Clinical Neurophysiology 128 (2017) 433–441

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Interrater reliability of visually evaluated high frequency oscillations

Aaron M. Spring [a,b,c], Daniel J. Pittman [b,c], Yahya Aghakhani [a], Jeffrey Jirsch [d], Neelan Pillay [a], Luis E. Bello-Espinosa [a,e], Colin Josephson [a], Paolo Federico [a,b,c,f,*]

[a] Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada
[b] Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada
[c] Seaman Family MR Research Centre, Foothills Medical Centre, Calgary, AB, Canada
[d] Department of Medicine, University of Alberta, Edmonton, AB, Canada
[e] Department of Paediatrics, University of Calgary, Calgary, AB, Canada
[f] Department of Radiology, University of Calgary, Calgary, AB, Canada

ARTICLE INFO

Article history:
Accepted 15 December 2016
Available online 30 December 2016

Keywords:
High frequency oscillations
Interrater reliability
Interrater variability
Spikes
Intracranial electroencephalography
Interictal epileptiform discharges

HIGHLIGHTS

- Epoched study design is an effective method of evaluating interrater reliability of visual assessment of high-frequency oscillations (HFOs).
- HFO identification agreement between two visual reviewers is poor (mean Cohen's Kappa = 0.403).
- Translation of HFOs to clinical practice requires a framework to reconcile important findings of existing HFO studies.

ABSTRACT

Objective: High frequency oscillations (HFOs) and interictal epileptiform discharges (IEDs) have been shown to be markers of epileptogenic regions. However, there is currently no 'gold standard' for identifying HFOs. Accordingly, we aimed to formally characterize the interrater reliability of HFO markings to validate the current practices.
Methods: A morphology detector was implemented to detect events (candidate HFOs, lower-threshold events, and distractors) from the intracranial EEG (iEEG) of ten patients. Six electroencephalographers visually evaluated these events for the presence of HFOs and IEDs. Interrater reliability was calculated using pairwise Cohen's Kappa (κ) and intraclass correlation coefficients (ICC).
Results: The HFO evaluation distributions were significantly different for most pairs of reviewers ($p < 0.05$; 11/15 pairs). Interrater reliability was poor for HFOs alone ($\kappa_{mean} = 0.403$; ICC = 0.401) and HFO + IEDs ($\kappa_{mean} = 0.568$; ICC = 0.570).
Conclusions: The current practice of using two visual reviewers to identify HFOs is prone to bias arising from the poor agreement between reviewers, limiting the extrinsic validity of studies using these markers.
Significance: The poor interrater reliability underlines the need for a framework to reconcile the important findings of existing studies. The present epoched design is an ideal candidate for the implementation of such a framework.

1. Introduction

In the treatment of medically-refractory focal epilepsy, successfully eliminating the occurrence of seizures is dependent upon the localization and removal of the regions generating the seizures. These epileptogenic regions may be identified by localizing interictal epileptiform discharges (IEDs) on the intracranial electroencephalogram (iEEG) (Blume et al., 2001; Ebersole and Wade,

* Corresponding author at: C1241a Foothills Medical Centre, 1403 29th Street NW, Calgary, AB T2N 2T9, Canada.

E-mail addresses: amspring@ucalgary.ca (A.M. Spring), djpittman@me.com (D.J. Pittman), Yahya.Agha-Khani@albertahealthservices.ca (Y. Aghakhani), jirsch@ualberta.ca (J. Jirsch), pillay@ucalgary.ca (N. Pillay), Luis.Bello@albertahealthservices.ca (L.E. Bello-Espinosa), colin.josephson@gmail.com (C. Josephson), pfederic@ucalgary.ca (P. Federico).

http://dx.doi.org/10.1016/j.clinph.2016.12.017
1388-2457/

1991). Recently, high frequency oscillations (HFOs) (Cho et al., 2014; Jacobs et al., 2010a) and HFOs occurring simultaneously with IEDs (HFO + IEDs) (Jacobs et al., 2008; Wang et al., 2013) have both been shown to be more effective than IEDs alone at delineating the epileptogenic regions.

However, there is presently no gold standard for the identification of HFOs. To use HFOs in clinical practice or to assess any given automated detection algorithm, each study or center must establish its own operational definition of HFOs and select a 'ground truth' for HFO identification. Typically, this 'ground truth' is selected as the concordant visual assessment of two reviewers (Amiri et al., 2016; Dümpelmann et al., 2012; Ferrari-Marinho et al., 2015; Frauscher et al., 2015; Jacobs et al., 2016, 2009, 2010a,b; Kerber et al., 2014; Pail et al., 2013; van Diessen et al., 2013; Zelmann et al., 2012, 2010, 2009; Zijlmans et al., 2009a), or the concordance of one visual reviewer and an automated detection algorithm (Crépon et al., 2010; Nagasawa et al., 2012; Sakuraba et al., 2016; Staba et al., 2002), or even the markings of a single visual reviewer (Bagshaw et al., 2009; Burnos et al., 2016; Chaitanya et al., 2015; Ellenrieder et al., 2012; Haegelen et al., 2013; Jacobs et al., 2008; Urrestarazu et al., 2007; Wang et al., 2013; Zijlmans et al., 2009b).

To date, the interrater reliability of visual HFO evaluations has not been formally investigated. HFO studies may currently be biased not only by the data available at any particular epilepsy center, but also by the tendencies of the selected reviewers (Worrell et al., 2012), in a manner that is heretofore undescribed. Previous studies have shown significant variability between reviewers in the evaluation of gamma oscillations (Gardner et al., 2007) and other EEG phenomena (Abend et al., 2011; Bendabis et al., 2009). A study using an animal model of epilepsy illustrated large variability between visual HFO markings of four reviewers, but the data under reviewed were limited to local field potentials, and no formal evaluation of variability or reliability was made (Salami et al., 2012).

The algorithmic identification of candidate HFO events for subsequent verification by visual reviewers constitutes an ideal platform for the investigation of the reliability of visual HFO evaluation. In particular, methods currently proposed (Lévesque et al., 2011; Salami et al., 2012) for the algorithmic identification of HFOs can be adapted to detect events at multiple thresholds, generating a spectrum of events. This spectrum of events may then be evaluated for HFOs by multiple visual reviewers, and the interrater reliability computed from their discrete evaluations.

To this end, the present study modified an HFO detection algorithm to identify events with varying probabilities of containing HFOs, and presented these to a set of epileptologists for visual review. The distributions of HFO and HFO + IED ratings were compared across reviewers, and Cohen's Kappa coefficients were computed for all pairs of reviewers to assess the interrater reliability of HFO and HFO + IED evaluation.

2. Methods

This study was approved by our local Research Ethics Board. Ten patients were recruited into the study, all with medically intractable epilepsies undergoing intracranial video-EEG monitoring (iVEM) for clinical purposes at our medical center (Supplementary Table S1). All data used in this study were collected primarily for clinical purposes, and were used retrospectively in this study without additional burden on the patients.

2.1. Preprocessing

All patients underwent multiple days of iVEM, with data sampled at either 1000 Hz or 2000 Hz. Twenty minutes of continuous iEEG data were selected preferentially to reduce the chance of artifact while allowing for sufficient postsurgical stabilization of the EEG. In particular, the data were typically selected beginning at midnight of the fifth day post-implantation; wherever such data were unavailable, the time nearest to midnight, and the day nearest to the fifth were selected. Data were derived differentially, using methods appropriate to the electrode types. Bipolar montages were used for strip and depth electrodes, where the differences between two adjacent contacts constituted the derived channel. Laplacian montages were used for grid electrodes, where a derived channel was obtained for each contact by subtracting the average of all adjacent contacts (i.e., two adjacent contacts for corner contacts, three adjacent contacts for edge contacts, or four adjacent contacts for all central contacts).

The montaged data were bandpass filtered (80–250 Hz) using a finite-impulse response filter, designed using the Remez-exchange (equiripple) algorithm to minimize the generation of artifact morphologically similar to HFOs (Widmann and Schröger, 2012). The filtered data were amplitude-normalized by the root-mean-square of sliding 1-second epochs, and three categories of events were detected from all channels of the normalized data.

2.2. Algorithmic identification of events

An automated detection algorithm was written based on criteria used by a previously published detector (Lévesque et al., 2011; Salami et al., 2012), and used to detect three types of events (EventTypes): candidate HFO events (CDT), low-threshold HFO events (LTR), and distractor events (DST), as outlined in Table 1 and illustrated in Fig. 1. Events were selected independently for each channel; in other words, an event of any type in Channel A would preclude the extraction of a nearby event in the Channel A, but would not affect the extraction of events in Channel B.

Both CDT and LTR events were selected based on the morphology of the normalized signal in order to represent potential HFOs as described in the literature. In both cases, the signal was deemed to have exceeded threshold in any given half-cycle only if it remained above the set threshold for between 2 and 7 ms. CDT events were selected first, according to strict criteria (see Table 1 for the detailed criteria); LTR events were selected where more broad criteria were met, provided they occurred at least 500 ms from any CDT event, in order to allow reviewers to determine their own internal threshold for marking HFOs.

DST events were identified where the criteria for CDT or LTR events were not met. These events were selected to provide reviewers with a representation of baseline activity less likely to contain EEG phenomena that may be interpreted as HFOs. The DST events were selected pseudorandomly, excluding data within 500 ms of any event of type CDT or LTR.

2.3. Visual reviewers

Six expert electroencephalographers were recruited into this study as visual Reviewers. Two Reviewers were previously trained to identify HFOs at the Montreal Neurological Institute (MNI; Montreal, QC), and another epileptologist was extensively trained by one of them, while the remaining three had no prior experience identifying HFOs. At the time of the study, four Reviewers were adult epileptologists, one was a pediatric epileptologist, and one was a senior epilepsy fellow. Five Reviewers were from Calgary, AB, and one (JJ) was from Edmonton, AB. All six were given a 20-min instructional video, containing a tutorial on using the evaluation program, the criteria used for HFOs, and examples of certain HFOs determined by two previously trained Reviewers. They were also given a printout summarizing the key information.