Signal-to-Noise Enhancement Through Instrumental Techniques

Part I. Signals, Noise, and S/N Enhancement in the Frequency Domain

G. M. HIEFTJE, Department of Chemistry, Indiana University, Bloomington, IN 47401

A consideration of some relatively simple techniques which enable better signal collection can result in improved analytical measurements. Since an increase in measurement time or a loss of resolution can occur simultaneously, a selected trade-off is necessary

According to communication theory, every measurement which is made involves the observation of a signal. To this way of thinking, all of analytical chemistry, indeed most of physical science, requires the examination of a vast number of signals in widely varying forms.

It is therefore important that we understand as fully as possible the nature of these signals and the way in which we can best observe them. From this unified approach, all signals can be examined collectively, and their common properties exploited to assist us in establishing rules and techniques for their measurement. It is often possible to improve signal measurement but only at some cost, such as an increase in measurement time or a loss of resolution, thereby necessitating a selected trade-off. The following discussion will examine signals, consider problems inherent in observing signals, and consider some relatively simple techniques which enable better signal collection.

Signals and Noise

All signals observed in real life carry with them an undesirable hitchhiker called noise. For our purposes, this noise may be considered to be any portion of the observed signal which we do not want. In many cases a signal can carry a great deal of information, much of which is useful. Nonetheless, any portion of the signal we wish to ignore or would rather not observe is noise, regardless of how useful it might be under different considerations. An example is the radiation emitted by a flame in flame spectrometry. Although the emission signal may contain information about a number of species present in a sample, if our only interest is in sodium, all other species can contribute to unwanted variations in the sodium emission and thereby constitute noise.

From this brief discussion, to most effectively observe any phenomenon of interest, it would be desirable to extract the signal from the noise. Although this is indeed important, it must be realized that noise can never be completely removed, because the devices used to reduce noise will themselves introduce a small but finite amount of noise onto the signal. Thus, it is necessary to employ a figure of merit, the signal-to-noise ratio (S/N), to define the relative freedom from noise in a system or to define the quality of a measurement.

Signal-to-Noise Ratio. Let us examine the signal-to-noise ratio more closely. To obtain the S/N ratio, we must first have ratioable quantities. Because a power ratio is generally most meaningful, we must therefore determine the power available in both the S and N waveforms. If the waveforms are dc, the power will be simply proportional to the respective dc value. If, however, the waveforms are ac, the power will be proportional to the root-mean-square (rms) value of the waveform. The rms value can be found from the following relationship:

$$A_{\text{rms}} = \left[\frac{\sum_i (\bar{A} - \Delta A_i)^2}{n}\right]^{1/2} \qquad (1)$$

where $\bar{A}$ is the mean value of the waveform, and ΔA_i is the deviation of the i^{th} point from the mean.

For a dc signal where the major noise is often a time variation of the signal, Equation 1 takes on added significance. In this case, the signal is the mean value ($\bar{A}$), and noise is just the excursion from this mean, so that Equation 1 will define the noise in the following way:

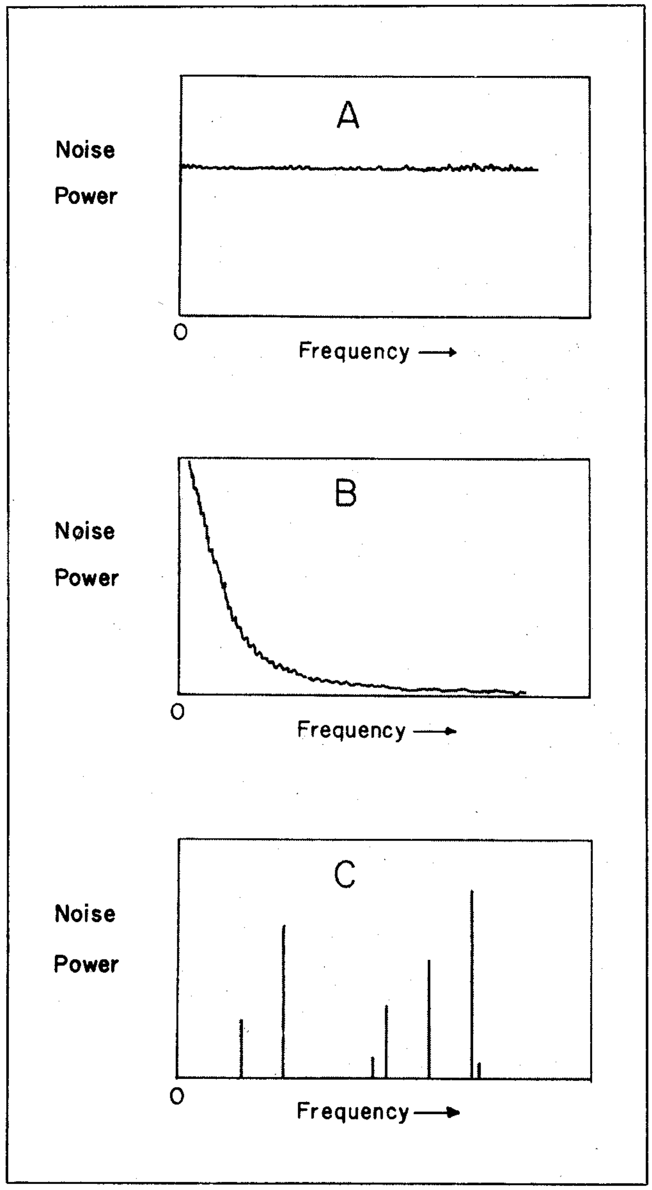

Figure 1. Noise-power spectra commonly found in chemical instrumentation

A. "White" noise
B. Flicker (1/f) noise
C. Interference noise

$$N = \left[\frac{\sum_i (S - \Delta S_i)^2}{i}\right]^{1/2} \quad (2)$$

Note that Equation 2 is the same as that used in computing the standard deviation (N) of a series of samples whose mean would be S. Therefore, N is equivalent to the standard deviation, and S is equivalent to the mean. Thus,

$$\frac{S}{N} = \frac{\text{mean}}{\text{standard deviation}} = \frac{1}{\text{relative standard deviation}} \quad (3)$$

This relationship emphasizes the importance of S/N in determining the precision of a measurement. Large S/N values obviously provide better precision (lower relative standard deviation). In general practice, the relative standard deviation is used to describe variations between discrete signal measurements made over a period of time; the signal-to-noise ratio is commonly used to express the variation in a single signal or waveform. However, Equation 3 shows the two to be related if proper evaluation is performed.

From this standpoint, S/N may be conveniently estimated in the common case of a dc signal trace on a recorder, oscilloscope, or other readout device. If the variation in the recorded signal is truly random, that is, if the noise is white, any excursion from the mean value (S) will be no greater than 2.5 times the standard deviation (N) within a 99% confidence limit (*1*). Because these excursions will be in both positive and negative directions, the standard deviation (N) can therefore be found by taking $^1/_5$ the peak-to-peak variations in the signal. From this, S/N can be easily calculated. In this estimation of S/N, it is important to recognize that a sufficiently long record of the signal must be obtained to provide a statistically reliable estimate of the peak-to-peak excursions (*1*). For example, the signal from an instrument having a time constant of 1 sec should be observed for about 2 min.

For ac signals, it is necessary to employ other methods to measure S/N. It is possible, of course, to convert the ac signal to dc by use of a conventional rectification technique and to measure S/N as described. However, during conversion some ac noise will also be rectified and can cause error in the measurement of S. A more accurate measurement of S/N for an ac signal can be made by employing a technique known as autocorrelation. This technique will be described in the second of the two articles in this series. When it is available, a special voltmeter designed to respond to the rms value of a signal can also be used to evaluate S/N; this method is subject to the same error as the ac to dc conversion process but generally involves a simpler measurement.

Noise-Power Spectra. Although signals can usually be characterized in terms of their waveforms (e.g., dc, sine wave), it is impossible to look at noise in the same way. Being inherently unpredictable, noise has no perceivably periodic character so that a noise waveform is almost meaningless. To characterize noise, it is more instructive to examine the frequency components in the noise waveform. This can be done with a noise-power spectrum, which is a plot of the noise power as a function of frequency.

For white noise, as its name implies, a flat power spectrum is found (Figure 1A). White noise is therefore a "hash" of all frequencies, the components of which are random in phase and amplitude. Examples of white noise are shot noise in photomultiplier tubes and Johnson noise caused by the random movement of electrons in a resistor. Johnson, Nyquist, or thermal noise, as it is variously termed, is particularly important in that it arises from the inherently random motion of charged species and therefore occurs in all real measurement situations. This fundamental thermodynamic noise has a flat power spectrum governed by the relationship:

$$E = (4\,RkT\Delta F)^{1/2} \quad (4)$$

E is the root-mean-square (rms) noise voltage contained in a bandwidth ΔF produced in a resistor of resistance R, k is the Boltzmann constant, and T is the resistor temperature in °K. The noise power, which is proportional to the rms voltage, increases as the square root of both the observed bandwidth and magnitude of the resistance. This characteristic is important in the selection of low-impedance filters to be discussed in the following section.

In contrast to white noise, another common form of noise, called "flicker noise," has a spectrum in which the power is approximately proportional to the reciprocal of the frequency. For this reason, it is often called $1/f$ (one-over-f) noise and has the power spectrum in Figure 1B. Flicker noise is common in most amplifiers and, in fact, in almost all instrumental systems, where it is usually termed "drift".

Another form taken by noise is "interference," often characterized by a line spectrum as shown in Figure 1C. Probably the most common form of interference noise is that arising from the 60-Hz power line. Therefore, a great deal of noise will usually be found at 60 Hz and all its harmonics (120 Hz, 180 Hz, etc.). However, other sources of interference noise, such as high-voltage spark sources, radio transmitters, or microwave devices, must be recognized. These high-frequency noise sources can be serious because interference noise is often added to a signal by transmission through the air, and noise-transmission efficiency increases with frequency.

Perhaps the most troublesome form noise can take is the form of an impulse. Impulse noise occurs erratically from such sources as the start-up of a large motor or a lightning flash, and because of its fast rise-and-fall times, has a broad frequency spectrum. The high peak power inherent in such impulse noise and its broad spectrum (similar to white noise) often make it difficult to eliminate.

Another type of noise which complicates certain measurements can be roughly termed distortion. Distortion is essentially the introduction of foreign frequency components onto the desired signal and can arise from a number of sources such as instrumental nonlinearities, transmission losses and reflections, and the interaction of the signal with various disturbing elements. Because of the unpredictability of some distortion sources, this form of noise can be quite vexing, especially in pulse-measurement systems where it is often encountered.

Most noise encountered in real situations will be a mixture of the noise types described. In general, the power spectrum will have a power proportional to $1/f^n$ where $0 < n < 1$. Additionally, the spectrum will often have line (interference) noise superimposed on it. With this in mind, we can investigate ways of eliminating or minimizing noise to enhance S/N.

Table I. Modulation Forms

Carrier-wave shape	Information-carrying property	Common name	Abbrev	Example
Dc	Amplitude	Analog		Meter reading
Sine wave	Amplitude	Amplitude modulation	SAM	Chopped-photometric signal
Sine wave	Frequency	Frequency modulation	SFM	Voltage-controlled oscillator Wavelength-modulation spectroscopy (2, 3)
Sine wave	Phase	Phase modulation	SPM	Photoelastic effect GC ultrasonic detector (4)
Pulse train	Amplitude	Pulse-amplitude modulation	PAM	Optical chopper Boxcar integrator
Pulse train	Repetition rate	Pulse-frequency modulation	PFM	Photon-counting (5) TV pictures from Mariner satellites
Pulse train	Pulse delay	Pulse-position modulation	PPM	Radar Remote Raman spectroscopy

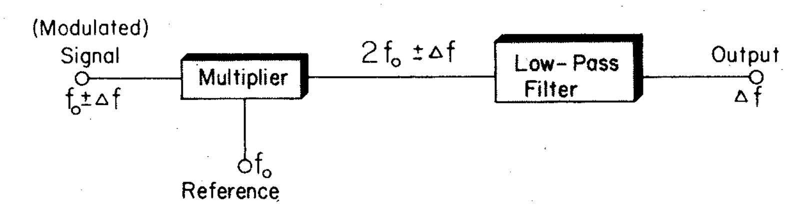

Figure 2. Instrumental components of lock-in amplifier

Signal/Noise Enhancement by Filtering

To extract a signal from any form of noise, some property of the signal must be employed to aid in its identification and separation from noise. Because signals possess an identifiable waveform and noise does not, all properties a waveform can possess—amplitude, frequency, and phase—can be used in this identification and extraction process. Perhaps the most commonly used property is frequency.

Dc Signal Filtering. For example, a dc signal can often be "smoothed" merely by using a long readout time-constant (damping) or by shunting a capacitor across an instrument's voltage output terminals. Note that this "smoothing" necessitates a trade-off between the response time of the device to a signal change and an improved signal-to-noise ratio. This procedure is nothing more than filtering, wherein all frequencies other than the frequency at which the signal is found are eliminated. In this case, the signal frequency is zero (dc) so that all higher frequencies are attenuated. This is analogous to the somewhat more versatile techniques of analog or digital integration of a signal over a fixed-time interval.

To carry information, the signal must have a finite bandwidth or frequency range (Δf), which means that the dc signal must be permitted to change in value to indicate a signal variation. Thus, the signal is not strictly dc (f_0) but has some additional, low-frequency (slowly varying) components (Δf) which are used to carry information. This necessary bandwidth makes perfect filtering impossible, since the filter must pass a finite band of frequencies to include all of the signal information. Because noise is found over a broad range of frequencies, some noise is almost certain to be included in this bandwidth and passed by the filter. For this reason, it is best to have the signal located in a frequency region which has little noise power.

Because most noise has a strong $1/f$ component, that is, because drift is generally significant, it is usually advantageous to locate the desired signal away from zero frequency (i.e., away from dc where $1/f$ noise is greatest). To do this, the signal information is impressed on a carrier wave at the desired frequency. This process is called modulation.

Modulation and Tuned Amplifiers. The process of modulation is employed both in everyday devices such as the radio and also in sophisticated electronic signal processing systems. All applications of this technique, regardless of their sophistication, depend upon the impression of one signal upon another. Usually the signal (the desired information) is impressed on another higher frequency wave (called the carrier) by modifying some property of the carrier in accordance with the characteristics of the signal. For example, in the common AM radio, a radio-frequency carrier wave is modulated in amplitude according to the frequency and amplitude of the audio signal which is impressed upon it. In a similar fashion, any property of a carrier wave can be modified in a way such that information from another signal is impressed on it.

Properties of a waveform are amplitude, frequency, phase (position with respect to a reference), and duty cycle (on-time compared to total duration). These properties can be varied on the following typical waveforms: sine wave, square wave, triangular wave, constant-level (dc) wave, and pulse train. Because the square and triangular waves are merely combinations of the fundamental and harmonics of a specific sine wave, these can be thought of as special cases of the sine wave and will therefore not be considered individually. By modulating various characteristics of the other waveforms, specific types of modulation can be effected. When this is done, the resulting modulated wave can be classified as shown in Table I.

From Table I, the carrier wave does not have to be at common electronically detectable frequencies but can range from dc to optical frequencies (10^{15} Hz). Nonetheless, when the signal information (perhaps chemical information about a sample) is forced upon the carrier, the process is called modulation.

Modulation is almost always used to improve the efficiency of transmission of a signal. In instrumental improvement in S/N, this is also the case; by impressing a signal on a carrier at the desired frequency, the information can be brought to a low-noise frequency region so that transmission efficiency and signal recognition are enhanced. In measurements on chemical systems, the modulation methods usually employed are sine-wave amplitude modulation and sine-wave frequency modulation, the former being the most common.

For example, in atomic absorption flame spectrometry if the intensity of a hollow cathode lamp is modulated by a light chopper at a frequency f_0, the lamp signal can be detected at f_0 to minimize the $1/f$ noise present in the system. This increases the effectiveness of filtering, since less noise will now be included in the necessary filter bandwidth. The filters which can be used to detect information at f_0 will necessarily be somewhat more complex than the simple capacitor shunt employed at dc. For example, specially designed "active filters" containing operational amplifiers provide narrow bandwidths having characteristics which can be mathematically defined. The Butterworth filter is such a device (*6*). Also, band-reject or "notch" filters can be used to reject specific frequencies (e.g., 60 Hz) where strong interference noise is known to exist.

Often, a filter is combined with an amplifier by use of appropriate feedback so that only a narrow band of frequencies is amplified. This system, called a tuned or frequency-selective amplifier, is often employed instead of a filter because of its ability to increase signal levels through amplification while actually reducing noise by appropriate filtering. Generally, this type of device provides an increase in S/N proportional to the reciprocal of the square root of the bandpass (the range of frequencies passed) of the amplifier. Therefore, the bandpass is made as narrow as is consistent with the necessary signal bandwidth discussed earlier. Unfortunately, this requires that the amplifier and the carrier frequency be stable so that the carrier and amplifier bands match exactly.

Practical experience has shown that bandwidths less than 1 Hz become troublesome in their frequency drift so that this provides a limit to the S/N enhancement attainable with a conventional tuned amplifier. Crystal filters provide one exception to this observation. The naturally stable resonance frequency of an oscillating crystal (e.g., quartz) can be used in a tuned system to furnish a stable but expensive narrow bandwidth amplifier. This unfortunately requires that the signal frequency remain stable at the crystal frequency. To solve the problem of obtaining a narrow bandwidth while maintaining a frequency match between the signal and filter bandpass, a device called a lock-in amplifier is often used.

Lock-In Amplifier

To eliminate the problem of frequency drift in the tuned amplifier and to further decrease the effective amplifier bandpass, a device called a lock-in or phase-sensitive amplifier was developed. This device discriminates against noise, not only on the basis of frequency as does the tuned amplifier, but also according to phase. Thus, the only signals (and noise) amplified by a lock-in amplifier are those having a specified frequency and phase with respect to a reference waveform.

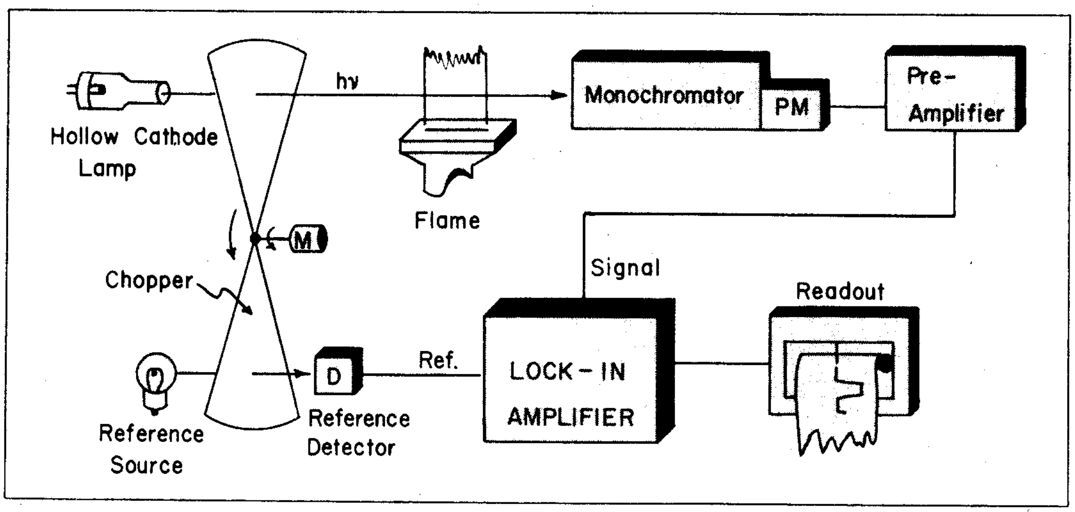

Figure 3. Schematic diagram of atomic absorption spectrophotometer employing lock-in amplification for signal-to-noise enhancement

Because the reference waveform can generate the signal (by modulation) or be obtained from the signal (during modulation), it can be phase-related to the signal. By suitable processing, only the signal and noise components having this phase relationship to the reference will be amplified, all others being rejected. This obviously provides even better noise rejection than the tuned amplifier and also, by using a reference waveform, eliminates the problem of frequency drift.

The components of a typical lock-in amplifier are shown in Figure 2. In this system the signal information (Δf) has been impressed on a carrier f_0 by a suitable modulation technique (e.g., optical chopper) and is phase-related to a reference waveform, also at f_0. When the signal (+ noise) is multiplied by the reference, a waveform at $2\ f_0$ carrying the signal information (Δf) results. Because Δf is generally low-frequency information, it can be efficiently extracted from the $2\ f_0$ carrier by a low-pass filter so that it alone appears at the output. The characteristics of this system are as follows:

The spectrum of signal frequencies (Δf) centered at f_0 are transformed to the same spectrum centered at dc.

If the signal and reference are of the same frequency, the output is dc, the dc value being largest when the signal and reference are in phase and smallest when they are 90° out of phase.

If the phase of the signal reverses, the sign of the dc output reverses.

Note that the lock-in amplifier performs essentially the reverse function of modulation. For this reason it is one member of a family of devices called demodulators. As an example of the use of a lock-in amplifier, consider an atomic absorption system such as shown in Figure 3.

In this system the reference is phase-locked to the chopper-modulated signal, which is a square wave of amplitude proportional to the hollow cathode emission intensity. Following this signal through the lock-in amplifier of Figure 2, we must first consider the multiplier output. In Figure 4 the multiplier output is shown for signals (or noise) which are in phase with or are 90° or 180° out of phase with respect to the reference wave. Peak-to-peak amplitudes of $2\ x$ and $2\ y$ are arbitrarily assigned to the signal and reference waveforms, respectively.

From Figure 4 the averaged values of the multiplier output are $+xy$ for the in-phase signal, $-xy$ for 180° phase difference, and 0 for 90° phase difference between the signal and reference. These are the values which would appear at the output of the low-pass filter, as predicted earlier. The $2\ f_0$ component discussed earlier is especially obvious in the 90° situation.

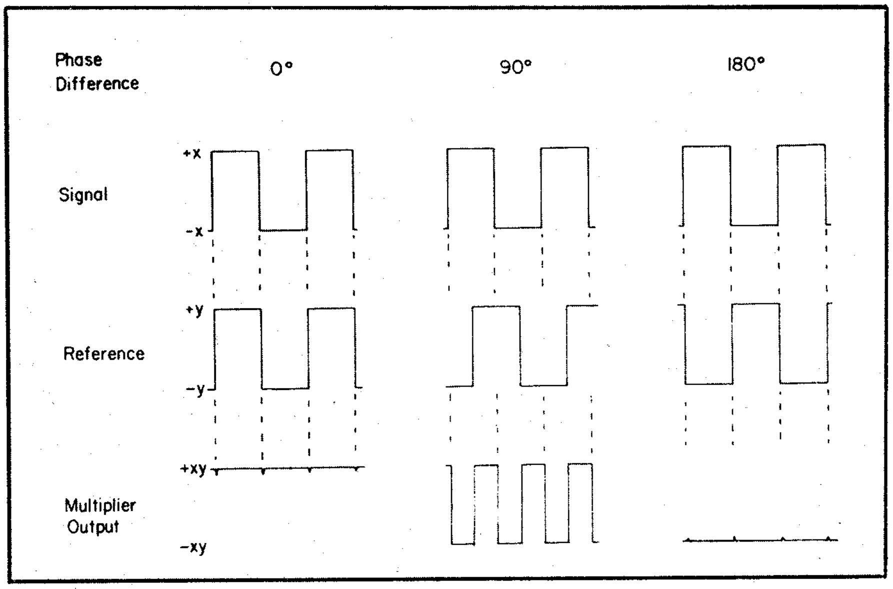

Figure 4. Waveforms illustrating multiplier output of lock-in amplifier for various phase differences between signal and reference inputs

As variations on this system, it is common to find both sine-wave signals and/or references, the considerations being the same as above. Note that any noise without a specified frequency and phase relationship to the signal will contribute little to the lock-in output. This provides an enhancement in S/N which, as with the tuned amplifier, will be approximately equal to the reciprocal of the square root of the amplifier bandpass. For the lock-in amplifier, the effective bandpass (Δf) can be approximated by

$$\Delta f = \frac{1}{4\ RC} \qquad (5)$$

where RC is the time constant of the low pass filter of Figure 2. This argues that the longest time constant should be selected which is consistent with the observation of signal variations (i.e., with the signal bandwidth). For properly designed lock-in amplifiers, it is therefore quite possible to obtain an effective bandpass of less than 0.01 Hz, much smaller than the value of 1 Hz obtainable with a tuned amplifier.

Obviously, neither lock-in nor tuned amplification should be utilized at a frequency which contains a strong interference noise component. If the signal naturally lies at such a frequency, appropriate modulation should be employed to move the signal information to another frequency region. Thus, one should never, for example, chop a light beam at 60 Hz or any of its harmonics or subharmonics. In a situation where considerable impulse noise is present, tuned or lock-in amplification may be ineffective in improving the signal-to-noise ratio. In these cases, the sporadic appearance and broad spectrum of impulse noise will cause it to be partially passed through these devices rather than eliminated.

Other names which have been used for the lock-in amplifier are synchronous detector (because of the necessary synchronization of signal and reference), phase-lock amplifier, coherent detector, and heterodyne detector. A limitation of these systems, as is apparent from the above discussion, is that the signal to be detected must be periodic or modulated in such a way as to be made periodic. When this is difficult or impossible, other signal enhancement techniques must be used. These techniques will be considered in the second article of this series.

In the second article, simple filtering and lock-in amplification will be compared to the more versatile techniques of signal averaging, boxcar integration, and correlation. With this information it should be possible to select the most appropriate method of signal-to-noise enhancement for any given measurement situation.

References

(1) V. D. Landon, *Proc. IRE*, 50 (February 1941).
(2) V. Svoboda, *Anal. Chem.*, **40**, 1384 (1968).
(3) W. Snelleman, *Spectrochim. Acta*, **23B**, 403 (1968).
(4) F. W. Noble, K. Able, and P. W. Cook, *Anal. Chem.*, **36**, 1421 (1964).
(5) M. L. Franklin, G. Horlick, and H. V. Malmstadt, *ibid.*, **41**, 2 (1969).
(6) Philbrick Applications Manual, Nimrod Press, Boston, MA, 1968. Available from Philbrick/Nexus Research, Dedham, MA 02026.

The author gratefully acknowledges the support of National Science Foundation grant NSF GP 24531.

Signal-to-Noise Enhancement Through Instrumental Techniques

Part II. Signal Averaging, Boxcar Integration, and Correlation Techniques

Techniques are described which can provide signal-to-noise enhancement for nonperiodic or irregular waveforms or for signals which have no synchronizing or reference wave

G. M. HIEFTJE
Department of Chemistry
Indiana University
Bloomington, IN 47401

In a previous article (*1*), it was indicated that noise is found in every experimental signal which can be measured. However, because signals have predictable behavior (i.e., because they are coherent), they can be partially extracted from the noise by use of suitable techniques, provided the necessary increase in measurement time can be tolerated. One such technique, filtering, is most effective if the signal information is located at a frequency or group of frequencies in which little noise is found. Modulation can be used to place the signal information at the desired frequency. However, even when the signal is located in a low-noise frequency region, some noise will always be passed by any filter because of the necessary bandwidth of the signal and of the difficulty involved in matching and maintaining a match between the signal frequencies and the filter (or tuned amplifier) bandpass. A lock-in amplifier was useful in solving these problems but required periodic signals or those which can be made periodic.

In this article, techniques will be described which can provide signal-to-noise enhancement for nonperiodic or irregular waveforms or for signals which have no synchronizing or reference wave. These techniques, signal averaging, boxcar integration, and correlation, will be discussed independently.

Signal Averaging—S/N Enhancement in Time Domain

If a signal is not periodic or repetitive, but is *repeatable*, it can be extracted from noise with the technique of ensemble averaging or signal averaging as it is often termed. This technique uses an averaging or integration procedure similar to that employed when multiple recorder or oscilloscope traces are superimposed to obtain an average or effective value.

Basically, signal averaging involves the instrumental superposition of a number of signal traces by sampling each signal record in the same way and storing the samples in either a digital or analog register. Because the records are each sampled in the same way and at the same corresponding times, the signals add coherently in register while the noise, being random, averages to zero. Averaging the ensemble of records thus provides increased S/N.

More quantitatively, the signal will add in register directly as the number of records sampled (n), whereas the noise will only add as the square root of this number ($\sqrt{n}$), assuming a $1/f^n$ noise spectrum (*1, 2*). Therefore, the S/N will increase as $n/\sqrt{n} = \sqrt{n}$.

To enjoy this advantage, the sampling of the signal records must meet certain criteria (*3*). For example, to extract all the information from the signal, it must be sampled at a frequency at least twice as great as its highest frequency component. However, the sampling frequency should not be significantly greater than twice the highest signal frequency. If this occurs, no additional information will be obtained, but more noise will be included than is desirable or necessary merely because the bandwidth is increased by faster sampling. Also, to prevent "aliasing" high-frequency noise down into the sampling frequency region, a high-frequency cut-off filter should be used at the signal input (*4*). Finally, if any interference (e.g., narrow band or line) noise (*1*) is expected to be present in the signal, it is important that the signal records not be scanned or sampled at a rate which is either a multiple or submultiple of the interference frequency. If this latter condition is not fulfilled, no improvement in S/N will occur for the interference noise.

In signal averaging, it is important to be able to properly sample and add the sampled records reproducibly. This requires a synchronizing signal or pulse which can serve to actuate the sampling system. Generally, this "sync pulse" can be derived from the signal or can be used to initiate the signal record. The synchronizing signal must obviously be reliable in predicting the

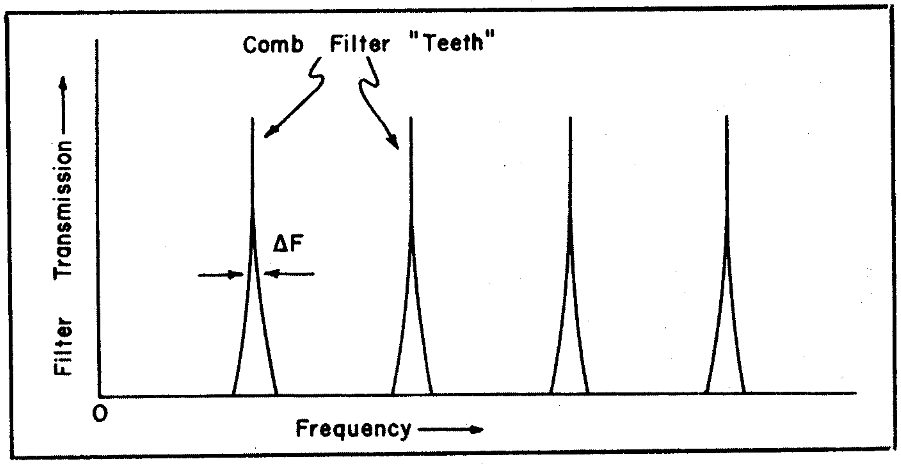

Figure 1. Frequency domain representation of signal averager characteristics. Teeth of comb filter located at frequency components of signal being measured

$$\Delta F = \text{tooth width} = \frac{0.886}{n\tau} \text{ (Hz)} \quad \text{(See text)}$$

Table I. List of Some Commercially Available Hardware Signal Averaging Instruments

Name	*Manufacturer*
Waveform eductor	Princeton Applied Research Corp.
Enhancetron	Nuclear Data, Inc.
Signal averager	Nicolet Instrument Co., Inc.
Time averaging computer	Varian Associates, Inc.
Spectrum accumulator	JEOLCO, Inc.
LAB-8	Digital Equipment Corp.

start of the signal for the records to add exactly and for S/N enhancement to be realized.

Because the signal averaging technique operates in the time domain, it is difficult to compare with the lock-in and frequency-selective amplifiers (*1*) for its capability to increase S/N. For a comparison, it is necessary to transform [via a Fourier transform (*5*)] the effect of the ensemble averager on a selected signal. When this is done, the averager acts like a comb filter (such as that portrayed in Figure 1) whose teeth are centered at the frequency components of the signal and whose teeth have a bandpass (Δf) of

$$\Delta f = \frac{0.886}{n\tau} \quad (1)$$

Here, τ is the time taken for the sampling of one record, and n is the number of scans. This type of filter, of course, passes the signal efficiently while removing the noise. For example, with a sampling time for each record of 20 msec and 10^6 scans, a tooth bandwidth of 5×10^{-5} Hz is obtainable—far better than would be possible with a lock-in amplifier (*1*).

In addition, the signal averager can provide signal-to-noise enhancement for nonrepetitive or nonperiodic signals, a feat which is impossible for tuned or lock-in amplifiers. Because the averager passes *all* the frequency components of the signal *in the proper phase*, the original signal is extracted from noise with its original waveform. This is important in many applications, such as nuclear magnetic resonance (nmr) spectroscopy (*2*), where the waveform of a signal conveys more information than its frequency components alone.

In nmr, as in a number of applications of signal averaging, it is interesting to note that the signal averager is often preceded by a lock-in amplifier. This combination is used because in situations where $1/f$ noise (i.e., drift) is significant, it is better to sweep a number of signals rapidly and average them, than to slowly sweep a single signal over the same time period by use of a correspondingly longer filtering time constant on the lock-in amplifier (*2*). Also, impulse noise will not be as troublesome to signal averaging as it was for lock-in or tuned amplification because for each impulse, only a single deviant point will occur in the final averaged signal. Generally, for a signal which is a smoothly varying time function, the deviant points can be easily discarded if simple but carefully chosen testing criteria are used.

A number of hard-wired signal averagers which are commercially available under various names are listed in Table I. However, note that unlike tuned or lock-in amplification, signal averaging can be conveniently performed under software control by a small laboratory computer. The increasing availability of these computers can be expected to render signal averaging the technique of choice for many routine signal enhancement chores.

From a practical standpoint, hardware ensemble averagers can be used for several purposes other than signal-to-noise enhancement. They can be employed to compute the average value or a more representative value of a signal. Also, because the storage register of the averager can be read at a faster or slower rate than that at which the signal was sampled, the device can be used to slow down or speed up the transmission of data.

Boxcar Integration—Poor Man's Signal Averager

A boxcar integrator is somewhat similar in design to both the lock-in amplifier and the signal averager and performs similar functions, depending on its application. The generalized schematic of a boxcar integrator, shown in Figure 2, indicates that the device is essentially a single-channel signal averager with a sampling gate which can be opened for a selected time during the passage of the signal to be measured. At a fixed delay the gate is synchronized to the signal through the reference channel, so that the same portion of the signal is sampled for each signal passage. The gated samples taken from each passage are then averaged by the low-pass (RC) filter to provide signal-to-noise enhancement *for the portion of the signal which is sampled.*

Clearly, operation of the boxcar integrator in this fashion provides no information about the signal except its average amplitude at a specified time. For complex signals, this is of limited utility. However, for a signal consisting of a square wave or a train of pulses, the average amplitude of the square wave or pulse can be of considerable importance and in many cases is the only information sought.

Several examples of pulsed signal detection important to chemical analysis can be cited (*6, 7*), and because of the high signal-to-noise enhancement attainable with pulsed or low duty-cycle signals (*8*), these applications can be expected to rapidly increase in number. Also, the attractiveness of pulsed lasers as sources for many spectrochemical techniques can be expected to make a gated detector, such as the boxcar integrator, increasingly useful as more and better laser sources are developed.

Operation of the boxcar integrator as a detector for pulsed signals requires only the optimization of the gate width and delay for the specific signal sought. As with the signal averager, the pulsed signal need not be periodic but must be repeatable and have a synchronizing reference which is time-locked to the signal. Any lack of synchronization between the signal and reference waveforms will cause "jitter" in the gate opening and resultant error or noise at the output. For this reason the gate width and delay in a commercial boxcar detector must be extremely precise and jitter-free, usually to within a few nanoseconds.

Compared to the measurement of a single pulse, the boxcar detector provides a signal-to-noise enhancement equal to the square root of the number of pulses integrated. However, if the boxcar integrator is compared to the commonly measured "average value" of a pulsed signal, an even greater S/N enhancement is realized. The degree of enhancement depends quantitatively, among other things, on the duty cycle of the signal (on-time compared to total duration) but can be qualitatively evaluated by considering the amount of noise which is detected in *each* case.

In measuring the average value, the detector must be continuously turned on and will therefore detect noise for the full duration of the signal. By contrast, the boxcar detector is "on" only during the gate pulse so that noise is detected only for that short time. Because both measurements include the total signal pulse, the reduction in detected noise in the boxcar integrator produces a dramatic signal-to-noise improvement over that of the average value measurement.

The similarity of the boxcar integrator of Figure 2 to a lock-in amplifier (*1*) is immediately apparent. In fact, the boxcar detector can serve as a lock-in amplifier simply by setting the gate width to pass the signal at half-cycle intervals. The added ability to gate low duty-cycle signals makes the boxcar integrator more versatile, however, and more sensitive in the detection of these signals. Like a lock-in amplifier, the boxcar integrator is especially susceptible to impulse noise and to interference noise which is at a multiple or submultiple of the sampling frequency. The sampling frequency should therefore be carefully chosen.

A second mode of operation of the boxcar integrator enables the enhancement of signals having a complex waveform. In this mode the boxcar integrator can again be thought of as a single-channel signal averager. Instead of being locked to a specific time channel, however, the boxcar gate is moved from channel to channel so that eventually the entire complex waveform is sampled. This can be done by automatically stepping or scanning the variable gate delay of Figure 2 so that the signal segments are scanned in sequence from zero delay to a delay which is greater than the signal duration.

In the scanning mode of operation the boxcar integrator provides signal-to-noise enhancement by sampling the signal a number of times at each value of time delay. In a stepped delay system this is accomplished by incrementing the delay after a desired number of signals have been sampled. A continuously scanned delay can be similarly employed, provided that the delay scan rate is significantly slower than the repetition rate of the signal. For irregular or nonperiodic signals the stepped delay is generally preferable, to ensure that the same number of signal waveforms are read at each delay position.

Compared to a signal averager, the boxcar integrator is a rather inefficient instrument for enhancing complex signals. Because the signal is sampled only briefly for each of its appearances, a great deal of information is thrown away. In fact, for a signal averager having *n* channels, signal enhancement will proceed *n* times faster than for a boxcar integrator with similar gate characteristics. From this, seemingly, instrumental simplicity is the only advantage of the boxcar integrator. However, because the boxcar integrator has only a single channel (gate), the gate can be made far more sophisticated than would be practical for the large number of channels in the signal averager. Generally, the gate in a boxcar integrator can operate much faster and more precisely than those in the signal averager, so that the boxcar detector is most useful in applications involving fast signals. Also, because signal improvement proceeds more slowly with the boxcar system, it is most advantageously applied to signals having a high repetition rate, usually above 100 Hz where suitable signal averagers would beome prohibitively expensive.

Like the signal averager, a boxcar integrator can be used for a number of purposes other than signal-to-noise enhancement. For example, fast signals with a high repetition rate can be sampled and scanned to provide readout on a much slower time scale. This is the principle on which the sampling oscilloscope operates and provides the basis for an even slower readout of fast signals on a device such as a servo recorder. Time resolved spectroscopy

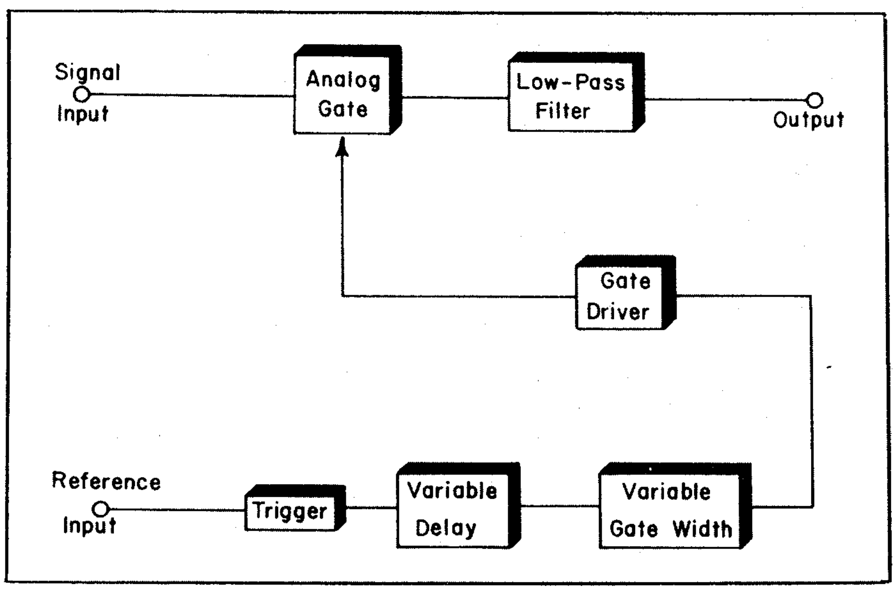

Figure 2. Schematic block diagram of boxcar integrator. Variable delay can be operated in either fixed, stepped, or scan mode

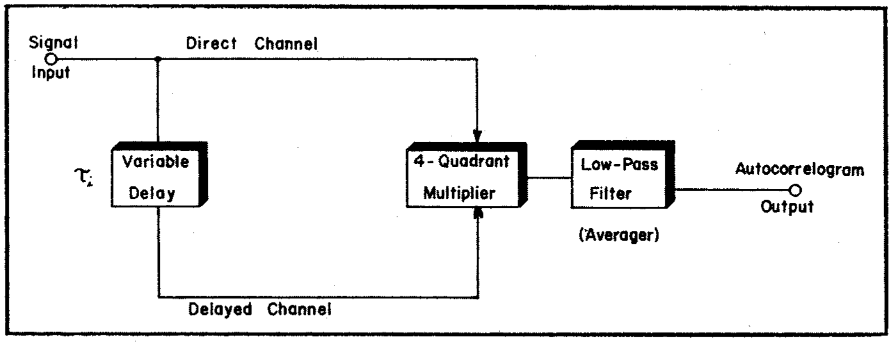

Figure 3. Functional diagram of autocorrelation computer. Conversion to cross-correlator possible by changing variable delay input to reference signal. 4-Quadrant multiplier accepts and outputs both positive and negative waveforms

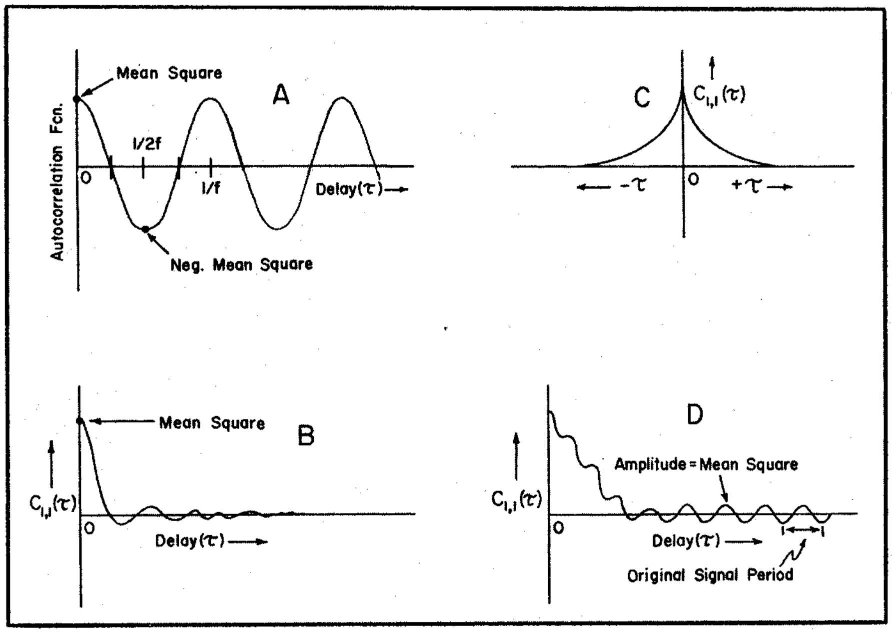

Figure 4. Autocorrelograms of various input signals. $C_{1,1}(\tau)$ = autocorrelation function

A. Autocorrelogram of sine-wave signal
B. Autocorrelogram of random signal (noise)
C. Autocorrelogram of band-limited noise showing both positive and negative delay (τ)
D. Autocorrelogram of extremely noisy sine-wave signal (S/N < 1)

can also be performed with a stroboscopic sampling unit similar in design to a boxcar integrator (*9*).

Correlation Techniques

The correlation techniques for S/N enhancement are an outgrowth from information theory and depend on the relationship between a signal and a delayed version of itself (autocorrelation) or of another signal (cross-correlation). Mathematically, if $V_1(t)$ and $V_2(t)$ are two functions (signals), the cross-correlation function $[C_{1,2}(\tau)]$ is

$$C_{1,2}(\tau) = \lim_{\tau \to \infty} \frac{1}{2T}\int_{-T}^{+T} V_1(t)V_2(t-\tau)dt \quad (2)$$

τ = delay

t = time

For the special case of $V_1(t) = V_2(t)$, a similar expression gives the autocorrelation function $[C_{1,1}(\tau)]$

$$C_{1,1}(\tau) = \lim_{\tau \to \infty} \frac{1}{2T}\int_{-T}^{+T} V_1(t)V_1(t-\tau)dt \quad (3)$$

From these equations, the correlation functions indicate whether coherence exists between two signals or within a signal with respect to a natural or artificial variable delay (τ). As an example of this, a strong cross-correlation exists between the New York blackout a few years ago and the later birth rate, with a natural delay time (τ) of nine months.

To implement and further elucidate the autocorrelation function, the hardware or software model of Figure 3 can be used. Figures 3 and 2 of Reference *1* show that the correlator is similar to a lock-in amplifier which incorporates a variable delay.

In the device of Figure 3 the signal is multiplied by a version of itself delayed by a fixed amount (τ), and the mean of this product is taken as one point of the correlation function $[C_{1,1}(\tau_i)]$. By taking more points in this fashion and plotting them as a function of the delay τ_i, the complete autocorrelation function or autocorrelogram can be traced out. As a simple example of this, refer to Figure 4A and assume a sine-wave input to the instrument in Figure 3.

When the delay (τ) is zero, the sine wave of frequency f_1 is fed directly to the multiplier along both channels so that the multiplier output will be a sine^2 wave. The averager will then output the mean of this wave which will be the mean square value of the input sine wave. Next, when τ is increased by a time equal to a quarter period of the sine wave, the sine wave will be 90° out of phase with its delayed version. From Figure 3 and the discussion of the phase-lock amplifier (*1*), the output of the averager will be zero under these conditions.

For the case of a delay $\tau = 1/2\ f_1$, e.g., ½ cycle or 180°, the sine waves will be exactly out of phase at the multiplier, so that a negative sine^2 wave will result. The averager will then output a value equal to the negative mean square of the input sine wave. From these points and Figure 4A, the correlogram in this case will be a cosine wave of frequency $1/\tau = f_1$

which is the same as the input frequency and of amplitude equal to the mean square of the input sine wave. Therefore, all information about the input signal is maintained except the phase information, since here a cosine wave was obtained whereas a sine-wave input was used. Mathematically, if the input waveform was

$$f(t) = A \cos (\omega t + \theta) \qquad (4)$$

the output would be

$$C_{1,1}(\tau) = \frac{A^2}{2} \cos \omega t \qquad (5)$$

For a more complex input waveform, the correlogram has quite a different appearance. If a broad band of frequencies (e.g., noise) is introduced into the autocorrelator (Figure 3), all frequencies will still be in phase when no delay is used, so that the averager output will be equal to the mean square, or $(\text{rms})^2$ of the input at $\tau = 0$.

As the delay is increased, however, the different frequencies present at the input will rapidly go out of phase with each other and average out to zero (Figure 4B).

The less coherence which exists in the correlator input, of course, the less will be the correlation between the input signal and its delayed replica so that the autocorrelation function will damp to zero more rapidly. In fact, for completely random, white noise, no time correlation exists. Therefore, for perfectly white noise, the autocorrelation function is a delta function with a value of the $(\text{rms})^2$ at $\tau = 0$ and with a value everywhere else equal to zero. For real, band-limited noise, such as found at an amplifier output, however, a function similar to Figure 4C is found. Referring to Figure 4C, the autocorrelogram is an even function, i.e., is symmetrical about $\tau = 0$. This can be appreciated by considering the negative delay ($\tau < 0$) to be effected by inserting the variable delay of Figure 3 into the direct channel rather than the delayed channel. Obviously, the mirror image correlogram will be produced.

If to a random waveform is added a coherent signal, e.g., a sine wave, the (one-sided) autocorrelation function will appear, as in Figure 4D where the random waveform (noise) contribution rapidly damps to zero, as in Figure 4C. The coherent signal, however, continues to contribute at much greater delays. This enables the signal (the sine wave) to be extracted *very* efficiently from the noise (random wave). This technique is especially useful when no means is available for synchronizing a sampling system to the signal, which is necessary with the tuned and phase-lock amplifier, with the signal averager, and with the boxcar integrator. Also, autocorrelation techniques are useful in situations where $S/N \ll 1$, i.e., when the signal is often unrecognizable or impossible to detect apart from the noise.

Signal-to-noise ratios can be conveniently measured from an autocorrelogram such as Figure 4D. The point at $\tau = 0$ is the $(\text{rms})^2$ value of $S + N$, but the peak value of the ac signal after the noise contribution has damped to zero is the $(\text{rms})^2$ value of the signal alone. From these quantities, the ratio of the rms values of S and N can be easily computed.

Cross-correlation is similar to autocorrelation except that the delayed signal arises from a second source, probably from the modulating function of the signal or as in phase-lock amplification, from a reference coupled to the signal modulator. In this case, the frequency components common to both waveforms will appear in the correlogram. For example, if a sine wave and square wave of the same frequency are cross-correlated, the correlogram would be a sine wave of that frequency and of amplitude equal to the average value of the product of the sine and square waves. This is because the sine wave fundamental of the square wave is the only frequency component common to both waveforms.

Although cross-correlation requires a reference as does phase-lock amplification, note that the reference and signal waveforms need not be sine or square waves but can take any form. Therefore, a signal can be modulated with any desired waveform, a capability which has advantage in certain applications (*10*). Like signal averaging, the correlation techniques are effective in reducing impulse noise by a simple elimination of obviously deviant points in the correlogram; interference noise effects can also be minimized with correlation.

Because the correlation techniques produce a phase-related representation of the common frequencies between two signals or within a signal, the correlogram may appear quite different from the original signal. As an example, the autocorrelation function of a square wave is a triangular wave of the same fundamental frequency. The triangular and square waves, of course, contain the same frequency components, although in the triangular wave the components are phase-related (i.e., are all cosine waves) and begin at $\tau = 0$. This indicates once more how the amplitude and frequency information is maintained in computing the correlation functions, although the phase information is lost.

Although the correlation techniques have not yet been extensively employed in chemical analysis, it is likely that their use will increase. Previously, it was necessary to calculate point-by-point correlograms or to use expensive hardware correlation computers. Now, with the increasing introduction of small digital computers into the chemical laboratory, it will become increasingly convenient to compute auto or cross-correlation functions with inexpensive software-based systems.

Summary

In this and in the previous article on instrumental methods of signal-to-noise enhancement (*1*), a number of old and new concepts have been introduced in a hopefully coherent form. No attempt has been made to maintain mathematical rigor in the treatment, and no technique has been treated exhaustively. Although several examples of the use of each technique were cited, these were not considered in detail and, of course, many additional applications could be listed. Rather, it is hoped that the reader with little background in communication theory has gained some familiarity with the subject and its instrumental implementation. Readers with a more extensive background have hopefully had their memories jogged and also benefited by a somewhat different treatment of the subject.

References

(1) G. M. Hieftje, *Anal. Chem.,* **44** (6) 81 (1972).
(2) R. R. Ernst, *Rev. Sci. Instrum.,* **36,** 1689 (1965).
(3) "Threshold Signals," J. L. Lawson, G. E. Uhlenbeck, Eds., Dover Publ., New York, NY, 1965, Chap. 3.
(4) G. Horlick, H. V. Malmstadt, *Anal. Chem.,* **42,** 1361 (1970).
(5) R. Bracewell, "The Fourier Transform and Its Application," McGraw-Hill, New York, NY, 1965, Chap. 2.
(6) G. M. Hieftje and H. V. Malmstadt, *Anal. Chem.,* **41,** 1735 (1969).
(7) L. M. Fraser and J. D. Winefordner, *ibid.,* **43,** 1693 (1971).
(8) V. G. Mossotti, Abstracts, 161st National Meeting of the American Chemical Society, Los Angeles, CA, April 1971, No. Anal. 27.
(9) M. L. Bhaumik, G. L. Clark, J. Snell, and L. Ferder, *Rev. Sci. Instrum.,* **36,** 37 (1965).
(10) R. R. Ernst, *J. Mag. Resonance,* **3,** 10 (1970).

The author gratefully acknowledges the support of National Science Foundation grant NSF GP 24531.