ECE 499

Prof. M. M. A. Salama

Contextual Analysis for Automatic Sleep Staging

Final Report

Jessie Y. Shen (ID 99271519)

Wednesday, April 28, 2004.

Summary

Review on automatic sleep staging reveals a new concept of incorporating contextual information to improve classification of ambiguous cases. Interpreting the context as human scorers do means that the computer must contain a comprehensive model of the sleep cycle.

Various models were considered from very simple to highly comprehensive. By analyzing their characteristics and their requirements, the simplest *one-cycle-duo-direction* and the most complex *duo-layer-multi-cycle* were analyzed further.

The simple model is adopted for immediate implementation and easy design of contextual analysis applications. The simple model proves that even its simplicity does not hinder its ability to improve certain stage differentiation.

The modelling process for the comprehensive model is designed to use genetic search algorithms. Various design considerations arise, and more analysis of these considerations is required. The algorithm is demonstrated to function even on a limited data set; however, for a representative model, a better collection of sleep data is required. One particularly important application of this model to contextual analysis is its ability to track the operation of particular stage detectors.

This project explored the area of contextual analysis for sleep staging, but many details remains to be studied.

Contents

List of Figures

List of Tables

1 Introduction

Sleep disorders affect the patients' quality of life in terms of education, career, personal interaction, etc. Appropriate treatment can only follow a correct diagnosis based on an overnight sleep study. The study collects various physiological signals to indicate to the doctor the type and quality of sleep experienced by the patient. The process of assigning the sleep type to a segment of PSG values, which is called sleep staging or scoring, is repetitive and time consuming. Therefore, the advantages of its automation are well recognized by both the medical and the engineering communities.

The last few decades have seen various algorithms aimed to automate sleep staging. The general approach applies signal processing and classification algorithms to capture PSG's visual information. While these algorithms have demonstrated encouraging performance, they are yet to be accurate enough to replace human scorers, because human scorers also consider contextual knowledge. Aside from performance improvements, the algorithms also need to be intuitively understandable by the medical professionals. Therefore, the approach should be modified to mimic closely human scorer's actions in the attempt to improve performance and encourage industrial adoption. Figure 1 shows the revised approach based on the sleep staging guidelines[1].

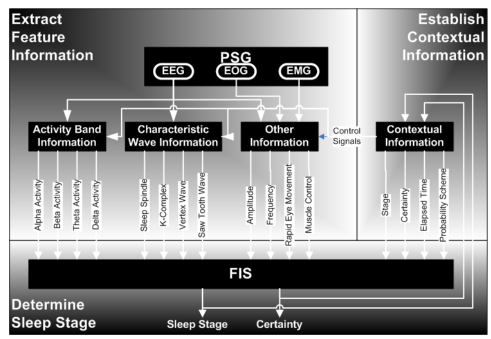


Figure 1: The revised layout of the sleep staging algorithm.

This report reviews the development of sleep staging hardware, methodology, and standards. The research for the automation of sleep staging is summarized with emphasis on contextual analysis. After showing that contextual analysis requires an accurate sleep cycle model, the rest of the report attempts to develop such a model. Finally, the model is applied to augment previous sleep staging algorithms with contextual information.

2 Sleep Staging

The evolution of sleep staging determines a great deal its current status. At first, the sleep staging primarily developed as relevant hardware was invented. It was followed by progressions in methodology, arriving at a generally accepted standard by Rechtschaffen and Kales, 1968. Since then, the focus turned to automating the process.

2.1 Hardware[1]

The earliest EEG machine was built by H. Berger, in 1928 [2]. His electrode-amplifier-recorder unit was physically limited to 10 Hz frequency resolution. Five years later, E. Adrian and B. Matthew adapted the oscilloscope for EEG recording such that 1000 Hz was achieved. For several years, W. G. Walter worked on topographical details by reduced the size of electrodes. By 1957, he designed a 22-electrode system strategically covering a human scalp. In 1960, A. Grass began experiments with computer interfacing. Recent years have seen equipment refinement, improved signal processing, shift to computer based analysis and storage.

2.2 Methodology

Sleep staging is not an exact science. Its practices and definitions have evolved as human learn more about the brain. The basic sleep stages were defined by 1930's. The last one associated with dreaming was identified in the 50's by E. Aserinsky. By 1968, Rechtschaffen and Kales compiled various publications to produced a sleep scoring manual [1], that remains the standard today.

This section first present some background on modern understanding of the human sleep cycle before presenting the actual sleep stage definition. It will emphasize the importance of

[1] A key PSG hardware is the EEG units; therefore, the discussion will focus on EEG equipment.

contextual information in the stage descriptions.

2.2.1 Human Sleep Cycle

Human sleep cycle is divided into multiple segments, where each demonstrates a set of characteristics associated with one of 6 stages. These stages are awake, NREM I to IV, and REM[2]. Their progression generally follow the sequence shown in Table 2.

NREM sleep provides the person with physiological rest and the brain's activities slow down. Since the overall electrical activity in the brain must remain constant, as the frequency slows down in NREM, the amplitude increases.

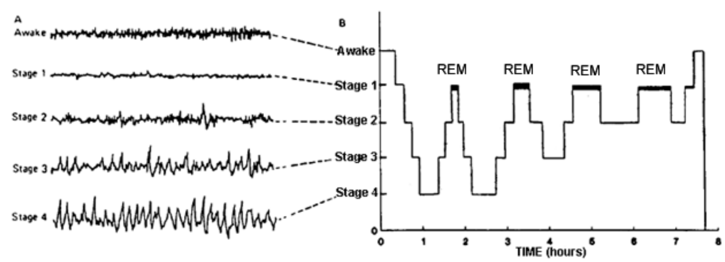


Figure 2: Sleep stages in a sleep cycle.

REM sleep provides the person with psychological rest by actively reorganize the brain to a better state. EEG during REM is similar to that of awake, with a reduction in amplitude. These two stages are best differentiated by EOG and EMG. EOG during REM shows rapid eye movement instead of the vision related movements associated with awake EOG. EMG under REM shows little or no activity, because without muscle tone, the physical actions appearing in the dreams cannot cause us actual harm.

Healthy humans with a regular night's sleep will follow these sleep stages in the particular pattern shown in Figure 2. Humans tend to enter into NREM sleep first. They progress deeper into successive stages, and then returns to lighter NREM sleep. The cycle of NREM to REM and back to NREM usually repeats every 90 to 100 minutes. As the night progresses, the portion of the cycle spent in NREM decreases and REM time increases. While this

[2] A 7th stage MT is generally added to represent the segments of time when the PSG signals are obscured by body movement

pattern may deviate from person to person, and from infancy to old age, the pattern repeats closely in the short term. This pattern translates to the contextual information that will become the focus of discussion later.

2.2.2 Sleep Stages Differentiation

Sleep staging attempts to identify the sleep stages discussed above based on the signals collected in PSG. Since the exact staging methods is covered in a 57-page manual, it is impossible to encapsulate the entire content. However, a summary of the manual is provided below.

Significant information are derived from EEG, EOG, and EMG. EEG[3] generally uses one of the two channels, C3 referenced against A2 or C4 referenced against A1. EEG reflects the brain activity. EOG is used to record eye movement. The electrodes are typically placed on both temples with a 1 cm offset from each other in lateral position. EMG measures the muscle activity on the chin. This position is often referred to as mental or submental EMG. PSG signals should be recorded with a minimum frequency of 100 Hz. It should be able to differentiate microvolts and to resolve maximum voltage of 1000 microvolts.

There are 8 EEG features(α-, β-, θ-, δ-activities, sleep spindles, k-complexes, vertex waves, sawtooth waves), 3 EOG features (rapid eye movement, slow rolling eye movement, and vision related eye movement), and 1 EMG feature (muscle tone). These 12 features compose the language with which to define the sleep stages.

While the actual sleep staging rules developed by Rechtschaffen and Kales are quite complex, the simple version listed below will cover most cases. However, these simple rules lose much of the contextual information used for staging in particular cases.

[3]The positioning of EEG electrodes follow the 10-20 system.

Stage Awake:	Mixed $\alpha, \beta, \theta, \delta$ activities.
Stage I:	α dominance exceeds 50% of time, and no ς and no κ.
Stage II:	α dominance less than 50% of time, and θ dominance the rest of the time, and ς and κ may be present.
Stage III:	δ dominance between 20% and 50%
Stage IV:	δ dominance greater than 50%
Stage REM:	Mixed $\alpha, \beta, \theta, \delta$ activities (lower amplitude than Awake), and ς and κ may be present.
Stage MT:	Very high amplitude, high frequency, but particular features.

2.2.3 Importance of Contextual Analysis

While the simple rules provided in the previous section covers many of the sleep staging cases, there still remain a portion that would be ambiguous or nonintuitive based only on the characteristic features. In particular, these cases occur around stage transition, where the characteristic features are further from nominal and require close observation.

In the manual, significant portions of the description focuses on these ambiguous cases and the contextual information that can correctly classify them. These descriptions can take one of three forms,

- *Context-based Description*: These descriptions single out a stage's particularities during some period. For instance, REM generally exhibits mixed frequency activity in EEG, rapid eye movement in EOG, and near zero EMG. However, at the beginning and the end of REM stage, these three characteristics may not start to exhibit at the same time. Therefore, special discussion is provided for the start and the end of REM stage. The first such rule, Rule REM-1A[4] says that epochs should be classified as REM if EEG and EOG show the correct characteristics and EMG does not. The condition being that EMG remains approximately constant before and long after the transition. However, the next rule, Rule REM-1B, says that if the EMG drops shortly after EEG

[4]Rule REM-1A is on the Page 9 of the manual[1].

and EOG fell into REM patterns, then only epochs after EMG has dropped count as REM.

- *Probability of Transitions* The transition likelihood is vague but present in the manual. For example, Stage I normally only lasts for 1 to 7 minutes during nocturnal sleep. In that case, if the characteristics are confusing, then scoring Stage I beyond 7 minutes would have very low probability.

- *Fix-it Rules* Fix-it rules are the most tedious in the Sleep Staging manual. Since the rules are not meant to connect small gaps in almost continuous classification, they are very counter-intuitive and difficult to grasp by inexperienced scorers. One example is the treatment of sleep spindles that occur during REM. When three consecutive epochs are REM, but one sleep spindle occurred in the middle epoch, it is still classified as REM.

This contextual information reduces ambiguity such that it reduces inter-observer error. However, this information can only be mastered through experience.

2.3 Sleep Staging Automation

This section first looks at some of the past publications on automation of sleep staging. A brief description of each article is provided in terms of its author, year, objective, method, and performance. Then, this section turns to literature discussing contextual analysis in automated sleep staging.

2.3.1 Past Publications

The complex nature of human sleep mechanism generated a large archive of articles dealing with various aspects of sleep staging. A selection of the most recent publications are listed below,

- The first group deals with direct sleep stage identification.
  - In 1998, Shimada[5] published a paper on Sleep EEG Recognition neural network that combines output from a Sleep Stage Diagnosis neural network and a Contextual Diagnosis neural network. In the report he claims that the performance of the improved unit is more consistently at 80%, whereas the previous performance may fall to 65%.
  - In 1998, Pacheco[6] generate the stage from 3 EEG channels, 2 EOG channels, and 2 EMG channels. He first extracts the features, finds the independent classification, and finally uses the context to validate the result.
  - In 1999, Oropesa[7] presented an automatic sleep stager that uses wavelet analysis followed by a neural network. His system reached high 90%'s for retrospective accuracy and 77.6% for prospective accuracy.
  - In 2000, Flexer[8] presented a continuous sleep stager that uses Hidden Markov Models (HMM). The system at that point had a performance of approximately 80% for wake, deep sleep stages, and only 26% for REM stage. In 2002, he [9] published that the REM detection rate has been improved to 68%.
  - In 2001, Van Hese[10] uses k-mean clustering with the Hjorth parameters, Harmonic parameters, and ratio band energy. The publication did not include effectiveness of the unit.
- The second group attempts to detect some EEG characteristics, such as sleep spindles, K-complex waves, etc. Their summary is in Table 1

While many studies were carried out to automate sleep staging, no standard algorithm has been accepted by the medical community. Many of the studies presented above have excellent results, but the remaining performance gap limits their usefulness and their acceptance. In

Table 1: Research for EEG characteristics detection.

Researcher	Year	Characteristic	Method	Performance
Bankman[11]	1992	K-Complex	ANN	Sensitivity 90% FPV 8%
Henry[12]	1994	K-Complex	Wavelet	N/A
Pohl[13]	1995	K-Complex	Neuro-Fuzzy	Accuracy 96%
Akin[14]	1998	SS	Wavelet	N/A
Gorur[15]	2002	SS	ANN	Accuracy 88.7%
Huupponen[16]	2001	α-Activity	FIS	TPV 85% FPV 13%
Shimada[17]	2000	α-Activity	MLP	Accuracy 80% to 90%

Section 2.3.3, the issues with these previous research is examined and some possible solutions are proposed.

2.3.2 Research on Contextual Analysis

While the EEG technicians have always placed significant weight on the contextual information during sleep staging, the engineers have only recently began to examine this wealth of information. In particular, Shimada have published some results that included contextual analysis.

In Shimada's article in 1998[5], he described a three part artificial intelligent sleep stager. Each part consisted of a multi-layer perceptron neural network (MLP) trained by backpropagation. The first neural network, called Sleep EEG Recognition Neural Network (SRNN), detects characteristic waves. The second one translates the characteristic waves into stages, and it is called Sleep Stage Diagnosis Neural Network (SSNN). These steps match actions by human scorers. The final network forces the contextual agreement, so it's called Contextual Diagnosis Neural Network (CDNN).

The CDNN adjusts the current outputs from SRNN by bringing contextual agreement with the last 2 to 14 segments[5]. CDNN has 2 hidden layers of 10 neurons each. Its output is also a set of probabilities for the five stages. Shimada applies a competitive post-filter

[5] It is unclear whether the last 2 to 14 segments' data is from SRNN, CDNN, or final decision.

and selects the stage with the highest adjusted probability. Shimada's new stager generally outperformed the conventional stager by less than 10% with maximum accuracy 82%.

In a later publication[17], Shimada attempted to bring in contextual information using internal mechanisms in neural network. He experimented with time-delay networks and all-connected networks, both of which would consider the contextual information without the significant overhead required by his 1998 system.

TDNN allows the engineers to feed outputs from an earlier time back into the neural network as inputs. Therefore, the new decision automatically takes into consideration the new input values and previous output values. ACNN starts in some steady-state, and new inputs will cause its output to change. Since these outputs are chained with the inputs, these connected values keep changing until a new steady-state is reached. Both methods lack control over intermediate steps and add overhead reducing performance.

While the performance achieved by Shimada lingered at the 80% mark, he did show the value of contextual information. His studies showed that context must be considered in a strictly controlled manner. Without proper supervision, bad context can confuse correct stage decision.

2.3.3 Issues and New Trends

The shortcoming of the attempts described in the previous section arises from the engineers failure to understand the nature of the sleep staging activity.

Figure 3: Gradual transition from Awake to Stage I during the middle of an epoch.

EEG, as a representation of the human consciousness, is characterized by continuous and gradual changes. However, the sleep stages are discrete classifications defined by neurophysiologists according to their comprehension of the human sleep patterns. Therefore, the match between the artificial stages and the natural states encoded in the signal segments is quite poor.

Some notable problems are:

1. The standard epoch size is 30 seconds, but the stages often changes in between. Figure 3 demonstrates this occurrence for an Awake to Stage I transition.

2. Amplitude and frequency changes that characterize the shift to another stage may be gradual. Figure 3 also shows these gradual changes as the stages change.

3. Characteristic features, such as SS, K-complexes, are only present some of the time.

4. Sleep staging rules, in particular the contextual ones, are difficult to translate into the rules accepted by the computer.

5. The output is compared against the stages classified by trained medical experts, who have an inter-observer agreement of less than 90%[18].

In the case of manual sleep scoring, the first three problems are resolved by many "fix" rules, which take an epoch's context to solve the ambiguity. The perplexity of these rules complicates the translation of the rule set. They are not intuitive to artificial intelligence (AI) algorithms, which means most agents will have difficulties learning these rules.

Since Shimada's research is the only stager that considers context, and its primary problem is an over-reliance on neural network. While neural networks are good for certain types of image recognition, the difficulty to understand the internal workings of neural networks reduces its acceptability by medical professionals. Another reason that neural network may not work is that it has difficulty grasping severely disjointed boundaries, which would be the case with the fix-it rules.

Alternatives such as rule based expert system or fuzzy inference system may prove more welcome with the medical experts. These systems give the engineers a lot more control to build in the three types of contextual information included in the sleep scoring manual.

3 Contextual Analysis and Sleep Cycle Modelling

Section 2.3.2 introduced by example of Shimada's research that contextual analysis can improve sleep stager performances. Section 2.3.3 discussed how context clarifies otherwise ambiguous classifications. This section first provides reasoning that successful contextual analysis requires an accurate sleep cycle model, then shows one set of steps to obtaining such a model.

3.1 Importance of Sleep Cycle Modelling

Section 2.2.1 already introduced the human sleep cycle. Based on the simple graphs shown in Figure 2, it should be relatively easy to develop a model that mimics this cycle. The factors that needs to be set are the estimated durations in each state and the variance thereof.

However, reality does not follow the simplified model that physicians present. For instance, the five graphs in Figure 4 shows the normal sleep patterns from subjects that had no sleep disorders. It is interesting to know that not only do they not resemble each other, many also do not look like the simplified medical model.

Of the subjects, Subject sc4002e0[6] and Normal III[7] showed the most standard cycle. Even so, the graph has difference from the example provided in Figure 2. From Normal I, it can be observed that a person can fell asleep and wake up for a long time before falling back to sleep. Normal II shows that the duration of REM does not always increase over night. Normal IV shows that Stage IV does not always disappear near the morning.

The above observations introduce some complexity to modelling the sleep cycle. It means that each stage's duration experiences significant variance. Also, the stages do not always transition smoothly, in that they may return to the previous stage before moving forward.

[6]This set of data is taken from the Sleep-EDF Database on `http://www.physionet.org/physiobank/database/sleep-edf/`.

[7]The Normal No. sets of data are provided by Dr. C. George of Western Ontario University.

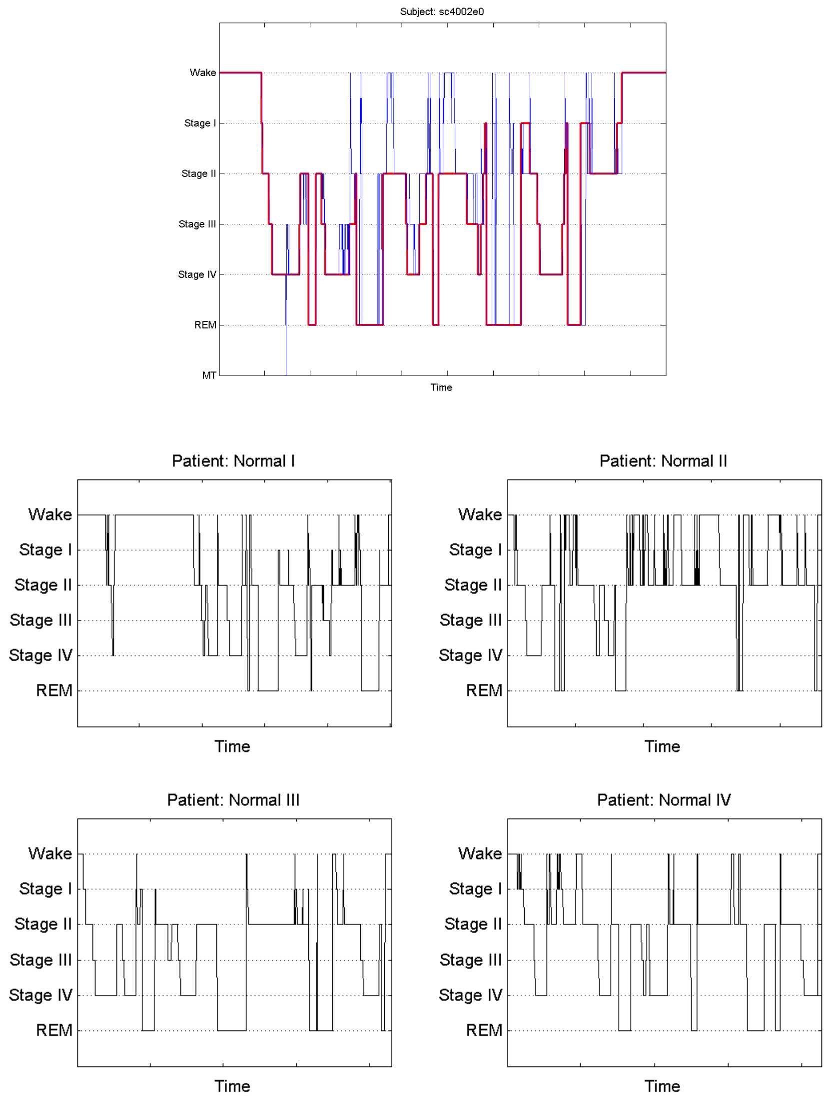


Figure 4: Five subjects with normal but different sleep patterns.

The shorter stages may be skipped, so the transition from state to state is not always clear. They may need to be associated with a time-referenced transition probability. In order to deal with these issues, an appropriate model must be designed.

4 Model Design

Several sleep cycle models were considered in this project. In general, as the models are improved in terms of better simulating reality, its complexity also increases. Rise in complexity translates to higher training data requirements, which is an issue discussed for each model.

4.1 One-Cycle-Duo-Direction Model

The first model, shown in Figure 5, is the basic model that is most commonly adopted. The patient starts Awake and proceeds through the NREM stages. Only Stage I and Stage II can precede to REM. All stages, except for Awake, can move into and leave from MT. This model is sufficient to generate the sleep stage transition graphs shown in Figure 2, but will not definitively show that pattern.

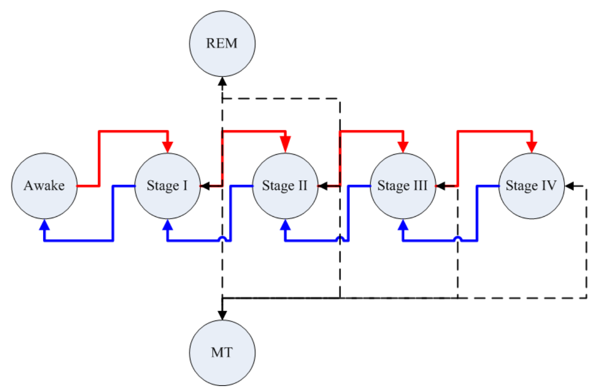


Figure 5: One-cycle-Duo-direction model.

4.1.1 Assumption

The model can handle skipped states. It is fairly common to see transient states such as Stage I and Stage III to be so brief that they are not scored at all. In which case, this model assumes that it can pass right through the transient states.

Occasional waking up during the night are ignored. Patients may wake up for various reasons for short bits of time. The patient generally return directly to the stage they were in. For example, a patient woken from Stage IV will relapse to Stage IV right after instead of starting in Stage I again. Due to the short duration and minimal interruption in the overall cycle, these segments can be ignored.

Transitions from a state into MT always returns to the first state. Subjects may move during any stage of their sleep but it generally doesn't bring about a change in states. Therefore, in actuality, a separate MT state should be observed for each state, except Awake, but in this model, they are lumped together. It may be worthwhile to count the probability of MT occurring under each state as well.

4.1.2 Shortcoming

This model does not track the direction of the NREM stage transitions. For instance if the patient is going through AWAKE→Stage I→Stage II, the next stage is most likely Stage III. However, if the sequence is Stage IV→Stage III→Stage II, the chance that Stage III occurs next is very low. However, this model cannot differentiate the difference between the two sequences because it does not record the direction. This issue also applies to entries into REM. The sequence Stage II→Stage I→REM is far more likely than Awake→Stage I→REM. Therefore, the model should be adjusted to look at the directions.

4.1.3 Data Requirement

An advantage of this model is its low requirement for data. Firstly, this model has very few parameters to set. Keeping to the assumptions listed above, there are a total of 22 transitions to parametize. (Awake and NREM states have a total of 8 one-direction transitions. There are 2 two-direction transitions with REM, and 5 two-direction transitions with MT.) In these 22 transitions, the 5 transitions from MT back to the source state should be considered as 1 based on the assumptions. Therefore, only 17 transitions needs to be found.

Each sleep study can expect to see 4 to 6 transitions between Awake and NREM stages. There are more transitions into REM and MT. These transitions will provide a good average.

The second reason for low data requirement is that the amount of ambiguity in the model surpasses errors from parameters. For instance, the transitions, whose probability change dramatically depending on the direction, would not be accurate regardless the amount of data provided. These transitions would observe one direction's probability as 90%+ and another direction's as 10%-. Since this model lumps them together, the data would suggest a 50% probability, which is not at all indicative of the actual two values. Therefore, the errors would come from the model as opposed to the parameters.

4.1.4 Implementation

This model is generally implemented as a discrete time Markov chain with 30-second epoch. The transition probability matrix is derived by counting the number of each type of transition leaving a certain state and calculating the ratio of each type to the total, which is the probability of that type of transition. Though the one-cycle-duo-direction model only considers a subset of the transitions represented by the transition probability matrix, the transitions outside of the model will almost equal zero.

Table 2: Sleep stage transition probabilities based on all subjects.

States	Awake	NREM I	NREM II	NREM III	NREM IV	REM
Awake	0.818	0.118	0.046	0.001	0.001	0.017
NREM I	0.074	0.491	0.421	0.000	0.000	0.014
NREM II	0.026	0.017	0.934	0.013	0.000	0.009
NREM III	0.003	0.000	0.171	0.594	0.148	0.000
NREM IV	0.013	0.003	0.024	0.007	0.786	0.001
REM	0.029	0.003	0.009	0.000	0.000	0.918

Table 2 contains the transition probability matrix extracted from the data sets provided by Dr. C. George of Western Ontario University. Figure 6 demonstrates the same example graphically. In this diagram, the transition arrows are weighted according to their significance relative to all transitions.

Despite the simplicity of this model, observations can be made to better understand the sleep cycle. Each state shows the highest probability of remaining in itself. As stated in the assumptions, NREM I and NREM III are transient states that are either short or skipped. NREM I's short duration is reflected by a lower probability of remaining in itself. The transition probability from Awake directly to NREM II indicates that NREM I may be skipped. Similar observations can be made about NREM III.

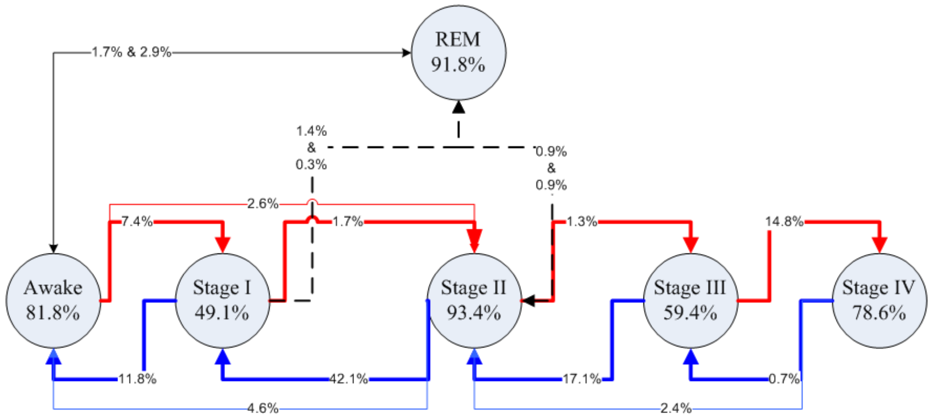


Figure 6: State transitions probabilities calculated by ratio of each type of transitions.

4.2 One-Cycle-One-Direction Model

The model in Figure 7 corrects the two direction issue in the previous section. It uses red to indicate the transition direction is towards deeper sleep and blue to indicate the opposite. Therefore, it differentiates the probability of advancing into the next sequential state from the probability of doubling back to the previous state. For instance, the probability $P(II \rightarrow I)$ in the previous model would be broken down to $P_{red}(II\ \mathit{relapsing\ into}\ I)$ and $P_{blue}(II \rightarrow I)$.

4.2.1 Shortcoming

This model does not look at the amount of time elapsed during the night. In this respect, subjects generally pass through this cycles 4 to 6 times. As the cycles progress, the configuration of the cycles change moderately. However, this model fails to look at these issues and possibly introduce high levels of error by lumping various cycles together.

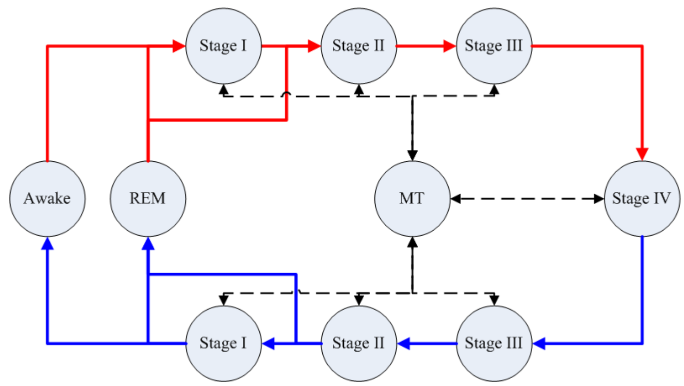


Figure 7: One cycle model where transition follows one direction.

4.2.2 Data Requirement

The data requirement for this model is relatively similar to the previous model. It has the same number of parameters to set. The only difference being that the training data must first be examined to identify the direction in which the transitions are destined. This process would be highly manual and time-consuming. Therefore, an implementation will not be provided for this model.

4.3 Four-Cycle and Six-Cycle Model

The one-cycle-one-direction model addressed the issue of useful direction information missing from one-cycle-duo-direction model. However, it fails to record the cycle in the context of the entire night's sleep. The four-cycle and six-cycle model aims to address that issue. Since it is well recognized that at least four cycles are experienced by most healthy subjects, this model looks at the first four cycles. Figure 8 shows the transition diagram. The six-cycle model just goes through two more cycles before returning to awake.

Despite that the diagram does not show the MT state, it is still considered in the model. The MT state is merely not shown graphically due to complications in display. Again, it should be noted that there should be an MT state for each non-Awake states. The reason for individual MT states is that following state i→ MT, returning to state i is far higher than going into another state.

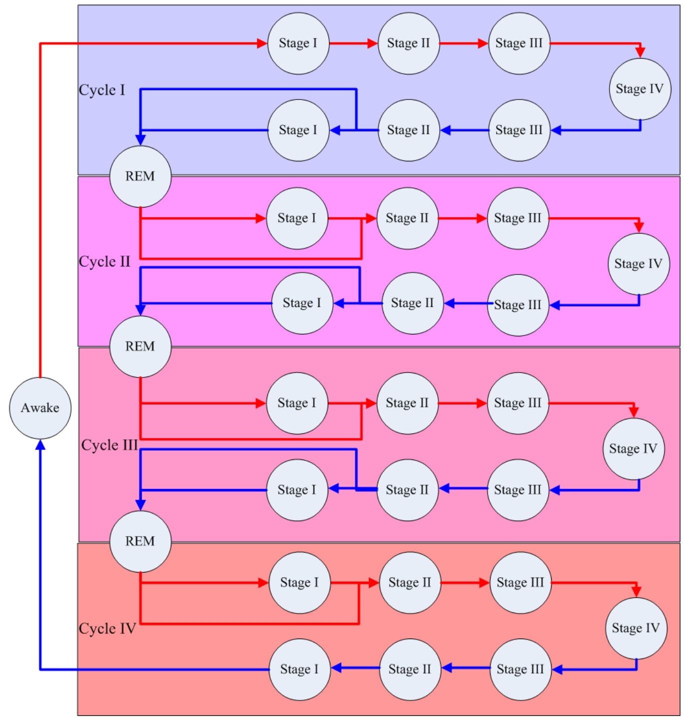


Figure 8: Four cycle model tracking both directions.

4.3.1 Shortcoming

The four-cycle model ignores the last two potential cycles. While it is difficult to obtain consistent patterns from the last two potential cycles, their frequent presence means it is important to include them in the model. However, a six-cycle model means that the last two cycles must be considered whether they occur. This issue can be dealt with by allowing optional cycles, but then the model would be more complex to construct.

These two models are very cumbersome. In order to make the model manageable, only the transitions vital to the cycle would be determined. Unlike the earlier models, such a complex model cannot have each state connecting to each other state, where many connections would never occur. For instance, there would not be a cycle 1 state 4 transition into cycle 3 state 3. However, the design choice, to only study the standard transitions, would mean that events such as skipped states, or doubling back to a previous, or waking intrusions, etc., are all being ignored. Therefore, some thought should be put into the model such that most possible and non-zero transitions are included in the model.

4.3.2 Data Requirement

The data requirement for these models is significantly higher. The number of parameters is approximately four and six times that of the one-cycle-duo-direction model. Since each sleep study can only provide one set of information, the quantity requirement jumps by a multitude. In terms of quality, the data sets must contain at least 4 continuous and relatively standard cycles.

Also the choice of data must look at the tradeoff between good sample space of training data and similarity to the base model. A large and representative collection of sample sleep studies will make the model more generalized. Their average will be a good starting point. However, if the samples are vastly different from the base model, then the variance exceeds the usefulness of the average value. Therefore, it is important to choose a representative yet coherent set of training data.

For the six-cycle model, further difficulty arises due to the inconsistent occurrence of the last two cycles. These cycles are also likely to be marred by multiple waking and dreaming.

4.4 Multi-Cycle-Connected Model

The last two models fail to contain the minor connections between states. For instance, state 1 is often skipped before entering REM, which means a direct transition from stage 2 to REM would be reasonable. As discussed previously, a fully connected model would be overly complicated and containing significant useless information. Therefore, this model aims to connect only the possible connections.

This model can be adapted to 4 or 6 cycles. It will include all the connection to Awake and to MT. It should assume possibilities of skipped states, of states that double back, etc.

4.4.1 Shortcoming

Many of the transitions added in this model are rare and they do not contribute significantly to the contextual information. For instance, consider a patient in NREM II for several minutes, shifts to I for one epoch, and returns to II immediately afterwards. The detection of that NREM I is not particularly vital. Therefore, to model the relatively insignificant transitions as equals to the main transitions would be placing undue emphasis on the added connections. Furthermore, due to their rarity, their transition probabilities will be too small to notice in the whole picture. One solution to this problem is to model these connections separately.

4.4.2 Data Requirement

It should be noted that the data requirement for this model is even higher. Since it not only models the basic transitions shown by physicians, it also models secondary transitions. It is clearly difficult to collect representative data sets to model the secondary transitions, which occurs in rare cases. Therefore, this model will require the significant preprocessing of the data sets. In particular, the secondary transitions must be identified to parametize separately.

4.5 Duo-Layer-Multi-Cycle Model

This model improves upon previous models by modelling main transitions separately from secondary transitions. The duo-layer approach is achieved by building a model with two levels of states, *cycle-level states* and *epoch-level states*.

4.5.1 Cycle-Level States

The cycle-level states demonstrate the overall behavior. It takes the form of the six-cycle model but acts like a continuous-time Markov chain. A plot of the transitions in cycle-level states would correspond to the Figure 2. The four-cycle equivalent transition diagram is in Figure 8[8].

Ideally, sleep data should be analyzed to identify the distributions that most suit each state. However, in this study, a normal distribution is assumed for the duration in each state. Therefore, each state must be associated with an average duration and a standard deviation. Only the transitions linked in the figure are assumed possible. In order to deal with transient states that are too short to score, these states can have duration of 0.

4.5.2 Epoch-Level States

From actual sleep studies, the plot in Figure 2 is not realistic. Figure 4 shows that aside from the main state transitions, many small spikes may occur in the behavior. These spikes may be short transitions into the previous states, next state, Awake, or MT. However, the cycle-level states will not be able to account for these spikes. Therefore, inside the cycle-level

Figure 9: 5-state structure.

[8]Note that the model being designed will assume it is not possible to skip NREM I between Awake and NREM II or REM and NREM II.

states, a 5-state structure is constructed to account for this kind of behavior. This structure is shown in Figure 9.

For these 5-state structures, the overall sleep study will still consider the subject as being in the *current state*. However, the subject may for short periods transition into the 4 transient states. The model contains a probability for such transitions and the expected duration in those states.

4.5.3 Shortcoming, Data Requirements, and Implementation

This model is designed to overcome the shortcoming of previous models. Despite encapsulating more information, its data requirements will be equivalent to that of the six-cycle model, because it can reuse the same data sets to extract the epoch-level states. Due to its comprehensive nature, its implementation is vastly more complex than previous models, and it will be discuss in Section 5.2.

5 Modelling Process

Based on the model in Section 4.5, this section designs the algorithm that parameterize the model. Before discussing the steps of the algorithm, first a quick look at the problem of modelling the duo-layer-multi-cycle model.

5.1 Problem Analysis

An algorithm that would develop the sleep cycle model described in the previous section must resolve several issues. In order to parameterize the average duration in the cycle-level states, the algorithm must overcome the lack uniformity between subjects. The algorithm also need to smooth individual curves such that individual durations can be extracted for variance calculation. The most significant obstacle is to select a set of data with sufficient quantity and adequate quality from which to build this model. First, take a closer look at the source of these problems and attempt to identify some solutions.

5.1.1 Lack of Uniformity

While the medical community presents a generic sleep cycle pattern, shown in Figure 2, high levels of inter-subject difference still exists. This fact is demonstrated in Figure 4. In fact, if the sleep cycle of the same subject on two different nights were observed, the cycles are still likely to be different.

Sleep Onset Variance

The first reason for the variance is that the onset of sleep may be different. One subject might have fallen asleep within 5 minutes of closing their eyes, while another patient can take 2 hours. This manifestation is related to the degree of tiredness the subject feels, whether the subject is excited or worried about the study and the new environment. There exist studies that determine the average wait before sleep onset. However, the number of uncertainty factors relevant to this issue is too high to model efficiently.

However, if the onset is different, the first transition between Awake and Stage I would be offset. In order to correct that issue, all data sets should preprocessed to start at the same epoch, meaning the first transitions occurring in a fixed epoch. The entire data segment would be shifted accordingly.

Fatigue Level

Physical and mental fatigue levels will determine partly the sleep cycle pattern. If subjects feel physically tired, they may have longer NREM IV periods or multiple occurrences of NREM IV. Emotionally burdened subjects may experience more REM periods. The idea behind these observation is that REM and NREM IV provide the best stages to get psychological and physiological rests, respectively.

There is no method by which to assess the degree of tiredness associated with the subject in each data set. Of course, a particular data set that contain tiredness information can be generated. For instance, a survey can be filled out by the subjects before going into the sleep study, or the subjects can be requested to perform certain tasks associated with a particular type of fatigue. Clearly, this type of data sets would be expensive and time-consuming. Though in a real continuous sleep monitoring situation, the fatigue information would be available through recent history analysis.

In the case where no fatigue level can be ascertained, this factor will be considered as an uncertainty.

Subject Profile

The sleep cycle also vary depending on the patient's profile in terms of age and sleep disorders. It is well-documented that as subjects become older, their sleep cycles change.

The older subjects tend to get lighter sleep as their circadian rhythm control mechanisms weaken. More specifically they experience shorter NREM III-IV and longer REM in the first cycle. In some cases, medications taken by elderly patients may affect sleep. Also, sleep disorders, such as the condition sleep apnea and restless legs are more common in older subjects.

This type of profile information can be derived relative easily from the medical records of the sample data set subjects. However, it is difficult to get significant number of cases with a particular profile, meaning that deriving a pattern is not easy. Another issue is that this information may not be available when the automatic sleep staging algorithms are used. One way of circumventing that issue is to adjust all data sets into a standard mold. However, such an adjustment function would not be easy to generate. Therefore, these factors must be considered again as a source of uncertainty.

Interpersonal Differences

Finally, even when two subjects with similar profile put under the same physiological stress, their sleep cycle still will not coincide with each other exactly. There will simply always be some interpersonal differences. There is no way to account for them. However, these differences should be minor enough that a prominent pattern can still be derived despite the differences. In the general situation, however, these interpersonal differences will be so small that they would not significantly contribute to the standard deviation.

5.1.2 Spikes in Data

Most scored sleep studies will contain spikes in data. These spikes occur for reasons including issues in sleep scoring, data selection, and natural causes.

Sleep Scoring

Sleep scoring attempts to impose a discrete set of states on a continuously and gradually changing biological process. This set of states are not proven to be perfectly correct, though they are widely accepted. In order to avoid ambiguity, the rules through which these states are meant to be scored are not always intuitive. Furthermore, the scoring accuracy depends on the scorer's experience. Of course being a tedious task, mostly technicians with little experience are given this task. These technicians may not understand the biological process or the nature of the rules such that they would score with error or inconsistency. Under such circumstances, spikes may occur either because the scoring rules failed to provide smooth transitions in the scoring, or because the scorer did not classify the borderline cases properly.

This issue is inherent to the sleep scoring problem, and it will exist both in the manual or the automated versions. As the goal of the automated version is to mimic the manual version, modelling with this issue is not a particularly important issue.

Data Selection

This issue is associated with the previous problem. Without selecting scores that were provided by an experience authority, the scores may contain inconsistencies or even errors. Also, without selection some of the particularly spiky data sets might be included as source data. This problem can be resolved by filtering the source data sets. In this process, scores that appear particularly noisy or inconsistent should be eliminated. Particular attention should be extended to data sets from experienced sleep scoring experts. Furthermore, it is worthwhile to select data sets that exhibit a range of behavior. One must be careful not to involve a data set that is so unique that the model is too biased from the standard.

Natural Causes

Aside from the external factors to cause spikes in sleep cycle, nature does play an important role as well. For instance, the subject may be disturbed at night by noise or movement causing the subject to transition from REM to Awake. When the stimulus is removed, the subject will most likely transition back. These will also appear as spikes. Similar explanation can be provided to transitions into movement time, into previous, and logical next states.

Since the model is supposed to capture this biological process, the manifestation of natural causes should not be ignored. The 5-state structure was designed to capture this kind of spiky behavior. Refer to the previous section for details on this structure and the specifics of how this structure can model these natural spikes. However, steps to prevent data sets that are artificially spiky should be taken in order to minimize the bias towards this behavior.

5.1.3 Data Set Shortage

The last problem facing modelling the sleep cycle is that there is a shortage of data sets. Each sleep study will only result in one set of data. Therefore, if getting a good statistical average requires some hundreds of samples, significant resources would have to be devoted to the data collection and data processing. Currently only two dozen data sets are available for construction and for validation. Therefore, this data shortage will present significant issues.

One issue is that there will be a bias to one or two scorers. While the eventual model may simply reflect the style of the participating scorers, it may be a problem if they are not particularly accurate. Another issue is that if the two scorers have contrasting methods, converging on a reasonable model may be difficult. The resulting model may also be inaccurate since there is no way to measure the goodness of the approximate averaging effect.

The serious nature of data set shortage means that this report will not be able to present a reasonably viable model. The algorithm will be designed and tested to deal with as many of the above issues as possible. It will be tested with the limited resources available to indicate the algorithm can be expanded to develop a real model with the right data sets.

5.2 Algorithm Design

The modelling process will identify the parameters of cycle-level and epoch-level states. The top view of the modelling process is Part A in Figure 10. The figure shows that there are 3 processes stemming from the same source data. Each process is aimed to identify a subset of the parameters. The first and second processes extract the average and variance in the cycle-level states duration, respectively. The third process extracts the epoch-level parameters. The output of these processes form the final model.

The most important step in this algorithm are the *average* procedure at the beginning of the first process. Its output feeds into all the processes. Given the lack of uniformity in data, a generic averaging algorithm would cause the sleep cycle pattern to be flattened, as seen in Figure 11. The resulting curve would not fit into the discrete sleep stages or the six-cycle model.

Since the average curve is meant to have the minimum distance to each of the samples, then a search algorithm can approximately find the optimal curve. This problem can be solved using genetic algorithms. Its flowchart is Part B in Figure 10. Using this method, the same function can be used by the second and third processes to smooth out each data set.

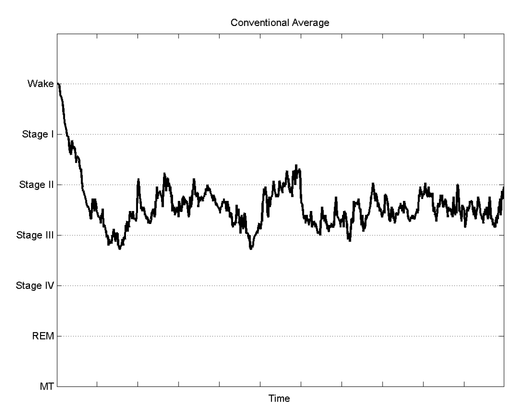


Figure 11: Conventional average algorithm flattens sleep cycle.

Search algorithm using genetic algorithms requires the following points to be addressed: the definition of the individuals, the initial population, the health measurement, the termination condition, and the birth process.

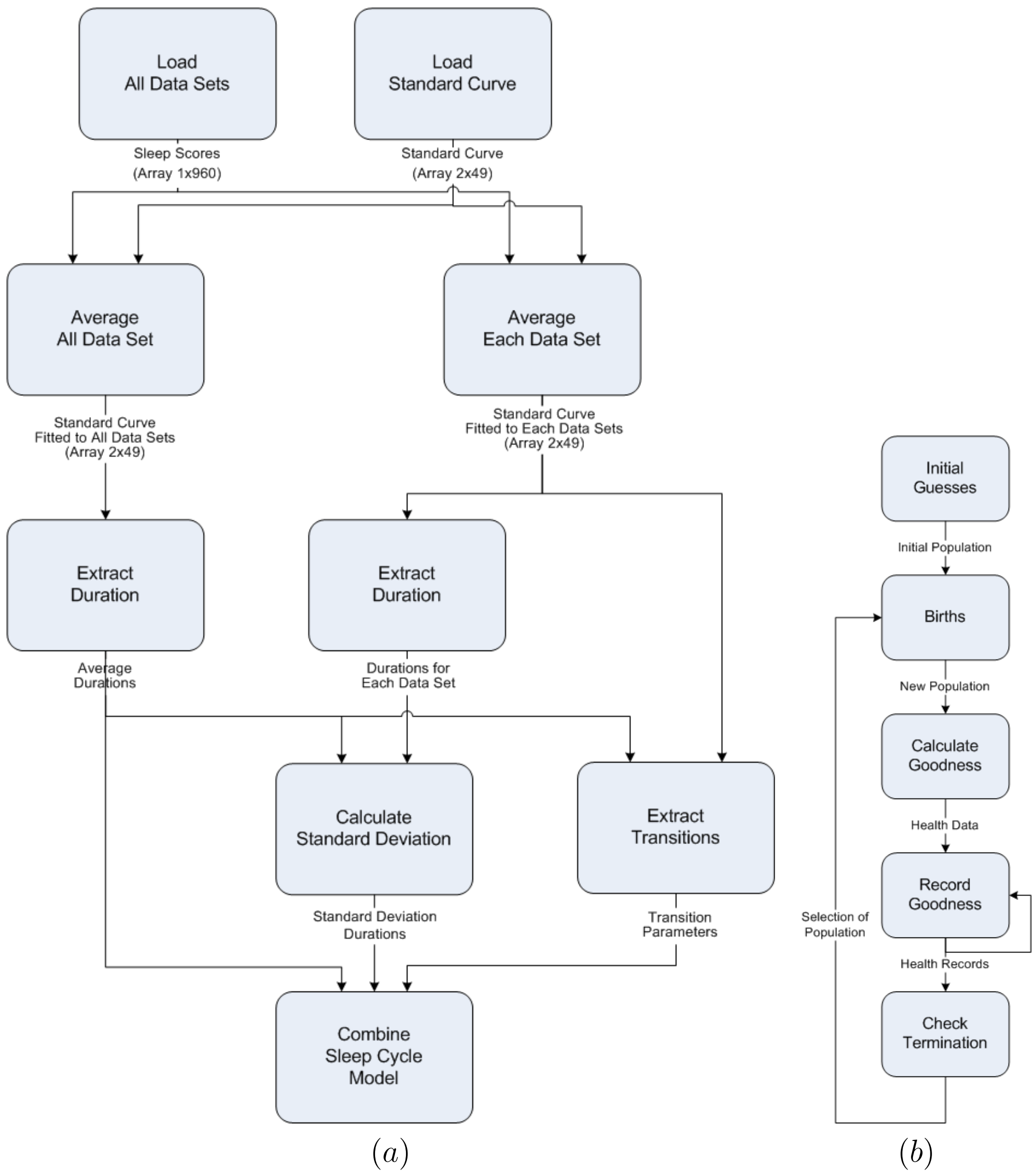


Figure 10: Flowchart (a) of overall modelling process. (b) of average procedure.

5.2.1 Definition of Individuals

The individuals in this context represents each possible solution. These solutions are assumed all viable but with varying fitness levels. In this context, these individual's fitness level means their distance to the data sets in general. The structure of the individuals determine the method by which the fitness can be found.

Epoch-Stage-based Definition

This definition identifies each individual based on an array of $2epoch/minute \times 60min/hour \times 8hours/study = 960$ elements, where each element indicates the stage scored. Figure 12 shows a small segment of this array. One advantage of this model is that the data sets are already stored in this format. The health can be easily defined as the minimum distance between each corresponding point-pairs from an individual and the reference data set. The disadvantage is that no obvious birthing processes can be applied. Since the individual must retain the six-cycle model, the design for good mutation and crossover algorithms is seriously limited.

Figure 12: Structure 1 to represent individuals.

Duration-based Definition

A different structure is to keep an array of duration spent in each states. The six-cycle model has a 49-element array. A sample structure is shown in Table 3. This structure simplifies the design of mutation and crossover algorithms, and those algorithms will be discussed in Sec-

Table 3: Duration table for the standard curve.

Stage	0	1	2	3	4	3	2	1	5
Cycle I	2	2	10	2	40	2	10	2	20
Cycle II		2	10	10	20	10	10	2	26
Cycle III		2	15	12	6	12	15	2	26
Cycle IV		2	18	5	0	5	18	2	40
Cycle V		2	5	0	0	0	5	2	46
Cycle VI		2	5	0	0	0	5	2	46

tion 5.2.5. Calculating the health directly from this structure is more complex than the epoch-stage-based definition. However, this form can easily be converted to the first structure.

5.2.2 Initial Population

There are three options in the design of the initial population.

1. The first option takes the sleep cycle curves provided by various medical sources as the base population. It is beneficial to start with the curve that approximates the ideal behavior, because it is already likely to be in excellent health. But these curves would be highly similar and they would not have enough variation to encourage new sample spaces. In this case, radical mutation or crossover can be attempted.

2. The second option is to randomly generate a population. The randomness in this step allows coverage by a significant sample space. However, it is possible that none of the future population will be close to the solution.

3. The final option is to visually analyze a subset of the data files, and record the pattern manually. Such an initial population starts closer to the data, but will migrate to converge all the data sets. Similarly, it may fall into the problem of being too close to some sets of data and never leaving the local minima in its area. Good design of mutation and crossover algorithms may overcome this issue.

5.2.3 Health Measure

Section 5.2.1 explained the basic method of determining the health by calculating the distance between an individual and a data set. While this method provides a basic measure, it is not most reflective of the health of an individual. The sleep stages are converted to discrete values on the computer, and this is a casual assignment. Therefore, the values are not reflective of the relative distance of one state to another.

For instance, a REM state may be classified as 5, but it is naturally next to NREM I, whose value is 1. So their distance should be 1 instead of 4. Furthermore, MT is classified as 6, but it is never more than 1 away from any non-awake states. Therefore, a weighting table that reflects the physiological distance is constructed as in Table 4. In practice, this table is normalized and de-referenced each time the distance between a data point and a curve point must be determined.

Table 4: Goodness weights.

Stages	0	1	2	3	4	5
0	0	1	1	1	1	1
1	1	0	1	1	2	1
2	1	1	0	1	2	1
3	1	2	1	0	1	2
4	1	2	1	1	0	3
5	1	1	1	2	3	0

The method of aggregating the distances of a data set to an individual is not determined. It is possible to use the average distance or the sum of all distances. The average method would give me a figure independent of the length of data set after lining up the arrays. On the other hand, it might be useful to have an indication of the amount of sleep based on the data length.

5.2.4 Termination Condition

There are multiple termination conditions that would make sense in this case. For instance, the algorithm can terminate once a pre-specified level of goodness is reached or when a certain number of generations was reached. A better criterion would probably be number of generations where performance has not been improved upon.

In this implementation, a particular method is used. An array of n elements is kept as a goodness record for the solutions already tested. Each element contains a field for the associated distance measure and a field for the array defining the individual. When a new individual with a shorter distance than the worst individual in the array, the new one is inserted in sorted order and the one with the worst distance is removed.

5.2.5 Birth Process

The birth process must be designed to complement the definition of individual. Using the duration-based definition, both mutation and crossover algorithms can be designed. Figure 13 shows one design of the birth process. This design random selects a subset of the population as input to *mutation one*, one subset for *mutation two*, and the rest for *crossover*. Each of these procedures will output its own next generation and *check for instant death* is executed to remove those individuals that fail the current definition of an individual. For instance, the individual that does not translate to 960 epochs would be instantly killed. Through this process only viable individuals are "born".

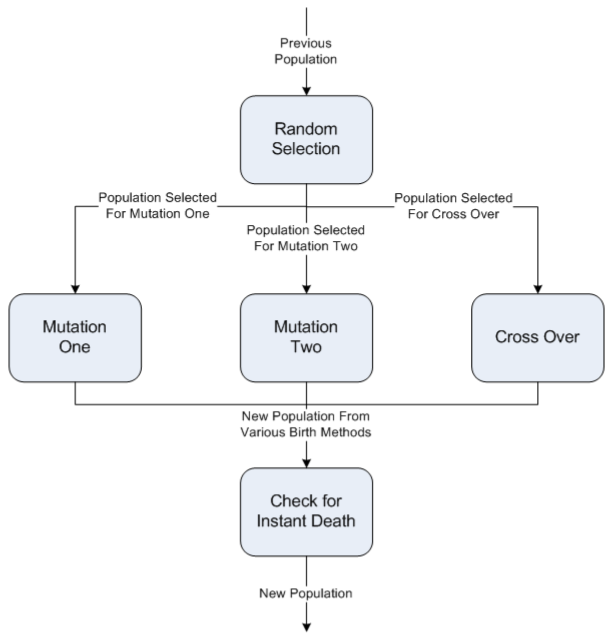


Figure 13: Flowchart for Birth Process.

There are 5 aspects of this process that require further study, the implementation of *mutation one*, *mutation two*, *crossover*, the combination of the latter three procedures, and the selection of parents.

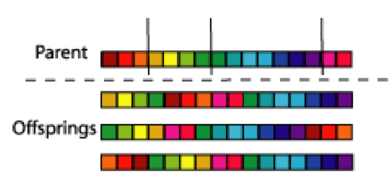


Figure 14: Mutation by cut and reattach segments.

Mutation One

Mutation One is designed to carry out drastic change from the parents to the offspring. It cuts the parents into segments and reattach the pieces. The location of the cut, the number of cuts, and the location of the reattachments would be randomly generated. The Figure 14 shows how one parent bore three different and healthy offsprings.

Mutation Two

Mutate Two aims to adjust the genetics by small amounts, for final tuning near the end. This method shifts the boundary between two neighboring segments. Figure 15 shows that but adjusting the boundary, many new individuals, highly related to the previous generation, will be generated. While this process is slow, it analyzes a region of sample space closely. To that end, it would be the best mechanism to use near the end of the search. Its attention to detail is likely to identify the closest solution.

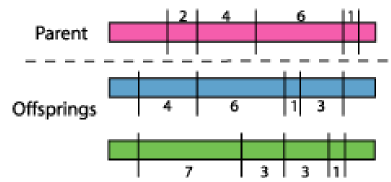


Figure 15: Mutation by shifting boundary.

Crossover

Crossover brings together two parents to produce the offsprings. The parents are cut at random points, and the segments are shuffled and rejoined. Figure 16 demonstrates crossover. The resulting offspring is also likely to explore a bigger sample space.

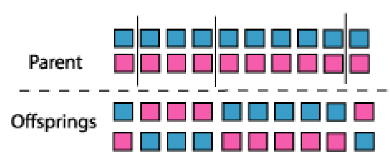


Figure 16: Crossover.

Mechanism Combination

The design in Figure 13 uses the three genetic algorithms in parallel. However, it is also possible to set these mechanisms in series. When these genetic operations occur in series, even more random or variation is created in the population. This type of set up would be suitable for a randomly created initial population, which is the second option described in Section 5.2.2. However, the initial population based on first or third option would lose performance with too much variation. The initial population was already close to the target, and the excessive variation may cause the algorithm to completely miss the nearby target.

Parent Selection

The selection of parents will have significant effect on the offspring. With the parallel layout of the genetic operations, two of the mechanisms are attempting to provide variation, while the third aims to refine the search. Therefore, it may be worthwhile to replace *random select* with a deterministic algorithm. This algorithm would direct very good parents to *mutation two*, and direct a combination of other parents to the other two mechanisms. Note that randomness is preserved because the genetic algorithms inherently contain the random factor.

5.2.6 Parameter Extraction

Though *average* described above does not provide a conventional average between a list of data sets, it does result in a pattern with minimum overall distance to all the data sets, while maintaining the standard sleep cycle shape. As indicated previously, this function is used also to smooth each data sets. Following these processing, the parameters needed in the duo-layer-multi-cycle model can be extracted.

Cycle-level State - Average Duration

The pattern resulting from *average* applied on all data sets will be an "average" of all patient data. The pattern is recorded as a set of durations, which can be extracted directly for the cycle-level states.

Cycle-level State - Standard Deviation

For each data set, the duration of the cycle-level states can be extracted from the smoothed data sets. These durations form a range, and based on the average determined previously, a distribution can be established.

Epoch-level State

The smoothed data sets are used to delimit the cycle-level states. With in each cycle-level state, the frequency of transitions into Awake, MT, previous, and next states can be calculated easily. At the same time, the average duration in each type of transition can be determined. Figure 17 demonstrates the process of identifying the transitions. It should be noted that for the cases where Awake is the previous or next state, the frequency and duration are split between the two situations.

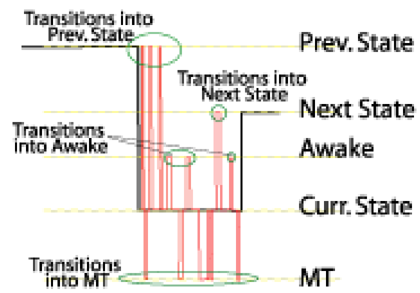


Figure 17: Epoch-level states determination.

5.3 Implementation and Result Analysis

The design in Figure 10 was implemented in Matlab v6.5. Due to a shortage of data sets, the full capability of this sleep cycle modelling algorithm cannot be illustrated. However, the available data sets were used to experiment with the basic functionality of the program. It should be noted that only *mutate two* was implemented, because the initial population was generated based on option one and three. Also, time limitations only allowed a small subset of the various design considerations discussed in Section 5.2 to be explored. The rest of this section demonstrate the function of this algorithm and its results. Then this section provides some analysis of the design considerations that were studied. It would be recommended to study the remaining design considerations at a later date.

5.3.1 Cycle-level State Parameterizations

Despite the lack of sufficient data sets, the algorithm was tested on existing data sets. Figure 18 shows the resulting curve compared with the curve provided as an initial guess. As can be seen from the diagram, the averaged curve has moved significantly from the initial guess. From the curve, the average durations are extracted and presented in Table 5.

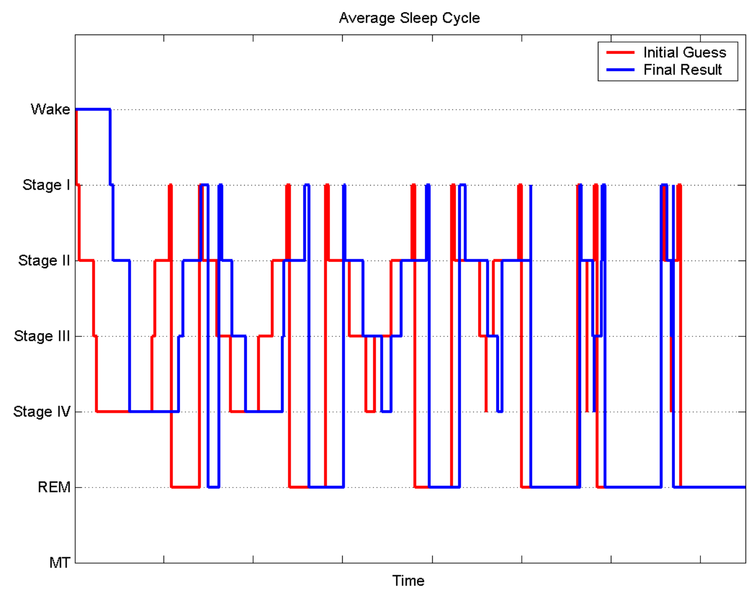


Figure 18: Average sleep cycle versus pattern from literature.

Interestingly, the characteristics discussed earlier can be observed in this averaged curve. As the night progresses, REM's duration increases and NREM IV's duration decreases. In fact, in the last two cycles, NREM IV is absent. NREM I and NREM III both have very short durations throughout the night reflecting that they are transient states.

Table 5: Average duration.

Stage	0	1	2	3	4	3	2	1	5
Cycle I	26	2	12	0	35	3	13	5	8
Cycle II		2	7	10	26	1	15	3	25
Cycle III		1	13	13	7	7	18	2	22
Cycle IV		4	16	7	3	0	21	0	35
Cycle V		1	8	1	0	5	1	2	40
Cycle VI		4	3	2	0	0	0	0	51

The standard deviation is also generated from the same data set. The values are presented in Table 6. By observation, the magnitude of the standard deviation in duration are proportional to the average duration. This observation indicates that the model for these cycle-level states can be simplified into an exponential distributions with only the average duration as parameter.

Table 6: Standard deviation in duration.

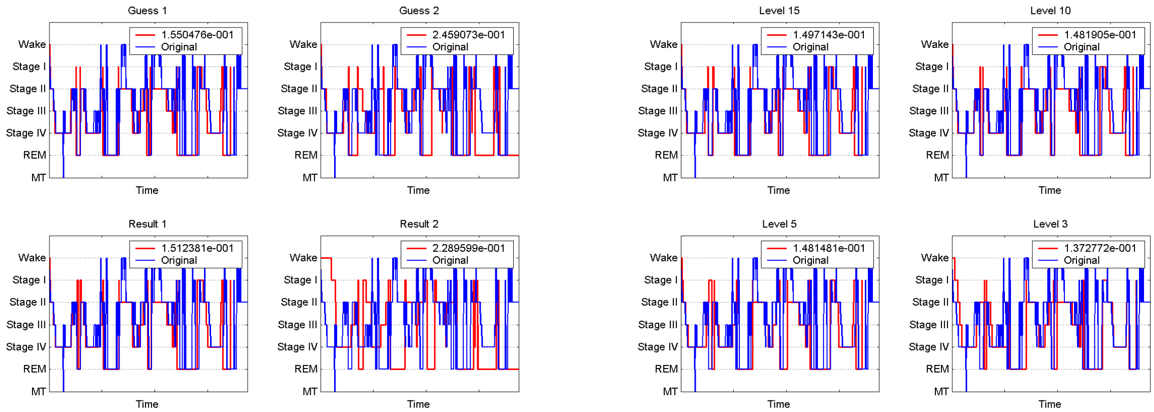
Stage	0	1	2	3	4	3	2	1	5
Cycle I	21.9547	12.557	10.344	4.2622	11.654	2.7767	8.0108	3.7563	8.7630
Cycle II		4.7039	8.9994	5.6883	7.6241	5.6666	10.063	2.1016	8.7936
Cycle III		2.7881	6.3048	5.5308	3.6955	4.8795	5.9279	2.8798	6.8418
Cycle IV		2.2121	3.9946	3.7430	3.2419	2.1602	8.3702	2.1856	5.9461
Cycle V		1.4059	1.9149	1.3441	2.2531	1.0801	4.8108	1.5631	7.5613
Cycle VI		1.6753	2.1409	0.52281	0.58310	0.77028	2.1276	0.47258	2.7970

5.3.2 Quality of Initial Population

A good initial guess of the standard curve will affect the smoothing performance. Figure 19 Part A demonstrates this effect. Guess 1 was generated based on manual observation and Guess 2 comes from a standard curve in literature. Guess 1 outperforms Guess 2 by a significant margin. Since the algorithm using the *mutate two* mechanism, can only improve the solution, both results shows improvement from the guess. However, Result 2 cannot match the performance of Guess 1 or Result 1. Therefore, unless the other genetic algorithms are implemented to bring in variance, the current algorithm relies heavily on a good initial guess.

5.3.3 Mutation Parameters

The current version of *mutate two* incorporates two parameters. The first parameter *mutation number* determines the average number of mutation spots to occur in each generation.



(a) (b)

Figure 19: Design considerations.(a) Quality of initial guess. (b)Mutation parameter.

The second parameter *mutation level* determines the average amount of boundary shift. Figure 19 Part B demonstrates the effect of changing *mutation level*. The levels used are 3, 5, 10, and 15. Note that the larger levels correspond to poorer performance. Since this mutation algorithm is meant to refine the search towards the target, large shifts in boundary will cause the algorithm to miss the target. However, the factor that cannot be seen from the graph is that the smaller levels require far more generations to find a local minima, where as larger levels work much faster. Experiment with changes in *mutation number* has similar effect.

Though these two parameters are adequate for this simple mutation algorithm, better adjustment parameters are required to improve the performance of the algorithm. One idea is to use adaptive parameters to introduce both speed and performance gains. Also, the parameters for the other two proposed genetic algorithms will have significantly different dynamics. These design consideration need to be studied at a later date.

6 Application of Contextual Analysis

This report showed the implementation of two sleep cycle models, the one-cycle-duo-direction model and the duo-layer-multi-cycle model. This section will first demonstrate the effect of contextual analysis using the one-cycle-duo-direction model. Though Section 5 documented the methods of developing the comprehensive model, data set shortage hinders the performance of the designed algorithm. Lacking a representative model, the application of this model to contextual analysis will be discussed and tested at a future date.

6.1 One-Cycle-Duo-Direction

This model can be applied to contextual analysis in various ways. This section will present two approaches. The first one is very intuitive, but the other required components are not present to test it. The second approach is adapted to classification methods developed in the ECE 710 project during the Fall of 2003.

6.1.1 Method One

The input would be the knowledge of the previous state or states, the state transition probability matrix [9], the current PSG signal patterns, and some fuzzy classification algorithm. The classification algorithm will produce probabilities of the current epoch of PSG signals matching each state's characteristics. They can be named $\{\phi_0, \phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6\}$. The previous state points to a set of transition probabilities, which is named $\{p_0, p_1, p_2, p_3, p_4, p_5, p_6\}$. Then the new probabilities, $\{q_0, q_1, q_2, q_3, q_4, q_5, q_6\}$, are defined as $q_i = \phi_i \times p_i$ for $0 \leq i \leq 6$. In effect, the context information is used to strength or detract from the belief of the classification of the current state.

Given a reasonable fuzzy classifier, this method would be simple and direct. The short-

[9] Recall that the transition probability matrix for this model is in Table 2.

coming of this approach includes all the disadvantages of this simple sleep cycle model and the fact that these transition probabilities are static. Despite these foreseeable problems, this approach is worthy of further analysis.

6.1.2 Method Two

This method is particularly adapted to the stage differentiating features designed for the ECE 710 course project. The set of features considered in this method is used to differentiate Awake and REM from the other states. The formula by which the context is considered for a feature that differentiates Stage i and Stage j is

$$f_{adjusted} = f \times \left(\frac{P\{\text{Current Epoch is i}\}}{P\{\text{Current Epoch is j}\}} \right)^n,$$

where f is the raw feature value and n is the emphasis factor. This formula considers the context, in that it uses the previous epoch's stage to provide a probability of current epoch being some stage i or j. The value n allows us to control how much context can influence the feature value.

The usefulness of this type of contextual information is demonstrated in Table 7. This table considers the Ratio feature's ability to differentiate the Awake state from the other states. The left side of the table shows the sensitivity and specificity of the feature on its own. The right side shows the performance of this featured adjusted with contextual knowledge. It should be noted that the context maintained or improved the performance.

Without context, this feature had low accuracy in terms of differentiating from NREM I and REM. One reason is that these two stages both bear some resemblance to the Awake stage. However, the context is able to improve awake and REM differentiation by nearly 16.7%. It also improves the performance for differentiating NREM I. However, the improvement is not as noticeable as REM, because NREM I, the transient state, may not have enough occurrence to have significant representation in contextual analysis.

Table 7: Ratio feature to differentiate Awake from other stages.

Stage	Orig. Sens.	Orig. Spec.	New Sens.	New Spec.
NREM I	0.9644	0.8475	0.9666	0.9153
NREM II	0.9936	0.9651	0.9920	1.0000
NREM III	0.9973	0.9787	0.9984	1.0000
NREM IV	1.0000	0.9803	1.0000	1.0000
REM	0.9352	0.8279	0.9947	0.9953

6.2 Duo-Layer-Multi-Cycle

This model contains significantly more information than the *one-cycle-duo-direction* model. Therefore, the strategic application of this data to contextual analysis can aide automatic sleep staging significantly. At the most basic level, two intuitive applications exist.

The first application uses the probability distribution parameterized with the average and standard deviation of duration to produce transition probabilities. This time correlated probability is more accurate than the discrete model used in Section 6.1.1. All other aspect of this application remains the same. It still uses these transition probabilities to bias the belief in the classification results.

The second application is relevant to particular feature differentiators. Based on the previous staging result, corresponding differentiators can be turned on. For instance, once reaching Cycle II red REM, the detector for NREM I and NREM II should be monitored. Emphasis should be placed on features that differentiate these three stages. By keeping track of the cycle-level states, the output of appropriate detectors are considered to find the next state transition. The advantage of this method is that the design of particular state detectors is far simpler than a detector that differentiates all the states. In order to avoid the case where one staging error causing a cascading effect, the contextual information must be validated by other events.

7 Conclusion

Research to automate sleep staging process has finally begun to emphasize the contextual analysis. The context can clarify ambiguous cases based on the previous state, the amount of time spent in the previous state, and the amount of time in the overall night. In order to conduct contextual analysis, an appropriate model of the sleep cycle must be developed.

A simple model, *one-cycle-duo-direction*, can contain the basic state transition probabilities. Its data requirement is low and its implementation is simple. Using very intuitive methods, its application to contextual analysis, thereby to sleep staging can be accomplished. It can also be adapted to particular stage differentiating features. In this application, it can improve the performance of otherwise weak cases by up to 16.7%. The resulting sensitivity and specificity are above 99% for differentiating Awake from NREM II, NREM III, NREM IV, and REM. Even differentiating Awake from NREM I has performances above 90%. However, this model's simplicity translates to certain weaknesses and ambiguity, such as lack of tracking time within a state, cycle, night, and direction differentiation.

A comprehensive model, *duo-layer-multi-cycle*, adopts a six-cycle model that tracks direction. Its upper level tracks the overall behavior in continuous time, and the lower level studies the short term transitions. A modelling procedure based on genetic search algorithms is designed and demonstrated to construct this model from source patient data. High data requirement meant that certain aspects of this model could be tested. Due to the amount of contextual information contained in this model, it can be projected to provide significant improvement to cases that are ambiguous to classify.

8 Recommendation

There are a number of related tasks that should be conducted to render this analysis complete. These tasks are

- Study the remaining design considerations. Apply the competing alternatives and select the better choice. Some combination of alternatives should be developed for some cases.

- Implement *mutate one* and *crossover* genetic algorithms. Identify appropriate parameters to control these algorithms. Compare and contract these two mechanism's ability to introduce variation.

- Improve *mutate two* by adopting adaptive parameters.

- Complete parameter extraction algorithms for epoch-level states.

- Collect more data sets to build a representative *duo-level-multi-cycle* model. Design the collection of data sets based on guidelines indicated in the report.

- Design the procedures to validate a given *duo-layer-multi-cycle* model. In particular, verify the appropriateness of the probability distribution of the cycle-level states.

- Develop better applications of the *duo-layer-multi-cycle* model in contextual analysis. Test these application either with a mock model or with the our current limited data set.

- Study approach 1 of one-cycle-duo-direction model's application to contextual analysis. Collect required components, such as fuzzy inference system or neural network, to test this approach.

9 Glossary

ACNN: All connected neural network.

AI: Artificial intelligence.

ANN: Artificial neural network.

EEG: Electroencephalogram.

EMG: Electromyogram.

EOG: Electrooculogram.

FIS: Fuzzy inference system.

MLP: Multi-layer perceptron

MT: Movement time.

NREM: Non rapid eye movement.

PSG: Polysomnogram.

REM: Rapid eye movement.

SS: Sleep spindles.

TDNN: Time-delay neural network.

References

[1] Rechtschaffen, A., and Kales, A., eds., "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects", NIH Publication No. 204, US Government Printing Office, Washington, DC., 1968.

[2] Hobson, J. A., Sleep, Scientific American Library, New York, 1989.

[3] Ogilvie, R. D., "The Process of Falling Asleep" Sleep Medicine Reviews, Vol. 5, No. 3, pp. 247-270, 2001.

[4] "History of Sleep Research" `http://library.thinkquest.org/25553/english/study/research/hist.shtml`

[5] Shimada, T., Shiina, T., and Saito, Y., "Sleep Stage Diagnosis System with Neural Network Analysis", Proc. of the 20th Ann'l Int'l Conf. of the IEEE Engineering in Medicine and Biology Society, vol. 20, no. 4, pp. 2074-2077, 1998.

[6] Pacheco, O. R., and Vaz, F., "Integrated System for Analysis and Automatic Classification of Sleep EEG", Proc. of the 20th Ann'l Int'l Conf. of the IEEE Engineering in Medicine and Biology Society, vol. 20, no. 4, pp. 2062-2065, 1998.

[7] Oropesa, E., Cycon, H. L., and Jobert, M., "Sleep Staging Classification using Wavelet Transform and Neural Network", TR-99-008, International Computer Science Institute, Berkeley, California, March 30, 1999.

[8] Flexer, A., Sykacek, P., et al., "Using Hidden Markov Models to build an automatic, continuous and probabilistic sleep stager", IEEE, pp.627-631, 2000.

[9] Flexer, A., Gruber, G., and Dorffner, G., "Improvements on Continuous Unsupervised Sleep Staging", IEEE,pp. 687-695, 2002.

[10] Van Hese, P., Philips, W., et. al., “Automatic Detection of Sleep Stages using the EEG”, 2001 Proc. of the 23rd Ann’l, EMBS Int’l Conf., pp.1944-1947, Istanbul, Turkey, Oct. 25-28, 2001.

[11] Bankman, I. N., Sigillito, V. G., et al., “Feature-Based Detection of the K-Complex Wave in the Human Electroencephalogram Using Neural Networks”, IEEE Trans. on Biomedical Engineering, vol. 39, no. 12, pp. 1305-1310, 1992.

[12] Henry, D., Sauter, D., and Caspary, O., “Comparison of Detection Methods: Application to K-Complex Detection in Sleep EEG”, IEEE, pp. 1218-1219, 1994.

[13] Pohl, V., and Fahr, E., “Neuro-Fuzzy Recognition of K-Complexes in Sleep EEG Signals”, IEEE, pp. 789-790, 1997.

[14] Akin, A., and Akgul, T., “Detection of Sleep Spindles by Discrete Wavelet Transform”, IEEE, pp. 15-17, 1998.

[15] Gorur, D., Halici, U., et al., “Sleep Spindles Detection Using Short Time Fourier Transform and Neural Networks”, IEEE, pp. 1631-1636, 2002.

[16] Huupponen, E., Lehtokangas, M., et al., “EEG Alpha Activity Detection by Fuzzy Reasoning”, IEEE, pp. 411-416, 2001.

[17] Shimada, T., Shiina, T., and Saito, Y., “Detection of Characteristic Waves of Sleep EEG by Neural Network Analysis”, IEEE Trans. on Biomedical Engineering, vol. 47, no. 3, pp. 369-379, March 2000.

[18] Norman, R., Pal, I., et al., “Interobserver agreementamong sleep scorers from different centers in alarge dataset”, Sleep, vol. 23, no. 7, pp. 901-908, 2000.

[19] Nayak, A., and Roy, R. J., “Time Frequency Spectral Representation of EEG”, IEEE, pp. 7-8, 1993.

[20] Scheuer, M. L., "Continuous EEG Monitoring in the Intensive Care Unit", Epilepsia, vol. 43, no. Supple. 3, pp. 114-127, 2002.

[21] Muthuswamy, J., and Thakor, N. V., "Spectral analysis methods for neurological signals", Journal of Neuroscience Methods, vol. 83, pp. 1-14, 1998.

[22] Cranstoun, S. D., Ombao, H. C., et al., "Time-Frequency Spectral Estimation of Multichannel EEG Using the Auto-SLEX Method", IEEE Transactions on Biomedical Engineering, vol. 49, no. 9, pp. 988-996, 2002.

[23] Laing, R. J. A., "A Sleep Spindle Detection Algorithm", Master Thesis.