Event Detection in Electrocorticogram Using Methods Based on a Two-Covariance Model

Jane E. Huggins, *Member, IEEE,* Victor Solo, *Fellow, IEEE,* Se Young Chun, *Student Member, IEEE,* Jeffrey A. Fessler, *Fellow, IEEE,* Chandan Damannagari, Simon P. Levine, *Member, IEEE,* Bernhard Graimann

Abstract

The University of Michigan Direct Brain Interface (UM-DBI) project seeks to detect voluntarily produced electrocortical activity (ECoG) related to actual or imagined movements in humans as the basis for a DBI. In past work we have used cross-correlation based template matching (CCTM) to detect event-related potentials (ERPs). This paper focuses on signal detection methods that exploit event-related changes in ECoG *power spectra*, signal characteristics that are ignored by the CCTM approach. We propose two new detectors based on covariance (autoregressive) models. The first, which we call the quadratic detector, is a classic fixed interval or window based likelihood ratio detector. The second, which we call the change-point detector, is a more sophisticated version of the quadratic detector that estimates the event time. Model order was chosen by applying the BIC criterion. Comparison of detection accuracy and response time for the CCTM, a basic feature-based band-power (BP) method, the quadratic detector and the change-point detector showed that the CCTM has the poorest performance, the quadratic detector and the BP have similar performance, and the change-point outperforms all three.

I. Introduction

A direct brain interface is an interface that accepts voluntary commands directly from the human brain (without requiring physical movement) to operate a computer or other assistive technology. The University of Michigan Direct Brain Interface (UM-DBI) project has focused on the detection of voluntarily produced event-related changes in electrocorticogram (ECoG) from subdural implanted electrodes. The short-term goal is an interface that is capable of operating single-switch assistive technologies. Longer-term goals are to increase DBI accuracy and the number of useable control channels to allow operation of more complex tasks.

In past work we used the cross-correlation based template matching (CCTM) method to detect event-related potentials (ERPs) [1]–[3]. However CCTM ignores event-related changes in the signal power spectrum, such as event-related desynchronization

Supported by Bioengineering Research Partnership RO1-EB002093 from the National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health (NIH), USA.

J.E. Huggins and S.P. Levine are with the Rehabilitation Engineering Program, Department of Physical Medicine and Rehabilitation, University of Mighican, Ann Arbor, MI 48109 USA and also with the Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109 USA.

V.Solo is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA and also with the School of Electrical Engineering, University of New South Wales, Kensington 2052, Sydney, AUSTRALIA.

S.Y. Chun and J.A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA.

C. Damannagari is with the Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109 USA.

B. Graimann is with the Institute of Automation, University of Bremen, GERMANY.

(ERD) and event-related synchronization (ERS) [4]. Power spectrum changes have been the basis for detection methods with electroencephalogram (EEG) (*e.g.*, [4]–[6]) and ECoG, (*e.g.*, [7]–[9]).

Most detection methods that exploit spectral changes have been designed using a *feature-based* strategy, which identifies one or more relevant spectral features and then applies a feature-based classifier such as linear discriminant analysis (LDA) or a neural network, *e.g.*, [4], [6], [8]. One such method, the band power (BP) method [4], which extracts the signal power in certain subject-dependent spectral bands is described in Section II-E and used for comparison purposes.

While feature-based methods can be appealing for their conceptual simplicity, it is difficult to establish optimality conditions for such methods. Instead, we focus here on a *model-based* approach in which we first postulate a signal model that is designed to capture key signal characteristics, and then develop a corresponding "optimal" detector based on that model. Of course, the optimality of the detector hinges on the accuracy of the model, and all models for biological signals are imperfect. Nevertheless, in a model-based approach the underlying assumptions are made explicit at the outset, rather than being hidden implicitly inside a feature extractor or neural network. This transparency can facilitate generalizations to new situations, such as multi-channel detection or variations in the signal characteristics of interest. For example, we have developed two versions of our model-based detector, the quadratic detector, which does not recognize the event time and the change-point quadratic detector, which does.

II. Methods

A. Data Collection

The data used for this analysis came from patients in the epilepsy surgery programs at the UM Health System in Ann Arbor and the Henry Ford Hospital in Detroit who were being treated for intractable epilepsy. ECoG was recorded from up to 126 subdural electrodes that had been implanted on the surface of the cerebral cortex of each patient for the sole purpose of recording seizure activity and mapping cortical function. The 4 mm diameter electrodes were arranged in grids or strips with a center-to-center distance of 1 cm. Electrode placement was selected solely for the purpose of epilepsy monitoring without regard for this research, and electrodes were not necessarily located over motor cortex. Due to time constraints in the hospital environment, each subject performed sets of about 50 repetitions of a simple movement. The movements were self-paced (unprompted) and spaced roughly five seconds apart. To enable off-line analysis for algorithm development, actual movements (instead of preferred motor imagery) were used so that the corresponding muscle activity recorded by electromyography (EMG) electrodes could be used as the "ground truth" reference. The EMG onset provided a "trigger", the only labeled instant for each event. Most of the data was collected at a sampling rate of 200 Hz, with some at 400 Hz.

Twenty datasets from 10 subjects, consisting of a total of 2184 channels were used for the study. These channels included data from many areas of the brain, some of which were unrelated to the action performed. All 2184 channels were used for the comparison of the detection methods. For algorithm optimization, however, we selected a subset of 233 "interesting" channels. Interesting channels were defined as the channels for which any of the detection methods considered had an HF-difference (our performance metric, described in Section II-B) greater than 70 for either the training or testing data, indicating that the channel contained brain activity related to the action performed.

For simplicity, in this study we focused on detecting event-related changes in ECoG recorded from a single electrode channel. The extension of the methods to multiple channels will be considered in future work.

B. Event Detection in Unprompted Experiments

Unprompted subject actions produce a realistic, but challenging detection task analogous to the manner in which a computer input device is expected to operate. The asynchronous nature of event detection when the event timing is self-paced (*i.e.*, user-paced) presents a challenge for any detection algorithm. While some labeling is provided for the training data (in this case, the triggers indicating EMG onset), the test data is entirely unlabeled and any sample could correspond to an event.

The task of a detection method is to identify the detection points, that is, the points that correspond to the user's intent to activate the interface. However, changes in brain activity related to initiation and performance of the action occur both before and after the trigger point, so the algorithm is not required to identify this point precisely. Instead, we define an expected response window around the trigger point labeling each event and accept any detection point occurring within the expected response window as a valid detection ('hit'). The length of the expected response window after each EMG trigger specifies the maximum allowed delay between the actual occurrence of an event and its detection.

A detection point is defined when the decision feature produced by a method rises above a detection threshold, selected using the training data. A hysteresis thresholding approach is used in which no further detections are then reported until the decision feature falls below a lower threshold, which is set to the mean of the decision feature over the training data. We determine the detection threshold (the upper hysteresis threshold) empirically from the training data so as to maximize our performance metric, the "HF-difference," *i.e.*, the difference between the "hit" percentage and the "false" detection percentage. The hit percentage is the percentage of events that were detected within the expected response window. The false detection percentage is the percentage of the detection points reported by the method that were false (*i.e.* not hits).

In all cases, we used the ECoG containing the first half (typically 25 out of 50 available repetitions) of the events for algorithm training, and the remaining ECoG for testing. We report the HF-difference for the test data because a classical ROC evaluation is infeasible due to the use of unprompted events and incompletely labeled data.

C. CCTM Approach - Modeling the Mean

Initially, the UM-DBI project used the CCTM method for signal detection [1]–[3]. Briefly, the CCTM method calculates an ERP template from the training data using triggered averaging. See Fig. 1 of [1] for examples of ECoG signals and ECoG templates. The decision feature is formed by cross-correlating that ERP template with the ECoG of the test data. For this comparison, detection points are identified using the hysteresis threshold, though historically other threshold strategies have been used *e.g.*, [2]. The CCTM method is the only method in this comparison that explicitly includes ECoG after the detection point in the identification of the detection point (*i.e.* a significant portion of the ERP template energy occurs well after the trigger). Elimination of the undesirable delay that this causes was one of the design criteria for the model-based methods presented here.

While the CCTM approach was not developed as a model-based method, it is equivalent to a likelihood ratio test under a simple two-hypothesis statistical detection model [10] assuming signal buried in white noise. Let $\boldsymbol{x}$ denote one block of ECoG

data and suppose that $\boldsymbol{x}$ arises from one of the following pair of hypotheses:

$$\begin{aligned} H_0 &: \ \boldsymbol{x} \sim N(0, \sigma^2 \boldsymbol{I}) \qquad \text{“rest state”} \\ H_1 &: \ \boldsymbol{x} \sim N(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}) \qquad \text{“event state,”} \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}$ denotes the ERP template, σ^2 is the noise variance, and $\boldsymbol{I}$ denotes the identity matrix. For this model, the Neyman-Pearson optimal detector, formed from the likelihood ratio, is the inner product $\boldsymbol{x}'\boldsymbol{\mu}$, where the $\boldsymbol{x}'$ denotes $\boldsymbol{x}$ transpose. In practice we must choose between rest and event states not just once, but at each time point, so we slide the signal block $\boldsymbol{x}$ along the ECoG data one point at a time, applying the template to each block and producing the CCTM decision feature.

D. Power Spectrum Changes

The “white noise” signal model (1) ignores the well documented event-related changes in the signal power spectrum of EEG that are also seen in ECoG. For example, Fig. 1 shows a moving-window power spectrum with a window size of 200 ms computed by fitting a 4th-order AR model to 51 events, then normalizing by subtracting a baseline power spectrum (corresponding to time $t = -3$ seconds relative to the trigger). The power spectrum changes significantly near event onset. The proximity of the spectral changes to the trigger time raises the hope of reduced detection delay. Further, Fig. 1 may give some indication that spectral changes may begin before the trigger time. Further, such spectral changes are visible even in moving window power spectra from individual events as shown in Fig. 2.

E. Bandpower (BP) method

Power values in specific frequency bands are one of the standard methods for extracting features describing oscillatory activity [11]. An additional advantage of using bandpower is that oscillatory activity in specific frequency bands is associated with specific cognitive or mental tasks in well-known brain areas [12]. Although we do not present such a spatio-temporal analysis here, we employed bandpower as a “feature-based” approach with which to compare the results of our model-based approaches. Bandpower features were extracted by filtering the data with Butterworth filters of 4th order for the following frequency bands: 0-4, 4-8, 8-10, 10-12, 8-12, 10-14, 16-24, 20-34, 65-80, 80-100, 100-150, 150-200, 100-200 Hz. The last three bands were used only for datasets having a sampling rate of 400Hz. The filtered signals were squared and smoothed by either a 0.75 or 0.5 seconds moving average filter. The latter was used for frequency bands in the gamma range. The signals were linearly combined by an evolutionary algorithm to produce the decision feature. An advantage of this approach is that state labels are not needed for training. The evolutionary algorithm uses the HF-difference directly to optimize the linear combination on the training set (See [7] for details).

F. Covariance Signal Model and Quadratic Detector

The quadratic detector [10] was developed as an alternative to (1) that accounts for power spectra changes. We now assume that each ECoG signal block $\boldsymbol{x}$ arises from one of the following two states:

$$\begin{aligned} H_0 &: \ \boldsymbol{x} \sim N(0, \boldsymbol{K}_0) \qquad \text{“rest state”} \\ H_1 &: \ \boldsymbol{x} \sim N(0, \boldsymbol{K}_1) \qquad \text{“event state,”} \end{aligned} \tag{2}$$

where we ignore the ERP component $\boldsymbol{\mu}$ for simplicity. By the Neyman Pearson lemma, the most powerful test for such a detection problem is given by the likelihood ratio. Under the model (2), the likelihood ratio simplifies (to within irrelevant constants) to the following quadratic form:

$$\Lambda(\boldsymbol{x}) = \boldsymbol{x}'(\boldsymbol{K}_0^{-1} - \boldsymbol{K}_1^{-1})\boldsymbol{x}. \tag{3}$$

We slide the signal block along the ECoG data to form the decision feature signal, and then identify the detection points as described in Section II-B.

1) Training: The covariance matrices $\boldsymbol{K}_0$ and $\boldsymbol{K}_1$ in (2) are unknown *a priori*, so one must estimate them from training data. If the length of the signal block is, say, 100 samples, corresponding to 0.5 seconds of ECoG data, then each covariance matrix is 100×100. This would be too many parameters to estimate from limited training data. Therefore, we assume a pth order autoregressive (AR) parametric model for the signal power spectrum as follows:

$$x[n] = -\sum_{m=1}^{p} a_q[m]x[n-m] + u[n], \tag{4}$$

where $n > p$, $q = 0, 1$ (each state) and

$$u[n] \sim N(0, \sigma_q^2).$$

As usual, we assume that the $u[n]$ are independent and identically distributed (i.i.d.). Thus, $\boldsymbol{a}_0$ and σ_0^2 fully describe the rest state while $\boldsymbol{a}_1$ and σ_1^2 fully describe the event state.

We used the Schwarz information criterion (BIC) [13] to choose model order. This is similar to AIC (Akaike's Information Criterion) [13] but penalizes the number of parameters in the model more heavily. The distribution of "interesting" channels among best BIC model order estimates are shown in Fig. 3.

As can be seen from the histograms, the estimated best model orders using BIC for the 200 Hz and 400 Hz datasets are 3 and 4 respectively. In order to have a common model order, we used p = 4 for all datasets. For a 4th order AR model, we must estimate 4 AR coeffecients and a driving noise variance σ_q^2 for each of the two signal states, for a total of 10 unknown parameters.

If each ECoG training data sample point were labeled as coming from a "rest" or "event" state, then it would be straightforward to find the maximum-likelihood (ML) estimates of the AR coefficients and driving noise variances using the Yule-Walker equations ([14], section 5.4). However, our ECoG experiments are unprompted and only a single time instant is labeled for each event. This incompletely labeled ECoG data complicates the training process. To "label" our training data for the purposes of estimating the AR model parameters, we must estimate which ECoG signal samples correspond to which state.

For labeling purposes, we assume that the brain is in the "event" state for some (unknown) period before and after each EMG trigger. We parameterize these event-state intervals using a variable w that describes the width of the interval around each trigger where the signal is assumed to be "event" state, and a variable c that describes the relative location of the center of each event-state interval relative to each EMG trigger time point. We assume that the remainder of the training data belongs to the "rest" state.

Fig. 4 illustrates how we label the training data, where $M_{q,k}(c,w)$ denotes the number of samples in the kth block under hypothesis q, and $x_{q,k}[n;c,w]$ indicates the nth data sample in the kth data block under hypothesis q. By construction, $M_{1,k}(c,w) = w$.

With this model we construct a joint probability density function for training data by adapting the procedure in [14]. The AR parameters $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$ represent the rest and event states.

$$\begin{aligned}\log \mathsf{p}\big(\boldsymbol{x}_{1,k}, \boldsymbol{x}_{0,k}, \forall k; \boldsymbol{a}_1, \sigma_1^2, \boldsymbol{a}_0, \sigma_0^2, c, w\big) \approx \\ -\frac{1}{2\sigma_1^2}\sum_{k=1}^{K-1}\sum_{n=p+1}^{M_{1,k}(c,w)} u_{1,k}^2[n;c,w] \\ -\frac{1}{2\sigma_0^2}\sum_{k=1}^{K}\sum_{n=p+1}^{M_{0,k}(c,w)} u_{0,k}^2[n;c,w] \\ -\sum_{k=1}^{K-1}(M_{1,k}(c,w)-p)\log\sqrt{2\pi\sigma_1^2} \\ -\sum_{k=1}^{K}(M_{0,k}(c,w)-p)\log\sqrt{2\pi\sigma_0^2},\end{aligned} \tag{5}$$

where for $q = 0, 1$:

$$u_{q,k}[n;c,w] \triangleq x_{q,k}[n;c,w] + \sum_{m=1}^{p} a_q[m]x_{q,k}[n-m;c,w].$$

The approximation in (5) is reasonable when $M_{q,k}(c,w)$ is large relative to the model order. Based on this model, we use a joint ML estimation procedure to estimate simultaneously the AR parameters and the center c and width w of the event-state interval as follows:

$$\begin{aligned}(\hat{c},\hat{w}) = \arg\max_{c,w} \max_{\boldsymbol{a}_1,\sigma_1^2,\boldsymbol{a}_0,\sigma_0^2} \\ \log \mathsf{p}\big(\boldsymbol{x}_{1,k}, \boldsymbol{x}_{0,k}, \forall k; \boldsymbol{a}_1, \sigma_1^2, \boldsymbol{a}_0, \sigma_0^2, c, w\big)\,.\end{aligned} \tag{6}$$

This joint labeling/training procedure requires an iterative search over the center c and width w parameters (outer maximization). To conserve processing time we have specified a maximum of 5 iterations. The inner maximization has simple analytical solutions based on modified Yule-Walker equations ([14], section 5.4, equation 5.22) to find the AR parameters. Fig. 5 shows an example of the spectra computed from estimated AR parameters corresponding to the rest and event states.

2) quadratic detector implementation: Direct implemention of the quadratic detector (3) would be inefficient due to the large matrix sizes. Fortunately, for AR signal models one can implement (3) using simple FIR filters:

$$\Lambda(\boldsymbol{x}) = \Lambda_0(\boldsymbol{x}) - \Lambda_1(\boldsymbol{x}), \tag{7}$$

where

$$\Lambda_q(\boldsymbol{x}) \triangleq \frac{1}{\sigma_q^2}\sum_{n=p+1}^{N} u_q[n]^2, \qquad q = 0, 1,$$

and where the innovation signals are defined by

$$u_q[n] \triangleq x[n] + \sum_{m=1}^{p} a_q[m]x[n-m]. \tag{8}$$

The block diagram in Fig. 6 summarizes the implementation of the quadratic detector (7) and a Matlab implementation is given in the Appendix. The ECoG signal is passed in parallel through two FIR filters, each the inverse of the corresponding AR model. The output of the filters is squared and normalized by the ML estimates of the driving variances. Next, the difference operation in essence compares "which model fits better." This is followed by the calculation of a moving average of length 2/3 seconds (*i.e.* 133 and 267 samples at 200 Hz and 400 Hz respectively). The output is the decision feature that is compared to a threshold as described in Section II-B.

Fig. 7 illustrates how the variance of the innovations process works as a decision feature by plotting individually the normalized variance of innovations $\Lambda_0(\boldsymbol{x})$ ("rest state") and $\Lambda_1(\boldsymbol{x})$ ("event state"). Near the trigger point the signal power spectrum becomes that of the event state, so the event state variance of innovations decreases whereas the rest state variance of innovations rises, leading to a large decision feature value.

G. Change-Point Detector

The quadratic detector does not recognize that the event of interest occurs at a particular time. The change-point detector, based on a body of work in the statistics and control engineering literature [15], builds on the quadratic detector as follows.

We introduce the change point time j and an alternative hypothesis H_{0j}: that H_0 (the null hypothesis in (2)) holds up to and including discrete time $j-1$ and then H_1 (the event hypotheses in (2)) holds. Consider testing a sequence of hypothesis H_0 against the alternative hypothesis H_{0j}. For a given j, the likelihood ratio test (LRT) statistic for testing H_0 versus H_{0j} turns out to be

$$S_j^k = \sum_{n=j}^{k} s_n,$$

where s_n is an "instantaneous" likelihood ratio given by

$$s_n = \frac{1}{2}\ln\frac{\sigma_0^2}{\sigma_1^2} + \frac{u_{0,n}^2}{2\sigma_0^2} - \frac{u_{1,n}^2}{2\sigma_1^2},$$

where $u_{q,n} = u_q[n]$, $q = 0, 1$ is defined in (8). (See section 2.2 of [15] for this result in a simpler setting, and section 8.3 of [15] for the autoregressive version needed here.)

But the change point time j is unknown and we estimate it by maximum likelihood as

$$T_c = \hat{j} = \arg.max_{1\le j\le k} S_j^k$$

Then the change-point statistic is defined by

$$S_{\hat{j}}^k = max_{1 \le j \le k} S_j^k$$

Note that S_j^k is a cumulative sum of "instantaneous" likelihood ratios making this a CUSUM test [15]. Finally then, the test is: reject H_0 the first time k that

$$max_{1 \le j \le k} S_j^k \ge h$$

where h is the fixed threshold. If H_0 is not rejected by the time a prespecified window of data is processed (*i.e.* $k = k_{max}$) then it is accepted.

The block diagram in Fig. 8 summarizes the implementation of the change-point detector and a Matlab implementation is given in the Appendix. Similar to the quadratic detector, the ECoG signal is passed in parallel through two FIR filters, each the inverse of the corresponding AR model. The outputs of the filters are squared and normalized by the ML estimates of the driving variances. The difference of the resulting outputs is added to the logarithm of the ratio of the standard deviations. This is followed by the maximization of a sequence of moving sums with lengths from (model order + 1) samples to k_{max} which was set to 2/3 seconds (e.g. moving sum lengths for the maximization at 200Hz are 5, 6, ... , 133 samples where each sample corresponds to 0.005 seconds). The output of the maximization is the decision feature, which is compared to a threshold as described in Section II-B.

III. Results

We calculated the performance of the CCTM, BP, quadratic detector and change-point methods for all 2184 channels from the 20 datasets and compared the number of channels at various HF-difference levels. The HF-differences for each channel were calculated for expected response windows that started 0.5 second before the triggers and ended 1 second, 0.5 second or 0.25 second after the triggers. The change-point, quadratic detector, and BP methods all produced better detection than the CCTM. As the maximum allowed delay was reduced, the number of channels at each HF-difference level decreased. Fig. 9 compares the performance of the change-point, quadratic detector, BP and CCTM detectors when the delay is constrained to be at most 1 second by comparing the number of channels at each HF-difference level. Fig. 10 shows the 0.5 second delay case. The effect of changing the maximum allowed delay for the quadratic detector and change-point detection methods are shown in Fig. 11 and Fig. 12 respectively.

IV. Discussion

The change-point and quadratic detector methods detect event-related changes in ECoG based on a two-covariance signal model that captures event-related changes in the ECoG power spectrum. Both have a simple implementation that is suitable for real-time use. The change-point is an extension of the quadratic detector and estimates the change-point time at which the event of interest occurs.

The results showed that both the quadratic detector and change-point offer improved detection accuracy relative to the CCTM method and can provide reduced detection delay. The BP method also offers improved detection accuracy relative to the CCTM method, confirming that capturing spectral changes in the signal is important for detection.

TABLE I

THE NUMBER OF SUBJECT/ACTION COMBINATIONS (AND NUMBER OF SUBJECTS) AT EACH HF-DIFFERENCE LEVEL FOR THE BP, QUADRATIC DETECTOR (QD), AND CHANGE-POINT (CP) METHODS.

	max 0.25 sec after			max 0.5 sec after			max 1 sec after		
HF >	BP	QD	CP	BP	QD	CP	BP	QD	CP
50	3(3)	3(3)	6(5)	13(8)	12(9)	13(9)	17(10)	17(9)	17(9)
70	1(1)	1(1)	3(3)	7(5)	7(5)	8(6)	12(8)	13(9)	12(9)
90	1(1)	1(1)	1(1)	1(1)	1(1)	3(2)	8(5)	5(4)	7(5)

While the number of channels at different levels of detection is interesting, the number of subject/action combinations (and the number of subjects) at each detection level is also important. The subjects represent people who could benefit from a DBI, while each subject/action combination represents a potentially independent DBI output. The 20 datasets on which these methods were tested contain 19 subject/action combinations from 10 subjects. As can be seen from Table I, the BP and quadratic detector methods have comparable numbers of subject/action combinations (and number of subjects) at each HF-difference level. The change-point method produces comparable subject/action combinations for the 1 sec delay case. However, for the preferred shorter maximum delays, the change-point method produces more subject/action combinations in more subjects than the BP and quadratic detector methods.

There are several modifications to the quadratic detector and change-point models that may improve these detection methods. First, the current quadratic detector and change-point models ignore the ERP component (upon which the CCTM is based). Integrating the mean into the model would potentially capture both the temporal and spectral event-related characteristics and produce improved detection accuracy. Second, the likelihood ratio is optimal for prompted experiments with a predetermined block of data, but is not necessarily optimal when applied with a sliding window, as required by our unprompted data. It would be desirable to develop "optimal" detectors for unprompted experiments. Third, time-varying models (*e.g.*, state-space or hidden Markov methods) might better capture how the spectral properties evolve over time [5]. Fourth, the power spectra shown in Fig. 1-2 suggest that there are at least three distinct sets of spectral characteristics. Training individual detectors for each set of characteristics may produce more accurate detection than lumping them into just an event and a rest state. Finally, extensions to multi-channel detection are also under consideration. Combining information from multiple channels may be especially relevant for the change-point method, because the combined information from several channels (even ones with HF-differences around 50) could produce accurate detection.

V. ACKNOWLEDGEMENT

The authors gratefully acknowledge Alois Schlögl for discussions about spectral changes.

REFERENCES

[1] S. P. Levine, J. E. Huggins, S. L. BeMent, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde, and D. A. Ross, "Identification of electrocorticogram patterns as the basis for a direct brain interface," J. Clin. Neuro., vol. 16, no. 5, pp. 439-47, Sept. 1999.

[2] J. E. Huggins, S. P. Levine, S. L. BeMent, R. K. Kushwaha, L. A. Schuh, E. A. Passaro, M. M. Rohde, D. A. Ross, K. V. Elisevich, and B. J. Smith, "Detection of event-related potentials for development of a direct brain interface," J. Clin. Neuro., vol. 16, no. 5, pp. 448-55, Sept. 1999.

[3] S. P. Levine, J. E. Huggins, S. L. BeMent, R. K. Kushwaha, L. A. Schuh, M. M. Rohde, E. A. Passaro, D. A. Ross, K. V. Elisevich, and B. J. Smith, "A direct brain interface based on event-related potentials," IEEE Trans. Rehab. Eng., vol. 8, no. 2, pp. 180-5, June 2000.

[4] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," Clinical Neurophysiology, vol. 110, no. 11, pp. 1842-57, Nov. 1999.

[5] G. Foffani, A. M. Bianchi, A. Priori, and G. Baselli, "Adaptive autoregressive identification with spectral power decomposition for studying movement-related activity in scalp EEG signals and basal ganglia local field potentials," J Neural Eng, vol. 1, no. 3, pp. 165-73, Sept. 2004.

[6] N.-J. Huan and R. Palaniappan, "Neural network classification of autoregressive features from electroencephalogram signals for brain computer interface design," J. Neural Eng., vol. 1, no. 3, pp. 142-50, Sept. 2004.

[7] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet packet analysis," IEEE Trans. Biomed. Engin., vol. 51, no. 6, pp. 954-62, June 2004.

[8] E.C. Leuthardt, G. Schalk, J.R. Wolpaw, J.G. Ojemann, D.W. Moran. "A brain-computer interface using electrocorticographic signals in humans." J Neural Eng., vol. 1, no. 2, pp. 63-71, June 2004.

[9] J. A. Fessler, S. Y. Chun, J. E. Huggins, and S. P. Levine, "Detection of event-related spectral changes in electrocorticograms," in Proc. IEEE EMBS Conf. on Neural Engin., 2005, pp. 269-72.

[10] Jane E. Huggins, Bernhard Graimann, Se Young Chun, Jeffrey A. Fessler, Simon P. Levine, "Electrocorticogram as a Brain Computer Interface Signal Source," in towards Brain Computer Interfacing, edited by Guido Dornhege, Jose del R. Millan, Thilo Hinterberger and Dennis McFarland, Klaus-Robert Muller: MIT Press, 2007 (in press).

[11] G. Pfurtscheller, C. Neuper, and N. Birbaumer, "Human brain-computer interface," in Motor Cortex in Voluntary Movements: A distributed system for distributed functions, A. Riehle and E. Vaadia, Eds. CRC Press, 2005, pp. 367-401.

[12] G. Pfurtscheller, B. Graimann, J. E. Huggins, S. P. Levine, L. A, and Schuh, "Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement," Clin. Neurophysiol., vol. 114, no. 7, pp. 1226-36, 2003.

[13] R. Shumway and D. Stoffer, Time Series Analysis: Prentice-Hall, 2000.

[14] S. M. Kay, Modern spectral estimation. New York: Prentice-Hall, 1988.

[15] M. Basseville and I. Nikiforov, Detection of Abrupt Changes: Prentice-Hall, 1991.

VI. Appendix

The matlab implementations of feature extraction using the quadratic detector and the change-point detector are shown below.

A. Input and Output

```
% Inputs:
% ECoGdata: The ECoG data for testing
% sigma0:   Variance for rest class
% H0:       Model Coefficients for the rest
%           class
% sigma1:   Variance for event class
% H:        Model Coefficients for the event
%           class
% k_max:    the width (or maximum width for
%           the change-point method) of the
%           moving average (or sum for the
%           change-point method) filter
%           (2/3 second in samples)
```

```
% model_order: the model order (4)

% Output:
% feature:  the decision feature.
```

B. Matlab Implementation of Feature Extraction for the Quadratic Detector

```
function[feature]=quad_feature(ECoGdata, H0,
     sigma0, H1, sigma1, k_max)

e0 = filter( H0, 1, ECoGdata );
e1 = filter( H1, 1, ECoGdata );

feature = filter(ones( 1, k_max) / k_max,
    1, e0 .^ 2 / sigma0 - e1 .^ 2 / sigma1);
```

C. Matlab Implementation of Feature Extraction for the Change Point Quadratic Detector

```
function[feature]=changepoint_feature
    (ECoGdata, H0, sigma0, H1, sigma1,
     k_max, model_order)

e0 = filter( H0, 1, ECoGdata );
e1 = filter( H1, 1, ECoGdata );

s = 0.5 * log( sigma0 / sigma1 )
    + e0 .^ 2 / 2 / sigma0
    - e1 .^ 2 / 2 / sigma1;

feature=filter(ones(1, model_order +1), 1, s);

for m_order = (model_order + 2) : 1 : k_max
    test_feature=filter(ones(1, m_order),1,s);
    feature = max(feature, test_feature);
end
```

Figure Captions

Figure 1: Power spectrum changes relative to the spectrum at $t = -3$ seconds, averaged across 51 events. Time $t = 0$ is the EMG trigger time.

Figure 2: Power spectrum changes for individual events.

Figure 3: Histograms of estimated best model orders for interesting channels using the BIC criterion. A) 200 Hz channels, B) 400 Hz channels.

Figure 4: Training data with $K - 1$ events.

Figure 5: Frequency responses for each state constructed from estimated AR parameters.

Figure 6: Quadratic detector implementation.

Figure 7: The average of the variance of innovations for the rest and event states around the trigger points.

Figure 8: Change-point detector implementation.

Figure 9: The number of channels at each HF-difference level for the CCTM, BP, quadratic detector, and change-point methods with maximum allowed delay of 1 second.

Figure 10: The number of channels at each HF-difference level for the CCTM, BP, quadratic detector, and change-point methods with maximum allowed delay of 0.5 second.

Figure 11: Quadratic detector performance at differing maximum allowed delays.

Figure 12: Change-point detector performance at differing maximum allowed delays.

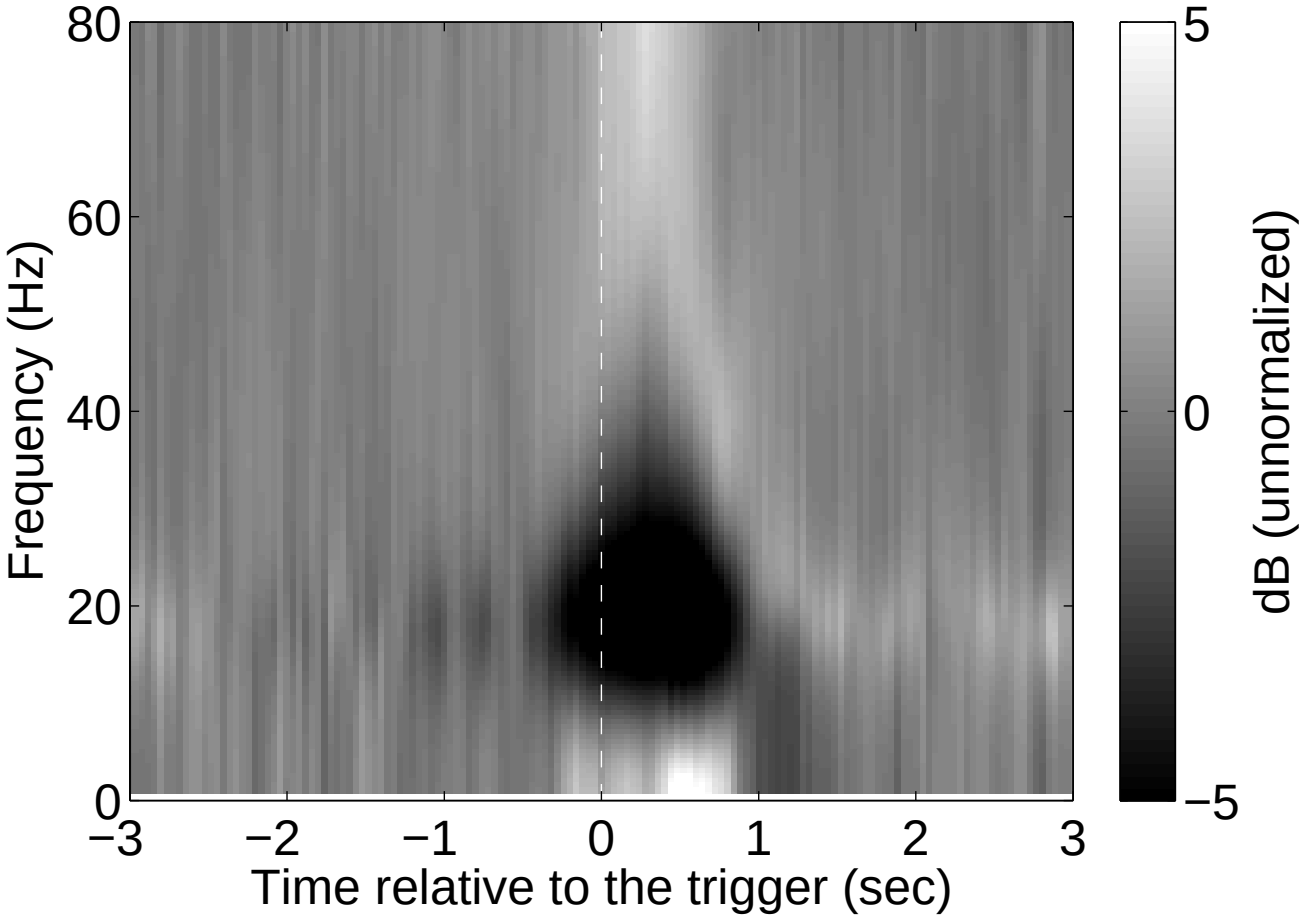
80
60
40
20
0
Frequency (Hz)
−3
−2
−1
0
1
2
3
Time relative to the trigger (sec)
5
0
−5
dB (unnormalized)

Fig. 1.

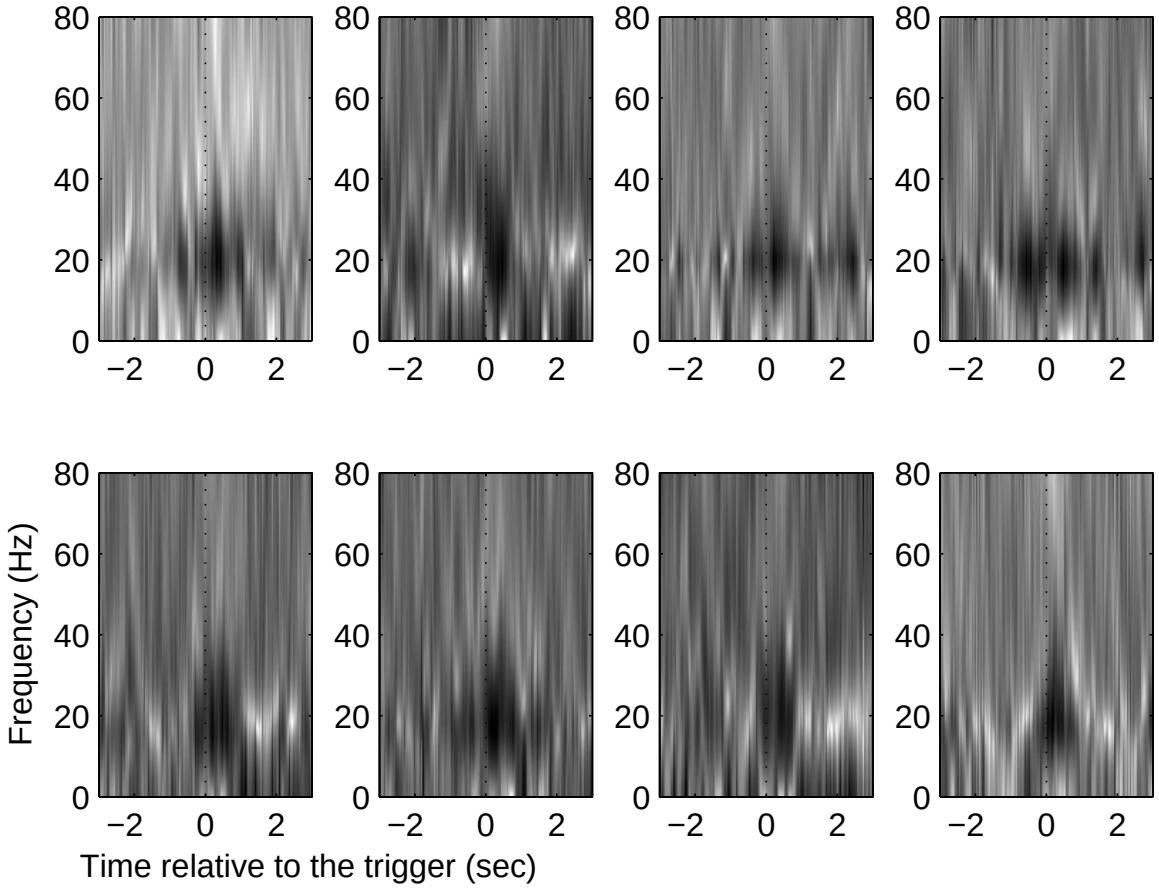
Frequency (Hz)
Time relative to the trigger (sec)
0
20
40
60
80
−2
0
2

Fig. 2.

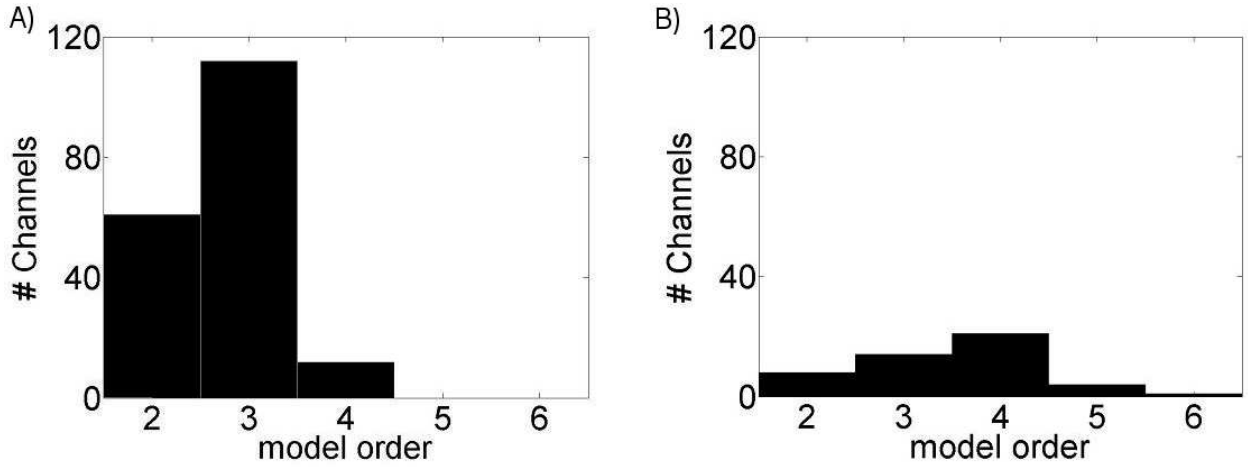
A)
120
80
40
0
Channels
2
3
4
5
6
model order
B)
120
80
40
0
Channels
2
3
4
5
6
model order

Fig. 3.

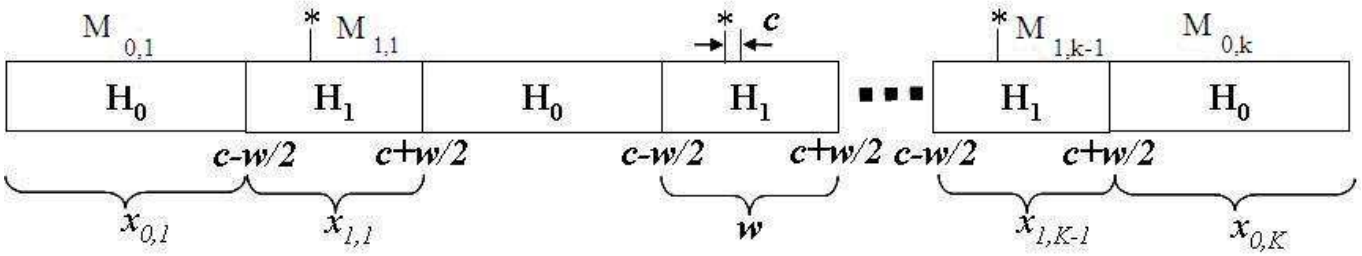
M 0,1
* M 1,1
* c
* M 1,k-1
M 0,k
H0
H1
H0
H1
H1
H0
c-w/2
c+w/2
c-w/2
c+w/2
c-w/2
c+w/2
x0,1
x1,1
w
x1,K-1
x0,K

Fig. 4.

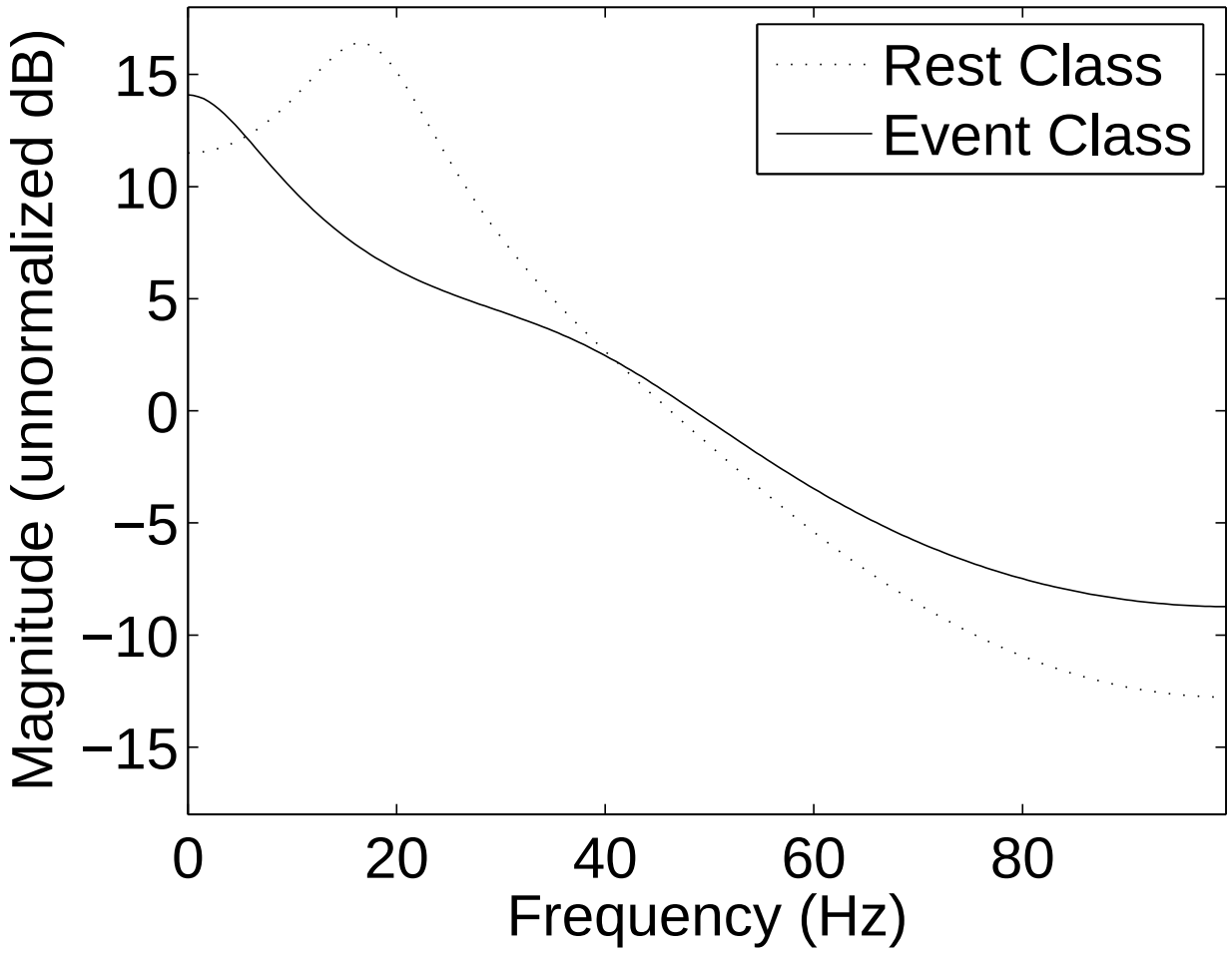
Rest Class
Event Class
Magnitude (unnormalized dB)
15
10
5
0
−5
−10
−15
0
20
40
60
80
Frequency (Hz)

Fig. 5.

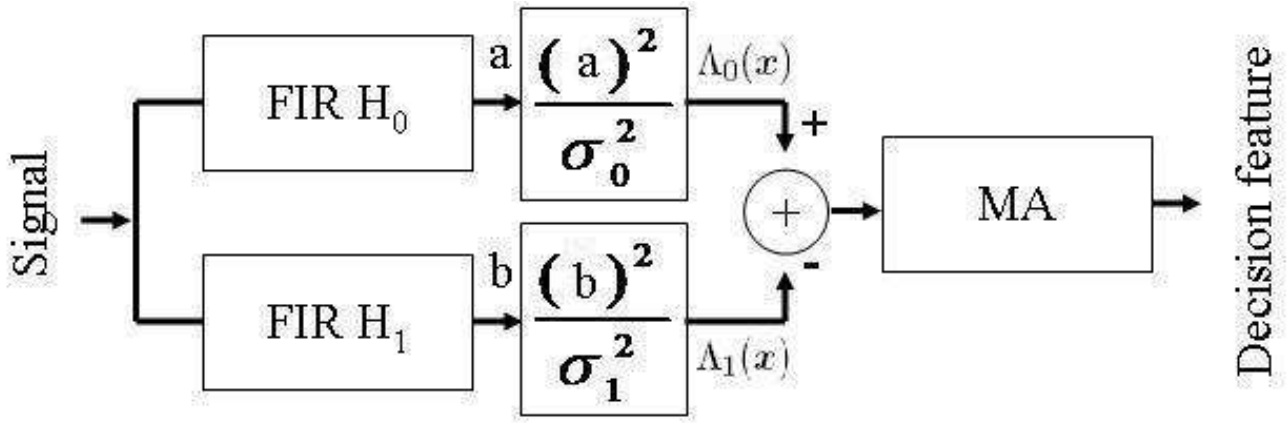
Signal
FIR H_0
FIR H_1
a
b
$\frac{(a)^2}{\sigma_0^2}$
$\frac{(b)^2}{\sigma_1^2}$
$\Lambda_0(x)$
$\Lambda_1(x)$
+
-
+
MA
Decision feature

Fig. 6.

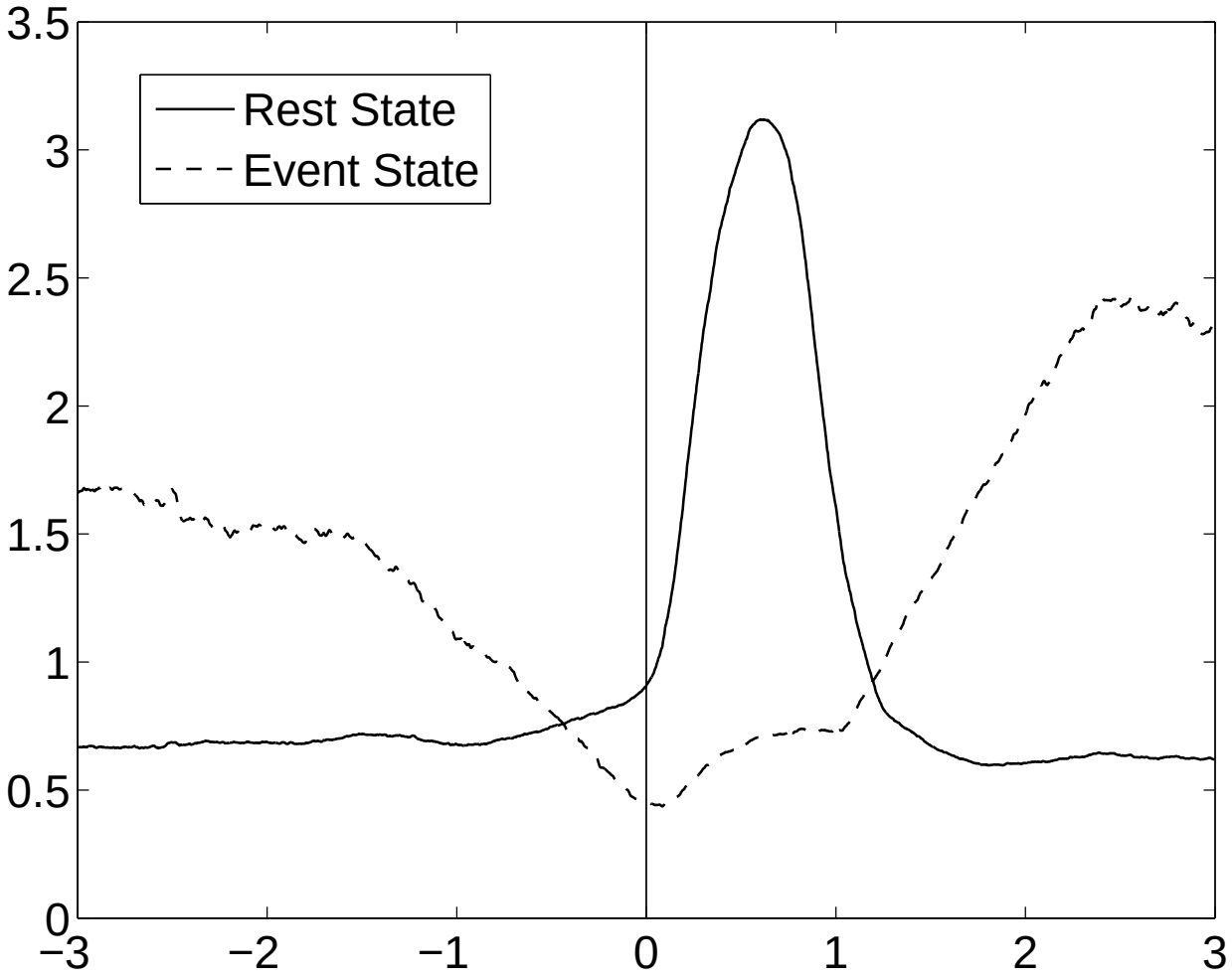

Fig. 7.

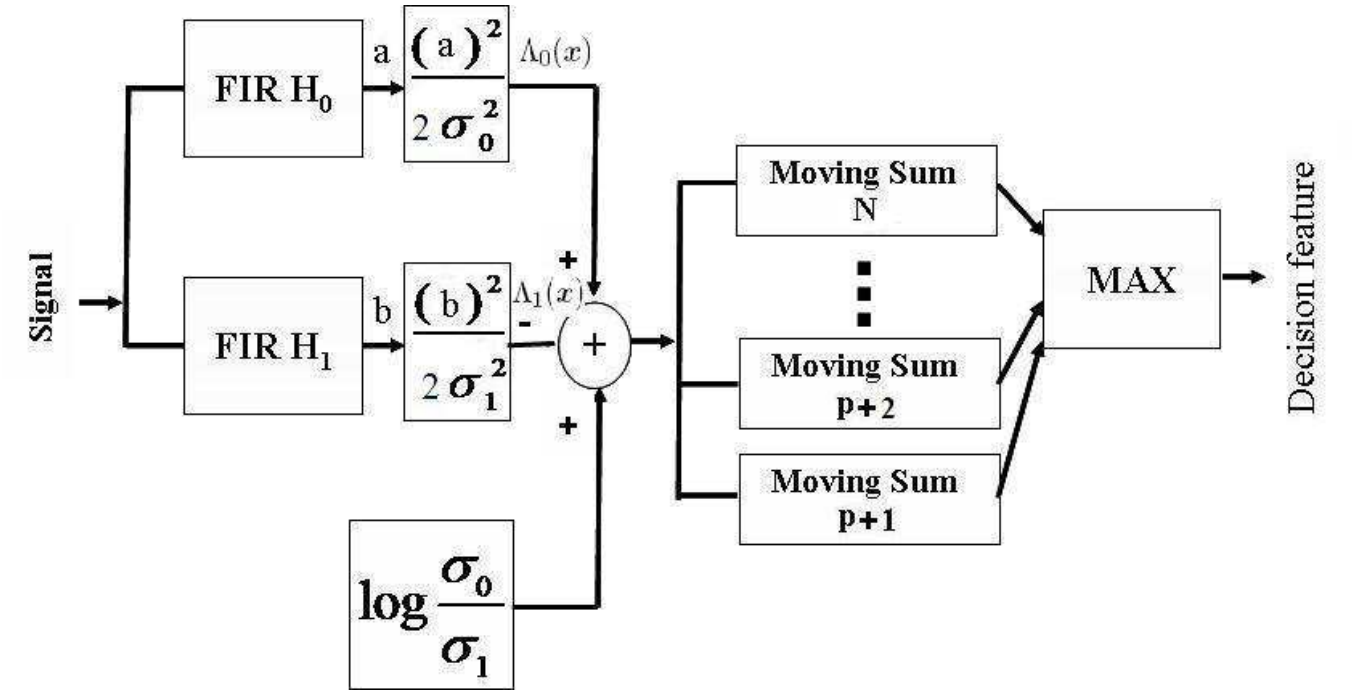
Signal
FIR H_0
FIR H_1
a
b
$\frac{(a)^2}{2\sigma_0^2}$
$\frac{(b)^2}{2\sigma_1^2}$
$\Lambda_0(x)$
$\Lambda_1(x)$
+
-
+
+
$\log\frac{\sigma_0}{\sigma_1}$
Moving Sum N
Moving Sum p+2
Moving Sum p+1
MAX
Decision feature

Fig. 8.

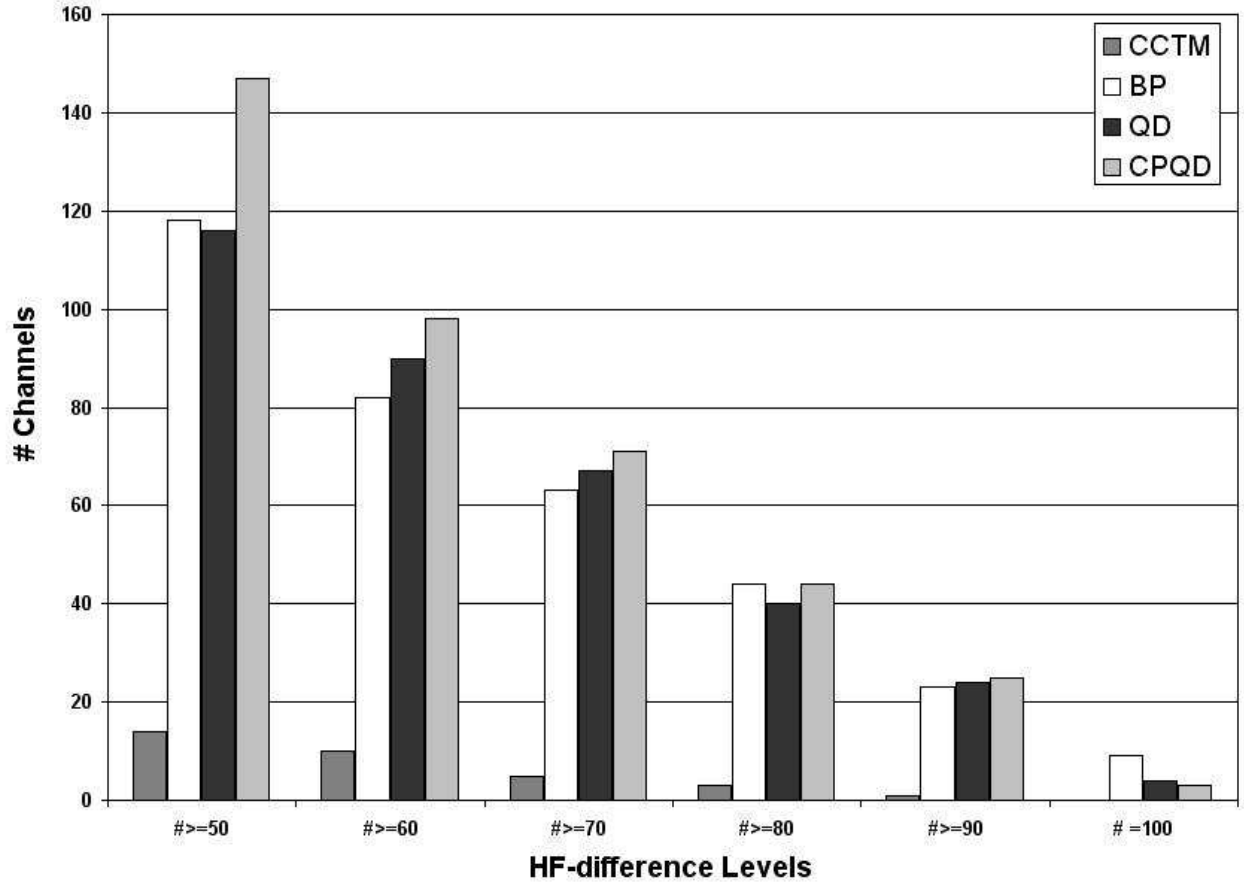
160
140
120
100
80
60
40
20
0
Channels
CCTM
BP
QD
CPQD
#>=50
#>=60
#>=70
#>=80
#>=90
=100
HF-difference Levels

Fig. 9.

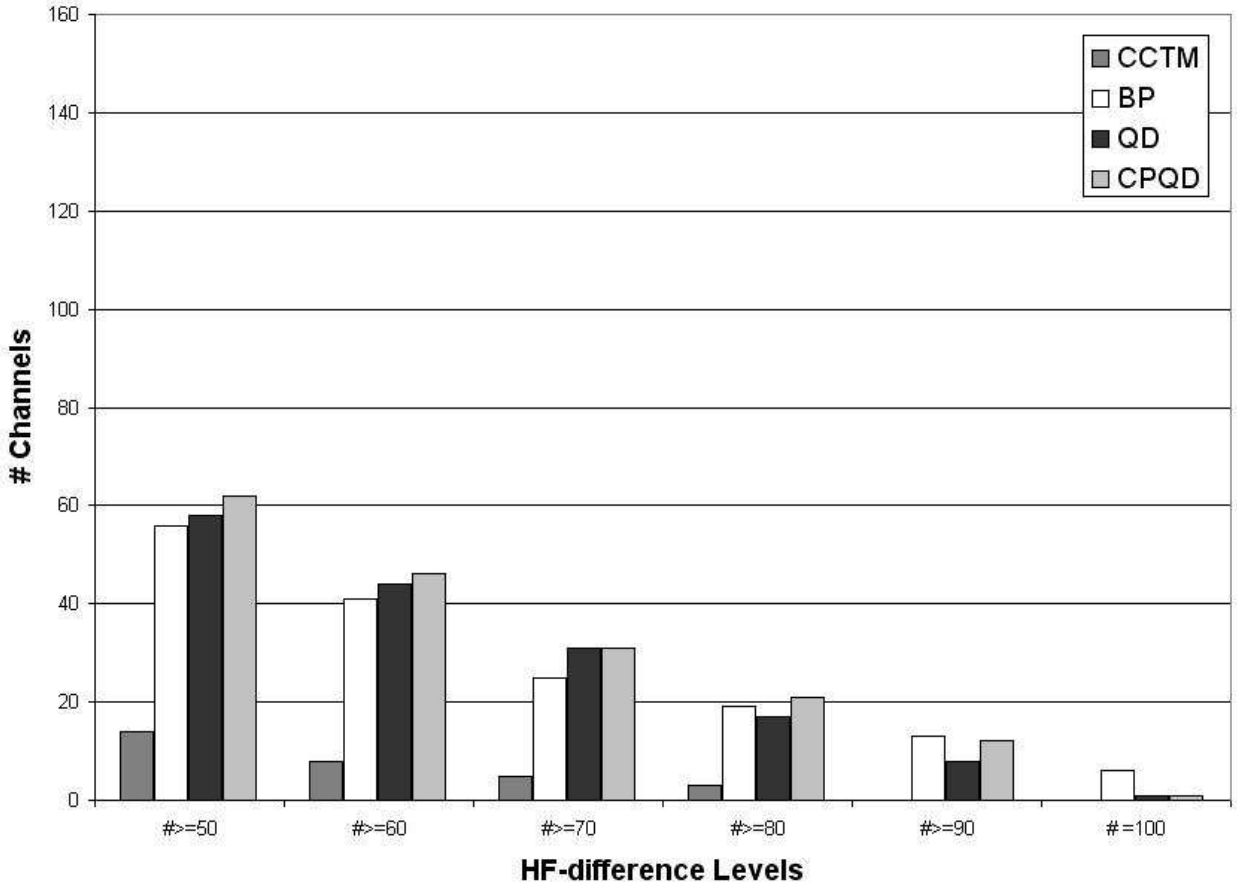
160
140
120
100
80
60
40
20
0
Channels
CCTM
BP
QD
CPQD
#>=50
#>=60
#>=70
#>=80
#>=90
#=100
HF-difference Levels

Fig. 10.

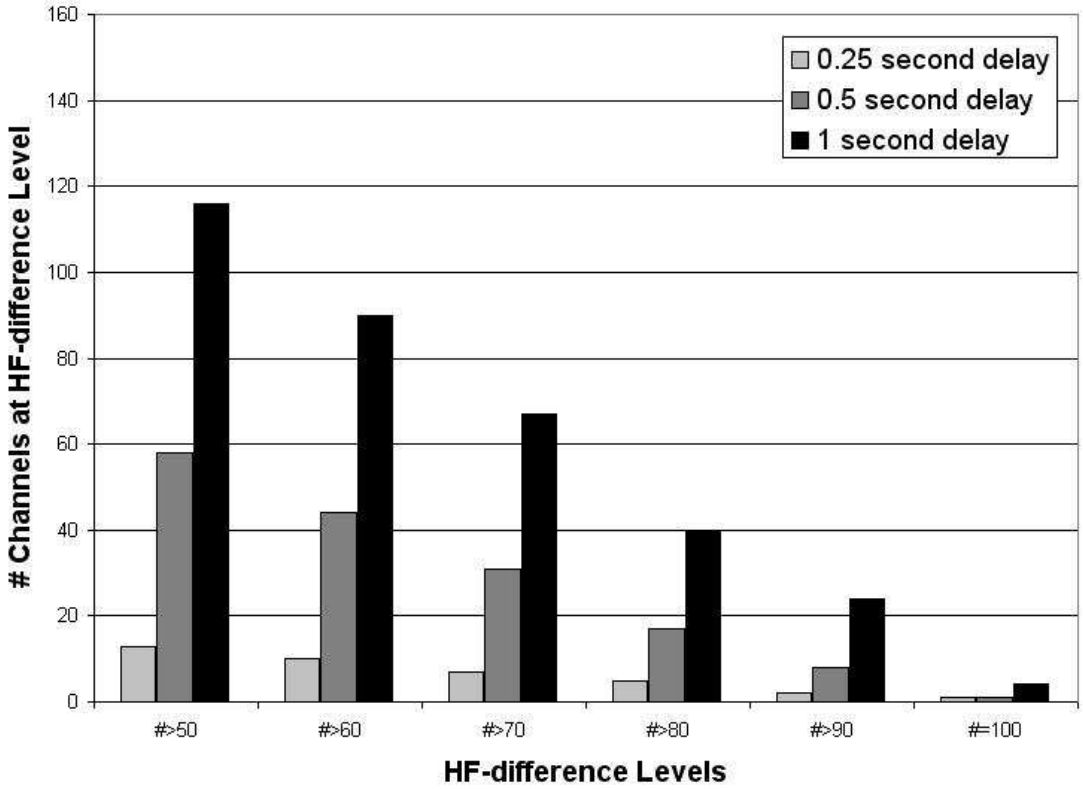
160
140
120
100
80
60
40
20
0
Channels at HF-difference Level
0.25 second delay
0.5 second delay
1 second delay
#>50
#>60
#>70
#>80
#>90
#=100
HF-difference Levels

Fig. 11.

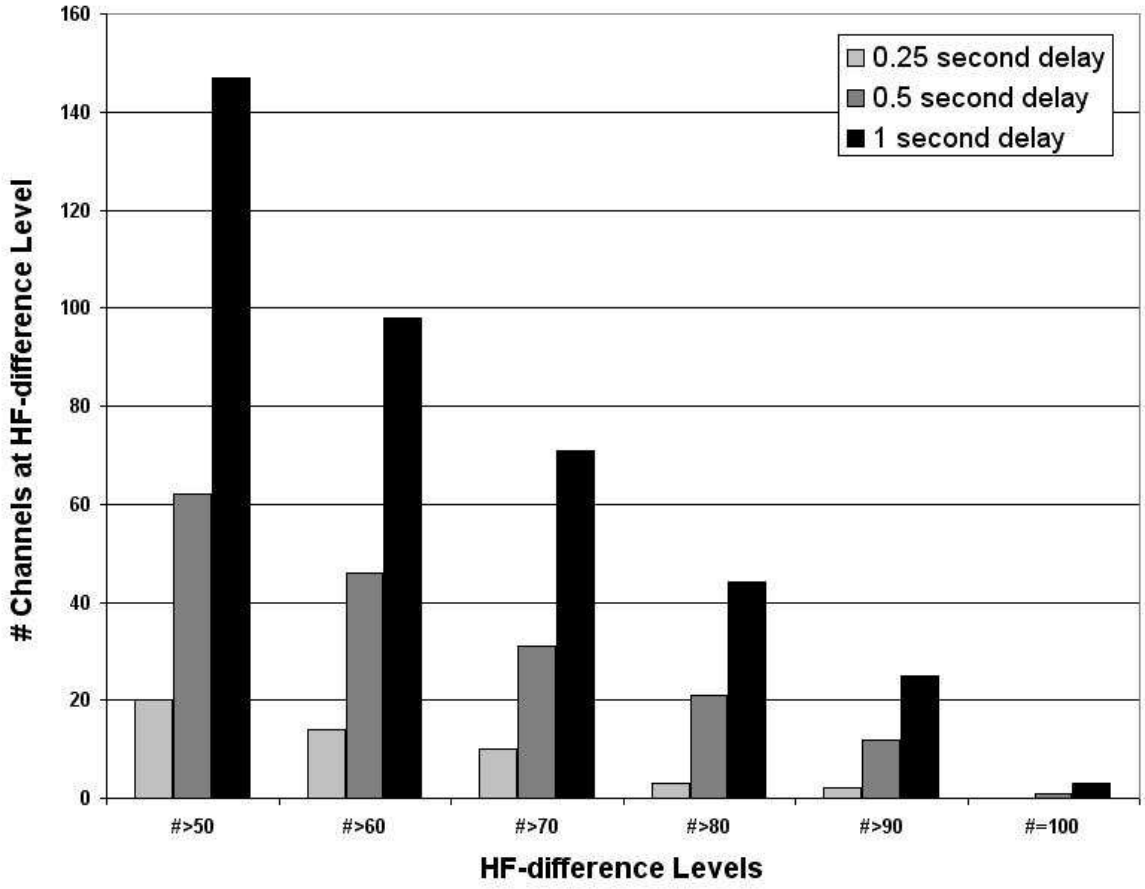
160
140
120
100
80
60
40
20
0
Channels at HF-difference Level
0.25 second delay
0.5 second delay
1 second delay
#>50
#>60
#>70
#>80
#>90
#=100
HF-difference Levels

Fig. 12.