TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

DIPLOMARBEIT

Spike Detection with Hidden Semi-Markov Event Sequence Models in EEG Data

Ausgeführt am

Institut für Analysis und Scientific Computing

Fakultät für Mathematik und Geoinformation

Technische Universität Wien

Und am

Department Safety & Security

Austrian Institute of Technology

unter der Anleitung von

Ao.Univ.Prof. Dipl.-Ing. DDDR. Frank Rattay

durch

Claudia Igler

Mohsgasse 17-19/2/22

1030 Vienna

Wien, Jänner 2013

Abstract

Epilepsy is one of the most common neurological disorders world-wide caused by excessive neuronal discharges in the brain. The affected regions of the brain and the resulting manifestations of the disease are very diverse and render diagnosis as well as treatment difficult. The main tool represents the electroencephalography (EEG) by visualizing epileptic patterns that can be either associated with a specific type of epilepsy or even show the area of origin, called epileptic focus. Patients that have a distinct epileptic focus are possible candidates for epilepsy surgery especially if anti-epileptic drugs fail to make them seizure free or result in serious adverse reactions.

As manual EEG viewing requires both, a high amount of experience and time, automated detection procedures appear to be provide valuable alternatives in this field. A short summary of the wide range of different algorithmic approaches that have been studied over the years are presented. The EpiScan module for online seizure detection with possible introduction of chirplet transform to the method is discussed in more depth. However, the main part focuses on a spike detection system based on statistical means through Hidden Semi-Markov Event Sequence Models (HSMESM). The main advantage of this approach lies with the statistical modeling of the duration probabilities of the possible states (for spikes: Rise, Peak and Fall).

During testing the algorithm provided better results than the current approach regarding detection of different spike morphologies and width. Variations in duration of the spike up to sharp waves could be recognized without adjustment of the parameter set or triggering of false alarms and even unusual shapes, like double-peaks, did not pose a problem to the HSMESM model. In addition, the ability of the HSMESM approach to detect other epileptiform EEG activities, such as polyspikes and vertex waves, was examined as well and showed promising outcomes. However, especially for polyspikes more false alarms were elicited in some cases. This issue could be addressed by adaption of the preprocessing as the windows used for building the event sequence are too large to yield appropriate events for polyspike detection.

Zusammenfassung

Epilepsie gehört zu den häufigsten neurologischen Krankheiten weltweit und wird verursacht durch exzessives Feuern von Nervenzellen im Gehirn. Die betroffenen Hirnregionen und die daraus resultierenden Erscheinungsformen der Krankheit variieren sehr stark, was sowohl die Diagnose, als auch die Behandlung erheblich erschwert. Das wichtigste Hilfsmittel stellt das Elektroenzephalogramm (EEG) dar, indem es epileptogene Muster sichtbar, welche entweder mit einem bestimmten Typ der Krankheit in Verbindung gebracht werden können oder möglicherweise sogar auf einen spezifischen Ursprungsort, den epileptischen Fokus, hindeuten. Patienten bei denen ein epileptischer Fokus ausgemacht werden kann, stellen möglichen Kandidaten für eine Operation dar, vor allem wenn Medikamente nicht zur Anfallsfreiheit führen oder schwere Nebenwirkungen hervorrufen.

Eine manuelle Durchsicht und Beurteilung von EEG Daten erfordern viel Zeit und Erfahrung, weshalb eine automatische Detektion eine vorteilhafte Alternative darstellen könnte. Im Folgenden wird ein Überblick über das große Angebot an verschiedenen Algorithmen, welche über die Jahre studiert wurden, gegeben. Das EpiScan Modul, das zur online Anfallsdetektion dient, und die mögliche Einführung der Chirplet Transformation in das Modul, werden im Detail beschrieben. Der Hauptteil beschäftigt sich aber mit einem Spikedetektionssystem das auf einem statistischen Modell, dem Hidden Semi-Markov Event Sequence Model (HSMESM), basiert. Der Hauptvorteil dieses Ansatzes besteht in der statistischen Modellierung der Dauer der möglichen Zustände (im Falle von Spikes: Anstieg, Spitze und Abfall).

In der Testphase zeigte der Algorithmus bessere Resultate als die jetzige Methode bezüglich der Detektion von unterschiedlichen Spikemorphologien und –weiten. Variation der Dauer des Spikes bis hin zu steilen Wellen konnten erkannt werden ohne dass der ursprüngliche Parametersatz angepasst werden musste oder falsche Alarme ausgelöst wurden. Sogar ungewöhnliche Formen, wie beispielsweise Doppelspitzen, stellten kein Problem für das HSMESM Modell dar. Außerdem wurde die Fähigkeit des HSMESM Ansatzes andere epileptiforme EEG Aktivitäten, wie etwa Polyspikes und Vertexwellen, zu erkennen, ebenfalls getestet und zeigt vielversprechende Ergebnisse. Speziell für Polyspikes wurden aber in manchen Fällen mehr falsche Alarme ausgegeben. Dieses Thema könnte durch Anpassung des Vorverarbeitungsprozesses behoben werden, da das Fenster, das in der Erstellung der Eventsequenz verwendet wird, zu groß ist, um geeignete Events für die Detektion von Polyspikes zu generieren.

Acknowledgement

I would like to thank Professor Frank Rattay for his contribution as a supervisor of this thesis.

Furthermore, I have to express my gratitude to Manfred Hartmann, Peter Dollfuß and Tilmann Kluge at the Austrian Institute of Technology for their willingness to help and support me during this project and who provided me with the opportunity to do this research topic.

Last but not least, I want to thank my friends and family for their constant patience and understanding which supported me during my studies.

Contents

1 Introduction

Epilepsy is one of the most common disorders of the nervous system and affects approximately 50 million people across the globe. The disease is characterized by unprovoked and recurring seizures caused by hypersynchronization of neural networks. Since this abnormal neural activity can occur in every region of the cerebral cortex – the outermost layer of the brain consisting of neuronal cells - the clinical correlates are manifold and diverse as well as in some cases even barely visible. Owing to this diversity a large number of different types of diseases or syndromes in which epilepsies are categorized exist having rather blurred distinctions at times.
The second chapter provides a brief glimpse into the complexity of epilepsy and its manifestations in addition to possible diagnosis and treatment opportunities in contemporary clinical application.

Moreover, despite decades of research the underlying cause of the disease remains elusive and is grasped only to limited extent. Therefore, diagnosis and treatment of epileptic patients remain challenging and are mostly based on experience. Accurate diagnosis of the type of disease is vital to determine the possible treatment options whereas diagnosis in turn is based on the visualization of abnormal epileptic neuronal activity in the cerebral cortex of the brain. The most important tool in this context constitutes the electroencephalography (EEG) which records action potentials in terms of electrical activity with electrodes placed on the scalp directly above the cerebral cortex.
However, EEG is not only useful for diagnostic approaches. Most epileptic patients are treated with anti-epileptic drugs and about 66% of the patients become seizure free through medication. On the other hand, about one third of those affected by epilepsy can only achieve a minimal reduction in seizures or suffer from severe adverse reaction. Epilepsy surgery constitutes a viable treatment alternative for those patients. Long-term video EEG monitoring is used to provide the best possible pre-surgical evaluation about the area from which the epileptic activity originates. If a so-called epileptic focus is found that has a circumscribed boundary and lies in a region that can be removed, epilepsy surgery augurs promising prospects.
Congenial to the versatile appearance of the disease the electrographic patterns recognized as epileptiform display a large variety of shapes as well. Analysis of EEG data is made even more intricate by the intrinsic range of EEG patterns that are considered as normal showing differences for wake and sleep states. Furthermore, technical or biological artifacts may obscure the recorded signal and are unfortunately quite common especially in long-term EEG monitoring.
In order to understand the issues EEG analyzers face – may they be human or computer-aided systems – a short overview about non-pathologic and epileptiform EEG activity is provided in this work.

The fourth chapter deals with the multifarious approaches made in automated detection of epileptic activity in EEG data. A concise summary lists the possibilities for ictal as well as

inter-ictal detection that were researched up until now. As the epileptiform patterns observed in ictal and inter-ictal periods differ, so do the considerations determining the methods for the respective algorithms. The determination of ictal phenomena or seizures again can be broken down into three tasks: seizure detection, seizure prediction and origin localization, each of them posing different challenges to researchers.
Several examples are explained in a little more depth to the end of visualizing the abstract algorithmic ideas somewhat better to the reader.

In the following the EpiScan algorithm which was developed at the Austrian Institute of Technology (AIT) is introduced and explained. It is intended to be coupled with Epilepsy Monitoring Units (EMUs) in hospitals in order to detect rhythmic epileptiform patterns in the EEG data that represent the most frequent characteristic during seizures in temporal lobe epilepsy (TLE) patients. The detection system is made up of six different modules the core of which is embodied by the periodic waveform analysis (PWA) building block. During tests with unselected data the EpiScan algorithm showed promising sensitivity while keeping the false alarm rate at an acceptable level.

The PWA module is based on a time-frequency method called wavelet transform for pattern detection. The wavelet transform decomposes the signal under analysis in so-called wavelet atoms thereby characterizing the spectrum of the data.
A study is presented investigating the advantages of replacing this method by chirplet transform. Chirplets are characterized by more parameters and thus have the ability of better adaption to the observed signal. Three different algorithms are examined for their efficiency and performance in detecting rhythmic epileptiform patterns (with or without chirp in the analyzed recordings).

The EpiScan system is intended for detection of rhythmic patterns during ictal periods whereas the purpose of the next module lies with spike detection. Spikes represent a very common epileptiform pattern that occurs in diverse shapes: individual spikes, polyspikes, spikes followed by a slow wave (called spike-wave complexes) and polyspike-wave complexes. The spike can be characterized by several features which are extracted from the original input signal (EEG records). These features are then used to classify the windowed data segments whether they contain a spike.

The characteristics used in the approach designed at AIT are based on morphological properties such as first and second derivation to determine parts of the spike called rise, peak and fall. However, the time frames for these sections are very narrow and do not allow for much variation. A possibility to introduce more variability into the duration of the spike components and the distance between them poses the use of a statistical model. In the following thesis, the Hidden Semi-Markov Event Sequence Model (HSMESM) is used as a statistical means for spike detection.

At first, the basic principles of Markov and Hidden Markov models are explained before exploring the specific type of HSMESM. After the model definition, the three computational problems faced in Hidden Semi-Markov models are detailed with emphasis on the first two of

them. The evaluation problem uses a forward-backward algorithm to determine the possibility of the analyzed observation sequence given a prescribed model, i.e. how well the model fits with the given observation. The second issue concerns itself with finding the optimal "hidden" state sequence correlating with the observed events. In this part an implementation of the Viterbi algorithm is employed. The last problem is dedicated to adjusting the model parameters to maximize the possibility of the observation sequence with the given model. The Baum-Welch algorithm uses an iterative procedure to find the optimal parameters at least locally.

Last but not least the results and conclusions obtained by the use of HSMESMs on EEG recordings of several patients and different types of spikes are illustrated. The method proved to be very promising and will be implemented in the current spike detection algorithm in the near future.

2 Epilepsy

Epilepsy is a neurological dysfunction of the brain which is caused by excessive, synchronized discharge of a group of neurons - which comprise a neural network - in the cortex. The disease is characterized by recurring, spontaneous and unpredictable seizures which are denoted by transient alterations in the perception or the behavior of the affected person.

With a prevalence of about 1-2% of the world's population Epilepsy is among the most abundant chronic diseases and dates back longer than the emergence of human beings as scientists were able to prove that every brain with enough complexity, e.g. that of apes, can display epileptiform discharges.

There exist many descriptions of human beings suffering from epileptic seizures in ancient history and over the course of time they were either regarded as being holy and in association with God or possessed by the devil. The first one to recognize the seizures as symptoms of a disease and to study it was Hippocrates 430 BC. In greater depth research of epilepsy started in the 17th and 18th century and many terms like the names of syndromes are still in use nowadays. The most important step, leading to a whole new understanding and possibilities for treatment of the disease, was the invention of the EEG in 1924 by Hans Berger. Over the last hundred years several pharmaceutical compounds were discovered that result in the reduction or even abscondence of epileptic seizures. However, only about two thirds of epileptic patients can be treated with antiepilectic drugs and although plenty knowledge was gathered in the last decades, the genesis of epilepsy still remains elusive. The probability of the occurrence of epileptic seizures is the sum of an inherent predisposition to epilepsy that is often a hereditary trait and past brain injuries like lesions or scars in the brain tissue.

Ordinary action potentials spread along the cell surface by propagating the voltage change along the membrane. If the summed voltage of the synapses of the nerve cell (dendrites) reaching the axon hill cell is large enough, a new action potential is created and propagates along the axon to other cells like brain or muscle cells. Other nerve cells can either be excited or inhibited by the short impulses.
During epileptic seizures however long voltage impulses originate in the cell body. They are translated into a train of maximal possible short impulses in the axon hill due to the refractory period, which results for example in muscle jerks. Additionally, cells involved in epileptic discharges are able to produce so-called pacemaker potentials without involvement or stimulation from other cells, which means they need no external stimulus to fire.
The functional activity of the brain is regulated by electrochemical signals of neurons to other cells and their responses to the signals. Certain areas of the brain have been identified as being responsible for specific cognitive functions for instance Broca's area and Wernicke's area which are in charge of language processing.
Epileptic activity disturbs and disrupts the regular function and the visible symptoms are correlated with the affected region in the brain. Epileptic discharges affecting the visual

processing centre in the occipital lobe can for example yield hallucinations or light flashes. If nearly the whole brain is encased in the epileptic activity the patient loses consciousness.

2.1 Classification

It is essential to differentiate between the terms epilepsy and epileptic seizure in the sense that epilepsy stands for the disease and epileptic seizure refers to the nature of the phenomenon or symptom. In addition, the concept of epileptic syndrome is used; however, there is a certain amount of discrepancy about the difference between syndrome and seizure type. The first international classification was proposed by the International League against Epilepsy (ILAE) and was revised several times since.
Seizures are dichotomously classified into focal or generalized. An additional category labeled unknown contains for example epileptic spasms as they exhibit generalized as well as focal manifestations. A generalized seizure is defined as “arising in and rapidly engaging bilaterally distributed networks”, whereas focal seizures “originate within networks limited to one hemisphere”.

In the latest edition focal seizures are not further differentiated but it is essential to note if the seizure does or does not impair the awareness of the patient. Moreover, initially focal seizures can develop into secondarily generalized ones.

Generalized seizures on the contrary are arranged into six different groups:

- Absence seizures: They are denoted by a short pause in awareness of the patient. This seizure type is often referred to as petit-mal.
- Atonic seizures: They are characterized by the sudden loss of muscle tension and are therefore often accompanied by a fall.
- Clonic seizures: Indications of this type are uni- or bilateral, rhythmic jerks of arms and legs caused by repeated alternating contraction and relaxation.
- Myoclonic seizures: They can be identified by the occurence of brief twitches of one muscle or a group of muscles.
- Tonic seizures: The muscle tone is greatly increased and leads to the stiffening of the body, arms and/or legs. This seizure type is often seen during sleep; however if the patient is awake falls are usually observed.
- Tonic-Clonic seizures: As the label indicates the seizure is designated by muscle stiffening followed by rapid, rhythmic jerks. Loss of consciousness is usually a result of this seizure type.

The classification of the disease epilepsy is divided in epileptic syndromes and nonsyndromic epilepsies. Approximately more than the half of epileptic patients can be diagnosed with a specific syndrome. The occurring syndromes can be most easily divided into groups by regarding the typical age at onset of the first seizures, e.g. West syndrome at some point in infancy or Lennox-Gastaut syndrome during childhood.

The nonsyndromic epilepsies are grouped into four classes: idiopathic focal or generalized and symptomatic focal or generalized. Symptomatic refers to epilepsies that can be associated with an underlying cause like a brain lesion whereas idiopathic means of unknown cause. As there are also types of epilepsy that cannot clearly be defined as either focal or generalized an additional group labelled undetermined epilepsies was added.

It is essential to note that seizures that can be associated with causes like fever, alcohol, drugs, acute injuries or sleep deprivation are not assessed as epilepsies. Equally, isolated seizures without recurrence do not classify as epilepsy even if the cause remains undetermined.

2.2 Diagnostic analysis

The diagnosis starts with the medical history of the patient and places special attention on the presence of epilepsy in other family members and pre-existing conditions that could cause symptomatic epilepsies.
The next step comprises a physical examination and laboratory tests with particular care to the nervous system and unbalances in nutrients or hormones.

The most important diagnostic tool in determining epilepsy is the electroencephalogram (EEG). Usually, during the routine EEG recording, lasting about 20-30minutes, hyperventilation and intermittent photic stimulation are applied to provoke a seizure. The EEG in epileptic patients shows abnormal activity during ictal periods (seizures) as well as inter-ictal. Therefore, EEG signature can be deployed to determine the type of seizures and the location and spread of epileptic activity. To get more accurate information about the genesis of the disease long-time EEG monitoring, often coupled with video recording of the patient, is necessary in most cases.

Psychogenic non-epileptic seizures are the most common differential diagnosis of epileptic diseases. However, they are not of organic but of psychological nature and the only way to differentiate them from epilepsies is through long-time EEG.

Additionally, especially in presurgical examinations magnetic resonance imaging (MRI) or functional magnetic resonance imaging (fMRI) is employed to determine the epileptic focus and to weigh the possible risks of surgery.

2.3 Treatment

As a result of the therapy the patient should become completely seizure free with acceptably few or no adverse reactions. If the onset takes place at during childhood or infancy the physical and cognitive development of the child should be unaffected.
Even after decades of research the evolution of epileptic seizures is largely unknown and mostly the therapy consists of anti-epileptic drugs that have proven effective by experience.

The drugs can only prevent seizures by influencing receptors or ion channels but they cannot eliminate the cause of the disease. If the patient has been seizure free for at least two years the gradual reduction and even termination of the drug dose can be considered and stands a chance of success of about 70 per cent. This led to the question if epileptic seizures can be learned or reversely forgotten. The fact that repetitive weak excitation ("kindling") often leads to epileptic activity supports this theory. However, conclusive prove for or against this speculation could not be supplied up until now.

However, only about 65% of epilepsy patients respond to the treatment with anti-epileptic drugs. Depending on the type of epilepsy, alternative treatment possibilities for medically refractory patients or patients with severe adverse reactions caused by the drug therapy include nerve stimulation (e.g. vagus nerve stimulation) and epilepsy surgery, which is only applicable in patients with distinctly localized epileptic areas in parts of the brain which can be surgically altered or removed without harming the patient's brain function. Surgical intervention is only possible if the epileptic seizures always originate in the same spot of the brain. The probability to achieve entire seizure freeness through epilepsy surgery varies between 50 and 80 per cent.

To guarantee the best possible outcome of the surgery the epileptic area has to be localized as accurate as possible to prevent the patient from brain dysfunctions. The most important tool in presurgical diagnostic is the application of long-term EEG or electrocorticography (ECoG). ECoG or intracranial EEG (iEEG) describes the deployment of electrodes directly on the exposed brain surface and usually requires a surgical procedure. The advantage of employing ECoG lies in the reduction of artifacts like movements and in the increased spatial resolution, because the recorded signal does not undergo attenuation by the skull.

Therefore, the proper and accurate interpretation of EEG or ECoG data is essential in presurgical diagnostic but does prove to be rather tricky and time-intensive due to the great variety of normal and abnormal EEG patterns and the large amount of data, especially in long-time EEG.

3 Electroencephalography (EEG)

The EEG measures the brain's electrical activity that is constituted by nerve impulses being sent and received to and from cortical neurons. This spontaneous activity is always present, even during sleep. Different regions of the brain are responsible for specific activities, e.g. the occipital lobe for vision processing, and the entirety of these areas is called the cerebral cortex. The functions of the cerebral cortex include abstract thought, reasoning, voluntary and involuntary control of skeletal muscles and the identification and differentiation of somatic, visceral and special sensory stimuli. The cerebral cortex is covered by the cranium, bones of the skull which cover and protect brain surfaces and in turn is cover by the scalp, a thin layer of skin.

Sensory information is received through receptors in peripheral body parts and then propagated through underlying centers of the brain to the cerebral cortex. The electrical activity of cortical nerve cells can be measured by electrodes placed directly on the skull. These electrodes show the activity in the region directly underneath them but this equals the activity from thousands of neurons as one millimeter of the cortex consists of more than 100.000 neurons. As electrical potential of single neurons would be too small to be detected by the electrodes, the EEG signal represents the synchronized activity of a large number of cortical nerve cells. Due to their well-aligned spatial orientation and because they fire together pyramidal neurons are thought to contribute most to the potentials that make up the EEG signal.
The electrical voltage fluctuation results from ionic current flows across neuronal membranes that are associated with the creation of action potentials. The most important extracellular current flows with regard to EEG signals are the synaptic potentials which can be either excitatory or inhibitory. They govern most of the waveforms seen in EEG as they have a longer duration than action potentials.
The standardized 10-20 electrode system uses anatomical landmarks on the scull and the distance between them is subdivided by 10% or 20% intervals. For clinical purposes a minimum of 21 electrodes is used, but an increase of the number to e.g. a 10-10 system is possible for more detailed studies. The electrodes have certain designations: Fp (frontopolar), F (frontal), P (parietal), T (temporal), O (occipital) and C (central). An additional z denotes placement on the middle line, whereas odd numbers are used for the left and even numbers for the right side.

In clinical application two different montages are used: referential or bipolar montage. In the referential montage the potential difference between each active electrode as initial input and one "neutral", passive electrode is determined. The advantage is that this measurement depicts the absolute voltage amplitude and the point of maximal amplitude can be identified as the place of origin.
In the bipolar montage two active, adjacent electrodes are compared and sites of maximal negativity or positivity are marked by a phase reversal.

3.1 Normal EEG signals

Normal rhythmic EEG activity can be divided in six simple waveforms that can be distinguished by frequency and amplitude.
Delta waves have the lowest frequency (<4Hz) but high amplitudes, ranging from 20 to 200µV. It appears during slow wave sleep in adults and babies as normal EEG pattern.

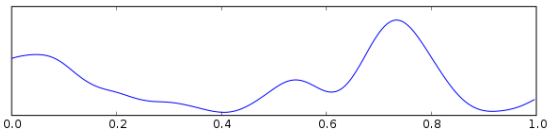


Fig. 3.1. Delta waves

Theta waves range from 4-8Hz with an amplitude of about 10µV and are also perceived in young children or during drowsiness or arousal in adults. They are also associated with idling or the inhibition of elicited emotional responses.

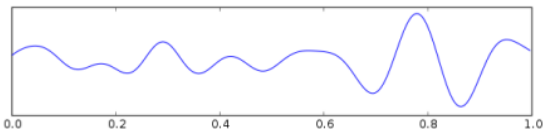


Fig. 3.2. Theta waves

Alpha waves have a frequency of 8-13 Hz and amplitude of 20-200µV. This pattern is most prominent in adults that are in a relaxed state with their eyes closed and has the highest amplitudes in the occipital region.

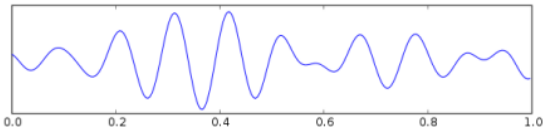


Fig. 3.3. Alpha waves

Mu waves are in the same frequency range and denote synchronized firing of motor neurons in the resting state.

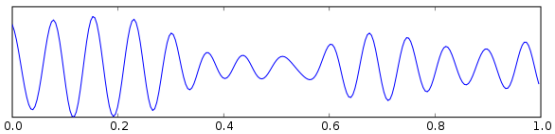


Fig. 3.4. Mu waves

Beta waves range from 13 to 30Hz and have amplitudes of 5 to 10µV. They are related to an active, busy, alert state and exerting mental activity as well as to REM (rapid eye movement) sleep. The lower amplitude is due to an alpha block or desynchronization that results in the counterbalancing of positive and negative potentials.

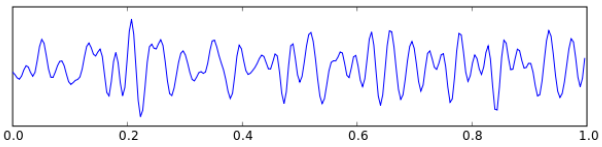


Fig. 3.5. Beta waves

Gamma rhythms have a frequency from 30 to up to 100Hz and seem to be correlated with cross-modal sensory processing.

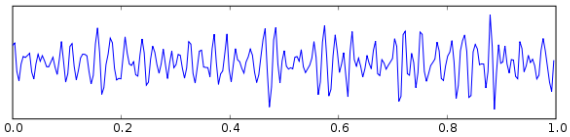


Fig. 3.6. Gamma waves

Most unusual EEG patterns that could be mistaken for abnormal activities are artifacts resulting from various extracerebral sources. Biological artifacts can be introduced through eye movement or blinks, the cardiac cycle (pulse), muscle movement or glossokinetic reasons (refers to tongue movements). To help in the differentiation of these artifacts an additional recording of EOG, ECG and EMG proves to be useful.

The eyelid movement or vertical eye movement during blinking results in a rather large potential (reaches amplitudes of mV) because the eye behaves like an electrical dipole that has a positive charge in the cornea compared to the retina. Eye movements, called saccades, on the other hand can produce spike potentials (SPs) due to electromyographic artifacts. Spike-and-wave artifacts can be provoked during photic stimulation due to EMG artifacts of the frontalis muscles of the scalp. Especially difficult to differentiate from abnormal EEG patterns are combinations of several artifacts (e.g. eye blink and muscle movement).

Artifacts that are not of a biological nature usually result from technical appliances. The most common source of disturbance is represented by the electrodes themselves and can be caused through movement or impedance changes. However, single electrode artifacts are easy to determine as they are limited to one channel in the EEG derivation. A 60Hz artifact as a result of poor grounding can obscure the EEG recording but the application of a 60Hz notch filter significantly reduces the problem. Even a simple telephone ring can cause confusion especially in long-term monitoring and often the source of the disturbance is quite elusive. Other medical devices that are attached to or even implanted in the patient like continuous positive airway pressure (CPAP) units, intravenous (IV) drips or vagus nerve stimulators are able to produce rhythmic, fast bursts of low voltage that are similar in appearance to spikes.

Obviously, in long-term monitoring the EEG signal is also recorded during sleep because many seizure types occur in the night or in the phases of awakening or drowsiness. The sleep EEG is divided into four stages of non-REM sleep (light sleep to deep sleep) and REM sleep where each stage is associated with specific patterns.

Characteristic features of stage I sleep are vertex waves (diphasic sharp waves) that are seen bilateral, synchronous and symmetrical. Moreover, attenuation of alpha rhythm, an increase in beta waves in the frontal areas, slow eye movements, vertex sharp waves and positive occipital sharp waves can be encountered during this sleep phase.
Sleep stage II shows the same characteristics as stage I but is associated with a slowing of the background frequencies and displays sleep spindles and K-complexes. Sleep spindles are defined as sinusoidal, transient 12-14Hz activity of varying amplitude. K-complexes exhibit an initial sharp wave followed by a slow wave of high amplitude (spike-wave complex) and may be connected with sleep spindles, especially in the fronto-central regions.
Prominence of slow wave sleep characterizes stages III (delta frequencies comprising 20-50% of the signal) and IV (delta is present for more than 50%).
During REM sleep eye movements, loss of muscle tone and waves resembling a saw tooth are noticeable.

Routinely performed EEG derivations include hyperventilation and intermittent photic stimulation as they often elicit abnormal responses in epileptic patients.
Hyperventilation leads to cerebral vasoconstriction and systemic hypocarbia and evokes a build-up in the EEG signature, especially notable in the frontal regions and often of high amplitudes. This build-up consists of a bilateral increase in delta and theta waves and should resolve itself after approximately one minute.
Intermittent photic stimulation is usually made up of 1-30Hz flashes continuing over ten seconds in a distance of less than 30 cm of the patient. The induced response (called photic driving) is usually time locked to the light stimulus and is prominent in the occipital locations. Typically, myoclonic artifacts due to movements of the frontalis or periocular musculature are present in the recorded signal.

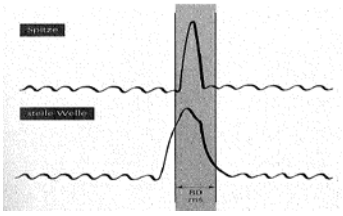
The variety of different normal patterns or seemingly abnormal patterns due to artifacts which can easily lead to overinterpretation renders EEG interpretation to be a difficult and time-consuming work.

3.2 Epileptiform abnormalities in EEG signals

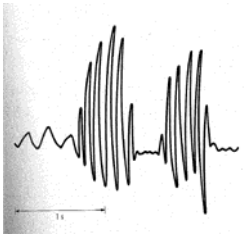
The EEG shows two different important abnormalities that aid in the detection and differentiation of epilepsies: ictal patterns (during seizures) and inter-ictal epileptiform discharges (IED). The shape and appearance of the observed abnormalities can be quite similar; however in IEDs those patterns are connected to construct transient waveforms whereas ictal discharges consist of polymorphic trains of waveforms. IEDs are displayed only occasional and short-term as opposed to discharges during seizures that last longer and exhibit a continuous form. Moreover, ictal periods are mostly accompanied by clinical correlates of the electrographic recordings.
This chapter will concern itself with the inter-ictal recordings which are easier obtained and have a higher occurrence than ictal runs. IEDs occur only rarely in patients without the expression of seizures and have been proved to be very useful in supporting clinical diagnosis

of epilepsy as well as provide relevant information for the following therapy. During routine sessions hyperventilation, photo stimulation and other methods are used to provoke epileptic seizures but it is important to keep in mind that artificially provoked seizures may show different behavior than spontaneous ones.
Scalp detected IEDs usually reflect radially oriented dipoles and although their site of expression may differ from the seizure origin, IEDs often assist significantly in the classification of the type of epilepsy or epileptic syndrome. At first differentiation into focal or generalized epilepsies can be gathered from the spread and type of IEDs. The discharges may also aid in the localization of the epileptic focus but their potential correlation with seizures varies according to the site of IED expression; e.g. frontal, anterior temporal and midline IEDs show the highest connection to seizures.
However, the absence of IEDs does not guarantee that epilepsy can be excluded as diagnosis as the EEG is based on volume conduction of potentials on the scalp level and may encounter difficulties in localizing deep-seated sources or small restricted foci as their amplitudes are easily obscured by artifacts or simply too low for detection.

Similar to seizures and epilepsies, IEDs are dichotomously classified into focal and generalized epileptic discharges.
The presence of focal interictal epileptiform discharges increases the probability of partial-onset seizures and their expression ranges from focal, lateralized and regional to secondarily generalized discharges. The morphological appearance ranges from sharp waves, spikes, spike-wave trains, polyspikes to polyspike-slow wave discharges. The polarity of spikes and sharp waves is usually negative at the surface producing a negative phase reversal. Spikes have a duration of 20-70msec whereas sharp waves have a blunter shape and last about 70-200msec. Secondarily generalized IEDs are called secondary bilateral synchrony (SBS) and characterize bilateral synchronous bursts that are often preceded by a focal “lead in” of 400msec or more.
Generalized epileptiform discharges are generically associated with generalized epilepsies and their form of expression can aid in the distinction between idiopathic and symptomatic types. They can appear with or without clinical symptoms, e.g. in absence epilepsies. The trademark of generalized epilepsies consists of bilateral, synchronous and symmetrical 3Hz spike-and-slow-wave complexes, usually with maximum amplitude in the fronto-central regions where the spike as well as the slow wave are surface negative. They are referred to as generalized spike-and-wave (GSW) and are called fast or slow if their frequency respectively exceeds or falls below 3Hz. Generalized IEDs can be of varying duration and mostly during sleep the GSWs may show fragmentation and polyspike emergence. Another form of generalized IEDs is hypsarrhythmia, usually associated with the West syndrome, and manifests itself as multi-focal IEDs in front of a high-voltage background of theta and delta waves. Generalized paroxysmal fast activity (GPFA) is typically encountered during sleep and is characterized by diffuse, bilateral burst of frequencies between 15 and 20Hz located in the frontal areas.

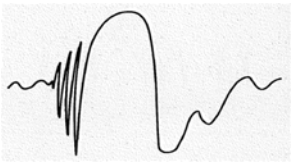


Spikes (upper signal), Sharp waves (lower signal)



Polyspikes

Spike-slow wave complex



Polyspike-slow wave complex

Fig. 3.7. Different types of epileptiform EEG patterns (Eba Homma, Leitfaden für die EEG-Praxis)

3.3 Epileptic seizure patterns in EEG signals

The appearance of seizures in the EEG derivation finally provides conclusive proof of the existence of some form of epilepsy as IEDs can also be seen in the recordings without the expression of seizures. The location and type of electrographic seizure correlates again aid in the classification of the epilepsy type, the localization of the epileptic focus and the treatment plan as well as in the prognosis of convalescence.

Similar to IEDs ictal discharges can be divided into two groups: focal or generalized seizures. In both cases the EEG may draw attention to seizures that would go unnoticed otherwise as there are no distinct clinical symptoms visible but the physical expression constitutes altered states of awareness.
Typically, focal epileptic seizures affect – at least at onset of the seizure - only a circumscribed part of the brain and the clinical symptoms are directly linked to the mental function of the epileptogenic zone. Generalized seizures on the other hand most often involve the entire brain; therefore resulting in bilateral manifestations sometimes accompanied by loss of consciousness or the passage to status epilepticus.

The morphology of focal seizures displays a wide variation of abnormal patterns in the EEG signature and depends strongly on the origin location. Prominent characteristics of ictal discharges are repetitive and rhythmic frequencies with alterations of frequency, amplitude, rhythmicity, progression and location. However, some focal ictal discharges, especially if associated with an “aura” (the patient feels the impending onset of a seizure), cannot be detected with the scalp electrodes.

If partial seizures are only expressed at one or two electrodes at onset, it is highly likely that the epileptogenic zone lies in the immediate surroundings of the recording site. Though even

lateralisation or regionalisation of ictal discharges are of significant relevance for presurgical evaluation.
Focal seizure complexion is strongly specific to the region of discharge origin which leads to quite clear correlations between electrographic manifestations and epilepsy types.

As opposed to focal ictal discharges, generalized seizures display a uniform behaviour especially in association with idiopathic epilepsies. Generalized discharges connected with symptomatic epilepsies can illustrate more diversity particularly in patients suffering from diffuse structural injuries.
The most prominent characteristic features of generalized seizures are GSW discharges often including polyspikes and usually depicting frequencies around 3Hz. GSW complexes with duration of less than 3sec mostly do not exhibit clinical correlates.
Myoclonic artifacts can obscure the EEG recording of ictal discharges and are most common in generalized tonic-clonic (GTC) seizures. This type of ictal discharges is initialized by repetitive alpha frequencies in the anterior head regions (called “recruiting rhythm”) which are then overlaid by movement artifacts resulting from the clonic phase.
Tonic seizures on the other hand are accompanied by a sudden onset of rhythmic 10Hz activity and paroxysmal fast frequencies mostly of low-voltage.
Just like focal ictal discharges, generalized seizures are rather specific of certain forms of epilepsy or epileptic syndrome and may provide relevant clinical information in terms of diagnosis and response to treatment.

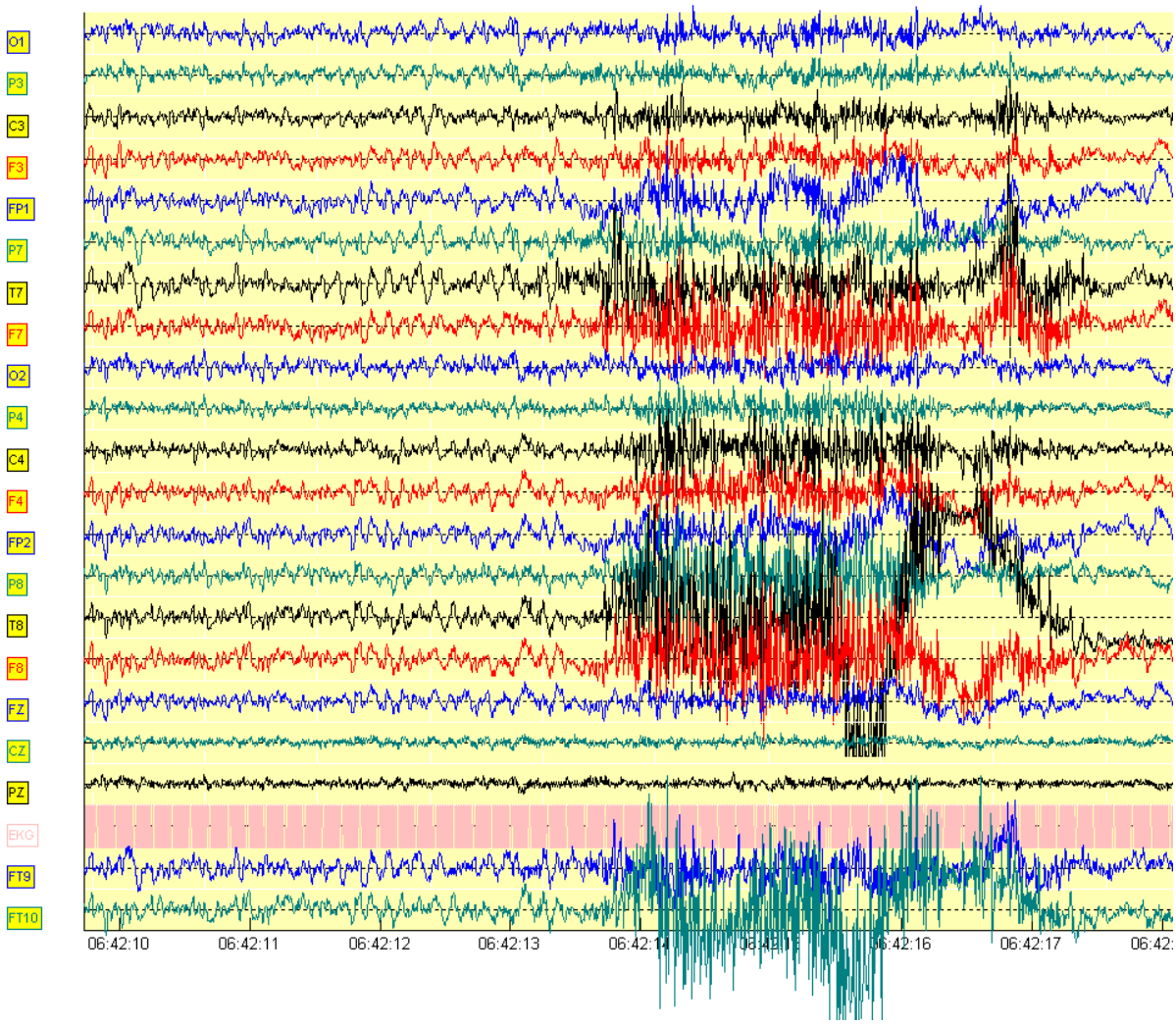


Fig. 3.8. Example of an EEG with onset of epileptic seizure at 06:42:13

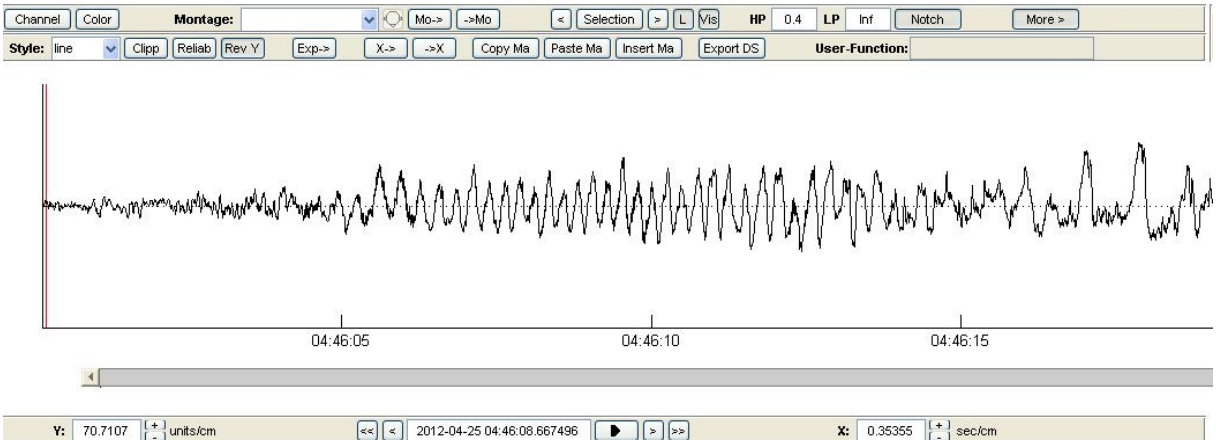


Fig. 3.9. Example of an epileptic EEG for one selected channel

4 Automated Detection of Epileptiform EEG activity

As already mentioned, detecting ictal and inter-ictal patterns in EEG signatures is of utmost importance for diagnostic and treatment purposes in epileptic diseases. The application of long-term EEG recordings coupled with video monitoring and subsequently the development of portable recording systems (ambulatory EEG) eventuated in a dramatic increase of clinical relevant information (inter-ictal as well as ictal accounts). Long-term video-EEG proved to be especially significant for the evaluation of possible candidates for epileptic surgery. Ambulatory EEG on the other hand provides more accurate ictal recordings as the frequency of seizures usually shows a decrease during (long and expensive) stays in special hospital units which aids both in the diagnosis of the type of epilepsy or epilepsy syndrome and in medical treatment considerations.

In former days the scrutiny of EEG data took place as visual inspection of the recordings by a specialized and experienced physician. This approach however proved to have several serious disadvantages. New techniques like long-term video-EEG monitoring result in a vast amount of data that render visual scanning methods ineffective and very time-consuming. Another difficulty lies in the similarity of epileptiform abnormalities with background noise and the obscuration by artifacts due to biological or environmental factors. These pitfalls and the wide variety of epileptic EEG activity frequently lead to misinterpretations and disagreements between neurophysiologists about one and the same location in the recorded data as their interpretations are rather subjective. Such being the case, a multitude of different methods for automated detection of inter-ictal or ictal epileptic discharges has been developed over the years to avoid mistakes owing to limited human abilities compared to computer-aided possibilities. The target lies with the development of expedient clinical units that are able to provide objective analysis of EEG signals and effort reduction thanks to computational efficient programming.

The strategies for automated analysis of EEG data can be divided into two main groups according to the scrutinized period:

1. Inter-ictal spike detection (spike detection analysis)
2. Ictal analysis (epileptic seizure analysis)

4.1 Automated inter-ictal spike detection

The objective of this kind of approaches to automated detection concerns itself with the problem of recognizing inter-ictal spikes in the channels of the EEG data while achieving highest possible sensitivity and selectivity. High sensitivity demands the ability of identifying

nearly all positive events concurrently with providing a minimum of false detections (high selectivity).

Considering the used criterion for spike detection, the different methods can be broken down into nine classes. However, each of these categories has to take into account two separate issues that make up the analysis of inter-ictal spikes: feature extraction and classification. The EEG data of one channel constitutes an N-dimensional space which is subsequently mapped onto an M-dimensional feature space where M is usually smaller than N. Approaches where M is equal to N choose not do preprocess the recordings to extract features because they want to avoid premature decisions regarding the ranking of importance of the parameters, in other words they use the raw EEG recordings. Oblivious if pre-processing has taken place the classification is then conducted on the M-dimensional feature space.

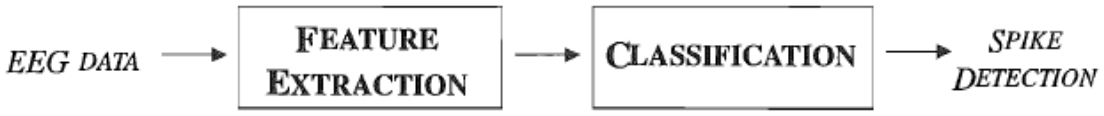


Fig. 4.1. The individual stages of automatic spike detection (James,1997)

Up until now only one channel was employed for the analysis at a time but the additional use of spatial and temporal information greatly increases the likelihood of a correct identification of candidate waveforms as actual spikes. Along these lines Figure 4.1. can be transformed to include contextual information. Feature extraction is performed for every channel of the recording and spatial information is therefore introduced through the multi-channel features, e.g. if only one channel shows a deflection it is necessary to make sure that artifacts due to electrode pops can be excluded. The adoption of temporal information can be achieved by taking into consideration the appearance of previous spikes in a certain time frame in the recorded signal in the classification step.

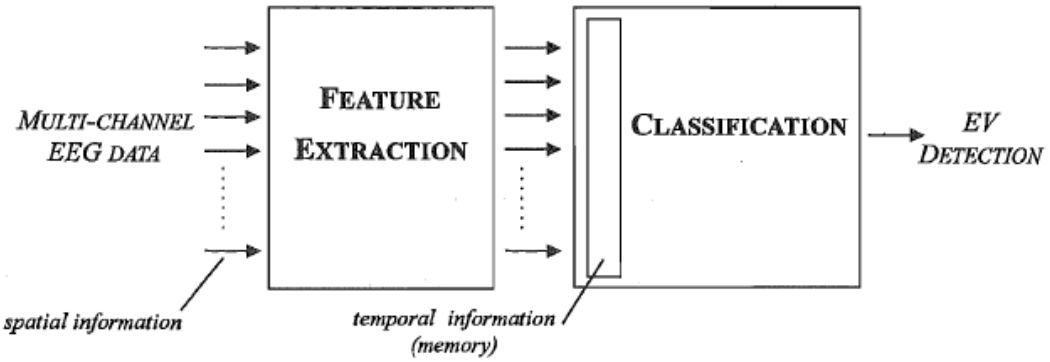


Fig. 4.2. Expansion to additional information and multi-channel analysis (James,1997)

In the following a summary over the nine categories for automated spike detection analysis will be provided:

a. Approaches using traditional recognition techniques called mimetic techniques

Mimetic methods extract several features from the raw data and employ the expertise of neurophysiologists to set a specific threshold value to distinguish normal and abnormal parameter displays. Usually the features taken from the EEG data define the "sharpness" of the spike candidate like peak amplitude, pre-peak and post-peak slope, durations and 2nd derivation but vary largely over the researchers taking this approach.
Gotman and Gloor (1976) for example refined their method by decomposing the transient waveform in two half-waves of opposite direction. Other authors introduced supplementary contextual information or ECG, EOG, EMG recordings to improve the performance of their technique (Glover et. al, 1989).

b. Approaches based on morphological analysis

Morphological characterization employs the use of waveform analysis as well as frequency-domain or time-frequency-domain methods to identify spike representations in the EEG signal. Thus the spectral content of the signal can be examined as to how it varies over time (e.g. using wavelet transform). Background activity and spike can be differentiated by decomposing the signal into distinct physical parts of interest.
Using the combination of mathematical morphology and wavelet transform Pon and coworkers (2002) were able to distinguish epileptiform signals from background activity and other transient phenomena regardless of the orientation of the spike (positive or negative). The mathematical component introduced a disk as a pre-defined structuring element as a comparative measure for smoothness.
Xu and coworkers (2007) optimize structure elements by building them from two parabolas and defining new optimization criterions for them. Moreover, they enhanced the morphological filter using open-closing and close-opening operations to get rid of amplitude deviations. (Opening and closing are basic operators used in the field of mathematical morphology to analyze and process geometrical structures.)

c. Approaches employing template matching

This method builds templates through averaging over several waveforms that have decisively been identified as spikes. Thereupon, this template is compared with segments of the EEG and high correlation between the two characterizes spike candidates.
El-Goharx, McNames and Elsas (2008) used for example the Mean Squared Error (MSE) test to match the templates to spikes in the signal under analysis.
Several researchers employed time-scale or time-frequency methods in this context to withdraw the parameters for template building from the EEG recordings (Senhadji and Wendling, 2002).

d. Approaches based on parametric techniques

The basic assumption of this method defines spikes as non-stationary deviations from the background EEG that is at least local stationary. Parameters make up a model of the normal EEG activity to build up an autoregressive (AR) filter.

Tzallas and coworkers (2006) employed Kalman filters to determine the parameters for the autoregressive model.
Equally applicable proved to be double differentiation with adjoined employment of linear predictive coding which makes use of the spectral signal composition (Birkemeier, Fontaine, Celesia and Ma, 1978).

e. Approaches deploying independent component analysis (ICA)

Independent component analysis breaks a multivariate signal down into subcomponents of supposedly statistical independent sources that add up to form the whole signal. These components are found by either maximizing the non-Gaussianity or minimizing the Mutual Information.
Ossadtchi and coauthors (2004) used ICA to identify spike candidates in the data and clustered them afterwards according to their location and time evolution, selecting thereupon only significant clusters as detected spikes.
Kobayashi and coworkers (1999) employed multidimensional ICA by generating mixing matrices composed of EEG background activity and epileptic discharges. They improved their method (2002) by applying RAP-MUSIC (recursively applied and projected multiple signal classification) to the resulting spatial patterns of ICA components corresponding to spikes for source localization.

f. Approaches using artificial neural networks (ANNs)

Artificial neural networks are mathematical models inspired by the human central nervous system. It consists of simple processing elements (nodes or neurons) that are interconnected to exhibit a complex global behavior. The strength of the connections adapts to the desired output (signal) and the artificial neurons “learn” new structures.
ANNs can be trained and fed with raw EEG data (Ko and Chung, 2000) or with extracted features. The selected parameters are either from mimetic approaches like durations, amplitude, slopes or even contextual information or are taken from time-frequency methods such as wavelet transform.
This approach can be even further extended to include three stages (Acir, Oztura, Kuntalp, Baklan and Guzelis, 2005). At first feature assessment produces three classes: epileptic discharges (EDs), no EDs and possible EDs. The latter are then further processed by a nonlinear ANN followed by the introduction of multichannel information. The group evaluated the performance of several different types of neural networks (e.g. multilayer perceptron, radial basis function ANN) through comparison to visual scrutiny of physicians.

g. Approaches employing clustering methods

Cluster algorithms divide the data under analysis in hierarchical or agglomerative groups and subgroups according to a specific similarity criterion. The difference to classification techniques lies in the formation of completely new categories instead of using predefined classes. Variation in similarity criteria, algorithmic complexity and tolerance of noise in the data leads to a large amount of different cluster analysis methods.

Sommer and Golz (2001) used self-organizing maps which produce discretized representations of the signal by deploying unsupervised learning and combined them with an image processing method (watershed transform) to yield separable clusters entirely automatically.
Other notable algorithms used for clustering are the nearest mean (NM) algorithm, the K-means algorithm and the fuzzy C-means (FCM) algorithm.
The FCM algorithm was implemented with and without using a pre-classifier by Inan and Kuntalp (2007). They also compared its performance to K-means clustering and found better results for FCM. In fuzzy clustering each data point belongs only to a certain degree to a specific cluster. At first the number of clusters and the membership of the points are initialized randomly. Then the procedure is repeated until the algorithm converges to yield the optimal clusters.

h. Approaches applying data mining and similar classification methods

Data mining (DM) is closely related to the last category as it attempts to split large datasets in new groups by transforming information in more comprehensible and easier applicable forms. It integrates aspects of statistics and database processing as well as neural networks and machine learning techniques to transcend simple raw data analysis to include pre-processing, modeling, drawing conclusion, management and update procedures. The advantage of applying DM to spike detection lies in the absence of the necessity to predetermine the spike morphology exactly.
This benefit was demonstrated by Valenti and coauthors (2006) by employing the DM techniques Decision Tree and Statistical Bayesian Classifier to detect inter-ictal spikes. Specific training of the algorithms proved to enhance the performance to recognize up-to-then invisible events.
Another classification method is constituted by support vector machines (SVMs) which are non-probabilistic linear classifiers based on supervised learning algorithms.
SVMs can again be coupled with a pre-processing step that distinguishes spike candidates from definitive non-spikes and using the SVM as post-processing to further analyze the first category (Acir and Guzelis, 2004). In SVMs the two possible categories are divided by a gap which was in this work adjusted by taking into account the distance of outliers that were grouped falsely.

i. Approaches making use of knowledge-based rules

The application of knowledge-based rules serves mainly the purpose of introducing contextual information like spatial and temporal details in addition to dealing with single EEG channels (as was done in most of the methods presented above, especially regarding the first four classes).
Liu and coworkers (2002) developed a rather complex multistage system incorporating adaptive filters, wavelet transform, ANN and expert systems to make use of temporal and spatial information with great success as they not only were able to locate spikes but slow waves as well.

Another knowledge-based technique consists of two separately trained and applied ANNs (Ozdamar and coauthors, 1991). The possible spike waveforms identified in the first component are spatially integrated in the second part of the system in groups of 16 first-level modules at a time.
James and coworkers (1999) came up with a three stage model combining mimetic approaches, an ANN called self-organizing feature map (SOFM) and fuzzy logic. After the features are extracted from the raw EEG using the mimetic component the SOFM allocates probabilities to the spike candidates and serves as a classifier. The first two steps are performed on single-channel data while the last part incorporates the spatial information by applying fuzzy logic.
The application of spatial information can even be directly involved into the preprocessing stage of a multi-level system (Webber and coauthors, 1994). Possible sharp transients are identified by thresholding the data in 4-channel bipolar chains. The parameterized candidates are then fed into a 3-layer ANN used as classifier.

Figure 4.3. summarizes the most important cornerstones making up the automated detection of inter-ictal spikes.
First a decision has to be made if the raw data should be used or if features should be extracted from the EEG. Moreover, if features are selected, which ones are of greatest importance? Finally, an appropriate algorithm or even a combination of them has to be chosen.

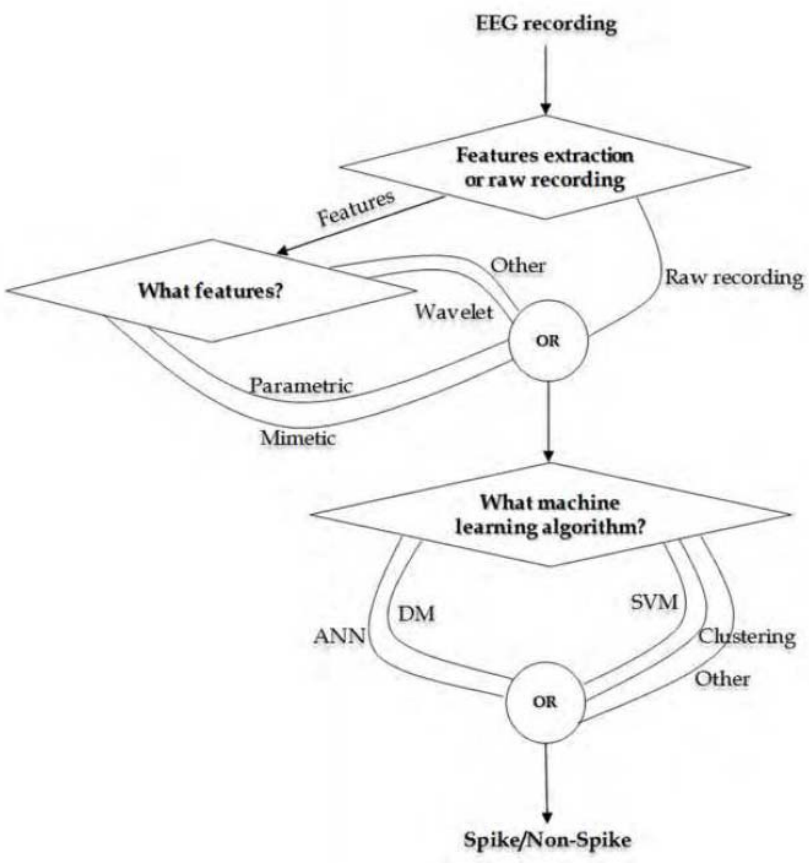


Fig. 4.3. Graphical representation depicting the decision nodes in the inter-ictal spike detection problem (James 1997)

4.1.1 Spike enhancement strategies

By selecting the right parameters and thresholds it would be comparatively easy to adjust every one of the spike detection methods in a way that all epileptic discharges would be recognized. However, this high sensitivity would come at the cost of misclassifying a large number of spike-like artifacts or background noise. On the other hand, it is also possible to maximize the selectivity but letting slip several true events. Several researchers advocate the opinion that it would be better to maximize sensitivity and let experienced physicians check the results for false detections. The disadvantage of this approach is that a relatively large number of incorrect alarms might lead to distrust in and disregard of the system by physicians and patients.

Therefore, the true challenge lies in finding the optimal balance between high sensitivity and high selectivity. To aid in this problem a pre-procession stage could be built into spike detection methods before feature extraction and classification take place. This step would not be commissioned with the detection itself but merely serve to enhance every spike-like waveform in the EEG data. An additional spike enhancement component would accentuate the sharp transients for further analysis with one of the spike detection methods already discussed above, thereby increasing the detection performance while keeping the number of misclassified events at an acceptable level.

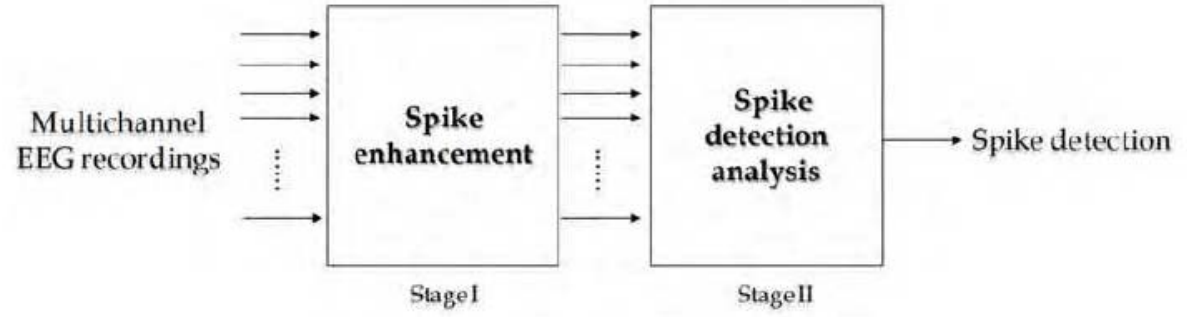


Fig. 4.4. Incorporation of spike enhancement previous to spike detection (Tzallas and coauthors, 2012)

A share of the techniques used for spike detection can also be applied to the enhancement stage of the system.
By assuming again the local stationarity of the background activity in the recordings and the non-stationarity of sharp transients an autoregressive model whose parameters are estimated by a Kalman filter can be used to reduce the amount of false detections (Oikonomou and coworkers, 2007).
James and coworkers (1997) applied multi-layer ANNs to implement a multi-reference adaptive noise cancelling (MRANC) which adaptively cancels the noise on the EEG channel that is currently investigated by taking advantage of the background activity on nearby channels in the multi-channel recordings. The continuous adaption of the filters makes it very stable with regard to deviations of the background activity like baseline drift and the deployment of ANN introduces a non-linear component that proved to be more successful than linear approaches.

4.2 Automated epileptic seizure analysis

Epileptic seizure analysis can be broken down into three main problems:

- Detection
- Prediction
- Origin location

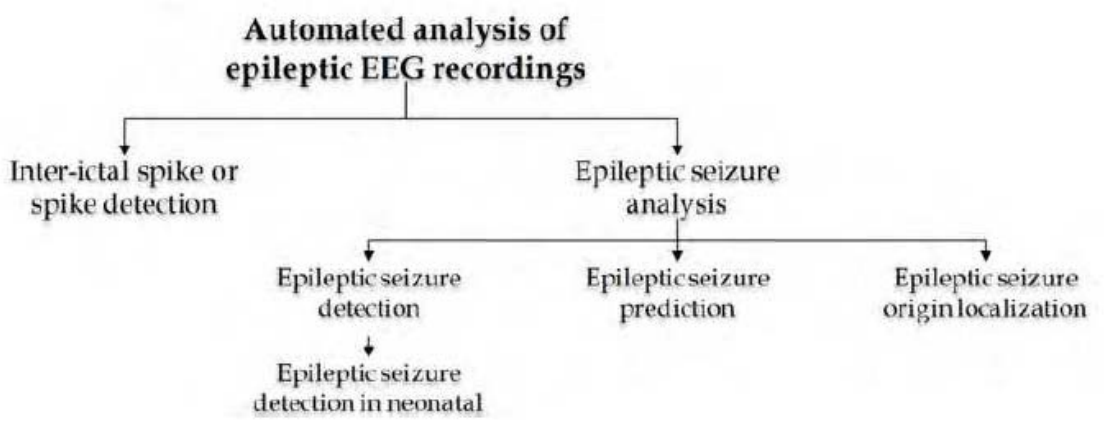


Fig. 4.5. The different approaches to analyzing epileptic EEG data with automatic procedures (Tzallas and coworkers, 2012)

a. Epileptic seizure detection

As opposed to seizure prediction, in seizure detection the goal is to give a warning as soon as possible after the first signs of an epileptic seizure have become evident in the EEG recordings.
As all of the developed approaches for automated epileptic seizure detection are based on the idea of recognizing specific patterns in the EEG data that are characteristic of epileptic seizures, the problem can again be broken down into the two steps feature extraction and classification. An observation window moves over the multichannel EEG recordings and collects prescribed features which are then fed into a classification procedure distinguishing inter-ictal or pre-ictal from ictal periods.

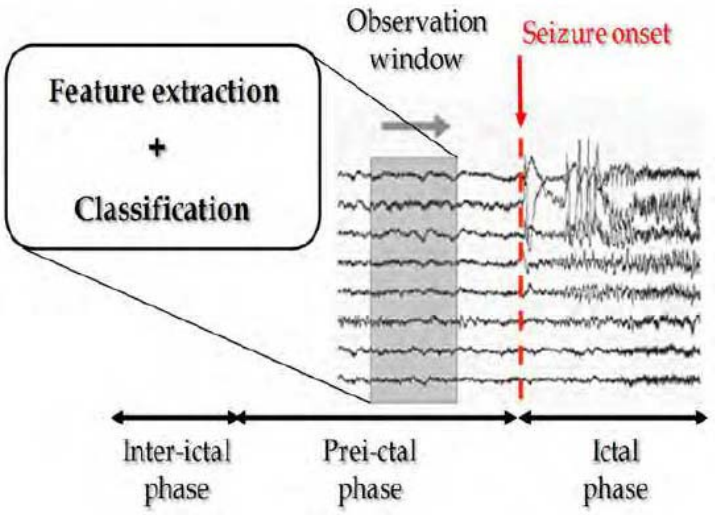


Fig. 4.6. Moving observation window to distinguish inter-ictal or pre-ictal phases from ictal ones (Tzallas and coworkers, 2012)

The choice of optimal features is crucial to the subsequent classification of the data under analysis and a huge variety of different approaches have been taken over time. The following will briefly summarize the volume of possibilities for feature extraction and present several examples in greater detail. Especially when addressing the problem of detecting rhythmic discharges some of the implemented algorithms are based on similar principles as inter-ictal spike detection.
Several authors based the selection of their features on the fact that epileptic seizures are caused by neural networks firing synchronously par example employing the autocorrelation function (Liu and coworkers, 2002).
Seizure onset was found to be correlated with an increase in EEG synchronization for newborns and features were chosen by making use of the synchronization likelihood (Altenburg and coauthors, 2003).
Van Putten (2003) proposed the calculation of nearest neighbor phase synchronization over several frequency bands and application of different thresholds in order to obtain receiver operating characteristic (ROC) curves for the evaluation of his approach and intuitive EEG interpretation.
Once again, the idea of morphological features such as durations, amplitude and slopes as well as the decomposition of the waveforms into two half-waves were used for spike or polyspike detection (Gotman, 1982, 1999).

To recognize rhythmic discharges in the EEG recordings methods were based on fast Fourier transform, frequency- or time-domain analysis or wavelet transform.
Polat and Günes (2207) implemented a two-level algorithm that used fast Fourier transform (FFT) as a feature extractor and afterwards employed decision trees as classifiers. The FFT provides a decomposition of the signal in its frequency components and the resulting parameters are then fed into a decision tree which is a probability model that calculates the expected values of each path.
Some methods used higher-order spectra (HOS) as compared to statistical tools such as first or second order moments (e.g. arithmetic mean and variance) to identify non-linear behavior

and deviations from Gaussianity in the EEG data (Chua and coworkers, 2008). The parameters were subsequently analyzed by a Gaussian mixture model and the results promised high classification accuracy.
Martinez-Vargas and coworkers (2011) concerned themselves with the comparison of several different time-frequency methods and found that simpler approaches proved to be more efficient whereas more sophisticated approaches might achieve the same performance only with further parameter tuning.

Some studies even tried to use a variety of features instead of concentrating on one specific type of them. A mixture of time-domain features, frequency-domain features, wavelet features, auto regressive coefficients and cepstral features was applied in neonatal spike detection. (Cepstral features result from taking the logarithm of the spectral representation of a signal and applying Fourier transformation.) The best features for differentiation were found to be time-domain ones including amplitude, duration and slope and frequency-domain parameters as the relative power between certain frequency bands.
Iscan and coworkers (2011) investigated the usefulness of combined time and frequency features. Time parameters were extracted by cross-correlation and frequency parameters by calculating the power spectral density. They were combined by divergence analysis to represent the feature distribution quantitatively.

A number of different time-domain features were used in many approaches including statistical features, nonlinear features, Lyapunov exponent, eigenvector methods, PCA, ICA, cross-correlation functions and entropy methods.
Lyapunov exponents characterize the rate of separation between infinitesimally close trajectories and describe the degree of chaos in a system. Coupled with wavelet coefficients they were used as feature extractors and classified by support vector machines or neural networks to show high accuracies (Güler and Ubeyli, 2007).
Wavelet coefficients were also used in connection with Shannon entropy to avoid exhaustive search and to calculate the lowest cost basis in terms of measuring the signal complexity against the wavelet coefficients with large entropy values indicating high complexity (Wang, Miao, Xie, 2011).
Cross-correlation describes a similarity measure based on finding patterns in a signal by comparing two waveforms as a function of a time-lag applied to one of them. It was used for feature extraction together with support vector machines with good results (Chandaka and coauthors, 2009).

Genetic programming (GP) is inspired by biological evolution and belongs to the group of machine learning algorithms. Guo and coworkers (2011) showed that GP could be employed for automated feature extraction with its tree structure representing the features and an implemented function automatically determining the number of chosen features. Therefore better performance of the classifier and a limitation of the dimensionality of necessary features could be attained.

As becomes evident from the short overview the list of different parameters is almost endless and yields a huge variety of implemented algorithms but one has to keep in mind that the features are of no physiological relevance.

In the following the chosen parameters determining the data under analysis have to be analyzed to differentiate epileptic seizures from the residual EEG. The number of methods proposed for classification purposes are equally numerous as those described for feature extraction.
McSharry and coworkers (2002) concerned themselves with the general question if linear approaches would yield a better result than non-linear ones based on the assumption that the dynamic processes underlying EEG recordings were constituted from low-dimensional chaos. However, they found no considerable difference between the two method groups except from fewer false results for non-linear detection.

The different types for classification include simple threshold, linear classifiers, rule-based decision but also complex decision boundaries as found in ANNs.
Altunay and coworkers (2010) used linear prediction filters to detect epileptic seizures through the energy of the prediction error. They found the energy value to be a very decisive tool that could be classified by simple thresholds.
A specific kind of ANN employed for classification with good results are multilayer perceptrons (MPs) (Mousavi and coauthors, 2008). MPs consist of multiple layers of nodes or neurons that are interconnected building a directed graph to map the input sets on an adequate output set. It is able to deal with non-linear data because each node has a non-linear activation function which means it is not a simple on-off switch. The model is trained through backpropagation by changing connection strength based on the error value produced in the output set.

Further methods for classification discussed in studies are SVM, k-nearest neighbor classifiers, quadratic analysis, logistic regression, naïve Bayes classifiers, decision trees, Gaussian mixture models, mixture of expert models and adaptive neurofuzzy interference systems.
A multi-level approach using PCA, ICA and linear discriminant analysis for dimensionality reduction before applying a support vector machine to the ends of building an intelligent device showed promising performances compared to other classification methods (Subasi and Gursoy, 2010).
Orhan, Hekim and Ozer (2011) applied k-means clustering to parameters resulting from wavelet transform and used the resulting probability distributions as input for a multilayer perceptron network.
Studies comparing simple statistical logistic regression (LR), which describes the probabilities of the possible outcome by a logistic function, and complex ANNs (MPs) found logistic regression methods to have less accuracy and inferior performance (Alkan and coworkers, 2005).

SVMs are the most acquainted representative of the class of Kernel machines (KMs). KMs employ a kernel function to avoid using the coordinates of the signal under investigation but computing the inner product of pairs of data in the feature space to achieve reduction of computational effort. Lima and Coelho (2011) used Gaussian and exponential radial basis functions as kernels to evaluate the performance of different types of support vector machines.
Mixture of expert (ME) models consist of a number of neural network experts that are divided into groups and a gate-bank. Experts in the same group receive the same input features whereas those in another group get different feature vectors. The local experts derive conditional probabilities for the target input which are then combined in the gate-bank by gating networks. The training of the gating networks selecting the expert network performing best can be coupled with an Expectation-Maximization (EM) to optimize the output by effectively making use of diverse features representing the signal under analysis (Übeyli, 2007).
Kannathal and coworkers (2005) used adaptive network based fuzzy inference systems (ANFIS) as classifiers when comparing different kinds of entropy estimators applied to EEG recordings. Fuzzy inference systems consist of a knowledge base that involves membership functions to assign membership values and fuzzy if-then rules to assign weights on which the following decision-making process is based. In ANFIS these fuzzy inference systems are further enhanced through supervised learning capabilities by employing adaptive neural networks.

b. Epileptic seizure prediction

Epileptic seizures occur mostly without warning (sometimes the patient feels an aura immediately before onset) and can lead according to the expressed clinical correlates to serious injuries and dangerous situation (e.g. while crossing a street, walking down stairs, swimming…). Evidently, it would be of great advantage for the patient to know the upcoming of a seizure in advance to take precautions, thereby immensely increasing the quality of life and reducing the patient's anxiety and constant need of alertness.
Moreover, advance warning could alert medical staff to deliver drugs or make arrangements for additional tests that could help in diagnosis or treatment considerations or even epileptic focus localization.
Prediction systems could be of huge benefit if coupled with treatment like drug delivery or electrical stimulation. In this way the treatment with anti-epileptic drugs could be reduced to targeted and short-time application which would possibly reduce the relatively high incidence of serious adverse effects.

Whereas epileptic seizure detection can also be applied offline, seizure prediction naturally has to take place online and should last a minimum of time to provide warning of the oncoming seizure as early as possible.

The approaches to predict epileptic seizures are much diversified and include time-domain methods, frequency-based analysis, nonlinear dynamics, methods of delay and intelligent strategies.
Lange and coworkers (1983) employed spatial (using only a subset of channels) and temporal information to extract statistical parameters like the rate of occurrence to determine specific patterns that could indicate the transition from inter-ictal to ictal periods in patients with temporal lobe epilepsy.
The possibility of discerning a distinct pre-seizure phase was also explored and confirmed by nonlinear EEG analysis (Lehnertz and coworkers, 2001).
Incorporating the time-frequency analysis tool of chirplets in a matched filter design proved to result in highly sensitive and specific markers for epileptic seizure activity (Schiff and coauthors, 2000).
The prediction of generalized epileptic seizures was attempted through wavelet transform and fuzzy logic by Geva and Kerem (1998). After features are extracted by the wavelet analysis an unsupervised optimal fuzzy clustering (UFOC) algorithm classifies them into several states as sleep, alert, active wakefulness, seizure and in about two-thirds of the test objects a pre-seizure state was detected too.

Osorio and Frei (2008) implemented an algorithm to distinguish seizure from non-seizure data by using a double-filter that employed two different timeframes, a short window and a longer window serving as reference. If the onset of a seizure was recognized, electrical stimulation was applied directly to the epileptogenic zone. The time-restricted and selective application of therapy showed to be overall beneficent with respect to seizure evolution. They used their method also for quantification of variables such as seizure intensity and duration, extend of spread, interdependencies and circadian rhythms to get a deeper insight into the workings of the epileptic brain.

For the evaluation of the performance of the diverse seizure prediction methods the application of the seizure prediction characteristic was introduced (Winterhalder and coworkers, 2003). A feasible prediction system should be able to provide an appropriate timeframe from the alarm to the actual onset of the seizure; this period is called seizure prediction horizon (SPH). Of course no system can foresee the exact time value of the start of the seizure. Therefore the seizure occurrence period (SOP) describes the time window in which the seizure is supposed to take place. In addition sensitivity and selectivity are investigated. Depending on the type of therapeutic intervention there is as minimum for the SPH and a maximum for the SOP. The minimum SPH is determined by the time it takes to successfully initialize the intervention method and other necessary precaution steps. The maximum of the SOP on the other hand is conditioned by the effects for the patient (avoiding prolonged fear) and the utmost time window that allows safe antiepileptic therapy without adverse effects.
To be effective and safe, the system has to detect nearly all oncoming seizures but a high false prediction rate would render the system ineffective and produce increased stress and anxiety for the patient.

As the type of therapy is often unknown during the development of the prediction method and to make the different systems comparable the values for SOP, SPH and the false prediction rate should be varied while calculating the according sensitivity. In this manner the appropriate seizure prediction technique can be selected for every treatment option and the assessment with the characteristic allows for performance outlook.

c. Epileptic seizure origin location

About one third of the patients suffering from epilepsy prove to be medically refractory and therefore treatment with anti-epileptic drugs is unsuccessful or comes at the cost of serious adverse reactions. If the impairment caused by the epileptic seizures is severe enough and comes from a consistent, specific area in the brain, epilepsy surgery can be considered as a viable therapy option. Additionally, the origin of the epileptic activity has to be in a circumscribed region of the brain that can be surgically removed or altered without damaging brain functions.
Of utmost importance for a successful surgery is the localization of the epileptic seizure origin. Detailed knowledge of the spot and spread of the focus can determine the outcome of the intervention. The necessary information is mainly obtained by analysis of ictal EEG data recorded either on the scalp or intracranial. Several researchers have developed different approaches to implement algorithms that assist in the gathering of the essential data for pre-surgical evaluation.

Parra, Spence, Gerson and Sajda (2005) took advantage of the assumption of a linear mixing model underlying the EEG recordings thereby avoiding the necessity of information about spatial correlations. The model employs statistical parameters like maximum power or maximum difference for linear integration of the multichannel data. Afterwards several analytical techniques as principal or independent component analysis serve to remove artifacts or extract responses that are of interest to the analyst.
The multichannel EEG data can be decomposed by applying wavelet transform to yield a tensor with three different modes as time samples, scale, electrodes (Acar and coworkers, 2007). Next, multiway analysis is employed to build a complex model of the epileptic seizure structure and artifacts are removed by multilinear subspace analysis. This approach uses highly nonlinear methods to implement an automated and robust algorithm for seizure origin localization.
Miwakeichi and coauthors (2004) used the same multiway array analysis called Parallel Factor Analysis (PARAFAC) as in the last approach to decompose the EEG data into a time-frequency-space representation. Thereafter they applied Source Spectra Imaging (SSI) to localize epileptic activity by estimating electric current sources from the spectral decomposition of the EEG.

5 EpiScan – Automated Online Seizure detection

The EpiScan detection system is intended for the use with Epilepsy Monitoring Units (EMUs) in hospitals in pre-surgical evaluation and was developed by the department Safety and Security at the Austrian Institute of Technology (AIT). As about a third of epilepsy patients can't become seizure free with anti-epileptic drugs or suffer from severe adverse reactions, epilepsy surgery presents an alternative therapy possibility to those patients. Accurate and detailed evaluation of the epileptogenic zone is vital to the outcome of the surgery and the most valuable tool in this context is represented by long-term EEG monitoring. During the time of the monitoring anti-epileptic drugs are reduced to provoke seizures in order to gain the desired information. Consequently, it is important to recognize nearly all seizures to be able to monitor them and to avoid risks for the patient's health.
However, manual scanning of the recordings would be extremely time consuming and arduous requiring experienced EEG technicians. Therefore, automated detection of seizures would provide a great benefit for several reasons. A distinction has to be drawn between systems that work offline and those that function online. Offline detection methods can considerably reduce the amount of time and work needed for reviewing the recorded data. Systems that work online on the other hand should be able to alert medical staff to the onset of a seizure in due time to provide medical treatment for the patients safety and to take measures for the performance of further tests that could possibly provide more insight into the disease.
According to their different purposes it is obvious that online and offline detection systems differ in their requirements. Online methods have to deal with a continuous data stream in real-time and have to keep the latency period as small as possible. As a result decisions are reached due to the analysis of data samples from the past only. Moreover, the EEG is the only input data to the system. EpiScan was developed for online use as a warning system working in combination with EMUs. The following specifications are the four main demands imposed on the EpiScan unit:

- High sensitivity: The detection system has to recognize as many epileptic seizures as possible to avoid dangers for the patients and the failure to notice relevant findings.
- Low latency: The alert should be raised as soon as possible after the first electrographical signs of the seizure onset are noticeable in order to allow for enough intervention time for the medical staff.
- High specificity: Especially in practical clinical applications it is extremely important to further the acceptance of the introduced system. A high number of false alarms would annoy the staff and reduce the trust in the method. In the last years a number of online seizure detection methods have been proposed and are even available as commercial software but the high rate of false alarms for unselected EEG records prevents them from effective use in clinical practice.

- Easy-to-use: For the same reason it is vital to make the system as user friendly as possible. Therefore, it shouldn't be necessary to adjust the method for every patient which means that EpiScan has to work parameter-free.

Several challenges have to be met in order to meet the requirements listed above. In the first place, the availability of an appropriately large pool of recordings which have been previously annotated by EEG experts is needed. To guarantee the relevance and comparability of the reviewed records and the detection system, the annotation has to be based on the EEG only. This data set then serves to identify the difficulties presented by the nature of the signals under analysis but also for performance testing of the implemented algorithm.
Since epileptic seizures are expressed in connection with a wide variety of different epileptic diseases and syndromes the corresponding EEG reflects this diversity and exhibits a large inter-patient variability that has to be dealt with devoid of individually adjustable parameters in the system.
The nature of long-term EEG monitoring poses further challenges like the difference in EEG recordings during sleep- and wake-states. Moreover, daily tasks performed by the patient result in artifacts, the most important of which are eyelid fluttering or blinks, chewing and movement artifacts. Changes in electrode impedance (also called electrode-pops) and other interference from technical equipment might also be visible in the recorded data. These issues are addressed by artifact detection and adaption methods.

The concept of the EpiScan algorithm consists of a periodic waveform analysis module with an additional baseline adaption to the ends of detecting rhythmic EEG patterns characterizing epileptic seizures.

The EpiScan detection system can be divided into six building blocks that make up the whole algorithm:

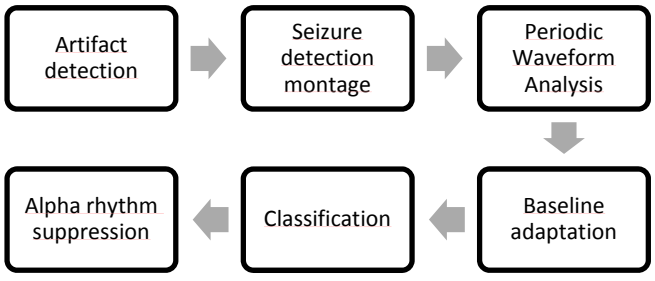


Fig. 5.1. Building Blocks of the EpiScan Algorithm (Hartmann, 2011)

1) Artifact detection:

First of all data samples obscured by technical or biological artifacts are recognized and labeled by using features as signal amplitude or variance or deviations in the power spectral density.

2) Seizure detection montage:

The usual montages used in clinical set-up for long-term EEG monitoring are the referential montage and two bipolar (longitudinal and transversal) ones. All electrodes from the referential one and a number of additional channels from the bipolar montages that were deemed to be most relevant for detection purposes were chosen for the EpiScan unit to work with.

3) Periodic Waveform Analysis:

The main part of the algorithm is the periodic waveform analysis (PWA) that detects rhythmic EEG discharges which characterize epileptic seizures especially in patients with temporal lobe epilepsy.

A rectangular time window ψ_t moves over the signal and selects the input sequence x_t for the analysis. The first step of the PWA consists of the calculation of the total harmonic energy E_τ of x_t for different cycle duration τ in a certain range.

$$E_\tau = \sum_{m>0} \left| \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x_t \, \psi^*_{t/\tau} e^{-j2\pi \frac{mt}{\tau}} dt \right|^2$$

Maximization of E_τ yields a dominant $\tilde{\tau}$ and $E_{\tilde{\tau}}$ is called the Periodic Energy Index (PEI). The Periodic Waveform Index (PWI) is then calculated by dividing the PEI by the signal energy $N_{\tilde{\tau}}$ with cycle duration $\tilde{\tau}$.

$$PWI = \frac{E_{\tilde{\tau}}}{N_{\tilde{\tau}}} \quad with \quad N_\tau = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} \left| x_t \psi^*_{t/\tau} \right|^2 dt$$

For entirely rhythmic signals the PWI is unity and for signals expressing no detectable rhythm the PWI approaches zero.

4) Baseline Adaption:

This module serves to take into account and counterbalance the inter-patient variability and the temporal changes in the recorded data.

For this purpose the PWI values are broken down into three different frequency bands PWI_δ, PWI_θ, and PWI_α. From these segments the P-th percentiles are calculated within a certain time window ranging over the past two to four hours in order to normalize the PWI with the obtained percentile values.

5) Classification:

For the classification part the PEI values are also split into the three frequency bands. Seizure alarms are triggered if the PWI values trespass a prescribed threshold and PEI values have to lie within a certain scope. The thresholds and scopes are defined for each frequency band individually but remain fixed from then on. Moreover, two alerts are only considered as separate if they distance between them exceeds 30 seconds.

6) Alpha rhythm suppression:

Rhythmic alpha activity can easily be mistaken as rhythmic epileptic patterns and accordingly may lead to a high number of false alarms.

In order to avoid this misinterpretation, the PEI from stage 2) is used to extract the highest amplitude in the alpha frequency band which is then employed to suppress alpha rhythms especially in occipital and parietal regions.

To provide an inside into the performance of the EpiScan algorithm several results from an off-line analysis simulating online processing of the EEG will be presented.
The recordings used for the test were taken directly from epilepsy monitoring units without further processing or alterations. Two EEG experts reviewed the data and annotated them according to 6 categories ranging from certainly a seizure (>90%) to certainly not a seizure (<10%). The first 3 classes (>50%) were taken to be markers of epileptic seizures that displayed a recognizable EEG recording.

The sensitivity of the EpiScan algorithm was determined patient-wise and defined as the ratio of true positives to the total number of recorded seizures. A true positive event is characterized by a seizure alert coinciding with an annotated marker. For evaluation of the sensitivities histograms and calculating the mean over all patients (with seizures) was employed.

Each seizure alarm is divided into sub-alerts of less than 30 seconds to facilitate reviewing purposes. A false event is defined by every sub-marker not intersecting with the basic truth defined by the annotated data. The false alarm rate is then given by the number of false events divided by the total amount of hours of EEG data of the patient under consideration. Again histograms and mean calculation over all patients (with seizures) make up the evaluation process.

In the simulation data from 48 patients with temporal lobe epilepsy 186 seizure markers were annotated as visible in the electrographic recording.
The mean sensitivity was found to be 83% and Figure 5.2. depicts a histogram of the sensitivities as there was a highly non-linear distribution seen over the patients. Notably, for about two thirds of the patients all seizures could be detected.

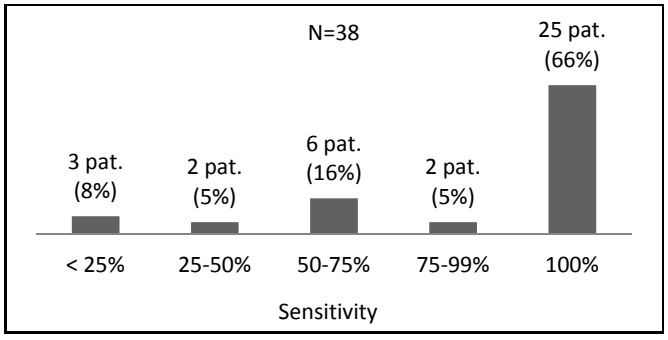


Fig. 5.2. Histogram of the sensitivities (Hartmann 2011)

Averaged over all patients a false alarm rate of 0.3 FA/h (false alarms per hour) or 7.2 FA within a day was obtained. Again the corresponding histogram was quite non-symmetric and shows that actually for about 40% the rate was even at less than 0.2 FA/h in other words less than one false alarm every five hours.

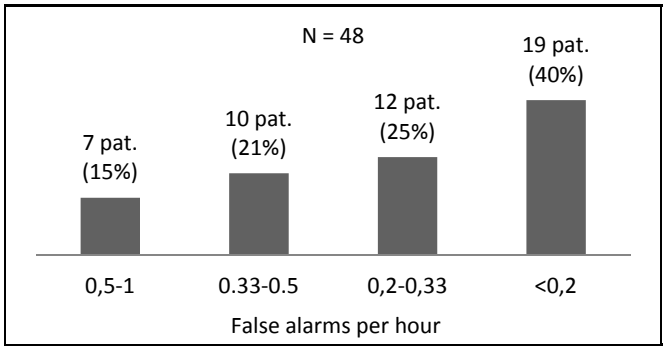


Fig. 5.3. Histogram of the false alarm rates (Hartmann, 2011)

5.1 Application of Chirplet Transform instead of PWA

Chirplet as well as wavelet transform belong to the family of time-frequency (TF) analysis methods which enable us to consider how the power spectrum of a non-stationary signal evolves over time. To this end the transform decomposes the input signal into its elementary components, called atoms, by building the inner product with a family of wavelets or chirplets.

Wavelets are obtained from a single Gaussian window, called mother wavelet, by applying translation and dilation. Those two mathematical operations are only one-dimensional which poses limitations on the representation of the signal under question.

Chirplets on the other hand are derived from their primitive, the mother chirplet, through application of two-dimensional affine coordinate transformations in the TF plane. These mathematical operations are translation, dilation, frequency modulation and chirp modulation whose parameters make up the index of every chirp atom. From this follows that each chirp atom $g_{(s,u,z,c)}$ can be characterized through its features scale s, time u, frequency z and chirp rate c.

Figure 5.4. shows the comparisons between a wave and a chirp and a wavelet and a chirplet respectively. Wavelets or chirplets are derived from a wave or a chirp by application of an elementary window. The notable difference between waves and chirps lies with the angle of rotation between the samples. Whereas it is consistent for waves, implying constant frequency, it is linearly increasing in the case of chirps.

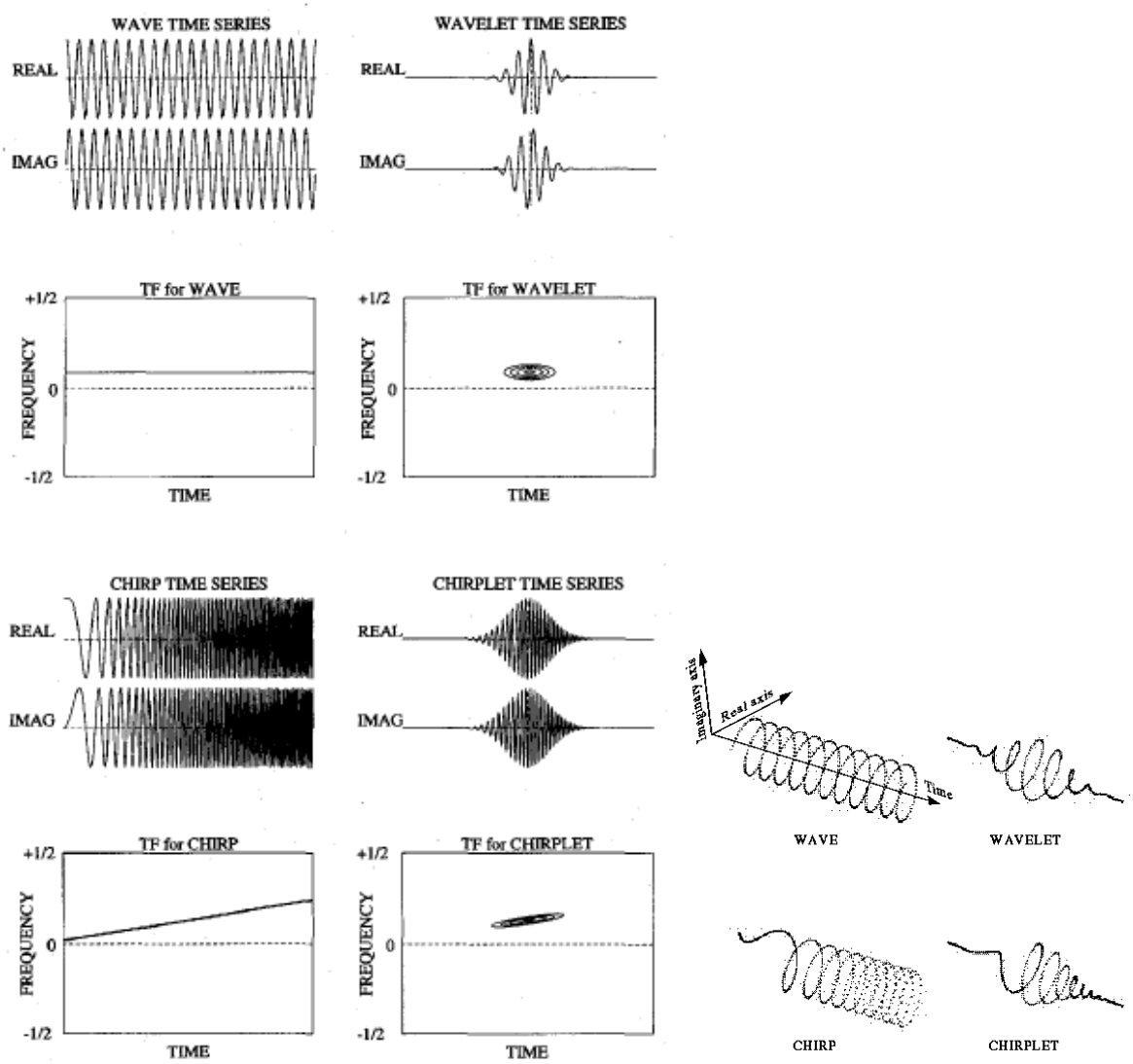


Fig. 5.4. Comparison of chirps and chirplets to waves and wavelets (Mann and Haykin, 1995)

Due to the doubled number of parameters and therefore an increased possibility for adaption to the signal under analysis, chirplets constitute improved options for signal decomposition and might be able to increase the performance of the EpiScan algorithm when applied in the third phase (PWA). For the evaluation of this possibility three implementations of algorithms using chirplet transform for decomposition are described in the following.

5.1.1 Standard matching pursuit with a multi-scale dictionary of Gaussian chirps

A matching pursuit algorithm splits the investigated signal into a linear expansion of waveforms or atoms which belong to a redundant dictionary.

This implies that the first step involves the building of a dictionary containing the chirp atoms. The discrete chirp dictionary D+ is built of all atoms $g_{(s,u,z,c)}$ so that (s,u,z,c) are in

$\Gamma+ := \{(a^j, na^j du, ka^{-j} dz, la^{-2j} dc), j,n,k,l \in Z\}$ where du, dz and dc are constants and a is called a scaling factor. The meaning of those four parameters lies therein that they define the sampling steps and thus the size of the dictionary. Several constraints like the Nyquist condition limit the sampling rate and the signal size and provide the boundaries for the parameters j, n, k, l which are implemented in the algorithm accordingly.

The chirp atoms making up the dictionary are then calculated by:

$$g_{(s,u,z,c)}(t) = \frac{1}{\sqrt{s}} * g(\frac{t-u}{s}) * \exp(\, i(z(t-u) + \frac{c}{2}(t-u)^2))$$

The algorithm is based on a standard matching pursuit which is unfortunately a greedy strategy to find a linear decomposition of the input and needs a lot of computational effort. The matching pursuit is initialized by $R^0 x = x$ and calculates iteratively a residual $R^{m-1} x$ of m-th order through the following steps:

- Computation of $\left|\left\langle R^{m-1} x, g_{(s,u,z,c)} \right\rangle\right|$ for all $g_{(s,u,z,c)}$ in D+.
- Selection of the chirp atom that yields the highest energy:
  $g_{(s,u,z,c),m} = \arg \max_{(s,u,z,c)} \left|\left\langle R^{m-1} x, g_{(s,u,z,c)} \right\rangle\right|^2$.
- Calculation of the new residual is achieved by removing the component pertaining to the selected atom:
  $R^m x = R^{m-1} x - \left\langle R^{m-1} x, g_{(s,u,z,c),m} \right\rangle g_{(s,u,z,c),m}$.

The algorithm does usually not converge and calls for a break condition or a prescribed number of iterations. After n iterations the energy of the input signal is split into the selected components and the residual: $\|x\|^2 = \sum_{m=1}^{n} |\left\langle R^{m-1} x, g_{(s,u,z,c),m} \right\rangle|^2 + \|R^n x\|^2$. However, the algorithm doesn't necessarily decompose x into the n best atoms. Furthermore, the computation time is very high and useless for practical application. For an N-point signal the total computational costs are $O(nN^2 \log N)$ for n iterations.
Although for the use with EpiScan actually only a few iterations are necessary to obtain the principal components, improvement upon the efficiency of the algorithm was inevitable.

5.1.2 Fast matching pursuit with a multi-scale dictionary of Gaussian chirps - The ridge pursuit

The basic idea underlying the improvement of the algorithm consists of the approximation of the best atom by exploiting local behavior instead of computing the inner product with every chirp atom of the dictionary.

As the chirp dictionary is an extension of the Gabor dictionary - which is built of the parameters s, u and z only - it can be shown (Gribonval, 2001) that the behavior of the inner product $\left\langle R^{m-1}x, g_{(s,u,z)} \right\rangle$ calculated with the best gabor atom $g_{(s,u,z)}$ will yield enough information to select a good chirp atom in the local vicinity. The best gabor atom is selected by the computation: $g_{(s^*,u^*,z^*),m} = \arg\max_{(s,u,z)} \left| \left\langle R^{m-1}x, g_{(s,u,z)} \right\rangle \right|^2$ and subsequently the "locally optimal" chirp atom is found through $g_{(s,u,z,c),m} = \arg\max_{(s,u^*,z^*,c)} \left| \left\langle R^{m-1}x, g_{(s,u^*,z^*,c)} \right\rangle \right|^2$. The time and frequency parameters u* and z* are kept constant because they have been proven to be very close to the optimal ones. To reduce exhaustive scanning for the scale and chirp rate parameters s and c the local behavior of z in the neighborhood of the optimal frequency rate z* is exploited.

The ensuing algorithm is called ridge pursuit and consists of the following stages:

- Selection of the best gabor atom $g_{(s,u,z)}$.
- Retaining of the parameters for time and frequency and optimization of scale and chirp rate parameters by using the local behavior of the optimal frequency rate.
- Update of the residual with the approximated chirp atom $g_{(s,u^*,z^*,c)}$.

The analysis of EEG signals is equivalent to dealing with real-valued signals. This observation leads to an adaption of the ridge pursuit algorithm that increases the performance of the decomposition but preserves the computational cost of complexity $O(nN\log^2 N)$ for both variants of the ridge pursuit.

Real-valued chirp atoms are defined by the introduction of an additional parameter φ characterizing the phase of the atom:

$$g_{(s,u,z,c)}(t) = \frac{1}{\sqrt{s}} * g\left(\frac{t-u}{s}\right) \times \cos\left(z(t-u) + \frac{c}{2}(t-u)^2 + \varphi \right)$$

.

The steps of the ridge pursuit for real-valued input signals are:

- Computation of the inner product $\left\langle R^{m-1}x, g_{(s,u,z)} \right\rangle$ with every complex Gaussian Gabor atom.
- Derivation of the optimal phase φ by using an orthogonal projection and subsequent selection of the best real-valued Gaussian Gabor atom $g_{(s^*,u^*,z^*)}$.
- The "locally optimal" values for s and c are then found through a parabolic interpolation.
- Next, the inner product $\left\langle R^{m-1}x, g_{(s,u^*,z^*,c)} \right\rangle$ is calculated and the corresponding optimal phase is selected to obtain the best real-valued chirp atom.
- Computation of the new residual with the approximated real-valued chirp atom.

5.1.3 Adaptive Gaussian chirplet decomposition

This approach is based on the assumption that the signal under analysis can be broken down into merely a few dominant leading terms. This proved to be the case during trials of the first two algorithms with epileptic patterns.

The simplification proposed for this algorithm is based on the principle of finding the parameters of the desired atom by solving a conventional curve fitting problem. The application of different testing points leads to a set of equations for the parameters and replaces the optimization process in previous approaches.

This algorithm can be broken down in the following steps:

- The calculation is based on a good initial estimation of the parameters and it seems that the method is independent of the initial chirp rate (Yin and coauthors, 2002). Therefore, the chirp rate can be set to zero in the beginning and once again the best Gaussian gabor atom is computed as a starting point.
- The optimal parameter set is estimated by building up a system of linear equations with six testing points. The points are obtained by varying the frequency around the initial position for two different points in time.
- Updating the residual with the atom corresponding to the parameters obtained by the curve fitting problem.

During every iteration stage only six inner products have to be computed consequently reducing the computational effort significantly.

Plots of the optimal chirp atoms for all of the algorithms have been realized for comparison with inherent Matlab spectrograms and spectrograms based on the calculation of the Wigner-Ville distribution of the found atoms.

Results

1) Standard matching pursuit

Variation of the parameters:

The performance of the algorithm improves for smaller values of the scaling factor a; however the computational costs increase considerably too. A good compromise offers the choice a=8.
Variations of the frequency or chirp rate sampling steps provide no significant enhancement. Good and realistic values are given by $dz = \frac{\pi}{2}$ and dc=0.004.
Values for the time sampling step that are bigger than 1 lead to markedly less accurate decompositions. However, once more much smaller values than 1 increase the computation time drastically.

The boundaries for the chirp rate given in the paper of Gribonval (2001) can be adapted to reduce the computation time because realistic values of the chirp rate parameter are very small for EEG signals. In test runs the values π and $-\pi$ were used, although even a limitation to $\pm\frac{\pi}{2}$ would be possible.

Computation time:

The computational effort was inacceptable for use in online detection systems.

Performance:

This algorithm proved to decompose periodic signals as well as signals containing a chirp. It was especially useful for periodic signals because the chirp atoms could adapt very nicely to the signal. However, spikes in the signal were recognized as well, although the standard matching pursuit needs more iterations to decompose such parts of the signal than the ridge pursuit. Moreover, the sensitivity was smaller than for the ridge pursuit.

Conclusion:

On the one hand, the performance of the algorithm proved to be very satisfactory but the computational costs can't be tolerated. A significant improvement would be the composition of a chirp dictionary beforehand but this solution would require a large space for memory location.

2) Ridge pursuit:

Variation of the parameters:

Basically, the same parameter values as in the previous algorithm proved to work best. The only notable difference in parameter variation was the fact that alterations of the scaling factor did not result in significant performance improvements.

Break condition:

The ridge pursuit was tested on its ability to decompose periodic signals and signals containing distinct chirps. To this end break conditions were applied ending the computation if a defined fraction of the original signal energy was achieved.

E1…Energy of the original signal
E2…Energy of the residual

Break condition	Number of iterations	
	Periodic signal	Signal with chirp
E2 < (E1)/2	8	16
E2 < (E1)/3	16	33
E2 < (E1)/4	32	47
E2 < (E1)/5	50	58

That EEG signals are only composed of a few dominant components was supported by the fact that the first couple of iterations reduced the energy by far the most. The periodic signals could be broken down into their principal components - leading to a reduction in their energy – much more quickly because a greater width of the decomposing chirp atom could be chosen. Signals holding one or several chirps had to be dismembered through a number of smaller individual chirps before the remaining periodic signal could be tackled with broader chirps. The table shows that the difference dwindles if the break condition is chosen smaller i.e. for a larger number of iterations.

Computation time:

Computation time is incomparably better i.e. smaller than for the matching pursuit. Decomposition to half of the signal took several minutes. Of course, this still seems rather long for online detection; however in application the signal doesn't need to be broken down to half its energy but two or three iterations are sufficient to determine the presence of a chirp in the recordings.

Performance:

The algorithm proved to be efficient for periodic signals as well as ones including chirps by adapting very sensitively to the data. Thus the ridge pursuit is able to decompose the underlying recordings significantly faster than the standard matching pursuit while simultaneously displaying more accuracy.

Conclusion:

All in all, the ridge pursuit proved to be the most useful algorithm of the three ones that were tested. The computational costs could be drastically reduced and the performance showed very good adaption to and representation of the signal under analysis. This method seems worthwhile to be explored further for possible incorporation into the PWA module in the EpiScan system.

3) Adaptive Gaussian chirplet decomposition

Variation of the parameters:

The same parameter values as in the first two approaches were employed and variations in all of them did not alter the performance of the method markedly.

Break condition:

The application of break conditions in terms of fractions of the energy proved to be futile as even a reduction to half of the energy took very long.

Computation time:

The computational costs could be once again improved by several minutes for a large number of iterations compared to the ridge pursuit.

Performance:

The algorithm usually detects the same first few components as the matching pursuit or the ridge pursuit (especially for signals containing a chirp) but was rather inefficient at composing them. The scaling factor s was mostly very small and as a result the decomposition of a signal component took more iterations.

Conclusion:

Although this approach showed the best computational time, its performance was the least accurate and satisfying. Even if only the first few components are desired, the algorithm doesn't always provide the right ones due to its under-performing adaption to the EEG recordings.

Summary

The ridge pursuit displayed the best compromise of computational effort and good signal representation. It proved to be a promising tool for EpiScan algorithm improvement and will be tested furthermore for introduction into the third module instead of the periodic waveform analysis.

6 Automated Spike detection

The following section discusses an automated spike detection algorithm developed by the Safety and Security Department of the Austrian Institute of Technology (AIT). It is intended to objectively and accurately recognize inter-ictal epileptic discharges in long-term EEG recordings employed in pre-surgical evaluation, thereby sparing medical experts time and effort.
This approach is based on a multi-level system and incorporates multi-channel methods as well as patient-adaptive processes making learning phases obsolete. Due to the combination of several methods, the algorithm is able to adapt to patient-specific spike morphologies and is able to detect a high variability of different spikes.
Figure 6.1. illustrates the setup of the method which comprises an initial spike detection followed by a clustering module and a final detection step.

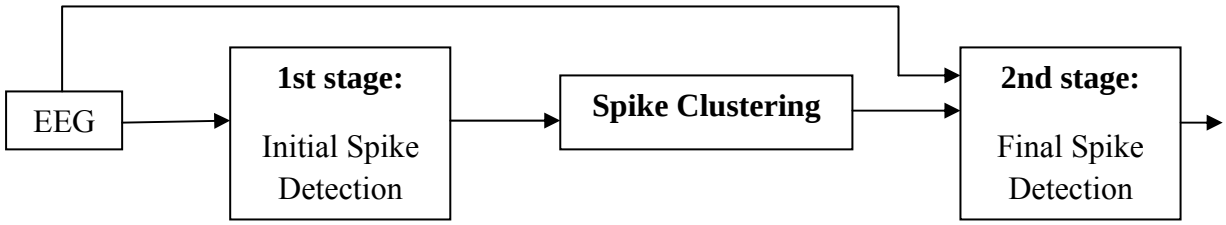


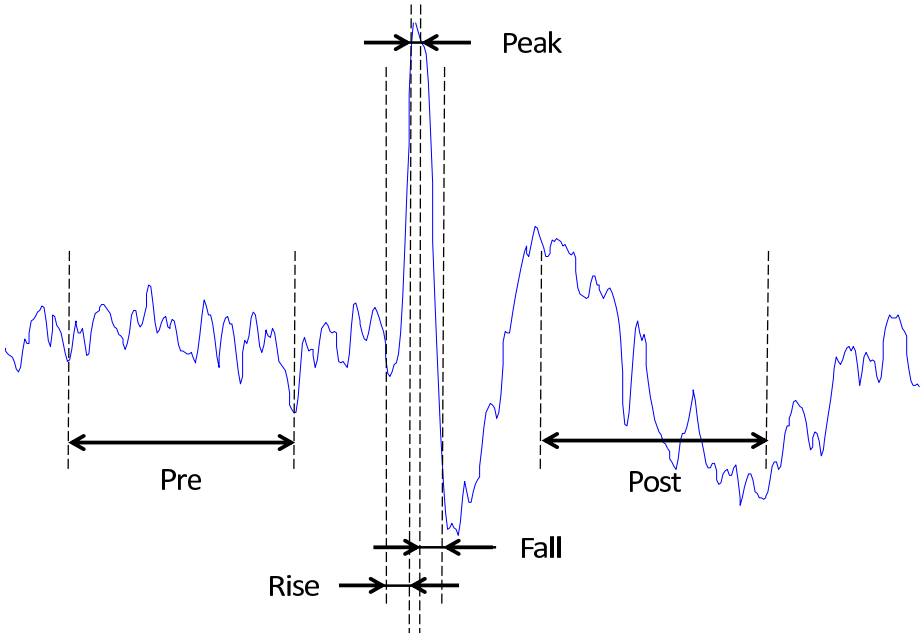
Fig. 6.1. Overview of the steps used in the automated spike detection algorithm at AIT

6.1 Initial Spike Detection

In the first step spike candidates are labeled by dividing the spike into several segments that can be identified through their individual features. Those segments have a specific temporal definition with respect to each other. This approach defines six different parts:

- Pre, characterizing the signal before the spike occurs
- Rise, declaring the ascent
- Peak, describing the tip of the spike
- Fall, designating the downward slope
- Wave, indicating a wave if a spike-wave complex is encountered
- Post, denoting the part of the signal after the spike

Each of these segments is characterized by predefined features that are extracted from the input signal which is designated by e.g. a filtered version of the EEG recordings. The features are calculated using window functions gliding over the data applying different norms for every individual segment. The input has to be dynamic but the module corrects variations in effective delays automatically.



Segments are characterized by the following parameters:

- their label (e.g. “Pre”)
- the name of the input signal (e.g. “derivation”)
- the clipping limits for the input (e.g. “Inf,-Inf”)
- the boundaries of the computation interval of the norm (e.g. “-10,10”)
- the type of norm which is used for the calculation (e.g. “2”)

The feature parameter set consists of

- the segment on which the conditions are applied (e.g. “Fall”)
- the boundaries of the feature vector (e.g. “1000,Inf”)
- a name if the feature shall be displayed (before thresholding)

The computation ensues individually for every channel and every specified pattern. First of all, the input signals are clipped according to the defined limits. Then the norm is calculated in a gliding window for each of the segments and the results are compared with their respective thresholds. The thresholding produces ones or zeros if the feature conditions are met or not. Those outputs are AND-connected for all segments of the pattern.

Final markers are generated by connecting all channels and templates with OR. All of the blocks resulting from the indicator function are shown separately and every channel containing at least one pattern within the predefined time interval is displayed.

The algorithm is summarized in Figure 6.3.:

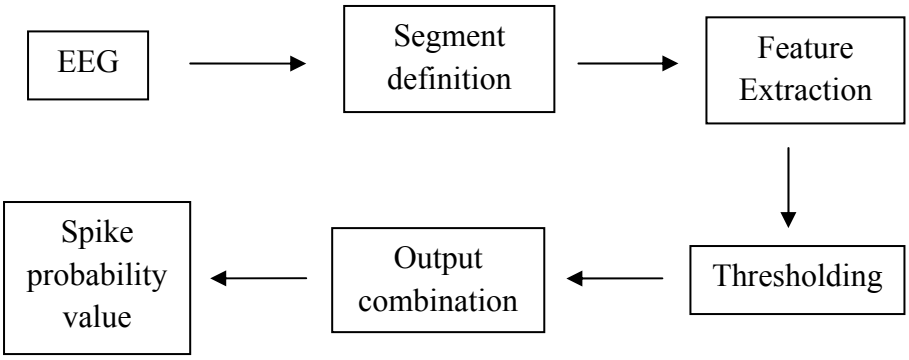


Fig. 6.3. Different stages in the initial spike detection

For every segment the delay of the input and the delay caused by the computation of the norm cumulate but they may have different values. Therefore, they have to be aligned before combination. The module uses an overlap-add approach for the cyclic processing of the EEG data.

In the following chapters Hidden Semi-Markov Event Sequence Models are applied to the thresholded features for more accurate spike detection as statistical models allow loosening of the rigid time definition of the segments and the duration intervals between them.

6.2 Spike Clustering

The spike markers that were detected in the first step of the algorithm are now clustered into various groups. Hierarchical clustering methods are used to assign spikes into specific groups according to selected features like source origin and morphology. Representative averages are built over each of these clusters and those that seem to be of interest can be selected as patient-specific templates for the next step.

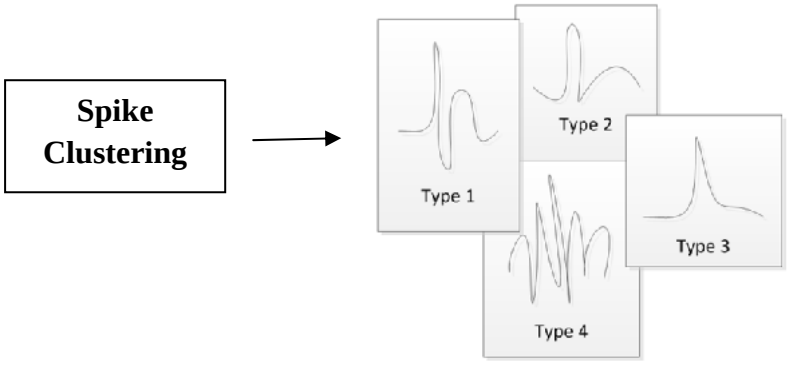


Fig. 6.4. Example of clustering into different groups according to spike morphology

6.3 Final Spike Detection

For the second detection stage statistical parameters are derived from the selected spike clusters and the EEG background noise. Statistical hypothesis testing of these arguments together with the normal EEG provide a test statistic value for every individual sample with higher values increasing the probability of spike detection. All channels of the reference montage are used for the calculation of the test statistic to recognize spike-like artifacts in the data. This distinction is possible by taking into account the distribution of the spike candidates over all electrodes. Patterns similar to the selected patient-specific templates increase the probability of an actual spike whereas arbitrary spread across the channels more likely indicates an artifact.

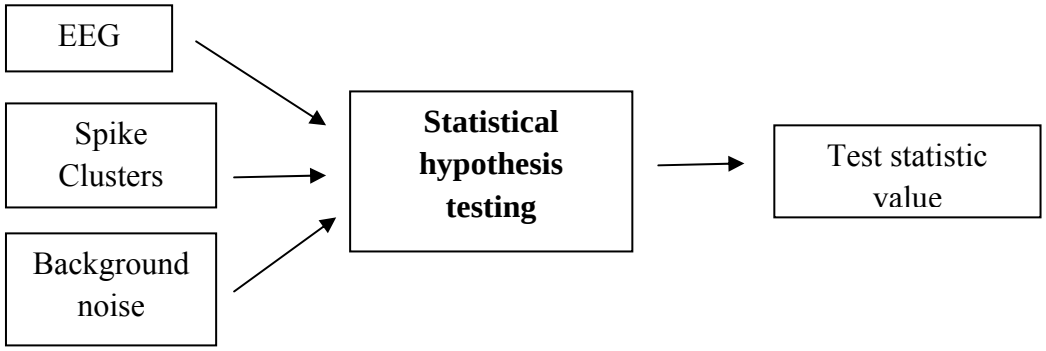


Fig. 6.4. Summary of the final spike detection steps

7 Application of Hidden Markov Models

Over the years signal models have proved to be very useful as many real-world dynamics can be characterized by observable outputs. Said models allow for theoretical descriptions of systems and procedures leading to better understanding of the underlying processes, improvement of the efficiency or costs and predictions about the outcome. Moreover, information about the signal source can be gained without actually employing the source for signal production which in some cases may be tedious or expensive.

Two types of model approaches are possible for signal analysis: deterministic models or statistical models. Whereas the first class makes use of distinctly determined parameters of the signal, stochastic modeling works only with statistical properties of the signal. The latter assume that the signal under analysis can be described as a parametric random process with parameters that can be well defined.

The first two sections will briefly introduce the ideas of Markov chains, Hidden Markov Models and Semi-Hidden Markov Models as well as the problems that have to be solved before moving to the extension to the more complicated Hidden Semi-Markov Event Sequence Models (HSMESMs).

7.1 Hidden Markov Models

A simple, discrete Markov chain consists of a set of N states that can be visited by the system at equidistant time intervals with specific probabilities (state transitions) related to each state.

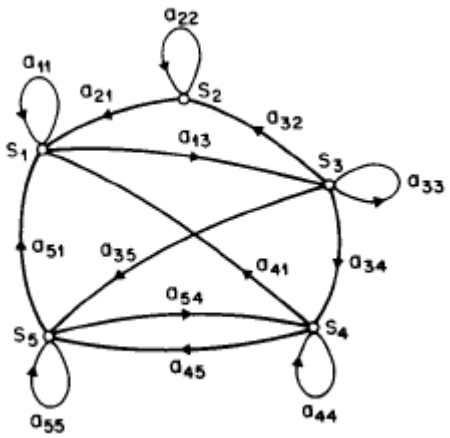


Fig. 7.1. A discrete Markov chain with states s_i and transitions a_{ij} (Rabiner, 1989)

Usually, the probability of being in a specific state q_t at time t requires the description of all predecessor states $q_1, \dots, q_{t-1}$. However, for first-order Markov processes the state transitions can be written as $a_{ij} = P[q_t = S_j | q_{t-1} = S_i]$ for $1 \leq i, j \leq N$ thereby assuming time

independence for a_{ij}. This leads to the following properties of the transition probabilities: $a_{ij} \geq 0$ and $\sum_{j=1}^{N} a_{ij} = 1$.

Discrete Markov chains are observable as they produce outputs that match with the states of the mode. An example would be a model of the daily temperature: the temperature could be classified into four states ($\leq -10, (-10,0), [0,10), \geq 10$) leading to a 4x4 transition probability matrix A=$\{a_{ij}\}\ for\ i,j = 1, \ldots, 4$. Then the probability for a specific observation sequence O=$\{S_2, S_2, S_1, S_2, S_3\}$ under the model would be given by $P(O|Model) = P[S_2] \cdot P[S_2|S_2] \cdot P[S_1|S_2] \cdot P[S_2|S_1] \cdot P[S_3|S_2]$ where $P[S_i] = P[q_1 = S_i], 1 \leq i \leq N$ denotes the initial probability of state i.

In Hidden Markov Models (HMMs) the states associated with the observations are not directly visible but the observation series is emitted by the underlying state sequence with a certain probability. Therefore, HMMs are doubly embedded stochastic systems consisting of the unobservable stochastic state process and the detectable observations which are emitted by each state according to probabilistic functions, i.e. the observation sequence is produced by a second stochastic process.

The idea of HMMs can be illustrated with a 2-stage model of the weather. The states are Rainy or Sunny with initial probabilities (dotted arrows) and transition probabilities between the states (block arrows) or self-transition (black arrows). The possible observations consist of Work, Sports or Shopping and the probabilities of the state Rainy or Sunny to emit them are visualized by blue or red dashed arrows.

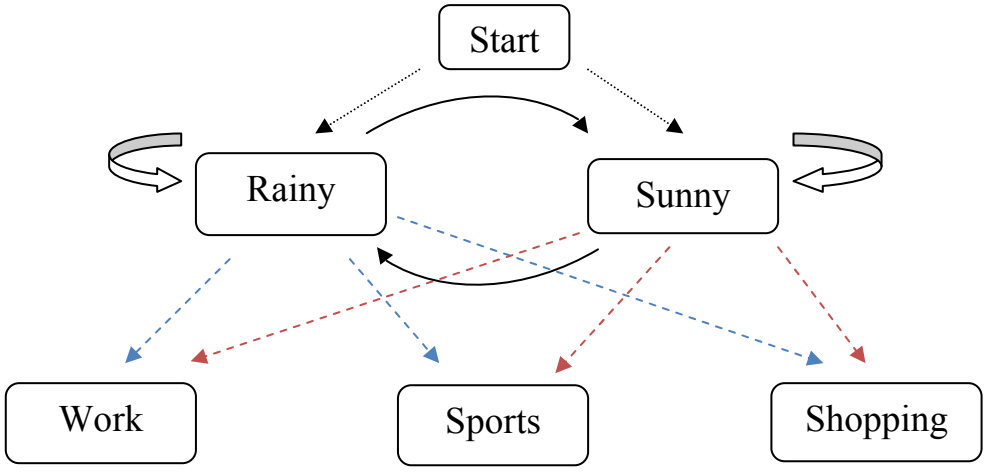


Fig. 7.2. Example of an HMM with the states sunny and rainy and the observations work, sports and shopping

In summary, it can be stated that an HMM consists of the following elements:

- The set of hidden states $S = \{S_1, \ldots, S_N\}$ with N denoting the number of possible states in the model (e.g. rainy and sunny). The state visited at time t is still labeled q_t. The

possible interconnections between the individual states can be defined through the state transitions and mostly describe some physical properties of the process under analysis.

- The set of observations that can occur in every state $O = \{o_1, \dots, o_M\}$ with M indicating the number of distinct observations (e.g. work, sports and shopping). They correspond to the visible output of the system that shall be modeled.
- The state transition probability matrix A=$\{a_{ij}\}$ where the state transitions are defined as $a_{ij} = P[q_t = S_j | q_{t-1} = S_i]$ for $1 \leq i, j \leq N$. If certain states in the model shall not be interconnected the respective a_{ij} is set to zero.
- The observation probability matrix B=$\{b_j(k)\}$ for state j with $b_j(m) = P[o_m \; at \; t | q_t = S_j]$ for $1 \leq j \leq N, 1 \leq m \leq M$.
- The initial state probability distribution $\pi = \{\pi_i\}$ where $\pi_i = P[q_1 = S_i]$ for $1 \leq i \leq N$.

If the described parameters of the model are chosen appropriately, they can be used either to produce an observation sequence or to describe how a given sequence was created by a model performing following steps:

1. Choice of a starting state $q_1 = S_i$ according to the initial probability distribution π.
2. Selection of an observation o_t in compliance with the observation probability distribution B.
3. Corresponding to the state transition matrix A the next state $q_{t+1} = S_j$ is visited.
4. Set t=t+1 and repeat the stages 2 and 3 until t=T.

Induced by the definition and shape of Hidden Markov Models three basic questions arise if applied to real-world systems. Subsequently, the problems and their solutions will be briefly summarized but more detailed descriptions will be given in the section concerned with HSMESMs.

1) The evaluation problem

For this question the observation sequence and the model $\lambda = (A, B, \pi)$ are given while the probability of the observation sequence given this specific model $P(O|\lambda)$ is sought-after. Therefore, information is obtained about how well the model does fit to the given observations.

The solution to this problem is usually calculated by a Forward-Backward algorithm which uses inference to compute the posterior marginals for every hidden state. The computation ensues by breaking the task down into subtasks; a procedure that is called dynamic programming.

2) The decoding problem

Again the observation sequence and the model are given, however now the goal is to uncover the hidden state sequence that produced the observations. The important choice in this part lies with the selection of the optimality criterion, e.g. should the individual best states be found or the most likely sequence of states? This decision depends on the plausibility of a specific criterion taking into consideration the real-world process to which the model shall be applied. However, in most applications the single state succession $Q = \{q_1, \dots, q_t\}$ shall be optimized, i.e. maximization of $P(Q|O, \lambda)$.

This problem can be solved using the Viterbi alignment which is very similar to the forward computation in the last problem except for the maximization procedure that substitutes the summation.

3) The learning problem

The last issue that has to be considered is the adjustment of the model parameters so that they optimally describe the generation of the observation sequence, i.e. maximization of $P(O|\lambda)$. Obviously, the training task is crucial to adapt the model optimally to the real-world system in order to produce useful and accurate predictions.

Training of the model parameters proves to be most difficult as no analytical solution has been found until now. One possible answer lies with local maximization of $P(O|\lambda)$ employing iterative computation such as the Baum-Welch algorithm. The Baum-Welch procedure is a type of expectation-maximization method and makes use of maximum likelihood estimates and posterior mode estimates for the model parameters.

It has to be noted that the computation, especially for problem 1 and 2, suffers from severe underflow as the forward variable tends to head exponentially towards zero. Such being the case scaling is necessary for feasible algorithm performance. Applying the scaling factor appropriately to the forward and backward variable leads to exact cancelling. However, caution has to be exercised when computing $P(O|\lambda)$ as the scaled forward values are used and the log function of $P(O|\lambda)$ has to be computed.
For the Viterbi algorithm the log function can be used from the beginning instead of scaling reducing computational effort and eliminating numerical issues.

The choice of initial parameters decides if the local maximum is also the global maximum of the likelihood function indicating the importance of this selection. Experience showed that random or uniform initial values are suitable for A and π whereas better B parameter estimates significantly improved the performance. In this case initial values can be obtained through segmentation of the observations into states followed by averaging over the observation subsequence within the hidden states.

Selection of the model type and consequently solution of the problems have always to be considered in relation to the real-world application to produce meaningful information about the process. An example for successful application of HMMs is the employment to specific tasks in speech recognition.
For many procedures the limitations of HMMs (e.g. independence of successive observations, representation of the observation distributions by Gaussian mixture densities) prove to be too restrictive, requiring the expansion to other models.

7.2 Hidden Semi-Markov Models

Equivalent to HMMs, Hidden Semi-Markov Models (HSMMs) are statistical models where the embedded "hidden" stochastic process is constituted by a Semi-Markov chain. A Semi-Markov chain still consist of a finite set of observable states S=$\{S_1, \dots, S_N\}$ and the transition probabilities a_{ij} between them; however, the time between the transitions is no longer equidistant but modeled by a probability distribution. Simplified, this means that for every state that is visited, a duration d_l is chosen from D=$\{d_1, \dots, d_L\}$ according to the duration density $p_{q_i}(d_l)$ for $1 \leq i \leq N, 1 \leq l \leq L$ in state i.

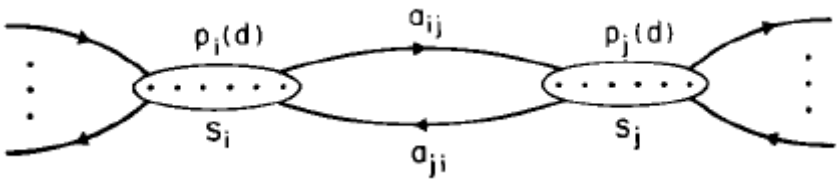


Fig. 7.3. Example of a Semi-Markov chain with transition probability $p_i(d)$ (Rabiner, 1989)

The application of this principle to extend HMMs leads to a variable duration for every state and the possibility of emitting a number of observations while being in one state.

Therefore, the elements of a general HSMM (without simplifying assumptions about state transitions, durations or observation distributions) are:

- The set of hidden states $S = \{S_1, \dots, S_N\}$ with N denoting the number of model states and q_t describing the state the system visits at time t.
- The set of observation symbols possible to be emitted in every state $O = \{o_1, \dots, o_M\}$ with M indicating the number of observations.
- The state transition probability matrix A=$\{a_{(i,d_l),(j,d_k)}\}$ where the state transitions are defined as $a_{(i,d_l),(j,d_k)} = P\left[q_{t+1}, \dots, q_{t+d_k-1} = S_j \middle| q_{t-d_l+1,\dots,} q_{t-1} = S_i\right]$ for $1 \leq i, j \leq N, 1 \leq l, k \leq L$ and $a_{(i,d_l),(i,d_k)} = 0$ for $1 \leq i \leq N, 1 \leq l, k \leq L$ because it is assumed that state i is visited for exactly d_l observations. The state transitions can be simplified

if additional assumptions are made, e.g. the transition to state j is independent of the duration of state i or the duration of state j is independent of state i (both assumptions together lead to explicit duration HMMs where $a_{(i,d_l),(j,d_k)} = a_{ij}p_j(d)$).

- The observation probability matrix B=$\{b_j(k)\}$ for state j with

  $b_{j,d_l}(m_1, \dots, m_{d_l}) = P\left[o_{m_1,\dots,m_{d_l}} \middle| q_t, \dots, q_{t+d_l-1} = S_j\right]$ for $1 \le j \le N, 1 \le l \le L, 1 \le m \le M$. If the observations are independent: $b_{j,d_l}(m) = \prod_{s=1}^{d_l} b_{j,d_l}(m_s)$.

- The initial state probability distribution $\pi = \{\pi_{i,d_l}\}$ where $\pi_{i,d_l} = P[q_1, \dots, q_{d_l} = S_i]$ for $1 \le i \le N$.

- The set of durations D=$\{d_1, \dots, d_L\}$ describing how long every state can be visited, i.e. how many observations are emitted.

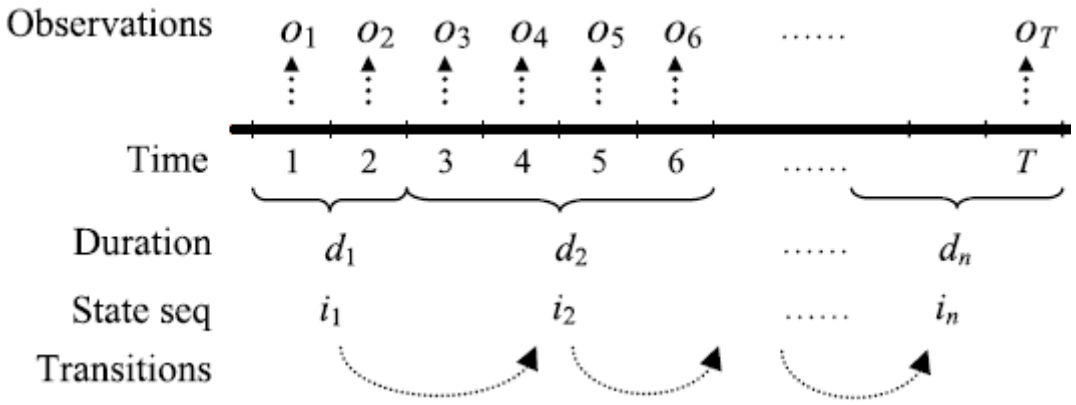


Fig. 7.4. Illustration of a general HSMMs with states $i_1, \dots, i_n$, respective durations $d_1, \dots, d_n$ and the emitted observations in every state (o_1, o_2 in i_1; $o_3, \dots, o_6$ in i_2; …) (Yu, 2009)

The generation of an observation sequence amounts to the following:

1. Selection of an initial state $q_1 = S_i$ with duration d_l according to π_{i,d_l}
2. Choice of an observation sequence $o_{m_1,\dots,m_{d_l}}$ corresponding to the joint observation probability $b_{j,d_l}(m_1, \dots, m_{d_l})$.
3. Transition to the next state $q_{d_l+1} = S_j$ with duration d_k due to $a_{(i,d_l),(j,d_k)}$.
4. Repeat steps 2 and 3 until q_T is reached.

The three basic problems remain the same for HSMMs and again the Forward-Backward algorithm, the Viterbi alignment and the Baum-Welch algorithm can be used to solve them if the respective formulas are adjusted accordingly. Equally, the same implementation issues as

underflow have to be addressed for HSMMs (e.g. implementing the algorithms in the logarithmic domain).

HSMMs have become more and more prominent in various fields over the last decade. Application areas are brain functional MRI sequence analysis, protein structure prediction, satellite propagation channel modeling, Internet traffic modeling, financial time series modeling, event recognition in videos and semi-conductor manufacturing to mention only a few examples. Depending on the real-world application there exist also manifold types of and extensions to HSMMs.

7.3 Hidden Semi-Markov Event Sequence Models

A disadvantage of HSMMs is the assumption of piecewise stationarity which makes them more suited for processes that are segmental in nature. However, especially biomedical signals like spike discharges in EEG or hemodynamic responses in functional MRI do not fit this scheme but are localized in time, nonstationary and often obscured by artifacts. Nevertheless HSMMs allow for the softening of strict shape and time specification in pattern detection. Therefore, a modification of HSMMs could prove to be very useful in biomedical areas dealing with event-based systems.

The idea is to place emphasis on the "true" events that emit only one observation and employ the principle of HSMMs to the duration between two events. Thus the stationarity assumption is limited to the time between events of interest during which observation sequences are emitted consisting of "false" events and missed observations. The models employing this principle are called Hidden Semi-Markov Event Sequence Models and will be described in detail subsequently.

In order to be able to apply an HSMESM, features have to be extracted from the signal in a pre-processing stage. The output of this step results in observations, called events e_l at time t_l leading to an event sequence $E = e_1, \dots, e_L$ of L recognized events at time points $T_E = \{t_1 = 1, \dots, t_L = T\}$. The events e_1 and e_L are introduced as they are needed for duration modeling but describe only fictive events. T_E is a subset of the time points of observations $T_O = \{1, \dots, T\}$ which corresponds to the complete observation sequence $O = o_1, \dots, o_T$. This means that E is made up of true positive events (tpes) and false positive events (fpes) whereas O also includes the missed observations or null events, e.g. $E = e_1, e_2, e_3, e_4, e_5$ and $O = e_1, \emptyset, \emptyset, \emptyset, e_2, \emptyset, e_3, e_4, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, e_5$.

7.3.1 Model construction

As HSMESMs are extensions to HSMMs they contain two stochastic processes, the hidden one, emitting the observed or null events over time, and the observable process, modeling the relation between the observations and the hidden states which produce them.

To sum up, after the preprocessing an observations sequence is obtained, constituting a mix of tpes, fpes and null events, representing the input to the model. The model may visit $C-2$ possible states and 2 fictive states (one at the end and one at the beginning). q_t labels the state that is visited at t and $q_t^{t'}$ denotes a state sequence with $q_t^{t'} = S_{ij}$ meaning $q_\tau = S_{ij}$ for $t \leq \tau \leq t'$. Accordingly, $o_t^{t'}$ defines a sequence of observations $o_t, \dots, o_{t'}$.

The parameter set of an HSMESM consists of the following elements:

- The set of hidden states $S = \{S_{tpe}, S_{agg}\}$ consisting of the true positive event and the aggregate states defined by

$$S_{tpe} = \{S_i, i = 1, \dots, C\}$$

$$S_{agg} = \{S_{ij}, i = 1, \dots, C-1, j = 2, \dots, C\}$$

  The S_i states correspond to a Markov chain which emits true positive events (tpes) at $t_l \; in \; T_E$. The start and final states ($i = 1$ and $i = C$) are introduced to model the fictive events e_1 and e_L respectively.

  The S_{ij} states on the other hand represent the semi-Markov process which outputs fpes at $t_l \; in \; T_E$ and zeros at $t \; in \; T_{E'}$. Those states again are comprised of two subsets: $S_{ij} = \{S_{ij}^0, S_{ij}^{\emptyset}\}$ denoting the fpe states (S_{ij}^0) and the missed observations ($S_{ij}^{\emptyset}$).

- The state transition probability matrix A=$\{a_{ij}\}$ where

$$a_{ij} = P\left[q_{t_{k+1}} = S_j \; or \; S_{ij} \middle| q_{t_k} = S_i\right] \text{ for } 1 \leq i \leq C-1, 2 \leq j \leq C, \forall t_k \; in \; T_E$$

- The set of durations between the tpe states D=$\{p_{ij}(.)\}$ where

$$p_{ij} = P[d_{k_l} | q_{t_l} = S_j, \; q_{t_{k+1}}^{t_{l-1}} = S_{ij}, q_{t_k} = S_i] \text{ for } 1 \leq i \leq C-1, 2 \leq j \leq C,$$
$$(t_k, t_l) \; in \; T_E \times T_E \; such \; that \; t_l - t_k = d_{k_l}, d_{k_l} \geq 1$$

- The set of tpe observation probability densities B=$\{b_j(.)\}$ where

$$b_j(o_{t_l}) = P\left[o_{t_l} \middle| q_{t_l} = S_j\right] \text{ for } 1 \leq j \leq C, t_l \; in \; T_E$$

  with the initializations $b_1(o_{t_1} = e_1) = 1$ and $b_C(o_{t_L} = e_L) = 1$

- The set of aggregate observation probability densities B'=$\{b_{ij}(.)\}$ where

$$b_{ij}(o_{t_l}) = P\left[o_{t_l} \middle| q_{t_l} = S_{ij}^0\right] \text{ for } 1 \leq i \leq C-1, 2 \leq j \leq C, t_l \; in \; T_E$$

- The emission probability matrix E=$\{e_{ij}\}$ where

$$e_{ij} = P\left[q_t = S_{ij}^0 \middle| q_t \in S_{ij}\right]$$
$$1 - e_{ij} = P\left[q_{t'} = S_{ij}^{\emptyset} \middle| q_{t'} \in S_{ij}\right]$$
$$\text{for } 1 \leq i \leq C-1, 2 \leq j \leq C, t \; in \; T_E, t' \; in \; T_{E'},$$

Notably, the parameter e_{ij} models the mean false alarm rate encountered between the tpes observed in the states i and j.

In HSMESMs aggregate states S_{ij} are inserted between the true positive events and can emit either fpes or null events. The S_{ij} characterize the transition between the states i and j and

have a duration d_{k_l} (actual states visited are $d_{k_l} - 1$) according to a certain duration distribution $p_{ij}(.)$ that models the time occupancy between two states i and j. During this time the observation subsequence $o_{t_{k+1}}^{t_{l-1}}$ is emitted for o_{t_k} emitted in S_i and o_{t_l} emitted in S_j. The output probability of S_{ij} is based on a latent Bernoulli process characterized by e_{ij}. This leads to the following probabilities for either an fpe or a null event:

$$P\left[o_t = e_l \middle| q_t = S_{ij}^0\right] = b_{ij}(e_l)$$

$$P\left[o_t = \emptyset \middle| q_t = S_{ij}^{\emptyset}\right] = 1$$

Due to combination with the emission probability matrix, the output probability in the aggregate states can be expressed as:

$$P\left[o_t \middle| q_t = S_{ij}\right] = b_{ij}(e_l) \cdot e_{ij} \cdot \mathbb{I}_{T_E}(t) + \left(1 - e_{ij}\right) \cdot \mathbb{I}_{T_E'}(t) = b_{ij}^{+}(o_t)$$

Assuming the conditional independence of observations, the output probability of the subsequence $o_{t_{k+1}}^{t_{l-1}}$ can be written as:

$$P\left[o_{t_{k+1}}^{t_{l-1}} \middle| q_{t_{k+1}}^{t_{l-1}} = S_{ij}\right] = \prod_{\tau=t_{k+1}}^{t_{l-1}} b_{ij}^{+}(o_\tau) = b_{ij}^{+}(o_{t_{k+1}}^{t_{l-1}})$$

Consequently, specification of an HSMESM requires the definition of A, B, B', D, E and the number of states C. Further on the parameter $\lambda = (A, B, B', D, E)$ will denote the whole set of model elements.

Observation: Setting $e_{ij} = 1, b_{ij}(.) = b_j(.), p_{ij}(.) = p_j(.)$ leads back to the standard HSMM formulation.

7.3.2 Generation of an observation\event sequence

Given the model set λ an observation sequence can be obtained by implementing following steps:

1. The process starts in the fictive tpe state S_1 at t_1 and emits e_1 according to $b_1(o_{t_1} = e_1) = 1$.
2. Being in the current state S_i at time t_k , $1 \leq k < L$, the next state S_j is selected in compliance with the state transition probability a_{ij}.
3. The inter-tpe duration d_{k_l} is determined by $p_{ij}(.)$ and S_j is reached at $t_l = d_{k_l} + t_k$, $l > k$.
4. The model remains in S_{ij} for $d_{k_l} - 1$ time points emitting fpes or null events corresponding to $b_{ij}^{+}(.)$.

5. Steps 2-4 are repeated until the final state S_C is visited at $t_L = T$ emitting e_L as prescribed by $b_C(o_{t_L} = e_L) = 1$.

The next three figures visualize the idea of HSMESMs:

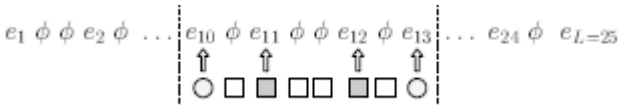


Fig. 7.5. Example of an emitted observation sequence consisting of tpes (circles), fpes (dark squares) and null events (white squares) (Thoraval, 2002)

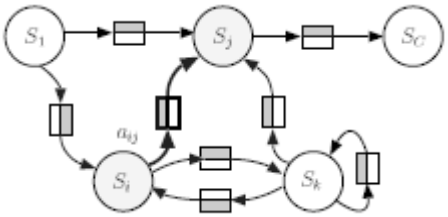


Fig. 7.6. Markov chain topology: transitions between tpe states during which observations (fpe or null events) take place (Thoraval, 2002)

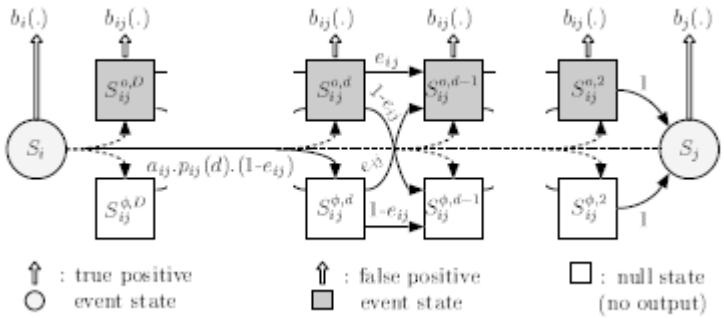


Fig. 7.7. Relation of HSMESMs to standard Markov chains: Modeling the aggregate states as two simultaneous Markov chains, one for fpes and null events respectively, leads to an equivalent standard Markov process. The fpes are emitted by the output probability matrix $b_{ij}(.)$ during the Markov states S_{ij}^0 and null events during the states $S_{ij}^{\emptyset}$ without relation to a probabilistic function. While remaining in S_{ij} the model switches between the two chains (depending on e_{ij}), e.g. in the first figure the process would start in the lower line (emitting a null event), then switching to the upper chain, then back again, remaining there for another time point before switching back again and so on. (Thoraval, 2002)

7.3.3 Computational Aspects

Similarly to HMMs or HSMMs three basic questions have to be addressed to obtain feasible information for real-world problems:

- The evaluation problem: The conditional probability $P[O|\lambda]$ of a specific observation sequence given the model parameters is sought-after.
- The decoding problem: Calculation of the state sequence q^* that best explains the observed events O employing the prescribed model λ.

- The learning problem: The model parameter set λ shall be optimized using the likelihood criterion $L(\lambda) = \prod_{v=1}^{V} P[O^v|\lambda]$ for mutually independent training sequences $\{O^v\}, v = 1, \dots, V$.

As already proposed for an HMM framework the Forward-Backward algorithm, the Viterbi algorithm and the Baum-Welch algorithm will be applied in order to solve the three problems respectively.

7.3.3.1 The Evaluation Problem

The Forward-Backward procedure is one of the most efficient solutions to the first question. The forward and backward variables are defined in the following way:

$$\alpha_{t_l}(j) = P[o_1^{t_l}, q_{t_l} = S_j|\lambda]$$

$$\beta_{t_l}(i) = P[o_{t_{l+1}}^{T}|q_{t_l} = S_i, \lambda]$$

$\alpha_{t_l}(j)$ describes the probability of observing the subsequence $o_1^{t_l}$ and ending in the state S_j at time t_l given the model set λ. $\beta_{t_l}(i)$ on the other hand characterizes the conditional probability of starting in state S_i at the time point t_l and observing from then on the subsequence $o_{t_{l+1}}^{T}$ with the given model λ.
The two variables can now be computed by initializing them, thereby employing the fictive events e_1 and e_L, and subsequently using an induction procedure. The formulas used in the induction step are given by:

$$\alpha_{t_l}(j) = \sum_{i=1}^{C-1} \sum_{\substack{k=1, \\ 1 \le d_{k_l} \le D}}^{l-1} \alpha_{t_k}(i) \cdot a_{ij} \cdot p_{ij}(d_{k_l}) \cdot b_{ij}^{+}(o_{t_{k+1}}^{t_{l-1}}) \cdot b_j(o_{t_l})$$

$$for\ l = 2, \dots, L\ and\ j = 2, \dots, C$$

with the initialization: $\alpha_{t_1=1}(i)\ for\ i = 1, 0\ otherwise$ for the forward variable and by:

$$\beta_{t_l}(i) = \sum_{j=2}^{C} \sum_{\substack{m=l+1, \\ 1 \le d_{l_m} \le D}}^{L} \beta_{t_m}(j) \cdot a_{ij} \cdot p_{ij}(d_{l_m}) \cdot b_{ij}^{+}(o_{t_{l+1}}^{t_{m-1}}) \cdot b_j(o_{t_m})$$

$$for\ l = L-1, \dots, 1\ and\ i = 1, \dots, C-1$$

with the initialization: $\beta_{t_L=1}(j)\ for\ j = C, 0\ otherwise$ for the backward variable.

Computation of these two variables enables the evaluation of $P[O|\lambda]$ for any t_l in T_E by deploying:

$$P[O|\lambda] = \sum_{j=1}^{C} \alpha_{t_l}(j) \cdot \beta_{t_l}(j)$$

In the special case of $t_l = T$ the calculation is reduced to:

$$P[O|\lambda] = \alpha_{t_L}(C)$$

It should be noted that HSMESMs need less computation time if $L \ll T$ (accordingly to the extent of the difference) as the forward and backward variables are only computed for the tpes and fpes at the corresponding time points t_l in T_E.

For the purpose of facilitating the understanding of the workings of the Forward-Backward procedure, the induction step will be illustrated in more detail.

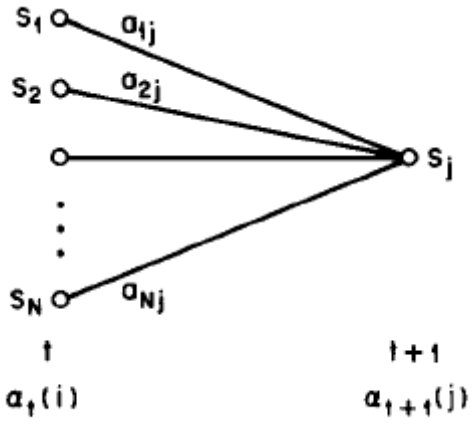


Fig. 7.8. Illustration of the induction step in the forward algorithm (Rabiner, 1989)

Figure 7.8. shows how the state S_j at time $t+1$ can be reached considering all possible pathways. Since $\alpha_t(i)$ denotes the probability of the observation subsequence o_1^t ending in S_i summation over $\alpha_t(i) \cdot a_{ij}$ for $i = 1, \dots, N$ leads to the probability of observing o_1^t and visiting S_j at $t+1$ over S_i at t. $b_{ij}^{+}\left(o_{t_{k+1}}^{t_{l-1}}\right)$ and $p_{ij}\left(d_{k_l}\right)$ then add the probability of observing $o_{t_{k+1}}^{t_{l-1}}$ in the aggregate state S_{ij} that has duration $d_{k_l} - 1$ between the time points t and $t+1$ during which the tpe states S_i and S_j are visited. Finally the probability $b_j(o_{t+1})$ is employed to obtain o_{t+1} in S_j which completes the observation sequence probability $P[o_1^{t+1}|\lambda]$.

Actually, the computation of the forward variable would suffice to solve the evaluation problem as the formula for $t_l = T$ implies. However, the backward variable is useful for the third problem and will therefore be visualized at this point as well.

The idea is essentially the same as for $\alpha_t(i)$ but now the assumption is that the model is in state S_i at the time t and the observation subsequence o_{t+1}^T from this point on until termination shall be calculated. a_{ij} and $\beta_{t+1}(j)$ account for the transition to S_j and the

subsequent probability to observe o_{t+2}^{T}. $p_{ij}(d_{l_m})$, $b_{ij}^{+}(o_{t_{l+1}}^{t_{m-1}})$ and $b_j(o_{t_m})$ capture the observation of $o_{t_{l+1}}^{t_{m-1}}$ in the aggregate state S_{ij} lasting $d_{l_m} - 1$ time points and finally observing o_{t_m} in S_j.

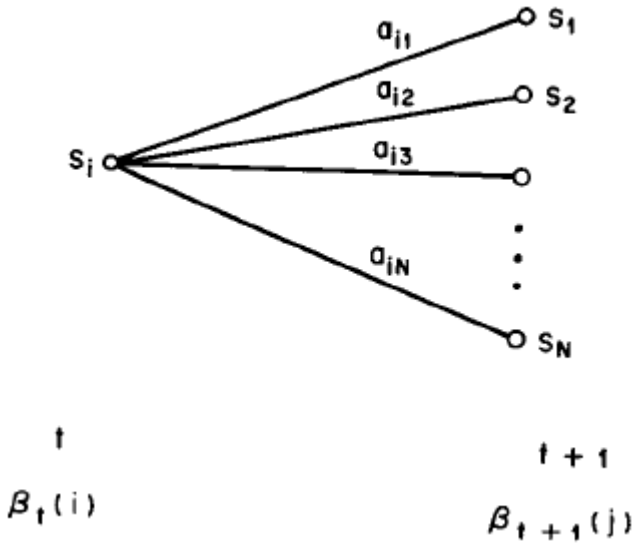


Fig. 7.9. Illustration of the induction step in the backward algorithm (Rabiner, 1989)

7.3.3.2 The Decoding Problem

As already in the HMM case, several solutions exist for this question depending on the application it is intended for and thus on the optimization criterion. Further on the Viterbi algorithm will be explained which serves to find the best overall state sequence $q^* = q_1^{T^*}$ (or the optimal pathway) through maximizing the conditional probability $P[O, q_1^T|\lambda]$ which is equal to maximizing $P[q_1^T|O, \lambda]$.

$$q^* = \max_{\{q_1^T\}} P[O, q_1^T|\lambda]$$

The Viterbi algorithm solves the problem of maximizing the a posteriori probability (MAP) by solving the identical problem of finding the shortest path through a specific graph through employing recursive methods. This principle is illustrated in the following graphs for a problem that has 4 states with certain allowed transitions (Fig. 7.10. a). Transformation into finding the most likely pathway among all possible paths leads to 7.10. b. The second figure visualizes an example by denoting certain weights or values to each transition (Fig. 7.11. a). The Viterbi algorithm determines the shortest path or the cheapest path through iteration.

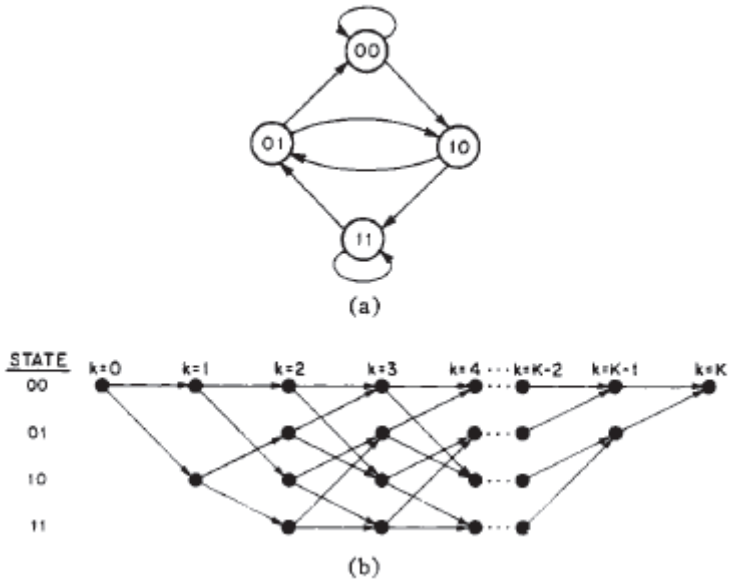


Fig. 7.10. Example graph for 4 states (a) and the according search for the most likely pathway (b).

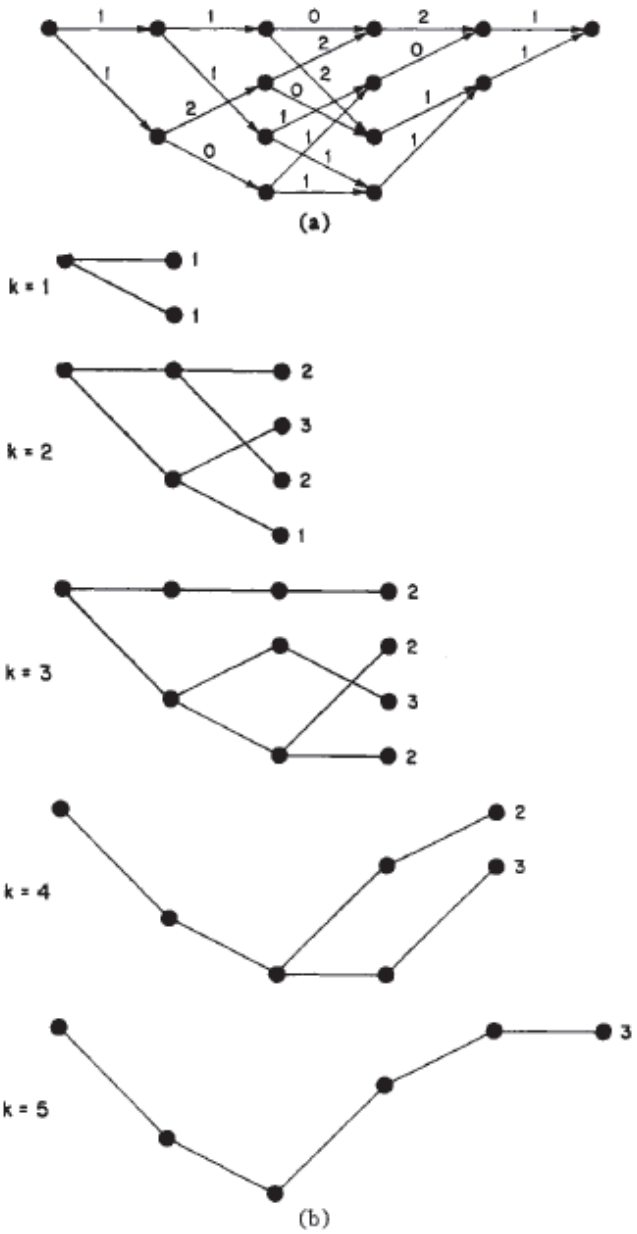


Fig. 7.11. Detailed depiction of the shortest pathway retrieval with specific weights (Forney, 1973)

The computation of the optimal state sequence given the observation sequence and the model parameter set requires the definition of the variable $\delta_{t_l}(j)$:

$$\delta_{t_l}(j) = \max_{\{q_1^{t_{l-1}}\}} P[o_1^{t_l}, q_1^{t_{l-1}}, q_{t_l} = S_j|\lambda]$$

The induction step of the maximization is rather similar to the one used for the forward variable:

$$\delta_{t_l}(j) = \max_{1\le i\le C-1} \max_{\substack{1\le k\le l-1 \\ 1\le d_{k_l}\le D}} \delta_{t_k}(i) \cdot a_{ij} \cdot p_{ij}(d_{k_l}) \cdot b_{ij}^{+}(o_{t_{k+1}}^{t_{l-1}}) \cdot b_j(o_{t_l})$$

$$l = 2, \dots, L \; and \; j = 2, \dots, C$$

with the initialization: $\delta_{t_1=1}(i)\ for\ i = 1, 0\ otherwise$.

In order to be able to extract the best state sequence it is necessary to employ an array $\varphi(j, l)$ that keeps track of the arguments (i, k) that correlate to the optimal states. For the backtracking the following notation will be used: $q_{\tau_r} = (i, k)$ for $q_{\tau_r} = S_i$ at $\tau_r = t_k$. The first step is to retrieve the best tpe state sequence $q^*_{tpe} = \left(q_{\tau_1}, \dots, q_{\tau_R}\right)^*$ of the hidden process. To this ends, starting at $q_{\tau_R} = (C, L)$ backtracking through the variable $\varphi(.,.)$ is employed for $r = R - 1, \dots, 1$:

$$q_{\tau_r} = \varphi(q_{\tau_{r+1}})$$

For practical applications the retrieval of q^*_{tpe} usually suffices as it provides the time points $\tau^*_{tpe} = \tau_1, \dots, \tau_R$ of tpe visited and the index sequence $i^*_{tpe} = i_{\tau_1}, \dots, i_{\tau_R}$ describing the tpe state class occupied at τ_r.

However, q^* can also easily be derived from q^*_{tpe} making use of the structure of HSMESMs:

$$q_t = q_{\tau_r}\ for\ t = \tau_r \text{ and } q_t = S_{ij}\ for\ \tau_r < t < \tau_{r+1} \text{ and } (q_{\tau_r}, q_{\tau_{r+1}}) = (S_i, S_j)$$

$$for\ t = 1, \dots, T\ and\ r = 1, \dots, R$$

Therefore, $\tau_1, \dots, \tau_R$ is a subsequence of the time points $t_1, \dots, t_L$ with $\tau_1 = t_1$ and $\tau_R = t_L$.

7.3.3.3 The Learning Problem

This part is concerned with the adjustment of the parameters of the model in order to optimally describe the realization of the observation sequence by the hidden state sequence. However, finding a solution to this problem proves to be rather difficult and leads only to local maxima of $P[O|\lambda]$.

For the special case of left-right models (e.g. modeling spikes in EEG signals), a single observation procedure is insufficient for training of the model parameters seeing that only very few observations are possible for every state. As a result, multiple observation sequences have to be applied in order to achieve meaningful parameter adjustments. This results in the maximization of the likelihood $L(\lambda) = P[O|\lambda] = \prod_{v=1}^{V} P[O^v|\lambda]$ for mutually independent training sequences $\{O^v\}, v = 1, \dots, V$.

The answer to this question is again provided by an iterative procedure, the Baum-Welch algorithm. It makes use of both, the forward and the backward variable, introduced in the chapter containing the evaluation problem. At first, the idea is illustrated for the HMM case to facilitate understanding. The variables $\alpha_t(i)$ and $\beta_t(i)$ are employed to ascertain the probability that the model is in state S_i at time t, having produced $o_1, \dots, o_t$, $(\alpha_t(i))$, transits to S_j at t+1, emitting o_{t+1}, $(a_{ij} \cdot b_j(o_{t+1}))$ and further on producing $o_{t+2}, \dots, o_T$ $(\beta_t(i))$. The next step involves the calculation of the estimation parameters describing how many times S_i

is visited and the number of expected transitions made from S_i to S_j. The obtained values are then used to reestimate the model parameters.

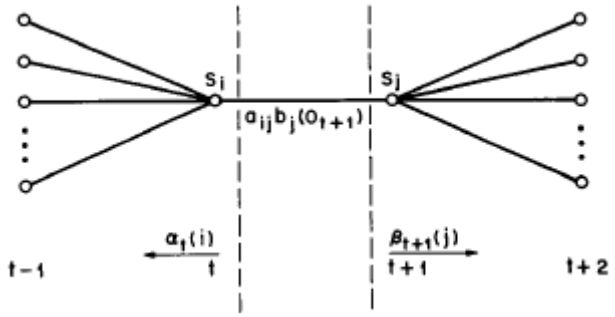


Fig. 7.12. Simplified visualization of the principle of the Baum-Welch procedure (Rabiner, 1989)

The derivation of the reestimation formulas for the HSMESM case can also be achieved directly through maximizing the auxiliary function $Q(\lambda, \lambda')$:

$$Q(\lambda, \lambda') = \sum_{v=1}^{V} \sum_{q^v} P[q^v, O^v|\lambda] \log P[q^v, O^v|\lambda']$$

with λ' being the auxiliary variable to λ and q^v, O^v being a particular state and observation sequence respectively. Baum and his colleagues proved that:

$$Q(\lambda, \lambda') \geq Q(\lambda, \lambda) \Rightarrow L(\lambda') \geq L(\lambda)$$

with the equation being equal if and only if $\lambda = \lambda$‘, which shows that the optimization of the auxiliary function does also lead to the (local) maximum of the likelihood function $L(\lambda)$. Iteratively replacing λ by λ' ultimately leads to convergence of the algorithm when some critical point is reached. The result of the optimization process is called a maximum likelihood estimate (MLE).

By defining the following variables:

$$\gamma_{ij}^{1v}(k) = P[q_{t_k^v} = S_i, q_{t_k^v+1} = S_j \; or \; S_{ij,} O^v|\lambda]$$

$$\gamma_i^{2v}(k) = P\left[q_{t_k^v} = S_i, O^v \middle| \lambda\right]$$

$$\gamma_{ij}^{3v}(t) = P\left[q_t = S_{ij}, O^v \middle| \lambda\right]$$

$$\gamma_{ij}^{4v}(k.l) = P[q_{t_k^v} = S_i, q_{t_k^v+1}^{t_l^v-1} = S_{ij,} q_{t_l^v} = S_j, O^v|\lambda]$$

the auxiliary function can be written as:

$$Q(\lambda,\lambda') = \sum_{v=1}^{V}\sum_{i=1}^{C-1}\sum_{j=2}^{C}\sum_{k=1}^{L^v-1}\gamma_{ij}^{1v}(k)\log\left(a'_{ij}\right) + \sum_{v=1}^{V}\sum_{i=1}^{C-1}\sum_{k=2}^{L^v-1}\gamma_{i}^{2v}(k)\log(b'_i(o_{t_k^v}^v))$$
$$+ \sum_{v=1}^{V}\sum_{i=1}^{C-1}\sum_{j=2}^{C}\sum_{k=2}^{L^v-1}\gamma_{ij}^{3v}(t_k^v)\log\left(b'_{ij}(o_{t_k^v}^v)\right)$$
$$+ \sum_{v=1}^{V}\sum_{i=1}^{C-1}\sum_{j=2}^{C}\sum_{m=1}^{2}\sum_{t\in T_m^v}\gamma_{ij}^{3v}(t)\log\left(e'_{ijm}\right)$$
$$+ \sum_{v=1}^{V}\sum_{i=1}^{C-1}\sum_{j=2}^{C}\sum_{k=1}^{L^v-1}\sum_{l=k+1}^{L^v}\gamma_{ij}^{4v}(k,l)\log(p'_{ij}(t_l^v - t_k^v))$$

with

$$T_1^v = T_E^v - \{1, T^v\}, T_2^v = T_{E'}^v - \{1, T^v\}\ and\ \ e_{ij1} = e_{ij}, e_{ij2} = 1 - e_{ij}\ .$$

The maximization procedure of $Q(\lambda,\lambda')$ can now be broken down into optimizing its individual terms separately. Employing Lagrange multiplier of differential calculus along with the following assumptions:

$b'_i(.)\sim N(\mu'_i,\sigma'_i), b'_{ij}(.)\sim N\left(\mu'_{ij},\sigma'_{ij}\right), p'_{ij}(.)\sim N(\mu d'_{ij},\sigma d'_{ij})$ and $o_{t_k^v}^v = e_k^v$

leads to the subsequent reestimation formulas for the model parameters:

$$a'_{ij} = \frac{\sum_{v=1}^{V}\sum_{k=1}^{L^v-1}\gamma_{ij}^{1v}(k)}{\sum_{v=1}^{V}\sum_{k=1}^{L^v-1}\sum_{l=2}^{C}\gamma_{il}^{1v}(k)}$$

$$\begin{cases} \mu'_i = \dfrac{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{2v}(k)\cdot e_k^v}{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{2v}(k)} \\ \sigma'_i = \dfrac{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{2v}(k)\cdot (e_k^v-\mu'_i)^2}{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{2v}(k)} \end{cases}$$

$$\begin{cases} \mu'_{ij} = \dfrac{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_{ij}^{3v}(t_k^v)\cdot e_k^v}{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{3v}(t_k^v)} \\ \sigma'_{ij} = \dfrac{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{3v}(t_k^v)\cdot (e_k^v-\mu'_{ij})^2}{\sum_{v=1}^{V}\sum_{k=2}^{L^v-1}\gamma_i^{3v}(t_k^v)} \end{cases}$$

$$e'_{ijm} = \frac{\sum_{v=1}^{V}\sum_{t\in T_m^v}\gamma_{ij}^{3v}(t)}{\sum_{v=1}^{V}\sum_{m=1}^{2}\sum_{t\in T_m^v}\gamma_{ij}^{3v}(t)}$$

$$\begin{cases} \mu d'_{ij} = \dfrac{\sum_{v=1}^{V}\sum_{k=1}^{L^v-1}\sum_{l=k+1}^{L^v}\gamma_{ij}^{4v}(k,l)\cdot d_{kl}}{\sum_{v=1}^{V}\sum_{k=1}^{L^v-1}\sum_{l=k+1}^{L^v}\gamma_{ij}^{4v}(k,l)} \\ \sigma d'_{ij} = \dfrac{\sum_{v=1}^{V}\sum_{k=1}^{L^v-1}\sum_{l=k+1}^{L^v}\gamma_{ij}^{4v}(k,l)\cdot (d_{kl}-\mu d'_{ij})^2}{\sum_{v=1}^{V}\sum_{k=1}^{L^v-1}\sum_{l=k+1}^{L^v}\gamma_{ij}^{4v}(k,l)\cdot d_{kl}} \end{cases}$$

Using the forward and backward variable, the probability parameters can be obtained in the following way:

$$\gamma_{ij}^{1v}(k) = \alpha_{t_k^v}(i)\cdot\left(\sum_{l=k+1}^{L^v} a_{ij}\cdot p_{ij}(d_{kl})\cdot b_{ij}^{+}\left(o_{t_k^v+1}^{v,t_l^v-1}\right)\cdot\ b_j\left(o_{t_l^v}^{v}\right)\cdot\beta_{t_l^v}^{v}(j)\right)$$

$$\gamma_{i}^{2v}(k) = \alpha_{t_l^v}^{v}(i)\cdot\beta_{t_l^v}^{v}(i)$$

$$\gamma_{ij}^{3v}(t) = \sum_{\substack{\{(t_m^v,t_n^v)\}\\ t_m^v<t<t_n^v}} \alpha_{t_m^v}^{v}(i)\cdot a_{ij}\cdot p_{ij}(d_{mn})\cdot b_{ij}^{+}\left(o_{t_m^v+1}^{v,t_n^v-1}\right)\cdot\ b_j\left(o_{t_n^v}^{v}\right)\cdot\beta_{t_n^v}^{v}(j)$$

$$\gamma_{ij}^{4v}(k,l) = \alpha_{t_l^v}^{v}(i)\cdot a_{ij}\cdot p_{ij}\left(d_{k_l}\right)\cdot b_{ij}^{+}\left(o_{t_k^v+1}^{v,t_l^v-1}\right)\cdot\ b_j\left(o_{t_l^v}^{v}\right)\cdot\beta_{t_l^v}^{v}(j)$$

$$\gamma_{ij}^{3v}(t) = \sum_{\substack{\{(t_m^v,t_n^v)\}\\ t_m^v<t<t_n^v}} \gamma_{ij}^{4v}(m,n)$$

An advantage of this algorithm is the fact that the stochastic constraints (listed below), to which each of the model parameters is objected to, are automatically maintained during each iteration.

$$\sum_{j=2}^{C} a'_{ij} = 1\ for\ 1\le i\le C-1$$

$$\sum_{k=1}^{L} b'_i\,(o_k) = 1\ for\ 1\le j\le C$$

$$\sum_{k=1}^{L} b'_{ij}\,(o_k) = 1\ for\ 1\le i\le C-1\ and\ 2\le j\le C$$

$$\sum_{d_{kl}=1}^{D} p'_{ij}\,(d_{kl}) = 1\ for\ 1\le i\le C-1\ and\ 2\le j\le C$$

7.4 Application of HSMESMs to spike detection in EEG signals

The spike detection algorithm currently employed at AIT suffers from the stringent duration intervals defined for the individual segments. Employing HSMESMs could overcome this disadvantage and statistical modeling could lead to detection of spikes of unusual shape and appearance.

A preprocessing step extracts certain features (first and second derivation) from the EEG signal and classifies three different observation symbols according to predetermined thresholds. The observations correspond to Rise, Peak and Fall in the signal under analysis. The number of possible observations is therefore L=5 (including the two fictive events). The result of this stage is one observation \ event sequence for each electrode channel of the EEG respectively.

The C-2 hidden states correspond to the Rise, Peak and Fall segments of the spike morphology. Therefore, a left-right model is applied ($a_{ij} = 0 \; if \; j > i + 1$) and self-transitions are not allowed ($a_{ii} = 0 \; \forall i$). The fpe and null events model signal parts correlating to artifacts or noise and normal EEG in that order. The observation probability distributions are one-dimensional Gaussians whereas the duration probability matrix is uniformly distributed except for the transitions between the states Rise and Peak as well as between Peak and Fall. For those two duration probabilities a normal distribution was better suited.

8 Results

The implementation of the HSMESM model was at first tested with fictive event sequences that were produced with a generator outputting random series of possible observations. As the model performed very well for the random processes, in the next step event sequences were generated through preprocessing and thresholding of EEG recordings. The preprocessing stage consisted of a slightly modified version of an already existing part of the current algorithm used at AIT. The HSMESM implementation was only applied to one channel at a time and does not take into account the spread of spikes over the whole montage.

The application value of the HSMESM algorithm in practical implementations was assessed in the following two categories:

1. Detection of a variety of spike morphologies
2. Detection of different types of epileptiform EEG patterns

8.1 Detection of different spike morphologies

This test series concerned itself with spikes of different width or unusual shapes. The ultimate goal was to show that the HSMESM algorithm provided better results than the currently used initial spike detection step. In this approach the width of the windows for Rise, Peak and Fall has to be fixed beforehand which is disadvantageous in patients with spikes of varying width as they cannot be found. The HSMESM algorithm provides the benefit of statistical modeling of the durations of the states and should be able to spot spikes that are wider or have atypical morphologies.

The following figures will show examples of a variety of spike morphologies that could be detected with the same parameter setting of the HSMESM algorithm. (The spikes are indicated by an arrow. They are depicted as negative – as they usually are – as opposed to the usual representation in EEG recordings as positive amplitude deflections.) The algorithm could clearly discern – the sometimes rather unusual - spike shapes from the rest of the EEG recordings. It could be seen that broader spikes provided no difficulties for the HSMESM model.

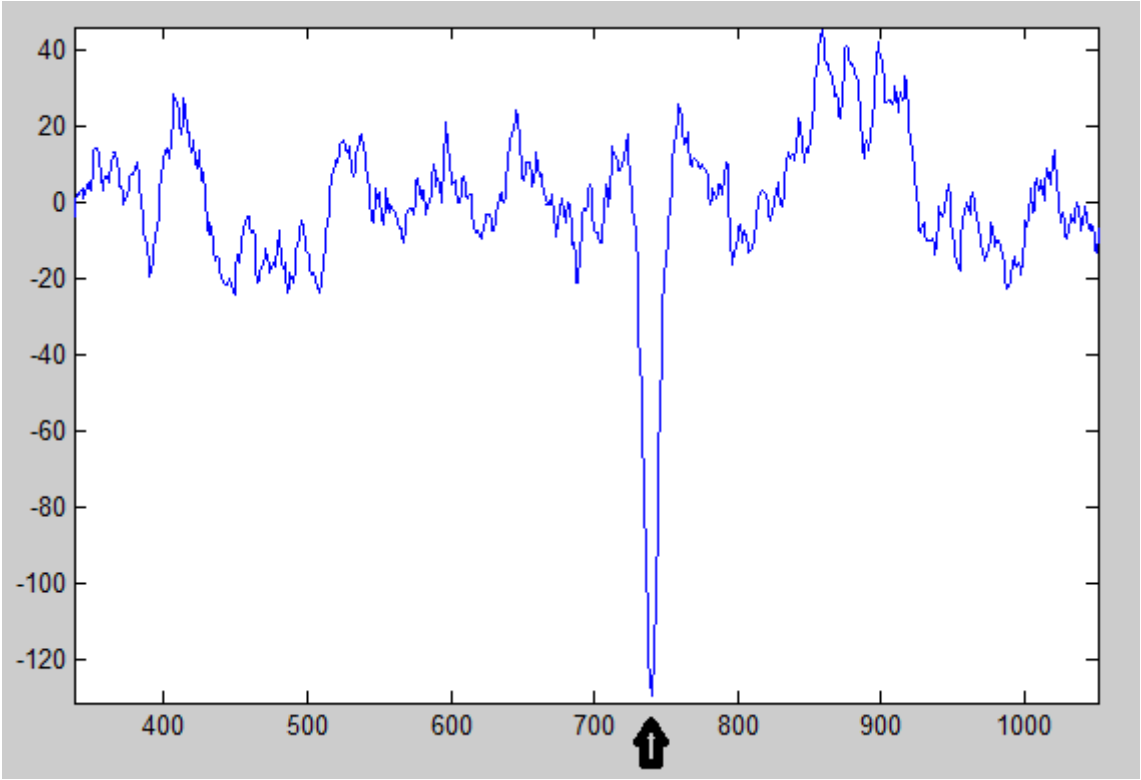


Fig. 8.1. This image shows a typical spike with low background activity, steep rise and fall and shallow width. The spike can be clearly distinguished from the rest of the signal.

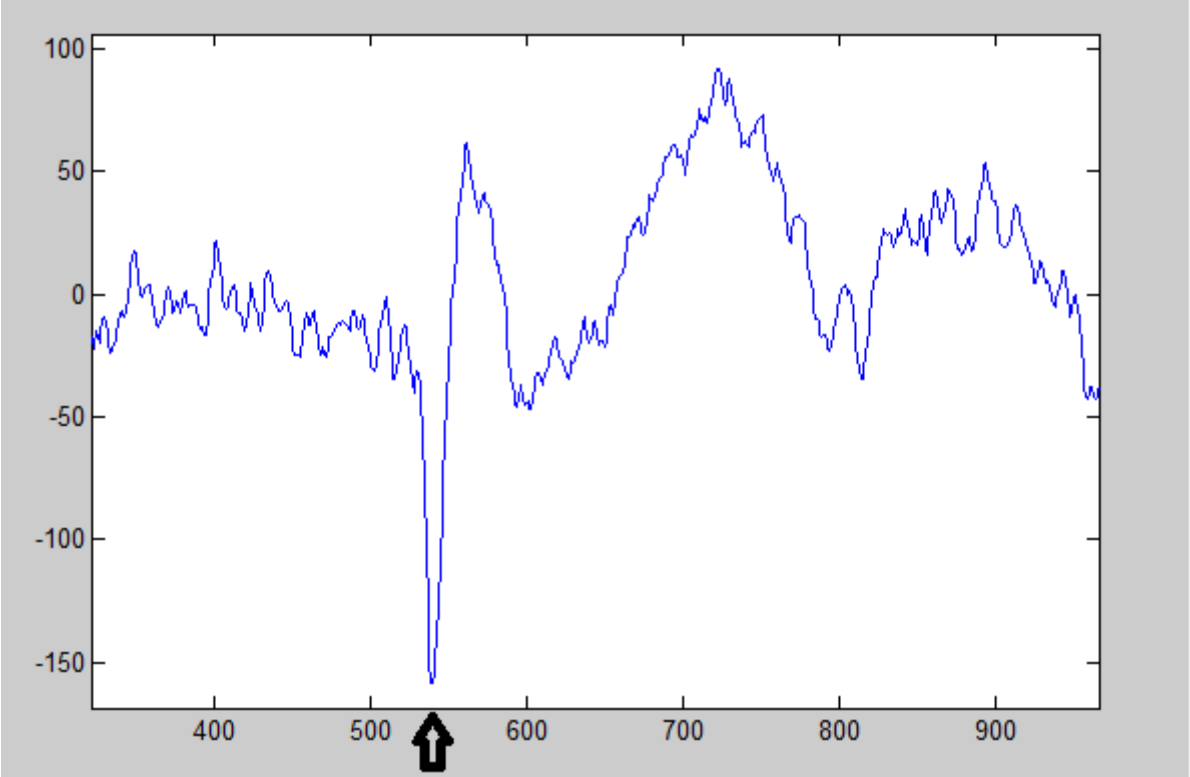


Fig. 8.2. The background signal after the spike shows more variety and erratic behavior but still the spike is clearly discernible and displays a typical shape.

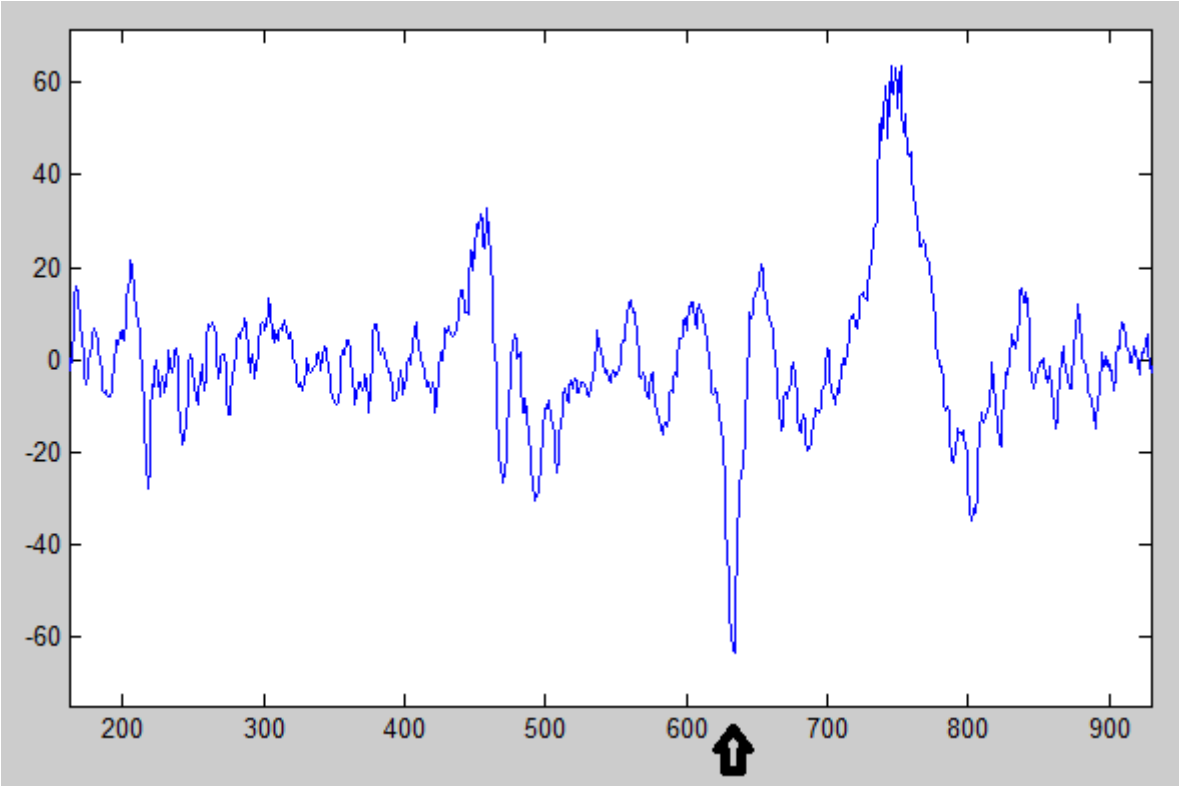


Fig. 8.3. This spike has considerably lower amplitude than the ones in Fig. 8.1. and 8.2. and widens slightly at the bottom. However this still provides no difficulty in detection for either method – the spike detector used at AIT nor the HSMESM algorithm.

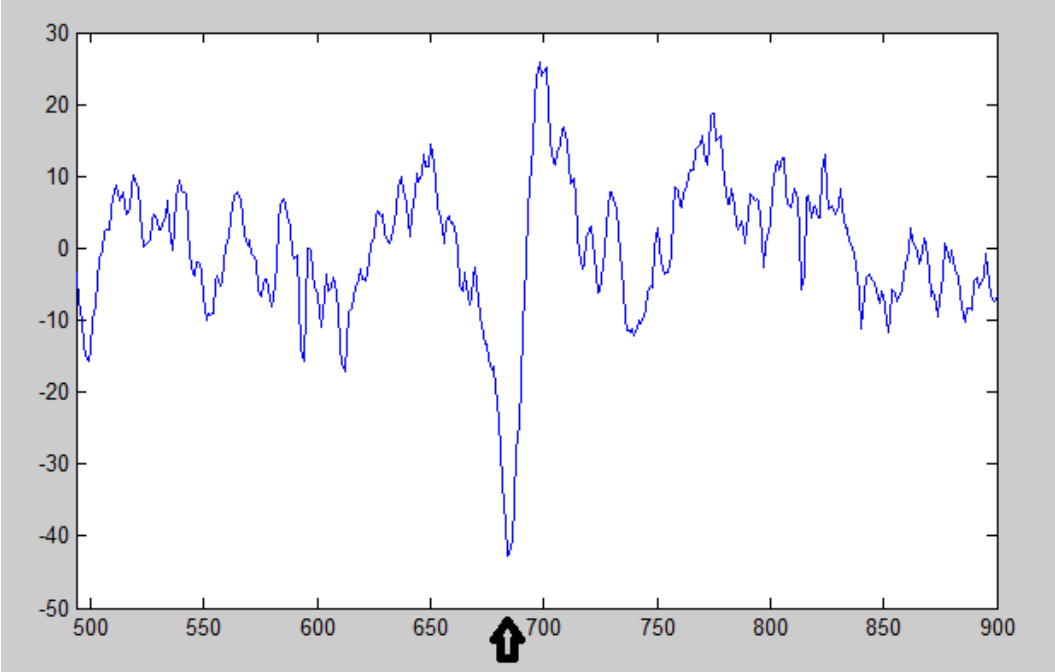


Fig. 8.4. This spike is rather similar to the one in Fig. 8.3. but shows a less distinct beginning of the Rise segment.

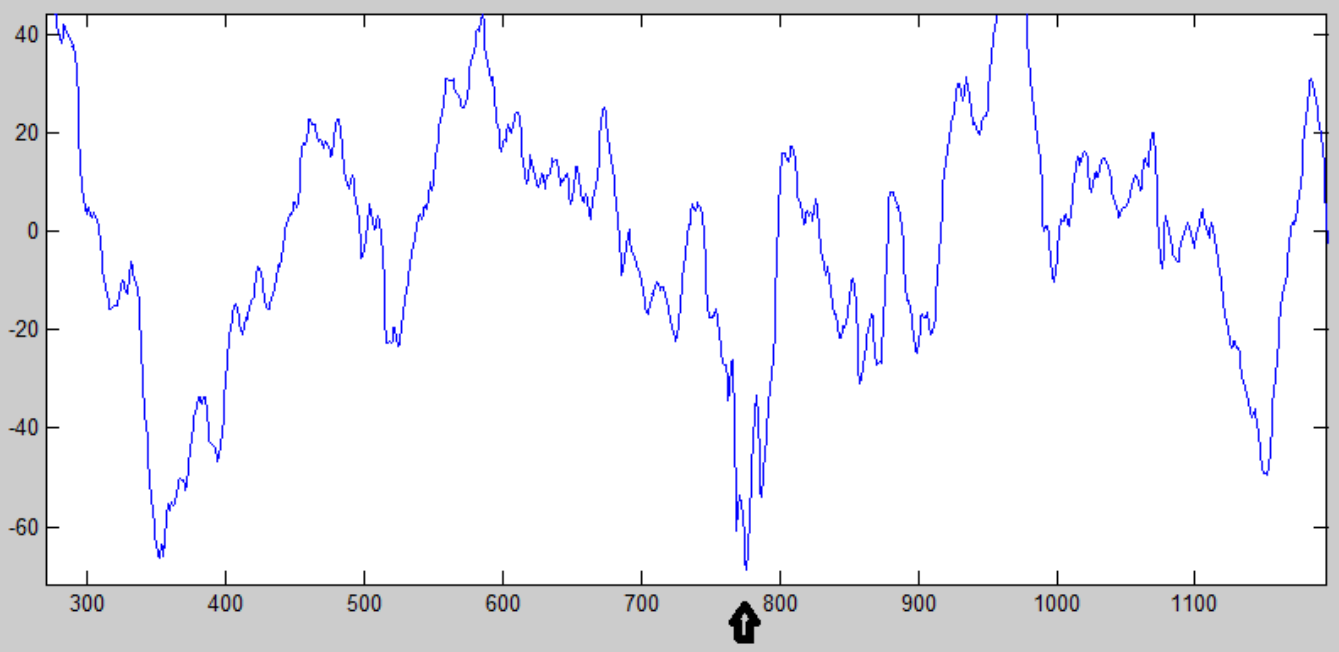


Fig. 8.5. In this signal part the spike is not clearly discernible for the untrained eye. It doesn't exhibit a clearly defined shape and has almost two peaks. The amplitude doesn't clearly distinguish the spike from other elements in the background EEG. Nevertheless, it could easily be detected by the HSMESM algorithm and moreover, differentiated from similar shapes that were no spikes.

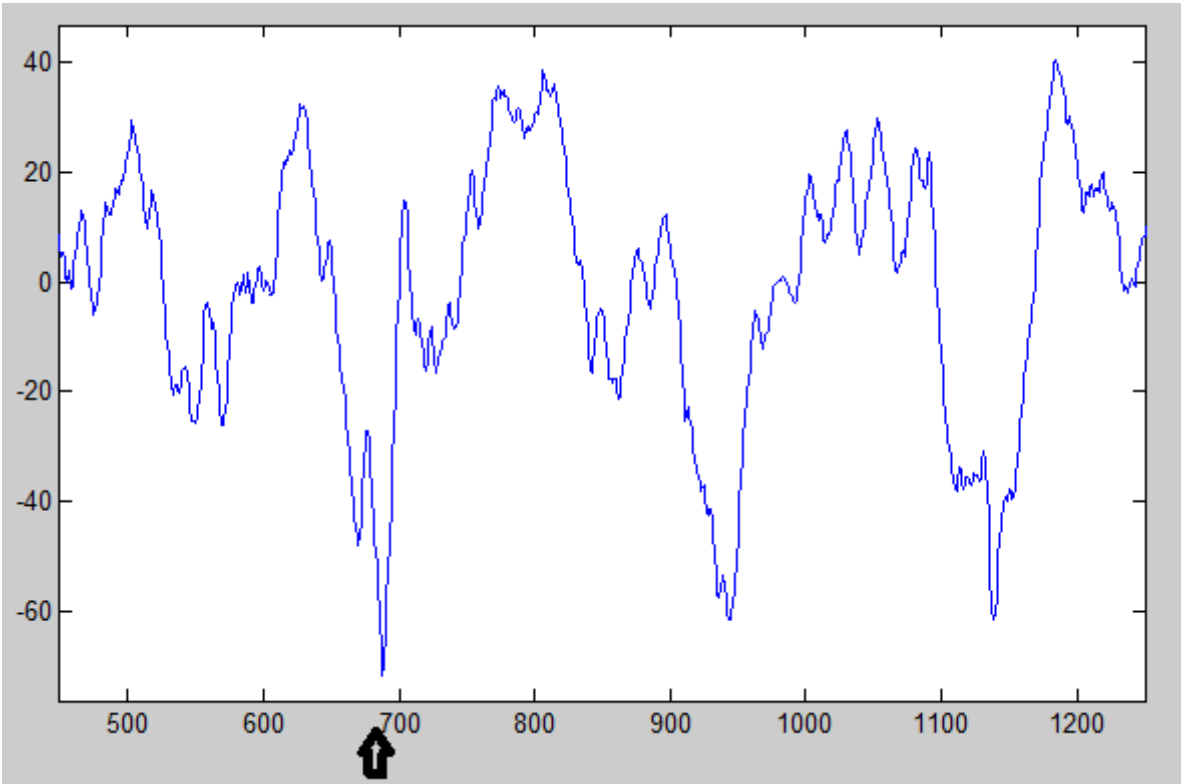


Fig. 8.6. This spike is comparable to Fig. 8.7. as it shows again a division of the actual peak and spike-like shapes in the surrounding signal.

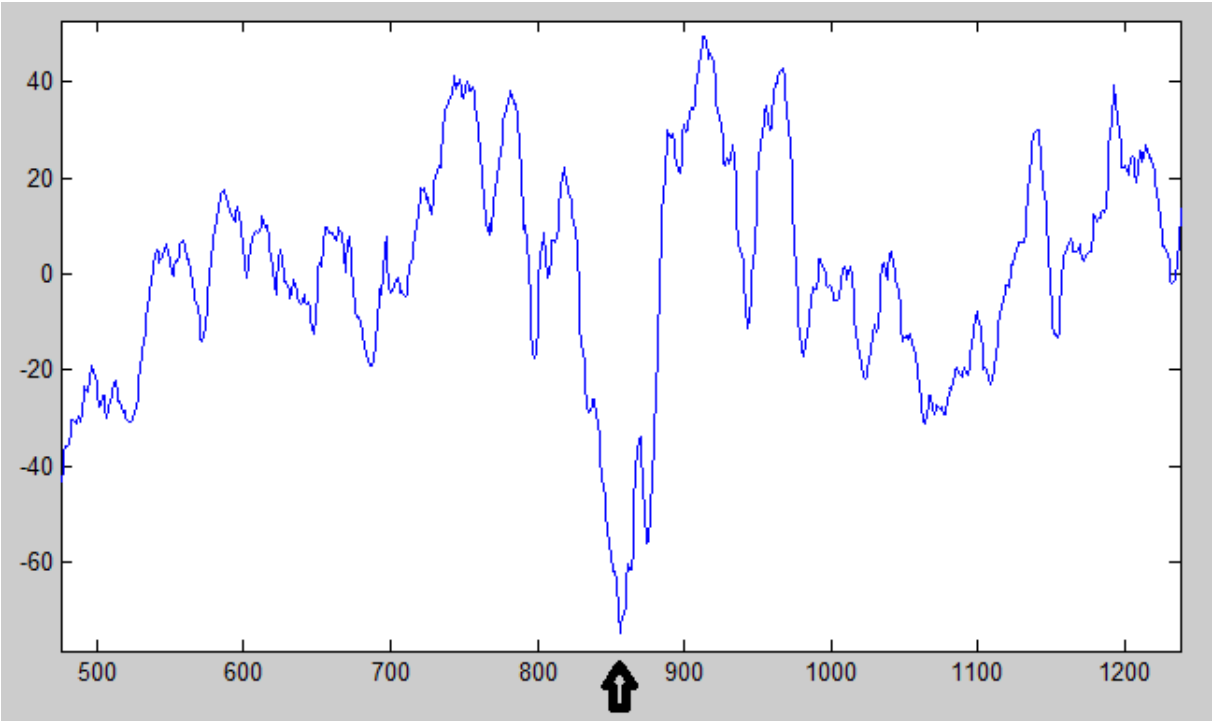


Fig. 8.7. The distance between the starting point of the 'Rise' segment and the end point of the 'Fall' segment is markedly larger than in the previous examples. Again the peak has not only one distinct tip but also a second smaller one in the 'Fall' segment. The broader width could be detected by the model due to the peak segment duration probability density without having to adjust the parameters differently than for example for the spike in Fig. 8.1. .

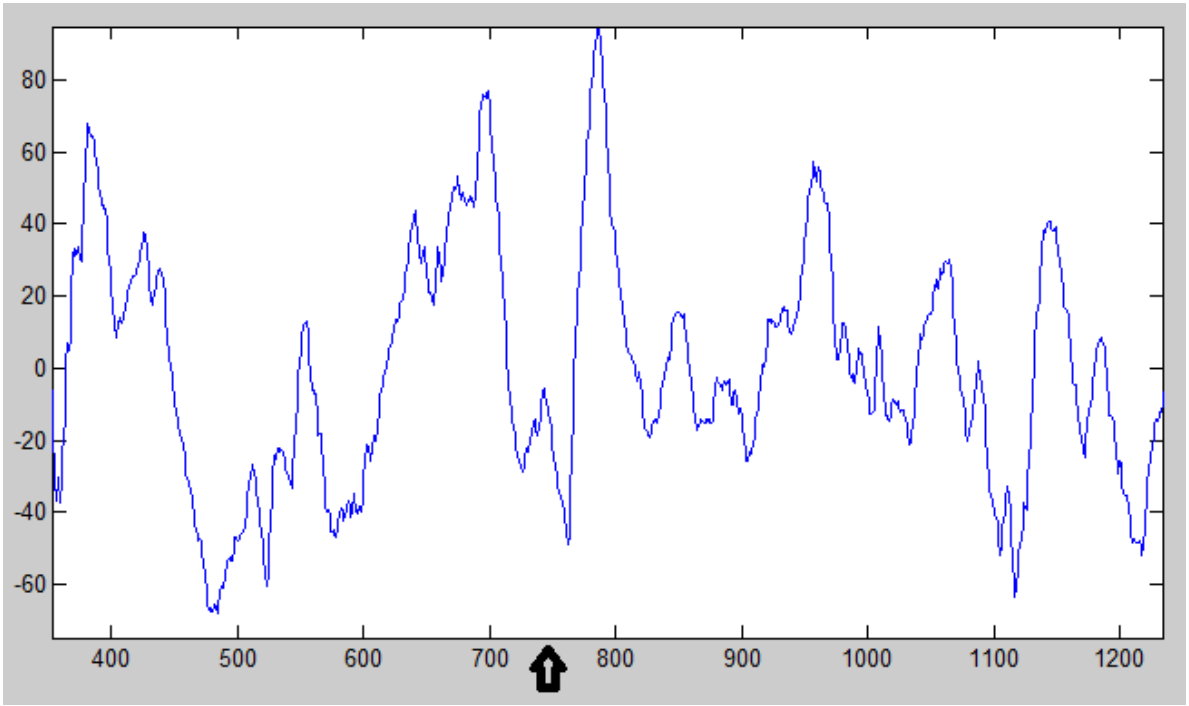


Fig. 8.8. This spike is similar to the one in Fig. 8.7. except that the first tip is the smaller one and the second tip the actual peak. Moreover, the background activity is noisier but didn't have any corrupting influence on the performance of the algorithm.

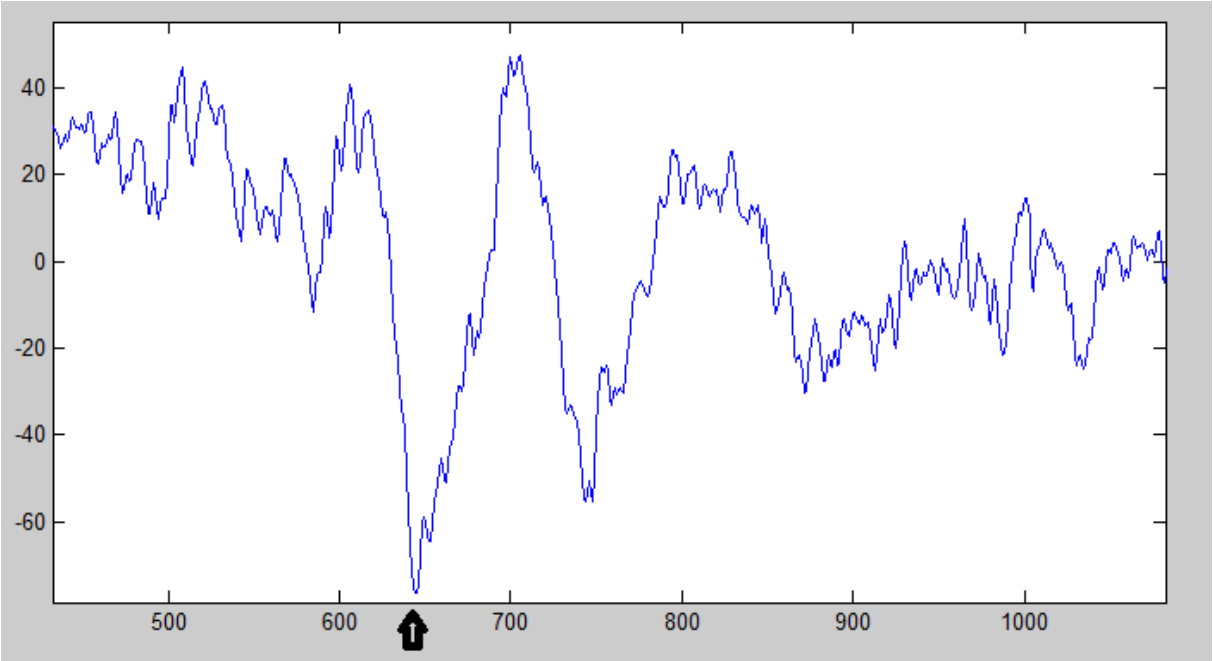


Fig. 8.9. This figure depicts again a spike of rather long duration but with a distinct peak this time.

The Figures 8.1. to 8.9 display a variety of spike morphologies – up to width that would rather define sharp waves and shapes like double peaks – that provided no difficulties for the HSMESM approach and could in some cases even be discerned from quite noisy background activity which often exhibited high amplitudes and spike-like shapes.

As the parameters needed no individual adjustments to these examples but rather a fixed parameter set was retained, this approach provides a promising alternative to the current spike detection in the initial phase of the algorithm developed at AIT.

To foster the claim of practical applicability further tests were undertaken to prove that other epileptiform EEG activities could be detected as well.

8.2 Detection of different epileptiform patterns in EEG data

The first additional epileptiform activities that should be detected in the test phase were polyspikes. If not all of the individual spikes in this pattern, so at least one of them could be detected in most cases. However, the difficulty lies with the narrow shape and the low amplitude. The windows used in the preprocessing step are too large to produce for every single Fall and Rise segment the correct observation. Therefore, the parameters of the HSMESM model need rather much variation possibility in the observations during the states as well as during the transitions. This on the other hand leads to several false detections in some test cases and the trade between sensitivity and specifity is challenging.

Figure 8.10. and 8.11. show samples of polyspikes that could be detected correctly without false findings in the surrounding background. (The arrow marks the first spike in the sequence.)

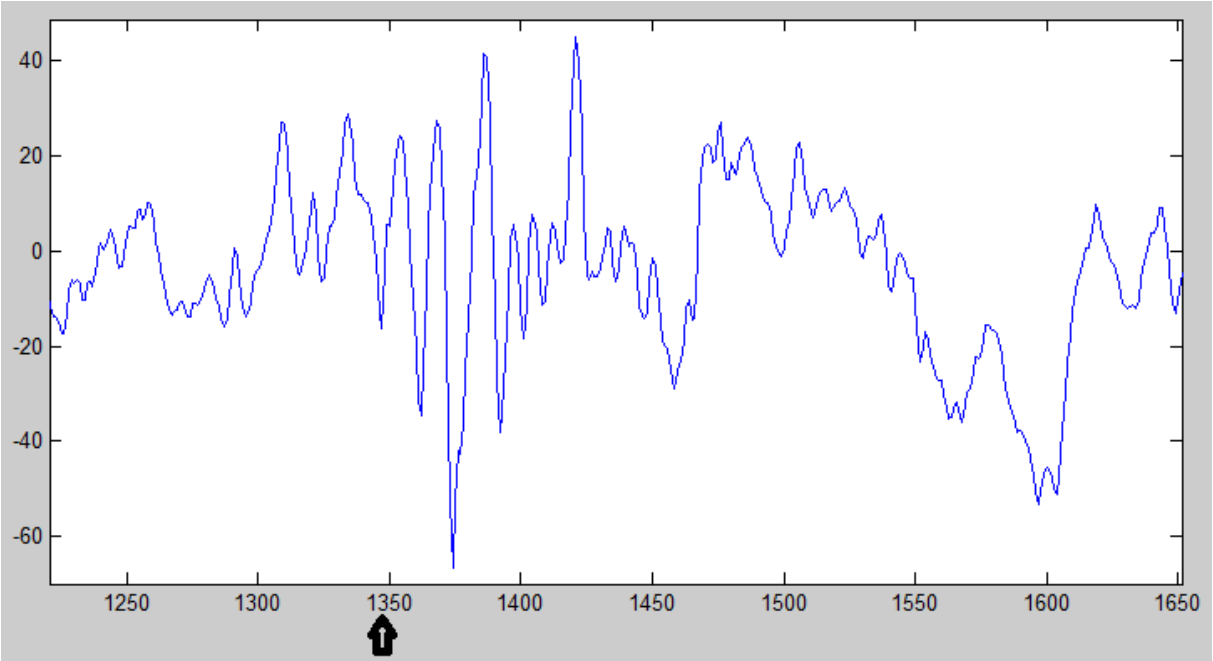


Fig. 8.10. This polyspike consists of four individual spikes whereas the first has rather low amplitude. The algorithm could detect two of them because they are too close together to yield good event sequences in the preprocessing phase.

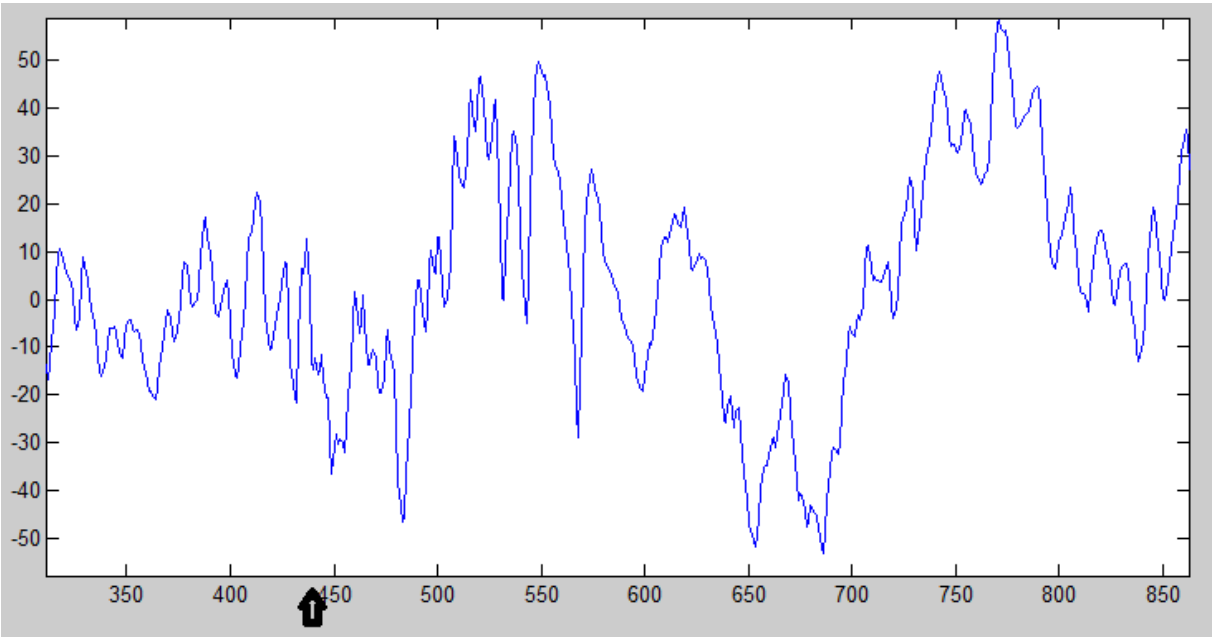


Fig. 8.11. The individual spikes in this polyspike pattern are not so easily discernible. However, eight peaks can be distinguished ending at about 700 and the same amount could be detected by the algorithm.

In conclusion, polyspikes could be detected very well by the HSMESM model if the distance between the individual spikes amounts to at least a few samples. In the case of several spikes in direct succession, adaption of the preprocessing stage could yield better results.

Vertex waves constitute another epileptiform EEG pattern of interest. They have a longer duration than real spikes (but can be equal to sharp waves), are usually symmetrical and mostly display a negative peak. However, the orientation of the wave is also influenced by the montage that is used and in the patients used for the tests, the waves showed a positive deflection from the baseline. Therefore, the preprocessing had to be adapted accordingly. Otherwise, the algorithm and the model remained unchanged.

As the Rise and Fall segments are not as steep as for spikes, there was usually only one observation triggered in the preprocessing step. Therefore, the observations during the hidden states could be well-defined with low variety but the transition observation symbols were rather similar to the background noise. Depending on the background activity, this fact could lead to an increase in false detections. The next three figures depict examples of vertex waves that could be easily detected by the HSMESM algorithm. (Vertex waves are marked with arrows in the figures.)

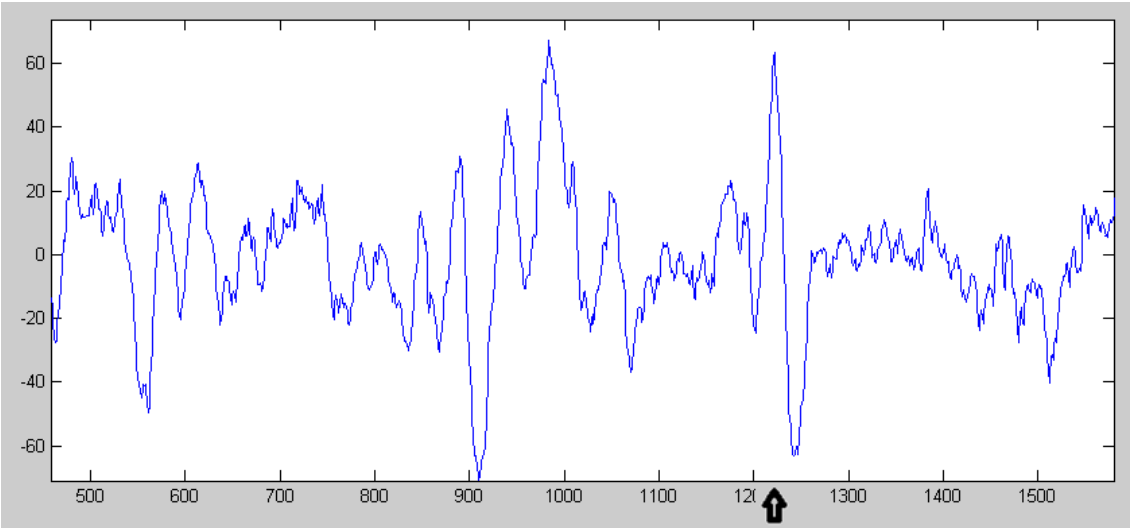


Fig. 8.12. This vertex wave displays a very neat and typical shape which makes differentiation from the background simple although its amplitude isn't much higher than some of the noise.

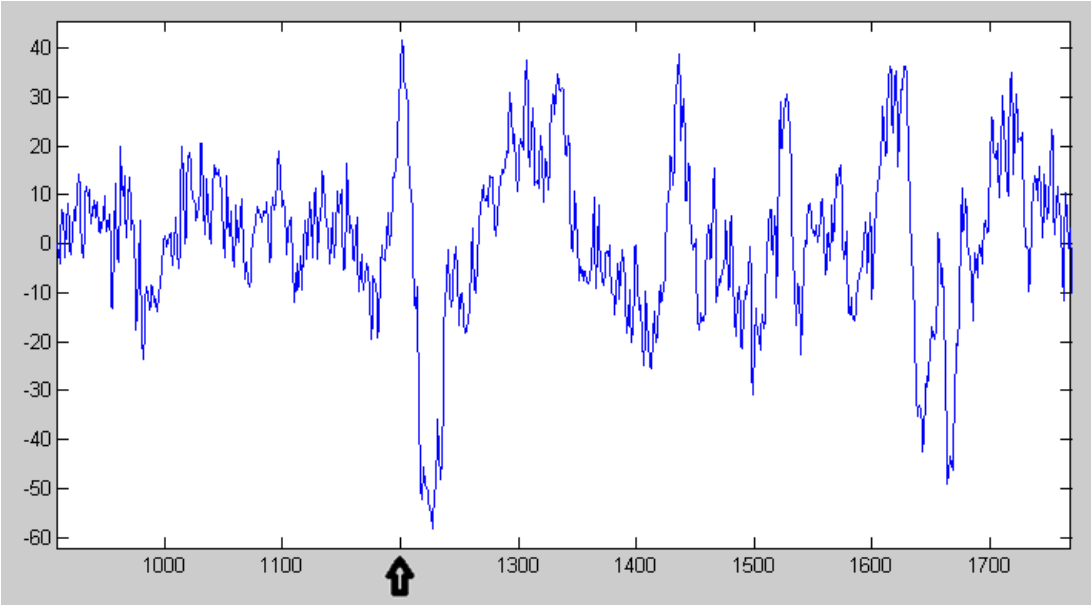


Fig. 8.13. This wave is somewhat more buried in the rest of the signal activity but was found nonetheless without additional false detections.

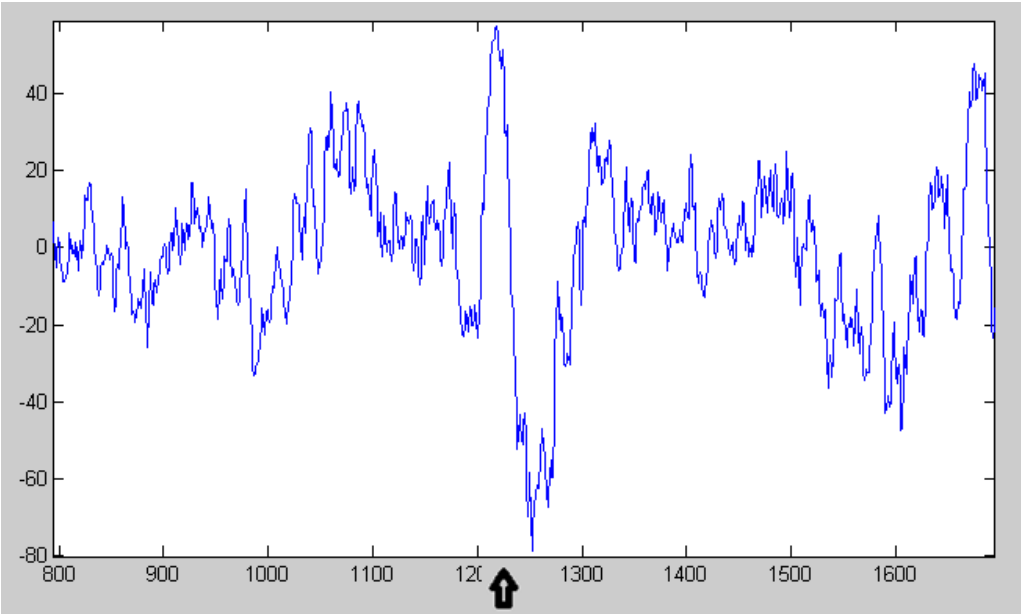


Fig.8.14. Another example of a vertex wave which is very similar to the Figures 8.12. and 8.13.. The only significant variation of vertex waves can be seen in the width but even that remains in a scope that can be covered by the model.

The remaining significant epileptiform EEG activity is constituted by spike-wave-complexes and polyspike-wave-complexes. However, they provide no new challenge to the model as the spikes and polyspikes can be detected as before and as the wave does directly follow the spike

or polyspike, it doesn't need extra detection. They do not interfere with the detection either as they do never display a sudden rise or fall or even a distinct peak.

Lastly, an example shall be provided of the algorithm dealing with a signal containing several spikes in a row but with significantly more distance than in polyspike patterns.

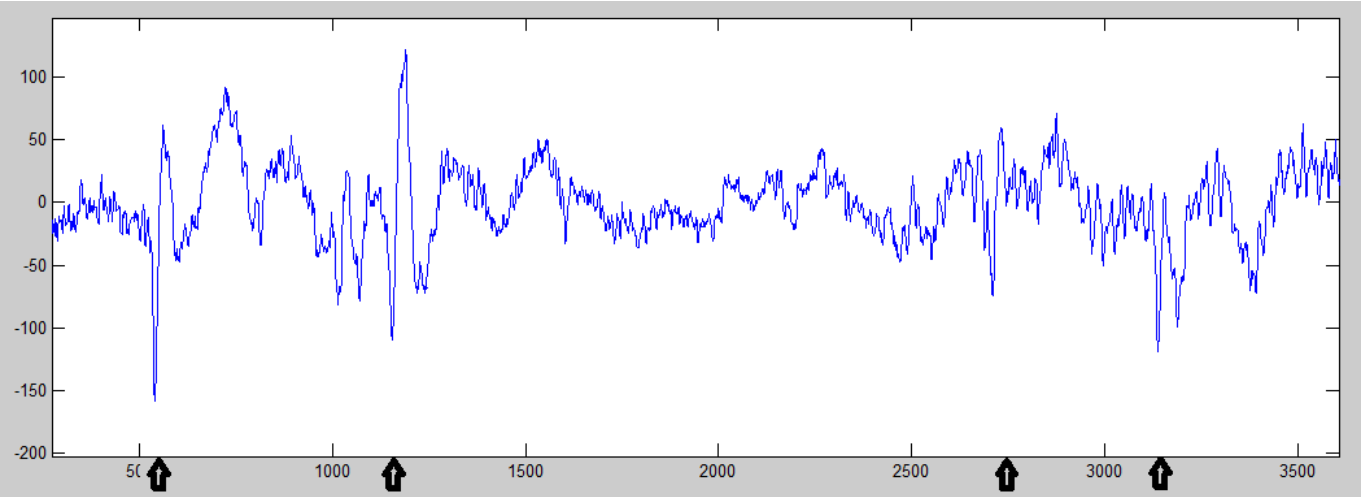


Fig. 8. 15. The figure depicts four spikes whereas the second resembles a polyspike. The HSMESM model was able to indicate all of them with false detections only in the case of the two spike-like tips which are followed by the second spike.

9 Conclusion

Epileptic diagnosis and therapy, including surgery, can benefit greatly from automated detection of epileptiform activity in EEG recordings. There exist already a wide range of different approaches which can be rudimentarily divided in the analysis of inter-ictal periods and the analysis of seizures themselves. The requirements are diverse according to the difference in typical epileptic patterns during these phases.

The aim of this thesis was the enhancement of already existing algorithms that were developed at AIT and are currently in practical application.

Section 5 deals with the replacement of the wavelet analysis in an online seizure detection system by chirplets. Although chirplets provided promising results regarding their detection ability, it is questionable if the additional computational effort is worthwhile. Especially, as online detection should keep the time between possible detection and warning as short as possible. However, the testing of the three variants of the algorithm showed that several assumptions and local estimations can be made without affecting the result while improving the computation time. If further steps were taken to increase the efficiency of the implementation, the chirplet transform could possibly enhance the performance of the current EpiScan algorithm.

However, a far more promising development is the employment of statistical models in automated spike detection during inter-ictal periods. Stochastic durations of the individual segments that shall be recognized in the signal could get rid of the limitation that is posed by the strict time definition of segment window functions that is currently applied. Hidden Semi-Markov Event Sequence Models were chosen as most appropriate to the task and subsequently implemented and tested. The algorithm displayed very promising results in the detection of spikes that varied in width as well as shape (e.g. double peaks) without the necessity of adapting the parameter set. Particularly concerning the different duration times of the spikes the model provided better findings than the existing initial spike detection. Moreover, additional epileptiform EEG patterns like polyspikes and vertex waves could be detected in most of the cases as well. The results for polyspikes could considerably improve through the adjustment of the preprocessing procedure.
On the whole, the HSMESM approach yielded very encouraging outcomes and might - after carrying out further long-term tests - be implemented in the initial stage of the automated spike detection algorithm at AIT to increase its present performance.

10 Appendix

This section contains the Matlab source codes of the implemented algorithms in the sections 5 and 7.

10.1 Standard Matching Pursuit

- Call and parameter definition: gamma = testdict(x)
- Matching pursuit algorithm: gamma = dictionary(x,N,a,du,dz,dc):
  Chirp atom computation: g = chirpatom(u,z,s,c,N)
  Spectral plot with Wigner-Ville distribution: w = plotmp1(y,N)

```
function gamma=testdict(x)

% Calls up the algorithm of the standard matching pursuit with a multiscale
dictionary of Gaussian chirps
%
% Input Variable: signal
%
% Output Variable: array, consisting of the chirps, the parameters related
% to the chirps and the residual of the signal

a=8;
du =1;
dz=pi/2;
dc=0.004;
N=length(x);
plot(real(x));

[gamma]=dictionary(x,N,a,du,dz,dc);
end


function [gamma]=dictionary(x,N,a,du,dz,dc)

% Implementation of the standard matching pursuit with a multiscale
% dictionary of gaussian chirps
%
% Input variables: signal, number of sampling points (length of the
signal), scaling factor, step
% width for the time, the frequency and the chirp rate
%
% Output variable: array, consisting of the chirps, the parameters related
% to the chirps and the residual of the signal

Rx(:,1)=x;
M=40;

for m=2:M

     E=0;
    c1=0;
    c2=0;
```

```
j2=floor(log(N)/log(a));

for j=0:j2  % scale

    disp(['j: ' num2str(j) ' of ' num2str(length(0:j2))])

    s=a^j;

    n2=floor((N-1)/(a^j*du));

    for n=0:n2   % time

       disp(['n: ' num2str(n) ' of ' num2str(n2+1)])
        drawnow

       u=n*a^j*du;

       k1=ceil((-pi*a^j)/dz);
       k2=floor((pi*a^j)/dz);

       for k=k1:k2   % frequency

          z=k*a^(-j)*dz;

l1=ceil(-pi);
l2=floor(pi);

                for l=l1:dc:l2     %chirp rate
       c=l;

             %  Berechnung des chirpatoms

                   g=chirpatom(u,z,s,c,N);
                   g=g(:);
                    gk=g';
         E1=0;

      % Suche nach der maximalen Energie und Speicherung der zugehörigen
      % Parameter

           E1 = gk * Rx(:,m-1);
        b=(abs(E1))^2;

        if E<b
             E=b;
             p=E1;
             u1=u;
             z1=z;
              s1=s;
              c1=c;
             g1=g;

        end
                  l=l+1;
                end

          k=k+1;
```

```
       end
       n=n+1;
    end
    j=j+1;
end

            if E==0
                u1=0;
                z1=0;
                s1=0;
                c1=0;
                p=0;
                g1=zeros(N,1);
            end

            y(:,m-1)=[u1,z1,s1,c1];

   % Berechnung des neuen Residuals

                     Rx(:,m)=Rx(:,m-1)-p*g1;
                     gall(:,m-1)=g1;

        % Abbruch des Algorithmus bei einem Bruchteil der Energie des
        % ursprünglichen Signals

                       P1=x'*x;
        P2=Rx(:,m-1)'*Rx(:,m-1);

        if P2<=P1*1/4
            break
        else
         m=m+1;
        end
end

gamma.g=gall;
gamma.y=y;
gamma.Rx=Rx;

w=plotmp1(y,N);
end


function [g] = chirpatom(u,z,s,c,N)

% Creation of a chirpatom using a gaussian window
%
% Input variables: time, frequency, scale and chirp rate parameters,
% number of sampling points
%
% Output variable: gaussian chirplet

    t = (0:N-1)';
    g = 1/(sqrt(s))*gauss((t-u)/s) .* exp(i*(z*(t-u)+c/2*(t-u).^2));
    g=g/norm(g);
end

% Gaussian window function
```

```
function y = gauss(t)
    y = 1/(sqrt(sqrt(pi)))*exp(-(t/2).^2);
end


function w=plotmp1(y,N)

% Spectrogram of the chirps using the Wigner-ville distribution
%
% Input variables: parameter sets of the chirps and the number of sampling
% points

n=size(y);
Nt = 1200;
vt=0:N/(Nt):N-1;
Nf = 800;
vf = -pi:2*pi/(Nf-1):pi;
w=zeros(Nt,Nf);
figure;

for m=1:n(2)
    for indf = 1:length(vf)
        f = vf(indf);
        for indt=1:length(vt)
            t = vt(indt);
                tk = y(1,m);
                alphak = 1/y(3,m)^2;
                wk = y(2,m);
                betak = y(4,m);
                w(indt,indf)=w(indt,indf)+2*exp(-(alphak*(t-
tk)^2+1/alphak*(f-wk-betak*(t-tk))^2));
        end
    end
end
    im = imagesc(vt,vf,w');
    set(get(im, 'parent'), 'YDir', 'normal');
end
```

10.2 Ridge Pursuit

- Call and parameter definition: gamma = testridgepursuit(x)
- Fast matching pursuit: gamma = ridgepursuit(x,N,a,du,dz,dc):

  Selection of the best gabor atom: [y phi] = bestGabor(Rx,N,a,du,dz,dc):
  - Computation of the gabor atom: g = chirpatom(u,z,s,0,N)
  - Computation of the optimal phase: phi = orthoprojector(g,Rx)
  - Computation of the real-valued gabor atom: g = realchirpatom(u,z,s,0,phi,N)

  Parabolic interpolation for the estimation of the scaling factor and chirp rate: [s c] = parabolicinterpolation(y,Rx,dz,N)

  Spectral plot with Wigner-Ville distribution: w = plotmp1(y,N)

```
function [gamma]=testridgepursuit(x)
```

```
% Calls up the algorithm of the algorithm of the fast matching pursuit with
a multiscale dictionary
% of gaussian chirps
%
% Input variable: signal
%
% Output Variable: array, consisting of the chirps, the parameters related
% to the chirps and the residual of the signal

a=8;
du = 1;
dz=pi/2;
dc=0.004;
N=length(x);
figure,plot(real(x))

[gamma]=ridgepursuit(x,N,a,du,dz,dc);
end


function [gamma]=ridgepursuit(x,N,a,du,dz,dc)

% Implementation of the fast matching pursuit with a multiscale
% dictionary of gaussian chirps (ridge pursuit)
%
% Input variables: signal, number of sampling points (length of the
signal), scaling factor, step
% width for the time, the frequency and the chirp rate
%
% Output variable: array, consisting of the chirps, the parameters related
% to the chirps and the residual of the signal

Rx(:,1)=x;
M=100;

for m=2:M

    [y1 phi1]=bestGabor(Rx(:,m-1),N,a,du,dz,dc);
    y(:,m-1)=y1;
    [sm,cm]=parabolicinterpolation(y(:,m-1),Rx(:,m-1),dz,N);
   um=y(1,m-1);
   zm=y(2,m-1);
                  g=chirpatom(um,zm,sm,cm,N) ;
                  g=g(:);
                   gk=g';
            [phi]=orthoprojector(g,Rx(:,m-1));
      [g2] = realchirpatom(um,zm,sm,cm,phi,N);
        g2=g2(:);
                    p=0;
           p = g2.' * Rx(:,m-1);
     gall(:,m-1)=g2;

          % Berechnung des neuen Residuals

        Rx(:,m)=Rx(:,m-1)-p*g2;
        y2(:,m-1)=[um,zm,sm,cm,phi];

        % Abbruch des Algorithmus bei einem Bruchteil der Energie des
```

```
        % ursprünglichen Signals

        E1=x'*x;
        E2=Rx(:,m-1)'*Rx(:,m-1);

        if E2<=E1*1/5
            break
        else
         m=m+1;
        end

end

gamma.g=gall;
gamma.y=y2;
gamma.Rx=Rx;

 w=plotmp1(y2,N);
end


function [y phi1]=bestGabor(Rx,N,a,du,dz,dc)

% Selection of the best real-valued Gaussian Gabor atom:
% 1) take the complex Gabor atom
% 2) calculate the optimal phase
% 3) calculate the real chirp atom and select the best
%
% Input variables: current residual, number of sampling points, scaling
% factor, step width for the time, the frequency and the chirp rate
%
% Output variables: parameter set of time, frequency and scale for the
% gaussian gabor atom, the optimal phase

j2=floor(log(N)/log(a));
  E=0;

for j=0:j2

    s=a^j;
    n2=floor((N-1)/(a^j*du));

    for n=0:n2

       u=n*a^j*du;
       k1=ceil((-pi*a^j)/dz);
       k2=floor((pi*a^j)/dz);

       for k=k1:k2
          z=k*a^(-j)*dz;
                  g=chirpatom(u,z,s,0,N)  ;
                  g=g(:);
                   gk=g';
             [phi]=orthoprojector(g,Rx);
             g2=realchirpatom(u,z,s,0,phi,N);
             E2=g2'*Rx;
        b=E2;
```

```
        if E<b
             E=b;
             p=E2;
             u1=u;
             z1=z;
              s1=s;
             phi1=phi;
             g0=g;
        end

          k=k+1;
       end
       n=n+1;
    end
    j=j+1;
end

if E==0
                u1=0;
                z1=0;
                s1=0;
                phi1=0;
                g0=zeros(N,1);
                p=0;
end
            %    g2(:,m-1)=g1;
              y=[u1,z1,s1];
end


function [phi]=orthoprojector(g,Rx)

% Calculation of the corresponding optimal phase of the real-valued
% (gabor\chirp) gaussian atom
%
% Input Variables: complex (gabor\chirp) atom, current residual of the
% signal
%
% Output variable: optimal phase

    P=g.'*g;
    g1=g';
     P1=g1*g1.';
    g2=(g-(P*g').')/(1-(abs(P1))^2);
    g2=g2(:);
    g2=g2';
    P2 = g2 * Rx;
      phi=imag(log(P2/abs(P2)));

end


function [g] = realchirpatom(u,z,s,c,phi,N)

% Creation of a real-valued chirpatom using a gaussian window
%
% Input variables: time, frequency, scale and chirp rate parameters, phase
% of the chirp, number of sampling points
%
% Output variable: real-valued gaussian chirplet
```

```
t=(0:N-1)';
    g = 1/(sqrt(s))*gauss((t-u)/s) .* cos(z*(t-u)+c/2*(t-u).^2+phi);
g=g/norm(g);

end

function y = gauss(t)
    y = 1/sqrt(sqrt(pi))*exp(-(t/2).^2);
end


function [sm,cm]=parabolicinterpolation(y,Rx,dz,N)

% Estimation of the locally optimal parameters for scale and chirprate
using a
% parabolic interpolation
%
% Input variables: set of parameters of the best real-valued gassian gabor
% atom, the current residual, the step with for the frequency and the
% number of sampling points
%
% Output variables: the optimal scale and chirprate parameters

% parabolische Interpolation

e1=y(2)+dz/y(3);
e2=y(2)-dz/y(3);
g0=chirpatom(y(1),y(2),y(3),0,N);
g1=chirpatom(y(1),e1,y(3),0,N);
g2=chirpatom(y(1),e2,y(3),0,N);

         g0=g0(:);
           gk0=g0';
         p0=0;
            p0=gk0*Rx;
        g1=g1(:);
           gk1=g1';
         p1=0;
            p1=gk1*Rx;
             g2=g2(:);
                   gk2=g2';
         p2=0;
           p2=gk2*Rx;

           % Zerlegung in Radius und Phase (Polarkoordinatendarstellung)

    r0=abs(p0);
    r1=abs(p1);
    r2=abs(p2);
    if imag(p0)>=0
        phi0=acos(real(p0)/r0);
    else
        phi0=-acos(real(p0)/r0);
    end
     if imag(p1)>=0
        phi1=acos(real(p1)/r1);
    else
        phi1=-acos(real(p1)/r1);
```

```
     end
      if imag(p2)>=0
        phi2=acos(real(p2)/r2);
    else
        phi2=-acos(real(p2)/r2);
      end

% Estimation of the optimal scale factor and chirp rate by using the
% local behaviour of the optimal frequency parameter

      Phi=(phi2-2*phi0+phi1)*(y(3)/dz)^2;
      LogA=log(r2*r1/(r0^2))*(y(3)/dz)^2;
      cm=-Phi/((LogA)^2+(Phi)^2);
      b=(-LogA)/((LogA)^2+(Phi)^2)+(1/(y(3))^2);
      sm=1/sqrt(b);
end
```

10.3 Adaptive Gaussian chirplet decomposition

- Call and parameter definition: gamma = testadaptiveDecomp(x)
- Adaptive gaussian chirplet decomposition: gamma = adaptiveDecomp(x,N,a,du,dz):
  Selection of the best gabor atom: y = Gabor(Rx,N,a,du,dz)
  Optimization of the chirp atom parameters through ‚curve-fitting': [w ln_a b] = linGLS(y,N,Rx,dz)
  Spectral plot with Wigner-Ville distribution: w = plotmp1(y,N)

```
function gamma=testadaptiveDecomp(x)

% Calls up the algorithm of the fast refinement for adaptive gaussian
chirplet
% decomposition
%
% Input Variable: signal
%
% Output Variable: array, consisting of the chirps, the parameters related
% to the chirps and the residual of the signal

a=8;
du =1;
dz=pi/2;
N=length(x);

plot(real(x));

[gamma]=adaptiveDecomp(x,N,a,du,dz);
end


function [gamma]=adaptiveDecomp(x,N,a,du,dz)

% Implementation of the fast refinement for adaptive gaussian chirplet
% decomposition
%
```

```
% Input variables: signal, number of sampling points (length of the
signal), scaling factor, step
% width for the time and the frequency
%
% Output variable: array, consisting of the chirps, the parameters related
% to the chirps and the residual of the signal

Rx(:,1)=x;
M=40;

for m=2:M

 [y]=Gabor(Rx(:,m-1),N,a,du,dz);

u0=y(1);
z0=y(2);
s0=1/y(3)^2;
c0=0;

[w ln_a b]=linGLS(y,N,Rx(:,m-1),dz);
w0=w;
ln_a0=ln_a;

% solve the linear system again for the next time step

u1=u0+du;
y(1)=u1;

[w ln_a]=linGLS(y,N,Rx(:,m-1),dz);
w1=w;
ln_a1=ln_a;

% computation of the optimal parameter set

r=(w0-w1)/(u0-u1);

if r==0
    c=0;
    s=1/(2*b)-s0;
    s=1/sqrt(s);
else
    c=c0+s0/(2*b*r*((s0/r)^2+1));
    s=(c-r)/r*s0;
    s=1/sqrt(s);
end

u=(1/s+1/s0)/(u0-u1)*(ln_a0-ln_a1)+(u0+u1)/2;
z=w0+r*(u-u0);

        g=chirpatom(u,z,s,c,N);
            g=g(:);
            gk=g';

     E=0;
            E = gk * Rx(:,m-1);
            p(:,m-1)=[u,z,s,c];

                    % update of the residual
```

```
                    Rx(:,m)=Rx(:,m-1)-E*g;
                    gall(:,m-1)=g;

        % Abbruch des Algorithmus bei einem Bruchteil der Energie des
        % ursprünglichen Signals

                              E1=x'*x;
        E2=Rx(:,m-1)'*Rx(:,m-1);

        if E2<=E1*1/2
            break
        else
         m=m+1;
        end
end

gamma.g=gall;
gamma.y=p;
gamma.Rx=Rx;

w=plotmp1(p,N);

end


function [w ln_a b]=linGLS(y,N,Rx,dz)

% Solution of a curve-fitting problem to get the parameters for the
% estimation of the best gaussian chirp atom
%
% Input variables: the parameter set of the best gaussian gabor atom,
% current residual, number of sampling points, step width for the frequency
%
% Output variables: Three parameters indicating the relationship between
the
% estimated gabor atoms and the optimal parameter set

dz=pi/2;
u0=y(1);
z0=y(2);
s0=y(3);
c0=0;

% use three different testing points

z1=z0+dz;
z2=z0-dz;

g0=chirpatom(u0,z0,s0,c0,N);
g1=chirpatom(u0,z1,s0,c0,N);
g2=chirpatom(u0,z2,s0,c0,N);

         g0=g0(:);
         gk0=g0';
         p0=0;
            p0=gk0*Rx;
         g1=g1(:);
```

```
        gk1=g1';
        p1=0;
           p1=gk1*Rx;
        g2=g2(:);
        gk2=g2';
        p2=0;
          p2=gk2*Rx;

A=[z0-z2 z2^2-z0^2; z0-z1 z1^2-z0^2];
d=[log(abs(p0)/abs(p2)); log(abs(p0)/abs(p1))];
x=A\d;
b=x(2);
w=x(1)/(2*b);
ln_a=log(abs(p0))+b*(w-z0)^2;
end
```

10.4 The Viterbi Algorithm

- Call: ev=hsmesmViterbi()
- Creation of the event sequence by preprocessing and thresholding: E=threshold()
- Definition of the model parameters: [B,P,C]=hsmesmParameter(E)
- Generation of a fictive event sequence for testing of the algorithm: [E,Z]=eventgenerator(B,P,T,C)

```
function [ev]=hsmesmViterbi()

E=threshold();
% [E,Z]=eventgenerator(B,P,T,C);
[B,P,C]=hsmesmParameter(E);
T=length(E);
E1=E;
e=find(E==0);
E1(e)=[];
L=length(E1);
f=[1:T]';
f(e)=[];

%maximization variable
delta=zeros(C+2,L);
ind=zeros(C+2,L,2);
delta(1,1)=1;
ind(1,1)=1;
d=0;

for l=2:L
    for j=2:C+2
        for i=1:C+1

            p=zeros(T,1);
            p(e)=1-P(i,j-1).emission;
            p(f)=P(i,j-1).observation(E1(find(f)))*P(i,j-1).emission;
            p=p*2.5;
            p(2:length(p)-1)=log(p(2:length(p)-1));
            k=1;
```

```
            while k<l-1
                if (f(l)-1)-(f(k)+1)>=0
                    m=f(k)+1:f(l)-1;
                    p2=cumsum(p(m));
                    b=p2(length(p2));
                else
                    b=0;
                end

                d=log(delta(i,k))+log(P(i,j-1).transition)+log(P(i,j-
1).duration(l-k))+b+log(B(j,E1(l)));
                d=exp(d);
                if d>delta(j,l)
                   delta(j,l)=d;
                   ind(j,l,1)=i;
                   ind(j,l,2)=k;
                end
              k=k+1;
            end
        end
    end
end

%retrieval of the best state sequence
l=1;
e=C+2;
r=L;
d2(l,1)=r;
q(l,1)=e;
l=l+1;

while r>1

    d2(l,1)=ind(e,r,2);
    q(l,1)=ind(e,r,1);
    e=q(l);
    r=d2(l);

    l=l+1;
end

tpe=q(length(q):-1:1);

ev.tpe=tpe;
ev.d2=d2;
ev.B=B;
ev.P=P;
ev.E=E;
ev.f=f;

end



function [B,P,C]=hsmesmParameter(E)

C=3;
L=5;
```

```
T=length(E);
E1=E;
e=find(E==0);
E1(e)=[];
L2=length(E1);

meanB=zeros(C+2,1);
varB=zeros(C+2,1);
meanO=zeros(C+1,C+1);
varO=zeros(C+1,C+1);
meanD=zeros(C+1,C+1);
varD=zeros(C+1,C+1);

meanB=[1,2,3,4,5];
varB=[0.01,0.01,0.01,0.01,0.01];

%set of tpe observation pdfs
    B=zeros(C+2,L);

    for i=2:C+1
        B(i,2:L-1)=pdf('n',2:L-1,meanB(i),varB(i));
    end

    B(1,1)=1;
    B(C+2,L)=1;

    meanO=[3,2,3,3.3;3,2,3,2;3,3,3,3;3,3,3,3.1];
varO=[0.4,0.4,0.4,0.4;0.4,0.15,0.4,0.4;0.4,0.4,0.1,0.1;0.4,0.4,0.4,0.8];

P= repmat(struct('transition', {0}, 'duration', {zeros(L2-1,1)},
'observation', {zeros(L,1)},...
'emission',{0}), C+1, C+1);

%state transition probability matrix
P(1,1).transition=0.95;
P(1,4).transition=0.05;
P(2,2).transition=1;
P(3,3).transition=1;
P(4,1).transition=0.8;
P(4,4).transition=0.2;

%set of inter-tpe state duration pdfs and set of aggregate observation pdfs
for i=1:C+1
        P(i,i).duration(1:L2-1)=pdf('u',1:L2-1,1,L2);
        P(i,i).observation(2:L-1)=pdf('n',2:L-1,meanO(i,i),varO(i,i));
end

for j=2:4
    P(1,j).duration(1:L2-1)=pdf('u',1:L2-1,1,L2);
    P(1,j).observation(2:L-1)=pdf('n',2:L-1,meanO(1,j),varO(1,j));
end

P(3,3).duration(1:L2-1)=pdf('n',1:L2-1,12,1.6);
P(4,1).duration(1:L2-1)=pdf('u',1:L2-1,1,L2);
P(4,1).observation(2:L-1)=pdf('n',2:L-1,meanO(4,1),varO(4,1));
P(2,2).duration(1:L2-1)=pdf('n',1:L2-1,4.7,0.01);

%emission probability matrix
```

```
P(1,1).emission=0.37;
P(1,4).emission=0.2;
P(2,2).emission=0.8;
P(3,3).emission=0.7;
P(4,1).emission=0.34;
P(4,4).emission=0.3;

end


function [E,Z]=eventgenerator(B,P,T,C)

Z=1;
E=1;
i=1;
m=size(P,1);
n=size(B,2);

for l=1:T
    for k=1:m
        y(k)=P(i,k).transition;
        k=k+1;
    end

    s=wuerfel(y');

    if s==3
        continue
    end

    d=wuerfel(P(i,s).duration);
    d=d-1;

    if length(E)+d>T-1
        d=(T-1)-length(E);
    end

    for j=1:d
        r=rand(1);
        if r>P(i,s).emission
            e=0;
        else
            e=wuerfel(P(i,s).observation);
        end

        E=[E;e];
        j=j+1;
    end

    if length(E)==T-1
        break
    end

    o=wuerfel(B(s+1,:));
    E=[E;o];
    i=s+1;
    l=l+1;
    Z=[Z;s+1];
```

```
end

E(T)=[n];
Z=[Z;C+2];

end
```

10.5 The Forward Algorithm

- Call: [obslikeli,alpha,B,P]=hsmesmForwardBackward()
- Creation of the event sequence by preprocessing and thresholding: E=threshold()
- Definition of the model parameters: [B,P,C]=hsmesmParameter(E)

```
function [obslikeli,alpha,B,P]=hsmesmForwardBackward()

E=threshold();
% [E,Z]=eventgenerator(B,P,T,C);
[B,P,C]=hsmesmParameter(E);
T=length(E);
E1=E;
e=find(E==0);
E1(e)=[];
L=length(E1);
f=[1:T]';
f(e)=[];

%forward variable
alpha=zeros(C+2,L);
alpha(1,1)=1;

for l=2:L
    for j=2:C+2
        for i=1:C+1
            p=zeros(T,1);
            p(e)=1-P(i,j-1).emission;
            p(f)=P(i,j-1).observation(E1(find(f)))*P(i,j-1).emission;
            p=p*2.5;
            p(2:length(p)-1)=log(p(2:length(p)-1));
            k=1;

            while k<l-1
                if (f(l)-1)-(f(k)+1)>=0
                    m=f(k)+1:f(l)-1;
                    p2=cumsum(p(m));
                    b=p2(length(p2));
                else
                    b=0;
                end

         a=log(alpha(i,k))+log(P(i,j-1).transition)+log(P(i,j-
1).duration(l-k))+b+log(B(j,E1(l)));
             a=exp(a);
             alpha(j,l)=alpha(j,l)+a;
```

```
                end
              k=k+1;
            end
        end
    end
end

obslikeli=alpha(C+2,L);
end
```

11 Bibliography

Aarabi, A., Wallois, F., & Grebe, R. (2006). Automated neonatal seizure detection: a multistage classification system through feature selection based on relevance and redundancy analysis. *Clin Neurophysiol, 117,* 328-340.

Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., & Yener, B. (2007). Multiway analysis of epilepsy tensors. *Bioinformatics, 23*(13), i10-18.

Acir, N., & Guzelis, C. (2004). Automatic spike detection in EEG by a two-stage procedure based on support vector machines. *Comput Biol Med, 34*(7), 561-575.

Acir, N., Oztura, I., Kuntalp, M., Baklan, B., & Guzelis, C. (2005). Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks. *IEEE Trans Biomed Eng, 52*(1), 30-40.

Alkan, A., Koklukaya, E., & Subasi, A. (2005). Automatic seizure detection in EEG using logistic regression and artificial neural network. *Neurosci Methods, 148*(2), 167-176.

Altenburg, J., Vermeulen, R. J., Strijers, R. L., Fetter, W. P., & Stam, C. J. (2003). Seizure detection in the neonatal EEG with synchronization likelihood. *Clin Neurophysiol, 114*(1), 50-55.

Altunay, S., Telatar, Z., & Erogul, O. (2010). Epileptic EEG detection using the linear prediction error energy. *Expert Systems with Applications,* (37), 5661-5665.

Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, J.H., van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., Moshé, S.L., Nordli, D., Plouin, P., Scheffer, I.E. (2010). Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009. *Epilepsia 51*, 676-85.

Birkemeier, W. P., Fontaine, A. B., Celesia, G. G., & Ma, K. M. (1978). Pattern recognition techniques for the detection of epileptic transients in EEG. *IEEE Trans Biomed Eng, 25*(3), 213-217.

Chandaka, S., Chatterjee, A., & Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications 36*(2), 1329-1336.

Chua, K. C., Chandran, V., Acharya, R., & Lim, C. M. (2008). Automatic identification of epilepsy by HOS and power spectrum parameters using EEG signals: a comparative study. *Conf IEEE Eng Med Biol Soc*, 2008, 3824-3827.

Cui, J., Wong, W., and Mann, S. (2004). Time-frequency analysis of visual evoked potentials by means of matching pursuit with chirplet atoms. *Conf IEEE Eng Med Biol Soc. 2004*, (1) 267-70.

Cui, J., Wong, W. (2008). Investigation of short-term changes in visual evoked potentials with windowed adaptive chirplet transform. *IEEE Trans Biomed Eng*, 55(4), 1449-54.

Cui, J, Wong, W. (2006). The adaptive chirplet transform and visual evoked potentials. *IEEE Trans Biomed Eng*, 53(7), 1378-84

El-Gohary, M., McNames, J., & Elsas, S. (2008). User-guided interictal spike detection. *Conf IEEE Eng Med Bio Soc*, 2008, 821-824.

Forney, G.D. (1973). The Viterbi Algorithm. *Proc IEEE 1973, 61*(3), 268-278.

Geva, A. B., & Kerem, D. H. (1998). Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering. *IEEE Trans Biomed Eng, 45*(10), 1205-1216.

Glover, J. R., Jr., Raghavan, N., Ktonas, P. Y., & Frost, J. D., Jr. (1989). Context-based automated detection of epileptogenic sharp transients in the EEG: elimination of false positives. *IEEE Trans Biomed Eng, 36*(5), 519-527.

Gotman, J., & Gloor, P. (1976). Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. *Electroencephalogr Clin Neurophysiol, 41*(5), 513-529.

Gotman, J. (1990). Automatic seizure detection: improvements and evaluation. *Electroencephalogr Clin Neurophysiol*, *76*(4), 317-324.

Gribonval, R. (2001). Fast Matching Pursuit with a multiscale dictionary of Gaussian Chirps. *IEEE Trans Signal Processing, 49(5),* 994-1001

Guler, I., & Ubeyli, E. D. (2007). Multiclass support vector machines for EEG-signals classification. *IEEE Trans Inf Tech Biomed, 11*(2), 117-126.

Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., & Pazos, A. (2011). Automatic feature extraction using genetic programming: An application to epileptic EEG classification. *Expert Systems with Applications, 38*, 10425-10436.

Hartmann, M., Fürbaß, F., Perko, H., Skupch, A., Lackmayer, K., Baumgartner, C, and Kluge, T. (2011). Episcan: Online seizure detection for epilepsy monitoring units. *Conf IEEE Eng Med Biol Soc*, 6096-9.

Hartmann, M., Fürbaß, F., Perko, H., Skupch, A., Lackmayer, K., Baumgartner, C., and Kluge, T., „EpiScan: Online seizure detection for epilepsy monitoring units“. *Conf IEEE Eng Med Bio Soc*.

Inan, Z. H., & Kuntalp, M. (2007). A study on fuzzy C-means clustering-based systems in automatic spike detection. *Comput Biol Med, 37*(8), 1160-1166.

Iscan, Z., Dokur, Z., & Tamer, D. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Systems with Applications, 38*, 10499– 10505.

James, C. J. (1997). Detection of epileptiform activity in the electroencephalogram using the electroencephalogram using artificial neural networks. *University of Canterbury, Christchurch.*

James, C. J., Jones, R. D., Bones, P. J., & Carroll, G. J. (1999). Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages. *Clin Neurophysiol, 110*(12), 2049- 2063.

Kannathal, N., Choo, M. L., Acharya, U. R., & Sadasivan, P. K. (2005). Entropies for detection of epilepsy in EEG. *Comput Methods Programs Biomed, 80*(3), 187-194.

Ko, C. W., & Chung, H. W. (2000). Automatic spike detection via an artificial neural network using raw EEG data: effects of data preparation and implications in the limitations of online recognition. *Clin Neurophysiol, 111*(3), 477-481.

Kobayashi, K., James, C. J., Nakahori, T., Akiyama, T., & Gotman, J. (1999). Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. *Clin Neurophysiol, 110*(10), 1755-1763.

Kobayashi, K., Akiyama, T., Nakahori, T., Yoshinaga, H., & Gotman, J. (2002a). Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. I: Principles and simulation study. *Clin Neurophysiol, 113*(5), 713- 724.

Korolyuk, V.S., Turbin, A.F. (1976). "Semi-Markov processes and their applications". *Kiev.*

Lange, H. H., Lieb, J. P., Engel, J., Jr., & Crandall, P. H. (1983). Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy. *Electroencephalogr Clin Neurophysiol*, *56*(6), 543-555.

Le Van Quyen, M., Martinerie, J., Navarro, V., Boon, P., D'Have, M., Adam, C., et al. (2001). Anticipation of epileptic seizures from standard EEG recordings. *Lancet, 357*(9251), 183-188.

Lehnertz, K., Andrzejak, R. G., Arnhold, J., Kreuz, T., Mormann, F., Rieke, C., et al. (2001). Nonlinear EEG analysis in epilepsy: its possible use for interictal focus localization, seizure anticipation, and prevention. *Clin Neurophysiol, 18*(3), 209-222.

Lima, C. A., & Coelho, A. L. (2011). Kernel machines for epilepsy diagnosis via EEG signal classification: A comparative study. *Artif Intell Med*.

Liu, H. S., Zhang, T., & Yang, F. S. (2002). A multistage, multimethod approach for automatic detection and classification of epileptiform EEG. *IEEE Trans Biomed Eng, 49*(12), 1557-1566.

Mallat, S.G. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Trans Signal Processing, 41*(12), 3397 - 3415

Mann, S. Haykin, S. (1995). The Chirplet Transform: Physical Considerations. *IEEE Trans Signal Processing, 43*(11), 2745-2761

Martinez-Vargas, J. D., Avendano-Valencia, L. D., Giraldo, E., & Castellanos-Dominguez, G. (2011). Comparative analysis of Time Frequency Representations for discrimination of epileptic activity in EEG Signals. *Conf IEEE Neural Eng*.

McSharry, P. E., He, T., Smith, L. A., & Tarassenko, L. (2002). Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings. *Med Biol Eng Comput*, *40*(4), 447-461.

Miwakeichi, F., Martinez-Montes, E., Valdes-Sosa, P. A., Nishiyama, N., Mizuhara, H., & Yamaguchi, Y. (2004). Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. *Neuroimage, 22*(3), 1035-1045.

Mousavi, S. R., Niknazar, M., & Vahdat, B. V. (2008). Epileptic seizure detection using AR model on EEG signals. *Conf Cairo Int Biomed Eng*.

Oikonomou, V. P., Tzallas, A. T., & Fotiadis, D. I. (2007). A Kalman filter based methodology for EEG spike enhancement. *Comput Methods Programs Biomed, 85*(2), 101-108.

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications* (38), 13475–13481.

Osorio, I., Frei, M. G. (2008). Realtime Detection, Quantification, Warning, and Control of Epileptic Seizures— The Foundations of a New Epileptology. *US Neurology, 4*(2), 67-70.

Ossadtchi, A., Baillet, S., Mosher, J. C., Thyerlei, D., Sutherling, W., & Leahy, R. M. (2004). Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatiotemporal clustering. *Clin Neurophysiol, 115*(3), 508-522.

Ozdamar, O., Yaylali, I., Jayakar, P., & Lopez, C. (1991). Multilevel neural network system for EEG spike detection. *Conf IEEE Comp society press Washington*.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *Neuroimage, 28*(2), 326-341.

Pon, L. S., Sun, M., & Robert, J. S. (2002). The bi-directional spike detection in EEG using mathematical morphology and wavelet transform. *Conf IEEE Trans Signal Processing*.

Polat, K., & Gunes, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation, 187*(2), 1017–1026.

van Putten, M. J. (2003). Nearest neighbor phase synchronization as a measure to detect seizure activity from scalp EEG recordings. *Clin Neurophysiol, 20*(5), 320-325.

Rabiner, L. (1989). A tutorial on Hidden Markov Models and selected application in speech recognition. *IEEE, 77*(2), 257-286.

Roger, J (1993). ANFIS——adaptive-network-based neurofuzzy inference systems, *IEEE Trans Syst Man Cybern. 20* (3), 665—685.

Schiff, S. J., Colella, D., Jacyna, G. M., Hughes, E., Creekmore, J. W., Marshall, A., et al. (2000). Brain chirps: spectrographic signatures of epileptic seizures. *Clin Neurophysiol*, *111*(6), 953-958.

Seino, M. (2006). Classification criteria of epileptic seizures and syndromes. *Epilepsy Research 70*, 27-33.

Senhadji, L., & Wendling, F. (2002). Epileptic transient detection: Wavelets and timefrequency approaches. *Clin Neurophysiol, 32*(3), 175-192.

Sommer, D., & Golz, M. (2001). Clustering EEG-segments using hierarchical agglomerative methods and self-organising maps. *Conf European Neural Network Soc*.

Subasi, A., & Gursoy, I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications 37*, 8659–8666.

Tatum, W. O., Husain, A. M., Benbadis, S.R., Kaplan, P. W. (2008). *Handbook of EEG Interpretation.*

Tzallas, A.T., Tsipouras, M. G., Tsalikakis, D. G., Karvounis, E. C., Astrakas, L., Konitsiotis, S., and Tzaphlidou, M. (2012). Automated Epileptic Seizure Detection Methods: A Review Study. *Mental and Behavioural Disorders and Diseases of the Nervous System.*

Tzallas, A. T., Oikonomou, V. P., & Fotiadis, D. I. (2006). Epileptic spike detection using a Kalman filter based approach. *Conf IEEE Eng Med Biol Soc, 1*, 501-504.

Ubeyli, E. D. (2007). Modified mixture of experts for analysis of EEG signals. *Conf IEEE Eng Med Biol Soc, 2007*, 1546-1549.

Valenti, P., Cazamajou, E., Scarpettini, M., Aizemberg, A., Silva, W., & Kochen, S. (2006). Automatic detection of interictal spikes using data mining models. *Neurosci Methods, 150*(1), 105-110.

Wang, D., Miao, D., & Xie, C. (2011). Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection. *Expert Systems with Applications*(8), 14314–14320.

Webber, W. R., Litt, B., Wilson, K., & Lesser, R. P. (1994). Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized EEG data. *Electroencephalogr Clin Neurophysiol, 91*(3), 194- 204.

Winterhalder, M., Maiwald, T., Voss, H. U., Aschenbrenner-Scheibe, R., Timmer, J. Schulze-Bonhage, A. (2003). The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy Behav., 4*(3), 318-25.

Xu, G., Wang, J., Zhang, Q., Zhang, S., & Zhu, J. (2007). A spike detection method in EEG based on improved morphological filter. *Comput Biol Med, 37*(11), 1647-1652.

Yin, Q., Qian, S, Feng, A. (2002). A fast refinement for adaptive Gaussian chirplet decomposition. *IEEE Trans Signal Processing, 50*(6), 1298-1306.

Yu, S.-Z. (2010). Hidden Semi-Markov Models. *Art Int, 174,* 215-243.