Automatic detection of spontaneous activity transients in preterm electroencephalography

Kirsi Palmu

Thesis for the Degree of Licentiate of Science
Helsinki, January 2013

Supervising professor: Prof. Risto J. Ilmoniemi
Thesis instructors: Adj. Prof. Sampsa Vanhatalo
Harri Valpola, PhD

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of licentiate thesis

Author Kirsi Palmu

Title of thesis Automatic detection of spontaneous activity transients in preterm electroencephalography

Department Department of Biomedical Engineering and Computational Science

Field of research Engineering Physics, Biomedical Engineering

Supervising professor Prof. Risto Ilmoniemi **Code of professorship** F036Z

Thesis instructors Adj. Prof. Sampsa Vanhatalo, Harri Valpola, PhD

Thesis examiner Doc. Ari Pääkkönen

Number of pages 57 **Language** English

Date of submission for examination 29.1.2013

Abstract

Very preterm infants may require neonatal intensive care for several months, and the developmental outcome of the care depends on how well brain function can be managed. Direct monitoring of brain function with electroencephalography (EEG) is currently not a part of routine care, since it is perceived challenging due to difficulties in its interpretation. Therefore, automated methods for EEG interpretation are needed in order to make brain monitoring part of the routine in neonatal intensive care.

This thesis investigates the detection of spontaneous activity transients (SATs), which form the majority of brain activity in preterm infants. Using manual markings by three doctors in 18 short recordings of preterm EEG, I show that SATs can be recognized by doctors in a consistent manner. A commercially available algorithm is then tested for its ability to detect SATs automatically. The performance of the algorithm is clearly insufficient and therefore it is developed further.

The parameters of the new, streamlined algorithm are optimized using unanimous markings by the three doctors as a gold standard. Estimates for the performance of the algorithm on unseen data are obtained by running the optimization 18 times, each time leaving out one of the recordings. The algorithm is then run on the EEG left out from the optimization using the optimized parameters. The estimated performance of the algorithm is found to be excellent, with sensitivity of 96.6 ± 2.8 % and specificity of 95.1 ± 5.6 %.

Segmentation of the EEG into SATs and periods between SATs is a starting point for further analysis. One promising direction for future studies is to use SAT%, the proportion of time covered by SATs, to detect cycles of different vigilance stages in preterm infants. Such cyclicity could become a marker of the brain's wellbeing.

The algorithm presented in this thesis may contribute to better care of preterm infants.

Keywords Preterm, neonate, prematurity, EEG, spontaneous activity transient, automated detection, algorithm, optimization, validation

Aalto-yliopisto
Perustieteiden
korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Lisensiaatin tutkimuksen tiivistelmä

Tekijä Kirsi Palmu

Työn nimi Spontaanien aktiviteettipurskeiden automaattinen tunnistus keskosten aivosähkökäyrästä

Laitos Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Tutkimusala Teknillinen fysiikka, lääketieteellinen tekniikka

Vastuuprofessori Prof. Risto Ilmoniemi **Professuurikoodi** F036Z

Työn ohjaajat Dos. Sampsa Vanhatalo, TkT Harri Valpola

Työn tarkastaja Dos. Ari Pääkkönen

Jätetty tarkastettavaksi 29.1.2013 **Sivumäärä** 57 **Kieli** englanti

Tiivistelmä

Erittäin ennenaikaisesti syntyneet keskoset saattavat tarvita teho-osastohoitoa jopa kuukausien ajan. Hoidon vaikutus lapsen kehitykseen riippuu paljon siitä, kuinka hyvin aivojen hoito onnistuu. Aivojen toiminnan jatkuva valvonta elektroenkefalografian (EEG) avulla ei vielä kuulu tavanomaiseen hoitokäytäntöön, koska EEG:n tulkintaa pidetään vaikeana. EEG:n tulkintaan tarvitaankin automaattisia menetelmiä, jotta aivojen tarkkailusta tulisi osa vastasyntyneiden tehohoidon rutiinia.

Tässä työssä tutkitaan spontaanien aktiviteettipurskeiden tunnistamista (engl. *spontaneous activity transient*, SAT). Keskosten aivotoiminta muodostuu suurelta osin aktiviteettipurskeista. Käyttämällä kolmen lääkärin käsin tehtyjä merkintöjä aktiviteettipurskeista 18 lyhyessä keskosilta mitatussa EEG:ssä todistan, että lääkärit tunnistavat aktiviteettipurskeet johdonmukaisesti. Tämän jälkeen testaan, sopiiko eräs myynnissä oleva algoritmi aktiviteettipurskeiden automaattiseen tunnistukseen. Algoritmin suorituskyky ei ole riittävä, joten kehitän siitä paremman version.

Uuden, parannellun algoritmin parametrit optimoidaan käyttämällä opetusaineistona niitä EEG-jaksoja, joiden luokittelusta kaikki kolme lääkäriä olivat yhtä mieltä. Algoritmin suorituskykyä arvioidaan suorittamalla optimointi 18 kertaa siten, että kullakin kerralla yksi mittauksista jätetään pois opetusaineistosta. Optimoitua menetelmää käytetään sitten aktiviteettipurskeiden tunnistamiseen poisjätetyssä mittauksessa. Algoritmin arvioitu suorituskyky on erinomainen; sen sensitiivisyys on 96,6 ± 2,8 % ja spesifisyys 95,1 ± 5,6 %.

EEG:n segmentointi aktiviteettipurskeisiin ja niiden välisiin jaksoihin tarjoaa pohjan jatkoanalyysille. Aktiviteettipurskeiden osuutta EEG:stä (SAT%) voidaan mahdollisesti käyttää keskosen vireystilan vaihtelujen seuraamiseen. Vireystilojen säännöllinen vaihtelu saattaa olla merkki aivojen hyvinvoinnista.

Tässä työssä esitelty algoritmi voi osaltaan edesauttaa keskosten hoidon kehittymistä entistä paremmaksi.

Avainsanat keskonen, vastasyntynyt, keskosuus, EEG, spontaani aktiviteettipurske, automaattinen tunnistus, algoritmi, optimointi, validointi

CONTENTS

LIST OF PUBLICATIONS

This Thesis is based on the following publications, which are referred to in the text by their Roman numerals.

I K. Palmu, S. Wikström, E. Hippeläinen, G. Boylan, L. Hellström-Westas, S.Vanhatalo. Detection of 'EEG bursts' in the early preterm EEG: visual vs. automated detection. *Clinical Neurophysiology*, 121:1015-1022, 2010.

II K. Palmu, N. Stevenson, S. Wikström, L. Hellström-Westas, S. Vanhatalo, J. M. Palva. Optimization of an NLEO-based algorithm for automated detection of spontaneous activity transients in early preterm EEG. *Physiological Measurement*, 31:N85-93, 2010.

The author has made a significant contribution to the publications. She had a major part in the planning of the analysis and optimization procedure and performed all calculations. She is also the principal writer of the publications.

LIST OF ABBREVIATIONS

ADR	average detection rate
aEEG	amplitude integrated EEG
APU	adaptive preprosessing unit
AR	auto-regressive
AS	active sleep
BBI	burst-to-burst interval
CA	conceptual age
DC	direct current
EEG	electroencephalography
EMG	electromyography
FLD	Fisher's linear discriminant
GLR	generalized likelihood ratio
GA	gestational age
IB%	interburst percentage, proportion of time not covered by bursts
IBI	interburst interval, time between two bursts
IBR	interburst-burst ratio
IEI	inter-event-intervals
inter-SAT	time between two SATs, also called IBI
LRTC	long-range temporal correlation
MBAT	multiband activity transient
NICU	neonatal intensive care unite
NLEO	non-linear energy operator
NN	neural network
PCA	post conceptional age
PMA	post menstrual age
rEEG	range-EEG
REM	rapid eye movement
SAT	spontaneous activity transient, also called burst
SAT%	SAT-percentage, proportion of time covered by SATs
SD	standard deviation
SEF	spectral edge frequency
SEM	spectral error measurement
SV	spontaneous variance
SWC	sleep wake cycle
TA	tracé alternant
QS	quiet sleep
WMI	white matter injury

1 INTRODUCTION

Motivation

Optimally, human babies are born after 37-41 weeks of gestation (Blencowe et al. 2012). At this time, many developmental processes have achieved a state which makes living and breathing in the outside world possible. Preterm babies, meaning infants born before 37 completed gestational weeks, face the challenges of extrauterine life in a less mature condition. Many of them need special care given in neonatal intensive care units (NICU). Due to continuous improvement of medical care, ever smaller and younger infants survive. However, survival does not always guarantee high quality of life. Every third child born prematurely suffers from neurocognitive problems (Mwaniki et al. 2012). For children born extremely preterm (before the 26^{th} week of gestation), the situation is even worse: in a recent study 80% of these children were found to be at least mildly disabled at the age of 6 years (Marlow et al. 2005).

In NICU, vital signs (heart and respiratory rate, blood pressure, blood oxygenation) are constantly monitored in preterm babies in order to enable an immediate reaction to any physiological problems. The ultimate goal of the monitoring is to protect the infants brain but direct monitoring of the brain's wellbeing is not yet part of the standard care procedure. This is unfortunate, as the time which preterm babies spend in the NICUs heavily overlaps with the period in which the main neural connections and sensory organization of the brain are formed (Vanhatalo, Kaila 2010). Adverse events in this period may cause irreversible deficits in the brain's wiring. Importantly, some of these events may even be caused by the care itself, such as unnecessarily high medication which prevents normal brain activity.

The preterm brain is anatomically and functionally different from brains of all other age groups including term babies. Its most salient feature is *spontaneous activity transient* (SAT, Vanhatalo et al. 2005) which are believed to be crucial for the development of correct nervous connections. Our hypothesis is that real time monitoring of SATs would contribute to a better understanding of the preterm brain and consequently to improved care of vulnerable preterm infants.

Problem statement

In this Thesis, I want to answer the following questions:

1) Can SATs be reliably recognized by human observers?

2) What are the characteristics of SATs?

3) How could we automatically detect SATs?

Outline

In the second chapter, I give a brief description of the origin and appearance of SATs. I also summarize present knowledge about measurable features of preterm EEG such as its frequency content.

The third chapter presents a thorough literary review of methods used for automated quantitative analysis of neonatal and especially preterm brain activity.

The fourth chapter presents data and methods used in Publications I and II. I utilize 18 short electroencephalograms (EEG) from preterm babies and visual rating of SATs by three doctors. In this way, it is possible to evaluate the agreement between the raters and also to use a more reliable set of ratings, based on the unanimous ratings of all three raters, as a gold standard for the calculation of SAT characteristics and the development of an automated SAT detection.

The fifth chapter is dedicated to results. I show that SATs are recognized by human raters with a relatively high overall inter-rater agreement. I also show that SATs are characterised by a duration of 1-10 seconds and increased amplitudes through the frequency spectrum. These results are partly utilized in the optimization of an algorithm for SAT detection. After optimization, SATs can be detected automatically with high accuracy.

In the sixth chapter, the results are discussed. Special attention is given to the pitfalls of the algorithm and future perspectives. Some preliminary results of work done after Publications I and II are shown.

The seventh chapter gives the conclusions.

2 BACKGROUND

Automated analysis of preterm EEG is a pattern recognition task: we want to recognize certain patterns in order to learn more about underlying data. A central part of this process is to *segment* the data into segments which only contain one pattern. As we do not want to do the segmentation manually, algorithms are needed. Rather than using just the raw data as input, they use some mathematically obtained *features* of the data which might help to differentiate between the patterns in question. The search for the best features and algorithms is then the main challenge in a pattern recognition task. After the segmentation, new information can be gained from the data. Again, there are several ways to quantify the segmented data and studies are needed to find the most useful measures.

In the first part of this chapter, the physiology of preterm EEG is described. The two following subchapters summarize the current knowledge about preterm EEG in terms of its frequency content (2.2) and the occurrence of SATs and inter-SATs, the hallmarks of preterm EEG (2.3). This knowledge is essential background information for the literature review of methods for automated EEG visualization and analysis in chapter 3.

2.1 Preterm EEG

EEG is the most commonly used method for functional measurements of the brain. It measures potential differences caused by synchronous neuronal activation in the brain by electrodes attached to the scalp. Preterm EEG is characterised by large bursts of activity, dominated by a low frequency wave with superimposed (nested) higher frequency oscillations (Figure 1). These bursts have been called by many names, including delta waves or delta brushes referring to a specific frequency range of 0.5-4 Hz named delta in the traditional EEG literature (for an overview on terminology, see Vanhatalo, Kaila 2010). Our group introduced the name *spontaneous activity transients* (SAT, Vanhatalo et al. 2005) to emphasise the endogenous nature of the transients. An additional reason for introduction of a new term is that these bursts contain activity in a wide frequency range not necessarily dominated by delta (Vanhatalo, Kaila 2010). A similar reasoning was followed by Hartley et al. (2012) who called the middle part of the same events “bursts of nested (high-frequency) oscillations within large slow-wave depolarisations” (BNO). In the remaining chapters, we use the terms “burst” and “SAT” interchangeably, as we do with “inter burst

interval" (IBI) and "inter-SAT period" meaning the time between these events. The terms burst and IBI are used especially in the literature review, as these have been the most commonly used descriptions in history, whereas SAT and inter-SAT are used when referring to our own studies.

SATs are a phenomenon that only exists for a certain period of development. In the immature brain, they are the main means of communication between brain areas and they are believed to be crucial for the development of correct nervous connections in the brain. Gradually, as the structure of the cortex approaches a more mature state, a qualitatively different, continuous oscillatory activity appears in the EEG of the babies. In the smallest preterm babies, the EEG is dominated by large SATs with nearly flat inter-SAT periods. In babies approaching term age, SATs have diminished in size whereas their structure has become more complex and the inter-SAT periods show oscillations with ever higher amplitudes. Connections between brain areas enable higher synchronisation of the brain activity (Vanhatalo, Kaila 2006). A schematic presentation of the development of SAT and inter-SAT periods is given in Figure 2.

In clinical EEG interpretation, preterm EEG is mostly described by means of continuity. In general, continuity implies EEG activity above a certain amplitude, whereas discontinuity is defined in terms of prominent low voltage activity or IBIs. The exact definitions of continuous vs. discontinuous pattern vary (for examples see table 1). Often a middle class is also defined for periods that do not fit into either of the main classes. This class might be called e.g. "semi-discontinuous tracing" (Andre et al. 2010) or "undifferentiated pattern" (Hayakawa et al. 2001).

Table 1: Some definitions of continuous / discontinuous preterm EEG in literature.

EEG tracing / pattern	**André et al. 2010**	**Haykawa et al. 2001**
continuous	"physiological for gestational age features, with minimal amplitude of 25 μV, lasting at least 1 min."	"EEG activity mainly consisting of delta waves > 100 μV that were continuously recognised for more than 20 seconds."
discontinuous	"Bursts of physiological activity according to age, separated by interburst intervals (IBI) of amplitude <25 μV lasting more than 3 s." For "discontinuous" tracing, the IBIs should cover at least 50% of a 1 min analysis period.	"bursts of EEG activity separated by low voltage activity < 30 μV for more than five seconds. Bursts were defined as EEG activity with amplitudes more than 100 μV lasting for 2–20 seconds in any of the channels."

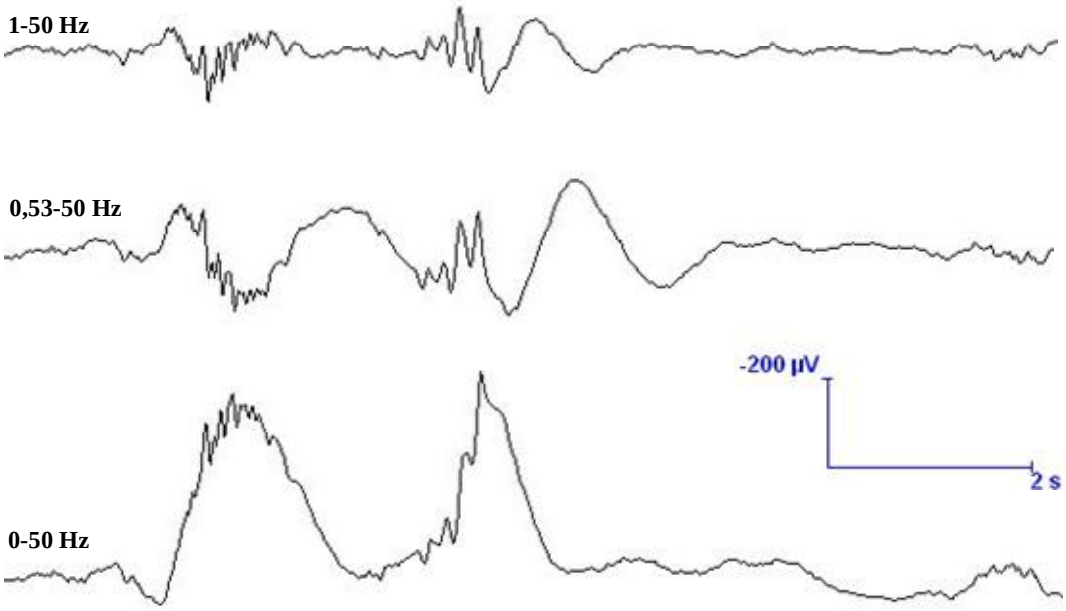

Figure 1. Preterm EEG with two SATs. EEG shown both without filtering (bottom) and with some conventional filter settings. Conventional high pass filtering heavily distorts the appearance of the SATs. (Palmu 2008)

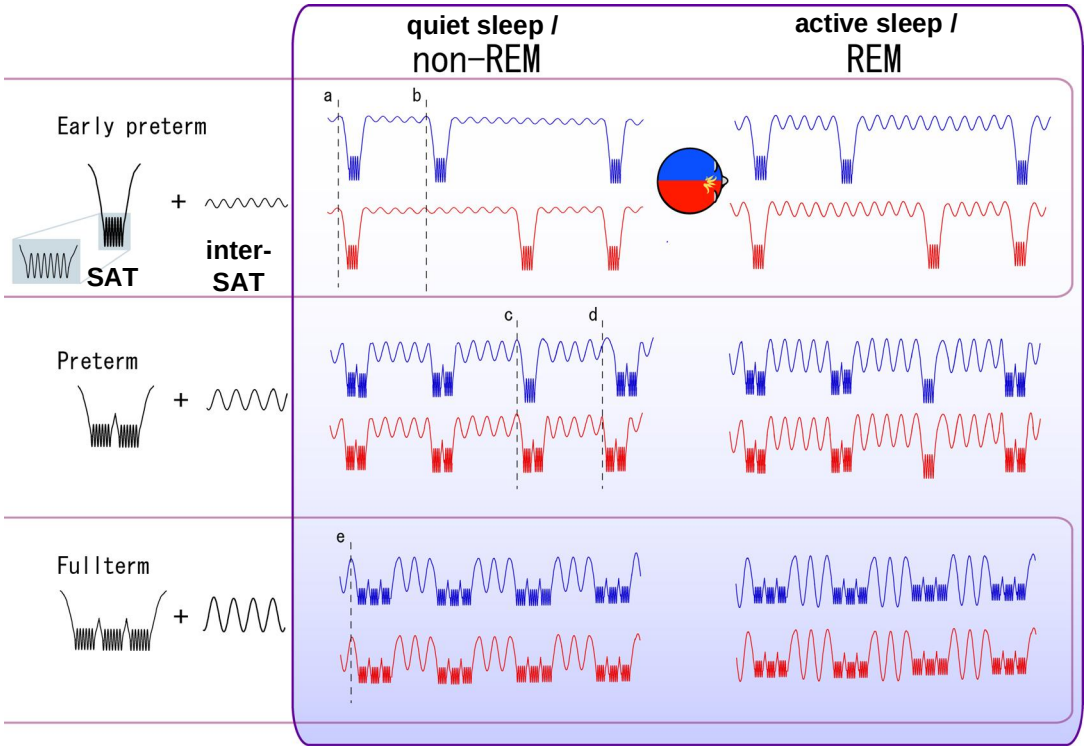

Figure 2. Development of SAT events and ongoing oscillatory activity during inter-SATs in preterm babies. Adapted from (Vanhatalo, Kaila 2006).

In preterm babies, the amount of continuity is also associated with vigilance stages. Traditionally, preterm sleep has been divided into periods of “active sleep” (AS) with more continuous tracings, and “quiet sleep” (QS) with more discontinuous tracings (Vecchierini, Andre & d'Allest 2007). Some researchers believe AS and QS are immature forms of the later recognizable sleep differentiation into rapid eye movement (REM) and non-REM sleep but this is still under debate (Grigg-Damberger et al. 2007).

Considering the physiological background of SATs as described above, the distinction between continuous and discontinuous activity seems somewhat arbitrary. In both continuous and discontinuous periods, SATs and inter-SAT periods follow each other – just their proportions are different: SATs appear with higher frequency during active sleep. Therefore, we believe that numerical measures are needed to describe the quantity of SATs as well as the quality of both SATs and inter-SAT periods.

As has been noted, SATs are a necessary, physiological phenomenon of brain development. Their appearance has a certain similarity with burst-suppression, a pathological EEG pattern seen in term babies after asphyxia (a period of deficient oxygen supply) and even in adults in some conditions. Both preterm EEG and burst-suppression EEG are characterized by an alternation of high amplitude EEG with rather low activity EEG with abrupt changes between these two states. This similarity is interesting as it might allow for methodological transfer from automated detection algorithms for burst suppression to automated detection of SATs.

Above, I have described the normal preterm EEG. Maybe the most important abnormal brain activity in preterm babies is seizures. It was estimated that at least 5% of very preterm babies suffer from seizures (Rennie, Boylan 2007). Development of seizure detection algorithms is an active field in EEG research (Temko et al. 2011, Deburchgraeve et al. 2008, Aarabi, Grebe & Wallois 2007) but is not in the scope of this thesis. It should be noted, however, that a system for continuous monitoring of EEG in preterm babies should include descriptors for both normal brain activity as well as for potential pathological events such as seizures.

2.2 Spectral characteristics of preterm EEG

A traditional way of quantifying EEG is to calculate its frequency content. The calculations are mostly done by Fast Fourier Transformation (FFT) in epochs of a few seconds, and averaged for the duration of the analysed EEG. The spectra are summarized as power values for certain frequency bands, most commonly defined as delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz) and beta (13-30 Hz) (Niemarkt et al. 2011). Sometimes, relative powers are calculated as the proportion of the power in a certain band in relation to the total power.

Division of spectra into the above mentioned frequency bands is based on observed characteristics of adult EEG. For example, oscillations in the alpha range are prominent in adult EEG when the person is awake with eyes closed. In preterm EEGs no such correlates exist. Recently it has even been claimed that neonatal EEG follows a scale-free frequency power distribution with no dominant frequency ranges (Fransson et al. 2012). Despite this qualitative difference in spectral characteristics of preterm EEG, most of the studies cited below report their results using the conventional band division.

The frequency content of EEG can also be described by a single value with a measure called spectral edge frequency (SEF), defined as the frequency below which either 90 % (Inder et al. 2003, West et al. 2006) or 95% (Bell et al. 1991a, Victor et al. 2005) of the total power reside. It is worth mentioning that in some studies (Inder et al. 2003, West et al. 2006) total power was calculated from a spectrum between 2-20 Hz, largely ignoring the most important very low frequencies in the preterm EEG. In other studies (Bell et al. 1991a, Victor et al. 2005), spectrum between 0.3-30 Hz was included in the calculation of total power.

Spectral characteristics are influenced by many temporal processes of different time scales. In preterm infants, at least four time scales of decreasing duration can be identified: the scales of maturation, extrauterine life, sleep stages, and individual SATs/inter-SATs. All these factors – maturation, extrauterine life duration, sleep stage and SAT/inter-SAT period – have a simultaneous influence on the spectra. Standardizing other confounding factors while studying one is not always possible, and sometimes it has not been even tried. Together with technical differences in the calculation of spectra, this leads to a rather heterogenic body of results. Selected re-

sults are summarized in the following subchapters. Table 2 shows a more comprehensive overview of the results.

2.2.1 Maturation

Maturation of the infants brain changes the spectral characteristics of EEG. In the studies summarized in this chapter, three different definitions of age are used: gestational age (GA), defined as the time from conception to the birth and estimated based on clinical data such as ultrasound imaging; postconceptional age (PCA) which is GA plus post-natal age (the time after the birth), as well as post menstrual age (PMA), being the time from last menstruation to the time of observation (Vecchierini, Andre & d'Allest 2007). The difference of these definitions, though clinically relevant, is not important in this context. Changes due to extrauterine life duration are described separately in next subchapter.

Recordings with direct current (DC) –coupled amplifiers have shown that the majority of power in preterm EEG resides well below 1 Hz. These lowest frequencies show a dramatic decrease in power with maturation (Vanhatalo et al. 2005, see Figure 3). In general, absolute power in the lower frequency range decrease with age. The relationship is strongest for delta band (Niemarkt et al. 2011, Bell et al. 1991b, Okumura et al. 2003)

In the higher frequency range, Niemarkt et al. (2011) showed a significant increase of absolute power in beta frequencies with age. Also in Schumacher et al. (2011), relative band power of beta band was higher in a group with older preterm babies. Changes in spectral contents were also reflected in SEF values, which increased with age (Bell et al. 1991a).

2.2.2 Extrauterine life

Neonatal EEG changes also as a function of extrauterine life in the very first days after birth, when the infant adapts to the new environment. Absolute power of the EEG and especially relative power in delta band increased during the first 3-4 days (Victor et al. 2005, Schumacher et al. 2011). In Victor et al. (2005), relative power in delta band increased from 68% on day 1 to 81% on day 4. In accordance with this result, SEF values were reported to decrease significantly during the first three days of life (West et al. 2006, Victor et al. 2005).

Table 2: Studies on dynamics of EEG spectra in preterm infants.

Study	age (n), recording time	filtering	epochs per baby	Changes due to
Vanhatalo 2005	PCA=32-46w (20)	0-NS	3x3min artefact free during discontinuous EEG	*maturation:* 0.01-0.2 Hz down
Niemarkt 2011	GA=29±0.3w (18), weekly recordings for at least 4 weeks starting with end of the first week	0.5-30 Hz	4h per recording except artefactual epochs	*maturation:* delta down, theta down, beta up
Bell 1991 Variation	GA=26-41w (alltogether 60), recording on day 3 GA <32w (20) GA 33-36w (20) GA 37-41w (20)	0.3-30 Hz	4x8s in active sleep	*maturation:* delta down, theta down, beta with quadratic correlation (initially up, then down)
Bell 1991 Spectral	GA=29-41w (51), recording on day3	0.3-30 Hz	4x8s per sleep stage (AS and QS)	*maturation:* SEF up *active -> quiet sleep:* SEF down
Okumura 2003	PCA=29-34w (30)	0.53-30 Hz	6*10s per recording from AS with continuous high voltage slow waves	*maturation:* delta down
Okumura 2006	PCA=29-35w (31)	0.53-30 Hz	6*10s per recording from AS with continuous high voltage slow waves	*maturation:* theta down, alfa no change, beta no change
Tolonen 2007	PCA=32-46w (16)	0-30 Hz	alltogether 797 epochs of 10 s, 515 of these epochs during discontinuous EEG and 282 epochs during continuous EEG. First part of each epoch inter-SAT, second part SAT.	*maturation:* theta up *SAT / inter-SAT:* SATs: in active sleep decrease in RMS due to maturation, inter-SATs: in quiet sleep increase with maturation
Schumacher 2011	GA=24-30w (48), recording started within 12h after birth, duration 3 days	0.5-70 Hz +notch SA: 1-30 Hz	all data (mean: 69.7h per patient) except artefactual epochs	*maturation:* relative beta up, all absolute powers (1-30 Hz) up *extra-uterine life:* total power up, relative delta up, relative theta, alpha, beta down
Victor 2005	GA=24-30w (53), recordings on days 1-4	0.3-70 Hz +notch SA: 0.5-30 Hz	60 min artefact free per recording, 4 recordings per baby	*extra-uterine life:* delta up, others nothing -> relative delta up (81 % on day 4), total power up
Pan et Ogawa 1999	GA=29-35w, PCA=30-39 w (14 babies, 26 recordings)	0.5-60 Hz	24h in 30s epochs	*active -> quiet sleep:* total power down, delta power down (for PCA<=36w)
Havlicek 1975	series with PCA=33, 34, 35, 37 w in one baby, other series (5) starting with 32-35w and ending with 35-40w	0.2-25 Hz	all 5.12 s epochs scored as QS (recording 60-120 min)	*maturation:* total power up *SAT / inter-SAT:* bursts: moderate increase due to maturation, flats: dramatical increase due to maturation, expecially 3.5-7.5 Hz.
Myers 1997	PCA=34-36w (10)	0.5-30 Hz	5-8 hours per infant except artectual epochs	*active -> quiet sleep:* 12-24 Hz down, 0.03-0.2 Hz (calculated on processed signal) up
Paul et al 2003	GA 31-40w (alltogether 21), recording in days 4...10 GA 31-32w (11) GA 39-40w (10)	0.4-50 Hz	5 min of QS and 5 min of AS	*active -> quiet sleep:* low voltage periods: delta, theta, alpha down, high voltage class: all up
West et al 2006	GA 24-31w (62), recording in days 1...7	2-20 Hz (SEF)	60 min per recording, 4 recordings per baby	*extra-uterine life:* SEF down (p=0.06)

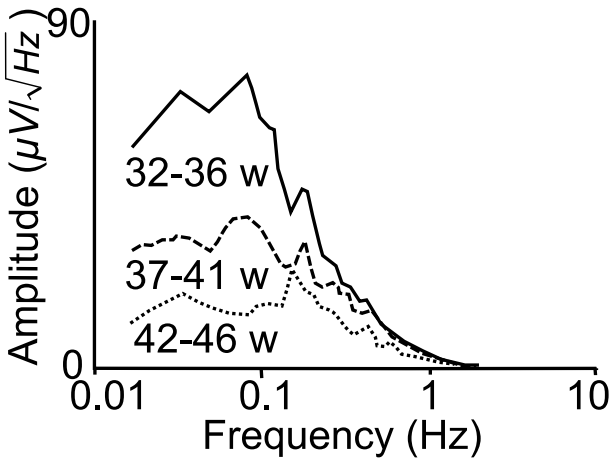

Figure 3. Changes in the lowest frequencies of EEG due to maturation. Note logarithmic frequency scale. EEG spectra are dominated by frequencies below 1 Hz especially in the preterm infants (conceptional age 32-36 weeks). Adapted from Vanhatalo et al. (2005).

2.2.3 Sleep stages

Spectral characteristics of EEG were shown to differ between periods of active and quiet sleep, however, the direction of change was not the same in all studies (see Table 2). Pan and Ogawa (1999) showed a significant decrease in total power from active to quiet sleep. In all infants with PCA≤36 weeks, there was also a significant change in delta power, with smaller values in quiet sleep.

Paul et al. (2003) studied several quantitative descriptors of neonatal EEG in order to find out which of them differed between the sleep stages. The EEG was first segmented into quasi-stationary segments and then the values of the descriptors, calculated from each segment, were averaged in three voltage classes. As preterm EEG consists of SAT and inter-SAT periods, the results of the voltage class with smallest voltages can be associated with inter-SATs and the voltage class with highest voltages can be associated with SATs, even though this was not the writers intention. In preterm babies, power in delta, theta and alpha bands was found to be smaller in quiet sleep than in active sleep in all channels in lowest voltage class (inter-SATs). This result corresponds to the schematic presentation in Figure 2 where the oscillations in inter-SAT periods have much smaller amplitudes during non-REM (quiet) sleep. In the highest voltage class (SATs), the power of alpha band was significantly higher in quiet sleep than active sleep in seven of the eight channels measured. Other bands had higher values in quiet sleep in 4-6 of 8 channels, none had smaller values in quiet sleep. This result could be interpreted as showing that SATs have higher amplitudes during quiet sleep, even though they appear more seldom.

2.2.4 SATs and inter-SATs

Havlicek, Childiaeva and Chernick (1975) studied spectra in manually scored epochs of bursts (SATs) and flats (inter-SATs) during quiet sleep in six preterm babies. In all babies, total power (1.5-25 Hz) increased with age both during the bursts and flats. A detailed study in one of the babies showed that as the powers during SATs increased moderately during the four weeks of study period, powers during inter-SATs increased very dramatically, with biggest change (45 x) in the frequency band 3.5-7.5 Hz. Havlicek et al. concluded that the results "support the notion of two different sources for burst-flat periodicity: bursts originating primarily from subcortical sources and flats reflecting cortical activity." Unfortunately, this early notion of the distinct physiological backgrounds of SATs and inter-SATs hasn't got much attention, with only one citation from this millennium.

Tolonen et al. (2007) studied manually selected SAT and inter-SAT epochs using a direct current (DC) -coupled amplifier in the recordings. Similarly to Havlicek, Childiaeva and Chernick (1975), they found an increasing trend of the RMS values calculated from inter-SAT periods during discontinuous (quiet) sleep with maturation. At the same time, the RMS values of SATs during continuous (active) sleep decreased with maturation.

2.2.5 Illness and medication

Additionally to the complex dynamics of the "normal" preterm EEG, the possibility of illnesses and medications as reason for spectral changes should be considered. Victor et al. (2005) couldn't show any differences in relative powers of EEG between infants with normal cranial ultrasounds and those infants with haemorrhages. Use of morphine as medication seemed to decrease relative power in delta band, but the differences did not reach significance. Norman et al. (2013) showed a general, long lasting depression of preterm background EEG after morphine used as premedication for intubation, however they did not study spectral measures. Additionally to pain medication, antiepileptic drugs are often used in neonates and might have adverse effects on the brain activity in preterm period (Rennie, Boylan 2007).

2.2.6 General remarks to spectral measures

The number of confounding factors in spectral analysis of EEG is large already from the physiological point of view. This makes use of normal values (see e.g. Victor et

al. 2005) questionable, as repeatability of the measures is not easy to achieve. We consider relative powers especially problematic, as they heavily depend on the amplifier as well as frequency bands used in the recording and analysis of the data. In preterm EEG, with most of the energy residing in the lowest frequency range, even small changes in low cut off frequency may cause significant changes in absolute powers of delta band, which again dominates the results of all relative band powers. Dominance of delta band makes relative powers in general difficult to interpret. E.g. Bell et al. (1991b) reported correlation of relative powers of theta, alpha and beta bands with age. However, as they concluded themselves, these are mainly caused by the decrease in the absolute powers of delta band.

Sometimes EEG can be influenced by surprising factors. Sahni et al. (2005) showed that sleeping position (prone vs. supine) can have a significant effect on power measures in preterm infants. Grieve et al. (2005) showed that tilting the bed of newborn infants (PCA 35-41 weeks) by 30 grades significantly increased the power of EEG in certain brain regions. Increase in power was evident throughout the frequency spectrum. These phenomena might be caused by changes in the position of brain relative to scull.

Barlow warned already 1985 (Barlow 1985) of using spectral analysis in studies of discontinuous neonatal EEG. Studying two phenomena (such as SATs and inter-SATs) together results in a spectrum which reflects a mixture of them and thus is useless. In his article Barlow thus proposed to use (automated) segmentation of the discontinuous EEG to calculate separate measures of the bursts and interbursts.

2.3 Occurrence of SATs and inter-SATs

Occurrence of SATs and inter-SATs can be quantified in many ways. The basis of the quantification is always the same: a visual or automated segmentation of the EEG into SAT and inter-SAT periods. If artefacts are not considered, normal preterm EEG is mostly considered to include only these two patterns. Therefore, there must be (almost) same number of SAT and inter-SAT epochs in each EEG.

Duration characteristics are the most widely used quantification methods of inter-SATs (IBI). Different measures are used: minimum, mean or maximum duration as well as percentage levels (e.g. 10^{th}, 50^{th}, 90^{th} percentile, Victor et al. 2005). Some

researchers calculated burst-to-burst intervals (BBI, Pfurtscheller et al. 2008). Bursts are not quantified as often as IBIs, but for them the same measures could be used.

Also the number of SATs or inter-SATs could be calculated. A third correlated measure is the amount of time occupied by SATs or correspondingly, the amount of time occupied by inter-SATs.

Below, general dynamics of occurrence of SATs and inter-SATs in preterm EEG are described. In most of the sited articles, segments were defined manually using certain amplitude and duration criteria. Both in manual and automated detection of SATs and inter-SATs, the exact definition of the patterns affects the segmentation results.

For example, if the definition of IBI requires the amplitudes to be below 15 μV (Hahn, Monyer & Tharp 1989), the resulting IBIs will be much shorter than in studies where IBIs were defined as epochs below 30 μV (e.g. Hayakawa et al. 2001, Victor et al. 2005, Selton, Andre & Hascoet 2000). Some studies to IBIs are summarized in Table 3.

Table 3. Mean and maximal IBI values and proportion of interbursts defined in some studies of preterm infants. Adapted from Victor et al. 2005 and Vecchierini et al. 2007. IBI results given as mean (range) except for Biagioni (mean±SD) and Victor (median, 10^{th}-90^{th} percentile).

Author	Year	Amplitude threshold (μV)	GA (weeks)	Number of infants	Mean IBI (s)	Maximal IBI (s)	IBI / recording time (%)
Connell	1987	ND	26-27	3	14 (9-17)	60 (35-80)	30 (20-50)
			28-29	4	12 (8-16)	50 (25-70)	20 (10-40)
Hahn	1989	< 15	26-27	5	5.4 (5-6)	12 (7-19)	
			28-29	15	5.6 (5-6)	12.5 (4-31)	
Biagioni	1994	<30	27-28	7		30.7±13.7	
			29-30	10		29.5±20.6	
Selton	2000	<30	26	4		46.4	
			27	9		36	
			28	4		26.6	
Hayakawa	2001	<30	25-26	6	13 (10-16)	44.2 (19-76)	48.4 (19-49)
Vecchierini	2003	<15	24-26	10		40 (23-59)	45
Victor	2005	<30	26-27	9	6 (4-13)	15 (11-29)	54 (17-66)
			27-28	10	6 (4-8)	14 (7-22)	39 (8-64)
			29-30	3	5 (5-8)	12 (11-23)	19 (18-56)

2.3.1 Maturation

There is a clear pattern in changes due to maturation. Maximum IBIs get shorter with maturation (Hayakawa et al. 2001, Victor et al. 2005, Hahn, Monyer & Tharp 1989, Selton, Andre & Hascoet 2000, Biagioni et al. 2007, Conde et al. 2005, Connell, Oozer & Dubowitz 1987), but interestingly Biagioni et al. (2007) found that the relationship between maximum IBI and PMA was not linear; it seemed much more logarithmic.

Also mean IBI gets shorter with age (Hayakawa et al. 2001, Victor et al. 2005, Hahn, Monyer & Tharp 1989, Connell, Oozer & Dubowitz 1987, Niemarkt et al. 2010b, Parmelee et al. 1969).

Minimum burst duration was not shown to vary with age (Biagioni et al. 2007) but mean burst duration increased with age (Hayakawa et al. 2001, Parmelee et al. 1969).

Number of IBIs (Hahn, Monyer & Tharp 1989) and, equivalently, number of SATs (Vanhatalo et al. 2005) in a given time window decrease with age.

Pfurtscheller et al. (2008) reported a decreasing trend in burst-to-burst intervals (BBI) with age. If BBI was considered as a time period consisting of a burst and an IBI, this result would be in conflict with results of Hahn, Monyer & Tharp (1989) and Vanhatalo et al. (2005). However, most likely the methodology used in Pfurtscheller et al. did not calculate BBI in this way, but also successive maxima in EEG were recognized as individual bursts. In this way, a continuous train of bursts was interpreted as several short BBIs.

Putting these results together, it is very logical that the percentage of time occupied by IBIs (Hahn, Monyer & Tharp 1989), also called mean interburst-burst ratio (IBR, Niemarkt et al. 2010b)), decreases with age. Or, to say it in one more different but related way: the proportion of discontinuous activity decreases with age, whereas the proportion of continuous activity increases (Vansweden et al. 1991).

2.3.2 Extrauterine life

Niemarkt et al. (2010b) showed that mean IBI decreased with extrauterine days. This was also reflected in decrease of mean interburst-burst ratio with extrauterine age. Van Sweden et al. (1991) showed that proportion of discontinuous activity decreased and proportion of continuous activity increased with extrauterine days.

2.3.3 Sleep state

Paul et al. (2003) used a relatively elaborate methodology to study differences in EEG measures between sleep states. They found that in preterm infants, the number of quasi-stationary segments was in general higher in active sleep than in quiet sleep. Fitting well with this results, quasi-stationary segments of the lowest voltage class (corresponding to inter-SATs, see also chapter 2.2.3) had significantly longer duration in quiet sleep than in active sleep.

Hartley et al. (2012) studied the long-range temporal correlations (LRTC) of "inter-event-intervals" (IEI) which effectively mean almost the same periods as IBI or inter-SAT. They found that even in the youngest preterm babies (GA 23-30 weeks) the IEI showed LRTCs. These temporal fluctuations could be attributable to varying vigilance stages in the study population.

2.3.4 Illness and medication

In preterm infants with major ultrasound brain lesions, mean and maximum IBI were shown to be longer than in an age matched control group, whereas mean and minimum burst durations were shorter than in control group (Conde et al. 2005). Mean IBI was longer in preterm babies with brain injury than in a group without brain injury (Wikström et al. 2008). Medication by morphine prolonged the IBIs (Norman et al. 2013).

3 AUTOMATED METHODS FOR ANALYSIS AND VISUALISATION OF PRETERM EEG

The main parts of this chapter deal with automated segmentation of preterm EEG and different ways to describe the segmentation results. A very influential EEG visualisation method, amplitude integrated EEG is shortly introduced as a prelude to the more advanced methodology. Later also methods utilizing mathematical signal processing but not based on segmentation are presented. The chapter is closed by a discussion of different attempts to cope with artefacts which form a big challenge for any automated method in the clinic.

3.1 Amplitude integrated EEG

In neonatal intensive care units, information of brain's wellbeing is needed constantly, during day and night. However, staff with expertise in EEG reading is not available all the time. One answer to this challenge is use of amplitude integrated EEG (aEEG), first introduced in the sixties by the name cerebral function monitoring (CFM, Maynard, Prior & Scott 1969). At the moment it is the most commonly used EEG measure in NICUs.

aEEG is a trend measure which describes the amplitude of EEG oscillations in a condensed form. In contrast to most methods presented in the following sections aEEG is not based on any segmentation. The EEG is first band pass filtered with an asymmetric filter with cut off frequencies of 2 and 15 Hz. The filtered data undergoes a semilogarithmic transformation which emphasises the low amplitude range. The signal is then rectified and the envelope of this processed signal is plotted with heavy time compression (e.g. 6 cm/h whereas normal EEG is plotted with 3 cm/s).

aEEG is mostly measured with two or four electrodes plus reference, giving one or two traces. In modern devices, not only the aEEG trend but also the raw (unprocessed) EEG is recorded. In this way, periods of suspected abnormal activity can be reviewed in more detail. aEEG can also serve as an additional trend measure in normal, multichannel EEG recordings. Example of an aEEG display is given in Figure 4.

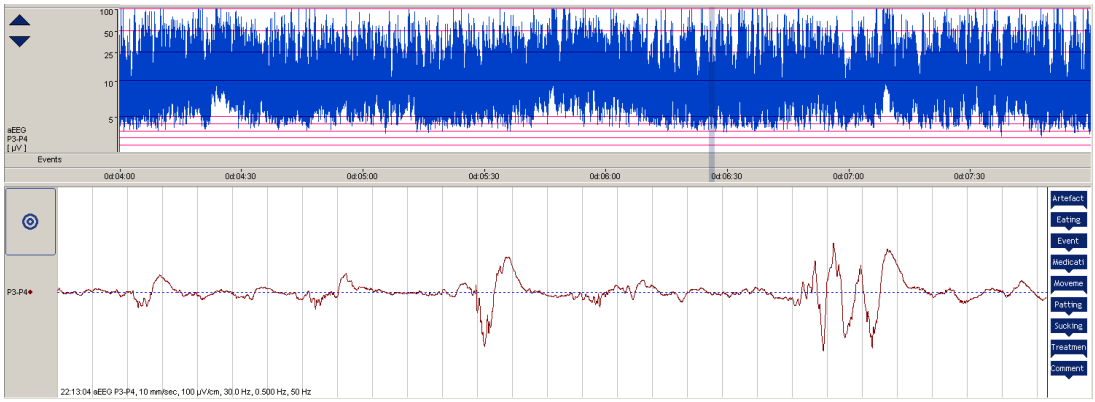

Figure 4: Example of aEEG trend (top panel), 4 hours of data. In most modern devices it is possible to review raw EEG, too (bottom panel, here about 30 seconds).

aEEG is mostly inspected visually but sometimes quantitative analysis is performed manually utilizing published aEEG values for different patterns to describe the aEEG in question (Hellström-Westas, Rosen 2006). Automated aEEG analysis exists but is not yet wide spread (for examples see Bowen, Paradisis & Shah 2010, Niemarkt et al. 2010a).

aEEGs absolute benefit is that it can be interpreted with relatively short training. However, the method is subjective and sensitive for artefacts which might lead to false interpretations. The heavy time compression and the low number of electrodes may leave e.g. neonatal seizures unnoticed (Rennie et al. 2004). The choice of filters also reveals the origin of the method in adult EEG: as frequencies below 2 Hz are filtered out, most of the power in preterm EEG is lost. Therefore, it seems that other methods should be developed instead of aEEG for the special task of preterm EEG analysis (Boylan 2011).

A specific answer to the problem of lost low frequency components in aEEG is range-EEG (rEEG, O'Reilly et al. 2012). In rEEG, total range of EEG values is calculated in 2 s epochs from raw EEG, meaning EEG with no additional filtering after acquisition. The authors claim that in visualization of eg. sleep wake cycles, rEEG shows superior differentiation in comparison to aEEG or other trend measures such as root-mean-square (RMS) calculated in 0.5 s epochs.

3.2 Segmentation based on simple thresholding

Segmentation of preterm EEG can be considered equivalent with detection of SAT and inter-SAT periods. Some methods state this explicitly, whereas others just search for high/low activity epochs, or pseudo-stationary epochs.

The simplest possible segmentation method is to segment EEG according to its amplitude. Wertheim et al. (1991) developed as early as 1991 an automated method to describe the "amount of discontinuity" in preterm EEG. The data was first filtered to 0.5-11 Hz. Then, intervals where absolute amplitude of the EEG remained below 25 µV were identified. A time threshold was also used: only such intervals which lasted at least six seconds where taken into account in the calculation of "interval duration", the percentage of the low amplitude intervals in 1 min analysis epoch. The value of interval duration was shown to correlate well with manually defined amount of discontinuity.

West et al. (2006) tried out four different thresholds on absolute values of preterm EEG filtered between 1-50 Hz: 10, 25, 50 and 100 µV. Opposite to Wertheim, their aim was to measure continuity, defined here as percentage time *above* the given threshold. Based on continuity results from 62 preterm infants they suggested that 50 µV might be best suited as threshold as it gave the continuity values the widest spread. Bowen, Paradisis and Shah (2010) used the same method to demonstrate that the continuity of EEG was decreased in infants with peri/intraventricular hemorrhage.

Pan and Ogawa (1999) calculated low amplitude epochs, defined as epochs of at least 4 s with amplitude below 25 µV, in order to define "discontinuity" as proportion of low amplitude epochs in a given analysis window. Bandwidth of their EEG device was 0.5-60 Hz, no further filtering is reported.

Jennekens et al. (2011) expanded the amplitude thresholding to multichannel data by addition of a channel threshold. Their dataset consisted of 8 infants of 29-34 weeks GA, with 9 recording electrodes. The main idea of the algorithm was as follows: First, the data were filtered to 0.5-32 Hz. Then, envelope of the filtered signal was calculated. Using a threshold, periods with high amplitude activity were defined and number of channels simultaneously showing high amplitude activity was calculated.

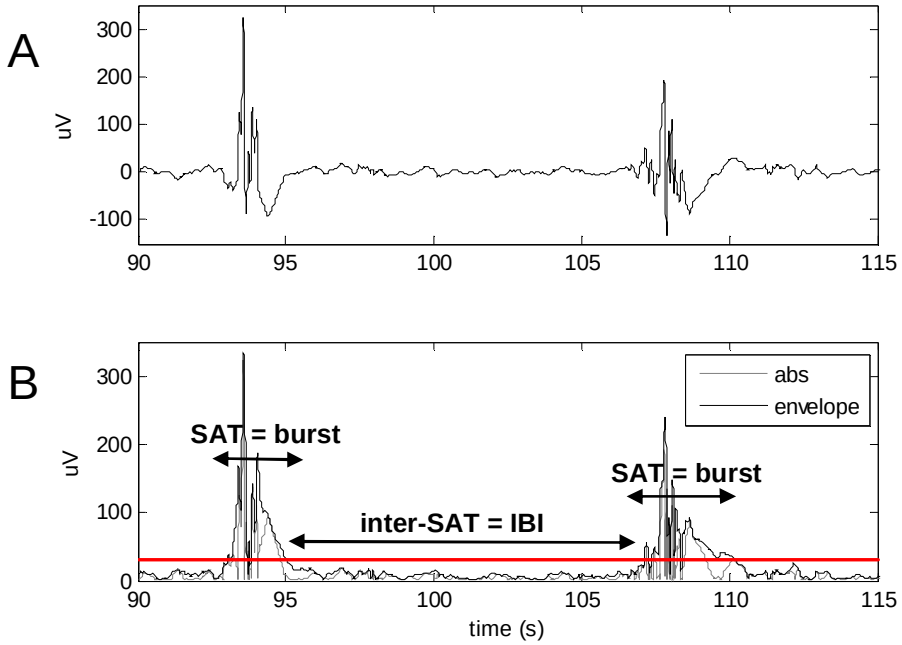

Figure 5. Schematic presentation of segmentation by thresholding. a) 15 s of filtered EEG in a preterm, b) absolute values and envelope of the EEG. An (arbitrary) threshold of 30 μV is shown in red. Periods above this threshold would be considered SATs, periods below the threshold as inter-SATs.

If a channel threshold was met, the segment was defined as either burst or continuous activity depending on its duration. The remaining segments were studied using a separate low amplitude threshold and if very low envelope values were found in all channels for at least 1 s, the segment was classified as IBI. Jennekens et al. used unanimous manual markings by two raters as their gold standard. Here, the methodology resembles the methods used in evaluation of the SAT detection algorithms in Papers I and II of this thesis. However, the optimization in Jennekens et al. was done somewhat simpler than in our Paper II and only utilized half of the available data. I also question the need for different classes for "bursts" and "continuous patterns" in analysis of preterm EEG.

A technically similar approach to Jennekens was followed by Vanhatalo et al. (2005), whereas in their case only one EEG channel was analysed, but this was filtered into 10 frequency bands. In each band, the amplitude envelope was estimated and then normalized with standard deviation (SD) after subtraction of the mean. Values above a threshold of 1.5 SD simultaneously in four or more frequency bands defined the duration of "Multiband activity transients" (MBAT), a mathematical

name for events, which they believed to correspond quite well with the SAT epochs. This method was developed further in my master thesis (Palmu 2008).

Often some preprosessing prior to thresholding is used in order to improve the segmentation results. Särkelä et al. (2002) found that in segmentation of adult burst-suppression pattern, use of non-linear energy operator (NLEO, Plotkin and Swamy 1992) improved the detector performance in comparison to pure amplitude criteria.

NLEO is defined as

$$\Psi_g(x(i)) = x(i-l)x(i-p) - x(i-q)x(i-s), \quad l + p = q + s, \qquad (1)$$

where *i*, *l*, *p*, *q* and *s* are sample indexes and *x(i)* is the signal of interest. Särkelä et al. (2002) used l=0, p=3, q=1 and s=2.

NLEO has often been used for EEG segmentation, however in a more adaptive manner (see also (Agarwal, Gotman 1999) (adults) and (Wong, Abdulla 2006)). The method of Särkelä et al. was the basis for burst-suppression detection implemented in commercial NicOne devices (Cardinal Healthcare, Nicolet Biomedical, Madison,WI) which is further studied in Publication I of this thesis. Shortly summarized, the method was based on calculating output of NLEO in a sliding window of 1 s length from EEG filtered into two frequency bands. The high frequency component (>47 Hz) was used for artefact detection, whereas the lower frequency component (<8 Hz) was used for detection of bursts and suppression with each of these having their own amplitude and time thresholds. In Särkelä et al. (2002), these thresholds were optimized by using short training data from anesthetised adult patients. The amount of suppression was described e.g. by burst-suppression ratio, defined as proportion of suppression time in the analysed epoch.

Niemarkt et al. (2010b) utilized the method in NicOne in a study of preterm infants and calculated additionally to mean IBI also an interburst-burst-ratio (IBR), proportion of IBIs in a sliding window of 10 minutes, as measure of continuity. They further thresholded the IBR and studied the epoch lengths of continuous (IBR<20%) and discontinuous (IBR>20%) EEG.

Hartley et al. (2012) detected bursts of nested oscillations (BNOs) by an algorithm that first calculated amplitude envelope of the signal filtered to three partly overlapping frequency bands (0.5-2 Hz, 8-22 Hz and 2-70 Hz). Each of the envelope signals

was normalized by a formula utilizing literature values of amplitudes in the frequency bands of interest. Product of the three normalized signals was then calculated after taking cube root of each of them. Hartley et al. called the final feature "confidence value on the presence of nested activity". In order to smooth out rapid fluctuations, moving average was counted in 0.5 s window. Finally, a threshold, empirically set at 0.8, was applied and epochs above the threshold were considered to be BNOs.

Pfurtscheller et al. (2008) utilized amplitude variability in segmentation. They calculated spontaneous variance (SV) in sliding window of 1 s length. A threshold, set at 30% percentile, was defined from cumulative density function of SV values from recordings of 34 preterm infants. In each epoch above the threshold, the timepoint with maximum value of SV was found. Burst-to-burst interval (BBI) was then defined as the time between two successive maxima.

3.3 Adaptive segmentation

Most methods discussed above use fixed thresholds for all data. Amplitude values in EEG are influenced by many factors including the electrode distance in bipolar recordings and the position of the electrodes relative to different brain regions, as shown for aEEG by Quigg and Leiner (2009). Also, dynamic changes in appearance of SATs and inter-SATs e.g. due to maturation are huge. This makes use of fixed thresholds suboptimal. Many studies hence proposed use of adaptive segmentation, which uses local characteristics of the signal to segment it in pseudo-stationary (also called quasi-stationary) segments.

Adaptive segmentation is based on comparison of two relatively short time windows, a test window and a reference window to which the test window is compared in some mathematical sense. If the dissimilarity between the two windows grows too large, a segment boundary is initiated, and the process is started from the beginning. Especially the definition of the reference window and how it changes as the pseudo-stationary epoch grows differ from case to case (see Figure 6).

Barlow (1985) was the first to use adaptive segmentation for neonatal EEG. The method used by him was first introduced by (Michael, Houchin 1979). In Barlow's method, the EEG was first filtered to 0.8-30 Hz. Then a difference measure based on autocorrelation functions in reference and test windows was calculated. The difference measure utilized both the changes in amplitude and frequency of the signal.

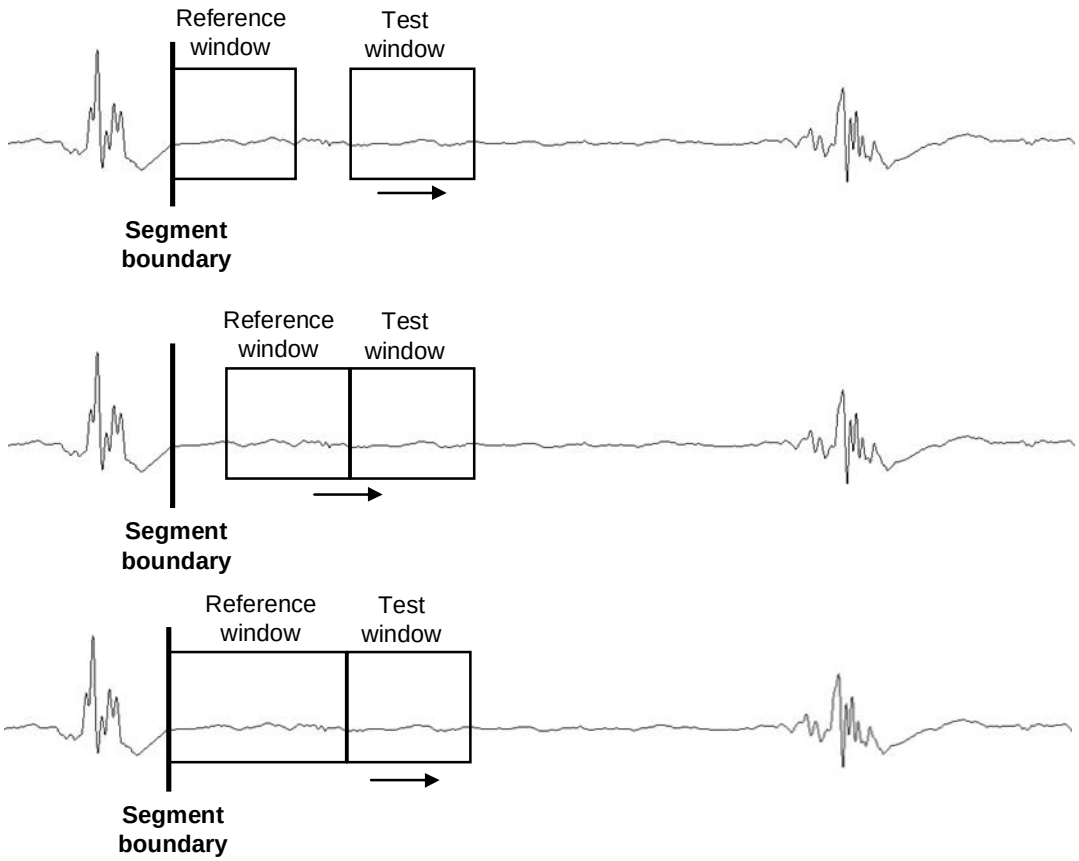

Figure 6: Different definitions of reference and test windows. Reproduced based on (Wong 2008)

Importantly, differences were expressed in percentage so that signals of different scales were handled equally. The reference window was fixed to the beginning of the epoch whereas the test window was sliding along the EEG. If the difference was above a given threshold, the algorithm searched back to find the point at which the change began. Segment boundary was then placed at this point, and a new search was initiated. Barlow reported that the algorithm correctly segmented the burst and interburst periods of discontinuous EEG.

Krajča et al. (1991) proposed an adaptive segmentation method related to that of (Michael, Houchin 1979). It was based on a difference measure, originally presented by (Värri 1988), which was a weighted sum of amplitude and frequency differences between test and reference windows which moved together over the EEG. Here, the amplitude measure ADIF was calculated as a sum of absolute amplitude values in the relevant window of length WL, and the frequency measure FDIF was estimated by the sum of differences of consecutive signal samples in the same window, as defined below:

$$ADIF = \sum_{i=1}^{WL} |x_i| \tag{2}$$

$$FDIF = \sum_{i=1}^{WL} |x_i - x_{i-1}| \tag{3}$$

The difference measure was then defined as

$$G = k_A |ADIF1 - ADIF2| + k_F |FDIF1 - FDIF2| \tag{4}$$

where *ADIF1* and *FDIF1* (*ADIF2* and *FDIF2*) refer to the values of the amplitude and frequency measures in the reference (test) window and coefficients k_A and k_F give the weighting of the amplitude and frequency differences, respectively.

Based on empirical testing, the frequency measure was weighted by a 7 times bigger coefficient than the amplitude measure. Boundaries of quasi-stationary segments were placed at the local maxima of the difference measure. In order to reject small, insignificant maxima in boundary detection, Krajča et al. (1991) introduced an additional adaptive threshold calculated as weighted average of amplitude and frequency measures in the 8 s period preceding the epochs under study. Only local maxima above this threshold were considered. This segmentation method was utilized on neonatal data by (Krajca et al. 2009) as well as (Paul et al. 2003), the results of which have been mentioned in the chapters above. Krajca et al. (1991) didn't report the filtering used in their studies, but in (Paul et al. 2003) and (Krajca et al. 2009) the EEG was filtered to 0.4-50 Hz prior to analysis.

Wong et al. (2006) compared three different segmentation algorithms in case of neonatal EEG. The paper was methodologically orientated and unfortunately did not report the ages of the infants studied. The methods used were spectral error measurement (SEM), generalized likelihood ratio (GLR) and NLEO. SEM and GLR are based on auto-regressive (AR) modelling, whereas NLEO is based on very simple calculations. The segmentation results were evaluated by comparing the segment boundaries with time-frequency presentations of the data. Wong et al. found that GLR was the most appropriate method for segmentation whereas NLEO performed almost as well. It is worth noting, though, that the evaluation method was

highly subjective and based on a very small dataset (11 min 40 s of data altogether). The NLEO method used by Wong was based on the influential papers by Agarwal et al. (1998, 2001) concentrating on adaptive segmentation and clustering of adult EEG. The segment boundaries were set at local maxima of a NLEO based difference function describing change in both amplitude and frequency characteristics of the signal.

3.4 Neural networks

Traditional pattern recognition techniques, such as neural networks (NN), are well suited for EEG segmentation. Galicki et al. (1997) trained a neural network to find the exact onsets of burst patterns in discontinuous neonatal EEG. They stated that finding the exact onset of the bursts was more challenging than finding the end of the bursts (beginning of IBI), which could be done with simple thresholding of power values. In training phase, the adaptive preprosessing unit (APU) and the NN were optimized. Prior knowledge about the "initial wave" of the burst was used in the choice of recording channel (Fp1 and Fp2), the range of possible passbands in the narrow band filter of the APU, and in the choice to use mean momentary power in 8 subsequent 0.5 s intervals of a 4 s sliding window as an input to the NN. Visually classified 4 s epochs (100 bursts and 100 IBIs from 4 neonates) were used as a training set. It could be shown that NN structure with 8 inputs, one hidden layer with 10 neurons and an output layer with 2 neurons were optimal for this task. The optimal mid-frequency of the narrow band filter was 2-5 Hz. Using these settings, all 33 bursts and 34 IBIs of the test set (all from same infant) were classified correctly. Galicki et al. compared their method with (Krajca et al. 1991) and stated a superior accuracy in the detection of burst onset.

Leistritz et al. (1999) proposed a method for automated segmentation of adult EEG into burst, suppression and "remaining" EEG. This method integrated the results of Galicki et al. in a more complex classifier. Suppression periods were first detected using a NN with two inputs (moving median and moving standard deviation of EEG filtered to 0.53-25 Hz in windows of 234 ms). The results of suppression detection were utilized in a more complex NN where other inputs are the SEF (at 95% power) and the momentary power within a narrow frequency band around 10 Hz. This NN had three outputs, one for each possible class burst, suppression and "remaining

EEG". Each timepoint was classified to the class with highest value of the corresponding output neuron. Leistritz et al. quantified and visualized their results by interval-duration-ratio (IDR), which was defined as the duration of the burst pattern divided by the length of the corresponding suppression-burst cycle. Even though technically interesting, the method of Leistritz et al. was optimized for a different type of EEG, where the mere existence of burst and suppression periods had to be detected. Also, the suppression periods of burst-suppression pattern are likely to have even lower amplitudes than the IBIs of preterm EEG, which with ongoing maturation resemble ever more the normal background EEG of mature brain.

3.5 Linear discriminant

Löfhede et al. (2010) presented a sophisticated system for automatic background classification of 8 channel EEG in full-term infants. This system is interesting as it contained as a subsystem a segmentation method for (pathological) burst-suppression pattern.

Epochs of 30 s were first classified into one of the states active awake, quiet awake, active sleep, quiet sleep and burst-suppression. In case of burst-suppression, the epochs were then further segmented into epochs of burst and suppression, which makes presentation of burst-suppression ratio and suppression length possible.

After bandpass filtering to 1.6-44 Hz, several features were used in training of Fisher's linear discriminant (FLD) separately for each of the background states. Löfhede et al. started with 22 measures of power, frequency distribution and entropy. For each of these, summary measures mean, variance, skewness and kurtosis were calculated, leading to 88 "metafeatures". A genetic algorithm was then used to find the optimal subset of 1-10 metafeatures for each of the FLD:s. For each class, the decision signal obtained from FLD was smoothed, and the 30 s epoch was then classified to the state with highest signal.

For burst-suppression segmentation, genetic algorithm was used to find an optimal subset of the original 22 features for an additional FLD. The optimal features incuded e.g. power in a 1 Hz band centered at 3 Hz and variance. Time resolution of the segmentation was 4 Hz.

Löfhede et al. reported a 100% correct separation of burst suppression EEG from the other types of EEG, and in segmentation of burst-suppression the probability of error

was around 4%. However, the dataset used for training and evaluation of burst-suppression detection was small, containing data from only 6 infants, with at least ten bursts each. Additionally, the results of the burst-suppression detection cannot be directly translated into the detection of physiological pattern of SATs and inter-SATs in preterm babies.

Fliesberg et al. (2011) used this method on a dataset of 22 full-term post-asphyctic infants. They used total suppression length per hour and mean IBI to describe the EEG of the babies and concluded that especially the total suppression length per hour discriminated the infants with poor outcome from infant with good outcome.

3.6 Clustering

In most methods summarized in the subchapter 3.1, the EEG was classified into two classes, bursts and IBIs, and the number or duration of epochs of either class or their proportion where used as descriptive factors. In following subchapters, we present ways to process the segmentation results further.

Clustering is based on the idea that there are separate groups in the feature space of some objects. In case of EEG, it is most commonly concerned with classification of (quasi-stationary) segments into some classes which might or might not have some physiological labels. Even the number of the clusters may be open in the beginning of the analysis process.

Barlow (1985) utilized a clustering method presented by (Barlow et al. 1981) to characterize tracé alternant (TA) and REM sleep patterns in neonatal EEG. The method was based on Ward's hierarchical cluster algorithm and produced both numerical and visual summaries of the data. However, the clustering was done for each of the classes TA and REM separately, and manual preselection of these was required. Hence, the method was not fully automated.

Krajča et al. (1991) utilized clustering of quasi-stationary segments to facilitate visual analysis of long-term EEG recordings in adults. In each segment, 10 different features as proposed by Värri 1988 were calculated: average amplitude, variability of segment amplitude, maximum positive and negative values in the segment, maximum values of first and second derivative and amplitudes in delta, theta, alpha and beta spectral bands. A fuzzy c-means clustering algorithm was then used to classify the segments to different clusters. Fuzzy clustering calculated the probability of each

segment to belong to the given cluster. In this way, it became possible to include only the most typical presentations of each cluster to the summary.

Agarwal et al. (1998, 2001) presented another method for the same task. In each quasi-stationary segment, three features per channel were calculated: average amplitude over the segment, sum of absolute differences in the segment (reflecting the average frequency, see also (Krajca et al. 1991)) and average output of the NLEO for the given segment. Using modified iterative k-means clustering, each segment was classified to one cluster. The number of clusters was restricted to 5 and often even less clusters were sufficient to present the variability of adult EEG. Similar numbers of clusters were deemed sufficient for adult EEG also by (Krajca et al. 1991)and (Barlow et al. 1981).

Krajca et al. (2009) used clustering as a pre-processing step for sleep state classification in neonates. Same 10 features as in (Krajca et al. 1991) were used, however, the fuzzy c-mean algorithm was replaced by k-means algorithm. The optimal number of the clusters was defined in an algorithmic way and was between 12 and 18. The clusters were ordered according to their amplitude, with clusters with high amplitudes having bigger cluster numbers. After this, temporal profiles based on momentary cluster number of each of the 8 recording channels were formed. After some signal processing steps including smoothing and averaging over channels, the resulting decision signal was thresholded and periods above the threshold were labelled quiet sleep whereas periods below the threshold were labelled active sleep. The procedure required decisions about several parameters of the classification algorithm. Here, the parameters were set experimentally, separately for the age groups of infants between 31-38 weeks and those with over 40 weeks of PCA. The authors reported high agreement between the algorithm and visual evaluation of the traces. However, these results were given as the number of recordings which agree with expert, whereas a more differentiated evaluation would have been appropriate.

3.7 Principal component analysis of segmentation results

Wong and Abdulla (2008) used the segmentation based on GLR (see above, (Wong, Abdulla 2006)) as pre-processing step for the segmentation of the EEG in epochs of discontinuous and continuous EEG. In a 10 minute sliding window, mean absolute amplitudes of each pseudo-stationary epoch were calculated. The distribution of the amplitude values was highly skewed and was modelled using a log-normal distribu-

tion. Two parameters, the estimated mean and standard deviation of the amplitude values in the log space, were then used for classification.

For classification, linear discriminant analysis was used. A training set, containing 50 manually classified 10 minute segments (25 for both continuous and discontinuous pattern) was used to learn the classifier. It turned out that continuous and discontinuous segments were well separated in the two dimensional parameter space. The classifier was then used on a test set, consisting of 60 recordings of approximately 2 h duration. Unfortunately, no numerical data on the classification accuracy in this dataset was given.

In her PhD (Wong 2008), Wong developed the method even further. Most interesting was the idea to use principal component analysis (PCA) to the above mentioned mean and standard deviation values obtained from a larger dataset of preterm infants with PCA=25...35 weeks. Utilizing the information of the age of the infants and visually defined continuity states (continuous, discontinuous and burst-suppression) in each epoch of the training dataset, Wong showed that the principal component of the transformation correlated with the amount of discontinuity in the EEG epoch, whereas the values along the minor component of the transformation correlated very strongly with the age of the infant. Hence, from each segment two different indices could be obtained: one describing the amount of continuity, and the other describing the maturation.

Wong also presented a way to visualize the results by plotting the mean ± standard deviation of the log transformed distribution over time. The resulting curve showed high similarity with dynamics of aEEG, however being much smoother and less prone to short artefacts.

The results by Wong are highly promising and it can be hoped that the ideas are developed further and also implemented in commercial devices. Considering the details, it may even be unnecessary to segment the data prior to calculation of the amplitude values. In our recent studies (publication in preparation) we have calculated root-mean-square (RMS) values in short, fixed length segments of preterm EEG and observed a very similar distribution of the RMS values as described by Wong for the mean amplitudes.

3.8 Sleep state classification

Myers et al. (1997) proposed to use the ratio of powers in low (0.03-0.20 Hz) and high (12-24 Hz) frequency band to segment preterm EEG in epochs of active and quiet sleep. The low frequency power was calculated from a signal first high pass filtered with cut-off frequency of 6 Hz and then rectified. In this way, the processed signal approximated the envelope of the high frequency bursts, whereas the low frequency deflections which normally appear at the same time were removed. In a dataset of 10 preterm infants recorded at 34-36 weeks PCA, Myers showed that the ratio of low and high frequency band powers was significantly higher in quiet sleep than in active sleep in all of the infants, both when considering a behavioural or an EEG based sleep scoring as reference. The shortcoming of this study is the study population. Infants of 34 weeks or later are quite near to full term and their EEG might be quite different to the extremely preterm infants most commonly seen in NICUs.

Pan and Ogawa (1999) tried out different measures for sleep stage differentiation in preterm infants. Minimum Akaike Information Criterion, total power of EEG, absolute power in delta band (0-3.5 Hz, with hardware filtering 0.5-60 Hz) and amount of discontinuity were found to best differentiate between active sleep, quiet sleep and awake. Using these parameters in multivariate discriminant analysis yielded significant differences between sleep stages in all age groups studied (CA=30...39 weeks).

3.9 Optimization and validation of algorithms

In all algorithms the number and type of inputs must be defined and often there are some parameters such as thresholds to be set. In many studies using automated analysis methods, the choice of inputs (features) and parameters used in the algorithms was reported in an insufficient manner.

Often parameter values such as thresholds were derived from literature (Hartley et al. 2012, Pan, Ogawa 1999, Wertheim et al. 1991) – some algorithms merely mimicked visual segmentation based on strict amplitude and duration criteria. Sometimes parameters seemed to be chosen by trial and error with no further definition of the process (Pfurtscheller et al. 2008, Krajca et al. 1991) whereas sometimes the thresholds were not even reported (Wong, Abdulla 2006). In some studies, a range of studied parameters was reported but the final choice was based on qualitative analysis such as visual inspection of the segmentation results (Vanhatalo et al. 2005, Barlow 1985, Särkelä et al. 2002). West et al. (2006) tried out four different thresholds and sug-

gested that the threshold value leading to largest spread in continuity values be used in future studies.

Sometimes segmentation results were compared with manual markings. In most cases (Särkelä et al. 2002) there was only a single manual marking of the data available. However, Jennekens et al. (2011) used unanimous markings by two raters as their gold standard. Comparisons were often done in indirect manner, e.g. looking at the correlations between average measures derived from both manual and automated classifications (Wertheim et al. 1991).

Jenneken et al. (2011) did proper optimization of their parameters by choosing threshold values that maximize the sensitivity of their algorithm for all three classes defined (burst, interburst, continuous activity).

Often there were no separate training/optimization and testing/validation steps. Performance values obtained in optimization were directly used as performance measures of the algorithm. This is in conflict with the very basics of pattern recognition theory which claims that only unseen data (data not used in the optimization) should be used in the validation. Using the same data for optimization of the algorithm and its validation gives a too optimistic view on the capabilities of the algorithm.

Jennekens et al. (2011) used their algorithm on a separate test data set. The segmentation results were inspected event vise by the human raters and each event was classified either as correct or incorrect. Accuracy of starting time or end time of events were not considered.

Särkelä et al. (2002) used a small subset of two patients as their training data and included these also in the test data. However, the test set was much larger than the training set. Validation results were given as confusion matrixes between automated and manual classifications. Basically each timepoint was treated separately but in case the segment boundaries in automated and manual segmentation differed in time less than 0.2 s the classification is treated as correct.

Wong and Abdulla (2006) validated three different segmentation algorithms by comparing the segment boundaries visually with time frequency distributions of the original signal. Segment boundaries at or near apparent discontinuity in time frequency distribution were considered correct.

3.10 Handling of artefacts

Artefacts are present in all real life EEG recordings, no matter how carefully these were conducted. Artefacts are also one of the main challenges for the wider use of automated analysis methods. Many if not most studies characterizing EEG quantitatively are based on manually selected, artefact free epochs. In such a setting, automated methods are just tools assisting the doctor in the offline analysis. For monitoring use, however, the method must cope also with epochs containing artefacts. Some authors claim that their methods are complex enough to work also with contaminated data (Flisberg et al. 2011) but most attempts are directed into methods automatically finding and either rejecting or correcting artefactual epochs.

Different types of artefacts present in preterm EEG were reviewed in (Walls-Esquivel et al. 2007). Several technical artefacts such as the line noise (50 Hz) are common in NICU where the infants are surrounded by a multitude of technical devices. Physiological artefacts arise from muscle or cardiac activity, respiration, hiccups, sucking, movements of the eyes or movements of the infant. Also handling of the baby causes artefacts.

The most straight forward idea for artefact rejection is to reject epochs with abnormally big amplitudes. Another similar attempt is to use standard deviation in a sliding window. Schetinin, Jakaite and Schult (2011) estimated the probability distribution of standard deviation in each recording using a 10 s sliding window. A threshold was then defined as the sum of the mode of the distribution and the maximal deviation from the mode times a constant, empirically defined to be 0.225. Segments with deviation above this threshold were rejected as artefacts.

Myers et al. (1997) used a spectral power measure to reject artefactual epochs. The EEG was divided into 1 min segments and in each the total EEG power was calculated. Utilizing the distribution of the total power values, segments where the total power exceeded the 75^{th} percentile by at least 1.5 times the interquartile range were rejected as artefacts. This is a commonly used threshold in outlier detection.

Schumacher et al. (2011) studied total power summed over 8 recording channels separately for four frequency bands. They compared median values of total power between a large dataset of preterm EEG with manually performed artefact rejection in the same dataset, when either 5, 10, 15 or 20% of highest total power values were rejected, a procedure also called trimming of the data. Schumacher et al. reported a

nonsignificant change in total power values when 5% trimming was used for artefact removal and hence used this method in their further analysis.

Agarwal et al. (1998) used several artefact rejection measures parallel. First, a fixed amplitude threshold of 300 μV was applied. Segments where absolute maximum amplitude in any of the recording channels of one lateral side exceeded this value were rejected. Second, a measure of frequency weighted energy was used, defined as the average output of the NLEO in a segment. A dynamic threshold not further specified was applied to find the artefacts . Finally, if the EEG segment was defined artefactual in one lateral side, it was rejected also in the contralateral side.

Interestingly, in (Agarwal, Gotman 2001) at least the second method was dropped and artefact rejection was based on amplitude criterion only.

Inder et al. (2003) attempted to reject technical artefacts by utilizing the continuous impedance measurements available in their recording device. All one minute intervals in which the electrode impedance was below 20 kΩ per pair and the total power between 2 and 20 Hz was above 10 μV^2 were included in the analysis. Also (Jennekens et al. 2011) and (Niemarkt et al. 2010b) rejected periods based on impedance measure, using an even lower threshold of 10 kΩ. West et al. (2005) used 15 kΩ and Schumacher et al. (2011) used 40 kΩ as rejection threshold.

Sometimes the algorithms are designed to be less prone for artefacts. For example, Jennekens et al. (2011) required at least 4 of 18 channels to show high amplitudes typical for bursts in order for a burst detection to take place. In this way, artefacts affecting just some single channel did not influence the detection. However, with this approach real local phenomena remain undetected.

Särkelä et al. (2002) included artefact as a separate outcome to their classification algorithm additional to burst and suppression. Artefact was the default output of the system. Using processed signals in two different frequency bands and separate thresholds for burst, suppression and artefacts, they achieved a high level of accuracy in the classification of the adult burst-suppression, including the correct definition of artefactual epochs.

Algorithms for neonatal seizure detection often include automated artefact detection. For example, Aarabi, Grebe and Wallois (2007) used several techniques parallel to detect different kinds of artefacts. Due to the need of manually selected artefact tem-

plates and/or multichannel recordings the techniques are not directly useable in NICU environment.

In summary, several methods for artefact identification have been proposed. Based on the special characteristics of the preterm EEG with very high amplitude bursts, the simple amplitude thresholding might not be a good choice for this kind of data (see also (Walls-Esquivel et al. 2007). Many of the methods build on statistical measures derived from the data and are therefore not directly suited for online use. A detailed study on artefact rejection methods would be needed but is not easy to conduct as this would need a large dataset with manually marked artefacts of different types. As the definition of an artefact might especially in the milder cases not be self evident the manual marking should be done by several raters in order to obtain a reliable data set for comparison and optimization of the methods. Additionally, artefacts are not really what doctors are interested in, they are just the annoying disturbance blurring their view to the real brain phenomena. This might partly explain the use of rather straightforward artefact rejection methods in many papers dealing with neonatal EEG analysis.

4 PATIENTS AND METHODS

Numerical analysis of the data was performed mostly with Matlab (Version 7.6.0, MathWorks, Natick, MA, USA) and in some cases with Excel (2003, Part of Microsoft Office).

4.1 Patients

The dataset used in both Publications I and II consists of EEG recordings from 18 preterm infants (GA=23-30 weeks, mean 26 weeks). EEG recordings were done during the first three days of life with P3-P4 derivation (midline reference) using a Nervus/NicOne 3.3 EEG system with U16 amplifier (Cardinal Healthcare, Nicolet Biomedical, Madison,WI). Passband of the amplifier started at 0.16 Hz, in our recordings sampling rate of 256 Hz was used. From each infant, an 11 minutes long epoch of good quality EEG was chosen for further analysis.

In Publication I, this dataset was divided into two groups with 12 extremely preterm infants (GA 23-27 weeks) and 6 very preterm infants (GA 28-30 weeks).

4.2 Markings

Three medical doctors, all experienced EEG readers, marked all SATs in the dataset on visual basis. The marking did not involve any prior epoch definition: the markings could start and stop anywhere in time. Inter-rater agreement of the doctors was assessed using confusion matrices. Additionally, proportion of overall agreement (Fleiss 1971) was evaluated. It is useful as a single measure of agreement between multiple raters.

Epochs where all three raters agreed on the definition of the EEG as either SAT or inter-SAT built the basis for the assessment of the spectral properties of SATs. These unanimously marked epochs were also used as gold standard in the validation and further development of the automated SAT detection algorithm in both Publications I and II.

4.3 Properties of SATs

In Publication I, distribution of SAT durations in manual markings as well as spectra of both SAT and inter-SAT epochs from unanimous markings were studied. Three indexes (number of SATs, average duration of SATs and proportion of time covered by SATs, SAT%) derived from the visual detections were calculated in order to com-

pare how robust they were for inter-rater differences. A significant linear correlation of index values between the raters was considered a sign for robustness of an index.

4.4 Validation of a commercially available method

In NicOne devices, a method originally developed for detection of adult burst-suppression during anaesthesia (see (Särkelä et al. 2002)) is implemented. This method has been used also for the assessment of SATs and inter-SATs in preterm babies (Niemarkt et al. 2010b, Wikström et al. 2008).

In Publication I, I implemented this method in Matlab as closely as possible imitating the NicOne implementation, and validated it utilizing the gold standard as described in 4.2. Sensitivity, specificity and accuracy of the automated detection were used as performance measures.

4.5 Further development of the method

In Publication II, I streamlined the above mentioned algorithm and solved some of its problems. The algorithm has several numerical parameters, the choice of which affects the detection results. I set one of these, the minimum duration of SATs, to one second using the knowledge from the manual markings. Other four parameters of the streamlined algorithm (low and high cut off frequency of the EEG filters, length of a smoothing window and threshold for SAT detection) were set by probing a four dimensional parameter space and by choosing the parameter combination which maximized the average detection rate (ADR, average of sensitivity and specificity). ADR was assessed by comparing the detection results to the gold standard. Real performance values of the algorithm when used on unknown data were estimated by using leave-one-out cross validation. In this publication, all 18 EEGs were considered as one group.

5 SUMMARY OF THE RESULTS

5.1 Inter-rater agreement

There was a high agreement on SATs between the raters. Proportion of overall agreement was 86% in the group of extremely preterm babies and 81% in the group of very preterm babies. 80% of the extremely preterm data were rated unanimously as either SAT or inter-SAT by all three raters. For the very preterm babies, the unanimous ratings covered 71% of the recording time. Epochs of unanimous ratings built the gold standard.

When studying the changes in different SAT indexes between the babies, it was noticed that only SAT% values showed strong (R^2 between 0.67 and 0.98), significant ($p<0.05$) correlation between all raters.

5.2 Properties of SATs

As already mentioned, very few SATs were shorter than 1 s. Mostly SATs lasted for less than 10 s (see Figure 5 in Publication I). With increasing maturation, the SATs became longer.

Spectra on SAT and inter-SAT periods revealed that SATs had remarkably higher amplitudes through all frequencies (see Figure 5 in Publication I). SAT spectrum also showed a distinctive “hump” in frequencies between 3-8 Hz.

5.3 SAT detection

5.3.1 Commercial algorithm

Accuracy (proportion of correctly classified samples) of the NicOne algorithm was on average 87% and 79% for extremely preterm and very preterm babies, respectively. Sensitivity was low, 64% and 69%. Specificity of the algorithm varied very much between the babies and was on average 96% and 88%.

5.3.2 Optimized algorithm

In leave-one-out cross-validation, the ADR of the optimized algorithm was estimated as 95.8% ± 2.3 % (mean ± SD). Sensitivity was estimated as 96.6 ± 2.8 % and specificity as 95.1 ± 5.6 %. In the optimization, all babies were studied as one group.

6 DISCUSSION AND FUTURE PERSPECTIVES

6.1 Inter-rater agreement

To our knowledge, we were the first group studying inter-rater agreement in visual detection of SATs. Our results show that the agreement is relatively high. However, for around 20 or 30% of time (extremely preterm and very preterm babies respectively), all three raters did not agree on the definition of EEG as either SAT or inter-SAT. The difference comes mainly from the exact definition of the onset and end of each SAT (see also Figure 2 in Publication I) and affects the duration of both the SATs as well as inter-SATs. This is interesting as lots of literature on preterm EEG is based on manual definition of inter-SATs, or IBIs. With reference to our study (Publication I), Jennekens et al. (2011) compared manual markings of bursts and IBIs by two observers. They report 66±3% and 70±6% agreement for bursts and IBIs, respectively. Exact definition of "agreement" is not given but the values seem comparable or slightly worse than those obtained in our study.

6.2 Properties of SATs and inter-SATs

In the Publications of this thesis, no inter-SAT values were reported. Some parameters calculated subsequently are presented in table 4. The values are calculated from recordings of each age group pooled together. The values show quite some differences between raters but in general they are comparable with previous results presented in Table 3.

Our own studies as well as (Jennekens et al. 2011) relied on visual perception in manual markings. Numerical amplitude criteria have also been proposed but most likely these, too, are originally based on visual analysis. The exact choice of amplitude criteria affects the definition of event onset times and therefore also the duration of IBIs.

Definition of trains of SATs as either several distinct events or one event of longer duration radically affects such descriptive values as mean SAT or inter-SAT duration, or the number of SAT events. This problem is true for both manual and automated detection of SATs and may be one reason for large variations in IBI values in literature (see e.g. Table 3). Figure 7 shows an example of this situation when using an automated SAT detection method (unpublished) for DC recorded EEG with

Table 4: Inter-SAT durations in the dataset of this thesis

Group	mean inter-SAT duration, s (rater A/B/C)	maximal inter-SAT duration in the dataset, s (rater A/B/C)
extremely preterm (GA 23-27 w)	10.1 / 9.4 / 6.4	77.7 / 76.2 / 51.9
very preterm (GA 28-30 w)	6.3 / 4.4 / 7.2	37.5 / 24.8 / 41.5

different thresholds. Lacking correlation of SAT numbers and SAT durations between the raters in our Publication I is one more example of the same problem.

We therefore feel that indexes based on mean duration or number of either SAT or inter-SAT are not the optimal choice for description of preterm EEG. We prefer SAT%, the proportion of time covered by SATs, which is more robust for small differences in visual or numerical thresholds. The robustness of SAT% was also seen in our Publication I as the SAT% values of different raters were highly correlated.

SAT% can be considered identical with measuring continuity (e.g. West et al. 2006). The opposing concept is that of discontinuity proposed e.g. by Wertheim et al. (1991, "interval duration"), Wikström et al. (2012, "IB%") and Niemarkt et al. (2010b, 2010a, "interburst-burst ratio"). Notably, the information content is the same no matter whether the proportion of SATs or inter-SATs is considered. It seems more logical to measure something that happens and not the time nothing happens but for the clinical use this choice is irrelevant. The only problem with competing definitions is the confusion they might bring about.

6.3 SAT detection

In Publication I, we validated an algorithm for SAT detection available in commercial devices. We found out that the sensitivity of the algorithm was very low. This was easily explained by the algorithm, as it did not classify the first 1-2 seconds of each SAT as SAT but remained in the preceding inter-SAT state.

In Publication II, we presented a streamlined algorithm which corrected this and some other problems of the commercial algorithm, and optimized it. The optimized algorithm showed excellent performance in the leave-one-out cross-validation which used separate data for training and testing. However, even the optimized algorithm has some serious limitations.

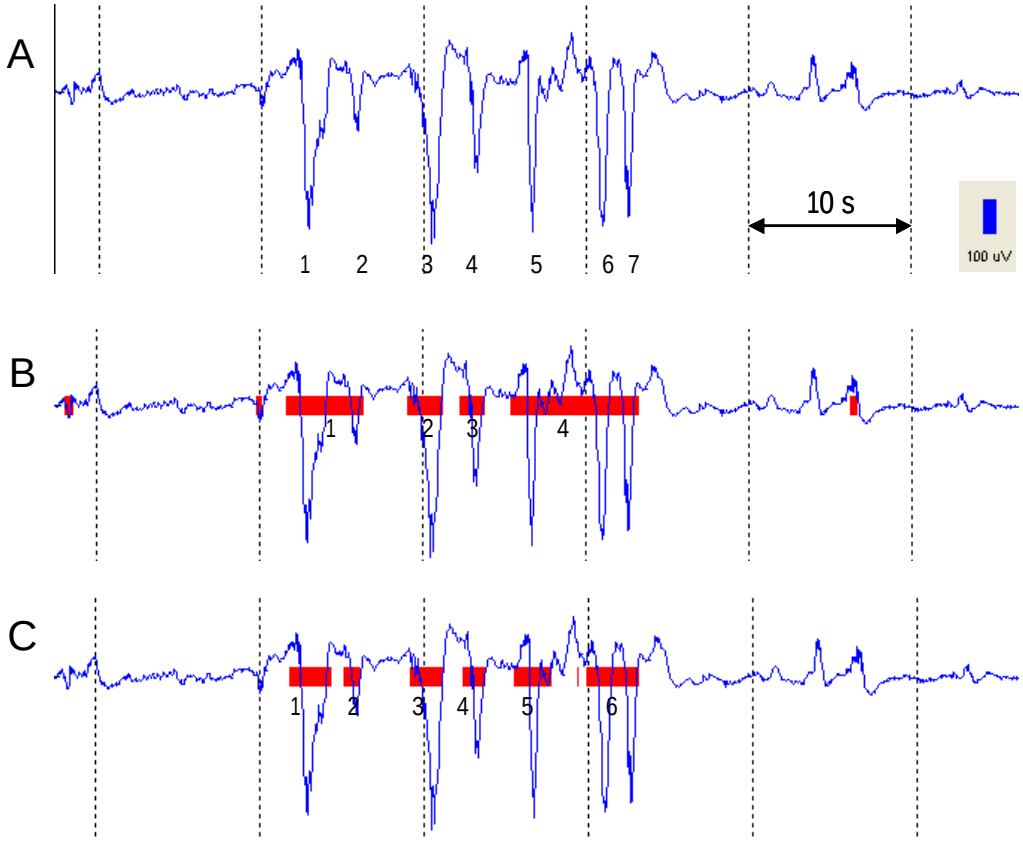

Figure 7: Example of automated detection of SATs with different thresholds. Threshold has a strong effect on SAT number and mean SAT duration but a smaller effect on SAT%. A: EEG with 7 visually clear SATs in the middle. B: Result of automated SAT detection algorithm (unpublished). Using a low threshold, some of the SATs are merged together in the automated detection. Four SAT periods lasting at least one second are detected. C: Higher threshold leads to separation of SATs in the automated detection. Six SAT periods are detected.

Most importantly, the detection algorithm is amplitude dependent and does not adapt to the data. It is optimized for data collected from a certain electrode derivation with a certain amplifier. It is known (see e.g. Quigg, Leiner 2009, Lamblin et al. 1999) that EEG amplitudes depend on electrode location, interelectrode distance, and the frequency band-pass of the amplifier.

The algorithm was optimized with recordings from extremely and very preterm babies. It might not work well in older babies where with the brain's maturation the difference between SAT and inter-SAT periods has diminished. However, it is exactly the extremely and very preterm babies that most urgently need brain monitoring. Therefore we consider our algorithm useful despite this limitation.

Our algorithm is based on the assumption that the EEG consists of SATs and inter-SATs only. In sick babies there might be also pathological EEG phenomena such as seizures. Detection of seizures in neonates is a task of its own and has been addressed recently by several groups, however mostly based on full term EEG (see e.g. Temko et al. 2011, Deburchgraeve et al. 2008, Aarabi, Grebe & Wallois 2007). In future, detections of seizures and SATs might run on monitoring devices in parallel.

The optimized algorithm proposed by us does not have any in-built artefact handling. The baseline correction of the algorithm attempts to minimize the effect of continuous artefacts such as electrocardiography (ECG) artefact or artefacts caused by mechanical ventilation. But most artefacts in neonatal EEG come from movements of the baby and will need a separate rejection scheme in the future.

It is not straightforward to compare the performance of our optimized algorithm with other algorithms developed for the same or similar task. Main reason for this is the general lack of properly done validation in the earlier published methods (see also chapter 3.9), or the different purpose of the algorithm (e.g. burst-suppression detection). The methodology in (Jennekens et al. 2011) is best comparable with ours which is understandable as they refer to Publication I in their study. Even though less stringent in definition of correct classification, the sensitivity of the algorithm in Jennekens et al. is reported to be lower than the sensitivity of our algorithm. This might be partly caused by the unnecessary division of preterm EEG into bursts, IBI and continuous pattern, whereas continuous pattern most likely just contains trains of bursts.

In general we question the use of “gold standards” based on a single visual detection in development of algorithms for SAT detection. In our case, we decided to use only those epochs rated unanimously for our gold standard. In this way, the optimization is based on “clean” samples of SAT and inter-SAT. The algorithm will define the border between these two classes with its own logic which is systematic and reproducible.

6.4 Future perspectives

Segmenting preterm EEG into SAT and inter-SAT epochs is just the first step in automated EEG analysis. As a task it could be compared to detection of R-peaks in QRS complexes of electrocardiography: a reliable detection is essential but actual analysis is done on features obtained by further processing of the detection results.

Calculating SAT% is already a step further. Wikström et al. (2012) showed that when measured at 24 h of age, a high IB% (basicly the same measure as SAT% from different view point) correctly prognosed the poor outcome at age of 2 years in 79% of the affected infants.

But one of the most promising directions for further research in our group is the use of SAT% to detect changes in vigilance stages. It has been shown that even very preterm infants (PCA=23-30 weeks) show long-range temporal correlations in their brain activity (Hartley et al. 2012). A stable sleep wake cycle (SWC) is more common in infants with good outcome (Wikström et al. 2012). In infants with intra-/periventricular haemorrhages, sleep wake cycles are more common with smaller haemorrhages (Olischar et al. 2007). SWC is also considered a sign for more advanced structural maturation of the brain. However, visual analysis of SWC based e.g. on aEEG is highly subjective and dependant on expertise of the staff available. Automated analysis of SWC could bring additional benefit from the use of long term EEG monitors in NICU. We have discovered that fluctuation of SAT% matches strikingly well with the fluctuation of vigilance stages defined by polysomnographic recording (publication in preparation). An example of SAT% fluctuation with time is shown in Figure 8.

We have also studied the relation of measures derived from automated SAT-detection with the structural development of the brain. In 21 preterm infants (CA=25-34 weeks), we found a significant negative correlation between maximal inter-SAT duration and several structural measures obtained from MRI studies done both preterm and at term equivalent age (publication in preparation). Less mature brain structure is thus correlated with longer inter-SAT periods, a logical result that corresponds well with earlier results about changes in IBI with maturation.

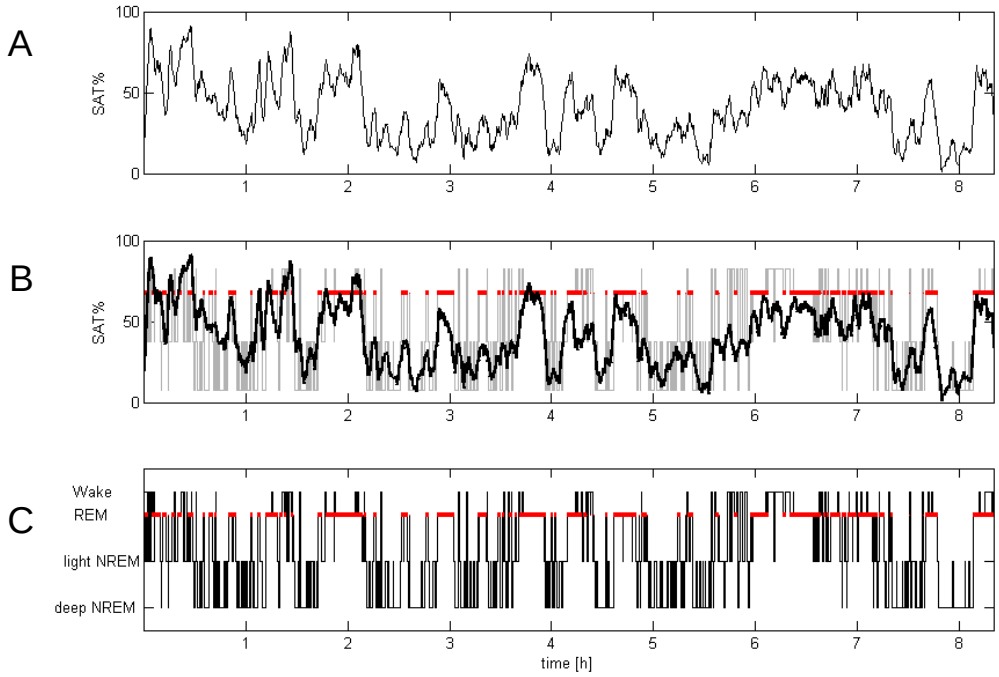

Figure 8: A. Example of SAT% fluctuation with time in an infant at conceptional age of 26 weeks. SAT% was calculated in 3 min epochs. C: Manual scoring of sleep stages. B: SAT% and scoring results overlaid. Note how SAT% increases during REM epochs and is very low during deep NREM epochs. The infant in this example is quite sick and demands mechanical ventilation. Sleep is fractionary with many transitions between sleep stages. Nevertheless several sleep cycles can be observed with cycle length of approximately half an hour.

For the usability of an automated detection of SATs some way of handling the artefacts is essential. I have developed a scheme for rejecting short periods of EEG based on RMS values. RMS values are first calculated in 5 s epochs and median RMS in whole epoch under study is evaluated. Five times this median value is then used as threshold and all 5 s epochs with higher values are rejected. In the dataset of this thesis, this threshold led to 3.8% of epochs to be rejected when manually marked artefacts were not considered in the analysis. The advantage of this method is that it adapts to the data and the threshold is not affected by artefacts as long as they cover less than 50% of the data. However, calculation of median as in my implementation is done afterwards and therefore the method is not directly applicable to real time monitoring. Further improvement of the method is thus needed.

7 CONCLUSION

Spontaneous activity transients (SATs) are a distinctive feature of preterm EEG. In this thesis I have shown that SATs are perceived in a relatively consistent way by different individuals. More importantly, I have also developed an algorithm that automatically detects SATs with high accuracy.

Reliable automated detection of SATs is a starting point for further development that is already under way. In future, detection of SATs may be a routine part of the monitoring paradigms in NICUs, with SAT% serving both as a trend measure and a feature for detection of sleep-wake cyclicity. Brain monitoring may become as normal as monitoring heart rate. This will make better care of vulnerable preterm infants possible. And better care means better prospects for a normal life without disability.

ACKNOWLEDGEMENTS

Work on this thesis was supported by the follwing foundations: Lastentautien säätiö and Emil Aaltosen säätiö. I also want to thank all co-authors in Publications I and II for their ideas, support and pleasant co-operation in general. Special thanks go to my instructor Sampsa Vanhatalo for his enthusiasm.

REFERENCES

Aarabi, A., Grebe, R. & Wallois, F. 2007, "A multistage knowledge-based system for EEG seizure detection in newborn infants", *Clinical Neurophysiology,* vol. 118, no. 12, pp. 2781-2797.

Agarwal, R. & Gotman, J. 2001, "Computer-assisted sleep staging based on segmentation and clustering", *Proceedings of the 23rd Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-4: Building New Bridges at the Frontiers of Engineering and Medicine,* vol. 23, pp. 1695-1698.

Agarwal, R. & Gotman, J. 1999, "Adaptive segmentation of electroencephalographic data using a nonlinear energy operator", *In Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, volume 4, pp. 199 – 202* .

Agarwal, R., Gotman, J., Flanagan, D. & Rosenblatt, B. 1998, "Automatic EEG analysis during long-term monitoring in the ICU", *Electroencephalography and clinical neurophysiology,* vol. 107, no. 1, pp. 44-58.

Andre, M., Lamblin, M.-., d'Allest, A.M., Curzi-Dascalova, L., Moussalli-Salefranque, F., Tich, S.N.T., Vecchierini-Blineau, M.-., Wallois, F., Walls-Esquivel, E. & Plouin, P. 2010, "Electroencephalography in premature and full-term infants. Developmental features and glossary", *Neurophysiologie Clinique-Clinical Neurophysiology,* vol. 40, no. 2, pp. 59-124.

Barlow, J. 1985, "Computer Characterization of Trace Alternant and Rem-Sleep Patterns in the Neonatal Eeg by Adaptive Segmentation - an Exploratory-Study", *Electroencephalography and clinical neurophysiology,* vol. 60, no. 2, pp. 163-173.

Barlow, J., Creutzfeldt, O., Michael, D., Houchin, J. & Epelbaum, H. 1981, "Automatic Adaptive Segmentation of Clinical Eegs", *Electroencephalography and clinical neurophysiology,* vol. 51, no. 5, pp. 512-525.

Bell, A., McClure, B., McCullagh, P. & McClelland, R. 1991a, "Spectral Edge Frequency of the Eeg in Healthy Neonates and Variation with Behavioral State", *Biology of the neonate,* vol. 60, no. 2, pp. 69-74.

Bell, A., McClure, B., McCullagh, P. & McClelland, R. 1991b, "Variation in Power Spectral-Analysis of the Eeg with Gestational-Age", *Journal of Clinical Neurophysiology,* vol. 8, no. 3, pp. 312-319.

Biagioni, E., Frisone, M.F., Laroche, S., Kapetanakis, B.A., Ricci, D., Adeyi-Obe, M., Lewis, H., Kennea, N., Cioni, G., Cowan, F., Rutherford, M., Azzopardi, D. & Mercuri, E. 2007, "Maturation of cerebral electrical activity and development of cortical folding in young very preterm infants", *Clinical Neurophysiology,* vol. 118, no. 1, pp. 53-59.

Blencowe, H., Cousens, S., Oestergaard, M.Z., Chou, D., Moller, A., Narwal, R., Adler, A., Garcia, C.V., Rohde, S., Say, L. & Lawn, J.E. 2012, "National, regional, and worldwide estimates of preterm birth rates in the year 2010 with

time trends since 1990 for selected countries: a systematic analysis and implications", *Lancet,* vol. 379, no. 9832, pp. 2162-2172.

Bowen, J.R., Paradisis, M. & Shah, D. 2010, "Decreased aEEG Continuity and Baseline Variability in the First 48 Hours of Life Associated With Poor Short-Term Outcome in Neonates Born Before 29 Weeks Gestation", *Pediatric research,* vol. 67, no. 5, pp. 538-544.

Boylan, G.B. 2011, "EEG monitoring in the neonatal intensive care unit: A critical juncture", *Clinical Neurophysiology,* vol. 122, no. 10, pp. 1905-1907.

Conde, J., de Hoyos, A., Martinez, E., Campo, C., Perez, A. & Borges, A. 2005, "Extrauterine life duration and ontogenic EEG parameters in preterm newborns with and without major ultrasound brain lesions", *Clinical Neurophysiology,* vol. 116, no. 12, pp. 2796-2809.

Connell, J., Oozer, R. & Dubowitz, V. 1987, "Continuous 4-Channel Eeg Monitoring - a Guide to Interpretation, with Normal Values, in Preterm Infants", *Neuropediatrics,* vol. 18, no. 3, pp. 138-145.

Deburchgraeve, W., Cherian, P.J., De Vos, M., Swarte, R.M., Blok, J.H., Visser, G.H., Govaert, P. & Van Huffel, S. 2008, "Automated neonatal seizure detection mimicking a human observer reading EEG", *Clinical Neurophysiology,* vol. 119, no. 11, pp. 2447-2454.

Fleiss, J. 1971, "Measuring Nominal Scale Agreement among Many Raters", *Psychological bulletin,* vol. 76, no. 5, pp. 378-382.

Flisberg, A., Kjellmer, I., Lofhede, J., Lindecrantz, K. & Thordstein, M. 2011, "Prognostic capacity of automated quantification of suppression time in the EEG of post-asphyctic full-term neonates", *Acta Paediatrica,* vol. 100, no. 10, pp. 1338-1343.

Fransson, P., Metsäranta, M., Blennow, M., Åden, U., Lagercrantz, H. & Vanhatalo, S. 2012, "Early Development of Spatial Patterns of Power-Law Frequency Scaling in fMRI Resting-State and EEG Data in the Newborn Brain", *Cereb Cortex.,* vol. Epub ahead of print.

Galicki, M., Witte, H., Dorschel, J., Eiselt, M. & Griessbach, G. 1997, "Common optimization of adaptive preprocessing units and a neural network during the learning period. Application in EEG pattern recognition", *Neural Networks,* vol. 10, no. 6, pp. 1153-1163.

Grieve, P., Myers, M., Stark, R., Housman, S. & Fifer, W. 2005, "Topographic localization of electrocortical activation in newborn and two- to four-month-old infants in response to head-up tilting", *Acta Paediatrica,* vol. 94, no. 12, pp. 1756-1763.

Grigg-Damberger, M., Gozal, D., Marcus, C.L., Quan, S.F., Rosen, C.L., Chervin, R.D., Wise, M., Picchietti, D.L., Sheldon, S.H. & Iber, C. 2007, "The Visual Scoring of Sleep and Arousal in Infants and Children", *J Clin Sleep Med.,* vol. 3, no. 2, pp. 201-240.

Hahn, J., Monyer, H. & Tharp, B. 1989, "Interburst Interval Measurements in the Eegs of Premature-Infants with Normal Neurological Outcome", *Electroencephalography and clinical neurophysiology,* vol. 73, no. 5, pp. 410-418.

Hartley, C., Berthouze, L., Mathieson, S.R., Boylan, G.B., Rennie, J.M., Marlow, N. & Farmer, S.F. 2012, "Long-Range Temporal Correlations in the EEG Bursts of Human Preterm Babies", *Plos One,* vol. 7, no. 2, pp. e31543.

Havlicek, V., Childiaeva, R. & Chernick, V. 1975, "Ontogeny of Eeg Power Characteristics of Quiet Sleep Periodic Cerebral Rhythm in Preterm Infants", *Neuropadiatrie,* vol. 6, no. 2, pp. 151-161.

Hayakawa, M., Okumura, A., Hayakawa, F., Watanabe, K., Ohshiro, M., Kato, Y., Takahashi, R. & Tauchi, N. 2001, "Background electroencephalographic (EEG) activities of very preterm infants born at less than 27 weeks gestation: a study on the degree of continuity", *Archives of Disease in Childhood,* vol. 84, no. 3, pp. F163-F167.

Hellström-Westas, L. & Rosen, I. 2006, "Continuous brain-function monitoring: State of the art in clinical practice", *Seminars in Fetal & Neonatal Medicine,* vol. 11, no. 6, pp. 503-511.

Inder, T., Buckland, L., Williams, C., Spencer, C., Gunning, M., Darlow, B., Volpe, J. & Gluckman, P. 2003, "Lowered electroencephalographic spectral edge frequency predicts the presence of cerebral white matter injury in premature infants", *Pediatrics,* vol. 111, no. 1, pp. 27-33.

Jennekens, W., Ruijs, L.S., Lommen, C.M.L., Niemarkt, H.J., Pasman, J.W., van Kranen-Mastenbroek, V.H.J.M., Wijn, P.F.F., van Pul, C. & Andriessen, P. 2011, "Automatic burst detection for the EEG of the preterm infant", *Physiological Measurement,* vol. 32, no. 10, pp. 1623-1637.

Krajca, V., Petranek, S., Mohylova, J., Paul, K., Gerla, V. & Lhotska, L. 2009, "Modeling the Microstructure of Neonatal EEG Sleep Stages by Temporal Profiles", *13th International Conference on Biomedical Engineering, Vols 1-3,* vol. 23, no. 1-3, pp. 133-137.

Krajca, V., Petranek, S., Patakova, I. & Värri, A. 1991, "Automatic Identification of Significant Graphoelements in Multichannel Eeg Recordings by Adaptive Segmentation and Fuzzy Clustering", *International journal of bio-medical computing,* vol. 28, no. 1-2, pp. 71-89.

Lamblin, M., Andre, M., Challamel, M., Curzi-Dascalova, L., d'Allest, A., De Giovanni, E., Moussalli-Salefranque, F., Navelet, Y., Plouin, P., Radvanyi-Bouvet, M., Samson-Dollfus, D. & Vecchierini-Blineau, M. 1999, "EEG in premature and full-term newborns. Maturation and glossary", *Neurophysiologie Clinique-Clinical Neurophysiology,* vol. 29, no. 2, pp. 123-219.

Leistritz, L., Jager, H., Schelenz, C., Witte, H., Putsche, P., Specht, M. & Reinhart, K. 1999, "New approaches for the detection and analysis of electroencephalographic burst-suppression patterns in patients under sedation", *Journal of clinical monitoring and computing,* vol. 15, no. 6, pp. 357-367.

Lofhede, J., Thordstein, M., Lofgren, N., Flisberg, A., Rosa-Zurera, M., Kjellmer, I. & Lindecrantz, K. 2010, "Automatic classification of background EEG activity in healthy and sick neonates", *Journal of Neural Engineering,* vol. 7, no. 1, pp. 016007.

Marlow, N., Wolke, D., Bracewell, M., Samara, M. & EPICure Study Grp 2005, "Neurologic and developmental disability at six years of age after extremely preterm birth.", *New England Journal of Medicine,* vol. 352, no. 1, pp. 9-19.

Maynard, D., Prior, P. & Scott, D. 1969, "Device for Continuous Monitoring of Cerebral Activity in Resuscitated Patients", *British medical journal,* vol. 4, no. 5682, pp. 545-546.

Michael, D. & Houchin, J. 1979, "Automatic Eeg Analysis - Segmentation Procedure Based on the Autocorrelation Function", *Electroencephalography and clinical neurophysiology,* vol. 46, no. 2, pp. 232-235.

Mwaniki, M.K., Atieno, M., Lawn, J.E. & Newton, C.R.J.C. 2012, "Long-Term Neurodevelopmental Outcomes After Intrauterine and Neonatal Insults: A Systematic Review", *Obstetrical & gynecological survey,* vol. 67, no. 6, pp. 345-346.

Myers, M., Fifer, W., GroseFifer, J., Sahni, R., Stark, R. & Schulze, K. 1997, "A novel quantitative measure of Trace-alternant EEG activity and its association with sleep states of preterm infants", *Developmental psychobiology,* vol. 31, no. 3, pp. 167-174.

Niemarkt, H.J., Andriessen, P., Peters, C.H.L., Pasman, J.W., Blanco, C.E., Zimmermann, L.J. & Oetomo, S.B. 2010a, "Quantitative Analysis of Amplitude-Integrated Electroencephalogram Patterns in Stable Preterm Infants, with Normal Neurological Development at One Year", *Neonatology,* vol. 97, no. 2, pp. 175-182.

Niemarkt, H.J., Andriessen, P., Peters, C.H.L., Pasman, J.W., Zimmermann, L.J. & Oetomo, S.B. 2010b, "Quantitative analysis of maturational changes in EEG background activity in very preterm infants with a normal neurodevelopment at 1 year of age", *Early human development,* vol. 86, no. 4, pp. 219-224.

Niemarkt, H.J., Jennekens, W., Pasman, J.W., Katgert, T., van Pul, C., Gavtlanes, A.W.D., Kramer, B.W., Zimmermann, L.J., Oetomo, S.B. & Andriessen, P. 2011, "Maturational Changes in Automated EEG Spectral Power Analysis in Preterm Infants", *Pediatric research,* vol. 70, no. 5, pp. 529-534.

Norman, E., Wikström, S., Rosen, I., Fellman, V. & Hellström-Westas, L. 2013, "Premedication for Intubation With Morphine Causes Prolonged Depression of Electrocortical Background Activity in Preterm Infants", *Pediatr Res.,* vol. 73, no. 1, pp. 87-94.

Okumura, A., Kubota, T., Toyota, N., Kidokoro, H., Maruyama, K., Kato, T., Hayakawa, F. & Watanabe, K. 2003, "Amplitude spectral analysis of maturational changes of delta waves in preterm infants", *Brain & development,* vol. 25, no. 6, pp. 406-410.

Olischar, M., Klebermass, K., Waldhoer, T., Pollak, A. & Weninger, M. 2007, "Background patterns and sleep-wake cycles on amplitude-integrated electroe-

ncephalography in preterms younger than 30 weeks gestational age with peri-/intraventricular haemorrhage", *Acta Paediatrica,* vol. 96, no. 12, pp. 1743-1750.

O'Reilly, D., Navakatikyan, M.A., Filip, M., Greene, D. & Van Marter, L.J. 2012, "Peak-to-peak amplitude in neonatal brain monitoring of premature infants", *Clinical Neurophysiology,* vol. 123, no. 11, pp. 2139-2153.

Palmu, K. 2008, *Spontaanien aktiviteettipurskeiden automaattinen tunnistaminen keskoslasten aivosähkömittauksesta (Detection of spontaneous activity transients in electroencephalygraphy of preterm infants)*, master thesis, Helsinki University.

Pan, X. & Ogawa, T. 1999, "Microstructure of longitudinal 24 hour electroencephalograms in healthy preterm infants", *Pediatrics International,* vol. 41, no. 1, pp. 18-27.

Parmelee, A., Akiyama, Y., Stern, E. & Harris, M. 1969, "A Periodic Cerebral Rhythm in Newborn Infants", *Experimental neurology,* vol. 25, no. 4, pp. 575-584.

Paul, K., Krajca, V., Roth, Z., Melichar, J. & Petranek, S. 2003, "Comparison of quantitative EEG characteristics of quiet and active sleep in newborns", *Sleep medicine,* vol. 4, no. 6, pp. 543-552.

Pfurtscheller, K., Bauernfeind, G., Mueller-Putz, G.R., Urlesberger, B., Mueller, W. & Pfurtscheller, G. 2008, "Correlation between EEG burst-to-burst intervals and HR acceleration in preterm infants", *Neuroscience letters,* vol. 437, no. 2, pp. 103-106.

Plotkin, E.I. & Swamy M.N.S. 1992, "Nonlinear signal processing based on parameter invariant moving average modelling", *Canadian Conference on Electrical and Computer Engineering,* vol.1, pp. TM.3.11.1.-11.4.

Quigg, M. & Leiner, D. 2009, "Engineering Aspects of the Quantified Amplitude-Integrated Electroencephalogram in Neonatal Cerebral Monitoring", *Journal of Clinical Neurophysiology,* vol. 26, no. 3, pp. 145-149.

Rennie, J. & Boylan, G. 2007, "Treatment of neonatal seizures", *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 92, no. 2, pp. F148-F150.

Rennie, J., Chorley, G., Boylan, G., Pressler, R., Nguyen, Y. & Hooper, R. 2004, "Non-expert use of the cerebral function monitor for neonatal seizure detection", *Archives of Disease in Childhood,* vol. 89, no. 1, pp. F37-F40.

Sahni, R., Schulze, K., Kashyap, S., Ohira-Kist, K., Fifer, W. & Myers, M. 2005, "Sleeping position and electrocortical activity in low birthweight infants", *Archives of Disease in Childhood,* vol. 90, no. 4, pp. F311-F315.

Särkelä, M., Mustola, S., Seppänen, T., Koskinen, M., Lepola, P., Suominen, K., Juvonen, T., Tolvanen-Laakso, H. & Jäntti, V. 2002, "Automatic Analysis and Monitoring of Burst Suppression in Anesthesia", *J Clin Monit Comput.,* vol. 17, no. 2, pp. 125-134.

Schetinin, V., Jakaite, L. & Schult, J. 2011, "Informativeness of Sleep Cycle Features in Bayesian Assessment of Newborn Electroencephalographic Maturation", *2011 24th International Symposium on Computer-Based Medical Systems (Cbms),* .

Schumacher, E.M., Westvik, A.S., Larsson, P.G., Lindemann, R., Westvik, J. & Stirs, T.A. 2011, "Feasibility of Long-Term Continuous EEG Monitoring During the First Days of Life in Preterm Infants: An Automated Quantification of the EEG Activity", *Pediatric research,* vol. 69, no. 5, pp. 413-417.

Selton, D., Andre, M. & Hascoet, J. 2000, "Normal EEG in very premature infants: reference criteria", *Clinical Neurophysiology,* vol. 111, no. 12, pp. 2116-2124.

Temko, A., Thomas, E., Marnane, W., Lightbody, G. & Boylan, G. 2011, "EEG-based neonatal seizure detection with Support Vector Machines", *Clinical Neurophysiology,* vol. 122, no. 3, pp. 464-473.

Tolonen, M., Palva, J.M., Andersson, S. & Vanhatalo, S. 2007, "Development of the spontaneous activity transients and ongoing cortical activity in human preterm babies", *Neuroscience,* vol. 145, no. 3, pp. 997-1006.

Vanhatalo, S., Palva, J., Andersson, S., Rivera, C., Voipio, J. & Kaila, K. 2005, "Slow endogenous activity transients and developmental expression of K+-Cl- cotransporter 2 in the immature human cortex", *European Journal of Neuroscience,* vol. 22, no. 11, pp. 2799-2804.

Vanhatalo, S. & Kaila, K. 2010, "Emergence of spontaneous and evoked electroencephalographic activity in the human brain" in *The Newborn Brain: Neuroscience and Clinical Applications*, eds. H. Lagercrantz, M.A. Hanson, L.R. Ment & D.M. Peebles, 2nd edn. edn, Cambridge University Press, pp. 229-242.

Vanhatalo, S. & Kaila, K. 2006, "Development of neonatal EEG activity: From phenomenology to physiology", *Seminars in Fetal & Neonatal Medicine,* vol. 11, no. 6, pp. 471-478.

Vansweden, B., Koenderink, M., Windau, G., Vanderbor, M., Vanbel, F., Vandijk, J. & Wauquier, A. 1991, "Long-Term Eeg Monitoring in the Early Premature - Developmental and Chronobiological Aspects", *Electroencephalography and clinical neurophysiology,* vol. 79, no. 2, pp. 94-100.

Värri, A. 1988, *Digital Processing of the EEG in Epilepsy*, Licentiate thesis, Tampere University of Technology.

Vecchierini, M.-., Andre, M. & d'Allest, A.M. 2007, "Normal EEG of premature infants born between 24 and 30 weeks gestational age: Terminology, definitions and maturation aspects", *Neurophysiologie Clinique-Clinical Neurophysiology,* vol. 37, no. 5, pp. 311-323.

Victor, S., Appleton, R., Beirne, M., Marson, A. & Weindling, A. 2005, "Spectral analysis of electroencephalography in premature newborn infants: Normal ranges", *Pediatric research,* vol. 57, no. 3, pp. 336-341.

Walls-Esquivel, E., Vecchierini, M.F., Heberle, C. & Wallois, F. 2007, "Electroencephalography (EEG) recording techniques and artefact detection in early

premature babies", *Neurophysiologie Clinique-Clinical Neurophysiology,* vol. 37, no. 5, pp. 299-309.

Wertheim, D., Eaton, D., Oozer, R., Connell, J., Dubowitz, L., Dubowitz, V., Willetts, R. & Wootton, R. 1991, "A New System for Cotside Display and Analysis of the Preterm Neonatal Electroencephalogram Rid G-4921-2011", *Developmental medicine and child neurology,* vol. 33, no. 12, pp. 1080-1086.

West, C., Harding, J., Williams, C., Gunning, M. & Battin, M. 2006, "Quantitative electroencephalographic patterns in normal preterm infants over the first week after birth", *Early human development,* vol. 82, no. 1, pp. 43-51.

West, C., Nolan, M., Williams, C., Harding, J., Dezoete, J. & Battin, M. 2005, "Comparison of quantitative measures and neurophysiologist assessment using cotside EEG monitors to predict outcome in preterm infants", *Pediatric research,* vol. 58, no. 2, pp. 425-425.

Wikström, S., Ley, D., Hansen-Pupp, I., Rosen, I. & Hellström-Westas, L. 2008, "Early amplitude-integrated EEG correlates with cord TNF-alpha and brain injury in very preterm infants", *Acta Paediatrica,* vol. 97, no. 7, pp. 915-919.

Wikström, S., Pupp, I.H., Rosen, I., Norman, E., Fellman, V., Ley, D. & Hellström-Westas, L. 2012, "Early single-channel aEEG/EEG predicts outcome in very preterm infants", *Acta Paediatrica,* vol. 101, no. 7, pp. 719-726.

Wong, L. & Abdulla, W.H. 2008, "Automatic detection of preterm neonatal EEG background states", *2008 Ieee International Conference on Acoustics, Speech and Signal Processing, Vols 1-12,* , pp. 421-424.

Wong, L. 2008, *Quantitative Continuity Feature for Preterm Neonatal EEG Signal Analysis*, PhD thesis, University of Auckland.

Wong, L. & Abdulla, W. 2006, *Time-frequency evaluation of segmentation methods for neonatal EEG signals, In Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1303 –1306*.