Publicly Accessible Penn Dissertations

Spring 2010

UNSUPERVISED CLASSIFICATION OF HIGH-FREQUENCY OSCILLATIONS IN NEOCORTICAL EPILEPSY AND CONTROL PATIENTS

Justin A. Blanco
University of Pennsylvania, blancoj@seas.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Biomedical Engineering and Bioengineering Commons

UNSUPERVISED CLASSIFICATION OF HIGH-FREQUENCY OSCILLATIONS IN NEOCORTICAL EPILEPSY AND CONTROL PATIENTS

Abstract

Quality of life for the more than 15 million people with drug-resistant epilepsy is tied to how precisely the brain areas responsible for generating their seizures can be localized. High-frequency (100-500 Hz) field-potential oscillations (HFOs) are emerging as a candidate biomarker for epileptogenic networks, but quantitative HFO studies are hampered by selection bias arising out of the need to reduce large volumes of data in the absence of capable automated processing methods. In this thesis, I introduce and evaluate an algorithm for the automatic detection and classification of HFOs that can be deployed without human intervention across long, continuous data records from large numbers of patients. I then use the algorithm in analyzing unique macro- and microelectrode intracranial electroencephalographic recordings from human neocortical epilepsy patients and controls. A central finding is that one class of HFOs discovered by the algorithm (median bandpassed spectral centroid ~140 Hz) is more prevalent in the seizure onset zone than outside. The outcomes of this work add to our understanding of epileptogenic networks and are suitable for near-term translation into improved surgical and device-based treatments.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Bioengineering

First Advisor
Brian Litt, MD

Second Advisor
Leif Finkel, MD PhD

Keywords
high-frequency oscillations, epilepsy, seizure localization, neocortex, automated EEG analysis, intracranial EEG

Subject Categories
Biomedical Engineering and Bioengineering

UNSUPERVISED CLASSIFICATION OF HIGH-FREQUENCY

OSCILLATIONS IN NEOCORTICAL EPILEPSY AND CONTROL

PATIENTS

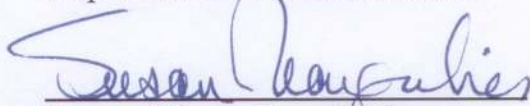Justin A. Blanco

A Dissertation

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2010

Brian Litt, MD
Supervisor of Dissertation

Susan Margulies, PhD
Graduate Group Chairperson

Leif H. Finkel, MD PhD
Co-Advisor of Dissertation

Dissertation Committee:
Abba M. Krieger, PhD
Diego Contreras, MD PhD
Gershon Buchsbaum, PhD
Douglas C. Maus, MD PhD

Unsupervised Classification of High-Frequency Oscillations in

Neocortical Epilepsy and Control Patients

To mom

For being a model of resiliency

Acknowledgment

Foremost, I owe thanks to Dr. Brian Litt for kindling my interest in HFOs and providing unflagging support and encouragement, clinical insights, and the freedom to pursue ideas in my own way – wrong turns, dead ends, and all. To the late Dr. Leif Finkel I am indebted for his accepting me into the Neuroengineering Lab, for connecting me with Brian, and for his many wisdom-filled chats about science, graduate school, and academia. His earnestness, humility, and intellect have left a permanent impression, and I miss him.

I thank Drs. Abba Krieger and Stephen Isard for tremendous technical advice and for devoting so much of their precious time to one-on-one scientific mentorship. I thank Drs. Diego Contreras, Gershon Buchsbaum, and Douglas Maus for the guidance and feedback they have given as members of my thesis committee, and Dr. Buchsbaum in particular for dutifully stepping in to play a role after Leif's passing.

I thank Drs. Greg Worrell and Matt Stead for their graciousness as hosts on my

research visits to the Mayo Clinic, and for the many hours they spent in the laboratory and operating room collecting data that they were willing to share.

Jonathan Viventi, my lab mate, has worn many hats in his tenure; the two for which I am most grateful are "friend" and "IT-guy." In the latter capacity he has been the creator of many a hack that got the job done, and without him I am certain I would still be trying to figure out how to properly span hard disks or synchronize files, rather than submitting this dissertation.

I thank Dr. Marc Dichter for his helpful suggestions in lab meetings and for endorsing my summer study in Woods Hole. I thank the Epilepsy Foundation for their grant support, which enabled a good portion of the work in these pages. I thank Carolyn Wilkinson and Ruthie Krieger for all their administrative efforts, Ed Nieh for helping condense and organize clinicians' reports on patients, and Drausin Wulsin for saving me some time by writing the code to render figure 6.1. I thank Drs. Bill Stacey, Brian Litt, and Eric Marsh for marking records, and Doug Maus and Matt Stead for passing along their data decompression routines, the use of which was a humble first step down a long road.

Thanks en masse to the many other collaborators, lab mates, and friends I've come to know during my PhD years, including Drs. Kareem Zaghloul, Stephen Wong,

ABSTRACT

UNSUPERVISED CLASSIFICATION OF HIGH-FREQUENCY

OSCILLATIONS IN NEOCORTICAL EPILEPSY AND CONTROL PATIENTS

Justin A. Blanco

Brian Litt, MD and Leif H. Finkel, MD PhD, Advisors

Quality of life for the more than 15 million people with drug-resistant epilepsy is tied to how precisely the brain areas responsible for generating their seizures can be localized. High-frequency (100-500 Hz) field-potential oscillations (HFOs) are emerging as a candidate biomarker for epileptogenic networks, but quantitative HFO studies are hampered by selection bias arising out of the need to reduce large volumes of data in the absence of capable automated processing methods. In this thesis, I introduce and evaluate an algorithm for the automatic detection and classification of HFOs that can be deployed without human intervention across long, continuous data records from large numbers of patients. I then use the algorithm in analyzing unique macro- and microelectrode intracranial electroencephalographic recordings from human neocortical epilepsy patients and controls. A central finding is that one class of HFOs discovered by the algorithm (median bandpassed spectral centroid $\sim$140 Hz) is more prevalent in the seizure onset zone than outside. The outcomes of this work add to our understanding of epileptogenic networks and are suitable for near-term translation into improved surgical and device-based treatments.

Contents

List of Tables

List of Figures

Chapter 1

Motivation

1.1 Problem and Clinical Significance

Epilepsy is the second most common neurological disorder after stroke, affecting approximately 50 million people worldwide [62]. For about 25% of this population, seizures cannot be controlled by any combination of traditional therapies [5] – medication and/or focal resection of brain tissue. The large proportion of patients without a viable treatment option has remained frustratingly constant over the past 15 years despite advances in surgical technology and substantial innovation in antiepileptic drug therapy, which has served primarily to mitigate the side-effects of treatment rather than make more people seizure-free [69]. This therapeutic plateau is motivating research into alternative treatment technologies, among the most promising of which are implantable devices designed to detect, predict, pre-

vent, and abort seizures.

Several seizure control devices are currently in clinical trials [52, 63, 4], and while preliminary safety results are encouraging, early efficacy results are more modest – NeuroPace's multi-center (safety) trial of a closed-loop brain stimulator, for example, reports less than 5% of patients achieving total seizure freedom – suggesting that the ultimate success and widespread clinical adoption of these devices will depend upon a more complete understanding of where, when, and how to deliver therapy within the brain.

This dissertation, whose specific contents we outline in the following section, aims to contribute to our understanding of the first of these three interconnected issues, *where* to direct treatment, which is arguably the most fundamental. It is difficult to imagine, for example, how the timing and structure of newly proposed therapies like controlled electrical [57] or magnetic stimulation [43], focal cooling [84], gene therapy [35], and drug-delivery [70] – or the efficacy of present day surgical resections – could be optimized without first having a reasonable picture of what the therapeutic targets should be.

Research aimed at finding therapeutic targets has been centered primarily on so-called "focal" epilepsies, in part because they are perceived to be the most clinically

tractable. Though diverse in etiology, these epilepsies share the characteristic that the earliest electrographic abnormalities preceding seizures are consistently observed on a small and usually compact subset of recorded electroencephalography (EEG) channels. This has led to a conceptual model of seizure generation that posits the existence of an epileptic focus: a time-stable and well circumscribed population of cells in which pathologic activity originates before propagating more globally within the brain. Moderate support for this conceptualization of focal epilepsies is provided by neurosurgical outcomes: in one study of medically refractory mesial temporal lobe epilepsy patients, for example, focal tissue resections guided by clinicians' EEG-based identification of the "seizure onset-zone" (SOZ), led to seizure freedom in about 50% of patients at 5 years post-operation [72].

This dissertation is grounded in the belief that the epileptic focus is a useful construct; at the same time, it permits a broader interpretation of what might comprise it. Rather than a single, compact and time-stable group of neurons, we consider that the epileptic focus could be a spatially distributed cluster of loci, possibly one whose members change across time. A more suitably general term for the epileptic focus viewed this way, which has traction in the epilepsy research community, is "epileptogenic network." Our central hypothesis is that, whatever its neural substrate, the epileptogenic network represents the origin of pathologic activity and emits a characteristic signature that can be detected using some measurement modality.

We focus our attention on the intracranial electroencephalogram (iEEG) because recent investigations using tools heretofore largely unavailable to clinicians – specifically high bandwidth recordings – suggest that a candidate biomarker for the epileptogenic network is emerging in this modality: transient, quasi-periodic activity with frequency between 100 and 500 Hz, collectively termed "high-frequency oscillations" (HFOs) to distinguish it from the slower activity (0.1-40 Hz) more commonly studied in the epilepsy community and traditionally used to make clinical decisions. Representative examples of HFOs are shown in figure 1.1.

HFOs have garnered clinical interest for their potential to provide more accurate information about seizure-generating brain areas than the state-of-the-art method of iEEG-based seizure localization: visual inspection of recordings for channels showing early seizure onset. If an improved localizing capacity can be demonstrated for HFOs, it will significantly impact clinical practice, potentially increasing both the efficacy of surgical resections to stop seizures and the number of medically refractory patients who become eligible for surgery.

In this dissertation we argue that the ability to mine HFOs from iEEG *automatically* is required to study their relationship with seizure onset regions effectively. As collateral benefits should HFOs prove practically useful, automation will dramat-
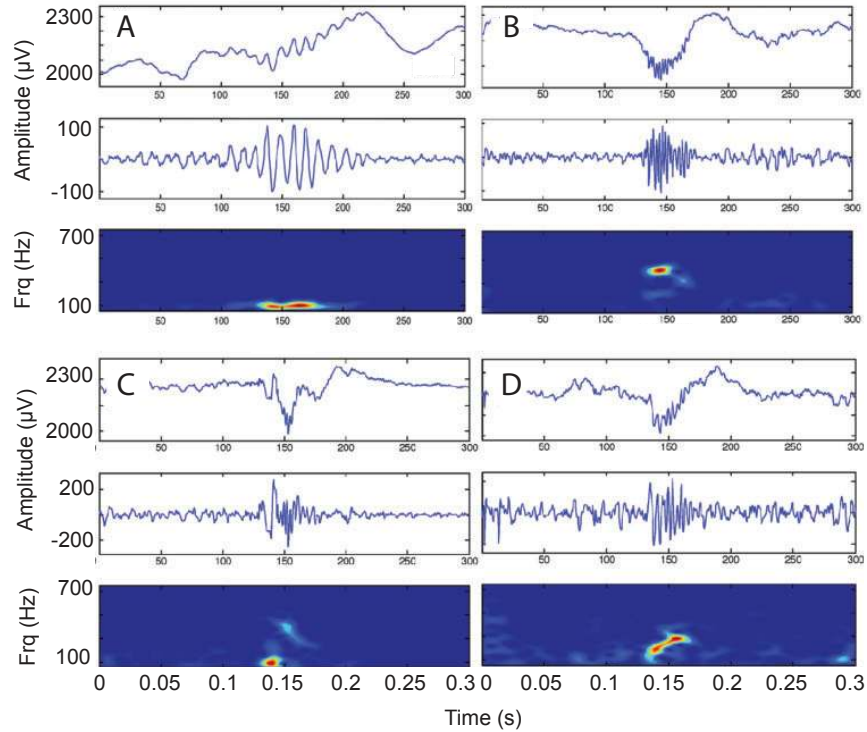
Figure 1.1. Four examples of 100-500 Hz high-frequency oscillations (A-D) recorded in patients with mesial temporal lobe epilepsy. Each HFO is depicted in three different views: unfiltered iEEG with the HFO centered at 0.15 s (top); the corresponding bandpass filtered iEEG (80-1000 Hz) (middle); and a spectrogram (2.6 ms window) of the same 0.3 s duration iEEG segment (bottom). Adapted from [83]; vertical scaling in the top and middle views is approximate.

ically reduce both clinician workload and the potential for human error and bias to impact the localization process. It will also influence the design of implantable seizure warning and control devices, which are a promising alternative for patients who are ineligible for surgery, either because their seizures are localized to eloquent cortex whose removal would lead to unacceptable neurological deficits, or because their epileptogenic networks are multi-focal, broadly distributed, or dynamically changing. In these cases, HFOs may provide the critical control signal for closed-loop devices, indicating where and when to deliver therapy for maximal efficacy.

1.2 Dissertation Goals and Roadmap

The clinical utility of HFOs as seizure localizing signals will ultimately be tested by correlating HFO properties with treatment outcomes and by conducting prospective clinical trials in which HFO measures are used to dictate treatment. Before such studies can be undertaken, however, more basic challenges must be met. These include developing reliable tools for the automated detection and classification of HFOs; comparing proposed machine methods with human performance; investigating whether there are preferred electrode geometries for recording HFOs; and improving our understanding of which attributes or classes of HFOs are characteristically epileptic and which are normal.

This dissertation takes up these challenges. It is organized in the following manner. Chapter 2 is a concise review of the scientific literature on HFOs[1]. It is not intended to be comprehensive, since further discussion of prior research comes in the introduction and discussion sections of succeeding chapters, which are organized similarly to scientific research articles. It is meant to fill in the gaps, in particular to provide the early history of HFO research leading up to the discovery of a link

---

[1]The review begins with a discussion of "ripples," originally a synonym for "high-frequency oscillations." More recently, the term "high-frequency oscillation" has come to signify the grouping of two putatively distinct classes of oscillations within the range 100-500 Hz: ripples (100-200 Hz) and "fast ripples" (250-500 Hz). Fast ripples are discussed in the review as well.

with epilepsy, and to review the body of work on HFOs in mesial temporal lobe structures, which motivates the present work in neocortex.

Chapter 3 (together with appendix A) presents the original engineering contributions of the thesis. It introduces a new algorithm for the fully automated detection and classification of HFOs. The main techniques employed – some classical and some lesser known – do not in themselves constitute the novel aspect of the work; although there are indeed some original computational features described, this thesis is not primarily about new theoretical developments in signal processing or machine learning. What is most unique is the manner in which the tools are assembled and their application to the HFO analysis problem. One specific new idea the thesis presents is treating the post-processing of an initial set of highly noise-polluted detections in two specialized filtering stages: the first is designed to address a particular theoretical (and practical) vulnerability of the state-of-the-art detector, and its role is newly framed as a frequency domain outlier detection problem that can be tackled using density estimation techniques; the second also involves a new problem formulation, being the first treatment of HFO candidates that uses the exploratory data analysis technique of clustering. Appendix A supplements chapter 3, describing in greater detail the computational features from which the inputs to our machine classifier are derived.

Chapters 4 and 5 contain the original scientific contributions of the thesis. Chapter 4 presents new findings about behavior: the degree to which human expert reviewers agree about what constitutes an HFO; as well as how our automated classifier both compares with their decisions and gives us insight into their biases. We make a data-driven argument for the downplaying of highly coveted "ground truth" data in these early stages HFO discovery. Chapter 5 is perhaps of greatest interest to clinicians and neuroscientists. It describes one of a very small number of investigations of neocortical, rather than mesial temporal, HFOs in humans; includes an even rarer description of spontaneous neocortical HFOs in patients with no history of seizures; and is the only study of which we are aware that directly compares the two groups. We investigate whether HFOs in any of the automatically detected clusters show changes in rate based on their locations within or outside the physician-labeled seizure onset-zone. We also ask whether electrodes with diameters on the micrometer scale tend to record HFOs of higher frequencies than standard, millimeter-scale clinical macroelectrodes. Chapter 6 concludes with ideas for future extensions of the work presented here, including transitioning the algorithm from the laboratory to the clinic.

Chapter 2

Background

2.1 Early Studies

Narrowband transient field-potential oscillations greater than 100 Hz and lasting tens of milliseconds, originally termed "ripples," were observed in *in vivo* extracellular microelectrode recordings from mammalian hippocampus more than thirty years ago [55]. Over the last 15 years, beginning with the work in rodents of Buszaki and colleagues [16], researchers have studied the physiology of these quasi-periodic events intensively.

Early hippocampal studies in rodents reported ripples spatially localized to CA1 and occurring predominantly during behavioral immobility, slow-wave sleep, and consummatory behaviors [16], often but not always in association with large 40-100

ms depolarizing events – called sharp waves [17] – induced by the synchronous discharge of CA3 pyramidal cells [85]. Ripple duration and peak frequency could be modulated by anesthesia [85], with Halothane in particular - a gap-junction blocker - leading to their nearly complete suppression while preserving sharp waves [25, 85]. The spatial characteristics of ripples and their correlation with single-cell activity were also investigated in these early experiments: ripples recorded at probes separated by distances as large as 4-5 mm within CA1 were found moderately coherent (for 3-5 cycles) [18]; and while the firing of most recorded cells showed no relationship to ripples, small proportions ($\sim$10-15%) of both pyramidal cells and interneurons fired with increased probability and in a phase-locked manner during the events – though only interneurons could fire at rates as fast as the ripple oscillations themselves.

Results from these early studies motivated different (though possibly complementary [24]) hypotheses about the cellular and network-level mechanisms underlying ripples, as well as their function. It is suggested, for example, that ripples in the hippocampal CA1 region reflect summed inhibitory postsynaptic potentials resulting from synchronously firing presynaptic interneurons [85]; that they are enabled by axo-axonal gap junction coupling of principal cells [25]; and also that they are formed by bursts of pyramidal cell population spikes [9]. Ripples are thought to be important in declarative memory consolidation [82], but have also been implicated

in pathological processes [10, 12].

Since studies in the early nineteen nineties in hippocampus, the ripple "story" has grown increasingly complex. Not only are ripples now reported in hippocampal areas outside of CA1, such as CA3, but also in parahippocampal areas [18] and in neocortex [33, 34] in animal models. In addition, they no longer appear to be confined to slow-wave sleep, immobility, and consummatory behaviors, having been observed during the active states of waking and REM sleep as well [33], albeit typically in smaller numbers. In a recent attempt to synthesize the spectrum of findings [75], Roger Traub begins by identifying two broad conditions under which ripple-like activity has been described in brain-slice, rodent, cat, and human experiments: when local field potentials are relatively small ($\sim$200 μV) and thus principal cells are presumed not to be particularly depolarized; and when large depolarizing events like sharp-waves are observed in conjunction with ripples. He then delineates at least eight subcategories of activity within these two broader classes, not necessarily disjoint, to which researchers have attached the "ripple" label. Factors distinguishing his categories include: spectral composition of the ripples; region of the brain in which they are observed; influence of the ionic composition of the extracellular fluid (in slice bath); whether they occur spontaneously or after tetanic stimulation; effect of chemical blockage of GABAa receptors by bicuculline; and the characteristics of the background activity upon which ripples are superimposed

(e.g. sharp-waves, evoked somatosensory potentials, synchronized bursts).

As Traub notes, whether a common mechanism unites all the varied examples of ripples that have been reported remains to be elaborated. But the heterogeneity of experimental paradigms, animal models, electrode-types and configurations, and data analysis methods used to record and make inferences about ripples in the years since the original rodent studies underscores the need to carefully contextualize both interpretations of findings in the literature and new hypotheses and experiments.

This dissertation, including the remainder of this review, is concerned with spontaneous, transient, quasi-periodic activity whose peak frequency falls in the band between 100 and 500 Hz, recorded *in vivo* using intracranial electrodes. Though our primary interest is in neocortical HFOs, we review findings from the hippocampal-entorhinal axis as well, with a focus on epileptic humans. In particular, we emphasize work investigating the relationship between ripples and seizure generation[1], giving the majority of attention to two types of papers: 1) those that give relatively strong statistical support for their claims; and 2) those that appear to have been particularly influential in inspiring further research and shaping current clini-

---

[1]In doing so, we deemphasize results pertaining to relationships between ripples and unit activity, including the many physiological and computational investigations of the mechanisms by which ripples are generated (for a review of this literature, see [24]). Typically, these studies are done in highly controlled environments, usually in anesthetized animals and/or slice preparation, with the full complement of chemical, electrophysiological, and histopathological tools for targeting and confirming placement of electrodes, and identifying cortical layers and specific neuron classes – techniques that cannot be used in the clinical environment for a variety of ethical and practical reasons.

cal opinion, even if their conclusions were based on less quantitative analyses[2].

2.2 Ripples and Fast Ripples in Epilepsy

The connection between ripples and the epilepsy research community began against the backdrop discussed above, in large part with the work of Bragin et al. [9, 10]. While earlier studies had investigated *broadband* activity above 100 Hz in human epilepsy patients, primarily in association with seizures [2, 30], Bragin was the first to report discrete, interictal ripple-like oscillations analogous to those that had been recorded in rodents, in a cohort of mesial temporal lobe epilepsy (MTLE) patients. In a departure from prior rodent studies, however, Bragin et al. conceptually subdivided high-frequency oscillations into two classes: "ripples," with peak frequency between 80 and 160 Hz, and "fast ripples," with peak frequency between 250 and 500 Hz[3], observing that the latter class appeared to be unilateral in a given patient while the former seemed to occur in both hemispheres. Fast ripples reportedly occurred during the same behavioral states and "randomly interspersed" with ripples on a given electrode, and in another study [13] did not have significantly different amplitude distribution from ripples. When Bragin et al. later mapped perforant

---

[2]We largely neglect single-patient case studies, studies whose findings essentially confirm previous results, and those whose claims are based on anecdotal rather than statistical evidence.

[3]Around the same time, Csicsvari et al., in rat experiments, also posited the existence of distinct ripple and fast ripple classes [21], though his proposed ranges – 100-130 Hz for the former and 140-200 Hz for the latter – were almost entirely subsumed by Bragin's human ripple definition.

path evoked and spontaneous fast ripples in 100 μm intervals within rat dentate gyrus, they found them more spatially constrained than prior reports of ripples, reporting fast ripples localized to a region 1 mm in length, which has led to the estimate that their volumetric extent surrounding a source is approximately 1 mm^3 [11].

2.2.1 HFOs and seizure onset regions in MTLE

In a separate paper discussing the same group of human subjects [9], Bragin et al. described their ripples and fast-ripples in more detail. Ripples were observed in 67% (6 of 9) of subjects, while fast ripples were observed in 5 of the 6 patients showing ripples. In 4 of the 5 patients with fast ripples – or 44% of the entire epileptic patient pool – fast ripples occurred only in areas identified as epileptogenic zones during presurgical evaluation[4]. In the remaining fast ripple patient, however, the opposite result was observed: fast ripples were recorded exclusively contralaterally to the presumed epileptogenic zone.

In an analysis of microelectrode recordings from the hippocampal-entorhinal axis in mesial temporal lobe epilepsy patients, Staba et al. provided some statistical support for dividing supra-100 Hz transients into ripple and fast ripple categories:

---

[4]No statistics were provided on the extent of the latter zones.

a model with relatively few parameters (a mixture of two Lorentzian distributions) that was a good fit to the apparently bimodal distribution of peak frequencies of detected oscillations [68]. To strengthen the support for discrete populations, the group discredited the idea that the observed bimodality might be the result of individual events themselves having two prominent spectral peaks, stating that double-peaks were found only rarely when they inspected power spectral density plots for a (presumably random) subsample of events ($< 5\%$ of $n = 200$). Having thus justified the binary classification of detections based on peak spectral frequency, they went on to provide a detailed quantitative description of ripples and fast ripples, drawing several important conclusions: 1) the ratio of fast ripple to ripple rates was significantly larger for channels ipsilateral to the clinically determined seizure-onset zone than for contralateral channels; 2) the latter difference was linked to the tendency for hippocampal atrophy to occur on the ipsilateral side; and 3) both ripples and fast ripples were significantly shorter in duration on the ipsilateral side.

In another study of temporal lobe epilepsy patients, Worrell et al. [83] did not find the bimodal distribution of peak frequencies that was observed by Staba et al. When they nonetheless classified events as ripples and fast ripples according to ranges similar to those proposed by Staba et al., they reached a second conflicting conclusion: whereas Staba et al. had found increases in fast ripples and *decreases* in ripples ipsilaterally, Worrell et al. found increases in the rates of both classes

ipsilateral to the seizure-onset zone. In addition, Worrell et al. called attention to the spectral complexity of detected events, which had been explicitly deemphasized by Staba et al., noting the difficulty of labeling detections as ripples or fast ripples based on a single value (peak frequency).

The finding in mesial temporal lobe structures of rate increases for both ripples and fast ripples in the seizure onset zone – in accord with Worrell et al. – has been reported in several other investigations [39, 40, 41]. However, the latter are more difficult to interpret due to heavy reliance on manual processing, aggregation of patients with neocortical and mesial temporal epilepsy when reporting significance, and the use of statistical tests that rely on distributional assumptions where nonparametric methods may be more appropriate.

2.2.2 Relationship to electrographic seizures

A few groups have studied the relationship between ripples and electrographic seizures. In anesthetized cats, Grenier et al. recorded ripples (80-200 Hz) from neocortex and showed that the ripples associated with spikes during seizures were larger in amplitude than those occurring in association with spikes during the interictal slow-wave oscillation [34]. Bragin et al. found a modest correlation between time to first detection of a 100-500 Hz high-frequency oscillation and time to first seizure in kainic-acid induced seizing rats [12]. No study of which we are aware,

however, yet supports the idea that any property of high-frequency oscillations, either ripples or fast ripples, can be used to predict the occurrence of seizures.

2.3   Conclusion

There is accumulating evidence that narrowband voltage transients with peak frequencies in the range between 100 and 500 Hz may be useful biomarkers for epileptogenic tissue. Whether the subdivision of this band into "ripple" and "fast ripple" oscillations is physiologically or clinically relevant, however, is less clear. Indeed, which properties of these transients carry the most information for distinguishing pathologic from non-pathologic tissue is still largely an open question, as is whether HFOs have any predictive value for signaling seizure occurrence.

The overwhelming majority of HFO analyses in the epilepsy literature and beyond concern mesial temporal lobe regions. Epileptic neocortex is by comparison a frontier. Further work must be done to quantitatively describe HFOs recorded there, and to compare results with the larger body of data from hippocampal and parahippocampal structures, as well as with non-seizing control patients. We review the relatively scant literature on neocortical HFOs in chapter 5, where we detail our own contribution to the area.

In that chapter we also address another topic in need of future research, the spatial scale over which high-frequency oscillations are generated, and in particular whether electrodes of different sizes preferentially record HFOs of different types. Currently there is controversy in the field, with some groups holding that pathological HFOs are most readily recorded with electrodes on the order of tens of micrometers or smaller [83, 64], while others contend that no useful information is lost using standard, millimeter-scale clinical electrodes [42, 20].

Many studies to date of high-frequency oscillations in epilepsy suffer from small sample-size, and substitute anecdotal evidence and descriptive statistics for more careful quantitative analysis. The few studies such as those by Staba et al. [68] and Worrell et al. [83] that have taken more rigorous approaches have yielded enticing results, but have nonetheless been hampered by selection bias arising primarily out of the need to reduce the volume of data available for analysis. Data reduction, in turn, is such an acute problem because automated HFO analysis is not yet good enough to eliminate the burden of human pre-screening and post-processing. The development of reliable automated HFO processors and their blinded deployment across long, continuous records drawn from larger numbers of patients will be required to make definitive statements about the relationship between HFOs and seizure generation. In the following chapters, we describe our contributions to this effort.

Chapter 3

Description of the Algorithm

3.1 Summary

High-frequency oscillations (HFOs) have been observed in animal and human intracranial recordings during both normal and aberrant brain states. It has been proposed that the relationship between subclasses of these oscillations can practically be used to identify epileptic brain. Studies of HFOs in epilepsy have been hampered by selection bias arising primarily out of the need to reduce the volume of data so that clinicians can manually review it. In this chapter, we introduce an algorithm for detecting and classifying these signals automatically, and demonstrate the tractability of analyzing a data set of unprecedented size, over 31,000 channel-hours of intracranial electroencephalographic recordings (iEEG) from micro- and macroelectrodes in humans. Using an unsupervised approach that does not pre-

suppose a specific number of clusters in the data, we show direct evidence for the existence of distinct classes of transient oscillations within the 100 to 500 Hz frequency range in a population of nine neocortical epilepsy patients and two controls. The number of classes we find, four – three plus one putative artifact class – is consistent with prior studies that identify "ripple" and "fast ripple" oscillations using human-intensive methods, and additionally identifies a less examined class of mixed-frequency events.

3.2   Introduction

In the epilepsy research community, there is mounting interest in the idea that high-frequency oscillations (HFOs) – narrowband transients recorded on the intracranial electroencephalogram (iEEG), having predominant frequency between roughly 100 and 500 Hz and lasting on the order of tens of milliseconds – can be used to identify epileptogenic brain tissue. From a clinical perspective, the utility of HFOs as biomarkers lies in their potential to improve the outcomes of surgical resections in patients with medication resistant epilepsy, either by augmenting or replacing signals currently used to delineate epileptogenic cortex. HFOs might also serve as input signals to controllers within implantable devices designed to deliver therapeutic intervention – for example, targeted electrical stimulation, drug delivery, or tissue cooling – in closed-loop fashion. Such devices, one of which is currently in clinical trials [4], promise to help the large population of medically refractory

epilepsy patients who are not candidates for resective surgery, either because their seizures appear to emanate from "eloquent" cortical areas with vital function, or because epileptiform activity appears diffuse or multi-focal within the brain.

To date, most studies of HFOs in epilepsy rely on highly human-intensive methods to extract the signals of interest from multichannel iEEG. With the goals of increasing the tractability of labor-intensive marking and analysis tasks and improving the interpretability of results, investigators typically perform dramatic data reduction steps before committing the data to statistical analysis. Examples of common data culling measures include pre-selecting: 1) subjects, electrodes, channels, and time-epochs that, upon visual pre-inspection, show prominent examples (e.g. high signal-to-noise ratio) of the activity to be analyzed; 2) electrodes or contacts which, according to MRI, appear located in gray matter or specific brain structures; 3) data recorded during certain states of consciousness, typically slow-wave sleep; 4) data recorded during specific brain states such as interictal, preictal, or ictal periods; 5) data recorded from periods deemed artifact-free, typically by human reviewers; and 6) data – or a subsample thereof, such as a 10-minute segment – that satisfy some combination of the above criteria. A smaller number of studies employ machine-based HFO detection in semi-automated procedures in which similar human screening occurs before the automatic detector is deployed, as well as after, in an attempt to reduce the number of spurious machine detections. Such quasi-

automated methods are designed with the primary aim of reducing the workload for human reviewers.

In addition, based in large part on human observation, most studies of HFOs in epilepsy presuppose two discrete categories of these oscillations, distinguished by the sub-band within 100-500 Hz in which the majority of the transient signal's energy lies. The two types of high-frequency oscillations are termed "ripples" and "fast ripples," and statistical analyses of HFO properties are typically preceded by the forced classification of all detected events as one of these two types.

In this chapter, we present a methodology for automatically detecting and classifying high-frequency oscillations in the human intracranial electroencephalogram. Differing from prior approaches to automated detection, we use no data pre-selection beyond our choice to study patients with epilepsy believed to be of neocortical origin and control patients, analyzing all recorded data from all available channels in all available patients. Furthermore, we do not presuppose the existence of two distinct subpopulations of HFOs. Instead, we treat this idea as a hypothesis, using it to inform the design of some of the features we use as inputs to our classifier. We develop an unsupervised, rather than supervised scheme for event classification because a) we believe ideas about what constitutes a physiologically or clinically meaningful HFO are still evolving, and the former framework is more natural for

exploratory analysis; and b) prior studies in which "expert" human reviewers have been asked to mark HFO-like activity on the iEEG have shown poor inter-rater reliability and reproducibility [32]. In a similar vein, we find no good reason to constrain the categorization of events to two types. Hence, our classification method requires no pre-specification of the numbers of HFO classes present in the data.

The primary purpose of this chapter is to present our method – its rationale as well as the technical details to replicate it – along with results that support its utility as a tool for automatically processing and organizing large quantities of neurophysiologic data containing spontaneous, transient events, in a manner that minimizes selection bias as well as human subjectivity and labor. Further, we show objective evidence for the existence of distinct classes of oscillations which have heretofore been taken largely for granted, strengthening the argument that "ripples" and "fast ripples" are meaningful physiologic labels (even if perhaps not the best terminology for avoiding confusion in the literature), and suggest that a third class of oscillations containing components of both frequencies should be investigated.

3.3 Methods

3.3.1 Patient population and data acquisition

Nine patients with medically refractory partial epilepsy believed to be of neocortical origin underwent implantation of subdural electrodes (Ad Tech Medical Instrument Company, Racine, WI) to localize the seizure onset zone after noninvasive monitoring was indeterminate. The location, number, and type of intracranial electrodes (grid and/or strip and/or depth electrodes) were determined by a multi-disciplinary team, including neurosurgeons, neurologists, neuroradiologists, and neuropsychologists, as part of routine clinical care. Standard clinical grid and strip electrodes were modified under an Institutional Review Board (IRB)-approved research protocol, by adding arrays of non-penetrating platinum-iridium microwires (40 μm diameter, with intra-array spacing of 0.5-1 mm center-to-center) between the clinical, 4 mm diameter contacts [78]. Standard clinical depth leads were similarly modified, in two ways: 1) by embedding microwires around the circumference of the lead-body between the 2 mm long clinical contacts; and 2) by passing a bundle of microwires within the lumen of the lead, so that they protruded by approximately 7-8mm from the distal tip [83]. Schematic depictions of the electrode modifications and an image of a typical implanted subdural grid are shown in figures 3.1A and 3.1B, respectively.

Two "control" patients, with chronic intractable facial pain but with no history of
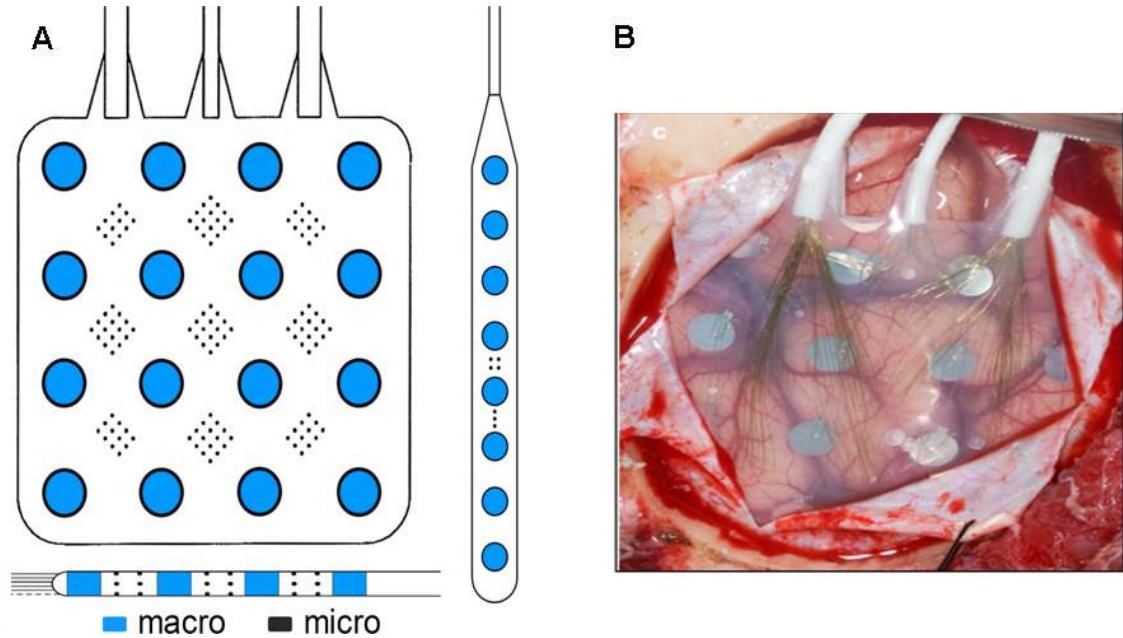
Figure 3.1. (A) Example schematics for typical hybrid grid and strip (above) and depth (below) electrodes. Arrays of microwire contacts are located at regular intervals between the larger clinical contacts. Drawings are not to scale. (B) Image showing subdural placement of a hybrid grid electrode.

seizures, were similarly implanted, as part of an unrelated, IRB-approved research protocol investigating electrical stimulation of motor cortex as a potential treatment for their condition.

In addition to intracranial electrodes, a limited montage of standard gold scalp electrodes (EOM1, EOM2, Fp1, Fp2, Fpz, C3, and C4) was placed on all patients, as well as electrodes to record electromyographic signals from the chin and tibialis anterior surface. Two stainless steel surgical sutures (Ethicon Inc., Somerville, NJ) were placed at the vertex region of the head at the time of surgery and used as reference and ground, respectively, for both intra and extracranial electrodes.

25

Continuous, long-term data were acquired using the Digital Lynx Data Acquisition System (Neuralynx, Inc., Bozeman, MT) at 32,556 samples per second and 20 bits per sample (stored), from up to 144 channels in each patient [14]. The input dynamic range was $\pm$ 132 mV and the noise level was $\sim$1.3 μV root-mean-square (RMS), yielding approximately 18 effective bits. Recordings were made using direct current (DC) capable amplifiers, and a 9 kHz analog lowpass filter was employed to minimize aliasing effects. The bandwidth of the raw recordings was thus approximately near-DC - 9 kHz.

3.3.2   Overview of HFO detection and classification

The automated HFO detection and classification algorithm is comprised of three major stages, depicted in figure 3.2.

In the first stage, candidate HFO events are detected in the bandpass filtered iEEG using a method designed by Staba et al. [68] to be sensitive, but which is highly non-specific. This initial pass over the data set dramatically reduces its size. In the second stage, a statistical model of the local background iEEG surrounding each candidate event is built, and events bearing too large a spectral similarity to the background activity according to the model are discarded from candidacy. In

26
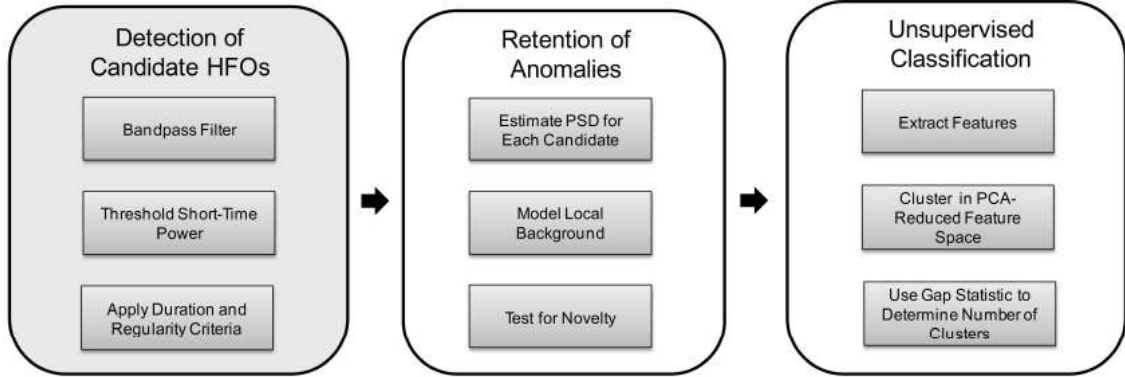
Figure 3.2. Block diagram showing the three-stage detection and classification algorithm. The first stage was described in [68]. The second and third stages are the contributions of this work.

the final stage, computational features are extracted from the retained candidates (which we label simply "anomalies," since they may or may not represent clinically meaningful events), and these features are used, after a dimensionality reduction step, as inputs to a classifier. Importantly, the classifier uses an automatic method of determining the number of clusters in the data, and the possibility that the data do not cluster at all (i.e. there is a single cluster) is not excluded. Below, we describe the method in detail.

3.3.3 Stage 1: Detection of candidate HFOs

HFOs are initially detected using methods closely adapted from those by Csicsvari et al. [21, 22] and Staba et al. [68]. Raw iEEG data are processed on a per-channel basis in 10-minute non-overlapping segments. Data are first decimated by a fac-

tor of 12 (8th order low pass Chebyshev Type I filter, cutoff frequency of 1085 Hz, forward and reverse filtered to eliminate phase distortion), to 2713 Hz, before band-pass filtering between 100 and 500 Hz (20th order Cauer filter, forward and reverse filtered; specifications: 65 dB minimum lower/upper stop band attenuation, 0.5 dB maximum pass band ripple, 25 Hz lower/upper transition width). Filters were de-signed and tested for stability using Matlab's filter design toolbox (The Mathworks, Natick, MA).

Using a 3 millisecond sliding window, a running RMS signal is computed from the bandpass data. The RMS signal is then compared against a threshold (the mean of the RMS signal plus 5 standard deviations) and successive samples exceeding this threshold for a minimum duration of 6 milliseconds are delimited by their upward and downward threshold crossing times. Marked events are then subject to the additional criterion that they must have at least six peaks greater than 3 standard deviations from the mean of the rectified bandpass signal. Finally, retained events separated by less than 10 milliseconds are merged to generate a candidate set of HFOs.

3.3.4 Stage 2: Retention of anomalies

Staba et al. [68] report high sensitivity for their detector when compared with a ground truth set of human markings, but the method suffers high false positive

rates [32], especially when no data pre-selection is performed. In pilot studies, we observed that a common putative failure mode for the detector was retaining transients with amplitudes larger than the average global background signal (on which the algorithmic threshold is based) but similar in morphology to the local[1] background iEEG. An example of these typical putative false positives is shown in figure 3.3A.

The design of the second stage of the algorithm therefore assumed that a key aspect of an HFO's saliency for a human reading iEEG is not only its relative amplitude but also its spectral distinctiveness - the degree to which its morphology, as opposed to simply its size, "pops out" from the background. All stage 1 detections that fail to meet a spectral novelty criterion, as described below, are eliminated.

Candidate HFOs passing through stage 1 are processed individually in stage 2^2, which operates on the decimated data. Twelve hundred milliseconds (after a 5 ms guard time) on either flank of a given candidate are used to build a model of the local background iEEG. Background data are segmented into clips of length equal to the candidate detection[3].

---

[1]By "local" we mean the amount of time typically displayed when screening for HFOs in raw or modestly filtered data (e.g. 1 second per screen with $\sim$0.5 Hz – 500 Hz bandwidth, $\sim$1.2 kHz sampling rate, and 7 μV/mm vertical scaling, on a monitor with 1280 x 1024 pixel resolution).

[2]To simplify notation, in equations (3.1) - (3.14) that follow, we have suppressed the indexing subscripts that indicate that all functions and variables are specific to a particular HFO candidate, but we here emphasize that these subscripts are implicit.

[3]Detections exceeding 50 ms ($\sim$14%) are first truncated to 50 ms to place a lower limit on the number of background segments available.
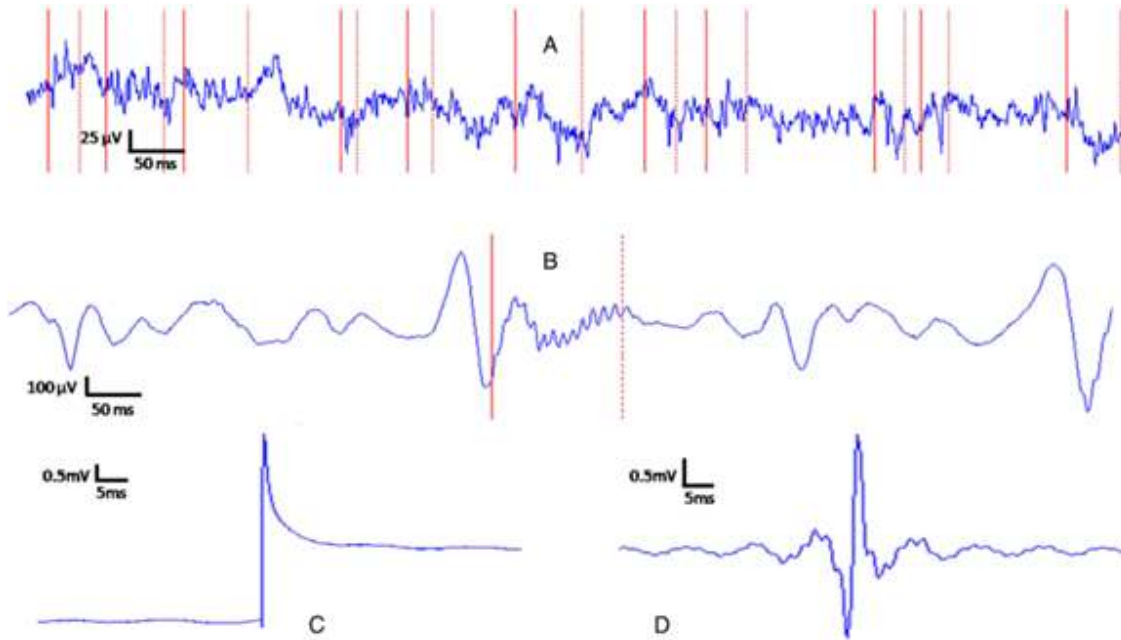
Figure 3.3. Examples showing the lack of specificity of the stage 1 detector – partial motivation for the design of stages 2 and 3. (A) A run of putative false positive stage 1 detections, none of which appear visually well distinguished from the surrounding iEEG. Individual detections are delimited by solid (start) and dotted (stop) red lines. These detections should be contrasted with the single detection in (B), a putatively valid HFO. (C) Typical sharp transient artifact detected by stage 1; unfiltered waveform. (D) Band-pass filtered version of C; this filter artifact is what the stage 1 detector actually "sees," giving insight into why the event is flagged as an oscillation. The algorithm we present is designed to discard in stage 2 events like those in A, and to aggregate into a single cluster in stage 3 events like that represented by C and D. More interesting events like that in B should group naturally into one or more clusters in stage 3.

For each background segment, as well as the candidate detection itself, the best linear fit is found using least-squares regression and removed by subtraction (i.e. the signal is detrended). Each segment is then energy-normalized by dividing by its Euclidean length, and a power spectral density estimate is computed using a multitaper technique from Thomson [73] (3 tapers; 512-point DFT (zero-padded); adaptive nonlinear combination method). The method in general involves multiplying the signal by n orthogonal data tapers having optimal spectral concentration properties – discrete prolate spheroidal sequences [66] – before computing the squared magnitude of the discrete Fourier transform in each case. This produces n independent spectral estimates, which are then additively combined using weighting factors that are adaptively chosen depending on local characteristics of the spectrum.[4]

The local background iEEG is modeled using a mixture of Gaussians, under the assumption that each background segment is an observation emitted from one of a small number (one to three) of multivariate Gaussian components. Thus, the probability of observing any local background segment, $p(\mathbf{x})$, is represented by the density function[5]:

---

[4]Multitaper techniques are particularly well suited to the analysis of short duration signals, as well as signals with spectra having a large dynamic range of amplitudes (see e.g. [59] and [53] for good statistical discussions of why), making them an appropriate choice for our application.

[5]In the GMM discussion below, we use the notation style of Bishop [8].

$$p\left(\mathbf{x}\right) = \sum_{k=1}^{K} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}} \cdot |\mathbf{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\mathbf{x} - \mathbf{u}_k\right)^T \mathbf{\Sigma}_k^{-1}\left(\mathbf{x} - \mathbf{u}_k\right)\right\} \qquad (3.1)$$

where $\mathbf{x}$ is the D x 1 vector representing the background segment, K is the number of mixture components, π_k is the weight of the kth mixture component ($\sum_{k=1}^{k} \pi_k = 1, \pi_k \geq 0 \; \forall \; k \in \{1, 2, \ldots, K\}$), D is the dimensionality of the data, $\mathbf{\Sigma}_k$ is the D x D covariance matrix for the kth mixture component, and $\mathbf{u}_k$ is the D x 1 mean vector for the kth mixture component.

The objective is to maximize the (log) probability of the collection of background segments surrounding each candidate HFO, assuming the background segments are independent and identically distributed. This quantity viewed as a function of the model parameters - the component weights, mean vectors, and covariance matrices (unconstrained, in our case) - is known as the (log) likelihood function, and in this case can be written as:

$$\ln p\left(\mathbf{X}|\mathbf{u}, \mathbf{\Sigma}, \pi\right) = \sum_{n=1}^{N} \ln\left\{\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\mathbf{u}_k, \mathbf{\Sigma}_k)\right\} \qquad (3.2)$$

where $\mathbf{X}$ is the N x D data matrix with N the number of background segments, $\mathbf{u}$ is the D x K matrix of horizontally concatenated mean vectors, $\mathbf{\Sigma}$ is the D x

32

$(D \cdot K)$ matrix of horizontally concatenated covariance matrices, π is the 1 x K vector of component weights, and $\mathcal{N}(\cdot)$ is the multivariate normal distribution (shown uncollapsed in (3.1), above – the expression to the right of π_k).

Maximization is performed using the Expectation Maximization (EM) algorithm [23], a two stage iterative numerical procedure. We set initial component weights equal to each other (i.e. $\pi_k = \frac{1}{K}$ for all k) and initialize mean vectors and covariance matrices using the k-means algorithm ([48]; also described below). The EM update steps, which guarantee an increase in the log-likelihood with each iteration (see e.g. [8] for a readable proof) are then performed as follows:

$$\text{E-Step:} \quad \gamma^{(t)}(z_{nk}) = \frac{\pi_k^{(t)} \mathcal{N}\left(\mathbf{x}_n | \mathbf{u}_k^{(t)}, \mathbf{\Sigma}_k^{(t)}\right)}{\sum_{j=1}^{K} \pi_j^{(t)} \mathcal{N}\left(\mathbf{x}_n | \mathbf{u}_j^{(t)}, \mathbf{\Sigma}_j^{(t)}\right)} \tag{3.3}$$

$$\text{M-Step:} \quad \mathbf{u}_k^{(t+1)} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma^{(t)}(z_{nk}) \, \mathbf{x}_n \tag{3.4}$$

$$\mathbf{\Sigma}_k^{(t+1)} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma^{(t)}(z_{nk}) \left(\mathbf{x}_n - \mathbf{u}_k^{(t+1)}\right) \left(\mathbf{x}_n - \mathbf{u}_k^{(t+1)}\right)^T \tag{3.5}$$

$$\pi_k^{(t+1)} = \frac{N_k}{N} \tag{3.6}$$

$$\text{where} \quad N_k = \sum_{n=1}^{N} \gamma^{(t)}(z_{nk}) \tag{3.7}$$

Note that z is K-dimensional random variable (called a "latent" variable because it is not explicitly observed) having K possible states, such that whenever a particular element is equal to 1, all others are equal to zero. Its marginal distribution is defined in terms of the mixture component weights, such that $p(z_k = 1) = \pi_k$. It is introduced in order to achieve a formulation of the likelihood function that is amenable to maximization using EM. $\gamma(z_{nk})$ is then defined as the posterior probability that $z_k = 1$ given that data segment $\mathbf{x}_n$ has been observed, or, more colloquially, the probability that mixture component k is "responsible" for generating observation $\mathbf{x}_n$. The EM algorithm allows us to learn the parameters of the model from the data. When it converges, it is guaranteed to converge to (at least) a local maximum of the likelihood function[6].

Given the small number of background segments relative to the dimension of the data, to combat the curse of dimensionality [6] we reduce the number of dimensions of each background segment (and the candidate segment) to two (i.e. D = 2), prior to learning the Gaussian Mixture Model (GMM). This is done using principal

---

[6]We take the algorithm to have converged when successive computed values of the likelihood function differ by less than or equal to 1 x 10^{-5}. In cases where numerical issues arise or where convergence is not achieved within a pre-specified number of iterations (500) – either due to encountering singularities of the likelihood function or to otherwise slow rates of convergence – the algorithm returns an "indeterminate" flag, and the HFO candidate in question is passed to the third stage of processing without interruption of the program flow.

components analysis (PCA) [37, 58] on the frequency-domain representations of the background segments discussed above, after removing the mean and dividing by the standard deviation of the background segments.

The principal components can be found by successively seeking out the spatial directions along which the lengths of the orthogonal projections of the data observations have maximal variance, subject to the constraint that each successive direction is orthogonal to its predecessors. These directions are exactly the eigenvectors of the covariance matrix, $\mathbf{C}$, of the data:

$$\mathbf{C} = \frac{1}{N} \sum_{n=1}^{N} \left(\mathbf{b}_n - \overline{\mathbf{b}}\right) \left(\mathbf{b}_n - \overline{\mathbf{b}}\right)^T \tag{3.8}$$

where $\mathbf{b}_n$ is the P x 1 power spectral density representation of background segment n, described above, and $\overline{\mathbf{b}}$ is the mean of all N background segments associated with a given HFO candidate (for us, $\overline{\mathbf{b}} = 0$). Since the number of background segments is smaller than their dimensionality, P, the data lie in a linear subspace whose maximum dimension is $N - 1$. Therefore, at least $P - N + 1$ eigenvalues (projection variances) must be zero, and this fact is used (see e.g. [8] pp. 569-570 for how) to increase the efficiency with which the relevant eigendecomposition is performed. The new coordinates for each background data segment are computed as:

$$\mathbf{X} = \mathbf{BU} \tag{3.9}$$

where $\mathbf{X}$ is the new N x D data matrix of background-segment representations, $\mathbf{B}$ is the N x P matrix whose i^{th} row is $\mathbf{b}_i^T$, and $\mathbf{U}$ is the P x D matrix whose columns are the unit-normalized eigenvectors of $\mathbf{C}$ corresponding to the D largest eigenvalues (for us, $D = 2$, as mentioned above). The D-dimensional projection for the HFO candidate segment itself is then computed, after removing the mean of the background segments and dividing by their standard deviation, using the same matrix $\mathbf{U}$.

Since there is no a priori reason to prefer a specific number of mixture components in the GMM (indeed, use of a GMM itself represents an assumption – about the appropriate model form), the parameters for $i = 1$, 2, and 3 component mixtures are learned, and the Bayesian Information Criterion (BIC) [65] is used to make the final model selection. The BIC, like other common information criteria (e.g. Akaike's [1]), contains a term that penalizes a model's complexity. It is used to guard against overfitting; that is, to adjust for increases in the data likelihood that occur only by virtue of having enough parameters to enable very precise tuning to the data set at hand, which presumably is just one sample from the space of all data sets generable by the process being modeled. The BIC can be derived starting from the Laplace approximation (see e.g. [47]) to the "model evidence" [8] – the

36

probability of the data given a particular model after marginalizing over all possible values of the parameters. With some nontrivial assumptions – namely, a) a broad (nearly uniform) Gaussian prior distribution over the parameters and; b) a Hessian matrix of the negative log-likelihood function (evaluated at the optimal parameter vector given the data) that is of full rank – the evidence for the ith model ($\mathcal{M}^i$), denoted by $p\left(\mathbf{X}|\mathcal{M}^i\right)$, can be approximated by:

$$\ln p\left(\mathbf{X}|\mathcal{M}^i\right) \approx \ln p\left(\mathbf{X}|\mathbf{u}^i_{\text{ML}}, \mathbf{\Sigma}^i_{\text{ML}}, \pi^i_{\text{ML}}\right) - \frac{1}{2}M^i \ln N \qquad (3.10)$$

[8] where the subscript ML stands for the "maximum likelihood" estimates found via EM, and the constant M^i in the second term on the right is the number of free parameters in the ith model. The computation in (3.10) is the Bayesian Information Criterion, and we select the model for which it is largest.

The goal in stage 2 is to assign a given HFO candidate to one of two classes, $\mathcal{B}$ (background) or $\mathcal{A}$ (anomaly), while minimizing the misclassification rate. This is theoretically done by assigning $\mathbf{h}$ to $\mathcal{B}$ whenever

$$p\left(\mathcal{B}|\mathbf{h}\right) > p\left(\mathcal{A}|\mathbf{h}\right) \qquad (3.11)$$

[26], where $\mathbf{h}$ is the 2-D representation of the HFO candidate, discussed above. Applying Bayes's Theorem, this condition can be shown to be equivalent to

$$p\left(\mathbf{h}|\mathcal{B}\right) > \frac{p\left(\mathbf{h}|\mathcal{A}\right)p\left(\mathcal{A}\right)}{p\left(\mathcal{B}\right)} \qquad (3.12)$$

The GMM describing $\mathcal{B}$ allows estimation of the quantity on the left directly, but there is no such model for $\mathcal{A}$ and one cannot reasonably be inferred given that there is at most a single observation from $\mathcal{A}$. The prior probabilities of $\mathcal{A}$ and $\mathcal{B}$, respectively, are similarly unknown.

To address these issues a heuristic criterion is used, which is based on the squared Mahalanobis distances, Δ_k^2, from the HFO candidate to the center of each GMM component:

$$\Delta_k^2 = \left(\mathbf{h} - \mathbf{u}_k\right)^T \mathbf{\Sigma}_k^{-1} \left(\mathbf{h} - \mathbf{u}_k\right) \qquad (3.13)$$

The squared Mahalanobis distances of a random sample drawn from a multivariate normal distribution (computed using the unbiased sample covariance matrix) will be distributed approximately as central chi-squared with D degrees of freedom, where D is the dimensionality of the data [51]. Using the assumption that $p\left(\mathbf{h}|\mathcal{A}\right)$ is a monotonic decreasing function of $p\left(\mathbf{h}|\mathcal{B}\right)$, so that the latter is high wherever the former is low, it is estimated that:

$$p\left(\mathcal{B}|\Delta_1^2, \ldots, \Delta_K^2\right) \approx \sum_{k=1}^{K} \pi_k \int_{\Delta_k^2}^{\infty} q\left(t\right) \mathrm{d}t \qquad (3.14)$$

where $q(x)$ is the central chi-squared density function with ν degrees of freedom:

$$q(x) = \begin{cases} \dfrac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} \exp\left(-\frac{1}{2}x\right) & x > 0 \\ \\ 0 & x < 0 \end{cases} \qquad (3.15)$$

The function $\Gamma(z)$ is the Gamma function, $\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)\mathrm{d}t$. A similar estimate was used by Roberts [61].

Procedurally, the percentage of the central chi-squared density lying to the right of the calculated Mahalanobis distance from each mixture component is found. A weighted average of these percentages, with weights equal to those of the corresponding mixture components is then computed. If the resultant estimated probability exceeds 5%, the HFO candidate is considered to have been generated by the local background process and it is removed from candidacy. All candidates for which the calculation in (3.14) - computed with respect to the candidate's unique local background model - falls below the 5% threshold are passed on to the final clustering stage. Figures 3.4 and 3.5 provide graphical depictions of stage 2: the former shows an event that was removed from candidacy, the latter an event that was passed on to stage 3.
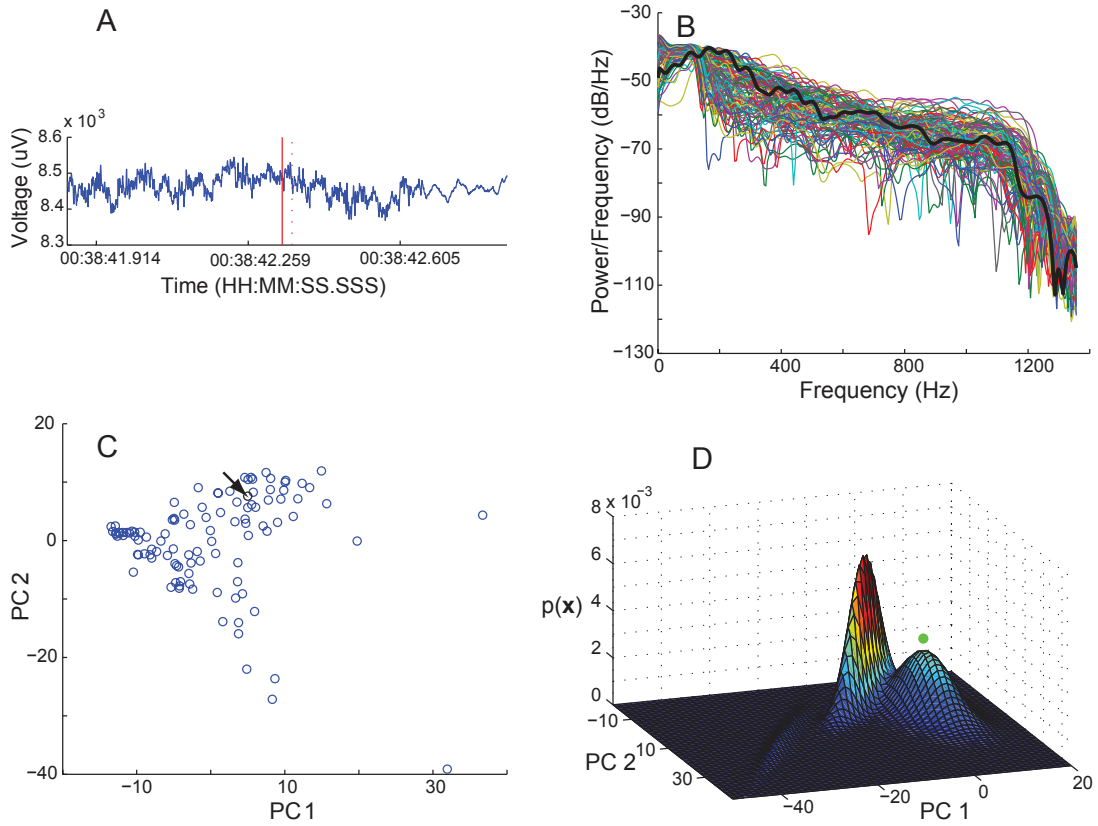
Figure 3.4. (A) Putative false positive detection, delimited by red lines. (B) Multitaper power spectral density estimates of the detection in A (black line) and of the signals obtained by segmenting the data flanking the detection in the manner described in section 3.3.4 (colored lines). Note that the black line, which is enhanced for visibility, falls within the "envelope" of the background spectra. (C) Projections (blue points) of the background spectra in B onto the plane defined by their first two principal components, along with the projection of the detection (black point). (D) Three-component Gaussian mixture model learned from the background data in C. Here the detection is shown in green and elevated above the density function to facilitate visualization. The test detailed in section 3.3.4 indicates that the detection is highly likely to have been generated from the background process described by the mixture model, and the detection is therefore discarded from candidacy.
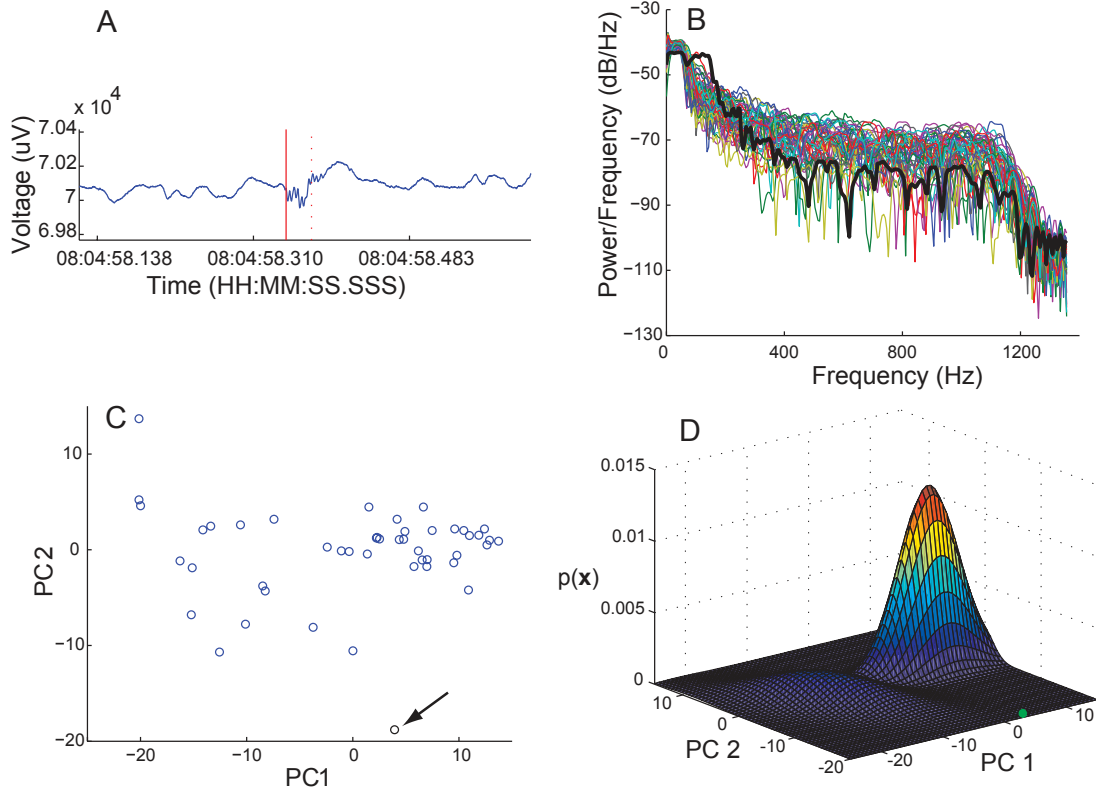
Figure 3.5. (A) Putative valid detection, delimited by red lines. (B) Multitaper power spectral density estimates of the detection in A (black line) and of the signals obtained by segmenting the data flanking the detection in the manner described in section 3.3.4 (colored lines). Note that the black line (in particular the values lying between roughly 100 and 200 Hz), which is enhanced for visibility, falls outside the "envelope" of the background spectra. (C) Projections (blue points) of the background spectra in B onto the plane defined by their first two principal components, along with the projection of the detection (black point). (D) Two-component Gaussian mixture model learned from the background data in C. Here the detection is shown in green to facilitate visualization. The test detailed in section 3.3.4 indicates that the detection is unlikely to have been generated from the background process described by the mixture model, and the detection is therefore passed on to stage 3 for clustering.

3.3.5 Stage 3: Unsupervised classification

3.3.5.1 Feature extraction

Before clustering, seven features are computed on the data corresponding to each event passing through stage 2. All features were designed or chosen with one of two goals in mind: 1) permit separation of "ripples" and "fast ripples," should there indeed be evidence for such compartmentalization (and not otherwise); 2) aggregate highly stereotypical, sharp transient recording artifacts that were observed in pilot studies (figures 3C and 3D). Features in category 1 were conceived a priori (blinded to all patient data), while features in category 2 were developed and refined using labeled artifacts from two patients (CT 01 and SZ 05). The features are described in detail in appendix A. The set of feature vectors $\{\hat{\mathbf{F}}_1, \ldots, \hat{\mathbf{F}}_I\}$, where I is the total number of anomalies passing through stage 2 (in our case, $I = 290,273$), is the data set that is clustered.

3.3.5.2 K-mediods clustering

Clustering was accomplished using a variant of the k-means algorithm [48]. The k-means algorithm seeks to minimize the quantity:

$$W_K = \sum_{k=1}^{K} \frac{1}{2N_k} \sum_{i,i' \in C_k} d_{ii'} \tag{3.16}$$

where K is the number of clusters, C_k denotes the set of indices of the observations

in cluster k, (with each observation assigned to a single cluster), and N_k denotes the number of observations in cluster k. $d_{ii'}$ is the squared Euclidean distance between P-dimensional observations x_i and $x_{i'}$:

$$d_{ii'} = \sum_{p=1}^{P} (x_{ip} - x_{i'p})^2 = ||(\mathbf{x}_i - \mathbf{x}_{i'})||^2 \tag{3.17}$$

The quantity W_K in (3.16) can be viewed as a modified version of the within-cluster point scatter [36] in which each cluster's contribution is now weighted inversely by its membership. Thus, roughly speaking, diffuse clusters with few members are not counted as energetically favorably as compact clusters with many. Minimizing this metric is exactly equivalent to minimizing the pooled within-cluster sum of squared distances from the cluster means (centroids), so that W_K can alternatively be written as[7]:

$$W_K = \sum_{k=1}^{K} \sum_{i \in C_k} ||\mathbf{x}_i - \bar{\mathbf{x}}_k||^2 \tag{3.18}$$

where $\bar{\mathbf{x}}_k$ is the mean of the elements in cluster k. An iterative descent algorithm for minimizing the above criterion can be obtained by first defining clusters via prototypes $\{\mathbf{u}_k\}_{k=1}^{K}$ and attempting to minimize, over parameters $\mathbf{u}$ and $\mathbf{Y}$, the error function:

---

[7]The technically inert factor of 2 in the denominator of (3.16) achieves this exact equivalence.

$$\varepsilon_K(\mathbf{u}, \mathbf{Y}) = \sum_{ki} \mathbf{Y}_{ki} ||\mathbf{x}_i - \mathbf{u}_k||^2 \qquad (3.19)$$

where $\mathbf{u}$ is the P x K matrix of prototypes and $\mathbf{Y}$ is a K x I indicator matrix of cluster assignments, with I the total number of data observations. That is, $\mathbf{Y}_{ki} = 1$ whenever $\mathbf{x}_i$ is assigned to cluster k, and 0 otherwise. The algorithm proceeds as follows:

1. Initialize $\mathbf{u}$ randomly.[8]

2. For each $\mathbf{x}_i$, compute the nearest prototype.

$$\text{set } \mathbf{Y}_{ki} = \begin{cases} 1 & \text{if } k = \underset{j}{\operatorname{argmin}} ||\mathbf{u}_j - \mathbf{x}_i||^2 \\ 0 & \text{otherwise} \end{cases} \qquad (3.20)$$

3. Recompute cluster prototypes.

$$\text{set } \mathbf{u}_k = \frac{\sum_i \mathbf{Y}_{ki}\mathbf{x}_i}{\sum_i \mathbf{Y}_{ki}} \qquad (3.21)$$

4. Return to step 2. Repeat until convergence.

As step 2 minimizes the error function with respect to $\mathbf{Y}$ and step 3 minimizes the error function with respect to $\mathbf{u}$, ε is guaranteed to decrease with each iteration so

---

[8]We use a randomly chosen k observations from the data set, $\mathbf{X}$.

that convergence – to a local minimum, at least – is assured.

Instead of using the L2 norm (Euclidean distance), our classification scheme uses the L1 norm (Manhattan distance) as distance metric, which yields the k-mediods algorithm. In step two of the algorithm (3.20), a given point is assigned to the prototype that is closest in the L1 sense, and in step three (3.21) the appropriate minimization is obtained by computing the mediod, rather than the centroid, for each newly formed cluster. The k-mediods method is used to achieve a clustering that is more robust against outliers than k-means, albeit at computational expense [36].

3.3.5.3 The gap statistic

Above, W has been subscripted with "K," the number of clusters, to emphasize that the k-means and k-mediods algorithms require that the number of clusters in the data be pre-specified. This is an obvious limitation in situations like ours, in which the number of clusters (i.e. "types" of anomalous events passing through stage 2) is a priori unknown, and is itself a major goal of the exploratory analysis.

W_k is a monotonically decreasing function of k; this can be intuited by considering the progression to the extreme case, in which each observation constitutes its own cluster and hence W_k is equal to zero. Plots of W_k versus k often show a character-

istic "elbow point," a value for k at which previously large decreases in the function become markedly flattened. Traditionally, this point has been taken to represent the transition between decreases that represent an evolution toward the data's true underlying structure and decreases that are simply attributable to the inevitably greater compactness of clusters that can be achieved by adding more cluster centroids. This elbow value of k is thus typically taken as the "optimal" number of clusters.

Tibshirani et al [74] address the issue of estimating the optimal number of clusters in a data set given a similarity metric and error measure, and present a statistical procedure to formalize the elbow heuristic concept described above. Their idea is to compare the graph of $\log(W_k)$ versus k with its expectation, $E_I^*[\log(W_k)]$, under a suitable null reference distribution of the data. They define the gap statistic for sample size I, $G_i(k)$, as:

$$G_I(k) = E_I^*[\log(W_k)] - \log(W_k) \tag{3.22}$$

and propose to choose the optimal k as that for which the gap statistic is largest after taking its sampling distribution into account. The null model assumed is a single component model, and the statistical procedure is designed to reject it in favor of a k-component model $(k > 1)$, if the strongest evidence for any such k under consideration merits it. The choice of the number of k's over which to probe

is of course still up to the experimenter. As recommended by Tibshirani et al., a uniform distribution over a box aligned with the principal components of the data is used as the form of the single-component reference distribution.

The clustering procedure begins with the data set $\{\hat{\mathbf{F}}_1, \ldots, \hat{\mathbf{F}}_I\}$ – the reduced feature-vector representations of anomalies 1 through I that have passed through stage 2. For number-of-cluster values $k = 1$ through 20, the k-mediods algorithm is run using 10 random initializations, and the run that yields the lowest $\log(W_k)$ is chosen for each k. Then, again for each k, an expected value for $\log(W_k)$ under the null reference is computed, using 20 reference data sets for each value of k. A given reference set is created by performing a PCA rotation of the data, sampling I vectors uniformly over the range of values along each dimension, and then back-rotating the samples into the original coordinate frame. Each reference set is then clustered in an identical manner to the true data set, and the expected value for each $\log(W_k)$ is taken as the average over the reference sets. In accordance with Tibshirani et al., a corrected standard deviation, s_k, that accounts for the simulation error is then computed as:

$$s_k = sd_k \sqrt{1 + \frac{1}{B}} \tag{3.23}$$

where sd_k is the standard deviation and B is the number of reference sets, in our case 20. Finally, again following Tibshirani et al., we take as the optimal clustering

of the data that which meets the following criterion:

$$\hat{k} = \min\{k \,|\, Gap(k) \geq Gap(k+1) - s_{k+1}\} \qquad (3.24)$$

where $\hat{k}$ is the estimated optimal number of clusters. (If the above set were empty, $\hat{k}$ would be set equal to the largest value tested, which in our case is 20).

3.4   Results

The number of channels and total number of iEEG hours processed per subject are shown in table 3.1. Across all eleven subjects, a total of 1,423,741 HFO candidates were detected in stage 1. The summed duration of all stage 1 detections was 13.7 hours, or 0.04% of the total data processed, giving an indication of the rarity of stage 1 detections. Among these stage 1 candidates, 290,273 ($\sim$20%) were flagged as anomalous events in stage 2 and passed on to stage 3 for clustering. To verify that the stage 2 anomaly detection algorithm was not performing the trivial task of taking a random subsample of post stage 1 events, we tested the null hypothesis that the proportion of events retained after stage 2 was the same for all patients, finding, as we expected, evidence to reject it ($\chi^2(10, N\!=\!1,423,741) = 57,088.23$, $p <\!\!<\!\!< 0.0001$).

Four clusters were found in stage 3 by the k-mediod/gap-statistic method. The

Table 3.1. Raw data summary

Subject	#Channels	Total iEEG Hrs
CT 01	144	1176
CT 02	128	1877
SZ 01	45	480
SZ 02	126	1890
SZ 03	69	977
SZ 04	90	2122
SZ 05	84	5460
SZ 06	115	10139
SZ 07	86	5615
SZ 08	104	812
SZ 09	110	735
11 subjects	**(1101)**	**(31283)**

Total iEEG hours is the sum over all channels of
the length of time recorded for each channel.

results of clustering are illustrated in figures 3.6 and 3.7. Figure 3.6A shows a clustering of 1,000 events selected randomly across all patients, plotted in the space defined by $\hat{f}_1, \hat{f}_2$, and $\hat{f}_3$ and colored according to class label. Figure 3.6B gives an alternative view of the clustering, projecting the points onto the planes defined by all possible pairs of principle components. Distinct boundaries are apparent between the red, the green, and the blue clusters taken together, while the separation between the light and dark blue clusters is less marked.

Figure 3.7 shows the cluster prototypes for each of the four clusters found by applying our algorithm to all 290,273 events passing through stage 2. The prototypes are found by computing the class members nearest (in the Mahalanobis sense) to the cluster mediods in feature space. In figure 3.7A, both the raw (top) and the
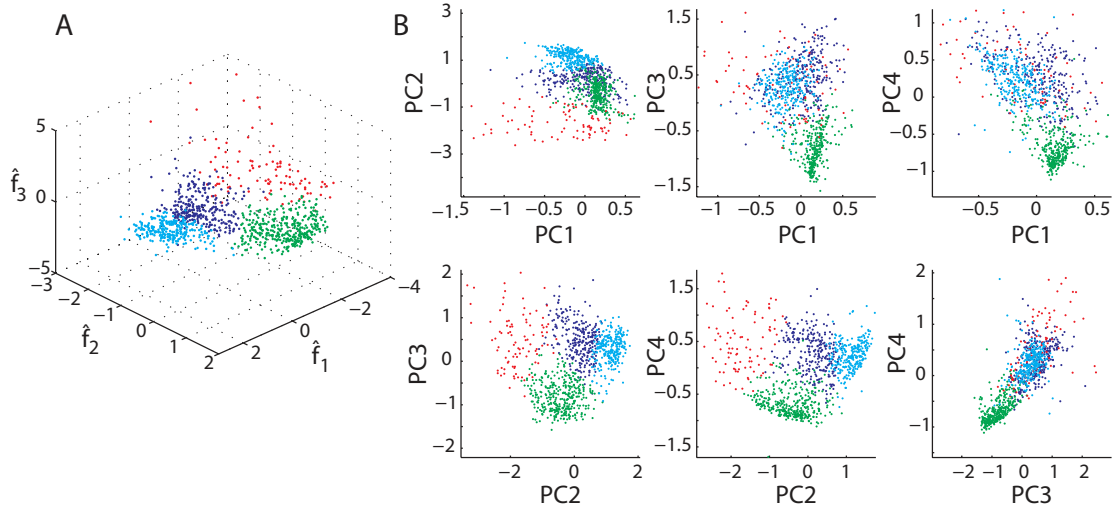
Figure 3.6. (A) Clustering of 1000 randomly selected events (across all patients) passing through stage 2, in the three dimensional space defined by $\hat{f}_1, \hat{f}_2$, and $\hat{f}_3$. Each of the four clusters found by the algorithm is colored uniquely. (B) The points on the left projected on to all possible pairs of principal components; axes are scaled to enable better visualization of cluster boundaries, so a small number of outlying points are not visible.

100-500 Hz bandpass filtered (bottom) detection for each prototype are displayed.

In figure 3.7B, we show five randomly selected members (bandpassed version only) from each cluster.

Qualitatively speaking, cluster 1 ($n = 70,671$) is comprised of irregular, mixed-frequency events; cluster 2 ($n = 92,744$) is comprised of sharp-rising transients that appear to be artifacts (i.e. whose oscillatory characteristics come from filtering the sharp rise); cluster 3 ($n = 38,945$) is comprised of fast, regular waveforms of relatively pure frequency; and cluster 4 ($n = 87,913$) is comprised of slow, regular waveforms of relatively pure frequency.
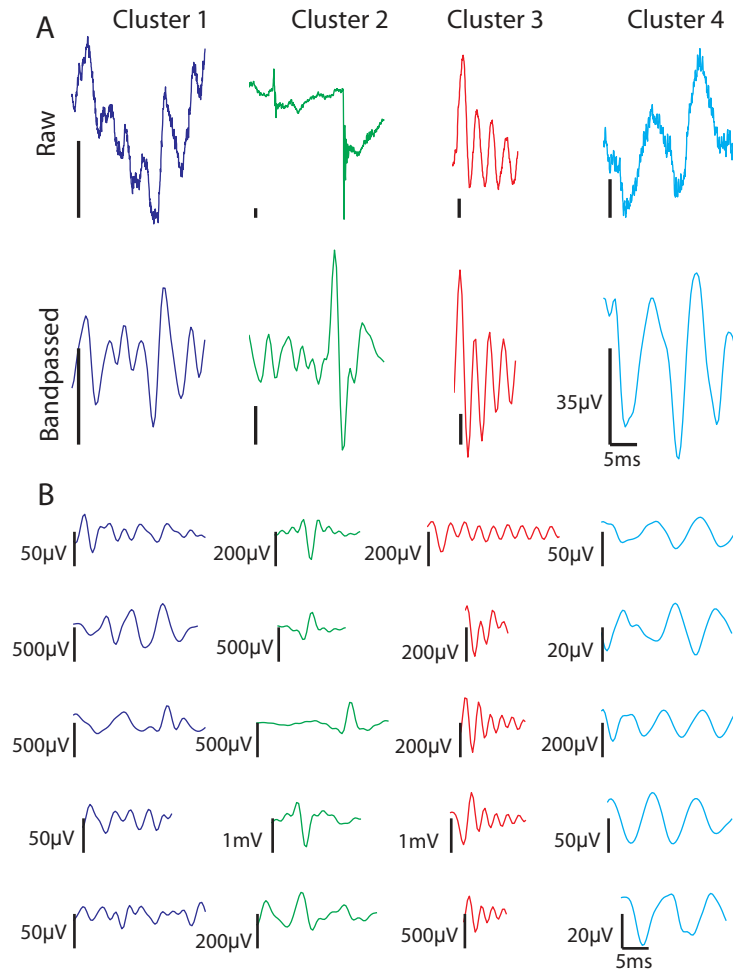
Figure 3.7. Results from clustering all 290,273 events, across all patients, that passed through stage 2. (A) Cluster prototypes found by taking the cluster member nearest in the Mahalanobis sense to the cluster mediod in feature space. Top images are unfiltered waveforms; bottom images are the corresponding 100-500 Hz bandpass filtered waveforms. Vertical scale bars are 35 μV in all cases. (B) Five randomly selected events from each cluster (bandpassed waveforms only). In both A and B, events are truncated to 25 ms, if necessary, to put all waveforms on the same time scale for easier comparison.

To investigate the stability of the clustering algorithm and test whether the aggregated clustering results we show in figures 3.6 and 3.7 were heavily influenced by individual subjects, we performed two additional experiments.

In the first, we applied the clustering algorithm to individual subjects, clustering 100 random samples drawn from the set of all events passing through stage 2 for each subject. The size of each of the 100 random samples was 25% of all the subject's post stage 2 events. Results are shown in figure 3.8, which contains one panel for each patient: a histogram summarizing the outcome of the 100-iteration sampling experiment. The lower right panel of the figure represents the results of a similar random sampling experiment performed on the data aggregated across subjects, with the only difference being that sample sizes of 5% (15,000 of all 290,273 events passing through stage 2 across all subjects) were used, due to the computational expense of clustering twenty-five percent one hundred times. Each individual subject's histogram is colored in proportion to the relative number of events passing through stage 2 for that subject. Thus, histogram color reflects the weight of the subject's contribution to each of the 100 random samples represented in the aggregate histogram in the lower right panel. The number in red indicates the mode of the distribution, which can be interpreted as the number of clusters the algorithm yields for that patient. The degree of concentration of the distribution for a given individual reflects the stability of the algorithm for that subject, which is a function

of the variability and shape of the subject's events in feature space, as well as the sample size clustered.

In the six of eleven subjects for whom the algorithm yielded four or fewer clusters, the clusters generally appeared to be subsets of those found in the aggregated data. In the remaining five subjects, for whom there were greater than four clusters, the additional clusters generally appeared to have resulted from over-splitting some combination of the four clusters found in the aggregated data. These trends are illustrated in figure 3.9, which shows examples of prototypes derived from clustering individual subject data. Prototypes in figure 3.9 are given a color corresponding to the aggregate prototype to which they bear the closest visual resemblance (cf figure 3.7A, bottom set of waveforms).

Subject SZ 01, for whom the algorithm yields strong evidence for three clusters, appears to exhibit the artifact, mixed, and slow oscillation clusters, but not the fast oscillations of the aggregated clustering. Subject SZ 08, for whom the algorithm weakly suggests two (over three) clusters, exhibits the artifact and slow oscillation clusters prominently. Subject SZ 03, for whom the algorithm performs relatively poorly, also appears to have clusters that are a subset of those found in the aggregate - namely, the mixed frequency and slow oscillation clusters - but they are often highly split by the algorithm. Figure 3.9 shows one particular iteration in which six
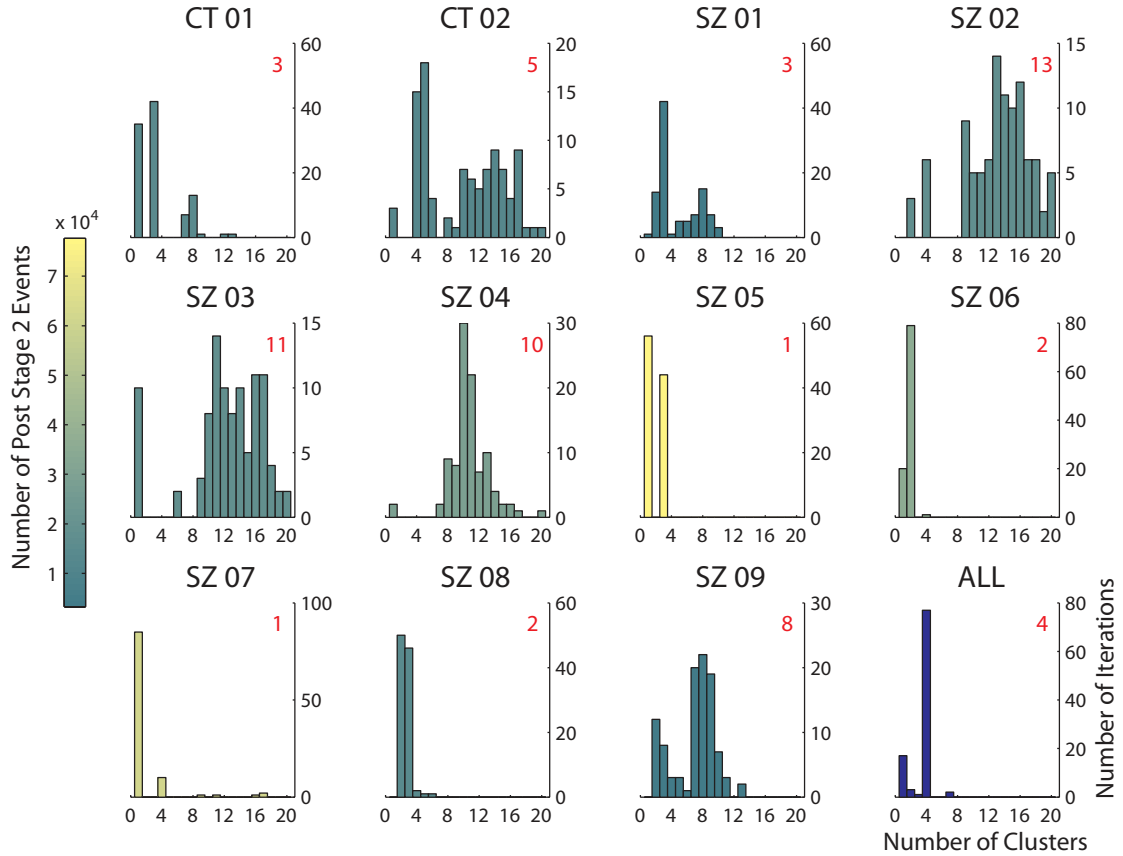
Figure 3.8. Individual subject clustering. Each panel (except the lower rightmost) gives the results, for a single subject, of an experiment in which 100 random samples of size equal to 25% of all events passing through stage 2 for the subject were clustered. Number of clusters found by the algorithm is plotted on the x-axis. Number of iterations is plotted on the y-axis. Show in red is the mode of each distribution, which can be interpreted as the number of clusters found by the algorithm for that subject. Histograms are colored in proportion to the relative number of events passing through stage 2 for each patient. The lower rightmost panel gives the results of a similar sampling experiment performed on the aggregated data, the only difference being that random samples of size 5% of the total were used, instead of 25%, to ease the computational burden.
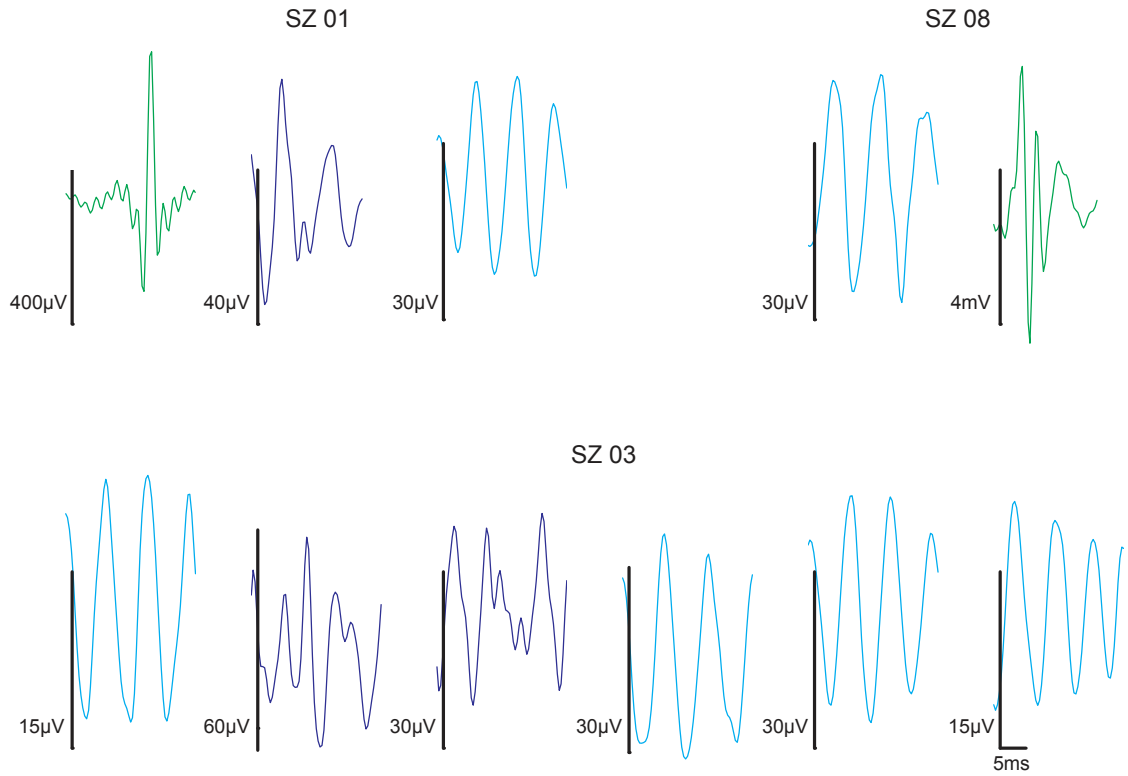
Figure 3.9. Example prototypes from individual subject clustering; 100-500 Hz band-passed waveforms only. Events are truncated to 25 ms, if necessary, to put all waveforms on the same time scale for easier comparison. Subjects SZ 01 and SZ 08 show clusters that appear to be subsets of the four clusters found in the aggregate. Subject SZ 03 does as well, but in this case it appears that two groups have been oversplit by the algorithm, yielding six clusters. The latter type of cluster stability issue was not seen in the aggregate clustering.

clusters were obtained from what appears to be these two.

To quantify these observations about the stability of the algorithm, we use Rand's c-statistic [60] (Rand's index), a common measure of agreement between two partitionings of a data set. Rand's index is simply the fraction of all pairs of observations that are treated similarly in both partitionings, where a "similarly treated" pair is one whose members either are in the same cluster or in different clusters in both partitionings:

$$
c = \frac{\binom{N}{2} - \left[\frac{1}{2} \left\{ \sum_{k=1}^{K} \left(\sum_{k'=1}^{K'} n_{kk'} \right)^2 + \sum_{k'=1}^{K'} \left(\sum_{k=1}^{K} n_{kk'} \right)^2 \right\} - \sum_{k=1}^{K} \sum_{k'=1}^{K'} n_{kk'}^2 \right]}{\binom{N}{2}}
$$
(3.25)

K is the number of clusters in the first partitioning, K' is the number of clusters in the second partitioning, and $n_{kk'}$ is the number of observations that were simultaneously in cluster k in the first partitioning and cluster k' in the second partitioning. Being a proportion, Rand's index varies between zero, when no pairs are treated similarly by the two clusterings, and one, when the clusterings are identical.

We were interested in how similar one partitioning was to another, within a subject, for iterations on which the algorithm returned more than a single cluster. Because each of the 100 clusterings for a given subject was based on an independent ran-

dom sample, the samples upon which any two clusterings were based almost always contained some non-overlapping observations. As the Rand index has nothing to say with regard to such unique observations, we considered only the intersection of the two clustered samples. For each individual subject, as well as for the aggregated clustering, we computed Rand's index for all pairs of clusterings in which both numbers of clusters were greater than 1. Results are shown in table 3.2. For completeness, we also show the results when the condition $(K, K') > 1$ is removed[9].

The high values of $\bar{c}$ in the second column of table 3.2 (and the accompanying low standard deviations in column 3) indicate that in cases where the algorithm returns a number of clusters greater than 1 on different iterations, the clusterings tend to have considerable agreement. This is consistent with the idea that the number-of-cluster-variability for individual subjects comes largely from the bulk splitting and merging of the clusters that were found in the aggregate[10][11].

---

[9]The latter result would be more interesting if, in cases where more than one cluster was found, the algorithm had a tendency to produce one very large cluster along with several very small clusters. However, since this was not our observation we felt the $(K, K') > 1$ case was most relevant.

[10]The results in table 3.2 should not be overstated. For example, one would not take seriously any multi-cluster result for subject SZ 05 despite its "high marks" in columns two and three of the table because the preponderance of its clustering results, as seen in figure 3.8, point toward a single cluster. (The low value for **n** in column three and the data in the last two columns of table 3.2 serve as a reminder of this.)

[11]As an alternative to the Rand index, which ranges from zero to one, one can compute the adjusted Rand index [38]. The latter can assume a larger range of values and is zero when the agreement between two partitionings is equal to its expectation under a chance model (i.e. each partitioning is picked at random from the set of all partitionings having the same number of clusters and members per cluster as the original). The average adjusted Rand indices for $(K, K') > 1$ in our case take values from $0.5 - 0.9$, from which it can be concluded that the similarities observed are substantially greater than would be expected by chance.

Table 3.2. Rand's c-statistic

Subject	$\bar{c}$	sd	n	$\bar{s}$	$\bar{c}_{(n=4950)}$	$sd_{(n=4950)}$
CT 01	0.843	0.114	2080	923.280	0.610	0.314
CT 02	0.873	0.060	4656	787.340	0.830	0.182
SZ 01	0.787	0.129	4851	203.597	0.778	0.146
SZ 02	0.893	0.092	4950	1102.662	0.894	0.092
SZ 03	0.936	0.020	4005	1139.078	0.782	0.329
SZ 04	0.923	0.024	4753	1825.557	0.891	0.159
SZ 05	0.995	0.002	946	4856.309	0.670	0.329
SZ 06	0.935	0.054	3160	2266.261	0.797	0.209
SZ 07	0.895	0.085	105	3919.238	0.793	0.350
SZ 08	0.934	0.054	4950	885.499	0.934	0.054
SZ 09	0.843	0.114	4950	240.703	0.843	0.115
ALL	0.950	0.074	3403	774.617	0.760	0.311

Rand's c-statistic (Rand's index) for all pairs of clusterings in which $(K, K') > 1$ (columns 2-5). Shown for each subject are the mean $\bar{c}$, standard deviation **sd**, size of the sample (i.e. number of clustering pairs) upon which the mean was based, **n**, and the average number of overlapping data observations in each pair of samples, $\bar{s}$. A value of 1 for any given $\bar{c}$ would indicate a perfect match between all **n** pairs of clusterings. Columns 6 and 7, included for reference, give the mean value for the Rand index when all 100-choose-2 (4950) pairs of clusterings are considered and its standard deviation, respectively. The average number of overlapping data observations in this case is so close to the value in column 5 (i.e. $\pm$ 1 observation, typically less) for all subjects that we have chosen not to replicate it.

Despite the expectedly variable stability on the individual subject data apparent in figure 3.8, the algorithm performs robustly on the aggregated data (bottom right panel), as well as on most patients for whom a relatively large number of events were clustered. Specifically, seventy-seven percent of the clustering iterations performed on the aggregated data yielded four clusters. Columns two and three of table 3.2 (last row) show that, furthermore, these 4-cluster partitionings

were highly similar to one another. Figure 3.10 shows that Shannon entropy[12] deceases with sample size for the aggregated data. There is a negative linear relationship (over the range of sample sizes considered) between the Shannon entropy of the distribution of number of clusters returned by the algorithm and sample size ($\beta = -5.71\text{x}10^{-5}, t(6) = -4.44, p = 0.004; r^2 = 0.77$). Distributions were formed at each sample size by making 100 random draws from all post-stage 2 detections. This result provides further quantitative evidence that the stability of the algorithm improves with the addition of data.
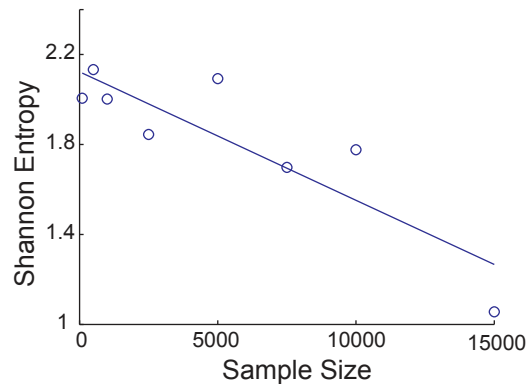
Figure 3.10. Scatter plot of sample size versus the Shannon entropy of the distribution of the number of clusters returned by the algorithm for the aggregated data. Distributions were formed at each sample size by making 100 random draws from all post-stage 2 detections. Equation for the regression line is $y = -5.71\text{x}10^{-5}x + 2.12$.

In the second experiment, to examine the influences of individual subjects on the aggregate result we performed "leave-one-out" clustering. Each subject, in turn, had his/her events removed from the pool available for clustering, and 5,000 events were

---

[12]Shannon entropy, S, is defined as $S = -\sum_i p_i \log_2 p_i$, where p_i is the probability of being in state i and $P \log_2 P := 0$ when $P = 0$.

randomly drawn without replacement from the remaining events, in 25 iterations. Figure 3.11 shows the results summarized as a bar plot in which the height of each bar represents the mode of the distribution of cluster numbers for the 25 iterations.
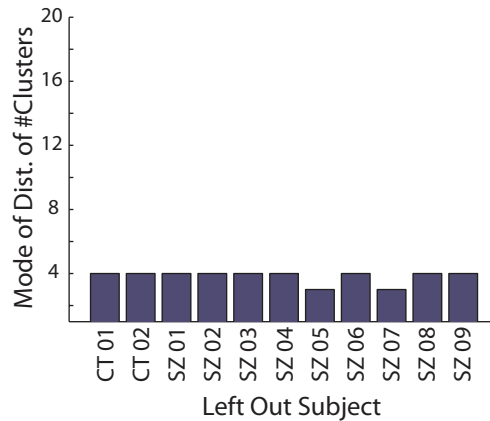
Figure 3.11. Results of an experiment in which each subject, in turn, was left out of the clustering. Five thousand randomly selected events from all those passing through stage 2 in the remaining patients were then clustered in 25 iterations. The x-axis shows the subject that was left out. The y-axis shows the mode of the distribution of number of clusters found in each of the 25 iterations.

Only patients 5 and 7 altered the result from the in toto result of four clusters, in both cases bringing the number of clusters to three, suggesting that these two patients were the predominant contributors to at least one of the four clusters. Indeed, subjects 5 and 7 account for 73% of all events in cluster 3, whereas they make an average combined contribution of only 45% in clusters 1, 2, and 4 (and 48% to the total number of events clustered).

Taken collectively, the results of both the individual patient and aggregated clustering experiments are objective evidence in support of the idea that human neocortex

is capable of generating *distinct classes* of transient oscillations within the 100-500 Hz frequency band, and that there are three salient classes (excluding the putative artifact class).

3.5 Discussion

3.5.1 Main contributions

The first contribution of this work is automation. We have introduced a new methodology, based on prior work, for automatically detecting and clustering high-frequency oscillations on the human intracranial electroencephalogram. We believe this is significant for several reasons. Perhaps the most pedestrian, but nonetheless valuable, is the fact that it decreases human labor substantially. Researchers have reported that it can take up to three hours to mark HFOs in a single channel of EEG that is just one minute in duration [76]. Assuming a generous per-channel marking rate of 1 hour per 10 minutes of data, it would take a superhuman reviewer (who did nothing but mark EEG all day at the same rate for eight continuous hours, never taking a weekend or holiday) over sixty-four years to manually mark the quantity of data processed for this study.

Though the number of subjects in our sample is comparable to many previous stud-

ies of HFOs in epilepsy, the total amount of iEEG data we analyzed, to the best of our knowledge, is vastly larger. Given the rarity of events, analyzing similarly large databases of iEEG - ideally with an even greater number of subjects - will, in our opinion, be required to make definitive statistical statements about the relationship between HFOs and seizure generation going forward. We consider the present effort a step in the direction of such large-scale analyses.

More importantly, in our automated approach we have done away with all data pre-selection. One could argue that this is not as beneficial as we would like to believe. After all, reducing the human workload is not the only motivation for performing data pre-selection; one might claim, for example, that choosing channels in the appropriate brain regions or epochs during certain states of wakefulness is critical to the interpretability of findings, or that interesting results will be entirely masked in the absence of these steps. But in fact, increasingly, researchers are finding exactly the opposite: that many of the pre-selection steps previously thought to be critical are most likely not. Below, we list just a few examples of beliefs that have restricted past HFO analyses but are no longer as widely held (in parentheses, we list examples of studies that have cast doubt on the original belief): 1) that HFOs occur only in hippocampal and parahippocampal areas (they also are seen in neocortex [33, 34, 42]); 2) that HFOs occur only during periods of slow-wave sleep (they also are observed during the active states of waking and REM sleep [3, 33]);

3) that HFOs occur only interictally (they are observed ictally [42] and preictally [40] as well; 4) that HFOs can only be recorded with microelectrodes (they can also be recorded with standard clinical macroelctrodes [42], albeit perhaps with less fidelity [83]).

Moreover, our aim is precisely to investigate the generalizability of some of the enticing findings that have recently been reported in the study of HFOs in human epileptic brain. We are concerned with how, if at all, conclusions about the practical utility of HFOs for localization of epileptogenic regions are modified when the many conditions implied by stringent data pre-selection are removed. If the answer turns out to be "not much," this will be particularly encouraging for the prospects of using HFOs as control signals on board implantable seizure therapy devices, for example, given that these devices, at least in early generations, will not have the same luxuries as human researchers when it comes to pre-identifying optimal data for processing.

The second major contribution of this work is that we have provided objective evidence for the existence, in the neocortex of epilepsy and control patients, of *distinct classes* of transient oscillations in the range between 100 and 500 Hz. We do not think this is splitting hairs. Most studies of HFOs in epilepsy of which we are aware simply presuppose the existence of discrete entities typically termed "ripples" and

"fast ripples" (albeit with experimental evidence, usually of the observational variety, for doing so). And at least one group has reported that their data did not clearly show distinct ripple and fast ripple modes [83]. One study that attempted to quantitatively verify the existence of discrete populations [68] did so by fitting functions to the distribution of peak frequency values of detected events, and arguing that the bimodality of the best fitting function was telling. The authors acknowledged that their analysis ignored the possibility of events which themselves had bimodal or more complex frequency spectra. Here, we show evidence that there actually is a cluster of mixed-frequency events, and furthermore that it appears to be separate from two other clusters that are comprised more purely of a single frequency.

That said, we have not explicitly demonstrated that the mixed-frequency class we observed (or, technically speaking, any of the other classes for that matter) reflects "real physiology." Indeed, a waveform with a similar time domain characteristic to our mixed-frequency class prototype could theoretically be generated by applying a 100-500 Hz bandpass filter to a white Gaussian noise burst, for example. But having examined the spectra of these mixed frequency events and often observed two prominent peaks, we feel confident suggesting that they should be further investigated as a legitimately hybrid class of oscillations. As for the other three clusters, we believe, as mentioned, that one is comprised largely of clinically uninteresting recording artifacts. The other two bear sufficient resemblance to what have been

labeled ripples and fast ripples to persuade us that we are recording the oscillations that have been recognized by other researchers in the community.

3.5.2   Conclusion

A flurry of recent papers suggests that the relationship between subclasses of high-frequency oscillations recorded using nonstandard, high bandwidth data acquisition systems could be a more reliable indicator of seizure-generating cortex than the waveform morphologies typically used by clinicians for seizure localization[13]. In this paper, we introduced an algorithm for detecting and classifying these signals in an automatic fashion and demonstrated the tractability of exploring a data set of unprecedented volume using the method. Using an unsupervised approach that did not presuppose a specific number of clusters in the data, we showed direct evidence for the existence of distinct classes of oscillations within the 100 to 500 Hz frequency range in a population of neocortical epilepsy patients and controls, addressing a current controversy in the field. The number of classes we found, four (three plus one artifact cluster), is consistent with prior studies that distinguish between ripple and fast ripple oscillations, and additionally suggests that a class comprised of hybrid ripple/fast-ripple events exists.

---

[13]HFOs are not, by default, part of the classical armamentarium of electrographic signatures used for localization because they are difficult or impossible to visualize in standard clinical iEEG due to their short timescales, small relative amplitudes, and spectral contents that can exceed the bandwidth that is traditionally recorded.

By minimizing selection bias and other human judgment in detecting and classifying HFOs, we believe that it will be possible to answer the important question of whether they can be used to delineate epileptogenic brain. This work also demonstrates the tractability of using machine learning techniques to analyze large streams of high resolution neurophysiologic data, which may become standard in the future as new devices and techniques for mapping and modulating function and dysfunction in neurological disorders, on multiple brain scales, evolve.

Chapter 4

Validation

4.1 Summary

Transient field-potential oscillations in the range of 100-500 Hz have been recorded from both normal and abnormal brain tissue in animals and humans. In epilepsy research, it is hypothesized that subclasses of these high-frequency oscillations (HFOs) are biomarkers for pathologic tissue. Understanding the potential relationship between HFOs and seizure generation requires the development of reliable automated tools for extracting HFOs from long continuous intracranial electroencephalographic records. Several authors have proposed automated HFO detectors, but none have designed them for use without stringent manual data pre-selection and post-processing, making conclusions difficult to generalize. Building upon a state-of-the-art detector, we developed an HFO classification algorithm to eliminate

the need for manual intervention. In this chapter, we compare the performance of our automated method with that of three expert human reviewers on an HFO verification task. We conclude that human reviewers are currently not in sufficient agreement about what constitutes an HFO to place high emphasis on ground truth data in detection benchmarking; and we find that our automated approach is statistically indistinguishable from humans in the classification task.

4.2   Introduction

Despite the proliferation of tools for automatically detecting seizures and other epileptiform activity, no algorithm yet exists for the fully automated extraction of 100-500 Hz transient high-frequency oscillations (HFOs) from intracranial EEG recordings. Several authors have reported on *semi-automated* approaches [20, 21, 22, 68] to HFO detection, which use intensive visual pre- and post- processing in conjunction with machine detection. As Gardner et al. discuss [32], none of these groups presents formal validation data for their automated methods; acceptable detection performance is either implicit or simply asserted. Staba et al. [68], for example, state without demonstration that "during development of [their] technique, it was found that it was effective in detecting greater than 84% of putative oscillatory events observable with visual EEG analysis."

4.2.1   Fool's Gold?: "Gold standards" for HFO detection

It is worth asking how disconcerted one should be about the scarcity of formal validation data for published machine HFO detection algorithms. Below we argue that, given the current state of the field, it would be misguided to place too strong an emphasis on classical performance metrics for existing or proposed automated detectors. As justification for this opinion, we offer the following three points. First, even in the absence of rigorous direct validation, current methods are undoubtedly *useful*. "Exhibit A" is their widespread acceptance: several research groups [20, 64, 67, 68, 83] have adopted methods similar to those originally presented without formal validation by Csicsvari et al [21], for instance, and have done so with success – where "success" means simply that detected events turned out to be related to outcome measures of scientific interest.

Second, as the results of Gardner et al [32] show, the task of having clinicians visually *identify* transient oscillations in iEEG – a requirement for generating ground truth data against which to evaluate automated methods – does not yield the complete set of events that is *verified* by the same reviewers when presented with a superset of their own markings containing those of a machine detector as well. Though the Gardner study involved oscillations in the gamma range, it seems likely that their conclusion that human reviewers make many false negative errors would apply to 100-500 Hz HFOs as well. In fact, one might expect the effect to be even

larger, both because HFOs are less familiar to clinical reviewers and because vigilance requirements for marking are even more demanding given the higher recording bandwidths. Findings like those of Gardner et al. call into question the sensibility of using a set of human markings as an absolutely rigid benchmark for automated detectors.

Lastly, insisting on perfect establishment of ground truth tends to raise the existential question "what is an HFO?" – a query that we find rings hollow. HFOs are, after all, essentially human constructs: categorizations we impose upon observed neurophysiological phenomena because they aid our understanding or ability to communicate about the workings of the brain. Far more critical than pinning down exactly what an HFO *is* is understanding what the concept embodies. Taking this tact lets us move the study of HFOs from the ill-posed realm of trying to extract unknown (or at least very difficult to agree upon) signals from a background noise whose statistical description is almost equally elusive, to the more tractable realm of exploratory data analysis driven by our ideas about what characterizes HFOs. A typical experimental question such as "do 'fast ripples' increase [in number] in the seizure onset zone" thus decomposes into several: Do we find evidence for transient oscillatory bursts at all? If so, is there evidence that they form well-separated populations in some frequency space? If so, do the populations characterized by higher frequencies tend to increase in seizure onset areas?

This empirical approach skirts the perilous business of first having to adjudge whether a particular waveform is worthy of being deemed an HFO – or worse, whether that waveform is a "ripple" or a "fast ripple" before we can hope to study anything. Instead, we settle for a crude detector – one that has been vetted by clinical opinion for its ability to find at least *some* things resembling what would catch the eyes of human reviewers – and we analyze its imperfect outputs. If we find results of clinical importance, we can use them to refine our understanding of what the critical properties of HFOs are and subsequently to optimize our detector, in the hope that we will extract more, or different, information in the next iteration. This evolutionary view of detector design is fundamental to the approach we have taken in this thesis.

What we should be asking of the earliest incarnations of fully automated HFO processing methods, then, is not whether they meet premature and arbitrary performance specifications for detection, but whether they can approximate the successes of semi-automated methods *without data pre-selection and post-processing by humans.* The latter limit the scientific interpretability of conclusions about HFOs and seizure generation – including the ability to assess the generalizability and clinical utility of those findings – to a far greater degree than the odd percentage point of sensitivity or specificity.

4.2.2 The present study

At the same time we offer this lengthy caveat, we also appreciate that it is helpful for practitioners who wish to evaluate our algorithm to understand it within the context of human performance. In the work we describe below, we asked three board-certified epileptologists to classify detected HFO candidates and compared their markings with the outputs of our automated classifier.

4.3 Methods

4.3.1 Reviewer labeling

Five thousand HFO candidates ($\sim$0.4%) were randomly selected among all those identified in stage 1, across all patients. As a conservative measure, we excluded all events from the two subjects' (CT 02 and SZ 05) files from which a subset of data had been used to develop artifact-distinguishing features. The remaining 4,773 randomly selected events, across ten patients (nine epilepsy and one control)[1], were then used in a human labeling experiment.

---

[1]For CT 02, 1 of 1 data file was used in artifact training; for SZ 05, only 1 of 8 total files was used in training. Thus, the number of patients from which events were drawn for the labeling experiment is only one less than the total number of patients, not two.

Three board-certified neurologists independently marked all presented events as either valid (positive) or invalid (negative) HFOs, according to the following criteria for what constitutes a valid HFO:

> *"Any transient, quasi-periodic voltage variation with predominant frequency between 100 and 500 Hz, lasting on the order of tens of milliseconds, standing prominently apart from the background signal, and having apparently physiologic origin."*

The criteria were intentionally somewhat vague to reflect the fact that there is currently no standard operational definition of an HFO.

Events were presented to reviewers via a custom Matlab graphical user interface (GUI), shown in Figure 4.1.

The GUI was comprised of four complementary views of each HFO candidate. The bottom view displayed roughly 1 second (0.5 seconds on either side of the candidate) of 5 Hz–1 kHz[2] single channel iEEG, sampled at 2,713 Hz, with vertical scaling of 21 μV/mm. The event under consideration was delimited by red lines (solid, start; dotted, stop) and the view could be scrolled for 30 seconds on either side of the default display window. The top, left view was of the raw data (near DC—9 kHz,

---

[2]Display distortion was in practice negligible at this default timescale due to the relatively low signal power above the effective Nyquist frequency of the display. As reviewers were free to zoom in (but not out), this compromise allowed us to faithfully represent the full bandwidth across nearly all available time scales without the need for zoom-adaptive filtering.
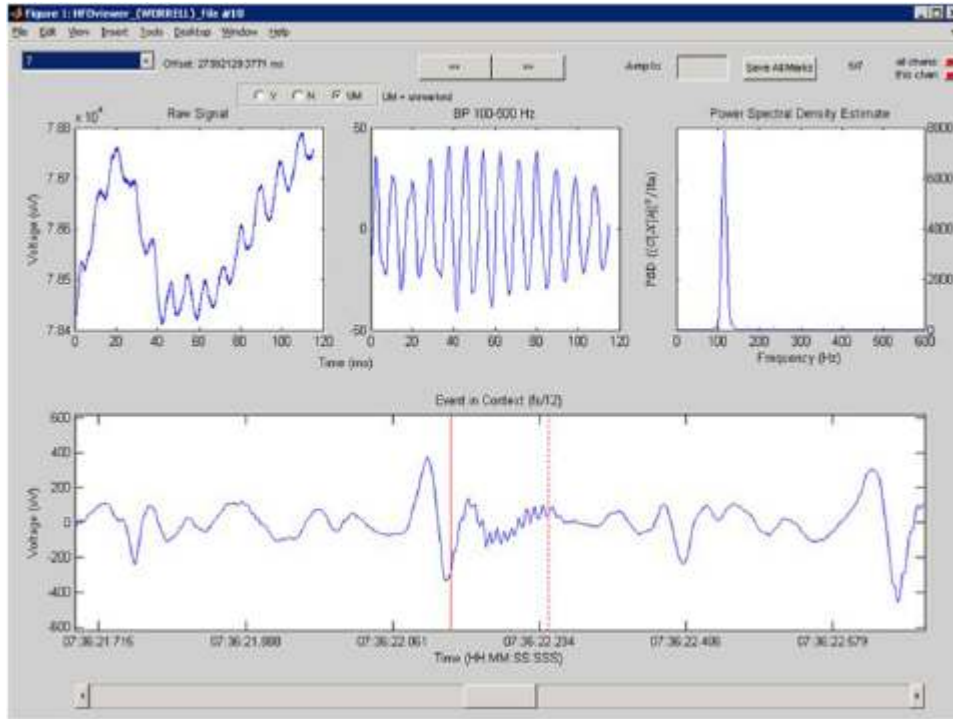
Figure 4.1. Screen shot of custom GUI used to present detections to clinical reviewers.

32,556 Hz sampling rate) corresponding to the detection; the top, middle view was

of the bandpass data corresponding to the detection (100-500 Hz, 2,713 Hz sampling

rate); and the top, right view was a frequency-domain representation of the middle

view. Unlike the bottom view whose vertical scaling was fixed, all top views were

auto-scaled to fit their viewing windows. Reviewers were free to edit their markings

until they had labeled every event and declared the task complete.

The human labeling task was binary, while the automated algorithm classified de-

tections into one of five groups: four clusters, plus a fifth group ("Cluster 0")

comprised of detections that were eliminated in stage 2. In order to compare hu-

man and machine performance directly, we took as machine-negative all events in cluster 0 and in cluster 2, whose centroid bore the closest qualitative resemblance to the artifacts we had designed features to identify. All other clusters were taken as machine-positive. This post-hoc labeling decision was made blinded to the human reviewers' markings; and while made manually for the present experiment, we note that it could readily be made automatically in the future if desired – for example by storing the coordinates of the cluster 2 centroid and assigning the negative HFO label to an automatic cluster whose centroid was sufficiently nearby.

4.3.2 Statistical analyses

We use the chi-squared test of homogeneity to test whether HFO counts are distributed identically across populations (where "population" is analysis-dependent), and the chi-squared test of independence to test whether marker labels are independent. The chance model we use for markers assigns a positive label to each event with probability $p = N_p/N$, with N_p the total number of events actually labeled positive by the marker and N the total number of marked events (4,773).

4.4 Results

In describing the results below, we use the term "reviewer" to refer specifically to humans and "marker," more generally, as a term that encompasses both humans and the machine algorithm. The terms "detection" and "event" are used synonymously.

4.4.1 Putative prevalence of valid HFOs by marker

Human reviewers were not in agreement about the overall prevalence of HFOs in the data set of candidates presented to them. The percentages of detections marked as positive HFOs by reviewers A, B, C, and the machine classifier (M) were 24.6%, 5.5%, 11.5%, and 13.0%, respectively. We rejected the null hypothesis that the proportion of detections marked as positive was independent of human reviewer ($\chi^2(2, N = 14,319) = 763.84, p \ll 0.0001$). Also apparent from these numbers is that the automated method's propensity to mark events as positive is not extreme relative to humans'.

4.4.2 Human reviewer preference by cluster

Reviewers had clear and differing cluster preferences. Figure 4.2A shows, for each human reviewer, all events falling into clusters 1-4 that were classified as positive

HFOs. For reviewer A, the majority of such detections (57.1%) fell into cluster 4. The largest clusters for Reviewer B were 3 and 4, with the former (44.3%) favored over the latter (27.1%). Reviewer C displayed yet a third pattern, splitting a majority fairly evenly between clusters 1 (42.0%) and 4 (41.4%). For all three reviewers, the smallest percentage was in cluster 2 (6.1%, 13.6%, and 1.9%, for A, B, and C, respectively), the putative artifact class. We reject the null hypothesis that the proportion of detections in each of the four clusters was the same across human reviewers $\chi^2(6, N=563) = 97.40, p \ll 0.0001$).
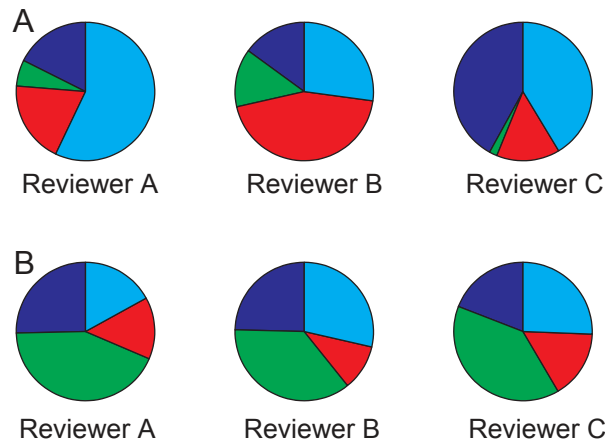
Figure 4.2. Each whole pie represents the total number of positively marked HFOs by a human reviewer (A: reviewer A; B: reviewer B; C: reviewer C) that were classified by the machine as belonging to clusters 1, 2, 3, or 4. Pie wedges represent the proportion of such marks falling into each cluster (blue: cluster 1; green: cluster 2; red: cluster 3; cyan: cluster 4).

Figure 4.2B shows, for each human reviewer, all events falling into clusters 1-4 that were classified as negative HFOs. As expected, the putative artifact cluster dominates for all three reviewers: 43.3%, 36.1%, and 39.3% for reviewers A, B, and C,

respectively. And as is the case for positive labels, we again reject the null hypothesis that the proportion of detections in each of the four clusters was the same across human reviewers ($\chi^2(6, N{=}2197) = 41.47, p \ll 0.0001$). Not shown in figure 2 are events that were marked as "0" by the automated detector – detections that were never classified into clusters 1-4 due to elimination in stage 2. These events are accounted for below, where we give standard performance metrics for the automated classifier against a ground truth set derived from the human reviewers' markings.

4.4.3 Inter-rater agreement

A question of fundamental importance in defining ground truth data is to what degree the independent human reviewers agree amongst themselves regarding what constitutes an HFO and what does not. Table 4.1 gives contingency tables, including the kappa score [19] and percentage agreement, for each of the three human-human marker pairs (left) and each of the three machine-human marker pairs. For all tables, we reject at the 5% significance level the null hypothesis that marker labels were independent, and the kappa values greater than one indicate that differences from chance were in the direction of agreement in all cases (AB: ($\chi^2(1, N{=}4773) = 260.94, p \ll 0.0001$); AC: ($\chi^2(1, N{=}4773) = 25.09, p \ll 0.0001$); BC: ($\chi^2(1, N{=} 4773) = 298.50, p \ll 0.0001$); MA: ($\chi^2(1, N = 4773) = 85.80, p \ll 0.0001$); MB: ($\chi^2(1, N = 4773) = 270.96, p \ll 0.0001$); MC: ($\chi^2(1, N = 4773) = 139.94, p \ll$

0.0001);). The average pairwise percentage agreement among human reviewers was 79%, while that for machine-human pairs was 80%. The average pairwise kappa score among human reviewers was 0.15, while that for machine-human pairs was 0.17. The latter average, however, and the individual kappa scores that comprise it are not straightforward to interpret given the different biases of the reviewers.

Table 4.1. Confusion matrices, all marker pairs

	B						A			
	0.17	1	0	77%			0.12	1	0	73%
A	1	174	1001	1175		M	1	245	374	619
	0	88	3510	3598			0	930	3224	4154
		262	4511	4773				1175	3598	4773

	C						B			
	0.07	1	0	72%			0.21	1	0	87%
A	1	183	992	1175		M	1	121	498	619
	0	367	3231	3598			0	141	4013	4154
		550	4223	4773				262	4511	4773

	C						C			
	0.23	1	0	88%			0.17	1	0	82%
B	1	117	145	262		M	1	159	460	619
	0	433	4078	4511			0	391	3763	4154
		550	4223	4773				550	4223	4773

Left side: human-human pairs; Right side: machine-human pairs. In red in each table are the kappa score (top left) and percentage agreement (top right).

We note that we have aggregated across subjects in computing these inter-rater agreement measures. Given the rarity of positive HFOs, sample sizes were too small to compute reliable statistics on an individual subject basis. But inspecting

kappa scores leads us to hypothesize that the degree of inter-marker agreement and the differences between human-human and machine-human pairs may vary with patient. For example, average human-human kappa for SZ 05 (1627 events) was 0.21 while the machine-human value was 0.27; for SZ 07 (1448 events) average performance was near chance for both human-human (-0.07) and machine-human (0.01) pairs; and for SZ 03 (295 events) average human-human kappa was 0.37, while average machine-human kappa was 0.19. It would be instructive to investigate these differences more systematically by conducting another marking experiment in which larger random samples of equal sizes were drawn from each subject.

The main conclusion we reach is that a given human is no more consistent with another human in his markings than he is with the machine.

4.4.4 HFO ambiguity

Ground truth looks very different depending on which of several plausible defining rules is adopted. 33.7%[3] of all detections were marked by at least one human reviewer as positive HFOs, while 39.6%[4] of all detections were marked by at least one marker as positive. 6.0%[5] of events were marked by at least two human reviewers

---

[3]Chance, which should be higher = 36.9%.
[4]Chance, which should be higher = 45.1%.
[5]Chance, which should be lower = 4.5%.

(i.e. majority consensus) as positive HFOs, while 10.3%[6] of all events were labeled positively by at least two markers. Only 2.0%[7] of events were marked by all three viewers (i.e. unanimous consensus) as valid HFOs, while 2.5%[8] of all events were marked by at least three markers as positive[9]. The range of these values, which is affected by both the marginal probabilities displayed by each marker and the degree to which they tend to actually agree, gives one view of the general uncertainty among reviewers about what counts as an HFO.

4.4.5 General classifier performance metrics

We formed a ground truth data set by labeling as positive all events marked positively by at least two human reviewers (i.e. majority human vote) and as negative all remaining events. The overall accuracy of the automated classifier against this benchmark was 86.7%. Sensitivity was a moderate 46.8%, reflecting the conservatism of Stage 2, which was designed to retain only events with large spectral dissimilarity from the background, a condition not explicitly enforced in the marking instructions for reviewers and to which we anticipated not all would adhere. Specificity was 89.2%, reflecting strong classification performance for negative events.

---

[6]Chance, which should be lower $= 8.7\%$.

[7]Chance, which should be lower $= 0.16\%$.

[8]Chance, which should be lower $= 0.74\%$.

[9]All reported values were significantly different at the 5% level (chi-squared test) from their chance values, which were computed using the marginal probabilities displayed by each marker. For brevity, we have omitted these results, as they are tangential to the point of the paragraph.

Given the relatively high marginal probability of negative events, however, precision was 21.5%. The F_1-measure, the harmonic mean of precision and sensitivity, was 0.30.

The precision metric reported above for the automated procedure should be viewed in light of the sparseness of positive events and in terms of its improvement on stage 1 alone. Moving from a data set that is 6% "pure"[10] to one that is 21.5% pure is an improvement of 258%. This large effective increase in signal-to-noise ratio improves the ability to detect differences in the rate at which true HFOs occur within and outside the seizure onset zone, a topic that we take up in the next chapter. It is also important to remember that precision, as well as the other performance metrics we report, is highly dependent on our definition of ground truth. If we consider a ground truth data set whose positively labeled examples are the union of all three human reviewers' positive markings, for example, the precision improves to 54.6% (with the F_1-measure improving slightly, indicating that this increase is not completely counterbalanced by a decrease in sensitivity). Also, the precision metric reported above is an aggregated measure with respect to the machine clustering. The precision for each of the four clusters considered individually is different and in some cases higher than this aggregate measure, as we discuss below.

---

[10]Six percent is the probability that a given event emerging from stage 1 would be declared a positive HFO by at least two human reviewers.

Table 4.2 shows the performance results obtained when we modify our ground truth definition in a manner consistent with the recommendations of Gardner et al. [32]. The modified ground truth set considers any event marked by at least two markers, human or automated, to be a positive HFO. The table compares the performance of each marker against this hybrid human-machine ground truth, and also gives the difference between each metric and that expected under a chance model.

Table 4.2. Performance metrics, modified ground truth

	A	B	C	M
Sensitivity (%)	78.1 (1.2)	44.2 (17.6)	55.2 (4.9)	69.1 (14.0)
Specificity (%)	81.5 (1.1)	98.9 (2.4)	93.5 (1.3)	93.4 (2.3)
Accuracy (%)	81.1 (1.0)	93.3 (2.9)	89.5 (1.0)	90.9 (3.0)
Precision (%)	32.5 (5.2)	82.4 (40.0)	49.1 (11.1)	54.6 (18.0)
F-measure	0.46 (0.057)	0.58 (0.253)	0.52 (0.087)	0.61 (0.167)

Modified ground truth data set formed by taking as positive HFOs all stage 1 detections marked positive by at least two detectors (human or automated). Values in parentheses are chance corrected.

Chance values, which can be computed exactly, were for convenience generated by simulation in the following way. For each rater, 100 random m x n marking matrices were generated, where m was the total number of marked events (4,773) and n was the total number of markers (4). Random marking matrices were drawn according to actual probability mass function displayed by each reviewer. For each trial, performance metrics were computed using the modified ground truth rule described above, and the 100 values in each performance metric category were averaged to yield a final expected value for each. Values in parentheses in the table are the differences between the observed values and these chance values.

4.4.6 Machine cluster purity

Given a machine-positive HFO cluster (i.e. 1, 3, or 4) the probability that one of its members was also marked positive by human reviewers was dependent on cluster. Table 4.3 shows these results for two cases, one in which ground truth positive is taken to be the union of all human reviewers' positive markings and one in which ground truth positive is taken to be a majority vote. For completeness, we also include the values computed for cluster 2. For both the majority ground truth ($\chi^2(2, N = 619) = 13.64, p = 0.0011$) and the union ground truth ($\chi^2(2, N = 619) = 33.88, p \ll 0.0001$), we reject the null hypothesis that the proportion of ground truth positive events occurring in each of the three machine positive clusters is the same.

Table 4.3. Cluster purity

	Majority	Union
Cluster 1	0.15	0.40
Cluster 2	0.01	0.11
Cluster 3	0.32	0.54
Cluster 4	0.21	0.67

Entries are the proportion of events marked as positive HFOs according to the ground truth rule specified by the column heading.

4.5 Discussion

The results of this marking study strongly reinforce the idea that we are in the nascent stages of describing high-frequency oscillations within the brain. Human reviewers do not agree on the prevalence of HFOs. Nor, relatedly, do they agree particularly well on what constitutes an HFO when they see one. Other researchers' results strongly suggest that, in addition to poor inter-rater agreement, intra-rater reliability is moderate at best [32]. Different reviewers demonstrate strong preferences for waveforms with differing characteristics. Nonetheless, the level of agreement does exceed chance – there is a core of commonality worth investigating more thoroughly. But the evidence makes it clear that, currently, "ground truth" HFO data are a false sense of security, and should be regarded as suggestive rather than authoritative.

The automated algorithm we introduce performs similarly to humans at the task of culling positive exemplars from a large set of candidate HFOs. Humans agree no more with each other than they do with the machine. The second and third stages of the automated algorithm, taken together, offer at least a threefold improvement in positive predictive value over the stage 1 detector alone – more if we consider individual clusters, some of which seem to capture waveform features that are more saliently HFO-like to humans than others. The automated approach provides the further advantages of being perpetually consistent in its application of detection

criteria and indefatigable in its marking effort.

The relative uncertainty among humans about what constitutes an HFO gives us confidence in framing our work as exploratory, and in the value of studying the outputs of our algorithm on their own merits. In the following chapter, we examine the relationship between the clusters our algorithm finds and putative areas of seizure generation.

Chapter 5

HFOs and Seizure Onset Regions

5.1   Summary

Transient high-frequency (100-500 Hz) field potential oscillations (HFOs) have been studied extensively in human mesial temporal lobe brain structures for their potential to delineate epileptic tissue. Comparatively little is known about their localizing properties in neocortical epilepsy and their prevalence in normal brain. We present a quantitative analysis of HFO subclasses and their rates of occurrence in a group of nine patients with epilepsy believed to be of neocortical origin and two control patients with no history of seizures. HFOs were detected and classified automatically in continuous long-term micro- and macro- intracranial recordings without performing data pre-selection. There are four main results: (1) A cluster of HFOs with power concentrated in the low end of the 100-500 Hz band (median spectral

centroid = 137 Hz) is increased in the seizure onset zone, while (2) a cluster with power concentrated in the high end of the band (median spectral centroid 305 Hz) is not; (3) We find no evidence that the rates at which HFOs are generated are different for control neocortex and non seizure onset zone neocortex in epilepsy patients; and (4) While HFOs recorded on parenchyma-penetrating microelectrodes have higher peak 100-500 Hz frequencies then penetrating macroelectrodes, this relationship does not hold for epipial electrodes on the neocortical surface.

5.2   Introduction

Epilepsy affects an estimated 50 million people worldwide [62], with 30% of patients having seizures that cannot be controlled by medication [45]. Current treatment of drug-resistant partial epilepsy assumes that the brain areas involved in initiating seizures are often well-circumscribed and unchanging, and hence amenable to neurosurgical resection. A common tool in surgical planning, used when noninvasive measures fail to unambiguously identify the seizure onset-zone (SOZ), is intracranial electroencephalograpy (iEEG). The state of the art in iEEG-based SOZ localization is labor-intensive and somewhat subjective: neurologists trained to recognize stereotypical epileptiform patterns visually review multichannel recordings to identify regions that consistently show the earliest departures from background activity prior to electrographic seizure onset. Epilepsy surgeries guided by these assessments

are moderately successful [72], but outcomes tend to be relatively poor for patients who do not have physical lesions apparent in magnetic resonance images [7], which provide supplementary localizing information.

The search for more reliable and quantifiable biomarkers for epileptic tissue has led to the recent study of transient, quasi-periodic field potentials within the 100-500 Hz frequency range. These signals, which last on the order of tens of milliseconds, have been observed in hippocampus and parahippocampal structures as well as in neocortex. They have been recorded in both normal and epileptic animals and in epileptic humans, using electrodes varying in size from the micron to the millimeter scale. In the epilepsy community (and outside) they are termed "high-frequency oscillations" (HFOs), to distinguish them from the slower activity (0.1-40 Hz) more commonly used in making clinical determinations of the SOZ.

HFOs are sometimes subdivided, conceptually and for analytical purposes, into "ripples" (100-200 Hz) and "fast ripples" (250-500 Hz), reflecting hypothesized differences in their cellular and network-level generating mechanisms, as well as their relationships to epileptogenic areas. It has been suggested, for example, that ripples in the hippocampal CA1 region reflect summed inhibitory postsynaptic potentials resulting from synchronous input from interneurons [85]; and also that they are formed by bursts of pyramidal cell population spikes [9]. They are thought to be

important in declarative memory consolidation [82], but have also been implicated in pathological processes [9, 12]. The mechanisms of fast-ripples, if different from ripples, are even less clear, and only recently being deliberately investigated [27, 31]. And though fast ripples are more frequently associated exclusively with pathology in the literature, authors have cautioned against inferring from this the unsupported mnemonic that "ripples equal normal and fast ripples equal epileptic" [28, 75].

The relationship between HFOs in more commonly studied mesial temporal lobe structures and the phylogenetically and structurally different neocortex also remains unclear, and systematic characterization of HFOs in normal brain for baseline comparison has not been done. A further complicating factor is that studies of HFOs in epilepsy have been hampered by selection bias arising primarily out of the need to reduce the volume of data so that clinicians can manually review it.

In chapter 1 we demonstrated an algorithm for the automated detection and classification of HFOs that required no human intervention and did not presuppose the existence of any specific number of classes, ripples and fast ripples included. Here, we quantitatively analyze the outputs of that algorithm in a group of nine neocortical epilepsy and two control patients implanted subdurally with hybrid macro/micro electrode arrays. We address three open questions about the relationship between neocortical high-frequency oscillations and seizure generation and the sensors re-

quired to record them: 1) Are the rates at which HFOs are generated different in control patient neocortical regions and the non seizure onset-zone (NSOZ) neocortical regions of epilepsy patients?; 2) Are the rates at which HFOs are generated different within and outside the SOZ in epilepsy patients?; and 3) Do microelectrodes tend to record HFOs of higher frequency than macroelectrodes, as might be expected if fast ripples were a more local phenomenon than ripples?

5.3   Methods

5.3.1   Data and subject description

The data set we analyze is comprised of 219,756 putative high-frequency oscillations recorded from the neocortices of eleven human subjects with subdurally implanted electrode arrays. Nine subjects had medically refractory partial epilepsy believed to be of neocortical origin and received electrode implants as part of routine clinical care. The location, number, and type of intracranial electrodes (grid and/or strip electrodes[1]) were determined by a multi-disciplinary team including neurosurgeons, neurologists, neuroradiologists, and neuropsychologists. Standard clinical grid and

---

[1]Depth electrodes were also implanted in some subjects; standard clinical depth leads were modified in two ways: 1) by embedding microwires around the circumference of the lead-body between the 2 mm long clinical contacts; and 2) by passing a bundle of microwires within the lumen of the lead, so that they protruded by approximately 7-8mm from the distal tip. To ensure that our conclusions pertain to neocortex, however, we excluded depth electrode data from all but one analysis (clearly described in the text). The purpose of the latter analysis was precisely to contrast depth and surface electrode results. Insufficient amounts of depth electrode data precluded comparative analyses elsewhere.

strip electrodes were modified under an Institutional Review Board (IRB)-approved research protocol by adding arrays of non-penetrating platinum-iridium microwires (40 μm diameter, with intra-array spacing of 0.5-1 mm center-to-center) between the clinical, 4 mm diameter contacts [78, 83]. Two "control" patients with chronic intractable face pain but no history of seizures were similarly implanted, as part of an unrelated IRB-approved research protocol investigating electrical stimulation of motor cortex as a potential treatment for their condition.

5.3.2 Signal acquisition and processing

Continuous, long-term data were acquired using the Digital Lynx Data Acquisition System (Neuralynx, Inc., Bozeman, MT) at 32,556 samples per second and 20 bits per sample (stored), from up to 144 channels in each patient [14]. The input dynamic range was $\pm$ 132 mV and the noise level was ~1.3 μV root-mean-square (RMS), yielding approximately 18 effective bits. Recordings were made using direct-current (DC) capable amplifiers, and a 9 kHz analog low pass filter was employed to minimize aliasing effects. The bandwidth of the raw recordings was thus approximately near-DC – 9 kHz.

Event data were extracted from the raw intracranial EEG (iEEG) using a three-stage process detailed in the methodology chapter. Briefly, in the first stage, candi-

date events were detected within non-overlapping ten-minute windows of 100-500 Hz bandpass filtered iEEG[2], on a per-channel basis. Amplitude and duration thresholds were applied to the short-time energy of the signal in each window; flagged segments were then subject to waveform shape criteria to ensure robustly oscillatory characteristics [21, 22, 32, 68]. In the second stage, a Gaussian mixture model of the local background iEEG ($\sim$2.5 sec) surrounding each candidate was learned using the Expectation Maximization (EM) Algorithm [23], and events bearing too large a spectral similarity to background according to the model were discarded from candidacy. In the final stage, events were clustered using the k-mediods algorithm [36], with the gap statistic [74] used to determine the optimal number of clusters. The approach yielded four clusters, which we analyze here.

5.3.3 Localization of the seizure onset-zone

For each patient, all recorded seizures were visually identified by a board-certified epileptologist prior to data analysis. The time of the earliest iEEG change was noted and the associated macroelectrode(s) were selected as the electrographic seizure onset zone. Time of earliest iEEG change was determined by identifying a clear electrographic seizure discharge in a macroelectrode recording and then looking backward in the record for the earliest change from background contiguously asso-

---

[2]Data were first decimated by a factor of 12, to 2,713 Hz.

ciated with the discharge. As microelectrode recordings were not used in clinical determination of the SOZ, their labels were extrapolated from those of the macroelectrodes. A microelectrode was given the label of SOZ if it belonged to a cluster immediately adjacent to a non-peripheral SOZ macroelectrode.

5.3.4 Statistical analyses

5.3.4.1 Control versus NSOZ groups

A permutation test was used to test the null hypothesis that there was no difference in the average median event rate between control and NSOZ groups. The permutation distribution was generated by taking all 55 (11-choose-2) possible groups of 9 and 2 subjects and computing the difference between the average median event rates for the two groups. Separate tests were carried out for each of the four clusters, and for macro- and microelectrodes, yielding eight total tests.

5.3.4.2 SOZ versus NSOZ channels

Because of the large variability in median event rates across seizure subjects, and because not all subjects were represented in all channel classes – for example, only two subjects had surface microelectrodes in the physician-labeled SOZ – we chose not to aggregate across subjects in studying the difference between SOZ and NSOZ

channels. Subjects were considered eligible for analysis if they had at least five SOZ channels and five NSOZ channels of a given channel class (i.e. macroelectrode or microelectrode). Of the nine seizure patients, five met this criterion for macroelectrodes (SZ 01, SZ 05, SZ 06, SZ 07, and SZ 09) and two for microelectrodes (SZ 02 and SZ 06). For each eligible patient, using event rate as the measured variable on channels, we tested the null hypothesis that the cumulative distribution function of event rates for SOZ channels was equal to that for NSOZ channels[3]. Separate analyses were carried out for each of the four clusters, and for macro- and microelectrodes, yielding a maximal total of eight tests per patient. The Mann-Whitney U test was used to test the null hypotheses, with the Bonferroni adjustment made to account for multiple comparisons.

5.3.4.3 Macro- versus microelectrodes

For each of the 219,756 neocortical surface events, we computed the peak frequency within the 100-500 Hz frequency band using a multitaper method from Thomson [73](3 tapers; 512-point Discrete Fourier Transform (zero-padded); adaptive non-linear combination method). We tested the null hypothesis that the cumulative distribution functions of event peak 100-500 Hz frequencies were equal for macro- and microelectrodes, with the alternative hypothesis being that the microelectrode

---

[3]This same test is also commonly, though less accurately, described as a test for a difference between the medians of two groups.

distribution was stochastically larger. The Mann-Whitney U test was used for the comparison. The same was then done for the 70,517 depth electrode events.

5.4   Results

5.4.1   HFO characteristics

Across eleven subjects, a total of 219,756 putative high-frequency oscillations (100-500 Hz) were automatically detected on neocortical surface channels and clustered by the algorithm reviewed above. The statistical clustering approach we used assumed no specific number of HFO groups in the data, only that the number was somewhere between one and twenty. Four clusters were discovered. The work detailed in the methodology chapter demonstrated the stability of the clustering algorithm and qualitatively described the waveform morphologies representative of each cluster. We briefly summarize those observations in the next paragraph. In this study, we analyze the clustering outputs, rather than the clustering method, quantitatively, and in relation to variables of clinical interest.

Twenty-nine percent (63,652) of all detections belonged to a cluster (#2) characterized by sharp, supra-physiological voltage transients; the sharp features yielded oscillations when bandpass filtered, thereby generating putatively false positive de-
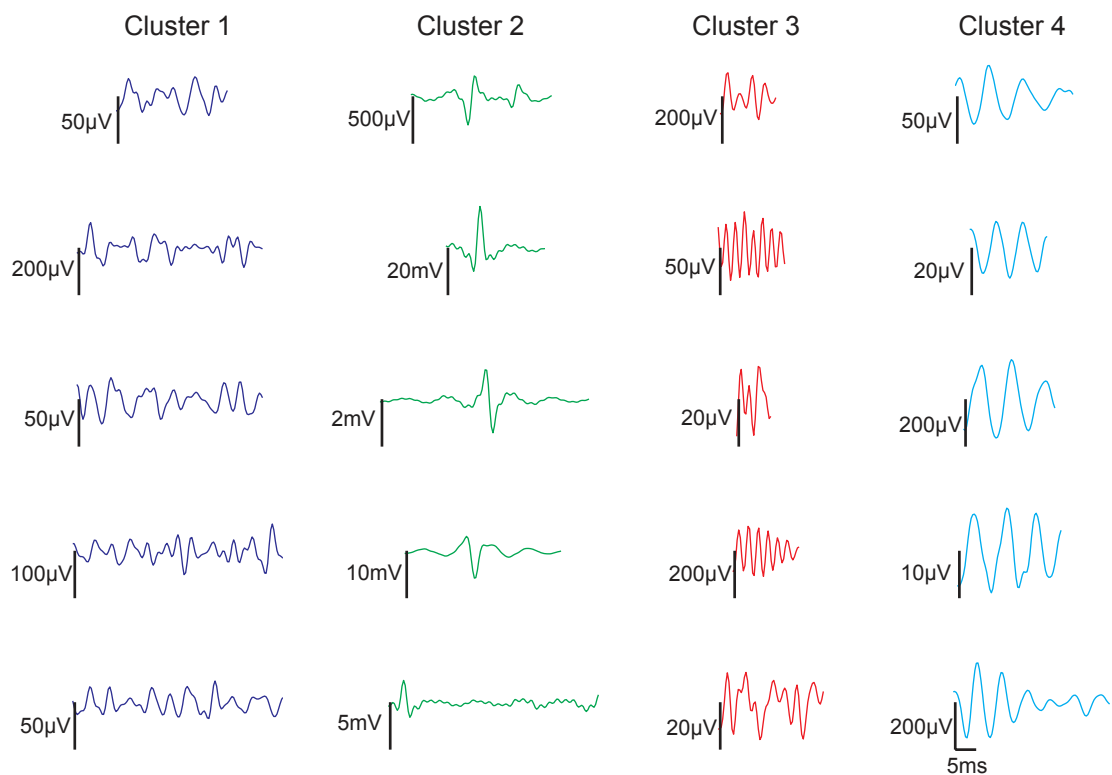
Figure 5.1. Five randomly selected waveforms from each of the four clusters found using the automated detection and unsupervised classification described in chapter 3. Waveforms are 100-500 Hz bandpass filtered segments corresponding to detections (truncated to 25 ms, if necessary, to put all waveforms on the same time scale for comparison). Note that these waveforms differ from those shown in figure 3.7B in that they are drawn exclusively from neocortical surface channels.

tections (see methodology chapter for further discussion). Of the remaining puta-

tively physiological events, 41% (64,392) belonged to a cluster (#1) characterized

by mixed-frequency, irregular waveforms; 11% (16,691) belonged to a cluster (#3)

characterized by relatively fast, regular waveforms; and 48% (75,021) belonged to a

cluster (#4) characterized by relatively slow, regular waveforms. Figure 5.1 shows

five randomly selected examples from each of the four clusters, illustrating these

properties.

Table 5.1 gives descriptive statistics for each cluster. Four quantities were computed

for each detection (i.e. cluster member): 1) the ratio of the integrated power spectral

density estimate (PSD) in the 250-500 Hz frequency band to that in the 100-200

Hz band (power ratio); 2) the centroid of the PSD within the 100-500 Hz frequency

band (spectral centroid); 3) the line length of the spectrally equalized time-domain

signal [32, 77], normalized by signal duration (line length); and 4) the peak of the

PSD in the 100-500 Hz band (spectral peak).

Table 5.1. Cluster statistics

Feature	Cluster 1 $(n=64392)$	Cluster 2 $(n=63652)$	Cluster 3 $(n=16691)$	Cluster 4 $(n=75021)$
PR	(0.823,0.986)	(0.833,1.018)	(3.725,3.196)	(0.0357,0.1069)
SC (Hz)	(219,49)	(220,50)	(305,37)	(137,35)
LL (AU)	(0.0313,0.0123)	(0.00657,0.00881)	(0.0242,0.0299)	(0.0263,0.0110)
SP (Hz)	(193,61)	(167,64)	(302,87)	(146,32)

Entries are (*median,interquartile range*); AU = arbitrary units; PR = power ratio; SC = spectral centroid; LL = line length; SP = spectral peak.

Since these quantities represent a subset of the features used to derive the clusters themselves, statistical comparison of their values across groups (using, for example, nonparametric analysis of variance methods) is not appropriate. Nonetheless, examining the values in table 5.1 gives a sense for the between-group differences that drove the clustering.

It can be seen, for example, that cluster 3 events tend to have higher peak 100-500 Hz frequencies than cluster 4 events. Furthermore, the relative proximity of the spectral centroid value to the spectral peak value in clusters 3 and 4, compared with clusters 1 and 2, suggests that the former pair have their power concentrated in a narrower band within 100-500 Hz than the latter – corroborating the descriptor "regular" for clusters 3 and 4 in the discussion above. Clusters 1 and 2 appear primarily distinguished by their line length values, which is consistent with the observation that cluster 2 events are characterized by sharp transients; that is, large sample-to-sample voltage differences that would contribute to high line-length values are present for a relatively smaller fraction of the overall segment duration in cluster 2 versus cluster 1 events. Cluster 1 events have spectral centroid values that are intermediate between those of clusters 3 and 4, supporting the observation that cluster 1 events tend to have relatively strong spectral components in both the 100-200 and 250-500 Hz band, and leading to their qualitative description as "irregular." Cluster 3 events have a median peak frequency that is consistent with

what prior studies have identified as "fast ripples," and cluster 4 events with what have been identified as "ripples."

5.4.2 Channel statistics by patient

Table 5.2 shows aggregated event rates for clusters 1,3, and 4 (i.e. all but the putatively artifact class), broken down by patient and by channel class. Each cell in the table gives the number of channels, the median aggregate event rate for those channels, and the interquartile range. The final column gives the total number of channel-hours of data recorded per patient.

Table 5.2. Channel statistics

| | Channel Class | | | | |
| | NSOZ | | SOZ | | |
Sub	Mac	Mic	Mac	Mic	**ChH**
CT 01	(**16**,26.87,17.69)	(**128**,14.29,19.73)	(**0**,-,-)	(**0**,-,-)	1176
CT 02	(**24**,13.54,8.71)	(**104**,8.71,11.27)	(**0**,-,-)	(**0**,-,-)	1877
SZ 01	(**34**,2.34,3.13)	(**6**,11.72,8.07)	(**5**,48.96,131.25)	(**0**,-,-)	480
SZ 02	(**22**,13.43,14.07)	(**79**,10.19,11.16)	(**1**,12.78,0.00)	(**24**,12.59,8.61)	1890
SZ 03	(**27**,31.37,12.25)	(**33**,37.84,29.61)	(**1**,32.55,0.00)	(**0**,-,-)	864
SZ 04	(**51**,18.54,7.86)	(**28**,9.44,8.51)	(**3**,18.87,4.72)	(**0**,-,-)	1921
SZ 05	(**33**,19.32,7.91)	(**0**,-,-)	(**11**,22.91,4.34)	(**0**,-,-)	2860
SZ 06	(**17**,2.65,2.72)	(**25**,8.03,8.32)	(**6**,7.75,1.73)	(**67**,8.35,7.27)	10139
SZ 07	(**57**,23.97,16.79)	(**22**,0.00,8.33)	(**7**,35.24,14.06)	(**0**,-,-)	5615
SZ 08	(**36**,5.05,1.49)	(**0**,-,-)	(**0**,-,-)	(**0**,-,-)	234
SZ 09	(**81**,2.03,1.22)	(**22**,0.00,1.63)	(**7**,2.85,0.71)	(**0**,-,-)	735

Triplets are (*number of channels, median event rate, interquartile range*), where an "event" is any member of cluster 1, 3, or 4 (i.e. putative artifact cluster 2 excluded). Units for the medians and interquartile ranges are $\times 10^{-4}$ *counts/sec.* ChH = Channel hours, the sum over all channels of the length of time recorded for each channel.

The table provides important context for the statistical analyses of the following

sections, making sample sizes clear and indicating where data are unavailable. For seven of the eight (all but SZ 02) seizure subjects for whom the comparison can be made, the median aggregate event rate is higher in the seizure onset zone than the non-seizure onset on macroelectrodes; the same is true on microelectrodes for two of the two subjects for whom the comparison can be made. Figure 5.2 complements table 5.2, showing the average proportion of events falling into each of the four clusters, broken down by patient and channel class. Clusters are colored in accordance with Figure 5.1.

Two main trends are apparent in subjects whose pie charts represent more than a single channel[4]: 1) cluster 3 (red) typically accounts for the smallest proportion of events; and 2) cluster 4 (cyan) typically accounts for a greater proportion of events in SOZ than in NSOZ channels. Below, we test whether the trends toward higher event rates in the SOZ apparent in Table 2 and Figure 5.2 are statistically significant, disaggregating clusters in order to discern which are responsible for any significant effects.

---

[4]SZ 02 and SZ 03 each have only one macroelectrode in the SOZ.
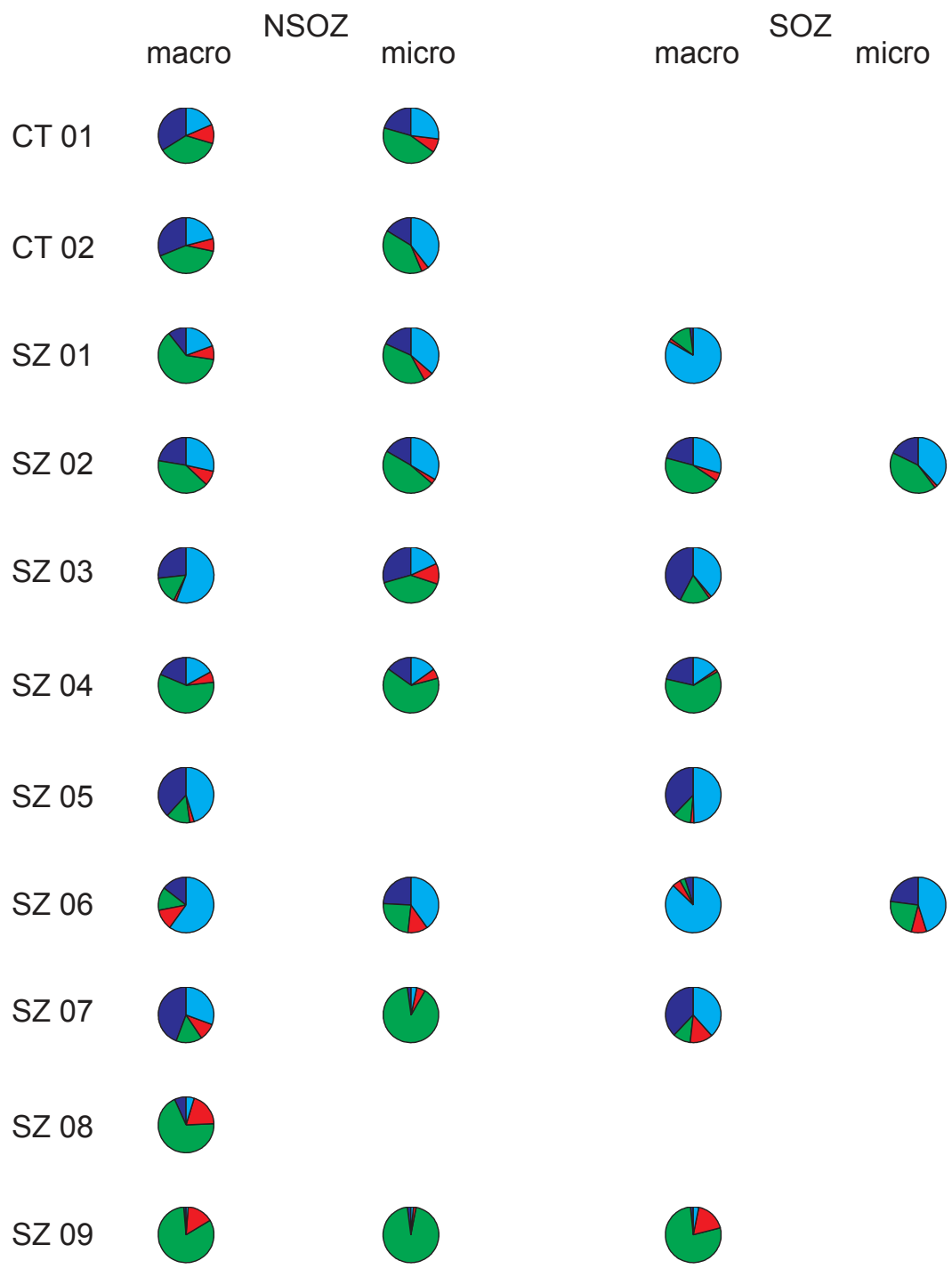
Figure 5.2. Caption on next page.

Figure 5.2. The area of each pie chart wedge corresponds to the mean proportion of events in a given cluster (cluster coloring scheme is the same as in figure 5.1: blue = cluster 1; green = cluster 2; red = cluster 3; and cyan = cluster 4), where the mean is over all channels in the category defined by the row and column of the pie. Empty cells in the table indicate that no data were available.

5.4.3   Control versus non seizure onset-zone (NSOZ) groups

Even at the liberal alpha level of 0.1, no significant differences were found between control and seizure patient-NSOZ groups. For macroelectrodes, p-values for permutation tests of differences between the average median event rates of control versus seizure patient NSOZ groups, before correcting for multiple comparisons, were: 0.25, 0.82, 0.56, and 0.69 for clusters 1, 2, 3, and 4, respectively. The corresponding p-values for microelectrodes were 0.92, 0.63, 0.92, and 0.94. Given the small number of control subjects these tests have little power, but they lead to the conclusion that there is no evidence to reject the null hypothesis of no difference between control and seizure-NSOZ groups.

5.4.4   Seizure onset-zone (SOZ) versus NSOZ channels

We compared event rates on SOZ channels to NSOZ channels on an individual subject basis, by electrode type (macro or micro) and by cluster. Five of nine seizure patients met our inclusion criterion (at least 5 channels in each class, SOZ and NSOZ) for macroelectrode comparisons, and two of nine met the same inclusion criterion for microelectrode comparisons. For microelectrodes, no significant differ-

ences were found between SOZ and NSOZ for any of the clusters, in either of the eligible patients. For macroelectrodes, no significant differences between SOZ and NSOZ were found in any of the patients for clusters 1-3.

For cluster 4, however, four of five eligible patients had higher median event rates on SOZ macroelectrodes. Uncorrected p-values for the Mann Whitney U test in these subjects – SZ 01, SZ 05, SZ 06, and SZ 07 – were 0.00036, 0.037, 0.00069, and 0.098. The marginally significant (at alpha = 0.1) results for patients SZ 05 and SZ 07 do not survive the Bonferroni correction.

The only eligible subject that did not have a higher median cluster 4 event rate for macroelectrode SOZ channels was SZ 09, who, anomalously, had median rates of zero on both channels. And though the median rate was not higher in SOZ channels for this subject, the mean rate was, by a factor of 1.7. High percentages of NSOZ and SOZ channels without any cluster 4 events was not surprising for this subject, given the uniquely poor quality of SZ 09's recordings, which is evidenced by the predominance of the color green (cluster 2) in that subject's plots in Figure 5.2. The results of the NSOZ to SOZ comparisons for macroelectrodes are summarized in the box plots of Figure 5.3.
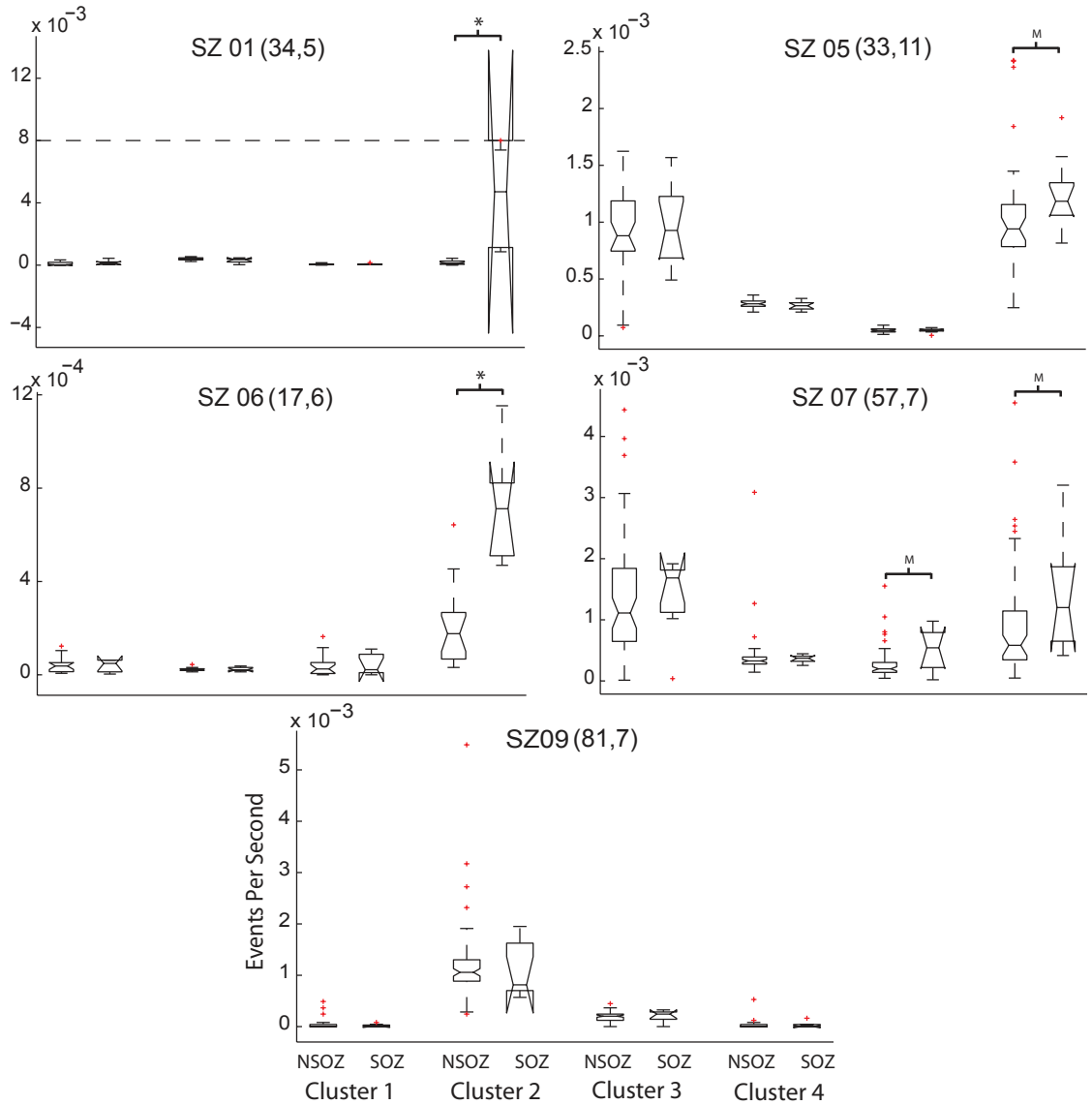
Figure 5.3. Caption on next page.

Figure 5.3. Each panel corresponds to results for one of the five subjects who met the inclusion criterion (at least 5 channels in each of the SOZ and NSOZ). Numbers of channels upon which box-and-whisker plots are based are shown in parentheses at the top of each panel, next to the subject name as (#NSOZ,#SOZ). There are two box-and-whisker plots for each of the four clusters, one corresponding to NSOZ channels (left) and one to SOZ channels. On each box, the central mark is the median event rate (counts/sec) and the edges are the 25th and 75th percentiles. Whiskers extend to the most extreme data points not considered outliers, where outliers are defined as points greater than 1.5 times the interquartile range above the 75th percentile or below the 25th percentile. Outliers are plotted individually as red crosses. Dotted line in top left panel represents an arbitrary value to which a single large outlier (34.7×10^{-3}) was clipped for visualization purposes. Notch widths are computed according to [50]; lack of notch overlap is a rough test for significant differences in medians at the 5% significance level; we use the Mann-Whitney U test on the distributions to more formally test significance. (* = Bonferroni corrected p-value for Mann-Whitney U test < 0.1, M = marginally significant at alpha level 0.1, but does not survive the Bonferroni correction).

5.4.5 Macro- versus microelectrodes

In contrast to findings from human temporal lobe [83], in neocortex we find no evidence ($p \approx 1$) to reject the null hypothesis of no difference in favor of the alternative hypothesis that the cumulative distribution function of event peak 100-500 Hz frequencies is stochastically larger for microelectrodes than macroelectrodes. As can be seen in figure 5.4A, the distributions for macroelectrode events ($n = 130,098$) and microelectrode events ($n = 89,658$) look qualitatively similar, and the medians are close (macro median, interquartile range: 170,69; micro median, interquartile range: 164,61). In fact, a post-hoc two-sided Mann Whitney U test finds evidence for the opposite effect ($p \ll 0.001$), namely that macroelectrodes have a tendency to record events with peak 100-500 Hz frequencies that are slightly higher than those for microelectrodes.
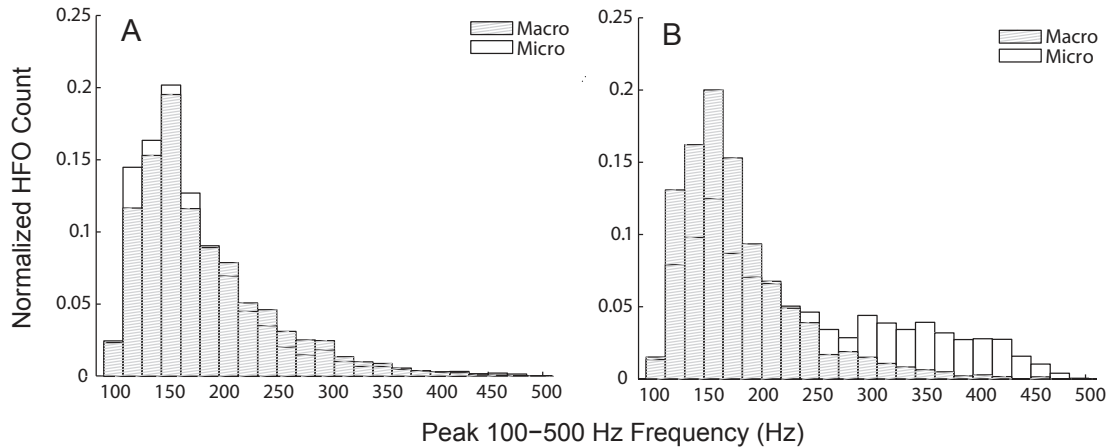
Figure 5.4. Histogram density estimates for the peak 100-500 Hz frequency of events (clusters 1-4, aggregated) on micro- (white) and macro- (hatched) electrodes, for surface (A) and depth (B) electrodes.

Figure 5.4B, on the other hand, shows a result that does agree with findings from the temporal lobe. It depicts the same two distributions as in figure 5.4A, but for depth electrodes, which were implanted in four of the nine seizure patients (SZ 03, SZ 04, SZ 05, and SZ 08) in addition to surface electrodes. Depth macroelectrodes yielded 11,179 events on 40 total channels in these four subjects. Depth microelectrodes, only present in two of the four subjects, yielded 59,338 events on 84 total channels. In figure 5.4B, the distributions show a strong qualitative difference. The medians are much farther apart than for the surface electrodes (macro median, interquartile range: 167,58; micro median, interquartile range: 209,154). Here, the one-sided Mann-Whitney U test, with alternative hypothesis that microelectrodes have stochastically larger peak (100-500 Hz) event frequencies than macroelectrodes, leads us to strongly reject the null hypothesis of no difference ($p <<< 0.0001$).

5.5 Discussion

There are four main findings of this work: 1) In a group of neocortical epilepsy patients, using automated detection and classification techniques without data pre-selection, we have identified a class of HFOs corresponding to ripples whose rate of occurrence is increased in the physician-labeled seizure onset-zone; 2) A class of oscillations corresponding to fast ripples is relatively rare, and does not show a rate increase in the neocortical seizure onset-zone; 3) We find no evidence that control neocortex is different from neocortex outside the seizure onset zone in epilepsy patients, when considering the rate at which HFOs are generated; and 4) While microelectrodes on the neocortical surface do not appear to preferentially record HFOs of higher frequency, microelectrodes embedded in the parenchyma do.

Spontaneous 100-500 Hz HFOs have rarely been systematically studied in human epileptic neocortex, and data from the neocortex of nonepileptic control patients is more scarce still. Here, we analyzed over two-hundred thousand automatically detected HFO candidates in neocortical recordings from nine epilepsy and two control patients, originating in over 27,000 channel-hours of iEEG. Though our number of subjects was comparable with prior studies of HFOs in human epilepsy, the overall volume of data we processed was significantly larger. We hypothesize that ana-

lyzing multi-hour continuous recordings permits more reliable estimates of global HFO rates across brain states than, for example, the short (e.g. 10-minute) iEEG segments typically examined under the constraints of human processing. Larger data sets also tend to produce more stable statistical estimates of the number of HFO classes (cf methodology chapter).

The finding of an increase in ripple-like oscillations in the neocortical SOZ is consistent with a prior study that used human-intensive processing [40]. The finding that fast ripple-like events are rare and not increased in the neocortical SOZ is in reasonable accord with one study that reported that >200 Hz HFOs were not observed in the four neocortical epilepsy subjects they studied [20]. Urrestarazu et al. also report fast ripples only rarely in neocortex [76]. They reported fast ripple SOZ rate increases in a small number of patients, but the rate increases did not reach statistical significance.

We observed events of all four classes in both control patients. Moreover, event rates for control regions were not particularly large or small for any cluster when compared with seizure-subject NSOZ regions. With only two control patients, our permutation tests had low power and these conclusions are very weak. But they are discouraging for the prospect of finding a universally "normal" rate of HFOs, which might serve as a baseline for patient-independent detection of the SOZ.

We are uncertain how to interpret the finding that microelectrodes record HFOs of higher frequency than macroelectrodes on depth but not surface electrodes. Given the cellular architectural differences between hippocampus, for example, and neocortex, it is plausible that this result reflects genuine physiologic differences. An alternative explanation might be related to the fact that surface microelectrodes are non-penetrating and rest atop the pia mater, which may act as a lowpass filter, while the depth microelectrodes actually penetrate the parenchyma.

Our data-mining approach (as well as subsequent analyses) treats all iEEG equally. On one hand this can be viewed as a limitation: our methods do not explicitly attempt to parse whether HFO detections are occurring ictally or interictally, or during specific states of arousal, or in conjunction with epileptiform events such as sharp waves, or within or outside of lesions, for example. As a result, interpreting the finding of an increase in cluster 4 event rates in SOZ must be done with suboptimal precision, and we cannot say with certainty that our detections provide information beyond that of potentially correlated phenomena. (That said, other authors have shed light on these issues, concluding that HFOs provide localizing information independent from that of interictal spikes [39], lesions [40], and seizures themselves ([83] and many others)).

On the other hand, this algorithmic naivete is a blessing, as it leads to several enticing ideas about HFOs in epilepsy. First, the fact that we find a signal that increases in the SOZ without special selection of patients, channels, or time-epochs for processing, suggests that the signal is strong. Second, it suggests that similar findings in prior studies of HFOs in mesial temporal lobe, which typically use restrictive data pre-selection criteria, may actually be more generalizable, and hence potentially more practically useful in the clinical setting. Finally, that we are able to detect and classify these signals *automatically*, without any human intervention, adds to the promise of practical clinical utility. Not only does it portend increases in the efficiency and reliability with which presurgical evaluations can be done; it also opens a door to the possibility that HFOs can be used as control signals in closed-loop implantable seizure therapy devices, which, at least in early generations will not have the same luxuries as human researchers to pre-identify optimal data for processing. Such devices, one of which is currently in clinical trials [4], offer hope to patients who are not surgical candidates.

Chapter 6

Conclusion

6.1   Summary of Findings

In this thesis, we introduced a new method for automatically detecting and classifying 100-500 Hz field potential oscillations in intracranial electroencephalographic records. We ran the algorithm on over 31,000 channel-hours of human iEEG, a data set orders of magnitude larger than any previously processed in the context of high-frequency oscillation analysis. We compared the outputs of the machine with human performance on the HFO classification task and found the algorithm as proficient as human experts at labeling ground truth events. Finally, we used the algorithm to investigate HFOs in neocortical epilepsy patients, providing one of the first systematic studies in this patient population, and the only that includes a direct comparison with control patients having no history of seizures.

112

We observed all four automatically identified classes of spontaneous neocortical HFOs in both epilepsy patients and controls, and found no significant differences between the rate at which HFOs occurred in control neocortex and that in non-seizure-onset neocortex in epilepsy patients. We found that one class of oscillations (median bandpassed spectral centroid $\sim$140 Hz) corresponding to what are known as ripples in the rodent and epilepsy literature was increased in rate in the seizure onset zone – the first time this finding has been demonstrated in an unselected data set without human intervention. A class of oscillations corresponding to fast ripples was rare in occurrence and, in contrast to findings from temporal lobe structures, showed no significant rate differences between non-seizure-onset and seizure-onset neocortices. Significant differences in neocortical surface HFO rates within and outside the SOZ were found only on macroelectrode recordings; none were found on microelectrode recordings. For depth electrodes, events detected by microelectrodes had stochastically higher peak frequencies than those detected by macroelectrodes, though this was not true for neocortical surface electrodes.

6.2 Insights and Future Directions

6.2.1 Near-term clinical transfer

The work in this dissertation is ready for immediate integration as a research tool into the clinical workflow for SOZ localization. For example, we can produce spatial maps of HFO rates like the one shown in figure 6.1, for each sensor device implanted in a patient, in a time frame sufficiently short for the data to inform the surgical decision, if desired.
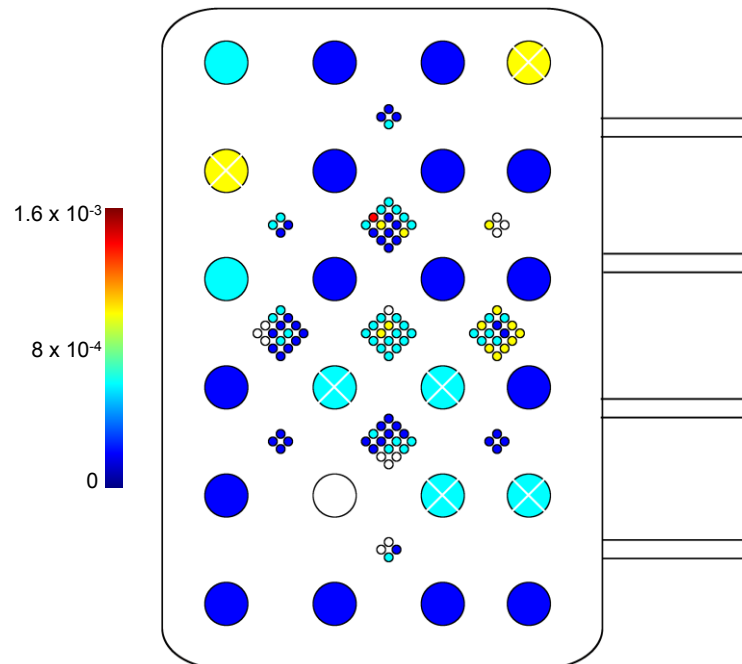
Figure 6.1. Spatial HFO map for subject SZ 06, showing cluster 4 event rate (event/sec) at each channel on a hybrid 6 x 4 subdural grid. White crosses mark macroelectrode channels for which underlying tissue was surgically resected. Unfilled circles indicate where no data were available.

Integrating similar displays with 3-D brain renderings generated from MRI images,

as well as producing time-series images of HFO-rate spatial maps, are also active projects.

6.2.2 Quantitative SOZ channel identification

Finding a signal that increases in the neocortical SOZ in the absence of stringent data selection and other human processing is highly encouraging. But whether HFO rates can reliably be used to identify SOZ is still in part an open question. There are at least two reasons for this. First, observing an increase in the rate of (cluster 4) HFOs relative to a "baseline" NSOZ rate may strongly suggest, but does not imply, that a channel is in the SOZ. It is possible, for example, that areas of secondary propagation also show HFO increases, but that the effect is masked by aggregating such channels with more numerous truly benign ones. Fortunately, others who have investigated the specificity of HFO rate increases in temporal lobe structures have reported them to be well-localized to primary seizure onset areas. The evidence is mostly anecdotal, however, and further tests should be done in neocortical patients to confirm it[1]. A straightforward approach would be to visually identify channels of secondary spread and conduct analyses similar to those of chapter 5, using three groups – secondary spread, SOZ, and NSOZ – instead of two, and substituting Kruskal-Wallis analysis of variance in place of Mann-Whitney U tests.

---

[1]At the same time, we note it is not clear that clinical consequences would be adverse if areas of secondary spread were confounded with areas of seizure onset in making decisions about which brain areas to resect.

The second and perhaps more serious practical challenge to using HFO rates as biomarkers for epileptogenic tissue is determining appropriate thresholds for declaring them abnormal. We found HFO rates both within and outside the SOZ to be highly variable among patients. Though our sample size of control patients was small, HFO rates in these subjects were not extreme relative to those in seizure patient NSOZ regions, making the prospect of using controls to establish global baseline "normal" absolute HFO rates unlikely. Before abandoning the idea completely, however, the correlation of event rates between cluster pairs (where observations are made on patients) should be examined. If a high correlation is found between cluster 4 rates and those of another cluster, it is possible that using a differential (or ratio) instead of absolute measurement would permit the establishment of a generic baseline.

As mentioned above, the significant increases we detected in SOZ were with respect to patient-internal baseline NSOZ rates. Establishing these baselines was done using prior knowledge, namely which channels were flagged as SOZ (and, implicitly, NSOZ) by the clinician. It is possible that prior knowledge can be similarly utilized in future automated attempts at prospectively identifying the SOZ. If clinicians can identify channels with high likelihood of being NSOZ, the parameters of a baseline distribution for NSOZ cluster 4 HFO rates can be learned, and SOZ channel iden-

tification can be treated in an outlier detection framework, controlling for multiple comparisons[2] Without prior information about an individual's baseline NSOZ rate distribution, identifying high-probability SOZ channels will be even more difficult. Monte-Carlo-type permutation methods would be worth investigating, especially if constraints could be placed on the sample space using assumptions on estimates of the marginal probability of being an NSOZ channel, for example, or on the spatial structure of NSOZ or SOZ channels. More naive approaches that attempt simply to find extreme values (e.g. "unusually" high rates of HFOs) in the absence of any other information, run the risk of missing a large number of SOZ channels in cases where they are relatively numerous.

An orthogonal approach to SOZ channel prediction using HFOs would be to attempt to learn a boundary separating NSOZ and SOZ channels in an HFO-derived feature space. Each channel in a training sample whose SOZ-NSOZ label was known could be summarized by a multivariate feature vector whose elements were, for example, median values of the features we used in unsupervised classification. Cluster labels could be ignored, used to determine which events were included in calculating the feature medians, or built into the construction of the feature vector itself – for example by representing each channel by a 12-dimensional (4 x 3, where 4 is the number of features used in clustering and 3 is the number of non-artifact clusters) feature vector. Augmenting this vector by including the rates of cluster 4 events,

---

[2]A nonparametric alternative might be to use permutation tests.

and perhaps those of other clusters, would seem a logical and potentially fruitful extension of the idea. Any of a wide array of well-known supervised learning algorithms (e.g. k-nearest neighbors, logistic regression, support vector machines, and many others) could be used to learn an effective input-output mapping, which could then be used to classify unseen channels as SOZ or NSOZ. This approach would permit a richer description of the relationship between HFOs and seizure generating areas than can be achieved by using the univariate measure of HFO rate alone. A major limitation, however, is that it relies heavily on accurate labeling of both the NSOZ and SOZ classes by humans, the latter of which by the very definition of the problem is particularly difficult to achieve.

6.2.3   Algorithmic improvements

Because of uncertainty in the specification of positive events at the outset, we made no explicit attempts to optimize stage 1 detection, instead choosing a detection scheme and parameter set that closely mirrored accepted approaches from the literature. Having now accumulated evidence that cluster 4 events are most important, a logical next step would be to tune stage 1 to detect this class with higher sensitivity and precision, under the assumption that the performance gains will lead to improved SOZ localization. Shrinking the window used by Staba et al. [68] for computing the per-segment RMS energy threshold (see chapter 3) has always

seemed intuitively promising[3]; now we can generate reasonable optimization criteria in manipulating this parameter. Using a nonparametric thresholding approach, as suggested by Gardner et al. [32], may also give some performance boost, though at increased computational cost.

Overhauling the stage 1 detection scheme is an alternative to making incremental improvements. With a large collection of cluster 4 exemplars now available, matched-filter type approaches could be tried. Another possibility would be to borrow from the literature on power law detectors, in particular the work of Nuttall [54] and Wang and Willett [80]. Nuttall addresses the problem of detecting transient, random signals of unknown location, structure, extent, and arbitrary strength in a background of stationary, white Gaussian noise with known power. Under some simplifying assumptions, he derives an easy-to-implement and empirically useful statistic for detecting such events: the sum of magnitude-squared Discrete Fourier Transform (DFT) coefficients, each raised to a power typically in the range of 2 to 3. Wang and Willett make natural extensions to Nuttall's original work in deriving their own transient-detector: they consider colored Gaussian noise, eliminate the assumption of known noise power, mitigate the effects of potential background nonstationarity by block processing, and leverage the observation that rather than being completely unknown in structure, transient signals tend to aggregate their energy in a band. These extensions seem promising for the HFO detection appli-

---

[3]Gardner et al. took a similar step in their automated analysis of high gamma oscillations.

cation, where transient, nonstationary signals with partially known structure but unknown duration, strength, and location, must be extracted from the typically physiologic "1/f" background. Wang and Willett also develop a further extension, which uses a wavelet, rather than Fourier decomposition in order to take advantage of the temporal concentration of transient signals. And importantly, they provide an approximation method for setting thresholds for the statistics they present in their paper.

6.2.4 Real-time implementations

Both real-time HFO rate display and real-time HFO-based SOZ identification seem realizable. The speed with which SOZ inferences[4] can be updated, however, is directly tied to the frequency with which events occur, and HFOs are rare: typically 3-10 per hour. The potential utility of intra-operative HFO monitoring therefore is questionable. The likelihood of providing instantaneous tracking or control using HFOs in implantable devices seems similarly low, but it is not clear that this would be necessary, as many patients have seizure reoccurrence periods on the order of days.

Improving our understanding of HFO dynamics – the way rates change over time

---

[4]whether made by human or machine

and the spatial trajectory of "hot" regions – is perhaps the biggest scientific challenge remaining, and will help clarify some of the above mentioned issues surrounding real-time applications. The combination of temporally sparse occurrence and high inter-patient variability of HFOs implies that progress will require analyzing long iEEG records in large numbers of patients – volumes of data for which analysis by humans using currently accepted methods is wholly untenable. We hope the automated approach developed in this dissertation will help point the way forward.

Appendix A

Computational Features

A.1 Introduction

Below, we provide a detailed description of the features that were discussed in chapter 3.

A.2 HFO distinguishing features

1. *Power ratio: (250-500Hz)/(100-200Hz)*

   Computed on the bandpassed data; frequency domain input. The ratio of estimated power in the hypothesized "fast-ripple" band to that in the hypothesized "ripple" band. If there are indeed two distinct populations of HFOs whose respective energy is concentrated in these two bands, one would expect the distribution of data along this feature to be strongly bimodal.

$$f_1 = \frac{\hat{P}_{[250,500]}}{\hat{P}_{[100,200]}} \tag{A.1}$$

where

$$\hat{P}_{[a,b]} = \sum_{\{k \in Z \,|\, \zeta(a) \le k \le \zeta(b)\}} |M[k]|^2, \quad (a < b) \in (0, Fs/2) \tag{A.2}$$

$M[k]$ can be a multitaper power spectral density estimate. In our case, to lessen computational burden we use the modified periodogram estimate, $M[k] = \sum_{n=0}^{N-1} w[n]\tilde{x}[n]e^{-j\frac{2\pi}{N}nk}$, $k \in \{0, 1, \cdots, N-1\}$ with an N-point Discrete Fourier Transform (DFT). $\{w[n]\}$ is the Hanning window [56] of length L and $\{\tilde{x}[n]\}$ is the detrended signal, obtained by subtracting the least squares fit of a line from the original length L signal, $\{x[n]\}$. Note the signal $m[n] = w[n]\tilde{x}[n]$, $n \in \{0, 1, \cdots, L-1\}$ is padded with $N-L$ zeros before computing the DFT. $\zeta(u) = \lfloor 0.5 + NTu \rfloor$, with T the sampling period, and $\lfloor \cdot \rfloor$ the "floor" operator, which as used in this context accomplishes rounding to the nearest integer. This handles cases when a and b do not coincide precisely with the frequency values sampled by the DFT. Fs is the sampling frequency.

2. *Spectral centroid*

Computed on the bandpassed data; frequency domain input. The frequency corresponding to the "center of mass" of the spectrum. This feature should distinguish events whose energy is relatively concentrated in either the ripple

or fast ripple sub-band alone from those with narrow components in both bands, with relatively broadband spectra, or with predominant energy in an intermediate band.

$$f_2 = \frac{\sum\limits_{k=0}^{N/2} \frac{k}{NT}|M[k]|^2}{\sum\limits_{k=0}^{N/2} |M[k]|^2} \tag{A.3}$$

3. *Spectral peak*

   Computed on the raw data; frequency domain input. Frequency corresponding to the peak of the estimated power spectral density. This feature should capture some differences in the spectral characteristics of the background signal upon which the detections are superimposed.

$$f_3 = \frac{1}{NT} \operatorname*{argmax}_{k} |M[k]|^2, \quad k \in \{0, 1, \cdots, N/2\} \tag{A.4}$$

4. *Line-length*

   Computed on the raw data, after first-order backward differencing [32, 77]; time domain input. Detections are first detrended and energy normalized by dividing by their Euclidean length. As detections are also of different durations, we normalize by the signal length (i.e. number of samples) as well. This measure was included because it has proven broadly applicable in auto-

124

matic EEG discrimination tasks, such as seizure [29] and gamma oscillation detection [32].

$$f_4 = \frac{1}{L} \sum_{n=0}^{L-2} |x[n+1] - x[n]| \tag{A.5}$$

A.3 Artifact distinguishing features

The three features below were all chosen empirically for their ability to separate a small training set of detections labeled as either "stereotypical artifact" or "other." A characteristic example of each group from this training set is shown in Figure A.1.
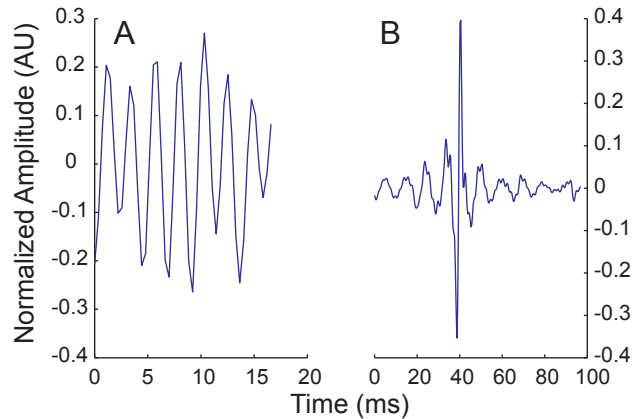
Figure A.1. 100-500 Hz bandpassed examples of human-labeled positive stereotypical artifact (A) and negative stereotypical artifact (B) classes. Note that the class in (B) is actually a heterogeneous "other" category consisting of all non-stereotypical artifacts and hence not as well represented by a single exemplar as the stereotypical artifact class. A small training sample ($n = 500$) of randomly selected stage 1 detections from two patients (CT 01 and SZ 05) was labeled and three features were empirically chosen for their strong ability to distinguish these two classes. The three features are described in the appendix.

1. *Global/average-local peak ratio*

Computed on the bandpassed data; time domain input. Ratio of the global maximum value to the average of all other local maxima (i.e. excluding the global maximum). Detections are first detrended and smoothed using a 3-point moving average filter.

$$f_5 = \frac{g^*}{\frac{1}{J-1} \sum_{\{g_i \in G \,|\, g_i < g^*\}} g_i} \tag{A.6}$$

where $G = \{g_1, \ldots, g_J\}$ is the set of all local maxima and $g^* = max\, G$.

2. *Entropy of the Squared and Normalized Teager Energy Vector*

Computed on the bandpassed data; time domain input. Detections are first detrended and energy normalized by dividing by their Euclidean length. For a signal composed of a single time-varying (or not) frequency and sampled at a sufficiently high rate, the Teager Energy sequence has a value at any time that approximates, to within a constant, the energy required to "generate" the signal using a mechanical spring-and-mass analogy. It is given by:

$$T[n] = \begin{cases} x[n]^2 - x[n+1]x[n-1] & n = \{1, \ldots, L-2\} \\ 0 & otherwise \end{cases} \tag{A.7}$$

[44], where we have imposed our own definition by necessity at the endpoints. We square the signal, making it everywhere positive, and normalize by its sum to make it a pseudo probability mass function before computing its entropy,

126

so that:

$$S[n] = \widehat{T}^2[n] = \frac{T^2[n]}{\displaystyle\sum_{n=0}^{L-1} T^2[n]}, \quad n = \{0, \ldots, L-1\} \tag{A.8}$$

$$f_6 = -\sum_{n=0}^{L-1} S[n] \log_2 S[n] \tag{A.9}$$

3. *Wavelet packet-based energy feature*

Computed on the bandpassed data; time domain input. The signal is first made even length, if necessary, by appending a single sample equal in value to the last sample (sample $L-1$). It is then made length-512 by either truncation or periodized extension on both sides [71], and energy-normalized by dividing by its Euclidean length. A 4-level wavelet packet decomposition using the wavelet packet system generated from the Daubechies ϕ_{D8} scaling function [81] is performed.

For a complete understanding of the wavelet packet decomposition, it is useful to have an appreciation for wavelet theory. The latter topic is too large and complex to attempt tutorial coverage here; excellent reviews can be found in [15, 49, 71]. Below, we attempt only to give sufficient background to make the operational description of our wavelet packet-based energy feature intelligible

to the typical reader with an engineering background.

We can express any continuous-time signal $f(t) \in L^2(R)$ using the expansion:

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0}(k) \, 2^{j_0/2} \, \varphi(2^{j_0}t - k) + \sum_{k=-\infty}^{\infty} \sum_{j=j_0}^{\infty} d_j(k) \, 2^{j/2} \, \psi(2^j t - k) \quad \text{(A.10)}$$

[15] where j_0 is an arbitrary initial scale and $\varphi(t)$, the basic *scaling function*, and $\psi(t)$, the *mother wavelet*, are non-unique functions that obey the following relations:

$$\varphi(t) = \sum_{n=-\infty}^{\infty} h_0(n) \, \sqrt{2} \, \varphi(2t - n), \quad n \in Z \quad\quad \text{(A.11)}$$

and

$$\psi(t) = \sum_{n=-\infty}^{\infty} h_1(n) \, \sqrt{2} \, \varphi(2t - n), \quad n \in Z \quad\quad \text{(A.12)}$$

for some sequences $\{h_0(n)\}$ and $\{h_1(n)\}$. The scaling and wavelet coefficients (i.e. c's and d's, respectively) in (A.10) are called the discrete wavelet transform of $f(t)$, and if the set of expansion functions forms an orthonormal basis (or tight frame [71]), they can be computed as the inner products[1]:

---

[1]In the more general case where the expansion set is only a frame, a *dual frame* can be specified

$$c_j(k) = \int_{-\infty}^{\infty} f(t)\, 2^{j/2}\, \varphi(2^j t - k)\, \mathrm{d}t \qquad\qquad \text{(A.13)}$$

and

$$d_j(k) = \int_{-\infty}^{\infty} f(t)\, 2^{j/2}\, \psi(2^j t - k)\, \mathrm{d}t \qquad\qquad \text{(A.14)}$$

In practice, the DWT coefficients can be computed without explicitly considering the functions in the expansion set, using Mallat's algorithm [49], which achieves the signal decomposition by passing it through a filter bank. The discrete-time wavelet transform (DTWT) coefficients, which arise from an analogous decomposition of a sequence rather than a continuous-valued function, and are what we typically calculate from sampled representations of signals stored in computers, are computed identically. We will not show the DWT/filter bank equivalence here (for a good reference see [79]), but rather simply state how the DWT (and DTWT) are implemented via filter banks, in order to illuminate the wavelet packet decomposition described below.

The scaling and wavelet coefficients at scale $j-1$ can be computed by passing the scaling coefficients at scale j through lowpass and highpass filters, respectively, followed by downsampling by a factor of 2. The impulse responses of the

such that the inner products in (A.13) and (A.14) are identical after substituting the dual frame functions for their counterparts in the original expansion set [15].

lowpass and highpass filters are the time-reversed recursion coefficients from (A.11), $\{h_0(-n)\}$, and from (A.12), $\{h_1(-n)\}$, respectively. At each stage, the wavelet coefficients are "left behind" and the filter bank is iterated on the scaling coefficients only, yielding the set of scaling and wavelet coefficients for the next lower (i.e. coarser) scale. The filtering tree continues down to scale $j = j_0$. For practical signals that are band-limited there will be an upper scale J above which the wavelet coefficients are negligible, and so by having as the initial input the fine scale coefficients c_J (which are approximated by the samples of the signal itself), the filter bank yields an approximation to the full decomposition in (A.10) [15].

If at each stage we iterate the filter bank on both the scaling and the wavelet coefficients – instead of only the scaling coefficients – the resultant full binary tree is known as a wavelet packet decomposition. The effective expansion set analogous to that in (A.10) is now overcomplete, but viewed as a time-scale decomposition of the input signal, the new procedure yields completely evenly spaced frequency (i.e. scale) resolution, in contrast to the logarithmic resolution of the DWT. The cost of this increase in resolution is increased computational complexity ($O(Nlog(N))$ rather than $O(N)$) [15].

If we consider as a "node" of the decomposition tree each sequence of coeffi-

cients that is the output of high- or lowpass filtering followed by downsampling of the signal at the parent node, then a 4-level tree has a total of 31 nodes. For each of the 31 nodes in the binary decomposition tree, we compute the ratio of the node's total energy to the cumulative energy of all nodes at the same level (for the top node, the input signal itself, this trivially equals 1). An energy vector, $\mathbf{E}$, is then formed by concatenating these node relative energies, traversing the tree from top to bottom and left to right. The feature returned, f_7, is the projection of this energy vector along a predetermined direction in 31-dimensional space, $\mathbf{V}_1$, found via PCA on a training set comprised of a subset of detections from two patients with putative artifacts labeled. An energy vector was formed for each detection using the method described above; the first principal component, $\mathbf{V}_1$, was then found and stored for later use in computing f_7:

$$f_7 = \mathbf{E}^T \mathbf{V}_1 \tag{A.15}$$

Similar metrics have been used in developing features for the classification of acoustic transients [46]. To ease the computational burden in the clustering stage, after forming the 7-dimensional feature vectors $\{\mathbf{F}_1, \ldots, \mathbf{F}_I\}$, where I is the total number of detections (anomalies) that pass through stage 2, we form $\{\hat{\mathbf{F}}_1, \ldots, \hat{\mathbf{F}}_I\}$ by projecting onto the first four principal components. This set of 4-dimensional feature vectors is the input to the clustering algorithm described in chapter 3.

Bibliography

[1] AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC-19*, 6 (1974), 716–723.

[2] ALLEN, P., FISH, D., AND SMITH, S. Very high-frequency rhythmic activity during sEEG suppression in frontal lobe epilepsy. *Electroencephalography and Clinical Neurophysiology 82* (1992), 155–159.

[3] BAGSHAW, A. P., JACOBS, J., LEVAN, P., DUBEAU, F., AND GOTMAN, J. Effect of sleep stage on interictal high-frequency oscillations recorded from depth macroelectrodes in patients with focal epilepsy. *Epilepsia 50*, 4 (2009), 617–628.

[4] BARKLEY, G. L., SMITH, B., BERGEY, G., WORRELL, G., CHABOLLA, D., DRAZKOWSKI, J., LABAR, D., DUCKROW, R., MURRO, A., SMITH, M., GWINN, R., FISCH, B., HIRSCH, L., AND MORRELL, M. Safety and preliminary efficacy of the RNS responsive neurostimulator for the treatment of intractable epilepsy in adults. 2006.

[5] BEGLEY, C. E., FAMULARI, M., ANNEGERS, J. F., LAIRSON, D. R., REYNOLDS, T. F., COAN, S., DUBINSKY, S., NEWMARK, M. E., LEIBSON, C., SO, E., AND ROCCA, W. A. The cost of epilepsy in the united states: An estimate from population-based clinical and survey data. *Epilepsia* (2000), 342–351.

[6] BELLMAN, R. *Adaptive Control Processes: A Guided Tour.* Princeton University Press, 1961.

[7] BIEN, C. G., SZINAY, M., WAGNER, J., CLUSMANN, H., BECKER, A. J., AND URBACH, H. Characteristics and surgical outcomes of patients with refractory magnetic resonance imaging negative epilepsies. *Archives of Neurology 66* (2009), 1491–1499.

[8] BISHOP, C. M. *Pattern Recognition and Machine Learning.* Springer Science and Business Media, LLC, New York, NY, 2006.

[9] BRAGIN, A., ENGEL JR., J., WILSON, C. L., FRIED, I., AND BUZSAKI, G. High-frequency oscillations in human brain. *Hippocampus 9* (1999), 137–142.

[10] BRAGIN, A., ENGEL JR., J., WILSON, C. L., FRIED, I., AND MATHERN, G. W. Hippocampal and entorhinal cortex high-frequency oscillations (100-500 Hz) in human epileptic brain and in kainic acid-treated rats with chronic seizures. *Epilepsia 40*, 2 (1999), 127–137.

[11] BRAGIN, A., MODY, I., WILSON, C. L., AND ENGEL JR., J. Local generation of fast ripples in epileptic brain. *The Journal of Neuroscience 22*, 5 (2002), 2012–2021.

[12] BRAGIN, A., WILSON, C. L., ALMAJANO, J., MODY, I., AND ENGEL JR., J. High-frequency oscillations after status epilepticus: Epileptogenesis and seizure genesis. *Epilepsia 45*, 9 (2004), 1017–1023.

[13] BRAGIN, A., WILSON, C. L., STABA, R. J., REDDICK, M., FRIED, I., AND ENGEL JR., J. Interictal high-frequency oscillations (80-500 Hz) in the human epileptic brain: Entorhinal cortex. *Annals of Neurology 52* (2002), 407–415.

[14] BRINKMANN, B. H., BOWER, M. R., STENGEL, K. A., WORRELL, G. A., AND STEAD, M. Large-scale electrophysiology: Acquisition, compression, encryption, and storage of big data. *Journal of Neuroscience Methods 180* (2009), 185–192.

[15] BURRUS, C. S., GOPINATH, R. A., AND GUO, H. *Introduction to Waveletes and Wavelet Transforms: A Primer*. Prentice Hall, 1997.

[16] BUZSAKI, G. High-frequency network oscillation in the hippocampus. *Science 256* (1992), 1025–1027.

[17] BUZSAKI, G., LEUNG, L., AND VANDERWOLF, C. Cellular bases of hippocampal EEG in the behaving rat. *Brain Research 287*, 2 (1983), 139–171.

[18] CHROBAK, J. J., AND BUZSAKI, G. High-frequency oscillations in the output networks of the hippocampal-entorhinal axis of the freely behaving rat. *The Journal of Neuroscience 16*, 9 (1996), 3056–3066.

[19] COHEN, J. A coefficient of agreement for nominal scales. *Education and Psychological Measurement 20*, 1 (1960), 37–46.

[20] CREPON, B., NAVARRO, V., HASBOUN, D., CLEMENCEAU, S., MARTINERIE, J., BAULAC, M., ADAM, C., AND LE VAN QUYEN, M. Mapping interictal oscillations greater than 200 Hz recorded with intracranial macroelectrodes in human epilepsy. *Brain 133* (2010), 33–45.

[21] CSICSVARI, J., HIRASE, H., CZURKO, A., MAMIYA, A., AND BUZSAKI, G. Fast network oscillations in the hippocampal CA1 region of the freely behaving rat. *The Journal of Neuroscience 19*, 16 (1999), RC20.

[22] CSICSVARI, J., HIRASE, H., CZURKO, A., MAMIYA, A., AND BUZSAKI, G. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *The Journal of Neuroscience 19*, 1 (1999), 274–287.

[23] DEMPSTER, A., LAIRD, N., AND RUBIN, D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodological) 39*, 1 (1977), 1–38.

[24] DRAGUHN, A., TRAUB, R. D., BIBBIG, A., AND SCHMITZ, D. Ripple (∼200 Hz) oscillations in temporal structures. *Journal of Clinical Neurophysiology 17*, 4 (2000), 361–376.

[25] DRAGUHN, A., TRAUB, R. D., SCHMITZ, D., AND JEFFERYS, J. G. R. Electrical coupling underlies high-frequency oscillations in the hippocampus in vitro. *Nature 394* (1998), 189–192.

[26] DUDA, R., AND HART, P. *Pattern Classification And Scene Analysis.* John Wiley and Sons, New York, 1973.

[27] DZHALA, V. I., AND STALEY, K. J. Mechanisms of fast ripples in the hippocampus. *The Journal of Neuroscience 24*, 40 (2004), 8896–8906.

[28] ENGEL JR., J., BRAGIN, A., STABA, R., AND MODY, I. High-frequency oscillations: What is normal and what is not? *Epilepsia 50*, 4 (2009), 598–604.

[29] ESTELLER, R., ECHAUZ, J., TCHENG, T., LITT, B., AND PLESS, B. Line length: An efficient feature for seizure onset detection. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2001), vol. 2, pp. 1707–1710.

[30] FISHER, R. S., WEBBER, W. R. S., LESSER, R. P., ARROYO, S., AND UEMATSU, S. High-frequency EEG activity at the start of seizures. *Journal of Clinical Neurophysiology 9*, 3 (1992), 441–448.

[31] FOFFANI, G., UZCATEGUI, Y. G., GAL, B., AND MENENDEZ DE LA PRIDA, L. Reduced spike-timing reliability correlates with the emergence of fast ripples in the rat epileptic hippocampus. *Neuron 55* (2007), 930–941.

[32] GARDNER, A. B., WORRELL, G. A., MARSH, E., DLUGOS, D., AND LITT, B. Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. *Clinical Neurophysiology 118*, 5 (2007), 1134–1143.

[33] GRENIER, F., TIMOFEEV, I., AND STERIADE, M. Focal synchronization of ripples (80-200 Hz) in neocortex and their neuronal correlates. *Journal of Neurophysiology 86* (2001), 1884–1898.

[34] GRENIER, F., TIMOFEEV, I., AND STERIADE, M. Neocortical very fast oscillations (ripples, 80-200 Hz) during seizures: Intracellular correlates. *Journal of Neurophysiology 89* (2003), 841–852.

[35] HABERMAN, R. P., SAMULSKI, R. J., AND McCOWN, T. J. Attenuation of seizures and neuronal death by adeno-associated virus vector galanin expression and secretion. *Nature Medicine 9*, 8 (2003), 1076–1080.

[36] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The Elements of Statistical Learning.* Springer-Verlag, New York, 2001.

[37] HOTELLING, H. Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology 24* (1933), 417–441.

[38] Hubert, L., and Arabie, P. Comparing partitions. *Journal of Classification 2* (1985), 193–218.

[39] Jacobs, J., LeVan, P., Chander, R., Hall, J., Dubeau, F., and Gotman, J. Interictal high-frequency oscillations (80-500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. *Epilepsia 49*, 11 (2008), 1893–1907.

[40] Jacobs, J., LeVan, P., Chatillon, C.-E., Olivier, A., Dubeau, F., and Gotman, J. High frequency oscillations in intracranial EEGs mark epileptogenicity rather than lesion type. *Brain 132* (2009), 1022–1037.

[41] Jacobs, J., Zelmann, R., Jirsch, J., Chander, R., Chatillon, C.-E., Dubeau, F., and Gotman, J. High frequency oscillations (80-500 Hz) in the preictal period in patients with focal seizures. *Epilepsia 50*, 7 (2009), 1780–1792.

[42] Jirsch, J., Urrestarazu, E., LeVan, P., Olivier, A., Dubeau, F., and Gotman, J. High-frequency oscillations during human focal seizures. *Brain 129* (2006), 1593–1608.

[43] Joo, E. Y., Han, S. J., Chung, S.-H., Cho, J.-W., Seo, D. W., and Hong, S. B. Antiepileptic effects of low-frequency repetitive transcranial magnetic stimulation by different stimulation durations and locations. *Clinical Neurophysiology 118* (2007), 702–708.

[44] KAISER, J. On a simple algorithm to calculate the energy of a signal. In *International Conference on Acoustics, Speech, and Signal Processing* (Albuquerque, NM, USA, 1990), vol. 1, pp. 381–384.

[45] KWAN, P., AND BRODIE, M. Early identification of refractory epilepsy. *The New England Journal of Medicine 342*, 5 (2000), 314–319.

[46] LEARNED, R. E., AND WILLSKY, A. S. A wavelet packet approach to transient signal classification. *Applied and Computational Harmonic Analysis 2* (1995), 265–278.

[47] MACKAY, D. J. Laplace's method. In *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2005, pp. 341–342.

[48] MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, CA, 1967), vol. 1, University of California Press, pp. 281–297.

[49] MALLAT, S. G. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*, 7 (1989), 674–693.

[50] MCGILL, R., TUKEY, J. W., AND LARSEN, W. A. Variations of box plots. *The American Statistician 31*, 1 (1978), 12–16.

[51] McLachlan, G. Mahalanobis distance. *Resonance 4*, 6 (1999), 20–26.

[52] Medtronic: Intercept epilepsy control system clinical trial, 2004. `http://www.epilepsycontrol.com`.

[53] Mitra, P., and Pesaran, B. Analysis of dynamic brain imaging data. *Biophysical Journal 76* (1999), 691–708.

[54] Nuttall, A. H. Near-optimum detection performance of power-law processors for random signals of unknown locations, structure, extent, and arbitrary strengths. *Technical Report ADA309568: Naval Undersea Warfare Center Newport Division; New London, CT, New London Detachment* (1996), 162.

[55] O'Keefe, J. Place units in the hippocampus of the freely moving rat. *Experimental Neurology 51* (1976), 78–109.

[56] Oppenheim, A., Schafer, R., and Buck, J. *Discrete-Time Signal Processing*, vol. 2nd. Prentice Hall, Upper Saddle River, NJ, 1999.

[57] Osorio, I., Frei, M. G., Sunderam, S., Giftakis, J., Bhavaraju, N. C., Schaffner, S. F., and Wilkinson, S. B. Automated seizure abatement in humans using electrical stimulation. *Annals of Neurology 57* (2005), 258–268.

[58] Pearson, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine 2*, 6 (1901), 559–572.

[59] Percival, D., and Walden, A. *Spectral Analysis For Physical Applications: Multitaper and Conventional Univariate Techniques.* Cambridge University Press, Cambridge, UK, 1993.

[60] Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association 66*, 336 (1971), 846–850.

[61] Roberts, S. J. Extreme value statistics for novelty detection in biomedical signal processing. *IEEE Proceedings Science, Technology, & Measurement 47* (2000), 363–367.

[62] Sander, J., and Shorvon, S. Epidemiology of the epilepsies. *Journal of Neurology, Neurosurgery, and Psychiatry 61* (1996), 433–443.

[63] SANTE: Stimulation of the anterior nucleus of the thalamus for epilepsy, 2003. `http://clinicaltrials.gov/ct/show/NCT00101933`.

[64] Schevon, C., Trevelyan, A., Schroeder, C., Goodman, R., McKhann Jr., G., and Emerson, R. Spatial characterization of interictal high frequency oscillations in epileptic neocortex. *Brain 132* (2009), 3047–3059.

[65] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics 6*, 2 (1978), 461–464.

[66] Slepian, D., and Pollak, H. Prolate spheroidal wavefunctions, Fourier analysis and uncertainty. I. *Bell System Technical Journal 40* (1961), 43–63.

[67] Staba, R. J., Fighetto, L., Behnke, E. J., Mathern, G. W., Fields, T., Bragin, A., Ogren, J., Fried, I., Wilson, C. L., and Engel Jr., J. Increased fast ripple to ripple ratios correlate with reduced hippocampal volumes and neuron loss in temporal lobe epilepsy patients. *Epilepsia 48*, 11 (2007), 2130–2138.

[68] Staba, R. J., Wilson, C. L., Bragin, A., Fried, I., and Engel Jr., J. Quantitative analysis of high-frequency oscillations (80-500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. *Journal of Neurophysiology 88* (2002), 1743–1752.

[69] Stacey, W. C., and Litt, B. Technology insight: Neuroengineering and epilepsy–designing devices for seizure control. *Nature Clinical Practice Neurology 4*, 4 (2008), 190–201.

[70] Stein, A. G., Eder, H. G., Blum, D. E., Drachev, A., and Fisher, R. S. An automated drug delivery system for focal epilepsy. *Epilepsy Research 39* (2000), 103–114.

[71] Strang, G., and Nguyen, T. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.

[72] Tanriverdi, T., Olivier, A., Poulin, N., Andermann, F., and Dubeau, F. Long-term seizure outcome after mesial temporal lobe epilepsy

surgery: Corticalamygdalohippocampectormy versus selective amygdalohip-pocampectomy. *Journal of Neurosurgery 108* (2008), 517–524.

[73] THOMSON, D. J. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE 9*, 9 (1982), 1055–1096.

[74] TIBSHIRANI, R., WALTHER, G., AND HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B 63*, 2 (2001), 411–423.

[75] TRAUB, R. D. Fast oscillations and epilepsy. *Epilepsy Currents 3*, 3 (2003), 77–79.

[76] URRESTARAZU, ELENA CHANDER, R., DUBEAU, F., AND GOTMAN, J. Interictal high-frequency oscillations (100-500 Hz) in the intracerebral EEG of epileptic patients. *Brain 130* (2007), 2354–2366.

[77] USUI, S., AND AMIDROR, I. Digital low-pass differentiation for biological signal processing. *IEEE Transactions on Biomedical Engineering BME-29*, 10 (1982), 686–693.

[78] VAN GOMPEL, J. J., STEAD, S. M., GIANNINI, C., MEYER, F. B., MARSH, W. R., FOUNTAIN, T., SO, E., COHEN-GADOL, A., LEE, K. H., AND WORRELL, G. A. Phase i trial: Safety and feasibility of intracranial electroencephalography using hybrid subdural electrodes containing macro- and microelectrode arrays. *Neurosurgical Focus 25*, 3:E23 (2008), 1–6.

[79] VETTERLI, M., AND HERLEY, C. Wavelets and filter banks. *IEEE Transactions on Signal Processing 40*, 9 (1992), 2207–2232.

[80] WANG, Z., AND WILLETT, P. K. All-purpose and plug-in power-law detectors for transient signals. *IEEE Transactions On Signal Processing 49*, 11 (2001), 2454–2466.

[81] WICKERHAUSER, M. V. *Adapted Wavelet Analysis from Theory to Software.* A.K. Peters, Ltd., Wellesley, MA, 1994.

[82] WILSON, M. A., AND MCNAUGHTON, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science 265* (1994), 676–679.

[83] WORRELL, G. A., GARDNER, A. B., STEAD, S. M., HU, S., GOERSS, S., CASCINO, G. J., MEYER, F. B., MARSH, R., AND LITT, B. High-frequency oscillations in human temporal lobe: Simultaneous microwire and clinical macroelectrode recording. *Brain 131* (2008), 928–937.

[84] YANG, X.-F., AND ROTHMAN, S. M. Focal cooling rapidly terminates experimental neocortical seizures. *Annals of Neurology 49* (2001), 721–726.

[85] YLINEN, A., BRAGIN, A., NADASDY, Z., JANDO, G., SZABO, I., SIK, A., AND BUZSAKI, G. Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: Network and intracellular mechanisms. *The Journal of Neuroscience 15*, 1 (1995), 30–46.