THE CITY UNIVERSITY
DEPARTMENT OF SYSTEMS SCIENCE

"NOISE POLLUTION ANALYSIS AND RECOGNITION"

by
Pantelis MOUKAS

A thesis submitted for the
award of the degree of
Doctor of Philosophy in Systems Engineering

February 1982

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

A	Matrix of augmented feature vectors
B	Matrix of weights
c	Velocity of sound
c_j	Class j
$c(0.5)$	Correlation coefficient for 50% overlap
C	Cepstrum of subrecord
d	Dimensionality of feature space
d^*	Dimensionality if augmented feature space
D_i	Fourier series coefficients
e_r	Mean standard error
f	Frequency
f'	Apparent frequency
f_c	Bandwidth
f_N	Nyquist frequency
f_s	Sampling rate
f_s'	Effective sampling rate
Δf	Frequency sampling interval (nominal resolution)
F	Output spectrum
F_p	Spectral ripples period in Hz
g_i	Discriminant function
h	Microphone height in metres
H	Flight path height in metres ; Transfer function
H_f	Frequency response of high-pass filter
j	$\sqrt{-1}$
K	Acoustic wave number
l	Loss function
l_d	Direct acoustic wave path
l_r	Refelcted acoustic wave path
Δl	Difference of wave paths
L	Window shift
L_f	Family of ground reflection ripple curves
L_s	Number of samples in signal record
L_{10}	10% sound level

L_{50}	50% sound level
L_{90}	90% sound level
L_{eq}	Equivalent sound level
m	Smoothing parameter
M_b	Number of bytes per "block"
n_c	Number of classes
N	Length of sequence (window size)
N_b	Number of bytes per record
N_i	Number of "blocks" per record
N_w	Number of windows in signal
N_w'	Number of windows in subrecord
N_w''	Smallest power of 2 > N_w'
p	Sound pressure; Probability density
p_o	Reference sound pressure
p_d	Sound pressure of direct wave
p_r	Sound pressure of reflected wave
P_x	Modif. periodogram (power spectrum) of sequence x
P_x	Average estimate of power spectrum from consequtive windows; Probability estimate
r	Overlap coefficient of windows
R	Conditional risk
r_R	Time base expansion coefficient
R_p	Playback speed
R_r	Record speed
R_s	'True' autocorrelation function of signal s
s	Sound signal
s^*	Sampled signal
S	Spectrogram
SPL	Sound pressure level
S_s	'True' power spectrum of signal s
t	Time
T	Sampling interval
T_b	Time taken for "block" transfer
U	Compensation factor for windowing
v	Velocity of flying source
v_a	Variation of average amplitude of subrecord

V_a	Discrete Fourier Transform of v_a
w	Window weighting sequence
$\underline{w}_i$	Weight vector
W	Discrete Fourier Transform of w; Matrix of weights
$\underline{W}_i$	Augmented weight vector
x	Sequence of samples from sampled signal s*
x_w	Windowed sequence x
X	Discrete Fourier Transform of x
$\underline{X}$	Feature vector
X_f	High-pass filtered spectrum
X_s	Scaled spectrum
$\underline{X}_s$	Average scaled spectrum;
	Average spectrum of subrecord
$\tilde{\underline{X}}_s$	Smoothed average scaled spectrum
X_m	Modified spectrum
X_t	Thresholded spectrum
X_w	Discrete Fourier Transform (spectrum) of x_w
γ	Window function (probability density)
δ	Dirac 'delta' function
ρ	Slope of tranjectory
σ	Standard deviation
ϕ	Angle
ψ	Angle
ω_o	Fundamental radian frequency

ACKNOWLEDGEMENT

I wish to express my gratitude to my parents. Without their support I would have never been engaged in this work, nor would I have been able to carry it out to whatever stage it has reached without their encouragement and patience.

I am deeply indebted to professor Ludwik Finkelstein for his authoritative supervision, guidance and invaluable help; also for his general direction of this project and the facilities he has provided as head of the Instrument Systems Centre, now Measurement and Instrumentation Centre.

My thanks are due to Mr L Norton-Wayne for his supervision and suggestions during the major part of the project. Also to the [redacted]

[redacted]

Finally, my thanks to everyone in the Department of Systems Science who make it one of the most hospitable and friendly places to work.

ABSTRACT

Instrumentation currently available for the automatic monitoring of noise nuisance has the shortcoming that although the intensity, duration and time of occurence of noises may be recorded, their source often cannot be identified.

This report describes research directed towards providing improved instrumentation which can identify noise pollution sources by exploiting the structure of the sounds they emit. The structure, which is due to the operating characteristics of the sources is revealed by the digital computation and display of Fourier spectrograms of digitized records of their sounds. In addition, the effect of ground reflections which introduces ripples in the spectra and is characteristic of the altitude of the source can be extracted by Cepstral analysis. Quasi-stationary fast periodicities of the signal appear in the spectrogram as quasi-straight lines along the time axis, whereas slow impulsive periodicities appear as isolated groups of spectra with increased total power and different structure.

Averaging spectra over a short periods (approximately 0.5 seconds) preserves the fast periodicities. The square root of the power in certain frequency bands of these average spectra can then be used as features in a feature space classifier. In a similar fashion, the cepstral "power" bands and the "energy" bands that result from the Fourier transform of the function representing the variation of the average amplitude with time can be used as additional features.

The scheme described above has been simulateded on a general purpose digital computer and has been tested with real sounds of jet aircraft, helicopters and trains. A suboptimal search has been done to select the "best" feature set, by sequencially designing and evaluating a linear classifier with one, two and so on features. Recogntition better than 95% has been achieved using 6 spectral features only. When the additional features were included the performance was consistently imroved by about 2% when 9 features were used. The effect of the width of the bands decreases with the number of features; however, 240Hz bandwidth was found optimum.

NOISE POLLUTION ANALYSIS AND RECOGNITION

to my parents

1 INTRODUCTION

Environmental noise pollution, although not an entirely new phenomenon, has been one of the principal nuisances that modern societies have to face. Noise pollution is caused by the same factors that cause all other environmental nuisances (air, water), i.e. by increased population, urbanization, technological change, but especially by the usual relegation of environmental considerations relative to economic ones. Since 1963, when the first major study of noise pollution policy was carried out in the United Kingdom, many measures have been taken by local authorities to monitor noise pollution in order to contain it in prescribed limits, provide development decisions etc. The latest monitoring system includes unattended instruments that record noise levels in problematic areas. As it will become clear in chapter 2, the data supplied by the existing system, which is incapable of identifying the offending sources, can be unreliable on one hand and give an incomplete 'noise picture' of the areas been monitored on the other. In this light, the Scientific Branch of the Greater London Council and the Instrument Systems Centre at the City University have started work towards the conceptual design of an instrument for the identification of sources causing noise pollution. The eventual task is the enhancement of the already existing noise monitoring instrumentation and hence improve environmental planning and control.

Solution to problems of identification is offered by pattern recognition techniques. The identification problem is formulated so that the identifiable sound signal is represented by 'meaningful' measurements, the so called features. Well developed classification methods exist (reviewed in chapter 3) that can assign a class to the features, hence the signal, using statistical decision theory. However, the success of all classification methods heavily depends on the 'quality' of the features representing the signal. The 'quality' of the features, in turn, depends on the thorough analysis of the problem and the signals to be recognized, so that their consistent properties, characteristic of their class, are fully defined and understood.

From the above discussion, it looks that feature extraction is more problem-oriented than classification. The investigation reported here has, therefore, concentrated mostly on the former and due to the iterative nature of the design process has employed digital simulation throughout. Digital simulation, of course, requires digitization of the data. Hence, this work has also been involved with specific problems arising from the properties and limitations of the available digitization system and the duration and high data rate of the acoustic records.

In brief, this work has involved digitization and spectral analysis of recorded sounds, using a general purpose digital computer with graphic and interactive facilities, to investigate methods of signal processing capable of identifying the sounds. Chapter 2 gives the framework of the noise pollution problem together with some definitions regarding its measurement, short and long-term effects etc. Chapter 3 gives a review of the available processing and classification methods that offer solution to problems of signal analysis and identification. The first part of pattern analysis is presented in chapter 4 where the properties of noise pollution sounds likely to enable the extraction of features for identification of the sources are examined. Chapter 5 is an exposition of the data acquisition system that was developed to enable digital analysis of the acoustic signals. The development of the digitization system together with the solution of the specific problems mentioned previously is also presented in this chapter. Chapter 6 is the continuation of pattern analysis and includes the spectral analysis methods developed for the extraction and display of the properties referred to in chapter 4. The definition and extraction of candidate features is given in chapter 7. In the same chapter the evaluation of candidate features leading to the selection of the 'best' feature set is also presented. Finally, in chapter 8 a scheme for the hardware implementation of a noise pollution source identifier is suggested. Chapter 9 is the summary and conclusion of the thesis including a few suggestions for further development of this project.

CHAPTER 2

2 THE NOISE POLLUTION PROBLEM

Noise is generally defined as an unwanted signal, that is a signal either carrying no information or whose information is unwanted. This definition applies to sound signals when referring to acoustic noise which is the result of the transfer of mechanical vibrations to air. What is striking in this definition is that the judgement whether a signal is noise or not depends on the human beings who are the recipients of the signal, in this case sound. As this 'subjectivity' leaves ground for misunderstandings, it is left to the law which regulate human conduct to define in 'objective' terms what is noise. With the aid of various areas of human activity and study, physics, physiology, psychology, sociology, engineering, etc., various attributes of noise are defined, other claiming some universality, others being applicable to certain cases only. This happens because the response of human beings to sounds varies with their cultural background, age, condition of health, profession, etc. It also varies with time. For example, tolerance to noise is lower when one wants to work, rest or sleep.

Generally, the degree of annoyance depends on the loudness of the sound. But it also depends, to a lesser degree, to the pitch (high pitched sounds are generally more disturbing than low pitched), duration, impulsive nature, harmonic content, etc. The complexity of the response of human beings to noise makes difficult to measure noise in a way that reflects human response to it. This has led to the definition of many measures tailored to particular types of noise or human response. These measures are treated in section 2.1.1 of this chapter.

The progress of technology and industry, and the consequent urbanization has created what is termed <u>environmental pollution</u>, noise pollution being a form of it. Monitoring pollution is a part of the activities taken to control and contain it using instruments tailored to the particular task. Monitoring noisy areas allows assessment of the sources of noise, regarding loudness, frequency content, behaviour, etc. The information obtained by monitoring is used for policing, for determining the effectiveness of the measures taken and for planning (redistribution or banning the offending sources, provision of licenses, planning of new residential areas etc.).

Unattended monitoring is the latest scheme employed by local authorities who are vested with the task of controlling noise. The noise levels and times of occurences of noise recorded by the unattended instruments do not guarantee, however, that the levels registered come from the sources of interest. This deficiency of the system is examined in section 2.1.2. In that section an extension of the system is suggested, namely its endowment with a capability of identifying the offending source, and the advantages offered by such a system are examined. Section 2.2 outlines the requirements and constraints of an instrument capable of identifying sources, to be used in conjunction with the existing system. Finally, in the same section, the approach to the construction of such an instrument is outlined.

2.1 Noise Pollution and Identification

Noise pollution, although not an entirely new phenomenon but rather a problem that has grown worse with time, is caused by the same factors which have brought us air and water pollution: population increase, industrialization, urbanization, technological change, all these combined with the usual relegation of environmental considerations relative to economic ones.

The term pollution indicates the degradation of the environment to a stage where normal activity is hindered. Ambient noise is always present, although it is sometimes low and possibly unnoticable. However, noise levels frequently occur which are high enough to constitute a nuisance and to provide a hazard to health and efficient work performance. Not only is, thus, the quality of life lowered and normal activity hampered, but danger signals and warning shouts can be masked with fatal results. Generally, communication by sound is made extremely difficult in noisy areas. Further, exposure to intense noise can lead to hearing loss which appears as a shift of the hearing threshold. Aircraft noise and surface transportation noise can cause abnormalities in sleep. In addition, there are a lot of psychological effects that can be caused by noise. According to the Environmental Ministers of the OECD countries at meetings in 1979 and 1980 [OECD, 1980], the general public is becoming less willing to accept excessive noise levels.

Surveys show that in terms of the number of people affected, road traffic noise causes most annoyance, but aircraft noise is generally considered to be the most intrusive source [SANDO, 1978]. In addition, trains, industrial processes, construction works, public entertainment such as speedway racing and pop concerts, noisy neighbours and even the hum of electrical transformers can cause an intolerable nuisance. Of course, the nuisance caused is subjective, in that the nature of the sound and the conditions under which it is heard rather than merely its amplitude and frequency content determine the extent of disturbance.

2.1.1 Measurement of Acoustic Noise

In air sound is transmitted as pressure fluctuations about the mean atmospheric pressure. To be audible, these fluctuations must have frequency (or frequency components) between 20 and 20,000 Hz. If the amplitude of the fluctuations is p the sound pressure level is given by

$$SPL = 10 \log_{10} (p/p_o) \text{ dB} \qquad (2.1)$$

where p_o is the reference pressure of $2x10^{-5}$ N/m^2.

The choice of a logarithmic scale was imposed by the fact that the range of audible sound pressures is extremely wide (from the reference pressure above to 20 N/m^2). In addition, the human ear has actually a logarithmic response. On the other hand, it has a variable sensitivity to sounds of different frequencies and levels: it is less sensitive at low frequencies, when these occur at low levels. This subjective evaluation of equal sound pressure levels, referred to as loudness, is measured in phons and is equal to the sound pressure level at 1000Hz.

This 'weighting' of the ear is taken into account when measurements are made through the weighting networks of the sound level meter. Measurements of this kind do not, therefore, indicate 'true' pressure levels but 'subjective' ones referred to as sound levels. Although the weighting of the ear varies for different sound pressure levels, to avoid complication, only three weightings, A, B and C are commercially used, corresponding to equal loudness contours of 40, 70 and 100 dB, respectively, as shown in figure 2.1 (D weighting, shown in the figure, is rarely used in aircraft noise measurements).

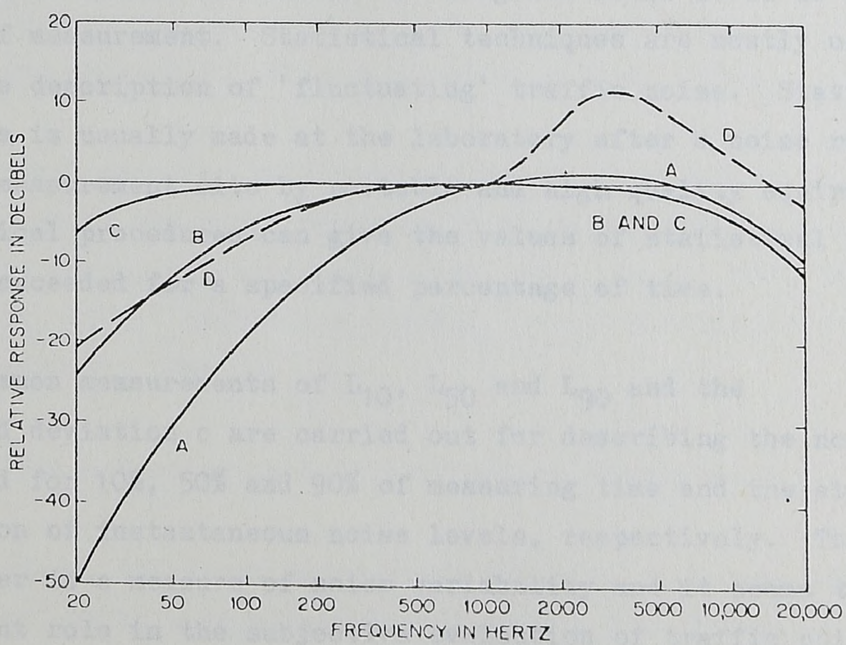


Figure 2.1. Frequency response of sound level weightings. (from Harris [1979])

The corresponding sound levels are depicted as dB(A), dB(B) and dB(C), respectively. In general, 'A' weightings are more suitable for levels under 55 dB, 'B' weightings for 55 to 85 dB and 'C' for high sound levels. Usually 'A' measurements are closest to subjective evaluation of loudness, 'B' and 'C' having fallen into disuse.

Typically, an ambient noise level of 50 to 60 dB can be expected in an office environment, whilst levels of 80 dB upwards (generated for example by low flying aircraft) provide a nuisance which is difficult to eliminate by sound insulation. A sound level of 120 dB is painful, and may lead to permanent damage to hearing.

When noise is not constant but varies with time, it is often important to have information of the time history of the noise so as to give information about the dose of noise received. This has led to different criteria of evaluating noise. The different criteria all have in common the use of 'A' weighted sound level as a basic scale of measurement. Statistical techniques are mostly used for a complete description of 'fluctuating' traffic noise. Statistical analysis is usually made at the laboratory after a noise recording at the measurement site by portable and high quality equipment. Statistical procedures can give the values of statistical noise levels exceeded for a specified percentage of time.

Common measurements of L_{10}, L_{50} and L_{90} and the standard deviation o are carried out for describing the noise level exceeded for 10%, 50% and 90% of measuring time and the standard deviation of instantaneous noise levels, respectively. The latter parameter is a measure of noise variability and it seems to have an important role in the subjective evaluation of traffic noise. Roughly, L_{10} represents the average of noise peaks, L_{50} is near the mean level for dense traffic and L_{90} may be considered as the background noise. For reasons of economy, samples of less than 60 min are taken over 24h and are found to give good results for homogeneous or static sources. In the U.K. the L_{10} has been further defined to be the arithmetic average of the 18 hour L_{10} levels between 06.00 and 24.00 hours on a normal weekday [VULKAN, 1975]. This measure only is used to assess the annoyance from traffic for the purposes of the Land Compensation act [1973, Amended 1975]. The measuring apparatus must be of sufficient high standard for ± 2dB(A).

A measure which is increasingly established is the <u>equivalent (continuous) sound level</u> given by:

$$L_{eq} = 10 \log_{10} \{ (1/100) \sum_i f_i \; 10^{(L_i/10)} \} \qquad (2.2)$$

where f_i is the percentage of measuring time in the i^{th} sound level class, usually a 5 db(A) interval, and L_i is the corresponding pressure level. The equivalent sound level is, in effect, the continuous sound level which would produce the same amount of energy at the measuring point as the actual fluctuating level during the measuring period. The equivalent level is used when the noise level fluctuates by more than ± 3dB(A).

Other measures have been suggested in an effort to approximate more closely the human response to noise in conditions of surface traffic noise, air transport noise, etc. (e.g. TNI, NNI, L_{NP} [STARR, 1972; BARDUCCI & CANNELI, 1975; MULHOLLAND & ATTENBOROUGH, 1981]).

2.1.2 Noise Monitoring & Identification of sources

In Britain, the first major study of noise pollution policy was carried out by a committee headed by Sir Alan Wilson in 1963 [COMMAND 2056, 1963] "to examine the nature, sources, and effects of the problem of noise, and to advise what further measures can be taken to mitigate it". The Wilson report, as it became known, suggested standards for maximum noise levels from motor vehicles and recommended various environmental standards of noise control. In the U.K. the Government have decided that action against environmental noise should generally be taken at local level by either strategic or local authorities. Various legislation is available to these authorities to deal with noise nuisance or to facilitate compensation. Powers are also vested at local level to provide planning and development decisions which can be used to improve the environment [MOUKAS et al, 1982]. The Greater London Council has taken an active interest in traffic noise with a policy statement in 1966 and a publication of a design bulletin on traffic noise [GLC, 1970].

The Noise Advisory Council succeeded the Wilson committee and has issued various recommendations from 1970 on surface and air traffic noise, industrial noise and acceptable levels of noise in residential areas: "In no circumstances should existing residential development be subjected, as an act of conscious public policy to L_{10} levels of 70 dB(A) unless some form of remedial or compensatory action taken by the responsible authority". Also, "An L_{10} level of 70 dB(A) constitutes, in our view, the limit of acceptable rather than a standard view of what is desirable. Wherever possible, planners should design to lower levels".

The first important step towards environmental noise control is the assessment of the magnitude of existing and potential noise nuisance. Traditionally, noise measurements have been made using attended instruments; either by tape recording to obtain permanent records for subsequent detailed spectral and statistical analysis, or by short-term sampling, using sound level meters to obtain percentile and equivalent sound levels. Both methods suffered from several problems: when batteries are used to power the instruments they have to be changed every 8 hours. Furthermore, the accuracy of tape recorders as a means of recording audible sound for later analysis is in question. Cold and humid weather may affect the accuracy of the recorder. Also, statistical processing at the laboratory is wasteful in manpower.

Recently, the collection of basic data has been aided by the development of automatic noise monitoring instruments. The Greater London Council in collaboration with London Boroughs make extensive use of an automatic monitoring system, namely the Local Authorities Aircraft Noise Monitoring System (LAANMS). LAANMS is being used to obtain long term data from aircraft using Heathrow Airport [VULKAN, 1979]. The data are used to provide basic information to assist both local and strategic planning and to determine the effects of measures taken to control aircraft noise. The equipment consists of a simple low capacity data logging device which records either unweighted or 'A' weighted noise levels sampled at predetermined intervals of 0.5 seconds or more [FREEDMAN <u>et al</u>, 1974]. For LAANMS the equipment is used in conjunction with noise level sensitive triggering. The data logger is the Micro Data Multi channel logger, 'Model 200'. The recorder consists of a small cassette tape deck powered by a 6 volt 7 ampere-hour nickel cadmium battery. The tape is moved on incrementally each time a 'word' of data is recorded and the sampling period is adjustable and can be preset by means of a switched crystal block. The output of the logger is recorded as a number between 0 and 255 giving a potential accuracy of better than 0.5% of full scale. At the end of a recording period, the cassette is removed from the logger and returned to the laboratory for translation, using a Microdata Translator unit 'Model 200', and processing of data for the computation of percentile levels and the equivalent sound level. The processing is done by a dedicated mini computer.

Freedman and his colleagues [1974] report that the computation of 18h L_{10} obtained from this system was in agreement within 0.4 dB(A) with measurements obtained by the conventional procedure using a level recorder and statistical analyzer, the microphones and amplifiers being the same in both measurements. According to Vulkan [1975] the advantages of this system are the simplified analysis without more time and the compactness, accuracy, ease of setting up and speed of analysis. The digital form of recording overcomes the problems of sensitivity to weather conditions in contrast with the conventional level recorders.

The disadvantage of the system is that when left unattended there is no way of checking that no spurious noises affect the recording. This happens because the instrument cannot discriminate between the various sources which may be present. Thus, long term measurements of an environmental noise source may be invalidated by the unexpected occurence of a spurious loud sound from a different source. Careful selection of the monitoring site can reduce the risk, but with no guarantee of the "purity" of the data collected in this way.

When used to monitor aircraft, it is at the sites near the airport that there is the highest confidence that all loud noise events are in fact due to aircraft overflights. Analysis of the time history of each event and of the pattern of occurence of events over a period of time, can be used to confirm identification where high signal to noise ratios occur and the operational attitudes of the aircraft are known [MOUKAS et al, 1982]. Where signal to noise ratios are small, such a simple system of recognition becomes unreliable, just when the inclusion of spurious events or the exclusion of valid events can have the greatest effect on the final result. At such sites the survey results are often most significant and may show unexpected conditions due to aircraft straying from the authorized flight track. If instruments can be pre-conditioned to respond only to one particular type of noise stimulus or to identify and record the type of source that has produced a particular noise record, then the conditions under which automatic noise monitoring can be reliably used will be considerably extended.

One possible early application may be to control the noise from heliport operation without the need to compare, under strict test conditions, the noise of each helicopter type against local noise standards [MOUKAS et al, 1982]. Heliports in London have restrictions placed on the number of movements that may be operated by each type of helicopter according to the peak noise that the helicopter makes with respect to the GLC helicopter guideline for the heliport [SIMSON, 1980]. These restrictions are designed to contain the noise exposure from the helicopter to an acceptable level and therefore require the planning authority to assess the noise of every helicopter type against its guideline. A development of the LAANMS could be used to identify helicopter movements from the general background noise and register the noise history of each event so that heliport operators can identify helicopters which are exceeding reasonable noise limits. Executive action can then be taken to surcharge or ban environmentally unacceptable operation. Other applications could include automatic noise monitoring where several sources are present simultaneously, or where intermittent noisy events occur in a high ambient noise environment.

2.2 Instrument Requirements & Design Methodology

The advantages of an enhancement of the LAANMS discussed in the previous section lead the Scientific Branch of the G.L.C. and the Department of Systems Science and the Instrument Systems Centre at the City University to define the requirements of an instrument capable of identifying the noise pollution sources, to be used in conjunction with LAANMS, as shown in figure 2.2. These requirements would guide the research directed towards investigating signal processing and recognition methods appropriate to this task. The requirements were:

1) The enhancement should be based on signals obtainable from the sensing apparatus of the existing system, that is its microphone. Of course the amplifier and the available signal conditioning and weighting networks could also be used (fig. 2.2).

2) Recognition of sounds should be done on-line with a processing rate fast enough to be synchronous with the current rates of one recognition decision per half a second.

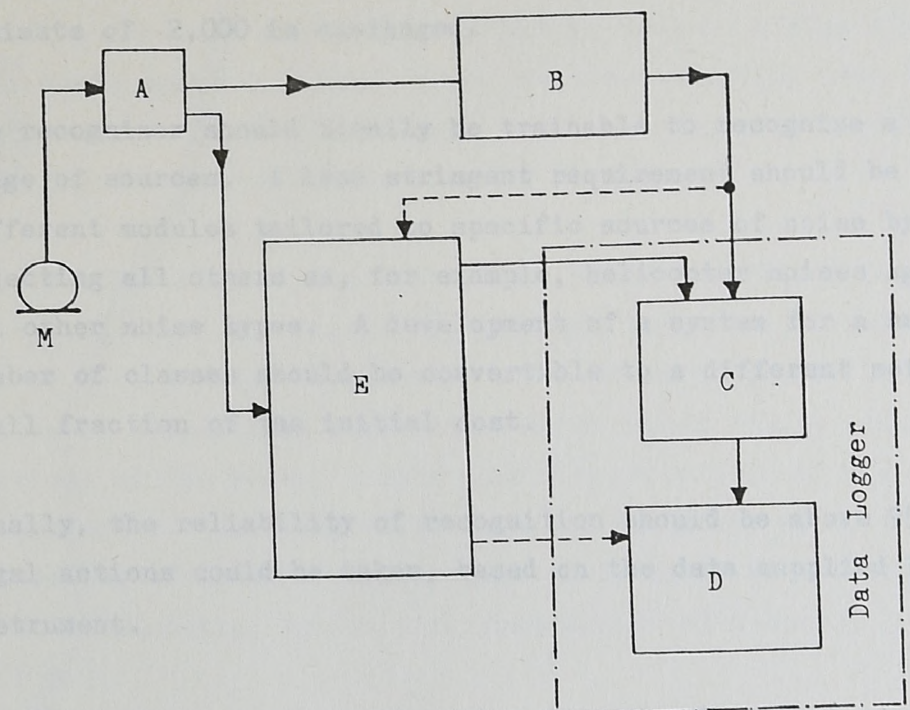


Figure 2.2. Source identification instrument.
--- Control signal, —— Data, M: Microphone, A: Amplifier and signal conditioner, B: Sound level meter, C: Digital encoder, D: Cassete unit, E: Source identifier.

3) The eventual hardware implementation of the recognizer should ideally be in the form of a 'plug-in' module whose output, suitably coded, should be recorded on the cassette unit, together with the 8-bit coded 'A' weighted sound level produced by the Microdata logger. Alternatively, when the recognizer is to be used to condition the instrument to respond to a particular type of noise, the output of the recognizer should act as a 'GO' signal to gate the output of the logger to the cassette unit, as shown in figure 2.2.

4) The cost of endowing the Microdata, or any similar system, with recognition capability should be commercially acceptable. Although this requirement is difficult to quantify, a rough estimate of 2,000 is envisaged.

5) The recognizer should ideally be trainable to recognize a wide range of sources. A less stringent requirement should be to have different modules tailored to specific sources of noise by rejecting all others as, for example, helicopter noises against all other noise types. A development of a system for a small number of classes should be convertible to a different set at a small fraction of the initial cost.

6) Finally, the reliability of recognition should be above 95%, since legal actions could be taken, based on the data supplied by the instrument.

The first three requirements arise from the fact that minimum modification of the existing system is desirable. These requirements impose serious restrictions to the approach for a solution of the recognition problem. Firstly, the system to be developed must use acoustic signals only for processing, and in fact the single signal obtained from the single microphone of the existing device. This implies that, no spatial information can be used to aid recognition from acoustic signals. For example, if, in a measuring site, traffic noise is predominantly on the left of the instrument, and rail noise on the right, an array of two or more microphones, properly positioned, would give additional information that could reduce the need for sophisticated spectral analysis of the acoustic signal.

The on-line requirement is another restriction. As mentioned in the previous section, the data logging system used by LAANMS, samples the sound level every 0.5 seconds. The recognizer must take decisions also every half second, by processing the incoming waveform. Assuming that frequencies above 8kHz are not informative, there are, at least, 16,000 samples to be processes. Assuming also that 8 bits are sufficient (they give an acceptable 50dB dynamic range) there are 128,000 bits to be processed. This, at first sight, does not seem a very serious restriction. However, to meet the cost and size requirements, the only processor that can be used is a slow micro. On the other hand, spectral computations are time consuming even for a larger processor. These restrictions have to be born in mind when investigating for appropriate processing methods.

The cost and size of the instrument, which are restrictions implied by the third and fourth requirement, limit the amount of store that can be used in the instrument. This, in turn, prohibits the utilization of information which could be obtained from a history of the behaviour of the noise, longer than 0.5 seconds or of any duration of the same order. Such information could, for example, be derived from processing of spectrograms as digital pictures to extract line structures representing the varying frequencies of a source [NICHOL, 1977 & 1979].

Requirement 5 implies that the technique employed for the design of a noise recognition system should not be tailored to a particular type of noise. It should rather use information which is likely to exist in all classes of noise of interest, or to as many as possible. The inclusion of a specific characteristic could only be justifiable when it improves the performance considerably, at no significant extra cost. Such an inclusion should not, ideally, be at the expense of the generality of the approach, so as to allow adaptability of the system to new environments with minimum effort and cost. The adaptability of the instrument should be dependent upon the adjustment of certain parameters of the system.

Finally, it must be emphasized that, since the monitoring instrument is deployed in a 'real' environment (in the sense that no restriction is imposed upon the behaviour of the noise sources, neither any restriction is possible) the design of the recognizer should be based on 'real' data, in other words on data recorded on locations where, either the existing monitoring system is already deployed, or on prospective monitoring sites. This is a serious restriction and unique in that all other noise identifying systems developed, or under development, require controlled conditions of monitoring in 'artificial' environments, as will be discussed in the next chapter.

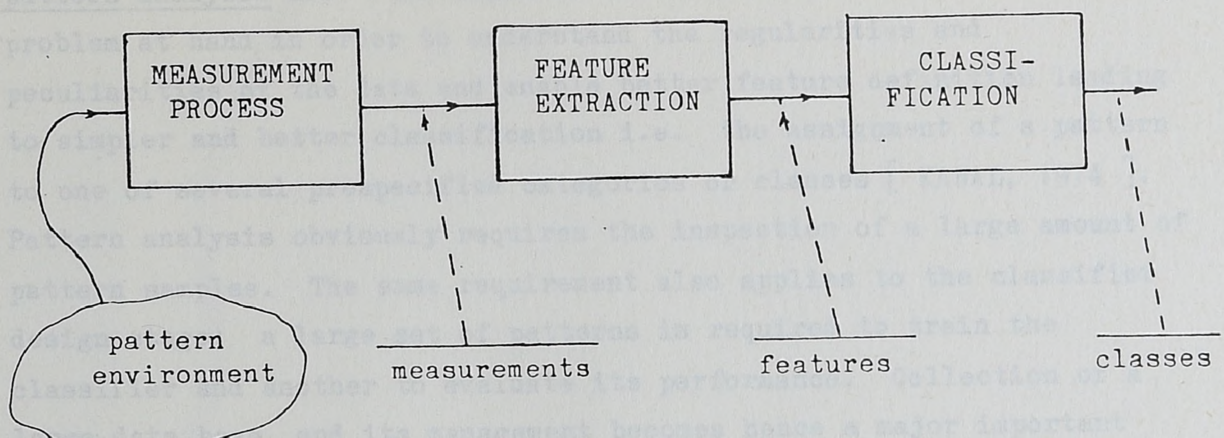


Figure 2.3. Schematic representation of a Pattern Recognition System.

Solution to identification problems is offered by pattern recognition. A model of a pattern recognition system may be schematically viewed as in figure 2.3, where three stages are involved, namely measurement process, feature extraction and classification. The boundaries between these three stages are not sharp and feedback between them is often needed. This is best seen in the design of a pattern recognition system which is a highly iterative process: Assuming that the measurements of the patterns are given, a set of features is derived from them and a classifier is designed which is also used to evaluate the features. A "bad" set of features requires a more sophisticated classifier and vice versa. This is a first indication of the interaction involved between these two stages in the design of a pattern recognition system. On the other hand, the assumption that we know a priori the necessary measurements requires elaboration: the initial measurements were decided upon by our intuition about the nature of the features; the insight gained during the design process may lead to re-examination of the necessity of some measurements (resulting in economies in sensor hardware and data processing) or may lead to new ones.

The features referred to previously are entities that are derived from the initial measurements. The derivation is led by what is termed pattern analysis i.e. the exploitation of whatever we know about the problem at hand in order to understand the regularities and peculiarities of the data and enable better feature definition leading to simpler and better classification i.e. the assignment of a pattern to one of several prespecified categories or classes [KANAL, 1974]. Pattern analysis obviously requires the inspection of a large amount of pattern samples. The same requirement also applies to the classifier design stage: a large set of patterns is required to train the classifier and another to evaluate its performance. Collection of a large data base, and its management becomes hence a major important factor in the design of a pattern recognition system.

2.3 Summary

In this chapter, it was shown that the increase of population, industrialization and urbanization caused, among other nuisances, noise pollution. To monitor this nuisance so as to contain it in acceptable standards, a need exists for an instrument capable of identifying the offending noise sources. The instrument should in fact be an enhancement of already existing monitoring systems. It has been decided to enhance the Local Authorities Noise Monitoring System which involves logging devices recording noise levels digitally on ordinary cassette tapes. The enhancement envisaged should be a commercially viable 'add-on' module using the sensing and signal conditioning units of the existing system and requiring minimal modification of it. The ~~the~~ output of the module should be a digitally encoded description of the source. This requirement implies that the recognizer should recognize the noise sources from the acoustic signal obtained from the single microphone of the existing system. Simple signal processing and pattern recognition methods should be used in order to comply with the requirement of low cost and real-time operation. An investigation and critique of the available methods, examined in this context, follows in the next chapter.

CHAPTER 3

3 SIGNAL PROCESSING AND CLASSIFICATION METHODS

The analysis of the noise pollution problem presented in the previous chapter led to the conclusion that a contribution to the solution of the problem can be offered by better monitoring of the offending sources by means of instruments capable of identifying them from the sound signals they emit. Clearly, the emitted signals contain the information sufficient for identification since, in most cases, humans can identify the sources by just hearing. Or is it not so? The process of recognition of the noise source in the human brain is not based on auditory data only; visual information, experience from the frequency of past occurences in the particular site, direction etc. play an important role in recognition. Nonetheless, the primary source of information is acoustic and this is the information that has to be used for the identification system of noise pollution sources.

The properties of the sources, in so far as their emitted sound signals is concerned, are treated in detail in the next chapter. Here we may only mention that the majority of them consist of revolving parts that generate periodic or quasi-periodic sounds. This observation immediately leads to the study of their spectra which give a concise description of the periodicities or frequencies involved and hence to the constructional characteristics of the sources i.e. their identity.

Spectral analysis has long been used for the study of the frequency content of signals. Although the fundamental theoretical tools used are orthogonal transforms, the Fourier transform in particular, their direct application to the analysis of actual signals is hindered mainly by the following two reasons: a) The signals to be analysed are not available for indefinitely long periods of time and b) are frequently of random nature and/or are contaminated by noise. Applicable techniques taking into account these limitations, i.e. techniques for the spectral analysis of samples of noisy signals have had a great development due to their relevance to the analysis of time series, the design of communication systems etc. Most of the well-

-developed techniques, however, are applicable under the assumption of ergodicity i.e. the ensemble statistics of the stochastic process modelling the signal are equal to time averages and of stationarity i.e. that the statistics of the process do not change with time. While the first condition is fulfilled in most practical problems, the second is difficult to assume, especially in the study of noise pollution sound signals which are the subject of this thesis. However, although these signals cannot be expected to be stationary, it is necessary to discuss the techniques available in order to assess their suitability for the treatment of these signals and select the conditions under which analysis is possible.

Section 3.1 of this chapter is dedicated to the discussion of modern digital techniques of power spectral estimation and the associated problems caused by digitization and truncation of the original acoustic data. Older techniques are briefly mentioned in order to introduce the modern ones where necessary and in order to provide the historical context of the latter. The section begins with a presentation of the digitization of analogue signals and the associated problems of aliasing and quantization errors; spectral analysis follows with references to windowing and smoothing and statistical problems and the section finishes with brief account of cepstral analysis techniques. Cepstral analysis was first introduced by Bogert et al [1963], as a technique for finding echo arrival times in a composite signal. Since then, it has been used in such diverse fields as radar and sonar, speech processing etc. As it will be shown in the next chapter, the environmental acoustic signals which are received by the sensor are composite signals consisting of direct and reflected components. Their cepstrum will reveal the reflections and thus help in providing information about the position of the source relative to the sensor and the ground.

The power spectrum is a means of representation of a signal in a compact form. It also provides a reduction of dimensionality by a factor of two (section 3.1.2.1) and it is invariant to time translation. The cepstrum has the capability of separating convolved signals and of revealing possible echoes in the waveform. As mentioned in the previous chapter, the classifier (which is the last stage of the pattern recognition system) assigns a class to an input pattern represented by a set of measurements. In the case of recognition of noise pollution sounds, one would expect that the information contained in the spectral and cepstral components, and possibly information regarding their time-variation, to be the information required by the classifier to perform its task. It is not certain, however, whether all components are necessary, perhaps some of them might lead to confusion between the types of noise to be classified. The use of all components, although might give some positive results, it would certainly not be economical if feasible at all. The problem is hence to take 'meaningful' measurements from the spectra and cepstra and feed these to the classifier. The extraction of meaningful measurements or features presupposes a detailed study of the nature of the problem which is the subject of the next chapters. No matter what the features might be, though, they must reflect the nature of the sources i.e. they must be representative of their identity. Their suitability can be ascertained, as it is argued in section 3.2.4, by their discriminating performance on the particular classifier on which they are to be used.

There are two general approaches to design a classifier. The linguistic or structural approach and the statistical feature space approach. So far no linguistic method has been used to classify spectra. Generally linguistic pattern recognition is applied where the pattern is well formed i.e. it has some systematic ordering and is not subject to noise and also when description rather than simple classification or labelling is required. One would expect that such a classifier would be unsuitable for the noisy patterns of this project. Feature space techniques, on the contrary, have good tolerance to noise and some of them do not require advance knowledge of the statistics of spectra. Also, the interfacing between the measurements fed and the structure of the classifier is much looser than in the case of a linguistic classifier. Section 3.2 is devoted to a brief review of feature space techniques with particular emphasis on a form of a linear classifier which is used for the classification of the spectra of this project because of its simplicity and ease of implementation. Finally, section 3.3 sums up and discusses the findings of this chapter and relates them with similar work in signature and speech analysis and recognition.

3.1 Digital Signal Analysis

To mention the advantages of digital techniques might look trivial, especially nowadays. Since 1975, when the first microprocessor was commercially available, anything that is subject to 'processing' has been affected. Signal processing, however, has been influenced earlier and the concepts of analogue to digital conversion digital filtering etc. are quite familiar. Hence, the purpose of this section is to merely introduce these concepts and stress those points only that are relevant to this project. Also to try and clarify the concepts which, due to an 'adventurous' past have had different definitions and interpretations in different fields and under different points of view. This especially applies to the techniques of spectral analysis which were revolutionized with the introduction of the Fast Fourier transform.

This section comprises three parts. The first deals with a brief exposition of digitization of analogue signals, the second with the spectral analysis of signals with particular emphasis on digital spectral analysis and related problems and the third deals with cepstral analysis which is relevant to composite signals i.e. signals that result from the convolution of two or more simpler signals.

3.1.1 Digitization

For the digital analysis of analogue data it is necessary that the data be converted into a series of discrete numbers. This process is termed <u>digitization</u> and may be separated into <u>sampling</u> and <u>quantizing</u>. This section presents general aspects of digitization and the related problems. The actual implementation of digitization to acoustic signals relevant to the work presented in this thesis and the solution of specific problems posed by the nature of these signals is dealt with in the chapter on data acquisition.

3.1.1.1 Sampling and Aliasing

Sampling is the process of defining the values of a signal at certain instants. If $s(t)$ is the analogue signal, the sampled version of it at times $t = kT$, for $k = 0, 1, \ldots$ is:

$$s^*(t) = \sum_k s(kT)\ \delta(t-kT) \tag{3.1}$$

where $\delta(t)$ is the Dirac delta function. The ratio $f_s = 1/T$ is called the <u>sampling rate</u>.

It is important to have a sufficient number of samples to describe the rapid variations that may exist in the signal. On the other hand, too many samples yield redundant and correlated data which increase the labour, length and cost of calculations with no profit. The <u>sampling theorem</u> states that the original signal $s(t)$ can be uniquely reconstructed from the sampled train $s^*(t)$, if $s(t)$ is band-limited to a frequency f_c (i.e. its Fourier transform is zero for $f > f_c$) and the sampling rate is chosen to be at least twice f_c. The frequency $f_N = f_s/2$ is known as the <u>Nyquist frequency</u> [BRIGHAM, 1974]. Frequencies in the signal that are higher than the Nyquist frequency are "folded" into the low frequency range from 0 to f_N and are confused with data in the lower range. This

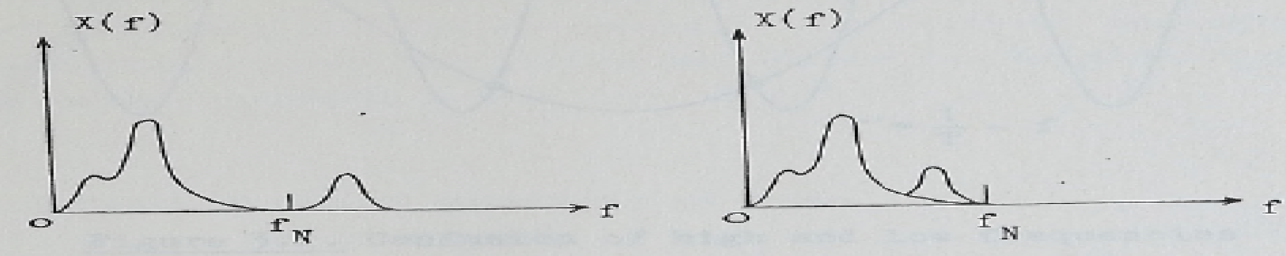


Figure 3.1. Effect of "folding" of high frequencies in the spectrum of an aliased signal

phenomenon is called aliasing and its effect on the spectrum of the signal is shown in fig. 3.1. A component of frequency $f > f_N$ will be confused with a lower component f' where $f' = f_N - (f - f_N)$, as shown in fig. 3.2 where the two sinusoids f and f' share the same sample points. In fact aliasing causes all frequencies given by $2nf_N \pm f$ to be indistinguishable from f, where $f < f_N$, with disastrous results. Figure 3.3 displays the aliasing effect on the spectrogram of a tone of linearly increasing frequency. Although the frequency of the tone increases above the Nyquist frequency, the spectrogram displays a "reflection" back towards the lower frequencies.

There exist two practical methods of overcoming the aliasing problem. The first is to choose the sampling interval T sufficiently small so that the Nyquist frequency is higher than the expected bandwidth of the signal. A sampling rate of 2.5

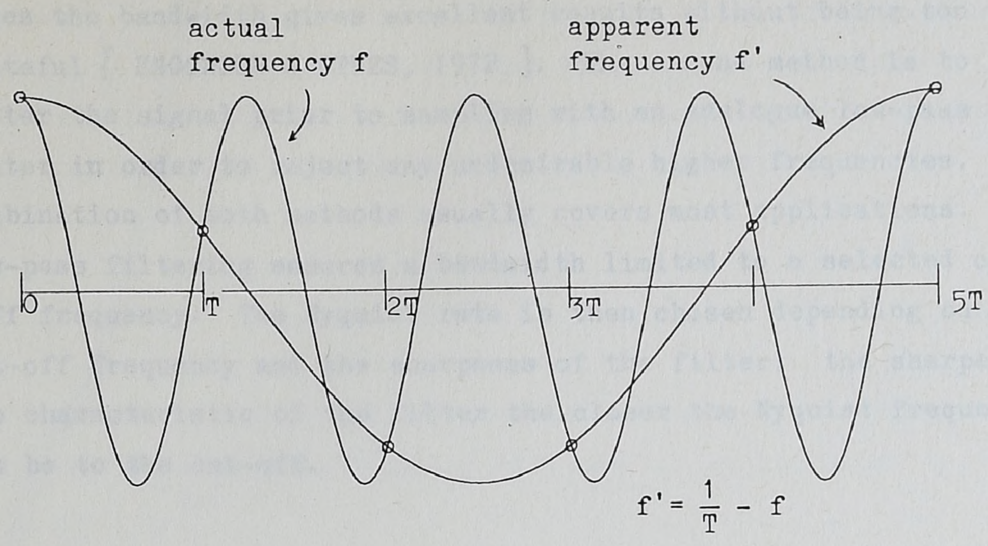


Figure 3.2. Confusion of high and low frequencies due to aliasing.

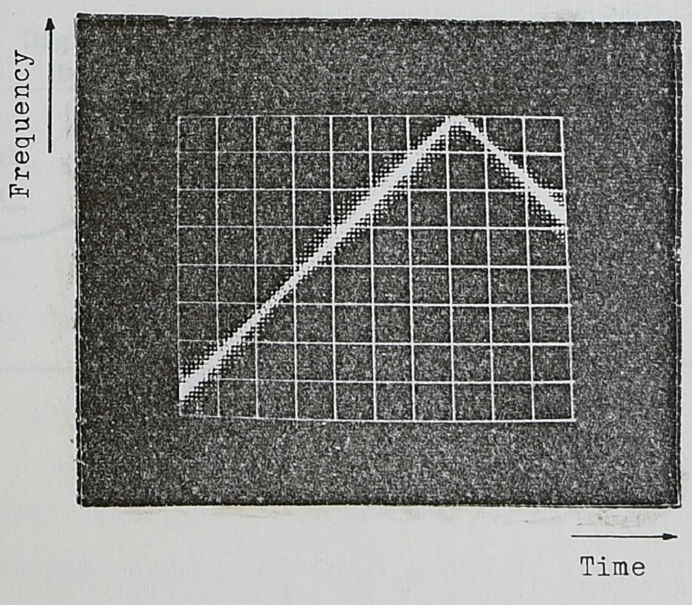


Figure 3.3. Spectrogram of an aliased tone of increasing frequency

times the bandwidth gives excellent results without being too wasteful [ENOCHSON & OTNES, 1972]. The second method is to filter the signal prior to sampling with an analogue low-pass filter in order to reject any undesirable higher frequencies. A combination of both methods usually covers most applications. Low-pass filtering ensures a bandwidth limited to a selected cut--off frequency. The Nyquist rate is then chosen depending on the cut-off frequency and the sharpness of the filter: the sharper the characteristic of the filter the closer the Nyquist frequency can be to the cut-off.

3.1.1.2 Quantizing and Quantization Noise

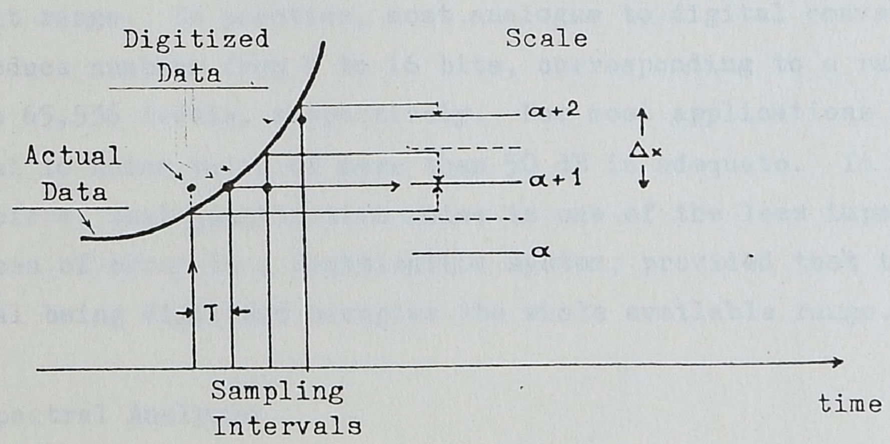


Figure 3.4. Quantization.

The second phase in the digitization process is quantization: the actual conversion of the analogue values of the sampled wave-form (which may occupy a continuum of values) to a series of numbers of finite length. No matter how many quantization levels are used, the sample level must be rounded to one of two adjacent levels, as shown in fig. 3.4. For ideal conversion, the quantization error will have a uniform probability distribution with a probability density function given by

$$p(x)=1 \ , \quad -0.5\Delta x \leqslant x \leqslant 0.5\Delta x$$
$$p(x)=0 \ , \quad \text{otherwise} \tag{3.2}$$

where Δx is the 'size' of the quantum in scale units. The mean value of the error is obviously zero and the variance is

$$\sigma^2 = \int_{-.5\Delta x}^{.5\Delta x} x^2 dx = \Delta x/12 \tag{3.3}$$

The standard deviation of the error may be viewed as the r.m.s. value of the quantization noise added to the original signal, i.e. $1/\sqrt{12}=0.29$ scale units. The signal to noise ratio for 12--bit conversion may then be calculated to be $2^{12}\Delta x/0.29\Delta x$ which is about 70 dB, provided that the signal occupies the whole 12-bit range. In practice, most analogue to digital converters reproduce numbers from 6 to 16 bits, corresponding to a range of 64 to 65,536 levels, respectively. For most applications a signal to noise ratio of more than 50 dB is adequate. It looks, therefore, that quantization noise is one of the less important sources of error in a digitization system, provided that the signal being digitized occupies the whole available range.

3.1.2 Spectral Analysis

Spectral analysis techniques have evolved together with the evolution of computers, having been influenced and directed, in a dialectic manner, by the computing power of the latter in combination with the effectiveness of the available algorithms. The most striking change has been the progressive abandonment of the mean-lagged product approach to power spectrum estimation [BINGHAM et al, 1967], as a result of the impressive reduction of computing time offered by the Fast Fourier Transform (F.F.T.) algorithm invented by Cooley and Tukey [1965].

The definition of the spectrum of a signal is mathematically simple: The <u>power spectrum</u> $S_s(f)$ may be defined as the square of the Fourier transform of the signal $s(t)$:

$$S_s(f) = \left| \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) \, dt \right|^2 \tag{3.4}$$

providing that the signal satisfies certain integration and continuity conditions [BRIGHAM, 1974]. In practice, however, discrepancies occur in in the definition not only because of the diversity of applications of Fourier analysis, but also because of the differing properties of the signals whose spectrum is studied and the degree of rigour required in each case. Since the introduction of distribution theory [PAPOULIS, 1962], the Fourier series and the Discrete Fourier transform (D.F.T.) may be directly derived from the Fourier transform and hence the spectra of discrete, periodic and aperiodic signals can be introduced in a unified way [BRIGHAM, 1975]. In the literature of time series and communication theory on the other hand, the subject is approached from a different viewpoint: the power spectrum of a stationary process is defined as the Fourier transform of the autocorrelation function of the process, whereas in the deterministic case this relation is only given as a property of the Fourier transform:

$$S_s(f) = \int_{-\infty}^{\infty} R_s(\tau) \exp(-j2\pi f\tau) \, d\tau \tag{3.5}$$

where $S_s(f)$ is the power spectral density or power spectrum and $R_s(\tau)$ is the autocorrelation function of the process $s(t)$. The autocorrelation function, which in practice is comp uted using mean-lagged products [BLACKMAN & TUKEY, 1959], has been the standard method for the estimation of power spectra of noisy signals and time series until the introduction of the F.F.T. The latter, which is just an algorithm for the fast computation of the D.F.T., has completely altered the approach to power spectrum estimation and unified - as the distribution theory did for digital, periodic and aperiodic signals - the deterministic with the stochastic case, not in terms of definition but in terms of practical estimation of spectra. Specifically, the power spectrum is estimated by directly applying the D.F.T. to a sample of the process, thus dispensing with the intermediate stage of the autocorrelation function. In fact, this was similar to the method employed before the use of mean-lagged products mentioned before called periodogram analysis in statistical literature. This term, though, fell into disrepute because of statistical errors due to the finite duration of the samples analysed [RICHARDS, 1967; BINGHAM et al, 1967]. The F.F.T. brought back the periodogram in a new form: the modified periodogram. The modification consists in weighting short segments of data with a weighting 'window' function which reduces leakage (section 3.1.2.2). In order to improve the statistics of the spectrum, the estimate is taken by time averaging over these short modified periodograms [WELCH, 1967]. This modification was the result of the knowledge acquired on the statistical behaviour of the spectra during the long use of lagged products.

As it was said in the introduction of this chapter, the main difficulties - apart from the purely computational ones - which caused these adventures were the finite duration and the noisy nature of the signals. The D.F.T., which is now generally used, is inherently finite and the problems associated with its use, namely leakage, windowing and statistical stability and reliability, follow its introduction in the next sections.

3.1.2.1 The Discrete Fourier Transform

The D.F.T. may be defined as follows: If $\{x(kT)\}$, $k=1, \ldots ,N$ is the sequence of samples representing the sampled waveform, as shown in fig. 3.5, then its D.F.T. is the (generally) complex sequence:

$$\{X(l)\} = \{{}^{1}/_{N} \sum_{k=1}^{N} x(k)\ \exp(-2\pi jkl/N)\}\ , \quad l=1,\ \ldots\ ,N \tag{3.6}$$

The sequence $\{X(l)\}$ is sometimes referred to as the spectrum of the sequence $\{x(k)\}$.

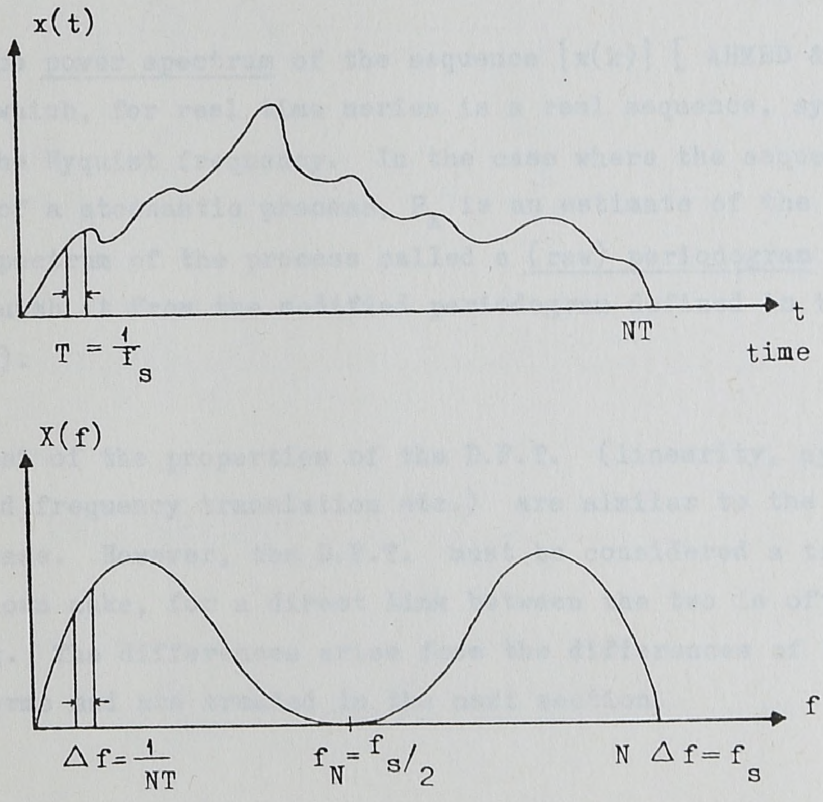


Figure 3.5. The Discrete Fourrier Transform

The definition of (3.6) assumes that the sequence $\{x(kT)\}$ is periodic with a period NT. The set of Fourier coefficients is also assumed to be periodic with a period $f_s=1/T$. The frequency spacing of the coefficients is clearly $\Delta f = f_s/N$.

This is also the minimum frequency resolution. When the time sequence is real (which is the case in the analysis of real signals) the real part of X(l) is symmetric about the Nyquist frequency (section 3.1.1.1) and the imaginary part is anti-symmetric. Thus, the second half of the transform may be seen as equivalent to the negative frequencies of the continuous case.

The square of the modulus of the D.F.T. given by:

$$\{P_x(l)\} = \{| X(l) |^2\} , l=1, \dots ,N \qquad (3.7)$$

gives the <u>power spectrum</u> of the sequence $\{x(k)\}$ [AHMED & RAO , 1975] which, for real time series is a real sequence, symmetric about the Nyquist frequency. In the case where the sequence is a sample of a stochastic process, P_x is an estimate of the power spectrum of the process called a <u>(raw) periodogram</u> (raw to distinguish it from the modified periodogram defined in the next section).

Most of the properties of the D.F.T. (linearity, symmetry, time and frequency translation etc.) are similar to the continuous case. However, the D.F.T. must be considered a transform in its own sake, for a direct link between the two is often misleading. The differences arise from the differences of the two transforms and are treated in the next section.

3.1.2.2 Leakage and Windowing

There are two major problems associated with the D.F.T. resulting from the fact that D.F.T. operates on sampled waveforms defined over finite intervals. <u>Aliasing</u>, which is a result of sampling, has already been referred to in section 3.1.1.1. <u>Leakage</u>, the second problem, is inherent in the Fourier analysis of any finite record of data formed by neglecting what happens before and after the period considered. A formal explanation of leakage associated with D.F.T. is given by Harris [1978], where the D.F.T. is defined as "the projection of the truncated waveform on an orthogonal basis set of trigonometric functions spanning the observation interval". Hence, "frequencies which coincide with the basis will project onto a single basis vector; all other frequencies will exhibit non-zero projections on the entire basis set".

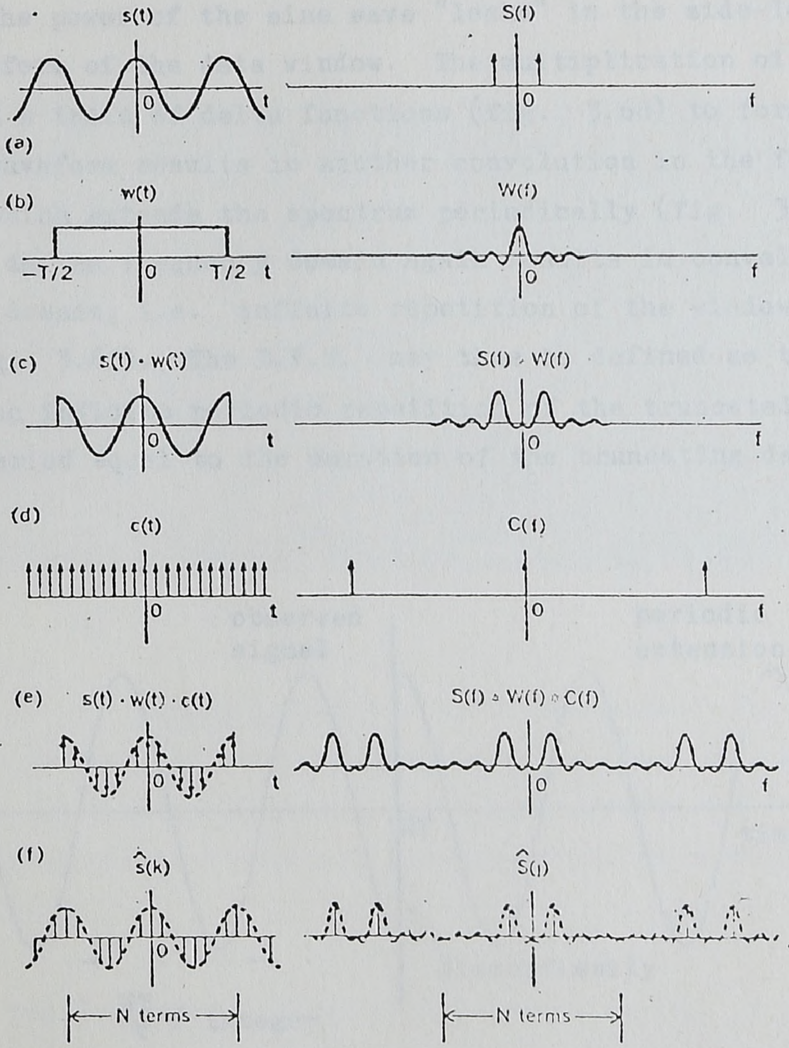


Figure 3.6. Derivation of the D.F.T. from the continuous Fourier Transform. (From Bergland [1969])

The meaning of leakage becomes clearer when the D.F.T. is derived from the continuous transform [BERGLAND, 1969; BRIGHAM, 1974]. The infinite duration sine wave of fig. 3.6a is effectively multiplied with a rectangular data window (fig. 3.6b) in the time domain. This multiplication is equivalent to a convolution in the frequency domain of their transforms. Thus, an impulse which is the transform of an infinite duration sine wave is convolved with the transform of the rectangular data window resulting in a function with a twin sinx/x shape (fig. 3.6c). The power of the sine wave "leaks" in the side-lobes of the transform of the data window. The multiplication of the sine wave with a train of delta functions (fig. 3.6d) to form the sampled waveform results in another convolution in the frequency domain, which extends the spectrum periodically (fig. 3.6e). Sampling in the frequency domain again results in convolution in the time domain, i.e. infinite repetition of the windowed waveform (fig. 3.6f). The D.F.T. may thus be defined as the transform of an infinite periodic repetition of the truncated signal with a period equal to the duration of the truncating data window.

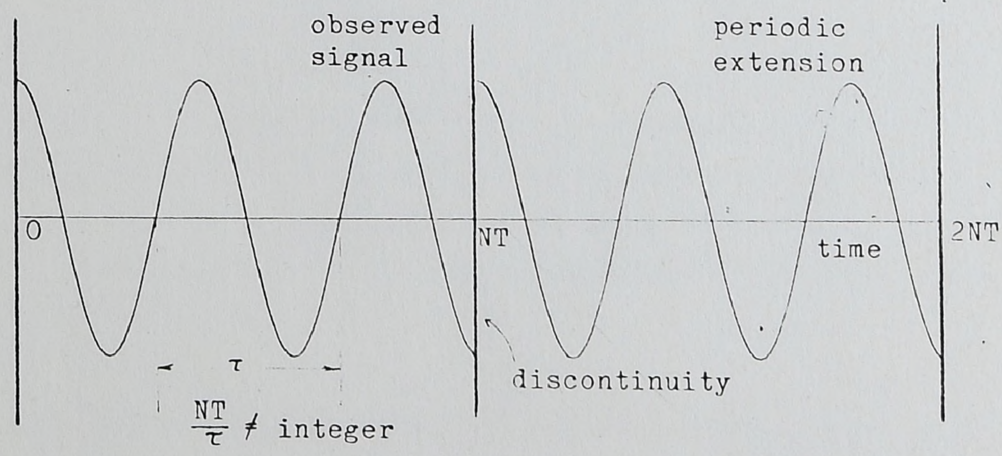
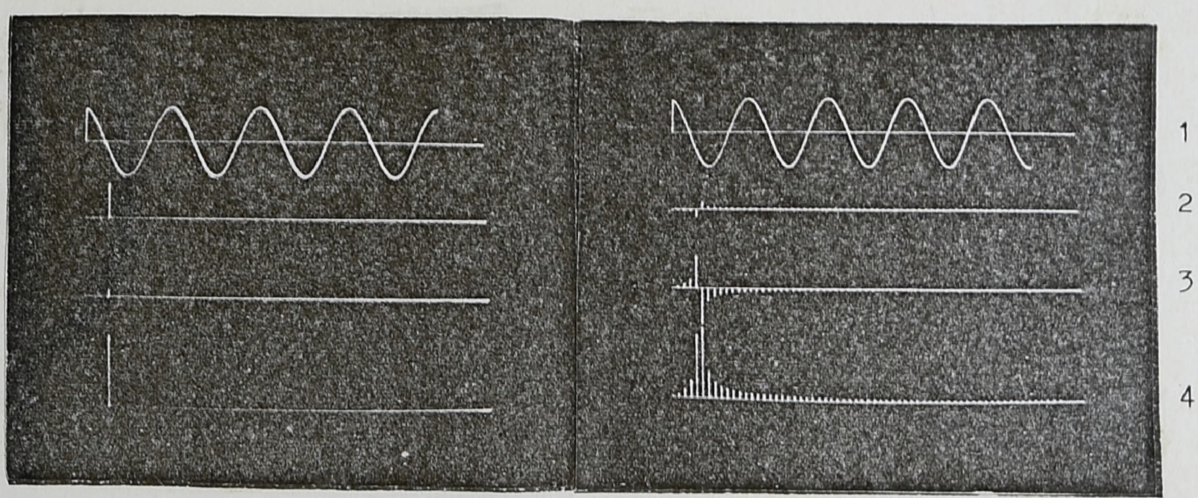


Figure 3.7. Discontinuities due to the implied periodicity of the D.F.T.

The above derivation can now lead to a somehow intuitive explanation of leakage: The implied periodic extension of the truncated waveform creates sharp discontinuities, as shown in fig. 3.7, which lead to additional frequency components in the spectrum (fig. 3.8b). Clearly, no discontinuity exists when the duration of the window is an exact multiple of all the components of the waveform, a condition rarely fulfilled in practice (fig. 3.8a).



(a) (b)

Figure 3.8. D.F.T. of a sinusoid

1. signal, 2. real, 3. imaginary, 4. power spectra.
   a) window duration exact multiple of period.
   b) window duration not exact multiple of period.

To reduce leakage, as it is understood above, it is necessary to reduce the order of discontinuity of the implied periodic waveform by windowing. Windowing is the multiplicative weighting of the truncated data in a way that smoothly reduces the data values at the boundaries to zero. The periodic extension of the data thus becomes continuous in many orders of derivative. From a different viewpoint, the so called data

<u>windows</u> may be seen as truncating functions with side-lobe characteristics of smaller magnitude than those of the rectangular window (fig. 3.6b), resulting in reduction of leakage. The computation of the D.F.T. is thus modified:

$$\{X_w(l)\} = \{{}^1/_N \sum_{1}^{N} w(k)\ x(k)\ \exp(-2\pi jkl/N)\}, \quad l=1,\dots,N \qquad (3.8)$$

where $w(k)$ is the window weighting sequence. This weighting, however, scales down the magnitude of the Fourier coefficients. The <u>modified periodogram</u> is the compensated estimate of the power spectrum and is defined as [WELCH, 1967]:

$$\{P_x(l)\} = \{{}^N/_U\ |X_w(l)|^2\}, \quad l=1, \dots, N$$

$$U = {}^1/_N \sum_{l=1}^{N} W^2(l) \qquad (3.9)$$

where U is the compensating factor, and $W(l)$ the D.F.T. coefficients of the window sequence $w(k)$. The modified periodogram approach, apart from being faster due to the speed of the F.F.T., has one more advantage over the mean-lagged product approach. The estimate of the autocorrelation function suffered from leakage in the same way that the raw periodogram did. Windowing for leakage reduction dates from this era but it did not wholly solve the problem: As the estimate of the power spectrum was calculated by multiplying the window with the autocorrelation estimate and transforming (see relation 3.5) the negative side-lobes of the window resulted in negative coefficients i.e. negative power. The modified periodogram avoids this by definition since it involves squaring of the coefficients (relation 3.9).

An extensive review and comparative assessment of most classic and modern windows that have been presented in the literature is given by Harris [1978]. Harris defined several "figures of merit" concerning the equivalent noise bandwidth, processing gain, overlap correlation etc. He concludes that in so far as harmonic energy detection is concerned, any window except the rectangular is nearly as good as any other. On the contrary, in discriminating very closely spaced frequency components not all windows are good performers. Harris recommends the use of modern windows which are more sophisticated such as the Kaiser-Bessel or the Blackman-Harris windows which he considers top performers. Very recently these windows were slightly modified and reassessed by Nuttall [1981] who also corrected some of the inaccuracies of Harris's plots. Tseng [1981] also presented a new 'parametric' window whose parameters can be adjusted to obtain optimal performance.

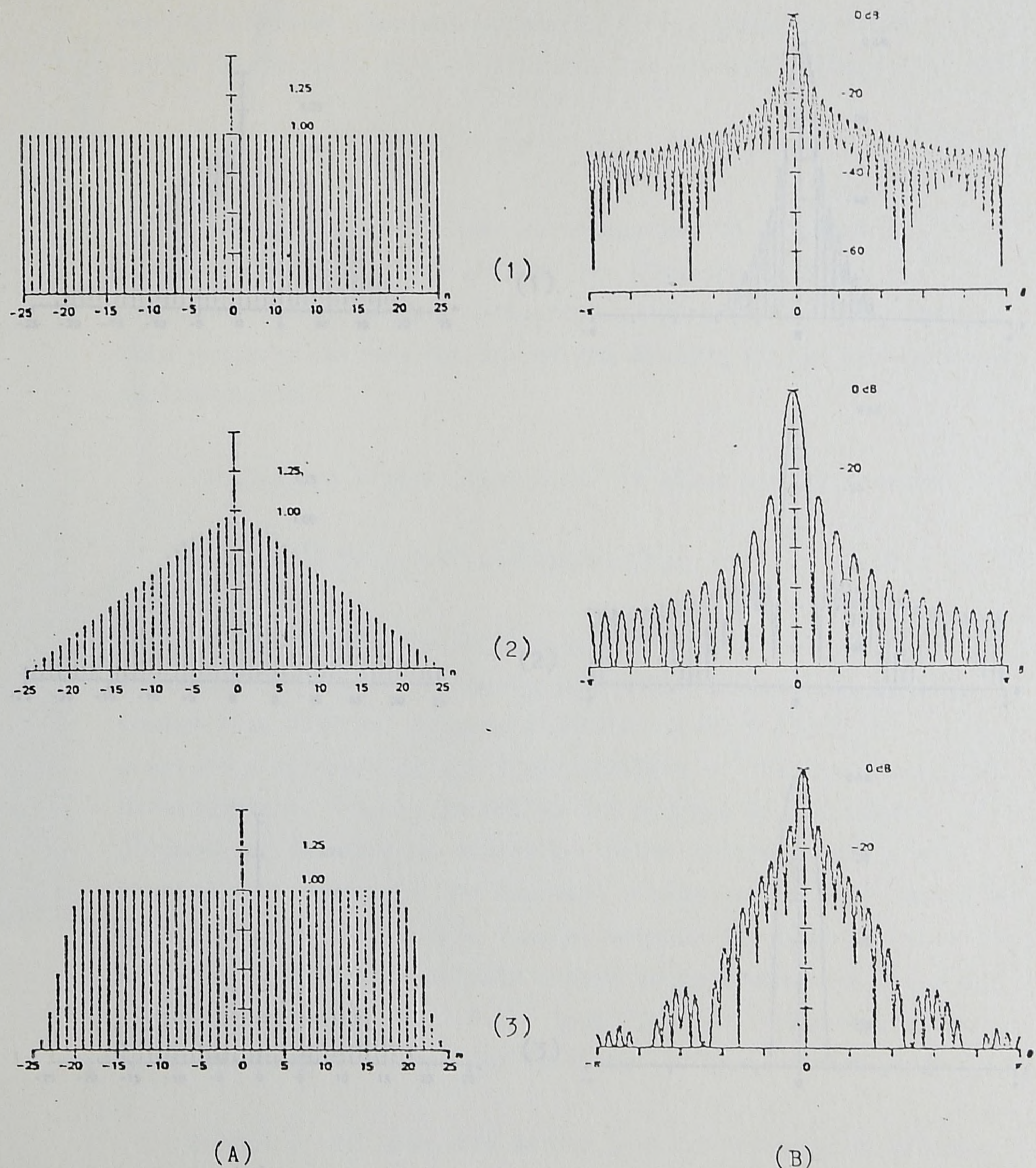
Nevertheless, in the literature of acoustic signal processing, even in recent papers, the windows mostly used are the classic ones: the triangular or Bartlett window, the Hanning and the Hamming windows (figs. 3.9) which were known and used prior to the F.F.T. era [BLACKMAN & TUKEY, 1959], probably because of familiarity and ease of implementation. Also because very closely spaced frequencies do not occur in speech and other acoustic signals.

The <u>triangular</u> or <u>Bartlett</u> window (fig. 3.9b) was selected by Thomas [1971] in his study of vehicular sounds, a work closely related to the work presented in this thesis. This window, although it has the advantage of nonnegative energy it has poor sidelobe characteristics (the highest sidelobe is only 27dB down). Thomas used autocorrelation estimates in the first part of his study to calculate power, hence the advantage of non-negative energy was critical.

The <u>Hanning</u> window (fig. 3.9c) is given by the formula:

$$\{w(k)\} = \{0.5 - 0.5 \cos[2\pi(k-1)/N]\} \quad , k = 1, \ldots ,N \qquad (3.10)$$

This window has a very interesting property, in addition to having a -32dB sidelobe level: Its spectrum (and hence each



<u>Figure 3.9.a</u> Leakage reducing windows
1) Rectangular 2) Bartlett 3) Tukey (25%)
A) Window B) Transform
(From Harris [1977])

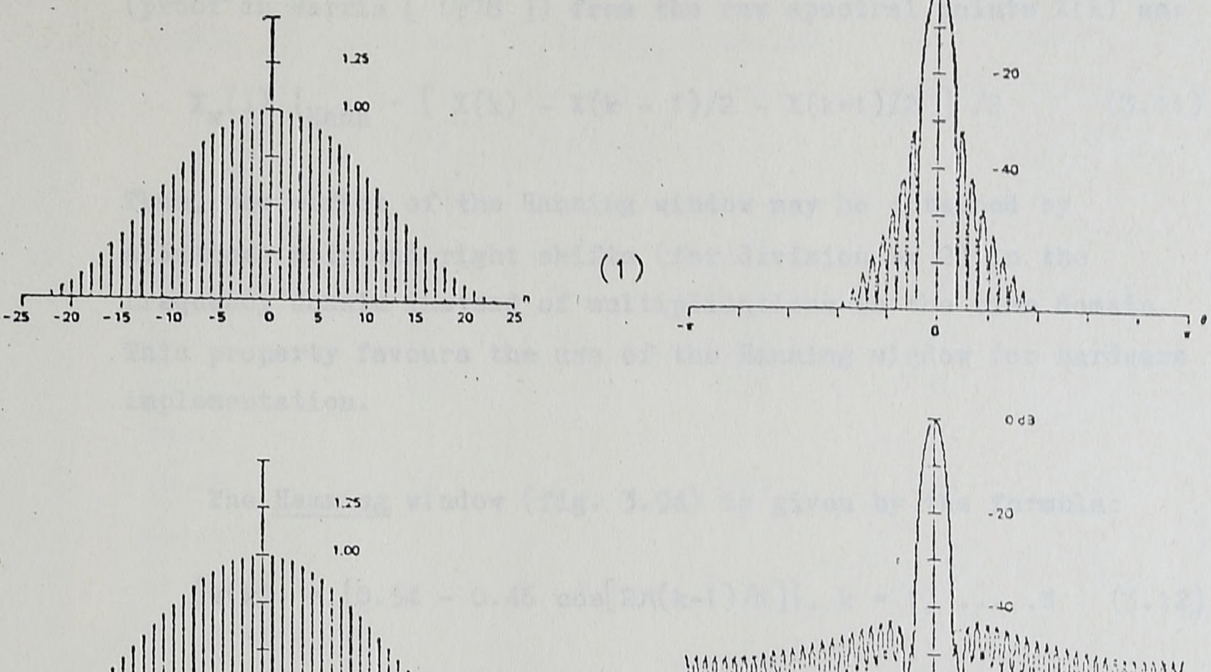


Figure 3.9.b Leakage reducing windows

1) Hanning 2) Hamming 3) Blackman-Harris (4 point)

A) Window B) Transform

(From Harris [1977])

component of the spectrum of a windowed signal) is nonzero at only three frequencies and the sample values are binary fractions, which can be implemented as right shifts. Thus, the Hanning-windowed spectral points $X_w(1)\,|_{Hann}$ may be found (proof in Harris [1978]) from the raw spectral points X(k) as:

$$X_w(1)\,|_{Hann} = [\ X(k) - X(k-1)/2 - X(k+1)/2\ ]\ /2 \qquad (3.11)$$

Thus, the effect of the Hanning window may be obtained by additions only and right shifts (for division by 2) in the frequency domain instead of multiplications in the time domain. This property favours the use of the Hanning window for hardware implementation.

The <u>Hamming</u> window (fig. 3.9d) is given by the formula:

$$\{w(k)\} = \{0.54 - 0.46\cos[2\pi(k-1)/N]\},\ k = 1, \ldots, N \qquad (3.12)$$

The similarity with the Hanning window is obvious. The adjustment of the coefficients gives this window (out of all cosine-type windows) minimum sidelobe levels (-43dB) by completely suppressing the first sidelobe of the Hanning window. Unfortunately, this is gained at the expense of the useful property of the Hanning window mentioned before. Still, the Hamming (and of course the Hanning) window may be implemented by multiplication in the time domain, without the need to store its samples, as these are already stored as trigonometric table for the computation of the F.F.T. Thus, practical implementation considerations force the use of either of these two windows.

Thomas [1971] in the second part of his work on vehicle sounds mentioned earlier alternatively used the rectangular, the Hanning and the Tukey window (mentioned below) in his spectral analysis which was a part of an experiment to measure the firing rate of combustion engines. The performance of the rectangular window was the worst, whereas the Hanning window the best.

Table 3.1

Window Characteristics

Window	Highest Sidelobe (dB)	Sidelobe Fall-off (dB/Oct)	Coherent Gain	6dB B/width	50% Overlap Correlation (%)
Rectangle	-13	- 6	1.00	1.21	50.0
Triangle	-27	-12	0.50	1.78	25.0
Hanning	-32	-18	0.50	2.00	16.7
Hamming	-43	- 6	0.54	1.81	23.5
Tukey (0.25)	-14	-18	0.88	1.38	44.4
Kaiser-Bessel (4-point)	-69	- 6	0.40	2.44	7.4
Blackman-Harris (4-point)	-74	- 6	0.40	2.44	7.4

Table 3.1 gives a summary of the properties of the windows mentioned and the two top performers preferred by Harris. The Tukey or cosine-tapered window is also included in the table, because of its popularity due to its high processing gain. This window is a combination of the rectangular and cosine bell windows: a percentage (25% in the table) of the data, at either end of the window, is weighted with a cosine bell. The coherent gain in the table is the normalised function U of relation 3.9 and the 6dB column refers to the bandwidth measured in units of fundamental frequency resolution ($\Delta f = f_s/N$) where the main lobe falls by 6dB.

Further considerations on the practical use of windows are presented in chapter 6 on spectrogram analysis. The next section deals with statistical consequences of windowing.

3.1.2.3 Statistics of Spectra

In the previous section, the windows were treated as means for leakage reduction. However, windows were initially used for the improvement of the statistics of spectra i.e. the means and variances of the estimated spectra.

It can be proved [SCHWARTZ & SHAW, 1975] that there exist a relation between the raw periodogram P_x of (3.7) and the true autocorrelation R_s, of the stochastic process s.

$$\{E[P_x(l)]\} = \{ \sum_{k=-N}^{N} R_s(k) \ (1-|k|/N) \ \exp(-2j\pi kl/N)\} \ , \ l=1, \ \dots \ ,N \qquad (3.13)$$

where $E[.]$ denotes expected value. The periodogram is computed from a sample $\{x(k)\}$, k=1, ... ,N of the process s. It may be seen that the estimate is biased for its mean value is not equal the real spectrum. However, it asymptotically approaches the real for $N \rightarrow \infty$. The term $1-|k|/N$ is equivalent to a triangular window weighting of the samples and is the result of the convolution of the rectangular data window with itself. (This shows why windows played an important part in the autocorrelation approach to the estimation of power spectra.)

It may be shown that [JENKINS & WATTS, 1968]:

$$\lim_{N \rightarrow \infty} \operatorname{var} \{P_x(l)\} = \{[S_s(l)]^2\} \ , \ l=1, \ \dots \ ,N \qquad (3.14)$$

where $S_s(l)$ is the true power spectrum of the process, regardless of the size of the sample of the process. This means that the <u>mean standard error</u> defined as:

$$e_r = \sqrt{\operatorname{var} [P_x(l)]} \ / \ S_s(l) \approx 1 \qquad (3.15)$$

is approximately one, or that the error at a particular frequency is as large as the power. This annoying property of the periodogram can be circumvented. One way is to average together several periodograms [WELCH, 1967]. The average estimate $\{\hat{P}_x(l)\}$ will be:

$$\{\hat{P}_x(l)\} = \{{}^1/_M \, P_x(l)\} \;, \quad l=1, \ldots ,N \tag{3.16}$$

where M is the number of periodograms averaged together. Provided that the periodograms are statistically independent the variance of this estimate will be:

$$\mathrm{var}\ \hat{P}_x(l) = \mathrm{var}\ P_x(l) \,/\, M \;\approx\; [S_s(l)]^2 \,/\, M \tag{3.17}$$

and a mean standard error:

$$e_r = 1 \,/\, \sqrt{M} \tag{3.18}$$

For a given number of periodograms the variance reduction that is actually achieved is less than what is given by the above formula since there is always a degree of dependence between consecutive periodograms unless the signal is gaussian white noise. The variance may be decreased by increasing the number of periodograms. A fixed sample size of L points may be divided into M segments of size N. It is obvious that M may be increased and hence the reduction of variance by decreasing N. Decreasing N, however, means degradation of the minimum spectral resolution of the estimate ($\Delta f = f_s/N$). In such a case the spectral estimates will have less variability but indeterminate frequencies. This uncertainty is more important where the spectrum has peaks and valleys and may be aleviated by <u>pre-whitening</u> the signal i.e. by preconditioning so that its spectrum becomes smooth.

The conclusion is that, for a given frequency resolution, the number of periodograms which is proportional to the duration of the sample record imposes a lower bound to the variance of the spectral estimates. However, for a given record the number of periodograms may be increased by taking <u>overlapping</u> segments. Overlap will also compensate for the possible loss of information at the two ends of the window where it exhibits small values. The resulting periodograms will have increased correlation and hence the overall reduction in variance will be less than that obtained by averaging uncorrelated periodograms, but still it will be better. Harris [1977] gives a formula derived from Welch [1967] relating the reduction of variance to the degree of overlap. The formula below gives the mean standard error for the special case of 50% overlap being used in most cases:

$$e_r^2 = {}^1/_M \left[1+2c^2(0.5)\right] - {}^2/_{M^2} \left[c^2(0.5)\right] \qquad (3.19)$$

M is the number of consecutive overlapping periodograms. The parameter c(0.5) is the correlation coefficient for 50% overlap. When modified periodograms are used c depends on the type of the window. Table 3.1 gives c(0.5) for various window types including the rectangular window. It may be noted that the latter has the highest correlation coefficient and hence more periodograms are needed to achieve the desired variance than by using any other window. This is an advantage of windowing in addition to that of reducing leakage. More, the reduction of leakage helps in the reduction of the bias of the estimate as well: lower sidelobes do not influence the magnitude of neighbouring estimates.

The reduction of leakage by windowing has an additional effect on the evaluation of spectra: that of <u>smoothing</u>. For example, as it was only hinted in the previous section, the effect of Hanning windowing is equivalent to substituting each spectral coefficient with an average of neighbouring ones. This, on one hand, results in reduction of the variance of spectral estimates in a way similar to that of averaging periodograms. In fact averaging M periodograms reduces variance by the same degree as averaging M consecutive spectral coefficients. On the other hand however, smoothing results in reduction of spectral resolution. The reduction may also be seen by the fact that the main lobes of all leakage-reducing windows are wider than the main lobe of the rectangular window. Table 3.1 gives the 6dB bandwidth of the spectral estimates for various window types as multiple of the fundamental spectral resolution Δf. The rectangular window has the smallest main lobe bandwidth, whereas the top ones have the widest.

To summarize, periodogram analysis involves the segmentation of a data record into overlapping segments. These segments are weighted using one of alternative window functions in order to reduce leakage and the variance of the spectral estimates. Each weighted segment is Fourier transformed using an F.F.T. algorithm. The periodogram is evaluated as given by formula (3.9) and the estimate of the spectrum by formula (3.16). The mean standard error of the spectral estimates is given by formula (3.19) using the corresponding correlation coefficient from table 3.1.

3.1.3 Cepstral Analysis

Cepstral analysis is suited to the analysis of data that contain echoes of a fundamental waveform. Historically, the cepstrum has its roots in the general problem of deconvolution of two or more signals. The power cepstrum was introduced by Bogert *et al* [1963] as a heuristic technique for finding echo arrival times in a composite signal. The authors showed that the effect of echo will manifest itself as a ripple in the log spectrum and hence the ripples will correspond to some 'frequency' peak in the 'spectrum' of the spectrum. Analytical explanation of this phenomenon will be given in the next chapter where reflections of environmental sounds are treated. Bogert and his colleagues defined the power cepstrum as the square of the Fourier transform of the logarithm of the power spectrum. They called the 'frequencies' in the cepstrum quefrencies having units of time.

Another broad category of problems where spectral ripples occur is when a periodic source excites a resonant system, as in the case of speech generation. The ripples in the speech spectrum are due to the harmonics of the fundamental frequency of the source [NOLL, 1967]. If $S(f)$ is the spectrum of the source, $H(f)$ is the transfer function of the system and $F(f)$ the output spectrum, then:

$$F(f)^2 = S(f)^2 \, H(f)^2 \tag{3.20}$$

By taking the logarithm of both sides of (3.13) the multiplication becomes an addition:

$$\log|F(f)|^2 = \log|S(f)|^2 + \log|H(f)|^2 \tag{3.21}$$

To obtain the cepstrum, we take the Fourier transform of the logarithms of the spectra. Because of the linearity of the transform the addition is preserved and the effect of the system on the output can be separated from the excitation, provided that their contributions yield components in different quefrency ranges, as shown in figure 3.9.

From a different point of view [COHEN et al, 1970] the effect of the logarithm is seen not as a separator of the convolved signals (or signal and impulse response) but as a spectral whitener. Cohen and his colleagues claimed that the logarithm is too severe a whitener for several applications and recommend other functions such as the square root of the power spectrum which in fact means direct application of the transform on the amplitude spectrum.

A relatively recent review of cepstral analysis is given by Childers et al [1977]. Childers and his colleagues define the cepstrum in terms of the z-transform because the power cepstrum does not exist for most signals; it is meaningful only when defined in a sampled data sense. The authors showed the estimation capability of echo delay time of the power cepstrum, starting from the convolution of two sequences whose spectra in the z-domain multiply to give the

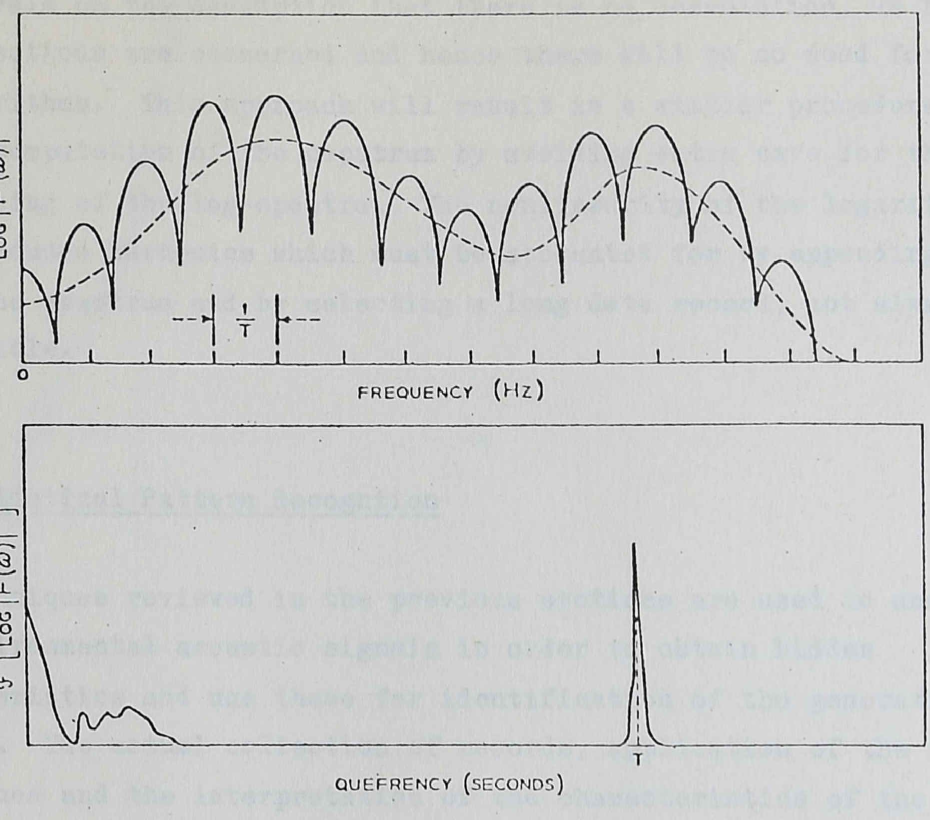


Figure 3.10. Spectrum and Cepstrum of speech signal.
(dashed line indicates vocal tract contribution)
(From Noll [1967])

spectrum of the output signal. By assuming that one of the input sequences is a composite one consisting of a simple 'wavelet' (an impulse for simplicity) and a delayed echo they go on to prove that the cepstrum of the output signal is rippled. Their model which involves convolution naturally leads them to taking the logarithm for deconvolution. It is of course natural to model the phenomenon of echo in its most general form, i.e. by taking into account the transmitting system. However, as Cohen [1970] pointed out there are cases where the logarithm is undesirable. It seems that this happens when the effect of the transmitting system is weak as is the case of simple reflections. In the next chapter we shall deal with composite signals resulting from ground reflection of the sound signals emitted by noise pollution sources. We shall base the analysis on the assumption that there is no convolution, as far as reflections are concerned and hence there will be no need for logarithms. This approach will result in a simpler procedure for the computation of the cepstrum by avoiding extra care for the aliasing of the log-spectra. The nonlinearity of the logarithm introduces harmonics which must be accounted for by appending zeros to the spectrum and by selecting a long data record, not always feasible.

3.2 Statistical Pattern Recogntion

The techniques reviewed in the previous sections are used to analyse the environmental acoustic signals in order to obtain hidden characteristics and use these for identification of the generating sources. The actual collection of records, application of the analysis techniques and the interpretation of the characteristics of the spectra and cepstra in terms of their relationship to the nature of the sources is the subject of next chapters.

Here we shall assume that meaningful measurements or features are available from a large set of records from all classes of interest. A noise record or pattern may be represented by a feature vector $\underline{X} = [X_1, \ldots, X_d]$, as shown in fig. 3.11. The pattern may belong to any class of the set of classes $\{c_1, \ldots, c_{n_c}\}$. Each class is governed by a class conditional probability density function (pdf) $p(\underline{X}/c_j)$, i.e. the probability that a pattern drawn from class c_j will be represented by a vector in the vicinity of $\underline{X}$. The law of total probability states that:

$$p(\underline{X}) = \sum_{j=1}^{n_c} p(\underline{X}/c_j)\, P(c_j) \tag{3.22}$$

where $P(c_j)$ is the a priori probability that the object will be drawn from class c_j.

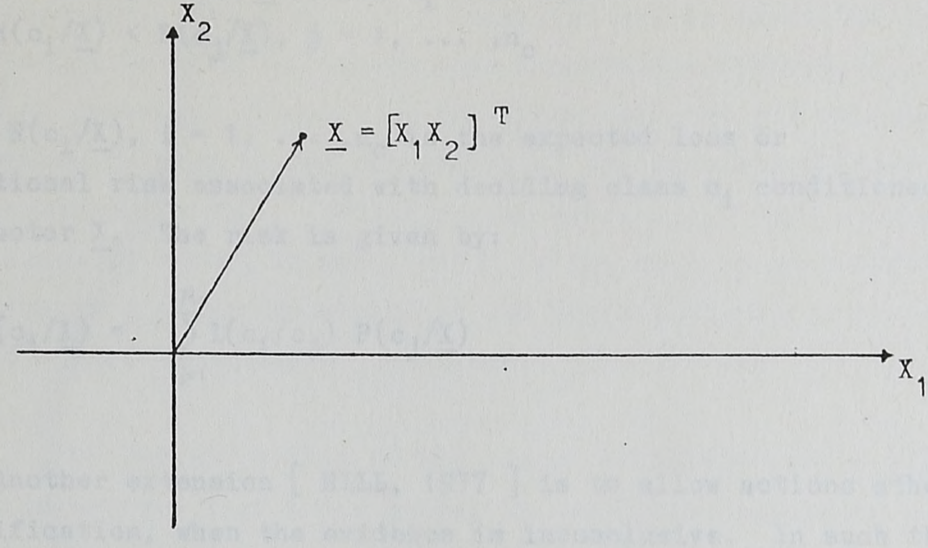


Figure 3.11. Feature Space

Bayes rule gives the a posteriori conditional probability $P(c_j/\underline{X})$ that pattern $\underline{X}$ was drawn from class c_j.

$$P(c_j/\underline{X}) = p(\underline{X}/c_j)\ P(c_j)\ /\ p(\underline{X}) \tag{3.23}$$

The set of a posteriori probabilities $P(c_i/\underline{X})$, $i = 1, \ldots, n_c$ defines a minimum error rate decision rule [DUDA & HART, 1973]:

Given a pattern $\underline{X}$ decide class c_i such that:
$P(c_i/\underline{X}) \geqslant P(c_j/\underline{X})$, $j = 1, \ldots, n_c$

This decision rule treats the errors of misclassification as incurring the same loss. This can be circumvented with the introduction of a loss function $l(c_i/c_j)$, $i, j = 1, \ldots, n_c$. The function is a measure of the loss incurred by deciding class c_i when c_j is the true class.

It can be shown [DUDA & HART, 1973] that the decision rule becomes:

Given a pattern $\underline{X}$ decide c_i such that:
$R(c_i/\underline{X}) < R(c_j/\underline{X})$, $j = 1, \dots, n_c$

where $R(c_i/\underline{X})$, $i = 1, \dots, n_c$ is the expected loss or conditional risk associated with deciding class c_i conditioned upon the vector $\underline{X}$. The risk is given by:

$$R(c_i/\underline{X}) = \sum_{j=1}^{n_c} l(c_i/c_j)\, P(c_j/\underline{X}) \qquad (3.24)$$

Another extension [HILL, 1977] is to allow actions other than classification, when the evidence is inconclusive. In such the case a repertoir of actions d_i, $i = 1, \dots, n_d$ such that $n_d > n_c$ and $d_i = c_i$, $i = 1, \dots, n_c$, is defined.
Then the decision rule becomes:

Given $\underline{X}$ decide d_i such that:
$R(d_i/\underline{X}) < R(d_j/\underline{X})$, $j = 1, \dots, n_d$

The definition of the conditional risk is modified accordingly:

$$R(d_i/\underline{X}) = \sum_{j=1}^{n_c} l(d_i/c_j)\, P(c_j/\underline{X}) \qquad (3.25)$$

3.2.1 Estimation of Probability Densities

Following the extended decision rule the classification problem may be defined in terms of the loss function $l(d_i/d_j)$ which is directly related to the nature of the specific application; and in terms of the a posteriori probability $P(c_j/\underline{X})$. The latter is given by equation (3.23) where it is expressed in terms of the class probabilities $P(c_i)$, the total probability $p(\underline{X})$ and the class conditional pdf's $p(\underline{X}/c_j)$. The first can be estimated as relative frequencies from the design set, if this is representative, or by the designer. The total probability $p(\underline{X})$ appears in both sides of the inequality of the decision rule and thus it can be neglected. The class conditional pdf's must be estimated from the design set. The probability densities may be estimated by either assuming that the variables have a certain distribution (e.g. gaussian) or not. In the first case the pdf estimation problem reduces to the estimation of the parameters (mean vector and covariance matrix). In the second case, i.e. when such an assumption cannot be made, which is the case in most applications where no a priori knowledge of the probability densities exists, the pdf's must be estimated otherwise.

For the estimation of the class conditional pdf's $p(\underline{X}/c_j)$ it is assumed that a design set is available containing samples from each class of interest. The problem may be split into separate estimations of the pdf's of each class. The basic idea of Parzen estimation [HILL, 1977] is that each sample (say the i^{th} sample of class j represented by the feature vector $\underline{X}_{ij}$) contributes to the estimate of the pdf throughout the space. The contribution is specified by a window function $\gamma(\underline{X},\underline{X}_{ij})$. The pdf estimate at any point $\underline{X}$ is formed by averaging over the complete set of window functions:

$$\hat{P}(\underline{X}/c_j) = {}^1/_{N_j} \sum_{i=1}^{N_j} \gamma(\underline{X},\underline{X}_{ij}) \tag{3.26}$$

where $\hat{p}(\underline{X}/c_j)$ denotes the estimate of $p(\underline{X}/c_j)$ and N_j is the number of samples of class c_j. Several window functions may be used, the most common being the multivariate spherical gaussian distribution. The standard deviation of the distribution determines the "peakiness" of the estimated pdf and is therefore called the smoothing parameter. The theory of Parzen estimation unfortunately does not offer any guidance for the selection of a proper smoothing parameter. The requirements for the realization of a classifier based on Parzen estimation is not practical for real time applications since it requires storage of the complete design set and substantial computation to reach a decision.

3.2.2 Linear Classifiers

Another family of classifiers also requiring the complete storage of the design set, as in the case of Parzen estimation mentioned in the previous section, is the family of the nearest neighbour classifiers. The decision rule in this case is to assign the sample represented by the feature vector $\underline{X}$ the label associated with its nearest labelled sample form the design set. A simplified form of this classifier is the minimum distance classifier where the design set is reduced to a set of prototype vectors, one for each class. The error rate of this family is always greater than the minimum possible, the Bayes rate.

The classifiers mentioned above are often called piecewise linear [FU, 1976] belonging to the family of linear classifiers. Linear classifiers are best described with the introduction of discriminant functions. A set of discriminant functions

$$g_i(\underline{X}), \; i = 1, \; \ldots \; , n_d$$

divides the feature space into n_d mutually exclusive regions. The regions R_j are defined such that:

$$g_j(\underline{X}) = \max g_i(\underline{X})$$
for all $\underline{X}$ in the j^{th} region R_j.

In a linear classifier the discriminant functions are linear, as shown in fig. 3.12, one for each class:

$$g_i(\underline{X}) = \underline{w}_i^T \, \underline{X} + w_{i0} \quad , \; i=1,\ldots,n_c \tag{3.27}$$

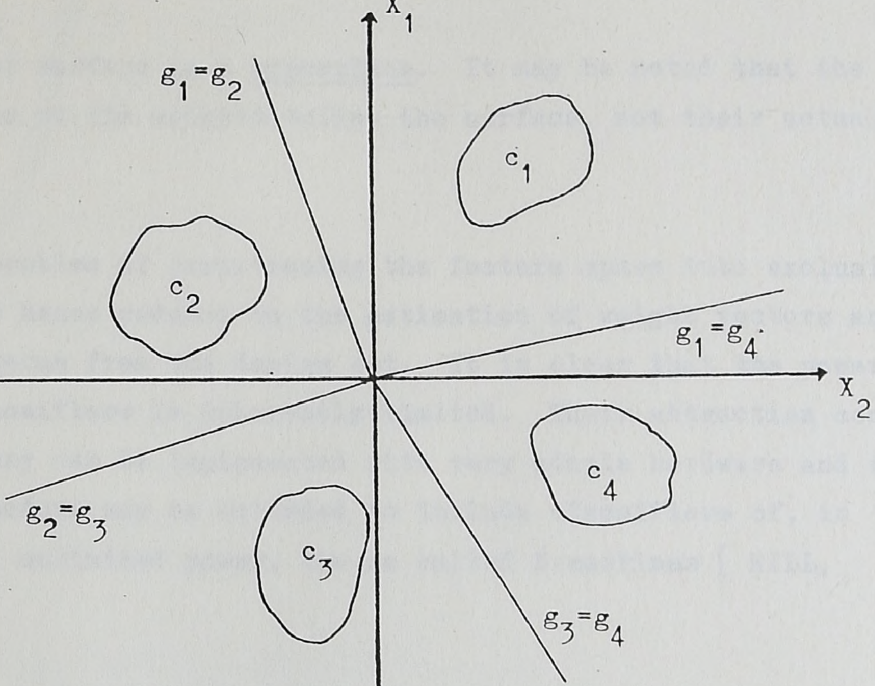


Figure 3.12. Linear Classifier

where

$$\underline{w}_i = [\ w_{i1},\ \cdots\ ,w_{id}\ ]^T$$

is a weight vector for class c_i.

The decision boundary between two contiguous regions R_i, R_j is defined by:

$$g_i(\underline{X}) = g_j(\underline{X})$$

and for a linear classifier this becomes:

$$(\underline{w}_i-\underline{w}_j)^T \underline{X} + (w_{i0}-w_{j0}) = 0 \qquad (3.28)$$

This linear surface is a <u>hyperplane</u>. It may be noted that the differences of the weights define the surface, not their actual values.

The problem of partitioning the feature space into exclusive regions is hence reduced to the estimation of weight vectors and constant terms from the design set. It is clear that the power of linear classifiers is inherently limited. Their attraction consists in that they can be implemented with very simple hardware and that their structure may be extended to include classifiers of, in principle, unlimited power, the so called Ø-machines [HILL, 1977].

The set of weight vectors can be estimated by using either of two methods. The <u>error correcting</u> procedures begin with an arbitrary set of weights, present the design samples in sequence for classification and modify the weight vectors when a design sample is misclassified by them. If a set of weights exists which will correctly classify the samples, then procedures exist which will converge to a solution in a finite number of corrections. Then the samples are said to be <u>linearly separable</u> [DUDA & HART, 1973]. If not, the procedure will not converge. Since it is unlikely that a design set is linearly separable in our case, these procedures are not favourable. In addition, they terminate with any solution which separates the classes, not necessarily the optimal. Their iterative nature on the other hand requires computations which must be performed on a large computer.

The <u>least mean square error</u> (LMSE) procedures offer a good compromise on both separable and nonseparable cases, but do not guarantee separation even when this is possible. In addition they closely relate to the Bayes's optimal discriminants in a least mean square sense. An introduction follows based on a review by Hill [1977].

Let us define an augmented feature vector

$$\underline{Y} = [\ 1,\ \underline{X}\ ]^T = [\ 1,\ X_1, \ldots, X_d\ ]^T$$

and similarly weight vectors

$$\underline{W}_i = [\ w_{i0},\ \underline{w}_i\ ]^T = [\ w_{i0},\ w_{i1}, \ldots, w_{id}\ ]^T$$

Then equation (3.27) may be rewritten in a homogeneous form:

$$g_i(\underline{X}) = \underline{W}_i^T\ \underline{Y}\ ,\ i=1,\ldots,n_c \tag{3.29}$$

Let $\{\ Y_{ij},\ i = 1,\ldots,N_j,\ j = 1,\ldots,n_c \}$ be the design set so that Y_{ij} is the i^{th} (augmented) vector of the j^{th} class. The complete design set may be represented as a matrix A of dimension (Nxd*), where d*=d+1 and N is the total number of samples in the set:

$$A = [\ Y_1\ Y_2\ Y_3\ \ldots\ ]^T$$

where Y_j is a matrix of dimension ($N_j$xd*) containing all vectors of j^{th} class:

$$Y_j = [\underline{Y}_{1j}^T\ \ldots\ \underline{Y}_{N_j j}^T]^T,\ j=1,\ldots,n_c$$

Let a matrix B be defined of dimensions (Nxn_c):

$$B = [\ B_1\ B_2\ B_3\ \ldots\ ]^T$$

where B_j is a matrix of dimensions $(N_j x n_c)$ consisting of N_j identical rows:

$$B_j = [\ \underline{b}_j^T\ \underline{b}_j^T\ \ldots\ ]^T,\ j = 1,\ldots,n_c$$

where:

$$\underline{b}_j = [\ b_{j1},\ \ldots\ ,b_{jn_c}]^T$$

Similarly the complete set of weight vectors may be represented as a matrix of dimension $(d^* x n_c)$:

$$W = [\ \underline{W}_1,\ \underline{W}_2,\ \underline{W}_3,\ \ldots\ ]$$

where $\underline{W}_j$ is the weight vector for the j^{th} class:

$$\underline{W}_j = [w_{j0},\ w_{j1},\ \ldots\ ,w_{jd}\ ]^T,\ j=1,\ldots,n_c$$

The solution to the problem is to determine the weight matrix such that:

$$AW = B$$

with minimum square error. In other words to minimize $||\ AW-B\ ||$ where $||\ C\ || = [\ \text{trace}\ C^TC\ ]^{1/2}$ = square root of the sum of all squared elements of C. The solution is given [WEE, 1968]:

$$W = (A^TA)^{-1}A^TB \qquad (3.30)$$

if $(A^TA)^{-1}$ exists. The discriminant functions this solution yields are:

$$g_i(\underline{X}) = \underline{W}_i^T\ \underline{Y},\ i=1,\ldots n_c \qquad (3.31)$$

If the vectors $\underline{b}_j$, $j = 1,\ldots,n_c$ are selected such that:

$$b_{jk} = -l(c_k/c_j),\ j,k = 1,\ldots,n_c$$

where $l(c_k/c_j)$ is the loss incurred by deciding class c_k when c_j is the true class, it can be shown [WEE, 1968] that the solution of (3.30) yields discriminant functions of equation (3.31) which are the minimum square error approximates to the Bayes optimal discriminants. The latter are given by:

$$g_i(\underline{X}) = -R(c_i/\underline{X})$$

$$= -\sum_{j=1}^{n_c} l(c_i/c_j)\ P(c_j/\underline{X})$$

The approximation is achieved regardless of the form of the class conditional pdf's, as the number of design samples tends to infinity. This procedure, therefore, incorporates the various error penalties assigned by the designer and also has a theoretical justification. Moreover, the solution may be obtained with acceptable computational demands and has been extended to account for indecision [HILL, 1977].

3.2.3 Comparison of Classifier Types

The effectiveness of the two forms of classifiers, namely the Parzen estimation and linear discriminants, in the recognition of defects on the surface of steel laminates has been studied by Hill [1977]. Hill rejected the use of an error correction procedure for the estimation of the weight vectors in favour of the LMSE procedure and compared the latter to two forms of Parzen estimation: a) using a spherical gaussian distribution for window function. This form he names "the exponential form" of Specht's classifier and b) the "polynomial form" of Specht's classifier [SPECHT, 1966], where the gaussian multivariate pdf is expanded via a Taylor series. Hill's expectation that Specht's classifier would be best suited to complex class conditional pdf's (which were anticipated due to the nature of his data) was not fulfilled. Specht's classifier in its polynomial form (which is a requirement for its implementation) showed a significant loss of performance relative to the exponential form, unless the smoothing parameter had a very large value. In that case, however, the polynomial form could easily be reduced to first order, without loss of performance. On the other hand, the performance of the LMSE classifier was as good as the best performance obtained from Specht's classifier. As far as normalization of features is concerned, normalization to zero means was a necessary requirement for the polynomial form of Specht's classifier. The Taylor series used to derive the polynomials are expansions about the origin of the space and their approximation accuracy is therefore highest close to the origin. On the contrary, the LMSE classifier avoids the need of normalization of the vectors. This classifier yields optimum transformation from a feature space to a decision space [HILL, 1977] which, in essence, automatically incorporates any normalization required. For this classifier, the only purpose of normalization is prevention of overflow/underflow problems during computation.

3.2.4 Feature Selection and Evaluation

The designer of a pattern recognition system is usually faced with a large number of candidate features. Selection of the 'best' subset of features is of major importance in the design process for two reasons: Firstly, a properly selected feature set may reduce the dimensionality of the feature space, resulting in savings in computation and hardware. Secondly, a properly selected subset may lead to better classification. This happens because the inclusion of a feature may be worse than merely useless; it may increase the confusion between classes.

The feature selection problem may be seen as a combinatorial optimization problem requiring a) a selection criterion and b) a search procedure. The criterion functions used are quite diverse because of the difficulty in making any assumptions about the class conditional distributions. The most common assumption is the gaussian distribution, which cannot be made about the distributions of spectra of noise pollution sounds. But even when the distributions are known, a straightforward analytical relationship between Bayes error rate and the features used is not available (unless the distributions are simple e.g. gaussian). For that reason, various measures of information and distance have been proposed to measure the effectiveness of a given set of features (probability of misclassification, average information content, divergence etc.). Kanal [1974] gives a summary of these measures concluding that the best is to try and estimate the error probability itself in a direct way. Hill [1977] suggests that the criterion should be the actual performance of the particular classifier designed on the selected feature subset.

The evaluation of performance is a problem of estimation of the behaviour of the classifier on new data i.e. data which were not part of the design set. Inevitably, the estimate will always involve some bias and the method to be used must ensure minimum bias. Intuitively, the risk is minimized when evaluation is done on data other than the design set. This means that the available data set must be divided into two subsets: one for training and one for testing. This method, though, requires a very large data set which is usually very expensive and labourious to acquire in practice. The cheapest alternative of course is the evaluation on the design set itself (the entire data set), but this gives a very optimistic bias. A compromise between the two extremes is the <u>leave-one-out</u> method which was recommended by Lackenbruch and Mickey [1968] in their comparative study of various alternative methods. In this method, given N samples, a classifier is designed on N-1 samples, tested on the remaining sample and then the result of all such partitions are averaged. This method takes N times the computation taken when partitioning into separate training and testing sets. It is obvious that such a method which is computationally expensive is combined best with a classifier which is easy and quick to design. This is a further incentive to use the LMSE procedure since it is possible to modify a design based on the complete data set so as to produce a design based on the complete set less one member, without repeating the entire design process [HILL, 1977].

Having selected the performance criterion and hence the evaluation criterion for the selection of a feature subset, the only way to ensure that the best subset of K features is chosen is to explore all $\binom{N}{K}$ possible combinations. The 'curse' arises from the fact that the set of K individually best features is not necessarily the best discriminating feature set of size K [KANAL, 1974]. Since an exhaustive search is not usually feasible, various suboptimal search procedures are in use. The most common is the so called <u>forward sequential</u> or <u>without replacement</u> selection procedure in which the best individual features are chosen on the first round and then the best pair including the best individual feature and so on. The <u>sequential rejection</u> or <u>backward sequential</u> procedure is the counterpart of the previous one. Here the n-1 best features are selected by rejecting the worst one, etc. The obvious disadvantage of the latter is that selection starts at a high-dimensional space when only a few features may be needed, whereas in the former method the search stops as soon as performance reaches a certain predefined threshold, or when a significant improvement is not obtained by adding more features. The major argument against both is that once a feature is selected (rejected) to extend (reduce) the feature subset, it will (not) be a member of all 'next' subsets. Hill [1977] suggests a combination of both which will somehow overcome the restrictions of either: additional features can be selected by "two steps forward and one step backward". Of course, any improvement would be obtained at the expense of increased computation.

Another consideration relevant to feature selection is the dimensionality of the feature set and its relation to the size of the training set. Experimentally, it has often been observed, given a finite training set, that as the number of features increases, the performance of the classifier first improves rapidly, later reaches a peak and finally falls off. This gives a guideline for the selection of the proper number of features. However, what is obtained by the experiment is an estimate of the actual performance of the classifier on unseen data. The reliability of the performance estimate depends on the size of the training set. The larger the ratio of its size to the dimensionality of the feature set, the better is the estimate. Furthermore, a sufficiently large number of samples per class is required in order to ensure low variance of the estimate. Again, conclusive statements can only be derived for features of known statistics. Seemingly, reasonable results are obtained when the above ratio (training set size to feature set dimension) is greater than 10 for each class [KANAL, 1974].

3.3 Summary, Related Work and Discussion

In the previous sections we mainly reviewed the underlying theories and techniques related to the analysis and recognition of signals. The literature sources for this review can be classified into the following two broad categories:

a) Theory and practice of signal analysis, particularly spectral analysis. This category is fairly diverse and encompasses a very broad spectrum of sources, both theoretical and applied, the main characteristic of which is rigour and detail. Section 3.1 was mostly based on literature from this category.

b) The second category includes the broad spectrum of papers on pattern recognition, with varying degrees of generality, detail and rigour. Section 3.2 of this chapter was a review of the most important classification, feature selection and evaluation techniques.

The assessment of the applicability of these theories and techniques to the analysis and recognition of noise pollution sounds, and hence the given emphasis, was based on personal judgement aided by conclusions drawn from the study of literature concerned with applications in related fields. This literature may be classified into two categories, additional to the ones mentioned above:

c) The third category, seemingly very relevant to the work reported in this thesis, includes papers on analysis and recognition of acoustic signals, excluding speech. These papers solely refer to spectral analysis with various degrees of involvement. They range from analysis of musical tones and transients [KEELER, 1972] to the short-time spectral analysis of the calls of whales [SINGLETON & POULTER, 1967]. One paper [MALING et al, 1967] on the spectral analysis of acoustic noise concludes that comparison between analogue and digital techniques gives practically similar results.

Papers on the recognition of acoustic signals have one common characteristic: they are interested in sounds generated under controlled conditions which inevitably restrict the relevance of their solutions to the particular problems they attempt to solve. They deal with what is called signature analysis and fault diagnosis of rotating machinery (electrical motors, gear boxes etc.) which is applied to product control. [DART, 1974; GUDAT, 1977 & 1978; DORPFELD & BELLINGER, 1978]. In these cases the details of spectral analysis are barely mentioned, the emphasis being put upon the identification of the generating phenomena. Many factors causing the specific patterns are attributed to a few constituent processes related to the cycle of the machine under test. Under these circumstances, the scatter of the parameters is small and representative templates of the waveforms resulting from different fault conditions are easily obtainable. This results to an exclusive use of minimum distance classifiers.

In some cases the analysis and hence the extraction of features is synchronized by signals obtained from marks on the flywheel [PAVLIDIS & FANG, 1972; USAMI et al, 1978]. Or signals other than acoustic ones are used. A recent workshop paper [FARAG & ROTHWEILER, 1980] reports a device for distinguishing trucks, planes flying, planes taking-off and planes taxiing with a recognition capability better than 80%, which uses seismic as well as acoustic signals. Another paper [PAU, 1977] reports a system for the adaptive classification of acoustic signals from aircraft engines using Kalman filtering. Here a parametrized model of the process to be identified is required.

The only work, known to the author, that involves extensive reports on both spectral analysis and classification is that of Thomas [1971 & 1978] and Thomas and Wilkins [1970 & 1972]. Thomas in his Ph.D thesis [1971] attempted to classify the sounds emitted by two diesel engines fitted to various vehicles. The conditions of recording were controlled in a sense that they were repeatable and were picked up by a microphone in an artificial environment. Nevertheless the control was much less than what applied in fault diagnosis. Thomas used two normalized central moments of spectra in his initial study to obtain a scatter diagram on a two dimensional feature space. The confusion of the classes seemed reasonable to justify further research in measuring the firing rate of the engines which was found to be a cause of confusion between the classes. The method used for the measurement of firing rate was cepstral analysis after unsuccessful attempts with autocorrelation. Cepstral analysis showed that periodicities may be extracted from a waveform with low signal to noise ratio and this may be used to detect the presence of an engine in relatively loud ambient noise. For this second part of his work Thomas used a nearest neighbour classifier (like all other workers in acoustic noise recognition) which did not give any conclusive results. The features used were again moments of spectra and cepstra and the firing period.

d) The fourth and final category of relevant papers comes from the field of speech processing and recognition which has had an enormous development in the last ten years in both academic and commercial research centres, because of the benefits offered by man-machine interaction through speech. All speech recognition systems involve, as a first stage, classification of 'phonemes'. Phonemes are elementary sounds into sequences of which any utterance can be reduced. Classification of phonemes is based on features which can either be fourier transform [SILVERMAN & DIXON, 1974] or linear prediction coefficients [WOHLFORD, 1980]. Cepstral analysis or linear prediction is also used to separate the informative resonances of the vocal cavity, called 'formants', from the exciting waveform of the vocal chords. The use of linear prediction requires that all sources be represented by a model having the same components but with differing parameters according to the speaker, the text spoken, and so on. Since it appears to be impossible to represent all noise pollution sources, or even a substantial number, by any one model, this approach is considered to be of little use in preliminary work. However, it may well prove useful for differentiating between sources whose sounds are similar structurally, for example different types of large, fixed-wing jet aircraft [MOUKAS et al, 1982].

Following classification at the phonemic level, speech is representable as a string of symbols (corresponding to phonemes) which is fed to the final stage of the speech recognition system. This stage can either be a template matching classifier, where templates are prototype words or phrases, or stochastic automata whose probabilities are assigned from the statistics of the language in the case of continuous speech [JELINEK, 1976]. What was said for the previous category can be, more or less, repeated here: because of the a priori knowledge of the position of phonemes in an utterance, extraction of templates is easy, either manually or automatically, and hence the types of classifiers used in phoneme classification are exclusively minimum distance

[SILVERMAN & DIXON, 1976].

As a concluding remark, it can be said that the frustrating impression the student of noise pollution gets, when involved in the study of the literature, as classified above, is that a) there is a small degree of overlap among the mentioned categories; b) there is no apparent relevance to the analysis and recognition of noise pollution and c) there are difficulties in assessing the importance of various approaches even within the same category. It may be mentioned, as an example, that there is no unifying theory, yet, incorporating both the leakage and statistical aspects of windowing. Extended studies on various windows exist, but treat them from the leakage point of view, only marginally referring to their statistical consequences. On the other hand, workers on the statistical aspects of spectral analysis seem to relegate the phenomenon of leakage and satisfy themselves with classical windows of the pre-F.F.T. era. The same happens in the literature on speech and signature analysis. The workers on these fields apparently ignore both modern data windows and the statistical behaviour of spectra. Also, in speech and signature recognition, the existence of classification techniques other than the various distance measures is ignored. Thus, the student of noise pollution keeps wondering about the practical importance of the theories and practices of categories a) and b) above being unable to assess them in a unified manner and evaluate their relevance and applicability in analysing and classifying noise pollution.

3.4 Conclusion

From the discussion of the previous section the following conclusions can be drawn: Spectral and cepstral analysis seems to be appropriate for minor information reduction and extraction of periodicities and reflections in acoustic noise pollution signals. The technique promising to provide spectral estimates of reasonable statistical stability and spectral resolution is that of modified periodograms, using a Hamming or Hanning data window for its relatively good performance and ease of implementation. The nature of noise pollution sounds, i.e. their noisiness and unpredictability, and also the lack of synchronization with the generating processes due to the need for remote monitoring and the fact that the techniques used should be general in order to accomodate a large variety of sources of differing properties, restricts the use of template matching for classification. The LMSE classifier, being versatile, easily implementable and fast seems to be appropriate for this case.

The next chapter includes a study of the properties of noise pollution sounds, their relation to the generating processes and the effects of ground reflections and the Doppler effect. This is the first step towards defining characteristics leading to the extraction of features to be described in the following chapters.

CHAPTER 4

4 PATTERN ANALYSIS I - PROPERTIES OF NUISANCE SOUNDS

As mentioned in the previous chapter, the extraction of features is led by pattern analysis. By pattern analysis we mean the use of whatever is known about the problem at hand to guide the gathering of data about patterns and pattern classes which may exist in the environment being examined. We also mean the subsequent subjection of the data to a variety of procedures for inferring probabilistic and deterministic structures that are present in the data. Pattern analysis is considered an intrinsic and important part of the design process of a pattern recognition system [KANAL, 1974]. Hence, the presentation of the work on the identification of noise pollution sound sources starts with the study of the properties of these sounds and their structure.

Sounds have structure i.e. relationships among their properties, arising chiefly from their mode of generation , e.g. the pulses of rotor noise emitted by helicopters, and the clicks resulting from wheels crossing joints in rails in train noise. In addition, the movement of the source relative to the sensor, the location of the latter and the propagation path may be of importance. The study of the properties of the sound signal received by the sensor can focus on those properties and especially on the ones that reflect the structural and behavioural characteristics of the source and the environment and thus help the extraction of features useful for the identification of the sources. The spectrum of the sound signal will reveal the possible periodicities of the source (section 4.1) and the cepstrum the possible interferences due to the multiple paths (section 4.2). Slow variations in the behaviour of the source and the environment are revealed in the short-time spectral analysis of the signal (section 4.3). The Doppler effect is studied in section 4.4.

4.1 Nature of Sources

The properties of vehicle noise sources as reflected in the sounds they emit has been studied by Thomas [1972], where various types of noises emanating from the different structures of the vehicle have been classified in two general categories: a) Direct air-borne sound which is directly radiated from individual parts of the vehicle and b) Indirect structure-borne sound which is a result of excitation of the bodywork of the vehicle by engine and road.

Most direct and structure-borne sounds can be compared to two other classes of complex sounds, namely musical instrument sounds and vowels. In all these cases the source is a complex tone exciting a resonant system coupled to the source. The resonant system has relatively fixed resonances which modify the exciting sound. The resulting combination of fundamental, harmonics and resonances or broad formants causes certain sets of harmonics to be emphasized relative to others [WEBSTER, 1969].

Thomas also classified the various sources as engine, exhaust and inlet, structure accessories, wheels and terrain, etc. as follows.

Engine noise is due to the explosion in the engine and to the rotation of the crankshaft. The frequency content reflects the constructional details of the engine.

Exhaust noise is due to the sudden release of gas into the exhaust system. The exhaust behaves like a resonant tube excited by impulsive noise. The rate of the impulses depends on the number of strokes and revolutions of the engine. This type of noise is dominant in vehicles by about 5 to 10dB [HILQUIST & SCOTT, 1975].

Diesel engines have a characteristic noise called diesel knock, this sound being attributed to the abrupt pressure rise produced by combustion [THOMAS, 1971]. This pressure produces a hump in the frequency response in the range 800 to 2500Hz and is perceived like a knock [PRIEDE, 1967].

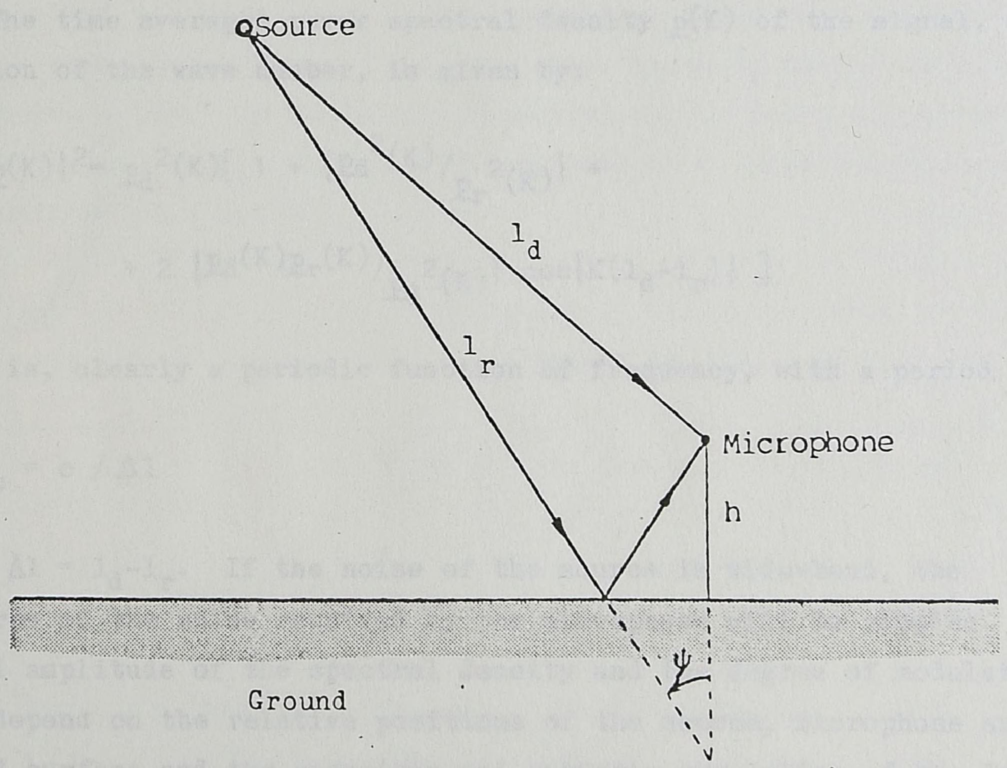
Wheels and terrain produce noise through their contact. This noise is transmitted through the vehicle so contributing to the total. Direct sound from the wheels is limited to a frequency range from 100 to 300Hz and appears to be independent of speed [HARRIS, 1979]. In this category the noise produced by the knocks of the wheels of the train carriage on the rail joints may be classified. In addition, the rails also radiate and in modern trains this is the major source of noise at high speeds.

The major source of noise in jet planes is due to the interaction of the jet exhaust with the surrounding atmosphere. The noise is wide-band and tones reflecting the number of blades in the turbine are superimposed.

In helicopters the characteristic source of noise is due to the blades of the main rotor. The interaction of the revolving blades with the air produces impulsive sounds referred to as the 'blade slap'. Blade slap occurs mainly in tandem rotor and two-bladed main rotor helicopters at a level about 20 dB above the average noise level emited. Practically all helicopters generate impulsive noise at a level 10 to 15 dB above the average level [LEVERTON, 1975]. The tail rotor noise is radiated as a 'whine' which is pronounced during cruising flight. The engine-compressor whine peaks at 8 to 10 kHz in the spectrum of the noise, outside of the range of practical analysis, as it will be seen in later chapters.

4.2 Propagation Effects

The environment and the medium modify the sound field by absorbing, reflecting and amplifying the sound wave emitted by the source, thus the structure of the sound when received is different from that generated originally. The extent of this effect depends on the complexity of the geometry of alternative sound paths and of the acoustic properties of the environment and the medium. Both the geometry and the acoustic properties can be time varying, unpredictable and not well defined and the latter may also be dependent on atmospheric conditions in a way not always well understood. The seriousness of this situation is exemplified by the complexity of the international standards for the noise certification of aircraft and the fact that tests made under the very strict conditions permitted in the regulations are not repeatable from day to day and from site to site [SMITH, 1977].



<u>Figure 4.1.</u> Multiple path reception due to ground reflection.

It seems that one of the prominent environmental effects and probably one that is best understood and consistent is the effect of <u>ground reflections</u>.

The acoustic signal received by a microphone positioned at a height h from the ground surface, as shown in figure 4.1, is expressed as:

$$p(t) = p_d(f)\ \exp\{jK(ct-l_d\} + p_r(f)\ \exp\{jK(ct-l_r)\} \qquad (4.1)$$

where $p_d(f)$, $p_r(f)$ are the sound pressure amplitudes as functions of frequency, for the direct and reflected waves, respectively; l_d and l_r are the path lengths of the direct and reflected waves, respectively; c is the ambient speed of sound, $K=2\pi f/c$ is the acoustic wave number, f is the frequency, t is time and $j=\sqrt{-1}$ [SYED et al, 1980].

The time averaged power spectral density $\underline{p}(K)$ of the signal, as a function of the wave number, is given by:

$$|\underline{p}(K)|^2 = \underline{p}_d^{\,2}(K)[\ 1 + \{\underline{p}_d^{\,2}(K)/\underline{p}_r^{\,2}(K)\} +$$

$$+ 2\ \{\underline{p}_d(K)\underline{p}_r(K)/\underline{p}_d^{\,2}(K)\}\ \cos\{K(l_d-l_r)\}\ ] \qquad (4.2)$$

which is, clearly a periodic function of frequency, with a period

$$F_p = c\ /\ \Delta l \qquad (4.3)$$

where $\Delta l = l_d - l_r$. If the noise of the source is wide-band, the spectrum of the noise received by the microphone will be rippled. The actual amplitude of the spectral density and the degree of modulation will depend on the relative positions of the source, microphone and ground surface and the geometric and acoustic properties of the latter and of the transmitting medium. If $l_d >> h$, that is when the source is far from the microphone, Δl may be approximated by $2h\cos\psi$, where ψ is the angle shown in figure 4.1. Thus, the period of spectral ripples becomes

$$F_p = c\ /\ 2h\ \cos\psi \qquad (4.4)$$

Ground reflection ripples can be removed by a technique due to SYED et al [1980], which uses Cepstral Analysis [CHILDERS et al, 1977]. Specifically, the power cepstrum of the signal, which is the Fourier Transform of its power spectrum, will feature a peak corresponding to the periodicity of the spectral ripples. This peak may then be "filtered" out - or liftered according to cepstral terminology. However, although ground reflection is an undesirable effect, when one is interested in estimating the 'free-field' spectrum of the source, it provides valuable information to aid the identification of aircraft, since it is related to the position of the source relative to the microphone. Specifically, the corresponding cepstral peak is a unique feature in that it discriminates between sources that lie very close to the ground and those that are very high above: $\cos\psi$ tends to zero (fig. 4.1) and hence F_p to infinity as the source approaches the ground (equation 4.2). Table 4.1 gives various values of F_p for normal atmospheric conditions and various combinations of h and ψ.

Table 4.1

Ground Reflection Ripple Periods

$F_p = c/2h \cos\psi$ (Hz)   c=343.25 (m/s)

h(m)	Angle (degrees) 0°	15°	30°	60°	89°
0.5	343.25	355.36	396.35	686.50	19671.
1.0	171.62	177.68	198.17	343.25	9835.
1.5	114.42	118.45	132.12	228.83	6557.
2.0	85.81	88.84	99.10	171.63	4917.
2.5	68.65	71.10	79.27	137.30	3934.
3.0	57.21	59.22	66.06	114.42	3278.
3.5	49.03	50.76	56.62	98.07	2810.
4.0	42.91	44.92	49.54	85.81	2458.

It may be noted that the smallest values of F_p occur when the source is directly above the microphone and that for $\psi=89^o$ the first peak occurs above 5kHz. This and the fact that signal attenuation rises rapidly with frequency suggests that the cepstra of wide-band sources near the ground do not exhibit the characteristic peak. The position of the microphone may be chosen in such a way that possible cepstral peaks due to reflections other than by the ground, or due to harmonics in the spectrum, do not interfere with the peak due to ground reflection.

The reflection effect is prominent in jet aircraft spectra and is superimposed upon discrete peaks due to narrow-band tones from the compressor and turbine of the jet engine. The wide-band noise in aircraft spectra arises from interaction between the jet exhaust and surrounding atmosphere. Helicopters do not emit their jet exhaust at such a high velocity. Narrow-band sounds such as turbine whine and bladeslap dominate the output, thus helicopter spectra do not exhibit the characteristic ripples. Sources close to the ground, also generate bands if they are wide-band and localized, though the bands are widely spaced. If, however, the source is distributed, as is the case of traffic noise, the bands cancel. However, the path of propagation does affect the spectrum of received sound. The precise relevance of propagation to identification will not be known until much more data have been examined, though the increase of attenuation with frequency in the atmosphere will probably ensure that low frequencies are worth most effort. The effect of long propagation paths is to reduce all sounds to an indistinguishable roar.

4.3 Time Variation

The above considerations, however, cover only one aspect of the problem. As previously mentioned, the operating modes of the sources and the geometry and acoustic properties of the environment and the medium are time varying, thus the signal received by the microphone exhibits a harmonic content changing with time.

A Fourier representation of the signal has the form:

$$s(t) = \sum_i D_i(t) \cos\{i\omega_o(t)t + \phi_i(t)\} \tag{4.5}$$

where D_i are the Fourier coefficients, ω_o is the fundamental frequency, and ϕ_i is the phase, for the ith Fourier component. All these parameters are time varying. This leads to the <u>Spectrogram</u> representation of the signal, where the power is represented as a function of frequency and time. Conceptually, the spectrogram is computed by imposing a sliding observation window on the signal and taking the Fourier Transform of the part revealed by the window. The signal is assumed to be stationary within the window; the size of the window is, therefore, chosen to justify this assumption.

The major variations arising from the operating modes of the sources which are inherent in the operation of the engine (vibrations, firing, resonances etc.) are normally quasi-stationary with a period usually shorter than the observation window. The corresponding frequencies will appear as straight lines along the time axis. Variations that occur slowly relative to the observation window can be periodic (helicopter main rotor noise), quasi-periodic (wheels crossing joints in rails in train noise), or aperiodic. These will appear in the spectrogram as isolated spectra (or groups of spectra) with distinctive structure and/or total power (in each spectrum) significantly differing from that of the neighbouring spectra in a way that cannot be masked by random variations of the power due to wind and similar effects.

The ground reflection effect is superimposed on the above structure and is stationary if the source does not move relative to the microphone. Otherwise, the angle ψ in expression 4.4 changes and with it the period of the spectral ripples.

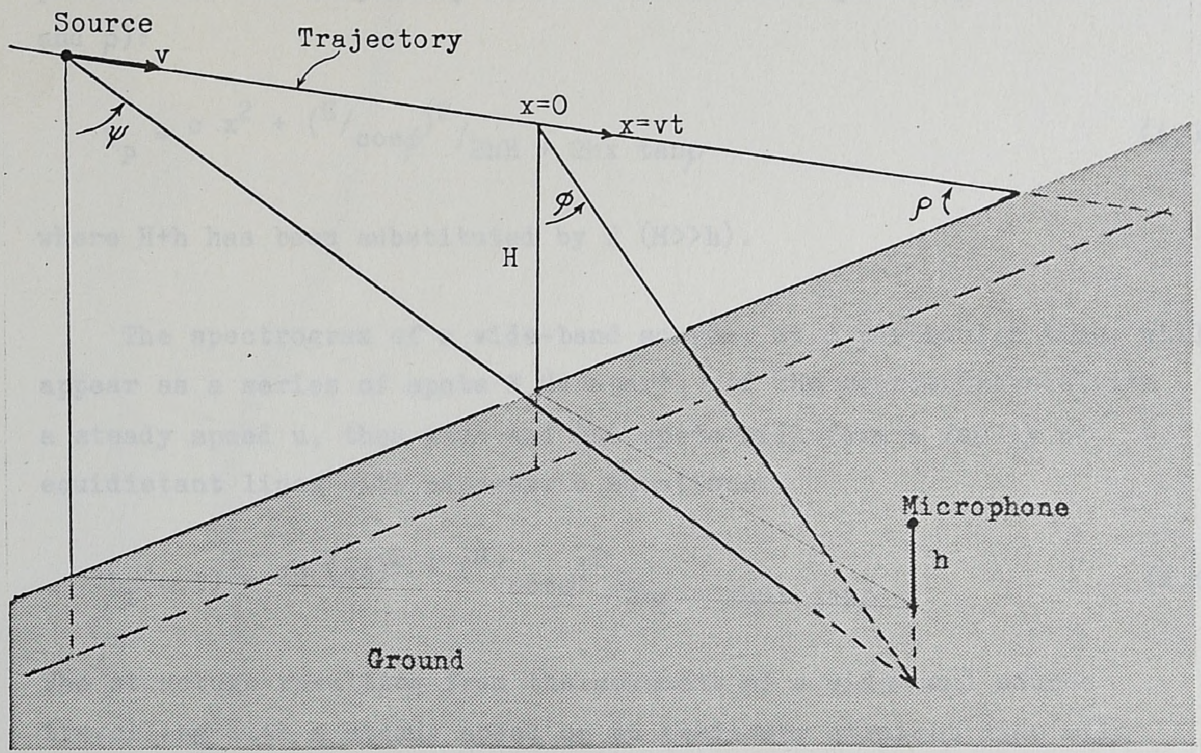


Figure 4.2. Source travelling over ground.

Let the source of fig. 4.1 move on an arbitrary straight line, as in figure 4.2. We can evaluate:

$$\cos\psi = {}^{x \tan\rho + H + h}/_{\,x^2 + ({}^{H + h}/_{\cos\phi})^2} \tag{4.6}$$

where x is the distance of the source from the point on the trajectory which is nearest to the microphone; H is the height of this point; ϕ is the value of ψ for x=0; and ρ is the slope of the trajectory.

By substituting the above value of $\cos\psi$ in (4.4) F_p becomes a function of the position of the source on its trajectory and of the position of the trajectory relative to the microphone (parameters H, ϕ and ρ):

$$F_p = c\, x^2 + \left({}^{H}/_{\cos\phi}\right)^2 /_{2hH} + 2hx \tan\rho \qquad (4.7)$$

where H+h has been substituted by H (H>>h).

The spectrogram of a wide-band source, at a particular time, will appear as a series of spots F_pHz apart. If the source travels with a steady speed u, then x=ut and the spots will form a family of equidistant lines with parametric equations:

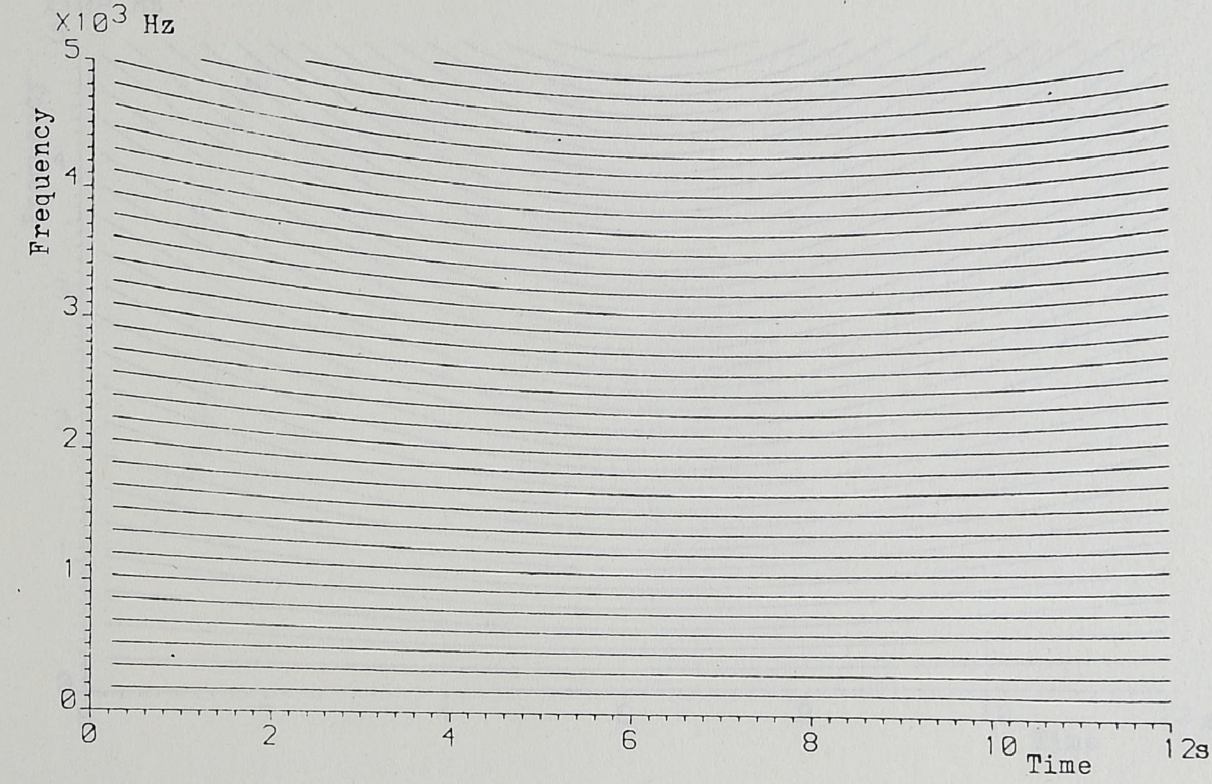
$$L_f(n,t) = nc\, (ut)^2 + \left({}^{H}/_{\cos\phi}\right)^2 /_{2hH} + 2hut \tan\rho \qquad (4.8)$$

The structure resulting from the movement of a wide-band source travelling with a steady speed on an arbitrary straight line will appear as shown in the simulated spectrograms of figure 4.3. The lines become nearly horizontal when the velocity and the inclination of the trajectory are small.

4.4 Doppler Effect

Another complication arising from the movement of the source is the Doppler effect. The shape of the spectrum is retained [YOUNG, 1968], but the frequency bands are shifted proportionally to the velocity of the source: frequency bands appear higher when the source is approaching and lower when it is receding, while the "true" frequencies are received when the source is closest. The characteristic pattern of figure 4.1(a) can be identified as due to Doppler effect when the shift from high to low frequencies coincides with the overhead position. The latter can be found by (a) the position of the maximum of the received power and (b) the position of the minimum period of the ground reflection ripples.

Figure 4.3.a. Simulated spectrograms of white noise source travelling over hard ground surface.
H=500m, v=200km/h, ϕ=20°, ρ=3°, h=1.2m.



<u>Figure 4.3.b.</u> Simulated spectrograms of white noise source travelling over hard ground surface.
H=500m, v=400km/h, ϕ=20°, ρ=3°, h=1.2m.

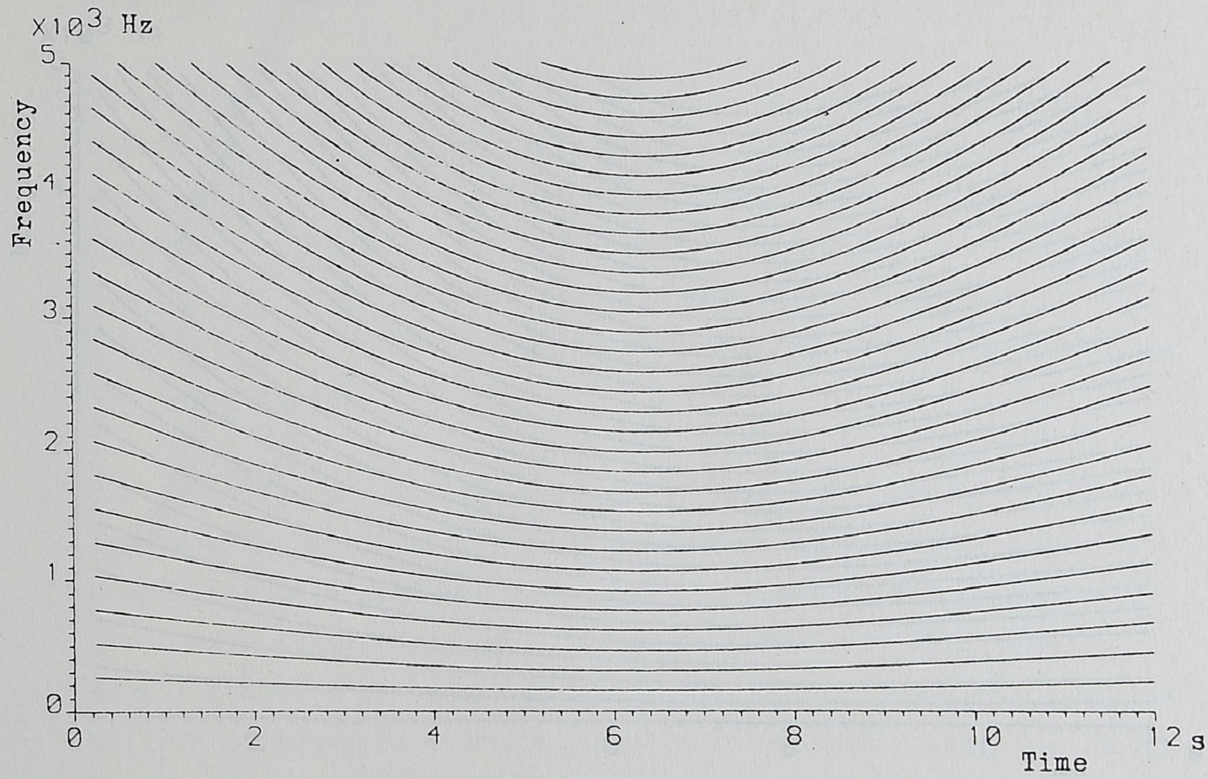
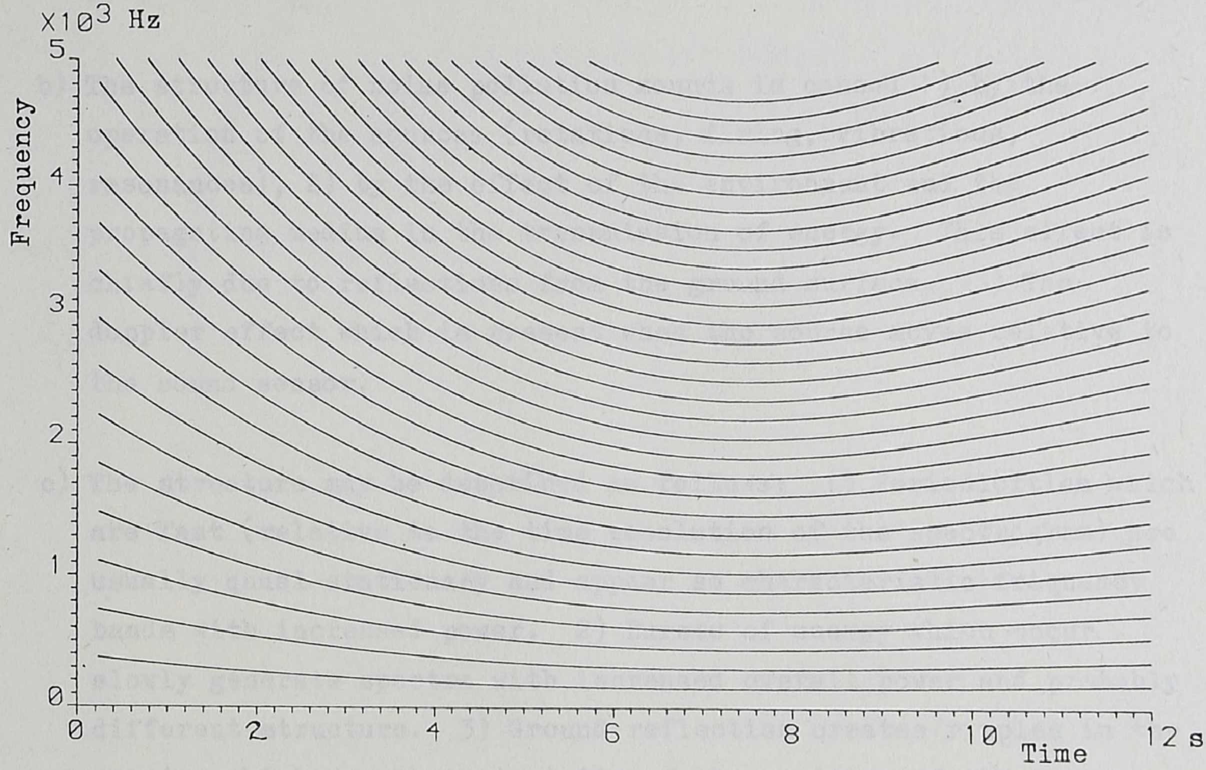


Figure 4.3c. Simulated spectrograms of white noise source travelling over hard ground surface.
H=500m, v=400km/h, $\phi=20^\circ$, $\rho=15^\circ$, h=1.2m.



4.5 Summary and Conclusion

The results of the analysis of this chapter may be summarized as

a) The behaviour of the noise pollution sources usually changes with time and hence the sounds they emit. Therefore, spectrogram analysis seems to be the most appropriate for the study of noise pollution sounds.

b) The structure of noise pollution sounds is caused 1) by the operation of the sources (rotations, firing, vibrations, resonances), 2) by the effect of the environment and the propagating medium in the transmission of energy. This effect is chiefly due to reflections from the ground surface. 3) The doppler effect which is present when the source moves relative to the sound sensor.

c) The structure may be described as follows: 1) Periodicities which are fast (relative to the time resolution of the spectrogram) are usually quasi-stationary and appear as characteristic frequency bands with increased power. 2) Bursts of energy which occur slowly generate spectra with increased overall power and probably different structure. 3) Ground reflection creates ripples in the spectra which are characteristic of the position of the source relative to the pick-up sensor and the ground. The ripples in the spectra appear as equidistant lines in the spectrogram and 4) The doppler effect shifts frequencies without changing their relative positions.

The conclusions of this analysis must be verified with actual data. The next chapter contains a description of the digitization system developed for the creation of a data base of real data. The actual generation of spectrograms is presented in chapter 6.

CHAPTER 5

5 DATA ACQUISITION & DIGITIZATION

In chapter 2 of this thesis we mentioned the iterative nature of the design of a pattern recognition system. The advantages of human interaction and intervention in all phases of the iterative design process and the important role of interactive computing and display in making this feasible are obvious. This work has, thus, employed digital simulation in all phases of the design process and consequently, it has involved the preparation and transformation of environmental noise records to a form suitable for digital processing. For this project, the analogue noise records were provided by the Scientific Branch of the Greater London Council, who made the recordings in situ. Editing, preprocessing and digitization were done at the Instrument Systems Centre Although the Centre had the equipment necessary for the digitization, no software was available. Section 5.2 includes a description of the hardware and discusses the specific problems that had to be solved for the development of software for the digitization of "long" acoustic signals, under restrictions imposed by the nature of the hardware. Editing of the tapes to remove spurious and degraded signals and anti-aliasing filtering is described in section 5.1. Finally, section 5.3 presents the data base which was set up to enable detailed analysis of a large number of noise records.

5.1 Recording and Editing of Analogue Data

For sophisticated analysis of sound, sound samples have to be recorded at the place of their occurence on magnetic tape and later be processed at the laboratory. Magnetic recording involves the conversion of one of the time variables of the sound field (sound pressure or particle velocity) to a magnetic variable of space. The microphone achieves the conversion from acoustic into electrical form, with an intermediate mechanical stage. Ideally, the conversion is accomplished without distortion of the time history representing the physical quantity being measured or recorded, i.e. sound pressure. In detail, if the input signal is a one-dimensional real function of time $s(t)$, the output of the ideal microphone must be $s'(t)=c \cdot s(t)$ where c is a calibration constant. This ideal situation, however, is difficult to achieve in practice, for gain and phase varies with frequency and distortion due to non-linearities is unavoidable. Nevertheless, in most cases this is not a serious problem. Microphones available commercially cover a wide range of applications and have characteristics optimized for each specific application. If used within the limits of their specifications, control and relative correction of errors is possible.

Further problems in the recording process occur due to non--linearities of the magnetization process which is the final stage of the recording process, caused by design compromises in magnetic tape recorders. In addition, low and high cut-off frequencies of tape recording are dependent on the magnetization process and the head gap width, respectively. All these effects may be dealt with adequately by using various modulation techniques which increase the frequency bandwidth and the signal to noise ratio and by increased tape speed which raises the high cut-off frequency [BENDAT & PIERSOL, 1971].

Beyond the limitations inherent in the magnetization process there are some other problems associated with the variations in tape speed called <u>time-base errors</u>. The main time-base error is flutter which may be defined as a variation of tape velocity from the nominal. The improvement of the design of the tape transport system e.g. servo--controlled motors, greatly reduces this error.

The original tapes were recorded by G.L.C, at various locations. The tapes were accompanied by data sheets with descriptions of their contents, environmental and recording conditions and operating conditions of the sources. The tapes were aurally inspected to locate <u>suitable events</u>. Events were regarded to be suitable when they could be clearly identified aurally by a human observer. Degraded events, i.e. events that were excessively noisy distorted, saturated etc., were discarded.

The suitable events were copied on a tape in groups of records of the same class at a speed of 7.5 ips. Although the signals had no significant power above 4 kHz, they were low pass filtered at 7kHz-12db/octave, during copying, to ensure no aliasing (section 5.2). Another feature of the signals was the concentration of most power in the low frequencies below 200Hz. In order to emphasize the middle frequencies which seem to contain most of the useful information pre-whitening is required [THOMAS, 1973]. The most appropriate form of pre-whitening for acoustic signals is the 'A' weighting used in acoustic measurements, which most closely approximates the weighting of the human ear [MOUKAS, 1976]. However, due to the unavailability of such a network no pre-whitening was applied to the signals. Instead, the middle frequencies were enhanced in the computation of the spectra, as will be explained in the next chapter, section 6.3. The majority of the events had a duration of approximately 13 seconds, although some were considerably longer. To cope with the blocking problem (section 5.2.3), the events were edited to a duration of approximately 12 seconds. Thus, longer events were copied in parts, and shorter events were extended. All records were preceded by identifiers and descriptions.

Two REVOX tape recorders were used for the editing and copying. The REVOX A77 is a High Fidelity 4-track stereo tape recorder with very good characteristics for sound recording and reproduction. Wow and Flutter is 0.08% @ 7.5 ips and 0.01% @ 3.75 ips. Frequency Response is 30 to 20,000 Hz +2/-3 dB and 30 to 16,000 Hz +2/-3 dB @ 7.5 and 3.75 ips, respectively.

5.2 Digitization System

Digital analysis of analogue data requires their conversion into a series of discrete numbers. This process is termed digitization and may be separated into sampling and quantizing. Sampling is the process of defining the values of a signal at certain instants. The sampling theorem states that the original data can be uniquely reconstructed from the sampled train if the original is band-limited and the sampling rate is at least two times the signal bandwidth. The Nyquist frequency is half the sampling rate and must therefore be higher than the bandwidth. Otherwise aliasing occurs, i.e. frequencies higher than the Nyquist frequency are "folded" into the lower frequency range and are confused with the data in this range, as explained in chapter 3, section 3.1.1.1. Aliasing is overcome by ensuring an upper bound in the frequencies of the analogue signal using low-pass filtering as mentioned in the previous section. The cut-off frequency of the filter depends on the actual frequency content of the signal and the available sampling rates. The latter depend on the speed of the hardware combined with the efficiency and complexity of the software that performs the digitization.

Quantization is the conversion of the analogue values of the sampled waveform which may occupy a continuum of values, to numbers of finite length. Quantization noise is inherent in this process and decreases with the length of the numbers, as explained in chapter 3, section 3.1.1.2. Optimal quantization is obtained by selecting the quantization levels to ensure minimum noise [MAX, 1980]. This is achieved by companding the signal before digitization [SMITH, 1957]. For 12-bit evenly spaced numbers the signal to noise ratio is more than 70dB and poses no problem. However, this applies when the signal uniformly occupies the whole 12-bit range and can be considerably lower with nonuniformly distributed signals and particularly at low signal values. When no companding is employed, automatic gain amplifiers should be used to ensure the full swing of the signal without clipping. For this project, a special program was developed to ensure this without the automatic gain which was unavailable (section 5.2.3).

The nature of the digitization system of the Instrument Systems Centre imposed limits on the sampling rate that was finally realised (section 5.2.2). In addition, the special problem caused by the duration and high data rate of the records combined with the properties of the digitization system had to be solved. The solution was not the most elegant but one which avoids sophisticated programing e.g. double buffering (section 5.2.4). This section, therefore, starts with the exposition of the hardware in order to make the development of the software comprehensible.

5.2.1 Equipment

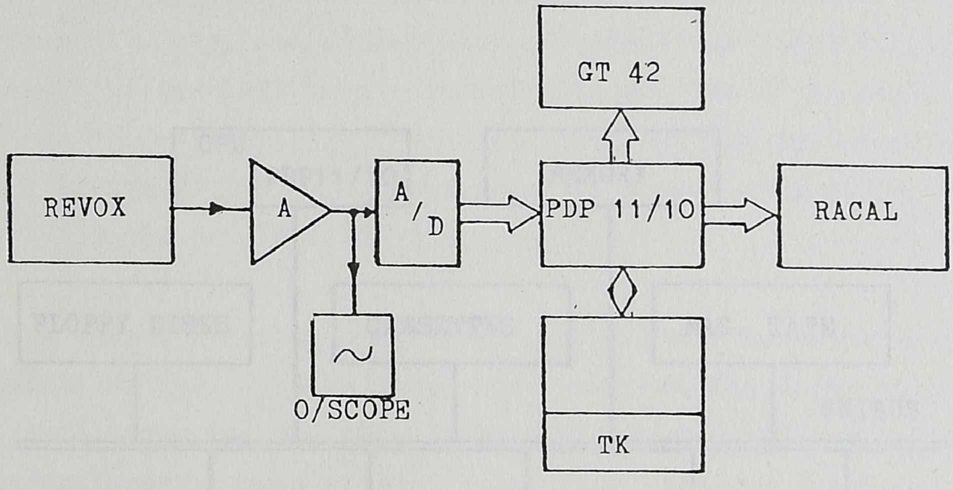


Figure 5.1. Digitization System.

The Digitization System is shown in fig. 5.1. It consists of the REVOX A77 analogue tape deck, the Analogue to Digital Conversion System, the PDP-11/10 Computer System and the RACAL T7000 7-track Digital Tape Deck.

The Analogue to Digital Conversion System (ADC) consists of an Analogic MP6812 Data Acquisition System, four voltage amplifiers and additional circuitry and power supplies. The MP6812 is a monolithic system, including a 12-bit A/D converter and logic to perform conversion on 16 channels. The system is wired for four channels and is switchable for 10 or 12 bit conversion and for positive or alternating waveforms. The four amplifiers provide a switchable gain of 10, 30 or 100. The input range of the converter is ±10.24 Volts. The corresponding range of the ADC output is 0 to octal 7777 in the 12-bit option. There is no automatic gain control.

The RACAL T7000 7-track Tape Deck (MT) consists of the tape transport mechanism and logic and the Data Formating Unit (DFU) with logic for data formating and control of the read/write process at three selectable densities of 200, 556, or 800 bpi. The speed of the tape movement is fixed at 37.5 ips. The seventh track is used for even or odd parity check [RACAL]

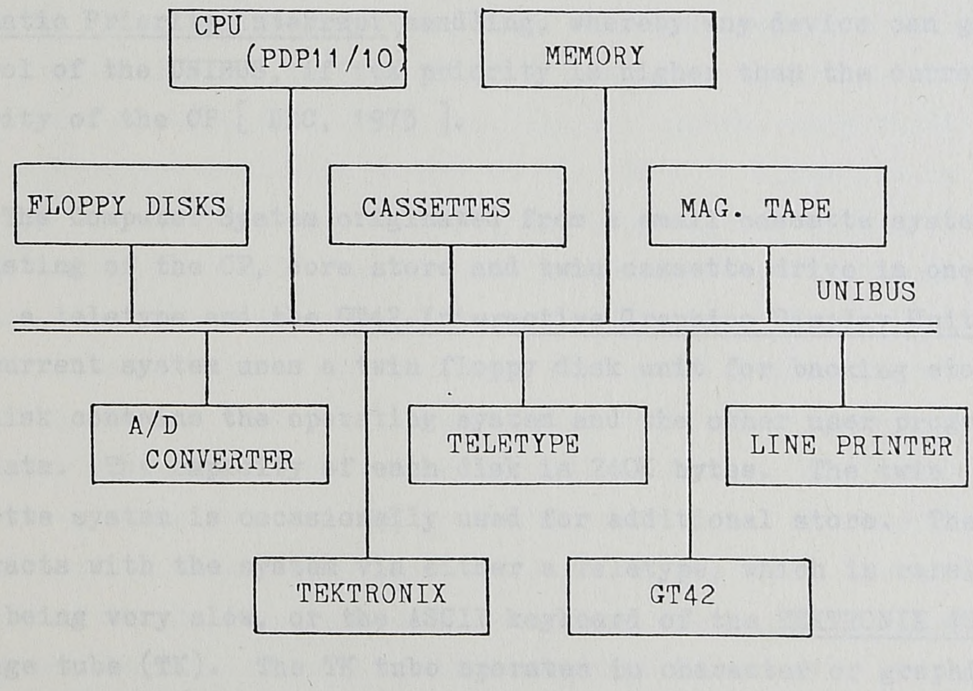


Figure 5.2. PDP-11 Computer System.

The digitization process is controlled by the PDP-11/10 Computer System of the Centre (fig. 5.2). The Digital Equipment Corporation (DEC) PDP-11 is a 16-bit minicomputer with 24K words of core memory. All computer system components and peripherals connect to and communicate with each other on a single high speed bus known as the UNIBUS. This configuration allows Direct Memory Access and asynchronous operation of all system components. Memory locations, processor registers and peripheral device registers are assigned an address on the UNIBUS, which may be manipulated as flexibly as main memory by the Central Processor (CP). The CP, which is connected to the UNIBUS as a subsystem, controls the time allocation of the UNIBUS for peripherals and performs arithmetic and logic operations and instruction decoding. The CP contains 8 general purpose registers, two of which are used as the Program Counter (PC) and Stack Pointer (SP) of the system. An additional register is the Processor Status Word (PSW) containing information on the current processor priority, operation modes and condition codes describing the result of the last executed instruction. The CP features Automatic Priority Interrupt handling, whereby any device can gain control of the UNIBUS, if its priority is higher than the current priority of the CP [DEC, 1973].

The Computer System originated from a small cassette system consisting of the CP, core store and twin cassette drive in one unit, a teletype and the GT42 Interactive Graphics Display Unit. The current system uses a twin floppy disk unit for backing store. One disk contains the operating system and the other user programs and data. The capacity of each disk is 240K bytes. The twin cassette system is occasionally used for additional store. The user interacts with the system via either a Teletype, which is rarely used being very slow, or the ASCII keyboard of the TEKTRONIX 4006-1 storage tube (TK). The TK tube operates in character or graphic mode to display hardware generated alphanumeric characters or stationary graphics, respectively. The GT42 graphic display allows interactive graphics with a light pen. It is particularly useful for the display and manipulation of fast dynamic processes. In alphanumeric mode it can simulate a VDU.

The computer system operates under the RT-11 operating system [DEC, 1976]. The RT-11 is a collection of system programs and routines coordinated by the Monitor, which communicates with the user via the keyboard. The user can call programs by typing certain commands. The system provides a number of program development aids such as Editor, Peripheral Interchange Program (File manipulator), Assembler, Linker, Librarian, FORTRAN IV compiler and other utilities. A comprehensive library allows the FORTRAN programer to use system routines for file and character handling, timing etc. Other libraries enable the use of GT42 and TK for graphics.

5.2.2 Constraints & Programing Considerations

The constraints imposed by the specifications of the available equipment determined the way digitization was to be performed. The noise records to be digitized were of about 10 to 20 seconds duration. Digitization of each record at a sampling rate of 20kHz would produce 200,000 to 400,000 12-bit samples requiring 150K to 300K words of storage, if they were closely packed, many times greater than the available memory of the PDP-11. Since timing is critical in real-time applications, the RT-11 system should be locked making direct storage on the floppy disk very difficult. Segmentation of the noise record would require instantaneous start/stop operation of the analogue tape under program control, to avoid transients. This is impossible due to the inertia of the tape transport mechanism. The only alternative left was the direct transfer of data from the ADC to the digital tape. However, this solution also imposed some restrictions: On one hand, the speed and density of the digital tape deck determined the possible sampling rates, as shown in table 5.1. The sampling rate is half the byte rate, since each sample read from the ADC corresponds to two bytes on the magnetic tape. On the other hand, the speed of the PDP-11, which undertakes the task of controlling and reading the ADC and of transferring the bytes to the DFU of the tape deck, imposed an upper bound to the sampling rate. The software that was developed safely copes with 10,425 samples per second corresponding to the 556 bpi density of the tape.

Table 5.1

Possible Sampling Rates at 37.5 ips

Density (bpi)	Byte Rate (bps)	Sampling Rate (sps)	Nyquist Freq. (Hz)
200	7500	3700	1875
556	20850	10425	5212.5
800	30000	15000	7500

This rate was low for the digitization of noise signals. Time base expansion had to be utilized, to compress the frequency range of the signals. Time base expansion occurs when data recorded at a speed R_r are played back at a speed R_p such that

$$r_R = R_p / R_r < 1.$$

The effective sampling rate f'_s is then

$$f'_s = f_s / r_R.$$

where f_s is the sampling rate. For the REVOX $r_R = 0.5$ ($R_r = 7.5$ ips and $R_p = 2.75$ ips), giving an effective sampling rate of 20850 Hz, sufficient for the task in hand.

The nature of the interfaces of the ADC and the RACAL dictated the programing technique to be employed for the development of the digitization software. The interfaces are Digital DR11-C general purpose, providing logic and buffer registers for program controlled transfers [DEC(b), 1973]. There are three addressable registers for each interface (fig 5.3), namely: The Control and Status

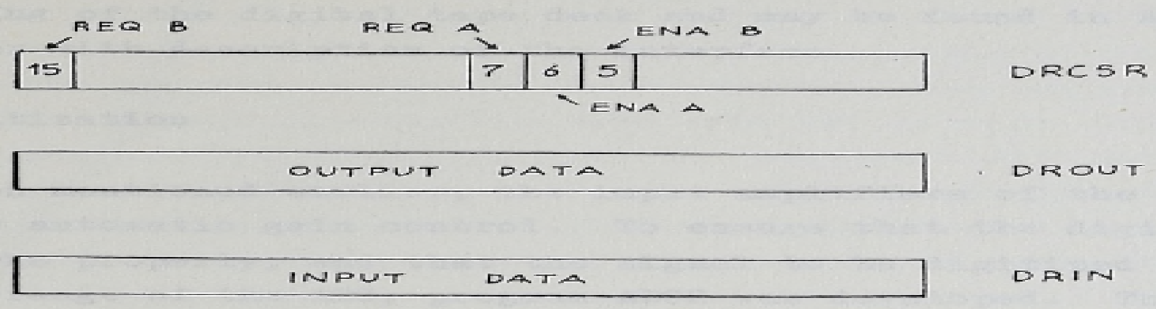


Figure 5.3. DR-11C Interface Registers.

Register (DRCSR), used to control two interrupt channels and to provide user-defined command and status functions for the external device; the read/write Output Register (DROUT); and the read-only Input Register (DRIN). Each interface is assigned two Interrupt Vectors, one for each channel. The vector is a pair of consecutive words where the PSW and the address of the first instruction of the Interrupt Service Routine of the device are loaded under program control. When an interrupt is requested on either of the two interrupt channels by user hardware, the corresponding Request bit (REQ) of the DRCSR is set. If the Interrupt Enable bit (INT EN) of that channel is set, then the interrupt is automatically passed to the processor. If the priority of the interrupt is higher than the current processor priority, the interrupt is accepted. The current contents of the PSW and PC are automatically pushed onto the processor stack, new values are loaded from the corresponding interrupt vector and the interrupt service routine is executed immediately. The "old" situation is restored upon execution of a special "return from interrupt" instruction by restoring the previous contents of PSW and PC from the stack.

Several Assembly language routines and MACROS were written for the handling of the digital tape deck and may be found in Appendix I, together with description of the interface.

5.2.3 Digitization

As has been mentioned earlier, the input amplifiers of the ADC do not employ automatic gain control. To ensure that the digitization system works properly, and that the signal to be digitized is within the input range of the ADC, program ADC2 was developed. The program operates in two modes: In the Examine Mode the signal is read in "blocks" of 1024 or more samples and the histogram of the block is displayed on the GT42. If the signal is clipped the bell of the keyboard rings. Thus, the user can adjust the gain of the amplifier of the REVOX to bring the signal within the input range of the converter. In addition, the waveform can be displayed on the TEKTRONIX screen to ensure that no aliasing occurs, by comparing the signal to that displayed on the CRT of the oscilloscope, connected to the output of the amplifier of the ADC. Moreover, the actual values of the waveform can be examined. During examination of each block, it is difficult to stop the analogue tape without loosing a part of the signal that may contain bursts with amplitudes outside the permitted range. To secure that the whole signal is tested for clipping the Test Mode is provided: the signal is digitized uninterruptedly and the bell rings whenever the waveform is clipped. In addition, the user may optionally examine whether the clipping is in the upper or lower extremes of the range. Program ADC2 is written in FORTRAN. The routines that read the ADC are written in assembler. Routine ADC reads data, stores them in a buffer and computes the histogram. It is called by the FORTRAN program in Examine Mode. Routine ADCLIP reads samples from the ADC continuously and checks whether the signal is clipped; the samples are discarded after the test. The waveform is assumed to be clipped, if three or more consecutive samples are all zero or all octal 7777. ADCLIP is called in Test Mode.

For the actual digitization of the signals, program DIG was developed. Interrupt programing was used to synchronize the processor with the tape movement, which provides the timing of the process. The program is divided in two parts: The main part initializes the digital magnetic tape unit (MT), loads the interrupt vector, locks out the RT-11, prompts the user to start the analogue tape and issues the WRITE command to the MT. Every time a new byte is needed, the MT interrupts the processor, causing a jump to the interrupt service routine. The routine is divided into two parts -it is in fact two separate interrupt service routines, each providing for the control of the program to be transferred to the other, when the next interrupt occurs. This technique was dictated by the fact that one ADC cycle corresponds to two MT cycles. The interrupt routines read the sample from the input register of the ADC interface, issue the command to start next conversion, split the 12-bit sample into two 6-bit parts, which are sent to the output register of the MT interface in the two MT cycles. Digitization is terminated if any key on the keyboard is hit. This causes an interrupt of higher priority and program control passes to a third interrupt routine, which initiates the process of program termination. Finally, two file marks are written on the tape and the tape is rewound. The program, which is written in assembly language, calls some of the tape handling routines and macros referred to in the previous section.

Listings of programs ADC2 and DIG may be found in Appendix II together with the description of the interface.

5.2.4 Tape Blocking

In all standard computer operating systems, as supplied by manufacturers, magnetic tapes can be read only if they are blocked, i.e. if the data consist of physical segments, called Blocks, separated from the neighbouring blocks by Inter-block Gaps, usually 0.75 of an inch. Blocking is necessary for the following reasons: Only a small part of the data, its size depending on the particular system, can be read and stored in memory for processing. This means that the tape must be stopped every time the part to be processed is read. The position of the tape is then undetermined, due to the inertia of the transport mechanism of the tape, combined with the high speed of the movement of the tape and the high density. The inter-block gaps account for this imprecision.

Tapes produced by program DIG are unblocked. Each record has a duration of approximately 10 to 20 seconds. This corresponds to one block of data of approximately 400,000 to 800,000 bytes, too large for most computer systems. Division in smaller blocks, while the analogue signal is being fed, would cause a loss of 208.5 samples every time an inter-block gap is written on the tape at the density of 556 bpi. (The gap is 0.75 of an inch, corresponding to 0.75 X D bytes, where D is the density). Interruption of the analogue signal is impossible without indeterminate time-base errors due to the transport mechanism of the analogue tape.

It was, thus, decided that the blocking of the digital tape should be done off-line. The idea was to read a number of bytes from the tape, stop the tape, and transfer the bytes onto the floppy disk. As the position of the tape is undetermined, because it is stopped at a position other than an end of block, the tape is rewound, the bytes previously read are skipped and the next "block", starting after the last byte of the previous reading, is read. The process is repeated until all bytes are transferred. Finally, the tape is rewritten in blocks of the desired size, with data transferred from the floppy disk.

The blocking of the tape done as described above is time consuming. If N_b is the total number of bytes on the tape and M_b is the number of bytes to be read each time, then

$$(i-1)M_b+1,\ (i-1)M_b+2,\ \ldots,\ iM_b$$

are the bytes transferred on the ith iteration. The total number of iterations is $N_i = N_b / M_b$. If T_b is the time taken for the M_b bytes to be read from the tape, then the total time needed is:

$$(1+2+3+\ldots+N_i)T_b = N_i(N_i+1)T_b / 2$$

T_b is proportional to the 'block' size and inversely proportional to the byte rate which is 2 times the sampling rate f_s times the time base expansion coefficient r_R. The total time needed for the 'blocks' of data to transferred from the tape will thus be:

$$N_i(N_i+1)M_b / 4f_s r_R$$

For a 12 second record, digitized at f_s=20,850 Hz and M_b= 20,850 bytes, the number of iterations N_i is 24 (12x20,850x2/20,850). The above formula yields 300 seconds or 5 minutes for the blocking of such a record. This duration excludes the time required for the tape to rewind at the end of each iteration and the time required for the data to be written on the floppy disk. Although the last two processes occur simultaneously, they nevertheless delay the process. The reverse process, i.e. the transfer of data from the floppy disk to the tape also requires time which is $1/r_R$ times the duration of the record, as far as writing on tape is concerned. In practice this is longer, due to the time required for the data to be read from the disk. A record of 12 seconds fills two floppy disks with data transferred from the tape.

Program BL6 was written to accomplish the blocking task. Part I reads the tape and writes onto disk. Part II is the reverse procedure. The user may optionally examine any block of data by listing the values or plotting the waveform on the TK screen. The program shares the same library of tape handling routines as DIG. In addition, it uses a special routine, NXRDMT, to skip and read the next "block" of data from tape. Listing with comments are in Appendix III.

5.3 Data Base

Apart from the technical problems (dealt with above) concerning the acquisition of data, of major concern in the design of a pattern recognition system is the setting up of a data base. Pattern analysis involves the inspection of a large set of data to enable definition of features for use in the classification stage. This set of data will also be used for the training and evaluation of the classifier.

Two principal considerations are important in the formulation of the data base: Firstly, the definition of the pattern classes and the selection of typical patterns samples. The data set must contain a representative selection of each class. The samples must not be especially atypical and must be sufficient in number. Secondly, the format of the records and the organization and management of the data base to enable easy retrieval of a particular sample, editing etc.

In this study, three classes of sound were considered, namely: aeroplane, helicopter and train noises for the reasons stated below:

a) They are of environmental importance in London and analysis of the noise for planning and other decisions would be aided by reliable automatic monitoring.
b) Recordings of them were readily available from existing attended noise monitoring programs.
c) The characteristics of these sources are relatively stationary with respect to a particular site, as opposed to other sources,

e.g. street noises.

d) Analyses of these sources are available in the literature.

The study of these classes was expected to reveal characteristic features, and provide an insight into the relationship of those features to the operating states of the sources and to those of the environment.

A number of records of the above classes was digitized at an effective sampling rate of 20,850 kHz. Two floppy disks were used for the blocking of each record. Thus, the size of one digital record was 480K 6-bit bytes, or 245,760 samples, corresponding to a duration of 11.79 seconds. The tapes with the digitized records were transferred to the University of London Computer Centre (ULCC), where they were copied on 9-track tape using program UPTAPE (Appendix IV). Record identifiers and descriptions were included at the beginning of each record. So, each record of the data base consists of:

a) Record Identifier
b) Class and Subclass Name (Operating Characteristics of Source)
c) Part
d) Label of G.L.C. Tape and Counter Index
e) Label of Analogue Data Base Tape and Counter Index
f) Recording site
g) Event
h) Sampling Rate
i) Comments
j) Data

The Identifier consists of 6 character of the form <u>Add.dd</u>, where A is alphabetic denoting the class, and d is digit. The first pair of digits denotes the subclass and the second pair is the number of the particular record. Descriptors (d) and (e) associate the digital record with the analogue tape. Descriptors (f) and (g) are the code names of the recording site and of the event, as reported in the G.L.C data sheet. Descriptor (c) is the part number of of the record, when the record is a part of a longer event. Table 5.2 gives a typical description of a record. Table 5.3 gives a description of the data base. No names were given in the data sheets for the helicopters.

Table 5.2

Typical Data Base Record Description

A02.02 CONCORDE TAKE-OFF
303.485-N1.716 SITE 23 EVENT 1 PART 2 LINEAR RECORDING

Table 5.3

Description of Data Base

Source	Aeroplanes	A	B	Helicopters	A	B	Trains	A	B
1	Concorde	2	5	Type 1	4	5	Br.Rail	3	4
2	Viscount	2	3	Type 2	3	7	Br.Rail Dsl	1	1
3	Trident	4	5	Type 3	4	5	Londn Trnsp.	1	1
4	I-11	1	1						
5	Airbus A300	1	2						
Total		10	17		11	17		5	6

Column A: Number of separate events

Column B: Number of 12 second records

An editing program AMTAPE was written to facilitate the editing of wrong identifiers, descriptors, or corrupted records (Appendix IV).

5.4 Summary and Conclusion

For the verification and further processing of the findings of the first part of pattern analysis of chapter 4 the need exists of collection of data, i.e. acoustic records of various noise pollution sources. Since, for reasons of economy, versatility and interactive development, digital simulation has been adopted throughout the work described in this thesis, digitization of the analogue acoustic records which were provided by the G.L.C. was necessary. The unavailability of proper hardware and software for the digitization of acoustic signals has led to the development of such a system based on existing hardware (a 'domestic' tape recorder, an A/D converter a minicomputer and a digital tape deck). On one hand, the duration and high data rate of the acoustic signals and the speed and capacity of the minicomputer system on the other, led to a solution which, although time consuming, is nevertheless unique in that it enables digitization and storage of any 'long' one-dimensional signal in a hardware environment not tailored to such applications. Such environments are most likely to be met in nonspecialist laboratories and, therefore, the solution offered here may be used as a guideline.

The digitization system was used to build up an extensive data base of environmental noise pollution sound records, namely jet aircraft, helicopter and train noise records, which are of specific environmental importance in the greater London area. The spectral analysis of those records, in the form of spectrograms, is the subject of the next chapter of this thesis. This analysis will reveal the structure of the sounds of the data base, the relationship of this structure to the characteristics of the sources and hence it will lead to the extraction of the specific spectral and temporal invariants that will eventually be used in the classification stage of this work.

CHAPTER 6

6 PATTERN ANALYSIS II - SPECTROGRAM GENERATION

In the examination of the properties of noise nuisance sounds in chapter 4 of this thesis we referred to the spectral analysis of the signals and the necessity for spectrogram representation of their spectral content, since the signals under examination vary with time. The second stage of pattern analysis has involved the development of a system for the generation of spectrograms of the noise pollution records of the data base presented in the previous chapter, in order to verify the results of the first stage of the analysis and thus lead to the definition and extraction of proper features.

Spectrograms have long been used to display the spectral structure of signals, traditionally in speech analysis. A special instrument has evolved for this purpose, namely the <u>sound spectrograph</u>, where the spectrum is obtained through a bank of analogue band-pass filters and is displayed by burnt marks on fascimile rolling paper. The density of the marks is roughly proportional to the logarithm of the spectral magnitude. The marks are capable of depicting an intensity range of 12dB [FLANAGAN, 1972]. The narrow dynamic range and the facts that neither additional processing of the spectrograms nor exact measuements of the spectra are possible limit the use of the instrument. In addition, the unavailability of such an instrument in the Instrument Systems Centre forced the development of software to simulate the generation of spectrograms digitally. The digital computation of spectrograms is given in section 6.1 where problems of resolution are also discussed. The tackling of the narrow dynamic range of the display medium is given in section 6.2 where the development of a system for the enhancement of spectrograms and for information reduction is given.

Programme FSGO3P implements spectrogram computation, enhancement and display. It has been developed initially at the CDC 7600 of the University of London Computer Centre (U.L.C.C.) which has the facility of plotting grey levels on microfilm with the use of a special package, namely PICPAC. The program was later modified for interactive use at the Science Research Council Prime 550 located at the Centre where various techniques and ideas could easily be applied and assessed and accordingly be applied at the CDC software, as well. A brief description of FSGO3P is given in section 6.3. Spectrograms of actual sounds obtained with FSGO3P are presented and discussed in section 6.4.

6.1 Computation of Spectrograms

The computation of spectrograms presented here is based on the spectral analysis techniques elaborated in chapter 3. In order to make the presentation comprehensible this section starts with some definitions leading to the formal definition of the spectrogram and the elaboration of the resolution problem.

6.1.1 Definition of the Spectrogram

Let s(t) be the time function representing the sound signal. The sampled data signal s* resulting from sampling s(t) every T seconds will be

$$s^*(t) = \sum_{k} s(kT)\,\delta(t-kT) \qquad (6.1)$$

where $\delta(t)$ is the Dirac Delta function. This sampled data signal may also be viewed as the sequence

$$\{s^*(k)\} = \{s(kT)\}, \quad k=1, \ldots, L_s \qquad (6.2)$$

where L_sT is the duration of s(t) and $\{.\}$ denotes a sequence. The sequence of (6.2) is one record of the data base.

Let us segment the sampled data signal $\{s^*(k)\}$ as follows: Each segment will comprise N samples and the i'th segment will be the sequence

$$\{x(i,k)\} = \{s[(i-1)L+k]\}, \quad k=1, \ldots ,N \tag{6.3}$$

where L is an integer such that $0<L\leqslant N$.

From this definition it follows that consecutive segments overlap if $L<N$ and the number of common samples is N-L. The ratio

$$r = N-L/N \tag{6.4}$$

is the <u>overlap coefficient</u> and is such that $0\leqslant r<1$.

Let us now weight each segment with a weighting sequence w(k). The weighted i'th segment called the i'th <u>window</u> will be

$$\{x_w(i,k)\} = \{x(i,k)\ w(k)\}, \quad k=1, \ldots ,N \tag{6.5}$$

The <u>Discrete Fourier Transform</u> (D.F.T.) of this window will be the sequence [AHMED & RAO, 1975]

$$\{X_w(i,l)\} = \{^1/_N \sum_{l=1}^{N} x_w(k) \exp(-2\pi jkl/N)\}, \quad l=1, \ldots ,N \tag{6.6}$$

where j= -1. The <u>modified periodogram</u> obtained from the i'th window (defined in section 3.1.2.2) is

$$\{P_x(i,l)\} = \{^N/_U\ |X_w(i,l)|^2\} \quad , l=1, \ldots ,N \tag{6.7}$$

where $|X_w(i,l)|$ denotes the amplitude of the D.F.T. of the i'th window and U is a scaling factor depending on the weighting sequence w(k) (defined in relation (3.9)).

The <u>Discrete Power Spectrogram</u> of the signal $\{s^*(k)\}$ may then be defined as the two-dimensional sequence or array

$$\{S\} = \{ \{P_x(i,l)\},\ l=1,\dots,N\},\quad i=1,\ \dots\ ,N_w \qquad (6.8)$$

where N_w is the number of windows in the signal. Obviously from (6.2) it follows that N_w is the largest integer such that

$$(N_w-1)L+N \leqslant L_s$$

The D.F.T. of each window is an approximation to the Fourier Transform (F.T.) of the corresponding section of the original signal s(t). The discrepancy between the two transforms arises from the sampling of s(t) and the fact that D.F.T. is finite whereas F.T. is infinite. Ideally, the square of each component of the sequence resulting from the D.F.T. should represent the power of the corresponding spectral band of width $\Delta f=1/NT$. However, due to the finite length of the transformed window this power "leaks" into neighbouring components, the degree and form of <u>leakage</u> depending on the weighting sequence. This sequence is used to reduce leakage by reducing the sample values at the ends of the segment. The weighting, however, scales down the amplitude of the transform. This effect is compensated for in the computation of the modified periodogram which is the 'best' estimate of the power spectrum of the corresponding segment of the signal. An exposition of the derivation of the D.F.T., its properties and its correspondence to the F.T. and references was offered in chapter 3 of this thesis.

6.1.2 Resolution Considerations

From the definitions given and the remarks made above it follows that the frequency resolution of the spectrogram is $1/NT$ and time resolution is LT or $(N-rN)T$ from relation (6.4). We see that time resolution depends on the overlap coefficient and the size of the window. Overlap of windows in the computation of spectrograms is a consequence of windowing: As mentioned before, a significant part of the signal is ignored due to the windowing sequence exhibiting small values at its two ends. On the other hand, the correlation of the spectra of consecutive windows increases with the overlap coefficient; also the computation required. Thus, r cannot but be small, e.g. 0.25 to 0.50 [HARRIS, 1978].

It follows, therefore, that the two resolutions are not independent. Assuming that T is fixed (from the sampling process), for a given N, i.e. frequency resolution, L is restricted by the correlation which increased overlap incurs. In fact, as frequency resolution improves ($1/NT$ decreases), time resolution deteriorates (LT increases).

6.1.3 Selection of Window Size.

Selection of N is thus a matter of trade-off between time and frequency resolutions. This can be based on the nature of the signal to be analysed and the objectives of the analysis. As very fine frequency resolution resulting from a large N is obtained at the expense of time resolution, it can only be applicable when the signal does not change significantly within the window.

As the helicopter noise exhibits 10 to 15 rotor beats per second, a window of 0.1 seconds duration would be required, which is too long for all the signals under consideration to be considered stationary. Nevertheless, impulsive slow periodicities are revealed as vertical equidistant lines, in the spectrogram, as in figure 6.1c. and their period can be measured in some other way (section

7.2.2). Thus, after some experimentation, a value of 512 was selected for N. This sets the fundamental frequency resolution at 40.72 Hz for the sampling rate of 20,850Hz. However, the actual resolution of the spectral components of the Discrete Fourier Transform is worse due to leakage. For the Hamming window function this is 1.81 times the fundamental (table 3.1), hence 72.7Hz. It can be found from table 4.1 that ground reflection ripples can be adequately revealed, provided the pick up microphone is positioned below about 2 metres from the ground. The overlap coefficient that was selected was 0.5. From the definitions above and (6.4) this corresponds to a time resolution of 12.2ms, which was thought to be short enough for the sounds under examination to be considered stationary.

6.2 Spectrogram Enhancement and Display

The procedure presented in the previous section computes a 'good' estimate of the power spectrogram of acoustic signals. For the purposes of this project, however, no exact measurement of the power of the signals is necessary. What is actually needed is a procedure to compute the distribution of the spectral content of the signals under consideration in a way that the potentially 'useful' attributes of the signals are enhanced. By 'useful' we mean those attributes that help to discriminate the classes. In this light, certain modifications in the computational procedure are desirable.

On the other hand, spectrograms obtained as described in the previous section have shown that most power of the environmental noises consistently occupied low frequencies and the ratio of the highest to the lowest amplitude was very large. This was in agreement with Thomas's findings about vehicle noises [1972]. As we mentioned in the previous chapter no pre-whitening was applied to the noise records of the data base. Hence, the modification to be applied to the computation of the spectrogram must also take into account this additional fact.

6.2.1 Scaling

The first form of whitening that may be applied to the spectrogram in order to compress its dynamic range is to plot the amplitude instead of the power. The amplitude being the square root of the power gives a 'mild' whitening [COHEN, 1970]. In addition, in order to compress the low frequencies, it was decided to weight the spectra with a sequence $\{H_f(l)\}$ representing the transmission characteristic of the high-pass filter. Ideally, low frequency compression should simulate the 'A' weighting of sound level metering (which account for the subjective evaluation of noise), likely to be employed in the implementation of the instrument envisaged. For reasons of simplicity, however, a fist order high-pass filter with a cut-off frequency of 1000 Hz (the same as the 'A' weighting filter) was simulated instead. The amplitude spectrum of the i'th window was thus made

$$\{X_f(i,l)\} = \{|X_w(i,l)|\ H_f(l)\}, \quad l=1, \ldots ,N \tag{6.9}$$

where $\{H_f(l)\}$ is the frequency response of the filter. This filtering prevents potentially useful peaks at higher frequencies being lost. In effect it is equivalent to pre-whitening mentioned in section 3.1. Furthermore, to cope with the narrow dynamic range of the display process, grey pixels with intensity proportional to the logarithm of the amplitude were displayed, as given in the relation below

$$\{X_s(i,l)\} = \{\log X_f(i,l)\}, \quad l=1, \ldots ,N \tag{6.10}$$

Other scaling functions can be used e.g. square root etc.

To summarize, the spectrogram that was finally computed and displayed is given by the formula:

$$\{S\} = \{ \{scl[\,|X_w(i,l)|\; H_f(l)]\},\quad l=1,\ \ldots\ ,N/2\},\ i=1,\ \ldots\ ,N_w \qquad (6.11)$$

where 'scl' denotes some 'whitening function'. It must be noted that in this formula any scaling factors that scale equally all frequencies have been dropped. Also, that the first N/2 spectral components are used, since the second N/2 are symmetric.

6.2.2 Thresholding

Although the smoothing of the logarithm is much too severe and lowers the contrast of the spectrogram, the difference in the structure of the noises is evident in figures 6.1. However, it can be seen that the characteristic structure is formed by fast variations of power in the spectrogram which is superimposed on a background in which power declines with frequency. The background is more or less the same for all classes and hence it does not help in discriminating between the classes. To observe the useful structure it was extracted from the background using the following variable thresholding scheme, developed for this purpose: The average scaled spectrum $\{\ \underline{X}_s(l)\ \}$, over all spectra in the spectrogram, was calculated.

$$\{\ \underline{X}_s(l)\} = \{{}^1/_N \sum_{i=1}^{N_w} X_s(i,l)\},\quad l=1,\ \ldots\ ,N/2 \qquad (6.12)$$

All frequency bands that consistently exhibit high power are thus preserved, whereas irregular peaks are smoothed out. In addition, a heavy smoothing is applied on this average spectrum

$$\{\ \tilde{\underline{X}}_s(l)\} = \{{}^1/_{2n+1} \sum_{q=-n}^{n} \underline{X}_s(l+q)\},\quad l=1,\ \ldots\ ,N/2 \qquad (6.13)$$

where each component in $\{ \underline{X}_s(l) \}$ is substituted by the average of 2n+1 neighbouring components and n is such that

$$n = m \quad \text{for } l-m > 0$$
$$n = l-1 \quad \text{for } l-m \leqslant 0$$

The value of m is a matter of choice according to the particular signal analysed. This yields the overall trend of the spectrogram along frequency, that is the background.

A modified spectrum is subsequently created by subtracting the smoothed average from each scaled spectrum

$$\{X_m(i,l)\} = \{X_s(i,l) - \tilde{\underline{X}}_s(l)\}, \quad l=1, \ldots ,N/2 \tag{6.14}$$

The thresholded spectrum $\{X_t\}$ is then computed as follows

$$\{X_t(i,l)\} = \{1 \text{ if } X_m(i,l) \geqslant t \text{ else } 0\}, \quad l=1, \ldots ,N/2 \tag{6.15}$$

where t is selected such that

$$\sum_{i=1}^{N_w} \sum_{l=1}^{N/2} X_t(i,l) = c\, N_w\, N/2 \tag{6.16}$$

that is, a fraction c of the spectrogram lies above the threshold.

This scheme was implemented in software and empirically, c=0.1 was found to be appropriate for most of the cases, as shown in figures 6.2. The function used for grey scaling (relation (6.11)) was the square root and the smoothing parameter m was 15. The method works well when the average power of the spectrum does not vary significantly within the spectrogram (figure 6.2c). The threshold should ideally be variable to accommodate time variations also.

6.3 Program FSGO3P

The computation, scaling, enhancement and display of spectrograms that were described in the previous sections is accomplished by program FSGO3P which has been developed for this purpose. FSGO3P comprises the main body (FSGO3P) and a number of routines that perform certain tasks. FSGO3P accepts a number of parameters such as the window size N, overlap L, type of scaling (logarithmic, linear, or square root), the smoothing parameter m etc. Routine READID accepts the identifiers of the records to be analysed and FINDID retrieves them from the data base one at a time. Subsequently, COMPG and COMPS compute auxiliary parameters combining the ones read with information retrieved from the data base (length of record, sampling rate etc.)

The actual computation of spectra is performed by TRANSIA. TRANSIA unpacks and segments the record (relation (6.3)), windows it (relation (6.5)) and performs the computation of the D.F.T. with the use of a base-2 Fast Fourier Transform (F.F.T.) algorithm (relation (6.6)). The amplitude of the transform is then calculated and whitened with the high-pass filter (relation (6.8)).

SCALIA scales the windows (relation (6.10)) and calculates the average scaled spectrum (relation (6.12)) which is subsequently smoothed (relation (6.13)) by routine SMOOTH.

MODPIC calculates the modified spectra (relation (6.14)) and TLEVEL evaluates the threshold (relation 6.16).

Finally PPLOT plots either the grey spectrogram produced by TRANSIA, SCALIA or MODPIC, or the binary one using the threshold calculated by TLEVEL.

Flow charts of FSGO3P and the routines may be found in APPENDIX V.

6.4 Results and Discussion

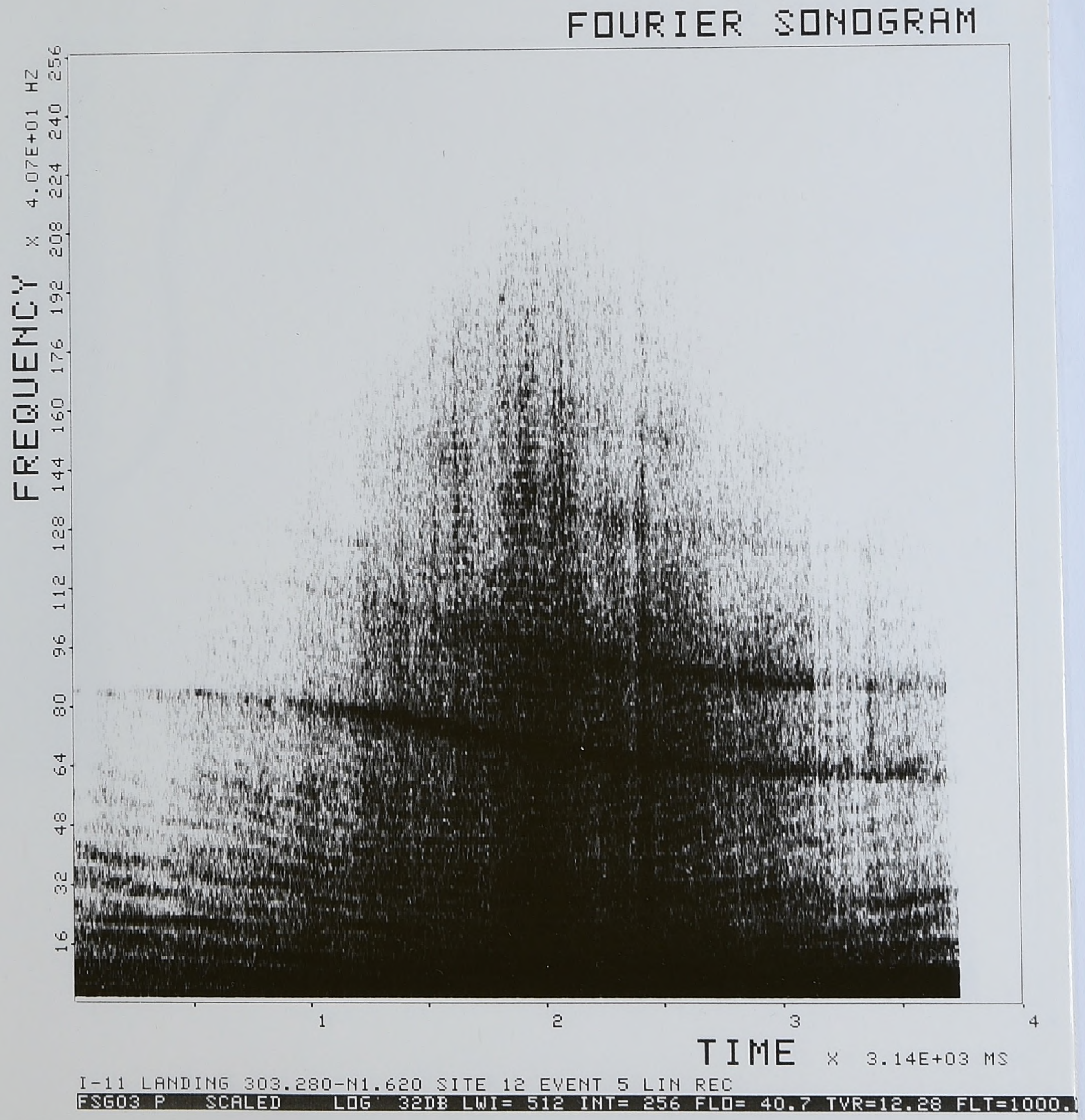
Figures 6.1 and 6.2 show spectrograms of typical records of the data base generated by program FSGO3P. The spectrograms a, b, c and d of figure 6.2 are the thresholded or binary versions of spectrograms a, b, c and d of figure 6.1, respectively.

In figure 6.1a, which was generated by the noise of a jet aircraft landing, two structures can by observed. Firstly, the ground reflection effect appears as a family of equidistant curves, as expected after the discussion of chapter 4 on the properties of pollution sounds. It is worth noting that the curvature of the lines is quite pronounced, indicating that, either the velocity of the jet was high or its distance from the pick-up microphone was small. Unfortunately, the data sheets that accompanied the noise records do not give any details on the speed and location of the jets relative to the microphone, nor any details about the recording site. However, it may be inferred that it is small distance rather than the velocity of the aircraft that creates the curvature, for no high speed is expected in take off.

The second noticable pattern is the set of horizontal lines above 2 kHz shifting to lower frequencies when the jet passes the overhead position. The shift is obviously due to the Doppler effect. The position of these lines, which are due to the compressors of the jet engines, depends on the number of the turbine rotor blades, hence on the type of the engines of the aircraft.

Figure 6.2a is the binary version of figure 6.1a, processed according to the method presented in section 6.2.2. The extraction of the structure (which is the purpose of the processing) is not very successful, mainly because of the temporal variation of the intensity of the sound signal. A threshold level variable with time would be more succesful in extracting the structure. Indeed, the variability of the threshold with frequency has helped in revealing the ground reflection ripples at low frequencies.

<u>Figure 6.1.a.</u> Computed spectrograms of jet landing.
Time window 24.56ms; Overlap 50%; Nominal time resolution 12.28ms;
Nominal frequency resolution 40.72Hz.



<u>Figure 6.1.b.</u> Computed spectrograms of jet *taking-off*. Time window 24.56ms; Overlap 50%; Nominal time resolution 12.28ms; Nominal frequency resolution 40.72Hz.

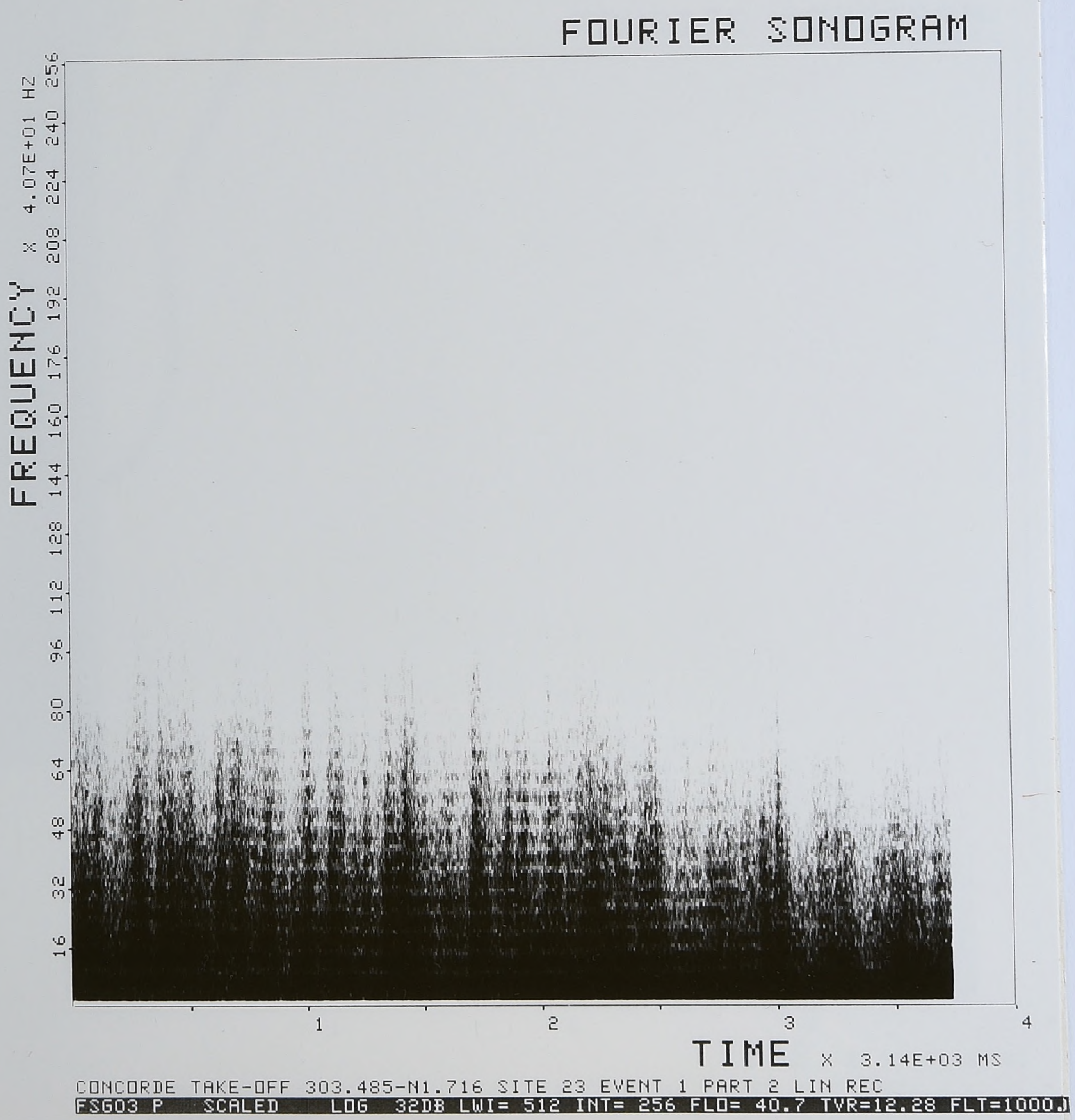


Figure 6.1.c. Computed spectrograms of helicopter hovering. Time window 24.56ms; Overlap 50%; Nominal time resolution 12.28ms; Nominal frequency resolution 40.72Hz.

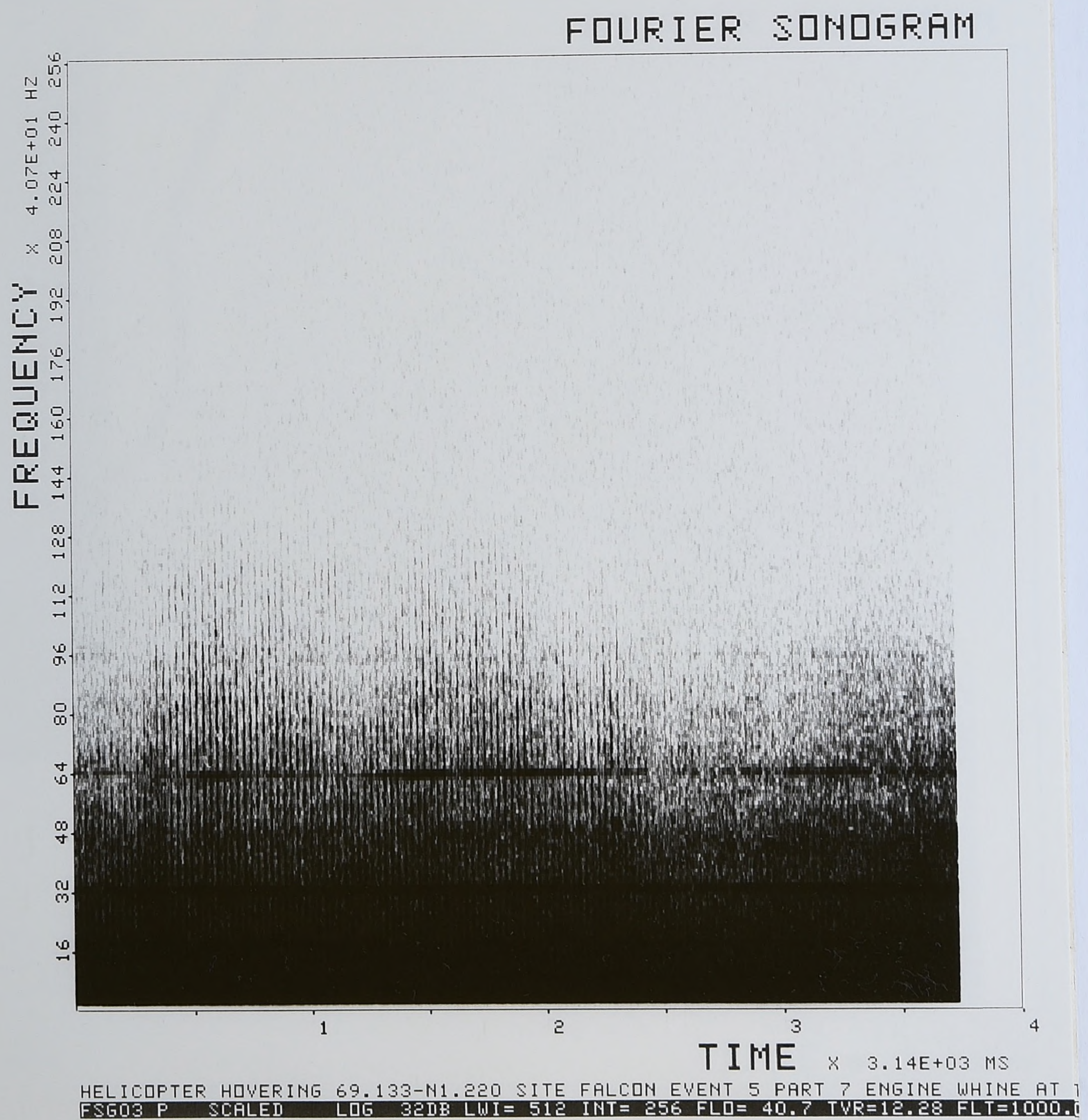
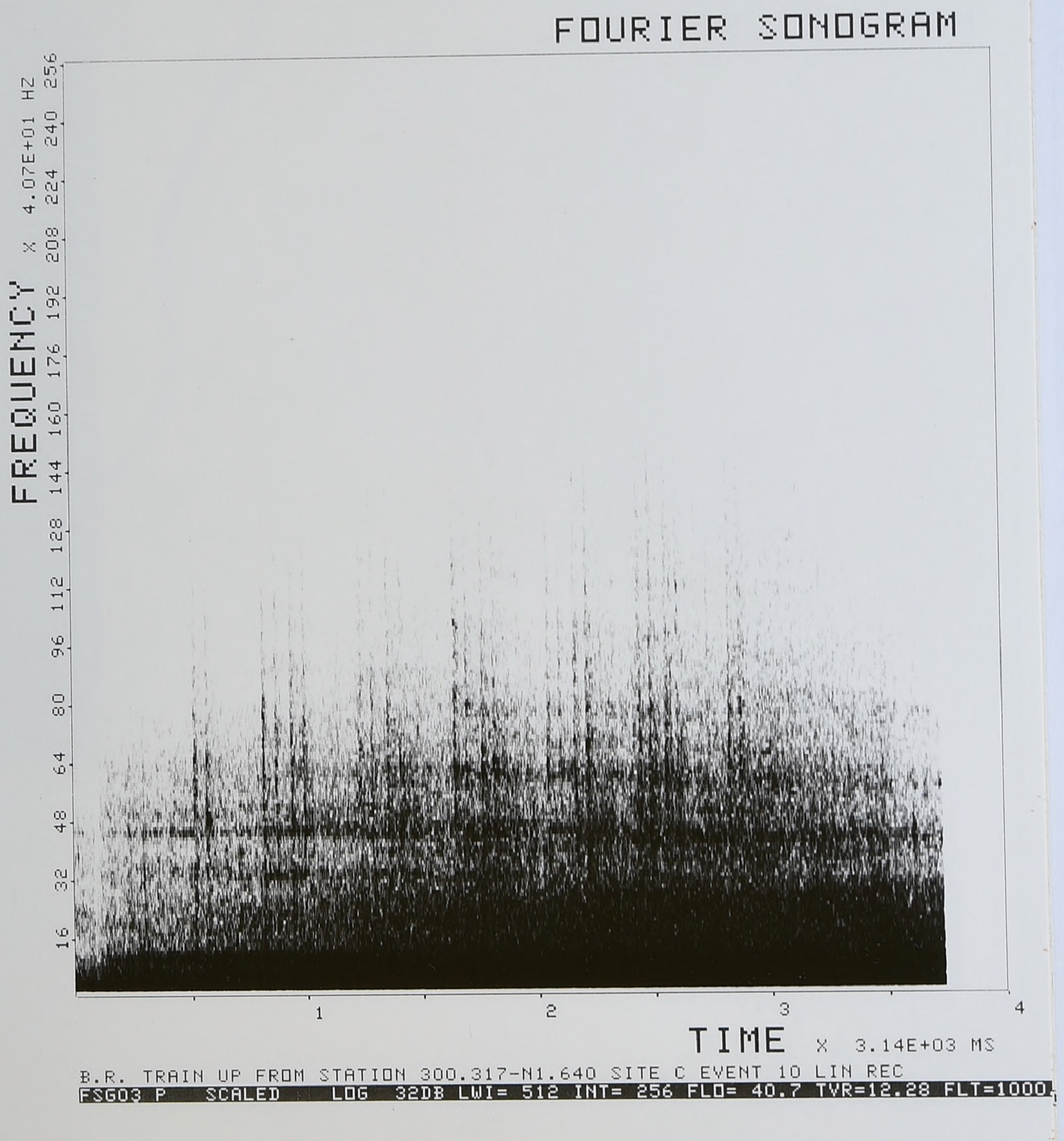


Figure 6.1.d. Computed spectrograms of train.
Time window 24.56ms; Overlap 50%; Nominal time resolution 12.28ms;
Nominal frequency resolution 40.72Hz.



The spectrogram of figure 6.1b (whose binary version is shown in figure 6.2b) is from a jet taking-off. The only structure seen here, as in all spectrograms of taking-off jets of the data base, is the ground reflection lines. The lines, in this case, are nearly horizontal indicating either that the speed of the aircraft is low, or that is at coniderable distance from the microphone. The lack of engine tones may be due to the operating condition of the aircraft or to the conditions of recording. All landing jets were recorded in a different site than the jets taking off. Again, the lack of information on flights and recording prevents any further insight into the causes of the structure of the spectrograms. Since the speed of the aircraft at landing cannot be different from its taking off speed, the hypothesis that, in this case, the distance of the source from the microphone was considerable is supported by the following two facts: First, high frequencies, as the jet tones, are attenuated with distance. Second, ground reflection lines are nearly horizontal in the spectrograms of distant sources.

The performance of the binary spectrogram is better than the previous one, since the overall variation of intensity is less in this case. Still, the fast random intensity variations create vertical gaps in the binary picture which could be mistakenly regarded as meaningful structure.

Contrary to jet noises, for the helicopter (figures 6.1c and 6.2c) bands due to ground reflection are not significant, although the source is wideband and multi-path propagation is possible. This is because the wideband noise from the jet exhaust is of much lower intensity. Instead, vertical lines are dominant due to the main rotor blades which create an impulsive waveform of high power and wide frequency content. The performance of the binary spectrogram in revealing the line structures is striking in this case. Worth noting are the horizontal lines at low frequencies which are completely hidden in the grey spectrogram. Lines due to compressor whine are also present.

The train spectrogram (figures 6.1d and 6.2c) shows a fairly steady horizontal pattern at three frequency bands. Although the source is wideband, it is effectively at the same height as the microphone so that the bands cannot result from reflection interferences. They most probably are due to resonances of the engine and the bodywork of the train. The regular vertical bars indicate the rail joint knocks of the wheels of the carriage.

6.5 Summary

In this chapter, spectrograms of actual sources were computed and displayed, as given by formulas (6.5), (6.6) and (6.11), on a general purpose digital computer. Hamming windowing was applied for leakage reduction. The window size was 24.56 ms with 50% overlap, giving nominal frequency and time resolutions of 40.72 Hz and 12.28 ms, respectively. A scheme for the processing of the spectrograms in order to enhance the characteristic structures was also presented.

The spectrograms obtained verify the results of the analysis of chapter 4 and suggest features which may be used for identification of the sources. The definition and extraction of features is the subject of the next chapter.

Figure 6.2.a. Binary spectrogram of jet landing..
(parameters as in figure 6.1.)

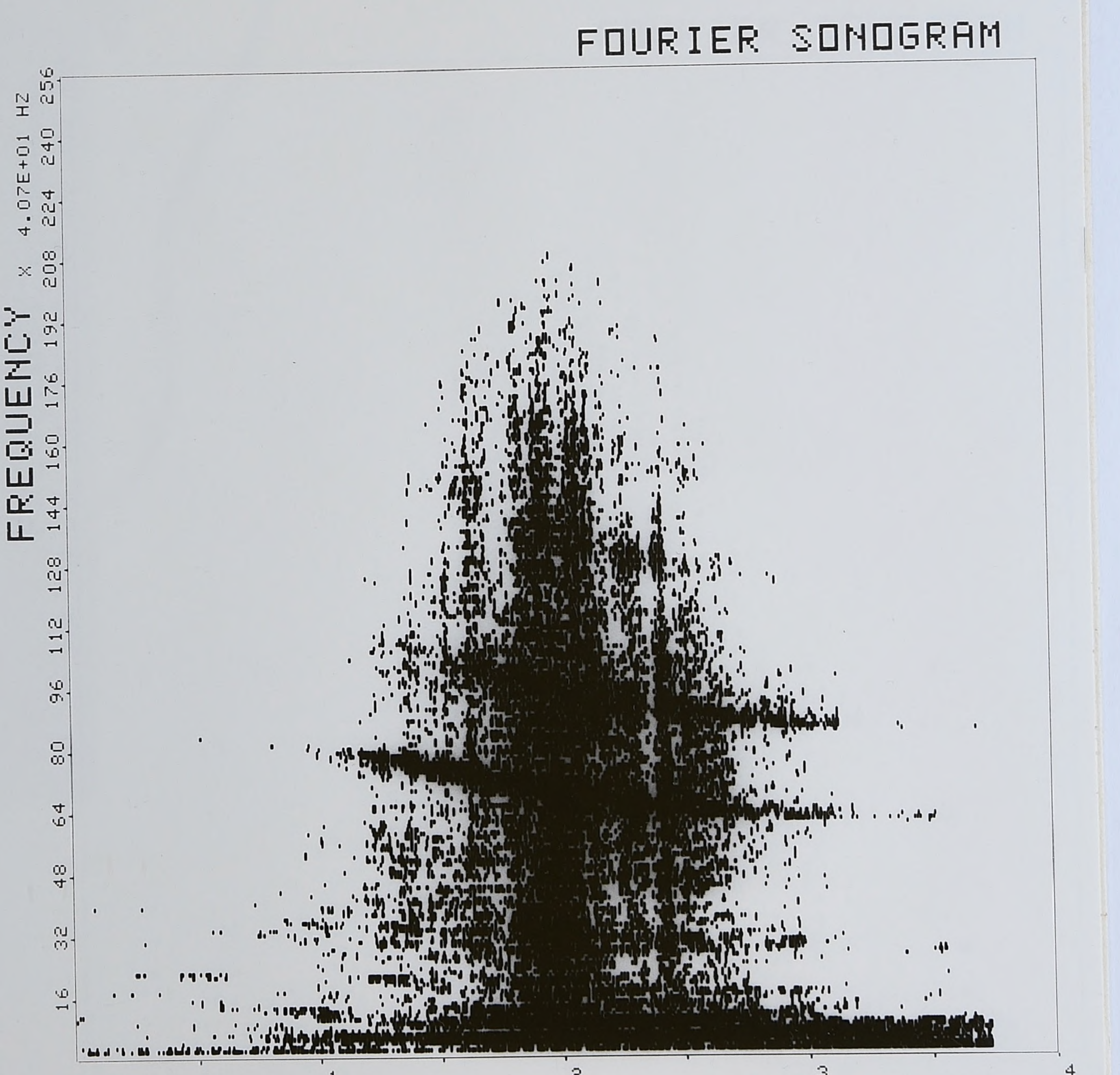
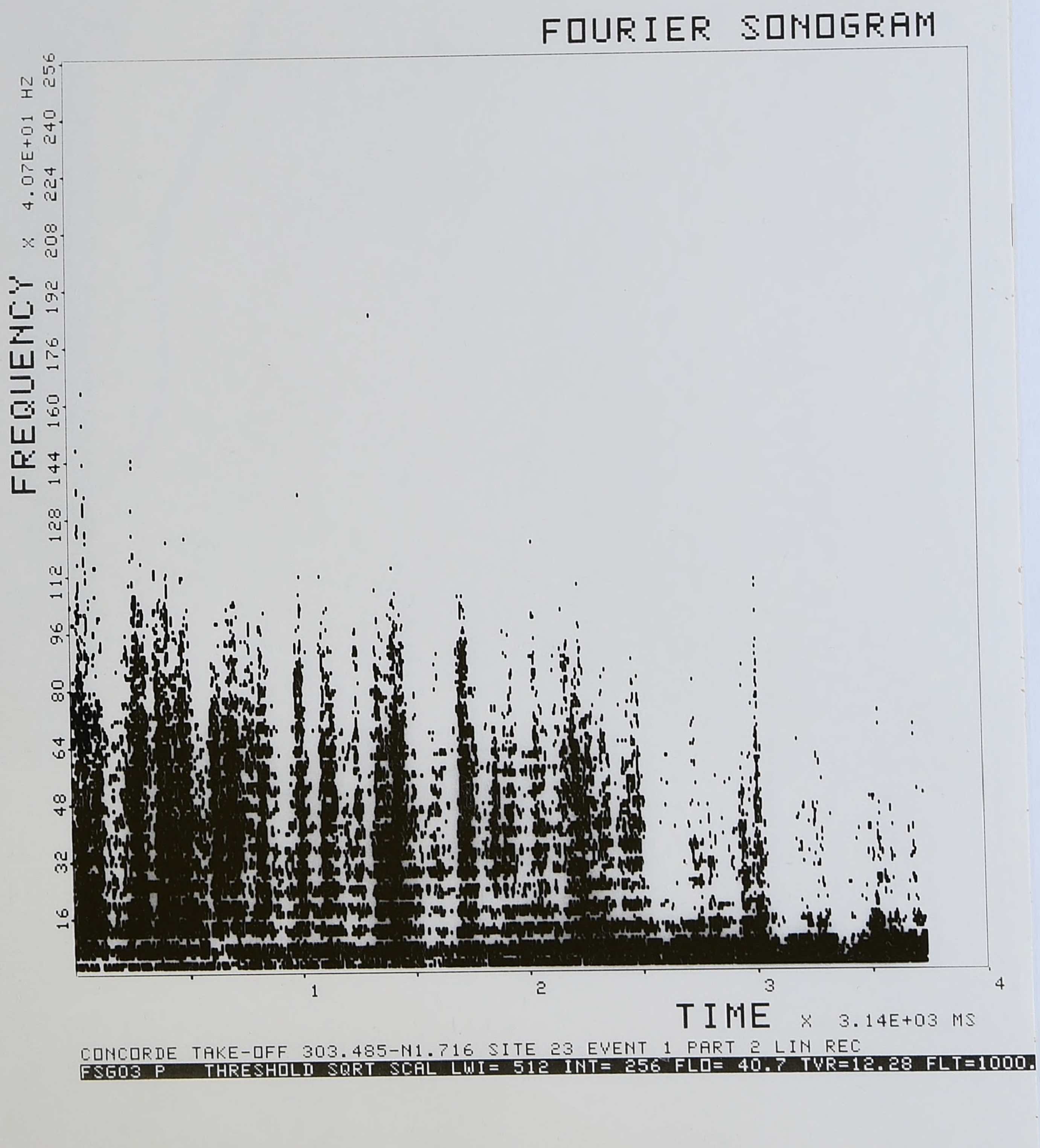


Figure 6.2.b. Binary spectrogram of jet taking off.
(parameters as in figure 6.1.)



<u>Figure 6.2.c.</u> Binary spectrogram of helicopter.
(parameters as in figure 6.1.)

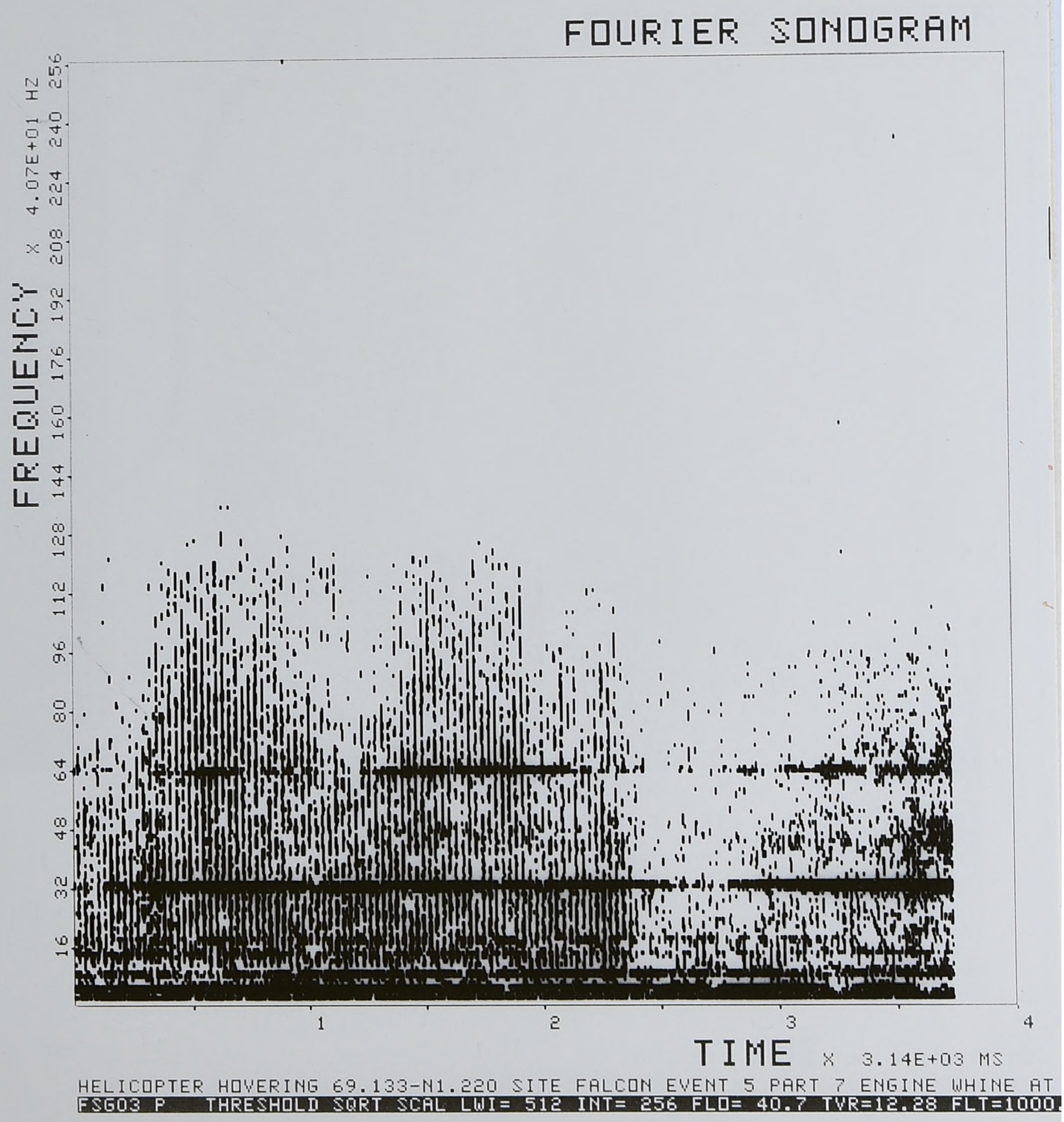
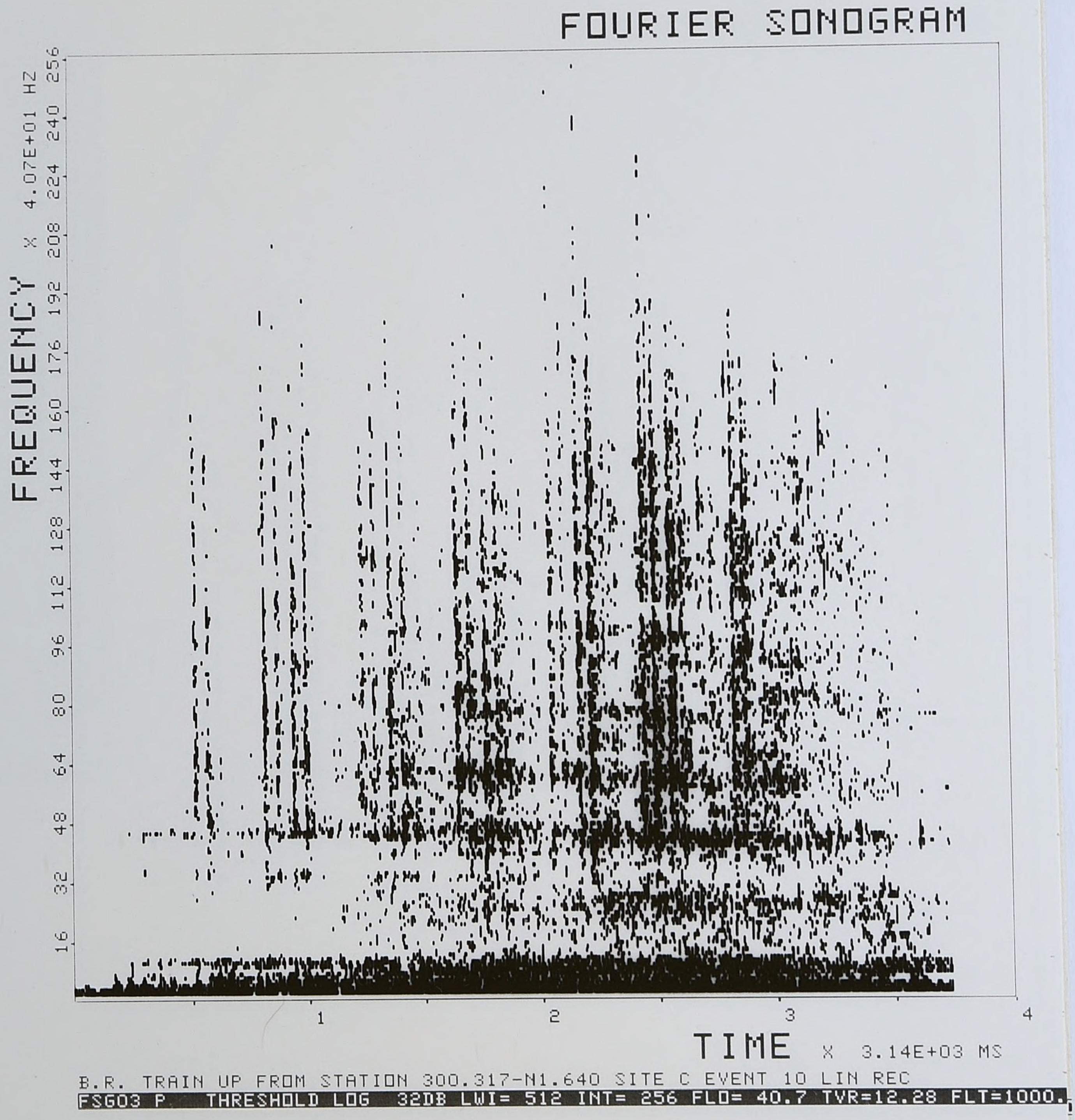


Figure 6.2.d. Binary spectrogram of train.
(parameters as in figure 6.1.)



CHAPTER 7

7 FEATURE DEFINITION & CLASSIFICATION

After the pattern analysis part of this work we now reach the stage of defining the features and designing the classifier whose task is threefold: to evaluate the features, select the "best" ones and finally perform the classification of unknown patterns. As we mentioned in chapter 2, the feature extraction stage is the most problem-oriented task of the design process and we have mostly concentrated on this. We shall not, therefore, be involved in the details of classifier design. Instead we shall examine the possibilities open for a suitable definition of candidate features in section 7.1. Section 7.2 gives an account of the evaluation and selection of the "best" feature set by using a linear classifier. Finally, the results are given and discussed in section 7.3.

7.1 Feature Definition

The properties of pollution sounds, as examined in chapter 4 and reflected in the spectrograms of the last chapter, both those due to the operating characteristics of the sources and those of the environment, suggest the features that are likely to discriminate the noise pollution classes. Firstly, the frequency bands where peaks are consistently exhibited through time suggest that these bands can be useful, as seen in figures 6.1 end 6.2 of the previous chapter. Secondly, the vertical pattern featured in helicopter spectrograms (figs. 6.1c and 6.2c) indicates this as a particular characteristic feature. Thirdly, the ripples in the spectra of jet aircraft (figs. 6.1a, 6.1b and 6.2a) are also characteristic for this class.

7.1.1 Alternative Features: Constraints

The problem of course lies in defining the characteristics mentioned above in terms of measurements that can be obtained from the spectrograms computed in the previous chapter. When looking for alternatives, the criteria for selection should be the instrument requirements presented and discussed in chapter 2, especially the requirement for on-line operation, limited store and the compatibility with the existing monitoring system. We may repeat here that these requirements prohibit any processing that would require extensive computation and the storage of a long part of the digitized acoustic signal, and that decisions about the noise type should be made every half second. The implication is that we have to dispense with information obtainable from any structure in the signal that extends over a half second interval. For example, although the structure (set of equidistant curves) in figure 6.1a is typical of the jets of the data base, its utilization becomes impossible under the circumstances mentioned above for it extends over the entire duration of the record. On the contrary, the structure of figure 6.1b is invariant over the duration of the record and it can be captured in a half second interval.

Considering this constraint, it was decided that the 12 second records of the data base should be divided into shorter <u>subrecords</u>. For technical reasons each record was divided into 26 subrecords of 0.45 seconds each. As noise pollution sounds cannot be expected to remain stationary for a duration more than that of a subrecord, we assumed that the subrecords were statistically independent. The only common characteristics of the subrecords would then be expected to be due to the generating source.

At first sight, what seem to be the measurements to be used as features are the series of spectra computed in the spectrogram. In each subrecord there are 36 spectra each of 512 spectral points. The number of independent spectral points, however, is 256 as mentioned in the previous chapter, covering a frequency range of

10,425 Hz. An observation of the spectrograms of figure 6.1 leads to the conclusion that there is no significant structure above 5 kHz. Therefore, it is justifiable to take the low 128 spectral components as candidates for the feature set, to cover the frequency range up to 5,212.5 Hz.

The utilization of the spectral estimates as features could then be achieved as follows: Each spectrum could be classified using the spectral estimates as features into one of a number of classes. The series of 34 classifications, represented as a string of symbols, could then be fed into a 'string' classifier in a form of either a template matcher or of a stochastic automaton. The idea of such a scheme is inspired from the speech recognition systems briefly mentioned in chapter 3 (section 3.4).

Certain problems, however, prohibit the realization of such a system. Classification of individual spectra presumes the existence of a finite number of classes of spectra and the possibility of defining these classes. Whereas in speech recognition systems the classes of phonemes is well defined and their number is finite, in the case of noise pollution there is no correspondence between perceptible 'elementary' noise 'phonemes' and spectra. Such a correspondence exists in speech enabling the labeling and classification of spectra and hence the extraction of prototypes or the definition of spectral classes. In addition, in speech, this procedure is aided by controlling the utterance of phrases (speed, speaker etc.) in anechoic chambers to avoid extraneous noises and so on. On the contrary, the noisy nature of noise pollution sounds and the impossibility to control the noise sources, apart form selecting the monitoring site, renders such a scheme extremely difficult.

It does not seem necessary to indulge more in arguments about the infeasibility of the scheme mentioned above. The following observation, however, alleviates the discouraging effects of the previous discussion. The purpose of a speech recognition system is to capture the sequence of phonemes so as to reconstruct the phrase being uttered. This is a lot more complicated than 'labeling' a segment of a noisy sound as, for example, 'helicopter' or 'train', etc. Thus, the search for a set of features should focus on capturing the 'whole' rather than describing the fine variations in the subrecord. Such an approach would also have the advantage of being simpler and hence faster so as to be implementable on a microprocessor.

Following this discussion, the next consideration is the definition of some average measures derived from the series of spectra comprising a subrecord.

7.1.2 Candidate Features: Definition and Extraction

In this section the average measures will be computed and the candidate features will be defined. The definitions will follow the symbolisms of the previous chapter on the computation of spectrograms.

7.1.2.1 Average Spectra

Let us define the average spectrum $\underline{X}_s$ of a subrecord as follows

$$\{ \underline{X}_s(l) \} = \{ {}^1/_{N_w'} \sum_{i=1}^{N_w'} X_s(i,l) \}, \quad l=1, \ldots ,N \tag{7.1}$$

where N_w' is the number of windows in a subrecord and X_s is the scaled spectrum as defined in relation (6.10). All persisting frequencies in the subrecord will be summed up in the summation of relation (7.1) and will contribute to a peak in the average spectrum. Minor changes in frequency of a particular tone will be tolerated by the bandwidth of the spectral components. On the other hand, all random changes due to phenomena such as wind or the random processes of the sources will be smoothed out by the summation. It must also be noted that the average spectrum is, generally, a better estimate of the spectrum of the source than the estimate obtained by a single transform (chapter 3, section 3.1.2.3).

Typical average spectra of jet aircraft, helicopter and train are shown in fig 7.1a, b and c, respectively. They were computed according to formula (7.1). The subrecords were taken from the records whose spectrograms are shown in figure 6.1b, c and d, respectively. It can be observed that the periodicities shown as horizontal lines in the spectrograms are preserved as peaks in the average spectra, and the noise has disappeared. Also worth noting is that the ripples due to ground reflection are preserved due to the fact that the pattern in figure 6.1b is nearly horizontal.

No average spectrum taken from the record of figure 6.1a is shown. The lines due to ground reflection are inclined and hence the rippled structure would vanish in the average spectrum. Hence, it was decided that no records of aircraft landing... should be used to evaluated the feature. This decision has a minor effect in the generality of the results, provided that the monitoring site is at the proper distance from the source. As discussed in the previous chapter, the differences between the two classes of spectrograms result from the distance of the sources from the microphone.

7.1.2.2 Average Cepstrum

The average cepstrum derived from the average spectrum by inverse D.F.T. may be defined as follows:

$$\{C(k)\} = \{1/_N \sum_{l=1}^{N} \underline{X}_s(l) \exp(j2\pi kl/N)\}, \quad k=1, \dots, N \qquad (7.2)$$

where $\underline{X}_s$ is the average spectrum defined in (7.1). No logarithm is included in the definition, as explained in chapter 2.

Average cepstra computed using the above definition are shown in figure 7.2a, b and c, derived from the average spectra of figure 7.1a, b and c, respectively. The jet cepstrum of fig.7.2a exhibits a pronounced peak as a result of the ripples caused by ground reflection (chapter 4). The peak for the jets of the data base is in the range 3.8 to 5.8 ms.

7.1.2.3 Average Amplitude and 'Pseudospectrum'

The candidate features defined so far are average measures representing the horizontal patterns observed in the spectrograms. However, the vertical patterns must also be utilized, especially the one featured in the helicopter spectrogram. The vertical lines in figure 6.1c are caused by impulses of increased power resulting from the interaction of the main rotor blades with the air. The fundamental frequency of the impulses is lower than the frequency resolution of the spectrograms. Thus, the impulses cannot be captured as horizontal lines in the spectrogram. Instead, when an impulse is included in a window, its contributions consists in raising the overall power in the corresponding spectrum. The power of the impulse is spread over the entire spectrum giving the characteristic vertical line pattern in the helicopter spectrogram.

The sequence v_a defined as:

$$\{v_a(i)\} = \{ \sum_{l=1}^{N} |X_s(i,l)|\}, \quad i=1, \ldots, N_W' \qquad (7.3)$$

represents the variation with time of the <u>average amplitude</u> of the subrecord. The average amplitude (which is the square root of the average power) is shown in figure 7.3a, b and c and 7.4a, b, and c for the records of figure 6.1b, c and d, respectively.

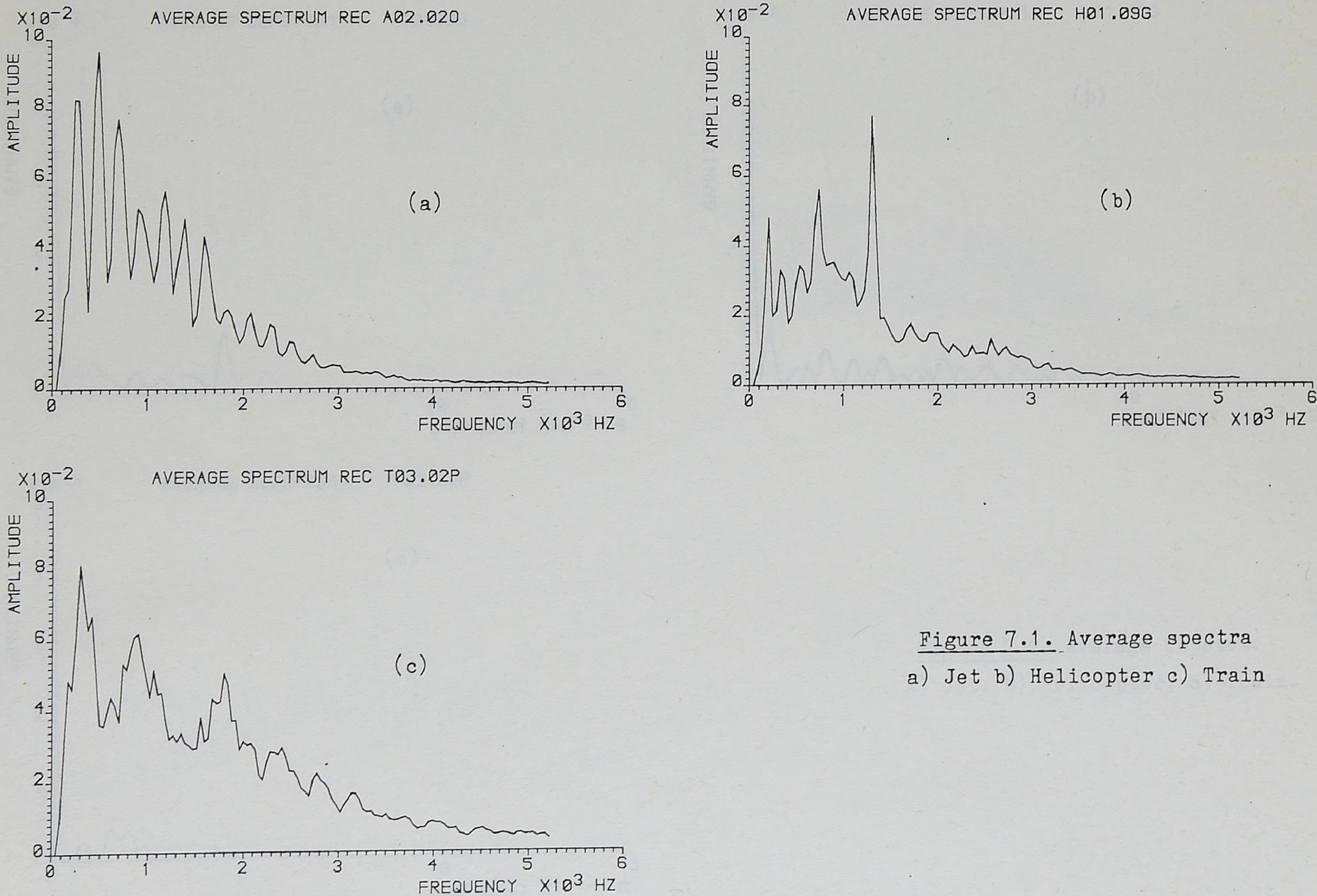
Note that whereas figures 7.4 show the variation for the duration of a subrecord (one for each class), in figures 7.3 the variation is shown over the entire duration of the record in order to emphasize the differences between the three classes. Clearly, the helicopter average amplitude of figures 7.3b and 7.4b varies in a periodic manner corresponding to the periodic repetition of the vertical lines displayed in the spectrogram. Similarly, the vertical patterns in the train spectrogram, caused by the wheel/rail knocks, are captured in the average amplitude. Unfortunately, not all trains feature this pattern which is anyhow difficult to extract, unlike the helicopter pattern, as will be shown below. As far as jets are concerned, no structure is seen in the average amplitude except random variations due to wind effects.

The periodicities in the average amplitude of the helicopter can be extracted by taking the D.F.T. of the sequence of (7.3):

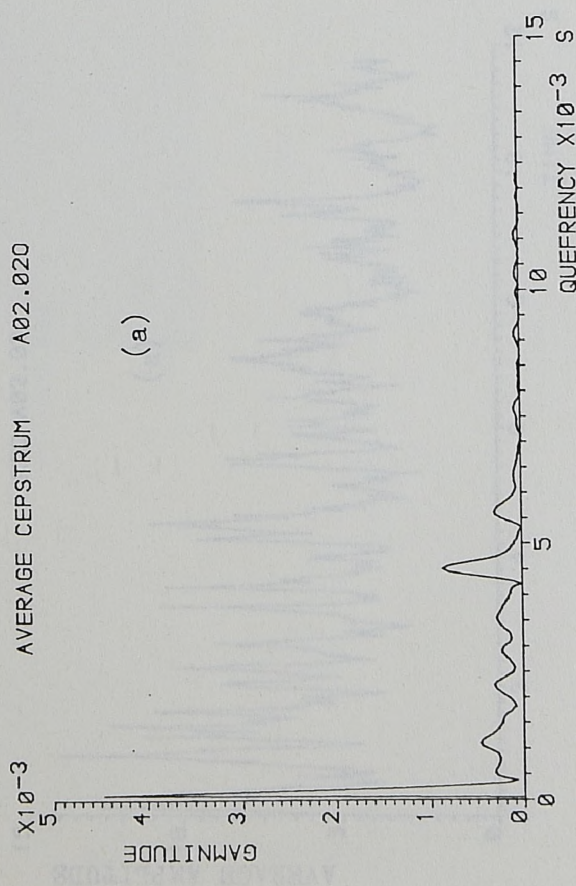
$$\{V_a(k)\} = \{{}^1/_{N''_W} \sum_{l=1}^{N''_W} v_a(l)\ \exp(-j2\pi kl/N''_W)\},\ k=1,\ldots,N'_W \tag{7.4}$$

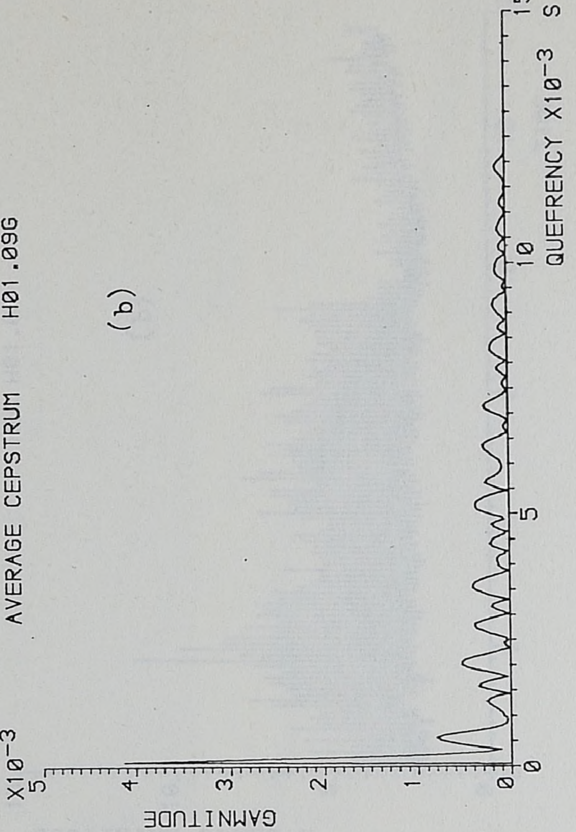
where N''_W is the smallest power of 2 greater then N'_W, in this case 64, the sequence v_a having been padded up with zeros. The result of the transformation, the sequence V_a is a <u>pseudospectrum</u> of the original sound signal. It is, in fact, the spectrum of a smoothed version of it. As the spacing between successive samples of v_a is 12.2 ms (the time interval between successive windows) the interval between successive samples of V_a is 1.23 Hz, for a sequence of 64 samples (section 3.1.2.1).

Figures 7.5a, b and c show pseudospectra derived from the average amplitude sequences of figure 7.3a, b and c, respectively. It can be seen that in fig. 7.5b, derived from the helicopter sound, a fundamental peak and its harmonics are featured. The fundamental peak is featured in the range 10 to 13



<u>Figure 7.1.</u> Average spectra
a) Jet b) Helicopter c) Train





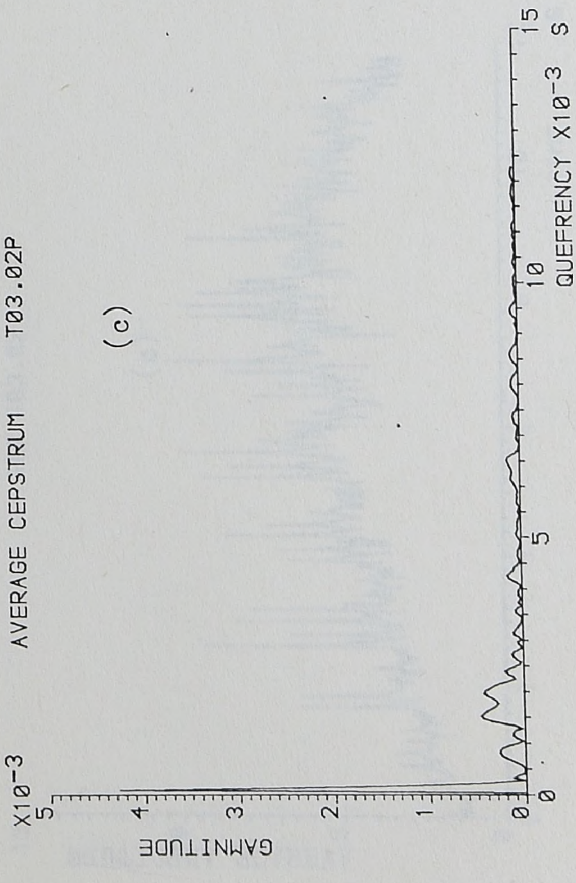
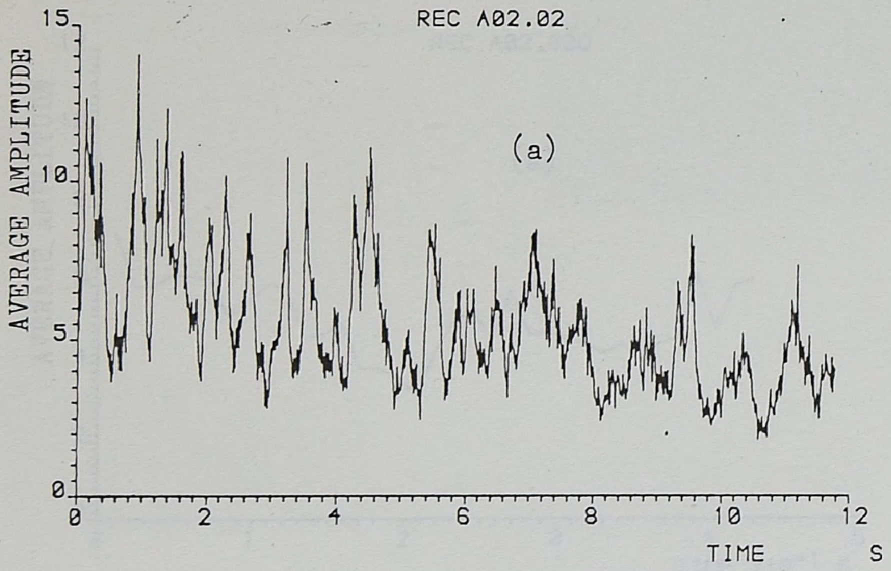
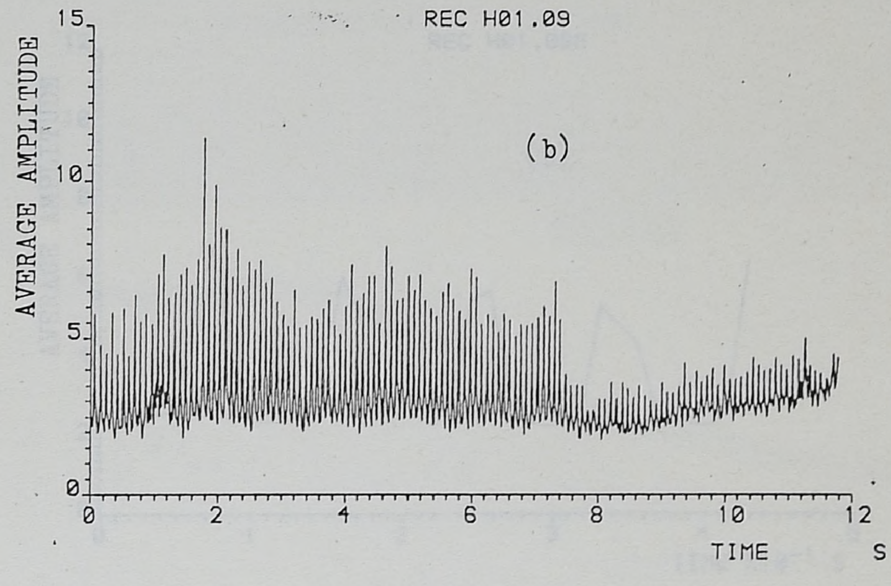
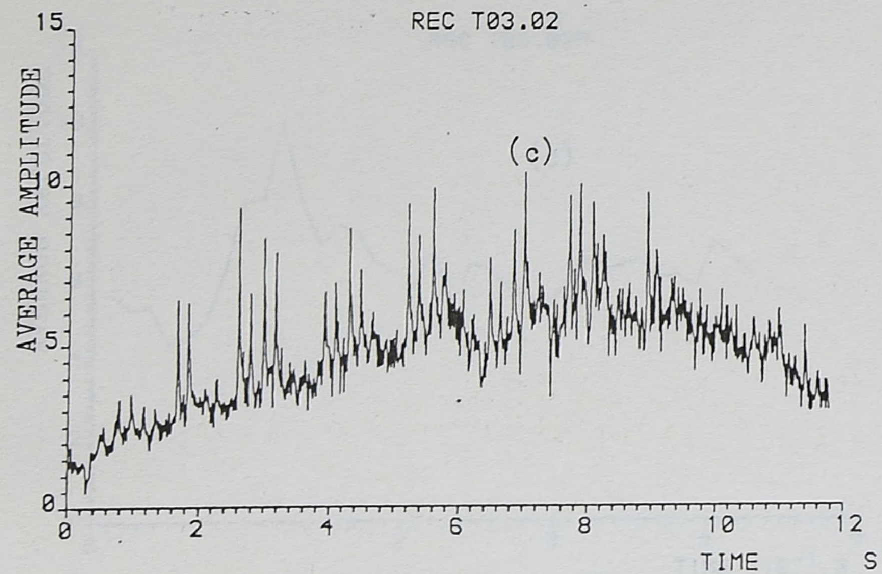


Figure 7.2. Average Cepstra.
a) Jet. b) Helicopter. c) Train.







<u>Figure 7.3.</u> Variation of average amplitude with time (12 s)

a) Jet b) Helicopter c) Train

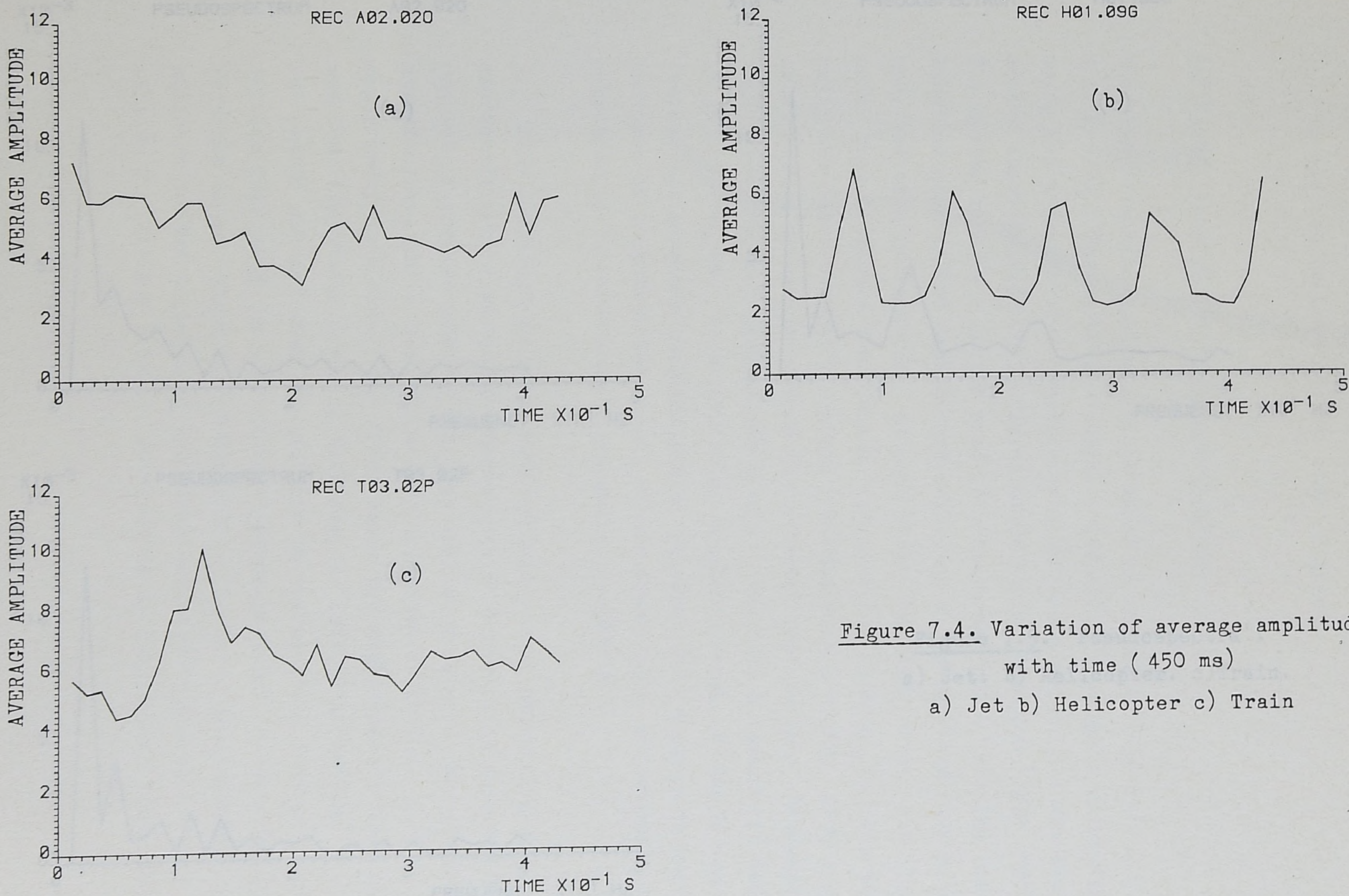


Figure 7.4. Variation of average amplitude with time (450 ms)
a) Jet b) Helicopter c) Train

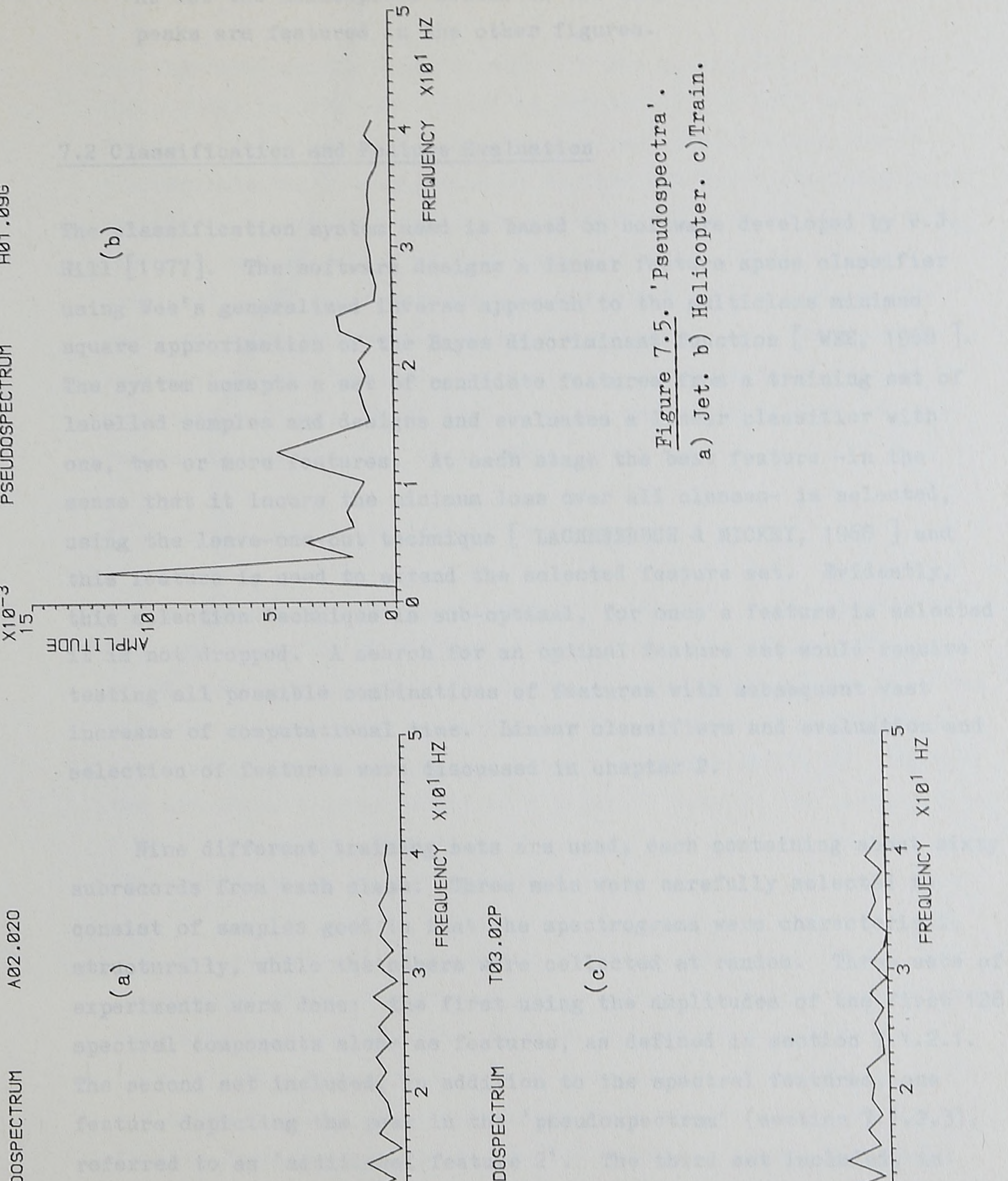


Figure 7.5. 'Pseudospectra'.
a) Jet. b) Helicopter. c) Train.

Hz for the helicopters sounds of the data base. No pronounced peaks are featured in the other figures.

7.2 Classification and Feature Evaluation

The classification system used is based on software developed by W.J. Hill [1977]. The software designs a linear feature space classifier using Wee's generalized inverse approach to the multiclass minimum square approximation of the Bayes discriminant function [WEE, 1968]. The system accepts a set of candidate features from a training set of labelled samples and designs and evaluates a linear classifier with one, two or more features. At each stage the best feature -in the sense that it incurs the minimum loss over all classes- is selected, using the leave-one-out technique [LACHENBRUCH & MICKEY, 1968] and this feature is used to extend the selected feature set. Evidently, this selection technique is sub-optimal, for once a feature is selected it is not dropped. A search for an optimal feature set would require testing all possible combinations of features with subsequent vast increase of computational time. Linear classifiers and evaluation and selection of features were discussed in chapter 2.

Nine different training sets are used, each containing about sixty subrecords from each class. Three sets were carefully selected to consist of samples good in that the spectrograms were characteristic structurally, while the others were collected at random. Three sets of experiments were done: the first using the amplitudes of the first 128 spectral components alone as features, as defined in section 7.1.2.1. The second set included, in addition to the spectral features, one feature depicting the peak in the 'pseudospectrum' (section 7.1.2.3), referred to as 'additional feature 2'. The third set included, in addition to the features of the first set, both features defined in section 7.1.2.2 and 7.1.2.3, referred to as 'additional features 1 and 2'.

7.2.1 Evaluation of Spectral Features

The objective of the first experiments was to establish whether a) the average spectra (as defined in relation (7.1) and shown in figure 7.1) of the subrecords alone, after normalization, could discriminate the classes; and b) to find out which frequency bands provide best discrimination. In addition, in view of the eventual hardware implementation of the classifier, the effect of modification of parameters such as the width of the bands on the performance of the classifier was investigated. The fundamental bandwidth of the spectral components was approximately 41Hz, as mentioned earlier. These were averaged in groups of 2 up to 12 to realise bandwidths from 41 to 488Hz.

7.2.2 Evaluation of Additional Features

The second set of experiments was to see whether incorporation of the cepstra and the variation of the average amplitude as features could improve the performance of the classifier significantly. The cepstrum was evaluated by taking the inverse Fourier transform of the average spectrum of each subrecord (definition (7.2)). Most of the aeroplane records exhibited a cepstral peak in the range 3.8 to 5.8ms. The average "amplitude" -gamnitude in cepstral terms- in this range was to be the cepstral feature.

The variation of the average amplitude with time (fig. 7.4) was thought to be another promising feature for noises that exhibit impulsive slow periodicities such as the helicopter blade noise (fig. 7.4b). For each subrecord the sequence of definition (7.3) was normalised, appended with zeros to form a 64 sample window and finally Fourier transformed. As the interval between successive samples of the function was 12.2ms the frequency components were 1.28Hz apart. Helicopter noises showed a peak in the range 10 to 13Hz and the average "amplitude" in this range constituted the corresponding feature.

7.3 Results

The scatter diagrams of figure 7.6 show how well the three classes are separated even when two features only are used. The two features in figure 7.6a are the amplitudes of the 33rd and 77th spectral components of the average spectra. It may be observed that the 33rd component (at about 1350 Hz) separates helicopters from the other two classes as the helicopters feature a peak in this frequency range (see figure 7.1b). Also, component 77 (at about 3200 Hz) separates the trains from the other classes for the same reason. This feature is also used in the diagram of figure 7.6b. The other feature in this diagram is the peak range in the Fourier transform of the average amplitude (additional feature 2). As expected, the latter separates the helicopters.

The graphs of figure 7.7 show the performance of the classifier as a function of the number of features. The curves formed by triangles show performance using spectral components only as features, those formed by squares show the performance when additional feature 2 is included. The x's show performance with all features, spectral and additional. The curves are drawn separately for the the 12 spectral bandwidths and are the average results of the nine training/testing sets.

7.3.1 Performance of Spectral Features

It is worth noting that, for bandwidths less than 400 Hz, recognition better than 80% was achieved with two features only. Recognition better than 95% was achieved using 6 spectral features only when the bandwidths are less than 290 Hz. When more than 6 features are used, the performance increases by a small percentage.

The frequency bands selected in each experiment were not always the same. The histograms shown in figures 7.8 depict the number of times a frequency band was selected. The histograms are drawn separately for each of the 12 bandwidths and give a 'quality' ranking to the frequency bands. A band selected in all experiments ranks 9, the highest rank. For small and medium bandwidths (figures 7.8a and b) the useful frequency range appears to extend up to 4 kHz. The highest ranking frequency bands are at about 1350 Hz, where the helicopter noise exhibits a peak in the average spectrum (figure 7.1b). For narrow bandwidths, the 250 Hz band is also high-ranking. This is where the jets exhibit a peak (figure 7.1a). However, the ranking of this peak decreases with increasing spectral bandwidth. The band at about 3200 Hz, where the trains exhibit a peak is also important, as well as the range between 200 and 2800 Hz.

For wide bandwidths, all bands are, more or less, selected without any significant preference, since the number of bands decreases (there are only nine bands for the widest bandwidth). Although for small bandwidths it is easy to explain the selection of certain bands (it can be attributed to some peak in the spectra), this becomes increasingly difficult as the bandwidth increases. The importance of narrow peaks diminishes, in the latter case, unless they are very pronounced, as is the helicopter peak at 1350 Hz.

The dependence of performance on the bandwidth of the spectral estimates is shown in figure 7.11a, where 8 curves are drawn for classification with 2, 3, 4 and so on, features. Overall, the performance drops as the bandwidth increases. The observed instability of the curves decreases with the number of features. It is interesting to note that the curves for 6 or more features show a local peak at the 250 Hz bandwidth. This peak is only a few percent lower than the maximum performance that is achieved with the narrowest bandwidth. It must be noted that the performance of narrow bandwidth spectral estimates should be assessed with some scepticism. It is very difficult to extrapolate their behaviour on unseen data, for any slight variations in the positions of the spectral peaks and valleys would cause misclassification. Therefore, the 250 Hz bandwidth must be considered optimum.

7.3.2 Performance of Additional Features.

When the additional features were included, the performance of the classifier always improved, as shown in the graphs of figure 7.7. The best performance is obtained with both additional features. The degree of improvement increases with the bandwidth. It is as high as 7% for the 450 Hz bandwidth, but only 2% for smaller ones. The improvement of performance starts early in the selection procedure, as seen in the same graphs.

It is impressive that the additional features are among the highest ranking ones, as shown in the histograms of figures 7.9 and The inclusion of the additional features had a negligible effect on the selection of the spectral bands. The bands selected were almost the same as in the first set of experiments.

Although the additional features ranked almost the same, their discriminating properties were different. Additional feature 1 (the 'cepstral feature') helped in decreasing the confusion of jet noises, whereas the additional feature 2 (the 'pseudospectral feature') discriminated helicopter noises.

The most interesting property of these features, when both are included in the feature set, is that they render the dependence of recognition on bandwidth negligible, as seen in figure 7.11b. (This also applies, to a lesser extend, for the 'pseudospectral feature', when it is included on its own (figure 7.11c.) This fact, allows some freedom in selecting the spectral bandwidths in the implementation stage of the recognition system. Wider bandwidths are easier to implement; they also preferable because they allow more tolerance the position of the spectral peaks. Nevertheless, it is doubtful whether the slight improvement in performance and the freedom to select wider bandwidths compensates the increased computation and complexity of the recognition system that is required.

7.4. Summary

This chapter was the presentation of the definition, extraction and evaluation of features for the classification of three classes of noise pollution sounds (jet aircraft, helicopters and trains). The records of the data base were split into subrecords of 0.45 seconds. The subrecords were used to train and evaluate the classifier. The candidate features were a) the average spectra of the subrecords, b) the average 'amplitude' of the average cepstra in the range 3.8 to 5.8 and c) the average 'amplitude' of the 'pseudospectra' in the range 10 to 13 Hz. Classification of better than 95% was achieved with 6 spectral features from the first candidate set, for spectral bandwidths of 41 up to 300 Hz. This figure was improved by 2%, on the average, when the features b) and c) above were included. The influence of the bandwidth of the spectral components was also examined and it was found that the bandwidth of 240 Hz is optimum.

The next chapter describes a suggested scheme for the implementation of a recognition system based on the findings of the simulation.

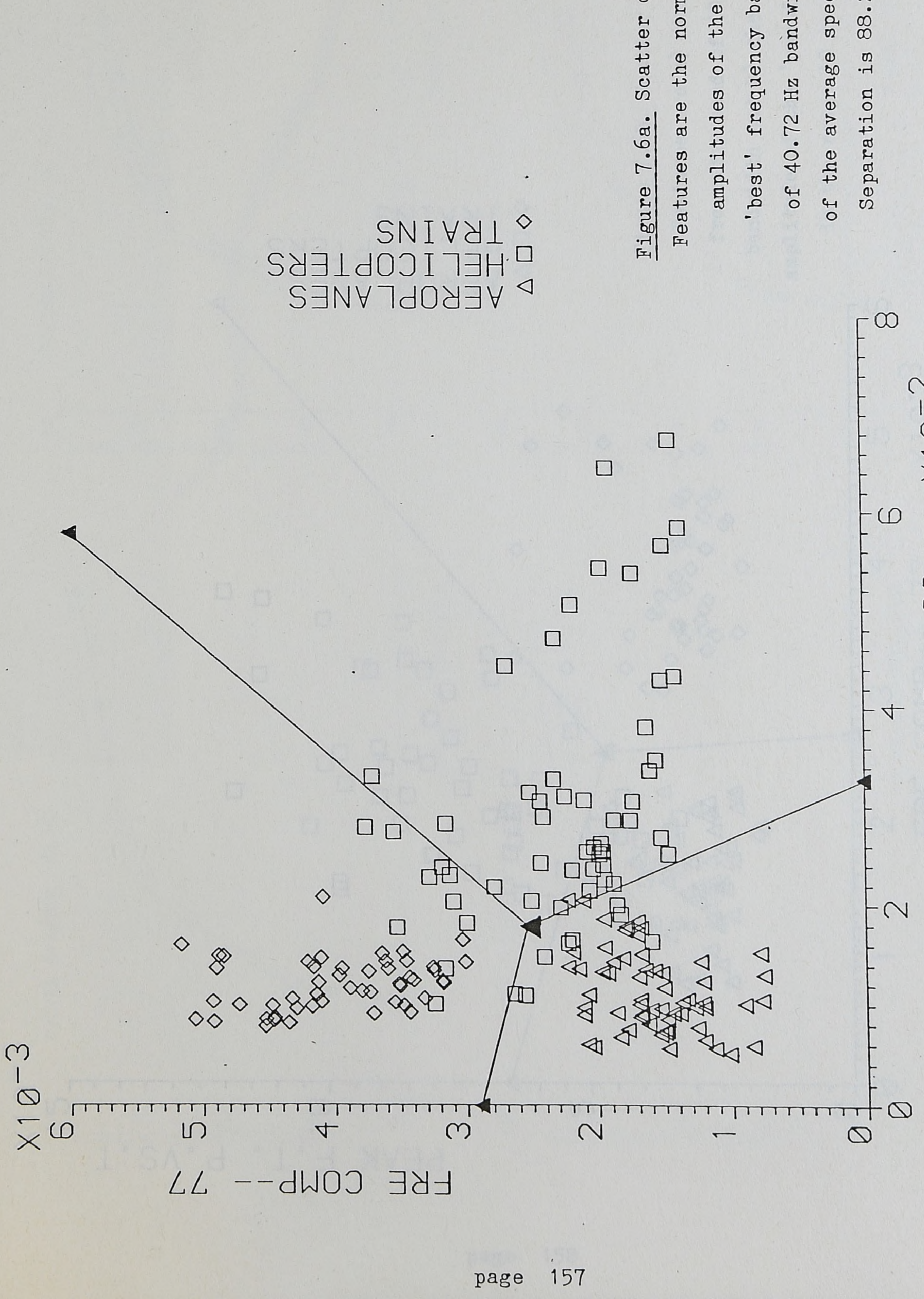


Figure 7.6a. Scatter diagram.
Features are the normalized amplitudes of the two 'best' frequency bands of 40.72 Hz bandwidth of the average spectrum
Separation is 88.2%.

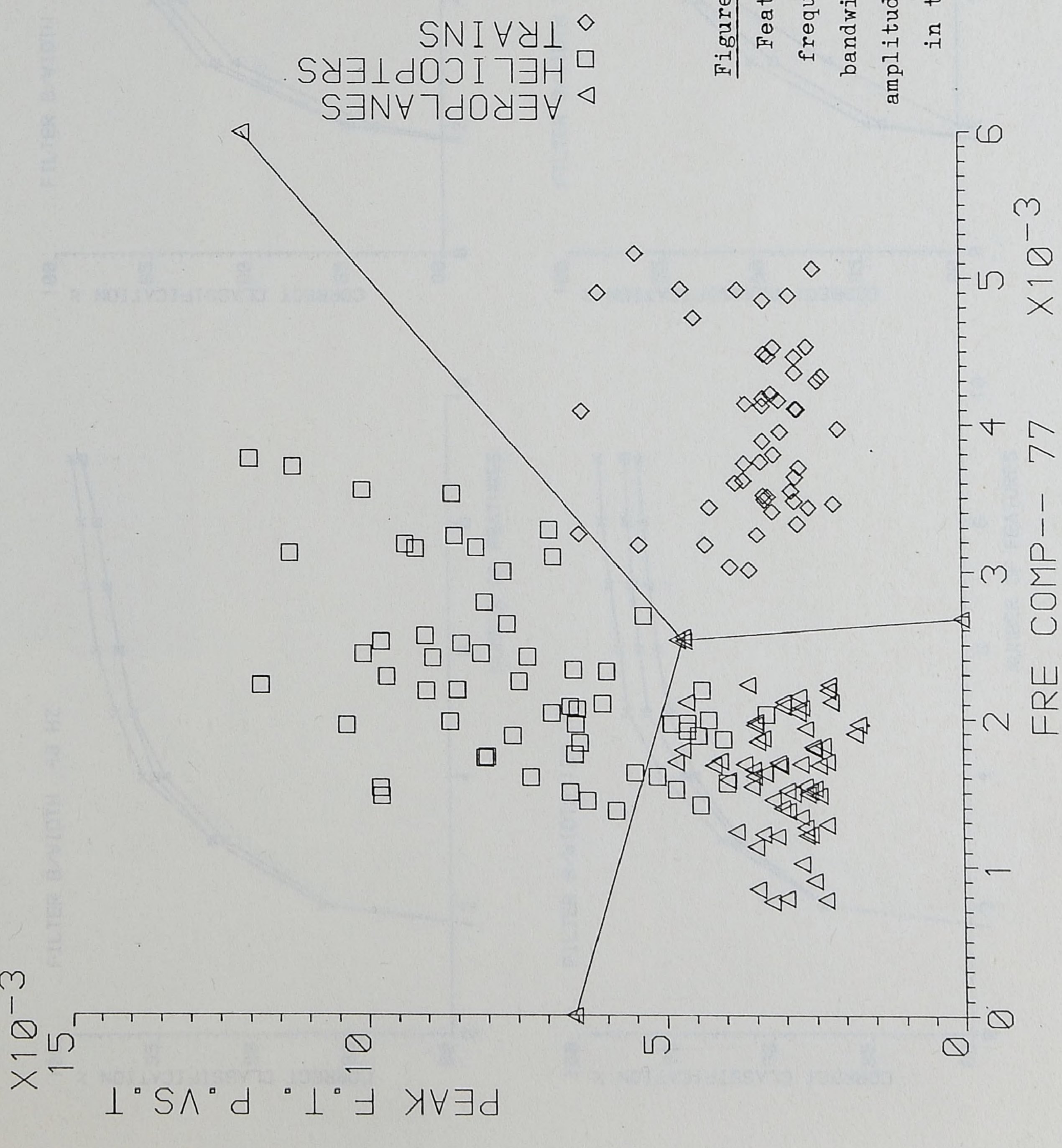


Figure 7.6b. Scatter diagram.
Features are a) the 77th frequency band of 40.72 Hz bandwidth and b) the average amplitude of the 'pseudospectrum' in the range 10 to 13 Hz.

<u>Figure 7.7.a.</u> Classifier performance.

(Δ spectral, □ spectral & 2, × spectral & 1 & 2 features)

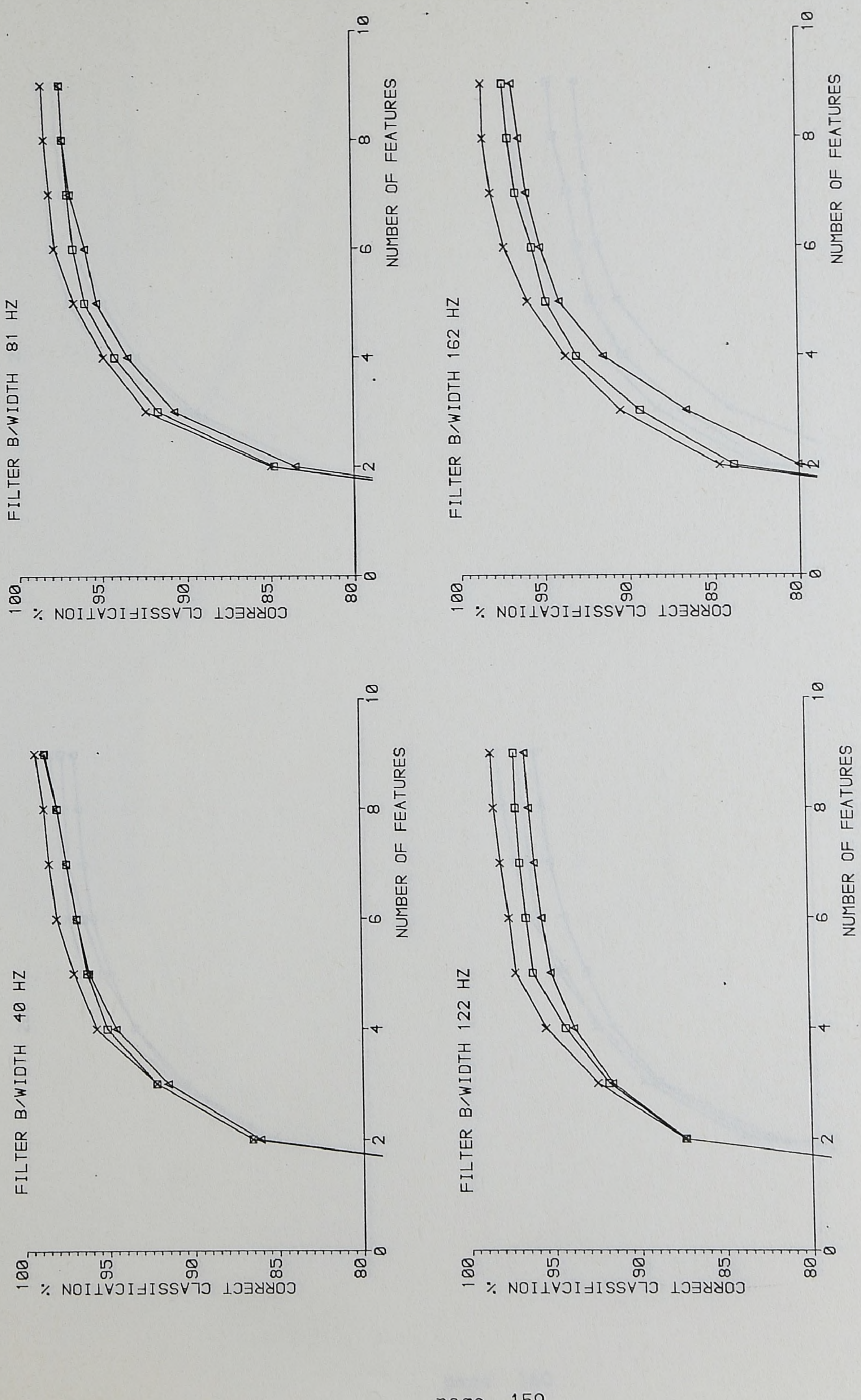


Figure 7.7b. Classifier performance.

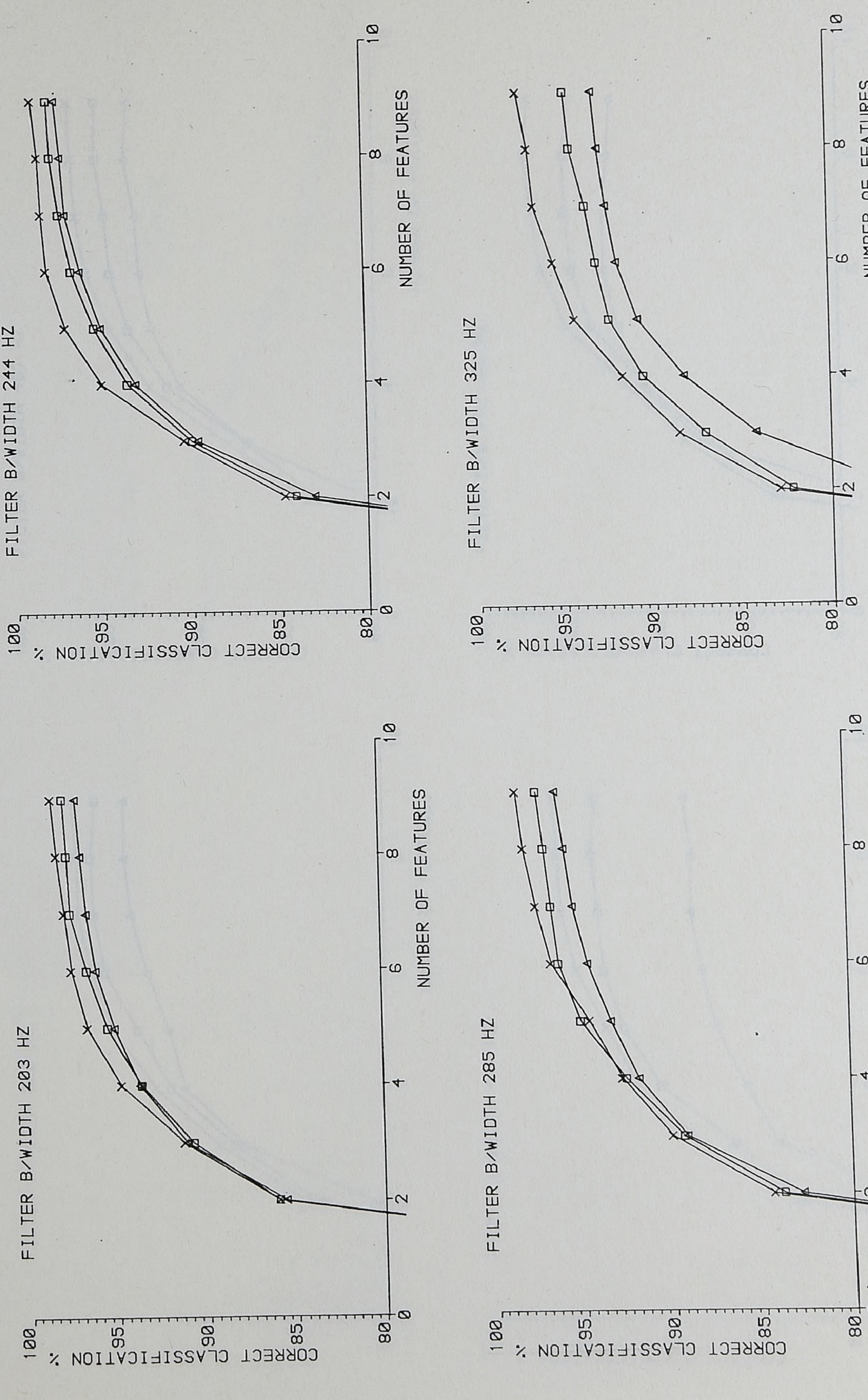


Figure 7.7c. Classifier performance.

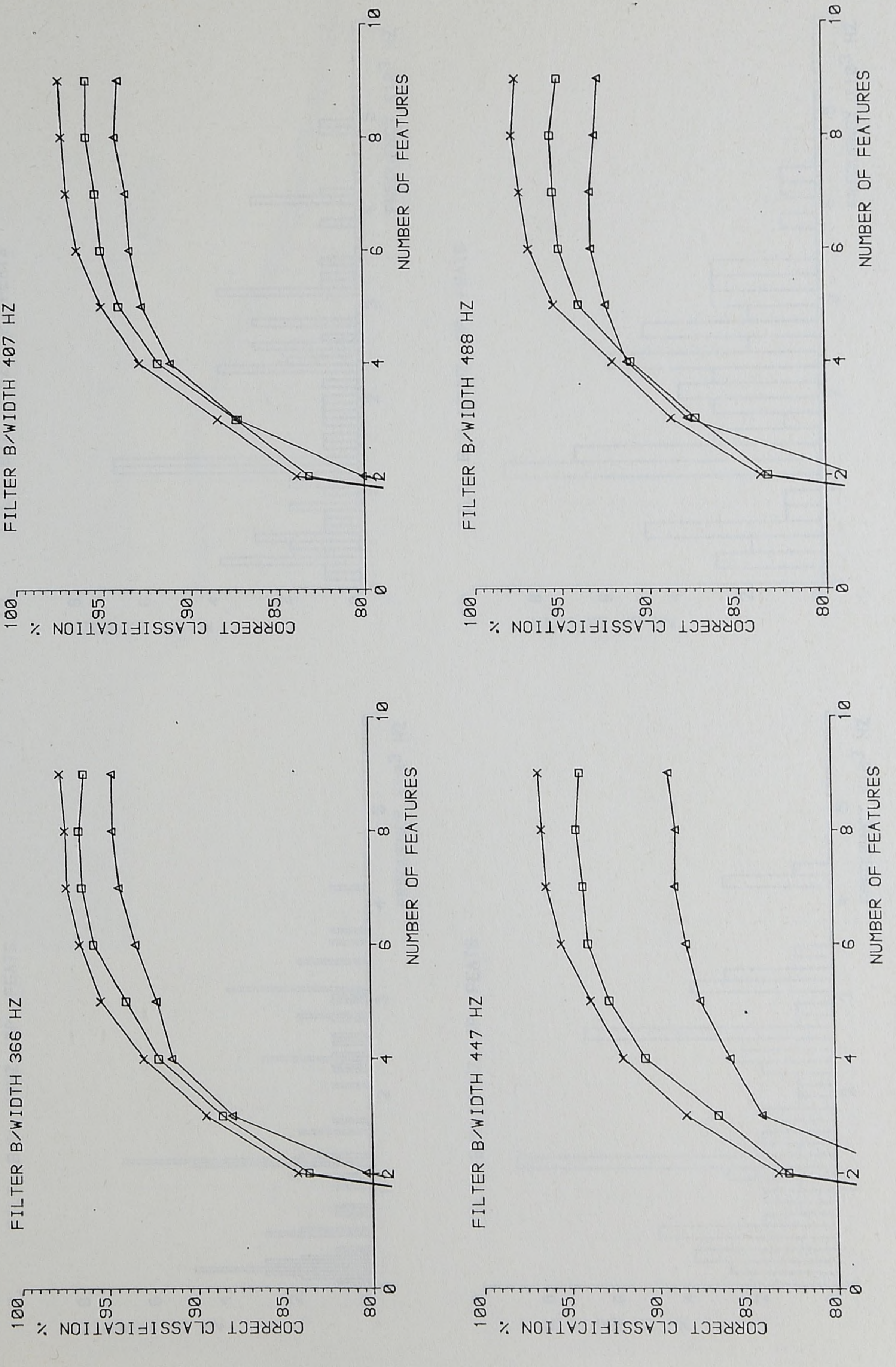


Figure 7.8a. Selected Frequency Bands.

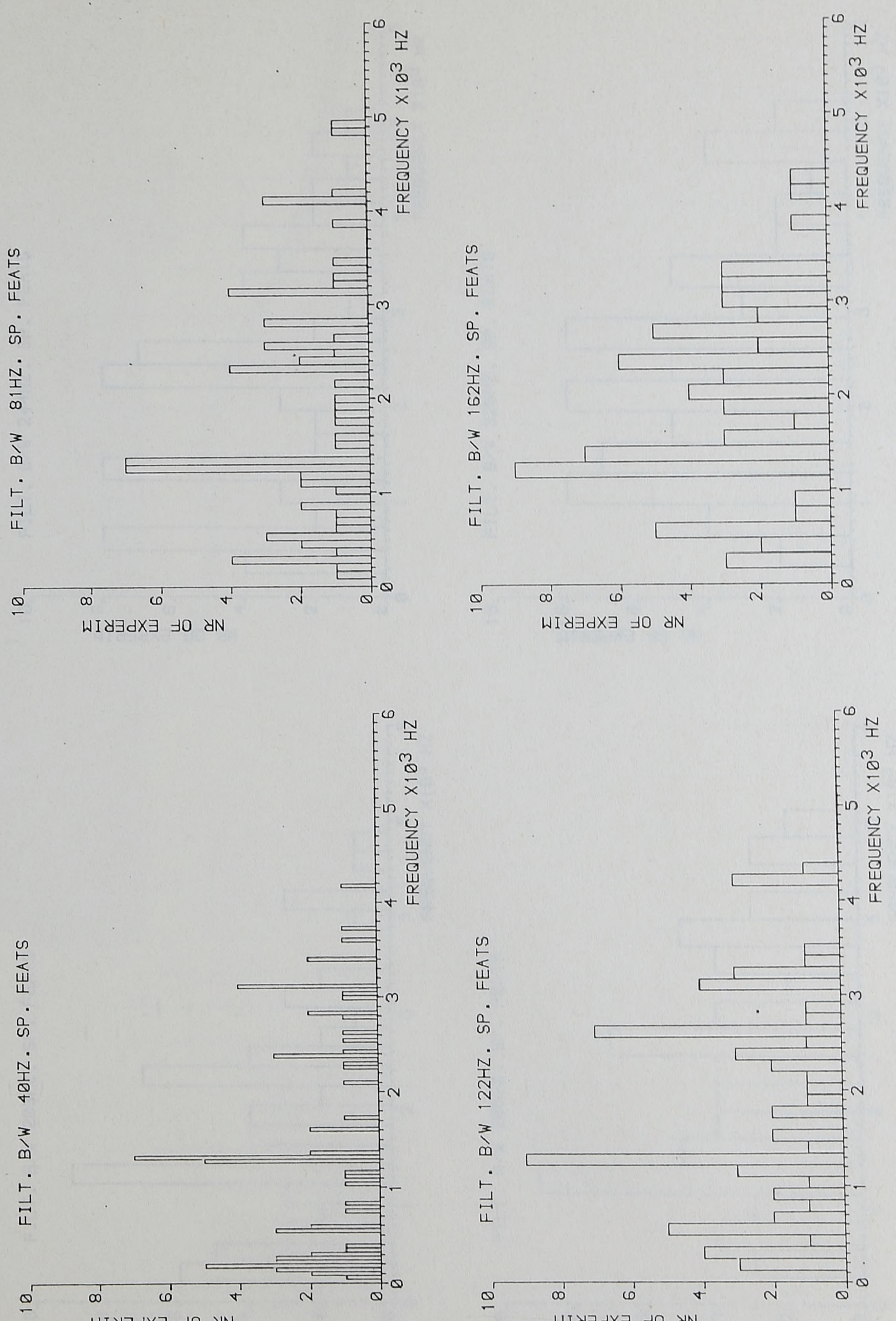


Figure 7.8b. Selected Frequency Bands.

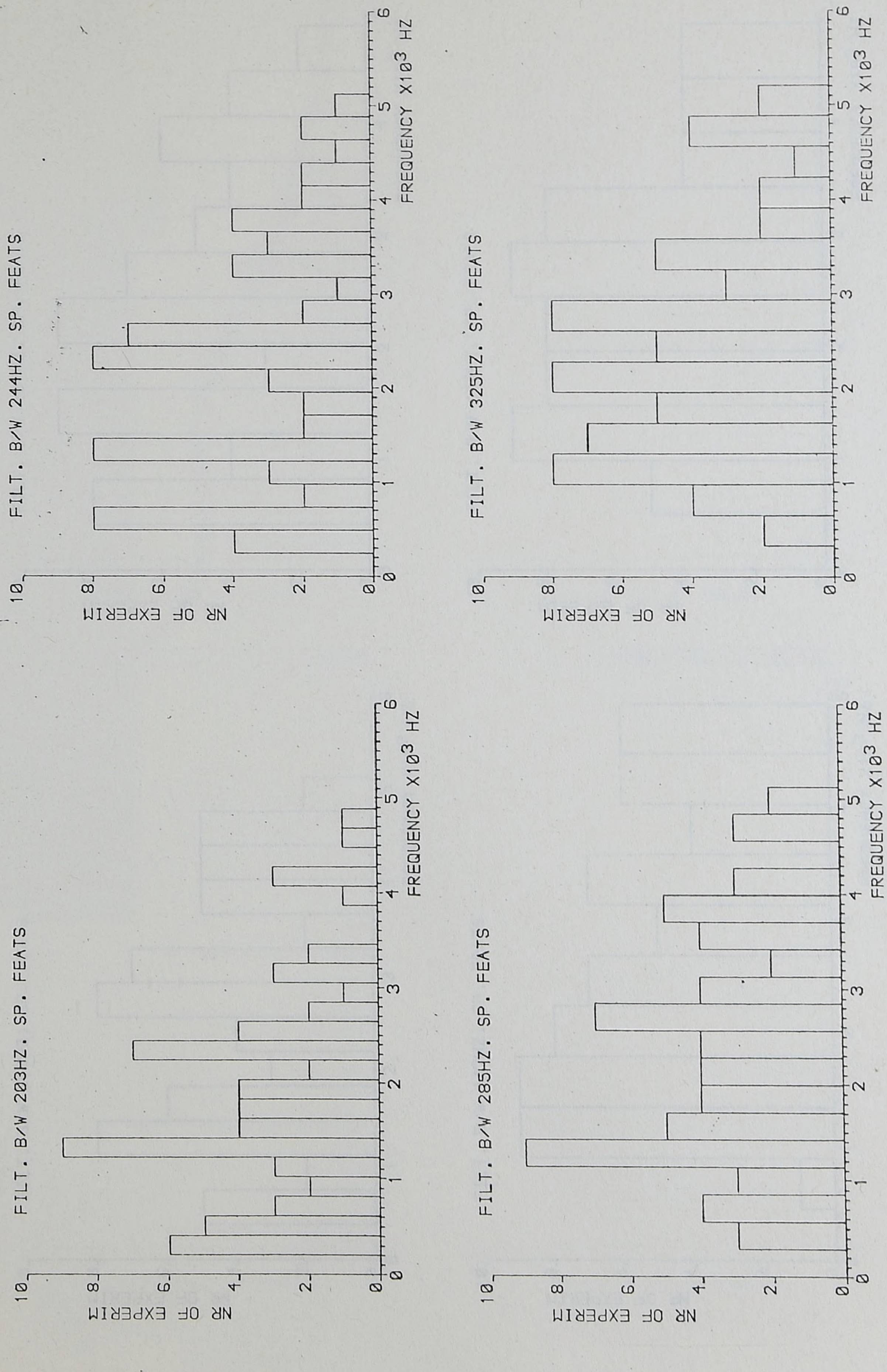


Figure 7.8c. Selected Frequency Bands.

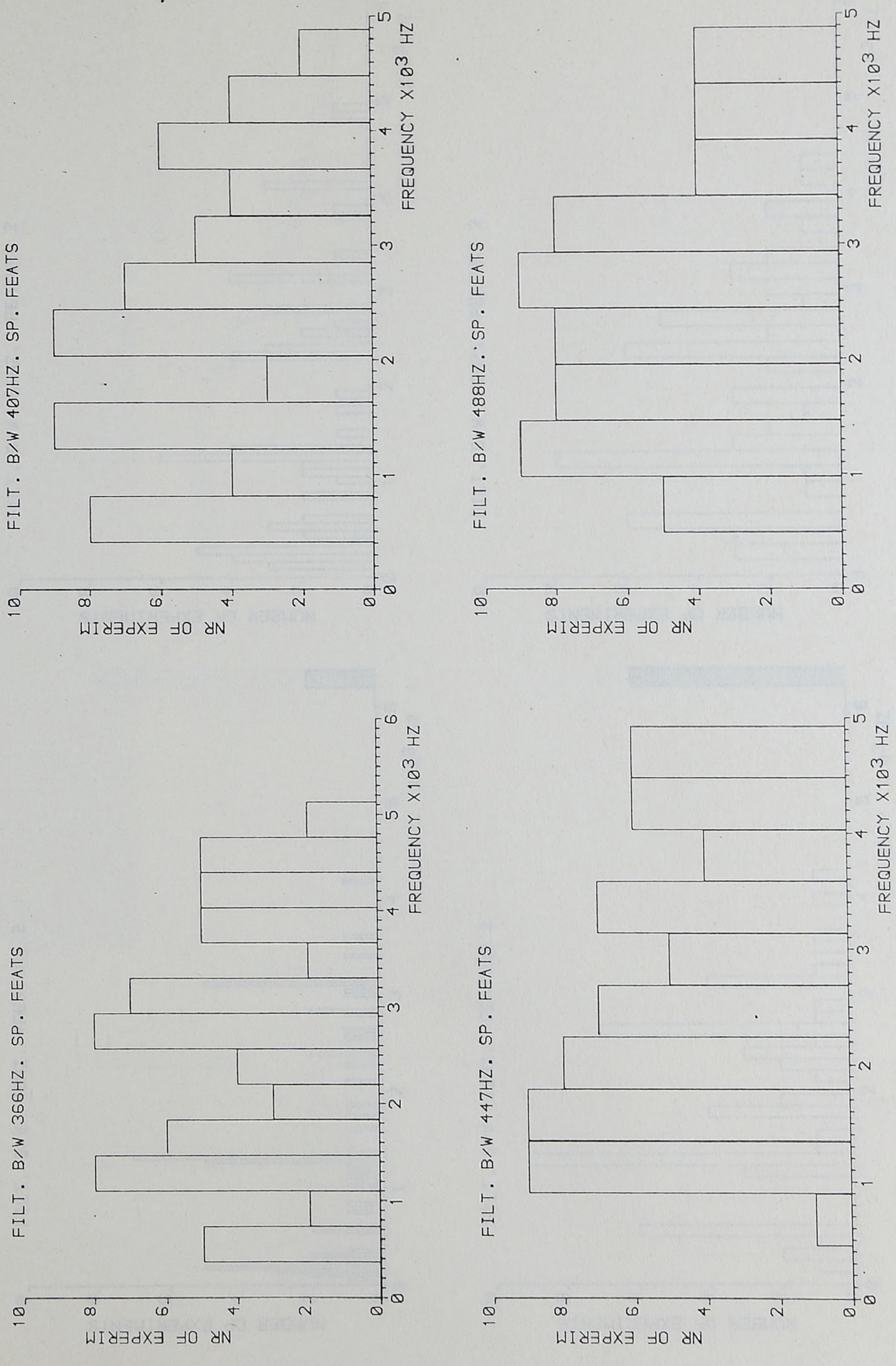


Figure 7.9a. Selected frequency bands & additonal feature 2.

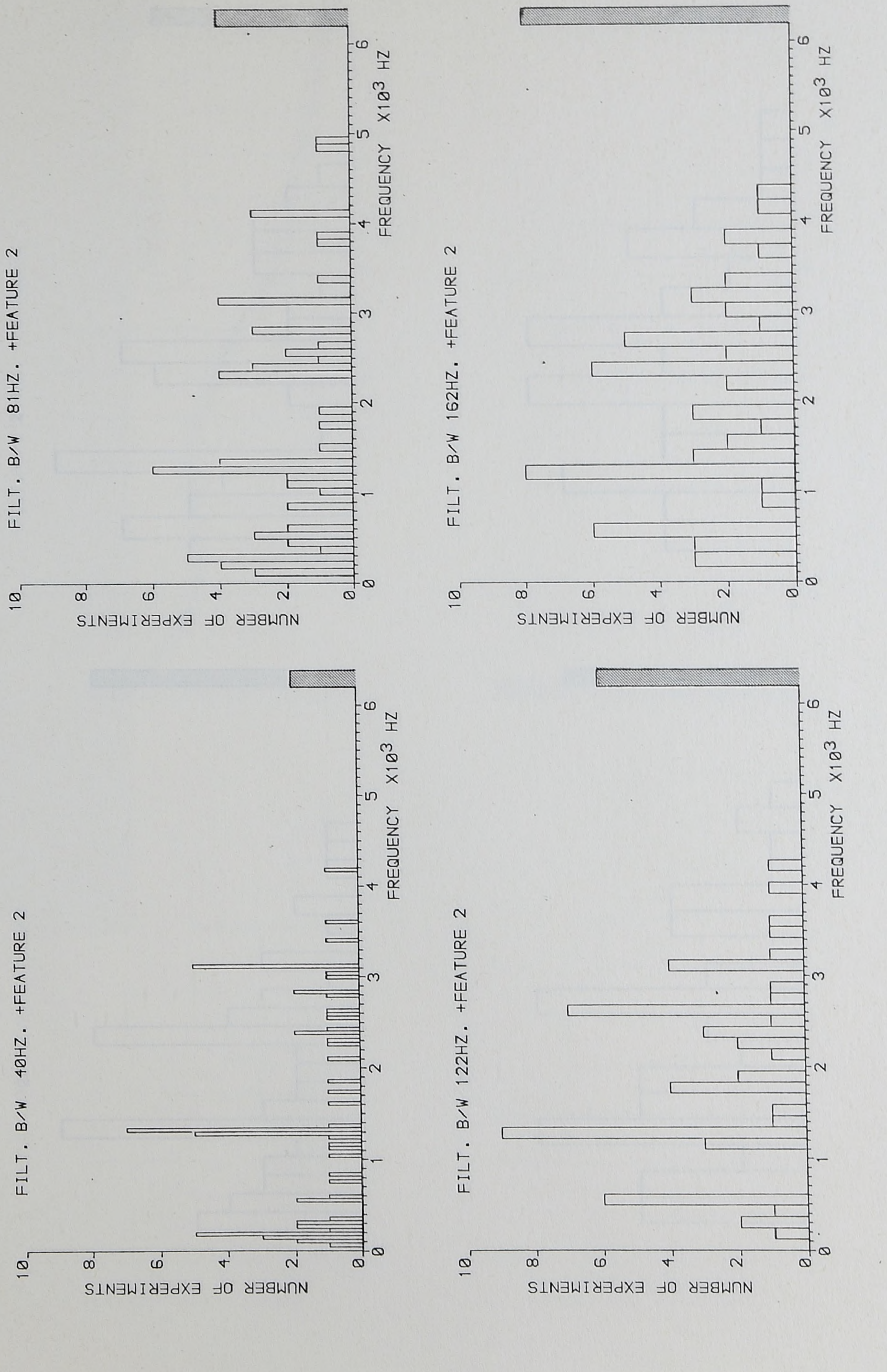


Figure 7.9b. Selected frequency bands & additonal feature 2.

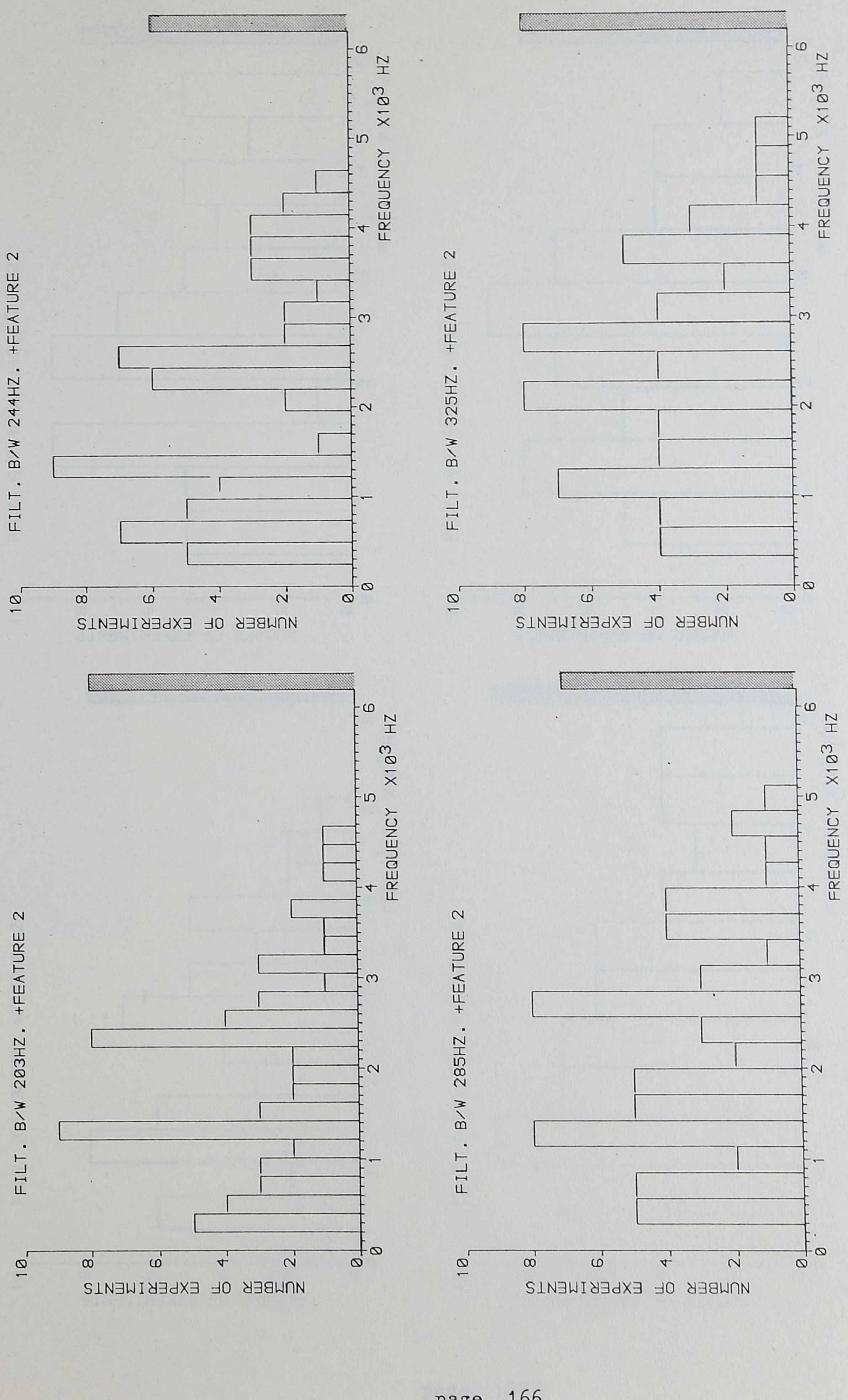


Figure 7.9c. Selected frequency bands & additonal feature 2.

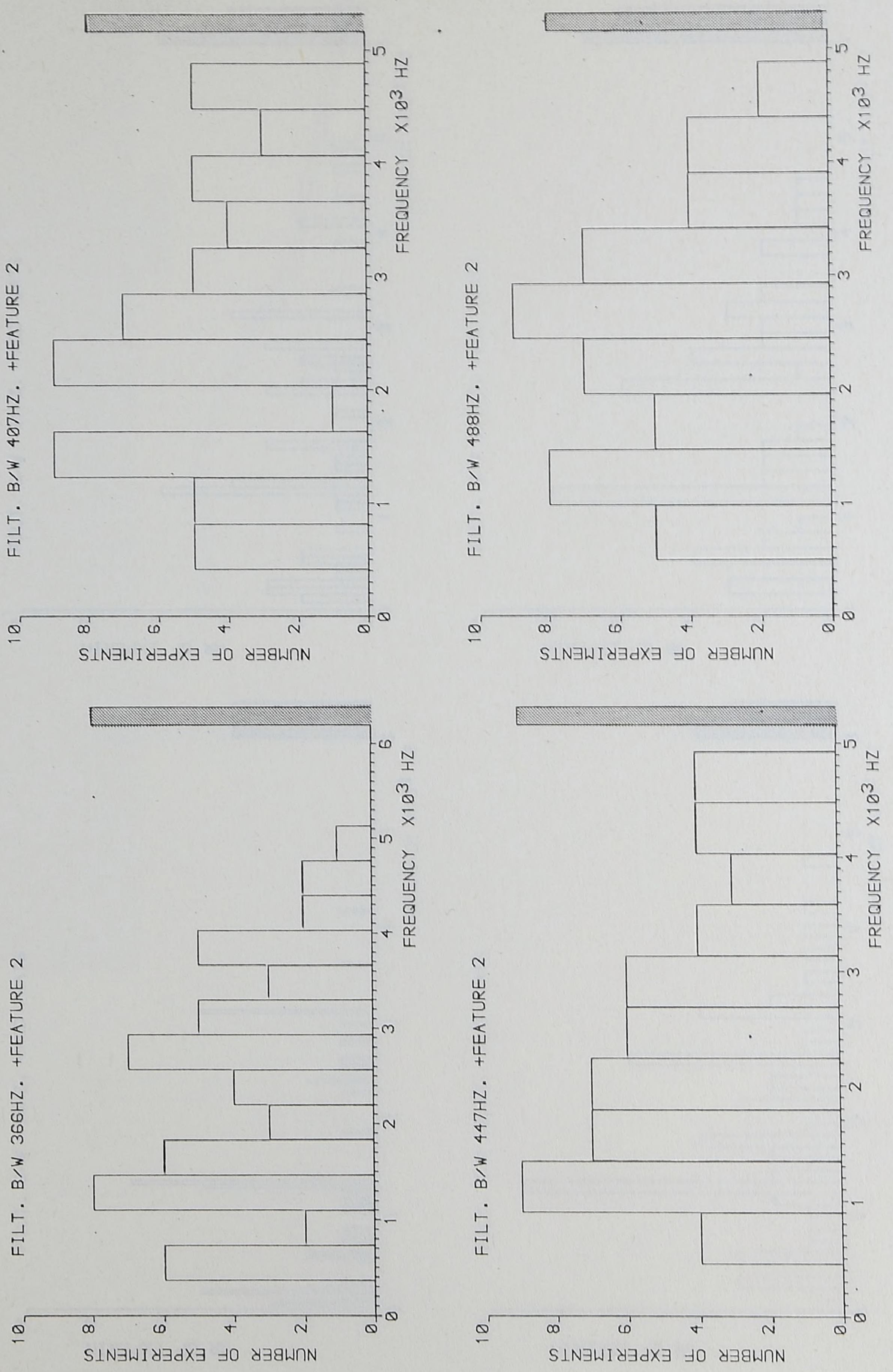


Figure 7.10a. Selected frequency bands & additional features.

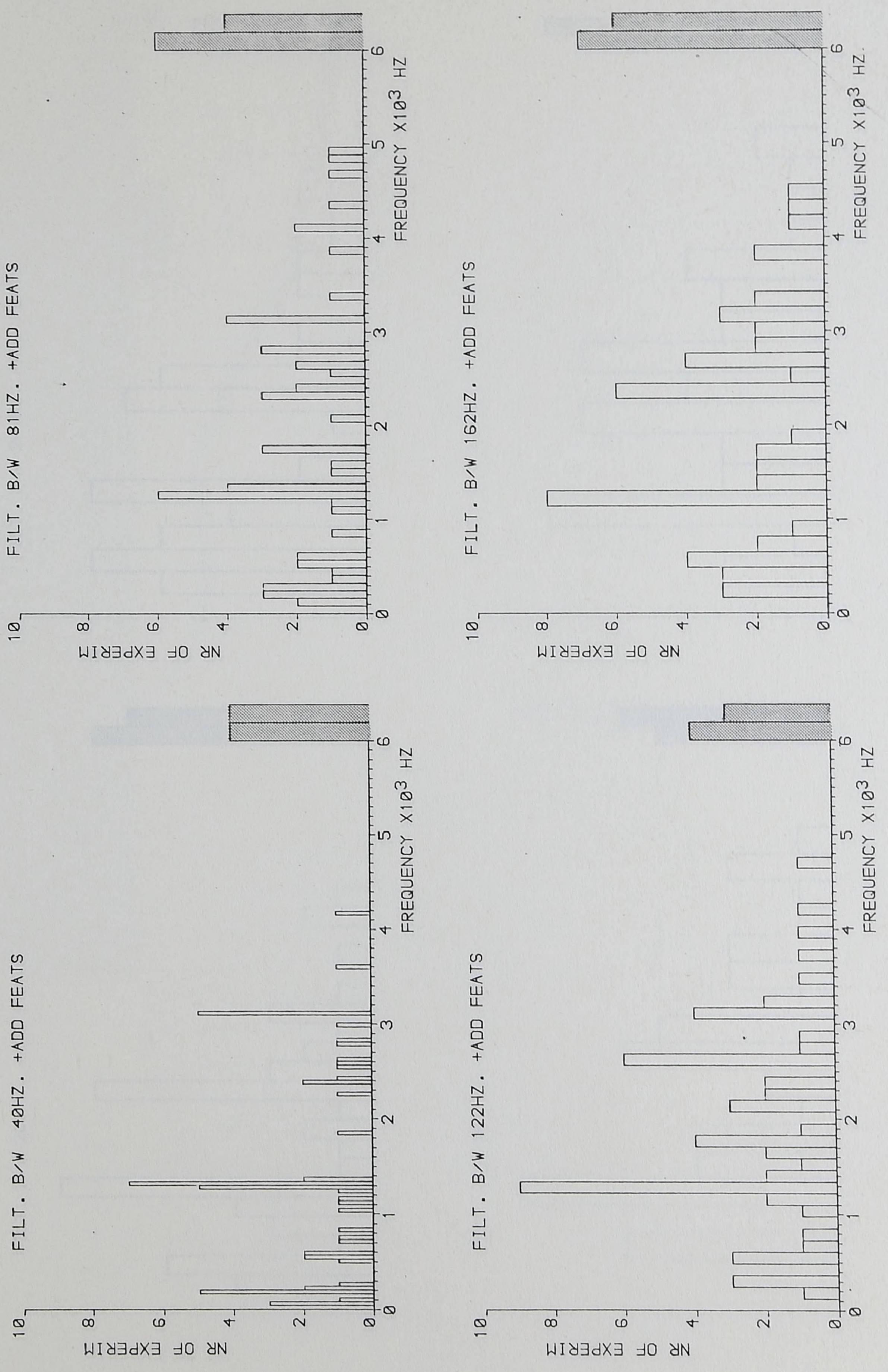


Figure 7.10b. Selected frequency bands & additional features.

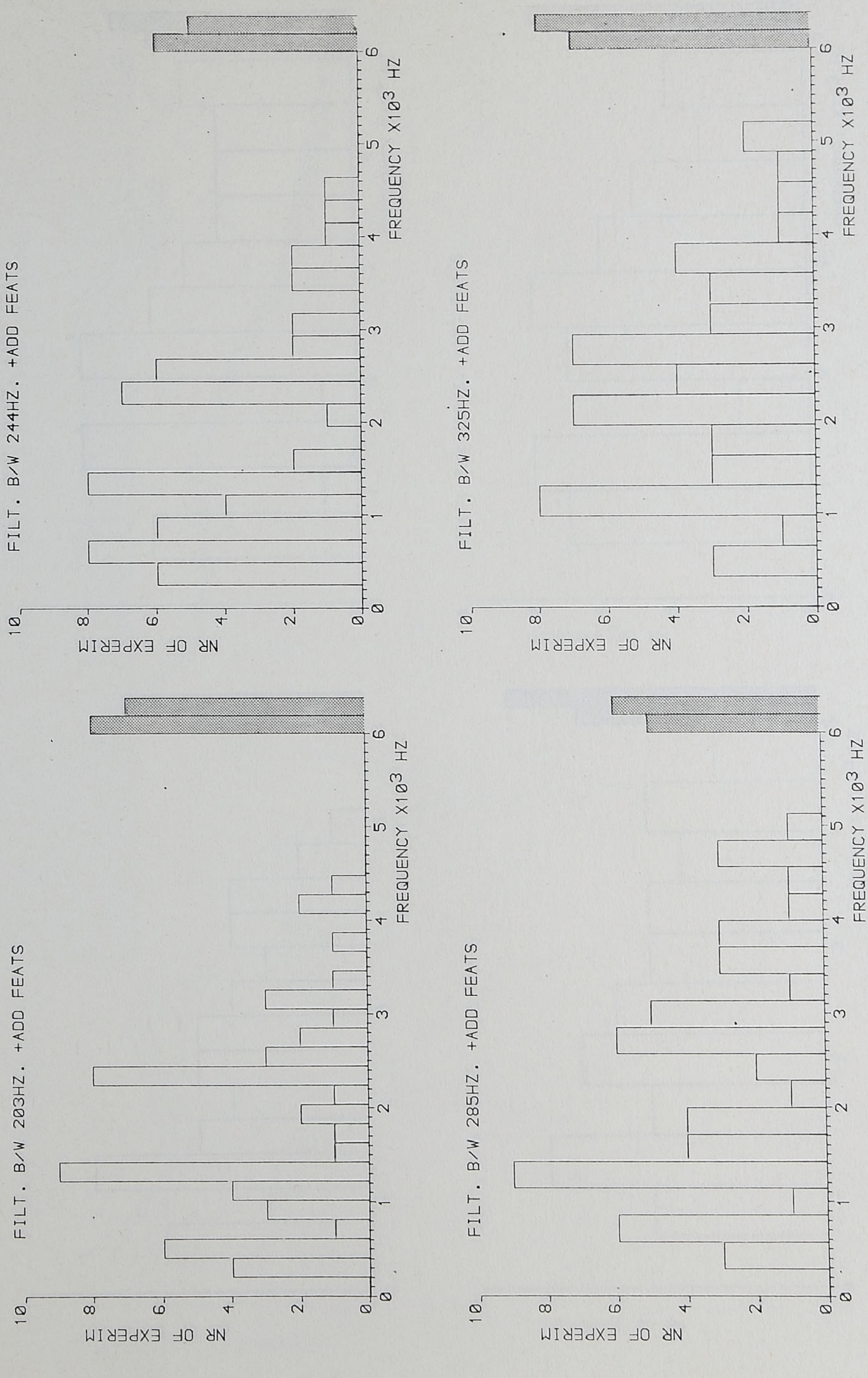


Figure 7.10c. Selected frequency bands & additional features.

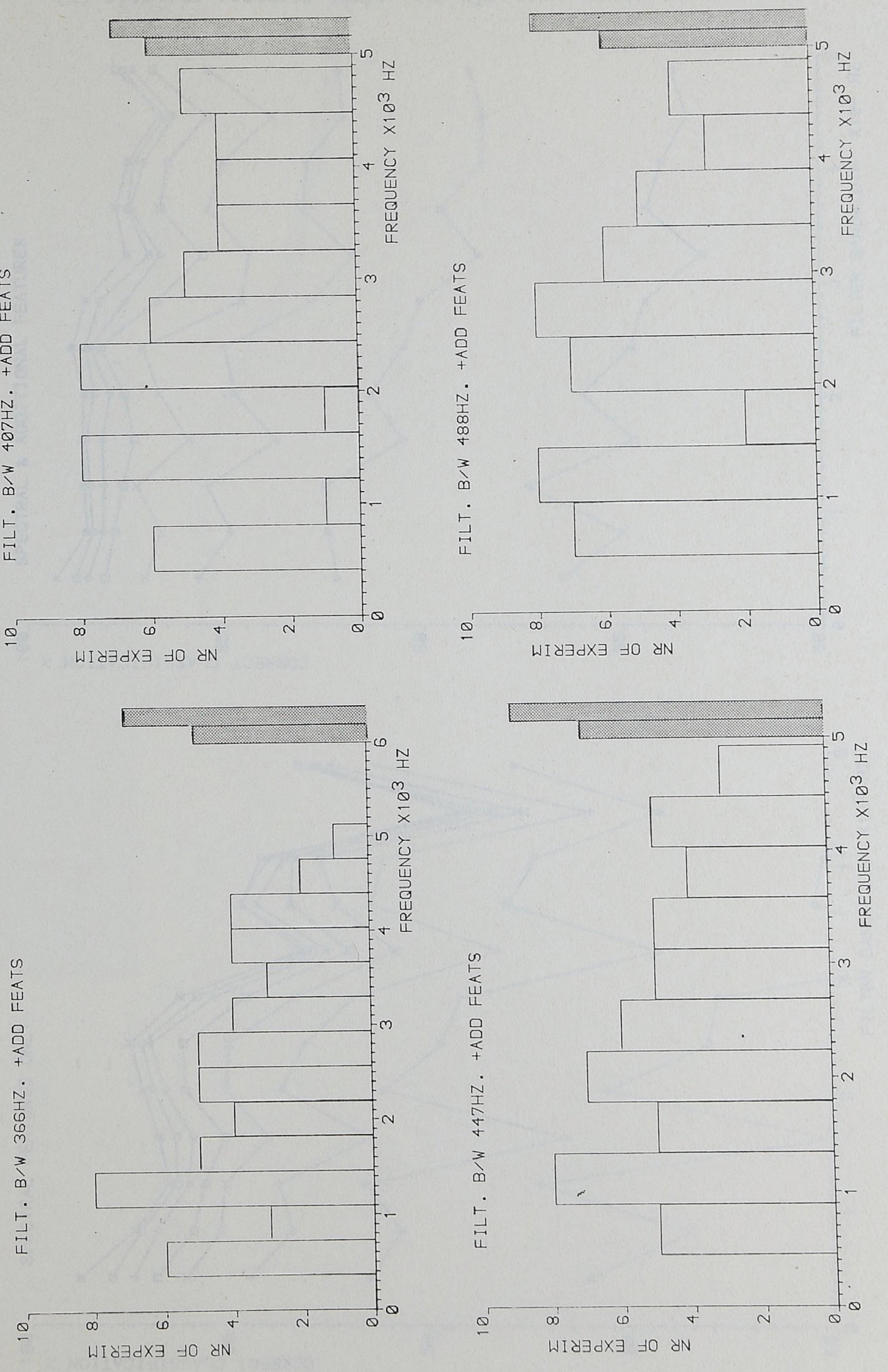


Figure 7.11. Effect of filter bandwidth on classifier performance.
(a) Spectral features only. (b) Spectral & additional features.

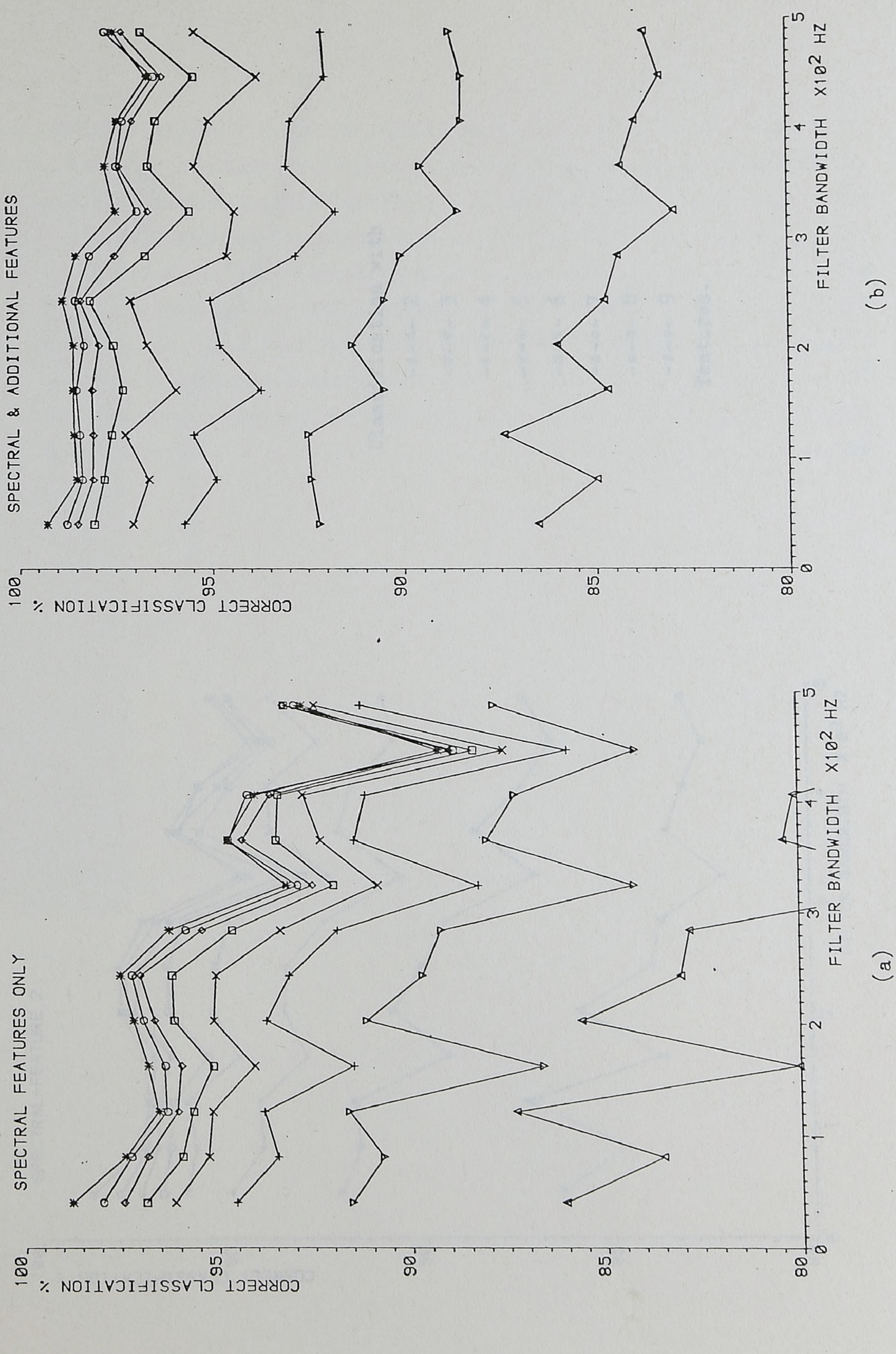
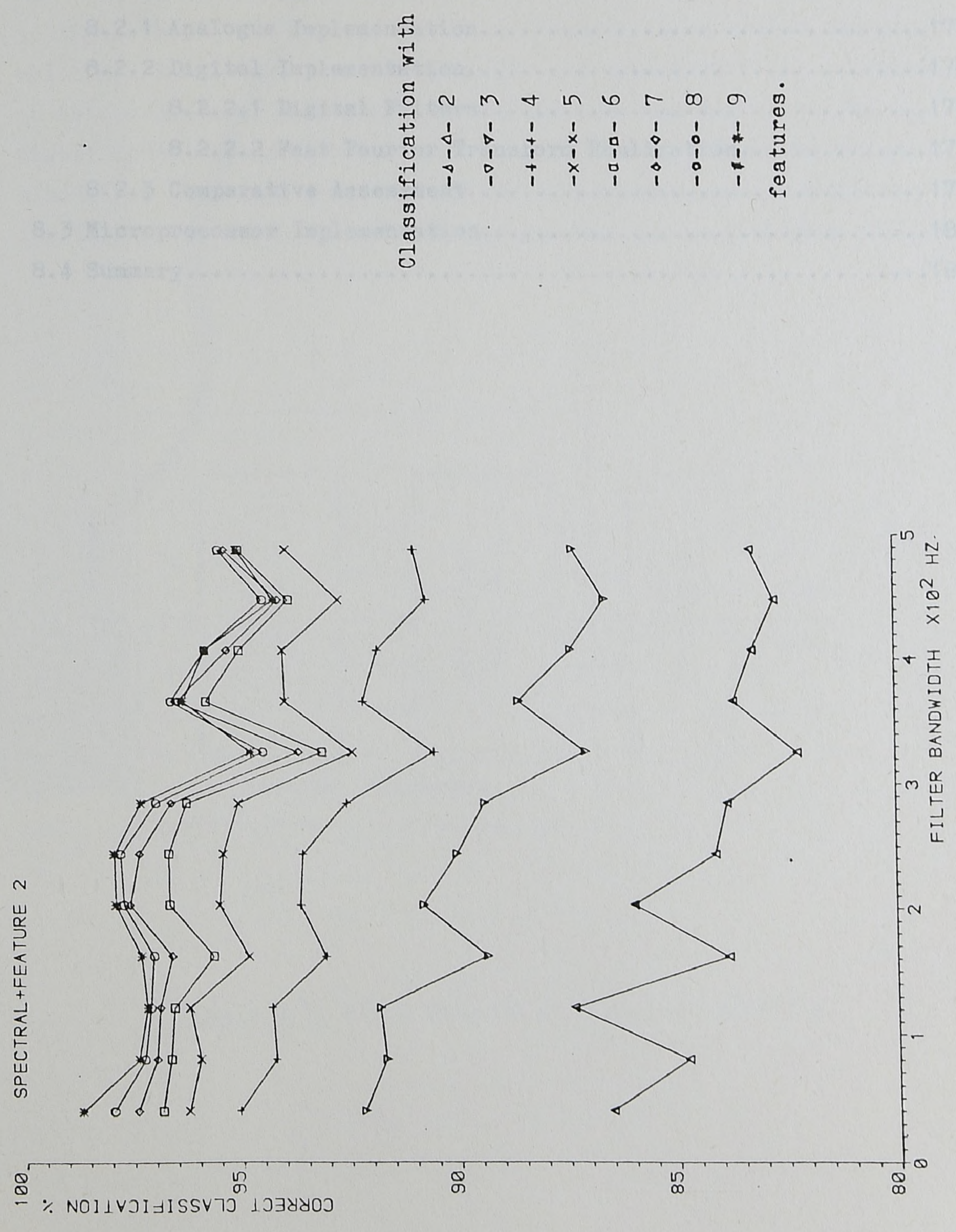


Figure 7.11c Effect of filter bandwidth on classifier performance.

Spectral & additional feature 2.



CHAPTER 8

8 IMPLEMENTATION OF A SOURCE IDENTIFIER

The results of the simulation study of the previous chapters indicate that some form of filter bank is the most important part of any implementation of a source identifier. The filter bank will produce the 'raw' features which will then be fed to the classifier as depicted in figure 8.1. The type of the filter bank will determine the way in which any additional processing (e.g. averaging) is to be performed in an intermediate stage between the filter bank and the classifier. Signal conditioning (low and high pass filtering and prewhitening, is assumed to have been performed by the amplifier/signal conditioner of the LAANMS, as shown in the configuration of figure 2.2 (page 28).

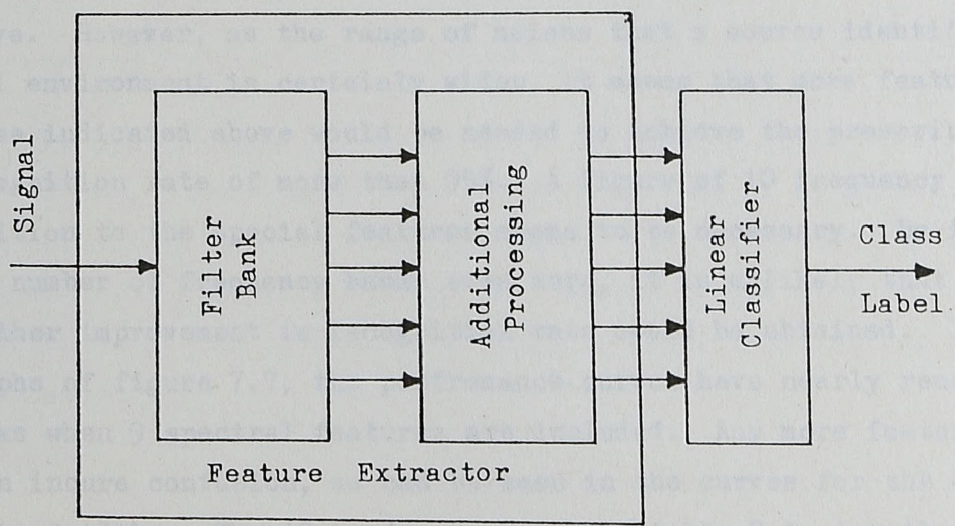


Figure 8.1. Block diagram of a source identifier.

8.1 Simulation Results: Implications

The results of the simulation of the previous chapter indicated that the features required for recognition better than 95% are either

a) the average amplitudes of 6 frequency bands, preferably of 250 Hz bandwidth, averaged over 450 ms.

b) the average amplitudes of 5 frequency bands, as above, in addition to the 'amplitude' of the 'pseudospectrum' (the Fourier transform of the time variation of the average amplitude of the signal over 450 ms) in the range of 10 to 13 Hz or

c) the average amplitudes of 4 bands as in a) above in addition to the 'amplitude' of the pseudospectrum and the 'gamnitude' of the average cepstrum in the range 3.8 to 5.8 ms.

An implementation of an instrument for the identification of noise pollution sources should be based on the results briefly mentioned above. However, as the range of noises that a source identifier in a real environment is certainly wider, it seems that more features than those indicated above would be needed to achieve the prescribed recognition rate of more than 95%. A figure of 10 frequency bands in addition to the special features seems to be necessary. By increasing the number of frequency bands even more, it is unlikely that any further improvement in recognition rate could be obtained. In the graphs of figure 7.7, the perfromance curves have nearly reached their peaks when 9 spectral features are included. Any more features might even incure confusion, as can be seen in the curves for the 450 and 490 Hz bandwidths. The 10 bands are given in table 8.1, together with their ranking (chapter 9, section 7.3). They have been taken from the histogram for the 244 Hz bandwidth of figure 7.8b.

Another consideration is that when more noise types are likely to occur in the vicinity of the identifier, the actual frequency bands that would incur the best recognition rate should be ascertained with a simulation study similar to the one presented in this thesis. The instrument must, therefore, have adjustable centre frequencies of the frequency bands.

Table 8.1

Frequency Bands for Source Recognition

(244 Hz Bandwidth)

	Frequency Range		
Band	Low	High	Rank
2	244	488	4
3	488	732	8
5	976	1220	3
6	1220	1464	8
9	1952	2196	3
10	2196	2440	8
11	2440	2684	7
14	3172	3416	4
15	3416	3660	3
16	3660	3904	4

The useful range of frequencies, as seen in table 8.1, is below 5 kHz. Hence, a sampling rate of 10 kHz, instead of 20 kHz, would be more economical to use in the implementation. This implies that 18 instead of 36 512-point windows overlapping by 50% would cover the 450 ms duration of the subrecord, offering a significant saving in computation.

8.2 Alternative Implementations

The source identifier can be implemented either using analogue or digital hardware. In the following sections the alternatives are briefly discussed and comparatively assessed.

8.2.1 Analogue Implementation.

An analogue implementation requires a filter bank of up to 10 passive or active band-pass filters of an order higher than two, so as sharp cut-offs can be realized. The linear classifier can then be constructed so as to accept as input the outputs of the filters, form an appropriate linear weighted sum of the filter values for each class and determine the largest sum for classification.

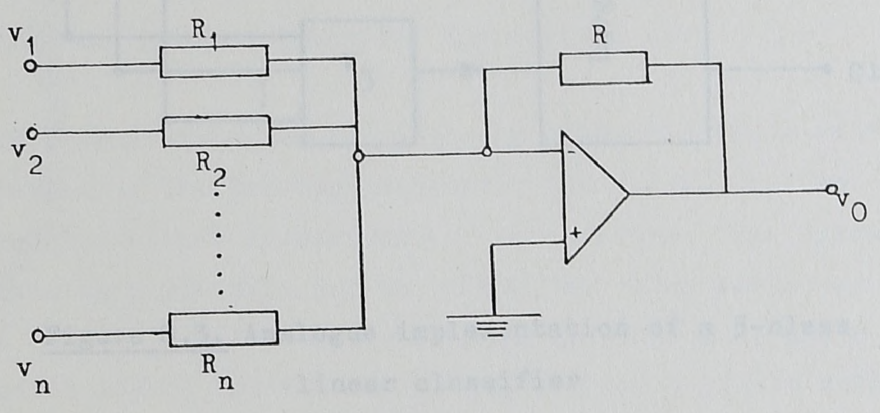


Figure 8.2. Analogue implementation of a weighted sum.

For the realization of the weighted sum, W.J. Hill [1977] proposes a circuit (fig. 8.2) with one operational amplifier and as many resistors as the number of filters. The voltage at the output of operational amplifier or figure 8.2, is given by the equation

$$-v_o = v_1 \cdot R / R_1 + v_2 \cdot R / R_2 + \dots + v_n \cdot R / R_n$$

One such circuit is needed for each class. The outputs of the weighting circuits are then fed into a 'largest value' block whose outputs are as many as the inputs. The output corresponding to the largest input takes the value logical one and all other outputs logical zero.

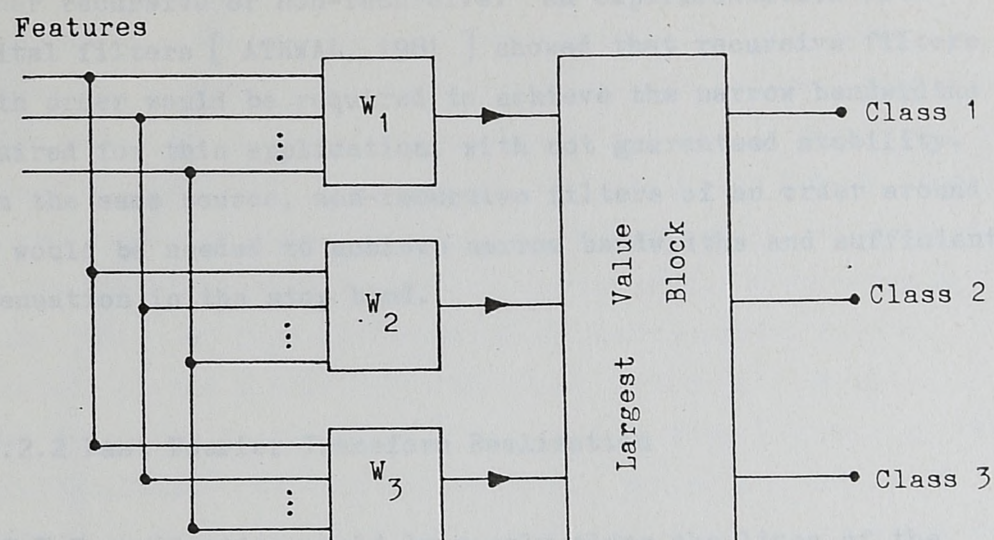


Figure 8.3. Analogue implementation of a 3-class linear classifier

Figure 8.3 depicts the analogue implementation of the linear classifier for three classes.

8.2.2 Digital Implementation

For the digital implementation, an analogue to digital converter (ADC) is required in addition to the filter bank. The ADC should be fed by an automatic gain controller to suppress the dynamic range of the input signal. ADCs, being nowadays standard building blocks, they are available as self-contained integrated circuits.

The linear classifier should also be implemented in digital form either in hardware or in software, depending on the form of filter bank that can either be a bank of digital filters or hardware or software F.F.T.

8.2.2.1 Digital Filters

The filter bank may be implemented by using digital filters either recursive or non-recursive. An experimentation with digital filters [ATHWAL, 1981] showed that recursive filters of 12'th order would be required to achieve the narrow bandwidths required for this application, with not guaranteed stability. From the same source, non-recursive filters of an order around 100 would be needed to achieve narrow bandwidths and sufficient attenuation in the stop band.

8.2.2.2 Fast Fourier Transform Realization

An F.F.T realization would be mostly along the lines of the simulation of the previous chapter. The F.F.T. may be implemented either in software or in hardware. The hardware approach is preferable for it offers the major advantage of speed which is very critical for this application. A hardware F.F.T. processor may be built either using digital hardware building blocks such as multiplier/accumulators, controller/sequencers, ROMs/RAMs and function generators; or by using a special purpose LSI package to perform the F.F.T.

8.2.3 Comparative Assessment

Analogue implementation offers the advantage of speed and low cost and ease of construction and debugging. However, it would be difficult to anticipate the performance of such a device, as there can be no direct comparison with the simulation which is entirely digital. Secondly, problems of signal drift and stability may arise with analogue hardware. Thirdly, an analogue implementation would be inflexible, unless redundancy is built into the device. But then this would be at the expense of cost and ease of construction.

The digital filter option is easily dismissed for, according to Athwal [1981], it would take about 8 minutes for the 10 filters to extract the features. Software computation of 18 F.F.T.s in double precision arithmetic could take up to 80 or 100 seconds on a Z80 microprocessor endowed with a hardware multiplier, including digitization, averaging and classification. In contrast with the software solution, a signal processing chip [AMI], is claimed to compute a single F.F.T. in under 50 ms with Direct Memory Access (DMA) and otherwise 300ms. Hence, recognition could be achieved in under one second or at the worst in 6.5 seconds. Very recently [BRAIN, 1981] an even faster hardware configuration for the realization of the F.F.T. is claimed to perform a 1024-point complex transform in less than 3 ms (or less than 1.5 ms for a 512--point transform). If this was employed, the entire set of 18 transforms in much less than 450 ms, the duration of the subrecord, would be possible. The software solution is cheaper, however, while the one incorporating the special chip is the costliest.

The conclusion from the comparative assessment is that the F.F.T. approach offers a realistic recognition time combined with flexibility and the advantage of direct relation to the simulation. Another advantage offered by the F.F.T. approach is the possibility to compute the cepstra by one additional transformation of the averaged spectra and a summation of the cepstral components in the range of the cepstral peak (section 7.1.2.2 and 7.2.1). The inclusion of the second additional feature of sections 7.1.2.3 and 7.2.2 could also be implemented by averaging the components of the average spectra, to compute the average amplitude. Then, by transforming the sequence of the average amplitudes to form the 'pseudospectrum' and by summing up the resulting components in the range of 10 to 13 Hz, the complete set of features, as defined and evaluated in the simulation could be utilized with minimal additional cost.

8.3 Microprocessor Implementation

A possible configuration of the identifier incorporating a hardware F.F.T. and a microprocessor for control and computations is shown in the block diagram of figure 8.4

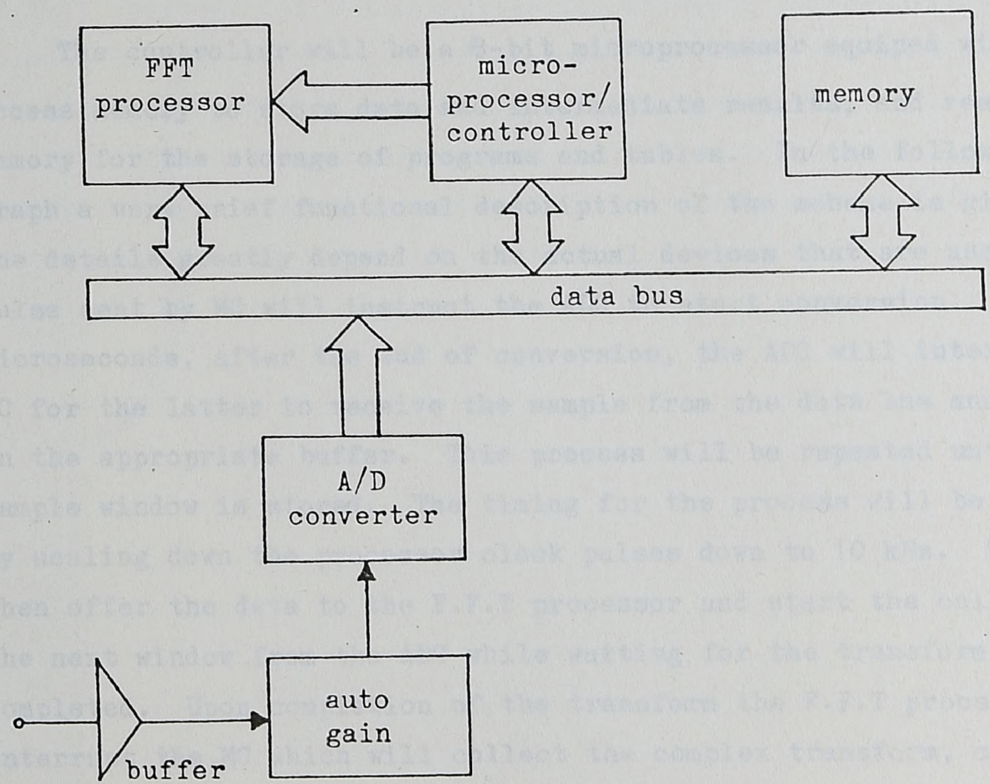


Figure 8.4. Microprocessor implementation of noise pollution source identifier.

As the instrument requirements specified in chapter 2, the amplifier/signal conditioner of LAANMS will provide for all necessary filtering and pre-whitening of the acoustic signal, before it is fed into the identifying module. A buffer stage is included in the suggested configuration for minimum loading of the amplifier.

The automatic gain stage will ensure that the signal lies within the prescribed input range of the analogue to digital converter (ADC). Thus, clipping of the signal at the two extremes of the range will be avoided. The ADC will preferably be a 8-bit device, giving approximately a 50 dB dynamic range. The choice of 8-bits will also facilitate the storage of the samples (1 sample per byte). The device will be controlled by the microprocessor/controller (MC). The F.F.T processor will either be an LSI package or a configuration of simpler devices, as mentioned in section 8.2.2.2.

The controller will be a 8-bit microprocessor equiped with random access memory to store data and intermediate results, and read only memory for the storage of programs and tables. In the following paragraph a very brief functional description of the scheme is given, as the details greatly depend on the actual devices that are used. A pulse sent by MC will instruct the ADC to start conversion. In a few microseconds, after the end of conversion, the ADC will interrupt the MC for the latter to receive the sample from the data bus and store it in the appropriate buffer. This process will be repeated until the 512 sample window is stored. The timing for the process will be supplied by scaling down the processor clock pulses down to 10 kHz. The MC will then offer the data to the F.F.T processor and start the collection of the next window from the ADC while waiting for the transform to be completed. Upon completion of the transform the F.F.T processor will interrupt the MC which will collect the complex transform, compute the amplitude spectrum, apply any additional processing (e.g windowing applied as smoothing of the spectral points, as explained in chapter 3 section 3.1.2.2) and store the amplitude spectrum. When the 450 ms of signal is transformed the MC will compute the average spectrum (and the additional features if required) and start classification by computing properly weighted sums of the selected frequency bands and finding the maximum sum. Whether one microprocessor can cope with both controlling and computation will depend on the actual devices used. . If this is not possible, a second processor can be included to perform the computations.

The construction of a similar scheme has already started by J.S Athwal [1981], incorporating two 8-bit microprocessors and an AMI signal processing peripheral for the computation of the F.F.T. The system has not yet been validated, although the software has been developed.

8.4 Summary

In this chapter, the possibilities for hardware implementation of a noise pollution source identifier were examined. For reasons of flexibility, speed and comparability with the simulation presented in earlier chapters, a microprocessor controlled scheme was suggested, incorporating a special purpose F.F.T processor. A brief functional description of the suggested scheme was also presented.

CHAPTER 9

9 SUMMARY & CONCLUSION

The necessity for a more efficient way of monitoring noise pollution has led to an investigation of the possibility of automatic identification of noise pollution sources. The outcome of the investigation would be an instrument to enhance the unattended monitoring system installed in the Greater London area. This eventuality directed the research to the analysis of the acoustic signals emitted by noise pollution sources by employing signal processing and classification techniques implementable at low cost and speed.

The investigation started with a study of the properties of noise pollution sounds in order to find exploitable characteristics useful for identification. This study suggested that the time-varying spectra of the sounds emitted by the sources contain information reflecting the operating characteristics of the sources and their relation to the environment in the following ways: Firstly, specific frequency bands corresponding to the constructional and operational details of the sources are prominent; secondly, the reflections caused by the ground surface produces ripples in the spectra indicative of the altitude of the source and thirdly that slow bursts of energy are revealed in the spectrograms as isolated spectra with distinctive total power. Verification of the ideas developed therein required the development of a data acquisition system compatible to digital analysis of sound signals and the creation of a data base of noise pollution sound records. Generation of spectrograms to reveal the structure of the sounds was subsequently simulated digitally and a system for their enhancement was also developed. The generated spectrograms of jet aircraft, helicopters and trains verified the results of the analysis and indicated that there exist structural differences in the distribution of the energy of the sound signals, both in frequency and time. Due to the constraints imposed by the cost and speed of the hardware that can be used to implement the feature extraction and classification process in the final instrument, it was decided to dispense with any structure in the signals that extends beyond 450 ms.

This meant that recognition should be based on short (450 ms) segments of sound.

Finally, it has been shown that for three noise pollution sounds which are structurally different (namely, fixed wing aircraft, helicopters and trains), recognition correct to better than 95% can be obtained using a linear classifier, the signal the square root of the signal power within certain bands being used as features. The performance was found to depend on the width of the bands, decreasing with increasing bandwidth. The bandwidth of 240 Hz was found to be optimum in the sense that it allows slight variations in the distribution of energy with frequency. The performance was found to improve to better than 98% at the cost of increased computation by incorporating additional features representing cepstral 'power', and the fluctuation with time of total power within the signal. The additional features also improved the dependance on the bandwidth and were found to be among the 'best' features.

A scheme for the implementation of an instrument, based on the above findings, was also presented. Digital implementation was preferred to analogue for reasons of versatility and comparison with the simulation. The introduction of fast hardware to perform the fast Fourier transform enables the whole simulation process to be repeated in microhardware and software.

For future development, the following are suggested:

a) Further extension of the data dbase to include a larger variety of pollution noises of interest, with emphasis on street noise. The performance of the developed system should then be tested on more realistic terms.

b) Exploration of the possibility to improve the recognition rate (which is likely to drop when more classes are considered) by concatenating consequtive decisions. Thus, the decision about a class will be based over longer observation of the signal..

CHAPTER 10

10 REFERENCES

AHMED, N. & RAO, K.R. (1975).
Orthogonal transforms for digital signal processing.
(Springer-Verlag).

AMI
S281A advanced product description. (American Microsystems, Inc.)

ATHWAL, J.S. (1981).
Design and development of a helicopter sound identification instrument.
M.Sc. thesis: Department of Systems Science, City University, London.

BARDUCCI, I. & CANNELLI, G.B. (1975).
Rating procedures for traffic noise. In: 8th INTERNATIONAL CONFERENCE ON ACOUSTICS, SYMPOSIUM ON NOISE IN TRANSPORTATION, Southampton, July 1974. Journal of Sound and Vibration, 43 (2) : 397-405.

BERGLAND, G.D. (1969).
A guided tour of the Fast Fourier Transform. IEEE Spectrum, July 1969 : 41-52.

BINGHAM, c; GODFREY, M.D. & TUKEY, J.W. (1967).
Modern techniques of power spectrum estimation. IEEE Trans on Audio & Electroacoustics, AU-15 (2) : 55-66.

BLACKMAN, R.B & TUKEY, J.W. (1959).
The measurement of power spectra. (N.Y.: Dover Publ.).

BOGERT, B.P.; HEALY, M.J.R. & TUKEY, J.W. (1963).
The frequency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In: Time Series Analysis, ed. by M. Rosenblatt. (N.Y.: Wiley).

BRAIN, D.J. (1981).
(Advanced Micro Devices (UK) Ltd.)

BRIGHAM, O.E. (1974).
The Fast Fourier Transform. (Prentice Hall).

CHILDERS, D.G.; SKINNER, D.P.; KEMERAIT, R.C. (1977).
The cepstrum: a guide to signal processing. Proc. IEEE 65 (10) :1426-1442.

COHEN, T. (1970).
Source depth determinations using spectral, pseudo-autocorrelation and cepstral analysis. Geophys. J. Roy. Astron. Soc.. 20 : 223-231.

COMMAND 2056 (1963)
Noise: Final report of the committee on problems of noise. (London: Her Majesty's Stationary Office).

COOLEY, J.W. & TUKEY, J.W. (1965).
An algorithm for machine computation of complex Fourier Series. Math. Computation, 19 : 297-301.

DART, R. (1974).
Real time high speed digital analysis for identification of rotating machinery noise. Electronique et Microelectronique Industrielle (EMI), 195 : 40-45. (French).

DEC (1973).
PDP-11/05/10/35/40 Processor Handbook. (Mayland, Massachusetts: Digital Equipment Corporation).

DEC(b) (1973).
Peripherals Handbook. (Mayland, Massachusetts: Digital Equipment Corporation).

DEC (1976).
RT-11 System Handbook. (Mayland, Massachusetts: Digital Equipment Corporation).

DORPFELD, W. & BELLINGER, J. (1978).
Fault detection by spectral analysis of sound as a Q.A. tool.

DUDA, R.O. & HART, P.E. (1973).
Pattern classification and scene analysis. (Wiley-Interscience).

ENOCHSON, L.D. & OTNES, R.K. (1972).
Digital Time Series Analysis. (Wiley).

FARAG, F.H.; ROTHWEILER, J.H. (1980).
Signal processing and pattern recognition for intelligent sensors. Paper presented at the IEEE Workshop on Digital Signal and Waveform Analysis, Miami, Florida, December 1980.

FLANAGAN, J.L. (1972).
Speech Analysis, Synthesis and Perception. (Springer-Verlag).

FREEDMAN, R.; HYDE, J.R.; REED, S.B. (1974).
Recording of noise and air pollution measurements. Environmental Health, 82 (1) : 3-5.

FU, K.S. (Editor) (1976).
Digital pattern recognition. (Springer-Verlag).

GUDAT, H.; SYRBE, M. & WILL, B. (ed) (1977).
Possibilities of quality control in industrial production through use of acoustic methods. In: AUTOMATISIERUNGSTECHNIK IM WANDEL DURCH MICROPROZESSOREN, INTERKAMA KONGRESS, 6-12 October. Proceedings. pp. 118-127. (German).

GUDAT, H. (1978).
The possibilities of noise testing of pattern recognition methods. Qual. und Zuverlassigkeit (Germany), 23 (3) : 63-67. (German).

HARRIS, C.M. (Editor) (1979).
Handbook of noise control. (Second edition, McGraw-Hill).

HARRIS, F.J. (1978).
On the use of windows for harmonic analysis with the Discrete Fourier Transform. Proc. IEEE 66 (1): 51-83.

HILL, W.J. (1977).
Defect recognition in automated surface inspection. Ph.D Thesis: The City University, Department of Systems Science.

HILLQUIST, R.K. & SCOTT, W.N. (1975).
Motor vehicle spectra: their characteristics and dependence upon operating parameters. Journal of the Acoustical Society of America, 58 (1) :2-10.

JELINEK, F. (1976).
Continuous speech recognition by statistical methods. Proc. IEEE, 64 (4) : 532-556.

JENKINS, G.M. & WATTS, D.G. (1968).
Spectral analysis and its applications. (San Francisco: Holden-Day).

KANAL, L. (1974).
Patterns in Pattern Recognition: 1968-1974. IEEE Trans on Information Theory, IT-20 (6):697-722.

KEELER, J.S. (1972).
Piecewise periodic analysis of almost periodic sounds and musical transients. IEEE Trans on Audio & Electroacoustics, AU-20 (5) : 338-344.

LACHENBRUCH, P.A. & MICKEY, M.R. (1968).
Estimation of error rates in discriminant analysis. Technometrics, 10 (2): 1-11.

LEVERTON, J.W. (1975)
Helicopter noise: Can it be adequately rated? In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SYMPOSIUM ON NOISE IN TRANSPORT, Southampton, July 1974. Journal of Sound and Vibration, 43 (2) : 351-361.

MALING, G.C.; MORREY, W.T. & LANG, W.W. (1967).
Digital determination of third-octave and full octave spectra of acoustic noise. IEEE Trans on Audio & Electroacoustics, AU-15 (2) : 98-104.

MAX, J. (1960).
Quantizing for minimum distortion. IRE Trans. on Information Theory, IT-6: 7-12.

MOUKAS, P. (1978).
Identification of sources of noise pollution. Report DSS/PM/154, Dept. of Systems Science, City University, London.

MOUKAS, P. & NORTON-WAYNE, L. (1981).
Computer recognition of nuisance noises. In: INTERNATIONAL CONFERENCE ON DIGITAL SIGNAL PROCESSING, Florence, Italy, 2-5 September. Proceedings. pp. 1103-1110.

MOUKAS, P.; SIMSON, J. and NORTON-WAYNE, L. (1982)
Identification of noise pollution sources. To be published in IEEE Trans. on Systems, Man and Cybernetics

MULHOLLAND, K.A. & ATTENBOROUGH, K. (1981).
Noise assessment and control. (Construction Press).

NICHOL, D.G. (1977).
Processing noisy line spectrograms as digital pictures. Pattern Recognition, 9 :137-146.

NICHOL, D.G. (1979).
Processing spectrograms as digital pictures to separate competing line signatures. Pattern Recognition, 11 :179-190.

NOLL, M.A. (1967).
Cepstrum pitch determination. J. of Acous. Soc. of America, 41 (2) : 293-309.

NUTTALL, A.H. (1981).
Some windows with very good sidelobe behaviour. IEEE Trans on Acoust., Speech & Signal Proc., ASSP-29 (1) : 84-91.

OECD (1980).
Part I - Conclusions of the Conference. CONFERENCE ON NOISE ABATEMENT POLICIES, Paris, 7-9 May 1980. Proceedings. Paris: O.E.C.D., pp. i-xvi.

PAPOULIS, A. (1962).
The fourier integral and its applications. (McGraw-Hill).

PAU, L.F. (1977).
An adaptive signal classification procedure: application to aircraft engine condition monitoring. Pattern Recognition, 9 : 121-130.

PAVLIDIS, T. & FANG, G.S. (1972).
A segmentation techique for waveform classification. IEEE Trans on Computers, C-21 : 901-904.

PRIEDE, T. (1967).
Diesel engine noise. I.S.V.R., University of Southampton.

RACAL
T7000 Digital magnetic tape system, Manual. Racal Thermionic Ltd., England.

RACAL
NRZ Digital formating unit, Manual. Racal Thermionic Ltd., England.

RICHARDS, P.I. (1967).
Computing reliable spectra. <u>IEEE Spectrum</u>, January 1967 : 83-90.

SANDO, F.D. (1978).
Road Traffic and the Environment. <u>Statistical News</u>, <u>40</u> : 40.24-40.28.

SCHWARTZ, M. & SHAW, L. (1975).
Signal processing: Discrete spectral analysis, detection and estimation. (McGraw-Hill).

SILVERMAN, H.F. & DIXON, N.R. (1974).
A paremetrically controlled spectral analysis system for speech. <u>IEEE Trans on Acoust., Speech & Signal Proc.</u>, <u>ASSP-22</u> (4) : 362-381.

SILVERMAN, H.F. & DIXON, N.R. (1976).
A comparison of several speech-spectra classification methods. <u>IEEE Trans on Acoust., Speech & Signal Proc.</u>, <u>ASSP-24</u> (4) : 289-295.

SIMSON, J. (1980).
Helicopter Noise Control in London. <u>In</u>: <u>SPRING CONFERENCE OF THE INSTITUTE OF ACOUSTICS, BEDFORD COLLEGE, London. Proceedings.</u> London: Institute of Acoustics.

SINGLETON, R.C. & POULTER, T.C. (1967).
Spectral analysis of the call of the male killer whale. <u>IEEE Trans on Audio & Electroacoustics</u>, <u>AU-15</u> (2) : 104-113.

SMITH, B. (1957).
Instantaneous companding for quantized signals. Bell System Technical Journal, 36:

SMITH, J.T. (1977).
International aircraft noise measurement procedures- Expensive acquisition of poor quality data. In: AIAA AEROACOUSTICS CONFERENCE, 4th, Atlanta, Georgia, Oct 3-5. New York: American Institute of Aeronautics and Aeroacoustics.

SPECHT, D.F. (1966).
Generation of polynomial discriminant functions for pattern recognition. Ph.D Thesis: Stanford University.

STARR, E.A. (1972).
Measuring noise pollution. IEEE Spectrum, June 1972 : 14-25.

SYED, A.A.; BROWN, J.D.; OLIVER, M.J.; HILLS, S.A. (1980).
The cepstrum: a viable method for the removal of ground reflections. Journal of Sound and Vibration, 71 (2): 299-313.

THOMAS, D.W. (1971).
Vehicle sounds recognition: a pilot study. Ph.D. thesis: Faculty of Engineering & Applied Science, University of Southampton.

THOMAS, D.W. (1978).
Vehicle sounds and recognition. In: Pattern Recognition: ideas in practice, ed. B.G. Batchelor. (Plenum Press).

THOMAS, D.W. & WILKINS, B.R. (1970).
Determination of engine firing rate from the acoustic waveform. Electronics Letters, 6 : 193.

THOMAS, D.W. & WILKINS, B.R. (1972)
The analysis of vehicle sounds for recognition. Pattern Recognition, 4 : 379-389.

TSENG, F.I.; SARKAR, T.K. & WEINER, D.D. (1981).
A novel window for harmonic analysis. IEEE Trans on Acoust., Speech & Signal Proc., ASSP-29 (2) : 177-188.

USAMI, T.; KOISUMI, T.; INARI, T. & OHNO, E. (1978).
Practical application of diagnostic signature analysis to testing or rotating machines. In: INFORMATION-CONTROL PROBLEMS IN MANUFACTURING TECHNOLOGY, 17-20 October, 1977. Proceedings. pp. 33-39.

VULKAN, G.H. (1975).
Developments in urban planning against noise. In: INTERNATIONAL CONFERENCE ON ACOUSTICS, SYMPOSIUM ON NOISE IN TRANSPORT, Southampton, July 1974. Journal of Sound and Vibration, 43 (2) : 439-449.

VULKAN, G.H. (1979).
Aircraft noise monitoring for local authorities. In: INTERNATIONAL CONFERENCE ON NOISE CONTROL ENGINEERING, INTER-NOISE '79, Warsaw, September 1979. Proceedings. Warsaw: Polish Academy of Science, pp. 713-716.

WEBSTER, J.C. (1969).
Identifying diesel engine sounds. Journal of Sound and Vibration, 9 (2): 241-246.

WEE, W.G. (1968).
Generalized inverse approach to adaptive multiclass pattern classification. IEEE Trans. on Computers C-17 (12): 1157-1164.

WELCH, P.D. (1967).
The use of FFT for the estimation of power spectra: a method based on time averaging over short modified periodograms. IEEE Trans. on Audio & Electroacoustics, AU-15 (2) : 70-79.

WOHLFORD, R.E. (1980).
A comparison of four techniques for automatic speeker recognition. _In:_ _INTERNATIONAL CONFERENCE ON ACOUSTICS SPEECH AND SIGNAL PROCESSING, Denver, Colorado, April 1980. Proceedings_. pp. 908-911.

YOUNG, R.W. (1968).
Doppler corrected spectra. _Journal of Sound and Vibration_ _7_ (1) : 119-120.

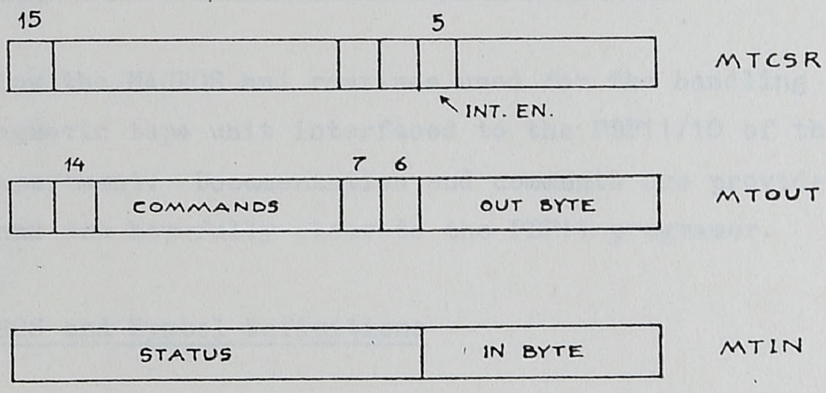
APPENDICES

The computer programs
(p.199-209, 211-240)
have been removed for copyright reasons

APPENDIX I: DIGITAL TAPE HANDLING

This appendix contains a description of the interface of the RACAL digital tape deck (MT) and the routines written for its handling.

1.1 MT Interface



The interface is schematically shown above. The MTIN register contains the 6-bit byte read from tape and information for the state of the deck. The MTOUT register is used to output a 6-bit byte to the tape and also to give it commands. The function of the particular bits of the registers is given in the listings of the next section. A general description of the read/write process follows:
A command is output to the high byte of MTOUT and afterwards the GO bit (bit 14) is set by the program (see MACRO MTOP). The program then must keep in pace with the tape by sending or accepting bytes when requested by the MT. Bit 15 of MTCSR is set when the transfer is done; this is used for synchronization.

When reading, the deck stops at the end of block. As the interface does not give an indication of this, the program must incorporate a time-out function. When writing, the end-of-block command is issued with the last byte of the block by setting bit 7 of MTOUT.

While reading or writing the processor must be at highest priority so that synchronization is not disrupted by any other process (e.g. clock).

1.2 MACROS and Routines for the Handling of the MT

There follow the MACROS and routines used for the handling of the RACAL digital magnetic tape unit interfaced to the PDP11/10 of the Systems Science Department. Documentation and comments are provided in the listings and are hopefully clear to the PDP11 programer.

1.2.1 MACROS and Symbol Definitions

[redacted]

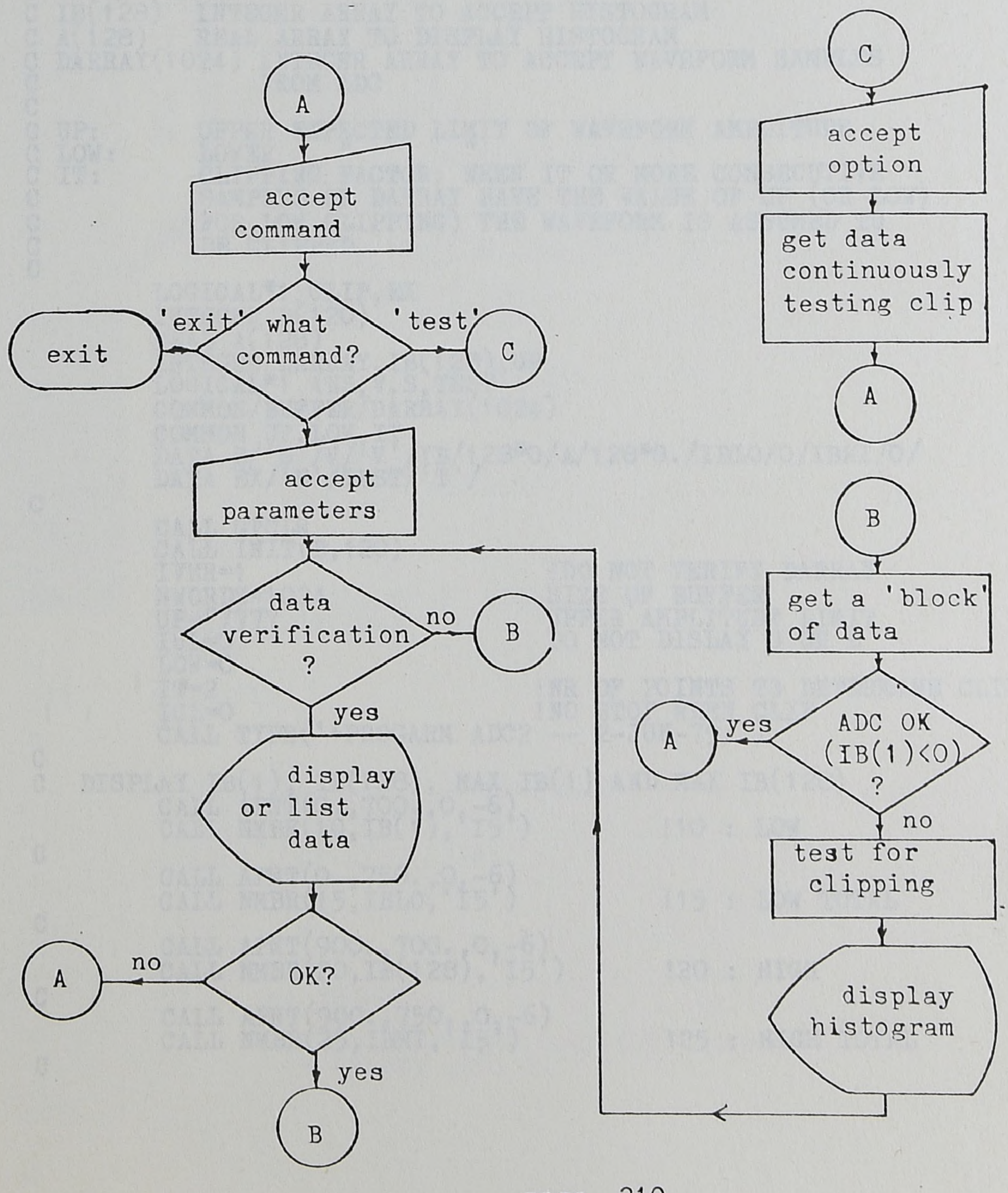
APPENDIX II: ADC & DIGITIZATION

This appendix includes information and programs for the testing of the Analogue to Digital converter and for digitizing analogue records

2.1 Program ADC2

There follows a description of program ADC2 that performs testing of the Analogue to Digital converter. Because the structure of the program is slightly complicated a flow chart is also given.

2.1.1 Flow-Chart

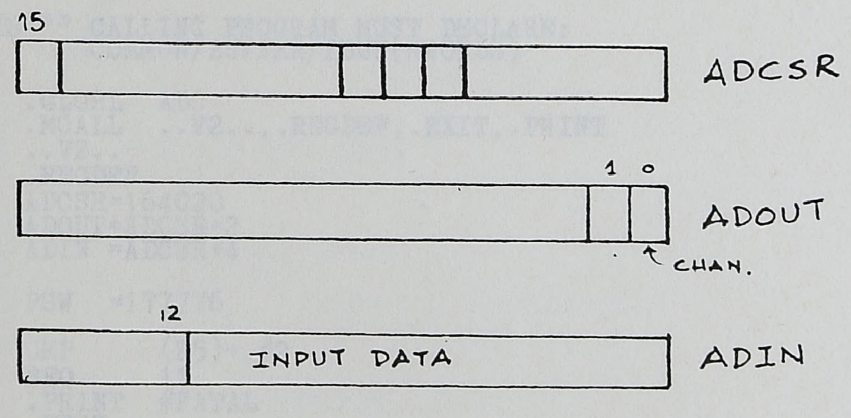


2.1.2 Listing

[redacted]

[redacted]

2.2 The ADC Interface



The figure above gives a schematic description of the interface. Conversion in one of the four ADC analogue channels starts when a number from 0 to 3 is output to the output register ADOUT. When conversion ends, the 15th bit of the ADCSR is set and then the data may be read from ADIN. Accessing ADIN also clears the ADCSR.

2.3 Routines ADC and ADCLIP

These routines are called by program ADC2. The listings include comments to help to their understanding.

[redacted]

2.4 Program DIG

Program DIG is used to digitize analogue records with a bandwidth up to 4kHz. A one block record is written on the tape which must be blocked (Appendix III). The program uses interrupts for synchronization with the tape. Interrupts are enabled by setting bit 5 of the MTCSR. The program starts with a display of the functions to be done by the user, locking out of the operating system, initialization of the MT and ADC, loading of the interrupt vectors of the MT and the Tektronix console (which is used for the termination of the process by hiting any of the keyboard keys). As said in Appendix I, the last byte to be output to the tape must have bit 7 of MTOUT set. Thus, the Tektronix interrupt service routine (KBINT) changes instruction at label INST from BIC (clear bits 6 and 7) to BIT (set bits 6 & 7). Thus, bit 7 of the last byte is set signaling the end of block to the MT. This change of the instruction at INST is also used by the main program in loop WAIT to detect the end of the transfer.

A listing of DIG follows. It contains information on how to run the program and comments to clarify the meaning of the instructions. The routines and MACROS that are not listed may be found in Appendix I which also includes description of the MT interface.

[redacted]

APPENDIX III: TAPE BLOCKING

This appendix includes program BL6 used for the blocking of tapes produced by program DIG, and the assembler routine called for reading from the tape.

3.1 Program BL6

There follows a listing of program BL6 including comments. Some of the routines of Appendix I are used. The user may break in any time to inspect the data transferred from tape or disk. The structure of the program is simple and a flow chart was thought unnecessary.

[redacted]

APPENDIX IV: DATA BASE HANDLING

This appendix includes the programs which create and edit the data base.

4.1 Program UPTAPE

UPTAPE appends new records at the end of the data base. The data base is in tape M8447H (local file name OLDATA) and the new records are in the PDP11 tape PMDAT1 (NUDATA). UPTAPE reads over to file UPDATA the existing data and copies NUDATA.

Each new record is preceded by the following descriptors which are read from cards:

card 1: NBL: Number of blocks of record (format I3)
card 2: SR: Sampling rate (F6.0)
card 3: IDENT: Identifier (A10)
card 4: ICLASS: Comments (8A10)
Reading terminates if NBL=0. This flags the end of the run.

Listing of UPTAPE follows, preceded by the control cards.

[redacted]

[redacted]

4.2 Program AMTAPE

This program performs editing of the data base created by the previous program. AMTAPE allows a) Change of identifier, b) change of comments c) change of sampling rate d) deletion of record. The program accepts up to 20 identifiers for accessing and modifying the corresponding records. Each identifier is followed by the new (if any) descriptors. The format is: Record identifier with flag (format A6,O2). When bits 0,1,2 or 3 of the octal flag are set the corresponding functions a,b,c or d described above will be done e.g. A01.0110 is delete record A01.01, A01.0107 is change identifier, comment and sampling rate of A01.01. If bits 0,1 or 2 of the flag are set the corresponding new items must follow in separate cards immediately and in order. Input is terminated upon reading a card with identifier 'STOP' or 'LIST'. The latter instructs listing of all records.

Listing of AMTAPE follows.

[redacted]

APPENDIX V: SPECTROGRAM GENERATION

This appendix includes the listings of the spectrogram generation program which was developed on the ULCC CDC 7600 computer under the SCOPE 2.1 operating system. The routines that are specific to this installation are not listed.

5.1 Program FSG03P

[redacted]

[redacted]

5.2 Subroutines

There follow the listings of the routines called by FSG03P. Routines FRID and CLOSE, open and close the plot file, respectively. Being special routines depending on the graphics package used (namely PICPAC), they are not included in the listings. The same applies to the routines FFRAMP and DATLST which initialize the frame and output the plot buffer, respectively.

UNPACK1, called by TRANSIA, also depends on the ULCC installation. It unpacks the next segment of signal samples from array IA in array S of the named COMMON /VECTOR/. The actual computation of the F.F.T. is done by the NAG (Nottingham Algorithms Group) routine CO6ABF.

[redacted]