THE UNIVERSITY OF MICHIGAN
COMMUNICATION SCIENCES LABORATORY
ANN ARBOR

SOME THEORETICAL AND EXPERIMENTAL ASPECTS OF ANALYSIS WITH THE SOUND SPECTROGRAPH

Hugh K. Dunn
Tiong Suy Yu
William D. Chapman

Distribution of This Document is Unlimited

Nonr 1224(22)
NR 049-122
Information Systems Branch
Office of Naval Research

August 1966

Communication Sciences Laboratory
The University of Michigan
August, 1966

Precis

Title: Some Theoretical and Experimental Aspects of Analysis With the Sound Spectrograph, Hugh K. Dunn, Tiong Suy Yu, and William D. Chapman, Communication Sciences Laboratory Report Number 7, August, 1966; Contract Nonr 1224(22), NR 049-122.

Background: The Communication Sciences Laboratory, under the sponsorship of the Information Systems Branch of the Office of Naval Research, has been conducting studies in speech analysis and synthesis. Techniques of spectrographic analysis have long been of basic importance in such research.

Condensed Report Contents: Before the introduction of the sound spectrograph in 1946, analysis of complex audio phenomena, such as speech, into time, frequency, and amplitude coordinates depended for the most part on a very time-consuming Fourier analysis of an oscillographic recording of the event. The techniques introduced with the sound spectrograph permitted the speeding-up of the analysis procedure by a large factor. These techniques and their underlying principles are discussed in the first few chapters of this report. A sound spectrograph developed at the Communication Sciences Laboratory of the University of Michigan is then described. This instrument includes some novel features, such as the automatic marking on the spectrogram of time, frequency, and amplitude scales.

For Further Information: The complete report is available in the major Navy Technical Libraries and can be obtained from the Defense Documentation Center. A limited number of copies is also available from the Director (G. E. Peterson) of the Speech Communications Research Laboratory, 115 West Figueroa Street, Santa Barbara, California 93104.

Acknowledgment

The sound spectrograph has become a standard tool in laboratories conducting technical research on speech. The versatility of the instrument in performing different types of analyses has provided considerable convenience in speech research. The spectrograph presents basic data about the acoustical characteristics of speech in a visual form that can be studied in detail. When suitable precautions are taken, relatively accurate measurements can be made from sound spectrograms.

In view of the extensive research information which has been obtained with the sound spectrograph, it is somewhat surprising that it has not been applied more widely in related fields. For example, those engaged in the description of the phonology of languages often rely on perceptual judgments in cases where the sound spectrograph could provide definitive answers. The sound spectrograph has considerable potential in the examination and evaluation of defective speech, but unfortunately it is used very little in clinical procedures. It may be hoped

that this situation will not endure indefinitely, but that eventually spectrographic and other instrumental techniques of speech analysis will be used more widely in the study of phonology and speech disorders.

Digital computers have clearly introduced many advances and new methods in the analysis of speech. Digital computer processing, however, is generally directed toward the data reduction of the speech wave and the extraction of particular, presumably significant, parameters. We believe that there will long be a need for the more detailed analyses which are achieved with the sound spectrograph and for the display of speech data in the formats provided by this type of instrument.

It is toward the continued and extended use of the sound spectrograph in both basic and applied work on speech that this report has been prepared. The principles of the analysis of speech with the sound spectrograph have been discussed briefly in a variety of sources, but generally not in detail. We believe that those who enter the field of speech research need some convenient reference and illustration of the processes of spectrographic analysis. One method of analysis is often more suitable than another in work on a particular problem, and it is hoped that the following pages will help indicate the potential of the various types of analyses which can be made with the machine. It is our impression that there is little that is new in this writing, but we hope that presenting the information about spectrographic principles in a single source will be of value to those interested in speech research.

The fact that this report bears the number 7 in the Communication Sciences Laboratory series is some evidence of the fact that we have planned to issue a report on spectrographic analysis for some time. The first six reports in the series were issued while the laboratory was entitled "Speech Research Laboratory", and with the completion of this present report, all remaining reports through number 15 have been issued by the Communication Sciences Laboratory.

In the latter pages of this report, analyses obtained with the Communication Sciences Laboratory sound spectrograph are discussed. The initial development of this instrument was begun at a time when there was a need in our laboratory for a high quality, wide range, and flexible sound spectrograph

which could be used in many different types of technical investigations on speech. The initial work on the instrument was supported by a grant from the National Science Foundation, G-3293. Later the work was continued under NSF Grant G-14482, entitled "Research Instrumentation for a Sound Spectrograph". These grants led to a variety of research studies in the laboratory which have been described elsewhere in the literature. More recently the work on the sound spectrograph and also the preparation of the more theoretical parts of this present report have been supported by the Information Systems Branch of the Office of Naval Research.

We are greatly indebted to Dr. George A. Hellwarth, formerly of the Communication Sciences Laboratory, for numerous and valuable suggestions about the circuits employed in the sound spectrograph which is discussed in the later chapters of this report. Dr. Hellwarth is currently a member of the Speech Analysis Group of the International Business Machines Corporation at Raleigh, North Carolina. Special thanks are also due Mr. Ralph H. Fertig who prepared the figures for the report and to Mrs. Gloria A. Mannlein who typed the manuscript.

Gordon E. Peterson, Director
Communication Sciences Laboratory

Table of Contents

		Page
1.	Introduction	1
2.	The Fourier Transformation and the Frequency Spectrum	13
	Fourier analysis of a periodic pulse	18
	Fourier transformation of a unit rectangular pulse	20
	Spectrum of a delta function	21
	Parseval's theorem	23
3.	Modulation and Detection Theory	26
	Modulation	26
	Single-sideband operation and balanced modulation	34
	Detection	36
4.	Signal Filtering	41
	Effect of a linear network on an informational signal	41
	Response from an ideal linear network	43
	Distortionless network	45
	Signal distortion	47
	Linear phase filter	47
5.	The Single-Filter Spectrograph	50
	Basic spectrographic operations	50
	Recording and reproduction	53
	The modulation and filtering processes	55

Table of Contents (continued)

		Page
5.	The Single-Filter Spectrograph (continued)	
	The three-dimensional spectrogram	67
	The amplitude sectioner	70
	Continuous amplitude display	84
	a. Direct oscillogram	90
	b. Low-pass filtering (averaging detector)	92
	c. Envelope detector	94
	d. Logarithmic circuit	96
	Equivalent low-pass filtering	101
	Remarks	113
6.	The CSL Sound Spectrograph	114
	Spectrograms	115
	a. Wide-band and narrow-band spectrograms	115
	b. Amplitude sections	117
	c. Continuous amplitude displays	119
	Operation of CSL sound spectrograph	121
	a. Recording and reproduction	125
	b. Analyzing circuits	129
	c. Pulsed switching	137
	d. Marking	142
7.	Some Spectrographic Reproduction Experiments	144
	Sound spectrogram	144
	Amplitude section	160

Table of Contents (continued)

		Page
7.	Some Spectrographic Reproduction Experiments (continued)	
	Continuous amplitude display	163
	Switching operation of the dual channels	180
	Automatic scaling	187
	a. Frequency scale	187
	b. Time scale	187
	c. Amplitude scale	189
8.	Further Spectrograph Design Considerations	192
Appendix		196
References		204

1. Introduction

An important new device for the acoustical analysis of speech was first described by engineers of the Bell Telephone Laboratories in 1946 (13). As the name "Sound Spectrograph" implies, the machine analyzes sound into its component frequencies, as the optical spectrograph does for light. To be even more descriptive, however, the name might be expanded to "Sound Spectro-Chonograph"; for one of the principal features of the instrument is that it shows, not just a fixed signal "spectrum", but also the changes in the spectrum with time.

In the original form, a 2.2-second sample of speech or other signal in the audio frequency range is recorded on a ring of magnetic tape. This recorded signal is then reproduced repeatedly and the signal is modulated (or "heterodyned") to higher frequencies by means of an oscillator. The oscillator frequency is changed slowly, so that the lower sideband of the modulated signal is swept across the pass-band of a fixed filter. The effect is as though successive reproductions of the signal were presented to a succession of bandpass filters of fixed width but with different center frequencies. The intensity of the signal passed by the filter is recorded as the darkness of a line marked on paper. This paper is mounted on a drum which rotates in synchronism with the tape, and the position of the marked line is shifted slightly for each successive reproduction. In about three minutes a complete analysis covering the range from 100 to 3500 cps can be made, in which the darkness of the record indicates the intensity at each location in time and frequency of the 2.2-second sound sample. For different purposes, different filter bandwidths may be used.*

In most analyses made with the sound spectrograph, the band of analysis is shifted only slightly in frequency for successive analyses. The patterns are made with many repetitions of the input signal, and in effect are constructed with highly overlapping filters. This provides a display of the details of the frequency composition which cannot easily be achieved otherwise.

Although the procedure described above requires some 80 times real time for a very good analysis of the sound sample, this is still a great advance in speed over previous methods of analysis. For comparable quality of analysis it had formerly been necessary to

---

*A commercial version of the sound spectrograph has now been produced for several years as the "Sona-Graph", manufactured by the Kay Electric Company, Pine Brook, New Jersey.

apply Fourier methods to each fundamental period of the sound, as recorded in an oscillogram. An example of such work on a speech sample about 1.5 seconds long was presented by Steinberg (25). Analyses of 58 different periods are shown in his report, each requiring between two and three hours to make. Thus the three minutes required by the sound spectrograph to analyze 2.2 seconds of speech represents a reduction of time to some 1/4000 of that required earlier.

The representation of intensity by the blackness of the record on the sound spectrogram adds a very useful third dimension to the time and frequency display of the plane figure. The various shades of darkness, however, do not give an accurate measure of intensity. Several methods of indicating sound amplitudes more accurately on spectrograms were proposed by Koenig and Ruppel (14). One of these methods, in which contour lines are marked on the spectrogram at fixed decibel intervals, has been developed more recently by Kersta and has been applied to the identification of individual voices (12).

The most widely used method of displaying amplitudes accurately with the sound spectrograph, however, does not normally include every point of time. Instead, two-dimensional plots of amplitude versus frequency are made at times selected by examining the three-dimensional spectrogram. The method has been described by Kersta (11), and the means for its use have been added to the various models of the spectrograph. Three or four intervals of time can be selected and analyzed in one run of about three minutes, so that on an average each "amplitude section" requires about 0.75 minutes. The records obtained are very similar to Fourier analyses made by older methods, although each display typically represents an integration over two or more periods of the speech signal. A set of such amplitude records, covering a speech sample as thoroughly as Steinberg did in the research cited above, can be made in about 1/200 of the time he required for an equal number of analyses.

Although the sound spectrograph has been used in the analysis of several different kinds of signals having changing spectra, its greatest use has been in research on speech. An early example of this use is provided by the book "Visible Speech" (21), which includes a comprehensive set of sound spectrograms of the phonemes of American English.

Modifications of the first sound spectrograph have been made in a number of laboratories, for the purpose of speeding up the analysis, extending the frequency range, or for performing desired functions not provided by the original machine. Some examples will be given below. A discussion of design parameters, and suggestions for spectrograph modifications to fulfill special purposes have been given by Peterson (18, 19).

As the source of the repeated signal which is required for analysis by a single filter in the sound spectrograph, Prestigiacomo (22) replaced the rotating magnetic ring and fixed reproducing head by a rotating head inside a stationary ring of magnetic tape. This ring is formed by moving any desired section of a normally recorded magnetic tape into the reproducing circle of the instrument. The linear speed of reproduction is 45 inches per second, which is six times greater than a recording speed of 7.5 inches/second. A recording 2.8 seconds long is analyzed from 50 to 7000 cps in about five minutes.

Many special purpose modifications of the sound spectrograph have been made. For the purpose of analyzing low level signals in noise, Peterson and Raisbeck (20) increased the integration period for amplitude sections. The device which they described had a number of different integrating periods, and more recent modifications of the spectrographs in the Communication Sciences Laboratory at the University of Michigan have provided for a continuous range of integrating times.

Huggins (9) replaced the usual bandpass filter of the spectrograph with three tuned circuits having slightly different peak frequencies. The phase relations of these tuned circuits cause the resonant frequencies of the speech formants to be shown more sharply, on the resulting "resonagram", than they are on the usual wide-band spectrogram.

The speed of spectrographic analysis was greatly increased by Mathes, Norwine and Davis (17), while retaining the features of magnetic recording and processing by a single filter. Their speed-up between recording and reproducing is 200 times, which results in reproduced frequencies up to 800,000 cps for recorded frequencies up to 4000 cps. The analyzing filter is centered at 1.2 megacycles per second, and the modulating oscillator moves between 1.2 and 2.0 Mc/s. Because of mechanical limitations in rotating the recording wheel at very high rim speeds, the length of sample analyzed is limited to 1/2 second. A total of 200 analyzing sweeps is made, so that with a speed-up of 200, the 0-4000 cps range can be analyzed in 1/2 second, the same as the sample length. Analysis results are displayed on a cathode-ray oscilloscope, with time and frequency on the x- and y-axes, and with amplitude applied to the brightness control. An amplitude section display is also provided, on another oscilloscope, at a selected time in the sample. For permanent record, the oscilloscope is photographed. In the device constructed a recorded sound can be transferred to the analyzing apparatus in 1/2-second portions, and each of these portions can be analyzed in 12 different equally spaced sections. The results of the analysis are then recorded on successive frames of a motion-picture film. The time required, allowing time for speed-up and slow-down of the recording wheel, and the twelve 1/2-second analyses of each 1/2-second period, is about 72 times real time. The resulting film, with 24 frames of amplitude sections per second, can be projected in synchronism with the recorded sound.

Spectrographic analysis in real time can be realized by the use of a bank of filters instead of sweeping the signal spectrum across the input of a single filter. One such arrangement was presented at the time of the development of the original sound spectrograph, in which twelve analyzing filters were used. The filter outputs controlled the intensity of the beam as it moved vertically in a rotating cathode-ray tube (23). The frequency display is not very accurate, of course, with such a small number of filters.

A real-time spectrograph employing 48 filters has been described by Fant (5) (with design attributed to H. Sund). The filters are mostly 300 cps wide, but overlapping, so that the total range is 0 to 10,000 cps. The filter outputs are scanned 500 times per second, with the results presented in a vertical line on a cathode-ray tube. Brightness of the cathode-ray spot at each location depends on the filter output voltage. The time axis is provided by a continuously moving film on which the tube face is focussed.

Another real-time spectrograph, by Harris and Waite (8), uses a bank of 54 Gaussian filters. The filter bandwidth is 70 cps up to a frequency of 1000 cps, and the bandwidth gradually increases for higher frequencies. The highest frequency reached is 16,800 cps. Filters overlap at the 3 db-down points. Filter outputs are scanned 500 times per second, and the result of each scan is displayed in the form of frequency versus amplitude, on a cathode-ray tube. A time axis is provided by a film which moves continuously in a direction perpendicular to the frequency axis. The side-by-side sweeps on the film give a good three-dimensional effect, though not quite the same as the conventional sound spectrogram.

The spectrograph of Gill (6) accomplishes real-time analysis without the use of a filter bank, and without a mechanical speed-up of a recorded portion of speech. By means of a recirculating digital delay line, 45 milliseconds of speech can be electronically stored and then played back in 150 microseconds, -- a speed-up of 300 times. The time-compressed signal is then analyzed repeatedly through a single heterodyne filter (with a bandwidth as small as 50 cps) in the same fashion as with the sound spectrograph. The input signal is fed into the delay line continuously to update the stored 45 ms sample at the same time that it is being recirculated and played back at high speed.

Weiss and Harris (28) also use delayed recirculation of signals in a self-correlation process. Their analyzer employs a carefully adjusted analog delay line and accomplishes the analysis in real time. The method is known as a coherent memory filter and is equivalent to the use of a continuous distribution of filters of a given bandwidth, this bandwidth being inversely proportional to the time interval used for the completion of each analysis. Thus an analysis over an 8000 cps range, made in 24 milliseconds interval, is equivalent to the use of a filter bandwidth of 63 cps over the entire range. Each such spectrum analysis is displayed on a cathode-ray oscilloscope, which can be photographed on motion picture film. They also convert the analysis data to digital form on a magnetic tape which is fed into a computer, and in about five times real time they obtain a complete record of formant frequencies and amplitudes, spectral power, plus the rates of change of these quantities, and voicing information.

Altogether, then, a variety of spectrographic analyzers have been designed and implemented. Unfortunately, it has not been possible in the above brief review to discuss all of the different machines which have been constructed, and some specialized machine operations have not been described. The various spectrographs employ different methods of signal storage, different types of waveform processing, and a variety of output displays. For analysis of the signal into its frequency components, the machines use a single filter or a bank of filters, or their equivalent in a self-correlation technique. A variety of circuits have been added to the basic spectrographic analyzer to perform special types of analysis or to analyze particular types of signals. In the two decades since the sound spectrograph was first described, spectrographic analysis has become a routine procedure in signal processing. A variety of spectrographic analyzers are now available in various laboratories for research on speech and other types of signals.

The purpose of this report is to present a further explanation of some of the principles of spectrographic analysis of sound, and to describe the sound spectrograph which has been designed and built in the Communication Sciences Laboratory of the University of Michigan. While the analysis concepts basic to the operation of the sound spectrograph are well known, it is hoped that a unified statement of the concepts will be helpful to those who are not fully familiar with spectrographic analysis. In Chapters 2 to 5 an attempt is made to present a statement of the principles of spectrographic analysis. These principles are generally well known, and for those who wish to pursue them further a number of references are given (1, 2, 3, 4, 7, 10, 15, 16, 24, 27, 29).

In Chapter 6 the CSL spectrograph will be described. Briefly, the instrument employs the original procedure of sweeping the lower sideband of the heterodyned signal across the passband of a single band-pass filter (or alternatively a set of low-pass filters). The analyzing speed-up of the instrument is greater than the original (6x), its time capacity is a little greater (2.9 sec), and its frequency range is larger (10,000 cps). The machine also has a greater flexibility in the choice of filter bandwidths, frequency range, and other parameters.

These and other features of the instrument will be described in some detail in Chapter 6, and the results illustrated in Chapter 7.

2. The Fourier Transformation and the Frequency Spectrum

Fourier analysis is particularly basic to the spectrographic analysis of speech wave forms. Consider first a class of real time functions f(t) which are <u>periodic</u> and <u>sectionally continuous</u>. That is, the form of the function is repeated endlessly in equal periods of time, T, and the function does not become infinite at any value of t, although its derivative may do so at isolated points. Then the function may be expressed as a sum of simple frequency functions, called a Fourier series. Such a series is defined as follows.

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_1 t}, \quad n=0, \pm 1, \pm 2, \ldots \qquad (2.1)$$

where

$$c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-in\omega_1 t} dt. \qquad (2.2)$$

As indicated above, T is the period of the function, and ω_1 is the repetition frequency in radians per second, $\omega_1 = \frac{2\pi}{T}$. Then $n\omega_1$ represents the frequency of one of the components of the Fourier series, and since n may assume negative as well as positive integer values, negative frequencies are implied. The more familiar trigonometric form of the Fourier series is

$$f(t) = a_o + \sum_{n=1}^{\infty} (a_n \cos n\omega_1 t + b_n \sin n\omega_1 t),$$

$$n = 1,2,3, \ldots \qquad (2.3)$$

where

$$a_o = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\, dt \qquad (2.4)$$

$$a_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \cos n\omega_1 t\, dt \qquad (2.5)$$

$$b_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \sin n\omega_1 t\, dt. \qquad (2.6)$$

Equation (2.3) can be reconciled with (2.1) and (2.2) if we put

$$a_o = C_o \qquad (2.7)$$

$$a_n = C_n + C_{-n} \qquad (2.8)$$

$$b_n = i(C_n - C_{-n}). \qquad (2.9)$$

C_n in general is a complex quantity, and C_{-n} is its conjugate. Thus a_n and b_n are real, and

$$C_n = 1/2\ (a_n - ib_n) \tag{2.10}$$

$$C_{-n} = 1/2\ (a_n + ib_n). \tag{2.11}$$

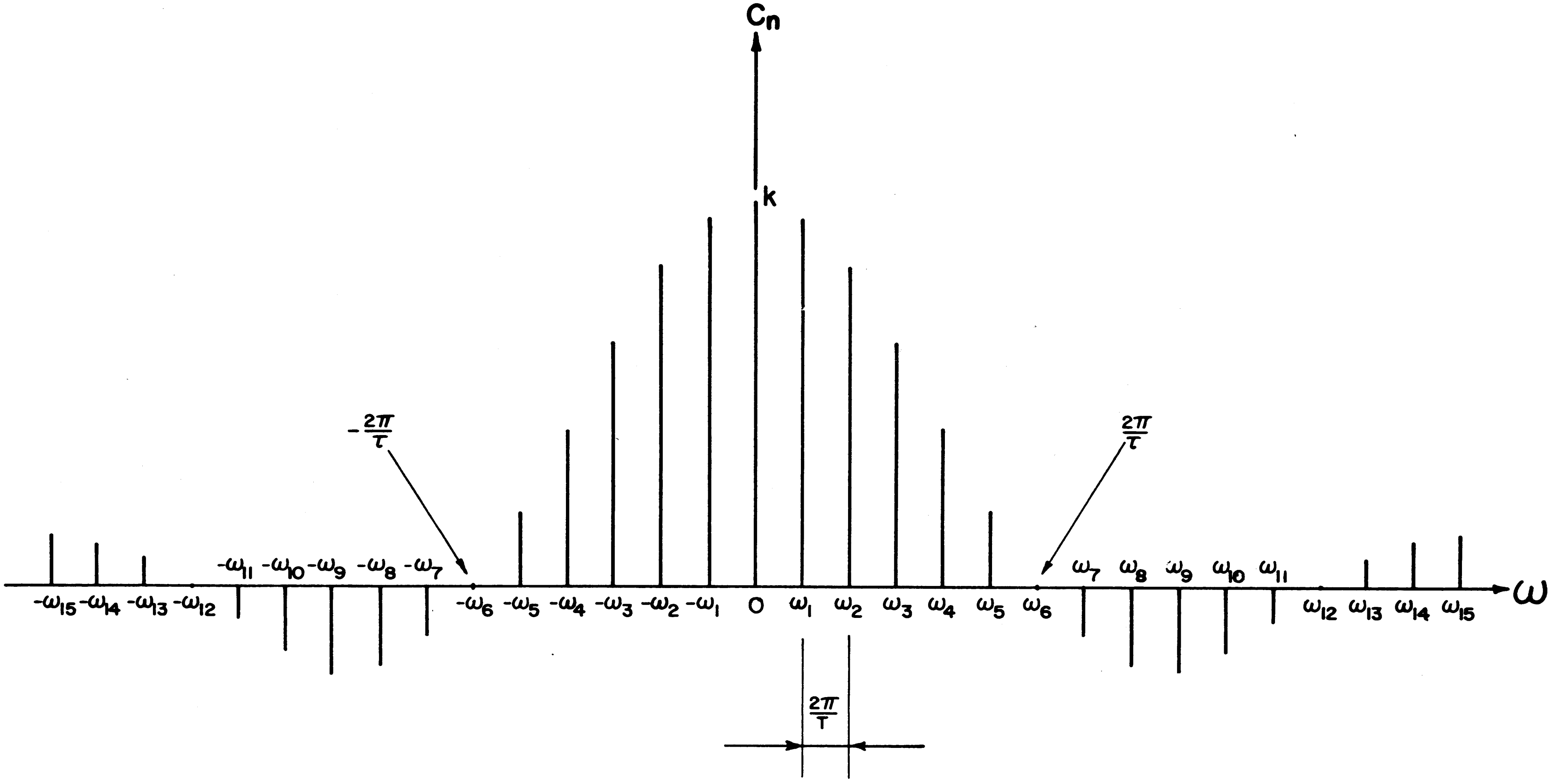
Also,

$$|C_n| = |C_{-n}| = (C_n C_{-n})^{1/2} = 1/2\ (a_n^{\ 2} + b_n^{\ 2})^{1/2}. \tag{2.12}$$

That is, the absolute value of C_n is one-half the amplitude of a frequency component in the trigonometric Fourier series.

By equations (2.1) and (2.2) a periodic function of time is transformed into a frequency function. The plot of $|C_n|$ versus frequency is known as the frequency spectrum of f(t). In the particular example shown in Figure 2.1, the values of C_n are all real, and these are the quantities plotted at both positive and negative frequency. The actual amplitudes of the real frequency components can be found in this case by doubling the amplitudes on the right side of Figure 2.1, but leaving C_o (the d.c. component) unchanged. The frequency spectrum is seen to be a set of discrete values in the frequency domain.

If f(t) is non-periodic it can also be subjected to a Fourier transformation, provided the sectionally continuous condition holds for every finite interval and the following convergent condition is also satisfied,



T IS THE PERIOD OF THE RECTANGULAR PULSE
τ IS THE DURATION OF THE PULSE

Figure 2.1. Frequency spectrum of a rectangular pulse train (for $\tau = \frac{1}{6} T$).

$$\int_{-\infty}^{\infty} \left| f(t) \right| \, dt < \infty . \tag{2.13}$$

The Fourier transformation can be written,

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} \, d\omega \tag{2.14}$$

where

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt . \tag{2.15}$$

Equations (2.14) and (2.15) are known as a Fourier transform pair. Equation (2.15) is often called the Fourier integral. It is the transformation of a non-periodic time function into a frequency function. The function $F(\omega)$, which takes the place of C_n in the periodic case, does not yield a set of amplitudes at discrete frequencies, but is continuous with frequency. It has both real and imaginary parts, and hence may be written

$$F(\omega) = \left| F(\omega) \right| \, e^{i\theta(\omega)} , \tag{2.16}$$

where $\left| F(\omega) \right|$ describes the amplitude spectrum and $\theta(\omega)$ represents the phase spectrum. Both $\left| F(\omega) \right|$ and $\theta(\omega)$ are continuous functions over the frequency domain.

By way of illustration, the Fourier analysis of a set of elementary functions of considerable importance in signal analysis will next be considered.

Fourier analysis of a periodic pulse

Consider a periodic train of rectangular pulses defined over the real time axis ($-\infty < t < \infty$), with a period T between pulses and a pulse width τ. In the period nearest t = o, let f(t) be defined as

$$f(t) \begin{cases} = 0, & -\frac{T}{2} < t < -\frac{\tau}{2} \\ = 1, & -\frac{\tau}{2} \leq t \leq \frac{\tau}{2} \\ = 0, & \frac{\tau}{2} < t < \frac{T}{2}. \end{cases} \tag{2.17}$$

The complete function can be defined if t in the above inequalities is replaced by t - nT, with $n = 0, \pm 1, \pm 2, \ldots.$

Substituting the f(t) of (2.17) into equation (2.2) gives

$$C_n = k \frac{\sin \pi nk}{\pi nk}, \tag{2.18}$$

where

$$k = \frac{\tau}{T}.$$

Then from (2.1) the Fourier series may be written

$$f(t) = k \sum_{n=-\infty}^{\infty} \frac{\sin \pi nk}{\pi nk} e^{in\omega_1 t}, \quad n=0, \pm 1, \pm 2, \ldots \tag{2.19}$$

In Figure 2.1, C_n (which in this case is real only) is plotted against ω, where

$$\omega_n = n\omega_1 = \frac{2\pi n}{T} \qquad (2.20)$$

and where k is given the particular value 1/6.

It is useful to note, from Equation (2.18), that the first zero of C_n occurs when

$$n = \frac{1}{k} = \frac{T}{\tau}. \qquad (2.21)$$

The corresponding value of ω is

$$\omega = \frac{2\pi}{\tau}. \qquad (2.22)$$

This zero will occur at an exact harmonic only if 1/k is an integer, as it is in the case plotted in Figure 2.1. But in any case the above relation shows that as pulse width decreases, the frequency components of the pulse train extend out over a greater frequency range before a zero amplitude is reached. Most of the energy of the pulse train will lie in the range

$$0 < |\omega| < \frac{2\pi}{\tau}. \qquad (2.23)$$

Thus the first zero crossing of the frequency spectrum is often regarded as the frequency spread of a signal. It is convenient to define a <u>bandwidth</u> as specifying the range of frequencies in which most of the signal energy is concentrated. In accordance with the above considerations, this bandwidth, Δf, can be defined as extending from zero frequency to the first zero crossing. That is,

$$\Delta f = \frac{1}{\tau} \qquad (2.24)$$

where

$$\tau \ll T.$$

Furthermore, it is important to note that any other criterion for the definition of bandwidth would still retain the inverse relationship between duration and bandwidth, so that

$$\Delta f = \frac{C}{\tau} \qquad (2.25)$$

where C is a constant depending on the criterion chosen.

Fourier transformation of a unit rectangular pulse

Instead of a periodic train, suppose there is one pulse only, centered at zero time.

$$f(t) \begin{cases} = 0, & -\infty < t < -\frac{\tau}{2} \\ = 1, & -\frac{\tau}{2} \leq t \leq \frac{\tau}{2} \\ = 0, & \frac{\tau}{2} < t < \infty \end{cases} \qquad (2.26)$$

The Fourier transform of this unit rectangular pulse can be obtained by substituting the particular f(t) into equation (2.15). This gives

$$F(\omega) = \tau \frac{\sin \omega(\frac{\tau}{2})}{\omega(\frac{\tau}{2})}. \qquad (2.27)$$

$F(\omega)$ is real only, and a plot of $F(\omega)$ versus frequency is given in Figure 2.2. The function is continuous with frequency, its first zero coming at $\omega = \frac{2\pi}{\tau}$. Thus, as with the periodic pulse train, the spread of the frequency spectrum is inversely proportional to pulse width. A shorter duration of the unit pulse gives a broader frequency spectrum.

Spectrum of a delta function

The delta function $\delta(t - t_o)$ is defined on the real time axis, as existing at $t = t_o$ but having zero value at all other points of time. Also by definition, it has the following relation

$$\int_{-\infty}^{\infty} \delta(t - t_o)\, dt = 1. \qquad (2.28)$$

Thus the delta function is a pulse of zero duration but infinite amplitude.

By equation (2.15), the Fourier transform of the delta function is

$$F(\omega) = \int_{-\infty}^{\infty} \delta(t - t_o)\, e^{-i\omega t}\, dt. \qquad (2.29)$$

Since the delta function is zero everywhere except at t_o, the exponential can be taken as constant and the remaining part of the integral is given by (2.28). Thus

$$F(\omega) = e^{-i\omega t_o}. \qquad (2.30)$$

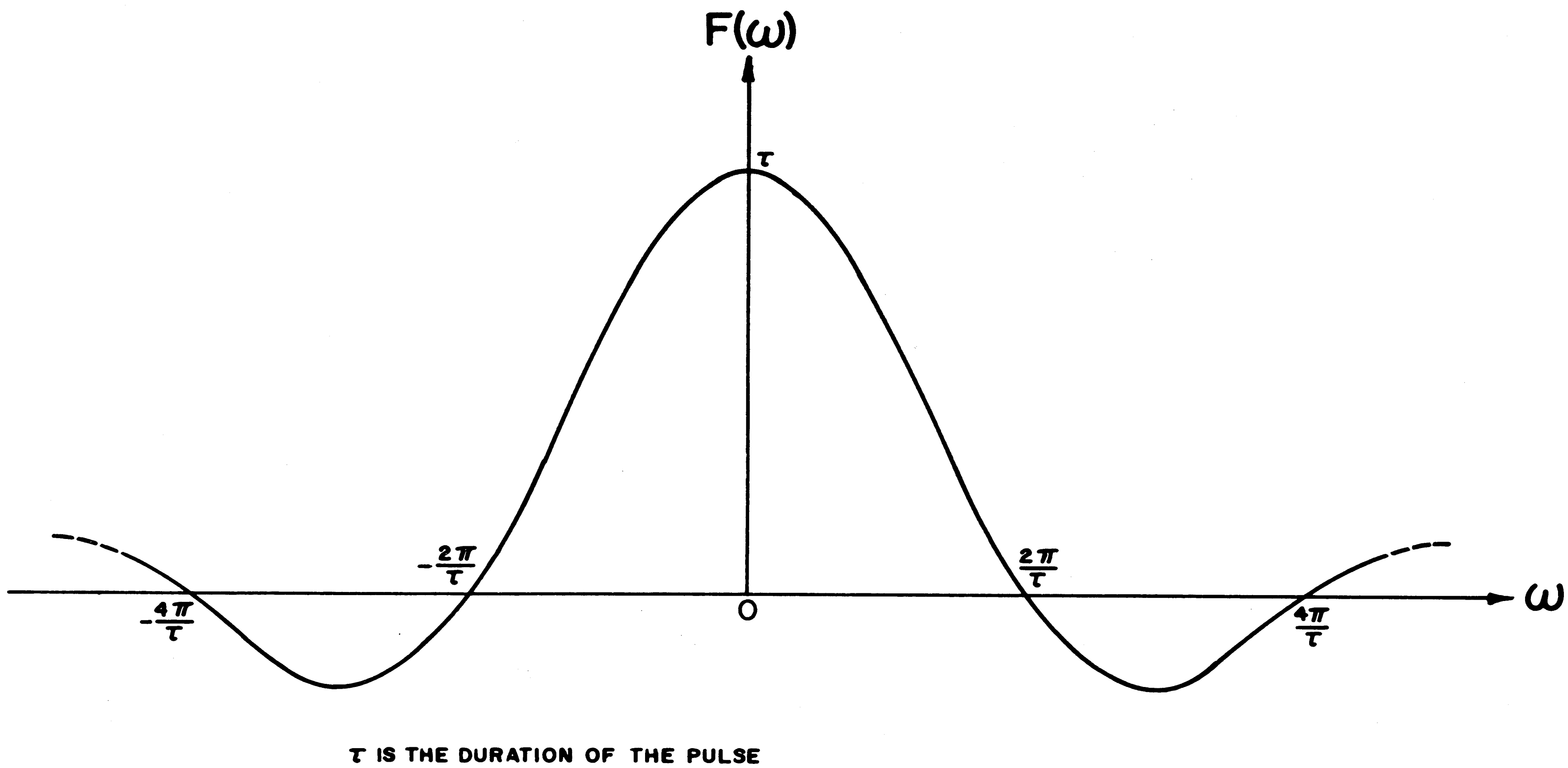


Figure 2.2. Frequency spectrum of a single rectangular pulse.

Separating $F(\omega)$ into amplitude and phase parts, as in equation (2.16), gives

$$|F(\omega)| = 1, \ \theta(\omega) = -\omega t_o. \qquad (2.31)$$

Equation (2.31) shows that the amplitude spectrum of a delta function is a continuous frequency function of unit height, extending over the whole frequency range. If the delta function occurs at zero time ($t_o = 0$), the phase spectrum in (2.31) vanishes, and equation (2.30) becomes

$$F(\omega) = 1. \qquad (2.32)$$

For any other value of t_o, the spectral vector rotates continuously with the angle $-\omega t_o$.

The frequency spectrum of the delta function $\delta(t)$ of equation (2.32) is shown in Figure 2.3.

Parseval's theorem

Parseval's theorem states that the energy of a periodic time function is proportional to the sum of the energies of the frequency components (i.e., the sum of the squares of the amplitudes $|C_n|$ of the function). Thus the energy in each period T of a periodic function f(t) is

$$E = K\int_{-\frac{T}{2}}^{\frac{T}{2}} f^2(t)dt = KT \sum_{n=-\infty}^{\infty} |C_n|^2, \ n = 0, \pm 1, \pm 2, \ldots, \qquad (2.33)$$

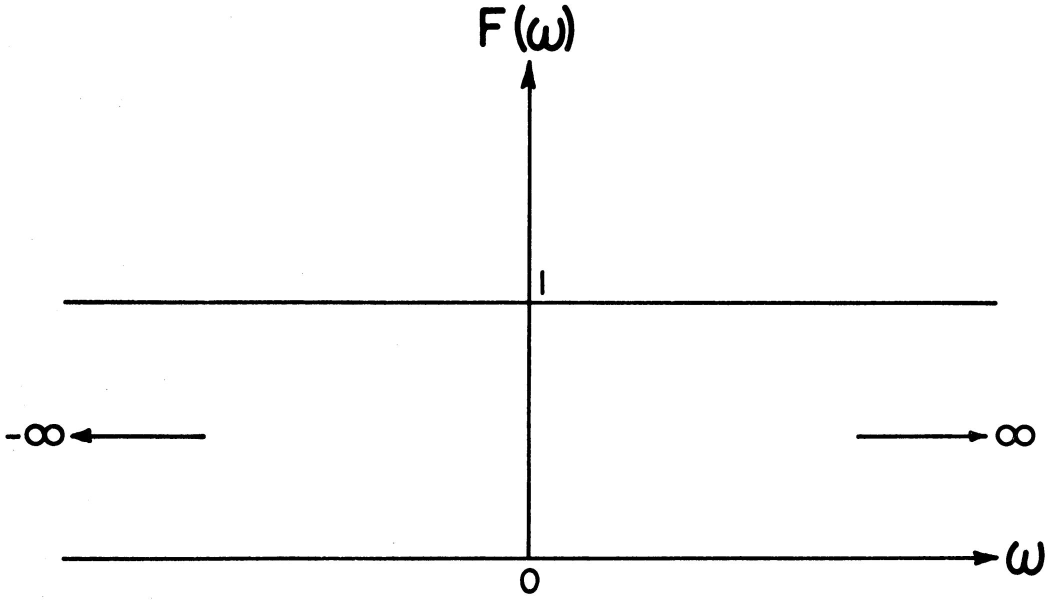


Figure 2.3. Frequency spectrum of a delta function.

where K is a proportionality constant. The above equation (2.33) is called the Parseval relation. It can be extended to the case of an aperiodic time function by letting the period T approach infinity. Equation (2.33) then becomes

$$E = K\int_{-\infty}^{\infty} f^2(t)\ dt = \frac{K}{2\pi} \int_{-\infty}^{\infty} \left|F(\omega)\right|^2 d\omega. \quad (2.34)$$

The quantities $\left|C_n\right|^2$ and $\left|F(\omega)\right|^2$ in (2.33) and (2.34) have the dimensions of spectral energy and spectral energy density, respectively. It may be stated, from these equations, that the energy of the time function f(t) is proportional to the square of its rectification.

3. Modulation and Detection Theory

Modulation

Modulation, as a process of frequency conversion, has been used extensively to transmit informational signals in a medium which separates a transmitter and a receiver. The medium may consist of a dielectric, a wave guide, wires, etc. When the parameters are appropriately chosen, modulation provides a more efficient means of signal transmission than can otherwise be realized. In general there are various methods of producing modulated signals, but they can be summarized into two general categories:

a. Continuous wave modulation, in which the amplitude, frequency, or phase of a sinusoidal carrier is varied continuously in accordance with the amplitude of the informational signal.

b. Discrete wave modulation (or pulse modulation), in which the height, the position, or the width of a set of carrier pulses is altered in a quantized way in accordance with a discrete informational signal.

Sometimes, for greater efficiency or for economical reasons, a combination of both methods is used. Each modulation technique has its own advantages and limitations. Since amplitude modulation is normally employed in spectrographic analysis, only amplitude modulation will be discussed here. As explained briefly in Chapter 1, the purpose of using modulation in the spectrograph is to shift the spectrum of the input signal continuously across a fixed filter. This procedure makes possible a much more detailed analysis than can be realized with a filter bank, unless the bank is very elaborate.

In general, an amplitude-modulated signal may be described by the following expression:

$$f_c(t) = k\,[1 + m\,f(t)]\cos\omega_c t \qquad (3.1)$$

where k is an arbitrary constant,

m is the modulation index,

f(t) is the informational signal,

and ω_c is the carrier frequency.

In explaining equation (3.1), it is said that the frequency ω_c "carries" the informational signal f(t). When ω_c is appreciably higher than the components of f(t), the form of the latter appears as the envelope of $f_c(t)$.

If f(t) is a periodic wave, then it can be expanded as a Fourier series, as in equation (2.1).

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{i\omega_n t}, \quad n=0, \pm 1, \pm 2, \ldots \qquad (3.2)$$

where

$$\omega_n = n\omega_1 = \frac{2\pi n}{T},$$

and T is the period of repetition of the wave. If there is a maximum significant frequency, ω_M, in f(t), such that

$$\omega_M = \frac{2\pi M}{T}, \qquad (3.3)$$

then the modulated signal $f_c(t)$ can be written

$$f_c(t) = k\left[1 + m \sum_{n=-M}^{M} C_n e^{i\omega_n t}\right] \cos \omega_c t. \qquad (3.4)$$

By observing the relations given in Chapter 2, among C_n, $|C_n|$, a_n and b_n, equation (3.4) can be reduced to

$$f_c(t) = k' \cos \omega_c t + mk \sum_{n=1}^{M} |C_n| \left\{ \cos[(\omega_c + \omega_n)t - \varphi_n] + \cos[(\omega_c - \omega_n)t + \varphi_n] \right\}, \qquad (3.5)$$

where $k' = k(1 + mC_o)$

and $\varphi_n = \tan^{-1} \dfrac{b_n}{a_n}$.

Thus $f_c(t)$ in its Fourier series form contains the carrier frequency ω_c and in addition sidebands, containing the difference frequencies $(\omega_c - \omega_n)$ on the lower side, and the sum components $(\omega_c + \omega_n)$ on the upper side.

The validity of equation (3.5), as a mathematical result of (3.1) and (3.2), does not depend on any particular relation among the quantities in (3.1). However, some practical considerations may set limits to these relations. In particular the ordinary physical means of implementing equation (3.1) do not permit the instantaneous value of $f_c(t)$ to be opposite in sign to that of $(\cos \omega_c t)$. That is, the quantity $[1 + mf(t)]$ should not be negative, or,

$$mf(t) \geq -1 \qquad (3.6)$$

for all values of t. In fact, most modulators generate a series of modulation products in addition to the first order sidebands of equation (3.5), and these higher products must be limited in amplitude in order that equation (3.1) be approximately satisfied. This requires that the amplitude of the informational signal be kept small compared with that of the applied carrier. The result is a more severe limitation than (3.6),

$$|mf(t)| \ll 1.$$

In a communication system it is often required that the envelope of the modulated wave carry the entire informational content of the signal f(t). This envelope is formed essentially by a "sampling" of f(t) at the frequency rate of the carrier, and this rate must then be at least twice that of the maximum frequency which carries information in f(t). That is,

$$\omega_c \geq 2\omega_M. \tag{3.7}$$

This restriction is not required in the sound spectrograph, since analysis by this instrument is performed on the modulated signal itself, not on the envelope.

It has been the usual practice in spectrograph design to observe a less severe restriction than (3.7),

$$\omega_c \geq \omega_M. \tag{3.8}$$

This condition serves to keep the two sidebands of the modulated wave in completely separate frequency regions. It will be shown in Chapter 5, however, that (3.8) need not be satisfied if the frequency region of the analyzing filter band is not adversely affected by the modulation products.

Another important aspect of the modulation process, as summarized in equation (3.5), is that each component of a sideband, $(\omega_c \pm \omega_n)$, has the same relative amplitude, $|C_n|$, as the corresponding component ω_n has in f(t), equation (3.2). A measurement of the f(t) component amplitudes can therefore be made using only the sideband components.

An example of a modulated signal is shown in spectrum form on the right side of Figure 3.1. For comparison, the spectrum of the original informational signal f(t) is shown on the left. In effect, the original spectrum is shifted bodily to form the upper sideband, and in addition to the shift the component frequencies are reversed in order in the lower sideband.

A plot like Figure 3.1 shows only the amplitudes and frequencies of the components of the spectrum and not the phases, i.e. the various values of ϕ_n in equation (3.5). The phase angle can be made to disappear for a particular component by choosing the zero of time to correspond with a positive maximum of this component. For some functions, like that plotted in Figure 2.1, all of the phase angles are zero, but in general this is not the case.

The constant k' is usually equal to k, since in the modulation process the informational signal usually has been passed through a transformer and the dc component (C_o) thereby removed.

When the informational signal f(t) is non-periodic, equation (3.1) still applies,

$$f_c(t) = k \cos \omega_c t + km\, f(t) \cos \omega_c t. \qquad (3.9)$$

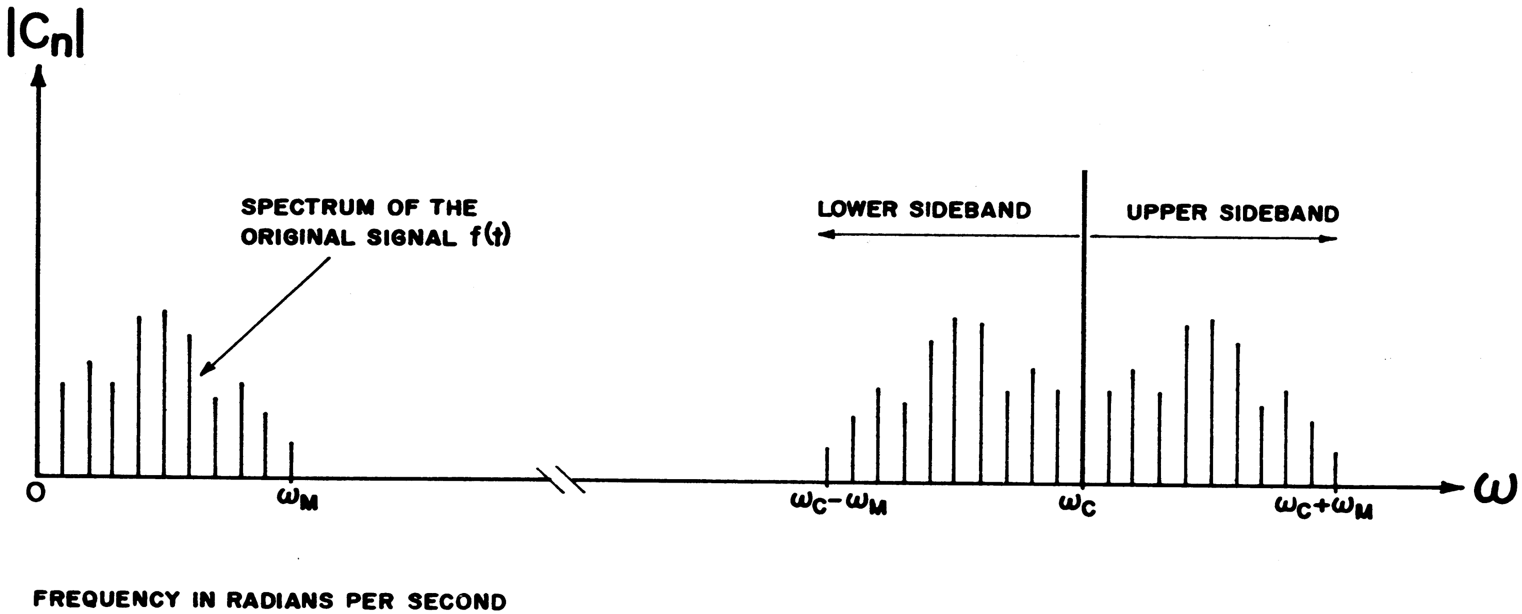


Figure 3.1. Frequency spectrum of a typical periodic, band-limited signal, and its modulated signal.

In the second term of (3.9) the <u>shifted</u> signal may be written as

$$f_s(t) = f(t) \cos \omega_c t = 1/2\ f(t)(e^{i\omega_c t} + e^{-i\omega_c t}). \quad (3.10)$$

The Fourier transformation of equation (2.14) can then be applied to $f_s(t)$,

$$f_s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_s(\omega) e^{i\omega t}\, d\omega \quad (3.11)$$

where

$$F_s(\omega) = \int_{-\infty}^{\infty} f_s(t) e^{-i\omega t} dt.$$

Taking $f_s(t)$ from (3.10), $F_s(\omega)$ becomes

$$F_s(\omega) = 1/2 \int_{-\infty}^{\infty} f(t)[e^{-i(\omega-\omega_c)t} + e^{-i(\omega + \omega_c)t}]dt. \quad (3.12)$$

Taking the two terms of the integral separately, each has the form of (2.15) except that the frequencies are shifted to the sidebands.

$$F_s(\omega) = 1/2\ [F(\omega - \omega_c) + F(\omega + \omega_c)] \quad (3.13)$$

If (3.13) is substituted into equation (3.11) and the result is then substituted into (3.9), and if in addition the frequencies in f(t) are limited to a maximum ω_M, then

$$f_c(t) = k \cos \omega_c t + \frac{km}{2\pi} \int_{-\omega_M}^{\omega_M} 1/2\ [F(\omega - \omega_c) + F(\omega + \omega_c)] e^{i\omega t}\, d\omega. \quad (3.14)$$

That is, when the informational signal is non-periodic, the modulated signal consists of the carrier frequency plus sidebands composed of the shifted spectrum of the informational signal. In this case the sidebands are continuous with frequency. An example of the spectrum of a non-periodic wave and of the resulting modulated spectrum is shown in Figure 3.2. In general $F(\omega)$ is complex, and only the real parts $F_r(\omega)$ are shown in the figure. These include negative frequencies. Absolute values, $|F(\omega)|$, are formed from a combination of the positive and negative frequencies.

Single-sideband operation and balanced modulation

It is evident from Figures 3.1 and 3.2 that either the lower or the upper sideband carries the entire informational content of the signal. Elimination of the carrier itself and one of the sidebands will therefore cause no loss in information. While this fact is not particularly important in spectrographic design, it does result in a saving of power and of frequency space. Thus in electrical communications it is the general practice to make the above indicated eliminations. Both can be accomplished by means of filters, but it is more efficient to eliminate the carrier (or reduce it to a low level) by the technique of balanced modulation. This removes

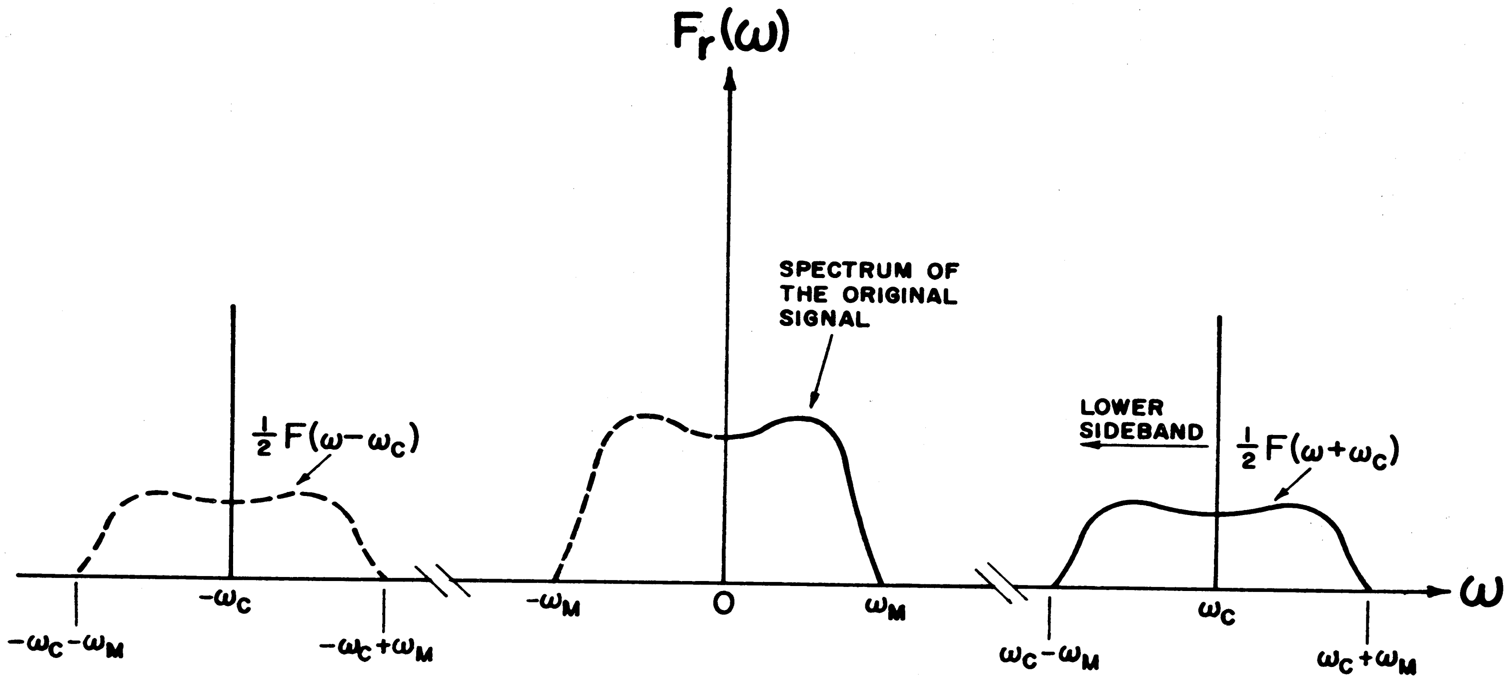


Figure 3.2. Frequency spectrum of a typical non-periodic, band-limited signal, and its modulated signal.

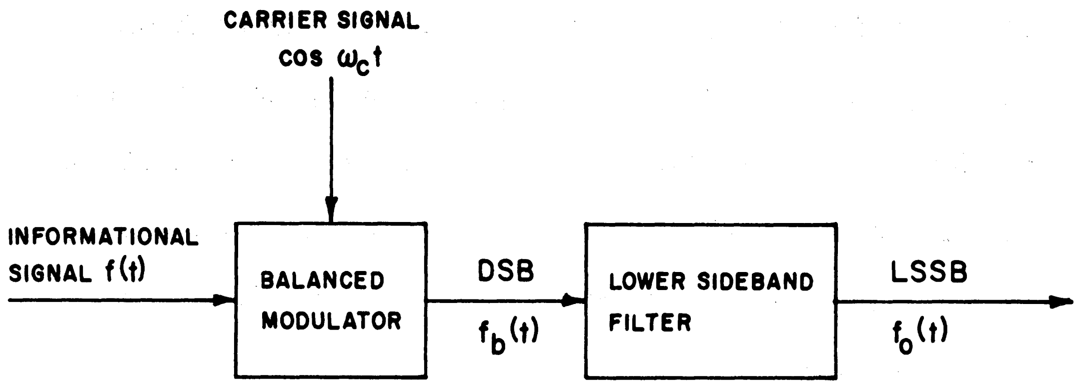
the first terms in equations (3.5) and (3.14). It is then relatively easy to filter out one of the sidebands, particularly if no extremely low frequencies are included in f(t). A block diagram of such a balanced modulation system is shown in Figure 3.3.

Detection

Although detection is not a part of the spectrographic process outlined in Chapter 5, it is used in the recovery of an original signal from the modulated wave received in a communication system. Amplitude detection will thus be considered briefly.

When the entire modulated signal (carrier and both sidebands) is being received, a rectification of the signal reduces the double (positive and negative) envelope to a single envelope of high-frequency pulsations. Further reduction to the envelope alone (which is just the informational signal) is then accomplished by low-pass filtering. Unless condition (3.7) has been followed in the modulation, information will be lost in this type of detection.

When a single sideband is received, the carrier frequency ω_c must be resupplied from an oscillator at the receiving terminal. Suppose the received signal is the lower sideband, with the frequencies $(\omega_c - \omega)$. The addition of the local ω_c produces new sidebands



WHERE $f_b(t) = \frac{k'}{T} \sum_{n=-M}^{M} c_n e^{j\omega_n t} \cos \omega_c t$

DSB = DOUBLE SIDEBAND SIGNAL

$f_o(t) = \frac{k'}{T} \sum_{n=1}^{M} |c_n| \cos(\omega_c + \omega_n) t$

LSSB = LOWER SINGLE SIDEBAND SIGNAL

Figure 3.3. Method of producing single sideband transmission.

with the sum and difference frequencies $(2\omega_c - \omega)$ and ω. The first of these is easily removed by filtering, if condition (3.8) was observed in the modulation, leaving only the original signal. A typical example of frequency shifting in modulation and demodulation for a single-sideband system is shown in Figure 3.4. Figure 3.5 shows a block diagram of the single-sideband detection process.

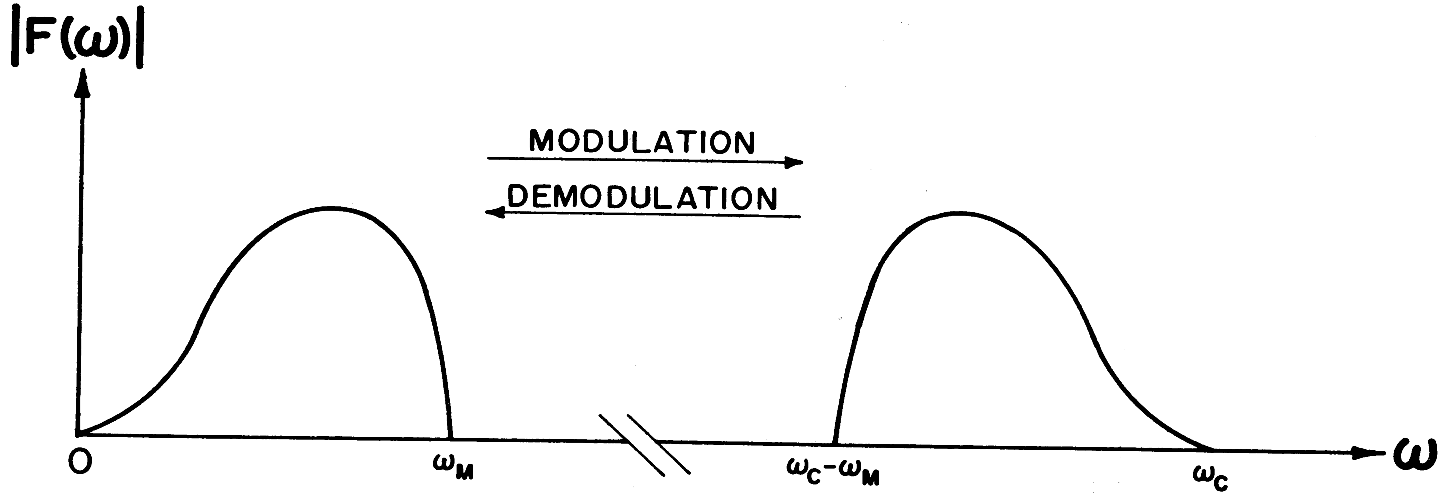


Figure 3.4. Single lower sideband demodulation.

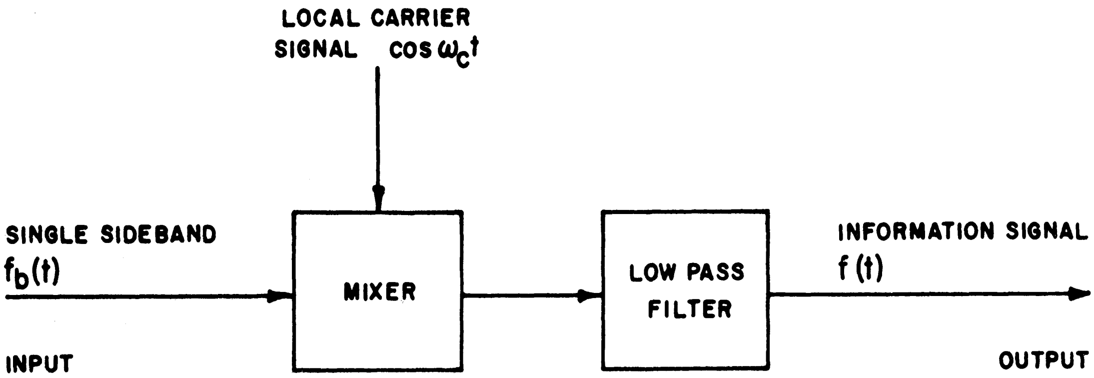


Figure 3.5. Block diagram of a single sideband detector.

4 Signal Filtering

Effect of a linear network on an informational signal

Suppose that information is to be represented by a series of rectangular pulses. Relatively high rates of information require a relatively large number of pulses per second, and therefore narrow pulses. It has been shown in Chapter 2 that there is an inverse relationship between pulse width and the spread of the frequency spectrum. Hence in a band-limited system there is a lower limit to pulse width, and therefore an upper limit to the rate at which the system can handle information. Expressed another way, a limitation in the frequency band of a system places a limit on the minimum time required for the system to respond.

Rectangular pulses in a band-limited channel will inevitably be rounded off at the corners. If information is carried only by the number of pulses and not by their shape, then a distorted version of the pulses at the output of a channel may be acceptable. Whether or not the pulses are correctly transmitted depends upon the time constant of the channel; that is, the time required, after a pulse is applied, for the output of the channel to reach a considerable fraction of its maximum value (an exact definition is not required here). The shorter this time constant, the more perfect the transmission will be. However, if only the number of pulses and not their shape is important, it will be sufficient if the time constant is of the same order of magnitude as the pulse width. That is,

$$\text{T.C.} \leq \tau, \tag{4.1}$$

where T.C. is the channel's time constant, and τ is the pulse width.

On the other hand, if information is carried by pulse shape as well as by the number of pulses, then the time constant of the channel must be much smaller,

$$\text{T.C.} \ll \tau. \tag{4.2}$$

Rather than using the time constant, it may be more convenient to speak of the bandwidth, Δf, of the channel. This bandwidth is closely related to the reciprocal of the time constant. Again without an exact definition, the case for transmission of pulses without regard to their shape may be stated as

$$\Delta f \geq \frac{1}{\tau}. \tag{4.3}$$

This inequality may be compared with equations (2.24) and (2.25).

The more restricted case for the transmission of pulse shape is given by

$$\Delta f \gg \frac{1}{\tau} . \qquad (4.4)$$

In work in pulse transmission, condition (4.4) is sometimes called the high fidelity criterion (e.g. 24, p. 30).

Response from an ideal linear network

Many problems in signal transmission through a linear electrical network can be solved more easily by application of the Fourier transformation technique.

Let a linear network have a response h(t) to the application of a unit impulse. That is, h(t) represents the output function of the system when a large potential is applied momentarily to the input, at zero time. The Fourier transform of h(t) will be designated $H(\omega)$, and called the transfer function of the system. It can be divided into amplitude and phase factors.

$$H(\omega) = |H(\omega)|\, e^{i\theta(\omega)} \qquad (4.5)$$

where $|H(\omega)|$ is the transfer amplitude characteristic, and $\theta(\omega)$ is the transfer phase characteristic.

If now a time function f(t) is applied to the input, the output g(t) of the system will be given by the convolution integral,

$$g(t) = \int_{-\infty}^{\infty} f(t')h(t-t')\, dt'. \tag{4.6}$$

That is, the output at any time (t) depends on the entire past history (t') of f(t'), modified by h(t-t'). Let $G(\omega)$ and $F(\omega)$ be the Fourier transforms of g(t) and f(t), respectively. It has been shown that the Fourier transform of the convolution integral (4.6) is equal to the product of the Fourier transforms of the convoluted parts. That is,

$$G(\omega) = F(\omega) \cdot H(\omega). \tag{4.7}$$

This relation makes the solution of (4.6) much easier in many cases. The output function g(t) is simply the inverse Fourier transfer of (4.7),

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega)F(\omega)e^{i\omega t}\, d\omega \tag{4.8}$$

When the spectrum of the output g(t) differs from that of the input f(t), the distortion is due to the band-limited and nonlinear characteristics of the system function $H(\omega)$. A discussion of the distortion caused by such a network will be presented in the following sections.

Distortionless network

The problem of distortionless operation is of course very important in communication systems. A system is said to be distortionless if the shape of the output time response is identical with that of the input signal. Such a system, in block form, is pictured in Figure 4.1.

Let the distortionless time relation be

$$g(t) = kf(t-t_o) \qquad (4.9)$$

where a constant time delay t_o in the output is not considered to produce distortion. The Fourier transform of (4.9) is

$$G(\omega) = k \int_{-\infty}^{\infty} f(t-t_o)e^{-i\omega t}dt = ke^{-i\omega t_o} \int_{-\infty}^{\infty} f(t')e^{-i\omega t'}dt' \qquad (4.10)$$

where $t' = t-t_o$. The integral in (4.10) is simply the Fourier transform, $F(\omega)$, of $f(t)$.

$$G(\omega) = ke^{-i\omega t_o} F(\omega) \qquad (4.11)$$

By comparison of (4.11) with (4.7), the transfer function of the distortionless system can be deduced,

$$H(\omega) = ke^{-i\omega t_o}. \qquad (4.12)$$

By comparison with (4.5), the amplitude characteristic of a distortionless network should be constant

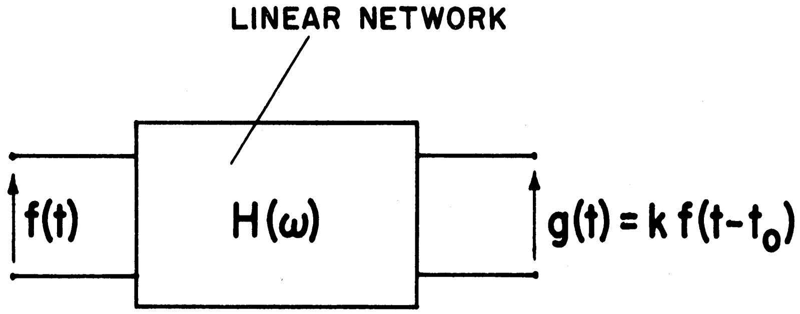


Figure 4.1. Block diagram representation of a distortionless network.

$|H(\omega)| = k$, over all frequencies contained in $f(t)$; and the phase characteristic should be linear with frequency, $\Theta(\omega) = -t_o\omega$, over the same range.

Signal distortion

In Figure 4.2 is shown an example of two stages in signal distortion. A rectangular input pulse is given in (a). In (b) is shown the output of a system which is limited in the transmission of high frequencies. Such a system will produce a change in the shape of the pulse but may leave it approximately symmetrical. It also introduces a constant time delay, causing the shift of the pulse center to t_o.

If, in addition, the phase of the system is not linear with frequency, the effect is to give each frequency a different delay time. The result may be something like that shown in Figure 4.2 (c), where the pulse is now markedly unsymmetrical. While symmetry is not a perfect measure of non-linear-phase response, it has been used in the above example for illustrative purposes.

Linear phase filter

In accordance with the above principles, if distortionless operation is desired, a filter placed

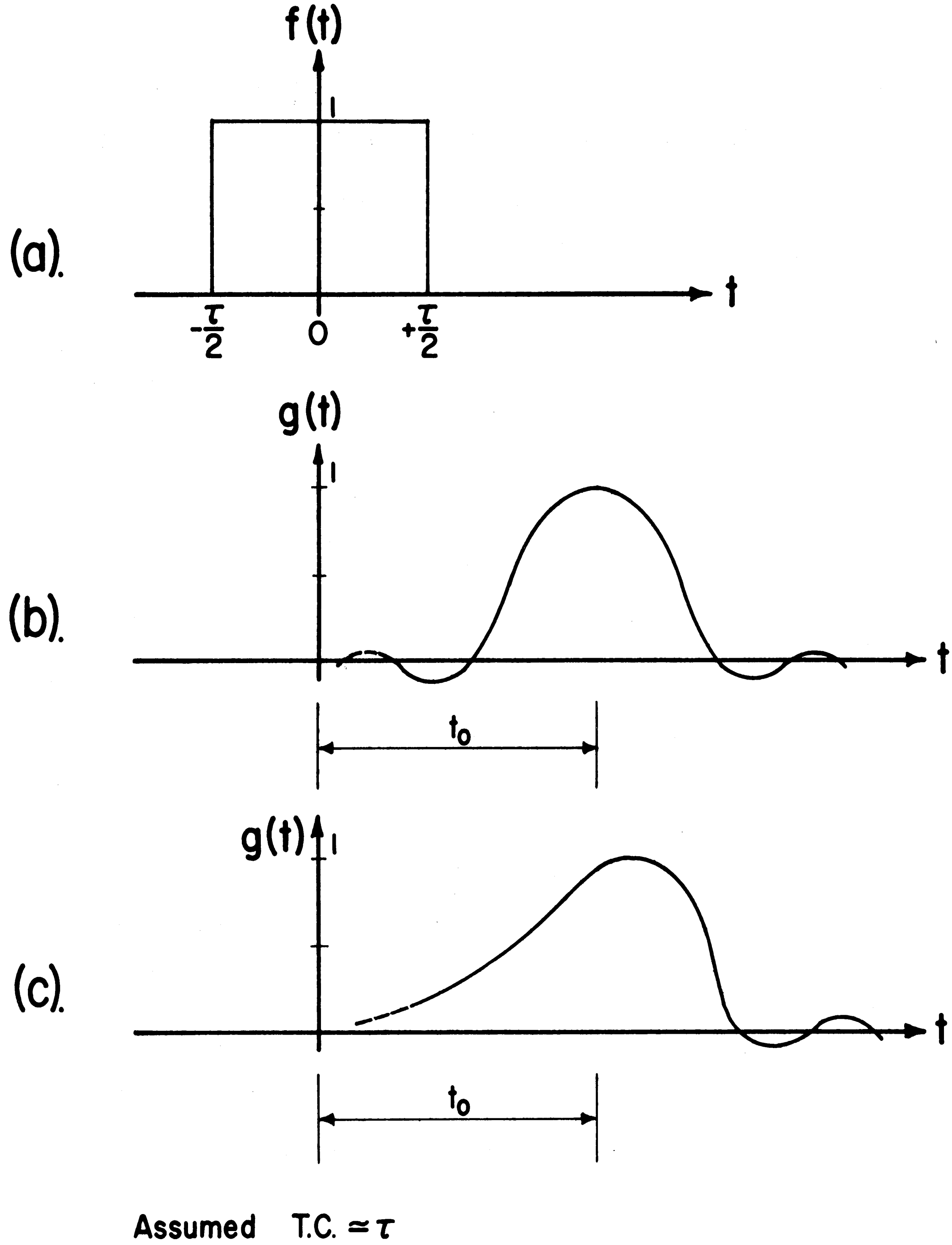


Figure 4.2. Effects of band limitation and phase variation on signal transmission:

(a) Rectangular pulse, input to network;

(b) Output response of a band-limited network, but with linear phase shift;

(c) Output response of network with non-linear phase shift.

in a system should have a flat amplitude characteristic over its assigned range of frequencies, and in addition should have a linear phase shift over this range. Realization of such a filter is difficult. In practice, variations in amplitude and in phase-shift characteristics from the ideal may occur and for certain purposes may be entirely acceptable. The effects of some of these variations, in a realizable filter, will be demonstrated by spectrographic analysis in Figures 7.1 and 7.6.

5. The Single-Filter Spectrograph

Basic spectrographic operations

In Chapter 1 an introduction to the sound spectrograph which uses only one band-pass filter for analysis was presented. The complete audio-frequency range is analyzed by placing the pass-band of the filter higher in frequency than the range to be analyzed, and using a heterodyne process for moving the different parts of the signal spectrum to the filter band. Repeated reproductions of the recorded signal are necessary. In this chapter the principles discussed in the preceding chapters will be used in a more detailed description of this type of sound spectrograph. With a few additions, the spectrograph to be described will be similar to those of References (13) and (11).

In the circuit of Figure 5.1, the signal which is to be analyzed is indicated by the input function f(t). We will consider f(t) to be an electrical wave which is a replica of an acoustic pressure due to <u>speech</u>; the process of conversion from sound by means of a microphone is omitted

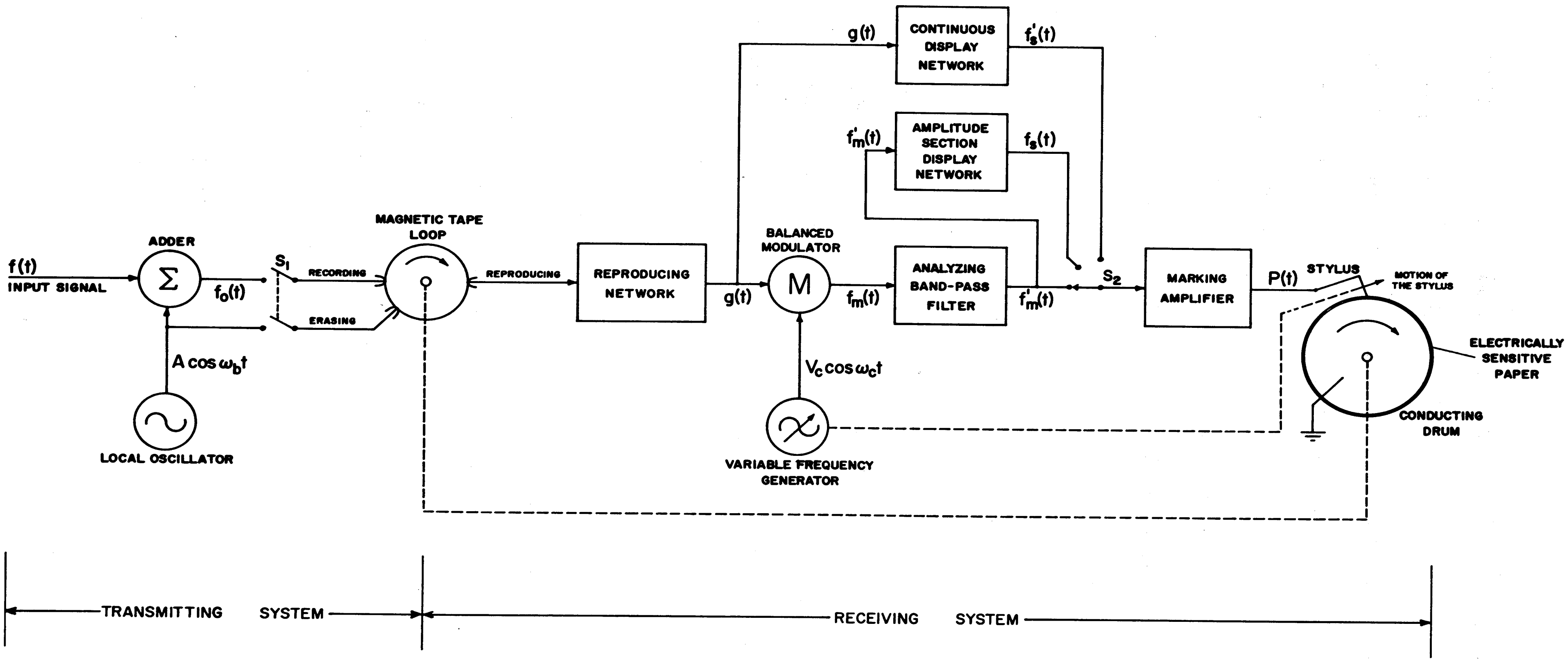


Figure 5.1. Block diagram of a sound spectrograph.

from the figure. It is not implied, however, that the sound spectrograph is limited to the analysis of speech. Any sound in the audio-frequency range can be analyzed with the same equipment. In fact, with suitable transducers any disturbance in the audio-frequency range can be analyzed with the sound spectrograph, regardless of the origin of the disturbance. The analysis can also be extended to other frequency ranges if the signals are first modulated to the audio range before insertion at the spectrograph input.

Whether the input f(t) is periodic or non-periodic, it has a spectrum of frequencies given by the Fourier Transform $F(\omega)$, and it is the purpose of the analyzer to measure the amplitude of each radian frequency (ω) present, at each time (t) of the input function. It should be noted, however, that not all signals can be analyzed profitably with the sound spectrograph: for example, signals in which the information is primarily contained in the phase relationships among the various components. While phase often affects the details of the patterns obtained with the sound spectrograph it is not designed to present a systematic display of phase relationships. It is, however, a unique accomplishment of the sound spectrograph that time, frequency, and amplitude are shown together in a meaningful display.

Recording and reproduction

The first step in the spectrographic process pictured in Figure 5.1 is the recording of a section of f(t), for future repeated reproduction. The recording medium shown in the figure is a circular loop of magnetic tape, which is rotated at a constant speed past the recording, reproducing, and erasing heads. For the most faithful reproduction of the input signal, the method of recording with an a.c. bias is used.* The biasing signal, $A \cos \omega_b t$, is added to the input signal, f(t), and the resultant $f_o(t)$ is given by

$$f_o(t) = C[f(t) + A \cos \omega_b t], \qquad (5.1)$$

where C and A are arbitrary constants, and ω_b is the biasing frequency. It is necessary to have

$$A \gg |f(t)| \text{ and } \omega_b \gg \omega_m, \qquad (5.2)$$

where ω_m is the maximum significant frequency in f(t).

From the biasing frequency generator an erasing field is also derived, and this field is applied to

---

*For this and other details of the magnetic recording and reproducing process, see reference (26).

the tape just before it passes the recording head. When the switch S_1 is opened, both recording and erasing are stopped. Almost the complete loop of the tape then remains filled with a recorded section of the signal f(t) -- that section which occurred during the last revolution of the tape before S_1 was opened.

The recording can be reproduced at the normal recording speed for purposes of monitoring. As with most instances of magnetic recording, it is necessary to use an equalizing network in the reproducing circuit, mainly to bring the lower, less strongly recorded frequencies into normal amplitude relation with the rest of the spectrum. The biasing frequency ω_b, although residually recorded on the tape, is not reproduced in strength because its wavelength on the tape is shorter than the gap of the reproducing head.

In reproducing for analysis purposes, a great deal of time can be saved by using much higher tape speeds than were used in recording. For example, suppose that the signal section recorded is 2.5 seconds in duration, and that 200 repetitions are required for the complete analysis desired. At normal reproducing speed this would require more than 8 minutes. With a reproducing speed six times that of recording, the time of analysis is reduced to 83 seconds. In general, let the reproducing speed be k times the recording speed. It is obvious that each frequency ω in the input f(t) will become $k\omega$ in the reproduction.

Of course account must be taken of these increased frequencies in the design of the analyzing apparatus. No change need be made in the gap width of the reproducing head, since it is the relation of this width to wavelengths on the tape which is important, and no change in wavelengths is made by the speed-up. It is necessary, however, to design the reproducing head so that there will not be winding losses at the higher frequencies. Also, a new equalizing network must be used for the higher speed.

Let us designate by g(t) the output of the reproducing network, in the speeded-up condition, but keep in mind that a value of t in this case is 1/k times the corresponding t of f(t). The Fourier Transform of g(t) is now $G(k\omega)$, where ω represents a radian frequency in the original signal f(t).

<u>The modulation and filtering processes</u>

As mentioned previously, the frequency analysis is usually performed by a filter which is set at a higher frequency than the highest frequency contained in the signal g(t) to be analyzed; (an exception to this procedure will be proposed later). This filter may have different bandwidths, as desired, but the bandwidth in any case is only a fraction of the range covered by the frequencies in g(t). To transform any particular portion of the spectrum to the frequency band passed by the filter, a local oscillator is used, supplying a carrier signal $V_c \cos \omega_c t$. This carrier signal is combined with g(t) in a balanced modulator, indicated by M in Figure 5.1, and shown in greater detail in Figure 5.2. This is a modulation or heterodyne process. The modulated output, $f_m(t)$, has a spectrum whose frequencies are given by $\omega_c - k\omega$ and $\omega_c + k\omega$, where ω still denotes frequencies in the original input f(t), and k is the speed-up factor. If the bridge of diodes in Figure 5.2 is properly balanced, the carrier ω_c will not appear in $f_m(t)$. In this modulation process the following conditions are assumed to be satisfied.

$$\left.\begin{array}{l} \omega_c \gg k\omega_m, \\ V_c \gg |g(t)|, \\ \text{R is large.} \end{array}\right\} \quad (5.3)$$

As before, ω_m is the maximum significant frequency

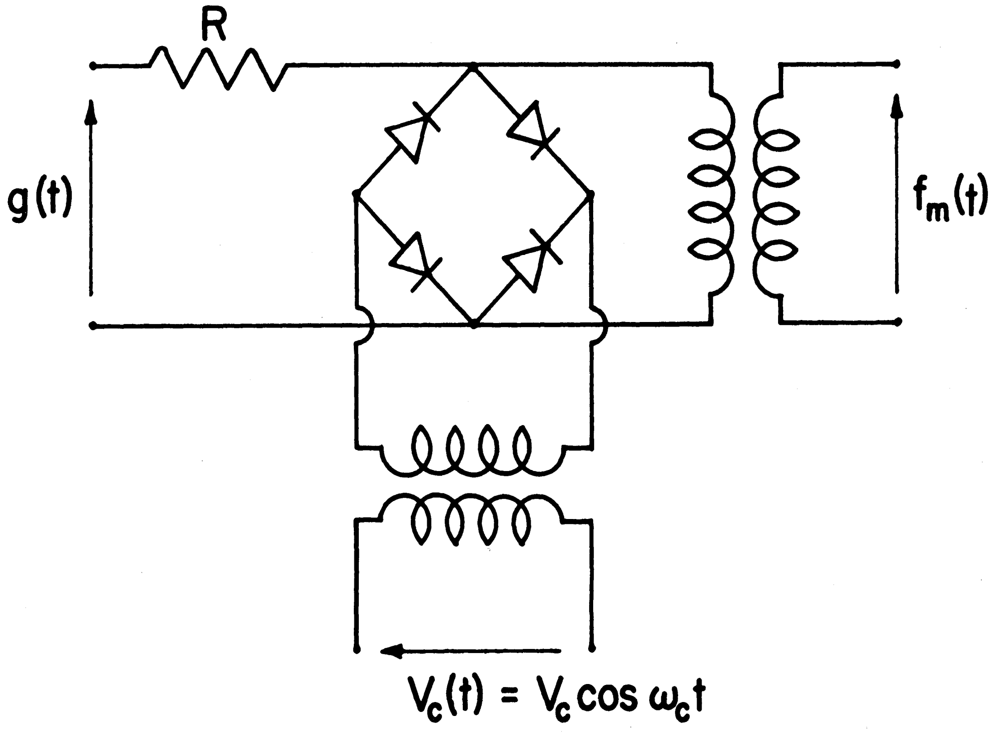


Figure 5.2. A simple diode balanced modulator.

contained in f(t). R refers to the series resistance shown in Figure 5.2.

Of the two sidebands which make up the spectrum of $f_m(t)$, either could be used for analysis. Let us follow through the process for the lower sideband, whose frequencies are given by $\omega_c - k\omega$. It will be assumed that the upper sideband always lies outside the filter passband. To bring different portions of the lower sideband into the filter passband, the carrier frequency ω_c is slowly varied (by means of a mechanically variable capacitor driven by an electric motor), so that there is a significant change in the location of the spectrum between repetitions of the signal. A possible course of variation of ω_c is shown in Figure 5.3, where it starts at a high value ω_h, at time t_1, and its subsequent values are given by

$$\omega_c = \omega_h - \mathcal{L}(t - t_1), \qquad (5.4)$$

where $\mathcal{L}$ is an arbitrary constant. It is also indicated that the variation stops at a low frequency ω_ℓ, which is placed just at the upper edge of the filter passband. Of course the variation could be in the other direction, with ω_h replaced in (5.4) by ω_ℓ, and the minus sign following it changed to plus. It is also possible to use other locations of the filter passband.

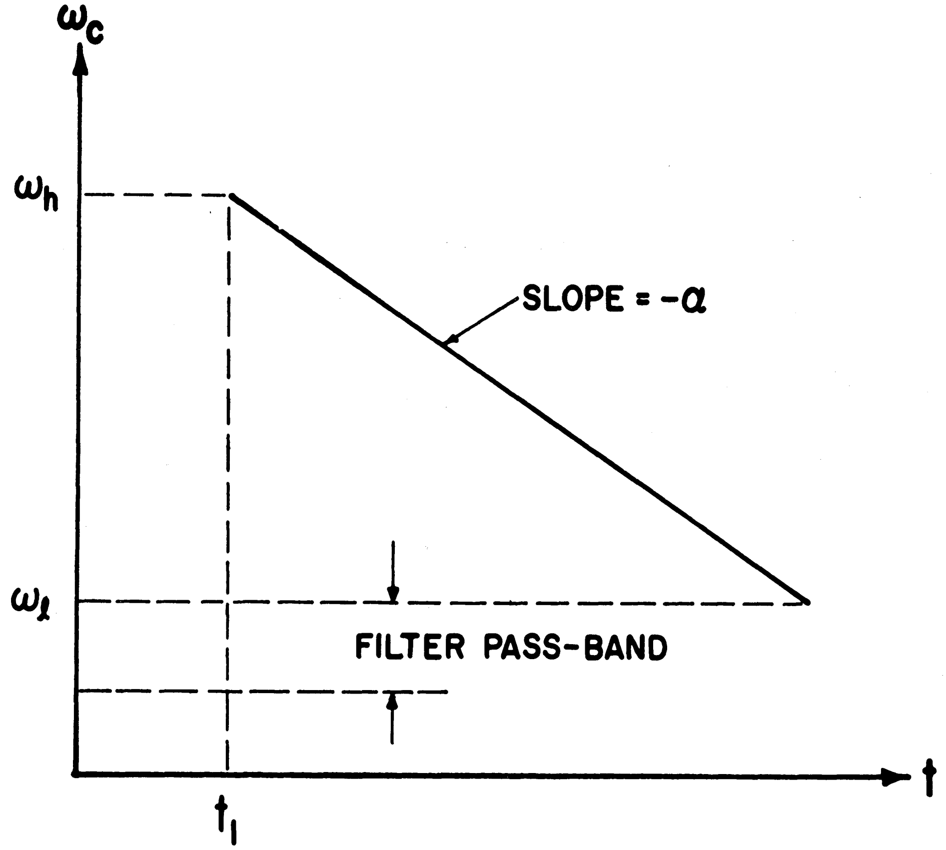


Figure 5.3. The variation of the carrier frequency $\omega_c(t)$ in the time-frequency plane.

Let us call the center frequency of the filter band ω_o, and the filter bandwidth $\Delta\omega$. Then under the conditions shown in Figure 5.3,

$$\omega_\ell = \omega_o + \frac{\Delta\omega}{2}. \tag{5.5}$$

With ω_c at this limit, then frequencies in the original signal lying between zero and $\frac{\Delta\omega}{k}$ radians per second result in lower sideband frequencies which lie inside the filter band. If ω_c were made any lower than the limit given in (5.5), some frequencies in the upper sideband would begin to pass the filter.

The upper limit of ω_c must be set to accommodate the maximum significant frequency ω_m.

$$\omega_h = \omega_o + \frac{\Delta\omega}{2} + k\omega_m = \omega_\ell + k\omega_m. \tag{5.6}$$

When ω_c is at this upper limit, only ω_m corresponds with the lower sideband frequencies which pass the filter. As ω_c is lowered from ω_h, however, other frequencies enter the analysis band, and $k\omega_m$ remains in the band until ω_c reaches $\omega_h - \Delta\omega$.

The speed of variation of ω_c is adjusted in accordance with the tape speed in the reproducing operation. A small change in ω_c takes place during each revolution of the tape, and this means a corresponding change in the values of $k\omega$ which make

$(\omega_c - k\omega)$ lie inside the filter band. For example, this incremental change in ω_c might correspond with 17 cps in the original signal, in which case 200 revolutions of the tape would be required to complete an analysis covering 3400 cps. The correspondence of the filter output at any moment with a particular band of frequencies in the original signal is easily established.

Figure 5.4 further illustrates the relation of the filter with the lower sideband spectrum during an analysis. The position of ω_c in frequency is shown at a fixed moment of time. Also, at the same moment the figure shows the entire lower sideband and a part of the upper sideband, corresponding with an assumed original spectrum. As ω_c moves to the left, in the figure, the sidebands move with it, and one after the other of the lower sideband frequencies pass through the location shown for the band-pass filter. The motion in frequency of ω_c stops at ω_ℓ, which coincides with the upper edge of the filter band. At this point no frequencies from the upper sideband have reached the filter band.

One type of ideal filter would have a transfer characteristic given by

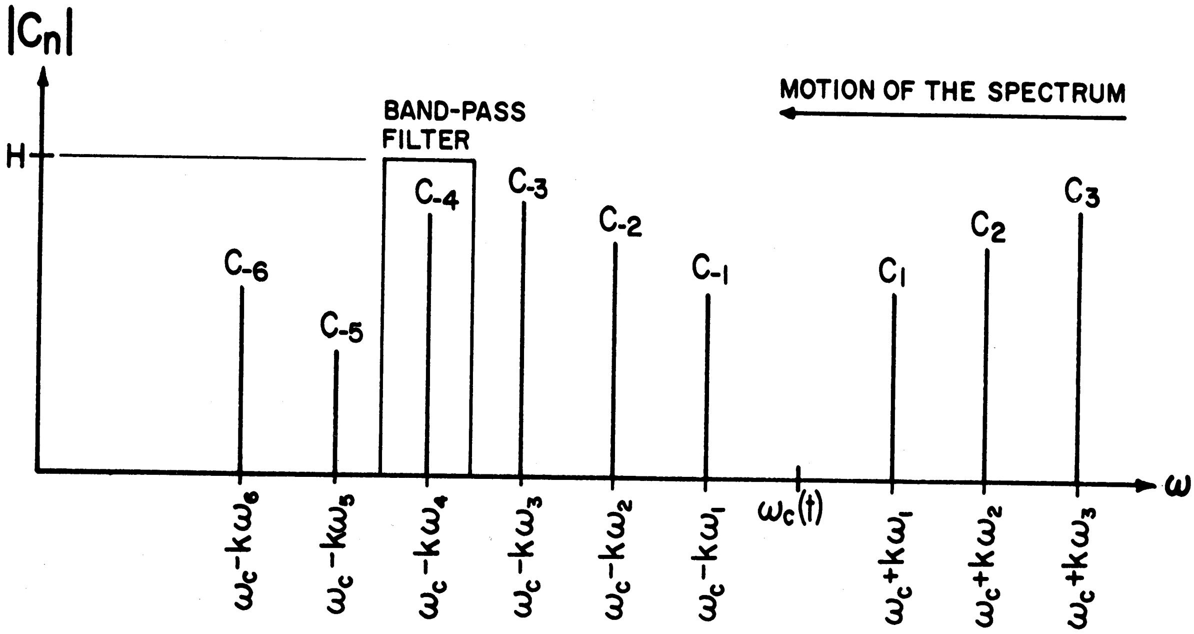


Figure 5.4. A double sideband frequency spectrum scanning across a band-pass filter.

$$H(\omega) = \begin{cases} 0 & , \quad \omega \leq \omega_o - \frac{\Delta\omega}{2} \\ He^{-i(\omega-\omega_o)t_o}, & \omega_o - \frac{\Delta\omega}{2} < \omega < \omega_o + \frac{\Delta\omega}{2} \\ 0 & , \quad \omega_o + \frac{\Delta\omega}{2} \leq \omega \ , \end{cases} \qquad (5.7)$$

where H is a constant amplitude characteristic and t_o is a constant phase characteristic. The conditions (5.7) describe a filter with completely sharp cutoffs, with constant transmission within the passband, and with linear phase shift within the band. It is not possible to build a filter with these ideal characteristics. However, satisfactory approximations to them can be realized in practice. A Gaussian type filter is acceptable, and in some respects has ideal response characteristics; other variations from the ideal band-pass filter will be considered in Chapter 7.

The bandwidth $\Delta\omega$ of the filter is a matter of choice, depending on the degree of frequency discrimination desired in the analysis. For example, $\Delta\omega$ may be made so small that only one harmonic of a voiced speech sound is included in the filter at any one time. With such a narrow filter, each harmonic appears separately in the output. A wider $\Delta\omega$ may include several harmonics at once, so that the output display shows only the relative strengths of different groups of harmonics. For example, regions of energy concentration which are associated with vowel formants and which are sufficiently separated in frequency may be displayed separately, without indicating the separate harmonics in each region of energy concentration.

Although the preceding account of the modulation and filtering process describes the usual arrangement of the sound spectrograph, it is possible to use a lower frequency location of a band-pass analyzing filter if this is desired. In the arrangement to be described, it will be found that the carrier frequency, ω_c, at one end of its range, fails to satisfy either condition (3.7) or (3.8), and the first condition of (5.3). This is made possible by the choice of location of the analyzing filter, and by the fact that only frequencies inside the band of this filter are important in the analyzing process.

The proposed settings are illustrated in Figure 5.5. On the left of the figure is an arrow representing the range of frequencies in the reproduced signal, from $k\omega_1$ to $k\omega_m$. To the right are the lower and upper sidebands, formed by seven different

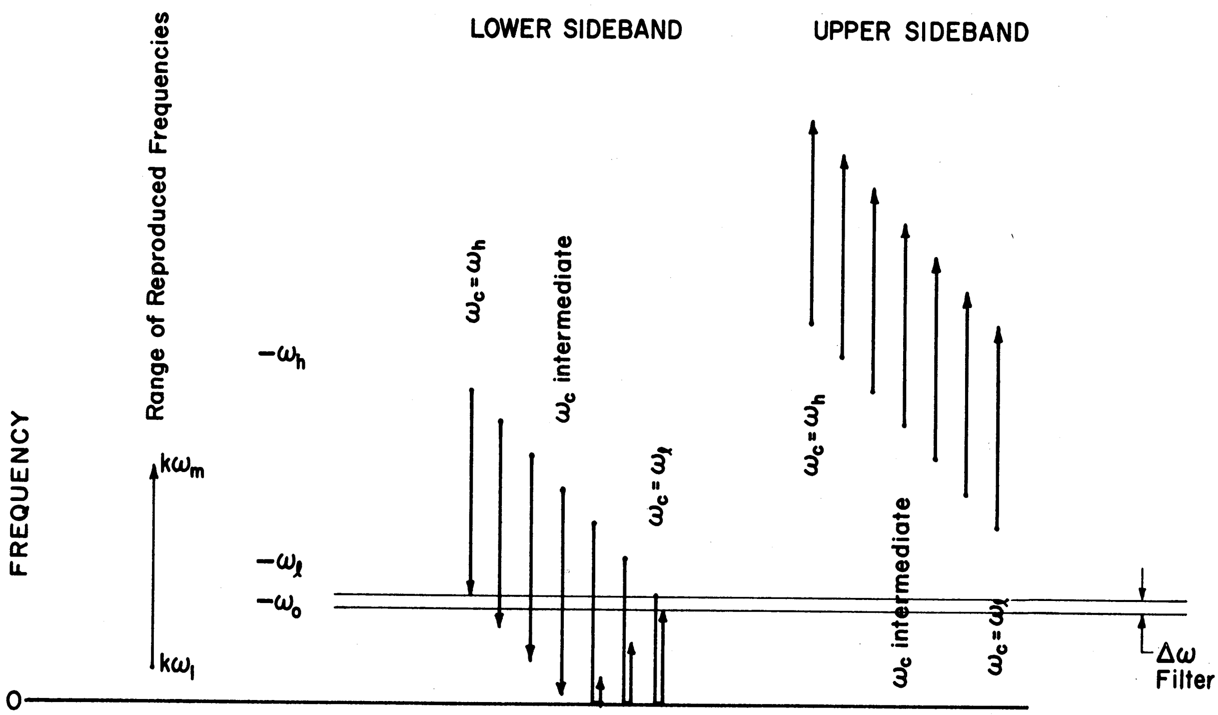


Figure 5.5. The lowest practical setting of a band-pass analyzing filter in a spectrograph.

values of the carrier, ranging from ω_h to ω_ℓ. At each such setting of ω_c, the base of the arrow corresponds with the lowest frequency in the reproduced signal, and the point of the arrow is derived from the highest frequency. The proposed location ω_o of the fixed band-pass filter is also shown, with any desired bandwidth $\Delta\omega$.

The following relation provides the lowest possible value of a band-pass filtering setting,*

$$\omega_o = \frac{k\omega_m - k\omega_1}{2} ,$$

and the filter limits are $\Delta\omega/2$ above and below ω_o. The highest and lowest values of ω_c are to be

$$\omega_h = \frac{1}{2} (3k\omega_m - k\omega_1 + \Delta\omega)$$

and

$$\omega_\ell = \frac{1}{2} (k\omega_m + k\omega_1 + \Delta\omega) .$$

The difference between these is just the range $k\omega_m - k\omega_1$. It will also be seen that ω_ℓ is above the upper limit of the filter by the amount $k\omega_1$, but even if $k\omega_1$ is quite small the balanced modulation will prevent ω_c itself from entering the filter.

---

*A modulator balanced for exclusion of input frequencies from the output is required.

The upper sideband in Figure 5.5 is well above the filter at all points. The lower sideband, for some of the values of ω_c, has components for which $(\omega_c - k\omega)$ is negative. Only the absolute values of these are shown, the arrow being folded back on itself at the zero line. With the arrangement described these folded back portions never get into the filter. This assumes, however, that ω_m is really the highest frequency in the input signal. This point could be assured by the use of a low-pass filter at $k\omega_m$ in the reproduced output circuit.

The three-dimensional spectrogram

It remains to display the analysis performed by the filter (with either wide or narrow bandwidth) in a meaningful form, showing the spectrum of f(t) at each moment of the time sample. For this spectrographic display a special conducting paper is normally used, such as "Teledeltos" or "Timefax", which shows a black trace when current from a moving stylus is passed through the paper to a metal backing. A sheet of this paper is wrapped around a metal cylinder, and a metal stylus of small diameter rests on the paper surface. The cylinder carrying the paper is rotated in synchronism with the tape record. The cylinder and the ring of magnetic tape should be mounted on the same rotating shaft, or else geared together in a 1:1 ratio.

With each revolution of the cylinder, the stylus is automatically moved a few mils in the direction parallel with the cylinder axis, so that each new frequency band is displayed in a slightly new position on the paper. Each revolution of the cylinder covers exactly the same period of time in the original recording, and as the analysis proceeds the display of the frequency being analyzed moves progressively across the paper. When the paper is removed and placed on a plane surface, time and frequency form the geometrical coordinates, while blackness of the record indicates intensity of the sound. An example may be seen in Figure 6.1.

In spectrograms of speech it is desirable that the darkness of the paper record show a direct dependence on the strength of the signal, over a range of at least 35 or 40 db of the input. The range provided by the paper, however, from a barely visible trace to full blackness, covers only about 12 db in the stylus voltage. To make the range of the signal meet that of the paper, a technique of amplitude compression of the signal before it is applied to the stylus is used. The compression should preferably be applied to the heterodyned signal before it reaches the filter, because compression in the filter output would, in effect, degrade the sharpness of the frequency cut-off of the filter.

In the preceding discussion of the three-dimensional spectrogram, it has been assumed that the switch S_2 in Figure 5.1 is in the lowest position, connecting the filter output directly with the "marking amplifier". No rectification of the signal is necessary before application to the stylus. The marking amplifier, however, should have sufficient gain and undistorted output capacity to supply some 200 volts (rms) to the stylus at a current of about 20 milliamperes through the paper.* This 4 watts of power produces approximately full blackness in the paper. The indicated paper impedance of 10,000 ohms becomes higher for lower applied voltages, which produce a lighter mark. It is essentially the power of the filter output which determines the blackness of the paper trace. This power is proportional to the integral, over the frequency band of the filter, of the square of the amplitude at each frequency.

---

*The figures given apply to the "Timefax" paper.

The amplitude sectioner

For the purpose of making accurate measurements of the amplitudes of the various frequencies in the spectrum, at chosen moments in the recorded sample of f(t), the switch S_2 of Figure 5.1 is moved to connect the marking amplifier to the output of the "amplitude display network". The input to this network is the filter output, with any desired bandwidth in the filter.

In using the amplitude sectioner, there are several operations which proceed just as they do in making a 3-dimensional spectrogram. These include repeated reproductions of the recorded signal, at a speed of k times the recording speed; the modulation of the reproduced signal with a slowly varying carrier frequency, ω_c, so that the lower sideband of the modulated signal is gradually swept across the pass band of the filter; and a slow movement of the stylus along the cylinder, so that the stylus occupies a slightly different position at each reproduction of the signal.

Different, however, is the action on the filter output. The amplitude display network of Figure 5.1 is shown in greater detail in Figure 5.6. The

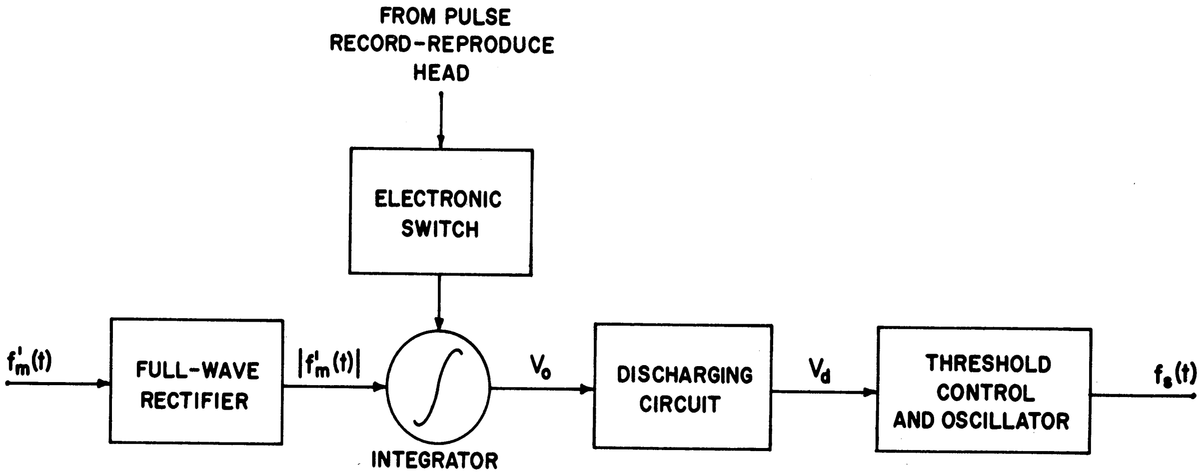


Figure 5.6. Block diagram of an amplitude display network.

filter output, $f_m'(t)$, is first full-wave rectified, and is then led to an integrating capacitance in a charge-discharge arrangement. It is only on discharge that a connection is made to the marking stylus, and this connection is from the capacitance which at that time is disconnected from the filter output.

The moments of discharge are determined by settings which have been made at selected points on the rotating magnetic ring and cylinder. The selection of these points is usually made after a 3-dimensional spectrogram has been produced and examined. In Figure 5.6 it is assumed that, with the cylinder in the position corresponding to a desired moment of the sample, a sharp pulse is recorded on an auxiliary recording track. The reproduction of this pulse at each revolution of the cylinder starts the charge and discharge cycle of the capacitance, by means of electronic switching.

An alternative arrangement is shown in Figure 5.7. A cam is set at the desired point on the cylinder; this cam operates a microswitch which causes a relay to connect the capacitance

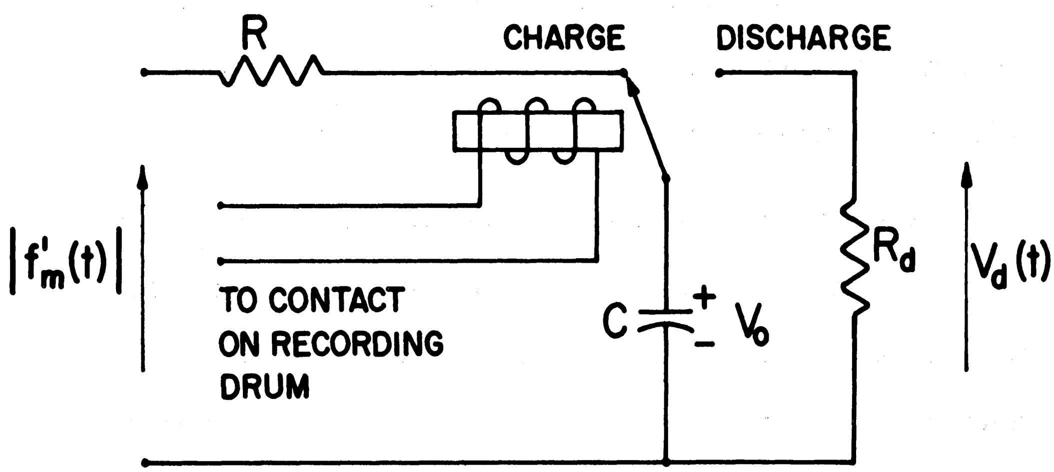


Figure 5.7. An integrating-discharging circuit.

to the "charge" side and to hold it there for a desired time. The relay is then released and the capacitance moves to "discharge". Although this second arrangement is more subject to error due to contact variations, it is perhaps simpler to follow in explaining the process.

The signal applied to the input, in Figure 5.7, is the rectified filter output. This may be designated $|f_m'(t)|$. Let the capacitance be connected to "charge" at time t' and remain there until t". Also suppose that the capacitance has no charge at t'. Then at t" the capacitance will have a potential V_o, given by

$$V_o = \frac{1}{RC} e^{-\frac{t''}{RC}} \int_{t'}^{t''} |f_m'(t)| \, e^{\frac{t}{RC}} \, dt, \qquad (5.8)$$

or, combining the two exponentials,

$$V_o = \frac{1}{RC} \int_{t'}^{t''} |f_m'(t)| \, e^{-\frac{(t''-t)}{RC}} \, dt. \qquad (5.9)$$

Suppose that in the design of the circuit, the product RC is made much larger than the whole charge time, t"-t'.

$$RC \gg t''-t' \qquad (5.10)$$

Then as t varies from t' to t" in the whole course of the integration represented in (5.9), the exponential will be close to unity, and for a close approximation may be removed.

$$V_o \approx \frac{1}{RC} \int_{t'}^{t''} \left| f'_m(t) \right| \, dt \qquad (5.11)$$

That is, if the condition (5.10) holds, the potential across the capacitance at the beginning of discharge is approximately proportional to the integral of the rectified filter output over the time t' to t". In the arrangement assumed in Figure 5.7, these times are set by the position and length of the cam on the rotating cylinder. In the arrangement assumed in Figure 5.6, t' is set by the position of the recorded pulse, while the charge interval t"-t' depends on the setting of a timing circuit attached to the electronic switch.

The integrating time may be made as long as desired for the purpose at hand. It would not normally be made shorter, however, than one period of the lowest frequency in f(t). That is, if ω_1 represents this lowest frequency (which in a periodic signal would be the fundamental), then

$$t''-t' > \frac{2\pi}{k\omega_1} . \qquad (5.12)$$

The V_o of equation (5.11) is a measure of the strength of the filter output at a given moment of f(t). The frequency of the analysis changes with each repetition of the reproduction, but the moment remains the same. Each successive value of V_o is recorded on the electrically sensitive paper as follows.

The first element in the discharging circuit is a resistance, indicated by R_d at the right side of Figure 5.7. The voltage across this resistance, $V_d(t)$, is equal to V_o at the beginning of a discharge. Subsequent values are given by

$$V_d(t) = V_o \, e^{-\frac{t}{R_d C}}, \qquad (5.13)$$

where the zero of t in this equation is the moment when C is switched to discharge.

The "oscillator" in Figure 5.6 is a constant frequency, constant amplitude oscillator, whose output $f_s(t)$ has a sufficiently high voltage (when passed through the marking amplifier) to cause the stylus to mark on the paper. The oscillations can take place, however, only when the input to the "threshold control" is higher than a threshold value, V_{th}. Marking can take place only while $V_d(t)$ is higher than V_{th}. We can then define a discharge time, T_d, during which marking occurs. This T_d is the value of t in equation (5.13) for which $V_d(t)$ has dropped to the value V_{th}.

$$T_d = R_d C(\ln V_o - \ln V_{th}) \qquad (5.14)$$

This equation is plotted in Figure 5.8 for three different values of V_o. The discharge time T_d is indicated in each case. It may be noted that equal increases in $\ln V_o$ produce equal increases in T_d.

For each revolution of the cylinder, marking begins at the moment of discharge, which is at the same point for each revolution. Marking then continues at a constant blackness for the time T_d. Because of the uniform speed of rotation, the length of line drawn is strictly proportional to T_d, which depends on the filter output in the manner given by equations (5.11) and (5.14). With the frequency of analysis changed slightly between repetitions, and the stylus moved slightly to correspond, a display of the signal spectrum at the moment chosen is built up on the paper. Since the length of line drawn is proportional to the logarithm of the ratio of V_o to V_{th}, this length can be calibrated in decibels.

In adjusting the constants for the amplitude sectioner there are some obvious restrictions. One is that the time required for the decay from the

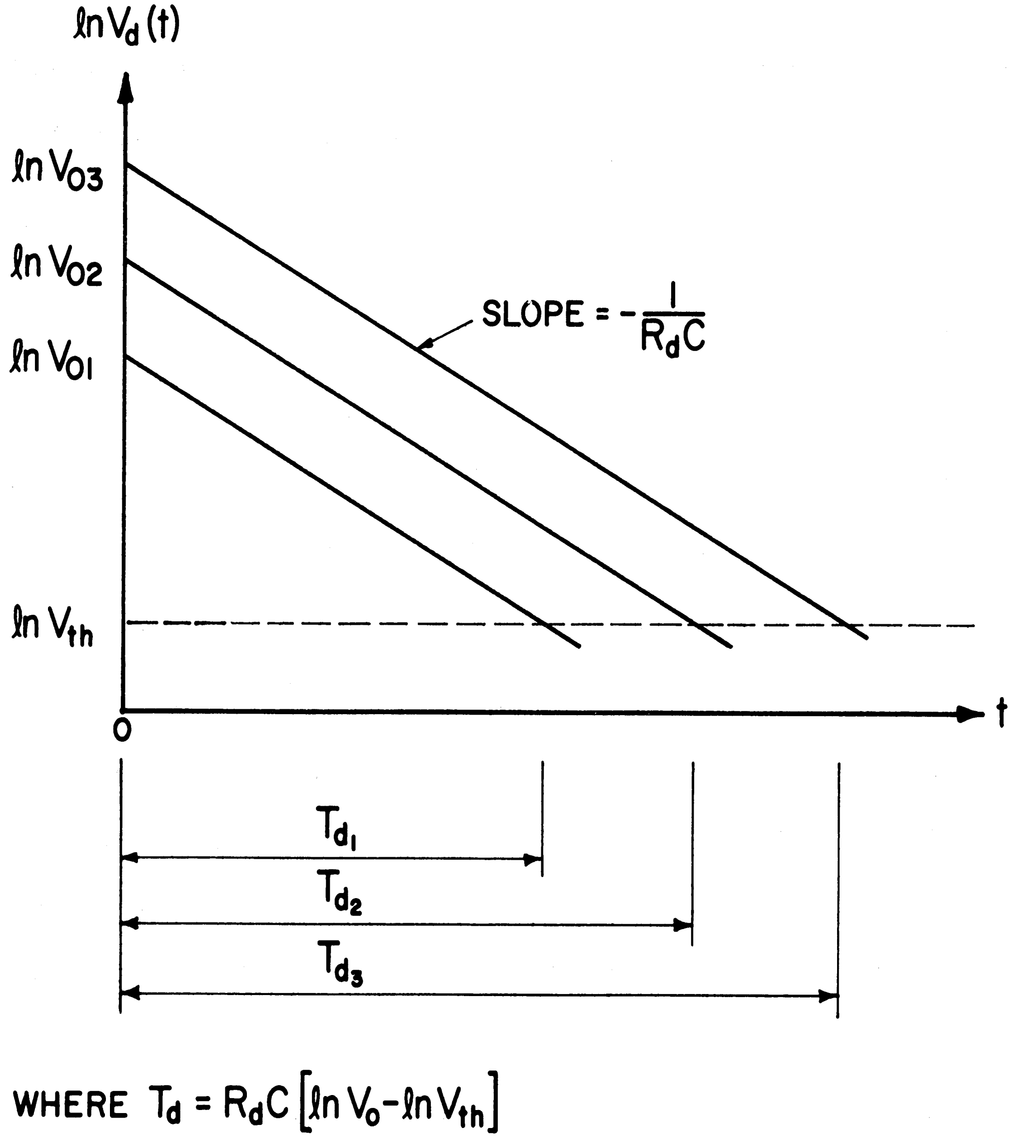


Figure 5.8. Relation of discharge time and logarithm of charge voltage.

highest value of V_o to the threshold V_{th} should be less than the time required for one revolution of the cylinder. Actually, it is usually found that a good aspect ratio of amplitude to frequency in the display is obtained when the highest amplitude is only a fraction of the cylinder circumference. This makes possible the analysis of additional moments in the signal, allowing time for the conversion of the highest amplitudes of the first moment to lines on the display. In fact, it may be possible to choose four or more points in a recorded sample, and to obtain amplitude spectra for all of the selected points in one operation.

One type of amplitude section to be considered is diagrammed in Figure 5.9. In this case the signal f(t) is assumed to be periodic, with a fundamental frequency ω_1, and a number of harmonic frequencies which are multiples of ω_1. It is assumed that a filter is used whose bandwidth is less than the frequency difference of consecutive harmonics in the speeded-up reproduction. That is,

$$\frac{\Delta\omega}{k} < \omega_2 - \omega_1 . \qquad (5.15)$$

The amount by which ω_c is advanced at each repetition, however, is much less than $\Delta\omega$, so that a single

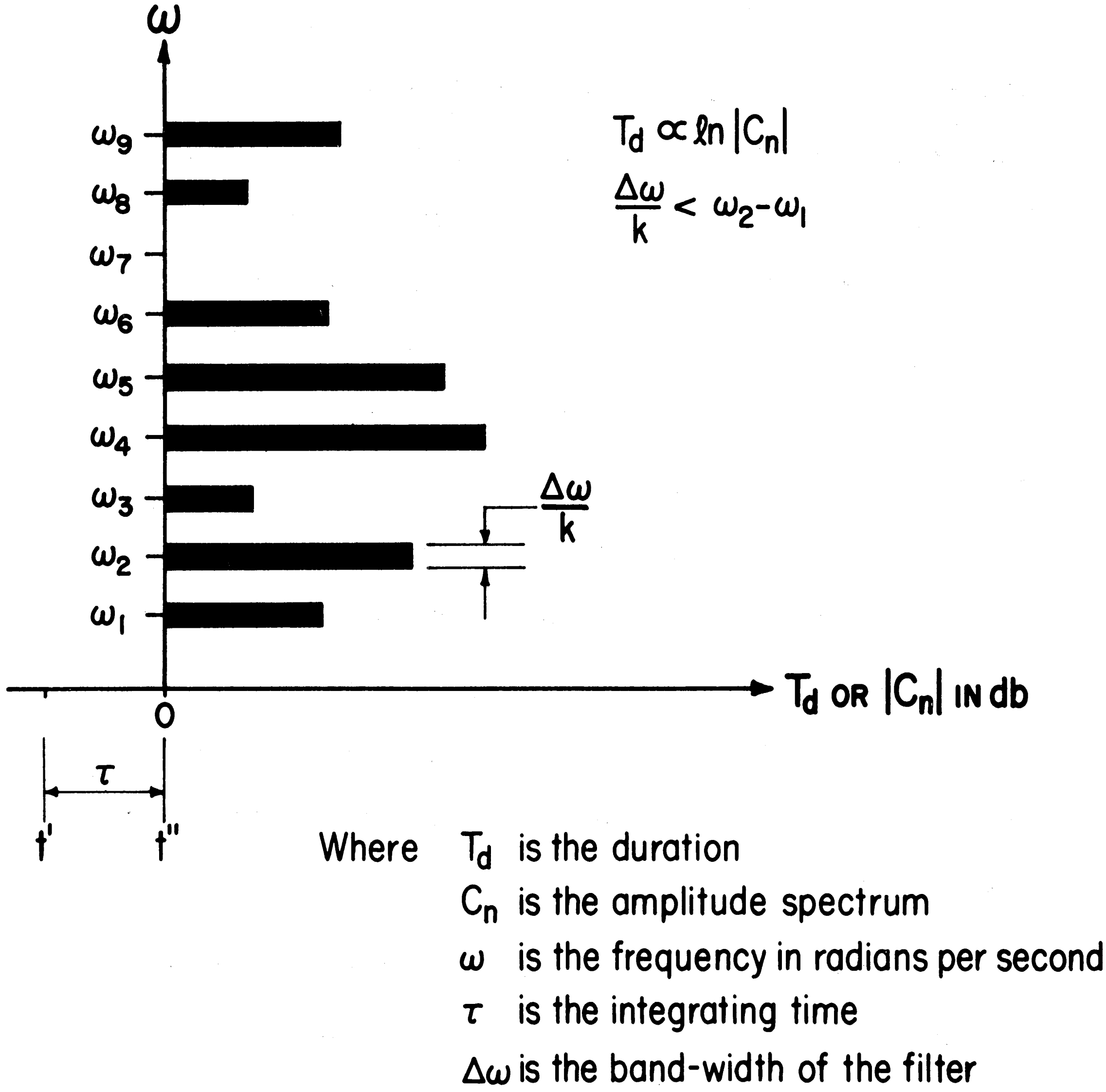


Figure 5.9. Idealized amplitude section, for a filter width less than the harmonic separation.

harmonic remains in the filter band for several repetitions. Correspondingly, the stylus draws several lines side by side, building up one of the rectangles shown in the figure. The square ends of these rectangles assume an ideal flat filter characteristic. With practical filters the ends of the dark areas would be rounded.

Figure 5.10 shows another example of an amplitude section, also constructed diagrammatically. Here a periodic input is again assumed, but the bandwidth of the filter is assumed to be somewhat more than twice the harmonic spacing in the reproduced wave. That is,

$$\Delta\omega > 2k\ (\omega_2 - \omega_1)\ . \qquad (5.16)$$

The assumed spectrum is shown in the upper part of Figure 5.10, together with the filter band which is to be swept across the spectrum. In the lower part of the figure is the outline of the display of the amplitude spectrum. The actual display would be entirely black under this outline. There are always two harmonics inside the filter band, and sometimes three. Individual harmonics show up only as short and narrow peaks superposed on the large dark areas, at points where a third harmonic is present in the

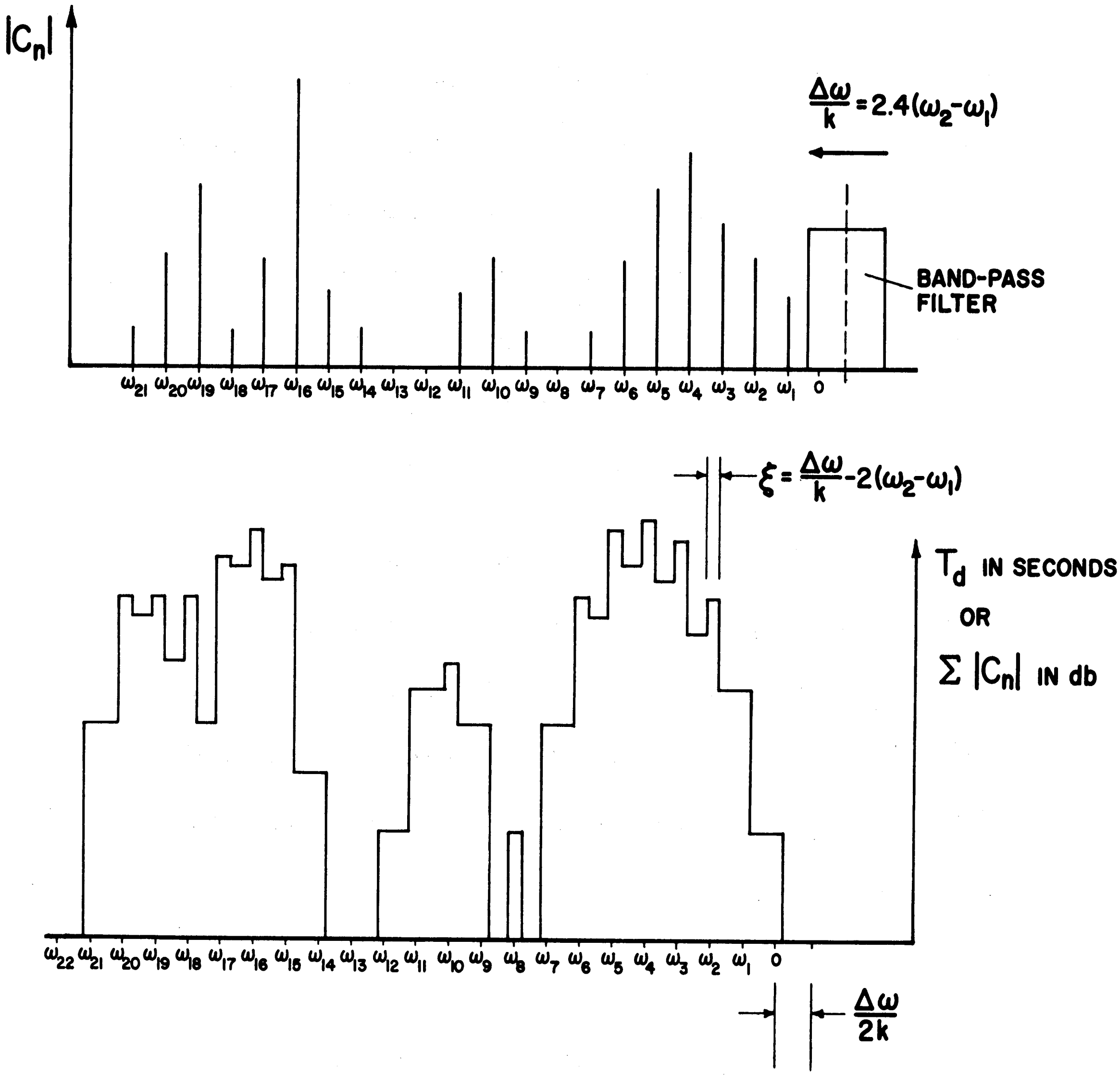


Figure 5.10. Idealized wide-band amplitude section.

band. Groups of harmonics (associated with formants) are indicated by the separated or partially separated dark areas.

The scale of frequency (like that shown at the bottom of Figure 5.10) is somewhat ambiguous, because of the bandwidth of the filter. It is convenient, however, to designate the scale value at any point as that frequency of the input signal which corresponds to the speeded-up and modulated frequency which lies at the center of the filter band, i.e., ω_o,

$$\omega = \frac{1}{k} (\omega_c - \omega_o). \tag{5.17}$$

Or, if equation (5.5) is satisfied,

$$\omega = \frac{1}{k} (\omega_c - \omega_\ell + \frac{\Delta\omega}{2}). \tag{5.18}$$

Chosen in this way, the scale will be the same whether the scanning of the spectrum is from low frequencies to high, or in the reverse direction. However, each harmonic will begin to show its effect on the filter output a distance of $\Delta\omega/2k$ to either side of its position, ω, on the scale.

The sharp rectangular corners of the curve in Figure 5.10 are due to the ideal flat characteristic assumed for the filter band. As with Figure 5.9, these sharp corners disappear with a practical filter.

For example, in Figure 5.11 the same original spectrum is assumed as in Figure 5.10. The filter, however, is assumed to have a trapezoidal characteristic, flat through the central part of the band, but sloping off at the edges. The effective filter bandwidth is wider than in Figure 5.10, although the flat portion is a little shorter than in Figure 5.10. In this case, the smooth curve at the bottom of Figure 5.11 is the expected edge of the dark area of the spectrum section. Roughly the display is a smoothed version of the curve in Figure 5.10.

Examples of experimental amplitude sections will be found in Chapters 6 and 7.

Continuous amplitude display

It is often convenient to have, along with a spectrogram, a continuous record of the unanalyzed average amplitude of the signal sample. Such a record can be made with the spectrographic equipment by the addition of the "continuous display network" shown in Figure 5.1. The record is printed on the electrically sensitive paper, and the time scale is conveniently the same as that of the spectrogram. In this case the switch S_2 of Figure 5.1 is moved to the third point, connecting the continuous display network

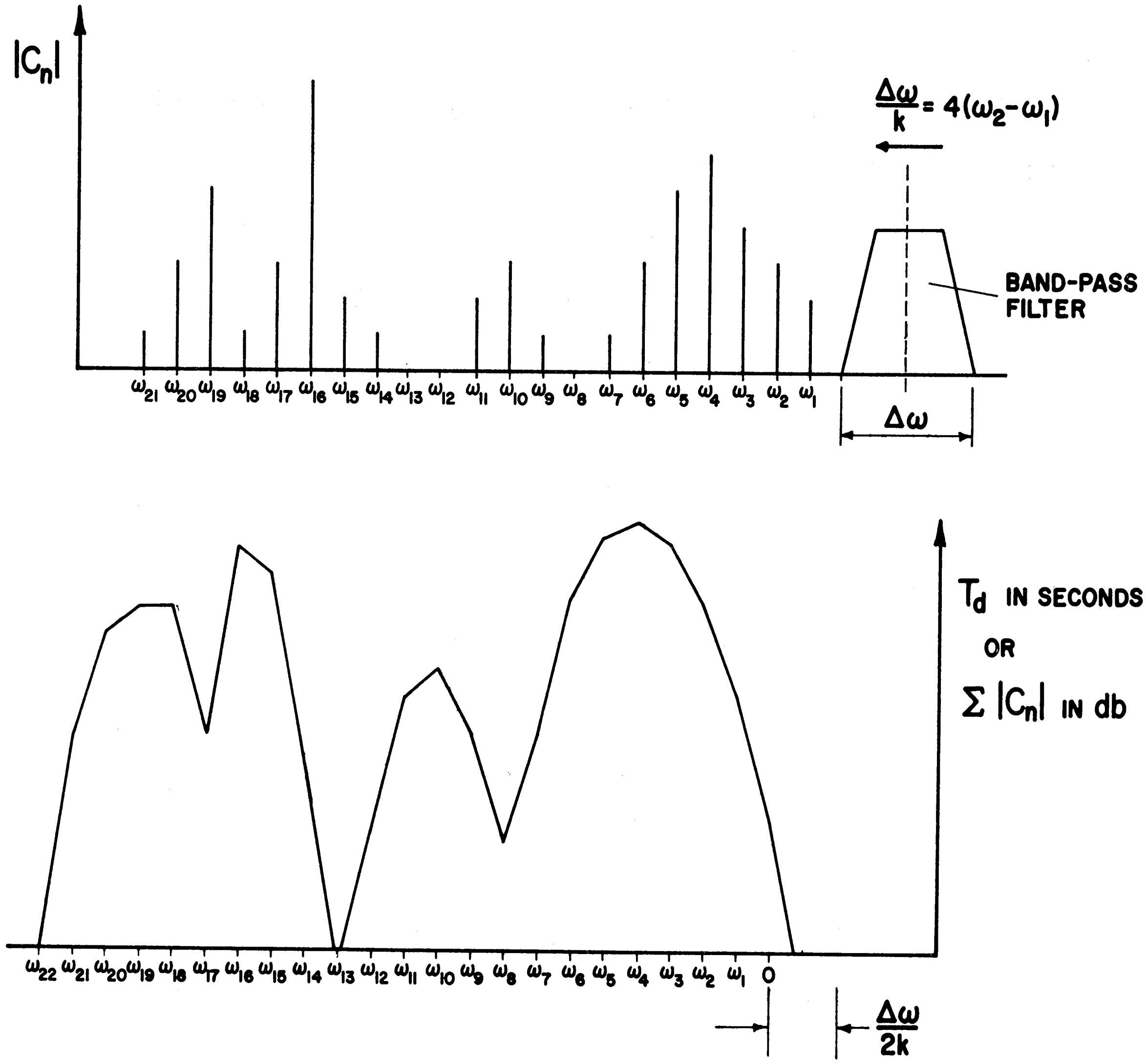


Figure 5.11. Wide-band amplitude section, as smoothed by a practical filter.

to the marking amplifier. The input to this network is the speeded-up reproduction, g(t), before modulation and filtering have taken place. No account need be taken of the speed-up in discussing the amplitude operations, except to remember to use the original time scale on the final record.

In the diagram of Figure 5.12, the first operation shown is that of rectification. Three choices are given for the type of rectification: half-wave on either the positive or the negative side of the signal, or full-wave.

Let us next discuss the output p(t), leaving the intermediate apparatus for later mention. A variable dc biasing voltage, $V_b(t)$, is first added to p(t) and the sum is then impressed on a threshold circuit and control oscillator exactly like those used with the amplitude sectioner. In fact, although separate units are indicated in Figure 5.1, the switching can be so arranged that the same threshold circuit and oscillator can be used for both the amplitude sectioner and the continuous amplitude circuits. It will be remembered that oscillations take place, and cause the stylus to mark on the paper, only when the input to the threshold circuit exceeds a threshold value V_{th}.

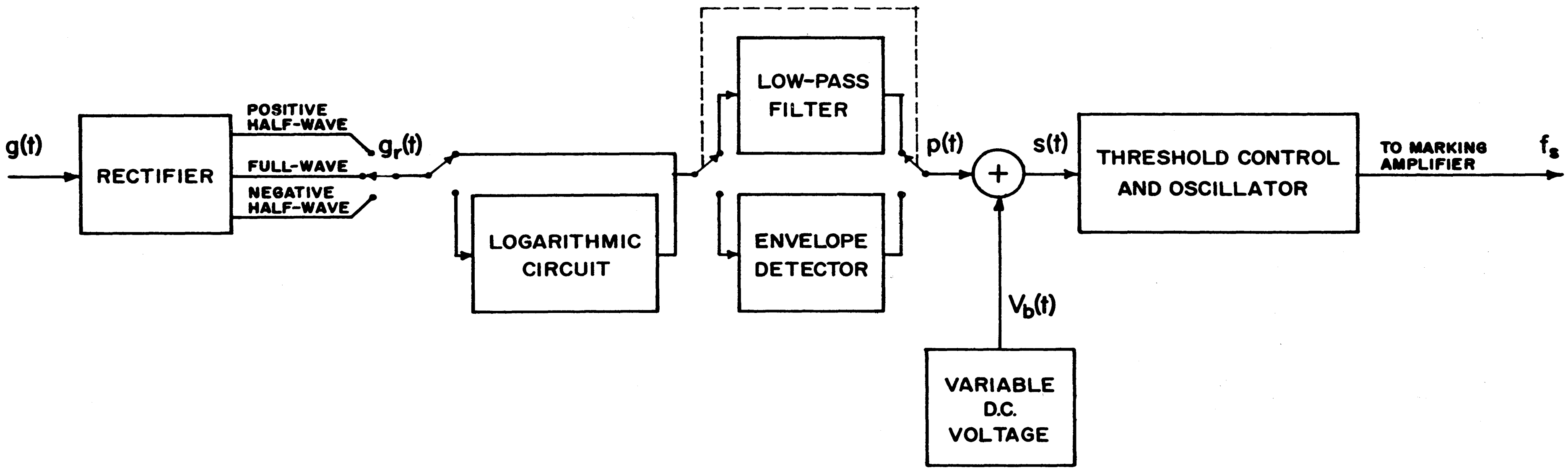


Figure 5.12. A network for continuous amplitude display.

Let s(t) be the sum of p(t) and the bias $V_b(t)$. Then marking takes places when

$$s(t) = p(t) + V_b(t) > V_{th}. \tag{5.19}$$

The bias $V_b(t)$ is slowly varied as the stylus moves along the cylinder. Thus the variation of $V_b(t)$ (and therefore of the amplitude in p(t) which causes marking) takes the place of the variation of frequency ω_c in the case of the 3-dimensional and sectional spectrograms. The same motor may be used for both types of variation. At each repetition of the complete recorded signal a new amplitude becomes the threshold of marking.

A course for the variation of $V_b(t)$ is given in Figure 5.13. Let is start at time t' at the negative value -V', and at subsequent times be given by

$$V_b(t) = \gamma(t-t') - V', \tag{5.20}$$

where γ is a constant. The variation should stop at the value V_{th}. The starting value -V' is so chosen that

$$\text{Max. } p(t) < V_{th} + V'. \tag{5.21}$$

Then if $V_b(t)$ is at this starting value, the condition for marking (5.19) will not be met by any of the amplitudes in p(t). As $V_b(t)$ begins to rise from -V', the

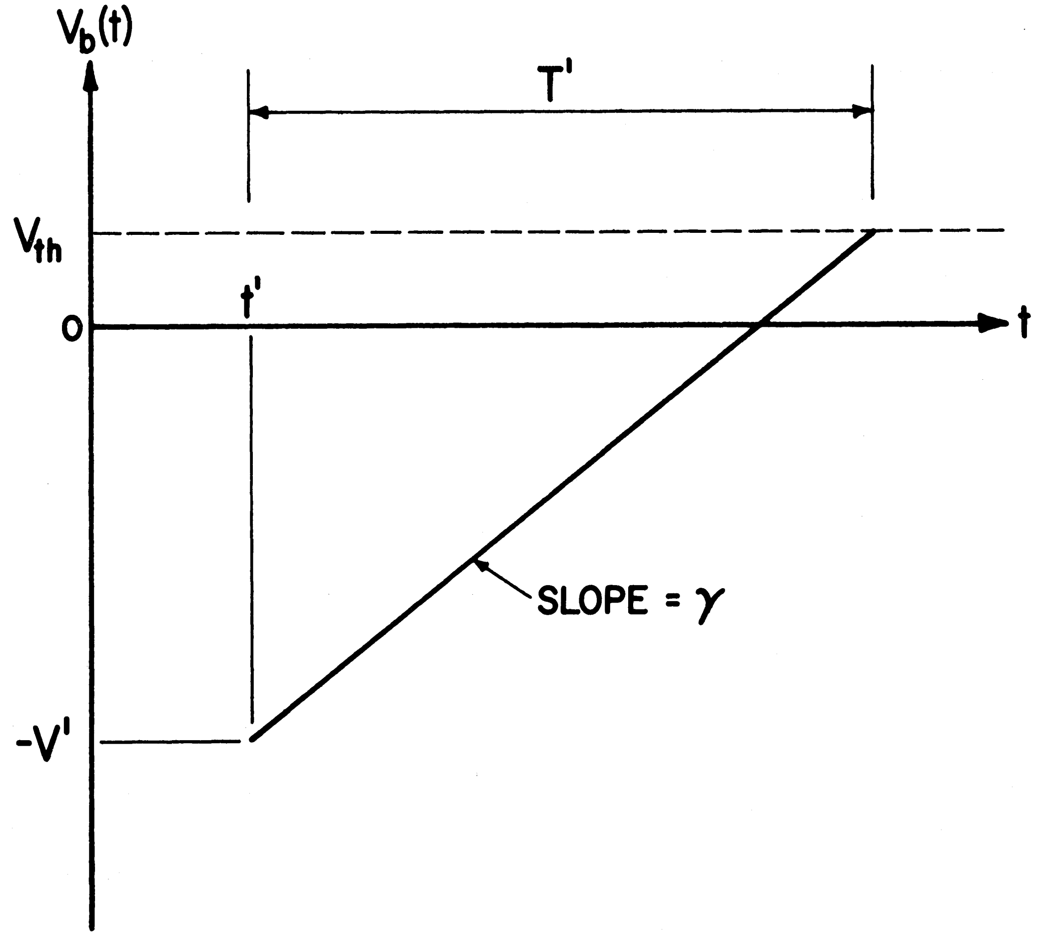


Figure 5.13. Variation of the dc bias $V_b(t)$ with time.

largest amplitudes in p(t) begin to mark. When $V_b(t)$ reaches V_{th}, all amplitudes in p(t) greater than zero satisfy (5.19), and so cause marking. It will be noted that all amplitudes in p(t) are positive, due to the rectification.

a. Direct oscillogram

Except for frequency limitations to be noted later, the continuous amplitude display is equivalent to an oscillogram of the input signal if the following procedure is observed. Let the stylus start at one end of its travel, while p(t) is the direct output of the positive half-wave rectifier, and $V_b(t)$ starts at -V' as described above. The variation of $V_b(t)$ takes place as the stylus moves, and when $V_b(t)$ reaches the value V_{th} a switch is thrown which reverses the direction of variation. At the same time, let p(t) be changed to the negative half-wave rectifier, while the stylus continues to move in the same direction as before. When $V_b(t)$ has moved back past the extreme value, the printed record will be an oscillogram of the signal, with the areas between the curve and the zero line filled in black.

The higher frequency variations will not show in the record, however, because of the compressed time scale which is suitable for the spectrogram type of record. In other words, the time response of the control oscillator and the marking process is not fast enough to follow the upper frequency variations in p(t). The resulting display, however, is not quite the same as would be obtained if the higher frequencies had been removed by filtering before rectification. If this had been done in such a manner as to leave only the fundamental frequency to be rectified, then the positive and negative spikes in a record like that of Figure 6.3 would alternate in time. When all frequencies of the signal are presented to the rectifier, however, peaks of energy are still seen at the fundamental frequency rate, but occur at almost the same time on the positive and negative sides of the wave. This is shown in the record of Figure 6.3.

Also, the amplitudes shown in a record like that of Figure 6.3 are not those of the fundamental frequency alone, but are those of the entire spectrum, peaking at the same rate as the fundamental frequency. The larger peaks on one side of zero in some speech sounds are typical of an oscillogram of undistorted speech, for which the condensation (positive) peaks are often larger than the rarefaction (negative) peaks.

b. Low-pass filtering (averaging detector)

The output of the rectifier, which has been labelled $g_r(t)$ in Figure 5.12, has a large dc component. In the case of a periodic g(t) and full-wave rectification, the dc component is given by equation (2.4),

$$\text{Av. } g_r(t) = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left| g(t) \right| dt, \qquad (5.22)$$

where T is the period of the complex wave. With an input wave like speech, this dc component will fluctuate at the syllabic rate.

Suppose that it is desired that the amplitude display show the syllabic fluctuations of the dc component, but not the ac components of $g_r(t)$. A low-pass filter of simple form, like that of Figure 5.14, can be used for this purpose. The output of the filter then becomes the p(t) of Figure 5.12. The conditions for excluding all ac components is that the R and C of the filter must obey the inequality

$$RC \gg T_1 = \frac{2\pi}{\omega_1}, \qquad (5.23)$$

where T_1 is the period and ω_1 the radian frequency of the lowest frequency component of g(t). To permit the syllabic fluctuation, however, we must have

$$RC \ll \tau', \qquad (5.24)$$

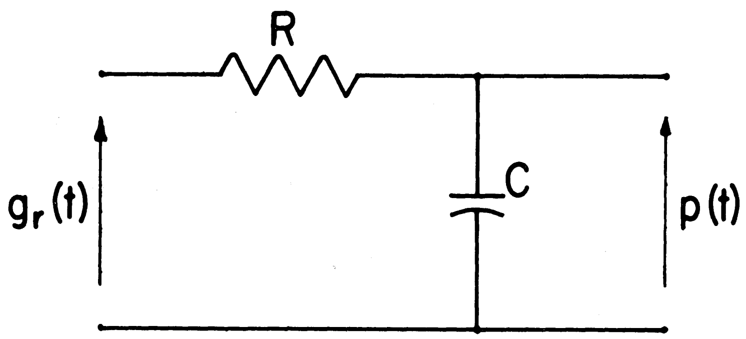


Figure 5.14. An averaging detector (low-pass filter).

where τ' is the shortest interval between the syllabic variations. In speech there is usually enough difference between T_1 and τ' that (5.23) and (5.24) can both be satisfied. This point will be discussed further in the next section.

If RC is made smaller than is indicated by (5.23), the lowest ac components of $g_r(t)$ will appear in p(t), followed by higher components as RC is further decreased. These components will be added to the dc component in the amplitude display, as far as the time response of the control oscillator and the marking process will permit.

c. Envelope detector

An alternative circuit for processing $g_r(t)$ is shown in Figure 5.12. It is called an envelope detector, and it differs from the averaging detector mainly in the relationship between the R and C of the filter circuit. Figure 5.15 applies to this envelope detector.

It is the nature of a rectifier circuit to have a low impedance in the forward direction but a very high impedance in the reverse direction. Thus the C of Figure 5.15 quickly takes on a positive charge, but tends to hold it as the voltage decreases.

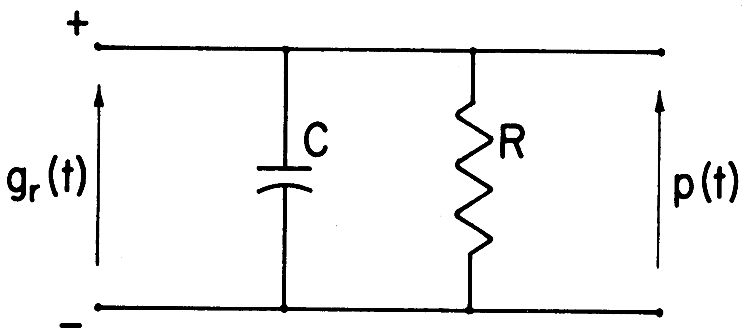


Figure 5.15. An envelope detector.

Assuming that the impedance which p(t) works into is also high, the resistance R is the chief control of the discharge of C. Thus the filter has a short time constant for increasing voltage, but a longer one represented by RC for decreasing voltage.

To follow the envelope of the rectified wave, the time constant RC should again satisfy the inequalities (5.23) and (5.24). It was indicated in the preceding section that for speech both these conditions could usually be satisfied. However, speech varies from person to person and from language to language, so that sometimes compromise values of RC must be chosen. The effects on the amplitude display of changing RC will be shown by examples in Chapter 7.

d. Logarithmic circuit

Instead of applying the rectifier output directly to the averaging or envelope-detecting filter, it may be desirable to interpose the logarithmic circuit shown in Figure 5.12. The purpose of this circuit is to make the voltages applied to the filter proportional to the logarithms of those in the rectifier output. An ideal arrangement would be given by the equations

$$g_{r\ell}(t) = \ell n\, \beta\, g_r(t) = \ell n\, \beta + \ell n\, g_r(t) \qquad (5.25)$$

or,

$$e^{g_{r\ell}(t)} = \beta\, g_r(t), \qquad (5.26)$$

where $g_{r\ell}(t)$ is the output of the logarithmic circuit and β is an arbitrary constant.

One possible circuit by which the above relation can be approximated is shown in Figure 5.16. The circuit consists of a series resistance R_s, and a number of shunts each containing a resistance, a diode, and a voltage source. If the sources, E_1, E_2, E_3 and so on are arranged in increasing order, the smallest voltages in $g_r(t)$ will cause current to flow only in the shunt R_o. The diodes in the other shunts prevent reverse currents from the voltage sources. Increasing voltages from $g_r(t)$ will first exceed E_1, then E_2, and so on, causing currents to flow in these shunts.

An exact plot of equation (5.25) is given in the solid curve of Figure 5.17. The real output of the logarithmic circuit is a broken line starting from zero and changing slope as each E is passed. The first section of this characteristic can be made to coincide with the desired curve at a point corresponding to g_1, the minimum expected value of

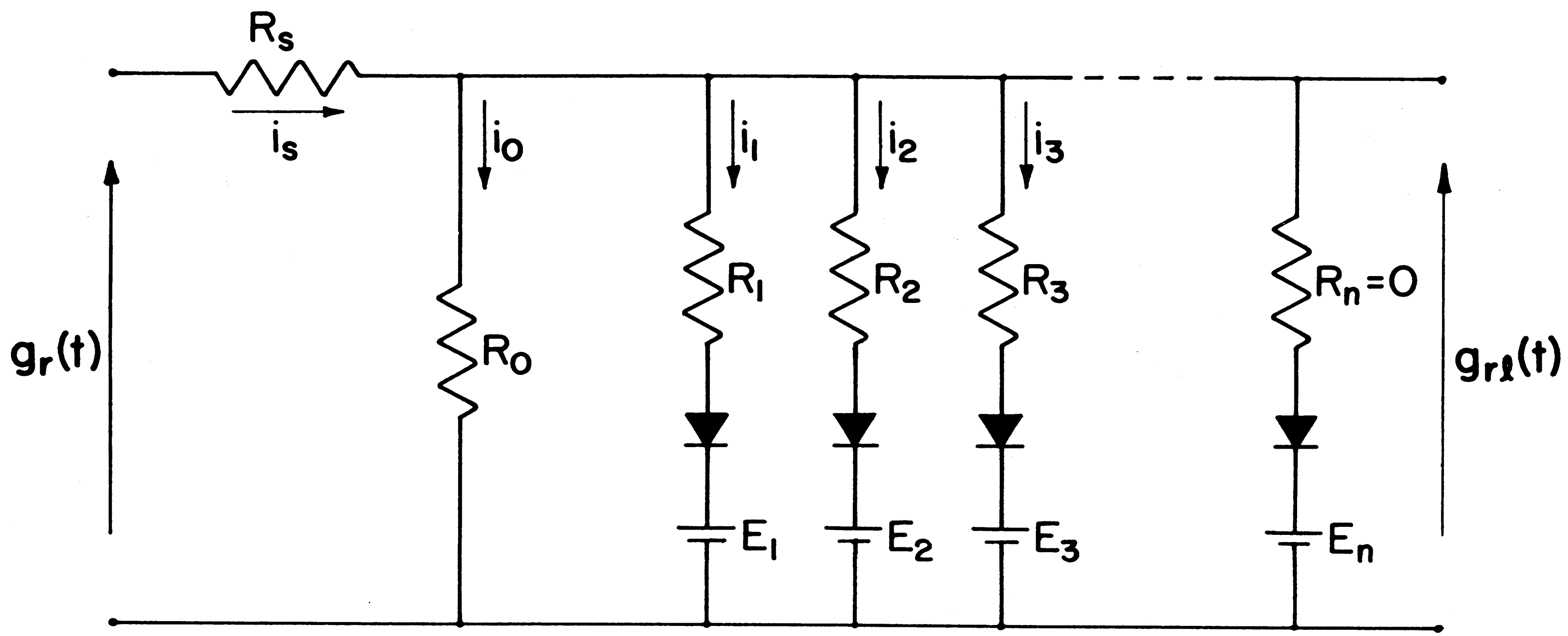


Figure 5.16. A logarithmic circuit.

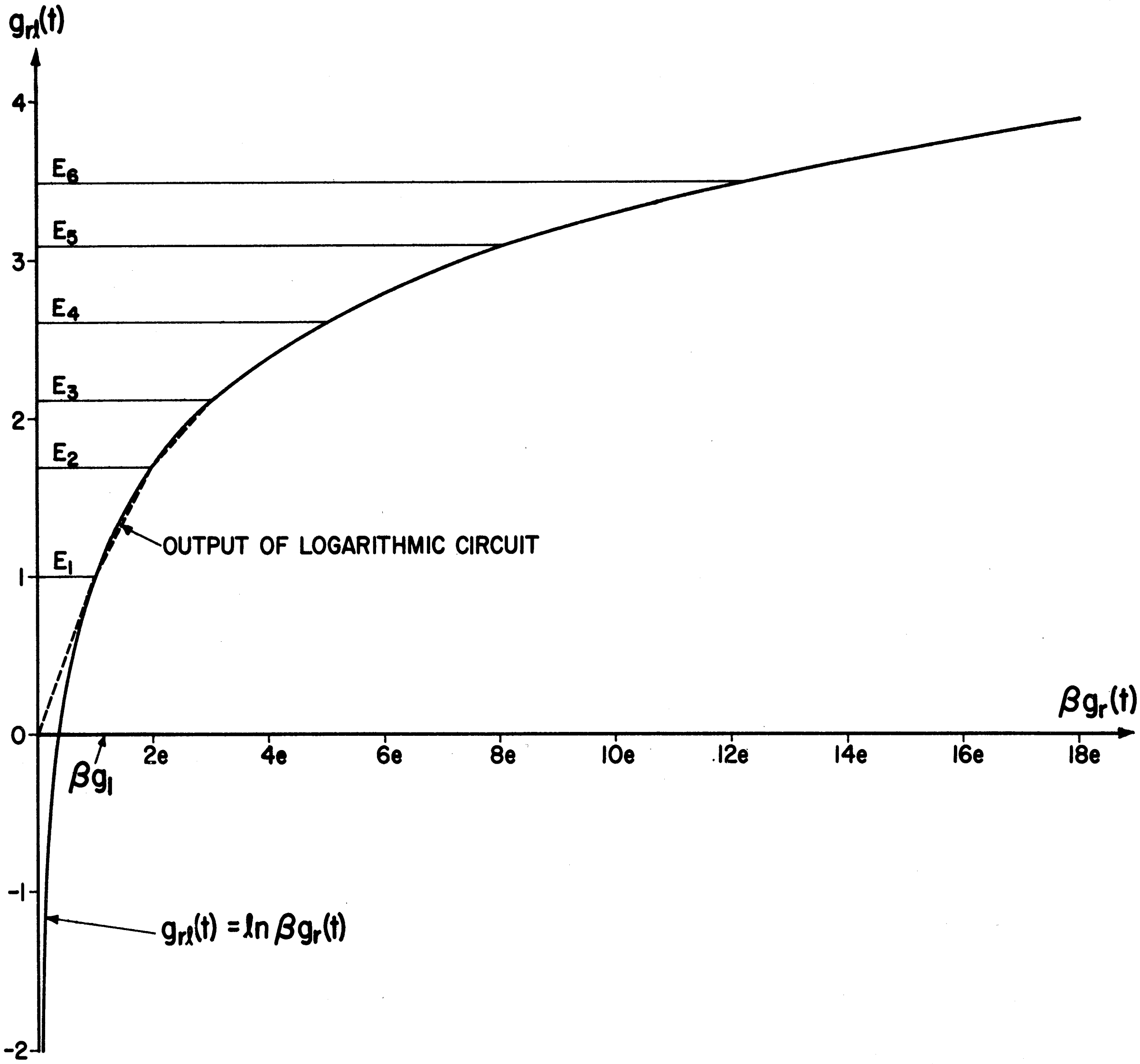


Figure 5.17. Response of a logarithmic circuit.

$g_r(t)$. R_s and R_o are first selected, with a moderately large (10 or more) ratio of R_s/R_o. E_1 is then given by

$$E_1 = \frac{g_1}{\frac{R_s}{R_o} + 1}. \qquad (5.27)$$

This relation states that E_1 is equal to the voltage across R_o which is produced by the g_1 value of $g_r(t)$. To make it also satisfy the logarithmic relation (5.25), we choose β from (5.26),

$$\beta = \frac{e^{E_1}}{g_1}. \qquad (5.28)$$

The subsequent sources E_2, E_3, and so on are chosen to keep the broken line of Figure 5.17 as close as desired to the logarithmic curve. The shunt resistances R_1, R_2, and so on are then calculated in succession, to make the upper end of each line section coincide with the curve. The process is continued until the highest expected value of $g_r(t)$ is exceeded. A relation for finding the particular shunt resistance R_k, from the E's and the previously found shunts, is given by

$$\frac{R_s}{R_k} = \frac{\frac{1}{\beta}(e^{E_{k+1}} - e^{E_k})}{E_{k+1} - E_k} - \left(1 + \frac{R_s}{R_o} + \frac{R_s}{R_1} + \cdots + \frac{R_s}{R_{k-1}}\right). \qquad (5.29)$$

In the above discussion it has been assumed that the output of the logarithmic circuit works into an impedance which is high compared with R_o.

Equivalent low-pass filtering

It has been indicated earlier that the lowest practical setting of a band-pass analyzing filter is at about half the maximum frequency contained in the reproduced signal being analyzed. If the filter were set lower than this, then at some point in the variation of the carrier frequency ω_c, components arising from entirely different parts of the signal spectrum would lie inside the filter band simultaneously.

If the analyzing filter is made low-pass, however, with a cut-off set at half the bandwidth of analysis desired, then the positive and negative frequencies passed by the filter arise from one continuous band in the signal spectrum. This process needs only one modulation,* with the carrier frequency varying from $k\omega_m + \Delta\omega/2$ down to zero, where $k\omega_m$ is the maximum frequency in the reproduced signal and $\Delta\omega$ is the bandwidth of analysis. The result is that the lower sideband of the modulation is swept past the filter band in an orderly manner,

*The footnote on p. 66 applies here also.

and the upper sideband frequencies lie outside the filter band except when $\omega_c < \Delta\omega/2$.

Let us suppose, however, that a variable carrier frequency already exists in the range necessary for modulation to a band-pass filter, but that it is found desirable to provide low-pass filtering because certain response characteristics are more easily achieved with low-pass filters than with their higher frequency band-pass equivalents. The necessity of supplying a variable carrier frequency in a new range can be avoided by a second modulation applied to the lower sideband resulting from the initial modulation process. For the second modulation a fixed carrier can be used, which is conveniently set at the frequency ω_ℓ, the lowest point in the variation of the first carrier, ω_c. Two sidebands are produced, their frequencies given by $\omega_\ell + \omega_c(t) - k\omega$ and $\omega_\ell - \omega_c(t) + k\omega$.

Assuming that all values of ω_c, including ω_ℓ, are higher than the maximum $k\omega_m$ of the reproduced frequencies $k\omega$, then the upper sideband of the second modulation is entirely outside the low frequency range of the low-pass filter. The lower sideband of the second modulation is the one which will be analyzed, containing the frequencies

$|\omega_\ell - \omega_c(t) + k\omega|$. Absolute values are indicated, since the quantity $\omega_\ell - \omega_c + k\omega$ can have either positive or negative values, depending on the values of ω_c and $k\omega$.

Figure 5.18 shows this doubly modulated signal, $\ell(t)$, as the input to the low-pass filter. The result of the first modulation, $f_m(t)$, is shown at the left in the figure as applied to a "lower sideband filter" before the second modulation, since it is only the lower sideband which is useful here. Actually this filter could be omitted, since the upper sideband of the first modulation would produce only frequencies outside the low-pass filter range in the second modulation.

In Figure 5.19, a set of signal frequencies from ω_1 to ω_m is assumed. The lower sideband of the first modulation, arising from these signal frequencies and a particular value of ω_c, is shown toward the right side of the figure. Also marked are ω_h and ω_ℓ, the extreme values of ω_c. At the left of Figure 5.19 the corresponding lower sideband of the second modulation is shown. Two of the frequencies in this sample spectrum are negative, and this fact is indicated in the diagram by placing the marks for these frequencies below the horizontal line.

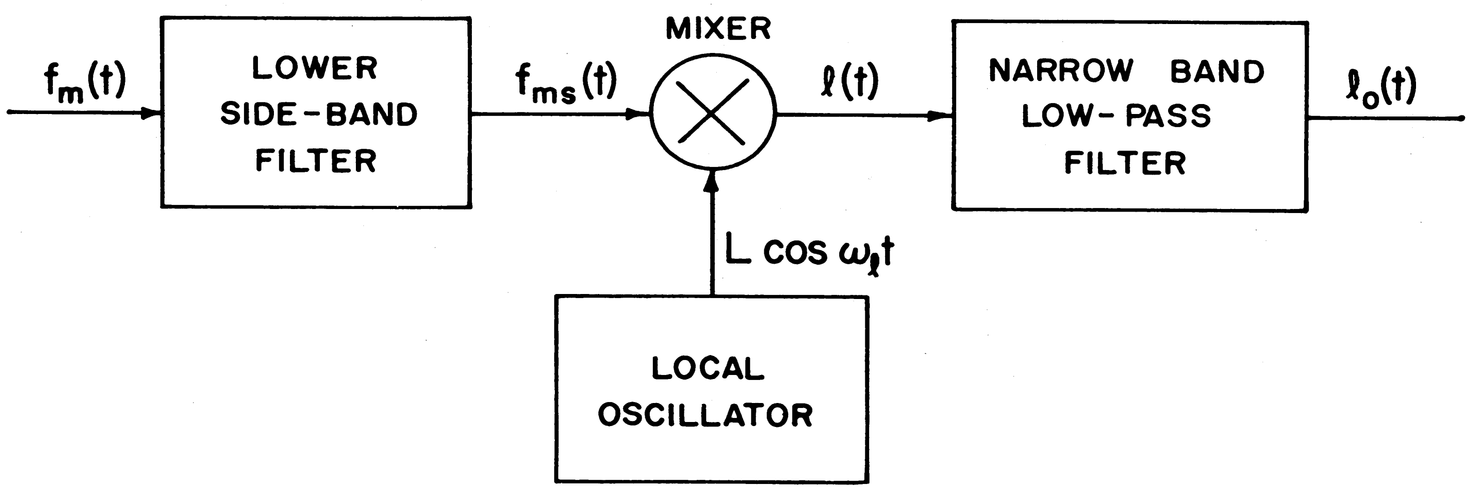


Figure 5.18. Double modulation for equivalent low-pass filtering.

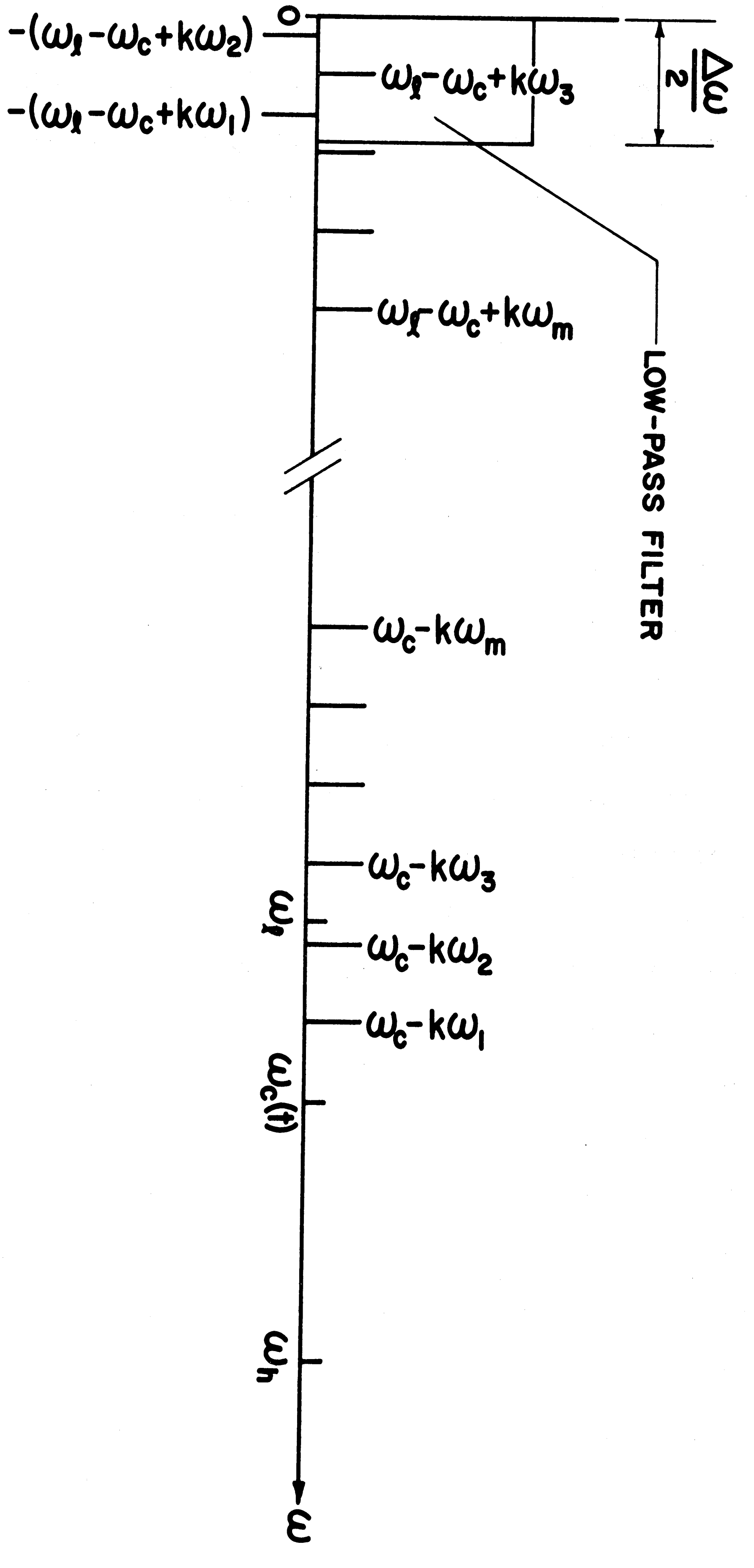


Figure 5.19. Positions of singly and doubly modulated signals on the frequency scale.

It can be seen which frequencies pass through the band of the low-pass filter as $\omega_c(t)$ is varied from ω_h to ω_ℓ. Note from equation (5.6) that

$$\omega_h - \omega_\ell = k\omega_m. \qquad (5.30)$$

Then at the start, when $\omega_c(t)$ equal ω_h, the component $|\omega_\ell - \omega_c + k\omega_m|$ is just at the zero mark in Figure 5.19, with the other frequencies strung out to the right of it, but below the horizontal line since they are negative. As $\omega_c(t)$ moves down in frequency from ω_h, the frequency due to $k\omega_m$ moves to the upper side of the line (becomes positive) and then to the right, and eventually out of the filter band. The other frequencies of the spectrum follow it. At the end, when $\omega_c(t)$ equals ω_ℓ, all of the doubly modulated frequencies are positive, and only those less than $\Delta\omega/2$ are still inside the filter band. Over most of the range of the original spectrum, the doubly modulated frequencies pass through the filter band twice. Only those for which $k\omega < \Delta\omega/2$ and those for which $(k\omega_m - k\omega) < \Delta\omega/2$ pass through the filter band less than two full times, and these pass through the band at least once.

At a given moment, the frequency at the zero value is that for which

$$k\omega = \omega_c - \omega_\ell. \qquad (5.31)$$

Frequencies less than this by amounts up to $\Delta\omega/2$ lie inside the filter band on the negative side, while those greater than this by amounts up to $\Delta\omega/2$ are inside the band on the positive side. Thus the total range of frequencies passed by the filter at any moment is $\Delta\omega$, the same as the filter bandwidth in the singly modulated case.

Unfortunately, there are two circumstances in which the technique described above, for a second modulation into a low-pass band, can fail to produce correct marking on the spectrogram for the actual energy present in the input signal. One of these arises from the fact that the changing frequency in the second modulation, due to any frequency ω in the input signal, must pass through a zero value. This occurs for that value of ω_c which will satisfy (5.31) for the particular ω being considered. Let us suppose that the output of the low-pass filter is full-wave rectified before being impressed on the marking amplifier. It can then be represented as in the upper curve of Figure 5.21 (discussed later), where the reductions to zero amplitude occur at intervals of half the period of the wave passing the filter. For moderate frequencies in the filter these amplitude fluctuations will be averaged out in the marking on the spectrogram, or can be integrated by the eye in viewing. But for frequencies near zero, there can be prolonged intervals of no marking.

The other circumstance that can produce incorrect marking is that two different original frequencies (e.g. two adjacent harmonics) may produce the same frequency in the low-pass filter. This can be seen from the explanation of how the frequencies move through the filter as ω_c varies, first on the negative side then on the positive (see Figure 5.19). Thus one component may be on the negative side and another positive at the same absolute value. This will happen for original frequencies ω_a and ω_b when ω_c reaches the value

$$\omega_c = \omega_\ell + \frac{k}{2} (\omega_a + \omega_b). \qquad (5.32)$$

The common resultant frequency will lie inside the filter band if

$$|\omega_a - \omega_b| < \frac{\Delta\omega}{k}. \qquad (5.33)$$

This is more likely to happen for wide band analysis than for narrow.

The effect of the coincidence of frequencies on the filter output depends on the amplitudes and relative phases of ω_a and ω_b. The effect is on the amplitude of the resultant frequency wave, which may vary from the sum of amplitudes of the two components down to zero. Thus at the moment of coincidence there would usually be a change in filter output, and this change might be a reduction to zero.

The effect of the first circumstance discussed above may not be entirely harmful, since the reduced marking may indicate the center position on the spectrogram of the frequency of interest. Both circumstances produce very local effects which are not greatly detrimental to the spectrographic display, even in the case of amplitude sections. There are various ways in which the effects in the two circumstances can be reduced or eliminated. One rather direct way in which the effects can be largely mitigated is to use two or more second modulating waves, of equal frequencies but with different phase constants. Both (or all) are applied to the lower sideband of the first modulation, and the separate products are then applied to separate low-pass filters. The outputs of these are full-wave rectified and then combined in such a way that at any moment the largest of them supplies the marking signal. Figure 5.20 pictures such a circuit, using four separate phase-shifted modulations. It will be noted in this circuit that each combination of direct and phase-inverted signals, through diodes, is actually a full-wave rectification.

The use of only two second modulations, if properly distinguished in phase, will insure that a cancellation of amplitudes of coinciding frequencies from two separate original frequencies will not occur. It can be shown that if such a cancellation takes place in one of the filters, it cannot also occur in the other filter.*

For a single component at zero frequency, the effect of the multiple modulations can be seen from the curves of Figure 5.21. The lowest of these curves is the combined filter output, obtained by using at all points the largest of the four phase-shifted curves. Thus the large variations in one of the separate filters are reduced to a small fluctuation in the combination. In this case, with a shift of $\pi/4$ radians between separate curves, the fluctuation of the combined curve is from .924 $|c_n|$ to 1.000 $|c_n|$, and the average value is .975 $|c_n|$. If only two filters were used, with a $\pi/2$ phase difference, the fluctuation would be

---

*See "Appendix" for details on frequency interactions.

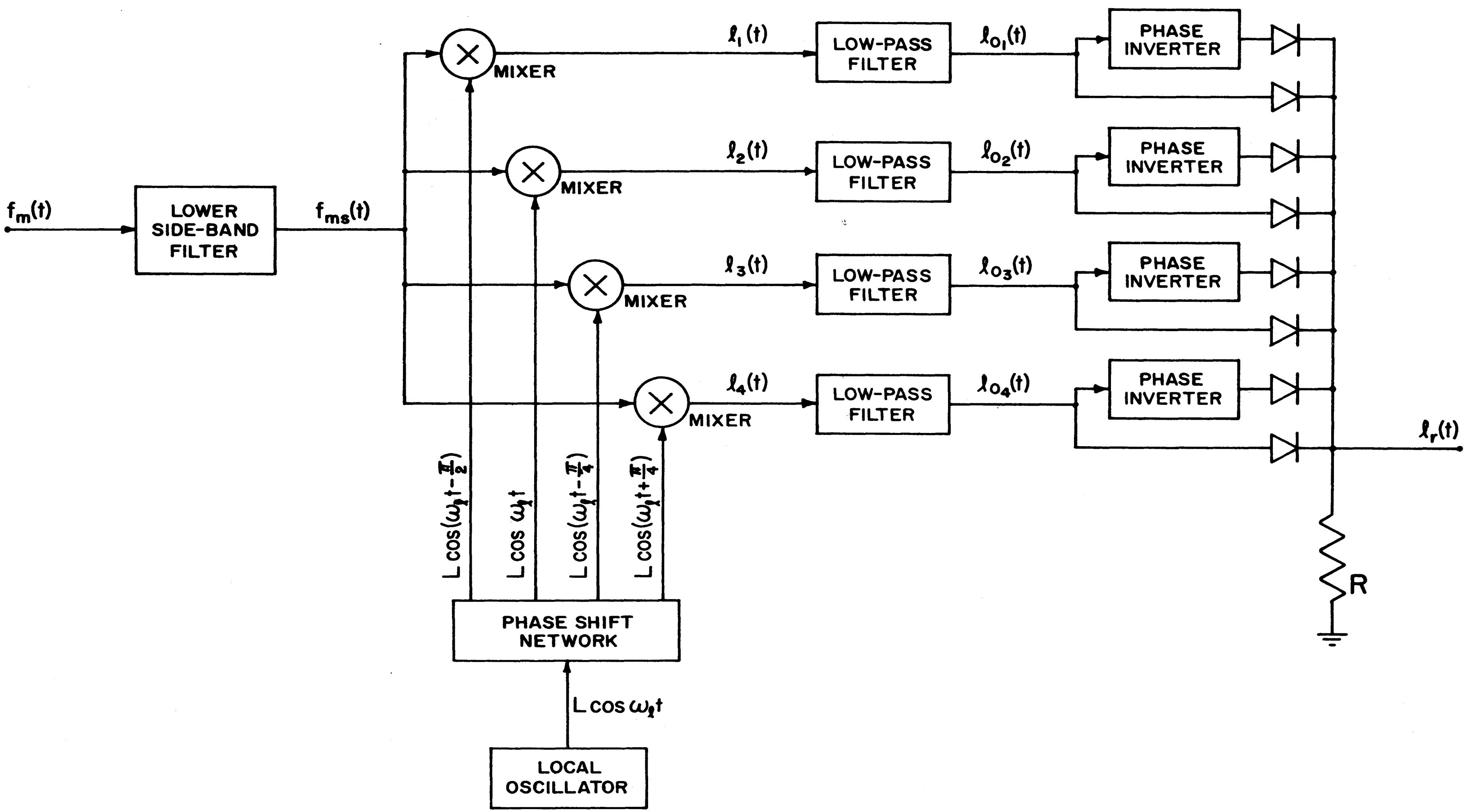


Figure 5.20. Multiple low-pass filter analysis, with phase-shifted carriers.

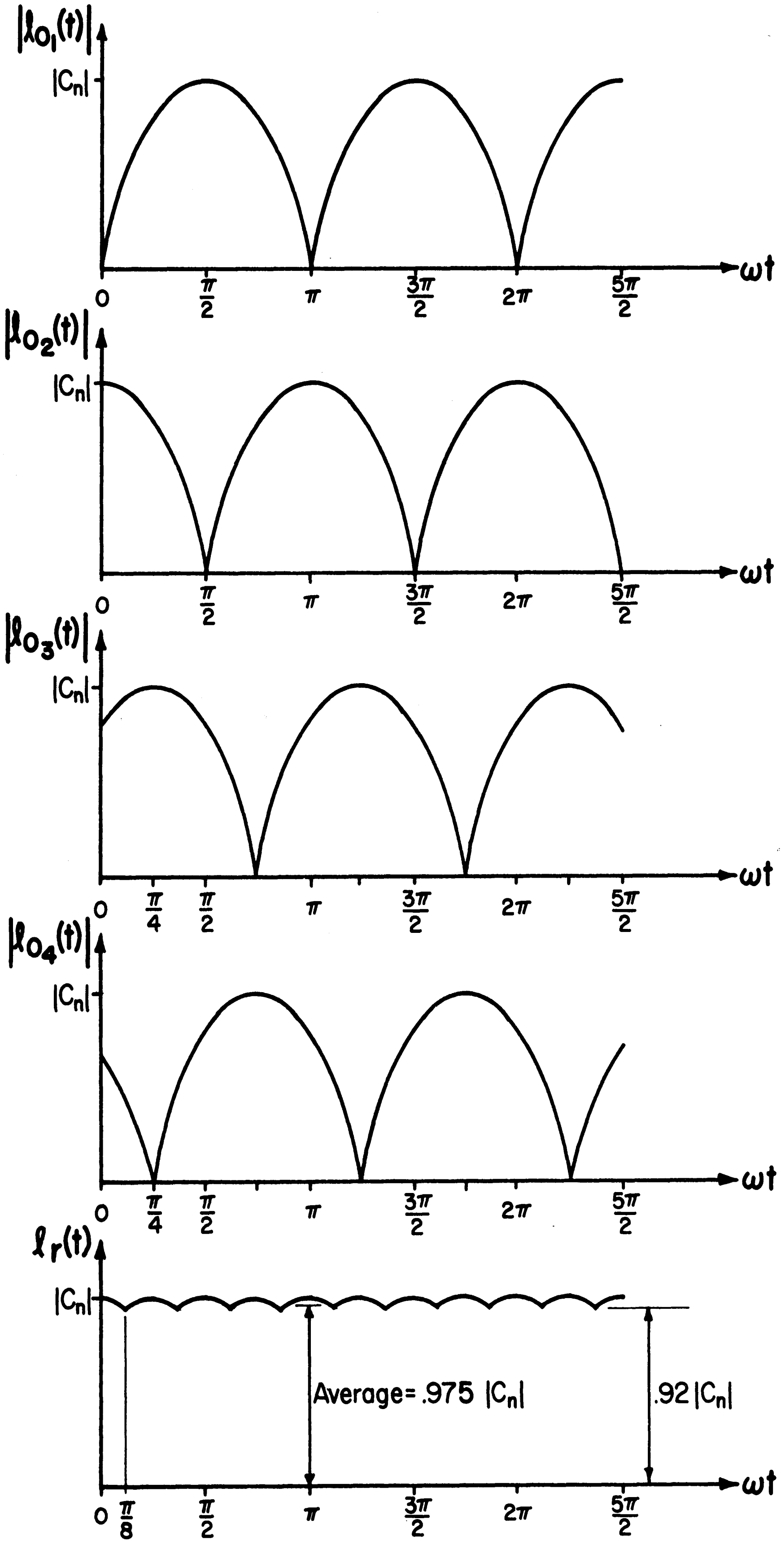


Figure 5.21. Single-frequency outputs of the phase-shifted low-pass filters, and their rectified combination.

from .707 $|C_n|$ to 1.000 $|C_n|$ with the average at .900 $|C_n|$. Thus the amount of the output fluctuation is reduced by increasing the number of phase-shifted modulations.

Remarks

Basic principles and circuits have been given in this chapter, for a sound spectrograph which uses a single fixed filter band for analysis. Considerable latitude has been left, however, in the actual design. One such design will be described in the following chapters, including some features which have not been mentioned in the foregoing presentation. Various examples of spectrograms will be given, showing how the spectrographic characteristics are changed by the bandwidth, the phase shift, and the time constant of the circuits used.

6. The CSL Sound Spectrograph

The development of a new sound spectrograph was undertaken at the Communication Sciences Laboratory because a need was felt for certain features not contained in available commercial models. The main objectives of the design may be listed as follows:

1. It should be possible to analyze sounds in the range from 30 to 10,000 cps.

2. The system should have two recording channels, either of which can be analyzed during reproduction of the signal. It should be possible to switch from one channel at any given instant in each cycle of reproduction, and the relative timing of the two channels should be adjustable.

3. Although the relatively slow analysis resulting from the use of a single scanning filter is acceptable, still the analyzing process should be as fast as possible, consistent with accuracy.

4. It should be possible to mark time, frequency, and amplitude scales automatically on the spectrograms.

5. Several scanning rates should be provided, at the choice of the operator.

6. Several different kinds of analyzing filters should be available, as well as several different filter bandwidths.

7. The amplitude display should have both averaging and envelope detectors, and these should have adjustable filter bandwidths and time constants.

Spectrograms

In addition to the above requirements, the spectrograph should of course have the usual features and capabilities outlined in Chapter 5. Examples of the three general types of display will be given here. A more general study of the effects of changing various parameters will be reported in Chapter 7.

a. Wide-band and narrow-band spectrograms

In Figure 6.1 are shown spectrograms of speech by a male voice, which were made with the CSL spectrograph. Three dimensions are shown on the flat display. Time and frequency are the horizontal and vertical linear dimensions, respectively, with the intensity of the signal shown by the darkness

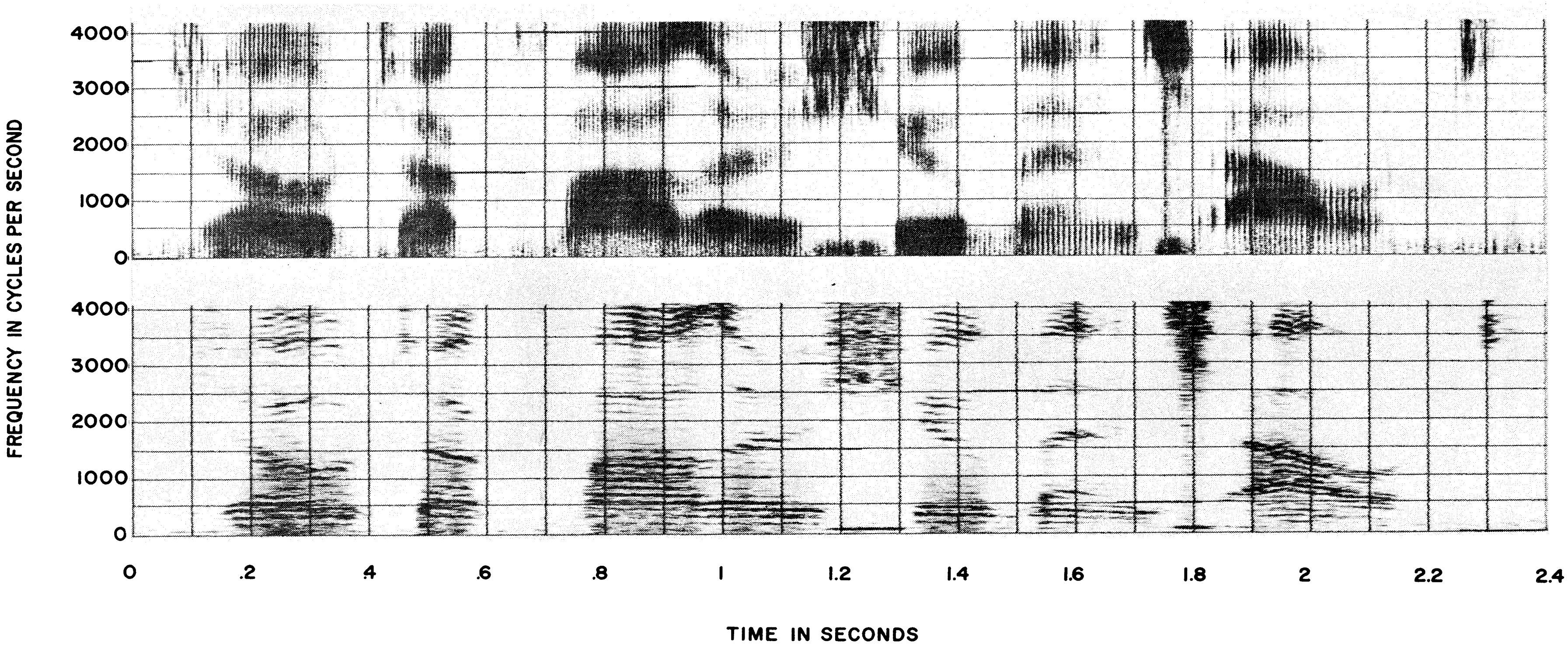


Figure 6.1. A wide band (300 cps) and a narrow band (50 cps) sound spectrogram of a male voice saying "Joe took father's shoe bench out." Time lines are shown each 100 milliseconds, and frequency lines each 500 cps.

of the record. In the upper spectrogram the analysis was made with a filter 300 cps wide, and the dark areas show the location in time and frequency of the main concentrations of energy, the so-called formants of the speech. In the lower spectrogram the filter bandwidth was only 50 cps., so that in the voiced portions only one harmonic of the fundamental frequency could be present in the filter at a time. Hence, in addition to the formants, the separate harmonics of the voice are also shown, rising and falling in frequency throughout the utterance. The grid of time and frequency lines was printed in these spectrograms as the spectrum analysis proceeded.

The close vertical lines in the wide-band spectrogram occur at the fundamental period of the voiced sounds of the speech. Their closeness of spacing corresponds with the height of the harmonic lines in the narrow-band spectrogram. The time lines do not appear in the latter because of the greater time constant of the narrow filter.

b. Amplitude sections

Figure 6.2 again shows, at the top, a wide-band spectrogram of speech. Below it are amplitude

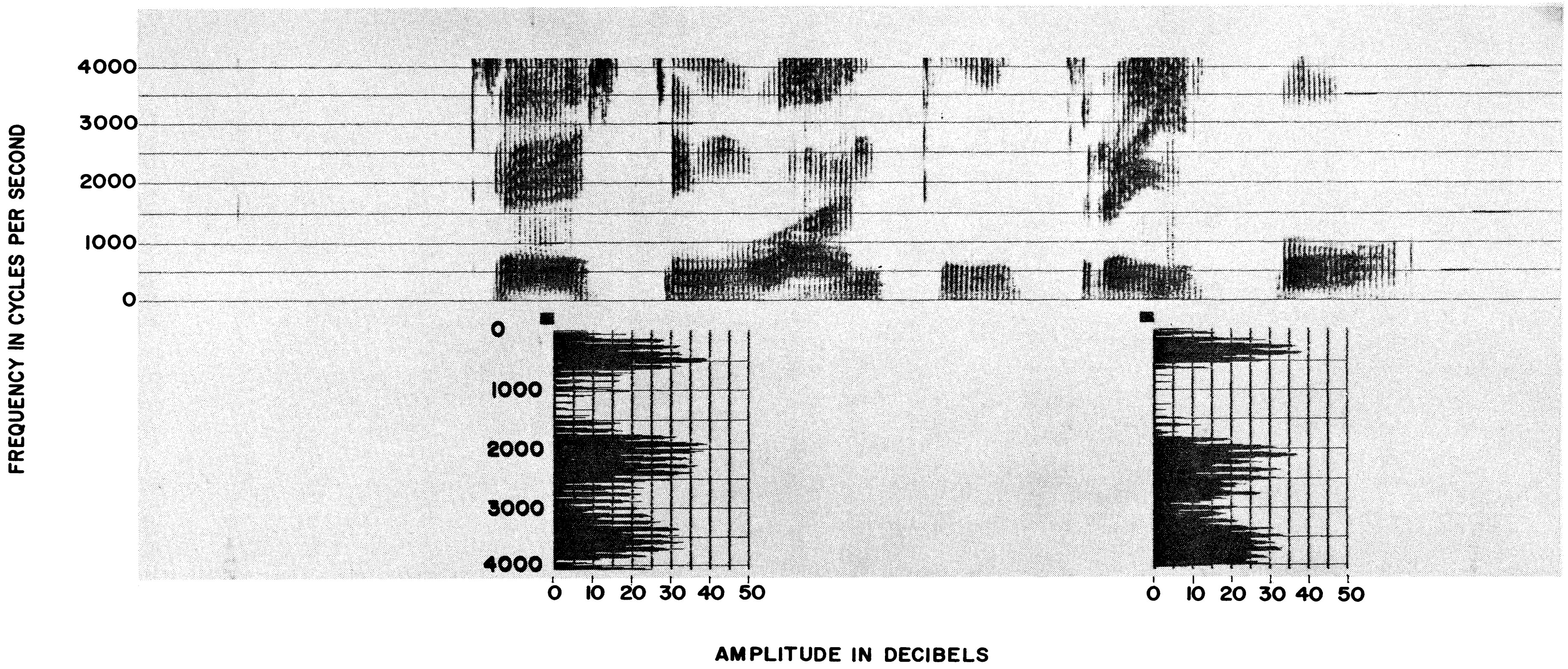


Figure 6.2. A wide band spectrogram of a male voice saying "testing one two three four", and narrow band amplitude sections at two points in the speech sample. The small rectangles show the position and integration time (about 20 milliseconds) of the sections. Amplitude lines are drawn on the sections each 5 db.

sections. These are two-dimensional plots of frequency versus amplitude, the time being fixed. Effectively, they are cross-sections of the spectrogram at fixed times. The small dark rectangles show the exact time span of the spectrogram which was integrated to produce the corresponding section. Although the spectrogram of Figure 6.2 is wide band, the sections were made with a narrow band filter, so that the separate harmonics of the voice are displayed. The grid on the sections shows linear frequency and logarithmic amplitude.

c. Continuous amplitude displays

In Figure 6.3 there is a wide-band spectrogram at the top and below it a continuous amplitude display. The latter was made with half-wave rectifiers; the part above the zero line is from the positive side of the signal and that below is from the negative side. An envelope detector was also used in constructing these continuous amplitude displays, with the time constant set at 8 milliseconds. Thus higher frequency variations are minimized, but the concentration of energy in successive cycles of the fundamental is shown. The amplitude scale is logarithmic, and the grid shows time versus amplitude in decibels.

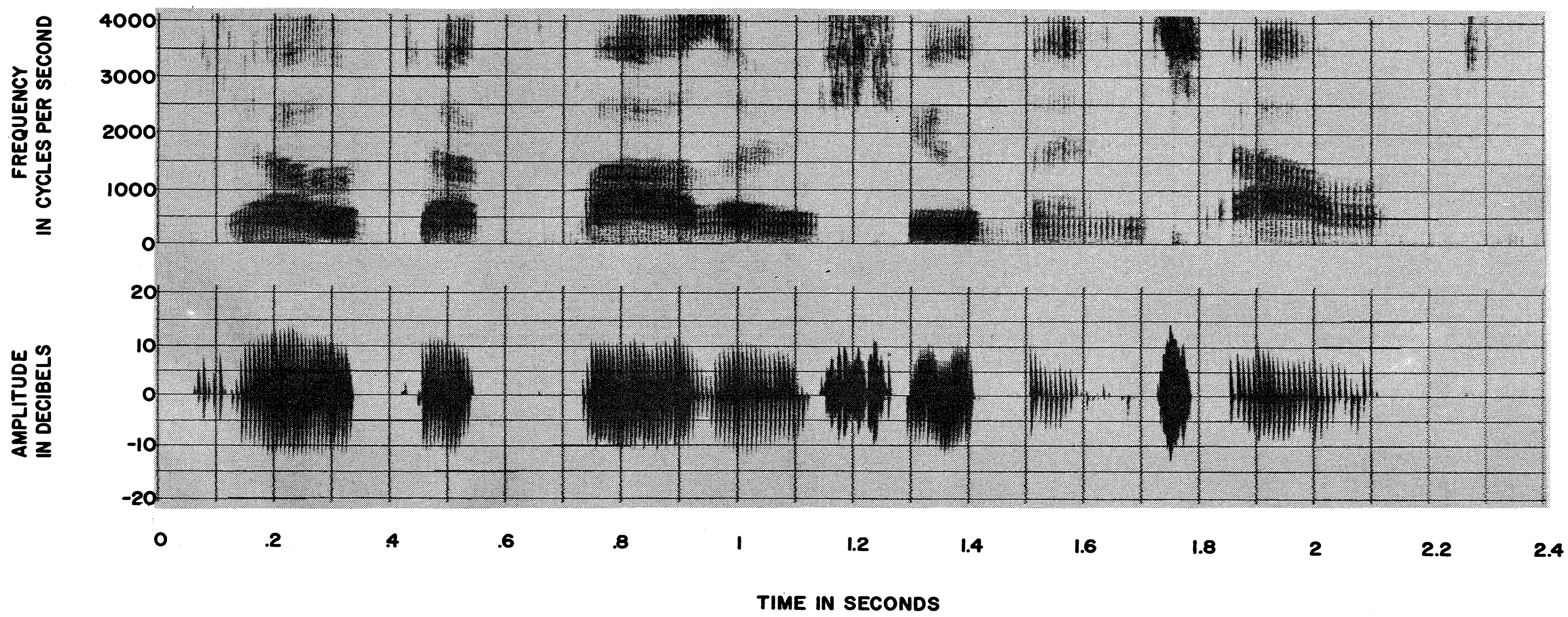


Figure 6.3. A wide band spectrogram of a male voice saying "Joe took father's shoe bench out", and a continuous amplitude display of the same speech sample. Amplitude lines are drawn each 5 db on the logarithmic scale of amplitude.

Operation of CSL sound spectrograph

A design which meets the objectives outlined at the beginning of this chapter must necessarily involve considerable complication, with a great deal of switching. The attempt has been made, however, to keep the operation as simple as possible. For this purpose, extensive use has been made of relays for automatic switching. Ease of maintenance has also been kept in mind.

A front-view photograph of the CSL spectrograph is shown in Figure 6.4, and a rear view in Figure 6.5. The arrangement is in panels, removable for test and repair. The main mechanical parts are located at the right in the front view, in the console standing in front of the racks which contain part of the electronic circuits. Recording paper almost nine inches wide can be placed on the revolving drum, which allows several associated displays to be made on the same sheet of paper. The drum circumference is 17 inches, so that this is the length of the paper record when the paper is removed.

The details of the circuitry will not be given here, but a functional block diagram is given in Figure 6.6, and will be described in the following pages.

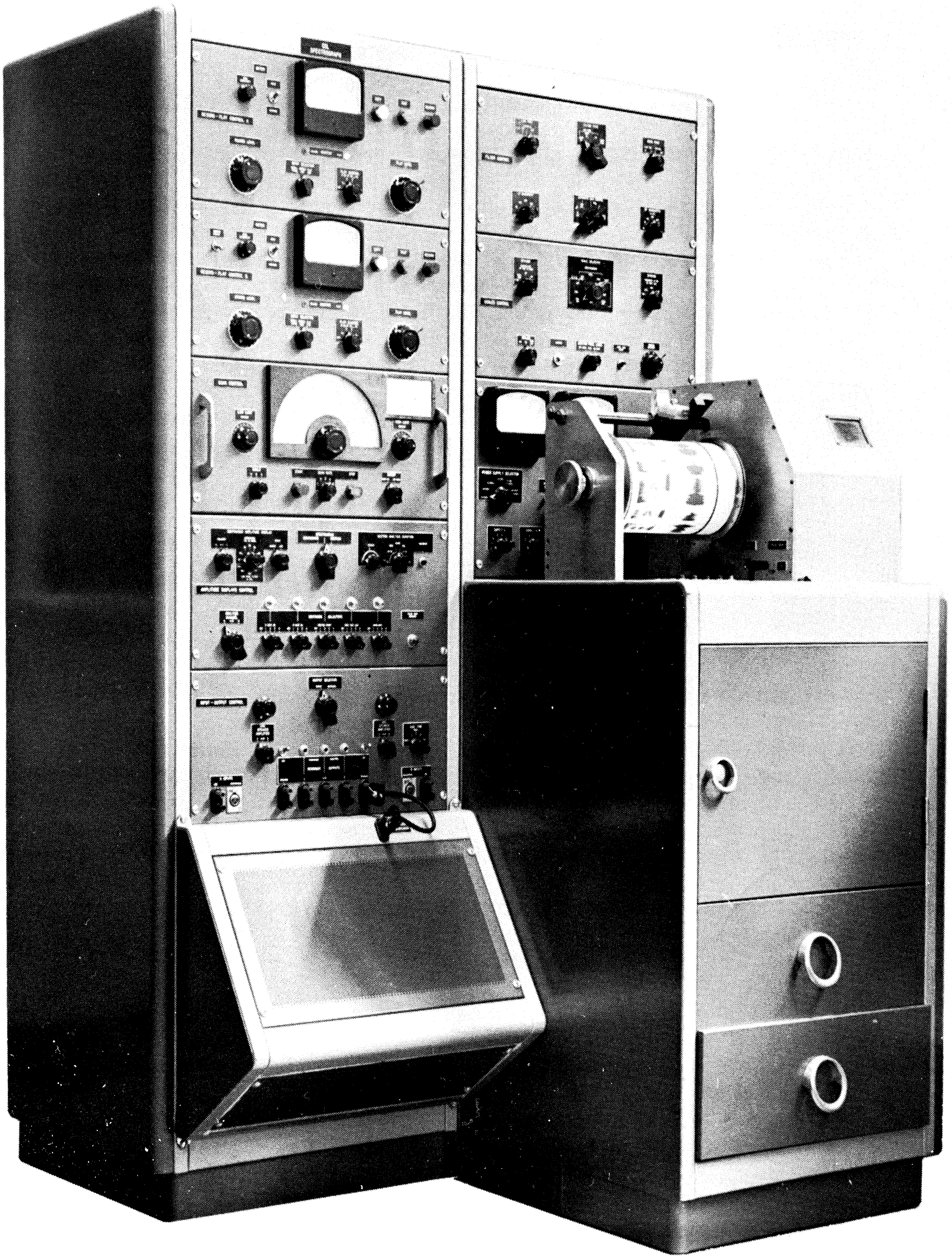


Figure 6.4. The CSL spectrograph -- overall front view.

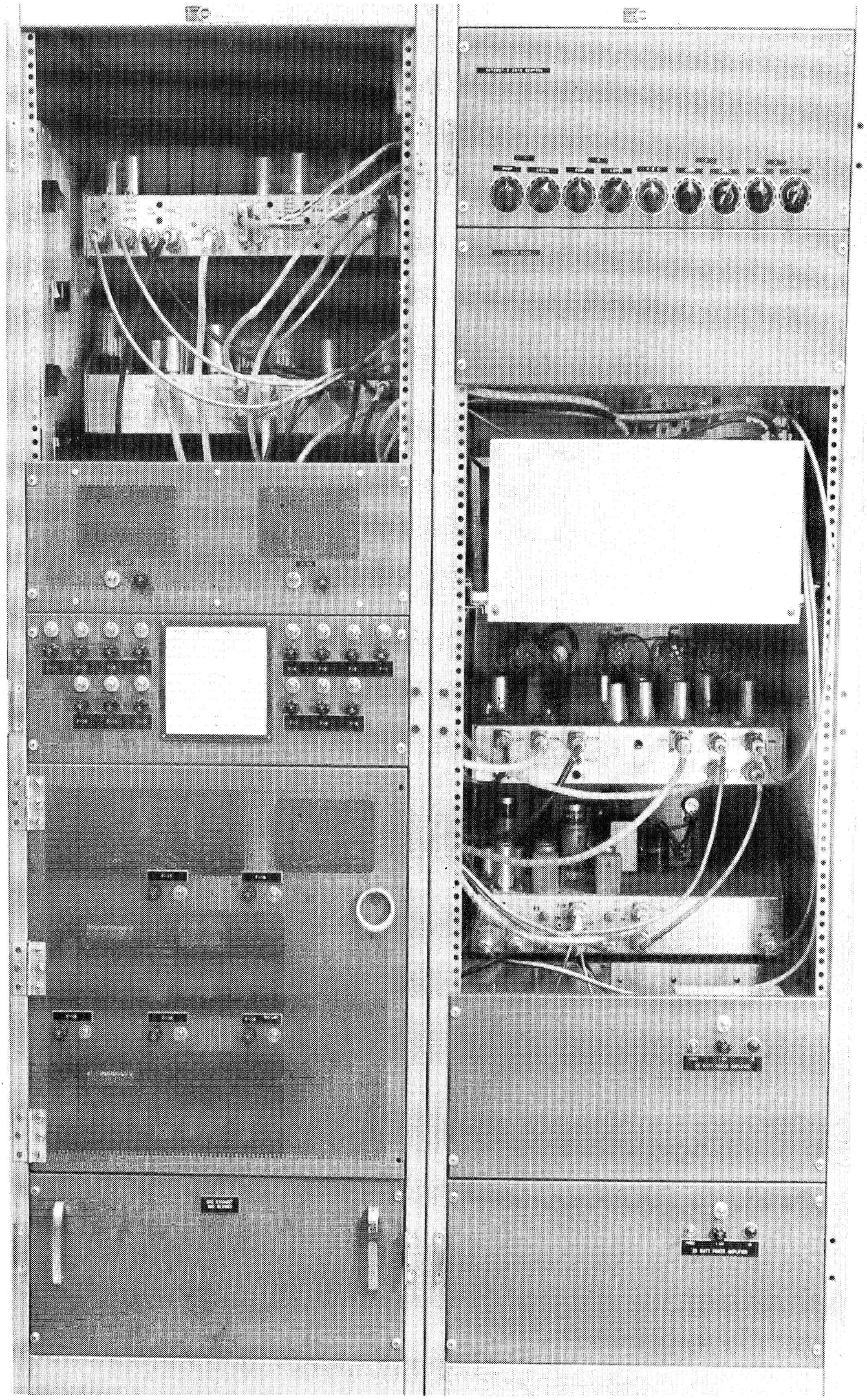


Figure 6.5. The CSL spectrograph -- rear view.

For simplicity, a number of features have been omitted from this diagram; the most obvious being the power supply and its distribution. All power comes from the usual 60-cycle supply, with the necessary fusing, switching, regulation, and voltage transformation. Other omissions will be mentioned as the description proceeds.

a. Recording and reproduction

At the left of Figure 6.6, five parts rotating on the same shaft are shown. The driving motor has two speeds, providing shaft rotation at 20 and 120 r.p.m. The two recording tapes are shown in the upper positions on the diagram. They consist of ordinary plastic magnetic tapes wrapped on rubber backing around aluminum drums and held securely. They are easily replaced when worn. Linear speeds of the tapes are 8.9 and 53.4 inches per second corresponding to the two motor speeds. In normal operation, the slow speed is used in recording and in monitoring the reproduction. The high speed is just six times the low speed and is used in reproduction for analysis. The recording-time capacity is three seconds, but a little of this is always lost at the beginning and the end of the recording.

The "A" and "B" recording and reproducing circuits are alike, and are fairly conventional, as shown in the figure. Some special features, however, are worthy of mention.

At the low speed of the motor, recording up to 10,000 cps is provided for. It is possible, however, to record at high speed, with recording up to 60 kc. To provide for this possibility, the bias and erase frequency is set high, at 200 kc.

The switching between "record" and "reproduce" is done by a relay, whose normal position is "reproduce". The "record" position exists only while a record button is depressed. A "test" button is also provided, which connects the recording circuit to the output amplifier for level adjustment purposes, but does not connect the record and erase heads. Still another "copy" button is provided, which connects the monitoring elements to the reproduce head instead of to the record head while recording is taking place. Monitoring at this position may sometimes be preferred for various purposes, including checking on the fidelity with which the signal is being recorded. Another automatic feature is that the relay which makes the actual connections to the record and erase heads cannot be energized unless the motor is running, at either speed. The "test" relay, however, works with the motor stopped. Although not shown, a "couple" switch, when closed, provides that both A and B circuits are put in the "record" position by pressing either of the record buttons.

Another function operated by either record button is the clamping of the indexing disc which normally rotates with the main shaft, and is located close to the paper drum. This disc is released to rotate again when the record button is released, and the mark on the disc then indicates the point on the drum where the recording ends. In wrapping paper around the drum, the overlap is placed at this point.

Equalizers are shown in both recording and reproducing circuits. The purpose of the record equalizer is to increase the strength of frequencies above about 1000 cps, so that the higher formants of speech are recorded at more nearly the levels of the lower formants. This provides better signal-to-noise ratios. The record equalization is compensated for in the reproducing equalizer, which provides as well for the deficiencies of the magnetic recording and reproducing process. The recording equalization can, however, be switched out, with corresponding changes in the reproduce equalizers. Since there are two motor speeds, separate reproduce equalizers must be provided. The change from one to the other is made automatically by relay, depending on the position of the motor speed switch.

In all cases the reproduced signal reaching the output amplifier is essentially flat with frequency in relation to the original input signal. In reproducing the signal, however, it is possible to introduce "shaping", to increase the amplitudes of higher frequency components. Several different rates, of increase of gain with increasing frequency, are provided and can be selected as desired. Some of the conditions of equalization are also very useful in analyzing signals other than speech.

VU meters are provided, which may be used in setting both recording and reproducing attenuators. The recording level may be set before recording begins, by the use of the "test" button described above. Not shown in the figure is a circuit which can be switched in before the VU meter, making it respond to peak rather than average instantaneous levels of the signal.

The outputs of the A and B circuits are led to separate "gates" (the operation of which will be discussed later), and thence to a switch by means of which either the A or the B channel or the sum of both may be connected to the analyzing circuit. The gates may be bypassed if desired.

Not shown in Figure 6.6 is a monitoring amplifier and loudspeaker, for the purpose of hearing the signals. This may be connected to any of the three outputs, A, B, or A+B. The monitor circuit is provided with a relay which automatically disconnects the loudspeaker under certain conditions. One of these is the case where a microphone is being used and the record or test button is pressed. Removing the loudspeaker in this case prevents acoustic feedback by interaction between the microphone and loudspeaker. The speaker is also disconnected when the motor switch is turned to high speed, and when the motor is slowing down from high speed. This automatically eliminates the annoyance of listening to the distorted and unintelligible speech in high-speed reproduction.

b. Analyzing circuits

Three different switches, as shown in the analyzing portion on the upper right of Figure 6.6, are involved in the choice among "spectrogram", "section", and "amplitude" displays. In the actual instrument, these three are mechanically combined into a single three-position control.

Analysis of the reproduced signals, whether from A, B, or A+B, proceeds essentially as outlined in Chapter 5. Since analysis to 10,000 cps is desired, and the motor speed in reproduction is six times the recording speed, frequencies up to 60 kcps may be encountered in analysis. To keep the modulation frequency higher than this, in accordance with condition (3.8), the scanning oscillator is set to work between 100 and 160 kcps.

The modulated signal can then be led to a band-pass filter centered at 100 kc, in which case the lower sideband of the modulation is swept past the filter band. Alternatively, the lower sideband goes to a set of four low-pass filters, after being demodulated by four 100-kc sine waves which have four different phase constants 45 degrees apart. The theoretical basis for employing more than one low-pass analyzing filter was discussed in Chapter 5. As the number of demodulation channels is increased, the accuracy of the amplitude representation of frequencies near zero is also increased. Four appears to be a very adequate number of such demodulation channels.

Switching has been provided for choosing between the band-pass and low-pass filters, for different types of filters in each of these classes, and for a number of different bandwidths of the filters. However, only a fraction of these possibilities are available at the present writing. Most of the work is being done with the less expensive low-pass filters, and for this ten sets of four filters each have been provided. These include five bandwidths (50, 100, 200, 300, 500 cps) and two types designated "Flat Top", and "Linear Phase". Some of the effects of filter type will be shown in Chapter 7 of this report.

Because of the relatively small range of levels provided by the electrical recording paper, two means of compressing the signal are provided, so that a range of 35 to 40 db of variation in the analyzed signal is presented within the 12 db of the paper. The first of these takes the entire unanalyzed signal, when a spectrogram or amplitude section is being made, rectifies it with a slow time constant, and applies the result to the marking amplifier in such a way that the gain of the latter is reduced when the signal is strong. This is similar to the "full band control" of the original Bell Telephone Laboratories spectrogram (see ref. 13), and increases the strength of weak phonemes in the speech with respect to the stronger ones.

The second type of compression consists of a non-linear circuit element shunting the input to each of the analyzing low-pass filters. Each of the filters is preceded by a control low-pass filter which has twice or more the bandwidth of the analyzing filter. This arrangement is similar to the "narrow band control" of the B.T.L. spectrograph. This type of compression is not quite as effective with the low-pass analyzing filters as it is with a high band-pass analyzing filter. Even with band limitation by the control filter, there can be modulation products generated by the non-linear element which lie inside the low-pass band. However, the cut-off characteristics of the analyzing filter are not impaired by the compression ahead of it, as it would be if the compression were applied to the filter output. By using only the level of that part of the spectrum close to the analyzing band for controlling the degree of compression, the system makes weak formants more prominent, as well as strengthening those portions of the signal which are weak in point of time. A similar control filter and compressor arrangement would be needed for the band-pass circuit if a good set of such filters were available.

Both types of compression can be removed completely, if desired. When in use, however, several switching combinations of the two types are provided. One of these uses the controlled compression without the full-band gain control, with choice of control filter widths of 100 or 200 cps, and respective analyzing filters of 50 or 100 cps. For the wide analyzing bands of 200 and 300 cps two choices are provided; one (No. 2) uses the full-band gain control alone and the other (No. 1) uses both types of compression, with the control filter width set at 500 cps for both of the analyzing band-widths.* All of the widths mentioned refer to original recorded frequency; the actual filter widths are six times as great.

The continuous amplitude display, i.e. the analysis in amplitude rather than frequency, is made essentially as outlined in Chapter 5. The switching provides a choice of positive half-wave, negative half-wave, or full-wave rectification. There is a logarithmic circuit which may be used or not, a choice of average or envelope detectors, a choice of five low-pass filter cutoffs ranging from 125 to 2500 cps, and a choice of seven recovery-time constants between 8 and 44 milliseconds. The various low-pass filters provide a choice of relatively rapid smoothings of the output of the average detector, and the alternative recovery-time constants for the envelope detector provide a further and slower smoothing for the amplitude marking, varying from voicing to syllabic rates. To the detector output a voltage is added which is varied by the scan motor to provide the amplitude analysis, and the combination

---

*When the No. 2 control setting is being used the analyzing bandwidth of 500 cps is also available.

is impressed on the threshold circuit. Although separate frequency and voltage scans are shown on the diagram, the voltages are actually derived from the frequencies by means of a frequency discriminator circuit.

The scanning circuit deserves some further mention. It is driven by a motor which is reversible and has five speeds in either direction. The variable condensers driven by the motor have plates shaped to provide a linear variation of oscillator frequency with angular position. The arrangement of speeds and scales is shown in the photograph of the "scan control" panel, Figure 6.7. The scale switch at lower left refers to the three scales of the semicircular dial. A is the outermost frequency scale, B the next (also frequency), and C the amplitude scale in db. The reversing switch is at the lower right, and the five motor speeds are chosen in the lower center.

The motor drives the condenser shaft through a clutch, which permits setting the scale by hand at any desired location before pressing the start button. A radial index line extends across the scales, and in addition a square of light shows behind the translucent panel, not only at the index position but also at the proper one of the three scales.

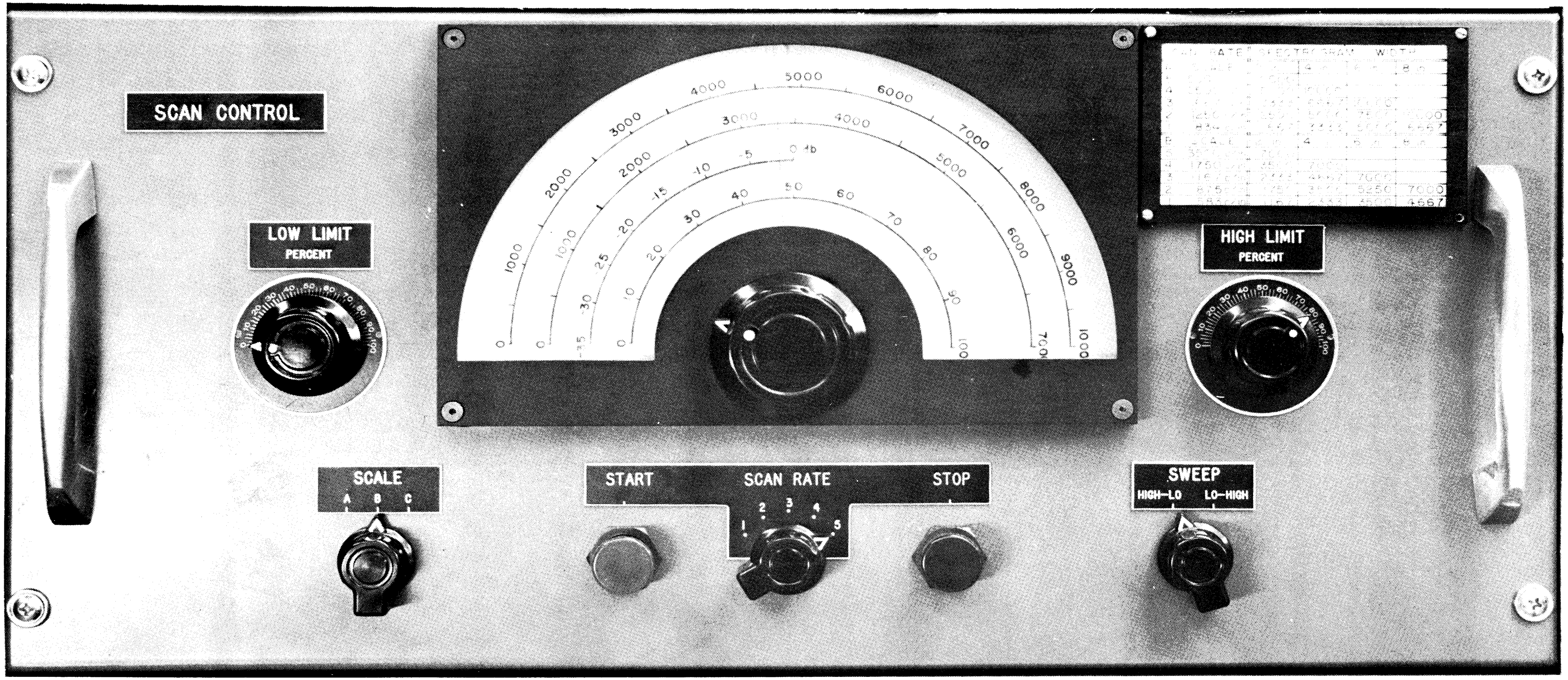


Figure 6.7. Scan control -- front panel.

This lighted portion moves with the shaft rotation, so that at any point of an analysis the frequency or amplitude being analyzed at the moment may be seen at a glance. A stop button is provided, but there are also upper and lower limit switches which may be set at desired positions, and which then stop the action automatically.

The frequencies shown on the scales are those of the original recording which are responsible for the analyzing filter output at each point of the scale. The actual frequencies of the variable oscillator, for the A scale, are 160 kc at the 10,000 cps position, and 100 kc at the zero position. For the B scale a change in the oscillator is made, so that the upper frequency is 142 kc, corresponding to 7000 cps at the top of the scale, and the lower limit is again 100 kc.

At the upper right of the panel is a table for the convenience of the operator, where the different scan rates are given in terms of cps per inch width of the spectrogram. The frequencies covered in 2, 4, 6, and 8 inches of spectrogram width are also given. The five scanning speeds (on scale A) are actually 1000, 1500, 2000, 3000, and 6000 cps per minute. The relation to width of spectrogram can be obtained from the fact that the stylus moves across the spectrogram at 0.01 inches per revolution of the drum, and since the latter makes 120 rpm in analyzing, the stylus then moves 1.2 inches per minute. A complete analysis cannot last longer than six and two-thirds minutes, since in this time the full practical paper width of 8 inches is covered. A more usual situation is the use of the A scale with speed three, requiring two minutes for analysis to 4000 cps, or five minutes for analysis to 10,000 cps. Corresponding widths of the spectrogram are 2.4 and 6.0 inches.

c. Pulsed switching

The third rotating member at the left of Figure 6.6 carries a recording tape to which is applied a single four-channel head. Erasing each channel is accomplished, with the 20 kc erase oscillator connected and the shaft turning at slow speed, by pressing and holding the correct button for one or more revolutions. DC pulses are then recorded with the shaft stopped in the desired position with respect to the stylus and the parts of the recorded sample. Then in the high-speed analyzing condition these pulses are picked up at definite times of each revolution and amplified in the four pulse amplifiers.

In the case where the A and B gates are being used, the A and B pulses are led to an electronic switch which is arranged to open the A gate when the A pulse arrives, but then when the B pulse comes the A gate is closed and the B gate opened. The order can be reversed, of course. Then, with the analyzing circuit connected to A+B, the analysis shifts from the A recording to the B at a definite point of time, and it does this for each revolution. The relative timing of the A and B recordings can be changed manually, by changing the B disc in angular position on the shaft with respect to the A disc.

Another useful application is to separate out a short interval of a single recording, with one pulse turning it on and the other turning it off. The effect can be listened to on the monitoring circuit if the shaft is rotated at slow speed. This can serve a useful purpose other than in making spectrograms; for example, in observing the auditory effect produced by a short segment of speech when isolated from its context. For further possible use an external connection is provided through which any one of the pulsed channels can be brought out.

The section pulses can be placed at one or several positions of shaft rotation. Each such pulse then causes an electronic connection in the "section gate" which puts the integrating condenser on charge. This connection is held for a length of time determined by the setting of the "section driver" and the condenser is then discharged into the threshold circuit. The charging time is continuously variable from 1 millisecond (of the original recording) to 3 seconds. The latter represents the complete recording time available.

A button is provided which connects the section driver to the threshold circuit, in such a way that marking is produced during the period of charge of the section condenser. By depressing this button for a few revolutions of the drum a rectangle is produced on the record (as in Figure 6.2), which shows clearly the integration period of the section.

The fourth pulse channel is used to provide a zero for the recording of a time scale on the spectrogram or continuous amplitude display. This pulse is applied to a gate, which opens to permit a 6000 pps generator to be connected to a "pulse divider". The latter has a three-position control to permit one pulse in 20, one in 100, or one in 200, to trigger the marking circuit. Since the 6000 pps correspond to 1000 pulses per second of the original recording, the marking pulses indicate 20, 100, or 200 milliseconds of the record. The counting starts at the same moment of each drum revolution, so that the time lines build up across the record as the analysis proceeds. Provision is also made, by using all three of the pulse division rates, for placing a time scale along the side of a spectrogram. This scale has marks every 20 milliseconds, with every fifth mark slightly longer and every tenth mark still longer.

When sections are being made, the section pulses rather than the time zero pulse initiate the pulse counting. The time rate of discharge of the section condenser is adjusted to the division of pulses so that each marked pulse corresponds to a 5 db fall of condenser voltage. Thus a series of amplitude lines is drawn across the section, spaced at 5 db intervals from the start of each section.

Frequency lines are generated by beating the scanning oscillator, first against a 100 kc signal, and the product of this against a frequency derived from the 6000 pps generator. The latter frequency may be set at 1200, 3000, or 6000 pps. Marking of the frequency line is triggered only when the frequency being analyzed is a multiple of this frequency, or in terms of recorded frequencies at 200, 500, or 1000 cps intervals. The latter two become 5 db and 10 db intervals when a continuous amplitude display is being made.

The frequency lines may be made to mark through the full extent of a spectrogram, or alternatively only at the two ends of the spectrogram. In a section they may mark either through the entire record, only through the separate sections, or only between sections.

All grid lines may be turned off completely, or they may be used together or either one alone in each of the three types of display: time and frequency for spectrograms, amplitude and frequency for sections, time and amplitude for the continuous amplitude display.

The order of setting the pulses may be as follows: The time zero may be set as soon as recording is finished, since the beginning of the record is shown by the index disc associated with the paper drum. That is, the pulse is recorded with the drum stopped at such a position that the stylus is near to or just beyond the index mark. A spectrogram is then made, and from it one chooses the several points where sections are to be made and records pulses at these points. If both A and B recordings are made, spectrograms of each are made side-by-side on the same paper. The relative timing of the two records is then adjusted by turning one recording disc on the shaft with respect to the other. The A and B gate pulses are then recorded at desired positions. The frequency lines or the amplitude lines for the continuous amplitude display require no pulse recording.

d. Marking

Marking of the electrically sensitive paper occurs only when a "main gate" is opened to allow a 100 kc sine wave with sufficiently high voltage to pass to the marking amplifier and stylus. Four different "trigger" circuits are shown in Figure 6.6, any one of which can initiate marking. One of these, the spectrogram trigger, must transmit a range of levels, and the main gate must respond proportionately to these, so that a range of grayness is produced on the recording paper. The other three trigger circuits operate in an "on" or "off" manner, with no range of "on" levels. Gains can be set in each circuit, however, to produce the desired darkness of record, from full black to a faint gray. Thus the scale lines can be made either darker or much lighter than the section or continuous amplitude records.

As mentioned earlier, the gain of the marking amplifier can be controlled by the general level of the reproduced signal, thus forming a part of the means of compressing levels within the paper range. Means are also provided (but not shown in Figure 6.6) for suppressing marking except when the main motor is on high speed.

The spectrograph is also provided with a blower which picks up the smoke produced by the paper marking, close to the point of origin, and conducts it outside the building in which the apparatus is housed.

7. Some Spectrographic Reproduction Experiments

This chapter will present some of the experimental results obtained with the CSL sound spectrograph. The output records (i.e., the sound spectrograms, amplitude sections, and continuous amplitude displays) and their dependence on the functional characteristics of the analyzing filters, detector time constants, filter band-widths, scanning rates, and other factors will be discussed in the following sections.

Sound spectrogram

A typical speech sound spectrogram made with the spectrograph is shown in Figure 7.1. From the band-width criterion, this spectrogram is classified as a wide-band sound spectrogram. The filter used was a low-pass flat-top Tchebychef-type filter, which had an equivalent pass-band of about 300 cycles per second. The normalized amplitude and the phase transfer characteristics of this filter are given in Figures 7.2 and 7.3 respectively.

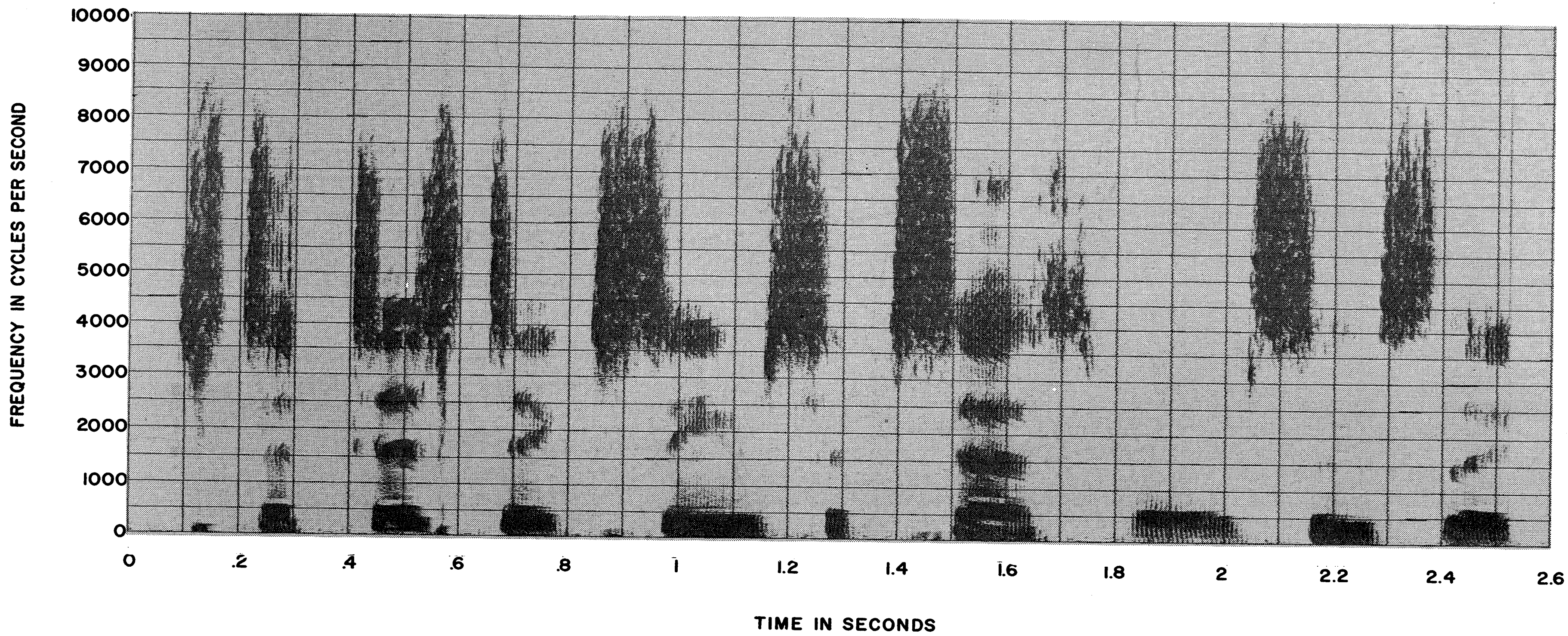


Figure 7.1. Wide band (300 cps) sound spectrogram of a male voice saying "Statistics seem successful since they --". The analysis extends to 10,000 cps at a scan rate of 1667 cps/inch. The flat-top filter has the amplitude and phase characteristics shown in Figures 7.2 and 7.3.

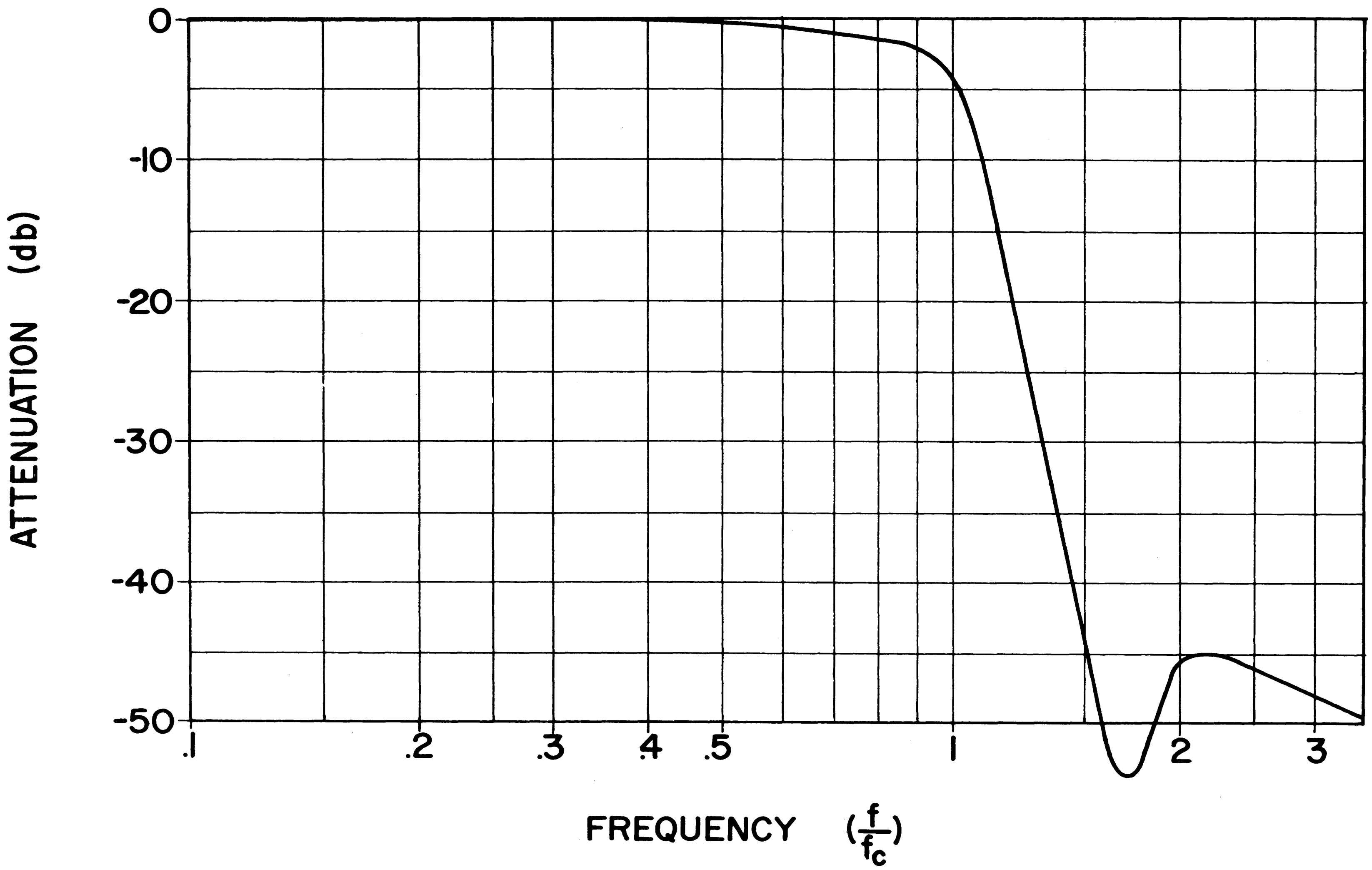


Figure 7.2. The transfer amplitude characteristic of a flat-top "Tchebyschef" type low-pass filter.

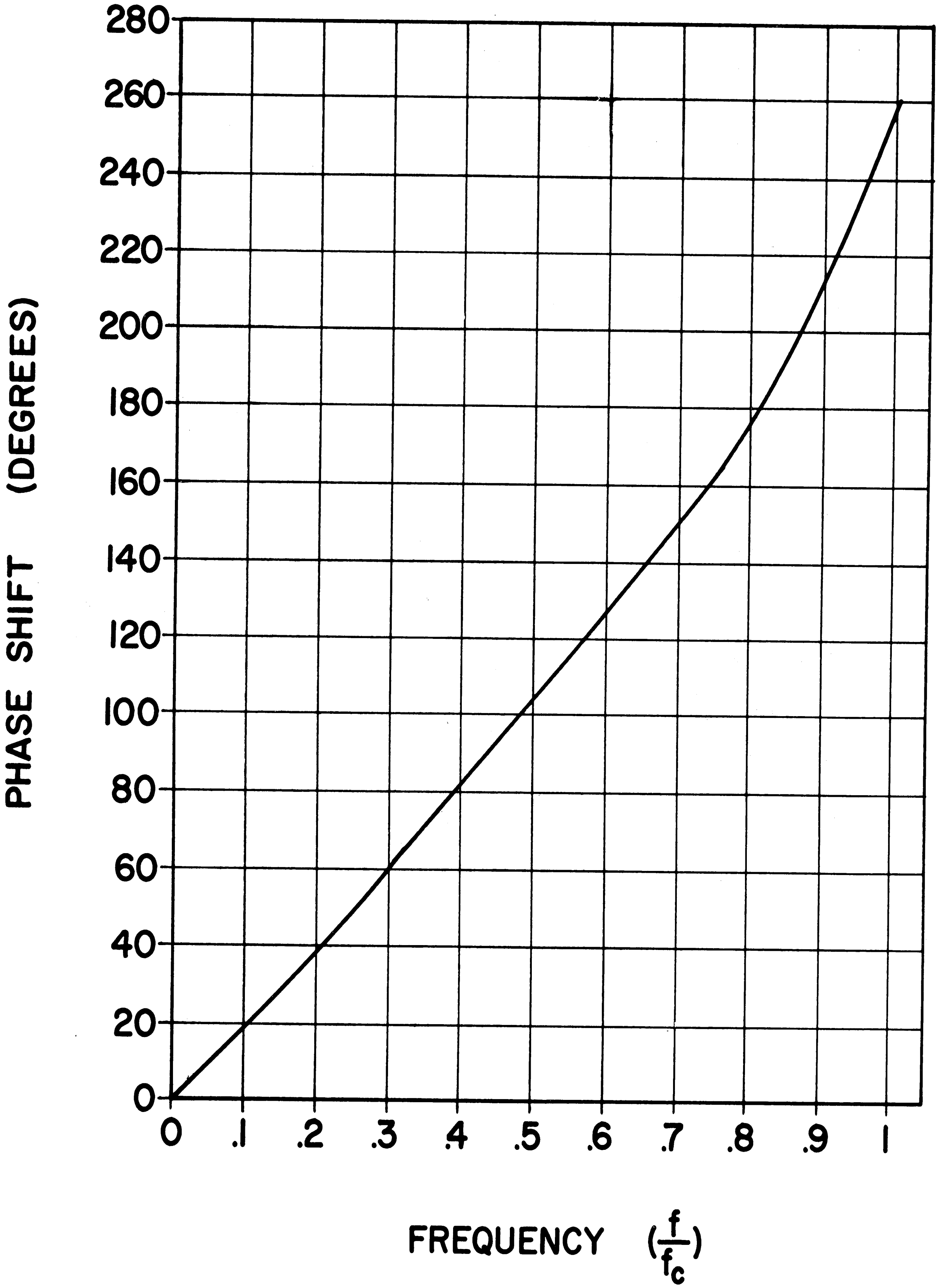


Figure 7.3. The transfer phase characteristic of a flat-top "Tchebychef" type low-pass filter.

In Figure 7.1, it can be seen that the vertical striations across the paper are not straight lines, but bend toward the right at both sides of each frequency band. This bending of the striations is caused by a longer time delay at the cutoff frequency of the filter. The difference in delay time at the output terminals of the filter is due to the non-linear transfer phase characteristic of the filter,* and can be calculated from the graph given in Figure 7.3. As an example, the delay time at the cutoff frequency of a 900 cps low-pass filter (i.e., 300 cps equivalent pass-band),** as shown in Figure 7.3, is calculated to be about 0.55 msec longer than that at the mid-band.

To avoid the type of distortion shown in Figure 7.1, a filter with a linear phase characteristic should be used. In practice a linear-phase low-pass filter is difficult to realize if a relatively flat-top amplitude characteristic is

---

*See the section "Signal distortion" in Chapter 4.

**The ratio of the reproducing speed to the recording speed of the CSL Sound Spectrograph is equal to 6. Therefore, the equivalent pass band for a 900 cps low-pass filter is

$$\frac{2 \times 900}{6} = 300 \text{ cps}.$$

also required. Nevertheless, a reasonable compromise can be made. Normalized amplitude and phase characteristics of a realizable linear-phase low-pass filter are given in Figures 7.4 and 7.5, respectively.

In Figure 7.6 is a sound spectrogram like that of Figure 7.1, except that the filter used was a low-pass linear-phase type. As with Figure 7.1, the filter has an equivalent pass-band of about 300 cycles per second. In Figure 7.6 it can be seen that the vertical striations are relatively straight, but that the frequency band marked is somewhat wider than that expected with the specified cutoff frequencies. This broader frequency band display is obviously due to the slow cutoff transfer characteristic of the filter, as shown in Figure 7.4.

A set of wide-band and narrow-band sound spectrograms produced at a constant scanning rate is given in Figures 7.7, and 7.8. The two filter types discussed above and a total of five different bandwidths were employed in making the spectrograms. The bandwidths and filter types are indicated with each of the spectrograms. All of the spectrograms in these figures are from a single recorded sample of the word "fish", spoken by a female voice. Only

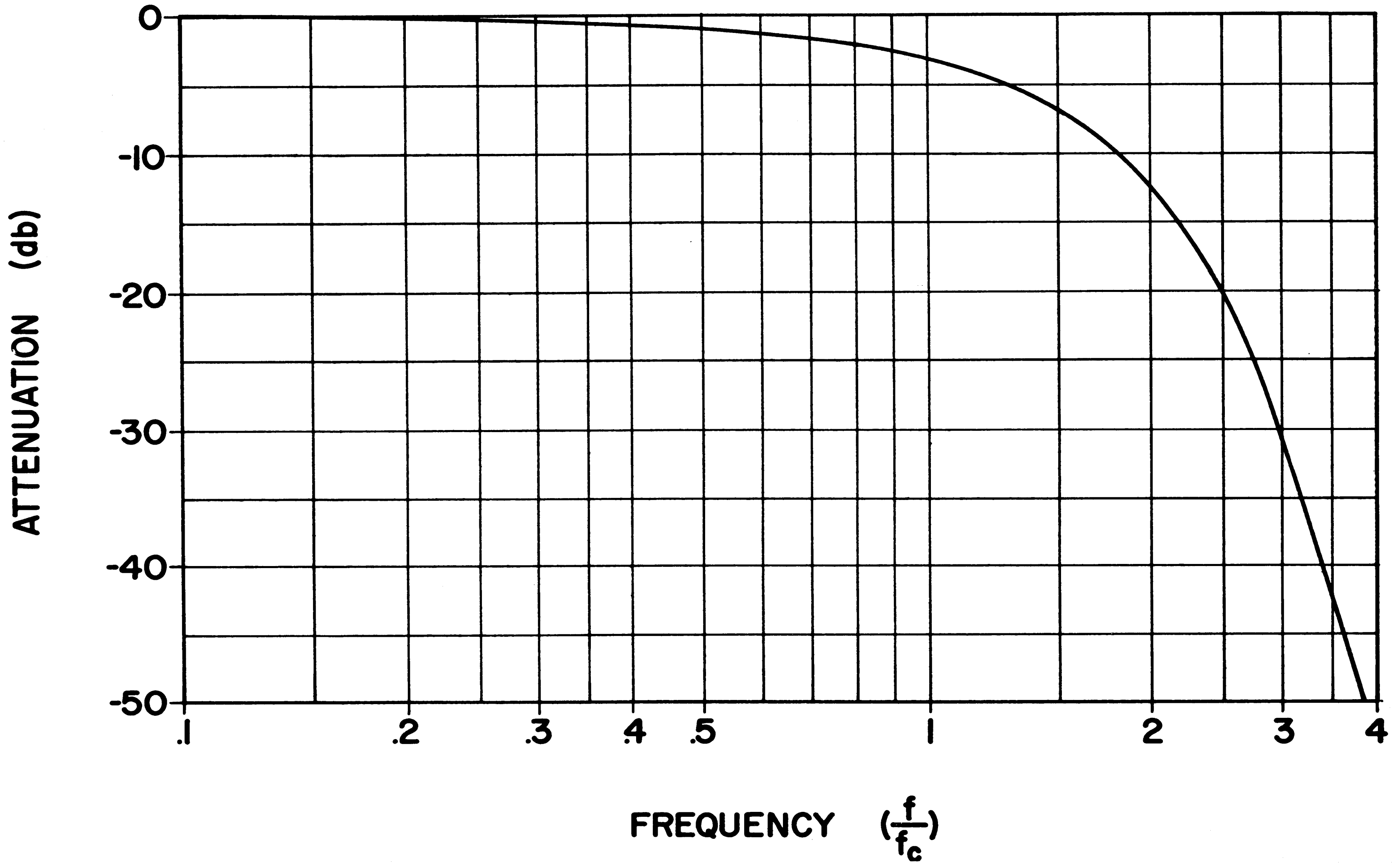


Figure 7.4. The transfer amplitude characteristic of a 9-pole linear-phase low-pass filter.

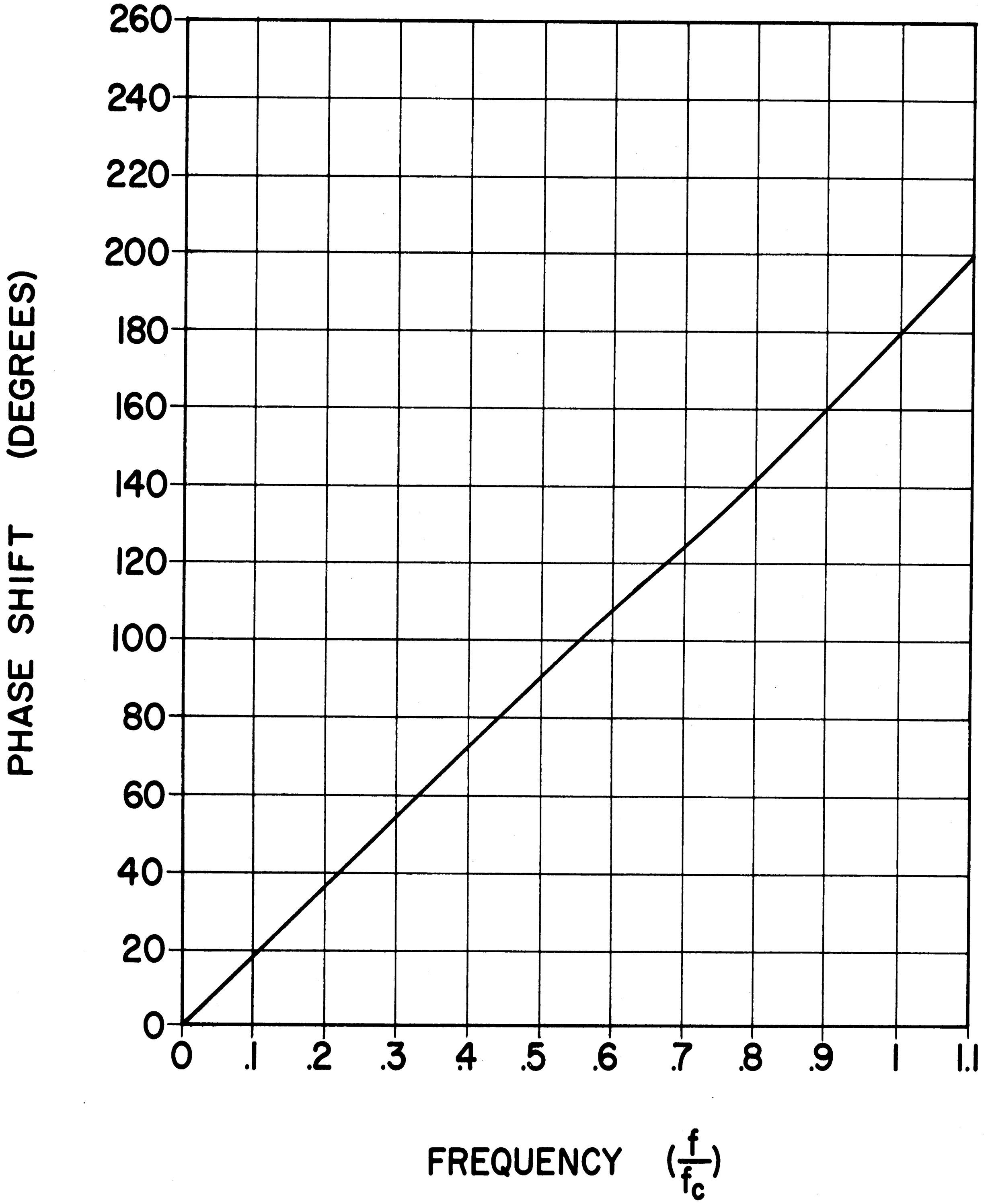


Figure 7.5. The transfer phase characteristic of a 9-pole linear-phase low-pass filter.

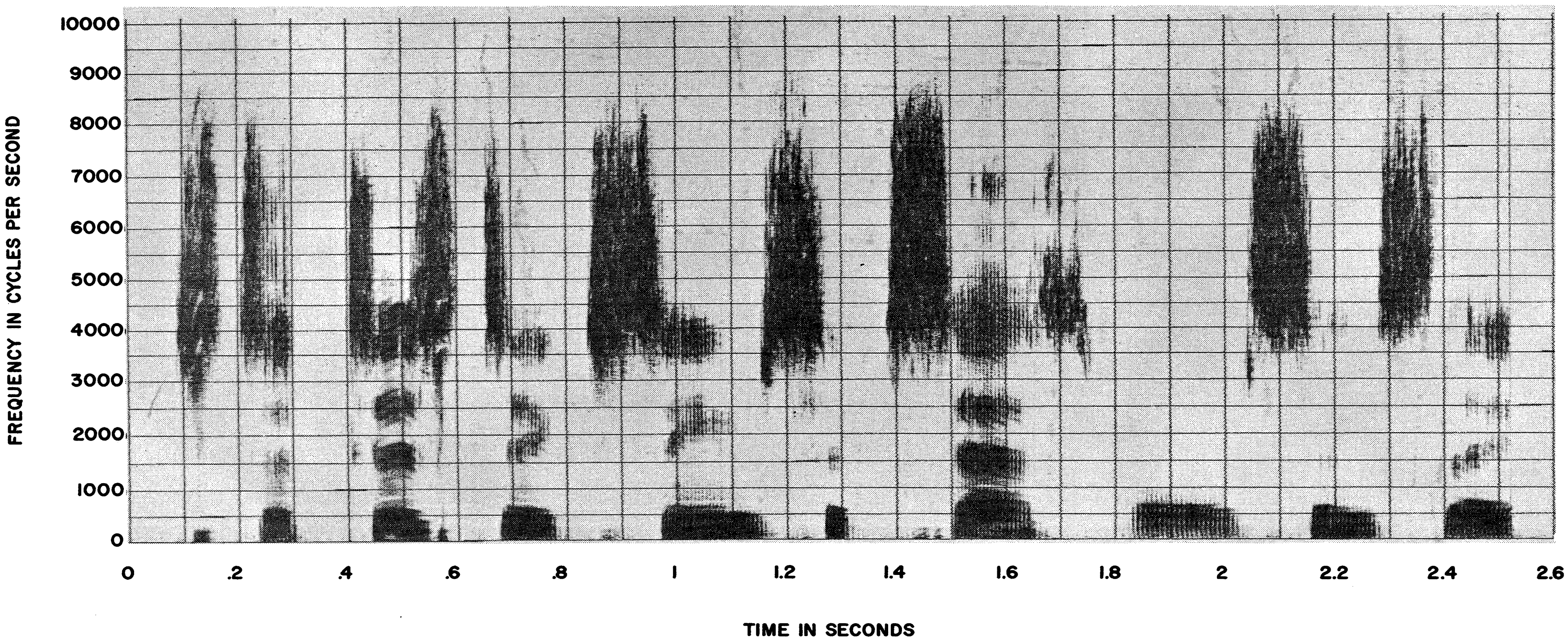


Figure 7.6. Sound spectrogram of the same speech sample used for Figure 7.1, with all conditions the same except that a linear-phase filter was used, with the characteristics shown in Figures 7.4 and 7.5.

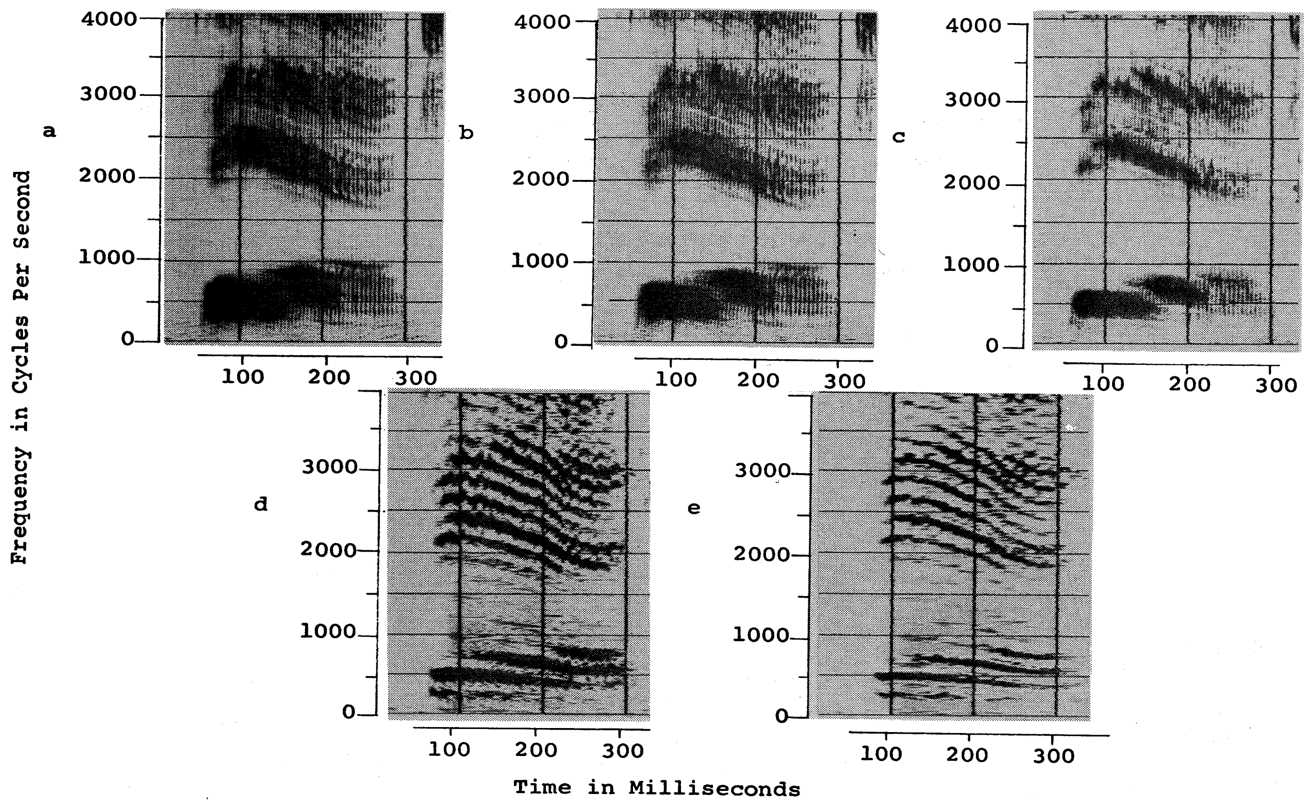


Figure 7.7. Spectrograms of the vowel in the word "fish", spoken by a female voice, and analyzed by a linear-phase filter having the following equivalent bandwidths: a, 500; b, 300; c, 200; d, 100; e, 50 cps.

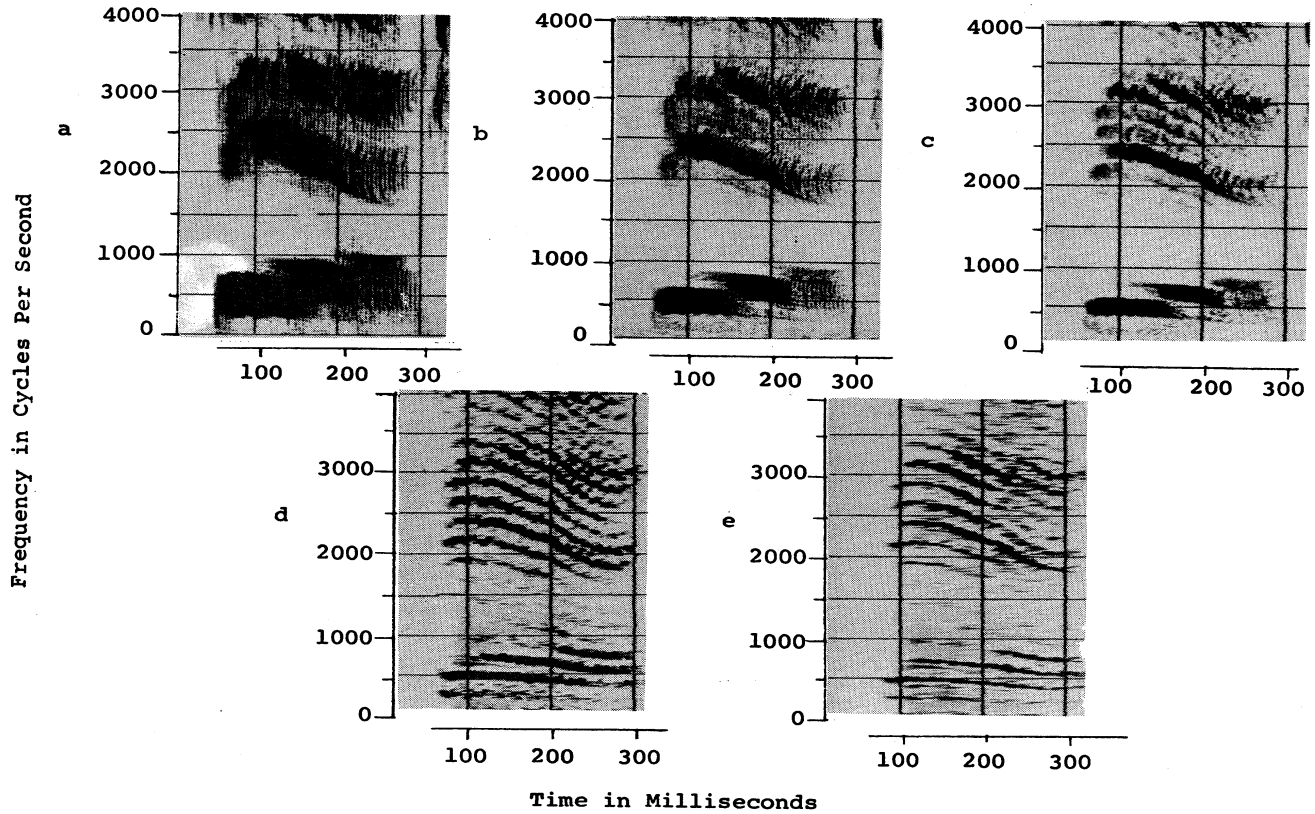


Figure 7.8. Spectrograms of the same speech sample and using the same nominal bandwidths used for Figure 7.7, but with a flat-top filter.

the vowel portion of the word is shown in the figures. It can be seen that the output resolution of the spectrogram is greatly dependent on the transfer characteristic of the analyzing filter. From such examples, the user of the sound spectrograph should be able to choose the filter which will provide the most effective display for any particular voice.

One of the features of the CSL sound spectrograph is its flexibility. An example of this is the 5-speed scanning motor discussed in Chapter 6. Figure 7.9 shows sound spectrograms produced by different scanning rates, with the same speech wave, filter type, and band-width. By the use of these different scanning rates, the spectrograms are made to appear either expanded or compressed in the frequency scale, thus providing a choice in the frequency/time aspect ratio.

The scanning rates of Figure 7.9 were obtained by using the "A" scale of frequency scanning, with a top frequency of 10,000 cps. If the "B" scale were used, with a top frequency of 7000 cps, the cps/inch rates of Figure 7.9 would be reduced by a factor of 0.7.

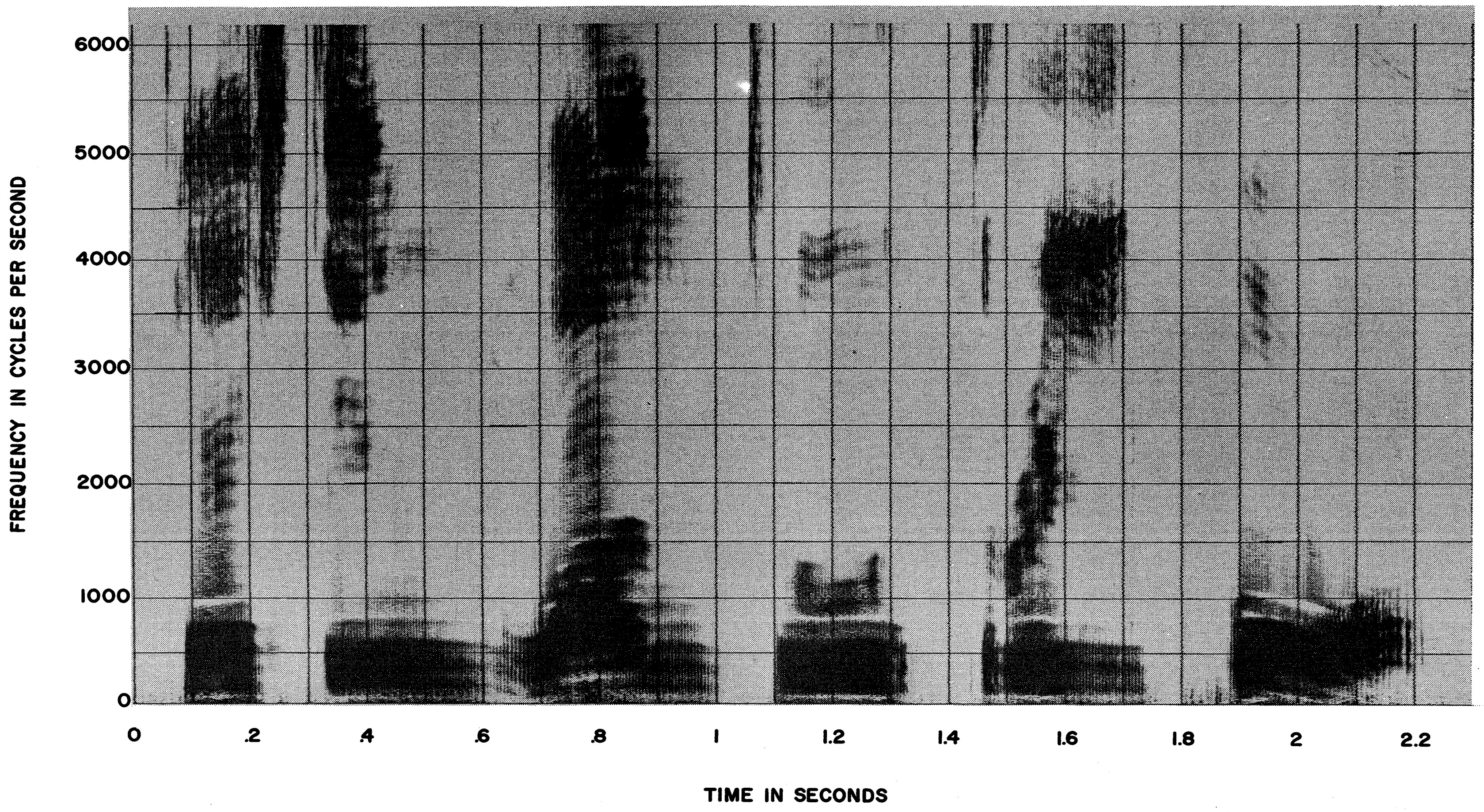


Figure 7.9a. Wide band sound spectrogram of a male voice saying "testing one two three four". The rate of scanning was 834 cps/inch.

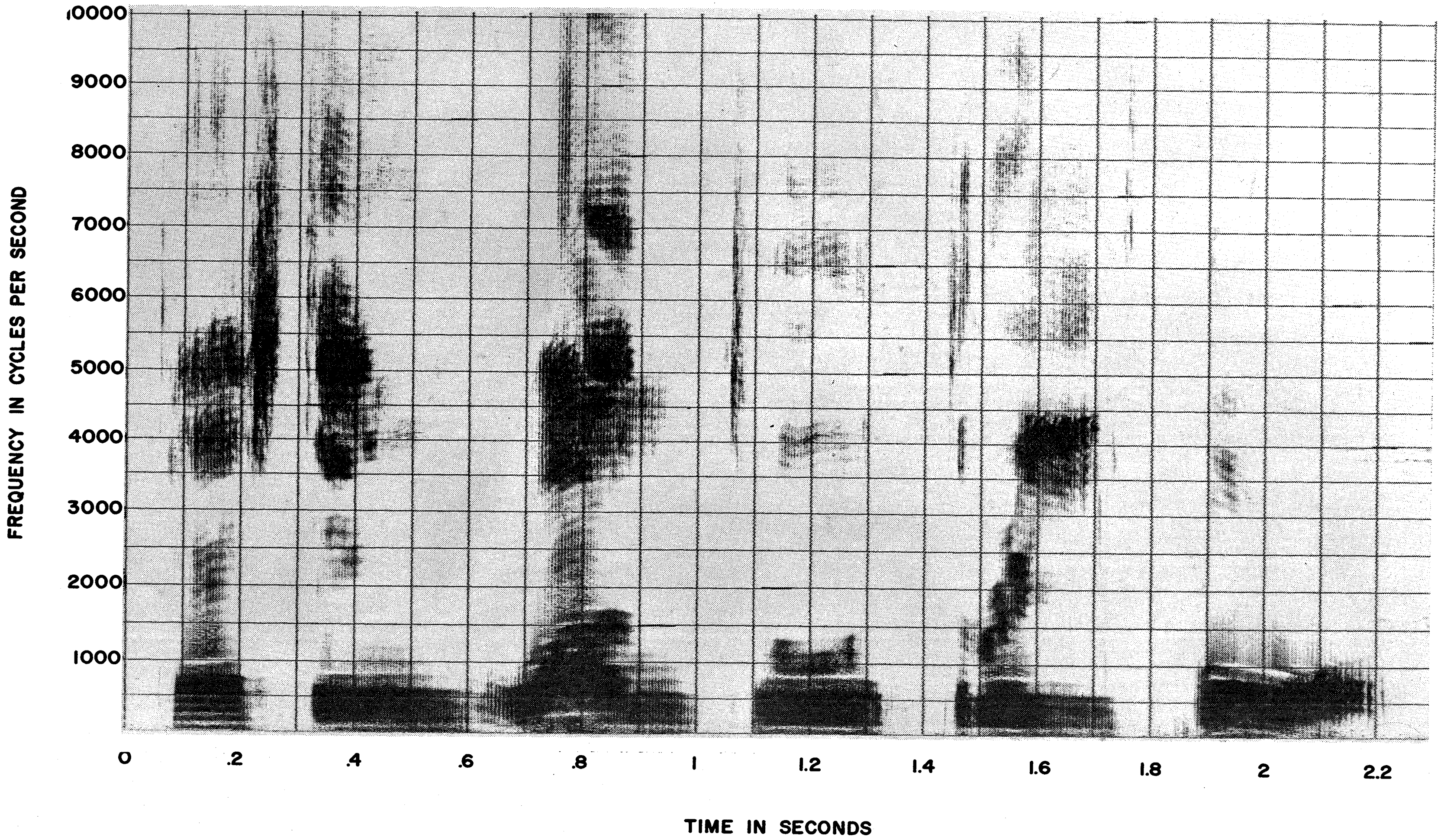


Figure 7.9b. The same speech sample used with Figure 7.9a, scanned at 1250 cps/inch.

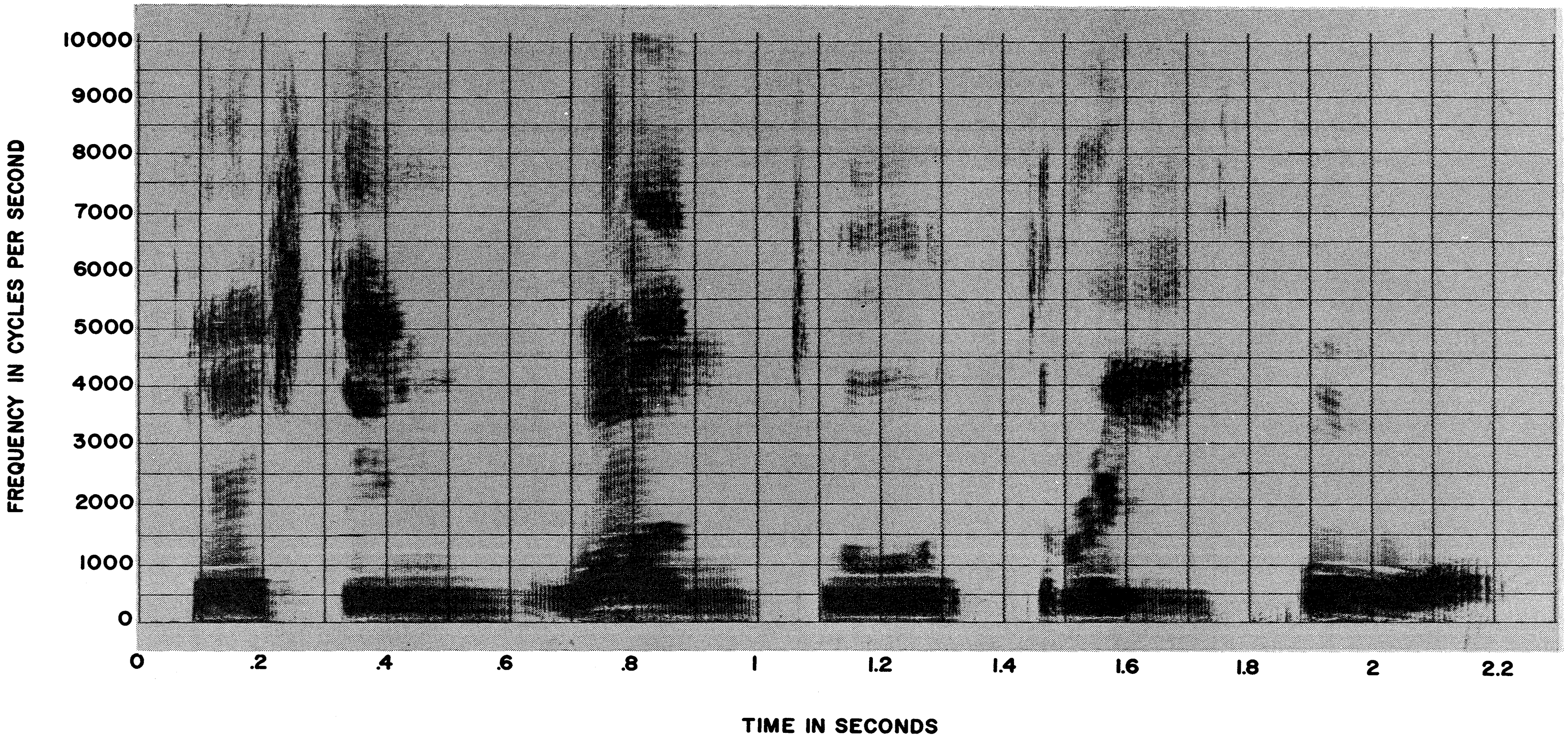


Figure 7.9c. The same speech sample used with Figure 7.9a, scanned at 1667 cps/inch.

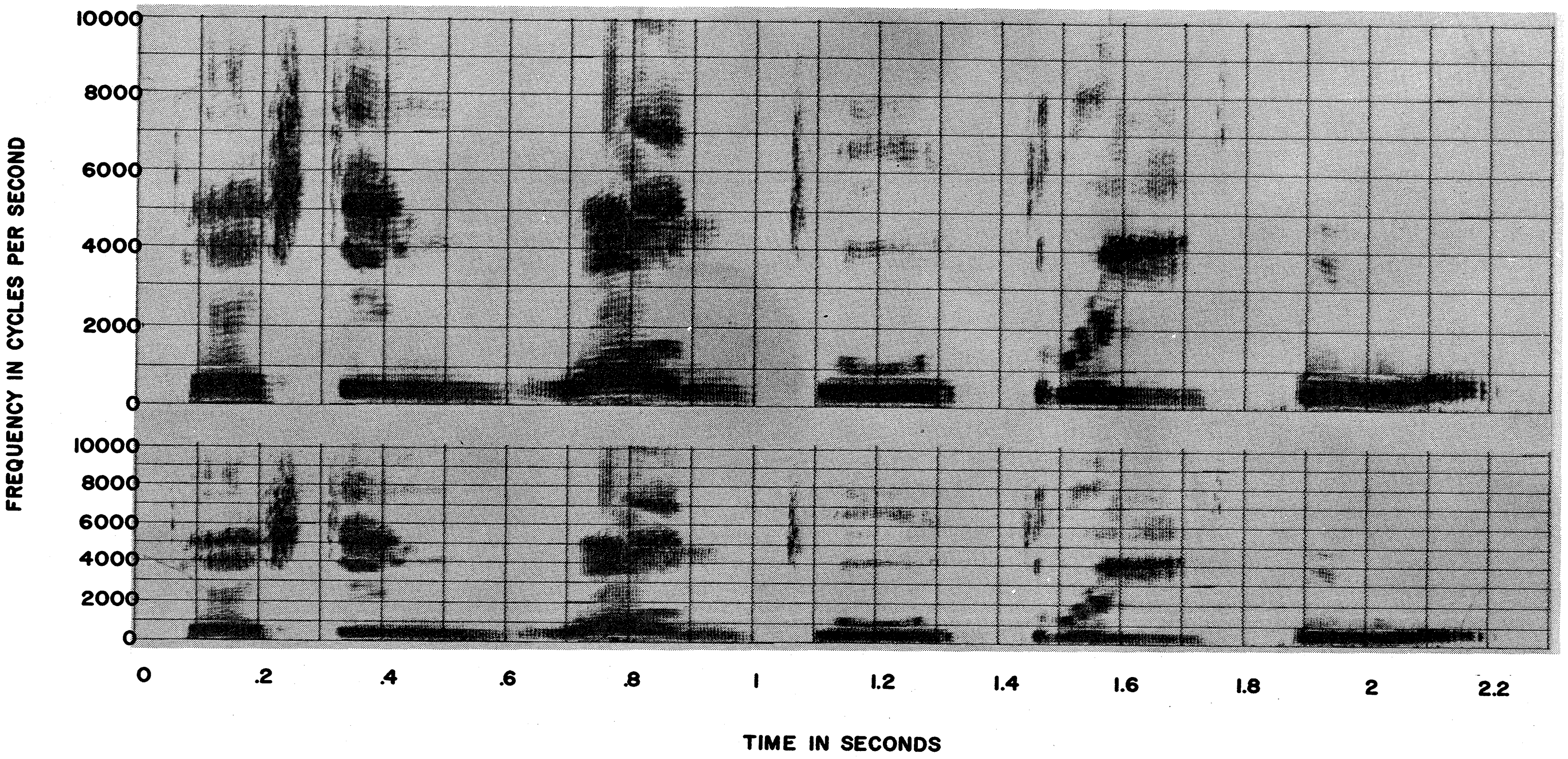


Figure 7.9d. The same speech sample used with Figure 7.9a, scanned at 2500 cps/inch (above) and 5000 cps/inch (below).

Amplitude section

The second kind of spectrogram produced by the CSL sound spectrograph is called an amplitude section (as defined in Chapters 5 and 6). An example of this amplitude versus frequency plot, at a fixed time in a typical speech wave, is given in Figure 7.10. At the top of the figure a spectrogram of a word is shown, made with a 500-cps flat-top filter. The black rectangle just below the spectrogram shows the span of time in the spectrogram, over which amplitudes were integrated in producing the section. As was discussed in Chapter 6, the duration of the integrating interval of the amplitude section is one of the features which is adjustable in the spectrograph. It can be varied from one millisecond up to about 3 seconds.

At the bottom of Figure 7.10 an amplitude section is given for the span of time indicated by the rectangle. This section was made with the same filter which was used for the spectrogram. Three other sections are shown in Figure 7.11, for the same interval and the same filter type, but with bandwidths of 300, 200, and 100 cps. The flat-top and sharp cutoff of the filter are reflected in the shape of the energy peaks of the wider band sections.

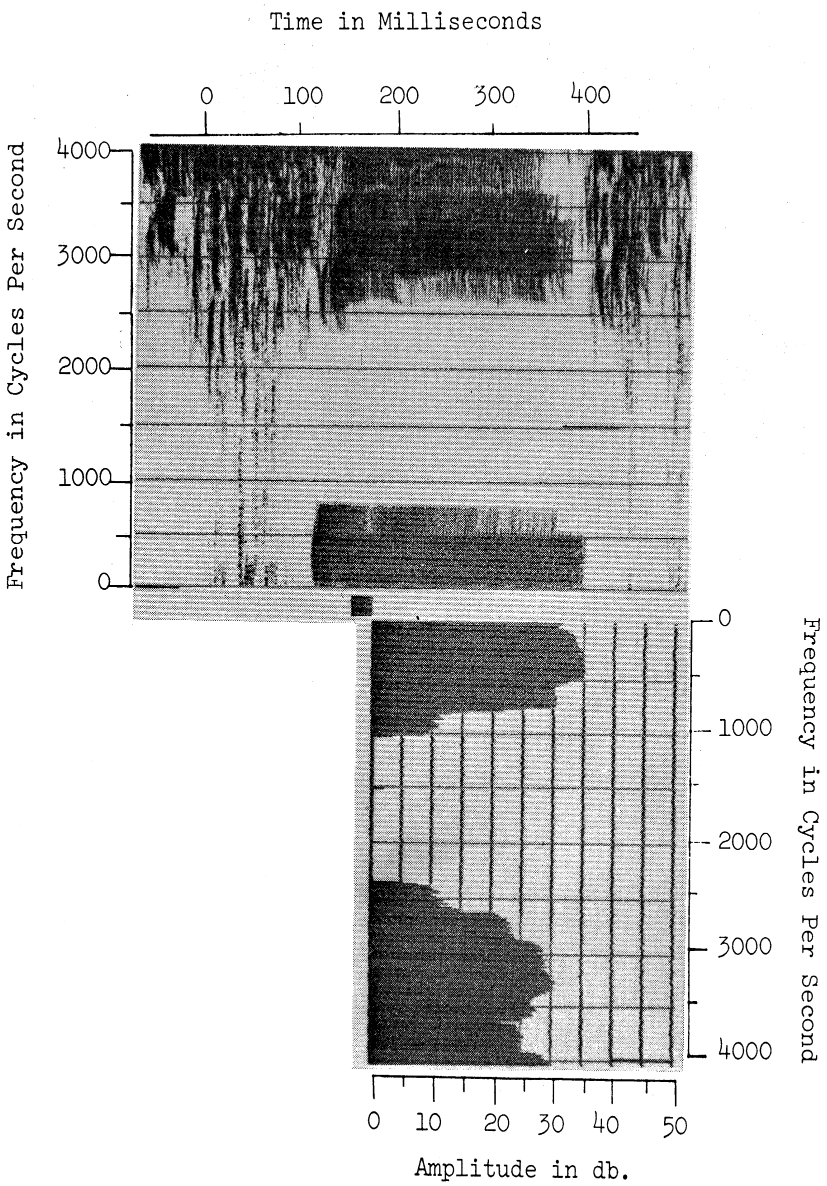


Figure 7.10. Spectrogram of "she s-" spoken by a female voice, and analyzed by a flat-top filter of 500 cps equivalent bandwidth. The amplitude section was made at the point shown, using the same filter.

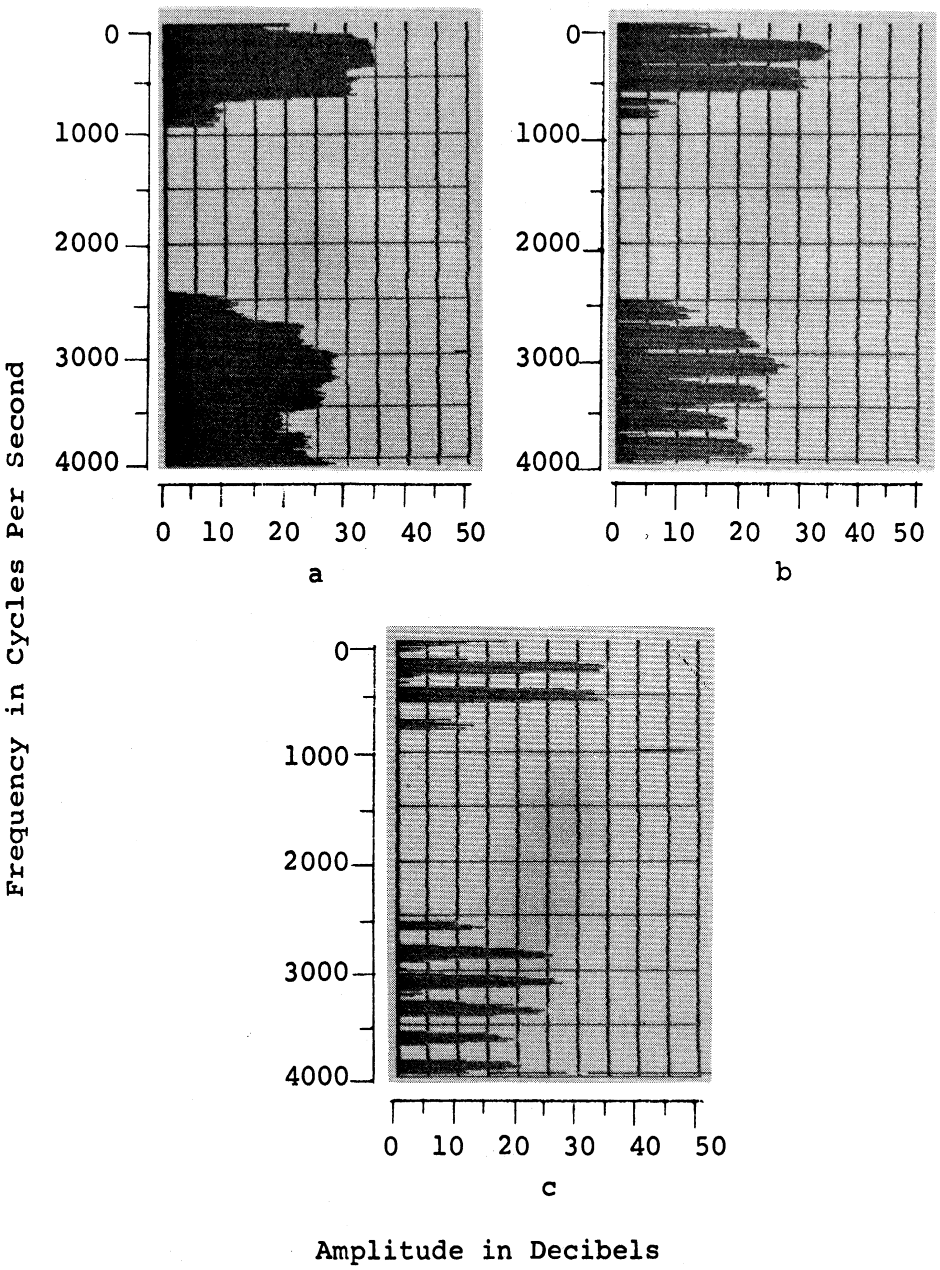


Figure 7.11. Amplitude sections made at the indicated point in the spectrogram of Figure 7.10, but with flat-top filters having equivalent bandwidths of 300, 200, and 100 cps.

In the two sections with the narrower filters, the individual harmonics of the voice fundamental frequency appear; these harmonics are more clearly separated in the section with the narrowest filter (100 cps).

In Figures 7.12 and 7.13 the same time interval of the same speech sample (as for Figures 7.10 and 7.11) was used, but the filters were changed to the linear-phase type,* for both the spectrograms and the sections. A fourth section at 50 cps bandwidth is shown in Figure 7.13, along with sections made with the same three filters used to construct Figure 7.11. The absence of sharp shoulders in the wider band sections seems to provide a truer representation of the envelope of the actual spectral distribution than was provided by the flat-top filters. The amplitude envelopes in the figures (particularly for the wide-band filters) are smooth curves, except for a small variation which is due to random noise.

Continuous amplitude display

The third kind of record produced by the CSL

---

*The equivalent pass-band approximates a Gaussian-type filter.

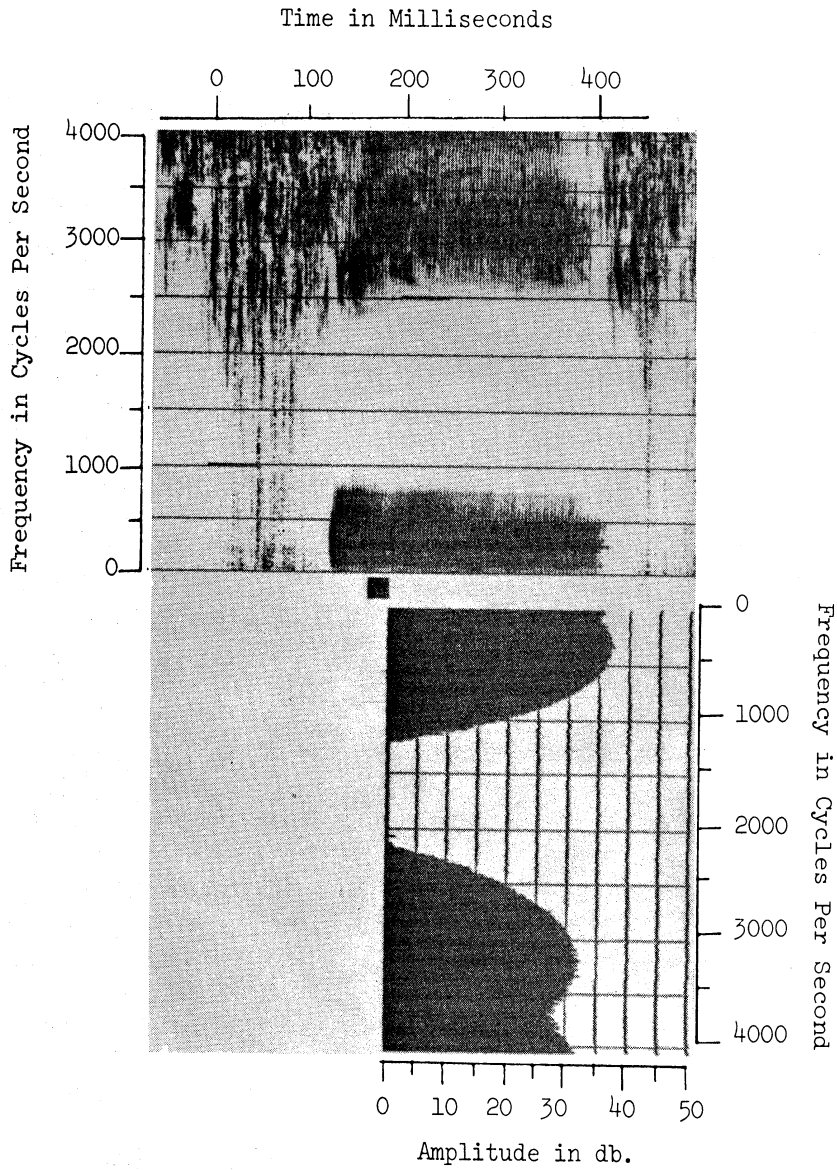


Figure 7.12. Spectrogram and amplitude section of the same speech sample used for Figure 7.10, but made with a linear-phase filter of 500 cps equivalent bandwidth.

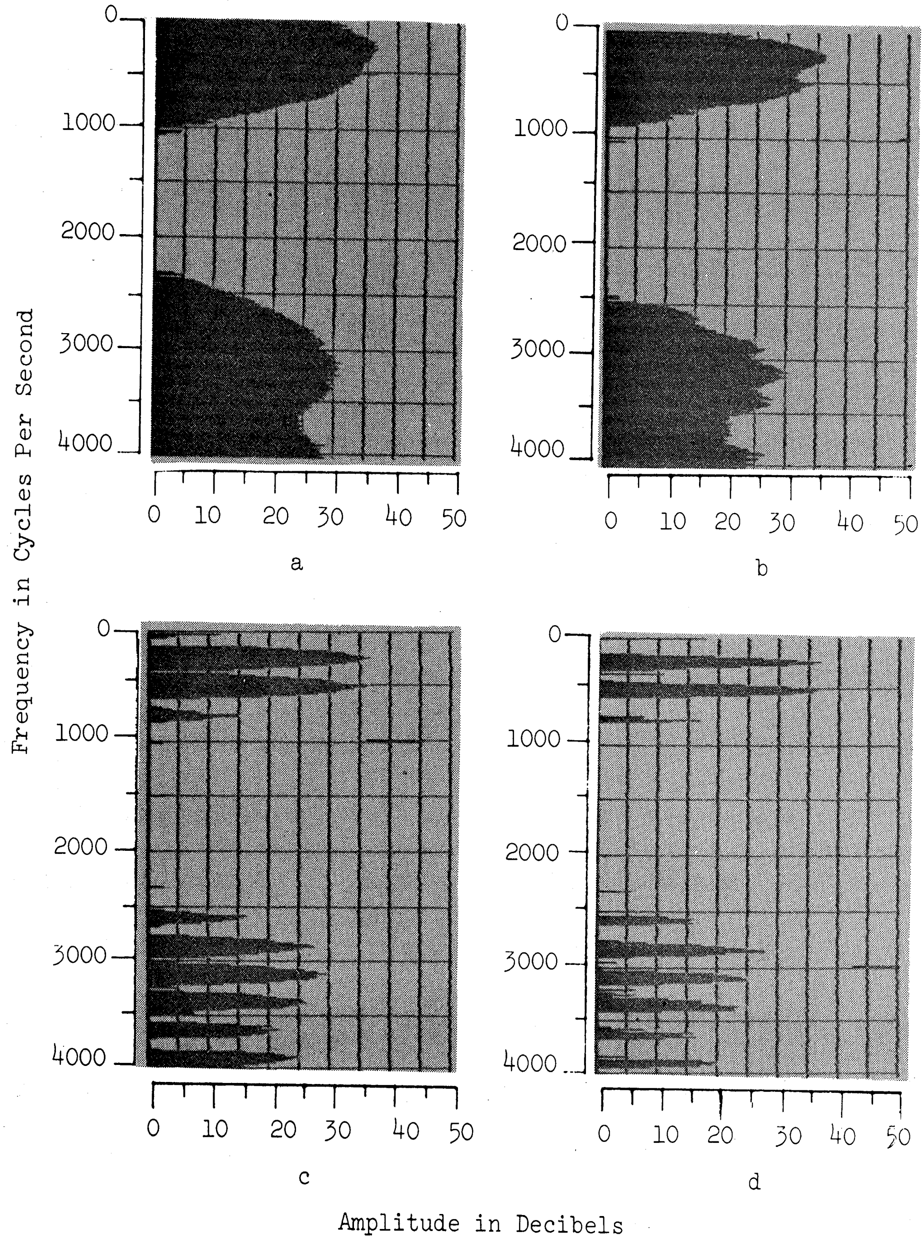


Figure 7.13. Amplitude sections of the same speech sample used for Figure 7.12, made with linear-phase filters having equivalent bandwidths of 300, 200, 100, and 50 cps.

sound spectrograph is called a continuous amplitude display. As we have pointed out in Chapters 5 and 6, this type of display is a two-dimensional graph which plots the amplitude of the entire spectrum versus time. Either an envelope detector or an averaging detector* (i.e., low-pass filter) may be used in the continuous display network, and the amplitude scale may be either linear or logarithmic. Figure 7.14 shows a typical speech wide-band sound spectrogram and its continuous amplitude display on a linear scale. The smallest time constant (8 ms) of the envelope detector was used. This figure shows the results of a scan from high amplitudes to low, through the rectified positive side of the wave, followed by a scan from low to high amplitudes through the negative side of the wave. The effect produced is like that of an oscillogram. The dissymmetry of positive and negative peaks is a fact often noticed in speech oscillograms. A strong low frequency noise of about 14 cps appears in the continuous amplitude display. Presumably this noise was present but unnoticed in the room at the time the speech recording was made.

---

*These detector types are described in Chapter 5.

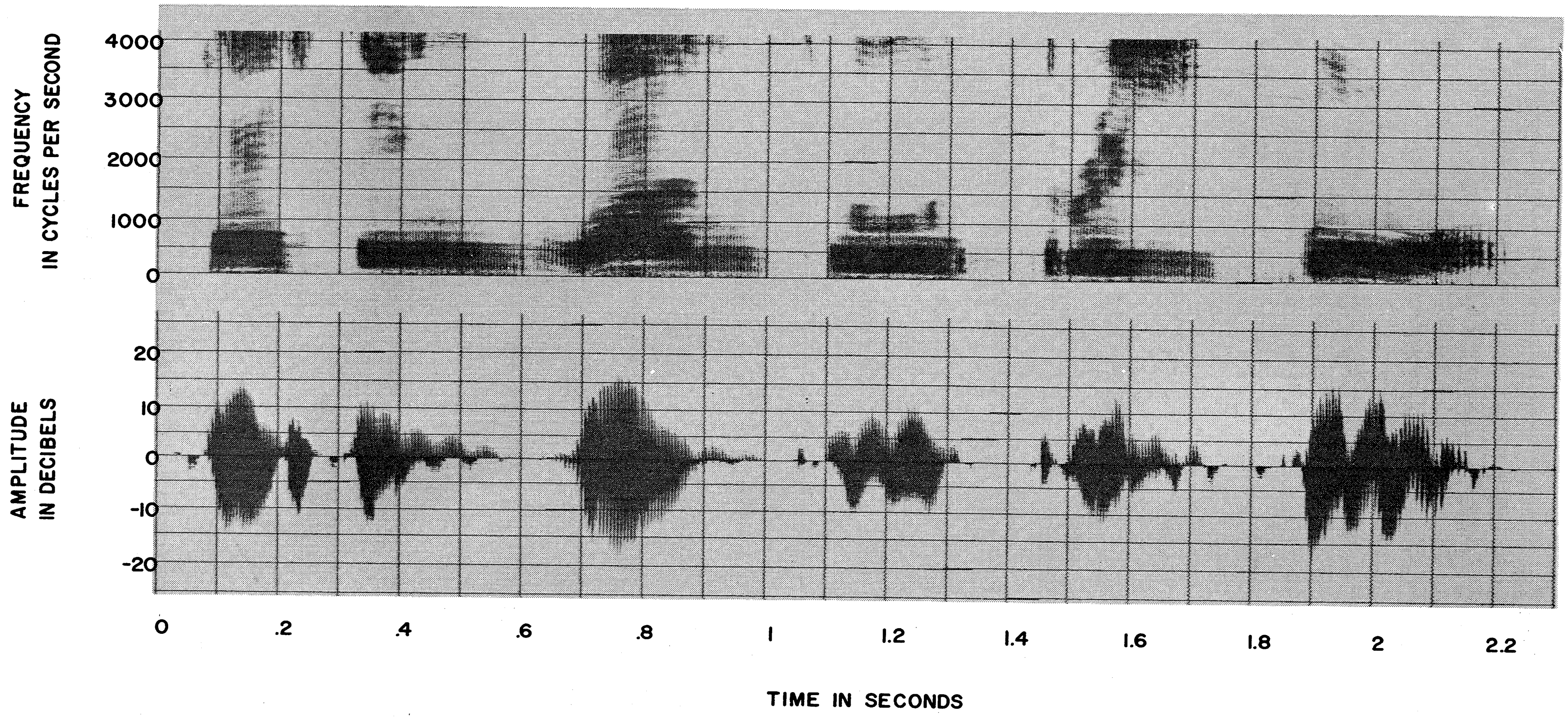


Figure 7.14. A wide band spectrogram of a male voice saying "Testing one two three four", and a continuous amplitude display having a linear amplitude scale. The envelope detector was set at the smallest time constant (8 ms).

The flexibility of the spectrograph control makes it possible to interchange the positive-negative polarities of the amplitude display, or to combine them by means of the full-wave rectifier. Examples of half-wave and full-wave continuous amplitude displays of a 46.7 cps sinusoidal signal are shown in Figure 7.15a and 7.15b respectively. The positive and negative alternations of this low frequency wave are easily distinguished in Figure 7.15a. The doubling of frequency in full-wave rectification is seen in Figure 7.15b.

The effect is somewhat different when the input is a complex wave like unfiltered speech, shown in Figures 7.16a and 7.16b. Whether the rectification is positive half-wave, negative half-wave, or full-wave, the peaks of the display show the concentration of energy in the fundamental periods of the signal. There is no easily identified or consistent time difference in these peaks between positive and negative displays, and no apparent doubling of peaks in full-wave rectification.

Figures 7.17, 7.18, and 7.19 show a set of the continuous amplitude displays, which were made by feeding the half-wave rectified speech wave to an envelope detector. Progressively larger

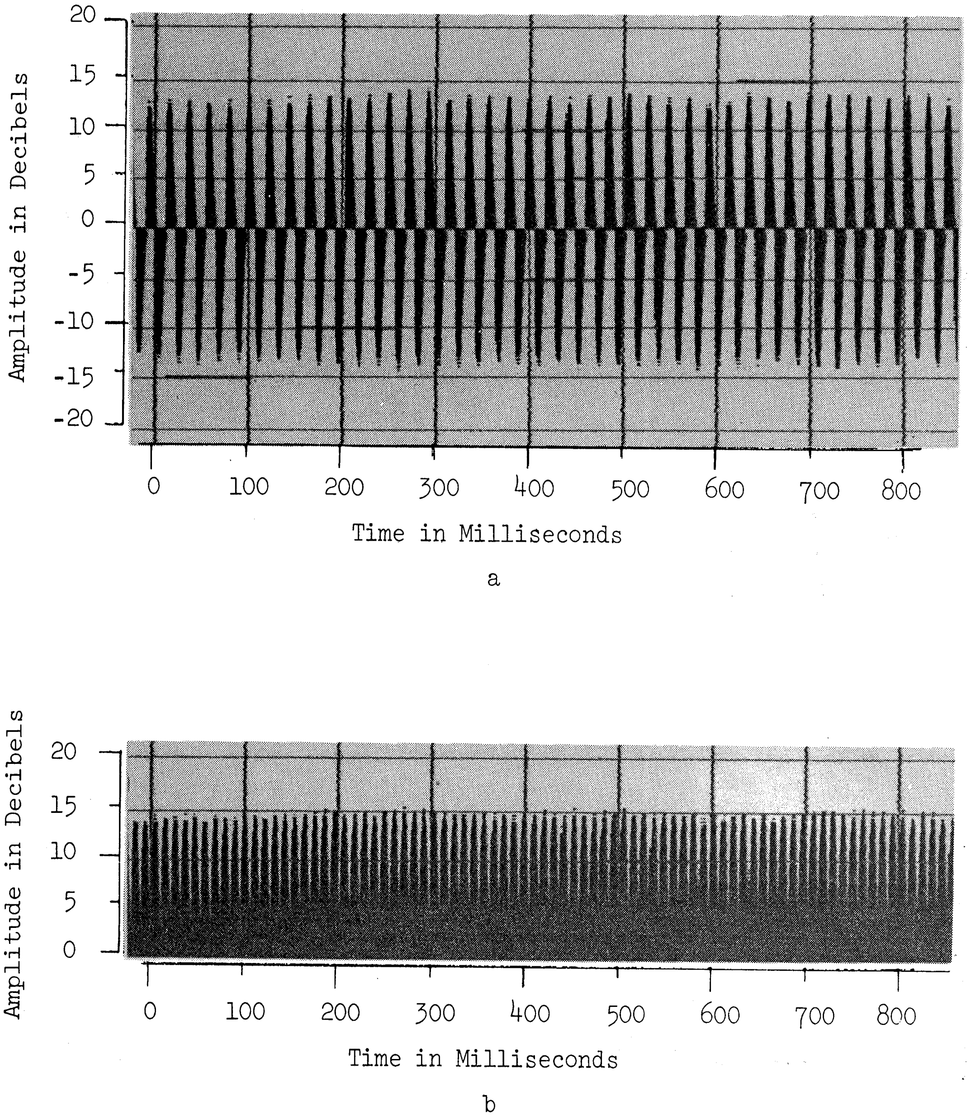
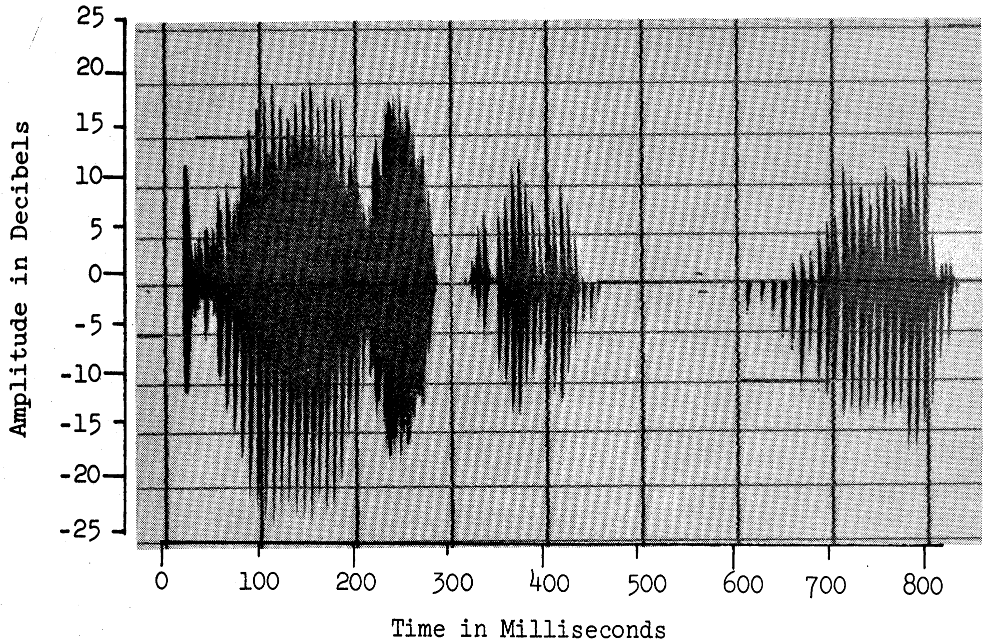
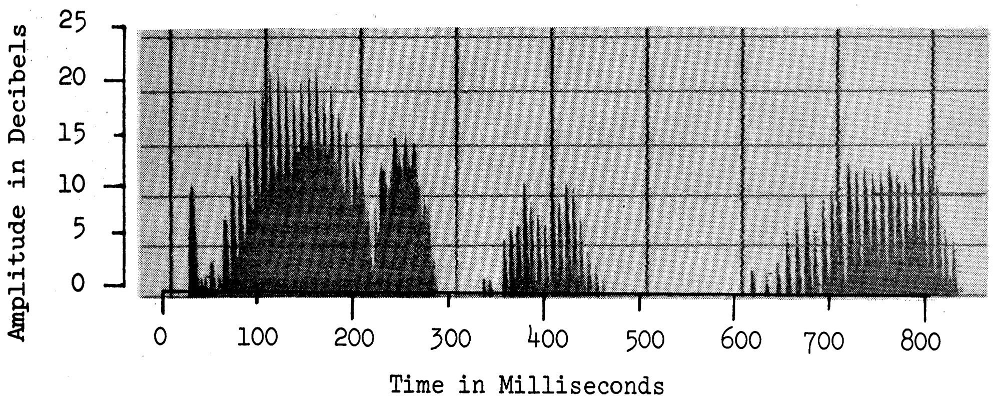


Figure 7.15. Continuous amplitude displays of a sine wave of 46.7 cps frequency. In (a) the scan was from high amplitude to low through the positive half-wave side, followed by low to high through the negative side. In (b) full-wave rectification was used.



a



b

Figure 7.16. Continuous amplitude displays of a speech sample, showing (a) the positive and negative sides of the signal and (b) the full-wave rectification.

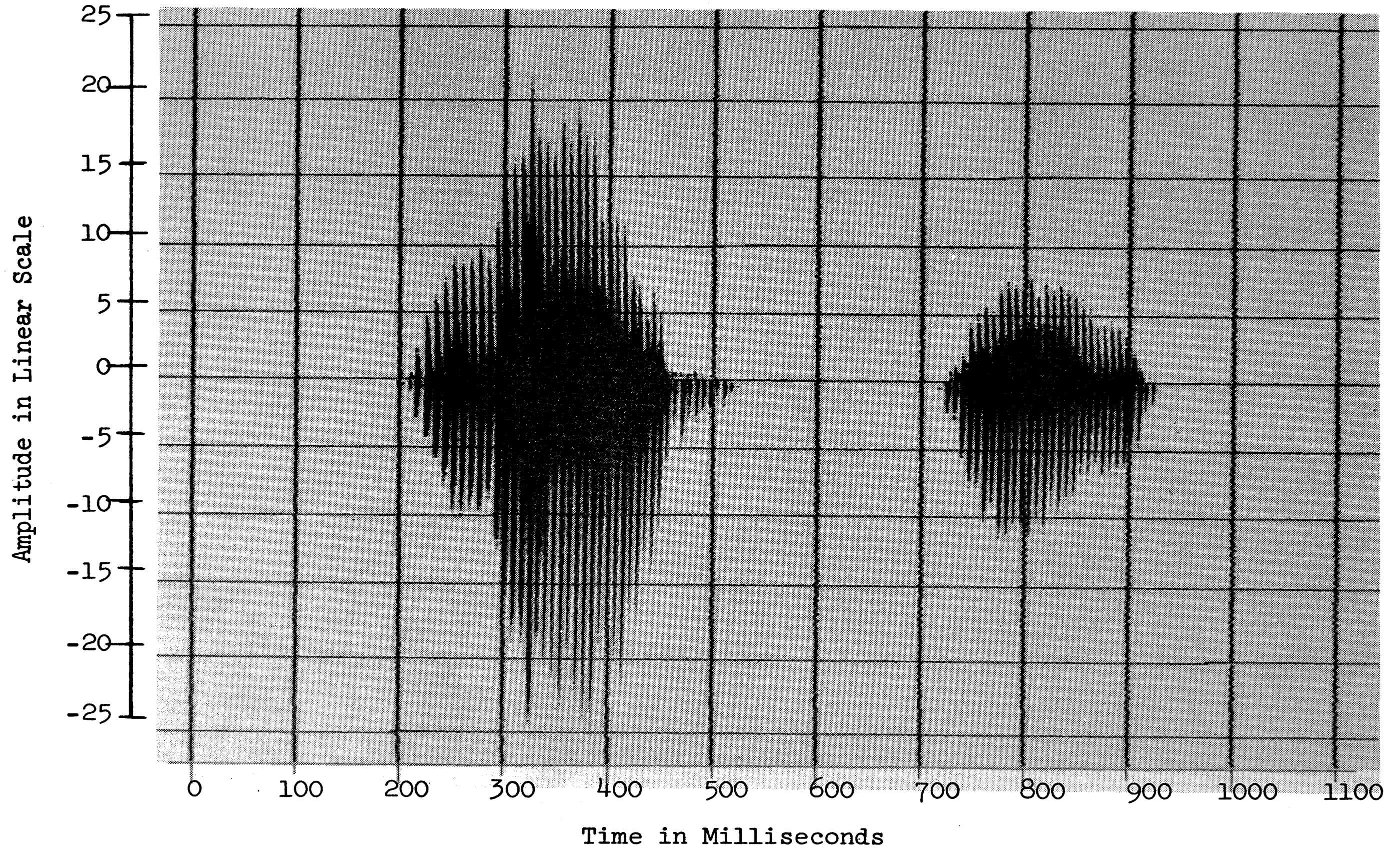


Figure 7.17. Continuous amplitude display of a sample of speech, using a linear amplitude scale. The recovery time constant of the envelope detector was 8 milliseconds.

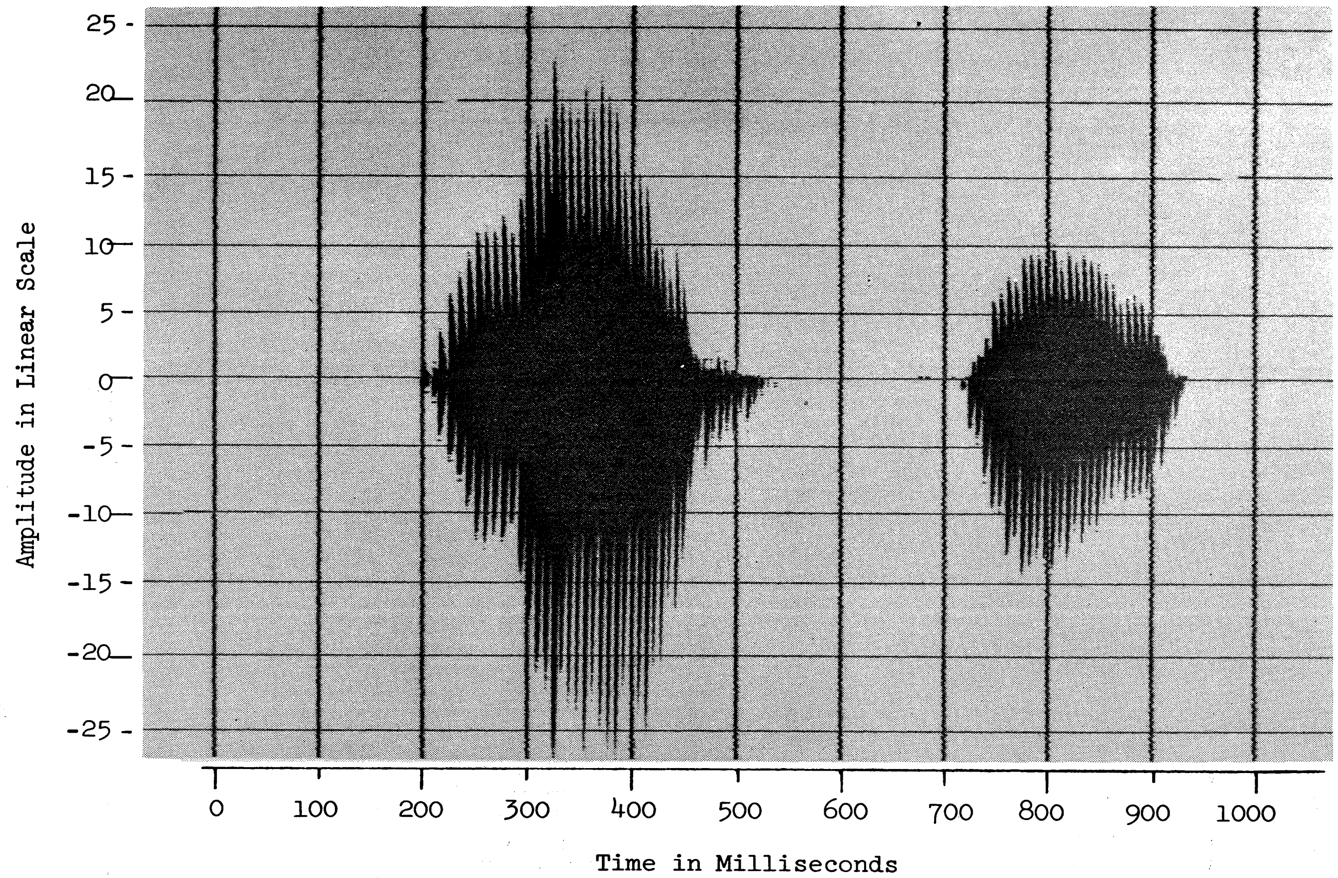


Figure 7.18. The same analysis and speech sample shown in Figure 7.17, except that the time constant was 14 ms.

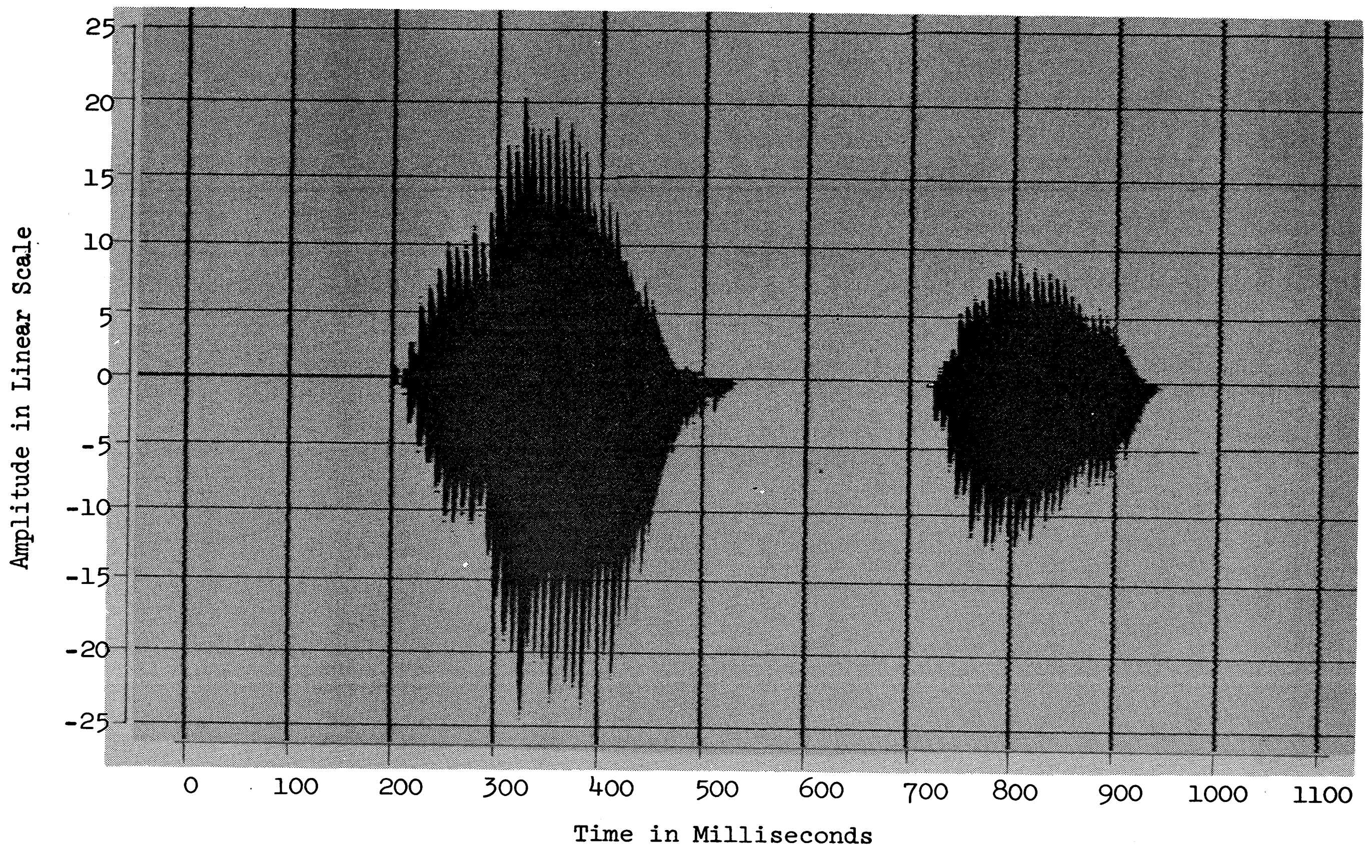


Figure 7.19. The same analysis and speech sample shown in Figure 7.17, except that the time constant was 38 ms.

recovery-time constants were used for the three figures. When the signal amplitude is falling rapidly, as on the right sides of the two words in these figures, the marked contour tends to follow the slower detector recovery curve if the circuit time constant is large. This is seen particularly in Figure 7.19.

In Figures 7.20, 7.21, and 7.22 the same speech sample was used as in the previous figures. The amplitudes, however, are in decibels instead of a linear scale, through the use of the logarithmic circuit.* Different time constants are again used in the three figures, and the distortion caused by a longer constant is again obvious, in Figure 7.22.

In Figures 7.23 and 7.24 a different pair of words is used, and the detector for the display is the low-pass filter (i.e., the averaging detector). Four different values of filter cutoff were employed in construcing the two figures. If the bandwidth of the low-pass filter is narrow enough to discriminate against the ac components of the rectified wave, then the output amplitude display will be the average value of the rectification. Thus

---

*See "logarithmic amplitude display", Chapter 5.

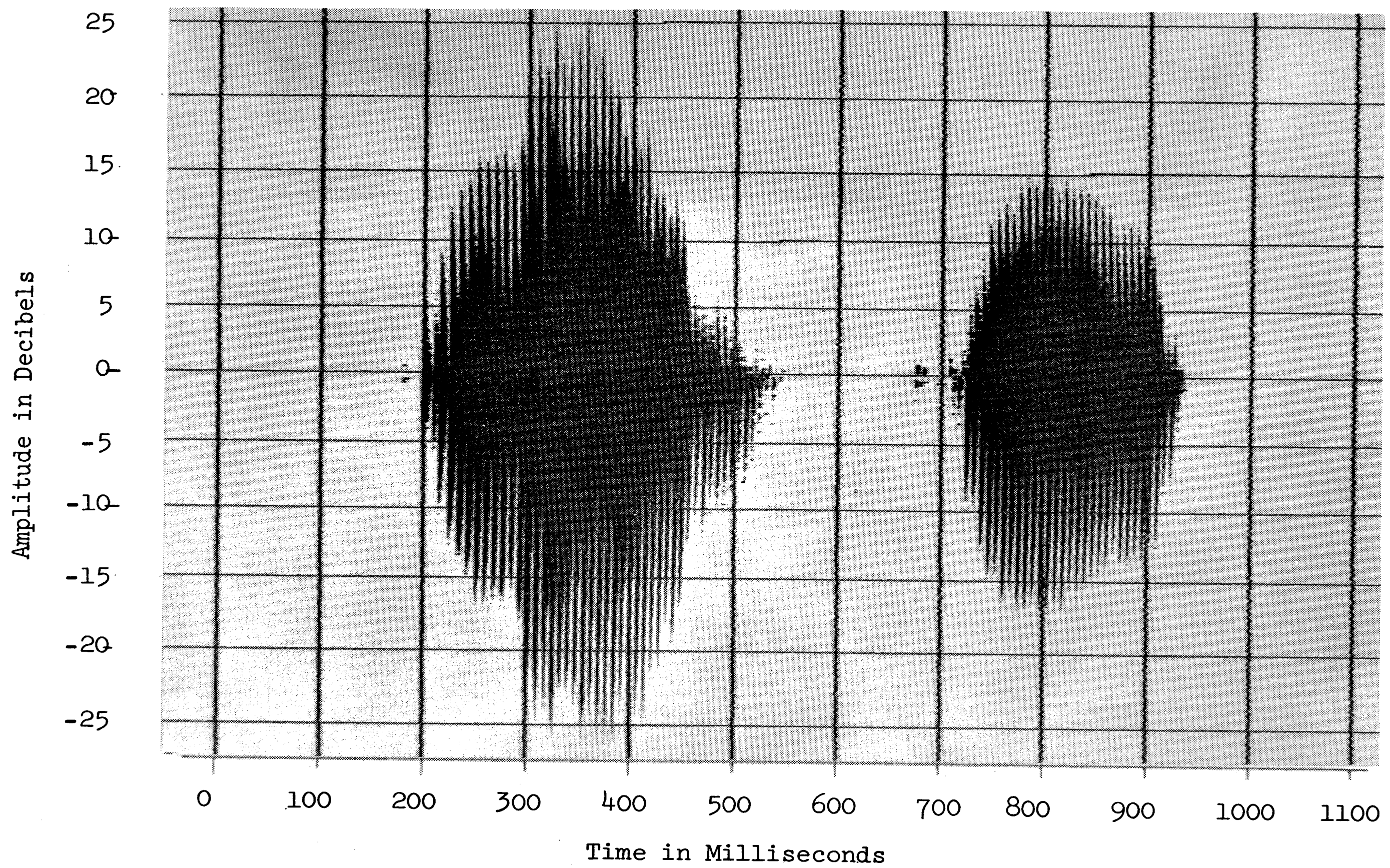


Figure 7.20. A continuous amplitude display with the same speech sample used for Figure 7.17, but with a logarithmic amplitude scale. The recovery time constant of the envelope detector was 8 ms.

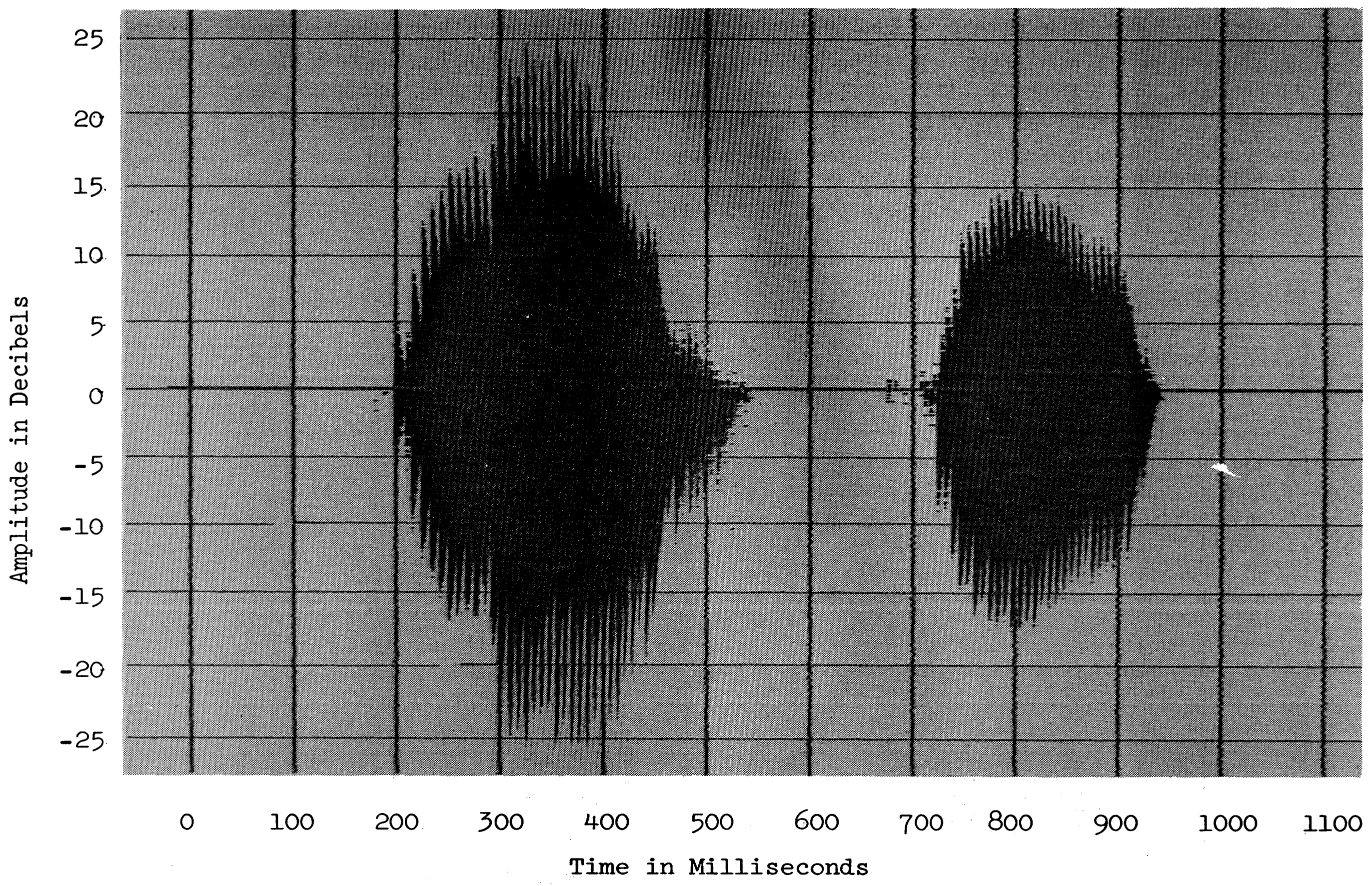


Figure 7.21. The same analysis and speech sample shown in Figure 7.20, except that the time constant was 20 ms.

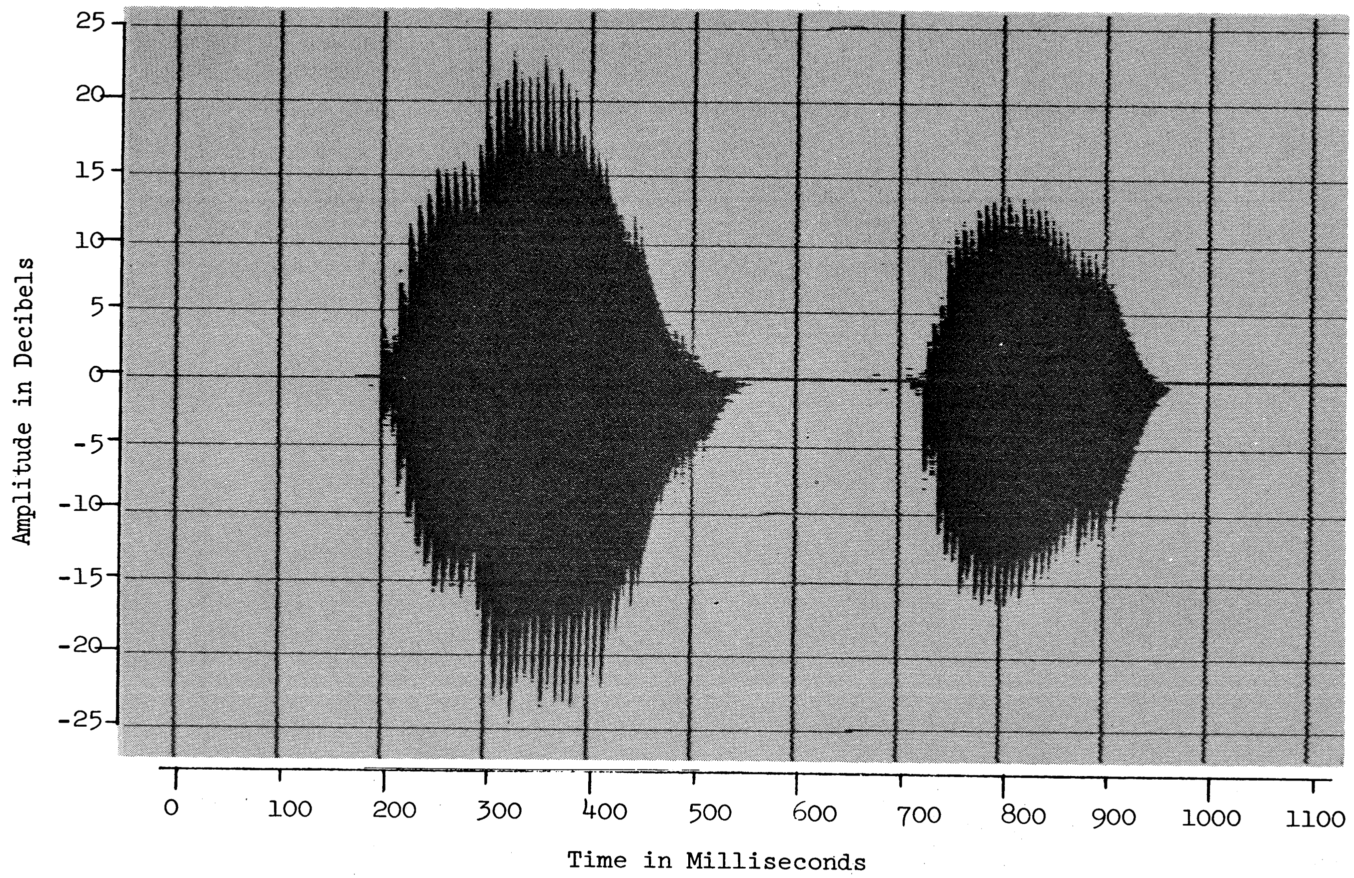


Figure 7.22. The same analysis and speech sample shown in Figure 7.20, except that the time constant was 44 ms.

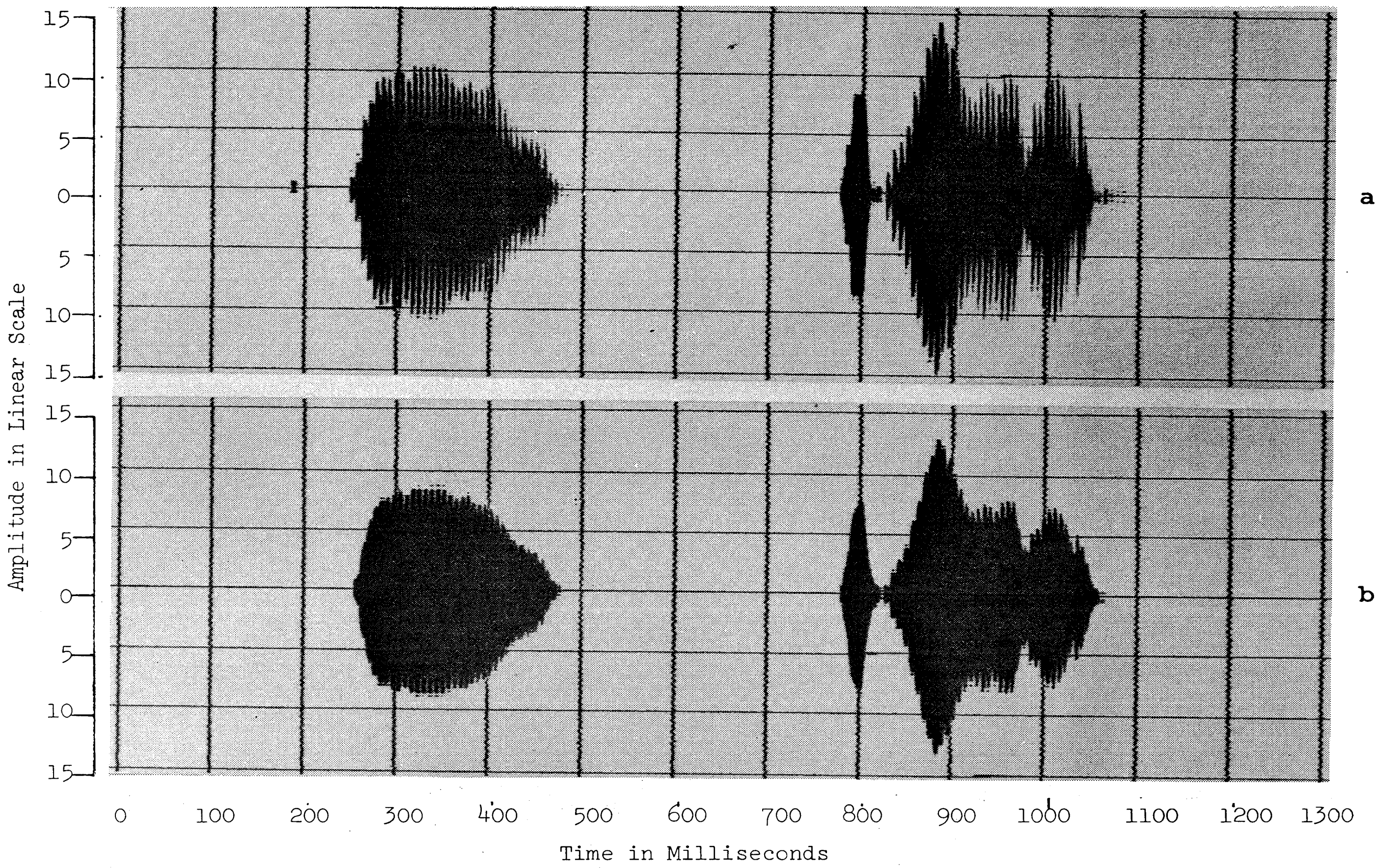


Figure 7.23. Continuous amplitude displays of a sample of speech, using a linear amplitude scale. In (a) the cut-off of the averaging detector was 2500 cps, in (b) it was 500 cps.

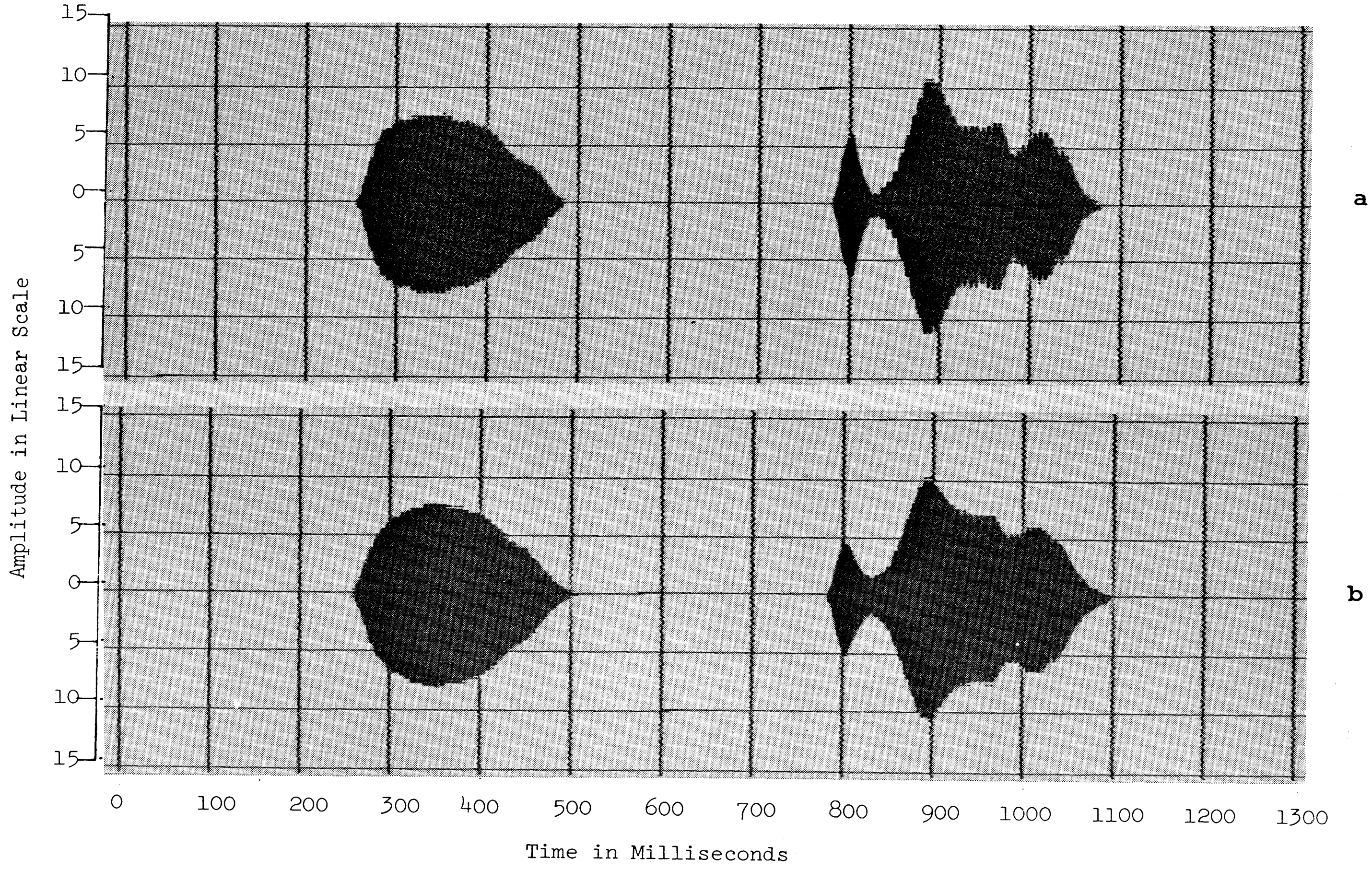


Figure 7.24. The same analysis and speech sample shown in Figure 7.23, except that the cut-off in (a) was 250 cps and in (b) it was 125 cps.

this type of amplitude display represents the average power flow of the speech wave. It can be seen from the figures that the narrower the bandwidth (within the range shown), the closer the amplitude display approaches the average value of the rectified signal.

Figures 7.25 and 7.26 repeat the analyses of the previous two figures, but with logarithmic instead of linear amplitude scales.

Switching operation of the dual channels

One of the features of the CSL sound spectrograph is the dual channel system. The two channels can be analyzed separately, simultaneously, or even alternatively with respect to each other. The alternative switching activity between the channels is controlled by a gating network which is activated by pulses from a third tape record. These pulses are previously recorded by the operator, as was explained in Chapter 6. The selection of the pulse positions is made by the operator, usually after first making separate spectrograms from both channels. The relation of the signals may be adjusted before selecting the positions of the switching pulses by examining the two spectrograms and changing

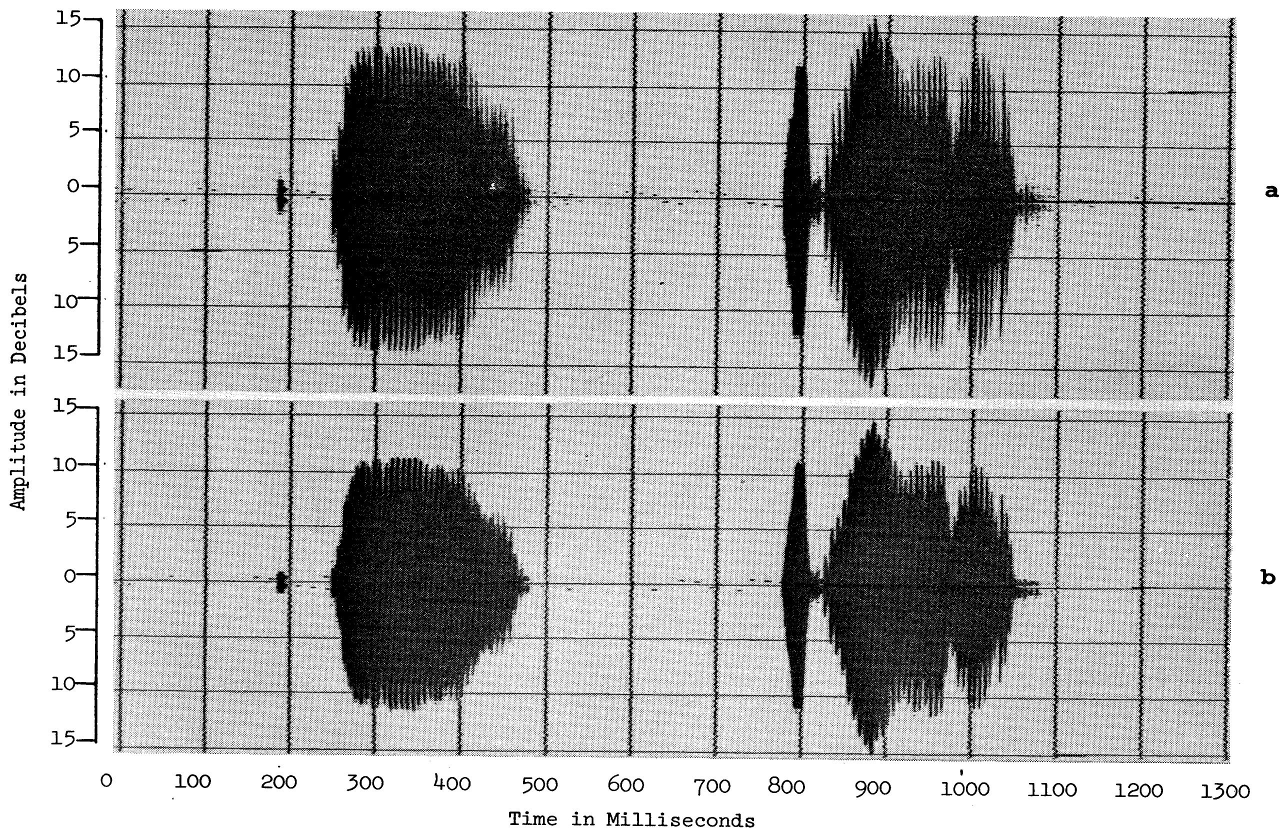


Figure 7.25. Continuous amplitude displays for the same sample of speech used in constructing Figure 7.23, but with a logarithmic scale of amplitudes. The averaging detector cut-off was (a) 2500 cps, and (b) 500 cps.

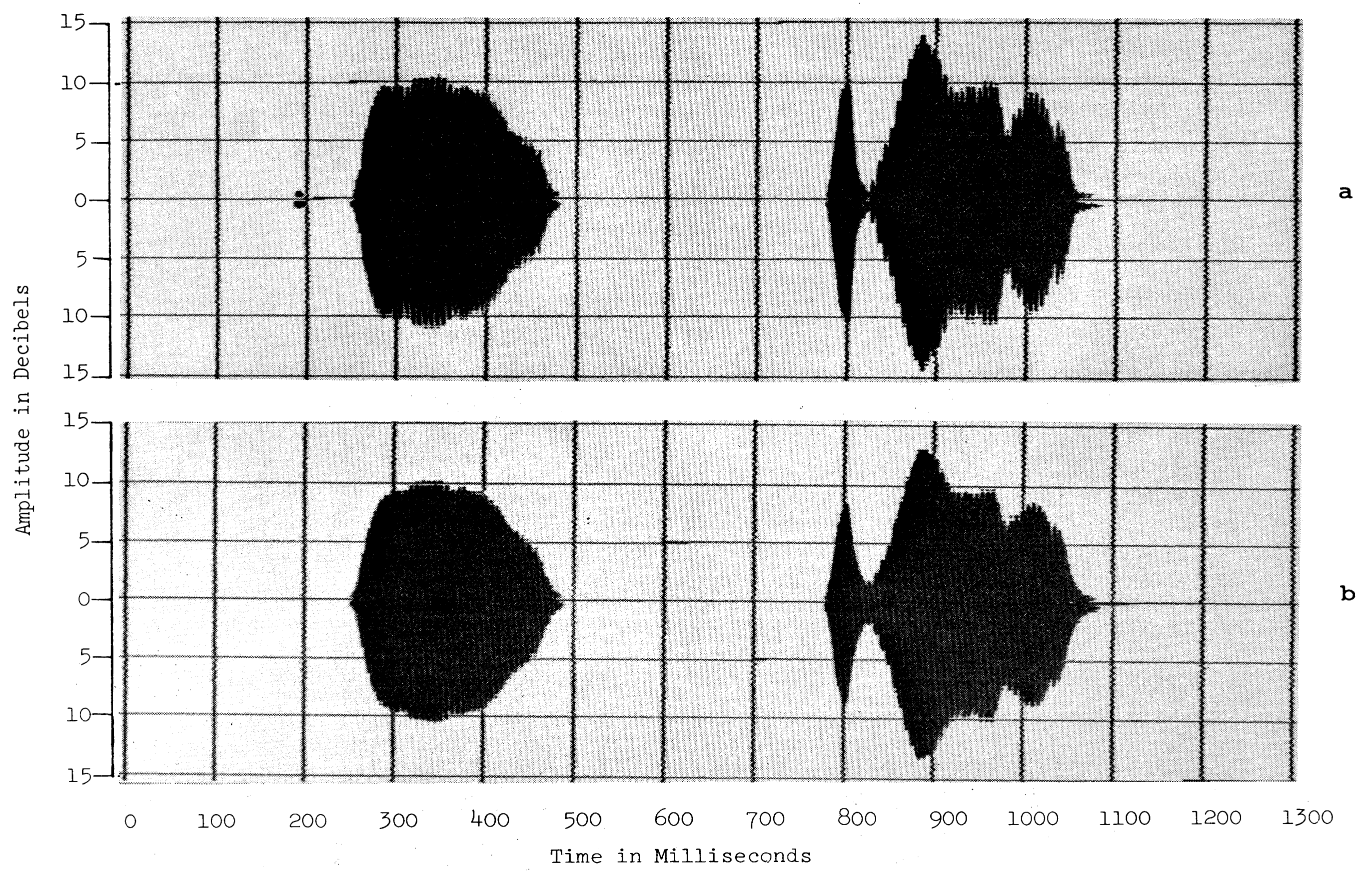


Figure 7.26. The same analysis and speech sample as shown in Figure 7.25, except that the averaging detector cut-off for (a) was 250 cps and for (b) it was 125 cps.

the angular position of one channel tape with respect to the other. Figure 7.27 shows a wide-band sound spectrogram produced by switching between the two channels. The time intervals occupied by each channel, in the bottom spectrogram, are identified below it. Separate spectrograms of the two channels are given above.

The channel-switching technique can also be applied to select sections of a single channel, for analysis. With accurately placed pulses, the gating network can control operations effectively even in very short intervals. This can be seen in Figure 7.28, where five sections of the upper spectrogram are selected for a separate display in the lower. Perhaps more significant than the visual presentation shown in this figure, is the fact that the separated sections can be heard in the monitoring system of the spectrograph or recorded on a separate tape for listening experiments. One of the original purposes of developing this circuit arrangement was to obtain dyad segments of natural utterances for synthesizing continuous speech.

In Figure 7.29 separate spectrograms from the two channels are again shown, and in the third spectrogram the channels are combined. No use of gates was made in this case.

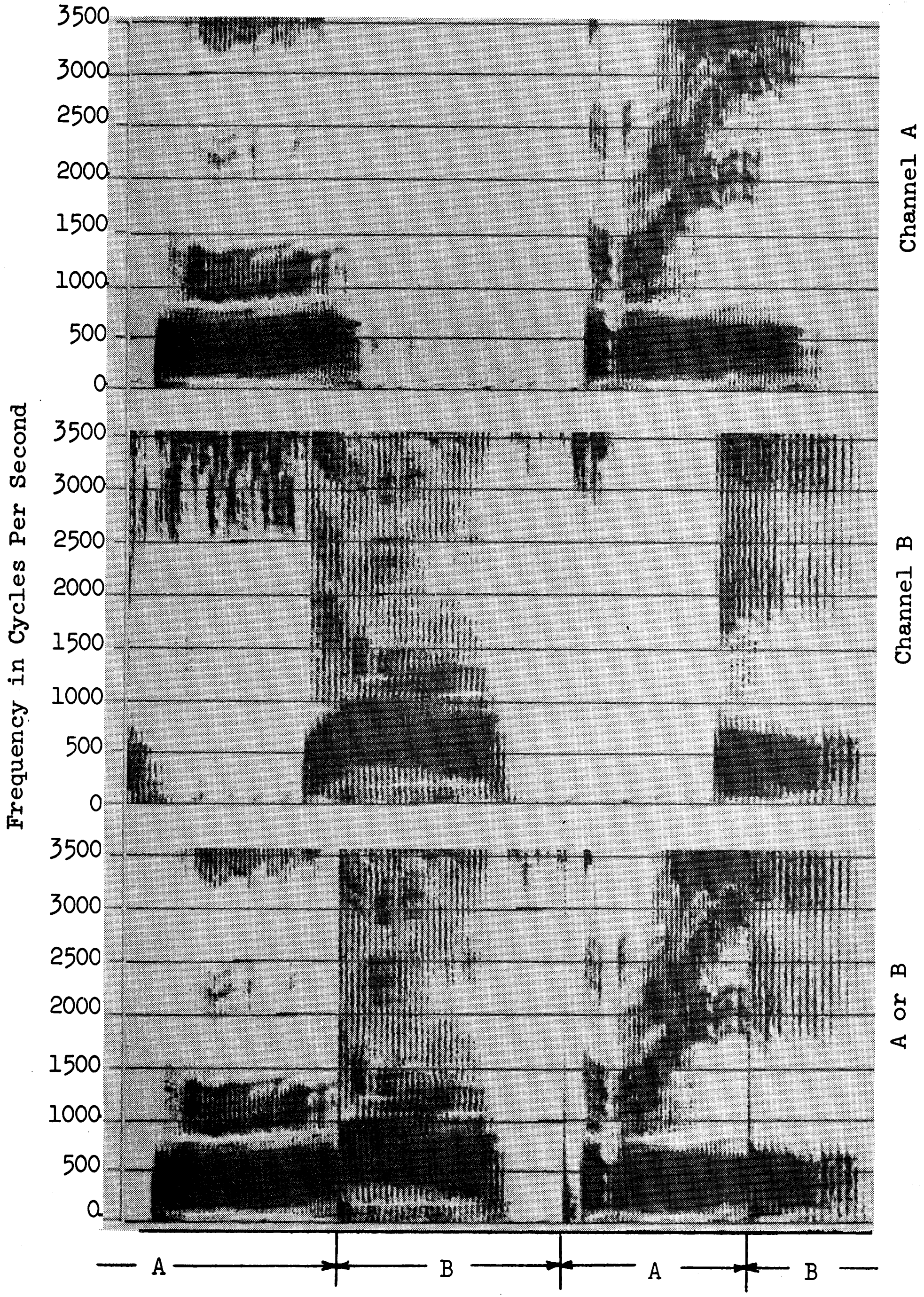


Figure 7.27. The two upper spectrograms were made separately from speech samples recorded in channels A and B of the spectrograph. The lowest spectrogram is a mixed sample made by pulsed switching between channels at the points indicated.

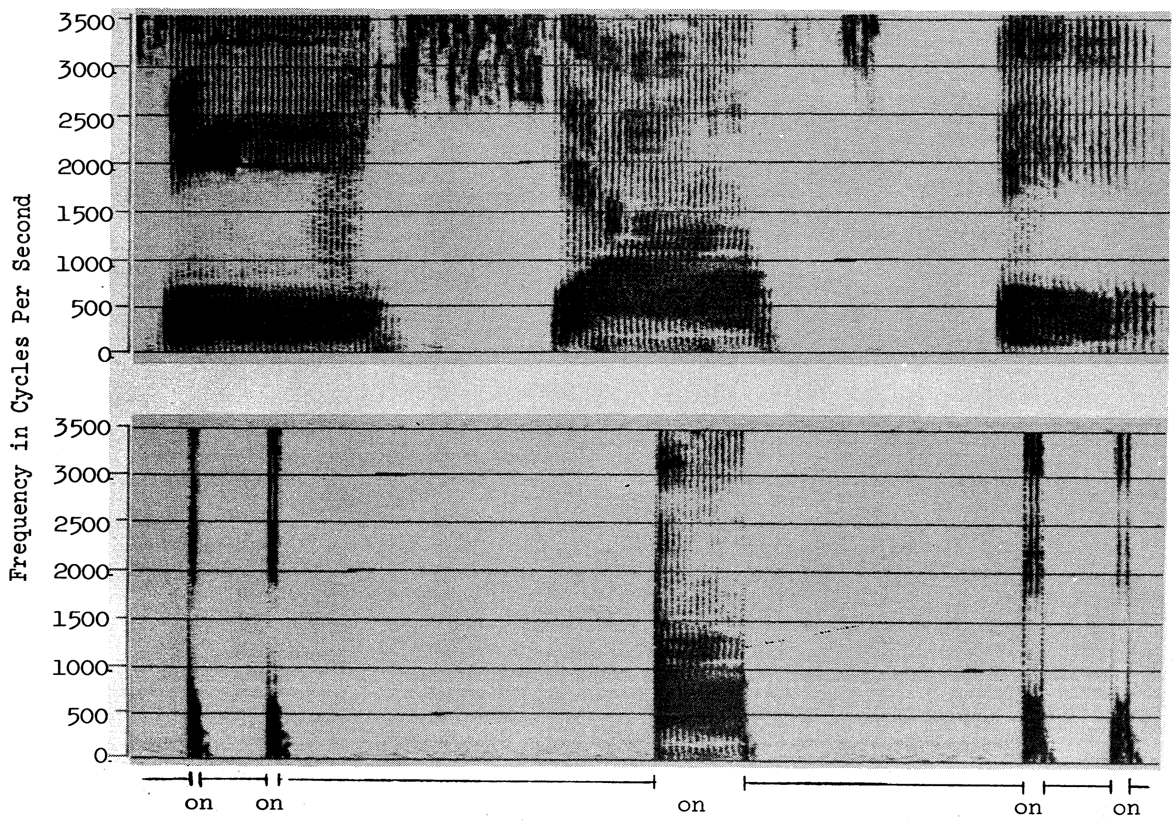


Figure 7.28. Short time sections were isolated from the speech sample shown in the upper spectrogram by means of pulsed switching; the isolated sections are shown in the lower spectrogram.

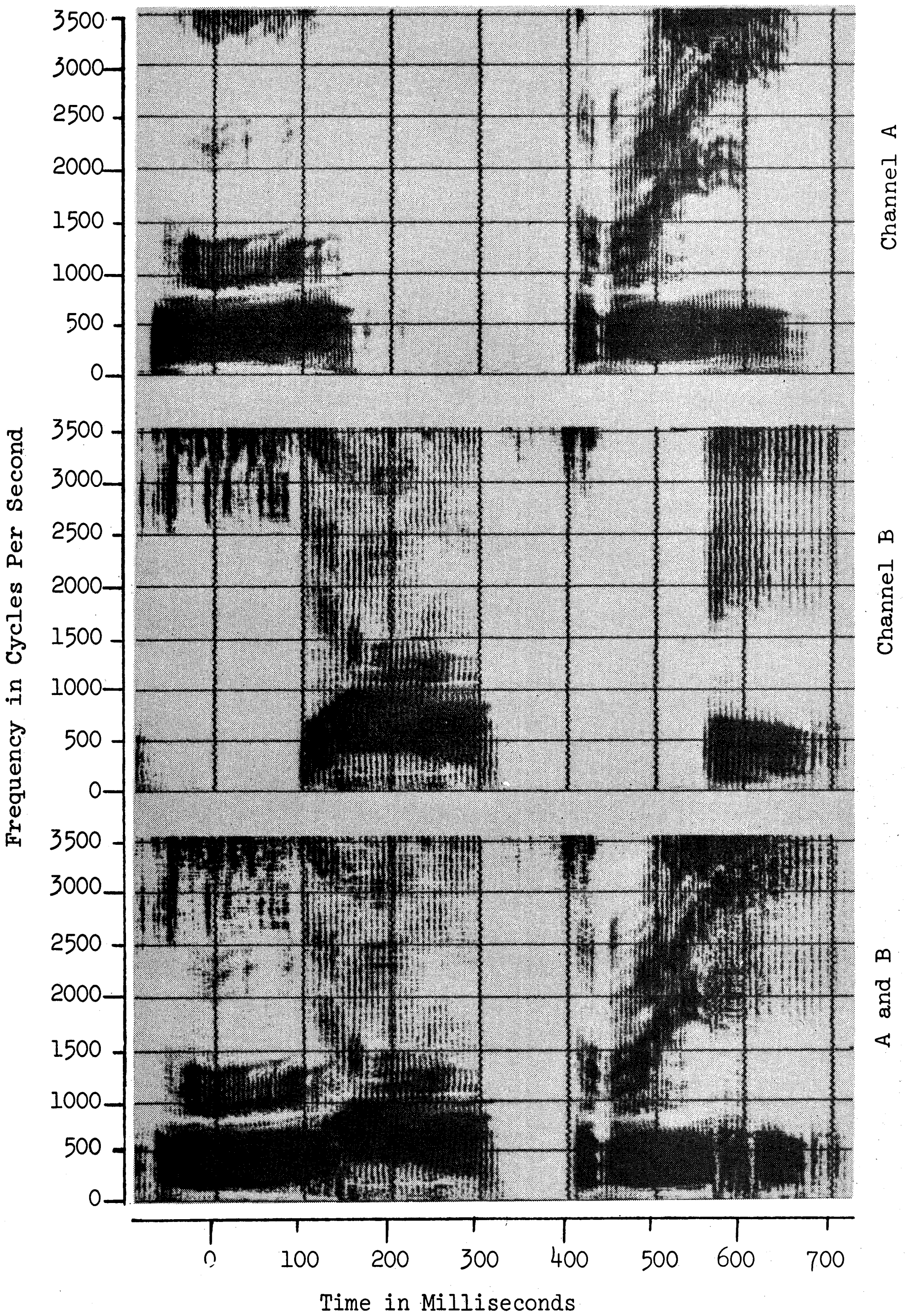


Figure 7.29. The two channels may be analyzed separately (as illustrated above), or completely superimposed (as illustrated below).

Automatic scaling

Another feature of the CSL sound spectrograph is the automatic scaling of a spectrogram. Scale lines marked on all previous figures were produced simultaneously with the spectrograms. The CSL sound spectrograms in general can be provided with frequency, time and amplitude lines.

a. Frequency scale

There are three selections of the automatic frequency lines: 1000, 500, and 200 cps spacings. Examples of these three frequency lines are given in Figures 7.30a, 7.30b, and 7.30c respectively.

b. Time scale

The same figures show the three time markings provided: 200, 100, and 20 millisecond spacings. In addition to these time lines produced with the spectrogram, a scale can be placed alongside the spectrogram as shown at the bottom of Figure 7.30b. In this scale every fifth of the 20-millisecond marks is automatically made slightly longer, and every tenth mark longer still.

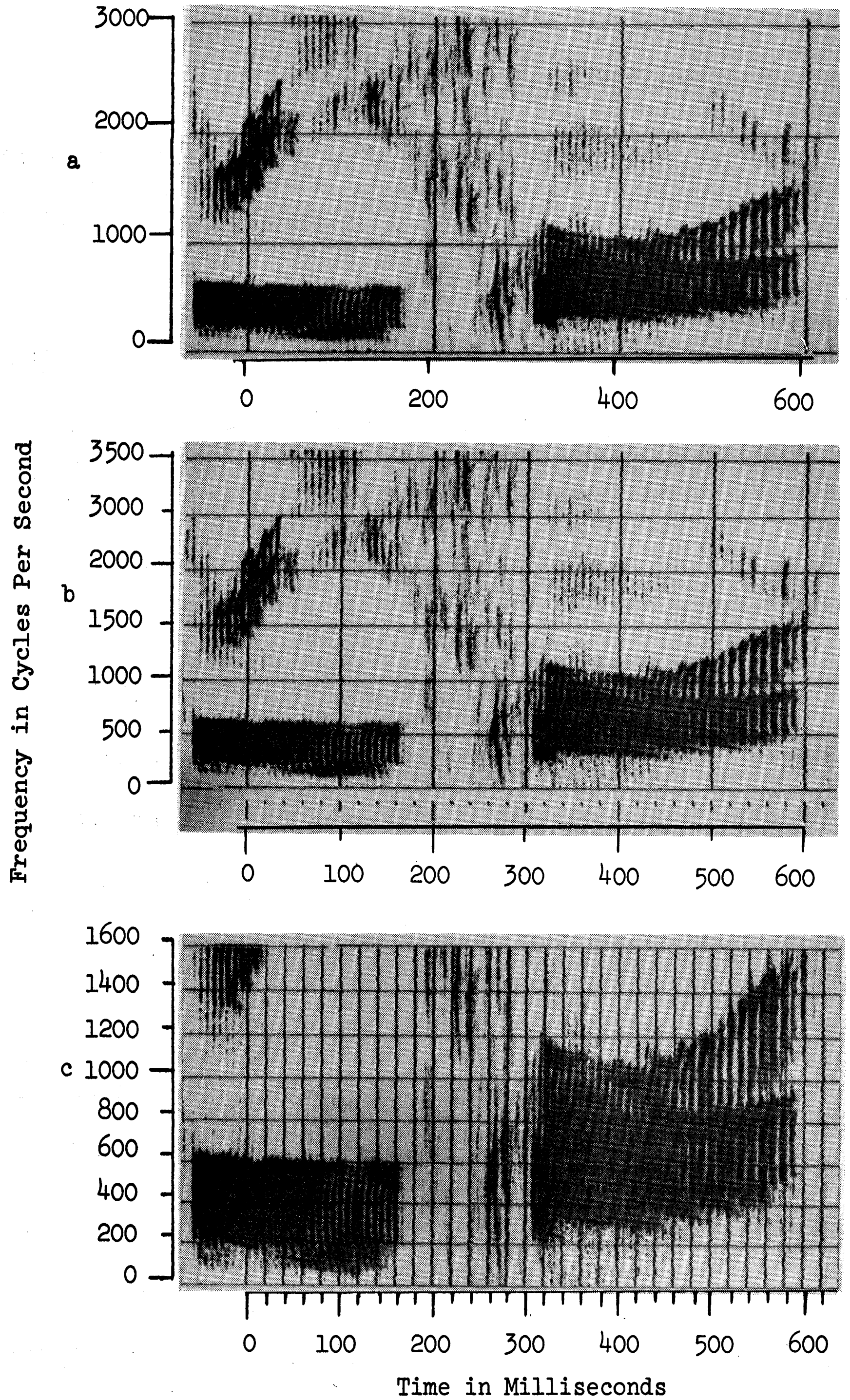


Figure 7.30. Scale lines are drawn automatically on the spectrograms: (a) at 200 milliseconds and 1000 cps spacings, (b) at 100 ms and 500 cps, (c) at 20 ms and 200 cps. A separate time scale can also be marked beside the spectrogram as shown below (b).

c. Amplitude scale

The CSL sound spectrograph can also produce 10 db and 5 db amplitude lines for the continuous amplitude display, and 5 db lines for the amplitude sections. Examples of these are shown in Figures 7.31a, 7.31b, and 7.31c.

The frequency scale for the amplitude section can be made either across the section only, outside the section only, or in full across both spaces. These three possibilities are shown in Figure 7.32.

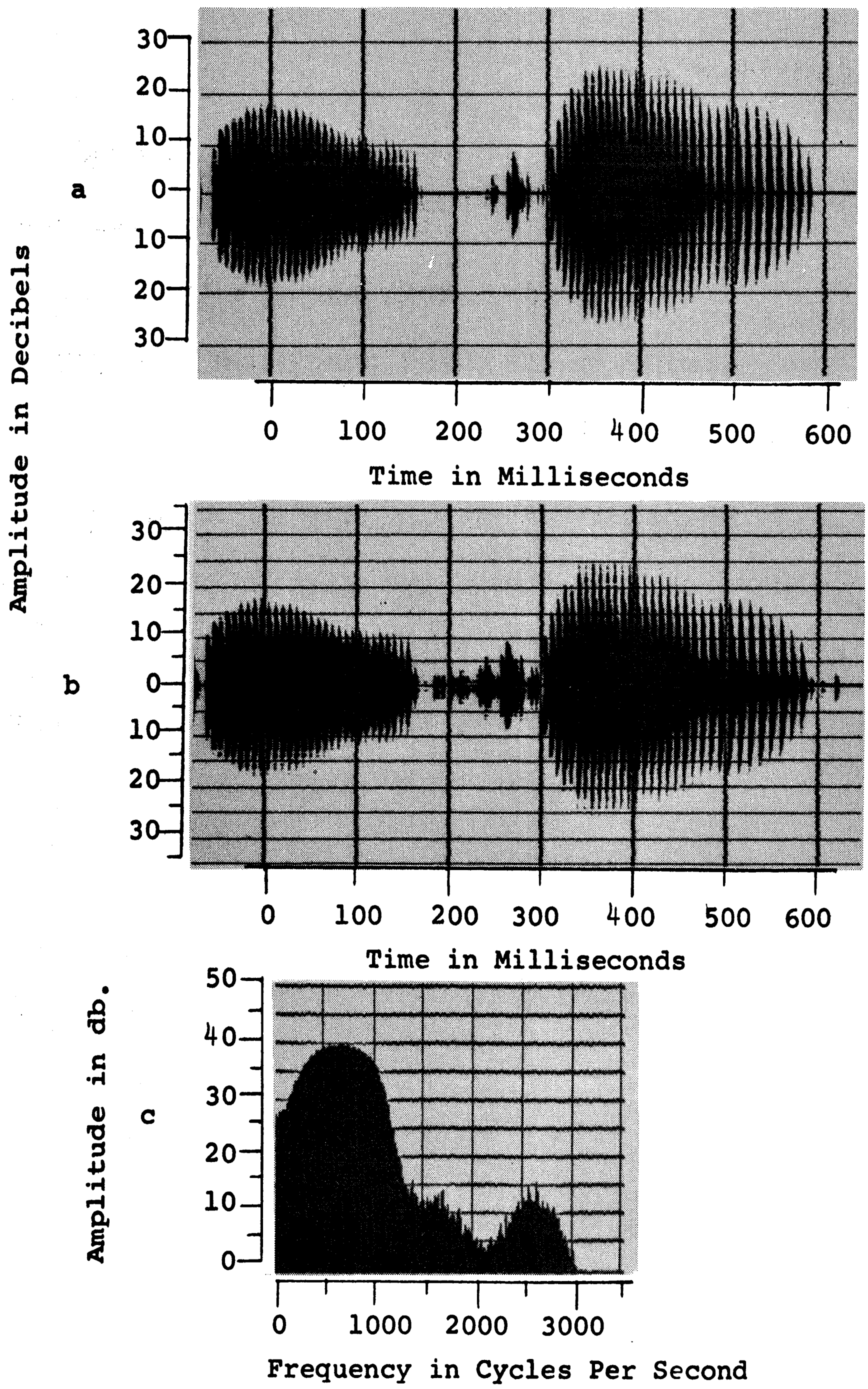


Figure 7.31. In (a) and (b) time lines were drawn automatically at 100 ms intervals on a continuous amplitude display. Amplitude lines were also drawn at 10 db intervals in (a) and at 5 db intervals in (b). On the amplitude section in (c) both frequency and amplitude lines were drawn.

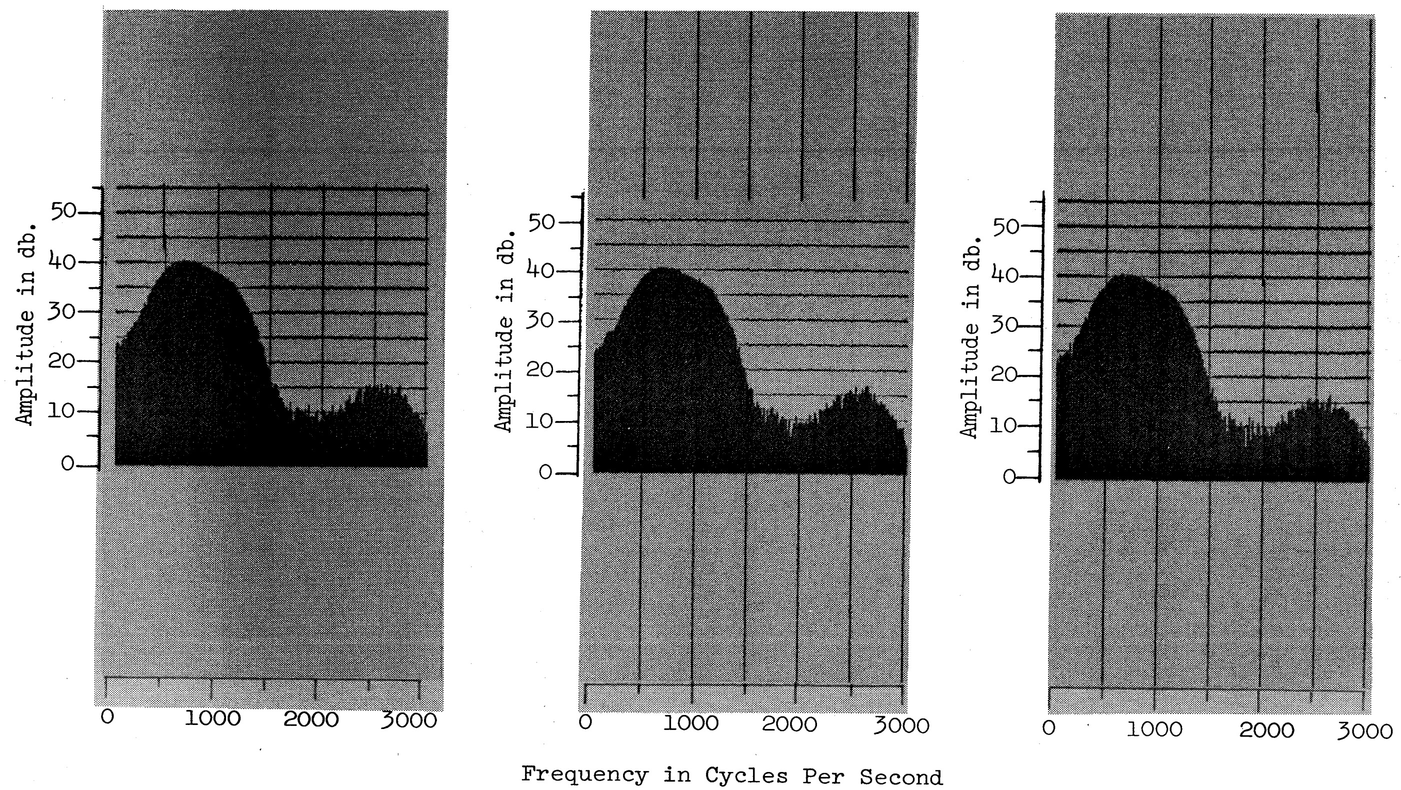


Figure 7.32. In an amplitude section, the frequency lines may be shown on the section only, outside the section only, or across the entire paper.

8. Further Spectrograph Design Considerations

As outlined at the beginning of Chapter 6, the CSL sound spectrograph was designed to have certain features not available in other spectrographs. These have been described in Chapter 6, and the examples shown in Chapter 7 illustrate the extent to which the instrument accomplishes the original purpose. There are some respects, however, in which improvements could be made in the instrument.

Originally, the CSL spectrograph was to be supplied with band-pass filters for analysis. However, because of the high frequency (100 kcps) at which it appeared necessary to operate the filters and because of the difficulty of achieving the desired filter characteristics at this frequency, attention has been concentrated on equivalent low-pass filters. As we have seen in Chapter 5, there are some theoretical difficulties with the use of such low-pass filters, which are partially alleviated by using multiple sets of filter circuits, having diversified phase constants.

If for no other reason than to make the entire circuit less complex, the use of band-pass filters would be desirable. In this connection the suggestion in Chapter 5, pp. 64-67, regarding the lowest practical frequency location of the band-pass filters, should be considered. If (as at present) the reproduced frequencies extend from zero to 60 kcps, the filter could be centered at 30 kcps. This would mean that the modulating frequency ω_c could be varied from 90 to 30 kcps, instead of from 160 to 100 kcps as is done at present. This is possible, however, only if the circuit is designed so that the output of the modulator is restricted to energy at the sideband frequencies and none of the energy of the original signal frequencies appears in the modulator output.

Marking is initiated in the CSL sound spectrograph by any one of four "trigger" circuits opening a gate to impress a 100 kcps signal on the marking amplifier. This trigger operation is ideal where marking is either "on" or "off", with no variation required in the darkness of the marking. This condition applies for the amplitude sections, the continuous amplitude displays, and the construction of scale lines. It does not apply in making a spectrogram, however, where a range of grayness in the marking is desired.

The trigger circuit for making spectrograms with the CSL instrument passes on a range of amplitudes from the analyzed signal to the marking amplifier, thus producing a range of darkness in marking the spectrograms. This action should be linear. That is, there should be no compression in that part of the system which is associated with the trigger circuit; the desired gain control is accomplished elsewhere in the circuit. It is not at present known, however, just how well the desired range and linearity of the spectrogram trigger circuit is attained. The question should be investigated, and if necessary the circuit for marking the broad and narrow band spectrograms could be redesigned so that the signal for marking spectrograms is led directly to the marking amplifier, by-passing the trigger and gate circuits. The 100 kcps oscillator would then not be needed for broad and narrow band analyses, since the signal would itself cause marking on the spectrogram when the signal amplitude is sufficiently great. If low-pass filters were used for the final analysis, care would need to be taken that the rectified low frequencies would be transmitted properly to the marking stylus.

The CSL spectrograph may seem complex in the number of controls which the operator must set correctly to obtain any given result. While an attempt has been made in the design to reduce the number of controls as much as possible, the remaining complexity must be taken as the price paid for the multiple features provided by the instrument.

APPENDIX

Interactions of Close Frequencies in Spectrographic Analysis

Single modulation to a band-pass filter

Let ω_a and ω_b represent the radian frequencies of two adjacent components of the signal f(t) which is to be analyzed by the sound spectrograph. If f(t) is a periodic function, ω_a and ω_b could be adjacent harmonics of the fundamental frequency. With a speed-up of k times, in reproducing the recorded signal, the components of the reproduced signal become $k\omega_a$ and $k\omega_b$. The reproduced function due to these two components only may be represented as

$$f_r(t) = A \cos (k\omega_a t + \varphi_a) + B \cos (k\omega_b t + \varphi_b), \quad \text{(A-1)}$$

where A and B are the amplitudes of the two reproduced components and φ_a and φ_b are assumed phase constants of the two.

To make the following analysis simpler, let us suppose that the two components are of equal amplitude. That is, let $A = B$. Then (A-1) can be written

$$f_r(t) = A\,[\cos(k\omega_a t + \varphi_a) + \cos(k\omega_b t + \varphi_b)]. \qquad \text{(A-2)}$$

To modulate the spectrum of the signal to a higher band of frequencies, multiply (A-2) by the function $\cos \omega_c t$, where ω_c is a variable carrier frequency. Let us assume that all values of ω_c are larger than $k\omega_b$, and that the latter is the larger of the two reproduced components,

$$\omega_c > k\omega_b > k\omega_a. \qquad \text{(A-3)}$$

There is no loss of generality in giving the carrier function an amplitude of unity and a phase constant of zero, since the other amplitudes and phase constants can be expressed in terms of those of the carrier.

As has been seen in Chapter 3, the result of the modulation is an upper sideband with frequencies above ω_c, and a lower sideband which can be expressed as follows,

$$f_{\ell s}(t) = \frac{A}{2}\left\{\cos[(\omega_c - k\omega_a)t - \varphi_a] + \cos[(\omega_c - k\omega_b)t - \varphi_b]\right\}. \qquad \text{(A-4)}$$

The sum of two cosines can be transformed to twice the product of two cosines, the arguments of the latter being half the difference of arguments of the former, and half the sum of the same. (A-4) becomes

$$f_{\ell s}(t) = A \cos[\frac{k}{2}(\omega_b-\omega_a)t+\frac{\varphi_b-\varphi_a}{2}]\cos[(\omega_c-k\frac{\omega_a+\omega_b}{2})t - \frac{\varphi_a+\varphi_b}{2}]. \qquad \text{(A-5)}$$

The second of the cosine factors of (A-5) has a frequency which is the average of the frequencies of the components of (A-4). By (A-3), ω_c is always larger than $\frac{k}{2}(\omega_a+\omega_b)$, so that the frequency of this factor of (A-5) cannot become zero. Also, if we assume that the frequencies of the components of (A-4) both lie inside the passband of the analyzing filter, their average also is in this passband.

The first cosine factor of (A-5) can be regarded as a varying amplitude factor, modulating the second function. This is, in fact, the origin of the phenomenon of <u>beats</u>. The frequency of this first factor is lower than that of the second. Actually the amplitude of the whole function (A-5) rises to maximum at twice the frequency of the first cosine factor, that is with a frequency of $k\omega_b-k\omega_a$. This <u>beat</u> frequency is at just the fundamental rate of the reproduced function, if ω_a and ω_b are adjacent harmonics of a periodic input wave.

When a wide band-pass spectrogram of "voiced" (i.e. periodic) speech is made, the beats described above produce striations in the record, which extend in straight lines across the frequency scale and are spaced in time at the fundamental period. The extension across the frequency scale arises from the fact that whatever frequency band is being examined at the moment, it contains frequency components which are spaced at the fundamental rate.

Second modulation to a low-pass filter

Let the lower sideband of the first modulation be given a second modulation which will place the component frequencies of the lower sideband of this second modulation inside the passband of a low-pass filter. That is, multiply (A-4) by the function L cos $(\omega_\ell t + \varphi_1)$, where ω_ℓ is the lowest frequency reached by ω_c in the course of its variation. The lower sideband of this second modulation will be given by

$$f_{2\ell s}(t) = \frac{AL}{4}\Big\{\cos[(\omega_c - k\omega_a - \omega_\ell)t - (\varphi_1 + \varphi_a)] + \cos[(\omega_c - k\omega_b - \omega_\ell)t - (\varphi_1 + \varphi_b)]\Big\}. \qquad \text{(A-6)}$$

Transforming (A-6) to a product of cosines,

$$f_{2\ell s}(t) = \frac{AL}{2}\cos[\frac{k}{2}(\omega_b-\omega_a)t+\frac{\varphi_b-\varphi_a}{2}]$$

$$\times\cos[(\omega_c-\omega_\ell-k\frac{\omega_a+\omega_b}{2})t-(\varphi_1+\frac{\varphi_a+\varphi_b}{2})]. \quad \text{(A-7)}$$

The first cosine factor of (A-7) is the beat factor, exactly the same as in (A-5). In the second cosine factor, however, there is a big difference from the corresponding factor of (A-5), in that ω_c may take a value which will make the frequency of this factor zero. This will happen even though (A-3) is observed, and is due to the presence of ω_ℓ in the frequency of the factor. Thus the frequency is zero if

$$\omega_c = \omega_\ell + \frac{k}{2}(\omega_a + \omega_b). \quad \text{(A-8)}$$

This condition for zero frequency of the average of the two components in (A-6) is exactly the same as the condition (5.32) for the coincidence of frequencies in the second modulation due to different frequencies in the input signal.

Now if we look at that moment of time when ω_c has the value given by (A-8), we find some singularities in the output of the low-pass filter. Assuming (A-8) to be true, the last cosine factor of (A-7) has a constant value, and can be made an amplitude factor modifying the beat function. Then (A-7) may be written

$$f_{2\ell s}(t) = \frac{AL}{2}\cos(\varphi_1+\frac{\varphi_a+\varphi_b}{2})\cos[\frac{k}{2}(\omega_b-\omega_a)t+\frac{\varphi_b-\varphi_a}{2}]. \qquad \text{(A-9)}$$

This amplitude factor may have any value between -1 and +1, depending on phase relations of the second carrier and the two reproduced components. It is zero and the filter output disappears if

$$\varphi_1 + \frac{\varphi_a+\varphi_b}{2} = n\frac{\pi}{2}, \qquad \text{(A-10)}$$

where n is any odd integer.

It is plain that if (A-10) is satisfied for a particular second carrier with the phase φ_1, then another second modulation could be applied with a phase φ_2 for which (A-10) would not be satisfied. It would only be necessary to make

$$0 < |\varphi_2-\varphi_1| < \pi . \qquad \text{(A-11)}$$

A value of $\pi/2$ for this difference would suffice. This is the remedy proposed in Chapter 5, p. 110, for the possible cancellation of two different components in the low-pass filter.

It is pertinent to inquire whether ω_c remains near the value in (A-8) long enough to give equation (A-9) significance. To answer this question let us start with the zero amplitude condition, i.e. let (A-10) hold, with $n = 1$, and let (A-8) be true at $t = 0$. Then let ω_c be changed at a constant rate R, which is expressed in cps per second. Then at time t,

$$\omega_c = \omega_\ell + \frac{k}{2}(\omega_a + \omega_b) + Rt. \qquad \text{(A-12)}$$

Substituting (A-10) and (A-12) into (A-7),

$$f_{2\ell s}(t) = \frac{AL}{2}\cos(\frac{\pi}{2} - Rt^2)\cos[\frac{k}{2}(\omega_b - \omega_a)t + \frac{\phi_b - \phi_a}{2}]. \qquad \text{(A-13)}$$

The first cosine factor in (A-13) takes the place of the similar factor in (A-9). It has the value zero at $t = 0$, and let us calculate its value at $t = T$, the fundamental period of the reproduced signal. The value of T can be found from

$$k(\omega_b - \omega_a)T = 2\pi ,$$

or

$$T = \frac{2\pi}{k(\omega_b - \omega_a)} = \frac{1}{k(f_b - f_a)} \qquad \text{(A-14)}$$

In the spectrograph described in Chapter 6, $k = 6$, and let us give the fundamental frequency a low value, 100 cps. Then $T = 1/600$. For the rate of change of ω_c, take the rate most often used in this same spectrograph, which is 10000 cps in 5 minutes, or 33 1/3 cps per second. This is in terms of input frequencies, so that for the rate of ω_c it must be multiplied by 2π and by k. Then $R = 400\pi$. Substituting these values of R and T into (A-13), the amplitude factor becomes

$$\cos\left(\frac{\pi}{2} - RT^2\right) = \sin(RT^2) = \sin\frac{400\,\pi}{600^2} = \sin 0.00111\pi$$

This angle is approximately 12 minutes of arc, and its sine is .00349. This is the value to which the amplitude factor of (A-13) has risen from zero, during one fundamental period of the reproduced signal. Since the maximum value of the factor is one, the above rise is very small. If we look at longer times, the rise of the factor at first will be approximately proportional to the square of the number of fundamental periods included. In ten periods the factor would be $\sin 20^{\circ} = .342$.

These considerations should be sufficient to show that when the condition represented by (A-9) occurs, it is relatively persistent.

References

1. H. W. Batten, R. A. Jorgensen, A. B. MacNee, and W. W. Peterson, "The Response of a Panoramic Receiver to CW and Pulse Signals", Proceedings of the IRE, Volume 42, Number 6, pp. 948-956 (June 1954).

2. L. Brillouin, Science and Information Theory, 2nd Edition, Academic Press, Inc., New York (1962).

3. R. V. Churchill, Fourier Series and Boundary Value Problems, McGraw-Hill Book Co., Inc., New York (1958).

4. C. L. Cuccia, Harmonics, Sidebands and Transients in Communication Engineering, McGraw-Hill Book Co., Inc., New York (1952).

5. C. G. M. Fant, "New Instruments for Speech Analysis and Synthesis", Proceedings of the VIII International Congress of Linguists, Oslo University Press, pp. 320-326 (1958).

6. J. S. Gill, "A Versatile Method for Short-Term Spectrum Analysis in 'Real-Time'", Nature, Volume 189, Number 4759, pp. 117-119 (January 14, 1961).

7. J. C. Hancock, An Introduction to the Principles of Communication Theory, McGraw-Hill Book Co., Inc., New York (1961).

8. C. M. Harris and W. M. Waite, "Gaussian-Filter Spectrum Analyzer", Journal of the Acoustical Society of America, Volume 35, Number 4, pp. 447-450 (April 1963).

9. W. H. Huggins, "A Phase Principle for Complex-Frequency Analysis and Its Implications in Auditory Theory", Journal of the Acoustical Society of America, Volume 24, Number 6, pp. 582-589 (November 1952).

10. W. Kaplan, Operational Methods for Linear Systems, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts (1962).

11. L. G. Kersta, "Amplitude Cross-Section Representation with the Sound Spectrograph", Journal of the Acoustical Society of America, Volume 20, Number 6, pp. 796-801 (November 1948).

12. L. G. Kersta, "Voiceprint Identification", Journal of the Acoustical Society of America, Volume 34, Number 5, p. 725 (May 1962) (abstract).

13. W. Koenig, H. K. Dunn, and L. Y. Lacy, "The Sound Spectrograph", Journal of the Acoustical Society of America, Volume 18, Number 1, pp. 19-49 (July 1946).

14. W. Koenig and A. E. Ruppel, "Quantitative Amplitude Representation in Sound Spectrograms", Journal of the Acoustical Society of America, Volume 20, Number 6, pp. 787-795 (November 1948).

15. Y. W. Lee, Statistical Theory of Communication, John Wiley and Sons, Inc., New York (1960).

16. S. J. Mason and H. J. Zimmerman, Electronic Circuits, Signals, and Systems, John Wiley and Sons, Inc., New York (1960).

17. R. C. Mathes, A. C. Norwine and K. H. Davis, "The Cathode-Ray Sound Spectroscope", Journal of the Acoustical Society of America, Volume 21, Number 5, pp. 527-537 (September 1949).

18. G. E. Peterson, "Parameter Relationships in the Portrayal of Signals with Sound Spectrograph Techniques", Journal of Speech and Hearing Disorders, Volume 17, Number 4, pp. 427-432 (December 1952).

19. G. E. Peterson, "Design of Visible Speech Devices", Journal of the Acoustical Society of America, Volume 26, Number 3, pp. 406-413 (May 1954).

20. G. E. Peterson and G. Raisbeck, "The Measurement of Noise with the Sound Spectrograph", Journal of the Acoustical Society of America, Volume 25, Number 6, pp. 1157-1162 (November 1953).

21. R. K. Potter, G. A. Kopp and H. C. Green, Visible Speech, D. Van Nostrand Co., Princeton, New Jersey (1947).

22. A. J. Prestigiacomo, "Plastic-Tape Sound Spectrograph", Journal of Speech and Hearing Disorders, Volume 22, Number 3, pp. 321-327 (September 1957).

23. R. R. Riesz and L. Schott, "Visible Speech Cathode-Ray Translator", Journal of the Acoustical Society of America, Volume 18, Number 1, pp. 50-61 (July 1946).

24. M. Schwartz, Information Transmission, Modulation, and Noise, McGraw-Hill Book Co., New York (1959).

25. J. C. Steinberg, "Application of Sound Measuring Instruments to the Study of Phonetic Problems", Journal of the Acoustical Society of America, Volume 6, Number 1, pp. 16-24 (July 1934).

26. W. E. Stewart, Magnetic Recording Techniques, McGraw-Hill Book Co., Inc., New York (1958).

27. F. E. Terman, Electronic and Radio Engineering, 4th Edition, McGraw-Hill Book Co., Inc., New York (1955).

28. M. R. Weiss and C. M. Harris, "Computer Technique for High-Speed Extraction of Speech Parameters", Journal of the Acoustical Society of America, Volume 35, Number 2, pp. 207-214 (February 1963).

29. E. M. Williams, "Radio-Frequency Spectrum Analyzers", Proceedings of the IRE, Volume 34, Number 1, pp. 18-22 (January 1946).

DISTRIBUTION LIST

Professor David Abercrombie, Head
Department of Phonetics
University of Edinburgh
Minto House
Chamber Street
Edinburgh, Scotland

Air Force Cambridge Research Labs.
Laurence G. Hanscom Field
Bedford, Massachusetts
Attention: Research Library,
CRMXL-R

Dr. Daniel Alpert, Director
Coordinated Science Laboratory
University of Illinois
Urbana, Illinois

Dr. Victor C. Anderson
University of California
Marine Physical Laboratory
San Diego 52, California

Dr. John W. Black
Director of Experimental Phonetics
Ohio State University
Columbus 10, Ohio

Miss Helen Brownsen
National Science Foundation
Office of Science
Information Services
Washington 25, D.C.

Mr. William D. Chapman
International Business Machines
Corporation
Raleigh, North Carolina

Chief of Naval Research (2)
Department of the Navy
Washington 25, D.C.
Attention: Code 437, Information
Systems Branch

Commander AFCRL
Detachment 2, AFRD
L. G. Hanscom Field
Bedford, Massachusetts
Attention: Technical Library

Commanding Officer
Harry Diamond Laboratories
Washington, D.C.
Attention: Library

Commanding Officer (10)
Office of Naval Research
Navy No. 100, Fleet Post
Office Box 39
New York, New York

Commanding Officer
ONR Branch Office
1000 Geary Street
San Francisco 9, California

Commanding Officer
ONR Branch Office
1030 E. Green Street
Pasadena, California

Commanding Officer
ONR Branch Office
207 West 24th Street
New York, New York

Commanding Officer and Director
U. S. Naval Training Device Center
Post Washington
Long Island, New York
Attention: Technical Library

Commander
Wright Air Development Division
Wright-Patterson Air Force Base,
Ohio
Attention: WCOSI

Dr. Franklin S. Cooper, President
Haskins Laboratories
305 E. 43rd Street
New York 17, New York

Professor James F. Curtis, Head
Department of Speech Pathology
and Audiology
State University of Iowa
Iowa City, Iowa

Dr. Edward E. David, Jr.
Director of Visual and Acoustics
Research
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

David Taylor Model Basin
Code 042 Technical Library
Washington, D.C.

Defense Documentation Center (20)
Cameron Station
Alexandria, Virginia

Dr. Pierre Delattre, Director
Experimental Phonetics Laboratory
The University of California
Santa Barbara, California

Department of the Army
Office of the Chief of Research
and Development
Pentagon, Room 3D442
Washington 25, D.C.
Attention: Mr. L. H. Geiger

Dr. H. H. Campaign
National Security Agency
Fort George G. Meade, Maryland

Director (6)
Naval Research Laboratory
Technical Information Officer
/Code 2000/
Washington 25, D.C.

Dr. Ladislav O. Dolansky
Department of Electrical
Engineering
Northeastern University
Boston, Massachusetts

Dr. James P. Egan
Hearing and Communication
Laboratory
Indiana University
Bloomington, Indiana

Dr. Med. K. Faaborg-Andersen
Centralsygehuset
Øreafdelingen
Nykøbing F., Denmark

Dr. Gunnar Fant
Speech Transmission Laboratory
Royal Institute of Technology
Stockholm 70, Sweden

Dr. James L. Flanagan, Head
Speech and Auditory Research
Department
The Bell Telephone Laboratories
Murray Hill, New Jersey

Dr. James W. Forgie
Lincoln Laboratory
Massachusetts Institute of
Technology
Lexington, Massachusetts 02173

Dr. George C. Francis
Computing Laboratory, BRL
Aberdeen Proving Ground, Maryland

Dr. Dennis B. Fry
Department of Phonetics
University College
Gower Street
London WC1, England

Dr. Paul L. Garvin
Ramo-Wooldridge
Division of Thompson Ramo-
Wooldridge Inc.
8433 Falbrook Avenue
Canoga Park, California

Dr. Louis J. Gerstman
College of Engineering
New York University
New York, New York

Dr. Vincent Giuliano
Arthur D. Little, Inc.
Acorn Park
Cambridge, Massachusetts

Dr. Bernard Gold
Lincoln Laboratories
Massachusetts Institute of
Technology
Lexington, Massachusetts

Dr. Martin Grützmacher
Physikalisch-Techniche
Bundesanstalt
Braunschweig, West Germany

Professor Cyril M. Harris
Department of Electrical
Engineering
Columbia University
632 West 125th Street
New York, New York

Dr. Joseph E. Hind
Associate Professor of
Neurophysiology
Laboratory of Neurophysiology
283 Medical Sciences Building
The University of Wisconsin
Madison, Wisconsin

Dr. Ira J. Hirsh
Central Institute for the Deaf
818 S. Kingshighway
St. Louis, Missouri

Dr. Harry Hollien
Communication Sciences Laboratory
Department of Speech
University of Florida
Gainesville, Florida

Mr. Robert A. Houde, Head
Speech Communication Section
Communication Sciences Laboratory
Research Division
General Dynamics/Electronics
1400 North Goodman Street
Rochester, New York

Dr. Arthur S. House
Department of Audiology and
Speech Science
Purdue University
Lafayette, Indiana

Dr. Frances Ingemann
Department of Linguistics
University of Kansas
Lawrence, Kansas

Dr. George A. Kopp
Director of Speech Clinic
Wayne State University
Detroit, Michigan

Dr. Karl D. Kryter, Head
Speech Research Group
Stanford Research Institute
Menlo Park, California

Dr. Peter Ladefoged, Director
Phonetics Laboratory
Department of English
The University of California
at Los Angeles
Los Angeles, California

Mr. Walter Lawrence
Department of Phonetics
University of Edinburgh
Minto House
Chambers Street
Edinburgh, Scotland

Dr. Ilse Lehiste, Chairman
Department of Linguistics
Ohio State University
Columbus, Ohio

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington 73, Massachusetts
Attention: Library

Dr. Bertil Malmberg
Institute of Phonetics
Lunds Universitet
Lund, Sweden

Miss Margaret Masterman
Cambridge Language Research Unit
20 Millington Road
Cambridge, England

Dr. Roger L. Merrill, Manager
Engineering Physics Department
Battelle Memorial Institute
505 King Avenue
Columbus, Ohio

Professor George A. Miller
Department of Psychology
Harvard University
Cambridge, Massachusetts

Professor Paul Moore, Chairman
Department of Speech
University of Florida
Gainesville, Florida

Dr. A. K. Smilow
National Bureau of Standards
Data Processing Systems Division
Room 239, Building 10
Washington 25, D.C.

Dr. A. M. Uttley, Superintendent
National Physical Laboratory
Autonomics Division
Teddington, Middlesex, England

National Security Agency
Fort George G. Meade, Maryland
Attention: Librarian, C-332

Naval Electronics Laboratory
San Diego 52, California
Attention: Technical Library

Naval Ordnance Laboratory
White Oaks
Silver Spring, Maryland
Attention: Technical Library

Professor Dr. Yoshiyuki Ochiai
Professor of Telecommunication
Laboratory of Audiology
Faculty of Engineering
Nagoya University
Furo-cho, Chikusa-ku
Nagoya, Japan

Mr. Douglas L. Hogan
National Security Agency
Fort George G. Meade, Maryland

Office of Naval Research
Branch Office Chicao
230 N. Michigan Avenue
Chicago, Illinois

Office of Naval Research
Branch Office Chicago
495 Summer Street
Boston, Massachusetts

Dr. Toshiyuki Sakai
Professor of Electrical Engineering
Kyoto University
Sakyo-Ku, Kyoto
Japan

Dr. Manfred R. Schroeder
Bell Telephone Laboratories
Murray Hill, New Jersey

Dr. Earl D. Schubert
Professor of Audiology
Division of Speech Pathology and Audiology
Stanford University School of Medicine
300 Pasteur Drive
Palo Alto, California

Professor Claude Shannon
Massachusetts Institute of Technology
Cambridge, Massachusetts

Dr. Eva Sivertsen
Møllebakken 36
Trondheim, Norway

Dr. Svend Smith
Svejagervej 52, Hellerup
Copenhagen, Denmark

Professor Dr. Antti Sovijarvi
Mantytie 17B
Helsinki, Finland

Dr. K. N. Stevens
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

Dr. S. S. Stevens, Director
Psycho-Acoustic Laboratory
Harvard University
Cambridge, Massachusetts

Mr. Peter Strevens
Language Centre
University of Essex
Wivenhoe Park
Colchester, Essex, England

Dr. D. L. Subrahmanyam, Assistant Director
Central Electronics Engineering Research Institute
Pilani (Rajasthan)
India

Mr. John Swaffield
British Post Office
Joint Speech Research Unit
Eastcote Road
Ruislip, Middlesex, England

Technical Library
Director Defense Research Engineering
Room 3C-128, The Pentagon
Washington, D.C.

Dr. G. H. Gleissner
Assitant Director for Computation
/Code K-4/
U. S. Naval Weapons Laboratory
Dahlgren, Virginia

Dr. Jw. van den Berg
Physiologisch Laboratorium
Der Rijks Universiteit
Te Groningen
Bloemsingel 1, Netherlands

Dr. Henning E. von Gierke
Aero-Space Medical Research Labs.
Wright-Patterson Air Force Base, Ohio

Dr. William S-Y. Wang
University of California
Department of Linguistics
Berkeley, California

Dr. John C. Webster
U. S. Navy Electronics Laboratory
San Diego 52, California
Code 2124

Dr. Ronald W. Wendahl
Department of Speech
The University of Minnesota
Minneapolis, Minnesota

Mr. Richard H. Wilcox, Head
Information Systems Branch
Office of Naval Research
Washington 25, D.C.
Code 437

Dr. Harold M. Wooster, Head
Information Sciences Division
Air Force Office of Scientific Research
Office of Aerospace Research
United State Air Force
Washington 25, D.C.
Attention: SRMI

Dr. E. Zwicker
Elektrotechnisches Institut der
Technischen Hochschule Stuttgart
Stuttgart W
Breitscheidstr 3, Germany

Professor Dr. E. Zwirner
Direktor
Institut fur Phonometrie
Munster/Westfalen
Stinfurter Strasse 107
Germany

None

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Communication Sciences Laboratory 180 Frieze Building The University of Michigan	2a. REPORT SECURITY CLASSIFICATION None 2b. GROUP

3. REPORT TITLE

Some Theoretical and Experimental Aspects of Analysis with the Sound Spectrograph

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (Last name, first name, initial)

Hugh K. Dunn
Tiong Suy Yu
William D. Chapman

6. REPORT DATE August 1966	7a. TOTAL NO. OF PAGES 206	7b. NO. OF REFS 29
8a. CONTRACT OR GRANT NO. Nonr 1224(22) NR 049-122 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Communication Sciences Laboratory Report Number 7 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	

10. AVAILABILITY/LIMITATION NOTICES

A limited number of copies are available from the Director of the Communication Sciences Laboratory

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Information Systems Branch Office of Naval Research Washington, D.C.

13. ABSTRACT

Before the introduction of the sound spectrograph in 1946, analysis of complex audio phenomena, such as speech, into time, frequency, and amplitude coordinates depended for the most part on a very time-consuming Fourier analysis of an oscillographic record of the event. The techniques introduced with the sound spectrograph permitted the speeding-up of the analysis procedure by a large factor. These techniques and their underlying principles are discussed in the first few chapters of this report. The use of a bank of filters to accomplish real-time speech analysis is not discussed, but that form of the sound spectrograph which achieves frequency analysis by sweeping the spectrum of the recorded signal past a single analyzing filter is discussed in detail. A sound spectrograph developed at the Communication Sciences Laboratory of the University of Michigan is then described, and samples of work with the machine are shown. This instrument includes some novel features, such as the automatic marking on the spectrogram of time, frequency, and amplitude scales.

None

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
analysis of speech filter analysis of speech Fourier analysis heterodyne filter spectrograms spectrographic analysis speech analysis sound spectrograms sound spectrograph						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

None

Security Classification