Clinical Neurophysiology 118 (2007) 1744–1752

www.elsevier.com/locate/clinph

Comparison of novel computer detectors and human performance for spike detection in intracranial EEG

Merritt W. Brown III [a], Brenda E. Porter [a], Dennis J. Dlugos [a], Jeff Keating [c], Andrew B. Gardner [c], Phillip B. Storm Jr. [b], Eric D. Marsh [a,*]

[a] *Division of Child Neurology, Children's Hospital of Philadelphia, Abramson Research Building, Room-502A, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA*

[b] *Pediatric Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA*

[c] *Department of Neurology and Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA*

Accepted 14 April 2007

Available online 1 June 2007

Abstract

Objective: Interictal spikes in intracranial EEG (iEEG) may correlate with epileptogenic cortex, but review of interictal iEEG is labor intensive. Accurate automated spike detectors are necessary for understanding the role of spikes in epileptogenesis.

Methods: The sensitivity, accuracy and reproducibility of three automated iEEG spike detectors were compared against two human EEG readers using iEEG segments from eight patients. A *consensus set* of detections was generated for detector calibration. Spike verification was calculated after both human EEG readers independently reviewed all detections.

Results: Humans and two of the three automated detectors demonstrated comparable accuracy. In four patients, automated spike detection sensitivity was >70% and accuracy was >50%. In the remaining four patients, EEG background morphology resulted in poorer performance. Blinded human verification accuracy was 76.7 ± 6.6% for computer-detected spikes, and 84.5 ± 4.1% for human-detected spikes.

Conclusions: Automated iEEG spike detectors perform comparably to humans, but sensitivity and accuracy are patient dependent. Humans verified the majority of computer-detected spikes.

Significance: In some patients automated detectors may be used for mapping spike occurrences in epileptic networks. This may reveal associations between spike distribution, seizure onset, and pathology.

Keywords: Spike detection; Intracranial EEG; Reproducibility; Reliability

1. Introduction

Epileptiform spikes and sharp waves during interictal EEG recordings are signs of cortical hyperexcitability, and suggest a tendency toward epileptic seizures. Characterization of the spatiotemporal distribution of spikes and sharp waves during presurgical intracranial EEG (iEEG) – especially with regard to seizure onset, underlying pathology, and outcome – may improve surgical planning and patient outcome. In order to study the distribution of these events, a valid iEEG spike detector must be employed. This paper reports the sensitivity, accuracy and reproducibility of three automated iEEG spike detectors (C1, C2, and C3), and compares their performance to two human EEG readers (H1, H2).

Development of a valid and reliable iEEG spike detector has been elusive. To date, almost all spike detectors have been designed for scalp EEG. Numerous challenges exist for spike detection in both scalp and intracranial EEG (Frost, 1985; Wilson and Emerson, 2002; Pang et al., 2003) including: difficulties with artifact rejection, state changes during a recording (Gotman and Wang, 1991, 1992), and variability of background activity in both

---

[*] Corresponding author. Tel.: +1 215 590 1719; fax: +1 215 590 3779. *E-mail address:* marshe@email.chop.edu (E.D. Marsh).

1388-2457/$32.00 © 2007 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.clinph.2007.04.017

normal and abnormal recordings. In addition, scalp EEG detectors must differentiate normal EEG transients morphologically similar to epileptiform discharges (e.g., wicket rhythm, mu rhythm, and sharp alpha). Algorithm design is made more difficult due to the lack of an universally accepted definition of iEEG or scalp spikes (gold standard), thereby making validation of an instrument challenging (Frost, 1985; Wilson et al., 1996; Dumpelmann and Elger, 1998; Gotman, 2001).

Many algorithms for spike detection have been proposed, including: mimetic and rule-based approaches (Wilson and Emerson, 2002), frequency domain methods (Gotman and Gloor, 1976), wavelet transforms (Dumpelmann and Elger, 1999), artificial neural networks (Webber et al., 1994; Ozdamar and Kalayci, 1998; Wilson et al., 1999), independent component analysis (Kobayashi et al., 2001), support vector machines (Acir and Guzelis, 2004), template matching (Sankar and Natour, 1992), topographic classification (Sankar and Natour, 1992; Feucht et al., 1997; Sugi et al., 2002; Adjouadi et al., 2004, 2005). Most techniques were developed specifically for scalp EEG, though some reports of their use (without validation) on iEEG have been made (Katz et al., 1991; Hufnagel et al., 2000; Asano et al., 2003). Only two studies generated new methods for iEEG spike detection (Dumpelmann and Elger, 1998, 1999; Valenti et al., 2006), with only partial comparisons to human performance for validation. Hence, with a lack of iEEG specific spike detectors available, two novel iEEG spike detectors were developed based on a combination of mimetic- and frequency-based approaches. The detectors performances were quantitatively validated against two humans and a commercial spike detector. Our results, which summarize the first comprehensive validation of computer-based spike detectors for iEEG, suggest that human-comparable automated spike detection algorithms can be used to analyze iEEG data with the goal of delineating the importance of interictal spikes for surgical planning.

2. Methods

2.1. Clinical data

This study was approved by The Children's Hospital of Philadelphia (CHOP) Institutional Review Board. Consent was obtained from patients' families and all clinical data, including recordings, were de-identified. Intracranial EEG recordings obtained from eight patients undergoing subdural electrode implantations as part of routine epilepsy surgical evaluations were utilized. Patients were selected from a group of 30 that had undergone iEEG recordings at the CHOP between 2003 and 2006. They represented a varied subset of phase II surgical cases for which detailed iEEG seizure marking and surgical pathology were present. All iEEGs were recorded with Grass-Telefactor 128 Channel CTE EEG machine using 16-bit amplifiers (Astro Med Corp., West Warwick, RI) sampled at 200 Hz per channel. An analog antialiasing bandpass filter (frequency cut-offs at 0.1 and 70 Hz) and notch filter (null at 60 Hz) were used for signal conditioning.

Recordings were reviewed in a referential montage, and marked prior to this experiment to determine seizure onset times and electrode locations. The reference electrode was placed contra-lateral to the recording grid and in the mid-temporal region. The reference electrode placement was consistent between patients to the degree that post-surgical placement allowed. A clinical epileptologist (E.M.), who did not participate in the marking experiments, selected two 10-min interictal epochs per patient for analysis. The selection of segments made sure that artifacts were excluded, particularly reference related artifacts that a bipolar montage would have removed. Each epoch consisted of two channels of iEEG containing spikes (total count ranged from 5 to 175 spikes per segment, or 0.25–8.75 spikes/min/channel). A representative single contact recording from four patients and highlighted spikes is shown in Fig. 1.

2.2. Human spike detection

All markings were performed using a graphical user interface (GUI), designed in MATLAB (The Mathworks, Inc., Natick, MA), which presented iEEG at 10 s per screen display, with a sensitivity of 200 μV per millimeter. The readers could adjust sensitivity but not the number of seconds displayed per page. Human reviewers, H1 and H2, marked onset and offset times of the spikes. For all markings, spike detections produced by two detectors occurring within 100 ms of each other were deemed equivalent. All markings were saved automatically to a database for further analysis.

Since there is no universal, precise definition of a spike, direct comparison of human and computer performance overly-penalizes automated algorithms (e.g., a single human is used as a gold standard, but it is known that humans exhibit significant interrater variability). To address this, we formed a consensus set of gold standard detections in the following manner: two expert readers reviewed the iEEG together using the MATLAB GUI and jointly identified spikes for each patient. The qualitative working definition of an iEEG spike required three features: (1) a negative polarity deflection with a sharp or spike morphology that stands out from the IEEG background, (2) an after-going slow wave, and (3) a duration less than 200 ms. Automated detectors were calibrated against the consensus set of detections to determine: true positives (TP), detections which agreed with the consensus set; false positives (FP), computer detections not present in the consensus set, and false negatives (FN), consensus events that went undetected. We did not consider true negatives (TN) for two reasons: (1) they unduly bias performance measures when the events of interest (e.g., spikes) are rare, and (2) without a perfect detector there is always some uncertainty surrounding true negative markings. The