A METHOD OF RAPID AND AUTOMATIC SEARCH FOR RADIO-SIGNALS WITHIN THE HF BAND

A Thesis submitted for the

Degree of Doctor of Philosophy

by

Emmanuel Mama Boye

Department of Electrical Engineering,

Imperial College of Science and Technology,

London, S.W.7.

May, 1968.

ABSTRACT

In the past monitoring of the hf radio spectrum used for communication purposes has been carried out by manual search techniques. This monitoring is necessary so that frequency bands suitable for new communication services can be located and, also, so that new transmissions of interest may be detected as soon as they arise. Many human operators are needed to carry out a comprehensive study of even a narrow frequency band and the work is extremely tedious. The need, therefore, for an automatic monitoring method has become increasingly large and in this thesis a description of an investigation into the problem of designing such a system is made.

The technique to be described is designed to search portions of the hf spectrum which covers the range of frequencies from 2-30 mHz. During the search procedure, parameters characterising the contents of that part of the spectrum being examined are statistically estimated by measuring certain properties of the spectrum. The parameters obtained are then used to establish the existence (or the non-existence) and location of the new transmission signals that may arise.

The parameters estimated are dependent on the forms of the functions that are generated by correlating successive spectrum samples of the same portion of the hf band. It is assumed that the contents of the spectrum vary randomly, so that if two such spectrum samples are $x(t)$ and $y(t)$ then the autocorrelation functions, ϕ_{xx} and ϕ_{yy}, and the cross-correlation function, ϕ_{xy} or ϕ_{yx}, are the

elementary functions from which the relevant parameters are derived. Simple test functions have been used to detect the emergence of new transmissions and the disappearance of old ones. In addition a theory of random points has been introduced and developed to show the probabilistic basis for the decision process.

The main feature of the automatic monitoring system is its conceptual simplicity, and it has a great potential for improving detection efficiency and reducing the necessary man-power. Because of the fundamental principles on which it depends, the system is highly versatile and its use in such fields as oceanography, radio astronomy and aeronautical engineering is envisaged.

Acknowledgements

The author should like to express his gratitude to his supervisor, Professor E. C. Cherry, whose help, guidance and encouragement were quite invaluable. The financial resources for the research made available by the Admiralty (A.S.W.E.), is gratefully acknowledged.

Many discussions with Dr. K.Rajappan. were extremely fruitful and I should like to thank him also. I am deeply indebted to Dr. A. H. Marsh for reading, correcting and criticising the manuscript. Thanks are due to Mr. A. White, "our" senior technician, Dr. D. Goodman, Dr. B. Travers and the many colleagues in the Electrical Engineering Department, whose friendship and help in many diverse ways will be cherished always. I also thank Mrs. Janette Dean who typed the thesis with the competence and the diligence of a real expert. Finally, I thank my uncle, Mr. A. A. Mensah, for his active constant encouragement.

TABLE OF CONTENTS

Page

Page

Chapter 1

INTRODUCTION

Chapter 1 INTRODUCTION

1.1 General Background

The congestion of traffic in the hf band, covering the frequency range between 3 and 30 mHz, is a problem which has engaged the attention of the communication engineer since the discovery of that band in the late twenties. In order to solve the problem, many international bodies have been set up with the authority to regulate communication in this band. These international bodies, armed with the advice of the communication engineer, have taken steps to explore different methods of providing accommodation for the extra traffic within the band. For example, an impressive saving of bandwidth has been realised by the use of single side band techniques, with the result that more and more radio signals can now be accommodated.

As an alternative solution, if the traffic density within the band can be precisely found, then a better knowledge of the degree of congestion can be obtained and used to plan the more efficient use of the entire band. Such a traffic density study will also lead to the discovery of 'empty spaces' which can be employed to accommodate more traffic. Hence, it has become necessary to design surveillance techniques by which wide frequency bands can be searched and the radio signals within the band located and identified. A method of assessing the times of initiation, or cessation, of radio signals in any portion of the hf band will also be valuable in determining the time-to-time occupancy of the region in question. It is the object of the investigation presented in this thesis to give a detailed study of this assessment problem and to expose some of the practical limitations of a surveillance system which can be constructed to search the hf band for transmission changes. Such a practical system, capable of conducting a rapid search, will be useful at air traffic centres, monitoring stations, and missile tracking and launching stations. Furthermore, a ready application of the search technique will be found in the service of radio communication between one land (i.e. fixed) station and many mobile stations where a quick way of seeking free channels and assigning them on demand to users is highly desirable. However, in order to fully appreciate the need for a surveillance system and the complex nature of the investigation, it will be necessary to give a brief account of hf communication and recount some of the inherent problems.

1.2 A Brief History of HF Communication

The first practical use[2] of radio communication was between ships and shore, following Marconi's first patent on wireless telegraphy in 1896 and the demonstration of this development to officials of the British Post Office in that same year. Marconi's early experiments showed that there was considerable attenuation of the propagating wave in the vicinity of the transmitter, but at this time the existence of "skip distance"(see Fig.2,Sect.1.4) and the concepts of maximum and , or lowest usable frequencies were not appreciated.

In succeeding years, a serious problem was the inability of transmitting systems to generate the required power at 'high' frequencies (frequencies in the neighbourhood of 15 kHz were then considered high). This fact, combined with the adverse effects of power absorption in the transmission medium at these 'high' frequencies, determined the maximum frequencies that could then be used for communication purposes.

The development of the vacuum tube and inproved circuit techniques led to the production of transmitters and receivers capable of functioning well above the previously considered maximum frequencies. It was also discovered that sky wave propagation (i.e. waves propagated by mirror-like reflection from the upper layers of the atmosphere) was possible in the 2 to 30 mHz region of the radio frequency spectrum and that waves in this frequency range could travel long distances without encountering high levels of atmospheric noise. The rapid development of long[4] distance world-wide radio communication then followed, and further advances were made during World War II when cables for line communication became unsuitable because of threat of their destruction by the enemy.

1.3 The Basic Hf Communication System

Any radio communication system is required to perform a function that involves the transmission of information, in digital or analog form, from one location to another. The volume of information that can be transmitted and the fidelity with which the transmitted message is reproduced at the receiving end are factors which determine the quality performance of the communication system.

Fig.1.1 is a block diagram of the basic radio communication system. It consists of two sub-systems, a transmitting system and a receiving system which are connected by the propagation medium. The original message to be transmitted is inserted in the transmitter where it is used to modulate an rf electromagnetic wave radiated from the transmitting antenna. During its passage through the transmission medium, the transmitted signal may become disturbed by the medium. Effects of fading (see Section 1.4) may also be noticeable in the received signal which, under normal operation of any practical system, is a noisy attenuated version of the transmitted signal. The process of demodulation and filtering which occur at different stages of reception further add to the noise already present in the signal. Hence, in addition to the channel noise whose sources are mainly atmospheric and man-made, the receiver noise becomes an additional problem.

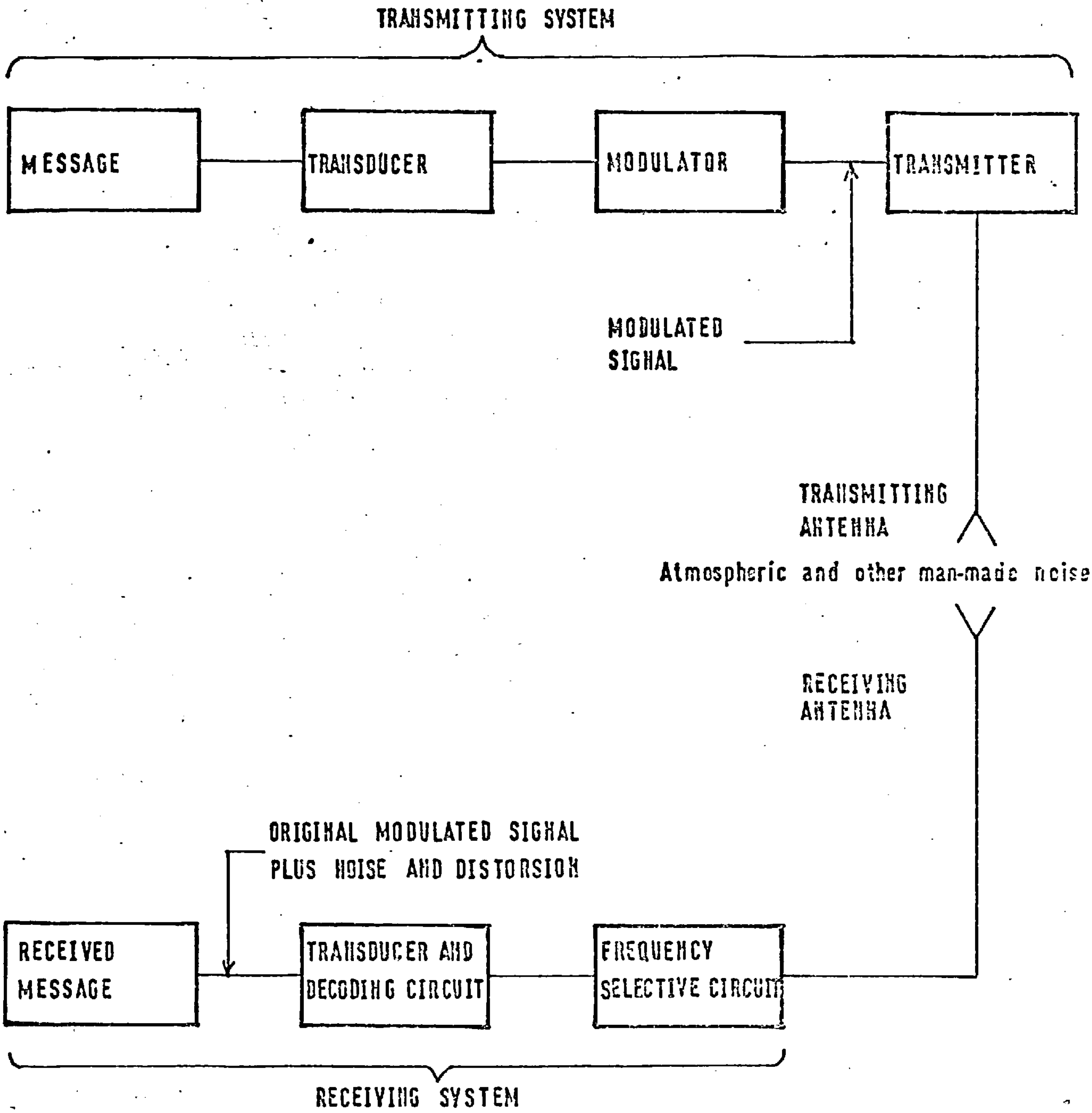


Fig.1.1 Transmitting and Receiving Systems

1.4 The Nature of the Propagation Medium and the Limitations Imposed by It

The efficient exploitation of the radio spectrum will not be possible until a reasonably complete understanding of the structure of the propagation[5] medium is attained. Research by many radio scientists and the development of the appropriate radio communication devices have led to the realisation that in the frequency range, 2 to 30 mHz, where sky wave propagation is possible, the propagation characteristics are largely dependent on the ionosphere, The propagated waves in this range of frequencies are guided within the earth's atmosphere by the earth's surface and the stratified ionosphere. Such sky waves are affected differently by the different strata of the ionosphere, resulting in ionospheric reflection, refraction, diffraction, polarisation, absorption and, or scattering of the waves.

The existence of the ionised layer of gas surrounding the earth was established by Appleton[6] in 1924. Since then, the investigation of the ionosphere has been undertaken on a world-wide scale, resulting in a detailed and an extremely complex picture. Experimental evidence has shown that the ionosphere is made up of three principal regions of varying degrees of ionisation. These are conventionally known, in an ascending order of height and ionisation density, as D-, E- and F- layer. The first measurements[7] to determine the height of the layers above the earth's surface were made by Appleton. Investigation into the diurnal, seasonal, and geographical variations of the ionospheric structure have also

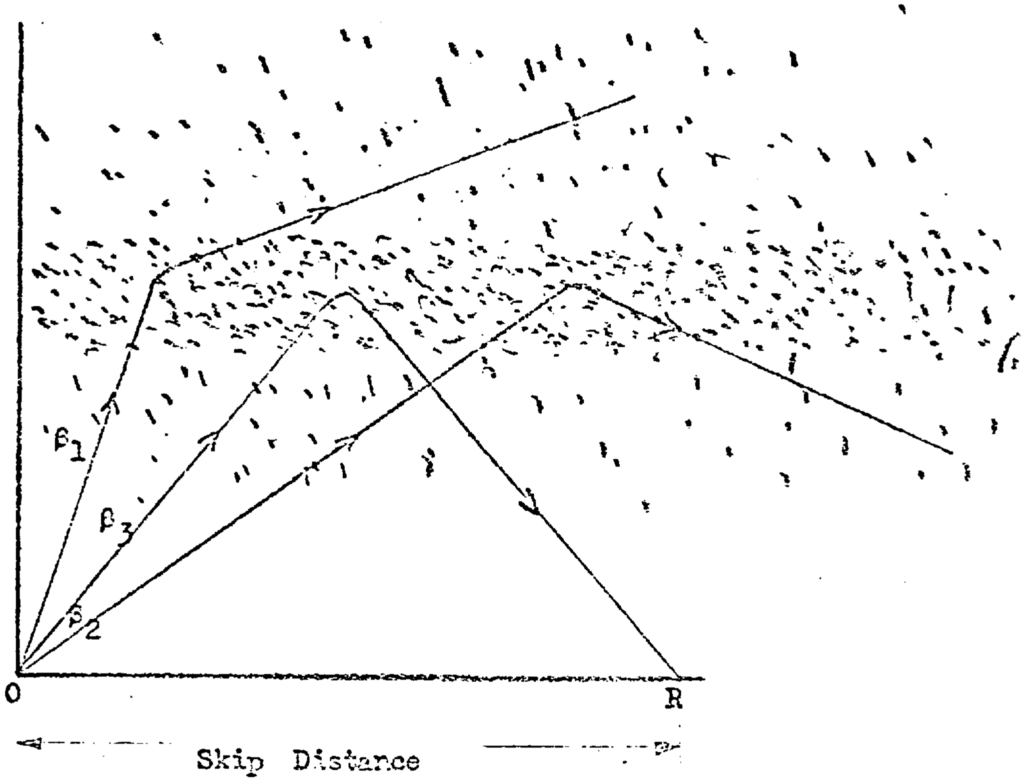


Fig. 1.2 Wave paths illustrating the concept of "skip distance".

been carried out and it is hence possible to determine and predict the values of the maximum useable frequency, the critical frequency (as defined below), the possible ranges of transmission and other relevant quantities.

The critical frequency, which is the maximum frequency at which waves sent vertically upwards are reflected back to the earth, is also related to the maximum ionisation density, N_{max}, by

$$f_c = 9N_{max}^{\frac{1}{2}}$$

(This relationship[8] is valid on the assumption that the earth's magnetic effects are negligible and that the earth's surface is flat). For waves subtending an angle of incidence, i, with the normal to the layer, the maximum frequency for total reflection is given by

$$f_{mn} = f_c \sec(i)$$

and is known as the maximum useable frequency.

Radio waves of frequencies less than the critical frequency of the F-layer will be returned to the earth irrespective of the angle of incidence. It has been shown[9] by Appleton and others[10] that when the frequency of the transmitted wave is higher than the critical frequency of the F-layer, then the only waves that will return to earth are those that will strike the ionosphere at an angle of incidence, β, such that $\cos\beta > \sigma_{max}$, where σ_{max} is the refractive index at the point of maximum ionisation density for the frequency involved. (See fig. 1.2).

Waves striking the ionosphere at an appreciable greater angle of incidence will pass through the ionosphere unreflected while waves of lesser angle of incidence will return to points beyond R (see fig. 1.2). For the latter waves, no sky wave energy is detectable at points closer to the transmitter than R.

Fig. 1.2 illustrates the two types of waves that are generally encountered. Wave 1 strikes the ionosphere at an angle of incidence, β_1, and suffers no reflection; wave 3 subtends an angle, β_3, with the normal and the energy in it is detectable at the point, R. The distance, OR, corresponding to the smallest distance from the transmitter for which waves, striking the ionosphere at oblique angles, is called the skip distance. The angle, β_2, is such that $\beta_3 < \beta_2$ and the corresponding wave is detectable beyond the skip distance.

Changes in the ionisation densities of the different layers of the ionosphere, which are known to be related to the sun spot cycle,[11,12] are the cuase of the corresponding changes in the following quantities: 1) the critical frequency, 2) the maximum useable frequency, 3) the lowest useable frequency, and 4) the skip distance. An empirical study of the changes in these quantities has been made[13] and it is now possible to predict the values of these quantities by comparison and extrapolation from data recorded during a previous sun spot cycle. There are, however, some unpredictable ionospheric changes, known to be caused by solar flares, which may have adverse effects on radio circuits. These sudden ionospheric disturbances (s.i.d.), as they called, are chiefly responsible for drop outs,(i.e.total loss of communication),which may persist from any period of a few seconds to several days.

Fading is another unpredictable phenomenon which imposes limitations on the full use of the available frequency range. The two main types of fading, slow and rapid fading, may occur together or separately and may also affect the different frequency components of a transmitted signal selectively or non-selectively.

In order to combat the effects of fading, the technique of space diversity reception[14] has been suggested and developed,[13] and it is quite common to have more than one receiving antenna at many receiving stations. In frequency diversity reception, use is made of the fact that some frequency components of the transmitted signal, in the presence of selective fading, may experience greater attenuation than other components. Therefore, by a correct choice of signal frequencies, frequency ranges subject to severe selective fading can be avoided so that a channel may be used most efficiently.

The combined effects of noise, fading and other sources of interference are made evident by the presence of errors in the received signal. Efficient means of detecting and correcting these errors have been developed, and an automatic error detection and correction system developed by Dr. Van Duuren of the Netherlands Post Office has paved the way to further development of more sophisticated systems designed to reduce the probability of error to a very low level.

Also, a theory of the probability of error in terms of the signal-to-noise ratio has been evolved[15] and is being used to reduce the effect of noise in the received signals.

1.5 Congestion in the Hf Band and the Purpose of the Present Investigation

The development of efficient, reliable and low cost hf devices and the extensive knowledge of the propagation medium have made hf communication especially popular. However, this very popularity has led to the acute traffic congestion in the band.

In 1947, the Administrative Radio Conference at Atlantic City, New Jersey, U.S.A., was convened to allocate the hf spectrum for the provision of the many required services. At this conference, a governing body was formed to assign portions of the hf spectrum to the individual stations which operated in the band. Prior to this time, the use of a frequency had been simply notified to the I.T.U.[x] for entry on the International Berne List with no attempt to coordinate the requirements of the individual stations. However, because of the steadily increasing congestion, in 1963, at a Geneva meeting under the auspices of the C.C.I.R.,[xx] a highly restricted use of the hf band

---

[x] I.T.U. stands for International Telecommunication Union.

[xx] C.C.I.R. stands for International Radio Consultative Committee.

was recommended.

Under these circumstances, it is desirable to develop searching techniques whereby the population of part of the hf band may be established automatically. An automatic search for signals within the spectrum will quickly provide valuable information about empty spaces which can accommodate new transmissions. In the past, empty spaces in the spectrum have been discovered at manual monitoring stations where field strength measurements and frequency measurements were also carried out. The skilled man-power required for this type of operation was large and, therefore, a means of performing the same task in an automatic fashion would be an invaluable asset. The object of the present investigation is to assess the feasibility of constructing an automatic searching 'machine' which will examine the spectrum and report its findings rapidly. The latter function can be performed by providing a visual or an aural indication of the presence of an 'empty space'. In this thesis, the theoretical basis of the construction of such a 'machine' is examined and the practical limitations are investigated by simulation and practical experiments.

Chapter 2

ANALYSIS OF THE SEARCH PROBLEM

Chapter 2 ANALYSIS OF THE SEARCH PROBLEM

2.1 Introduction

Virtually, all of the hf spectrum is occupied by either noise or active transmission signals.[x] The various transmissions and the noise, if considered as time functions, are inseparable, one from another. Fig.2.1 is an overall representation of hf signals and noise. The noise may be assumed to be a stochastic process and, similarly, the transmissions are also assumed to be random processes because any transmission, represented in the time space as $m(t)$, can be characterized by the probability that at times $t_1, t_2, \ldots, t_n$, the corresponding values, $m(t_1), \ldots, m(t_n)$, lie in some specified intervals. Therefore, a combined wave-form which is the sum of the waveforms shown in figs.2.1a and 2.1b, is also a random process. As a time function, it is impossible to break the combined waveform of fig.2.1c into its component time functions and thereby inspect any one desired component function.

In order to distinguish one transmission from another, or from noise, it is necessary to change the basis of description of the signals and use a frequency rather than a time basis of description.

---

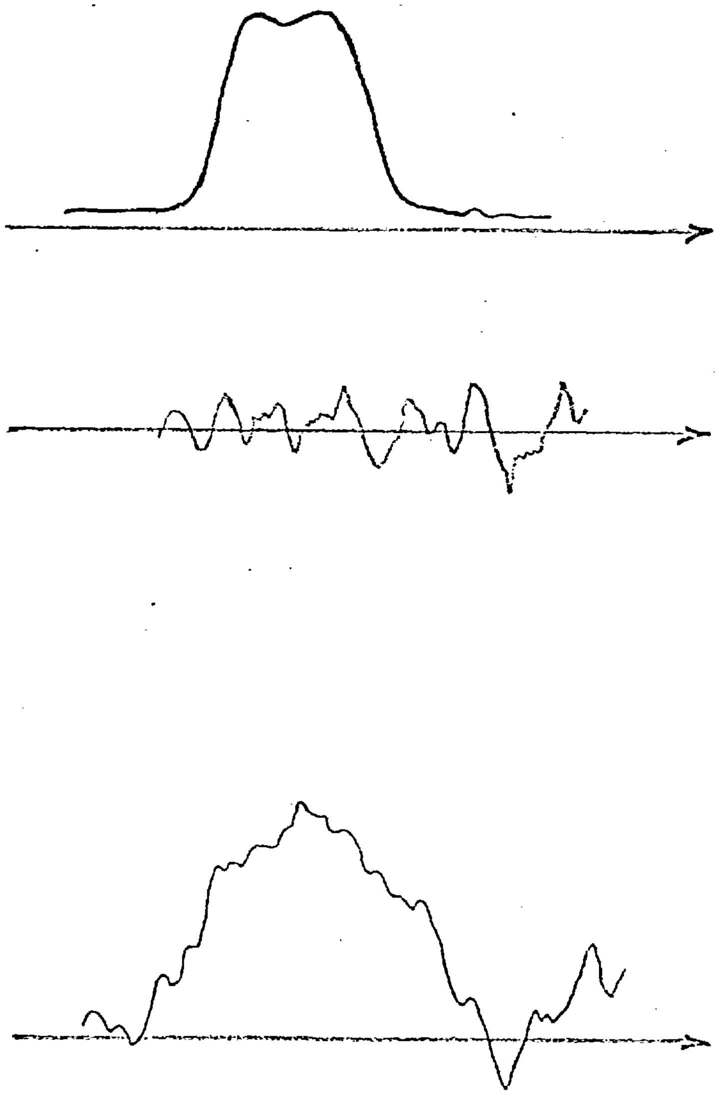
[x]For the purposes of the present investigation, it is necessary to distinguish between a 'transmitted signal' and a 'transmission signal". A transmission signal may comprise a number of signals which are transmitted as a composite wave radiating from one transmitting antenna. Each of these signals will be known as a transmitted signal. The distinction becomes necessary when it is recalled that there are recently developed systems which, for the maintenance of secrecy, are made to transmit pieces of intelligence deliberately hidden in noise.

Fig. 2.1 Two representative time functions of

a) a well defined signal

b) a random signal and

c) a combined signal-plus-random signal



(a)

(b)

(c)

The salient principle of the transformation is that if all the transmissions occupy different parts of the spectrum, then it is possible to separate them in the frequency space.

2.2 Frequency Domain Representation of Radio Signals

The Fourier representation of a signal is perhaps the most important tool which is widely used in radio engineering. In this representation, the functions, e(t) and E(f), are said to be Fourier transforms of each other if they are related in the following manner:

$$E(f) = \int_{-\infty}^{+\infty} e(t) \, \epsilon^{j2\pi ft} \, dt \qquad \ldots (0.1)$$

$$e(t) = \int_{-\infty}^{+\infty} E(f) \, \epsilon^{-j2\pi ft} \, df \qquad \ldots (0.2)$$

The frequency function, E(f), which is known as the spectrum of the time function, e(t), is a frequency distribution which shows the amount of each complex component, $\epsilon^{j2\pi ft}$, necessary for the construction of e(t). It has been shown[16] that this representation of a time function, e(t), in terms of its spectrum, E(f) is valid only if e(t) satisfies certain conditions*, which are generally satisfied by all practical radio signals. Furthermore, if the frequency composition of the radio signal is such that the function, $E(f)$, is practically zero outside a certain band of frequencies, say, between f_L and f_H, then

$$e(t) = 2 \int_{f_L}^{f_H} E(f)\, \epsilon^{-j2\pi ft}\, d\mu \quad \ldots (0.3)$$

Fig 2.2 is a diagram of a signal, $e(t)$, and its spectrum, $E(f)$. If a radio signal whose Fourier transform is limited to W Hz and lasts for T seconds then, according to the Sampling Theorem[17], it can be specified by 2TW equidistant samples of it. The quantity, TW, is the product of signal duration and signal bandwidth, and it is an important parameter in the description of the signal.[18]

The outstanding advantage of the Fourier representation is that if the signal whose spectrum is $E(f)$ is passed through a stationary, linear filter whose system function is $H(f)$, then the spectrum of the

---

* These conditions are that the function may have a finite number of maxima and minima in any interval and that $\int_{-\infty}^{+\infty} e(t)\, dt$ shall be finite. This latter condition might at first glance seem to be serious, since it rules out functions with constant dc component. However, if the dc component is only present for a finite length of time, which is always the case, then the condition is satisfied.

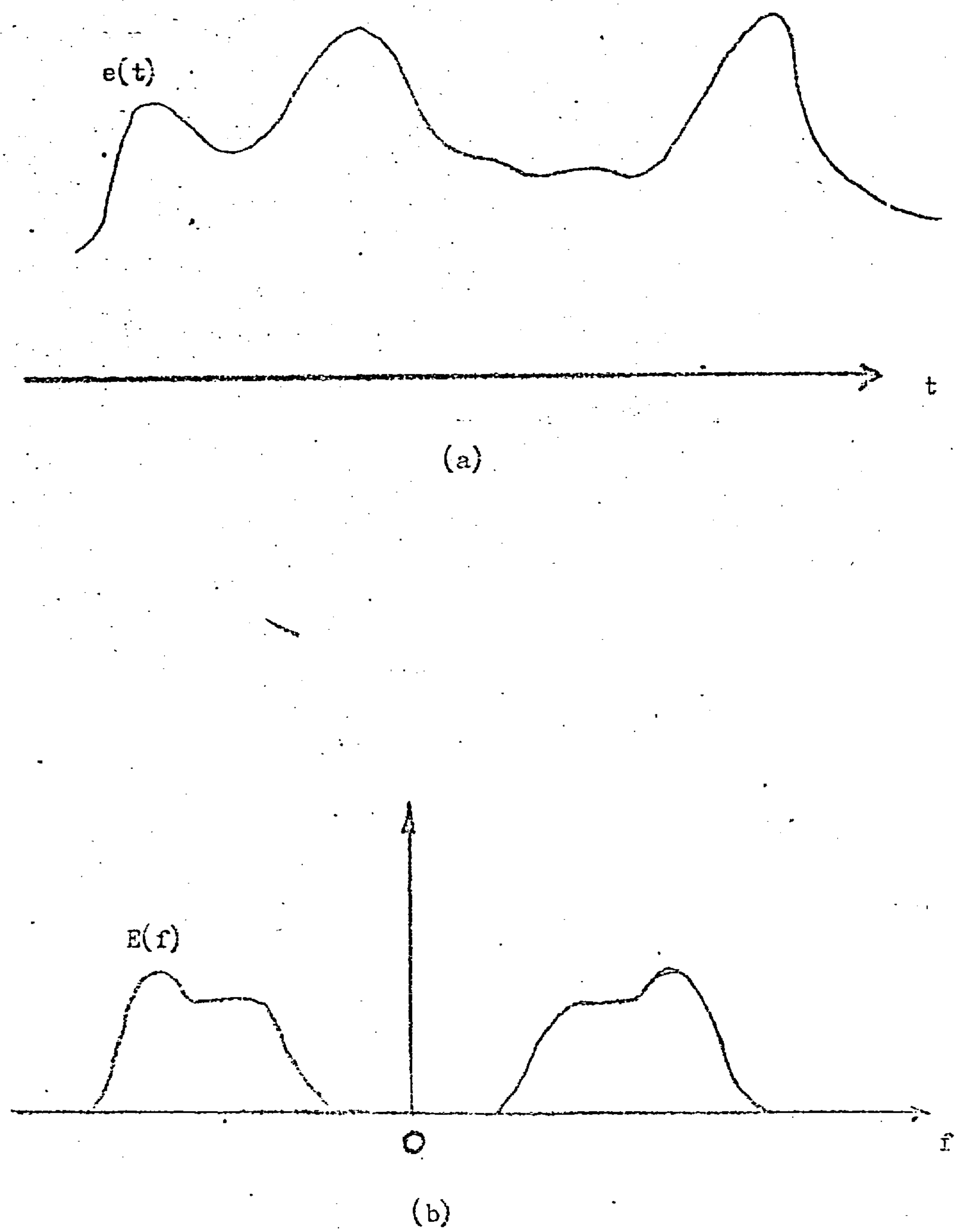


Fig. 2.2 (a) A time function, e(t), and (b) its spectrum, E(f)

output[*] signal is given by the product.

$$G(f) = H(f)E(f) \qquad \ldots (0.4)$$

Hence, if we have a finite scheme of non-overlapping filters, $H_1(f)$, $H_2(f), \ldots\ldots, H_N(f)$, then the corresponding output spectra, $G_1(f)$, $G_2(f), \ldots\ldots\ldots, G_N(f)$, give evidence of the existence of $X_1(f)$, $X_2(f), \ldots\ldots, X_N(f)$, the input spectra. This is the basic principle of a monitoring device which is capable of tracking a flux of many transmissions and presenting their spectra as evidence of their operational existence. Clearly, the simultaneous reception of a finite number of active transmissions is possible by the use of a monitoring device which consists of a bank of filters, each of which is assigned and tuned to a particular frequency range. In addition to the filters, the monitoring device may incorporate some other processing device whereby the intelligence that is transmitted may be identified.

If the presentation of the spectra is made on the screen of a CRT,[**] then observation of the screen will indicate where the transmissions are in the frequency space. However, the presence of noise which combines with the transmissions in some unspecified manner may limit the ability of the observer to detect the exact

---

[*] If the impulse response of the filter is h(t) then the output, g(t), when the input is e(t) is given by the convolution of h(t) and e(t). That is,

$$g(t) = \int_{-\infty}^{+\infty} h(\tau)\, e(t-\tau)\, dt$$

[**] Cathode Ray Tube

number and locations of the various transmissions. It is, therefore, necessary to make a precise study of the noise and then develop appropriate techniques whereby the noise effects may be eliminated or reduced to a tolerable minimum.

2.3 Frequency Spectrum in the Presence of Noise

A simple block diagram of a transmission system is shown in fig. 2.3. Each individual contributor to the overall noise has been isolated for the purpose of clarity and, although signal degradation produced by each could be examined separately, it is the joint effect on the final received signal which is of interest. Since noise is usually a random process it can only be described statistically.

One important statistical characteristic of the noise is its auto-correlation function. If this characteristic is time-invariant, that is, stationary (or quasi-stationary), then it is possible to derive the power spectrum of the noise from the knowledge of the auto-correlation function. The power density spectrum of any signal representing a random process is related to the statistical parameters of that signal and is equivalent to a frequency density distribution of the average power contained in the random process. The power spectrum and the auto-correlation function of the noise are, therefore, related and an expression of this relationship is revealed in the following equations:

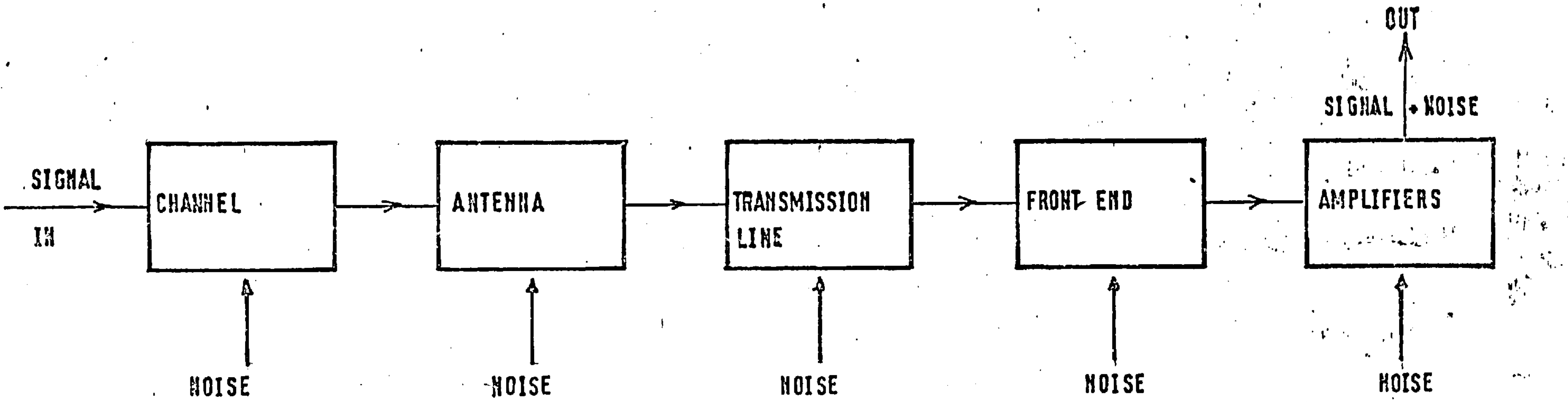


Fig. 2.3. A Block diagram of a conventional receiving system showing the various contributors to the overall noise

$$Q(\tau) = \int_{-\infty}^{+\infty} \phi(f) e^{j2\pi f\tau} \, df \qquad \ldots (0.1)$$

$$\phi(f) = \int_{-\infty}^{\infty} Q(\tau) e^{-j2\pi f\tau} d\tau \qquad \ldots (0.2)$$

Hence, $Q(\tau)$ and $\phi(f)$, the autocorrelation and the power density spectrum, respectively, of the noise are observed to be Fourier transforms of each other. (cf. eqns. 0.1 and 0.2 of Sect. 2.2. The statement in the form of a theorem, has been rigorously established by Wiener and is generally called the Wiener theorem for autocorrelation function).

A device which resolves the radio signals into components in different frequency ranges will also provide information about the frequency distribution of the noise power in all the frequency ranges covered by the total frequency band under investigation.

Since the monitoring device to be employed in the present analysis is essentially a spectrum analyser, the frequency composition of the transmissions and the noise will be presented for analysis. It will be observed that wherever there is no signal, noise may be found. However, because of the randomness of the times of initiation and the locations of the transmissions, the appearance of the signal under investigation may exhibit no definite spatial pattern. Furthermore, the space between the spectral components of the signal will be filled with the evidence of the noise.

It is clearly not possible to ascertain the number of transmissions active within any portion of the hf band at a given time by a simple process of counting because of the presence of noise. It has been suggested[19] in the past that, by measuring the energy content of any portion of the spectrum and comparing this with a pre-specified threshold, it is possible to discriminate between the existence of noise alone and signal plus noise. The pre-specification of the threshold was found difficult because of the various factors involved. In the method of analysis to be described, the possibility of a technique which is insensitive to noise and considers transmission changes only will be examined. In order to make this examination, good receiving systems must be available, and in the next section the desirable qualities of such receivers will be discussed.

2.4 Receivers for Monitoring

The receivers required must be capable of good frequency setting accuracy (better than 15 parts in 10^7) and must be able to accept both AM and FM signals. For the more specialised task, the monitoring device must possess receivers for i.s.b. (i.e. independent side band) telegraphy or telephony. The ability of rapid tuning and minimum waveband switching is also desirable. Special attention must be given to the suppression of receiver oscillator radiation affecting associated receivers either directly or indirectly (e.g. via the aerial system) and also to the suppression of spurious frequencies generated within the receivers. It is also useful to include a receiver input attenuator in order to eliminate spurious inputs caused by high level signals. Receivers with good rf selectivity are also to be preferred.

A block diagram of the type of a receiver to be employed in the present investigation is shown in fig. 2.4 . The aerial loading of the receiver is designed for wide band operation. In addition, there are six double tuned aerial coils one of which can be pre-selected for optimum performance in a chosen wave band. All transmissions whose frequency bands lie anywhere between 0.98 mHz and 30 mHz can be received and analysed in a manner appropriate to the requirements of the search procedure.

The receiver remains tuned to within 50 Hz of the selected frequency under conditions of constant voltage and ambient temperature.

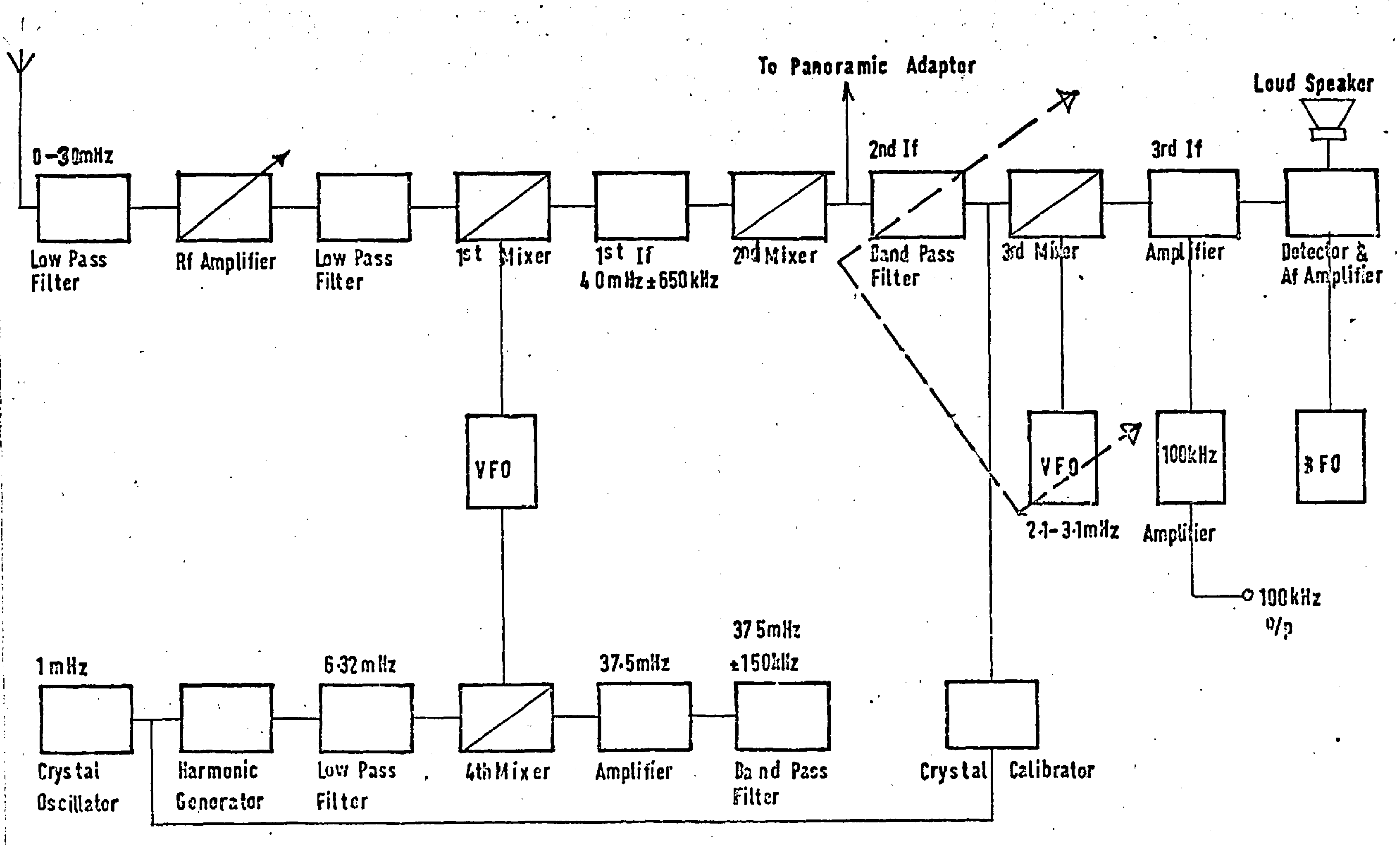


Fig.2.4 BLOCK DIAGRAM OF THE MONITORING RECEIVER

By means of a selector switch, i.f.[x] bandwidths of 13, 6.5, 3.0, 1.2, 0.3 and 0.1 kHz are obtainable. Such a range of selectivity, combined with the 20 db improvement of the signal-to-noise ratio which is made possible in this particular equipment, by the use of an automatic volume control (a.v.c), permits the reception of signals whose strengths are only 1μ volt.

It is also possible to control the levels of exceptionally strong signals when it is likely that such strong signals would cause over-loading in the early stages of the reception operation. Also, strong signals which cannot be rejected sufficiently by tuning the aerial can be suppressed by controlling their levels. There is also noise limiter which reduces the effect of noise peaks exceeding the level of a modulated signal by about 30% .

It is interesting to note that this type of receiver can be used in conjunction with a panoramic adaptor. The use of the panoramic adaptor, another proprietary equipment by Racal Ltd., makes it possible to obtain a visual display of any selected frequency region. The panoramic adaptor is a form of a spectrum analyser and consists of an adjustable narrow band receiver which repetitively scans a fixed frequency range.

---

[x] Intermediate frequency

2.5 Synopsis of the Searching Technique

Using the monitoring receiver in conjunction with the panoramic adaptor, records of a selected frequency region can be obtained and analysed. These records reveal the time-to-time composition of that part of the hf spectrum being examined.

Fig.2.5 is a diagramatic representation of two spectrum records which, for the sake of clarity, are essentially noise-free. A direct comparison of the two records will show which signals are present in one record but not in the other. If one of these records is obtained at time t_1, and the other at time t_2, then any peak (or a group of peaks) in the second record which is not in the first is an indication of a new transmission which has appeared in the time interval, $\{t_1, t_2\}$. In a similar manner, knowledge about the cessation of a transmission can be obtained. Furthermore information about the signal strength of a particular transmission may be obtained by measuring and estimating the appropriate parameters. If noise is present, such an algebraic comparison is impossible. Under these conditions an efficient method of comparing any two spectrum records involves the cross-correlation function of the two records. The reasons for choosing the correlation function as the desired means of comparison will be discussed in chapters 3 and 4.

It will further be shown in chapter 5 that the fundamental problem is the real-time analysis of a stochastic process. It will also be demonstrated that maximisation of the signal posterior probability

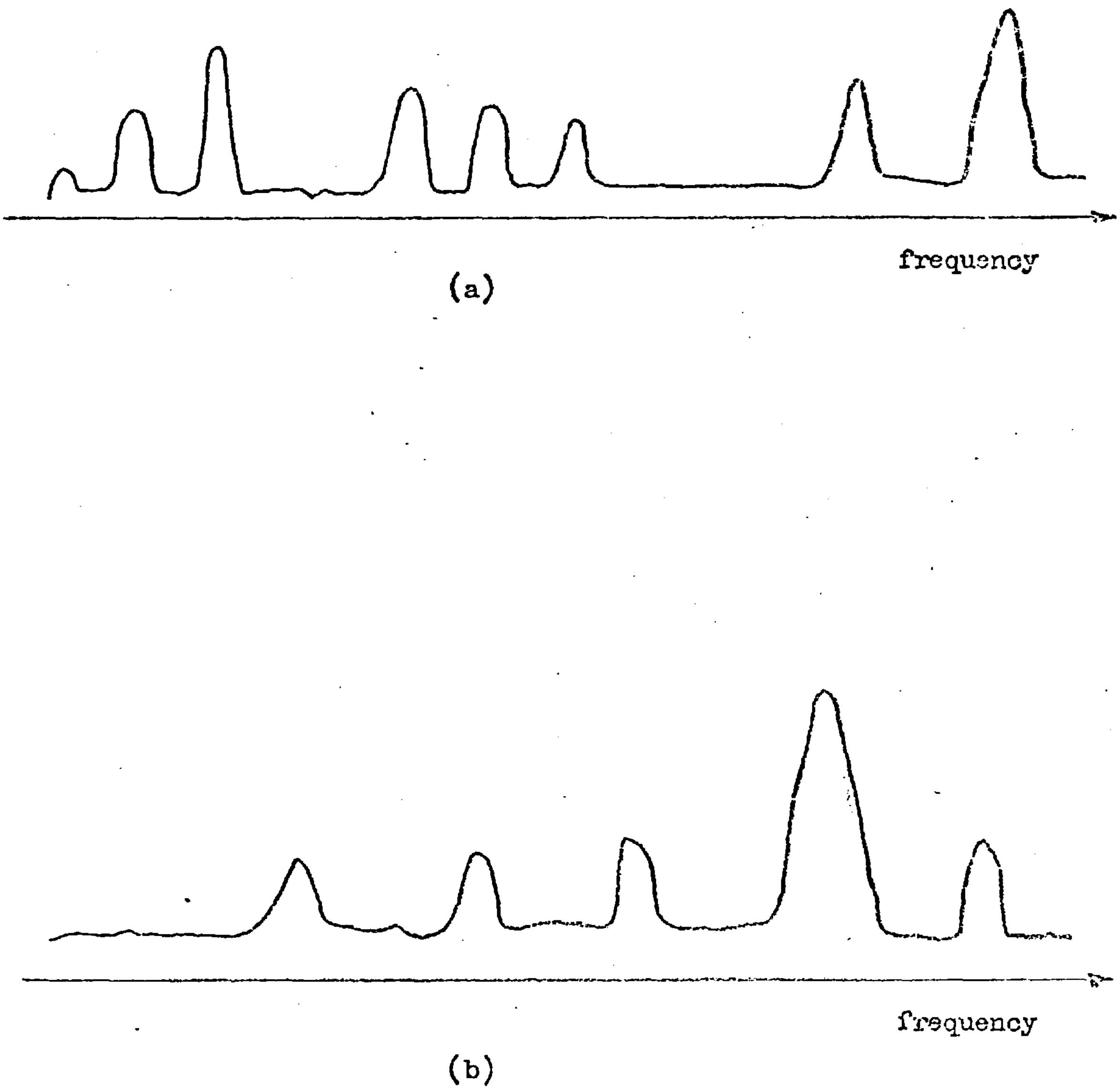


Fig. 2.5 Noise-free spectra of signals located in the same frequency band but occurring at different times.

function is more fundamental to the process of signal detection than the maximisation of the signal-to-noise ratio and that this also leads to the maximum gain [x] of information. The concept of the "distance" between one probability measure and another is introduced and developed in chapter 4.

In chapter 6, the experiments designed to verify the points raised in earlier chapters are described, and the results are discussed in chapter 8 with suggestions for further research.

---

[x] Information about an event is gained when the difference between two types of uncertainties, expressible in terms of some probability measures, is known.

Chapter 3

A METHOD OF SEARCHING FOR SIGNALS LOCATED IN THE FREQUENCY SPACE

Chapter 3 A METHOD OF SEARCHING FOR SIGNALS LOCATED IN THE FREQUENCY SPACE

3.1 Introduction

It has been mentioned in the previous chapters that a number of real-time transmissions signals which are simultaneously active and whose emissions are band-limited* can usually be separated by the use of a frequency analyser. This type of separation is possible if the frequency components of one transmission signal are not coincident with those of another transmission signal. Although it is generally difficult to separate adjacent transmission signals which overlap in the frequency space, techniques do exist for separating and thereby reducing any interference effects caused by the overlapping transmission signals[20]. However, these techniques will not be considered rigorously here.

The present investigation will be confined to the examination of portions of the hf spectrum occupied by non-overlapping transmissions and to the identification of the locations of these transmission signals. Any portions of the frequency space not occupied by a transmission signal will be assumed to be empty and hence available for use by new transmission signals.

Since the time-to-time changes within the hf spectrum are the prime object of the search, apparatus capable of providing records

---

*An emission of a radio signal is said to be band limited when the frequency components of the signal lie within a certain frequency range which is sufficient to ensure the faithful transmission of the necessary information.[21]

of the search space is required. From these records information about the number of active transmission signals and their locations at any one time can then be obtained. Furthermore, a means of detecting the times of initiation, or cessation, of the transmission signals can be readily established by a proper utilisation of the available records.

3.2 The Panoramic Survey

When the Racal panoramic adaptor is employed in conjunction with the radio communication receiver, a visual display of the flux of signals that are picked up in any chosen frequency band is produced on a screen of the panoramic adaptor in the form of a spectral display. It is, therefore, possible to investigate the radio carrier waves which are operating within any selected range of frequencies in the hf band by scanning that portion. The largest portion of the spectrum that can be viewed on the screen of the panoramic adaptor is 1 mHz; and the smallest is about 100 kHz. The band of frequencies which is selected for examination can be scanned at one of two rates. A sweep time of 2 seconds is employed when scanning wide bands, but a sweep time of .2 seconds is sufficient for small bandwidths of 1000 kHz and less.

The spectral picture produced on the screen of the panoramic adaptor will generally provide some evidence of the frequencies which are being employed at any one time for the transmission of messages. The various characteristics of each spectrum must be examined in order to obtain the necessary information about the types of transmission signals that are operating.

As an example of a typical spectrum that can be produced on the screen of the panoramic adaptor, it is always helpful, for the purposes of simplicity, to consider the line spectrum. A schematic representation of a line spectrum is shown fig.3.1. From such a model it is relatively easy to find the characteristic* frequency or frequencies of the transmission signals operating in any particular band. In this picture, the part of the transmitted power which is received at a given frequency is represented by a line in the spectrum, and the vertical height of the line above the zero mark is a relative measure of the received power at that frequency. A careful study of the amplitude variations and the position variations of a line in the spectrum will provide valuable information about the types of transmission signals which are operative at any given time. (In some modes of transmission the carrier frequency is suppressed and the existence of the transmission signal must be determined, therefore, by the detection of some other characteristic frequency component such as the side bands of the transmitted signal). Also, the various initiation times of new transmission signals can be estimated by

---

*A characteristic frequency is the frequency which can easily be identified and measured in a given emission.[22]

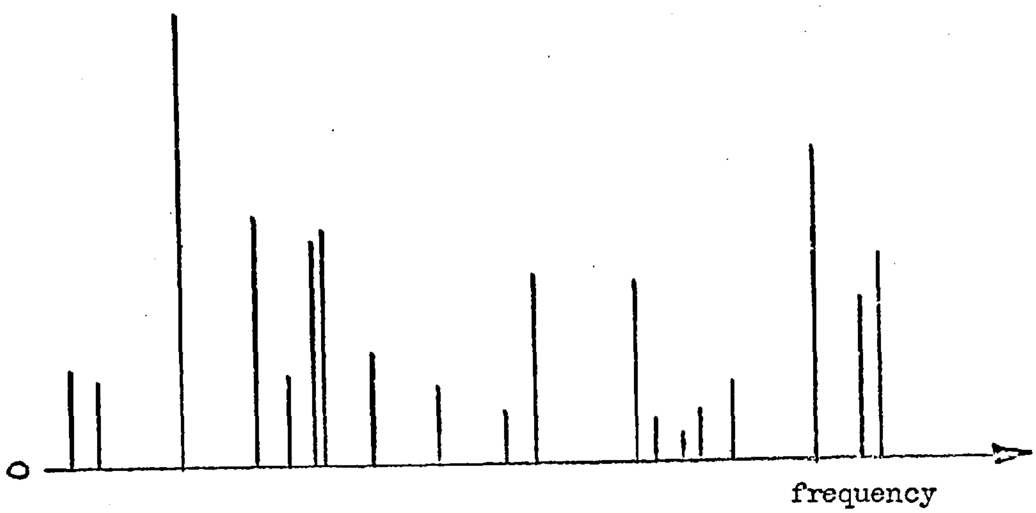


Fig. 3.1 An example of a line spectrum

comparing any two scan pictures during some interval of the search period.

By making a comparative study of the number of carriers within a given bandwidth at various instants of time, knowledge about the number* of transmission signals per some unit of the frequency space and changes in this number which result from the appearance of new transmission signals or the disappearance of old ones will be obtained. For example, if the traffic densities at times, t_1 and t_2 $(t_2 > t_1)$, are TD_1 and TD_2, respectively, then the difference, $TD_2 - TD_1$, will be used to indicate the transmission changes and the time interval, $[t_1, t_2]$, will determine the time in which the changes occur. Again, in order to ascertain the characteristic behaviour of all transmission signals occurring within a given band, it is important to try and recognise any relationship that may exist between them. The parameters, in terms of which these relationships can be expressed, will then provide information about the types of transmission signals that are likely to be detected simultaneously. (It may commonly occur that amplitude modulated (AM) transmission signals are found in one particular portion of the hf band and frequency modulated (FM) signals in another.) Obviously, knowledge about the manner in which the transmission signals have been distributed within the band will also be invaluable in determining the locations of the transmission

*This number of transmission signals per unit frequency space will be known as the traffic density. The unit of frequency space which will be adopted will be determined generally by the nature of examination being carried out and in particular by the type of resolution required.

signals, and in this investigation, a study of the characteristics of the different portions of the search space will be undertaken.

3.3 An Example of a Real-Time Spectrum

The real-time spectrum(fig.3.2)that can be produced on the screen of the panoramic adaptor is seldom like the one shown in fig.3.1. This is because noise and fading tend to make the detection of the wanted signals an uncertain process. Hence, from a real-time spectrum it is only possible to obtain an incomplete knowledge about the population bands of frequencies and the signal amplitudes that are transmitted; and because of this, the task of parameter estimation which has been briefly described above becomes a problem of great complexity.

The problem of accurate parameter measurement and the correct assessment of signal behaviour within a given band of frequencies can be solved if the major sources of interference can be eliminated. Indeed, most problems in communication engineering are often complicated by the presence of man-made and natural interfering phenomena. One obvious solution, therefore, involves a means of assessing the level of noise and making sure that the signal energy is well above the noise level. However, in the present investigation, evidence of the existence of weak signals will also be required in order to ensure that an 'empty space' in the hf band contains nothing but noise.

Since the problem of assessing the level of the noise and the

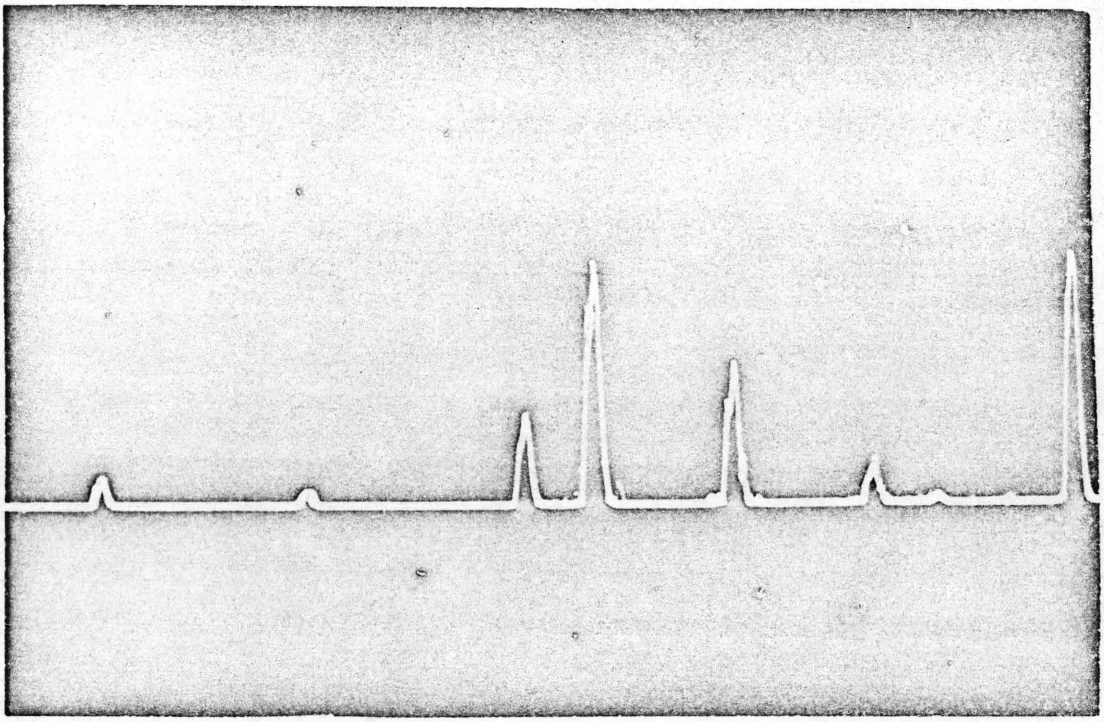


Fig. 3.2 An example of a real-life spectrum produced by an artificial spectrum generator.

other interfering phenomena can best be described in a statistical sense, it has proved useful, at least in the present investigation, to combine results derived from transmission theory and those obtained from the theory of random point estimation. For example, by using the techniques of diversity reception and employing means of computing error probabilities, the problem of correct parameter measurement may be simplified and solved.

3.4 An Operational Search Technique

The primary object of the present analysis is the finding of a model for a searching system which is capable of identifying the locations and the duration times of a number of transmission signals existing within a specified portion of the hf spectrum in the presence of noise. The overall function of the system can be divided into three distinct operational sections:

1) the function of finding the frequency of a transmission signal;

2) the function of assessing the duration times, and

3) the function of determining the number of transmission signals occupying a selected portion of the hf band. It is obvious that each section of the whole model is capable of independent operation and of providing an answer to a specific question about the nature of that part of the hf band which is being examined at any given time. For

instance, section 1) will give the required information about the characteristic frequencies of any number of the transmission signals under examination. However, because of the constant presence of the various forms of interference, especially noise and fading, the functional behaviours of each section must take account of the probabilistic nature of the interference phenomena. Only when this is done will the system be capable of providing correct answers a high percentage of the time to any of the questions concerning, for example, the duration times of the signals. It will also be useful to assign a probability measure to the performance of the proposed model in terms of the success rate at which correct answers are given.

3.5 Distribution of Transmission Signals within the Search Space

At any given time, the loss[x] of any number of transmissions in a given bandwidth tends to change the arrangement of the transmission signals within the search space. For example, if the configuration of the search space is characterised by the number of transmission signals present and the locations of these signals, then changes in the configuration of the space are obviously caused by the corresponding changes in the locations and the number of transmission signals.

---

[x]The advent of new transmission signals will indicate a "gain" of signals within the band of frequencies being examined.

It will be shown presently that by measuring certain invariant or semi-invariant properties of the search space and continually comparing one set of these properties with another set of similar properties, evidence of the occurrence of new transmission signals or the disappearance of old ones will be obtained. Also, by a proper utilisation of the properties a determination of the times of initiation or the disappearance of transmission signals can be made. It is important to note that the invariant properties of the search space to be considered are those which have some functional relationship with the number and locations of the transmission signals within the search space.

3.6 Basis of the Comparative Study of Spectra

Any portion of the hf spectrum which is examined at any given time can be divided into a number of frequency space intervals each of which will be referred to as a "band slot" or a "cell". The band slots making up the search space are scanned successively in time and the decision on the presence or the absence of a transmission signal in the i-th band slot is made on the basis of the observations made and stored during some previous time interval. This stored information can usefully exist in the form of, or be represented as, a wave-form, f_i. (The wave-form, which will be discrete, will be obtained by scanning and sampling a selected portion of the hf spectrum which is being examined). The analysis is begun by comparing the wave-form, f_i, in some special manner with another wave-form, f_{i+k}, which is obtained by scanning and sampling the same band slot but at a later time. Here, k (= 1, 2, ...) is an index of the order of the scan and k=1 indicates that f_{i+1} is the waveform sampled immediately after f_i.

The waveform, f_{i+k}, (k = 1,2,....) which are compared with f_i are rank-ordered in a fashion which suggests their similarity (or dissimilarity) with f_i. All waveforms which possess a pre-specified amount of similarity with f_i are accepted as being identical with f_i. The various properties of the waveform in terms of which similarity is measured can be regarded as some dimensions of a space in which a point, designated as f_i, is located. Obviously, each waveform is a point in the space and each dimension of the space is an expression of a property of the waveform. If the waveform is specified by H distinct properties and is, therefore, a point in an H-dimensional space, then the coordinates, a_1, a_2, ... a_H, of the point have the numerical values which correspond to the amount of each property of the waveform. The set of points which belong to a particular class and, therefore, exhibit definite similarity among themselves correspond to an ensemble of points within some particular small region of the space. Another set of points will cluster in some other region of the space, and the difference between classes of points will be expressed in terms of the "distance" between clusters. For example, if f_i and f_j belong to two different classes then the distance between them, expressed quantitatively in some suitable manner, will determine the difference in their properties. Hence the present analysis will be restricted to the determination of differences between observations which are represented by the waveforms. These differences, as will be seen later, depict changes that occur during the various scan periods.

It has been found useful in this investigation to use the correlation function of f_1 and f_2 as the desired measure of the difference between them. The important point to note is that even in the presence of noise, fading or both, any two waveforms will be functionally identical only if they contain the same number of transmission signals which are operating on the same frequencies. However, when the transmission signal contents of the two waveforms are different from the point of view of the signals' locations and or their numbers, a correlation function of the two waveforms will show their lack of similarity. The type of correlation function to be employed in this investigation will, therefore, provide a running indication of the similarity between, say, f_1 and f_2. The problem, therefore, consists of obtaining the correlation function which will be used to separate dissimilar waveforms and at the same time to group waveforms which are similar. The correlation function to be used should take into account only those properties of the waveforms which are relevant to the transmission changes. For simplicity, it has always been assumed[23,24] that the sources of noise within any frequency range are independent, and that the signal-to-noise ratio, although unspecified, remains substantially unchanged, or at best, changes only slowly. With this assumption, the mathematical analysis in this investigation will become tractable.

The correlation of random functions is encountered in communication problems on many occasions[25-27]. It is generally used to express the similarity or dissimilarity between one group of events and another. Also, it allows a realistic ordering of a number of events according to the degree of similarity. For example, the cross-correlation between the output and the input of a linear system provides information regarding the amount of the input message that is retained in the output message. In radar systems,[28] the received echo is compared with an attenuated replica of the transmitted signal, and the desired information concerning the range or the bearing of a reflecting object may be ascertained. A process of iterated autocorrelation has also been used[29] to determine that frequency which contains a greater energy density than any other frequency.

However, there is a fundamental difference between the conventional application of correlation techniques and that presented in the present investigation. For example, in radar and other data transmission problems where correlation techniques are employed, an output event is compared with a known and a readily available set of events and the similarity of an output event to anyone of the known events is computed. The similarity is said to be great when there is a high correlation between the output event and a known event. In the present analysis, however, all the events which are available for comparison are output events and the correlation technique is used to obtain information about some common properties that are possessed by them. For this reason, any two waveforms are said to be identical if they contain the same transmission signals (which are located within the same frequency bandwidth) in spite of the presence of the various effects of noise and the other interfering phenomena. Such identical waveforms will "correlate" well not when the quantitative value of the correlation is high but when the correlation function exhibits some special features which will be discussed later in the thesis.

3.7 The Correlation Method of Search

By retaining records of the various scan pictures, or time histories, as they are sometimes called, an efficient means of detecting changes in the number of transmission signals can be provided. In addition, the locations of new transmissions which have occurred or old ones which have ceased operation within the frequency bandwidths under an examination can be ascertained. It is the changes within the search space, caused by the advent of new transmission signals, or the disappearance of old ones, that are of greatest interest in the present analysis. Hence any two scan pictures, or time histories, of the same spectrum will be identical if they contain the same transmission signals which are located at the same frequency positions within the search space. Transmission signals which appear in one time history and not in a subsequent one are regarded as "missing" signals and the application of the correlation method of search will, in a manner to be described, lead to a gain of information concerning the

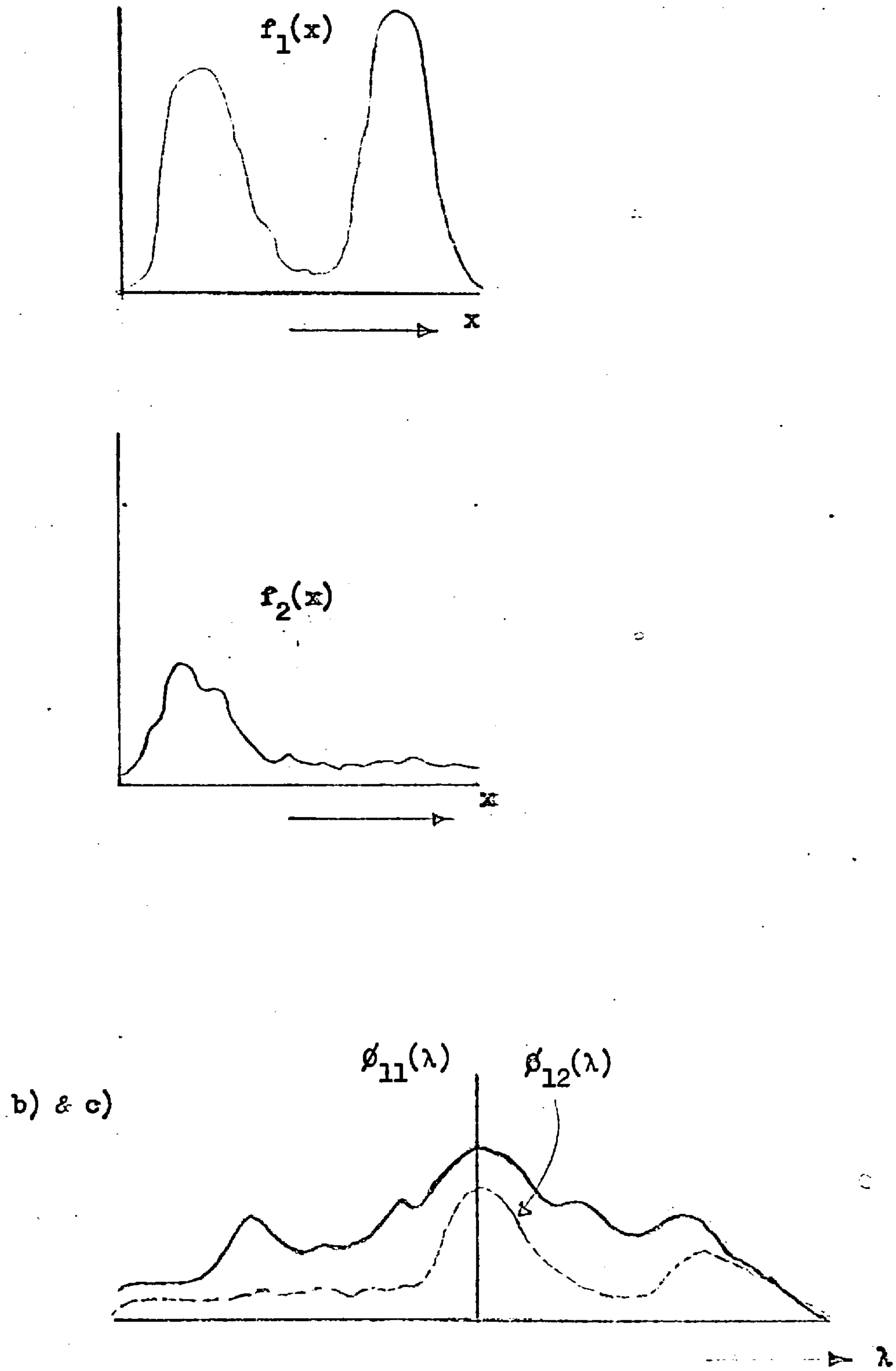


Fig. 3.3 Two functions, f_1 and f_2, and their cross-correlation function, ϕ_{12}

times and the locations of the "missing" signals. Similarly, transmission signals which appear in a later time history and not in an earlier time history are said to be the newly initiated transmission signals.

Briefly, the basis of the technique is that if two time histories which are correlated with each other are identical (or nearly so on the basis of some prescribed set of rules), then the resulting space function will possess the distinctly symmetrical features of an autocorrelation function. If the space function that is generated is $\bar{\underline{g}}(\lambda)$ where the space argument, λ, signifies frequency, then its symmetrical features will be revealed by the fact that

$$\bar{\underline{g}}(\lambda) \quad = \bar{\underline{g}}(-\lambda)$$

with respect to some appropriately chosen axis. On the other hand, two non-identical time histories will produce a correlator output function which is non-symmetrical. These ideas are illustrated in fig.3.3. In fig.3.3c, an autocorrelation function of f_1 is shown; fig.3.3b is the cross-correlation function of f_1 and f_2. Since f_1 and f_2 are not identical, the cross-correlation function is not symmetrical and, as will be expected, is different from the autocorrelation function. Hence, the difference between the two types of the correlation function can be used to discriminate between two time histories of a search space which are not identical,

3.8 Construction of the Correlation Technique

3.8.1 Introduction

If the correlation technique of processing any two observations of the search space is to be useful, then it should provide an efficient means of detecting transmission changes in that portion of the spectrum which is under examination. That is, the changes that will be looked for are those due solely to the advent of new transmissions or the cessation of old ones. It has been stated, in section 3.5 that two identical, or nearly identical, time histories which contain the same transmission signals will, after the appropriate processing, produce a correlator output function which is symmetrical. It has also been stated that a non-symmetrical function will be the output of the correlator when two non-identical time histories are processed.

3.8.2 A Mathematical Model of the Technique

A function such as a(x) can be defined in a way to describe the relevant characteristics of any one transmission signal which is operating in the search space. The argument, x, covers a range over which the frequency components of the transmission signal are distributed.

In this range of x, $a(x)$ may be assumed to be continuous, or piece-wise.[x] continuous. But in situations where the transmission signal may be characterised only by its carrier frequency, $a(x)$ becomes a Dirac function and in such a case any two transmission signals with carrier frequencies at X_1 and X_2 Hz, respectively, may be represented by

$$m(x) = a_1\delta(x - X_1) + a_2\delta(x - X_2) \; \ldots\ldots \; (2.1)$$

In the general situation where the carriers and the other characteristic frequencies - such as the side band frequencies - are present, $a(x)$ will be assumed continuous or piece-wise continuous over some range of x where the whole of the transmission signal is distributed. The transmission function then takes the form

$$m(x) = \sum_{i=1}^{K} a_i(x - X_i) \qquad \ldots\ldots \; (2.2)$$

when there are K transmission signals in the space and a_i is the amplitude of the i-th transmission signal; x covers the range of the space where the i-th transmission is distributed.

---

[x]Intuitively, the notion of continuity of a function at a given point means that the value of the function throughout a neighbourhood of the point will differ from its value at the given point by as little as desired if the neighbourhood is sufficiently small. A function is said to be piece-wise continuous if it is only continuous in a finite number of intervals obtained by dividing a given interval with a finite number of successive points.

In the absence of any transmission signals in the search space, the space configuration can be represented by

$$f(x) = n(x) \qquad \dots (2.3)$$

where $n(x)$ represents the interfering noise. If the action of the noise is purely additive, then the function of the space configuration when both noise and transmissions are present is, from equations (2.2) and (2.3)

$$f(x) = \sum_{i=1}^{K} a_i(x - x_i) + n(x) \qquad \dots (2.4)$$

It follows from equation (2.4) that

$$f(x + \lambda) = m(x + \lambda) + n(x + \lambda) \qquad \dots (2.5)$$

Multiplying (2.4) by (2.5)

$$f(x)f(x+\lambda) = \sum_{i=1} \sum_{j=1} a_i(x - x_i)a_j(x + \lambda - x_j) + n(x)n(x + \lambda) +$$

$$+ \sum_{i=1} a_i(x - x_i)n(x + \lambda) + \sum_{i=1} a_i(x + \lambda - x_i)n(x + \lambda) \qquad \dots (2.6)$$

The correlation function, $\phi(\lambda)$, then becomes

$$\phi(\lambda) = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(x)f(x + \lambda)\,dx$$

$$= \lim_{T \to \infty} \left[\frac{1}{T}\int_0^T \sum_i \sum_j a_i(x-x_i)a_j(x+\lambda-x_j)dx + \frac{1}{T}\int_0^T n(x)n(x+\lambda)dx + \right.$$

$$\left. + \frac{1}{T}\int_0^T \sum_i a_i(x-x_i)n(x+\lambda)dx + \frac{1}{T}\int_0^T \sum_i a_i(x+\lambda-x_i)n(x)dx \right] \qquad \dots (2.7)$$

Now, interchanging the order of integration and summation,

$$\phi(\lambda) = \lim_{T \to \infty} \left[\sum \sum \frac{1}{T} \int_0^T a_i(x-x_i)a_j(x+\lambda-x_j)dx + \frac{1}{T}\int_0^T n(x)n(x+\lambda)dx \right.$$

$$\left. \sum \frac{1}{T}\int_0^T a_i(n-x_i)n(x+\lambda)dx + \sum \frac{1}{T}\int_0 a_i(x+\lambda-x_i)n(x)dx \right] \qquad \ldots (2.8)$$

The meaning of equation (2.8) can be visualised by referring to fig. 3.4, in which there are two functions representing noise and a transmission signal. The third and fourth terms of the right-hand side

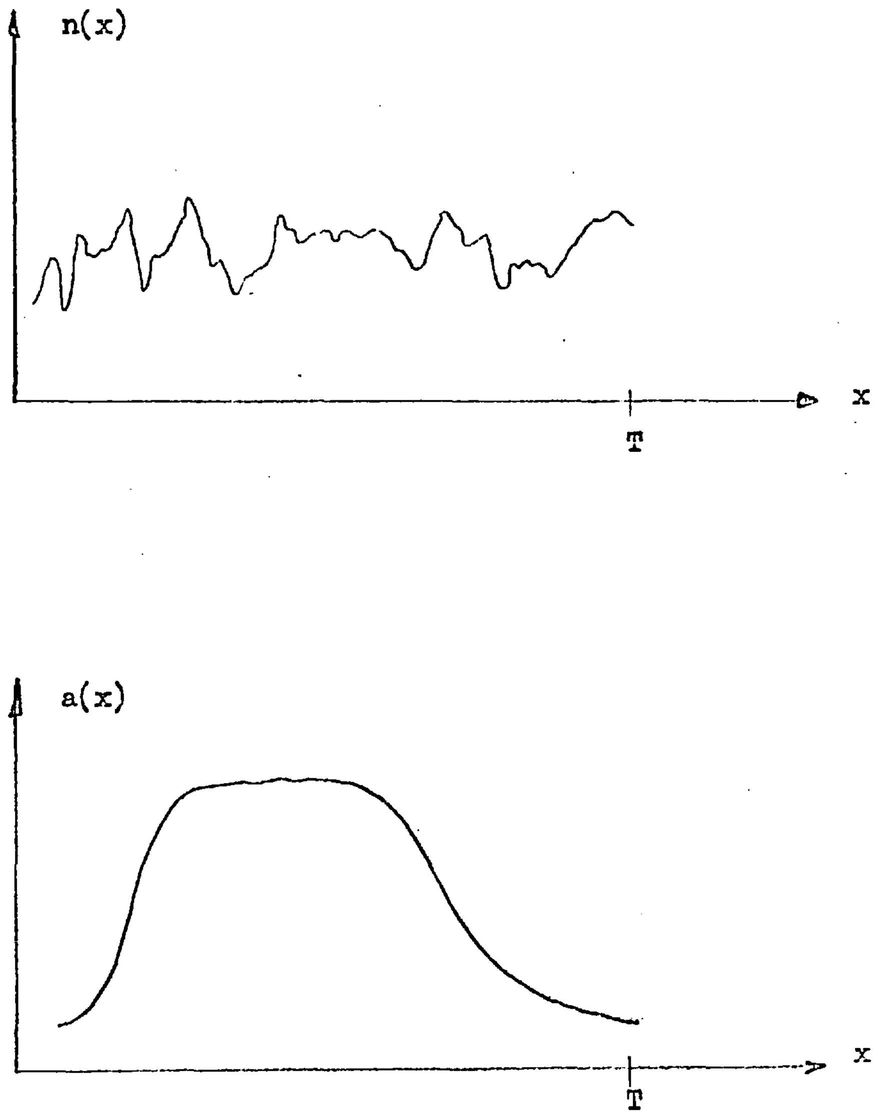


Fig. 3.4 A sketch of two functions illustrating
a) noise signal and
b) transmission signal

of equation 2.8 contains the sums of terms, each term being the cross-correlation of noise and a transmission signal. Referring to fig. 3.4, this correlation of noise and a transmission signal can be visualised as a process involving multiplication and addition of two space functions. If the noise, $n(x)$, and the transmission signal, $a_i(x)$, are uncorrelated, then

$$\frac{1}{T}\int_0^T a_i(x-x_i)n(x+\lambda)\,dx = \frac{1}{T}\int_0^T a_i(x+\lambda-x_i)n(x)dx$$

$$= 0 \qquad \ldots (2.9)$$

and the required information about the space configuration is derived only from the first two terms of equation 2.8 . Clearly, the first term is the auto-correlation of the various transmissions, $\{a_i\}$ and will be a symmetrical function of the argument, λ. Under the assumption* that the noise is a stationary, random process but with an unspecified amplitude distributions,

$$\frac{1}{T}\int_0^T n(x)\,n(x+\lambda)dx$$

is also an even, symmetrical function of λ. Hence the right hand side of equation 2.8 becomes the sum of two space functions which are both symmetrical about the same axis. It can, therefore, be concluded that $\phi(\lambda)$ is an even symmetrical function when the noise and the transmission signals are uncorrelated, a condition which is generally met in the search space.

---

* The assumption is that the noise and the transmission signal are "uncorrelated".

If the assumption concerning the correlation between the noise and the transmission function is relaxed, then

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T a_i(x-x_i)n(x+\lambda)\, dx = u \qquad \ldots (2.10)$$

and

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T a_i(x+\lambda-x_i)n(x)dx = v \qquad \ldots (2.11)$$

where both n and v may be non-zero, ~~and they~~ and will represent the amount of correlation between the transmission signal function, $a_i(x)$, and the noise[++] function $n(x)$. If T (cf. fig. 3.4) is sufficiently large so that the law[x] of large numbers is applicable, then u and v will be each equal to a constant within the correlation range (i.e. the allowed value of λ). This is made clear by referring to fig. 3.4 and observing that the operations involved in the computation of the quantities are as follows:

i) Multiply $n(x)$ by $a(x)$ point by point

ii) Add the products

iii) Take the average

iv) Translate $a(x)$ or $n(x)$ by amount, λ, and repeat operations i) to iii).

[++] It will be assumed that the statistical structure of noise function, $n(x)$, will remain substantially fixed from scan to scan of the same frequency bandwidth. It will however be different for the different portions of the whole hf band.

[x] Simply, this law states that the time and distribution averages of random variables are one and the same thing.(30-31)

Thus if the law of large numbers holds, then the results of step (iii) will be a constant or nearly so for all the allowed translations, λ. In conclusion, it can be said that the symmetry of $\phi(\lambda)$ is unaffected by the amount of correlation between the noise waveform and the transmission function because the quantities, u and v, are effectively constant.

The consequence of the above results can now be applied to the analysis of two observations $f_1(x)$ and $f_2(x)$, of the search space. $f_1(x)$ and $f_2(x)$ are defined in the following manner:

$f_1(x) \triangleq$ the time history of the search space at time t_1

$f_2(x) \triangleq$ the time history of the search space at time t_2

where $t_2 > t_1$.

The model function of the space configuration then becomes

$$f_1(x) = \sum_{i=1}^{K} a_{1_i}(x - x_i) + n_1(x) \qquad \dots (2.12)$$

and

$$f_2(x) = \sum_{i=1}^{K'} a_{2_i}(x - x_i) + n_2(x) \qquad \dots (2.13)$$

Proceeding in the same way as for equation (2.6),

$$\phi_{12}(\lambda) = \lim_{T \to \infty} \left[\frac{1}{T} \int \left\{ \sum_i a_{1_i}(x-x_i) + n_1(x) \right\} \left\{ \sum a_{2_i}(x+\lambda-x_i) + n(x+\lambda) \right\} dx \right]$$

$$\dots (2.14)$$

Expanding and re-arranging equation (2.14)

$$\phi_{12}(\lambda) = \lim_{T\to\infty} \sum_{i=1} \sum_{j=1} \frac{1}{T}\int a_{1_i}(x-x_i)a_{2_j}(x+\lambda-x_i)dx + \frac{1}{T}\int n_1(x)n_2(x+\lambda)dx$$

$$+ \sum_i \frac{1}{T}\int a_{1_i}(x-x_i)n_2(x+\lambda)dx + \sum_{i=1} \frac{1}{T}\int a_{1_i}(x+\lambda-x_i)n_1(x)dx \qquad \ldots (2.15)$$

If the noise is uncorrelated with the transmissions then the last two terms of the right-hand side of eqn.2.15 will be zero. However, any amount of correlation between the noise and the transmission functions will make

$$\frac{1}{T}\int_0^T a_{1_i}(x-x_i)n_2(x+\lambda)dx \text{ and } \frac{1}{T}\int_0^T a_{2_i}(x+\lambda-x_i)n_1(x)dx$$

non-zero. But as has been shown earlier, these two quantities are constants and hence will not affect the symmetry of the correlation function, $\phi_{12}(\lambda)$.

The second term on the right-hand side of equation (2.15) expresses the cross-correlation of $n_1(x)$ and $n_2(x)$. Hence, if $n_1(x)$ and $n_2(x)$ are independent random stationary processes with identical distributions, then

$$\frac{1}{T}\int_0^T n_1(x)n_2(x+\lambda)dx$$

will be an even function of the argument, λ, and will not affect the symmetry of the final function, $\phi_{12}(\lambda)$. Therefore, the function, $\phi_{12}(\lambda)$, will be symmetrical or non-symmetrical about the chosen axis if and only if the first term,

$$\sum_{i=1} \sum_{j=1} \frac{1}{T} \int_{0}^{T} a_{1_i}(x-x_i) a_{2_i}(x+\lambda-x_j)dx$$

is symmetrical or non-symmetrical. Clearly, if the statistics of the transmission signals, $\{a_{1_i}\}$ and $\{a_{2_j}\}$ are the same so that the distributions of the transmission signals in the two independent scan periods are identical, then the first term is also a symmetrical function and $\phi_{12}(\lambda)$ is, therefore, symmetrical. If on the other hand, new transmissions occur or old ones disappear, then the distribution of the transmissions will change correspondingly with the result that the first term of $\phi_{12}(\lambda)$ will not be symmetrical. Hence, a test for symmetry can be the basis of the search technique. In other words, the test statistic is one which reflects the fact that the correlation function, $\phi_{12}(\lambda)$, is symmetrical or not. If $\phi_{12}(\lambda)$ is symmetrical then the transmission distributions in the two observations of the same spectrum are the same and no change in the number of transmissions has thus occurred. A change in the number of transmissions and the locations of transmission signals will be indicated by the fact that the function, $\phi_{12}(\lambda)$, will be non-symmetrical.

In order to fully exploit the advantages of the correlation technique, it will be useful to construct the two types of the correlation functions, the auto- and the cross-correlation and compare them. Thus, having obtained the two wave-forms, f_1 and f_2, of any two observations, the two types of correlation functions will be formed by correlating f_1 with itself to obtain $\phi_{11}(\lambda)$ and f_1 with f_2 to obtain $\phi_{12}(\lambda)$. The difference between the two correlation functions can then be utilised to discriminate between the transmission contents of f_1 and f_2. These ideas have been illustrated in fig. 3.5 below

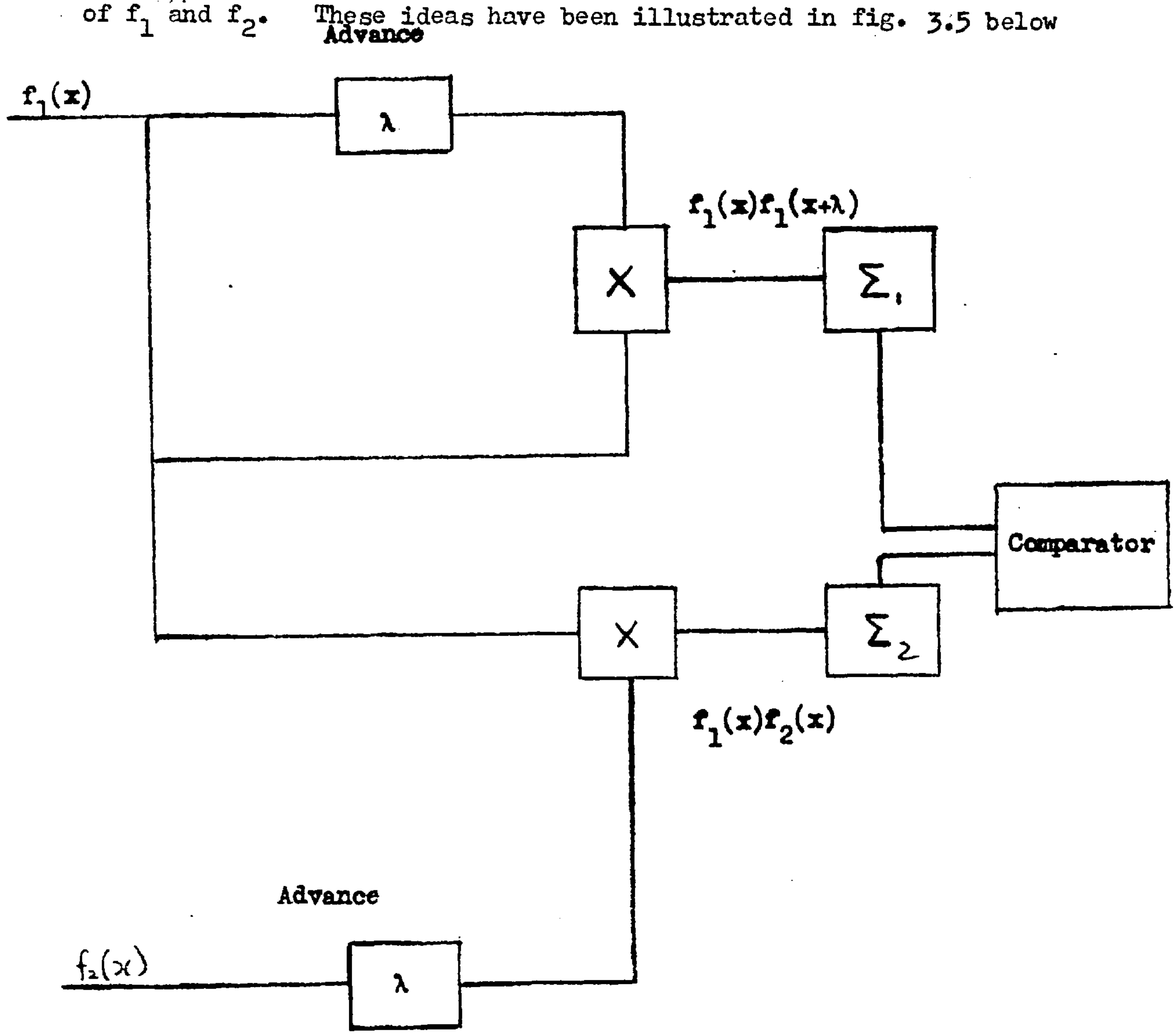


Fig. 3.5 A comparison between two observations, $f_1(x)$ and $f_2(x)$, using correlation technique

The function, $f_1(x)$ is multiplied by a delayed (or an advanced) version of itself and then by a delayed version of $f_2(x)$. The amount of delay, λ, is pre-determined and related to the rate of sampling of the functions, $f_1(x)$ and $f_2(x)$. The products, $f_1(x)f_1(x+\lambda)$ and $f_1(x)f_2(x+\lambda)$ are summed separately in the blocks marked, $\sum_1$ and $\sum_2$. Finally the comparator is used to compare the differences between the two sums.

3.9 A Method of Difference-Correlation

If f_1 is correlated with $f_1 - f_2$, the resulting function is

$$f_1 \quad (R) \quad (f_1 - f_2) = \phi_{11}(\lambda) - \phi_{12}(\lambda) \qquad \ldots (0.1)$$

where the sign, (R), denotes the correlation operation. Hence, by postponing the correlation operation until the difference, $f_1 - f_2$, has been obtained, it will be possible to generate a correlation function which contains the two correlation functions, the auto-correlation and the cross-correlation function. Clearly, if

$$f_1 = f_2$$

then

$$\phi_{11}(\lambda) = \phi_{12}(\lambda) = \phi(\lambda)$$

so that

$$\phi_{11}(\lambda) - \phi_{12}(\lambda) = 0 \qquad \ldots (0.2)$$

Zero is trivially a symmetrical function, and the result of correlating

f_1 with the difference, $f_1 - f_2$, clearly satisfies the symmetry condition which must be met if the two observations, f_1 and f_2, contain the same transmission signals.

When $f_1 \neq f_2$, the difference, $\phi_{11}(\lambda) - \phi_{12}(\lambda)$, can be obtained by combining equations 2.7 and 2.15 of Section 3.8. Thus,

$$\phi_{11}(\lambda) - \phi_{12}(\lambda) = \lim_{T \to \infty} \left[\sum_i \sum_j \frac{1}{T} \int_0^T a_{1_i}(x-x_i) a_{1_i}(x+\lambda-x_i)dx \right.$$

$$- \sum_i \sum_j \frac{1}{T} \int_0^T a_{1_i}(x-x_i) a_{2_j}(x+\lambda-x_j)dx$$

$$+ \frac{1}{T} \int_0^T n_1(x) n_1(x+\lambda)dx - \frac{1}{T} \int_0^T n_1(x) n_2(x+\lambda)dx$$

$$+ \sum_i \frac{1}{T} \int_0^T a_{1_i}(x-x_i) n_1(x+\lambda)dx +$$

$$- \sum_i \frac{1}{T} \int_0^T a_{1_i}(x-x_i) n_2(x+\lambda)dx +$$

$$+ \sum_i \frac{1}{T} \int_0^T a_{1_i}(x+\lambda-x_i) n_1(x)dx$$

$$- \sum \frac{1}{T} \int_0^T a_{2_i}(x+\lambda-x_i) n_1(x)dx \qquad \ldots (0.3)$$

As previous results have already shown, the symmetrical features of equation (0.3) will be dependent upon whether or not the first two terms are symmetrical. The case when a_1 and a_2 are the same has been treated already.

However, when $a_1 \neq a_2$ it can be shown quite easily that the symmetry of $\phi_{11}(\lambda) - \phi_{12}(\lambda)$ will be destroyed. Hence, if the function, $\phi_{11}(\lambda) - \phi_{12}(\lambda)$, is not symmetrical about the chosen axis, then it can be concluded that $a_1 \neq a_2$ and that either new transmissions have occurred or old ones have ceased operation.

The difference, $f_1 - f_2$, may contain additional information concerning the locations of the transmissions which have either appeared or ceased operation in the search space. This is seen by referring to figs.3.6a and 3.6b where two noiseless scan pictures of the one spectrum are shown. The transmission signal in fig.3.6b which does not appear in fig.3.3b is marked (A) and can be isolated by 'subtracting' the function represented in fig.3.6a from that in fig.3.6b However, since any time history contains a noise component whose instantaneous power can considerably affect the desired output of a differencer, the limits of the usefulness of a differencer will be realised. In fact, using the same notation as in equations (2.2) and (2.3)

$$f_1 - f_2 = (m_1 - m_2) + (n_1 - n_2) \qquad \ldots (4)$$

From this last equation, it is clear that if the power density of the noise is much greater than the signal power, then the transmission signals will generally be completely masked by the noise and no information about them may be gained. This implies, therefore, that a

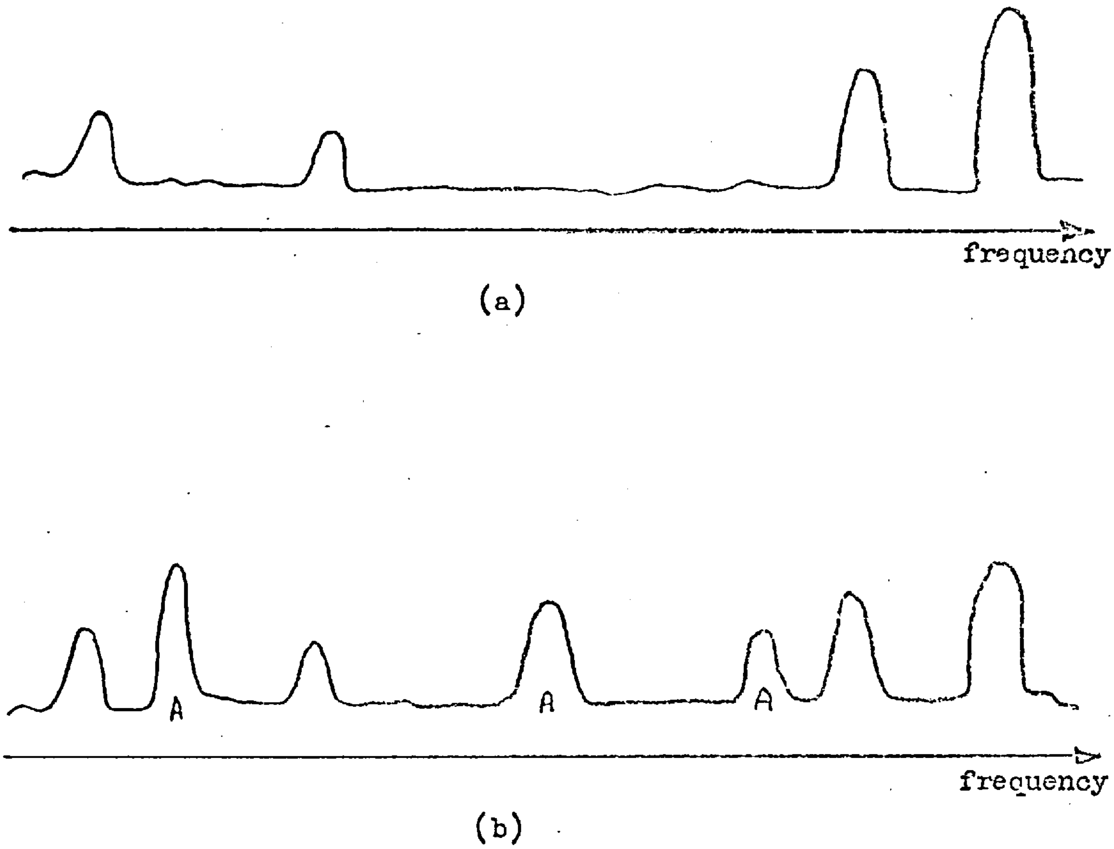


Fig. 3.6 A diagram illustrating the emergence of "new" transmission signals

differencer is suitable only in situations where the signal-to-noise ratio is high. At low signal-to-noise ratios, the difference-correlator will be incapable of providing an effective means of detecting the transmission changes which, after all, are the object of the search procedure.

3.10 Symmetry as Basis of Discrimination

If the premise that the observations, f_1 and f_2, are random processes is accepted, then the probability that f_1 and f_2 are identical is directly given by a measure of the correlation between them. The correlation function, $\phi_{11}(\lambda)$, which is obtained by correlating f_1 with itself is by definition an even function, symmetrical about an appropriately chosen axis. When, however, f_1 and f_2 are correlated the resulting function may not be an even, symmetrical function; it will only be symmetrical when f_1 and f_2 are identical in the sense that they each contain the same transmission signals plus noise. Hence, by comparing the two sides of the correlation function about the chosen axis and testing for symmetry, it is possible to obtain a measure of the correlation between the two functions which have been selected as time histories of the search space and consequently discover any transmission changes between them. The degree of symmetry of the correlation function can be measured in terms of the following expressions:

$$I' = \left| \phi^{+}(\lambda) - \phi^{-}(\lambda) \right| \qquad \ldots (1)$$

and

$$I'' = \int_{-\infty}^{0} |\phi(\lambda)| \, d\lambda - \int_{0}^{\infty} |\phi(\lambda)| \, d\lambda \qquad \ldots (2)$$

I' and I" are not the only indices of the degree of symmetry, but for the purpose of the present investigation they will be accepted as sufficient indices. $\phi^{+}(\lambda)$ and $\phi^{-}(\lambda)$ are, respectively, the positive and the negative parts of the correlation function, $\phi(\lambda)$.

It is important to observe that in order to use I" as a test for symmetry, the function, $\phi(\lambda)$, should be square integrable,* a condition which is generally satisfied in most practical situations. The test for symmetry, as given by expression I", is very revealing in the sense that it provides basis for point-to-point analysis of the search space. It can also be used to introduce the notion of distance, d, which will be defined by

$$d(\phi^{+}, \phi^{-}) = \left| \phi^{+} - \phi^{-} \right| \qquad \ldots (3)$$

d may be said to represent the distance between ϕ^{+} and ϕ^{-}. This expression for d satisfies the customary axioms(32) for the measure

---

* A function, g(t), is said to be square integrable in the interval $[0, T]$ if

$$\int_{0}^{T} |g(t)|^2 \, dt < \infty$$

of distance which are

$$d(\phi^{+}, \phi^{-}) = d(\phi^{-}, \phi^{+}) \qquad \ldots (0.4a)$$

$$d(\phi^{+}, \phi^{+}) = d(\phi^{-}, \phi^{-}) = 0 \qquad \ldots (0.4b)$$

$$d(\phi_1^{+}, \phi_2^{-}) + d(\phi_1^{-}, \phi_2^{+}) = d(\phi_1, \phi_2) \qquad \ldots (0.4c)$$

$$d(\phi_1, \phi_2) + d(\phi_2, \phi_3) = d(\phi_1, \phi_3) \qquad \ldots (0.4d)$$

0.4a and 0.4c express, respectively the symmetry of $\phi(\lambda)$ and the triangle inequality 0.4b is an expression of an identity.

3.11 Consideration of Fading in the Analysis

3.11.1 Introduction

Briefly, fading is the phenomenon which affects the amplitudes of the received waveform in a manner which causes the amplitude to vary either slowly or rapidly. The distorting effects of fading may be selective in the sense that different frequency components of the received signal may be affected differently. In this way, a received signal may be rendered unintelligible by a complete or a partial deletion of some frequency components. In the normal radio engineering practice, devices such as the automatic volume (or gain) control[33,34] are employed to combat the effects of fading, especially non-selective type.

Figs.3.7a - 3.7c illustrate, in a simple manner, the action of selective and non-selective fading in any portion of the search space. Fig.3.7a is the given spectrum of a signal which is not affected by any type of fading. The waveform representing this 'pure' signal is f_a. The disturbed versions of f_a which are produced in the presence of selective and non-selective fading are shown in figs. 3.7b and 3.7c, respectively. It is evident from the waveforms depicted in the diagrams that the transformation which produces f_b from f_a is linear and that f_b is related to f_a by

$$f_b = cf_a \qquad \text{... (2.1)}$$

where $0 \leqslant c < 1$

In a way, the effects of non-selective fading can be said to exhibit some well-known features of a transmission device whose frequency response is flat. Since the quantitity, c, is generally less than unity the amplitudes of the frequency components of the signal, f_a, suffer the same amount of attenuation in the transmission medium. In other words, the shape of f_a is preserved by the transformation which produces f_b from f_a.

Now, consider the situation in which the two observations to be compared in the analysis are represented by f_a and f_b. It must be remembered that f_a is considered different from f_b only when its transmission signal contents are different from those of f_a. In fact, the full expressions for $f_a(x)$ and $f_b(x)$ can be deduced from that for $f(x)$ in equation 2.4 (Sect.3.8.2) and when this is done the analysis leading to an

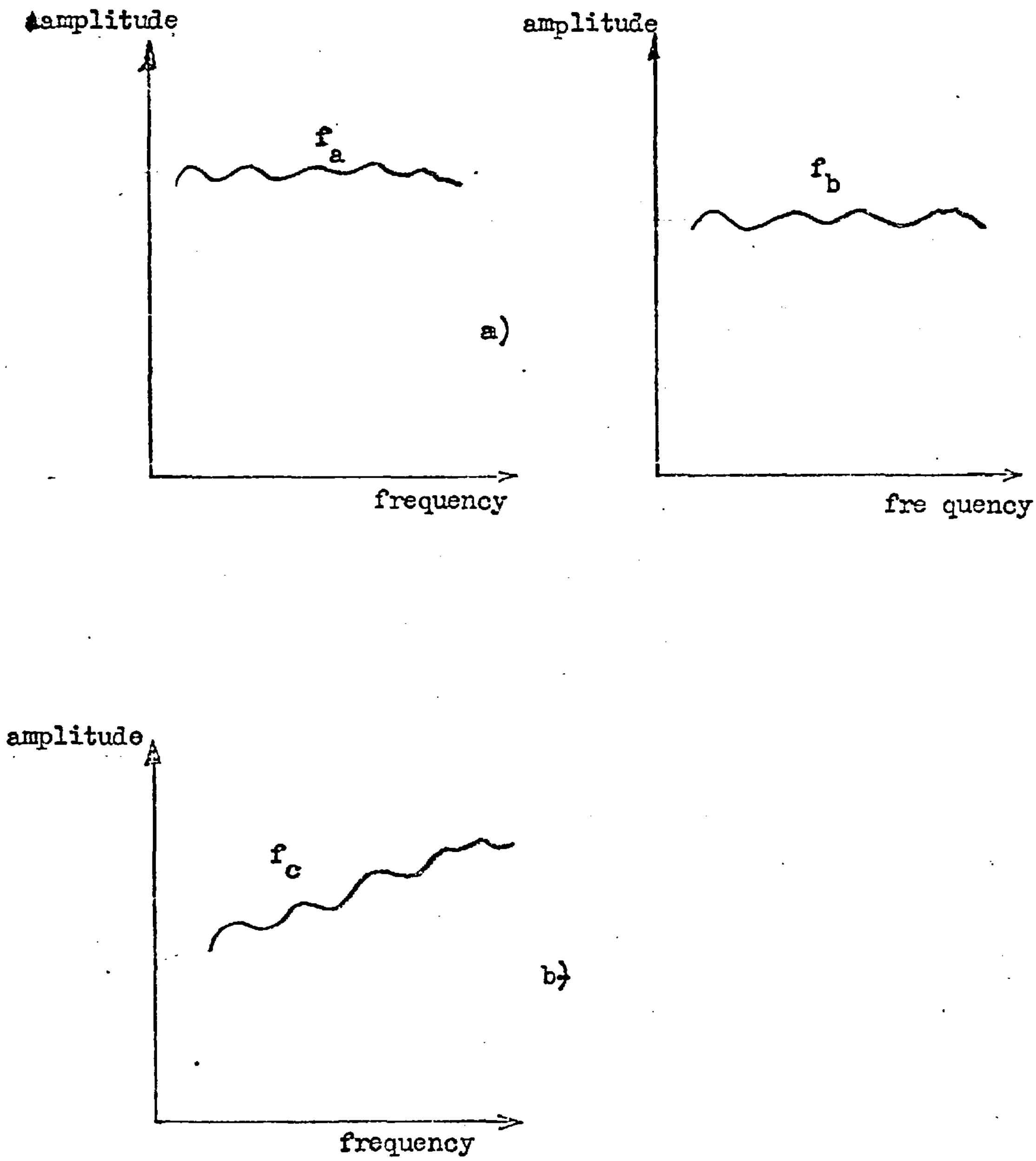


Fig. 3.7 A diagram illustrating the phenomenon of
a) non-selective fading and
b) a form of selective fading

equation similar to equation 2.15 (Sect. 3.8) can be made. Furthermore, it is revealing to visualise f_b as the output of a linear device whose input and system functions are, respectively, f_a and c.

Forming the difference, $f_a - f_b$,

$$f_a(x) - f_b(x) = (1 - c)f_a(x)$$

$$= c'f_a(x) \qquad \ldots (2.2)$$

where $c' = (1 - c)$.

The next step in the analysis is the computation of the correlation of f_a and f_b. That is,

$$f_a \; Ⓡ \; f_b = f_a \; Ⓡ \; c'f_a$$

$$= c'\phi_{aa} \qquad \ldots (2.3)$$

Therefore, apart from the multiplicative factor, the result of the process is the same as would be obtained when f_a is correlated with itself. The function, $c' \Phi(\lambda)$, is a symmetrical function and, by applying the test for symmetry, it is easy to show that

$$d\,(c'\phi^{+}_{aa},\ c'\phi^{-}_{aa}) = c' \left| \phi^{+}_{aa} - \phi^{-}_{aa} \right|$$

$$= 0 \qquad \ldots (2.4)$$

This is the desired result of the search process when the transmission contents of the two observations, f_a and f_b, are identical in the presence of non-selective fading.

It can also be shown that when the transformation of f_a to f_b is due not only to non-selective fading but also to a change in the transmission contents, then the result of the search process will not be a symmetrical function. This can be seen by computing the correlation function of f_a and f_b when the transmission functions, m_a and m_b, are different. Thus,

$$\phi = f_a \; R \; f_b \qquad \ldots (2.5)$$

If the noise components of f_a and f_b are, respectively, n_a and n_b, then

$$\begin{aligned} \phi &= \left[m_a(x) + n_a(x) \right] \; R \; \left[m_b(x) + n_b(x) \right] \\ &= m_a(x) \; R \; m_b(x) + n_a(x) \; R \; n_b(x) \\ &+ m_a(x) \; R \; n_b(x) + m_b(x) \; R \; n_a(x) \\ &= \phi_{m_a m_b} + \phi_{n_a n_b} + \phi_{m_a n_b} + \phi_{m_b n_a} \qquad \ldots (2.6) \end{aligned}$$

As has been demonstrated earlier,* all the terms in the above expression for ϕ, except the first, are either zero or symmetrical about the chosen axis. Hence, the fact that f_a and f_b are not identical as far as their transmission contents are concerned can be deduced from the fact that 2.6 (4) is not a symmetrical function. Clearly, the search method described in the analysis is insensitive to disturbances which are caused by non-selective fading and takes account only of the distortions that are due to real transmission changes.

*See Section 3.8.2.

3.11.3 Selective Fading and its Effects

Where the fading is selective, the waveform, f_c, which is a disturbed version of f_a, is not a simple well-defined function and its relationship with f_a is not simple. Generally, the complex nature of f_c is such that the simple analysis pertinent to non-selective fading is no longer applicable. When the fading is selective it is seldom possible to determine the fading pattern and utilise the knowledge to the best effect in only two observations. However, by combining a consecutive series of observations into one composite observation, the general analysis of the fading may become tractable, and two such composite observations may be analysed satisfactorily. The technique of combining diverse observations of the same event is employed in diversity reception in which two copies of the same signal, which experience different fading patterns are used to "augment" each other.

As an example, consider a simple selective fading pattern whose effect varies linearly with frequency, so that the low frequency components of the transmission signal are more severely affected than the high frequency components. Such a fading pattern will transform f_a (fig.3.7a) into a waveform such as the one illustrated in fig.3.7c. In such a simple case, a method of equalisation might be used to re-transform f_c into a flat-topped spectrum and the analysis then becomes the same as that in the case of non-selective fading.

3.12 The Presence of Impulse Noise

One type of noise, known as low-level noise, is usually below the normal signal level and has the appearance of Gaussian noise superimposed on the harmonics of some very low frequencies(37). The level and the character of this type of noise are such that the effects on the performance of most communication receivers is very small. A second type of noise, impulse noise, differs from the low-level noise in two important ways. First, its appearance(38) when viewed on an oscilloscope is that of a rather widely separated bursts of relatively short duration, about 5 to 50 ms. Second, the level of the impulse noise may be as much as several dbs above the normal signal level.

Since its duration is short, the impulse noise has a flat spectrum which extends over an infinitely wide band odf frequencies. When this type of noise is present in the search space under examination, its effect is to raise the normal level of the signal by an amount proportional to its power density. In effect, the impulse noise changes the d.c. level of the signal.

If f_e (fig.3.8) is the resulting waveform when the effects of the impulse noise are superimposed on f_a, then because of the properties of the impulse noise the relationship between f_a and f_e is

$$f_e = f_a + G \qquad \ldots (0.1)$$

where G is a positive real quantity. The correlation of f_a with the difference, $f_a - f_e$, is given by

$$f_a \circledR (f_a - f_e) = f_a \circledR \left(- G\right) \ldots (0.2)$$

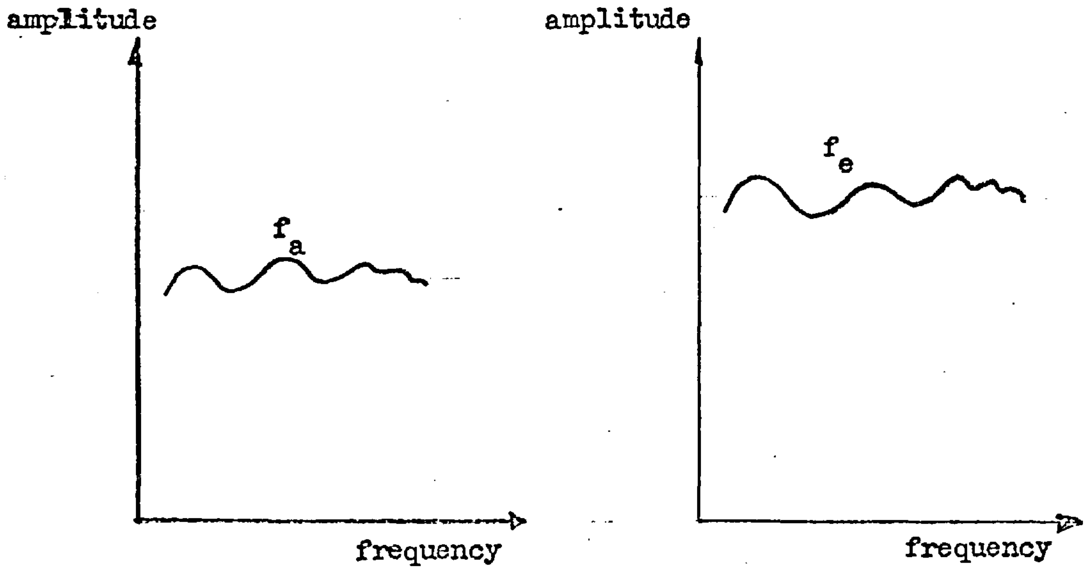


Fig.3.8 A diagram illustrating the instantaneous effect of impulse noise

But since

$$f_a \; R \; (-G) \;=\; G'\int f_a(x)dx \qquad \ldots (3)$$

the output of the difference correlator is

$$Q \;=\; G' \int f_a(x)dx \qquad \left[G' = -G\right] \qquad \ldots (4)$$

which is a constant and consequently a symmetrical function. The condition for symmetry is, therefore, satisfied and the only conclusion is that no transmission signal changes occur when the f_a is transformed to f_e by the advent of an impulse noise within the portion of the frequency band being examined.

f_e can be expressed wholly in terms of a transmission function, $a(x)$, and an additive noise function $n(x)$. The analysis is then similar to that depicted in section 3.8.2 , and the conclusion is that in the presence of impulse noise, the correlation search technique is capable of reliable results, even in the presence of impulse noise. It is also revealing to note that other search techniques [14] which rely on energy difference in order to detect transmission signal changes will produce erroneous results in the presence of impulse, because of the energy increase which will accompany the onset of impulse noise. Similar mistakes will occur when the observations are disturbed by the action of fading. Hence the correlation search technique has clear advantages when dealing with fading and impulse noise.

3.13 The Need for a Threshold

Practically, when noise and fading are present, although f_1 and f_2 may represent two space configurations which contain the same transmissions, d (or D) will seldom be exactly zero. Practical situations like this necessitates a modification of the decision rule based on the zero value of d. Hence, by setting a threshold so that all values of d below this pre-specified threshold can be taken as sufficient to ensure that f_1 and f_2 are identical, it will be possible to make decisions with an arbitrarily low error rate. This is generally done by studying the statistics of d (or D) during simulation tests or in practical experiments. Such a study will be described in the appropriate chapter of this thesis.

Chapter 4

A RADIO SURVEILLANCE SYSTEM

Chapter 4 A RADIO SURVEILLANCE SYSTEM

4.1 Introduction

A practical surveillance system has an important function to perform and that is to provide an easily understood information about an object, or a system of objects, which is being examined in some scientific manner. In such a system various methods are tried during simulation or experiment in order to remedy the system deficiencies and attain some sort of acceptable performance. Where radio surveillance is concerned, the general problem is to set up a simple strategy whereby any one, or several, of a small number of transmissions starting up in any selected portion of some frequency band (like the HF band) can be tracked. The word track has been used here in a restricted sense, and a transmission is said to have been tracked when its position or location in a spectrum has been identified and some information about its probable duration has been obtained.

The presence of noise in the system tends to make the task of correct identification difficult, and spurious location reports are not entirely avoidable. Furthermore, the measurements that are made on the system include some unavoidable error. In this chapter, we attempt to establish a probabilistic foundation for the technique described in chapter 3.

4.2 A Statistical Model

In order to completely identify a transmission in a sense consistent with our objective, we ought to know its location, x, in a frequency spectrum and its duration, t. These important attributes of the system (i.e. the entire frequency band which has been selected for examination) are directly or indirectly related to some measurable quantities. Precise measurements of these quantities, however, is laborious and difficult and the amount of measurement error, inescapably associated with the measured quantities, makes the problem of precise identification even more difficult. Again, the possibility of wrongly associating some measured information with a non-existent transmission cannot be ignored, and a measure of the chance of this event ever occurring is an index of the success of the strategy. Known simply as the false alarm, this event (i.e. identifying a measurement with a non-existent transmission) and the rate at which it is known to occur specify the success performance of the strategy.

In order, therefore, to formulate a statistical model which will adequately describe the behaviour of our system and give credence to the reliability of our strategy, we must first postulate certain probabilities and define their distributions. These are:

(i) the probability that n transmissions are observed in the system.

(ii) the probability that the n transmissions have durations t_i $(i = 1, 2, \ldots, n)$

(iii) the probability that n_i of the observed transmissions are spurious. (This is an index of the false alarm rate).

(iv) the probability that n_j transmissions have escaped our observation. (This is the probability of false dismissal).

4.3 The Distribution of the Transmission Signals and the Duration Times

In this investigation we are dealing with a small number of transmissions which are randomly located in the frequency space. The locations of these transmission signals are assumed random because each transmission signal has a finite probability, not known a priori, of appearing anywhere in the frequency band. Noting the time of initiation of a transmission and observing its total time of persistence, a complete track of it can be established. Our strategy will be to compute the probability that a given subspace of the total "free" space will be occupied by a transmission signal whose probability of persisting for a time t is p(t). If the probability of that particular transmission appearing at x is $\omega(x)$, then the success of the strategy will be dependent on some compound probability, P, which is proportional*

---

* Without loss of any clarity the constant of proportionality can be made equal to unity, recalling that $\int_{-\infty}^{\infty} dP = 1.$

to the probabilities $p(t)$ and $\omega(x)$. That is

$$P = k\, p(t)\omega(x) \qquad \text{... (0.1)}$$

Admittedly, it is generally difficult to know the probabilities, $p(t)$ and $\omega(x)$, a priori and it is best to circumvent this particular difficulty by generating other easily obtainable quantities that are related to $p(t)$ and $\omega(x)$ in some obvious and meaningful fashion. Strictly speaking, $p(t)$ and $\omega(x)$ should be replaced, respectively, by $p_m(t)$ and $\omega_m(x)$, which reflect the fact that these probabilities are conditioned on the existence of a transmission. Furthermore, if the probability of falsely associating a measurement with a non-existent transmission is ν and that of dismissing the existence of a transmission is μ, then the compound probability, P, will also depend on ν and μ. But for a specified ν and μ we can represent P by the product of $p(t)$ and $\omega(x)$, so that

$$P = kp(t)\omega(x) \qquad \text{... (0.2)}$$

4.4 The Effects of Scanning the Space

If the space, or more precisely, the bandwidth being searched is W then we can accomplish our search objective by one of two methods. By employing N observers with N fixed-tuned receivers, each with a relatively small bandwidth of W/N Hz, the whole search operation* can be satisfactorily

*There will be N concurrent observations aimed at detecting any new transmissions that will appear in any of the N band slots.

performed. Obviously, the number, N, of observers, or the i.f. channels, required for this operation is quite large. A second method, which has the inherent capability of reduced man-power by employing one receiver or one i.f. channel, consists in scanning the total space and obtaining information about new transmissions by comparing the time histories of some two scans which are selected in some prescribed fashion. (The prescribed manner of scan selection has been fully described and a more thorough discussion of this method can be found in Chapter (6) of this thesis). Using this second method of comparing two selected scans, transmission signals which appear in one time history (i.e. in one scan period) and not in the other are the transmissions which we seek and about which some information is desired.

With this scanning strategy, it is possible that a transmission whose duration time (or time of persistence) is less than the period of one scan may escape identification and no information about its existence will be gained. It is, therefore, necessary to require that the duration time, t_i, should at least satisfy the following relationship:

$$t_i \geqslant \frac{W}{V}$$

where W = total bandwidth (or total space) in Hertz

V = sweep rate of scan given in Hertz per second

Hence if the two time histories are separated by an integral number of the scan time, $\frac{W}{V}$, and the first of these is obtained at time t, then the second time history is obtained at t + T where,

$$T = k\frac{W}{v}, \quad k = 1, 2 \ldots\ldots n$$

We have assumed that the "blank" time between two consecutive scans is negligible.

Now let it be assumed that the second time history contains a transmission which is absent in the first. Then this new* transmission is identifiable if some period of its persistence time is wholly contained in the time interval, $\left[t + T,\ t + T + \frac{W}{v}\right]$. (See Fig.4.1). This requirement is necessary if the new transmission is not to escape detection.

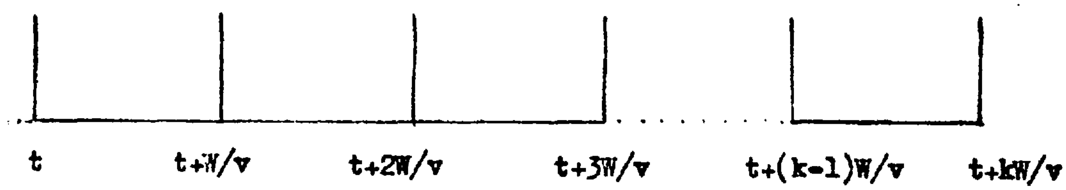


Fig.4.1 A diagram illustrating blocks of observation time

---

* A transmission which appeared in the first scan and not in the second is a missing transmission and should be extinct for a time $t > \frac{W}{v}$.

If the time of persistence of a transmission is less than $\frac{W}{v}$, then it may not come under observation during any one scan period and, therefore, will not be recorded for further examination. Hence, whatever time a transmission starts up and at whatever time it exits, it will not come under observation to be recorded so long as its total time of persistence is less than $\frac{W}{v}$. In the terminology of mathematical statistics, a new transmission which starts up and persists for a time t can be observed and recorded if the probability p(t) is large enough. (See Fig. 1). That is, if the transmission lasts for a time $t \ll \frac{W}{v}$ then the probability of detection of that transmission is negligible. On the other hand a transmission lasting for a time $t > \frac{W}{v}$ has a significant probability of being detected. (See fig. 4.2).

4.5 Random Elements of a Frequency Space

4.5.1 Introduction

Regarding the different transmission signals as some elements of a frequency space, we can attempt to determine the probability that a small number of them can be found, on the average, within a selected band. If the frequency space elements are generated independently so that there is no correlation whatever between the locations and the times of generation of these elements, then we can associate the distribution of these elements with some appropriate distribution law. In fact, in the absence of strict correlation between the space elements, a Poisson law of distribution may be quite applicable. However, in some

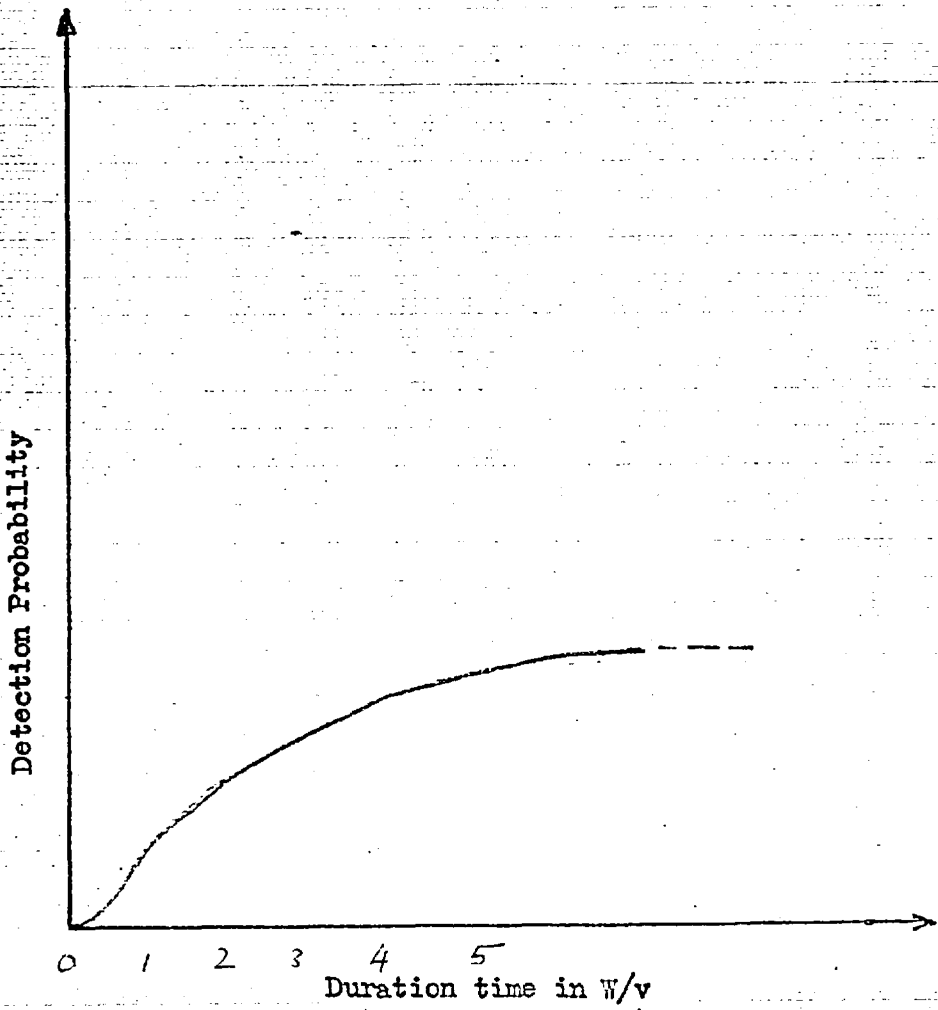


Fig. 4.2 A diagram ~~a~~ showing that the probability of detection varies with the period of scan.

practical situations the elements are somewhat correlated and a simple Poisson law is no longer valid. In the present analysis we shall not specify the amount of correlation that may exist between a given group of space elements.

Fig.4.3 is an illustration of a frequency space in which a number of transmission signals exist. Each spike in the scan picture of the space can represent one transmission signal but if the emission of a particular transmission is of the FM type then a collection of many spikes will constitute one FM transmission signal and hence the spikes within that group will be correlated. Similar situations arise in radar where, apart from the main echo, there exist a number of higher order echoes. Obviously, the main echo and its associated higher order echoes are harmonically related but any two main echoes, arising from two distinct objects, may not be related. Therefore, in the present analysis we shall not specify the amount of correlation that may exist between a given group of space elements. We then ensure that application of the results of the present analysis may be found in situations where the elements of interest are either correlated or not.

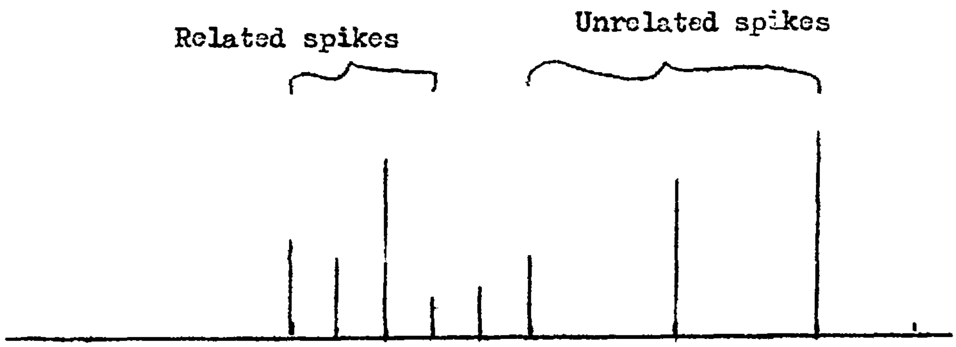


Fig.4.3 A space illustrating the existence of related and unrelated elements.

4.5.2 Configuration Modes of the Space

The frequency space, S, in which a transmission signal or signals are being sought is divided into a number of small intervals or cells, each of which can contain at least one space element. A space element, as has been described previously, is a point or a collection of points with which we can associate one hf transmission. In Fig. 4.4 , for instance, the number of cells is C = 3, and each of these is occupied by one or more transmissions as indicated by the number of m's written in each cell. Our objective is to obtain a relevant statistic of the information contained in one cell* and utilise this statistic in a data association (39) process in which a given statistic is associated either with the fact that a transmission is present

*The words cell and interval will be used interchangeably in this discussion

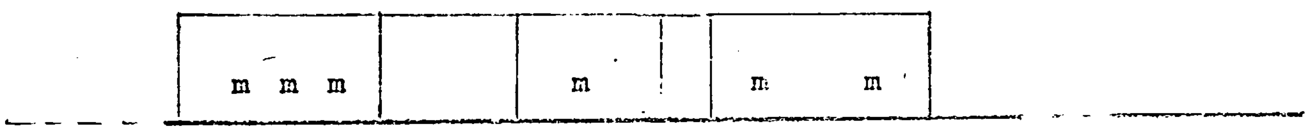


Fig. 4.4 A hypothetical space of cells containing different numbers of elements

or the fact that it is absent in the given cell. A convenient and a desirable statistic, for our purpose, is one which is sensitive to changes in the transmission signal content of a cell. In the meantime, we shall assume that this statistic is determinable and is invariant under changes that do not affect the relevant quantities in the space interval. Such a statistic will provide the necessary information about the nature of the random space elements.

Suppose the space is quantized uniformly, or non-uniformly, into a number of intervals designated by s_1, s_2, ... s_N. If the space is one-dimensional then s_i is a one-dimensional subspace and is contained between, say, the points x_i and $x_i + \Delta x_{i+1}$; its length is then approximately equal to Δx_i. A conceptual extension to an n-dimensional space is possible if we represent x_i by an n-tuple, or a vector, so that

$$x_i = (x_{i_1}, x_{i_2}, \ldots x_{i_n})$$

For our immediate purposes, we can simplify matters by associating the contents of a cell, s_i, with the value 1 or 0, depending on whether the information contained is indicative of the presence of a transmission signal or not. Hence a functional such as $U(s_i)$ is either 1 or 0. Mathematically put,

$$U(s_i) = \begin{array}{l} 1 \text{ if } s_i \text{ contains a transmission} \\ 0 \text{ if } s_i \text{ does not contain a transmission.} \end{array}$$

Thus, by continually examining the contents of one cell, or a collection of cells, and observing the transition times(i.e. times when $U(s_i)$ changes from 1 to 0 or vice versa) we are able to detect the times of emergence of new transmissions or the disappearance of old ones. When we are dealing with a collection of cells we may have to consider a secondary functional, $w(U_1 \ldots U_N)$, which depicts the overall estimate of the space. The functional arguments, U_i, are either 1's or 0's. In this manner, three principal modes of space configurations can be distinguished, viz:

(1) A space configuration whose functional, w, has only 1's as its arguments. We shall call this the "positive mode".

(2) A space configuration whose functional, w, has only 0's as its arguments. This is the "null node".

(3) A space configuration whose functional, w, has a random mixture of 1's and 0's as its arguments. This is the "mixed mode".

We can specify that an interval whose contents reveal the presence of a transmission has a functional U = 1. If a transmission does not exist in a particular cell then its associated U is equal to 0.

A more general terminology may be introduced by considering that each cell contains some space element. A "positive space element" exists in a cell, s_i, if $U(s_i) = U_i = 1$. If $U(s_i) = U_i = 0$, it may be said that s_i contains a "null space element". Mathematically, $U(s_i) = 1$ if s_i contains a "positive space element", or $U(s_i) = 0$ if s_i contains a "null space element".

A sufficiently general mode of the entire space is the "mixed node" and we shall be particularly interested in its configuration. Such a mode contains an unspecified number of space elements and is represented by the functional

$$W(U_1, U_2 \ldots U_N).$$

A diagrammatic representation of a typical space has been given in Fig.4.5, and in this figure, the "space element" functionals, which correspond to various portions of the space, have been shown. In this particular example the overall functional, w, is given by

$$w = w(U_1, U_2, U_3, U_4) = w(1, 1, 0, 1)$$

It is important to point out that a group of detectors which examine the contents of the cells and decides whether U is 1 or 0, operate non-parametrically[40,41] in the sense they make little or no use of the statistics of the transmission or the noise field. This is largely because a priori knowledge about these fields are either scantily available or non-existent. Additionally, the number of changes that can be expected in a space of a finite number of elements is quite large and cannot be indexed by a finite number of real parameters. As an example, we can consider the problem of accommodating telegraph channels, telephone channels and broadcasting channels in any given broad band. Where the size of the broad band is known, we can calculate the number of channels of one particular kind, but we shall find it a very difficult task to compute the number of the different types of channels that can be accommodated within the same given broad

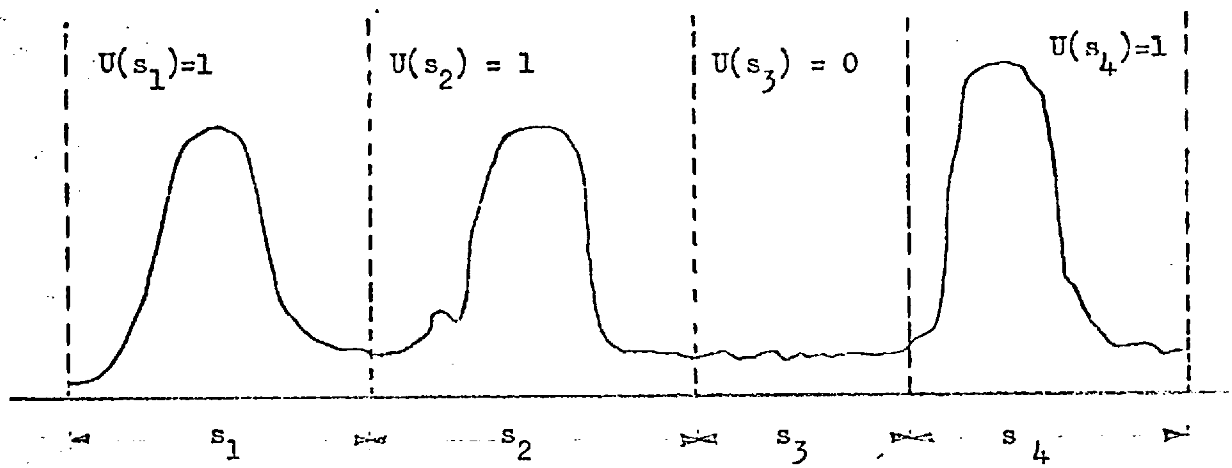


Fig. 4.5 A diagramatic sketch of signals in a hypothetical space

band. Our problem is, therefore, non-parametric in this sense, but the functionals, w, and their representation as detector outputs will suffice to depict which space intervals or cells are occupied by transmissions. In this way we are able to describe the "local" behaviour of the transmission signals within different portions of the broad band being examined.

4.0 Distribution of Elements in the Search Space

In a mixed mode configuration, we do not know the composition of the space a priori but we can assume that there are n positive elements and m null elements so that $m + n = N$. The arrangement of these N elements is not ordered in any particular manner and, in general, we shall expect to find a space element of one kind randomly followed (or preceded) by a space element of the same type or that of the other type. Combinatorial techniques exist for calculating the number of the different possible combinations of the two types of N elements.[42,43] Each possible combination, depicting a possible space configuration, will be assigned a finite probability measure which is an index of the chance that that particular configuration will occur. If we assume that every combination is equally likely to occur, then the probability* associated with each combination is $\binom{N}{m}^{-1}$.

* This represents the "worst" possible configuration. In terms of information theory this type of configuration maximises the uncertainty of the situation under examination.

At this point it is appropriate to make use of an assumption we made earlier when we postulated the existence of the probability, $p_n(s_1 \dots s_n)$, that there are n positive elements in the cells, s_1, $s_2, \dots, s_n$. This is true if $U(s_i) = 1$ for $i = 1, 2, \dots n$. But since we have assumed that all combinations of m null elements and n positive elements are equally likely, the probability of a given configuration of m + n space elements is equal to the probability of a transformed configuration in which the n positive elements occupy the first n cell positions. Before we make use of $p_n(s_i \dots s_n)$ we shall consider the distribution density function, $F(s_1 \dots s_N)$, in terms of which we can calculate the probability that each of the cells $s_i(s = 1, 2, .. N)$ contains at least one space element. Some of the different forms that $F(s_1, s_2 \dots s_N)$ can take are illustrated in Fig. 4.6. When a transmission is not active or is absent its cell is occupied by null element, that is, by noise alone. However, as it becomes active a positive element appears in the cell. If $F(s_1, s_2 \dots . s_N)$ is a well-behaved, continuous function in almost all regions of interest, then the probability that there are N space elements in a given space is

$$\triangle P(s_1, s_2 .. s_N) = F(s_1, s_2 \dots s_N) \prod_{i=1}^{N} \Delta s_i \qquad \dots (0.1)$$

It will be illuminating, at this point, to show briefly the difference between the two probability expressions, $p(\varepsilon_1, s_2 \dots s_n)$ and $P(s_1, s_2 \dots s_N)$. The latter provides information about all cells of the space (i.e. a volumetric description) in each of which one space

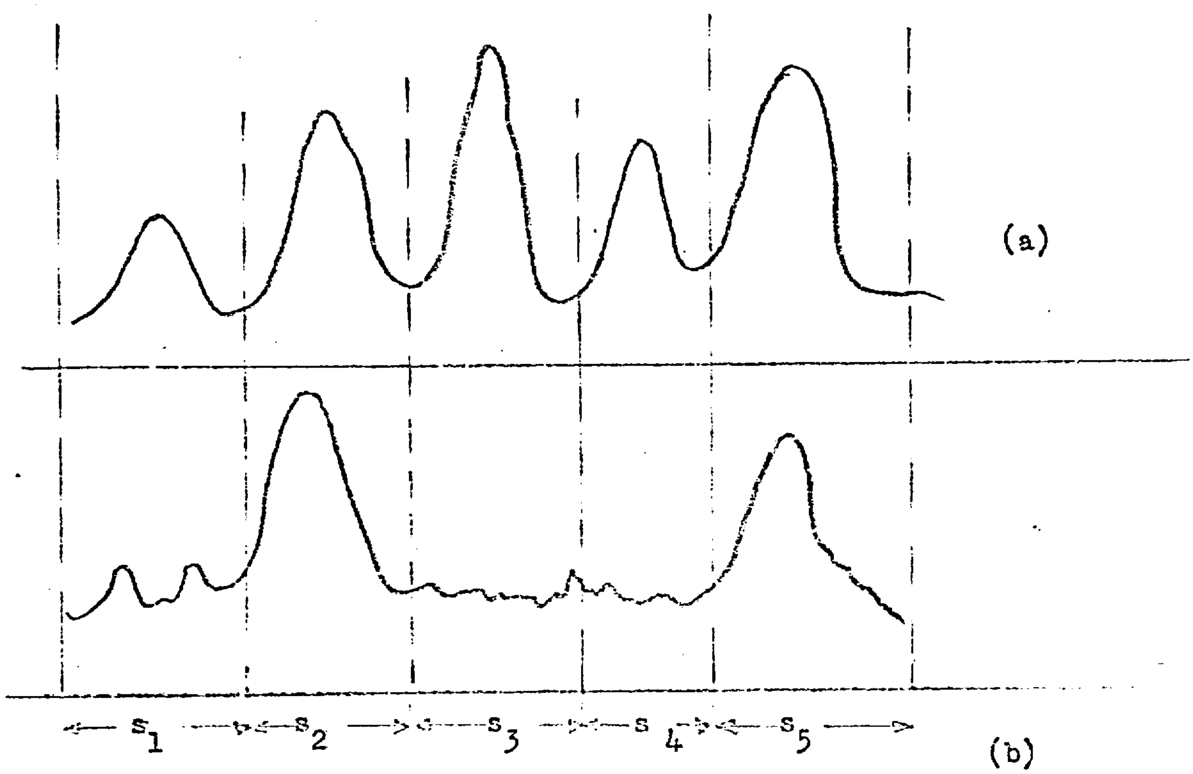


Fig. 4.6 Some forms of $F(s_1, s_2, \ldots\ldots s_N)$

element (positive or null) can be expected to exist, while the former provides a localised statistical description of the set of cells which contain positive elements only.

We shall now show how the probability function $p_n(s_1, s_2, \ldots s_n)$* is derived from the knowledge of $P(s_1, s_2, \ldots s_N)$; and we shall further observe that the distribution density function $F(s_1 \ldots s_N)$ could be related to the instantaneous picture of the space under examination.

As already mentioned, a given space interval, s_i, contains a positive element or a null element if the associated functional, U, has the value 1 or 0, respectively. Since the two events represented by U = 1 and U = 0 are mutually exclusive, it follows that

$$Pr(U = 1) + P(U = 0) = 1 \qquad \ldots (0.2)$$

Considering a space configuration in which there are n positive elements and m null elements, we can determine the associated $p_n(s_1 \ldots s_n)$ and $P(s_1 \ldots s_N)$ subject to the condition that $m+n = N$.

Let Q_* represent the event that there are a total of n positive

---

*As will be shown later, $p_n(s_1, s_2 \ldots s_n)$ must be written as $p_n(s_1, s_2, \ldots s_n/N)$ so that the conditioning of it on the existence of N cells may be reflected. Also, written in the form, $p_n(s_1, s_2, \ldots s_n/N)$, its behaviour as a posterior probability is revealed. However, in all the discussions we shall write $p_n(s_1 \ldots s_n)$ for $p_n(s_1 \ldots s_n/N)$.

elements, disregarding their cell positions within the given space. Then the notation for $P(s_1, s_2 \ldots s_N)$ appropriately changes to

$$P(s_1, \ldots s_N) = Pr(Q_n, U_1' = 0, U_2' = 0, \ldots U_m') = Pr(Q_n, U_1', U_2', \ldots U_m)$$

In this notation $U_i'(i = 1, 2, \ldots m)$ is the functional associated with a null element, and in this particular case there are m of such elements.

From equation (0.2)

$$Pr(Q_n, U_1' = 1) + Pr(Q_n, U_1' = 0) = Pr(Q_n) \qquad \ldots (0.3)$$

Hence $Pr(Q_n, U_1' = 0) = Pr(Q_n) - Pr(Q_n, U_1' = 1) \qquad \ldots (0.4)$

The L.H.S. of eqn.(0.4) is the probability of the joint event that Q_n and the existence of one null element occur simultaneously; that is, there are n positive space elements and one null element. Also, it follows from equation (0.4) that the joint probability $Pr(Q_n, U_1' = 0)$ can be expressed in terms of $Pr(Q_n)$ and $Pr(Q_n, U_1' = 1)$, the latter giving the probability that there are n + 1 positive elements. In fact,

$$Pr(Q_n, U_1' = 1) = Pr(Q_{n+1}) \qquad \ldots (0.5)$$

may be regarded as an appropriate recurrence equation, and by using equation (0.5), we are able to convert a null space element to a positive element. This means that we can operate on a space which contains n positive elements and one null element to obtain the quantitity representing $Pr(Q_{n+1})$. The condition which must be fulfilled before this operation can be performed physically is that there should be n positive elements and one null element in the space.

Hence, for the conversion of a null element the emergence of an active transmission in a cell which was previously occupied by noise alone must occur. When this happens $U_1' = U_1 = 1$.

We can now proceed and derive the expression for $Pr(Q_n, U_1' = 0, U_2' = 0)$ which is the probability that in addition to the event Q_n there exist two null space elements in the given space. Using similar arguments as in the case of Q_n plus one null space element, we have

$$Pr(Q_n, U_1' = 0, U_2' = 0) + Pr(Q_n, U_1' = 0, U_2' = 1) = Pr(Q_n, U_1' = 0 \quad \ldots (0.6)$$

Rewriting equation (0.4) we have,

$$Pr(Q_n, U_1' = 0) = Pr(Q_n) - Pr(Q_n, U_1' = 1) \quad \ldots (0.7)$$

and using the same reasoning which lead to equation (3), we see that

$$Pr(Q_n, U_1' = 0, U_2' = 1) + Pr(Q_n, U_1' = 1, U_2' = 1) = Pr(Q_n, U_2' = 1) \quad \ldots (0.8)$$

Substituting (0.7) and (0.8) into (0.6), we have

$$Pr(Q_N, U_1'=0, U_2'=0) = Pr(Q_n) - Pr(Q_n, U_1'=1) - Pr(Q_n, U_2'=1) + Pr(Q_n, U_1'=1, U_2' = 1) \quad \ldots (0.9)$$

In this case equations (0.7) and (0.8) represent the conversion operations which are performed under the condition stated above.

So far we have considered two distinct space configurations, one of which has one null element and the other two null elements. We can now consider the general case when the configuration contains m null elements. It is not difficult to show that, in this general case,

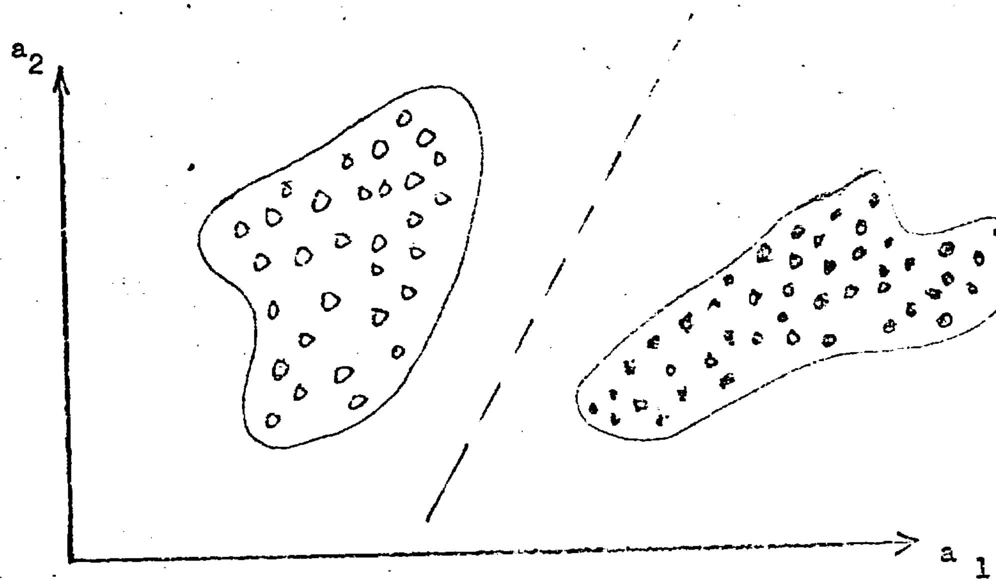
$$\Pr(Q_n, U_1' = 0, U_2' = 0, \ldots, U_m' = 0)$$
$$= \Pr(Q_n) - \Pr(Q_n, U_1' = 1) + \ldots\ldots$$
$$+ (-1)^m \Pr(Q_n, U_1' = 1 \ldots\ldots\ldots U_,' = 1) \qquad \ldots (10)$$

This expression, written in full, contains 2^m terms and converts the probability $\Pr(Q_m, U_1 = 0, \ldots U_m = 0)$ of n positive elements and m null elements to a probability of positive elements only in a sense described earlier. This artifice of conversion has been adopted in order to make use of the density function, $F(s_1 \ldots s_N)$, in the calculation of $P(s_1, \ldots s_N)$, remembering that

$$\Delta P(s_1, \ldots s_N) = F(s_1, \ldots s_N) \prod_{i=1}^{N} \Delta s_i$$

In this expression we have assumed that each cell contains at least one space element.

Our main objective is to determine the probability, $p_n(s_1 \ldots s_n)$, that each of the intervals, $s_i (i = 1, \ldots n)$, contains one positive element, and that the rest of the space contains null elements only. In order to find $p_n(s_1 \ldots s_n)$ we partition the total space into two main sub-regions, one of which contains all the positive elements and the other, all the null elements. This idea is illustrated in Fig.4.7 where elements of two different classes are contained in the different regions labelled R and R'. These classes are linearly separable in fig.4.7a but not in the more complex situation, shown in fig.4.7b. It can be seen that the boundaries shown as dotted lines a)



b)

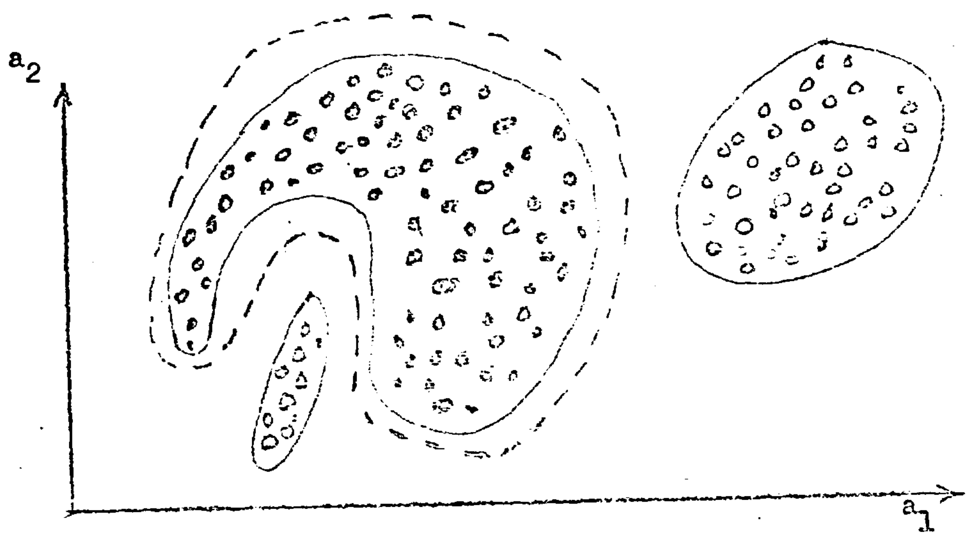


Fig. 4.7 Separation of classes by decision rules.

a) Linear Decision rule ; b) Non-linear Decision rule in fig.4.7 separate members of different classes and are constructed by some decision rules appropriate for the particular set of classes. It is important to note that in these diagrams each member of R or R' is located by a set of two numbers, a_1 and a_2, because the space of search is two dimensional. The dichotomy is only artificial and may not be physically possible. But it is a helpful procedure in any situation where the objects of analysis belong, or can be made to belong, to some well-defined categories as they are in our present case.

An effective way of classifying all the elements of the space in accordance with their two allowable states is to derive an appropriate functional from the function $F(s_I, \dots s_N)$. This is done by removing, through appropriate integration procedures, one group of quantities associated with either the positive or the null space elements. Since we are primarily interested in the positive space elements - at least, more so than in the null space elements - we shall integrate away all the quantities belonging to the null elements. We are then left with a resultant functional which is dependent only on the positive space elements or quantities associated with them.

If we momentarily allow a continuous variation[x] in the U's then we can find the probability that a particular value of U is contained in

[x] The change from, say U = 1 to U = 0 can be regarded as being gradual. This is only a mathematical necessity in order to make the analysis tractable.

the interval $(U, U + \Delta U)$. It will then be possible to calculate the ratio

$$\frac{\Delta P(U_I, \ldots\ldots U_N)}{\Delta U_I \ldots\ldots\ \Delta U_n}$$

where $n \leq N$.

Because of the simple linear relationship[x] between U and s we can replace $P(U_I, \ldots\ U_N)$ by $P(s_I, \ldots\ s_N)$, and the small increments in the U's can be replaced by the corresponding increments in the s's. Therefore, the above ratio becomes

$$\frac{\Delta P(s_I, \ldots\ldots s_N)}{\Delta s_I \ \ldots\ldots\ \Delta s_n}$$

where we have assumed that the s's also can be varied continuously. In the limit, as the individual s's approach zero, the ratio tends to a limit given by

$$p(s_I, \ldots, s_n) = \lim_{s_i \to 0} \frac{\Delta P(s_I, \ldots\ldots, s_N)}{\Delta s_I \ldots\ldots\ \Delta s_n} \quad \ldots (0.11)$$

We should remember that $P(s_I, \ldots, s_N)$ is the probability that of the N space elements n are positive and m are null. That is,

[x] An interval, s, containing a positive space element has an associated $U(s_1)$=U=I and so the probability that s contains a positive space element is equivalent to the probability that U=I. More mathematically, the variable, s, and U, have a one-to-one mapping and the Jacobian of the transformation can be assumed constant and normalised to unity.

$$P(s\ ,\ \ldots\ldots\ s) = P(s_1,\ \ldots\ s_n,\ s_1',\ \ldots\ s_m')$$
$$= P(S_1,\ \ldots\ S_n,\ S_{n+1}\ \ldots\ S_{n+m})$$

where, again, the prime sign denotes a null element as distinct from a positive element which is represented by an s without the prime sign.

We shall now attempt to write equation (0.10) in a more compact form and then substitute the resulting expression for $P(s_I, \ldots\ s_N)$ into (0.11). Referring to equation (0.10), we see that the second term (in the summation) is a sum of terms containing one U each. The third term is a sum of terms each of which contains two U's. The sign preceding a particular term is dependent upon the number of U's to be found in that term of that particular sum and it is a plus sign if the number of U's is even, otherwise it is a minus sign. These considerations lead us to write

$$\Pr(Q_n, U_1' = 0,\ \ldots\ U_m') = \Pr(Q_n) + (-1)^m \sum_{k_1 < k_2 < \ldots < k_m} \Pr(Q_n, U_{k_1}' = 1, \ldots, U_{k_m}' = 1)$$
$$= (-1)^m \sum_{0 < k_1 < \ldots < k_m} \mathrm{Pe}(Q_n,\ U_{k_1}' = 1,\ \ldots, U_{k_m}' = 1) \qquad \ldots\ (0.12)$$

It is important to note that $\Pr(Q_n, U_1' = 0, \ldots, U_m' = 0)$ is the probability that the space under consideration contains only n positive elements and that the rest of the space is occupied by null elements. However, the probability that there are n+m = N space elements at all time is written as

$$Pr(Q_n, U'_1 = 1, \ldots, U'_m=1) = F(s_1,\ldots,s_n,s'_{n+1},\ldots,s'_{n+m}) \prod_i^n \Delta s_i \prod_{j=n+1}^{n+m} s_j$$

by our basic definition. ... (0.13)

Thus, substituting equation (0.13) into (0.12), we have

$$Pr(Q_n,U'_1=1,\ldots,U'_m=1) = (-1)^m \sum F(s_1,\ldots,s_n,s'_{n+1},\ldots,s'_{n+m}) \prod_i^{n+m} s_i \prod_j^n \Delta s_j \quad \ldots (0.14)$$

To obtain the probability $p(s_1,\ldots,s_n)$, we divide equation (0.14) by the product $\prod \Delta s_i$ and proceed to the limit as the Δs_i $(i = 1,2,\ldots,n)$ go to zero.

Thus,

$$p(s_1,\ldots,s_n)=\lim_{\Delta s_i \to 0} = Pr(Q_n,U'_1=1,\ldots,U'_m=1)/ \prod_i^n \Delta s_i$$

$$= \lim_{\Delta s_i \to 0} \frac{\Delta P(s_1,\ldots,s_n,s_{n+1},\ldots,s'_{n+m})}{\prod_i^n \Delta s_i}$$

$$= \lim_{\Delta s_i \to 0} \frac{(-1)^m \sum_j F(s_i,\ldots,s'_{n+m}) \prod_{i=n+1}^{n+m} \Delta s'_i \prod_{i=1}^{n} \Delta s_i}{\prod^n \Delta s_i}$$

$$= \lim_{\Delta s_i \to 0} (-1)^m \sum_j F(s_1,\ldots\ldots s_{n+m}) \prod \Delta s_j$$

$$= \sum (-1)^m \int_{S} \int F(s_1,\ldots\ldots s_{n+m}) \prod_j ds. \quad \ldots (0.15)$$

where the integration is performed over the whole space of search, S.

It is immediately evident from the mathematical procedure delineated above that it is always feasible to remove - through an integration procedure-the quantities associated with the null elements. However, since there are different possible combinations of m null elements (in fact, m! possible combinations do exist) among the n positive elements and we assume that each combination is equally likely to occur, a more accurate expression for $p_n(s_1, \ldots s_n)$ is:

$$p_n(s_1 \ldots s_n) = \sum \frac{(-1)^m}{m!} \int_S \!\!\cdots\!\! \int F(s_1, \ldots s_{n+m}) \prod_{i=n+1}^{n+m} ds_i \qquad \ldots (16)$$

If the statistical characteristics of noise and the signal fields are assumed known then the form of $F(s_1, \ldots s_{n+m})$ will be completely known and the integration is straight forward. However, in the present analysis the form of $F(s_1, \ldots s_{n+m})$ is unknown and unspecified. Any temporal representation of the spectrum (i.e. the search space) is regarded as $F(s_1 \ldots s_{n+m})$ and from this some inference is made concerning the arrival or the disappearance of transmissions. Furthermore if we restrict ourselves to detecting changes in the space configuration, then our search procedure should aim at exploiting the difference between one set of probabilities and another. Hence, probabilities such as $p_n(s_1 \ldots s_n)$ and $p_k(s_1 \ldots s_k)$ $(k \neq n)$ could be manipulated in such a way as to bring out their differences. The concept of distance between p_n and p_k can be developed and shown to form the basis of our search for the "arrival" and the "departure" times of the transmissions. It must also be observed that the different $p_n(s_1 \ldots s_n)$ $(n = 0,1,2,\ldots)$ determine the various uncertainties about the search space and by computing the differences between them we are able to gain some information about the existence (or non-existence) of elements in the search space.

4.7 Application of the Concepts of Random Point Theory

4.7.1 Introduction

We shall now apply the concepts which have been illustrated in the above discussion to our search problem. Suppose the sample function, $f(x)$, at the output of a receiver is an additive combination of noise and an unspecified number of transmission signals whose locations in the search space (i.e. the frequency band being examined) are represented by the points, x_i $(i = 1, \ldots\ldots r)$. We denote by Q_r the event that a definite number, r, of points will fall in the region, S, and that one point will fall in each of the intervals,* $s_i (i = 1, \ldots\ldots\ldots, r)$, of the space. A typical sample function is shown in fig.4.3(page 101) in which each spike or a collection of consecutive spikes constitutes one hf transmission. Generally speaking, one spike in a frequency space like this will represent one AM (amplitude modulated) transmitted signal but more than one spike in a recognisable, definite spacial arrangement will reveal the presence of some other mode of transmission, FM, say.

From the sample function alone we can hardly gain substantial information about the transmissions. The sample function may be the output of a receiver whose normal criterion of excellent performance is the maximised signal-to-noise ratio (SNR) which is obtained by means

---

* The intervals is regarded as a collection of points in the neighbourhood of x_i.

of some suitable filtering. The motivation for this type of reception by filtering lies in the fact that noise is what ultimately limits the sensitivity of the receiver and the less there is of it the better. However, Woodward[45] and others[44] have shown that a detector which only maximises the SNR may not face up to the problem of extraction of maximum information and that a mere observation of the end product of such a detection procedure does not ensure maximum gain of information. Hence any detection procedure whose goal is to maximise the information gain is more desirable, and the communication engineer generally aims at extracting maximum information from any given sample function.

4.7.2 Extraction of Information

In the present problem the wanted signal is usually in the form of a pulse of a definite but unknown spatial duration, and the information required for its identification is that concerning the position and the amplitude. The signal, as an output of some receiver, usually contains noise which, depending upon the magnitude of its power relative to that of the signal, may partially or wholly obscure the signal position and destroy information about the amplitude. Our problem, therefore, is to operate on the received signal, f, which is the sum of the wanted signal, $m(x)$, and the unwanted obiquitous noise, $n(x)$, and thereby obtain some information concerning the position and amplitude of the signal pulse. (The information already contained in the sample function, f, is never increased by any amount of operation on f, at best, we can hope to conserve information about the wanted signals which have been contaminated by channel and receiver noise.) Any other information in f is not very important to the observer whose primary objective is to gain information only about the signal pulses and their spacial positions. Therefore, a procedure which eliminates as much unwanted information as possible is optimally appropriate in our search problem. From a knowledge of the sample function we should like to compute the conditional probabilities of the existence (or non-existence) of the various useful signals. It will be shown (see section 4.7.4) that these conditional probabilities, known as the posterior probabilities, cannot generally be computed exactly without the knowledge of the prior probabilities with which the wanted signals occur.

4.7.3 Bayes Theorem

The probability that two events X and Y occur simultaneously is denoted by P(X,Y) and satisfies the relation

$$P(X,Y) = P(X)\ P_X(Y) \qquad \dots (3.1)$$

$$= P(Y)\ P_Y(X) \qquad \dots (3.2)$$

We observe that fromequations (3.1) and (3.2)

$$P_X(Y)\ P(X) \quad = \quad P_Y(X)\ P(Y)$$

$$P_Y(X) \quad = \quad \frac{P(X)\ P_X(Y)}{P(Y)} \qquad \ldots (3.3)$$

Equation (3.3) is a mathematical representation of Bayes Theorem; $P_Y(X)$ is the prosterior probability of the event, X, given that Y has occurred; P(Y) is the prior probability of the event, Y; $P_X(Y)$, known as the likelihood function, is the probability that X causes Y and P(X) is the prior probability of X.

In communication, Y is usually the received pattern of a message, X, transmitted over some communication channel. Since noise in the channel and the receiving apparatus is unavoidable, the received pattern is never a true reproduction of the transmitted message. However, the received pattern, Y, contains a retrievable evidence of X and the general communication problem is to infer from the knowledge of the received pattern, Y, that X is the transmitted message. This inference is the object of the computation of posterior probability which makes it possible for us to extract maximum information about X.

4.7.4. The Posterior Probability

From equation (3.3) of the preceding section we see that the posterior probability of X given Y is

$$P_Y(X) \; = \; P(X)\ P_X(Y)/P(Y) \qquad \ldots (4.1)$$

Now let us consider a complete system of events, X_i(i I, ... N) in which one and only one event, X must occur as result of some experiment. Such a finite scheme of events is said to be mutually exclusive, and if another event, Y, can occur only if one event of the system, X, has occurred, then the probability that Y has occurred is given by the expression

$$P(Y) = \sum_i P(X_i)P_{X_i}(Y) \qquad \text{... (4.2)}$$

Substituting the expression for P(Y) from equation (4.1) we see that

$$P_Y(X_i) = \frac{P(X_i)P_{X_i}(Y)}{\sum_i P(X_i)P_{X_i}(Y)} \qquad \text{... (4.3)}$$

Since the system, X_i, (i I, ...N) is complete, in the mathematical sense

$$\sum_i P_Y(X_i) = 1 \qquad \text{... (4.4)}$$

For our purposes X_i (i = 1, ... N) is some description of a state in which a signal plus noise or noise alone exists. Equation (4.3) tells us that the posterior probability of any X_i, given a particular Y, is obtained by considering equation (4.3) in which the summation is taken over all X's. Y is the given data from which we wish to extract maximum information about X. Hence if the criterion of excellence or optimality is the maximum gain of information then the correct method is to choose an X_i corresponding to the maximum $P_Y(X_i)$. Since Y is known we can assume that P(Y) in equation (4.1) is known and, in the worst possible situation, equal to a constant.

Thus

$$P_Y(X) = KP(X)\, P_X(Y) \qquad \ldots (4.5)$$

Whenever it is impossible to know the prior probability, P(X), the dependence of $P_Y(X)$ on P(X) is relaxed by assuming that P(X) is a uniform distribution. A uniform distribution means that all events of the finite scheme are equally likely to occur. The events with which we are concerned in this analysis are: (1) that a given space in the spectrum is occupied by signal plus noise and (2) that a given space in the spectrum is occupied by noise alone. Without some a priori knowledge* about the composition of the spectrum we shall assume that the two events of our scheme are equally likely. In a situation where this assumption cannot be justified, the dependence of the posterior probability only on the likelihood function is assumed and employed in the hope that the results will not deviate markedly from those that would be obtained if the prior probability were known and used. The computation of the likelihood function, rather than the posterior probability, has been found in many practical situations to sufficiently conserve as much as possible information about unknown states, X_i ($i = I \ldots\ldots N$). Woodward[46] and Davies[47] have shown that it is not possible to select a value of X which maximises the information gain unless a definite prior distribution is employed because "information gain" cannot be measured in the absence of the prior probability.

---

* A reference to the Berne list will provide some information regarding the frequencies which are being used.

4.7.5 The Decision-making Process

An optimum system which, from the available information, Y, computes the likelihood function $P_X(Y)$, or the posterior probability, $P_Y(X)$, is an indispensable device if the information to be gained is to be a maximum. Such an optimum system may leave the decision making to an intelligent observer (or to a different device) who will decide on one state of X from among two or more alternatives. For instance, the observer may decide that a transmission signal is present or absent and this decision, not always without some bases may reflect the observer's prejudice concerning the states of the events being examined. In general, the smaller the prior probability (or the observer's prejudice) of the occurrence of the signal, the more the signal must exceed the noise in order to be sure of its existence. The signal, as an event, is said to exist if the likelihood function corresponding to its existence is larger than the likelihood function corresponding to its absence. The signal, as an event, is said to occur if the likelihood function (or the prior probability) corresponding to its existence is larger than the likelihood function (or prior probability) corresponding to its absence. This statement generally is true when the prior probabilities of these two events are equal. For, if X_1 represents the event that a signal exists and X_2, the event that a signal does not exist, then making use of equation (4), we observe that

$$\frac{P_X(X_1)}{P_Y(X_2)} = \frac{K\ P(X_i)\ P_{X_1}(Y)}{K\ P(X_2)\ PX_2(Y)}$$

$$= \frac{P(X_1)P_{X_1}(Y)}{P(X_2)P_{X_2}(Y)} \qquad \ldots (5.1)$$

Thus if $\frac{P(X_1)}{P(X_2)} = 1$ then

$$\frac{P_Y(X_1)}{P_Y(X_2)} > 1 \quad \text{if and only if}$$

$$\frac{P_{X_1}(Y)}{P_{X_2}(Y)} > 1$$

Therefore, if there are two possible and equally likely states then a more useful quantity is the likelihood ratio which is the ratio of the likelihood function corresponding to the signal's presence to the likelihood function corresponding to its absence.

The decision scheme can be made automatic in a rather simple way by continually comparing various values of the calculated likelihood ratio with a preset value, known as the threshold. Before the threshold is fixed a limit is set to the false alarm probability (i.e. the probability of deciding on the existence of a signal when it is in fact absent) which can be tolerated in order to ensure a satisfactory detection of the signal. Hence when the proper constraints are set by the false alarm probability a decision scheme can be established so that the distinction is made between one hypothesis (that a signal exists) and an alternative (that a signal does not exist).

For our purposes we consider testable hypotheses which will reflect our interest in different configurations of our search space. As far as the different space elements (i.e. positive and null space elements, as defined in section (4.5), are concerned, a space configuration remains unchanged until a new transmission (or transmissions) occurs or an old one exits. We are able to detect the arrival of new transmission(s) or the departure of old ones by means of this change in the state of the space configuration, and our primary objective is to detect this change, whenever it occurs, in as short a time as possible, subject to the prescribed condition on false alarm probability. The detection problem is now reduced to a simple hypothesis testing, the results of the test indicating whether the hypothesis C that a change has occurred is true or not. The alternative, C^-, is that no change has occurred and that the general parameters which characterise the configuration of the space have suffered no change.

Our decision will, therefore, be based on the posterior probability of the parameters of the space configuration. As noted earlier, a space, occupied by n positive elements (i.e. n transmissions) and m null elements is described by:

$$P(s_1, s_2 \dots s_n) = \sum_m \frac{(-1)^m}{m!} \int \dots \int F(S_1 \dots S_n, S_{n+1} \dots S_{n+m}) \prod_{k=1}^{m} ds_{n+k} \qquad \dots (5.2)$$

where S is the space of all elements. The expression, $F(S_1 \dots S_{n+m})$, in the integrand is the assumed distribution of all elements in the given space. After the appropriate integration over the whole space, we obtain a derived distribution of the elements in which we are primarily interested, that is, the positive elements. It is, therefore, reasonable to regard the result of the integration [that is, $p(s_i, \ldots s_n)$] as a measure of the posterior probability distribution of the n positive elements in the given space. A comparison of the posterior probability that the space is occupied by n positive space elements and the posterior probability that there are k ($\neq$ n) positive space elements in the same given space reveals the existence of a test statistic for the proposed hypothesis. The likelihood ratio test[48,49] is such a comparative test statistic because the likelihood ratio is, in the main, an expression of the relative difference between two likelihood functions corresponding to two different situations which are describable in terms of some probability measures.

If $p_n(s_1 \ldots\ldots s_n)$ and $p_k(s_1 \ldots\ldots s_k)$ are, respectively, the posterior probabilities that the space contains n and k positive space elements then a test statistic for our hypothesis can be chosen from among the following:

$$\left.\begin{aligned} \beta &= \frac{P_n(s_1 \ldots\ldots s_n)}{P_n(s_1 \ldots\ldots s_k)} && \ldots(5.3) \\ \rho &= \int_{S'} |P_n - P_k| \; dp(s) && (5.4) \\ \alpha &= \int_{S'} (P_n - P_k)^2 \, dp(s) && (5.5) \end{aligned}\right\} P_i = p(s_1 \ldots s_i), n{=}n,k.$$

The merit of a test statistic is reflected by the ease with which it can be computed and by its sensitivity to small differences between the posterior probabilities. It is obvious that test statistics in (5.4) and (4.5) illustrate the concept of the distance between the two posterior probabilities which are being examined and they are widely used. In fact, the test statistic in (4.4) is analogous to the mean-square error[14] statistic which is often employed when the essence of testing is to indicate how far a known result of some trial has departed from the desired or the true result.

4.7.6 Approximation to the Posterior Probability

The statistic, ρ, can be regarded as an index of the 'distance' between $P_n(s_1 \ldots s_n)$, the probability that there are n transmissions in the given space, and $P_n(s_1 \ldots s_k)$ the probability that there are k transmissions in the same space. In other words, ρ, is a measure of the difference between the two probabilities and it is equal to zero when $n = k$. In Chapter 3, the concept of distance between the two sides of a cross-correlation function was introduced and developed; in order to show how useful it is as a discriminant of two time histories of one and the same frequency spectrum. The greater the distance between the two sides of the correlation function, the greater the likelihood that the two spectral configurations of the same frequency space under examination are different with respect to the transmission signal contents. If the distance between the two sides of the correlation function has the same probability of occurrence as the distance between the posterior probabilities which describe the frequency space, then we can use the distance, ρ, to show a variation in the configuration of the frequency space and, therefore, a change in the transmission content. This is the justification, based upon the concepts of probability theory, for using the cross-correlation function as a sufficient means of detecting changes in the space configuration.

As a very useful tool, the correlation technique has found many useful and general applications in the different branches of communication engineering. The obvious difference between the classical use of correlation techniques and the one being presented in this thesis lies in the fact that the classical correlation techniques are associated with the computation of the likelihood ratio. In this thesis, however, we have presented a unified theory of random points in which a new concept of the distance between the two sides of the correlation function has been introduced and developed in order to discriminate between two time histories of a given space. In this development the cross-correlation function has been directly related to the posterior probability distribution, in the sense that the "distance" between the two sides of cross-correlation function is comparable with the distance between the two posterior probabilities and, furthermore, express the changes that have occurred in the search space as far as the transmission contents are concerned. It should be noticed that the relationship between any two time histories of a frequency space can be described adequately in terms of appropriate probability functions. The correlation function can be regarded as another form of the mathematical representation of similarity (or dissimilarity) between any two time histories under consideration. The cross-correlation can, therefore, be considered as a useful approximation to the more analytic function (from the point of view of probability theory) of probability distribution, which is generally more difficult to generate. Hence, using the cross-correlation function in the same way as a histogram we have established a new technique of surveying a given space and detecting changes that are likely to occur in it. These changes are usually those that affect some important parameters of the space which, in this problem, are the transmission contents.

The test statistic given in equation (5.3) is the well known likelihood ratio method of discriminating, in a linear way, between two or more regions of a decision space, and a great deal about this is already known from the literature on it.[50,52] The test statistic, β is the familiar minimum error test and the reader is referred to the appropriate texts on this subject.[53,54]

When the test statistic given in equation (5.4) is used the result of the comparison procedure, expressed in terms of the distance between p_n and p_k, is a pure number. One assumption implied in the integration which is depicted in equation (5.4) is that a change in the configuration of the space is equally likely in any portion which is being examined. This assumption is made for the sake of simplicity.

In the ideal situation, the integral, ρ, is different from zero only when n is different from k; otherwise it is always zero. But in any practical system it is natural to expect that the results will never be ideal and that ρ will always be some positive number regardless of whether a change in the space configuration has occurred or not. In other words, there is always a finite probability of inferring from the results of the integration that a change in the number of transmissions has occurred when, in fact, the number has remained unchanged. This finite probability, conventionally known as the false alarm rate, must always be specified in any given problem in such a way that the probability of making the correct decisions is always maximised. Furthermore, if our decision making process is to be made automatic, then it is appropriate to lay down a simple rule like the following:

if $\rho > \rho_x$ then the space configuration has changed and

if $\rho < \rho_x$ then the space configuration has not changed.

In this case ρ_x is some appropriately chosen threshold consistent with the specified false alarm probability; ρ_x is also chosen in such a way that the probability of making an incorrect decision is always very small.

Fig. 4.8 shows a block diagram of a simple system that will detect changes in the configuration of the space being examined. The two sources shown in the diagram generate, in some fashion, the probabilities which describe the locations of the space elements in the neighbourhood of some point X_i. The output of the device marked ρ is the statistic defined by

$$\int \left| P_n - P_k \right| \; dp(s) = \rho$$

This test statistic, computed in a manner approximating equation (5.4) is applied to the input of the threshold detector which makes the binary decision on ρ.

The decision rule is that the space configuration has changed if $\rho > \rho_x$ and that $\rho < \rho_x$ implies no change in the space configuration.

4.8 Summary

In chapter 3, it is shown that changes in the configuration of the search space are revealed by the lack of symmetry in the correlation function that is generated by correlating two time histories of the space. The correlation technique is used in comparing two time histories, f_1 and f_2, and then by generating the function ρ, we are able to gain information about the appearance or disappearance of transmission signals in the search space. The random point theory developed in this chapter has been presented to indicate the justification for the suggested search technique. It shows that

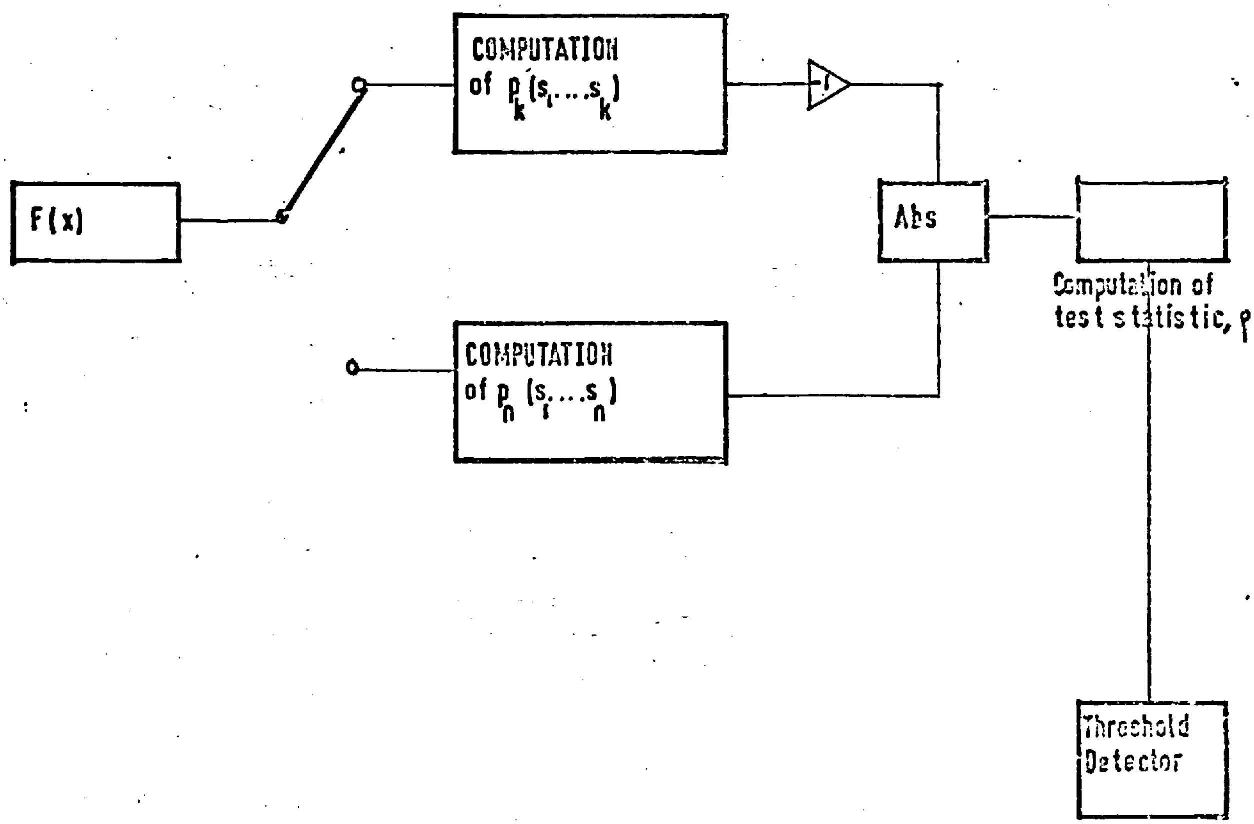


Fig 4.8 A Detection system using a test statistic

we can find it mathematically appropriate and computationally feasible to remove a set of unwanted quantities from a larger set of many quantities by means of a integration procedure. Thus $P_n(s_1 \ldots s_n)$ is derived from $F(s_1 \ldots s_{n+m})$ by performing the mathematical operations indicated in the equation below.

$$P(s_1 \ldots s_n) = \sum \frac{(-1)^m}{m!} \int \cdots \int F(s_1 \ldots s_{n+m}) \prod_{k=n}^{n+m} ds_k$$

The function, F, which is the total probabilistic description of the search space is equivalent in many important ways to the sample function, f, which provides us with a temporal picture of the search space. A linear operation on two sample functions to produce the correlation function is conceptually identical with that involved in the generation of $P(s_1, \ldots s_n)$ from $F(s_1 \ldots s_{n+m})$.

As we see in chapter 3 and again in section 4.7 of this chapter, a gain of information about the arrival or withdrawal of transmissions in the given search space is provided by a direct comparison of the two sides of the correlation function. In terms of the posterior probabilities, $p_j(s_1 \ldots s_j)$ $(j = 1,2,\ldots)$, a gain of information about similar events is derived by directly computing the 'distance' between $P_n(s_1 \ldots s_n)$ and $P_k(s_1 \ldots s_k)$.

The notion of distance between P_n and P_k on which we base our decision about the probable changes in the configuration of the search space is shown to be equivalent to the likelihood ratio employed in constructing a Bayes' solution to a decision problem.

The likelihood ratio of $P_n(s_1 \ldots s_n)$ to $P_k(s_1 \ldots s_k)$ expresses the relative probability that the events described by these quantities are similar. In the same manner, the 'distance' between P_n and P_k is a measure of closeness, and distance in this context is not to be understood in the ordinary Euclidean sense. For our purposes the concept of distance is useful because it is a real valued function and readily allows the ordering of the events according to the degree of their closeness to each other. We therefore, see that our technique of searching the space for the arrival times(or withdrawal times) of the various transmissions within a given space has been based on the appropriate concepts which are strongly suggested by the probability considerations.

Chapter 5

NOISE STUDIES

Chapter 5 NOISE STUDIES

5.1 Introduction

When the transmissions of the type described earlier impinge on the antenna of a receiver their detection is made uncertain by the simultaneous presence of the random fluctuations called noise. If the voltage or current at some stage in the reception process is recorded as a function of time, the record would display an irregular appearance and there would be no simple way of predicting the values of the fluctuating voltage or current. Furthermore, records which are obtained from the same receiver at different instants of time and records which are obtained at corresponding stages of many receivers would differ in their detailed structure. However, certain average properties of these records would be nearly the same, and by studying a large number of these records the average behaviour of the fluctuating phenomenon could be described statistically. Such a study is important because from it a quantitative measure of the 'expected departure' from the average behaviour can be computed and used to predict values of the fluctuating voltage or current. Also an assessment can be made of the noise structure which is an important factor if the noise effects in the proposed search technique are to be made negligible.

In the present analysis, noise will be treated as one composite source of interference and no distinction will be made between the various types of noise as discussed in Chapter 2. For a full assessment of the influence of composite noise on radio communication it is necessary to know the variation, as a function of time and frequency, of both the level and structure. In order, however, to assess the noise structure and evaluate the noise effects on radio communication the time scale of amplitude variation and the cumulative amplitude distribution must be known.

5.2 Sources of Noise in Radio Communication

One of the main obstacles to obtaining reliable signal transmission through the ionosphere is the atmospheric noise. This type of noise is caused largely by the small electric discharges which take place in the upper layers of the atmosphere and thereby generate radio waves in the form of very sharp sudden pulses.[55] Extensive studies[56] have shown that atmospheric noise is an additive disturbance and extrinisic to the physical communication medium. When atmospheric noise is picked up on the receiving aerial it gives rise to short crackles of varying intensity and incidence; viewed on an oscilloscope it appears as series of pulses with short but varying durations.

Spectral examination of the pulses reveals the fact that atmospheric noise has a distribution of energy which decreases at high frequencies. It is, therefore, not surprising that the disturbing effects of atmospheric noise on radio signals are not as serious at the low frequencies as at the high frequencies. In fact, at frequencies below 30 mHz, atmospheric noise is not a very substantial source of interference in radio communication.

In the tropics where thunderstorms occur with a frequency which is dependent on the times and the seasons, the incidence of atmospheric noise is also known to exhibit corresponding variations with the times and the seasons. A quick and a reliable method of determining the course of atmospheric noise consists in taking bearings on the atmospherics received at distant points(57,58); such a method can be used to locate the origin of the atmospheric noise. Tremellen and Cox,(59) by means of an empirical study, have shown that the estimable levels of atmospheric noise caused by thunderstorms decrease with increasing latitude.

At frequencies between 20 and 100 mHz extraterrestial noise is a significant contribution to the total received noise. As the name suggests, this noise component is due chiefly to electric fields produced by disturbances which originate outside the earth or its atmosphere The sources of the disturbance are known to include the stars, the ionised interstellar matter and the invisible concentrated sources called radio stars.(60) An examination of the extraterrestial noise has shown that it has the characteristics of random noise with a continuous frequency spectrum and that its intensity at the earth's surface is determined mainly by ionospheric absorption and the direction of arrival.

Another type of noise encountered in radio communication is that generated by electric machinery in industrial plants.[61] Transients effects produced by the making and breaking of a current are the direct causes of this type of noise which has the characteristics of an impulse noise and, because of its coherent frequency components, differs from random noise. The peak amplitude is proportional to the bandwidth of the receiver and not to the square-root of the bandwidth as in the case of background atmospheric noise.[62] Reference to appendix A will make this point clearer.

5.3 Effects of Noise on Radio Communication

In any radio communication system like the one illustrated in the block diagram of fig.5.1, the receiver serves as stage where evidence of a transmitted message is gathered and then analysed in order to recover the message. The source selects, in accordance with a prescribed alphabet, the messages which are encoded and transmitted as physical signals usually in the form of waveforms. The transmitted signals travel through the communication medium to their destination where they are received, operated upon and interpreted, again in accordance with an established alphabet.

In the absence of noise or any other disturbance the recovery of the desired message from the transmitted physical signals can be achieved

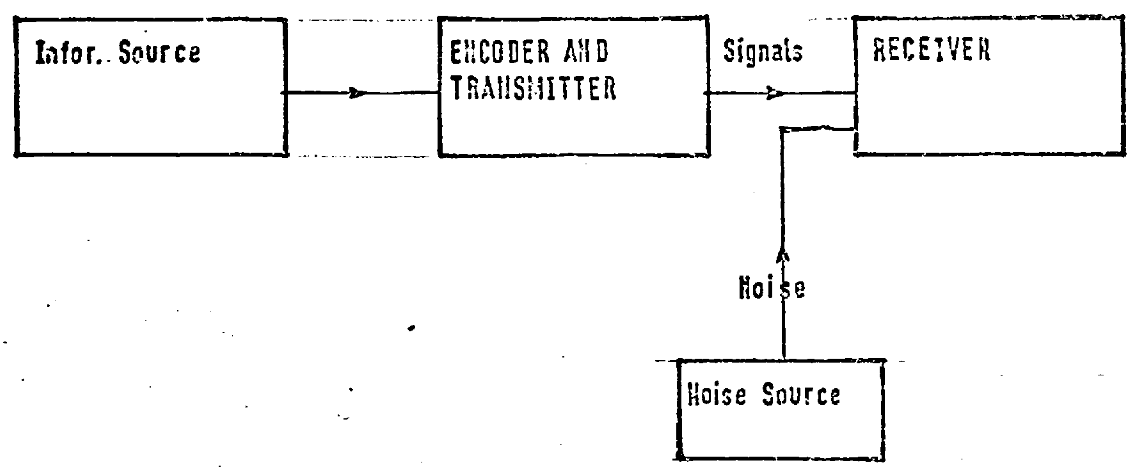


Fig 5.1 Communication in the presence of noise

without errors. However, the presence of noise is unavoidable and all the real-time physical signals are subject to noise disturbances which are beyond the control of the transmitter or the receiver. The effect of the noise on the transmitted signal is destructive and sets the ultimate limitation on the information capacity* of the radio communication channel. Any knowledge about the information source, or signals received from it, provides no information about the moment-by-moment noise values. However, communication theory has demonstrated the fact that solely from the knowledge of the statistical characteristics of the noise source, the average rate of information loss can be determined.(63) Obviously, by reducing the rate of information loss, the maximum extraction of the transmitted message can be obtained.

Another way in which the performance of a communication system is limited by noise is expressed in terms of the signal-to-noise ratio of the receiver which is the terminal of the 'travelling message'. Generally, if the noise voltage induced at the aerial of the receiver is less than a certain amount, the interference introduced by it is negligible. Expressed in terms of the signal-to-noise ratio, the performance of a received communication system is poor when the

---

*The information capacity, C, of a channel of width W Hz is mathematically expressed as(64)

$$C = W\log_2 (1 + P/N) \text{ bits/sec}$$

where P and N are the average signal and noise powers, respectively. Generally this represents the maximum number of bits which a channel can transmit in one sec with a vanishingly small probability of error. It is generally true for unconstrained systems and provides a standard for assessing the efficiencies of practical systems.

ratio is small. For any communication system there is a critical signal-to-noise ratio at the receiver input below which the ratio at output falls rapidly. Thus frequency modulated systems are capable of a much reduced output noise level compared with amplitude modulated systems for a given input.

5.4 Measurement of Radio-Noise Influence

The essence of any scientific attack on any problem is the measurement of a 'magnitude'. In the case of radio communication, attempts have been made to measure the noise power and assess its average behaviour in a chosen channel.

The simplest way to measure the noise power is to use a linear distortionless amplifier (so that the output is linearly dependent on the input voltage) and a quadratic detector which measures the output noise power. However, any non-linear amplitifier in combination with a non-quadratic detector can be used if the proper precautions are taken. In this case,it is necessary that the output measuring device be calibrated directly in terms of the output noise power. The output noise power can then be read in a straight-forward way from a calibration curve.

The bandwidth of the noise amplifier should be small compared to the bandwidth of the input circuit, but it should not be too small as this would then require a large time constant of the detector.[65] At times, the small selective bandwidths that are required for accurate measurement are not obtainable at the frequencies at which the measurement is performed. It then becomes advisable to provide the amplifier with a mixer stage that converts the noise signal to a lower frequency at which selectivity is more obtainable.

Noise power is an important parameter in the field of radio communication and is used in assessing the interfering effects of radio noise on a world wide basis.[66] The basic parameter used by the C.C.I.R. is the effective noise figure of the aerial which is related to the root-mean-square value of the noise envelope by the following expression:[x]

$$F_a = E + 66.8 - 20\log f$$

where E is the r.m.s. value of the noise envelope in decibels above 1 microvolt over a frequency band of 10 kHz; f is the frequency in mHz and F_a' is the ratio of the available noise power to the thermal noise which would be obtainable if the aerial were at a reference temperature arbitrarily set at 288°K.

To study the diurnal variations of the noise level the day was divided into six blocks of four hours each. Using an omnidirectional

---

[x]This was first described by a committee[67] which reported its findings to the C.C.I.R.

antenna coupled to a specially adapted receiver twelve observations were made. A reading of the noise level was taken every 20 minutes and from an assembly of such readings the curves shown in fig. 5.2 were drawn.

The generalised curve of the diurnal characteristic shows that from 1 to 5 mHz the night-time noise becomes definitely less than the day-time noise. However, it is dangerous to over-generalise since variations in noise characteristics accompany observations which are made under similar conditions at different geographical locations.

The arrangement of the apparatus employed in making the noise observations can be seen schematically represented in the block diagram of fig. 5. 2 . The noise within frequency band of 10 kHz at a particular centre frequency in the Hf spectrum is tuned in on the receiver and the output level is measured in the stage following the receiver. To determine the actual level of the received noise the output of a noise generator is fed into the receiver through a bank of attenuators which are used in adjusting the level until the same reading as that produced by the aerial noise is obtained.

5.5 Measurement of the Amplitude Probability Distribution

In order to obtain the amplitude probability distribution, a scheme represented diagramatically in fig. 5.3 is used. The noise envelope

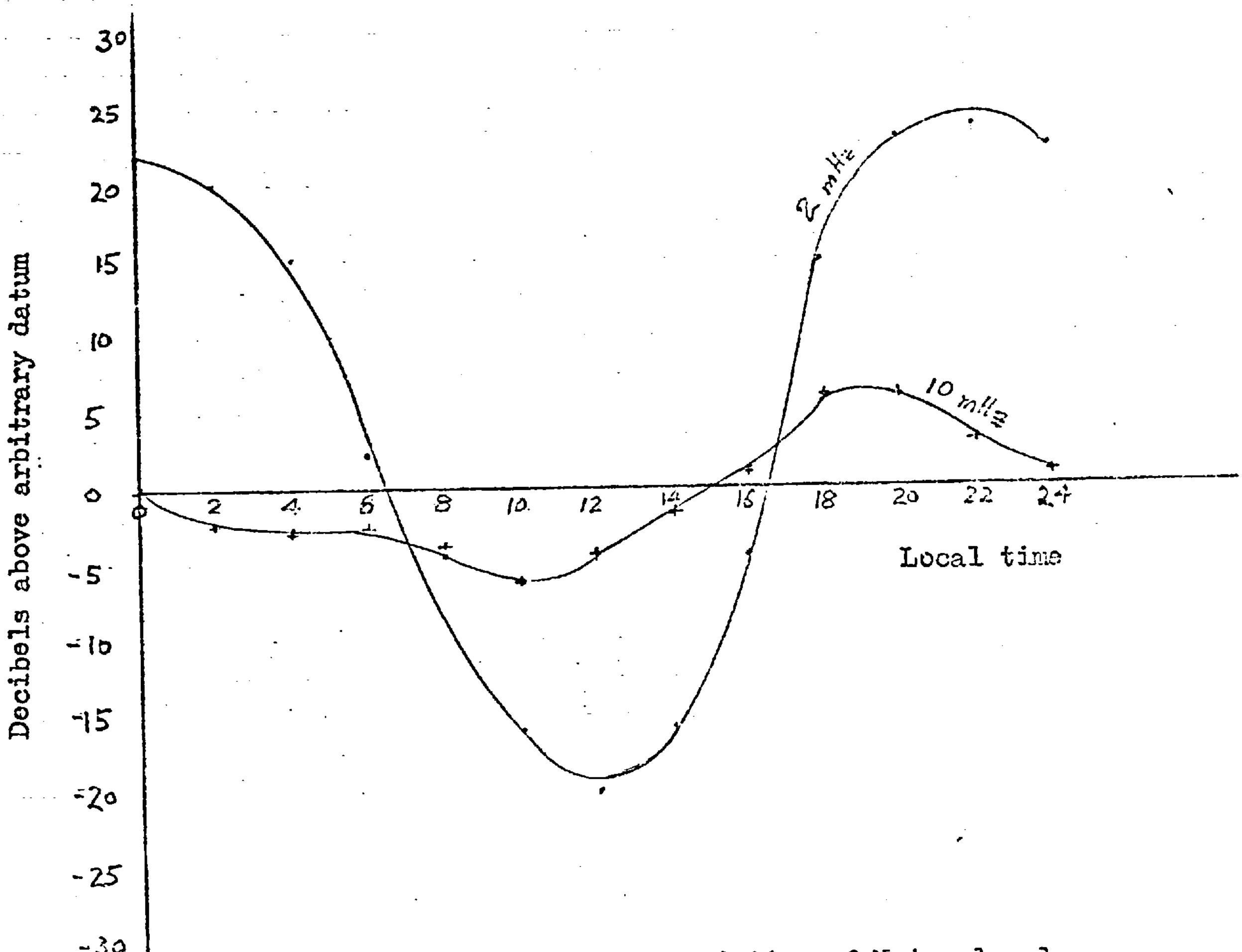


Fig. 5.2 Variation Characteristics of Noise levels

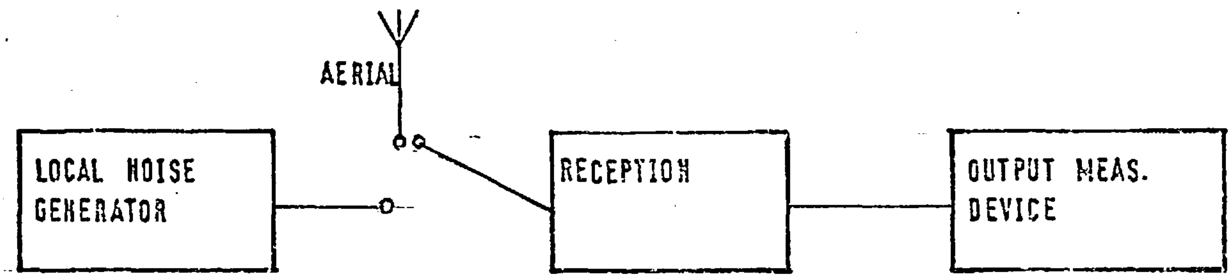


Fig.5.2(A)Measurement of noise level

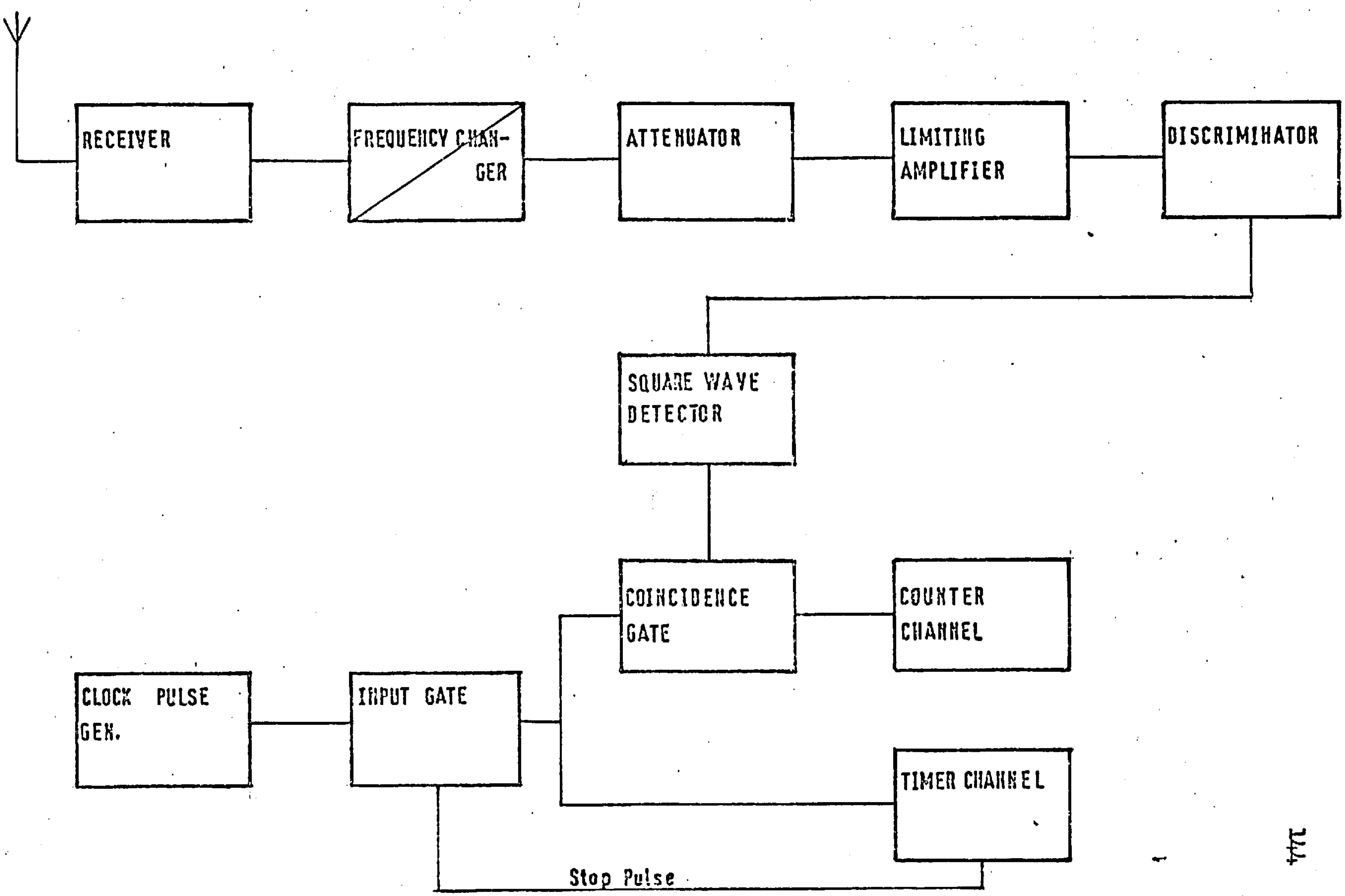


Fig 5.3 BLOCK DIAGRAM OF THE NOISE MEASURING EQUIPMENT

within a band of 10 kHz is obtained by the use of a receiver whose final i.f. output is 10 kHz wide at 3 db points. The receiver used is a commercial type with all the a.g.c.[x] circuits disconnected and provided with an emitter follower stage at the output of the final i.f. stage. The aerial used is a vertical omnidirectional type and an aerial attenuation pad is provided in order to prevent overloading of the whole receiver set-up by exceptionally strong signals.

To compute the amplitude probability distribution (a.p.d.) the received signal is applied to an amplitude discriminator and a limiting amplifier. The discriminator is operated at a fixed threshold but the signal amplitude may be varied by the aerial attenuator. The output from the discriminator passes to a square wave generator (i.e. a box-car generator) where those sections of the noise envelope which exceed the threshold are converted to rectangular pulses as illustrated in fig. 5.4 . The resulting picture is that of a chain of noise pulses. The amplitude of the noise pulses varies randomly from zero to infinity and a digital computer can be used to count those pulses which have an amplitude above a certain threshold level. The higher the level is, the lower the counting result, τ_c, will be. Because of the random nature of the noise, the measurements are associated with an error; the relative standard deviation of the error can be calculated. This can be made arbitrarily small percentage-wise by increasing the counting time. In practice the error is small (percentagewise) when BT $>>$ 1, B being the bandwidth of the noise and

---

[x]a.g.c. stands for automatic gain control.

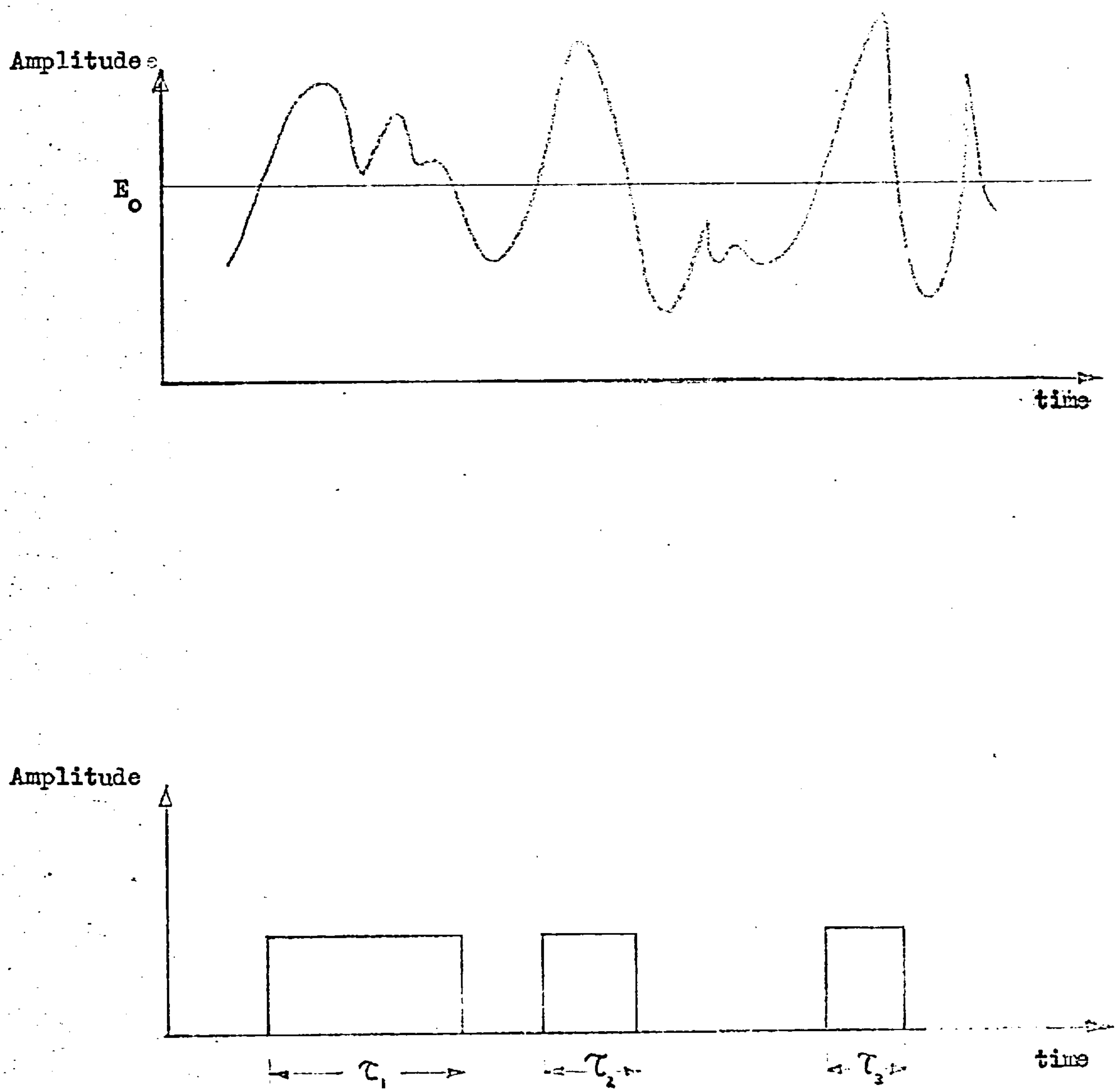


Fig. 5.4 A method of calculating amplitude distribution density (a.p.d.)

T the counting time.

It must be remembered that the object of the experiment is to calculate the probability that the amplitude, E, of the noise is greater than E_o, a reference level that is pre-set by the use of the aerial attenuator. By measuring the durations of the times, $t_i (i=1,2,..)$, during which the noise level is equal to or greater than E_o and comparing this to the total observation time, the probability, $Pr(E \quad E_o)$ can be calculated. If the total time of observation is T, then

$$Pr(E \quad E_o) \quad = \quad \frac{t_i}{T}$$

Also, noting that

$$Pr(E \quad E_o) + Pr(E \quad E_o) \quad = \quad 1$$

it is concluded[x] that

$$Pr(E \quad E_o) \quad = \quad 1 - Pr(E \quad E_o)$$

Errors that are to be introduced by the time constant of the detector can be avoided by converting the received noise at any frequency to an intermediate frequency at 100 kHz.

In fig. 5.5 the fraction of the time during which the noise level is equal to or above a given reference level is plotted as a function of the reference level. The plots shown have been obtained for noise at 1, 5, 7.5 and 20 mHz, using ten reference levels.

---

[x]For a more rigorous account of the distribution and its type see Appendix B.

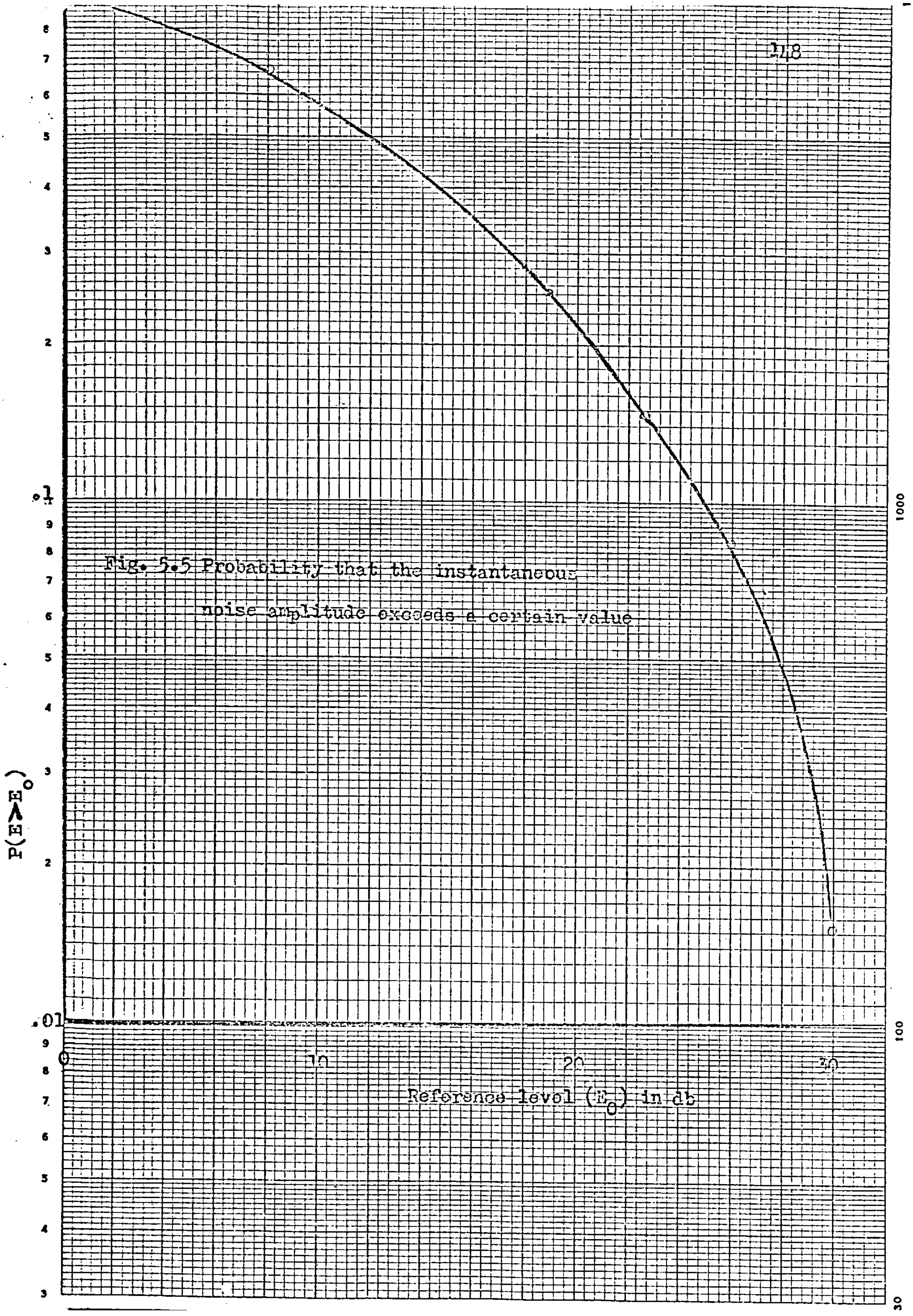


Fig. 5.5 Probability that the instantaneous noise amplitude exceeds a certain value

5.6 The Ratio of the Signal Power to the Noise Power

When the noise and a signal appear at the output of the receiver the disturbing effect, or the annoyance value, of the noise depends not only on its characteristics and the signal-to-noise ratio, but also on certain subjective and objective factors involved in the method of observation. Steundel(68) who studied the aural sensation produced by a succession of impulses discovered that the apparent loudness of an impulse is determined by the pressure integrated over a small fraction of a second in the region of its peak value.

However, when the properties of the ear are ignored, as done in the case of instrumental observation, the fluctuation of the indication due to noise is the relevant measure of the interfering effect. In the presence of a steady signal, unperturbed by noise, the indication is ideally steady and only related to the signal energy through the dynamics[x] of the energy measuring device. Therefore, a measuring device constructed with a hypothetical bypass for fluctuations due to noise can be used to measure only the steady indications due to steady (or slowly varying) signals. An arrangement for such measurements has been shown in the block diagram of fig.5.6 , and it is obvious that an instrument of this type will be

---

[x] The dynamics of a measuring device is usually expressed by the output-input relationship. Thus if u is the output and i is the input, then $u = f(i)$ expresses the dynamic relationship between u and i.

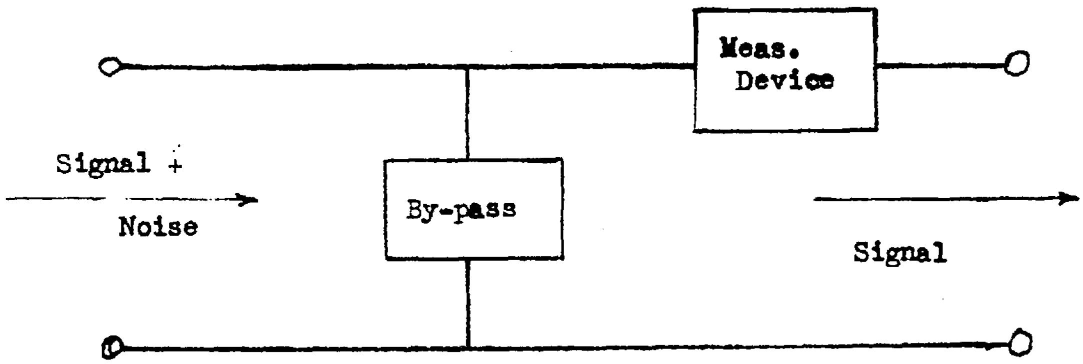


Fig. 5.6 An arrangement for hypothetical Noise-free Measurement

quite unresponsive to noise, indicating only the presence of wanted signals. With this arrangement the ratio of the power of the signal to that of the noise becomes unimportant since all of the noise is fed into the by-pass and hence does not affect the desired output.

In the present investigation an attempt has been made to construct an equipment capable of ignoring the continuous presence of the noise and responding only to the signal which "falls into the field of search". This field of search comprises the total frequency band in which the time of initiation or the cessation of a signal is to be determined. The indicating instrument, for all practical purposes, will remain in some particular state irrespective of the noise intensity, but will immediately change into its second possible state whenever there is a change in the signal content of the search field. In the following chapters practical steps that are taken perform this operation will be discussed.

Chapter 6

DESIGN OF EXPERIMENTS

Chapter 6 DESIGN OF EXPERIMENTS

6.1 A Biased Sampling Procedure

In the conventional Signal Sampling Theorem* (Whittaker, Gabor, Shannon, Nyquist, etc.), the signal is considered "band-limited" to W Hz (a Fourier infinite time concept). If the band-limited signal has a time duration of T seconds then 2TW samples of it, uniformly spaced at successive intervals of 1/2W seconds, are required for its specification.

Many signals employed in communication fluctuate considerably, having intervals of dc, zero or slowly varying values, intersperesed with regions of rapid change. Hence for communication purposes, signals representing, say, speech or video are regarded as stochastic processes and knowledge of their moment-by-moment fluctuating properties are therefore essential. Unfortunately, by averaging over infinite time as is done in Fourier analysis the moment-by-moment fluctuating properties of the signals are concealed.

In communication science, problems which cannot be tackled non-empirically so long as the temporal fluctuations are so concealed are sometimes encountered. For example, in some search problems, interest is focussed only on some particular events and it then becomes expedient to adjust the rate of sampling so that, in the areas where the objects are likely to be found, more sampling may be

---

*See references (69-70).

done than in those areas where the events are not likely to occur. It is, therefore, feasible to vary the rate of sampling of video signals, using an instrument called a detail detector.(71) This instrument examines a waveform of a video signal continuously and samples it at rates depending upon whether there is little or great picture detail. The different rates of sampling are determined by, and based on, the statistical measurements of picture detail of different types of television picture.

Obviously, what is required in situations like this, is a non-uniform rate of sampling, an idea which strongly suggests the use of a "variable bandwidth". The bandwidth required in any search procedure should not be a fixed quantity but a variable determined only by the characteristics of the objects being sought. In other words, the size of the total search space is only a probable measure of the quantity of objects that can be expected.

Fig.6.1 is a diagramatic illustration of a search space the width of which is λ. If interest were to be momentarily focussed on the ill-defined pulses shown in the diagram (the rough widths of the pulses are L_t ($t = 1, 2, \ldots$) then a biased form of sampling may be employed. The transmission signals represented in the frequency space as pulses can be adequately specified not by taking samples from the whole search space at some calculable rate determined by the width of the whole search space. In fact, since the signals in the frequency space are specified by samples taken only from regions occupied by them, the sampling rate can be made variable. A

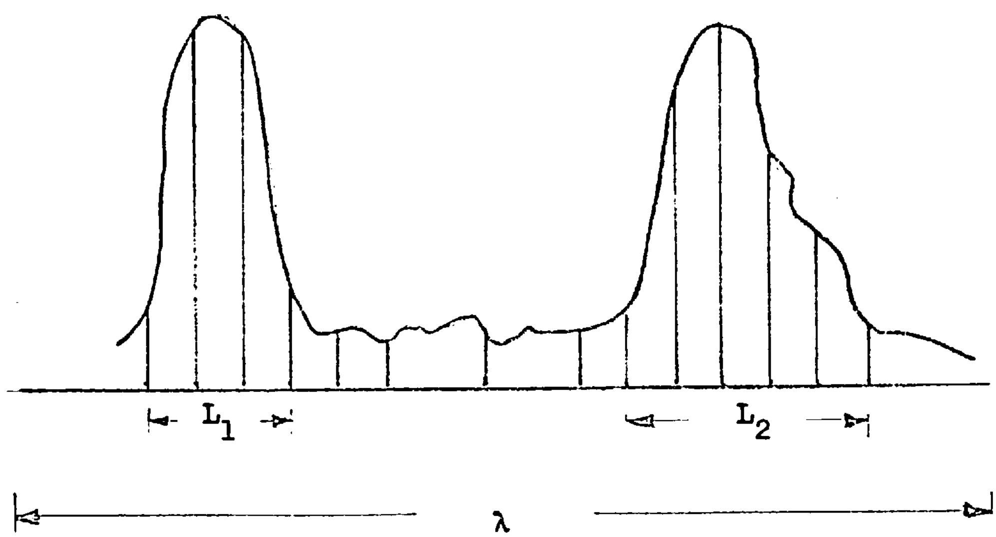


Fig. 6.1 A search field in which biased sampling can be performed

satisfactory variable rate of sampling can be achieved by first assessing, in some appropriate manner, the statistical properties of the search space. When the statistical measurements have been made, the biased form of sampling in the search space can be performed. The specification of the transmission signals located in the space can then be accomplished within some calculable margin of error. A practical criterion for the success (or failure) of the whole operation can be made to take account only of the regions occupied by the signals, that is, a microscopic and not volumetric assessment.

One practical way of achieving biased sampling is by sampling the "important areas" more often than the other regions of the search space where transmission signals are not likely to be found. In situations where many cycles of sampling are desired and feasible, the first cycle of sampling can be done at one appropriate rate; subsequent cycles containing fewer numbers of samples each can then be obtained by making use of the information derived from the first cycle of samples. However, for reasons suggested by time saving considerations and other practical factors it is not always advisable to allow more then one cycle of sampling. Other forms of biased sampling must then be explored in order to derive the full benefits of biased sampling.

Fig. 6.2 shows a portion of the Hf spectrum in which the important areas containing transmission signals as well as noise are clearly visible and marked A^+. The "unimportant area " appearing between two successive "areas of importance" contains noise alone and

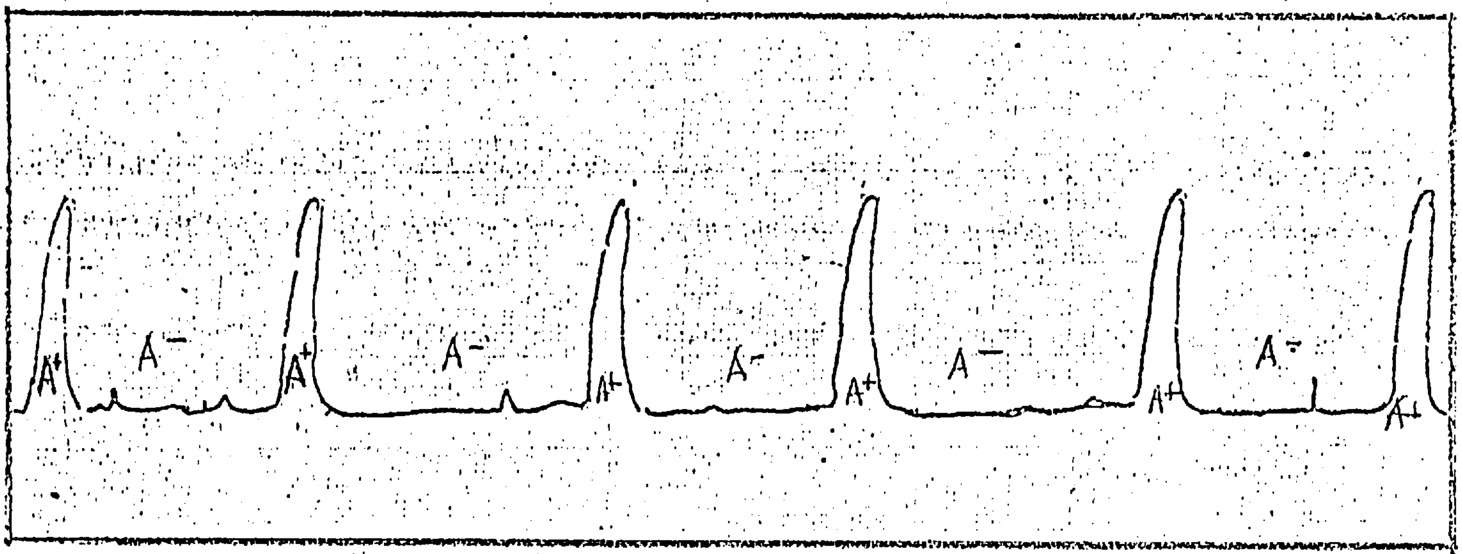


Fig. 6.2 A plot of a spectrum showing 'important' and 'unimportant areas'

is marked A^-. If the locations of the "important areas" are known a priori then it is possible to design the biased sampling technique in such a way that more samples are taken from them than from the unimportant areas. (In the figure referred to this idea is illustrated by the closeness of the samples taken from the important areas). But since the a priori information is usually not available, a compromise technique of biased sampling becomes necessary. In this method of sampling, samples are taken at a rate which lies between the rate required for the "important areas" and that for the "unimportant areas". The sampling is then performed at a uniform rate which, although slow for the "important areas", is still good enough for the specification of the signal contained in the important area. Also, since it is computationally easy and usually more convenient to sample at a uniform rate, the sampling is performed at some uniform rate determined by the appropriate number of samples required to specify, not the whole search space, but the narrowest possible band-limited transmission signal.

Furthermore, the need to save time, though an important consideration, is not the only reason for performing empirically biased sampling. For it can be shown (see Section 6) that the chances of extracting maximum information about the objects "hidden in the search space" are also optimised when biased sampling techniques are properly employed. Again, from the point of view of information theory, the process of biased sampling helps to conserve information about the desired objects of search and at the same time destroys only the bogus information about the unwanted objects.

By long term spectral analysis or cross-correlational examination of the samples from the search space, the appropriate information concerning the essential characteristics of the space will be obtained and then used to identify the wanted signals. The identification is achieved by examining changes in the space characteristics which reflect corresponding changes in the degree of correlation between the various groups of samples collected from the one and the same search space. When the degree of correlation falls below some acceptable value it is concluded that either a new transmission signal has occurred within, or an old one has disappeared from, the search space being examined.

In most cases of interest, it is not known a priori whether or not any number of transmission signals exist in the search space. A preliminary search procedure is then required in order to "learn" the state of the space. This learning process takes time and may be difficult but, as will be shown later, this difficulty is not too great and there is ultimately a considerable amount of time saved by employing this general method of biased sampling. It is also important to note that the method of biased sampling has the inherent ability of reducing the number of samples required to specify a transmission signal in the frequency space and hence can be regarded as a process of removing redundant samples. The actual number of dimensions of the signal is not affected.

6.2 Choice of Scanning Time of a Sector when the Observation Time is Fixed

In searching for signals located in one or several small intervals of a frequency space which are scanned successively, the overall time, T, of observation may be a fixed quantity. The time, t, allotted for the scanning of the ith interval of the space will then be related to T in such a way that if there are m frequency space intervals, then

$$\sum_{i=1}^{m} t_i = T$$

The time, t_i, can also be sub-divided into n_i equal intervals, x_i, so that

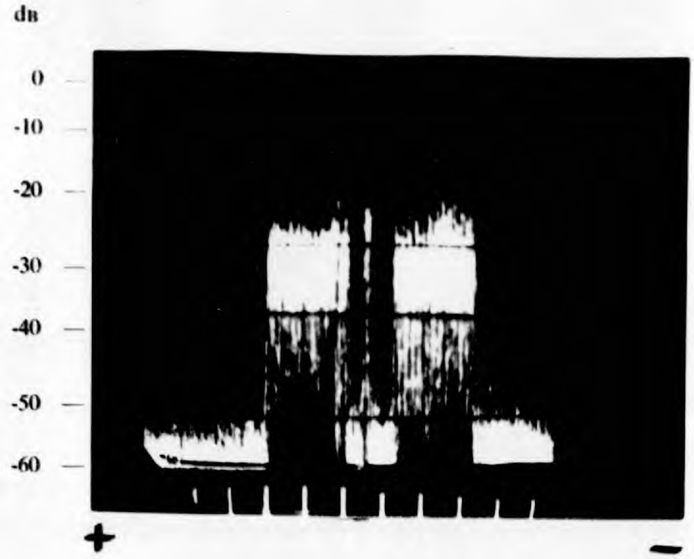
$$n_i \, x_i = t_i \quad (n \geqslant 1)$$

Having done this the decision on the presence or the absence of a signal in the ith space interval can be taken either on the basis of the observation stored during the time t_i or after each of the n_i observations. However, when the time t_i is sub-divided into the n_i equal intervals ($n_i x_i \simeq t_i$) the probability of detection is observed to be unaffected by the sub-division if the signal to be detected is coherent and undergoes fluctuations which are fast compared to the ratio t_i/n_i (i.e. x_i). Thus if in the detection of constant amplitude signals the coherence is not preserved from observation cycle to observation cycle then the sampling (i.e. the sub-division of the time t_i) will lead to a decrease in the detection probability. If coherence is preserved, however, by utilizing the evidence obtained in a previous observation time the probability of detection can be made to increase or, at worst, remain constant.

6.3 A Study of the Frequency Space Characteristics

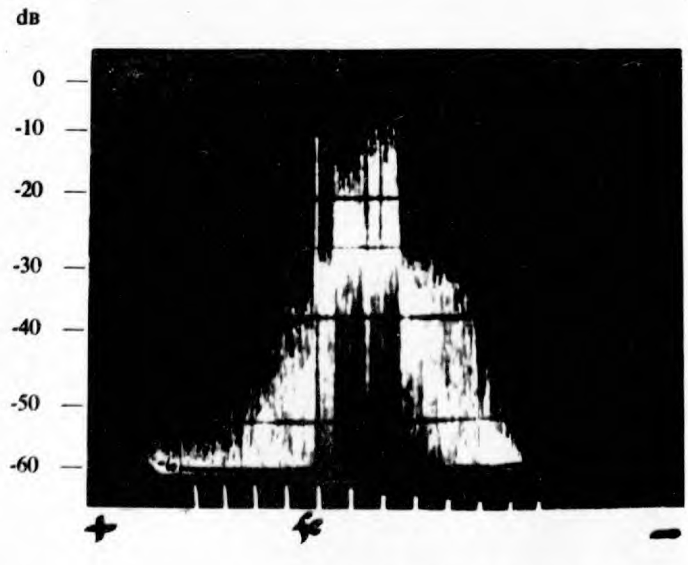
Figs. 6.3 and 6.4 illustrate some spectral configurations of which can exist in a frequency space under examination. These different configurations, in the usual terminology of signal transmission, represent only a few of the many types of radio emissions which are employed in radio communication. (To obtain pictures of these spectral displays a photograph of each spectral configuration is taken in three stages. First, the top part of the display, 0 to -30 db, is photographed; the gain of an associated amplifier is then adjusted to + 30 db and the lower half of the display is obtained. Finally, a standard frequency modulated with 1 kHz harmonics replaces the signal under examination to provide an accurate scale at the base of the display. The frequency range can be extended and the sensitivity also increased by the addition of suitable frequency changing circuits.)

As an example of the picture of spectral composition of a frequency space, fig. 6.3a is a representation of an amplitude modulated signal field (a telephone signal). The two independent side bands of the signal can be seen on either side of the unsupressed carrier. Fig. 6.3b



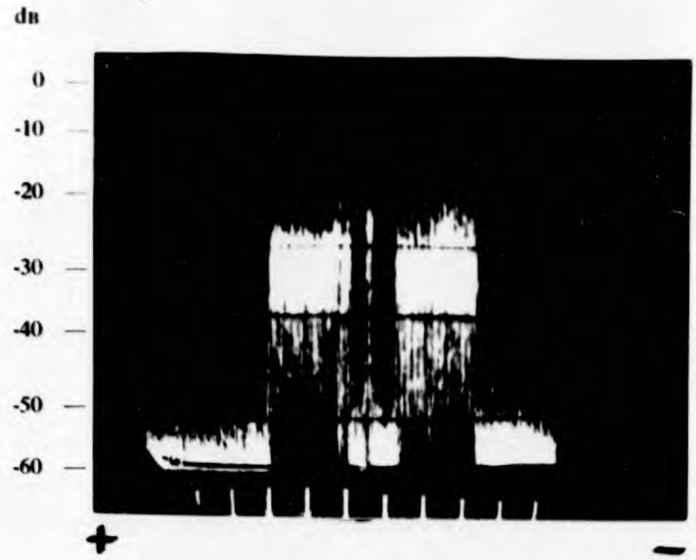
Frequency 1 Kc/s per division.

Fig. 6.3a An AM signal field with unsuppressed carrier



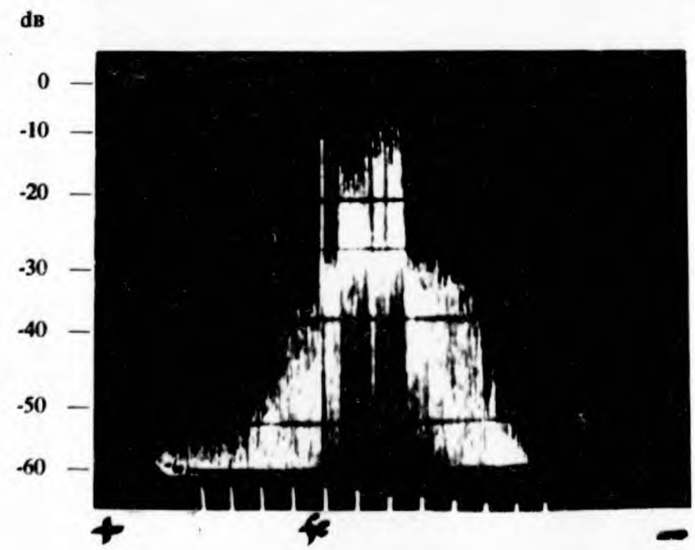
Frequency 1 Kc/s per division.

Fig. 6.3b An AM signal field with suppressed carrier (s.s.b)



Frequency 1 Kc/s per division.

Fig. 6.3a An AM signal field with unsuppressed carrier



Frequency 1 Kc/s per division.

Fig. 6.3b An AM signal field with suppressed carrier (s.s.b)

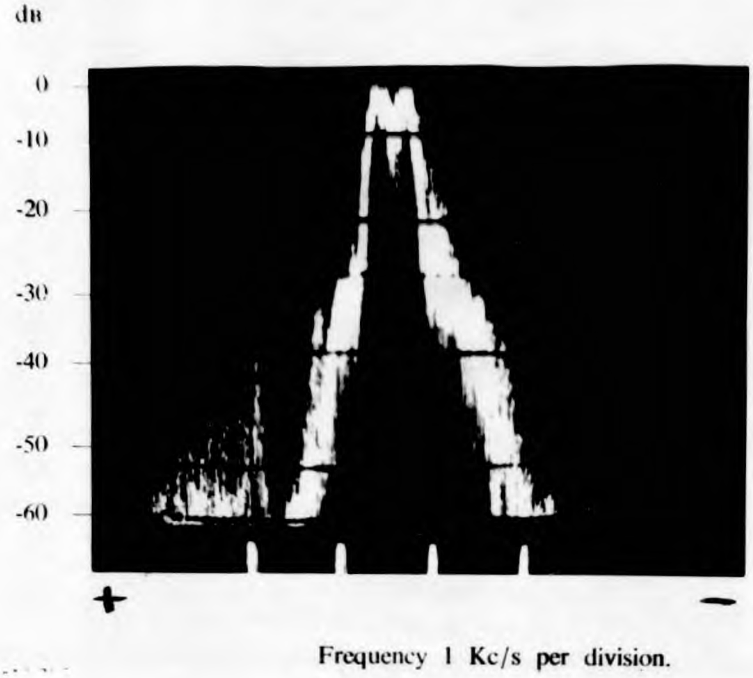


Fig. 6.4a An FM signal field

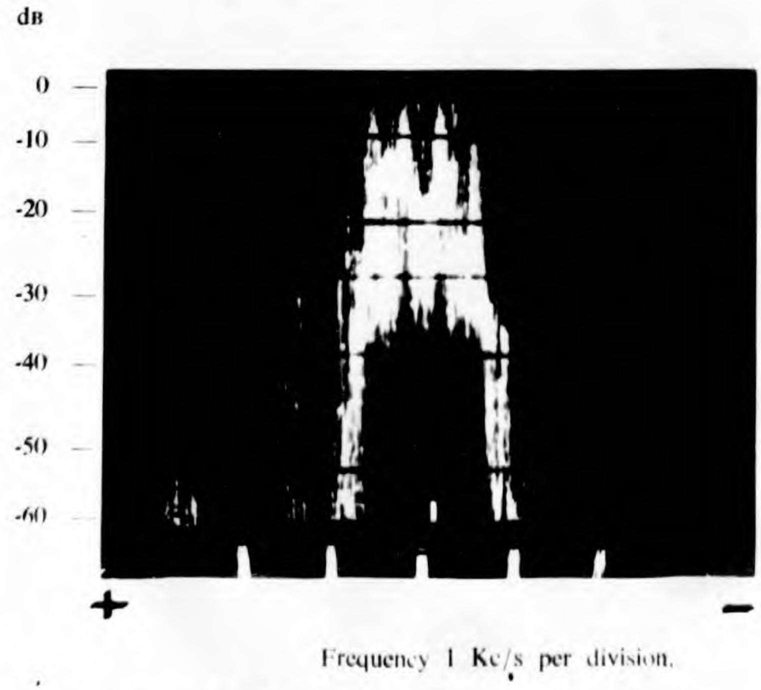
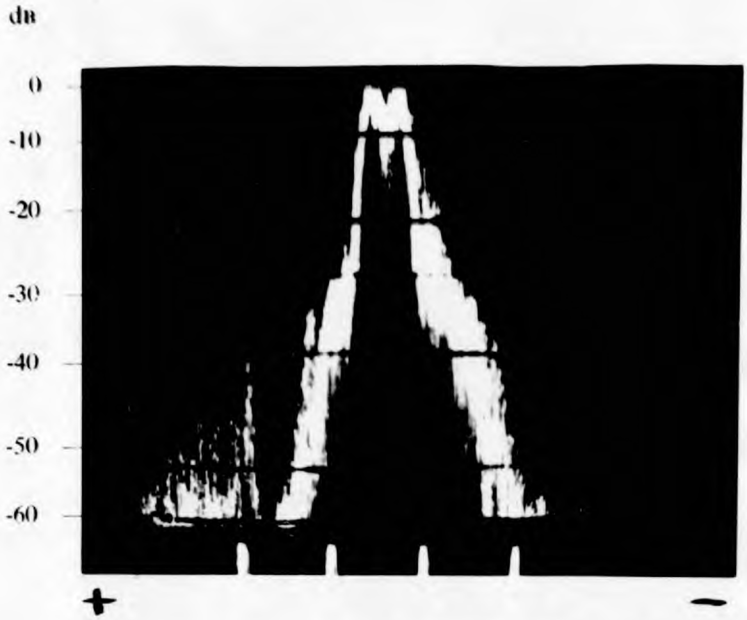


Fig. 6.4b An FM (telegraph) signal field



Frequency 1 Kc/s per division.

Fig. 6.4a An FM signal field

Frequency 1 Kc/s per division.

Fig. 6.4b An FM (telegraph) signal field also shows an amplitude modulated signal but with a single side band and illustrates a typical signal field of a multi-channel voice frequency telegraphy system. The spectral configuration of a frequency modulated signal field is shown in fig.6.4a and fig.6.4b is the field state of a frequency modulated signal, with four-frequency diplex as in telegraphy. Before the emergence of any of the types of transmissions that are possible in the frequency space being examined the field state of the space represents the amount of noise that can be detected by the receiving equipment and so be found in the accompanying measuring devices. Such a noisy field state is shown in fig.6.5.

In the absence of any of the active transmissions mentioned above, any frequency space under examination can be characterised by parameters which are associated with noise alone. Hence in order to provide for the possibility of detecting an active transmission which may 'emerge', the frequency space must be examined continually. It will then be possible to detect as soon as practically possible when the noise parameters of the 'noisy' field state have changed sufficiently to indicate the emergence of an active transmission (or a group of active transmissions). Conversely, a field known to contain an active transmission can also be examined continually in order to detect when the transmission ceases operating in its allotted portion of the frequency space.

In the present analysis as previously explained, it was found necessary and expedient to scan the field and obtain sample values from it. The samples thus obtained would specify the particular field

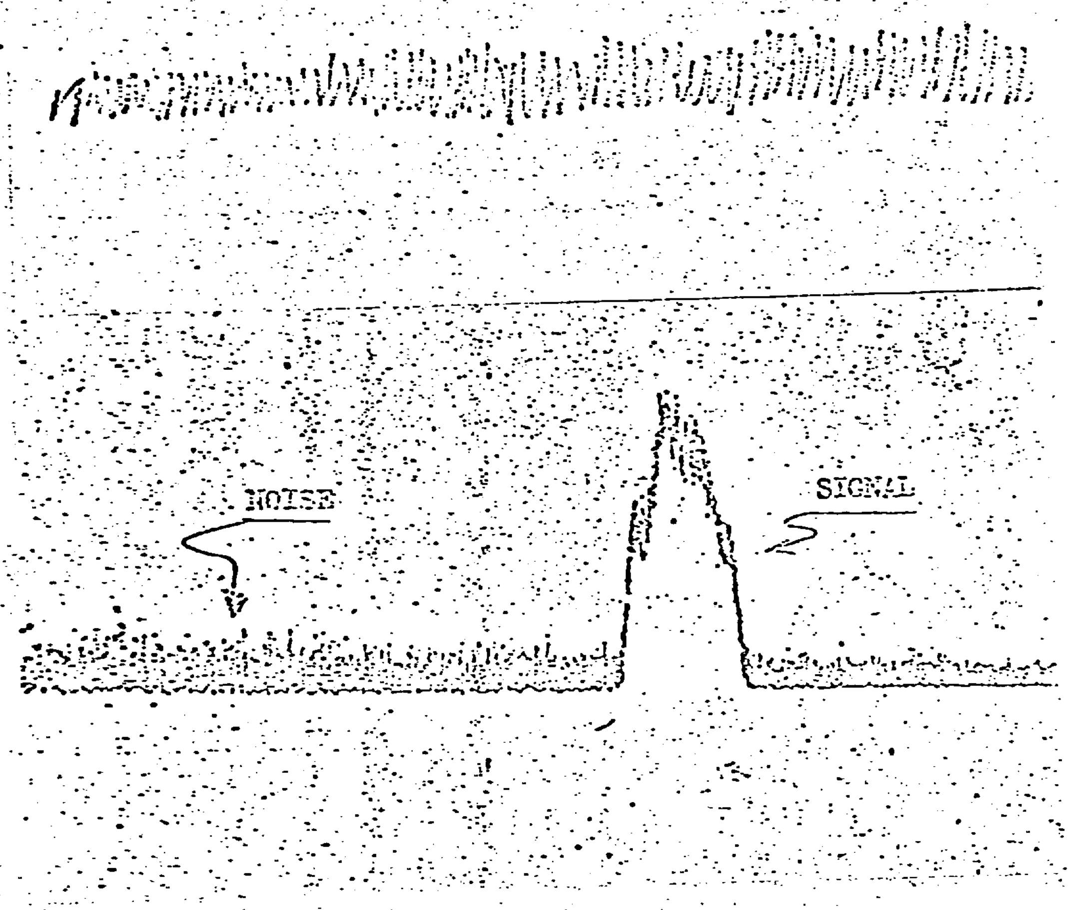


Fig. 6.5 A spectrum of noise and a transmission signal

under examination at the particular time when the samples are obtained. At a later time the field is scanned for the second time and another group of samples assembled for the analysis. It is found that by comparing (by means of correlation analysis) the two groups of samples assembled from the frequency space under examination changes in the transmission content of the space can be computed. The essential object of the procedure is to detect a calculable change (in the field) whenever it occurs in as short a time as possible, subject to the normal, prescribed condition on false alarm probability. The detection problem is, therefore, one of simple hypothesis testing. The hypothesis that there has been no parameter change is tested against the alternative that there has been a parameter change and the appropriate decision is then made.

As long as the state of the field remains unchanged the sample groups assembled from it can be analysed in arbitrarily selected pairs and shown to produce a definite form of a correlation function. (cf. analytic continuation[73]). Hence the persistence of the one form of correlation function is an indication of the fact that the transmission content of the field has not suffered any change. The arrival, or the cessation, of a transmission can, therefore, be announced when the form of the correlation function is sufficiently different from the expected form "of no change". In chapter 4 the expected form of no change has been given a precise, mathematical interpretation and a measurable deviation from it is used in the process of deciding whether or not new transmissions have appeared in the frequency space under examination.

In the present investigation two selected groups of samples from one and the same field are correlated and the properties of the resulting correlation function are examined and interpreted in order to define the state of the particular field. Thus, if the correlation function is an even function so that the two sides of it are identical with respect to an appropriately chosen axis than the two groups of the samples assembled from the field specify the same state of a field. However, when the correlation function obtained by processing the two groups of the samples is non-even then the state of the field being examined has suffered a change during the second sampling. Also when the correlation function is even but appears as a damped exponential function the indication is that the two groups of samples are from a random or a 'noisy field'.

More generally, the correlation function consists of alternating crests and troughs. In the instances when the distances between successive crests (or troughs) are uniform over a reasonable number of cycles and the correlation function itself regularly drops to some terminal amplitude beyond which there is no significant further decrease, the two groups of samples would appear to have come from the same transmission field. A distinct variation of the above correlation function which was encountered frequently during the investigation is the one in which the distances between successive crests (or troughs) are not uniform and the correlation function is characterised by alternating growth and decay. In most situations this later type of correlation function is not even and hence implies a change in the state of the field from which the two groups of samples are assembled.

6.4 The Computation Process

A proper choice of the interval between successive samples of the field was made after some amount of preliminary study has been made employing artificially generated transmission signals and noise.[x] In order to simplify the relevant computations the nature and the form of the transmission signals used were assumed known.

The samples were obtained at the appropriate rate[xx] of one sample value per 1 kHz[x] by using an analogue-to-digital converter[xxx]

---

[x] White noise was used in these preliminary studies simply because means of generating this type of noise was easily and readily available. Moreover, white noise lends itself easily to analysis.

[xx] By confining the sample space to being quite large but finite, the problem associated with defining a uniform density over an infinite set can be avoided. The rate of sampling in this investigation was equivalent to about four samples per one voice channel which is about 3.5-4.0 kHz wide.

[xxx] The analogue-to-digital converter used in this investigation is another proprietary equipment manufactured by Mullard Co., England. It is used in converting continuous current waveforms into discrete ones by a method of sampling.

in conjunction with a tape perforator which would punch out the digitised amplitudes on a punch tape. The analogue-to-digital converter is designed to sample analogue voltages in the range 0-10 volts and provides the corresponding digital outputs. The digitising is performed by a shift register, and the relevant circuits for producing the digital equivalent of an analogue input voltage are also part of the converter.

For the purposes of the present investigation a sampling pulse of not less than 1 millisecond was required. This satisfied the condition that t_i/n_i is long compared with the signal fluctuation. Digitising at the rate of ten digits per one sample value, a digitising frequency of about 1 kHz was found quite suitable and convenient from the point of view of ease of operation. Also, clock pulses, obtained as an auxiliary output from the panoramic adaptor, were used to synchronise the sampling signals and in this way the problem of having to provide a separate sampling signal facility was avoided. Another related problem which was overcome by using the clock pulses from the panoramic adaptor was that of non-synchonous functioning of the panoramic adaptor and the analogue-to-digital converter. In fact, it was not found necessary to build any separate unit which would provide the required synchronism apart from the cascade of simple multi-vibrator circuits the inputs of which are the clock pulses from the panoramic adaptor. Suitable division of the clock pulse period can then be accomplished by tapping from the appropriate sections of the cascade. In fact, to meet the simple requirements of the method of sampling evolved for the investigation, a suitable arrangement of multi-vibrator circuits (including one Schmitt trigger circuit and one amplifying stage) was used to obtain pulses whose repetition rates are a 50th and, or a 100th, of the rate of the clock pulse derived from the panoramic adaptor.

Assuming the two groups of M samples assembled from the frequency field are X_i and X_j (i, j 1,2, ...,N) the analysis begins with the computation of the serial products

$$\delta_p = X_1Y_{p+1} + X_2Y_{p+2} \quad \ldots\ldots +X_{M+p}Y_M$$

A close inspection of the above equation will show that δ_p is the pth coordinate of the discrete cross-correlation function after the pth translation of the X group of samples with respect to the Y group. A simple method of normalising* each product of the series was employed during the early stages of the investigation but was abandoned later because normalization was found to be contributing very little to the results of the analysis. In fact, the time involved in the normalisation process was out of proportion to its usefulness.

Other sets of serial products, C_p and D_p, were computed,

---

* In order to normalise the product, δ_p, we multiply δ_p by 1/(M-p). Thus, the normalised product, γ_p, becomes

$$\frac{1}{M-p} \quad (X_1Y_{p+1} \quad \ldots\ldots\ldots\ldots\ldots\ldots \; X_{M\ p}Y_M)$$

performing ~~the~~ similar operations as in the computation of δ_p. Thus,

$$C_p = X_1X_{p\ 1} + X_2X_{p\ 2} + \quad \ldots\ldots + X_{M-p}X_M$$

and

$$D_p = Y_1Y_{p\ 1} + Y_2Y_{p\ 2} + \quad \ldots\ldots + Y_{M-p}Y_M$$

As before C_p and D_p are unnormalised and for the same reasons as for δ_p. It is important to note that C_p and D_p ($p \quad 1,2,\ldots\ldots\ldots,2M$) are, respectively, the auto-correlation discrete functions of the X and the Y groups of samples assembled from the field under examination. A direct comparison between D_p and δ_p (or between C_p and δ_p) was made in order to analytically study the differences the transmission contents of the two states of the frequency field under examination. An indication of the location of any transmission which appeared in one state and in the other state was necessary in order to establish the identity of that transmission. Although this indication was provided by the type of comparative study undertaken in the present investigation, it soon became clear that the results were dependent upon the resolution capability of the instruments being used. It is worth mentioning that with this equipment it was possible to detect the emergence of transmission(s) within a frequency range of 10 kHz in a total sweep range of 100 kHz.

6.5 Choice of Test Functions

Having obtained a pair of spectra in a manner described earlier (section 6.3) the detection of transmission changes is begun by correlating the waveform representations of the two spectral records. The correlator employed is designed in such a way that its output is either the autocorrelation function of any one waveform representation of a spectrum or the cross-correlation function of two different waveform representations. Assuming that the spectral records are represented by f_i and f_j, the correlation function which can be evaluated are ϕ_{ii}, ϕ_{jj} and ϕ_{ij}, where

ϕ_{ii} = autocorrelation function of f_i

ϕ_{jj} = autocorrelation function of f_j

ϕ_{ij} = cross-correlation function of f_i and f_{ji}

In order to establish any differences between the transmission signal contents of the spectrum represented by f_i and those of another spectrum represented by f_j, a general function $T(\phi_{ii}, \phi_{ij}, \phi_{jj})$, is conceived. Three main classes of the generalised function can be evolved, and by using each class function, the differences between f_i and f_j can be detected. The three classes of functions, to be known as "test functions", are the following:

(1) $T_1(\phi_{ii}, \phi_{ij})$

(2) $T_2(\phi_{ii}, \phi_{jj})$

(3) $T_3(\phi_{ij})$

Other combinations of the ϕ's can be formed to generate some more classes of the test functions, but in the present investigation, use will be made of only three depicted above, largely because of their simplicity in form.

The test function, $T_1(\phi_{ii},\phi_{ij})$, is derived from the autocorrelation function, ϕ_{ii}, and the cross-correlation function, ϕ_{ij}, and is defined as the "absolute" difference between ϕ_{ii} and ϕ_{ij}. In mathematical terms,

$$T_1(\phi_{ii}, \phi_{ij}) = \left|\phi_{ii} - \phi_{ij}\right| \qquad \ldots (0.1)$$

Similarly,

$$T_2(\phi_{ii}, \phi_{jj}) = \left|\phi_{ii} - \phi_{jj}\right| \qquad \ldots (0.2)$$

It is evident from the form of equation (0.2) that the evaluation of $T_2(\phi_{ii}, \phi_{jj})$ requires the use of the autocorrelation functions of the two spectra represented by waveforms f_i and f_j.

The derivation of the test function, $T_3(\phi_{ij})$, involves only the use of the cross-correlation function, ϕ_{ij} and is defined by the following relationship:

$$T_3 = \left|\phi^{+}_{ij} - \phi^{-}_{ij}\right| \qquad \ldots (0.3)$$

Clearly, the evaluation of $T_3(\phi_{ij})$ implies a comparative study of the two sides (i.e. positive and negative) of the cross-correlation function, ϕ_{ij}, formed from the waveforms, f_i and f_j, which represent the two spectra under examination. In equation (0.3), ϕ^{+}_{ij} is the positive side of ϕ_{ij} and ϕ^{-}_{ij} is the negative side. Actually, T_3 is the "absolute" difference between the two sides of the cross-correlation function, ϕ_{ij}.

6.5.1 Nature of the Test Functions

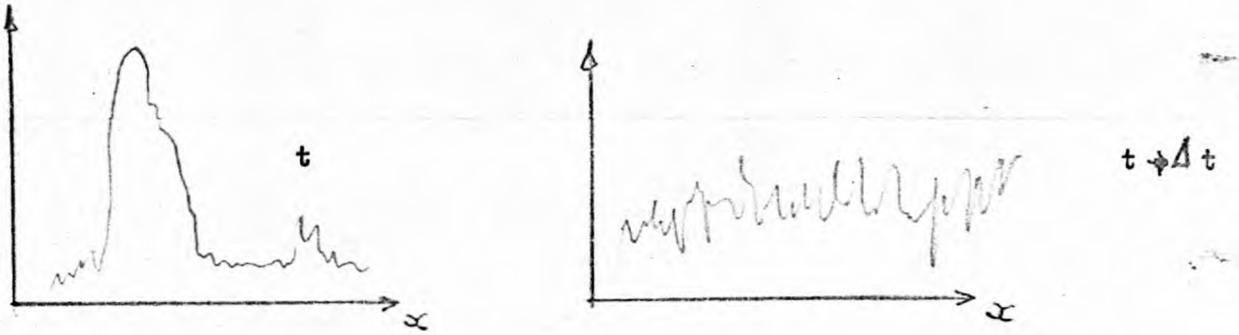
It can easily be seen that the test functions, T_1, T_2 and T_3, are ideally zero* functions; that is, whenever the two spectral records being compared have the same transmission signal contents then their "difference" is ideally zero. Otherwise, they have finite values at all, or some, of the points in their space of definition. The space of definition is equal** to either the swept bandwidth (for T_3) or two times the swept bandwidth (for T_1 and T_2). Since most of the practical situations encountered are far from being ideal, the test functions are not usually zero functions, but have a finite value almost everywhere in the space in which they are defined. This is true even when the signal contents of the two spectral records under examination are nominally the same.

Fig.6.6a shows a representative waveform of a spectrum which might be observed at some time t. At a later time, t + Δt, the spectrum is

---

*A zero function is a function whose value at all points is zero.

**The swept bandwidth from which f_i and f_j are constructed is equal to the size of the space in which T_3 is defined. This is because for every point in T_3 there are two points in ϕ_{ij} from which T_3 is derived. It should also be remembered that the correlation functions, ϕ_{ii}, ϕ_{jj}, and ϕ_{ij} are defined in a space equal to two times the swept bandwidth and that the "fold-over" effect inherent in T_3 is absent in T_1 or T_2.



Fig, 6.6 Represaentative waveforms of a spectrum produced at a) t and b) t + t

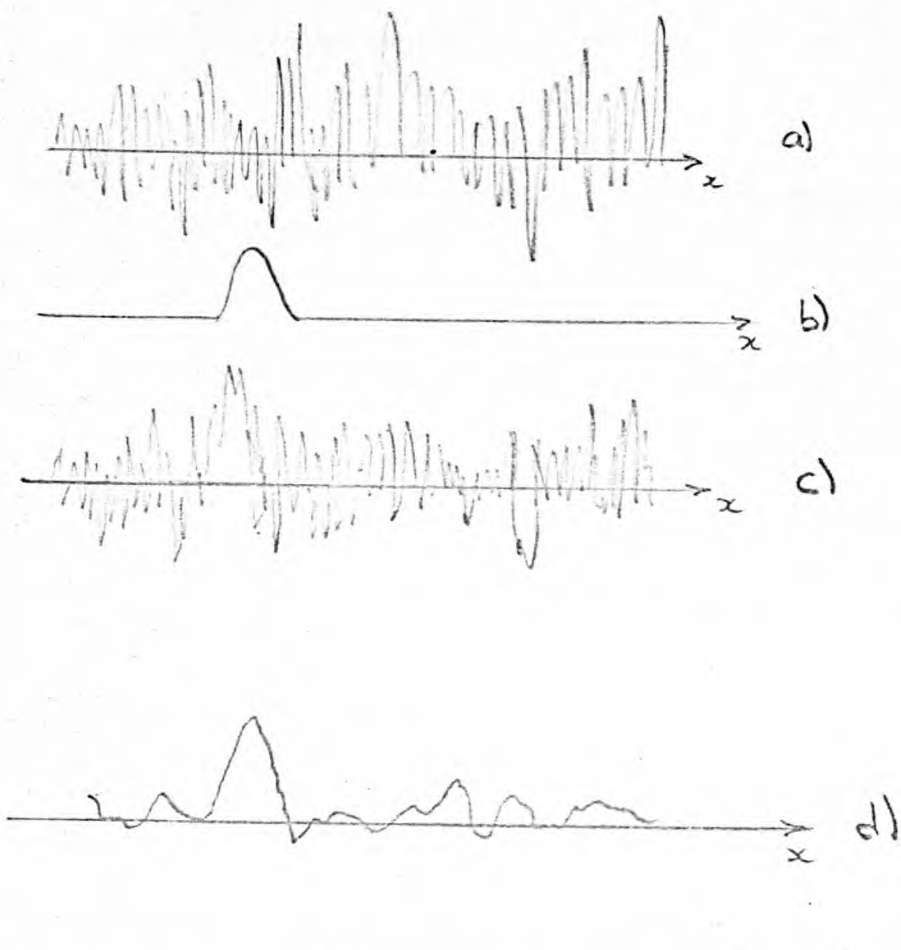


Fig. 6.7 a) noise, b) signal, c) signal + noisee d) correlation function of a) and d).

observed again and represented by the waveform shown in Fig. 6.6b. In the analysis of such a pair of spectral records, some important assumptions are made. First, it is assumed that the locations of the signals which appear in the spectral records do not vary with time. Second, it is assumed that all the noise in the spectrum can be regarded as an addition to the transmission signals that may exist in the frequency band under observation.

Fig. 6.7a is a typical representation of a noise spectrum found within a portion of the hf band. When an active signal transmission signal whose spectrum is similar to that shown in fig. 6.7b appears anywhere in the given region, the spectrum of the combined signal and noise is obtained (fig. 6.7c). From the two waveforms shown in figs. 6.7a and 6.7c, the correlation function ϕ_{ii} is formed. In this function, it will be observed that the noise has assumed a structure (fig. 6.7d) similar to that of the signal which it "cloaks". This equilization of structures is accompanied by a corresponding improvement in the amplitude discrimination between the signal and the noise. In fact, this is one advantage of the correlation process.

The peak that appears in the plot of the test function, T_3, is an indication of the difference, in signal contents, between the two spectral records under examination. The test function would have been a zero function were it not for the signal located in the search band slot and revealed in one of the spectral records.

6.6 Setting of Threshold

6.6.1 Introduction

When the effects of noise and fading are absent, it is generally possible to attain theoretical perfection in a transmission signal detection process. However, in the presence of noise and when the signals are affected by fading, the existence of the signal within the search band becomes an uncertain event and can be established only statistically. The problem, therefore, of determining the existence of a transmission signal can be solved in a manner which attaches a measure of uncertainty to the achievable solution. In fact, in "spotting" the precise location of a transmission signal in a given frequency band, the observation time required is inversely proportional to the bandwidth* of the signal and, for a signal with a small bandwidth ($<$ 1Hz), the detection time is very long.

To the human observer who is required to make decisions on the existence (or non-existence) of the transmission signals, a criterion based on the signal-to-noise ratio of the detection process is normally sufficient and the important requirement, therefore, is a maximised signal-to-noise ratio level. Such a procedure based only on the

---

*Here, the principle involved assumes that time and frequency are conjugate variables and that, in order to sharpen a characteristic in terms of one variable, it is necessary to increase its breadth in terms of the conjugate variable. This is the basis of the uncertainty principle in modern quantum mechanics used in atomic theory.

signal-to-noise ratio may be informationally adequate, but the human observer is usually left with the task of making a mental estimation and hence the final decision. It will, therefore, be more convenient and, of course, more appropriate to provide, not information to be further assimilated by a human observer, but the information which is the decision itself. The problem is a complex statistical one and can only be solved satisfactorily by judicious reference to the appropriate statistical theories on decision making.(74)

6.6.2 Amplitude of Peaks in Test Functions

The appearance of peaks in the plot of a test function has been shown (section 6.5) to be the basis of the decision concerning the differences between any pair of spectra being analysed. However, the occurrence of peaks alone cannot be taken as a sufficient criterion since the noise[x] effects and other disturbances may produce peaks and hence lead to the erroneous conclusion that some relevant change in the frequency space has occurred.

In the present investigation, therefore, a criterion based on the amplitude of a peak has been adopted as an adequate "figure of merit" and used as a basis for the decision making procedure. The choice of this criterion was suggested by the results of some empirical studies, the description of which can be seen in Section 7.2.2. In

---

[x]The noise under consideration is that introduced during the analysis and that caused by errors in the computation processes.

these studies, the amplitudes of the various peaks appearing in the plots of the test function are measured and those peaks whose amplitudes exceed a "certain figure" indicate situations in the analysis where transmission signal changes in the spectrum have occurred. This figure, with which the peak amplitude is compared, is known as the threshold of the search system, and thus sets the minimum limit to the detectable transmission signal changes that can occur in a spectrum. The setting of the threshold is never perfect and this also means that when a change is detected by the search procedure, there is a small but finite probability that the detection result is false.

6.6.3 Logarithmic Assessment of Peaks

The test functions derived from the auto- and cross-correlation functions of the waveforms which represent the spectral records, are plotted on a logarithmic basis. To do this, the values at all points of the correlation functions, which are employed in the computation of a test function, are compared to the value at the origin. In other words, each correlation function is normalised with respect to the value at the origin. After this, the modified (see Section 7.) point values are converted into corresponding values in decibels by the use of the appropriate ~~unitriplicative~~ multiplicative* factor. The test function is then

*Strictly speaking, the unit of the decibel has been introduced for the particular purpose of measuring power ratios. Hence, a power ratio of P_1/P_2 is equivalent to $10 \log_{10} P_1/P_2$ decibels.

plotted with an ordinate measured in decibels. The choice of 0 db as the datum of all the measurements has been based on the fact[xx] that

$$\log_{10} \frac{\phi(0)}{\phi(0)} = \log_{10} 1 = 0$$

Indeed, when an event is compared with itself the condition of no change will be observed, and if a "logarithmic comparator" is postulated in such a situation, its output will be zero. The only point in any of the correlation functions which is transformed unchanged into a point in a test function is the origin and hence the condition of no change is always observed there.

According to Shannon(75), the choice of the logarithmic basis of description is usually prompted by the following reasons:

(1) The logarithmic value is more useful. Parameters of engineering importance tend to vary linearly with the logarithm of the number of possibilities .... Doubling the time roughly squares the number of possibilities, or doubles the logarithm.

(2) The logarithmic value is nearer to our intuitive feeling as to the proper measure. One feels, for example, that two punched cards should have twice the information capacity of one for information storage.

(3) The logarithmic value is mathematically more suitable.

[xx]Since the power spectrum is the representation of the autocorrelation function in the frequency domain, the relation between them is a linear one. In fact, if $r(f)$ is the power spectrum and $R(\tau)$ is the autocorrelation function of some time function $g(t)$, then

$$R(\tau) = \quad r(f)\, e^{j2\pi f} \quad df \quad \text{and}$$

$$r(f) = \quad R(\tau)\, e^{-j2\pi f} \quad d\tau.$$

Hence, just as we can measure the ratio of the powers at any two arbitrarily chosen frequencies, we can also obtain the ratio $\frac{R(\tau_1)}{R(\tau_2)}$ which is a "power" ratio at some two points, τ_1 and τ_2, in the correlation space of R( ).

Chapter 7

PRELIMINARY EXPERIMENTS ON REAL AND SIMULATED SIGNAL

Chapter 7 PRELIMINARY EXPERIMENTS ON REAL AND SIMULATED SIGNAL.

7.1 Introduction

Preliminary tests were carried out with the purpose of studying certain factors which may affect the results of the search process. The topics studied were the necessary time separation of the spectral records, and the effect of selective fading and non-selective fading. A general study of the ability of a test function to detect any changes in the spectrum was undertaken, and methods of deciding on a system threshold were also investigated. All these tests involved experiments on both real and simulated spectral records.

7.2 Suitability of Test Functions

7.2.1 Experimental Technique.

Theoretical considerations (section 3.) have shown that it is generally possible to detect transmission signal changes in any given spectrum by the application of a test function defined in section 6. The changes in a spectrum can arise from a number of causes, the important ones being:

(1) Actual "loss" or "gain" of transmission signals;

(2) Selective and non-selective fading.

(3) Very low signal-to-noise ratio of signals located in the spectrum.

In order to judge the suitability of a test function when any of the above mentioned causes affect the transmission signal contents of a spectrum, preliminary experiments and simulations were needed. Fig. (7.1) shows the experimental arrangement employed. Four AM[≠] signal generators and a white noise source were used. The signals and the noise were combined in an adding circuit whose output was fed into the (Racal) receiver. The panoramic adaptor and the (Mullard) analogue-to-digital converter then produced, respectively, a visual display and a digital record of the composite spectrum of the various input signals and noise. Two spectra were produced for the analysis, and each of these contained a number of AM signals centred on different frequencies within the selected frequency band. The centre frequencies were chosen in such a way that for any two adjacent signals, the highest frequency component of the lower-frequency centred signal differed by at least[x] 4 kHz from the lowest frequency component of the higher-frequency centred signal.

One spectrum of four such signals, mixed with white noise, was produced, sampled and stored; another spectrum whose signal contents were changed[xx] according to the type of study in progress was also produced and sampled. The spectral records produced then became the discrete waveform representations of the spectra of signals (and noise) injected into the receiver.

The analysis of the various spectral records that are generated

---

[x]The choice of this was determined by the width of the conventional voice channel.

[xx]The changes are effected by any of the three causes mentioned in sect.7.2.1.

[≠]Amplitude modulated.

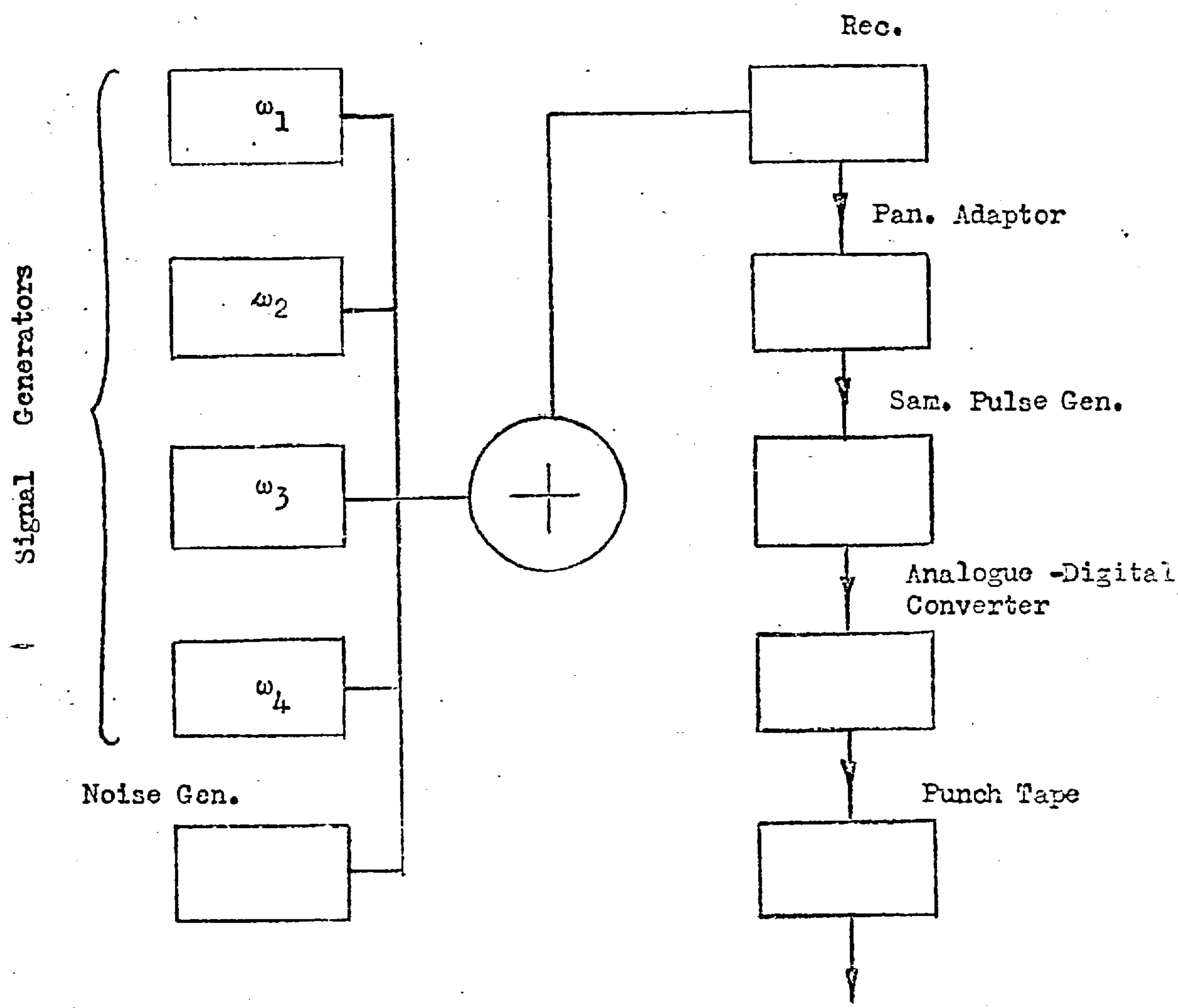


Fig. 7.1 A block diagram of the arrangement used in assessing suitability of test functions

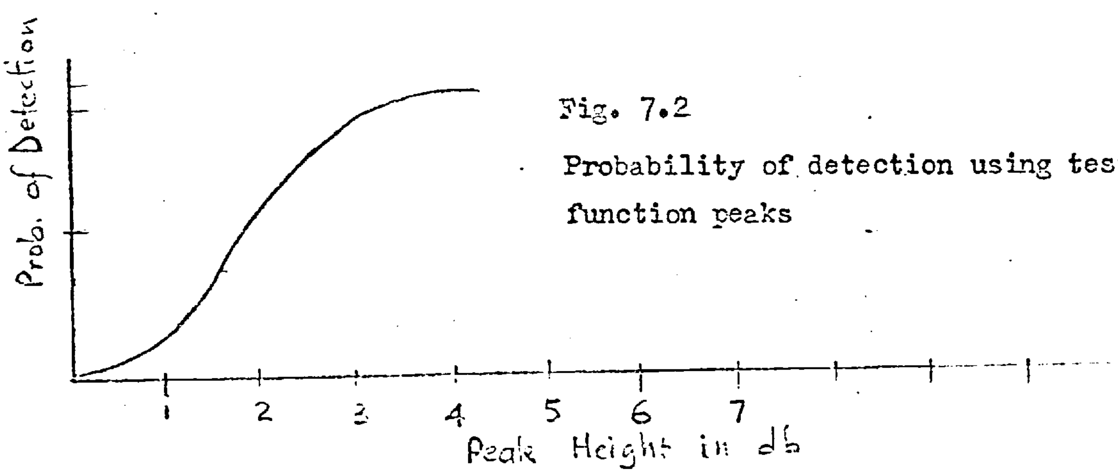


Fig. 7.2

Probability of detection using test function peaks follows the same trend as those discussed in section 6.4. Thus, having obtained a pair of spectra represented as waveforms, f_i and f_j, the correlation functions, ϕ_{ii}, ϕ_{ij} and ϕ_{jj}, ~~are~~ were evaluated.

At times, the differences between f_i and f_j were so great that they could be observed by visual comparison of the plots of ϕ_{ii} and ϕ_{ij}. But at other times, it became necessary to actually compute and plot a test function in order to obtain the relevant information about the transmission changes.

7.2.2 Threshold Setting

Theoretical considerations have shown that the peaks in the test function, which rise above the 0 db level indicate that a difference in the pair of spectra being analysed has been detected. However, in practical situations to be described, it was observed that even though any two spectra contained the same transmission signals, the test function, $T_3(\phi_{ij})$, was 0 db only at the origin. This departure from the ideal is attributable to the presence of noise, fading and other sources of disturbance in the hf spectrum and suggests also the need of a threshold different from the 0 db level.

In order to show practically that the height of the peaks in the test function can be used in the decision making process, pairs of spectra which were known to contain the same transmission signals were produced. Another set of pairs in which the two spectra of a pair were known to contain different and unequal numbers of transmissions were also produced for analysis. The method of analysis and the experimental arrangement employed were similar to those discussed in section 7.2. In other words, real-life conditions were simulated by generating AM signals and combining them with artificial noise in order to produce the spectra which were analysed. The object of this present exercise, was to show experimentally that the test function was capable of detecting any differences between a given pair of spectra which actually contained different transmission signals.

In these experiments, the average signal-to-noise* ratio of the signals in each spectrum was maintained constant but with a small difference (3-5 db) between the two averages. (A way of controlling the level of the signals and noise has been discussed in section 7.2.3.). The test function was computed for each pair of spectra and the mean peak-height of the test function was calculated. In situations where there was no "genuine" difference between the signal contents of the two spectra being analysed, peaks of the test function were observed to be below the 3 db level.** With this level as threshold, 96% of the genuine differences were detected. The plot of fig. (7.2) shows that for peaks in the test function whose levels were higher than 3 db, the percentage of differences detected was smaller. It is important to note here that, with the levels of signal-to-noise ratio employed (1-25 db) the results were not affected appreciably.

Real-life spectra were also analysed in exactly the same way as

---

*In the simulation tests, the signal-to-noise ratio was assessed by a direc comparison of the strength of the signals injected into the receiver with the power of the noise with which they are mixed.

**This is the "level" of dissimilarity. See section 8.1.

that described above. The spectra were all chosen from a region of the hf spectrum where the signals contained were known a priori. The spectra were obtained at different times of the day and night, and the separation time between a pair was, on the average, 45 minutes (see section 7.4). The signal-to-noise ratio of the signal in each spectrum was unknown and hence unspecified. However, the results with artificial signals have shown that the signal-to-noise ratio, provided it remained constant, did not influence the direction of the results. With real-life transmission signals, 90 % of the genuine differences were detected when the threshold level of 3 db was used.[xx]

7.2.3 Sensitivity of System to Changes in Signal Levels. (Effects of Non-selective Fading).

By keeping the level of the four signals in one spectrum at a constant level and varying all the signal levels in the other spectrum, different sets of paired spectra were generated. For each pair, the three test functions were computed and used to detect any differences that might exist between the pair concerned. A means[x] of measuring the noise average power in the selected frequency band was devised, and knowing the signal power level, an estimate of the signal-to-noise ratio was made. (It is important to note that the noise generator was capable of producing noise bandwidths ranging from 50 Hz to 500 kHz.)

---

[x]A random noise volt-meter (Brüel and Kjaer, type 2417) was used.

[xx]Fifty pairs of spectra were examined in any given situation.

Also, with a suitable circuit arrangement, a noise bandwidth of 5-50 kHz could be translated to any high frequency range by a heterodyne process.

In the experiments carried out, a maximum difference in signal level of 20 db between the two spectra was achieved. Some representative plots of a test function for various signal level differences have been shown in fig. 7.3. A table of results is also given (see Table 1). In this table, a result of the search analysis is obtained by selecting a test function from the first row and the level difference from the first column. The column in which the particular test function is found and the row corresponding to the selected level difference meet at Y or N. A result, Y, implies that there is a difference in signal contents of the two spectra under examination; N implies the opposite decision.

From the plots and the tables provided, it is evident that, although the test functions are derived from simple similar functions (i.e. the correlation functions), the capability to detect "genuine" differences between any two spectra vary from one test function to another. In any case, the results have shown that when the signals and noise in the spectrum are affected by disturbances which resemble non-selective fading, it is generally possible to detect the genuine transmission changes caused by the advent of new signals or the cessation of active operation by a transmission.

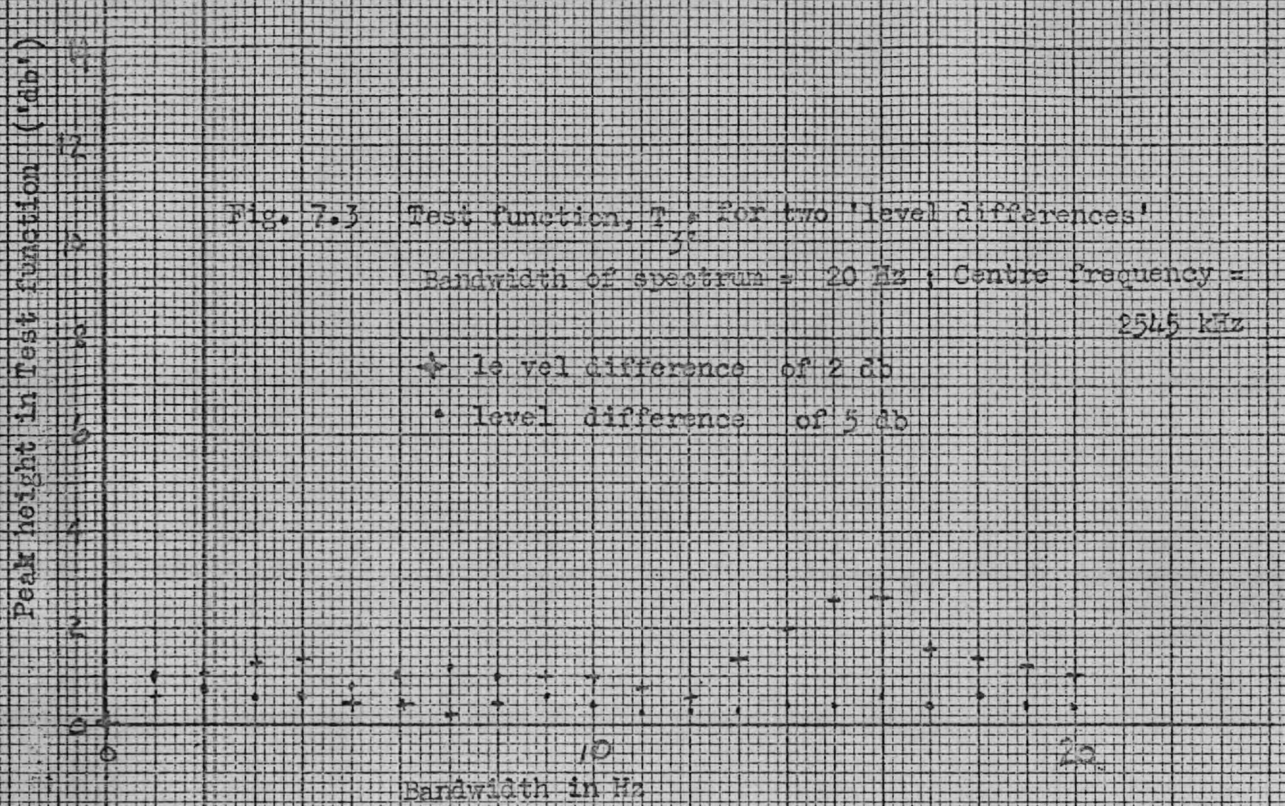


Fig. 7.3 Test function, T_3, for two 'level differences'

Bandwidth of spectrum = 20 Hz ; Centre frequency = 2545 kHz

+ level difference of 2 db

• level difference of 5 db

TABLE 1

Level Diff (db) \ T. Func.	T_1	T_2	T_3
2	Y	Y	Y
3	Y	Y	Y
4	Y	Y	Y
5	Y	Y	Y
6	N	Y	N
8	Y	Y	Y
10	Y	N	Y
12	Y	N	Y
15	N	Y	Y
18	Y	N	Y
20	Y	Y	Y

Y = 'change' condition

N = 'no change' condition

7.2.4 Simulation of the Effects of Selective Fading

Another series of experiments were performed in order to study the results of simulated selective fading. Selective fading patterns* were generated by adjusting the levels of the individual signals in the selected frequency band. (Provision for controlling the levels was available in the instruments used).

As before, the test function corresponding to a pair of spectra was computed, and a table of the results has been given in Table 2. Ten test functions obtained did not exhibit any definite shape or pattern. It was also observed that by using relatively high level signals in one spectrum of the pair, the results of the analysis were misleading. Differences between the spectra were observed when., in fact, the two spectra contained the same number of signals centred on the same frequencies. The table, however, shows that the test functions give reasonably good results when the average difference in signal levels is small, say, 3-8 db.

7.3 Emergence and Disappearance of Signals

A method of investigating changes that take place when a new transmission starts up and ceases in any given portion of the hf band was devised, involving the arrangement of the experimental apparatus

*Selective fading patterns and their probability distribution are fully discussed by S. O. Rice in the reference 76.

Diff. in ave. level of signals the two spectra \ Test Function	T_1	T_2	T_3
2db	Y	Y	Y
10db	Y	N	N
20db	N	N	N

TABLE 2

An example when some tests yield the answer 'yes'.

Number of signals emerging \ Test Function	T_1	T_2	T_3
1	Y	Y	Y
2	Y	Y	Y
3	Y	Y	Y
4	Y	Y	Y

TABLE 3

An example when all tests yield the answer 'yes'.

Y ='change' condition

N ='no change' condition

similar to that used in the previous experiments (Section 7.2). In these experiments also, AM signal generators and a white noise source were employed. A provision for quenching one or more of the signals permitted the insertion of a specified number of signals. Their locations within the chosen frequency band were known and their levels were also pre-set. Various S.N.* ratio levels (see section 7.2.1) were employed, and the test functions were used as means of detecting the differences in the signal contents of the two spectra being examined.

Results of these simulations are given in the form of a table (see Table 3). In the first column, the difference between the number of signals contained in one spectrum and the other is given. To see the results of the search analysis when there are two more signals in one spectrum than in the other, a test function is selected. If the test function selected is T_3, the result is seen to indicate that, in fact, there was a genuine difference between the two spectra. Also, a rough estimate of the location of the "difference" signals can be made from the locations of the peaks in the plot of the test function (see fig. 7.4).

By using the test function, $T_3(\phi_{ij})$, it was possible to detect differences between a pair of spectra when the signal-to-noise ratio in each was as low as 5 db. The application of the results of this exercise to the real-life problem of detecting the emergence of new transmission signals was carried out by using pairs of the spectra chosen from a given portion of the hf band.

---

*S.N. Signal-to-noise.

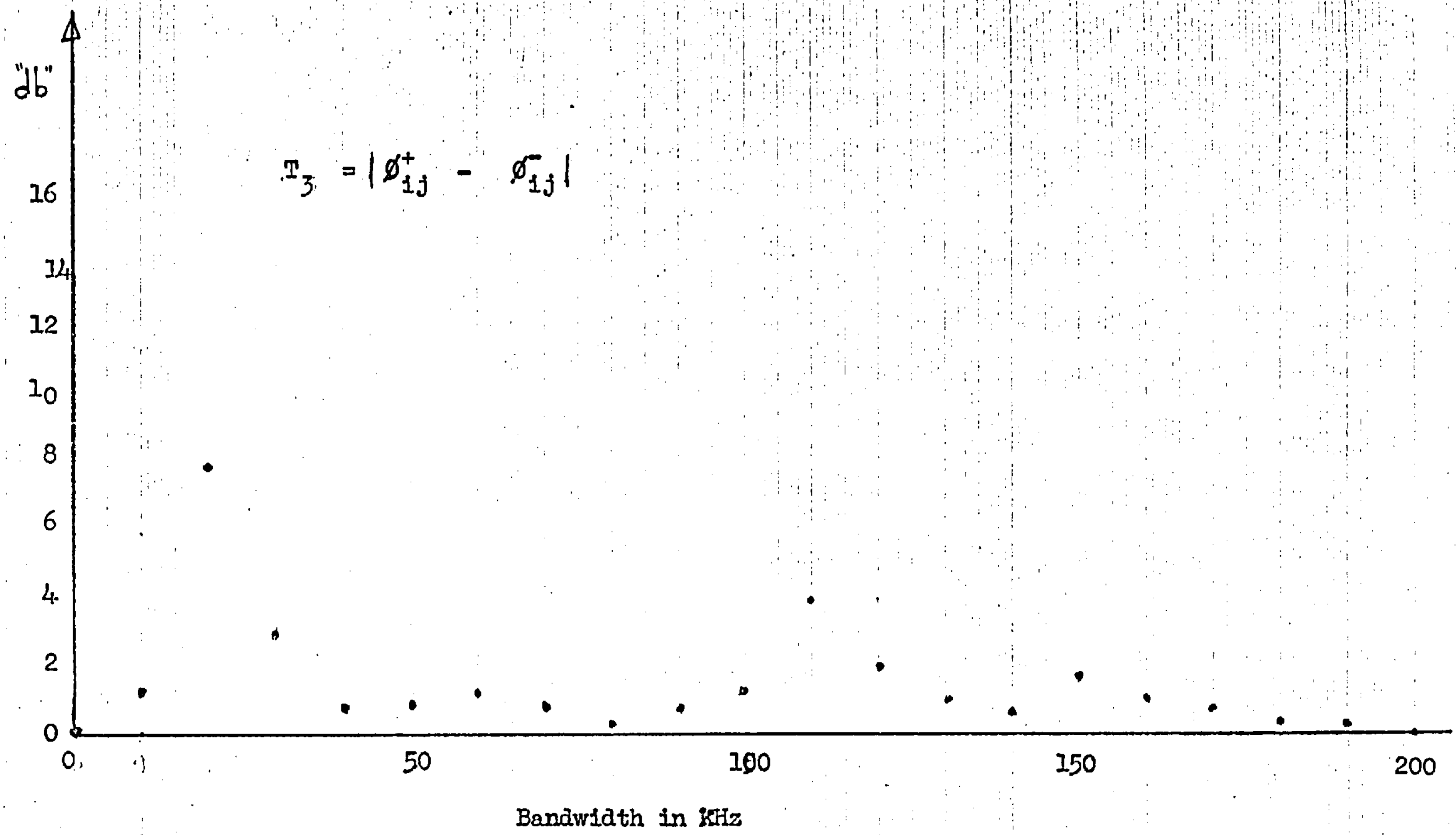


Fig. 7.4 A test function showing a pproximate location of the "difference signals".

7.4. The Effects of the Period Between Spectra

7.4.1 Experiment to determine choice of separation time

In the simulation experiments, and those involving known regions of the hf band, the time separation of the spectral records used was not an important factor. However, in practice, when the characteristics of the signals under investigation are unknown, this period becomes significant. Preliminary tests were carried out to determine an optimum separation time between the two spectra. Pairs of spectra were produced with different time intervals between them. The correlation analysis, as previously described, was then carried out for each pair of spectra. In this way, results for different values of separation time was obtained.

The time lapse used in the experiment ranged from 2 seconds to as much as 85 minutes. However, these tests called for another preliminary study designed to obtain information about the signal population densities of the various portions of the hf band. For this study, a British Post Office (B.P.O.) scanning receiver was used. This equipment, shown in a form of a block diagram (Fig. 7.5), has a sensitive communication type receiver (with a non-slip slow motion drive) which is swept over a pre-specified frequency band once every 2 minutes by a drive unit which is mechanically coupled to the tuning spindle of the communication receiver. In synchronism with the receiver tuning control, a recording device moves steadily over a chart of an electrolytic paper and records a mark whenever a signal is picked up by the receiver at a strength higher than the minimum at which the equipment is pre-set to operate. The swept band is

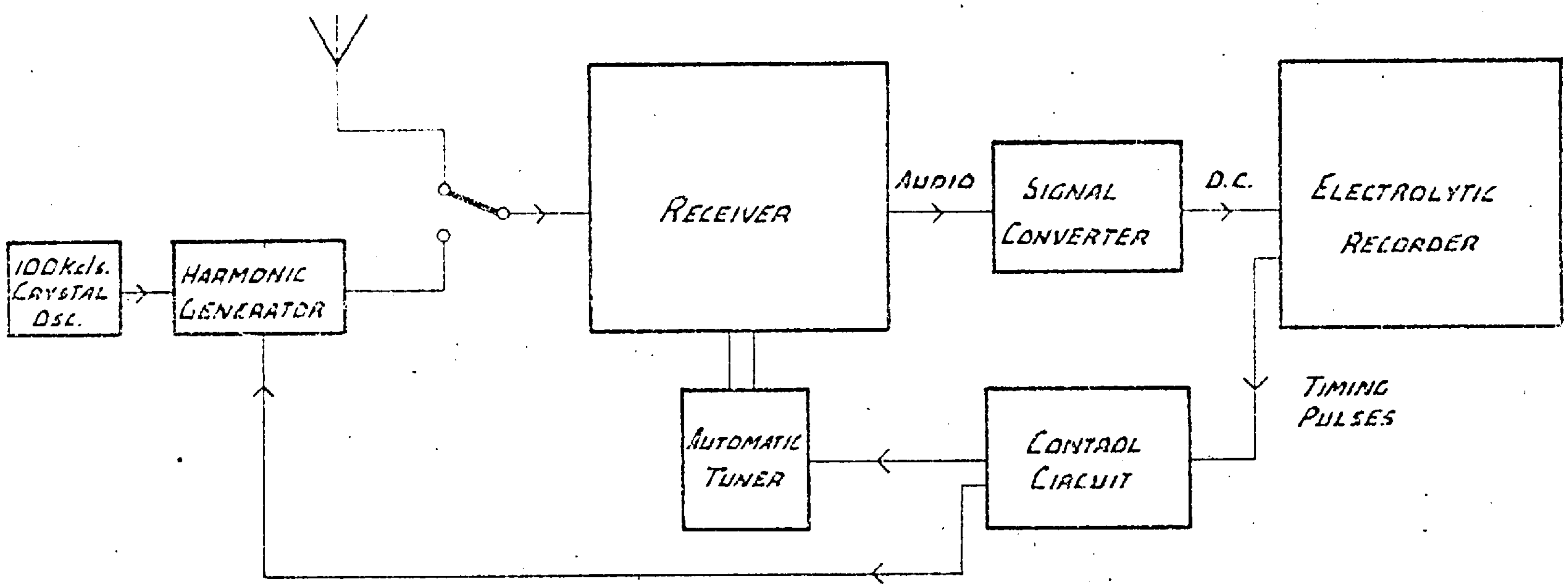


Fig. 7.5 A block of the Scanning Receiver

AUTOMATIC SCANNING RECEIVER
BLOCK SCHEMATIC

represented by the width of the chart, the smaller the frequency swept the larger the spacing between the traces of adjacent transmission signals. A normal operating condition gives a swept band of approximately 100 kHz for an 8 in. width of the chart. The normal chart speed is 1/2 in. per hour. At two hourly intervals, the input of the radio receiver is automatically disconnected from the aerial and is switched to receive frequency calibration signals from a 100 KHz harmonic generator. A typical record is shown in Fig. 7.6.

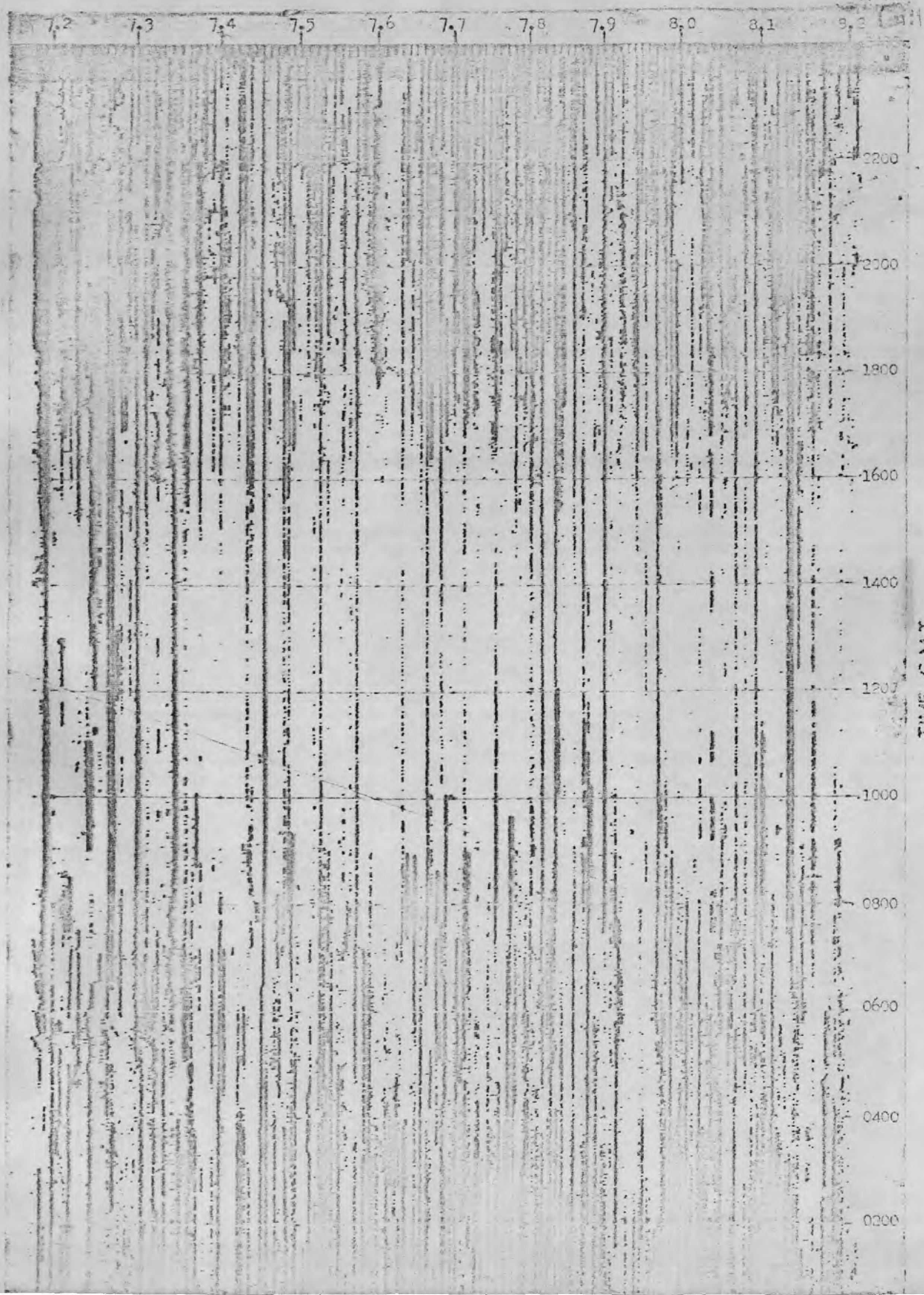
From such a record, it was possible to identify the locations of the active transmission signals. A choice of the portion of the hf band to be investigated can then be made on the basis of the knowledge of the space occupancy provided by the record. Having selected such a portion for the preliminary studies, pairs of spectra of that portion were produced. The shortest separation time was 0.2 sec. and is set by the speed of operation of the Racal panoramic adaptor. The slowest available rate of scanning is once every 2 seconds, and at this slow rate, spectra whose bandwidths are much greater than 100 kHz are scanned without loss of any significanct information about the contents.

By varying the separation time and studying the effects produced, it was possible to estimate the optimum separation time that may be employed in the selection of the pairs of real-life spectra. If the frequency band under examination is scanned at the rate of once every 2 secs., then a separation time of about 40 minutes, disregarding the "blank" time between consecutive sweeps, is equivalent to a frequency

FIG. 7.6

TYPICAL SCANNING RECEIVER RECORD

WITH RECEIVER INPUT LEVEL > 10μV [redacted]



space separation of more than 1000 times the bandwidth of the spectrum being examined. In fact, 40 minutes is a very long time for studying small portions of the hf spectrum, and the results of such studies will show whether the random processes comprising the spectrum are stationary* (e.g. short-term stationarity) or not.

Another important consideration which influenced the choice of the separation time was the particular characteristics of the various transmission signals that could come under observation during the scanning period. If data were available, giving information about the characteristics of the various signals likely to be found in any given spectrum, then use of this information could be made in deciding on the optimum separation time. For example, a frequency band containing broadcast signals, which have an average continuous "active life time" of 10 ± 2 hours, can be examined by analysing a pair of spectra which are separated by as much as 3 to 4 hours. This sort of separation time would not be suitable in the analysis of a search band containing telegraph signals whose average life time of continuous activity is much shorter.

As may be realised, the problem of choosing an optimum separation time becomes fairly complex when it is remembered that a combination of different communication services can be accommodated in a given broad band of frequencies and that the choice of the band slot isolated for analysis is arbitrary. However, the results obtained showed that, for normal broadcast signals, an average separation time of up to 45 minutes was adequate if significant transmission changes were to be

*Generally speaking, a random process with non-changing characteristics is said to be stationary.

satisfactorily detected.

Fig. 7.7 shows a block diagram of the experimental arrangement employed in analysing the spectra that are produced when the Racal receiver is connected to a vertical aerial. This arrangement is similar to that shown in Fig. 7.2, except that in the latter, AM signal generators, instead of real-life signal sources were used. The experiments were performed both at night and during the day.

The spectral records of a given portion of the hf band (1-12 mHz) were selected in pairs with different separation times for the different pairs. For each pair, the correlation analysis was performed. The test function, $T_3(\phi_{ij})$, was applied and plots of it are given in Fig. 7.8. for the various pairs of spectra. A table of results (Table 4) has also been given, and in this table the results of analysing two real-life spectra, separated by c seconds, is given. Y implies that a difference was observed which N implies the opposite. Certain results indicate that with real-life signals in the hf spectrum, especially during daytime, one has to wait for a considerable time before any significant change occurs. This is generally due to the fact that the transmission signal contents of any given portion of the hf spectrum do not change frequently. The results also confirm the established fact[77] that this part of hf band is congested during the daytime.

For this portion of the hf band, it would thus appear that a long period between samples is appropriate and that 45 minutes would certainly not be too long. It must be emphasized, however, that in general the optimum separation time will depend very much on the type

Table 4

Separation time \ Test Function	T_1	T_2	T_3
0.2 sec.	N	N	N
.2 sec.	N	N	N
5 "	N	N	N
10 "	N	N	N
15 "	N	N	N
20 "	N	N	N
25 "	N	N	N
30 "	N	N	N
60 "	N	N	N
5 min.	N	N	N
10 "	N	N	N
15 "	N	N	N
20 "	N	N	N
25 "	N	N	N
30 "	N	N	N
35 "	N	N	N
40 "	N	Y	N
45 "	N	Y	N
50 "	Y	Y	N
55 "	Y	Y	N
60 "	N	Y	N
70 "	N	Y	Y
80 "	N	N	Y
90 "	Y	N	Y
120 "	Y	N	N
150 "	Y	Y	Y
180 "	Y	Y	Y
200 "	Y	Y	Y
240 "	Y	Y	Y
300 "	Y	Y	Y

Y ≡ "change" condition

N ≡ "no change" condition

of signals occupying a band.

7.5 A Proposed Scheme of Operation

An application of the results of this investigation will now be discussed. A complete block diagram of the proposed scheme is shown in Fig. 7.9. Real-life transmission signal and noise, as time functions, are picked up by the aerial of a communication receiver and then transformed by the panoramic adaptor into events in frequency space. The continuous waveform representations of the frequenc space description are digitised by a converter which produces discrete versions of the continuous waveforms. The result is a set of "discret spectra" of the input signals and noise.

The rate of sampling performed by the analogue-to-digital converter, is determined by the particular transmission signals of interest. For this reason, a provision for a variable rate of sampling made available in the design of the converter. The first of the two spe produced for the analysis is stored as a discrete waveform. As the second spectrum is produced, also in its discrete form, it is immediately correlated with the first spectral record. A choice of a test function is made and derived from the output of the correlator. If, as is likely, test function, $T_3(\phi_{ij})$ is chosen, T_3 is then made available, for "inspection", to the "assessor" which assesses the heights of peaks in the test function and compares them with a preset threshold value. An indication of the result of the search procedure is

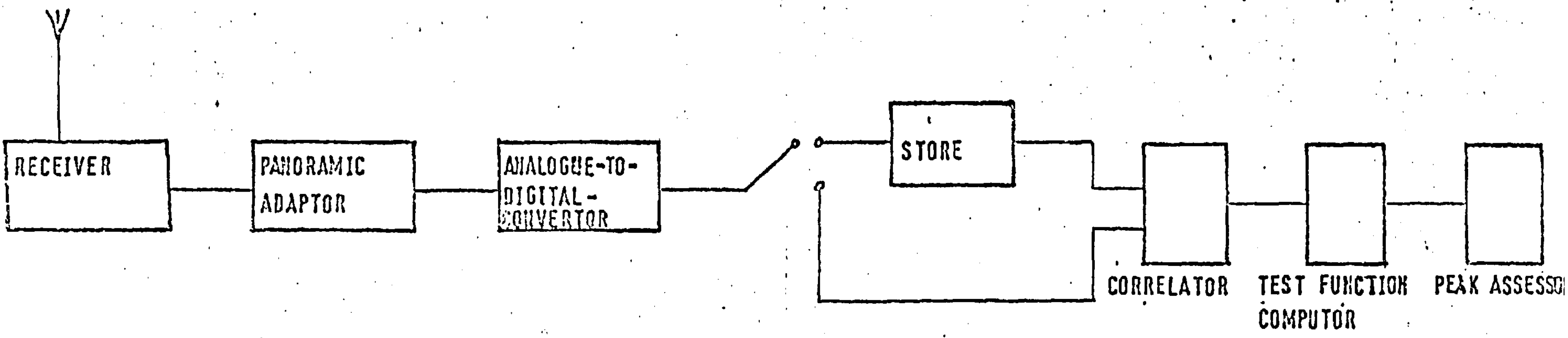


Fig. 7.9 A Schematic Represention of the Automatic Search system

produced at the output of the assessor. The output generally consists of two binary numbers, say, 1 for a change in the spectrum and 0 for no change. Once a change in the spectrum is observed, a careful examination must be made, if required, to obtain the intelligence in the new transmission or to identify a transmission which has ceased normal operation.

Depending upon the characteristics of the particular signals being sought, the separation time between the two spectra can be made as small as desired. Therefore, if the separation time is c seconds, say, then the changes recorded by the search system will be defined within the time interval of c seconds. Usually it will be known a priori what types of signal are being sought.

The operation of the system can then proceed unsupervised, with considerable man-power saving. However, a period of "learning" is required, probably under manual supervision, in order to set a reasonable threshold for the system. Once a threshold has been set, only periodic checks are required to ensure that the whole search system is functioning satisfactorily.

Chapter 8 CONCLUSION

8.1 Concluding Discussion

The results of the present investigation have shown that it is possible to detect the initiation of a new transmission, or the cessation of an active transmission, in a given portion of the hf band by comparing two spectral records. The time interval during which a transmission change is observed or detected is defined by the two instants of time at which the two spectral records are made.

The method of analysis adopted is basically simple and involves the use of statistical parameters which are derived from a complete set of correlation functions. These correlation functions are obtained by correlating one waveform representation of the spectrum with itself or with the waveform representation of other spectral record. The time interval between the making of the two records is an important parameter, affecting the results of the analysis. A preliminary study suggested that the optimum time separation between the spectral records depended very much on the signal characteristics. Therefore, a consideration of the signal characteristics, such as bandwidth and average life-time, was necessary. The results of the analysis showed that for broadcast signals in the 1-2 mHz band, a time separation of up to 45 minutes was sufficient.

Simulation tests also showed that for signal-to-noise ratios above 5 db, the search technique was insensitive to signal-to-noise ratio changes of up to 12 db. Within these limits, the effects of non-selective fading were therefore slight, but selective fading proved more of a problem.

At first sight, the results of the search analysis involving selectively fading signals suggest that the technique is not so useful in this case. Indeed, the correlation analysis showed that, although one of the spectra produced might be known to contain signals which were a selectively fading version of the signals in the other spectrum, an analysis of the two spectra showed that the signal contents had "changed". However, it must be remembered that a signal may "lose its identity" completely under the influence of some types of fading, especially, deep selective fading. As far as the simple search technique is concerned, two spectra, one of which is influenced by selective fading, are not the same even if, in fact, they contain the same transmission signals. A direct comparison of two such spectra must reveal differences between them. In the light of this explanation, the experimental results are seen to be as satisfactory as could be expected. Only a search process which involved an actual identification of the signals could overcome this difficulty. Alternatively, a method of combating selective fading in such situations might involve use of diversity techniques.[78]

The correlation of two spectral records of a given portion of the hf band minimizes the effects of noise. The results also showed that the noise process within a given spectrum was, at least, quasi-stationary and does not change its structure. However, it was found that the occurrence of impulse noise produced effects whose duration was either too short to be noticeable or so long that every frequency component within the spectrum was intensified by an equal amount. In the latter case, a situation exists similar to that of non-selectively fading signals, and analysis has shown that the transmission signal contents are not effectively altered by the occurrence of an impulse noise.

The ordinates in the plot of the test function can be related to the "level" of dissimilarity of the two spectra being analysed. Hence, 0 db "level" of dissimilarity indicates the condition of no change while any level of a finite value shows a departure from the "no change" condition. An ideal choice of a threshold can be made by taking into consideration all the factors which affect the signal contents of the spectrum. In this investigation, the threshold was set with the view of achieving a practical optimum detection probability, in the shortest time possible, for the transmission changes that occur within a given portion of the hf spectrum.

The efficiency of a search technique can be expressed in terms of its ability to detect such transmission changes with an appropriate false alarm rate. A properly designed system would necessarily reflect the "costs" involved in the decision procedures. Any extra information available, such as the precise frequency location of transmission changes is also of interest. The choice of test functions used in the described system, will depend on such considerations.

A further factor is the practical realisation of the system. Most of the work done in this research involved operations which were performed in non-real time, and this was made possible by the use of a digital computer (the "Atlas"). However, in practical situations where quick decisions are required, the operations must be in real time. Only short periods of time may be available for the computation of correlation functions and test functions, and this will require highly sophisticated storage facilities.

8.2 Suggestions for Future Development

Up to now a method of identifying transmission changes within a given spectrum has been given. It will, however, be desirable to "trap" that portion of the spectrum, in which changes have been detected, for further analysis. The subsequent analysis, after "trapping", will aim at extracting the intelligence in the signal which has caused the change in the spectrum. Hence, a phase locking device or some "trapping" mechanism in the final stages of the search process would be a considerable advantage. Such a device should be capable of reconstructing the "trapped" signal from the available data and presenting it for further examination. It is, therefore, hoped that further work in this direction will be undertaken.

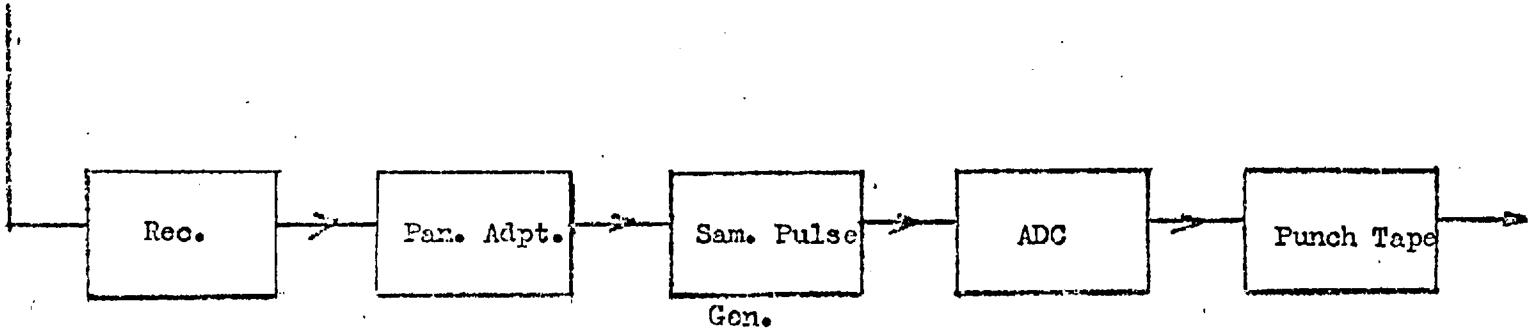
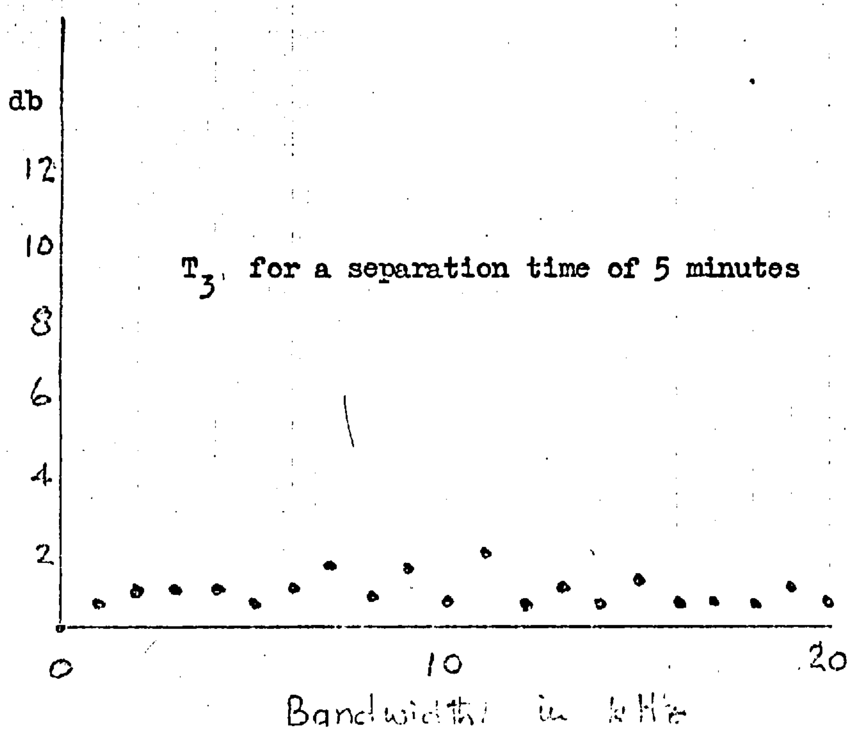


Fig. 7.7 A block diagram of the arrangement used in the "separation time" experiment

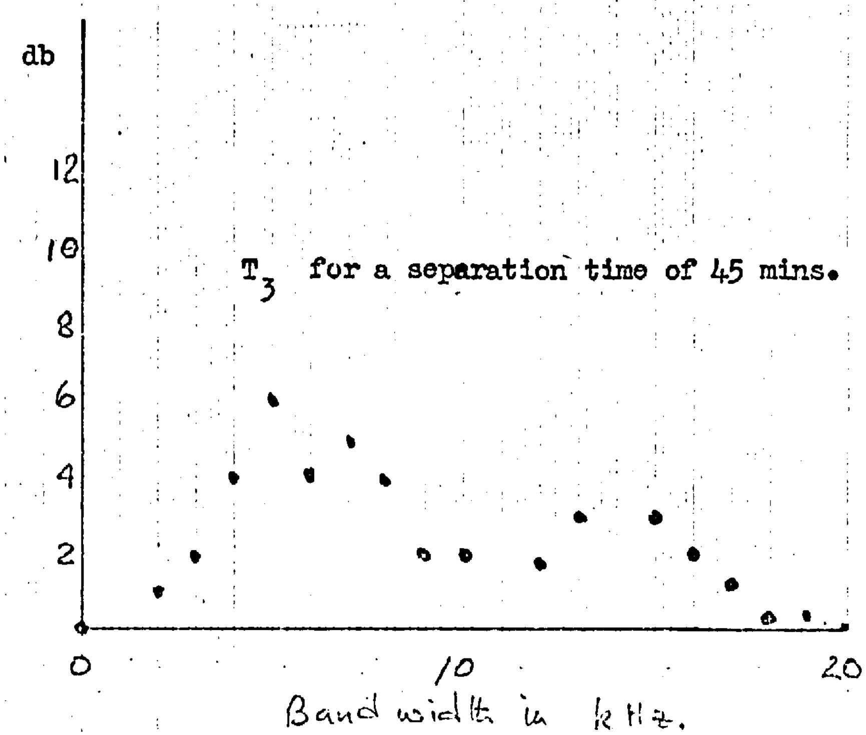
Fig. 7.8 Test function for various separation times

a)



T_3 for a separation time of 5 minutes

b)



T_3 for a separation time of 45 mins.

APPENDIX A

A Characteristic Bandwidth of Random Noise

Random noise arises in an electrical circuit when a large number of elementary events are superimposed with random time, or phase relationship. Such noise has a continuous frequency spectrum due to the random occurrence of the elementary events and, therefore, the noise power in any small frequency interval is directly proportional to the size of the interval. To see this consider a receiver with a frequency response specified in terms of the complex transfer function g(f) between the input and the output. g(f) may be represented as

$$g(f) = \int_{-\infty}^{\infty} G(f)\, e^{j\phi(f)} \qquad \ldots (0.1)$$

where g(f) is the modulus and is hence the quantity determined by measurement of the gain versus frequency characteristics. $\phi(f)$ expresses the phase information of the transfer function. The bandwidth, B, of the receiver is the width of g(f) between 3 db points below the midband value, G_o. The power bandwidth is customarily defined as

$$B_o = \frac{\int_o^{\infty} G^2(f)df}{G_o^{\ 2}} \qquad \ldots (0.2)$$

This is the average bandwidth of the curve referred to the maximum, $G_o^{\ 2}$. In general, the amplitude and phase of the frequency components of a voltage applied to an amplifier are modified on passage through the receiver. Thus if the gain modulus of amplifier is G and the mean square value of the input random noise voltage is $e^2(f)df$, then the mean square value of the output noise is

$$E^2 \quad = \int_0^{\infty} G^2 \; e^2(f)df. \qquad \ldots (0.3)$$

Combining (0.2) and (0.3) and making use of the mean value theorem*(L-4).

$$E^2 \;=\; e^2(f_m)G_o{}^2B$$

where f_m is frequency lying within the pass-band. If, as is usually the case, the input noise is only a slowly varying function of frequency, f_m and f_o can be considered identical so that

$$E^2 \;=\; e_o{}^2 \; G_o{}^2 \; B_o.$$

Hence the r.m.s. value of the output voltage is proportional to the square root of the power bandwidth(79).

* See I.S. Sokolnikoff and R.M. Redheffer, "Mathematics of Physics and Modern Engineering" pg 380, McGraw-Hill Book Co. Inc. New York, N.Y.

APPENDIX B

The distribution of the amplitude of random noise has been investigated both theoretically and experimentally.(80-81) The root-mean-square (r.m.s.), or the effective value, E, of the noise having an instantaneous voltage, V, is defined by

$$E = \sqrt{\lim_{T \to \infty} \frac{1}{T} \int_0^T V^2(t)dt.}$$

It can be shown that the probability that V lies in the range V, V + dV is given by

$$dP_v = \frac{1}{E\sqrt{2\pi}} \; e^{-V^2/2E^2}, \; dV$$

which is the Gaussian or the normal distribution. Hence the probability that the magnitude of the instantaneous voltage exceeds a certain voltage V_o is, therefore,

$$P_v = 2\int_{V_o}^{\infty} \frac{1}{E\sqrt{2\pi}} \; e^{-V^2/2E^2} \, dV = 1 - \operatorname{erf}\left(\frac{V_o}{\sqrt{2}\;E}\right)$$

in which $\operatorname{erf}(V_o/\sqrt{2}\,E) = \frac{1}{\sqrt{2\pi}} \int_{V_o/\sqrt{2E}}^{\infty} e^{-t^2/2} \, dt$

in which erf is the error function.

The probability that the envelope lies between A and A + dA is given by

$$dP_A = \frac{A}{E^2} \; e^{-A^2/2E^2} \cdot dA$$

which is the Raleigh distribution. Therefore, the probability that the envelope exceeds a certain value, A_o, is given by

$$P_A = \int_{A_o}^{\infty} \frac{A}{E^2} e^{-A^2/2E^2} dA$$

$$= e^{-A_o^2/2E^2}$$

Fig. B.1 shows the relative probability densities $E\frac{dP_v}{dv}$ and $E\frac{cP_A}{ds}$ for instantaneous voltage and envelope, respectively, and Fig. shows that V or A shall exceed a certain value.

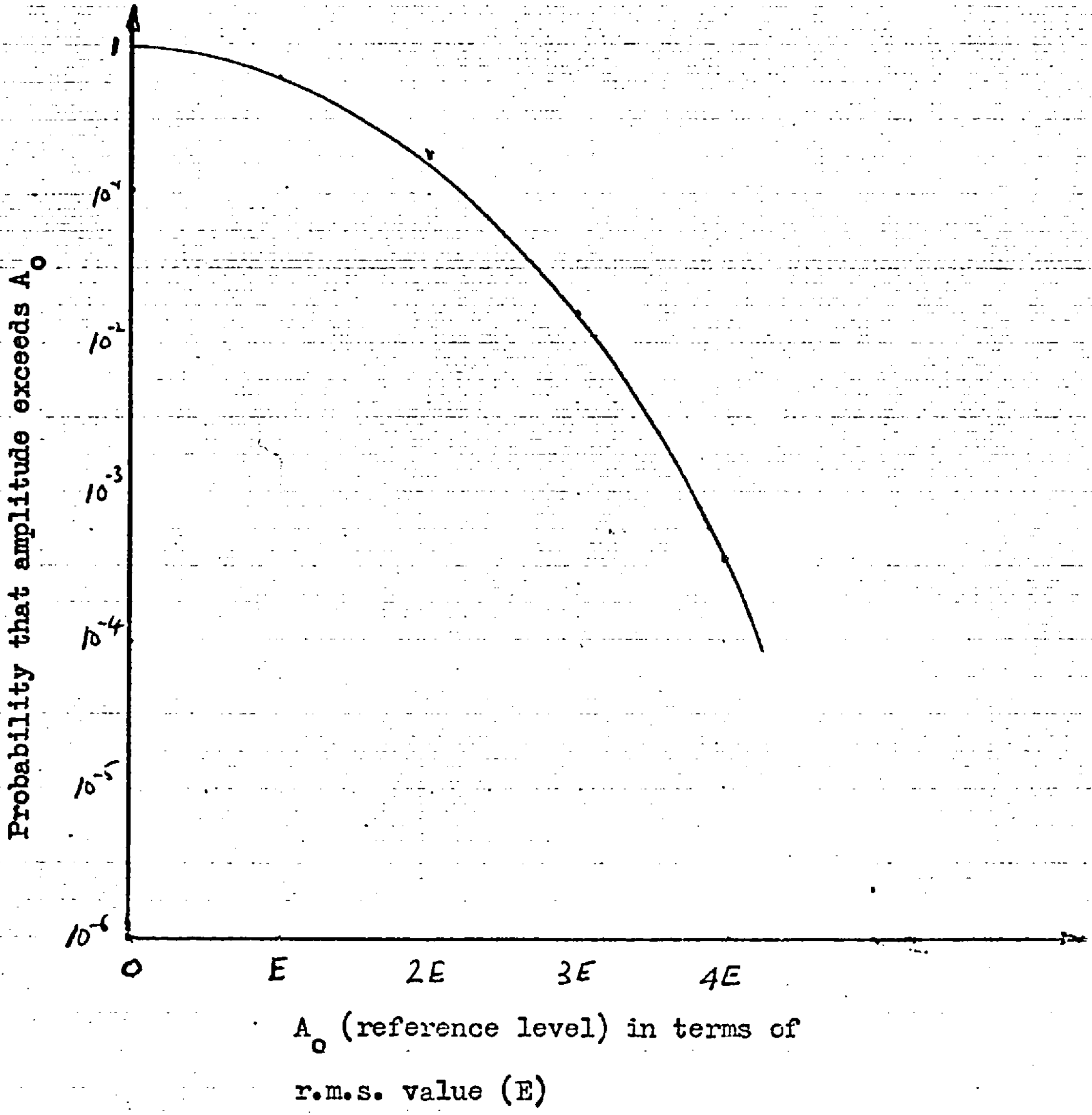


Fig. B1 Probabiliäy function of the instantaneous noise fluctuations

APPENDIX C

A Peak-holding Circuit Using an Operational Amplifier

A peak holding circuit senses and remembers the highest positive (or negative) peak in a varying signal over some specified period of time, Fig. C.1.

When a negative portion of the waveform is applied to the input of the circuit, it is amplified and inverted by the operational amplifier. If the positive charge appearing on C is less than the negative value of the input value of the waveform at that instant, the operational amplifier will provide the necessary gain to overcome the forward-diode drop in D1 and will also provide current to charge C to a value equal to the input but opposite in polarity.

If the charge on C is already greater than input, the operational amplifier output will be negative. When the input signal goes positive the output of the amplifier will not cause a change of charge on C.

The emitter-follower serves to isolate C and thus allows it to hold its charge for an extended period.

It should be noticed that the peak voltage is across the feedback resistor R_3 and not across the storage capacitor.

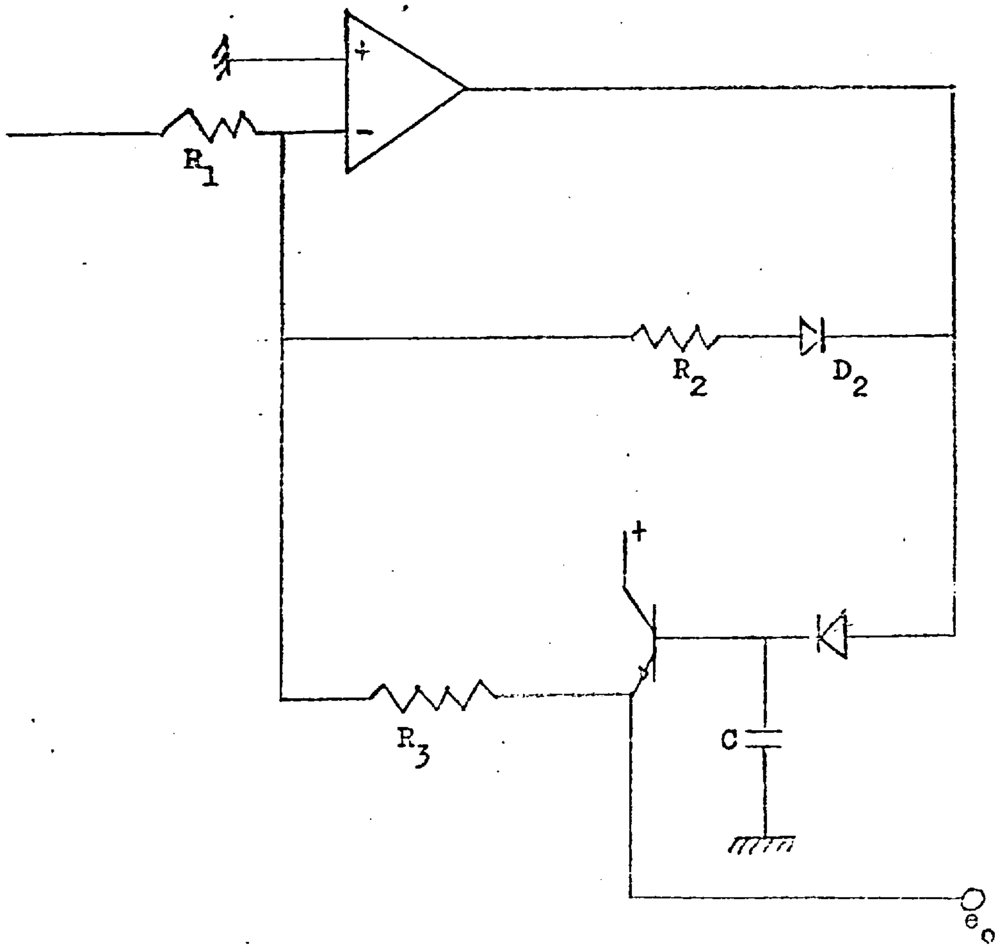


Fig. C1 A peak holding circuit

LIST OF REFERENCES

(1) E. C. Cherry, "Pulses and Transients in Communication Circuits", pp.222-262, Chapman and Hall, London, 1949.

(2) "Radio Spectrum Utilization", A Report of the Joint Tech. Advisory Comm., Inst. of Elect. & Electronics Engrs and Electronic Industries Assoc., 1964.

(3) F. E. Terman, "Radio Engineers Handbook", pp.

(4) W. H. Lee and R. W. Cannon, "The Development of Long Distance Point-to-Point Communication, since 1947", I.E.E. Conv. on hf Communication, E.D.4, pp.417-421, March 1963.

(5) E. V. D. Glazier and H. R. L. Lamont, "Transmission and Propagation", pp.437-495, HMSO, 1966.

(6) E. V. Appleton, "A Method of Measuring Upper Atmospheric Ionisation", Nature, 127, p.197, 1931.

(7) E. V. Appleton and R. Naismith, "Some Further Measurements of Upper-Atmospheric Ionisation", Proc. of Roy. Soc., vol. 50, Series A, pp.685-708, 1935.

(8) H. G. Booker and S. L. Seaton, "Relation between Actual and Virtual Ionospheric Height", Phys.Rev., vol.57,p.87,Jan.1945.

(9) E. V. Appleton and R. Naismith, "Some Measurements of Upper-Atmospheric Ionisation", Proc.Roy.Soc., vol.137, Series A, pp.36-54, 1932.

(10) A. S. Namba, "General Theory on the Propagation of Radio Waves in the Ionised Layer of the Upper-Atmospheric",Proc.I.R.E., vol.21, p.238, Feb. 1933.

(11) L. W. Austin, "Solar Activity and Radiotelegraphy", Proc.I.R.E., vol. 20, p.280, Feb. 1932.

(12) K. Steenivasan, "On the Relation between Long-Wave Reception and Certain Terrestial and Solar Phenomena", Proc.I.R.E., vol.17, p.1793, Oct. 1929.

(13) R. Naismith, et al, "A New Method of Ionospheric Forecasting", Proc. I.R.E., vol. 109B, No. 44, pp.125-130, Mar. 1962.

(14) E. J. Baghdady, "Diversity Techniques", Lectures on Communication System Theory", Chapt. 7, McGraw-Hill Book Co. Inc., New York, N.Y.

(15) P. M. Woodward, "Probability and Information Theory with Applications to Radar", pp.129-130, Pergamon Press, New York, N.Y., 1964.

(16) S. Goldman, "Frequency Analysis, Modulation and Noise", pp.55-56, McGraw-Hill Book Co., Inc., New York, N.Y., 1948.

(17) E. T. Whittaker and G. N. Watson, "A Course of Modern Analysis", Cambridge Univ. Press, New York, N.Y.

(18) D. Gabor, "Theory of Communication", J. Inst. Elect. Engrs., 93 (III) p.429.

(19) K. Rajappan, "Automatic Detection of Radio Signals in the HF Band", Ph.D. Thesis, Imperial College of Science and Technology, London, 1965.

(20) T. Y. Young, "Statistical Epoch Detection of Overlapping Signals", John Hopkins Univ., 1963.

(21) "Radio Regulations" General Secretariat of the Int. Telecom. Union, Geneva, 1959.

(22) As in (21).

(23) I. A. Bolshakov, "Distinguishing Signal Parameters during Extraction from Noise".

(24) B. M. Dwork, "The Detection of a Pulse Superimposed on Fluctuation Noise", Proc. I.R.E., 38, p.771.

(25) J. J. Bussgang, "Cross-correlation function of Amplitude Distorted Signals," Tech. Report No. 216, M.I.T., March 1952.

(26) T. Kailath, "Correlation Detection of Signals Perturbed by a Random Channel", I.R.E. Trans. on Inf. Theory, vol. IT-6, pp.361-366, June 1960.

27) M. Ezckiel, "Methods of Correlation Analysis", John Wiley and Sons, Inc., New York, N.Y., 1960.

(28) E. G. Bowen, "A Textbook of Radar", Camb. Univ., Press, Cambridge.

(29) P. Kleinrock, "Detection of Energy Peaks of an Arbitrary Signal", Lincoln Lab. (M.I.T.) Tech. Rept. No. 325, Aug. 1963.

(30) W. B. Davenport, Jr., and W. L. Rost, "An Introduction to the Theory of Random Signals and Noise", p.67, McGraw-Hill Book Co. Inc., New York, N.Y., 1958.

(31) Y. W. Lee, "Statistical Theory of Communication", p.208, John Wiley and Sons, Inc., New York, N.Y., 1963.

(32) D. A. S. Fraser, "Non-parametric Methods in Statistics", John Wiley and Sons, Inc., New York, N.Y., 1957.

(33) H. Sohon, "Supervisory and Control Equipment for Audio-frequency Amplifiers", Proc.I.R.E., vol.21, p.228, Feb. 1963.

(34) S. J. Mason and H. J. Zimmermann, "Electric Circuits, Signals and Systems", John Wiley and Sons, Inc., New York, N.Y., 1960.

(35) F. A. Gunther, "Topospheric Scatter Communication - Past, Present and Future", IEEE Spectrum, pp.79-100, Sept. 1966.

(36) D. G. Brennan, "Linear Diversity Combining Techniques", Proc.I.R.E., vol. 47, pp.1075-1102, June 1959.

(37) J. L. Holsinger, "Digital Communication over Fixed Time-Continuous Channels with Memory - with special Application to Telephone Channels", Lincoln Lab.(M.I.T.) Tech.Rept., No.366, Oct. 1964.

(38) As in (37).

(39) R. W. Sittler, "An Optimal Data Association Problem in Surveillance Systems", IEEE Trans. on Mil. Electronics, p.125, April 1964.

(40) R. F. Daly and C. F. Rushforth, "Non-parametric Detection of a Signal of Known form in Additive Noise", IEEE Trans. on Inf.Theory, vol. IT-11, pp.70-76, Jan. 1965.

(41) As in (32).

(42) H. Cramer, "Mathematical Methods of Statistics", Princeton Univ. Press, 1945.

(43) W. Feller, "An Introduction to Probability Theory and its Applications", John Wiley and Sons, Inc., New York, N.Y., 1964.

(44) L. A. Weistein and V. D. Zubakov, "Extraction of Signals from Noise", Translated by R. A. Silverman, Prentice-Hall Inc., Englewood Cliffs, N.J., 1962.

(45) As in (15).

(46) As in (15).

(47) I. L. Davies and P. M. Woodward, "Information Theory and Inverse Probability in Telecommunication", Proc. IEE, vol. 99, Part III, p.37, 1952.

(48) D. Middleton, "Introduction to Statistical Communication Theory", McGraw-Hill Book Co. Inc., New York, N.Y., 1957.

(49) A. M. Mood and F. A. Graybill, "Introduction to the Theory of Statistics", McGraw-Hill Book Co. Inc., N.Y., N.Y., 1963.

(50) G. S. Sebestyen, "Recognition of Membership in Classes", I.R.E. Trans. on Inf. Theory, vol. IT-6, pp.44-50, Jan. 1961.

(51) T. R. Goodman, "System Identification and Prediction - An Algorithm Using a Newtonian Iteration", Proc. Q. Appl. Math., Oct. 1966.

(52) R. L. Grinsdale, "A System for the Automatic Recognition of Patterns", Jour. of I.E.E., 106, pt. 3, pp.210-221,

(53) C. Helstrom, "Statistical Theory of Signal Detection", Pergamon Press, New York, N.Y., 1960.

(54) S. M. Sussman, "Least-Square Synthesis of Radar Ambiguity Functions", I.R.E. Trans. on Inf. Theory, vol. IT-8, p.246, April 1962.

(55) J. H. Reyner, "Radio Interference and its Suppression", Chapman and Hall Ltd., London, pp.6, 1936.

(56) H. A. Thomas and R. E. Burgess, "Survey of Existing Information and Data on Radio Noise in the Frequency Range, 1-30 mHz". Special Report, Radio Research Board, H.M.Stationary Office, London, 1947.

(57) R. A. Watson and J. F. Herd, "An Instantaneous Direction Reading Goniometer", Journal I.E.E., vol.64, pp.611, May 1926.

(58) A. E. Harper, "Some Measurements on the Directional Distribution of Static", Proc. I.R.E., vol. 17, p.1214, July 1929.

(59) R. W. Tremellen and J. W. Cox, "The Influence of Wave Propagation on the Planning of Short Wave Communication", Journal I.E.E., 94, Part IIIA, p.200, 1947.

(60) M. Ryle, "Radio Astronomy", The Physical Society Reports on Progress in Physics, 13, pp.184, 1950.

(61) Aldert van der Ziel,"Noise", Chapman and Hall Ltd., London, pp. 1955.

(62) C.C.I.R. Rept. No. 65, Revision of Atmospheric Radio Noise Data", I.T.U., 1957.

(63) E. C. Cherry, "On Human Communication", John Wiley & Sons Ltd., 1961.

(64) C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication", The Univ. of Illinois Press, Urbana, 1963.

(65) As in (61).

(66) C. Clarke, "Atmospheric Radio-Noise based on Amplitude-Probability Measurements at Slough, England, during the International Geophysical Year", Proc. I.R.E., Vol.109, No.47, part B, Sept. 1962.

(67) As in (2).

(68) U. Steudel, "The Sensation of Loudness and its Measurement", Hochfrequen and tech. u. Elektroakust, 41,pp.116-128, 1933.

(69) R. E. Kalman and J. E. Bertram, "A Unified Approach to the Theory of Sampling Systems", Jour. of the Franklin Inst., No.V, vol.267, May 1959.

(70) J. W. Tukey, "Sampling Theory of Power Spectrum Estimate", in Symposium on Applications of Autocorrelation Analysis to Physical Problems", Woods Hole, June 1949. (Office of the Naval Research).

(71) E. C. Cherry, M.H.Kubba and D.E.Pearson, "An Experimental Study of the Possible Bandwidth Compression of Visual Image Signals", Proc. IEEE, vol.51, No.11, pp.1507-1517, Nov. 1963.

(72) N. I. Akliezer, "Theory of Approximations", Frederick-Ungar, New York, N.Y., 1956.

(73) M. Schwartz, W. Bennett, S. Stein, "Communication Systems and Techniques", pp.48-54, McGraw-Hill Book Co. Inc., New York, N.Y., 1966.

(74) L. Weiss, "Statistical Division Theory", McGraw-Hill Book Co.Inc., New York, N.Y., 1961.

(75) As in (64).

(76) S. O. Rice, "Distribution of the Duration of Fades in Radio Transmission", Bell Syst. Tech. Jour., vol. 37, pp.581-635, May 1958.

(77) As in (2).

(78) As in (35) and (36).

(79) G. A. Campbell and R. M. Foster, "Fourier Integrals for Practical Applications", Bell Syst. Monog., 584, 1931.

(80) K. Franz, "The Amplitude of Noise Voltages", Elect. Nachr. Tech., 1942, 19, pp.166-173.

(81) V. D. Landon, "The Distribution of the Amplitude with Time in Fluctuation Noise", Proc.I.R.E., 1941, vol.29, pp.50-55.