30th Annual International IEEE EMBS Conference
Vancouver, British Columbia, Canada, August 20-24, 2008

On data reduction in EEG monitoring: comparison between ambulatory and non-ambulatory recordings

Alexander J. Casson and Esther Rodriguez-Villegas

Abstract— **Objective: To compare the performance of an EEG data selection/reduction algorithm for epileptic EEGs on ambulatory and non-ambulatory recorded data to confirm that acceptable performance is achievable in ambulatory recordings despite the presence of overt artifacts. Methods: A total of 167 hours of EEG data containing 899 marked interictal events is analysed to determine the percentage of events correctly recorded (the sensitivity) and the amount of data reduction achieved. Results: A better sensitivity-data reduction trade-off is found in the ambulatory recorded data. This may be unexpected as ambulatory recordings are known to contain large numbers of artifacts, but is accounted for by these artifacts being easily detected and discarded, improving the data reduction. Conclusions: Satisfactory performance levels are found in both data types, no degradation is present with ambulatory recordings. Significance: Demonstrates that the processing of EEG data for wearable EEG applications is feasible without a loss in performance compared to traditional inpatient EEG usage.**

I. INTRODUCTION

The electroencephalogram, or EEG, is a classic method for measuring a person's *brainwaves*. Electrodes are placed on the scalp and these detect the micro-Volt sized signals that result outside the head due to the accumulated neuronal action within the brain. See [1] for an overview.

EEG has long been used in the diagnosis of epilepsy, a common neurological disorder that affects approximately 1% of the population and is characterised by unprovoked seizures [2]. These seizures manifest in the EEG allowing them to be recorded and accurate diagnoses made. Other features, such as *interictal spikes* (spike features in-between seizures) are also present in the EEG of epileptic patients and these can also be useful in diagnosis and the localisation of the epileptic loci within the brain. This is not least because seizures can be rare events and so recording them on a relatively short EEG can be difficult. Interictal events give secondary features that can be informative.

The trend in epilepsy diagnosis has thus progressed towards longer EEG recordings. However, this gives a large amount of data to be analysed by a human and so there has been a large amount of interest in automating both spike and seizure detection (see, for example, [3], [4]). At the same time, long term inpatient EEG recordings are resource intensive and not universally available [5]. As a result ambulatory EEG (AEEG) recordings are also available during which the patient has their EEG recorded on a portable unit while undertaking their normal daily life. These recordings cost approximately 50% of their inpatient counterparts [5] and are potentially integrated into body area networks for personalised medicine applications.

However, as the patient is not restricted to a single location for the recording of an AEEG there is the possibility for additional artifacts, not present in an inpatient recording, to occur. For example the quality of the recording may reduce as an electrode comes loose over time and there is no technician present to spot and correct this. Multiple studies, such as [6], have investigated the human interpretation of ambulatory and non-ambulatory recordings, demonstrating the clinical utility of AEEG. However, it is also essential to show that automated algorithms can perform satisfactorily upon AEEG data sets.

This is especially important as signal processing begins to be incorporated into the AEEG recording device. For example, the battery life of un-cumbersome, wireless, AEEG units is very limited, [7], and the authors have proposed the use of an onboard detection algorithm to select only *candidate* interesting sections of EEG to be recorded. This offers significant data reduction and so power savings and an increased operational lifetime [8].

The feasibility of detection algorithms applied to the AEEG has been demonstrated in, for example, [9]. This paper investigates the direct comparison of an algorithm performance when applied to ambulatory and non-ambulatory recordings. An algorithm developed by the authors for online epileptic EEG data reduction, described in Section II, is taken and its performance with both ambulatory and non-ambulatory recordings investigated. Methods are described in Section III with the analysis being broken into two parts: one to measure the number of events correctly recorded and one to quantify the amount of data reduction, with this approach allowing more data to be analysed. The results are presented in Section IV and discussed in Section V where it is found that AEEG data sets can be analysed without a degradation of performance. Indeed, despite the potential extra artifacts present in ambulatory records the algorithm performance is found to be superior in these data sets and this effect is commented upon.

The results presented here are of course specific to the algorithm considered, which is designed for the online data reduction of epileptic EEG traces, but AEEG is also of critical use in fields such as augmented cognition where the EEG is processed to determine the user's cognitive state. These applications will also require online signal processing and so the result that the automated analysis of AEEG traces is not significantly degraded is very significant for these

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ {acasson, e.rodriguez}@imperial.ac.uk.

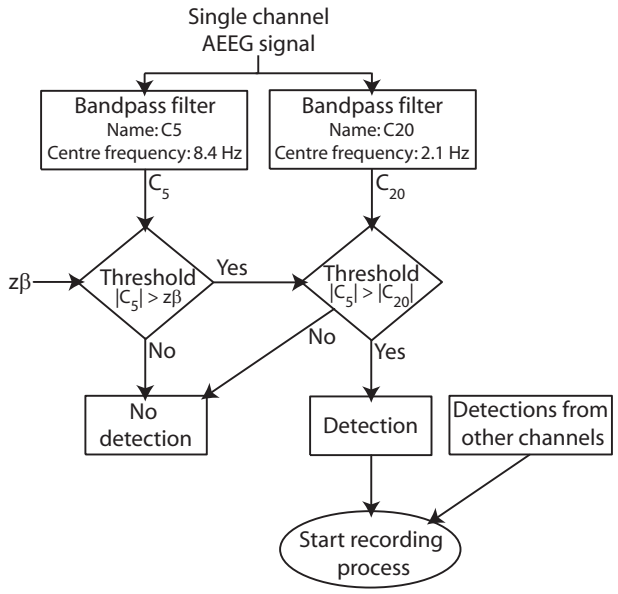

Fig. 1. A high level overview of the algorithm to be investigated. Detections in any monitored channel cause the algorithm to start recording.

fields as well.

II. ALGORITHM OVERVIEW

The algorithm to be investigated here is a developed version of the one proposed in [10]. A high level overview of the procedure is given in Fig. 1. The overall aim of the algorithm is to correctly record interictal epileptic events as identified by an expert marker whilst not recording background signals. This gives a significant data reduction, reducing the amount of data to be stored and hence reducing the system power consumption as this data storage is power intensive. In turn this offers greater operational lifetimes. As the algorithm aim is only data reduction, not event quantification, significant data reduction can still be achieved even with a number of false positives present [8].

The algorithm version considered here operates on 10 EEG channels in parallel with a detection in any one channel causing all of the channels to record. The 10 channels for analysis are selected purely as the channels common to all of the data sets available and are: F7, F8, Fp1, Fp2, O1, O2, T3, T4, T5, and T6.

The core of the processing is the extraction of frequency content in two bands by wavelet based, bandpass filtering. This information is then thresholded to determine whether a detection is made. The external detection threshold $z\beta$ is given by a user set parameter β and z which is an automatically generated normalizing parameter to correct for broad level amplitude differences in different EEG traces. The user is free to sweep β to obtain a range of performances from the algorithm.

III. METHODS

The comparison carried out here is broken down into two parts corresponding to comparing the two performance metrics of interest separately. These two metrics are the sensitivity—the percentage of expert marked events which are correctly recorded—and the percentage of the total data that is marked for recording, which is related to the data reduction. The data available for analysis, which is obtained from two testing sites, is summarised in Table I. Note that in general the total data for each patient may be made up of more than one non-continuous data set, but for the long term ambulatory recordings (patients 19–21) the records are continuous to reflect any potential electrode contact issues. The presence of any medication taken by the patients during testing is not normalized.

Section IV-A presents the sensitivity analysis using the data from patients 0–15. The aim of this is to show that expert marked events can be correctly detected in both ambulatory and non-ambulatory data. The analysis uses 173 events in approximately 26 hours of AEEG data and 726 events in 9 hours of non-AEEG data. The disparity in the amounts of data analysed is robustly compensated for by using a time/event weighted average to calculate the sensitivity [11]. This is calculated as: if there are M records and the i^{th} record has a duration T_i, N_i marked events and D_i correctly detected events the time/event weighted sensitivity is:

$$\text{Sensitivity} = \frac{1}{\sum_{i=1}^{M} \frac{T_i}{N_i}} \sum_{i=1}^{M} \frac{D_i}{N_i} \cdot \frac{T_i}{N_i}. \qquad (1)$$

This calculation method prevents short records or ones with large numbers of events from over-weighting the results, normalizing for the amount of data tested. A similar weighting is applied to the percentage of data transmitted figures.

The results are presented on a ROC curve to illustrate the trade-off between the sensitivity of the algorithm and the amount of data reduction that can be achieved as the β value is varied. This is required as although the aim of this analysis is only to compare the sensitivity of the algorithm on the two data types, of course any sensitivity can be achieved if no data reduction is required, and so this *cost* must also be illustrated. A *chance* performance line is also shown: if events are randomly distributed it may be expected as a first pass that sending 10% of the data will result in a 10% sensitivity, and so the algorithm performance should always be better than this line.

Nineteen thresholds equally spaced between $\beta^2 = \{0.1 - 1\}$ are used. Following each automated detection a section of data 1.25 seconds before and afterwards (2.5 seconds in total) is selected for recording with the remainder of the data being discarded. An event is deemed to be correctly recorded if there is a detection within one second of the marked event.

Section IV-B then presents an analysis of long term data sets for which only a total data reduction (the weighting factor given above is not applied) is calculated. In all other respects the setup is the same as above. The aim of this analysis is to illustrate on larger data sets, which more accurately reflect the long term use of the algorithm, that acceptable levels of data reduction can be achieved with the two data types. Sixty-six hours of ambulatory data is analysed with 68

TABLE I

THE DATA AVAILABLE FOR ANALYSIS. NOTE THAT THE DATA FOR EACH PATIENT MAY BE MADE UP OF MORE THAN ONE NON-CONTINUOUS TEST.

Patient	Age at test	Gender	Type of recording	Testing site[a]	Marked interictal events	Recording duration
0	Unknown	Unknown	Non-AEEG	NSE	644	00:36:55
1	24	Male	Non-AEEG	NSE	49	03:58:15
2	47	Female	AEEG	NSE	7	02:00:11
3	33	Female	AEEG	NSE	52	06:00:33
4	51	Unknown	AEEG	NSE	12	04:00:22
5	23	Female	AEEG	NSE	11	04:00:22
7	43	Male	Non-AEEG	NSE	2	00:18:30
8	46	Male	AEEG	NSE	26	02:00:11
9	45	Male	Non-AEEG	NSE	30	04:00:22
10	23	Female	AEEG	NSE	45	04:00:22
11	53	Female	AEEG	NSE	8	02:00:11
13	21	Female	AEEG	NSE	12	02:00:12
15	33	Unknown	Non-AEEG	NSE	1	00:10:53
16	33	Female	Non-AEEG	Freiburg	N/A	23:10:05
17	56	Female	Non-AEEG	Freiburg	N/A	34:18:41
18	41	Female	Non-AEEG	Freiburg	N/A	11:00:00
19	Unknown	Unknown	AEEG	NSE	N/A	22:57:05
20	Unknown	Unknown	AEEG	NSE	N/A	22:06:23
21	Unknown	Unknown	AEEG	NSE	N/A	21:05:43

[a]NSE is the National Society for Epilepsy in the UK and Freiburg is the Freiburg University Hospital in Germany.

hours of non-ambulatory data. It is known that some marked, interictal events are present in the analysed data but this is ignored and the data is analysed essentially prospectively: the data reduction is found regardless of what actually happens in the data set, exactly as if the algorithm was being used online. The data for patients 16, 17 and 18 was provided in approximately hour long epochs with some discontinuities present. Only continuous, seizure free epochs have been considered for analysis.

IV. RESULTS

A. Sensitivity comparison

The results from the sensitivity comparison are shown in Fig. 2. In general a high sensitivity with a small percentage of data being transmitted is wanted so lines should be as far to the top left of the graph as possible. It is clearly seen that better performance is obtained with the ambulatory data than with the non-ambulatory data, with (as may be expected) the analysis of all of the data lying between the two. At first this result seems counter-intuitive, better performance is expected with the non-ambulatory data, but a range of explanations are possible and these are discussed in Section V.

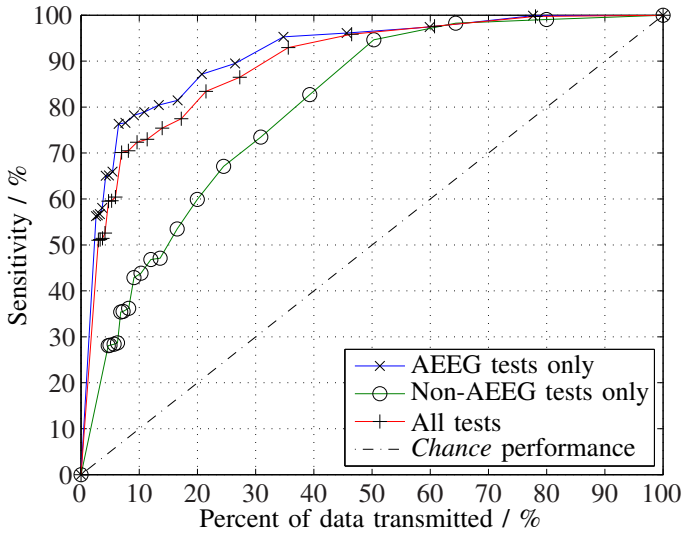

Fig. 2. Results for the sensitivity comparison show that a better sensitivity-data reduction trade-off is achieved when ambulatory data is analysed.

B. Data reduction comparison

Table II shows the amount of data that is transmitted when the long term data is analysed for several different β^2 threshold levels, although not all of the thresholds used in Fig. 2 are shown. To achieve a reasonable sensitivity-data reduction trade-off it is anticipated that values of β^2 between 0.2 and 0.3 would be most suitable for use. At these threshold levels it is again found that the ambulatory data achieves a better data reduction than the non-ambulatory data, although the non-ambulatory data does give better performance when high thresholds are used.

Two important conclusions can be drawn from this. Firstly, the level of data reduction that is achieved by the algorithm

TABLE II

RESULTS ILLUSTRATING THE DATA REDUCTION ACHIEVED WITH THE LONG TERM DATA SETS ANALYSED.

Threshold (β^2)	Percentage of total data transmitted / %	
	AEEG data: Patients 19–21	Non-AEEG data: Patients 16–18
0.10	66.3	71.1
0.15	48.9	55.6
0.20	37.1	43.1
0.25	29.1	33.3
0.30	23.3	25.9
0.35	19.2	20.4
0.40	16.2	16.3
0.45	13.8	13.2
0.50	12.0	10.7

is maintained across testing sites indicating that different test procedures and equipment do not unduly affect the results. Secondly, the ambulatory data again achieves *better performance* than the non-ambulatory data which may not be initially expected. This is discussed in Section V.

V. DISCUSSION

The results from the previous section clearly indicate that the algorithm can achieve good results for both ambulatory and non-ambulatory recorded data: overall approximately 90% of marked events are correctly recorded while transmitting only 30% of the total data. This performance level is confirmed with similar data reductions being achieved in the long term analysis. It is also seen that in most cases a better data reduction is achieved when ambulatory data is used, a potentially unexpected result.

There are several potential explanations, however. Firstly, and it cannot be ruled out entirely, the algorithm is currently implemented in MATLAB code and a simple coding error may be present. However, the algorithm has been checked at multiple levels of abstraction and as the result appears in data from two testing sites, and is essentially independent of which data is analysed, it is unlikely that a coding error would result in such a systematic error in the results. Similarly the amount of data tested is not identical between the two cases but this has been corrected for by the weighting applied, as have systematic differences in the recording process (different equipment is necessarily used for the two types of recording) by analysing data from two testing sites. Different markers may interpret the EEG differently, potentially affecting the sensitivity results, but this would not affect the data reduction one. The result is thus deemed to be a real effect, rather than an artifact of a non-ideal testing procedure.

The explanation proposed here is quite simply that as expected ambulatory recordings have lots of obvious artifacts. However, these obvious *bad recording periods* are readily identified by the algorithm and so not recorded, giving a large data reduction. In turn this readily leads to the improved performance of the ambulatory data over the non-ambulatory data. This is a good result for the data reduction algorithm proposed by the authors: good data reduction is achieved in all cases, and is better in the ambulatory case where the algorithm is intended to be used. However, it does raise questions as to whether interictal events can be correctly identified if they were to occur in a *bad recording period* which relates to the overall clinical utility of AEEG monitoring. At this point, however, this is not in doubt: AEEG has been shown to be clinically useful in 75% of patients tested with it, abnormalities are found in 12–25% of patients where the initial EEG was normal or non-diagnostic and this costs just 50% of traditional inpatient monitoring [5].

From these results two important conclusions can be drawn. Firstly, it is possible to achieve good sensitivity-data reduction trade-offs when using ambulatory data. Any comparatively *poor quality* of the recordings due to the presence of artifacts or electrode connection issues does not affect the ability to correctly identify candidate events. Indeed, in some respects this is beneficial as these artifacts are easily detected and removed so that less data needs to be saved, reducing power consumption. The second conclusion is that it is valid to test the algorithm on non-AEEG data as well as AEEG data, greatly increasing the amount of data that is available. If non-ambulatory data is analysed the actual results on ambulatory data should only be better than the ones found. Although these conclusions are inevitably application specific, it is a good result for further demonstrating the feasibility of the online signal processing of the AEEG which may be of use in many other applications.

VI. CONCLUSION

This paper has investigated the performance of an EEG signal processing algorithm on ambulatory and non-ambulatory recorded data. In general ambulatory recordings are less controlled than non-ambulatory ones with more artifacts expected to be present, and this may affect the signal processing performance. It is found, however, that superior performance is achieved on the ambulatory recordings, potentially as any additional artifacts are easily identified. This indicates that the automated analysis of ambulatory EEGs, possibly online in the AEEG device, is feasible without a loss in performance.

VII. ACKNOWLEDGEMENTS

The authors gratefully acknowledge and thank the National Society for Epilepsy in the UK and the University of Freiburg in Germany without whom the analysis presented here would not have been possible.

REFERENCES

[1] C. D. Binnie, A. J. Rowan, and T. Gutter, *A manual of electroencephalographic technology*. Cambridge: Cambridge University Press, 1982.

[2] P. E. M. Smith and S. J. Wallace, *Clinicians' guide to epilepsy*. London: Arnold, 2001.

[3] S. B. Wilson and R. Emerson, "Spike detection: a review and comparison of algorithms," *Clin. Neurophysiol.*, vol. 113, no. 12, pp. 1873–1881, 2002.

[4] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalogr. and Clin. Neurophysiol.*, vol. 54, no. 5, pp. 530–540, 1982.

[5] E. Waterhouse, "New horizons in ambulatory electroencephalography," *IEEE Eng. Med. Biol. Mag.*, vol. 22, no. 3, pp. 74–80, 2003.

[6] J. S. Ebersole and S. L. Bridgers, "Direct comparison of 3- and 8-channel ambulatory cassette EEG with intensive inpatient monitoring," *Neurology*, vol. 35, no. 6, pp. 846–854, 1985.

[7] D. C. Yates and E. Rodriguez-Villegas, "A key power trade-off in wireless EEG headset design," in *IEEE EMBS NER*, Hawaii, May 2007.

[8] A. J. Casson and E. Rodriguez-Villegas, "Data reduction techniques to facilitate wireless and long term AEEG epilepsy monitoring," in *IEEE EMBS NER*, Hawaii, May 2007.

[9] A. Shoeb, J. Guttag, S. Schachter, D. Schomer, B. Bourgeois, and S. T. Treves, "Detecting seizure onset in the ambulatory setting: Demonstrating feasibility," in *IEEE EMBC*, Shanghai, September 2005.

[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 *IEEE EMBC*, Lyon, August 2007.

[11] A. J. Casson, E. Luna, and E. Rodriguez-Villegas, "Performance metrics for the accurate characterisation of interictal spike detection algorithms," in preparation.