Detection of Epileptic Seizures Through Audio Classification

G.R. de Bruijne[1], P.C.W. Sommen[1] and R.M. Aarts[1,2]

[1] Eindhoven University of Technology / Electrical Engineering, Eindhoven, The Netherlands
[2] Philips Research, Eindhoven, The Netherlands

Abstract — **Patients that undergo treatment in the epilepsy clinic Kempenhaeghe in the Netherlands are being monitored with different sensory signals, including audio. In this paper a new patient monitoring system for the detection of epileptic seizures through audio classification is proposed. The proposed system enables automated detection of epileptic seizures, which can have a large positive impact on the daily care of epilepsy patients. This system includes three stages. First, the signal is enhanced by means of a microphone array, followed by a noise subtraction procedure. Secondly, the signal is analyzed by audio event detection and audio classification. When an audio event is detected, features are extracted from the signal. Bayesian decision theory is used to classify the feature vector based on a discriminant analysis. Finally, it is decided whether or not to trigger an alarm. The performance is tested with audio signals obtained from measurements with epileptic patients. The results show that, with a limited set of features, good classification results can be achieved.**

Keywords — **Epileptic seizures, automated detection, patient monitoring, signal enhancement, audio classification.**

I. INTRODUCTION

Epilepsy is a neurological disorder affecting approximately one out of 150 people in the Netherlands. The brain is made up of a vast number of neurons that communicate with each other through electrical signals. Epileptic seizures happen when an abnormal electrical discharge occurs in the brain, disturbing its normal functioning. The type of seizure depends upon where it takes place, and how much of the brain is affected [1].

As a result of an epileptic seizure, sounds can be involved, e.g. when a patient screams during or after the seizure. Sounds can also come indirectly from a patient, e.g. when the seizure causes twitching of the muscles when the patient lies in bed. In the epilepsy clinic Kempenhaeghe the sounds in the room of a patient are captured by a ceiling loudspeaker that also serves as a microphone. An alarm is triggered in a control room, if the level of this signal exceeds a predefined threshold for a certain period of time.

Two shortcomings of the current system at the epilepsy clinic Kempenhaeghe are the number of false alarms and the number of potentially missed detections. The false alarms might come from surrounding noises, e.g. opening or closing doors, footsteps, sounds originating from outside the room, etc. A number of epileptic seizures are missed as a result of low level sounds.

In this paper an improvement of the current patient monitoring system is proposed. The system is introduced in Section II. In order to evaluate the system, an audio corpus is built, which is described in Section III. The features that are extracted from the signals in order to classify the sounds are described in Section IV. In Section V we show and discuss the experimental results. Conclusions and future works summarizes the paper in Section VI.

II. THE PATIENT MONITORING SYSTEM

The proposed system is depicted in Fig. 1. The sounds picked up from the patients room are first sampled with an analog-to-digital converter at $f_s = 10$ kHz.

A. Signal enhancement

The incoming audio signal is first enhanced by means of a microphone array. A microphone array is a spatial filter that separates signals that have overlapping frequency content, but originate from different spatial locations [2]. It is assumed that the monitoring system is active during the night, so the patient should normally be situated in bed. With the microphone array, we can focus on the sounds originating from around the bed. The spatially filtered signal is high-pass filtered at approximately 90 Hz to remove any low-frequency hum or noise components which might be present in the audio signal.

Our dataset includes audio signals obtained from measurements with epileptic patients. These signals are corrupted by broadband noise as a result of the air conditioning system. However, we can estimate the original signal by means of a noise reduction technique known as magnitude spectral subtraction [3]. The basic principle of this technique is to subtract the magnitude spectrum, obtained during a pre-estimated noise only period, from that of a noisy signal.

J. Vander Sloten, P. Verdonck, M. Nyssen, J. Haueisen (Eds.): ECIFMBE 2008, IFMBE Proceedings 22, pp. 1450–1454, 2008

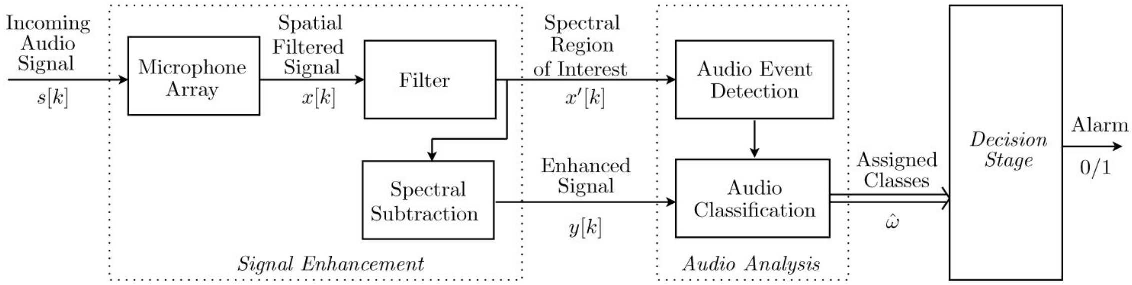

Fig. 1: Blockscheme of the proposed patient monitoring system.

B. Audio analysis

From the daily patient monitoring at night, a sound flow will continuously be analyzed, including audio event detection and audio classification.

1) Audio event detection: In order to have a reliable monitoring system, every event in the noisy audio signal should be detected. Therefore, we segment the continuous signal in frames of 25 ms. The power is calculated and compared to a predefined threshold Th_d after third order median filtering. The threshold is determined during one second of background noise only, and it is calculated by [4]

$$Th_d = BNP_{mean} + k \cdot BNP_{std}\,, \tag{1}$$

where BNP denotes the background noise power, and k is a tunable parameter.

2) Audio Classification: For the classification of sound types, two-class classifiers are used. Each classifier is trained to classify the sound as being the seizure-related sound of interest, e.g. screams, or not. Prior to classification, the enhanced signal is segmented in analysis frames of one second, features are extracted from half-overlapping subframes of 25 ms and the resulting feature vector is classified.

There are many types of features that can be extracted from the (sub)frames. To reduce the computational load, we restrict this research to the use of low-level features, like zero-crossings rate, band-energy ratio, etc. We use the Bhattacharyya distance as a performance measure to select powerful features [5]. We further assume that the probability density functions follow a Gaussian distribution with mean μ_c, and standard deviation σ_c, with c denoting the class index. As a result, the decision boundaries form quadratic surfaces in the feature space. This is referred to as quadratic discriminant analysis [6].

C. Decision stage

In this final stage, it is decided whether or not to trigger an alarm. In this paper, the focus is on the signal enhancement stage and the audio analysis stage.

III. AUDIO MATERIAL

In order to train and validate the monitoring system, real recorded sounds obtained from real epileptic patients in the Kempenhaeghe clinic are used. This real dataset originates from 2003. These sounds were recorded with a directional microphone (Beyerdynamic, type MCE86 N(C).02). For one night, nine patients were monitored with audio, video and EEG signals. A subset from a total of 428 video files recorded from 17 different patients was analyzed and annotated. Sound was observed during 61 out of 95 seizures.

Because of the limited real dataset, extra recordings were collected during this research. The sounds were recorded with ten volunteers in two different (quiet) rooms, with different microphones: a Creative Labs headset HS-300 and a Behringer ECM800. Extra environmental sounds were collected from different sound libraries. From the real dataset, the average signal-to-noise ratios (SNR) of each sound class is estimated, so that the energy level of the simulated sounds can be adapted to the expected levels in practice [7]. Our total corpus contained 253 files (78 real, 175 simulated), with a total duration of 14 minutes, and it includes the following sounds:

- Seizure-related sounds: screams, smacking of lips, movements of the bed due to shaking of the patients during the seizures, noises due to bronchial secretion, snores, moans, sighs, and heavy breathings.
- Non-seizure related sounds: speech, bed squeaks, pillow sounds, coughs, and other patient related sounds.
- Environmental sounds: opening/closing doors, footsteps, flushing toilets, ticking clock, traffic and weather-related sounds.

IV. Features

In our patient monitoring system a subset of the observed sounds during or after an epileptic seizure is being classified. For each sound class, features are carefully selected that describe the sounds, based on temporal and spectral information.

A. Screams

Four features are extracted from the subframes in order to detect screams.

Short-time energy: The short-time energy in a subframe is obtained by the sum of the squares of the signal. The mean of the measured short-time energies in a frame is used for classification. Compared to the other expected sounds, screams have high signal energies.

Pitch: Here, the pitch is the frequency that corresponds to the maximum in the normalized autocorrelation function within a limited range of delays. To optimize the accuracy of the pitch detection, the input signal is modified by means of soft center clipping [8], and the pitch detector is followed by a fifth order median filter. For screams, the pitch in subsequent subframes has relatively small variation. As a result, if we measure the number of subsequent frames such that $|\Delta \text{pitch}[m]| = |\text{pitch}[m+1] - \text{pitch}[m]|$ is smaller than a predefined threshold, and $m=0,1,\ldots,M-1$ is the subframe index where M is the number of subframes in an analysis frame, we expect a high score.

Pitch strength: The strength of the pitch is determined by the value corresponding to the maximum in the normalized autocorrelation function. We use the average pitch strength as a feature for classification.

Spectral flatness of the residual after linear prediction: In speech processing, a simplified model that is often used for describing the signal is a time-varying all-pole spectral shaping filter with a variable (glottal) excitation source: a white noise source for unvoiced sounds, and periodic pulses for voiced sounds [4]. If we use linear prediction to estimate the spectral shaping filter, the spectrum of the residual signal for voiced sounds is not flat. We can measure the spectral flatness by calculating the ratio of the geometric mean to the arithmetic mean of the power spectrum. It is one for a constant spectrum. As a result, voiced subframes will have a lower score than unvoiced subframes. The average spectral flatness in a frame is used as a feature for classification.

B. Smacking of the lips

Three features are extracted from the subframes in order to detect these sounds: short-time energy, spectral centroid, and the first linear prediction coefficient. Due to the repetitive short transitions of the signal, the variations of the feature values are of interest.

Low short-time energy ratio (LSTER): It is defined as the ratio of the number of subframes whose short-time energy is less than 0.5 times of the average short-time energy in a one second frame. There are a relatively high number of silence subframes in lipsmack sounds. Therefore it is expected that it has a high *LSTER* score.

Temporal behavior of the spectral centroid: The spectral centroid is the center of mass of the power spectrum. Due to the repetitive patterns of the feature values during smacking of the lips, the modulation spectrum of this feature has a peak around the repeating frequency, and decaying harmonics. The modulation energy between 2-10 Hz appears to be discriminative.

Temporal behavior of the first linear prediction coefficient: Similar to the temporal behavior of the spectral centroid, this feature has most of its modulation energy between 2 Hz and 10 Hz.

C. Movements of the bed due to shaking of the patients during the seizures

When a patient has a tonic-clonic seizure, the person first stiffens (tonic) and then shakes (clonic). As a result, the bed of the patient will shake. In order to detect a shaking bed in Kempenhaeghe, small bells are often attached to the bed of the patient. This way, the ringing bell sounds can trigger the currently used monitoring system. The specific ring bell sounds observed in the real data shows high spectral peaks. We can use the band-energy ratio to discriminate between this sound and other expected sounds. The band-energy ratio is defined as the relative amount of energy present in a specified frequency range. Here, the average and standard deviation of B_r between 2.2 kHz and 2.6 kHz in a frame are used as features for classification.

D. Noises due to bronchial secretion

Often due to epileptic seizures, bronchial secretion with noticeable respiration is observed. Three features are extracted from the frames in order to detect these sounds: the average band-energy ratio between 2.5 kHz and 5 kHz, the average zero-crossing rate and the average first linear prediction coefficient.

V. Results and discussion

In the experimental results the microphone array was not incorporated, because the real data obtained from the Kempenhaeghe clinic was recorded with a single microphone.

A. Audio event detection

From the real dataset it follows that for *k=3*, the number of missed detections and the number of false detections is minimized. In order to validate the detection algorithm, we used screams, smacking of the lips, and noises due to bronchial secretion from the simulated dataset. We corrupted these clean signals with background noise recorded in the Kempenhaeghe institute and measured the number of missed detections. Furthermore we measured the number of false detection when only background noise was present. It follows that no false detections or missed detections occurred (Table I).

Table I Performance of the audio event detection algorithm

	Seizure-related sound	Background noise
Detection	72	0
No detection	0	72

B. Audio classification

For the training and validation of the audio classification stage, each sound in the simulated dataset was repeated three times with varying SNR, based on the measured average and standard deviation of the SNR from the real dataset. Furthermore, the signals were randomly corrupted with four different recordings of background noise in different rooms.

We used 10-fold cross validation to calculate the sensitivity (SE), specificity (SP) and the positive predictive value (PPV). The latter performance measure uses prior probabilities, which we estimated from the real dataset. The classification results are shown in Table II and Table III. Each entry represents the average value ± the standard deviation.

Table II Audio classification results

	Screams		Smacking of the lips	
SE	SP	PPV	SP	PPV
0.98	0.96 ± 0.01	0.30 ± 0.11	0.96 ± 0.00	0.02 ± 0.00
0.95	0.99 ± 0.01	0.53 ± 0.23	1.00 ± 0.00	0.75 ± 0.40
0.90	0.99 ± 0.01	0.63 ± 0.23	1.00 ± 0.00	0.76 ± 0.39
0.85	0.99 ± 0.01	0.72 ± 0.23	1.00 ± 0.00	0.77 ± 0.37

Table III Audio classification results continued

	Movements of the bed		Noises due to bronchial secretion	
SE	SP	PPV	SP	PPV
0.98	0.97 ± 0.01	0.40 ± 0.09	0.63 ± 0.01	0.02 ± 0.00
0.95	0.98 ± 0.01	0.53 ± 0.16	0.72 ± 0.01	0.02 ± 0.00
0.90	0.99 ± 0.01	0.60 ± 0.19	0.95 ± 0.01	0.11 ± 0.04
0.85	0.99 ± 0.01	0.66 ± 0.20	0.96 ± 0.01	0.14 ± 0.06

Screams can be detected with a high accuracy; with a sensitivity of SE=0.98, the average specificity is SP=0.96. With an estimated prior probability of 1.13%, this results in PPV=0.30±0.11, implying that approximately 1 out of 3 detections is a scream.

Sounds produced during smacking of the lips can be detected with a high accuracy (SE=0.98, and SP=0.96). However, due to a low prior probability (0.09%), the positive predictive value is low (PPV=0.02).

The results of the detection of noises due to bronchial secretion shows low performance. With a sensitivity of SE=0.95 and an average specificity of SP=0.72, a lot of false detections occur (approximately 98 out of 100 alarms are false). This low performance is mainly due to a diversity of sounds available for training of the classifier. The sounds are very typical and hard to imitate. The real data was recorded during the day and included a mixture of sounds with surrounding noises. As a result, only the simulated data is used for both training and validation of the classifier.

The detection of movements of the bed due to shaking of the patients during the seizures shows high accuracy (SP=0.97±0.01, and PPV=0.40±0.09 at SE=0.98). However, these results are based upon only three real recordings, with a total duration of 35 seconds. Nevertheless, we expect that the accuracy would still be high after validation with more data, because the band-energy ratio is very discriminant due to the typical peaks in the spectrum of the bell sounds.

VI. Conclusion and future work

In this paper we have presented a patient monitoring system for the automated detection of epileptic seizures. Good performance is obtained for a subset of the typical observed sounds during or after an epileptic seizure. It is evaluated with a limited dataset and still needs to be validated in real conditions together with a microphone array. The proposed system can be expanded such that more seizure-related sounds can be detected. Moreover, the decision stage enables to further improve the accuracy of the system, e.g. by looking at the order in which sounds occur.

Acknowledgment

The authors would like to thank J. Arends of the epilepsy clinic Kempenhaeghe for providing the essential data.

REFERENCES

1. Wilson S (1928) Epileptic variants. J Neurol Psychopath. 31:223–240
2. Van Veen B, Buckley K (1988) Beamforming: a versatile approach to spatial filtering. IEEE–ASSP 5:4–24
3. Boll S (1979) Suppression of acoustic noise in speech using spectral subtraction. IEEE–ASSP 27:113–120
4. Childers D (2000) Speech processing and synthesis toolboxes. John Wiley & Sons, New York
5. McKinney M, Breebaart J (2003) Features for audio and music classification. Proc. ISMIR
6. Duda R, Hart P, Stork D (1993) Pattern classification and scene analysis (2nd edition). John Wiley & Sons, New York
7. Virone G, Istrate D (2007) Integration of an environmental sound module to an existing in-home activity simulator. IEEE–EMBS Proc. vol. 29, Engineering in Medicine and Biology Society, Lyon, France, 2007, pp 3810–3813
8. Sondhi M (1968) New methods of pitch extraction. IEEE Trans. Audio Electroacoust. 16:262–266