Technical University of Denmark

DTU

Automatic multi-modal intelligent seizure acquisition (MISA) system for detection of motor seizures from electromyographic data and motion data

Conradsen, Isa; Beniczky, Sándor; Wolf, Peter; Kjaer, Troels W.; Sams, Thomas; Sørensen, Helge Bjarup Dissing

DTU Library
Technical Information Center of Denmark

Automatic multi-modal intelligent seizure acquisition (MISA) system for detection of motor seizures from electromyographic data and motion data

Isa Conradsen[a,b,*], Sándor Beniczky[b,c], Peter Wolf[b], Troels W. Kjaer[d], Thomas Sams[a], Helge B.D. Sorensen[a,**]

[a] Technical University of Denmark, Department of Electrical Engineering, Orsteds plads bygn. 349, 2800 Kgs. Lyngby, Denmark
[b] Danish Epilepsy Centre, Department of Neurophysiology, Kolonivej 1, 4293 Dianalund, Denmark
[c] University of Southern Denmark, Institute of Regional Health Services Research, Winsløwparken 19, 3, DK-5000 Odense C, Denmark
[d] Rigshospitalet University Hospital, Department of Clinical Neurophysiology, 2100 Copenhagen, Denmark

ARTICLE INFO

ABSTRACT

The objective is to develop a non-invasive automatic method for detection of epileptic seizures with motor manifestations. Ten healthy subjects who simulated seizures and one patient participated in the study. Surface electromyography (sEMG) and motion sensor features were extracted as energy measures of reconstructed sub-bands from the discrete wavelet transformation (DWT) and the wavelet packet transformation (WPT). Based on the extracted features all data segments were classified using a support vector machine (SVM) algorithm as simulated seizure or normal activity. A case study of the seizure from the patient showed that the simulated seizures were visually similar to the epileptic one. The multi-modal intelligent seizure acquisition (MISA) system showed high sensitivity, short detection latency and low false detection rate. The results showed superiority of the multi-modal detection system compared to the uni-modal one. The presented system has a promising potential for seizure detection based on multi-modal data.

1. Introduction

Epilepsy is a functional disorder of the brain caused by excessive discharges of groups of neurons clinically characterized by repeated unprovoked seizures lasting from seconds to minutes. About 1% of the world's population has epilepsy. Seizure manifestations can be motor (tonic, clonic, tonic–clonic, etc.), sensory, psychic or vegetative, and consciousness may be retained or altered, sometimes with automatic behavior. In spite of much progress with pharmacological, surgical and other treatments, about 25% of epilepsy patients continue to have seizures. For many of these patients, seizure onset is unpredictable, impairing independent living and increasing the risk of injuries, e.g. by falls or burns. Therapy resistant patients with generalized tonic–clonic seizures have an increased risk of dying as a consequence of a seizure, especially when they live alone and the seizures occur during sleep [1,2]. An automatic seizure detection system that alerts relatives or other helpers of an on-going seizure would alleviate several of these problems. The earlier a seizure is detected,

the more useful the system would be. It could also be beneficial in determining therapeutic success or failure in patients who live alone and cannot reliably report whether they still have seizures. A clinically feasible detection system needs to be both reliable and comfortable.

Today the diagnostic gold standard in epilepsy is electroencephalography (EEG) with simultaneous video surveillance. EEG is known to be reliable for detection of seizures [3–5]. EEG-recordings require, however, either an invasive recording (intracranial electrodes) or the placement of several scalp-electrodes, which is less stable over time. The patient might also be uncomfortable wearing electrodes on the scalp, which are very noticeable for others, and thereby stigmatizing the patient further. Despite the EEG method being the gold standard, it does not necessarily seem to be the best option for a seizure alarm outside the hospital. It has been attempted to use video recordings for seizure detection [6], but they had too many restrictions and limitations (obstacle-free area of movement covered by light and camera).

Nijsen et al. [7] used accelerometers (ACM) for seizure detection. The visual analysis of the movement data recorded with these sensors showed promising results (91% of the seizures with motor phenomena were detected), and was considered feasible for detection of seizures. Others [8,9] have tried to detect seizures based on ACM data, but an ideal method has not yet been presented.

Earlier studies [7–9] on detecting seizures from ACM data did not report on aspects concerning the time between seizure onset and the detection, but only on detection versus no detection. It is highly desirable to achieve an early detection of seizures (i.e. with only a few seconds of delay) to make possible an intervention to stop the seizures and/or prevent injuries during the seizures. To make such a system reliable for detection of seizures we decided to work with multi-modal data, so we extended the system from using only ACM to combine it with sEMG and gyroscopes (angular velocity (ANG) data). Gyroscopes provide information on the rotation of each joint, so this data covers e.g. movements where the limbs are accelerated less, but still rotated. In a preliminary study [10] we found that the three modalities sEMG, ACM and ANG provided complementary information with potential improvement of classification accuracy. The next issue was to identify the most promising features to distinguish between seizures and normal activities and furthermore identify the most appropriate classifier to automatically differentiate between the two classes based on the feature vectors. Nijsen et al. [11] showed through a visual analysis that the continuous wavelet transformation (CWT) seems to be a better feature than short time Fourier transformation (STFT) for ACM data. Seizure detection from sEMG signals is a rather unexplored field, but from a visual inspection of the data it seems that both the amplitude and the frequencies of the signal during seizures are different from normal activities. The discrete wavelet transformation (DWT) seems to be a good choice as a feature extraction method, since it provides a good frequency resolution at low frequencies and furthermore a good time resolution at the high frequencies. Based on this we used DWT for feature extraction and support vector machines (SVM) as a classifier in a pilot study [12], including both sEMG, ACM and ANG data, with very promising results

on distinguishing between seizures/simulated seizures and normal activities.

In this paper we search for the best feature extraction method based on the wavelet transformation to separate simulated seizures from normal activity. The wavelet transformation is good at describing both the morphology and the spatial distribution in the movement signals. Compared to the DWT, the wavelet packet transformation (WPT) provides equal time and frequency resolution for all frequencies. Besides DWT we have therefore also tested the WPT as a method for extracting features for all modalities in this automatic multi-modal intelligent seizure acquisition (MISA) system. To classify our data into the two groups, seizures and normal activities, we used SVM [13] (as in our pilot study [12]) as a binary classifier trained on feature vectors from both classes, since it is well known to function better than other classifiers when the data classes are of unequal sizes. We used data from healthy subjects who simulated seizures (as instructed by a physician) to develop our algorithm upon. To assess the similarity between the simulated seizures and a real one, we have visually compared the raw data from the simulated seizures with a real seizure from a patient for all modalities.

This paper is organized as follows: the recordings are presented in Section 2; data presentation is given in Section 3; the method in details in Section 4 and the results in Section 5. At last Sections 6 and 7 encompass the discussion and the conclusion, respectively.

2. Recordings

The goal of the project was to detect simulated seizures from multi-modal signals based on movement data (sEMG, ACM and ANG). To be able to statistically explore whether the automatic detection algorithm is functioning, the number of simulated seizures for each healthy subject had to be more than five. The reason for initially using healthy subjects (who simulated seizures) instead of epileptic patients was the difficulty in the patient recruitment. Most patients with more than five seizures with motor manifestations within a few days are mentally retarded and therefore have difficulties in cooperating, when wearing the suit containing the movement sensors. Therefore it has yet only been possible to collect seizure data from one patient and we only succeeded in obtaining one seizure from this patient. Ten healthy subjects who were instructed to simulate seizures are therefore monitored with all modalities and used for the project.

The project had been approved by the ethics committee of Region Zealand, Denmark. All subjects involved received information on the project and gave their written consent to participate in the study.

The recordings on healthy subjects were made at the Danish Epilepsy Centre in Dianalund, Denmark. Ten healthy subjects aged 23–30, both male and female, were included. It is assumed that there is no effect of gender. The measurements lasted 1.5–3 h for each healthy subject. All of the healthy subjects were asked to simulate three types of seizures and some normal activities. They were given a description of the seizures, and they watched seizures on a video. Before the recording the healthy subjects trained simulating the seizures

Table 1 – Information about the healthy controls.			
Subjects	Gender	Age	Test file length [h]
1	F	26	1.5
2	M	30	1.25
3	M	26	0.85
4	F	24	1
5	F	26	1
6	M	24	1
7	M	27	0.75
8	M	25	1
9	M	24	1.25
10	M	23	1

while assisted by a physician. During the recordings a physician was present to check that the simulated seizures were visually similar to real ones, when looking at the patients. If they were not, the subject was corrected and asked to simulate a new one. The normal activities were biking, use of mobile phone, computer and TV, eating and gambling with dices. The last activity was chosen because of the movement's similarity to clonic seizures or the clonic phase of the tonic–clonic seizures, whereas the rest were activities which the patients as well as gambling have access to during a normal admission in the epilepsy monitoring unit (EMU). Each of the seizures was simulated five times for each healthy subject. The times for the simulated seizures were annotated like during a normal admission. The three types of simulated seizures and their descriptions are as follows:

Myoclonic is a very short lasting twitch in a single muscle. The healthy subject is asked to make as short a contraction of the right biceps brachii as possible, which will cause a very short lasting movement of the right lower arm.

Versive–asymmetric tonic seizure is characterized by a turn of the head to an almost uncomfortable angle, where the healthy subject is looking upwards and to the side. This is followed by an isometric contraction in an asymmetric posturing, where the arm, on the same side towards which the head is turned, will be placed above the head.

Tonic–clonic starts as an isometric contraction of all the muscles. After a while it changes to rhythmically repetitive jerks made by alternating contraction and relaxation of the muscles.

The gender and age of the healthy subjects are listed in Table 1, as well as the length of the signals for the testing phase of the classification. The sEMG data were sampled at frequency of 1024 Hz, whereas ACM and ANG were sampled at frequency of 120 Hz.

The epileptic patient was admitted to the EMU at Rigshospitalet (Copenhagen University Hospital) for a diagnostic indication, as stipulated by the ethics committee. The admission lasted 3 days, where the patient, a 29 years old male had one seizure (of the generalized tonic–clonic seizure type). The time for the seizure has been clinically annotated by the neurophysiology technicians and later checked by a physician. The sEMG data were sampled at frequency of 1000 Hz, whereas the ACM and ANG were sampled at a frequency of 120 Hz as for the healthy subjects. The reason for the use of different sampling frequencies for the sEMG data is that the two participating departments use different recording programs with

different setup possibilities for the sampling frequencies. Furthermore the sEMG data from the EMU at the Danish Epilepsy Centre were filtered before exportation, so to equalize the frequency bands of the sEMG signals for the different subjects all data was filtered with a low-pass filter with a cut-off at 70 Hz.

The recordings all included both EEG, video, sEMG, electrocardiography (ECG) and motion sensors (ACM and ANG), but for this study only sEMG data and the data from the motion sensors were used. The motion sensors used are the system by Xsens [14] called Xsens MVN, which is a wireless system consisting of a suit with 16 sensors. Each of these sensors includes 3D ACM, 3D gyroscopes and 3D magnetometers. Based on the recordings, the Xsens MVN software system performs necessary biomechanical calculations, which provide data from 7 extra positions on the body (shown as position 2, 3, 4, 5, 6, 19 and 23 in Fig. 1). The output therefore is 3D ACM and 3D ANG from 23 different locations on the body, see Fig. 1. With these 23 placements on the body a full body system able to outline practically all movements of the body parts during seizures is obtained. For each of these two modalities we have 69 channels. For the recording on the patient we were not able to use the head sensor (position 7 in Fig. 1), hence, due to the biomechanical calculations in the software, data are useless for three positions (position 5, 6 and 7 in Fig. 1). The third modality, sEMG, is applied as 14 surface electrodes, each accompanied with its own reference electrode, placed on nearby bone or tendon. The sEMG electrodes are placed on the center of the belly of the muscle and symmetrically on the body on the following muscles: sternocleidomastoid, deltoid, biceps brachii, triceps brachii, biceps femoris, quadriceps femoris, and tibialis anterior. The muscles are chosen by a physician based on the knowledge on which parts of the body are most active during seizures, and to ensure full body coverage.

The recordings were performed by starting all conventional measurements in the EMU (i.e. all modalities except for the motion sensor system). When this was up and running, the motion sensor system was started and the time in the sEMG sampling system was annotated by the neurophysiological assistant, as precisely as possible. All data types were then used from this point and on, whereby they were synchronized.

The sEMG electrodes are connected to the EEG amplifier, which, in this case, is not wireless. It is the plan to implement the system as fully wireless in the future, allowing for the patients to move around more freely, while wearing the alarm and monitoring device. The hope is to identify which sensor positions are better at distinguishing seizures from normal activities and thereby being able to lower the number of sensors.

3. Data presentation

To assess the reliability of using simulated seizure data from healthy subjects instead of epileptic patients, the raw data from the simulated seizures were compared visually to an epileptic one for all modalities. Since we only could record one patient with epilepsy, a statistical comparison of the quantitative data/parameters was not possible. A representative simulated tonic–clonic seizure is shown in this paper for comparison with the real seizure. Fig. 2 shows the time
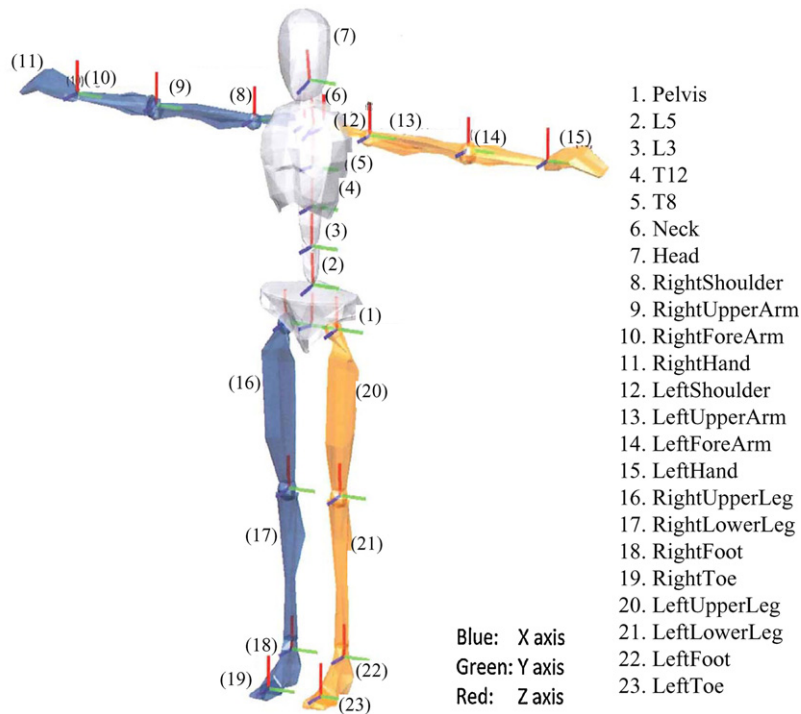
Fig. 1 – The placement of the sensor positions from the Xsens MVN system, see the Manual for Xsens MVN, revision D, June, 2008 [14].

Fig. 2 – The sEMG data for a real seizure and a representative simulated seizure are shown in a and b, respectively. The matching spectrograms (for a normalization of the signals) are shown in c and d, respectively, where the red color means high power, blue color means low. The data is from the right biceps. The seizure and the simulated seizure are both surrounded by normal activity data, 1.5 min prior and 1 min later. The black vertical lines represent onsets and offsets of seizures and simulated seizures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 3 – The ACM data for a real seizure and a representative simulated seizure are shown in a and b, respectively. The matching spectrograms (for a normalization of the signals) are s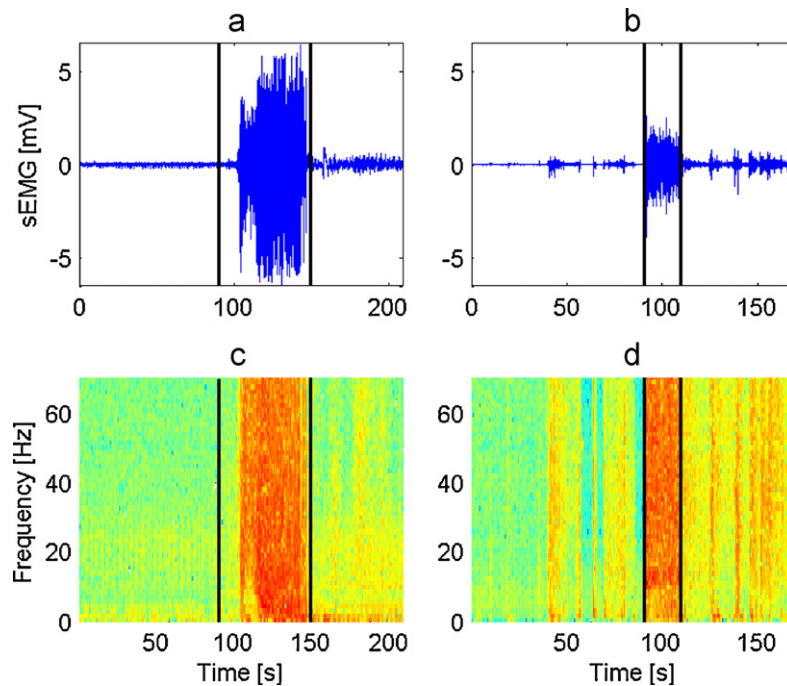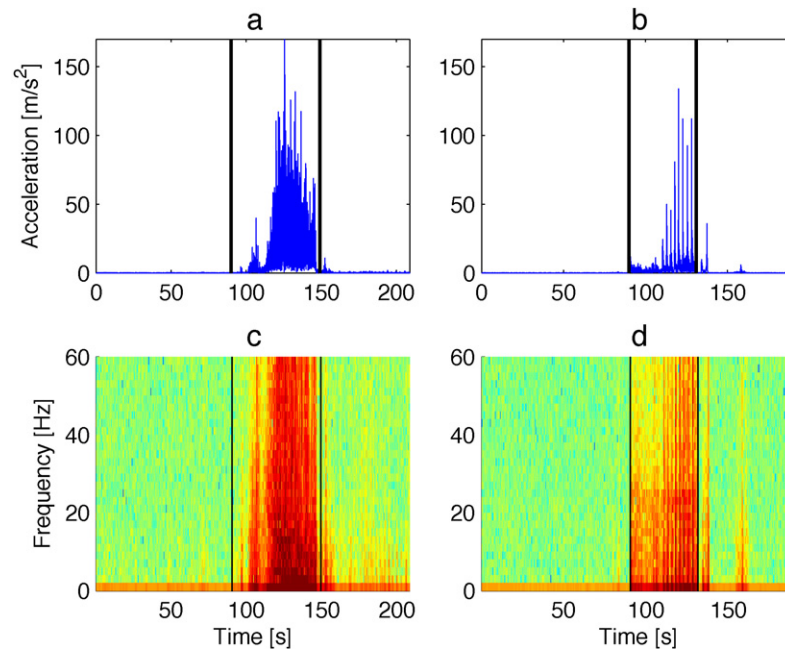hown in c and d, respectively, where the red color means high power, blue color means low. The data is from the right lower arm. The seizure and the simulated seizure are both surrounded by normal activity data, 1.5 min prior and 1 min later. The black vertical lines represent onsets and offsets of seizures and simulates seizures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

plots and spectrograms of a seizure/simulated seizure and the surrounding normal activity for the patient and the healthy subject, respectively. The data is sEMG from the right biceps brachii. The start and end of the seizure/simulated seizure is marked by the black vertical lines. For the patient it can be seen that the seizure starts prior to the muscle activity, so the first signs of the seizure was only visible in the EEG or other muscles and first a bit later did the tonic–clonic part start in the biceps brachii. The starting point for the simulated seizure is defined as were the muscle activity starts. For some patients the start of a seizure might as well be when the muscle activities are started, so this will not be seen as a difference. A clear difference is the amplitude of the signals, but it should be noted that this characteristic depends among other on the strength of the subject and the thickness of the skin/fat layer between electrode and muscle. The spectrograms are made based on normalized signals (to make sure the amplitude differences will not influence our interpretation) and plotted with the same color bar. From the spectrograms it is revealed that for both the seizure and the simulated seizure the power contained in the signal is increased for all frequencies through a longer period compared to the normal activities. In a future study we might examine the higher frequencies (above 70 Hz), since the signals do not seem to be unimportant above this limit. We cannot exclude that besides the difference in amplitude there are other differences too between the sEMG signals from the seizure and the simulated seizures.

Figs. 3 and 4, which show the raw ACM and ANG data from the right forearm, respectively, for both the patient and a representative healthy subject, are visually similar.

The amplitude, however, is also a problem for these modalities. The movements during the real seizure seem to have larger acceleration and angular velocity and furthermore both movements seem to be more confounded in the real seizure, whereas most of the healthy subjects have lower accelerations and angular velocities especially. There are though also differences among patients with epilepsy and the one we recorded from may have had faster movements than the average patient. We will have to trust that the acceleration of the simulated seizures were similar to real ones seen visually, when looking at the healthy subjects, since this was what the physician concluded during the simulations. The spectrograms show that the real seizure has a larger power in the higher end of the frequencies, than the simulated seizure, but the simulated seizures do though show a higher power in some frequencies (above 15 Hz for the ANG signal and above 1 Hz for the ACM signal) than the normal activities. These spectrograms are as well as for the sEMG generated based on the normalized signals to avoid power differences based on the amplitude of the signal. There are smaller differences in the frequencies between ACM and ANG that seems to give some useful complementary features to our algorithm.

The movements simulated by the healthy subjects visually (when looking at the healthy subject) closely resembled those occurring during the seizures, therefore it is reasonable to assume that the signals recorded by the motion sensors are similar to what we would have recorded from patients with epilepsy. This is also what we observed when we compared the data from the simulated seizures to the real one, though with some differences in the strength of the seizures. How-
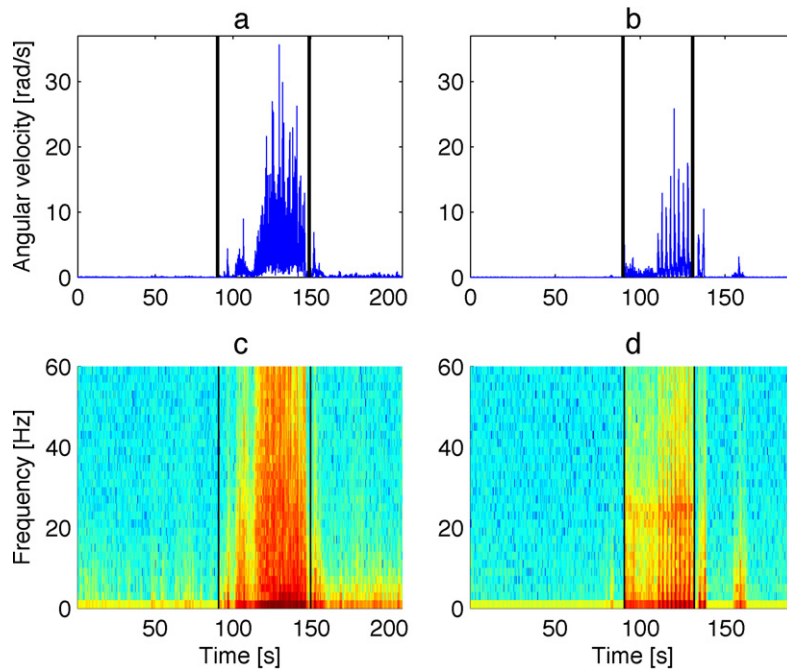
Fig. 4 – The ANG data for a real seizure and a representative simulated seizure are shown in a and b, respectively. The matching spectrograms (for a normalization of the signals) are shown in c and d, respectively, where the red color means high power, blue color means low. The data is from the right lower arm. The seizure and the simulated seizure are both surrounded by normal activity data, 1.5 min prior and 1 min later. The black vertical lines represent onsets and offsets of seizures and simulated seizures. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

ever, these differences make the real seizure stand out even more from the normal background activity, suggesting that the algorithm might work even better on the real seizure data than on the simulated ones.

4. Method

The method for detection of seizures based on multi-modal data is split in several steps as outlined in Fig. 5. The first step is to extract appropriate features and the second is to classify the data based on these features. Prior to these steps it is, however, necessary to take a look at the data and how it may

be divided into training and test sets for the classification. All of the signal processing is performed in MATLAB 7.6.

4.1. Data partitioning

Data are partitioned due to the fact that during the recordings, for practical reasons, all simulated seizures, were simulated within a short time with the healthy subjects practicing the simulations in between. It is therefore not possible to make a causal spilt of the data into training and test periods, where the first part would be used for training and the last part for test.
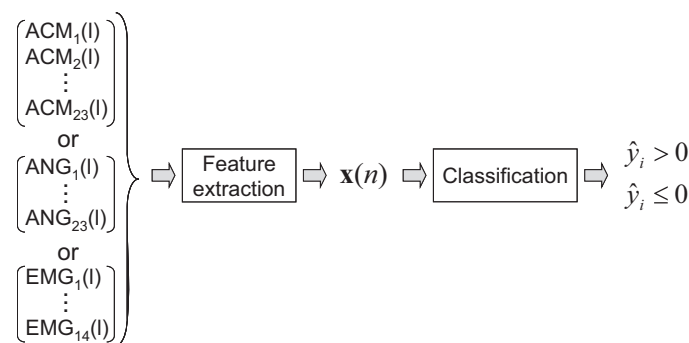

Fig. 5 – Method for detection of simulated seizures based on multi modal data. Three types of data are used, from which features are extracted. The feature vector is sent through to a classifier, which outputs y_i. A positive y_i classifies as a simulated seizure, whereas a negative y_i belongs to the normal activity class.

Fig. 6 – Diagram of the segmentation of the data strings from files containing simulated seizures. Between each segment a period of 5 s of data is left unused. Each normal activity segment lasts 1 min.

For the classification, both a training set and a test set of data is needed. Therefore data is divided into smaller segments and split randomly into the training and test phases. By splitting data in smaller segments of simulated seizures and normal activity data there are more segments to choose from, when randomizing the training and test phases related to the classification. This ensures that both the training and test phases contain segments from all the different activities performed. For each healthy subject several files are processed. A file containing seizures is divided in subparts as shown in Fig. 6, where the data parts between the simulated seizures are left unused, since, as earlier mentioned, the healthy subjects might have been practicing for the simulation of simulated seizures in between the actual simulations. The simulated seizures are split in separate segments as to contain each simulated seizure as a whole. The data period after the simulated seizures is split into segments of 1 min. This length ensures that the movements within the segments make sense, and that a sufficient number of segments are obtained for training and test. Between each segment a sequence of 5 s is left unused to reduce the correlation between two successive periods as much as possible without too much loss of data. A file without simulated seizures is treated in the same way as the period following the simulated seizures. The file is split into segments of 1 min, with 5 s sequences left unused between each–just as explained above.

4.2.   Preprocessing

With 14 sEMG channels and 69 channels of ACM and ANG, respectively, we have 152 channels in total. In order to decrease this number and thereby the computational load regarding the feature extraction, for ACM and ANG we used the length of the direction vector instead of the three dimensional (3D) coordinates, x, y and z (e.g. for ACM):

$$ACM = \sqrt{ACM_x^2 + ACM_y^2 + ACM_z^2} \tag{1}$$

Naturally since we only have one signal for each sEMG electrode, the raw sEMG data will be used.

This preprocessing of data leaves us with 60 channels (preliminary features) of data in total.

4.3.   Feature extraction

In classification problems the choice of features is often more important than the choice of classifier [15], since the features outline the details to discriminate between two groups,

whereas one classifier might provide a similar result as another, based on the same set of features.

The features for discriminating between *simulated seizures* and *normal activities* should therefore be chosen based on how well they distinguish between the two groups. Based on a visual inspection of data Nijsen et al. [11] found that a wavelet decomposition with the fifth Daubechies as a mother wavelet was the most appropriate feature compared to the STFT for ACM data. Consequently, we have decided to use the fifth Daubechies as a mother wavelet for our data; ACM as well as sEMG and ANG. Compared to the STFT where a signal is split in sine functions with different frequencies, the wavelet transformation divides the signal into shifted and scaled versions of a mother wavelet. The discrete wavelet decomposition is basically two filters that are applied sequentially to the input signal again and again (one time for each step), the filters are composed as low- (g) and high-pass (h) filters based on the mother wavelet. From each filtration an approximation (A) and a detail (D) signal is achieved. Each approximation signal can be further filtered into a new level with both an approximation and a detail signal, see Fig. 7. The black squares mark the division by the DWT, whereas the WPT is demonstrated by all squares, where also the detail signals are filtered. A mother wavelet is defined by a scaling function $\varphi(x)$ and a wavelet function $\psi(x)$ [16], described by the low-pas filter, g, and the high-pass filter, h [17]:

$$\varphi_{j,m}(l) = 2^{j/2} \cdot g_j(l - 2^j m) \tag{2}$$

$$\psi_{j,m}(l) = 2^{j/2} \cdot h_j(l - 2^j m), \tag{3}$$

where j is the resolution or scale parameter, m is the translation parameter and the inner product normalization is described by $2^{j}/2$. The decomposition is then described as the discrete approximation, $A_j(m)$, and detail, $D_j(m)$, signals given by [17]:

$$A_j(m) = u(l) * \varphi_{j,m}(l) \tag{4}$$

$$D_j(m) = u(l) * \psi_{j,m}(l), \tag{5}$$

Each window of each channel (ACM, ANG or sEMG) is applied in the wavelet transformation as $u(l)$. By the extension of the DWT to further filtering on each detail signal as well, the WPT is, as stated above, obtained. Thereby the signal is split up in uniform frequency bands with equal frequency and time resolutions for all frequencies. This means that no matter which frequency band shows the largest difference between simulated seizures and normal activity in the movement data, an appropriate resolution is achieved for both time and frequency. So a good time resolution is not compromised by a bad frequency resolution and correspondingly a good frequency resolution is not compromised by a bad time resolution.

Each DWT and WPT is determined from a window of 0.75 and 1 s, respectively, both with an overlap of 50%. The windows should be short enough to capture the important details of the seizures and at the same time, long enough to keep a good frequency resolution. The window lengths are chosen based on the results of our assessment of the optimal value for the two methods (DWT and WPT), respectively. Before the
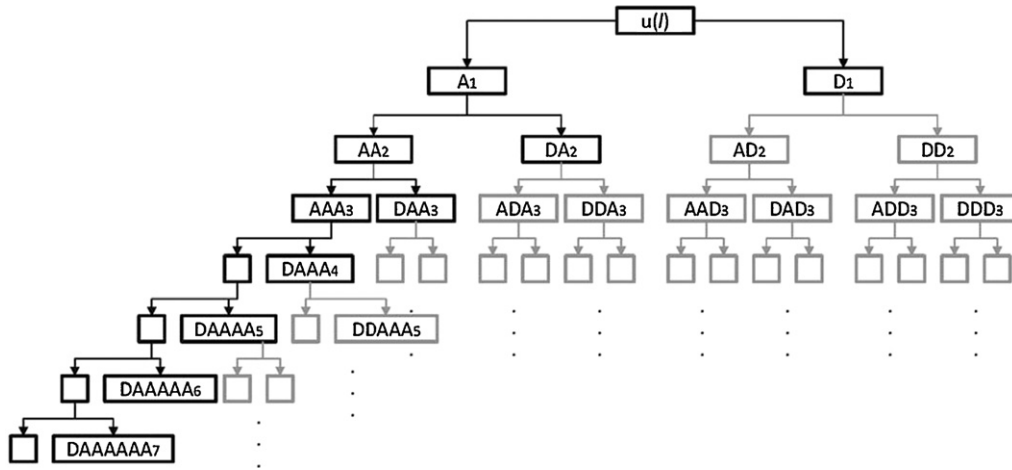
Fig. 7 – The signal, *u*(*l*), is filtered and thereby split in approximation and detail signals. The scheme with the black boxes shows the decomposition with a normal wavelet, whereas the total scheme shows the decomposition with wavelet packets. The decomposition is in both cases made to level 7 (seven layers). According to this scheme the detail bands we use for the DWT would be named: DAAA$_4$, DAAAA$_5$, DAAAAA$_6$ (and DAAAAAA$_7$ for sEMG signals). These names are long, which is why we use the short terms instead: D$_4$, D$_5$, D$_6$ (and D$_7$ sEMG signals).

windows are divided in approximation and detailed signals, they are filtered by multiplying a Hann window of the same length as the signal window to smoothen the spectrum. All feature extractions are processed in MATLAB with the Wavelet Toolbox.

4.4.   DWT *feature extraction*

The DWT can be made with an optional number of layers. We found that for the sEMG signals with a sampling frequency of 1024 Hz it would be most efficient to use 7 layers, whereby the last bands had a resolution of 4 Hz. For the ACM/ANG signals we found that 6 layers were to be used, whereby the last band had a resolution of ~1 Hz. From a visual inspection of the features extracted from the different bands in the 7 (6) layers, the detail signals layer 4–6 (ACM/ANG signals) and 4–7 (sEMG signals) turned out to provide larger differences (for the log-sum/energy parameter introduced below), when comparing the simulated seizures to randomly chosen normal activi-

ties. For the ACM/ANG signals the frequencies extracted are 0.94–7.5 Hz and for the sEMG signals they are 4–64 Hz.

To evaluate these signals and decrease the amount of data entering the feature vector we are interested in a measure for each signal indicating how much "energy" they contain. This can be evaluated by calculating a "log-sum" measure of the signals as shown in Fig. 8 and given in (6):

$$x_{j-3} = \log \left(\sum_{m=1}^{L/2^j} |D_j(m)| \right), \tag{6}$$

where L is the number of samples in the signal $u(l)$, j is the resolution (4, 5, 6 for ACM/ANG and 4, 5, 6, 7 for sEMG) and $D_j(m)$ is the detail signal. By applying the logarithm, it is ensured that the smaller differences between feature vectors from different classes are enhanced, while the larger differences between feature vectors are reduced. The influence on the system by possible outliers is thereby reduced. This means that the sys-
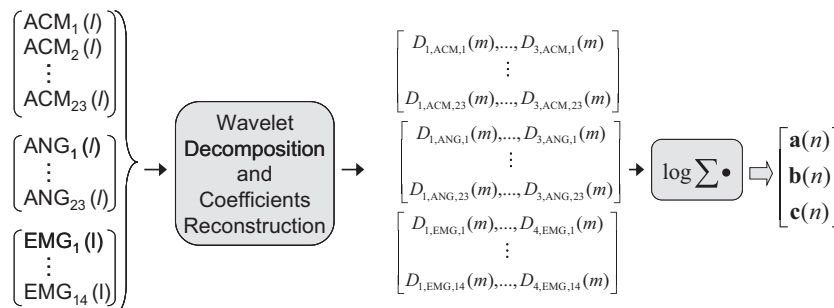
Fig. 8 – Flowchart of the feature extraction from Fig. 5. One window of data is analyzed at a time. *l* is the sample number. The chosen sub-bands are reconstructed, which for the DWT are D$_4$, D$_5$, D$_6$ and D$_7$ (only for sEMG signals). For the WPT the sub-bands used are DDA$_3$ and ADD$_3$ for the ACM and ANG signals, whereas AAAAD$_5$ and DDAAA$_5$ are used for the sEMG signals (The names are given as illustrated in Fig. 7.). A "log-sum" measure is calculated from the used bands as input to the feature vector.

tem is assumed to be less affected by outliers in the movement signals.

The feature vector, **x**, is then collected from the vectors **a**, **b** and **c**, with three (ACM/ANG) or four (sEMG) "log-sum" measures for each data window for all channels in the different modalities:

$$\mathbf{a} = [x_{1,\text{ACM}_1}, x_{2,\text{ACM}_1}, x_{3,\text{ACM}_1}, x_{1,\text{ACM}_2}, \dots, x_{1,\text{ACM}_{2B}}, x_{2,\text{ACM}_{2B}}, x_{3,\text{ACM}_{2B}}]$$

$$\mathbf{b} = [x_{1,\text{ANG}_1}, x_{2,\text{ANG}_1}, x_{3,\text{ANG}_1}, x_{1,\text{ANG}_2}, \dots, x_{1,\text{ANG}_{2B}}, x_{2,\text{ANG}_{2B}}, x_{3,\text{ANG}_{2B}}]$$

$$\mathbf{c} = [x_{1,\text{EMG}_1}, x_{2,\text{EMG}_1}, x_{3,\text{EMG}_1}, x_{4,\text{EMG}_2}, x_{1,\text{EMG}_2}, \dots, x_{1,\text{EMG}_{14}}, x_{2,\text{EMG}_{14}}, x_{3,\text{EMG}_{14}}, x_{4,\text{EMG}_{14}}]$$

$$\mathbf{x}_n = [\mathbf{a}_n, \mathbf{b}_n, \mathbf{c}_n]^T$$

(7)

where ACM_1 means ACM channel 1 and so on and n is the time index. For convenience the time index, n, is omitted in the previous equations. The concatenation of the measures into a feature vector is shown as the last step in Fig. 8.

4.5. WPT feature extraction

As with the DWT, the WPT can be made with an optional number of layers. We used the same number of steps as for the DWT. This divides the signal into frequency bands of 4 Hz. From a visual inspection of the reconstructed sEMG signals we found the reconstruction signals that contained the largest differences between simulated seizures and normal activities. It turned out to be the second and the fourth band in the fifth step, corresponding to frequency bands of 16–32 Hz and 48–64 Hz, respectively. So it showed unnecessary to decompose it into seven steps. For the ACM/ANG data, because of the lower sampling frequency, the decomposition was made in six layers as used for the DWT. This gave frequency bands for the reconstructed signal s of 0.94 Hz. A visual inspection as described above was conducted with the result that the fourth (22.5–30 Hz) and seventh (45–52.5 Hz) band of the third step contained the larger differences between the simulated seizures and normal activities for both ACM and ANG.

As for the DWT, we calculate "log-sum" measure of the signals, as given in (8) (for sEMG data) and (9) (for ACM/ANG data):

$$x_k = \log \left(\sum_{m=1}^{2^{L/5}} |R(m)| \right), \text{ where } R = \text{AAAAD}_5 (k = 1),$$

$$R = \text{DDAAA}_5 (k = 2)$$

(8)

$$x_k = \log \left(\sum_{m=1}^{2^{L/3}} |R(m)| \right), \text{ where } R = \text{DDA}_3 (k = 1),$$

$$R = \text{ADD}_3 (k = 2)$$

(9)

where L is the number of samples in the signal $u(l)$ and $R(m)$ is the reconstructed signal for the given sub-band. As earlier explained, the logarithm is applied to ensure that smaller differences between feature vectors from different classes are

enhanced and the influence by possible outliers is assumed to be reduced.

The feature vector, **x**, is then collected from the vectors **a**, **b** and **c**, with two "log-sum" measures for each data window for all channels in the three modalities:

$$\mathbf{a} = [x_{1,\text{ACM}_1}, x_{2,\text{ACM}_1}, x_{1,\text{ACM}_2}, \dots, x_{1,\text{ACM}_{2B}}, x_{2,\text{ACM}_{2B}}]$$

$$\mathbf{b} = [x_{1,\text{ANG}_1}, x_{2,\text{ANG}_1}, x_{1,\text{ANG}_2}, \dots, x_{1,\text{ANG}_{2B}}, x_{2,\text{ANG}_{2B}}]$$

$$\mathbf{c} = [x_{1,\text{EMG}_1}, x_{2,\text{EMG}_1}, x_{1,\text{EMG}_1}, \dots, x_{1,\text{EMG}_{14}}, x_{2,\text{EMG}_{14}}]$$

$$\mathbf{x}_n = [\mathbf{a}_n, \mathbf{b}_n, \mathbf{c}_n]^T$$

(10)

where ACM_1 means ACM channel 1 and so on and n is the time index as described above. As earlier noted the time index, n, is omitted in the previous equations for convenience.

4.6. Final feature vectors

All possible combinations (**a**, **b**, **c**, **a** and **b**, **a** and **c**, **b** and **c** and **a**, **b** and **c**) of the three modalities are sent through the classifier, to explore which combination would be better for an alarm system. Eqs. (7) and (10) represent the combination where all data are used. The entering of the feature vector into the classifier is shown as the final step in the classification procedure (see Fig. 5).

4.7. Classification

We decided to see the problem as a binary classification problem with the classes *Seizure* and *Normal activity*. One could also have chosen to classify the simulated seizures into different groups, but in this study we wanted to examine the possibility of making one classifier for all motor seizures. The class, *Seizure*, contains different kinds of simulated seizures with motor manifestations, whereas the class *Normal activity* contains anything but the simulated seizures. The amount of data in the two classes is very different, since we have more normal activity data than simulated seizure data, which makes the SVM algorithm attractive compared to, e.g. neural network classifiers [18]. When using the SVM one can also be sure to find a global and unique solution to the classification problem (quadratic problem), compared to neural network where there are multiple local minima and thereby multiple solutions [19]. This means that one can be sure that an optimal solution is obtained using SVM. A third reason to choose SVM is that it is less disposed to overfitting, since it chooses a specific hyperplane (with the largest margins) to separate the two classes [20].

The classification is the last part in the detection algorithm. Data are divided into two groups, *training* and *test*, see Fig. 9, where the classifier is trained on the data from the *training* group. The data from the *test* group can then be classified with
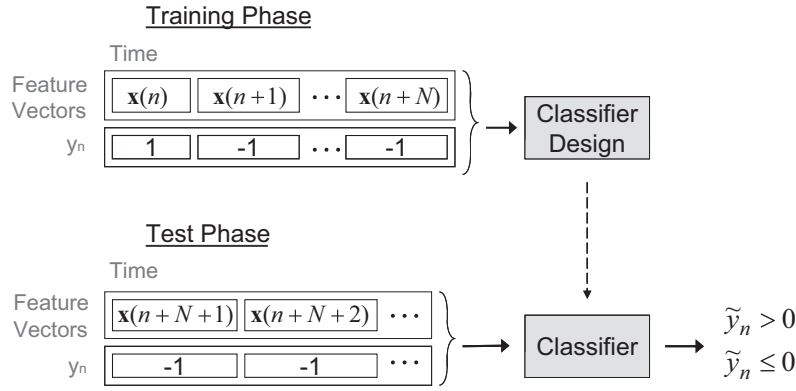
Fig. 9 – The classification part of the algorithm is split is two; the training and the testing phase. During the training phase the classifier is trained on feature vectors and their corresponding target (−1 (*normal activity*) or 1 (*simulated seizure*)). In the testing phase the "new" data is classified as *simulated seizure* or *normal activity*.

the classifier trained for the purpose. The classifier will return a positive or negative value for each *test* vector, dependent on whether it is classified as a simulated seizure or not.

The divisions of the data into these groups are made randomly, for both simulated seizure and normal activity data, ensuring close to equal amounts of each data type in each group. It is ensured that each simulated seizure type is represented in both phases (training and test).

For the training, data is labeled:

$$\{\mathbf{x}_n, y_n\}, \quad n = 1, \ldots, k, y_n \in \{-1, 1\}, \mathbf{x}_n \in \Re^d, \tag{11}$$

where k is the number of training examples, d is the dimension, $\mathbf{x}_n$ is the feature vector (n is the time index) and y_n the matching target, indicating which of the classes the feature vectors belong to, −1 for normal activity and 1 for simulated seizure.

A two-class linearly separable data set (where $d > 2$) can be classified by a hyperplane described by:

$$f(\mathbf{x}_n) = \mathbf{w} \cdot \mathbf{x}_n + b = 0, \tag{12}$$

where $\mathbf{w}$ is the normal to the hyperplane and b is a shifting constant.

The hyperplane is computed based on support vectors, which are the feature vectors placed closest to the hyperplane separating the two classes. These feature vectors from the two classes must satisfy:

$$y_n \cdot (\mathbf{w} \cdot \mathbf{x}_n + b) \geq 1 - \xi_n, \text{where } \xi_n \geq 0 \forall n, \tag{13}$$

where ξ_n, a positive slack variable, is introduced to handle data, due to the fact that most classification problems are not completely separable. Data points assigned to the wrong side of the margin (defined by (13)) thereby have a penalty that increases with the distance to the margin.

To separate the two classes, the problem of finding the optimal parameters, $\mathbf{w}$ and b, can be reduced to minimizing the performance function [13]:

$$\frac{1}{2}\|w\|^2 + C\sum_{n=1}^{l} \xi_n \text{ subjected to } y_n \cdot (\mathbf{w} \cdot \mathbf{x}_n + b) \geq 1 - \xi_n, \tag{14}$$

where C is a factor which sets the trade-off between the size of the margin and the penalty of the slack variable, ξ_n [13]. From tests we found that the most optimal value of C for our algorithm is 0.8, which is used for the results presented later in this paper.

For (14) to be minimized, each term should be minimized. Minimizing the first term means maximizing the margin between the support vectors of the two classes. The second term, which encompasses the slack variable, is minimized by keeping the distance from incorrectly classified feature vectors to the margin as small as possible. When a feature vector is correctly classified ξ_n is set to 0, whereby the second term in (14) will be 0. For a feature vector correctly classified, but placed on the wrong side of the margin, ξ_n is between 0 and 1, whereas it is above 1, if the feature vector is wrongly classified. In the two latter cases the margin is attempted placed as close to these incorrectly classified feature vectors as possible in order to minimize the second term in (14).

To solve (14) Lagrange multipliers are applied and we obtain [13]:

$$L(\mathbf{w}, b, \xi, \alpha, \mathbf{r}) = \frac{1}{2}\langle \mathbf{w} \cdot \mathbf{w} \rangle + C\sum_{n=1}^{l} \xi_n$$

$$-\sum_{n=1}^{l} \alpha_n \left[y_n \left(\langle \mathbf{x}_n \cdot \mathbf{w} \rangle + b \right) - 1 + \xi_n \right] - \sum_{n=1}^{l} r_n \xi_n, \tag{15}$$

where $\alpha_n \geq 0$ and $r_n \geq 0$. Eq. (15) is then transformed from its primary form to the dual form by differentiating it with respect
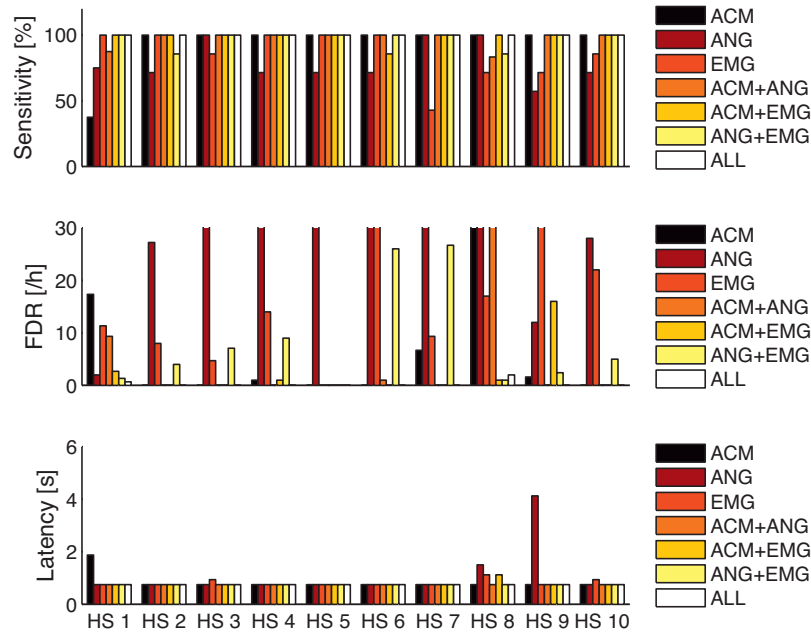
Fig. 10 – The results from the DWT feature extraction method. HS means healthy subject.

to $\mathbf{w}$, ξ_n and b and substitute the obtained relations into the primal form to obtain [13]:

$$L(\mathbf{w}, b, \xi, \alpha, \mathbf{r}) = \sum_{n=1}^{l} \alpha_n - \frac{1}{2} \sum_{n,j=1}^{l} y_n y_j \alpha_n \alpha_j \left\langle \mathbf{x}_n \cdot \mathbf{x}_j \right\rangle \qquad (16)$$

The hyperplane which separates the two classes the best could then be found by maximizing (16) with respect to $\sum_{n=1}^{l} y_n \alpha_n = 0$, $C \geq \alpha_n \geq 0$, $n = 1, \dots l$.

In (15) $\langle \mathbf{x}_n \cdot \mathbf{x}_j \rangle$ may also be written as $K(\mathbf{x}_n \cdot \mathbf{x}_j)$, where K is a kernel. If the two classes are not linearly separable a non-linear kernel may be applied. A kernel is a function which transforms a signal from one space (input space) into another space of a higher dimension, called the feature space. Thereby a linear hyperplane, separating the two classes may be found in the new feature space. In our case, we have been able to separate the classes linearly, so a non-linear kernel has been considered unnecessary.

The steps explained above are all performed in Matlab by the SVM$^{\text{light}}$ package from Joachims [21]. The package returns a classification-model based on the given training set, which can then be used to classify a test set.

4.8. Test methodology

To evaluate how well the detection algorithms function, certain measures may be calculated for each healthy subject. The test measures used in this article are:

- *Sensitivity (SEN)* is the fraction of seizures that are correctly classified.
- *Latency (LAT)* is the time from seizure start to the detection.
- *False detection rate (FDR)* is the number of falsely detected simulated seizure onsets per hour.

When the content of a window is classified as a simulated seizure an alarm will be generated. The latency is measured as the delay from simulated seizure start till the alarm is generated (first window with a positive outcast). This means that the shortest possible latency will correspond to the length of the window (0.75 s for the DWT method and 1 s for the WPT method, respectively). Only the first window, in a row of successive detections, will generate an alarm. This means that when successive *normal activity* windows are detected as a *simulated seizure* only the first one will generate a false alarm, and thereby it will only count for one FP.

The FDR is a better measure than the often used specificity, when evaluating results on seizure/simulated seizure detection. To obtain valuable results for FDR the measurements should contain several hours for testing. For practical reasons we only measured for 1.5–3 h for the healthy subjects, which may influence our results.

5. Results

The results vary depending on the feature extraction method. The results for the DWT method are shown in Fig. 10, whereas the results for the WPT method are shown in Fig. 11. To compare the results of the different combinations of modalities, the median and 95% confidence level of all results (both methods) are given in Table 2.

5.1. DWT method

The DWT method (Fig. 10) shows an almost perfect result, when combining all modalities, while the detection is less accurate when only one modality is used. For the first healthy subject (HS) the worst result is seen, when only the ACM data is included, whereas the worst result for healthy subject 2, 4,
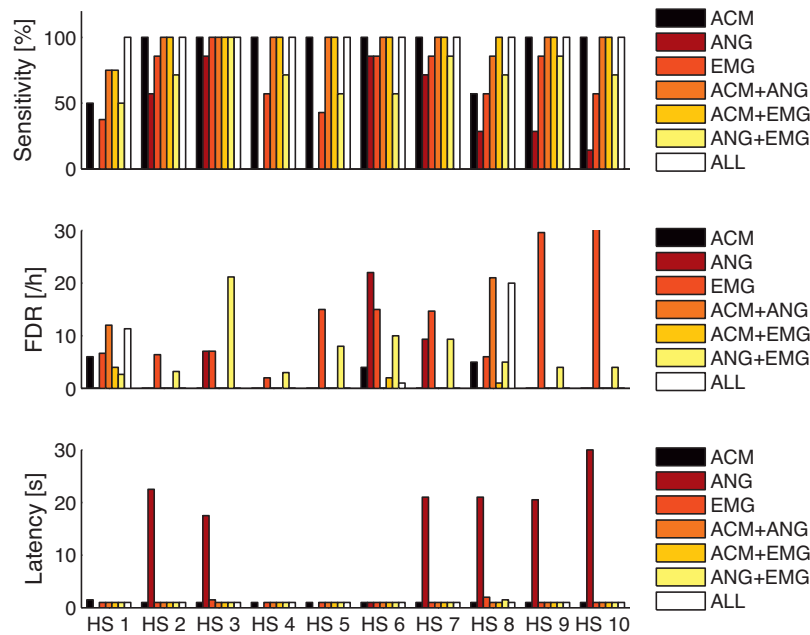
Fig. 11 – The results from the WPT feature extraction method. HS means healthy subject.

5, 6, 9 and 10 are seen when only the ANG data is used. For the last three healthy subjects (3, 7 and 8) the worst result is achieved when only the sEMG data is used. The latencies are short for all tests; the longest latency is seen for the ninth healthy subject, where only the ANG data is used. The FDR shows that for several tests the number of false detections per hour is truly high. For 7 of the 10 healthy subjects the highest FDR is observed when only the ANG data is used for the remaining three it is when only sEMG or ACM is used. When all modalities are used, the FDR is though equal to 0 for 8 of the 10 healthy subjects, the last two (HS 1 and 8) have an FDR of 0.67 and 1.7, respectively. In Table the results are presented to easy compare the different combinations of modalities. This shows that the ANG modality alone performs the worst and that clearly a combination of all modalities performs the best.

5.2. WPT method

For all healthy subjects except for healthy subject 6, the accuracy is the lowest when only the ANG data is used, and for some the sensitivity is as low as 0%. For the healthy subject

6 the results are the worst when the ANG data is combined with the sEMG data. For all healthy subjects the best results are obtained, when all modalities are combined. The latency is seen to be short for all tests except for healthy subject 2, 3, 7, 8, 9 and 10, when only the ANG data is used. The FDR is as low as 0 for about half of the tests, for a few it is as high as 30, and for the rest the FDR is around 10. When all modalities are used for eight of the 10 healthy subjects it succeeded in keeping an FDR of 0, but for the remaining two (HS 1 and 8) the FDR is 11 and 18, respectively. Looking into Table 2 it is seen that the ANG modality alone performs the worst with a too low sensitivity and much too high median latency. A combination of all modalities is shown to provide the best results.

5.3. Comparison

The results for the healthy subjects on multi-modal data (sEMG, ACM and ANG) clearly show that the algorithm performs better when all three modalities are used (see Table 2). This is independent on whether the DWT or the WPT feature extraction method is applied. From Table 2 it is clearly seen

Table 2 – Median values (and in parentheses the 95 confidence level). SEN: sensitivity; FDR: false detection rate; LAT: latency.

	Discrete wavelet transform						Wavelet packet transform					
	SEN [%]		FDR		LAT [s]		SEN [%]		FDR		LAT [s]	
ACC	100	(52–100)	0.5	(0–70)	0.75	(0.75–1.6)	100	(52–100)	0	(0–5.8)	1	(1–1.4)
ANG	71	(60–100)	49	(4.3–77)	0.75	(0.75–3.5)	29	(0–86)	0	(0–19)	19	(1–28)
sEMG	93	(49–100)	13	(1.1–50)	0.75	(0.75–1.1)	71	(39–97)	11	(2.9–33)	1	(1–1.9)
ACC, ANG	100	(82–100)	0	(0–38)	0.75	(0.75–0.75)	100	(77–100)	0	(0–19)	1	(1–1)
ACC, sEMG	100	(89–100)	0	(0–13)	0.75	(0.75–1.0)	100	(81–100)	0	(0–3.5)	1	(1–1)
ANG, sEMG	100	(86–100)	4.5	(0.2–27)	0.75	(0.75–0.75)	71	(52–97)	4.5	(2.7–19)	1	(1–1.4)
All	100	(100–100)	0	(0–1.7)	0.75	(0.75–0.75)	100	(100–100)	0	(0–18)	1	(1–1)

that, when all modalities are used the two methods provide similar results, with the only exceptions of the latency, where the lower bound is dependent on the window length, and the FDR which has a wider 95% confidence range for the WPT method. This difference is caused by two healthy subjects (1 and 8) who are seen to have much larger FDR in Fig. 11, than in Fig. 10, when all modalities are used. Besides from using all modalities it is difficult to say which method provides the best result. It depends on the individual subject and whether the sensitivity or the FDR is the most important. The DWT gives the highest sensitivity for all, whereas the WPT provides a lower FDR.

6.   Discussion

The best results, for distinguishing between simulated seizures and normal activities based on the two wavelet methods, are clearly obtained when all three modalities are included. However, if the number of modalities or sensors/electrodes could be reduced, without worsening the results too much, it would be preferable considering the usability for the patients.

The ANG modality alone is not useful, but the best results are obtained when it is combined with the ACM and the sEMG modalities. For both methods it would though be the ANG modality that would be eliminated, if one wanted to base a system on only two modalities, since the combination of ACM and sEMG, show the next best result for both methods.

Based on the results it seems evident to combine all three modalities, but it does not allow us to determine which wavelet method is the best. Beside the different approaches, DWT and WPT, it should also be noted that the two are based on different frequencies, so this might as well influence the results. When we examined which frequency bands gave the largest differences between simulated seizures and normal activities, it resulted in different bands for the two methods. Based on that examination we ended up investigating different frequencies for the two methods. We expect that future tests on patient data will reveal which wavelet method is preferable for a final detection system. Also, as mentioned earlier, we chose to look at the classification problem as binary, but when real seizures are collected it might improve the results further if the seizures are split up in different groups, dependent on the type. Furthermore we will focus on using more patient-friendly measuring equipment, suitable for long term monitoring.

The prime limitation in getting the adequate patient data was the way the motion sensors were attached to the patient. They were placed in pockets of a specially designed suit. Wearing this suit did not constitute a problem for the well functioning patients. However, these patients rarely have generalized tonic–clonic seizures. The patients who frequently have this seizure type are typically mentally retarded, and they could not tolerate the suit.

The aim of our study was to determine whether an algorithm for seizure detection based on multimodal data can be developed and further which combination of the modalities that would perform the best. As our results on healthy subjects who could tolerate the suit are encouraging for using a combination of all three modalities, it is worthwhile to focus on further development of a sensor setup, which could be tolerated by the patients. Fewer and smaller sEMG electrodes and/or motion sensors, attached to the patient, with wireless communication could solve this problem. To make such a change it would be helpful to investigate which places on the body that are more suited to wear these sensors and with how few sensors and/or electrodes is it possible to achieve an acceptable result.

7.   Conclusion

The automatic MISA system implemented offers a new approach for use of movement data with feature extraction from discrete wavelet components or wavelet packet components combined with an advanced classifier, to detect epileptic seizures. Based on the present studies both feature extraction methods provided equally promising results, and for both the best result was obtained for all healthy subjects, when combining all modalities. Future studies are needed to reveal which feature extraction method is the best choice for patient data.

Our data show the superiority of the multi-modal approach as compared to a uni-modal approach, especially compared to the ANG modality alone. At the moment, the device is a prototype for research use only. We have experienced that some patients feel uncomfortable wearing the suit containing the sensors. The superiority of the multi-modal results encourage us to develop a more patient-friendly multi-modal equipment containing the sensors, suitable for long term monitoring of the patients with epilepsy. It is convenient to base this new equipment on knowledge of which sensor/electrode placements alone or combined can provide acceptable results, with the presented algorithm. The next step is therefore to test the algorithm on a feature vector containing fewer sensor/electrode places. It is furthermore important to research on improving the biomedical signal processing for either a patient specific or a patient generic system using our multi-modal seizure onset detection approach.

Conflict of interest statement

The authors declare that they do not have any conflict of interests.

involvement, than the sponsorship. The financial support is greatly appreciated.

REFERENCES

[1] L. Nashef, P. Ryvlin, Sudden unexpected death in epilepsy (SUDEP): update and reflections, Neurologic Clinics 27 (4) (2009) 1063–1074.

[2] J.R. Hughes, A review of sudden unexpected death in epilepsy: prediction of patients at risk, Epilepsy and Behavior 14 (2) (2009) 280–287.

[3] K.M. Kelly, D.S. Shiau, R.T. Kern, J.H. Chien, M.C.K. Yang, K.A. Yandora, J.P. Valeriano, J.J. Halford, J.C. Sackellares, Assessment of a scalp EEG-based automated seizure detection system, Clinical Neurophysiology 121 (11) (2010) 1832–1843.

[4] A. Schad, K. Schindler, B. Schelter, T. Maiwald, A. Brandt, J. Timmer, A. Schulze-Bonhage, Application of a multivariate seizure detection and prediction method to non-invasive and intracranial long-term EEG recordings, Clinical Neurophysiology 119 (1) (2008) 197–211.

[5] A.M. Chan, F.T. Sun, E.H. Boto, B.M. Wingeier, Automated seizure onset detection for accurate onset time determination in intracranial EEG, Clinical Neurophysiology 119 (12) (2008) 2687–2696.

[6] Z. Li, A.M. da Silva, J.P.S. Cunha, Movement quantification in epileptic seizures: a new approach to video-EEG analysis, IEEE Transactions on Biomedical Engineering 49 (6) (2002) 565–573.

[7] T.M.E. Nijsen, J.B.A.M. Arends, P.A.M. Griep, P.J.M. Cluitmans, The potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy, Epilepsy and Behavior 7 (1) (2005) 74–84.

[8] G. Becq, S. Bonnet, L. Minotti, M. Antonakios, R. Guillemaud, P. Kahane, Collection and exploratory analysis of attitude sensor data in an epilepsy monitoring unit, in: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 2775–2778.

[9] K. Cuppens, L. Lagae, B. Ceulemans, S. Van Huffel, B. Vanrumste, Detection of nocturnal frontal lobe seizures in pediatric patients by means of accelerometers: a first study, in: 31st Annual International Conference of the IEEE

[10] I. Conradsen, S. Beniczki, P. Wolf, D. Terney, T. Sams, H.B.D. Sorensen, Multi-modal intelligent seizure acquisition (MISA) system—a new approach towards seizure detection based on full body motion measures, in: 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 2591–2595.

[11] T.M.E. Nijsen, P.J.M. Cluitmans, P.A.M. Griep, R.M. Aarts, Short time Fourier and wavelet transform for accelerometric detection of myoclonic seizures, in: Engineering in Medicine and Biology Society, EMBS Benelux Symposium, 2006, pp. 155–158.

[12] I. Conradsen, S. Beniczki, P. Wolf, J. Henriksen, T. Sams, H.B.D. Sorensen, Seizure onset detection based on a uni- or multi-modal seizure acquisition (UISA/MISA) system, in: 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2010, pp. 3269–3272.

[13] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods, 1st ed., Cambridge University Press, 2000.

[14] Xsens, P.O. Box 559, 7500 an Enschede, The Netherlands, http://www.xsens.com.

[15] M.F. McKinney, J. Breebaart, Features for audio and music classification, Proceedings of the ISMIR (2003) 151–158.

[16] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7) (1989) 674–693.

[17] A. Subasi, EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Systems with Applications 32 (2007) 1084–1093.

[18] C.Y. Yang, S. Yang Jr., J.J. Wang, Margin calibration in SVM class-imbalanced learning, Neurocomputing 73 (1–3) (2009) 397–411.

[19] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2) (1998) 121–167.

[20] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297.

[21] T. Joachims, Learning to classify text using support vector machines: methods, theory, and algorithms, Computational Linguistics 29 (4) (2002) 656–664.