



Published in final edited form as:

J Clin Neurophysiol. 2009 August ; 26(4): 218–226. doi:10.1097/WNP.0b013e3181b2f29d.

A Multi-stage System for the Automated Detection of Epileptic Seizures in Neonatal EEG

Joyeeta Mitra,

Dept. of Radiology, University of Massachusetts Medical School, Worcester, MA

John R. Glover,

University of Houston, Dept. of Electrical & Computer Engineering, Houston, TX.

Periklis Y. Ktonas,

University of Houston, Dept. of Electrical & Computer Engineering, Houston, TX

Arun Thitai Kumar,

University of Texas M.D.Anderson Cancer Center, Dept. of Experimental Diagnostic Imaging, Houston, TX

Amit Mukherjee,

Phd student at ECSE Department, Rensselaer Polytechnic Institute, Troy, NY

Nicolaos B. Karayiannis,

University of Houston, Dept. of Electrical & Computer Engineering, Houston, TX

James D. Frost Jr.,

Peter Kellaway Section of Neurophysiology, Departments of Neurology and Neuroscience, Baylor College of Medicine, Houston, TX

Richard A. Hrachovy, and

Peter Kellaway Section of Neurophysiology, Departments of Neurology and Neuroscience, Baylor College Medicine, and Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX

Eli M. Mizrahi

Peter Kellaway Section of Neurophysiology, Departments of Neurology and Neuroscience, Baylor College of Medicine, Houston, TX

Abstract

This paper describes the design and test results of a 3-stage automated system for neonatal EEG seizure detection. Stage I of the system is the initial detection stage, and identifies overlapping 5-s segments of suspected seizure activity in each EEG channel. In Stage II, the detected segments from Stage I are spatiotemporally clustered to produce multi-channel candidate seizures. In Stage III, the candidate seizures are processed further using measures of quality and context-based rules to eliminate false candidates. False candidates due to artifacts and commonly occurring EEG background patterns such as bifrontal delta activity are also rejected. Seizures at least 10 s in duration are considered for reporting results.

The testing data consisted of recordings of 28 seizure subjects (34 hrs of data) and 48 non-seizure subjects (87 hrs of data) obtained in the neonatal intensive care unit. The data were not edited to remove artifacts and were identical in every way to data normally processed visually. The system was able to detect seizures of widely varying morphology with an average detection sensitivity of

almost 80% and a subject sensitivity of 96%, in comparison to a team of clinical neurophysiologists who had scored the same recordings. The average false detection rate obtained in non-seizure subjects was 0.74 per hr.

Keywords

EEG; neonatal; epileptic; seizure; detection; automated

1. Introduction

Prolonged EEG-Video monitoring is currently recognized as a useful and highly specific procedure for the detection, classification, and management of seizures occurring in the neonatal period (Mizrahi et al, 2008). It is now routinely available in many centers, and its utilization is expanding in neonatal intensive care units. Because of the increasing demand, and the time-consuming nature of routine human visual analysis, the need for a reliable automated system of analysis has been widely recognized. In recent years there have been a number of attempts to automate the process of neonatal EEG seizure detection, but the problem is very challenging since the characteristics of neonatal seizures, as observed clinically and electrographically, are unique compared with those of older children and adults. In addition, neonatal EEG seizure patterns vary widely, and diagnostically specific interictal patterns are rare (Mizrahi et al., 2004). Consequently, while considerable progress has been reported, current techniques do not yet provide sufficient accuracy to permit routine clinical application (Liu et al., 1992; Gotman et al., 1997; Celka and Colditz, 2002; Thitai Kumar, 2003; Navakatikyan et al., 2006; Green et al., 2007; Aarabi et al., 2007). In the system described here, prior approaches to the recognition of neonatal seizures based on measures of EEG rhythmicity have been improved upon by the use of a new detection scheme based on wavelet analysis coupled with improved neural network-based classification and novel contextual rules. Extensive testing on an independent data set consisting of unedited neonatal recordings demonstrated enhanced sensitivity in comparison to most prior methods, and improved specificity in comparison to all earlier reports.

2. Methods

2.1 System Overview

A block diagram of our 3-stage system for neonatal EEG seizure detection is shown in Fig. 1. The input to the system consists of 12 EEG channels plus EKG, EMG, and respiration channels. Full details on the data used for testing are provided in Section 3.1 below. The system output consists of the detected multichannel seizures.

In the first stage, for detection purposes only, seizure waveforms are considered to belong to two broad morphological categories¹, Type A - pseudo-sinusoidal discharges (Fig. 2 (a)), and Type C - discharges with spike or sharp wave components, including spike and slow wave complexes (Fig. 2 (b)), multi-phasic spikes, and single spikes (Fig. 2 (c)). The Stage I Type A and Type C algorithms have been designed to detect in individual channels seizure activity of these types in 5-s segments of EEG, overlapped by 2.5 s. The objective of the Stage I algorithms is to detect all probable seizure waveforms in all EEG channels. Thus, the algorithms are tuned for maximizing sensitivity and as a consequence exhibit higher false detection rates. These short seizure segments detected in Stage I are then spatiotemporally clustered (in time, and

¹In early publications we discussed Type B “complex morphology” activity, but we have since determined that Type B activity is adequately detected by the Type A and Type C algorithms.

across channels) in Stage II. True seizures typically cause multiple segment detections by one or more Stage I algorithms for the required minimum length of time (at least 7.5 s). Spatiotemporal clustering serves to fill gaps between segment detections, rejects isolated detections and clustered detections with duration less than 7.5 s, and thus enhances both the sensitivity and specificity. It also helps to “build” the multi-channel seizure that can be compared to the electroencephalographer’s scored seizures for performance evaluation. In Stage III, the candidate multichannel seizures from Stage II undergo post-processing to remove artifacts, based on the existence of certain conditions such as bifrontal delta waveforms, or evidence in non-EEG channels of commonly occurring physiological artifacts. Candidate seizures are also evaluated by measures of quality and other context-based rules to make a decision on the authenticity of the seizure and whether the subject in question is a definite seizure subject. Stage III thus serves to lower further the false detection rate.

2.2 Stage I: Seizure Segment Detection

The Stage I block diagram is shown in Fig. 3. Both Type A and Type C detection algorithms have a neural network (NN) classification block following the rule-based classification block. The outputs of Stage I are those EEG segments in each channel that are qualified after the NN-classification block as being either Type A or Type C. The following sections present an overview of the methods that are used by the Type A algorithm and a more detailed explanation of the Type C algorithm, since the latter has not been previously published.

2.2.1 Type A Detection Algorithm—The algorithm to detect Type A seizure discharges has a two-step classifier. In the first step, Type A-like segments are gathered using threshold-based rules similar to those used by Gotman et al. (1997). The features computed for each segment are:

- First and second dominant frequencies: the frequencies corresponding to the two tallest peaks in the spectrum.
- Widths of dominant frequencies: the difference between the frequency corresponding to half of the peak amplitude of the dominant peak in its falling slope and the frequency corresponding to half of this amplitude in the rising slope.
- Percentage of power in first dominant frequency: ratio of the power in the first dominant frequency within its width to the total power in the spectrum.
- Percentage of power in the first two dominant frequencies: ratio of the power in the first two dominant frequencies within their widths to the total power in the spectrum.
- Peak ratio: the ratio of the peak value at the first dominant frequency to that at the second dominant frequency.
- Stability ratio: The segment is divided into four sub-segments and the mean absolute value of the data in each is computed. The stability ratio is computed as the ratio of the maximum sub-segment mean to the minimum sub-segment mean.

The rules then state that the segment is qualified as a Type A seizure candidate if:

1st and 2nd dominant frequencies are each between 0.5 Hz and 30 Hz, AND widths of 1st and 2nd dominant frequencies are both < 0.6 Hz, AND stability ratio < 2, AND (percentage of power in 1st dominant frequency > 30%, OR percentage of power in 1st and 2nd dominant frequencies > 45%, OR peak ratio > 1.5).

The thresholds above are picked to provide very high sensitivity so that very few true seizure segments are missed.

In the second step, the qualified segments are passed through a quantum neural network with 5 nodes in the hidden layer and a single output. The output response, or membership value, represents the degree to which the segment belongs to the Type A seizure class. The cutoff for the membership value is then chosen so that the true-to-false detection ratio is optimum in training subjects. This algorithm is described in greater detail in (Thitai Kumar, 2003; Karayiannis et al., 2006).

2.2.2 Type C Detection Algorithm—Spike-like activity in neonatal electrographic seizures consists of single-phasic or multi-phasic spikes and also spike & slow wave complexes with prominent spike components. This type of activity characterizes the majority of neonatal epileptic seizures, and the Type C detection method described here provides a unified approach that accounts for the great variability in the morphology, frequency and amplitude of such activity. Wavelet transforms are particularly appropriate for spike-detection due to their time and frequency localization property. One approach has been to view spikes as a type of singularity, where the singularities correspond to the wavelet coefficient modulus maxima at fine scales, and a method for singularity detection using wavelets was reported by Mallat and Hwang (1992). A robust method of spike detection using this approach was proposed by Nenadic and Burdick (2005).

In our algorithm, the continuous wavelet transform (CWT) using the Daubechies order four wavelet is used to obtain the locations of seizure discharges containing spikes. This orthogonal family of wavelets has been studied previously for analysis of EEG seizure discharges by Adeli et al. (2003). Fig. 4 illustrates the basic idea and motivation behind our use of wavelets in this algorithm. A single spike-like seizure discharge often has multiple peaks, as shown in Fig. 4 (a) for three simulated spike-like bursts. Thus in the wavelet transform domain, multiple wavelet coefficient maxima are seen within a small neighborhood. Multi-phasic spike discharges and spike & slow-wave complexes, being wider in duration, contain significant power in the wavelet coefficients at the coarser scales. Thus, wavelet coefficients in a single high-frequency band are not sufficient to obtain a good frequency characterization for detecting neonatal spike-like activity. We found that such activity is adequately characterized by mid-range (3–33 Hz) frequencies, which correspond approximately to scales 4 (33 Hz) to 40 (3.3 Hz) for a sampling frequency of 185 Hz. Five wavelet scales (4, 8, 16, 32, and 40) were chosen for characterizing the scale content (frequency signature) of spike-like activity. The absolute values of the wavelet coefficients are smoothed over a span appropriate to the wavelet scale and expected duration of the spike-like seizure discharges, as shown in Fig. 4(b) for simulated data, and autocorrelation analysis of the smoothed wavelet coefficients is used to determine which wavelet scales have sufficient rhythmicity. The height of the first peak to the right of the zero-lag point of the normalized autocorrelation function is used as a measure of rhythmicity (the percentage of power in the rhythmic component), and the average Inter-spike Interval (ISI) is obtained from the position of this peak relative to the zero-lag point. A true Type C segment is expected to be rhythmic in multiple scales with similar ISI. Thus, the algorithm proceeds to find the center of seizure discharges and subsequent features for classification only if the segment has more than one wavelet scale with at least 40% power in the rhythmic component with same ISI. The center of the spike-like seizure discharges is then estimated as the mean of the locations of the significant wavelet maxima in the chosen scales.

To characterize spike-like seizure activity, three distinct groups of features are computed for each 5-s segment of EEG data: rhythmicity-based, wavelet scale-based and amplitude-based. The rhythmicity-based features are first used to “loosely” pre-classify a segment as possible seizure or as non-seizure (immediately rejected) in a threshold-based rule. The thresholds chosen are low enough to gather a broad range of rhythmic segments, with the goal of including all true seizure segments at the expense of including some non-seizure segments as well. If a segment survives this pre-classification, it is then processed by a Levenberg-Marquardt

feedforward neural network (LM-NN) with 4 nodes in the hidden layer, which has been trained using all three groups of features to more accurately separate the seizure segments from the non-seizure segments. The segments classified as non-seizure are rejected, and those classified as seizure provide the final output of the algorithm. The features used are described here:

Rhythmicity-based features: *PerRhy* (Percentage of power in the rhythmic component) - The data segment is represented as a triangular pulse sequence with pulses centered at the (possibly multi-phasic) spike discharge locations and height equal to the amplitude of the spike discharge. The normalized autocorrelation function of this sequence is computed, and *PerRhy* is given by the height of the first peak to the right of the zero-lag point.

PerDur (Percentage of rhythmic ISI duration) is the length of the rhythmic portion (matching ISIs) of the segment expressed as a percentage of total duration of the segment. In the special case of arrhythmic spike discharges which may cause the ISIs to be non-rhythmic, *PerDur* is computed as the sum total of the ISIs expressed as a percentage of total duration of the segment.

Wavelet Scale-based features: *ScaleContVec* (Scale-content vector) characterizes the scale/frequency profile of the spike-like discharges. It is a 5-element vector, where each element is the sum of the absolute values of wavelet coefficients at the spike discharge locations in each scale normalized by the number of estimated locations.

Amplitude-based features: *BlockRange* is the interquartile range of amplitude of the block (1-min. duration, 12 EEG channels) of data being processed that contains the current segment. *BlockRange* is used to normalize the following two amplitude-based features.

AvPeakAmp is the ratio of the average absolute amplitude of the data-segment in a small window around the locations of spike-like discharges, to *BlockRange*.

SigAmpRange is the ratio of the interquartile range of amplitude of the segment, to *BlockRange*.

2.3 Stage II: Spatio-temporal Clustering

In Stage II the Type A and Type C candidate seizure segments from all channels from Stage I are pre-selected for spatio-temporal clustering on the basis of their NN-response values. This provision prevents clusters from being formed out of segment detections with low NN-response values that are unlikely to be seizure candidates. The clustering algorithm used is an adaptation of the one described by Klatchko et al. (1998) and is illustrated in Fig. 5. The pre-selected candidates are first clustered temporally, based on a *Time Continuity test*, i.e., whether they are "close" in time (gap of less than 5 s). The gap allowed between the segments is equivalent to one missed non-overlapping segment. Temporal clusters longer than 7.5 s are immediately qualified. This duration is small enough to qualify groups of two candidate segments with an overlap of 2.5 s between them. The temporally clustered candidate seizure segments in each channel are then examined using a *Spatial Continuity test*. This test determines whether there exist overlapping or contiguous temporal clusters in channels that satisfy the criterion for being "affined" or "connected," meaning that the channels are in the same region of the head. Such temporal clusters are then grouped together. Resulting candidate seizures are multi-channel or single-channel clusters that can either be complete seizures themselves, or portions of a larger seizure. This clustering procedure serves to enhance both sensitivity - by filling gaps in segment detections - and specificity - by rejecting isolated segment detections, and it additionally creates the candidate multichannel seizures for further scrutiny in Stage III.

2.4 Stage III: Post-processing with Rules

Stage III involves further processing of the candidate seizures for rejection of artifacts and for quality-based measures to reduce the false detection rate. The rules for artifact rejection and quality are applied in two steps: Rules for *artifact rejection*, followed by rules for *overall seizure quality*.

2.4.1 Rules for Commonly Occurring Artifacts—Neonatal scalp EEG contains artifacts due to a variety of sources. Here the term “artifact” is used to refer to all non-seizure electrographic patterns that interfere with the seizure detection process and cause false detections. Some artifacts are normal or abnormal EEG background patterns; some are physiological artifacts from non-EEG sources such as chewing, sucking, and movement. Of the EEG background patterns found in our data, only bifrontal delta activity caused enough false detections to warrant specific measures to remove them. In Stage III, the first post-processing step is, therefore, to remove candidate seizures that correspond to artifacts. The approach used for all of the following artifact rejection routines is to reject an entire candidate seizure if a significant percentage of its constituent segment detections are found to be artifacts.

A. Removal of Bifrontal Delta (BFD): Characteristics of BFD activity: BFD activity is a high-voltage rhythmic EEG activity with the following characteristics:

- Dominant frequency is in the range (0.5 – 4 Hz).
- Occurs in frontal regions (Fp1-T3, Fp2-T4, Fp1-C3, Fp2-C4 and Fz-Cz).
- Occurs bilaterally.
- Intermittent in nature with continuous periods less than 20 s in duration.
- May be mixed with frontal sharp transients.

BFD rejection algorithm: The algorithm for rejecting false candidate seizures due to BFD activity is as follows:

1. The candidate seizures are pre-screened to determine if they are likely to be caused by BFD patterns. Such patterns mostly cause short false candidates because of their intermittent nature. The false candidate seizures due to BFD activity also include one or more frontal channels. To avoid the rejection of bona fide generalized seizure activity involving frontal channels as BFD artifact, candidate seizures spanning only a few channels (less than 6) are tested for this artifact. The above conditions form the pre-screening criteria for rejection of BFD activity.
2. The Type A and Type C segment detections constituting the selected candidate seizures and occurring in frontal channels are then examined for BFD-like characteristics. This is done using two features, the percentage of power in the delta activity range (0.5 – 4 Hz) and the dominant frequency of the segment. If the segment detections examined have more than 75% power in the delta range and the dominant frequency is in the range 0.5 to 4 Hz, then they are suspected to be due to BFD activity. In the case of a Type C detection, the repetition frequency is also checked to see if it is within the delta range. This is necessary because BFD activity is sometimes observed to occur mixed with frontal sharp transients. The detected segments are also labeled as BFD activity if the data segments from any pair of homologous frontal channels in the same time window as that of the detected segments are found to have BFD characteristics.

B. Removal of Artifacts from Non-EEG Sources: Characteristics of the artifacts: The non-EEG channels used for our context-based rules are: EMG, EKG, and Respiration. The artifacts

caused by physiological processes such as chewing, sucking, etc., or by patient movement such as patting are accompanied by synchronized rhythmic activity in the EMG channel.

The non-EEG artifacts are removed by correlating the EEG data with each of the non-EEG channels, or by matching the fundamental period of repetition. The algorithm is as follows:

1. The EEG data corresponding to each of the constituent segment detections (Type A and Type C) in the candidate seizure and the non-EEG channel (EMG, EKG and Respiration) data in the same time window are retrieved. The data are filtered to remove the baseline drift (high-pass, cutoff 0.2 Hz), and de-noised using the *Universal Threshold* procedure (Donoho, 1995). The normalized cross-correlation of the signals is then obtained.
2. If the activity in the EEG and non-EEG channel data is synchronized, with matching fundamental periods, the cross-correlation function computed above is expected to have a high value close to zero-lag. Therefore, the maximum value of the cross-correlation function in a narrow time-span (~ 0.3 sec) around the zero-lag is obtained as a discriminator for synchrony and is called DegC (degree of correlation). If DegC is greater than 0.4 (the maximum possible value of DegC is 1 for complete synchrony), then the segment detection corresponding to the EEG channel is labeled as artifact.
3. If DegC is below 0.4, a second approach to determine any synchrony is used. This involves comparing the fundamental period of the non-EEG channel with that of the detected segment. As the segments are detected in the first place due to rhythmicity of the underlying data, it is expected that in the case of an artifactual event, similar rhythmicity will be seen in the non-EEG channel. Thus, autocorrelation-based rhythmicity analysis is used to determine the periodicity of the non-EEG activity. The method used is the same as that used in the Type C algorithm for rhythmicity analysis of wavelet coefficients. If the fundamental periods of the two signals match (within 0.1 s), then the event is considered artifactual.

2.4.2 Rules for Seizure Quality—The following rulesets were devised to reduce false detections due to reasons other than the presence of artifacts. They consist of rules that are based on one or more characteristics of neonatal seizures with the purpose of rejecting seizure candidates that are not of sufficient quality to be retained as true seizures. Fig. 6 shows the scheme of application of the first three rulesets in Stage III.

Ruleset 1: For identifying seizures with overall good quality: *Good quality seizures* are defined as those satisfying this ruleset which is applied to the candidate seizure as a whole. It is based on three distinct criteria of seizure quality, namely, Stage I *segment quality*, *frequency consistency*, and *duration*. The criterion of Stage I segment quality is given the most weight and thus stricter thresholds were used for the corresponding rules. Single-channel candidate seizures were qualified with a more stringent set of percentage-based thresholds than multi-channel candidate seizures, as they were more likely to be false positives clustered together with a smaller number of segment detections in them.

A candidate seizure's Stage I segment quality is assessed by the percentage duration of *good quality segment detections* in the candidate seizure, which must be significant (75 % or more for multi-channel, 80 % or more for single-channel candidate seizures). *Good quality segment detections* were defined as those having NN response exceeding experimentally determined thresholds, ≥ 0.85 for Type A detections and ≥ 0.4 for Type C detections.

Frequency consistency of a candidate seizure is determined by the percentage of segment detections *within a channel* with matching *repetition frequencies* (50 % or more for multi-

channel, 80 % or more for single-channel candidate seizures). In addition, for multi-channel seizures, the *repetition frequencies between at least two of the channels* must match. The repetition frequencies of two Type A segment detections are matched by comparing their 1st and 2nd dominant frequencies, and for two Type C segment detections by comparing the ISIs (inverse of repetition frequency). A match is said to occur when the values are sufficiently close with a tolerance based upon the computational error involved and the proximity of the segments, i.e., whether they are overlapping or adjacent. The repetition frequencies of Type A and Type C segment detections are compared only if the segments overlap in time. When a pair of segments is successfully compared for a match, the repetition frequency of the segment occurring first is called the *matching frequency* of the pair; or, if the two segments begin at the same time, for example, a segment classified both as Type A and as Type C, the *matching frequency* is the average of the two repetition frequencies. We can now compute the percentage of segment detections that are consistent in terms of repetition frequency, within a single channel and between two channels of a multi-channel seizure candidate:

Within a channel: The percentage of segments with matching repetition frequencies is computed as the ratio of the number of frequency-matched pairs in the channel to the total number of segment pairs that were compared for a match.

Between two channels: For each pair of channels in a multi-channel seizure candidate, say channel 1 and channel 2, two features are computed, PctMatch12 and PctMatch21, which are computed as follows: PctMatch12 is defined as the percentage of the *matching frequencies* of channel 1 that are within 0.3 Hz of the neighboring (within 5 s) matching frequencies within channel 2. PctMatch21 is defined similarly. The percentage used for determining frequency consistency is the minimum of these two.

The minimum duration requirement for a candidate seizure to qualify by Ruleset 1 was set to be 10 s.

Ruleset 2: For identifying more seizures in subjects with some overall good quality seizures: Short or inferior quality seizures in subjects with at least 15 s of overall good quality seizures (qualified by Ruleset 1) are detected by this ruleset. It is to be noted that this ruleset does not change the number of subjects classified as having seizure, as it is applied only to subjects with some already detected overall good quality seizures. This ruleset is meant to improve the sensitivity within a particular subject by matching characteristics of an unqualified candidate seizure with overall good quality seizures using (i) repetition frequencies, and (ii) the channels of occurrence. The Mahalanobis distance measure is used to obtain a measure of *similarity* in terms of the repetition frequency. Two samples are formed, one consisting of the *matching frequencies* from all channels of a candidate seizure qualified by Ruleset 1, and the other consisting of the same from candidate seizures not yet qualified. If the samples are determined to be similar, and if the candidate seizure compared has 50% channels in common and has duration greater than 10 s, then it is qualified as seizure by Ruleset 2.

Ruleset 3: A special case rule for identifying focal seizures with weak rhythmicity: This ruleset is meant for detecting seizures in those subjects where seizures are focal, mostly occur in one or two channels, and consist of short rhythmic fragments (10–15 s) interrupted by bursts or other non-seizure activity. This ruleset is designed to qualify candidate seizures in subjects that have no seizures meeting overall good quality requirements defined by Ruleset 1, but for whom sufficient evidence of the presence of a true seizure can be obtained based upon the predominant occurrence of good quality segment detections in only a few channels which are intermittent and thus fail to qualify the duration requirement.

Each channel is graded on a scale of 0–1, where the grade is 0 for the occurrence of no candidates, and 1 for maximum number and duration of candidate seizures with Average Quality (AQ) greater than 0.5. AQ is computed as:

AQ of a candidate seizure = (Number of good quality Type A and Type C segments) / (Total number of segment detections in the candidate seizure), where good quality segments are obtained as in Ruleset 1. If the seizure activity is localized to only a few channels, then the number of channels with a high grade as computed above is found to be small indicating that the Stage I algorithms detected seizure segments predominantly in those channels. To further confirm that the activity in these channels is seizure, the *matching frequencies* from the candidate seizures used for grade computation are compared for similar rhythmicity characteristics using the Mahalanobis distance measure as in Ruleset 2. If a significant percentage of the candidate seizures of high AQ are similar in terms of *matching frequencies*, they are qualified by Ruleset 3.

Ruleset 4: For growing candidate seizures: Ruleset 4, not shown in Fig. 6, is used to grow any of the seizures qualified by Rulesets 1, 2, and 3. It does so by merging with them previously rejected segments under certain criteria:

- a. It looks 20 s on either side of every channel of a qualified candidate seizure to find if there are more Type A or Type C segments that were rejected earlier for low quality.
- b. These segments are compared for matching repetition frequency with segments of the same type currently within the candidate seizure.
- c. If one of these segments matches 50% of the segments of same type already within the seizure candidate, it is added to the seizure candidate.

Note that Ruleset 4 can only lengthen a detected seizure, and will not introduce any new false seizures.

3. Results

3.1 Data Description

The EEG data used in this study were obtained from a large database containing a wide variety of neonatal seizures. This database was developed by the Clinical Research Centers for Neonatal Seizures (CRCNS) under a previous NIH grant, and contains video/EEG/polygraphic recordings of several hundred individual clinical seizures, which have been fully characterized and classified in terms of their electrographic and behavioral features (Mizrahi et al., 2001). Each patient was recorded continuously in the neonatal intensive care unit (NICU) of Texas Children's Hospital for approximately 2 hours under the supervision of an experienced EEG technologist. Each recording was later interpreted by a team of 4 clinical neurophysiologists (electroencephalographers) who provided a consensus characterization of the first 10 seizures that occurred during the 2-hour session (if more than 10 seizures occurred during the recording, scoring was terminated after the 10th seizure; if fewer than 10 seizures occurred, the entire recording was analyzed). Information used to evaluate the automated system performance included the starting and ending times of each seizure documented, as well as the region of the brain in which the seizure originated.

A Nihon Kohden electroencephalograph with a low-frequency cut-off of about 0.5 Hz was used for data acquisition. The multi-channel EEG data were passed through anti-aliasing filters with cut-off frequencies at 70 Hz and digitized at 185 Hz with a system developed at Baylor College of Medicine. The channels recorded were: Fp1-C3, Fp2-C4, C3-O1, C4-O2, Fp1-T3, Fp2-T4, T3-C3, C4-T4, C3-Cz, Cz-C4, Fz-Cz, Cz-Pz, EOG, EMG, ECG, chest respiration,

pulse, airflow, body movement, oxygen saturation, CO₂. Of these, 12 EEG channels and 3 non-EEG channels (EMG, EKG, and respiration) were used as input to the system.

Evaluation of an automated seizure detection system for clinical application must be based on a wide variety of seizure morphologies and subjects. Test data must include artifacts arising from various sources, e.g., normal or abnormal seizure-like EEG patterns not corresponding to seizures, physiological sources (EKG, EMG), movement of electrodes with respect to the subject due to sucking, chewing, and other activities, etc. A useful seizure screening device must be evaluated for high *subject sensitivity*, i.e., a small number of subjects missed as seizure subjects, in contradistinction to other methods of grading where sensitivity is reported based on the total number of seizures in the dataset. This is because, for subjects having one or very few seizures in the entire record, the cost of missing a seizure is much higher than missing a seizure of equal duration in a subject that has multiple recurring seizures. Therefore, we have graded our system based on the false detection rate and both the subject sensitivity and the overall sensitivity of seizure detection.

The training set consisted of 21 seizure subjects and 20 non-seizure subjects. Most commonly-observed artifacts in the training set were bi-frontal delta activity and movement. Other artifacts observed were EMG, EKG, respiration, chewing, sucking, suppression-burst and tremor. Some training data were termed as “abnormal EEG”, and these records were marked by the presence of intermittent rhythmic background patterns.

After training, the testing data set consisted of 28 seizure subjects and 48 non-seizure subjects. The duration of each recording used for testing the system corresponded to the entire portion of the recording that was scored by the team of electroencephalographers. Seizures ranged in duration from 10 s to 20 min. The data sets used for testing the automated system were not edited to remove artifacts and were identical in every way to data normally processed visually in the NICU.

3.2 Performance Metrics

The performance metrics, namely, sensitivity, subject sensitivity, and false detection rate were computed as follows:

1. True Seizures (TS) = number of electroencephalographer-scored seizures overlapping one or more system-detected seizures. The seizures detected were classified as true seizures if they overlapped in time with the seizures marked by the electroencephalographers and also overlapped with at least one channel in the spatial region of the brain where the seizure started, as determined by the electroencephalographers.
2. False Seizures (FS) = number of system-detected seizures not overlapping any electroencephalographer-scored seizures.
3. Total Scored Seizures (TSS) = total number of seizures scored by the electroencephalographers in the subject.
4. Total Data Duration (TDD) = the duration in hours of continuous EEG data over which the system performance was evaluated.
5. Sensitivity (%) = $100 \times TS / TSS$
6. False detection rate (per hour) = FS / TDD
7. Subject sensitivity (%) = percentage of seizure subjects in which sensitivity was non-zero.

3.3 System Performance

3.3.1 Testing data—Table 1 shows the performance of the system in testing seizure subjects and Table 2 in testing non-seizure subjects. Designed to be a screening device, the system is tuned for minimum false detection rate in non-seizure subjects at the expense of a few missed seizures in seizure subjects. In testing subjects as seen from Table 1, 16 seizure subjects were detected with sensitivity greater than 90% and 7 additional subjects were detected with sensitivity greater than 60%. Subject sensitivity, i.e., the percentage of seizure subjects in which at least one seizure was detected, was 96%. Only 1 subject out of 28 seizure subjects was missed completely. The false detection rate in some seizure subjects was higher due to the presence of rhythmic seizure-like patterns in between scored seizures. In the case of non-seizure subjects (Table 2), no seizures were detected in 27 out of 48 subjects. The average false detection rate obtained for non-seizure subjects was 0.74 per hr

3.3.2 Analysis of misses and false detections—The missed seizures are summarized in Table 3 in the following categories:

- (a) Slow pseudo-sinusoidal: Seizures that were unusually slow (fundamental period ≥ 2.5 s), generally with very high amplitude, were missed because they were difficult to check for periodicity within our 5-s segment length in Stage I.
- (b–C) Low amplitude: Very low-amplitude, short seizures in depressed EEG were missed because the Stage I algorithms were not well-tuned for them. For example, the only seizure subject missed completely had a single short seizure that was of very low amplitude and high frequency.
- (b–d) Short duration: The most prevalent characteristic by far was that of being short, with duration less than 20 s. These seizures simply had insufficient time to establish their rhythmicity to the satisfaction of the clustering algorithm and Ruleset 1.

The false seizure detections are summarized in Table 4 in the following categories:

- a. Short duration: With respect to seizure duration, short false seizures were the principal problem, as they were for missed detections above, particularly those shorter than 20 s.
- b. Ruleset 1: Correspondingly, it is Ruleset 1 where most of the mistakes are being made. This would be expected, as it is difficult to judge seizure quality adequately when the seizure is so short and cannot exhibit a long period of rhythmicity.
- c. Single-channel; rhythmic interictal background: Also not unexpected, the largest category of false detection was the single channel false seizure, where Ruleset 1 has to make a decision without support from multiple channels. In particular, there is no opportunity to find frequency synchronization between channels. Also, it is interesting that for seizure subjects, 28% of the false detections were rhythmic background, whereas the figure was only 10% for non-seizure subjects. This implies that the interictal background has more latent rhythmicity than the non-seizure background, perhaps an interesting area for further study. Artifacts caused only 10% of the false detections, indicating that our artifact rejection measures are effective.

In summary, misses and false detections were predominantly “brief” both temporally (short) and spatially (single-channel), with borderline rhythmicity, causing Ruleset 1 to make an incorrect classification.

4. Discussion

The work described here follows over a decade of work by multiple researchers toward the development of a reliable neonatal seizure detection system. An early method described by Liu et al. (1992) assumed that the essential characteristic in newborn seizure EEG was periodicity. The autocorrelation of a short epoch of EEG data was scored to quantify periodicity and used in a rule-based algorithm to perform classification. This approach was shown to be promising, although the data set used for testing was very limited. Gotman et al. (1997) performed the periodicity analysis in the frequency domain, and also included steps to detect arrhythmic spikes and slow (<0.5 Hz) complex discharges. Their results indicated that only 2/3 of the seizures were being detected, while there were 2.3 false detections per hour, resulting in about half of the reported detections being false alarms.

In more recent work by Navakatikyan et al. (2006), seizure detection was performed by parallel fragmentation of EEG into waves, followed by classification using features based on parameters such as amplitude, wavelength, correlations between the shapes of the waves, and features derived from wave sequences. The algorithm detected seizures based on increased regularity quantified by the linear correlation between adjacent wave shapes and various other threshold-based rules. This system was reported to have a sensitivity of 89.7% and a false detection rate of 2 per hr. It may be noted that pooling the detected seizures from all subjects for computation of sensitivity has the disadvantage of overshadowing the performance in subjects with a smaller number of scored seizures.

Greene et al. (2007) recently reported a new algorithm that combines EEG and EKG analysis in the detection of seizure events, with results for a patient-dependent algorithm and a patient-independent algorithm. The patient-independent algorithm, which would be needed in the neo ICU, had a seizure sensitivity of 81.44% with a false detection rate of 3.15 per hr. The authors felt that this level of performance was not reliable enough for routine use in the neo ICU.

Aarabi et al. (2007) reported on a multistage system that is architecturally similar to ours, employing artifact rejection, EEG segmentation, neural network classification, and rules. Their system was tested on 10 newborns between 39 and 42 weeks, with a reported seizure detection sensitivity of 74% and a false detection rate of 1.55 per hr.

In comparison, the system described here was able to detect one or more seizures in all but one of the 28 seizure subjects (96% subject sensitivity). Most of the missed seizures (70%) were short in duration (10–20 s), with additional misses due to low frequency data. The overall seizure sensitivity of this system in seizure subjects (79.8%) is higher than that reported by Gotman et al. (67%) and Aarabi et al. (74%), and lower than that reported by Navakatikyan et al. (89.7%). With respect to the false detection rate, the present system performs better than any of those previous systems, with less than half the number of false detections (0.78 per hr overall average, 0.74 per hr in 48 non-seizure subjects). Again, it is short seizure candidates that cause the most problems. Plans for improving the system will involve special treatment for low-frequency and low-amplitude segments in Stage I, and for short seizure candidates in Stage III.

Acknowledgment

We acknowledge support of this research from the National Institutes of Health, grant no. R01-NS40577.

Funding from: National Institutes of Health, grant # R01-NS40577

References

- Aarabi A, Grebe R, Wallois F. A multistage knowledge-based system for EEG seizure detection in newborn infants. *Clinical Neurophysiol* 2007;118:2781–2797.
- Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods* 2003;123:69–87. [PubMed: 12581851]
- Celka P, Colditz P. A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison. *IEEE Trans Biomed Eng* 2002;49(5):455–462. [PubMed: 12002177]
- Donoho DL. De-noising by soft-thresholding. *IEEE Trans on Information Theory* 1995;41:613–627.
- Greene BR, Boylan GB, Reilly RB, de Chazal P, Connolly S. Combination of EEG and ECG for improved automatic neonatal seizure detection. *Clin Neurophysiol* 2007;118(6):1348–1359. [PubMed: 17398146]
- Gotman J, Flanagan D, Zhang J, Rosenblatt B. Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalog. Clin. Neurophysiol* 1997;103:356–362.
- Karayiannis NB, Mukherjee A, Glover JR, Ktonas PY, Frost JD, Hrachovy RA, Mizrahi EM. Detection of pseudo-sinusoidal epileptic seizure segments in the neonatal EEG by cascading a rule-based algorithm with a neural network. *IEEE Trans on Biomed Engr* 2006;53(4):633–641.
- Klatchko A, Raviv G, Webber WRS, Lesser RP. Enhancing the detection of seizures with a clustering algorithm. *Electroencephalography and Clinical Neurophysiology* 1998;106:52–63. [PubMed: 9680165]
- Liu A, Hahn JS, Heldt GP, Coen RW. Detection of neonatal seizures through computerized EEG analysis. *Electroencephalog. Clin. Neurophysiol* 1992;82:30–37.
- Mallat S, Hwang WL. Singularity detection and processing with wavelets. *IEEE Trans on Information Theory* 1992;38:617–643.
- Mizrahi EM, Clancy RR, Dunn JK, Hirtz D, Chapieski L, McGuan S, Cuccaro P, Hrachovy RA, Wise MS, Kellaway P. Neurologic impairment, developmental delay, and postneonatal seizures 2 years after EEG-video documented seizures in near-term and term neonates: Report of the clinical research centers for neonatal seizures. *Epilepsia* 2001;42:102–103.
- Mizrahi, EM.; Plouin, P.; Clancy, RR. *Epilepsy: A comprehensive textbook*. In: Engel, J., Jr; Pedley, TA., editors. *Neonatal Seizures*. Vol. 2nd edition. Philadelphia: Lippincott Williams and Wilkins; 2008. p. 639-658.
- Mizrahi, EM.; Hrachovy, RA.; Kellaway, P. *Atlas of Neonatal Electroencephalography*. Vol. 3rd ed. Lippincott Williams & Wilkins; 2004.
- Navakatikyan MA, Colditz PB, Burke CJ, Inder TE, Richmond J, Williams CE. Seizure detection algorithm for neonates based on wave-sequence analysis. *Clin Neurophysiol* 2006;117(6):1190–1203. [PubMed: 16621690]
- Nenadic Z, Burdick JW. Spike detection using the continuous wavelet transform. *IEEE Trans on Biomed Engr* 2005;52:74–87.
- Thitai Kumar, A. M.S. Thesis. University of Houston, TX: Department of Electrical and Computer Engineering; 2003. Automated detection of narrowband rhythmic segments in neonatal seizure EEG.

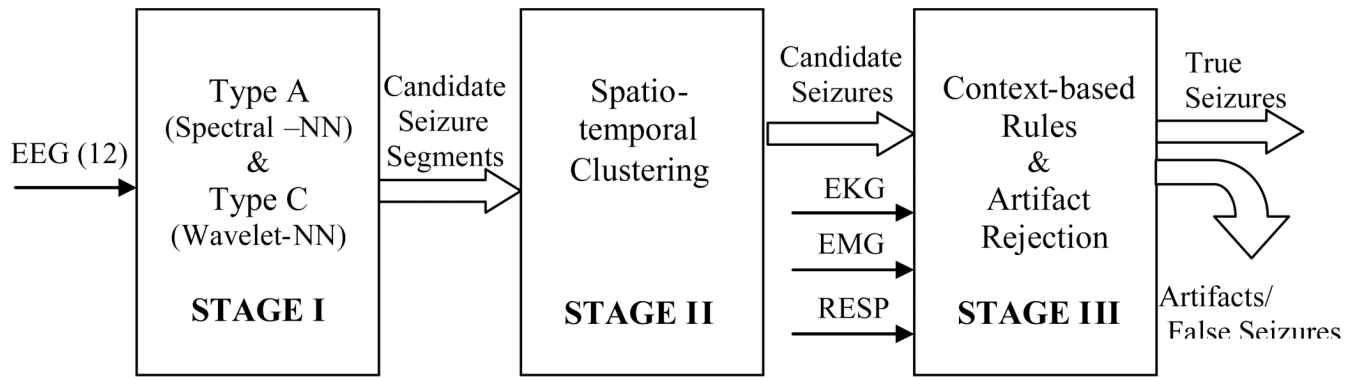


Figure 1.
3-stage system architecture

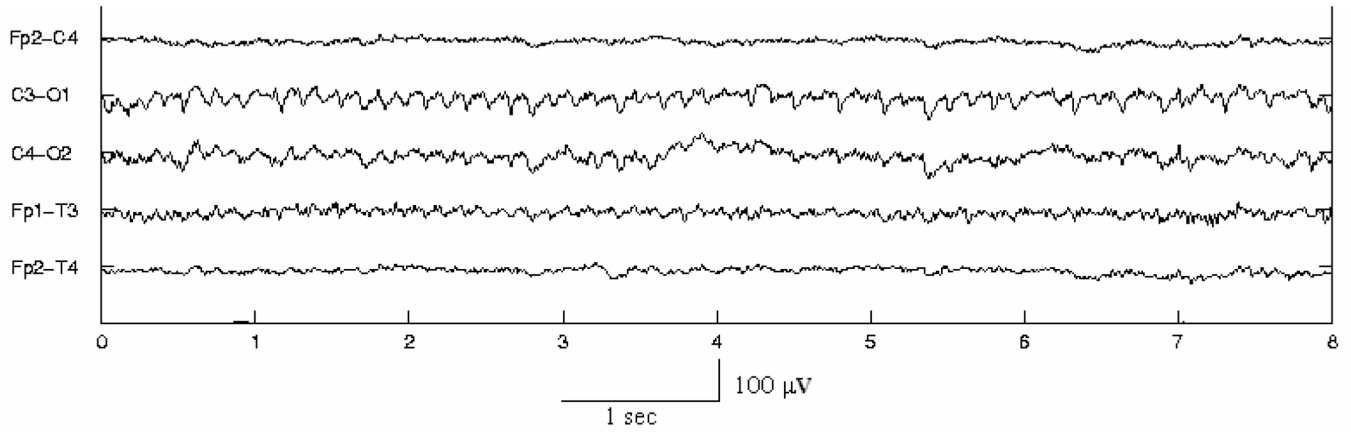


Figure 2a.

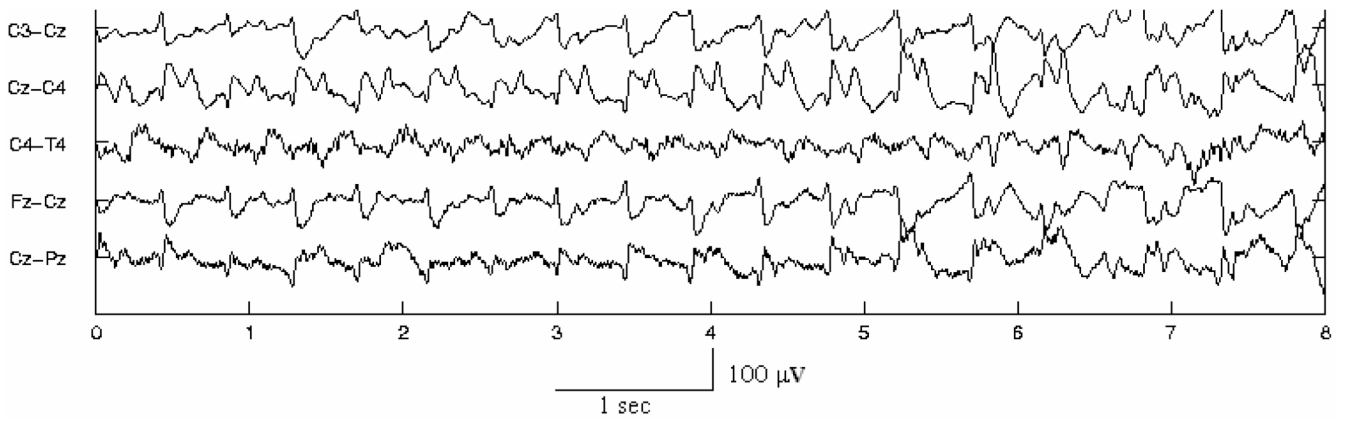


Figure 2b.

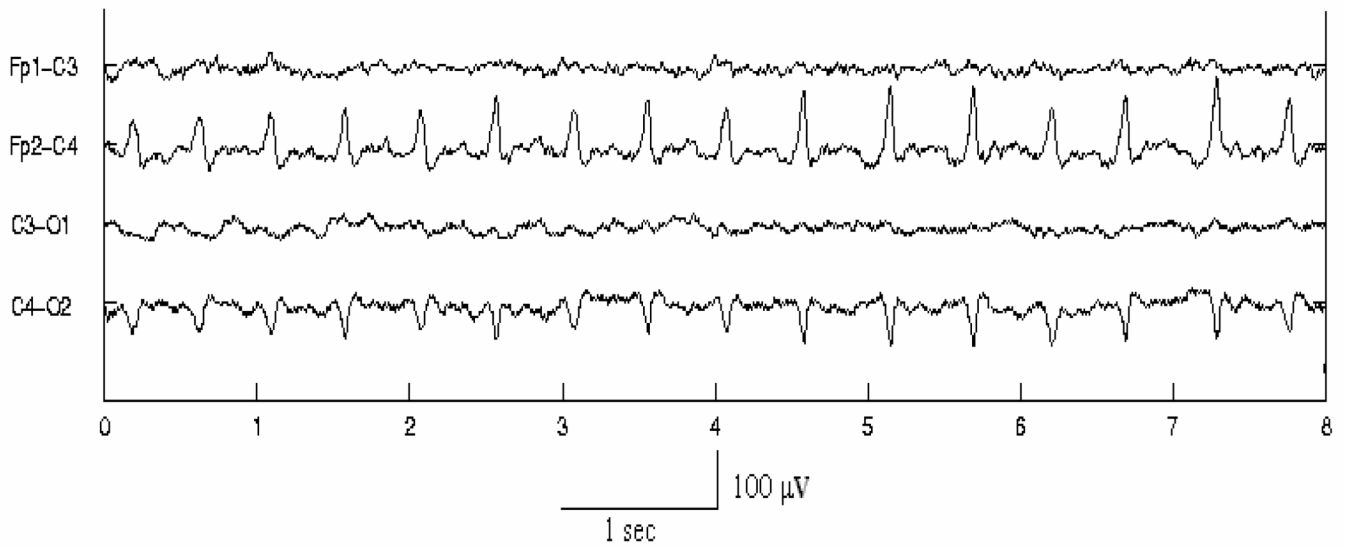


Figure 2c.

Figure 2.

Figure 2a. Example of Type A seizure activity in channels C3-O1, Fp1-T3 & to some extent in C4-O2.

Figure 2b. Example of complex morphology Type C seizure activity with spike components in channels with Cz as the common electrode.

Figure 2c. Example of spike-like Type C seizure activity in channels Fp2-C4 and C4-O2.

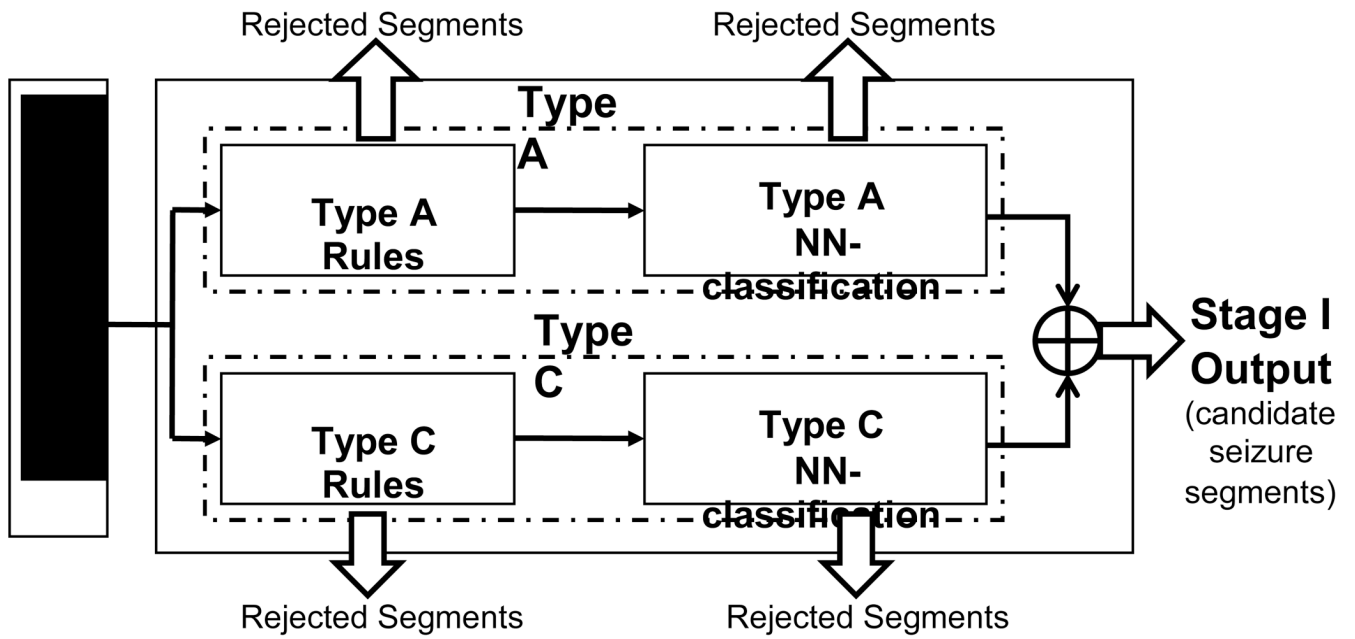


Figure 3.
Stage I block diagram.

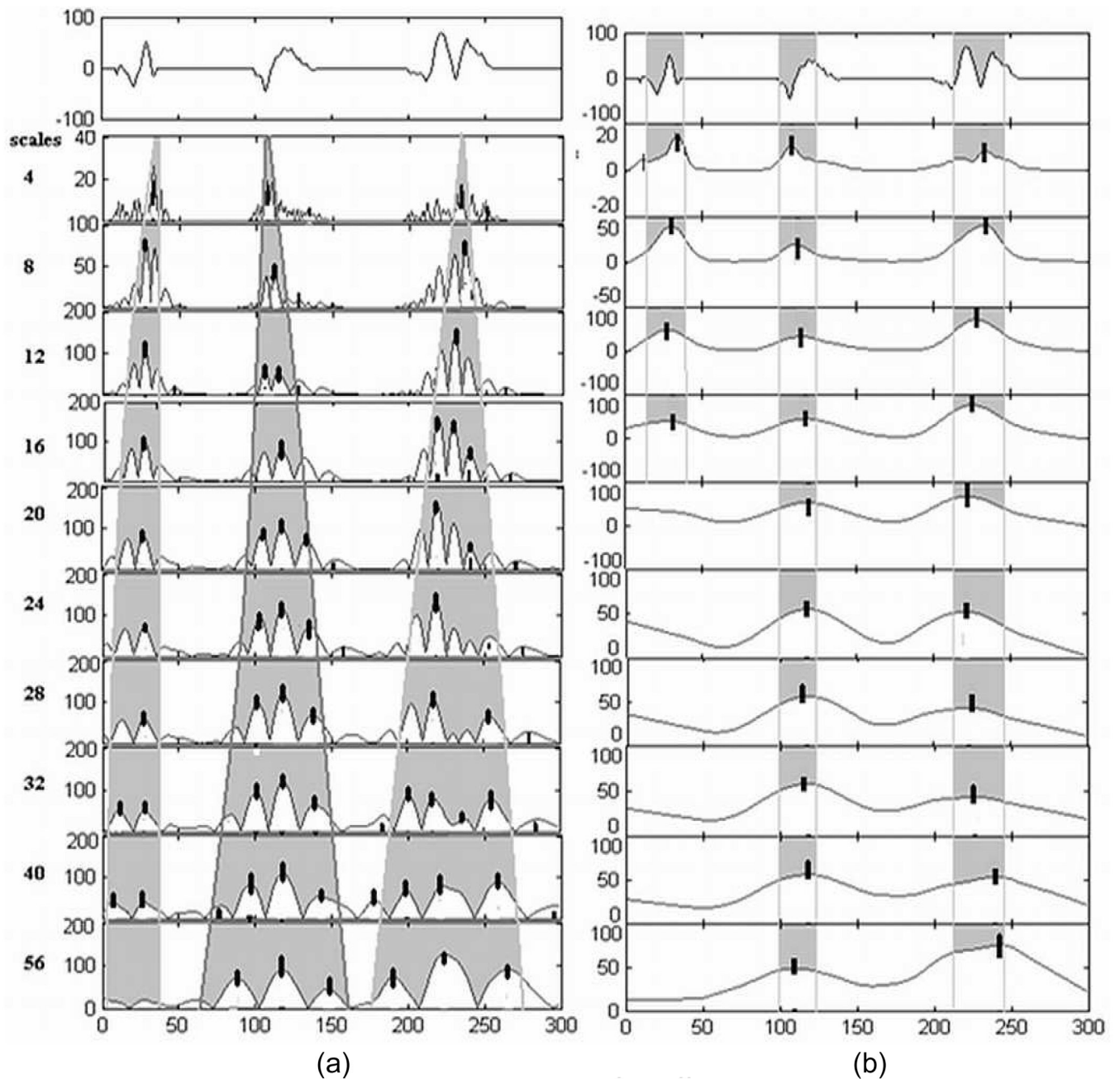


Figure 4.

Evolution of modulus maxima of wavelet coefficients across scales for different types of simulated spike-like activity shown at the top of each figure, (a) before smoothing, and (b) after smoothing. The black bars indicate locations of significant wavelet maxima determined at each scale. The shading indicates the spread of wavelet coefficients corresponding to each discharge in different scales. The duration of the segment is $1.6 \text{ s} = 300$ samples (sampling rate = 185 Hz).

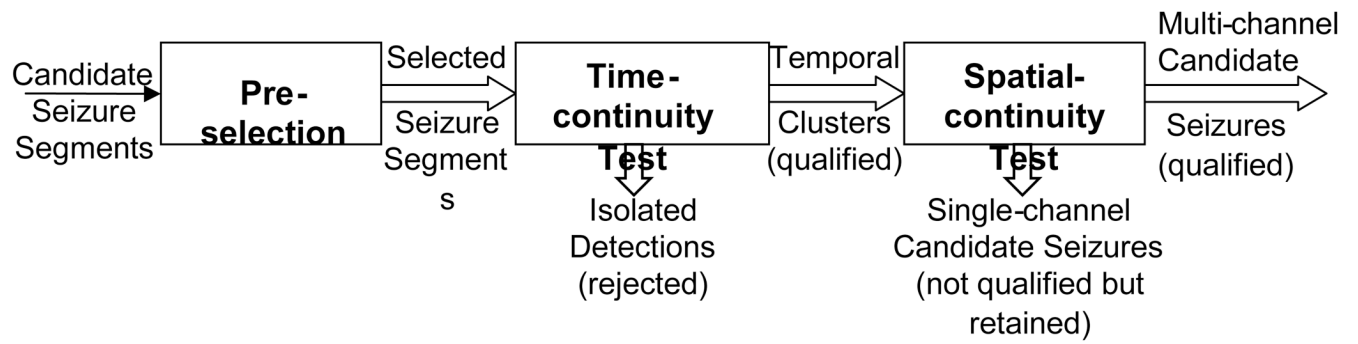


Figure 5.
Stage II block diagram.

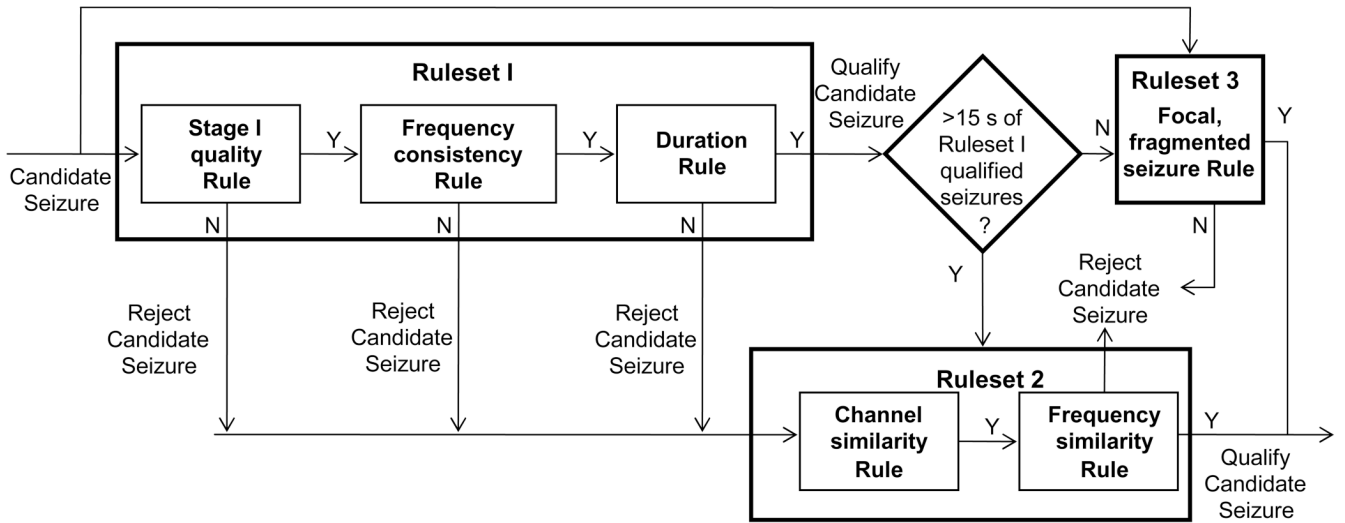


Figure 6.

Stage III block diagram showing the operation of Rulesets 1, 2, and 3. Ruleset I tests three different criteria to assess overall seizure quality. Ruleset 2 uses similarity measures to improve sensitivity within a subject with some good quality seizures. Ruleset 3 has special case rules for focal and fragmented seizures

Table 1

Performance of the system for seizure subjects.

Subjects	True detections	Missed Seizures	False detections	% Sensitivity	Data duration (hr)	False detection rate (/ hr)
1	10	0	0	100	2.01	0
2	9	1	2	90	1.5	1.33
3	9	1	1	90	0.78	1.28
4	8	1	5	89	0.9	5.56
5	2	0	1	100	0.75	1.33
6	6	0	0	100	1.67	0
7	10	0	2	100	1.75	1.14
8	6	4	1	60	1	1
9	2	3	1	40	1.83	0.55
10	5	0	2	100	1.5	1.33
11	8	2	1	80	0.95	1.05
12	11	0	0	100	1.43	0
13	9	1	2	90	1	2
14	10	0	1	100	1.33	0.75
15	6	0	0	100	2.03	0
16	1	0	0	100	2	0
17	1	0	1	100	1	1
18	9	1	2	90	1.03	1.94
19	0	1	0	0	0.82	0
20	5	5	0	50	0.7	0
21	7	3	1	70	0.9	1.11
22	3	7	2	30	0.94	2.13
23	1	0	0	100	0.75	0
24	8	2	1	80	1.65	0.61
25	1	0	0	100	0.17	0
26	6	3	0	66.7	1.77	0
27	3	5	0	37.5	0.93	0
28	7	3	0	70	0.5	0

Subjects	True detections	Missed Seizures	False detections	% Sensitivity	Data duration (hr)	False detection rate (/ hr)
Total	163	43	26	79.8 (av.)	33.6	0.86 (av.)

Table 2

Performance of the system for non-seizure subjects.

Subject	False Seizures	Total Data Duration (hr)	False Detection Rate (/hr)
1	0	2.05	0.00
2	0	2.05	0.00
3	0	1.97	0.00
4	1	2.16	0.46
5	0	0.23	0.00
6	0	2.06	0.00
7	2	2.01	1.00
8	0	2.13	0.00
9	0	2.04	0.00
10	2	2.03	0.99
11	0	2.21	0.00
12	0	0.32	0.00
13	0	2.03	0.00
14	0	2.12	0.00
15	2	2.00	1.00
16	0	2.04	0.00
17	4	1.54	2.60
18	0	2.07	0.00
19	0	1.78	0.00
20	4	1.98	2.02
21	0	2.03	0.00
22	9	2.12	4.25
23	3	0.96	3.13
24	0	2.07	0.00
25	1	2.30	0.43
26	1	1.85	0.54
27	0	2.09	0.00
28	1	2.04	0.49
29	0	1.68	0.00
30	1	1.73	0.58
31	0	2.01	0.00
32	0	2.08	0.00
33	13	2.01	6.47
34	0	2.03	0.00
35	0	1.12	0.00
36	1	2.18	0.46
37	0	1.73	0.00
38	3	1.68	1.79

Subject	False Seizures	Total Data Duration (hr)	False Detection Rate (/hr)
39	1	2.03	0.49
40	0	2.25	0.00
41	1	0.84	1.19
42	1	0.42	2.38
43	1	1.81	0.55
44	2	0.83	2.41
45	0	2.01	0.00
46	0	2.06	0.00
47	4	1.91	2.09
48	0	2.05	0.00
Total	58	86.73	0.74 (av.)

Table 3

Summary of missed seizures in testing data.

	Types of Misses	Description	% of misses	Average length (s)
a.	High-amplitude pseudo-sinusoidal discharge	Missed completely if slower than 0.4 Hz, detected otherwise as low quality	30%	Variable (30 s or longer)
b.	Depressed EEG, high-frequency seizure	Very low amplitude, spike-like, and very fast activity (> 6 Hz)	2% (one subject)	15 s
c.	Depressed EEG seizure with arrhythmic spikes	Very low-amplitude seizure, short in duration, and consisting of burst-like spike activity	15%	< 15 s
d.	Short-duration seizure discharges in subjects with longer seizures	No specific morphology; these seizures account for the percentage rejected by Ruleset 2 in Stage III due to short duration or low quality.	53%	< 20 s

Table 4

Summary of false seizure detections in testing data.

Non-seizure Subjects		Seizure Subjects				
a. Durations of false detections (s)						
<20	20-30	30-60	>60	20-30	30-60	>60
41	23	27	7.7	58	9	24
% of total						9

b. False detections caused by each Ruleset						
	Ruleset 1	Ruleset 2	Ruleset 3	Ruleset 1	Ruleset 2	Ruleset 3
% of total	62	7.7	30	57	34	8.6

c. Types of false detections (if they belong to specific categories)									
	Artif.	Rhythmic back-ground	Single-channel	Noisy Data	Artif.	Rhythmic interictal	Low amp.	Single-channel	Noisy Data
% of total	10	10	30	5	9	28	20	30	10