 67-14,140

PASKEWITZ, David August, 1942-
THE QUANTIFICATION OF NOCTURNAL ELECTRO-
ENCEPHALOGRAPHIC PATTERNS IN MAN.

The University of Oklahoma, Ph.D., 1967
Psychology, experimental

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE QUANTIFICATION OF NOCTURNAL ELECTROENCEPHALOGRAPHIC PATTERNS IN MAN

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

DAVID AUGUST PASKEWITZ

Norman, Oklahoma

1967

THE QUANTIFICATION OF NOCTURNAL ELECTROENCEPHALOGRAPHIC

PATTERNS IN MAN

APPROVED BY

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

I wish to thank each member of my committee for his time and effort. Special thanks go to Dr. Arnold E. Dahlke for his patience and understanding during the preparation of this dissertation.

I wish to express my thanks to Dr. Harold L. Williams, Dr. Boyd K. Lester, and the rest of the staff of the Whiteman House for Mental Health Research, Department of Psychiatry, Neurology and Behavioral Sciences, University of Oklahoma Medical Center for the many hours of time and the considerable effort which they put forth during the collection and analysis of the data.

My appreciation is also extended to the University of Oklahoma Medical Center Computer Facility and to Mr. Gary Haskin in particular. The many hours which he spent programming, running data, and coordinating efforts at the Computer Facility were invaluable.

I would further like to acknowledge the encouragement and support of my wife, Carol, without which I would not have found it possible to complete this research.

The research reported in this dissertation was supported in part through the Office of the Surgeon General, United States Army, Contract No. DA-49-193-MD-2758.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

THE QUANTIFICATION OF NOCTURNAL ELECTROENCEPHALOGRAPHIC PATTERNS IN MAN

CHAPTER I

INTRODUCTION

Sleep is a basic state of physiological functioning. Across mammals from man to the Virginia Oppossum (Snyder, 1964) striking similarities have been obtained in physiological sleep patterns. Much of the present knowledge about the functional anatomy of the brain has been discovered through a study of the sleeping state in man and in lower animals. The many articles by Jouvet and his associates on sleep mechanisms in the cat and the work on the brain stem reticular formation by Moruzzi and Magoun are but cases in point. During sleep an individual is relatively free from influences generated by the surrounding environment as well as from transient psychological states resulting from the experimental situation. While the individual is asleep, physiological events like those associated with psychopathic disorders (Lester & Burch, 1965) or those occurring as the result of the administration of drugs (Muzio, Roffwarg, & Kaufman, 1964; Rechtschaffen & Maron, 1964) can be studied in this environment comparatively free of external stimuli.

In addition to the study of physiological events taking place within the organism, sleep researchers have been interested in the behavioral properties of sleep and drowsy states. As noted by Lindsley (1960) and others (Williams, Hammack, Daly, Dement, & Lubin, 1964), the behavioral efficiency of an individual is related in part to the position he occupies on a waking-sleeping continuum (Lindsley, 1960). In assessing this relationship between behavior and the depth of sleep (or its converse, degree of alertness) it is necessary to specify depth of sleep in some form which is independent of the behavioral measures themselves. The primary instrument which has been used to determine the level of alertness of the organism is the electroencephalograph (EEG).

Berger (1930) first recorded brain potentials as the EEG, taking some of his recordings during sleep. Since that time many others have used the technique, among them Kleitman (1963) and Dement (1955; 1958). From the beginning, recurrent configurations of wave amplitude and frequency have been used to classify EEG patterns into behaviorally relevant states. The first to do so for sleep were Davis, Harvey, Loomis, and their associates (Davis, Davis, Loomis, Harvey, & Hobart, 1937; Harvey, Loomis, & Hobart, 1937; Loomis, Harvey, & Hobart, 1937). They assigned the first five letters of the alphabet to the successive stages of sleep from wakefulness to deep sleep as follows (after Kleitman, 1963): A, to the interrupted alpha rhythm pattern, 9-11/second at about 60 microvolts in amplitude, seen during relaxed wakefulness; B, to a low voltage, irregular pattern, seen during the passage into sleep; C, to a spindle pattern, with spindles of 14-15/second waves of from 20-40 microvolts in amplitude superimposed on an irregular pattern of slower waves; D, to a pattern much like C except for the appearance of delta waves of about 1-3/second frequency and of up to 300 microvolts in amplitude and;

E, to a random pattern made up primarily of very slow and large delta waves.

Although this system has found continuing usage, a classification scheme devised more recently by Dement and Kleitman (1957) will be used throughout the remainder of this paper. A simplification of the earlier system, the Dement-Kleitman system proposed four stages of sleep: 1, low voltage activity with irregular frequency; 2, sleep spindles of about 14/second frequency and high voltage, sharply diphasic K-complexes in a background of low voltage, fast activity; 3, random delta waves with some spindling and; 4, predominantly large, slow delta waves. As with the earlier classification, the successive stages correspond to deeper phases of sleep.

With the addition of a waking state, the pattern found in a person who is relaxed with eyes closed, consisting primarily of alpha activity, five states may be clinically distinguished by reference to the oscillograph records. One further state may be distinguished on the basis of eye movements which occur during sleep. In the course of a night, the stages of sleep vary in a cyclic pattern (Kleitman, 1963). Every sixty to ninety minutes the subject emerges out of deeper sleep into a pattern resembling stage 1 sleep. During this emergent stage 1 sleep, there occur rapid eye movements which have led some authors to term this phase "REM sleep" (Luce, 1965). This eye movement phenomenon, first studied by Aserinsky and Kleitman (1953; 1955), was found related to dreaming. Subjects awakened during periods of REM activity indicated often that they had been dreaming whereas if awakened during other periods of sleep they rarely reported dreams. The question of the depth of REM sleep is open to controversy. The brain is highly active and both metabolism and EEG activity may resemble the state found during periods of intense waking concentration. Peripheral musculature, on the other hand, lacks tonus, although the eyes move in their characteristic spurts (Luce & Segal, 1966). Williams, Morlock, and Morlock (1966) found that their subjects responded to few auditory stimuli during REM sleep when simply instructed to respond but when failure to respond resulted in aversive stimulation they did as well as during "lighter" stages of sleep. In this paper no differentiation will be made between initial and emergent (REM) stage 1 classifications since the EEG patterns are so very similar and since the initial stage 1 period is usually short-lived and easily identified through its occurrence at the beginning of the night.

Accurate information as to the depth of sleep or classically defined stages of sleep is important not only in studies of the behavioral properties of sleep and drowsy states but also in studies of the effects of various drugs on sleep patterns throughout the night. Gresham, Webb, and Williams (1963), for instance, found that alcohol diminishes the duration of REM periods during the first hours of sleep. Rechtschaffen and Maron (1964) found that dexedrine can reduce the amount of the night spent in REM sleep. Other studies which have a need for accurate recognition of EEG sleep stages are those in which the investigators attempt to deprive the subject of a specific stage of sleep for some portion of the night, usually by waking him as soon as he enters that particular stage. Dement (1960) pioneered the investigation of REM deprivation and produced evidence to indicate that this stage of sleep may be important to a person's psychological well-being. The important problem in studies of this sort is the identification of the exact moment of transition from some other stage of sleep into that stage of which the subject is to be deprived. Typically, studies of specific stage deprivation have involved a tremendous amount of work on the part of the investigators who must monitor the recordings continuously so as not to miss these transitions (Luce, 1965).

Work by Aserinsky (1965) indicates that during the REM stage of sleep measures of respiration show differing but consistent patterns between periods of ocular quiescence and those periods during which the eyes are making the movements typical of this stage of sleep. Studies of this sort also point up the need for accurate information about changes in the state of a sleeping subject.

The classification of an EEG record into the stages of sleep and the identification of transitions from one state to another has been and remains primarily a job for a clinical encephalographer or some other person who has received the training necessary to reliably distinguish the patterns of activity. The degree and type of experience at classification which a person has often bears heavily on his reliability. The place at which he received his training, the person under whom he learned, and the use to which the records are put may all affect the amount of agreement which may be anticipated between a classifier and other persons classifying the same record (Burch, Vorderman, & Dossett, 1966; Kozhevnikov & Meshcherskiy, 1964). Monroe (1967) indicates that high variability can be expected between raters with respect to the total number of minutes a subject spends in a particular stage of sleep. He obtained standard deviations ranging from 3.88 minutes to 18.45 minutes for the stages with data from 27 independent raters. Furthermore, as Burch et al. (1966) indicate, trained personnel have only a limited amount of time to study records or to classify and interpret the EEG as it occurs.

Personal communication with a number of researchers trained in the analysis of sleep records has indicated that the identification of transitions between stages of sleep is particularly difficult. Often the researcher must wait for several minutes before he is sure that a transition has taken place. Amplitude differences, since they are more readily apparent than are subtle frequency changes, also carry a disproportionate weight in the judgment of whether or not the sleep has changed (Burch et al., 1966; Kozhevnikov & Meshcherskiy, 1964). Due to these difficulties, those investigators interested in REM deprivation have often employed measures other than those obtainable from the EEG, such as muscle relaxation, as indicators of entry into the REM stage (Dement, 1965; Kales, Hoedemaker, Jacobson, & Lichtenstein, 1964).

In light of the problems facing a person attempting to analyze EEG data visually, as outlined above, there has been an increasing use of objective means of characterizing EEG patterns during the past few years. Most of this work has used electronic means, notably digital computers (Burch, 1959; Burch, Nettleton, Sweeney, & Edwards, 1964; Farley, 1961; Kozhevnikov & Meshcherskiy, 1964; Shapiro & Fink, 1966). Some of the techniques have been applied to sleep data. Fink and Shapiro (1965) used several methods, notably period analysis, to classify sleep stages. Burch et al. (1966) have used period analysis to classify levels of consciousness in records of Gemini astronauts. Johnson, Nute, Austin, and Lubin (1966) used an auto-spectral analysis to investigate the changes which occur in the various spectra with changes in the depth of sleep.

Many of these techniques require the use of a high capacity digital computer in association with other complex digitizing and recording equipment. An added disadvantage, particularly in the case of establishments with insufficient funding to employ a computer on demand, is the fact that much of the analysis is thereby necessarily off site and the results of the treatment are available only after the raw data are transported to the facility for analysis and that analysis can be completed. In the case of some of the techniques the analysis itself is complex enough to prohibit its use with ongoing records unless samples are taken only at widely spaced intervals. It is the purpose of the present paper to develop and evaluate a technique suitable for on-site, continuous, real-time analysis of EEG sleep records and their classification into the Dement-Kleitman sleep stages.

A system to aid or replace the clinician in the classification of sleep records should be convenient and reasonably inexpensive. In addition, it should meet three criteria: 1) The system should, within limits, be as reliable as are several clinicians among themselves in classifying sleep stages; 2) The system should show more responsiveness to changes than can the clinician and; 3) It should maintain its reliability not only within one night on one subject but across several nights on that subject and across several subjects.

Any system must first express the EEG record in some set of characteristic values and then proceed to classify sections of the record on the basis of those values. These two functions, analysis and classification, will be discussed separately.

The most frequently encountered methods of analysis are analog bandpass filters, auto-spectral or power spectral density analysis, amplitude analysis, random shape analyses, and period analysis. Analog bandpass filters may be of several types. Walter (1943), one of the pioneers in automatic analysis of the EEG, used electro-mechanical resonators for his analyzer. Other analyzers have since been built using both electro-mechanical and purely electronic resonant filters, including those using operational amplifiers in active filter designs. Kozhevnikov and Meshcherskiy (1964) review quite thoroughly the design of these sorts of filters. They point out that the principal advantage in this type of analysis is the very sharp cutoff characteristic of the filters, allowing maximal separation of the various frequencies of interest. They also report, however, a difficulty in maintaining the exact frequency characteristics from run to run, particularly in the case of narrow bandwidths. The output of the filters is typically integrated over time to yield an average power function for the frequency band under observation.

Auto-spectral techniques are essentially a discrete form of Fourier analysis, the process of breaking down a complex, regular wave into the component waves which make it up. The principle involves the digital filtering of an autocorrelation function to uncover the periodicies it contains. Autocorrelation is the process of multiplying the record by a displaced copy of itself (Walter, 1963). Regular waves produce functions which oscillate in a regular fashion as the lag between the two correlated samples is increased while aperiodic waves produce functions which rapidly go to zero as the lag increases. Any regular wave with concomitant "noise" riding on it will result in a regular autocorrelation function of reduced amplitude relative to that of a purely regular wave. Thus, the autocorrelation function accentuates regular activity at the expense of random activity. Once the autocorrelation function has been obtained, it is then multiplied by other functions specifically weighted to yield synchrony with the autocorrelation function at a given frequency. In this way the output of the filters is taken as the sum of the products of the time (autocorrelation) function and the filtering function. This sum is a maximum when the activity is regular and at the exact center frequency of the filter. Frequencies above or below the center frequency result in lessened outputs. Spurious peaks may occur at frequencies other than the center frequency. These peaks can be smoothed by weighting the estimates for the filtering function through procedures known as "hanning" and "hamming." The outputs of the filters may be expressed as a proportion of the total variance of the wave, given as the autocorrelation function of the wave at zero lag, or as values of the filters relative to each other. While the auto-spectral method does give its user a frequency-wise breakdown of an EEG signal, certain considerations must be taken into account as regard its usefulness in a sleep stage identification system.

Assuming on-line analysis as desirable, the technique requires the use of a reasonably large general-purpose digital computer. Even so, analysis time is sufficiently long to make the omission of a considerable portion of the record necessary during the time that the analysis is being carried out. Furthermore, as Walter (1963) points out, the autospectrogram does not respond to aperiodic activity nor does it take into account changes in parameters occurring within the analysis period.

Some authors, notably Brazier (Brazier & Barlow, 1956; Brazier & Casby, 1952) have used analog computers for some of the calculations involved in auto-spectral analysis. Gauss (1964) and Whittlesey (1964) have developed methods of obtaining the power spectrum directly by means of digital filtering through a replacement of original data points with weighted sums of neighboring points but these methods involve more effort and correspondingly more analysis time than do the autospectral methods discussed above.

Amplitude measures of the EEG may be used to quantify the waves in a record. Perhaps the simplest method of doing this is simply to integrate the raw EEG over time. Drohocki (1948) has developed an integrator which provides a continuous cumulative measure of the area under successive brain waves, irrespective of frequency. The output consists of a series of uniform pulses the rate of which is proportional to the total cumulated area under the EEG waves. Other methods of determining amplitude are possible, depending for the most part on the type of frequency analysis to be performed. If one is performing a spectral analysis, the total variance of the wave may be taken as a measure of amplitude. In a period analysis the time taken between successive baseline crossings may be used, however, as indicated by Shapiro and Fink (1966), this measure is dependent to a large extent upon the frequency of the primary wave. One may also take a discrete sample of the record and produce a closed figure by joining the end points of the sample to a baseline taken as the greatest negative value in the sample wave. The area of the figure so produced may be considered as a measure of amplitude. A considerable amount of time is required to arrive at this measure.

A random shape analysis (Vanderplas, Sandison, & Vanderplas, 1965) involves the calculation of a number of attributes of a closed figure produced from a discrete sample of the EEG activity. For example, one second of EEG activity is taken from the record and an irregular polygon made by joining the endpoints of the sample to an artificial baseline. This baseline is usually a value at or somewhat greater than the largest negative value which the wave assumes during that one second. Any measures appropriate to the irregular figure may then be calculated. Some common measures used in this sort of analysis are area, perimeter, area to perimeter ratio, number of angles in the figure, and the relationships between angles. The procedure can be applied to any number of samples and the resulting values compared to uncover similarities and differences. Shapiro and Fink (1966) report, however, that a random shape analysis is extremely time consuming and therefore not suited to on-line applications.

Period analysis was developed by Neil Burch and his associates (Burch, Nettleton, Sweeney, & Edwards, 1964). The analysis is based on a mathematical model of cortical functioning using a Gram-Charlier series, the parameters of which are evaluated through coding the analog EEG signal and its first and second derivatives into square-wave trains corresponding to the baseline crossings of the analog functions (Saltzberg & Burch, 1959). The square-wave trains are produced by a Schmidt trigger which is activated as the wave assumes a positive (or negative) value and turns off as the wave crosses the baseline again, yielding a square-wave whose duration is the time during which the signal is positive (or negative). In period-analytic parlance the square-wave trains resulting from the coding of the primary function, the first derivative, and the second derivative are called the major, intermediate, and minor periods, respectively. The intermediate and minor periods give indications of the "riding" activity, that is, activity of lower amplitude and differing in phase and/or frequency from that of the primary signal. It is clear that for a regular wave the time that the signal is above the baseline can be translated into frequency since, for example, a wave of ten cycles per second would complete one cycle in a tenth of a second during which it would be positive in value for half of the time, or five one-hundredths of a second. This relationship between the period of a wave and its frequency can be used to design digital filters to indicate activity occurring within a certain band of frequencies. Any wave which can meet upper and lower limits with respect to its period is classified as activity within the band defined by those limits. By setting the upper and lower limits the width of a band, the degree of overlap between bands, and the center frequency of the band can be made to correspond to classical patterns in the EEG or to any other criterion desired. Similar or differing sets of filters may be applied to the major, intermediate, and minor periods.

The result, then, is a form of analysis which is primarily frequency sensitive, designed to be relatively insensitive to amplitude, which provides considerable flexibility and is extremely well suited to on-line analysis if special-purpose coding and filtering devices are used. In addition to special-purpose devices, the period analysis technique is applicable to the use of a general-purpose digital computer (Burch et al., 1964). The most typical way of expressing the output of period analysis is as a total count of the number of waves meeting the frequency criteria for each band as accumulated during some time period (data epoch). It has been historically popular to use ten seconds as the epoch length (Burch et al., 1964; Shapiro & Fink, 1966). It is possible, however, either to increase or decrease the epoch length, making it convenient to view trends over longer periods of time or to look at very short-term changes associated with specific stimuli.

In choosing a system of analysis to quantify EEG records preceding their classification into sleep stages, several considerations must be made. These considerations were made in the present case with the goal in mind of arriving at an on-line, real-time classification system for a reasonable price. The first consideration is that of analysis time necessary to achieve resolution which will enable accurate classification. Although relatively simple to obtain, the usual amplitude measures in and of themselves are usually insufficient to allow unequivocal classification of EEG patterns following differing dosages of amobarbital. Agnew, Parker, Webb, and Williams (1967) report correlations (Eta) ranging from 0.72 to 0.96 between scored stages of sleep and the output level of a Drohocki (1948) integrator. An examination of their results leads to the conclusion, though, that this sort of analysis would allow only minimal differentiation between the waking state, stage 1, and stage 2. Random shape analysis, as indicated earlier, requires a considerable amount of time to perform and hence is not well suited to real-time analyses. Analog filters, an auto-spectral analysis, or period analysis all fall within reasonable limits with respect to analysis time. Analog filters process the record continuously and the resolution is dependent upon the time base used for integration of the outputs. An auto-spectral analysis, as noted earlier, is sufficiently difficult so as to require the omission of a portion of the ongoing record in the analysis. Period analysis can be a continuous process provided special-purpose equipment is used. When a general-purpose digital computer is used for the analysis, the analysis time is sufficiently long so as to require the omission of a certain amount of the record. Shapiro and Fink (1966) find that they lose about one second between epochs. With the special-purpose equipment available, all of the record is included in the analysis and the resolution available in terms of output is dependent upon the epoch length chosen.

A second consideration concerns the equipment necessary to complete the analysis. In the case of analog filtering, considerable special equipment in the shape of separate filters for each band of frequencies is necessary. Auto-spectral analysis requires that a general-purpose digital computer be available continuously during the course of the night while period analysis may either be performed on relatively simple special-purpose equipment or on a general-purpose digital computer with an increase in analysis time.

A third consideration involves the accentuation or attenuation of certain portions of the EEG activity relative to the others. In this respect auto-spectral analysis, as mentioned earlier, emphasizes regular activity while period analysis codes and filters the wave without regard to regularity, that is, a wave of a certain period will register within the frequency band for that period regardless of the activity which precedes it. The output with period analysis is thus potentially more variable. Auto-spectral analysis also tends to emphasize low frequency components in the EEG since they are usually of greater amplitude and are thus more powerful. In contrast, period analysis gives equal valence to counts in any frequency range, regardless of amplitude. In light of the randomness usually found in fast activity and its lack of amplitude relative to slower activity, period analysis gives a better look at fast activity than does an auto-spectral analysis.

A fourth consideration concerns the ease of interpreting the output of the analysis. Auto-spectral analysis typically is expressed as a percentage or proportion of the total variance of the wave. Period analysis is expressed as the total count of waves within a band during the analysis epoch. Either output is easily amenable to further operations. Period analytic filters are more easily set with respect to bandwidth, band overlap, and center frequency to correspond to the clinically recognized delta, alpha, beta, and other frequency bands than are those filters programmed into an auto-spectral analysis.

After consideration of all these factors involved in an analysis system, it was decided that a period analytic approach had the most to offer, particularly with respect to on-line capability and cost of acquisition. An attempt was made to set up a special-purpose analyzer which would yield a maximum amount of clinically relevant information within six bands of output. Major period bands were calibrated to correspond to delta, theta, and alpha activity and intermediate period bands to correspond to alphoid (riding or very regular alpha), spindle or sigma, and beta activity. This analysis system was used throughout the study to provide the data used in classification.

It is the goal of this study to develop a system of classification using the period analytic data which will enable accurate assignment of sleep record epochs to the Dement-Kleitman sleep stages when compared to a group of trained human interpreters. Furthermore, the system should indicate a potential for development into an on-line analog special-purpose computer. A possible first step in classifying epochs into sleep stages is the calculation of a multiple discriminant function analysis (Cooley & Lohnes, 1962). This analysis uses the original scores on a number of measures (in this case the counts for each of the six bands) to determine scores along lines (discriminant functions) so placed in the score space as to maximally separate several predefined groups from each other. Orthogonal discriminants exist up to as many as there are measures or one less than the number of groups, whichever is the lesser number. A particular individual combination of original scores, then, yields a combination of discriminant scores along lines on which the groups are overlapped as little as possible. As Cooley and Lohnes (1962) point out, the primary advantage in calculating a discriminant function analysis is that it often leads to a significant reduction in the dimension of the predictor space without a significant loss in information. There are approximate tests of the significance of discriminability of the analysis and indicators of the relative contributions and significance of the contributions of the original variables to the analysis to enable the investigator to evaluate the usefulness both of the analysis itself and his original measures.

The classification itself, whether carried out in the M-dimensional space generated by the original measures or in the reduced space resulting from a multiple discriminant function analysis, may be based on one of two separate criteria. The first of these is a minimum distance criterion. To be classified as a member of a particular group, the sum of the squared deviations of the individual scores from that group's means on the measures (or that group's centroids on the discriminant functions) must be at a minimum for that group. Shapiro and Fink (1966) have had some success in using this criterion to classify responses to sodium pentothal. They used a learning technique whereby initial values of the group means were chosen and then modified by recalculating new means including each new set of values once a particular epoch had been classified as belonging to the group.

The second criterion which may be used in classifying individual epochs into groups is the maximum likelihood criterion. Several decision functions may be applied to the data in meeting this criterion depending on the variables which one wishes to control. Cooley and Lohnes (1962) cite two basic functions. The first of these is based entirely on the centour (centile contour) concept. This concept may be likened to the ellipse enclosing the scores for a group from a normal bivariate population. Similar ellipses can be generated enclosing any specified percentage of the scores in the group. These ellipses are the centours. In a space of more than three dimensions, these centours become hyperellipsoids. Any score vector (set of scores for an individual on the measures) to be classified falls on a centour for each group. The score vector is classified as belonging to that group where the centour on which it falls encloses the smallest percentage of the cases. This decision rule will result in the minimum number of misclassifications if the group dispersions are equal and the _a priori_ probabilities of group membership (number of cases which can be expected to fall into each group) are equal. If dispersions and/or _a priori_ probabilities cannot be assumed equal, a second decision function will result in fewer misclassifications.

The second decision rule offered by Cooley and Lohnes (1962) is based on Bayes's theorem and takes into account unequal dispersions and *a priori* probabilities. Application of this rule results in a probability statement for each score vector with respect to each group. The individual is classified as belonging to the group for which the probability of group membership is highest. Burch et al. (1966) applied this decision rule to EEG data gathered during the flight of the Gemini 7 astronauts and analyzed through period analysis into thirty measures. They achieved considerable success in classifying the record into ten states of consciousness. As with other Bayesian techniques, the choice of the *a priori* probabilities will affect the degree of misclassification. These probabilities may be adjusted to reflect the relative importance of misclassification of any one group at the expense, however, of overall success in classification. Although the decisions based on Bayes's theorem result in the fewest number of misclassifications overall, groups with low *a priori* probabilities will be underassigned by this method. Depending upon the relative importance of misclassifications in the groups, the probabilities may be adjusted to correct for this tendency.

To summarize the rationale for the present study, it was proposed that through period analysis six variables could be measured in the EEGs of sleeping subjects which would allow the classification of those patterns which occurred into five states of consciousness corresponding to the Dement-Kleitman sleep stages. The goal of the study was to evaluate such a system as a potential model for on-line, real-time classification of EEG sleep patterns at a reasonable cost, relieving the human classifier of this task. It was hypothesized that the system would: 1) be, within limits, as reliable as are several clinicians among themselves in classifying sleep stages; 2) be more sensitive to changes than the clinicians and; 3) maintain its reliability across several nights on one subject and across several subjects.

CHAPTER II

METHOD

Subjects. The subjects were four male students from the University of Oklahoma, ranging from 21 to 24 years of age.

Apparatus. The EEG was recorded on a Grass Model 6 electroencephalograph. The electrodes were Grass E5G gold-plated cup, surface type, suspended in a bentonite paste and covered with a gauze patch saturated in collodion. The standard left parietal to left occipital and left central to left frontal bipolar arrangements were used for electrode placements. For two of the subjects the anterior EEG signals were recorded on an Ampex FR-1300 F. M. analog tape recorder. In addition to writing out the EEG signals on paper, the posterior signals were subjected to an on-line period analysis and the frontal signals were reproduced from tape and analyzed the next day. The signals to be analyzed were fed to a Biophysical Model 102A analog to pulse width converter which produced square-wave trains corresponding to the baseline crossings of the analog signal (major period) and the baseline crossings of the analog first derivative (intermediate period). These square-wave trains were then fed to a Biophysical Model 909 digital data processor which employs six digital filters, three each for the major and intermediate periods, set as indicated in Table 1. The output pulses from the filters are accumulated during a ten-second epoch and are then fed to a digital to analog conversion circuit which varies its output voltage in relation to the number of pulses during the epoch. These voltages were used to drive three dual-channel Texas Instruments recti/riters and appeared as histograms of counts per ten-second epoch for each of the six bands. The six output records were visually interpreted and punched on standard IBM cards for subsequent analysis by an IBM 1620 digital computer.

Table 1

Digital Filter Settings in the Biophysical Model 909

Period	Band	Activity	Frequency Low	Frequency High	Total Count
	M1	Delta	0.5	2.5	20
Major	M2	Theta	3.0	6.8	68
	M3	Alpha	8.0	12.0	100
	I1	Alphoid	7.0	11.0	90
Intermediate	I2	Spindle	11.0	15.0	110
	I3	Beta	18.0	30.0	230

Procedure. The subjects were instructed to appear at the laboratory at 8:30 p.m. and soon after they arrived electrode application was begun. In addition to the four electrodes mentioned earlier, the subjects wore left and right eye electrodes, a forehead eye reference electrode, and a forehead grounding electrode. They further word EKG (electrocardiograph) disk electrodes at the back of the neck and on the left side, an elastic strain gauge around the lower thoracic region, and disk skin potential electrodes on the lower leg and palmar surface of the foot. When all electrodes were in place and had been checked, the subject was allowed to get into bed, the eye leads were checked for correct polarity, the lights were turned out, the door was closed, and the subject was allowed to fall asleep. The sleep room was air-conditioned with a moderate degree of soundproofing. The subjects slept from about 10:00 p.m. for eight hours. They were then awakened, told to rest quietly for five minutes of waking baseline, the electrodes were removed, and the subjects were allowed to leave. During the night the equipment was constantly monitored to assure that it was functioning reliably. All records used in the present study were obtained from subjects who had slept under the same conditions for several nights previous to the analysis and were not under the influence of drugs or alcohol.

The EEG records from the Grass Model 6 were scored, minute by minute, by three staff members, one of them an assistant assigned to the scoring and the other two senior research staff members. The scoring was done independently. The records for computer analysis were chosen by one of the senior staff members on the basis of the technical quality and lack of unusual patterns. Three records were chosen initially where the period analysis was available only for the posterior record. Two of the records were from two different subjects while the third record was from a second night on one of the subjects. After viewing the results of the classification procedure with these records, two additional records from two different subjects on whom frontal analyses were available were chosen. For each of the chosen records, the same senior staff member went through the record and picked those minutes which he felt were "classic" examples of each of the five states of consciousness into which the record was to be classified, awake, stage 1 (emergent or REM), stage 2, stage 3, and stage 4. An attempt was made to choose as many of these "classic" minutes for each category as would correspond proportionally to the amounts of these states in the total record since the proportion of cases in each "classic" group formed the a priori probabilities for most of the subsequent classifications.

The cards containing the scores for the six frequency bands, one card for each ten-second epoch, were processed by the IBM 1620 computer to yield cards containing six means, one for each band, for each minute. These minute cards were then used in all further analysis. The cards for those minutes defined as "classic" minutes were used as input to the discriminant function analysis program.

The multiple discriminant function analysis program used to reduce the test space and to examine the groups (states of consciousness) with respect to the six variables was based on the program given in Cooley and Lohnes (1962), modified to yield the F-ratios testing the significance of group differences on the six variables. Some reprogramming was also necessary to adapt the original program to the 1620 computer. The results of the discriminant function analysis were used in computing the group centroids and dispersion matrices in the reduced discriminant space by means of the program RSPACE (Cooley & Lohnes, 1962). The resulting centroids and dispersion matrices were then available for input to the classification program.

Classification of all available minutes of the record was accomplished by means of the program CLASSIF (Cooley & Lohnes, 1962), modified so as to avoid output of the centour information in the form of Chi-squares as originally written. The listed output consisted of the minute number, the titles of the three most likely groups, ordered from most to least likely, and the probabilities associated with those groups.

As mentioned, the _a priori_ probabilities given to the computer for the classification were based on the number of minutes chosen as "classic" for that record. In only one case was this not true, that where the classification of both of the frontal records was based on discriminants developed from the combined "classic" minutes for both records. In this case only the numbers of minutes in each group for the first of the two records were used since it was felt that the others were non-representative of the night as a whole. The entire eight-hour record was classified for all three of the posterior records. Tape footage limitations forced the classification of the records for the two subjects with anterior data only for the first 356 and 372 minutes.

The number of discriminants available for computation of the group centroids and dispersions, that is, the dimensionality of the reduced space, was limited to four, one fewer than the number of groups. An examination of the percentage each root (discriminant contribution) was of the trace (total discriminating power of the analysis) provided an indication of the number of discriminants necessary to account for most of the discriminability. In all of the analyses, the first three discriminants proved to contain all but a very minor portion of the total and therefore only the first three were used to compute the statistics relevant to the several spaces of reduced dimensionality.

The first record to be subjected to a discriminant function analysis and classified was one of the three for which only posterior data was available. Upon assurance that the technique would classify this record, a similar procedure was carried out with one of the two records from the other subject with posterior data only. To investigate the generality of the procedure, all of the "classic" minutes for both records were then combined, a discriminant analysis performed, and all minutes for both records were classified on the basis of these combined discriminants. The second night for the one subject was then classified on both the discriminants developed from the data on that subject's first night and the discriminants resulting from the combination of the "classic" data for the two subjects to see if general properties existed both across nights and across subjects. In light of the potentially greater variability in the intermediate bands with frontal records, evident from an examination of the band raw score means for the five groups, discriminant function analyses were run for frontal data from two additional subjects on whom such data was available. All available minutes for these two records were then classified on discriminants developed from combined "classic" data as well as on discriminants developed from each subject's own "classic" data.

Each minute for each record, then, was classified by three independent clinical scorers and by two sets of discriminants. The results were listed together, minute by minute, for each record and each classification compared to every other classification within each record. A $\underline{k}$ coefficient of nominal agreement (Cohen, 1960) was computed for each resulting contingency table. The $\underline{k}$ coefficient indicates the proportion of total agreement between any two judges after chance agreement has been removed. The statistic is designed for use in situations where no criterion of "correctness" exists and no restriction can be placed on the marginal distributions. Discrepancies between judges are considered equal to one another in that categories are not arranged in any particular order. Unlike the Chi-square contingency test, the k coefficient does not deal with the off-diagonal frequencies, that is, with the frequencies of disagreement. When chance agreement is equal to observed agreement the coefficient is zero. When perfect agreement occurs between judges the coefficient is a positive 1.00. This perfect agreement can occur only when all cases fall on the diagonal and this, in turn, requires that the marginal distributions for the two judges are identical. The maximum value which k can assume for a given set of marginals, k-max, indicates the limitations imposed by the judges' lack of agreement on the distribution of the cases into categories. The ratio of k to k-max serves to describe the proportion of marginally permitted agreement realized between the judges. In most cases, however, any disagreement has negative consequences.

CHAPTER III

RESULTS

<u>The results of the discriminant analysis</u>. In as much as the purpose of the multiple discriminant function analysis was to create vectors on which the five groups were maximally separated, it is of interest to see how well the analysis fulfilled this purpose. The first step was to explore the total set of raw score differences for discriminability between the groups. Table 2 lists the F-ratios for each of the six analyses performed. For this table and all succeeding tables in this chapter, the following abbreviations will be used: Posterior 1 (P1), Posterior 2 (P2), Combined Posterior (Comb. P), Anterior 1 (A1), Anterior 2 (A2), Combined Anterior (Comb. A), First Night (N1), Second Night (N2). In every case, the hypothesis, that group differences as large as those obtained would result from drawing five random groups from a six-dimensional multivariate swarm, could be rejected at much better than the one-percent level. The groups could not be regarded as having similar profiles on the six bands. It was also of interest to examine the significance of the differences between group means for each variable individually to see if all variables were contributing significantly to the discriminability. Table 3 gives these individual F-ratios for each of the six analyses. All six of the variables in each analysis are significant at greater than the

one-percent level. Appendix A contains the group means and standard deviations for the six variables for each analysis.

Table 2

F Tests for the Total Discriminability of the Measures

		Degrees of Freedom	
Analysis	F-ratio	Between	Within
P1	79.16	24	647
P2	64.02	24	797
Comb. P	113.91	24	1477
A1	123.85	24	859
A2	75.45	24	455
Comb. A	143.74	24	1348

It was contended earlier that an anterior electrode placement would yield greater variance in the measures than would posterior placement. To test this contention, F-ratios were calculated between the total variances for the combined posterior "classic" minutes and the combined anterior "classic" minutes for each of the six bands. Table 4 displays the variances and F-ratios. All of the F-ratios were significant at better than the one-percent level. The anterior variances for all bands except M3, the alpha band, were larger in the anterior minutes than in the posterior minutes. The greater variance in alpha with a posterior placement is not surprising in light of the fact that waking alpha is severely attenuated in anterior records.

Table 3

Individual F-Ratios for the Six Bands

Analysis	Band	Among Mean Square	Within Mean Square	F-ratio
P1	M1	4697.22	26.16	179.52
	M2	1373.27	8.69	158.03
	M3	2370.46	7.72	307.15
	I1	981.43	16.51	59.45
	I2	215.74	4.29	50.25
	I3	939.77	7.95	118.23
		degrees of freedom = 4 & 190		
P2	M1	3988.89	13.30	299.87
	M2	686.36	7.06	97.28
	M3	1300.63	14.53	89.53
	I1	851.47	14.80	57.54
	I2	307.60	11.86	25.94
	I3	639.96	9.41	68.00
		degrees of freedom = 4 & 233		
Comb. P	M1	8796.07	20.59	427.28
	M2	1904.23	9.82	193.97
	M3	3575.65	16.41	217.85
	I1	1911.36	35.53	53.78
	I2	553.66	10.21	54.24
	I3	1593.55	9.42	169.07
		degrees of freedom = 4 & 428		

Table 3 (continued)

Analysis	Band	Among Mean Square	Within Mean Square	F-ratio
A1	M1	11475.29	23.79	482.27
	M2	1801.30	12.08	149.08
	M3	1039.86	9.02	115.31
	I1	2157.17	6.77	318.80
	I2	2509.95	20.96	119.77
	I3	6866.69	25.55	268.78
	degrees of freedom = 4 & 251			
A2	M1	2109.32	14.18	148.74
	M2	427.82	9.37	45.65
	M3	221.31	6.19	35.73
	I1	5502.55	13.72	400.98
	I2	1513.25	8.24	183.70
	I3	3476.09	9.19	378.16
	degrees of freedom = 4 & 135			
Comb. A	M1	12741.94	28.81	442.29
	M2	2053.23	14.79	138.84
	M3	1055.70	12.06	87.55
	I1	7051.26	42.33	166.55
	I2	4003.76	25.89	154.62
	I3	10448.28	38.26	273.06
	degrees of freedom = 4 & 391			

Each analysis resulted in four orthogonal discriminants. To determine how many of these discriminants to use as a basis for classification, it was necessary to examine the percentage of the total discriminating power contained in each discriminant. These percentages are listed by the program as "Percentage which each root is of trace" and appear as Table 5. It is clear that consideration of more than the three discriminants with the highest percentages would add little to the classification since in every case the first three account for more than ninety-nine percent of the total discriminability.

Table 4

Standard Deviations, Variances, and F-Ratios for the Six Bands on Combined Posterior and Combined Anterior "Classic" Minutes

	Variance			
Band	Posterior	Anterior	F-ratio	p
M1	101.81	157.50	1.55	$<.01$
M2	27.35	35.40	1.29	$<.01$
M3	49.42	22.66	2.18	$<.01$
I1	52.85	113.21	2.14	$<.01$
I2	15.21	66.10	4.35	$<.01$
I3	24.11	143.52	5.95	$<.01$

degrees of freedom = 432 posterior and 395 anterior

Table 5

Percentage Which Each Root is of Trace

	Discriminants			
Analysis	1	2	3	4
P1	54.09	33.84	11.76	.29
P2	73.94	18.79	7.10	.15
Comb. P	59.99	32.27	7.58	.14
A1	65.46	27.44	6.81	.27
A2	74.57	20.39	4.32	.69
Comb. A	74.55	17.80	7.22	.41

Often a discriminant function analysis is performed primarily to explore the differences which may arise. In such a case it is of greatest interest to examine the relative contributions of the measures to each of the discriminants. These relative contributions appear as scaled vectors and are exhibited as Table 6. While it is not the purpose of this paper to examine these differences in detail, some of the information contained in the scaled vectors may help to shed light on certain of the classification difficulties which were encountered.

An examination of the group centroids in the three-dimensional discriminant space (Table 7) is especially important in explaining any lack of success in classification. In the posterior analyses, for example, the stage 3 and stage 4 centroids are not particularly different from each other while in the anterior analyses the centroids are well separated. One could thus expect difficulty in differentiating the two stages on posterior classifications and not on anterior classifications.

Table 6

Scaled Vectors Showing the Relative Contributions of the Bands to the Three Chosen Discriminants for the Six Analyses

		Band					
Analysis	Discriminant	M1	M2	M3	I1	I2	I3
P1	1	- 2.14	.59	35.09	7.23	-10.66	2.05
	2	- 9.02	28.38	0.00	4.99	3.11	26.80
	3	8.66	- 5.70	2.12	.17	-23.49	20.76
P2	1	28.84	-16.50	-36.78	19.84	11.10	- 3.95
	2	- 3.87	-26.22	15.09	19.88	-31.36	9.21
	3	2.75	22.14	4.00	27.40	-35.01	8.46
Comb. P	1	-42.03	26.35	45.77	12.50	-16.24	32.70
	2	3.95	45.56	-30.71	-35.18	34.16	- 9.22
	3	-31.45	-10.89	-18.33	- 3.39	45.26	-37.13
A1	1	-16.74	17.81	20.38	-28.84	26.09	17.07
	2	9.46	11.27	-18.06	12.67	49.35	39.31
	3	.94	11.67	19.84	-25.79	22.53	-43.31
A2	1	22.39	7.85	- 7.37	26.51	10.19	-13.70
	2	- 1.19	10.16	5.41	-17.84	25.49	-12.51
	3	- 4.03	- 8.00	11.23	-17.99	-12.94	-24.00
Comb. A	1	29.92	-35.01	-41.01	71.37	-19.57	- 9.59
	2	2.87	- .15	-31.43	26.28	73.67	56.29
	3	13.72	13.11	-38.18	62.83	16.57	75.25

Table 7

Centroids of the Five Groups in Reduced Discriminant

Space for the Six Analyses

		Groups				
Analysis	Discriminant	Awake	Stage 2	Stage 3	Stage 4	REM
P1	1	35.69	8.77	3.57	2.12	9.12
	2	28.56	40.94	28.27	24.47	45.52
	3	- 5.65	- 9.11	- 5.98	- 3.51	- 2.27
P2	1	-17.85	- 8.03	11.35	19.08	-15.93
	2	- 4.31	-20.86	-17.82	-15.71	-17.77
	3	5.96	11.17	13.80	14.03	16.07
Comb. P	1	34.19	22.90	5.45	.37	29.58
	2	- 3.21	20.07	15.96	14.21	18.71
	3	- 9.88	-10.78	-13.62	-15.00	-16.72
A1	1	22.56	18.34	6.88	-10.03	23.51
	2	18.99	37.62	38.24	34.19	36.29
	3	- 5.32	- 5.18	- 5.53	- 9.97	-13.79
A2	1	- .46	25.26	42.54	50.16	7.50
	2	4.06	18.48	11.43	5.21	9.04
	3	-32.94	-37.02	-37.70	-37.45	-40.75
Comb. A	1	-22.37	-16.26	- 2.84	11.66	-21.67
	2	15.00	28.58	28.21	24.88	26.63
	3	17.51	18.81	20.23	23.74	28.91

The classification results were examined to shed light on the three major hypotheses concerning the system: 1) that it would, within limits, be as reliable as would the three clinical scorers in classifying sleep stages; 2) that it would be more sensitive to changes than the clinicians and; 3) that it would maintain its reliability across nights on a subject and across subjects.

The first hypothesis. The first of these hypotheses would be supported if agreement as to the classification of the minutes was as great between the clinicians and the computer as among the clinicians themselves. The first results which were examined were those indicating the computer classification of the "classic" minutes for each record. The percentages of correct classification (correct, that is, on the assumption that all "classic" minutes are actually members of the group to which the clinician assigned them) are given in Table 8. The number of minutes in each group and the associated a priori probabilities are given in Table 9. Seven of the ten overall percentages are above 90. The other three are above 80 percent. On certain of the minutes where the two classifications did not agree, furthermore, the clinician felt compelled to admit, upon reexamination of the record, that the computer classification was probably more representative than his own. For the anterior records, individual stage percentages are all greater than 75, with all but one greater than 80. For the posterior records, on the other hand, disregarding the second night, four percentages were less than 65. These low percentages occur within the slow-wave stages 3 and 4. The anterior data seems to provide a better basis for classification of this slow-wave sleep than does the posterior data.

Table 8

Percent of Computer Agreement with "Classic" Assignation for the Ten Classifications Carried Out

		States					
Minutes	Discriminants	Awake	Stage 2	Stage 3	Stage 4	REM	Overall
P1	P1	100.0	95.0	82.8	63.6	88.9	90.3
	Comb. P	100.0	93.0	58.6	63.6	82.2	84.6
P2	P2	85.7	99.4	87.5	91.7	91.8	96.2
	Comb. P	100.0	95.5	100.0	58.3	87.8	92.4
P2(N2)	P2(N1)	90.0	96.0	100.0	57.1	92.8	92.1
	Comb. P	80.0	85.5	100.0	15.0	91.3	81.9
A1	A1	100.0	85.6	90.5	100.0	91.4	90.2
	Comb. A	100.0	79.2	81.0	100.0	93.1	86.7
A2	A2	91.7	96.9	88.2	95.2	96.0	95.0
	Comb. A	83.3	95.4	88.2	90.5	84.0	90.7

Turning to the results of the classifications for all of the minutes in a record, the _k_ coefficients, _k-max_ coefficients, and the ratios between the two coefficients are given in Tables 10-14. The contingency tables from which the statistics were derived appear in Appendix B.

Table 9

Number of Minutes in the "Classic" Groups and A Priori Probabilities*

Record	Awake	Stage 2	Stage 3	Stage 4	REM	Total
P1	10(.05)	100(.51)	29(.15)	11(.06)	45(.23)	195(1.00)
P2	7(.03)	154(.65)	16(.07)	12(.05)	49(.20)	238(1.00)
Comb. P	17(.04)	254(.65)	45(.10)	23(.05)	94(.22)	433(1.00)
P2(N2)	10	173	5	21	69	278
A1	16(.06)	125(.49)	21(.08)	36(.14)	58(.23)	256(1.00)
A2	12(.09)	65(.46)	17(.12)	21(.15)	25(.18)	140(1.00)

*Anterior 1 _a priori_ probabilities used in combined anterior classification.

A summary of Tables 10-14 is given in Table 15 where the mean _k_ coefficients, _k-max_ coefficients, and ratios for each record and over all records are displayed. All of the _k_ coefficients are significant at greatly beyond the chance level, as might be expected. It can be seen that the overall _k_ mean between raters (.783) is somewhat higher than that between the raters and the computer classifying on the basis of discriminants developed from the "classic" minutes for the same record (own discriminants, .711). The two means are closer, however, than one might expect. If one omits the means for the Posterior 2, second night, the overall means become .779 between raters and .719 between raters and the computer; the computer-rater agreement approaches even more closely the between-rater agreement. The values of marginal agreement, _k-max_, are even more nearly the same (.896 and .892). If the second night for Posterior 2 is again omitted, the values become .895 between raters and

Table 10

The k and k-max Coefficients and Ratios for Posterior 1 Minutes

Statistic	Judge	Rater 2	Rater 3	Discriminants P1	Discriminants Comb. P
k	Rater 1	.806	.730	.708	.632
	Rater 2		.769	.728	.678
	Rater 3			.678	.676
	P1 Discrim.				.810
k-max	Rater 1	.954	.938	.940	.963
	Rater 2		.955	.954	.924
	Rater 3			.935	.918
	P1 Discrim.				.903
Ratio	Rater 1	.845	.778	.753	.656
	Rater 2		.805	.763	.734
	Rater 3			.725	.736
	P1 Discrim.				.897

.894 for the computer-rater mean, equal for all intents and purposes. The ratios, as might be expected, differ in favor of the raters as do the k coefficients. Again, omitting the second night, the means become .870 between raters and .804 between raters and the computer own discriminants.

Agreement between the raters and the computer own classification is greater for the two anterior records than for the posterior records.

Table 11

The _k_ and _k-max_ Coefficients and Ratios for

Posterior 2 Minutes, First Night

Statistic	Judge	Rater 2	Rater 3	Discriminants P2	Discriminants Comb. P
_k_	Rater 1	.948	.668	.588	.511
	Rater 2		.696	.608	.534
	Rater 3			.730	.707
	P2 Discrim.				.830
k-max	Rater 1	.955	.796	.799	.721
	Rater 2		.839	.842	.764
	Rater 3			.942	.895
	P2 Discrim.				.906
Ratio	Rater 1	.993	.839	.736	.709
	Rater 2		.830	.722	.699
	Rater 3			.775	.790
	P2 Discrim.				.916

Two variables may account for this difference. It may be that the greater variability in the measures with the anterior placement leads to better classification by the computer. It is also likely, however, that a technical artifact of some importance can account for at least part of the difference. Synchronization of the analog record with the epochs and minutes which appear as output from the period analyzer was enhanced

Table 12

The k and k-max Coefficients and Ratios for

Posterior 2 Minutes, Second Night

Statistic	Judge	Rater 2	Rater 3	Discriminants P2	Discriminants Comb. P
k	Rater 1	.774	.812	.662	.540
	Rater 2		.818	.712	.621
	Rater 3			.672	.598
	P2				.677
k-max	Rater 1	.864	.886	.873	.676
	Rater 2		.954	.887	.804
	Rater 3			.888	.778
	P2				.726
Ratio	Rater 1	.896	.916	.758	.799
	Rater 2		.857	.803	.772
	Rater 3			.757	.769
	P2				.932

through an automatic record marker just prior to the gathering of the anterior data.

The second hypothesis. The second hypothesis, that the system would be more sensitive to changes than would the clinicians, was evaluated by counting the changes within each record, both for the raters and for the two computer classifications. These numbers are displayed in

Table 13

The k and k-max Coefficients and Ratios for Anterior 1 Minutes

				Discriminants	
Statistic	Judge	Rater 2	Rater 3	A1	Comb. A
k	Rater 1	.833	.862	.738	.710
	Rater 2		.724	.804	.755
	Rater 3			.693	.688
	A1				.886
k-max	Rater 1	.871	.939	.863	.842
	Rater 2		.820	.948	.894
	Rater 3			.804	.790
	A1				.938
Ratio	Rater 1	.956	.918	.855	.843
	Rater 2		.883	.848	.845
	Rater 3			.862	.871
	A1				.945

Table 16. For each of the records and on the means for all the records the number of computer changes is greater. It is interesting to note that rater 1, the less experienced assistant, showed the fewest changes, both on the individual records and overall. Perhaps the experienced rater is more willing or able to look at the separate minutes apart from their surroundings.

The third hypothesis. The third hypothesis, that the system

Table 14

The k and k-max Coefficients and Ratios for Anterior 2 Minutes

Statistic	Judge	Rater 2	Rater 3	Discriminants A2	Discriminants Comb. A
k	Rater 1	.764	.733	.702	.596
	Rater 2		.814	.845	.686
	Rater 3			.803	.700
	A2				.770
k-max	Rater 1	.888	.862	.829	.875
	Rater 2		.919	.940	.875
	Rater 3			.936	.866
	A2				.845
Ratio	Rater 1	.860	.850	.847	.681
	Rater 2		.886	.899	.784
	Rater 3			.858	.808
	A2				.911

would maintain its reliability across nights and across subjects, was examined separately across nights and across subjects. Support for the contention that generalization across nights was possible would be indicated if the classification of minutes for the second night could be accomplished as accurately from discriminants developed on data from the first night as could minutes for the first night. The overall percentage of computer agreement with the "classic" assignation for the posterior 2

Table 15

Mean k and k-max Coefficients and Ratios Between Raters and Between Raters and Discriminant Ratings for Each Record and Over All Records

Statistic	Record	Raters	Discriminants: Between Raters and Own	Discriminants: Between Raters and Combined
k	P1	.768	.705	.662
	P2 (N1)	.771	.642	.584
	P2 (N2)	.801	.682	.586
	A1	.806	.745	.718
	A2	.770	.783	.661
	Overall	.783	.711	.642
k-max	P1	.949	.943	.935
	P2 (N1)	.863	.861	.793
	P2 (N2)	.901	.883	.753
	A1	.877	.872	.842
	A2	.890	.902	.872
	Overall	.896	.892	.839
Ratio	P1	.809	.747	.709
	P2 (N1)	.887	.744	.733
	P2 (N2)	.890	.773	.780
	A1	.919	.855	.853
	A2	.865	.868	.758
	Overall	.874	.797	.767

Table 16

The Number of Changes Within Each Record for the Three Raters and for the Two Computer Classifications

Record	Rater 1	Rater 2	Rater 3	Computer 1	Computer 2
P1	55	73	74	98	105
P2(N1)	57	60	91	119	128
P2(N2)	43	73	60	114	151
A1	34	45	35	66	66
A2	37	56	59	62	94
Mean	45.2	61.4	63.8	91.8	108.8

second night (Table 8) is 92.1, an acceptable degree of accuracy. The accuracy of classification of stage 4 sleep on these discriminants is relatively low, however, only 57.1 percent. Partial explanation for this low percentage may come from an examination of the relative number of cases in the stage 4 group for the first night's "classics" when compared to the second night (Table 9). Over 7.5 percent of the cases for the second night fall in the stage 4 category while only 5 percent fall in the stage 4 category for the first night. Since groups with low *a priori* probabilities tend to be underassigned in the first place, it is likely that some of the minutes in the "classic" stage 4 group for the second night had classification probabilities which were greater for another stage of sleep when classified on the basis of the *a priori* probabilities for the first night. When classified on the combined posterior discriminants, the second night exhibited 81.9 percent correct overall classification. When it is considered that this record was classified on discriminants from two other nights, this percentage is quite respectable. Again, the correct classification of stage 4 sleep is low, an unacceptable 15.0 percent. The comments regarding a priori probabilities made earlier also apply here. The difficulty appears to be in differentiating stage 3 from stage 4. If these two categories are grouped together as "slow-wave sleep" for this classification the percentage of correct classifications within this new group jumps to 100.0 and the overall percentage to 88.1.

Turning to the k and k-max coefficients and the ratios, it is clear, with regard to reliability across nights, that the agreement between the raters and the two computer classifications (.682, .586) is considerably less than the agreement between the raters (.801) (Table 15). It is interesting to note, however, that both in the case of the k coefficients and in the case of the ratios the agreement between the raters and the computer classification on the first night's discriminants is greater than the corresponding value for the first night's minutes, as it is for the agreement between the raters and computer classification on combined posterior discriminants. It would appear, then, that agreement on the second night was at least as good as on the first night. Between-rater agreement for the second night was higher than for the first night, however.

The contention that general discriminants could be developed across subjects would be supported if classification were as accurate on combined discriminants as on the record's own discriminants. The percentages of correct classification on the "classic" minutes (Table 8) indicate that while no set of "classic" minutes was classified as accurately with the combined discriminants as with the record's own discriminants, all of the combined classifications were better than 80.0 percent and with the exception of the second night on posterior 2, none are further than ten percentage points away from those for the record's own classification.

The agreement coefficients are all lower between raters and combined classifications than between the raters themselves. The combined classifications also agree less with the raters than do classifications based on the record's own discriminants. Figures 1 and 2 illustrate the best and the worst rater-computer agreement achieved within the study. Figure 1 presents the sleep stage profiles, smoothed over five-minute blocks, for rater 2 and for the computer own classification on anterior 2. Figure 2 presents similar profiles for rater 1 and combined classification on posterior 2, first night.

Agreement between the two computer classifications ranges from .677 to .886, the lowest value resulting from the two classifications of the second night on the second posterior subject. The mean agreement between the two computer classifications is .795, better than the mean between-rater agreement. The mean <u>k</u> to <u>k-max</u> ratio between the two computer classifications is .920, perhaps indicating that much of the disagreement is in the marginal distributions.

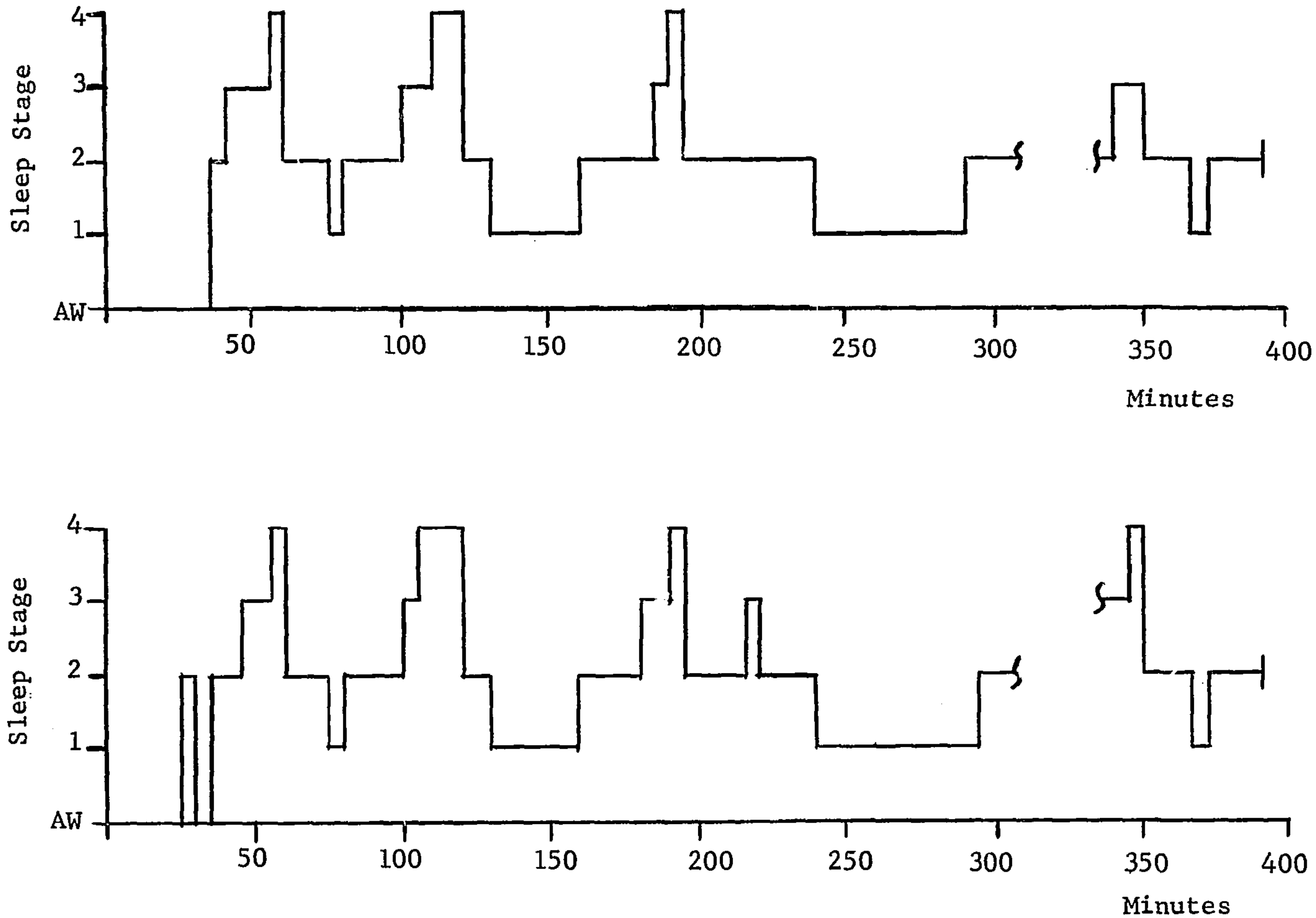

Fig. 1. Smoothed sleep stage profiles for the anterior 2 record. Upper half: rater 2. Lower half: computer own discriminants.

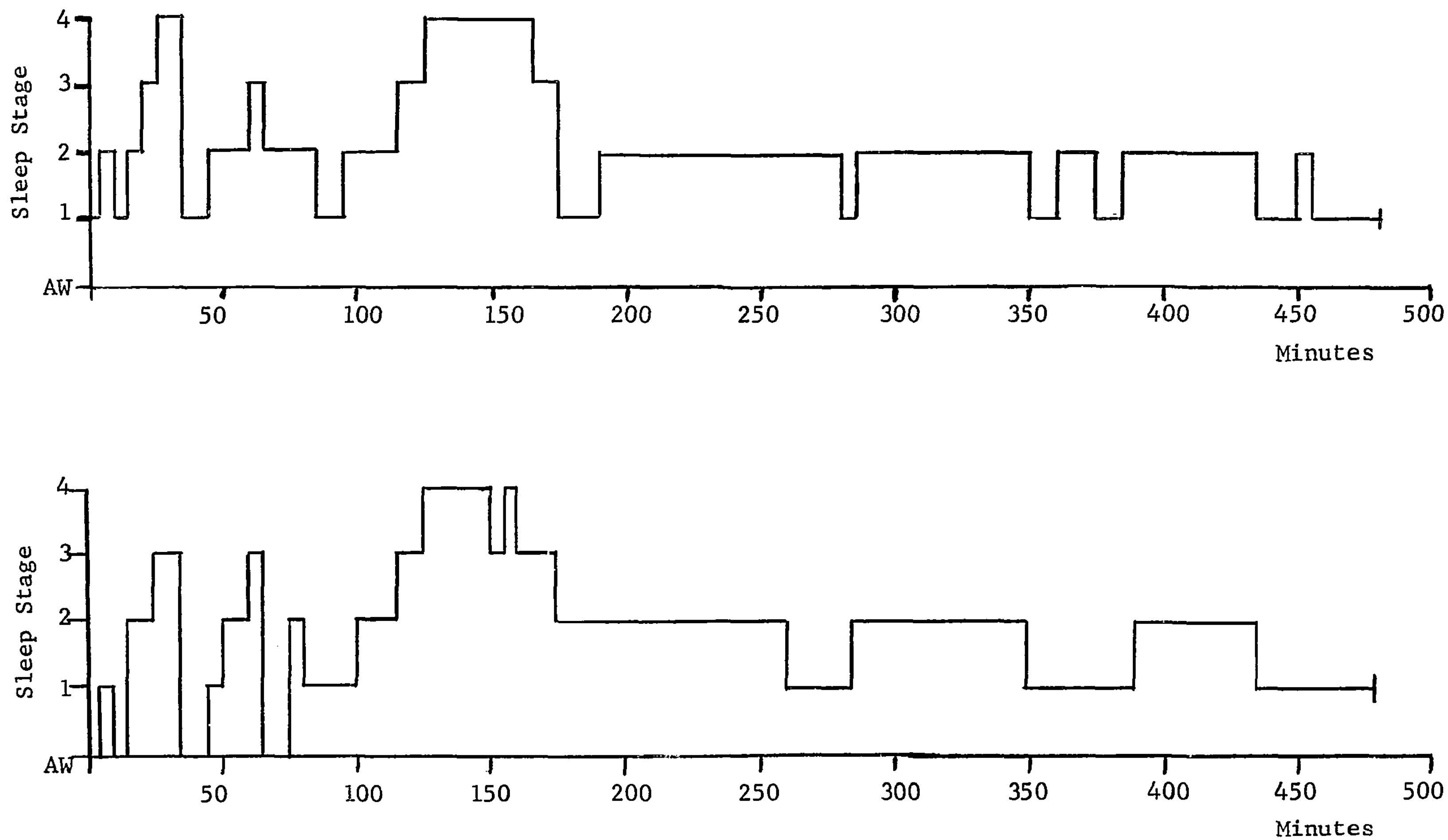

Fig. 2. Smoothed sleep stage profiles for the posterior 2 record, second night. Upper half: rater 1. Lower half: combined classification.

CHAPTER IV

DISCUSSION

The results of the discriminant output indicate that sufficient differences exist between the five sleep stages on the six measures to generate significant separation. Each of the measures taken added significantly to the total effectiveness of the analysis, a fact which argues strongly for retention of all six bands. The question of whether or not six bands alone are sufficient to accurately classify the record remains open. The classification achieved with the system, while surprisingly good, would be improved in reliability somewhat with the addition of parameters such as total counts of baseline crossings for the major, intermediate, and minor periods. There is some doubt, however, as to whether the addition of other measures would produce an increase in reliability commensurate with the added expense of obtaining and analyzing such measures.

Both the increased variability and the increased reliability gained by using the anterior electrode placement raises a question concerning the usual procedure of analyzing posterior EEG. In light of the findings, it is suggested that anterior EEGs should be used as a basis for classification, particularly when period analysis is employed. If possible, the posterior record can be filtered separately to obtain the alpha frequency information and this posterior alpha count used in place of that resulting from anterior placement. This posterior information would be of particular importance where the detection of waking in the subject is a primary consideration such as studies which stimulate sleeping subjects.

The classification results indicate that the computer can duplicate the clinical process of scoring sleep records to a large extent. Not only were the obtained reliabilities much better than a random agreement, but they closely approached the reliabilities to be expected from clinicians themselves. It must also be remembered that the clinical raters employed in the study were working in the same laboratory and that their agreement is probably greater, therefore, than it would be for raters employed in diverse settings. The results lend considerable support to the hypotheses concerning the reliability and usefulness of the data reduction and classification system.

While the raters and the computer disagree to a certain extent on a minute-to-minute basis, so do the raters themselves. The problem of reliability is tied up in the question of whether or not the automatic classification ought to exactly match the clinical classification. It is possible, particularly when the computer seems more willing to change stages than does the human scorer, that the computer classification more closely mirrors the state of the subject. The greater number of minute-to-minute changes in the computer classification, when compared to clinical classifications, leads to the conclusion that the computer system is sensitive to changes in the data which are not detected by the human scorers. While some computer changes may result from the introduction of artifact such as subject movement into the record, changes of the sort occurring between stages 3 and 4 or even between stages 2 and 3 are not likely to be the result of such artifact.

The classification procedures used by the computer judge each minute as a separate entity, apart from the surrounding record. The minute cards could be presented for classification in any random order with the same resulting classifications. Such a procedure, however, when applied to the human interpreter, would result in somewhat different classifications for the minutes than those which would be produced from a continuous record. Such a study should be undertaken to evaluate the degree to which the surrounding record influences the clinician's judgments regarding sleep staging.

It is entirely possible to program into the system a probability based on a form of the Markov process in which the system would take into account the classification of the minute, or several minutes, preceding the one being classified. In order to fully duplicate the clinical process, however, it would be desirable to have the weighting of such a probability reduced by a factor corresponding to the duration of the stage. Such a procedure would increase reliability at the expense of computer sensitivity to minor changes in the record. The decision regarding such an addition would depend to a large extent on the relative importance of detecting change.

The number of stage changes may itself serve as a measure in sleep research. Williams and Williams (1966), in a study of the relationship between EEG profiles and performance, counted transitions from stage to stage. Defining quiet and restless sleepers on the basis of differing transitional probability matrices, they found a significant difference between the two groups on this measure and found it related to performance under conditions of sleep deprivation. An automatic analysis of sleep data could lead to a finer discrimination between the groups and perhaps enable investigators to more thoroughly study the nature of sleep disturbances.

The selection of the "classic" minutes largely determines the reliability realized between a computer classification and that of a human scorer. The computer classification is no better than the degree to which the data on which the discriminants are developed is representative of the record one wishes to classify. While agreement between raters on the "classic" minutes was excellent overall, certain discrepancies occurred as the result of the clinician's choosing some minutes within each group which were "marginal" to increase the group dispersions slightly for the sake of generality across subjects. This increase in the group dispersions may have led to certain misclassifications as the product of overlapping group boundaries. It is clear that the percentages of correct classification of "classic" minutes show that a number of the minutes chosen were not representative of the groups into which the clinician placed them. To refine the group parameters, then, and reduce overlap between groups, these incorrectly classified minutes should be discarded. The method of comparing "classic" minutes to the groups from which they came is admittedly "bootstrapping." In any developmental context, however, such a procedure yields the information necessary for refinement of the classification. Ideally, of course, all minutes could be correctly placed in their "classic" groups. In a general system designed to accept data from any subject and classify it into stages, it is a good idea either to arrive at the parameters for a particular group from data on minutes agreed upon by a number of diverse clinical scorers and from a large number of subjects or, alternatively, to modify the parameters as the result of computer experience with classification after the manner of Shapiro and Fink (1966).

The *a priori* probabilities of group membership also influenced the agreement reached between automatic and human classifiers. The differing distributions of the minutes into the five states by the classifying parties, human and machine, indicate that this problem is as great as that posed by a minute-by-minute classification. Differing distributions force disagreement on a minute-to-minute basis. Especially where the automatic classification is designed to yield information as to the percentages of the total night spent in the various stages of sleep, such as in drug studies, differences in the distributions can become a central factor. An adjustment of the *a priori* probabilities can overcome differences between the computer classification and a clinical classification but radical adjustment would lead not only to a loss of generalizability but also to a loss in minute-by-minute agreement. It is probably true that a general set of *a priori* probabilities based on an average over several subjects and several nights would lead to optimal long-term classification accuracy.

The special problems encountered in research on sleep deprivation or drug effects, where distributions differing from baseline probabilities are to be expected, could call for another set of probabilities for use on these occasions. This, in essence, is probably what the clinician, consciously or unconsciously, does when confronted with a sleep record that he knows resulted from some experimental manipulation. He forms an hypothesis about the changes in the record which he expects to see and he is likely to err in his scoring of this hypothesis. The solution in clinical settings has been to use a blind scoring technique. An automatic analysis using altered probabilities would only lend human biases to a system. A conservative approach to the problem would suggest that significant changes in stage distributions observed while using baseline probabilities would constitute a powerful argument in favor of an hypothesis of change.

The problem of the generalizability of any set of discriminants used for classification of EEG patterns, either across subjects or across nights, is tied not only to within-night reliability but to the between-record reliability of the human classifiers and their scoring strategies as well. There is some evidence, both anecdotal and empirical, to indicate that two types of individual scoring strategies may exist. One type of individual has an internalized concept of how a particular stage of sleep should appear and classifies epochs as that stage only to the extent that they meet this internal criterion. The other type of individual uses a "rubber ruler" in that he has a preconceived idea of the stages that should occur and when they should occur and classifies the most likely epochs as members of the respective groups. Monroe's (1967) findings lend support to the contention that two strategies exist. For the twenty-seven raters employed in that study, the number of minutes classified as stage 4 sleep ranged from none to 83.0. Some raters, in other words, saw little or no stage 4 sleep while others saw as many as 83.0 minutes. Similar, though less striking, differences occurred for other categories.

It may be desirable to standardize criteria for sleep record scoring to a much greater extent than has so far been accomplished. It is sometimes difficult to interpret the results of studies using sleep profiles without a clear statement of the particular set of criteria used in scoring. It would appear that if an exact and standard procedure could be applied to all sleep records, a considerable gain in clarity and a corresponding reduction in the amount of supplementary information in reports could be realized. One approach to such a standardization would be to apply precise period analytic procedures to the EEG record and submit the results to computer analysis and classification based on a set of universally available discriminants. The classification would then be regarded as the "true" profile for that record. There are several difficulties apparent in such a procedure. The first difficulty is the assumption that analytic facilities are universally available. A second and perhaps more important difficulty concerns the equivalence of patterns across subjects. The basic question is whether or not the slow, regular, and rolling delta activity of classic stage 4 sleep in one subject is the same, physiologically and behaviorally, as is a much more broken, random pattern resembling stage 3 for another subject for whom this state is his deepest. A computer system would classify the first instance as stage 4 and the second stage 3. In effect, then, the computer classification would indicate that the second subject showed no stage 4 sleep. What is needed before such a question can be resolved is a series of carefully done studies exploring the physiological and behavioral concomitants of specific EEG patterns. That this problem is far from simple is pointed up in the Williams et al. (1966) study on instrumental behavior during

REM sleep. It may be that the currently employed Dement-Kleitman classification scheme contains too few categories and that a greater number of behaviorally relevant states could be found. It may also be, on the other hand, that reducing classification to two sleeping states, slow-wave or low voltage, fast activity, will serve to differentiate all possible behaviors with the other stages simply gradations within these categories.

The system developed within this paper is capable of conversion into an on-line, real-time classification system. There are a number of ways in which this might be accomplished. The various approaches to the problem differ at the point of interface between digital and analog components of the system. Since period analysis is essentially a digital filtering system, the primary interface occurs at the point where the analog EEG signal is converted to square-wave trains. The next point at which a decision must be made is at the output of the digital counters. Here the digital output may be left in its BCD (Binary Coded Decimal) format or converted to an analog voltage through a resistance ladder. The digital data can either be analyzed with a special-purpose digital computer (an expensive proposition) or an on-line analysis could be performed by a general-purpose digital computer on a time-sharing arrangement with a remote data input connection. A better solution would seem to be the use of an analog computer programmed to weight the output voltages of the period analyzer to yield discriminant scores. These scores would then be further weighted by functions corresponding to group dispersions, a priori probabilities, and other factors such as transitional probabilities. At this point a graphic representation of the probabilities of each group could be recorded and interpreted visually or digital decision logic could be employed to identify the most likely group or groups.

CHAPTER V

SUMMARY

The study of sleep has become one of the principal fields of study in the last decade. Numerous researchers are using sleep as a tool to study psychological and physiological problems such as mental illness, drug effects, and performance. Although less than forty years old as a clinical technique, the electroencephalograph (EEG) has become a standard criterion for states of consciousness. Several schema for the labeling of EEG patterns during sleep have been developed, notably that of Dement and Kleitman. The classification of sleep records into categories has long been a job for persons with the necessary training to interpret the nocturnal patterns. Such classifications are less than perfect, not entirely reliable, and consume a great deal of time on the part of the individual. There has been a great deal of effort in recent years directed toward the development of automatic systems of analysis and classification of nocturnal EEGs. The majority of these techniques employ a general-purpose digital computer for all or a part of the data-reduction process. In light of the expense and delay inherent in the use of such a computer, this study sought to develop and evaluate a technique suitable for on-site, real-time analysis and classification of EEG sleep records into the Dement-Kleitman sleep stages. It was felt that the technique should be:

1) as reliable as are several clinicians among themselves in classifying records; 2) more responsive to change than is the clinician, and; 3) maintain reliability across subjects and across nights.

Several quantification procedures for expressing the EEG were discussed, among them analog filters, auto-spectral analysis, amplitude analysis, and period analysis. Due to considerations of analysis time, expense, and the greater responsiveness of period analysis to random, high-frequency activity, this method was chosen to quantify the EEG. The method consists, essentially, of a set of on-line, special-purpose digital filters and associated digital and analog circuitry.

The use of multiple discriminant function analysis can significantly reduce the dimensionality of the raw score space and provide weights for the measures which produce maximal separation between criterion groups. Based on the discriminant functions, classification can use a minimum-distance criterion or a maximum-likelihood criterion. The maximum-likelihood criterion, when a decision rule based on Bayes's theorem is employed, results in optimal classification if group dispersions and a priori probabilities cannot be assumed equal.

The subjects were four male students from the University of Oklahoma. They were run on separate nights with standard left parietal to left occipital and left central to left frontal electrode placements. For two subjects the anterior signals were recorded and analyzed the next day. The signals were analyzed by an on-line special-purpose period analyzer set to yield information about frequencies corresponding to the clinical waveforms of delta, theta, alpha, regular alpha, sleep spindles, and beta activity.

The EEG records were scored, minute by minute, by three independent raters. One rater chose "classic" minutes of each of the five stages used. These minutes formed the basis for the multiple discriminant function analysis which was performed on an IBM 1620 computer. All minutes for a night were classified both on discriminants developed for that night and on combined discriminants. A second night's record was classified on discriminants developed from the first night for one of the subjects. Contingency tables resulting from a comparison of each classification with every other classification for each record were tested for agreement by a coefficient designed for use with nominal scale data.

The results indicated significant separation between groups on each of the variables and overall. Anterior electrode placement was found to give greater variance in all measures except alpha activity.

The results further indicated that considerable agreement between clinicians and the computer classifications was possible. In all cases, greater than 80 percent of the "classic" minutes were correctly classified into their respective groups. Computer-rater agreement on the coefficients approached closely the agreement found among the clinicians themselves. A count of the number of changes in classification for each of the raters, human and computer, showed that the computer classification was probably more sensitive to change than were human raters. Generalization across nights and across subjects, while possible, did not result in reliability as high as that achieved within a night or subject. In general, however, the results indicate that the technique has sufficient potential to warrant further refinement.

The results were discussed in terms of increasing the reliability achieved in the study and adding to the generalizability across subjects and across nights, including non-baseline nights. Possible refinements included additional measures, the use of anterior electrode placements, the addition of transitional probabilities, and more representative methods of selecting "classic" minutes and _a priori_ probabilities of group membership. Consideration was given to the use of the number of stage changes as a measure of sleep. It was suggested that two scoring strategies may exist for clinicians classifying sleep records. One type of individual uses invariant criteria for stage patterns while the other adjusts his criteria for each subject according to expectations based on previous experience and experimental treatment. The feasibility of standardizing record scoring through automatic analysis and classification of sleep records according to a standard procedure was considered. The principal drawback was held to be the lack of sufficient knowledge concerning the physiological and behavioral concomitants of sleep stages. The system was then discussed as a model for an on-line, real-time system of automatic sleep stage classification with consideration of the digital-analog interface problem.

REFERENCES

Agnew, H. W., Jr., Parker, J. C., Webb, W. B., & Williams, R. L. Amplitude measurement of the sleep electroencephalogram. Electro-enceph. clin. Neurophysiol., 1967, 22, 84-86.

Aserinsky, E. Periodic respiratory pattern occurring in conjunction with eye movements during sleep. Science, 1965, 150, 763-766.

Aserinsky, E., & Kleitman, N. Regularly occurring periods of eye motility and concomitant phenomena during sleep. Science, 1953, 118, 273-274.

Aserinsky, E., & Kleitman, N. Two types of ocular motility occurring in sleep. J. appl. Physiol., 1955, 8, 1-10.

Berger, H. Ueber das Elektroenkephalogramm des Menschen. J. Psychol. Neurol., 1930, 40, 160-179.

Brazier, M. A. B., & Barlow, J. S. Some applications of correlation analysis to clinical problems in electroencephalography. Electro-enceph. clin. Neurophysiol., 1956, 8, 325-331.

Brazier, M. A. B., & Casby, J. U. Cross-correlation and auto-correlation studies of EEG potentials. Electroenceph. clin. Neurophysiol., 1952, 4, 201-211.

Burch, N. R. Automatic analysis of the electroencephalogram: A review and classification of systems. Electroenceph. clin. Neurophysiol., 1959, 11, 827-834.

Burch, N. R., Nettleton, W. J., Sweeney, J., & Edwards, R. J. Period analysis of the electroencephalogram on a general purpose digital computer. Ann. N. Y. Acad. Sci., 1964, 115, 827-843.

Burch, N. R., Vorderman, A., & Dossett, R. Period analysis of the electroencephalogram from the orbital flight of Gemini 7. NASA Project Report, July 14, 1966.

Cohen, J. A coefficient of agreement for nominal scales. Educat. Psychol. Measmt., 1960, 20(1), 37-46.

Cooley, W. W., & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N., & Hobart, G. Changes in human brain potentials during the onset of sleep. Science, 1937, 86, 448-450.

Dement, W. C. Dream recall and eye movements during sleep in schizophrenics and normals. J. nerv. ment. Dis., 1955, 122, 263-269.

Dement, W. C. The occurrence of low voltage, fast, EEG patterns during behavioral sleep in the cat. Electroenceph. clin. Neurophysiol., 1958, 10, 291-296.

Dement, W. C. The effect of dream deprivation. Science, 1960, 131, 1705-1707.

Dement, W. C. Further studies on the function of rapid eye movement sleep. Amer. Psychiatric Assn., New York, 1965.

Dement, W. C., & Kleitman, N. Cyclic variations in EEG during sleep and their relations to eye movements, body motility, and dreaming. Electroenceph. clin. Neurophysiol., 1957, 9, 673-690.

Drohocki, Z. L'intégrateur de l'électroproduction cérébrale pour l'électroencéphalographie quantitative. Rev. Neurol., 1948, 80, 619-624.

Farley, B. G. Recognition of patterns in the EEG. Electroenceph. clin. Neurophysiol., Suppl. 20, 1961, 49-55.

Fink, M., & Shapiro, D. Computer analytic classification of EEG sleep stages. Psychiatric Research Foundation of Missouri, Publication 65-10, 1965.

Gauss, E. J. Estimation of power spectral density by filters. J. A. Comput. Mach., 1964, 11, 98-103.

Gresham, S. C., Webb, W. B., & Williams, R. L. Alcohol and caffeine: Effects on inferred visual dreaming. Science, 1963, 140, 1226-1227.

Harvey, E. N., Loomis, A. L., & Hobart, G. A. Cerebral states during sleep as studied by human brain potentials. Sci. Monthly, 1937, 45, 191-192.

Johnson, L. C., Nute, C., Austin, M. T., & Lubin, A. Spectral analysis of the EEG during waking and sleep. Paper read at American Electroencephalographic Society, Denver, October 6-9, 1966.

Kales, A., Hoedemaker, F. S., Jacobson, A., & Lichtenstein, E. L. Dream deprivation: An experimental reappraisal. Nature, 1964, 204, 1337-1338.

Kleitman, N. Sleep and wakefulness. (Rev. ed.) Chicago: University of Chicago Press, 1963.

Kozhevnikov, V. A., & Meshcherskiy, R. M. Current methods of electroencephalographic analysis. State Publ. Hs., Moscow, 1963. (Translation, U. S. Dept. Commerce, JPRS 23, 590, 1964 [OTS 64-21749]).

Lester, B. K., & Burch, N. R. Psychophysiological studies of sleep in schizophrenic and control populations. Paper read at American Psychiatric Assn., New York, 1965.

Lindsley, D. B. Attention, consciousness, sleep, and wakefulness. In Handbook of physiology. Sect. 1, Neurophysiology, Vol. III. Washington: Amer. Physiol. Soc., 1960.

Loomis, A. L., Harvey, E. N., & Hobart, G. A. Cerebral states during sleep as studied by human brain potentials. J. exp. Psychol., 1937, 21, 127-144.

Luce, G. G. Current research on sleep and dreams. U. S. Public Health Service, Publication No. 1389, 1965.

Luce, G. G., & Segal, J. Sleep. New York: Coward-McCann, Inc., 1966.

Monroe, L. J. Inter-rater reliability of scoring EEG sleep records. Paper read at APSS, Santa Monica, April, 1967.

Muzio, J., Roffwarg, H., & Kaufman, R. Alteration in young adult human sleep EEG configuration resulting from d-LSD-24. Paper read at APSS, Palo Alto, 1964.

Rechtschaffen, A., & Maron, L. The effect of amphetamine on the sleep cycle. Electroenceph. clin. Neurophysiol., 1964, 16, 433-445.

Saltzberg, B., & Burch, N. R. A rapidly convergent orthogonal representation for EEG time series and related methods of automatic analysis. IRE Wescon Convention Record, Part 8, 1959.

Shapiro, D., & Fink, M. Quantitative analysis of the electroencephalogram by digital computer methods. Publication of the Department of Psychiatry at the Missouri Institute of Psychiatry, University of Missouri School of Medicine, January 10, 1966.

Snyder, F. The REM state in a living fossil. Paper read at APSS, Palo Alto, 1964.

Vanderplas, J. M., Sandison, W. A., & Vanderplas, J. N. Statistical and associational characteristics of 1100 random shapes. Percept. mot. Skills, 1965, 21, 414.

Walter, D. O. Spectral analysis for electroencephalograms: Mathematical determination of neurophysiological relationships from records of limited duration. Exper. Neurol., 1963, 8, 155-181.

Walter, W. G. An automatic low frequency analyser. Elec. Eng., 1943, 16, 9-13.

Whittlesey, J. R. B. A rapid method for digital filtering. Commun. A. Comput. Mach., 1964, 7, 552-556.

Williams, H. L., Hammack, J. T., Daly, R. L., Dement, W. C., & Lubin, A. Responses to auditory stimulation, sleep loss, and the EEG stages of sleep. Electroenceph. clin. Neurophysiol., 1964, 16, 269-279.

Williams, H. L., Morlock, H. C., Jr., & Morlock, J. V. Instrumental behavior during sleep. Psychophysiology, 1966, 2, 208-216.

Williams, H. L., & Williams, C. L. Nocturnal EEG profiles and performance. Psychophysiology, 1966, 3, 164-175.

APPENDIX A

MEANS AND STANDARD DEVIATIONS ON THE RAW SCORE MEASURES FOR EACH OF THE "CLASSIC" GROUPS, BY ANALYSIS

POSTERIOR 1 ANALYSIS

Means

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	3.09	14.63	41.59	31.84	20.63	19.80
Stage 2	14.70	32.36	14.36	24.25	23.40	22.47
Stage 3	35.68	23.93	7.80	28.64	19.87	16.52
Stage 4	39.96	20.40	5.79	26.84	17.62	15.98
REM	12.92	33.34	13.63	16.37	19.13	29.48
Overall	18.23	29.75	14.13	23.62	21.42	22.70

Standard Deviations

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	1.58	5.29	5.86	9.03	1.49	3.70
Stage 2	4.71	2.79	2.32	3.73	2.19	2.66
Stage 3	7.59	3.04	2.22	3.12	2.18	2.70
Stage 4	5.54	2.91	1.04	2.42	1.78	1.28
REM	4.32	2.53	3.27	3.96	1.85	3.23
Overall	11.06	6.07	7.51	6.03	2.94	5.21

POSTERIOR 2 ANALYSIS

Means

	Variable					
Group	M1	M2	M3	I1	I2	I3
Awake	4.58	14.15	38.54	18.64	26.34	23.22
Stage 2	13.76	28.73	18.01	13.46	25.77	24.55
Stage 3	32.20	24.15	10.90	23.30	23.78	17.83
Stage 4	40.60	21.01	8.49	25.50	22.40	16.46
REM	8.06	31.20	20.74	11.54	20.21	29.16
Overall	14.91	28.11	18.22	14.48	24.34	24.60

Standard Deviations

	Variable					
Group	M1	M2	M3	I1	I2	I3
Awake	1.84	7.68	18.46	17.04	10.58	11.43
Stage 2	3.86	2.27	2.52	2.84	3.05	2.59
Stage 3	5.05	2.50	1.83	2.30	2.61	1.11
Stage 4	3.79	1.95	1.06	1.22	2.42	.97
REM	2.28	2.75	2.50	2.77	3.21	2.69
Overall	8.97	4.30	6.02	5.38	4.10	4.48

COMBINED POSTERIOR ANALYSIS

Means

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	3.70	14.43	40.33	26.40	22.98	21.21
Stage 2	14.13	30.16	16.57	17.70	24.84	23.73
Stage 3	34.41	24.01	8.90	26.74	21.26	16.99
Stage 4	40.30	20.72	7.20	26.14	20.11	16.23
REM	10.39	32.22	17.34	13.85	19.69	29.31
Overall	16.41	24.85	16.38	18.60	23.03	23.74

Standard Deviations

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	1.80	6.15	12.22	14.12	7.18	7.73
Stage 2	4.23	3.05	3.02	6.18	2.98	2.80
Stage 3	6.94	2.83	2.55	3.83	2.98	2.34
Stage 4	4.61	2.42	1.72	1.97	3.21	1.13
REM	4.18	2.85	4.59	4.15	2.69	2.95
Overall	10.09	5.23	7.03	7.27	3.90	4.91

ANTERIOR 1 ANALYSIS

Means

Group	Variable M1	M2	M3	I1	I2	I3
Awake	4.97	21.20	18.82	1.00	5.56	34.80
Stage 2	17.76	27.08	41.81	11.84	23.78	31.97
Stage 3	29.81	22.43	11.48	18.52	26.14	22.12
Stage 4	50.49	15.53	3.73	21.97	18.65	13.82
REM	9.58	32.53	11.46	5.80	11.52	47.21
Overall	20.70	25.94	12.47	11.77	19.34	32.24

Standard Deviations

Group	Variable M1	M2	M3	I1	I2	I3
Awake	2.78	2.61	2.02	.78	1.53	3.07
Stage 2	5.72	4.05	3.85	3.17	6.05	6.14
Stage 3	6.68	3.10	2.04	2.86	3.35	3.01
Stage 4	3.70	1.76	.95	1.57	2.51	1.80
REM	2.68	3.20	2.07	1.81	1.99	4.77
Overall	14.26	6.34	5.02	6.36	7.74	11.52

ANTERIOR 2 ANALYSIS

Means

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	12.55	28.23	15.20	3.42	10.85	37.69
Stage 2	17.51	29.34	17.85	21.47	31.11	21.53
Stage 3	25.73	27.10	15.31	37.98	28.26	12.21
Stage 4	36.93	22.54	12.38	42.10	24.14	9.98
REM	14.28	34.52	11.92	7.75	18.38	38.02
Overall	20.42	28.88	15.43	22.57	25.71	23.00

Standard Deviations

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	3.52	3.66	1.51	.74	2.74	3.46
Stage 2	3.70	3.28	3.05	4.82	3.46	3.66
Stage 3	4.23	2.14	1.93	3.24	2.20	1.89
Stage 4	4.14	2.10	2.13	2.59	1.90	1.47
REM	3.33	3.32	1.58	1.52	2.13	2.48
Overall	8.63	4.63	3.52	13.10	7.18	10.43

COMBINED ANTERIOR ANALYSIS

Means

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	8.22	24.21	17.27	2.04	7.83	36.03
Stage 2	17.67	27.85	15.85	15.14	26.29	28.40
Stage 3	27.98	24.52	13.19	27.22	27.09	17.69
Stage 4	45.49	18.11	6.92	29.38	20.68	12.41
REM	10.99	33.13	11.60	6.38	13.58	44.44
Overall	20.60	26.98	13.52	15.59	21.59	28.97

Standard Deviations

Group	Variable					
	M1	M2	M3	I1	I2	I3
Awake	4.89	4.67	2.55	1.43	3.38	3.50
Stage 2	5.11	3.95	3.87	5.95	6.34	7.34
Stage 3	6.01	3.56	2.75	10.25	3.05	5.60
Stage 4	7.63	3.89	4.46	9.99	3.51	2.51
REM	3.60	3.34	1.94	1.94	3.75	5.97
Overall	12.55	5.95	4.76	10.64	8.13	11.98

APPENDIX B

CONTINGENCY TABLES FOR ALL PAIRED CLASSIFICATIONS

POSTERIOR 1 CLASSIFICATIONS

	Rater 1					
Rater 2	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	42	0	0	0	2	44
Stage 2	1	235	12	0	3	251
Stage 3	0	14	34	4	0	52
Stage 4	0	0	6	12	0	18
REM	1	16	0	0	97	114
Total	44	265	52	16	102	479

	Rater 1					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	36	2	0	0	1	39
Stage 2	0	230	9	0	12	251
Stage 3	0	25	28	3	0	56
Stage 4	0	0	15	13	0	28
REM	8	8	0	0	89	105
Total	44	265	52	16	102	479

Computer Own	Rater 1 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	34	3	0	0	0	37
Stage 2	2	230	7	1	19	259
Stage 3	0	10	37	8	0	55
Stage 4	0	0	4	7	0	11
REM	8	22	4	0	83	117
Total	44	265	52	16	102	479

Computer Combined	Rater 1 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	33	8	2	0	1	44
Stage 2	4	228	9	1	29	271
Stage 3	0	15	28	7	0	50
Stage 4	0	0	13	8	0	21
REM	7	14	0	0	72	93
Total	44	265	52	16	102	479

	Rater 2					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	36	1	0	0	2	39
Stage 2	0	225	6	0	20	251
Stage 3	0	20	36	0	0	56
Stage 4	0	0	10	18	0	28
REM	8	5	0	0	92	105
Total	44	251	52	18	114	479

	Rater 2					
Computer Own	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	34	3	0	0	0	37
Stage 2	3	224	10	0	22	259
Stage 3	0	7	38	10	0	55
Stage 4	0	0	3	8	0	11
REM	7	17	1	0	92	117
Total	44	251	52	18	114	479

Computer Combined	Rater 2					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	33	6	1	0	4	44
Stage 2	5	224	8	0	34	271
Stage 3	0	10	35	5	0	50
Stage 4	0	0	8	13	0	21
REM	6	11	0	0	76	93
Total	44	251	52	18	114	479

	Rater 3					
Computer Own	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	33	3	0	0	1	37
Stage 2	2	219	17	1	20	259
Stage 3	0	4	34	17	0	55
Stage 4	0	0	1	10	0	11
REM	4	25	4	0	84	117
Total	39	251	56	28	105	479

Computer Combined	Rater 3 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	33	7	2	0	2	44
Stage 2	3	220	17	1	30	271
Stage 3	0	7	35	8	0	50
Stage 4	0	0	2	19	0	21
REM	3	17	0	0	73	93
Total	39	251	56	28	105	479

Computer Combined	Computer Own Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	35	3	0	0	6	44
Stage 2	1	248	1	0	21	271
Stage 3	0	6	41	3	0	50
Stage 4	0	0	13	8	0	21
REM	1	2	0	0	90	93
Total	37	259	55	11	117	479

POSTERIOR 2 CLASSIFICATIONS, FIRST NIGHT

	Rater 1					
Rater 2	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	0	0	0	0	3
Stage 2	0	258	1	0	0	259
Stage 3	0	5	29	0	0	34
Stage 4	0	0	0	49	0	49
REM	0	9	0	0	126	135
Total	3	272	30	49	126	480

	Rater 1					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	14	0	1	12	30
Stage 2	0	203	4	0	3	210
Stage 3	0	14	19	8	0	41
Stage 4	0	0	7	40	0	47
REM	0	41	0	0	111	152
Total	3	272	30	49	126	480

	Rater 1					
Computer Own	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	15	0	1	15	34
Stage 2	0	203	6	0	16	225
Stage 3	0	3	21	17	0	41
Stage 4	0	0	3	31	0	34
REM	0	51	0	0	95	146
Total	3	272	30	49	126	480

	Rater 1					
Computer Combined	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	29	0	1	21	54
Stage 2	0	184	7	0	16	207
Stage 3	0	6	23	23	0	52
Stage 4	0	0	0	25	0	25
REM	0	53	0	0	89	142
Total	3	272	30	49	126	480

	Rater 2					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	13	0	1	13	30
Stage 2	0	201	3	0	6	210
Stage 3	0	9	24	8	0	41
Stage 4	0	0	7	40	0	47
REM	0	36	0	0	116	152
Total	3	259	34	49	135	480

	Rater 2					
Computer Own	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	15	0	1	15	34
Stage 2	0	199	9	0	17	225
Stage 3	0	2	22	17	0	41
Stage 4	0	0	3	31	0	34
REM	0	43	0	0	103	146
Total	3	259	34	49	135	480

Computer Combined	Rater 2					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	3	28	0	1	22	54
Stage 2	0	181	10	0	16	207
Stage 3	0	5	24	23	0	52
Stage 4	0	0	0	25	0	25
REM	0	45	0	0	97	142
Total	3	259	34	49	135	480

Computer Own	Rater 3					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	24	5	0	0	5	34
Stage 2	3	187	13	0	22	225
Stage 3	0	0	25	16	0	41
Stage 4	0	0	3	31	0	34
REM	3	18	0	0	125	146
Total	30	210	41	47	152	480

Computer Combined	Rater 3					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	27	10	0	0	17	54
Stage 2	3	176	12	0	16	207
Stage 3	0	1	27	24	0	52
Stage 4	0	0	2	23	0	25
REM	0	23	0	0	119	142
Total	30	210	41	47	152	480

Computer Combined	Computer Own					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	33	5	0	0	16	54
Stage 2	0	202	1	0	4	207
Stage 3	0	3	39	10	0	52
Stage 4	0	0	1	24	0	25
REM	1	15	0	0	126	142
Total	34	225	41	34	146	480

POSTERIOR 2 CLASSIFICATIONS, SECOND NIGHT

	Rater 1					
Rater 2	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	5	3	0	0	3	11
Stage 2	2	255	0	0	4	261
Stage 3	0	11	15	18	0	44
Stage 4	0	0	4	10	0	14
REM	2	16	0	0	132	150
Total	9	285	19	28	139	480

	Rater 1					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	7	3	0	0	0	10
Stage 2	0	265	2	0	7	274
Stage 3	0	2	17	20	0	39
Stage 4	0	0	0	8	0	8
REM	2	15	0	0	132	149
Total	9	285	19	28	139	480

Computer First Night	Rater 1					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	6	6	0	0	6	18
Stage 2	2	250	0	0	31	283
Stage 3	0	14	16	15	0	45
Stage 4	0	0	3	13	0	16
REM	1	15	0	0	102	118
Total	9	285	19	28	139	480

Computer Combined	Rater 1					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	7	14	0	0	19	40
Stage 2	2	201	0	0	9	212
Stage 3	0	13	19	25	0	57
Stage 4	0	0	0	3	0	3
REM	0	57	0	0	111	168
Total	9	285	19	28	139	480

Rater 3	Rater 2: Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	5	4	0	0	1	10
Stage 2	0	251	11	1	11	274
Stage 3	0	1	30	8	0	39
Stage 4	0	0	3	5	0	8
REM	6	5	0	0	138	149
Total	11	261	44	14	150	480

Computer First Night	Rater 2: Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	8	4	0	0	6	18
Stage 2	1	238	4	0	40	283
Stage 3	0	7	36	2	0	45
Stage 4	0	0	4	12	0	16
REM	2	12	0	0	104	118
Total	11	261	44	14	150	480

Computer Combined	Rater 2					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	9	11	0	0	20	40
Stage 2	0	196	4	0	12	212
Stage 3	0	6	39	12	0	57
Stage 4	0	0	1	2	0	3
REM	2	48	0	0	118	168
Total	11	261	44	14	150	480

Computer First Night	Rater 3					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	7	1	0	0	10	18
Stage 2	3	244	0	0	36	283
Stage 3	0	14	29	2	0	45
Stage 4	0	0	10	6	0	16
REM	0	15	0	0	103	118
Total	10	274	39	8	149	480

Computer Combined	Rater 3 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	7	5	0	0	28	40
Stage 2	3	201	0	0	8	212
Stage 3	0	13	37	7	0	57
Stage 4	0	0	2	1	0	3
REM	0	55	0	0	113	168
Total	10	274	39	8	149	480

Computer Combined	Computer First Night Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	18	11	0	0	11	40
Stage 2	0	210	2	0	0	212
Stage 3	0	1	43	13	0	57
Stage 4	0	0	0	3	0	3
REM	0	61	0	0	107	168
Total	18	283	45	16	118	480

ANTERIOR 1 CLASSIFICATIONS

	Rater 1					
Rater 2	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	21	0	0	0	0	21
Stage 2	0	143	0	0	4	147
Stage 3	0	34	25	0	0	59
Stage 4	0	0	6	40	0	46
REM	0	0	0	0	99	99
Total	21	177	31	40	103	372

	Rater 1					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	19	1	0	0	3	23
Stage 2	1	172	13	0	4	190
Stage 3	0	0	15	4	0	19
Stage 4	0	0	3	36	0	39
REM	1	4	0	0	96	101
Total	21	177	31	40	103	372

Computer Own	Rater 1 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	18	0	0	0	1	19
Stage 2	0	133	3	0	7	143
Stage 3	0	32	18	1	0	51
Stage 4	0	0	10	39	0	49
REM	3	12	0	0	95	110
Total	21	177	31	40	103	372

Computer Combined	Rater 1 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	19	2	0	0	8	29
Stage 2	0	128	4	0	3	135
Stage 3	0	26	16	0	0	42
Stage 4	0	0	11	40	0	51
RFM	2	21	0	0	92	115
Total	21	177	31	40	103	372

Rater 3	Rater 2 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	19	1	0	0	3	23
Stage 2	1	138	47	0	4	190
Stage 3	0	0	12	7	0	19
Stage 4	0	0	0	39	0	39
REM	1	8	0	0	92	101
Total	21	147	59	46	99	372

Computer Own	Rater 2 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	18	0	0	0	1	19
Stage 2	0	124	13	0	6	143
Stage 3	0	8	41	2	0	51
Stage 4	0	0	5	44	0	49
REM	3	15	0	0	92	110
Total	21	147	59	46	99	372

Computer Combined	Rater 2					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	19	2	0	0	8	29
Stage 2	0	116	17	0	2	135
Stage 3	0	5	36	1	0	42
Stage 4	0	0	6	45	0	51
REM	2	24	0	0	89	115
Total	21	147	59	46	99	372

Computer Own	Rater 3					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	18	0	0	0	1	19
Stage 2	0	135	0	0	8	143
Stage 3	1	40	9	1	0	51
Stage 4	0	1	10	38	0	49
REM	4	14	0	0	92	110
Total	23	190	19	39	101	372

Computer Combined	Rater 3					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	22	2	0	0	5	29
Stage 2	0	130	0	0	5	135
Stage 3	0	34	8	0	0	42
Stage 4	0	1	11	39	0	51
REM	1	23	0	0	91	115
Total	23	190	19	39	101	372

Computer Combined	Computer Own					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	19	1	1	0	8	29
Stage 2	0	129	6	0	0	135
Stage 3	0	0	42	0	0	42
Stage 4	0	0	2	49	0	51
REM	0	13	0	0	102	115
Total	19	143	51	49	110	372

ANTERIOR 2 CLASSIFICATIONS

	Rater 1					
Rater 2	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	25	9	0	0	2	36
Stage 2	1	152	8	0	0	161
Stage 3	0	16	28	0	0	44
Stage 4	0	1	13	10	0	24
REM	2	5	0	0	84	91
Total	28	183	49	10	86	356

	Rater 1					
Rater 3	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	24	0	0	0	2	26
Stage 2	1	158	6	0	2	167
Stage 3	0	16	18	0	0	34
Stage 4	0	0	25	10	0	35
REM	3	9	0	0	82	94
Total	28	183	49	10	86	356

Computer Own	Rater 1 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	26	7	0	0	3	36
Stage 2	1	145	7	0	0	153
Stage 3	0	18	19	0	0	37
Stage 4	0	0	23	10	0	33
REM	1	13	0	0	83	97
Total	28	183	49	10	86	356

Computer Combined	Rater 1 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	24	5	0	0	2	31
Stage 2	1	144	6	0	23	174
Stage 3	0	33	21	1	0	55
Stage 4	0	0	22	9	0	31
REM	3	1	0	0	61	65
Total	28	183	49	10	86	356

Rater 3	Rater 2 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	23	0	0	0	3	26
Stage 2	3	154	4	0	6	167
Stage 3	0	5	28	1	0	34
Stage 4	0	0	12	23	0	35
REM	10	2	0	0	82	94
Total	36	161	44	24	91	356

Computer Own	Rater 2 Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	31	1	0	0	4	36
Stage 2	0	148	3	1	1	153
Stage 3	0	6	30	1	0	37
Stage 4	0	0	11	22	0	33
REM	5	6	0	0	86	97
Total	36	161	44	24	91	356

	Rater 2					
Computer Combined	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	28	1	0	0	2	31
Stage 2	4	139	3	0	28	174
Stage 3	0	21	30	4	0	55
Stage 4	0	0	11	20	0	31
REM	4	0	0	0	61	65
Total	36	161	44	24	91	356

	Rater 3					
Computer Own	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	25	1	0	0	10	36
Stage 2	0	146	7	0	0	153
Stage 3	0	8	23	6	0	37
Stage 4	0	0	4	29	0	33
REM	1	12	0	0	84	97
Total	26	167	34	35	94	356

Computer Combined	Rater 3					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	23	0	0	0	8	31
Stage 2	0	145	3	0	26	174
Stage 3	0	20	27	8	0	55
Stage 4	0	0	4	27	0	31
REM	3	2	0	0	60	65
Total	26	167	34	35	94	356

Computer Combined	Computer Own					
	Awake	Stage 2	Stage 3	Stage 4	REM	Total
Awake	31	0	0	0	0	31
Stage 2	2	137	0	0	35	174
Stage 3	0	16	37	2	0	55
Stage 4	0	0	0	31	0	31
REM	3	0	0	0	62	65
Total	36	153	37	33	97	356