

# Toward Online Data Reduction for Portable Electroencephalography Systems in Epilepsy

Alexander J. Casson\*, *Student Member, IEEE*, and Esther Rodriguez-Villegas, *Senior Member, IEEE*

**Abstract**—Portable EEG units are key tools in epilepsy diagnosis. Current systems could be made physically smaller and longer lasting by the inclusion of online data reduction methods to reduce the power required for storage or transmission of the EEG data. This paper presents a real-time data reduction algorithm based upon the discontinuous recording of the EEG: noninteresting background sections of EEG are discarded online, with only potentially diagnostically interesting sections being saved. MATLAB simulations of the algorithm on an EEG dataset containing 982 expert marked events in 4 days of data show that 90% of events can be correctly recorded while achieving a 50% data reduction. The described algorithm is formulated to have a direct, low power, hardware implementation and similar data reduction strategies could be employed in a range of body-area-network-type applications.

**Index Terms**—EEG, epileptic spike detection, local signal processing, online data reduction, wavelet analysis.

## I. INTRODUCTION

THE electroencephalogram (EEG), which records the microvolt-sized signals that result on the scalp from brain activity, has long been a key tool in epilepsy diagnosis [1]. It is used in determining whether the disorder is present, the disorder type if it is, and the locus within the brain. Traditional diagnostic EEG monitoring is performed in the inpatient setting via a 20- to 30-min EEG test [2] with indicative activity found in approximately 50% of patients [3]. To increase diagnostic yield, longer duration tests are frequently used, and it is not uncommon for patients who are candidates for epilepsy surgery to undergo many days of continuous monitoring. However, long-term testing, or multiple short tests, are expensive and are not universally available [4].

As a result, portable outpatient-based monitoring known as ambulatory EEG (AEEG) is a popular alternative. AEEG allows monitoring in the patient's home environment and is estimated to be clinically useful in 75% of patients and to be 50% cheaper than comparable inpatient monitoring [4]. As such, increased use of AEEG for long-term outpatient monitoring could be of significant use in reducing the epilepsy misdiagnosis rate, currently estimated at 25% [5]. Recent questionnaire results have also indicated that there is significant desire from the medical

community to have miniaturized, long-lasting, wearable AEEG systems in place [6].

For such long-term outpatient AEEG monitoring to be successful the AEEG unit must be physically small, discrete, comfortable and socially acceptable. Assuming that practical and acceptable long-lasting electrode technologies are developed (see [6]) still required is device miniaturization that will place stringent requirements on the available battery size, and so on the power consumption. It is estimated that the total AEEG system power consumption must be less than 140  $\mu$ W.

Many biomedical sensors suffer from similar long-term power issues, and it is agreed that the only way of reducing the device power consumption is the introduction of local signal processing, or *intelligence*, that runs on the portable device itself [7], [8]. The idea is that in any system setup, some automated processing will be carried out on the collected physiological signal, and that storage or transmission of the signal is power intensive.<sup>1</sup> Thus, if the automated analysis is carried out on the portable device, only the results of this analysis need to be transmitted, not the entire collected signal. This can significantly reduce the amount of data to be transmitted, and hence, the required power consumption. This is provided, of course, that the local signal processing has lower power consumption than simply transmitting all of the data. A full-power tradeoff of this idea has been detailed elsewhere [9].

For AEEG systems for use in epilepsy diagnosis, analysis of the EEG produced is still done by human interpreters, although software tools (feature detection algorithms) are available to help identify areas of potential interest. The key local signal processing required is thus simply data reduction, rather than data interpretation, and this must operate at very low power levels and in real time for use in long-term systems.

This paper presents an algorithm designed to carry out this data reduction. The algorithm is described in Section II, and is based around a real-time feature detection algorithm to simultaneously reduce the amount of data that are presented for human analysis. The data reduction performance of the algorithm is presented in Section III and Section IV via MATLAB simulations on real EEG data. The low-power implementation of the algorithm lies beyond the scope of this paper. Finally, the results are discussed and analyzed in Section V.

Preliminary versions of the algorithm reported here have been presented in [10] and [11]. These analyses have not used the full algorithm considered here, or given full algorithm details,

<sup>1</sup>For simplicity, all arguments in this paper are formulated in terms of transmitting data off the portable device wirelessly. Local flash memory power consumption is also power intensive and arguments could be reformulated in terms of this, if desired.

Manuscript received March 11, 2009; revised June 3, 2009. First published July 28, 2009; current version published November 20, 2009. Asterisk indicates corresponding author.

\*A. J. Casson is with the Department of Electrical and Electronic Engineering, Circuits and Systems Research Group, Imperial College, London SW7 2AZ, U.K. (e-mail: acasson@imperial.ac.uk).

E. Rodriguez-Villegas is with the Department of Electrical and Electronic Engineering, Circuits and Systems Research Group, Imperial College, London SW7 2AZ, UK (e-mail: e.rodriguez@imperial.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBME.2009.2027607

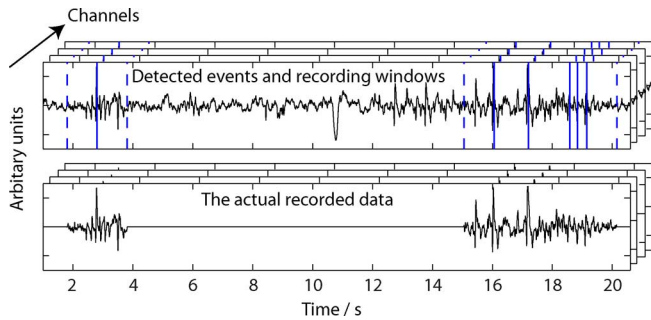


Fig. 1. Discontinuous recording procedure only saves data for a brief period (dashed vertical lines) on either side of an automated detection (solid vertical lines). This is seen to provide significant data reduction.

and have used a smaller test database, without performing any statistical testing of the results.

## II. DATA REDUCTION ALGORITHM

### A. Methodology

A standard AEEG unit may record up to 32 channels, with a sampling rate of 200 Hz and 16-bit resolution. This gives a data rate of 12.5 KBytes/s to be transmitted off the portable device. The aim of online data compression is to reduce this average data rate while consuming as little power as possible. There are several ways of providing such data compression, and a comparison of different methods with their advantages and disadvantages is given in [12].

The compression method investigated here is based around discontinuous monitoring of the AEEG [12]. Epileptic EEG traces can be broken down into two phases: ictal (seizure activity) and interictal (spikes and spike-and-waves) [2]. Interictal activity usually contains isolated events along with normal background signals. By recording only the ictal and interesting interictal activity, significant data reduction can be achieved [13].

This procedure is illustrated in Fig. 1, where three interictal spikes, along with a window of data around them, are selected for transmission. Even with some false detections present, a significant reduction in the amount of data to transmit is seen.

Discontinuous monitoring does introduce some data interpretations issues as the full EEG record is no longer available, and these are discussed in [12]. Nevertheless, discontinuous recording methods have long been implemented in offline systems for epilepsy diagnosis [14], and are available through devices, such as the Reveal Heart Monitor [15], for ECG investigation.

It thus seems natural to extend such systems into the AEEG realm, and indeed, if long-term recordings—of the scale of months—are ever to be performed to capture rare epileptic events, it is unlikely that a human interpreter could score all of the data. Some form of feature detection algorithm is required, and implementing this on the AEEG unit itself can simultaneously improve battery life and device size.

For the initial algorithm development investigated here, only the recording of interictal spikes is considered.<sup>2</sup> The extension of the algorithm and its evaluation on recording seizure epochs is left to future work.

<sup>2</sup>This paper treats all interictal events, such as spikes, sharp waves, and spike-and-waves, under the umbrella term spikes.

### B. Topology Motivation

The obvious method to drive a discontinuous data reduction mechanism for recording interictal spikes is to use an interictal spike detector. An attempt is made to detect when the features of interest occur, and then, a window on either side of these is recorded.

Over the past 40 years, a huge number of interictal spike detection algorithms have been published in the academic literature and made available as software tools. A complete review of these lies beyond the scope of this paper, but a list of 70 such papers is given in [16], and [17] provides an excellent review from 2002.

The aim here is not to generate a new algorithm with superior performance to all of these algorithms. Instead it is to develop, and report the performance of an algorithm that meets some very specific needs.

- 1) It must detect events with medically acceptable accuracy.
- 2) It must operate in real time.
- 3) It must be of low computational complexity (so that the power required for the signal processing does not exceed the transmission power, and leads to an overall reduction in the system power consumption).
- 4) It should map to a hardware implementation.

This last point is important for minimizing the power consumption required to implement the algorithm. While a micro-processor running a software algorithm generally offers a more flexible solution, it cannot approach the performance of a full hardware implementation.

For optimal results, the need for this hardware implementation must drive the design of the entire algorithm. For example, only algorithm blocks that correspond directly to circuit elements, such as filters, rectifiers, and adders, should be considered for use. This does, however, hugely reduce the range of signal processing choices available.

The algorithm proposed is based upon the continuous wavelet transform (CWT), which many studies have shown to be of significant use for analyzing and detecting epileptic EEG transients [18]–[20]. In addition, hardware CWT implementations have been reported with quoted power consumptions ranging down to 100 nW [21], [22], readily suitable for the low-power aim of the algorithm. Preliminary estimates of the power consumption for all the blocks required shortly can be found in [23].

### C. Operation

Given the limited choice of signal processing constructs available, the algorithm itself if based upon [19], [24], and [25], which report CWT-based algorithms that obtain suitable detection performance. With some modifications, which are detailed shortly, all of the required blocks can be made to operate in continuous time, thus ensuring real-time operation, and can be mapped to either analog or digital domain circuits, giving the maximum possible flexibility at the hardware implementation stage.

The final algorithm is shown in Fig. 2, and its operation can be broken down into several stages.

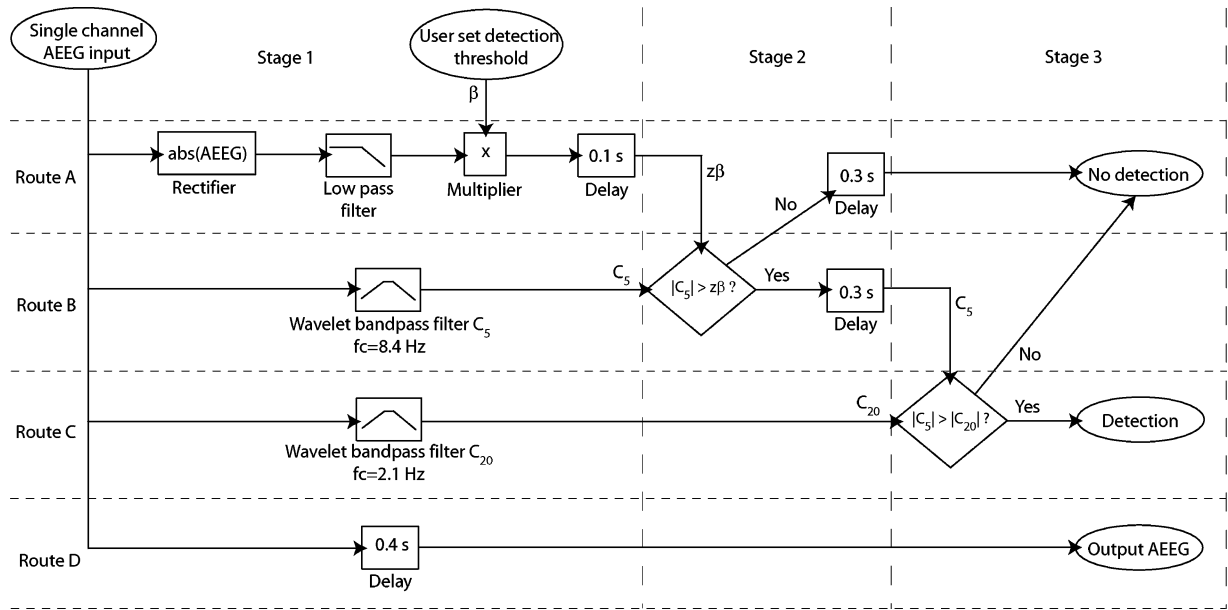


Fig. 2. Detailed operation of the AEEG data reduction algorithm proposed. The core algorithm is based around signal normalization (route A) and frequency extraction by two wavelet filters (routes B and C) with thresholding used to determine whether a detection is made. Not shown: detections in multiple channels are then combined so that a detection in any one causes all of the channels to record, and a memory buffer is incorporated to allow AEEG from before and after a detection to be recorded.

*Stage 1:* Stage 1 of the algorithm has three continuous time filters acting on the input AEEG data in parallel, thus forming three processing routes, A–C.

Route A is used for signal normalization to correct for broad-level amplitude changes in the AEEG signal that may occur in different parts of the head, between different patients, and due to amplification or attenuation due to the quality of the electrode contact. This normalization allows the use of a fixed detection threshold later in the algorithm regardless of the amplitude of the input signal.

As the wavelet transform is linear, normalization of the wavelet coefficients requires an output that is directly proportional to the input signal amplitude. This is provided by an envelope detector. A full-wave rectifier ( $\text{abs}(\text{AEEG})$ ) followed by a second-order low-pass filter (Butterworth approximation, cutoff 0.16 Hz) is used. This produces an output  $z$ , which is multiplied by a user-set threshold  $\beta$  to provide user control over the algorithm operation.

In comparison, stage 1 for signal routes B and C consists of performing the CWT at two discrete scales. Mathematically, the CWT at a single scale is equivalent to a single bandpass filter [26]; if the mother wavelet to be used is defined by  $\psi(t)$ , the required bandpass filter has impulse response

$$h(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{-t}{a}\right). \quad (1)$$

The operation thus requires two continuous time bandpass filters that can readily be implemented in the analog or digital domain. Following [19], [24], and [25], the Mexican hat mother wavelet, which is illustrated in Fig. 3(a), is used.

In reality, however, in order to be suitable for ultralow power on-chip implementation following techniques, such as those in [21] and [22],  $h(t)$  cannot be implemented directly. Instead,

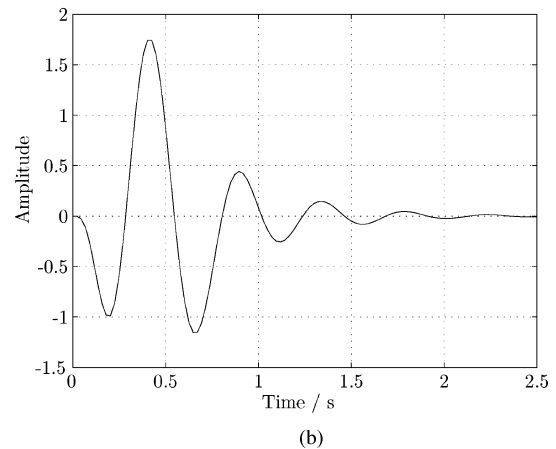
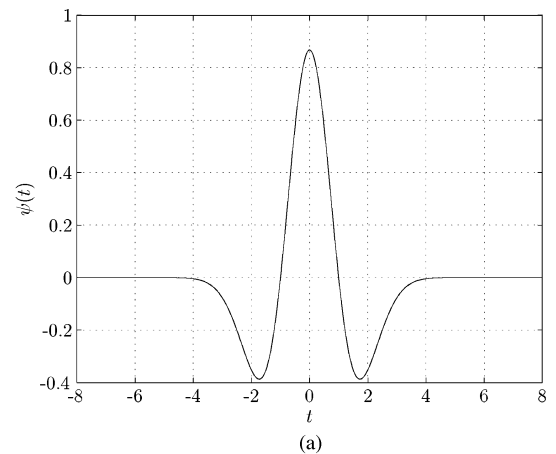


Fig. 3. Comparison of the shapes of the Mexican hat mother wavelet and the implemented approximation. Note that numerical values differ due to a change in wavelet scale between the two figures. (a) Mexican hat mother wavelet (scale = 1). (b) Implemented wavelet function corresponding to a bandpass filter with center frequency 2.1 Hz (scale = 0.1).

TABLE I  
s-DOMAIN TRANSFER FUNCTIONS OF TWO CWT BANDPASS FILTERS USED

$C_5$ :	
$H(s)$	$= \frac{-2.15 \times 10^{-4} s^2}{1.43 \times 10^{-12} s^7 + 3.23 \times 10^{-10} s^6 + 3.61 \times 10^{-8} s^5 + 2.65 \times 10^{-6} s^4 + 1.35 \times 10^{-4} s^3 + 0.005 s^2 + 0.10 s + 1}$ (2)
$C_{20}$ :	
$H(s)$	$= \frac{-6.88 \times 10^{-3} s^2}{2.34 \times 10^{-8} s^7 + 1.34 \times 10^{-6} s^6 + 3.70 \times 10^{-5} s^5 + 6.79 \times 10^{-4} s^4 + 8.67 \times 10^{-3} s^3 + 0.075 s^2 + 0.40 s + 1}$ (3)

an approximation is required in order to make a finite order, stable filter. For conciseness, the approximation used here is not repeated in depth, but can be found in [27].

Essentially,  $h(t)$  is delayed by an amount  $T$ , and then approximated by a truncated Maclaurin series. The end result is for each analysis scale a transfer function describing a two-zero, seven-pole bandpass filter in the  $s$ -domain. Transfer functions for both the Mexican-hat-based wavelet filters are given in Table I. For comparison with the shape of the Mexican hat mother wavelet, the route C filter impulse response is shown in Fig. 3(b).

Two wavelet transform filters are thus used:  $C_5$ , which is at scale 0.025, with a center frequency of 8.4 Hz and transfer function (2) (in Table I); and  $C_{20}$ , which is at scale 0.1, with a center frequency of 2.1 Hz and transfer function (3) (in Table I). The aim of route B, which uses the  $C_5$  filter is to identify possible spikes. Route C, using the  $C_{20}$  filter provides a simple rule to reject artefacts and incorrect detections.

Note that each wavelet filter uses a different value for the delay  $T$ , and this must be compensated for, as shown in Fig. 2, to ensure that all coefficients are calculated at the correct time.

*Stage 2:* In stage 2 of the algorithm, the information from the  $C_5$  wavelet filter is used to determine whether a candidate spike is present by performing the comparison

$$|C_5| > |z\beta|. \quad (4)$$

A magnitude comparison is performed to allow the detection of spikes of different polarities. Note that this is equivalent to performing the comparison

$$\left(\frac{C_5}{z}\right)^2 > \beta^2 \quad (5)$$

where  $(C_5/z)^2$  corresponds to the normalized wavelet power. For this reason, detection threshold values in the remainder of this paper are given in terms of  $\beta^2$ .

*Stage 3:* In stage 3 of the algorithm, the information from the  $C_{20}$  wavelet filter is used to reject some of the candidate spikes from stage 2. It thus provides a simple rule to reject artefacts and incorrect detections by ensuring that the normalized power in the signal band ( $C_5$ ) is larger than the normalized power in this artefact band

$$|C_5| > |C_{20}|. \quad (6)$$

If both conditions (4) and (6) are satisfied a detection of a spike event to be recorded is made. This corresponds to a yes/no flag.

All of the components required in stages 1–3 operate in continuous time, and so, intrinsically in real time, although there is

a 0.4 s delay between the input data and an output decision due to the wavelet approximation method used. Route D is present to delay the AEEG signal itself by a similar amount so that it is correctly passed on to the data recording unit.

Note that this organization is highly useful for the eventual hardware implementation of the algorithm. The AEEG that is recorded (route D) is separate from the AEEG that is analyzed by the algorithm (routes A–C). Thus, although route D has to operate with sufficient performance to allow a high-quality AEEG recording (in terms of dynamic range, sampling frequency, and similar characteristics), the data reduction algorithm itself does not, thus potentially allowing much simpler hardware implementations.

*Multiple Channel Operation and Recording Process:* The algorithm considered here operates on ten EEG channels in parallel and a simple OR gate of the detection output flag from each single channel algorithm is used so that a detection in any one channel causes all of the channels to record. The ten channels for analysis are selected purely as the channels common to all of the datasets used (see Section III-C) and are: F7, F8, Fp1, Fp2, O1, O2, T3, T4, T5, and T6.

When a detection is made, it triggers the recording process to record a set period of the AEEG trace from before and after the flag (the recording window), which requires using some buffer memory to store data from before the detection. This process is applied to all of the channels in the AEEG recording, regardless of whether they have been passed through the detection algorithm.

### III. DATA REDUCTION TEST SETUP

#### A. Analysis Method

For the discontinuous AEEG recording unit, the key performance factors of interest are the sensitivity and the percentage of data transmitted. (See [16] for a full discussion of optimal performance metrics.) The sensitivity is the fraction of expert marked events that are correctly recorded

$$\text{sensitivity} = \frac{\text{number of correct detections}}{\text{total number of marked events}} \times 100\% \quad (7)$$

and illustrates how much of the wanted AEEG information is correctly recorded. A high sensitivity is wanted for good performance.

The percentage of data transmitted provides a measure of the amount of data reduction achieved, illustrating how much of the AEEG is selected for recording and how much is discarded. A low value is wanted for good data reduction performance.

The algorithm of Fig. 2 is tested by analyzing EEG datasets as the value  $\beta^2$  is varied over the range 0.1–1 in 0.05 steps. At each value of  $\beta^2$ , values for the sensitivity  $S$  and percentage of data reduction  $C$  are found, thus allowing the performance to be illustrated.

In this study, two methods for assessing the average level of performance are used. If  $M$  EEG records are analyzed and the  $i^{\text{th}}$  record has a duration  $T_i$ , with  $N_i$  expert marked events and  $D_i$  correctly detected events, the sensitivity in any record is given by  $D_i/N_i$ . At each  $\beta^2$  threshold used with the algorithm

the value  $D_i$  changes. The overall sensitivity for all tests at each threshold value can then be reported as the

$$\text{total sensitivity} = \frac{1}{\sum_{i=1}^M N_i} \sum_{i=1}^M D_i \times 100\% \quad (8)$$

to accurately represent the total number of detections made, or as the

$$\text{time/event weighted sensitivity} = \frac{1}{\sum_{i=1}^M \frac{T_i}{N_i}} \sum_{i=1}^M \frac{D_i T_i}{N_i N_i} \times 100\% \quad (9)$$

which provides robust normalization for the amount of data tested [16]. The percentage of data transmitted is calculated analogously. As the threshold value is varied, a range of  $S$  and  $C$  values are produced, and these can be plotted on a tradeoff curve.

This analysis is repeated several times using different recording windows, allowing the user to choose the amount of background information that is saved in response to each detection of a candidate spike. Events are deemed to be successfully recorded if a detection occurs no more than 4, 2, 1, or 0.4 s away from the marked position for the 10, 5, 2.5, and 1 s recording windows, respectively. This ensures that a reasonable amount of background signal is recorded in response to any detection.

In addition to these curves, the graphs of Section IV also illustrate the performance of an algorithm that simply selects arbitrary data sections. For example, if, as a first-pass approximation, interictal events are assumed to occur at random times, sending 10% of the raw data will result in a sensitivity of 10%. This line thus represents the level of chance detections and any data selection scheme based upon detecting spikes should always achieve better performance than this.

## B. Results Produced

Testing of the algorithm is complicated slightly by the fact that not all EEG records contain events of interest to record. As a result, it is not possible to obtain sensitivity values for these records. In the results of Section IV up to five separate results are plotted, and these are calculated as follows.

1) *Total Sensitivity Average*: An average value for each  $\beta^2$  threshold used with the algorithm is calculated. The  $S$  value is generated by analyzing the only tests that contain marked events and the values are combined using (8). The  $C$  value is calculated using all of the data with the results between tests being combined using an equivalent of (8).

2) *Time/Event Weighted Average*: Again, an average value is calculated for each value of  $\beta^2$  used. In this case, both the  $S$  and  $C$  values are calculated using only the data that contain marked events with results combined between different tests using (9).

3) *Individual Results*: If the dataset being analyzed contains one or more marked events each time the algorithm is run at a particular threshold  $\beta^2$ , an individual  $S$  and  $C$  value is found. This pair of values is then plotted. These results are thus only shown for the tests that contain marked events.

4) *Max–Min Performance Limit*: This is the convex hull of the individual results as defined earlier.

5) *Constant  $\beta^2$  Contour*: This is a line connecting all of the individual results that were calculated at the same  $\beta^2$  value. It is drawn from the highest sensitivity point to the lowest sensitivity point (so cannot *double back* in this direction).

In the earlier cases, the drawing of the averages and the max–min performance limit include the *known* points that if no data are sent, the sensitivity is 0%, and similarly that if all of the data are sent, the sensitivity is 100%.

## C. EEG Data Used

The EEG data used in this analysis are summarized in Table II. A total of 982 interictal events in just over 4 days of data from 18 patients are analyzed. Note that, in general, the total data for each patient may be made up of more than one noncontinuous dataset, with 30 data files being analyzed in total. A range of sleep and wakefulness data are analyzed. To find the sensitivity value, interictal events that should be recorded have been marked by one expert from the National Society for Epilepsy in the U.K., with the interpretation of the markings approved by a second person.

All recordings are performed with a referential montage (FCz reference) and are analyzed in the recorded montage, in line with the online aim of the algorithm. Most of the recordings have more than the ten channels used in the algorithm procedure, but no analysis is performed on these excess channels. The ten channels used are selected as the subset of channels common to all of the available EEG records, and are not based on any *a priori* assumption. It is possible that a different algorithm performance will be obtained if this subset of channels is changed. The sampling rate varies between 200 and 256 Hz depending on the type of recording, and the implemented algorithm is independent of this.

Overall, this dataset should be sufficient for testing of the algorithm (see Section III-D), although it is possible to note a number of limitations that are present, and so should be kept in mind when interpreting the results.

1) *Data Type*: An algorithm for use in AEEG systems is being developed, and so, ideally, all test data should come from AEEG recordings. However, to form a suitably large test dataset, it has not been possible to analyze only ambulatory collected data. Data are available from a range of fMRI/EEG, long-term, routine, and ambulatory recordings. Different types and frequencies of artefacts may be present in the different types of recording that may affect the performance. Nevertheless, a preliminary version of the algorithm detailed in [11] analyzed the ambulatory and nonambulatory recordings separately, concluding that the algorithm performance is not overestimated by analyzing all of the available data together.

2) *Number of Events*: No restriction is placed on the number of events needed in a file for it to be suitable for analysis, although low numbers of events limit the potential resolution of the sensitivity found in any one test.

3) *Medication*: Any medication taken by patients has not been normalized.

TABLE II  
DATA AVAILABLE FOR ANALYSIS (NOTE THAT DATA FOR EACH PATIENT MAY BE MADE UP OF MORE THAN ONE NONCONTINUOUS TEST)

Patient	Age at test	Gender	Type of recording	Sampling rate / Hz	Marked interictal events	Recording duration
0	Unknown	Unknown	fMRI/EEG	200	644	00:36:55
1	24	Male	Long term monitoring	256	49	03:58:15
2	47	Female	AEEG	208	7	02:00:11
3	33	Female	AEEG	208	52	06:00:33
4	51	Unknown	AEEG	208	12	04:00:22
5	23	Female	AEEG	208	11	04:00:22
6	44	Female	Routine EEG	256	0	00:46:19
7	43	Male	Routine EEG	256	2	00:18:30
8	46	Male	AEEG	208	26	02:00:11
9	45	Male	Long term monitoring	256	30	04:00:22
10	23	Female	AEEG	208	45	04:00:22
11	53	Female	AEEG	208	8	02:00:11
12	27	Female	Long term monitoring	256	0	02:00:11
13	21	Female	AEEG	208	12	02:00:12
15	33	Unknown	Routine EEG	256	1	00:10:53
19	Unknown	Unknown	AEEG	200	28	22:57:05
20	Unknown	Unknown	AEEG	200	15	22:06:23
21	Unknown	Unknown	AEEG	200	40	21:05:43
Total	-	-	-	-	982	104:03:00

Also, some of the recordings used are d.c. coupled and so contain noticeable d.c. offsets. In reality, it is anticipated that an AEEG unit would be a.c. coupled, and so, all of the EEG data are high-pass filtered using a first-order filter with cutoff frequency 0.16 Hz (as recommended in [28]) before being processed by the algorithm.

#### D. Confidence Intervals

In order to avoid overcomplicating the graphs in Section IV, confidence intervals are not plotted on each result graphic. Instead, following the procedure from [29], confidence intervals for the total sensitivity average results are calculated, and these are used to indicate that the algorithm has been tested with sufficient data to ensure confident results, given the assumption that the test database is representative. As in [29], confidence intervals for the performance (sensitivity) and cost (percentage of data transmitted) are plotted separately.

1) *Sensitivity*: Confidence intervals are generated from a binomial distribution: if there are  $N$  events and the reported sensitivity is  $S$  a distribution  $B(N, S/100)$  is used to estimate the 95% two-tailed confidence intervals.

Given an observed sensitivity  $S$ , these intervals show that range that the true value of  $S$  could reasonably occur and still produce the observed result by chance. The intervals are illustrated in Fig. 4(a) for a reported sensitivity of 80% [29]. Note that the binomial distribution assumes that the detection of any one event is independent from the detection of any other, but this seems reasonable when testing a large amount of data from multiple people.

The graph from Fig. 4(a) can be replicated for different reported sensitivity values, and the largest confidence intervals are found for a reported sensitivity of 50%. At this sensitivity, for the 982 events used here, it is possible for the reported sensitivity to be overestimated by up to 3.17%. For most  $S$  values,

however, the confidence intervals will be smaller than this. The confidence intervals required are tabulated in [30].

2) *Percentage of Data Transmitted*: A binomial distribution is not suitable for use in this case, as there is no analog of the value  $N$ . Instead, in lack of a better probability distribution suitable for use, a procedure based on the false detection rate is used to approximate the confidence intervals.

Confidence intervals for the false detection rate are first found from a Poisson distribution [29]: false detections are assumed to be rare, occur at a fixed rate, and could occur at any instant in time. For this, 95% two-tailed confidence intervals are thus generated. Assuming that false detections are rare, when recording the EEG around one false detection, it will not *overlap* with the EEG recorded from another false detection. There is thus a fixed ratio between the false detection rate and the amount of data transmitted, allowing the wanted confidence intervals to be calculated.

Results for a  $C$  value of 20% are shown in Fig. 4(b). With four days of data and 20% reported  $C$ , the confidence intervals range from approximately 19.4% to 20.4%, showing that the amount of data transmitted will not be underestimated by more than 0.4%.

Both of the  $S$  and  $C$  confidence interval figures show that sufficient data are being tested in order to have reasonable confidence that the results produced are representative of the algorithm performance, and are unlikely to have occurred by chance.

## IV. RESULTS

### A. Average Results

The total sensitivity performance of the algorithm is shown in Fig. 5. This shows a good level of average performance: using any recording window, it is possible to correctly record over 90% of the expert-marked events while only sending 50% of the entire EEG record. If only 80% sensitivity is deemed acceptable

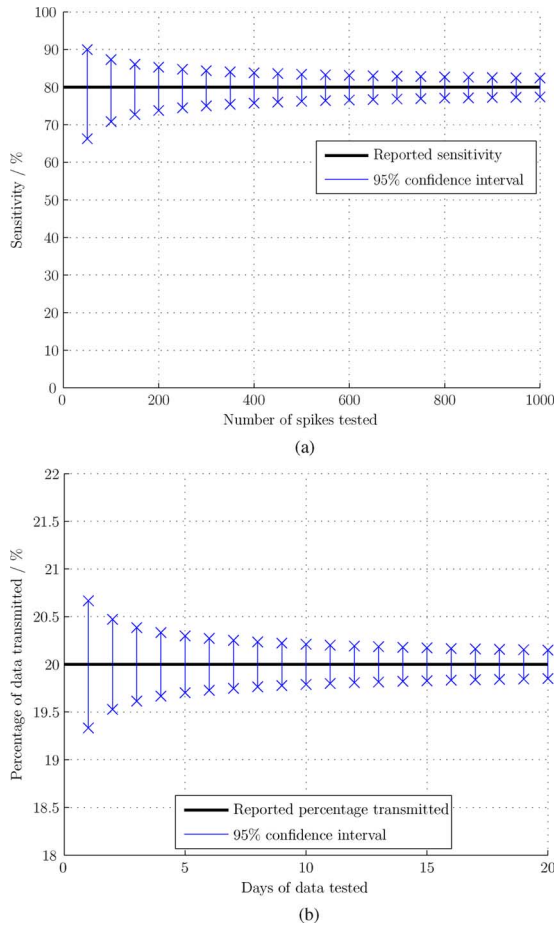


Fig. 4. Estimated 95% two-tailed confidence intervals showing the range that given the observed result, the true value could reasonably lie within. There is a 95% chance that the true sensitivity is within the interval shown. (a) Sensitivity intervals for 80% performance. (b) Percentage of data reduction intervals for 20% performance.

data reductions of 70% or more are achievable. For comparison, Fig. 6 shows the time/event weighted average. Overall sensitivities comparable to those in Fig. 5 (over 90%) are possible, but they involve sending more of the raw data, up to 73% for the 10-s window.

As an illustration of the algorithm complexity, the algorithm (including calculation of the performance metrics) operates approximately 40 times faster than real time when simulated in MATLAB on a machine with an Intel Core 2 Duo processor and 3 GB of RAM.

### B. Variance Between Tests

Fig. 7 illustrates the variance of the algorithm performance between different tests using the 5-s recording window. At a threshold value of  $\beta^2 = 0.25$  (empirically chosen), a range of performances are present. Some tests achieve 100% sensitivity with a good amount of data reduction, while others require nearly all of the data to be transmitted. Still other tests have a very low sensitivity and require significant amounts of data to be sent. Some variance in the performance is naturally expected, although a desired amount has not yet been specified. This is discussed further in Section V.

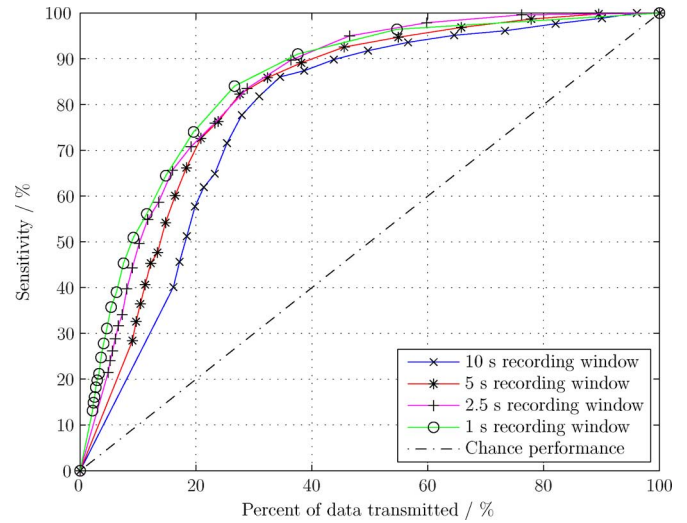


Fig. 5. Performance of the algorithm using the total sensitivity averaging method.

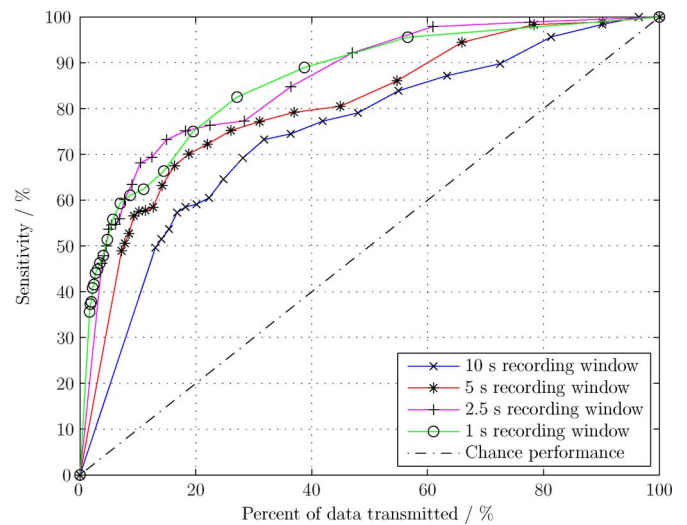


Fig. 6. Performance of the algorithm using the time/event weighted averaging method.

### C. Overtransmission

Fig. 8 complements Fig. 7 to illustrate the amount of *unnecessary* data that are marked for transmission by the algorithm. This is done by considering the percentage of EEG data that would be transmitted if only the expert marked events were transmitted with 100% accuracy, producing a value  $C_I$ .

The  $x$ -axis of the results graph is thus modified to show  $C - C_I$ , or the overtransmission difference, which illustrates the excess data transmitted. From this it is clear that a large number of false detections are present resulting in excess data being transmitted. Comparing Fig. 8 with Fig. 7, however, it can be seen that all of the individual results in the top right-hand corner, those for which good sensitivities are achieved, but at the cost of sending lots of data, are moved to the left in Fig. 8. This indicates that these records have roughly the same number of false detections as the other records, but more events are present and this intrinsically requires more data to be sent.

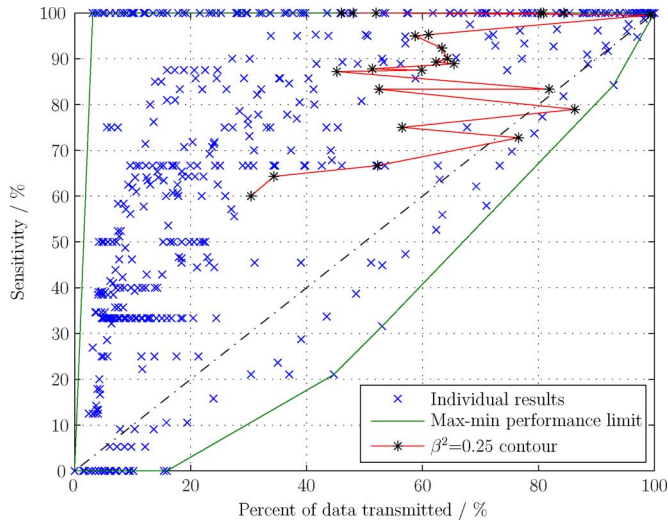


Fig. 7. Results showing the max–min performance limit and constant  $\beta^2$  contour to show how the algorithm performance varies between tests with a 5-s recording window.

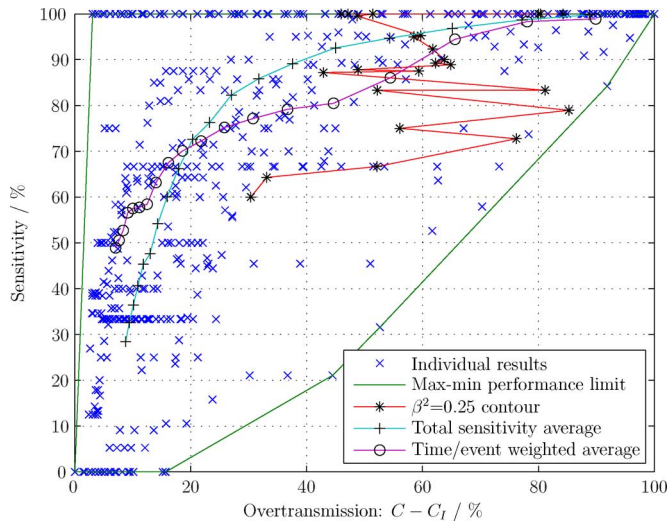


Fig. 8. Overtransmission difference showing the amount of unnecessary data transmitted.

## V. ANALYSIS AND DISCUSSION

### A. Statistical Validation

Ideally, the results curves from Section IV need to answer the following question: is the algorithm performance statistically good enough? Unfortunately, this is not easily done, and it is likely that the closest possible test at this point is showing that the algorithm performance is statistically better than chance performance. This would illustrate that the algorithm does have some skill and acts as a basis point for further improvement.

To do this, the Mann–Whitney  $U$ -test is used. This is a non-parametric test and is frequently used for testing the areas under receiver operating characteristic (ROC) curves [31]. For each curve in Section IV,  $y$ -axis points corresponding to sensitivity values are used to form the set  $s$ , and  $x$ -axis points corresponding to percentage of data transmitted values to form the set  $c$ . The following null hypothesis is then considered [32]:

TABLE III  
 $U$  VALUES FROM MANN–WHITNEY  $U$ -TEST APPLIED TO FIGS.5 AND 6

Recording window size / s	Total sensitivity graph (Fig. 5)	Time/event weighted graph (Fig. 6)
1	35	14
2.5	44	24
5	49	40
10	54	65

“There is no tendency for members of set  $s$  to exceed members of set  $c$ .”

From the aforementioned hypothesis, if  $S$  values tend to be the same as  $C$  values, the algorithm performance would be along the chance line ( $y = x$ ). If they are statistically different, the performance must be better than chance. Note that the Mann–Whitney  $U$ -test assumes that samples in  $s$  and  $c$  are independent, which seems reasonable provided that spikes are rare events: with any level of transmission it is possible to get any sensitivity; it is not necessary to have a large  $C$  value to get a large  $S$  value.

For calculation, the  $x$  and  $y$  values from the result curves are extracted to form sets  $s$  and  $c$  with the known end points at 0% and 100% excluded. Each set, thus, has 19 entries ( $n_1 = n_2 = 19$ ) corresponding to the thresholds used in the algorithm.  $U$  values are then calculated following the procedure detailed in [32] and compared to the significance levels in the same reference. Calculated  $U$  values for the total sensitivity and time/event weighted results are shown in Table III.

All the  $U$  values given in Table III are below the critical  $U$  value, and so, the null hypothesis is rejected ( $p = 0.05$  two-tailed test,  $n_1 = n_2 = 19$ ,  $U_{crit} = 113$ ,  $U_{max} = 65$ ). It is, thus, concluded that the performance of the algorithm is statistically superior to that of a chance classifier.

### B. Discussion

Overall, the performance level seen in Fig. 5, correctly recording more than 90% of the expert marked events while only sending 50% of the entire EEG record, is deemed to be a good starting performance level. Considering the graph of Fig. 5 more closely, the performance shown can be roughly broken down into two sections.

First, starting in the bottom left of the graph, the sensitivity increases very quickly for little extra data being sent. This must correspond to the easy events being detected with relatively few false detections. In the second section, the performance levels off between 80% and 90% sensitivity, and significantly more data are required to capture the remaining events.

A similar shape is seen for the time/event weighted results in Fig. 6, although the overall performance level is relatively poor for the larger recording windows, with data reductions of the order of 30% being achieved. This result is in line with the conclusions from [16], whereby the time/event weighting provides robust normalization for the amount of data tested, but can underestimate the actual number of events detected.

Even with this metric, though, suitably high sensitivities can be obtained while providing data reduction. If the algorithm operates in suitably low power, turning off the high-power



transmission stage, if even for only a short time, can still give significant overall power savings.

To determine whether such detection performance is acceptable, Wilson *et al.* [33] note that when multiple human interpreters (the current gold standard) are asked to interpret and mark the same EEG recording, they disagree with each other anywhere between 0% and 90% of the time. Likewise, as suggested in [14], it is unlikely that all spikes need to be detected by an automated algorithm to enable accurate diagnosis. The aim of diagnosis is to pool all of the available information to enable a decision based on the balance of probabilities to be made. The presence or absence of a small number of spikes should not be a critical factor in this decision process.

Given these facts, a 90% sensitivity and 50% data reduction, is likely sufficient for an initial implementation. Sending 50% of the data still leaves a large amount of background material to help with the diagnosis, and allows for relatively low *trust* in the system; significantly, more than the minimum data required are still presented to the interpreting neurologist allowing them to be in control of the diagnosis procedure.

The overtransmission graph of Fig. 8 and common sense, however, indicate that a 30%–50% data reduction still corresponds to a large number of false detections, and there is potentially significant room for improvement. Comparison of the performance with offline spike detection algorithms is difficult as most algorithms report a rate of false detections rather than a data reduction, and the link between false detection rate and data reduction is not necessarily simple.

For example, in the algorithm considered here, each spike is actually detected multiple times at instances close to the spike center. This is desirable as it helps ensure the total spike duration is recorded. For comparison with other methods, however, are all of these correct detections, or should only one be considered correct with the rest being false? It is not possible to ensure that the same criteria apply across all studies from the literature.

A rough comparison of algorithm performances is possible along the following lines, however. From a 2002 review [17], false detection rates for algorithms that achieve sensitivities over 90% range between 1.4 and 4 per minute. Assuming no overlap between the false detection recording periods, the data reduction for a 5-s recording window varies between 67% and 88%, and for a 10-s recording window between 33% and 77%. It thus ranges from reasonably similar to the performance here, to noticeably better. Similarly, reported EEG data compression techniques (which aim to represent the EEG information in fewer digital bits, rather than the discontinuous method investigated here) report data reductions of approximately 40% using lossless encoding (based on Huffman coding) [34], and up to 90% data reduction when lossy transform-based compression (based upon thresholding wavelet coefficients) is used [35].

Results such as these, however, are entirely expected. It is obvious to some extent, that if more signal processing *power* is used it should be possible to achieve a better tradeoff between the sensitivity and the data reduction. The algorithm proposed here has traded off this performance for an algorithm that has an intrinsic, and direct, low-power implementation. For a fair comparison of algorithms, the three-way tradeoff between sensitivity, data reduction, and power consumption should be

considered. Given that figures are not available at this point, the initial performance here does not seem unreasonable. Also, the use of the online system presented here to facilitate device miniaturization and longer term monitoring does not preclude the use of an offline spike detection algorithm to further improve the performance prior to clinical review, if so desired.

More questionable, with regards to the overall performance, is whether the amount of variance in performance seen in Figs. 7 and 8 is acceptable. As noted earlier, due to the anticipated large asymmetry between the required algorithm power consumption and the wireless transmitter power consumption, only a very modest data reduction is required to begin achieving significant power savings. The limiting factor in the variance is thus deemed to be the sensitivity achieved.

For a 5-s recording window (see Fig. 7), 6 of the 24 individual results achieve less than 80% sensitivity. The algorithm performance is thus unacceptably low for 25% of test cases. With the lack of more detailed assessment criteria at this point, it is deemed that this level of variance is acceptable for the first-pass algorithm.

Nevertheless, there are certainly improvements and extensions to the algorithm that are possible and should improve this performance. For example, observations of the wavelet coefficients  $C_5$  and  $C_{20}$  in the cases where spikes are not correctly recorded indicate that the origin of the performance variance seen lies with suboptimal choices for the wavelet shape and scale. Optimizing these, possibly on a patient-specific basis, could lead to a decrease in the sensitivity variance.

However, before proceeding on algorithm improvement and optimization, it is first essential to answer the following questions.

- 1) What power level can the algorithm be implemented in, and is this in line with expectations?
- 2) What is the diagnostic impact of the method? Can accurate diagnoses still be made from the reduced datasets created? If so, is it possible to more accurately fix required performance levels for the performance to be deemed acceptable?

In the authors' opinions, it is essential to answer these questions first before trying to optimize the algorithm performance from its already satisfactory point, and an amount of iteration will necessarily be required.

It is thus proposed to accept the current, promising level of algorithm performance in order to allow some of the research questions highlighted earlier to be addressed. Undoubtedly, there is still a significant amount of work required in order to realize online data compression with power consumption of the order of microwatts and that is fully verified through clinical testing. The algorithm proposed and tested here, however, can act as a sound starting point for this, moving us closer to wearable, long-lasting AEEG devices.

## VI. CONCLUSION

This paper has presented an online data reduction methodology for use in AEEG systems for epilepsy diagnosis. By discarding uninteresting, background, EEG signals and only transmitting potentially interesting signals for human analysis significant power can be saved in the wireless transmitter.

An algorithm for this, based around the CWT, has been presented with its performance evaluated through MATLAB simulations. This found that a 90% sensitivity, with a 50% data reduction is achievable and this is deemed to be an acceptable starting level. Before attempting to improve this level, it is essential to verify that the algorithm can indeed be implemented on-chip using the submicrowatt power levels that are required.

#### ACKNOWLEDGMENT

The authors would like to acknowledge and thank S. Patel and D. Yates for their contributions to earlier iterations of the algorithm presented here, and the National Society for Epilepsy in the U.K. for their assistance.

#### REFERENCES

- [1] C. D. Binnie, A. J. Rowan, and T. Gutter, *A Manual of Electroencephalographic Technology*. Cambridge, U.K.: Cambridge Univ. Press, 1982.
- [2] P. E. M. Smith and S. J. Wallace, *Clinicians' Guide to Epilepsy*. London, U.K.: Arnold, 2001.
- [3] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *JNNP*, vol. 76, no. 2, pp. ii2-ii7, 2005.
- [4] E. Waterhouse, "New horizons in ambulatory electroencephalography," *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 74-80, May/June 2003.
- [5] D. Smith, B. A. Defalla, and D. W. Chadwick, "The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic," *QJM*, vol. 92, no. 1, pp. 15-23, 1999.
- [6] A. J. Casson, S. Smith, J. S. Duncan, and E. Rodriguez-Villegas, "Wearable EEG: What is it, why is it needed and what does it entail?," in *Proc. IEEE EMBC*, Vancouver, Aug. 2008, pp. 5867-5870.
- [7] E. Jovanov, J. Price, D. Raskovic, K. Kavi, T. Martin, and R. Adhami, "Wireless personal area networks in telemedical environment," in *Proc. IEEE EMBS ITAB*, Arlington, Nov. 2000, pp. 22-27.
- [8] K. Baert, B. Gyselinckx, T. Torfs, V. Leonov, F. Yazicioglu, S. Brebels, S. Donnay, J. Vanfleteren, E. Beyne, and C. Van Hoof, "Technologies for highly miniaturized autonomous sensor networks," *Microelectron. J.*, vol. 37, no. 12, pp. 1563-1568, 2006.
- [9] D. C. Yates and E. Rodriguez-Villegas, "A key power trade off in wireless EEG headset design," in *Proc. IEEE EMBS Neural Eng. (NER)*, Kohala Coast, HI, May 2007, pp. 453-456.
- [10] A. J. Casson, D. C. Yates, S. Patel, and E. Rodriguez-Villegas, "Algorithm for AEEG data selection leading to wireless and long term epilepsy monitoring," in *Proc. IEEE EMBC*, Lyon, France, Aug. 2007, pp. 2456-2459.
- [11] A. J. Casson and E. Rodriguez-Villegas, "On data reduction in EEG monitoring: Comparison between ambulatory and non-ambulatory recordings," in *Proc. IEEE EMBS*, Vancouver, BC, Canada, Aug. 2008, pp. 5885-5888.
- [12] A. J. Casson and E. Rodriguez-Villegas, "Data reduction techniques to facilitate wireless and long term AEEG epilepsy monitoring," in *Proc. IEEE EMBS Neural Eng. (NER)*, Kohala Coast, HI, May 2007, pp. 298-301.
- [13] J. Gotman, "Automatic detection of seizures and spikes," *J. Clin. Neurophysiol.*, vol. 16, no. 2, pp. 130-140, 1999.
- [14] J. Gotman, "Automatic recognition of interictal spikes," in *Long-Term Monitoring in Epilepsy*, J. Gotman, J. R. Ives, and P. Gloor, Eds. Amsterdam, The Netherlands: Elsevier, 1985, pp. 93-114.
- [15] Medtronic. (2006). Medtronic: Home page [Online]. Available: <http://www.medtronic.com>
- [16] A. J. Casson, E. Luna, and E. Rodriguez-Villegas, "Performance metrics for the accurate characterisation of interictal spike detection algorithms," *J. Neurosci. Methods*, vol. 177, no. 2, pp. 479-487, 2009.
- [17] S. B. Wilson and R. Emerson, "Spike detection: A review and comparison of algorithms," *Clin. Neurophysiol.*, vol. 113, no. 12, pp. 1873-1881, 2002.
- [18] L. Senhadji and F. Wendling, "Epileptic transient detection: Wavelets and time-frequency approaches," *Neurophysiol. Clin.*, vol. 32, no. 3, pp. 175-192, 2002.
- [19] M. Latka, Z. Was, A. Kozik, and B. J. West, "Wavelet analysis of epileptic spikes," *Phys. Rev. E*, vol. 67, no. 5, pp. 902-904, 2003.
- [20] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, no. 4, pp. 626-638, Apr. 1996.
- [21] S. A. P. Haddad, S. Bagga, and W. A. Serdijn, "Log-domain wavelet bases," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 52, no. 10, pp. 2023-2032, Oct. 2005.
- [22] P. Agostinho, S. Haddad, J. De Lima, and W. Serdijn, "An ultra low power CMOS pA/V transconductor and its application to wavelet filters," *Analog Integr. Circuits Signal Process.*, vol. 57, no. 1, pp. 19-27, 2008.
- [23] A. J. Casson and E. Rodriguez-Villegas, "Generic vs. custom; analogue vs. digital: On the implementation of an online EEG signal processing algorithm," in *Proc. IEEE EMBC*, Vancouver, BC, Canada, Aug. 2008, pp. 5876-5880.
- [24] N. Coninx, "Automated detection of epileptic events in the interictal EEG using the wavelet transform," Bachelor's thesis, Univ. Maastricht, Maastricht, The Netherlands, 2005.
- [25] S. Patel, "Intelligent low power circuit for electroencephalography," Master's thesis, Imperial College Lond., London, U.K., 2006.
- [26] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. San Diego, CA: Academic, 1999.
- [27] A. J. Casson, D. C. Yates, S. Patel, and E. Rodriguez-Villegas, "An analogue bandpass filter realisation of the continuous wavelet transform," in *Proc. IEEE EMBC*, Lyon, France, Aug. 2007, pp. 1850-1854.
- [28] M. R. Nuwer, G. Comi, R. Emerson, A. Fuglsang-Frederiksen, J.-M. Guerit, H. Hinrichs, A. Ikeda, F. J. C. Lucas, and P. Rappelsberger, "IFCN standards for digital recording of clinical EEG," in *Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiology (Electroencephalography and Clinical Neurophysiology Supplement 52)*, 2nd ed., G. Deuschl and A. Eisen, Eds. Amsterdam, The Netherlands: Elsevier, 1999, pp. 11-14.
- [29] J. C. Sackellares, D.-S. Shiau, K. M. Kelly, and S. P. Nair, "Testing a prediction algorithm: Assessment of performance," in *Seizure Prediction in Epilepsy*, B. Schelter, J. Timmer, and A. Schulze-Bonhage, Eds. Weinheim, Germany: Wiley, 2008, pp. 249-259.
- [30] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, 6th ed. Edinburgh, U.K.: Oliver and Boyd, 1963.
- [31] S. J. Mason and N. E. Graham, "Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation," *Q. J. R. Meteorol. Soc.*, vol. 128, pp. 2145-2166, 2002.
- [32] M. Bland, *An Introduction to Medical Statistics*, 3rd ed. Oxford, U.K.: Oxford Univ. Press, 2000.
- [33] S. B. Wilson, R. N. Harner, F. H. Duffy, B. R. Tharp, M. R. Nuwer, and M. R. Sperling, "Spike detection. I. Correlation and reliability of human experts," *Electroencephalogr. Clin. Neurophysiol.*, vol. 98, no. 3, pp. 186-198, 1996.
- [34] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 2, pp. 105-114, Feb. 1997.
- [35] J. Cardenas-Barrera, J. Lorenzo-Ginori, and E. Rodriguez-Valdivia, "A wavelet-packets based algorithm for EEG signal compression," *Med. Inf. Internet Med.*, vol. 29, no. 1, pp. 15-27, 2004.



**Alexander J. Casson** (S'07) received the M.Eng. degree in engineering science from the University of Oxford, Oxford, U.K., where he was a Lord Crewe scholar, in 2006. He is currently working toward the Ph.D. degree in low-power circuits and systems design at Imperial College London, London, U.K.

In particular, he has been engaged in algorithm and circuit design for online EEG data reduction.

Mr. Casson was awarded an Institution of Engineering and Technology postgraduate scholarship in 2008.



**Esther Rodriguez-Villegas** (SM'08) is currently a Senior Lecturer (Associate Professor) with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K., where she leads a team of nine Ph.D. students, specializing on ultralow power electronic circuits and systems for truly wearable physiological monitoring. She is the author or coauthor of more than 80 peer-reviewed papers and a book on floating-gate MOSFET (FGMOS) transistors, which was published by the Institution of Engineering and Technology in October 2006.