



Generalizability of High Frequency Oscillation Evaluations in the Ripple Band

Aaron M. Spring^{1,2,3}, Daniel J. Pittman^{2,3}, Yahya Aghakhani¹, Jeffrey Jirsch⁴, Neelan Pillay¹, Luis E. Bello-Espinosa^{1,5}, Colin Josephson¹ and Paolo Federico^{1,2,3,6*}

¹ Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada, ² Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada, ³ Seaman Family MR Research Centre, Foothills Medical Centre, Calgary, AB, Canada, ⁴ Department of Medicine, University of Alberta, Edmonton, AB, Canada, ⁵ Department of Paediatrics, University of Calgary, Calgary, AB, Canada, ⁶ Department of Radiology, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:

Fernando Cendes,
Universidade Estadual de Campinas,
Brazil

Reviewed by:

Benjamin Brinkmann,
Mayo Clinic, United States
Andy P. Bagshaw,
University of Birmingham,
United Kingdom

*Correspondence:

Paolo Federico
pfederic@ucalgary.ca

Specialty section:

This article was submitted to
Epilepsy,
a section of the journal
Frontiers in Neurology

Received: 03 December 2017

Accepted: 11 June 2018

Published: 28 June 2018

Citation:

Spring AM, Pittman DJ, Aghakhani Y, Jirsch J, Pillay N, Bello-Espinosa LE, Josephson C and Federico P (2018) Generalizability of High Frequency Oscillation Evaluations in the Ripple Band. *Front. Neurol.* 9:510. doi: 10.3389/fneur.2018.00510

Objective: We examined the interrater reliability and generalizability of high-frequency oscillation (HFO) visual evaluations in the ripple (80–250 Hz) band, and established a framework for the transition of HFO analysis to routine clinical care. We were interested in the interrater reliability or epoch generalizability to describe how similar the evaluations were between reviewers, and in the reviewer generalizability to represent the consistency of the internal threshold each individual reviewer.

Methods: We studied 41 adult epilepsy patients (mean age: 35.6 years) who underwent intracranial electroencephalography. A morphology detector was designed and used to detect candidate HFO events, lower-threshold events, and distractor events. These events were subsequently presented to six expert reviewers, who visually evaluated events for the presence of HFOs. Generalizability theory was used to characterize the epoch generalizability (interrater reliability) and reviewer generalizability (internal threshold consistency) of visual evaluations, as well as to project the numbers of epochs, reviewers, and datasets required to achieve strong generalizability (threshold of 0.8).

Results: The reviewer generalizability was almost perfect (0.983), indicating there were sufficient evaluations to determine the internal threshold of each reviewer. However, the interrater reliability for 6 reviewers (0.588) and pairwise interrater reliability (0.322) were both poor, indicating that the agreement of 6 reviewers is insufficient to reliably establish the presence or absence of individual HFOs. Strong interrater reliability (≥ 0.8) was projected as requiring a minimum of 17 reviewers, while strong reviewer generalizability could be achieved with < 30 epoch evaluations per reviewer.

Significance: This study reaffirms the poor reliability of using small numbers of reviewers to identify HFOs, and projects the number of reviewers required to overcome this limitation. It also provides a set of tools which may be used for training reviewers, tracking changes to interrater reliability, and for constructing a benchmark set of epochs that can serve as a generalizable gold standard, against which other HFO detection algorithms

may be compared. This study represents an important step toward the reconciliation of important but discordant findings from HFO studies undertaken with different sets of HFOs, and ultimately toward transitioning HFO analysis into a meaningful part of the clinical epilepsy workup.

Keywords: high frequency oscillations (HFOs), generalizability, generalizability theory, interrater reliability, interrater variability, epilepsy, intracranial electroencephalography (iEEG)

INTRODUCTION

In the treatment of drug-resistant focal epilepsy, the localization and removal of the region generating the seizures is critical to the successful elimination of seizures. It has been proposed that this region—the so-called epileptogenic zone (EZ)—may be identified using high frequency oscillations (HFOs) (1, 2), or interictal epileptiform discharges (IEDs) occurring simultaneously with HFOs (HFO+IEDs) (3, 4). Extensive research has been performed with regards to the characteristics of these HFOs and their correlation with epileptogenicity and the EZ. Nonetheless, the lack of generalizability and comparability between the various studies remains an obstacle to the implementation of HFOs prospectively in clinical practice (5).

There is clear and mounting evidence of a correlation between the EZ and HFOs. A meta-analysis revealed a significant correlation between the resection of tissue exhibiting HFOs and the absence of seizures following surgery (6). However, the findings of the individual retrospective studies within the meta-analysis, as well as subsequent studies, have varied substantially. Some found that the resection of HFOs in the ripple range (~80–250 Hz) correlated more strongly with positive outcomes than did the resection of HFOs in the fast ripple range (FR; ~250–500 Hz) (1), while others found that the opposite was true (7). There were studies that found a significant effect of resecting either ripples or FRs (2, 8), and more still that did not find any significant effect for either (9, 10). One study identified ripples to be superior at positive prediction, and FRs at negative prediction (11). More recently, it was shown that in some brain regions, the amplitudes or rates of HFOs are insufficient to distinguish the EZ from baseline activity (12).

There is also contradicting data regarding the role of HFOs occurring in isolation or those co-occurring with IEDs. There is some evidence that resection of HFO+IEDs are correlated with positive outcomes, while that of HFOs in isolation are not (13). There is also evidence that all HFOs are correlated with the EZ regardless of morphology (14), and other studies still have aimed to exclude HFOs generated by filtering IEDs and other sharp transients (15, 16). A recent study also provided evidence that the co-occurrence of ripples and FRs would be more useful than either alone in delineating the EZ (17), though this has yet to be reported in other studies.

Therefore, while there is overwhelming evidence confirming that HFOs are correlated with the EZ, it remains unclear which HFO characteristics would be most useful in prospectively identifying the EZ for surgical resection. Though prospective

studies are ongoing (9, 18), a Cochrane review recently cited lack of evidence in concluding that “no reliable conclusions can be drawn regarding the efficacy of using HFOs in epilepsy surgery decision making at present” (19). Indeed, the different findings of all of these studies should first be reconciled.

That the results from the many studies undertaken in this field differ is not surprising – these studies have been undertaken using variable definitions of HFOs, methods of detection, and datasets. Many have relied upon visual review by one (3, 4, 14, 20–27) or two (1, 15, 16, 28–41) experts, and up to four experts have been used in rodents (42). However, while it has been shown that experts have moderate to strong agreement with regards to classifying candidate events as neural or artefactual in origin (43), it has also been shown that they have poor interrater reliability for classifying candidate events as HFOs (44) or gamma oscillations (45) in the first place.

In an effort to decrease the subjectivity and increase the speed of HFO identification, several automated or semi-automated detection algorithms have been developed. Some have been used in conjunction with expert review to obtain a set of HFO markings (43, 46–51), but many have been developed for use in isolation.

Several features have been proposed to quantify high frequency oscillatory activity, including amplitude threshold (52) or envelope (53), line length (33, 45), instantaneous frequency (33), conventional energy (45, 54), wavelet transforms (32), Stockwell entropy (11), non-harmonicity (50) and Teager energy (55). These features may be considered alone, or may be incorporated into a machine learning algorithm (33). Other algorithms have been developed based on pattern matching (23, 34, 42, 53); integration of data from multiple channels (11, 53); empirical mode decomposition (56); Gabor atoms (41); Gaussian mixture model clustering (57); topographic analysis (16, 51); independent component analysis (51); and integration of a rigorous artifact rejection step following HFO detection (43).

The performance of these algorithms may only be evaluated against some determined baseline. In theory, the ideal algorithm would be able to identify HFOs that perfectly delineate the EZ in prospective studies. In practice, this is not a widely feasible comparison, and even comparisons with retrospective surgical outcomes (11, 50) are rarely performed, so existing benchmarks are often used. Whether these benchmarks are other algorithms (43, 52), visual review (16, 32–34, 36, 41), or some combination thereof, the result is the same—the comparison of an algorithm against a set of markings that itself has not been validated in a generalizable, reliable manner, and therefore, a comparison of unknown clinical significance.

In the present study, we seek to establish a protocol for overcoming this critical limitation of both visual and algorithmic HFO identification. To accomplish this, we turn to a statistical framework called generalizability theory. The use of generalizability theory has been commonplace in the education literature for decades, but has only recently been implemented in neurology (58). It not only allows the characterization of interrater reliability in a manner that is less subject to bias than traditional metrics (59), but also predicts how changing the sample size could affect the generalizability (60). It may be used to determine how to create a dataset that would achieve strong generalizability in practice.

The present study makes use of generalizability theory to establish a framework for optimizing visual HFO markings, in order to reconcile the findings from diverse HFO studies and facilitate the transition of HFO evaluation to routine clinical care.

METHODS

This study was approved by our local Research Ethics Board. Forty-one consecutive adult patients (mean age: 35.6 years) were recruited, all with drug-resistant focal epilepsy and undergoing intracranial video-EEG monitoring (iVEM) at high sampling rates (1,000–2,000 Hz) for possible surgical candidacy at our epilepsy center. All patients were included in the study, regardless of the types or locations of the implanted intracranial electrodes, or of the type or presumed localization of the epileptogenic activity.

Data were initially processed as per the methodology detailed in Spring et al. (44). Twenty minutes of iEEG data were selected, filtered (80–250 Hz), derived (bipolar or Laplacian), and normalized (sliding 1 s root-mean-square). As in the previous study, the data were selected as close as possible to midnight, as close as possible to the fifth day postimplantation, in order to reduce artifact. No preference was given to stage of sleep or wakefulness due to the lack of concurrent scalp EEG recordings. Notably, this study focuses on HFOs in the ripple band (80–250 Hz), and the terms HFOs and ripples are used synonymously throughout this text, except where noted otherwise.

Three types of events were algorithmically detected from the normalized data: candidate HFO events, low-threshold HFO events, and distractor events. From each dataset (patient), 64 events of each type were pseudorandomly selected, resulting in a total of 7,872 events (64 events per type \times 3 event types \times 41 patients = 7,872 events). Filtered data (80–250 Hz) were presented to six visual reviewers as a series of epochs (250 ms), each containing the entirety of one event. Three seconds of unfiltered data were also provided simultaneously, and included the same 250 ms epoch, as well as the preceding and following 1,375 ms of data. In each case, the data presented included the target channel (in which the event was detected), the two nearest neighboring channels, and four channels randomly selected from those remaining.

A detailed overview of the evaluation program and process are published in our previous work (44), and a screenshot of the program is shown in **Figure 1**. The six reviewers (YA, JJ, NP,

LB, CJ, PF) hailed from two epilepsy centers, and had varying degrees of experience evaluating HFOs, as detailed previously (44). They were instructed to identify HFOs that stood out from the surrounding baseline for at least 3 consecutive cycles. They also had the opportunity to mark the presence of any artifacts that they believed affected the presence or interpretation of an HFO.

HFO Rating

Each reviewer registered an HFO rating for each epoch. The rating was on a scale from -5 to $+5$, with the magnitude corresponding to the given confidence rating (1–5, with 5 indicating complete certainty), and the sign corresponding to the HFO evaluation of present (+) or absent (–). This rating reflects the likelihood of each epoch containing an HFO, as determined by each reviewer, and enables a more robust comparison of the relative “stringency” of the reviewers. In essence, stringency describes the probability of a reviewer positively marking an HFO of a given threshold. For example, a more stringent reviewer would typically assign lower ratings than a less stringent colleague, as they have a higher internal threshold for marking HFOs.

Evaluation Time

The evaluation program recorded the time each reviewer spent on each epoch evaluation. The mean evaluation time was then calculated for each reviewer, before and after eliminating outliers (lower and upper 5% of each reviewer’s times).

Generalizability Studies

An initial generalizability study (G-study) was performed to characterize the observed interrater reliability and generalizability. It was also performed to determine the consistency of the internal threshold employed by each reviewer in their evaluation of potential HFOs. In all cases, the HFO ratings made for each epoch by each reviewer were used as the G-study “measurement.”

Models

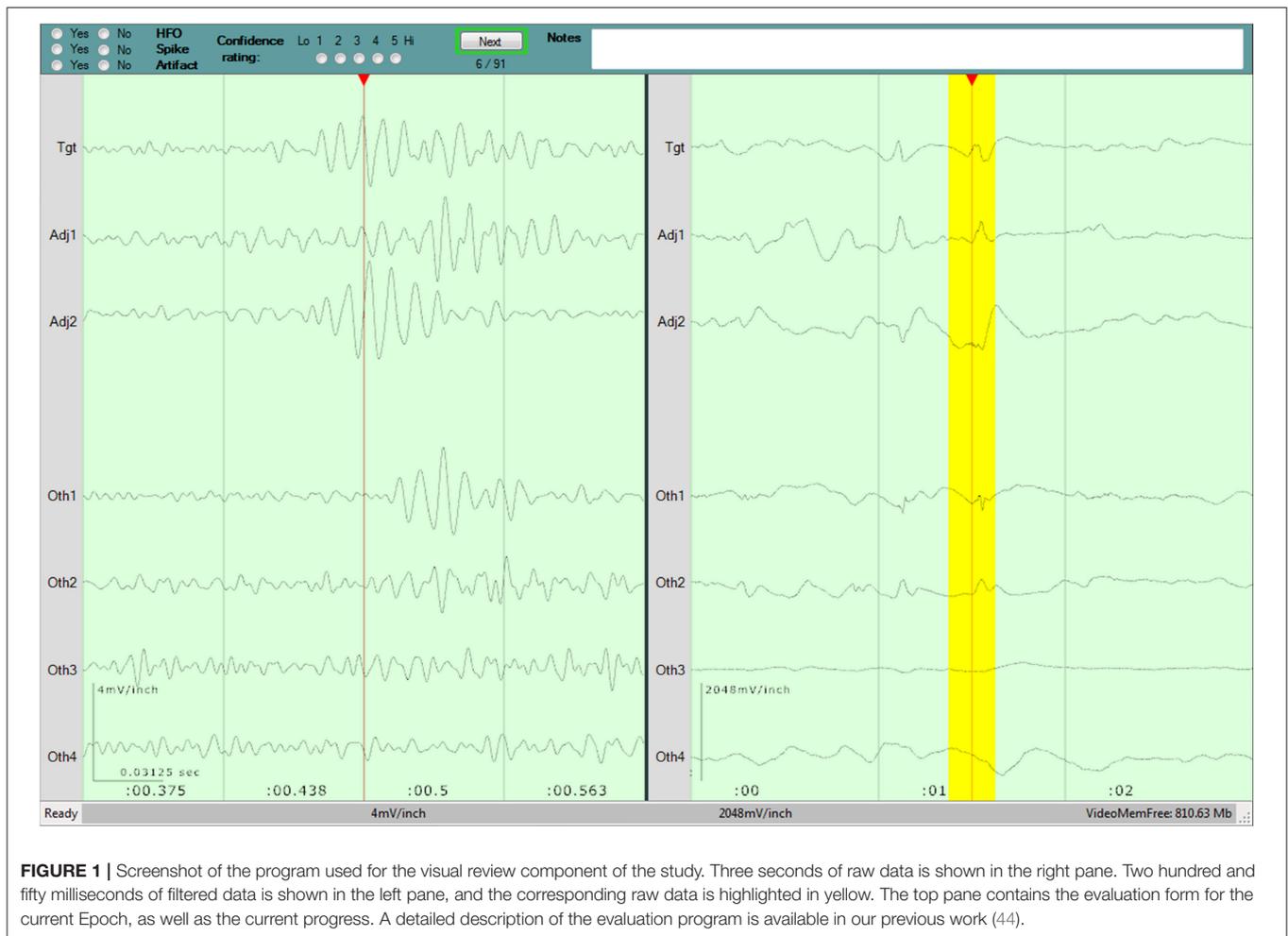
A nested mixed-effects model was used for the G-study analyses. The Image 1 in the Supplementary Material shows a Venn diagram depicting the primary effects and their interactions:

- Reviewer (r), the clinician evaluating the data for HFOs
- Dataset (d), the patient from whom the data were collected
- EventType (t), the type of event used to generate the epoch
- $e:[d \cdot t]$, the individual Epoch, nested within Dataset and EventType
- The interactions amongst the above effects

The generalizability theory model is expressed as:

$$X_{rdte} = \mu + \nu_r + \nu_d + \nu_t + \nu_{r,d} + \nu_{d,t} + \nu_{r,t} + \nu_{r,d,t} + \nu_{e:[d,t]} + \nu_{r,e:[d,t]}$$

where X_{rdte} is the HFO rating given by Reviewer r to Epoch e of Dataset d and EventType t ; μ is the grand mean HFO rating; and ν_α is the score effect for any arbitrary effect α .



Objects of Measurement

In generalizability theory, any one of the effects can be declared the “object of measurement,” while the other effects constitute the “facets” of variability. As the names imply, the object of measurement is what we are measuring, while the facets are effects to which we can attribute the variability of our measurements. The assignment of the object of measurement depends upon the question of interest. This study was designed to determine how well one could answer two questions:

1. What is the likelihood that Epoch e_i is an HFO?
2. What is the stringency of Reviewer r_i in evaluating Epochs for HFOs?

In order to address these questions, two separate objects of measurement were used. For Question 1, the apparent object of measurement was Epoch; however, since Epoch is nested within Dataset and Type, the object of measurement was actually $e:[d \cdot t]$. For Question 2, the object of measurement was simply Reviewer (r). It is important to note the primary objective of the study is not to provide answers to these questions with the present sample sizes, but rather to quantify and predict how well the answers obtained would generalize to the population of epochs, reviewers, or patients.

Generalizability Coefficient Calculation

A minimum norm quadratic unbiased estimator (61, 62) was used to calculate the variance components, which were in turn used to derive the generalizability coefficients. Details regarding the calculations are provided in the Generalizability Theory Details section. A threshold of 0.8 was used to indicate high generalizability or interrater reliability (44).

Epoch Generalizability—Interrater Reliability

For Epoch generalizability, the coefficient is a measure of how well the consensus HFO rating given by a set of Reviewers would generalize to the universe of potential Reviewers. It is an indicator of interrater reliability: a coefficient of 1.0 indicates that the HFO rating would generalize completely, while a coefficient close to 0.0 would indicate that it would not generalize. Optimizing this coefficient would reflect increasing the overall interrater reliability, and allow for the generation of a set of “universal” HFO Epoch evaluations.

Reviewer Generalizability—Internal Reviewer Threshold

The generalizability coefficient of the Reviewer object is a measure of how well the relative observed stringency of

Reviewers would generalize to the universe of potential Datasets and Epochs. A coefficient approaching 1.0 indicates complete generalizability of Reviewer stringency, which would permit the confident relative ranking of Reviewers in terms of their probability of positively identifying an HFO of a given threshold. A low coefficient, on the other hand, would indicate that such a ranking could not be reliably made. Such rankings could be useful for clinical purposes, as in ensuring that a selected set of Reviewers would not be skewed toward a high or low stringency. They would be even more valuable for training purposes: Reviewers with a lower stringency could be trained to increase their internal threshold, while those with a higher stringency could be trained to decrease it.

Decision Study Projections

In order to determine how the generalizability of each object of measurement would be affected by changes in the other facets, a decision study (D-Study) projection was performed for each object of measurement. The generalizability of the Epoch rating was projected across 5–300 Datasets and 2–20 Reviewers, and the generalizability of the Reviewer rating was projected across 1–3,000 Epochs and 5–300 Datasets. This enables us to predict how many datasets, reviewers, or epochs would be needed for us to achieve strong generalizability; in other words, it is used to predict the number of reviewers that would be needed for studies interested in determining the presence of HFOs, or the number of epochs and datasets that would be needed for studies interested in determining the internal stringency of reviewers.

Generalizability Theory Details

The following section contains the details regarding the definitions and calculations involved in the derivation of the generalizability coefficients.

Variance Components

The variance components (σ_a^2 for facet α) were estimated within SPSS. Specifically, the VARCOMP procedure was implemented using a minimum norm quadratic unbiased estimator [MINQUE; (61, 62)], with an intercept term and uniform weight assignment. These variance components were independent of the object of measurement.

The normalized variance components (σ_A^2 for facet α) were then computed by normalizing the variances component by the sample sizes (n_a) of all effects in the given facet, regardless of whether the effects are crossed or nested. Sample formulae are illustrated below:

$$\sigma_A^2 = \frac{\sigma_a^2}{n_a} \quad \sigma_{A:B}^2 = \frac{\sigma_{a:b}^2}{n_a \cdot n_b} \quad \sigma_{A \cdot B}^2 = \frac{\sigma_{a \cdot b}^2}{n_a \cdot n_b}$$

The sample size for the object of measurement was set to 1, so this process was repeated for each object of measurement.

Object and Residual Variances

The object of measurement variance (σ_e^2) is the variance attributable to the object of measurement. The object variances of all objects were estimated by summing the variance component of

the object of measurement with those of any interactions between the object of measurement and fixed facets only:

$$\sigma_e^2(e) = \sigma_{e:[D:T]}^2 \quad \sigma_e^2(r) = \sigma_r^2 + \sigma_{r:T}^2$$

The relative residual variance (σ_δ^2) is the variance attributable to the interaction between the object of measurement and the random effects. The relative residual variances of all objects were estimated by summing the variance components of interactions between the object of measurement and at least one random facet:

$$\sigma_\delta^2(e) = \sigma_{R:e[D:T]}^2 \\ \sigma_\delta^2(r) = \sigma_{r:D}^2 + \sigma_{r:D:T}^2 + \sigma_{r:E[D:T]}^2$$

The absolute residual variance (σ_Δ^2) is the variance not solely attributable to the object of measurement. The absolute residual variances of all objects were estimated by summing the variance components of all facets excluded from σ_e^2 :

$$\sigma_\Delta^2(e) = \sigma_R^2 + \sigma_D^2 + \sigma_{R:D}^2 + \sigma_{R:T}^2 + \sigma_{D:T}^2 + \sigma_{R:D:T}^2 \\ + \sigma_{R:e[D:T]}^2 \\ \sigma_\Delta^2(r) = \sigma_D^2 + \sigma_{r:D}^2 + \sigma_{r:T}^2 + \sigma_{D:T}^2 + \sigma_{r:D:T}^2 + \sigma_{E:[D:T]}^2 \\ + \sigma_{r:E[D:T]}^2$$

Generalizability Coefficients

Two generalizability coefficients may be estimated for each of object of measurement: A relative generalizability coefficient (ρ^2) which describes how well the relative measurements amongst the objects generalize to the universe of facets, and an absolute dependability coefficient (φ^2) which describes how well the absolute scores of the objects generalize to the universe. These may be estimated from the object variance as well as the relative or absolute residual variance, respectively, by:

$$\rho^2 = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_\Delta^2} \quad \varphi^2 = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_\Delta^2}$$

In the present study, the relative measurements are of more interest compared to the absolute measurements: the likelihood of an epoch containing an HFO compared to other epochs, or the likelihood of a reviewer marking an HFO compared to other reviewers, are of greater interest than those likelihoods relative to some arbitrary value. As such, the following discussions are limited to the relative generalizability coefficients.

RESULTS

In total, 41,065 individual Epoch evaluations were made. A summary of the completed and missing evaluations is provided in the Supplementary Material (Data Sheet 1).

Evaluation Time

The trimmed mean evaluation time (excluding upper and lower 5% of values as outliers) for the six Reviewers were 7.15, 14.51, 4.32, 11.45, 4.93, and 8.43 s per Epoch, (overall mean = 7.43 s).

Generalizability

The derivations of the G-study coefficients are outlined in **Table 1**. The generalizability coefficients were calculated to be 0.588 for Epoch and 0.983 for Reviewer.

Decision Study

Epoch Generalizability

The D-study projections for Epoch generalizability are displayed in **Figure 2**. The projections increased with the number of Reviewers, reaching the threshold of 0.8 with 17 Reviewers, and were independent of the number of Datasets.

Reviewer Generalizability

The D-study projections for Reviewer generalizability are displayed in **Figure 3**. The Reviewer generalizability projections increased with the number of Epochs and Datasets. A trend of diminishing returns was demonstrated, wherein the generalizability reached a plateau beyond which increases in the number of Epochs did not yield an appreciable increase in generalizability. For example, the threshold of 0.8 was achieved with 1 Epoch per EventType per Dataset in 10 Datasets, or with 3 Epochs per EventType per Dataset in 5 Datasets.

DISCUSSION

The generalizability studies reaffirm the poor interrater reliability typically observed between two reviewers (44), and the projections indicate that a prohibitive number of reviewers are required to achieve strong generalizability. The results also illustrate the strong Reviewer generalizability and the relatively

small evaluation times within the novel epoched framework. Together, these findings provide further evidence for the need for a gold standard for HFO identification, and a framework for its establishment.

Evaluation Time

There was a notable variability amongst the Reviewers' mean evaluation times. This is unsurprising, as evaluation times also vary in other medical imaging applications (63, 64). Further investigation into the nature of the differences between Reviewers could assist the optimization of HFO evaluations for any given Reviewer.

It is equally unsurprising but notable that evaluation of HFOs using this supervised framework is substantially faster than continuous review. One of the Reviewers reported that prior to this study, he often required 8 h to review 5 min of continuous EEG data for HFOs in multiple channels. Using the trimmed mean evaluation time of 7.43 s, he would be able to evaluate 3876 Epochs in those same 8 h. Even considering our most active HFO patient, who exhibited 278.7 algorithmically-detected events per minute (across all event types, and all electrodes), this Reviewer would only have to evaluate 1394 Epochs from 5 min of data, which would take only 2.87 h. In the case of our second most active patient, he would only have to evaluate 350 Epochs (<45 min). Thus, he would complete the evaluations in substantially less time compared to continuous review in most cases. Furthermore, it is likely that he would have to evaluate only a fraction of the Epochs, further reducing the time burden.

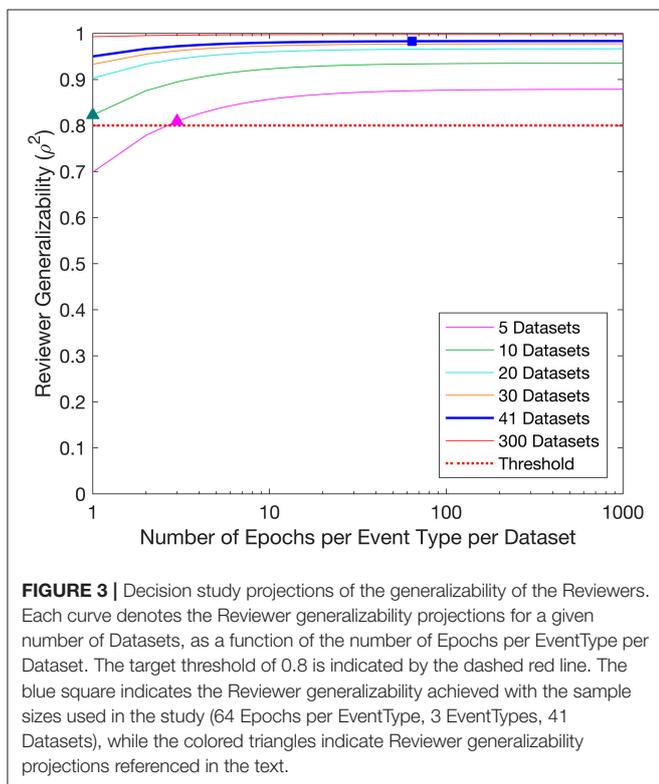
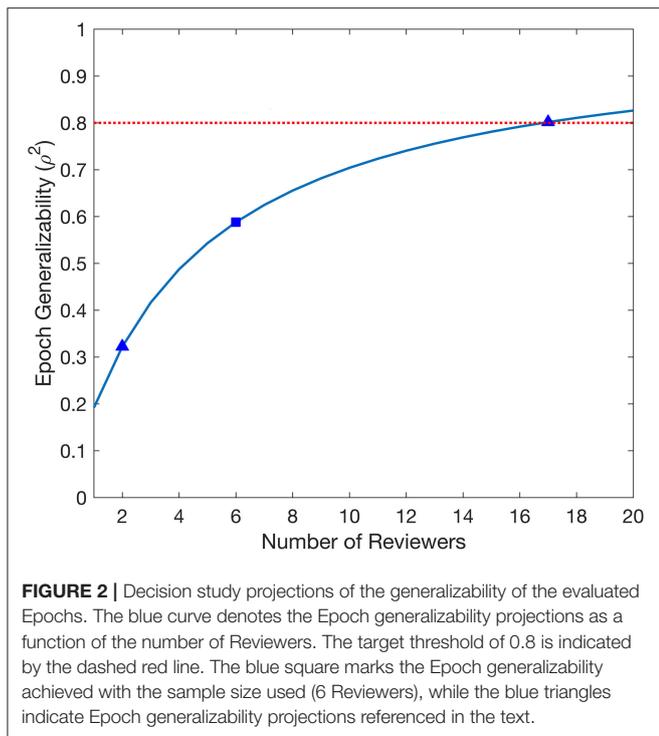
This anecdotal evidence is further supported by a recent study reporting that approximately 5 h were required review

TABLE 1 | Derivation of the generalizability coefficients for model considering all EventTypes.

(A)							
Facet α	Object: Epoch					Object: Reviewer	
	σ_a^2	n_a	σ_A^2	Term	n_a	σ_A^2	Term
r	1.134	6	0.189	Δ		1.134	τ
d	0.195	41	0.005	Δ	41	0.005	Δ
$r \cdot d$	0.743	246	0.003	Δ	41	0.018	Δ, δ
$r \cdot t$	0.638	18	0.035	Δ	3	0.213	τ
$d \cdot t$	0.161	123	0.001	Δ	123	0.001	Δ, δ
$r \cdot d \cdot t$	0.553	738	0.001	Δ	123	0.004	Δ, δ
$e:[d \cdot t]$	1.409		1.409	τ	7,872	0.000	Δ
$r \cdot e:[d \cdot t]$	5.925	6	0.988	Δ, δ	7,872	0.001	Δ, δ

(B)					
Object	σ_τ^2	σ_δ^2	σ_Δ^2	ρ^2	φ^2
Epoch	1.409	0.988	1.222	0.588	0.536
Reviewer	1.347	0.024	0.030	0.983	0.978

(A) Overview of variance components. Each row corresponds to the effect or interaction listed in the first column. The sets of columns outline the variance components (leftmost column set), as well as the normalized variance components where the object of measurement is set to Epoch (second column set) or Reviewer (third column set). For each object of measurement, the effective sample sizes (n_a), normalized variance components (σ_A^2), and term(s) corresponding to the variance components are listed. **(B)** The object variances (σ_τ^2), relative residual variances (σ_δ^2), absolute residual variances (σ_Δ^2), relative generalizability coefficients (ρ^2), and absolute dependability coefficients (φ^2), for each object of measurement. The relative generalizability coefficients are indicated in bold.



10 min of continuous EEG for spikes and ripples in a single channel derivation (27). This 2.5 h for 5 min of continuous data is somewhat less than the worst-case 2.87 h for epoched

review noted above. However, the epoched review would provide markings for every channel, while the continuous review completed in that time constitutes markings for only a single channel.

Epoch Generalizability—Interrater Reliability

The D-study projections for Epochs predicted how many Reviewers would be required to achieve a desired Epoch generalizability. Each Epoch is selected from exactly one Dataset, and is entirely independent of other Datasets; therefore, D-study projections are only affected by the number of Reviewers. Only a moderate Epoch generalizability (0.588) was achieved with the six Reviewers used in this study. Notably, 17 Reviewers would be required to achieve a threshold of 0.8. This is a prohibitive amount for any individual epilepsy center, and would therefore not be feasible for the routine evaluation of any individual patient.

One solution to the requirement of such a large number of Reviewers would be to initiate a multi-center study, including at least 17 Reviewers, all evaluating the same Epochs. This would result in highly generalizable evaluations of individual Epochs, featuring both positive and negative HFOs of varying characteristics, with and without artifacts, constituting a benchmark set of universal HFO evaluations. This benchmark could be used as a reliable standard for testing any number of HFO detection or classification algorithms, or for the testing or training of Reviewers. The development of such a benchmark dataset is currently underway, featuring epileptologists from multiple centers across western Canada.

The Epoch generalizability for two Reviewers was estimated to be 0.322. Notably, this is in agreement with our previous study, which found poor interrater reliability when using a smaller dataset (44). This supports the use of the Epoch generalizability coefficient as a measure for interrater reliability, and provides further evidence against the use of agreement between two visual reviewers as a “gold standard” for HFO markings.

Reviewer Generalizability—Internal Reviewer Threshold

The D-study projections for Reviewer predicted how many Epochs and Datasets would be required to achieve a desired Reviewer generalizability. For example, to achieve a Reviewer generalizability of 0.8, each Reviewer could evaluate 3 Epochs per EventType in 5 Datasets ($3 \times 3 \times 5 = 45$ evaluations), or 1 Epochs per EventType in 10 Datasets ($1 \times 3 \times 10 = 30$ evaluations). A study interested only in establishing a ranking of Reviewers by achieving a Reviewer generalizability of 0.8 could therefore be constructed with as few as 30 evaluations per Reviewer. Such a study could be completed for any given set of Reviewers, and would require an average time commitment of just under 4 min, assuming the trimmed mean observed Epoch evaluation time of 7.43 s.

Given this high Reviewer generalizability, there need not be 17 Reviewers to produce highly generalizable Epoch evaluations. Rather, the Epoch generalizability could be increased by

increasing training with respect to the HFO evaluations. In the present study, training was limited to the presentation of an instructional video and companion document, which defined HFOs and provided clear examples. This training was designed to avoid influencing the internal threshold of the Reviewers. Conversely, training Reviewers to adjust their internal thresholds would decrease the variability between Reviewers (65–67), increasing the generalizability of their Epoch evaluations. Our group is currently conducting one such study using the principles of the Delphi method (68), whereby Reviewers first conduct the study with no *a priori* information, and then again using feedback information regarding the stringency of their evaluations relative to that of their peers. An increase in the Epoch generalizability could be expected with the Delphi method, which would decrease the number of Reviewers needed to obtain universal HFO evaluations for any given Epoch.

Such training procedures might facilitate the feasibility of on-site evaluation of patient data for HFOs as an informative component of the epilepsy pre-surgical workups. The proposed threshold training could also reduce the number of Reviewers needed to construct a benchmark dataset, or to further increase its generalizability. These possibilities should be thoroughly evaluated following the implementation of this Delphi study.

HFO Evaluation Framework

Overall, this study has established a novel framework for efficient and controlled evaluation of HFOs. The Epoch design allows for the pre-selection of events based on any number of algorithmic criteria, and can be tailored to the particular application. It also ensures that all Reviewers can evaluate the exact same events of interest in the exact same order, in a manner that requires only an average of 10 s per evaluation. The evaluation software program provides Reviewers with an intuitive and versatile interface to complete any number of evaluations. It enables Reviewers to not only mark the presence of HFOs, but also the absence of HFOs, and the self-reported certainty in these determinations.

The present study has established that evaluations conducted using this supervised method are not sufficient to be used immediately in clinical practice. Rather, the framework itself may be extended to any number of automated detectors, to any number of reviewers, or to any form of standardized training, producing a comprehensive set of evaluations accompanied by descriptive information regarding the generalizability of the findings.

External Applicability

Notably, detection of electrographic signals such as HFOs are often performed using a visual review of a continuous record, or using an automated detector alone, rather than a supervised framework. However, the calculated generalizability coefficients and the predicted number of reviewers apply only to studies undertaken in this framework. It is entirely possible that HFO markings made on a continuous record would require fewer reviewers to achieve strongly generalizable findings, due to influences such as the additional context provided by the continuity of the record. Likewise, it is intuitive that the context

provided by a continuous record may bias the markings—such as cases where clear or frequent HFOs in one channel preclude more ambiguous or occasional HFOs in another channel from being noticed. Furthermore, the sheer volume of continuous data may result in different reviewers focusing their search of HFOs on different channels or time windows, increasing the number of reviewers required to achieve strongly generalizable findings. Regardless of such points of speculation, it should be unambiguously emphasized that one cannot assume that HFO markings obtained through continuous review are generalizable—rather, it must be assumed that they are not generalizable, until it can be proven that they are.

The findings of this study are inherently applicable to the oscillations detected using two thresholds of an established detection algorithm, along with baseline EEG segments. The generalizability may differ as the threshold or method used to detect candidate HFOs varies, as might the clinical meaning of such oscillations. Further studies conducted within this framework may indeed be used to compare the meaning and generalizability of HFOs identified at various thresholds, in other frequency bands, or using different detection algorithms.

Alternatively, this framework could allow for the evaluation of HFO detection algorithms in several capacities, including comparing the performance of countless HFO detection algorithms against generalizable markings obtained by a large set of visual reviewers. A detection algorithm that is shown to be consistent with strongly generalizable visual markings could then be used as a standard, either alone or in conjunction with one or two reviewers, replacing the ongoing need for such a large number of visual reviewers.

Furthermore, while the waveforms of interest in this study were HFOs identified on intracranial EEG, studies have shown that other electrographic markers are subject to varying degrees of interrater reliability. It would be interesting to extend the methodology presented herein to other EEG markers independently of HFOs, such as epileptiform discharges, or even other electrographic markers, such as U waves on ECG, to determine the generalizability of those visual markings. It would also be interesting to apply this methodology to simultaneous EEG and magnetoencephalography, which has recent shown promising results for the source localization of HFOs and IEDs (69). The present study focused on the generalizability of HFO detection. As a next step, we are currently assessing how the algorithmically detected HFOs correlate with location of the seizure onset zone and post-surgical outcome.

Limitations and Future Direction

The findings of this study establish a framework for further studies, which can efficiently reduce or eliminate many factors that currently limit conclusions that may be drawn in HFO studies.

One notable consideration is the degree of *a priori* training of the various Reviewers. All of the Reviewers in the present study were familiar with HFOs, and all were given both an instructional video and a short written document to help them become acquainted with the software used and the criteria used to evaluate HFOs in this study. However, the training was

not exhaustive and was not designed to influence the internal threshold of the Reviewers with respect to HFO evaluations—only the clearest HFOs were presented as examples. It is reasonable to assume that two reviewers trained at the same center would be more similar than reviewers with different training backgrounds, and that therefore the markings made by two such reviewers would be more generalizable. However, our previous work found the pairwise interrater reliability between two similarly and extensively trained reviewers to be as poor as the overall interrater reliability (44). Furthermore, it is likely that across the global population of HFO reviewers, the individual training received would also vary greatly. Nonetheless, standardized interventional training such as the simple Delphi study described above may be effective in using feedback between evaluation sessions to encourage Reviewers to align their internal threshold to a desired target, thus increasing the generalizability of evaluations, whatever the background of any given Reviewer.

There are other confounding factors that may affect how representative our sample Epochs, Datasets, or Reviewers were, which could influence the generalizability. Additionally, Reviewers may be affected not only by *a priori* training, but also by fatigue; fortunately, the framework presented herein is designed to preclude fatigue, by reducing the time commitment of HFO evaluations, and future studies aimed at optimizing interrater reliability could further reduce the workload. Also, the Datasets or Epochs may be affected by time of day, admission date, and evaluation order: steps were made to control for time of day and admission date, by making the data selection across patients consistent, and the evaluation order was standardized for all Reviewers. Nonetheless the impact of any of these effects is unknown, but the proposed framework is well suited to study the effect sizes.

Another effect that may be addressed in future studies is that of reproducibility, which examines the variability in HFO evaluations of the same Epochs by the same Reviewers on multiple occasions. Such internal consistency is currently being assessed by incorporating multiple occasions into the aforementioned Delphi study of a large number of Reviewers.

CONCLUSION

We have reaffirmed that the current practice in visual HFO identification, which uses the agreement between only two visual reviewers to identify HFOs, is unreliable due to significant variability between reviewers. We have also projected that the large number of visual reviewers required to produce reliable HFO markings would be an impractical barrier to undertaking visual HFO review in a clinical setting without first reducing the variability amongst reviewers.

We have also outlined a set of tools, including an evaluation program and a statistical framework, which may be used for training visual reviewers and tracking changes to interrater

reliability, all while reducing the time burden of HFO analysis. Ongoing studies at our research center are using established training protocols within the framework presented herein to standardize the internal HFO threshold of visual reviewers, further reducing the time burden. It may alternatively serve as the basis of a multicenter study to construct a benchmark set of epochs that can serve as a new, highly generalizable gold standard, against which any number of HFO detection algorithms may be compared.

Ultimately, this study represents an important step toward the reconciliation of important but discordant findings from HFO studies undertaken with different sets of HFOs, and ultimately toward transitioning HFO analysis into a feasible and meaningful part of the epilepsy pre-surgical workup.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the University of Calgary Conjoint Research Ethics Board with written informed consent from all subjects. All subjects were given written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Calgary Conjoint Research Ethics Board.

AUTHOR CONTRIBUTIONS

AS, DP, and PF contributed conception and design of the study. AS and PF performed the initial data collection. AS wrote the morphology detector. DP created the evaluation program. YA, JJ, NP, LB, CJ, and PF performed the primary data analysis. AS performed the statistical analyses. AS, DP, and PF interpreted the data. AS wrote the first draft of the manuscript. All authors contributed to manuscript review, and read and approved the submitted version.

FUNDING

This work was supported by the Canadian Institutes of Health Research (MOP-230809). AS was supported by Alberta Innovates Health Solutions and the Canadian Institutes of Health Research.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Houman Khosravani and the staff of the Seizure Monitoring Unit (Foothills Medical Centre) for their invaluable help acquiring the EEG data used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2018.00510/full#supplementary-material>

REFERENCES

- Jacobs J, Zijlmans M, Zelmann R, Chatillon CE, Hall J, Olivier A, et al. High-frequency electroencephalographic oscillations correlate with outcome of epilepsy surgery. *Ann Neurol.* (2010) 67:209–20. doi: 10.1002/ana.21847
- Cho JR, Koo DL, Joo EY, Seo DW, Hong SC, Jiruska P, et al. Resection of individually identified high-rate high-frequency oscillations region is associated with favorable outcome in neocortical epilepsy. *Epilepsia* (2014) 55:1872–83. doi: 10.1111/epi.12808
- Jacobs J, LeVan P, Chander R, Hall J, Dubeau F, 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* (2008) 49:1893–907. doi: 10.1111/j.1528-1167.2008.01656.x
- Wang S, Wang I, Bulacio J, Mosher JC, Gonzalez-Martinez J, Alexopoulos AV, et al. Ripple classification helps to localize the seizure-onset zone in neocortical epilepsy. *Epilepsia* (2013) 54:370–6. doi: 10.1111/j.1528-1167.2012.03721.x
- Cimbalnik J, Kuciewicz MT, Worrell GA. Interictal high-frequency oscillations in focal human epilepsy. *Curr Opin Neurol.* (2016) 29:175–81. doi: 10.1097/WCO.0000000000000302
- Höller Y, Kutil R, Klaffenböck L, Thomschewski A, Höller PM, Bathke AC, et al. High-frequency oscillations in epilepsy and surgical outcome: a meta-analysis. *Front Hum Neurosci.* (2015) 9:574. doi: 10.3389/fnhum.2015.00574
- Akiyama T, McCoy B, Go CY, Ochi A, Elliot IM, Akiyama M, et al. Focal resection of fast ripples on extraoperative intracranial EEG improves seizure outcome in pediatric epilepsy. *Epilepsia* (2011) 52:1802–11. doi: 10.1111/j.1528-1167.2011.03199.x
- Okanishi T, Akiyama T, Tanaka S, Mayo E, Mitsutake A, Boelman C, et al. Interictal high frequency oscillations correlating with seizure outcome in patients with widespread epileptic networks in tuberous sclerosis complex. *Epilepsia* (2014) 55:1602–10. doi: 10.1111/epi.12761
- van 't Klooster MA, Leijten FS, Huiskamp G, Ronner HE, Baayen JC, van Rijen PC, et al. study group H. High frequency oscillations in the intra-operative ECoG to guide epilepsy surgery (“The HFO Trial”): study protocol for a randomized controlled trial. *Trials* (2015) 16:422. doi: 10.1186/s13063-015-0932-6
- van Klink NE, Van't Klooster MA, Zelmann R, Leijten FS, Ferrier CH, Braun KP, et al. High frequency oscillations in intra-operative electrocorticography before and after epilepsy surgery. *Clin Neurophysiol.* (2014) 125:2212–19. doi: 10.1016/j.clinph.2014.03.004
- Fedele T, van 't Klooster MA, Burnos S, Zweiphenning WJ, van Klink N, Leijten F, et al. Automatic detection of high frequency oscillations during epilepsy surgery predicts seizure outcome. *Clin Neurophysiol.* (2016) 127:3066–74. doi: 10.1016/j.clinph.2016.06.009
- Guragain H, Cimbalnik J, Stead M, Groppe DM, Berry BM, Kremen V, et al. Spatial variation in high-frequency oscillation rates and amplitudes in intracranial EEG. *Neurology* (2018) 90:e639–46. doi: 10.1212/WNL.0000000000004998
- Wang S, So NK, Jin B, Wang IZ, Bulacio JC, Enatsu R, et al. Interictal ripples nested in epileptiform discharge help to identify the epileptogenic zone in neocortical epilepsy. *Clin Neurophysiol.* (2017) 128:945–51. doi: 10.1016/j.clinph.2017.03.033
- Burnos S, Frauscher B, Zelmann R, Haegelen C, Sarnthein J, Gotman J. The morphology of high frequency oscillations (HFO) does not improve delineating the epileptogenic zone. *Clin Neurophysiol.* (2016) 127:2140–8. doi: 10.1016/j.clinph.2016.01.002
- Amiri M, Lina J, Pizzo F, Gotman J. High frequency oscillations and spikes: separating real HFOs from false oscillations. *Clin Neurophysiol.* (2016) 127:187–96. doi: 10.1016/j.clinph.2015.04.290
- Waldman ZJ, Shimamoto S, Song I, Orosz I, Bragin A, Fried I, et al. A method for the topographical identification and quantification of high frequency oscillations in intracranial electroencephalography recordings. *Clin Neurophysiol.* (2018) 129:308–18. doi: 10.1016/j.clinph.2017.10.004
- Fedele T, Burnos S, Boran E, Krayenbühl N, Hilfiker P, Grunwald T, et al. Resection of high frequency oscillations predicts seizure outcome in the individual patient. *Sci Rep.* (2017) 7:13836. doi: 10.1038/s41598-017-13064-1
- Hussain SA, Mathern GW, Sankar R, Wu JY. Prospective and “live” fast ripple detection and localization in the operating room: Impact on epilepsy surgery outcomes in children. *Epilepsy Res.* (2016) 127:344–51. doi: 10.1016/j.eplesyres.2016.09.017
- Gloss D, Nevitt SJ, Staba R. The role of high-frequency oscillations in epilepsy surgery planning. *Cochrane Database Syst Rev.* (2017) 10:CD010235. doi: 10.1002/14651858.CD010235.pub3
- Urrestarazu E, Chander R, Dubeau F, Gotman J. Interictal high-frequency oscillations (100–500 Hz) in the intracerebral EEG of epileptic patients. *Brain* (2007) 130:2354–66. doi: 10.1093/brain/awm149
- Bagshaw A, Jacobs J, LeVan P, Dubeau F, Gotman J. Effect of sleep stage on interictal highfrequency oscillations recorded from depth macroelectrodes in patients with focal epilepsy. *Epilepsia* (2009) 50:617–28. doi: 10.1111/j.1528-1167.2008.01784.x
- Zijlmans M, Jacobs J, Zelmann R, Dubeau F, Gotman J. High-frequency oscillations mirror disease activity in patients with epilepsy. *Neurology* (2009) 72:979–86. doi: 10.1212/01.wnl.0000344402.20334.81
- Ellenrieder N, Andrade-Valença L, Dubeau F, Gotman J. Automatic detection of fast oscillations (40–200Hz) in scalp EEG recordings. *Clin Neurophysiol.* (2012) 123:670–80. doi: 10.1016/j.clinph.2011.07.050
- Châtillon CE, Zelmann R, Hall JA, Olivier A, Dubeau F, Gotman J. Influence of contact size on the detection of HFOs in human intracerebral EEG recordings. *Clin Neurophysiol.* (2013) 124:1541–6. doi: 10.1016/j.clinph.2013.02.113
- Haegelen C, Perucca P, Châtillon CE, Andrade-Valença L, Zelmann R, Jacobs J, et al. High-frequency oscillations, extent of surgical resection, and surgical outcome in drug-resistant focal epilepsy. *Epilepsia* (2013) 54:848–57. doi: 10.1111/epi.12075
- Chaitanya G, Sinha S, Narayanan M, Satishchandra P. Scalp high frequency oscillations (HFOs) in absence epilepsy: An independent component analysis (ICA) based approach. *Epilepsy Res.* (2015) 115:133–40. doi: 10.1016/j.eplesyres.2015.06.008
- Chu CJ, Chan A, Song D, Staley KJ, Stufflebeam SM, Kramer MA. A semi-automated method for rapid detection of ripple events on interictal voltage discharges in the scalp electroencephalogram. *J Neurosci Methods* (2017) 277:46–55. doi: 10.1016/j.jneumeth.2016.12.009
- Jacobs J, Zelmann R, Jirsch J, Chander R, Châtillon CE, Dubeau F, et al. High frequency oscillations (80–500 Hz) in the preictal period in patients with focal seizures. *Epilepsia* (2009) 50:1780–92. doi: 10.1111/j.1528-1167.2009.02067.x
- Zelmann R, Zijlmans M, Jacobs J, Châtillon CE, Gotman J. Improving the identification of high frequency oscillations. *Clin Neurophysiol.* (2009) 120:1457–64. doi: 10.1016/j.clinph.2009.05.029
- Zijlmans M, Jacobs J, Zelmann R, Dubeau F, Gotman J. High frequency oscillations and seizure frequency in patients with focal epilepsy. *Epilepsy Res.* (2009) 85:287–92. doi: 10.1016/j.eplesyres.2009.03.026
- Jacobs J, Zijlmans M, Zelmann R, Olivier A, Hall J, Gotman J, et al. Value of electrical stimulation and high frequency oscillations (80–500 Hz) in identifying epileptogenic areas during intracranial EEG recordings. *Epilepsia* (2010) 51:573–82. doi: 10.1111/j.1528-1167.2009.02389.x
- Zelmann R, Mari F, Zijlmans M, Chander R, Gotman J. Automatic detector of high frequency oscillations for human recordings with macroelectrodes. *Conf Proc IEEE Eng Med Biol Soc.* (2010) 2010:2329–33. doi: 10.1109/IEMBS.2010.5627464
- Dümpelmann M, Jacobs J, Kerber K, Schulze-Bonhage A. Automatic 80–250 Hz “ripple” high frequency oscillation detection in invasive subdural grid and strip recordings in epilepsy by a radial basis function neural network. *Clin Neurophysiol.* (2012) 123:1721–31. doi: 10.1016/j.clinph.2012.02.072
- Zelmann R, Mari F, Jacobs J, Dubeau F, Gotman J. A comparison between detectors of high frequency oscillations. *Clin Neurophysiol.* (2012) 123:106–16. doi: 10.1016/j.clinph.2011.06.006
- van Diessen E, Hanemaaijer J, Otte W, Zelmann R, Jacobs J, Jansen FE, et al. Are high frequency oscillations associated with altered network topology in partial epilepsy? *Neuroimage* (2013) 82:564–73. doi: 10.1016/j.neuroimage.2013.06.031
- Pail M, Haláček J, Daniel P, Kuba R, Tyrliková I, Chrastina J, et al. Intracerebrally recorded high frequency oscillations: simple visual assessment versus automated detection. *Clin Neurophysiol.* (2013) 124:1935–42. doi: 10.1016/j.clinph.2013.03.032
- Kerber K, Dümpelmann M, Schelter B, Le Van P, Korinthenberg R, Schulze-Bonhage A, et al. Differentiation of specific ripple patterns helps to

- identify epileptogenic areas for surgical procedures. *Clin Neurophysiol.* (2014) 125:1339–45. doi: 10.1016/j.clinph.2013.11.030
38. Ferrari-Marinho T, Perucca P, Mok K, Olivier A, Hall J, Dubeau F, et al. Pathologic substrates of focal epilepsy influence the generation of high-frequency oscillations. *Epilepsia* (2015) 56:592–8. doi: 10.1111/epi.12940
 39. Frauscher B, von Ellenrieder N, Ferrari-Marinho T, Avoli M, Dubeau F, Gotman J. Facilitation of epileptic activity during sleep is mediated by high amplitude slow waves. *Brain* (2015) 138:1629–41. doi: 10.1093/brain/awv073
 40. Jacobs J, Vogt C, LeVan P, Zelmann R, Gotman J, Kobayashi K. The identification of distinct high-frequency oscillations during spikes delineates the seizure onset zone better than high-frequency spectral power changes. *Clin Neurophysiol.* (2016) 127:129–42. doi: 10.1016/j.clinph.2015.04.053
 41. Jrad N, Kachenoura A, Merlet I, Bartolomei F, Nica A, Biraben A, et al. Automatic detection and classification of high-frequency oscillations in depth-EEG signals. *IEEE Trans Biomed Eng.* (2017) 64: 2230–40. doi: 10.1109/TBME.2016.2633391
 42. Salami P, Lévesque M, Gotman J, Avoli M. A comparison between automated detection methods of high-frequency oscillations (80–500Hz) during seizures. *J Neurosci Met.* (2012) 211:265–71. doi: 10.1016/j.jneumeth.2012.09.003
 43. Gliske SV, Irwin ZT, Davis KA, Sahaya K, Chestek C, Stacey WC. Universal automated high frequency oscillation detector for real-time, long term EEG. *Clin Neurophysiol.* (2016) 127:1057–66. doi: 10.1016/j.clinph.2015.07.016
 44. Spring AM, Pittman DJ, Aghakhani Y, Jirsch J, Pillay N, Bello-Espinosa LE, et al. Interrater reliability and variability of visually evaluated high frequency oscillations. *Clin Neurophysiol.* (2017) 128:433–41. doi: 10.1016/j.clinph.2016.12.017
 45. Gardner A, Worrell G, March E, Dlugos D, Litt B. Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. *Clin Neurophysiol.* (2007) 118:1134–43. doi: 10.1016/j.clinph.2006.12.019
 46. Staba RJ, Wilson CL, Bragin A, Fried I, Engel Jr. J. Quantitative analysis of high-frequency oscillations (80–500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. *J Neurophysiol.* (2002) 88:1743–52. doi: 10.1152/jn.2002.88.4.1743
 47. Crépon B, Navarro V, Hasboun D, Clemenceau S, Martinerie J, Baulac M, et al. Mapping interictal oscillations greater than 200 Hz recorded with intracranial macroelectrodes in human epilepsy. *Brain* (2010) 133:33–45. doi: 10.1093/brain/awp277
 48. Nagasawa T, Juhász C, Rothermel R, Hoechstetter K, Sood S, Asano E. Spontaneous and visually driven high-frequency oscillations in the occipital cortex: Intracranial recording in epileptic patients. *Hum Brain Mapp.* (2012) 33:569–83. doi: 10.1002/hbm.21233
 49. Sakuraba R, Iwasaki M, Okumura E, Jin K, Kakisaka Y, Kato K, et al. High frequency oscillations are less frequent but more specific to epileptogenicity during rapid eye movement sleep. *Clin Neurophysiol.* (2016) 127:179–86. doi: 10.1016/j.clinph.2015.05.019
 50. Geertsema EE, van 't Klooster MA, van Klink NE, Leijten FS, van Rijen PC, Visser GH, et al. Non-harmonicity in high-frequency components of the intra-operative corticogram to delineate epileptogenic tissue during surgery. *Clin Neurophysiol.* (2017) 128:153–64. doi: 10.1016/j.clinph.2016.11.007
 51. Shimamoto S, Waldman ZJ, Orosz I, Song I, Bragin A, Fried I, et al. Utilization of independent component analysis for accurate pathological ripple detection in intracranial EEG recordings recorded extra- and intra-operatively. *Clin Neurophysiol.* (2018) 129:296–307. doi: 10.1016/j.clinph.2017.08.036
 52. Charupanit K, Lopour BA. A simple statistical method for the automatic detection of ripples in human intracranial EEG. *Brain Topogr.* (2017) 30:724–38. doi: 10.1007/s10548-017-0579-6
 53. Chen-Wei C, Chien C, Shang-Yeong K, Shun-Chi W. Multi-channel algorithms for epileptic high-frequency oscillation detection. *Conf Proc IEEE Eng Med Biol Soc.* (2016) 2016:948–51. doi: 10.1109/EMBC.2016.7590858
 54. Worrell G, Jerbi K, Kobayashi K, Lina J, Zelmann R, van Quyen M. Recording and analysis techniques for high-frequency oscillations. *Prog Neurobiol.* (2012) 98:265–78. doi: 10.1016/j.pneurobio.2012.02.006
 55. Nelson R, Myers SM, Sinmonotto JD, Furman MD, Spano M, Norman WM, et al. Detection of high frequency oscillations with Teager energy in an animal model of limbic epilepsy. *Conf Proc IEEE Eng Med Biol Soc.* (2006) 1:2578–80. doi: 10.1109/IEMBS.2006.259694
 56. Huang L, Ni X, Ditto WL, Spano M, Carney PR, Lai YC. Detecting and characterizing high-frequency oscillations in epilepsy: a case study of big data analysis. *R Soc Open Sci.* (2017) 4:160741. doi: 10.1098/rsos.160741
 57. Liu S, Sha Z, Sencer A, Aydoseli A, Bebek N, Abosch A, et al. Exploring the time-frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy. *J Neural Eng.* (2016) 13:026026. doi: 10.1088/1741-2560/13/2/026026
 58. Blood A, Park Y, Lukas R, Brorson J. Neurology objective structured clinical examination reliability using generalizability theory. *Neurology* (2015) 85:1623–9. doi: 10.1212/WNL.0000000000002053
 59. Kline TJ. *Reliability of Raters. Psychological Testing: a Practical Approach to Design and Evaluation.* Thousand Oaks, CA: Sage (2005). p. 185–200.
 60. Brennan RL. *Generalizability Theory.* New York, NY: Springer-Verlag (2001).
 61. Rao CR. Estimation of variance and covariance components MINQUE theory. *J Multivar Anal.* (1971) 1:257–75. doi: 10.1016/0047-259X(71)90001-7
 62. Rao CR. Minimum variance quadratic unbiased estimation of variance components. *J Multivar Anal.* (1971) 1:445–56. doi: 10.1016/0047-259X(71)90019-4
 63. Roos JE, Chilla B, Zanetti M, Schmid M, Koch P, Pfirrmann CW, et al. MRI of meniscal lesions: soft-copy (PACS) and hard-copy evaluation versus reviewer experience. *AJR Am J Roentgenol.* (2006) 186:786–90. doi: 10.2214/AJR.04.1853
 64. Saur SC, Alkadhhi H, Stolzmann P, Baumüller S, Leschka S, Scheffel H, et al. Effect of reader experience on variability, evaluation time and accuracy of coronary plaque detection with computed tomography coronary angiography. *Eur Radiol.* (2010) 20:1599–606. doi: 10.1007/s00330-009-1709-7
 65. Kondo Y. Examination of rater training effect and rater eligibility in L2 performance assessment. *J Pan Pacific Assoc Appl Linguist.* (2010) 14:1–23.
 66. Fahim M, Bijani H. The effects of rater training on raters' severity and bias in second language writing assessment. *Ir J Lang Test.* (2011) 1:1–16.
 67. Davis L. The influence of training and experience on rater performance in scoring spoken language. *Lang Test.* (2016) 33:117–35. doi: 10.1177/0265532215582282
 68. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. *Manage. Sci.* (1963) 9:458–67. doi: 10.1287/mnsc.9.3.458
 69. Papadelis C, Tamilia E, Stufflebeam S, Grant PE, Madsen JR, Pearl PL, et al. Interictal high frequency oscillations detected with simultaneous magnetoencephalography and electroencephalography as biomarker of pediatric epilepsy. *J Vis Exp.* (2016) 118:e54883. doi: 10.3791/54883
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Spring, Pittman, Aghakhani, Jirsch, Pillay, Bello-Espinosa, Josephson and Federico. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.