Abnormality detection in noisy biosignals

Emine Merve Kaya and Mounya Elhilali
Department of Electrical and Computer Engineering
The Johns Hopkins University
Baltimore, MD 21218, USA
Email: {merve, mounya}@jhu.edu

Abstract—**Although great strides have been achieved in computer-aided diagnosis (CAD) research, a major remaining problem is the ability to perform well under the presence of significant noise. In this work, we propose a mechanism to find instances of potential interest in time series for further analysis. Adaptive Kalman filters are employed in parallel among different feature axes. Lung sounds recorded in noisy conditions are used as an example application, with spectro-temporal feature extraction to capture the complex variabilities in sound. We demonstrate that both disease indicators and distortion events can be detected, reducing long time series signals into a sparse set of relevant events.**

I. INTRODUCTION

Computer-aided diagnosis (CAD) research focuses on the detection of disease and its precursors. While this is an important goal in computational processing of medical signals, studies are generally based on relatively clean data with no detrimental amount of noise. Such an approach may not always extend its results to real data in practical applications. Although some imaging based areas of CAD have shown results significant enough to be now standard procedure in many hospitals [1]; diagnosis on many types of temporal signals remains an ongoing investigation. The standard approach in the literature is to extract features from the data and use a clustering algorithm to divide the data according to the likelihood and types of disease [2]–[7]. However, many of the employed machine learning algorithms are not sufficiently robust to noise and results may be significantly affected in the presence of distortions.

In this work, we tackle this issue by switching the CAD focus from diagnosis to detection of events of interest. The strategy is to identify occurrences of abnormalities in temporal signals. Once an event is detected, it can be examined by a physician or put through further computational processing to determine what it is. In the first case, the proposed approach is desirable because it turns a very long signal into a sparse and easily manageable sequence of events; while still keeping the physician in control of decision making. In the latter case, the mechanism provides a guided search method: Rather than blindly searching the entire signal for signs of disease, the problem is reduced to classifying what kind of abnormal event has occurred. We adopt a similar framework to one used to model auditory deviance detection processes in the brain [8], and apply this deviance detection scheme to find abnormalities in auscultation signals applied to lung sounds. The proposed scheme employs recursive tracking of temporal patterns in the signal using Kalman filtering, a popular choice in many medical applications [9]. Kalman filtering is fast, efficient, applicable to real-time tracking and robust to recording noise. Importantly, unlike other sophisticated artificial intelligence algorithms, Kalman filters are easy to interpret and adapt to the desired application.

The outline of this paper is as follows. First, we introduce the abnormality detection framework, which is the main component that can be applied to a variety of biosignals. Then, we explain the data considered in this work and the selection of features appropriate for the data. Last, the ability of our model to match expert labeling of the data is demonstrated.

II. METHODS

A. Detection scheme

The proposed scheme starts with mapping of the biosignal onto an appropriately chosen feature space. The choice of this space depends on the nature of the quantity of interest, and incorporate prior information about the modality under study. This issue shall be discussed in the following section. Once the transformation of the signal is achieved, the goal is to track the evolution of this signal in this new space and detect deviations from "normal" behavior. Following this concept of deviance detection, we first define what standard or normal behavior is, and the acceptable variance around this standard; in order to determine "abnormal" behavior. Based on this definition, the recursive scheme iterates through the temporal signal, predicting at each time instant how the signal should evolve. Any deviations from this prediction results in an alert or abnormal event.

The scheme employs tracking based on Kalman filtering. A Kalman filter is a discrete-time linear dynamical model with assumed Gaussian noise. It advances recursively through the signal to predict the value of a state or quantity, along with error covariance (measure of the estimate accuracy), and refines the estimate based on measurements of the state. The underlying process of a Kalman model is given in (1). X is the state that is being tracked, while Y is the real measurement. The measurement is inherently a function of the real state value. Both the state and measurement are perturbed by noise, respectively w and v. F, and H represent system matrices.

$$X_t = F_t X_{t-1} + w_t$$
$$Y_t = H_t X_t + v_t \quad (1)$$

Details of the implementation are presented in [8] but are summarized here. The system matrices are assumed to be constant (2). Given Q, the system noise covariance and R, measurement covariance, K_t, the Kalman gain, can be found. $\hat{P}_t$ is the state prediction error covariance matrix providing a measure of accurate fit. $\hat{X}_t$ represents the a posteriori estimate of the state X_t. Putting these together, we have the final recursive system that will hereafter be referred only as the "Kalman", presented in (3). Intermediate values such as a priori state estimate and innovation have been incorporated into the equations for conciseness.

$$F_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad H_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \qquad (2)$$

At the beginning of each Kalman, the initial values of the system are calculated from a short time segment, selected 1 second in the application. The variances of the noise parameters are arbitrarily set based on prior information of the feature at hand. Here we use $\sigma_w = 0.001$, and $\sigma_v = 0.06$.

$$
\begin{aligned}
K_t &= (F\hat{P}F^T + Q)H^T(H(F\hat{P}F^T + Q)H^T + R)^{-1} \\
\hat{X}_t &= F_{t-1}\hat{X}_{t-1} + K_t(Y_t - HF\hat{X}_{t-1}) \qquad (3) \\
\hat{P}_t &= (I - K_tH)(F\hat{P}_{t-1}F^T + Q)
\end{aligned}
$$

The detection scheme assigns one set of Kalmans to each feature, as shown in Fig.1. After finding the standard of the incoming sound feature within the initial buffer, a separate Kalman is allocated to each stream. At each consecutive time instance, the incoming value in that feature is compared with every prediction in the Kalman set of that feature. If no filter has predicted this value, the time is recorded to be of interest, and a new stream is initialized so that the value may form a new stream if it occurs again not long after. If the value does match one of the predictions, the matching Kalman is updated to new estimates. Finally, a memory cleanup exists to remove the tracking of the streams that have not been updated for a long time (8 seconds in our implementation).

The decision of whether an input value fits into the prediction of a Kalman depends on the measurement innovation covariance, a metric of how well the predictions are matching the input. This is a direct function of the estimated error covariance, which is updated at every time step. Consequently, if the innovation is always small, the tolerance for fitting a Kalman will get smaller over time. But if the input signal is very noisy, the tolerance will grow. The validation gate used here is an adaptation from [10] with $\gamma = 4$ resulting in about 97% of probability mass inside validation gate.

$$|Y - \hat{X}_n| <= \sqrt{\gamma(\hat{P}(1) + \sigma_v^2)} \qquad (4)$$

This mechanism produces a spike train as output, indicating locations of interest. The amplitude of each spike is the distance of the real value at that time to the validation gate. For multiple spikes at the same time location, only the maximum spike is recorded. The last processing step filters spikes with a threshold depending on which feature they originated from.
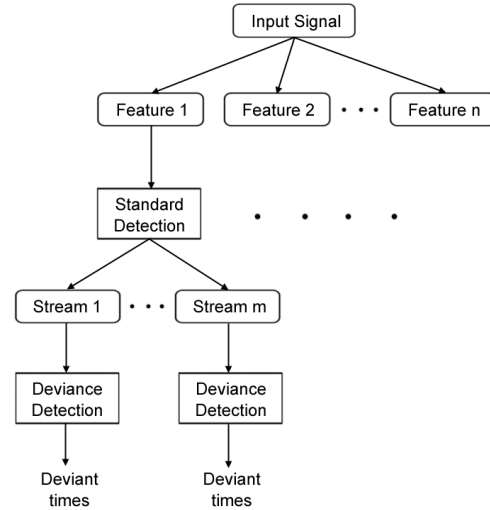

Fig. 1. The full detection model. Features extracted from the biosignal are put through the mechanism in parallel. Standard detection divides the feature into streams from which deviants are detected. New streams may be created, and unused ones erased as necessary.

B. Data

The data we used to demonstrate our framework is of paediatric auscultations recorded in a noisy environment. Lung sounds provide a marker of lung pathologies or airway obstruction. The base of the sound signal is a cyclic pattern of the airflow during breathing. If there is existence of a disease, anomalous patterns are superimposed onto the periodic breathing cycle. Two of the most studied patterns belong to stationary wheezes and transient crackles, which are markers of specific pathologies. These patterns are not trivially described within a signal processing framework. Characteristically, lung sounds span 50-2500Hz, wheezes are slightly higher frequency of about 100-2500Hz, and crackles span about 100-500Hz.

Previous studies have mostly dealt with recordings from adults in a controlled environment, resulting in relatively clean data where noise was not considered a significant effect; however, this might not always be the case in practice. The data used in this work was obtained from outpatient paediatric clinics in Kathmandu, Nepal and Lima, Peru [11]. The data contains artifacts from background chatter, child crying, and environment noise. Acquisition was made from a digital recording stethoscope connected to an MP3 player, sampled at 44100Hz. For the Nepal data, annotations were made by two physicians at 15 second segments, labeling the segment as normal, crackle or wheeze (with no specification of exact timing). For the Peru data, annotations were available for every event other than regular breathing, labeled by type of event in addition to onset and offset times.

C. Feature extraction

Due to the acoustic nature of our dataset, we have selected spectro-temporal features in our application of the detection scheme. It should be noted again that the detection scheme can be used on many different types of features; the ones presented

here are merely an appropriate selection for the given task.

We aim to capture the subtle variances in the sound signals with a thorough analysis of their spectro-temporal properties. Mimicking the information processing of the early auditory system [12], the signal is bandpass filtered and put through a mechanism including high-pass filtering, non-linear compression, and low-pass filtering, in simulation of inner hair cells. The final output is the integration of a lateral inhibitory network (see [13] for implementation details of the spectrogram). Our spectrogram is computed with time windows of length 8ms with no overlap and 128 channels. In our experiments, we use data with a sampling rate of 8kHz, which gives us center frequencies ranging logarithmically between approximately 100Hz and 4kHz. Every channel of the spectrogram is treated as a separate feature in deviance detection.

Bandwidth information is computed from the spectrogram by filtering it with cortical bandpass filters, mimicking the response of neurons at the mammal auditory cortex, which are tuned to a range of spectral resolution and temporal modulation [14]. The characteristic ripple frequencies to compute the bandwidth feature are selected logarithmically between 2^{-2} and 2^4 cycles per octave. Rate information is computed from the spectrogram by frequency binning of overlapping windows across time. In this implementation, power is summed for 10 bins among 200ms windows.

Lastly, we include envelope information of the original feature waveform. Loudness difference can be indicative of wheezes (which typically appear louder) or other interesting information in the signal. The envelope is computed by taking the magnitude of the Hilbert transform of the data, and running a Butterworth filter of cutoff 60Hz and order 6 through it.

There is two types of normalization during feature extraction. First, before the time-frequency analysis, the sound signal is normalized by its standard deviance at every 1s segment. Second, after the features are extracted, they are processed to reduce sampling artifacts. With our Kalman trackers, we want to see the overall level change rather than point-to-point level change, which is highly dependent on the original sampling rate and further complicated with resampling. To alleviate this problem, we use a non-overlapping windowing scheme where at each window we assign the mean of the window to every sample. Window length is arbitrarily selected to be 0.8s. Following, the features are downsampled with a rate of 1/8. The resulting feature streams have the necessary information represented in a significantly smaller number of samples than the original sound waveform.

Thresholds used for filtering spikes at the output of the deviance mechanism are as follows. Envelope: 0.5; Spectrogram: 0.5, Bandwidth: 0.6, Rate: 0.55. Thresholds are application specific and determined based on results from the Nepal dataset. Same threshold values are used for the Peru dataset.

III. RESULTS

For each of the two datasets at hand, we extract the clearly labelled trials to test the model on. For the Nepal dataset, there are 24 trials of varying length. Annotation is available at 15

TABLE I
DETECTION RATES ON THE NEPAL DATASET ON A SEGMENT BASIS

	# Segment	# Miss	Detection
Total "normal"	38	11	71.05%
Wheeze	36	0	100.00%
Crackle	29	6	79.31%
Wheeze & crackle	15	0	100.00%
Total "abnormal"	80	6	92.50%
Total	118	17	85.59%

TABLE II
DETECTION RATES ON THE PERU DATASET ON AN EVENT BASIS

	# Event	# Miss	Detection
Stethoscope	88	4	95.45%
Verbal	122	8	93.44%
Child movement	78	10	87.12%
General noise	30	3	90.00%
Child crying	30	3	90.00%
Ambient sounds	14	2	85.71%
Total "normal"	362	30	91.71%
Wheezing	41	0	100.00%
Rhonchi,crackle,ad.sound	14	2	85.71%
Total "abnormal"	55	2	96.36%
Total	417	32	92.33%

second increments. There is no information about where in the 15 second segment the abnormal event occurs. Four possible labels are assigned: Normal, crackle, wheeze, both crackle and wheeze. Our task is not the differentiation of diseases, but merely the signs of abnormality. Thus, we group the latter three labels into one label, "abnormal". There are a total of 38 "normal" and 80 "abnormal" segments throughout the 24 trials.

We run each trial separately, so the system is initialized only at the beginning of the trial, and consecutive segments of the same trial are fed into the already running system. If there is a spike that is higher than the threshold in any feature during the duration of the annotated event, the segment is treated as "abnormal". If there is no spike in any feature, it is treated as "normal". Calculation of performance is done on a segment basis to match the annotations. Results are presented in Table I.

The Peru dataset has a more detailed labeling scheme in which environmental sounds are also annotated, reflecting the increased complexity of sounds represented in the recordings. Trials with complete annotation and no cutoffs in recording are selected for testing the model. There are 47 such trials in total, each at about 1m20s in length. Six of these trials are from patients with bronchiolitis, and five are from patients with asthma. There are six different types of noise distinctly noted: Stethoscope movement, verbal communication, patient movement, noise, child crying, and ambient noise such as background chatter or television. For abnormal lung sounds, there are four classes: Wheezes, crackles, adventitious sounds, rhonchi. Sounds except wheeze are grouped due to their low number of occurrence. Time annotations are of beginning and ending times of one of the events. Events may be overlapping. The number of events in one trial varies between 2-33.

Again, each trial is run separately. The output of spikes

are compared in time with the time annotations of their corresponding trial. If there is a spike during the time of an event, that event is marked as detected. If there is no spike in any feature during an annotated event, the event is marked as missed. In the case of overlapping events, when there is a spike during their time range, it is counted as a hit for both events; as our goal is simply to drive attention to time instances. Spikes that occur but do not correspond to any annotated event are marked as an error. Results are presented in Table II. Additionally, there are a total of 61 spikes unaccounted for, averaging to 1.3 spike per trial.

We can see from the results that there is a high rate of detection for relevant events in both experiments. In both sets of data, we find the wheezing with 100% accuracy, whereas the performance drops for crackles. This is not unexpected due to wheezes being more conspicuous from a signal processing perspective, as they are louder and last longer; while crackles have the disadvantage of being overpowered in both energy and time. Although the Nepal dataset has less noise than the Peru dataset, the noise is not explicitly annotated. There are some segments where there is stethoscope movement and loudness artifacts, which are apparent upon listening. However, we have kept the annotation of the physician as a reference. The low classification rate of eventless segment classification reflects this discrepancy between apparent signal problems and annotation. To summarize, we are not teaching the system that the expert labeled trials are ground truth and comparing other trials to that; we are rather providing an unsupervised detection scheme that will find any signal abnormality, and this is reflected in the results of both datasets.

In the Peru dataset results it is much more clear that the results of Nepal dataset is due to annotation discrepancies. Since signal abnormalities are labeled more thoroughly, we are able to show that our model indeed finds any type of noise in the signal: Be it abnormal lung sounds, or abnormal environmental artifacts. The amount of false alarms is low, corroborating our claim that most of the deviants our model finds are in fact events of interest.

IV. CONCLUSION

We have proposed a method of abnormality detection in biosignals following an alternate interpretation of computer-aided diagnosis. Rather than the traditional scheme of making decisions on whether a given trial of medical test is indicative of disease or not; we follow the approach of marking points of interests among time for further processing, either by a physician or computationally. We have aimed at capturing a more sparse representation of the biosignal by reducing it to only events of potential interest. This approach allows us to make more versatile decisions by being prepared for any type of distortions and artifacts from recording noise.

Results on noisy datasets were presented, illustrating the complications of making signal-based diagnosis in the presence of significant noise. The results indicate that we can successfully reduce the signal into its significant events. Our results might lead to a shift in computer-aided diagnosis

methodology, especially in one dimensional temporal signals, where learning on small segments is complicated with issues such as lack of knowledge on whether the event exists in part or in whole in the segment. The processing we have demonstrated takes care of such problems by pinpointing the precise location of the abnormal event in the signal.

Further, although our presented results are on lung sounds, our mechanism has wide application due to its lack of dependence on the constraints of a specific modality. A large set of features is supported by the model, inviting application on different types of biosignals.

REFERENCES

[1] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198 – 211, 2007.

[2] U. Acharya, O. Faust, S. Sree, F. Molinari, R. Garberoglio, and J. Suri, "Cost-effective and non-invasive automated benign and malignant thyroid lesion classification in 3d contrast-enhanced ultrasound using combination of wavelets and textures: A class of thyroscan algorithms," *Technol Cancer Res Treat.*, vol. 10, pp. 371–380, 2011.

[3] F. Dehghan, H. Abrishami-Moghaddam, and M. Giti, "Automatic detection of clustered microcalcifications in digital mammograms: Study on applying adaboost with svm-based component classifiers," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 4789 –4792.

[4] O. Faust, U. R. Acharya, L. C. Min, and B. H. C. Sputh, "Automatic identification of epileptic and background eeg signals using frequency domain parameters." *International Journal of Neural Systems*, vol. 20, no. 2, pp. 159 – 176, 2010.

[5] M. Karimi, R. Amirfattahi, S. Sadri, and S. Marvasti, "Noninvasive detection and classification of coronary artery occlusions using wavelet analysis of heart sounds with neural networks," in *Medical Applications of Signal Processing, 2005. The 3rd IEE International Seminar on (Ref. No. 2005-1119)*, 2005, pp. 117 – 120.

[6] L. Pesu, P. Helisto, E. Ademovic, J.-C. Pesquet, A. Saarinen, and A. Sovijarvi, "Classification of respiratory sounds based on wavelet packet decomposition and learning vector quantization," *Technology and Health Care*, vol. 6, no. 1, pp. 65–74, 1998.

[7] S. Jadhav, S. Nalbalwar, and A. Ghatol, "Artificial neural network based cardiac arrhythmia classification using ecg signal data," in *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, vol. 1. IEEE, 2010, pp. V1–228.

[8] E. M. Kaya and M. Elhilali, "A model of auditory deviance detection," March 2013, manuscript submitted for publication.

[9] H. Poor, *An introduction to signal detection and estimation*, 2nd ed. Springer-Verlag, 1994.

[10] E. Arnaud, E. Memin, and B. Cernuschi-Frias, "Conditional filters for image sequence-based tracking - application to point tracking," *Image Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 63 –79, 2005.

[11] L. E. Ellington, R. H. Gilman, J. M. Tielsch, M. Steinhoff, D. Figueroa, S. Rodriguez, B. Caffo, B. Tracey, M. Elhilali, J. West, and W. Checkley, "Computerised lung sound analysis to improve the specificity of paediatric pneumonia diagnosis in resource-poor settings: protocol and methods for an observational study," *BMJ Open*, vol. 2, no. 1, 2012.

[12] X. Yang, K. Wang, and S. Shamma, "Auditory representations of acoustic signals," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 824 –839, 1992.

[13] T. Chi, P. Ru, and S. A. Shamma, "Multiresolution spectrotemporal analysis of complex sounds," *The Journal of the Acoustical Society of America*, vol. 118, no. 2, pp. 887–906, 2005.

[14] S. Shamma, H. Versnel, and N. Kowalski, "Ripple analysis in ferret primary auditory cortex. i. response characteristics of single units to sinusoidally rippled spectra," 1994.