General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

NASA CR-
144377

STEPWISE ITERATIVE FOURIER TRANSFORM: THE SIFT

Prepared by:
Vernon A. Benignus
and
Gayla Benignus

BEHAVIORAL TECHNOLOGY CONSULTANTS, INC.
8641 Colesville Road, Silver Spring, Md.

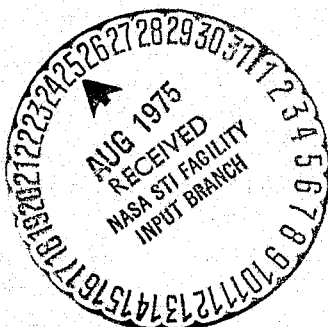**July, 1975**

STEPWISE ITERATIVE FOURIER TRANSFORM: THE SIFT

Prepared for: '

NASA LYNDON B. JOHNSON
SPACE CENTER
Under Contract·
~~NAS9-13707~~
NAS 9-13703

Prepared by:

Vernon A. Benignus
and
Gayla Benignus

Behavioral Technology Consultants, Inc.
8641 Colesville Road
Silver Spring, Maryland

July, 1975

STEPWISE ITERATIVE FOURIER

TRANSFORM: THE SIFT

Vernon A. Benignus and Gayla Benignus
Behavioral Technology Consultants, Inc.

SUMMARY

Spectrum analysis of biomedical data offers some unique advantages over classical statistical analyses and over visual analysis by trained clinicians, particularly for applied biomedical studies of physiological functioning in flight and simulation chamber environments. The data obtained in these settings often take the form of time series, which typically present complex waveforms composed of various periodicities. The application of spectrum analysis techniques to biological data still presents enough uncertainties and constraining factors to make spectrum analysis results less than straightforward with respect to interpretations. The present study program was designed specifically to study the respective effects of some common data problems on results obtained through stepwise iterative Fourier transformation of synthetic data with known waveform composition. Included in this group were the problems of gaps in the data, different time-series lengths, periodic but nonsinusoidal waveforms, and noisy (low signal-to-noise) data. Results on sinusoidal data were also compared with results obtained on narrow-band noise with similar characteristics. The findings showed that the analytic procedure under study can reliably reduce data in the nature of (1) sinusoids in noise, (2) asymmetric but periodic waves in noise, and (3) sinusoids in noise with substantial gaps in the data. The program was also able to analyze narrow-band noise well, but with increased interpretational problems. The procedure was shown to be a powerful technique for analysis of periodicities, in comparison with classical spectrum analysis techniques. However, informed use of the stepwise procedure nevertheless requires some background of knowledge concerning characteristics of the biological processes under study. Uninformed use of the procedure can lead to obvious inferential errors. It is also recommended that the program should be subjected to further tests involving comparisons of its performance across a range of different life systems measures of periodic processes before it is accepted as a standard analytic tool for general biomedical data analysis applications.

INTRODUCTION

This report describes a method of spectrum analysis that was developed in the course of testing and modifying Rummel's program for multiple-regression estimation of period domain spectra (Rummel, 1971). It follows an earlier report addressed to evaluation of the original program (Benignus, 1972).

The findings described in this report are relevant to the many theoretical and practical complexities confronting research investigators who wish to analyze biorhythms data. While a number of spectrum analysis techniques are available for the study of biorhythms, application to biological time-series continues to yield results that are often ambiguous. For applied research purposes such as studies of spaceflight effects on circadian rhythms in astronauts, several specific problems assume considerable significance. The amount of data required for reasonably precise estimates of circadian phenomena is one issue of special importance to space medicine and biology. The task of obtaining data in flight simulation or spacecraft environments over a period of many days can be very expensive. Standard spectrum analysis procedures may require testing subjects over periods of time that are too long to be feasible. Another significant problem faced in applied research on biorhythms is the problem of spectral resolution. Resolution capacity of standard forms of spectrum analysis is a function of record length. Therefore, if it is possible to obtain data only over short periods of time, changes in the characteristics of the biorhythms under study may not be detected because of coarse spectrum resolution.

In chamber studies, as well as in manned spaceflights, other problems can arise that affect spectrum analysis results in ways that have not yet been clearly delineated. The problem of missing data effects on spectrum results is a very common occurrence and questions still exist concerning the selection of interpolation procedures for filling gaps in the data. The necessity for collecting biological samples such as blood or urine samples at unequal intervals presents significant questions. One particularly serious problem for standard spectrum analysis procedures is the problem of analyzing biorhythms which are unstable with respect to amplitude or phase. Research investigators could greatly profit from new spectrum analysis techniques capable of usefully characterizing such unstable rhythmicities.

The study program described in this report provided an opportunity to explore the Rummel program's performance with respect to these significant problems and then to refine his basic approach, to the extent of creating a new program for stepwise iterative Fourier transformation, designated as SIFT.

The results obtained not only have methodological implications but also provide an empirical basis for reconceptualizations of spectrum analysis philosophy.

It is assumed that the reader has previously acquired a basic understanding of multiple regression, stepwise regression, and spectrum analysis, although references to basic literature are provided for review purposes.

SYMBOLS

T time series length

b least squares regression coefficient

x_n time series of n length

$\mathbf{b}$ vector of regression coefficients

$\mathbf{R}_{11}$ matrix of correlations among the independent variables

$\mathbf{R}_{12}$ vector of correlations between independent variables and the dependent variable

Δf frequency increment

ΔP period increment

A_i amplitude estimate at frequency i

a_i sine component of amplitude at frequency i

b_i cosine component of amplitude at frequency i

$\underline{t}$ statistical criterion for estimating predictive value (student's $\underline{t}$)

$\underline{R^2}$ multiple squared correlation coefficient

$\underline{z}$ Fisher's $\underline{z}$-transform

RMS   root mean square

SNR   signal-to-noise ratio

L length of time series containing narrow-band noise (in terms of number of observations)

f frequency

P period

p probability

k number of bands used for spectrum analysis

4

BACKGROUND AND RATIONALE

Resolution of the DFT

In methods of spectrum analysis using the Discrete Fourier Transform (DFT) or its fast algorithm, the Fast Fourier Transform (FFT), the attainable frequency domain resolution is $1/T$ Hz, where T is the length of the time series being analyzed (Hinich and Clay, 1968).

This reports shows that, for some cases, it is possible to achieve considerably better resolution by a method of multiple regression spectrum analysis. For present purposes the DFT is viewed as a series of single-variable, orthogonal, regression analyses. This view permits generalization to a method of multiple estimation which can yield better spectral resolution. In the following discussion the usual assumptions are made about time series, in terms of their randomness, Gaussian distribution, and stationarity (Bendat & Piersol, 1966).

By solving a least-squared-error equation using a cosine wave as an independent variable to estimate a time series, it may be shown that the least-squared estimation error is obtained when the regression coefficient is

$$b_i = \frac{2}{N} \sum_{n=1}^{N} x_n \cos (2\pi nf_i t_n) \qquad [1]$$

Here x_n is an N-long time series estimated by a cosine wave of frequency f_i. The least-squares regression coefficient, b_i, is the Fourier coefficient for the cosine component at frequency f_i. Equation [1] is the Euler equation for estimating the Fourier coefficient using the usual DFT. By minimizing squared errors of estimation, the usual Euler expression may be derived for the sine components as well.

It has been demonstrated (Bendat and Piersol, 1966) that two cosine Fourier coefficients, b_i and b_j, are independent so long as the frequencies of the consine estimator waves, f_i and f_j, are harmonic multiples. This is true when the lowest frequency to be estimated in the data is $1/T$ and the next higher frequencies estimated are spaced in a harmonic progression and separated by $1/T$, i.e., $2(1/T)$, $3(1/T)$, etc. It may be easily shown that two cosine independent variables of frequency f_i and f_j are linearly uncorrelated when they are harmonically related. When two independent variables are uncorrelated and the dependent variable is a Gaussian random variable, the corresponding two regression coefficients are also unrelated (Guilford, 1965). From least-squares regression analysis the DFT may then be viewed as a series of single-variable, orthogonal, regression estimates, each performed at a different frequency with frequencies spaced at the harmonic intervals of $1/T$. At each frequency there are, of course, two orthogonal estimators: the sine and the cosine waves.

If two independent variables, say cosine waves, are spaced more closely than $1/T$, it can easily be shown that they are no longer uncorrelated. When two independent variables in a regression analysis are correlated, their corresponding regression (Fourier) coefficients are also correlated (Guilford, 1965).

It is for this reason that, when spectrum estimates are spaced more closely than 1/T, they are no longer fully resolved from one another.

Multiple Regression Spectra

It is not at all unusual in multiple regression analysis to treat correlated independent variables. As discussed above, if a series of single-variable regressions is run in such a situation, the regression coefficients will be correlated. Indeed, the estimates of regression are highly misleading (Guilford, 1965). The usual solution is to use multiple-regression analysis in such a way that the regression coefficients are each calculated with the effects of the correlated estimators "taken out," i.e., considered. Such regression coefficients are called "partial" regression coefficients, and they are much less misleading (Guilford, 1965).

It is often desirable to estimate a spectrum in which estimates are spaced more closely than 1/T and hence would (in the usual DFT) be overresolved. If, however, a multiple-regression scheme is used, it is possible to account for the inter-correlations among independent-variable waves and obtain more meaningful sample regression (Fourier) coefficients. It is this method of analysis which is treated in this report.

An equation for the computation of multiple regression coefficients can be written in matrix form as

$$b = R_{11}^{-1} R_{12}$$

[2]

where the boldface letters indicate matrices (Cooley and Lohnes, 1971). In equation [2], b is the vector of regression coefficients which are being computed, R_{11} is the matrix of correlations among the various independent variables, and R_{12} is the vector of correlations between the independent variables and the dependent variable.

A first impression would suggest that for a multiple-regression spectrum analysis one could simply compute R_{11} for the full spectrum of independent variables (the sine and cosine waves); compute R_{12} (the correlations between the sine and cosine waves and the time series); then solve equation [2] for b, the vector or regression coefficients, and obtain a highly resolved spectrum. In fact, the matrices of correlations are easily computed, and this is done in the method reported here.

When, however, an attempt is made to solve equation [2], a complication occurs. When the order of R_{11} is high and when many inter-correlations are large, the inversion of R_{11} becomes increasingly inaccurate for any fixed level of computational precision. While this is a soluble difficulty, solutions are costly in terms of memory and computation time.

6

There is another, more compelling reason for not proceeding as indicated in equation [2], i.e., for not computing a spectrum containing all possible frequencies of interest simultaneously. If such a simultaneous computation were made, many of the spectrum estimates (regression coefficients) would have highly unreliable values. In the life sciences, most time series have at least a moderately peaked spectrum; consequently, most of the significant activity in a time-series record can be accounted for by considerably fewer than all possible frequencies. If only a few frequencies are required, it is implied that the simple correlations in matrix R_{12} would, in general, be low for all except the required frequencies. The standard error of estimation for a partial regression coefficient varies inversely with the correlation between the corresponding independent and dependent variables (Guilford, 1965). Thus, for any independent variable (sine or cosine wave of a particular frequency) that does not correlate well with the time series, the estimate of spectrum energy (regression coefficient) will be very unstable, especially in a multiple-regression case. In practice it has been found that, for frequencies that do not contribute significantly to the time series, spectrum estimates of 10 to 100 times larger than reasonable values actually occur under common circumstances. Not only is it computationally costly to perform [2], but also many of the values in b are wildly unreliable.

A Stepwise Method

When faced with many possible independent variables of unknown predictive value, a common statistical procedure is to employ stepwise selection of those independent variables (Draper and Smith, 1966). This implies the performance calculations described in equation [2] but with fewer variables used. Most regression programs compute R_{11} and R_{12} for all possible independent variables and then select independent variables, singly and in order of their predictive importance, to build a set of regression solutions in order of increasing complexity. The first regression equation executed contains one (the most predictive) independent variable. The next solution usually contains two independent variables: the first found to be highly predictive plus the next most predictive one. This procedure continues to include, one at a time and in order of predictive salience, the remaining independent variables.

In a true stepwise procedure, all variables previously entered are checked at each step to see if they still contribute to prediction in a significant way after other variables have subsequently been entered. When the program reaches some a priori criterion of predictive accuracy, the addition of new independent variables stops.

In the case of spectrum analysis, the final solution of a true stepwise procedure would yield a vector, b, of regression coefficients which would specify the characteristics of the first k most important frequencies, k being the number of frequencies entered in order to achieve the desired predictive accuracy. There are several presumed advantages of this method over some

alternative procedure such as simply picking the first k biggest spectrum peaks. As discussed above, after the first variable is entered, the others are computed with the correlations between independent variables <u>considered</u>; and therefore, a much more accurate result is achieved (Guilford, 1965; Draper and Smith, 1966). (This is true, of course, only for the overresolved spectrum case). Variables are also selected on the basis not of their amplitudes, but of their predictive reliability. Especially in cases where several predictors have already been entered, as discussed above, the reliability of the regression coefficient can be very low; and its computed amplitude can, therefore, be erroneously very high. In the latter case, the frequency chosen would not be a good candidate as an independent variable.

When a vector, b , of spectrum estimates has been computed by a stepwise procedure, it is certainly not equivalent to a comparable spectrum as estimated by the standard DFT. The DFT computes k spectrum estimates, equally spaced and orthogonal to each other. The true stepwise program computes k spectrum estimates which are the first most important ones and which might be unequally spaced and nonorthogonal. In this stepwise method, no estimates are made in spectrum regions which have insufficient activity to enter into the prediction scheme. The method of stepwise spectrum analysis has the advantages that (1) it yields a more accurate and resolved result and (2) its spectrum is still parsimoniously stated (and may indeed be more parsimoniously stated) than a less resolved DFT. The disadvantages of the method are (1) it yields no estimates for some parts of the spectrum, so that the usual continuous spectrum series is not obtained, and (2) stepwise procedures can be seriously capricious. The latter difficulty will be discussed extensively in the next paragraph and in context throughout this report.

The first difficulty cited, that of noncomparability to DFT results, is only one of form or convention. It is easily arguable that in an overresolved spectrum these other frequency estimates are either redundant in the case of the DFT or would be highly inaccurately estimated in the case of the stepwise Fourier Transform. In any case, the argument of parsimony is a very powerful one from both a statistical and a philosophical point of view.

The point is often made (and with great vehemence in some quarters) that any method of stepwise regression analysis such as the stepwise Fourier Transform is a capricious procedure which capitalizes on small sampling fluctuations (Efroymson, 1962). In all cases of stepwise procedures, which are in the nature of general "fishing expeditions," it is necessary to replicate the procedure at least once (or until stable results are achieved) to insure that chance results did not occur. The probability of chance results increases with (1) the number of independent variables available for selection by a program and (2) the number finally selected. A great deal of attention has been paid to such capriciousness in the evaluation of this method.

8

Two Domains

The usual DFT is a transformation of data measured in time (time-domain data) into data expressed as a function of frequency (frequency-domain data).

It is possible to express the results of a Fourier transformation as a "period" spectrum where period, P, is defined as $P = 1/f$, the reciprocal of the frequency (f). The period of a wave is simply its duration. Spectra computed in this manner will be said to be expressed in the "period domain."

Usually the frequency-domain analysis is performed by calculating equally spaced spectrum estimates at some interval of $f = 1/T$. Similarly, the period-domain estimates are also calculated at equal period increments of P, where P is expressed as time. It should be pointed out that equally spaced period estimates are not equally spaced in the frequency domain, as illustrated in table 1.

TABLE 1. COMPARISONS OF FREQUENCY AND
PERIOD SPACING FOR SPECTRUM ANALYSIS

P	$\triangle P$	f	$\triangle f$
1		1	
	1		.5
2		.5	
	1		.17
3		.33	
	1		.08
4		.25	

The fact that the period domain represents an unequally spaced frequency domain gives rise to somewhat undesirable properties. The main problem is that, since the correlations among spectrum estimates are determined by their spacing in the frequency domain, the period domain spectrum has a varying amount of correlation between each pair of the adjacent frequencies, depending upon where in the spectrum the pair is located. Thus, if $\triangle P$ is set to produce resolved (uncorrelated) estimates at the short-wave (high-frequency) end of the spectrum, then the long-wave (low-frequency) end of the spectrum will be grossly overresolved, i.e., estimates will be highly correlated.

In this report, the theory underlying methods is always discussed in terms of the more usual frequency domain. The method of spectrum computation is always evaluated in both the frequency and the period domains. Further comparisons and discussion of the differences are made.

Some Particulars of Computation

There are many different forms of stepwise regression analysis. It is, there-
fore, necessary to specify the particulars of the approach used in this report.
Further, since this report describes how stepwise regression is applied to
analyze spectra, it has also been necessary to discuss more specifically the
computational routine used here. Many of the features of computation have
been decided upon in a somewhat arbitrary fashion, when no obvious advantage
was offered by alternate but similar procedures. In such cases, choice of
procedures was governed by (1) programming convenience or (2) convention.

Sine and cosine characteristics are found at each frequency in a spectrum,
and the inclusion of these two specific phase angles of a frequency will
provide all possible information about the amount of that frequency in a
spectrum. It was arbitrarily decided that, if either component of a given
frequency was to be entered into (used in the analysis of) the spectrum,
the other component of that frequency would be entered as well. In the
procedure used here, the sine and cosine components were combined first and
then evaluated for predictive utility. Combinations of sine and components
were handled in the usual way: The amplitude, A_i, at frequency i is

$$A_i = a_i^2 + b_i^2 \qquad [3]$$

where a_i is the amplitude of the sine component and b_i is the amplitude of
the cosine component. It is possible that a better procedure might consist
of evaluating sine and cosine components separately and entering them only
if they meet some pre-set criterion. This alternative choice might be
especially important at a late stage of selection of predictors where either
the sine or the cosine component could be of almost no predictive worth and,
therefore, only contribute to a higher error in estimation. Separate selec-
tion of sine and cosine components has not been previously explored in any
known procedure, nor is it included here.

In this program two criteria for selection are used at various stages of
computation for assessing the utility of an independent variable. One of
the criteria, the t-value associated with an independent variable as a
predictor, evaluates the reliability of that variable's estimate. The
other criterion evaluates the importance of the independent variable as
an estimator in terms of reduction of uncertainty. This latter criterion
is R^2, the multiple squared correlation between the group of independent
variables and the dependent variable. A discussion of how these criteria
are used follows.

PRECEDING PAGE BLANK NOT FILMED

At any given stage of selection, the previously unselected variable that has the most reliable estimate (highest t-value) is selected for entry into the spectrum. Just after a new variable is entered, an attempt is made to optimize the R^2 value by moving all of the previously entered independent variables plus the newly entered variable up and down in the spectrum. For example: Suppose frequencies 12, 14, and 32 Hz were entered into a spectrum where 0.5 Hz resolution was being attempted, and suppose that 14 was the one last entered. The next step would be to move the 12 Hz band down to 11.5 Hz and recompute R^2. If an increase in R^2 occurred, the next step would be to move the 12 Hz band down to 11 Hz, etc., until a decrease in R^2 occurred, at which point the frequency just previous to the decrease of R^2 would be used. If, when the first downward movement was made, R^2 had decreased, an attempt would then be made to move the 12 Hz band up to 12.5 Hz and recompute R^2. When the lowest frequency band has been moved down and/or up to optimize R^2, the next highest band (in this case the newly entered 14 Hz band) would be moved up and down in a similar fashion. In this way all bands are adjusted for optimum R^2 in relation to specified frequencies. This up and down movement of the frequencies will be called "iterative R^2 optimization." If, during iterative R^2 optimization, an attempt is made to move an independent variable at some frequency into a frequency band already occupied by a previously entered independent variable, the optimization is halted there; and the "merger" of two independent variables is prevented. If an attempt is made to move an independent variable into either the highest or lowest frequency band in a spectrum, during iterative R^2 optimization, this, too, is prevented. This general procedure of stepwise variable selection and iterative R^2 optimization will be called the stepwise iterative Fourier Transform (SIFT).

The SIFT Computational Procedure

The SIFT procedure has evolved from a trial-and-error procedure using Monte Carlo test runs. New variables are entered until no more variables with sufficient t-values remain. The t-value which is used as a criterion by the SIFT is a severe problem and will be discussed later.

Further details of the program's computational procedure are spelled out in the step-by-step list in table 2.

Computational SIFT Options

The procedure described in table 2 may be implemented in either the frequency or the period domain. The program's output can be controlled via various print/plot options to provide step-by-step output of considerable detail. Less detailed output consisting, in the extreme, of only the final solution can also be obtained. Input data are obtainable from either real data or Monte Carlo simulation subroutines. The whole program may be executed any number of times under do-loop control. Further computational details are given in the program documentation.

TABLE 2. COMPUTATIONAL PROCEDURES FOR EXECUTION OF SIFT

1. Read in the parameters of the analysis.

2. Generate a matrix of estimator waves consisting of the sine and cosine waves of the frequencies of interest.

3. Read in the dependent variable wave **Y**.

4. Compute a matrix of intercorrelations between the estimator waves, R_{11}, and between the estimator waves and the dependent variable wave, R_{12}.

5. Compute a usual DFT spectrum by first multiplying the correlations in R_{12} by the standard deviations of the dependent-variable wave and then combining sine and cosine components.

6. Compute a vector of t-values associated with each amplitude (regression coefficient) in the DFT spectrum.

7. Select the most stable estimator wave (the one with the largest t-value), and attempt iterative optimization of R^2 for that estimator, as discussed on page 12.

8. Compute a new spectrum using a 2-variable estimation scheme where the amplitude at each frequency is estimated using equation [2], with the frequency entered first as one variable, and each frequency in the spectrum entered, in turn, as the other variable.

9. Evaluate the t-values of the new 2-variable-estimation spectrum to find the next highest t-value. Also check previously entered t-values for some minimum value. Again, attempt iterative R^2 optimization for all predictor waves.

10. Loop through steps 8 and 9 until no new variables can be added (because no new t-values are large enough) or until some maximum number of variables have been added.

11. Compute amplitudes and phase angles only for the frequencies selected by the above SIFT procedure.

The performance of the SIFT program has been documented both by Monte Carlo simulation and by analysis of real physiological data. The results of these tests are reported here.

Some Criteria of Performance

Since the output from the SIFT program has a format inherently different from that of the DFT, it was necessary to devise new criteria for performance evaluation. The basic criteria of performance involved (1) measures of resolution and (2) measures of error rate. Program performance was tested by executing the SIFT program 100 times, using Monte Carlo data as input. The output from SIFT, at each execution, printed a list of frequencies which were found to be significant, along with their amplitudes, phases, $\underline{t}$-values, and $\underline{R}^2$ value. These output values were then collected, over all executions, into distributions. Means and standard deviations for these outputs were also computed.

The frequency resolution of SIFT was described by making a probability distribution of the frequencies found in each execution. Thus, if the Monte Carlo signal contained one sinusoidal wave added to noise, one would expect the SIFT to find one frequency per execution to be significant; and, over all executions, the frequencies found should be distributed around the original frequency put into the Monte Carlo data. If two sinusoids were put into the Monte Carlo data, the probability distribution of all frequencies should then show two peaks at the correct frequencies; and the tails of the two peaks should overlap minimally.

The amplitude accuracy of the program is evaluated in similar terms. A probability distribution of amplitudes found by the SIFT, collected over all executions, is constructed. If all sinusoids in the Monte Carlo data had the same amplitude, a single-peak distribution should occur. If two different amplitudes were used for the sinusoid(s) in the Monte Carlo data, two peaks should occur in the amplitude data.

For each series of Monte Carlo runs, the mean $\underline{R}^2$ value (via Fisher's $\underline{z}$-transform) was also computed. In certain cases, means and standard deviations were also computed on other output values.

Records were kept on the number of errors of various kinds that the program made. On any individual execution the SIFT could have found either more or less than the correct number of frequencies in the Monte Carlo data. A count was kept of this type of error. It should be pointed out that this kind of error is not detectable when real data are being analyzed. The other kind

of error which was recorded is a detectable type of error that can be spotted
even in real data. For example, in any execution the program occasionally
found amplitudes which were much too large in view of the root mean
square (RMS) amplitude of the input wave. On each run of data, the SIFT
program computed the amplitude of the input wave and compared it to the
amplitude of the wave reconstructed by an inverse Fourier transform of the
SIFT results. If the reconstructed wave's amplitude was greater than the
input wave's amplitude by more than a factor of two, the results were
deleted from the probability distribution and an error count was made. Since
this is a detectable error, it is proper to delete the results of the Monte
Carlo execution from the probability distributions describing program perfor-
mance, so long as the error count is considered.

The following is an evaluation of SIFT's performance when various kinds of
Monte Carlo data were used. The criteria of performance outlined above were
used to document the results.

The t-Value Criterion

One of the major problems in a stepwise-regression analysis is the proper
utilization of the "significance test" criterion that is used. It was there-
fore decided to study this problem using a Monte Carlo procedure. It is
clear that if the t-criterion is set too small, then spurious peaks due to the
Gaussian noise may become significant, and the possibility of finding too
many significant peaks begins to rise. Similarly, if the t-criterion is set
too large, then some of the salient peaks in the Monte Carlo data will be
rejected.

Another problem, not unique to the SIFT, is that, as more estimators are
added to the regression equation, the value of the residual becomes pro-
gressively smaller, and the calculated t-value becomes larger. Again, for
certain stages of a SIFT procedure, spurious peaks might appear significant.
A solution to this problem would be to use some other criterion such as (1)
accepting a significant decrease in the residual sum of squares or (2) enter-
ing that variable with the highest partial correlation. The SIFT uses two
t-value criteria: one for the first selection and another for a later stage
of the selection. While some other procedure could be better, the pressure
of computational time and the availability of the double t-criterion solu-
tion made this method most attractive.

A study using Monte Carlo runs was performed to evaluate the effect of the
value of the t-to-enter criterion when more than one sinusoid was present
in a signal. For each Monte Carlo run, the program was executed 100 times
on 100 independently generated signals. Monte Carlo runs were made for
various values of t-to-enter and for various signal-to-noise ratios (SNRs).
For each SNR, the "best" t-to-enter was selected. As t-to-enter was made
larger, the program found proportionally fewer significant frequency bands,

16

and as t-to-enter was made smaller, too many bands were found significant. It was possible, therefore, to select the value of t-to-enter so as to minimize the probability of these kinds of errors, which (again) are undectable in real data. The detectable kind of error discussed above (that of finding too much amplitude in the reconstituted wave) was a monotonically decreasing function as t-to-enter increased in value. Therefore, the "best" value of t-to-enter was selected as the one which minimized the undetectable error of finding too few or too many bands in the data.

The criterion for entry of the first frequency band for the Monte Carlo data was arbitrarily set at a t-value of 1.65, corresponding to $p < 0.1$. Using this value of t, the program never entered an estimator in 100 Monte Carlo trials using only Gaussian random noise. Yet, even with relatively large proportions of noise added to single sinusoids, at least one frequency was always entered as a predictor.

Table 3 shows the "best" values of t-to-enter for the case where two sinusoids of equal amplitudes (of 1.0) at periods of $P_1 = 23.0$ and $P_2 = 27.0$ were mixed with Gaussian noises of amplitudes 0.5 or 1.0 (SNR = 1.0:0.5 or SNR = 1.0:1.0). As may be seen from inspection of table 3, the best t-to-enter value decreases as the amplitude of the noise increases. This is a reasonable result, since more noise implied that the entered sinusoidal independent variables account for less of the dependent variable wave. This finding is disturbing in that the value of the SNR is never known a priori in real data. The implication of this fact is that one should probably do several analyses on each piece of real data.

TABLE 3. BEST VALUE OF t-TO-ENTER AS
A FUNCTION OF SNR, USING TWO SINUSOIDS
OF P = 23.0 AND P = 27.0

Amplitude of Noise	Best t-to-enter
0.5	10.0
1.0	5.0

Early analyses should be used to (1) estimate the SNR and (2) more informedly set the t-to-enter criterion on subsequent runs. The value of t-to-enter must eventually be set by some rather arbitrary method involving the criteria of parsimony and reasonableness. In summary, this consideration means that the SIFT produces non-unique solutions; but, then, this is true of other stepwise statistical techniques and is one of their disadvantages in general.

Table 4 shows the error rate for the two previously discussed kinds of errors as a function of noise level for the optimum t-to-enter values when the SIFT was used in the period domain. When the frequency-domain performance was evaluated, the t-to-enter values were optimum at the same values and the error rates were almost identical

TABLE 4. PROBABILITY OF TWO KINDS OF ERROR
AS A FUNCTION OF SNR, USING THE SIFT
IN THE PERIOD DOMAIN

SNR	Probability of Detectable Errors	Probability of Undetectable Errors
1.0:0.5	0.15	0.03
1.0:1.0	0.16	0.10

Figure 1 shows plots of the probabilities of finding various periods in synthetic data (sinusoids combined with Gaussian random noise), produced as described above.

It can be seen that the discrimination between the two periods is quite good for both SNR values. Even when t-to-enter values are not optimum in terms of various error rates (as discussed above), the curves of period discrimination are not greatly degraded.

In subsequent sections, comments continue to be made about the optimum t-to-enter value. The problem of how to set t-to-enter values is addressed in the "Concluding Remarks" section.
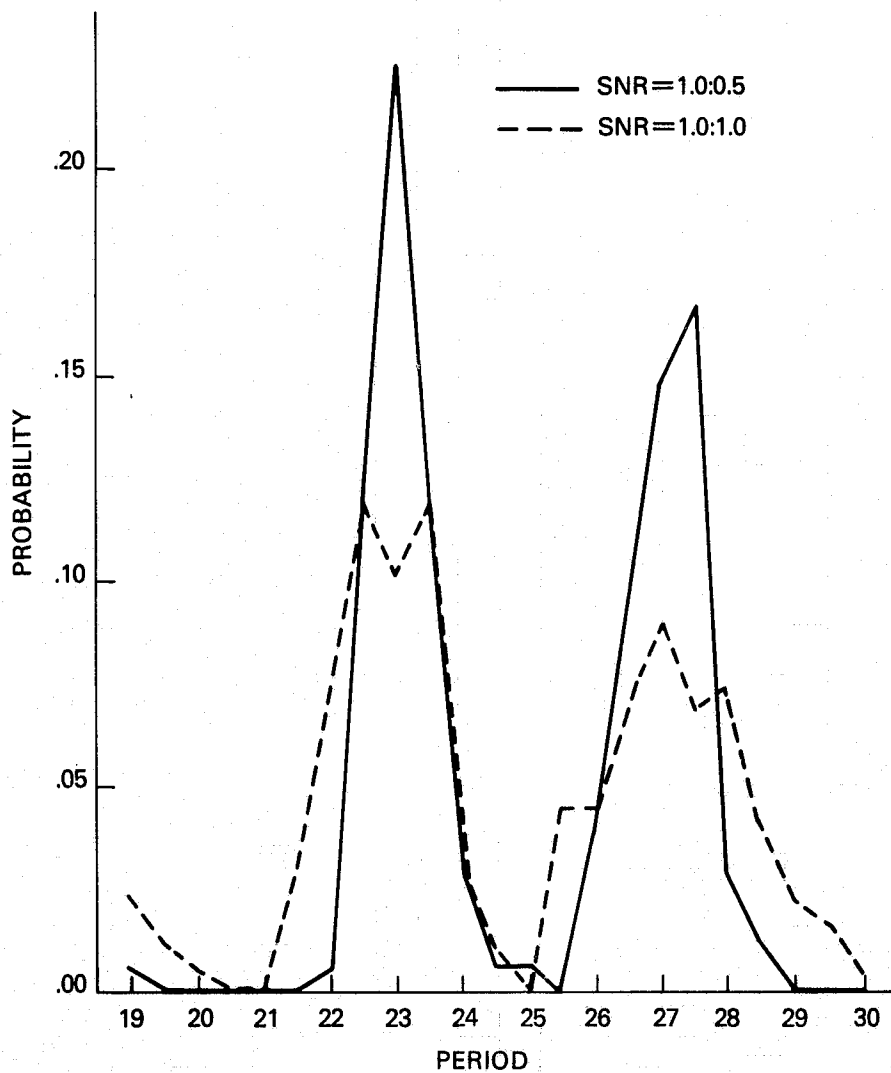
Figure 1. Period discrimination at two levels of SNR, using SIFT with optimized t-to-enter values, two periods in Monte Carlo data, $P_1 = 23.0$, $P_2 = 27.0$.

Period and Amplitude Discrimination

The curves for discrimination between two periods of equal amplitude are shown in figure 1. Since both of the periods in that graph had the same amplitude, it would be expected that the amplitude probability distribution for the output from SIFT would have a single peak centered about 1.0, the correct amplitude.

Figure 2 shows that this is indeed the case for both levels of SNR. The 1.0:1.0 SNR value, of course, produced a wider distribution of amplitudes.

The period and amplitude discrimination study was also carried out for two sinusoids which had different amplitudes. Figure 3 shows the period-discrimination curve for two sinusoids of $P_1 = 23.0$ and $P_2 = 27.0$ with amplitudes of 2.0 and 1.0 respectively. Figure 4 shows the corresponding amplitude-discrimination curves. Note that the sinusoid of $P_1 = 23.0$ has an amplitude (A_1) of 2.0, and the noise amplitude is 1.0, so that the SNR for this wave is 1.0:0.5. The peak in the frequency-discrimination curve of figure 3, corresponding to $P_1 = 23$, should be compared in shape to the $P_1 = 23.0$, SNR = 1.0:0.5, curve of figure 1. It is apparent that the shape is almost the same. Similarly, the shape of the peak in the amplitude discrimination curve of figure 4 corresponding to $A_1 = 2.0$ should be compared to the shape of the peak for SNR = 1.0:0.5 of figure 2. Except for absolute probabilities, the curves are reasonably similar. Obviously the absolute probability values in figure 4 ought to be lower, since the events are divided into two peaks.

The peak in figure 3 which corresponds to $P_2 = 27.0$ has a shape which approximates the corresponding peak in figure 1, SNR = 1.0:1.0. This is reasonable, since the amplitude of the $P_2 = 27.0$ component was 1.0, thus making the SNR for the $P_2 = 27.0$ sinusoid 1.0:1.0. Similarly, the amplitude-distribution peak for the wave with $P_2 = 27$, $A_2 = 1.0$ is similar to the amplitude distribution for SNR = 1.0:1.0 in figure 2.

From these observations the reasonable conclusion is that, when two or more sinusoids are present in a signal, the discrimination of those sinusoids is based upon the ratio of their amplitudes to the noise in the signal.

The best t-to-enter criterion for the runs considered above was $t = 7.5$, which is halfway between the best value for SNR = 1.0:0.5 and the best value for SNR = 1.0:1.0, when amplitudes of the two sinusoids are equal. Apparently for one of the components, $P_2 = 27.0$, the t-to-enter is too large; and occasionally the SIFT does not enter one of the P_2 waves. This contention was borne out by an examination of the particular periods found in 100 Monte Carlo runs by SIFT. This examination revealed that, when an undetectable error occurred, it was due to an omission of the $P_2 = 27.0$ band. Since the t-to-enter value for $P_1 = 23.0$ was too small, none of these bands were ever omitted. Consequently, the overall undetectable error rate of the Monte Carlo runs for two amplitudes was lower than that for either of the
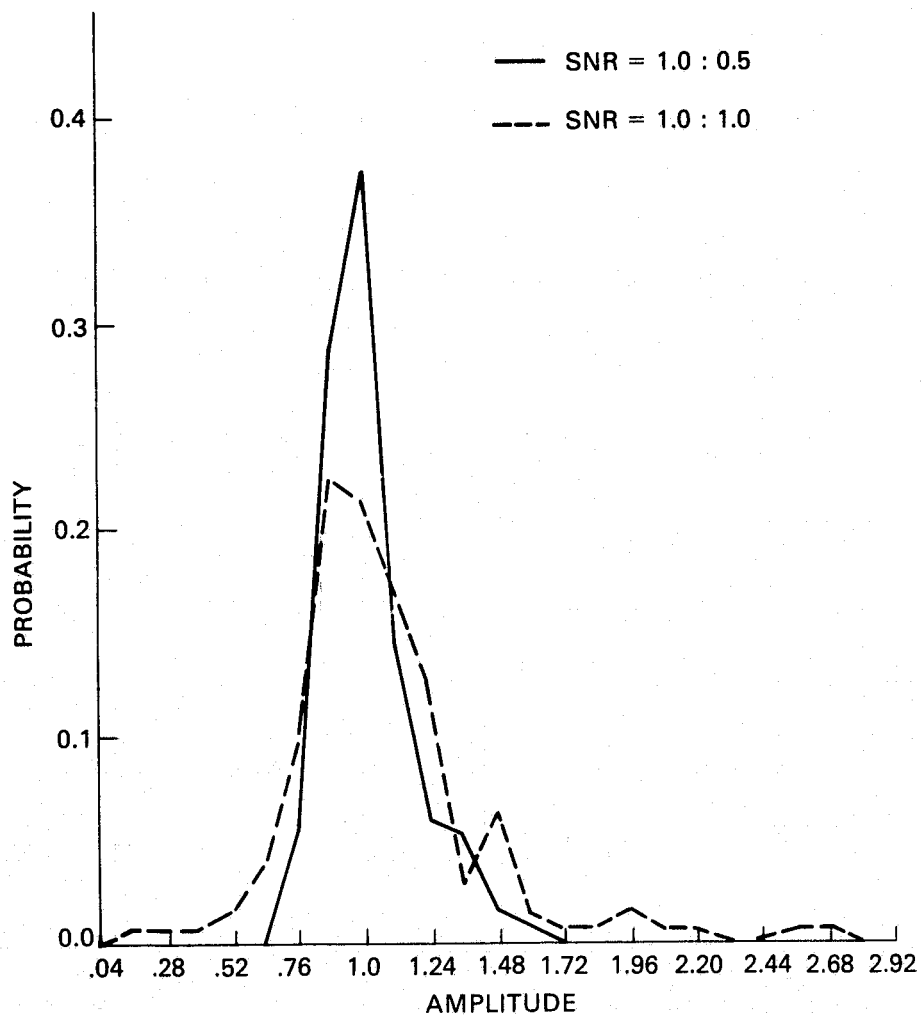
Figure 2. Probability distributions of amplitudes in the period domain found by SIFT for two levels of SNR, single amplitude in Monte Carlo data.
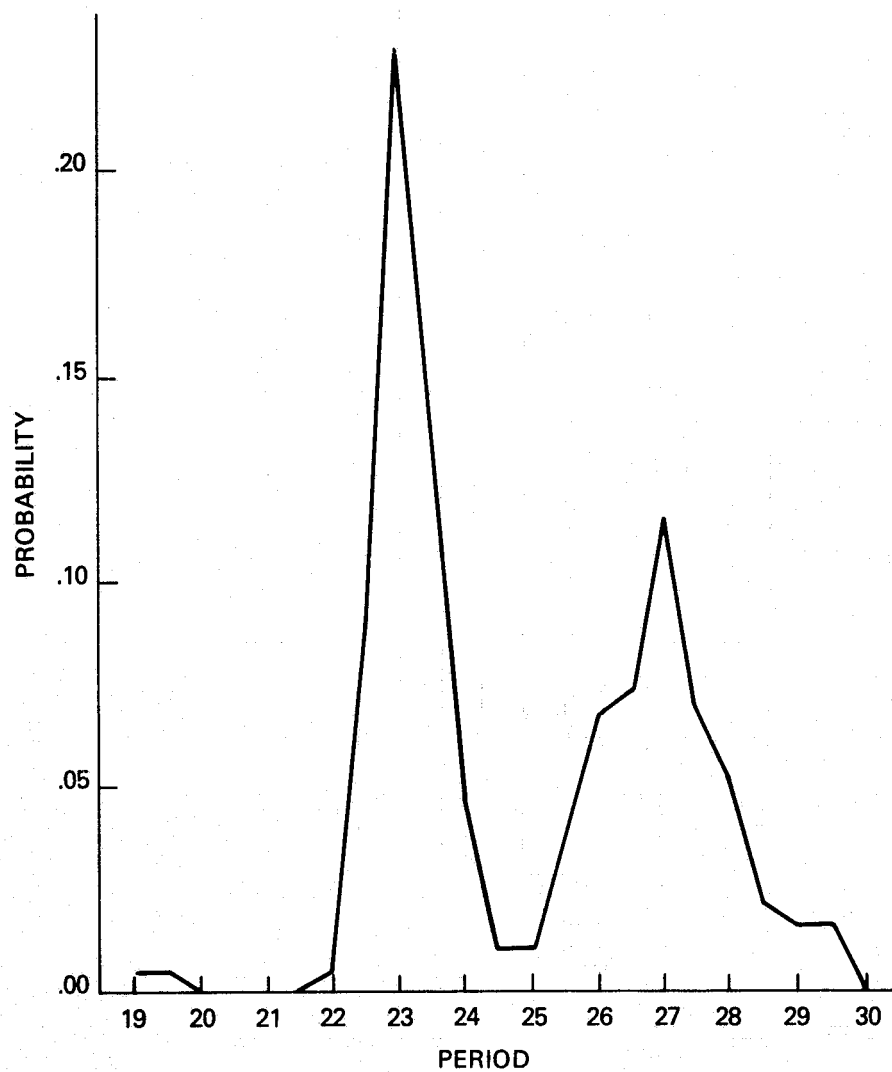
Figure 3. Period (P) discrimination for two sinusoids of
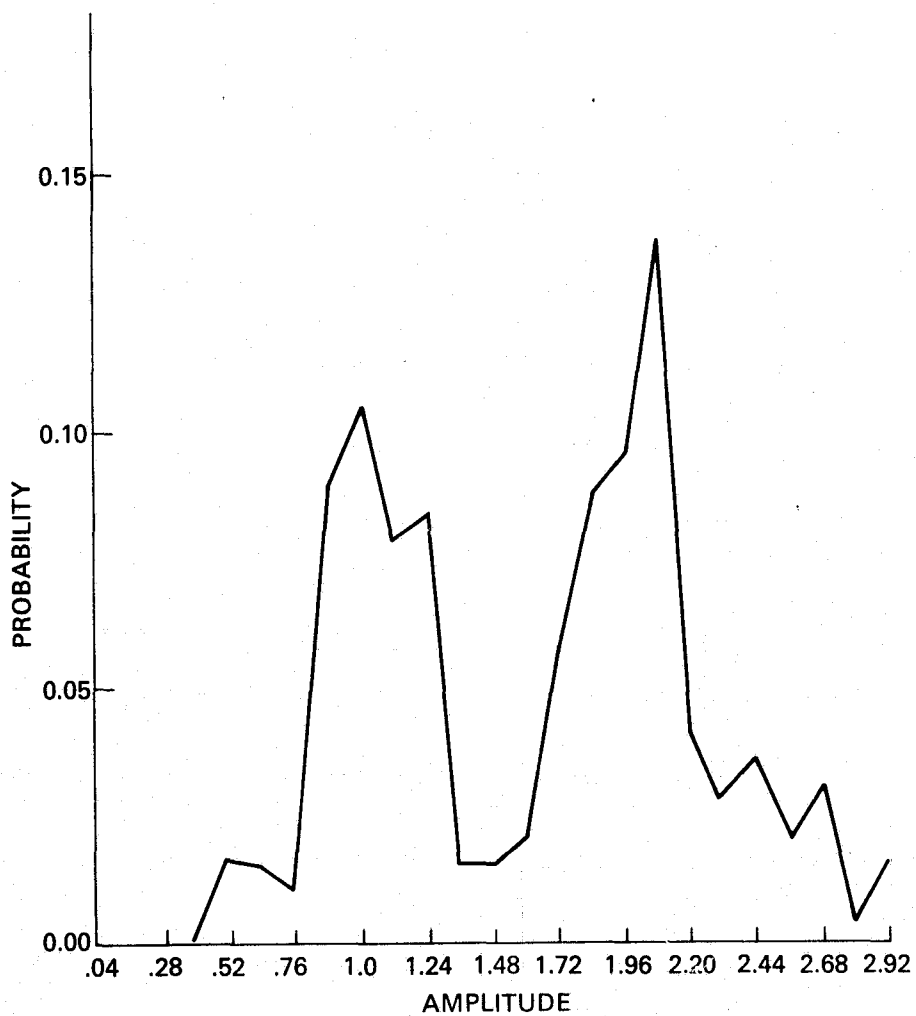P = 23.0 and P = 27.0 with amplitudes of 2.0 and 1.0 respec-
tively. Noise level = 1.0.

Figure 4. Amplitude discrimination in the period domain by SIFT for two sinusoids of P = 23.0 and P = 27.0 with amplitudes of 2.0 and 1.0 respectively. Noise level = 1.0.

23

SNR = 1.0:1.0 runs. Table 5 shows the probability values for the two kinds of error. It is probable that the detectable error rate for this case was also lower, simply because detectable error rate falls as t-to-enter is increased. These runs were repeated for the frequency domain, and the results were similar.

TABLE 5. PROBABILITY OF TWO KINDS
OF ERRORS FOR $\underline{t}$-TO-ENTER = 7.5[a]

Probability of Detectable Errors	Probability of Undetectable Errors
0.04	0.06

[a] P_1 = 23.0, A_1 = 2.0, P_2 = 27.0, A_2 = 1.0, noise amplitude = 1.0.

The Effects of Time-Domain Data Length

It was decided to evaluate the effect of the length (L) of the time domain data upon the period discrimination and error rate of SIFT. Theoretically, if all factors were adjusted appropriately, the output from the SIFT ought to be predictable.

Three values of L were used in Monte Carlo tests: L = 50, L = 100, and L = 200 data points. In all of these cases, it was assumed that the sampling rate was held constant so that longer data vectors correspond to longer time intervals. The value of P was adjusted to an appropriate size for a given value of L as shown in columns 1 and 2 of table 6.

TABLE 6. VALUES OF $\triangle P$, P_1, AND P_2 FOR VARIOUS VALUES OF L

L	$\triangle P$	P_2	P_1	P_2-P_1	Number of Cycles at Mid-Frequency
50	1.00	13.5	11.5	2	4
100	0.50	27.0	23.0	4	4
200	0.25	54.0	46.0	8	4

Preliminary results during this phase seemed to indicate that the number of cycles of the periodic components that could be completed during the course of the time-series segment was a critical factor. If more cycles were completed, better resolution was obtained. In order to make useful comparisons, this factor (the number of cycles used) had to be held constant. In previous studies, about 4 cycles of the mid-frequency between the two periodic components were completed, i.e., the mid-frequency between $P_1 = 23.0$ and $P_2 = 27.0$ was $P = 25.0$. In the case of $L = 100$, a wave of $P = 25$ could complete 4 cycles. Then, if L was doubled, P for the periodic components was also doubled to allow the waves to complete the same number of cycles. Table 6 shows the values of the two periodic components, P_2 and P_1, for each value of L. The effect of using a different number of cycles will also be shown.

Figure 5 shows three period-domain discrimination curves for $L = 50$, $L = 100$, and $L = 200$. In this figure P_1 and P_2 are not specifically labeled nor are any other periods. Examination of figure 5 enables the reader to assign specific period values to the graph. The nonspecific values were used to permit superposition of the graphs for comparative purposes. It appears from figure 5 that, even when the various above-mentioned factors are accounted for, longer pieces of time-series data possibly yield better resolution.

This finding was further emphasized by an examination of failure rates. Figure 6 shows the probability of detectable and undetectable failures as a function of L. Even when the various parameters were accounted for, short time series produced more failures of both kinds.

It is not entirely clear why long time series can be more accurately analyzed even when $\triangle P$ and $(P_2 - P_1)$ are adjusted for theoretical expectations. It appears, however, that not only does the SIFT provide greater resolution than would be theoretically expected, but as L increases, the performance improves increasingly over theoretical expectation. This probably has something to do with the value of L with respect to the size of the correlation matrix. This conclusion is supported by the fact that, as L increased, the decrease in program failures was more dramatic than the increase in resolution.

A test was performed of the hypothesis that the width of the spectrum window at $L = 200$ had not really narrowed appreciably over expectation. A Monte Carlo run was performed with $P_2 = 50$ and $P_1 = 47$, which produced a frequency-domain separation of 0.0007 rather than 0.0032, as shown in figure 5. In the 0.0007 frequency separation run, the program separated the two signals in only one case out of 100. In all other cases, it found only one frequency. Thus, we may conclude that, as L increases, the program makes fewer errors; but the resolution is approximately an inverse function of L. This observation is consistant with theory.
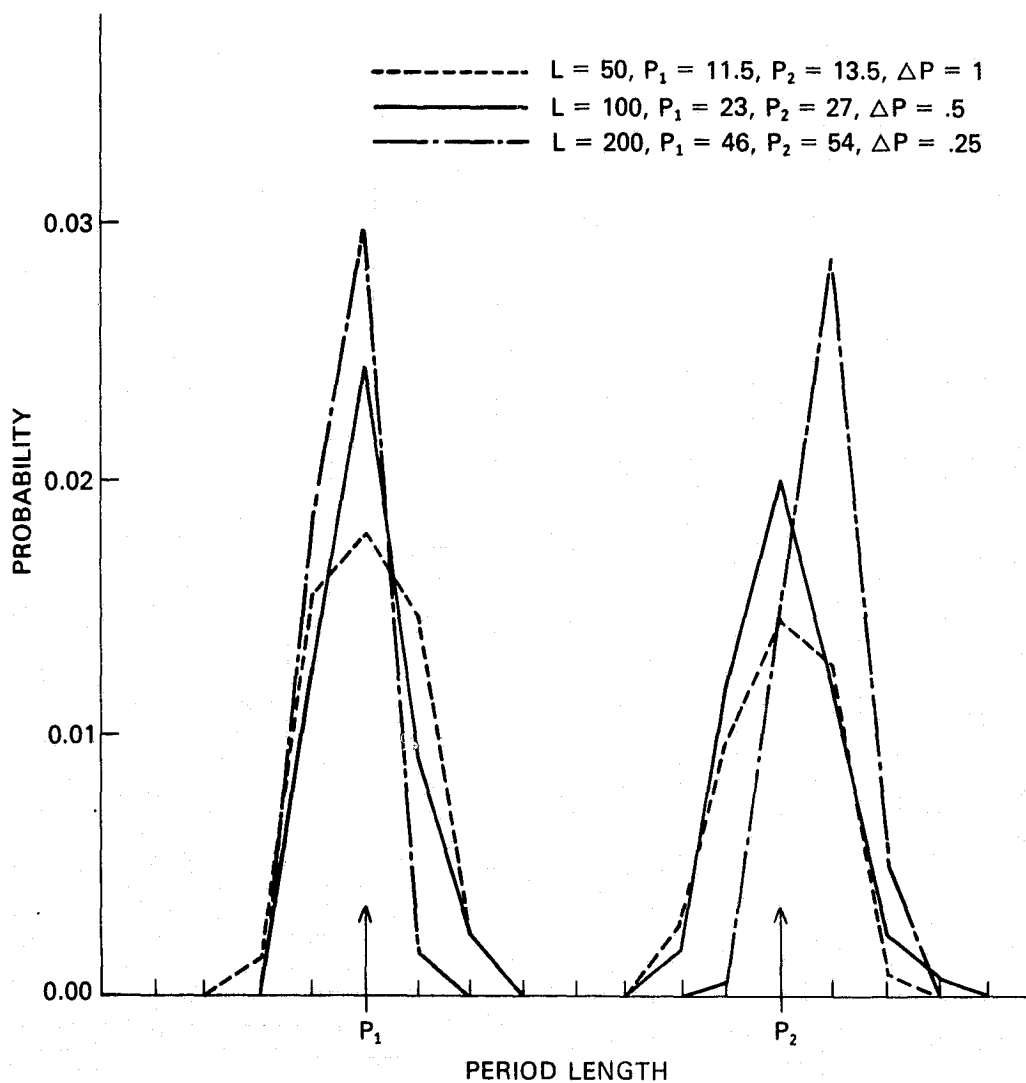
25

Figure 5. Period discrimination curves for three time-series lengths. Data normalized for expected resolution and number of cycles in time series.
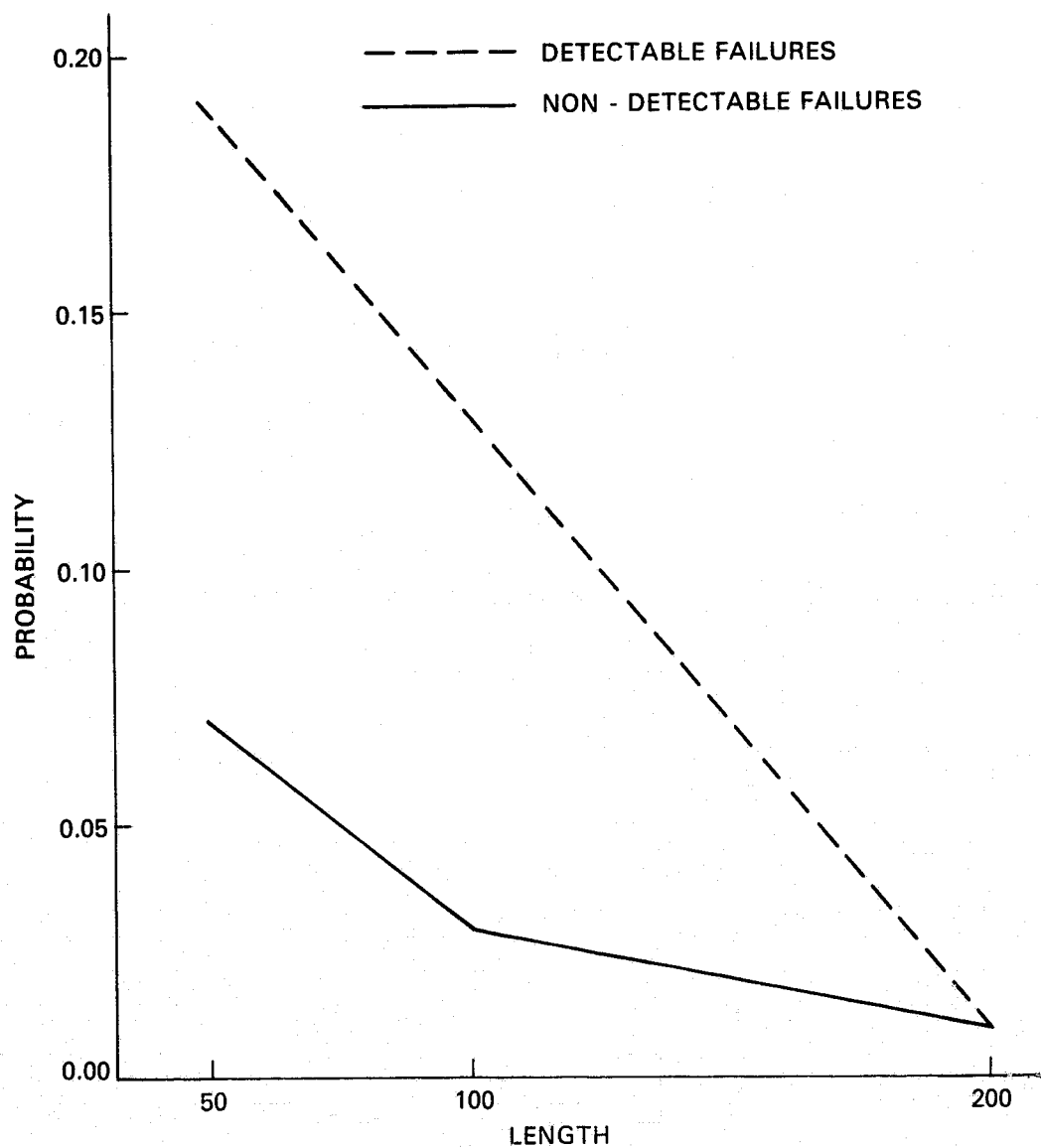
Figure 6. Probability of program failures with three different data lengths.

In summary, it has been shown that the SIFT method has resolution which is approximately an inverse function of L. This finding is entirely consistent with theory. It appears, however, that as L increases, the number of program failures decreases, even when the three variables, ΔP, $(P_2 - P_1)$ and the number of cycles per time-series segment are accounted for.

Asymmetric Waves

It is well known that some biological waves (such as the circadian rhythm) do not present symmetrical positive and negative half-cycles. It was decided to investigate the performance of the SIFT on such asymmetric data. For this purpose, periodic waves were generated with certain asymmetries. Specifically, an asymmetric wave was constructed by splicing a long, positive, sine wave with a short, negative, sine wave at the 180^0 point. Several cycles of this wave were generated. The asymmetry coefficient was defined as the proportion of the whole cycle occupied by the positive half-cycle. Thus, a wave with an asymmetry coefficient of 0.5 is an ordinary sine wave. Figure 7 shows the plot of a single cycle of a wave with P = 24.0 and an asymmetry coefficient of 0.75.

The initial study of asymmetry was carried out with a single periodic wave of amplitude = 1.0 added to a Gaussian noise of amplitude = 1.0. Figure 8 shows the width of the probability peaks for three asymmetries. It can readily be seen that, as asymmetry becomes more extreme (the asymmetry coefficient increases), the period discrimination is degraded. However, an asymmetry coefficient of 0.75 is probably more extreme than is usually found in actual biological data. At 0.75 the discrimination accuracy is still quite good. Figure 9 shows a plot of the amplitude discrimination of SIFT for the same kind of signals with three levels of asymmetry. While the discrimination curves have only slightly greater width for an asymmetry coefficient of 0.9 than for an asymmetry coefficient of 0.5, there is a definite shift in the mean values such that the more asymmetric a wave becomes, the more its amplitude is "underestimated." The calculated RMS amplitude of an asymmetric wave is exactly correct, so that the underestimation of amplitude is apparently due to the fact that the dependent variable wave is non-sinusoidal while the predictor wave is sinusoidal. The power in the harmonies of the dependent-variable wave is therefore not included in the P = 24 band. The estimate is, therefore, an accurate estimate of the amplitude of the fundamental.

It was also decided to investigate the two-period case where both of the waves in the signal were asymmetric. It is difficult to judge the suitability of such a model to real data situations. In real data, if two asymmetric waves occurred, they would probably be separated widely in the period domain. Alternatively, if two closely spaced waves occurred, they
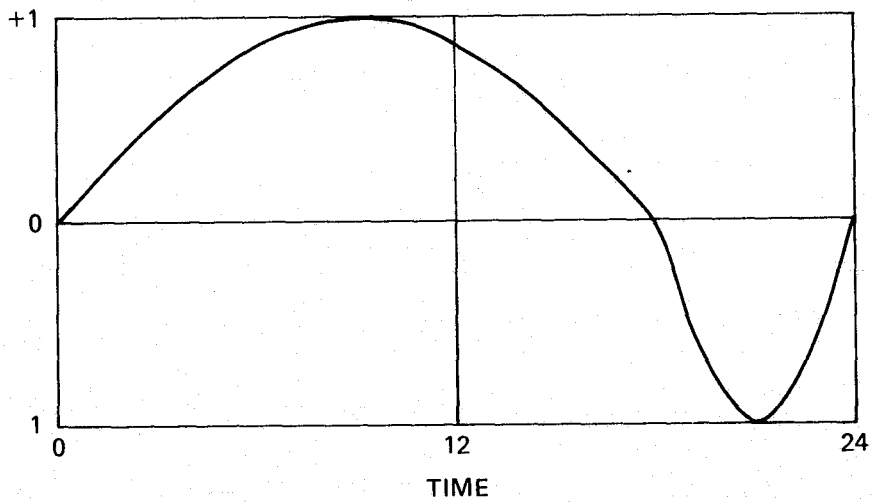
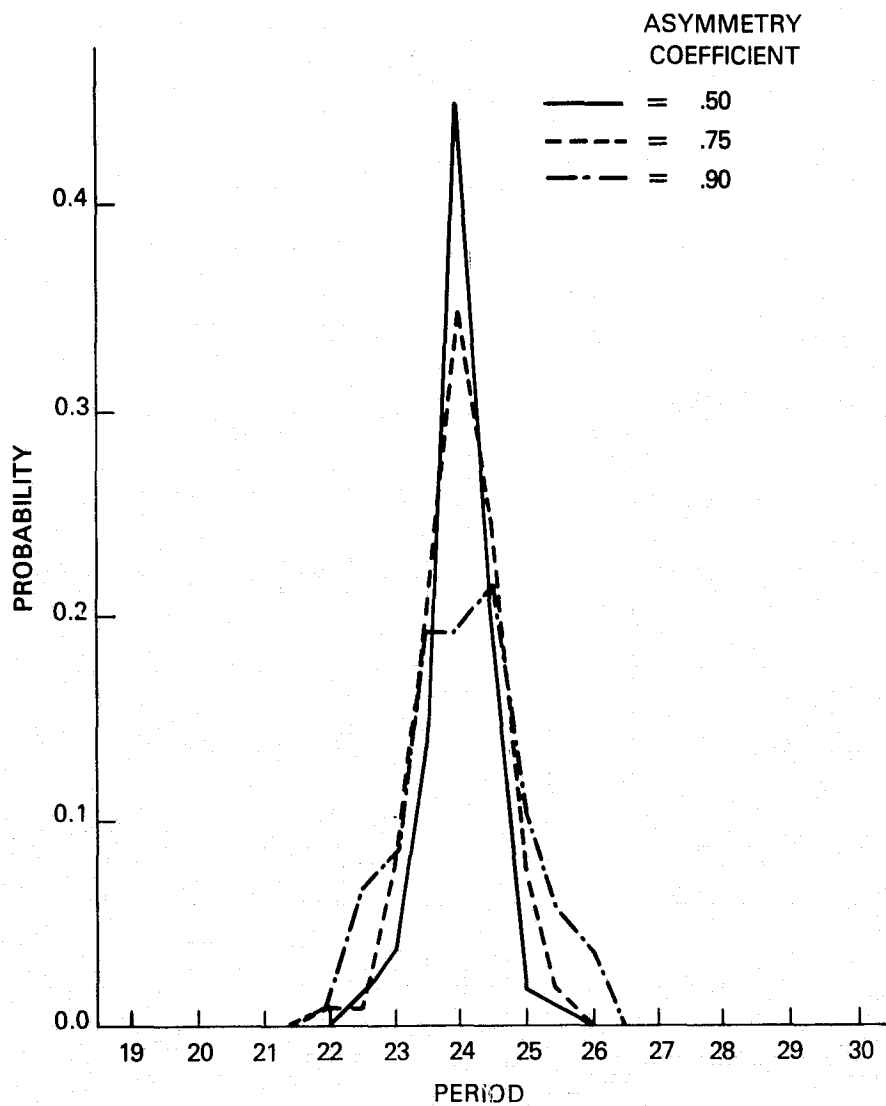Figure 7. A periodic wave with asymmetry = .75, P = 24.0.

Figure 8. Period discrimination for three single
asymmetric waves in noise of equal amplitude. P = 24.0,
A = 1.0, SNR = 1.0:1.0. Each curve represents the result
of a separate Monte Carlo run with a different level of
asymmetry.
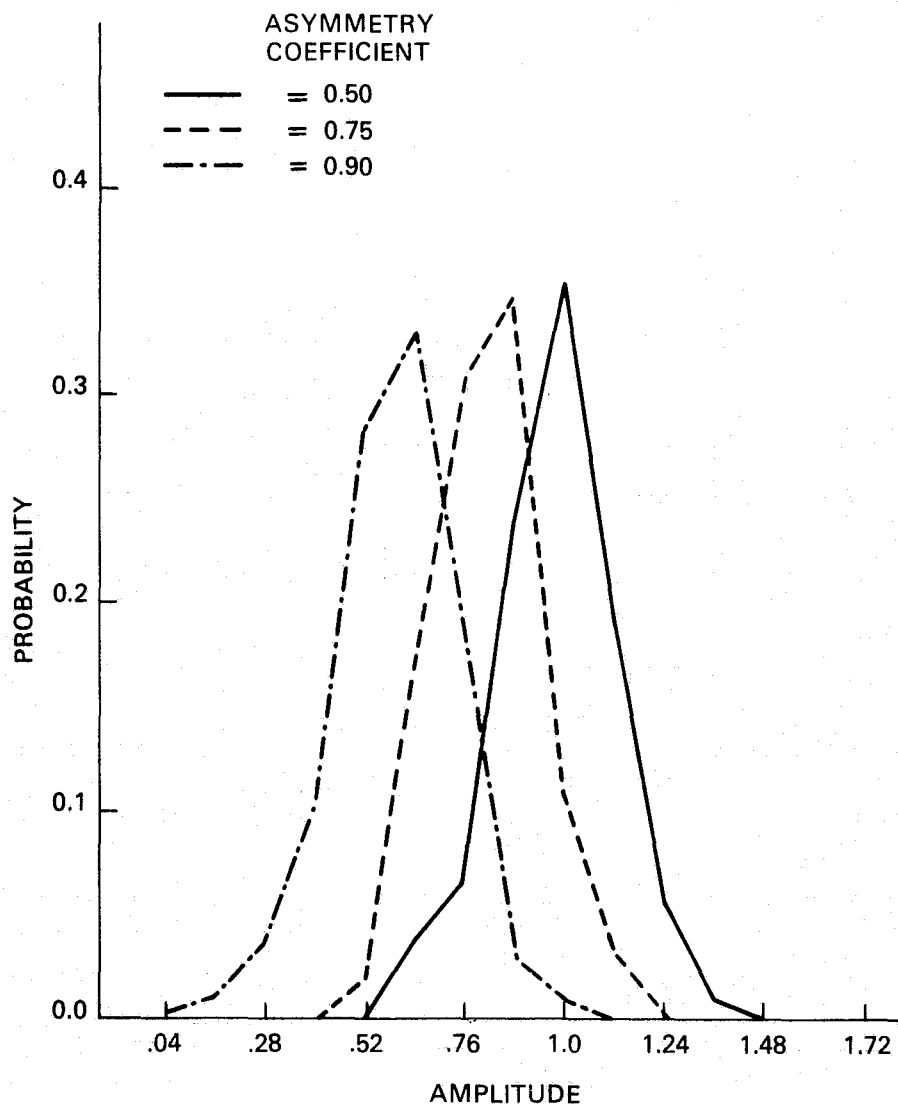
Figure 9. Amplitude discrimination accuracy for three single, asymmetric waves in noise of equal amplitude. P = 24.0, A = 1.0. Each curve represents the result of a separate Monte Carlo run with a different level of asymmetry.

would not necessarily both be asymmetric. Nonetheless, it was decided to run a Monte Carlo study to compare a signal with two asymmetric waves, each with an asymmetry coefficient of 0.75, to a signal with two sinusoids. In both cases a noise level of 1.0 was used (SNR = 1.0:1.0).

Table 7 shows the two kinds of error rates for the two kinds of signals.

TABLE 7. PROBABILITY OF TWO KINDS OF ERROR FOR A SIGNAL CONTAINING TWO SINUSOIDS AND FOR A SIGNAL CONTAINING TWO ASYMMETRIC WAVES, ASYMMETRY = 0.75[a]

Signal	Probability of Detectable Errors	Probability of Undetectable Errors
Two Symmetric Waves	0.16	0.10
Two Asymmetric Waves	0.11	0.15

[a] All sinusoids had an amplitude of 1.0, noise amplitude was 1.0, and t-to-enter was 5.0.

Apparently, the asymmetric pair of waves lead to identifying too-large amplitudes less frequently than is the case for symmetric pairs. However, with asymmetric pairs, too many or too few bands are found more frequently than for sinusoidal waves. Inspection of the results reveals that, with two asymmetric waves, the SIFT routine too often finds too few bands. Experiments done for this report have shown that if t-to-enter is reduced so that the SIFT enters two bands more frequently, the probability of finding too much power is prohibitively increased.

Figure 10 compares the period discrimination of the SIFT for Monte Carlo signals with two periodic waves for two levels of asymmetry. When two periodic asymmetric waves of asymmetry = 0.75 are found in a signal, it is apparent that the period discrimination of SIFT is degraded as compared to SIFT's performance on sinusoidal waves. The tendency of SIFT is to find more extreme bands (periods either too long or too short) in a signal containing a pair of asymmetric waves.
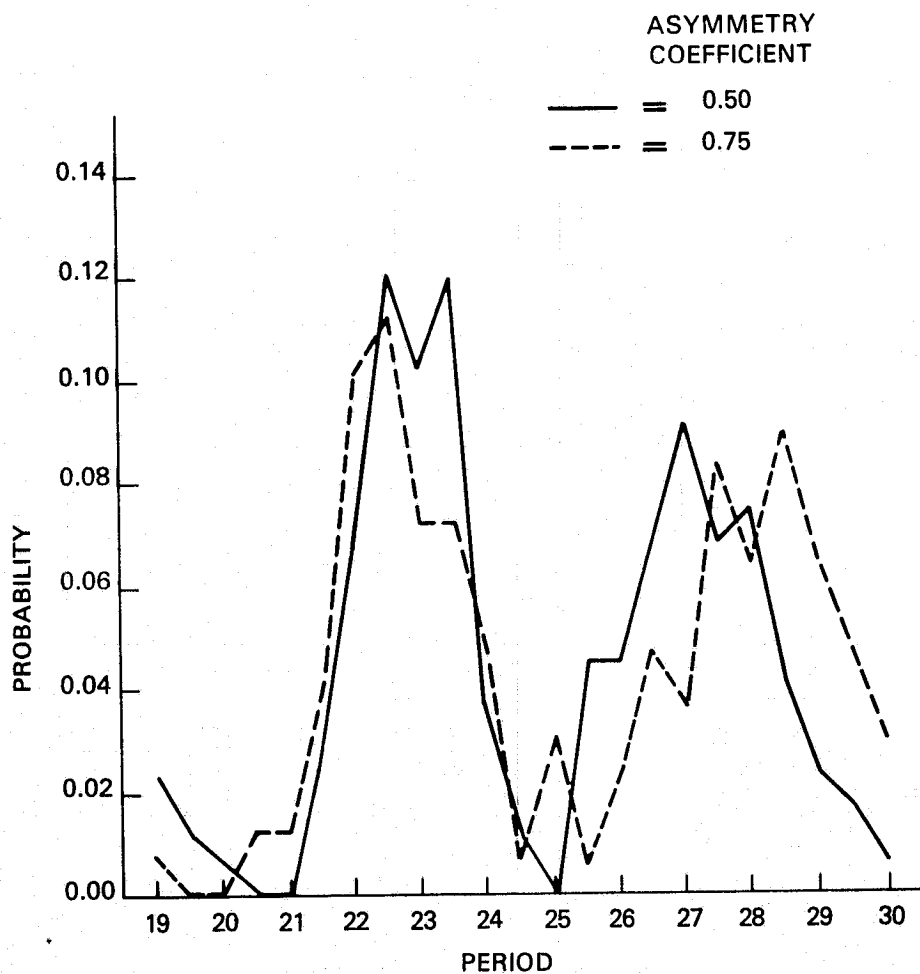
32

Figure 10. Period discrimination at two levels of asymmetry; two waves; $P_1 = 23.0$, $P_2 = 27.0$; SNR - 1.0:1.0; t-to-enter = 5.0.

Figure 11 compares the amplitude estimation accuracy for the two cases of
a signal containing two periodic (sinusoidal) waves. Again, due to the
non-sinusoidal nature of the asymmetric wave, its amplitude is "under-
estimated." The extra power is, of course, found in the harmonics. The
widths of the amplitude discrimination peaks are approximately equal, the
asymmetric wave peak being perhaps a bit narrower.

It appears, in summary, that (1) unless the asymmetry becomes quite extreme
or (2) unless there are two asymmetric waves closely spaced in the period
domain, the SIFT does a quite acceptable analysis. Frequency discrimina-
tion suffers little, and amplitudes are accurately estimated (except for
their harmonics).

Narrow-Band Noise

It is reasonable to argue that biological data are never strictly periodic.
There is usually variability in the peak-to-peak amplitude from cycle to
cycle, and there is also usually variability in the cycle length from cycle
to cycle. If these two variabilities are small (but non-zero) and Gaussian
in nature, the signal is not a strictly periodic one but is called a "narrow-
band Gaussian noise." As these two variabilities increase, the bandwidth of
the noise increases. Since biological signals are usually represented as
narrow-band-noises (however narrow the band might be), it was considered
important to test the performance characteristics of SIFT on effects of these
kinds of signals. In addition to the problem of narrow-band noises in "real"
biological data, many of the usual assumptions in spectrum analysis are
actually violated by the presence of periodic signals in time-series data.

A convenient way of generating narrow-band noise is to simultaneously modu-
late the amplitude and frequency of a sinusoid ("carrier signal") with
filtered Gaussian random noise. The procedure is as follows: (1) generate
two Gaussian, random, time series of length L, where L is the same as the
length of the narrow-band noise to be generated, (2) low-pass filter the
two Gaussian, random, time series so as to remove frequencies above about
1/10 of the frequency of the "carrier signal," and (3) use one of the two
low-pass, filtered, Gaussian time series to frequency-modulate the "carrier
signal" and the other to amplitude-modulate it. The results of such pro-
cedures are shown in figure 12.

The upper and lower waves are the same except for the "gain" of the modulators.
In subsequent discussion, the two waves of figure 12 are called "low-gain
modulated" and "high-gain modulated" narrow-band noises. The reader might
find it interesting to compare the waves of figure 11 with some real bio-
logical time series. Aschoff (1965), for instance, shows a plot of body
temperature across time which is quite similar in wave shape to the high-
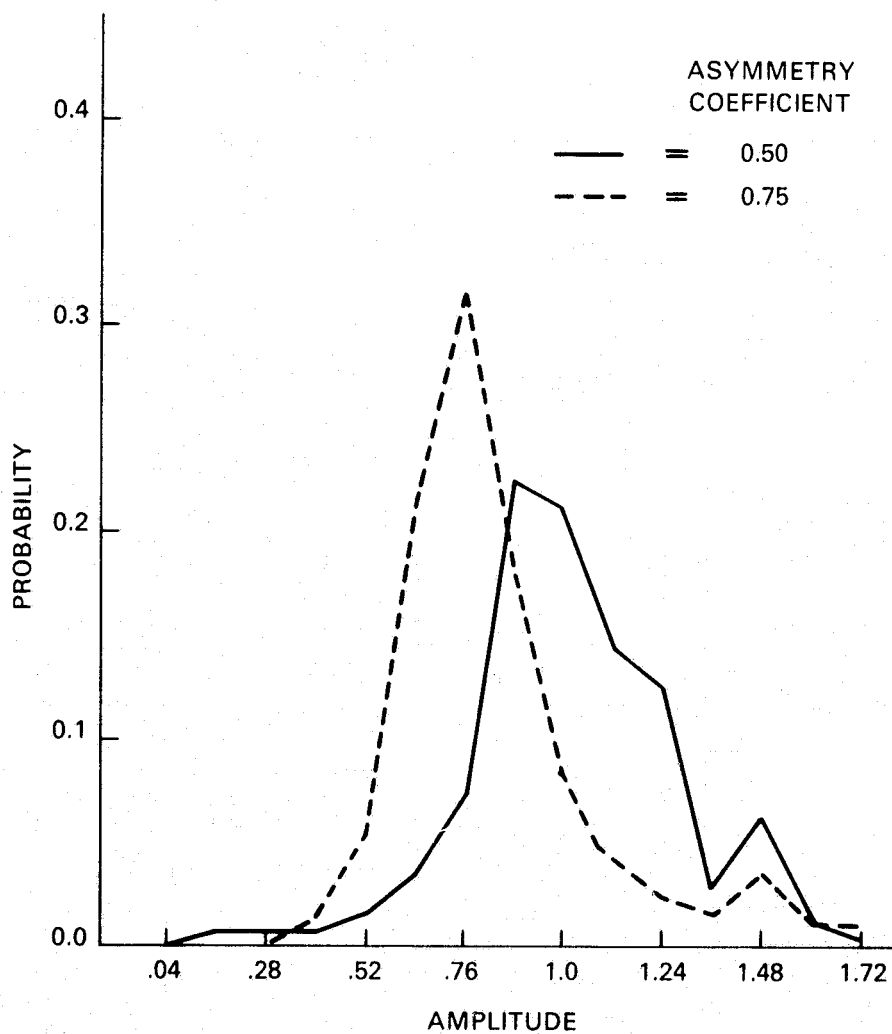gain, modulated, narrow-band noise.

Figure 11. Amplitude accuracy for Monte Carlo data
containing two periodic waves of equal amplitude
(A = 1.0); P_1 = 23.0; P_2 = 27.0; Noise = 1.0;
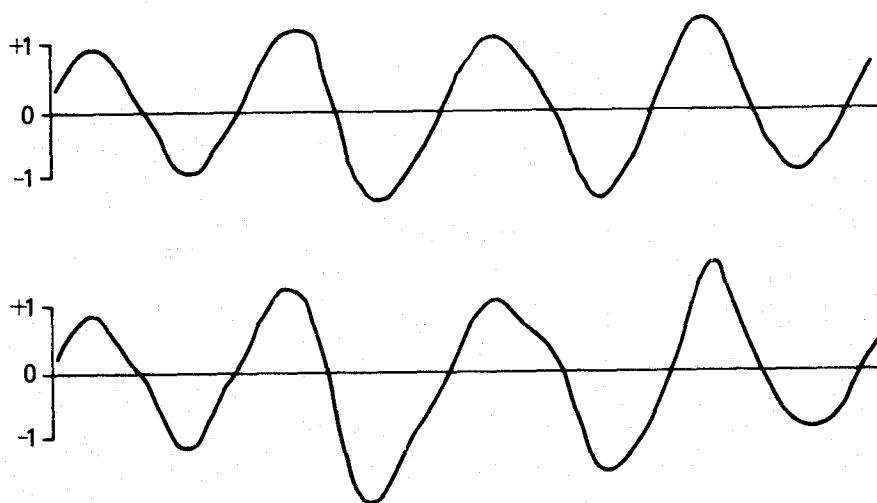t-to-enter = 5.0. Two levels of asymmetry.

Figure 12. Two narrow-band noises generated by simultaneous
amplitude and frequency modulation of a sinusoid by low-pass,
filtered, Gaussian, random time-series. The upper wave and
lower wave are generated from the same processes except that
the modulation indices are higher for the lower wave.

The high-gain, modulated, narrow-band noise was used in a series of Monte Carlo studies with SIFT. At values of t-to-enter which worked well for sinusoids of SNR = 1.0:0.5, the narrow-band noise produced alarmingly frequent instances of entering too many bands. When t-to-enter was increased, as shown in table 8, this type of error decreased. Since the narrow-band noise signals contain a wave of constantly changing period, the SIFT apparently properly attempted to fit several sinusoids, all of which account for different parts of the wave.

TABLE 8. SIFT PERFORMANCE FOR SINGLE SINUSOIDAL SIGNAL IN GAUSSIAN NOISE AND NARROW-BAND NOISE OF THE SAME PERIOD[a]

t-to-enter	Probability of Detectable Errors		Probability of Undetectable Errors	
	Narrow-band Noise	Sinusoid SNR 1.0:1.0	Narrow-band Noise	Sinusoid SNR 1.0:1.0
5.0	0.23	0.05	0.57	0.13
7.5	0.06	0.02	0.37	0.06
10.0	0.01	0.02	0.24	0.02
15.0	0.01	0.01	0.06	0.02
20.0	0.01	0.01	0.03	0.01
30.0	0.01	0.01	0.01	0.00

[a]P = 24.0, SNR - 1.0:1.0.

As shown in figure 13, however, if the SIFT is given a high enough t-to-enter value (in this case 20), the program most probably fits the correct period. When operating on a narrow-band noise, the discrimination of the SIFT is poorer than when it is operating on a sinusoid with SNR = 1.0:1.0. For those segments where the SIFT did fit waves, but not at P = 24.0, inspection of a few individual narrow-band noise runs indicated that the SIFT had selected from the signal those cycles about the period value which had the highest amplitude.

The amplitude-discrimination curve for the narrow-band noise data is shown in figure 14. The fact that SIFT usually "seized upon" the highest amplitude cycles to fit the first wave selected is further illustrated by the fact that most of the cycle amplitudes found were higher than 1.0. Thus, it is to be expected that SIFT will usually overestimate such amplitudes. As can be seen from figure 14, the program also occasionally underestimated the amplitude.
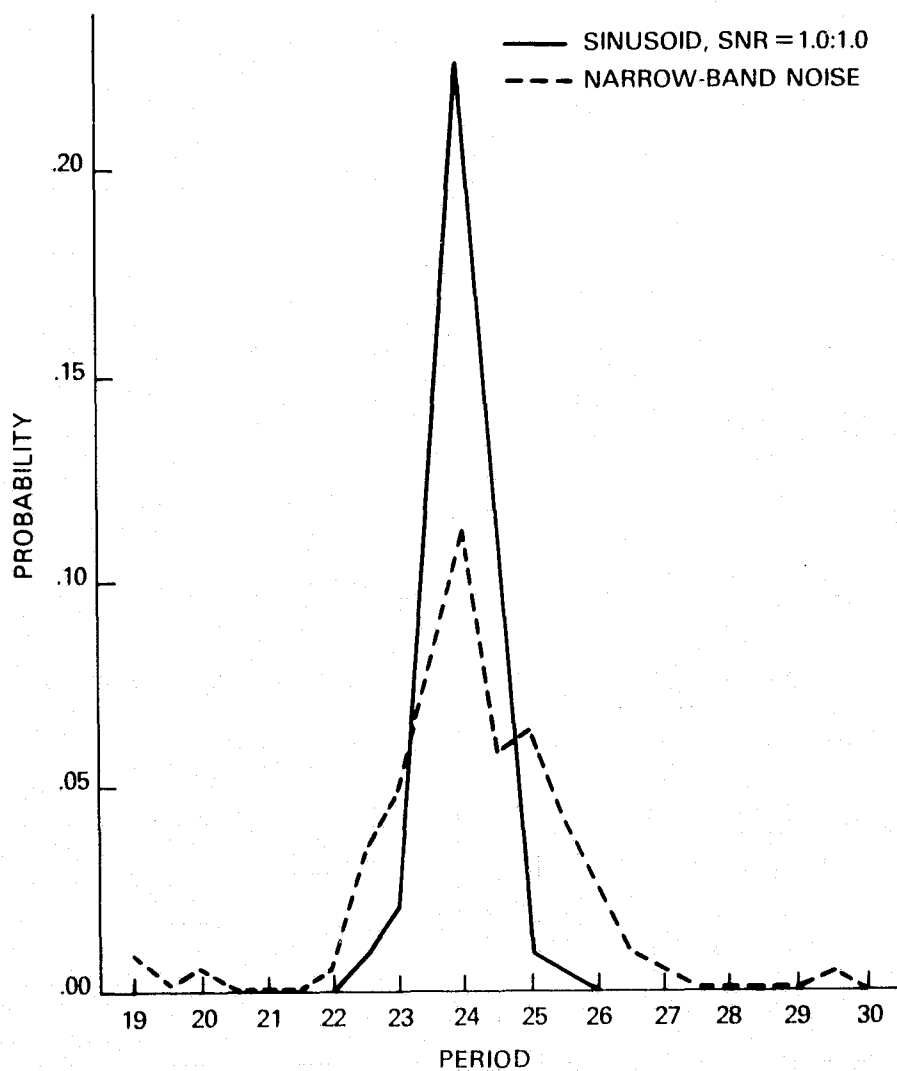
Figure 13. Period discrimination curves for a narrow-band
noise and for a sinusoid with SNR = 1.0:1.0. Narrow-band
noise waves generated from same process which produced
high-gain, modulated narrow-band noise of figure 12. Value
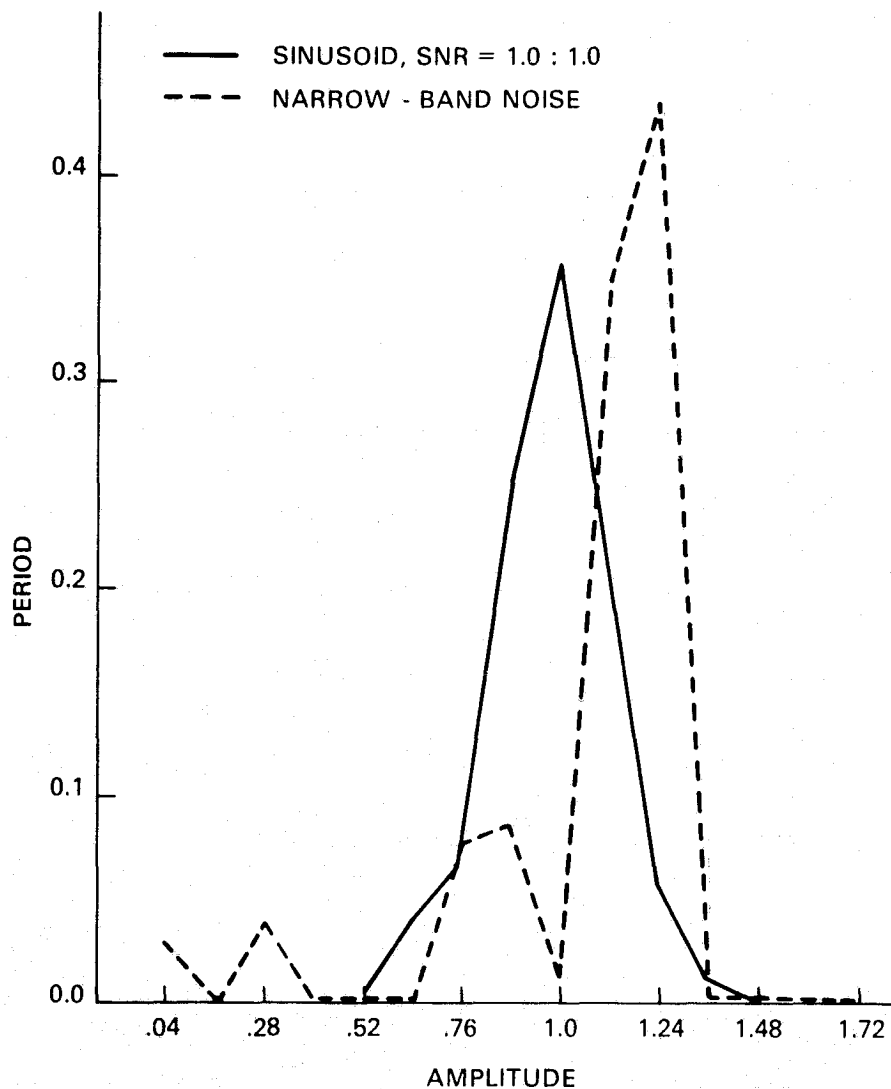of t-to-enter for narrow-band Monte Carlo run was 20.0.

Figure 14. Amplitude discrimination curves for a narrow-band
noise and for a sinusoid with SNR = 1.0:1.0. Narrow-band
noise waves generated from same process which produced high-
gain modulated, narrow-band noise of figure 12. Value of
t-to-enter for narrow-band noise Monte Carlo run was 20.0.

Although it is difficult to estimate the degree of generalizability of such a model to real biological data, an attempt was made to study the performance of SIFT with two narrow-band noises spaced more closely in ther period domain than are theoretically resolvable.

Since the effort to analyze two high-gain, modulated, narrow-band noises ended in almost complete failure to find consistent data, two low-gain modulated waves were generated and analyzed in a Monte Carlo run. The t-to-enter values were found here to be broadly optimized around 20.0, but, as shown by table 9, the probability of finding the wrong number of peaks was still prohibitively high. Figure 15 shows the frequency-discrimination curves for the 2-period case. From this curve it may be seen that the lower period is estimated with about the same accuracy as two sinusoids, with an SNR = 1.0:1.0. The longer period is found less often and is usually estimated as too long.

TABLE 9. OPTIMIZED ERROR PROBABILITIES OF SIFT
WHEN ANALYZING TWO LOW-GAIN, MODULATED, NARROW-BAND NOISES

Probability of Detectable Errors	Probability of Undetectable Errors
0.04	0.52

The empirical results from the narrow-band noise Monte Carlo runs demonstrated a basic property of the SIFT program. With these runs the band-width of the population of all possible signals was not zero (as with a sinusoid) but rather a population of time series with a bandwidth greater than zero. When the SIFT is used to analyze such data, it will pick the period containing the cycle with the largest amplitude in the particular segment being sampled and analyzed.

If the t-to-enter value is low enough, it will then pick the period of the next most prominent cycles in the particular sample segment analyzed. This is indeed an accurate analysis of a narrow-band noise segment, but it must be interpreted correctly by the user. From all that has been determined empirically, if the SIFT finds two periods in close proximity in the period domain, it could mean that some cycles were of one period and other cycles were of the other period. If such a result is found in real biological data, the implication is that the data are either composed of two closely spaced sinusoids or of a narrow-band noise containing individual cycle lengths, as
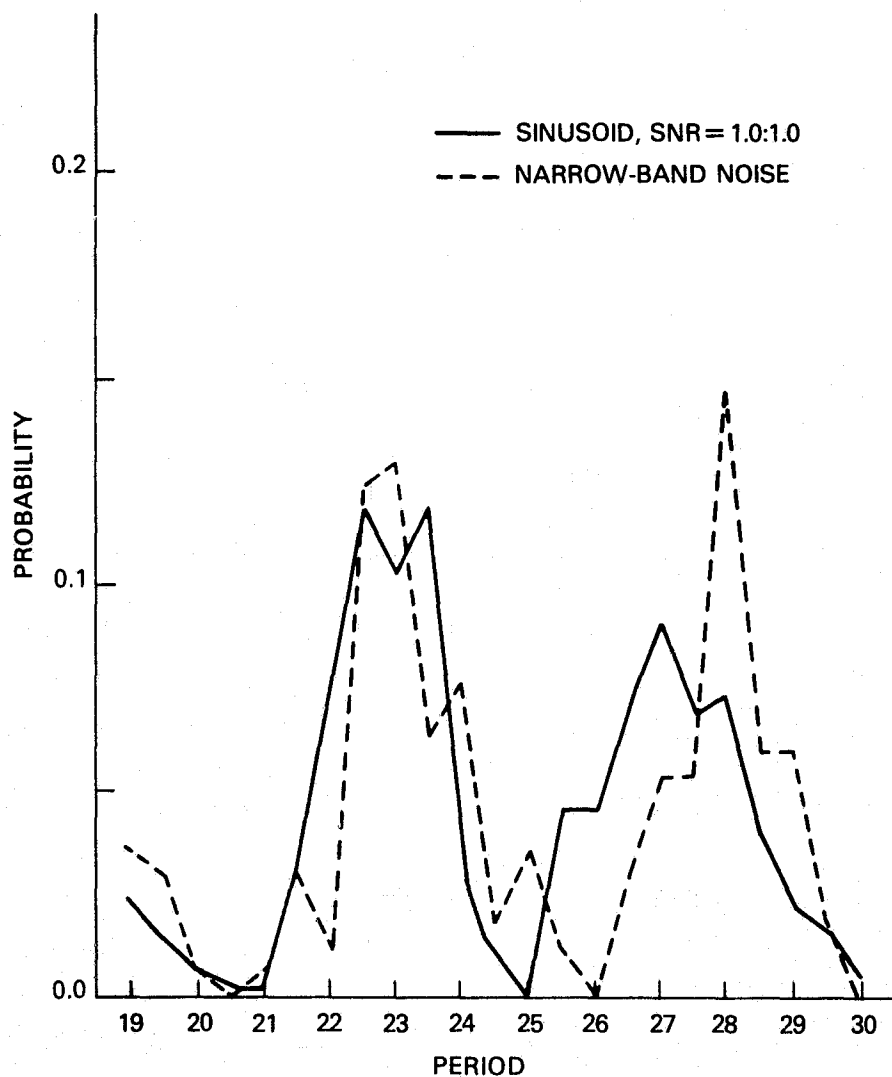
40

Figure 15. Period discrimination for two sinusoids with SNR = 1.0:1.0 and for two narrow-band noises. Mean period values for narrow-band noise waves were the same as for sinusoids; P_1 = 23.0; P_2 = 27.0. Sinusoids were mixed with noise: SNR = 1.0:1.0, narrow-band noises were constructed without added noise and had amplitudes of 1.0.

found by the SIFT. Furthermore, the fact that the SIFT frequently finds the "wrong" number of peaks in a Monte Carlo narrow-band noise probably only implies that the cycle lengths of that particular segment were more or less homogeneous and the SIFT (at least approximately) reported them correctly. The utility and interpretation of any SIFT results, then, depend upon the philosophical position that the investigator holds with respect to life-sciences data. If narrow-band noise is considered the appropriate model, then SIFT probably draws the user's attention to a great deal of irrelevant detail about a particular sample segment from a population, which would not necessarily be repeated in exactly the same way on another segment sampled from the same population. This is especially true if the t-to-enter value is set too small. The "best" t-to-enter figure is, of course, determined by (1) the amount of white, Gaussian noise in the signal, (2) the bandwidth of the narrow-band noise, and (3) the amount of detail that the user wishes to use in describing a wave. It is to be stressed that these "unknowns" are not readily determined when time-series data collected from biological systems are analyzed.

Missing or Asynchronously Sampled Data

In many biological data acquisition systems, the probability of losing data is relatively high. Transitory instrumentation failure, "noise" from arti-facts such as muscle potential, atmospheric noise in telemetry systems, etc., can all produce segments of "missing data" in a time series.

In some cases of biological time series, it would be extremely difficult (and often artificial) to sample values at regular time intervals. Urine volume measures are such a variable. Usual forms of spectrum analysis assume that data are sampled synchronously and continuously. The SIFT was written, however, in such a manner as to circumvent this problem.

At the suggestion of Dr. John Rummel (1974), the subroutine in the SIFT which generates the sine/cosine estimator waves for the least-squares spectrum equation was modified as follows: Instead of internally generating a "time axis" of regularly spaced points which serve as the arguments for sine and cosine function calls, the subroutine reads a series of time points as input data. The time points that are utilized are only the read-in times at which the corresponding samples in the data were collected. This time vector is then used as a series of arguments for the sine/cosine function calls. For the purposes of the Fourier (regression) analysis, the important aspect is that the sine/cosine estimator wave is sampled at the same time that the dependent variable is sampled.

Monte Carlo studies were run to explore the effects of several types of asynchronous sampling, as follows: (1) "jitter" in the sampling rate, (2) sampling in "clumps" separated by times when no samples were collected, and (3) relatively long gaps in the data (to be called the "missing-data" case). The jittered sampling was achieved by generating a synchronous sample

series with a random number of $\pm$ 1.0 sampling unit (maximum) added to each time point. The clumped sampling was done by generating 10 samples for $\underline{t}$ = 0.0 to $\underline{t}$ = 5.0, then skipping (not taking samples) from $\underline{t}$ = 5.0 to t = 10.0, sampling from t = 10.0 to $\underline{t}$ = 15.0, etc. The missing data sampling was done in three ways: (1) 10% missing from the middle, (2) 10% missing during the first third, and (3) 10% missing during the final third. All of these tests were performed with sinusoidal data having two periods (P_1 = 23.0; P_2 = 27.0) and an SNR = 1.0:1.0. In all cases, a total of 100 points was used.

Optimization runs on the above kinds of asynchronously sampled data showed that optimum $\underline{t}$-to-enter was around 10.0, the optimum value also for synchronously sampled data of those specifications. The comparisons to follow, however, are made at a $\underline{t}$-to-enter of 5.0 because error rates were so near zero in the synchronously sampled data that it was felt that a slightly higher rate would show differences better.

Table 10 shows the error probabilities for synchronously sampled data, jittered sampling, and clumped sampling. In the two cases of asynchronous sampling shown, the error rates are actually lower than those for synchronous sampling. Why this should occur is not clear, but differences of this size, where (in fact) slightly different parts of the same data are used, are probably not important. Figure 16 shows the period-discrimination curves for the three cases shown in table 10. Obviously, only minor differences are seen. For jittered or clumped sampling, then, it may be concluded that no deleterious effects occur over synchronous sampling.

TABLE 10.   ERROR RATES FOR SYNCHRONOUS,
JITTERED, AND CLUMPED SAMPLING

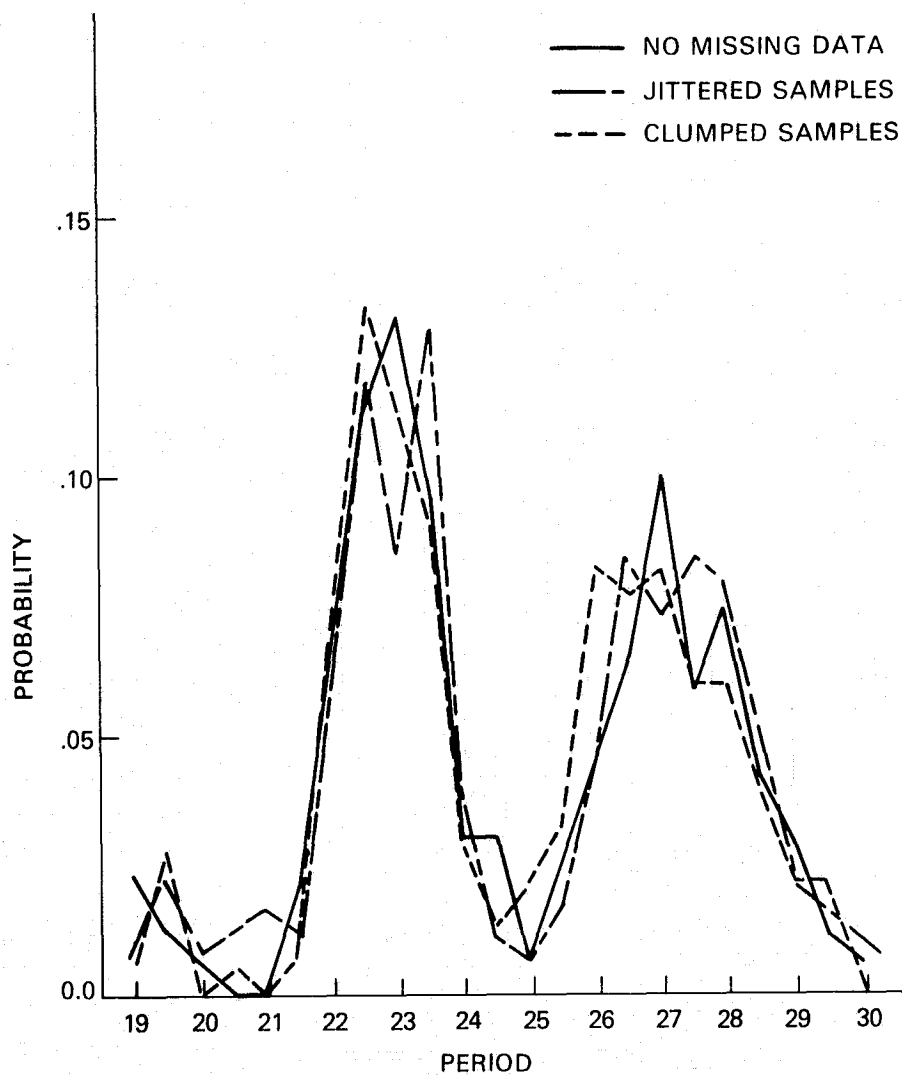| Type of Sampling	Probability of Detectable Errors	Probability of Undetectable Errors
Synchronous	0.14	0.20
Jittered	0.15	0.09
Clumped	0.12	0.06

Figure 16. Period discrimination for two kinds of asynchronously sampled time series. In all runs: $P_1 = 23.0$, $P_2 = 27$, SNR = 1.0:1.0, and t-to-enter = 5.0.

Table 11 and figure 17 show comparisons for the three kinds of missing data, i.e., where 10% of the data was missing in one of three places in the time-series wave. Considering that the time-series data used in this test were composed of two sinusoids, closely spaced in the period domain, the differences observed in table 11 and figure 17 might easily be due to the amount of redundancy between the two waves (which almost overlap in the time domain in some parts of the record). Under any circumstances the error scores are always better for the missing data (again no possible explanation is offered), and the period discrimination scores overlap well with the synchronously sampled case.

TABLE 11. ERROR RATES FOR 10% MISSING DATA IN ONE
OF THREE PLACES OF A TIME-SERIES RECORD

Type of Sampling	Probability of Detectable Errors	Probability of Undetectable Errors
Synchronous	0.14	0.20
Missing data in middle	0.09	0.08
Missing data in 1st third	0.04	0.08
Missing data in 3rd third	0.07	0.10

In conclusion, it is probably safe to say that missing data and asynchronous sampling are not important considerations in considering the use of the SIFT on "real" life-sciences data. The practice of not attempting to analyze for any period shorter than twice the shortest inter-sample interval should be followed. Also, if no smoothing or accumulators are used during data collection, the minimum sampling rate must be twice that of the highest frequency fluctuation present in the data.

Real Data Analysis Test

In order to test the SIFT program on real data, numeric data were obtained by manually scaling the record of urine volume over a 10-day period, as reported in Aschoff (1965). Figure 18 shows a plot of this record interpreted
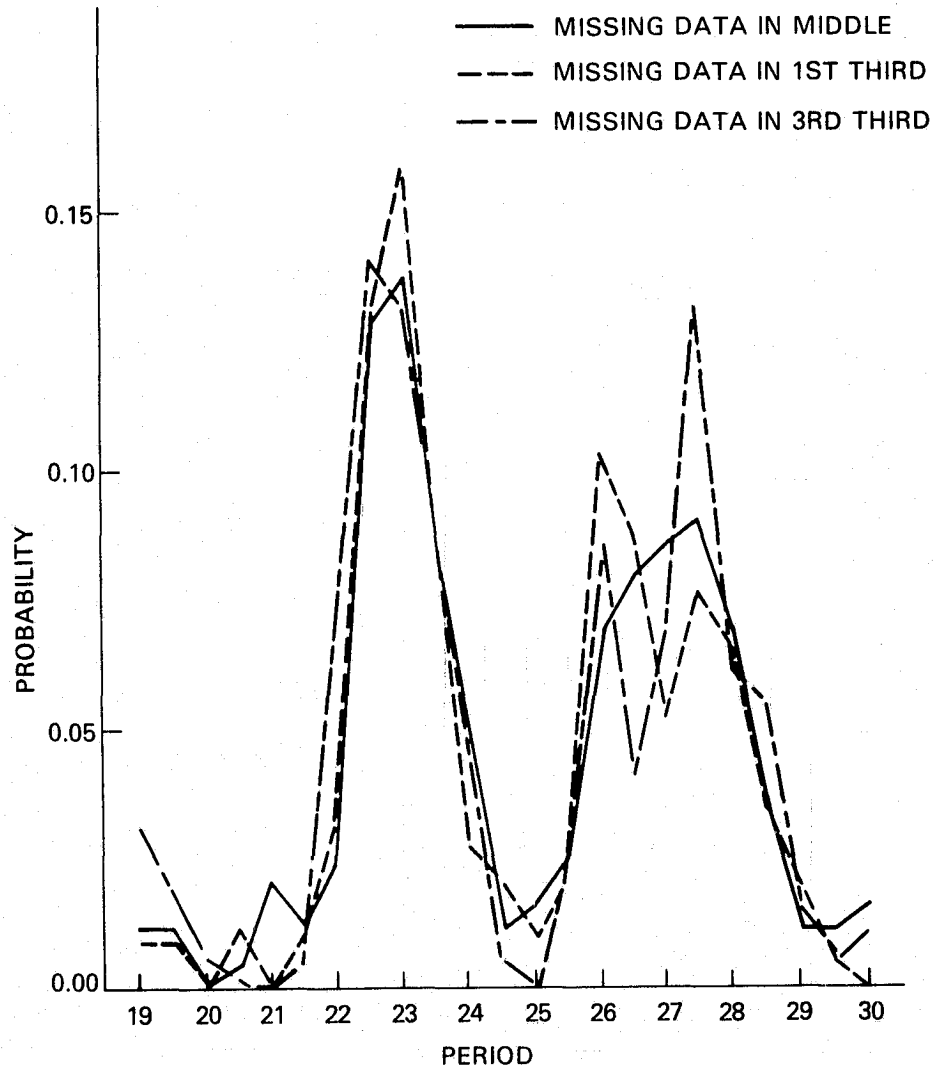
Figure 17. Period discrimination for 10% missing data
in one of three places of a time series record. In all
cases: P_1 = 23.0, P_2 = 27.0, SNR = 1.0:1.0, and $\underline{t}$-to-
enter = 5.0.

as an interpolated curve rather than as the step function shown by Aschoff. It is readily seen that each day the urine volume peaks at some time which varies across days. Not only do the widths and times of occurrence of urine-volume cycles vary across days, but the height of the peaks also varies. The cycles in figure 18 are not sinusoidal and might, therefore, fit the asymmetric, periodic data model with random noise added. Alternatively, the appropriate model might be a narrow-band noise model. The narrow-band noise model clearly does not, however, account for the apparent cycle asymmetry. Possibly some model involving asymmetry plus random-cycle width and height might be appropriate. Such a wave might be made symmetrical by performing a log transform before SIFT analysis.

It was decided to perform several different types of SIFT analyses on these data. It would be possible to analyze them in the frequency domain by making frequency spectrum estimates beginning at 0.1 cycle per day, spacing other estimates at 0.1 cycles per day. This frequency spectrum would be computed at appropriate resolution, i.e., at frequency intervals of 1/T. The highest frequency that could be analyzed in the spectrum would be about 2 cycles per day. This limitation is due to the fact that one or two of the measures of urine volume were spaced at 8-hour intervals, although most of the inter-sample times were shorter than 4 hours. Inferences about frequencies higher than 2 cycles per day should be viewed with caution, because most of the days contained 4 to 5 samples.

It was decided, therefore, to perform a SIFT on the Aschoff data in the period domain. The shortest analyzable period would be about 0.5 day and the longest one possible, about 10 days. Since it was felt that the major interest would lie between 0.5-day to about 2.5-days, it was decided to analyze for periods in the range of 0.5- to 2.75-day length.

The space between adjacent period estimates, $\triangle P$, was set at 0.05 day. Table 12 shows the equivalent frequency-domain spacing for three points in the period-domain spectrum. Clearly, with $P = 0.05$ and $T = 10.0$ days, the period-domain spectrum is well overresolved at $P = 2.75$, is approximately properly resolved at $P = 1.0$, and is quite a bit underresolved at $P = 0.5$. Perhaps the most serious implication of these resolution difficulties is that there are considerable gaps in the spectrum in the region of $P = 0.5$. It is possible that much energy could be missed by this type of analysis if a narrow peak existed between, say, $P = 0.50$ and $P = 0.55$. This analysis ($\triangle P = 0.05$, $T = 10$) was performed as a "first cut" at the data. Subsequent runs will be discussed where the problem of resolution is less extreme.
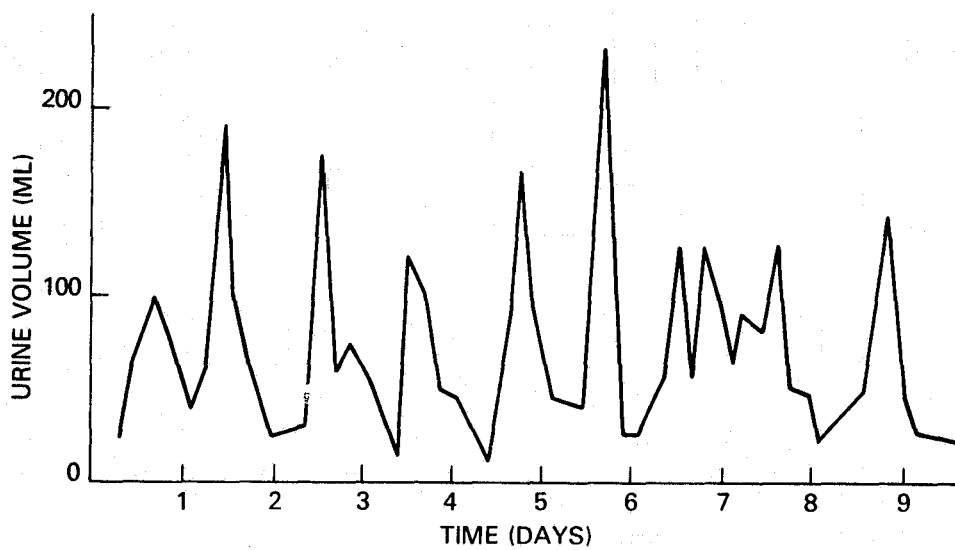
Figure 18. Urine volume as a function of time over a
10-day period. After Aschoff (1965).

TABLE 12. FREQUENCY-DOMAIN SPACING OF SPECTRUM ESTIMATES
FOR VARIOUS PERIODS IN A PERIOD-DOMAIN SPECTRUM

P	f	Δf
0.50	2.000	
0.55	1.818	0.182
1.00	1.000	
1.05	0.952	0.048
2.70	0.370	
2.75	0.356	0.014

Figure 19 is a plot of the raw period-domain spectrum, computed as described above. This plot represents a DFT result which is the first spectrum computed by the SIFT program. It can readily be seen that the spectrum in figure 19 is relatively smooth and has broad peaks toward the long-wave end, while the peaks at the short-wave end are sharp and narrow. From the discussion above regarding the resolution in various parts of the period spectrum, it is seen that the broad, long-wave peaks are composed of many highly redundant estimates and might, therefore, represent a narrow-band spectrum element even though the peaks are quite broad. Similarly, the short-wave peaks are jagged and narrow because each peak represents a unique, nonoverlapping piece of information with, in this case, a gap of noncovered spectrum data between. The spectrum estimates around P = 1.0 represent approximately properly resolved data.

The program next performed a SIFT on the Aschoff data. For this analysis, the t-to-enter value was set at 15.0. This value is quite possibly too low if the data at hand really are narrow-band noise. On the other hand, if SIFT is permitted to print out its stepwise solutions, even if more periods are fitted in the final solution than the user believes reasonable and parsimonious (whatever ground there might be for such belief), then the user can always select that step which he considers most desirable. Table 13 shows the periods with their amplitudes and t-values at each step during the SIFT.
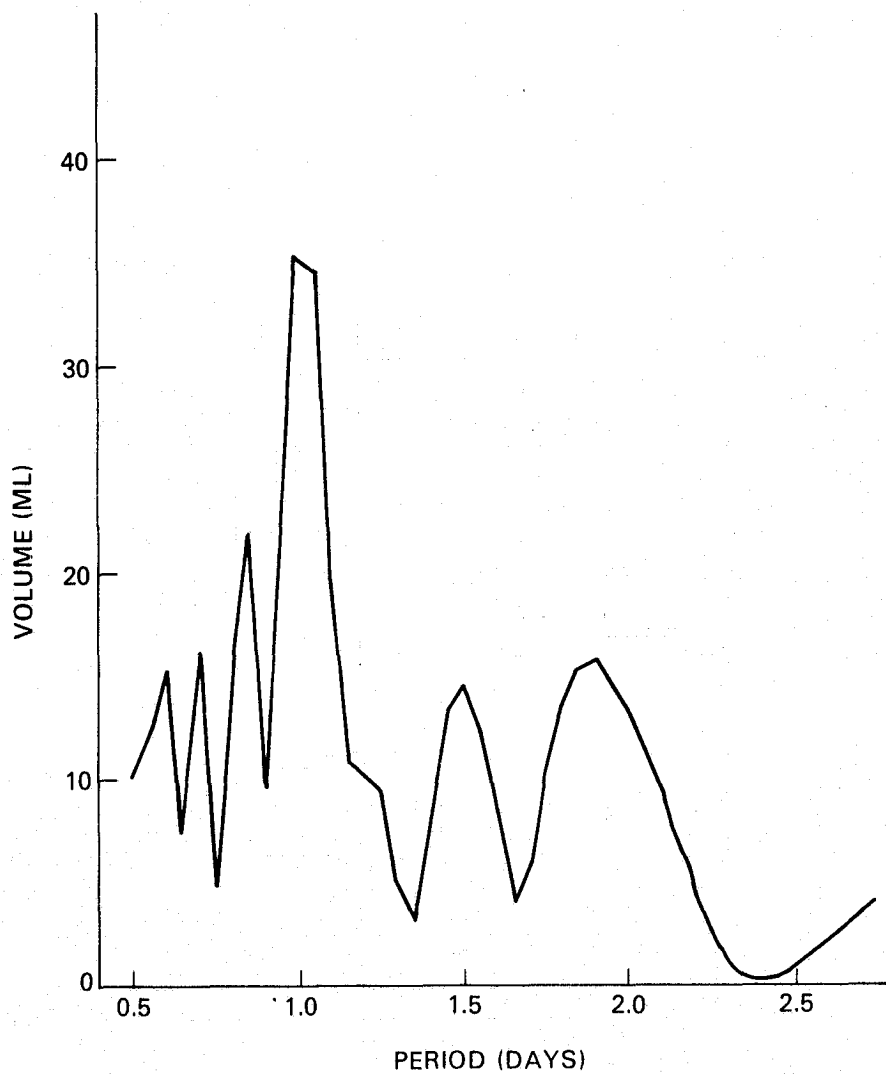
Figure 19. Plot of raw period-domain spectrum of the urine volume data shown in figure 18, as computed by the DFT; $\Delta P = 0.05$.

TABLE 13. STEPWISE PERIOD-ANALYSIS RESULTS FROM
A SIFT OF THE TIME SERIES SHOWN IN FIGURE 18[a]

Step	Period	Amplitude	t-Value
1	1.00	35.77	29.35
2	1.00	39.49	33.54
	1.90	21.77	16.21
3	1.00	39.73	34.19
	1.10	18.96	18.20
	1.90	22.53	21.05
4	0.85	20.82	16.24
	1.00	42.97	37.84
	1.15	17.40	8.74
	1.90	20.26	16.84

[a]Period domain; $\triangle P = 0.05$; t-to-enter = 15.0.

Steps 1 and 2 showed logical and interpretable results. The first period entered
was the 24-hour (circadian) rhythm, and it was clear from Figure 19 that there
was a great deal of power in that band. The second period entered, P = 1.9,
is at a nearly two-day period length. Some subjects, when placed in sensory
isolation, spontaneously go into a 48-hour rhythm. This might be a rhythm
coexisting with the circadian but of lower amplitude. A sufficiently imagina-
tive look at figure 18 might permit a viewer to "see" that the volume output
around the end of even-numbered days is lower than that around the end of
odd-numbered days. Perhaps this is the 2-day rhythm that SIFT has found.

Steps 3 and 4 added components that are not easily explained. Perhaps they
are due to statistical deviations in cycle width for this sample segment
taken from a population of narrow-band noise segments, and for this reason
ought to be ignored.

On the other hand, they might represent actual periodic processes. At this
point there is no way of knowing which is the "true" case from an empirical
point of view. If a number of repetitions of the study on the same subject
(or a group of subjects) always show these periods, then it would be worth-
while to try to explain them. If not, they must be attributed either to
random variability or to an uncontrolled variable. Of course, all periods
must stand up to the criterion of replicability.

At this point in the analysis it was decided to do something about the gaps between adjacent spectrum estimates for short period lengths as well as to improve resolution around $P = 1.0$. One way of doing this might have been to run a new period domain spectrum where $\triangle P = 0.02$ rather than $\triangle P = 0.05$. This would have resulted in some overresolution at $P = 0.5$ and would thus have improved the resolution at the shorter period lengths. The problem with this would be that, for a fixed-spectrum vector length (limited by memory requirements), a shorter long-wave limit would have obtained. Because the $P = 1.9$ band contained considerable power and would have been eliminated by this procedure, an alternative was sought.

Since the period domain spectrum had already computed a highly overresolved, long-wave spectrum ($1.0 < P < 2.75$), it was decided to run a frequency-domain SIFT on the data of figure 18. Figure 20 shows this raw frequency-domain spectrum. This spectrum is computed with a $\triangle f = 0.05$, which represents about twice maximum resolution. This resolution is, for the frequency domain, uniform for all parts of the spectrum. Another advantage of the frequency domain spectrum is that it entirely covers all possible frequencies without gaps.

As can be seen from figure 20, the two major peaks are at 1.0 and about 1.9 cycles per day. The width of the peaks is about equal (as would be expected from theory).

Table 14 shows the frequencies, amplitudes, and t-values at each step of the SIFT.

TABLE 14. STEPWISE FREQUENCY-ANALYSIS RESULTS FROM
A SIFT OF THE TIME SERIES SHOWN IN FIGURE 18[a]

Step	Frequency	Amplitude	t-Value
1	1.90	32.79	29.76
2	1.00	31.08	27.22
	1.90	27.31	25.96
3	1.00	30.04	27.25
	1.75	21.09	20.53
	1.90	27.70	27.95
4	0.50	21.73	21.14
	1.00	33.61	32.11
	1.75	22.59	23.28
	1.90	26.94	28.85
5	0.50	22.49	22.58
	0.90	17.63	16.58
	1.00	36.87	36.32
	1.75	22.92	24.67
	1.90	24.14	25.16

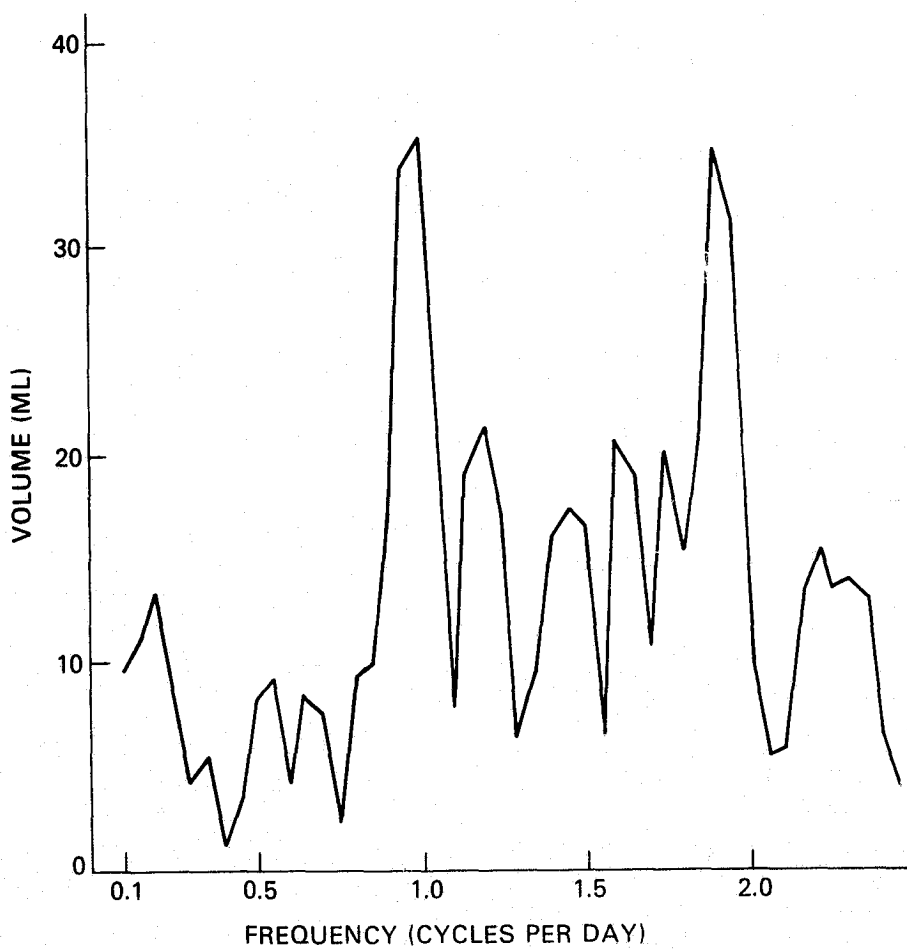[a]Frequency domain; $\triangle f = 0.05$; t-to-enter = 15.0.

52

Figure 20. Plot of raw frequency-domain spectrum of
urine volume data shown in figure 18, as computed by
DFT; $\Delta f = 0.05$.

Here (as with the period-domain analysis) the first few steps are clearly believable from examination of the raw spectrum and, again with some imagination, from the time-series record. The circadian rhythm has the highest amplitude, and the peak at 1.9 cycles per day is possibly related to the first harmonic of the circadian frequency. The fact that the actual frequency is not 2.0, however, militates against this notion. As with the period-domain spectrum, periods entered after the first two steps are more difficult to interpret, although the 0.5 cycle per day (P = 2.0) did eventually make itself evident as salient. The fact that the 0.5-cycle-per-day frequency was entered on the 4th step in the frequency-domain SIFT but was entered instead on the second step in the period-domain SIFT is difficult to explain. One clue is that in the period spectrum the band of P = 2.0 was in an area of the spectrum where the data were highly overresolved and perhaps, therefore, over-represented. Another possibility is that the different results occurred because the frequency spectrum has a wider coverage and different resolution. In any event, the two domains appear to provide perhaps complementary pieces of information about the time series. Probably the frequencies of highest reliability are (1) the circadian, (2) a 2-day period, and (3) the first harmonic (0.5 day) of the circadian.

It should be pointed out that the SIFT's output must be considered in a statistical light, just like any other descriptive procedure. The output from an analysis of one subject, one trial, is only one observation in k-space, where k is the number of bands in the output of the SIFT. In order to interpret SIFT results, one must have many observations (both trials and subjects), and one must compute the usual summary data, such as means and significance tests. However, these further computations, particularly the problem of hypothesis testing, represent the subject of an entirely different discourse.

CONCLUDING REMARKS

Studies of biological processes such as the circadian rhythm during manned space flights and flight simulation are usually subject to significant constraints. Testing may be restricted to single individuals, or small groups at best, and massive amounts of data may be acquired for analysis purposes over a period of days. The purpose of biomedical studies may be exploratory, in that they are often addressed to detection of any medically significant deviations from normal physiological functioning. For these kinds of purposes, time-series analysis techniques can be appropriately used, and their use can permit detection of changes in biological processes that cannot be discovered through visual analysis or conventional statistical treatment of the data.

The correct application of classical spectrum analysis techniques involves taking into account some specific design considerations in initial construction of the experimental design. Even when this is done, however, spectrum analysis of the resulting data can present problem situations for which the appropriate treatment has not been clearly established.

The present program was undertaken to examine some problems commonly experienced in spectrum analysis of biomedical data, such as gaps in the data, different time-series lengths, and periodic but nonsinusoidal processes in the data being analyzed. Through this work, it has been possible to develop an improved type of spectrum analysis procedure and to examine how results obtained with this procedure are specifically affected when the problems identified above occur. The findings have shown that the SIFT (Stepwise Iterative Fourier Transform) can reliably reduce data of the nature of (1) sinusoids in noise, (2) asymmetric but periodic waves in noise, and (3) sinusoids in noise during which sampling was asynchronous and/or data were missing. The program was also able to analyze narrow-band noise well, but substantial interpretational problems became apparent. Specifically, on any "real" data analysis, it would be very difficult to determine the appropriate model for the data from the analysis results; and, unless the model is known a priori, it is difficult to set such parameters as t-to-enter.

The t-to-enter problem must be handled via philosphical methods and considerations. The results must be "reasonable," for instance, in terms of consistency with previously developed knowledge concerning the processes under study. The final criterion of any analysis is, of course, that of replication, and this is complicated because different t-to-enter values can influence the number of spectrum peaks found in the data. With this type of program, it is not advisable to leave all the decisions to the computer. The indiscriminate use of SIFT could lead to very serious inferential errors.

The SIFT has been shown to be a powerful data reduction technique for elucidating the main structure of time-series records in the frequency or period domain. Before the utility of the SIFT can be fully assessed, more experience with large volumes of real data must be acquired.

REFERENCES

Rummel, J. A.   Rhythmic variation in heart rate and respiration rate during space flight - Apollo 15.   Proceedings of the International Symposium on Chronobiology.   Little Rock, Ark., October   1971.

Benignus, V. A.   Statistical assessment of a unique time series analysis technique.   Behavioral Technology Consultants, Inc., Final Report (NASA contract NAS 9-11784), 1972.

Hinich, M. J. & Clay, C. S.   The application of the discrete Fourier Transform in the estimation of power spectra, coherence and bispectra of geophysical data.   Review of Geophysics, 1968, 6, 347-363.

Bendat, J. S. & Piersol, A. G.   Measurement and analysis of random data. New York: Wiley, 1966.

Guilford, J. P.   Fundamental statistics in psychology and education. 4th edition. New York: McGraw-Hill, 1965.

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

Draper, N. R., & Smith, H.   Applied regression analysis.   John Wiley, New York, 1966.

Efroymson, M. A.   Multiple regression analysis.   In Ralson, A., & Wilf, H. S., (Eds). Mathematical methods for digital computers.   John Wiley, New York, 1962.

Aschoff, J.   Circadian rhythms in man.   Science, 1965, 148, 1427-1432.

Rummel, J. A.   Personal Communication.   1974.