Clinical Neurophysiology 124 (2013) 2328–2335

Contents lists available at SciVerse ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Inter-ictal spike detection using a database of smart templates

Shaun S. Lodder [a,*], Jessica Askamp [a], Michel J.A.M. van Putten [a,b,*]

[a] Clinical Neurophysiology, MIRA-Institute for Biomedical Technology and Technical Medicine, University of Twente, The Netherlands
[b] Dept. of Neurology and Clinical Neurophysiology, Medisch Spectrum Twente, Enschede, The Netherlands

ARTICLE INFO

Article history:
Accepted 27 May 2013
Available online 20 June 2013

Keywords:
Electroencephalography (EEG)
Epilepsy
Epileptiform discharges
Automated detection
Template matching
Computerized interpretation

HIGHLIGHTS

- A novel spike detection method is presented making use of >2000 templates to find inter-ictal epileptiform activity in EEGs.
- Vulnerability to the variability in spike morphology is avoided by containing examples of most common spike morphologies in the database.
- Certainty values paired with each detection allows the reviewer to view most probable events first, further reducing the load on visual analysis.

ABSTRACT

Objective: Visual analysis of EEG is time consuming and suffers from inter-observer variability. Assisted automated analysis helps by summarizing key aspects for the reviewer and providing consistent feedback. Our objective is to design an accurate and robust system for the detection of inter-ictal epileptiform discharges (IEDs) in scalp EEG.
Methods: IED Templates are extracted from the raw data of an EEG training set. By construction, the templates are given the ability to learn by searching for other IEDs within the training set using a time-shifted correlation. True and false detections are remembered and classifiers are trained for improving future predictions. During detection, trained templates search for IEDs in the new EEG. Overlapping detections from all templates are grouped and form one IED. Certainty values are added based on the reliability of the templates involved.
Results: For evaluation, 2160 templates were used on an evaluation dataset of 15 continuous recordings containing 241 IEDs (0.79/min). Sensitivities up to 0.99 (7.24 fp/min) were reached. To reduce false detections, higher certainty thresholds led to a mean sensitivity of 0.90 with 2.36 fp/min.
Conclusion: By using many templates, this technique is less vulnerable to variations in spike morphology. A certainty value for each detection allows the system to present findings in a more efficient manner and simplifies the review process.
Significance: Automated spike detection can assist in visual interpretation of the EEG which may lead to faster review times.

1. Introduction

Epilepsy is estimated to affect around 50 million people worldwide. With a high temporal resolution that can capture inter-ictal epileptiform discharges (IEDs), EEGs play an important role in the diagnosis of epilepsy. A major drawback however, is that reviewing them is time-consuming. Also, the diagnostic yield is low, partially due to the relative short duration of routine EEG recordings, and due to this, multiple routine recordings are typically required before signs of inter-ictal epileptiform activity are found (Doppelbauer et al., 1993). Given that reviewers have different levels of training and experience, a reasonably high inter-observer reliability is also known to exist (Azuma et al., 2003; Abend et al., 2011).

Longer recordings have shown to increase the chances of finding inter-ictal epileptiform activity (Faulkner et al., 2012; Agbenu et al., 2012; Leach et al., 2006; Friedman and Hirsch, 2009). In some cases this can lead to fewer follow-up recordings and an overall increase in diagnostic efficiency. Unfortunately, longer recordings also result in more time required for visual analysis, a burden that

* Corresponding authors. Address: Dept. Clinical Neurophysiology, Building Carre, Science and Technology, P.O. Box 217, 7500 AE Enschede, The Netherlands. Tel.: +31 53 489 4599.
E-mail addresses: S.S.Lodder@utwente.nl (S.S. Lodder), M.J.A.M.vanPutten@utwente.nl (M.J.A.M. van Putten).

1388-2457/$36.00
http://dx.doi.org/10.1016/j.clinph.2013.05.019

is best avoided. Computerized assistance with the detection of IEDs can lessen the burden of visual analysis, and as an added benefit, ensure more consistency between reviews that will lower inter-rater variability. A large number of detection algorithms have already been proposed over the last four decades, ranging from template matching and parametric methods, to mimetic analysis, spectral analysis and artificial neural networks (see Wilson and Emerson, 2002; Halford, 2009). A review by Halford (2009) shows that some techniques have achieved very promising results, but due to a lack of a common dataset, comparisons between methods are still hard to make.

One of the main obstacles in IED detection is not only to find the inter-ictal events, but also to minimize the number of false detections per minute. If this number is too large, a reviewer will still be required to inspect most of the data and automated detection will be of no use. Most methods for this reason use a multi-step process to first detect possible events, and then discard artifacts and non-epileptiform events with additional rules from the set of potential candidates. Although difficult to compare without a common dataset, a good benchmark for the feasibility of spike detection methods in practice are their sensitivities and false detection rates. Low sensitivities show that reviewers can miss important IEDs if any are present, and as mentioned, high false detection rates make evaluating the output of the algorithm more time consuming than a visual review itself. Although selectivity measures also provide useful information about the algorithm, it does not take into account the number of IEDs per minute in the dataset, which is critical when comparing methods from different datasets. Recent methods report average sensitivities of 0.90 (Nonclercq et al., 2012), 0.65 (De Lucia et al., 2008), 0.82 (Argoud et al., 2006), 0.93 (Subasi, 2006), 0.78 (Halford et al., 2012), 0.69 (Ji et al., 2011a), and 0.92 (Van Hese et al., 2008), and average false detections per minute of 6 (De Lucia et al., 2008), 13.4 (Argoud et al., 2006), 0.92 (Indiradevi et al., 2008), and 0.09 (Ji et al., 2011a). In most studies that report high sensitivities, a high false detection rate is present if given, otherwise data epochs are used instead of continuous EEG, or the number of false detections are not reported (Nonclercq et al., 2012; Subasi, 2006; Halford et al., 2012; Indiradevi et al., 2008; Van Hese et al., 2008). On the other hand, studies with low false detection rates show lower sensitivities. This shows the trade off for all methods between choosing high sensitivities with more false detections, or lower sensitivities with fewer false detections.

Regardless of the many algorithms available, few implementations have made it to clinical practice. With the aim of achieving a practical and reliable system for IED detection, we introduce an automated detection method based on template matching to find IEDs in EEG recordings. The method is unique in the sense that it uses a large database of templates extracted from a training set of example EEGs. Unlike other template matching algorithms that only rely on a single or a small number of templates to search for inter-ictal events, our system can match more variations in spike morphology and is less vulnerable to variability. In addition, the algorithm is designed to incorporate experience into the templates from past classifications during training, which can be extended even further so that they can gain additional experience with repeated use. The presented method is intended to assist instead of replace visual inspection, with the aim of significantly reducing review time for clinicians and neurologists.

2. Methods

2.1. Subjects and data

An EEG dataset for this study was obtained from the Department of Clinical Neurophysiology at the Medisch Spectrum Twente (MST) in the Netherlands. Recordings were made with a standard 20–30 min protocol using a Brainlab EEG system and standard Ag–AgCl electrode caps placed according to the international 10–20 guidelines. Impedances were kept below 5 kΩ to reduce polarization effects. A sample rate of either 250 Hz or 256 Hz was used for each recording. Afterwards it was band-pass filtered between 0.5–30 Hz, and downsampled to 100 Hz to increase efficiency of the algorithm. IEDs were marked and reviewed by experienced electroencephalographers on the channels where they were clearly visible (MvP and JA). Each IED consisted of one of the following patterns: (i) spike, (ii) sharp wave, (iii) spike and slow wave, (iv) sharp wave and slow wave (v) slow wave with absent or very small preceding spike, or (vi) polyspike and slow wave. IEDs were marked in three montages: common reference, bi-polar and source.

The dataset consisted of 23 records with 723 IEDs in total and a combined recording length of 481 min. The patient group consisted of 15 males and 8 females with ages ranging from 4 to 52 (mean 23.2 ± 17.1) years. Thirteen subjects were diagnosed with generalized epilepsies (4 absence, 1 juvenile myoclonic, 8 idiopathic), and the remaining 10 with focal epilepsies (5 temporal lobe, 1 rolandic, 4 other). The dataset was split into two parts, one for template collection and training, and the other for an evaluation set. Eight EEGs were used for training and 15 for testing. The training set contained 482 marked IEDs with a total recording length of 175 min. The testing set contained 241 marked IEDs with a total recording length of 306 min.

2.2. Preprocessing

Independent component analysis was used to reduce the influence of eye blink artifacts. After calculating the independent components with the *FastICA* algorithm,[1] each component was compared to the electrooculogram (EOG) channel recorded with the EEG. If a component showed a substantial correlation with the EOG channel (>0.5), it was removed by setting all its values to zero. The remaining components were projected back to their channel space by applying the inverse transform.

2.3. Method outline

The presented method uses a database of templates resembling typical waveforms of epileptiform activity, and finds matches between the templates and inter-ictal events in the EEG. The main concepts that make this method different from others are the following: (i) a substantially larger set of templates is used, (ii) the templates are designed to have experience by remembering past classifications and using this to improve future predictions, and (iii) an agreement system is used to combine detected events from multiple templates and thus lower false detections. The sections that follow give a detailed outline of the various parts of the method, which can be divided into template collection, learning, and detection. An illustrative summary of the template collection and learning phase is shown in Fig. 1, and one for the detection phase in Fig. 2.

2.4. Template collection

Starting with a blank database, templates are added using a training dataset. For each EEG in the training set, a collection of epochs is extracted at the locations where IEDs were marked by the reviewers. Each epoch represents a template and is added to the database. Three montages are used during collection: common reference, bi-polar, and source. Templates are extracted per chan-

[1] Available at http://research.ics.aalto.fi/ica/fastica/.

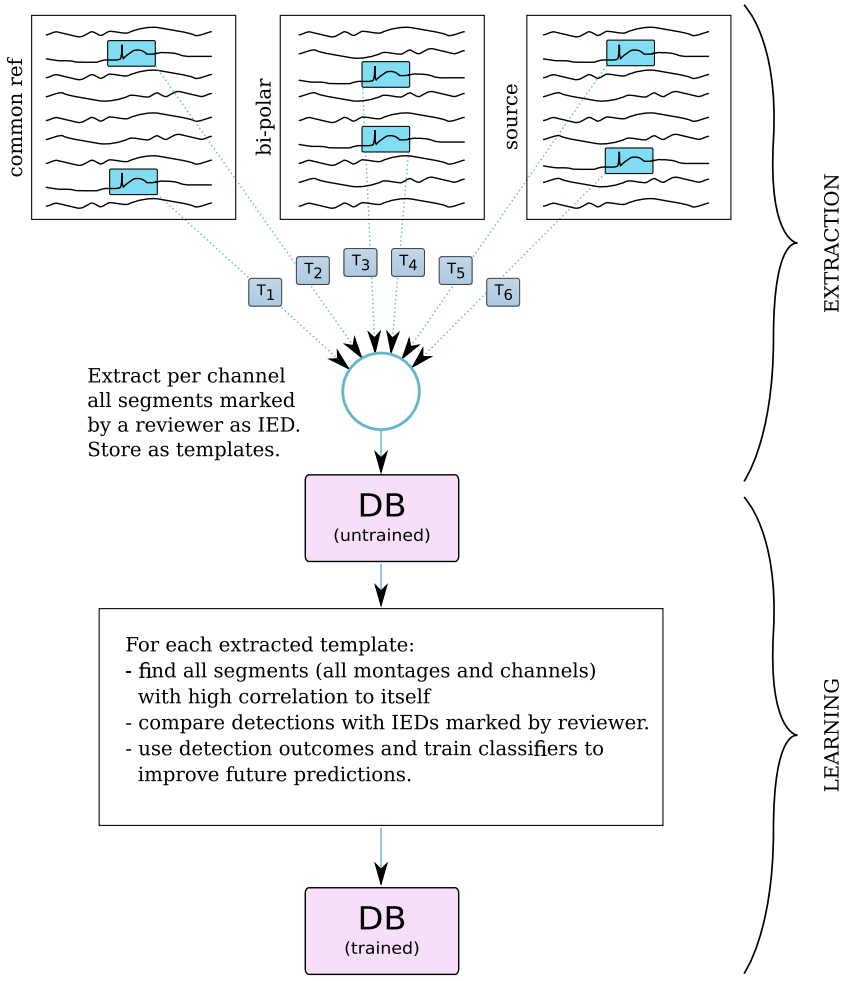

Fig. 1. Outline of the template extraction and learning steps. After the templates are trained, each of them contributes to the detection by nominating possible IEDs.

nel, i.e. if an IED is marked over two channels in the common reference montage and one channel in the bi-polar montage, three templates will be created. Single spikes or sharp waves were not included as templates, due to a high degree of false detections. An example of extracted templates is shown in Fig. 3. Note that they also have differences in magnitude (amplitudes were not normalized, but used 'as recorded').

2.5. Learning

After extracting the templates, their ability to find and discriminate between other IEDs from the same or other recordings and non-epileptiform activity is determined. This is done by finding a time-shifted correlation between each template and all EEG channels for all EEGs in the training set with the same montage. Due to the computational cost involved, this part of the algorithm was implemented with parallel processing. Locations are found where the templates have correlations above 0.85, and the underlying EEG segment is extracted to calculate additional properties to further determine the relationship between them and the template by which they were detected. For this, let a template consisting of N samples (N differs per template) be defined as $S_{tem}(n)$: $n \in [1 \ldots N]$, and each segment it finds as $S_{seg}(n)$: $n \in [1 \ldots N]$, as illustrated in Fig. 4. In addition, let the preceding N samples of each segment (same length as segment) be defined as $S_{prec}(n)$: $n \in [1 \ldots N]$ and the variance of the channel on which the segment is located as σ^2_{ch}. Then, the following properties are calculated:

$$f_{CRR} = \mathrm{corr}(S_{tem}, S_{seg}), \tag{1}$$

$$f_{DCRR} = \mathrm{corr}\left(S'_{tem}, S'_{seg}\right), \tag{2}$$

$$f_{MAD} = \frac{1}{N}\sum_{n}^{N} \|S_{tem}(n) - S_{seg}(n)\|, \tag{3}$$

$$f_{VRCHAN} = \frac{\sigma^2_{seg}}{\sigma^2_{seg} + \sigma^2_{ch}}, \tag{4}$$

$$f_{VRTEM} = \frac{\sigma^2_{seg}}{\sigma^2_{seg} + \sigma^2_{tem}}, \text{ and} \tag{5}$$

$$f_{VRPREC} = \frac{\sigma^2_{seg}}{\sigma^2_{seg} + \sigma^2_{prec}}, \tag{6}$$

where

$$\sigma^2_{seg} = \mathrm{var}(S_{seg}), \tag{7}$$

$$\sigma^2_{tem} = \mathrm{var}(S_{tem}), \text{ and} \tag{8}$$

$$\sigma^2_{prec} = \mathrm{var}(S_{prec}). \tag{9}$$

In addition to having a high correlation with the template ($f_{CRR} \geqslant 0.85$), these properties are used to determine the similarity between each detected segment and its template. Property f_{DCRR} measures the correlation between the derivatives of S_{seg} and S_{tem}, f_{MAD} finds the mean amplitude difference between S_{seg} and S_{tem}, and f_{VRCHAN}, f_{VRTEM} and f_{VRPREC} calculates the variance ratio between σ^2_{seg} and the sum of σ^2_{seg} with σ^2_{ch}, σ^2_{tem} and σ^2_{prec} respectively.

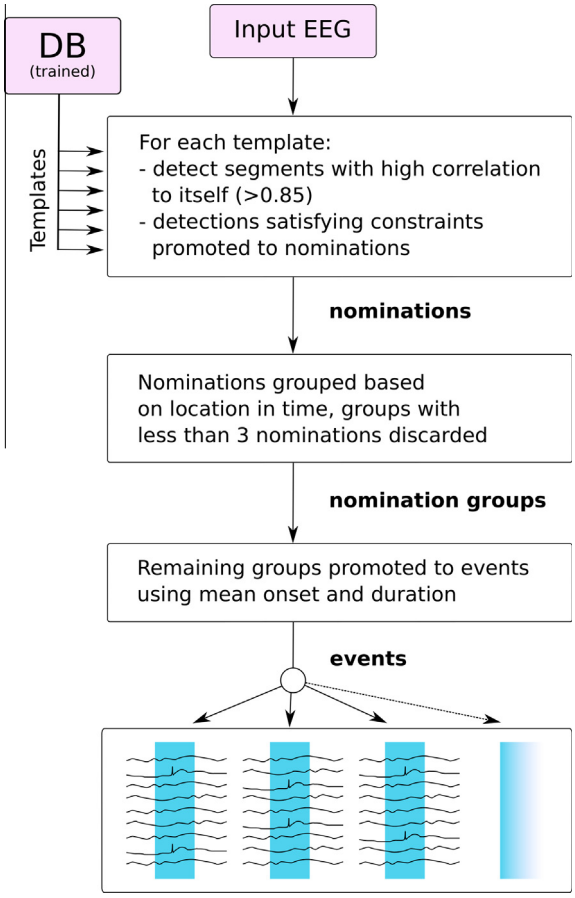

Fig. 2. Outline of the detection method. Trained templates detect and nominate possible IEDs. The system combines these nominations and identifies IEDs based the mutual agreement of multiple templates.

Using these properties, the location of each segment is converted to a "nomination" of an IED event if it satisfies both of the following conditions: (i) $f_{MAD} < 20$ μV, and (ii) $f_{VRCHAN} > 0.75$ or $f_{VRPREC} > 0.67$. Given that the correlation between two segments is scale invariant, the first requirement ensures that their magnitude is similar. The second requirement is based on the assumption that, for scalp EEG, the amplitude changes within an IED segment will be significantly larger than its preceding segment and the channel variance. Threshold values were chosen arbitrarily. Given that the true locations of IEDs are known in the training set, each nomination is either marked as correct or as a false detection, depending on if any of its samples overlap in time with a marked IED. At this point, the false detection rate of a template can be calculated:

$$FDR_{tem} = \frac{\text{false detections}}{\text{true detections} + \text{false detections}}. \quad (10)$$

Using only the criteria above to nominate events as IEDs lead to many false detections. Therefore, a linear support vector machine (SVM) is trained using the nominations in the training set to give the template additional experience in discriminating between true IEDs and false predictions. Vectors consisting of $\{f_{CRR}, f_{DCRR}, f_{VRTEM}\}$ are used as input. By having an SVM trained on past classifications, the template can now find events with high correlation to itself and predict if it is in fact an epileptiform event or not. After a template's SVM is trained, its reliability (R_{tem}) is calculated as the accuracy of re-classifying its own training vectors. The reliability reflects a templates ability to separate IEDs from non-epileptiform events, and is used as a weight factor during detection.

After all training steps are complete, templates that do not comply with the following conditions are discarded: (i) Templates with zero detections, (ii) templates with only false detections, (iii) templates with a small number of correct detections and no false detections. The last condition avoids templates that only find themselves in the training set and which have no use in finding other IEDs.

2.6. Detection

For detection, a similar approach is followed as during learning: For each template in the database, a time-shifted correlation is calculated on each EEG channel. Locations with correlations above 0.85 are nominated as IEDs if they also satisfy the conditions: (i) $f_{MAD} < 20$ μV and (ii) $f_{VRCHAN} > 0.75$ or $f_{VRPREC} > 0.67$, and using each template's own SVM, a prediction is made if the event is in fact an IED or only a similar looking artifact. False predictions are discarded and the remaining events are converted to nominations and added to a global pool. Nominations are stored with the following information about themselves: (a) its onset and duration, (b) the channel on which it was found, (c) f_{CRR}, and (d) R_{tem}.

Next, nominations are grouped together if they reside on the same channel and overlap with more than 75%. Groups with fewer than three nominations are considered unreliable and are discarded. Each of the remaining groups is considered a detected event, and the onset and duration is taken from the mean of the nominations in the group. The event is also assigned a detection certainty, which is done by calculating the mean product of each nomination's correlation and template reliability:

$$certainty = \frac{1}{M}\sum_{m}^{M} R_{tem}(m) f_{CRR}(m), \quad (11)$$

with M the number of nominations in that group. Lastly, events from individual channels are merged into one if they overlap in time, and once again the mean onset, duration and certainty is used to describe the event.

3. Results

A total of 2632 templates were collected during training, of which 472 were discarded for not passing the constraints described in Section 2.5. To gain some insight on the behavior of the remaining 2160 templates, histograms of FDR_{tem} and R_{tem} are provided in Fig. 5(a) and (b) respectively. The mean false detection rate before training SVMs is 0.74 and the mean template reliability after SVM training 73.8%. The two histograms show that the templates by themselves have high false detection rates and suboptimal accuracies in finding IEDs. Fig. 5(b) shows however how the use of SVMs and additional properties can help reduce the number of false detections and improve accuracy. Fig. 6(a) gives an ROC curve from all detected events in the test EEGs, and it shows that although low accuracies are achieved at template level, the combined information from individual templates lead to reliable detections on a global level. As seen in Fig. 6, the certainty threshold can be adjusted to either maximize the sensitivity or minimize the false detection rate. Fig. 6(b) and (c), respectively show the sensitivity and false detection rate as a function of minimum certainty, and Table 1 gives corresponding values to a number of points on the curves in Fig. 6. For all test EEGs combined, a maximum sensitivity of 0.99 is reached at a false detection rate of 7.24 false positives per minute. Note that the IED rate in the evaluation set is only 0.79 IEDs per minute, and so a false detection rate of 7.24 will be unacceptable for practical use. Although far from ideal, we consider a more acceptable false detection rate for reviewing to be below three per minute. The mean sensitivities obtained after

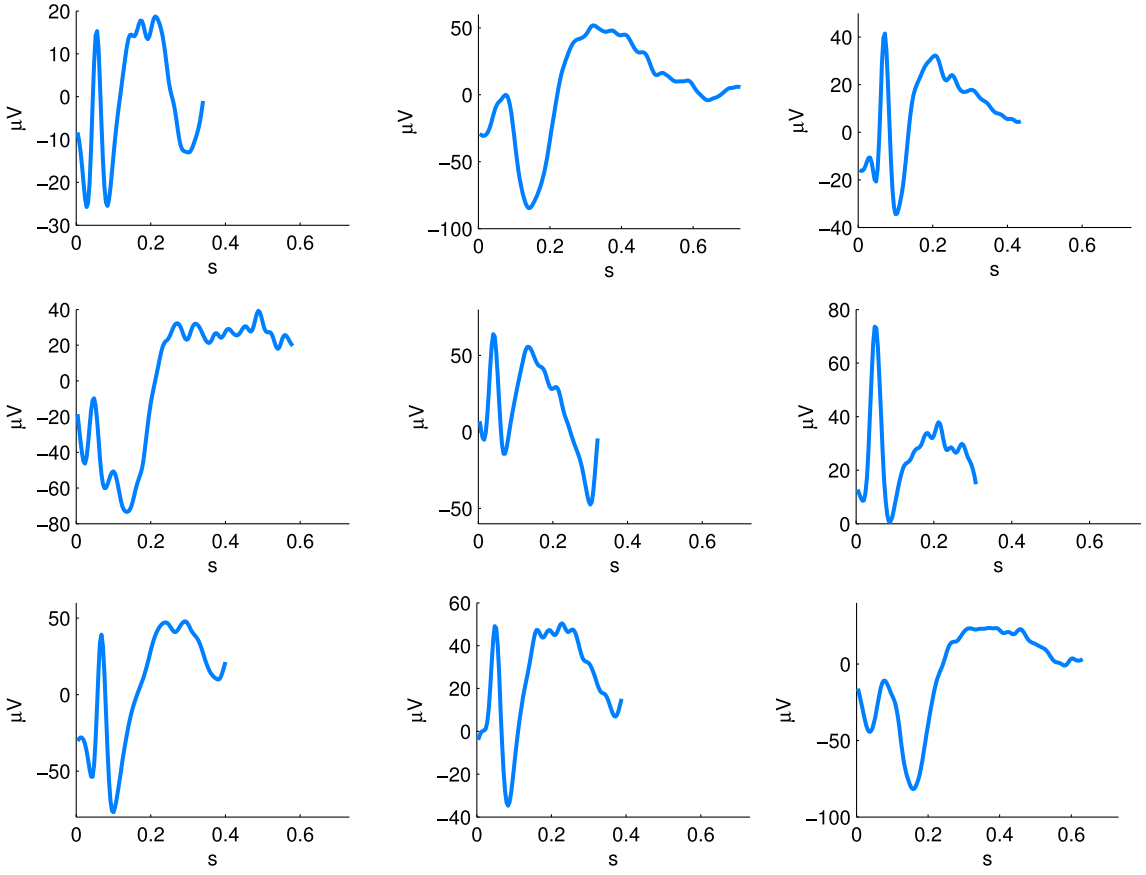

Fig. 3. Nine templates taken from the trained database consisting of 2160 templates. Template lengths are allowed to differ and amplitudes are preserved.

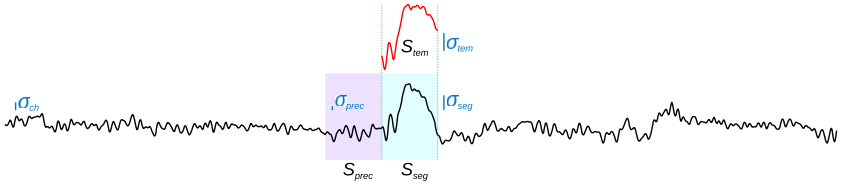

Fig. 4. Template features: To find matching events, templates use correlation coefficients, amplitude differences and variance ratios between a template and its detected segment. Variance ratios add temporal context to a detected event by comparing it to its surrounding activity.

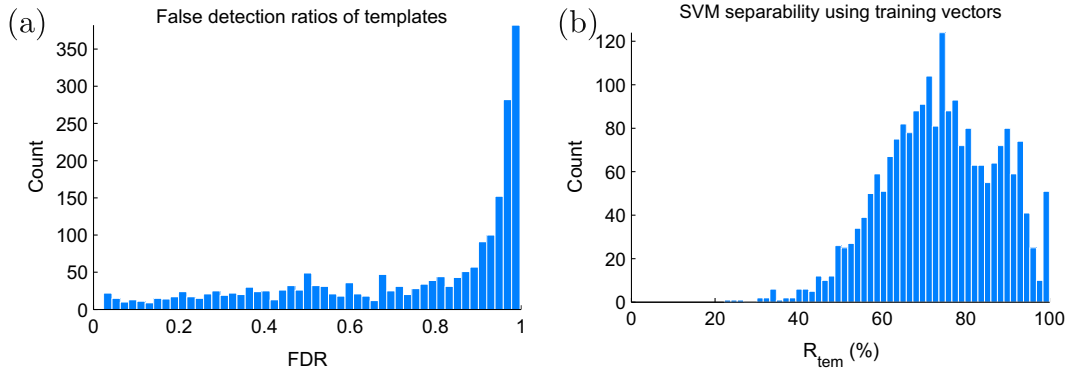

Fig. 5. Histograms combining individual template statistics: (a) false detection ratios; and (b) R_{tem}: the ability of each template's SVM to separate its training vectors. The two histograms show that by themselves the templates have high false detection rates and in many cases poor separability between IEDs and non-IED events. The power of the system lies in combining the information and using a network effect to allow accurate predictions.

adjusting the certainty threshold of each EEG individually is shown in Table 2. The thresholds were chosen in such a way as to limit the false positives per minute to no more than three. After this, a mean sensitivity of 0.90 and mean false detection rate of 2.36 min^{-1} is

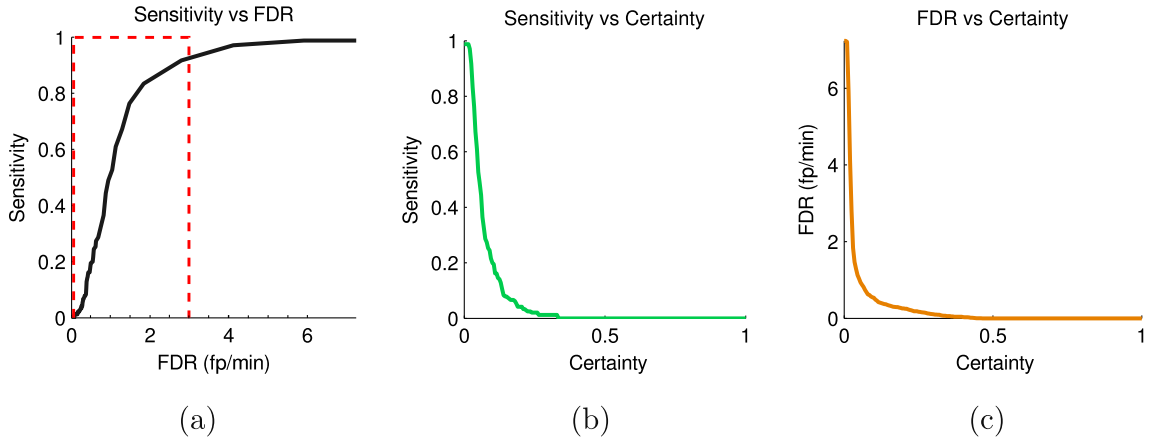

Fig. 6. Evaluation of the described method on a test set of 15 EEGs: (a) an ROC curve shows how the sensitivity of detecting IEDs can be improved at the price of more false detections. The dashed rectangle gives an approximate range of false detection rates that we consider as more feasible for practical use; (b) higher confidence values reduce the sensitivity, (c) but also lowers the false detection rate.

Table 1
Choosing to maximize the sensitivity or minimize the false detection rate can be done by adjusting the minimum confidence threshold. This allows a reviewer to view the most probable IEDs first, and then to look at additional events if desired. Values below correspond to points on Fig. 6(a–c).

Min confidence	Mean sensitivity	Mean Fp/min
0	0.99	7.24
0.02	0.97	4.12
0.04	0.67	1.29
0.06	0.44	0.87
0.08	0.27	0.62
0.1	0.2	0.54
0.12	0.15	0.41
0.14	0.08	0.37
0.16	0.07	0.31
0.18	0.06	0.29
0.2	0.04	0.26
0.22	0.03	0.22
0.24	0.02	0.19
0.26	0.02	0.16

Table 2
Mean sensitivities and false detection rates per EEG in the test set using a varying confidence threshold value per EEG and limiting false detections to a maximum of three per minute.

EEG	Dur (min)	IEDs	SEN	Fp/min
S1	16.3	13	1.00	1.59
S2	22.5	11	0.82	2.93
S3	20.0	5	1.00	1.40
S4	20.0	36	0.97	2.60
S5	22.3	19	0.79	1.93
S6	19.5	6	1.00	2.72
S7	20.3	79	0.97	1.82
S8	21.2	5	0.80	1.98
S9	20.0	2	1.00	2.95
S10	20.0	12	1.00	2.00
S11	21.5	7	0.38	2.88
S12	20.3	7	0.86	2.95
S13	20.7	19	0.95	2.08
S14	21.0	14	1.00	2.95
S15	20.0	6	1.00	2.60
Mean	20.4	16.07	0.90	2.36

achieved. The selection of a threshold cannot be chosen a priori to limit the false detection rate to a known maximum, but this shows that detections with higher certainties are more reliable. In a typical usage scenario, a user may choose to view events with the highest certainties first, and then gradually reduce the threshold if he wishes to see more.

In addition to the detection algorithm, the described method was also implemented into an EEG viewer. Fig. 7 shows how each of the detected events are presented, and using the options provided, the reviewer can decide to either confirm, reject, or indicate doubt concerning each detected event. In such a way, the method can be used to assist during a review by pointing out areas of interest in the EEG and allowing the reviewer to give the final verdict.

4. Discussion

Automated detection of inter-ictal epileptiform discharges is a crucial step towards improving the efficiency of epilepsy diagnostics and monitoring. Not only does it reduce the review time significantly and make longer recordings feasible, but it also allows for more objective analysis with less inter-observer variability. The presented method uses a large collection of templates to detect inter-ictal epileptiform discharges. It keeps track of past classifications during training in order to provide the templates with additional statistical experience, and the size of the database allows it to contain example spike-and-slow-wave patterns of many morphologies so that IEDs can be detected across multiple patients and recordings, regardless of whether their inter-ictal patterns look the same. In addition, using the experience of each template, the system can determine its reliability during detection. This approach resembles the way in which humans search for IEDs: by finding waveforms that have properties associated with IEDs, and mentally comparing them to patterns observed in the past. In addition to high correlations between templates and detected EDs, additional features were chosen to add information and improve detection accuracy. Given that correlations are scale independent, the mean amplitude difference (f_{MAD}) ensures that matching events are on the same scale, and the variance ratios (f_{VRCHAN}, f_{VRTEM}, f_{VRPREC}) ensure that IEDs are significantly different from its surrounding activity. By adding these properties, the method is capable of not only using wave morphology, but also temporal context.

To improve accuracy, the system removes uncertainties of single templates by using a network effect to make decisions. In other words, even though individual templates by themselves miss events and make many false detections, the collective value of their votes result in accurate detections. Using this approach, the system is less vulnerable to variability in spike morphology than other methods which rely on single features or a small number of templates. Still, the system is highly dependent on the templates

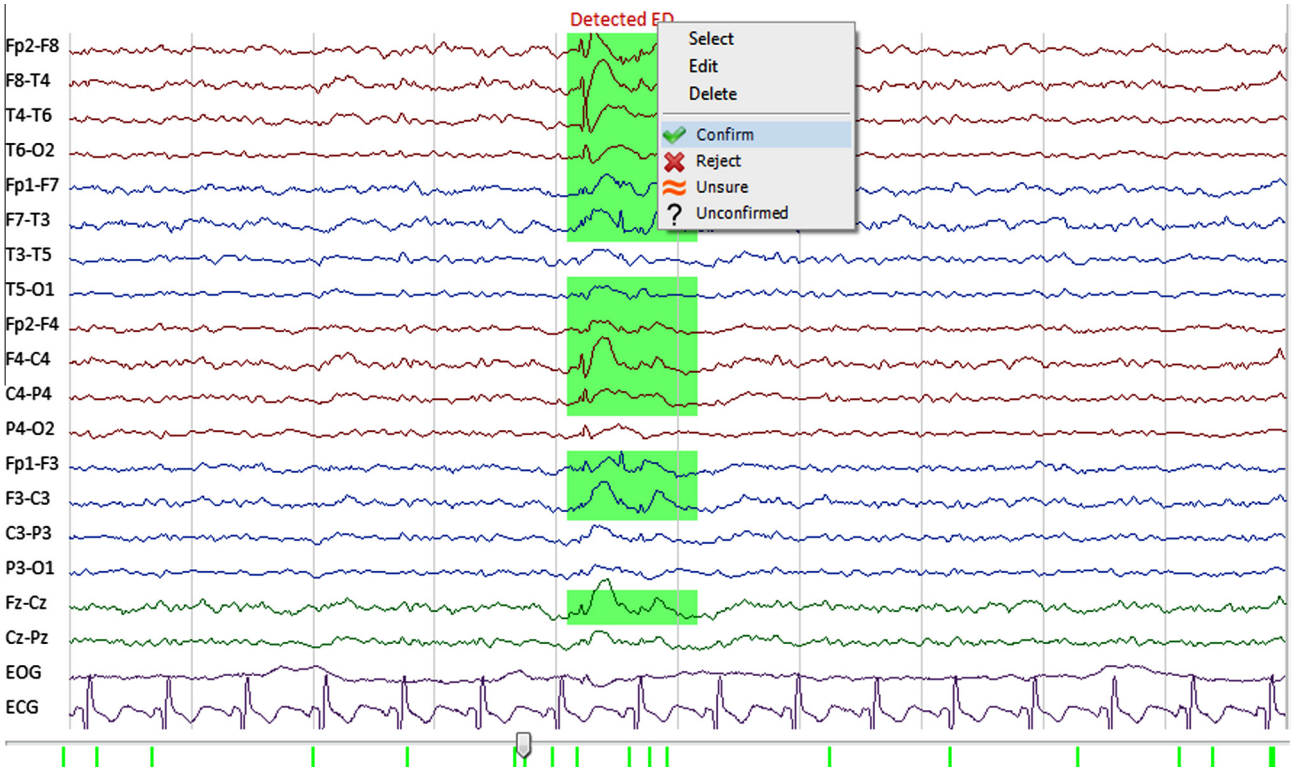

Fig. 7. A detected IED highlighted on top of the EEG, as displayed to the reviewer in a bi-polar montage.

contained within its database. Our results show that most IEDs can be detected with the current set of templates and thresholds described. However, the ideal set (and number) of templates to find all common inter-ictal patterns is not known, and it could be that additional templates with more variations in spike and wave morphology will allow the system to match events more accurately, allowing higher thresholds for individual detections, and thereby lowering the overall false detection rate.

A major advantage of this technique is that the system remains dynamic and that more templates can be added if required. By further extending the method and adding a feedback loop during reviews, the database can become adaptive. As an example, missed events can be added as templates if a reviewer finds any IEDs visually during a review. Another example is if template reliabilities change when new EEGs are reviewed and false detections are marked by a reviewer. As currently implemented, templates search for matching IEDs in parallel to each other. Although not strictly necessary, this speeds up the detection process significantly and makes the system scalable and fast enough for clinical use. Our implementation of the method runs on a standard desktop computer (Intel i7-2600 3.40 GHz CPU, 4 GB RAM) and requires no additional hardware.

Reviews by Wilson and Emerson (2002) and Halford (2009) summarize most of the reported techniques over the last four decades. In addition to these, newer methods have also been reported (Nonclercq et al., 2009, 2012; Scherg et al., 2012; Ji et al., 2011a,b; Igasaki et al., 2011; Argoud et al., 2006; Nenadic and Burdick, 2005; Wang et al., 2010; Zhou et al., 2012). Comparisons between different methods are hard to make because there is no common dataset currently available. Without a benchmark dataset, reported results rely heavily on the data used. For example, some studies use standard 20 min recordings to evaluate their methods, whereas others use longer recordings or smaller EEG segments simply categorized as either containing IEDs or not. If the number of IEDs per minute are not within the same range between datasets, measures such as the specificity, selectivity and the false detection rate cannot be compared in a fair manner. As suggested by Halford, 2009, a multi-center project is needed to create a dataset for reliable benchmarks. The author later presents such a dataset, and although still under development, it may suffice for benchmark testing and future comparisons (Halford et al., 2011, 2012). Regardless of using different datasets, our method compares well to others. Depending on the confidence threshold chosen, the system achieves sensitivities of up to 0.99. With thresholds chosen as to limit the false detection rate, a mean sensitivity of 0.90 and mean false detection rate of 2.36 is achieved. We feel that the threshold should be set interactively by the reviewer during runtime, which can start by viewing the most probable IEDs first with a high threshold value, and then gradually view more detections (with greater false detection rate) if needed by lowering it.

Apart from achieving high accuracies in epileptiform event detection, automated methods should also be responsible for presenting their findings to the reviewer in a fast and efficient manner. Although some commercial applications already exist, automated detection is mostly ignored by practicing neurologists. One reason for this could be that the software is still too complicated and time consuming to use. Fig. 7 shows an example of how a detected event is presented to a clinician. To make review time faster, events can be grouped by location or morphology and reviewing can take place on a macro scale to save time. This is already shown in some studies (and some products) who use clustering techniques to group similar looking events together (Nonclercq et al., 2012; Scherg et al., 2012; Ji et al., 2011a).

Although our method shows promising results, it is still far from ideal and some reported methods appear to perform better. Many of them however, have already been refined to the point where further improvement is limited. Our method, as presented here, is only a first step in testing a novel approach where many improvements can still be made. In work to follow, we will try to use more of the spatial context available to discard artifacts and lower the

false detection rate. One way that we can achieve this is by introducing event clustering. Events with similar properties can be clustered, and each cluster's validity as IEDs can be determined based on the spatial distribution and frequency of occurrence of the events it contains. Apart from clustering, a number of ad hoc rules can also be applied as a post-detection step to mimic the procedures followed by reviewers when determining the validity of IEDs. Further work will also focus on testing the system on long-term EEG recordings such as in-home ambulatory monitoring, where the burden of visual inspection is even greater and more problematic for the reviewer.

In summary, automated detection of inter-ictal epileptiform activity will become invaluable to current reviewing techniques that rely mostly on visual analysis. A reliable system for a wide range of EEGs with different spike morphologies is needed, which should be tested with a gold-standard dataset for comparisons with other techniques. With minor improvements and an intuitive interface, the presented method is capable of providing more efficiency with reviewing and in turn improve the diagnostic efficiency of epilepsy.

Acknowledgments

The authors would like to extend their gratitude to the reviewers for their valuable comments and suggestions that contributed to the quality of this work.

References

Abend NS, Gutierrez-Colina A, Zhao H, Guo R, Marsh E, Clancy RR, et al. Interobserver reproducibility of electroencephalogram interpretation in critically ill children. J Clin Neurophysiol 2011;28:15–9.

Agbenu J, Newton RW, Martland T, Ismayl O, Hargreaves S. Effect of reducing the recording time of standard EEGs on the detection of EEG-abnormalities in the management of the epilepsies of childhood. Seizure 2012;21:422–5.

Argoud FIM, De Azevedo FM, Neto JM, Grillo E. SADE3: an effective system for automated detection of epileptiform events in long-term EEG based on context information. Med Biol Eng Comput 2006;44:459–70.

Azuma H, Hori S, Nakanishi M, Fujimoto S, Ichikawa N, Furukawa TA. An intervention to improve the interrater reliability of clinical EEG interpretations. Psychiatry Clin Neurosci 2003;57:485–9.

De Lucia M, Fritschy J, Dayan P, Holder DS. A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis. Med Biol Eng Comput 2008;46:263–72.

Doppelbauer A, Zeitlhofer J, Zifko U, Baumgartner C, Mayr N, Deecke L. Occurrence of epileptiform activity in the routine EEG of epileptic patients. Acta Neurol Scand 1993;87:345–52.

Faulkner HJ, Arima H, Mohamed A. Latency to first interictal epileptiform discharge in epilepsy with outpatient ambulatory EEG. Clin Neurophysiol 2012;123:1732–5.

Friedman DE, Hirsch LJ. How long does it take to make an accurate diagnosis in an epilepsy monitoring unit? Clin Neurophysiol 2009;26:213–7.

Halford JJ. Computerized epileptiform transient detection in the scalp electroencephalogram: obstacles to progress and the example of computerized ECG interpretation. Clin Neurophysiol 2009;120:1909–15.

Halford JJ, Pressly WB, Benbadis SR, Tatum WO, Turner RP, Arain A, et al. Web-based collection of expert opinion on routine scalp EEG: software development and interrater reliability. J Clin Neurophysiol 2011;28:178–84.

Halford JJ, Schalkoff RJ, Zhou J, Benbadis SR, Tatum WO, Turner RP, et al. Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis. J Neurosci Methods 2012;212:308–16.

Igasaki T, Higuchi T, Hayashida Y, Murayama N, Neshige R. Proposal for patient-specific automatic on-line detection of spike-and-wave discharges utilizing an artificial neural network. In: 2011 Fourth international conference on biomedical engineering and informatics (BMEI). IEEE; 2011. p. 813–7.

Indiradevi KP, Elias E, Sathidevi PS, Dinesh Nayak S, Radhakrishnan K. A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. Comput Biol Med 2008;38:805–16.

Ji Z, Sugi T, Goto S, Wang X, Ikeda A, Nagamine T, et al. An automatic spike detection system based on elimination of false positives using the large-area context in the scalp EEG. IEEE Trans Biomed Eng 2011a;58:2478–88.

Ji Z, Wang X, Sugi T, Goto S, Nakamura M. Automatic spike detection based on real-time multi-channel template. In: 2011 Fourth international conference on biomedical engineering and informatics (BMEI); 2011b. p. 648–652.

Leach JP, Stephen LJ, Salveta C, Brodie MJ. Which electroencephalography (EEG) for epilepsy? The relative usefulness of different EEG protocols in patients with possible epilepsy. J Neurol Neurosurg Psychiatry 2006;77:1040–2.

Nenadic Z, Burdick JW. Spike detection using the continuous wavelet transform. IEEE Trans Biomed Eng 2005;52:74–87.

Nonclercq A, Foulon M, Verheulpen D, De Cock C, Buzatu M, Mathys P, et al. Spike detection algorithm automatically adapted to individual patients applied to spike-and-wave percentage quantification. Neurophysiol Clin 2009;39:123–31.

Nonclercq A, Foulon M, Verheulpen D, De Cock C, Buzatu M, Mathys P, et al. Cluster-based spike detection algorithm adapts to interpatient and intrapatient variation in spike morphology. J Neurosci Methods 2012;210:259–65.

Scherg M, Ille N, Weckesser D, Ebert A, Ostendorf A, Boppel T, et al. Fast evaluation of interictal spikes in long-term EEG by hyper-clustering. Epilepsia 2012;53:1196–204.

Subasi A. Automatic detection of epileptic seizure using dynamic fuzzy neural networks. Expert Syst Appl 2006;31:320–8.

Van Hese P, Vanrumste B, Hallez H, Carroll GJ, Vonck K, Jones RD, et al. Detection of focal epileptiform events in the EEG by spatio-temporal dipole clustering. Clin Neurophysiol 2008;119:1756–70.

Wang C, Zou J, Zhang J, Wang M, Wang R. Feature extraction and recognition of epileptiform activity in EEG by combining PCA with ApEn. Cogn Neurodyn 2010;4:233–40.

Wilson SB, Emerson R. Spike detection: a review and comparison of algorithms. Clin Neurophysiol 2002;113:1873–81.

Zhou J, Schalkoff RJ, Dean BC, Halford JJ. Morphology-based wavelet features and multiple mother wavelet strategy for spike classification in EEG signals. In: 2012 Annual international conference of the IEEE engineering in medicine and biology society. IEEE Engineering in Medicine and Biology Society; 2012. p. 3959–62.