### **AUTOMATIC SEIZURE DETECTION USING SUPERVECTORS**

*M. Golmohammadi, S. Lopez, A. Gross, M. G. Andrade, V. Shah, I. Obeid and J. Picone*

Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA

{meysam, silvia.lopez, aaron.gross, guedesm, vinit.shah, iobeid, picone}@temple.edu

**Abstract**

Manual analysis of EEG signals is a tedious, time-consuming task that requires a highly trained neurologist. Interrater agreement amongst experts is low, and automatic processing techniques suffer from high false alarm rates. It is a challenging task because the events of interest occur infrequently and often occupy less than 0.1% of the duration of the signal. In this paper, we present an automatic seizure detection algorithm based on hidden Markov models that uses a supervector approach to identify and localize seizures. Principal component analysis is used for dimensionality reduction. Results are presented on the TUH EEG Corpus ­– the largest publicly available corpus of clinical EEG recordings. The proposed supervector model produces a 10% relative reduction in specificity compared to our baseline system that uses channel­-independent features. The proposed algorithm is computationally inexpensive, making it suitable for real-time critical care applications such as Intensive Care and Epilepsy Monitoring Units.

*Index Terms*— hidden Markov models, principal component analysis, electroencephalography, seizure detection

# Introduction

Electroencephalograms (EEGs) are used in a wide range of clinical settings to record electrical activity along the scalp. EEGs are the primary means by which physicians diagnose brain-related illnesses such as epilepsy and seizures [1]. However, detection of seizure events in a clinical setting requires a highly trained, board-certified neurologist. It is a time-consuming and expensive process since identifying rare clinical events requires analysis of long data streams. EEG event detection poses an interesting signal processing problem for several reasons including: (1) the duty cycle of the signal of interest is extremely low – events of interest occupy less than 0.1% of the duration of the signal, which means the prior probability of an event is very low; and (2) interrater agreement between experts is low [2][3] – epileptiform activities can often be considered ictal, inter-ictal or postictal depending on a clinician's experience [4]; (3) patients are often sedated with anesthetics which tend to reduce EEG amplitude and frequency, increasing confusion amongst expert readers; and (4) around 19% of patients have electrographic seizures without any clinical signs [4].

Automatic seizure detection can reduce the time to diagnosis, reduce errors and enhance real-time applications by auto scanning EEG signals and flagging sections of the signal that need further review by a clinician. Though many algorithms have been applied to this problem over the years [5]­­-[7], commercial systems perform poorly in critical care settings [8]. These systems suffer from a very high false alarm rate, which in turn overwhelms healthcare workers who must service these alarms. Therefore, most alerts from these types of systems are ignored in clinical settings. The goal of our work is to achieve a false alarm rate of 1 per 8 hours per patient. Current systems operate at rates at least two orders of magnitude higher than this. Hence, there is a clearly defined technology need and motivation to evaluate new approaches to the problem.

Machine learning has made tremendous progress over the past three decades due to rapid advances in low-cost highly-parallel computational infrastructure, powerful machine learning algorithms, and, most importantly, big data. Although contemporary approaches for automatic seizure detection have employed more modern machine learning approaches such as neural networks and support vector machines [9], state of the art machine learning algorithms [10][11] that employ high dimensional models have not previously been utilized in EEG analysis because there has been a lack of sufficiently large databases to support training of these models. These models incorporate a large number of parameters and sophisticated training techniques that require orders of magnitude more data than is typically found in publicly available databases [12][13]. Further, most available databases do not accurately characterize clinical settings where there are a host of practical problems that often result in false alarms (e.g., electrode connectivity, patient movement).

Recently, a significant big data resource, known as the TUH EEG Corpus [14], has become available, offering a unique opportunity for machine learning research. This data collection project is an ongoing effort and currently includes over 30,000 EEG sessions from over 16,000 patients. It spans 14 years and includes virtually all EEGs administered at Temple University Hospital. The database includes detailed physician reports and patient medical histories, which are critical to the application of deep learning. In this study we are using a subset of the data, described in Section 5, that has been manually reviewed for seizure events.

# Applications of Hidden Markov Models

An overview of our baseline system architecture for automatic EEG interpretation is shown in Fig. 1. This architecture is based on a low-level event detection system that uses hidden Markov models (HMMs), and borrows heavily from our previous work in speech recognition [15][16]. There are four major features of the system: (1) feature extraction that integrates cepstral and energy measures that are enhanced by differential energy measures; (2) epoch-level hidden Markov model technology to achieve high performance classification [17]; (3) a self-training process to ingest and classify (data wrangling) large amounts of EEG data [18]; and (4) two levels of postprocessing that integrate spatial and temporal context.

This system identifies six types of events: (1) spike and sharp wave (SPSW), (2) generalized periodic epileptiform discharge and triphasic waves (GPED), (3) periodic lateralized epileptiform discharge (PLEDs), (4) eye movement (EYEM), (5) general artifact (ARTF) and (6) background activity (BCKG). The first three events relate to signal events of interest while the last three are used to improve classification of the background signal. This system delivers 90% detection accuracy and a 5% false alarm rate on clinical data. Our focus in this study is to adapt this system to the task of seizure detection for critical-care applications in Intensive Care Units (ICU) and Epilepsy Monitoring Units (EMU).

The first pass of processing performs event detection using a feature extraction approach [15] that parallels the well-known cepstral filter bank approach used in speech recognition [19]. One small variation of note is that we use a linear frequency scape rather than the mel scale. We refer to this system as channel independent since our HMMs are not trained on specific EEG channels. The same model is used across all channels. Our experiments with other popular approaches to feature extraction, most notably wavelets, have shown very little advantage over MFCCs on the TUH-EEG Corpus.

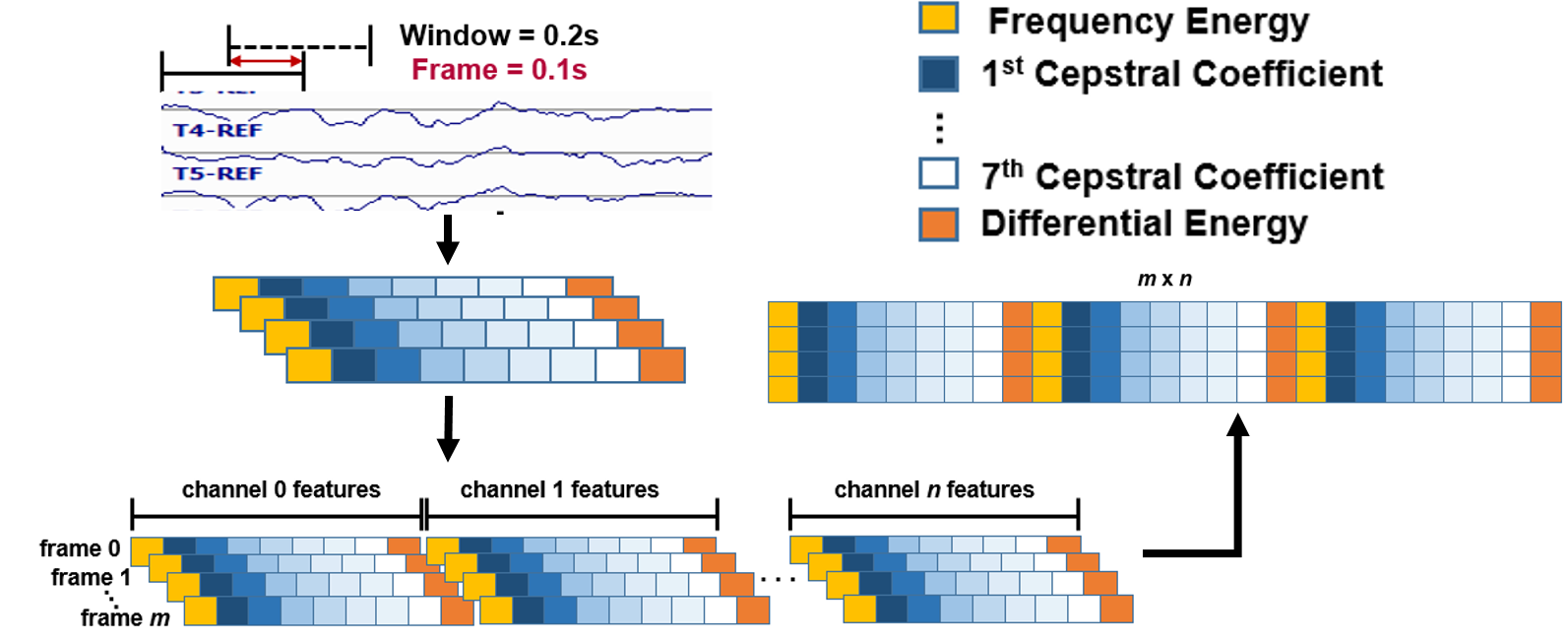


Fig. 2. Supervector feature stacking

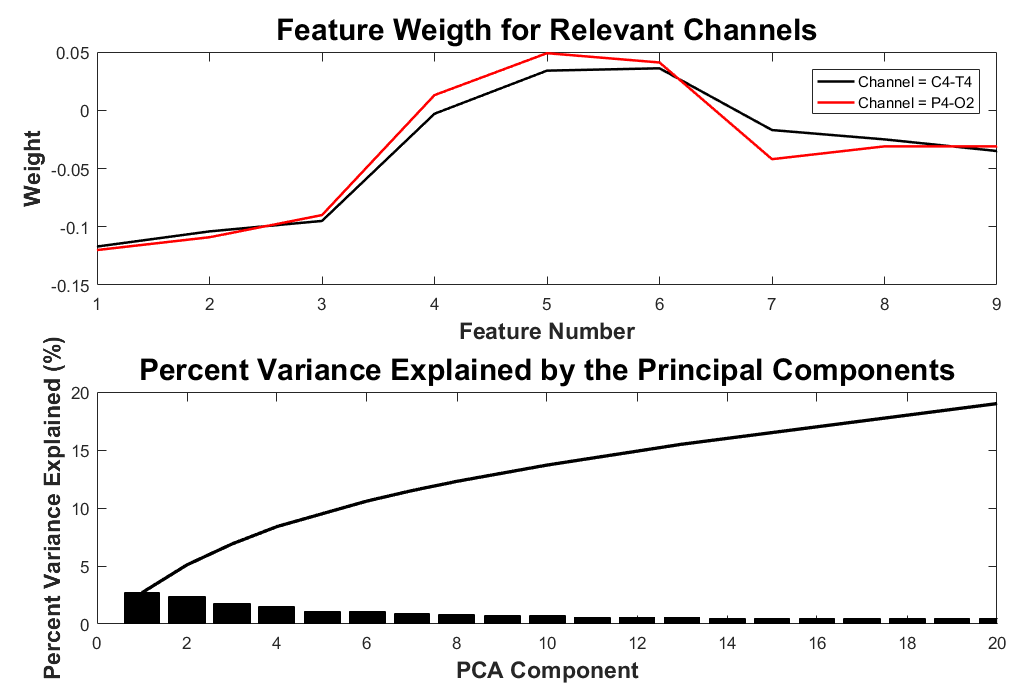


Fig. 3. Feature weights are shown for channels C4-T4 and P4-O2 (top) along with the percent of the variance explained by the first 20 principal components.

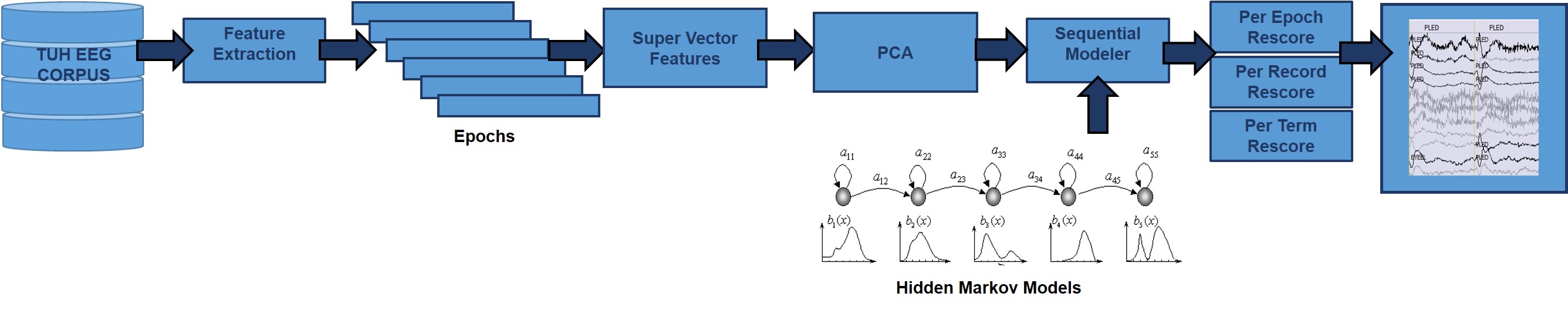


Fig. 1. An overview of a seizure detection system based on hidden Markov models and supervector features.

Since neurologists typically review EEGs on a standard PC monitor using a 10-second per view window, which often translates to a temporal resolution of 1 second for annotations, we designed our pattern recognition system to classify data on 1-second epochs. We further subdivide these 1 sec epochs into frames that are 0.1 secs in duration, so that we can model each epoch as a sequence of 10 frames. In a previous study, we have systematically calibrated the feature extraction parameters and we showed that the use of a novel differential energy feature improved performance for absolute features, but that benefit diminishes as first and second order derivatives are included [15]. We have shown there is benefit to using derivatives and there is a small advantage to using frequency domain energy. The output of the feature extraction component of the system is a feature vector of length 26 generated 10 times a second. There are 22 channels of data in a standard 10/20 EEG [1], which means the overall dimension of this feature vector is 22x26 = 572.

Our channel-independent baseline configuration classifies a group of frames as an event on a per-channel basis using an HMM based classifier. We use standard three-state HMM models that employ a left-to-right topology [16] to encode the temporal evolution of the signal. In this study, a binary classification is used in which there are two models: seizure and non-seizure. The system was implemented using HTK [20]. An extension was developed that calculates the likelihood of the data given a model for a defined segment of the signal. Labels are output for each epoch for each channel, and then postprocessed to produce a single decision based on the individual channel outputs. Discriminative training was not used though preliminary experiments have shown only marginal gains for using such approaches.

# SUPERVECTORS

Some seizure events, known as focal, tend to appear on a specific set of channels. Other events, known as general, appear across all channels. There are systematic differences in the nature of the signal for seizure and non-seizure events. Therefore, to obtain a baseline level of performance, we explored a simple Principal Components Analysis (PCA) [21] approach to modeling and analyzing the stacked feature vectors produced by the appoach shown in Fig. 2. In this study, we explored only stacking across channels, though in related work, we are developing deep learning techniques for stacking in both space and time. The goal of this study was to determine if spatial correlations would impact seizure detection performance. The supervector was then represented in a lower dimensional space through the implementation of PCA.

The weight of the channel independent feature vectors was analyzed and plotted for two relevant channels: C4-T4 and P4-O2. These two channels were selected due to the fact that they represent the channels with the most (C4-T4) and the least (P4-O2) seizure observations in our dataset. The weights for the eigenvector corresponding to the largest eigenvalue are shown in the top of Fig. 3. The variance explained by the first 20 principal components is also shown in the bottom of Fig. 3.

Note that in this picture the weights of the feature vectors are higher in the 4th and 5th cepstral coefficients, which correspond to the mid-frequency range alpha and beta waves that are commonly found in awake adult EEGs (~12Hz-16Hz). It is also evident that the weights of the features are similar across channels, emphasizing the mid-frequency range and attenuating the energy term and lower frequencies. These eigenvectors are clearly accounting for spectral variations between the signal events of interest and the background channel, and are also creating a focus on the frequency range most critical to seizure detection. Neurologists routinely consider alpha and beta wave behavior in assessing a seizure, and PCA is attempting to learn something similar.

The first 20 principal components, plotted in the bottom plot of Fig. 3, explain 19% of the variance present in the data. This is relatively small and indicates that no single subset of the input feature vector explains a large percentage of the variance. In fact, the amount of additional variance explained by each feature above rank 20 is fairly small, with approximately 95% of the variance being explained by the first 300 eigenvalues. Hence, it is clear that the discriminating power of these features can be improved.

# Evaluation MetrICS

We characterized the performance of our seizure detector in terms of sensitivity and specificity using three evaluation methodologies: per epoch, per record and per term. Sensitivity refers to the percentage of test seizures correctly identified as seizures. Specificity refers to the percentage of declared seizure activity in the absence of an actual clinical seizure. In addition to the core evaluation procedure, we also present results using a Detection Error Tradeoff (DET) curve. Since performance evaluation is not consistent throughout the research literature, we present results using three common evaluation metrics that have appeared in the literature.

**4.1. Per Epoch Evaluation Methodology**

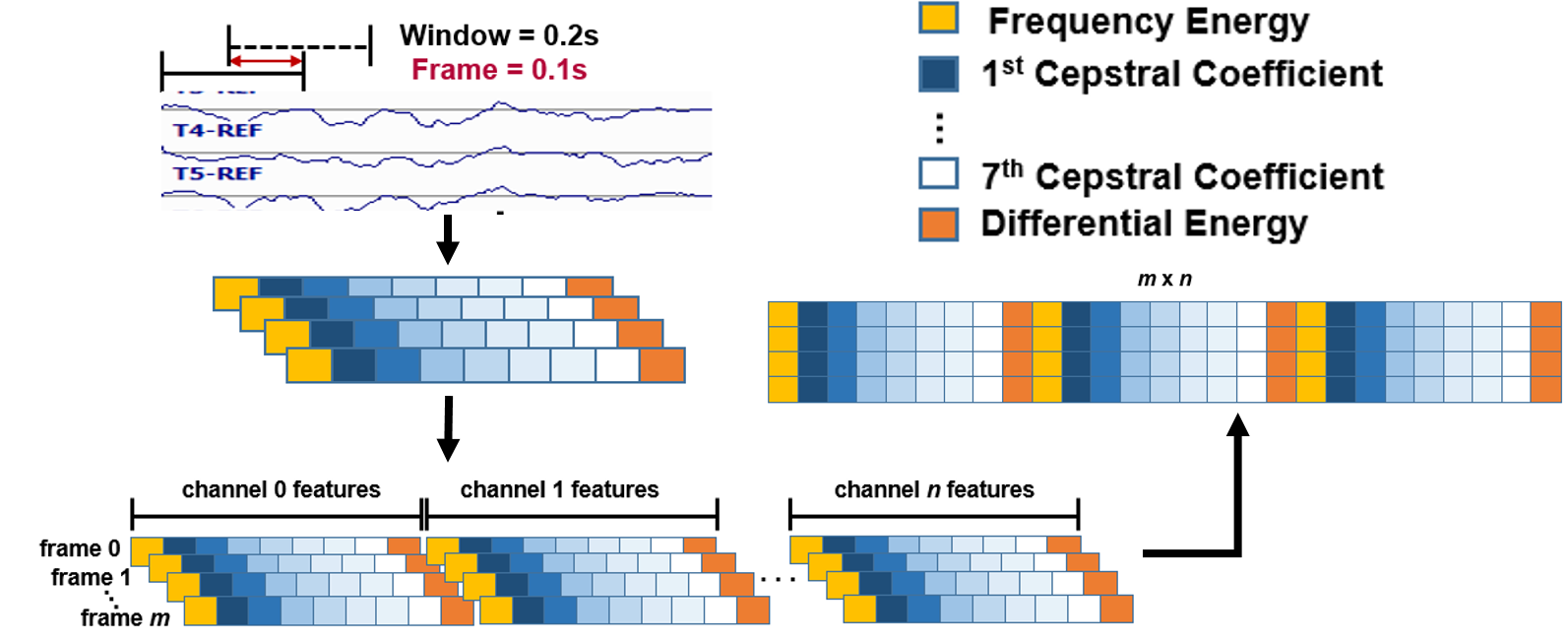


Fig. 2. Supervector feature stacking

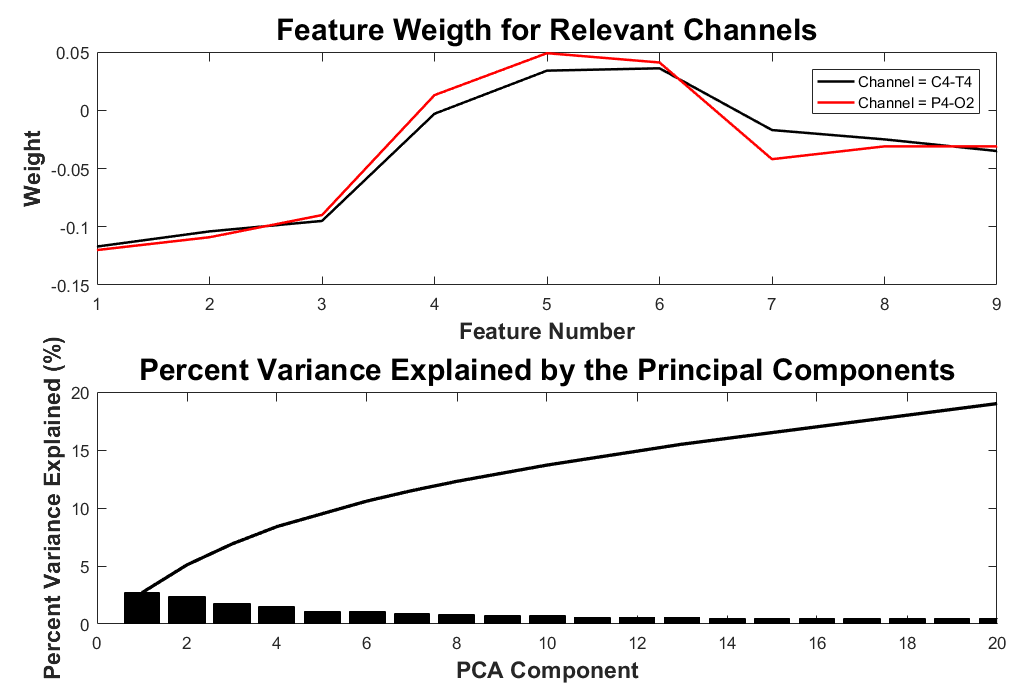


Fig. 3. Feature weights are shown for channels C4-T4 and P4-O2 (top) along with the percent of the variance explained by the first 20 principal components.

Our initial approach to evaluating seizure detection was based on signal epochs. Evaluation is channel specific meaning that each channel was assessed and included in the overall performance. Correct and spurious detections are verified by finding matching hypothesis and reference epochs and comparing their class labels. The raw comparison is a simple evaluation method without any post-processing steps. The epoch duration is defined as 1 second in this study. The advantage of this approach is that it evaluates a system’s ability to assess the locality of a seizure.

**4.2. Per Record Evaluation Methodology**

An option of performing classification, and hence evaluation, over the raw output data is to collapse spatial (i.e. channel) information. Based on this idea, an evaluation approach of records was implemented. A single judgement was made for each epoch that applied to all channels. To collapse channels, label files for the input data are processed to represent information in a single time axis. The idea is to make classification space independent, classifying a time interval of the whole EEG recording into seizure or background. Since the data is labeled for supervised learning, collapsing reference information is a simple task. It is defined a priority value for classes to decide how a time interval must be classified. Intervals of the same recording are evaluated in different channels through a priority queue. The priority values are specified as parameters, allowing maintainability. For seizure detection, the seizure class has the highest priority.

The hypothesis case is analogous, although an averaging process over channels is introduced to deal with likelihoods. In addition, the process accepts a time smoothing operation using neighboring time events. The procedure can be summarized as the application of a simple box filter over the system output hypothesis files. The data for a recording can be visualized as a rectangular matrix, with height equals to the number of channels and width equals to the number of epochs. The box filter has the same height as the data matrix itself and a variable width value accepted as a parameter. The output hypothesis vector entries follows this formula:

Where n is the number of channels in the EEG data, and W is the number of neighboring epochs to be used. For this paper, n=22 and after calibrating results we selected W=3.

With reference time delimitation, hypothesis epochs correspondent to a reference record are averaged. The class likelihoods for background and seizure are compared and a class label is stipulated for the hypothesis record.

Since spatial information is collapsed and records are in a greater scale than epochs, the cost of strictly evaluation (i.e. counting spurious and correct detected events) is significantly reduced. In the other hand, the process introduces post-processing steps that increases the overall cost when compared to the epoch-based methodology.

**4.3. Per Term Evaluation Methodology**

Opposing to per epoch and per record performance metrics, in per term evaluation methodology, terms are defined as maximized single-class time intervals of signal. In this basis, spatial information is abstracted and terms can be computed either for collapsed or expanded channel information data. The goal is to achieve a general methodology for seizure detection, taking into account peculiarities of the problem domain.

The procedure begins with term generation for reference data. A merging algorithm joins together consecutive time intervals of signal labeled with the same class. Reference terms are used to generate hypothesis terms, giving time intervals to delimit macro terms. Hypothesis sub terms are computed through the same merge algorithm used for reference, although limited by macro term duration. In this case, the hypothesis term data structure has one more dimension than the reference one, as if each reference term was expanded into hypothesis sub terms.

To be independent of spatial information, the merging algorithm accepts lists of 3-element (i.e. start time, stop time and class label) data events. The list is parsed grouping consecutive events of the same class. Terms related to each group are added to the returning term list by selecting the start time of the earliest event in the group, the stop time of the latest and the group class.

Correct and spurious detected terms are counted based on the reference term class using two rules. The rule for seizure uses a threshold ratio to define a correct match of a hypothesis term. If the summed duration of hypothesis seizure sub terms of a reference seizure term is greater than the threshold ratio times the reference duration, a correct detection is accounted. For background terms, a threshold duration of seizure is used instead. A background term is accepted as a correct detection if and only if there are no seizure sub terms longer than the specified threshold duration.

In summary, in per term evaluation methodology, basic detection performance will be characterized via standard detection error tradeoff (DET) curves of sensitivity versus specificity. Sensitivity and specificity are functions of the detection threshold, theta, and will be computed separately for each search term using this formula:

Where N\_correct(term, theta) is the number of correct (true) detections of term with a score greater than or equal to theta, N\_spurious(term, theta) is the number of spurious (incorrect) detections of term with a score greater than or equal to theta, N\_true(term) is the true number of occurrences of term in the corpus, and N\_nt(term) is the number of opportunities for incorrect detection of term in the corpus.

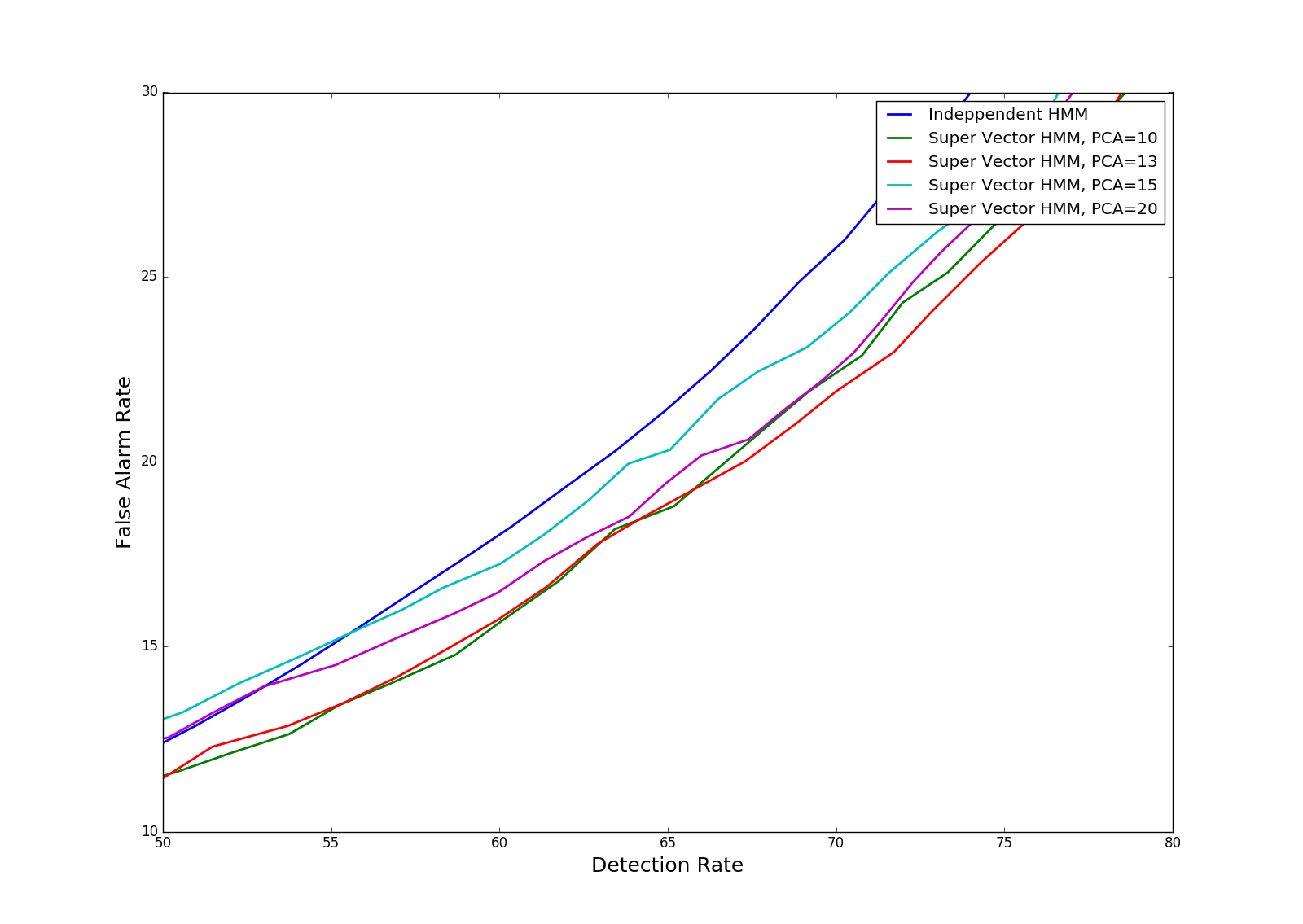


Fig. 4. Performance as a function of supervector dimension

# Experiments

In this section, some experimental results will be presented. The TUH-EEG corpus has been used. The dataset in TUH-EEG mostly have two different montages: REF and LE. In this paper, just those sessions that have REF montages are used for training and evaluation. A major goal of this study was to generate a large annotated corpus of seizure events that can support machine learning experiments. For this goal, we implemented a semi-automated strategy to label seizures. In the first step, TUH-EEG reports are analyzed using natural language processing techniques like NegEx [20] to locate sessions most likely to contain seizures. Then these sessions manually reviewed by an expert. The statistics of session and patient information of TUH\_EEG database annotated with seizures for this study is presented in Table 1.

The process of annotating based on NLP resulted in 346028 seconds of seizure data and 371423 of data that do not include seizures which simply referred as background in this paper. The statistics for total, training and evaluation dataset are represented in Table 2.

The first series of experiments were run on this dataset using independent channel feature based on HMM to calibrate the parameters. Then the second set of experiments were run on the same dataset using super vector features based on HMM to evaluate this approach against independent channel feature method.

Table 1. An overview of the seizure detection subset

|  |  |  |  |
| --- | --- | --- | --- |
| **Description** | **Files** | **Sessions** | **Patients** |
| TUH\_EEG (v0.6.0) | 38,408 | 16,982 | 10,868 |
| NLP Pool | 2,969 | 946 | 669 |
| Manually Verified | 270 | 174 | 118 |
| Seizures with REF Montage | 216 | 125 | 69 |
| Seizures with LE Montage | 54 | 49 | 49 |

Table 2. Total duration of the data in seconds

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Train** | **Evaluation** | **Total** |
| Seizure | 292,689 | 53,339 | 346,028 |
| Background | 308,240 | 63,182 | 371,423 |

Table 3. Seizure detection results for the TUH EEG subset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **System** | **Per Epoch** | | **Per Record** | | **Per Term** | |
| **Sens.** | **Spec.** | **Sens.** | **Spec.** | **Sens.** | **Spec.** |
| Indep. | 78.3% | 35.1% | 94.3% | 25.0% | 94.3% | 15.4% |
| SuperV | 72.8% | 24.1% | 86.4% | 13.5% | 86.4% | 23.1% |

A DET curve for calibrating of PCA size for super vector feature experiments is presented in Fig. 4. Four different PCA size used in these experiments. The results of these experiments can be compared with independent channel approach. As you can see all the experiments deliver a better performance in comparison with independent channel approach. The best result for super vector feature can be achieved by PCA size of 13.

Additionally the results of the seizure classification task on TUH-EEG for both of the independent channel feature and super vector feature based on HMM are presented in Table 3. The results for super vector in this table is related to PCA size of 13. As you can see super vector approach drops specificity by 10% based on per epoch evaluation methodology. The improvement for per record and per term methodology is about 2%.

# Conclusions

In this paper we have introduced a new approach for seizure classification based on HMMs that uses a supervector feature vector. We demonstrated that this approach improved performance by decreasing the error rate associated with sensitivity by 10% relative. In addition, this approach gives us the ability to localize seizures by channel and significantly decreases the computational requirements.

Our main challenge now is significantly reducing the false alarm rate. We are currently exploring deep learning approaches to doing this. Integrating spatial and temporal context and improving discrimination between background and signal events will be critical to achieving substantial reductions. Some form of real-time adaptive patient normalization might also be required.

# AcknowledgmentS

This material is based in part upon work supported by the   
National Science Foundation under Grant No. IIP-1622765. Any opinions, findings, and conclusions or recommendations   
expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Research reported in this publication was also supported by the National Institutes of Health under Award Number U01HG008468. The TUH EEG Corpus work was funded by (1) the Defense Advanced Research Projects Agency (DARPA) MTO under the auspices of Dr. Doug Weber through the Contract No. D13AP00065, (2) Temple University’s College of Engineering and (3) Temple University’s Office of the Senior Vice-Provost for Research.

# References

1. T. Yamada and E. Meng, Practical Guide for Clinical Neurophysiologic Testing: EEG. Philadelphia, Pennsylvania, USA: Lippincott Williams & Wilkins, 2009.
2. J. J. Halford, D. Shiau, J. A. Desrochers, B. J. Kolls, B. C. Dean, C. G. Waters, N. J. Azar, K. F. Haas, E. Kutluay, G. U. Martz, S. R. Sinha, R. T. Kern, K. M. Kelly, J. C. Sackellares, and S. M. LaRoche, “Inter-rater agreement on identification of electrographic seizures and periodic discharges in ICU EEG recordings,” Clin. Neurophysiol., vol. 126, no. 9, pp. 1661–1669, 2015.
3. H. E. Ronner, S. C. Ponten, C. J. Stam, and B. M. J. Uitdehaag, “Inter-observer variability of the EEG diagnosis of seizures in comatose patients,” Seizure, vol. 18, no. 4, pp. 257–263, 2009.
4. J. Claassen, S. a Mayer, R. G. Kowalski, R. G. Emerson, and L. J. Hirsch, “Detection of electrographic seizures with continuous EEG monitoring in critically ill patients.,” Neurology, vol. 62, no. 10, pp. 1743–1748, 2004.
5. J. Gotman, “Automatic recognition of epileptic seizures in the EEG,” Electroencephalogr. Clin. Neurophysiol., vol. 54, no. 5, pp. 530–540, Nov. 1982.
6. F. Sartoretto and M. Ermani, “Automatic detection of epileptiform activity by single-level wavelet analysis,” Clin. Neurophysiol., vol. 110, no. 2, pp. 239–249, 1999.
7. K. Gadhoumi, J. M. Lina, F. Mormann, and J. Gotman, “Seizure prediction for therapeutic devices: A review,” Journal of Neuroscience Methods, vol. 260. Elsevier, pp. 270–282, 2016.
8. P. Hu, “Reducing False Alarms in Critical Care,” in Working Group on Neurocritical Care Informatics, Neurocritical Care Society Annual Meeting, 2015.S. Ramgopal, “Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy,” Epilepsy Behav., vol. 37, pp. 291–307, Jan. 8AD.
9. T. Alotaiby, S. Alshebeili, T. Alshawi, I. Ahmad, and F. Abd El-Samie, “EEG seizure detection and prediction algorithms: a survey,” EURASIP J. Adv. Signal Process., vol. 2014, no. 1, pp. 1–21, 2014.
10. G. Saon, H.-K. Kuo, S. Rennie, and M. Picheny, “The IBM 2015 English Conversational Telephone Speech Recognition System,” in Proceedings of INTERSPEECH, 2015, pp. 1–4.
11. C. Farabet, C. Couprie, L. Najman, and Y. Lecun, “Learning hierarchical features for scene labeling,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1915–1929, 2013.
12. A. Shoeb and J. Guttag, “Application of Machine Learning To Epileptic Seizure Detection,” Proc. 27th Int. Conf. Mach. Learn., pp. 975–982, 2010.
13. D. Wulsin, J. Blanco, R. Mani, and B. Litt, “Semi-Supervised Anomaly Detection for EEG Waveforms Using Deep Belief Nets,” in International Conference on Machine Learning and Applications (ICMLA), 2010, pp. 436–441.I. Obeid and J. Picone, “The Temple University Hospital EEG Data Corpus,” Front. Neurosci. Sect. Neural Technol., vol. 10, p. 00196, 2016.
14. I. Obeid and J. Picone, “The Temple University Hospital EEG Data Corpus,” Front. Neurosci. Sect. Neural Technol., vol. 10, p. 00196, 2016.
15. A. Harati, M. Golmohammadi, S. Lopez, I. Obeid, and J. Picone, “Improved EEG Event Classification Using Differential Energy,” in Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium, 2015, pp. 1–4.
16. J. Picone, “Continuous Speech Recognition Using Hidden Markov Models,” IEEE ASSP Magazine, vol. 7, no. 3, pp. 26–41, Jul. 1990.
17. B.-H. Juang and L. Rabiner, “Hidden Markov Models for Speech Recognition,” Technometrics, vol. 33, no. 3, pp. 251–272, 1991.
18. Yang...
19. J. Picone, “Signal modeling techniques in speech recognition,” Proc. IEEE, vol. 81, no. 9, pp. 1215–1247, 1993.
20. S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollagson, D. Povey, V. Valtchev, and P. Woodland, “The HTK Book,” Cambridge, UK, 2006.
21. R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification, 2nd ed. New York City, New York, USA: John Wiley & Sons, 2003.
22. L.. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.
23. S. Baldassano, D. Wulsin, H. Ung, T. Blevins, M.-G. Brown, E. Fox, and B. Litt, “A novel seizure detection algorithm informed by hidden Markov model event states.,” J. Neural Eng., vol. 13, no. 3, p. 036011, 2016.
24. W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan, “A simple algorithm for identifying negated findings and diseases in discharge summaries.,” J. Biomed. Inform., vol. 34, no. 5, pp. 301–10, 2001.
25. P. H. Da Rocha Garrit, A. Guimaraes Moura, I. Obeid, and J. Picone, “Wavelet Analysis for Feature Extraction on EEG Signals,” in NEDC Summer Research Experience for Undergraduates, Department of Electrical and Computer Engineering, Temple University, 2015, p. 1.
26. A. Harati and J. Picone, “A Doubly Hierarchical Dirichlet Process Hidden Markov Model with a Non-Ergodic Structure,” IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 24, no. 1, pp. 174–184, 2016.
27. S. Lu and J. Picone, “Fingerspelling Gesture Recognition Using A Two-Level Hidden Markov Model,” in Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICPV), 2013, pp. 538–543.