An algorithm for the automatic detection of seizures in neonatal amplitude-integrated EEG

CML Lommen (c.lommen@mmc.nl)1,4, JW Pasman2, VHJM van Kranen3, P Andriessen1, PIM Cluitmans4, LGM van Rooij5, S Bambang Oetomo1

1. Maxima Medical Centre Veldhoven, Department of Neonatology, The Netherlands
2. Radboud University Nijmegen Medical Centre, Clinical Neurophysiology, The Netherlands
3. University Hospital Maastricht, Clinical Neurophysiology, The Netherlands
5. University Medical Centre Utrecht, Neonatology, The Netherlands

Abstract

Aim: To develop and evaluate an algorithm for the automatic screening of electrographic neonatal seizures (ENS) in amplitude-integrated electroencephalography (aEEG) signals.

Methods: CFM recordings were recorded in asphyxiated (near) term newborns. ENS of at least 60 sec were detected based on their characteristic pattern in the aEEG signal, an increase of its lower boundary. The algorithm was trained using five CFM recordings (training set) annotated by a neurophysiologist, observer 1. The evaluation of the algorithm was based on eight different CFM recordings annotated by observer 1 (test set observer 1) and an independent neurophysiologist, observer 2 (test set observer 2).

Results: The interobserver agreement between observer 1 and 2 in interpreting ENS from the CFM recordings was high (G coefficient: 0.82). After dividing the eight CFM recordings into 1-min segments and classification in ENS or non-ENS, the intraclass correlation coefficient showed high correlations of the algorithm with both test sets (respectively, 0.95 and 0.85 with observer 1 and 2). The algorithm showed in five recordings a sensitivity ≥ 90% and approximately 1 false positive ENS per hour. However, the algorithm showed in three recordings much lower sensitivities: one recording showed ENSs of extremely high amplitude that were incorrectly classified by the algorithm as artefacts and two recordings suffered from low interobserver agreement.

Conclusion: This study shows the feasibility of automatic ENS screening based on aEEG signals and may facilitate in the bedside interpretation of aEEG signals in clinical practice.

INTRODUCTION

For the neurological assessment of critically ill newborns electroencephalography (EEG) is an important tool. The clinical relevant characteristics of the EEG are the background pattern and electrographic neonatal seizures (ENS) (1).

Seizures are one of the most distinct signs of cerebral dysfunction in newborns (2). They may be caused by most common cerebral pathologies of the newborn like hypoxic ischemic encephalopathies, intracranial hemorrhages, metabolic encephalopathies, cerebral infarctions or cerebral infections (2). They often have an early postnatal onset and may be first indicators of cerebral disease (2,3). Repetitive seizures may cause disturbances in ventilation and a rise in blood pressure (2), and may cause damage to the developing brain (4,5).

For the detection of ENSs in newborns electroencephalography is a sensitive method and considered as the golden standard. Clinical signs of seizures in newborns are often subtle, difficult to differentiate from normal newborn movements, or even absent (silent seizures) (2,6). Especially after medication seizures are often clinically silent (7). In neonatal EEG the electrical cortical activity is measured using 9–21 electrodes connected to the scalp of the newborn (8,9). ENSs are characterized by rhythmic discharges of waveforms of similar morphology. They have a relatively sudden onset and termination as well as an evolution in frequency, topology and amplitude (9,10). They may either be focal, multifocal or global and are generally defined to have a minimum duration of 5 or 10 sec (1).

Practical considerations make the EEG less suitable for long-term monitoring of the brain function in the neonatal intensive care unit (NICU) environment. This is mainly due to the time consuming interpretation of the numerous signals. The complex nature of the signals makes interpretation by experts necessary. Therefore, EEG measurements at the NICU are generally performed for the duration of about 45–90 min.

The cerebral function monitor (CFM) (11) facilitates long-term monitoring of cerebral electrical activity at the NICU (12). The CFM measures only one channel, biparietal EEG. This EEG is converted into an amplitude-integrated EEG (aEEG), which shows clinically useful information even when displayed at 6 cm/h, resulting in a clear overview of several hours on a monitor. The amplitude of the signal shows the relative power of cerebral electrical activity. The CFM is relatively easy to install and to interpret by doctors
and nursing staff with limited expertise, and therefore suitable for long-term monitoring (13).

CFM is predominantly used to detect ENSs outside the period of the EEG measurement and outside office hours. Using CFM only generalized and focal parietal ENSs will be measured. However, focal ENSs often migrate or spread to other areas in the cortex (8) and then can be detected by CFM, as reported in a study by Toet et al. (14).

Still some difficulties in the interpretation of the CFM recording remain, especially concerning the detection of ENSs. Because of hypersynchronisation, i.e. the simultaneous discharge of large groups of neurons, seizures cause a rise in the amplitude of the aEEG signal (6,13). A similar rise is known to be caused by different types of artefacts (15,16). Furthermore, it is often difficult to detect ENSs of short duration and/or low amplitude (17,18).

The objective of this study was to develop an algorithm as a screening tool for the automatic detection of ENSs in aEEG signals of newborns. The algorithm should be able to recognize artefacts that can be mistaken for ENSs. The purpose of the algorithm is to assist doctors and nurses in the interpretation of the aEEG signal and as an alarm function on the CFM monitor. This study describes the algorithm and its evaluation.

**METHODS**

**Subjects**

In this study 13 CFM recordings were included of 13 newborns. Ten of these newborns were full term and 10 of them were diagnosed with asphyxia. The clinical details of these newborns are given in Table S1.

**Recordings**

Recordings have been obtained using the CFM6000 (Olympic Medical, Seattle, WA, USA). These digital CFM recordings contain both the EEG and aEEG signal. For these measurements needle electrodes were placed parietal at locations P3 and P4 according to the international 10–20 system (15).

CFM measurements were a part of standard clinical monitoring of severely ill neonates, the main indication being detection of epileptic seizure activity. Therefore, the medical ethics committee did not consider it necessary to evaluate this study. Measurements have been performed between June 2003 and November 2004 at the Máxima Medical Centre Veldhoven, the Radboud University Nijmegen Medical Centre, and at the Wilhelmina Children’s Hospital Utrecht, the Netherlands and one recording at the Hammersmith Hospital in London, UK. Inclusion of the recordings was based on their variability in types of background patterns (19) and presence of ENSs, given for each recording in Table S1.

The presence of ENSs is specified in either repetitive ENSs or status epilepticus, i.e. an ENS that is so prolonged or frequently repeated as to create a fixed or lasting epileptic condition (9). The duration of the recordings varied from 9 to 24 h (total: 222 h). Figure 1 shows 100 min of recording 1, containing repetitive ENSs, characterized mainly by the sudden increase of the lower boundary.

**The algorithm**

In this study only ENSs with a duration of at least 60 sec were included. The minimum duration of ENSs is generally defined to be 5 or 10 sec (1). However, due to the compressed nature of the aEEG signals, these ENSs may not be recognizable. Moreover, ENSs of short duration may be difficult to differentiate from bursts. The goal of the algorithm is to be a screening tool for the detection of ENSs, to help neonatologists in detecting non-clinical seizures that might have damaging effects on the brain. Therefore we arbitrarily decided to start with only including ENSs with a duration of at least 60 sec.

The algorithm has been written in the mathematical program MATLAB® (The MathWorks, Massechusetts, USA).

Before detecting ENSs, artefacts were detected and rejected in a way similar to artefact detection of the CFM6000 monitor. In the monitor amplifier saturation and high impedance are marked as artefact. During amplifier saturation the aEEG signal reaches its maximum amplitude. In the algorithm these parts of the signal were classified as artefact as well. The signal is also classified as artefact when the impedance exceeds 20 kΩ, like in the monitor. Since an impedance beneath 5 kΩ is preferred (13), our algorithm marked the aEEG signal corresponding to impedances between 5 and 20 kΩ.

The ENS detection was based on the main characteristic change in the aEEG signal due to ENSs; a sudden increase of the lower boundary of the signal (Fig. 1). A lower boundary is preferred (13), our algorithm marked the aEEG signal corresponding to impedances between 5 and 20 kΩ.

The ENS detection was based on the main characteristic change in the aEEG signal due to ENSs; a sudden increase of the lower boundary of the signal (Fig. 1). A lower boundary was defined every 10 sec of the signal as the 10th percentile of the samples, see Figure 2A. The reference boundary, i.e. the lower boundary of the background pattern, was calculated for each segment of 10 sec. It was defined as a weighted average of the lower boundaries over the last 6 min. ENSs and artefacts that occurred during the last 6 min were not included in the calculation of the reference boundary. Figure 2B gives an example of this reference boundary.

Subsequently, a sudden rise of the lower boundary of the aEEG signal was classified as a pattern of interest. The
algorithm detected a pattern of interest whenever the lower boundary of the aEEG signal was at least an empirically determined threshold (EDT) higher than the reference boundary for a period of at least 60 sec. Whenever an artefact was detected during this increase, the parts of the pattern of interest not classified as artefact needed to have a total duration of at least 60 sec. Otherwise, the increase would be considered as artefact. The EDT value has been determined using a training set of five CFM recordings (recordings with ID A–E) annotated by a neurophysiologist.

Patterns of interest in the training set were caused by either ENSs or muscle activity. Muscle activity is characterized in the EEG signal by its high frequent fluctuations, containing frequencies reaching up to over 100 Hz (9). Cerebral electrical activity only contains frequencies of up to 50 Hz (1). Accordingly a pattern of interest containing in the EEG signal a high percentage of frequency power above 50 Hz was classified as muscle artefact. All other patterns of interest were classified as ENSs.

Evaluation of the algorithm
In this study it was intended to evaluate the ability of the algorithm to detect ENSs in CFM recordings, not the ability of CFM recordings to detect ENSs compared to standard EEG. Therefore evaluation was based on the CFM recording, containing both 1-channel EEG and aEEG signals.

Two neurophysiologists, observer1 and observer2, visually examined the CFM recordings for the presence of ENSs with a minimum duration of 60 sec. The examinations were performed independently with no preliminary consultation between the observers. They were not aware of the diagnosis of the newborns. Clear ENSs were annotated accordingly as well as patterns where the observer doubted if they were ENSs. Annotations were based mainly on the one-channel EEG signal displayed at 2.5 cm/sec and the aEEG and impedance signals, displayed at 6 cm/h. There were no options for scaling or frequency analysis.

From the two neurophysiologists, observer1 annotated the full length of the recordings and observer2 the first 6 h of each of the recordings. The recordings were divided into a training set and two test sets (Table S1):

- Test set 1: recordings 1–8, annotated by observer1.
- Test set 2: the first 6 h of recordings 1–8, annotated by observer2.

The degree of agreement for these sets between the detections of the observers and the algorithm were found by comparing timeframes of 10 sec. The training set was used to train the algorithm and set its EDT value. Hypothetically this set would therefore have the highest agreement with the detections of the algorithm. The CFM recordings in the test sets were different from those in the training set. Test set 1 was annotated by the same observer as the training set. Test set 2 was annotated by an independent observer, observer2. This set had the least correlation with the training set, and would therefore hypothetically have the lowest agreement with the detections of the algorithm.

Statistical analysis
The interobserver agreement was determined based on notable ENS activity (NEA) for each of the first 6 h of the 13 recordings. We used an arbitrary threshold for NEA of at least three ENSs or an ENS of at least 3 min in 1 h recording (21). Subsequently we used the G coefficient (22) for the calculation of interobserver agreement. The G coefficient divides the variance of the data into different components related to their source. In our case the different observers, the different patients (or CFM recordings) and for each patient the time within the recording. The G coefficient is then calculated as the extend to which the observers influence the total variance. The value of the G coefficient is therefore between 0 and 1, where 0 will be calculated when all the variance between the data is related to the variance between the observers, and 1 when the variance is not at all related to
the variance between the observers, i.e. there is no variance between the observers.

To evaluate the algorithm the sensitivity and false positives per hour (FPh) for each recording of each of the three sets have been calculated. However, sensitivity and FPh do not take into account the variability of ENS length and amount of ENSs between the CFM recordings. Therefore the recordings have been divided into segments of 1 min, each classified as ENS or non-ENS for each of the observers and the algorithm. This was used to calculate an intraclass correlation coefficient (ICC) (23,24) to compare the different sets. The ICC uses components of variance to calculate the correlation between pairs of observations, i.e. the reliability of these observations. The ICC was calculated separately for each of the sets explained above, compared to the algorithm.

RESULTS

Interobserver agreement
Two neurophysiologists, observer1 and observer2, independently analyzed the CFM recordings. We determined an interobserver agreement based on NEA for each of the first 6 h of the 13 recordings. Results are shown in Table S2.

A full interobserver agreement of NEA was found in nine of the 13 recordings. A high G coefficient of 0.82 has been found as overall interobserver agreement.

All interobserver disagreements were based on a higher amount of ENSs detected by observer1 compared to observer2. These disagreements were mostly caused by patterns in the EEG signals of high frequency and some rhythmic character. Especially in recordings 4 and 8 these patterns caused a large fluctuation in the lower boundaries of the aEEG signals (Figs 3B and C).

Evaluation of the algorithm
The algorithm is evaluated with the three sets described in the methods. The results are summarized in Table S3 A, B and C, respectively.

The values of ICC for comparison with each of these sets are given in Table S4.

Evaluation with the training set (Table S3A)
Since the algorithm has been trained on this set, the ICC value was very high, 0.99. For recordings A, C and E there was almost full agreement between observer1 and the algorithm, with sensitivities of 99–100% and FPh 0–0.42. Recording C showed a neonatal status epilepticus of frequently repeating ENSs that were all but one detected by the algorithm. Recording D, a recording with a relatively low interobserver agreement, had the lowest sensitivity, but still 81%. Recording B showed no ENSs. However this recording contained many muscle artefacts. Most of these artefacts were detected as such by the algorithm, but a mean of 1 FPh remained.

Evaluation with test set 1 (Table S3B)
This set of different CFM recordings had still a high ICC value of 0.95. In recording 1, a recording of 12 h that showed
Automatic seizure detection in neonatal aEEG

no ENS, only one pattern was falsely detected as ENS by the algorithm. Of the remaining seven recordings, four had a sensitivity of 93–100%, including one recording that showed a status epilepticus. The FPh was in four recordings at highest 0.58 and overall at highest 1.89.

Three recordings had a low sensitivity. Recordings 4 and 8 had low interobserver agreements as well. These recordings showed large fluctuations in the lower boundaries of the aEEG signals, as shown in Figures 3A and B. For the algorithm these large fluctuations caused, depending on the durations and amplitudes, either false positives, or a rise in the reference boundary. Rises in the reference boundary may cause following ENSs to remain undetected. Recording 6 showed five ENSs with in the aEEG signal an immediate rise in amplitude to the maximum value as shown in Figure 3C. This so-called amplifier saturation caused these patterns to be detected as artefacts by the algorithm.

**Evaluation with test set 2 (Table S3C)**

This data set, containing different CFM recordings than the training set and annotated by an independent observer, was the least related to the training set. The ICC value of the ENS detections of the algorithm compared to this set was, as expected, the lowest value but was still good, 0.85. The set only included the first 6 h of each of the recordings.

The recordings with a full interobserver agreement of NEA had sensitivities of 90 to 100%. This was comparable to the values found for test set 1 in Table besides for recording 6. In recording 6 the ENSs that were falsely detected by the algorithm as artefacts, were not within the first 6 h of the recording, so were not included in test set 2. Therefore, the sensitivity was in this case 100%. The values for FPh were in all but one recording equal or slightly higher than in Table S3B, but never exceeding 1.33. For the recordings with low interobserver agreements, in recording 4 observer2 did not annotate any ENS and in recording 8 only one, which was not detected by the algorithm. The amount of FPh for these two recordings was as high as 4, caused by the large fluctuations in the lower boundaries of the aEEG signals.

**DISCUSSION**

In the present study, an algorithm for automatic screening of ENSs in neonatal CFM recordings is proposed. The purpose of the algorithm is to assist doctors and nurses in the interpretation of the aEEG signal and as an alarm function on the CFM monitor. A part of the algorithm detected artefacts, in order to distinguish them from ENSs. ENSs were detected based on their characteristic pattern in the aEEG signal, i.e. an increase of the lower boundary. Muscle activity, that may cause a similar rise in the lower boundary of the aEEG signal, was differentiated from ENSs by using the high frequency content of its pattern in the EEG signal. Other artefacts were detected by using the amplitude of the aEEG signal and the magnitude of the impedance.

Two neurophysiologists, observer1 and observer2, analyzed 15 CFM recordings for training and evaluation of the algorithm. For nine of the 13 recordings there was a full interobserver agreement of NEA for each hour. An overall interobserver agreement of G coefficient 0.82 was found. Five CFM recordings annotated by observer1 were used to train the algorithm, the so-called training set. Eight recordings were used for further evaluation of the algorithm, test set 1 and test set 2 annotated by observer1 and an independent second observer, observer2, respectively.

For the recordings with full interobserver agreement of NEA the evaluation of this automatic seizure detection algorithm showed very high sensitivities and low FPh for most CFM recordings. All but one showed a sensitivity of 90% or higher when compared with either of the observers. The recording with a low sensitivity contained ENSs where the amplitude of the aEEG signal immediately rose to the maximum value. The algorithm therefore classified these patterns falsely as artefacts. The amount of FPh was only for two of the 10 recordings higher than 1, but never more than 1.33.

Of the four recordings with a lower interobserver agreement, two had sensitivities below 80% and FPh higher than 1 when compared to the annotations of observer1. These two recordings, both from the test set, showed large fluctuations in the lower boundaries of the aEEG signals. This caused the missed ENSs and false detections of the algorithm. Observer2 only detected 0 and 1 ENS in these recordings. In a second review of these two recordings, the observers advised that clarification about the presence of ENSs should be given by standard EEG measurements.

The ICC values for the different sets were high and decreased for sets that were less related to the training set. This relationship was expected and indicates the reliability of the evaluation method.

Automatic ENS detection methods based on standard EEG recordings have been described earlier, based on minimum ENS durations of 10 sec. Liu et al. (25) described an automatic ENS detection based on the repetitive character of seizure induced EEG signals. He found a sensitivity of 84% and a specificity of 98%. Another important study of Gotman et al. (26) was more comprehensive, and also included detection of spikes that may not be rhythmic. His thorough evaluation resulted in a sensitivity of 69% and 2.3 false positives per hour (27). More recent algorithms were based on time-frequency (28) or wavelet analysis (29), that detect the repetitive structures by their characteristic evolution of frequency in time. These algorithms were not evaluated. An algorithm based on synchronization likelihood, detected statistical interdependencies between EEG channels. A sensitivity of 65.9% and a specificity of 89.8% were found in its evaluation (30). Finally Navakathiyan recently published an interesting algorithm comparing intervals, amplitudes and shapes of consecutive waves to detect ENSs. The evaluation showed a sensitivity of 83–95% with 2.0 false positives per hour (31).

EEG is the golden standard for the detection of ENSs. However, for long-term monitoring and ENS detection outside office hours when EEG may not be available, CFM is a useful tool. Automatic ENS detection in EEG signals is extremely complex due to the variation in morphology and
frequency between and within ENS patterns. This makes algorithms for their automatic detection complex and still in need of validation by medical specialists. CFM processes the EEG signal into a relatively easy to interpret aEEG signal. In these signals ENS patterns have similar morphologies: the rise of the lower boundary of the aEEG signal. Remarkable is the relatively straightforward detection of these patterns, including artefact rejection, resulted in high sensitivities in combination with low false positives per hour for ENSs of at least 60 sec.

Some remarks need to be made here. First, although high sensitivities and low false positives per hour were reached, the detections of the algorithm are still in need of validation by clinicians, i.e. the final decision about the presence of seizure activity still needs to be made by the clinicians. The second remark concerns the limited amount of channels measured in CFM. Due to this type of measurement some focal ENSs can be missed. However, CFM measurements are not intended to be a substitute for standard EEG. The standard EEG measurements for detailed registrations and a more extensive spatial examination of the brain function are complemented by the CFM measurements that facilitate long-term monitoring. Third, we arbitrarily chose to include only ENSs of at least 60 sec in this study, due to the compressed nature of the aEEG signal and the aim to use the algorithm as a screening tool. Future research will aim to decrease this minimum duration of included ENSs. Finally, some false or missed episodes occur in recordings with extremely high fluctuations in the lower boundaries of the aEEG signals and ENSs that cause immediate amplifier saturation.

Clinical relevance of automatic ENS detection is related to the question if seizures cause damage to the brain of the newborn. The newborn brain is more susceptible to seizures, but also more resistant to brain injury. Furthermore, neurological sequelae following neonatal seizures are mainly due to the aetiologies of these seizures (32). However, due to their energy-consuming mechanisms, seizures may effect activity-dependent changes in the developmental processes of the brain (33). Both Scher (5) and Ben-Ari (34) indicate that prolonged or frequent seizures do have a long-term effect on the newborn brain, mainly due to the interference with its development.

This study introduced an algorithm for the screening of neonatal ENSs in aEEG signals with rather good results. The algorithm facilitates the interpretation of aEEG signals in clinical practice and can be used as an alarm function on the CFM monitor. However, some points of improvement could increase the reliability of the algorithm in terms of sensitivity and FPh. First, ENSs with an extremely high amplitude (Fig. 3C) need to be detected as true ENSs. Second, a function should be included to recognize recordings with a large fluctuation in the lower boundary of the aEEG signal, which reduces the reliability of the ENS detection. In this case human validation is necessary and possibly the measurement of a standard EEG. Third, in the training of the algorithm as well as in the evaluation, more recordings should be included. These recordings should be annotated by more than one neurophysiologist, since the algorithm replicates the detections of these neurophysiologists. Besides these points of improvement, also ENSs with a duration less than 60 sec should be included. These are new goals for our future research.

CONCLUSIONS
An automatic detection algorithm for neonatal seizures in aEEG signals was developed and evaluated in this study. In the evaluation rather good results were found for recordings with clear ENSs. The algorithm developed, facilitates the interpretation of aEEG signals in clinical practice and may be used as an alarm function on the CFM monitor.

ACKNOWLEDGEMENTS
We thank Ir Joep Kierkels, Eindhoven University of Technology, as well as Ir Chris Peters and Ad Smets, Máxima Medical Centre Veldhoven for their help in the development of the algorithm; Ted Weiler PhD, Olympic Medical, for his help in the understanding of the CFM algorithm; Professor Dr Martin Offringa, Academic Medical Centre Amsterdam, and Dr Caroline Terwee, VU University Medical Centre Amsterdam, for their help in finding the suitable statistical analysis methods; Dr Djien Liem, Radboud University Nijmegen Medical Centre and Dr Denis Azzopardi, Imperial College London for providing CFM recordings.

References


**Supplementary material**

The following supplementary material is available for this article:

**Table S1** Clinical details of the patients and their CFM recordings that were used in this study. The aEEG background patterns are categorized as continuous normal voltage (CNV), discontinuous normal voltage (DNV) or burst suppression (BS) (19). The ENSs are categorized as either repetitive ENSs (RS) or status epilepticus (SE)

**Table S2** Notable ENS activity (NEA) in the first 6 h of the 13 recordings according to observer1 and observer2, and their disagreement. For NEA we used an arbitrary threshold of at least three seizures or a seizure of at least 3 min/h.

**Table S3** Comparison of the ENS annotations by one of the observers with the ENS detections of the algorithm, given in number of ENSs (nr) and amount of ENS minutes (min). Sensitivity and false positives per hour (FPh) were calculated from the number of ENSs. (A) The training set compared to the algorithm. (B) Test set 1 compared to the algorithm, containing the full recordings 1–8, annotated by observer1. (C) Test set 2 compared to the algorithm, containing only the first 6 h of each recordings, annotated by observer.

**Table S4** The intraclass correlation coefficient (ICC) calculated for the ENS detections of the algorithm compared to the training set and the two test sets. Test set 1 was annotated by the same neurophysiologist as the training set, test set 2 by an independent neurophysiologist.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1651-2227.2007.00225.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.