

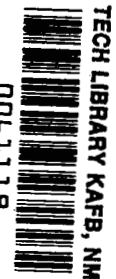
# NASA CONTRACTOR REPORT



NASA CR-11

C.1

0061119



NASA CR-1840

LOAN COPY: RETURN TO  
AFWL (DOGL)  
KIRTLAND AFB, N. M.

## A TECHNIQUE FOR AUTOMATIC REAL-TIME SCORING OF SEVERAL SIMULTANEOUS SLEEP ELECTROENCEPHALOGRAMS

*by R. W. Becker, J. S. Lukas, M. E. Dobbs, and F. Poza*

*Prepared by*  
STANFORD RESEARCH INSTITUTE  
Menlo Park, Calif.  
*for Langley Research Center*

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • AUGUST 1971



0061119

1. Report No. NASA CR-1840		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle A Technique for Automatic Real-Time Scoring of Several Simultaneous Sleep Electroencephalograms				5. Report Date August 1971	
				6. Performing Organization Code	
7. Author(s) R. W. Becker            M. E. Dobbs J. S. Lucas             F. Poza				8. Performing Organization Report No.	
				10. Work Unit No.	
9. Performing Organization Name and Address  Stanford Research Institute Menlo Park, California				11. Contract or Grant No. NAS1-9286	
				13. Type of Report and Period Covered Contractor Report	
12. Sponsoring Agency Name and Address  National Aeronautics and Space Administration Washington, D.C. 20546				14. Sponsoring Agency Code	
				15. Supplementary Notes	
16. Abstract In the course of an ongoing program to assess the effect of aircraft noise on human sleep, the need for some automatic technique for analyzing and scoring the electroencephalographic (EEG) sleep records became apparent. Extensive EEG records are obtained in this program since groups of four subjects each are tested nightly. At the typical rate at which encephalographic paper records are made (177 ft of paper per hour, per subject), a great deal of technical manpower is required for visual scoring and analysis of such an experiment. An additional, and possibly more critical, problem involves the imperfect reliability of visual analysis of EEGs: inter-scorer reliabilities range between 0.80 and 0.95, even among "expert" scorers.  In light of the above, Stanford Research Institute, while under contract to the National Aeronautics and Space Administration, during an eight-month period developed and tested the electronic devices for filtering and processing the electrophysiological signals necessary to discriminate between stages of sleep. Included in the system is a small computer (XDS CEL6) used to score and analyze the processed electrophysiological signals.					
17. Key Words (Suggested by Author(s))  EEG - Computer scoring of sleep data			18. Distribution Statement  Unclassified - unlimited		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 30	22. Price* \$3.00



## CONTENTS

---

LIST OF ILLUSTRATIONS . . . . .	iv
LIST OF TABLES . . . . .	v
I INTRODUCTION . . . . .	1
II AN OVERALL DESCRIPTION OF THE SYSTEM . . . . .	2
III SLEEP SCORING SYSTEM ACCURACY . . . . .	9
A. Method . . . . .	9
B. Results . . . . .	11
IV DISCUSSION AND CONCLUSION . . . . .	17
REFERENCES . . . . .	20
Appendix--DETAILED DESCRIPTION OF THE COMPUTER PROGRAM . . . . .	21

## ILLUSTRATIONS

---

Figure 1	Schematic Diagram of Automatic Sleep Stage Scoring System. . . . .	3
Figure 2	Analog Preprocessor Block Diagram for One Subject . .	4
Figure 3	Timing Charts for Sampling and Processing Analog Signals. . . . .	5

TABLES

---

Table I	Sleep Stage Classifications by the SRI Laboratory of the 3322 Epochs Used in this Study . . . . .	10
Table II	Agreement (in percent) between Sleep Stages as Classified Visually by the SRI Laboratory and as Classified by the University of Oklahoma Sleep Laboratory . . . . .	12
Table III	Percent Agreement between Sleep Stages as Classified Visually by the SRI Laboratory, the Oklahoma Laboratory, and the Computer for each of 12 Subjects. . . . .	13
Table IV	Agreement (in percent) between Sleep Stages as Classified Visually by the SRI Laboratory and as Classified by the Computer . . . . .	15
Table V	Agreement (in percent) between Grouped Sleep Stages as Classified by the Authors, the University of Oklahoma Sleep Laboratory, and the Computer for Subjects in Three Age Groups. . . . .	18
Table A-I	Relative Thresholds Used in the Computer Program to Discriminate between Sleep Stages. . . . .	26

A TECHNIQUE FOR AUTOMATIC REAL-TIME SCORING OF  
SEVERAL SIMULTANEOUS SLEEP ELECTROENCEPHALOGRAMS

By

R. W. Becker, J. S. Lukas, M. E. Dobbs, and F. Poza

I INTRODUCTION

In the course of an ongoing program to assess the effect of aircraft noise on human sleep, the need for some automatic technique for analyzing and scoring the electroencephalographic (EEG) sleep records became apparent. Extensive EEG records are obtained in this program since groups of four subjects each are tested nightly. At the typical rate at which encephalographic paper recordings are made (177 feet of paper per hour per subject), a great deal of technical manpower is required for visual scoring and analysis of such an experiment. An additional, and possibly more critical, problem involves the imperfect reliability of visual analysis of EEGs: inter-scorer reliabilities range between 0.80 and 0.95, even among "expert" scorers.<sup>1,2,3,4\*</sup>

In light of the above, Stanford Research Institute, while under contract to the National Aeronautics and Space Administration, during an eight-month period developed and tested the electronic devices for filtering and processing the electrophysiological signals necessary to discriminate between stages of sleep. Included in the system is a small computer (XDS CE16) used to score and analyze the processed electrophysiological signals.

This report documents that development effort. It includes (1) a description of the electronic devices, (2) a detailed description of the computer program, as an appendix, and (3) the results of a study of the accuracy of sleep-stage classifications by the system as compared to classifications by two independent sleep laboratories.

---

\* References are listed at the end of the text.

## II AN OVERALL DESCRIPTION OF THE SYSTEM

A functional description of the sleep analysis system is shown in Figure 1. The four electrode outputs from each subject pass into the electroencephalograph (EEG) for amplification and recording on paper, and the amplified signals are sent in parallel to the analog preprocessor.

In the analog preprocessor, described schematically in Figure 2, the EEG signal is filtered into five components: 0-2.5 Hz, 3-7 Hz, 8-12 Hz, 13-16 Hz, and 16-40 Hz. The 0-2.5 Hz band is simply amplified and sent to the multiplexer-converter. The 3-7, 8-12, and 13-16 Hz bands are each full-wave rectified and smoothed in order to produce a slowly-varying voltage proportional to the envelope of the filter outputs. The 16-40 Hz band and the electromyographic (EMG) band are squared and smoothed. The 2 electro-oculogram (EOG) channels are simply amplified without filtering or smoothing. In total, 32 channels of information (5 filtered EEG signals, 2 EOG signals, and 1 EMG signal for each of four subjects) are available simultaneously for sampling and digitizing. These 32 outputs of the analog preprocessor are sampled and digitized under computer control as shown in Figure 3.

The epoch timer (Figure 1) causes a pulse every 20 seconds that (1) creates a visual mark on the timing channels of the 2 EEGs, and (2) by activating an interrupt line on the computer, informs it that a new epoch has begun. Each 20-second epoch is broken into 400 timing blocks, each 50 milliseconds long. Approximately 10 milliseconds of the block are used to sample, store, and process the 32 channels. The remaining 40 milliseconds are used in the very first block of an epoch to make the stage-scoring decision for the previous epoch. In the remainder of the



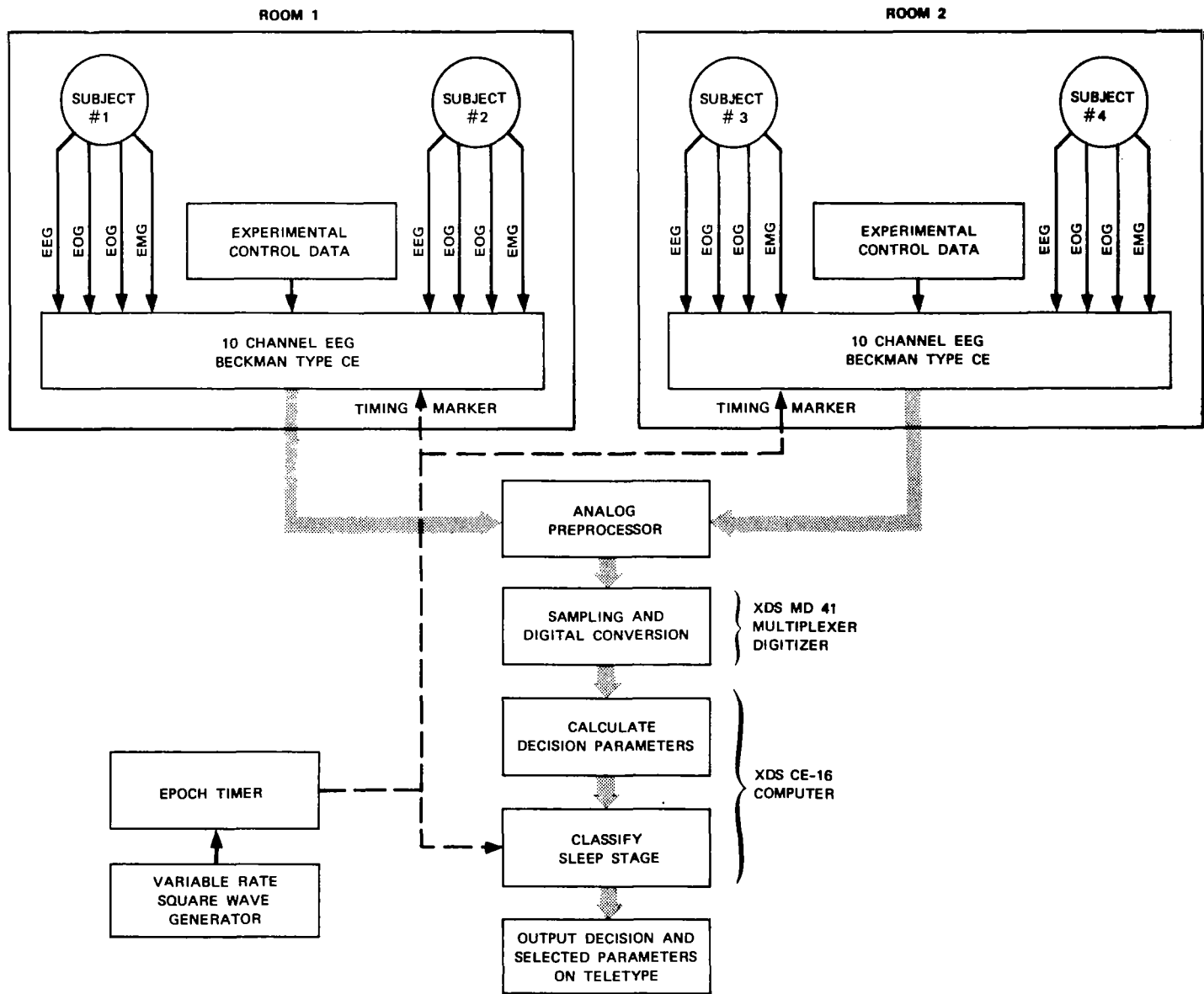


FIGURE 1 SCHEMATIC DIAGRAM OF AUTOMATIC SLEEP STAGE SCORING SYSTEM

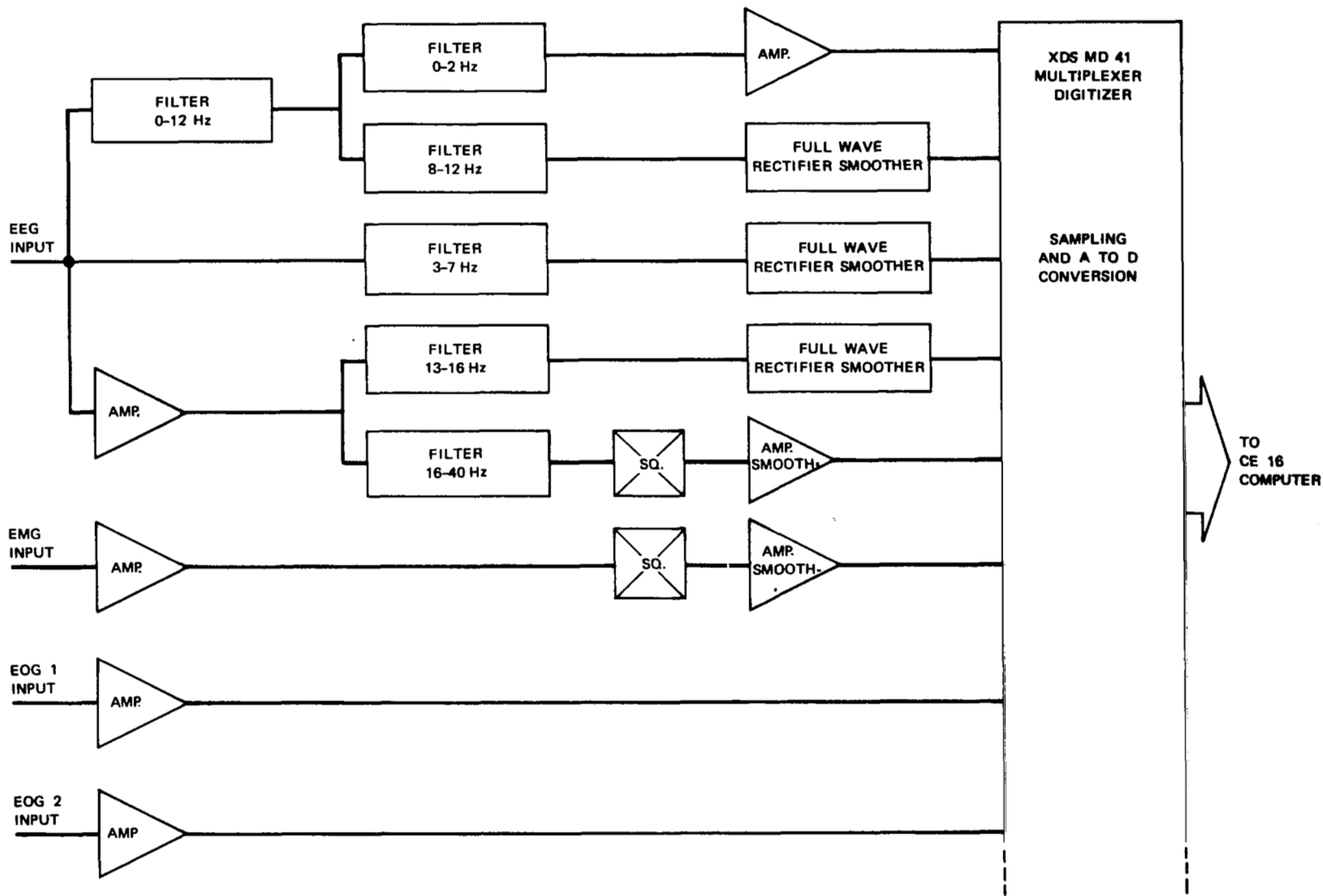


FIGURE 2 ANALOG PREPROCESSOR BLOCK DIAGRAM FOR ONE SUBJECT

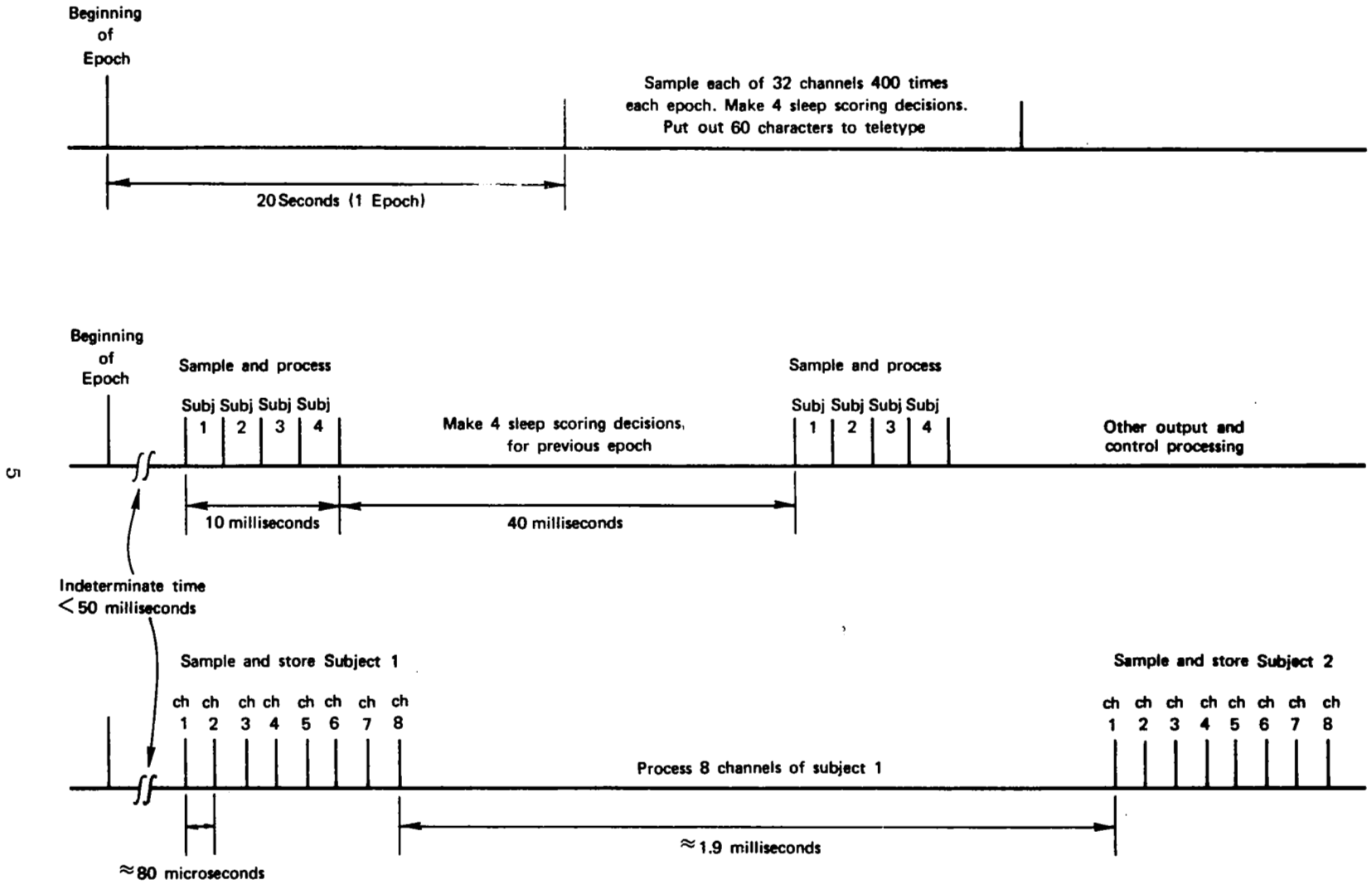


FIGURE 3 TIMING CHARTS FOR SAMPLING AND PROCESSING ANALOG SIGNALS

blocks, the 40 millisecond period is used to put out a character to the teletype whenever necessary, or simply waiting in a loop for the next block to begin. In the 10-millisecond sampling sections, the eight channels of the first subject are sampled (at a rate of one channel approximately every 80 microseconds) and processed for 2 milliseconds; then the eight channels of the second subject are sampled and processed, etc. The channels are sampled and processed in the order 0-2.5 Hz, 3-7 Hz, 8-12 Hz, 13-16 Hz, 16-40 Hz, EMG, EOG, EOG.

Each of the 32 analog voltages (constrained by the analog preprocessor to be between -10 V and +10 V) is digitized by converting the voltage to a 15-bit binary number. If the number is negative, it is represented in 2's complement form. During the 80-microsecond sampling period, the 15-bit word is made into a CE-16 16-bit word and all numbers are converted to absolute values except for the 0-2.5 Hz output and the EOG channels.

Approximately 0.6 millisecond is required to sample the eight channels of one subject and process them as is described above. The next 2 milliseconds are used to develop a set of numbers associated with each epoch that are later used to determine the stage of sleep. These numbers are of two kinds. One kind of number is simply the sum of the absolute values of a given signal. Thus, we calculate the sum of the absolute values from the rectified and smoothed 3-7 Hz EEG band, 8-12 Hz EEG band, 13-16 Hz EEG band, and the sum of the absolute values of the squared and smoothed 16-40 Hz EEG band, and the EMG signal. These numbers depend only upon the output from one channel of the analog preprocessor.

The other kind of number represents the number of times that the overall pattern of signals suggested the existence of a certain kind of activity such as spindling or rapid eye movement. These numbers include (1) amount of potential delta activity in the epoch, (2) amount of

potential alpha activity, (3) amount of potential rapid eye movements. These numbers are generally a function of several analog signals; for example, a given sample of eight signals from a particular subject is considered indicative of potential alpha activity only if (1) the 8-11 Hz sample is above a threshold for the subject and (2) both the 3-7 Hz and 13-16 Hz samples are below specified thresholds for the subject. In this way the patterns of analog signals are used to prevent some misclassifications of stage of sleep that would occur if only positive responses from single filters were used to indicate a particular kind of activity.

During the last 40 milliseconds of the first 50-millisecond timing block of an epoch, the numbers described above that have been calculated for the previous epoch are further processed as follows:

First, the sums of absolute values of given signals are converted to average values, and the number of sampling intervals indicating a certain kind of activity are converted to percent of the epoch spent in that kind of potential activity. This conversion is necessary because the number of sampling periods in an epoch may vary slightly from epoch to epoch because of the asynchrony of the epoch-timer and the on-going real-time program. Secondly, the mean 3-7 Hz level divided by the mean 8-11 Hz level is calculated. This ratio parameter is used in a number of stage-scoring decisions later in the program. Finally, these calculated parameters on each epoch are used to score the sleep stage of each of the four subjects. A non-detailed version of the scoring algorithm may be stated as follows:

1. Are both EMG and beta levels very high? If so, classify as movement time.
2. Is delta activity greater than 50 percent? If so, classify as stage 4.

3. Is delta activity greater than 20 percent? If so, classify as stage 3.
4. Is spindle activity greater than a specified threshold for this subject? If so, note spindle occurred this epoch. Are spindle level, beta level, EMG level, and theta/alpha ratio all consistent with stage 2 for the subject? If so, classify as stage 2. If not, examine stage REM possibility.
5. If not stage REM, are three of the four above parameters consistent with stage 2, and has there been a spindle within last three minutes? If so, classify as stage 2.
6. Otherwise use theta/alpha ratio to decide between stage awake and stage 1. A high ratio corresponds to stage 1.
7. Stage REM is scored if significant eye movements have occurred during the last three minutes and two successive epochs have not had high EMG levels.

After the classifications have been made, the sleep stage for each subject and decision parameters for one of the four subjects are printed on the teletype. Although printing of the information takes 11 seconds, time-sharing permits concurrent printing of data from the epoch just concluded with sampling and calculation of the data during the ongoing epoch. If any subject has changed stage of sleep from the last epoch, this information is added to a stage-transition matrix that will be printed out at the end of the evening for each subject.

### III SLEEP SCORING SYSTEM ACCURACY

#### A. Method

Two scorers of "medium" experience<sup>2</sup> from each of two sleep laboratories\* scored the EEGs using criteria described by Rechtschaffen and Kales.<sup>5</sup> It is estimated that the scorers had, on the average, four years of scoring experience, with a range of about three to six years. Use was made of specially-devised scoring forms that were keyed with the EEG record and the printout of the computer such that the 20-second epochs scored by the computer were identical to those scored visually by both pairs of scorers.

It should be noted that most of the experience of two of the four scorers was with college-age subjects, while the EEGs scored in this study were from other age groups. (The ages of the subjects may be found in Table III.) Our experience indicates that there are pronounced differences between age groups with respect to both frequency and amplitude of the EEG. For example, children (6 to 8 years of age) typically had delta waves of 600-700 microvolts, while those waves of the old group (69 to 75 years of age) were typically at or below 100 microvolts and in some cases below 75 microvolts. The differences between the scoring criteria espoused by Rechtschaffen and Kales and those found in the EEGs of some of our 12 subjects made scoring by the inexperienced a difficult task, as will be seen below. Any scoring disagreements

---

\* We are grateful to Dr. Harold L. Williams of the University of Oklahoma Medical School for his participation.

between scorers within a laboratory were resolved through discussion of details of the epochs in question.

The EEG records were obtained near the conclusion of a study concerned primarily with the effects of aircraft noises on sleep; they generally had a duration of about six hours, beginning when the subjects first went to bed, and included the effects of the test stimuli on the subjects. A computer routine selected 16 contiguous 20-second epochs from a single subject for presentation of detailed information of the EEG parameters (see Step 14 of Appendix 1) and then automatically shifted to the next subject. The epochs analyzed visually corresponded to those assigned by the computer for detailed data presentation.

Table I shows how the 3322 epochs used in this study were scored visually by the SRI laboratory. It is believed that the sleep stages

Table I

SLEEP STAGE CLASSIFICATIONS BY THE SRI LABORATORY  
OF THE 3322 EPOCHS USED IN THIS STUDY

Sleep Stage as Scored by SRI	Age Group					
	Young (5 to 8 years)		Middle-Aged (45 to 57 years)		Old (69 to 75 years)	
	N	%	N	%	N	%
0	73	6.0	61	5.9	153	14.2
1	48	3.9	97	9.4	90	8.4
2	561	46.1	577	56.0	530	49.3
3	174	14.3	91	8.8	124	11.5
4	93	7.7	24	2.3	57	5.3
REM	267	22.0	180	17.5	122	11.3
Total	1216		1030		1076	



as assigned in our laboratory are probably accurate since we have had considerable experience in scoring the sleep of individuals in the age groups of interest.<sup>6</sup>

It should be pointed out that the data in Table I should not be mistaken for the distribution, in terms of relative duration, of the different sleep stages during a night of sleep. The samples studied were not representative of normal sleep patterns because sleep-disturbing stimuli were presented.<sup>7</sup>

Four signals were recorded from each subject. They were:

1. EEG--Left or right central electrode (C<sub>3</sub> or C<sub>4</sub>) monopolar with respect to the contralateral mastoid
2. Two eye-movement electrodes, one slightly above and the other slightly below the outer canthi of the eye and both monopolar with respect to single electrode just above the nasion
3. Bipolar electrodes on the lower chin, one to two cm to the right and left of the midline.

## B. Results

### 1. Comparisons of Sleep Stages as Scored Visually by Two Pairs of Scorers

In the following tables the data indicate that the amount of agreement between the two groups of scorers depends not only upon the individual being scored but also upon the stage of sleep being scored. Table II shows that agreement varied between 97 percent in the scoring of stage 4 for the children to 22 percent in the scoring of stage awake in the middle-aged group, while agreement between the scorers over all

Table II

AGREEMENT (IN PERCENT) BETWEEN SLEEP STAGES AS CLASSIFIED VISUALLY  
BY THE SRI LABORATORY AND AS CLASSIFIED BY THE UNIVERSITY  
OF OKLAHOMA SLEEP LABORATORY

Age Group	Sleep Stage as Scored Visually in the SRI Laboratory	Sleep Stage as Scored in the Oklahoma Laboratory							Row Total (N)
		0	1	2	3	4	REM	?*	
Children	0	87.7 81.0	12.3 15.5						73
	1	25.0 15.2	41.7 34.5	16.7 2.1	6.3 1.2	4.2 0.9	4.2 0.8	2.1 100	48
	2		3.7 36.2	63.3 92.2	29.4 65.7	2.9 7.9	0.7 1.7		561
	3		0.6 1.7	0.6 0.2	44.8 31.1	54.0 46.5			174
	4				3.2 1.2	96.8 44.6			93
	REM	1.1 3.8	2.6 12.1	7.9 5.5	0.7 0.8		87.6 97.5		267
Column Total (N)		79	58	385	251	202	240	1	1216
Middle-Aged	0	21.3 81.3	55.7 21.5	6.6 0.8			4.9 1.5	11.5 46.7	61
	1	2.1 12.5	78.4 48.1	6.2 1.1	1.0 1.0		10.3 5.1	2.1 13.3	97
	2	0.2 6.3	8.3 30.4	87.0 94.5	2.9 17.2		1.2 3.6	0.4 13.3	577
	3			19.8 3.4	74.7 68.7	4.4 26.7		1.1 6.7	91
	4				54.2 13.1	45.8 73.3			24
	REM			0.5 0.2			97.8 89.8	1.7 20.0	180
Column Total (N)		16	158	531	99	15	196	15	1030
Old	0	74.5 89.8	17.0 10.9	1.3 0.5				7.2 23.4	153
	1	10.0 7.1	41.1 15.5	13.3 3.5			11.1 8.5	24.4 46.8	90
	2	0.8 3.1	31.3 69.5	60.2 92.2	6.0 21.2			1.7 19.2	530
	3		0.8 0.4	10.5 3.8	88.7 72.8				124
	4				15.8 6.0	84.2 100			57
	REM		7.4 3.7				88.5 91.5	4.1 10.6	122
Column Total (N)		127	239	346	151	48	118	47	1076

This and subsequent similar tables can be read as follows: for any given row the numbers (in percent) above the diagonal line should be read, and for any given column read the numbers below the diagonal lines. Thus, for the children, of the 73 epochs scored 0 (rows) in the SRI laboratory 87.7 percent were scored 0 and 12.3 percent were scored 1 by the Oklahoma group. In contrast, of the 79 epochs scored 0 (column) in the Oklahoma laboratory 81 percent were scored 0, 15.2 percent were scored 1, and 3.8 percent were scored REM in the SRI laboratory.

\* For the old group, the large majority of entries in this column are scores of X. See text.

stages for each subject, as shown in Table III, ranged between 38 and 88 percent.

Table III

PERCENT AGREEMENT BETWEEN SLEEP STAGES AS CLASSIFIED VISUALLY  
 BY THE SRI LABORATORY, THE OKLAHOMA LABORATORY, AND THE  
 COMPUTER FOR EACH OF 12 SUBJECTS

Group	Age (Years)	Subject	Agreement (in Percent) Between SRI Lab and:	
			Oklahoma Laboratory	Computer
Children	5	LP	68	38
	5	DP	69	70
	6	SP	69	59
	8	KP	70	57
All Children			69	56
Middle- Aged	45	VS	82	78
	45	DM	85	53
	55	EJ	84	62
	57	RY	79	58
All Middle-Aged			82	63
All Old	69	OH	68	60
	70	GA	38	35
	74	JR	88	64
	75	HC	83	70
All Old			68	60
All Subjects			73	59

Clearly the sleep records of some subjects were more difficult to score than those of other subjects. For example, the scorers from the Oklahoma laboratory thought it necessary to assign a new stage, "X", to some epochs of subject OH. These epochs feature a particularly "flat" EEG with relatively long bursts (about 2 seconds) of spindle-like low-voltage activity (about 25 microvolts) but at a frequency (17-20 Hz) that could not be considered spindles.

Another factor in the disagreement between scorers was the amplitude of the EEGs. In particular the slow-wave activity (less than 2.5 Hz) of children was of very high voltage (above 600 microvolts) typically, while being relatively low (below 100 microvolts) in the old men. This precluded the direct use of standard criteria such as are presented in Rechtschaffen and Kales.<sup>5</sup> Rather, it was necessary for each scorer on the basis of his experience to establish amplitude criteria for each subject. It is clear that different criteria were chosen by different pairs of scorers, as is indicated by the disagreements in Table II in the scoring of stages 4, 3, and 2.

## 2. Comparison of Automatic Scoring with Visual Scoring by the SRI Laboratory

Examination of Tables III and IV shows that when the automatic scoring technique is compared to visual scoring, much higher agreement was found for some subjects and for some sleep stages than for others. Least agreement between automatic and visual scoring was found for stages 1 and REM. While developing this scoring system, little effort was put into making the stage awake versus stage 1 decision, and it is now clear that the parameter (mean alpha to mean theta ratio) chosen to make this discrimination was inadequate. The disagreement in scoring REM was partly due to missing some REM periods entirely, and partly due to disagreements

Table IV

AGREEMENT (IN PERCENT) BETWEEN SLEEP STAGES AS CLASSIFIED VISUALLY  
BY THE SRI LABORATORY AND AS CLASSIFIED BY THE COMPUTER

Age Group	Sleep Stage as Scored Visually in the SRI Laboratory	Sleep Stage as Scored by the Computer							Row Total (N)	
		0	1	2	3	4	REM	Movement		
Children	0	64.4	19.2	8.2				8.2		73
		24.6	7.5	1.7				2.3		
	1	35.4	31.2	16.7	8.3	2.1		6.3		
		8.9	8.0	2.2	2.1	3.7		1.1		
	2	21.6	14.8	52.9	3.4			7.3		
		63.3	44.4	82.9	10.1			15.5		
	3	1.7	10.3	20.7	55.2	2.9		9.2		174
		1.6	9.6	10.1	50.8	18.5		6.1		
	4			4.3	73.1	22.6				93
				1.1	36.0	77.8				
	REM	1.1	21.2	2.5	0.7			75.0		267
		1.6	30.5	2.0	1.1			75.0		
Column Total (N)		191	187	358	189	27	264			1216
Middle-Aged	0	45.9	26.2	14.8	3.3				9.8	61
		40.0	11.6	1.6	1.6				85.7	
	1	6.2	30.9	51.5	1.0			9.3	1.0	
		8.6	21.7	8.7	0.8			9.6	14.3	
	2	3.8	10.2	76.1	9.9					
		31.4	42.8	76.2	44.5					
	3	13.2	4.4	15.4	62.6	4.4				91
		17.1	2.9	2.4	44.5	23.5				
	4				45.8	54.2				24
					8.6	76.5				
	REM	1.1	16.1	35.6				47.2		180
		2.9	21.0	11.1				90.4		
Column Total (N)		70	138	576	128	17	94	7		1030
Old	0	54.9	22.9	7.8			3.3	11.1		153
		60.9	16.1	2.6			6.2	89.5		
	1	20.0	32.2	31.1			16.7			
		13.0	13.3	6.0			18.8			
	2	4.5	20.1	68.7	4.0		2.6			
		17.4	49.3	78.1	19.4		17.5			
	3	1.6	15.3	19.4	61.3	2.4				124
		1.4	8.8	5.2	70.4	6.2				
	4	1.8			19.3	78.9				57
		0.7			10.2	93.8				
	REM	7.4	22.1	31.1			37.7	1.6		122
		6.5	12.4	8.1			57.5	10.5		
Column Total (N)		138	217	466	108	48	80	19		1076

concerning the beginning and ending of REM periods. Of 42 REM periods lasting 6 epochs or longer as scored by the authors, the automatic system completely missed 14. These misses probably were due to an overly conservative criterion for EMG magnitude in the automatic system.

Disagreements in the scoring of stages 3 and 4 possibly are of interest because this classification should be the easiest to automate. The scoring manual<sup>5</sup> states that a slow wave, of less than 2 Hz, having a magnitude greater than 75  $\mu$  volts represents delta activity (providing it is not a K-complex). If an epoch has more than 50 percent delta activity, it should be scored stage 4, and if an epoch contains between 20 and 50 percent delta activity, it should be scored stage 3. This rule was followed exactly in the computer except that different voltage thresholds had to be used because of the wide disparity with respect to amplitude between subjects. The visual scorers presumably also had to use different thresholds for different subjects, and it is clear that, if the human was using the same threshold as the computer, agreement should be nearly 100 percent in the scoring of stages 3 and 4. It is equally clear that, if the human and computer thresholds were different, consistent misclassifications would result. That is, if the computer required a higher voltage for a slow wave to be labelled delta than did the human scorer, the automatic system would sometimes misclassify stage 4 as stage 3, and stage 3 as stage 2, but never in the opposite direction. However, analysis of the data showed that misclassifications in both directions were made. This finding suggests that the disagreements between the automatic system and visual scoring were due not only to the choice of the wrong threshold levels in the automatic system, but also to inconsistencies in choice of thresholds among the human scorers.

#### IV DISCUSSION AND CONCLUSION

It is clear that neither the EEG sleep records scored visually by the Oklahoma group nor those scored by computer showed high agreement with scores visually assigned by the authors of this paper. The agreement (about 73 percent over all age groups and sleep stages) found between records scored visually is of the order that might be expected on the basis of other studies<sup>4</sup> and in light of the fact of the limited experience of the scorers with the sleep EEGs of the age groups studied (see Ref. 2, Table 3). Our experience suggests that the criteria described by Rechtschaffen and Kales<sup>5</sup> are not applicable to all age groups of subjects. A case in point, possibly extreme, may be the two-second burst of spindle-like beta activity seen in one of the 69-year-old men as was noted above. Similar activity was seen also in two of the other old subjects, but the segments of their records containing the activity were not analyzed as a part of this report.

The agreement between sleep stage assignments by the automatic system and those by human scorers was not sufficiently high to justify use of the automatic system (in its current state) in a sleep research program. In searching for ways to improve the system it is reasonable to ask how well the automatic scoring system does if we ignore the stage awake versus stage 1 discrimination, which was not of particular interest in this study, and also ignore the stage 3 versus stage 4 discrimination, which we know can be made by computer as well as it is made by humans for a specified amplitude criterion of delta activity. These data, shown in Table V, indicate a marked superiority in agreement between human scorers

Table V

AGREEMENT (IN PERCENT) BETWEEN GROUPED SLEEP STAGES  
AS CLASSIFIED BY THE AUTHORS, THE UNIVERSITY OF OKLAHOMA  
SLEEP LABORATORY, AND THE COMPUTER FOR SUBJECTS  
IN THREE AGE GROUPS

Group	Sleep Stage As Scored By Authors	Sleep Stage as Scored in Oklahoma Laboratory					Row Total (N)	Sleep Stage as Scored by Computer					Row Total (N)
		0 and 1	2	3 and 4	REM	? and X		0 and 1	2	3 and 4	REM	? and X	
Children	0 and 1	86.8	6.6	4.1	1.7	0.8	121	76.9	11.6	4.1	7.4		121
		76.6	2.1	1.1	0.8	100	24.6	3.9	2.3	3.4			
	2	3.7	63.3	32.3	0.7		561	36.5	52.9	3.3	7.3		561
		15.4	92.2	40.0	1.7		54.0	82.9	8.8	15.5			
	3 and 4	0.4	0.4	99.2			267	7.9	15.0	71.2	5.9		267
		0.7	0.3	58.5			5.5	11.2	88.0	6.1			
	REM	3.7	7.9	0.8	87.6		267	22.5	2.6	0.7	74.2		267
		7.3	5.4	0.4	97.5		15.9	2.0	0.7	75.0			
Column Total (N)		137	385	453	240	1	1216	378	358	216	264		1216
Middle-Aged	0 and 1	79.1	6.4	0.6	8.2	5.7	158	50.6	37.4	1.9	5.7	4.4	158
		71.8	1.9	0.9	6.6	60.0	38.5	10.2	2.1	9.6	100		
	2	8.5	87.0	2.9	1.2	0.4	577	14.0	76.1	9.9			577
		28.2	94.5	14.9	3.6	13.3	38.9	76.2	39.3				
	3 and 4		15.7	83.5		0.8	115	13.9	12.2	73.9			115
			3.4	84.2		6.7	7.7	2.4	58.6				
	REM		0.6		97.8	1.6	180	17.2	35.6		47.2		180
			0.2		89.8	20.0	14.9	11.2		90.4			
Column Total (N)		174	531	114	196	15	1030	208	576	145	94	7	1030
Old	0 and 1	76.5	5.8		4.1	13.6	243	68.3	16.5		8.2	7.0	243
		50.8	4.1		8.5	70.2	46.8	8.6		25.0	89.5		
	2	32.1	60.2	6.0		1.7	530	24.7	68.7	4.0	2.6		530
		46.4	92.2	16.1		19.2	36.9	78.1	13.5	17.5			
	3 and 4	0.6	7.2	92.2			181	12.2	13.3	74.5			181
		0.3	3.7	83.9			6.2	5.2	86.5				
	REM	7.4			88.5	4.1	122	29.5	31.2		37.7	1.6	122
		2.5			91.5	10.6	10.1	8.1		57.5	10.5		
Column Total (N)		366	346	199	118	47	1076	355	466	156	80	19	1076



(80 percent over all subjects) over the agreement between the automatic system and human scorers (66 percent over all subjects).

The problems in the automatic scoring system can probably be solved in several steps. First of all, more positive identification of stage awake and sleep stage 1 should be implemented. This may require the use of additional EEG leads placed to provide good evidence of alpha activity as is found in the resting subject with closed eyes, and also the examination of the eye channels for "slow rolling" activity, which tends to occur in sleep stage 1.

Secondly, stage 2 can be discriminated more positively by identifying K-complexes and ensuring that they are not mistaken for delta activity. When using the system with many different ages of subjects, accurate scoring of stage 2 may require tuning the center-frequency of the spindle activity filter (set at about 14.5 Hz for all subjects in this study) for each subject.

Finally, detection of stage REM must be improved. A significant improvement probably can be obtained by filtering the EMG channel so as to improve its signal-to-noise ratio and by using a more elaborate algorithm to determine those epochs marking the beginning and end of the REM periods.

The evidence obtained in this study suggests that with the improvements outlined above, agreement comparable to that found between human scorers should be obtainable in an automatic real-time classification system.

## REFERENCES

1. A. Lubin, L. C. Johnson, and M. T. Austin, "Discrimination Among States of Consciousness Using EEG Spectra," Psychophysiology, Vol. 6, pp. 122-132 (1969).
2. L. J. Monroe, "Inter-Rater Reliability and the Role of Experience in Scoring EEG Sleep Records, Phase I," Psychophysiology, Vol. 5, pp. 376-384 (1969).
3. J. R. Smith and M. Negrin, "Automatic Analysis of Sleep Electroencephalograms by Hybrid Computation," Technical Paper No. 421, Florida Engineering and Industrial Experiment Station, University of Florida, Gainesville, Florida (September 1968).
4. H. L. Williams, University of Oklahoma Medical School, (personal communication, 1969).
5. A. Rechtschaffen and A. Kales (eds.), A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects, Publication No. 204, National Institute of Public Health, U.S. Department of Health, Education, and Welfare (1968).
6. E. Kahn and C. Fisher, "The Sleep Characteristics of the Normal Aged Male," J. Nervous Mental Disease, Vol. 148, pp. 477-494 (1969).
7. J. S. Lukas, Mary E. Dobbs, and K. D. Kryter, "Disturbance of Human Sleep by Subsonic Jet Aircraft Noise and Simulated Sonic Booms," prepared for the National Aeronautics and Space Administration, Contract NAS1-9286, SRI Project 8027, Stanford Research Institute, Menlo Park, California (August 1970).

Appendix

DETAILED DESCRIPTION OF THE COMPUTER PROGRAM

Function	Computer Program Description
<u>Step 1</u> --To control sampling and digitizing	Each of the eight analog channels from each subject is sampled and digitized (see Figure 3 and accompanying discussion). The 15-bit 2's complement number from the digitizer is shifted to produce a 16-bit word, including sign, in the CE-16. All data (32 words) are stored in reserved locations in page 0 of memory.
<u>Step 2</u> --To detect and measure delta activity (0-2.5 Hz) and to eliminate artifacts due to movement, GSR, etc. for subject 1	The sign of a newly-arrived sample from the subject of interest is compared with the sign of his previous 0-2.5 Hz sample. If the sign is the same, a time counter is incremented (a one is added to the previous number) and the absolute amplitude of the sample is compared with the largest amplitude found during the period in which the samples have had this sign. If the sample in question has a greater amplitude, the maximum amplitude is changed to this new value. However, if the sign is different from that of the previous sample, the time counter is checked to determine (1) that at least 150 milliseconds and not more than 1 second was spent in the previous time state, and (2) that the absolute maximum amplitude obtained during the previous sample period is greater than a particular threshold* (H01), set uniquely for each individual. If both conditions are met, it is assumed that one half of an EEG slow wave of frequency .5 to 3.0 Hz has been detected and the amount of time spent in the previous sample period is added to a reserved word describing the amount of time spent by this person in sleep stage delta in each epoch. If either the time or maximum amplitude criteria are not met, then the time in the previous sign state is considered not to have been delta time (it may have been an artifact or some other activity) and the step is finished.
<u>Step 3</u> --To detect alpha activity (9-12 Hz) and discriminate it from activity in the 3-7 Hz and 13-16 Hz bands in subject 1	First, the amplitude of a newly-obtained 8-12 Hz sample is added to a double precision number (a number of 31 bits rather than 16 bits) containing the sum of the amplitudes in the 8-12 Hz samples of each epoch. Next, the 8-12 Hz sample is compared to an individual's threshold (H8) for that frequency band. If the 8-12 Hz sample amplitude is greater than its threshold, in order to assure that the amplitude of the 8-12 Hz signal is not due to crosstalk from adjacent bands, the 3-7 Hz sample is compared to its (3-7 Hz) threshold (L83), and the 13-16 Hz sample is compared to its (13-16 Hz) threshold (L813). If the amplitudes in both adjacent bands are less than their thresholds, the word keeping track of the number of samples in an epoch that show significant alpha activity is incremented.
<u>Step 4</u> --To detect spindle activity (13-16 Hz) and discriminate it from alpha activity in subject 1	This step is very similar to Step 3. First, the 13-16 Hz sample amplitude is added to a double precision number containing the sum of such numbers in the epoch. Second, the 13-16 Hz sample and 8-12 Hz sample are compared to the thresholds for those bands. If the 13-16 Hz sample is higher than its threshold (H13) and the 8-12 Hz signal is lower than its threshold (L138), a counter indicating spindle activity is incremented.
<u>Steps 5, 6, 7</u> --To add the amplitudes obtained in the samples of the 3-7 Hz, 16-40 Hz bands, and the EMG activity to those obtained earlier in an epoch (subject 1)	The sampled amplitudes of signals in the 3-7 Hz and 16-40 Hz bands are added to double precision numbers containing the sums of each kind of activity during an epoch. In like manner, the amplitude of the EMG signal is added to its previously obtained level.

\* It is important to note that the threshold values vary as some function of the individual and the stage (step) of data processing. In order to illustrate this point the threshold values used in this study are presented in Table A-I. The values are arbitrary since they depend upon the particular amplifiers, attenuators, filters, etc., used in any laboratory. They are presented here to illustrate the relative magnitudes of threshold voltages observed between the three age groups and between individuals within an age group.

Step 8--To detect EOG (electro-oculographic) signals representative of those found during the REM (rapid eye movement) sleep stage in subject 1

The signs of the two EOG samples are compared to determine (1) if they are of opposite polarity, and (2) if their amplitudes exceed a threshold (HEY). If both criteria above are met, and if the EMG sample is below a specified threshold (LYE), a temporary counter is incremented. If the criteria are not met, the same temporary counter is checked to determine if the criteria were met in the immediately preceding samples and if so to determine if less than one second was spent in the desired state. That is, we check whether an eye movement has occurred and whether it was a rapid eye movement (REM). If there has been a REM, the amount of time spent in the REM is added to a word that keeps track of potential REM time in an epoch. The temporary counter is then set to zero and the step is finished.

Step 9--To obtain and process data from subjects 2, 3, 4

Repeat Steps 1 to 8 for 2nd, 3rd, and 4th subjects

Step 10--To check for a new epoch

Epoch timing is accomplished via an external interrupt system (see Figure 1). Every 20 seconds a pulse is generated externally and is put on one of the interrupt lines of the CE-16. The pulse causes the computer to stop its ordinary schedule of processing for 100 microseconds and to increment a special-reserved word in page 0 of memory. At step 10 in the processing schedule, this special word is examined to determine if it has changed since the last time Step 10 was executed. If so, processing is continued with Step 14; otherwise Steps 11, 12, and 13 occur.

Step 11--To monitor, control, and command printing of information on the teletype

When the sleep-scoring program is operated in the mode in which sleep stages for each of the four subjects are printed out and in which detailed information is printed out for one subject, approximately 60 characters of information must be printed out in each epoch. Since the teletype printer can print at the rate of ten characters per second, six seconds are required to print the 60 characters. However, it is not necessary to cut this much time out of each 20-second epoch, since once the computer has told the teletype what to print--which requires only a few microseconds--it can continue to do other processing while the teletype is printing the character. Thus this step in the program is occupied by a routine that (1) determines if anything more needs to be printed from the last epoch, (2) if so, it checks to see if the teletype is still busy typing the last character, (3) if not, it sends the next character to be printed to the teletype.

Step 12--To control the interval between successive repetitions of Step 1

A delay is inserted to cause the time between two successive executions of Step 1 to be 50 milliseconds, thus creating a sampling rate of 400 times per second.

Step 13--To return processing to Step 1 for subjects 2, 3, and 4

Command the computer to return and execute Step 1.

Step 14--To calculate the parameters used in deciding the sleep stage of subject 1

In this step certain percentages and averages to be used in Step 16 to determine the stage of sleep are calculated. The parameters calculated and the steps from which they were obtained are: (a) percentage of time spent in potential delta activity--based upon Step 2; (b) percent time spent in potential alpha activity--based upon Step 3; (c) percent time spent in potential spindle activity--based upon Step 4; (d) percent time spent in potential REM--based upon Step 8; (e) average absolute amplitudes of activity in the 3 to 7 Hz, 8 to 12 Hz, 13 to 16 Hz, 16 to 40 Hz bands, and in the EMG channel; and (f) average absolute amplitude of 3 to 7 Hz activity divided by the average absolute amplitude of 8 to 12 Hz activity (the mean theta to mean alpha ratio). Percentages and averages, rather than simple sums, are required since slightly different numbers of samples may be obtained in different epochs. Although the parameters listed above are calculated for each of the four subjects, they are printed out only for a single subject because of the limited typing speed of the teletype. Clearly, these data could be stored on magnetic or paper tape.

Step 15--To calculate decision parameters for subjects 2, 3, and 4

Repeat Step 14 for 2nd, 3rd, and 4th subject.

Step 16--To calculate parameters for subjects 2, 3, and 4

Classification is based on the parameters calculated in Step 14 (Step 15 for subjects 2, 3, and 4), and proceeds as follows:

- (a) Compare the average EMG value in the epoch with the individual's threshold (EM). If it is less than EM, go to Step d.
  - (b) Compare the average beta (16 to 40 Hz) level in the epoch with the individual's threshold (BM). If it is less than BM, go to Step d.
- Decision: (c) Store the number 10 in two epoch counters (CNT2 & CNT5). Call the state of sleep 6 (movement time) and go to Step ee.
- Movement Time
- (d) Compare the percentage of potential spindle activity in the epoch to the individual's threshold (S2). If it is less than S2, go to Step f.
  - (e) Put the number 0 in the epoch counter CNT2.
  - (f) Compare the percentage of delta activity in the epoch to the individual's threshold (D4--It is equal to 50 percent for all 12 subjects). If it is less than D4, go to Step j.
  - (g) Compare the average beta level to the individual's threshold (B4). If it is greater than B4, go to Step n.
  - (h) Compare the average theta to average alpha ratio in the epoch to the individual's threshold (R4). If it is less than R4, jump to Step n.
- Decision: (i) Call the stage of sleep 4 and go to Step ee.
- Sleep Stage 4
- (j) Compare the percentage of delta activity in the epoch to the individual's threshold (D3--It is equal to 20 percent for all 12 subjects). If it is less than D3, go to Step n.
  - (k) Compare the average beta level to the individual's threshold (B3). If it is greater than B3, go to Step n.
  - (l) Compare the average theta to average alpha ratio to the individual's threshold (R3). If it is less than R3, go to Step n.
- Decision: (m) Call the stage of sleep 3 and go to Step ee.
- Sleep Stage 3
- (n) Store the number zero in the characteristic counter, NCHAR. Compare the percentage of potential spindle activity to the individual's threshold (S2). If it is greater than S2, increment NCHAR.
- Compare the theta to alpha ratio to the individual's threshold (R2). If it is greater or equal to R2, increment NCHAR.
- Compare the average EMG value with the threshold (E2). If it is less than or equal to E2, increment NCHAR.
- Compare the average beta output with the individual's threshold (B2). If it is less than or equal to B2, increment NCHAR.
- If NCHAR is not equal to 4, go to Step p.
- Decision: (o) Store the number zero in the epoch counter CNT2. Store the number ten in the epoch counter CNT5. Call the stage of sleep 2 and go to Step ee.
- Sleep Stage 2
- (p) Compare the average EMG to the individual's threshold E5. If it is greater than E5, go to Step u.
  - (q) Store the number -1 in the state counter PREV.
  - (r) Compare the percentage of potential REM activity to the individual's threshold I5. If it is less than I5, go to Step w.
  - (s) Store zero in the epoch counter CNT5.
- Decision: (t) Call the stage of sleep REM and go to Step ee.
- Sleep Stage REM
- (u) Add 1 to the state counter PREV.
  - (v) If PREV = 0, go to Step (r), else go to Step x.

(w) If the epoch counter CNT5 is less than 10, go to Step t.  
 (x) If NCHAR is less than 2, go to Step aa.  
 (y) If the epoch counter CNT2 is greater than 9, go to Step aa.  
 Decision: (z) Call the stage of sleep 2 and go to Step ee.  
 Sleep  
 Stage 2 (aa) Compare the percentage of potential alpha activity with the individual's threshold AO. If it is less than AO, go to Step cc.  
 Decision: (bb) Call the stage of sleep 0 and go to Step ee.  
 Sleep  
 Stage 0 (cc) Compare the average theta to average alpha ratio to the individual's threshold RO. If it is less than RO, go to Step bb.  
 Decision: (dd) Call the stage of sleep 1.  
 Sleep  
 Stage 1 (ee) Increment the epoch counters CNT2 and CNT5.

Step 17--To classify the sleep stages of subjects 2, 3, and 4

Repeat Step 15 for subjects 2, 3, and 4.

Step 18--To determine the subject about whom detailed information is to be printed for this epoch

Has detailed information been provided on the same subject for the last 16 successive epochs? If not, transfer the current subject's parameters to the output buffer as follows: (1) percentage of epoch spent in delta activity, (2) percentage of epoch spent in alpha activity, (3) percentage of epoch spent in spindle activity, (4) percentage of epoch showing eye movement, (5) mean theta level during epoch, (6) mean alpha level during epoch, (7) mean spindle level during epoch, (8) mean beta level during epoch, (9) mean EMG level during epoch, (10) theta to alpha ratio during epoch.

If 16 epochs have been devoted to this subject, rotate to new subject in circular order 1 → 2 → 3 → 4 → 1. A change can also be accomplished by manually setting a switch that causes a computer interrupt that informs it to transfer to the next subject in order.

Step 19--Augment the state-transition matrix

A matrix is maintained in memory that is printed out at the end of the night. This matrix contains for each subject the number of times in the night in which a subject went from any given stage to any other stage including the same stage. Thus, in this step, the i, j, k cell of the three-dimensional matrix is incremented where i is the subject number, j is subject i's prior state of sleep, and k is his current stage of sleep. In this study the epochs occurring in the three minutes following the presentation of an acoustic stimulus to a subject were not included in the matrix.

Step 20--Finish prior epoch processing

Make all the memory cells containing sums and amounts of time spent in various activity for the previous epoch 0 and go to Step 1.

Table A-1

RELATIVE THRESHOLDS USED IN THE COMPUTER PROGRAM  
TO DISCRIMINATE BETWEEN SLEEP STAGES

Threshold Designator	Subject Number and Age (in Years)												Threshold Designator
	Old				Middle-Aged				Young				
	1-70	2-69	3-75	4-74	1-45	2-53	3-45	4-57	1-8	2-5	3-5	4-6	
H01	1.367	1.367	1.367	1.367	1.367	1.367	1.367	1.367	1.094	1.172	1.172	1.172	H01
H8	2.500	2.500	2.500	2.500	2.656	2.500	2.500	2.500	2.500	2.500	2.500	2.500	H8
L83	4.375	4.375	4.375	4.375	4.375	4.375	4.375	4.375	4.375	4.375	4.375	4.375	L83
L813	2.188	2.188	2.188	2.188	2.109	2.188	2.188	2.188	2.188	2.188	2.188	2.188	L813
H13	1.084	1.250	1.094	1.094	1.094	1.016	1.172	1.094	0.625	1.250	0.625	0.625	H13
L138	1.250	2.500	1.250	1.250	1.250	1.328	1.250	1.250	1.250	1.250	1.250	1.250	L138
HEY	0.625	0.664	0.625	0.625	0.625	0.781	0.781	0.625	0.625	0.625	0.625	0.625	HEY
LYE	0.313	0.313	0.313	0.313	0.313	0.293	0.313	0.313	0.313	0.313	0.313	0.313	LYE
EM	2.500	2.734	2.734	2.961	2.681	2.891	2.969	2.813	2.813	2.813	2.813	2.813	EM
BM	2.891	2.891	4.375	3.828	3.984	3.047	3.984	4.180	3.906	3.906	3.906	3.906	BM
B4	2.500	2.559	3.828	2.813	2.969	2.656	1.875	3.984	1.563	2.266	1.563	2.373	B4
B3	2.500	2.559	3.828	2.813	2.969	2.656	1.875	3.984	1.563	2.266	1.563	2.373	B3
B2	1.250	2.924	3.281	2.500	2.422	2.627	2.031	3.594	1.099	2.266	1.099	1.099	B2
E2	0.625	0.781	1.250	0.938	0.469	0.625	1.094	1.094	0.908	0.693	0.776	1.245	E2
E5	0.391	0.186	0.313	0.313	0.273	0.181	0.181	0.273	0.234	0.234	0.234	0.234	E5
R4	2.56	2.256	1.984	1.904	1.600	1.472	1.440	1.120	2.56	2.56	2.56	2.56	R4
R3	2.56	2.256	1.984	1.904	1.600	1.472	1.440	1.120	2.56	2.56	2.56	2.56	R3
R2	2.084	1.904	1.792	1.328	1.536	1.456	1.312	1.024	1.664	1.664	1.664	1.664	R2
R0	2.096	1.280	1.664	1.408	.960	.944	1.795	1.024	2.56	2.432	1.696	1.616	R0
D4	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	D4
D3	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	D3
S2	12.80	24.00	24.00	24.00	24.00	20.80	25.60	25.60	12.80	12.80	12.80	12.80	S2
I5	3.20	3.20	3.00	3.20	3.20	3.40	3.00	3.00	3.20	3.20	3.20	3.20	I5
A0	.3	.3	.3	.3	.3	.3	.3	.3	.3	.3	.3	.3	A0

Voltage Thresholds

Ratio Thresholds

Percentage Thresholds