A REAL - TIME AUDIO SPECTRAL ANALYSER USING ACTIVE FILTERS WITH ADJUSTABLE PARAMETERS

A Thesis submitted by Ronald Stanley THOMAS for the degree of Doctor of Philosophy in the Faculty of Engineering of the University of London.

The Electrical Engineering Department,
Imperial College,
London, S.W.7.

August, 1964.

ABSTRACT

During the past twenty-five years spectral analysis has played an important part in the study of speech signals. It has aided research workers in determining many of the speech parameters most important in perception, in building speech bandwidth compression systems and, more recently, in constructing instruments which will recognize automatically some speech sounds. During this time a large number of electronic spectral analysers have been built, many of which operate in real-time. Most of the real-time analysers comprise a bank of band-pass filters with fixed characteristics chosen to satisfy the individual requirements of research workers. In this thesis a real-time analyser of the filter bank type in which the filtering characteristics can be arbitrarily chosen is described and demonstrated.

Quenched lossless tuned circuits are used as the band-pass filters. Any desired filtering characteristic is achieved by weighting the input to the analyser with periodic functions which determine both the effective bandwidths and frequency responses of the filters. The outputs from the filters are observed at fixed intervals of time determined by the weighting functions.

Speech spectrograms are shown in which three different filter frequency characteristics are used. Their

differences are noted and they are also compared with results from the Kay Electric Sonagraph, a commercial analyser which does not operate in real-time. The results compare favourably. Since filter bandwidths can be changed easily, a simple system is shown whereby the filter bandwidths in the analyser can be switched automatically between two alternate values depending upon characteristics of the signal being analysed. In addition, further application of such an analyser is suggested for research into a speech bandwidth compression scheme.

ACKNOWLEDGEMENTS

The author wishes to thank his supervisor, Professor E.C. Cherry, for his encouragement and guidance during the course of the work. Colleagues at the Imperial College have also been helpful in several ways, particularly Mr. L.S. Moye for many fruitful discussions and Mr. M. Lecours for helping to obtain the photographs in Figures 5.4 and 5.5. Thanks are also due to Dr. H.B. Voelcker, who in the beginning suggested building a quenched lossless tuned circuit with the possibility of using it in a spectral analyser.

Help obtained from the Electronic Workshops, especially Mr. A. White for winding the inductors in the filters, is appreciated.

The Kay Electric Sonagraph used for comparison with the results from the Imperial College analyser was that in the Phonetics Dept. at University College. The writer wishes to thank Dr. A.J. Fourcin for his help in this respect. Thanks also goes to Mr. J.C. Wells, a phonetician, for voicing the speech sounds in Figures 4.7 and 4.8.

The author wishes to thank the Beaverbrook Foundation for a scholarship from 1960 to 1962, and the Defence Research Board of Canada for financial assistance from Sept., 1962 to the present.

Mr. J.J. Sparkes is thanked for reading the proofs.

CONTENTS

INTRODUCTION AND BACKGROUND

1.1. Introduction.

The variations with time of the short-term energy spectrum of signals such as speech can be measured with a bank of electrical filters. In normal banks of filters the bandwidths and frequency characteristics are very difficult to change, since physical changes in the circuits are necessary. In this thesis a new type of spectral analyser constructed at Imperial College is described, which has banks of filters where the bandwidths and frequency characteristics are easily changed by purely electrical means. Since there is an arbitrary choice of filtering characteristics, such an instrument has considerable advantages over other analysers of the filter bank type for research problems.

The versatility of the Imperial College analyser is demonstrated with speech and other signals where the results are compared with spectrograms (time-frequency-intensity displays) from the Kay Electric Sonagraph, the main instrument commercially available for speech analysis. Suggestions are also made for further application of the IC analyser in variable bandwidth filtering and bandwidth compression studies.

During the past twenty-five years a large number of spectral analysers have been developed and applied in many research and measurement problems. As the importance of short-term spectral analysis has increased, these analysers have also increased in versatility and complexity. Although the spectral analysis of speech signals is of most concern here, many other signals have been profitably analysed in this way as well. Before considering details of the IC analyser, it is in order to make some comments on spectral analysis, some of its uses, and the types of analysers developed previously.

1.2. Spectral Analysis.

Any communication signal may be represented either as a function of time or as a function of frequency. As a function of time a signal may be expressed as a varying voltage, pressure, position, etc., and its value will be known precisely at any instant in time. This representation is the one with which most engineers are more familiar, since almost all communication signals are originally produced as time functions. As a function of frequency a signal has a more limited use.

The transformation between the two representations can be accomplished by the Fourier integral transform pair

$$\left.\begin{aligned} S(f) &= \int_{-\infty}^{\infty} s(t) \quad e^{-j2\pi f t} dt \\ s(t) &= \int_{-\infty}^{\infty} S(f) \quad e^{j2\pi f t}\, df \end{aligned}\right\} \qquad \ldots(1.1)$$

where s(t) is the time function and S(f) is its frequency spectrum, a function only of frequency. These two equations form the basis of spectral analysis.

The representation of a signal either as a time function or as a frequency function is an idealization, and intuitively we regard most audio signals as functions of both time and frequency. For example, a note from a piano has a definite tone or frequency, but we are also very much aware of the time at which it occurs. With this in mind, Gabor (1946) published a paper setting forth a mathematical method for expressing any signal as a summation of functions of both time and frequency which could be denoted by discrete coefficients on the time-frequency plane. (This is discussed more fully in Chapter 2). Gabor pointed out that the original time function could be recovered from the time-frequency representation.

The process of reaching a time-frequency representation of a time function is normally referred to as short-term spectral analysis, and it can be performed by mechanical, graphical or purely mathematical means, although electronic devices are most often used. Because of the difficulty of instrumentation and the lack of use for some of the spectral information, practical spectral analysers never present enough information on the time-frequency plane for exact resynthesis of the time function. At the very least phase information is not available. However, this incomplete presentation of the spectral structure is still very useful in showing up interesting characteristics of the processes involved in the production of the signal, and often it provides enough information for approximate resynthesis of the original time function.

One of the most fruitful applications of short-term spectral analysis has been in the study of speech signals. The first electronic spectral analyser was used by Dudley (1939) in his vocoder. By first representing a speech signal in spectral form, Dudley was able to transmit it in that form with a reduced bandwidth before approximately resynthesizing the temporal waveform. Since that time many other analysers have been built and used in speech analysis. From the knowledge gained in this way, speech has been transmitted over communication channels with low bandwidth capacities, and machines have been built which will automatically recognize speech sounds. It has been shown that time-frequency-intensity patterns obtained by spectral analysis of speech signals can be recognised by trained persons (Potter et al, 1947) and these patterns have even found some application in the instruction of the deaf. However, the cost and complexity of most suitable analysers limit the use somewhat in this instance.

Spectral analysis finds application in many other fields where periodic or quasi-periodic phenomena are of interest. Some of these fields are aerodynamics, seismology, oceanography, economics, geomagnetism, and meteorology to mention a few. In all these cases spectral representations of the time functions emphasize certain characteristics not before obvious and often allow predictions to be made about future events.

1.3. Spectral Analysers.

Short-term spectral analysis is most commonly accomplished by means of electr~~onic~~ical band-pass filters. For a time varying source applied to such a filter, the output is a function of both time and frequency as required by the concept of a short-term (or time variable) spectrum. Thus, by means of filters with suitable temporal and frequency characteristics it is possible to obtain a desired time-frequency representation of any signal.

Many spectral analysers, varying greatly in versatility and complexity, have been developed. Some of them provide only a rudimentary analysis in which a small percentage of the spectrum is sampled. Others present up to half the theoretically available spectral information on the time-frequency plane. (Phase data, which are not presented in practice constitute half the available spectral information from the communication point of view).

Many analysers comprise one or more (usually one) banks of contiguously tuned band-pass filters (Dudley, 1939; Admiraal, 1960; Harris, 1963; Liljencrants, et al, 1962; Campanella, 1962). These devices carry out their analysis in real-time, i.e. at the same rate the original signal is produced. They can be designed to give relatively complete coverage of the spectrum of interest. (This depends upon the number, bandwidths and frequency responses of the band-pass filters). As a signal is fed in parallel to a filter bank, it is possible, at any moment in time, to observe a spectral section (intensity versus frequency) by interpolating between voltage amplitudes in the contiguous filters. Since the outputs from the filters are continuously varying, the spectral section is itself a function of time. It is possible to provide means for observing or recording all the filter outputs continuously or else to observe spectral sections at fixed time intervals.

Since this type of analyser has a number of filters, it often suffers from the disadvantage of being rather bulky. It also lacks versatility owing to the fact that filter characteristics are fixed in the sense that they cannot be altered without physical changes in the circuits. However, these analysers do provide the simplest method of obtaining real-time spectral analysis, and they have proved very popular in speech analysis. The design and construction of most such analysers have been carried out in research laboratories according to their specific requirements.

Other analysers normally have a single filter which can have its tuned frequency moved across the spectrum of interest either manually (as in wave analysers) or automatically. One filter can sample only a small portion of a spectrum at any time, so that in devices such as wave analysers or analysers with an automatically swept filter frequency, the spectral coverage is very small. They are suitable only for the analysis of signals which do not change their spectral structure rapidly with time. If more complete coverage of the time-frequency plane is required, the single filter must sample the spectrum repeatedly. This can be achieved by storing the temporal waveform and presenting it to the filter again and again while the filter frequency is swept across the spectrum. It is possible in some cases to speed up the presentation of the stored signal to many times its original speed and thus allow the single filter to perform the function of a bank of filters in real-time.

The main advantage of a single filter analyser lies in the fact that it is easier to change a single filter than a bank of filters. Some reduction in physical size is also possible, but if real-time analysis or good spectral coverage is desired, additional complications are introduced in the storage of the temporal waveform.

1.4. Kay Electric Sonagraph.

The most important and almost only instrument commercially available at present for spectrographic studies of speech, the Kay Electric Sonagraph, is a single filter analyser. It is based on an instrument developed at the Bell Telephone Laboratories and described in a series of articles by Potter et al, 1946 and Prestigiacomo, 1957. Owing to the importance of the Sonagraph in speech analysis and the fact that it was the only analyser available for direct comparison with the IC analyser, a description of its operation will be given.

The instrumentation of the Sonagraph comprises a recording and playback unit giving a closed-loop magnetic memory storage with a 2.4 second capacity. A frequency range of 8,000 cycles per second can be covered by means of a heterodyne wave analyser which scans the signal spectrum slowly and linearly as the magnetic loop is continuously replayed. Two alternative filter bandwidths of 45 c/s and 300 c/s are available by manual switching. The time required for a complete time-frequency-intensity analysis of a 2.4 second speech sample is five minutes. The "Sonagram" is printed on a sheet of teledeltos paper with the spectral intensity represented by the density of a grey-black marking. To ensure synchronization of the signal for each replay the paper is placed on a drum mechanically attached to, and centred on the axis of rotation of, the magnetic loop. Frequency is shown on the vertical axis and time along the horizontal axis. Included is a spectral sectioning device which can display spectral sections, (intensity versus frequency) at any preselected instant in time by the setting of a microswitch to be activated on each replay. The time-frequency-intensity displays obtained from the Sonagraph are mainly used in qualitative studies of speech signals. Examples of Sonagrams are shown in Chapter 4 where comparison is made between them and the results from the IC analyser.

1.5. Analysers using Delay Lines.

As mentioned above, it is possible to achieve real-time analysis with a single filtering device if the stored time function can be speeded up sufficiently. There are two spectral analysers which provide real-time analysis in this way by using delay lines to store a time function for a short interval of time. In one (Gill, 1961) a single filter is swept across the spectrum of interest. In the other (Brookner, 1961) an unconventional method of filtering, based on a phase interference principle, is used. Since both devices are a departure from conventional analysers it is worth noting some of their characteristics.

In Gill's analyser the input waveform is first sampled at a fixed rate, and the samples are then converted to digital form before being put on a closed loop memory consisting of a delay line. The samples circulate round the loop at high speed and each time round a new sample is added as the "oldest" sample on the loop is dropped out. The stored samples (300 in number) then pass any point on the delay line at a much higher rate than that at which they were originally produced. At one point on the loop, the samples are reconverted to analogue form and then processed by a single band-pass filter which is swept across the frequency range of interest. In this way, the original samples are processed many times by the filter, which effectively performs the function of a bank of filters. This analyser finds its main use in the presentation of time-frequency-intensity displays of speech signals.

The analyser described by Brookner can, within limits, simulate the transfer characteristic of a bank of arbitrary band-pass filters. It is based on what is called the Coherent Memory Filter (CMF) which is illustrated in Fig. 1.1.

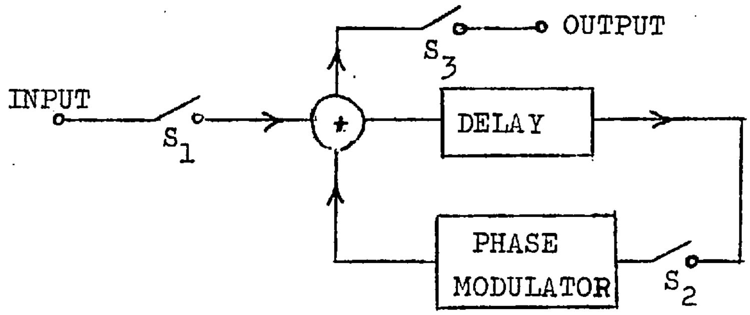


Fig. 1.1. Coherent Memory Filter.

The feedback loop in the CMF has unity gain and the addition of the delay line provides a closed loop memory for the system. Phase modulation in the feedback loop provides a variable time delay, which allows scanning of the frequency spectrum when a suitable modulation function is used. By a process of phase interference and adding, the necessary integration for filtering is performed. The output from the CMF consists of spectral sections at fixed intervals of time. After each spectral section is read out the contents of the delay line memory are wiped out by opening S_1 and S_2, and the filtering process is begun again. The mathematical description of this device is rather complicated and can be found in Brookner's paper. Brookner describes how arbitrary filtering characteristics can be achieved by complex modulation of the signal entering the CMF, and also by additional filtering of its output.

1.6. Digital Computers in Spectral Analysis.

Recently, with the coming of advanced programming techniques and high speed digital computers, it has become feasible to use digital computers for spectral analysis. Since then, digital computers have been used for studies in bandwidth compression, speech recognition, and adaptive filtering (Glaser, 1961; Golden, 1963; Manley, 1963; Sholtz and Bakis, 1962).

Before the analysis is performed the temporal waveform is first sampled at a suitable rate and the discrete samples are converted to digital form. The digital data are then fed into the computer, which carries out the mathematical operations performed in filtering. This is a convenient method for spectral analysis since arbitrary filtering characteristics can be chosen and phase information is easily obtained as well. Although analysis is not performed in real-time, digital computers are useful for the evaluation of hypothetical systems before the construction of complicated instruments is undertaken.

1.7. <u>Analysis by Means of the IC Analyser</u>.

The analyser constructed at the Imperial College comprises two banks 32 of active band-pass filters in which the transfer characteristics can be arbitrarily chosen. Before the speech signal is fed in parallel to the filter banks, its amplitude is modulated by a weighting function which determines the effective characteristics (bandwidths and frequency responses but not tuned frequencies) of the band-pass filters. The amplitudes of the voltages from the filters are then observed at fixed time intervals. These intervals are determined by the chosen filter bandwidths. Interpolation between contiguous filter outputs can be undertaken such that the output is presented as a sequence of spectral sections of intensity versus frequency. The analysis takes place in real-time as in all analysers of the filter bank type. The spectral output is suitable for time-frequency-intensity displays and would also be suitable for bandwidth compression studies, where speech synthesis occurs, or for speech recognition studies.

Not all signals are analysed in an equally suitable manner with fixed filter characteristics. The choice of arbitrary filtering characteristics in the IC analyser without physical circuit alterations is then a significant advantage over other analysers using filter banks. The analyser may be "matched" to the signal either manually or, perhaps eventually even automatically if desired.

Two points of similarity exist between the IC analyser and that described by Brookner - the choice of arbitrary filtering characteristics and the necessity of observing the output at fixed time intervals. When the same filtering characteristics are used, however, the two filter banks in the IC analyser make it possible to obtain twice as much spectral information (from the communication point of view) as that from Brookner's analyser.

1.8. The Application of Spectral Analysis in the Study of Speech Signals.

1.8.1. Bandwidth Compression of Speech Signals.

The principal application of spectral analysis of speech signals has been in the development of bandwidth compression systems. Early work was done in speech bandwidth compression by Homer Dudley, 1939, who devised the first "Vococer" at the Bell Telephone Laboratories. This device transmitted intelligible speech through 11 channels of only 25 c/s each, operating with a method of spectral analysis and synthesis. The spectrum of speech was roughly analysed into 10 bands of 250 c/s each, and the intensity of each band transmitted through separate channels of 25 c/s. The transmitted intensities were then used for modulating buzzers at the receiving end, roughly reproducing the original spectrum. The eleventh channel was used for transmitting the pitch (larynx frequency).

The resulting speech transmission was then achieved in a bandwidth of less than 300 c/s as compared to 3,000 c/s for normal telephone quality speech. Although the quality of the synthesized speech was not very good, it served to demonstrate the principle of spectral analysis and resynthesis, and since Dudley's Vocoder many other bandwidth compression systems using analysis - synthesis techniques have been developed (Gabor, 1947; Fant, 1959; Lawrence, 1960; Harmuth, 1960; Bell, et al, 1961; Campanella, 1962; Golden, 1963).

* Strictly speaking Gabor's method is not analysis as such but a form of temporal distortion.

Most of the subsequent vocoders at the Bell Telephone Laboratories have been similar in principle to Dudley's original, having only more refinements. One of the most recent attempts has been the synthesis of a hypothetical vocoder with a digital computer by Golden, who claims good quality in the resulting speech.

Other systems are Partially based on ~~electrical~~ rough analogues of the vocal tract (Stevens, et al, 1953). In these systems several parametric functions derived from the original speech waveform and its spectrum are used to control the ~~analogue~~ synthesizer. For example, the control functions may represent the main formant frequencies, formant amplitudes, hiss, larynx pitch, etc., and may be derived automatically from the output of a real-time spectral analyser, or manually through visual observation of a time-frequency-intensity display. The total bandwidth occupied by the control functions is considerably less than that of the original speech waveform.

Although fairly good quality speech can be obtained from these parametric synthesizing devices with a high bandwidth compression ratio, there are still problems to be solved before they become commercial propositions. As research tools, however, these instruments have been very valuable in indicating the important parameters in speech perception. Perhaps the best known systems of this type are the Parametric Artificial Talking Device (PAT) developed by W. Lawrence and the device known as OVE developed by C.G.M. Fant.

1.8.2. Machine Recognition of Speech.

The time-frequency-intensity displays (spectrograms) of speech which have proved so helpful in the development of bandwidth compression systems were first introduced at the Bell Telephone Laboratories during the war years (Potter, et al, 1947). It was demonstrated that the distinctive spectral patterns could be recognised and read by trained persons even when speakers were changed. It was later shown at the Haskins Laboratories (Borst Liberman and Cooper, 195~~7~~0 (1)) that the gross spectral patterns even when stylized by hand painted approximations contained most of the necessary information for perception. Stylized hand painted spectrograms were converted to temporal waveforms by means of a vocoder playback to give intelligible speech.

Since the gross spectral structure contains the information necessary for the identification of most speech sounds, it is a convenient way of specifying speech. This specification has led to the construction of machines which recognize the spoken word under limited Conditions (Davis, et al, 1952; Sholtz and Bakis, 1962; Olson, et al, 1962; Olson, 1964). The important first step in such machines is spectral analysis in real-time, normally carried out by a bank of filters. The outputs from the spectral analysers are processed by special circuits which compare the resulting spectral patterns to stored reference patterns. In this way speech can be recognised as a sequence of phonemes. Further processing can occur to convert the phonemes to words, type the words on paper, translate to another language or resynthesize audible speech.

The work in speech recognition is still in its early stages and most of the devices which have been constructed use only rudimentary spectral analysers. These devices can recognize only a small number of words, and in some cases they must be adjusted for individual speakers. Even where analysis is more elaborate, differences between speakers and accents cause difficulties. Theoretically, however, the problems can be overcome by improved specification and analysis.

Olson, et al, feel that recent results indicate that speech processing systems will eventually come into common use. Equipment with a capacity of 100 syllables and having translation facilities and a speech synthesizer has been demonstrated and shown to perform well even when speakers were changed. Olson has pointed out that a machine having a large vocabulary will require a very extensive memory to hold reference information. For real-time operation of such a device, high speed computer techniques will have to be applied. In spite of present limitations, the feasibility of speech recognition and processing systems has been demonstrated and research is likely to continue until the main problems are solved.

SIGNAL ANALYSIS AND FILTERING

2.1. Simultaneous Time-Frequency Representation of Temporal Waveforms.

There are many ways of representing a temporal signal on the time-frequency plane depending upon the method of analysis used in obtaining the spectral structure. The representation may contain all the information necessary to reconstitute exactly the original temporal waveform, or only part thereof. Discrete samples or continuous functions along either the time or frequency axis may be used in the above representation. Of most interest in this thesis will be the cases where ~~discrete~~ separate spectral samples are used.

Depending upon the choice of temporal and spectral "window functions" there are unlimited possibilities for time-frequency portrayals (Blackman & Tukey, 1958; Gabor, 1946). A mathematically simple one suggested by Gabor (1946, 1953) uses the Gaussian elementary function:-

$$\mathrm{CIS}(f_n, t_k, t) = \frac{2^{1/4}}{(\Delta t)^{\frac{1}{2}}} \exp(j\beta) \exp\left[-\frac{\pi(t-t_k)^2}{(\Delta t)^2} + 2\pi\, j\, f_n t\right] \quad \ldots(2.1)$$

The Fourier integral of the CIS function (that is, its spectral window) is curiously of the same form, but Δt is replaced by Δf where $\Delta f \;\; \Delta t = 1$.

Gabor has shown that any complex signal, $f(t)$, can be expressed as the sum of Gaussian elementary functions as follows:-

$$f(t) = \sum_n \sum_k C_{nk} \, CIS(f_n, t_k, t) \qquad ...(2.2)$$

where $f_n = n\Delta f$, $t_k = k\Delta t$, and C_{nk} is a complex number.

The parameters Δf and Δt are considered to be the 'effective spectral width' and 'effective duration' respectively of the elementary function. Gabor defines these parameters as the square roots of the normalized second moments about the mean of the spectral and the time windows, and shows that, for this definition, the product $\Delta f \ \Delta t$ is greater than one for all functions other than the Gaussian.

In practice, however, it is not usually convenient to analyse a waveform to obtain its time-frequency representation in terms of the Gaussian elementary function, and other finite (time-limited) functions may be preferable. Unfortunately, for time-limited functions Gabor's definition of Δf and Δt leads to a diverging integral in the calculation of Δf so that the $\Delta f \ \Delta t$ product becomes meaningless.

Many alternative definitions for Δf and Δt exist (Fant, 1959; Wainstein & Zubakov, 1962; Slepian et al, 1961), however, one which will be used throughout the remainder of this thesis when referring to temporal and spectral windows. (Δf and Δt) will be defined as the widths of normalized rectangles having the same areas as those under the spectrum envelope and time function envelope respectively (Fig. 2.1). The product $\Delta f \ \Delta t$ is always equal to one in this case (Wainstein & Zubakov, 1962).

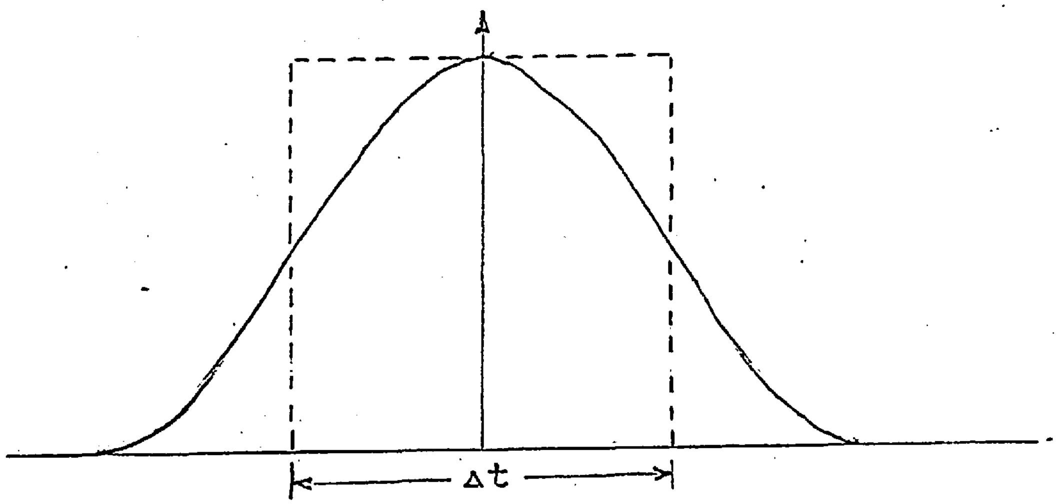


Fig. 2.1 Illustration of the definition of effective duration, Δt, of a temporal window function.

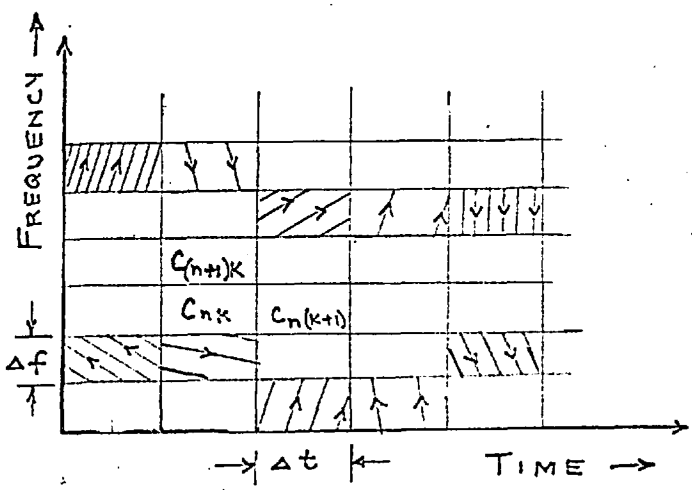


Fig. 2.2 Time-frequency representation of a temporal waveform.

This definition is particularly applicable in cases where most of the areas under the spectral and time functions are confined to the Δf and Δt intervals.

Since the Δf Δt product is one it can be seen from Equation 2.2 that for a signal f(t) which is (essentially) confined to frequency 0-W and time 0-G, there will be WG coefficients C_{nk} on the time-frequency plane. (G is used to represent time here rather than the conventional symbol T, to prevent confusion in later equations where T is used to represent the integration time in the filters of the I.C. analyser). Inasmuch as the coefficients are complex, this is in agreement with the fundamental theorem in communications (Shannon, Gabor, Woodward) that 2 WG samples are necessary to specify precisely a signal limited to time G and bandwidth W. In the spectral representation the coefficients C_{nk} can be regarded as effectively independent, although this is strictly true only when they are orthogonal.

An illustration of a time-frequency representation using Gaussian elementary window functions is given in Fig. 2.2 where line shading represents the moduli of the coefficients C_{nk} and the line directions phase. It must be stressed that for complete specification of the temporal signal the Δf Δt areas remain constant at one, but the "aspect ratio" $\Delta f/\Delta t$ can vary. Since the original waveform can be built up from the information contained in the C_{nk} coefficients, it is possible to transform from one aspect ratio to another by mathematical manipulation.

2.2. Time-Frequency Analysis by Filtering.

The analysis to obtain the necessary coefficients for a time-frequency representation such as that of Fig. 2.2, can be achieved with a suitable bank of band-pass filters. The temporal windows have the same form as the envelope of the response of a band-pass filter to a Dirac pulse applied to its input. The spectral windows, or filter frequency responses, are the Fourier integrals of the filter impulse responses.

The parameters Δf and Δt are very useful in specifying filter bandwidths and 'memory' times. The impulse response of a filter can be regarded as its memory function and Δt as the time during which the filter output responds significantly to an input, while Δf is the effective bandwidth of the filter frequency response curve.

The required WG complex spectral coefficients for complete signal specification can be obtained by feeding the signal into a parallel bank of filters tuned contiguously at intervals of Δf cycles per second, and periodically sampling both modulus and phase of all the filter outputs simultaneously at intervals of Δt seconds. All the necessary information is obtained providing the band-pass filters have suitable characteristics.

In practice, however, phase information is normally not recorded owing to difficulty of measurement and lack of suitable means for visible presentation. This means that, where WG coefficients are obtained, half of the information in the communication sense of the word is lost (Gabor, 1946). Modulus samples obtained at Δt intervals are sufficient only to show the power spectral structure of a signal. If a signal encoded into its spectral coefficients is to be transmitted over a communication channel, it is desirable that phase information be present to enable exact resynthesis of the temporal waveform at the receiver. However, transmission in this form is not often considered for other than speech-like signals where a prior knowledge of the signal structure is available.

A prior knowledge of speech may make it possible to compensate at the receiver for the lack of phase information in a signal transmitted in terms of its power spectrum. Advantages in bandwidth compression may then justify transmission in this form where the information loss after compensation at the receiver is not enough to impair perception significantly.

Since transmission ~~considerations are of~~ characteristics offer little ~~importance~~ difficulty in visible spectral representations, one need not be so concerned with the exact number of spectral coefficients available, where extra samples only serve to enhance detail. This philosophy is adhered to in the Kay Sonagraph where the analysing filter output amplitude is read continuously, presenting effectively many more than WG modulus coefficients. Due to the 'minimum phase' characteristics of the vocal tract (Fant, 1959) phase in speech waveforms can be approximately inferred from this time-frequency display as will be seen in Chapter 4. Of course, it would be undesirable to transmit directly the spectral information in the form obtained from the Sonagraph, since the transmission channel would need a bandwidth capacity much in excess of the bandwidth of the original speech waveform.

Summarizing, a temporal waveform limited to time G and bandwidth W may be specified almost exactly by WG complex spectral coefficients, while half the information is lost when phase information is omitted to make the coefficients real. In signals such as speech where a prior knowledge exists, some of the phase information can be inferred from the spectral display aided by increasing the number of real spectral coefficients, although half the information is theoretically still not available.

2.3. Filters.

2.3.1. Operation of Passive Low-Pass and Band-Pass Filters.

To understand fully the theoretical basis of spectrographic analysis, the formal relationships between Fourier integral analysis and the function of band-pass and low-pass filters should be considered. From this study follows the explanation of the function of sampling filters where periodic time-weighting is used. Passive filters will be considered from the viewpoint chosen by Fant, 1959.

Given a band-pass filter with a response to a Dirac pulse (impulse response) of $h(t)$ and Fourier transform $H(f)$, the exact expression for the output $g(t)$ of the filter when the input is connected to a source $f(t)$ is given by the convolution integral

$$g(t) = \int_0^t f(\tau)\, h(t-\tau) d\tau = \int_0^t f(t-\tau)\, h(\tau) d\tau \qquad \ldots(2.3)$$

If the band-pass filter is symmetrical the impulse response $h(t)$ can be expressed as an amplitude modulated carrier

$$h(t) = h_0(t) \cos(\omega_0 t + \emptyset_0) \qquad \ldots(2.4)$$

It is convenient to consider $h_0(t)$ to be the impulse response of a low-pass filter from which the band-pass filter has been derived using Laurent's (1952) em-transformation with a factor $\frac{1}{2}$ and involving the substitution of ω for $\omega/2 \left[1 - \omega_o{}^2/\omega^2\right]$. Band-pass filters can then be treated as special cases of low-pass filters having half the bandwidth, with a resulting simplification in the analysis.

The convolution of Equation 2.3 comprises a folding back from the instant t of the function $h(\tau)$, followed by a multiplication of $h(\tau)$ with corresponding values of $g(t - \tau)$ of the past history of the input signal, and finally an integration. Fig. 2.3 illustrates this where a low-pass filter is considered, the integration occurring after the multiplication. The function $h_o(t)$ is the impulse response of a low-pass filter and would normally be a smoother function in practical filter circuits.

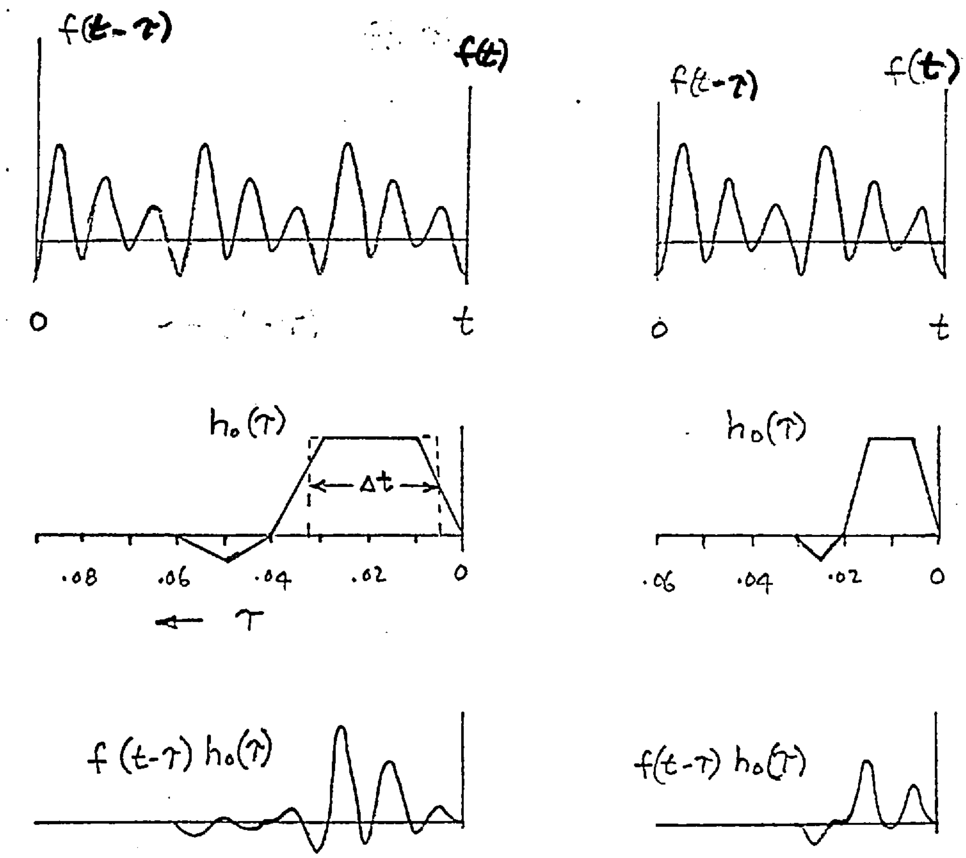


Fig. 2.3 Convolution of a speech waveform by means of a narrow-band filter (weighting function of longer duration) and a wide-band filter(weighting function of shorter duration).

The function of a band-pass filter (Kupfmuller, 1949; Fant, 1959) can be broken down into two successive operations. The first is the weighted sampling of the input resulting in the product $f(t - \tau)\, h_o(\tau)$ and the second is the application of a formal Fourier integral transform to generate a spectral term of frequency ω_o. In normal filter circuits the amplitude and phase of this component are continuously time-varying.

Consider the expanded expression for the output of a band-pass filter.

$$g(t) = \int_0^t f(\tau)\, h_o\,(t - \tau) \cos\left\{\omega_o(t - \tau) + \phi_o\right\} d\tau$$

$$= \int_0^t f(t - \tau)\, h_o(\tau) \cos\,(\omega_o\,\tau + \phi_o)\, d\tau \qquad \ldots(2.5)$$

On solution of Equation 2.5 the result is

$$g(t) = a(\omega_o, t) \cos\,(\omega_o t + \phi_o) + b(\omega_o, t) \sin\,(\omega_o t + \phi_o) \qquad \ldots(2.6)$$

where

$$\left.\begin{aligned} a(\omega_o, t) &= \int_0^t f(\tau)\, h_o\,(t - \tau) \cos \omega_o\, \tau d\tau \\ b(\omega_o, t) &= \int_0^t f(\tau)\, h_o\,(t - \tau) \sin \omega_o\, \tau d\tau \end{aligned}\right\} \qquad \ldots(2.7)$$

The coefficients in Equation 2.7 are identical to the amplitude coefficioients of a Fourier integral if the weighting function $h_o (t - \tau)$ is given a constant value of $1/\pi$ and the integration limits are extended to plus and minus infinity.

The spectrum of the filter output is obtained by applying a Fourier integral transformation to the convolution integral of Equation 2.3.

$$G(f) = F(f) \cdot H(f) \qquad \text{...(2.8)}$$

Where G(f) is the filter output spectrum consisting of the spectrum of the input F(f), weighted by the spectrum of the filter impulse response, H(f).

2.3.2. Modifying Filter Characteristics by Time-Weighting.

Time-weighting denotes here the modulation or multiplication of the signal which is being fed into a filter by some arbitrary function in order to modify the output from the filter. Within certain limitations the effective frequency response of a filter may be given any desired shape by this means. The main limitation is that only certain prescribed instants in time are allowed for observation of the filter output for a required frequency response.

Before considering time-weighting, observe what happens when when a filter's characteristics are altered (normally accomplished by modifying its circuit). Suppose that the alteration is equivalent to multiplying the filter impulse response by a weighting function $K(t)$ to obtain a new impulse response $h(t)\ K(t)$. The output $g(t)$ of this filter when connected to a source $f(t)$ is then given by

$$g(t) = \int_0^t f(\tau)\ h(t-\tau)\ K(t-\tau)\ d\tau = \int_0^t f(t-\tau)\ h(\tau)\ K(\tau)\ d\tau \quad ...(2.9)$$

Equation 2.9 is illustrated in Fig 2.4 where, after the impulse response is modified by the weighting function, normal convolution of the speech waveform by the new impulse response occurs.

Now consider the case where the source $f(t)$ is multiplied by the weighting function $K(T-t)$. Suppose that the output from a filter with impulse response $h(t)$ is observed at time T after the weighted signal $f(t)\ K(T-t)$ has been applied to its input. The result is given by

$$g(T) = \int_0^T f(\tau)\ K(T-\tau)\ h(T-\tau)\ d\tau = \int_0^T f(T-\tau)\ K(\tau)\ h(\tau)\ d\tau \quad ...(2.10)$$

which is illustrated in Fig. 2.5.

The weighted input signal is convolved by the filter impulse response $h_o(\tau)$ in the normal way. However, the weighting function $K(T-t)$ <u>can</u> be regarded as a part of the filtering process, modifying the effective filter characteristics (rather than the signal) in the integration interval 0 - T. This also

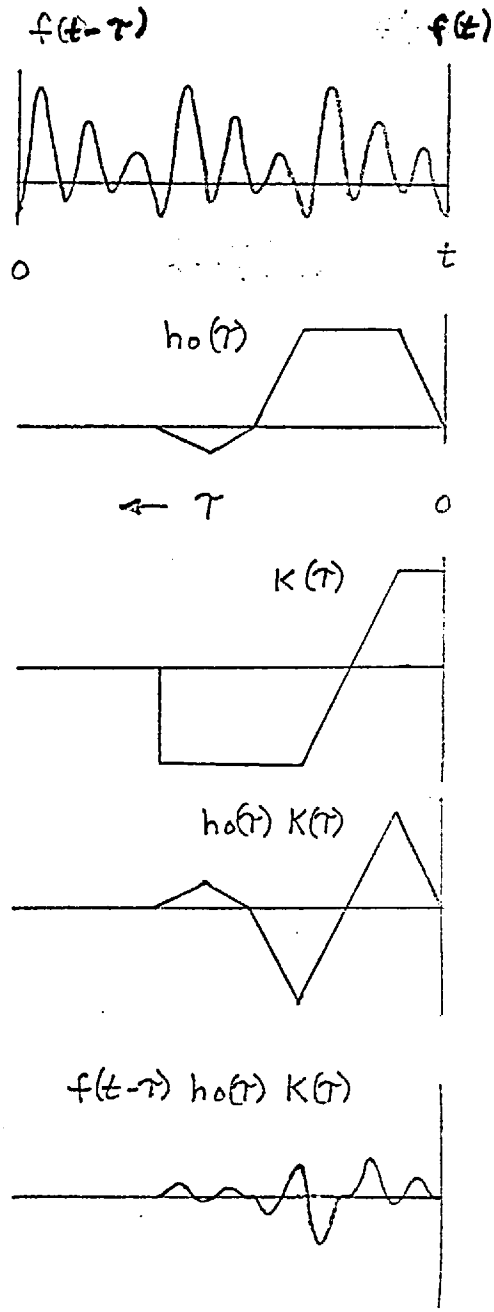


Fig. 2.4 Convolution of a speech waveform by means of a filter whose impulse response $h_o(\tau)$ has been modified by multiplication with a weighting function $k(\tau)$.

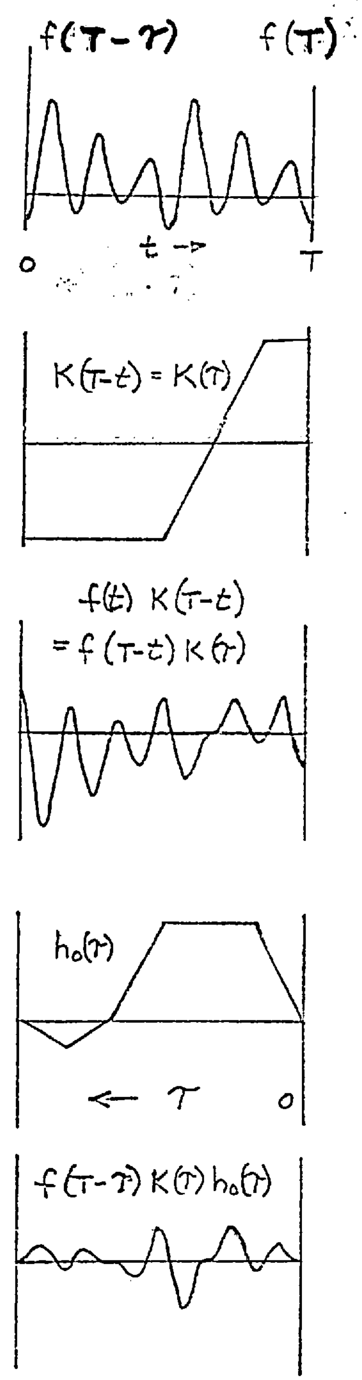


Fig. 2.5 Convolution of a speech waveform weighted by the function $K(T-t)$ by means of a filter with an impulse response $h_o(\tau)$.

means that if the filter output is observed at a time other than T, the effective characteristics will again be different.

By comparison of Equations 2.9 and 2.10 and Figures 2.4 and 2.5, it can be seen that at time $t = T$, the output from the filter with impulse response $h(t)\ K(t)$ will be identical to that in the case where the source is weighted by the function $K(T - t)$ before being applied to a filter with impulse response $h(t)$.

Then to effectively modify a filter impulse response in a definite time interval $0 - T$, the procedure can be (a) choose the function which when multiplied by the filter impulse response gives a desired response, (b) fold the function back in time from the instant T, and (c) multiply the input to the filter by the folded back function. Further, if both a weighting function $K(t)$ and a filter impulse response $h(t)$ go to zero, after a time T, the output from a filter with impulse response $h(t)$ where its input has been weighted periodically by the function $K(T - t)$, will be identical at intervals of T to the output from a filter with impulse response $h(t)\ K(t)$ where no input weighting is applied. It can be seen from the foregoing that periodic sampling filters with easily variable effective impulse responses can be synthesized through time-weighting the inputs to conventional filters with suitable periodic weighting functions.

2.4. Filters in the Imperial College Analyser.

The filters in the spectral analyser are active and designed to act as lossless parallel resonant LC circuits where the impulse response can be forced to zero (quenched) after any time T. Since the circuit is lossless, its effective impulse response envelope when quenching is used, is a rectangular temporal window, and any weighting function used in conjunction with the circuit simulates an effective impulse response of the same form. One need then only generate a desired periodic "impulse response envelope" for weighting the source signal to obtain the required sampling filter.

The circuit and its description are included in the Appendix, but the properties of a parallel resonant tuned circuit will be discussed here (Robin et al, 1963).

Consider first a parallel resonant circuit L, C and R, where R is the parallel resistance, driven from a current source I sin ωt. The sum of the currents in the components as a function of the voltage on the capacitor is written below as a differential equation.

$$C \frac{dv}{dt} + \frac{v}{R} + \frac{1}{L}\int v\,dt = I \sin \omega t \qquad \ldots(2.11)$$

on solving for v the result is

$$v = A \exp(-at) \sin(\omega_d t + B) + bI \sin(\omega t + \phi) \qquad \ldots(2.12)$$

where $a = 1/2\,CR$

$$b = \frac{R}{\left[(1 + \omega^2C^2R^2)(1 - \omega_o^2/\omega^2)^2\right]^{\frac{1}{2}}}$$

$$\phi = \tan^{-1} 2\,CR\,(\omega - \omega_o)$$

$$\omega_d = (1/LC - 1/4\,C^2R^2)^{\frac{1}{2}}$$

$$\omega_o = 1/(LC)^{\frac{1}{2}}$$

A and B are constants depending on initial conditions.

It can be seen that the solution contains two components:

(a) the steady state solution which is a continuous sine wave of the same frequency as the driving waveform and amplitude proportional to b,

(b) the transient component which is an exponentially decaying sinusoidal waveform of frequency ω_d and time constant 1/a.

If the circuit becomes lossless $R = \infty$, $a = 0$, $\omega_d = \omega_o$ and $\phi = \pi/2$, so that Equation 2.12 reduces to

$$v = A \sin(\omega_o t + B) + \frac{I\omega}{C(\omega^2 - \omega_o^2)} \sin(\omega t + \pi/_2) \qquad \ldots(2.13)$$

Choosing the initial conditions such that $v = dv/_{dt} = 0$ at $t = 0$ we find $A = -\dfrac{I\omega}{C(\omega^2 - \omega_o^2)}$ and $B = \pi/2$ so that Equation 2.13 reduces again to

$$v = \frac{2I\omega}{C(\omega + \omega_o)(\omega - \omega_o)} \sin \frac{\omega + \omega_o}{2} t \cdot \sin \frac{\omega - \omega_o}{2} t \qquad \ldots(2.14)$$

Putting $\omega - \omega_o = \Delta\omega$ Equation 2.14 becomes

$$v = \frac{It}{2C} \frac{4\omega}{t(2\omega_o + \Delta\omega)\Delta\omega} \sin \frac{\Delta\omega}{2} t \sin \frac{(2\omega_o t + \Delta\omega)}{2} t \qquad \ldots(2.15)$$

In the limiting case where $\Delta\omega = 0$ the output is given by

$$v = \frac{It}{2C} \sin \omega t \qquad \ldots(2.16)$$

The envelope of the output rises linearly with time when $\Delta\omega = 0$. Equation 2.15 is illustrated in Fig. 2.6 for three values of $\Delta\omega$.

If we assume that $\Delta\omega \ll \omega_o$ in Equation 2.15, the amplitude of the voltage output of the lossless resonant circuit at time $t = T$ is given by

$$v_T \approx \left| \frac{IT}{2C} \frac{\sin \frac{\Delta\omega}{2} T}{\frac{\Delta\omega}{2} T} \right| \qquad \ldots(2.17)$$

This is the effective frequency response of a lossless resonant circuit when the output is observed after a finite time T. This frequency response is illustrated in Fig. 2.7.b where the output of a sampling filter is repeatedly observed at intervals of T seconds as $\Delta\omega$ is varied linearly. The operation frequency of the analyser filters are high enough that the relationship $\Delta\omega \ll \omega_o$ holds.

It should be noted that the frequency response of a lossless parallel tuned circuit can be obtained more easily by observing that its response to a Dirac pulse is $K \cos \omega_o t$ where $\omega_o = 1/(LC)^{\frac{1}{2}}$. That is, when the lossless resonator is pulsed it rings for an indefinite time at a constant amplitude and frequency

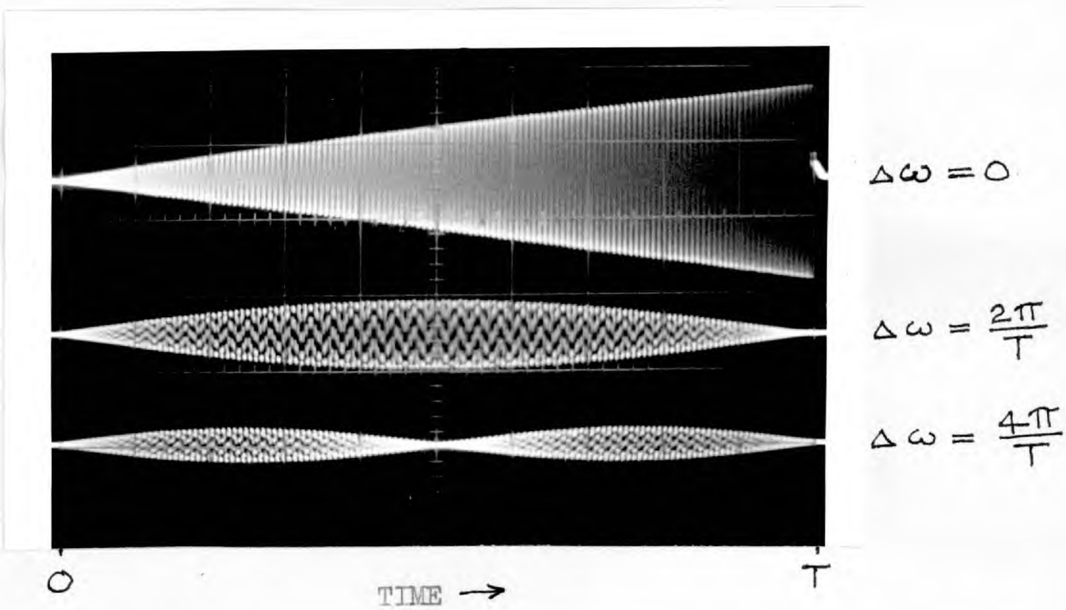
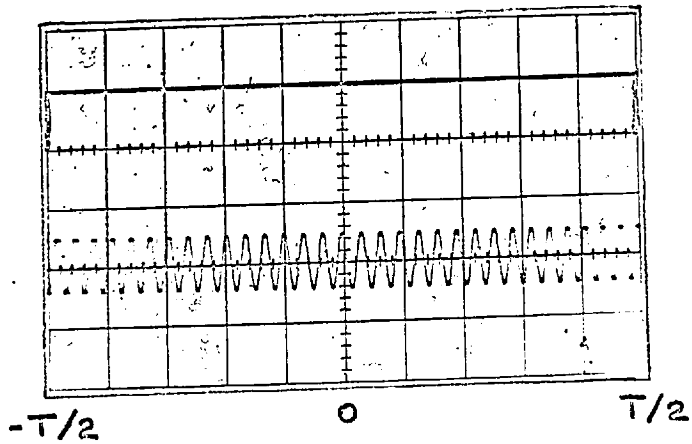
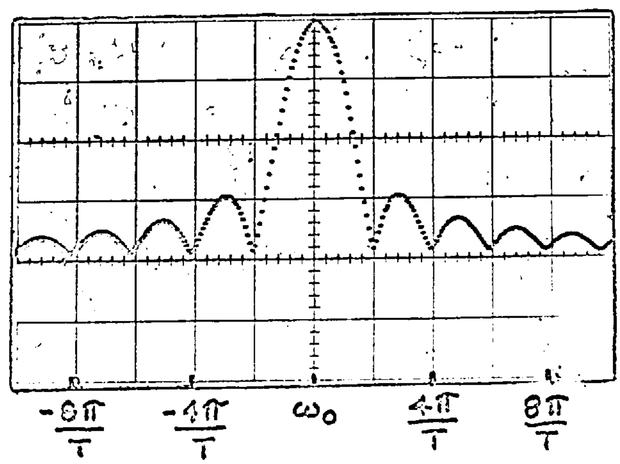


Fig. 2.6 Tuned circuit response with time for three different frequencies applied to input.



(a)



(b)

Fig. 2.7 Frequency response of the lossless quenched resonant circuit. *(Amplitude Scale linear).*

ω_o. The frequency response after a definite time T can be found by considering the impulse response to consist of a time-limited sinusoid of duration T as in Fig. 2.7.a and computing the Fourier integral transform directly. The amplitude of the frequency response is then given by

$$v_T = K\frac{T}{2} \left| \frac{\sin\left(\frac{\omega - \omega_o}{2}\right) T}{\left(\frac{\omega - \omega_o}{2}\right) T} + \frac{\sin\left(\frac{\omega + \omega_o}{2}\right) T}{\left(\frac{\omega + \omega_o}{2}\right) T} \right| \qquad \ldots(2.18)$$

where the first term in Equation 2.18 is identical to Equation 2.17 when K = I/C. In the practical case where $\Delta\omega \ll \omega_o$ the second term in Equation 2.18 is negligibly small.

2.5. Time-Weighting with Analyser Filters.

To illustrate the effects of time-weighting used in conjunction with a lossless resonator, some examples of photographs are shown in Fig. 2.8 to 2.17. Figures 2.9.a to 2.17.a show one cycle of a periodic weighting function having just below it a weighted sinusoidal waveform. The weighted sinusoid (temporal window function) chosen for convenience can be regarded as a synthesized filter impulse response, differing only in that the effective impulse response is reversed (folded back) in time from that shown. (Where the temporal windows are symmetrical functions this is of no consequence.)

The frequency response curves in Figures 2.9.b to 2.17.b (spectral window functions) consist of amplitude samples of the filter output voltage observed at intervals of T seconds, while the frequency of a sinusoidal source is swept slowly and linearly through the tuned region of the filter. (It must be remembered that the filter memory is quenched at intervals of T seconds immediately after each output sample is observed). The smoothed spectral window functions of the form shown can easily be calculated, since they are equal to the moduli of the Fourier integral transforms of one cycle of the weighting functions.

Fig. 2.8. shows the effect of weighting a sinusoidal input to one of the lossless filters with a raised cosine function. In this case $\Delta\omega$ is equal to zero. The waveform in the filter during the integration period 0 - T is shown above the raised cosine function which is used for weighting the input. The frequency response resulting from such a weighting function is shown in Fig. 2.9.

The weighting functions shown in Figures 2.9.a and 2.10.a are of interest in spectral analysis because they are easily used in mathematical calculations, easily generated in ordinary sinusoidal oscillators and their spectral windows have low sidelobes. It can be seen from Figures 2.9.c and 2.10.c that the Hanning window gives spectral sidelobes rapidly decreasing as $\Delta\omega$ increases while the Hamming window gives sidelobes at a lower level, but remaining relatively constant with increasing $\Delta\omega$. The maximum sidelobe for the Hanning window is approximately 32 db below the main lobe and that for the Hamming window approximately 43 db down.

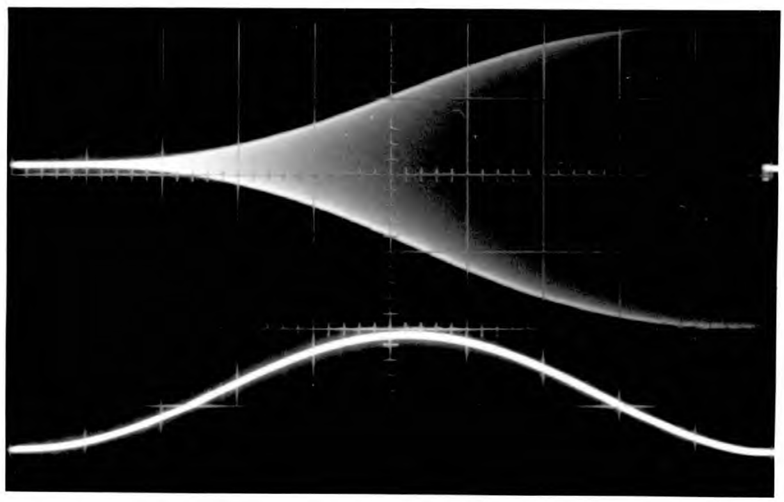


Fig. 2.8 One integration period in a lossless tuned circuit when a sinusoidal source at frequency ω_o is weighted with a raised cosine function.

In many applications (particularly in visible spectrograms) it will be desirable that an analysing filter response have low sidelobes, in which case the Hanning and Hamming window functions can prove useful. It should be noted that the Hanning window provides a tolerable approximation to a truncated elementary Gaussian window function.

Fig. 2.11 is of interest since it resembles closely the response of a tuned RLC circuit. The impulse response of such a circuit is an exponentially decaying sinusoid, and the resonance curve is of the form shown in Fig. 2.11.b. The main difference is the rippled appearance of the spectral window in the photograph, due to truncating the exponential weighting function after time T.

From Figures 2.12 and 2.13 it can be seen that adding a constant to the temporal window has the effect of adding a sin x/x function to the spectral window function. If the truncated sin x/x weighting function in Fig. 2.12 could be extended, the spectral window would more and more closely resemble a rectangular function.

The remaining Figures (2.14, 2.15, 2.16, 2.17) serve to illustrate the versatility of the active filters in the IC analyser and the ease with which a large variety of spectral characteristics are obtained. Although they may not be very suitable for obtaining visible presentation of speech spectrograms owing to the sidelobe structures, they might prove interesting for studies in signal encoding (such as transmission in terms of

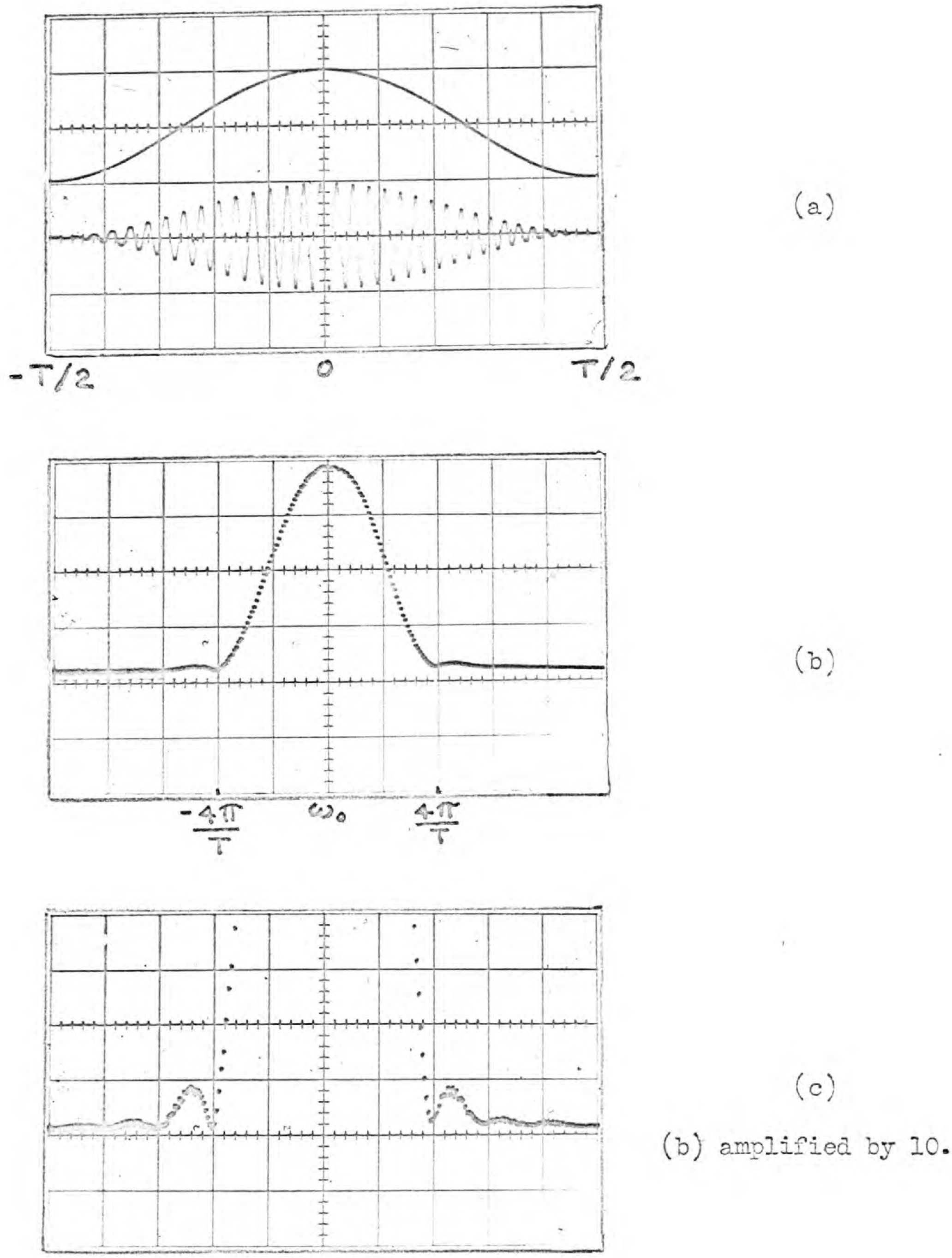
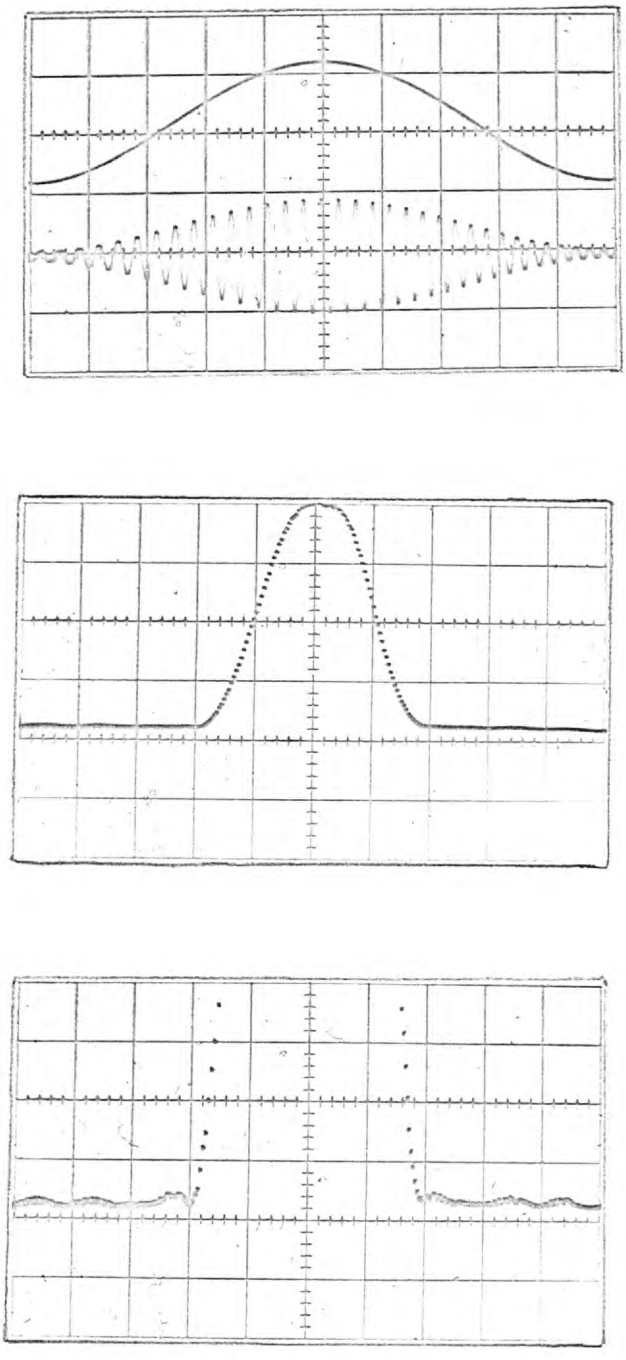


Fig. 2.9 Effect of weighting with a raised cosine or 'Hanning' window, i.e. $0.5(1 + \cos\pi t/T)$ from $-T/2$ to $T/2$.



(a)

(b)

(c)
(b) amplified by 10.

Fig. 2.10 'Hamming' window, $0.54 + 0.46\cos\pi t/T$.

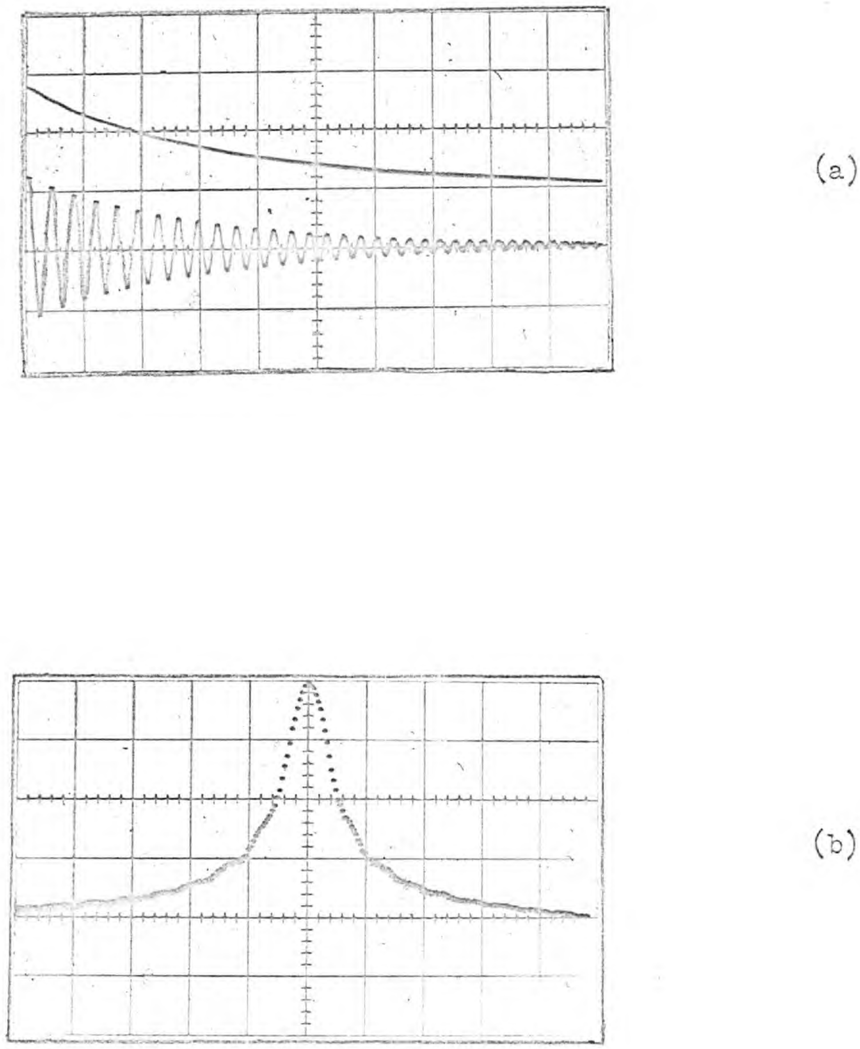
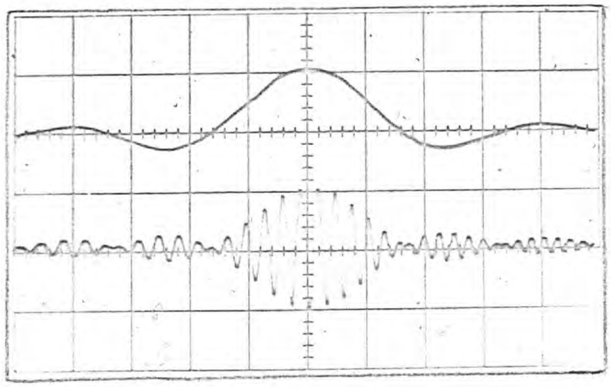
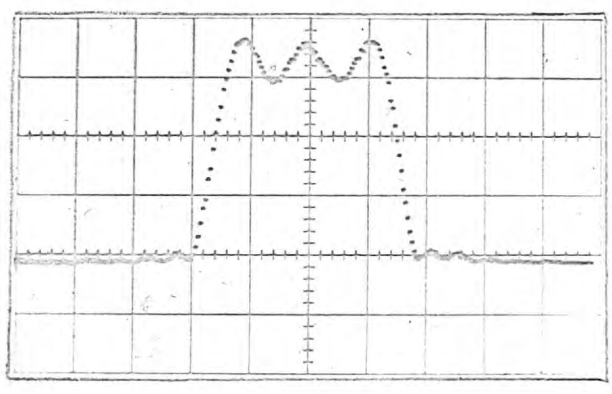


Fig. 2.11 Truncated exponential window.

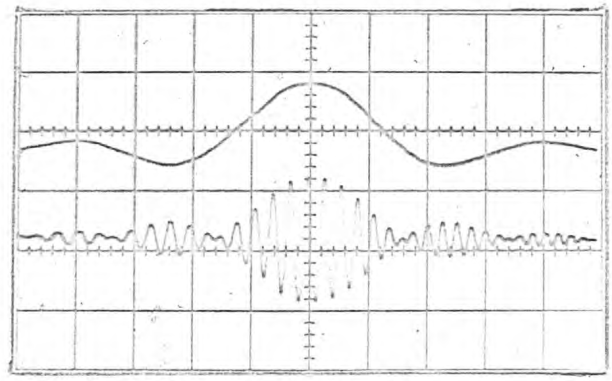


(a)

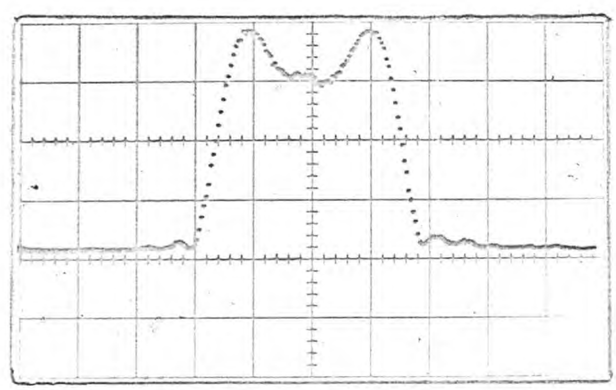


(b)

Fig. 2.12 Truncated sinx/x window.

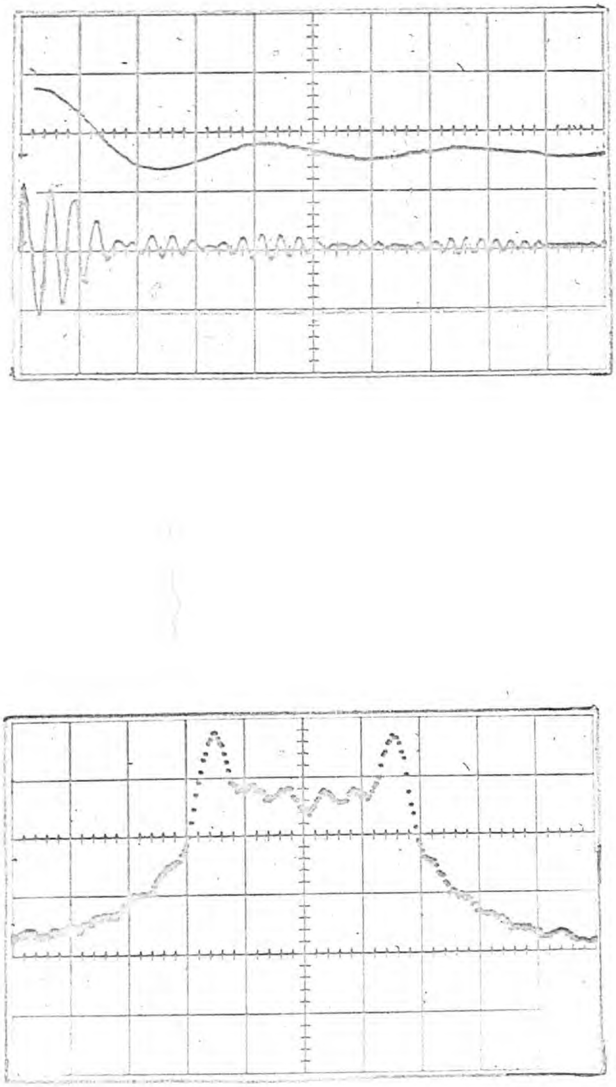


(a)



(b)

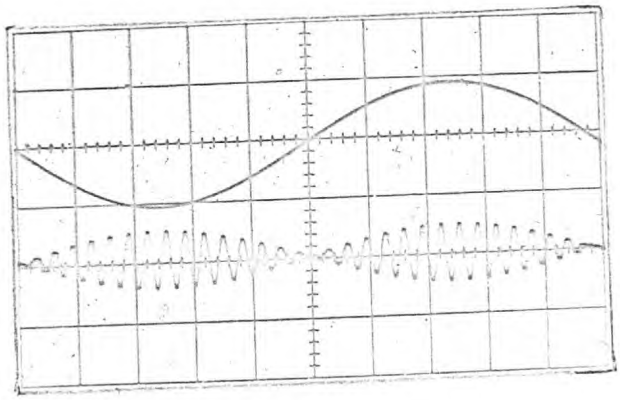
Fig. 2.13 Truncated sinx/x - constant.



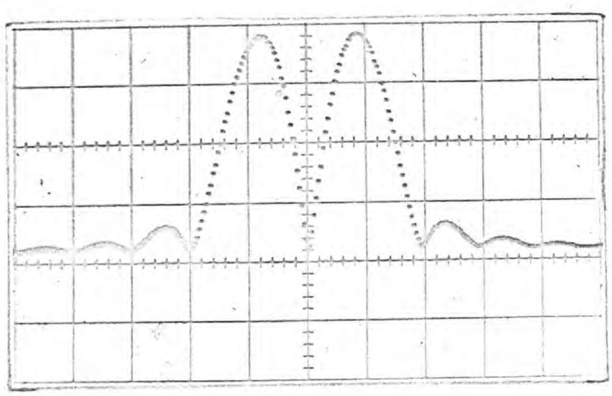
(a)

(b)

Fig. 2.14 Half of truncated sinx/x + constant.

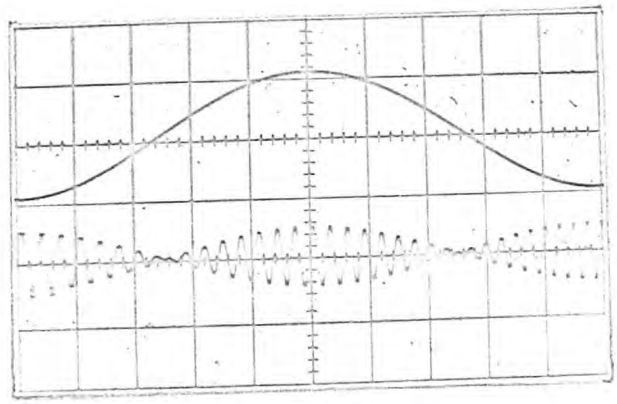


(a)

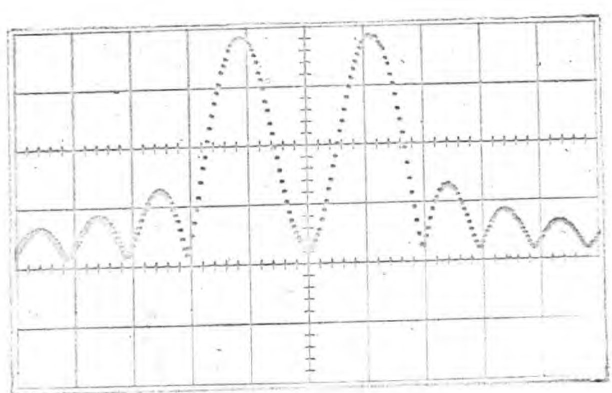


(b)

Fig. 2.15 One cycle of sine wave.

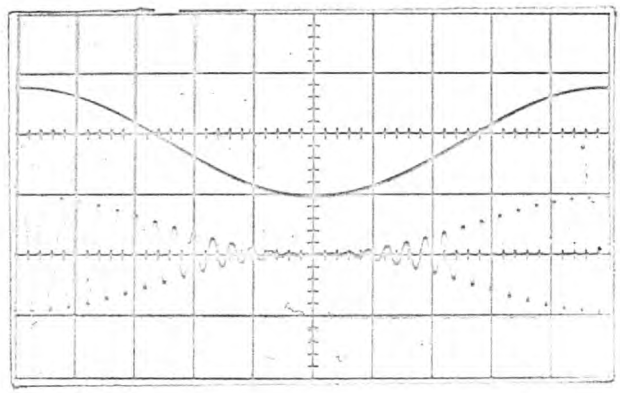


(a)

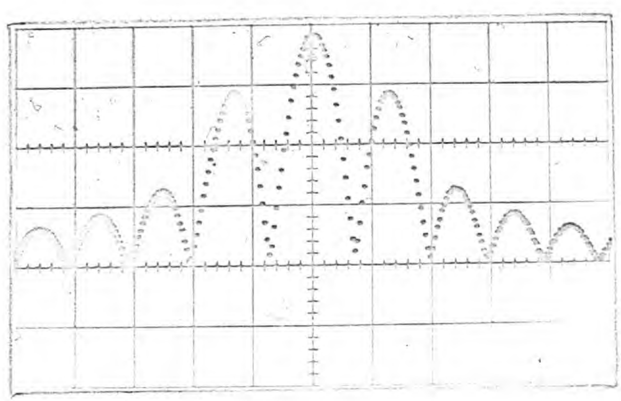


(b)

Fig. 2.16 One cycle of cosine wave.



(a)



(b)

Fig. 2.17 Inverted raised cosine, $0.5(1 - \cos\pi t/T)$.

orthogonal functions or consideration of the best shape for sonar pulses).

2.5. Time-Frequency Representation with the Imperial College Analyser.

As mentioned previously, the outputs from a parallel bank of filters should be observed at intervals of Δt seconds to obtain the necessary coefficients for a spectral representation. When time-weighting is used with the active filters in the IC analyser, the filter outputs can be observed only at intervals of T seconds, and since in most temporal windows Δt is less than T, one bank of filters tuned at intervals of Δf cycles per second is not sufficient to obtain WG spectral coefficients for a suitably band-limited and time-limited signal.

Consider the case where the Hanning window, with Δt equal to T/2 is used for time-weighting. If two identical banks of filters are now employed in the analysis their outputs may be sampled alternately, while the Hanning windows are applied periodically and out of phase to the separate filter bank inputs as illustrated in Fig. 2.18. Waveform A will be applied to the input of filter bank A and waveform B to the input of filter bank B. Each filter bank output is observed at the end of its appropriate T intervals, combining to provide a set of spectral samples at every Δt interval.

The associated spectral windows for the weighted filters

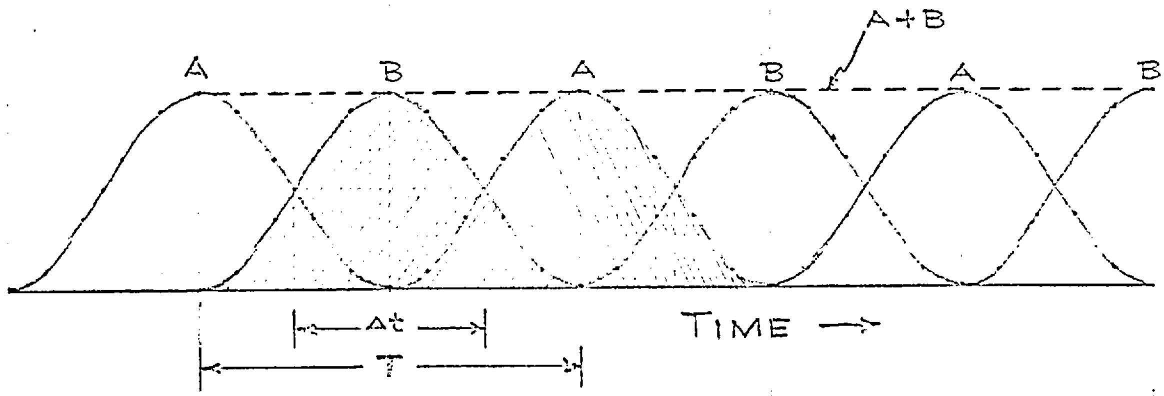


Fig. 2.18 Hanning temporal window functions staggered in time.

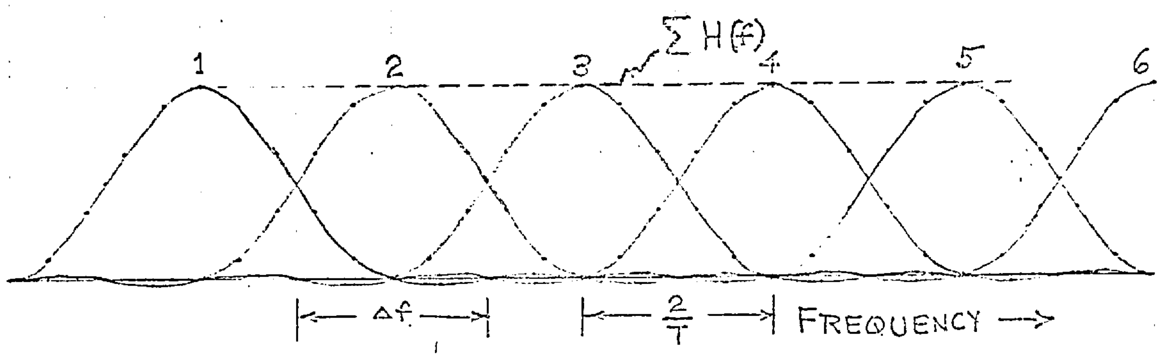


Fig. 2.19 Effective spectral response of a parallel bank of filters after time-weighting with the Hanning window.

spaced at Δf cycle per second intervals are shown in Fig. 2.19. It can be seen that none of the spectral coefficients will be completely independent of the immediate neighbours either on the time or frequency axes due to the overlapping.

Since two banks of filters are required in the above case, identical banks have been provided in the IC analyser, in order that WG real spectral samples may be obtained. It should be noted that Δt is not often equal to T/2 and, in fact, is equal to T in the case where the temporal window is rectangular as in Fig. 2.8.a. In this case only one bank of filters is required to give the necessary number of spectral coefficients. An interesting property of the rectangular time window is that, where the filters are tuned at Δf C/S intervals, the resulting sin x/x spectral windows are orthogonal, (Harmuth, 1960) and the spectral coefficients obtained from such filters are strictly independent of each other. In other cases where Δt is less than T/2, it is not possible to obtain complete analysis along the time axis with two filter banks.

In the foregoing discussion no mention has been made of losses or inaccuracies, owing to physical limitations of the analyser circuits. Deserving mention is the loss sustained due to the finite time required to quench the energy in the active filters which effectively shortens the integration time T. In practice the time lost in one integration interval is about 250 microseconds which is a small percentage of the total which can vary between 2.5 and 20 milliseconds. Although the effective temporal window function is slightly shortened, compensation can easily be made in any theoretical considerations, if desired.

IMPERIAL COLLEGE ANALYSER

3.1. Construction.

The spectral analyser constructed in the Communications Section at Imperial College to analyse a signal into 32 or 64 spectral bands is shown in Figures 3.1 and 3.2. The band-pass filters and all the associated control circuits are transistorized and mounted on plug-in modules for easy servicing. The modules are housed in eight Lektrokit circuit boxes having hinged front panels for easy access from the front as well as from the rear. Connections between boxes are made by multicore shielded cables for control and timing pulses, and by coaxial cables for analogue speech signals and spectral coefficients.

From Figures 3.1 and 3.2 the positions of the main circuit sections in the analyser can be seen from top to bottom as follows:-

(a) Boxes 1 to 4 - Filter banks with associated sampling and readout circuitry.

(b) Box 5 - Stages 1 to 16 of a readout shift register and some circuitry for control and synchronization.

(c) Box 6 - Stages 17 to 32 of the readout shift register and some circuitry for presenting the spectral coefficients to the display oscilloscope.

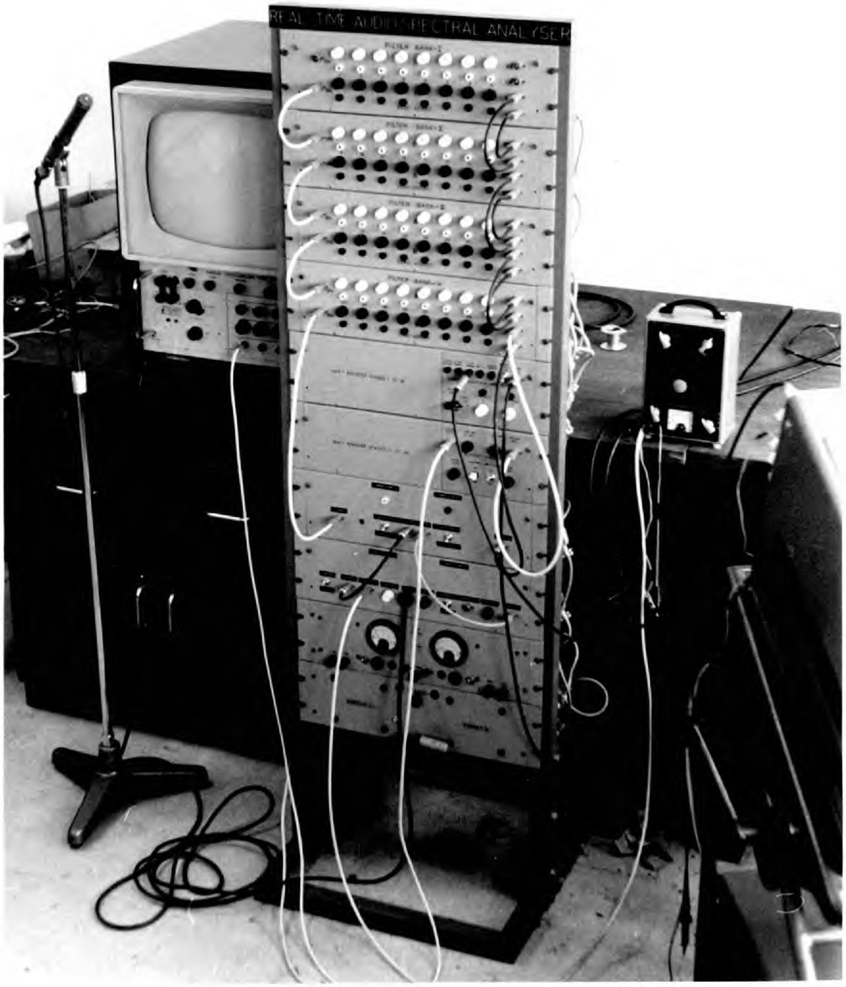


Fig. 3.1 Spectral analyser - front view.

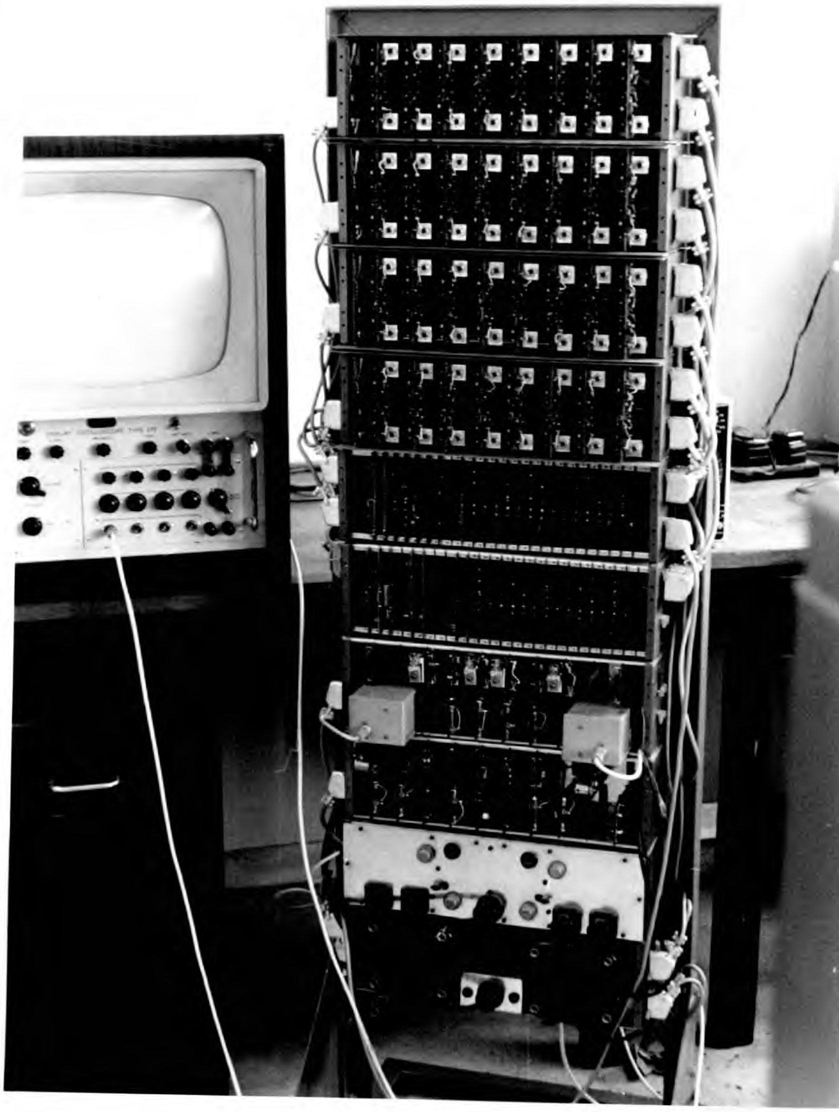


Fig. 3.2 Spectral analyser - back view showing plug-in modules.

(d) Box 7 - Two modulators with local oscillators for translating the speech signal to the frequency range where the filter banks are tuned.

(e) Box 8 - Two multipliers for time-weighting.

Below the eight boxes are the power supplies. To the left of the analyser rack is shown a display oscilloscope for the time-frequency-intensity representation of speech signals, and to the right is an audio oscillator used for time-weighting.

3.2. Block Diagram and General Operation of the Analyser.

The flow diagram of Fig. 3.3 is sufficient to outline the general operation of the spectral analyser. It will be shown in Chapter 5 that more complex analysis is possible by automatically varying filter bandwidths, but it will not be treated here.

The steps taken in speech spectral analysis are as follows:-

(a) The speech signal, which may have been recorded on magnetic tape or else spoken directly into a microphone, is passed into an equalization circuit where its higher frequency components are emphasized.

(b) From the equalizer the speech goes into two multipliers, A and B, where time-weighting

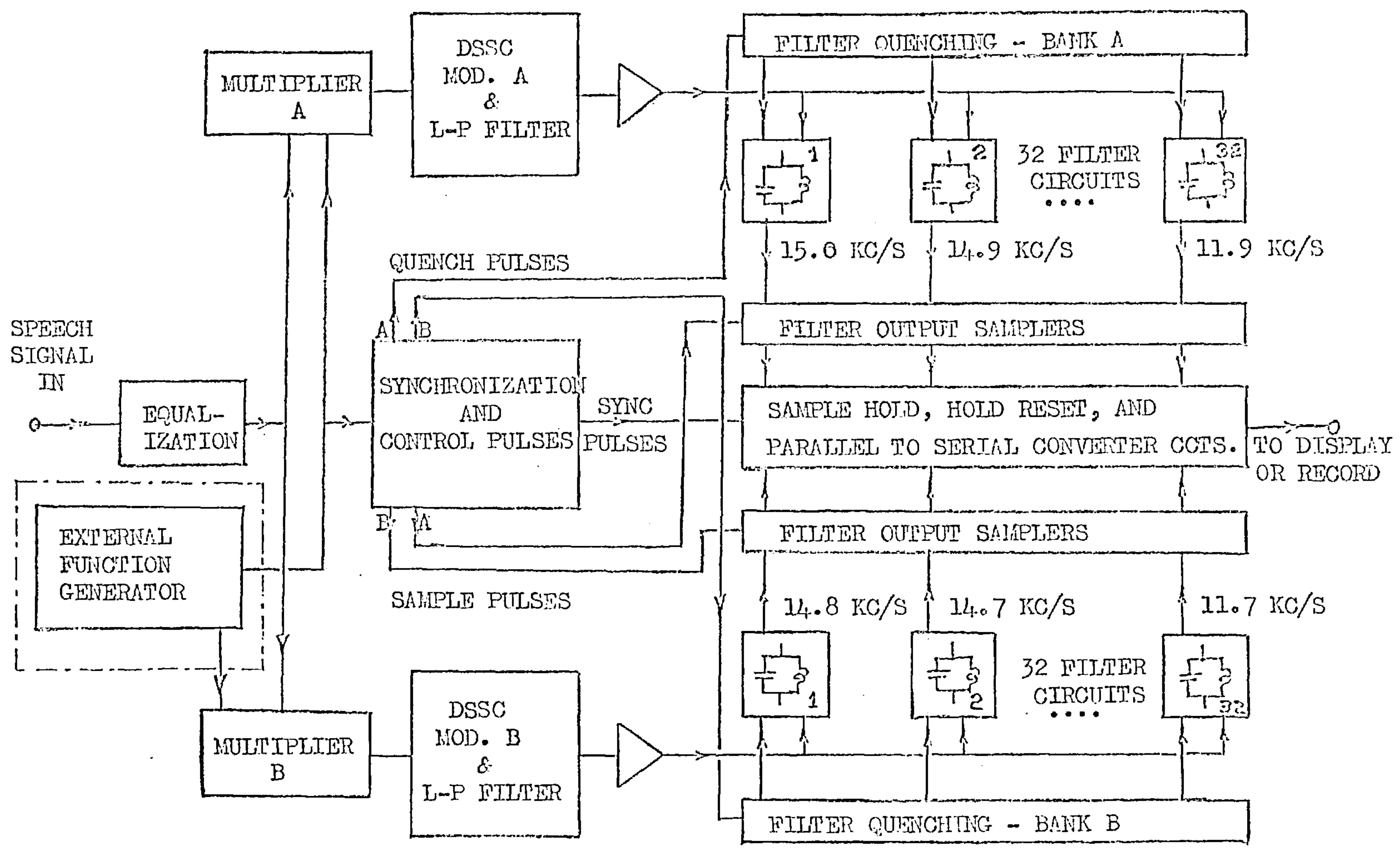


Fig. 3.3 Flow diagram of the Imperial College analyser.

occurs. A periodic weighting function is applied in one multiplier and the same function is applied out of phase in the other.

(c) The weighted speech signals from multipliers A and B go to two modulators A and B respectively. In each modulator the weighted speech signal is translated in frequency to the range where the appropriate filter bank is tuned. The carrier is suppressed and the upper sideband removed by filtering.

(d) The lower sidebands of the translated speech signals are fed in parallel to the appropriate filter banks - modulator A to Filter Bank A and modulator B to Filter Bank B.

(e) The voltage amplitudes in the filters of Bank A are sampled simultaneously at a time suitably synchronized to the weighting function. The samples are stored temporarily as voltages on a bank of capacitors and the filters in Bank A quenched. Finally, the stored spectral samples are read serially in order to be presented at the analyser output.

After this operation the filter outputs from Bank B are sampled, stored, and read out serially after the filters have been quenched. The whole process is periodically repeated.

(f) The serially presented spectral samples may be operated upon by output circuits such as a logarithmic amplifier or a low-pass filter. Normally they will be passed through a low-pass filter with a Gaussian impulse response in order to obtain smooth interpolation between the contiguous spectral coefficients, giving spectral sections as mentioned in Chapter 2.

(g) The smoothed waveform made up of spectral coefficients obtained from the analyser can then be used to modulate the intensity of a CRT where a raster (somewhat like that in television, but much slower) synchronized to the readout rate is provided, enabling visible presentation of the short-term spectral structure of the speech signal.

3.3. Comments on the Individual Analyser Sections.

Many details of the analyser circuits may be found in the Appendix and in a manual left in the Signals Laboratory with the Spectral Analyser. However, a few comments on each section of the analyser and the analysis process will be made here.

3.3.1. Equalization.

Equalization is performed by a CR circuit which emphasizes the frequencies from 500 C/S to 6 KC/S at the approximate slope of 6 db/octave (see Fig. 4.1 for the frequency response curve). This raises the energies in the speech formants to an approximately common level, which greatly aids in the discrimination of the formant structure of speech in visual spectral presentations.

3.3.2. Function Generation.

The time-weighting functions are generated by equipment external to the spectral analyser. Not only are these functions multiplied by the input speech signal to set effective filter characteristics and bandwidths, but they provide synchronization for all the operations in the analyser.

One of the most useful functions, the raised cosine or Hanning window, can be obtained from an audio oscillator. The out of phase function is then simply obtained by inversion of the sinusoid. Other weighting functions can be generated in special circuits, or else by means of a photoelectric function generator. The Signal Processing Engine in the Signals Laboratory of the Communications Section can be modified to generate functions which have been first painted or cut out on a card (Appendix 3).

3.3.3. Multiplication.

The multipliers used for time-weighting are four-quadrant time division analogue multipliers in which the product is represented by the net area under a square wave having simultaneous modulation of pulse width and pulse amplitude. They are based on the principles described by Atkins, 1958, where transistorized switching circuits are used throughout. Low distortion (approx.1%)

in the product is obtained for signals with a sum frequency (sum of weighting function and speech signal frequencies) of 12 KC/S. Each multiplier input will accommodate signals with frequencies between 10 C/S and 12 KC/S.

3.3.4. Modulation and Low-Pass Filtering.

The modulators are of a Double Sideband Suppressed Carrier, DSSC, type. Square wave carriers are used to facilitate linear modulation of the carrier waves by the speech signals with simple transistor switching circuits. Two local oscillators are included with each modulator, generating carriers at frequencies in the region of 15 KC/S and 18.2 KC/S. By switching between the two carrier frequencies the analysis in each filter bank can be performed over two ranges of the speech spectrum, 0 - 3.2 KC/S or 3.2 - 6.4 KC/S. Thus, one filter bank can be stacked above the other in the frequency spectrum as will be seen in Chapter 4, or if desired both banks may be moved to the higher range.

The upper sidebands in the modulator outputs are removed by sharp cutoff low-pass filters. In a sense the filter banks are themselves equivalent to low-pass filters. In practice, however, removal of the upper sideband first eliminates interference effects in the first two or three band-pass filters (those covering the first two or three hundred cycles of speech) in the filter banks.

3.3.5. The Banks of Band-Pass Filters.

Each of the two filter banks comprises 32 active band-pass filters of the type described in Chapter 2, tuned contiguously at 100 C/S intervals and fed in parallel from a single power amplifier. The gain or sensitivity of each filter is adjustable, and each has been set such that the response of the filter bank over the spectrum is flat for a signal applied at the input without equalization. The effects of the variations in the pass-bands of the low-pass filters used for removing the upper sidebands are thus overcome.

Filters of the same number in each bank are mounted on the same plug-in modules in order to share the readout circuits. To keep interference between them to a negligible level (approx. 60 db down) their centre tuned frequencies are separated by 200 C/S as can be seen in Fig. 3.3. The physical positioning of the modules is also such that interference between oontiguous filters in the same bank is a minimum.

After quenching the filters, a small residue of energy is left in the tuned circuits which takes the form of constant amplitude oscillations. This energy is, however, very low and the usable dynamic range of each filter is greater than 60 db.

As mentioned in Chapter 2, the minimum time T between spectral samples in a filter bank is about 2.5 milliseconds. This corresponds to maximum filter bandwidths, Δf, of approximately 800 C/S when time-weighting is used. A minimum bandwidth of approximately 25 C/S can be obtained, but neither the minimum nor the maximum bandwidths are likely to be used much in practice.

3.3.6. Synchronization and Control.

Control pulses are generated in synchronism with the weighting functions through the use of a Schmitt trigger and pulse delay circuits. These pulses operate transistor switches to discharge (reset) the capacitors holding the spectral coefficients, sample the voltage envelopes from the filters, and quench the energy in the filter circuits in that order. A waveform is derived to route the control pulses to filter banks A or B as desired. Provision is made for alternately sampling the two banks or for omitting one bank completely in the analysis by sending all control pulses to the other.

3.3.7. Sample Hold, and Parallel to Serial Converter Circuits.

Since it is necessary in practice to sample all the filters in one bank simultaneously, the samples must be stored and held before they can be presented at the analyser output in serial form. The spectral samples are stored on a bank of capacitors as a result of opening transistor gates for a period of time just before the end of the integration interval T, long enough to detect the peak voltages in the filters. The capacitors on which the voltage samples are stored are shared by both filter banks, the samples being stored for alternate time intervals.

To obtain the spectral coefficients in serial form a 32-stage shift register is used as a ring counter. A pulse is inserted at one end of the shift register just after the filters have been sampled and then moved one stage at a time down the register by means of shift pulses. At each stage the pulse opens one section of a 32-way OR gate, allowing one spectral coefficient at a time to reach the output of the analyser.

3.3.8. Displaying and Recording.

The output from the analyser takes the form of a series of 32 spectral coefficients at every interval of Δt seconds. The proper interpolation between the contiguous spectral coefficients would give the correct (as determined by the spectral window functions) spectral sections for each Δt interval. In practice, for visible presentations, only an interpolation function which is easily applied can be chosen. For this reason, the serially presented coefficients are passed through a low-pass filter having a Gaussian impulse response which gives a smooth transition between the contiguous discrete samples(Chapter 4). The result is a smoothed waveform as a function of time which can be used in visible displays or for obtaining permanent records.

In addition to the Gaussian low-pass filter, the output circuits include a logarithmic amplifier which, when used, emphasizes the low level spectral coefficients. With the proper logarithmic law the spectral sections can then be presented as log power v. frequency rather than the normal spectral intensity v. frequency.

For visible displays the spectral sections are used to modulate the intensity of the electron beam in a CRT while a raster is provided as the beam is swept slowly and linearly horizontally across the CRT. This gives a three dimensional (time-frequency-intensity) spectral display with frequency on the vertical axis and time on the horizontal axis. In addition, the spectral sections may provide some deflection of the electron beam on the horizontal axis, making the display on the CRT a series of spectral sections. However, the net effect of combined intensity modulation and horizontal deflection in this way is to give the visible spectrograms a bas-relief appearance (Campanella, 1962) and thus increase their dynamic range. The results of both types of display are shown in Chapter 4.

Permanent records of the spectrogram can be obtained on film from the CRT displays by means of a single frame camera or a continuous film camera. In cases where computer analysis of spectra might be desired, an analogue-to-digital converter could be connected to the analyser output to obtain the coefficients in their digital form. If necessary, the analyser output can also be obtained in parallel form suitable for speech synthesis studies.

SPECTRAL RESULTS

4.1. Preliminary Comments.

In this chapter are shown a variety of spectrograms obtained from the IC analyser. To demonstrate the effects of changing filter characteristics on the spectrograms three weighting functions have been chosen. In addition, filter bandwidths have been varied. A number of short sentences, which include a comparatively large group of speech sounds, have also been analysed and the patterns can be compared to Sonagrams of the same sentences.

Since results from the IC analyser and the Sonagraph are being compared the main differences in the displays should be noted. In the Sonagraph the output is recorded continuously from a filter which samples the spectrum repeatedly while being shifted in frequency by 15 c/s on each replay. The output from the IC analyser is in the form of successive spectral sections each consisting of a number of discrete samples (between which interpolation can be performed) from a bank of filters tuned contiguously at 100 c/s intervals.

The Sonagrams have grey-black markings on teledeltos paper to represent spectral intensity, while spectrograms from the IC analyser are shown on a CRT where spectral intensity can be represented by the spot brightness. Both displays have a low dynamic range. In the case of the Sonagrams, the marking range of the teledeltos paper is 6 db. In the case of the CRT display, intensity modulation is very non-linear. Further, the human eye is limited in its ability to distinguish accurately a range of intensities. To improve the display from the IC analyser, the spectral sections can also provide some horizontal modulation of the spot on the CRT (i.e. along the time axis) which, when applied along with intensity modulation, has the effect of giving a bas relief appearance to the spectrogram. The high energy portions of the spectrum 'stand out' increasing the dynamic range of the display.

For speech signals, the higher frequencies are emphasized in both analysers by means of equalization networks. From Fig. 4.1 it can be seen that the network in the IC analyser emphasizes the frequencies above 3 kc/s slightly more than that in the Sonagraph.

4.2. Filtering Characteristics.

Although many filter characteristics could have been used for the spectrograms from the IC analyser, only three have been chosen. The main reason for the choice of the particular characteristics is that their temporal and spectral windows comprise functions under which the areas are confined mainly to central positions in time or frequency. The positions of resonances or harmonics are visibly most clearly defined in spectrograms where such filter characteristics are used.

The three temporal window functions and their spectral windows are shown in Fig. 4.2 and Fig. 4.3 respectively. The

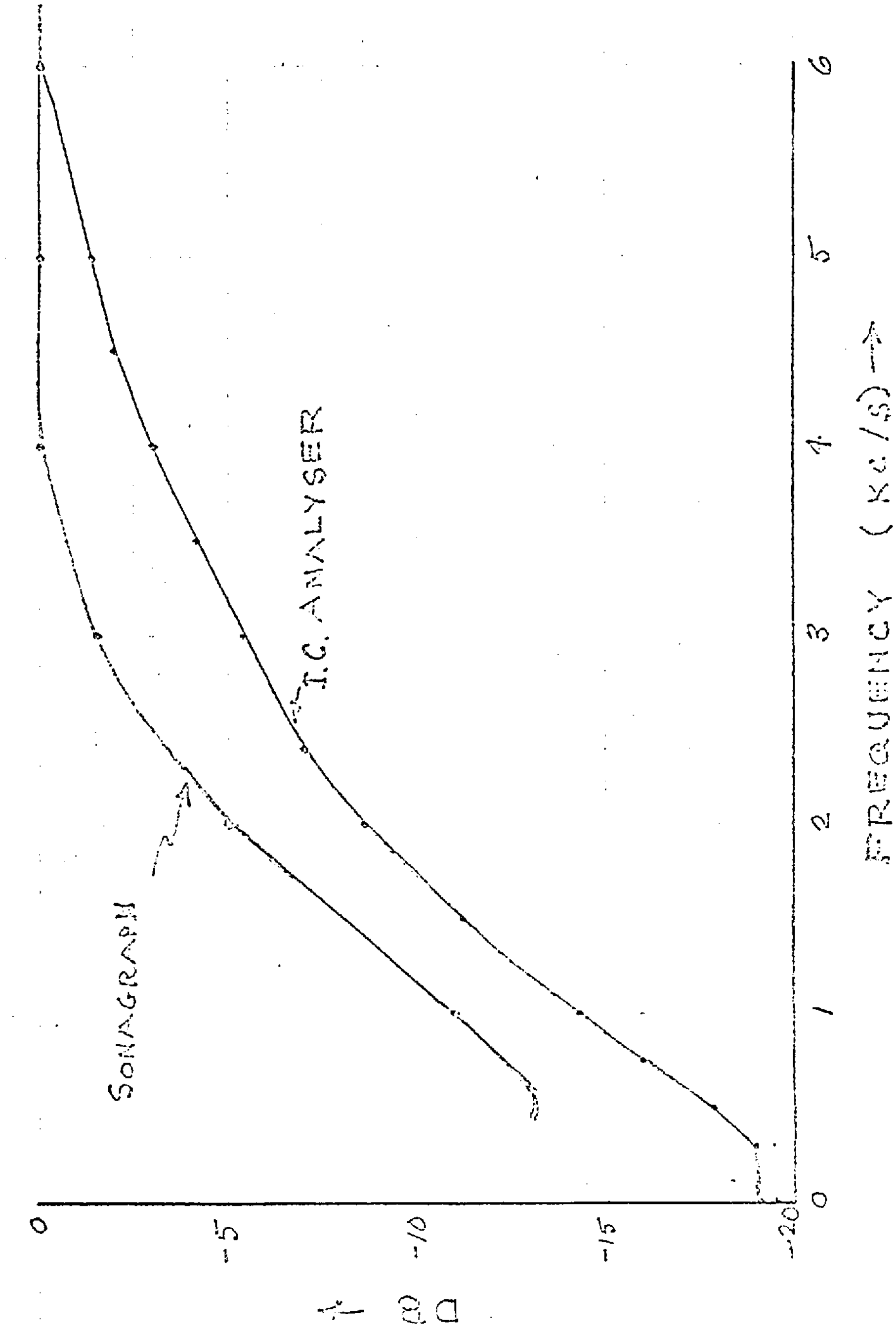


Fig. 4.1 Equalization in the I.C. analyser and the Sonagraph.

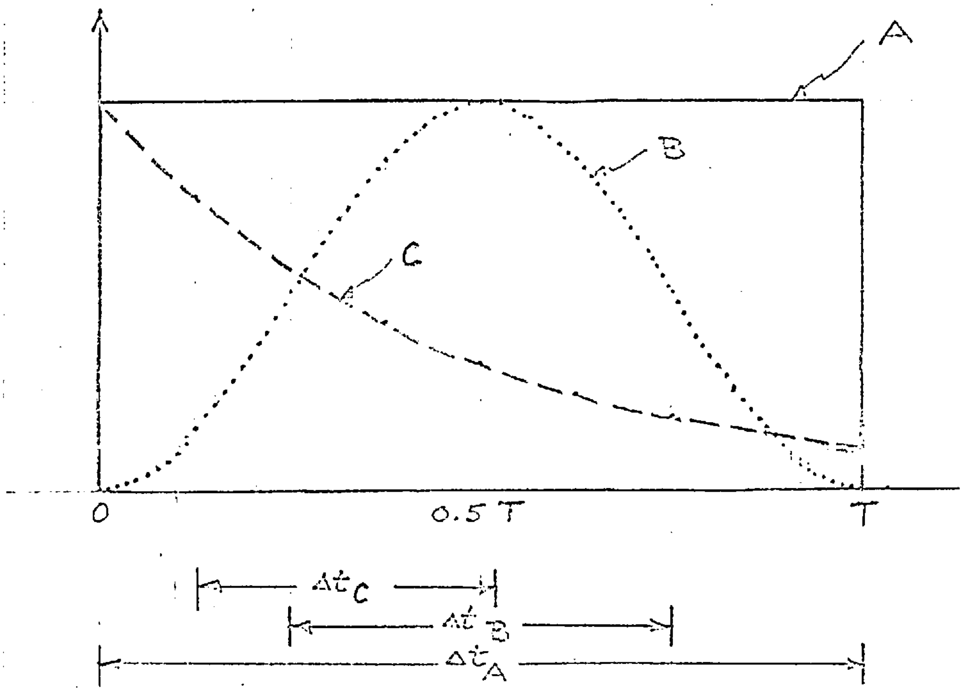


Fig. 4.2 Time-weighting functions used to obtain sample spectrograms.

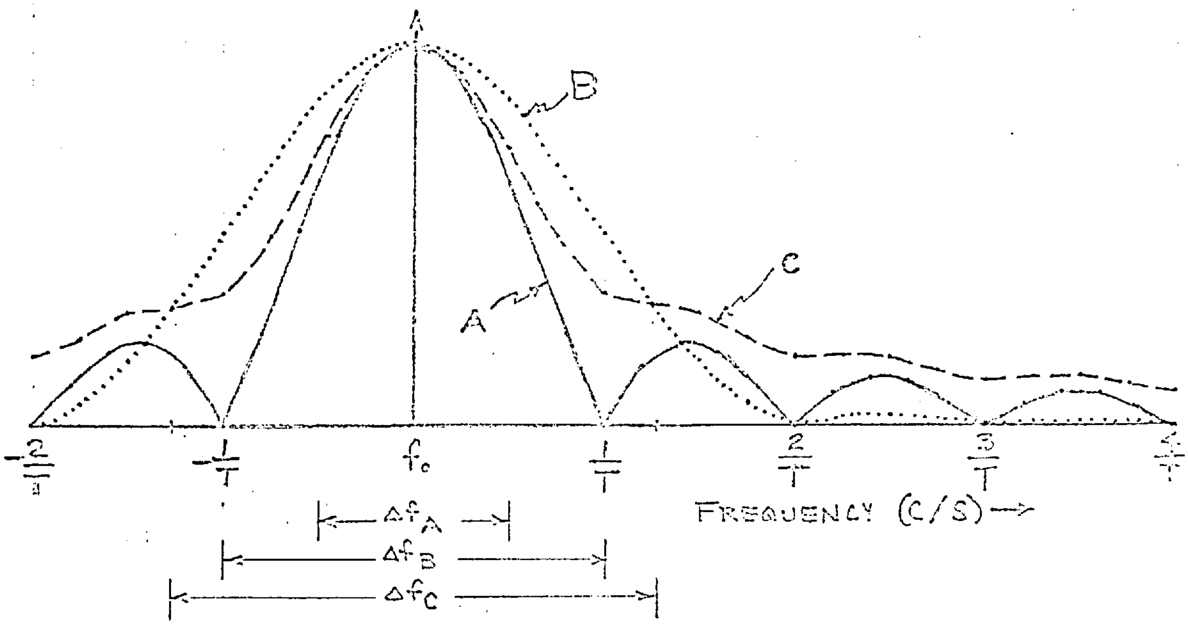


Fig. 4.3 Spectral window functions obtained with the functions shown in Fig. 4.2.

rectangular temporal window function A has the shape of the effective impulse response of the lossless filters without time-weighting. It is thus the most convenient response to use, but the side-lobe structure in its spectral window is quite pronounced.

The raised cosine function B is a reasonable approximation to a Gaussian function. The Gaussian function has been suggested by Gabor as a suitable one for the time-frequency representation of signals. In this case, the side-lobes in the spectral window are very low in amplitude and rapidly die away as the distance from f_o increases.

The third temporal window function C, a truncated exponential, is an approximation to the envelope of the impulse response of a single tuned RLC circuit. Simple tuned circuits have found application in spectral analysers (Admiraal, 1960). The frequency response curve has a relatively sharp main peak, but its amplitude dies away quite slowly as the distance from f_o increases.

The characteristics of the two filters (with bandwidths of 45 c/s and 300 c/s) in the Sonagraph are nominally roughly Gaussian in shape. Their bandwidths are defined as the approximate frequency band between the 3 db down points on their frequency response curves. This definition normally leads to a lower valve for defining bandwidths than the values of Δf used in defining bandwidths in the IC analyser. In the Sonagraph available for comparison, the wide-band filter had a spurious spike in its

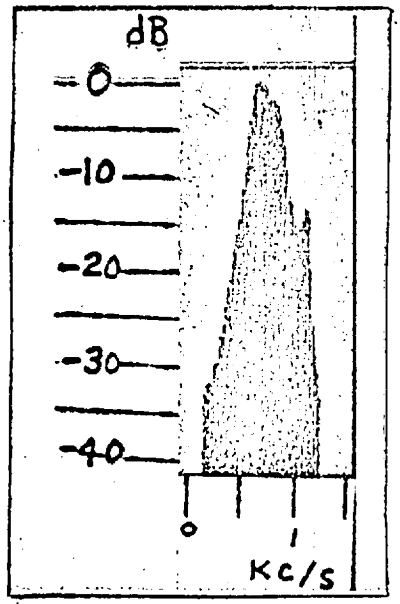


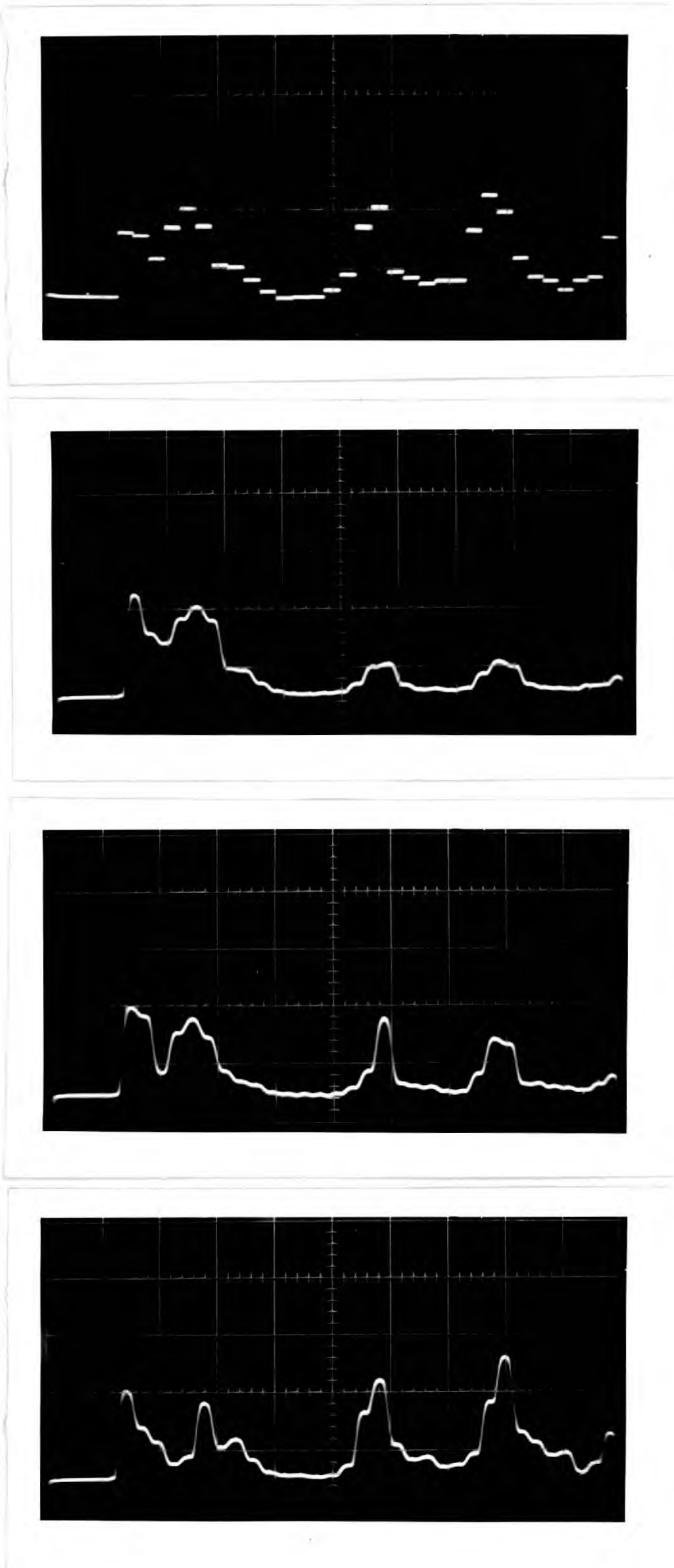
Fig. 4.4 Response of the 300c/s filter in the Sonagraph to a sinusoidal signal.

frequency response curve as shown in Fig. 4.4. Fig. 4.4 is a spectral section of a sine wave obtained from the Sonagraph using the 300 c/s filter to show up the response of the filter itself. The effect of the spike is most noticeable in Sonagrams of sinusoidal signals.

4.3. Spectral Sections from the IC Analyser.

To demonstrate the effect of equalization and the operations which can be performed on the spectral output from the IC analyser a few sections (intensity v. frequency) are shown in Fig. 4.5. They are spectral sections of the vowel sound ε, but do not represent exactly the same time instant in each case. Weighting function B is used where Δf is 100 c/s, the first filter being tuned at 100 c/s and the last at 3.2 kc/s. There are 32 spectral coefficients making up each section preceded by a longer zero reference sample.

Interpolation or smoothing between contiguous samples can be seen to overcome the quantized appearance to a great extent. Without the equalizer the higher formants in the vowel sound are very much lower in amplitude than the pitch frequency and the first formant. Thus, in spectrograms of limited dynamic range the higher formants would be difficult to observe. A logarithmic amplifier can be used to compress the range of amplitudes of the spectral coefficients of lower amplitudes with respect to those of higher amplitudes. This step improves time-frequency displays. The best results, however, for visible



(a)
With equalization.

(b)
With interpolation.

(c)
With interpolation and logarithmic amplification.

(d)
With interpolation and equalization.

Fig. 4.5 Spectral sections of the vowel 'ɛ' using weighting function B and filter bandwidths of 100 c/s.

presentations are obtained by equalization to raise the energy in all the formants to an approximately common level. For this reason, in all the speech spectrograms shown later, equalization is used.

4.4. Spectrograms with Differing Filter Characteristics and Bandwidths.

In Figures 4.6 to 4.8 a number of spectrograms from the IC analyser are shown in which the filter characteristics illustrated in Figures 4.2 and 4.3 are used. From these can be seen the effects of different filter bandwidths and frequency characteristics in a test signal and in some speech sounds. Sonagrams with markers at 3.2 kc/s to indicate the position of the upper limit of the spectrograms from the IC analyser are also included for comparison.

As the bandwidths of the filters increase the number of spectral sections, and therefore spectral coefficients, making up a spectrogram in a given time also increases since no filters are removed from operation. Although there is theoretically no more information available on the time-frequency plane, the extra samples provide better interpolation along the frequency axis. In all cases there are two displays of each spectrogram, one where horizontal modulation of the CRT spot is used in addition to the intensity modulation, and the other where only intensity modulation is used.

The details of the analysis from Figures 4.6 to 4.8 are shown in Table 4.1. Where the A characteristic is used Δt is equal to T and only one bank of filters is necessary. However, in both the other cases two filter banks are used in the analysis. In case B Δt is equal to T/2, but in case C Δt is less than T/2 and in the narrow-band examples there are .95 WG spectral coefficients making up the spectrogram in a time G and bandwidth W.

Figures 4.6 to 4.8	Weighting Function	Δt msec	Δf c/s	T msec	Spectral Sections
a,b	A	10	100	10	100/sec
c,d	"	5	200	5	200
o,p	"	3.3	300	3.3	300
e,f	B	10	100	20	100
g,h	"	5	200	10	200
i,j	C	8	125	21	95
k,l	"	4	250	10.5	190
m	Narrow - band Sonagram				
n	Wide - band Sonagram				

TABLE 4.1

The test signal shown in Fig. 4.6 is made up of the sum of three sinusoidal waveforms, two having fixed frequencies of 100 c/s and 3 kc/s while the third has its frequency modulated over the range 1,500 ∴ 400 c/s and is switched off and on at intervals of time. From the spectrograms the frequency responses of the filters can actually be inferred, particularly where the wide bandwidths are used.

In case A, the side-lobes of the sin x/x spectral window function can be seen where the filter bandwidths are set at 200 c/s and 300 c/s. They are most easily observed in Fig. 4.6.c and 4.6.o. These side-lobes are undesirable in the analysis of sinusoidal signals since, in visible displays, they appear as spurious signals of low amplitude.

The exponential weighting function (Case C) gives a spectral window having a relatively sharp main peak with a gradual fall-off on both sides. It can be seen, particularly in Fig. 4.6.k. that a large number of filters react to a single sinusoid. This is not too objectionable as long as the signals are separated in frequency by a reasonable distance.*

Case B provides spectrograms where there are no visible side-lobes and where the responses of the filters are not spread

* In the 100 c/s component of the test signal there are phase interference effects owing to the frequency translation and imperfect removal of the upper side-band. This makes the output from the filters sensitive to the phase of low frequency signals in the region of 100 c/s giving the appearance of amplitude modulation at a very low frequency. This effect is most pronounced with asymmetrical weighting functions.

over a wide frequency band. For these reasons, it appears to be the most suitable of the three characteristics for defining the positions of sinusoidal signals.

For a direct comparison the following groups should be taken together, a,e,i; b,d,j; c,g,k; and d,h,l. In the Sonagrams of Fig. 4.6.m and 4.6.n the 100 c/s component of the test signal does not appear. The base line at the bottoms of the Sonagrams is the zero reference line, and the 100 c/s component seems to be below the range of the Sonagraph. The effect of the spurious spike in the wide-band filter can be seen as a shadow at 3.7 kc/s just above the 3 kc/s component in Fig. 4.6.n.

The speech sounds in Fig. 4.7 and Fig. 4.8 are ti:, tɛ, tɔ: and li:, lɛ, lɔ: respectively. These sounds were spoken by a phonetician and recorded on magnetic tape before performing the analysis.

Of these spectrograms from the IC analyser the B filter characteristic provides the clearest displays. The successive spectral sections in this case are more consistent and less dependent upon the phase of the fundamental pitch in vowel sounds which falls under the integration interval T. In addition the filters between formants do not respond at low levels as they do in both cases A and C.

Case B shows up most clearly resonances with low energy levels falling near formants of high energies. The side-lobes in

Case A and the wide frequency responses in Case C obscure the lower energy resonances. Perhaps the clearest illustration of this is in the sound lɔ:. On Fig. 4.8.c and 4.8.f can be seen a resonance with a low energy level at about 1 kc/s just above the second formant. This is still just discernible in Fig. 4.8.g and 4.8.h where the filter bandwidths have been doubled. However, in both cases A and C this particular resonance cannot be observed. Even the Sonagraph does not show up the resonance with any clarity.

The next low energy resonance falling at approximately 2 kc/s in the sound lɔ: (also tɔ:) is far enough away from high energy formants to be discernible with all three filter characteristics. Even so, this resonance is most clearly defined in Case B. Another example where a low energy formant is shown most clearly in Case B is the third formant (about 2.4 kc/s) in the sounds tɛ and lɛ.

When the bandwidths of the filters are broadened the number of spectral sections making up a visible display is increased for a given time. This means higher resolution on the time axis making it easier to observe temporal events. (The time axis in the spectrograms can be stretched by speeding up the horizontal time base on the CRT to make successive spectral sections easier to observe). The extra resolution on the time axis is paid for by less resolution on the frequency axis making it more difficult to separate formants close in frequency.

From the wide-band Sonagrams (Fig. 4.7.n and Fig. 4.8.n) it can be seen that the positions of the larynx pulses are shown as vertical striations. This is possible because the output from the analysing filter is recorded continuously rather than being sampled at fixed time intervals as in the IC analyser. From this information the phase of the speech waveform can be inferred.

It is impossible to locate the positions of larynx pulses with the same accuracy in the visible displays from the IC analyser. However, by using a very wide bandwidth and stretching the time axis, it is possible to determine the positions of such pulses approximately. In Fig. 4.9 is shown a spectrogram of the sound tε where the filter bandwidths are 330 c/s making the time between successive spectral sections approximately 3 milliseconds. It can be seen that one larynx period covers approximately three spectral sections or about 9 milliseconds in the vowel sound ε.

In the narrow-band Sonagrams the frequency resolution is very good, and it is possible to see the harmonics of the larynx tone in the vowel sounds. Such resolution is impossible in the IC analyser owing to the low number of filters in the banks. For the same resolution on the frequency axis in the IC analyser a total of 3,200/15 or 214 filters would be required to cover a band of 3.2 kc/s. This would be impractical and would provide many more than WG spectral coefficients in a time G and frequency band W.

The 32 filter frequencies in the IC analyser are, however, sufficient to show the pitch harmonics in most speech spoken by females. An illustration of this is the phrase "Tea or coffee?" spoken by a female in Fig. 4.10. (Note that just before the word "or" there is a noise transient which was on the magnetic tape where the signal was recorded). In the word "tea," the pitch frequency is about 300 c/s and harmonics can be seen at 300 c/s intervals on the frequency axis. In male voices the pitch is normally much lower in frequency and its harmonics cannot be observed.

4.5. Spectrograms of Connected Speech.

In order to provide spectrograms of a variety of speech sounds eleven short sentences have been analysed and are shown in Figures 4.12 to 4.22. (These sentences were spoken by the author, who is Canadian). Wide-band Sonagrams are included. Narrow-band Sonagrams are not used much in practice since most of the desired information is available in the wide-band analysis. One narrow-band Sonagram of the first sentence is shown in Fig. 4.11.

Each sentence is analysed three times by the IC analyser. In one case weighting function B is used where Δt is 10 milliseconds and filter bandwidths are 100 c/s. This analysis covers the spectrum up to 3.2 kc/s. The sentences are analysed twice using no time-weighting as in Case A and with filter bandwidth of 100 c/s. However, since Δt is equal to T in this case only one bank of filters is required to cover a 3.2 kc/s frequency band, and one bank has been stacked above the other to provide a total coverage of 6.4 kc/s. This allows the inclusion of spectral data above 3.2 kc/s which is omitted in the previous spectrograms. The fourth formant in vowels and hissing sounds as in 's' usually fall above 3.2 kc/s, as does the greater part of the energy in 't' sounds.

It is not the purpose here to discuss the spectral structure of speech sounds since this subject has been quite extensively treated in the well-known book by Potter, Kopp and Green (1947) and by other researchers since then. The sentences do, however, provide a wide range of speech sounds which can be used for the comparison of results from the IC analyser to Sonagrams.

In these spectrograms as in those shown previously, the increased dynamic range obtained by using horizontal modulation of the spot on the CRT as well as intensity modulation is obvious. In such cases the display has a much greater dynamic range than that in the Sonagrams, and it could be of more use in quantitative studies.

Apart from the larynx pulses, the spectral patterns of the speech sounds appear to be shown up as well by the IC analyser as by the Sonagraph although, unlike the Sonagraph, the IC analyser operates continuously in real time, and in effect displays many fewer spectral coefficients.

FIG 4.6 TEST SIGNAL


(a)
Δf = 100 c/s
(b)
(c)
Δf = 200 c/s
(d)
A
(e)
Δf = 100 c/s
(f)
(g)
Δf = 200 c/s
(h)
B
(i)
Δf = 125 c/s
(j)
(k)
Δf = 250 c/s
(l)
C
(m)
N-B
SONAGRAM
(n)
W-B
SONAGRAM
(o)
A
Δf = 300 c/s
(p)

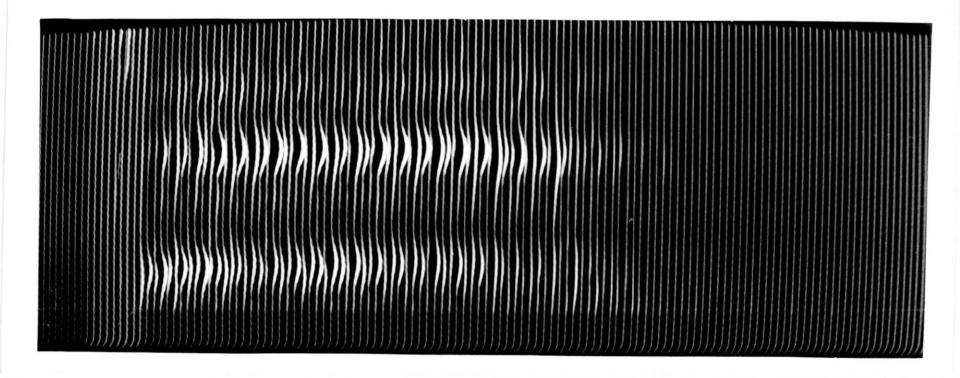
FIG 4.7 ti:, tε, tɔ:


(a)
Δf = 100 c/s
(b)
(c)
Δf = 200 c/s
(d)
A
(e)
Δf = 100 c/s
(f)
(g)
Δf = 200 c/s
(h)
B
(i)
Δf = 125 c/s
(j)
(K)
Δf = 250 c/s
(l)
C
(m)
N-B
SONAGRAM
(n)
W-B
SONAGRAM
(o)
A
Δf = 300 c/s
(P)
SECONDS →

FIG 4.8 Ii:, Iε, Iɔ:

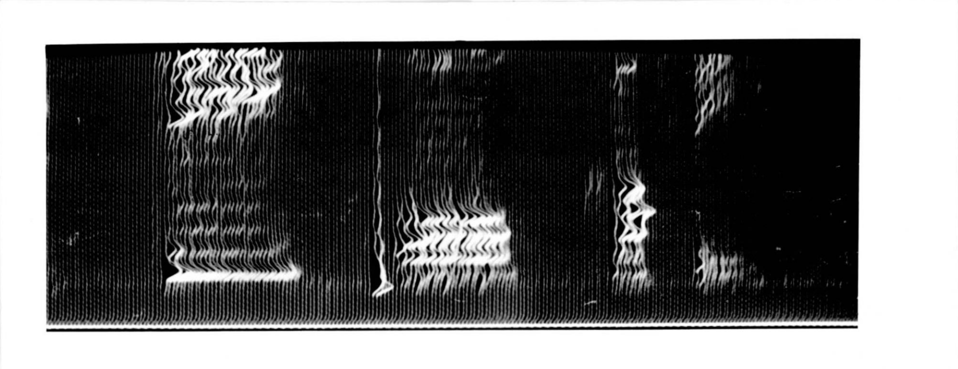

(a)
Δf = 100 c/s
(b)
(c)
Δf = 200 c/s
(d)
A
(e)
Δf = 100 c/s
(f)
(g)
Δf = 200 c/s
(h)
B
(i)
Δf = 125 c/s
(j)
(k)
Δf = 250 c/s
(l)
C
(m)
N-B
SONAGRAM
(n)
W-B
SONAGRAM
(o)
A
Δf = 300 c/s
(p)

FIGURES 4.9 TO 4.14



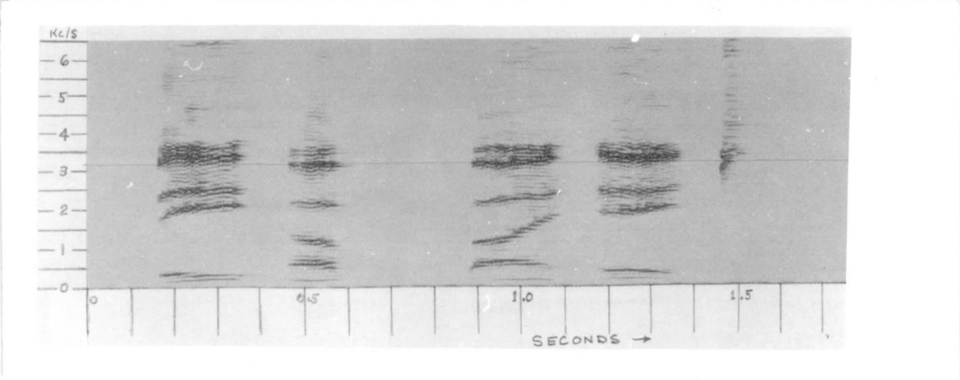
t ɛ

FIG 4.9



TEA OR COFFEE ?

FIG 4.10



BE UP BY EIGHT

FIG 4.11

B

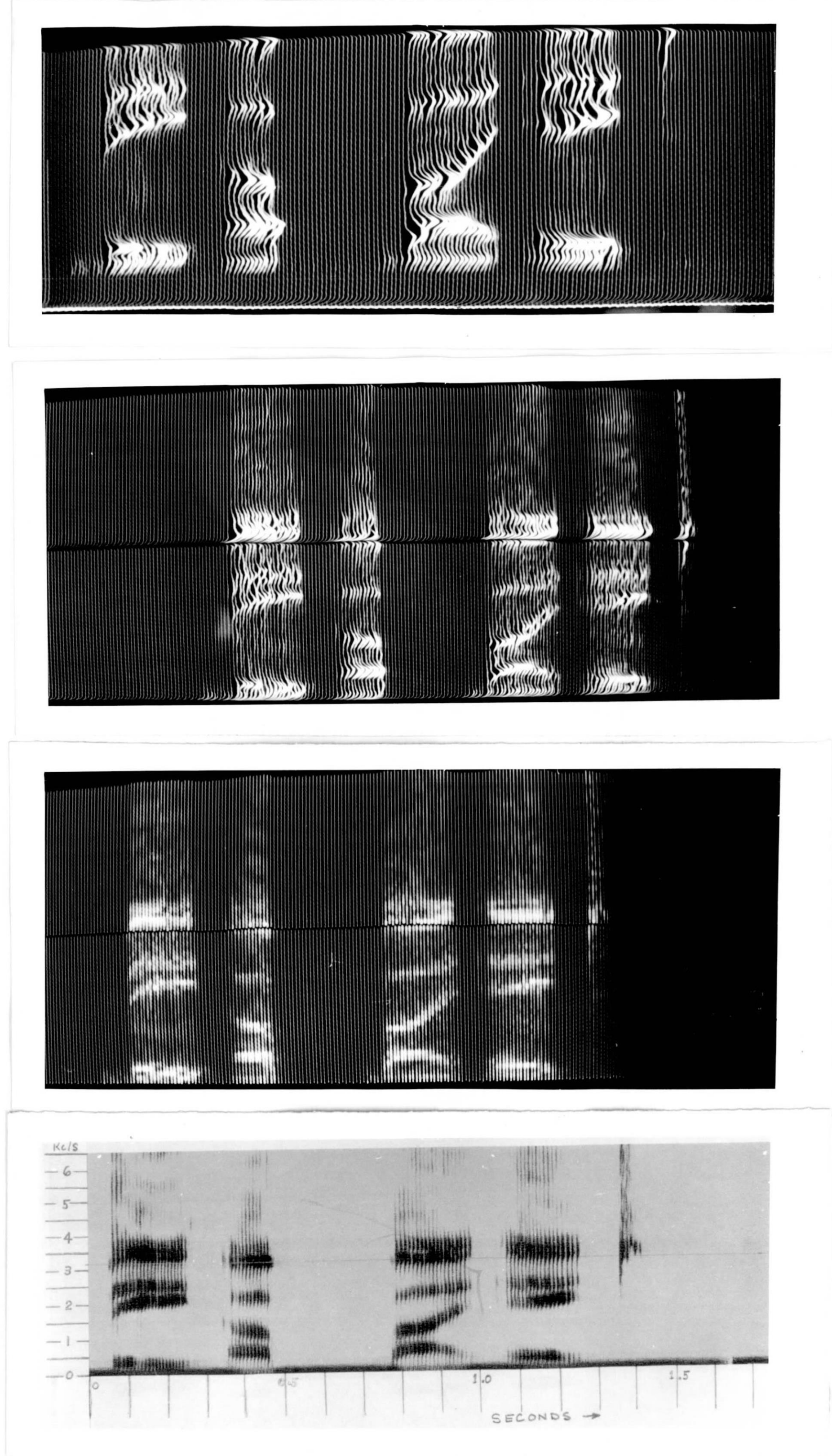
Δf = 330 c/s

A

Δf = 100 c/s

N-B

SONAGRAM



BE UP BY EIGHT

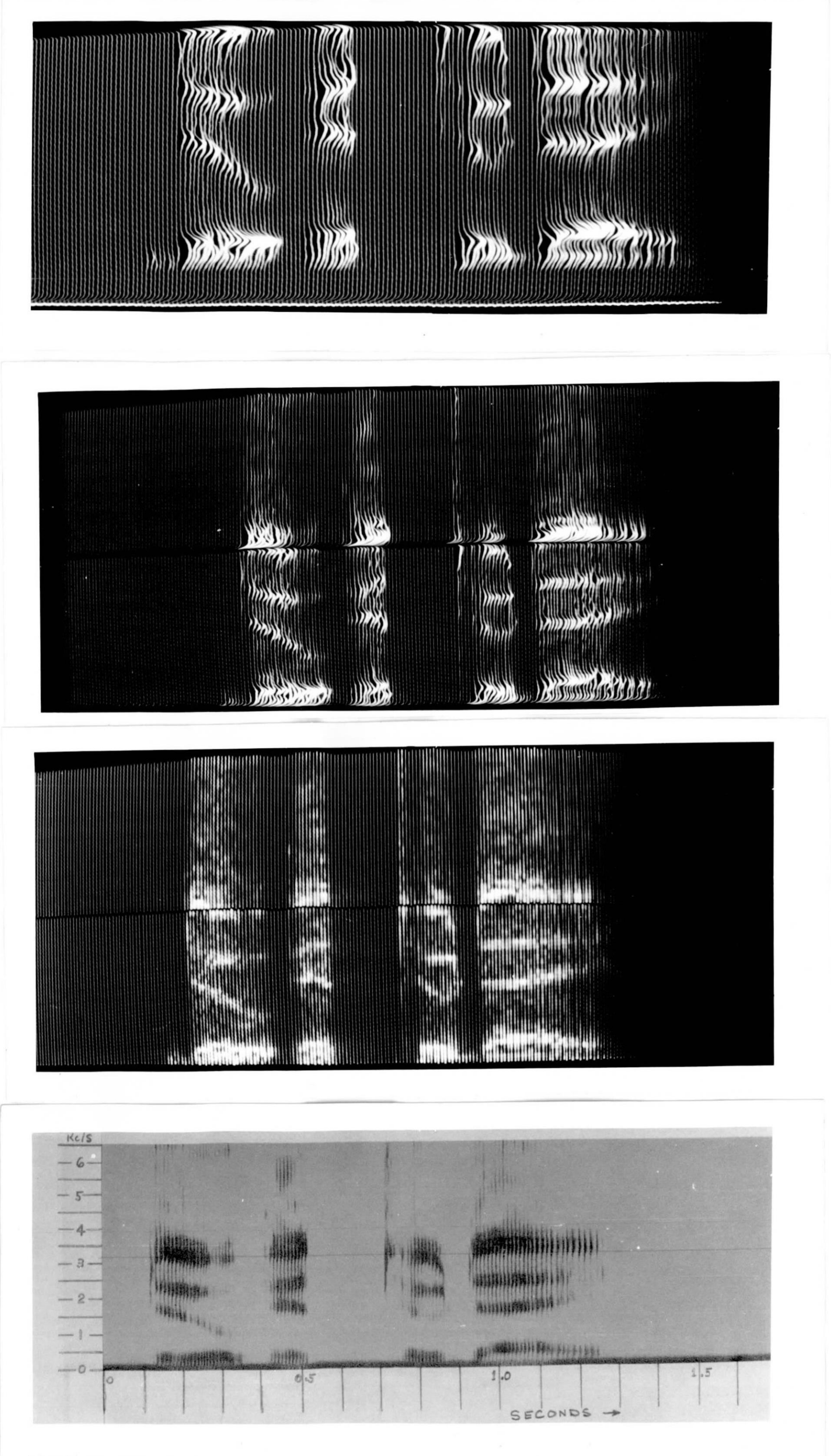
FIG 4.12

B

Δf = 100 c/s

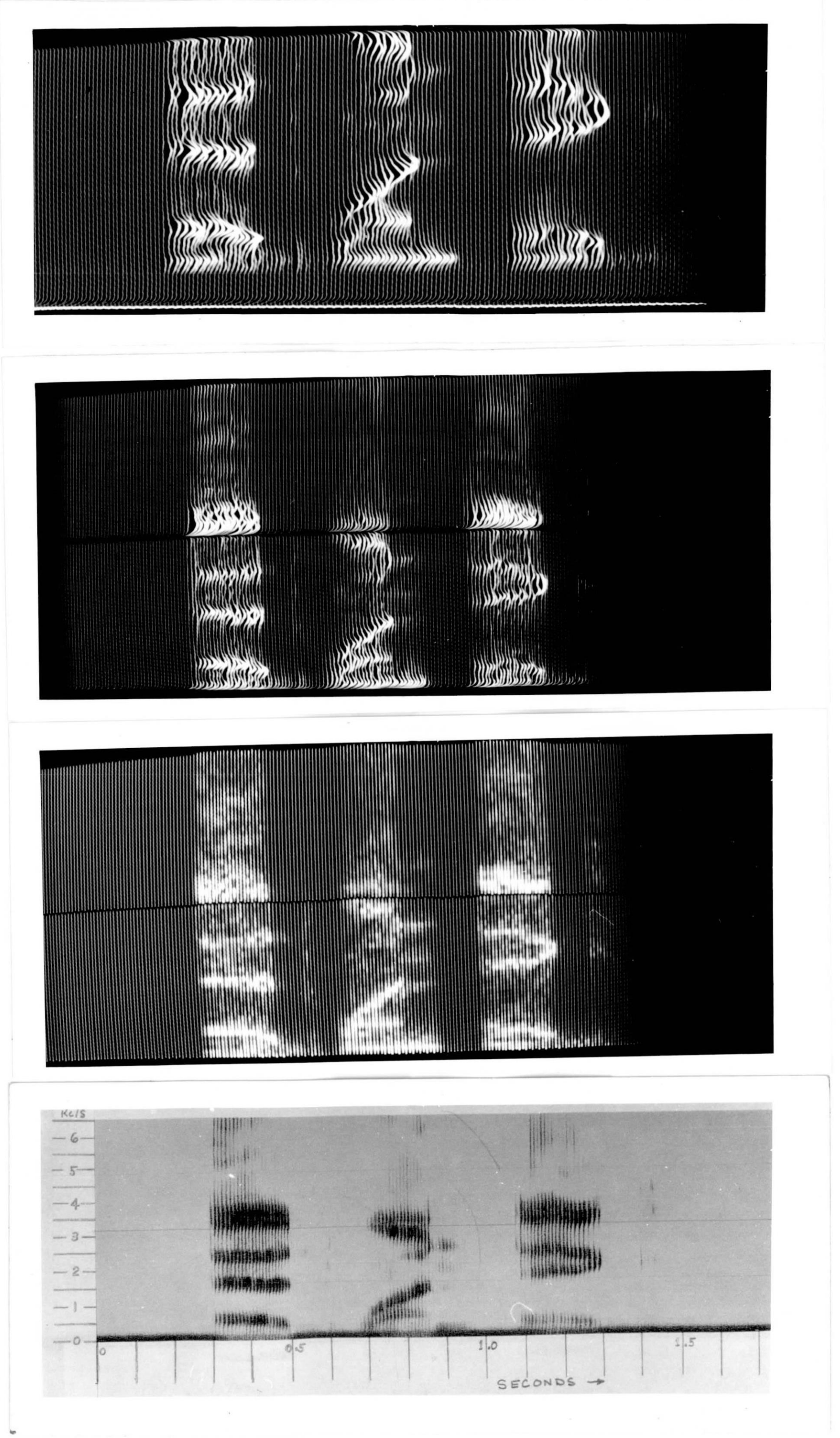
A

Δf = 100 c/s



DO IT TODAY

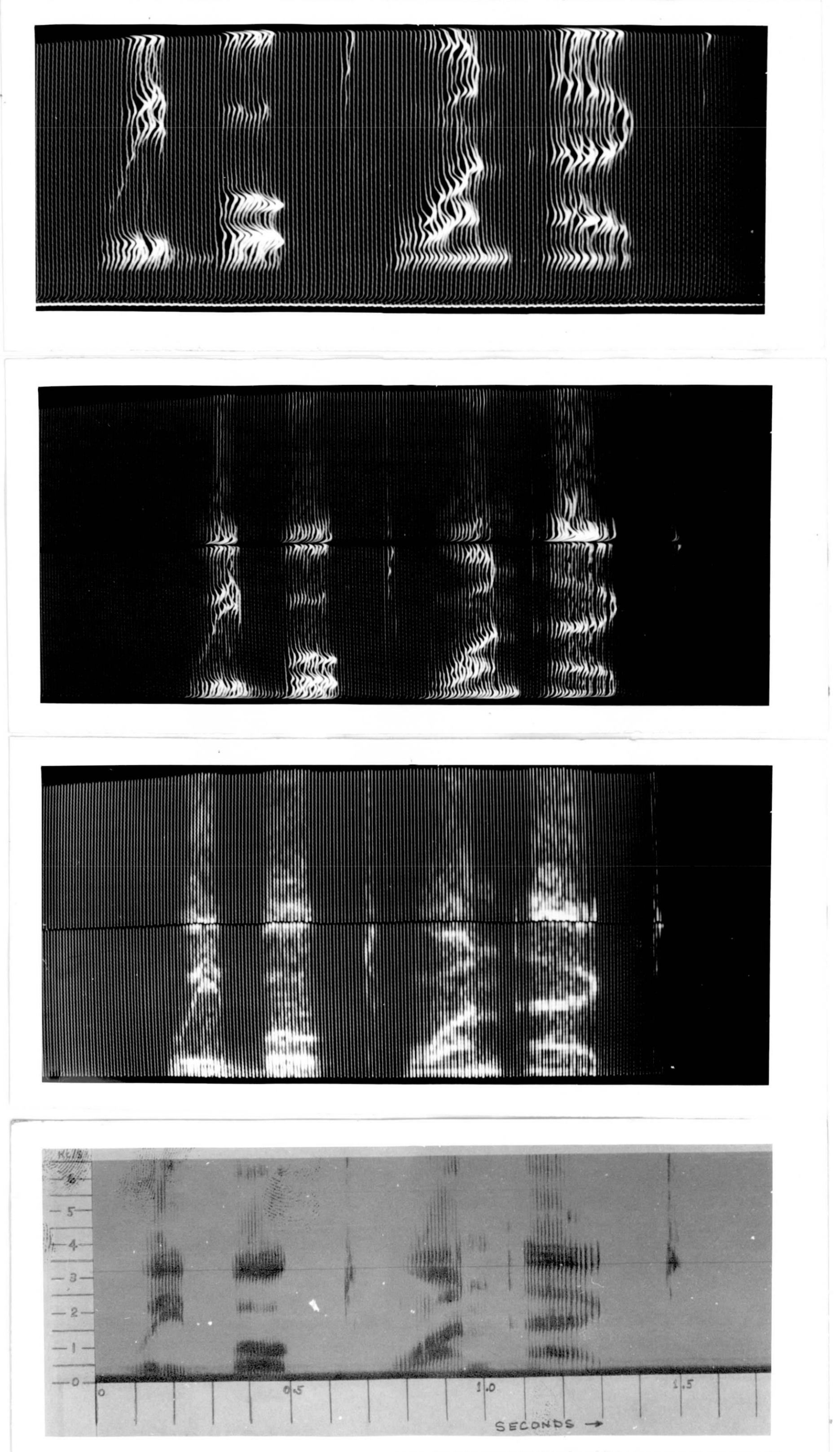
FIG 4.13



ADD ONE EGG

FIG 4.14

FIGURES 4.15 TO 4.18



WE BOOKED ONE ACT

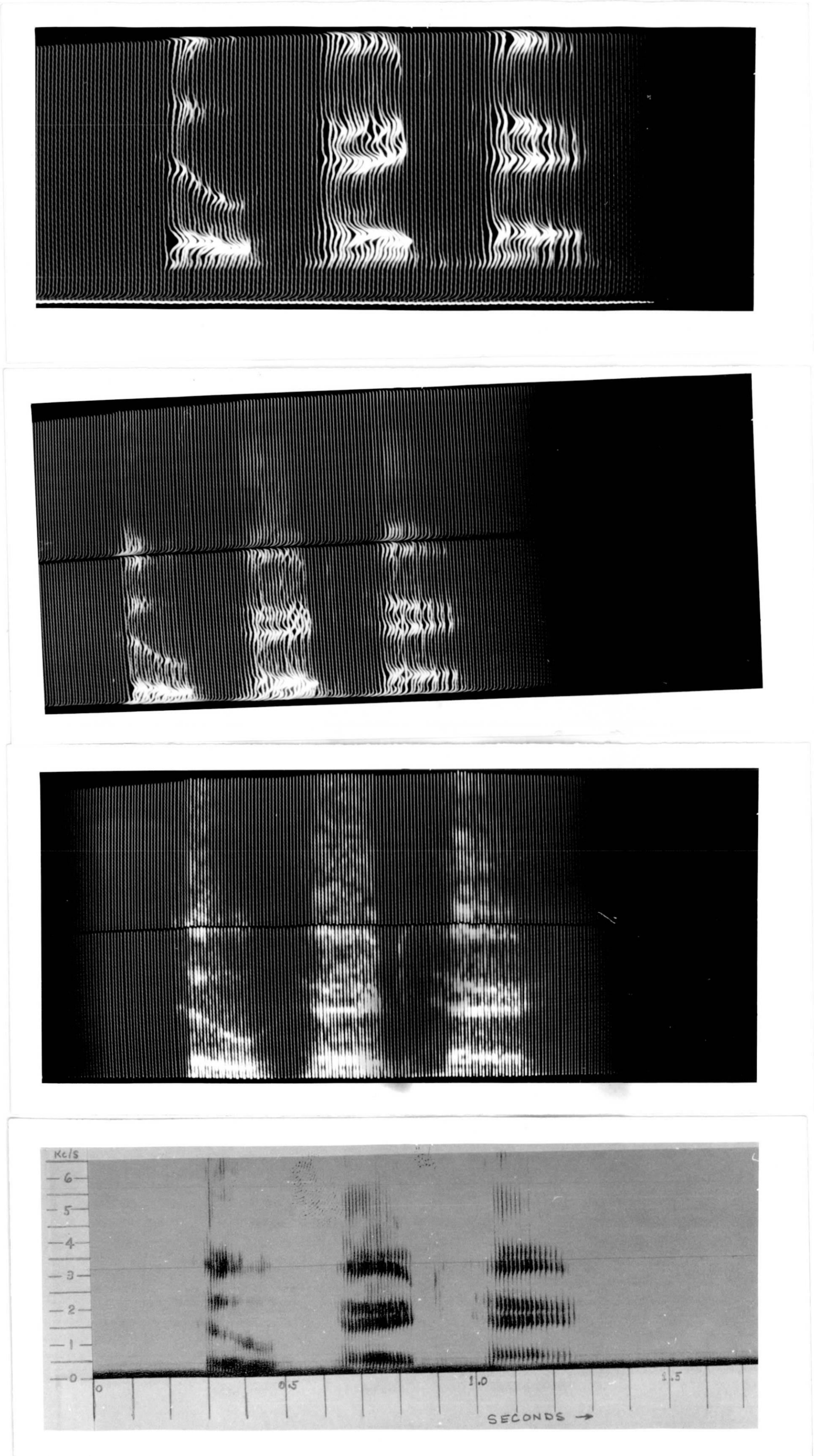
FIG 4.15

B

$\Delta f = 100$ c/s

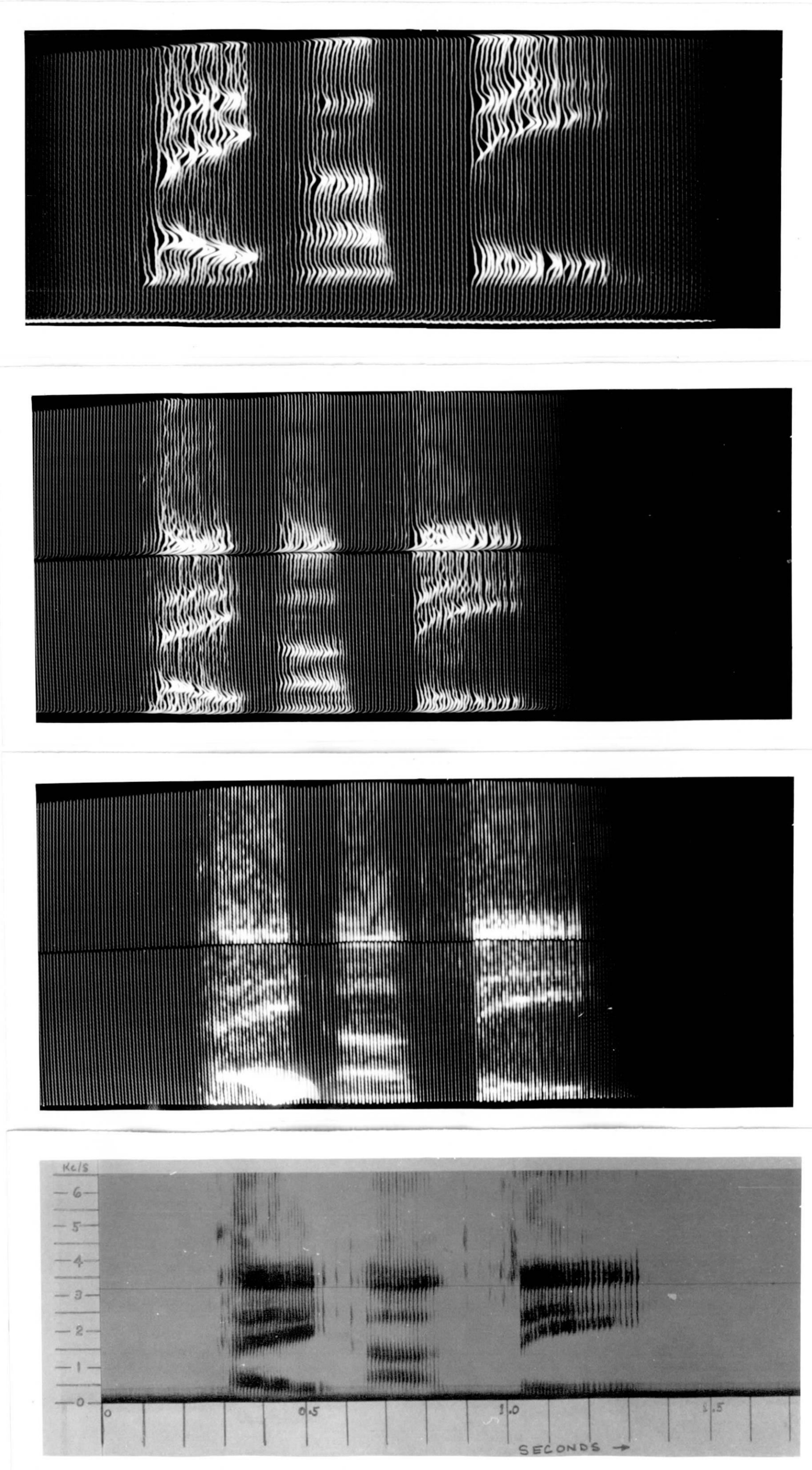
A

$\Delta f = 100$ c/s



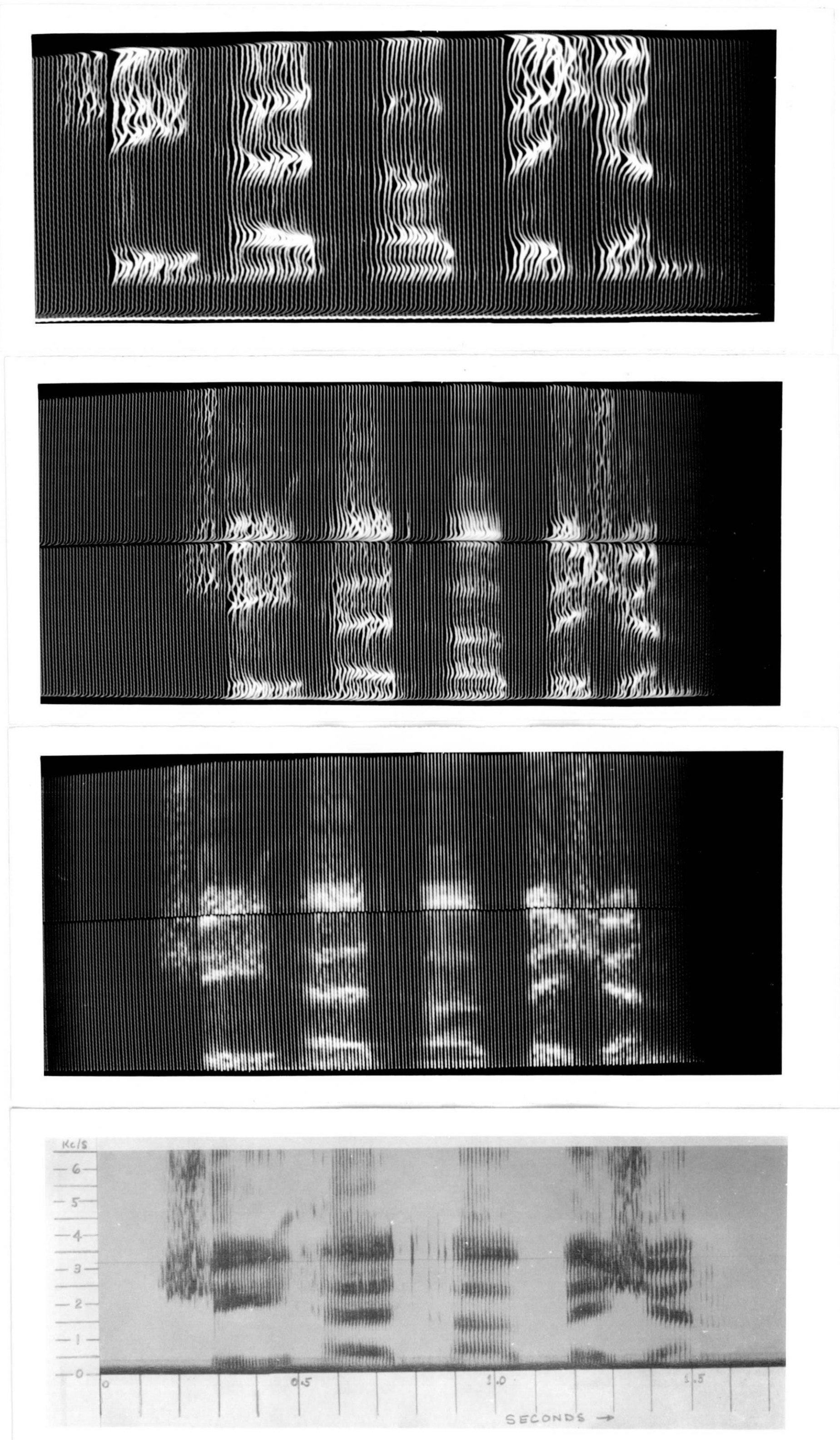
WHO HEARD HER ?

FIG 4.16



PAY A FEE

FIG 4.17



SHE HAD A VISION

FIG 4.18

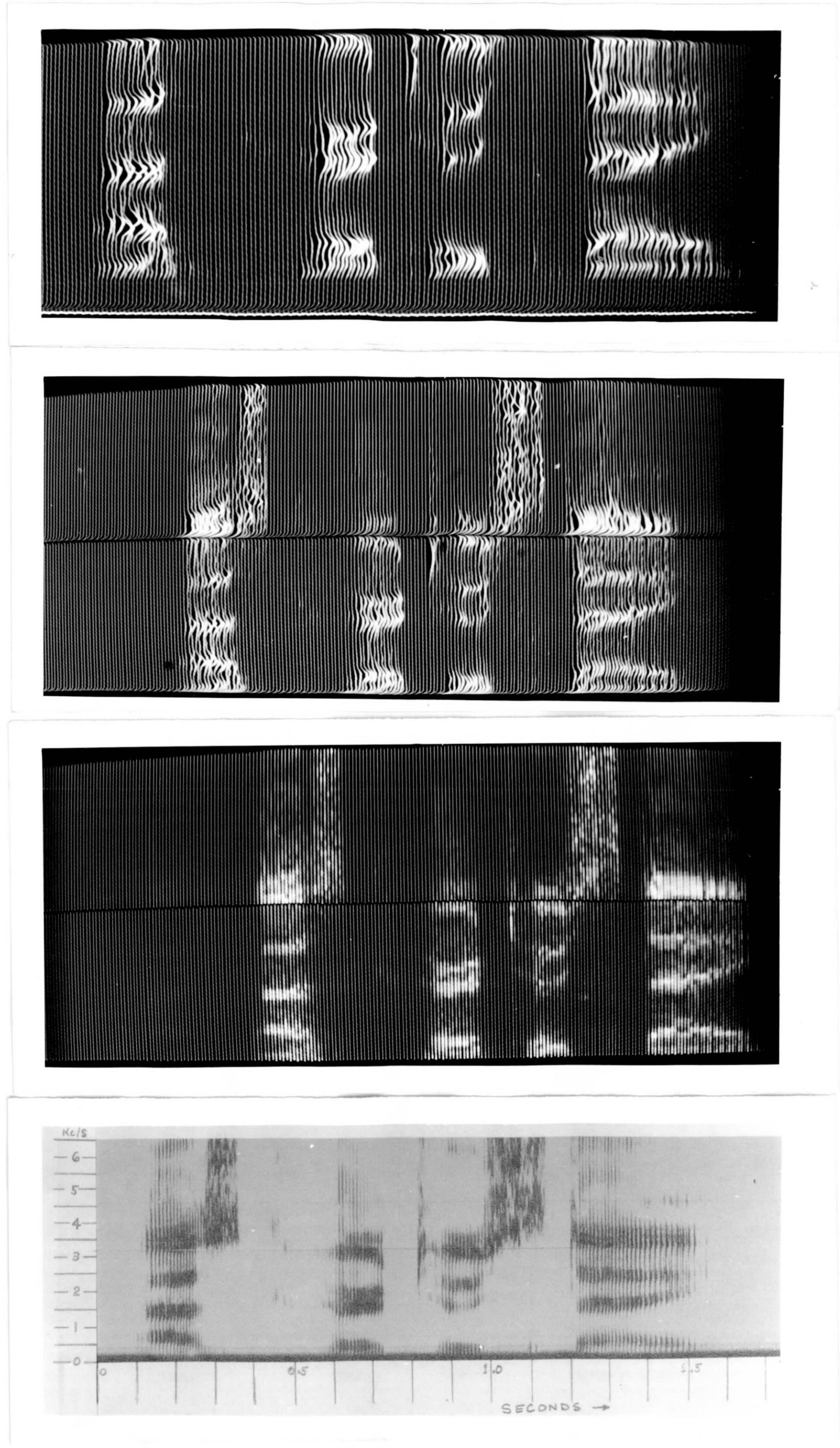
FIGURES 4.19 TO 4.22

B

$\Delta f = 100$ c/s

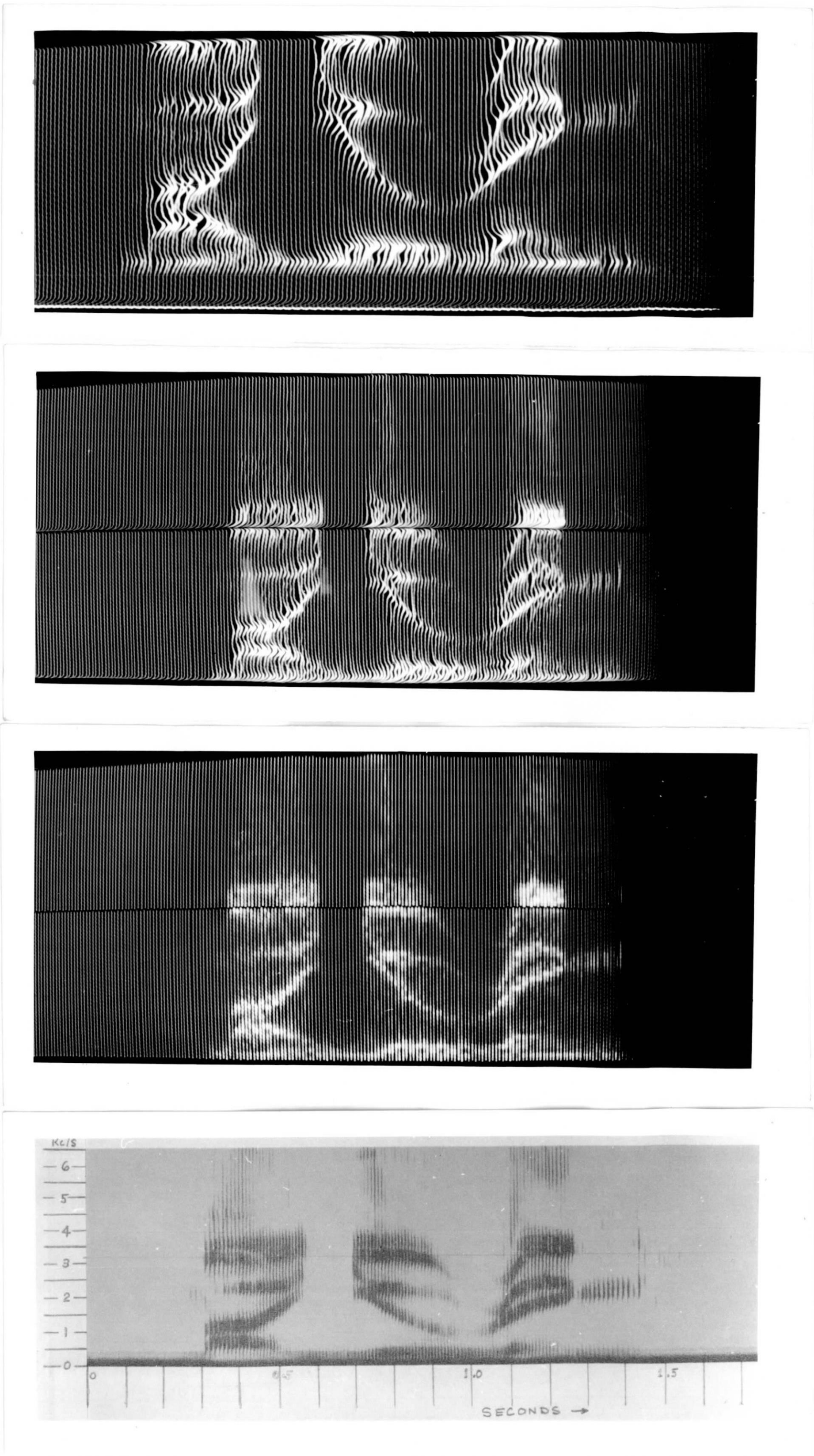
A

$\Delta f = 100$ c/s



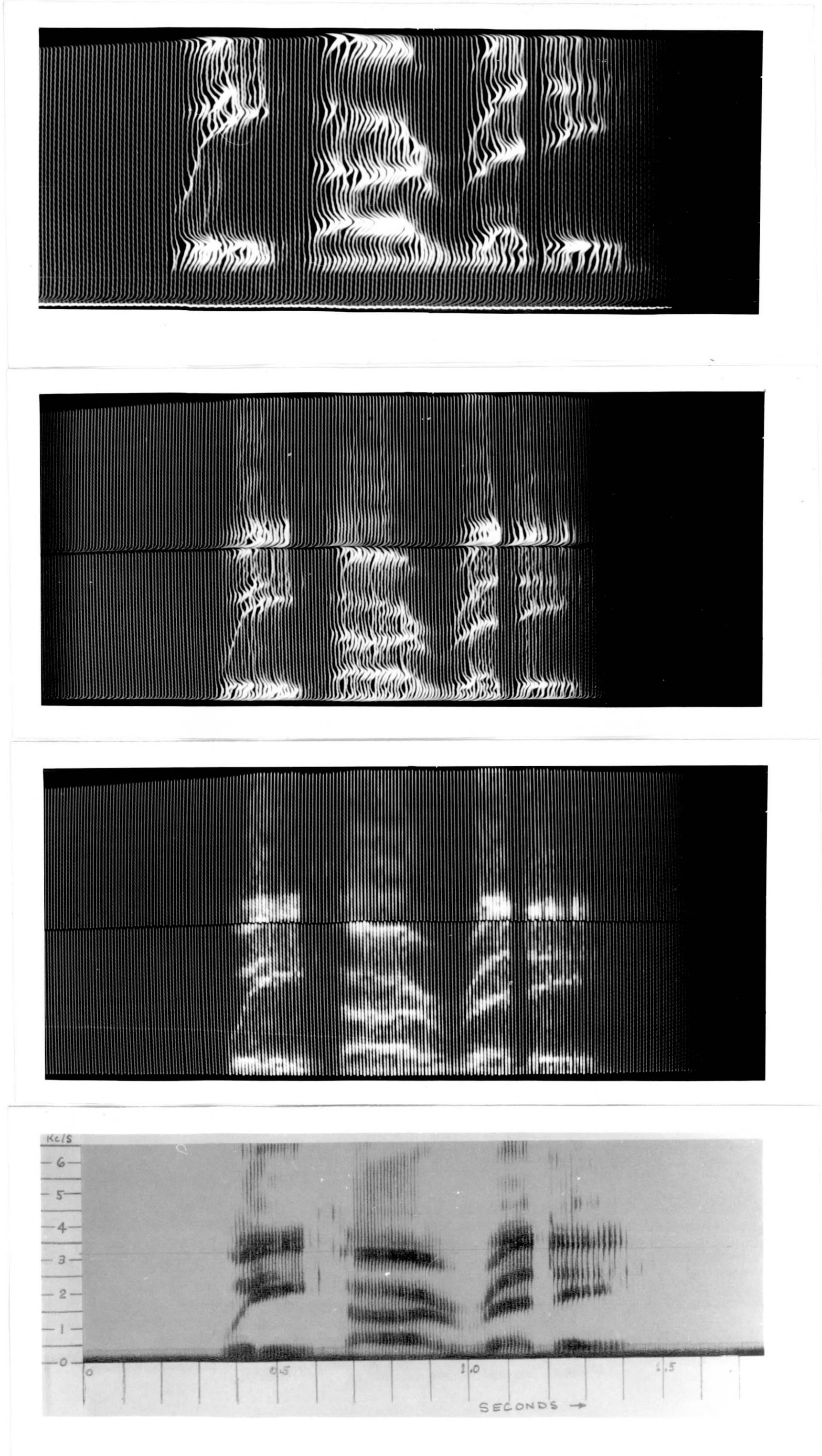
ASK HER TO STAY

FIG 4.19



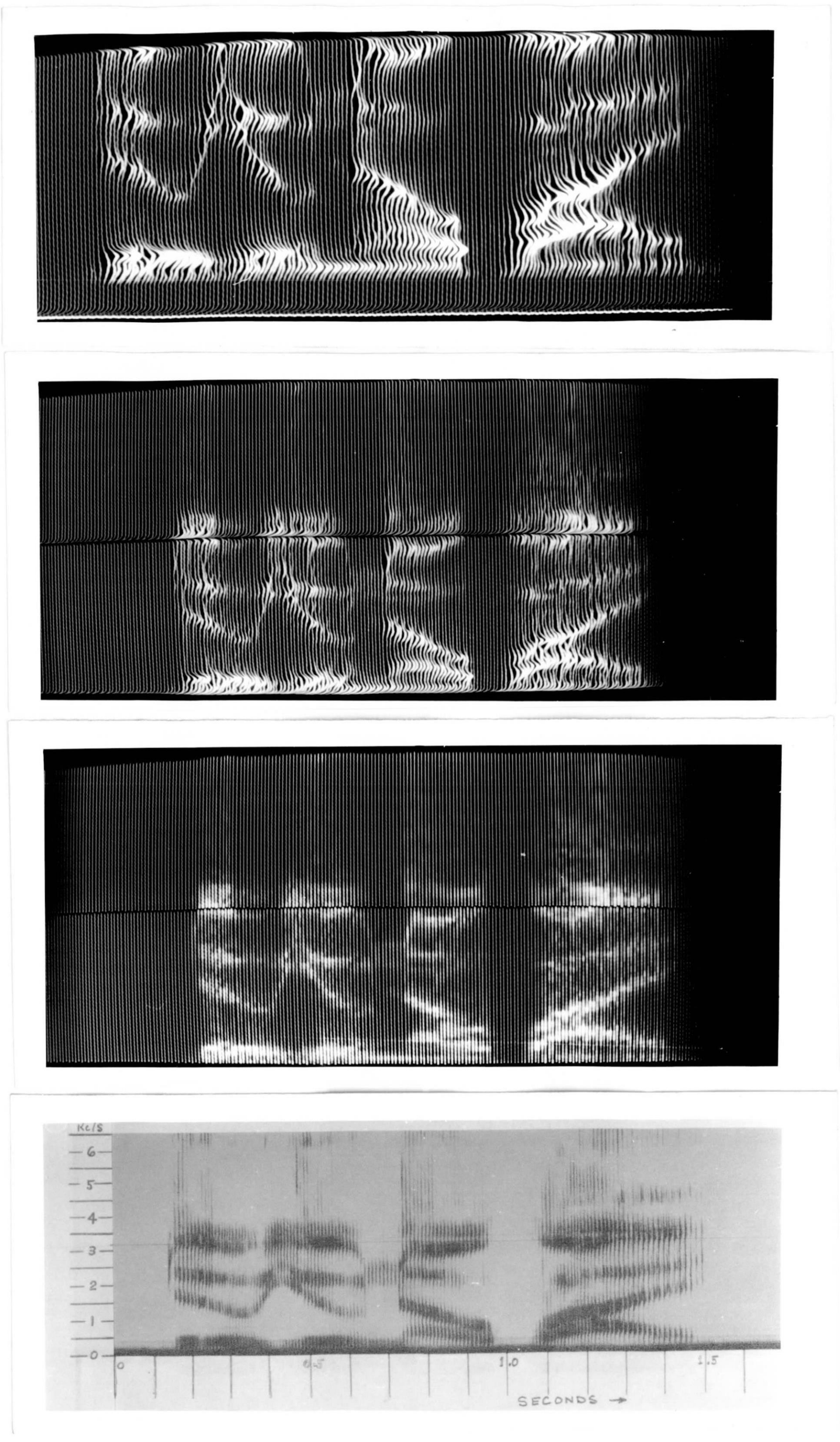
MY NEW RING

FIG 4.20



WE ARE READY

FIG 4.21



DO YOU KNOW WHY ?

FIG 4.22

AUTOMATIC ADJUSTMENT OF FILTER BANDWIDTHS

5.1. Introduction.

Not all signals are analysed in an equally desirable manner with filters having fixed characteristics. For this reason, it would be advantageous to have the filter characteristics adjusted automatically to best suit the signal being analysed. In this chapter a method is described where the bandwidths of the filters in the IC analyser may be switched between two values during the course of a single analysis. This is a very simple case of a signal automatically controlling one of the parameters of the analysing filters.

Before the filter bandwidths can be adjusted a control function must be derived from the signal. A device has been built which, when the signal is applied to it, generates a parametic function segmenting the signal such that either one or other of two bandwidths can be chosen.

5.2. Automatic Segmentation of Signals.

A block diagram of the device used for automatic segmentation is shown in Fig. 5.1. It is similar in principle to one suggested by Liljencrants (1962). The output from the segmenter is given by

$$U(t) = \frac{d\ Y_{hp}\ (t)}{dt} + \frac{d\ Y_{ap}\ (t)}{dt} \qquad \ldots(5.1)$$

where $Y_{hp}\ (t)$ and $Y_{ap}\ (t)$ are the rectified voltage outputs from

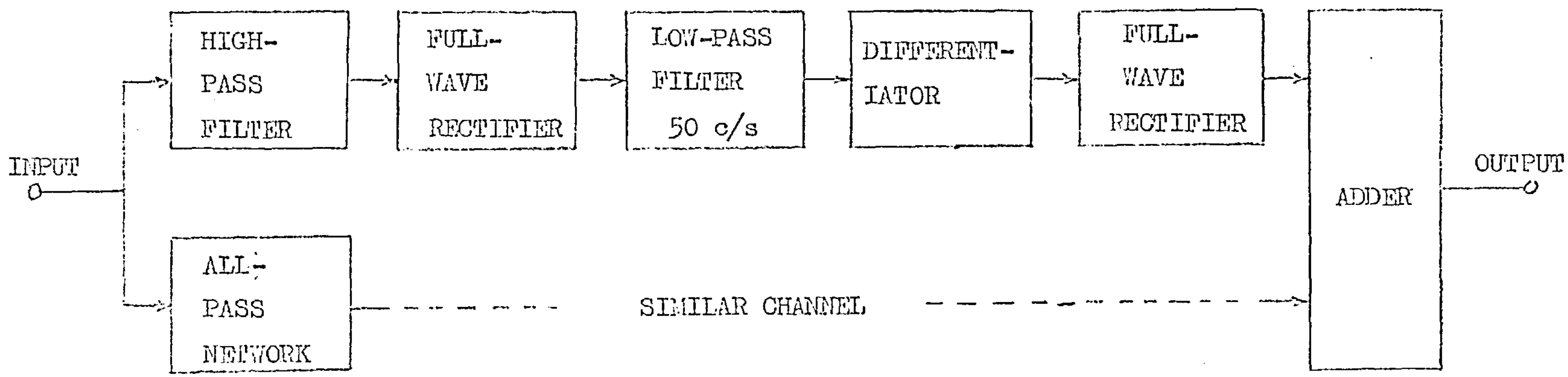


Fig. 5.1 Device for the automatic segmentation of speech.

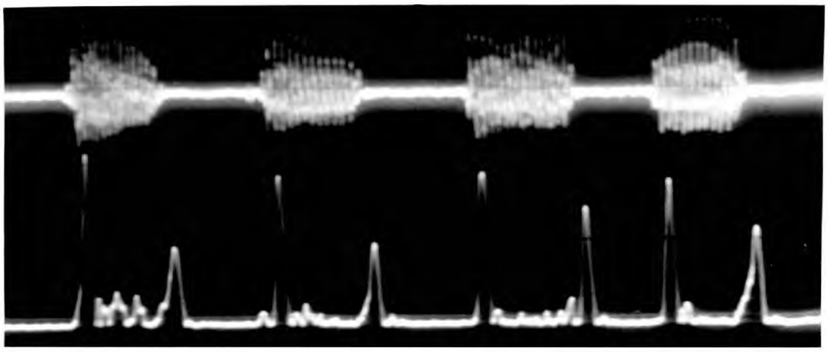
the two channels averaged over a time of approximately 15 milliseconds by means of the low-pass filters. The long integration time is used to prevent any output due to the larynx pitch in voiced speech sounds. Other channels preceded by band-pass filters can be added to such a device to give more complex modes of operation if desired, but two are sufficient in this case.

Effectively, the output from the segmenter becomes a maximum where the structure of the incoming signal changes most rapidly (for example, changes in the envelope). The two channels allow the contributions from the high or low frequency parts of the spectrum to be adjusted as desired. Where the signal structure is changing rapidly, wide-band filtering in spectral analysis can show up the temporal structure better than can narrow-band filtering.

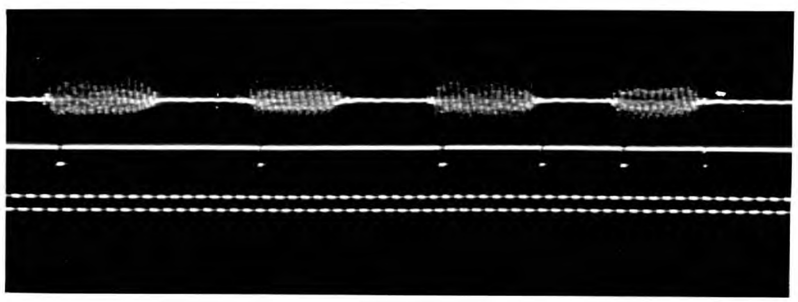
Two photographs of the output from the segmenter are shown in Fig. 5.2 for a sample of speech into it. In Fig. 5.2.a the direct output is shown below the speech waveform, and in Fig. 5.2.b the output is shown quantized into two levels between the speech waveform and a 50 c/s reference square wave. A Schmitt trigger is used for quantization. It can be seen that there is a delay (about 15 milliseconds) in the segmenter output with respect to the envelope changes in the speech waveform, which is caused by the averaging in the low-pass filters.

5.3. Spectral Analysis with two Alternative Filter Bandwidths.

The procedure followed for the spectral analysis is



(a)



(b)

PE-TER PI-PER

Fig. 5.2

illustrated by the block diagram in Fig. 5.3. In order that the analyser filters can take up the desired bandwidths in anticipation of the signal, some delay (24 milliseconds in this case) is introduced. If a sinusoidal signal of constant amplitude were fed into one of the filters tuned at the same frequency, the output amplitude of the filter would be inversely proportional to its bandwidth. For example, to increase Δf the integration time T is decreased proportionately, which leaves less time for integration to build up the voltage amplitude in the filter. To compensate for this, the amplitude of the signal entering the analyser is changed when the filter bandwidths are switched.

To illustrate the results from such a system, two sample signals are analysed in Figures 5.4 and 5.5. In Fig. 5.4 is shown the analysis of a periodic signal having a second harmonic and which is switched on and off at intervals of time. The normal filter bandwidths are 100 c/s, but for short intervals of time bracketing the switching points of the signal, the filter bandwidths are set to 200 c/s. When the bandwidths are doubled twice as many spectral sections per unit time are given. It is then possible from looking at the spectrogram to specify the positions in time at which the signal is switched with twice the certainty that would be possible if bandwidths of 100 c/s were used throughout. On the other hand, the <u>frequency</u> of the signal can be specified with twice the certainty using 100 c/s bandwidths. Since no time-weighting has been used in the analysis, it is

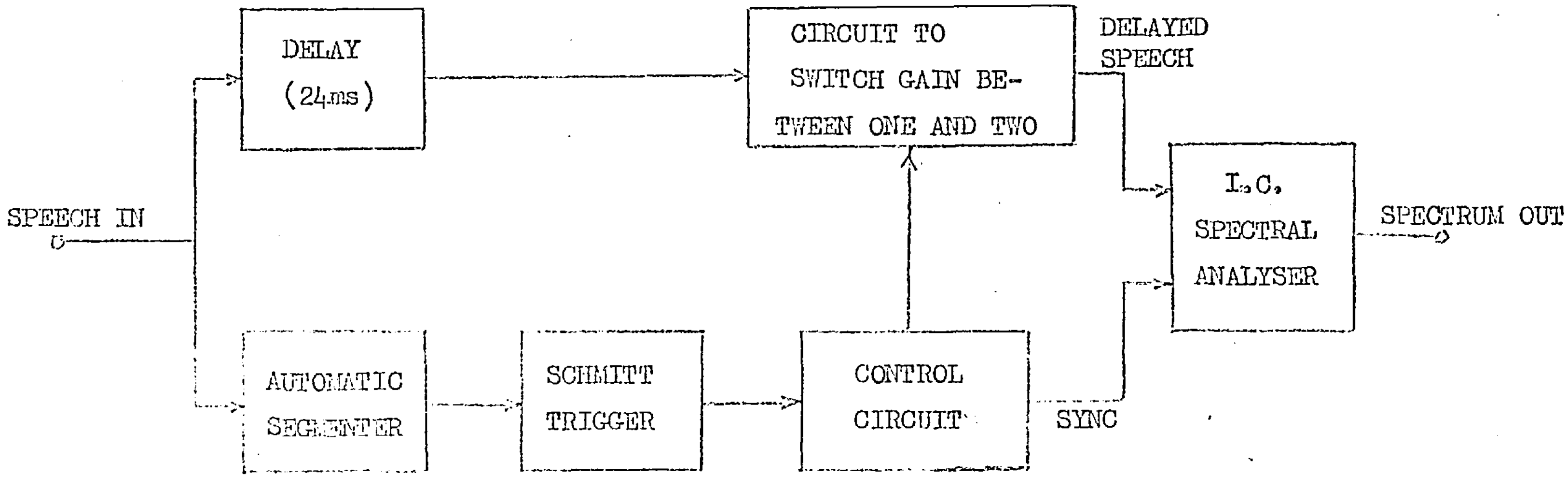
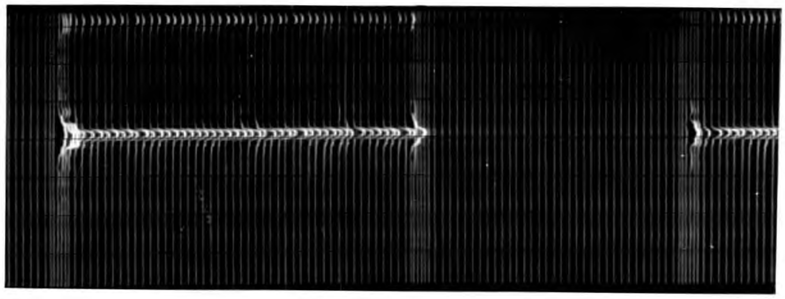


Fig. 5.3 System used to provide analysis with automatic switching between two alternative filter bandwidths.



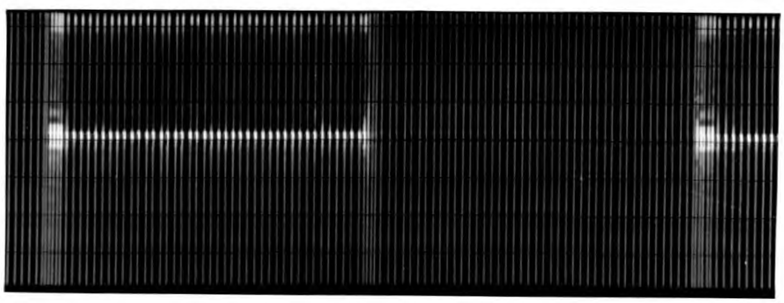
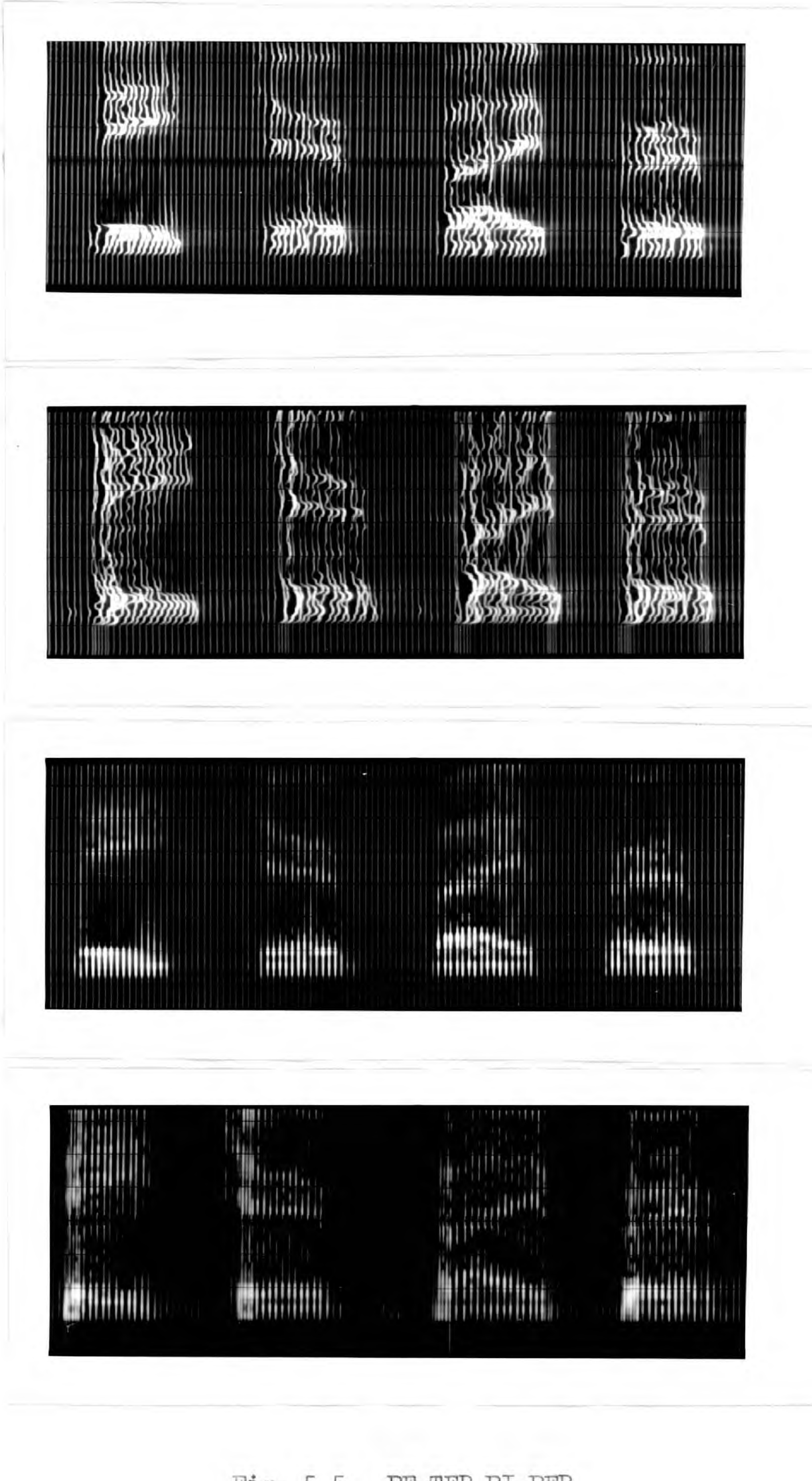


Fig. 5.4



(a)

(b)

(c)

(d)

Fig. 5.5. PE-TER PI-PER

possible to see the side lobes in the frequency response where the bandwidths are 200 c/s.

A short sample of speech is analysed in Fig. 5.5, first with fixed filter bandwidths of 100 c/s (Fig. 5.5.a and c) and second with the two bandwidths of 100 c/s and 200 c/s (Fig.5.5.b and d). Bandwidths are doubled from 100 c/s to 200 c/s during times of rapid change in the speech envelope. As in the above example, temporal events are emphasized more where the bandwidths are wide and the frequency structure is emphasized more where bandwidths are narrow.

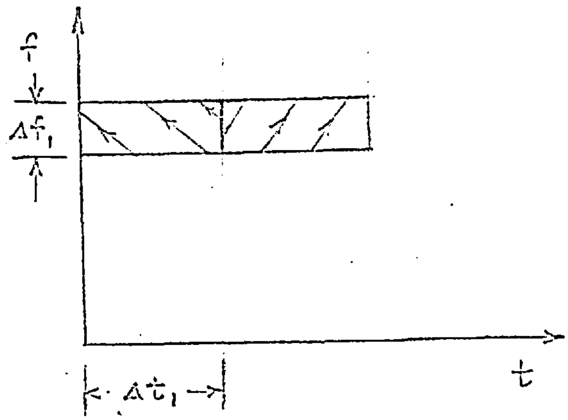
5.4. Possible Uses for Automatically Adjusted Filter Bandwidths.

The examples shown above are sufficient only to illustrate very simple switching between two filter bandwidths and do not in themselves show much justification for such a method of analysis. However, there are two possible advantages to automatically varying filter bandwidths. One is the possibility of minimizing the amount of data on the time-frequency plane, i.e. minimizing the number of spectral coefficients specifying the signal which have a significant amplitude. (This should be possible with signals such as speech where some of the spectral coefficients on the time-frequency plane have negligible values).[2,3] This could be of use for bandwidth compression where a signal was encoded and transmitted in terms of its spectral coefficients. The other possibility is that of emphasizing temporal events at one time and the frequency structure at others in visible presentations of spectrograms. It is likely that for signals like speech the most suitable spectrograms can be obtained by varying filter bandwidths continuously, rather than by discrete switching.

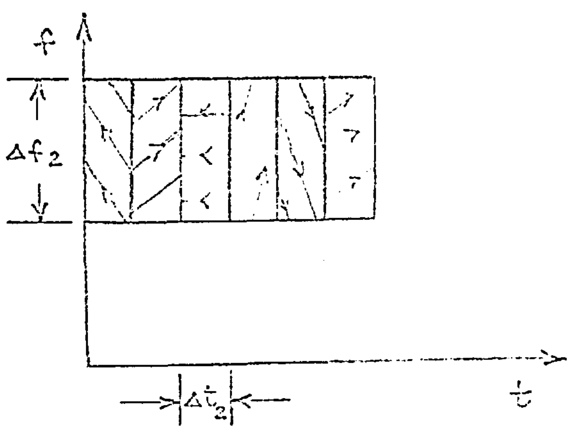
Consider the time-frequency representations of a sinusoidal signal in Fig. 5.6. In one case the aspect ratio $\Delta f/\Delta t$ is much less than (Fig.5.6.a) that in the other. Only two complex coefficients are enough to specify the signal over the same period of time in which six coefficients are needed with the larger aspect ratio. Not only is the data on the time-frequency plane less, but the representation more closely approximates our intuitive concept of a sinusoid, since it is well known that a sine wave of infinite time duration can be represented as a spectral line. For transients having a time duration in the order of Δt_2 the larger aspect ratio would give less data on the time-frequency plane.

In more complex signals like speech, it could be determined whether any reduction in spectral data occurs when the analysis is performed with the system of Fig. 5.3 by counting the coefficients having significant amplitudes. This has not yet been done.

Where the IC analyser is considered for minimization of spectral data, there is some limitation owing to the fact that the tuned frequencies of individual filters are fixed with respect to each other. To keep the $\Delta f \, \Delta t$ areas from overlapping



(a)



(b)

Fig. 5.6 Time-frequency representation of a sinusoidal signal. Line shading indicates modulus and line direction indicates phase.

it would be necessary to switch bandwidths in discrete steps and remove some filters from operation as bandwidths increase. It is possible to achieve this with the IC analyser in its present form.

In analyses for visible presentation, it is unimportant if the amount of data on the time-frequency plane is more than theoretically necessary. It is then acceptable to change filter bandwidths continuously if desired and without removing any filters from operation.

CONCLUSIONS AND SUGGESTIONS

6.1. Conclusions.

A new type of spectral analyser with arbitrary filtering characteristics has been described and demonstrated. It has been shown that weighting the input to a bank of filters by a periodic function can have the same effective result as changing the filter characteristics providing that the outputs from the filters are observed at prescribed instants in time (Chapter 2). If such a function is used in conjunction with a lossless quenched resonator, the envelope of the effective impulse response of the resulting sampling filter is the same shape as the weighting function over one period. It is only necessary to multiply a periodic weighting function by the input to a bank of such resonators to obtain desired filter bandwidths or frequency responses.

A test signal and some speech samples have been analysed (Chapter 4) using three different weighting functions in order to demonstrate some of the differences which can occur in spectrograms. For visible time-frequency representations of signals the most suitable temporal and spectral window functions are those in which the envelopes comprise a main central lobe confined chiefly to a central position in time or frequency, and having very low or non-existent side lobes. Of the three weighting functions used in the illustrations, the raised cosine most suitably fulfills these conditions and as such has been chosen as the one giving the best filtering characteristics for spectral analysis of speech.

The speech spectrograms obtained with the IC analyser have been compared with results from the Sonagraph. It was found that temporal events are not shown up as clearly by the spectrograms from the IC analyser as in the wide-band Sonagrams. This is due to the fact that the outputs from the bank of filters cannot be observed continuously as is done for the filter output in the Sonagraph. Other than this, however, the spectral patterns are shown up as clearly by the IC analyser operating, unlike the Sonagraph, in real-time. The use of successive spectral sections rather than a continuous output also makes possible a 'bas relief' type of display which increases the dynamic range of the spectrograms.

In Chapter 5 it has been suggested that analysis with varying filter bandwidths may prove useful, either in visible displays or in reducing the amount of data on the time-frequency plane. A simple example of the adaptability of the IC analyser in this respect has been shown with suggestions for further experiment. The number of occupied ($\Delta f \; \Delta t$) areas can be counted to determine which form of analysis provides the least data for the same resolution on the time-frequency plane. The filter bandwidths can also be made continuously variable and the subjective effects on spectrograms assessed. A form of automatic speech segmenter which can be used for widening filter bandwidths at times where the structure of a signal changes rapidly has been suggested.

There are many applications in which the versatility of the filtering characteristics in the IC analyser could prove useful. Two examples of such applications will be mentioned here.

Recently Harris and Waite (1963) carried out a study using a digital computer to calculate spectral sections of vowel sounds. The purpose was to compare the results that would be obtained from analysers of the filter bank type having various filter characteristics. The IC analyser operating in real-time could have carried out the actual tests.

Another situation in which the arbitrary filtering characteristics can be useful is where it is desired to obtain a spectral section by averaging over a time longer than the filter memory or integration period. In some speech sounds (such as vowels) the spectral structure remains effectively constant over a relatively long period of time and <u>average</u> spectral patterns will define them most clearly. Fluctuations between successive short term spectral sections can be averaged out in a way similar to that shown in Fig. 6.1. In this case an integrator was added to the output of a Sonagraph to produce the spectral sections in Fig. 6.1.b and 6.1.d. which are from the same signals as shown in the sections of Fig. 6.1.a. and 6.1.c. where no averaging is done. (These results were obtained in the

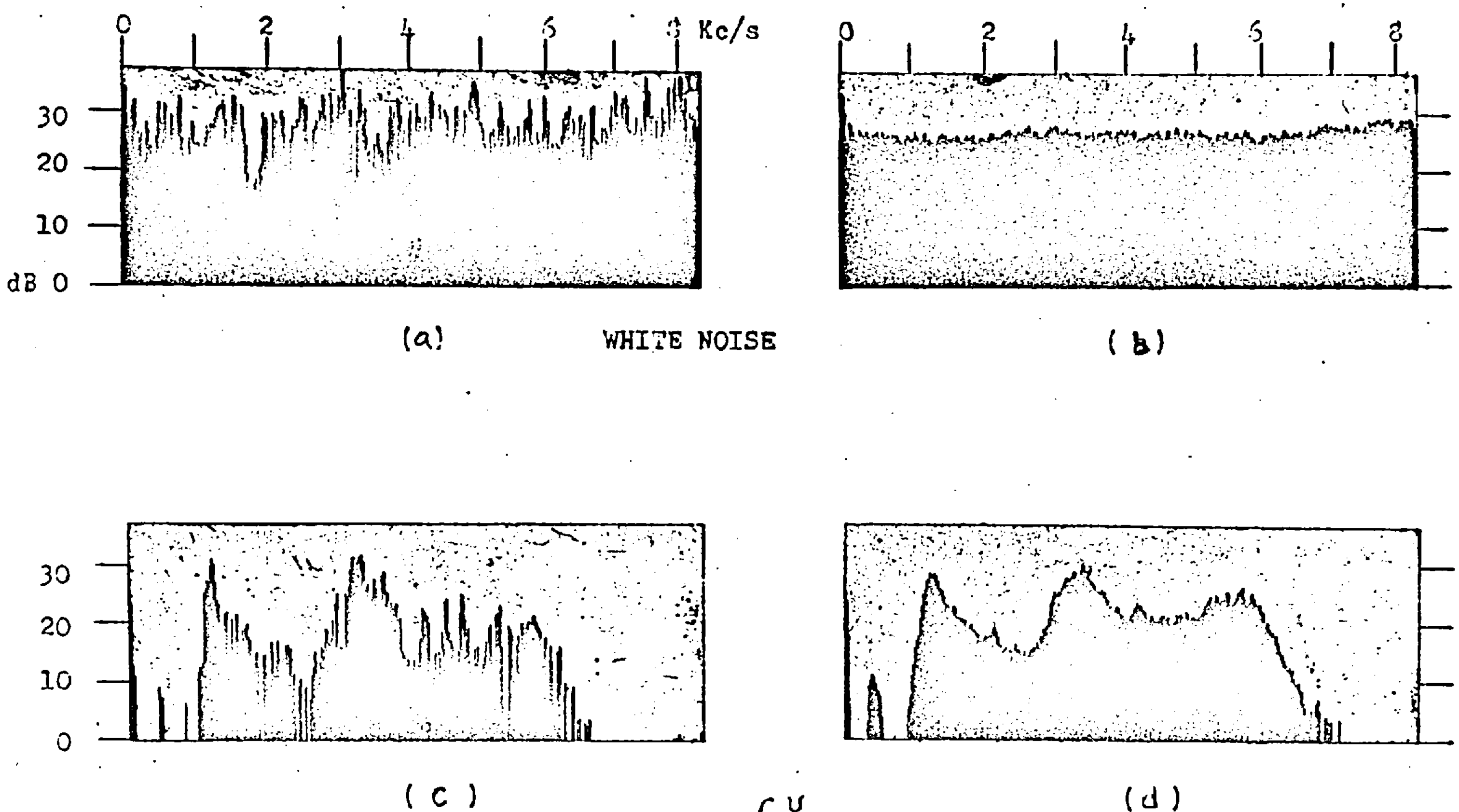


Fig. 6.1

Phonetics Department at University College). Wood (1964) mentions an integrator which will average an arbitrary number of spectral sections as they are presented from an analyser. Such a device could well be used with the IC analyser. Since the bandwidths and frequency characteristics of analysing filters largely determine the shapes of spectral sections it would then be advantageous to have the arbitrary choice of filter characteristics.

6.2. Suggestions for Further Work.

6.2.1. Demonstration of the Fourier Integral Transformation.

In Figures 2.9 to 2.17 are shown displays of the moduli of the Fourier integral transformations of a single period of several weighting functions. The curves are presented as a series of samples obtained from a filter in the IC analyser as a sinusoidal source is swept slowly and linearly in frequency, but interpolation between the samples can easily be performed. In this case, the envelope of the voltage waveform in the filter was sampled. However, phase data is also present in the filter waveform, and if both phase and modulus could be measured a true Fourier integral transformation of the weighting function could be plotted.

Since most students find the concept of the Fourier integral transformation intuitively difficult at first, a Fourier transformer of this form could prove useful in demonstrations. All that would be necessary is (a) paint or cut out the desired function on a card, (b) insert the card into a photo electric function generator, (c) apply the resulting periodic time function to the Fourier transformer, and (d) display the Fourier transformation on a CRT or plot with a pen recorder.

Such a device could also be useful in research where many functions are not easily expressed mathematically. For these functions the Fourier transformations are difficult to obtain by other means. A simple example of a case where a Fourier transformer would be useful is in demonstrating to students the relationship between the response of a circuit to a Dirac pulse and its frequency transfer characteristics. The circuit impulse response could first be obtained experimentally and, using the Fourier transformer, the frequency characteristics then plotted. A check could then be obtained by using a conventional signal generator as a source and plotting the frequency characteristics directly.

In Fig. 6.2 is shown a block diagram for a Fourier transformer. The oscillator which generates a sine wave can have its frequency slowly and linearly swept across the range bracketing the tuned frequency of the lossless filter by means of a ramp generator. Its output can then be weighted by the (time) function for which the Fourier transformation is desired. The quenching of the filter must be synchronized such that the sinusoid entering the filter has the same phase at the start of

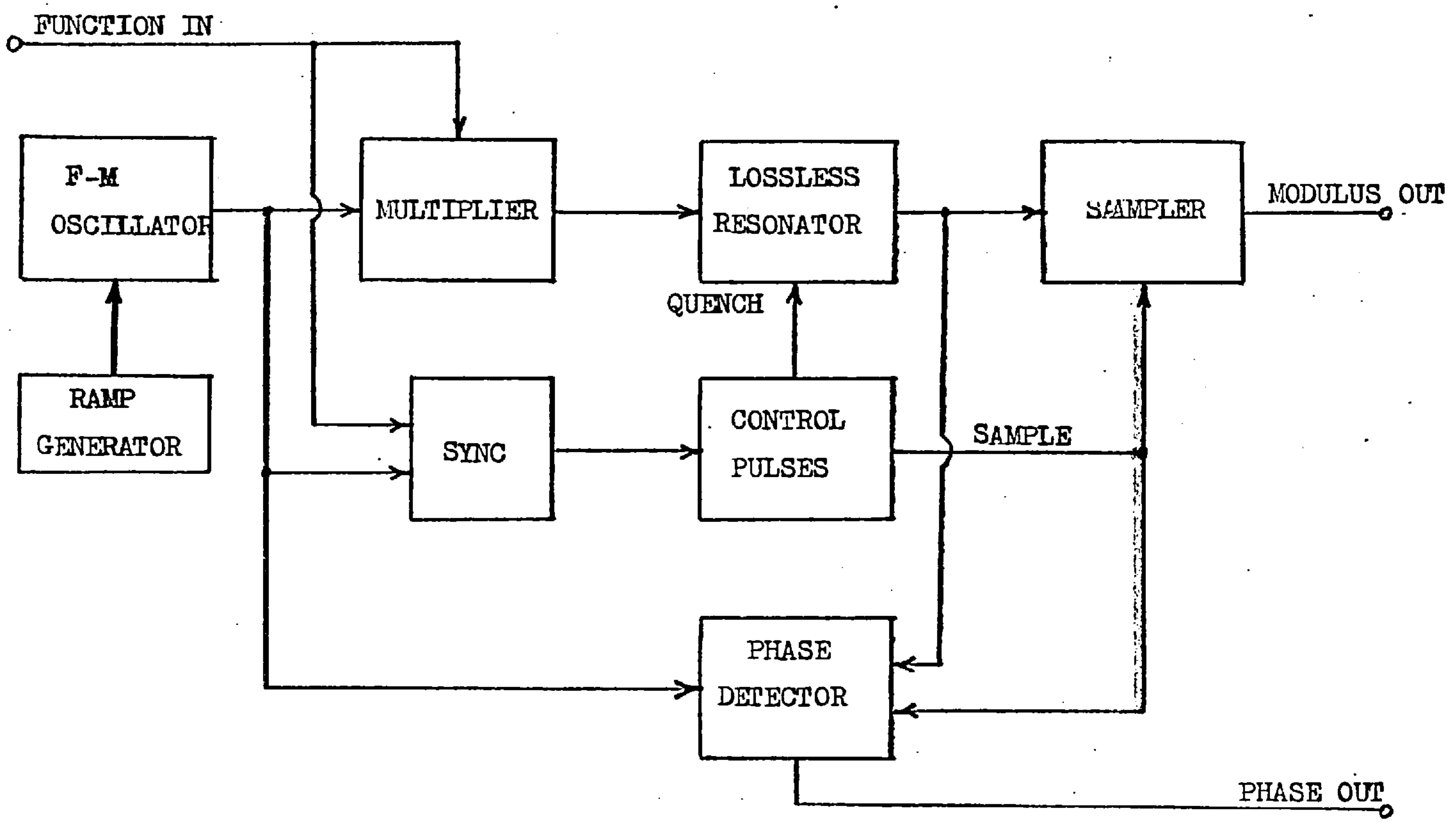


Fig. 6.2. Fourier Transformer.

each fixed integration interval, T. At the ends of the integration intervals both the envelope and the phase of the waveform in the filter can be detected and then displayed on a CRT or recorded. Since there is no imaginary component in the Fourier transform of a symmetrical function the phase detector is unnecessary in such cases. Thus, where the functions are symmetrical, the sampling circuit could be set to sample a point on the filter waveform itself rather than the envelope.

Except for the phase detector the designs for the necessary circuits in such a Fourier transformer are standardized, and most are available in the Signals Laboratory at Imperial College.

6.2.2. Possibilities in Bandwidth Compression.

As mentioned previously, a speech signal can be expressed in terms of complex spectral coefficients in such a way that the original temporal speech waveform can be recovered from the coefficients. In its spectral form a speech signal can theoretically be transmitted over a communication channel with the same bandwidth as that required for transmitting the temporal waveform. After resynthesis at a receiver the results will be the same.

In practice, phase data are not obtained from spectral analysers, so spectral coefficients are real rather than complex and exact resynthesis cannot be carried out. However, in some situations where speech is synthesized from spectral information this is not very important. Golden (1963) claims to have synthesized good quality speech where no phase data were present above 750 c/s when he simulated a vocoder with a digital computer. (He transmitted the spectrum from 250 to 750 c/s intact, thus retaining phase information in this region).

From spectrograms of speech signals it can be seen that at any moment there are certain frequency bands where very little energy lies. The result on the time-frequency plane is areas where little or no energy is observable. Since these 'empty' areas cover a large part of the time-frequency plane it would be advantageous for purposes of transmission to redistribute the spectral coefficients having a significant amplitude by means of encoding in such a way that no empty areas remained. This would result in a considerable saving of bandwidth in speech transmission.

Cherry and Gouriet in 1953 discussed the possibility of bandwidth compression of television signals by recoding, and since that time research has been carried out to evaluate a bandwidth compression scheme (Cherry, et al, 1963). In this scheme advantage is taken of correlation between groups of successive picture elements in a line of the television picture. Only one of a group of picture elements of the same amplitude may be transmitted along with position information. Similarly the spectral sections from the IC analyser could be treated in the same way as lines in a television picture to obtain speech compression.

One of the problems in television bandwidth compression is the construction of reasonably large high-speed storage units for coding. This has meant that practical advantage can be taken of the statistical redundancy in only one line at a time. Since speech signals are much lower in frequency and each spectral section has many fewer data than a line in a television signal, coding problems will be lessened, and storage units can be built which will hold several spectral sections at once rather than only one. In this case, it is possible that one of a group of spectral samples covering an <u>area</u> on the time-frequency plane can be transmitted, while position information is sent to show where the area lies. This could result in much greater compression than treating one spectral section at a time.

The above is only one possibility for recoding speech signals and further study can be carried out to decide on a suitable method. It should be noted that the above method of bandwidth compression does not discard some of the spectral data (except phase) as happens in other speech compression schemes.

Estimates of the possible compression of speech samples can be obtained by measuring the statistics of the spectral coefficients on the time-frequency plane. Correlation between coefficients within a single spectral section or between successive sections can be obtained by converting the coefficients to digital form and using comparator circuits and counters. One of the simplest statistics to measure is the number of spectral coefficients falling above or below a chosen pedestal which represents a point below which the amplitude of a coefficient is considered negligible. Only a Schmitt trigger and a counter are required.

Before coding systems are devised, however, the quality of speech synthesized from the spectral coefficients obtained from the IC analyser should be evaluated. A synthesizer can be constructed which is effectively the mirror image of the analyser. Since the best results in Vocoders at the Bell Telephone Laboratories have been obtained where part of the lower end of the speech spectrum has been transmitted intact, it is likely that this approach is most promising. The analysis, coding, decoding and synthesis can then be carried out for the higher frequency part of the spectrum above say 700 c/s, and the base band below 700 c/s can be added after synthesis has taken place. A simple block diagram of a suggested bandwidth compression system of this type is shown in Fig. 6.3.

A synthesizer should take the form shown in Fig. 6.4. For each filter in the analyser there is one channel in the synthesizer. For two banks of filters there must then be two banks of channels in the synthesizer. Each channel consists of a storage circuit to hold the spectral coefficient for the time interval T, a simple multiplier, and an oscillator. The amplitude

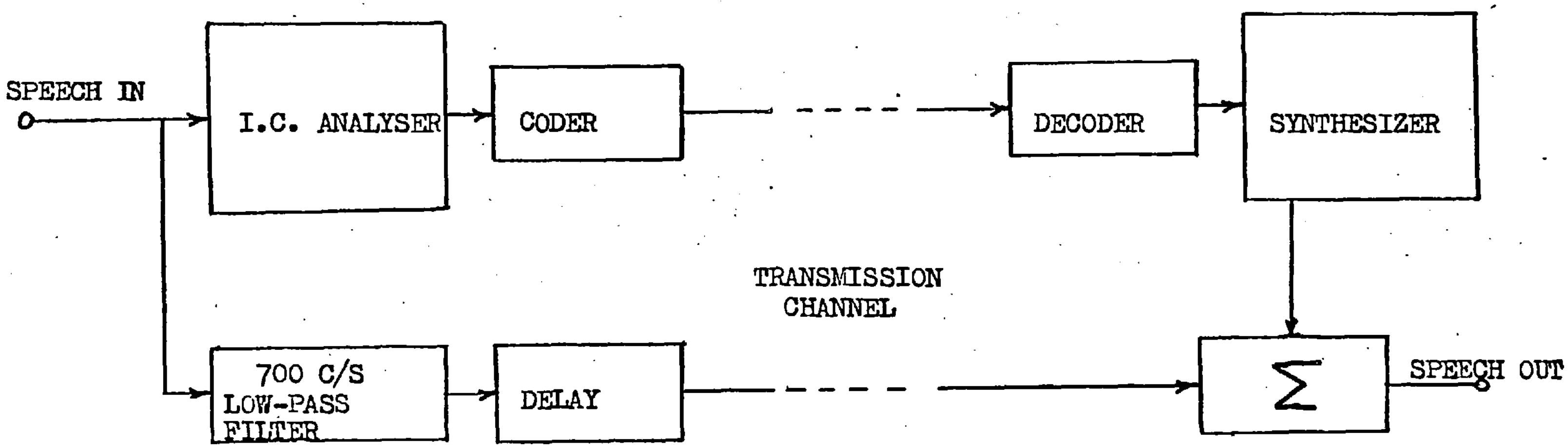


Fig. 6.3. Proposed speech bandwidth compression system.

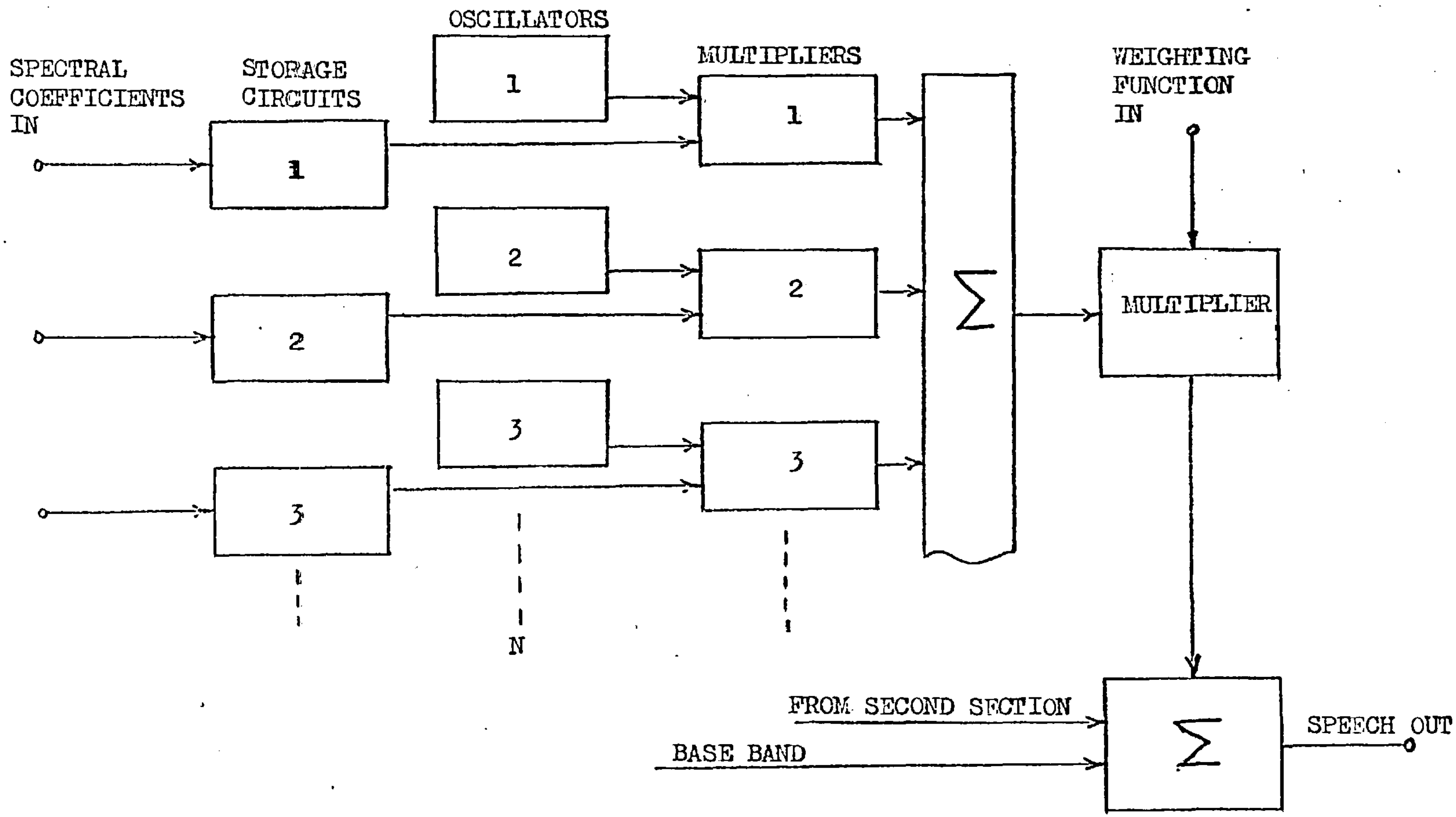


Fig. 6.4. Proposed speech synthesizer.

of the oscillator output is set by the amplitude of the spectral coefficient. All outputs from the multipliers in one bank are added together in a simple resistor network and passed into a time division multiplier of the form used in the IC analyser. Here the sum is weighted by the same weighting function as used in the analysis.

Each oscillator can be shared by two channels, one from each bank. There must also be a second time division multiplier so the out of phase weighting function in the second filter bank can be applied. The complete synthesis is obtained by adding the outputs from the two sections of the synthesizer and the 700 c/s base band.

Square wave oscillators can be used in the synthesizer so that the simple multipliers in each channel can be simple transistor switching circuits. In this case, the synthesis must take place over an intermediate frequency range above the normal speech spectrum so that the high frequency components introduced by the square waves can be removed by low-pass filtering just after the summing. Storage circuits can take the same form as those in the IC analyser and, for preliminary experiments, the actual circuits in the analyser can be used. There are 32 storage circuits in the IC analyser which is sufficient for a synthesizer having 16 oscillator frequencies spaced at 200 c/s intervals.

In a synthesizer like the one suggested here, there will be some noise introduced mainly due to the sampling of spectral coefficients on the time axis. The choice of temporal window functions will then be quite important. If no time-weighting is used in the analysis and synthesis there will be sharp transitions (because of the rectangular temporal windows) between successive sets of samples (spectral sections) which will cause a buzz at the sampling rate in the synthesized speech. Analysis and synthesis in this way is the same as that suggested by Harmuth (1960) where he discusses speech transmission in terms of orthogonal function sets (rectangular envelope temporal windows and sin x/x spectral windows).

When time-weighting is used, however, the transition between successive sets of spectral coefficients can be smoothed in any desirable way. A good example is the raised cosine function. If analysis is carried out with filters spaced at 100 c/s intervals there will be one hundred Δt intervals (and thus one hundred sets of samples) per second. In the synthesis the transition from one set of temporal windows to the next is smooth and the sampling noise will thus be confined mainly to a 100 c/s component. If this component proves objectionable it can be removed by filtering before the base band is added. However, the noise cannot be so easily removed when no time-weighting is used.

The alternative to building analogue circuits for evaluating speech from the suggested synthesizer is to program a digital computer to simulate the system. The spectral coefficients from the IC analyser can be quantized, recorded, and fed into the computer where the mathematical manipulations are carried out. A study of a suitable method for coding can also be carried out by computer if desired. In the end, however, only systems which are practical to build will prove desirable in any future applications.

APPENDICES

APPENDIX I: ACTIVE FILTERS IN THE IC ANALYSER

7.1. Circuit Description.

The active filter circuit used in the IC analyser is shown in Fig. 7.1. The resonant circuit consists of the inductor and capacitor C. T_2 is used as an amplifier to provide feedback to the tapped inductor which acts much like an auto-transformer. The feedback is adjusted by the 1K Ω potentiometer such that the loop gain is equal to one and the resonant circuit behaves as if it were lossless. T_1 is turned off during the integration period of the filter and is switched on for short intervals to remove the energy from (quench) the resonant circuit after each integration period. The resistive pad at the input allows the sensitivity of the circuit to be adjusted without disturbing the effective loop gain and thus the Q of the resonant circuit. The 500 K Ω potentiometer is set to leave the minimum residue in the circuit after quenching. The gain of the amplifier T_2 as an emitter follower stage is very stable which ensures high stability in the loop gain or circuit Q.

7.2. Q - Multiplication.

The circuit in Fig. 7.1 is similar to one suggested by Bachmann (1958) for the purpose of Q - multiplication. The addition of quenching makes it useful as a sampling filter. The analysis to show how the resonant circuit is made effectively

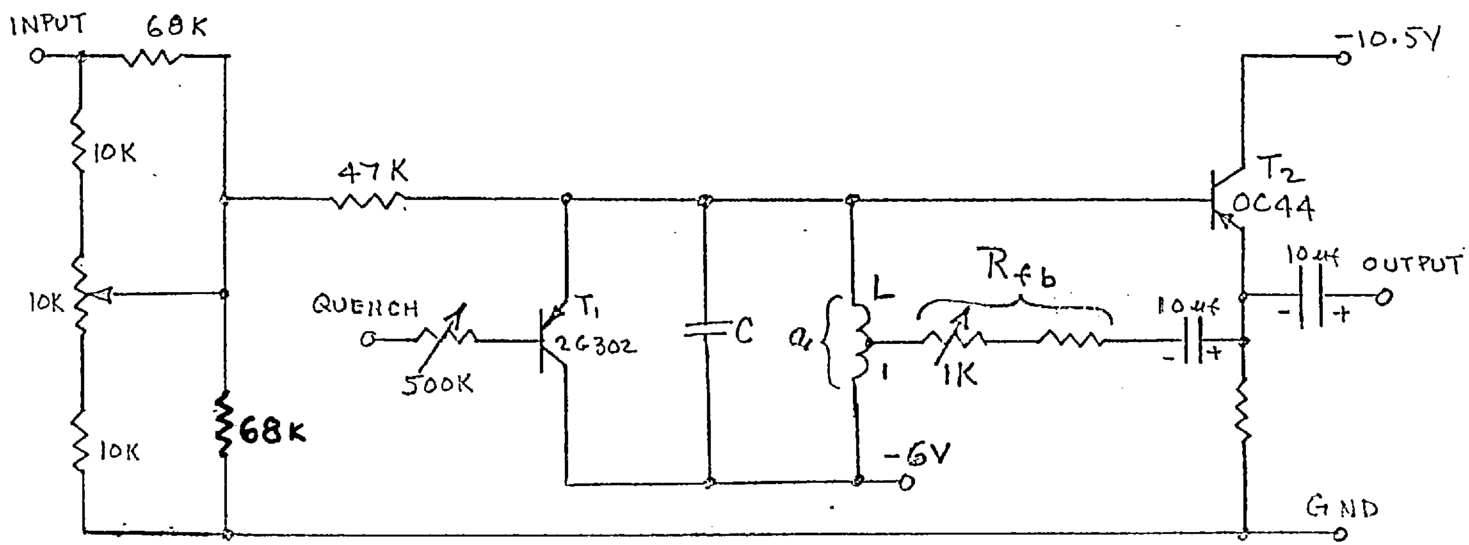


Fig. 7.1 Active filter circuit in the I.C. analyser.

lossless, i.e. given an infinite Q, is done here.

A parallel resonant circuit may be shown in either of the configurations in Fig. 7.2. The Q of such a circuit is defined as

$$Q = \frac{\omega_o L}{R_s} = \frac{1}{\omega_o C R_s} = \frac{1}{G_p \omega_o L}$$

$$= \frac{\omega_o C}{G_p} \qquad \ldots(1)$$

where G_p is the parallel conductance and R_s is the series resistance in the inductor.

$$\omega_o^{\,2} = \frac{1}{LC}\left(1 - \frac{L}{R_p^{\,2} C}\right) \approx \frac{1}{LC} \qquad \ldots(2)$$

for $R_s \ll \omega_o L$

The Q of this circuit can be increased by adding a negative conductance G_n to G_p or a negative resistance R_n to the series resistance R_s. The Q then becomes

$$Q = \frac{\omega_o L}{R_s + R_n} = \frac{1}{\omega_o C \,(R_s + R_n)}$$

$$= \frac{1}{(G_p + G_n)\omega_o L} = \frac{\omega_o C}{G_p + G_n} \qquad \ldots(3)$$

Note that $R_n \neq 1/G_n$.

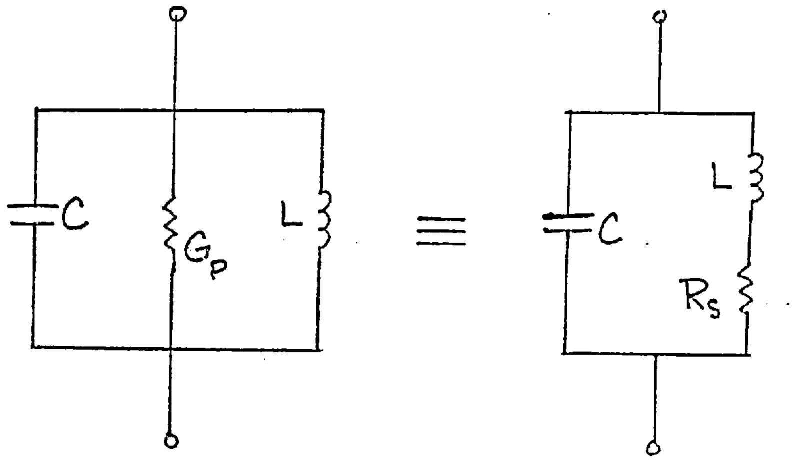


Fig. 7.2 Parallel resonant circuit.

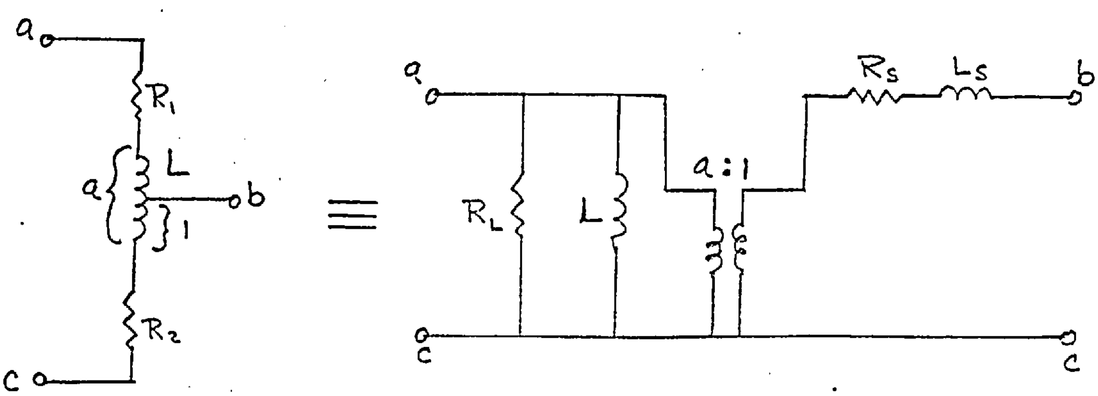


Fig. 7.3 Equivalent circuit for a tapped inductor.

There are three conditions depending upon the values of G_n and R_n.

1. When $|G_n| < G_p$ or $|R_n| < R_s$, the Q is positive and finite and the circuit is damped or lossy.
2. When $|G_n| = G_p$ or $|R_n| = R_s$, the Q is infinite and the circuit is lossless.
3. When $|G_n| > G_p$ or $|R_n| > R_s$, the Q is negative and the circuit is underdamped or oscillatory.

Condition 2 is the desired condition for the filters in the IC analyser.

7.3. Equivalent Circuits.

The circuit of Fig. 7.1 may be redrawn in an equivalent form to simplify the analysis. To do this, an equivalent circuit must also be used for the tapped inductor. Such an equivalent circuit which proves suitable is one suggested by Page (1958) and illustrated in Fig. 7.3. For a good quality inductor such an equivalent circuit is quite accurate.

Using the equivalent circuit of Fig. 7.3 it is possible to draw the low frequency AC equivalent circuit of the filter as in Fig. 7.4. In this case R_T is the shunt resistance of the switching transistor T_1, R_i is the source resistance, R_a is the input impedance of the transistor amplifier T_2, and R_o is the output impedance of the amplifier.

The equivalent circuit in Fig. 7.4 can be further

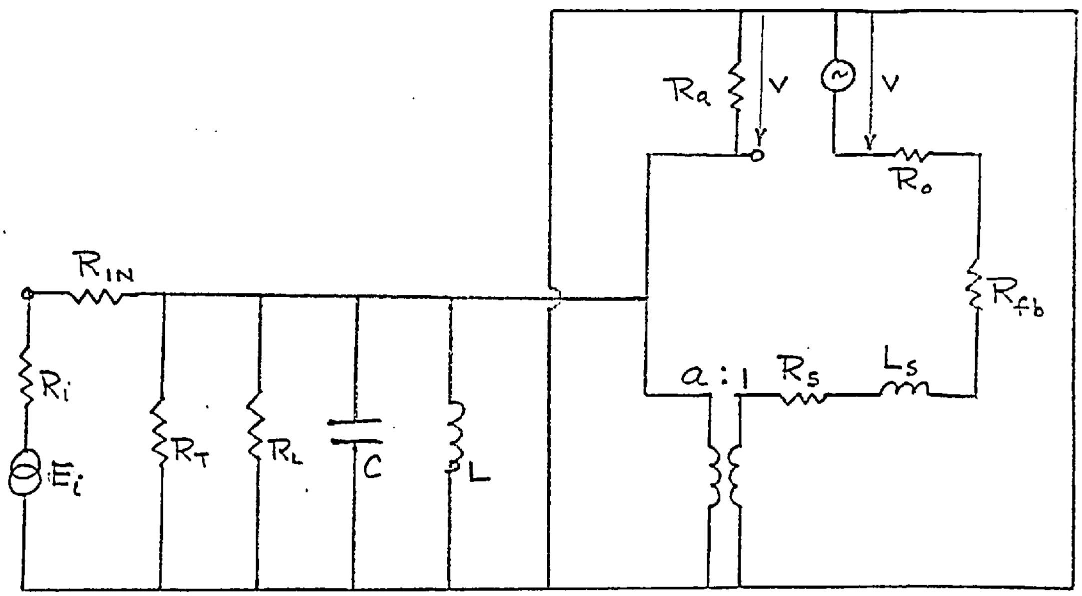


Fig. 7.4 Equivalent circuit for the active filter.

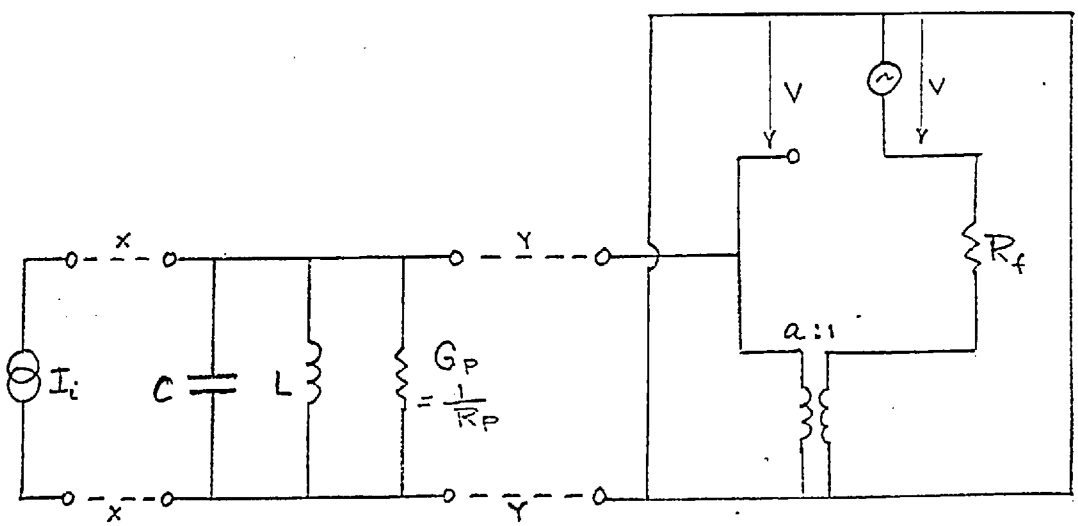


Fig. 7.5 Simplified equivalent circuit for the active filter.

simplified by transforming the voltage source to a current source by Norton's theorem and by lumping all the shunt resistors (R_a, R_L, R_T, $R_{in} + R_i$) together. In the practical circuit $W_o L_s << R_{fb}$ and can be neglected while R_o, R_{fb} and R_s are lumped together. The resulting equivalent circuit is shown in Fig. 7.5. Between the terminals x - x and y - y is the damped resonant circuit. The current entering the circuit to the right of terminals y - y can be expressed as

$$I = GV = \frac{V}{a^2 R_f} - \frac{V}{a\ R_f} \qquad \text{...(4)}$$

where G is the conductance of the circuit. This yields

$$\frac{I}{V} = G = \frac{1}{a^2 R_f} (1 - a) \qquad \text{...(5)}$$

If $a > 1$ the conductance seen looking to the right of terminals y - y is negative. This is in fact the case and

$$G = G_n \qquad \text{...(6)}$$

The total conductance in parallel with the LC resonant circuit is $G_p + G_n$ and the Q is given by

$$Q = \frac{1}{\left[G_p + (1-a)/a^2\ R_f\right]\ \omega_o L}$$

$$= \frac{\omega_o C}{G_p + (1-a)/a^2\ R_f} \qquad \text{...(7)}$$

Decreasing R_f increases the Q of the circuit, and when the denominator goes to zero the Q becomes infinite and the resonant circuit appears lossless.

During quenching, transistor T_1 is turned on such that R_T drops to a very low value. G_p then takes a new value G'_p where

$$|G_n| << G'_p \qquad \ldots(8)$$

The new value of the circuit Q is given approximately by

$$Q = \frac{1}{G'_p \omega_o L} = \frac{\omega_o C}{G'_p} \qquad \ldots(9)$$

Since G'_p is large the Q is low and the energy in the LC circuit is quickly dissipated. Also the transistor T_1 effectively shorts out the input to the circuit.

APPENDIX II: FURTHER DETAILS OF THE CIRCUITS

7.4. Equalizer Circuit.

The RC equalizer circuit used for emphasizing the higher frequency components in speech is shown in Fig. 7.6. A plot of the output VS input frequency is shown in Fig. 4.1. This circuit was chosen because its response curve flattens off at the low frequency end, and it has a slope of approximately 6db per octave over the spectrum of interest before beginning to flatten off again at the higher frequency end.

7.5. Synchronization and Control Pulses.

The method used for synchronization and for generating control pulses is shown in Fig. 7.7. Synchronization can be achieved by applying the time-weighting function to the Schmitt trigger. The two pulse delay circuits are emitter-coupled monostable multivibrators. The first one sets the time at which filter bank A will be sampled and the second the time at which filter bank B will be sampled. They can be adjusted as desired to ensure that sampling occurs at the correct point on the weighting function.

The pulses from the emitter-coupled monostable multivibrators trigger MMV-1, which is a conventional monostable multivibrator. MMV-1 provides a pulse to quench the spectral

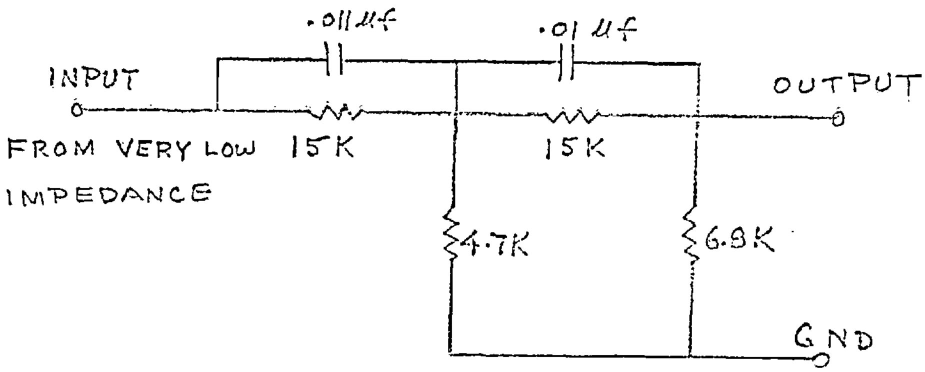


Fig. 7.6 Equalizer circuit.

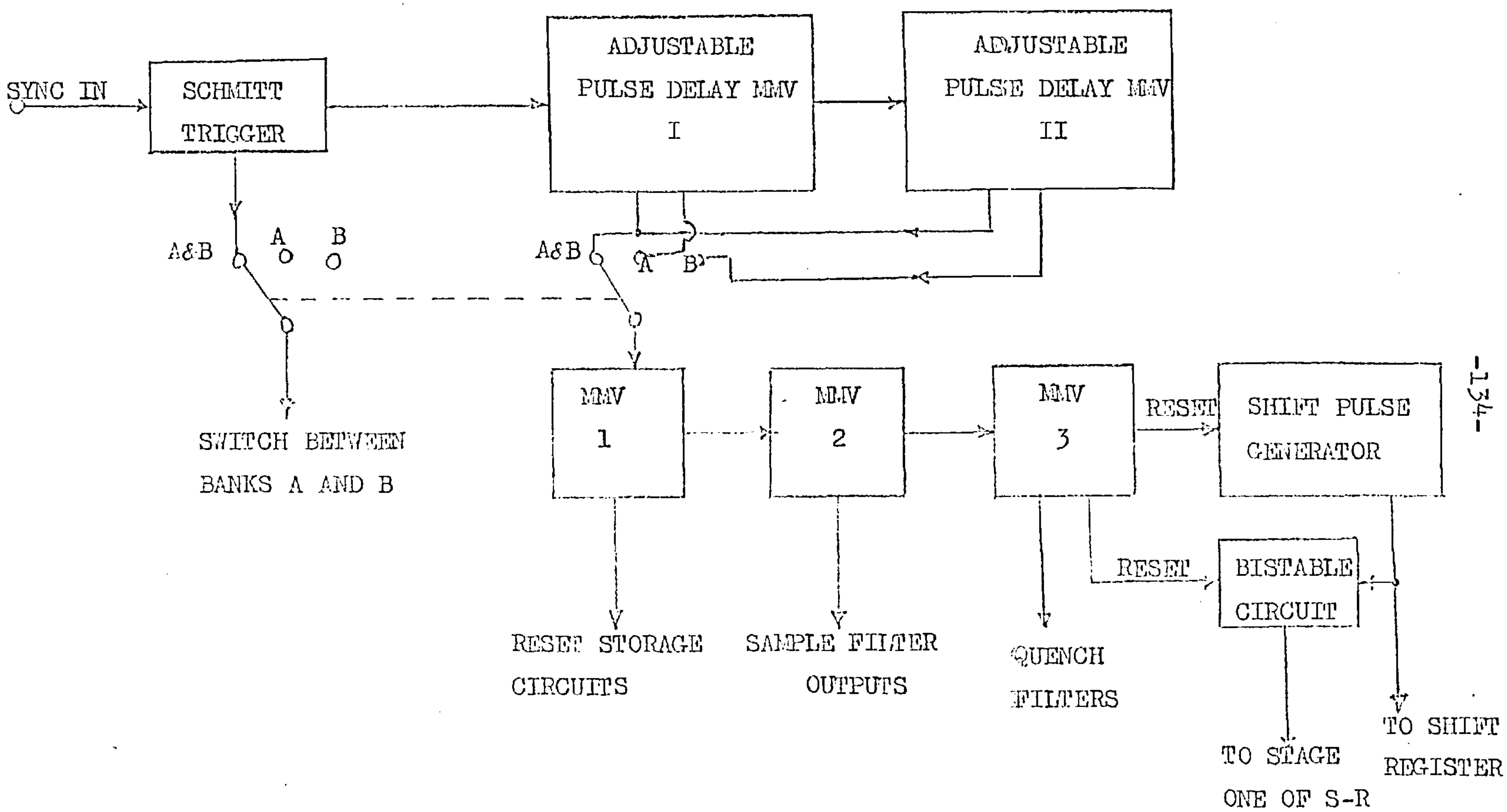


Fig. 7.7 Circuits providing synchronization and control pulses.

samples in the storage capacitors and to trigger MMV-2 with its trailing edge. MMV-2 then provides a pulse to sample the filter outputs which are stored, while MMV-3 is triggered in turn. MMV-3 provides a pulse to quench the energy in the active filters and to initiate the circuits which control the shift register readout circuits. The bistable circuit inserts a one into the first stage of the shift register and the shift pulse generator moves it along one stage at a time to provide a serial readout from the spectral analyser.

The sampling pulses and filter quench pulses can be routed to either bank A or bank B by means of OR gates which are controlled from the Schmitt trigger. Care must be taken that a complex weighting function does not switch the Schmitt trigger in such a way that the pulses are incorrectly routed or generated too frequently. In addition, it is possible by manual switching to choose one filter bank at a time by halving the number of control pulses and routing them all to one bank rather than alternately between the two banks.

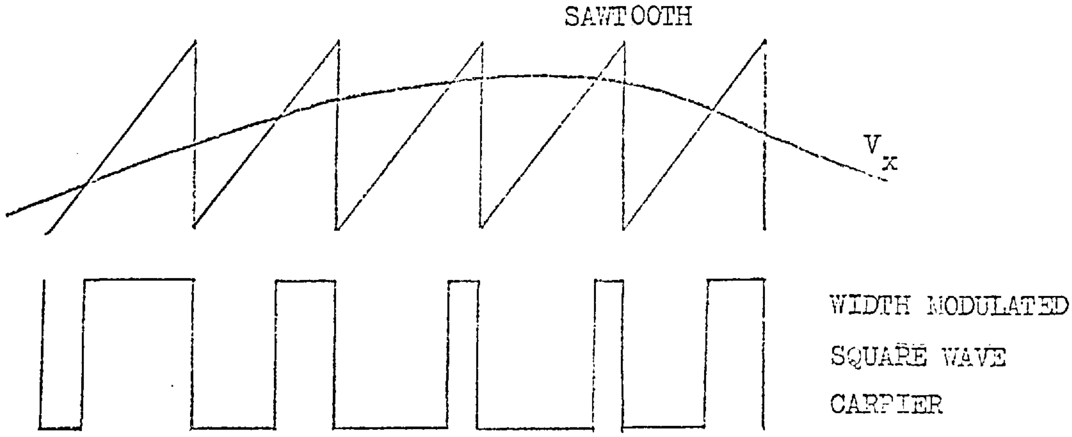
7.6. Multipliers.

In order to accomplish time-weighting of signals entering the analyser it was necessary to have multipliers with four quadrant multiplication, i.e. operation in which each multiplicand may be of either sign. It was desirable to have an accuracy within about 1% in the product and a total product frequency range up to the maximum speech frequencies likely to be analysed plus the weighting function frequencies.

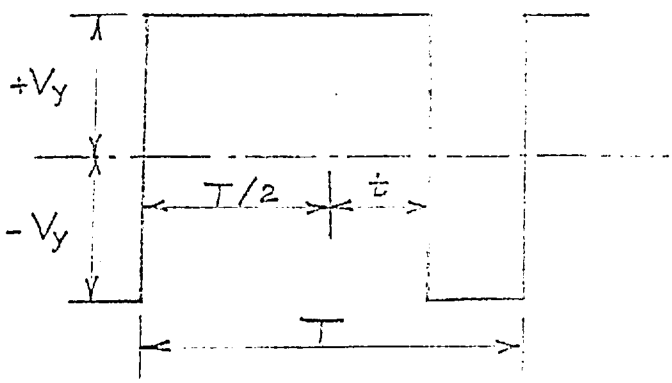
Most commercial multipliers do not fulfill these requirements and they tend to be rather expensive. For these reasons, it was decided to build transistor time division multipliers similar in principle to that suggested by Atkins (1959). The product in such a multiplier is given as the net area under a square wave after simultaneous amplitude and width modulation (see Fig. 7.8.).

Most time division multipliers have a ratio of square wave fundamental frequency to input frequency of 100 or more. Since for input frequencies of several kilocycles per second this means a square wave frequency in the order of megacycles with associated difficulties in the high speed switching circuits, Atkins suggests a much lower ratio. He has calculated the inaccuracies that could result and has chosen as a reasonable limit that the square wave carrier frequency should be four times the maximum product frequency or more. In the multipliers constructed for the IC analyser a square wave carrier frequency was chosen at 70 kc/s which is adequate for an estimated maximum desired product frequency of about 12 kc/s.

The form of the multipliers is shown in Fig. 7.9. A sawtooth waveform is generated at 70 kc/s by the ramp generator. A coincidence, or crossover, detector determines the points at which the ramp and the multiplicand at the x input are at the



(a) Pulse width modulation.



MULTIPLICANDS V_x & V_y

NET AREA

$$= V_y(T/2 + t) - V_y(T/2-t)$$

BUT $t = KV_x$

THEREFORE, NET AREA

$$= 2KV_xV_y$$

(b) Principle of multiplication.

Fig. 7.8

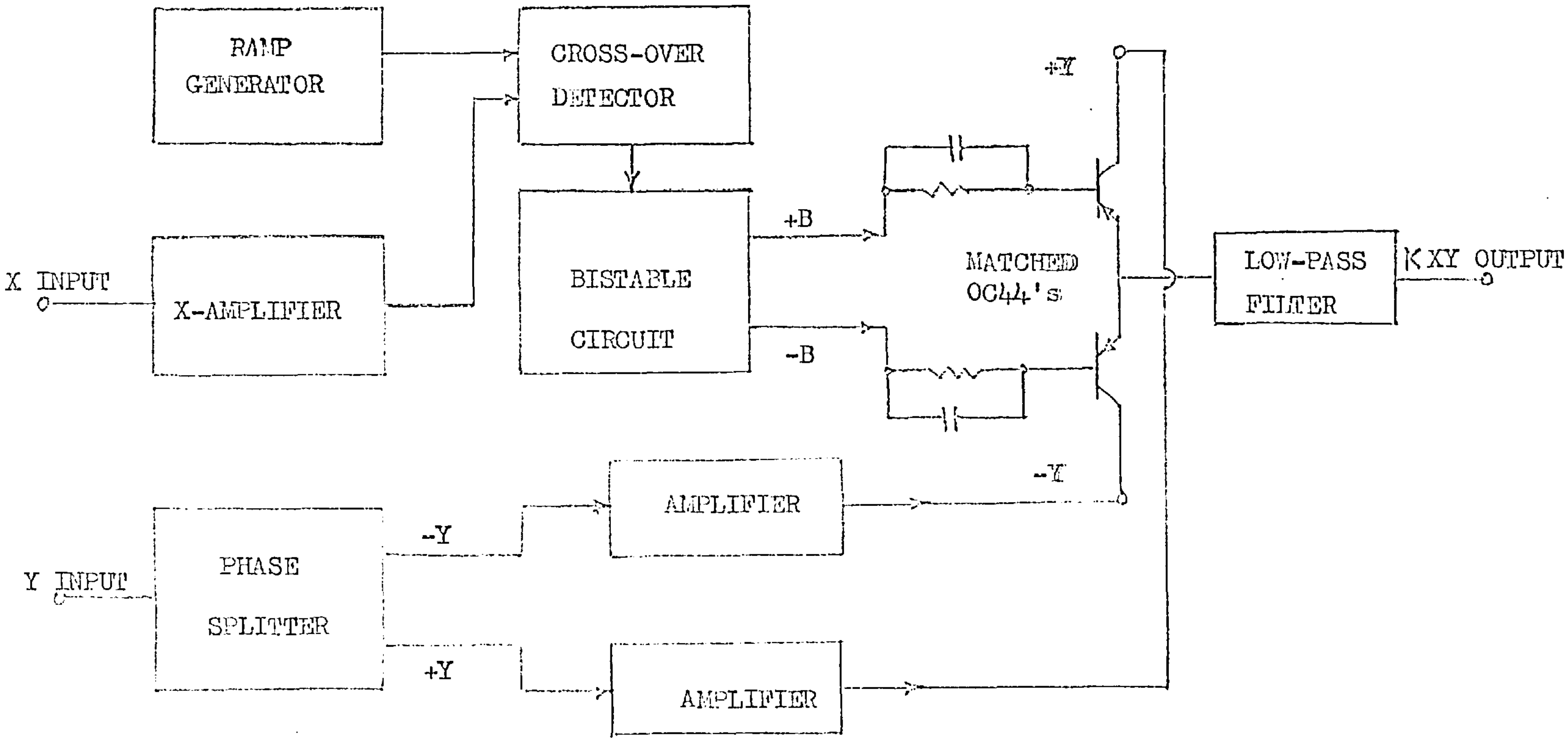


Fig. 7.9 Four-quadrant time division multiplier.

same voltage. At each coincidence point a bistable circuit is triggered. The resulting waveform from the bistable circuit is a square wave carrier having its width modulated proportional to the x-multiplicand.

The width modulated square wave carrier and its inverted component are used to switch the two matched OC44's such that the +Y - multiplicand and the - Y - multiplicand are alternately presented to the low-pass filter. The waveform entering the low-pass filter is a width and amplitude modulated square wave carrier containing an XY product component as well as the carrier frequency and some harmonics. The low-pass filter removes the high frequency components leaving only the component proportional to the XY product at the output.

There is a small pure delay in the formation of the product, but it is unimportant in this case. AC coupling has been used in both channels in the multipliers used here, with a range extending down to about 10 c/s in each channel. This is not a fundamental limitation, however, and DC operation can be achieved with small changes in the circuits. A product containing frequencies up to 12 kc/s can be obtained within an accuracy of 1%. The multipliers have a dynamic range greater than 60db which is adequate for speech signals. Circuit details and zeroing procedures are given in the manual left with the IC analyser.

7.7. DSSC Modulators.

The DSSC modulators used for translating the speech spectrum to the intermediate frequency range where the filter banks are tuned use square wave carriers. This allows the use of switching circuits for the modulation with a resulting suppression of the carrier fundamental frequency in the product.

The block diagram is shown in Fig.7.10. The output from a stable local oscillator is converted to a square wave switching waveform by a Schmitt trigger. This waveform controls a balanced linear gate such that the modulating waveform and its inverted component are alternately presented at the output. The result is an amplitude modulated square wave containing components which are products of the square wave fundamental (and its harmonics) and the input waveform. A sharp cut-off low-pass filter can be used to remove the upper side-band at the carrier fundamental frequency and all higher harmonics, leaving only the lower side-band to go to the filter bank. Low distortion (less than 1%) can be achieved in this type of modulator.

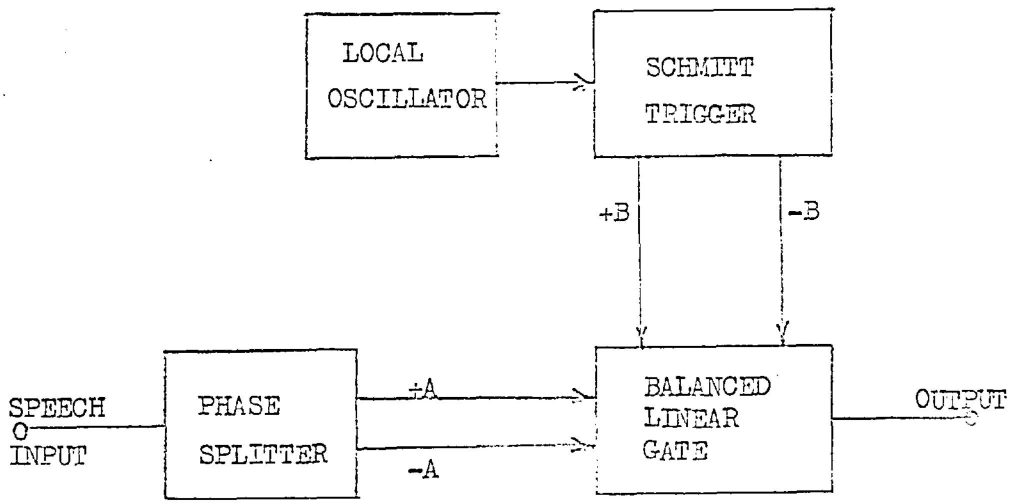


Fig. 7.20 DSSC modulator.

APPENDIX III: GENERATION OF WEIGHTING FUNCTIONS

7.8. Photo Electric Function Generator.

Many simple functions for weighting the input to the IC analyser can be generated by relatively simple circuits, but more complex functions require more versatile generators. One method is to use photo electric means. In a photo electric function generator it is required to prepare only a mask of the desired function and place it in contact with the face of a CRT.

The Signal Processing Engine in the Signals laboratory, Imperial College (H.B. Voelcker, 1961) has been modified as shown in Fig. 7.11 to generate weighting functions. The Engine was originally designed for optical coding of signals, and it uses coding masks placed in front of a CRT. By connecting a negative feedback loop from the photo multiplier, it is possible to make the spot on the CRT follow the edge of a function which has been cut out on a mask. A voltage proportional to the amplitude of the function can then be obtained. An electrical function waveform is produced each time the spot is swept horizontally across the CRT.

A maximum function repetition rate in excess of 1 kc/s is possible with this equipment. A simple RC equalizer circuit in the feedback loop maintains stability. The function masks can be cut in cards or painted on transparent material and inserted into the slot on the front of the CRT which is normally used for coding masks. The electrical function waveform can be observed at one of the vertical deflection plates of the CRT.

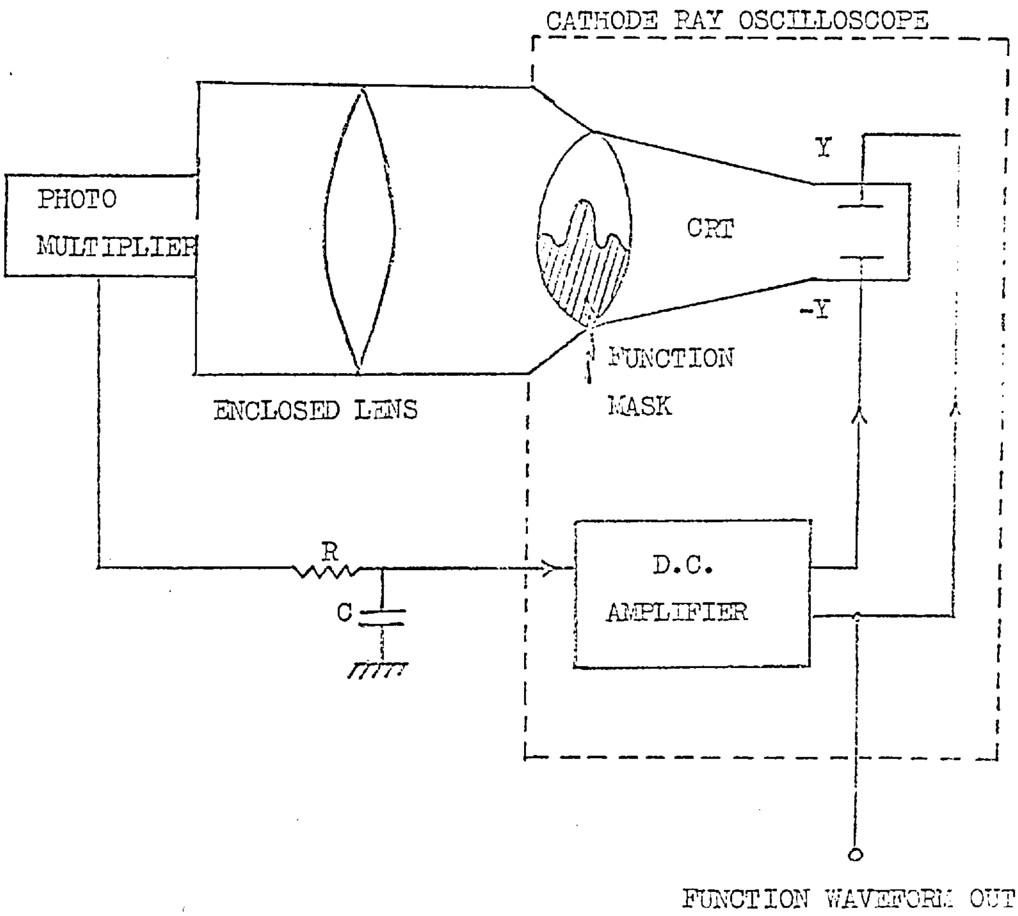


Fig. 7.11 Photo electric function generator.

BIBLIOGRAPHY

1. D.J.H. Admiraal, An Acoustic Spectrum Analyser with Electronic Scanning, Philips Technical Review, Vol. 21, No. 12, pp. 349-356: October, 1960.

2. F. Anderson, An Experimental Pitch Indicator for Training Deaf Scholars, J Acoust. Soc. Amer., Vol. 32, pp. 1065-1074: 1960.

3. W.R. Atkins, Pulse Width Modulation and Analogue Multiplication with Transistor Circuits, M. Sc. Thesis, University of London: 1959.

4. Bachmann, Q-Multiplication mit Transistoren, Archiv der Elektrischen Uebertragung, Vol. 12, pp. 368-370: 1958.

5. C.G. Bell, et al, Reduction of Speech Spectra by Analysis-by-Synthesis Techniques, J Acoust. Soc. Amer., Vol. 33: December, 1961.

6. R.B. Blackman and J.W. Tukey, The Measurement of Power Spectra, Dover: 1958.

7. E. Brookner, Synthesis of an Arbitrary Bank of Filters by Means of a Time-Variable Network, IRE International Convention Record, Part 4, p. 221: March, 1961.

8. S.J. Campanella, A Spectrographic Study of Formant Tracking, Proc. Speech Communication Seminar, Stockholm: August 29 - September 1, 1962.

9. E.C. Cherry, On Human Communication, Wiley: 1957.

10. C. Cherry and G.C. Gouriet, Some Possibilities for the Compression of Television Signals by Recoding, Proc. IEE, Vol. 100, pp. 9-18: January, 1953.

11. C. Cherry, et al, An Experimental Study of the Possible Bandwidth Compression of Visual Image Signals, Proc. IEEE, Vol. 51, pp. 1507-1517: November, 1963.

12. F.S. Cooper, et al, Some Experiments on the Perception of Synthetic Speech Sounds, J Acoust. Soc. Amer., Vol. 24: November, 1952.

13. K.H. Davis, et al, Automatic Recognition of Spoken Digits, J Acoust. Soc. Amer., Vol. 24, p. 637: November, 1952.

14. H. Dudley, Re-making Speech, J Acoust. Soc. Amer., Vol. 11, p. 169: 1939.

15. H. Dudley, The Automatic Synthesis of Speech, Bell Telephone System Tech. Publ., Monograph B-1169: 1939.

16. C.G. Fant, Modern Instruments and Methods for Acoustic Studies of Speech, Proc. VIII International Congress of Linguists, Oslo, p. 282: 1958.

17. C.G. Fant, Acoustic Analysis and Synthesis of Speech with Applications to Swedish, Ericsson Technics, No. 1: 1959.

18. D. Gabor, Theory of Communication, Journal IEE, Vol. 93, Part III, pp. 429-457: 1946.

19. D. Gabor, New Possibilities in Speech Transmission, Journal IEE, Vol. 94, Part III, pp. 369-390: 1947.

20. D. Gabor, A Summary of Communication Theory, Comm. Theory, Ed. Willis Jackson, Butterworth: 1953.

21. J.S. Gill, A Versatile Method for Short-Term Spectrum Analysis in 'Real-Time', Nature, Vol. 189, No. 4759, pp. 117-119: January, 1961.

22. E.M. Glaser, Signal Detection by Adaptive Filtering, IRE Trans. on Information Theory, Vol. IT-7, pp. 87-98: April, 1961.

23. R.M. Golden, Digital Computer Simulation of a Sampled-Data Voice-Excited Vocoder, J Acoust. Soc. Amer., Vol. 35, pp. 1358-1366: September, 1963.

24. M. Halle and K. Stevens, Speech Recognition: A Model and a Program for Research, IRE Trans. on Information Theory, Vol. IT-8, pp. 155-159: February, 1962.

25. H.F. Harmuth, Audio Communication with Orthogonal Time Functions, IEE Monograph No. 405E: October, 1960.

26. C.M. Harris and W.M. Waite, Gaussian-Filter Spectrum Analyser, J Acoust. Soc. Amer., Vol. 35, pp. 447-450: April, 1963.

27. C.M. Harris and W.M. Waite, Response of Spectrum Analysers of the Bank-of-Filters Type to Signals Generated by Vowel Sounds, J Acoust. Soc. Amer., Vol. 35, pp. 1972-1977: December, 1963.

28. R.D. Hawkins, Vibrating Optic Fibers - A New Technique for Audio-Frequency Information Processing and Pattern Recognition, Symposium on Optical Processing of Information, Washington, D.C.: October 23-24, 1962.

29. W.H. Huggins, A Phase Principle for Complex-Frequency Analysis and its Implications in Auditory Theory, J Acoust. Soc. Amer., Vol. 24, pp. 582-589: November, 1952.

30. W.E. Kock and R.L. Miller, Dynamic Spectrograms of Speech, J Acoust. Soc. Amer., Vol. 24: November, 1952.

31. K. Küpfmüller, Die Systemtheorie der Elektrischen Nachrichtenübertragung, 386pp., Zurich: 1949.

32. H.J. Landau and H.O. Pollak, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - III, The Bell System Technical Journal, Vol. 41, pp. 1295-1336: July, 1962.

33. T. Laurent, Delay Time and Transient Time in Electrical Filters with Phase Distortion, Comm. Theory, Ed. Willis Jackson, Butterworth: 1953.

34. W. Lawrence, Parametric Artificial Talking Device, The 60th Meeting of the Acoustical Society of America, Session G: October 20-22, 1960.

35. J. Liljencrants, et al, 51 Channel Analyser for Spectrum Sampling, Proc. Speech Communication Seminar, Stockholm: August-September, 1962.

36. J. Liljencrants, A Few Experiments of Voiced-Voiceless Identification and Time Segmentation of Speech, Report STL-QPSR-3/1962, Royal Institute of Technology, Stockholm: October, 1962.

37. H.J. Manley, Analysis-Synthesis of Connected Speech in Terms of Orthogonalized Exponentially Damped Sinusoids, J Acoust. Soc. Amer., Vol. 35, pp. 464-474: April, 1963.

38. H.F. Olson, et al, Demonstration of Speech Processing System Consisting of a Speech Analyser, Translator, Typer, and Synthesizer, J Acoust. Soc. Amer., Vol. 34: October, 1962.

39. H.F. Olson, Speech Processing Systems, IEEE Spectrum, pp. 90-102: February, 1964.

40. D.F. Page, A Design Basis for Junction Transistor Oscillator Circuits, Proc. IRE, Vol. 46, pp. 1271-1280: June, 1958.

41. G.E. Peterson, Design of Visible Speech Devices, J Acoust. Soc. Amer., Vol. 26, pp. 406-413: 1954.

42. R.K. Potter, et al, Technical Aspects of Visible Speech, J Acoust. Soc. Amer., Vol. 18: 1946.

43. R.K. Potter, A.G. Kopp, and H.C. Green, Visible Speech, D. Van Nostrand Co., Inc., New York, 441pp.: 1947.

44. A.J. Prestigiacomo, Plastic-Tape Sound Spectrograph, J Speech and Hearing Disorders, 22, 89pp.: 1946.

45. A.J. Prestigiacomo, Amplitude Contour Display of Sound Spectrograms, J Acoust. Soc. Amer., Vol. 34, pp. 1684-1688: November, 1962.

46. H.K. Robin, et al, Multitone Signalling System Employing Quenched Resonators for use on Noisy Radio-Teleprinter Circuits, Proc. IEE, Vol. 110, pp. 1554-1568: Sept., 1963.

47. C. Shannon and W.E. Weaver, The Mathematical Theory of Communication, University of Illinois Press: 1949.

48 P.N. Sholtz and R. Bakis, Spoken Digit Recognition using Vowel-Consonant Segmentation, J Acoust. Soc. Amer., Vol. 34: January, 1962.

49. D. Slepian, H.O. Pollak, and H.J. Landau, Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, Bell Telephone System, Monograph 3746: 1961.

50. K.N. Stevens, S. Kasowski, and C.G. Fant, An Electrical Analogue of the Vocal Tract, J Acoust. Soc. Amer., Vol. 25: July, 1953.

51. H. Urkowitz, Analysis and Synthesis of Delay Line Periodic Filters, IRE Trans. on Circuit Theory, Vol. CT-4, pp. 41-53: June, 1957.

52. H.B. Voelcker, Experiments in Signal Analysis and Coding, Report in Signals Laboratory, Imperial College of Science and Technology: 1961.

53. L.A. Wainstein and V.D. Zubakov, Extraction of Signals From Noise, Prentice-Hall, Inc.: 1962.

54. D.E. Wood, New Display Format and a Flexible-Time Integrator for Spectral-Analysis Instrumentation, J Acoust. Soc. Amer., Vol. 36, pp. 639-643: April, 1964.

55. P.M. Woodward, Probability and Information Theory, with Applications to Radar, Pergamon Press: 1953.

(1) F.S. Cooper, A.M. Liberman, J. Borst, Synthetic Speech: A Study of the Auditory Perception of Complex Sounds, Science, Vol. 112, p. 426; Oct. 13, 1950 - Abstract of paper presented at National Academy of Sciences.

(2) C.P. Smith, Selective Compression of Speech Sounds, J. Acoust. Soc. Amer., Vol. 25; 832A, 1953.

(3) C.P.Smith, Speech Sound Coder, J. Acoust. Soc. Amer., Vol. 28; 159A, 1956.