Clinical Neurophysiology 128 (2017) 1246–1254

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

A novel scheme for the validation of an automated classification method for epileptic spikes by comparison with multiple observers

Niraj K. Sharma [a,*], Carlos Pedreira [b], Maria Centeno [a], Umair J. Chaudhary [a], Tim Wehner [a], Lucas G.S. França [a], Tinonkorn Yadee [a], Teresa Murta [a], Marco Leite [a], Sjoerd B. Vos [a,c], Sebastien Ourselin [a,c,d], Beate Diehl [a], Louis Lemieux [a]

[a] Dept. of Clinical and Experimental Epilepsy, UCL Institute of Neurology, London, United Kingdom
[b] Dept. of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[c] Translational Imaging Group, Centre for Medical Image Computing, UCL, London, United Kingdom
[d] Dementia Research Centre, UCL Institute of Neurology, London, United Kingdom

ARTICLE INFO

Article history:
Accepted 19 April 2017
Available online 4 May 2017

Keywords:
Interictal spike classification
Intracranial EEG
Automated spike classification
Information theory

HIGHLIGHTS

- We created a validation method for the evaluation of automated classification of interictal spikes.
- We used a modified version of *Wave_clus* (WC) to automatically classify the data of 5 patients.
- WC classification was similar to EEG reviewers providing an unbiased evaluation of the clinical data.

ABSTRACT

Objective: To validate the application of an automated neuronal spike classification algorithm, *Wave_clus* (WC), on interictal epileptiform discharges (IED) obtained from human intracranial EEG (icEEG) data.
Method: Five 10-min segments of icEEG recorded in 5 patients were used. WC and three expert EEG reviewers independently classified one hundred IED events into IED classes or non-IEDs. First, we determined whether WC-human agreement variability falls within inter-reviewer agreement variability by calculating the variation of information for each classifier pair and quantifying the overlap between all WC-reviewer and all reviewer-reviewer pairs. Second, we compared WC and EEG reviewers' spike identification and individual spike class labels visually and quantitatively.
Results: The overlap between all WC-human pairs and all human pairs was >80% for 3/5 patients and >58% for the other 2 patients demonstrating WC falling within inter-human variation. The average sensitivity of spike marking for WC was 91% and >87% for all three EEG reviewers. Finally, there was a strong visual and quantitative similarity between WC and EEG reviewers.
Conclusions: WC performance is indistinguishable to that of EEG reviewers' suggesting it could be a valid clinical tool for the assessment of IEDs.
Significance: WC can be used to provide quantitative analysis of epileptic spikes.

1. Introduction

As part of standard practice for assessing patients with epilepsy, clinical neurophysiologists are able to detect interictal epileptiform discharges (IED or 'epileptic spikes') during interictal EEG recordings. Although there is no gold standard as to what constitutes an epileptic spike, they tend to comprise a high amplitude deflection event lasting approximately 40–100 ms (De Curtis and Avanzini, 2001). Some patients evaluated for resective surgical treatment for epilepsy are investigated with intracranial EEG (icEEG) usually when there is strong evidence of an epileptogenic focus but not sufficient information to define a surgically resectable area using non-invasive methods. These patients may be implanted with multiple electrodes targeting deep areas of the brain or placed on the cortex to record epileptic activity (Fernández and Loddenkemper, 2013). In these patients, evidence suggests that a good postsurgical outcome is associated with the removal of the region generating the most frequent epileptic spikes (Asano et al., 2003; Marsh et al., 2010). However, detection of

* Corresponding author.
E-mail address: niraj.sharma.14@ucl.ac.uk (N.K. Sharma).

http://dx.doi.org/10.1016/j.clinph.2017.04.016
1388-2457/

epileptic spikes on icEEG has shown a low level of agreement (<50%) for both the intra-rater (Brown et al., 2007) and the inter-rater comparisons between clinical neurophysiologists (Dümpelmann and Elger, 1999; Barkmeier et al., 2012; Gaspard et al., 2014). To reduce this subjectivity, computational algorithms designed for the automated detection of IEDs on icEEG have been implemented (Dümpelmann and Elger, 1999; Bourien et al., 2005; Valenti et al., 2006; Brown et al., 2007; Barkmeier et al., 2012; Gaspard et al., 2014). However, to our knowledge, the work on IED classification has been limited (Bourien et al., 2005; Yadav et al., 2011; Janca et al., 2013).

Classification of IEDs into various IED 'populations' generally relies on clinicians distinguishing between different IED types by assessing the EEG waveform which often takes into account the epileptic spike's field distribution (Gotman, 1999; James et al., 1999), which may also help highlight the boundaries of the region responsible for generating them (the so-called irritative zone). A previous study by our group (Pedreira et al., 2014) demonstrated the successful use of an automated neuronal spike classification algorithm, *Wave_clus* (WC) (Quian Quiroga et al., 2004), to classify epileptic spikes on scalp EEG for the purpose of modelling the concurrently acquired functional MRI. In this study we present and apply a validation framework for a similar application of WC to icEEG recordings (for the purpose of modelling concurrent fMRI data; which will be the topic of future work).

Our aim was to compare human expert IED classification as it is performed in normal ('optimal') conditions against the automated classification method to be used with WC. To our knowledge no formal comparison of automated vs human observer classification of epileptic spikes on icEEG has been published to date. Our approach targets the following questions:

- Does WC-human epileptic spike classification agreement variability fall within inter-human classification agreement variability?
- Looking at the classification labels (or clustering groups) of individual spikes; are WC results similar to those of human observers?

To validate this framework we used data from 5 patients reviewed by 3 human observers for the comparison with WC. We hypothesise that WC can produce similar IED classification results to that of human EEG reviewers whilst also providing additional information.

2. Data and methods

2.1. Patients, icEEG recording and pre-processing

We analysed icEEG signals recorded in 5 right-handed men (24–39 years) who were undergoing simultaneous intracranial EEG-fMRI (Table 1). The five patients were selected based on the small number of polyspikes observed during the recording. All patients underwent intracranial EEG recordings for clinical purposes to delineate the ictal onset zone and/or to perform direct electrocortical stimulation following a recommendation of a multidisciplinary team meeting. Patients were invited to undergo simultaneous intracranial EEG-fMRI (icEEG-fMRI) recordings at the end of their clinical evaluation. This study was approved by the Joint UCL/UCLH Committees on the Ethics of Human Research, and the patients gave written informed consent. The icEEG recording obtained during the simultaneous icEEG-fMRI study was used since we ultimately want to apply WC in the analysis of icEEG fMRI data however, no fMRI data was analysed for the purpose of this study.

In each patient there were between 31 and 84 implanted electrode contacts on configurations including grid electrodes, depth electrodes or both. The electrodes were connected to an MR-compatible amplifier system (Brain Products, Gilching, Germany). icEEG signals were acquired at a sampling rate of 5 kHz. After recording, we applied offline correction for MR scanning artefacts (Allen et al., 2000) and the resulting EEG was down sampled to 250 Hz. The EEG was band-pass filtered (2–70 Hz) and the same referential montage was used for all 4 EEG reviewers.

2.2. IED detection

The 5 icEEG recordings were inspected by EEG reviewer 'H1' for clinical purposes using *BrainVision Analyser* (Brain Products, Germany). During this procedure H1 placed a marker close to the negative/positive peak of each IED event (across the entire recording) that had a single sharp component. We then randomly selected 100 IEDs, using a random number generator, from each recording for this study (see Fig. 1; step 1).

2.3. IED classification by human observers (H2, H3 and H4)

Reviewers H2 (10 years of experience in icEEG interpretation), H3 (4 years of experience in icEEG interpretation) and H4 (2 years of experience in icEEG interpretation) independently classified the IED events selected by H1 through visual inspection of the waveforms in a 300 ms time window using BrainVision Analyzer. H2-4 performed the classification by visualizing the EEG activity in all recorded channels, in order to replicate their standard modus operandi. For each patient they were asked to classify the events into IED classes or as non-IEDs. H2-4 were free to define and use as many IED classes as they felt appropriate for each recording. Of the three EEG reviewers, two (H2 and H3) were trained at the same institution. Implantation diagrams, showing the position of the electrodes in relation to the brain, were provided.

2.4. Automated IED classification (WC)

The automated classification method *Wave_Clus* is a modification of the one described in Pedreira et al. (2014) and summarised in a flowchart (see Fig. 1; step 2). First, between 8 and 14 channels of interest were selected for each patient based on channels in which the IEDs were noted in the clinical EEG report as being most prominent and frequent. Second, we modified the IEDs' temporal marking (by H1) by automatically adjusting them to the peak of the sharp wave across the channels of interest (details of this process can be found in Supplementary Methods 1.0).

The IEDs were segmented in 300 ms epochs around the peak of the sharp wave (100 ms pre-peak to 200 ms post-peak) and concatenated across the channels of interest to form meta-IEDs (Pedreira et al., 2014). WC was then used to perform automated classification on the meta-IEDs similarly to our previous work (Pedreira et al., 2014). Based on the morphology and distribution of the IEDs, the algorithm automatically determined the number of classes per case and the events assigned to them. Then, the user performed a visual verification of the final classes obtained; including some events which were labelled as 'non-IED'.

2.5. Automated IED classification validation

We wanted to answer the question: can the results of the automated classification be distinguished from those obtained from humans? More specifically, we compared the two types of IED classification in two ways: first, we determined whether WC-human reviewer agreement variability falls within inter-human reviewer agreement variability; second, we compared *Wave_Clus* and