國立交通大學

生醫工程研究所

碩 士 論 文

利用單通道腦電波自動睡眠分期之快速動眼期睡眠剝奪

Automated Sleep Staging using Single EEG Channel for REM Sleep Deprivation

研 究 生：李郁萱

指導教授：陳永昇　教授

中 華 民 國　九 十 八 年 六 月

利用單通道腦電波進行自動睡眠分期之快速動眼期睡眠剝奪
Automated Sleep Staging using Single EEG Channel for REM Sleep Deprivation

研 究 生：李郁萱　　Student：Yu-Shiuan Lee

指導教授：陳永昇　　Advisor：Yong-Sheng Chen

國 立 交 通 大 學
生 醫 工 程 研 究 所
碩 士 論 文

A Thesis
Submitted to Institute of Biomedical Engineering
College of Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in

Computer Science

June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

摘　　要

本論文目的為發展以單通道腦電波為分析訊號來發展即時自動睡眠分期系統，並利用此系統進行快速動眼期之睡眠剝奪實驗。醫學文獻顯示，針對憂鬱症的睡眠腦波，其中一項特徵為憂鬱症患者的快速動眼期發生頻率會比正常人頻繁，因此在他們所服用的抗憂鬱劑具有抑制快速動眼期的效果。相關醫學文獻指出，服用抗憂鬱藥物和快速動眼期睡眠剝奪的機制是相同的，在過去有少量針對憂鬱症患者進行睡眠剝奪的實驗，其實驗結果也證實了睡眠剝奪對於內生型憂鬱症(endogenous depression) 的改善是具有一定效果的。由於在進行睡眠剝奪的實驗中，睡眠分析師必須整夜監測受試者的睡眠狀態，利用人工方式判讀睡眠分期，並在受試者的睡眠狀態處於快速動眼期時將之吵醒。但由於費時費力的缺點，過去僅有少數研究提出睡眠剝奪實驗的治療方式對於憂鬱症治療的效果。因此，本論文欲發展以單通道腦電波為基礎的即時自動睡眠分期系統，以即時偵測到快速動眼期並對受測者進行睡眠剝奪。

在此系統中，我們利用支援向量機作為分類器，對擷取之腦電波特徵做分類。我們利用二十五位受測者的睡眠腦波作為挑選腦電波特徵的測試資料，在嘗試過各種不同的特徵擷取方式後，採用可以得到最佳睡眠分期準確度的特徵擷取方法。利用本論文所建構之自動睡眠分期系統，針對二十五位不同受測者，平均可達到百分之八十五的分期準確度。我們對此自動睡眠分期系統進一步發展成即時自動睡眠分期系統，一旦偵測到快速動眼期，系統即會自動發出聲音以剝奪受測者之快速動眼期的睡眠。在睡眠剝奪的實驗中，我們以六位健康狀況良好，無失眠狀況及憂鬱傾向的受試者來進行本實驗。本實驗目的在於驗證我們所設計的自動睡眠剝奪機制是否能夠達到一定的睡眠剝奪效果，未來預期能運用此系統以進行憂鬱症患者的睡眠剝奪，並探討睡眠剝奪對於憂鬱症的療效。

Abstract

The objective of this study is to develop an on-line automated sleep staging system based on single EEG analysis to assist REM sleep deprivation. Polysomnographic sleep research has demonstrated that the increased rapid eye movement (REM) density is one of the characteristics of depressed sleep. Some experiments were conducted to confirm that REM sleep deprivation (REM-SD) for a period of time is therapeutic for endogenous depressed patients. However, because of its intensive labor requirement, validity of this therapy has not yet been assessed by a sufficient amount of depression patients. Therefore, we propose to develop an automated sleep staging system using only single EEG channel to achieve on-line detection for REM state during sleep. For our sleep staging system, it is based on the supervised classification method with support vector machine. In order to select the feature extraction method which can achieve the best classification result within single EEG channel, we have implemented some feature extraction methods and tested with 25 sleep records. Feature sets derived from different feature extraction method which can achieve the best classification accuracy is the one we will adopted in the proposed system. The average accuracy of classification of all 25 records can achieve 85% that is feasible to REM sleep deprivation. After the algorithm of sleep staging is established, we extend the system to on-line staging which can output the scoring result right away as every 30-second epoch is acquired. Once the REM state is detected by the system, the system will make a shrill sound to disturb the subject until the REM sleep state is changed to other states. Six healthy subjects enrolled in this experiment and have verified the feasibility of the procedure of REM sleep deprivation assisting with the on-line automated sleep staging system. The experimental results demonstrate that our on-line automated sleep staging system is reliable and applicable for on-line REM sleep deprivation.

致　　謝

碩士班兩年的生活很快的就過去了，在此要感謝我的指導教授——陳永昇老師及陳麗芬老師的指導，讓我能在研究上更為精進；也要謝謝我的摯愛家人還有 BSP 實驗室一起奮鬥的夥伴們：乙慈、亮亮、發哥、sheep、育宏以及已經畢業的士暐學長，還有實驗室的學長姐及學弟妹們，因為有你們的陪伴，讓我兩年的碩士生涯既豐富又精彩。另外，要特別感謝台中聖美醫院的林昌宏醫師以及黃杏娟睡眠分析師，在睡眠醫學的知識上為我指點迷津，提供睡眠資料供我們實驗；還有杏娟不辭辛勞的幫忙，提供睡眠人工分期的結果，以當作我們自動分期結果的比較基準；還有，感謝三軍總醫院內湖院區精神科的毛衛中醫師，提供睡眠實驗室以進行睡眠剝奪實驗並提供一些專業上的意見；也感謝陳宜榛助理小姐一個多月來的陪伴。

碩士班的結束也代表我即將踏入職場，這兩年的難忘回憶將為我的學生生涯畫下完美的句點。

謹以此論文獻給我最摯愛的家人

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter we introduce some background knowledge of the thesis. We briefly describe the motivation of the proposed work at the first section. Next in section 1.2, we introduce human sleep from the definitions of sleep, the methods of sleep examination, and the standard used for sleep staging. In section 1.3 we describe about the development of automated sleep staging but the details will introduce in chapter two. In section 1.4 we give a brief introduction to sleep deprivation and demonstrated the effects of REM sleep deprivation for depression . Finally in section 1.5 is the overview of the thesis.

1.1 Motivation

Sleep is important and essential for human, not only for the function of body restoration but also enhance our memories. However, busy life in nowadays cause people living under pressure and impairing the sleep quality to reduces the efficiency of learning and working in consequence. Furthermore, some studies suggest that chronic insomnia is probably a risk factor for depression [6]. For above reasons, bad sleep quality may cause negative physical even mental effect [16] and this problem is becoming one of the common problems for seeking medical attention. In sleep medicine, the polysomnogram (PSG) is a multi-parametric test used as a diagnostic tool in the study of sleep. Among the PSG, sleep electroencephalogram (EEG) is the main reference for sleep staging. The physician can find out the anomalies in sleep by the sleep staging result. In recent studies, it was proposed that the relation between depression and insomnia has a strongly bi-directional linkage. The depressed sleep is characterized by a disinhibition of REM sleep including increased REM sleep. Thus, most of effective antidepressant agents suppress REM sleep. Base on the assumption of REM sleep suppression is the necessary prerequisite for antidepressants effect [9], some experimental studies had been conducted; Early in the year of 1975, Vogel et al. [8] proposed that REM sleep deprivation is therapeutic for depressed patient. Due to the experiment of sleep deprivation is so labor-intensive that only few data had been confirmed.

For a long time, the method of automated sleep staging were proposed in succession. Although many PSG systems on the market have the function of automated sleep staging

embedded, however, physicians would rather scored manually due to the result is more reliable. Therefore, to develop a reliable automated sleep staging system is an essential task for sleep medicine. Moreover, the extension of automated sleep staging system to on-line scoring may give a big help for applying REM sleep deprivation.

1.2 Sleep EEG

Sleep is a complex physiological and behavioral process. There are two separate states have been defined within sleep, which are rapid eye movement (REM) and non-rapid eye movement (NREM). REM sleep is defined by EEG activation, muscle atonia, and episodic bursts of rapid eye movements. The mental activity of REM sleep is associated with dreaming. Dreaming has important psychological effects, helping us to put "things in order". NREM sleep, by contrast, is defined along the measurement axis of the EEG and conventionally subdivided into four stages, which are stage 1, 2, 3 and 4, The four NREM stages parallel a depth of sleep, with arousal threshold generally lowest in stage 1 and highest in stage 4. Usually, stage 1 and 2 sleep are referred as light sleep and investigators refers to combine stage 3 and 4 as slow-wave-sleep (SWS) as deep sleep [17]. The continuous alternation of NREM sleep and REM sleep through the night sleep is a cyclic process. Across the nocturnal sleep, REM sleep generally become longer across the night; The average period of the NREM-REM sleep cycle is approximately 90 to 110 minutes.

Within sleep, brain activity is mainly divided into four rhythms [20]. According to different frequencies, it can split to delta wave (<4 Hz), theta wave (4-8 Hz), alpha wave (8-12 Hz) and beta wave (>13 Hz). Beta waves are defined as high frequency and low amplitude; alpha waves, which occur during relaxed state, are regular rhythms with higher amplitude than beta waves. Theta waves are typically of even greater amplitude and slower frequency than alpha waves. Delta waves, slowest EEG rhythms, generally have the highest amplitude. When we fall into deep sleep, theta waves and delta waves becomes more and more obvious in the EEG background activities. The predominant rhythm of EEG is different in each sleep stages which are used to discriminate transitional stages. The defined rules will be introduced in the next section.

Many factors may affect the distribution of sleep stages and the most strongest factor affecting the cyclical alternation of sleep is age. As the brain structure mature with age, it achieves a level that can support high-voltage slow wave EEG activities. Thus, SWS is maximal in younger children and decreases gradually with age. In the contrast, arousals during sleep increased with ages. Figure 1.1 illustrated the age-related change in human sleep.

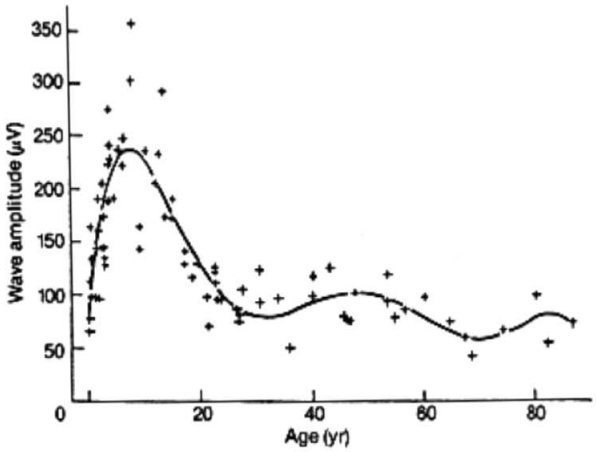
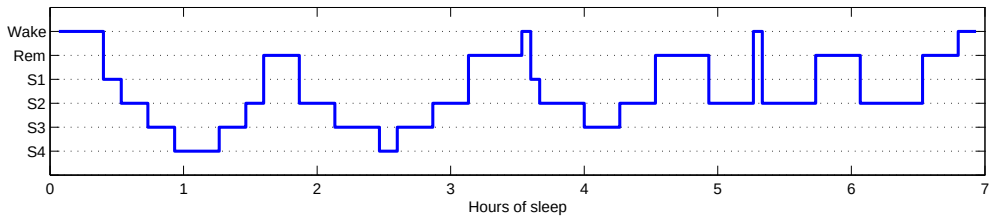
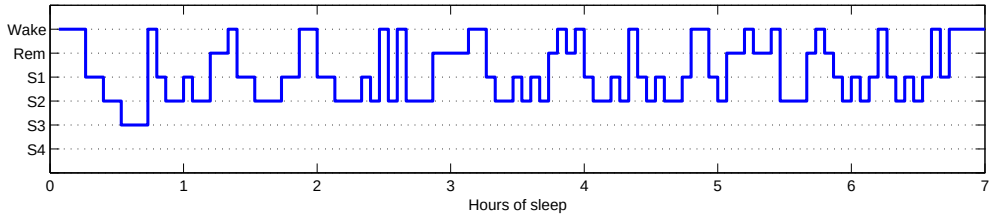


Figure 1.1: The graph illustrated age-related changes in human EEG amplitude. Higher frequency have the higher amplitude. (Figure source: Normal Human Sleep: An Overview)

We can observe cyclical alternation of sleep by a diagram which is called sleep hypnogram. As shown in Figure 1.2, it appears different stages following with different time in sleep process. The upper figure is the hypnogram of the one who is healthy and young. There are 4 to 8 cycles in an 7-hour sleep; the lower figure is the hypnogram of an old one. As the figure shown, sleep process of the elder is more irregular than the younger, which the sleep state usually stay in the light sleep and have more arousal. Generally speaking, the elder has the worse sleep quality compared to younger. Therefore, an objective basis for people to judge sleep quality is starting from observing the sleep stages.



(a) younger sleep structure



(b) elder sleep structure

Figure 1.2: Comparisons with hypnogram of healthy younger adult versus elderly one. The upper figure is the hypnogram of the healthy younger adult. The sleep cycle is more complete; The lower figure is the hypnogram of an elder. The sleep cycle is indiscriminate and hardly enter deeper sleep stages

1.3 Sleep staging

1.3.1 Polysomnography

The dynamic structure of sleep can be examined by polysomnography (PSG). It records physiological signals such as electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), blood pressure, respiratory efforts, oxygen saturation (SaO2) and so on for monitoring long-term changes of brain activity during sleep. Recently, sleep centers were established in many hospitals, which have professional equipments and good environments for sleep examinations. The environment of total isolation for any electric waves can reduce noises while signal acquiring. The sleep center use the computer-based polysomnographic system to record dynamic sleep structure. The computer-based PSG system comprises sensors, amplifiers, filters, analog-digital converters and computer with peripheral devices as illustrated in Figure 1.3. The system should analyze EEG, EOG, EMG in terms of sleep stages; respiration, snoring and oxygen satura-

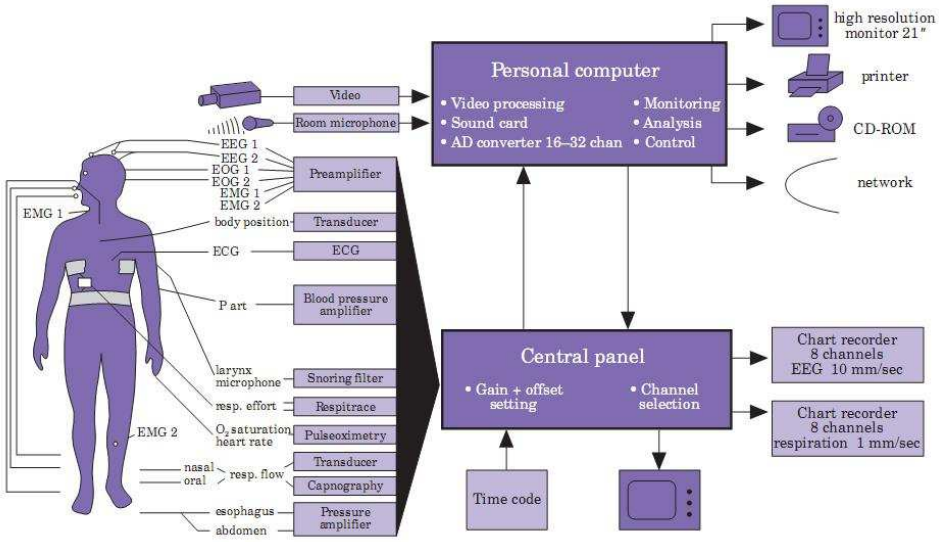


Figure 1.3: An overview of PSG system. A computer based system consists of sensors, amplifiers, filters, analog to digital converters and a computer with peripheral devices. On the right side of the diagram all peripherals essential for documentation can be seen. (Figure source: sleep medicine reviews, vol. 4, No. 2, T. Penzel et al: The Computer based sleep recording and analysis, p 131-148, 2000)

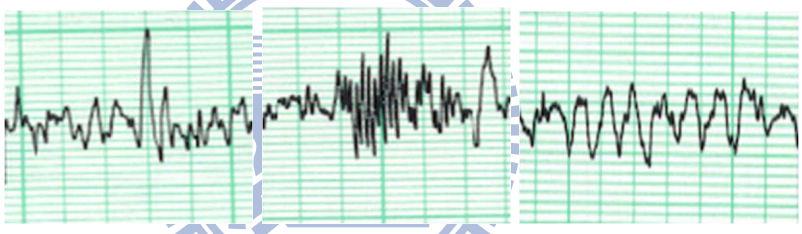
tion in terms of sleep related breathing disorders; and EMG tibialis in terms of movement disorders. Other parameters recorded such as body temperature, ECG, blood pressure may require additional analysis. Although most PSG system has the function of sleep scoring, most sleep specialist would rather score manually by R&K rules and abandon the result generated by the system. Due to sleep scoring system is unreliable, PSG system is only for signal acquisition without any analysis.

1.3.2 Rechtschaffen and Kales (R&K) rules

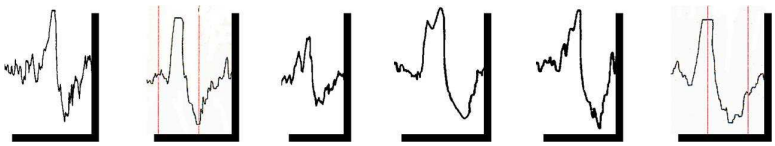
The standard of sleep staging is R&K rules, which is proposed by Rechtschaffen and Kales in 1968 [25]. This rule is a reference method for sleep stage scoring in sleep medicine. According to Rechtschaffen & Kales' suggestion, sleep scoring depends on the characteristics of electroencephalogram (EEG), electrooculogram (EOG) and electromyogram (EMG). Central leads of EEG (C3-A2 or C4-A1) are the minimal requirement and

occipital leads (O2-A1 or O1-A2) are additionally recorded. Two EOG leads are placed at the outer corners of the eyes to detect vertical and horizontal eye movements. EMG leads are placed in the mental and submental areas to monitor the variation of voltage between the two. Table 1.1 is the arrangement of the standard sleep staging criteria in adults according to the EEG, EMG and EOG.

For manual staging, EEG signal is the major reference, with signals of EMG and EOG assisted. The continuous signal are segmented as an epoch for every 30 seconds. According to the frequency, amplitude and morphology of EEG signal, sleep process can divided into six stages: Stage W, Stage 1, 2, 3, 4 and REM. The characteristics of EEG signals and corresponding sleep stages defined are presented in the following list [18]. Figure 1.5 are EEG tracings of six sleep stages. The EEG pattern is also described with such characteristic waveform as vertex sharp wave, spindles, sawtooth wave and K-complexes as shown in Figure 1.4. These EEG patterns may appear in different sleep stages.



(a) vertex sharp wave (b) spindle (c) sawtooth



(d) K-complex

Figure 1.4: (a) vertex sharp wave; (b) spindle wave; (c) sawtooth wave; (d) K-complex wave

- **Stage W (Wakefulness):** The EEG contains alpha activity and/or low voltage, mixed frequency activity.
- **Stage 1:** it is characterized by the low voltage and mixed frequency EEG. The alpha

Table 1.1: Outline of sleep scoring standard according to R&K standard

Stage		EEG	EOG	EMG
Wakefulness		**Eyes Closed:** rhythmic alpha (8-13 cps); prominent in occipital; attenuates with attention; **Eyes open:** relatively low voltage, mixed frequency	Voluntary control; REMs or none; blinks; SEMs when drowsy	Tonic activity, relatively high; voluntary movement
NREM	**Stage1**	1.Relatively low voltage, mixed frequency; 2.May be theta (3-7 cps) activity with greater amplitude 3.Vertex sharp waves; 4.Synchronous high-voltage theta bursts in children	SEMs	Tonic activity, may be slight decrease from waking
	Stage2	**Background:** relatively low voltage, mixed frequency; **Sleep spindles:** waxing, waning, 12-14 cps (>0.5sec); **K complex:** negative sharp wave followed immediately by slower positive component (>0.5sec); spindles may ride on KCs; KCs maximal in vertex; spontaneous or in response to sound	Occasionally SEMs near sleep onset	Tonic activity, low level
	Stage3	20%-50% high amplitude (>75 uV), slow frequency (<2cps); maximal in frontal	None, picks up EEG	Tonic suppression; phasic twitches
	Stage4	>50% high amplitude, slow frequency	None, picks up EEG	Tonic activity, low level
REM		Relatively low voltage, mixed frequency	Phasic REMs	Tonic suppression; phasic twitches

Adopted from Principles and Practice of Sleep Medicine, 2nd Ed. W.B. SAUNDERS company, which is modified from Rechtschaffen A, Kales A (eds): A Manual of Standardized Terminology: Techniques and Scoring System for Sleep Stages of Human Subjects. Los Angeles, UCLA Brain Information Service/Brain Research Institute, 1968.

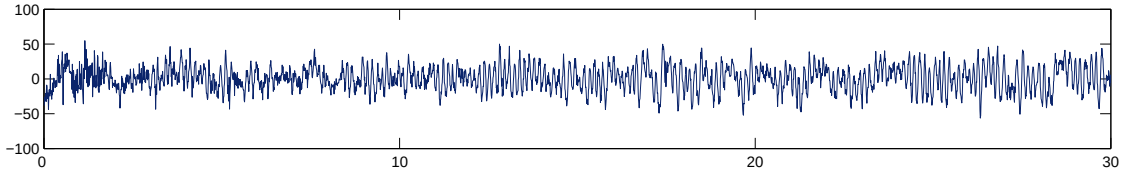
wave is less than 50% in overall epoch. Stage 1 occurs most often in the transition from wakefulness to the other sleep stages. Thus, an increased amount of Stage 1 sleep represents that sleep process is disrupted. At the end of Stage 1, the slow wave with sharp negative voltage which is called vertex sharp wave may appear in the EEG.

- **Stage 2:** there is a gradual appearance of high-voltage slow wave activities. Sleep spindles and/or K complexes may appear as a characteristic waveform. Spindle wave is an EEG oscillations of 12-15 Hz with a duration of 0.5-1.5 seconds. K-complex wave is a high-amplitude biphasic wave which begins with an initial sharp positive voltage and then followed by a negative deflection slow wave. The duration is at least 0.5 second.

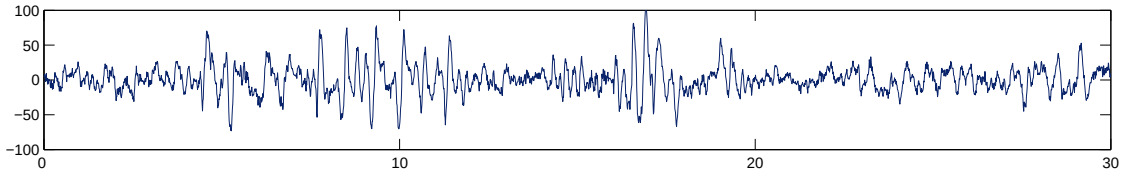
- **Stage 3:** Moderate amount of slow delta waves. Slow delta wave is defined as EEG activity slower than 2 Hz with peak-to-peak amplitude greater than 75 microvolt. If the slow delta waves of EEG activity occupy more than 20% but less than 50% of an epoch.

- **Stage 4:** Large amount of slow delta waves which occupy more than 50% of an epoch.

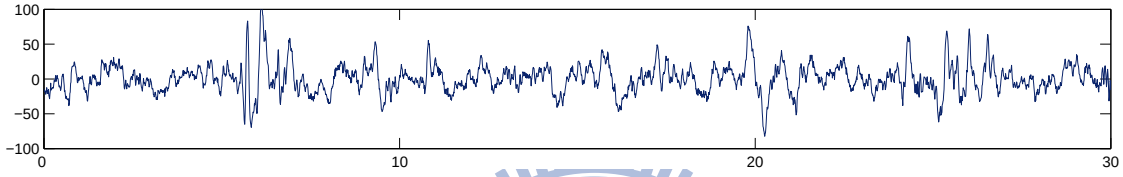
- **Stage REM:** is defined by a relatively low voltage, mixed frequency EEG in conjunction with episodic REM with low amplitude EMG tone. In this state, it may appear sawtooth wave, which is low amplitude and looks just like a sawtooth.



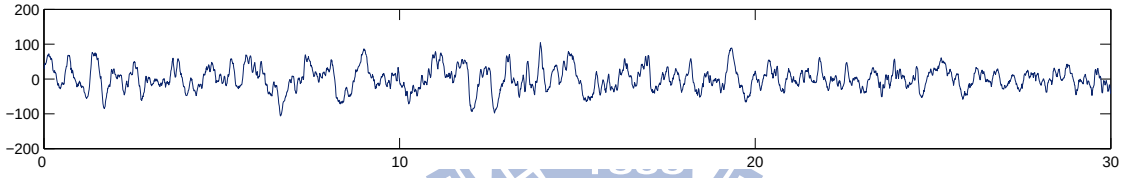
(a)



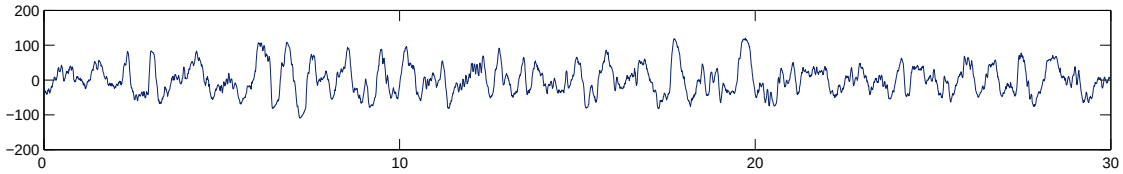
(b)



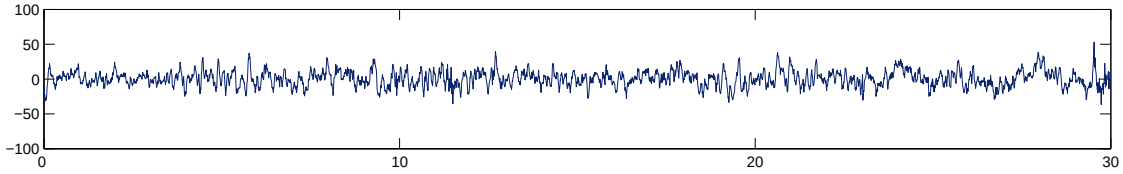
(c)



(d)



(e)



(f)

Figure 1.5: From (a) to (f) are the EEG tracings of Stage W, Stage1, 2, 3, 4 and REM relatively. The EEG tracing was recorded from a referential lead (C3/A2) recorded on a Neuroscan NuAmps.

1.3.3 Necessity for automated sleep staging

The development of automated sleep staging was driven by two objectives. Firstly, visual sleep staging is very time consuming and subjective, especially in patients with disturbed sleep; automated sleep staging should help to reduce the time needed by an experienced sleep specialist and reduce the inter-rater variability. Secondly, the limitations of the R&K rules in describing certain sleep disorders [26] have become clearer during the past decade. The automated sleep staging can give new quantative measures which correspond better to the extent of these disorders. So far, with the automated sleep staging, we can solve the problems we met before and to achieve objective, effective and precision judgment for sleep staging. However, the results of automated sleep staging systems currently available are not good enough to replace the manual scoring by R&K rules.

1.4 Sleep deprivation

In some medical literatures, the relation between sleep and depression was discussed. They suggested that more than 90% of depression patients may suffer from insomnia (e.g. Mendelson et al., 1977) and chronic insomnia is probable a trigger to cause depression relatively [6]. Disturbances of sleep are typical for most depression patients and belong to the core symptoms of the disorder. Polysomnographic sleep research has demonstrated that besides disturbances of sleep continuity, in depression sleep is characterized by a reduction of slow wave sleep and a disinhibition of REM sleep, with a shortening of REM latency, a prolongation of the first REM period and increased REM density. These findings have stimulated many sleep studies in depressive patients.

Most antidepressant drugs suppress REM sleep and it has therapeutic effects to improve endogenous depression [7]. Related researches were proposed that the mechanism in the treatment of REM sleep deprivation (REM-SD) similar with antidepressant drugs [28]. A study [8] experimented with 34 endogenous depressive patients. They were divided into two parallel groups, which were deprived of REM sleep and non-REM sleep respectively. After three weeks both groups were somewhat improved but the REM-deprived group improved significantly more than the group awakened from non-REM sleep. Therapeutic

efficacy of REM sleep reduction appeared similar to reported efficacy of antidepressant drugs treatment of depression. Eight of nine endogenous patients which unimproved by REM-SD also did not improve with antidepressant drugs.

The study of therapeutic effect with REM sleep deprivation does not have any new progress since Vogel's work was proposed in 1975 which was not repeated for a long time. One of the possible reason is that in general patients won't stay too long in the hospital so it is difficult to trace the effects in long term. The other reason is that since there always need a person who must keep awake to do sleep scoring for this patient and alarm him/her when the REM state is detected, it is a laborious task for the investigators. Due to these two difficulties, only few experiments were conducted with depression patients and used to confirm the effect of REM sleep deprivation. REM sleep deprivation may not apply popularly to the clinical therapy for depression. Thus, by the on-line automated sleep staging system, we can apply REM sleep deprivation for depression patients easily and conveniently.

1.5 Thesis overview

In this chapter, we introduce our objective of the thesis at first and then introduce about human sleep from sleep definitions to the scoring method in sleep medicine and extends to implementation of sleep deprivation. Chapter 2 begins to introduce the automated sleep staging. In this chapter, we introduce several related works of automated sleep staging using single EEG channel. By overview other's work, we will recognize many different methods for data preprocessing, feature extraction and classification. We may analysis the pros and cons of these algorithms, and extract some useful parts as reference method for our work. At the final section of Chapter 2, we will introduce our proposed work briefly and the details will be expounded in the next chapter. Chapter 3 provides the methods of our proposed automated sleep staging system. It will be introduced session by session following the order of data preprocessing, feature extraction and classification. In Chapter 4, we will present the experiment results. There are two parts of our experiments: the first part is the off-line testing with 25 records and the second part is the on-line testing. In

the on-line part, we implement the experiment of REM sleep deprivation with the on-line automated sleep staging system. The efficacy will present in the results. In Chapter 5, we summarize this work with the results and explain some possible reasons of misclassification of our method. In Chapter 6 we give some conclusions to this work and future works for extended research.


E S A
1896

Chapter 2

Automated Sleep Staging

2.1 Introduction to automated sleep staging

Since R&K rules were proposed in 1968, many automated sleep staging method had been proposed. Because of the labor-intensive and high inter-rater variability of manual scoring, researchers hope to use automated sleep staging to achieve the same task accurately and efficiently. In automated staging, we process the sleep staging as a classification problem. The variation of relative band energy is different according to different sleep stage. Also, the specific waveform may appear in some sleep stage, like sleep spindle and K-complex are features of Stage2. Characteristics of sleep stages were defined in R&K rules and the feature set which used for classification is usually referred to R&K rules. With automated staging, it is hard to decide the definite threshold value for scoring the sleep stage. Therefore, we process the automated sleep staging as a classification problem. As the conception of data mining, we extract "meaningful information" from sleep signals and transform the information into feature sets. The predicted result is more objective compared with manual scoring.

In automated sleep staging, there are two aspects of development. The one in that which physiological signals are used for automated staging and the other one is what kind of features of the analyzed signal can achieve the best classification. As mentioned in 1.3.2, signals of EEGs, EOGs, and EMGs are the three main requirements for scoring. Therefore, at the beginning of the automated analysis of sleep, researchers attempt to directly reference to R&K rules and translated these rules into programming codes to achieve scoring automatically. Afterward, researchers process the sleep scoring as a classification problem and many techniques of data mining were adopted to solve this classification problem. People tried to extract representative features of the signal from different sleep stages and classified by the methods of data mining. Also, they use the combination of EEG and EMG signals or the EMG and EOG signals for analysis. The correlation between these signals can as useful features for classification. Moreover, researchers intend to use only single channel of EEG for automated sleep staging. Many automated sleep staging methods base on single EEG signal analysis were proposed in recent years [19] and it is verified that using data from a single EEG is reliable for automated sleep staging [1] [4].

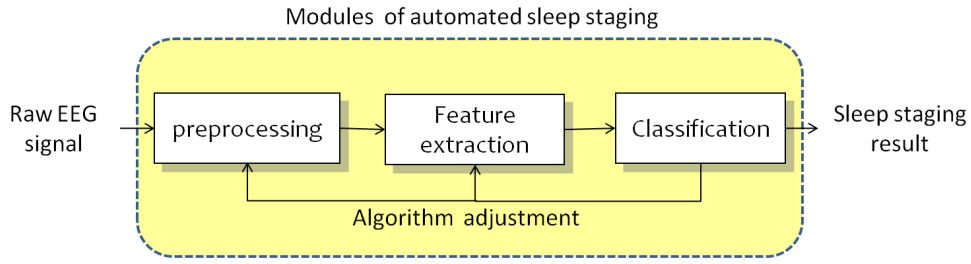


Figure 2.1: Framework of sleep stage classification

2.2 Survey of automated sleep staging methods

Generally, automated sleep staging follows a framework of classification: preprocessing, feature extraction and classification [27]. This framework is shown in Figure 2.1. The differences are in the selection of feature sets and the classification method. With various combination of different feature sets and classification methods, many methods had been tried in the past and we can make some coordination. In this section, we will introduce several methods of feature extraction and classification.

2.2.1 Feature extraction method

In the literature, many EEG features were tried in the past times [5]. Generally, EEG feature extraction method can be coordinated to four different kind which are frequently used. There are temporal analysis, spectral analysis, time-frequency analysis, and non-linear parameter analysis. Table 2.1 is an arrangement for the three different analysis methods corresponding to their parameters and descriptions.

Temporal analysis

In the time domain analysis, time signal is computed, like Hjorth's parameter [10]. It is based on the variance of the signal x[n] and its first and second derivative in an epoch

Table 2.1: Arrangement for the four different EEG analysis methods.

Method	Parameter	Description
Temporal	Hjorth parameter	Differentiate the continuous EEG signal for their variability.
Spectral	Power spectral analysis, Relative band energy	Generally split signal into four different frequency band, such as delta, theta, alpha and beta band.
Time-frequency	Wavelet transformation	The wavelet transform of a signal can be represented in terms of both time and frequency.
Non-linear	Approximate entropy, Spectral entropy	Calculate the signal complexity as non-linear parameters to quantify the cortical function at different sleep stages.

and derived three parameters: Activity, Mobility and Complexity for the quantification of EEG.

Spectral analysis

In the frequency domain analysis, the most representative feature extraction methods are the FFT-based spectral analysis. Powers of different frequency bands are commonly used as features. Usually, the range of frequency band in 0.5-50 Hz is divided into four frequency bands which are: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), and beta (13-30 Hz) band. For example, relative band energy is defined as the ratio of respective band energy to the total energy [20].

Time-frequency analysis

The Fourier transform (FT) can provide the frequency contents of the signal but it can't provide any time information. In contrast, time-frequency analysis can get the temporal information and the spectral information of the signal simultaneously. Thus, time frequency analysis is useful for the analysis of signals containing multiple time-varying frequencies. Wavelet transform (WT) is a function which can extract time-frequency pattern from the signal. Jobert et al. [15] illustrated advantages of the wavelet analysis over the Fourier analysis in sleep research. Further research in 1999, by the adaptive time-frequency lo-

calization property of the Wavelet Transform (WT), they proposed a method based on a specific Wavelet Packet Transform(WPT) and use the neural network for the classification task [11].

Non-linear analysis

Methods from nonlinear dynamics were introduced into the analysis of EEG signals. Since the EEG signal is non-stationary and noisy, non-linear analysis provide an effective method to study the dynamics of its complex behaviors. Nonlinear analysis is based on quantifying the signal complexity and yield additional information which is not redundant to the information gained by spectral analysis [14]. Rajendra et al. [24] proposed several non-linear analysis methods for quantifying different sleep stages, including correlation dimension, fractal dimension, largest Lyapunov entropy, approximate entropy, Hurst exponent, phase space plot and recurrence plots. With different values of each sleep stage, these values can be a useful feature sets for classification.

2.2.2 Classification method

For automated sleep staging, we expect to construct a stage classifier which can predict the new coming data. Thus, model learning with features sets extracted from the training data is a essential procedure. Two major categories of model learning techniques are unsupervised and supervised learning.

Supervised training

In supervised model learning, the training data consist of input vectors and each input vector may correspond to a class label. Supervised learning was implemented in neural network and it is commonly used for automated sleep staging [11] [13]. The structure of neural network construct with basically three layers of neurons: input layers, hidden layers and output layers. N feature vectors are input as N neurons in the input layer. Each neuron getting one of N feature. M neurons in the hidden layer fully connected to the first layer and P neurons in the output layer, with a defined target vector of zeros and a one in the

position according to an specific sleep stage. The output layer was fully connected to the hidden layer. The goal of the network was to correctly classify the 30 seconds epochs of sleep recording characterized by the N parameters.

Other supervised classification method like support vector machine had been used for automated sleep staging [21] [22]. It it very important whether the selection of feature sets is include enough information for classification. With the adequate feature sets, it will perform good classification in most cases.

Unsupervised training

In unsupervised training, by contrast, the model is trained with unlabeled data set. The example of unsupervised training for the sleep stage classifier is Hidden Markov Model (HMM). Arthur F. et al. [1] proposed a method of HMM for automated sleep scoring with single EEG channel. HMM is a method based on the transition probabilities between different stages. Rules of Hidden Markov Chain was proposed by Rabiner and Juang [23] and it is characterized as

1. Having a finite number of N states;
2. A new state is entered based upon a transition probability distribution A which depends on the previous state;
3. After each transition an observation output symbols is produced according to a probability distribution B which depends on the current state.

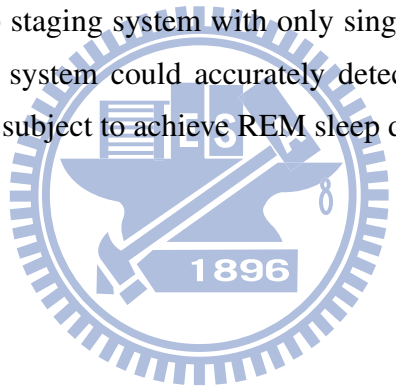
In the previous research, they produce three continuous probability taces P(wake), P(deep) and P(rem) with one second resolution based on HMM. With three class classification, it can achieve around 80% accuracy.

2.3 Thesis overview

After reviewing related researches for automated sleep staging, we prefer to adopt the method of support vector machine (SVM) which training the classifier with a supervised

learning method. Therefore, we need to find a set of features from single EEG which correspond with the properties of SVM and can achieve the better classification. The proposed automated sleep staging system will test by 25 subjects to verify its reliability and efficacy. Afterwards we may develop the original staging system to on-line staging that can output the staging result epoch-by epoch in real time. We perform an experiment of REM sleep deprivation using this on-line automated sleep staging system. Subjects enroll the experiment of REM sleep deprivation are all healthy people rather than depressed patients, since we want to test the feasibility of the on-line automated staging system help for REM sleep deprivation.

Figure 2.2 is the flowchart of our work, including the off-line testing with 25 recordings and extend system of on-line automated sleep staging. The algorithm of the staging in the off-line part is adopted in the on-line part. The objective of our work is to construct an on-line automated sleep staging system with only single EEG channel for assisting REM sleep deprivation. The system could accurately detect the REM state in real time and automatically alarm the subject to achieve REM sleep deprivation.



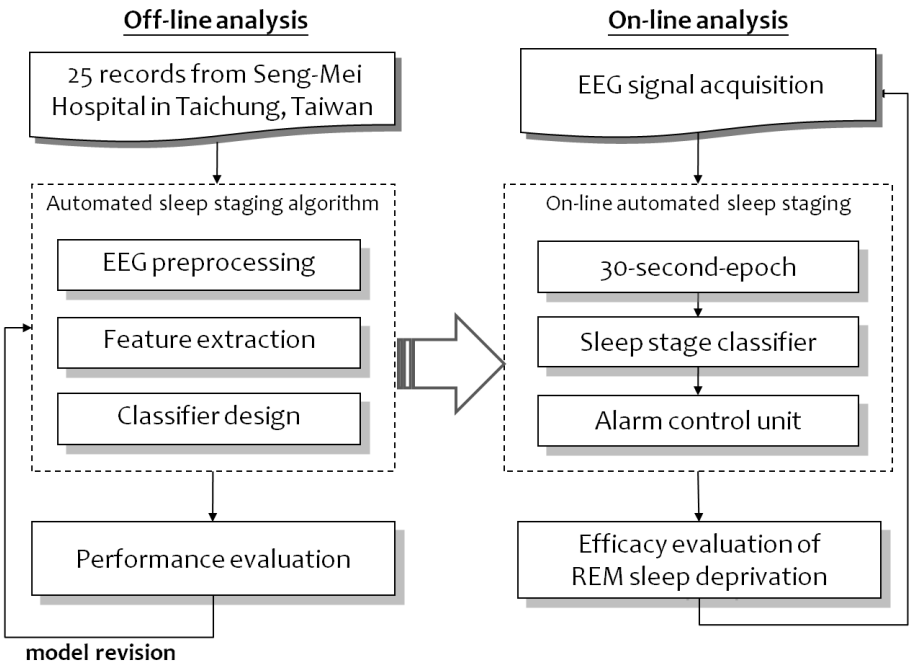


Figure 2.2: Flowchart of our work

Chapter 3

The Proposed Methods

The procedure of automated sleep staging can be split into three main steps: (1) EEG preprocessing, (2) feature extraction, and (3) classification. Raw EEG signals acquired from the device need preprocessing through the bandpass filter at first. Next, the preprocessed EEG signals will transform to feature vectors which are extracted in the frequency domain and take for the input sets for classification. For classification, we choose support vector machine (SVM) [2] as the classification method. In SVM, parameter adjustment is an essential step to obtain the better accuracy of training model . After the automated sleep staging system is constructed, we make some extension to develop the on-line automated sleep staging system which can detect the REM sleep in real time for assisting REM sleep deprivation.

3.1 Signal preprocessing

At first, we need to downsample the raw EEG signal to 100 Hz. After downsampling, the continuous time EEG signal are segmented with every 30-second interval. We choose the segment length to be 30 seconds which is as same as R&K standard. For an instance, an 8 hour records will split into 960 epochs.

3.2 Feature extraction

Since the purpose of our work is to implement automated sleep staging in real time, the efficiency of the signal processing time is the essential consideration. Moreover, as we only use a single EEG signal for staging, feature sets which can represent the difference of each stage is also another important consideration. For above reasons, the selection of feature extraction method is the most important task of automatic staging. Four feature extraction methods are mentioned in Section 2.2. With the time domain analysis, it is hard to present the signal variability between the different stages since scoring is based on single EEG analysis; With non-linear analysis, because of its high computational complexity, the computation time for the feature set may extend and thus not feasible for the on-line system. Due to the specific features of sleep EEG in some frequency bands is defined in R&K rule

and the computational complexity is moderate for on-line staging, we extract the feature sets in the frequency domain.

We divide the frequency band into seven sub-bands between 0.5 Hz and 45 Hz. Based on the four main rhythms of sleep EEG, we further subdivide them into lower and upper band within each range except for the alpha band. The δ band is split into $\delta 1$ (0.5-2.5 Hz) and $\delta 2$ (2.5-4 Hz); The θ band is split into $\theta 1$ (4-6 Hz) and $\theta 2$ (6-8 Hz); The β band is split into $\beta 1$ (13-25 Hz) and $\beta 2$ (25-35 Hz). Besides, we define a Total band with the range of 0.5-45 Hz. The EEG signal is filtered using the butterworth filter with the bandwidth of previously defined ranges in each frequency bands. The arrangement of these eight frequency bands is shown in Table 3.1.

Table 3.1: Eight frequency bands with its frequency range respectively

Frequency band	**Range (Hz)**
1. Total band	0.5-45
2. Lower delta band	0.5-2.5
3. Upper delta band	2.5-4
4. Lower theta band	4-6
5. Upper theta band	6-8
6. Alpha band	8-12
7. Lower beta band	13-25
8. Upper beta band	25-35

The raw EEG signal is denoted by $x(t)$ which is a time domain signal. After $x(t)$ is band pass filtered with eight frequency bands respectively, the *i th* filtered signal is denoted by $\widetilde{x_i}$(t). Thus, the mean amplitude of an epoch with a specific frequency band is computed by

$$\overline{m_i} = \frac{1}{l}\sum_{t=0}^{l}\widetilde{x_i}(t) \tag{3.1}$$

where *l* denoted sample points of an epoch that depends on the sampling rate and epoch length.

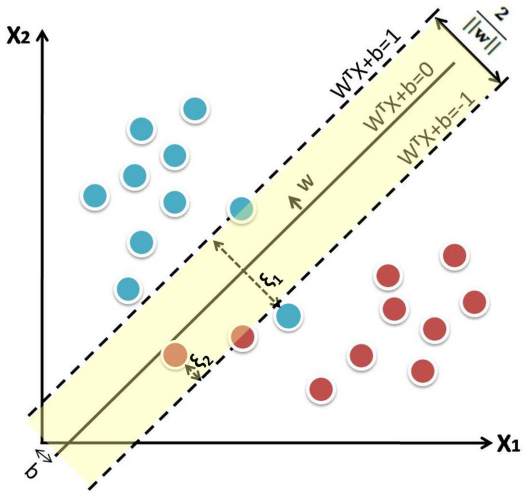


Figure 3.1: Illustration of SVM classifier. Here are two different data and we need to find a decision boundary (solid line) to separate two classes of data and the margin (dashed line) between the two is maximal. Data locates on the margin is defined as support vector; ξ is the error value of misclassified.

Thus, the feature vector of the *n th* epoch is composed of mean amplitudes with eight frequency bands which can be represented as $[\overline{m_{1n}}, \overline{m_{2n}}, \overline{m_{3n}}, \overline{m_{4n}}, \overline{m_{5n}}, \overline{m_{6n}}, \overline{m_{7n}}, \overline{m_{8n}}]$,

3.3 Classification

Introduction to SVM

We use support vector machine as our classifier. It is a statistical-based classification method that many applications had been proved by their results. High dimensional data set can be easily processed and since the analysis of physiological signals is a high complexity computation, it is suitable to use SVM for analysis. The goal of SVM is to train a model which predicts target value of data instances in the testing set which are given only the attributes. The basic idea underlying SVM is to determine a hyperplane which can separate the two different classes so that the margin between the training data and decision boundary is maximal. Figure 3.1 illustrate the conception of SVM.

The classification problem can be stated as follows: Assume that there are M different classes of objects, given a new object assign it to one of the M classes. Given a training set of instance-label pairs (x_n, y_n), n=1,...,l which x_n denotes a feature vector of instance n, y_n denotes a known class label. It can be solved as a optimization problem:

$$\begin{aligned} \min_{w,b,\xi_i} \quad & \frac{1}{2}w^T w + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad i = 1, ..., l \end{aligned} \tag{3.2}$$

$\sum \xi_i$ is an upper bound on the number of training errors and C is a parameter that controls trade-off between margin and error. The feature vector x_n is mapped into a higher dimensional space by the function ϕ in order to find a linear separating hyperplane with the maximal margin in this higher dimensional space. The function ϕ is a kernel function. There are many kinds of kernel functions and radial basis function (RBF) is in general used since it maps nonlinear separable data into a linear space where they are almost linearly separable. RBF is defined by

$$K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2), \gamma > 0 \tag{3.3}$$

where x_i and x_j are two different feature vectors and γ is a kernel parameter defined by user.

Due to the specificity of different subjects, every subject need to uses their own training data to construct a subject-specific classification model and predict the new given data by this model. The predicted result then compare to the ground truth which is scoring by specialists.

Parameter selection

The regularization parameter C and the kernel parameter γ are important to the model training. Different (C, γ) values are tried to see which one gives the highest cross validation accuracy. We then use the best parameter to train the whole training set and generate the

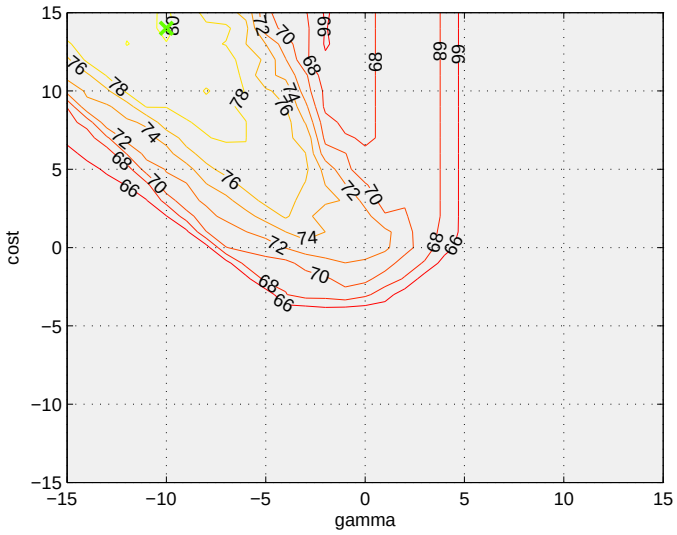


Figure 3.2: Parameter searching with grid search. The exponent of (C ,γ) are searching in the range of [-15,15]x[-15,15]. The green cross represent the global optima in the range of [-15,15]x[-15,15] and it located in (14, -10).

final model. Cross-validation (CV) is often used to evaluate classifier accuracy. We split the training data into five disjoint subsets of equal size. The classifier is trained five times. Each time a different subset is left out for testing and the classifier trained on the remaining four subsets. The average accuracy over all the test sets is an estimate of the true classifier accuracy. Also, the cross-validation procedure can prevent the overfitting problem.

For the parameter searching, we use the simple grid search [3] on the following points of (C, γ): [2^{-15}, 2^{-13}, 2^{-11}, ..., 2^{13}, 2^{15}]$\times$ [2^{-15}, 2^{-13}, 2^{-11}, ..., 2^{13}, 2^{15}] and the one with the best cross-validation rate is picked, for example, (2^{-5},2^{7}). Later, we conduct a finer grid search in the sub-range which the exponent add one and subtract one, that is [2^{-4}, $2^{-4.25}$, $2^{-4.5}$, ..., $2^{-5.75}$, 2^{-6}]$\times$ [2^{6}, $2^{6.25}$, $2^{6.5}$, ..., $2^{7.75}$, 2^{8}]. While the best accuracy in this sub-range is found, the CV rate then compare to the one which is found before to decide which pair of (C, γ) is used. After the best (C ,γ) is found, the whole training set is trained again to generate the final classifier.

SVM applying on the automated sleep staging

In our work, the sleep stage classification problem can be stated as follows: There are three different classes of stage: WAKE, REM, and NREM. Each epoch is computed with their values of mean amplitude in eight different frequency bands which form the feature vector. After the training data was transformed by feature extraction method, it can be denoted by (x_1, y_1), ..., (x_n, y_n), which x_n denotes a feature vector of epoch n, $x_n \equiv [\overline{m_1}_n, \overline{m_2}_n, \overline{m_3}_n, \overline{m_4}_n, \overline{m_5}_n, \overline{m_6}_n, \overline{m_7}_n, \overline{m_8}_n]$, y_n denotes a known class label obtained from sleep specialist, $y_n \in$ {Wake, REM, NREM}.

3.4 On-line system for REM sleep deprivation

The flow chart of the on-line system for REM sleep deprivation is shown in Figure 3.3. The EEG signal continuously acquiring from the device. While the 30-second buffer is occupied, it will send to the classifier for scoring. As the REM state is detected, the system will send a stimulus to disturb the subject. The stimulus we use is a shrill sound which may continue ringing until the sleep state of the subject transform to the other state from the REM state. REM sleep will continue for a period of time. The unavoidable phenomenon of the automated sleep staging is that the oscillation between the two state usually appear in the scoring result. Therefore, we set a temporal constraint which is defined as follows: consecutive epoch prior to this epoch are all belong to the REM state, then the system send an alarm to disturb the subject; otherwise, the subject may not incur any disturbance during sleep. With this constraint, the number of false alarm of the system will reduce to avoid excessive disturbance to the subject.

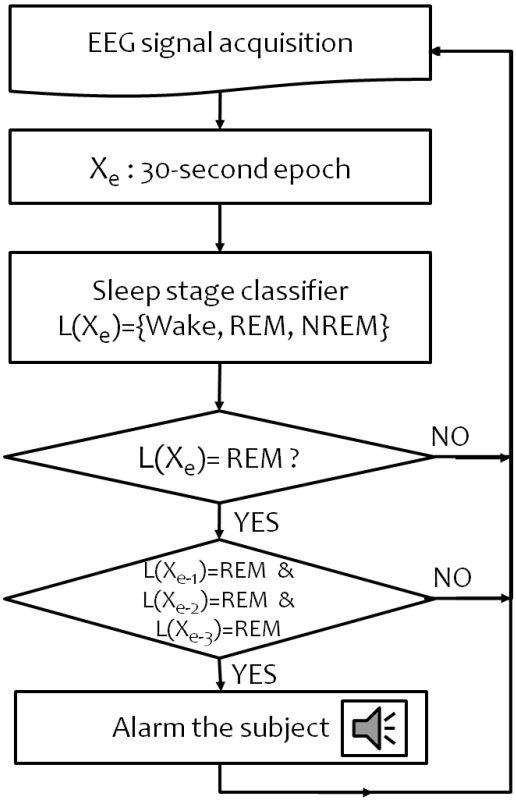


Figure 3.3: Flowchart of the on-line automated sleep staging system for assisting REM sleep deprivation

Chapter 4

Experiments

4.1 Off-line analysis

4.1.1 Data preparation

Subject

Before implementing the on-line automated sleep staging system, we analyze data set with off-line analysis to decide the methods of automated sleep staging in each step and verify the accuracy of staging. The data sets were obtained from the sleep center of Sang-Mei Hospital in Taichung, Taiwan. There are 25 records from 10 males and 15 females. These subjects probably have some kind of sleep disorders but it will not affect our experiment results.

Records

The PSG was recorded digitally with $Sandman^{®}\ Digital20^{TM}$ system. Data were collected by nocturnal polysomnography, including four EEG channels (C3/A2, C4/A1, O1/A2, O2/A1), EMG of chin, EOG of the left side and the right side, EKG, SpO_2, airflow and sleep position. These records were saved as a EDF (European Data Format) file which is a standard format in sleep research. Each subject has two nocturnal sleep records about 6 to 8 hours. In the second night sleep, continuous positive airway pressure (CPAP) is added to the subject so the subject sleeps better than the day before, which means the sleep cycles may include more complete stages. Therefore, the recording of which the CPAP is added is used as the training data while the other recording as testing data. These records have been previously scored by sleep specialists and assumed as ground truth for comparison with results of the proposed automated staging method. We extracted EEG signal of C3/A2 channel from PSG records as the data set for automated scoring.

4.1.2 Experimental results

The data set include 25 subjects. All are well scoring by sleep specialists as our ground truth and compare to the results from our system. For the off-line analysis, we classified sleep stages into three classes: Wake, REM and NREM.

While we classified as three stages, the accuracy of the proposed automated sleep staging system is 69.5% to 93.8% and 85.04% in average. Table 4.1 shows the result of 25 subjects. It shows the percentage of total accuracy and the REM accuracy for every records. Total accuracy means that if predicted label is same as ground truth and it takes into count. The REM accuracy means that both ground truth and predicted label is labeled as REM state and it takes into count. The REM accuracy of some records are very low and this will be discussed in the next chapter.

Table 4.1: List of total accuracy and REM accuracy of 25 subjects. Total accuracy means that if predicted label is same as ground truth and it takes into count; The REM accuracy means that both ground truth and predicted label is labeled as REM state and it takes into count. The REM accuracy of Subject 3 is labeled as – which means there is no REM state in this record.

Subject	**Total accuracy**	**REM accuracy**	**Subject**	**Total accuracy**	**REM accuracy**
s0102	84.7% (728/859)	77.0% (117/152)	s2102	82.9% (636/767)	100.0% (63/63)
s0302	88.6% (616/695)	87.3% (89/102)	s2202	81.7% (686/840)	54.4% (136/250)
s0402	77.7% (483/622)	–	s2302	84.9% (699/823)	56.4% (119/211)
s0502	91.4% (620/678)	90.5% (143/158)	s2402	86.8% (688/793)	63.4% (102/161)
s0602	76.4% (616/806)	40.4% (88/218)	s2502	86.8% (665/766)	83.7% (41/ 49)
s0802	80.9% (637/787)	100.0% (81/ 81)	s2602	93.7% (731/780)	85.7% (96/112)
s1202	81.1% (664/819)	58.6% (163/278)	s2702	90.0% (764/849)	85.0% (142/167)
s1402	86.2% (702/814)	89.0% (219/246)	s2802	90.7% (640/706)	74.5% (82/110)
s1502	90.9% (687/756)	86.1% (99/115)	s2902	86.9% (672/773)	64.5% (107/166)
s1702	88.3% (722/818)	67.8% (99/146)	s3102	84.9% (656/773)	96.4% (135/140)
s1802	83.7% (769/919)	58.9% (201/341)	s3202	86.2% (676/784)	58.8% (100/170)
s1902	93.8% (804/857)	77.6% (142/183)	s3302	78.6% (582/740)	65.0% (104/160)
s2002	83.5% (641/768)	69.0% (172/228)			

Table 4.3 is the confusion matrix of all 25 records. Totally there are 19592 epochs in 25 record. An element in row i and column j counts the number of times class j was classified as i. Diagonal elements count the number of correct classifications and off-diagonal ele-

ments count the number of misclassifications. The last column of the matrix is the positive predictive value (PPV) and used to evaluate the accuracy of classification. The relationship among terms are shown as Table 4.2. PPV is defined by the ratio of TP and the sum of TP and FP.

Table 4.2: Relationships among terms. Definitions of True positive (TP), False positive (FP), True negative (TN) and False negative (FN)

		Condition(ground truth)	
		positive	**negative**
Test outcome	**positive**	True Positive (TP)	False Positive (FP)
	negative	False Negative (FN)	True Negative (TN)

Table 4.3: Confusion matrix of 25 Subjects. An element in row i and column j counts the number of times class j was classified as i. Diagonal elements count the number of correct classifications and off-diagonal elements count the number of misclassifications. The last column of the matrix is the positive predictive valus of the classification result which is defined by the ratio of true positive and total epoch of a state. There are totally 19592 epochs in 25 subjects.

		Manual scoring				
		Wake	**REM**	**NREM**	**Total**	**PPV**
Automated scoring	**Wake**	**1567**	161	290	2018	77.7%
	REM	65	**2840**	880	3785	75.0%
	NREM	404	1006	**12379**	13789	89.8%
	Total	2036	4007	13549	19592	**85.6%**

Figure 4.1 shows the hypnogram of the sleep structure for Subject s1902. The accuracy of the record is 93% which is the highest accuracy of 25 records. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line represents the results of manual scoring which we assumed as the ground truth, and the red line is the results of automated scoring by our system. Table 4.9 is the confusion matrix for subject19. The numbers in bold face represent those correctly classified.

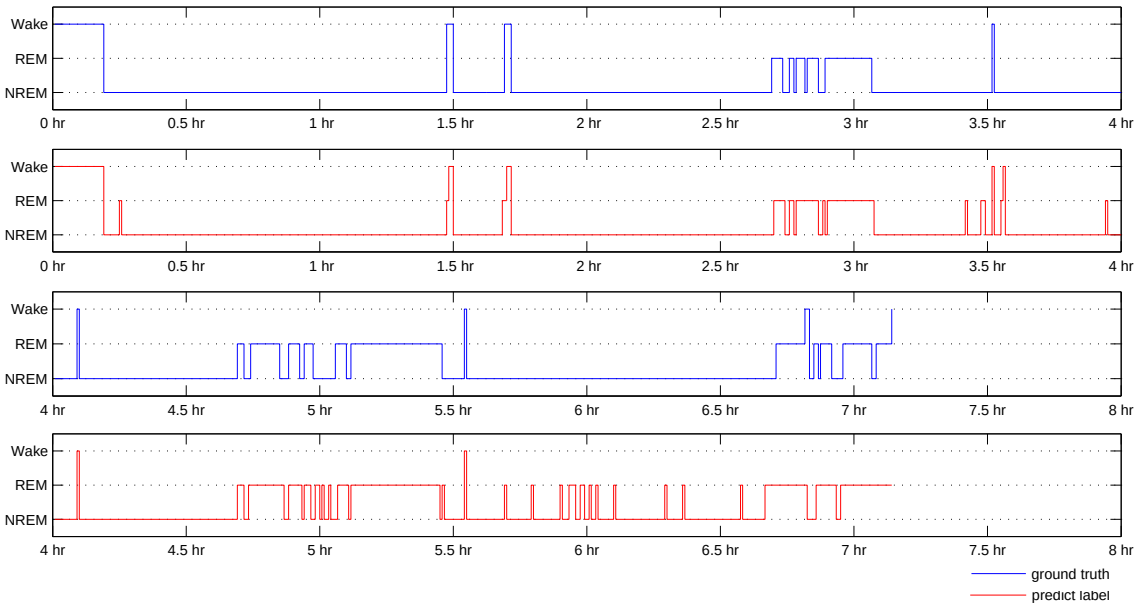


Figure 4.1: Hypnogram of Subject s1902. The accuracy of the record is 93%. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line and the red line are represented as the result of manual scoring and the result of automated scoring by our system respectively.

Table 4.4: Confusion matrix of Subject s1902.

		Manual scoring				
		Wake	**REM**	**NREM**	**Total**	**PPV**
Automated scoring	**Wake**	**28**	2	4	34	82.4%
	REM	0	**142**	6	148	95.9%
	NREM	2	39	**634**	675	93.9%
	Total	30	183	644	857	93.8%

4.2 On-line analysis

4.2.1 Experiment setup

Subjects

Nine volunteers (aged 26±3 years, 2 men and 7 women) were enrolled in the study and tested with three questionnaires: Beck Depression Inventory-II (BDI-II), Beck anxious inventory (BAI) and Multidimensional Fatigue Symptom Inventory-Short Form (MFSI-SF). Followings are introduction about three questionnaires:

- **Beck Depression Inventory (BDI):** The Beck Depression Inventory (BDI-II), created by Dr. Aaron T. Beck, is one of the most widely used for measuring the severity of depression. BDI-II contains 21 questions, each answer being scored on a scale value of 0 to 3. Higher total scores indicate more severe depressive symptoms. The cutoffs are as follows:
  - 0-13: minimal depression
  - 14-19: mild depression
  - 20-28: moderate depression
  - 29-63: severe depression

- **Beck Anxiety Inventory (BAI):** The Beck Anxiety Inventory (BAI), created by Dr. Aaron T. Beck, is used for measuring the severity of an individual's anxiety. BAI contains 21 questions, each answer being scored on a scale value of 0 to 3. Higher total scores indicate more severe depressive symptoms. The cutoffs are as follows:
  - 0-7: miminal level of anxiety
  - 8-15: mild anxiety
  - 16-25: moderate anxiety
  - 26-63: severe anxiety

- **Multidimensional Fatigue Symptom Inventory (MFSI):** is a 30-item short form of the MFSI that yield scores only for the empirically derived subscales. Preliminary research suggests that it has acceptable psychometric properties and may be used as a substitute for the MFSI when time constraints and scale length are of concern. The cutoffs are as follows:
  - General scale = sum of items 10, 12, 14, 17, 18, and 28
  - Emotional scale = sum of items 3, 8, 13, 21, 23, and 30
  - Physical scale = sum of items 2, 4, 6, 16, 19, and 26
  - Mental scale = sum of items 1, 11, 15, 20, 25, and 27
  - Vigor scale = sum of items 5, 7, 9, 22, 24, and 29
  - Total score = (General + Physical + Emotional + Mental) - Vigor

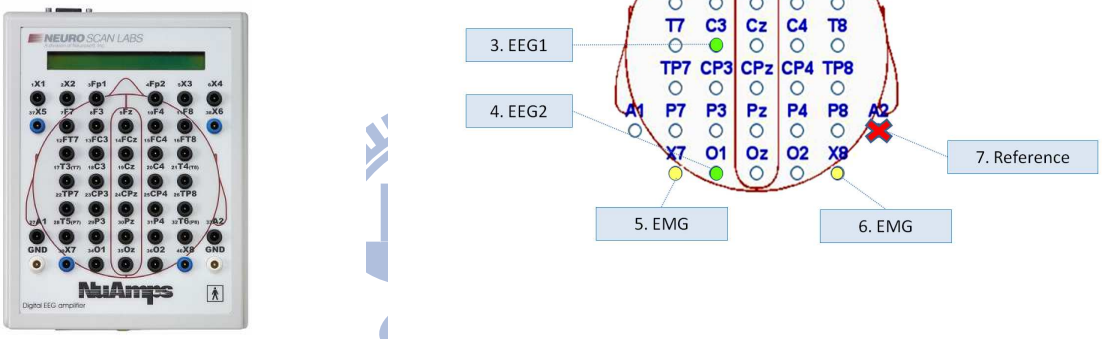
These questionnaires were used for selecting healthy subjects which have good sleep quality and without any mental disorders. We acquired sleep records in the sleep laboratory at Tri-Service General Hospital in Taipei, Taiwan. Arrangement of scores are shown in Table 4.5.

Table 4.5: Arrangement of scores of 9 subjects.

Subject	**BDI**	**BAI**	**MFSI-SF**						R/A
			Gen	Phy	Emo	Men	Vig	Total	
s001	0	1	2	0	0	0	18	-16	R
s002	7	5	4	2	6	5	15	2	R
s003	8	1	2	2	0	3	5	2	R
s004	24	28	12	13	3	8	6	30	A
s005	4	2	1	1	2	4	12	-4	A
s007	7	29	1	2	0	3	11	-5	A
s008	29	3	2	1	2	1	2	4	A
s009	6	15	7	6	10	10	12	21	A
s010	6	3	4	1	0	3	10	-2	A

Electrode placement

In each subject, we recorded two full-night polysomnography using a recording device (NuAmps, Compumedics Neuroscan, CO, USA; figure is shown in Figure 4.2(a)) with a bandwidth of 0.5-70 Hz. NuAmps is a DC amplifier designed to record a wide variety of multichannel neurophysiological signals, such as EEG, EOG, ECG and EMG.



(a) NuAmps (b) Electrode placement

Figure 4.2: (a) NuAmps; (b) Electrodes placement of NuAmps. There are total seven channels: 2 EEG, 2 EOG , 2 EMG and a reference(A2).

Data was recorded by 2 EEG, 2 EOG, and 2 EMG channels. EEG was recorded according to the to the international 10-20 international standard [12] with central EEG (C3) and occipital EEG (O1); Right EOG is placed slightly lateral and 1 cm up from the outer canthus and left EOG is placed slightly lateral and 1 cm up from the outer canthus; 2 EMG is placed in submental area. The electrode placement is shown in Figure 4.2(b).

On-line sleep deprivation

Every subject have two nocturnal sleep records. In the first night, we use the the software *Acquire* to record the sleep. This full-night sleep record will then scoring by sleep

specialist to obtain the class labels. With the first night recording as the training data and the obtained class labels, we train a classifier and this model will then apply to the on-line automated staging. In the second night, the acquired EEG signal need of real-time analysis. We develop a program which can record the signal continuously and process the sleep staging every in thirty seconds. Once the system have been detected the REM state, it may alarm the subject automatically to deprive his/her REM sleep. In the on-line automated sleep staging system, the system will record the epoch number and the time of occurrence of the alarm. As the hypnogram shown, we can see that the REM state will change after the alarm and the REM sleep may reduced in full-night sleep.

Since the proposed algorithm was based on single EEG channel, we use only EEG signal of C3/A2 leads with the record length of 6 to 10 hours for the automated sleep staging. For each subject, the recording of the first night sleep is used as the training data and used for training his own model. In the second night sleep, we acquire the sleep records with a real-time analysis for every 30-second-epoch by his own training model. Afterward, the second night record is scored by sleep specialists and this result is as a ground truth for comparison with automated scoring results.

4.2.2 Experimental results

In the on-line analysis, there are 6 records. We apply sleep deprivation to the second night to the subject. Table 4.6 outline the total accuracy and the REM accuracy of six records. Table 4.7 is the confusion matrix of total six records.

When applying REM sleep deprivation, the control of the alarm is an important issue. We use two factors to evaluate the correctness of the alarm occurrence while the REM state was detected. We defined *true alarm* rate to evaluate the correct detection of the REM state and alarm the subject; another factor is *false alarm* rate. It counts for alarm occurring when the system detect as REM sleep but actually subject is in Wake state or NREM sleep. The definition of two factors are presented in Table 4.8.

The evaluation of the true alarm rate is defined as:

$$\frac{\mid C_A \cap C_M \mid}{\mid C_A \mid} \tag{4.1}$$

Table 4.6: List of total accuracy and REM accuracy of 6 subjects.

Subject	Total accuracy	REM accuracy
s00402	89.6% (549/613)	83.3% (90/108)
s00502	87.2% (1070/1227)	88.8% (198/223)
s00702	83.0% (995/1199)	78.6% (286/364)
s00802	75.6% (914/1209)	65.0% (119/183)
s00902	84.4% (579/686)	75.4% (46/61)
s01002	80.8% (600/742)	69.6% (112/161)

Table 4.7: Confusion matrix of totally 6 records of on-line analysis.

		Manual scoring				
		Wake	**REM**	**NREM**	**Total**	**PPV**
Automated scoring	**Wake**	**585**	61	194	841	69.7%
	REM	48	**750**	201	999	75.1%
	NREM	227	238	**2870**	3335	86.5%
	Total	861	1049	3309	5175	86.1%

Table 4.8: An element in row i and column j represent that class j is classified as i. False alarm is defined as REM state is misclassified as other two states; false detection is defined as the sleep state is misclassified as REM state.

		Manual scoring		
		Wake	**REM**	**NREM**
Automated scoring	**Wake**			
	REM	*false alarm*	*true alarm*	*false alarm*
	NREM			

where C_A denotes the set of REM epochs determined by automated scoring and C_M denotes the set of REM epochs determined by manual scoring.

The evaluation of the false alarm rate is defined as:

$$\frac{\mid C_A \cap C_{NM} \mid}{\mid C_A \mid} \tag{4.2}$$

where C_{NM} denotes the set of Wake and NREM epochs determined by manual scoring.

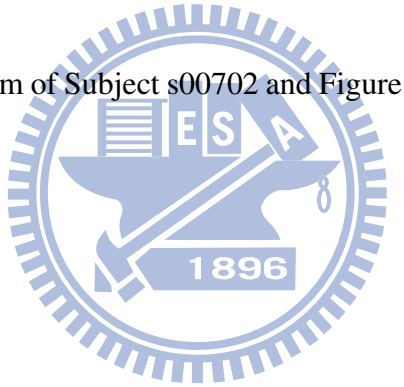
It doesn't matter that false alarm occurring in the wake state; by contrast, false alarm occurring in the NREM state will disturb the subject's sleep and it need to be reduced as possible. Thus, the temporal constraint was added for the on-line system which could effectively reduce the false alarm. Automated staging usually have the problem of state oscillation that the stage transit quickly between the two states. The truth is that the time of REM sleep remains for thirty minutes or even longer, not just last for 30 seconds or 1 minutes (1 or 2 epochs). There are two requirements in the temporal constraint. One is that the processing epoch is scored as REM state and the other is that three epochs prior to the processing epoch are all scored as REM state. The system will alarm the subject if these two requirements are both satisfied. Therefore, the total numbers of alarms will decrease by adding the temporal constraint in the automated staging system. The alarm only occurs when the subject is in a stable REM sleep. Table 4.9 demonstrated the comparisons between the result of "with temporal constraint"and "without temporal constraint". We can see that the percentage of true alarm increased and false alarm decreased which illustrate the temporal constraint has improved the REM accuracy for REM sleep deprivation.

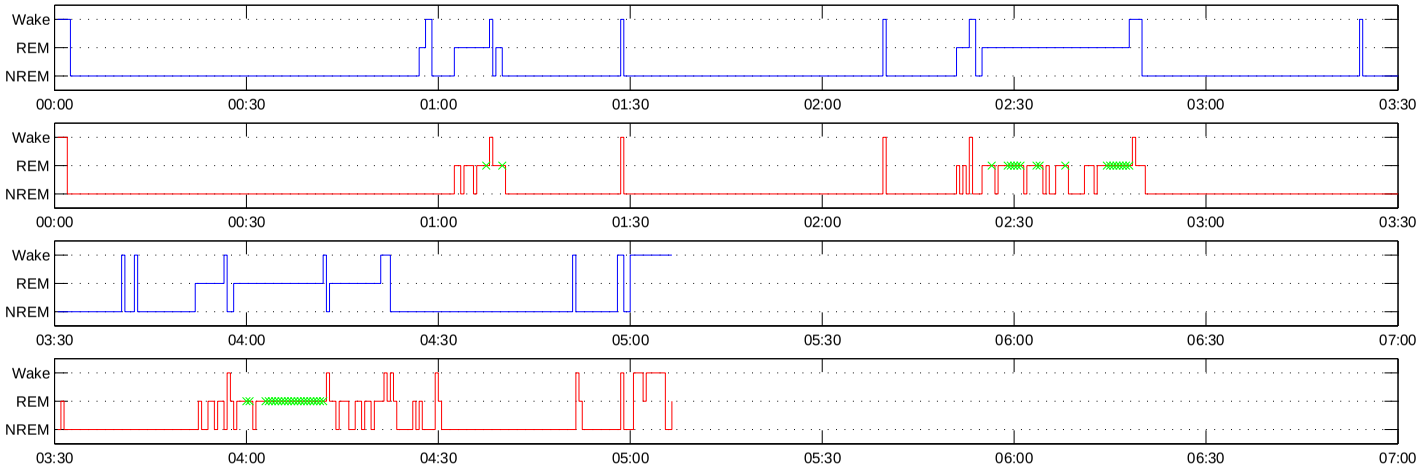
Following are the hypnograms and confusion matrices of three different subjects. Figure 4.3 is the hypnogram of Subject s00402. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line represents the results of manual scoring which we assumed as the ground truth, and the red line is the results of automated scoring by our system. The green dots represent that the alarm occurring. As the figure shown, once the system detect the REM state, the system will not alarm the subject except three consecutive epochs prior to this epoch are judged as REM state and thus the false alarm may reduce. The alarm produced by system will disturb the subject and the REM state may either change to the Wake state or NREM state in his sleep.

Table 4.9: Comparisons between the results of "with temporal constraint"and "without temporal constraint". True alarm means it is predicted as REM state in accordance with the ground truth. False alarm means that it is predicted as REM state but actually it is NREM state.

	W/O temporal constraint		**W/ temporal constraint**	
Subject	**True alarm**	**False alarm**	**True alarm**	**False alarm**
s00402	83.3% (90/108)	9.3% (10/108)	92.5% (37/40)	2.5% (1/40)
s00502	72.8% (198/272)	23.5% (64/272)	80.4% (181/225)	16.0% (36/225)
s00702	81.7% (286/350)	13.1% (46/350)	89.7% (157/175)	4.6% (8/175)
s00802	54.6% (119/218)	39.0% (85/218)	60.0% (12/20)	0.0% (0/20)
s00902	80.7% (46/57)	8.8% (5/57)	60.0% (6/10)	0.0% (0/10)
s01002	83.0% (112/135)	7.4% (10/135)	89.2% (66/74)	0.0% (0/74)

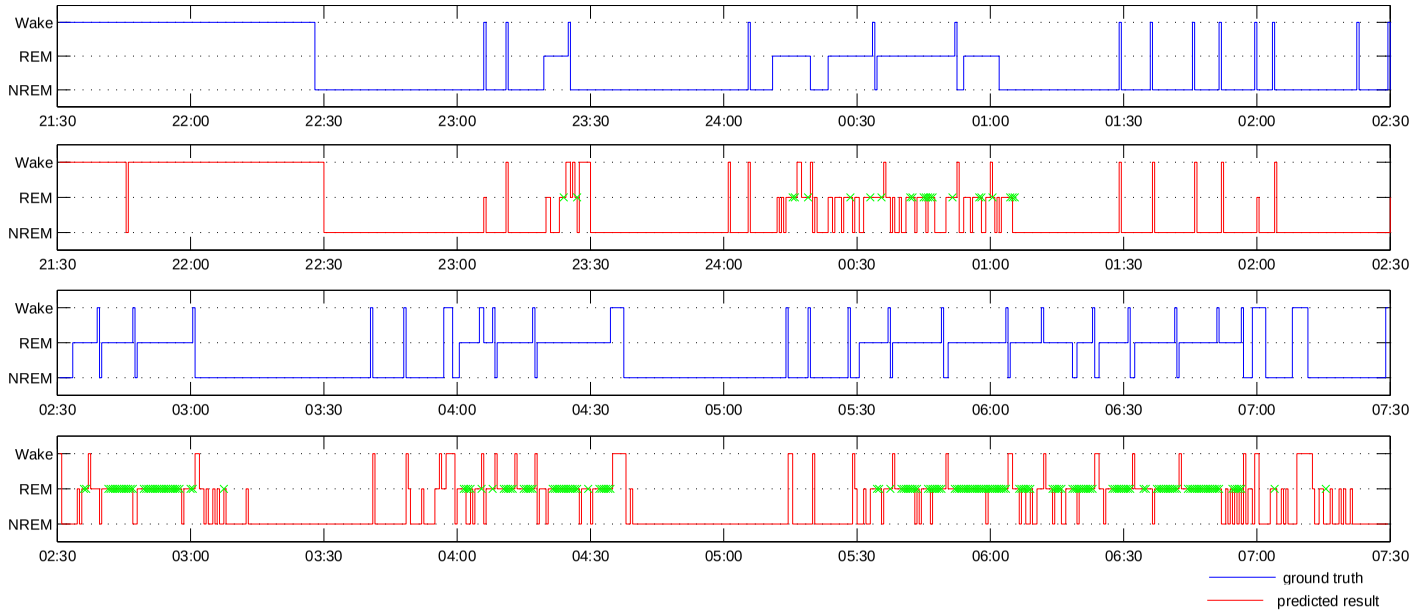
Figure 4.4 is the sleep hypnogram of Subject s00702 and Figure 4.5 is the sleep hypnogram of Subject s01002.





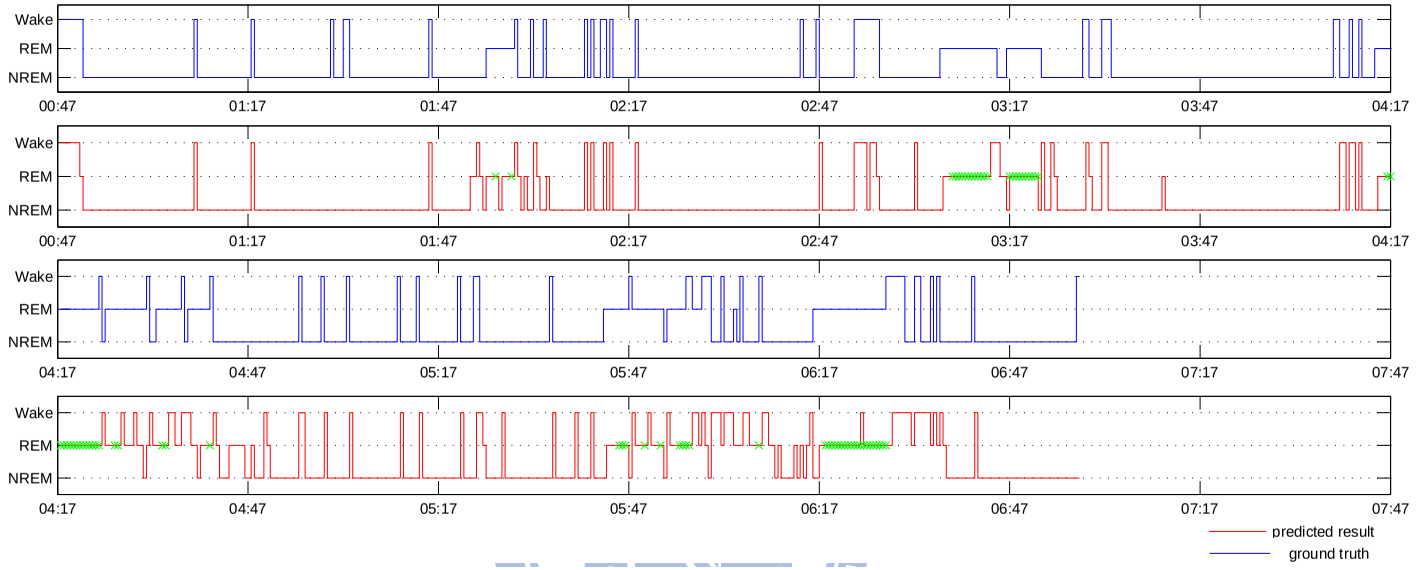
		Manual scoring				
		Wake	**REM**	**NREM**	**Total**	**PPV**
Automated scoring	**Wake**	**440**	28	13	481	91.5%
	REM	10	**90**	8	108	83.3%
	NREM	5	0	**19**	24	79.2%
	Total	455	118	40	**613**	89.6%

Figure 4.3: Hypnogram of S00402. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line represents the results of manual scoring which we assumed as the ground truth, and the red line is the results of automated scoring by our system. The green dot represent the alarm occurring.



		Manual scoring				
		Wake	**REM**	**NREM**	**Total**	**PPV**
Automated scoring	**Wake**	**134**	11	39	184	72.8%
	REM	28	**286**	42	356	80.3%
	NREM	23	67	**575**	665	86.5%
	Total	185	364	656	**1205**	82.6%

Figure 4.4: Hypnogram of S00702. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line represents the results of manual scoring which we assumed as the ground truth, and the red line is the results of automated scoring by our system. The green dot represent the alarm occurring.



		Manual scoring				
		Wake	**REM**	**NREM**	**Total**	**PPV**
Automated scoring	**Wake**	**36**	15	25	76	47.4%
	REM	13	**112**	10	135	83.0%
	NREM	45	34	**452**	531	85.1%
	Total	94	161	587	**742**	80.9%

Figure 4.5: Hypnogram of S01002. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line represents the results of manual scoring which we assumed as the ground truth, and the red line is the results of automated scoring by our system. The green dot represent the alarm occurring.


E S A
1896

Chapter 5

Discussions

5.1 Parameter searching

In SVM, parameter selection is an essential procedure. The γ in the kernel function and the C in the objective function are two parameters need to select. We use the grid search to find a (C,γ) pair which can achieve the best cross-validation rate. The searching range of (C, γ) referred to [2] is recommended to be [-10, 10] but we found that for some cases it may exceed this range. Thus, we expand the searching range of (C, γ) both to [-15, 15]. Although higher CV rate represent the training model is more accurate but the CV rate is not proportional to the predicted accuracy. With the grid search, system will select a pair of (C, γ) with the highest CV rate as parameters for training the model. However, the testing data predicted with the model with highest CV rate does not represent that can achieve the highest classification accuracy. For example, as shown in the figure, the x-axis represent the 140 different (C, γ) pairs have been tried. At the 54*th* (C, γ) pair, the CV rate is 72.88% and the classification accuracy is 83.93%; at the 55*th* (C, γ) pair. Figure 5.1 shows the relationship between the CV rate and predicted accuracy.

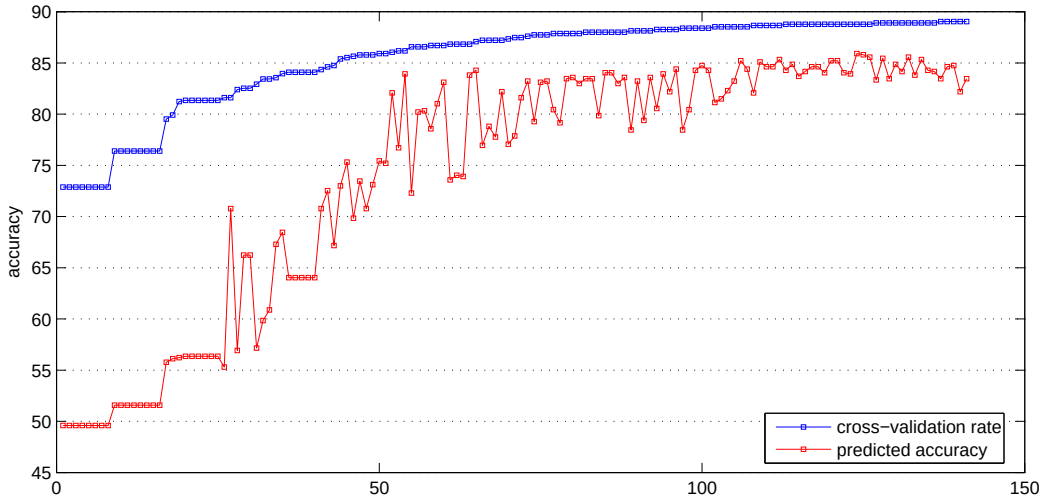


Figure 5.1: The relation between predicted accuracy and cross validation rate. The predicted accuracy is not proportional to the cross validation rate.

5.2 Problems of automated staging

5.2.1 Reasons for bad classification

From the results of our experiment, the effectiveness of our automated staging system is promising. The accuracy of 3-stage classification can achieve 85% in average with 25 normal subjects. By the manual scoring with specialists, they score visually based on EEG, EOG and EMG signals. Since we use only single EEG channel for scoring, the quality of the EEG signal is very important. Following are the reasons which will impact the accuracy of classification.

Signal quality

For physiological signal analysis, the analysis result often depends on the signal to noise rate. Due to the sleep EEG acquisition, it is unavoidable that the signal interferences cause by the body movement. Those movement may cause an obvious signal oscillation as to affect the power distribution of the original signal and may affect the scoring result. This artificial interference will be rejected when scoring by manual, but it is hard to define a specific threshold to reject them.

Limited information from single EEG

The features obtained from the EEG signal is limited especially for the information of the REM state. Figure 5.2 is the statistical plot of mean amplitude value of eight frequency band with respect to three stages. The blue, red and green line represent stage W, REM and NREM respectively. We can see that in the Wake group, the α (8-12 Hz), $\beta 1$ (13-25 Hz) and $\beta 2$ (25-35 Hz) are more significant in three stages. By contrast, $\theta 1$ (4-6 Hz) and $\theta 2$ (6-8 Hz) are more significant in the NREM group. However, there is no any significant values in the REM group.

Problems of supervised classification

Following we will discuss about the reasons that SVM perform the bad accuracy of classification. In the results of off-line analysis with 25 subjects, the variation of the pre-

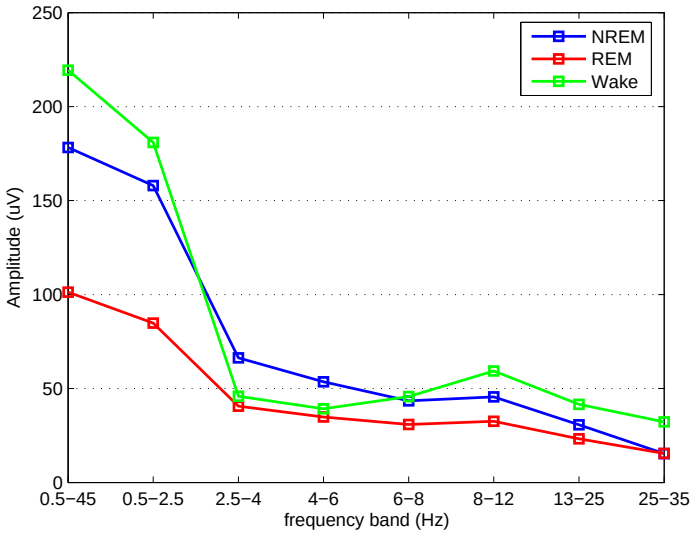


Figure 5.2: The statistical plot of mean amplitude value of eight frequency band with respect to three stages. The blue, red and green line represent stage W, REM and NREM respectively.

dicted accuracy is low to 76% and high to 93%. The procedure of the automated staging is the same to every subject but the result is in divergence. We conclude two factors of this phenomenon.

In the off-line analysis, our data were from 25 subjects with sleep disorders. Sleep recordings acquired with CPAP added were used as testing set. Thus, the sleep structure between training set and testing set are a little bit different. This will effect the classification accuracy. However, we have tried splitting the same recording with two parts: the odd epochs were used as training data while the even epochs were as testing data; and on the contrary, the even epochs were used as training data while the odd epochs were as testing data. The total accuracy is averaged by this two parts. We compared the results with the results of predicting by trained models as shown in Table 5.1. We can see that to the same recording our feature sets were reliable. The accuracy of classification also depends on the variation between training data and testing data.

Another reason is that if there are many differences of signal patterns between the model and newly predicted record, then the predicted result will be bad. If some stage lacks in the

Table 5.1: Comparisons of the classification accuracy with two different training methods. One is training with a sleep record and testing with another one record for the same subject; the other method is using only one sleep record and training with its odd epochs and testing with its even epochs. We can see that the accuracy of testing with another recording by a trained model will effect by the variations of testing data and training data.

Subject	**Trained with whole record**	**Trained with even(odd) epoch**
s0102	84.7% (728/859)	88.1% (757/859)
s0302	88.6% (616/695)	92.7% (644/695)
s0402	77.7% (483/622)	89.7% (558/622)
s0502	91.4% (620/678)	94.5% (641/678)
s0602	76.4% (616/806)	88.1% (710/806)
s0802	80.9% (637/787)	86.4% (680/787)
s1202	81.1% (664/819)	87.7% (718/819)
s1402	86.2% (702/814)	87.7% (743/814)
s1502	90.9% (687/756)	92.1% (699/756)
s1702	88.3% (722/818)	91.4% (748/818)
s1802	83.7% (769/919)	92.2% (847/919)
s1902	93.8% (804/857)	95.8% (821/857)
s2002	83.5% (641/768)	89.8% (690/768)
s2102	82.9% (636/767)	93.5% (717/767)
s2202	81.7% (686/840)	87.5% (735/840)
s2302	84.9% (699/823)	91.7% (755/823)
s2402	86.8% (688/793)	93.8% (744/793)
s2502	86.8% (665/766)	95.7% (733/766)
s2602	93.7% (731/780)	95.3% (743/780)
s2702	90.0% (764/849)	92.8% (788/849)
s2802	90.7% (640/706)	92.2% (651/706)
s2902	86.9% (672/773)	89.1% (689/773)
s3102	84.9% (656/773)	91.9% (710/773)
s3202	86.2% (676/784)	93.8% (735/784)
s3302	78.6% (582/740)	83.8% (620/740)

training data but exist in testing data, of course that the trained model can not classify this kind of stage in the testing data. Thus, the accuracy of classification is bad. This situation was shown in Figure 5.3.

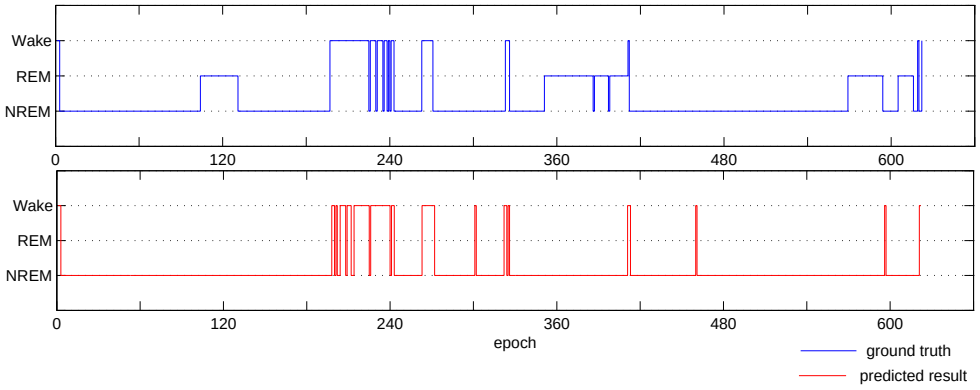


Figure 5.3: Example of record S004. Due to DAY1 recording is as training sets and lack of REM state, the trained model of S004 can not identify the REM state. The testing set is the DAY2 record of S004. The figure illustrate that the REM state are all misclassified.

Table 5.2: Confusion matrix of Subject S004.

		Manual scoring			
		Wake	**REM**	**NREM**	**Total**
Automated	**Wake**	48	0	10	58
scoring	**REM**	0	0	0	0
	NREM	9	121	434	564

5.2.2 Compared with general model and subject-specific model

In our result, we trained a subject-specific model which means to the same subject using one of his recording as training data and the other one for testing data. The accuracy of automated staging depends on the similarity of training data and testing data. Generally, sleep EEG of different people is different and thus we will train a specific model with his own training data. However, if there is a big difference of the sleep structure between the training data and testing data to the same subject, then we will get the bad predicted result;

by contrast, if the sleep structure between the training data and testing data is similar, even using a general model and test with recordings from different subjects can get the good predicted result. Table 5.3 show the comparisons between the result of testing with the subject-specific model and general model. The general model was trained with four recordings that are s1501, s1701, s1801, and s1901 and the testing data are s0101, s0301, s0401, s0501, and s0601. The accuracy of predicting by general can achieve 70% to 80% but for most cases the classification accuracy were worse than the result of predicting by subject-specific model.

Table 5.3: Comparisons between the results of testing with subject-specific model and general model. The general model was trained with four recordings that are s1501, s1701, s1801 and the testing data are s0101, s0301, s0401, s0501, and s0601.

Subject	**subject-specific model**	**general model**
s0101	84.7% (728/859)	70.2% (603/859)
s0301	88.6% (616/695)	80.3% (558/695)
s0401	77.7% (483/622)	77.7% (483/622)
s0501	91.4% (620/678)	74.9% (508/678)
s0601	76.4% (616/806)	81.1% (654/806)

Although other feature extraction method can achieve higher classification accuracy, however, computation speed is one of the important consideration for implementing the on-line system. Feature sets which has high computation complexity is not suitable for on-line analysis even it can get the better classification result. Our feature extraction method is suitable for the on-line system since the computation time is not too long. It takes about 1 second to score the 30-second epoch and send the alarm to the subject and therefore the delay time is about 1 second.

5.3 Effect on REM-SD with the proposed system

Since the anomaly in depressed patient is that increased REM density, the objective of the therapy of sleep deprivation is to reduce those “redundant” periods of REM sleep rather

than deprived them all. Therefore, it is tolerable for some missing detection of REM states. However, the percentage of REM sleep to be deprived for good therapeutic effect requires further clinical studies. In our automated sleep staging system, the REM detection rate is 70% to 80%. Thus, not every REM state will be detected and send alarm to the subject. However, we need to reduce the false alarm as possible. As Figure 5.4 shown, we found that in the upper night it is more easier to disturb the subject with the alarm and harder to wake the subject up in the lower night compared to the upper night. The true reason is that in the upper night the subject is disturbed continuously and thus the SWS (deep sleep) are increased to "compensate" his sleep. Therefore, it is harder to disturb him by the same alarm level in the lower night. One possible way to solve this problem is to gradually increasing the volume of the alarm.

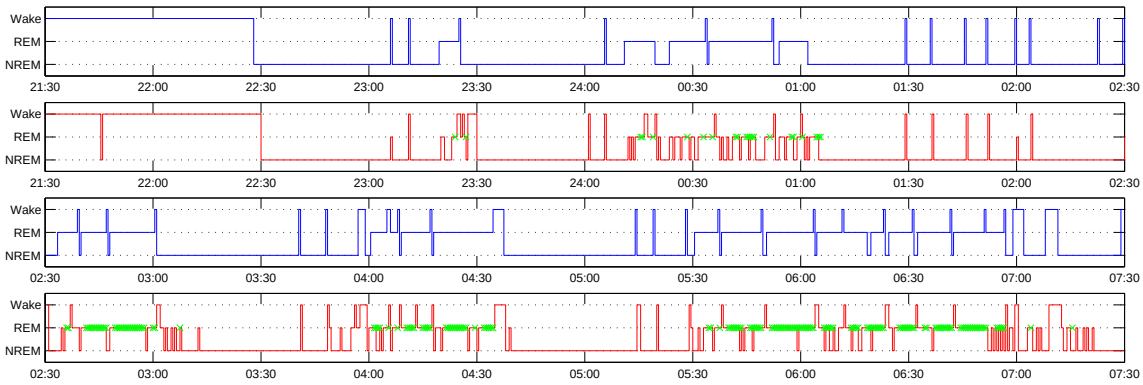


Figure 5.4: Hypnogram of S00702. The first and second field is the upper night sleep and the third and fourth field is the lower night sleep. The blue line represents the results of manual scoring which we assumed as the ground truth, and the red line is the results of automated scoring by our system. The green dot represent the alarm occurring.

Chapter 6

Conclusions and Future Works

6.1 Conclusions

For a long time, most researches of automated sleep staging focus on improving the classification accuracy with various feature extraction and classification methods. Sometimes the researcher may affirmed that their proposed method was feasible with high classification accuracy but it is hard to compare with each other since the different testing data is used for analysis. Though many automated systems have been developed but none are universally applicable. Therefore, rather than devoting to achieve the extremely high accuracy of the proposed system, we emphasize the practicality instead. We have proposed an automated sleep staging system using single EEG channel which can achieve accurate classification for three sleep stages, that are Wake, REM and NREM. With the proposed feature extraction method, our system can achieve precise scoring with fewer processing time. Due to its moderate computation speed for the on-line staging, we further extend this system to an on-line system which is distinctive that compares to other related works. In the experiment of on-line REM sleep deprivation, it is verified that this on-line system is feasible and reliable for assisting REM deprivation. The purpose of the on-line automated sleep staging system is to assist REM sleep deprivation for depressed patient to reduce their intensive REM sleep and consequently a few epochs of miss detection of REM state is tolerant.

6.2 Future works

Experimented on depression patient

So far we only applying the REM sleep deprivation on healthy people in order to test our proposed system. There is no need to discuss the effects of REM sleep deprivation on healthy subject. In the future, we expect to apply this on-line automated sleep staging system to depression patients for REM sleep deprivation in long-term. Thus, it may easier for neurologists to process a great quantity of experiments to see the effects after the REM sleep deprivation and verify the viewpoints which were proposed in the paper. Moreover, they can confirm whether this method is suitable for depression patients in clinical therapy.

For achieving the therapeutic effect, the therapy of REM sleep deprivation may continue for a period of time but the patient usually may not stay too long in the hospital. Thus, we hope they can easily use the system in home. This is the reason why we use only single EEG channel in our proposed system. It is more easier to implement the REM sleep deprivation to depression patient in long term and to confirm the therapy.

Materialization of the system

In further development, we expect to develop our system to a home depression cure machine. Base on the single channel EEG for sleep staging, it is more convenient for the patient to operate at home. Following the increased time of the therapy, the model of sleep scoring will update for every period of time. By this machine, depression patients can receive the long-term therapy in home.


E S A
1896

Bibliography

[1] Flexer Arthur, Gruber Georg, and Dorffner Georg. A reliable probabilistic sleep stager based on a single eeg signal. *Artificial Intelligence in Medicine*.

[2] Chang Chih-Chung and Lin Chih-Jen. Libsvm: a library for support vector machines. 2007.

[3] Chih-Chung Chang Chih-Wei, Hsu and Lin Chih-Jen. A practical guide to support vector classification. 2008.

[4] Berthomier Christian, Drouot Xavier, Stoica Maria, Herman, Berthomier Pierre, Prado Jacques, Bokar-Thire Djibril, Benoit Odile, Mattout Jèrèmie, and d'Ortho Marie-Pia. Automatic analysis of single-channel sleep eeg: Validation in healthy individuals. *SLEEP*, 30(11), 2007.

[5] F. Lopes da Silva. Computer-assisted eeg diagnosis: Pattern recognition techniques. *Electroencephalography, Basic Principles, Clinical Applications and Related Fields.*, 54, 1987.

[6] Riemann Dieter, Berger Mathias, and Voderholzer Ulrich. Sleep and depression - results from psychobiological studies: an overview. *Biological Psychology*, 57, 2001.

[7] W. Vogel Gerald, Buffenstein A., Minter K., and Hennessey Ann. Drug effects on rem sleep and on endogenous depression. *Neuroscience & Biobehavioral Reviews*, 14, 1990.

[8] W. Vogel Gerald, A. Thurmond, P. Gibbons, and K. Sloan. Rem sleep reduction effects on depression syndromes. *Arch Gen Psychiatry*, 32, 1975.

[9] Giedke Henner and Schwärzler Frank. Therapeutic use of sleep deprivation in depression. *Sleep Medicine Reviews*, 6(5), 2002.

[10] B. Hjorth. Time domain descriptors and their relations to a particular model for generation of eeg activity. 1975.

[11] INTERNATIONAL COMPUTER SCIENCE INSTITUTE. *Sleep Stage Classification using Wavelet Transform and Neural Network*, 1999.

[12] H. H. Jasper. Reports of the committee on methods of clinical examination in electroencephalography. *Electroencephalography and Clinical Neurophysiology*, 10, 1958.

[13] Cafferel Jennifer, John Gibson G., Phil Harrison J., J.Griffiths Clive, and J. Drinnan Michael. Comparison of manual sleep staging with automated neural network-based analysis in clinical practice. *International Federation for Medical and Biological Engineering*, 44, 2006.

[14] Fell Jürgen, Röschke Joachim, Mann Klaus, and Schäffner Cornelius. Discrimination of sleep stages: a comparison between spectral and nonlinear eeg measures. *Electroencephalography and Clinical Neurophysiology*, 98, 1996.

[15] Jobert Marc, Timer Christian, Oiseau Eric, and Schulz Harmut. Wavelets - a new tool in sleep biosignal analysis. *Journal of Sleep Research*, 3, 1994.

[16] W. Mahowald Mark and H. Schenck Carlos. Insights from studying human sleep disorders. *Nature*, 437(27), 2005.

[17] A. Carskadon Mary and C. Dement William. *Normal human sleep: an overview*. Philadelphia: Saunders, 1994.

[18] B. Russo Michael. Sleep stage scoring.

[19] Proceedings of the 23rd Annual EMBS International Conference. Istanbul, Turkey. *Automatic detection of sleep stages using the EEG*, 2001.

[20] Proceedings of the 26th Annual International Conference of the IEEE EMBS. San Francisco, CA, USA. *EEG feature extraction for classification of sleep stages*, 2004.

[21] Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference Intelligent Agents, Web Technologies and Internet Commerce IEEE computer society. *Automatic Sleep Staging using Support Vector Machines with Posterior Probability Estimates*, 2005.

[22] Proceedings of world academy of science, engineering and technology. *Harmonic parameters with HHT and Wavelet transformation for automatic sleep stages scoring*, 2007.

[23] L.R. Rabiner and B.H. Juang. An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1), 1986.

[24] Acharya U. Rajendra, Faust Oliver, Kannathal N., Chua TjiLeng, and Laxminarayan Swamy. Non-linear analysis of eeg signals at various sleep stages. *Computer Methods and Programs in Biomedicine*, 80(1):37–45, 2005.

[25] A. Rechtschaffen and A. Kales. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. 1968.

[26] Himanen Sari-Leena and Hasan Joel. Limitations of rechtschaffen and kales. *Sleep Medicine Reviews*, 4(2), 2000.

[27] Penzel Thomas and Conradt Regina. Computer based sleep recording and analysis. *Sleep Medicine Reviews*, 4(2), 2000.

[28] Gerald W. Vogel. Evidence for rem sleep deprivation as the mechanism of action of antidepressant drugs. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 7, 1983.