected and higher functions like relating a visual impression to a remembered name, and then pronouncing the name involve many parts of the cerebral cortex.

2.2.2 The cerebellum

The cerebellum is an important center for coordinating movements, ordered for example by conscious planning performed in the cerebrum, or regulating unconscious movements such as controlling posture and balance. The cerebellum receives information from various parts of the body, such as muscles, skin, eyes and the parts of the brain that are involved in control of movements.

2.2.3 The brain stem

The brain stem consists of the *pons* and the *medulla oblongata*. The brain stem contains all nerve fibers passing between the spinal cord, the cerebrum and the cerebellum, and most of the neuronal bodies of the cranial nerves. It also contains centers for vital functions such as respiration and circulation.

2.2.4 The thalamus

The thalamus is a part of the diencephalon (Figure 5), which together with the cerebrum forms the forebrain. The thalamus is believed to translate incoming information and relay it to the appropriate parts of the cerebral cortex and therefore plays an important part of the generation of EEG signals.

Chapter 3

The Electroencephalogram

The word ‘electroencephalogram’ (EEG) originally denoted the graphs showing the signals obtained by registering potential differences between electrodes placed on the scalp. Over the years, however, the term EEG has been used when referring to the signal itself, the technique to register the signal, or the printed graph. In this thesis EEG is used to denote the signal, unless otherwise stated.

The potential differences between the electrodes are generally believed not to be caused by the action potentials that carry information between neurons, but rather by the postsynaptic potentials that appear in the synapses of the dendrites of the large pyramidal neurons [7]. These neurons are located in layers three and five of the cerebral cortex, but their dendrites stretch throughout several layers towards, and approximately orthogonal to, the cortical surface. The postsynaptic potentials are caused by the release of neurotransmitter substances into the synaptic clefts of the neurons, and each neuron can have tens of thousands of synapses. The postsynaptic potentials can be both excitatory and

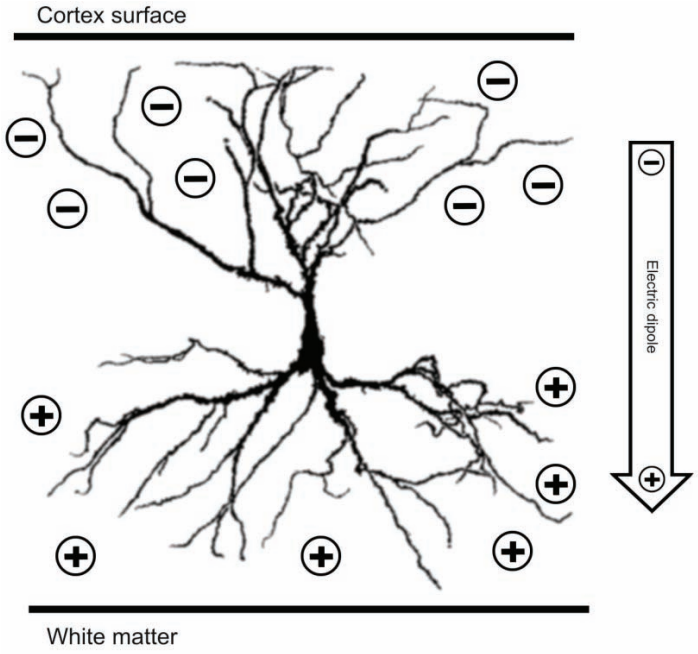


Figure 6: A cortical pyramidal cell with net charges marked by + and -. A relatively larger amount of inhibitory synapses close to the cell body gives a surplus of positive ions, while a relatively larger number of excitatory synapses in the dendrites closer to the surface give a surplus of negative ions. The result is an electrical dipole pointed inwards.

inhibitory. The former type is associated with an extracellular surplus of negative ions while the latter is associated with a surplus of positive ions, giving them different electrical potential. There is a relative difference in the distribution of excitatory and inhibitory synapses between the part of the neuron that is close to the cortex surface and the deeper part which creates an electrical dipole pointed inward, as illustrated by Figure 6. For these signals to be large enough to make a measurable contribution to the electric field on the scalp, large areas of the cortex need to be synchronously active. This synchronicity is due to the fact that groups of neurons are simultaneously stimulated by impulses originating in the thalamus.

The electrodes used when registering EEG can be either metal plates, attached to the skin with the help of a conductive paste, or needle electrodes, stitched through the skin. When using electrodes placed on the exposed surface of the cortex, the recorded signal is called an electrocorticogram (ECoG). The signal levels are usually in the range of 20-100 μV when measured at the scalp, but can be a few millivolts when measured invasively at the surface of the cortex.

The temporal resolution of the EEG is very good, and changes in the activity of the brain can be detected instantly. The spatial resolution is however bad, because signals from different sources are mixed through superposition. This means that the timing of EEG activity can be

determined very accurately, but it is more difficult to determine where in the brain the activity is taking place.

3.1 The 10-20 system

The electrodes are often placed according to the international 10-20 system [8] which defines a number of electrode locations by dividing the head into 10% and 20% intervals using the *nasion* and the *inion* as landmarks for the front to back direction and the *preaurical points* for the side to side direction (Figure 7), which defines 21 electrode positions. The first letter in the electrode name indicates which region it is placed over: *F* for frontal lobe, *C* for the central line dividing the head in a rear and front half, *P* for parietal lobe, *O* for occipital lobe, and *T* for temporal lobe. Numbers in the electrode name are odd for the left hemisphere and even for the right, and increase with increased distance from the midline. A *Z* refers to an electrode placed along the midline.

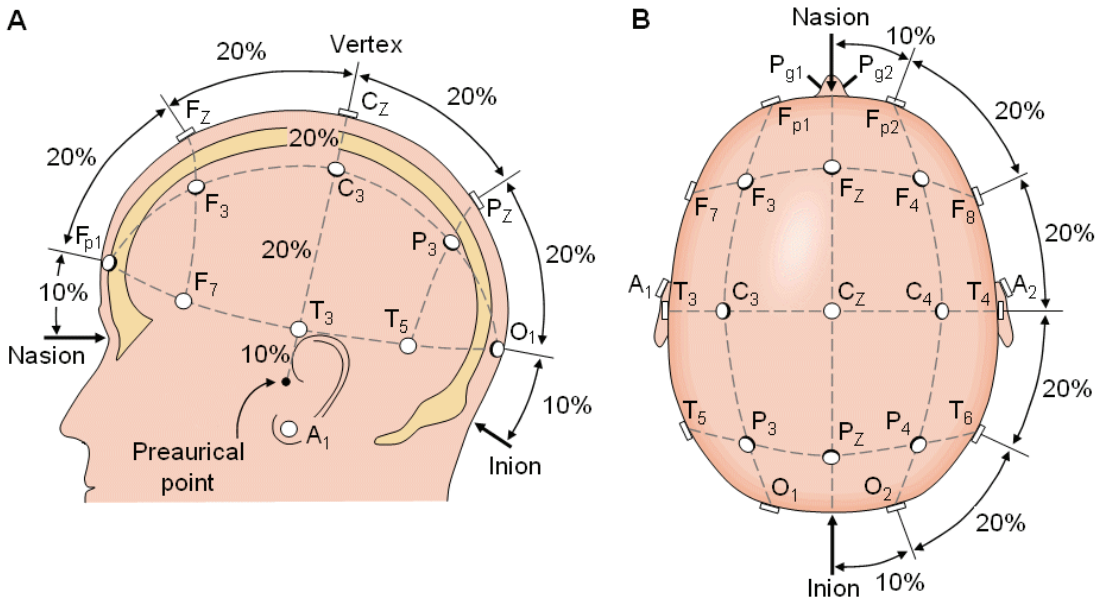


Figure 7: Landmarks and electrode locations of the 10-20 system (Figure from [9]).

3.2 Montages

When measuring one voltage signal at least two electrodes are needed: an active electrode and a ground. However, in practice EEG signals are measured using differential amplifiers where the difference between two electrodes is amplified, and the ground electrode is separate. This is done because the EEG signals are very small, and interference from electrical appliances can cause large problems by making the potential of the patient vary in relation to the measuring equipment, and not involving the ground in the actual measurement decreases this problem. The ground electrode is still needed for connecting the amplifier ground to the patient ground; otherwise potential differences may arise that cause problems with

common-mode interference. It is also needed because of the small bias currents that enters the amplifier through the sensing electrode inputs, and has to return to the patient.

When measuring multiple channels, as is usually the case when measuring EEG, one of the two inputs from each of the differential amplifiers can be connected to a single reference electrode. All the channels then have the same reference when measured, but when viewing the signals the channels can be combined in different ways. The way the channels are combined when displayed is called the montage. Some common montages are listed below.

- The *referential* montage uses the designated reference electrode. The location of the referential electrode varies, but it is usually placed somewhere along the midline of the head so that no emphasis is placed on any one of the hemispheres. Another alternative is "linked ears", where electrodes on the ears are connected to each other and used as reference.
- The *bipolar* montage displays the difference between pairs of usually adjacent electrodes.
- The *common average reference* (CAR) montage displays the difference between the sensing electrode and the average of all channels.

Note that in the bipolar and CAR montages the reference input is mathematically eliminated when the difference of different electrodes is calculated.

In this work the common average reference montage has been used. An advantage of this montage is that interference occurring at all channels is cancelled by the subtraction of the average. Another advantage is that each displayed channel shows the local activity compared to the total activity of the brain, thus improving the spatial localization of the activity as compared to the bipolar montage where the displayed activity is the difference between two adjacent electrodes.

3.3 Types of EEG activity

The EEG activity can be divided into different groups, of which some can be labeled normal and some are considered abnormal, or *pathological* (indicative of disease).

However, normal is a very broad statement in the case of EEG. The signal picked up by EEG electrodes can have many different characteristics and still be labeled as normal, depending on for example sleep stage or age. Some frequency-based categories of EEG are described in Table 1. These designations originally arose because rhythmic activity within certain

frequency bands was found to have biological significance, or associated with certain regions of the scalp.

Most of the cerebral activity is traditionally thought to be found in the range 1-20 Hz, but recent research suggests that important information can be found in the extremely low frequencies that most EEG amplifiers filter away [10].

Name	Frequency limits	Location	Properties
δ (delta)	0.5 – 3.5 Hz	Widespread	Occur in infants and during deep sleep or anesthesia.
θ (theta)	3.5 – 7.5 Hz	Mainly in parietal and temporal lobes	Most prominent in small children and during drowsiness or sleep.
α (alpha)	7.5 – 13 Hz	Rear half of the head	Occur during awake and resting state, high amplitude when eyes closed. Mostly sinusoidal shape.
β (beta)	above 13 Hz	Most common in frontal and central regions	Often divided in two sub-bands, of which the higher frequencies appear during tension and intense activation of the CNS and the lower are attenuated during mental activity.

Table 1: Properties of some common EEG rhythms. The first four frequency bands are not overlapping and cover the whole EEG spectrum, even though the higher frequencies of the β band are today usually named γ rhythms.

The above-mentioned types of activity are “continuous” in the sense that they describe more or less rhythmic activity that goes on for some time, until they are changed by some change in mental state or sleep stage. Many of the types listed below are intermittent, meaning that one kind of activity is interrupted by sudden outbursts of other kinds of activity. In the seizure case, one type of activity (seizures) can also be superimposed on another (*e.g.* burst suppression).

Seizures (Figure 8) are the result of abnormal synchronization of groups of neurons and may or may not give rise to clinical symptoms (symptoms that are easily noticed in the clinic). The type of symptoms depends on which part of the brain that is affected. If it is a motor area of the brain, then the result can be wild and uncontrollable motion of the body. On the other hand, if it is a sensory area in the brain that is affected, the result may be that the person experiences *e.g.* visual flashing or unpleasant odors. There may also be *sub-clinical* seizures that do not cause any detectable symptoms, but are present in the EEG. Some newborn children

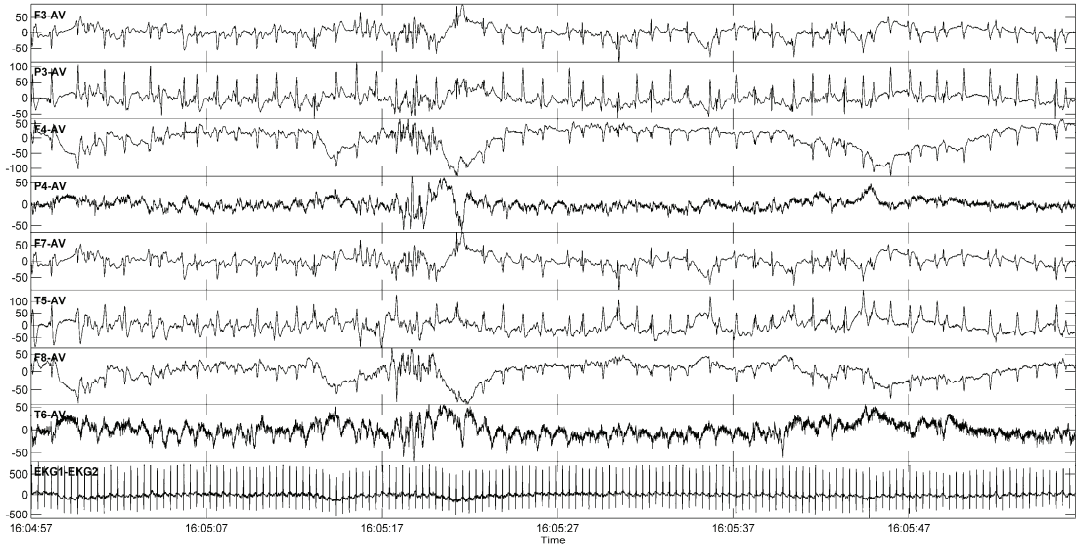


Figure 8: One minute of seizure activity from an asphyctic baby. The seizure is concentrated to the left part of the rear of the brain, and combined with a BS pattern with a burst around 16:05:17. The high frequency noise in P4 and T6 is probably due to that the baby was lying with the right side down. The ECG signal (bottom graph) is included for comparison.

have seizures, the majority of which are sub-clinical. Even sub-clinical seizures may be harmful to the brain, implying that there is a need to detect and classify this type of activity so that children having seizures which are not apparent can be given the appropriate treatment.

Burst-suppression (BS, Figure 9) is one of several indicators of severe pathology in the electroencephalogram (EEG) signal that may occur after brain damage, caused by *e.g.* asphyxia (insufficient oxygen and nutrient supply) around the time of birth [11, 12]. Certain characteristics of this pattern can provide clinicians with important information about the prognosis of the patient, and are thus important in the adjustment of the

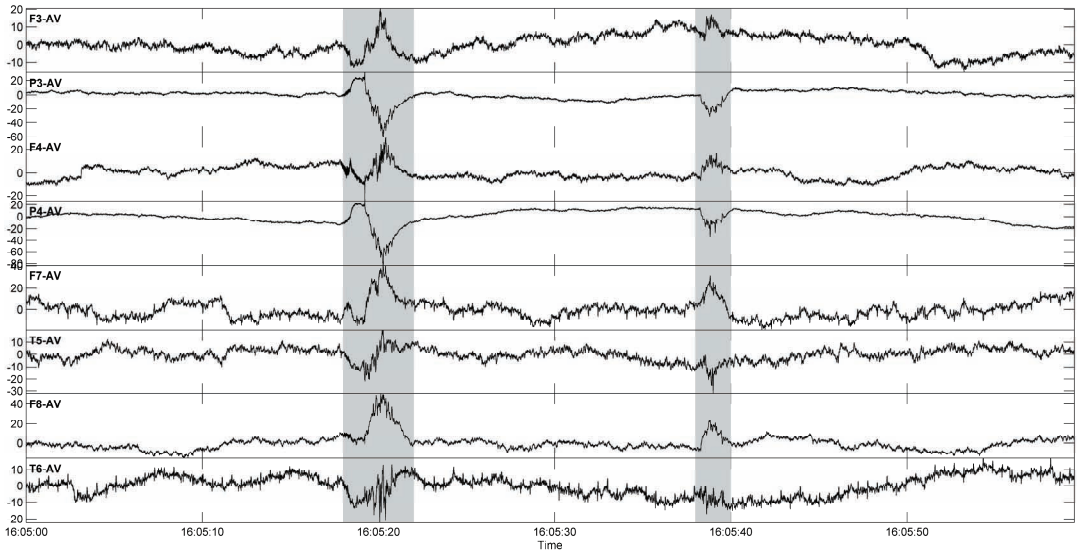


Figure 9: One minute of burst suppression from an asphyctic baby, with two visually classified bursts marked with shading.

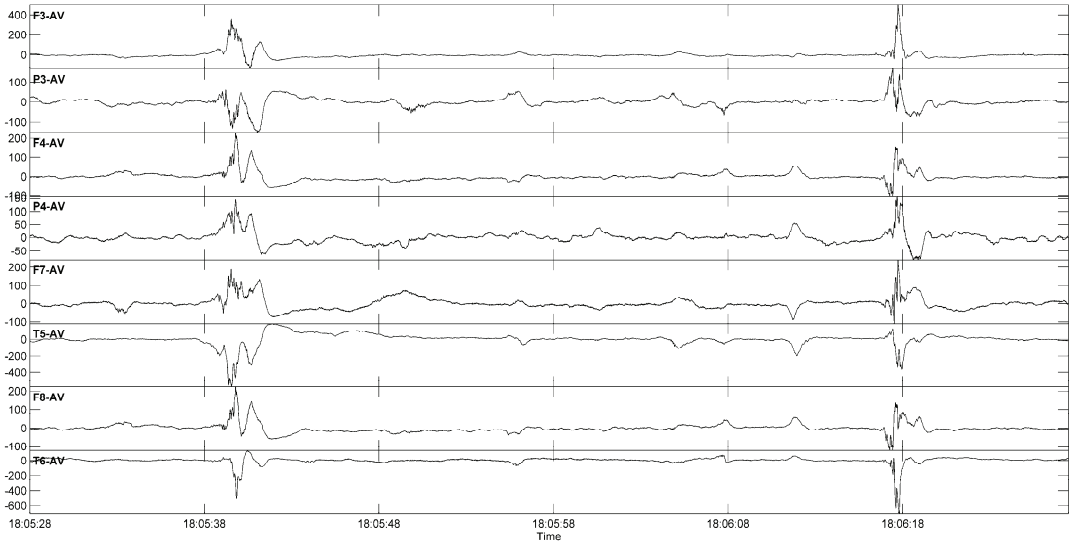


Figure 10: One minute of *tracé discontinue* from a preterm baby.

treatment. Examples of important characteristics of the BS pattern are the length of the burst and suppression intervals, the percentage of suppression activity in a recording, and the spectral contents of the bursts [5, 10, 13].

Tracé discontinue (Figure 10) is visually somewhat similar to burst suppression, but is normal in premature babies [14]. Periods with low amplitude or inactivity alternate with activity with higher amplitude and mixed frequency content. This type of activity dominates most recordings from premature babies, without depending on state, but is most marked during quiet sleep. The interburst durations decrease as the infant matures and the properties of the interburst activity change, and the activity during quiet sleep evolves into *tracé alternant*.

Tracé alternant (Figure 11) is a pattern with alternating active and less active periods that is seen in healthy full-term children during quiet sleep [14]. Instead of the suppression or inactivity that is seen in BS or *tracé discontinue* there are low activity periods that contain low frequency activity. The low activity is interrupted by random high activity periods containing transients with higher frequency and amplitude. *Tracé alternant* usually emerges 34-36 weeks after gestation, but there is a significant overlap between *tracé discontinue* and *tracé alternant* before the baby reaches full term.

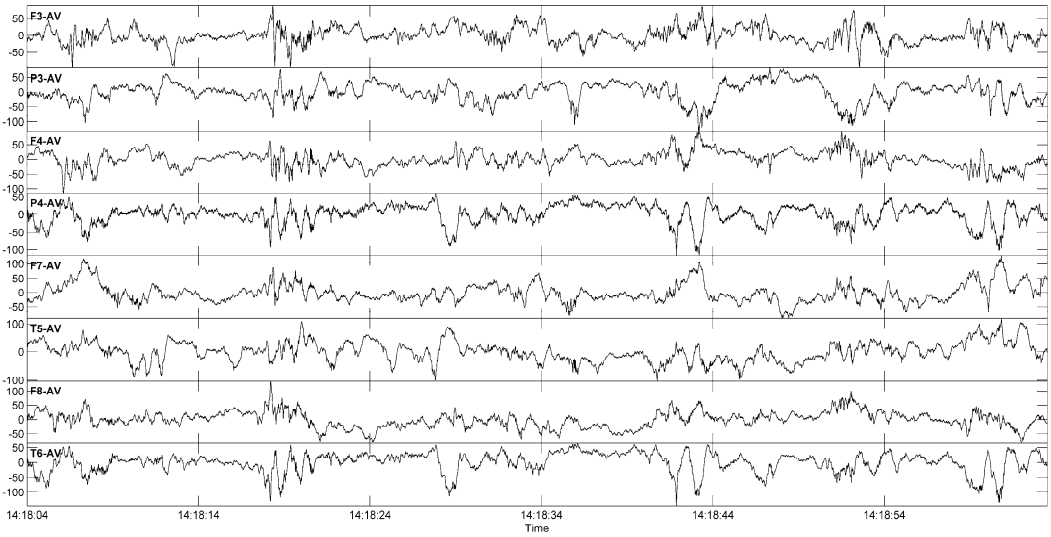


Figure 11: One minute of *tracé alternant* recorded from a healthy baby during quiet sleep.

Figure 8 - Figure 11 were produced using a notch filter at 50 Hz, a low-pass filter at 70 Hz and a linear detrend filter. Often a high-pass filter is used to remove slow baseline fluctuations, but that would change the appearance of the low-frequency components in *e.g.* many of the bursts.

Chapter 4

Data and applications

This section describes the three categories of data that have been used in the project. The data were collected at the Queen Silvia Children's Hospital, which is a part of the Sahlgrenska University Hospital in Göteborg. Details regarding the collection of the EEG signals can be found in the respective paper.

4.1 Data from healthy babies

The common denominator for all papers included in the thesis is solving problems that are of interest when building a monitoring device for the NICU. The most important function of such a monitor is to be able to distinguish activity recorded from a healthy baby with no brain-related problem from activity from a sick baby. Therefore EEG signals were collected from 20 healthy full term newborn babies that had uneventful deliveries. Data were collected during a few hours so that the four behavioural states active awake (AW), quiet awake (QW), active sleep (AS) and quiet sleep (QS) [15] could be included in most of the

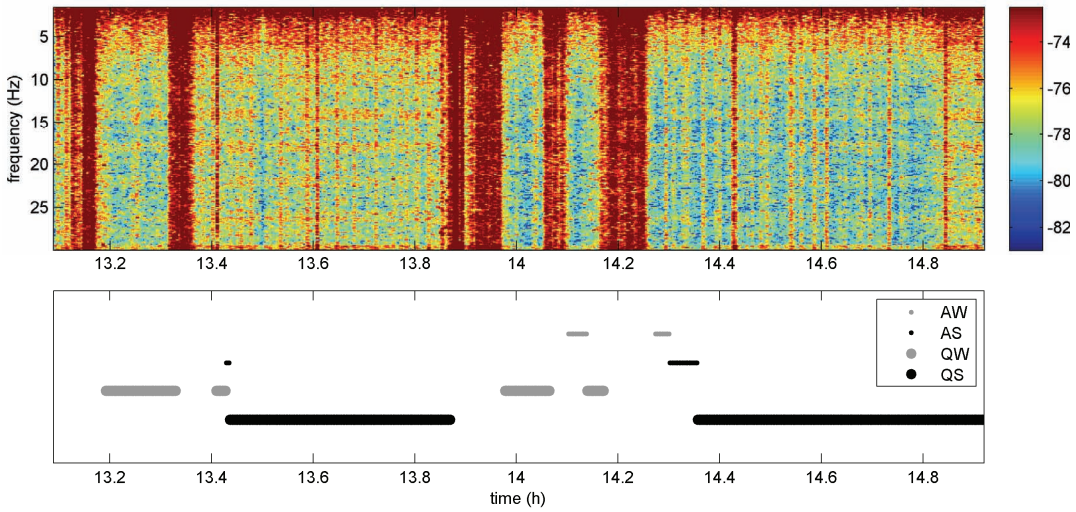


Figure 12: Time-frequency plot of 2h of EEG from a healthy newborn baby, with the behavioural states marked. The unmarked high-power red strips are mainly episodes of crying, with the high power mainly due to muscle artefacts from the scalp.

recordings. These data were divided into the four states based on the observations made by the technician performing the recording and on the classification made by an experienced electroencephalographer. For quiet sleep, EEG of the *tracé alternant* type was chosen. Figure 12 shows a time-frequency plot of one of the recordings. The most prominent features of this plot are the red high power strips, but these are not caused by the actual EEG signal but are mostly due to artefacts produced by the muscles on the baby's scalp during crying.

4.2 Data from post-asphyctic babies

The second group of babies consists of six full-term neonates that were suffering the after-effects of asphyxia during birth. This condition includes lack of oxygen, leading to a build-up of carbon dioxide and a lowered pH value in the blood, and can, among other things, lead to brain damage. These babies all exhibited a severe burst suppression pattern in their EEG. Continuous EEG recordings, of between 6 and 40 minutes in length, were made for each of the six babies. The recordings were then visually classified by an electroencephalographer. The length of each recording was chosen to include at least 10 bursts, and all artefacts were visually identified and marked for later exclusion from the analysis. The total amount of data in this category was 77 minutes.

For evaluation of the BS-related methods in a setting as close as possible to the clinical one, a 32 hour recording from one of the six babies was used. The baby had to be resuscitated after birth and was then intubated and put on a ventilator. The EEG recording was started six hours after birth and continued for 32 hours with a short break around 18 hours after start. The ventilator frequency was set to 40/min (0.7 Hz) initially and was

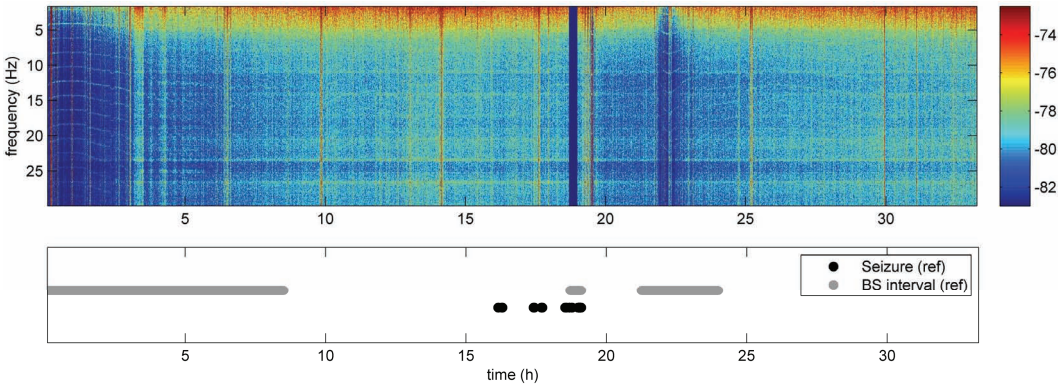


Figure 13: Time-frequency plot of 32h of EEG from a post-asphyctic baby with episodes of BS and seizures marked. The dark blue strip around 18h is a gap in the recording.

changed to 30/min (0.5 Hz) 16 hours after start. At 18 hours after the start of the recording, a dose of Phenobarbital was given to treat seizure activity. The baby was later diagnosed with cerebral palsy. Figure 13 shows a time-frequency plot of this recording, where two long BS episodes can be seen as depressions in the mean power of the signal. Some seizure episodes are also marked, but are too short compared to the scale of the figure to be visible in the plot.

4.3 Data from preterm babies treated with indomethacin

While the two preceding groups of babies all where full term, the babies in the final group were born prematurely, between 25-33 weeks of gestation, and were displaying the intermittent *tracé discontinue* pattern that is normal for this patient group. These babies were all diagnosed with a clinically persistent *ductus arteriosus*, meaning that the shunt that allows blood in the unborn fetus to by-pass the lungs did not close, which it normally does autonomously shortly after birth (Figure 14 and Figure 15). This imposes circulatory disturbances that increase the risk of brain damage. To induce closure of the *ductus arteriosus* the drug indomethacin can be used. There have however been concerns raised to whether this may have a negative effect on the brain because of the drug's vasoconstricting properties, *i.e.* it tends to make the blood vessels temporarily contract and thereby reduce the blood flow. To examine this, seven premature neonates with clinically significant persistent *ductus arteriosus* were recruited. EEG signals were recorded before, during and after an intravenous infusion of indomethacin over 20 minutes, and the effect on the brain was estimated by automatic segmentation of the EEG and measuring the length of all low activity periods.

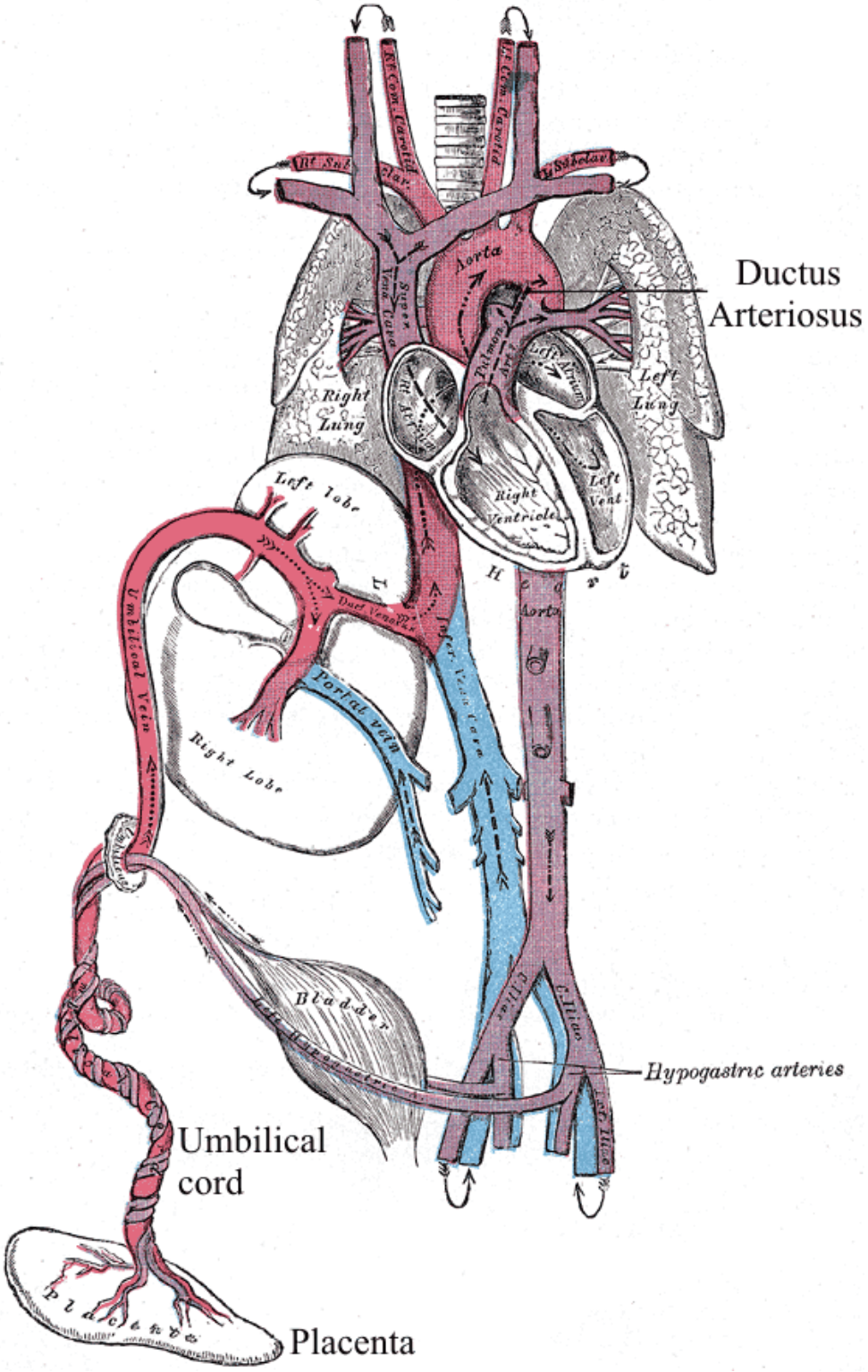


Figure 14: The fetal circulatory system. Since all the oxygen the fetus needs is supplied by the mother via the placenta and the umbilical cord, the lungs are not vital for gas exchange before birth, and the blood is partly allowed to bypass them through the *ductus arteriosus* (figure from [16]).

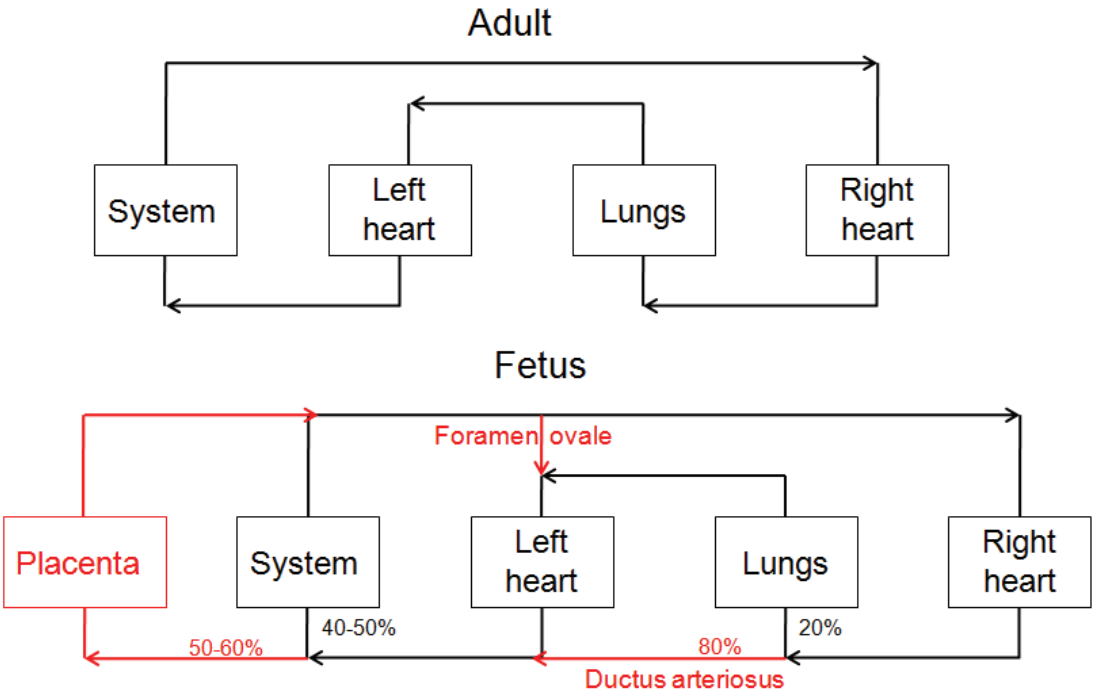


Figure 15: Schematic of the circulatory system in adults and fetuses. The foramen ovale is a hole that allows passage of blood from the right to the left atria.

Chapter 5

Signal processing methods

In this chapter the signal processing methods used in the project are described. All methods mentioned are based on sampled, discrete signals as opposed to analog continuous signals. Many of the measures used to describe these signals are statistical, and related to the shape of the distribution of the signal. Others estimate the frequency contents, the power of the signal, or the information content. These measures are here termed *features*, because this term is common in classification literature, and because the measures are meant to describe different characteristics, or features, of the underlying EEG signal.

5.1 Filtering and pre-processing

The amplitude of the EEG is very low when measured from the scalp, in the range of tens of microvolts, making it very sensitive to interference from surrounding electrical fields created by common electrical appliances. These interferences are typically *common mode*, meaning that they appear on all leads simultaneously. Using the common average

reference (CAR) montage described in section 3.1.2 helps to suppress this interference because the mean of all channels is subtracted.

Usually 50 Hz noise caused by interference from surrounding electrical equipment is present in the EEG to some extent, and therefore a notch filter at this frequency has been used in all cases. In some cases when the EEG has been extremely suppressed, the 50 Hz was still dominant after notch filtering and a low-pass filter with cut-off frequency at 44 Hz with zeros that coincided with the 50 Hz peak was used. In other cases cut-off frequencies of 20 and 70 Hz have been used, depending on the application.

The data was also high-pass filtered. It has been shown that important information in neonatal EEG is present in the very lowest frequencies [10, 13], but because of interference sources, some of the sick babies were *e.g.* ventilated at a frequency of 0.3-0.7 Hz, a high-pass filter was used. In some cases a cut-off frequency of 0.5 Hz was enough, while others required a cut-off at 1.6 Hz. If information from these low frequencies is needed in a monitoring device, once the classification process is finished it would be possible to go back to the raw EEG signal and apply a different processing on the segments of interest using the classification results.

In one of the BS patients (patient 3, Paper II and onwards), an LMS (least mean square) adaptive filter [17] with a separate ECG channel as reference was used to suppress ECG interference in the EEG signal.

5.2 Artifact removal

For the BS segmentation process, periods that were manually identified as artifacts were removed from the set after the feature extraction step, and were not included in the training or evaluation of the classification methods. The reason for including them in the feature extraction is that cutting a signal may introduce sudden steps in the resulting signal when the remaining parts are merged. These steps are of a high-frequency nature and would influence many of the features described in section 5.3. Therefore all relevant parameters should be extracted before any cutting of the signal is performed.

For the classification into different states, no artifacts were removed. Because much longer windows were used, most artifacts were drowned in the real EEG activity and were therefore judged to be negligible.

5.3 Features

The EEG signal is not random, but it is complex enough to be described in stochastic terms as a random process. Medical EEG specialists mainly use visual inspection of the waveforms in the time domain to classify the

activity. However, when building a signal processing system for classifying signals, well-defined measurable features that can be implemented using mathematical functions are needed. Examples of features are the total power of a signal or the distribution of the power with respect to frequency. These two features measure two different properties of the underlying signal and can be independent, because two signals with the same total power could have totally different power spectra.

All the features used here are extracted from *sliding windows*, because it is not possible to calculate spectra or statistical measures on single samples. A sliding window is an interval of the signal, for example one second long. With a sampling frequency of 200 Hz, 200 samples will fall into this window, giving a statistical basis for calculating different parameters. Each parameter yields one value for each window. When the parameters have been calculated for a window, it is moved a certain distance along the signal, for example 0.25 s, and the parameters are calculated for the samples that now are inside the window. Moving the window a distance that is smaller than the width of the window results in a certain overlap, in this example 0.75 s. The operation also results in a reduction in effective sampling frequency, from the original 200 Hz to 4 Hz, since now four samples are used to describe one second of the signal instead of the original 200. This reduces the temporal resolution for detection, but since the quality of the parameters calculated from the data in the window generally increases with increasing number of samples, this results in a trade-off between time resolution and feature estimation accuracy.

When using sliding windows, the length of the window and the distance it is moved in each step have to be decided. If the window is too short, there is too little information in each window to calculate a spectrum of sufficient resolution, or to calculate statistical parameters with high enough reliability. If it is too long, short periods of activity in the EEG could be drowned by other activity surrounding it. In the case of burst-suppression, a window length of one second was chosen, determined by the fact that the shortest bursts in the available material were one second long, and in the state classification case metafeatures, *i.e.* features of the features were used, with a window length of 30 s (see section 5.4.1). The step length that the window is moved is not as critical as the window length. A short step length results in a large overlap between consecutive windows, and produces smooth output signals. A longer step length reduces the effective sample rate of the output signal, thus reducing the computation time for the following processing steps.

5.3.1 Spectral Edge Frequency

The spectral edge frequency (SEF) [18] of a signal is the frequency under which a certain percentage (*e.g.* 95%) of the power resides. This gives a measure of the shape of the frequency distribution, because for an EEG signal there will always be power in the low frequency range. SEF is a common measure used in EEG monitoring, and has for example been used for estimating the depth of general anesthesia, when a patient is made unconscious for surgery.

5.3.2 Three-Hz power

This feature measures the power in a 1 Hz wide band centered around 3 Hz, and was inspired by earlier work [19] in which it was used for detecting BS under anesthesia. BS is often characterized by low frequency content [10, 13] making low frequency features simple and natural choices.

5.3.3 Median

The median [20] is the number that divides a distribution in two equal parts. It can be found by sorting the numbers and taking the middle one. For a normal distribution the median is equal to the mean, but for *e.g.* the exponential distribution it is not. When estimating parameters from a limited set of samples, the median is less sensitive to outliers (extreme values) than the mean.

The behavior of the median for BS segmentation depends on the filtering. The median will act as a low-pass filter with frequency characteristics depending on the window length, and probably capture the low frequency part of the bursts. However, if the signal is high-pass filtered (by processing or by bad electrode coupling) the bursts would be transformed into zero-mean signals, and the median would not be usable as a feature.

5.3.4 Shannon Entropy

The Shannon Entropy [21] is a measure of uncertainty of a random variable, or in other words the information content. It is estimated by defining a set of bins that divide the amplitude range into disjoint intervals $I_1,..,I_U$ and then estimating the probabilities $p(I_1)\ldots\ p(I_U)$ by counting the number of samples which fall into each bin.

$$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$$

In the present implementation, 20 bins distributed between +/-σ were used, where σ is the standard deviation of the EEG amplitude in each window of the signal.

When applied on a BS signal the entropy decreases during the burst intervals. This is due to the fact that the suppression activity is mainly just noise, having high entropy, while the bursts are comparatively more ordered with a combination of low and high frequencies.

5.3.5 Zero Crossings

The rate of zero crossings [18] was investigated as a way to measure frequency contents of a signal in the era before inexpensive computer chips, and it is implemented by measuring how often the signal crosses the zero level. This is not simply related to the frequency of the signal, since a high-frequency component superposed on a low-frequency component may not cross the zero level very often.

5.3.6 Variance

The variance [22] is defined by

$$Var(X) = E\left\{\left[X - E(X)\right]^2\right\}$$

where X is a random variable and $E(X)$ is the expected value of X. The variance is also denoted σ^2, where σ is the standard deviation of the probability density function of X. The standard deviation is a measure of the degree of spreading of the distribution around the expected value. In the time domain, a large standard deviation would imply that the signal contains a large fraction of samples with amplitudes that are far away from the mean, while a low standard deviation implies that the samples are mostly close to the mean.

Numerically, the variance is estimated as the mean of the squared difference between each sample and the sample mean:

$$s^2(x) = \frac{1}{N-1}\sum_{n=1}^{N}(x[n] - \mu)^2$$

where μ is the sample mean. Often the signal is zero-mean, which makes the variance equal to the mean power of the signal (the mean of the squared sample values).

Since burst periods above all are characterized by having higher power than suppression periods, the variance is a natural starting point for BS segmentation.

5.3.7 Skewness

The skewness [22] of a distribution is defined by

$$Skew(X) = \frac{E\left\{\left[X - E(X)\right]^3\right\}}{\sigma^3}$$

and is a measure of how symmetric the distribution is. In the time domain for a zero-mean distribution, a high skewness value would imply that most of the samples with amplitude deviating from the mean are positive.

Numerically, the skewness is estimated as the cube of the sample mean of each sample's deviation from the sample mean, normalized by the cube of the standard deviation:

$$g_1(x) = \frac{\frac{1}{N}\sum_{n=1}^{N}(x[n]-\mu)^3}{\sigma^3}$$

The skewness was mainly used as a meta-feature (section 5.4.1) applied on the other feature signals. For example, the skewness of the residual energy variance was included when classifying BS from other types of EEG (Paper V). The skewness was also considered for BS segmentation in Paper I.

5.3.8 Kurtosis

The kurtosis [22] of a distribution is defined by

$$Kurt(X) = \frac{E\left\{\left[X - E(X)\right]^4\right\}}{\sigma^4}$$

and is a measure of how "peaky" a distribution is. Higher kurtosis means that more of the variance is due to infrequent extreme deviations.

Numerically, the kurtosis is estimated as the mean of each sample's deviation from the sample mean raised to the power of four, normalized by the standard deviation raised to the power of four:

$$g_2(x) = \frac{\frac{1}{N}\sum_{n=1}^{N}(x[n]-\mu)^4}{\sigma^4}$$

The kurtosis was mainly used in the same way as the skewness, as a meta-feature. For example, the kurtosis of the spectral roll-off was used for classifying BS from other types of EEG (paper V). The skewness was also considered for BS segmentation in Paper I.

5.3.9 Spectral centroid

The spectral centroid [23] is commonly used for characterizing sound. It is the "centre of mass" of the spectrum, calculated as a weighted mean of the frequencies in the signal with their magnitudes as weights:

$$c = \left(\sum_{k=1}^{K} kX[k]\right)\left(\sum_{k=1}^{K} X[k]\right)^{-1}$$

where X is the Fourier transform of the signal and K is the number of points in the estimated spectrum. Since the value of K is not physical frequency but depends on the sampling frequency and the number of points used for estimating the spectrum, the value of c will have to be scaled to get the value of the centroid in hertz. When used as a feature, the actual value is usually not of interest, but rather the distribution of the feature for the different classes that are to be classified.

5.3.10 Residual energy variance

The residual energy variance is a measure of how accurately the signal can be predicted by a filter of a given order, and is related to the entropy of the signal. The feature was implemented by finding the eight coefficients of the linear prediction filter that minimized the prediction error in the least squares sense. The order of the filter was determined by estimating the maximum number of peaks in a typical EEG signal. The residual was calculated as the output of filtering the signal through the prediction filter, and then the variance of the residual was calculated.

5.3.11 Spectral flux

The spectral flux [24] measures the change in the spectrum between consecutive windows using the squared Euclidian distance (2-norm) between the spectra:

$$sf = \sum_{k=1}^{K} \left(|X_2[k]| - |X_1[k]| \right)^2$$

where X_1 and X_2 are spectra for two consecutive windows of the signal, and K is the number of points in the estimated spectrum.

5.3.12 Delta flux

The delta flux feature measures the rate of change in the signal by taking the square of the Euclidian distance between consecutive windows of the signal in the time domain:

$$df = \sum_{n=1}^{N} \left(x_2[n] - x_1[n] \right)^2$$

where x_1 and x_2 are the signals in the time domain and N is the number of samples in the window.

5.3.13 Spectral flatness

The spectral flatness [23] is a measure of how flat the spectrum is. A high spectral flatness indicates that the spectrum has a similar amount of power in all bands – like white noise. A low spectral flatness indicates a spiky spectrum, like a mixture of sinusoids.

The spectral flatness is calculated by dividing the geometric mean of the power spectrum with the arithmetic mean:

$$flat = \left(\prod_{k=1}^{K} X[k] \right)^{1/K} \left(\frac{1}{K} \sum_{k=1}^{K} X[k] \right)^{-1}$$

where X is the Fourier transform of the signal and K is the number of points in the estimated spectrum.

5.3.14 Spectral roll-off

The spectral roll-off is based on the same principle as SEF95 and measures how wide the spectral distribution is, but uses the frequency for 85 % of the energy instead of 95. This feature is common in sound processing [25], hence the different name.

5.3.15 Cepstrum-based coefficients

The cepstrum is defined as the inverse Fourier transform of the logarithm of a spectrum and contains information about the rate of change in the different frequency bands. In the current implementation, ten triangular overlapping windows were applied on the spectrum before transformation, resulting in ten coefficients that each is related to a frequency band.

5.4 Feature characteristics and post-processing

After feature generation some steps need to be taken to prepare the feature signal for the classification algorithms. Some of these steps, such as the normalization procedure, were chosen to reduce the differences between recordings from different patients that were found to be a problem when segmenting BS signals (5.4.2). Others, such as the channel combination and feature smoothing, were chosen based on prior knowledge of the expected characteristics of the BS activity.

5.4.1 Meta-features

For the state classifier (paper V), the features were summarized by applying the four statistical measures mean, variance, skewness and kurtosis on the feature signals from each 30 s non-overlapping epoch in the data, resulting in 88 features based on the original 22. This was done because sleep-stages and BS go on for some time, at least a couple of minutes, while *e.g.* a single burst can be as short as one second. These measures describe different properties of the feature signals distributions. For example, the mean and the variance are measures of where the distribution is located and how wide it is, while skewness and kurtosis measure its shape. These measures are in this paper called *metafeatures*, because they are features of the features of the EEG.

5.4.2 Normalization

The feature signals represent different properties of the EEG and have different amplitude ranges. For many classifiers it is an advantage to have features with approximately the same dynamic range, because of the way the thresholds or weights are adapted in multiple dimensions. A standard way to achieve this is to normalize all feature signals by subtracting the mean from the signal and dividing it with its standard deviation, producing signals with zero mean and unit standard deviation. This method works well for signals where there is approximately the same amount of the different classes in the different data sets, but in burst-suppression signals this is not the case. Data from different patients contain different amounts of bursts, and just subtracting the mean would make the mean for the suppression part end up at different levels.

To cope with this, a different normalization procedure was developed for the BS segmentation problem. The first and 99^{th} percentile of the histogram were found, and fitted into the interval zero to one, giving all feature signals an equal magnitude range. The percentiles were chosen, as opposed to just using minimum and maximum values, to make the normalization procedure less sensitive to extreme values. The method has the drawback that if the signal to be normalized does not contain both classes, for example if the whole recording only contains suppression, the mean of the suppression will be pulled towards the value the burst part would have had. Therefore it must be known that a signal actually contains *e.g.* both burst and suppression before this normalization procedure is carried out.

In the state classification case, the signals were not normalized because the signals were previously divided into short epochs of a few minutes each, making it hard to do a good normalization without removing the feature characteristics that are needed for classification. An alternative could have been to go back to the uncut signals and normalize it, and then cut the normalized signal. However, the question is how this would be done in a practical bed-side system. If the system is to be operational at the instance it is turned on, it cannot first collect data from all states before it starts to do classification. Because of these reasons, the normalization was left out.

5.4.3 Feature smoothing

Because of the sliding window used in the extraction of the feature values, the feature signals have a lower effective sampling frequency than the original EEG signal. The value of the features for each segment of the EEG varies a lot, but when detecting events that are a few seconds long it is better to have a smoother signal to capture trends in this time range. Therefore the feature signals were smoothed using a six-point triangular

window function. This is performed by convolving the signal with the window function, meaning in practice that for each sample in the signal, the triangular function is centered at that particular sample, and the six samples surrounding the centre sample are multiplied by the corresponding window values. The mean of the six resulting values is the output value for that position in the signal. This procedure smooths the signal by averaging over a small interval of the input signal, but with emphasis on the center value because of the triangular shape of the window function.

5.4.4 Channel combination

After feature generation, the feature signals were combined per feature by forming a new signal as the median value of the eight feature channels. This was motivated by the fact that both BS and the different behavioral states are considered to be global phenomena with certain characteristics appearing in a majority of the EEG channels. Using the median over the channels also removes disturbances present in one or a few of the channels.

Reducing the number of channels involved in the classification also reduces the computational complexity for the feature selection and classification steps.

Chapter 6

Classification methods

Classification methods are algorithms that use properties of objects to place the objects in different classes. The objects can be of many different kinds, we could, for example, use the chemical composition of bones to determine where a murder victim comes from as in the tv-series *Bones*, or we could use a thermometer and a glance out the window as features for classification of the weather to be able to choose the appropriate clothes to wear in the morning. In this thesis the features described in the preceding chapter are used to classify the activity found in the EEG activity of newborn babies.

Classification algorithms can be purely statistical, and apply a threshold on the data based on prior knowledge of the distribution of the data with respect to the class membership. They can also be based on machine learning, and iteratively learn patterns from examples. In this chapter, the classification algorithms known as Fisher's linear discriminator (FLD), artificial neural networks (ANN), support vector machines (SVM) and hidden Markov models (HMM) are described. All except HMM have the

properties that they typically are trained using pre-classified multidimensional input training data, and form classifiers by trying to find the optimal boundary between the classes in this multidimensional space. The ANN does this by iteratively moving a number of linear planes based on the error in each step, while the SVM optimizes a nonlinear boundary that maximizes the margin between the classes. FLD is a linear projection and, based on scatter matrices for the classes, the projection onto the line that gives the best linear separation is found analytically. HMM on the other hand works in a different way, and includes information about how the process develops in time described as state and transition probabilities. The HMM classifier is described in the end of the chapter.

All these methods are based on finding a classifier using training data, and then applying it to test data. The reason for choosing this type of algorithms is that we envision a system that is trained on a number of manually classified training cases, once and for all. The system could then be used for classifying new cases without the need of any manual adaptation when it is applied to a new patient, something that would be of great benefit in a hospital setting.

The method of maximum likelihood for classification based on one-dimensional inputs is presented for comparison, and because it is a part of the FLD method.

6.1 Classifier accuracy estimation

The accuracy of a classifier can be determined by dividing the data into two sets, one for training the classifier and one for testing it. If a very large data set was available it could be divided into *e.g.* 25% test data and 75% training data. Because the test set is often used to find the best setup of the classifier and the best set of features for the task, there is often a need for a validation set, that is only used to evaluate the performance of the classifier after all parameters have been fixed.

However, it is often the case that there is only a limited amount of data available and more efficient methods to estimate the accuracy are needed. A common way is to use *m-fold cross-validation*, where the data are divided into m sets of equal size n/m, where n is the total number of data samples. Then the classifier is trained m times, each time with a different set held out for validation, and the final accuracy is the mean of these m instances. When using genetic algorithms for feature selection (described in Chapter 7.3) this is however not practical, because the feature set is changing in each iteration and the result would be that different sets of features are used for different parts of the data.

6.2 Performance measures

Different performance measures have been used during the project. Measures that are easy to relate to are sample sensitivity and specificity. Sensitivity was defined as the percentage of class 1 (*e.g*. the burst samples) that were correctly classified, and specificity as the percentage of class 2 (*e.g*. the suppression samples) that were classified as class 2.

Based on the sensitivity and specificity measures, receiver operating characteristic (ROC) curves [26] were formed by taking the raw output from the respective algorithm and using a range of thresholds covering the range of the outputs. These curves show the potential of a detector, displaying all possible trade-offs between sensitivity and specificity. The area under the curve (AUC) provides a single number for comparing the classifiers, with the value one representing the perfect classifier.

Because of the small number of patients, the individual performance measures were not averaged, but kept separate for comparison of the performance of the methods for the different patients in papers I-IV.

In paper V when the genetic algorithm was used for feature selection, a single number performance measure was needed. Therefore, the probability of error was used, calculated as

$$P_{err}=P(\text{class 1})P(\text{class 2}|\text{class 1})+P(\text{class 2})P(\text{class 1}|\text{class 2})$$

where P(class 1) is the probability of class 1 and P(class 1|class 2) is the probability of classifying a given sample as class 1 when it does in fact belong to class 2, and the other way around. The formula gives a weighted sum of the two misclassification probabilities, where the weights are the proportions of the two classes in the data.

The performance is also presented as a confusion matrix, where the probabilities for misclassifying each of the classes can be found. Table 2 gives an example of how a confusion matrix can be defined.

Table 2. Definition of the confusion matrix entries.

		True class				
		BS	QS	QW	AS	AW
Predicted class	BS	P(BS|BS)	P(BS|QS)	P(BS|QW)	P(BS|AS)	P(BS|AW)
	QS	P(QS|BS)	P(QS|QS)	P(QS|QW)	P(QS|AS)	P(QS|AW)
	QW	P(QW|BS)	P(QW|QS)	P(QW|QW)	P(QW|AS)	P(QW|AW)
	AS	P(AS|BS)	P(AS|QS)	P(AS|QW)	P(AS|AS)	P(AS|AW)
	AW	P(AW|BS)	P(AW|QS)	P(AW|QW)	P(AW|AS)	P(AW|AW)

For example, P(BS|BS) is the probability of classifying a sample as BS when the true class really is BS, and P(BS|QS) is the probability for classifying a sample as BS when the true class is QS.

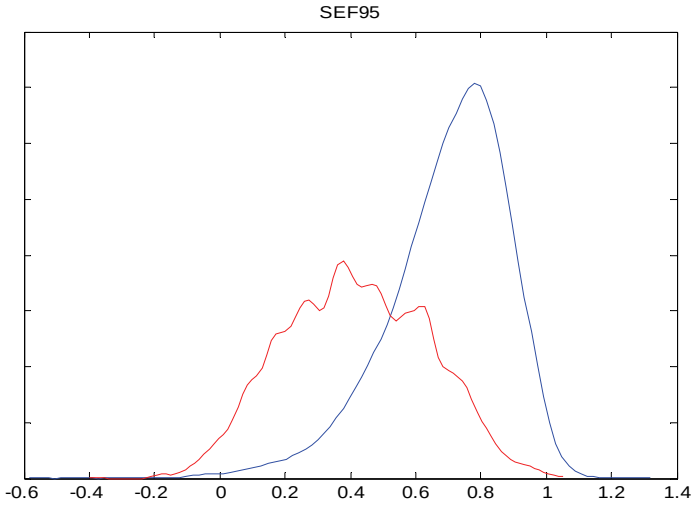


Figure 16: Example of pdf for the feature SEF95 divided into burst (red) and suppression (blue).

6.3 Maximum Likelihood Classification

Maximum likelihood (ML) [27] classification is a simple method for classifying samples based on their probability density functions (pdfs) [28]. A pdf is a statistical description of a signal that shows how probable different values are, and can be estimated by using a histogram (Figure 16).

In a two-class case, the conditional probabilities are compared:

$$\text{Decide } \omega_1 \text{ if } p(x|\omega_1) > p(x|\omega_2); \text{ otherwise decide } \omega_2$$

The above decision rule is interpreted as follows: for each sample, examine the value and look up the probability for that value in the pdfs for the two classes. Then choose the class with the highest probability.

The ML algorithm described above is equivalent to the *Bayesian decision rule* with equal prior probabilities [26]. Prior probability in this case means that the classification rule is weighted with the probability of each class, which can be estimated by calculating the proportion of the samples in the training data that belong to each class. This was not used in the BS segmentation case because bursts generally are much less likely than suppression, and weighting the decision rule would decrease the sensitivity significantly.

The ML classifier was implemented by using training data to estimate the pdfs by *e.g.* calculating and smoothing the histograms for the two classes (Figure 16). The areas of the histograms were then normalized to one. Unseen data are classified by taking each sample and comparing it with the two pdfs. The one that gives the maximum probability value is chosen as the most likely class for the sample. The input value that produces the

same probability value for the two classes is the threshold of the ML classifier.

The ML classifier differs from the other ones in that it cannot handle multidimensional feature data without estimating parametric multi-dimensional pdfs, which requires large amounts of data. A way to avoid this is to use a projection like Fisher's Linear Discriminant (see next section) to reduce the dimensionality of the feature data. Another way is to use the ML classifier on the different features separately, and then combine the results by *e.g.* voting. These methods, however, have the drawback that any nonlinearities in the high-dimensional feature data are disregarded. Therefore, ANN and SVM were used for exploring the benefits of using classifiers that can adapt to nonlinearities in the data.

6.4 Fisher's Linear Discriminant

Fisher's linear discriminant (FLD) is an analytical way to derive a linear projection [29] of a multidimensional dataset onto the line that gives the maximum separation of two classes. This reduces the dimensionality to one, and ML classification can be used to calculate the final classification.

A projection of a vector **x** onto a line in the direction of the vector **w** is written

$$y = \mathbf{w}_F^T \mathbf{x}$$

where $\mathbf{w}_F$ is a vector, **x** is d-dimensional signal and *y* a one-dimensional signal. In the present case, each vector **x** contains one value from each of the five features and represents one second of the EEG signal in feature space. The goal here is to find the line defined by **w** that gives maximum separation of the classes in the output one-dimensional signal *y*.

It can be shown [26] that the vector **w** is given by

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

where $\mathbf{m}_i$ is a vector in which each element is the mean of the corresponding feature signal for class *i*, and $\mathbf{S_W}$ is the within-class *scatter matrix* which is proportional to the sample covariance matrix for the entire dataset (both classes).

The projection is then applied to the test data, resulting in a mapping from the N-dimensional feature signal to a one-dimensional signal. Applying a threshold to this signal produces the detections. The threshold can for example be found through maximum likelihood classification.

The FLD was used both for segmenting BS activity, segmenting *tracé discontinue* and for classification of BS versus normal activity. In the two latter cases it was used together with a genetic algorithm for feature selection. The method was found to be not quite as good as the more

advanced methods described later in this chapter. However, it has the advantage that it is much quicker to train, making it the classifier of choice for feature selection when many different combinations of features have to be evaluated.

6.5 Artificial Neural Networks

Artificial neural networks (ANN) is a collection of methods used for training a classification system using pre-classified training data [30]. The inspiration for the model and the names of the structures comes from neurology, and the model is thought to mimic the behavior of the neural networks in the human brain. The neural network is relatively simple to implement, and is represented by one matrix per layer.

6.5.1 Multi-layer feed-forward neural network

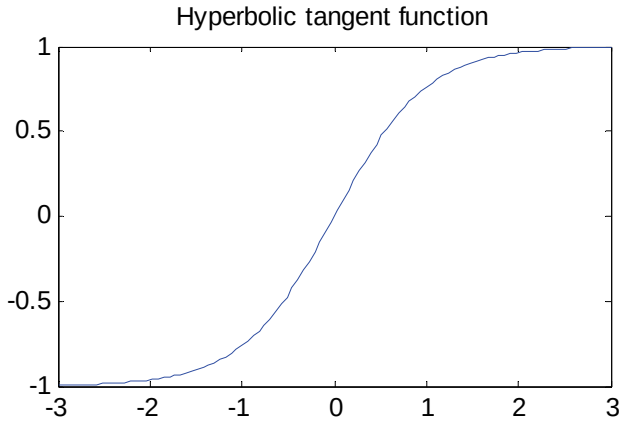
The perceptron, or artificial neuron, is the simplest type of ANN. It is a simplified model of a biological neuron, having a structure with any number of inputs and one output. There is also a bias input. The neuron implements a linear decision boundary in a space of the same dimensionality as the number of inputs. The neuron can be written

$$y = \varphi(\mathbf{w}_P^T \mathbf{x} + b)$$

which is the output of the function φ when the input is the inner product of the d-dimensional input signal $\mathbf{x}$ and a weight vector $\mathbf{w}_P$, plus the bias term b. The weight vector describes the orientation of the plane, and the bias term specifies the position.

Compared with the FLD, the neuron equation is equivalent to the FLD projection except for the bias term and the activation function. The bias in the perceptron acts like a threshold; the projected signal is offset so that taking the sign, or thresholding at zero, gives the detections. The bias for the perceptron is found through a learning rule and the threshold in the FLD is found by *e.g.* ML, but in both cases they are found using the training data.

There are many alternative activation functions, for example the simple signum function that acts as a zero-level threshold and outputs a binary signal. Another common class of activation functions are the *sigmoid* functions. These functions have an S-shape that introduces a nonlinear property in the network. The S-shape gives the functions the property that they act linearly around the origin, while large inputs are compressed into the interval zero to one. In the implementation used in this project, the hyperbolic tangent function was used (Figure 17).



$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Figure 17: The hyperbolic tangent function, an example of a sigmoid function that is commonly used as activation function in neural networks.

The neuron can be graphically visualized as in Figure 18. In this example, it has three inputs and one output, but the model allows for any number of inputs. Each input i is associated with a weight w_i and the bias term is a constant unit input associated with the weight w_0.

The multilayer perceptron (Figure 19) consists of any number of layers of neurons, although two (or three, if the inputs are considered to be a layer) is the most common number of layers. The outputs from the first layer are connected to the inputs of the next layer in a strictly forward manner, hence the designation “feed-forward” network.

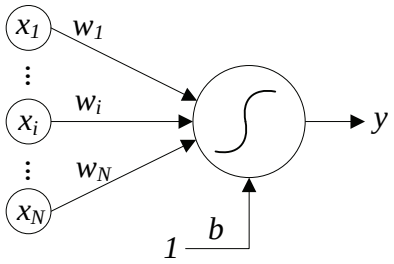


Figure 18: A neuron with N inputs, bias and one output. The S-shape symbolizes a sigmoid activation function.

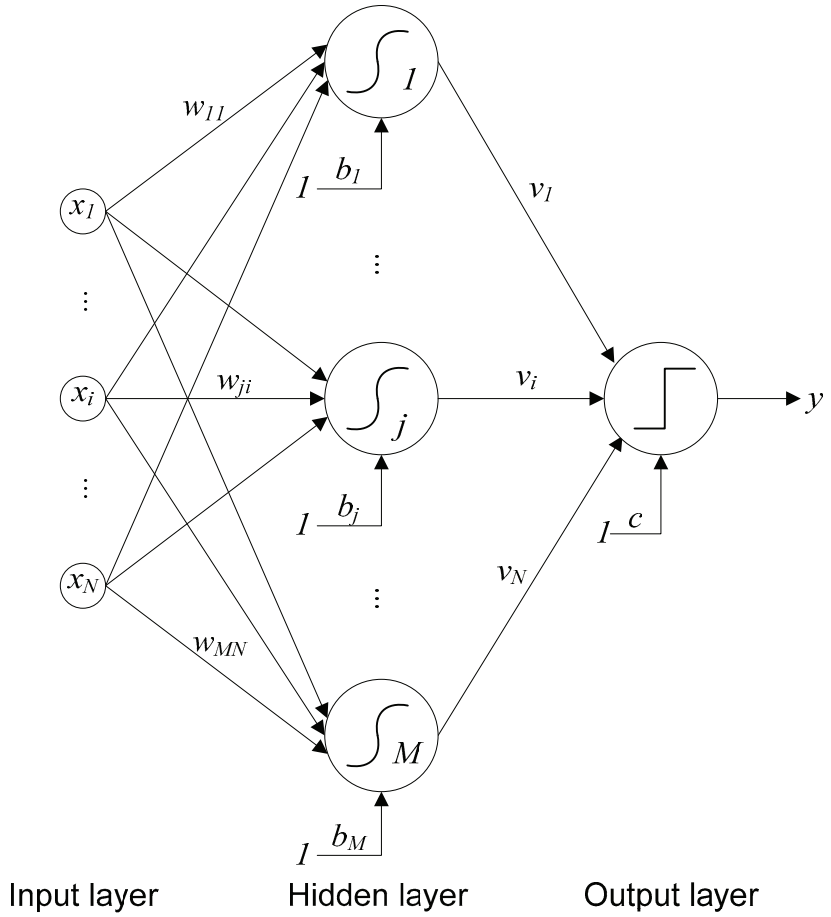


Figure 19: Three-layer perceptron with input layer to the left and output to the right. The hidden layer has sigmoid activation functions and the output neuron has a signum (threshold) activation function. Each arrow, connecting one layer with the next, is associated with a weight. Each input value to a neuron is multiplied with its respective weight before all inputs to the neuron are added. The bias term can be seen as a constant input with value one, and the bias value is the weight of the bias input.

The input and output layers are the only ones that need to be observed by the user; the layers in between are called *hidden* layers. Usually the network is designed for a two-class problem, with an output layer consisting of one neuron that sums and thresholds the inputs from the hidden layer and gives +/-1 as output. The number of neurons in the hidden layer depends on the nature of the problem, *i.e.* how complex the boundary is which is needed to divide the classes in feature space.

Physically the feed-forward network can be interpreted as a combination of a number of linear classifiers that together form a nonlinear boundary in the multidimensional feature space. Each of the neurons in the hidden layer implements a linear hyperplane and the output neuron combines them into a nonlinear decision boundary.

6.5.2 Back-propagation training

The network is trained by using an algorithm called error back-propagation. During training, data are presented at the input of the

network, and the resulting output is compared with reference data. If the output is wrong, the error is propagated back through the network, and all weights are updated by an amount proportional to the error [26]. The training process can be viewed as an iterative multidimensional optimization, and the training algorithm searches a multidimensional error surface for the global minimum. During this process it is in some instances possible that the algorithm stops in a local minimum instead of the global one. The network is initialized with a set of randomly selected weights, and depending on the starting point the final network may converge to the optimum or stop in a sub-optimal state. A way to reduce the risk of using a sub-optimal classifier is to train a number of instances of the network and choose the one with the lowest training error.

A problem that is common to all classifiers is the risk of overfitting, which means that the classifier is so well trained that it perfectly fits the training data, even with regard to noise. This leads to poor generalization: the classifier performs badly when presented with unseen examples. When training a feed-forward ANN, a common way to avoid overfitting is to use a separate test set that is passed through the network at certain intervals during the iterative training process. When the error starts to increase for this test set, the training is stopped, a method called early stopping.

There are some heuristics commonly used to help the training process to get to the right solution more quickly and to reduce the complexity of the classifier. Adding a momentum factor to the training algorithm helps the classifier to speed up when moving along nearly planar error surfaces, and may help in moving out of local minima. Weight decay is a technique that reduces the influence of weights that do not help to reduce the training error, and involves the multiplication of each weight with a fraction after each training step. Weights that are not reinforced by the training will then gradually decay to zero.

6.5.3 Using the neural network for BS segmentation

The neural network was applied on the burst-suppression segmentation problem using five pre-selected features. It was implemented as a feed-forward network with one hidden layer containing ten neurons, a number empirically found to be “large enough”. The output layer contained one neuron and the input layer contained five inputs, one per feature. The network was implemented in Matlab using basic operations such as for-loops and matrix multiplications.

The training of the network was performed using error back-propagation. The influence of redundant neurons was reduced using weight decay.

Overfitting was avoided by using early stopping, meaning that the error on the evaluation set was monitored and a training length that resulted in a reasonable trade-off for all different patients was chosen. Figure 20 shows

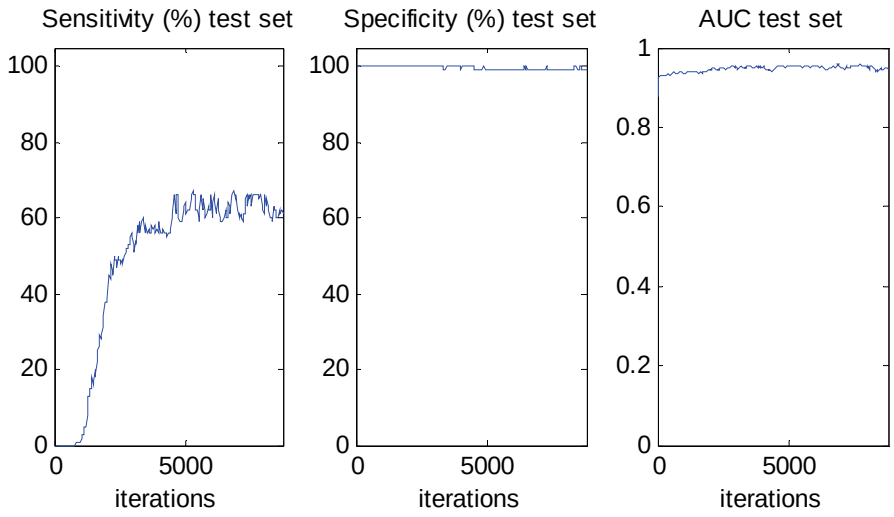


Figure 20: Sensitivity, specificity and AUC as functions of the number of training iterations in the ANN.

an example of how the sensitivity, specificity and AUC evolve during training. Initially, all samples are classified as suppression, resulting in 100% specificity and zero sensitivity. The sensitivity is then gradually increased while maintaining a high specificity. This relation is due to the uneven distribution of burst and suppression in the data: there are many more suppression samples than burst samples. The AUC is found by sweeping a threshold over the entire range of the output signal, and is very high except for the very first iterations. The contradiction between the gradual increase in sensitivity and the very quick increase in AUC arises because the AUC calculation does not depend on a specific threshold value.

Figure 21 is a visualization of the network weights. These determine how large an influence the inputs have on each hidden neuron, and how much influence each hidden neuron has on the final summation in the output neuron. As can be seen in the figure, about half of the neurons have weights close to zero, meaning that they contribute very little to the final result. This effect is achieved through weight decay, and show that some neurons (*i.e.* features) could be removed to reduce the computational complexity.

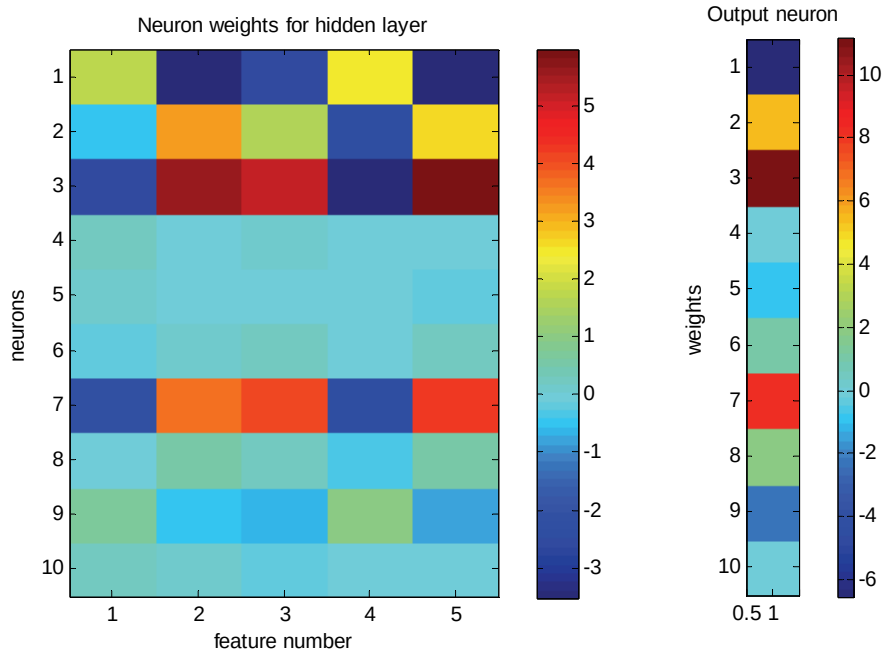


Figure 21: Visualization of the weights of the trained network. Each of the ten hidden neurons is connected to all input features via weights, and each hidden neuron is connected to the output neuron via another set of weights. Each training of the network results in a unique set of weights. In this case, weight elimination has driven six of the hidden neurons to very low values, meaning that they do not contribute much to the result.

6.6 Support Vector Machines

While ANNs are based on minimizing the empirical risk by iterative training, and use a number of heuristics to improve performance and avoid stopping in local error minima, the support vector machine (SVM) [31] has a more solid theoretical foundation. It consists of two main parts:

- Using kernel functions to map the input data to a higher-dimensional space
- Maximizing the margin of classification

In short, the first part means that the SVM uses a function to transform the problem nonlinearly and then construct a nonlinear decision boundary using linear techniques. The second part means that while a classifier such as an ANN settles for any boundary that separates two classes in the training data from each other, the SVM finds the one that maximizes the margin (Figure 22), a property that often will increase the generalization ability of the classifier.

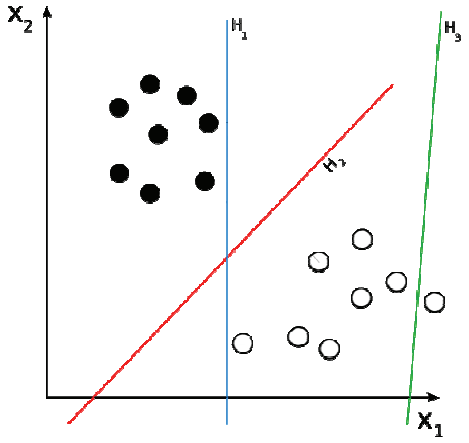


Figure 22: Two classes can be separated by many boundaries, but only one maximizes the margin.

The SVM also has the property that only the samples near the decision boundary are involved in the computations. While the ANN is trained by updating its weights with an amount proportional to *any* randomly chosen misclassified training sample, the SVM can be trained by choosing the current *worst*-classified pattern. When the training is finished, these patterns would be the *support vectors*, the patterns on the margins that define the optimal separating hyperplane (Figure 23). This method of training the SVM, however, is feasible only for very small datasets, because it would mean that all training samples have to be searched in each iteration in order to find the worst-classified sample. Instead other more efficient algorithms are used.

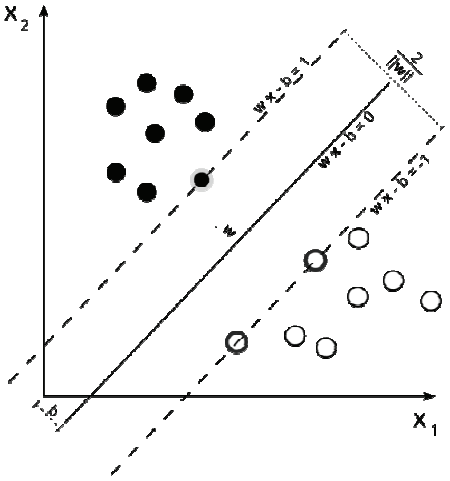


Figure 23: The support vectors are the training samples that fall on the margin.

6.6.1 Kernel trick

Kernel functions are used for transforming a classification problem nonlinearly into a higher-dimensional space. Theoretically, if a problem is not separable in the basic feature space, there exists a space of higher dimension where the problem is separable. This is, for example, always true in the extreme case where the number of dimensions is equal to the

number of samples. There are infinitely many possible kernel functions, but the one used here is the Gaussian radial basis (RBF) kernel:

Gaussian RBF kernel: $$K(\mathbf{x},\mathbf{y}) = \exp\left(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{d}\right)$$

In the Gaussian RBF kernel, *d* is an input parameter that sets the width of local Gaussian function centered at each support vector.

The transformation is never explicitly calculated, because the transformed space may be of very high dimensions, and require complex computations. Instead, the calculations are done implicitly by using the kernel function with the feature values as input.

6.6.2 Training and training parameters

The SVM maximizes the margin between separable classes; but in reality, classes are rarely completely separable. The regularization parameter *C* controls how many of the samples are allowed to be misclassified, thus determining the smoothness and complexity of the classification boundary. Without regularization the algorithm could adapt a very complex boundary, correctly classifying all the training samples, but this would in many cases lead to overfitting and bad generalization ability.

The Gaussian kernel also has a control parameter, *d*, that determines the radius of the Gaussian function, thus controlling the radius of influence for each training pattern.

The SVM is a deterministic minimization problem, and has the advantage over *e.g.* neural networks that the same data and control parameters always produce the same result, while the randomly initiated neural network may stop in a sub-optimal local minimum. The kernel and regularization parameters are not analytically derivable, however, and need to be found through iteration. This can be done through a grid search: the performance of the classifier is measured at a number of (*C*, *d*) combinations, and the combination with the best performance is chosen.

6.6.3 Using the support vector machine for BS segmentation

When the SVM was applied to the BS classification problem, the RBF kernel was used with width parameter and misclassification weight determined by a grid search. The performance values varied significantly for the different patients, so the final values were chosen to give decent performance for all cases, rather than selecting a parameter setting that would have given almost perfect performance in some cases and zero performance in others. Figure 24 shows an example of AUC values plotted in a grid of parameter values. Note that the surfaces for the different patients are located at different heights, and that the gradients for

increasing performance point in different directions. This makes the optimization of the parameter values difficult, and it will end up as a trade-off that gives decent performance in all patients.

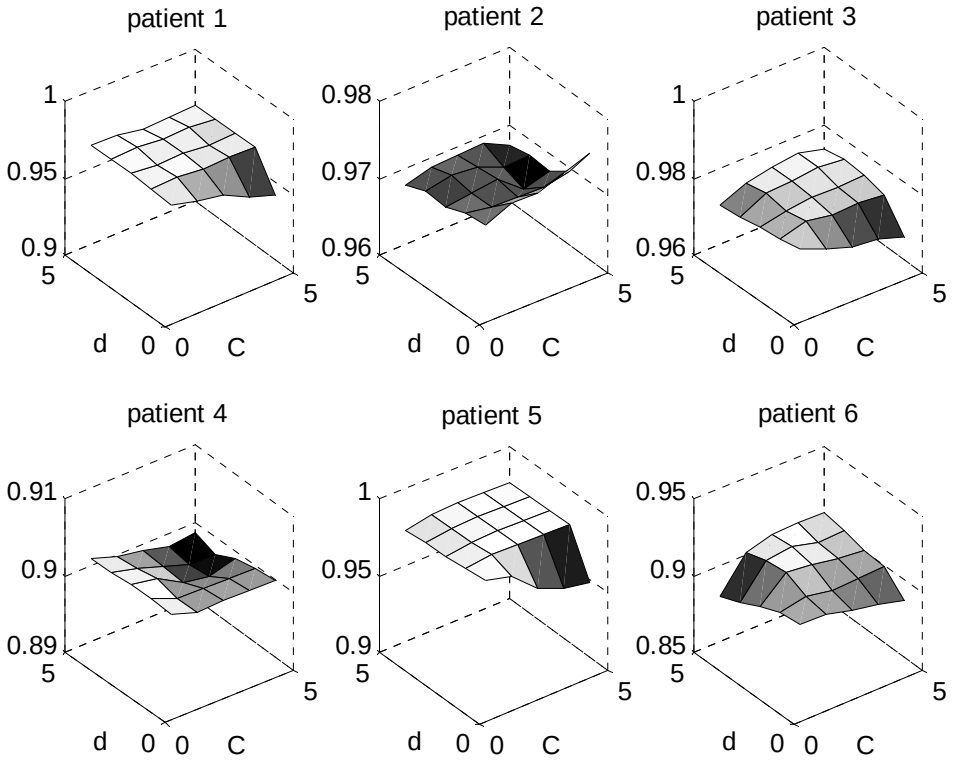


Figure 24: Example of AUC plots from a grid search for an optimal (C, d) combination. Note that the gradients in the subplots all point in different directions, meaning that increasing the performance for some patients would decrease the performance for other patients.

6.7 Hidden Markov models

The previously described classification methods treat each time instance (feature sample) as a separate event without regard to how the sequence of samples evolves in time. Since each activity state in the brain goes on for some time and not all transitions from one state to another are equally likely it makes sense to include this probability of transition in a model used for classification of these states. A hidden Markov model (HMM) [32] is a collection of a number of states. Each state is associated with probabilities of transition to each of the other states, and a probability density function that specifies the probability of emitting an observable signal. The model has also the attractive property that it can be trained in an unsupervised way, without using pre-classified training data.

The first order HMM is based on the first order Markov model (Figure 25), a random process characterized by that it has the Markov property

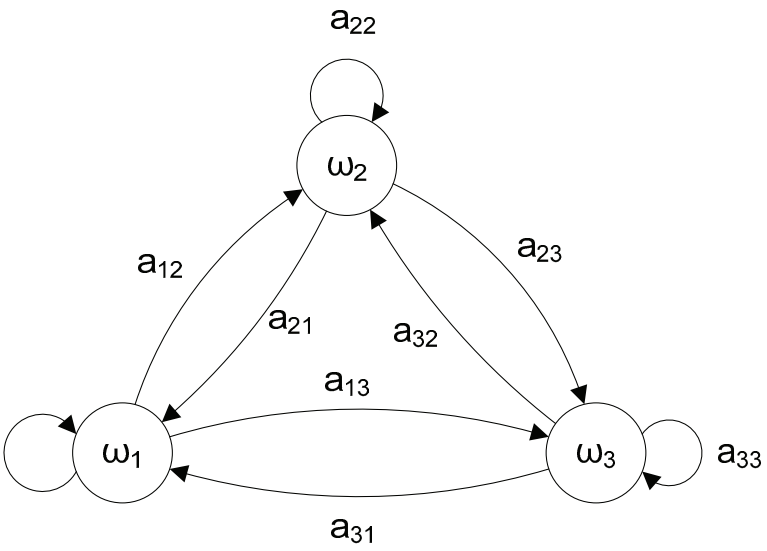


Figure 25: A three state Markov model. a_{xy} is the probability of transition from state ω_x to state ω_y.

[22]. This property states that the state of the model at time *n* is directly influenced by the state at *n-1* but not by any of the earlier states:

$$P(X_n=s \mid X_0=x_0, X_1=x_1, \ldots, X_{n-1}=x_{n-1}) = P(X_n=s \mid X_{n-1}=x_{n-1}) \text{ for all } n \geq 1$$

The symbol *n* is here used for discrete time. Even though time is continuous we can consider it to be discrete, especially when using features of sampled signals. If the features are calculated using a one-second window with 75 % overlap, we have a time resolution of 0.25 s, *i.e.* the process can only change state four times per second.

The Markov model is stochastic, meaning that the future state cannot be predicted in detail, only in probabilistic terms. Each state is associated with a number of transition probabilities that describes the probability that the process jumps to one of the other states in the next step.

The word “hidden” in HMM refers to the property that the state in a HMM cannot be directly observed. Instead it emits some symbol or signal that is visible to the observer. For example, HMM is popular in speech recognition, where each word is represented by a separate HMM. Each state then represents a phoneme (or some other sub-word element), and fitting a recorded sequence of phonemes to the appropriate model gives the classification. The emitted signals are incorporated in the model as a pdf, estimated as a histogram or a parametric model, of the emission probabilities.

The HMM theory can be divided into three main problems:

- The *evaluation* problem. For a HMM, complete with transition and emission probabilities: Determine the probability that a given sequence of observations was generated by the model.

- The *decoding* problem. For a HMM and a sequence of observations: Determine the most likely sequence of hidden states that generated the observations.
- The *learning* problem. If the structure of the model is known but not the transition- and emission probabilities: Estimate these parameters given a training set of observations.

The evaluation problem is important in *e.g.* speech recognition when a sequence of phonemes is to be compared with models of a number of words, but is not of interest here. What we are interested in is the decoding problem: we can measure EEG and calculate features from it, but we want to know what the underlying states were that made the brain produce these signals. The learning problem is important because in order to get a HMM that can be used for this, we must first train it so that it learns the appropriate parameters.

Training of the HMM can be done using the *Baum-Welch* or *forward-backward* algorithm that solves the learning problem. The decoding problem is solved using the *Viterbi* algorithm.

6.7.1 Baum-Welch algorithm

The Baum-Welch algorithm is a generalized expectation-maximization algorithm, and is used to estimate the parameters of the HMM. These parameters are the transition and emission probabilities, and are found based solely on the observed training data. The algorithm improves the parameters of the model by calculating the probability of the model producing the training data and iteratively updates the parameters until some stop criterion is met.

6.7.2 Viterbi algorithm

The Viterbi algorithm uses dynamic programming to find the sequence of hidden states that is most likely to have produced a sequence of observed events. The algorithm assumes that the state is a (time) sequence. It also assumes that the most likely sequence leading to the state at time n only depends on the observed event at time n and the most likely sequence at time $n-1$, as is the case in a first-order Markov model.

6.7.3 Initialization

Instead of just initializing the models randomly, the k-means clustering algorithm [26] was used to estimate the initial pdfs, with the number of clusters $k = 2$. For the initial “seeds” the extreme points in the data were chosen, with max or min set to burst and suppression using prior knowledge of the feature distributions. These seeds were used as initial cluster means. The algorithm then proceeds by comparing the distance between each point in the data and all the cluster means. The point is then

associated with the closest cluster, and the mean of that cluster is recomputed. The process is iterated until all points belong to a cluster and the cluster centres are stabilized.

This initialization was used to reduce the risk that the training ends up in a local error minimum instead of a global minimum; something that random initialization might lead to. Another way to reduce the risk of this is to run the algorithms a few times and choose the best instances, but that might be a weakness if implemented in a real monitoring system.

6.7.4 Using the hidden Markov model for BS segmentation

In paper IV, a two-state HMM was used for burst suppression segmentation. A number of features of the EEG were used as visible symbols, and the two states represented burst and suppression, respectively (Figure 26).

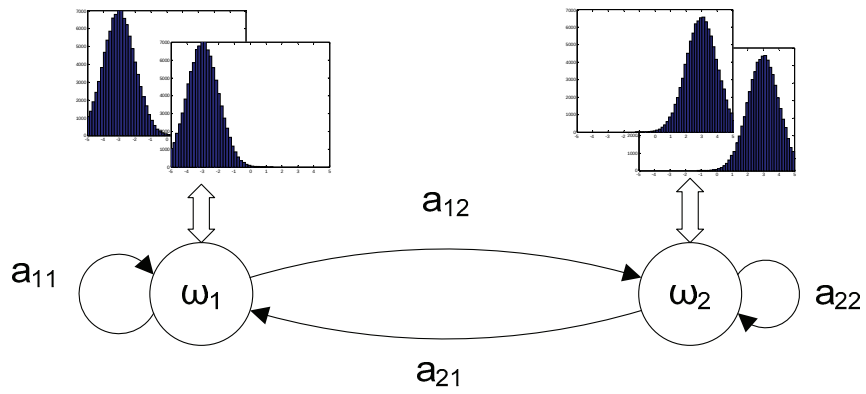


Figure 26: HMM with two states (burst and suppression). Each state is associated with an output vector that is observed, here illustrated by four distributions. The goal is to estimate the model parameters and distributions based on unmarked training data, and then estimate the sequence of states that is most likely to have generated an observed mixture.

The same set of five features which was used as in paper III was also used for SVM so that the two methods could be compared. To enable the HMM to use more than one feature the emission probabilities were multiplied together. This can be done if the different emission probabilities can be assumed to be independent. That is probably not true in the case investigated here, since the features all are generated by the same EEG signal, but the method was nevertheless found to work satisfactorily.

The results show that given identical input feature data, the unsupervised HMM classifier performs in principle just as well as the supervised SVM. In the present case, the only input parameter to the HMM is the number of states in the data, all other HMM parameters are estimated automatically by the algorithms. Because patients are individuals, an unsupervised learner that estimates a patient-specific model of EEG could be very

useful in a monitoring system. The method was considered for paper V, but because the computational complexity makes it infeasible to use together with feature selection using genetic algorithms the idea was dropped in favour of the simple FLD, and saved for future work.

6.8 Output signal smoothing

The output from the classification procedures was in some cases found to be noisy, and using knowledge of the typical length of a burst, a smoothing algorithm was applied that uses a five-point sliding window and checks if the majority of the samples in the window are burst detections. If they are, all samples in the window are converted to burst detections; if they are not, all samples in the window are converted to suppression. This has the result that groups of short detections are converted into longer continuous detections, while short isolated detections are removed. This increases the performance in some cases by reducing the number of false positives and negatives, and gives the detections more realistic lengths.

6.9 Patient sample equalization

The availability of the different types of activity differs between different patients, causing the chosen EEG epochs to vary in length. In paper V the risk of skewing the classification results because of this was reduced by equalizing the amount of data from the patients and classes by random sampling of the data. The median of the number of samples from the different patients was determined, and then samples were drawn from each patient that equalled the median. In the cases when the available number of samples was lower than the median, samples were drawn with replacement.

The data from the different patients were then concatenated into one matrix, and the sample order was randomized. The resulting matrix was divided into three parts; one half of the data was used for training, one quarter for testing and one quarter for validation. The training set was used for calculating the FLD projection matrix and the testing set for selecting the best one. When all selection was finished, the resulting classifier was tested on the validation set, which had not been involved in the process. If the result from the validation set is much worse than on the testing set, the classifier has probably been overfitted during training, and will not be able to generalize to unseen data.

Chapter 7

Feature selection

In Chapter 5, 22 features were introduced, which together with the four meta-feature functions resulted in a total of 88 meta-features that potentially could be used for classifying EEG signals. In chapter 6 some different classifiers were introduced, all of which can be used with several features in parallel. So why not just use every feature we can think of, and get the optimal classification? No, it is not quite that simple. All features cannot be appropriate for all classification problems. When segmenting a burst suppression signal we need features that enhance the difference between the activity in the bursts and the suppressions, while if we want to classify behavioral states we want features that enhance the differences between *e.g.* quiet sleep and active awake. These two sets of features would probably be different, since the BS activity is so different from the activity found in a healthy baby. If we add features that are not related to the classification we actually want to do, they can only contribute with noise that makes the classification more difficult.

Another reason for restricting the number of features is the *curse of dimensionality* described below.

7.1 Curse of dimensionality

The curse of dimensionality [33] describes how the number of dimensions, *i.e.* included features, is related to the number of examples needed to train the classifier in a satisfactory way. When the number of dimensions increases, the volume of the feature space increases and if the number of examples is constant they will be increasingly more spread out in this space. Consider *e.g.* the example of ten samples that are evenly distributed over a ten-point interval in a one-dimensional space, and thus cover 100 % of that space. If we then increase the number of dimensions to two, the number of discrete points in the space increase to 100, and our ten samples only cover 10 % of it. In three dimensions they only cover 1 %!

If, on the other hand, we use as many dimensions as we have training examples they could be spread out so that one example is present in each dimension. A classifier would for this example deteriorate into a look-up table and we would certainly be able to get a perfect score on the training data, but we would have learnt nothing about the actual distribution of the data and hence not be able to classify unseen data. Thus, using more features is not automatically better, the number of dimensions should always be much smaller than the number of training examples, and it is essential to remember to validate the classifier and the feature selection on unseen data as carefully as possible.

When choosing features to include in some sort of analysis the properties of the activity that is being analyzed would ideally be well known, and the feature functions would be designed to bring out these properties. In reality, however, these properties are not always known; they might even be what are ultimately sought for. When doing research on two types of activity we might know that they are different because the subject behaves differently during each type of activity, but we might not know how this is reflected in the EEG. We would therefore need some blind way for feature selection which could help us to find these differences.

In the present work, two main methods have been used for feature selection: exhaustive search and restricted search using genetic algorithms. Both are used as so called *wrapper* algorithms, *i.e.* the actual classifier is used for evaluation of the tested feature combinations. An alternative is the *filter* approach, where some other function is used for evaluation of the feature combinations. While being potentially quicker than the wrapper type, the filter has the drawback that because it does not test the actual classifier it is not guaranteed to return optimal solutions.

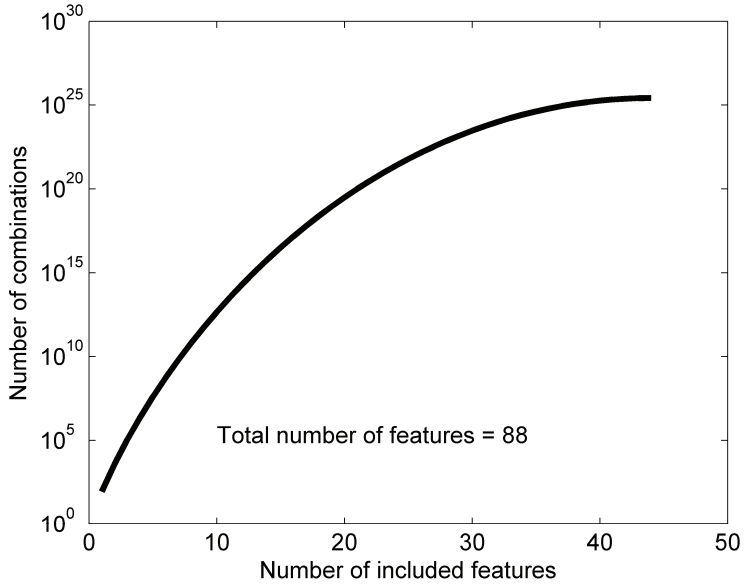


Figure 27: The number of feature combinations when selecting from a set of 88 features. Note that the y-axis is logarithmic.

7.2 Exhaustive search

The simplest and safest way to ensure that the optimum feature combination is selected is to make an exhaustive search. That means that every possible combination of features is evaluated by training and testing using the selected classifier. The method was used in paper III, but is only an alternative for small sets of features because with increasing numbers the computational complexity rises very quickly to impractical levels.

The number of combinations can be calculated using the binomial coefficient applying the following formula, where n is the total number of features to choose from, and k is the number of selected features:

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 1} = \frac{n!}{k!\,(n-k)!}, \qquad 0 \leq k \leq n$$

If we assume that the number of features n is 88 and calculate the binomial coefficient for $k \in \{1,2,\ldots,44\}$ we get the graph in Figure 27. Thus, for combinations of ten features chosen from a set of 88, our classifier has to be trained and evaluated more than 10^{12} times. If the classification problem that is evaluated in each step is not absolutely trivial, this approach will take too long.

7.3 Restricted search using genetic algorithms

Genetic algorithms (GA) [34] are inspired by the natural selection taking place in nature. Instead of methodically trying every possible combination without regard to performance, a random set of combinations is first tried,

and then the ones that seems most promising are combined in pairs to form new combinations. Random mutation is also allowed. The process is repeated until some stop condition is met, *e.g.* the error is below a threshold or the number of iterations reaches a pre-determined maximum. The reason for using genetic algorithms for feature selection is that it is faster than an exhaustive search when the number of features is large. Feature selection using genetic algorithms was used in paper V and VI.

When using genetic algorithms for feature selection, the set of features that are selected for evaluation is encoded in a chromosome, which can simply be a vector with ones representing selected features and zeros representing features that are not selected. When the genetic algorithm is initialized a number of chromosomes, or individuals, are randomly created. This population of possible solutions is then evolved for a number of generations. The basic evolution functions are combination of promising individuals and random mutation, *i.e.* some of the chromosome entries are flipped at random. In each generation, the fitness of the different individuals is evaluated using some fitness function, and then a number of the best ones are combined in pairs into offspring that are passed on to the next generation.

To reduce the probability that the search quickly stops in a local minimum a number of techniques are used. For example, some of the individuals with low fitness are kept to ensure that the search area is not reduced too soon. Random mutations are used to send individuals out into areas that may have been missed during the initialization. Some of the best individuals can be kept unchanged from one generation to the next one (so-called *elitism*), because combining two good individuals may not always produce an individual that is better than its parents.

In feature selection it is important to be able to restrict the number of selected features in the chromosome, because using too many features may reduce the performance and also be infeasible in a practical system where all these features must be extracted in real time. The number of features was limited using the method described in [35]: If the desired number of features is m and the number of features chosen by the GA is p, the following is done in each generation: If $p<m$, $(m-p)$ features are randomly added to the chromosome. If $p>m$, $(p-m)$ features are randomly removed from the chromosome.

Because the result of the GA may depend on the randomly initiated starting set, the algorithm was run five times for each number of included features, and the one with the lowest error on the test set was included in the evaluation. The observed differences in probability of error between different runs were less than one percent.

Chapter 8

Conclusions and future work

The presented methods for separating BS signals from activity recorded from healthy neonates and the BS segmentation, all work well. So far, we have developed methods for segmentation and classification of mainly intermittent activity, and we have made a preliminary test by calculating burst suppression ratios and mean suppression lengths on a long recording (paper V). A natural next step is to continue to test the methods on long recordings that are as close to the clinical reality as possible, and to find properties of the segmented BS signals that have clinical relevance. The next project is therefore to apply the methods on a database containing long recordings from a number of post-asphyctic newborns. These recordings are many hours long (as the example in paper V); the BS signals used so far are merely short typical examples of this data. One of the goals of the project would be to find parameters that can predict the outcome of the hypoxic events, much like was done in a piglet model [1], but this time improving the analysis by analysing the activity in the burst and suppressions separately.

The functionality developed here should also be implemented in the SACS® system [4]. This would allow its testing by the clinicians working with this kind of patients, both at the NICU and at the neurophysiologic labs. The SACS system is developed using a modular approach that makes it very flexible. Implementing a classifier as a SACS module makes it possible for the user to combine it in any way he likes with the other modules and easily set up customized applications within the system.

At the NICU the new EEG based functionality should be tested as an alternative or complement to the aEEG that currently is used for monitoring of some infants. At the lab it could be used as a tool that simplifies the work of the neurophysiologist by highlighting areas of interest that should be further studied. The quantification of different pathologic aspects of the EEG might also add to the diagnostic value of the findings presented in this study. So far, the segmentation methods have been tested on a database containing six patients and the classification has been tested on the same six patients and recordings from 20 healthy babies. To make sure that the methods work on a broad range of subjects they should be verified in a much larger trial with more patients and longer recordings.

An important class of pathological activity that can be found in the EEG of newborn babies is seizures. Seizure detection should also be included in the system, for instance one of the methods that have been developed elsewhere. This is an important part of the other main goal of the project; to develop methods that give information of ongoing pathological processes. This would enable the clinician to introduce therapy to prevent the negative effects of this activity, thereby probably improving the prognosis for the neonate.

Some ideas have popped up during the research process that have not yet been tested. For example, a method that has potential to become a useful tool in this field is using the hidden Markov models that were tested in paper IV. The method proved to work well for segmenting BS signals, without supervised training, and produces patient specific models that could be of great use in other applications. The method was initially tested on the classification problem, because the models were thought to be useful for discrimination between the two intermittent classes BS and *tracé alternant*. However, since the problem could be solved using simpler classifiers, and because the HMM is too computationally demanding to be used with the feature selection using genetic algorithms, this idea was discarded.

A possible application for the HMM could be to use it for classification of sleep states. This problem was touched upon in paper V, but satisfying results were only reached for the quiet sleep class (*tracé alternant*). Possibly, features derived from signals other than EEG could be included

in the classification process. Easily accessible examples are heart rate variability calculated from the heart rate signal that usually is present in the existing monitoring equipment, or some movement sensor added to the system.

References

[1] N. Löfgren, "The EEG of the Newborn Brain - Detection of Hypoxia and Prediction of Outcome," in *Department of Signals and Systems*. vol. Doctor of Philosophy Göteborg: Chalmers University of Technology, 2005.

[2] N. Löfgren, "Remote sessions and frequency analysis for improved insight into cerebral function during paediatric and neonatal intensive care," in *Department of Signals and Systems*. vol. Licentiate of Engineering Göteborg: Chalmers University of Technology, 2002.

[3] I. Rosén, "The Physiological Basis for Continuous Electroencephalogram Monitoring in the Neonate," *Clin Perinatol*, vol. 33, pp. 593-611, 2006.

[4] K. Lindecrantz, R. Bågenholm, F. Göthe, A. Hedström, N. Löfgren, S. Nivall, and J. Ouchterlony, "A general system used in monitoring of cerebral and circulatory function in neonatal intensive care," *Med Biol Eng Comput*, vol. 37, pp. 888-9, 1999.

[5] C. C. Menache, B. F. D. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology*, vol. 27, pp. 93-101, 2002.

[6] E. P. Widmaier, H. Raff, and K. T. Strang, *Vander's Human Physiology*, 10th ed. New York: McGraw-Hill, 2006.

[7] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain*, 2nd ed. Baltimore: Lippincott Williams & Wilkins, 2001.

[8] H. H. Jasper, "The Ten-Twenty Electrode System of the International Federation," in *International federation of societies for electroenchephalography and clinical neurophysiology*, Brussels, 1957, pp. 371-375.

[9] J. Malmivuo and R. Plonsey, *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*, Web version ed. New York: Oxford University Press, 1995.

[10] M. Thordstein, N. Löfgren, A. Flisberg, R. Bågenholm, K. Lindecrantz, and I. Kjellmer, "Infraslow EEG activity in burst periods from post asphyctic full term neonates," *Clinical Neurophysiology*, vol. 116, pp. 1501-1506, 2005.

[11] G. Holmes, J. Rowe, J. Hafford, R. Schmidt, M. Testa, and A. Zimmerman, "Prognostic value of the electroencephalogram in neonatal asphyxia," *Electroencephalography and Clinical Neurophysiology*, vol. 53, pp. 60-72, 1982.

[12] D. B. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. M. T. Robertson, "EEG and long-term outcome of term infants with

neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology,* vol. 110, pp. 655-659, 1999.

[13] M. Thordstein, B. G. Wallin, A. Flisberg, R. Bågenholm, I. Kjellmer, N. Löfgren, and K. Lindecrantz, "Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates," *Clinical Neurophysiology,* vol. 115, pp. 2461-2466, 2004.

[14] G. L. Holmes and C. T. Lombroso, "Prognostic value of background patterns in the neonatal EEG," *Journal of Clinical Neurophysiology,* vol. 10, pp. 323-352, 1993.

[15] H. F. R. Prechtl, "The behavioural states of the newborn infant (a review)," *Brain Research,* vol. 76, pp. 185-212, 1974.

[16] H. Gray, *Anatomy of the Human Body,* 20 ed. Philadelphia: Lea & Febiger, 1918.

[17] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: John Wiley & Sons, Inc., 1996.

[18] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology,* vol. 89, pp. 980-1002, 1998.

[19] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *Journal of Clinical Monitoring and Computing,* vol. 15, pp. 357-367, 1999.

[20] J. A. Rice, *Mathematical Statistics and Data Analysis,* 2nd ed. Belmont: Duxbury Press, 1995.

[21] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* vol. 27, pp. 379-423, 623-656, 1948.

[22] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford: Oxford University Press, 2005.

[23] G. Peeters, "A large set of audio features for sound description," Ircam, Paris2004 04 23 2004.

[24] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing,* vol. 13, pp. 1035-1046, 2005.

[25] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing,* vol. 10, pp. 293-302, 2002.

[26] R. O. Duda, Hart, P. E., Stork, D. G., *Pattern Classification,* 2nd ed. New York: John Wiley & Sons, Inc., 2001.

[27] C. J. Huberty, *Applied discriminant analysis*. New York: John Wiley & Sons, 1994.

[28] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall, 1993.

[29] G. Strang, *Introduction to Linear Algebra*, 3rd ed. Wellesley MA: Wellesley-Cambridge Press, 2005.
[30] S. Haykin, *Neural networks*, 1 ed. New York: Macmillan College Publishing Company, 1994.
[31] F. Rosenblatt, "The Perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65 p. 23, 1958.
[32] L. R. Rabiner, "Tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, pp. 257-286, 1989.
[33] R. Bellman, *Adaptive control processes: A guided tour*. Princeton, New Jersey: Princeton University Press, 1961.
[34] W. Siedlecki and J. Sklansky, "A note on genetic algorithms for large-scale feature selection," *Pattern Recognition Letters*, vol. 10, pp. 335-347, 1989.
[35] S. Salcedo-Sanz, G. Camps-Valls, F. Pérez-Cruz, J. Sepúlveda-Sanchis, and C. Bousono-Calzón, "Enhancing genetic feature selection through restricted search and Walsh analysis," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 34, pp. 398-406, 2004.

Part II

Appended papers

Summary of papers

This section summarizes the most important points of the six papers included in the thesis. In short, paper I-IV develops the methodology for burst suppression segmentation, and different classifiers and features are evaluated. Paper V is focussed on classification of background activity, where the time scale is in minutes instead of seconds as in the previous papers. The burst suppression segmentation is included as a tool for automatic analysis of the parts of the background activity that is classified as BS and is tested on a very long recording.

In paper VI the developed methods are applied on a practical research problem: intermittent tracé discontinue signals are automatically segmented in order to evaluate the impact on the brain activity in preterm babies by a certain medication.

Paper I: Detection of bursts in the EEG of post asphyctic newborns

In this paper eight features were evaluated with regard to their ability to separate bursts from suppression in neonatal EEG. For the evaluation an artificial neural network (ANN) implemented using the neural network toolbox in Matlab was used. With the best feature as a starting point, additional features were added one by one. Sample-wise sensitivity, specificity and their sum were used as performance measures.

It was found that different features were best for different patients, making an optimal selection difficult. For example, simply using the variance gave usable result for all but one patient. Adding more features increased the sensitivity for this patient up to a point where the sensitivity for another patient started to decrease. It was concluded that additional features and classification techniques should be tested.

Paper II: Classifying burst and suppression in the EEG of post asphyctic newborns using a support vector machine

In this paper a support vector machine (SVM) with a radial basis function (RBF) kernel was used for segmentation of burst-suppression signals. The data set, as well as the features, were slightly changed as compared to the previous paper, so the results are not directly comparable. A fixed set of five features was used.

It was found that the SVM had the advantage over the ANN that it gives a classifier with a maximum margin, that the error on the training set can be controlled during training and that it is deterministic, and eliminates the problem of risking ending up in a local error minimum. It was also found that the sample-wise sensitivity and specificity measures are somewhat misleading, and should be complemented with a hit/miss measure that places more emphasis on detection of whole events instead of just counting the samples.

Paper III: Classification of burst and suppression in the neonatal electroencephalogram

In this paper, the three classifiers Fisher´s linear discriminant (FLD), an artificial neural network and a support vector machine were compared with regard to their performance when separating burst from suppression. These classifiers all use supervised learning. For performance measurements, sample-wise sensitivity and specificity were used to create receiver operating characteristic (ROC) curves for the classifiers that reveal their full potential.

The same five features were used as in paper II, but now the optimal subset was chosen by exhaustive search. It was found that the performance for this set of features started to decrease after two features. The SVM performed best, and could handle higher numbers of features best, but the ANN and FLD performed almost as well.

Paper IV: Comparing a supervised and an unsupervised classification method for burst detection in neonatal EEG

In this paper a Hidden Markov Model (HMM) and a Support Vector Machine (SVM) using unsupervised and supervised training, respectively, were compared with respect to the same problem as in the previous papers. The study showed that the SVM and the HMM exhibit similar performance, despite their fundamental differences.

It was concluded that HMMs could be used for automatic generation of patient-specific models of intermittent EEG that in turn could be used for classifying entire signals as well, by using the model´s parameters as features.

Paper V: Automatic classification of background EEG activity in healthy and sick neonates

In this paper the emphasis of the analysis was moved from segmentation to classification. Instead of segmenting signals that are known to contain exclusively pathological burst suppression the analysis starts with deciding what type of EEG it is dealing with. Along with the BS EEG from the same six patients as in the previous papers, signals from twenty healthy babies were included, divided into four behavioural states. The states were active awake, quiet awake, active sleep and quiet sleep, and the signals were divided into these states by a combination of observation of the baby during recording and by off-line visual examination of the EEG. A FLD was used for deciding what type the EEG belonged to, using a feature subset chosen from a larger set using a genetic algorithm. When BS activity was detected the signal was segmented using the previously developed methods. The methods were tested on long recordings as well, a two-hour recording from one of the healthy babies and a 32-hour recording from one of the sick patients.

We managed to achieve 100% correct classification when separating burst suppression EEG from all four healthy EEG types and 93% true positive classification when separating quiet sleep from the other types. The other three states could not be classified.

Paper VI: Does indomethacin for closure of patent ductus arteriosus affect the cerebral function?

In this paper the methods developed in the previous papers were used in a practical application. The purpose of our investigation was to see if the drug indomethacin, used in conventional dose for closure of patent *ductus arteriosus* (a shunt allowing most of the blood to bypass a fetus' lungs), affects brain function. This was represented by EEG evaluated by automatically estimated quantitative measures.

The impact on the brain function was estimated by measuring the length of all low activity periods (LAP, similar to the suppressions in the previous papers). These lengths were automatically measured using two methods, one using fixed thresholds and one using Fisher´s linear discriminant together with a genetic algorithm for feature selection.

Neither of the two methods identified any change of the amount of LAPs in the EEG after as compared to before the indomethacin infusion. It was concluded that indomethacin in conventional dose for closure of patent ductus arteriosus does not affect the brain function as evaluated by quantitative EEG.

Paper I

Detection of bursts in the EEG of post asphyctic newborns

J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Detection of bursts in the EEG of post asphyctic newborns

J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Abstract

Eight features inherent in the electro-encephalogram (EEG) have been extracted and evaluated with respect to their ability to distinguish bursts from suppression in burst-suppression EEG. The study is based on EEG from six full term infants who had suffered from lack of oxygen during birth. The features were used as input in a neural network, which was trained on reference data segmented by an experienced electroencephalographer. The performance was then evaluated on validation data for each feature separately and in combinations. The results show that there are significant variations in the type of activity found in burst-suppression EEG from different subjects, and that while one or a few features seem to be sufficient for most patients in this group, some cases require specific combinations of features for good detection to be possible.

1. Introduction

The burst-suppression (BS) pattern is one of several indicators of severe pathology in the electroencephalogram (EEG) signal that may occur after brain damage, caused by e.g. asphyxia [1, 2]. Certain characteristics of this pattern can provide the clinicians with important information about the recovery of the patient and it is thus important in the adjustment of the treatment. Examples of important characteristics of the BS pattern are the length of the burst and suppression intervals, and the spectral content of the bursts [3, 4].

In practice, most EEG recordings are evaluated by experienced neurophysiologists through visual inspection of the unprocessed signal [5]. Some attempts towards automatic detection have been made, most of them targeting burst-suppression induced by anesthesia [6, 7].

Our goal is to develop tools that can be used for automatic classification of BS caused by various causes, for example perinatal asphyxia (insufficient oxygen/dioxide ventilation around the time of birth). The target patient group is newborns, and a possible future application is a monitoring system for neonates in intensive care.

This paper is focused on evaluating a number of features inherent in the EEG signal, with respect to their ability to distinguish between burst and suppression activity in BS EEG. These features are then used as inputs to a neural network for the actual classification.

2. Methods

This study was performed on EEG from six full term infants having suffered from perinatal asphyxia. Each subject contributed with a continuous recording of 6-40 minutes, selected and classified by an experienced electroencephalographer (MT). The length was chosen to include at least 10 bursts. Eight channels were used, and the data was digitized at a sampling rate of 200 Hz. The mean of the EEG signal was subtracted, the signal was band pass filtered 0.5 to 20 Hz, and notch filtered at 50 Hz to reduce power line interference, before feature generation. The reference data had a time resolution of one second.

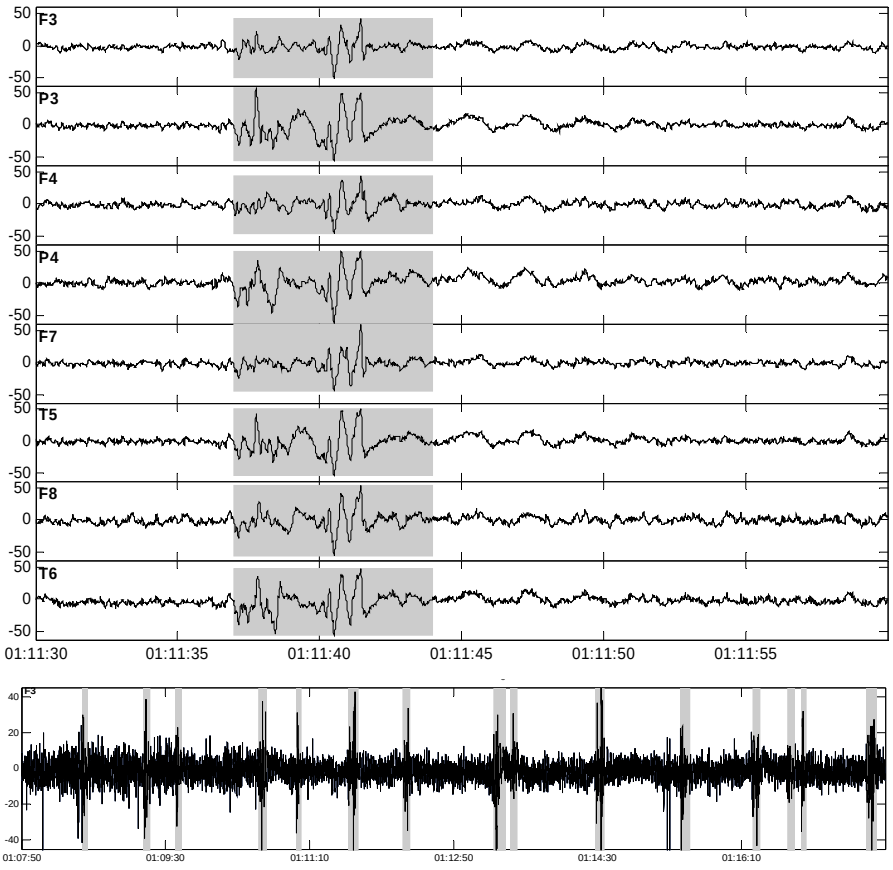


Fig. 1. The top plot is an example of 30 seconds of burst-suppression EEG. The bottom plot shows ten minutes of the same signal, but only one channel. The shaded areas are bursts identified by an electroencephalographer.

A selection of the recorded burst-suppression EEG is displayed in Fig. 1. Eight features, most of which have been used for detection of BS in adult subjects under anesthesia [7, 8], were selected (Table I). The features

were calculated for overlapping segments of the raw signal. The segment length was one second, motivated by the fact that bursts were considered to be at least one second long. The overlap was 0.75 s, producing an output feature signal with an effective sampling rate of 4 samples per second.

Feature	Description
Spectral Edge Frequency (SEF95) [6]	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on rectangular windows of the signal
3 Hz power	Power in a one Hz wide band centered at three Hz
Median [9]	Median absolute value
Variance (s2(x)) [9]	$$s^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x-\mu)^2$$ Where x is a time series, μ is the sample mean of x
Skewness (E3(x)) [9]	$$E^3(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\mu)^3}{\sigma^3}$$ σ is the standard deviation of x. x and μ as above
Kurtosis (E4(x)) [9]	$$E^4(x) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x-\mu)^4}{\sigma^4}$$ x, μ and σ as above
Zero crossings [6]	Rate of zero crossings
Shannon entropy (HSh) [10]	$$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$$ p(I1)… p(IU), pU is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals I1,..,IU. 20 intervals were used evenly distributed between the maximum and minimum value of the signal in the window. HSh is a measure of uncertainty of a random variable.

Table I: Features selected for evaluation of their ability to distinguish bursts from suppression in burst-suppression EEG.

After feature generation, the feature signals were combined per feature by taking the median over the eight channels. This was motivated by the fact that BS is considered to be a global phenomenon appearing in all EEG channels, and using the median over the channels removes disturbances present on one or a few of the channels. This also reduces the computational complexity as compared to feeding the network with all channels.

The feature signals were used as inputs to a feed-forward neural network with error back-propagation training [11], which was implemented using the neural network toolbox in Matlab. The training of the network was performed using the segmented reference data. The feature signals were fed to the network one at a time, and then in combinations. The number of input nodes was set by the number of features used in each experiment. Ten hidden nodes were used, chosen by empirical testing, and one output node. Ten instances of the network were trained for each feature, and the best one was chosen for the evaluation. This technique was used to reduce the risk of using a network that failed to converge when trained. The training of neural networks uses a randomly initiated starting network, which then is trained using randomly selected samples from the training set. The training algorithms move in a complex landscape, trying to find a minimum error state. However, this landscape usually contains multiple local minima, and to reduce the risk of "getting stuck" in a local minimum, a number of separate instances of the network are trained, and then the instance with the lowest classification error is used.

In order to maximize the use of the limited number of patients, the data was not divided into fixed training- and testing sets. Instead, leave-one-out cross-validation was used, meaning that the data from one patient was used for testing, and the other five were used for training. This process was then repeated for all patients. In order to give each patient an equal chance to influence the network training, the shorter records were repeated so that they got the same length as the longest one. This equalizes the amount of data from each patient, but it does not take into account the number of bursts provided by the patients.

An output sample from a neural network depends only on one input feature vector, and does not take previous or subsequent samples into consideration. The detection signal is therefore noisy, with a lot of short detections. Using knowledge of the typical length of a burst, an iterative smoothing algorithm was designed that converts groups of short detections into longer continuous detections, while short isolated detections are removed. This was found to increase the performance considerably.

For performance evaluation, sample wise sensitivity and specificity were used as measures. Sensitivity was defined as the percentage of the burst

samples that were correctly classified as bursts, and specificity as the percentage of the remaining samples that were not classified as bursts. As a measure of performance for choosing which networks to keep when training multiple instances, the sum of the sensitivity and specificity was used.

The features were ordered with the best single feature as a starting point. Initial experimental results showed that most patients were not affected very much by adding features, with the exception of patient six. Therefore, features were added in an order that increased the performance for patient six, while trying not to degrade the performance for the other ones.

3. Results

In Fig. 2 the eight feature signals are compared. The mean and standard deviation of the feature signals have been calculated for the burst and suppression parts separately. The feature signals have been normalized by subtracting the mean and dividing by the standard deviation to simplify comparison. The figure shows that although the estimated means for burst and suppression of most feature signals differ, their standard deviations overlap considerably.

The results in Fig. 3 show that there are large differences between the patients with regard to how well the features work in terms of sensitivity. Using these results, the best feature was used as a starting point for the feature order for the test shown in Fig. 4. Here, both individual and total sensitivity and specificity are shown for increasing numbers of features.

Fig. 5 demonstrates the difference between sample wise sensitivity and burst wise sensitivity. Although the sample wise sensitivity in this case is merely 50%, all but two bursts have been detected.

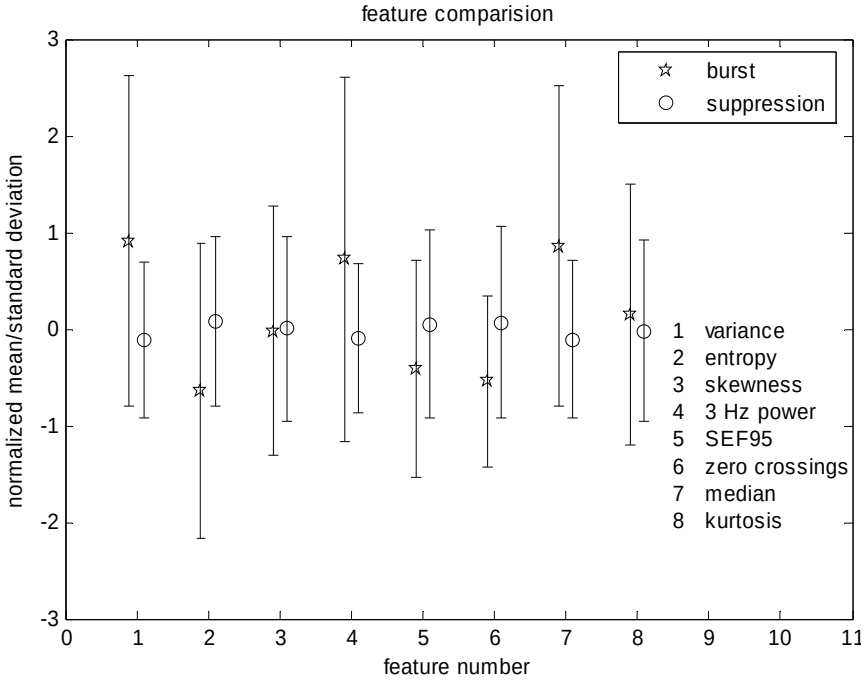


Fig. 2. The triangles and stars represent the mean of the feature signals for suppression and burst periods respectively. The lines show the standard deviations around the means. Data from all six patients are included.

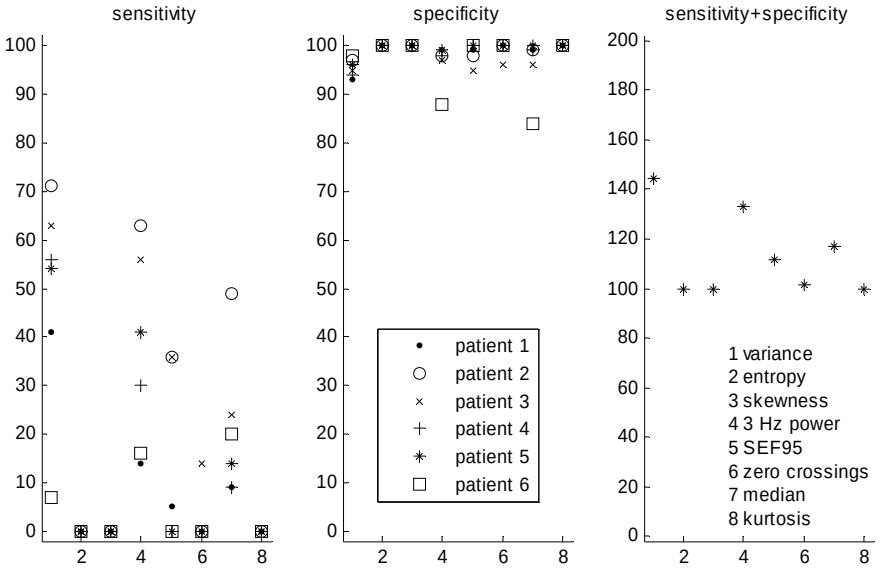


Fig. 3. Sensitivity and specificity achieved when the feature signals have been run through a neural network, one at a time. In each case, five networks have been trained, and the best one has been chosen. The sensitivity is in many cases very low, but this is partly due to the way the sensitivity is calculated. See Fig. 5 for comparison.

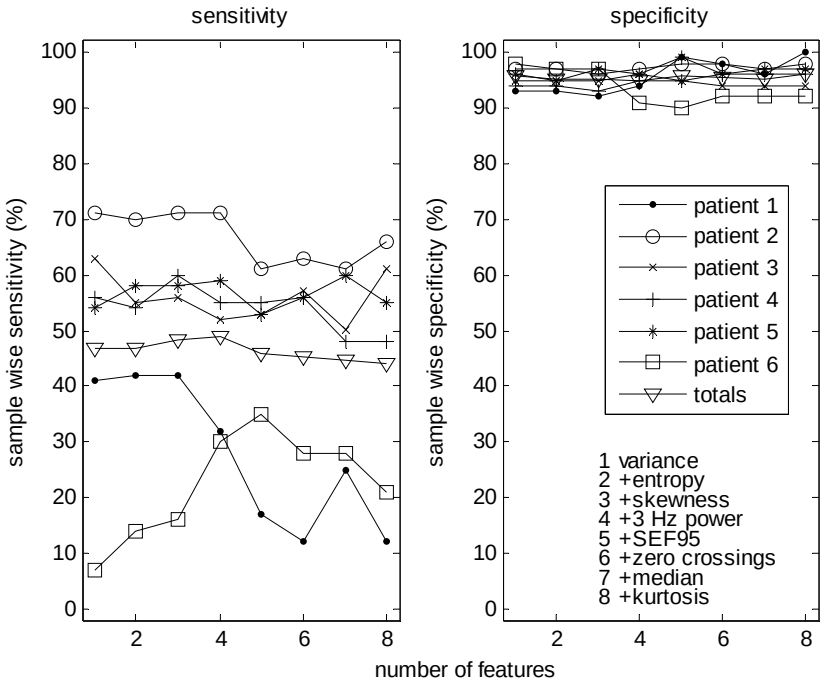


Fig. 4. Sensitivity and specificity after the feature signals were run through a neural network. The curves show the performance for the six patients for increasing numbers of features. The curve with the triangles shows the total performance for the six patients. In each case, ten networks were trained, and the best one with regard to the training set was chosen. The first point along the x-axis shows the result when feeding the network with the best feature, the variance. The subsequent points show the results when more features are added one by one until all eight features are included.

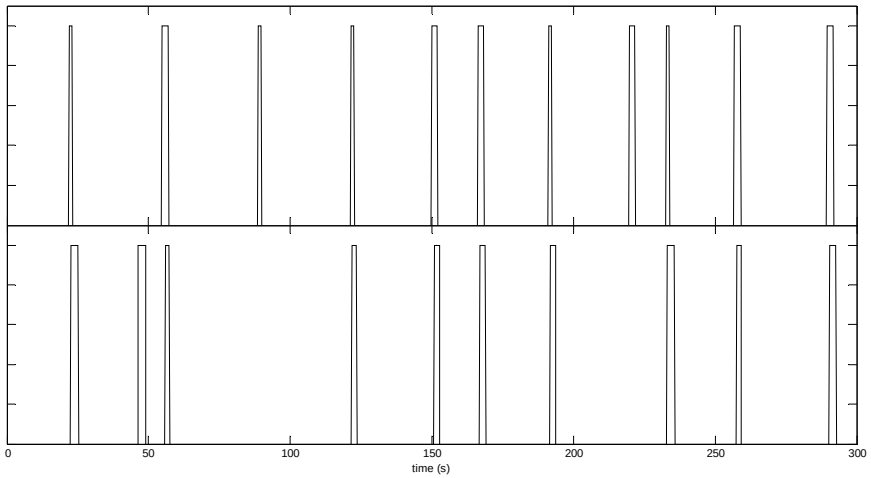


Fig. 5. Example of detection using four features on patient 5. The top plot shows the reference data, the plot below shows the detections. It can be seen that 9 out of 11 bursts in the reference data have been detected and that there was one false detection.

Patient Number	Burst Percentage
1	24 %
2	13 %
3	6 %
4	27 %
5	26 %
6	4 %

Table II: Patient BS ratio.

Table II shows the burst contents in the signals from the different subjects. Note that the data from patient six, who has the worst detection performance, has a significantly lower burst content that most of the others.

4. Discussion

In all calculations of this study, all eight available channels have been used. However, since bursts are mainly global phenomena, it should in principle be possible to detect them using only one channel. The drawback of using fewer channels is that using less data makes the system more sensitive to disturbances affecting only some of the channels, like bad electrode-skin coupling and artifacts. As an alternative to using the median of the input feature signal, one network could be trained for each channel, and then the final decision could be made by counting the number of detections from the different channels. This would however increase the computational complexity for the network training.

The method to calculate sensitivity and specificity used here is rather strict; every missed sample lowers the measured performance. The one-second time resolution used in the reference data probably contribute to a lowering of the calculated performance. Alternatively, the number of bursts detected divided by the total number of bursts could be used to define sensitivity, which would lead to higher scores. However, this would not at all take into consideration the length of the bursts, or the accuracy of the detection of the start- and endpoints.

The specificity for many of the features is very low (Fig. 3), but they still contribute to a higher sensitivity for patient six when used in combination with the variance (Fig. 4). However, adding the features in descending sensitivity order does not necessarily improve the result. An explanation for this may be that the best features share a lot of common information,

and that combining them does not add useful information for the burst detection.

For most cases, the specificity is higher than the sensitivity. This is probably due to the fact that the burst periods are generally much shorter than the suppression periods, which means that the network has more examples of suppression to learn from than it has examples of bursts. The detections are often short, meaning that even a large number of false detections contribute little to a worsening of the specificity.

Given the small amount of data available, not much can be said about the actual performance of the neural networks. For most patients in the data set the detection works well when using only one feature, but for patient six the performance is very low when using one feature, and increases drastically when adding more features. At the same time, the performance for patient one decreases, while all other remain almost constant, even though the EEG from all patients are similar with respect to presence of burst-suppression patterns. Apparently there are characteristics of BS as detected by an experienced neurophysiologist that are not captured by the features selected here, no matter what combination used. In order to make the detection work equally well for all patients, other features or different classification techniques are needed.

In order to produce a complete burst-suppression detection- and quantization system, artifact rejection need to be included. It could also be useful to have a pre-classifier to first decide if the data is an example of burst-suppression or another type of periodic EEG activity, before the segmentation is made.

When more data is obtained, it will be possible to do a more statistically reliable training and evaluation of the neural network. The results of this will hopefully provide an easily used and reliable tool for automatic segmentation of burst-suppression EEG in newborns that has suffered from asphyxia. This tool may be a part of a monitoring system for neonatal intensive care units, as well as a tool for segmenting burst-suppression data used for off-line analysis.

5. References

[1] G. Holmes, J. Rowe, J. Hafford, R. Schmidt, M. Testa, and A. Zimmerman, "Prognostic value of the electroencephalogram in neonatal asphyxia," *Electroencephalography and Clinical Neurophysiology,* vol. 53, pp. 60-72, 1982.

[2] D. B. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. M. T. Robertson, "EEG and long-term outcome of term infants with neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology,* vol. 110, pp. 655-659, 1999.

[3] C. C. Menache, B. F. D. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology*, vol. 27, pp. 93-101, 2002.

[4] M. Thordstein, B. G. Wallin, A. Flisberg, R. Bågenholm, I. Kjellmer, N. Löfgren, and K. Lindecrantz, "Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates," *Clinical Neurophysiology*, vol. 115, pp. 2461-2466, 2004.

[5] M. D. Lamblin, M. André, M. J. Challamel, L. Curzi-Dascalova, A. M. D'Allest, E. De Giovanni, F. Moussalli-Salefranque, Y. Navelet, P. Plouin, M. F. Radvanyi-Bouvet, D. Samson-Dollfus, and M. F. Vecchierini-Blineau, "EEG in premature and full-term newborns. Maturation and glossary," *Neurophysiologie Clinique*, vol. 29, pp. 123-219, 1999.

[6] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology*, vol. 89, pp. 980-1002, 1998.

[7] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *Journal of Clinical Monitoring and Computing*, vol. 15, pp. 357-367, 1999.

[8] C. Jeleazcov, S. Egner, F. Bremer, and H. Schwilden, "Automated EEG preprocessing during anaesthesia: New aspects using artificial neural networks," *Biomedizinische Technik*, vol. 49, pp. 125-131, 2004.

[9] J. A. Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. Belmont: Duxbury Press, 1995.

[10] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, 1948.

[11] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation*. Redwood City, CA: Addison-Wesley Publishing Company, 1991.

Paper II

Classifying burst and suppression in the EEG of post asphyctic newborns using a support vector machine

J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Classifying burst and suppression in the EEG of post asphyctic newborns using a support vector machine

J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Abstract

A Support Vector Machine (SVM) was trained to distinguish bursts from suppression in burst-suppression EEG, using five features inherent in the electro-encephalogram (EEG) as input. The study was based on data from six full term infants who had suffered from perinatal asphyxia, and the machine was trained with reference classifications made by an experienced electroencephalographer. The results show that the method may be useful, but that differences between patients in the data set makes optimization of the system difficult.

1. Introduction

The burst-suppression (BS) pattern is one of several indicators of severe pathology in the electroencephalogram (EEG) signal that may occur after brain damage, caused by e.g. asphyxia (insufficient gas and nutrient supply around the time of birth) [1, 2]. Certain characteristics of this pattern can provide clinicians with important information about the prognosis of the patient and is thus important in the adjustment of the treatment. Examples of important characteristics of the BS pattern are the length of the burst and suppression intervals, the percentage of suppression activity in a recording, and the spectral content of the bursts [3-5].

In practice, most EEG recordings are evaluated by experienced neurophysiologists through visual inspection of the unprocessed signal [6]. Some attempts towards automatic detection have been made, most of them targeting burst-suppression induced by anesthesia [7, 8].

Our goal is to develop tools that can be used for automatic classification of BS induced by various causes, for example perinatal asphyxia. The target patient group is newborns, and a possible future application is a monitoring system for neonates under intensive care.

This paper focuses on using a set of features derived from the EEG as input data to a Support Vector Machine (SVM) [9], and training it to classify bursts and suppressions in BS EEG. Previously a similar approach

was made using a Neural Network [10], with varying results, and it seemed as further work was needed. The present work is based on partly the same data set: One patient from the old set was excluded because the burst activity was judged by the electroencephalographer to be “much less distinct” than the one seen in the others, and therefore not suitable for use with an algorithm that require training. Another patient has been added, thus producing a set of data from six patients.

2. Method

This study was performed on EEGs from six full term infants having suffered from perinatal asphyxia. Each subject contributed with a continuous recording of 6-40 minutes, selected and classified by an experienced electroencephalographer (MT). The length was chosen to include at least 10 bursts, and all artifacts were manually identified and removed from the data. A common average montage with eight channels was used, with electrodes placed according to the international 10-20 system at the positions F7, F3, T5, P3, F8, F4, P4 and T6, and the data was digitized at a sampling rate of 200 Hz. A straight line was fitted to each signal and subtracted to remove any trends, the signal was band pass filtered between 0.5 and 20 Hz, and notch filtered at 50 Hz to reduce power line interference, before feature generation. In one case, a LMS adaptive filter [11] with a separate ECG channel as reference was used to suppress ECG interference in the EEG signal. The manually classified BS reference data had a time resolution of one second. When the start or end of a burst did not coincide with even seconds, the time was rounded off so that the entire burst was included. This meant that in many cases, up to two seconds of suppression activity was included with every burst.

An example of the recorded burst-suppression EEG is displayed in Fig. 1. Five features, some of which have been used for classification of BS in adult subjects under anesthesia [8, 12] and that have shown promising results earlier [10], were selected (Table I). The features were extracted from overlapping segments of the EEG signal. The segment length was one second, motivated by the fact that bursts were considered to be at least one second long. The overlap was 0.75 s, producing an output feature signal with an effective sampling rate of 4 samples per second. All feature signals except variance were smoothed by convolution with a triangular window. Finally the feature signals were normalized by fitting their distributions between the first and the 99th quartile into the interval zero to one, thus giving all features the same amplitude range.

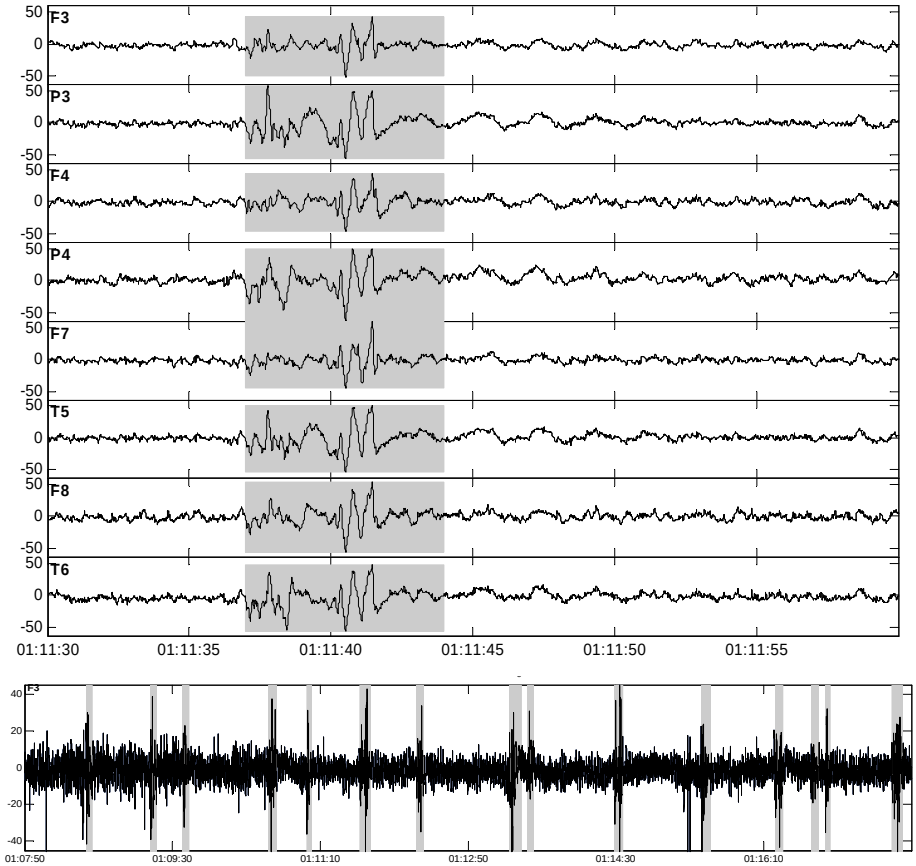


Fig. 1. The top plot is an example of 30 seconds of burst-suppression EEG. The bottom plot shows ten minutes of the same signal, but only one channel. The shaded areas are bursts manually classified by an electroencephalographer.

Feature	Description
Spectral Edge Frequency (SEF95) [7]	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on Blackman-Harris windows of the signal. Output signal smoothed by a six-point triangular window.
3 Hz power	Power in a one Hz wide band centered at three Hz. Output signal smoothed by a six-point triangular window.
Median [13]	Median absolute value. Output signal smoothed by a six-point triangular window.
Variance ($s^2(x)$) [13]	$s^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x-\mu)^2$ x is a time series, μ is the sample mean of x.

Shannon entropy (H_{Sh}) [14]	$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$	$p(I_1)\ldots p(I_U)$, p_U is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals $I_1,..,I_U$. 20 intervals were used evenly distributed between +/- two standard deviations of the signal in the window. H_{Sh} is a measure of uncertainty of a random variable. Output signal smoothed by a six-point triangular window.

Table I: Features used for classifying burst and suppression in bs EEG

After feature generation, the feature signals were combined per feature by taking the median over the eight channels. This was motivated by the fact that BS is considered to be a global phenomenon appearing in all EEG channels, and using the median over the channels removes disturbances present in one or a few of the channels. This also reduces the computational complexity as compared to feeding the classifier all channels.

The classification was performed using a SVM implemented using a freeware Matlab toolbox [15] running in Matlab 7.3. As kernel function the Radial Basis Function (RBF) [16] was used, with parameters set by optimizing the output result on the test set, using the parameter that produced the best total result for all patients in a leave-one-out scheme. The data was not divided further into a separate validation set because of the limited number of patients.

In order to maximize the use of the limited number of patients, the data was not divided into fixed training- and testing sets. Instead, leave-one-out cross-validation [17] was used, meaning that the data from one patient was used for testing, and the other six were used for training. This process was then repeated for all patients. To illustrate the differences found between the patients, individual performance measures are presented.

An output sample from a support vector machine depends only on one input feature vector, and does not take previous or subsequent samples into consideration. The input samples are represented by points in a multidimensional feature space, where the time order is not considered. The detection signal is typically noisy, with a lot of short detections. Using knowledge of the typical length of a burst, a smoothing algorithm was applied that uses a five point sliding window and checks if the majority of the samples in the window are burst detections. If they are, all samples in the window are converted to burst detections, if they are not; all samples in the window are converted to suppression. This results in

that groups of short detections are converted into longer continuous detections, while short isolated detections are removed, and increases performance in most cases.

For performance evaluation, sample sensitivity and specificity were used as measures. Sensitivity was defined as the percentage of the burst samples that were correctly classified as bursts, and specificity as the percentage of the remaining samples that were classified as suppression. As an alternative performance measure, a binary hit/miss measure was used, counting any burst interval that contained at least one detection as detected. This makes sense in a situation where the absolute burst limits are not so important, but rather a measure of the occurrence of bursts is desired.

The percentage of a recording that consists of suppression is an important measure of the recovery of a post-asphyctic newborn, and was also calculated, both from the reference data and from the detections for comparison.

3. Results

In Fig. 2 the five feature signals are compared. The figure shows that although the medians for burst and suppression of the feature signals differ, their distributions overlap considerably.

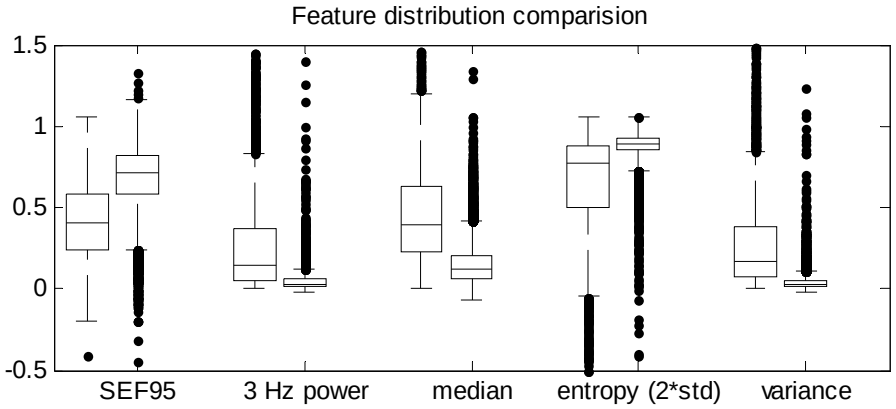


Fig. 2. The boxes show the median and quartiles for the feature signals and the whiskers show the rest of the data. The dots represent outliers. The boxes to the left show burst, and the ones to the right shows suppression. Data from all six patients are included.

Patient	BSR (%)	Median burst length (s)	Median supp. length (s)
1	13	5	1
2	6	1	2
3	27	4	3
4	30	3.5	4
5	1	3	5
6	75	7	6

Table II: Patient BS data

Table II shows the burst contents in the signals from the different subjects, and Table III shows the performance achieved for the different subjects. Fig. 3 shows an example of what the BS signal with both manual and SVM classification can look like.

Patient	Se (%)	Sp (%)	H/M	Hit	Miss	SR (%)	DSR (%)
1	74	98	95	20	1	87	89
2	89	94	100	21	0	94	89
3	75	99	100	15	0	72	79
4	63	97	83	15	3	70	78
5	97	83	100	10	0	99	81
6	35	100	83	24	5	24	74
Mean	72	95	94	18	2	74	82

Table III: Performance measures

Se: Sensitivity, the proportion of burst samples that are correctly classified.
Sp: Specificity, the proportion of suppression samples that are correctly classified.
H/M: Hit/Miss sensitivity, if at least one sample of the burst detected, the burst is considered to be detected.
Hit: The number of bursts that were detected.
Miss: The number of bursts that were not detected.
SR: Suppression Ratio, the percentage of the recording that is classified as suppression by the human expert.
DSR: Detected Suppression Ratio, the percentage of the recording that is classified as suppression by the algorithm.

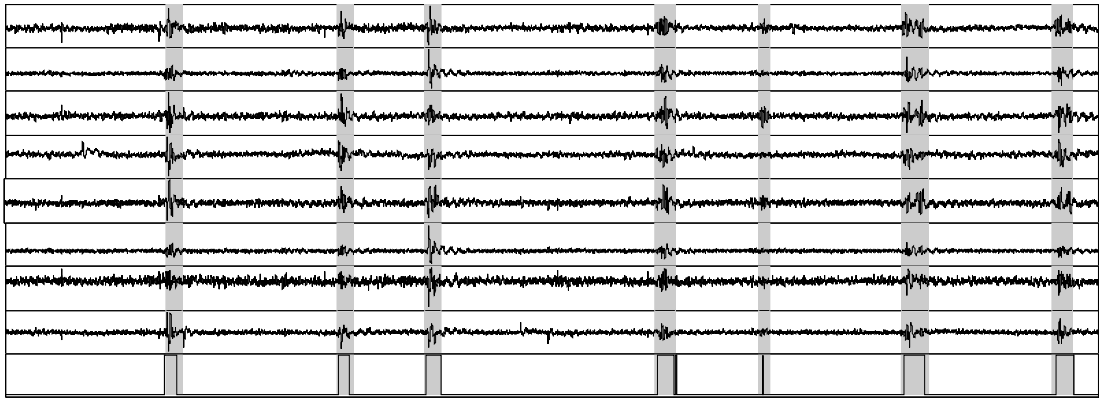


Fig. 3. Four minutes of EEG data taken from Patient 1. The shaded areas show the reference classification made by the expert, and the bottom graph show the classifications made by the algorithm. All six bursts are detected by the algorithm and there are no false detections that are not close to the actual bursts.

4. Discussion

In a previous study [10] bursts were detected using a Neural Network (NN). An important difference between the two techniques is that SVM finds a maximum margin between the classes, while the NN settles for any solution that separates them. When the classes are non-separable, the SVM should also give more deterministic results, because it is possible to control how large error that is allowed on the training set. The SVM is also more deterministic in the sense that a given set of input training data and parameters results in a specific classifier, as compared to the stochastically initiated NN training that may produce sub-optimal classifiers depending only on the random initiation. For this paper, both the data set and some of the preprocessing steps have been modified as well as the detection algorithm, so unfortunately the results are not directly comparable.

The method to calculate sensitivity and specificity used here is rather strict; every missed sample lowers the measured performance. The one-second time resolution used in the reference data probably contribute to a lowering of the calculated performance, because the algorithm often make much tighter detections when compared to the reference. The hit/miss sensitivity measure gives a complementary measure of performance. Comparing the amount of suppression in the manual classification and the SVM classification also gives an indication of the usefulness of the method. Measuring the amount of suppression can be used for diagnosing the recovery of a sick infant, and monitoring it automatically at bedside could be useful.

The optimization of the input parameters to the SVM is difficult when looking at the six patients separately. When adjusting a parameter, some

of the patients increase their performance while others decrease. The optimization ended up as a tradeoff that gives decent performance for all the patients in the group.

In order to produce a complete burst-suppression detection- and quantization system, artifact rejection should be included. It could also be useful to have a pre-classifier to first decide if the data is an example of burst-suppression or another type of periodic EEG activity, before the segmentation is made. During this study it has become apparent that the characteristics of the data vary significantly between different patients, and as mentioned in the introduction, one patient even had to be excluded because the burst activity was much more indistinct than the one found in the others, even though it still is categorized as burst suppression. This could be caused by varying degrees of cerebral dysfunction or other types of individual variations. That the patients in this study suffer from different degrees of illness is indicated by Table II, the amount of bursts in the EEG ranges from one percent up to 75 percent. For properly characterizing the apparently different groups of burst-suppression, more data is needed, especially when using this type of algorithms that require lots of data for training. However, this type of data is not easily obtained.

The results of this work will hopefully provide an easy to use and reliable tool for automatic segmentation of burst-suppression EEG in newborns that has suffered from asphyxia. This tool may be a part of a monitoring system for neonatal intensive care units, as well as a tool for segmenting burst-suppression data used for off-line analysis.

References

[1] G. Holmes, J. Rowe, J. Hafford, R. Schmidt, M. Testa, and A. Zimmerman, "Prognostic value of the electroencephalogram in neonatal asphyxia," *Electroencephalography and Clinical Neurophysiology*, vol. 53, pp. 60-72, 1982.

[2] D. B. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. M. T. Robertson, "EEG and long-term outcome of term infants with neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology*, vol. 110, pp. 655-659, 1999.

[3] C. C. Menache, B. F. D. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology,* vol. 27, pp. 93-101, 2002.

[4] M. Thordstein, B. G. Wallin, A. Flisberg, R. Bågenholm, I. Kjellmer, N. Löfgren, and K. Lindecrantz, "Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates," *Clinical Neurophysiology*, vol. 115, pp. 2461-2466, 2004.

[5] M. Thordstein, N. Löfgren, A. Flisberg, R. Bågenholm, K. Lindecrantz, and I. Kjellmer, "Infraslow EEG activity in burst periods from post asphyctic full term neonates," *Clinical Neurophysiology*, vol. 116, pp. 1501-1506, 2005.

[6] M. D. Lamblin, M. André, M. J. Challamel, L. Curzi-Dascalova, A. M. D'Allest, E. De Giovanni, F. Moussalli-Salefranque, Y. Navelet, P. Plouin, M. F. Radvanyi-Bouvet, D. Samson-Dollfus, and M. F. Vecchierini-Blineau, "EEG in premature and full-term newborns. Maturation and glossary," *Neurophysiologie Clinique*, vol. 29, pp. 123-219, 1999.

[7] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology*, vol. 89, pp. 980-1002, 1998.

[8] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *Journal of Clinical Monitoring and Computing*, vol. 15, pp. 357-367, 1999.

[9] V. N. Vapnik, *Statistical Learning Theory*. New York: John Wiley & Sons Inc., 1998.

[10] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Detection of Bursts in the EEG of Post Asphyctic Newborns," in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 2179-2182.

[11] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: John Wiley & Sons, Inc., 1996.

[12] C. Jeleazcov, S. Egner, F. Bremer, and H. Schwilden, "Automated EEG preprocessing during anaesthesia: New aspects using artificial neural networks," *Biomedizinische Technik*, vol. 49, pp. 125-131, 2004.

[13] J. A. Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. Belmont: Duxbury Press, 1995.

[14] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, 1948.

[15] A. Schwaighofer, "SVM toolbox for Matlab," 2.51 ed Berlin: Frauenhofer FIRST, Intelligent Data Analysis, 2002.

[16] S. R. Gunn, "Support Vector Machines for Classification and Regression," Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, University of Southampton1998.

[17] R. O. Duda, Hart, P. E., Stork, D. G., *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, Inc., 2001.

Paper III

Classification of burst and suppression in the neonatal EEG

J. Löfhede[1,2], N. Löfgren[1], M. Thordstein[3], A. Flisberg[4], I. Kjellmer[4] and K. Lindecrantz[1]

Classification of burst and suppression in the neonatal EEG

J. Löfhede[1,2], N. Löfgren[1], M. Thordstein[3], A. Flisberg[4], I. Kjellmer[4] and K. Lindecrantz[1]

[1] School of Engineering, University College of Borås, Borås, Sweden.
[2] Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden
[3] Department of Pediatrics, Göteborg University, The Queen Silvia Children's Hospital, Göteborg, Sweden.
[4] Institute of Neuroscience and Physiology, Section for Neuroscience and Rehabilitation, Sahlgrenska University Hospital, Göteborg, Sweden.

E-mail: johan.lofhede@hb.se

Abstract

Fisher's linear discriminant (FLD), a feed-forward artificial neural network (ANN) and a support vector machine (SVM) were compared with respect to their ability to distinguish bursts from suppressions in electroencephalograms (EEG) displaying a burst-suppression pattern. Five features extracted from the EEG were used as input. The study was based on EEG signals from six full term infants who had suffered from perinatal asphyxia, and the methods have been trained with reference data classified by an experienced electroencephalographer. The results are summarized as the area under the curve (AUC), derived from receiver operating characteristic (ROC) curves for the three methods. Based on this, SVM performs slightly better than the others. Testing the three methods with combinations of increasing numbers of the five features shows that the SVM handles the increasing amount of information better than the other methods.

1. Introduction

During delivery there is always a risk of insufficient circulation or blood gas exchange to the baby, something that may lead to hypoxia (lack of oxygen) and potentially cause brain damage. Babies at risk are kept under close surveillance during delivery and afterwards at a neonatal intensive care unit (NICU), but it is difficult to determine at an early stage if the babies are recovering, if there is a threat of brain damage or if it already has occurred [1]. Parameters such as heart rate, blood pressure and oxygen saturation are monitored regularly, but they are only measures of the general condition of the baby. If the function of the brain itself is to be monitored, a more direct way is to measure the electrical signals

spontaneously produced by the brain, the electroencephalogram (EEG). This offers a possibility for continuous monitoring over long time with high time resolution.

Earlier studies have investigated how certain parameters calculated from the EEG signal can be used for detecting hypoxia in the brain, and predicting the outcome after a hypoxic event [2-4]. In practice, however, most EEG recordings are evaluated through visual inspection of the unprocessed signal by a clinical neurophysiologist, an expertise typically not available at the NICU. Even though methods for remote consultations have been developed [5], this methodology only allows for evaluation at distinct time instances and is not suitable for continuous bedside monitoring.

One attempt to simplify long-time monitoring of brain function that can be used for bedside monitoring is the amplitude-integrated EEG (aEEG), for example implemented as the cerebral function monitor (CFM) [6]. This method displays a processed version of a usually one-channel EEG on a compressed time scale, and provides the clinician with a simple way to monitor the brain activity of a patient. However, the compressed signal can sometimes mask interference and artefacts and cause them to be mistaken for brain activity, and there are also examples of missed seizure activity [7]. Because of these limitations, neurophysiologists argue that the unprocessed EEG signal has to be taken into consideration when interpreting the aEEG.

This paper addresses the problem of classifying epochs of periodic burst-suppression (BS) EEG as burst or suppression and compares three different classification algorithms. Burst-suppression is intermittent activity, usually consisting of longer low amplitude activity segments (suppression) and shorter periods of high amplitude activity (bursts), indicating severe pathology in a newborn baby. The relative amount of suppression, expressed as for example interburst interval (IBI) or burst-suppression ratio (BSR), provides a measure of the status of the brain and may also give prognostic information [8]. Following trends over time in these parameters gives an indication of whether a baby is recovering or not. Some attempts towards automatic classification of burst-suppression have been made, most of them targeting burst-suppression induced by anaesthesia [9, 10], but also for monitoring neonates [11].

The intended setting for this classification problem is that previous classifiers have already decided that the EEG indeed consists of BS activity. The task of the present classifier is then to segment the EEG into burst and suppression given BS EEG. An ongoing study is addressing the problem of comparing BS with other types of intermittent EEG in neonates.

The methods and results presented here have been partly presented earlier in conference proceedings and a licentiate thesis [12-15]. The methods include three different classification algorithms, the feature functions that are used as inputs to the classifiers, and the pre-processing that has been used. Results are presented for a data set containing EEG data from six full-term babies whose EEG exhibit burst-suppression caused by asphyxia during birth.

2. Methods

2.1. Data

All data were collected at a neonatal intensive care unit (NICU) using the SACS® [16] system. An experienced electroencephalographer then chose suitable periods of EEG and exported them to the Matlab environment, where all subsequent processing was performed.

The data that have been used consist of EEG recordings from six full-term newborn infants having suffered from perinatal asphyxia (impaired gas transfer and circulation around the time of birth). Eight channels were used, with electrodes placed according to the international 10-20 system [17] at positions F7, F3, T5, P3, F8, F4, P4 and T6. The number of electrodes used in the study is a compromise between the desire to get high spatial resolution and practical issues regarding the size of the babies' heads and the number of cables. We found that eight channels evenly distributed over the head gave a sufficient resolution and gained acceptance from parents and clinicians. The data were digitized at a sampling rate of 200 Hz.

Each subject contributed with a continuous recording of 6-40 minutes, selected and manually classified as burst or suppression by the electroencephalographer. The length was chosen to include at least 10 bursts, and all artefacts were manually identified and marked for later exclusion from the analysis. The number of bursts in each signal is stated in Table 1. Because the signals only contain burst and suppression, the number of suppression segments in each signal is equal to the number of bursts plus one (all signals start and end with suppression).

The reference classifications were manually recorded with a resolution of one second. When the start or end of a burst did not coincide with even seconds, the time entry was rounded so that the entire burst was included. This meant that in some cases up to nearly two seconds of suppression activity were included with the burst.

Suppression periods contain very little activity, while the bursts are short periods of higher activity. Figure 1 shows an example of a few very

pronounced bursts in a longer suppression segment. As can be seen in the figure, the bursts do have higher amplitude than the suppressions, but the whole burst would not be detected by a simple amplitude threshold.

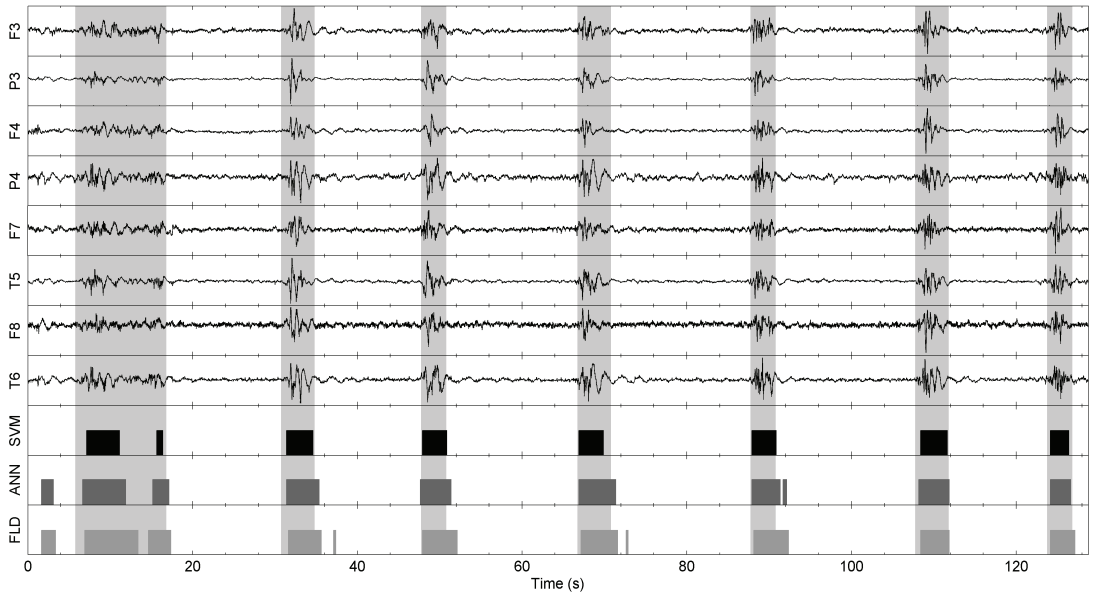


Figure 1: 2.5 minutes of BS EEG from patient 3 in a CAR montage after band-pass filtering and 50 Hz-interference suppression. This is an example of very pronounced burst-suppression activity, where the shaded intervals are bursts marked by a human expert. The three bottom fields show classifications made using the three methods.

Table 1 shows some characteristics of the signals from the different patients, and illustrates that while the signals all comprise burst-suppression EEG, the characteristics vary significantly. For example, patient 5 has a much longer median interburst interval than the other four, and patient 6 has lower burst-suppression ratio and higher mean power in the suppression parts than the others.

Table 1: Characteristics of the raw EEG signals (after filtering)

Pat.	BSR	TL (s)	mBL (s)	mIBI (s)	mBP (μV^2)	mSP (μV^2)	nB
1	87	775	5	32	72	9	23
2	94	483	1	25	110	7	23
3	73	265	4	13	162	14	15
4	71	307	3.5	13	61	10	22
5	99	2299	3	127	28	4	11
6	25	329	7	2	94	23	31

BSR: burst suppression ratio, the percentage of suppression in the record [18].
TL: total signal length, after removal of artefacts
mBL: median burst length
mIBI: median interburst interval (median suppression length)
mBP: mean burst power
mSP: mean suppression power
nB: Number of bursts in signal (number of suppressions is nB plus one)

The long term outcome was that patient 6 had only minor damages, while all the others were diagnosed with different degrees of cerebral palsy. Patient 2 died at the age of ten months.

2.2. Performance measures

For performance evaluation, sample sensitivity and specificity were used as measures. Sensitivity was defined as the percentage of the burst samples that were correctly classified as bursts, and specificity as the percentage of the suppression samples that were classified as suppression.

Receiver operating characteristic (ROC) curves [19] were formed by taking the raw output from the respective algorithm and using a range of thresholds covering the range of the outputs. These curves show all possible trade-offs between sensitivity and specificity. The area under the curve (AUC) provides a single number for comparing the classifiers, with the value one representing the perfect classifier.

Because of the small number of patients, the individual performance measures were not averaged, but kept separate for comparison of the performance for different patients. When all different combinations of features were tested, the mean of the results for the different patients were used to enable a simple visualisation of the results.

2.3. Training set

Data from six patients were available, making it unfeasible to divide the data in fixed training- and test sets. Instead leave-one-out training was used in the sense that one patient was held out and used for validation while all the data from the remaining five were used for training. This was repeated for each patient, thus producing six ROC curves for each classifier, and no classifier instance was ever tested on data from a patient that was used for training it. This scheme was used to maximise the use of the limited dataset, while also making sure that the classifiers were tested in a situation that mimicked a real scenario. In a hospital setting it would be unpractical if the classifier would need individual training for each new patient it was used on.

2.4. Pre-processing

The signal level of the EEG is very low when measured from the scalp, in the range of tens of microvolts, making it very sensitive to interference from surrounding electrical fields created by common electrical appliances. These common mode disturbances are reduced by using the CAR (common average reference) montage, meaning that the mean of all channels is subtracted from each channel.

The signals were also band-pass filtered between 0.5 and 20 Hz to remove low- and high-frequency interference before feature generation. In one case, an LMS (least mean square) adaptive filter [20] with a separate ECG channel as reference was used to suppress ECG interference in the EEG signal.

2.5. Features

Features are mathematical functions that are applied on the samples of the EEG signal that falls in a sliding window. The feature signal represents a certain characteristic of the underlying signal, and should in the classification case enhance the characteristics that differ between the classes. When multiple features are used they should complement each other, and be as independent as possible.

The features that were used in this work are listed in table 2, and include measures of power, frequency distribution and entropy of the signal. These features were chosen because they were thought to complement each other. More advanced ways of selecting features exist, but the scope of this work has mainly been to evaluate the classifiers, not the features.

Table 2: Feature functions

Feature	Description
Spectral Edge Frequency (SEF95) [18]	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on rectangular windows of the signal
3 Hz power	Power in a 1-Hz wide band centred at 3 Hz
Median [21]	Median absolute value
Variance ($s^2(x)$) [21]	$s^2(x) = \frac{1}{N-1}\sum_{n=1}^{N}(x[n]-\mu)^2$ where x is a time series, and μ is the sample mean of x
Shannon entropy (H_{Sh}) [22]	$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$ $p(I_1)\ldots p(I_U)$ is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals $I_1,..,I_U$. 20 intervals were used evenly distributed between the maximum and minimum values of the signal in the window. H_{Sh} is a measure of uncertainty of a random variable.

A sliding window with a length of one second, corresponding to 200 samples, was used, with an overlap of 0.75 seconds. The window length was chosen to match the one-second resolution used in the reference data,

which in practice set the minimum burst length to one second. This produces feature signals with an effective sample rate of 4 Hz. The choice of window length results in a trade-off between time resolution and accuracy. All feature signals (except variance) were smoothed by convolution with a six-point triangular window. This was performed because we know that a burst is at least one second long and we want to suppress short spikes or dips in the feature signal to get longer continuous detections. Finally the feature signals were normalized by fitting their distributions between the first and the 99th percentile into the interval zero to one, thus giving all features the same amplitude range. This removes the bias caused by different features having different magnitudes. The proposed normalization method is suitable when the signal is known to contain both burst and suppression activity. Figure 2 gives an example of what the five features can look like in the time domain.

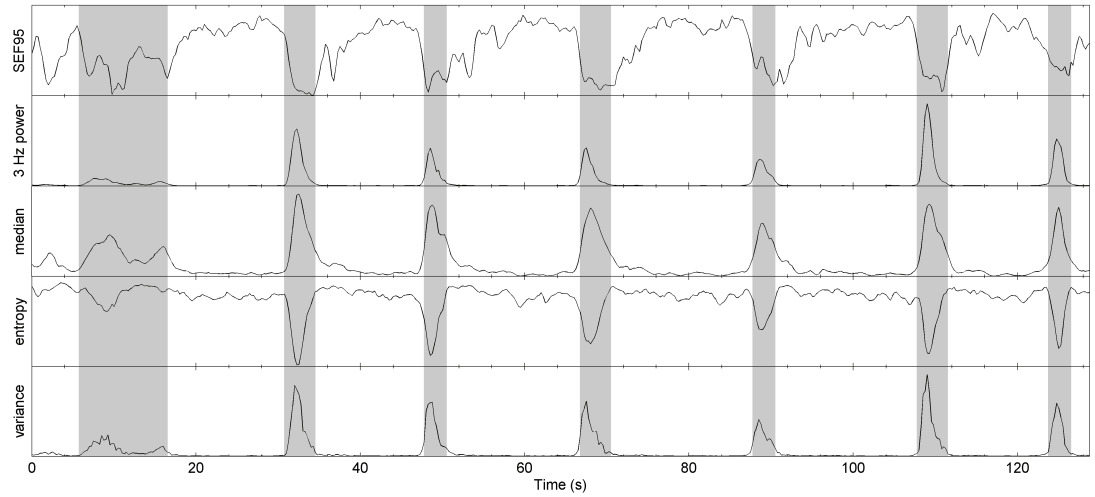


Figure 2: An example of the five features used in this work. The features were extracted from the 2.5 minute long EEG epoch shown in figure 1. The shaded areas were marked as bursts by an electroencephalographer.

In order to give all patients an equal weight in the training of the respective classifier, the data from each patient was repeated so that all records were the same length as the longest one. Since the classifiers are trained on data from five patients in ways that minimises the classification error in a statistical sense, this approach is used to make sure that the training does not put higher emphasis on the data from the patients with longer recordings. An alternative could have been to instead set the limit to the shortest recording and draw samples from the longer ones, but this was judged to be a waste of data.

2.6. Feature selection

When using nonlinear classifiers such as neural networks or support vector machines, nonlinear relationships in the feature data that are not obvious may be found. To find the best combination of features for a

given problem, *exhaustive search* was used. The exhaustive search guarantees an optimal result and is performed by testing all possible combinations of increasing numbers of included features and then choosing the best one. This is only possible when a limited number of features are available, since the number of combinations grows dramatically when more features are added.

The results described in figure 1-4 were generated using all five described features in parallel, while Figure 5-8 shows results from the exhaustive search of all combinations of the five features.

2.7. Classifiers

The classification algorithms Fisher's linear discriminator (FLD), artificial neural networks (ANN) and support vector machines (SVM) were compared in this work. They are all trained using supervised learning and hence require pre-classified training data. They can handle multidimensional input training data, i.e. several features in parallel, and form classifiers by trying to find the optimal boundary between the classes in the multidimensional space defined by the feature vectors.

All of these methods are based on training a classifier using pre-classified training data, and then applying it to test data. The reason for choosing this type of algorithms is that we envision a system that is trained on a number of manually classified training cases, once and for all. Then this system can be used for classifying new cases without the need of any manual adaptation when it is applied to a new patient, something that would be of great benefit in a hospital setting.

2.7.1. Fisher's Linear Discriminant

This method is based on scatter matrices [19, 23] formed from the training data from the two classes. These are used for deriving the projection of the multidimensional input space onto the line that gives the maximum ratio of between-class scatter to within-class scatter. The projection is then applied to the test data, resulting in a mapping from the five-dimensional feature signal to a one-dimensional decision function. The detections are then acquired by applying a threshold, in this case chosen implicitly using maximum likelihood (ML) [19].

2.7.2. Artificial Neural Network

An artificial neural network (ANN) [19, 24] adapts a decision boundary to the data during a training phase by iteratively moving a number of linear planes based on the error in each step. The network was set up as a three-layer feed-forward network with ten neurons in the hidden layer and one neuron in the output layer. The number of neurons was determined to be large enough through empirical testing. The hidden neurons were

equipped with sigmoid activation functions, while a linear output neuron was used for enabling the generation of a smooth ROC curve. The training of the network was performed by error back-propagation using randomly selected training examples. For each set of training data, five instances of the network were trained and the one with the best AUC score was selected. This is done because the neural network is initialized randomly, and the training can get stuck in a sub-optimal state.

2.7.3. Support Vector Machine

The SVM [19, 25, 26] optimizes a nonlinear boundary that maximizes the margin between the classes. The system was implemented using a freeware Matlab toolbox [27]. As kernel a Gaussian radial basis function (RBF) [28] was used, with a width parameter that has to be chosen. The SVM setup also includes the choice of a weight for misclassified samples, and these two parameters were chosen by performing a grid-search over a suitable area and choosing the combination that produced the best AUC score for all patients.

3. Results

3.1. Performance when using all five features

In figure 1 an example of detections from the different classification methods is displayed together with the EEG. The grey blocks in the bottom part of the figure represent the classifications made with the three methods, while the shading shows the reference classification made by the human expert. All bursts in the epoch are detected and the differences between the methods are small. These results are just examples of detections in the time domain, and represent just one point along the ROC curve. Specific results depend on what parameters are chosen when training the classifier.

Figure 3 shows the results of the methods in the form of ROC curves, one plot for each method with one curve for each patient. These curves reflect the full potential of a method by showing all possible trade-offs between sensitivity and specificity.

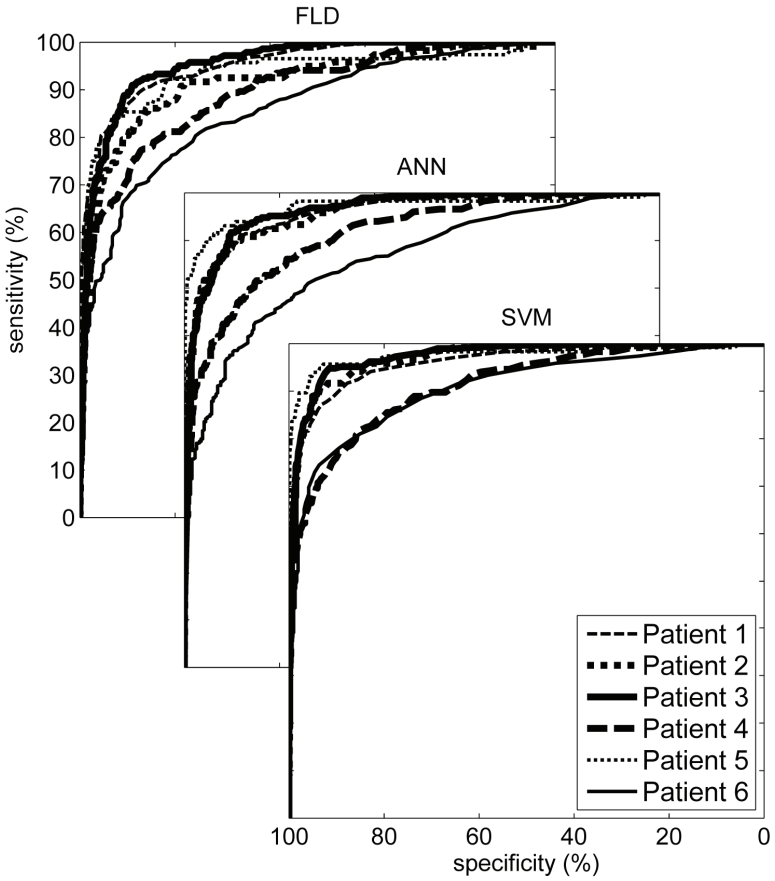


Figure 3: ROC curves for the three detection methods, with separate curves for the six patients.

The AUC values for the classifiers are summarized in figure 4. This summary shows that the patients fall into two groups: one group where the methods perform well, i.e. for patients 1, 2, 3 and 5, and one group where the methods perform less well, i.e. numbers 4 and 6.

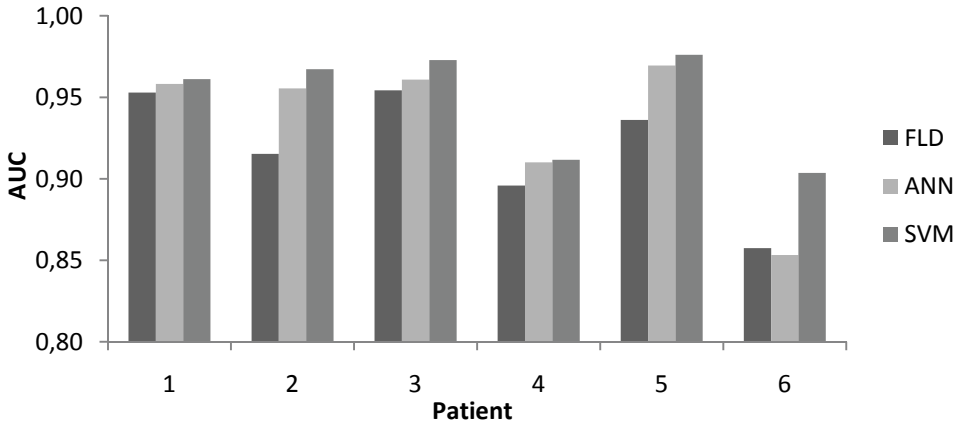


Figure 4: Comparison of the AUC results for the different methods, displayed for the six patients individually. The values are the areas under the curves in figure 3; with one representing the perfect classifier and 0.5 representing random classification (observe that the AUC-axis starts at 0.80).

The comparison also shows the tendency that the FLD performs worst in general, and SVM performs best. The statistical material (six patients) is however too small to calculate statistics or draw solid conclusions about the performance of the methods.

3.2. The performance when varying the number of features

The result of evaluating all possible combinations of the five features is shown in figures 5 to 8. The figures show the performance for the three methods when including increasing numbers of features in the classification. For each number of included features the combination that produced the highest individual AUC score was used. Figure 5 to 7 show the performance for the individual patients while figure 8 shows a comparison of the mean AUC for the patients in the same plot. The plots show that all three methods have a tendency to decrease in performance when more than two features are used. Figure 8 show that FLD decreases the most, while SVM only decreases slightly and ANN has performance values between the two other methods. It should be noted that the function of the classifier algorithm in practice is equal to determining a threshold in the case when only one feature (one dimension) is used, reflected by the essentially identical results for the three classifiers in the one-feature case.

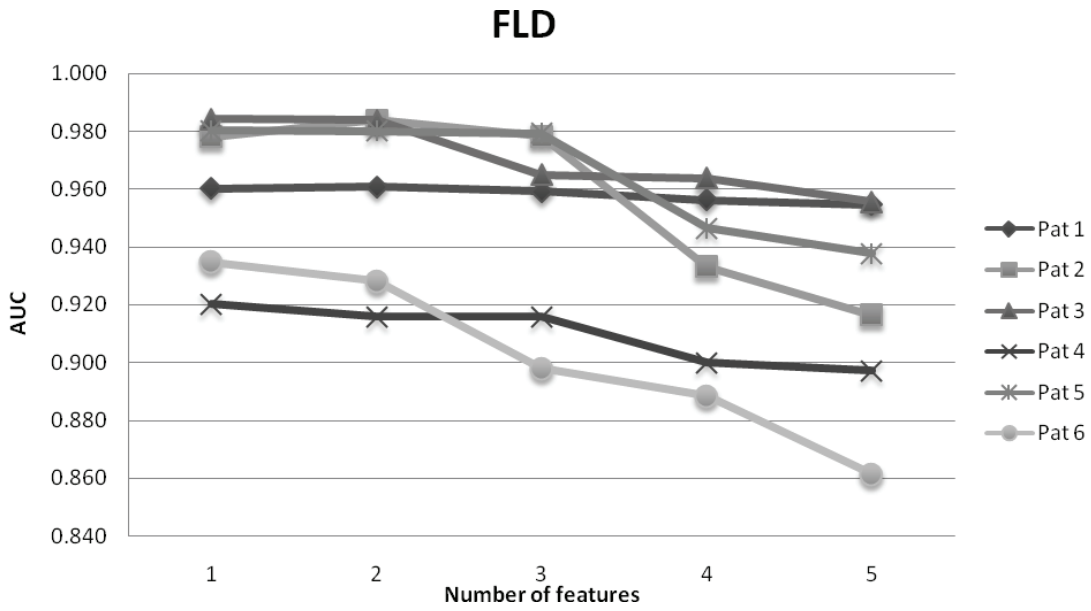


Figure 5: AUC performance for FLD as a function of the number of included features.

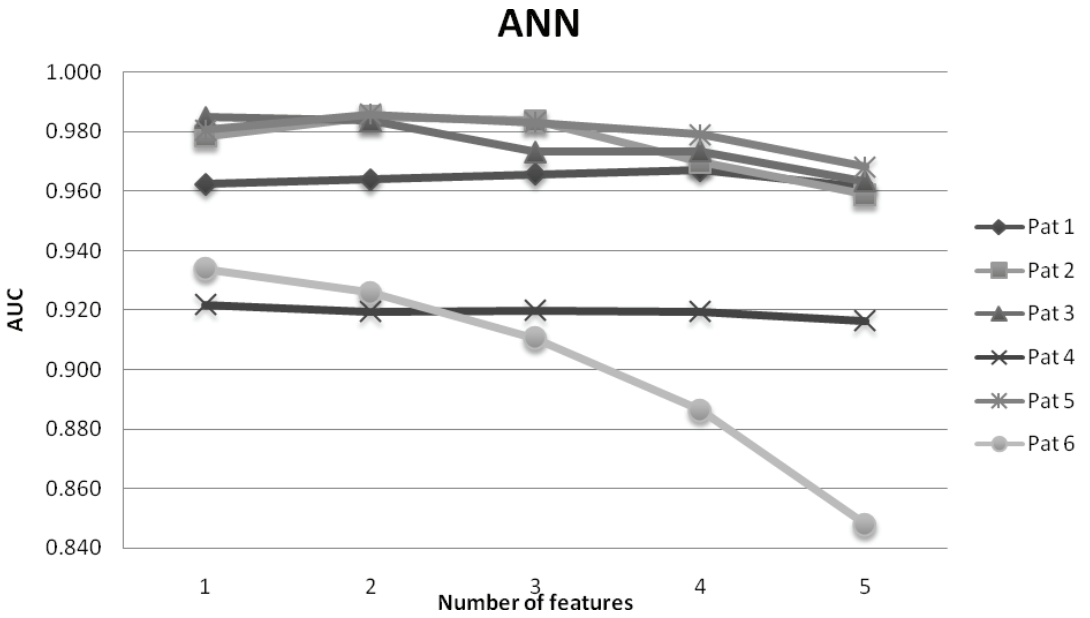


Figure 6: AUC performance for ANN as a function of the number of included features.

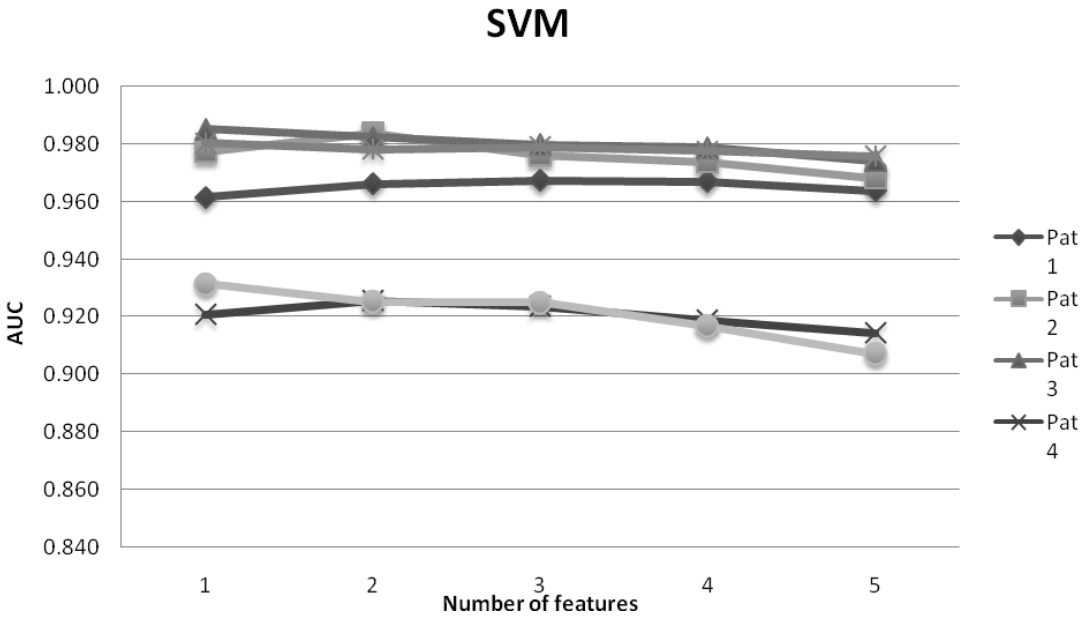


Figure 7: AUC performance for SVM as a function of the number of included features.

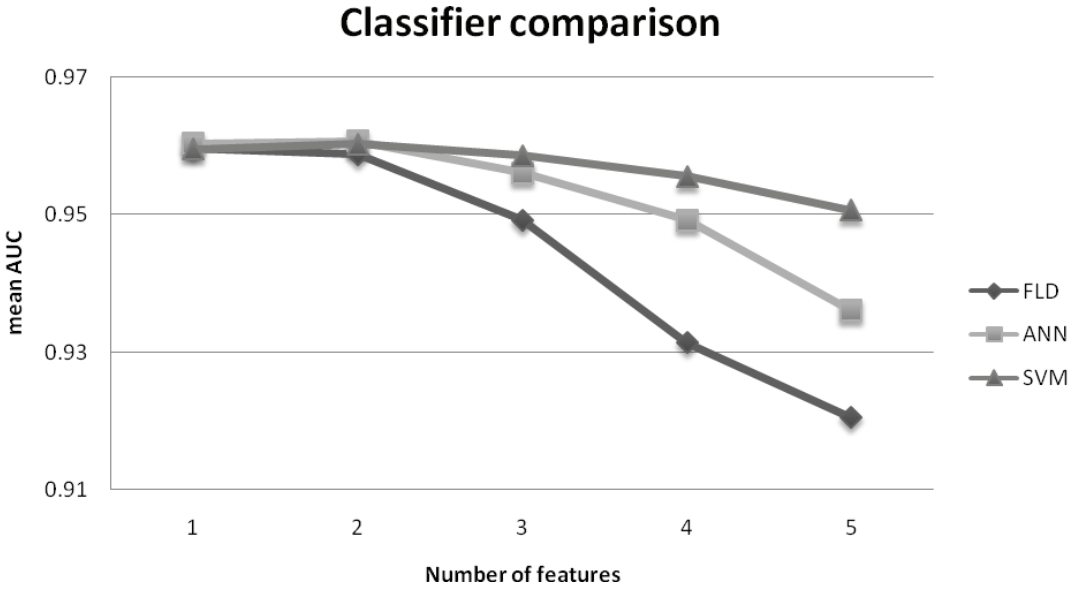


Figure 8: Comparison of the mean AUC values as a function of the number of included features. (Note that the AUC-axis has different limits than in the previous figures).

4. Discussion

Three methods for detecting bursts in burst-suppression EEG have been compared, all of which allows for many variations in implementation. Fisher's linear discriminant (FLD) is the most straight-forward of the three; it provides the user with a well-defined projection of the data that can be thresholded automatically by using maximum likelihood. The support vector machine (SVM) on the other hand requires the user to choose a kernel and to find appropriate values for the misclassification weight and the kernel parameter. The artificial neural network (ANN) has to be set up with respect to network topology and types of activation functions and the training can be adjusted using parameters that control step length, momentum and weight elimination. All these variables make the task of making a completely fair comparison of the methods difficult.

The trade-off between sensitivity and specificity is determined by the way the final threshold is chosen. When the threshold is chosen implicitly by ML estimation, normalized probability density functions for the two classes are used, meaning that the proportion of burst and suppression in the training data does not matter. When using ANN, the number of samples matter because each time a sample of either kind is encountered on the "wrong side" of the border in the training process the border is pushed in the appropriate direction. The situation is similar in SVM, where (in the current implementation) equal weights are put on misclassified samples from both classes. This inequality is partly eliminated when the final thresholding is removed from the algorithms,

and the raw output is used for generating the ROC curves. This procedure moves the border generated by the respective algorithms, and is essentially equal to changing the penalty for misclassifying samples from the two classes, depending on in which direction the border is moved.

The method to calculate sensitivity and specificity used here is rather strict; every missed sample lowers the measured performance. The one-second time resolution used in the reference data probably contribute to a lowering of the calculated performance, because the algorithms often make much tighter detections when compared to the reference. It can be argued that exact detection of the time limits of individual bursts may not be the most important part of a monitoring system. Instead, parameters extracted from the segmented signal, for example the burst-suppression ratio, could be compared to the outcome of the patient, because the most important aspect of a monitoring device should be that it displays information that is relevant to the health status of the patient. This type of testing has not been possible so far, but field-tests at the NICU are being planned.

The detection results for the different patients tend to fall into two groups, with results for patient 4 and 6 at a lower level than the others. These differences may be explained by varying degree of cerebral dysfunction, variations in signal quality or some other type of individual variation. For example, patient 6 had significantly higher power in the suppression parts of the signal and a lower BSR than the others (table 1). This could indicate a lower degree of cerebral dysfunction, corroborated by the fact that patient 6 only had minor injuries at the long-term follow up while all others were diagnosed with cerebral palsy. Other possible reasons may be that the electrodes can have been applied slightly different on the patients, giving them different frequency properties. That this factor may be at play here is indicated by the fact that the extremely low frequency activity that often is associated with bursts [29, 30] could only be found for some of the patients (data not shown). It could also be that the present features do not represent what is common in the group of patients in an optimal way. The patients were however chosen to be similar in terms of appearance of the EEG signal when analyzed visually, and all were severely ill at birth after a full-term pregnancy.

The problem of different patients having different characteristics could possibly be solved by using the right combination of multiple features. The idea is to use many features, some of which provide useful information for some of the patients, and other features that provide this for other patients. The job of the classifier is then to combine the information so that the large amount of dependencies does not degrade the classification performance.

The exhaustive feature search in figure 5 shows a decrease in performance when using more than two features. The difference between the three classifiers also tends to increase with the number of features. This is probably due to that the present set of features includes a significant amount of redundant information, meaning that increasing the number of features in the classification does not add more useful information regarding the class membership. All data that is introduced in the system does however include noise, thereby reducing the signal to noise ratio. The fact that the SVM has a smaller decrease in performance than the others indicates that it is more robust with respect to noise and dependencies within the feature data set.

Burst-suppression measures are commonly used in measures of the depth of anaesthesia in adults undergoing surgery, for example as a part of the bispectral index (BIS) [18], a composite parameter that is a combination of multiple parameters. Most published work on burst suppression detection has focussed in this area, and some methods that have been explored are the use of single features as decision functions [9] or using multiple neural networks with various input features [10]. When compared to the method described in [9], with the conditions set as equal as possible (the decision function was generated for all eight available channels and then the median was applied), the SVM was slightly better for some of these neonatal patients in terms of ROC curves.

Other possible applications of classification algorithms in EEG could for example be recognizing short events such as artefacts and seizures, but they could also be used for classifying longer epochs into different types of ongoing activity such as burst-suppression, tracé alternant or continuous EEG. This could possibly be used to summarize the full EEG in a more detailed form than what is possible with methods such as aEEG/CFM, and be used as decision support in a monitoring system at the NICU.

A monitoring system will have to contain functionality for detecting artefacts and excluding them from analysis. It should also be able to recognize seizures. With some further effort, this work will hopefully provide an easy to use and reliable tool for automatic segmentation of burst-suppression EEG in newborns, for example suffering the after-effects of asphyxia. This tool may be a part of a monitoring system for neonatal intensive care units, as well as a tool for segmenting burst-suppression data used for off-line analysis.

5. Conclusion

The AUC results in figure 4 indicate that the SVM has the best performance, followed by ANN and FLD. SVM and ANN are non-linear

methods while FLD is purely linear, something that might give FLD some disadvantage. However, the difference in performance is small, and although there are differences in the level of computational complexity, all methods would work on line when implemented in commonly available computer hardware. As shown in figure 8, SVM is the most robust method when dealing with multidimensional feature data.

Acknowledgements

This study was supported by a grant from the BIOPATTERN EU Network of Excellence, EU Contract 508803, and the Margarethahemmet Foundation.

References

[1] M. Scher, "Perinatal asphyxia: timing and mechanisms of injury in neonatal encephalopathy," *Current neurology and neuroscience reports,* vol. 1, pp. 175-184, 2001.

[2] N. Löfgren, K. Lindecrantz, A. Flisberg, R. Bågenholm, I. Kjellmer, and M. Thordstein, "Spectral distance for ARMA models applied to electroencephalogram for early detection of hypoxia," *Journal of Neural Engineering,* vol. 3, pp. 227-234, 2006.

[3] G. L. Holmes and C. T. Lombroso, "Prognostic value of background patterns in the neonatal EEG," *Journal of Clinical Neurophysiology,* vol. 10, pp. 323-352, 1993.

[4] M. I. Shevell, A. Majnemer, and S. P. Miller, "Neonatal neurologic prognostication: the asphyxiated term newborn," *Pediatric Neurology,* vol. 21, pp. 776-784, 1999.

[5] N. Löfgren, K. Lindecrantz, M. Thordstein, A. Hedström, B. G. Wallin, S. Andreasson, A. Flisberg, and I. Kjellmer, "Remote Sessions and Frequency Analysis for Improved Insight Into Cerebral Function During Pediatric and Neonatal Intensive Care," *IEEE Transactions on Information Technology in Biomedicine,* vol. 7, pp. 283-290, 2003.

[6] P. F. Prior and D. E. Maynard, *Monitoring Cerebral Function*. Amsterdam: Elsevier, 1986.

[7] I. Rosén, "The Physiological Basis for Continuous Electroencephalogram Monitoring in the Neonate," *Clin Perinatol,* vol. 33, pp. 593-611, 2006.

[8] C. C. Menache, B. F. D. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology,* vol. 27, pp. 93-101, 2002.

[9] M. Särkelä, S. Mustola, T. Seppänen, M. Koskinen, P. Lepola, K. Suominen, T. Juvonen, H. Tolvanen-Laakso, and V. Jäntti,

"Automatic analysis and monitoring of burst suppression in anesthesia," *Journal of Clinical Monitoring and Computing,* vol. 17, pp. 125-134, 2002.

[10] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *Journal of Clinical Monitoring and Computing,* vol. 15, pp. 357-367, 1999.

[11] G. Hauksson, "Automated Analysis of Newborn EEG," in *Electrical Engineering*. vol. Master of Science Stockholm: Royal Institute of Technology, 2006, p. 93.

[12] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Detection of Bursts in the EEG of Post Asphyctic Newborns," in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 2179-2182.

[13] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Classifying Burst and Suppression in the EEG of Post Asphyctic Newborns using a Support Vector Machine," in *3rd International IEEE EMBS Conference on Neural Engineering*, Kohala Coast, Hawaii, 2007.

[14] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Comparison of three methods for classifying burst and suppression in the EEG of post asphyctic newborns," in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007.

[15] J. Löfhede, "Classification of Burst and Suppression in the Neonatal EEG," in *Department of Signals and Systems*. vol. Licentiate of Engineering Göteborg: Chalmers University of Technology, 2007, p. 55.

[16] K. Lindecrantz, R. Bågenholm, F. Göthe, A. Hedström, N. Löfgren, S. Nivall, and J. Ouchterlony, "A general system used in monitoring of cerebral and circulatory function in neonatal intensive care," *Med Biol Eng Comput,* vol. 37, pp. 888-9, 1999.

[17] H. H. Jasper, "The Ten-Twenty Electrode System of the International Federation," in *International federation of societies for electroenchephalography and clinical neurophysiology*, Brussels, 1957, pp. 371-375.

[18] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology,* vol. 89, pp. 980-1002, 1998.

[19] R. O. Duda, Hart, P. E., Stork, D. G., *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, Inc., 2001.

[20] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: John Wiley & Sons, Inc., 1996.
[21] J. A. Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. Belmont: Duxbury Press, 1995.
[22] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* vol. 27, pp. 379-423, 623-656, 1948.
[23] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics,* vol. 7, pp. 179-188, 1936.
[24] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.,* vol. 5, pp. 115-133, 1943.
[25] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2 ed. New York: Springer-Verlag New York Inc., 2000.
[26] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "Training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, Pittsburgh, PA, USA, 1992, pp. 144-152.
[27] A. Schwaighofer, "SVM toolbox for Matlab," 2.51 ed Berlin: Frauenhofer FIRST, Intelligent Data Analysis, 2002.
[28] S. R. Gunn, "Support Vector Machines for Classification and Regression," Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, University of Southampton1998.
[29] T. Lipping, P. Loula, V. Jantti, and A. Yli-Hankala, "DC-level detection of burst-suppression EEG," *Methods of Information in Medicine,* vol. 33, pp. 35-38, 1994.
[30] M. Thordstein, N. Löfgren, A. Flisberg, R. Bågenholm, K. Lindecrantz, and I. Kjellmer, "Infraslow EEG activity in burst periods from post asphyctic full term neonates," *Clinical Neurophysiology,* vol. 116, pp. 1501-1506, 2005.

Paper IV

Comparing a Supervised and an Unsupervised Classification Method for Burst Detection in Neonatal EEG

J. Löfhede, J. Degerman, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Comparing a Supervised and an Unsupervised Classification Method for Burst Detection in Neonatal EEG

J. Löfhede, J. Degerman, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, K. Lindecrantz

Abstract

Hidden Markov Models (HMM) and Support Vector Machines (SVM) using unsupervised and supervised training, respectively, were compared with respect to their ability to correctly classify burst and suppression in neonatal EEG. Each classifier was fed five feature signals extracted from EEG signals from six full term infants who had suffered from perinatal asphyxia. Visual inspection of the EEG by an experienced electroencephalographer was used as the gold standard when training the SVM, and for evaluating the performance of both methods. The results are presented as receiver operating characteristic (ROC) curves and quantified by the area under the curve (AUC). Our study show that the SVM and the HMM exhibit similar performance, despite their fundamental differences.

1. Introduction

The burst-suppression (BS) pattern is one of several indicators of severe pathology in the electroencephalogram (EEG) signal that may occur after brain damage, caused by *e.g.* asphyxia (insufficient gas and nutrient supply) around the time of birth [1]. Certain characteristics of this pattern can provide clinicians with important information about the prognosis of the patient and is thus important in the adjustment of the treatment. Examples of important characteristics of the BS pattern are length of burst and suppression intervals, percentage of suppression activity in a recording, and spectral contents of bursts [2-4].

In practice, most EEG recordings are evaluated by experienced neurophysiologists through visual inspection of the unprocessed signal. Concerning the analysis of burst suppression EEG, some attempts towards automatic classification of burst-suppression have been made, most of them targeting burst-suppression induced by anaesthesia [5, 6], but also for monitoring neonates [7].

Our goal is to develop tools that can be used for automatic classification of BS caused by *e.g.* perinatal asphyxia. The target patient group is newborns, and a possible future application is a monitoring system for neonates under intensive care.

This paper compares Support Vector Machines (SVM) and Hidden Markov Models (HMM) with respect to their performance when classifying burst and suppression in neonatal EEG. An important difference between the methods is that the SVM uses supervised training while the HMM uses unsupervised training. Another difference is that the SVM examines the data samples without regard to order, while the HMM includes the state of the previous sample (burst or suppression) in the classification. The underlying process is assumed to be a Markov process when deciding the most probable state of the present sample (the Markov condition [8]). The word hidden in HMM refers to that the states and the transition probabilities are not directly observable in the data, but each state is associated with an observable output, in our case a set of features.

2. Methods

2.1. Data

The study was performed on EEGs from six full term infants having suffered from perinatal asphyxia. The data set is the same as was used in [9]. Each subject contributed with a continuous recording of 6-40 minutes, selected and classified by an experienced electroencephalographer (MT). The length was chosen to include at least 10 bursts, and all artifacts were manually identified and removed from the data. A common average montage with eight electrodes was used, placed according to the international 10-20 system at positions F7, F3, T5, P3, F8, F4, P4 and T6. The signals were digitized at a sampling rate of 200 Hz and band pass filtered between 0.5 and 20 Hz before feature generation. In one case (patient 2), an LMS (least mean square) adaptive filter [10] with a separate ECG channel as reference was used to suppress ECG interference in the EEG signal. The manually classified BS reference data had a time resolution of one second. When the start or end of a burst did not coincide with even seconds, the time was rounded off so that the entire burst was included. This meant that in many cases up to nearly two seconds of suppression activity was included with every burst.

2.2. Features

Five features were selected (Table I), some of which have been used for classification of BS in adult subjects under anesthesia [6, 11] or has shown promising results earlier [12]. The features were extracted from overlapping one-second segments of the EEG signal. The segment length was motivated by the fact that bursts were considered to be at least one second long. Combining this segment length with an overlap of 0.75 s produced an output feature signal with an effective sampling rate of 4

samples per second. All feature signals (except variance) were smoothed by convolution with a six-point triangular window. This was performed because we know that a burst is at least one second long and we want to suppress short spikes or dips in the feature signal to get longer continuous detections. Finally the feature signals were normalized by fitting their distributions between the first and the 99th percentile into the interval zero to one, thus giving all features the same amplitude range. This removes the bias caused by different features having different magnitudes. The proposed normalization method is suitable when the signal is known to contain both burst and suppression activity.

Feature	Description
Spectral Edge Frequency (SEF95) [13]	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on Blackman-Harris windows of the signal. Output signal smoothed by a six-point triangular window.
3 Hz power	Power in a one Hz wide band centered at three Hz. Output signal smoothed by a six-point triangular window.
Median [14]	Median absolute value. Output signal smoothed by a six-point triangular window.
Variance ($s^2(x)$) [14]	$s^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x-\mu)^2$ x is a time series, μ is the sample mean of x.
Shannon entropy (H_{Sh}) [15]	$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$ $p(I_1)\ldots p(I_U)$, p_U is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals $I_1,..,I_U$. 20 intervals were used evenly distributed between +/- two standard deviations of the signal in the window. H_{Sh} is a measure of uncertainty of a random variable. Output signal smoothed by a six-point triangular window.

Table I: Features used for classifying bursts and suppression in bs EEG.

2.3. Performance measures

For performance evaluation, sample sensitivity and specificity were used as measures. Sensitivity was defined as the percentage of the burst samples that were correctly classified as bursts, and specificity as the percentage of the suppression samples that were classified as suppression.

Receiver operating characteristic (ROC) curves were formed for the SVM by taking the raw signal that is proportional to the distance for each sample from the separating hyperplane (this signal is normally thresholded at zero) and sweeping it with a range of thresholds that covered the range of the signal. For the HMM the ROC was generated by sweeping a range of *a priori* probabilities for the burst and suppression states. The ROC curves show the potential of a detector, displaying all possible tradeoffs between sensitivity and specificity. The area under the curve (AUC) provides us with a single number for comparing the classifiers, with the value one representing the perfect classifier. For the HMM this measure is not as accurate as for the SVM, because only partial ROC curves that had to be linearly interpolated could be obtained. The interpolation was performed by adding the extreme values to the ROCs, *i.e.* the coordinates (100, 0) and (0, 100).

2.3.1. Support Vector Machine

The SVM [16] optimizes a nonlinear boundary that maximizes the margin between the classes. It uses (manually) pre-classified training data to generate this boundary, and is therefore a classifier that is trained in a supervised manner. The system was implemented using a freeware Matlab toolbox [17]. As kernel a Gaussian radial basis function (RBF) [18] was used, with a width parameter that has to be chosen. The SVM setup also includes the choice of a weight for misclassified samples, and these two parameters were chosen by performing a grid-search over a suitable area and choosing the combination that produced the best AUC score for all patients. The training was performed using the leave-one-out method; the system was tested on each patient using the other five patients for training.

2.3.2. Hidden Markov Model

The HMM [16] does not require any manual pre-classification of the training data; only the feature data and the number of states are needed as input. The number of states here is two, one for burst and one for suppression. Other parameters such as transition probabilities and emission distributions () are initialized randomly. After initialization, the algorithm works in two steps. First the Baum-Welch algorithm is used to re-estimate the parameters of the model. Then the Viterbi algorithm is used to find the sequence of states that is most likely to have produced the sequence of feature samples in the data. The *a priori* state probability in the performance measure evaluation was incorporated in the re-estimated distributions before decoding the states.

Because the HMM parameters are initialized randomly there is a risk that the training ends up in a local error minimum instead of in the global minimum. To reduce the risk of this, the data was run through the

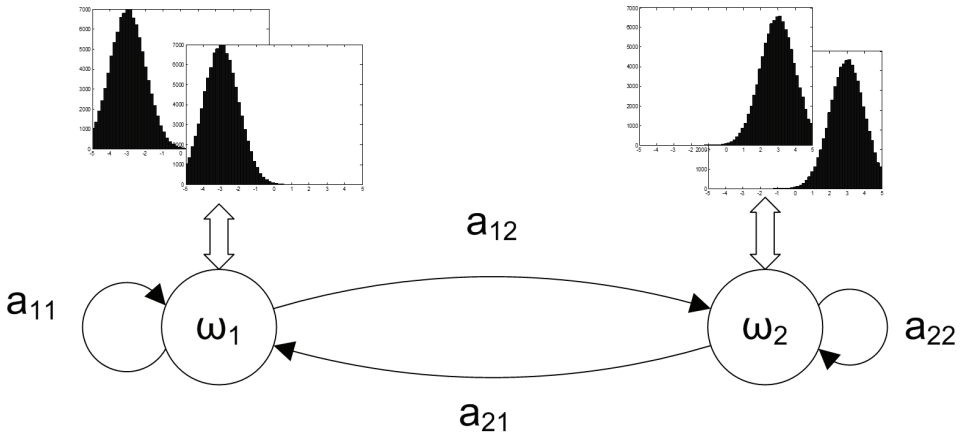


Figure 1: HMM with two states (burst and suppression). Each state is associated with an output vector that is observed, here illustrated by four distributions. The problem is to estimate the model parameters and distributions based on unmarked training data, and then estimate the sequence of states that is most likely to have generated an observed mixture.

algorithm ten times (empirically found to be enough), and the instance with the highest sum of sensitivity and specificity was chosen.

3. Results

Figure 2 shows a comparison of the ROC curves acquired when running the SVM and the HMM on data from the six patients. Each of the six plots is associated with one of the patients, and shows ROC curves for the two classification methods. The ROC for HMM is plotted with dots to illustrate that most points are grouped together.

Figure 3 is an example of an EEG signal for one of the patients, and shows some of the detections for both methods in the time domain.

Table II shows the classification results as AUC values. Because of the linear interpolation of the ROC curves for the HMM, the ROC itself should be taken into account when studying these numbers. The table also shows the sensitivity and specificity achieved for the two methods without sweeping the threshold or *a priori* distributions.

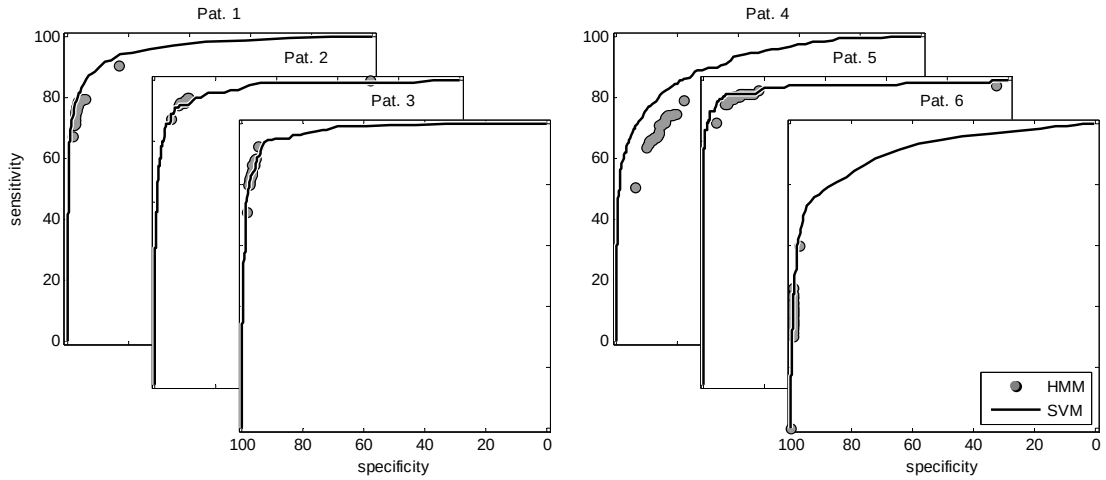


Figure 2: ROC curves for SVM and HMM. The HMM curves are not complete due to limitations in the methods. It should be noted that in five out of six cases the curves from the two methods are virtually on top of each other, even though the HMM uses unsupervised training and SVM uses supervised training. The rings show each point in the HMM ROC and illustrates that most points are closely grouped together.

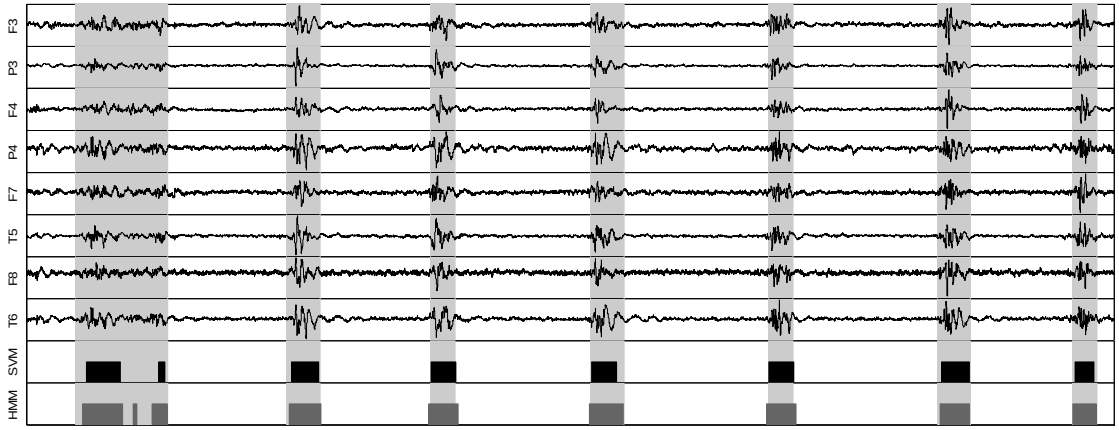


Figure 3: Example of an EEG signal from patient three, with detections made by the SVM and the HMM algorithms, respectively. The sensitivity and specificity tradeoffs are the ones inherent in the classifiers (Table II).

	SVM	HMM*	SVM**		HMM**	
Pat.	AUC	AUC	Sens.	Spec.	Sens.	Spec.
1	0.96	0.92	70	99	76	97
2	0.97	0.95	89	94	93	91
3	0.97	0.95	65	99	81	98
4	0.91	0.82	61	98	69	86
5	0.98	0.95	96	86	94	90
6	0.90	0.79	32	100	38	99

Table II: Classification Results

*** Calculated by linearly interpolating the ROC curve**
**** These numbers refer to the inherent choice of the classifier**

4. Discussion

The results show that given identical input feature data, the HMM classifier produces performance values that in terms of ROC curves are very close to the ones produced by the SVM. The results of the inherent settings show that the HMM generally gets a slightly higher sensitivity at the cost of a lower specificity. A significant difference between the two is that the SVM uses supervised training while the HMM uses unsupervised training. The only input parameter to the HMM is the number of states in the data, all other HMM parameters are estimated automatically by the algorithms.

An earlier study (publication in progress) indicates that the SVM is very good at handling this multidimensional feature data, and generalizes well. The HMM on the other hand does not have to generalize, because it is trained in an unsupervised way on each patient and creates a patient-specific model. Because all burst and suppression properties differ between individuals, an unsupervised classifier that estimates a patient-specific EEG model could be very useful in a monitoring system.

Patient-specific models could also be used for other classification tasks in a monitoring system. Burst-suppression is a type of periodic activity that indicates severe pathology in a neonate. However, tracé alternant is a similar type of periodic activity that is normal during this period of life. By automatically building models of the periodic signals using the HMM algorithms, features that describe the entire signal could be extracted from the models, *i.e.* transition probabilities and the properties of the emission distributions. These features could then be used for classifying the entire signal segment as *e.g.* burst-suppression or tracé alternant.

Conclusion

It has been demonstrated that the HMM classifier is able to accurately distinguish the two types of activity in burst-suppression EEG, without supervised training. The performance of HMM and SVM, as showed in Table II, does not differ significantly, although SVM is slightly more accurate when compared to the gold standard. This finding supports the theory that discrete time signals generated by the EEG can be modeled by a first-order Markov chain, and that the unsupervised HMM can be useful for characterizing periodic EEG.

References

[1] D. B. Sinclair, M. Campbell, P. Byrne, W. Prasertsom, and C. M. T. Robertson, "EEG and long-term outcome of term infants with

neonatal hypoxic-ischemic encephalopathy," *Clinical Neurophysiology,* vol. 110, pp. 655-659, 1999.

[2] C. C. Menache, B. F. D. Bourgeois, and J. J. Volpe, "Prognostic value of neonatal discontinuous EEG," *Pediatric Neurology,* vol. 27, pp. 93-101, 2002.

[3] M. Thordstein, B. G. Wallin, A. Flisberg, R. Bågenholm, I. Kjellmer, N. Löfgren, and K. Lindecrantz, "Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates," *Clinical Neurophysiology*, vol. 115, pp. 2461-2466, 2004.

[4] M. Thordstein, N. Löfgren, A. Flisberg, R. Bågenholm, K. Lindecrantz, and I. Kjellmer, "Infraslow EEG activity in burst periods from post asphyctic full term neonates," *Clinical Neurophysiology,* vol. 116, pp. 1501-1506, 2005.

[5] M. Särkelä, S. Mustola, T. Seppänen, M. Koskinen, P. Lepola, K. Suominen, T. Juvonen, H. Tolvanen-Laakso, and V. Jäntti, "Automatic analysis and monitoring of burst suppression in anesthesia," *Journal of Clinical Monitoring and Computing,* vol. 17, pp. 125-134, 2002.

[6] L. Leistritz, H. Jäger, C. Schelenz, H. Witte, P. Putsche, M. Specht, and K. Reinhart, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation," *Journal of Clinical Monitoring and Computing*, vol. 15, pp. 357-367, 1999.

[7] G. Hauksson, "Automated Analysis of Newborn EEG," in *Electrical Engineering*. vol. Master of Science Stockholm: Royal Institute of Technology, 2006, p. 93.

[8] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford: Oxford University Press, 2005.

[9] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Comparison of three methods for classifying burst and suppression in the EEG of post asphyctic newborns," in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007.

[10] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: John Wiley & Sons, Inc., 1996.

[11] C. Jeleazcov, S. Egner, F. Bremer, and H. Schwilden, "Automated EEG preprocessing during anaesthesia: New aspects using artificial neural networks," *Biomedizinische Technik,* vol. 49, pp. 125-131, 2004.

[12] J. Löfhede, N. Löfgren, M. Thordstein, A. Flisberg, I. Kjellmer, and K. Lindecrantz, "Detection of Bursts in the EEG of Post Asphyctic Newborns," in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 2179-2182.

[13] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *Anesthesiology,* vol. 89, pp. 980-1002, 1998.
[14] J. A. Rice, *Mathematical Statistics and Data Analysis,* 2nd ed. Belmont: Duxbury Press, 1995.
[15] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* vol. 27, pp. 379-423, 623-656, 1948.
[16] R. O. Duda, Hart, P. E., Stork, D. G., *Pattern Classification,* 2nd ed. New York: John Wiley & Sons, Inc., 2001.
[17] A. Schwaighofer, "SVM toolbox for Matlab," 2.51 ed Berlin: Frauenhofer FIRST, Intelligent Data Analysis, 2002.
[18] S. R. Gunn, "Support Vector Machines for Classification and Regression," Faculty of Engineering, Science and Mathematics, School of Electronics and Computer Science, University of Southampton1998.

Paper V

Automatic classification of background EEG activity in healthy and sick neonates

Johan Löfhede, Magnus Thordstein, Nils Löfgren, Anders Flisberg, Manuel Rosa-Zurera, Ingemar Kjellmer and Kaj Lindecrantz

Manuscript submitted to the Journal of Neural Engineering.

Automatic classification of background EEG activity in healthy and sick neonates

Johan Löfhede[1,2], Magnus Thordstein[3], Nils Löfgren[4], Anders Flisberg[5], Manuel Rosa-Zurera[6], Ingemar Kjellmer[5] and Kaj Lindecrantz[1]

[1] School of Engineering, University College of Borås, Borås, Sweden.

[2] Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden

[3] Institute of Neuroscience and Physiology, Section for Neuroscience and Rehabilitation, Sahlgrenska University Hospital, Göteborg, Sweden.

[4] Neoventa Medical AB, Göteborg, Sweden.

[5] Department of Pediatrics, Göteborg University, The Queen Silvia Children's Hospital, Göteborg, Sweden.

[6] Department of Signal Theory and Communications, University of Alcalá, Alcalá de Henares, Spain.

E-mail: johan.lofhede@hb.se

Abstract

The overall aim of our research is to develop methods for a monitoring system to be used at neonatal intensive care units. When monitoring a baby, a range of different types of background activity needs to be considered. In this work, we have developed a scheme for automatic classification of background EEG activity in newborn babies. EEG from six full term babies who were displaying a burst suppression pattern while suffering from the after-effects of asphyxia during birth was included along with EEG from 20 full term healthy newborn babies. The signals from the healthy babies were divided into four behavioural states: active awake, quiet awake, active sleep and quiet sleep. By using a number of features extracted from the EEG together with the Fisher's linear discriminant classifier we have managed to achieve 100% correct classification when separating burst suppression EEG from all four healthy EEG types and 93% true positive classification when separating quiet sleep from the other types. The other three sleep stages could not be classified.

When the pathological burst suppression pattern was detected, the analysis was taken one step further and the signal was segmented into burst and suppression, allowing clinically relevant parameters such as suppression

length and burst suppression ratio to be calculated. The segmentation of the burst suppression EEG works well, with a probability of error around 4 %.

1. Introduction

When a baby is being born there is a risk for complications that reduce the blood gas exchange to the fetus during delivery. This can lead to hypoxia, acidemia and potentially cause brain damage. When complications are expected, the baby is monitored closely during the delivery and is then, if necessary, transferred to a neonatal intensive care unit (NICU). It is however difficult to determine at an early stage if the baby's brain is recovering, if brain damage is still a threat or if it already has occurred [1]. At the NICU parameters such as heart rate, blood pressure, oxygen saturation and other measures are monitored regularly [2], but these parameters only assess the general condition of the baby. To monitor the function of the brain itself, a more direct way is to measure the electroencephalogram (EEG); a measure of the electrical signals produced by the brain. This makes possible continuous monitoring of the brain over long time with high time resolution.

In earlier studies [3-5], various parameters have been used for the detection of hypoxia in the brain and for the prediction of outcome after a hypoxic event. Concerning the EEG, clinical neurophysiologists evaluate most recordings through visual inspection of the unprocessed signal. This service is however usually not available at the NICU. Even though methods for remote consultations have been developed [6], as a rule, only evaluations at certain distinct time instances are possible. This way of analysing data is of course suboptimal when it comes to continuous bed side monitoring.

An existing method for simplified long-time monitoring of brain function is amplitude-integrated EEG (aEEG), for example implemented as the Cerebral Function Monitor (CFM) [7]. This method filters the EEG in a certain way and then displays it on a compressed time scale, giving the clinician a simple way to monitor the brain activity of a patient. There have however been reports of missed seizure activity [8, 9], and the compressed signal can mask interference and artefacts and cause them to be mistaken for brain activity. Because of these drawbacks, neurophysiologists argue that the unprocessed EEG signal has to be taken into consideration when interpreting the aEEG [8].

Burst suppression (BS) is one type of activity that can be seen in babies after asphyxia, and it indicates that the baby is in a serious condition. In earlier work [10] we have demonstrated a scheme for segmenting an EEG signal into periods of burst and suppression. This scheme assumes that the

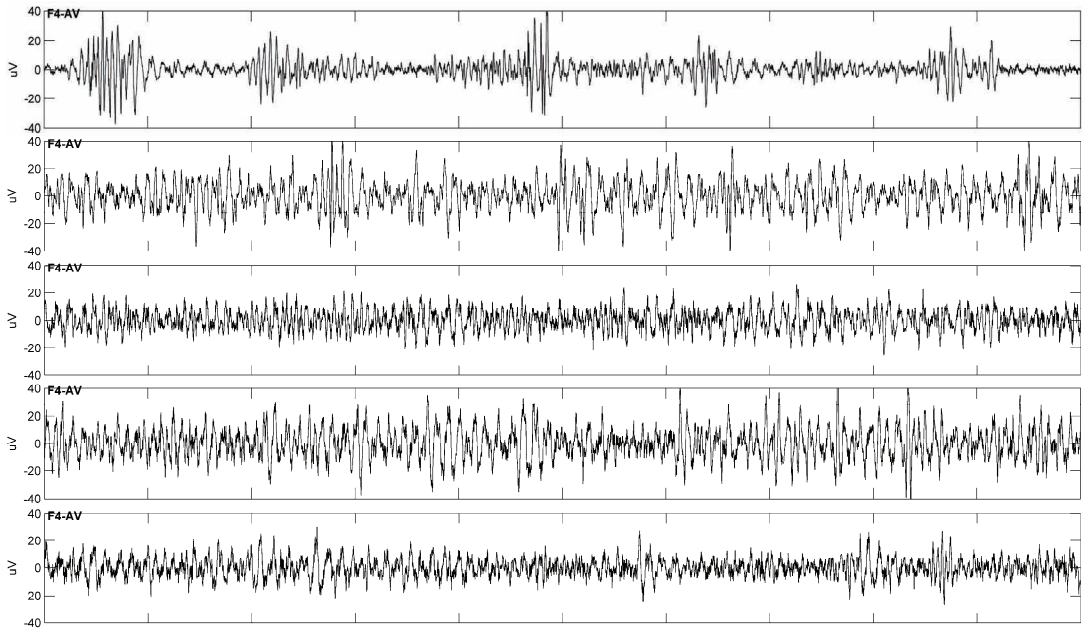


Figure 1. Examples of EEG data from the five classes. The classes are from the top: BS, QS, QW, AS and AW. 50 s are displayed and the amplitude scale is +/- 40 µV. The signals were band-pass filtered between 1.6 and 44 Hz and the average of all channels was subtracted from each channel, as in the rest of the paper. Note that only channel F4 from each patient is displayed here, while eight channels from each subject was used for extracting the features used for classification.

EEG is known to contain a BS pattern. As a next step towards the development of a complete system for automatic classification of neonatal EEG we have now developed a scheme for determination if an EEG is of BS type. The BS detector was tested on EEG known to contain BS activity, and to check for any false positives it was validated against normal EEG. In this study we have chosen to divide the normal data, acquired from healthy newborn babies, according to the behaviour of the baby as described by Prechtl [11]. The first four states in this scale correspond to quiet sleep (QS), active sleep (AS), quiet awake (QA) and active awake (AW). The fifth state, crying, has not been included. The behavioural characteristics that the Prechtl scale is defined by can not be observed [4] during BS. Quiet sleep EEG can have some resemblance with BS when it has a periodic form, the so called *tracé alternant* (TA) pattern whereas the EEG during the other states is less periodic. Examples of signals from the five classes are displayed in figure 1.

When the high level (long time scale) classification in the scheme encounters BS in the EEG, the signal is processed further on a lower level (short time scale) by segmentation into burst and suppression, using methods similar to what we describe in [10]. However, the choice of features has been optimized by adding feature selection using genetic algorithms, and the feature set has been expanded from five features to 22. The segments are then used for extracting the suppression length and the

burst suppression ratio, two parameters that may give prognostic information about the brain [12].

Most publications on BS classification address the problem of segmenting adult EEG recorded during deep anaesthesia, while only a few deal with neonatal BS induced by hypoxia. The publications regarding burst suppression detection or classification usually deal with what we in this paper call "segmentation", i.e. given a signal known to be burst suppression, classify it into burst and suppression, which also is the topic in [10].

The scheme was developed using data from full term neonates: 20 healthy patients and six sick ones that had suffered perinatal asphyxia, expressing a BS pattern in their EEG. The scheme was also tested on long continuous recordings, one from each group.

2. Methods

2.1. Data

All data were collected using the SACS® [13] system, using sampling frequencies of 200 Hz (post-asphyctic patients) or 256 Hz (healthy babies). The data sampled at the higher frequency were downsampled to 200 Hz to make sure that all data were comparable. An experienced electroencephalographer (MT) then chose suitable periods of EEG and exported them to the Matlab environment, where all subsequent processing was performed.

Eight channels were used, with electrodes placed according to the international 10-20 system [14] at positions F7, F3, T5, P3, F8, F4, P4 and T6. Disturbances were reduced by using the CAR (common average reference) montage, meaning that the mean of all channels was subtracted from each channel. Moreover, the data were notch filtered at 50 Hz to reduce power line interference, and then band pass filtered between 1.6 and 44 Hz to remove low- and high frequency noise. These limits were chosen especially because a large amount of energy was present in the low frequencies, possibly due to ventilator interference. Although we realize that the normal neonatal EEG contains a large proportion of its energy at very low frequencies [15, 16], the present recordings were often obtained when the baby was artificially ventilated at frequencies around 0.5 Hz and in the most suppressed BS periods 50 Hz interference required heavy filtering to be sufficiently damped. These filter limits were chosen because they were found to work best with the classification and segmentation methods. When the classification and segmentation is done, the segment limits can be used together with the original signal for extracting low frequency information.

2.1.1. Data from healthy babies

The first group of subjects consisted of 20 full-term healthy newborn babies who were recorded after uneventful deliveries. Data were collected during a few hours so that the four behavioural states could be included in most of the registrations. Then epochs of median length of 1.4 to 2.3 minutes for each state were chosen. These data were divided into the four states based on the observations made by the technician performing the recording and on the classification made by an experienced electroencephalographer (MT). For quiet sleep, EEG of the TA type was chosen.

For evaluation of the method in a situation close to the clinical setting, one of the recordings (2 h) was processed by the system as it was, without selecting any particular periods.

2.1.2. Data from post-asphyctic babies

The second group consisted of EEG recordings from six full-term newborn infants having suffered from perinatal asphyxia (impaired gas transfer and circulation around the time of birth). These babies all exhibited a severe BS pattern in their EEG. Each subject contributed a continuous recording of 6-40 minutes, selected and visually segmented into burst and suppression segments by the electroencephalographer. The length was chosen to include at least 10 bursts, and all artefacts were visually identified and marked for later exclusion from the analysis. The total amount of data in this category was 77 minutes, before artefact removal and equalization of the number of samples from the different patients.

For evaluation of the BS-related methods in a setting as close as possible to the clinical one, a 32 hour recording from one of the six babies was used. The baby had to be resuscitated after birth and was then intubated and put on a ventilator. The EEG recording was started six hours after birth and continued for 32 hours with a short break around 18 hours after start. The ventilator frequency was set to 40/min (0.7 Hz) initially and was changed to 30/min (0.5 Hz) 16 hours after start. At 18 hours after the start of the recording, a dose of Phenobarbital was given to treat seizure activity. The baby was later diagnosed with cerebral palsy. This recording was not manually segmented because of its length.

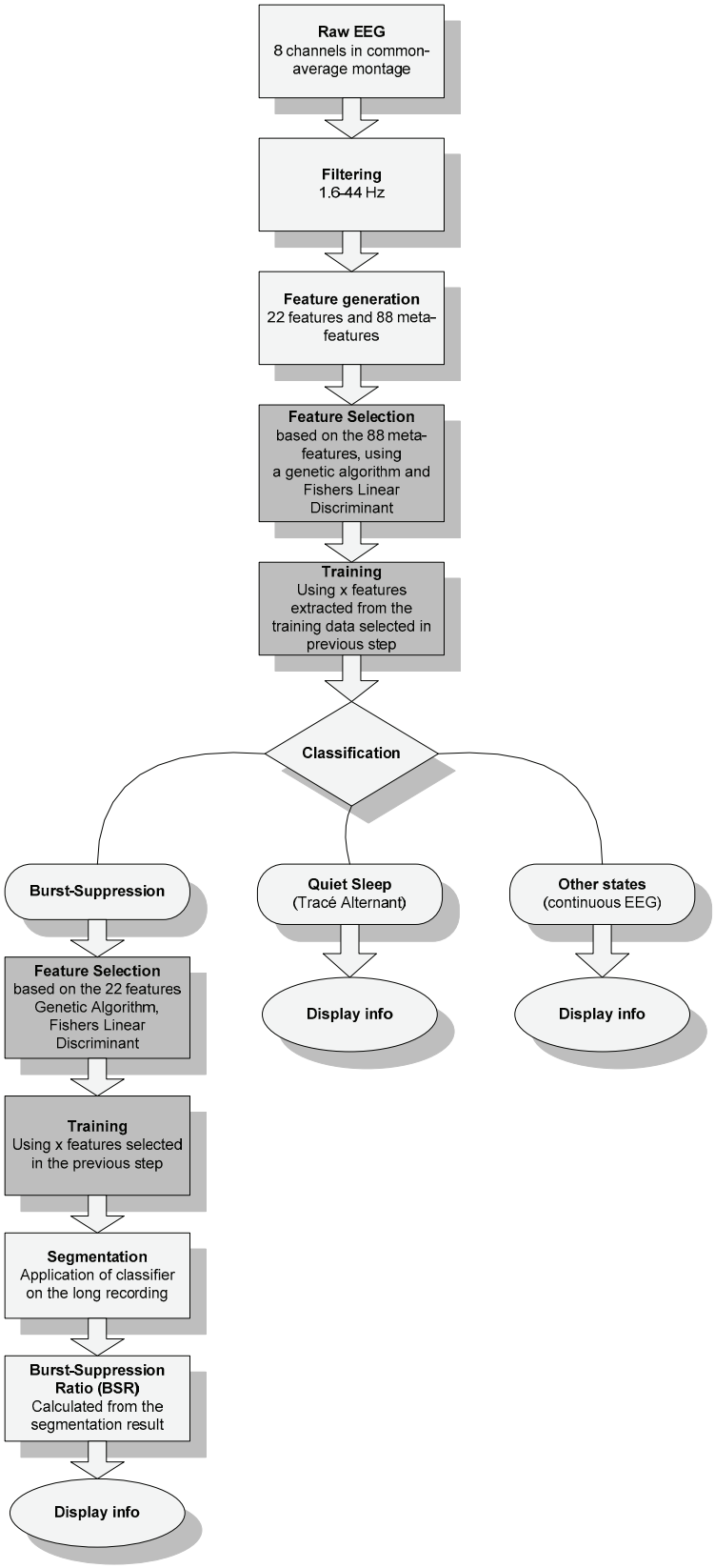


Figure 2. Flowchart describing the process from the raw EEG to classification into different states and segmentation of BS. The darker boxes represent steps that are only carried out during training.

2.2. System overview

The analyzing scheme includes many processing steps, from pre-processing of the raw signal to the final result. The first step is running the raw EEG through a band-pass filter described above, and then a number of features are extracted and fed to the classifier that separates BS from normal EEG. When BS is detected, another classifier is used for segmenting the EEG into burst and suppression. The segmented output can be displayed as *e.g.* the burst-suppression ratio (BSR), or used for further processing. All the different steps are included in the flowchart in figure 2. Note that some of the steps in the flowchart are only made during training or development of the system, and will not be included in a future system for clinical use.

2.3. Features

Features represent certain characteristics of the underlying signal, and should in the classification case enhance the characteristics that differ between the classes. When multiple features are used they should complement each other, and be as independent as possible. The 22 features that were used in this work are listed in table 1. These features include measures of power, frequency distribution and entropy and were chosen so that they should represent a broad spectrum of different characteristics of the signal.

A sliding window with a length of one second, corresponding to 200 samples, was used, with an overlap of 0.75 seconds. This produces feature signals with an effective sample rate of 4 Hz. The choice of window length is a trade-off between time resolution and accuracy.

Because both bursts and suppressions on one hand and the characteristics of different behavioural states on the other are largely global phenomena, the activity that is typical for them should appear in a majority of the channels. This criterion was implemented by taking the median over the eight channels after the features were generated. This enhances the common activity, *e.g.* the bursts, while attenuating activity that appear in less than half of the channels, *e.g.* some types of seizures and artefacts.

For the BS/sleep stage classifier, the features were summarized by applying the four statistical measures mean, variance, skewness and kurtosis on the feature signals from each 30 s non-overlapping epoch in the data, resulting in 88 features based on the original 22. This was done because sleep-stages and BS goes on for some time, at least a couple of minutes, while *e.g.* a single burst can be as short as one second. These measures describe different properties of the feature signals distributions. For example, the mean and the variance are measures of where the

distribution is located and how wide it is, while skewness and kurtosis measure its shape. These measures are in this paper called *metafeatures*, because they are features of the features of the EEG.

Table 1. Features used for classification and segmentation.

Feature	Description
Spectral Edge Frequency (SEF95)	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on rectangular windows of the signal.
3 Hz power	Power in a 1-Hz wide band centred at 3 Hz.
Median	Median absolute value.
Shannon entropy (H_{Sh})	The Shannon entropy is a measure of uncertainty of a random variable and is calculated using the equation $H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$ where $p(I_1)\ldots p(I_U)$ is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals $I_1,..,I_U$. 20 intervals were used evenly distributed between the maximum and minimum values of the signal in the window.
Zero crossings	The number of zero crossings in each window.
Variance ($s^2(x)$)	The variance is calculated using the equation: $s^2(x) = \frac{1}{N-1}\sum_{n=1}^{N}(x[n]-\mu)^2$ where x is a time series, and μ is the sample mean of x.
Spectral centroid	The "centre of mass" of the frequency distribution: $c = \left(\sum_{n=1}^{N} nX[n]\right)\left(\sum_{n=1}^{N} X[n]\right)^{-1}$ where X is the Fourier transform of the signal and N is the number of samples in the window.
Residual energy variance	The variance of the residual after applying an eight-coefficient linear predictor on the signal window.

Spectral flux	Measures the change in the spectrum between consecutive windows using the distance between the spectra.
deltaF	Measures the rate of change by taking the Euclidian distance between consecutive windows of the signal.
Spectral flatness	Calculated by dividing the geometric mean of the power spectrum with the arithmetic mean: $F = \left(\prod_{n=1}^{N} x[n] \right)^{1/N} \left(\frac{1}{N} \sum_{n=1}^{N} x[n] \right)^{-1}$ A high spectral flatness indicates that the spectrum has a similar amount of power in all bands – like white noise. A low spectral flatness indicates a spiky spectrum.
Spectral roll-off	Equivalent to SEF85.
Cepstral coefficients	Ten cepstral coefficients were extracted by applying the Fourier transform on a window of the signal, mapping the logarithm of the spectrum onto a linear scale using ten triangular overlapping windows, and then taking the discrete cosine transform of the list of amplitudes. The cepstral coefficients are the amplitudes of the resulting spectrum, and are related to the rate of change in the corresponding frequency bands.

2.4. Feature selection

For the segmentation, the 22 feature functions listed in table 1 have been available, and for the classification the four metafeature functions were used, resulting in a total number of 88 metafeatures. In order to select the optimal set of features, a *genetic algorithm* (GA) has been used [17], with the number of features limited using the method described in [18, 19]. The algorithm searches the space of possible combinations using methods inspired by natural selection and can usually find a combination that is close in performance to the optimal one. The advantage is that it is much quicker than an exhaustive search because only a subset of all possible combinations is tested. The algorithm was used to try combinations of one to ten features selected from the total set of 22 features or 88 metafeatures.

The number of feature data points from the different patients was equalized by drawing a number of samples from each patient that was equal to the median number of samples available. In the cases when the available number of samples was lower than the median, samples were drawn with replacement. The data from the different patients were then concatenated into one matrix, and the sample order was randomized. The resulting matrix was divided into three parts; one half of the data was used for training, one quarter for testing and one quarter for validation (see table 2 for data set sizes). The training set is used for calculating the projection matrix and the testing set is used for selecting the best one. This is repeated for many combinations using the genetic algorithm. When all selection is finished, the resulting classifier is tested on the validation set, which has not been involved in the process. If the result from the validation set is much worse than on the testing set, the classifier has probably been overfitted during training, and will not be able to generalize to unseen data.

Because the result of the GA may depend on the randomly initiated starting set, the algorithm was run five times for each number of included features, and the one with the lowest error on the test set was included in the evaluation. The observed differences in probability of error between different runs were less than one percent.

	Training	Test	Validation
Classification	4848	2424	2424
Segmentation	208	96	96

Table 2. Dataset sizes. The values are the number of feature samples used for training.

2.5. Classifier

Fisher's linear discriminant (FLD) was used for both classification and segmentation. It has good performance when used with linearly separable features, and is very quick compared to more advanced classifiers as *e.g.* support vector machines or neural networks. When using genetic algorithms it is unwise to use more advanced classifiers since the classifier is trained and evaluated thousands of times using different combinations of features and the processing time would make it impractical.

FLD is based on scatter matrices [20, 21] formed from the training data from the two classes. These are used for deriving the projection of the multidimensional input space onto the line that gives the maximum ratio of between-class scatter to within-class scatter. This projection is then applied on the test data, resulting in a mapping from the five-dimensional

feature signal to a one-dimensional decision signal that is thresholded at zero.

In the classification case, after the feature signals were combined by using the FLD classifier, the resulting (non-thresholded) classification signals were smoothed by convolving with a ten sample rectangular window. The final classification was then achieved by selecting the largest of the five classification signals for each time instance.

2.6. Performance measure

The performance was measured in terms of the probability of error calculated as

$$P_{err}=P(\text{class 1})P(\text{class 2}|\text{class 1})+P(\text{class 2})P(\text{class 1}|\text{class 2})$$

where P(class 1) is the probability of class one and P(class 1|class 2) is the probability of classifying a given sample as class 1 when it does in fact belong to class 2, and the other way around. The formula gives a weighted sum of the two misclassification probabilities, where the weights are the proportions of the two classes in the data.

The performance is also presented as a confusion matrix, where the probabilities for misclassifying each of the classes can be found. The confusion matrix entries are defined in table 3.

Table 3. Definition of the confusion matrix entries.

		True class				
		BS	QS	QW	AS	AW
Predicted class	BS	P(BS|BS	P(BS|QS	P(BS|QW	P(BS|AS	P(BS|AW
	QS	P(QS|BS	P(QS|QS	P(QS|QW	P(QS|AS	P(QS|AW
	Q	P(QW|BS	P(QW|QS	P(QW|QW	P(QW|AS	P(QW|AW
	AS	P(AS|BS	P(AS|QS	P(AS|QW	P(AS|AS	P(AS|AW
	A	P(AW|BS	P(AW|QS	P(AW|QW	P(AW|AS	P(AW|AW

For example, P(BS|BS) is the probability of classifying a sample as BS when the true class really is BS, and P(BS|QS) is the probability for classifying a sample as BS when the true class is QS.

2.7. Burst suppression segmentation

When the first classifier in the system detects BS, the data is passed along to a second classifier that works as a segmenter. This segmenter is also an instance of a FLD, but is trained on pre-segmented bursts and suppressions. Even though it was found in [10] that the support vector machine classifier was better than the FLD at segmenting BS the

difference was not very large, and the larger number of features and the feature selection process used in this case made the FLD a better choice. The segmenter works with the 22 base features that have a sampling rate of 4 Hz, giving the segmentation a time resolution of 0.25 s.

The trained segmenter was applied on the parts of the 32 h recording that had been classified as BS, and the segmented signal was then used for extracting information about the underlying EEG and the BS pattern. Parameters such as the length of suppression and the burst-suppression ratio that are commonly used to characterize BS were calculated.

3. Results

The results are grouped into results from the two main parts of the scheme: the classification into the five classes (BS and normal states) and the results of segmenting BS signals and extracting information from the segments. The classification is tested on both a long (2 h) recording of a healthy baby and a long (32 h) recording of a sick baby.

3.1. State classification

Figure 3 shows the probability of error for the different numbers of features used in the feature selection process. The error is presented for both the test set, which is used during the selection process to choose the best candidate selections, and for the validation set, which is only used when the training and selection process is finished. The selected feature combinations are displayed in table 4.

Figure 4 shows the distributions (estimated by histograms) of the output of the Fisher projection for each class. The classification is acquired by thresholding this signal at zero. If the distributions are completely separated as in the BS case, perfect classification is possible on this specific validation set. The confusion matrix in table 5 shows the probabilities for correct classification and for the different types of misclassification.

Figure 5 shows a time-frequency plot of the two hour long recording from a healthy baby used for testing the final sleep stage/BS classifier. The lower plot shows the classification done by the technician who monitored the recording. During the parts of the recording that are not marked as a behavioural state the baby is crying, marked by high power strips in the time frequency plot that mainly are due to muscle artefacts.

Figure 6 shows the output of the five classifiers when running the long sleep stage recording through it. The BS classification is perfect, with no epochs erroneously classified as BS. The QS classifier does fairly well during the first half of the recording, but misses a large part of the second

QS period. The QW and AS classifiers also have low precision during the second half of the recording, and the AW classifier even more so.

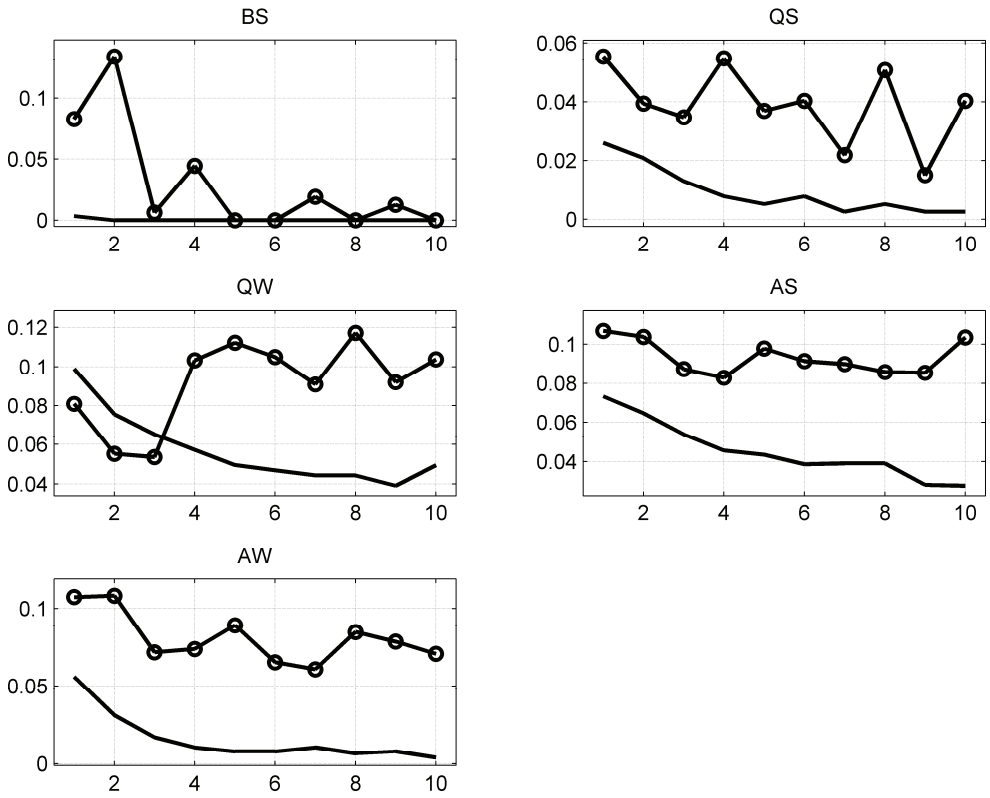


Figure 3. Probability of error as function of the number of features included in the training. The lines without circles show the error on the test set and the lines with circles show the error on the validation set.

Burst Suppression
median mean
entropy variance
residual energy variance skewness
spectral roll-off kurtosis
cepstral coeff 9 mean

Quiet Sleep
SEF95 kurtosis
median variance
variance skewness
centroid kurtosis
spectral flux skewness
deltaF skewness
deltaF kurtosis
spectral flatness mean
cepstral coeff 7 mean

Quiet Wake
SEF95 skewness
SEF95 kurtosis
entropy variance

Active Sleep
median skewness
centroid mean
centroid skewness
cepstral coeff 10 mean

Active Wake
SEF95 skewness
median mean
median skewness
spectral flatness variance
spectral roll-off mean
cepstral coeff 7 mean
cepstral coeff 7 kurtosis

Table 4. Features selected for the five classes.

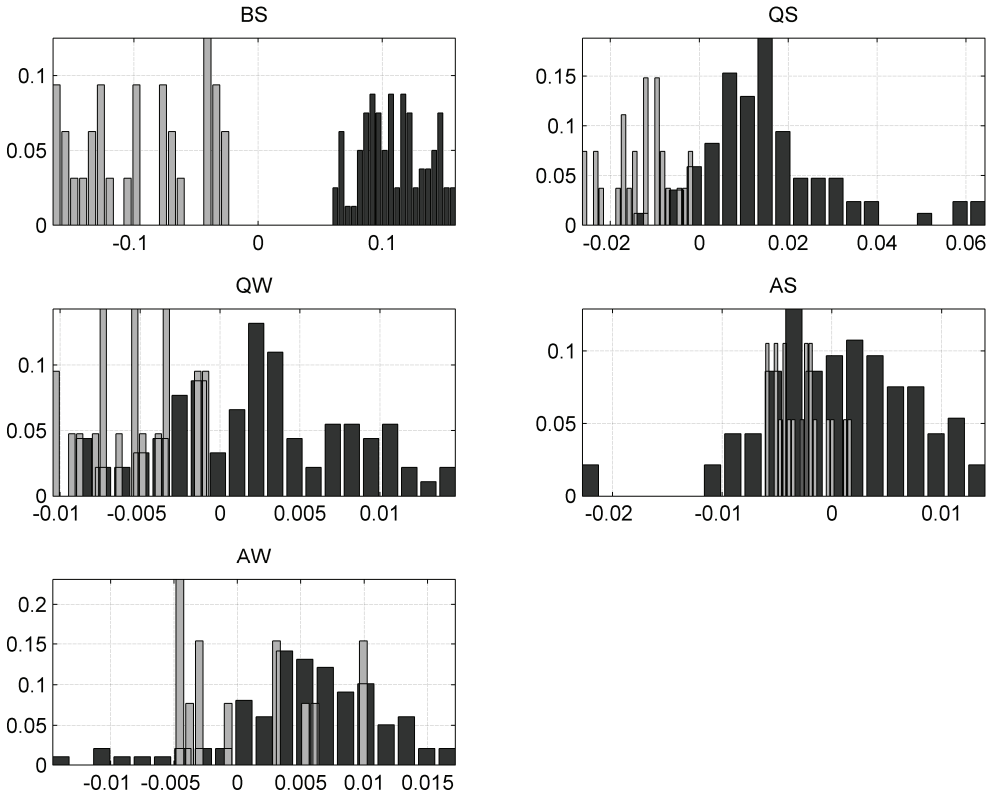


Figure 4. Histograms showing the distributions of the output from the Fisher projection for the validation data. The solid lines are the single classes and the dashed lines are the combination of the other classes for each binary classification case.

		true values				
		BS	QS	QW	AS	AW
predicted values	BS	100	0.0	0.0	0.0	0.0
	QS	0.0	92.6	0.0	5.3	7.7
	QW	0.0	3.7	47.6	47.4	38.4
	AS	0.0	3.7	28.6	42.1	30.8
	AW	0.0	0.0	23.8	5.2	23.1

Table 5. Confusion matrix for the BS/sleep stage classification. The diagonal elements are the probabilities for correct classification and the other elements the probabilities of confusing the different classes.

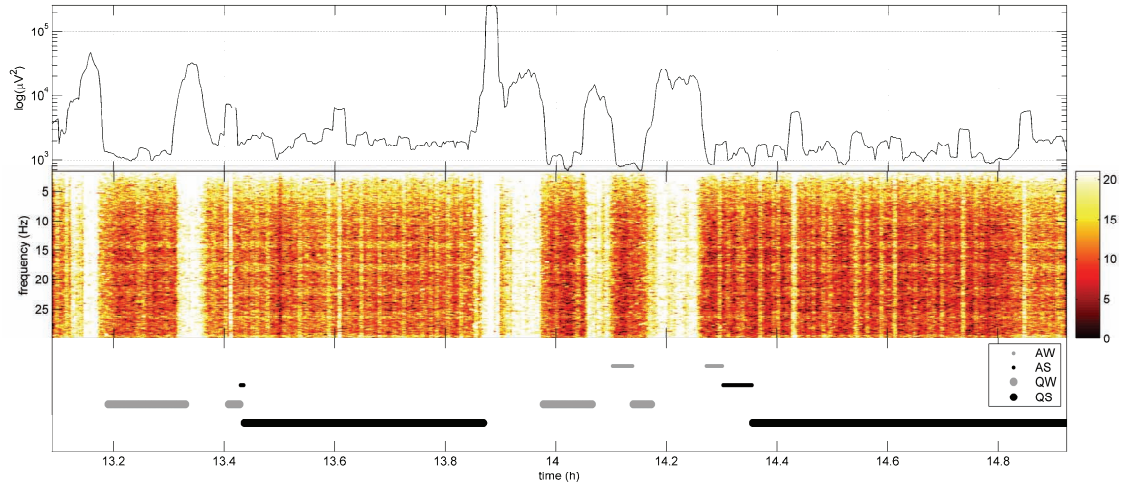


Figure 5. Average power and time-frequency plot of the long sleep stage recording of a healthy baby. The colours show the signal power in decibel, with the lowest power (dark red) as reference. The markings below the figure shows the different sleep stages, marked by the technician doing the recording.

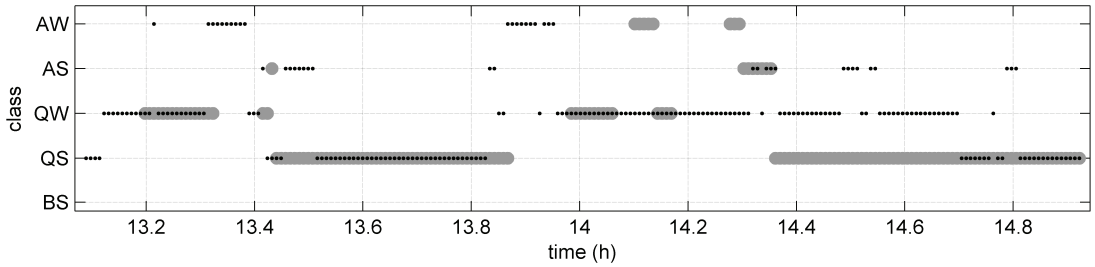


Figure 6. Result of running the long sleep stage recording through the classifier. The black dots represent the classification of 30 s epochs, and the grey thick lines show the reference classification done by the technician doing the recording.

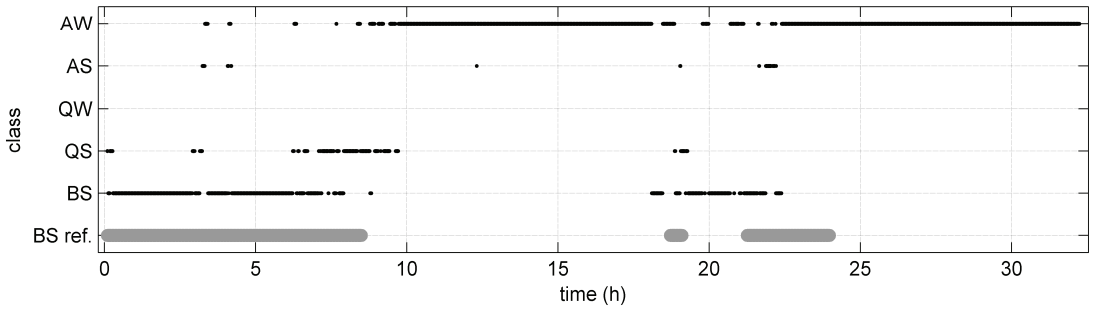


Figure 7. Result of running the 32 h hypoxia-registration through the BS/sleep stage classifier. Each dot represents the classification of a 30 s epoch. The grey thick lines are classified as BS periods by the electroencephalographer.

3.2. Burst suppression segmentation

Figure 8 shows the results from the feature selection process where the manually pre-classified training and validation sets were used. The evolution of the probability of error is shown when increasing the number of included features from one to ten, for both the test data and the validation data. In the figure the probability of error on the validation set

seems lower than the probability of error on the test set. Theoretically, the opposite is expected, but the size of the error is essentially the same. The scale emphasizes the small random difference that in this case happens to be in opposition to the theory for large sets of data. The limited size of the current data set caused the results to vary slightly depending on how the data was distributed in the training, testing and validation sets, but the error range did not vary significantly. Table 6 shows the selected features, and figure 10 gives an example of the segmented EEG.

Figure 9 contains a smoothed graph of the mean power and a time-frequency plot that gives an overview of the 32 h recording from one of the babies with burst suppression. In the plot it is clearly visible that the power of the signal is very low in the beginning, and then gradually increases. Half way through the recording there is a relapse into lower power before the final stabilization at a more normal level. The time-frequency plot also shows the frequency distribution of power which is relevant because many of the features considered measure different aspects of the frequency distribution.

Figure 11 and 12 show the results of applying the automatic classification and segmentation methods on the long 32 h recording. The estimated length of each suppression segment and the estimated burst-suppression ratio for 30 s windows are displayed. For the first BS episode the BSR is low, corresponding to a small amount of bursts compared to the amount of suppression. During the later BS periods the BSR is higher.

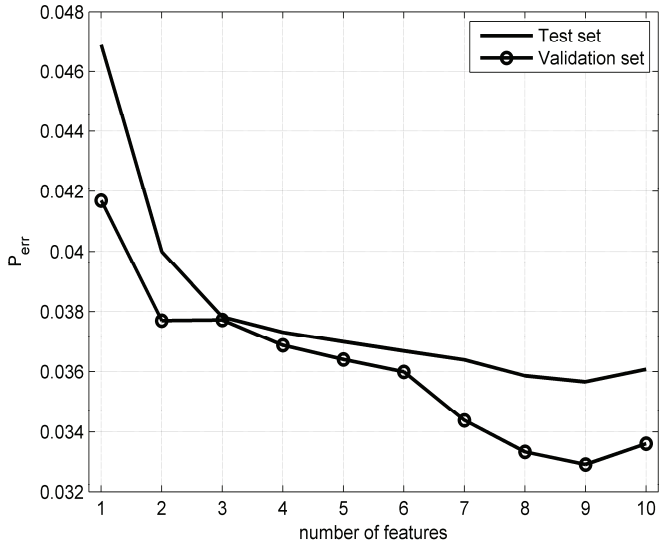


Figure 8. Probability of error as a function of the number of features included in the training.

Table 6. Features selected for BS segmentation by the genetic algorithm.

Feature name
3 Hz power
Variance
Centroid
Residual energy
deltaF
Spectral flatness
Cepstral coeff 2
Cepstral coeff 3
Cepstral coeff 10

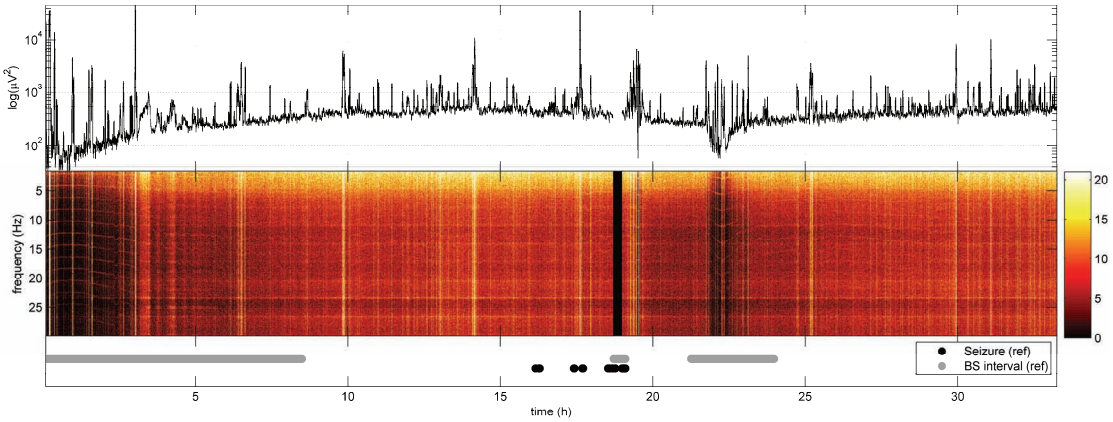


Figure 9. Average power and time-frequency plot for the long burst suppression recording. The dark blue strip around 18 hours is a gap in the recording. The colour shows the signal power in decibel, with the lowest power (dark red) as reference. The lower subplot shows visual classifications made by an electroencephalographer.

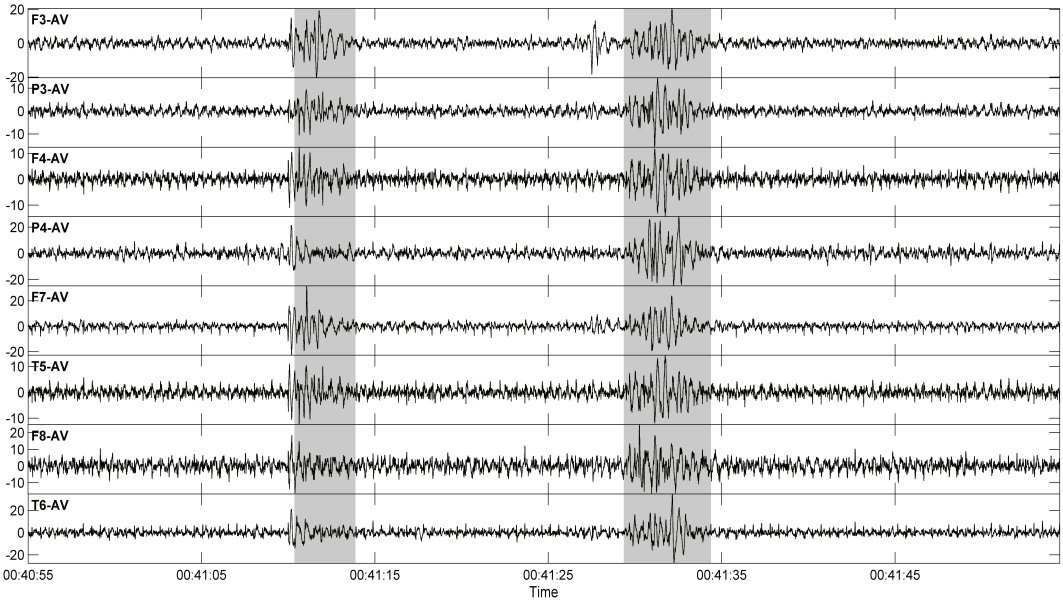


Figure 10. A 60 s example of segmented EEG.

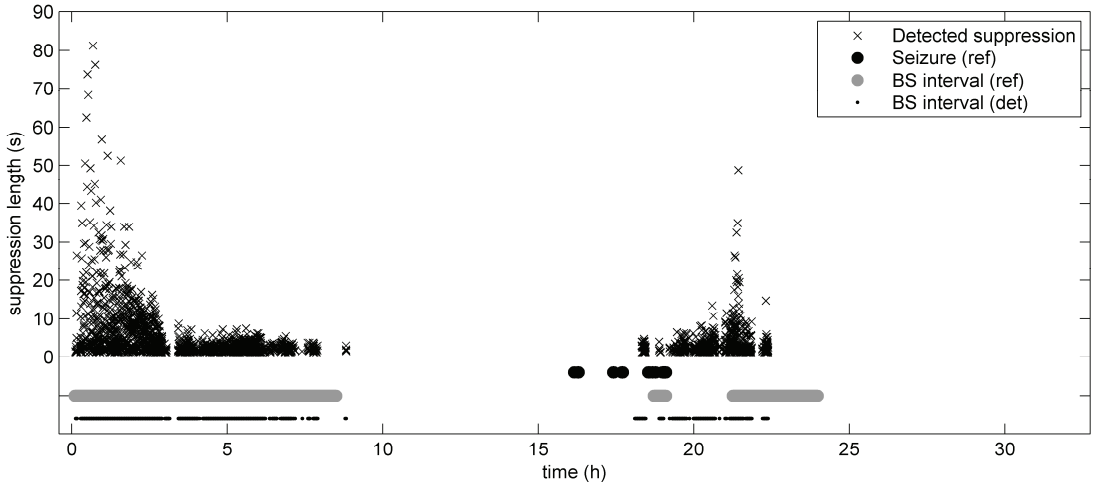


Figure 11. The x-marks show the estimated lengths of the suppression for the BS episodes that were detected by the method. The large black and grey dots show the seizure and BS episodes that were found by the human expert. The small black dots show the BS episodes that were detected by the method. The x-coordinate shows the position in time and the y-coordinate the length of the interval. Seven hours after the start, the BS pattern is gradually replaced by TA. After 18 h there is a 20 min gap in the record, and after the gap there is some seizure activity which after treatment turns into a short BS episode followed by a longer one that starts at 21:15 and gradually stops before 24 h.

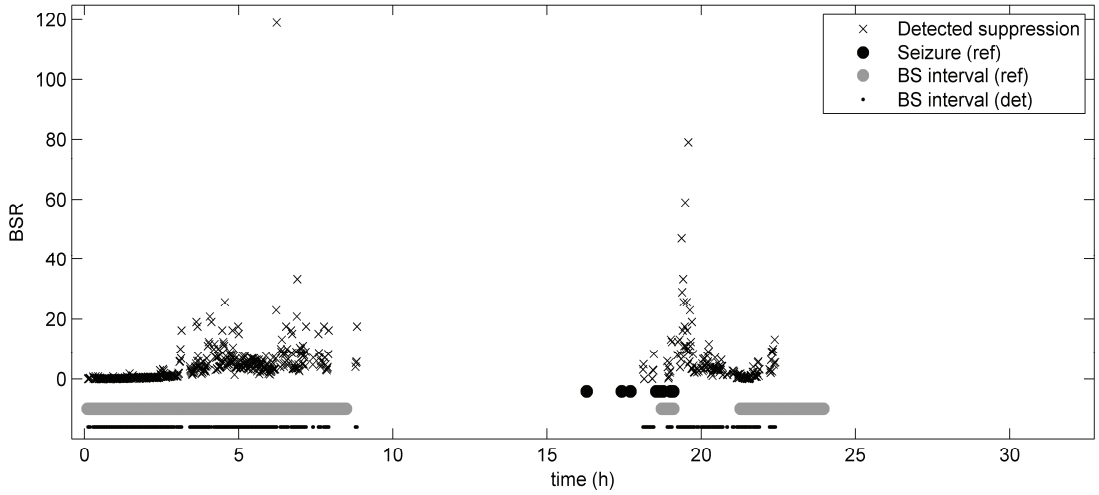


Figure 12. The x-marks show the estimated burst suppression ratio for the BS episodes that were detected by the method. The large black and grey dots show the seizure and BS episodes that were found by the human expert. The small black dots show the BS episodes that were detected by the method. The x-coordinate shows the position in time and the y-coordinate the length of the interval. Seven hours after the start, the BS pattern is gradually replaced by TA. After 18 h there is a 20 min gap in the record, and after the gap there is some seizure activity which after treatment turns into a short BS episode followed by a longer one that starts at 21:15 and gradually stops before 24 h.

4. Discussion

In this paper we have described a possible scheme for automatic detection and quantification of BS patterns and the performance of the scheme if presented with EEG known to contain BS as well as normal EEG during different behavioural states. The scheme does well in distinguishing BS from the normal EEGs. In the case of BS, the EEG is also segmented into burst and suppression, and the BSR and the suppression lengths are calculated.

The classification of BS from post asphyctic patients when compared to EEGs from the healthy babies is 100 % accurate. This is not too surprising, since the BS pattern differs from the likewise periodic TA pattern by its interburst activity having much lower amplitude [4]. The classification of quiet sleep, containing TA activity, is also highly accurate with a true positive classification rate of 93 %. This can be compared to recent works where true positive classification rates of 89.2 % have been achieved when using the behavioural state classification of quiet sleep [22] and 80-90 % when using the definition that TA is equal to quiet sleep [23]. Whatever definition is used, it is clear that quiet sleep is the easiest state to classify, indicated *e.g.* by the finding that the spectral edge frequency and the entropy of the EEG between quiet and active sleep differs significantly [24-26].

The training of the classifier is illustrated in figure 3, where the probability of error is shown for increasing numbers of features included in the classification process. When classifying BS compared to all other classes, the error goes down to zero for five features. The other classes do not reach this accuracy, and in some cases there is a tendency for the error to level out or even increase when the number of features is increased over a certain level. This is probably due to overfitting, *i.e.* the classifier starts to model the noise in the data resulting in poor generalization ability. Overfitting is common in machine learning and is avoided by monitoring the error and limiting the number of included features to the one that minimizes the error on the validation set.

The selected metafeature combinations are shown in table 4, and show that the individual metafeatures usually do not show up in more than one classification case. This is because the chosen feature combinations enhance what makes the individual classes stand out from the rest. The job of the classifier is to combine the chosen metafeature signals into a single classification signal that can be thresholded. These classification signals are displayed in figure 4. In the BS case there is no overlap between the signal distributions, so perfect classification is possible. In the other cases the amount of overlap varies from minimal in the QS case to complete overlap in the AS and AW cases.

When the classifier was tested on the long sleep stage recording from a healthy baby (figure 5 and figure 6), BS is not detected a single time. QS detection performs well for the first QS episode, but in the second a lot is confused with QW and some with AS. QW detection has a lot of false positives, and AS and AW detection works less well. When the classifier was tested on the long recording from a sick baby, all three BS episodes are detected.

The BS segmentation part was trained in a similar way as the classification part, and a minimum probability of error of 3.3 % on the validation set was achieved when using nine features. This corresponds to correct classification of 87 % of the suppression epochs and 94 % of the burst epochs, where epochs in this case refer to the one-second time resolution set by the feature extraction process.

Figure 11 and figure 12 shows the results of running the 32 h recording through the classification and segmentation parts and calculating the suppression lengths and the burst suppression ratio from the resulting BS pattern. These are two possible ways to display the acquired information regarding the state of the brain. For example, a suppression length of more than 30 seconds has been shown to have a high specificity for predicting an unfavourable neurological outcome [12], making it an important parameter for a future monitoring system.

5. Conclusion

We aim at developing one important component in a system for monitoring neonatal EEG. At this stage, the classifier easily can distinguish burst suppression from normal EEG in a neonate, and quiet sleep also can be separated from the other sleep stages. When the EEG is classified as BS, it is segmented into burst and suppression, and parameters such as suppression length and BSR can be calculated. This enables the clinician to follow brain function in quantitative terms. These results could be useful to include in a monitoring device for continuous monitoring of babies at neonatal intensive care units.

Acknowledgements

This research project was supported by a grant from the Margarethahemmet foundation, the Swedish state under the ALF-agreement and the BIOPATTERN EU Network of Excellence, EU Contract 508803.

We acknowledge the skilled acquisition of high quality EEG recordings from healthy neonates by Lisa Sellersjö and Maria Karlsson.

References

1. Scher M 2001 Perinatal asphyxia: timing and mechanisms of injury in neonatal encephalopathy *Current neurology and neuroscience reports* **1**(2) 175-84
2. Nicklin S, Wickramasinghe Y A and Spencer A S 2004 Neonatal intensive care monitoring *Current Paediatrics* **14**(1) 1-7
3. Löfgren N, Lindecrantz K, Flisberg A, Bågenholm R, Kjellmer I and Thordstein M 2006 Spectral distance for ARMA models applied to electroencephalogram for early detection of hypoxia *Journal of Neural Engineering* **3**(3) 227-34
4. Holmes G L and Lombroso C T 1993 Prognostic value of background patterns in the neonatal EEG *Journal of Clinical Neurophysiology* **10**(3) 323-52
5. Shevell M I, Majnemer A and Miller S P 1999 Neonatal neurologic prognostication: the asphyxiated term newborn *Pediatric Neurology* **21**(5) 776-84
6. Löfgren N, Lindecrantz K, Thordstein M, Hedström A, Wallin B G, Andreasson S, Flisberg A and Kjellmer I 2003 Remote Sessions and Frequency Analysis for Improved Insight Into Cerebral Function During Pediatric and Neonatal Intensive Care *IEEE Transactions on Information Technology in Biomedicine* **7**(4) 283-90
7. Prior P F and Maynard D E. 1986 *Monitoring Cerebral Function* (Amsterdam: Elsevier)
8. Rosén I 2006 The Physiological Basis for Continuous Electroencephalogram Monitoring in the Neonate *Clin Perinatol* **33** 593-611
9. Toet M C and Lemmers P M A 2009 Brain monitoring in neonates *Early Human Development* **85**(2) 77-84
10. Löfhede J, Löfgren N, Thordstein M, Flisberg A, Kjellmer I and Lindecrantz K 2008 Classification of burst and suppression in the neonatal electroencephalogram *Journal of Neural Engineering* **5**(4) 402-10
11. Prechtl H F R 1974 The behavioural states of the newborn infant (a review) *Brain Research* **76**(2) 185-212
12. Menache C C, Bourgeois B F D and Volpe J J 2002 Prognostic value of neonatal discontinuous EEG *Pediatric Neurology* **27**(2) 93-101
13. Lindecrantz K, Bågenholm R, Göthe F, Hedström A, Löfgren N, Nivall S and Ouchterlony J 1999 A general system used in monitoring of cerebral and circulatory function in neonatal intensive care *Med Biol Eng Comput* **37** 888-9
14. Jasper H H 1957 The Ten-Twenty Electrode System of the International Federation *International federation of societies for electroenchephalography and clinical neurophysiology* Brussels.

15. Thordstein M, Löfgren N, Flisberg A, Bågenholm R, Lindecrantz K and Kjellmer I 2005 Infraslow EEG activity in burst periods from post asphyctic full term neonates *Clinical Neurophysiology* **116**(7) 1501-6
16. Thordstein M, Wallin B G, Flisberg A, Bågenholm R, Kjellmer I, Löfgren N and Lindecrantz K 2004 Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates *Clinical Neurophysiology* **115**(11) 2461-6
17. Siedlecki W and Sklansky J 1989 A note on genetic algorithms for large-scale feature selection *Pattern Recognition Letters* **10**(5) 335-47
18. Salcedo-Sanz S, Camps-Valls G, Pérez-Cruz F, Sepúlveda-Sanchis J and Bousono-Calzón C 2004 Enhancing genetic feature selection through restricted search and Walsh analysis *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* **34**(4) 398-406
19. Alexandre E, Cuadra L, Rosa M and LÃ³pez-Ferreras F 2007 Feature selection for sound classification in hearing aids through restricted search driven by genetic algorithms *IEEE Transactions on Audio, Speech and Language Processing* **15**(8) 2249-56
20. Fisher R A 1936 The use of multiple measurements in taxonomic problems *Ann. Eugenics* **7**(2) 179-88
21. Duda R O, Hart, P. E., Stork, D. G. 2001 *Pattern Classification.* 2nd ed (New York: John Wiley & Sons, Inc.)
22. Gerla V, Paul K, Lhotska L and Krajca V 2009 Multivariate analysis of full-term neonatal polysomnographic data *IEEE Transactions on Information Technology in Biomedicine* **13**(1) 104-10
23. Piryatinska A, Terdik G, Woyczynski W A, Loparo K A, Scher M S and Zlotnik A 2009 Automated detection of neonate EEG sleep stages *Computer Methods and Programs in Biomedicine* **95**(1) 31-46
24. Zhang D, Ding H, Liu Y, Zhou C and Ye D 2009 Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram *Physiological measurement* **30**(5) 491-504
25. Löfgren N A, Outram N and Thordstein M 2007 EEG entropy estimation using a Markov model of the EEG for sleep stage separation in human neonates *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering* Kohala Coast, HI.
26. Korotchikova I, Connolly S, Ryan C A, Murray D M, Temko A, Greene B R and Boylan G B 2009 EEG in the healthy term newborn within 12 hours of birth *Clinical Neurophysiology* **120**(6) 1046-53

Paper VI

Does Indomethacin for Closure of Patent Ductus Arteriosus Affect Cerebral Function?

A. Flisberg, I. Kjellmer, J. Löfhede, N. Löfgren, M. Rosa-Zurera, K. Lindecrantz and M. Thordstein

Manuscript submitted to Acta Paediatrica.

Does Indomethacin for Closure of Patent Ductus Arteriosus Affect Cerebral Function?

A. Flisberg[1], I. Kjellmer[1], J. Löfhede[2], N. Löfgren[3], M. Rosa-Zurera[4], K. Lindecrantz[2] and M. Thordstein[5].

1. Department of Pediatrics, Division of Neonatology, The Queen Silvia Children's Hospital Sahlgrenska University Hospital - Östra, Göteborg, Sweden.
2. School of Engineering, University of Borås, Borås, Sweden and Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden
3. Neoventa Medical AB, Göteborg, Sweden.
4. Dept. of Signal Theory and Communications, University of Alcalá, Alcalá de Henares, Spain.
5. Institute of Neuroscience and Physiology, Department of Clinical Neurophysiology, Sahlgrenska University Hospital, Göteborg, Sweden.

Correspondence: Anders Flisberg, Department of Pediatrics, Division of Neonatology, The Queen Silvia Children's Hospital, Sahlgrenska University Hospital - Östra, Göteborg, Sweden.

Tel: + 46-31-3436707

Fax: + 46-31-3436696

Email: anders.flisberg@vgregion.se

Abstract

Objective: To study if indomethacin used in conventional dose for closure of patent ductus arteriosus, affects cerebral function measured by EEG evaluated by quantitative measures.

Study design: Seven premature neonates with clinically significant persistent ductus arteriosus were recruited. Electroencephalograms (EEG) from eight electrodes were recorded before, during and after an intravenous infusion of 0.2 mg/kg indomethacin over 20 minutes. The EEG was analysed by two methods with different degrees of complexity for the amount of low activity periods (LAP, "suppressions") as an indicator of cerebral strain.

Results: Neither of the two methods identified any change of the amount of LAPs in the EEG after as compared to before the indomethacin infusion.

Conclusion: Indomethacin in conventional dose for closure of patent ductus arteriosus, does not affect cerebral function as evaluated by quantitative EEG.

1. Introduction:

The clinical use of indomethacin for closure of symptomatic patent ductus arteriosus was established almost thirty years ago (1,2). Prophylactic therapy for patent ductus arteriosus in very-low-birth-weight infants started after Ment et al. (3,4) demonstrated that the incidence rate of intraventricular hemorrhage decreased after giving prophylactic indomethacin.

However, side effects from indomethacin were observed early on, not only related to the expected effect of reduced synthesis of prostaglandins but also resulting from a direct vasoconstrictor effect of indomethacin (5). Direct measurements of blood flow velocities in cerebral arteries revealed significant reduction of blood flow velocity to the brain of preterm infants starting 5-30 min after infusion of indomethacin and lasting at least one hour (6,7,8). Indirect estimates using near-infrared spectroscopy confirmed the reduction of blood flow after indomethacin together with a diminution of oxygen availability to the brain tissue (9, 10).The observations that indomethacin in therapeutic doses causes reduction of blood flow to the brain together with the reported lack of protection from cerebral palsy, deafness and blindness by prophylactic indomethacin in another large controlled trial (11), led to substitution of indomethacin with ibuprofen in many clinical centres. But conclusions from current literature are by no means straight forward. Follow-up studies claim decreased white matter injury (12) and favourable neurodevelopmental outcome (13) after indomethacin. In a systematic review indomethacin is recommended for use during the first day of life but ibuprofen on the following days (14) and when the impact of two multicenter randomized controlled trials (4,11) on clinical practice is evaluated, the authors arrive at advocating the use of indomethacin (15).

But the fact that indomethacin reduces blood flow and oxygen delivery to the brain remains. It would thus be reasonable to find out whether these effects of indomethacin are severe enough to cause a deterioration of the brain function. We therefore investigated the influence of indomethacin on the EEG activity in preterm neonates treated with indomethacin because of clinically significant patent ductus arteriosus. The EEG of very preterm infants typically demonstrates a pattern alternating between low

and high amplitude activity. Generally, strain of the central nervous system affecting cortical function, leads to increased proportions of low amplitude activity and changes in the frequency content of the EEG (16, 17). Therefore, the characteristics of this pattern were evaluated in preterm infants with patent ductus arteriosus immediately before and during treatment with indomethacin.

2. Material and methods

2.1. Patients

The inclusion criteria were prematurity, defined as an age of less than 34+0 gestational weeks at birth and a hemodynamically significant patent ductus arteriosus.

We included seven newborn premature babies with a gestational age between 25-33 weeks at birth. Three of the babies were SGA (<-2SD) and four were AGA (± 2 SD). All seven babies had a hemodynamically significant ductus arteriosus and were treated with indomethacin, 0.2 mg/kg .The drug was given as an intravenous infusion over 20 minutes. Treatment was initiated between day 1-7 of life. The number of doses for each patient was between two and three given with an interval of 12 hours. Indomethacin treatment was successful in six of the babies. In one baby the ductus arteriosus was later surgically closed. Six babies had RDS (respiratory distress syndrome). Three of the babies were treated with ventilator, three treated with CPAP-(continues positive airway pressure) and one baby was without any respiratory support during the EEG recording. Oxygen saturation was measured continuously and was kept above 92%. No medication was added during the recording. The three babies on the ventilator had intravenous infusion of morphine in a dose between 8-20 μg / kg / h.

All but one patient had invasive arterial blood pressure continuously measured. Mean arterial blood pressure was above 35 mm Hg for all babies during the registration. pCO_2 was measured before and after indomethacin. In the group it varied between 3.5-8.0 kPa. Blood glucose was measured before and after indomethacin and was > 3mmol/l in all babies except one were the values were between 1.7-2.9 mmol/l. Ionized calcium was above 1.1 mmol/l and hemoglobin levels were above 140 g/l in all babies.

Head ultrasound was normal prior to the treatment with indomethacin in all the babies. A follow up head ultrasound was done in five babies and was normal. Two babies did not have a second ultrasound.

Routine follow up was made. All the babies except one had a normal follow up at 18 corrected months of age. The baby with abnormal follow up had a delayed sensory-motor development.

2.2. Recordings

EEG was recorded from eight electrodes positioned according to the 10-20 system at F3, P3, F4, P4, F7, T5, F8 and T6 and displayed using a common average montage. The signals were high pass filtered at 1.6 Hz, low pass filtered at 70 Hz, amplified and stored digitally using a sample frequency of 200 Hz. The system used for recording, SACS® (Signal Archiving and Communication System), is a software system which runs on a standard PC under Windows operating system (18) with a custom built amplifier.

2.3. Analysis

A period from 40 minutes before to 2 h after start of the first dose of indomethacin infusion was selected for analysis.

2.3.1. Method 1

The time period between 40 and 20 minutes before administration of indomethacin was chosen as a baseline period. The expected time for maximal effect of indomethacin on cerebral perfusion (7), 30-60 minutes after administration, was chosen for evaluation of a possible effect on cerebral function. These periods were first visually analysed by a clinical neurophysiologist (MT), specialised in newborn EEG interpretation to exclude periods of recording with artefacts. Next, the periods were divided into 60 s segments with 30 s overlap. The samples in each segment were classified as belonging to a high amplitude period (HAP) or low amplitude period (LAP) by using a simple amplitude threshold and time constrains:

First step, channel wise classification;

LAP is a period of at least one second where no absolute value of any sample exceeds 15 μV. HAPs are all samples not classified as LAPs.

Second step, combining all electrodes;

For a sample to be classified as HAP in the second step, it has to be classified as HAP in at least five out of the eight channels during the first step and be at least one second long in all these channels. All samples not fulfilling the above requirements are classified as LAPs.

Based on this classification, mean LAP time was calculated for each 60 s segment not containing artefacts. If an artefact was present in any channel within the segment, the whole segment was discarded from the analysis.

This produced a feature signal containing an estimate of the mean LAP time with one sample each 30 s.

2.3.2. Method 2

A set of 22 feature signals were extracted from the EEG (cf. appendix table), using a one-second sliding window with 0.75 s overlap producing feature signals with an effective sampling rate of 4 Hz. Each feature signal describes a specific property of the underlying EEG signal, for example power or frequency distribution, and can be used for classification of different types of activity in the EEG. The features were extracted from the eight EEG channels separately, but then the median over the channels were taken, producing a single signal for each feature. This was done because HAPs were defined to be present in at least half of the channels.

Ten minutes (t=[-20,-10] min) of the signal from each patient was segmented by visual analysis (MT) and was used as training data for the classification algorithm. To find the best features to use, the data was fed to a genetic algorithm that tested combinations from one to ten features, using the classification algorithm Fisher's linear discriminant (19). The process was repeated independently for the different patients. Half of the pre-segmented data was used for training and half was used for testing the classifier. The feature combination that gave the lowest probability of error on the test data was then used for segmenting the entire signal (Appendix 2).

The output signal from the classification algorithm was filtered for removing short detections or gaps in longer detections in the detection signal. This was done using a 4-point (1 s) sliding window, setting the output of each window to the majority class in the input. Fig 1 shows an example of the result of automatic segmentation together with the manually segmented training sequence.

Then the lengths of all LAPs were calculated for twenty minute epochs, including a baseline period defined as 40-20 minutes before the infusion of indomethacin. For each epoch and patient the median LAP length was calculated, resulting in seven values for each epoch. The median and confidence interval (5^{th} and 95^{th} percentile) was then estimated using bootstrapping (20).

3. Results

In the analysis based on the threshold of amplitude (the first method) the mean LAP at baseline was 4.6 +/-3.2 seconds, and during the period 30-60 minutes after administration of indomethacin, 4.9 seconds, -/+ 5.2 seconds (mean +/- SD, Fig 2).

The results from the more elaborate analysis using feature extraction in combination with classification according to Fisher's linear discriminant algorithm (the second method) are displayed in Fig 3. Compared to baseline there was no difference in LAP time after indomethacin was given.

Thus, in neither of the analyses an effect on cerebral function measured as LAP time of the EEG after indomethacin administration of 0.2 mg/kg, could be found.

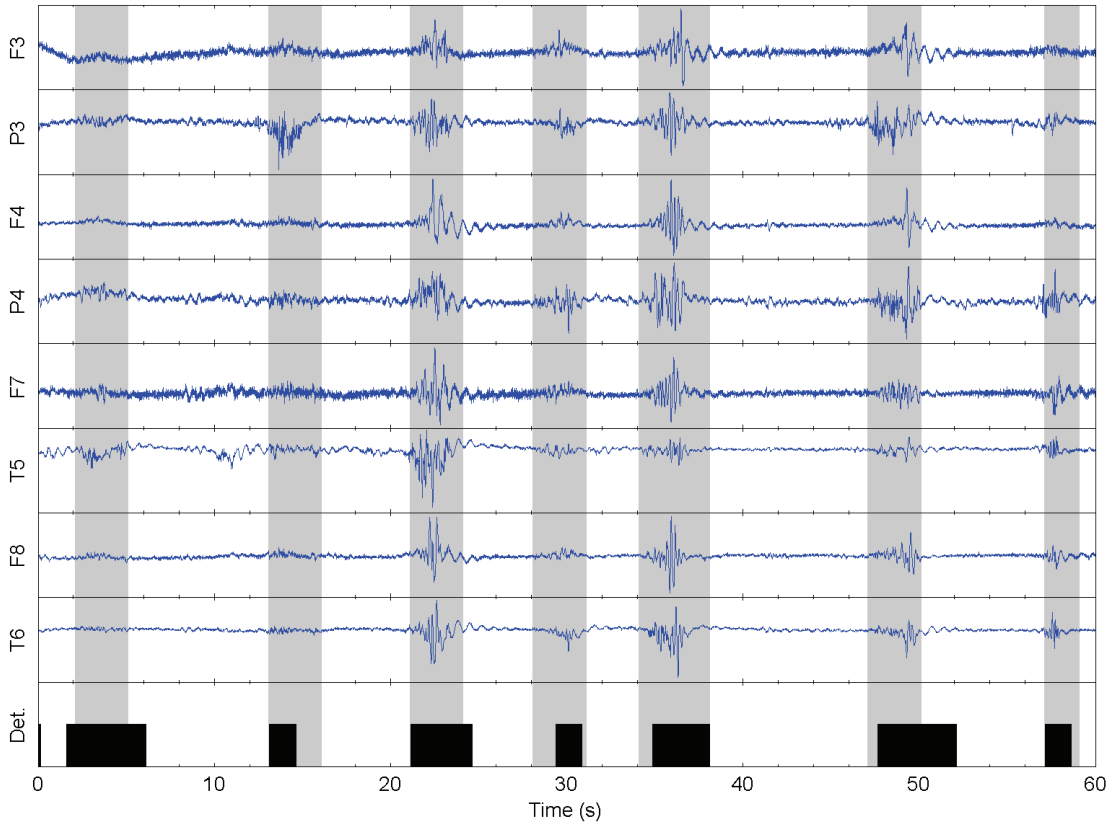


Fig. 1: Sixty seconds of eight channels of EEG from one patient, filtered from 1.6 to 70 Hz. The shaded areas have been manually classified as high activity periods (HAPs) by a neurophysiologist (MT). The black blocks show the parts of the signal that has been classified as HAPs by the machine learning algorithm.

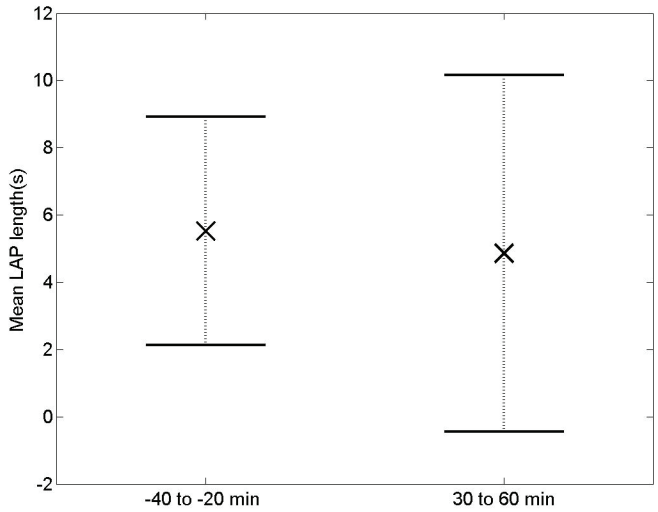


Fig. 2: Mean time for low activity periods over one minute EEG segments at 40-20 minutes prior to and 30-60 minutes after the onset of indomethacin infusion respectively. The crosses represent the mean of these mean suppression lengths and the horizontal bars show the mean +/- the standard deviation.

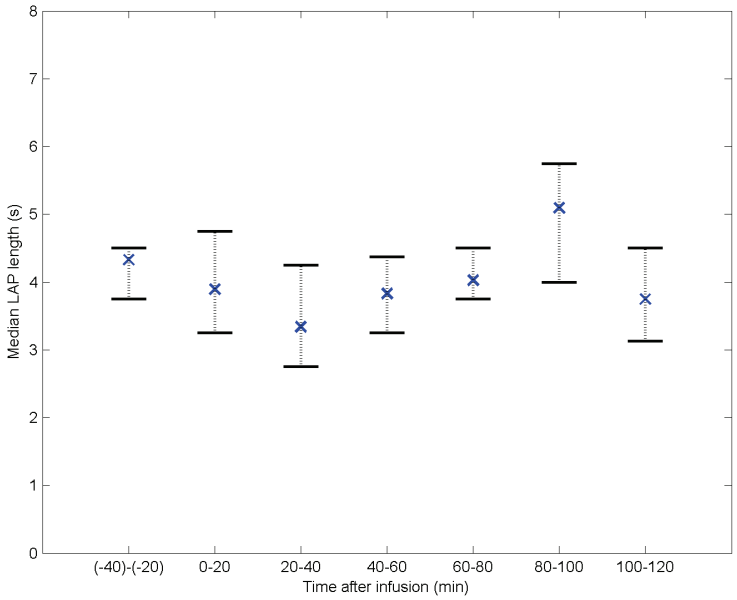


Fig. 3: Median low activity period (LAP) length (marked with X) for 20 minute epochs. The horizontal lines indicate the 5th and 95th percentile. The infusion of Indomethacin was started at t = 0, and the period from 40 to 20 minutes before the infusion is used as base line.

4. Discussion

Patency of the ductus arteriosus is a common problem in the very-low-birth-weight infant. The pulmonary consequences of this complication have been studied extensively and pharmacological closure of symptomatic ductus arteriosus in preterm babies is clinical practice.

It has been suggested that an altered cerebral blood flow (CBF) after the administration of indomethacin could be a contributory cause of ischemic cerebral damage (21).

We used EEG to measure the effect of indomethacin on the cerebral function. Two different methods were applied to analyze the original EEG signal for the degree of intermittency. We were unable to detect any effect on the EEG after indomethacin was given with either one of these methods (Fig. 2 and Fig. 3).

The intermittent EEG, typical for the very preterm neonate (16) reacts to adverse influences including reduced cerebral perfusion (22) with increased intermittency recorded as prolonged LAPs, often referred to as prolonged interburst intervals.

A lack of effect creates problems for interpretation. Could there have been an effect that was missed in our recordings? A recent report describes how an EEG cap with 20 recording electrodes detects spontaneously occurring focal and localized so called spontaneous activity transients in the EEG of the very preterm infant (23). Our recording system used fewer recording electrodes – eight – and therefore did not provide the same degree of resolution. On the other hand, we wanted to study generalized effects from an intravenous infusion of indomethacin, which most probably would have been detected in our system. It is therefore reasonable to assume that any effect of indomethacin would have been detected by our system.

Conclusions: Indomethacin in a dose of 0.2mg/kg, given intravenously over 20 minutes to preterm infants does not increase the degree of intermittency of the EEG, indicating that any possible cerebral vasoconstriction is not severe enough to affect cerebral cortical function.

Acknowledgments

This study was supported by grants from the Sahlgrenska Academy, from the Göteborg Medical Society, the Swedish state under the ALF agreement and from the Margarethahemmet Foundation.

Abbreviations

SEF90, Spectral edge frequency 90; EEG, Electroencephalogram; MABP, Mean arterial blood pressure; ECG, Electrocardiogram; FIR, Finite impulse response; SACS, Signal Archiving and Communication System.

References

1. Yeh TF, Luken JA, Thalji A, Raval D, Carr I, Pildes RS. Intravenous indomethacin therapy in premature infants with persistent ductus arteriosus--a double-blind controlled study. J Pediatric 1981;98:137-145.

2. Gersony WM, Peckham GJ, Ellison RC, Miettinen O, Nadas A. Effects of indomethacin in premature infants with patent ductus arteriosus: results of a national collaborative study. J Pediat 1983;102:895-906

3. Ment LR, Duncan CC, Ehrencranz RA, Kleinman CS, Pitt BR, Taylor KJ, Scott DT, Stewart WB, Gettner P. Randomized indomethacin trail for prevention of intraventricular hemorrhage in very low birth weight infants. J Pediat 1985;197:937-43

4. Ment LR, Oh W, Ehrencranz RA, Philip AG, Vohr B, Allan W, Duncan CC, Scott DT, Taylor KZ, Katz KH, Schneider KC, Makuch RW. Low-dose indomethacin and prevention of intraventricular hemorrhage: a multicenter randomized trial. Pediatrics 1994;93:543-50 (alla förff. Kommer inte med i slutlistan – hur många beror på vilken tidskrift)

5. Wennmalm Å, Carlsson I, Edlund A, Eriksson S, Kaijser L, Nowak J. Central and peripheral haemodynamic effects of non-steroidal anti-inflammatory drugs in man. Arch Toxicol Suppl 1984;7:350-9

6. Lundell BP, Sonesson SE, Cotton RB. Ductus closure in preterm infants. Effects on cerebral hemodynamics. Acta Paediatr Scand Suppl 1986;329:140-7

7. Saliba E, Chantepie A, Autret E, Gold F, Pourcelot L, Laugier J. Effects of indomethacin on cerebral hemodynaics at rest and during endotracheal suctioning in preterm neonates. Acta Paediatr Scand 1991;80:611-5

8. Mosca F, Bray M, Lattanziio M, Fumagalli M, Tosetto C. Comparative evaluation of the effects of indomethacin and ibuprofen on cerebral perfusion and oxygenation in preterm infants with patent ductus arteriosus. J Pediat 1997;131:549-54

9. Benders MJ, Dorrepaal CA, van der Bor M, van Beel F. Acute effects of indomethacin on cerebral hemodynamics and oxygenation. Biol Neonate 1995;68:91-9

10. Patel J, Roberts I, Azzopardi D, Hamilton P, Edwards D. Randomized double-blind controlled trial comparing the effects of ibuprofen with indomethacin on cerebral hemodynamics in preterm infants with patent ductus arteriosus. Pediatr Res 2000;47:36-42

11. Schmidt B, Davis P, Moddemann D, Ohlsson A, Roberts RS, Saigal S, et al. Long-term effects of indomethacin prophylaxis in extremely-low-birth-weight infants. N Engl J Med 2001;344:1966-72

12. Miller SP, Mayer EE, Clyman RI, Glidden DV, Hamrick SE, Barkovich AJ. Prolonged indomethacin exposure is associated with decreased white matter injury detected with magnetic resonance

imaging in premature newborns at 24 to 28 weeks´ gestation at birth. Pediatrics 2006;117:1626-31

13. Ment LR, Vohr B, Allan W, Westerveld M, Sparrow SS, Schneider KC, et al. Outcome of children in the indomethacin intraventricular hemorrhage prevention trial. Pediatrics 2000;105:485-91
14. Aranda JV, Thomas R. Systematic review: intravenous ibuprofen in preterm newborns. Semin Perinatol 2006;30:114-20
15. Clyman RI, Saha S, Jobe A, Oh W. Indomethacin prophylaxis for preterm infants: the impact of 2 multicentered randomized controller trials on clinical practice. J Pediatr 2007;150:46-50.
16. Scher MS. Electroencephalography of the newborn: Normal and abnormal features. In Niedermeyer E and Lopes da Silva F (eds): *Electroencephalography; Basic principles; Clinical applications and related fields*. p 896-946, Lippincott, Williams & Wilkins, 1999 17
17. Thordstein M, Flisberg A, Löfgren N, Bågenholm R, Lindecrantz K, Wallin BG, Kjellmer I. Spectral analysis of burst periods in EEG from healthy and post-asphyctic full-term neonates. Clin Neurophysiol 2004; 115: 2461-24618.
18. Lindecrantz K, Bågenholm R, Göthe F, Hedström A, Löfgren N, Nivall S et al. A general system used in monitoring of cerebral and circulatory function in neonatal intensive care. Med Biol Eng Comp 1999; 37 Suppl 2 : 888-889.
19. Duda RO, Hart PE Stork DG. Pattern Classification.2001 (New York; John Wiley& Sons,Inc.).
20. Rice JA. Mathematical Statistics and Data analysis (Belmont: Duxbury Press).
21. Liem KD, Hopeman, JCW, Kollee, Oeseburg B. Effects of repeated indometacin administration on cerebral oxygenation and haemodynamics in preterm.infants combined near infrared spectrophotometry and Doppler ultrasound study. Eur J Pediatr 1994;153:504-509.
22. Victor S, Marson AG, Appleton RE, Beirne M, Weindling AM. Relationship between blood pressure, cerebral electrical activity, cerebral fractional oxygen extraction, and peripheral blood flow in very low birth weight newborn infants. Pediatr Res 2006;59:314-319
23. Vanhatalo S, Metsäranta M, Andersson S. High-fidelity recording of brain activity in the extremely preterm babies: Feasability study in the incubator. Clin Neurophysiol 2008;119:439-445

Appendix 1

List of features used for segmentation in method 2.

Feature	Description
Spectral Edge Frequency (SEF95)	Frequency under which 95% of the signal power resides, based on the Fourier transform (FT) calculated on rectangular windows of the signal
3 Hz power	Power in a 1-Hz wide band centred at 3 Hz
Median	Median absolute value
Shannon entropy (H_{Sh})	$H_{Sh} = -\sum_{u=1}^{U} p(I_u)\log p(I_u)$ $p(I_1)\ldots p(I_U)$ is a discrete set of probabilities, which are estimated by counting the samples falling in the disjoint amplitude intervals $I_1,..,I_U$. 20 intervals were used evenly distributed between the maximum and minimum values of the signal in the window. H_{Sh} is a measure of uncertainty of a random variable.
Zero crossings	The number of zero crossings in each window.
Variance ($s^2(x)$)	$s^2(x) = \frac{1}{N-1}\sum_{n=1}^{N}(x[n]-\mu)^2$ where x is a time series, and μ is the sample mean of x
Spectral centroid	The ''centre of mass'' of the frequency distribution $c = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$ Where x(n) represents the magnitude of bin number n, and f(n) represents the centre frequency of that bin.
Residual energy variance	The variance of the residual after applying an eight-coefficient linear predictor on the signal window.
Spectral flux	Measures the change in the spectrum between consecutive windows using the Euclidian distance (2-norm) between the spectra.
deltaF	Measures the rate of change by taking the Euclidian distance between consecutive windows of the signal.

Spectral flatness	Calculated by dividing the geometric mean of the power spectrum with the arithmetic mean. $$F = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}}$$ A high spectral flatness indicates that the spectrum has a similar amount of power in all bands – like white noise A low spectral flatness indicates a spiky spectrum
Spectral roll-off	Equivalent to SEF85
Cepstral coefficients	The cepstral coefficients were extracted by applying the Fourier transform on a window of the signal, mapping the logarithm of the spectrum onto a linear scale using ten triangular overlapping windows, and then taking the discrete cosine transform of the list of amplitudes. The cepstral coefficients are the amplitudes of the resulting spectrum, and are related to the rate of change in the corresponding frequency bands.

Appendix 2

Features selected (x) by the generic algorithm and used for classification of HAP and LAP.

Feature	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Pat. 6	Pat. 7
SEF95	x					x	
3 Hz power		x	x	x	x	x	x
median		x			x	x	
entropy			x	x	x	x	x
zero crossings	x	x	x		x	x	
variance						x	
centroid		x		x		x	
residual energy var.		x			x	x	
spectral flux					x		
deltaF						x	
spectral flatness		x		x	x		
spectral roll-off				x			
cepstral coeff 1	x		x	x			
cepstral coeff 2				x			x
cepstral coeff 3		x			x		x
cepstral coeff 4		x					
cepstral coeff 5	x			x			
cepstral coeff 6				x	x		
cepstral coeff 7		x					
cepstral coeff 8	x	x					
cepstral coeff 9	x						
cepstral coeff 10				x		x	