Computerized Analysis of Pupillograms in Studies of Alertness

Jay W. McLaren, Jay C. Erie, and Richard F. Brubaker

When alert subjects sit in a quiet, darkened room for 10–15 min, their pupils remain dilated. If the subjects become sleepy or fatigued, their pupils become miotic and oscillate. Pupillometry, the recording of pupil diameter, has been used to study alertness. Pupillograms are typically graded by subjective assessment on the basis of oscillations of pupil diameter (hippus) and miosis. In this study, numeric parameters were calculated from pupillograms and were compared to subjective ratings by clinicians. Pupil diameters were recorded for 15 min at five samples per second with a custom built video pupillometer. Subjects sat in the dark and fixated a small red light. Digital filtering techniques and Fourier analysis were used to calculate several parameters designed to report hippus and miosis. The same set of pupillograms was graded by three physicians familiar with pupillometry. Grades were assigned on a scale of 1 to 4, 1 being most alert and 4 least alert. A linear combination of three of the numeric parameters had the highest correlation with the average subjective grade (r = 0.92). These techniques provide a quantitative way to evaluate pupillograms that will be used in the assessment of alertness. Invest Ophthalmol Vis Sci 33:671–676, 1992

When alert subjects sit in a quiet darkened room, their pupils remain dilated. If they are fatigued or sleepy, however, their pupils become miotic and oscillate.[1-10] The pupillogram (recording of pupil diameter) has been used to study alertness of individuals with sleep disorders[1,4-7,9,10] and to evaluate the efficacy of medication for treatment of hypersomnolence.[1] Some have suggested that pupillography could be used in diagnosing narcolepsy[1] and testing the ability of automobile drivers[11] and commercial airline pilots to remain alert.[12]

Pupil diameters typically have been recorded for 10–15 min while the subject sits in a darkened room and looks at a small fixation light.[1,4] Investigators have appraised pupillograms subjectively by assessing hippus (oscillation of pupil diameter) and the severity and onset of miosis.[9,10] Their appraisal could be made simpler and more objective if a computerized system were available to calculate numeric parameters based on the intensity and time course of hippus and miosis.

In this study, we developed and implemented several algorithms to analyze pupillograms numerically.

The efficacy of these methods was tested by comparing numeric parameters to subjective assessments of the same pupillograms by three physicians.

Materials and Methods

Pupillograms were recorded for 15 min in the dark with a custom designed video pupillometer. One hundred and sixty seven subjects were recruited among employees and students at Mayo Clinic and from the population of Rochester, Minnesota. Each completed a questionnaire to evaluate sleep habits and was asked to wear a wrist actigraph (Ambulatory Monitoring Inc., Ardsley, NY) for three days before the study to evaluate the quantity and efficiency of sleep. This group included 75 males and 92 females. Ages ranged from 7–82 yr with a median age of 33 yr. Consent to participate in pupillometry was obtained from each subject after the nature of the procedure had been explained fully.

Subjects were seated in the pupillometer in a quiet darkened room. Ambient light was less than 0.01 footcandles. They were allowed to dark adapt for 2 min before pupil diameter was measured continuously for 15 min. Subjects fixated a small red spot of light approximately 2 m from their eyes. The alignment of the infrared cameras was maintained with the help of a bridgepiece that met the bridge of the nose.

Although two or more pupillograms were recorded on separate occasions, only one was used for comparison of quantitative parameters with subjective grade.

671

Subjective Assessment of Pupillograms

Pupillograms were graded on an integer scale of 1 (most alert) to 4 (least alert) by three physicians. Each physician was familiar with pupillometry and had used pupillograms in their evaluations of alertness. All identifying information had been removed from the printed graphs so subjects could not be identified from their pupillograms. After at least one week, the set of pupillograms was graded a second time. The order of the pupillograms was changed between the first and second grading. The overall grade assigned to each pupillogram was the average of the six grades (two from each physician, three physicians).

A pupillogram was classified as uninterpretable if the recording was incomplete, the pupil diameter was obliterated by excessive eyelid closures, or for any technical reason the data could not be evaluated. These were excluded from further analysis.

Computerized Analysis of Pupillograms

Eyelid closure (blink) artifact: Several algorithms were designed to extract information from recordings of pupil diameter. Before these could be applied, data recorded during eyelid closures had to be removed. An algorithm was written to search for the beginning and end of each closure, and linear extrapolation was used to replace pupil diameters during this interval. The technique is described here briefly. Each pupil diameter was compared to the previous, and if the difference was greater than the maximum of 0.4 mm or 1.5 standard deviations of the previous six measurements, the point marked the beginning of a closure. The end of the closure was found by searching for the next sequence of six measurements that had a standard deviation of less than 0.4 mm. The first point after the beginning of the closure that differed by less than 0.4 mm from the mean of these six points was assumed to be the first valid measurement after the eyelids reopened. A simple linear extrapolation was used to approximate pupil diameters while the eyelids were closed, and these were substituted for the invalid data in later analyses. The criteria (0.4 mm and 1.5 standard deviations) were determined empirically. Any pupil diameter less than 1.0 mm was assumed to be invalid because it was outside the linear range of the instrument. These data were not used in searching for stable measurements after the closure.

Parameters to evaluate hippus: Hippus is the oscillation of pupil diameter. The magnitude of oscillation can be calculated as a function of frequency by using the Fourier transform. The frequency spectrum of each pupillogram was computed by using a digital fast Fourier transform algorithm. This transform operated on a sequence of N measurements, where N is an

integer power of 2.[13,14] A time of 51.2 sec (256 measurements at 0.2 second intervals) was the time segment chosen to be transformed. With these parameters, the frequency resolution was 0.02 Hz and the maximum detectable frequency was 2.5 Hz. A 15 min pupillogram was composed of 17 full 51.2 sec segments and a remainder of 29.6 sec. Data from this final short interval was not included in the averages of any parameters.

The first parameter designed to report hippus was derived from the Fourier-transformed pupillogram and was the area of the frequency spectrum in a selected band of frequencies. This parameter will be denoted $FFT(f_1,f_2)$ where f_1 and f_2 represent the lower and upper frequency bounds. $FFT(f_1,f_2)$ was calculated over several frequency ranges on each 51.2 sec segment and averaged over all 17 segments. The parameter was calculated over several frequency ranges. The Fourier transforms of three subgroups of subjects consistently graded 1, 2, and 3 by all three physicians were studied to determine which frequencies were most important in judging pupillograms.

A second parameter to report hippus also was investigated. In principle, this parameter was similar to $FFT(f_1,f_2)$, but was simpler to calculate because it did not require the use of the Fourier transform. Pupillograms were filtered with a digital bandpass filter (third order elliptical filter[15]), and the energy of the filtered sequence was calculated:

$$SS(f_1,f_2) = \sum_{i=1}^{k} d_i^2 \qquad (1)$$

where $SS(f_1,f_2)$ is the energy of the sequence in the passband of the filter with lower and upper frequency cutoffs f_1 and f_2, d_i is the ith filtered pupil diameter, and k is the number of points in the sequence (256 for 51.2 sec). The term "SS" (sum of squares) was adapted because this parameter is numerically equal to the sum of pupil diameter squared, after bandpass filtering. SS (f_1,f_2) was calculated over each time segment, and the average was determined over the 17 full segments of each pupillogram. For comparison, the same 51.2 sec time segments used in the Fourier analysis were used when calculating this and other parameters.

Parameters to evaluate miosis: Miosis is another subjective cue that has been used as an indicator of decreased alertness. Three parameters were designed to report miosis. All were based on the average pupil diameter calculated over each 51.2 sec segment.

First, a parameter that will be referred to as "cumulative miosis" (CM) was determined:

$$CM = \sum_{i=1}^{17} (1 - PDR_i) \qquad (2)$$

where PDR_i is the ratio of the average pupil diameter during the ith time segment to the average pupil diameter during the first time segment. Subtraction of PDR_i from 1 forces this parameter to increase as miosis becomes stronger.

A second parameter was used to describe how soon the average pupil diameter decreased. This simply was the number of intervals that remained after the average pupil diameter dropped to below 80% of the initial diameter (PDR_i dropped below 0.8). This parameter ranged from 0 to 17 and increased as miosis occurred sooner.

Third, the minimum ratio of average pupil diameter to the initial pupil diameter was calculated and will be denoted by PDR_{min}. All programs used to calculate parameters were written in Quick BASIC (Microsoft Corporation, Redmond, WA) on an IBM model PS2 microcomputer (IBM Corporation, Boca Raton, FL).

Comparison Between Quantitative Parameters and Average Grade

Our goal was to find numeric parameters consistent with subjective assessment. The relationship between the numeric parameters and average subjective grade was studied by using linear regression analysis. In some cases, the log of the parameter was used to linearize the relationship. A stepwise regression was used to determine if the combination of more than one parameter would provide a better prediction of grade than any single parameter. All statistics were calculated on the IBM PS2 with the program STATPAC (Walonick Associates, Inc., Minneapolis, MN).

Results

Subjective Grades

Pupillograms of only 146 of the 167 subjects were considered by all graders to be valid. The remaining 20 could not be interpreted by at least one grader and were not included in this analysis. The primary reasons for rejection were excessive eyelid activity that

Table 1. Agreement between first and second grade of same grader

| | Percent of total grades assigned | | | |
| | Grade assigned first | | | |
Grade assigned second	1	2	3	4
1	38.4	3.7	0.2	0
2	7.5	29.5	2.5	0.2
3	0.2	5.5	10.7	0.9
4	0	0	0.7	0

Table 2. Agreement between graders

| | Percent of all comparisons | |
Difference between grades	First grade assignment	Second grade assignment
0	68.0	69.4
1	30.6	30.1
2	1.4	0.5
3	0	0

obliterated pupil diameter and failure to complete the 15 min session.

The subjective grade used in the correlation analysis was the average of six grades. Grades assigned by an individual physician were not always consistent with those assigned by the others, and grades assigned on the first assessment were not always the same as those assigned by the same physician on the second assessment. One consequence of this disagreement was that some of the averaged grades assumed noninteger values.

Consistency between the first and second grade assigned by each grader is illustrated by the matrix shown in Table 1. Percentages on the diagonal of the matrix indicate agreement between the first and second grade, and off-diagonal percentages indicate differences. Approximately 79% of the grades did not change, and 21% differed by 1 grade. Fewer than 1% differed by 2 grades.

Table 2 shows differences between grades assigned to the same pupillogram by each pair of graders (grader 1 vs 2, 1 vs 3, and 2 vs 3). Approximately 69% of the grades were identical to those of other graders. About 30% differed by one grade and fewer than 1% differed by two grades.

Hippus

The averaged frequency spectra of each of three groups of subjects consistently graded 1, 2, and 3 are shown in Figure 1. Spectra differed by the most at low frequencies. At higher frequencies, differences diminished. For this reason, several low frequency bands were used when calculating $SS(f_1,f_2)$ and $FFT(f_1,f_2)$.

The correlation between average grade and $FFT(f_1,f_2)$ and $Log[SS(f_1,f_2)]$ was highest at low frequency bands and decreased as frequency increased. For example, the correlation coefficient of average grade with FFT (0.02, 0.04) was 0.773, but decreased to 0.493 when f_1 was 0.25 and f_2 was 0.27 Hz. Correlations of grade with FFT (0.02, 0.04) and Log [SS(0.02, 0.04)] are illustrated in Figures 2 and 3.

Miosis

Cumulative miosis had a slightly stronger correlation with grade than either of the parameters designed
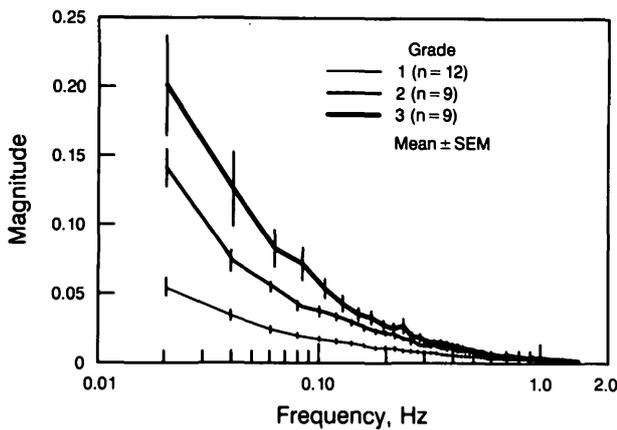
AVERAGE SPECTRA

Fig. 1. Frequency spectra of pupillograms graded consistently 1, 2, or 3. At low frequencies spectra differed by the most. As frequency increased, spectra converged. For this reason, parameters for reporting hippus were calculated over low-frequency bands.

AVERAGE GRADE vs. LOW-FREQUENCY OSCILLATIONS
Sum of Squares

Fig. 3. Sum of squares, a simplified parameter to report hippus. This parameter reports the energy in the pupillogram after band-pass filtering between 0.02 Hz and 0.04 Hz. It is closely related to the parameters based on the Fourier transform, but simpler to calculate.

to measure hippus (Fig. 4). Minimum average pupil diameter was greater than one half of the initial average pupil diameter in all subjects. Unlike the other parameters, PDR_{min} was inversely related to average grade (Fig. 5).

Time remaining after PDR dropped to below 0.8 was not linearly related to average grade (Fig. 6). In subjects graded 1.5 and below, the average diameter remained greater than 80% of the initial diameter throughout the 15 min trial. This parameter showed little relation to average grades above 1.5.

Combination of Parameters

A stepwise regression was used to search for a combination of parameters that would better correlate

with average grade than any single parameter. A combination of FFT (0.02, 0.04), PDR_{min}, and cumulative miosis (Sum[1-PDR]) had a higher correlation with average grade than any single parameter. The best combination was given by:

$$R_1 = 0.114 \times CM - 2.44 \times PDR_{min}$$
$$+ 1.85 \times FFT(0.02, 0.04) + 3.27 \quad (3)$$

where R_1 is the linear combination of the other three parameters. The correlation coefficient of this new parameter with average grade was 0.921 (Fig. 7). When the simpler parameter log[SS(0.02,0.04)] was substituted for the Fourier parameter, new regression coefficients were obtained and the correlation coefficient

AVERAGE GRADE vs. LOW-FREQUENCY OSCILLATIONS
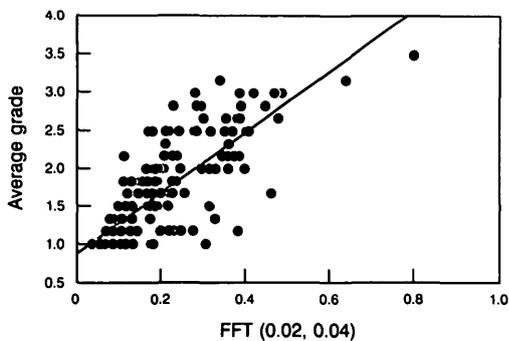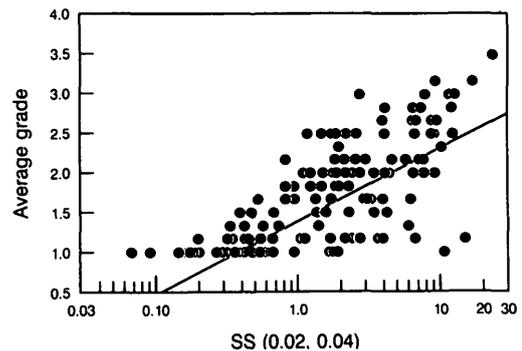### Fourier Transform

Fig. 2. FFT(0.02,0.04), a parameter based on the Fourier transform of the pupillogram. This number is the sum of magnitude between 0.02 Hz and 0.04 Hz (period of oscillation between 51 sec and 25 sec). This passband represents the range of frequencies where the greatest differences between spectra were found. The average grade assigned by physicians increased as FFT(0.02, 0.04) increased.

AVERAGE GRADE vs. CUMULATIVE MIOSIS
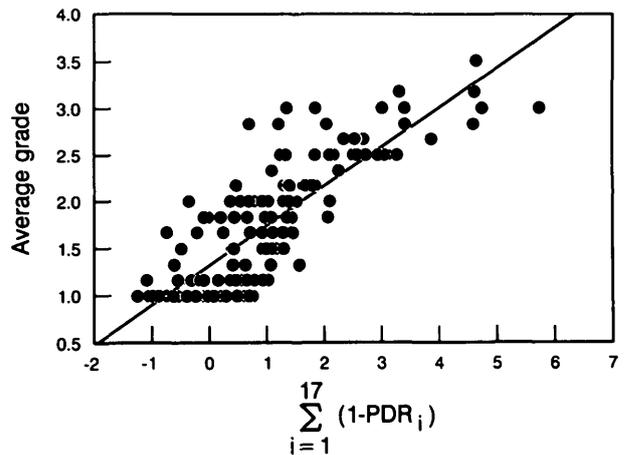

$$\sum_{i=1}^{17} (1\text{-}PDR_i)$$

Fig. 4. Cumulative miosis was greatest for pupillograms that were graded highest.

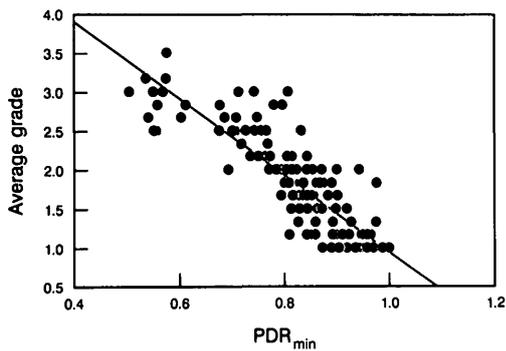## AVERAGE GRADE vs. MINIMUM PUPIL DIAMETER RATIO

Fig. 5. Minimum pupil diameter ratio. PDR_{min} was the minimum ratio of average pupil diameter for any 51.2-sec time segment to the average pupil diameter during the first segment.

was only slightly less, 0.914. The regression formula in this case was:

$$R_2 = 0.141 \times CM - 2.42 \times PDR_{min}$$

$$+ 0.375 \times Log[SS(0.02, 0.04)] + 3.55 \quad (4)$$

Correlation coefficients of average grade with several numeric parameters are shown in descending order in Table 3. The highest correlations were with R_1 and R_2, parameters derived from the linear combination of other parameters.

Discussion

The use of pupillometry to evaluate alertness of normal subjects and patients with sleep disorders has

AVERAGE GRADE vs. NUMBER OF TIME SEGMENTS REMAINING

Time segments remaining after miosis to
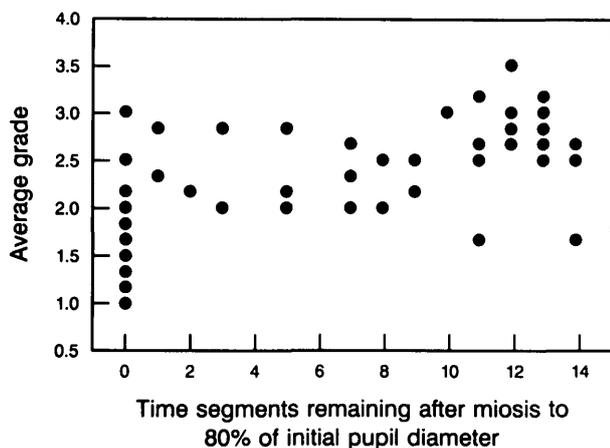80% of initial pupil diameter

Fig. 6. Number of time segments remaining after miosis to 80% of initial pupil diameter. PDR_{min} was greater than 0.8 in all pupillograms with an average grade of 1.5 or less.

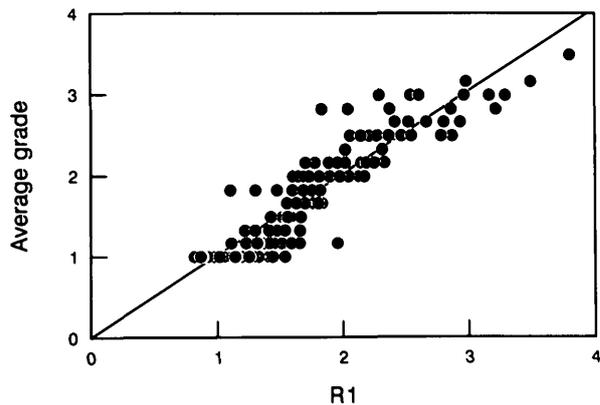## AVERAGE GRADE vs. LINEAR COMBINATION OF FFT (0.02, 0.04), PDR_{min} AND CM

Fig. 7. R_1, parameter based on linear combination of FFT(0.02,0.04), CM, and PDR_{min}. The average grade had a higher correlation with this parameter than with any of the individual parameters.

been based on subjective appraisal of graphs of pupillograms. The appearance of hippus and the presence and degree of miosis have been used as indicators of decreased alertness.[1,4,6,7,9,10] The algorithms used in this study and presented here were designed to calculate numeric parameters that grade hippus and miosis.

An ideal numeric parameter should have several characteristics. It should reduce the information available from the pupillogram to a few numbers. These must use information from the entire pupillogram and report some aspect that is relevant to alertness. The parameter must be easy to obtain. Numeric parameters should be calculated without the use of subjective input from the investigator, other than to check for obvious problems, such as excessive eyelid activity. All useful parameters must be consistent with the collective judgment of trained physicians.

The parameters designed in this study satisfy these requirements. Each is a single number that summa-

Table 3. Correlation of parameters with average subjective grade, in order of highest to lowest correlation coefficient

Parameter	Correlation coefficient
Linear combination, R_1	0.921
Linear combination, R_2	0.914
PDR_{min}	−0.880
CM	0.831
FFT (0.02, 0.04)	0.773
Log [FFT (0.02, 0.04)]	0.768
Log [SS (0.02, 0.04)]	0.714
SS (0.02, 0.04)	0.637
Mean pupil diameter	−0.516

rizes 14.5 min of the pupillogram. Parameters are easy to calculate from the digitized pupillogram. If appropriate algorithms are incorporated in programs that measure pupil diameter, the programs could print the parameters as part of the pupillogram and summary report.

The strong correlation between the subjective grade and numeric parameters indicates that the numeric parameters are consistent with the graders. Therefore the same conclusion should be arrived at when using numeric parameters as when using subjective assessment. Numeric parameters have some advantages over subjective assessment. Because they are continuous rather than discrete, numeric parameters are more precise than subjective parameters. This is important when following alertness of a subject after treatment. Numeric parameters also are consistent, while subjective grades vary from grader to grader and for an individual grader vary from one grading to another, as demonstrated in Tables 1 and 2.

The greatest disadvantage to using computer programs to extract parameters related to alertness is the potential of applying these techniques to invalid data. For example, numeric parameters would be meaningless if the subject's eyes were closed during most of the measurement. Investigators must be alert to these problems. The use of numeric parameters should not replace examination of the pupillogram but should supplement it.

We have not addressed the usefulness of these parameters in distinguishing between normal subjects and subjects with disorders of alertness. If this judgment can be based upon the subjective criteria used by the graders in this study, quantitative parameters can be substituted for subjective appraisal. Numeric parameters should provide a better appraisal of alertness because they represent a continuous grade rather than discrete. A numeric parameter can be followed over time to determine if alertness of a subject under treatment has changed. Subtle differences between groups of subjects with sleep disorders and control subjects may be more efficiently detected by using numeric parameters than by subjective assessment.

Before these parameters can be used to establish a specific diagnosis, carefully controlled studies of appropriate groups of subjects must be performed.

References

1. Yoss RE, Moyer NJ, and Ogle KN: The pupillogram and narcolepsy. A method to measure decreased levels of wakefulness. Neurology 19:921, 1969.
2. Kollarits CR, Lechman J, Kollarits FJ, and Gillin JC: The pupil dark response in narcolepsy. Curr Eye Res 2:261, 1983.
3. Löwenstein O and Friedman ED: Pupillographic studies. I. Present state of pupillography; its method and diagnostic significance. Arch Ophthalmol 27:969, 1942.
4. Schmidt HS and Fortin LD: Electronic pupillography in disorders of arousal. In: Sleeping and Waking Disorders: Indications and Techniques, C Guilleminault, editor. Menlo Park, CA, Addison-Wesley, 1982, pp. 127-143.
5. Schmidt HS and Schrier SS: Comparison of MSLT and pupillometry in assessment of narcolepsy and idiopathic CNS hypersomnolence. Sleep Research 12:281, 1983.
6. Schmidt HS, Jackson EI, and Knopp EW: Electronic pupillography (EPG): Objective assessment of sleepiness and differentiation of disorders of excessive somnolence. Sleep Research 10:48, 1981.
7. Newman J, Broughton R, and Dunkan DW: Pupillometric assessment of excessive daytime sleepiness (EDS) in narcolepsy-cataplexy. Sleep Research 18:276, 1989.
8. Hertz G, Spielman AJ, Hakerem G, and Pressman M: Pupillometry and MSLT: The effects of napping on pupil indication of sleepiness. Sleep Research 17:22, 1988.
9. Löwenstein O, Feinberg R, and Loewenfeld IE: Pupillary movements during acute and chronic fatigue. A new test for the objective evaluation of tiredness. Invest Ophthalmol 2:138, 1963.
10. Yoss RE, Moyer NJ, and Hollenhorst RW: Pupil size and spontaneous pupillary waves associated with alertness, drowsiness, and sleep. Neurology 20:545, 1970.
11. Yoss RE: The sleepy driver: A test to measure ability to maintain alertness. Mayo Clin Proc 44:769, 1969.
12. Yoss RE, Moyer NJ, Carter ET, and Evans WE: Commercial airline pilot and his ability to remain alert. Aerospace Medicine 41:1339, 1970.
13. Press WH, Flannery BP, Teukolsky SA, and Vetterling WT: Numerical Recipes, The Art of Scientific Computing. Cambridge, UK, Cambridge University Press, 1986.
14. Ludeman LC: Fundamentals of Digital Signal Processing. New York, Harper and Row, 1986.