Brain Topogr (2015) 28:172–183
DOI 10.1007/s10548-014-0379-1

ORIGINAL PAPER

Detection of Interictal Epileptiform Discharges Using Signal Envelope Distribution Modelling: Application to Epileptic and Non-Epileptic Intracranial Recordings

Radek Janca · Petr Jezdik · Roman Cmejla · Martin Tomasek · Gregory A. Worrell · Matt Stead · Joost Wagenaar · John G. R. Jefferys · Pavel Krsek · Vladimir Komarek · Premysl Jiruska · Petr Marusic

Received: 19 August 2013 / Accepted: 27 May 2014 / Published online: 27 June 2014

Abstract Interictal epileptiform discharges (spikes, IEDs) are electrographic markers of epileptic tissue and their quantification is utilized in planning of surgical resection. Visual analysis of long-term multi-channel intracranial recordings is extremely laborious and prone to bias. Development of new and reliable techniques of automatic spike detection represents a crucial step towards increasing the information yield of intracranial recordings and to improve surgical outcome. In this study, we designed a novel and robust detection algorithm that adaptively models statistical distributions of signal envelopes and enables discrimination of signals containing IEDs from signals with background activity. This detector demonstrates performance superior both to human readers and to an established detector. It is even capable of identifying low-amplitude IEDs which are often missed by experts and which may represent an important source of clinical information. Application of the detector to non-epileptic intracranial data from patients with intractable facial pain revealed the existence of sharp transients with waveforms reminiscent of interictal discharges that can represent biological sources of false positive detections. Identification of these transients enabled us to develop and propose secondary processing steps, which may exclude these transients, improving the detector's specificity and having important implications for future development of spike detectors in general.

Premysl Jiruska and Petr Marusic wish it to be known that, in their opinion, the last two authors should be regarded as joint Senior Authors and joint Corresponding Authors.

R. Janca · P. Jezdik · R. Cmejla
Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

M. Tomasek · P. Jiruska · P. Marusic (✉)
Department of Neurology, 2nd Faculty of Medicine, Motol University Hospital, Charles University in Prague, Prague, Czech Republic
e-mail: petr.marusic@fnmotol.cz

G. A. Worrell · M. Stead
Mayo Systems Electrophysiology Laboratory, Mayo Clinic, Rochester, MN, USA

J. Wagenaar
Department of Neurology and Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

J. G. R. Jefferys
Neuronal Networks Group, School of Clinical and Experimental Medicine, University of Birmingham, Birmingham, United Kingdom

J. G. R. Jefferys
Department of Pharmacology, University of Oxford, Oxford, United Kingdom

P. Krsek · V. Komarek
Department of Paediatric Neurology, 2nd Faculty of Medicine, Motol University Hospital, Charles University in Prague, Prague, Czech Republic

P. Jiruska (✉)
Department of Developmental Epileptology, Institute of Physiology, Academy of Sciences of Czech Republic, Prague, Czech Republic
e-mail: jiruskapremysl@gmail.com

Keywords Spike detection · Interictal epileptiform discharges · Intracranial recording · Automatic detection · Hilbert transform · Principal component analysis

Introduction

Interictal epileptiform discharges (spikes, IEDs) are electrographic markers of epileptic tissue (de Curtis and Avanzini 2001). They are used to diagnose epilepsy, monitor disease activity and localize epileptogenic tissue. Localizing properties of interictal discharges are extensively used during the presurgical examination (Rosenow and Lüders 2001) to obtain information about the epileptic network organization and to plan the site and extent of resection (Asano et al. 2003; Bautista et al. 1999; Marsh et al. 2009; Wilke et al. 2009). Currently, multichannel intracranial recording represents the standard technique from which the majority of information about the functional organization of epileptic network is obtained. However, clinical demands for more precise and better delineation of epileptogenic cortex result in utilization of new recording techniques which enable large-scale recording from wide areas of brain, providing increased spatiotemporal resolution (Brinkmann et al. 2009; Van Gompel et al. 2008). The downside of this approach is that the amount of recorded data is massive (Brinkmann et al. 2009). Visual analysis of long-term recordings from hundreds of channels is extremely difficult, virtually impossible, and introduces human bias (Gotman 1999).

Accurate detection of interictal discharges still represents an unmet need in clinical neurophysiology and presents a challenge for future improvement in presurgical evaluation and surgical outcome. The main principles of detection and labelling of electrographic epileptiform phenomena can be classified into three main groups: manual review, supervised (semiautomatic) detection and unsupervised (automatic) detection (Wilson and Emerson 2002; Worrell et al. 2012). In particular, automatic detection represents an important step to facilitate processing of large amounts of clinical data obtained from long-term multichannel invasive recordings. Various automatic spike-detecting algorithms have been developed and are extensively reviewed by Frost (1985), Gotman (1986), Wilson and Emerson (2002) and Halford (2009). Unfortunately, the practical implementation of spike detection is still limited. The validation of spike detectors is complicated by low inter-rater agreement in marking IEDs (Barkmeier et al. 2012; Webber et al. 1993; Wilson et al. 1996). An alternative strategy is to consider the output of unbiased automatic (or at least semiautomatic) algorithms as true detections. This approach requires detailed identification of sources of false detections and definition of features which would enable their classification and separation by higher-level processing. In scalp EEG biological sources of false detections include muscle artefacts, eye movements, alpha waves etc. and several methods were designed to eliminate them (Jung et al. 1998; Yuan et al. 2012). In contrast to scalp EEG intracranial recordings are considered to be largely (but not always) immune to myogenic or oculogenic artefacts. However, studies (Dumpelmann and Elger 1999; Lodder et al. 2013) which applied spike detectors to intracranial recordings still demonstrate the presence of false detections. Therefore, the design of reliable detectors of intracranial spikes requires better understanding of the most common sources of false positives (FP).

In our article, we present a new adaptive automated spike detector based on modelling of statistical distributions of signal envelopes containing spikes and background activity. To maximize the performance and identify the crucial components that an effective intracranial spike detecting algorithm should possess we applied the detector to both epileptic recordings and non-epileptic data obtained in patients with intractable facial pain implanted with intracranial electrodes for the purpose of invasive brain stimulation (Stead et al. 2010). In addition, this approach allowed us to identify common sources of false detections in intracranial signals.

Methods

Patient Selection and Data Acquisition

Data used in this study were recorded in patients with refractory epilepsy who underwent invasive exploration as a part of the presurgical examination. Data from 15 adult and 15 pediatric patients were collected. Research procedures and data collection were approved by the institutional ethical committee and patient or parent informed consent was obtained. Signals from subdural and/or depth macroelectrodes (Ad-Tech Medical Instrument Corporation, Racine, USA) were amplified (Schwarzer GmbH, Heilbronn, Germany), filtered using antialiasing filters at 1/3 of sampling frequency and sampled at frequency 200, 250 or 1,000 Hz (Stellate Inc., Montreal, Canada). Signals were recorded in reference mode. One of the implanted electrodes without epileptiform activity was selected as a reference electrode.

Evaluating automatic detection techniques requires gold standard events reliably identified by human readers. To this end we used an automated method to blindly select recordings to avoid those with very large numbers of IEDs, which complicate human labelling, and those with few or no IEDs. This approach was designed to avoid the substantial bias that can arise from subjective selection of the recordings (Gotman 1999). Specifically, we used the

Table 1 Patients characteristics

Dataset	Sex	Age	Epilepsy	Duration (years)	Pathology
Patient 1	F	24	Right frontal	6	No abnormality
Patient 2	F	37	Left frontal	28	No abnormality
Patient 3	F	17	Left temporal	2	FCD Ib
Patient 4	F	8	Right frontal	3	FCD IIb
Patient 5	F	14	Right multilobar	8	HS, FCD Ia
Patient 6	M	31	Right frontal	27	FCD IIb
Patient 7	M	10	Left multilobar	3	FCD IIa

FCD focal cortical dysplasia, *HS* hippocampal sclerosis

following automatic procedure to blindly select recordings from our database of 30 patients. The procedure used signal thresholding to detect high amplitude interictal discharges and it was applied to all interictal recordings in the database. Mean and standard deviation of the incidence of detected discharges in each dataset was determined and normalized with respect to its duration. Recordings with maximal IED rates below or above two times standard deviation were excluded. From the remaining datasets, we randomly selected interictal recordings from seven patients. This represented manageable sample necessary to achieve the goals of this study, i.e. to perform human labelling, identification of gold standard IED and then for detector optimization, validation and testing. To disguise the information about spike propagation which could bias human labelling, only 15 channels with the highest spike rate were selected from each dataset that originally contained 66.9 ± 28.7 channels (median 65 channels). In total, 105 channels out of original 460 were analysed and the complete procedure generated seven interictal 15-channel datasets, each with duration of 5 min. The patients characteristics are included in Table 1.

Non-epileptic data were obtained in patients with intractable chronic facial pain who underwent intracranial electrode implantation for experimental brain stimulation treatment (Lima and Fregni 2008; Stead et al. 2010). Electrodes were positioned in primary motor and premotor cortices. Data were acquired on a DC capable 320 channel system, sampled at 32 kHz, with a dynamic range of ±132 mV at 1 μV resolution (Neuralynx, Inc.). The averaged signal was used as the common reference for each channel.

Human Spike Labelling

For the purpose of expert spike labelling we have developed signal viewer which shares similar features with any standard EEG viewer but facilitates spike labelling and spike classification. Three experienced neurophysiologists independently reviewed the selected data and labelled the spikes. They have 10 (Reader 1), 13 (Reader 2) and 10 (Reader 3) years of experience with human and animal intracranial EEG. During the labelling process reviewers were allowed to adjust the amplitude and time resolution of the plotted signal, but not the filter settings. Signals were assessed only in reference montage with negativity plotted upwards. All reviewers were blind to any information about the datasets (location of electrodes, clinical data etc.).

If a spike was identified its entire waveform was labelled. Spikes in each of the 15 channels were marked separately and immediately classified into one of two groups—obvious spikes or ambiguous ones. Waveforms which were associated with any kind of hesitation and/or doubts whether they should be classified as an obvious spike were classified as ambiguous. Polyspikes were labelled as a single event. In case of episode of repetitive interictal discharges experts labelled each spike of the episode. Obvious and ambiguous spikes on which two or more readers agreed were used as a "gold standard" spikes for detector performance testing (Barkmeier et al. 2012; Webber et al. 1993).

Spike Detection Algorithm

The principle of the spike detection is based on distinct morphological and statistical properties between spikes and background activity; schematically given in Fig. 1. Traditionally, a spike is clinically defined as a sharp transient with duration between 20 and 70 ms and clearly distinguishable from background activity (Gloor 1975; Niedermeyer and Lopes da Silva 2004). In the frequency spectrum it is characterized as a local energy increase particularly in

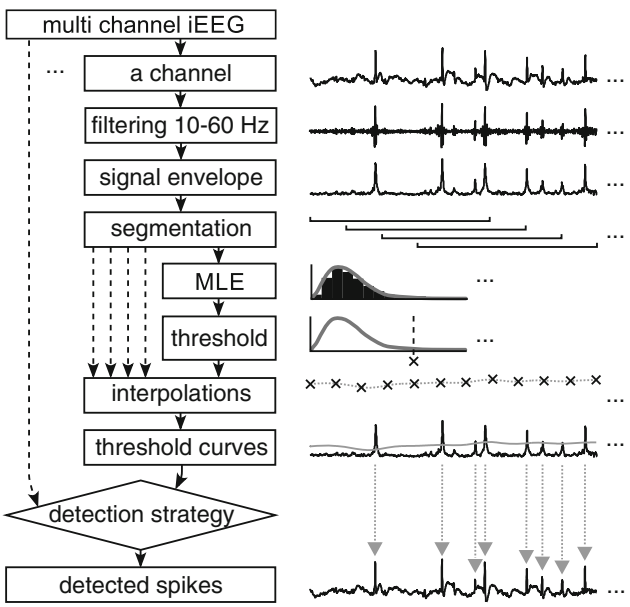

Fig. 1 Schematics of detection algorithm. *MLE* maximum-likelihood estimation method

14.3–50 Hz frequency band (inverse value of spike duration $f = T^{-1}$).

Initially, signals with higher sampling frequency were resampled to 200 Hz to maintain filter characteristics constant. Each channel was zero-phase filtered in 10–60 Hz band using combination of high-pass and low-pass 8th order type II Chebyshev digital filters (with stop-band ripple). Notch biquad filter with poles on 0.985-circle were applied to reduce hum noise (50 or 60 Hz) using a 4 Hz band (Fig. 2b). Instant envelope of each filtered channel was calculated using absolute value of Hilbert transform (Cieslak-Blinowska et al. 2011; Durka 2004). Spikes induce an increase in energy, which manifests in the 10–60 Hz frequency band as a peak in the envelope (Fig. 2c). The algorithm estimates the statistical distribution of the envelope and identifies a threshold value, which enables discrimination of spikes from background activity. The signal envelope was analysed using a moving window with a segment width of 5 s and 80 % overlap between consecutive segments. The statistical distribution of the envelope was calculated for each segment and approximated with best-fitting statistical model using a maximal likelihood algorithm (MLE) (Myung 2003). The ideal model should be able to describe the distribution of both segments with background activity and segments with spike. Twenty different models were meticulously evaluated, for example log-normal, Rayleigh, alpha-stable, Poisson, extreme value etc. Log-normal provided the most satisfactory statistical model, characterized by small residuals, low computational demands and advantage of being continuously differentiable. It is characterized by following equation:

$$y = f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}; x > 0 \quad (1)$$

where y is probability density, x is envelope amplitude, μ and σ are conjugated parameters of Gaussian distribution: $\mu = \frac{1}{N}\sum_{i=1}^{N} \ln(x_i)$ and $\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} [\ln(x_i) - \mu]^2}$, where N is length of envelope segment.

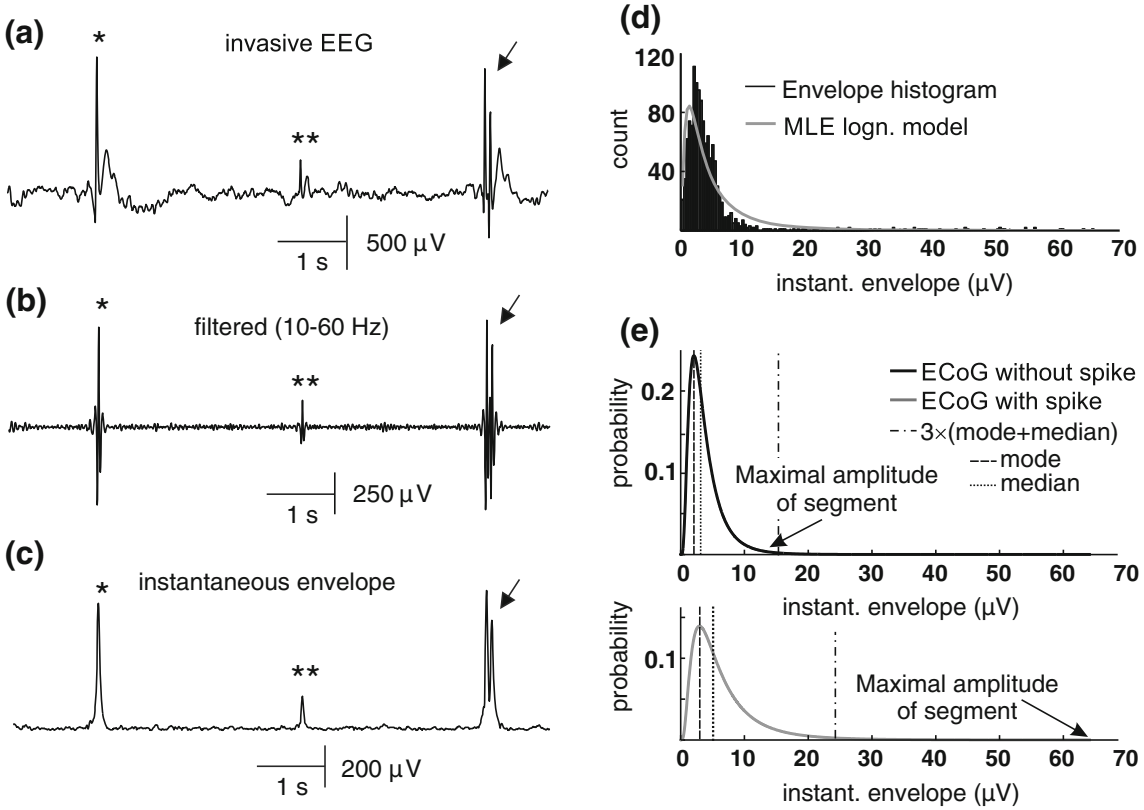

Fig. 2 Envelope distribution modelling. **a** Raw data (negativity up) containing obvious spike-and-wave discharge (*single asterisk*), low-amplitude spike-and-wave discharge (*double asterisks*) and polyspike (*arrow*). **b** Band-pass (10–60 Hz) filtered signal. **c** Envelope of the filtered signal. **d** Histogram of the envelope distribution of signal segment containing interictal spike and corresponding log-normal model fitted to histogram using minimal likelihood estimate technique (MLE). **e** Examples of log-normal envelope distribution models of the segment containing background activity and spike. Presence of spike in the signal segment is associated with a biased (positive skew) log-normal distribution model characterized by larger variance and long right tail. Mode, median and threshold value for $k_1 = 3$ are plotted for each distribution model

The presence of spikes in the signal is associated with a biased (positive skew) log-normal envelope distribution characterized by larger variance and long right tail (Fig. 2d, e). In contrast, a spike-free signal has a distribution less biased (Fig. 2e). Distinct log-normal distribution properties enable discrimination of spikes from background activity. In this study we implemented this feature as the main criterion for spike detection. We have empirically determined that a combination of mode and median of the log-normal distribution provides optimal parameters that discriminate spike envelope distribution from the background activity (Fig. 2e). Therefore, mode and median of the log-normal distribution was used to define a threshold that discriminates segments with spikes from the segments with background activity (Fig. 2e). The threshold value *th* was calculated for each segment using the following formula, where μ and σ were estimated using MLE:

$$th = k_1 \cdot [Mode + Median] \quad (2)$$

$$Mode = e^{\mu - \sigma^2} \quad (3)$$

$$Median = e^{\mu} \quad (4)$$

Coefficient value k_1 was optimized using gold standard spikes (see "Results" section). Variance, probability density values and several quantile measures were also tested to determine the threshold, but detection results were not satisfactory due to high number of FP (i.e. low selectivity).

Threshold values in each segment were interpolated using a cubic spline algorithm and a threshold curve was created (Fig. 1). Spikes with very high amplitudes induced large deviation of the threshold curve. To compensate for these deviations the threshold curve was smoothed with a moving average filter with order corresponding to segmentation window time. Each channel had its own envelope and its own threshold curve. Local maxima at intersections between envelope and threshold curves marked detected spike (Fig. 1). Detected events with separations less than 120 ms were merged together as single events.

The presence of beta and mu rhythms can lead to FP detections, particularly in recordings obtained from the motor cortex (see results section). Episodes of these rhythms can be excluded from the detection by incorporating the secondary processing steps which identify their distinct spectral features. Unfiltered signals were segmented into 20 s sections with 50 % overlap. A linear prediction filter (12-order; Sanchez et al. 2006) was used to calculate the power spectrum of each segment. If a local spectral maximum exists in the 10–25 Hz band, the signal contains beta or mu rhythm and detected events within this segment are excluded.

A Matlab file with the detector is available for free download on the Intracranial Signal Analysis Research Group Prague web page (http://isarg.fel.cvut.cz) and the International Epilepsy Electrophysiology Portal (http://www.ieeg.org). Annotated data used in this study can be found at the same locations.

Evaluation Methodology

The position of spikes in the gold standard datasets had to be corrected because readers did not mark the exact times of labelled spikes. For each labelled spike, the maximal value of the envelope was determined as an exact time index of the spike. To evaluate detector performance we used standard metrics (Casson et al. 2009) and defined the following parameters: true positive (TP)—correct detection of a spike labelled by reader; FP—incorrect detection, when a detected spike does not have a corresponding event labelled by human reader; false negative (FN)—missed detection when a spike was labelled by reader but missed by the detector. Sensitivity (SEN) represents percentage of detected spikes and selectivity (positive predictive value; PPV) is percentage of correct detections. They are calculated according the following equations.

$$SEN = \frac{TP}{TP + FN} \cdot 100 \quad (5)$$

$$PPV = \frac{TP}{TP + FP} \cdot 100 \quad (6)$$

Due to high variability in spike counts in each dataset we also determined total-sensitivity (*tSEN*) which incorporates averaging by a number of datasets, *M* (Casson et al. 2009):

$$tSEN = \frac{100}{\sum_{i=1}^{M} (TP_i + FN_i)} \sum_{i=1}^{M} TP_i \quad (7)$$

Detector performance was examined using receiver-operating characteristic (ROC) curves (Casson et al. 2009). To find the best detector settings (threshold coefficient k_1, filter bandwidth boundaries and segmentation window size), an optimal operating point on the ROC curve was determined by fitting the ROC curve with two first order polynomials. The optimal operating point was characterized by the lowest value of the sum of squares of the residuals. The performance of the detector was tested against a published and freely available algorithm, which detects spikes strictly in the time domain according to the shape of the spike waveform (Barkmeier et al. 2012).

Evaluation of False Detection

To investigate sources of FP detections we used two strategies. First, we used principal component analysis to decompose data to non-correlated components (Jung et al. 1998). Signal sections with detected events (30 ms before

and 330 ms after time index) were extracted to find the decomposition transform. The first five components represented >80 % of the information and their waveforms were visually examined by readers. Secondly, we applied the detector to non-epileptic data obtained from two patients with intractable facial pain who underwent implantation of subdural electrodes over the sensorimotor cortex for experimental pain treatment (Stead et al. 2010). Each dataset contained 1 h recording of spontaneous brain activity. An average reference was calculated to reject common mode signals. Channels which contained large numbers of artefacts were manually removed. In total, 25 out of 33 channels were analysed. All results in this study are expressed as mean ± SD (median).

Results

Characteristics of Human Labelling

Human readers labelled 6,518 spikes in total, 53 ± 21 % were classified as obvious spikes and 47 ± 21 % as ambiguous. The proportions varied between readers (Fig. 3a, b). Two-reader agreement was found in 50 % of obvious spikes and three-reader agreement was observed in 25 % of obvious spikes only (Fig. 4b, c). Consensus between two readers was found in 25 % of ambiguous spikes and it dramatically dropped to only 5 % of spikes with consensus of all three readers. Agreement between pairs of readers (for both ambiguous and obvious spikes) was quantified by Cohen's kappa coefficient (Zijlmans et al. 2002) which revealed low values of agreement for all three reader pairs (kappa 0.01, 0.23 and 0.17; Fig. 4a). When only obvious spikes were considered kappa values were 0.29, 0.81 and 0.26. For evaluation of the detector, gold standard datasets were determined from all spikes in which at least two readers were in agreement (n = 1,956).

Gold standard obvious spikes had amplitude 378 ± 389 μV (median 276 μV) and spike width at level of 20 % of amplitude was 44 ± 22 ms (median 37 ms). Gold standard ambiguous spikes had amplitude 196 ± 207 μV (median 190 μV) and duration 52 ± 23 ms (median 47 ms). The signal-to-noise ratio (SNR) was determined as the ratio between power of the spike and power of the baseline activity. Spike power was calculated as squared median of maximal amplitude of the labelled spike, while baseline power was estimated from the 0.3 s signal section before onset of spike. SNR was determined separately for raw signals and for 10–60 Hz band pass filtered signals. SNR of obvious spikes was 22.4 ± 29.5 dB (median 15.3 dB) in raw signals and 32.5 ± 38.3 dB (median 22.9 dB) in filtered signals. Ambiguous spikes had SNR 18.4 ± 21.5 dB (median 13.5 dB) in raw and 24.8 ± 26.6 dB (median 22.7 dB) in filtered data (Table 3).

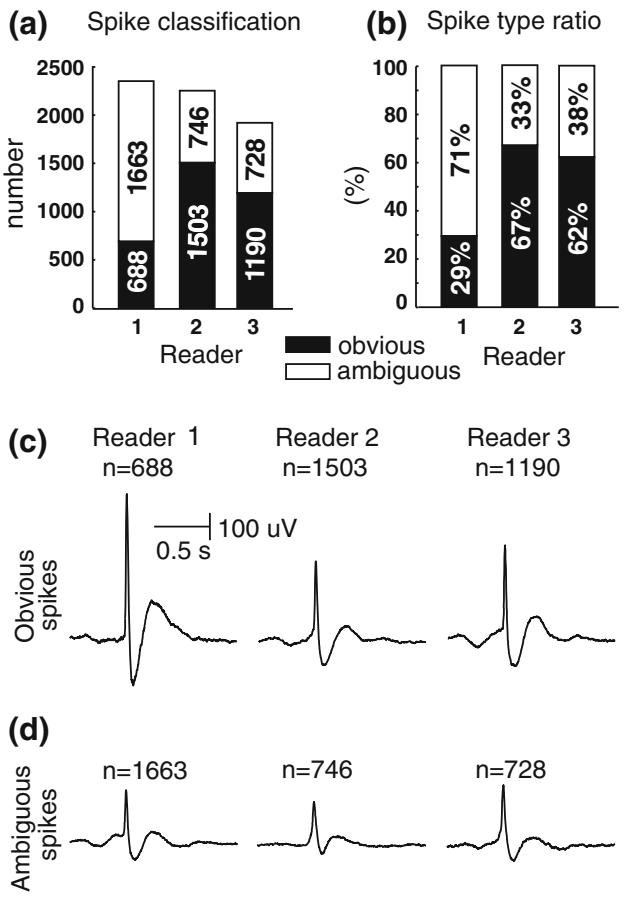

Fig. 3 Reader spike labelling. **a** Number of spikes classified by each reader as obvious or ambiguous spikes. **b** Proportions of individual spike subclasses marked by each reader. **c** Waveform average of obvious spikes labelled by each reader. **d** Average of ambiguous spikes

Detector Optimization and Spike Detection

Performance of the detector depended on the following settings: threshold coefficient k_1, filter bandwidth boundaries and segmentation window size. Impact of frequency boundaries was tested between 8 and 20 Hz for the high pass filter and 30–80 Hz for the low pass filter. The best performance was achieved with band-pass filtering at 10–60 Hz. Optimal window size of signal segmentation was 5 s with 80 % overlap. Both parameters had low impact on the performance of the detector, in contrast to the threshold coefficient k_1. Optimal values of k_1 were determined by a cross-validation procedure (Golub et al. 1979; Kohavi 1995). The first step involved random selection of two thirds of channels (70 out of 105) to generate the ROC curve and estimate its error function. The optimal k_1 value was found as a minimum of the error

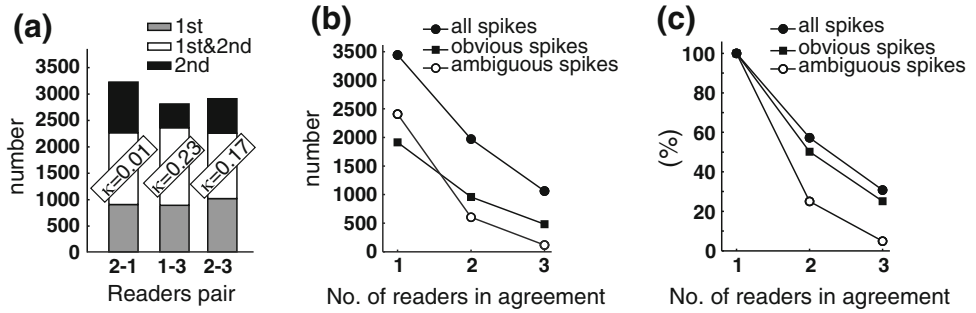

Fig. 4 Inter-reader agreement. **a** Agreement between pairs of readers: intersection and Cohen's kappa κ. **b** Number of spikes in agreement between two and three readers for obvious, ambiguous and all spikes. **c** Proportions of spikes as a function of agreement between numbers of readers. Agreement between three readers is particularly low for ambiguous spikes

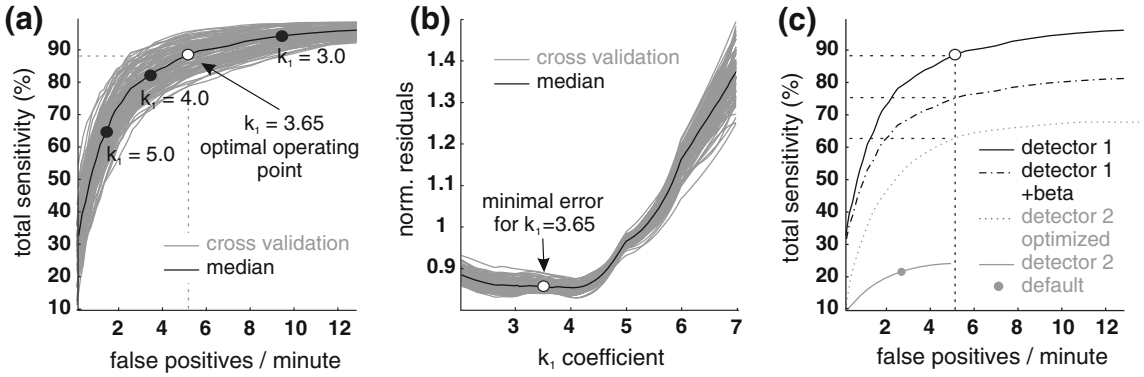

Fig. 5 Detector performance. **a** ROC curves obtained during cross-validation procedure (*grey lines*). Median ROC curve (*black line*). ROC curves were calculated in 0.1 steps for values of coefficient k_1 ranging from 1.8 to 10. Examples of operating points for three selected values of coefficient k_1 ($k_1 = 5.0$, $k_1 = 4.0$, $k_1 = 3.0$) are shown by the *black circles*. **b** Error function calculating the sum of residuals for each ROC curve obtained during the cross-validation (*grey lines*). Median of error function (*black line*). The minimal median value of the residuals is at with $k_1 = 3.65$. This value was set as optimal operating point of the detector. **c** Comparison of the envelope distribution modelling algorithm (detector 1) with algorithm that uses spike shape for detection (detector 2). ROC curves of both algorithms. Performance of a freely available detector (detector 2) is shown with default (*grey circle*) and with optimized settings (*grey dotted line*). At the same level of false positive rate, our detector had higher total sensitivity than the optimized detector 2. Incorporation of a post-processing step, that removes beta oscillations, decreases *tSEN* by 12 % in optimal working point (detector 1 + beta, *dash dot line*)

function (Fig. 5b). The performance of the detector (tSEN and FP rate) was then evaluated on the remaining one third of channels (35 out of 105)—ROC curve of testing dataset (Fig. 5a). The process was repeated one hundred times. This cross-validation procedure was performed for values of coefficient k_1 ranging from 1.8 to 10 (0.1 steps). Optimal performance was achieved for $k_1 = 3.5 \pm 0.6$ (median 3.65) when tSEN was 88.9 ± 4.4 % (median 89.3 %) and rate of FP detections was 5.2 ± 0.5 (median 5.2) min^{-1}. Median of k_1 was selected as the optimal value and the optimised detector was then applied to the entire dataset (Fig. 5c). Grouped results showed average SEN 91.4 ± 12.2 % (median 94.8 %; Table 2) with mean FP rate 8.2 ± 7.4 (median 7.5) min^{-1}. At this threshold, shape of the averaged detected event highly correlated with the averaged gold standard spike waveform (Pearson correlation 98 %, $p < 0.001$). The amplitude of detected events was 217 ± 312 μV (median 164 μV), duration of their averaged waveform was 35 ms and SNR of detected events in raw signals was 19.9 ± 27.7 dB (median 13.1 dB) and 29.2 ± 36.3 dB (median 21.4 dB) for filtered signal (Table 3).

We compared the performance of our detector with a published detector (Barkmeier et al. 2012). Initially we used its recommended settings but later modified them to maximize its performance when applied on our datasets (Fig. 5c). The most important changes in settings involved decrease of the multiple of standard deviation, change of values for minimal half-wave amplitudes (1–800 μV), and duration and slope of the half-waves. With default settings tSEN was 21.3 % with FP rate 2.9 min^{-1}. At this FP rate tSEN of our algorithm was 80.0 %. After optimization of published detector its maximal tSEN was 62.1 % with 5.3 FP min^{-1}. Meanwhile, *tSEN* of our detector was 88.9 %. In

Table 2 Results of detection per each dataset—gold standard spikes (GS), false positive (FP) and false negative (FN) detections

Dataset	No. of GS	Sensitivity	Selectivity	FP (min^{-1})	No. of FP	No. of FN
Patient 1	613	94.8	63.2	7.5	339	32
Patient 2	453	65.6	59.8	4.4	200	156
Patient 3	21	100.0	5.5	8.0	361	0
Patient 4	3	100.0	13.6	0.4	19	0
Patient 5	319	98.8	22.5	24.1	1085	4
Patient 6	335	92.8	47.9	7.5	338	24
Patient 7	212	87.7	41.9	5.7	258	26
Mean ± SD	279 ± 221	97.4 ± 12.2	36.3 ± 22.7	8.2 ± 7.4	371 ± 336	35 ± 55
Median	319	94.8	41.9	7.5	338	24

Table 3 Properties of gold standard spikes (GS) and other events—true positive (TP), false positive (FP) and false negative (FN)

	GS		TP + FP + FN	FP	FN
	Obvious	Ambiguous			
n	1,496	460	4,314	2,600	242
Amplitude (μV)	378 ± 389 (276)	196 ± 207 (190)	217 ± 312 (164)	118 ± 212 (107)	154 ± 178 (141)
Duration t20 % (ms)	44 ± 22 (37)	52 ± 23 (47)	35*	38*	84 ± 81 (77)
SNR (dB)	22.4 ± 29.5 (15.3)	18.4 ± 21.5 (13.5)	19.9 ± 27.7 (13.1)	16.0 ± 21.5 (11.2)	15.5 ± 21.0 (11.1)
$SNR_{10–60\ Hz}$ (dB)	32.5 ± 38.3 (22.9)	24.8 ± 26.6 (22.7)	29.2 ± 36.3 (21.4)	24.7 ± 32.0 (20.3)	17.8 ± 17.7 (16.6)
Correlation with GS	NA	NA	0.98	0.93	0.83
			$p < 0.001$	$p < 0.001$	$p < 0.001$

Median values are in brackets. Star marks cases where mean or median values could not be calculated due to low signal-to-noise ratio (SNR). Values were obtained from the averaged waveform

both cases envelope modelling approach demonstrated higher reliability to detect spikes with 3.8 and 1.4 times higher total SEN respectively.

Source of False Detections

We examined the source and properties of falsely detected events in detail. We first examined properties of false detections from recordings in patients with epilepsy and then applied the detector to non-epileptic invasive recordings.

The total number of FP detections was 2,600 with a mean of 371 ± 336 (median 338) FP detections (Table 2) per dataset. To determine whether false detections included missed spikes, three hundred randomly selected FP detections were presented independently to four neurophysiologists (Readers 1–3 and one additional). They classified as spikes 18 ± 8 % of FP detections. To decompose the sources of FP detections from epileptic recordings we utilized principal component analysis which extracted the five most common components from detected events (Fig. 6b, c). These components were then visually analysed by neurophysiologists to determine the potential sources of false detection according to the waveform morphology of extracted components (Kirsch et al. 2006). Neurophysiologists agreed that waveforms of three components resembled spikes and thus based on principal component analysis 52 % of FP detections could be classified as spikes. We examined the relationship between the number of gold standard spikes and the number of detected events in each channel using regression analysis. Results showed a linear relationship between the number of gold standard (labelled) spikes and detected events ($r^2 = 0.55$, Fig. 6d) with a significant positive Spearman correlation (76 %, $p < 0.001$). The mean ratio between the number of detected events and labelled spikes was 6.6 ± 10.3 (median 2.6) − zero (no detected events) and infinite ratio (no spikes labelled by readers) channels were excluded. This suggests that channels with higher numbers of expert labelled spikes had also higher numbers of detected events.

Undetected gold standard spikes FN were characterized as waveforms with amplitude 154 ± 178 μV (median 141 μV), duration 84 ± 81 ms (median 77 ms) and SNR is 15.5 ± 21.0 dB (median 11.1 dB). Waveform features and lower correlation with spikes suggest that FN detection include mainly discharges which share features of sharp waves (Fig. 6a).

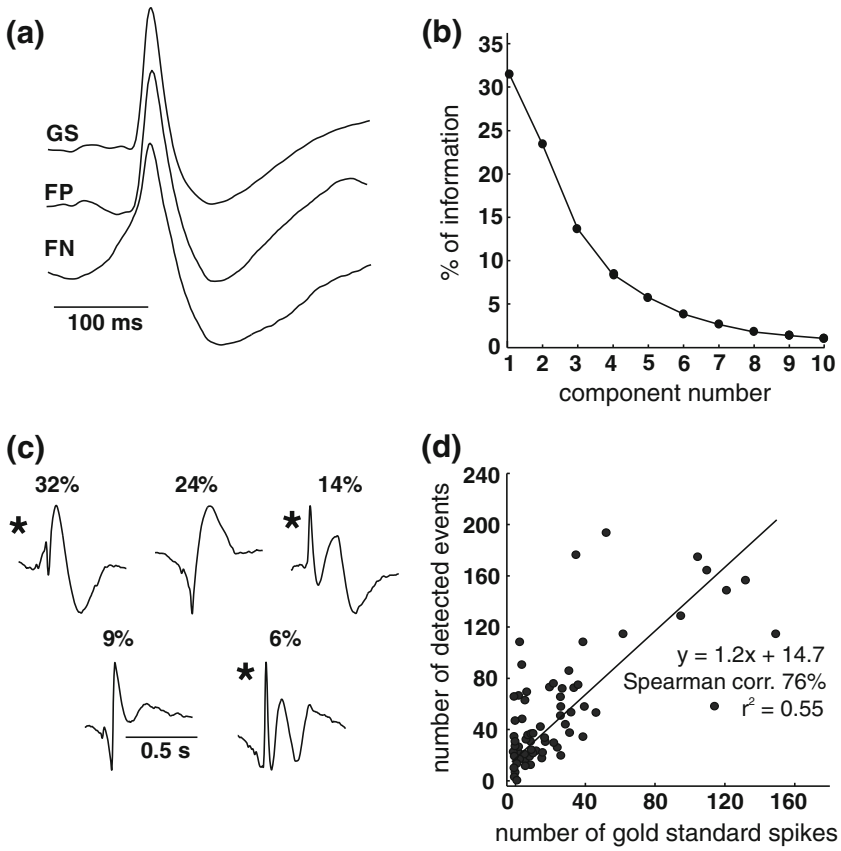

Fig. 6 Sources of false detections. **a** Normalized average waveform of gold standard (GS), false positive (FP) and false negative spikes (FN). Note the wider shape of false negative detections. **b** Principal component analysis of false positive detections. First five components represent 85 % of all FP detections. **c** Five major components waveforms. Components labelled with *star mark* waveforms on which readers agreed to resemble spike. **d** Relationship between number of gold standard spikes and number of detected events in each channel demonstrate significant positive correlation. Linear trend is estimated by equation $y = 1.2x + 14.7$ ($r^2 = 0.55$), where x is number of gold standard spikes and y is number of detections

When the detector was applied to the non-epileptic data it detected 2.4 ± 2.4 (median 1.5) events min^{-1}. High-rate detections were present particularly in channels with prominent brain rhythms overlapping in frequency with spikes—mainly beta and mu rhythms (Fig. 7a, b). Episodes of these rhythms can be excluded from detection by incorporating secondary processing steps (see Methods). The post-processing step decreased the number of detections in non-epileptic data by 27 % and decreased tSEN of the detector by 12 % when applied to epileptic data. After removing rhythmic activities, isolated non-epileptic sharp transients (Fig. 7c, d) represented the majority of false detections. We examined their morphological and spatial properties. Average waveforms demonstrate that these transients are characterized by the lack of a subsequent slow-wave typically found in IEDs (Fig. 7e). Next we examined whether non-epileptic sharp transients can be discriminated from IEDs by their spatial profile (pattern of propagation). For each detected event we obtained signal envelope values in all electrodes and normalized them by envelope amplitude of the detected event. The resulting histogram demonstrated that spikes tend to have significantly larger spatial distribution than non-epileptic sharp transients ($p < 0.001$, Fig. 7f).

Discussion

The detector developed in this study belongs to class of detectors that distinguish spikes from the background activity according to signal envelope modelling. The time domain in which the adaptive envelope modelling algorithm operates allows more precise spike time indexing. Decision threshold for spike detection is continuously and automatically adjusted according the statistical parameters of background activity surrounding the spike. The algorithm is resistant to non-stationarity of brain signals (Wong et al. 2006) and adaptive estimation of background minimizes effect of long term interference (gain fluctuations due changes in electrode impedance, system noise, etc.) on number of false detections. Alternative methods of spike detection based on characterization of background have been published. These methods model background activity by adaptive prediction, for example Kalman filtering (Oikonomou et al. 2007) or autoregressive moving average modelling (known as ARMA) (Schack and Grieszbach 1994). However, these methods require a priori knowledge about the studied system and settings (Kalman gain, covariance matrix, model order etc.) must be adjusted accordingly, which results in lower robustness.

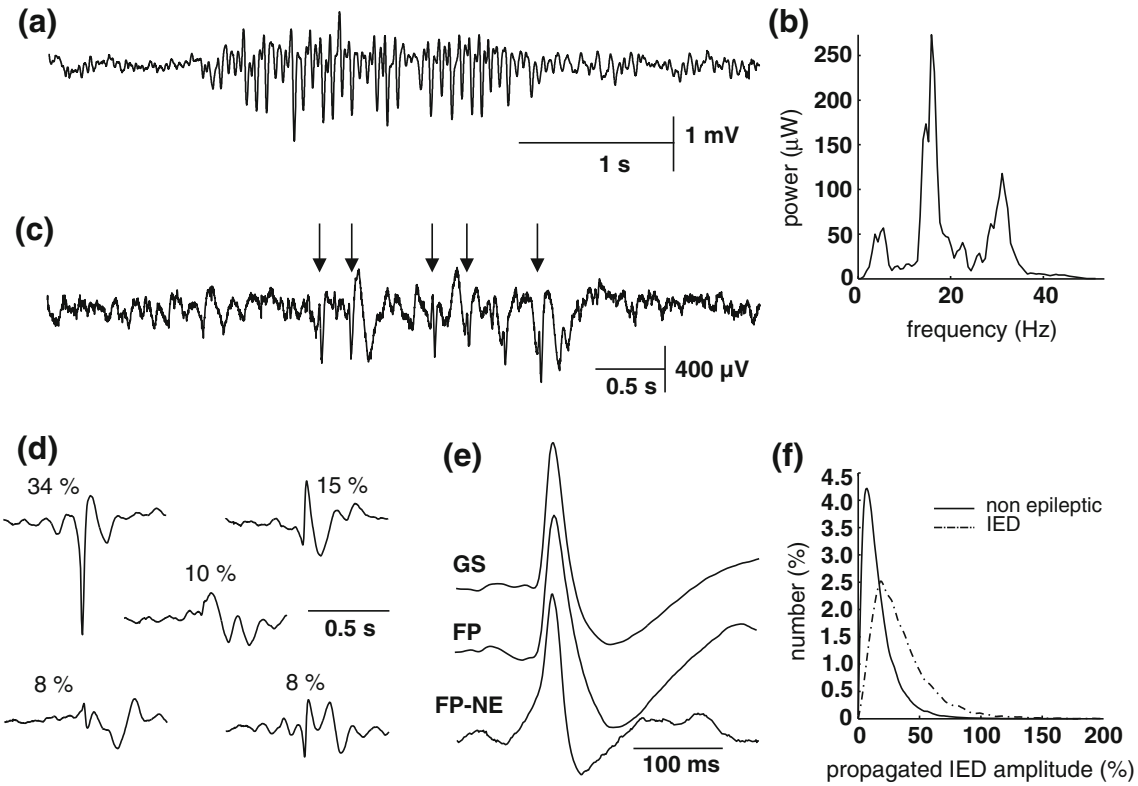

Fig. 7 Detections in non-epileptic data. **a** Example of 16 Hz oscillations and corresponding power spectrum **b**. The presence of obvious peaks in power spectrum enables design of post-processing steps to eliminate these oscillations. Note the sharp shape of each cycle. **c** Isolated non-epileptic sharp transients. *Arrows* mark detected events. **d** Five major components of events detected in non-epileptic datasets. **e** Comparison of averaged waveform of gold standard spikes (GS), false positive detections (FP) and detections from non-epileptic datasets (FP-NE) after beta rhythms were excluded. Average of non-epileptic detections is characterized by a less prominent following wave. **f** Spatial profile of non-epileptic sharp transients and of interictal epileptiform discharges. Histograms show normalized envelope values evoked by detected events in adjacent channels. Non-epileptic transients are rather local phenomena with limited spatial propagation in contrast to interictal epileptiform discharges

A major drawback of automatic detectors is low selectivity PPV and high rate of FP detections (Wilson and Emerson 2002). To propel development of automatic detectors requires understanding of the sources of false detections including the biological ones. We showed that in artefact-free recordings FP detections contain substantial numbers of spikes, which were not identified by human reviewers, probably due to low amplitude or low SNR. These low-amplitude spikes may represent an important additional source of information. During the presurgical examination location of the irritative zone is derived mainly from the spike rate. We showed positive correlation between gold standard spikes and detected events suggesting that channels which contained high numbers of gold standard spikes also contained high number of detections. It seems that the detector can achieve similar results to human reviewers if spike rate is considered. Unfortunately, the results also included channels in which readers did not label any spikes but detector did. To determine the exact clinical significance of falsely detected low-amplitude spikes will require more detailed studies in future. However, our preliminary results suggest that they may provide important information about propagation, spatial distribution profile of IEDs and identification of their site of origin (Janca et al. 2013). Identification of biological sources of false detections required artefact-free recordings. This algorithm is not resistant to technical artefacts. If the envelope properties of a technical artefact closely mimic envelope properties of spikes it will be detected. However, implementation of post-processing steps can eliminate this kind of false-detections. Yadav et al. (2011) demonstrated that waveform clustering of detected events followed by visual inspection of waveform clusters can be used to identify distinct spikes waveforms. This approach can be also applied to eliminate artefacts. In addition, clustering according to the spatial profile of detected events can be used to separate random artefacts from spikes (Janca et al. 2013).

Other potential sources of FP detections include non-epileptiform sharp transients and/or cerebral rhythms

which have shape and frequency content similar to low amplitude spikes (Gotman and Gloor 1976; Rakhade et al. 2007). The brain generates physiological oscillations at frequencies similar to those in spikes (beta oscillations, mu rhythm) which can be pronounced in certain regions. Their oscillatory patterns and distinct spectral profile, however, enable separation of these events and thus secondary processing steps can be integrated into the algorithm to exclude them. Isolated sharp non-epileptiform transients represent a more difficult task, because we do not know what is their biological or clinical significance and what are the cellular mechanisms generating them. In scalp electroencephalogram several benign epileptiform variants (transients) were described, e.g. benign epileptiform transients of sleep (White et al. 1977), wicket spikes (Reiher and Lebel 1977), "14-" and "6-" Hz positive spikes and others (Santoshkumar et al. 2009). These electrographic phenomena are not considered pathological and often occur during drowsiness or sleep. Some studies show high prevalence of these electrographic phenomena amongst healthy subjects while others demonstrate low prevalence. If these phenomena are recorded in scalp EEG and have biological sources, it is plausible that they will be also recorded in invasive EEG. Unfortunately, studies focused on human physiological brain activity recorded invasively are extremely rare or nearly absent (Niedermeyer and Lopes da Silva 2004). Experimentally one of the best described physiological sharp transients are so called hippocampal sharp waves, which can occur isolated or with superimposed ripple activity, and are involved in memory processes. Therefore, the presence of benign epileptiform or physiological sharp transients may complicate the process of automatic spike detection. This problem is not unique for spikes but also complicates studies focused on automatic detection of pathological high-frequency oscillations, which demonstrate frequency overlap between physiological and pathological forms, as well as regionally specific frequency profiles (Jacobs et al. 2012). To be able to discriminate non-epileptic phenomena requires understanding their functional significance, regional distribution and identification of the specific discriminating features. At the moment, our results suggest that one of the features that could differentiate benign epileptiform variants from spikes is the absence of a prominent following slow wave and the local nature of non-epileptic sharp transients.

These observations provide important information for any study focused on development of spike detectors. Secondary processing steps that reject physiological sharp transients and sharp rhythms should represent an integral component of future spike detecting algorithms to maximally suppress false detections of neuronal origin. Nevertheless, to minimize false detection in general, algorithms should obey several conditions. In the majority of cases, the application of IIR filters is preferable. To avoid reference electrode contamination by spikes or interference of electrocardiographic activity utilization of bipolar recording can be considered. Feasible alternatives to unsupervised detectors are methods of supervised (semiautomatic) detection methods. These may represent compromise between unsupervised approach and manual review. Although not being fully automatic they would still substantially decrease human labour of reviewing multichannel data. Major components or clusters of detected events in each channel could be plotted for subsequent visual analysis. Non-epileptic components would be identified by human reader and manually excluded from the final quantification of detected events (Yuan et al. 2012).

Acknowledgments This work has been supported by the Ministry of Health of the Czech Republic Grant IGA MZ CR NT/11460-4.

References

Asano E, Muzik O, Shah A, Juhász C, Chugani DC, Sood S, Janisse J, Ergun EL, Ahn-Ewing J, Shen C, Gotman J, Chugani HT (2003) Quantitative interictal subdural EEG analyses in children with neocortical epilepsy. Epilepsia 44:425–434. doi:10.1046/j.1528-1157.2003.38902.x

Barkmeier DT, Shah AK, Flanagan D, Atkinson MD, Agarwal R, Fuerst DR, Jafari-Khouzani K, Loeb JA (2012) High inter-reviewer variability of spike detection on intracranial EEG addressed by an automated multi-channel algorithm. Clin Neurophysiol 123:1088–1095. doi:10.1016/j.clinph.2011.09.023

Bautista RE, Cobbs MA, Spencer DD, Spencer SS (1999) Prediction of surgical outcome by interictal epileptiform abnormalities during intracranial EEG monitoring in patients with extrahippocampal seizures. Epilepsia 40:880–890. doi:10.1111/j.1528-1157.1999.tb00794.x

Brinkmann BH, Bower MR, Stengel KA, Worrell GA, Stead M (2009) Large-scale electrophysiology: acquisition, compression, encryption, and storage of big data. J Neurosci Methods 180:185–192. doi:10.1016/j.jneumeth.2009.03.022

Casson AJ, Luna E, Rodriguez-Villegas E (2009) Performance metrics for the accurate characterisation of interictal spike detection algorithms. J Neurosci Methods 177:479–487. doi:10.1016/j.jneumeth.2008.10.010

Cieslak-Blinowska K, Zygierewicz J, Blinowska KJ (2011) Practical biomedical signal analysis using MATLAB. CRC Press, Boca Raton

de Curtis M, Avanzini G (2001) Interictal spikes in focal epileptogenesis. Prog Neurobiol 63:541–567. doi:10.1016/S0301-0082(00)00026-5

Dumpelmann M, Elger CE (1999) Visual and automatic investigation of epileptiform spikes in intracranial EEG recordings. Epilepsia 40:275–285. doi:10.1111/j.1528-1157.1999.tb00704.x

Durka PJ (2004) Adaptive time-frequency parametrization of epileptic spikes. Phys Rev E 69:051914. doi:10.1103/PhysRevE.69.051914

Frost JD (1985) Automatic recognition and characterization of epileptiform discharges in the human EEG. J Clin Neurophysiol 2:231–249

Gloor P (1975) Contributions of electroencephalography and electrocorticography to the neurosurgical treatment of the epilepsies. Adv Neurol 8:59–105

Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21:215–223. doi:10.1080/00401706.1979.10489751

Gotman J (1986) Computer analysis of the EEG in epilepsy. In: Lopes de Silva FH, Storm van Leeuwen W, Remond A (eds) Clinical applications of computer analysis of EEG and other neurophysiological signals. Elsevier, Amsterdam, pp 171–204

Gotman J (1999) Automatic detection of seizures and spikes. J Clin Neurophysiol 16:130–140

Gotman J, Gloor P (1976) Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. Electroencephalogr Clin Neurol 41:513–529. doi:10.1016/0013-4694(76)90063-8

Halford JJ (2009) Computerized epileptiform transient detection in the scalp electroencephalogram: obstacles to progress and the example of computerized ECG interpretation. Clin Neurophysiol 120:1909–1915. doi:10.1016/j.clinph.2009.08.007

Jacobs J, Staba R, Asano E, Otsubo H, Wu JY, Zijlmans M, Mohamed I, Kahane P, Dubeau F, Navarro V, Gotman J (2012) High-frequency oscillations (HFOs) in clinical epilepsy. Prog Neurobiol 98:302–315. doi:10.1016/j.pneurobio.2012.03.001

Janca R, Jezdik P, Cmejla R, Krsek P, Jefferys JGR, Marusic P, Jiruska P (2013) Automatic detection and spatial clustering of interictal discharges in invasive recordings. In: Medical measurements and applications proceedings (MeMeA), pp 219–223. doi:10.1109/MeMeA.2013.6549739

Jung TP, Humphries C, Lee TW, Makeig S, McKeown MJ, Iragui V, Sejnowski TJ (1998) Removing electroencephalographic artifacts: comparison between ICA and PCA. Neural Netw Signal Process VIII:63–72. doi:10.1109/NNSP.1998.710633

Kirsch HE, Robinson SE, Mantle M, Nagarajan S (2006) Automated localization of magnetoencephalographic interictal spikes by adaptive spatial filtering. Clin Neurophysiol 117:2264–2271. doi:10.1016/j.clinph.2006.06.708

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Int Jt Conf Artif Intell 14:1137–1145

Lima MC, Fregni F (2008) Motor cortex stimulation for chronic pain systematic review and meta-analysis of the literature. Neurology 70:2329–2337. doi:10.1212/01.wnl.0000314649.38527.93

Lodder SS, Askamp J, van Putten MJ (2013) Inter-ictal spike detection using a database of smart templates. Clin Neurophysiol 124:2328–2335. doi:10.1016/j.clinph.2013.05.019

Marsh ED, Peltzer B, Brown MW III, Wusthoff C, Storm PB Jr, Litt B, Porter BE (2009) Interictal EEG spikes identify the region of electrographic seizure onset in some, but not all, pediatric epilepsy patients. Epilepsia 51:592–601. doi:10.1111/j.1528-1167.2009.02306.x

Myung IJ (2003) Tutorial on maximum likelihood estimation. J Math Psychol 47:90–100. doi:10.1016/S0022-2496(02)00028-7

Niedermeyer E, Lopes da Silva F (2004) Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins, Philadelphia

Oikonomou VP, Tzallas AT, Fotiadis DI (2007) A Kalman filter based methodology for EEG spike enhancement. Comput Methods Programs Biomed 85:101–108. doi:10.1016/j.cmpb.2006.10.003

Rakhade SN, Shah AK, Agarwal R, Yao B, Asano E, Loeb JA (2007) Activity-dependent gene expression correlates with interictal spiking in human neocortical epilepsy. Epilepsia 48:86–95. doi:10.1111/j.1528-1167.2007.01294.x

Reiher J, Lebel M (1977) Wicket spikes: clinical correlates of a previously undescribed EEG pattern. Can J Neurol Sci 4:39–47

Rosenow F, Lüders H (2001) Presurgical evaluation of epilepsy. Brain 124:1683–1700. doi:10.1093/brain/124.9.1683

Sanchez JC, Mareci TH, Norman WM, Principe JC, Ditto WL, Carney PR (2006) Evolving into epilepsy: multiscale electrophysiological analysis and imaging in an animal model. Exp Neurol 198:31–47. doi:10.1016/j.expneurol.2005.10.031

Santoshkumar B, Chong JJ, Blume WT, McLachlan RS, Young GB, Diosy DC, Burneo JG, Mirsattari SM (2009) Prevalence of benign epileptiform variants. Clin Neurophysiol 120:856–861. doi:10.1016/j.clinph.2009.03.005

Schack B, Grieszbach G (1994) Adaptive methods of trend detection and their application in analysing biosignals. Biom J 36:429–452. doi:10.1002/bimj.4710360404

Stead M, Bower M, Brinkmann BH, Lee K, Marsh WR, Meyer FB, Litt B, Van GJ, Worrell GA (2010) Microseizures and the spatiotemporal scales of human partial epilepsy. Brain 133:2789–2797. doi:10.1093/brain/awq190

Van Gompel JJ, Stead SM, Giannini C, Meyer FB, Marsh WR, Fountain T, So E, Cohen-Gadol A, Lee KH, Worrell GA (2008) Phase I trial: safety and feasibility of intracranial electroencephalography using hybrid subdural electrodes containing macro- and microelectrode arrays. Neurosurg Focus 25:E23. doi:10.3171/FOC/2008/25/9/E23

Webber WR, Litt B, Lesser RP, Fisher RS, Bankman I (1993) Automatic EEG spike detection: what should the computer imitate? Electroencephalogr Clin Neurophysiol 87:364–373. doi:10.1016/0013-4694(93)90149-P

White JC, Langston JW, Pedley TA (1977) Benign epileptiform transients of sleep. Clarification of the small sharp spike controversy. Neurology 27:1061

Wilke C, Drongelen WV, Kohrman M, He B (2009) Identification of epileptogenic foci from causal analysis of ECoG interictal spike activity. Clin Neurophysiol 120:1449–1456. doi:10.1016/j.clinph.2009.04.024

Wilson SB, Emerson R (2002) Spike detection: a review and comparison of algorithms. Clin Neurophysiol 113:1873–1881. doi:10.1016/S1388-2457(02)00297-3

Wilson SB, Harner RN, Duffy FH, Tharp BR, Nuwer MR, Sperling MR (1996) Spike detection. I. Correlation and reliability of human experts. Electroencephalogr Clin Neurophysiol 98:186–198. doi:10.1016/0013-4694(95)00221-9

Wong KFK, Galka A, Yamashita O, Ozaki T (2006) Modelling non-stationary variance in EEG time series by state space GARCH model. Comput Biol Med 36:1327–1335. doi:10.1016/j.compbiomed.2005.10.001

Worrell GA, Jerbi K, Kobayashi K, Lina JM, Zelmann R, Le Van QM (2012) Recording and analysis techniques for high-frequency oscillations. Prog Neurobiol 98:265–278. doi:10.1016/j.pneurobio.2012.02.006

Yadav R, Shah AK, Loeb JA, Swamy MNS, Agarwal R (2011) A novel unsupervised spike sorting algorithm for intracranial EEG. In: Annual international conference of the IEEE Engineering in Medicine and Biology Society, pp 7545–7548. doi:10.1109/IEMBS.2011.6091860

Yuan Y, Yang C, Si J (2012) The M-Sorter: an automatic and robust spike detection and classification system. J Neurosci Methods 210:281–290. doi:10.1016/j.jneumeth.2012.07.012

Zijlmans M, Huiskamp GM, Leijten FS, Van Der Meij WM, Wieneke G, Van Huffelen AC (2002) Modality-specific spike identification in simultaneous magnetoencephalography/electroencephalography: a methodological approach. J Clin Neurophysiol 9:183–191