A WARNING SIGNAL IDENTIFICATION SYSTEM (WARNSIS) FOR THE HARD OF HEARING AND THE DEAF

Kwok Wing Chau

B. A. Sc. University of Windsor

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF ELECTRICAL ENGINEERING

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July 1989

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of ELECTRICAL ENGINEERING

The University of British Columbia
Vancouver, Canada

Date July 31, 89

Abstract

The objective of this project has been to design a reliable warning sound recognition system for hard of hearing and deaf people. Commercially available auditory warning devices use simple technologies, which are not able to produce the performance required. The demand for a versatile WARNing Signal Identification System (WARNSIS) that satisfies the needs of hard of hearing and deaf individuals has been well established. This WARNSIS must be "teachable" in order to cope with the many different sounds, and diverse noisy environments. Relevant sounds are telephone rings, sirens, and smoke and fire alarms, and noise includes all other sounds including radio-music, conversation, machinery, etc.

In the absence of published data, we studied extensively both timing and spectral characteristics of warning sounds. We found that the average short-time absolute amplitude of warning sounds is useful in providing timing information, and that the short-time spectra yield characteristic patterns for signal classification.

The WARNSIS operates in real-time, and embodies two parts: the timing analyzer and the spectral recognizer. The timing analyzer continuously monitors the variations of environmental sounds, from which important timing features are derived. If a potential warning sound is detected, the spectral recognizer is activated to analyze its spectral patterns. When these patterns match one of the learned and pre-stored templates, a warning sound is identified with the known warning sound associated with that template. An advantage of such a recognition scheme is that it avoids unnecessary and computationally intensive spectral analysis work when only noise is present.

Evaluation results show that the WARNSIS can reliably recognize warning sounds in random noise with no false alarms. In loud music and conversation backgrounds the WARNSIS can still achieve a high recognition rate, but more false alarms are generated. In household environments where conditions are less demanding than our evaluation criteria, our system is expected to produce very satisfactory results. Since the WARNSIS can be taught to learn and recognize new warning sounds, it may be used in other applications such as noisy industrial sites and traffic light control.

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to thank my supervisor, Dr. C.A. Laszlo for his patience, encouragement, and input during this project. I am greatly indebted to my colleagues, Darrell Wong and Sammy Yick for their invaluable discussions and advice. Special thanks are due to Angela Choi and Michael Slawnych for their comments and suggestions to improve the presentation of this thesis. Finally, very deep gratitude is directed to my family for their generous financial support.

This project was funded by Natural Sciences and Engineering Research Council of Canada grant A67012.

Chapter 1

Introduction

1.1 Background

Auditory communication is vital to normal life. Such communication often focuses on speech which is one of the most effective means of conveying ideas, opinions or information among people. Auditory communication also plays an important role in associating people with their environment. In particular, auditory warnings are of great importance. Such warnings include baby cries, telephone rings, doorbells, door knocks, fire or smoke alarm bells, burglar alarms, car horns, sirens, and electronic buzzers commonly used in household appliances and office equipment.

Generally, auditory warnings are achieved by special sounds. Firstly, warning sounds are usually loud, strident and insistent to effectively cut through speech and background noises, and to command people's attention. Secondly, different warning sounds convey different "messages" which demand responses of varying urgency. Some warning sounds are used to "announce" a condition, or an event; for example, an incoming telephone call, or a visitor at a door. Other warning sounds alert people to potential life-threatening situations such as a fire, or intruders inside a house. Failure to respond to these warning sounds may result in serious harm.

Unfortunately, hearing-disabled people have difficulty in hearing warning sounds and in many cases cannot hear even very loud alarms. This problem extends to many different situations of everyday life. For such individuals, many common household

sounds go undetected (sounds produced by oven buzzers, bathroom fans, stove hood fans, or running water) causing inconvenience and occasional danger in homes. In noisy environments, hearing-disabled individuals cannot discriminate different types of sounds. For example, they cannot hear the sounds that indicate automobile malfunctions such as worn brakes, bad wheel bearings, or noisy mufflers.

In addition, hard of hearing individuals who wear hearing aids can only detect warning signals if their hearing aids are operating and are sensitive enough. Specifically, unless the hearing aid is worn during sleep, hard of hearing people usually cannot hear the sound of burglar alarms, or fire and smoke alarm bells. Furthermore, in tornado-prone states of the U.S. (Kansas, Texas and Arkansas), the general public is usually alerted of approaching tornadoes by loud siren sounds. Missing such warning sounds can be fatal! But hearing-disabled people often cannot hear such sounds, and their utmost concern and their urgent need for special devices to warn of such impending disasters have been forcefully stated [1].

Indeed, the invisible disability of deaf and hard of hearing people creates serious inconveniences, frustrations, fears, and hazards in their daily life. In particular, the vulnerability to missing auditory warnings contributes significantly to the lack of mobility, independence, and security of hearing-disabled persons. In response to the obvious need to help hearing-disabled people to cope with this problem, a number of special alert aids have been designed and marketed.

1.2 Auditory Warning Aids for Hearing Impaired Persons

A range of systems, signalling and wake-up devices are currently available to alert hearing impaired individuals to telephone rings, doorbells, door knocks, fire or smoke alarm bells, and general emergency signals in diverse environments [1,2,3,4]. Some

systems are simple sound amplitude amplification devices, which increase the volume of warning sounds to a level detectable by hearing aid wearers. Other, more sophisticated systems, are capable of driving external visual modules and tactile actuators.

Three major types of auditory warning aids for the hearing-disabled are in use: directly activated hard-wired systems, acoustic threshold detector systems, and hearing ear dogs.

1.2.1 Hard-wired Systems

Such systems require direct electrical connection to sound generating sources. They are reliably activated by the electric signal that drives the warning sound generator, and alert the hearing-disabled by either flashing lights, or by vibratory actuators. To increase the operational range, and to eliminate the need for long cables, an intermediate AM or FM transmitter can be integrated into such systems. Single or multiple remote receivers distributed throughout the home or office can pick up the transmitted signal, and subsequently turn on actuators.

A characteristic example of such systems is the Sonic Alert, which will produce light flashes to alert the hearing impaired to telephone calls. The device can be used with any telephone, and is easily installed by plugging it into any modular telephone jack and electrical outlet. Both the plug-in and a remote radio-transmitter version are available from the Special Needs Department of the British Columbia Telephone Company.

Some hard-wired devices are simple enough to be installed by users without extensive electronic skills (e.g., Sonic Alert). Other, more sophisticated devices, are custom designed, and require permanent installation by a technician at a considerable cost. As reported, these custom designed devices often must be left behind when hearing-disabled individuals move from house to house [1]. In addition, as the number of sound

generating devices increases in homes or offices, the cost of hard-wired systems escalates due to both the wiring required, and the increased complexity. Finally, before any remote warning device is installed, hearing-disabled people have to check if there are similar remote systems installed in neighboring houses. Due to "cross-talk", such systems in close proximity are very prone to generating false warnings.

1.2.2 Threshold Detector Systems

Since warning devices produce sounds that are louder than normal environmental sound levels, threshold detector systems are designed to respond to changes in loudness. Instead of direct connection to sound generating sources, threshold devices employ a microphone, or special electromagnetic field sensor for signal acquisition. With sensitivity adjustment, a threshold device can be adapted to operate with various types of alarms, for example horns, sirens, and telephones, under different acoustic conditions. When the signal level from any source exceeds the preset threshold value of the system, such a device will automatically activate the actuator to alert a hearing impaired individual.

Since these devices cannot selectively identify the sources of the loud sounds, in acoustic systems the microphone is positioned in close proximity to the warning sound generator for maximum system sensitivity and selectivity to the desired inputs. A hearing impaired person can adjust the device sensitivity according to the acoustic background noise level. Such a device is simple to operate, and is used to monitor crying babies, telephones, doorbells, and burglar or smoke alarms.

While threshold devices are generally more flexible than hard-wired systems, proper setting of the device sensitivity is frustrating to many users. Adjusted too high, the device is likely to miss the occurrence of warning sounds. A low threshold setting makes the device vulnerable to false triggering.

Threshold detection systems using electromagnetic field sensing detect only warning sounds emitted by electromechanical actuators, for example telephones and doors equipped with electromechanical bells. When an electromechanical bell is activated, a strong time-varying electromagnetic field is produced to activate an internal electromechanical vibrating system. Consequently, this vibration generates a loud sound. For the purpose of warning sound detection, the stray electromagnetic field emitted by many devices may be utilized. For example, with a suction cup electromagnetic field pickup coils may be attached to the telephone or bell housing to intercept part of the time-varying magnetic field. The output of the pickup coil is amplified and fed to an appropriate threshold detection circuit. To alert hearing disabled individuals, such systems provide outlets for lamps and external vibratory actuators.

Since some warning devices are usually installed out of reach inside houses and offices (for example, fire alarms), the installation of the field pickup coils may be difficult. Due to low signal levels, special care is needed in handling the wiring connection between the pick-up coil and the threshold detector circuit. In addition, many newer appliances use solid-state buzzers which do not generate any magnetic field. Nevertheless, electromagnetic field sensing is a reliable method if employed under the appropriate circumstances.

1.2.3 Hearing Ear Dogs

While a Hearing Ear dog is not a technological device, it is included here to underscore the seriousness of the problem, and the complex and expensive solutions that are being offered. The Hearing Ear dog program was originally funded by the U.S. Government to meet the special needs of hard of hearing and deaf people. An affiliated program was established in Ontario, Canada and is named the Hearing Ear Dogs of Canada. Only mature hearing-disabled individuals are qualified recipients of Hearing Ear dogs. In the

U.S., the expenditure involved in dog selection, veterinary care, housing, training, and placement are fully subsidized by the U.S. Government. Hearing Ear dogs are trained to alert their owners to warning sounds commonly found in the living environment.

Dogs chosen from pet adoption offices are extensively screened prior to the rigorous four to five months of training. During this training, the Hearing Ear dog learns obedience, and how to respond to sounds emitted by household appliances and warnings. The Hearing Ear dogs can reliably recognize warning sounds they are trained for, and will skillfully alert their owners. In addition, the Hearing Ear dog usually is an ideal companion for elderly people.

The Hearing Ear dog approach to the problem also has some negative aspects. The training and dog placement processes are lengthy and costly, and the program often has a very long list of applicants wanting dogs. Moreover, since the training of Hearing Ear dogs requires special skills, once a placement is made recipients cannot teach their dogs to learn new warning sounds. The maintenance of the dogs is a costly proposition, and their transportation also creates problems. Furthermore, the presence of animals is not always tolerated in offices, hotels and other public places.

1.3 Project Objectives

Existing auditory warning aids for hearing-disabled people suffer from various functional deficiencies. Such deficiencies include lack of portability, lack of flexibility in recognizing warning sounds, and the propensity for false-alarms. In a recent survey [1], hearing-disabled people have expressed their desire for personal warning sound recognition systems which are easy to operate, and which are able to distinguish different household warning and emergency sounds. The demand for a versatile WARNing Signal Identification System (WARNSIS) which satisfies such needs is well established.

Motivated by this demand, by recent advances in speech recognition technology, and by the availability of specialized VLSI processors, it has been our objective to develop a real-time, adaptive WARNSIS which meets the following design criteria:

1. To be "teachable", which means that the device must be able to learn new warning sounds, and recognize them after a training procedure;

2. Have a recognition performance that is similar to that of normally hearing adults in very noisy environments; and

3. To produce acceptable positive and negative false-alarm rates in use.

In order to achieve this goal, work was undertaken to :

1. Investigate the characteristics of the warning sounds commonly used in office and living environments;

2. Utilize the results obtained in 1. to develop a recognition technique which has high reliability under noisy conditions;

3. Implement a prototype WARNSIS embodying the recognition technique developed in 2; and

4. Evaluate its overall performance in different noisy environments.

1.4 Thesis Outline

In Chapter 2 the literature on the various warning devices is reviewed. Industrial standards for the output power and spectral characteristics of warning sound generators are also discussed. Chapter 3 investigates the timing and spectral features of some common auditory warning sounds. Chapter 4 reviews different speech recognition techniques,

with detailed discussion of the filter-bank approach used in this work. The details of our WARNSIS implementation are presented in Chapter 5, and the evaluation of the system performance is contained in Chapter 6. Chapter 7 gives the conclusion and recommendations for further improvement in system performance.

Chapter 2

Warning Sounds and Generating Devices

2.1 Types of Warning Signal Generating Devices

Devices which generate audible warning signals employ either electro-mechanical or solid state transducers. Electro-mechanical warning devices generally include a metallic gong, hammer and coil assembly. To activate such a device, its coil is electrically energized, causing the hammer to vibrate and to strike the gong. The tonal quality and loudness of these devices depend upon the various components in the electro-mechanical assembly. Such are the shape and size of the gong(s), the force with which the hammer strikes the gongs, and the mounting and housing enclosure. In addition, in the manufacturing process, the mechanical components are assembled with fairly large tolerances. Therefore, the characteristics of the sound generated by such devices vary significantly, even for different units of the same model.

In the devices which employ solid-state transducers, warning sounds are elicited by applying electric voltage waveforms to these components. The tonal quality and loudness of such devices depend on the characteristics of these waveforms, and of the frequency response of the transducers. The waveforms are produced by electronic circuits, and therefore their characteristics can be easily manipulated. Since transducers are manufactured to close tolerances, the characteristics of the sounds generated by these electronic warning devices vary very little, even for different units of the same model.

2.2 Industrial Standards for Warning Devices

2.2.1 Sound Output Power

Conceptually, warning sounds should be sufficiently loud to be effective in generating attention among people in the vicinity of the warning device. Based on this concept, various standard organizations [1] established recommendations for the sound output power of smoke alarm detectors [5], household fire warning and burglar alarm systems [6], vehicle alarm systems [7], telephone rings [8,9,10] and general audible signalling devices used for life safety and property protection [11]. In general, it is recommended that in non-industrial environments an auditory warning device operated at rated voltage, and mounted in its intended position, be capable of providing an output sound pressure level (SPL) at least 85 dBA (with reference to 20 μ Pa) measured at a distance of 10 feet from the device [12].

More specifically, the minimum recommended SPL for warning devices depends on the environment where these devices are installed. If the warning devices are used in public places, a minimum of 15 dBA SPL above the average ambient sound level is required. If the devices are intended to be used in private residences, these devices should produce a minimum of 10 dBA SPL above the average ambient sound level [11].

2.2.2 Frequency Specification

Our survey of the publications of five major standard associations led us to conclude that no specific guidelines on frequency content of general warning sounds has been established. The only exception is the telephone, whose required acoustic output power and frequency content are specified by the CSA, EIA, ANSI and Bell Laboratories.

[1] Canadian Standards Association (CSA), the Electronic Industries Association (EIA), the Underwriters Laboratories Incorporated (UL), the American National Standards Institute (ANSI), and the National Fire and Protection Association (NFPA) of the U.S.

2.3 Literature on Warning Sound Characteristics

2.3.1 Telephone Rings

Telephone ringers are designed to produce easily recognizable alerting sounds. The available standards are applicable to telephones with electromechanical, or bell-type, alerting ringers, and with modern electronic tone ringers [8,9,10]. The important performance characteristics specified by these standards are summarized for our purposes as follows:

1. The alerting signal of a telephone with an electro-mechanical alerting device shall contain two or more major frequency components (f1 and f2) in the 500 – 6000 Hz range, with at least one having a mean power level of $\geq$ 73 dB, relative to 1 pW. The second major component shall have a mean sound power level of $\geq$ 68 dB, relative to 1 pW;

2. The total mean acoustic power level shall be $\geq$ 80 dBA, relative to 1 pW. These power levels apply with the volume control set for maximum volume;

3. At least one of the major component (f1) shall be below 2000 Hz. The nominal frequency of the higher major frequency component (f2) shall be equal to or greater than 5/4 of the lower major frequency component (f1), i.e., f2 $\geq$ 5/4 f1;

4. The alerting signal of a telephone with an electronic alerting device that does not produce an acoustic spectrum rich in overtones shall meet the criteria in 1), with the exception that f1 and f2 shall each have a mean power level of $\geq$ 73 dB, relative to 1 pW;

5. A telephone shall have a loudness adjustment accessible to the user that produces at least of a 6 dBA total attenuation when operated from its high to low volume position; and

6. With regard to ringing cycles, ringing current supplied by telephone company central office shall belong to one of the following sequences :

   - Repetitive bursts of 2 seconds out of every 6 seconds where an individual burst may be as short as 0.8 second;
   - Repetitive bursts of 1 second out of every 4 seconds where an individual burst may be as short as 0.6 second; or
   - Repetitive bursts of at least one ringing burst of a minimum 0.5 second duration in any 4 second period.

2.3.2 Smoke Detector Alarm Sounds

Smoke alarms are used to alert people to the presence of smoke and to the potential of fire. Generally, this warning sound is very strident and insistent. In a study of alarm sound attenuation inside residential buildings Halliwell and Sultan [13] investigated the spectral content of the sounds produced by a number of smoke detectors. Using a 2-channel FFT analyzer connected to two microphones, they obtained the short-time spectra, and for each sound 64 of these short-time spectra were averaged to give the spectrum. The narrow-band spectrum was subsequently converted to a third-octave spectrum by simply summing the energy within the third-octave bands. Their results for various smoke detectors show two or more strong spectral components in all computed spectra [Table 2.1]. Unfortunately, this work did not include the investigation of the variation of the short-time spectra obtained from consecutive samples.

Table 2.1: Spectral analysis results for different smoke detectors [13]

Detector Type†	1/3 Octave Frequency Bands (kHz)										
	0.5	0.63	0.8	1.0	1.25	1.6	2.0	2.5	3.15	4.0	5.0
A1	38‡	39	39	39	**63**	57	73	**96**	84	63	50
A2	37	38	38	38	44	56	70	**98**	92	67	56
B1	**82**	**82**	60	71	74	81	79	**95**	**95**	**95**	88
B2	79	**81**	66	72	76	81	77	93	94	**96**	92
C1	44	44	44	45	45	50	61	79	**102**	90	69
C2	44	44	44	45	45	50	62	79	**102**	91	70
D1	46	46	46	46	47	52	63	80	**103**	93	71
D2	44	44	44	45	45	50	62	80	**102**	88	68
E1	**84**	70	69	**85**	76	**92**	88	**96**	92	91	80
E2	76	**83**	63	69	80	**87**	85	97	**100**	91	89
F1	61	60	72	70	70	74	**86**	75	83	**90**	82
F2	58	61	69	70	72	77	**90**	81	82	**89**	82
G1	37	37	37	38	39	50	63	88	**95**	69	55
G2	38	38	38	38	39	48	61	84	**95**	71	56

† : Detectors with same letter denote identical model.
‡: Maximum Sound Power Output in dB

2.3.3 Warning and Alarm Sounds Generated by Vehicles and Traffic Control Devices

Miyazaki and Ishida [14] have studied the spectral characteristics of traffic alarm sounds commonly used in Japan. Such include sounds produced by electric horns used in passenger cars, small, middle size buses and trucks; air horns used in large buses, heavy duty trucks, and trailers; sirens used in emergency vehicles; horns used in rail-road crossing; and traffic noises.

Their observations have only limited value for us since they neither give description of the techniques used nor do they specify the type (short-time or long-time average) of the spectra obtained. Table 2.2 summarizes their results. They conclude that traffic-alarm-sounds have sharp line spectra, whereas ambient traffic noise is wide-band random noise.

Table 2.2: Summary of spectral analysis results for traffic alarm sounds [14]

Traffic Alarm Devices	Installed Vehicles	Major Frequency Features
Electric horn	Passenger cars, small, middle size busses trucks	basic resonant frequency at 300 Hz - 500 Hz, dominant harmonics at 2.0 - 4.0 kHz
Air horn	large busses, heavy duty trucks, trailers	dominant peaks at 300 - 500 Hz
Siren	Emergency vehicles	dominant peaks at 700 - 2000 Hz
Rail-road crossing		2 - 3 dominant peaks at 2.0 - 4.0 kHz
ambient traffic noise		broadband noise below 300 Hz

In British Columbia, and typically in North America, three types of emergency vehicle siren sounds are used: the "hi/lo" sound, the "yelp" sound, and the "wail" sound.

The hi/lo sound is usually found on most ambulances. It consists of two alternating tones, and with the pattern repeating about once per second. Two commonly used tone pairs are 690/920 Hz and 520/1520 Hz. The wail sound is a slow changing tone between two preset tone frequencies. A typical example is the wail sound used by police motorcycle sirens with preset tone frequencies at 500 Hz and 1460 Hz, and a repetition rate of 10 cycles per minute [15]. The yelp sound is a fast changing tone between two preset tone frequencies. A typical example is the electronic siren produced by Southern Vehicle Products Inc., which provides a yelp sound with preset tone frequencies at 600 Hz and 1350 Hz, and a repetition rate of 3 to 5 cycles per second [16]. The yelp and wail sounds are used by both fire-trucks and police cars.

2.4 The Emerging Scientific basis for Generating Warning Sounds

While warning sounds have been used for a long time, many of these are based on subjective opinions as to what is "best". Only recently was any scientific work done to determine what sound characteristics will elicit optimal responses under varying circumstances. Such work is particularly relevant for us, since in the future warning devices may follow a more systematic approach to sound generation than it has been the case until now.

2.4.1 A Generic Warning Sound Generating Scheme

According to the work of Patterson and his colleagues, a warning sound need not to be excessively loud, but its amplitude must depend on the background noise level. They have demonstrated, that in order to hear sounds reliably in noise, some spectral components must be between 15 dB and 25 dB above the masked threshold [17,18]. Lower and Wheeler [19] has developed a desk-top computer program to estimate this

background threshold. With the estimated background threshold, the spectral component amplitude of the warning sound can be determined. This approach had been used to study the intense background noise of military helicopters in the U.K. [20]. With regard to the frequency content of the warning sound, Patterson [17] limits it to the range between 0.5 kHz and 5.0 kHz.

Based on these spectral amplitude and frequency limits of the warning sounds, a pattern of pulsative sounds which is distinctive and resistant to undesirable noise contamination, was constructed by Patterson [17,22,23]. As shown in Fig. 2.1 , this prototype warning sound basically consists of a sequence of bursts each of which is made up of a sequence of pulses. Different degrees of perceived urgency can be manipulated by simply varying the characteristics of the pulse sequences.

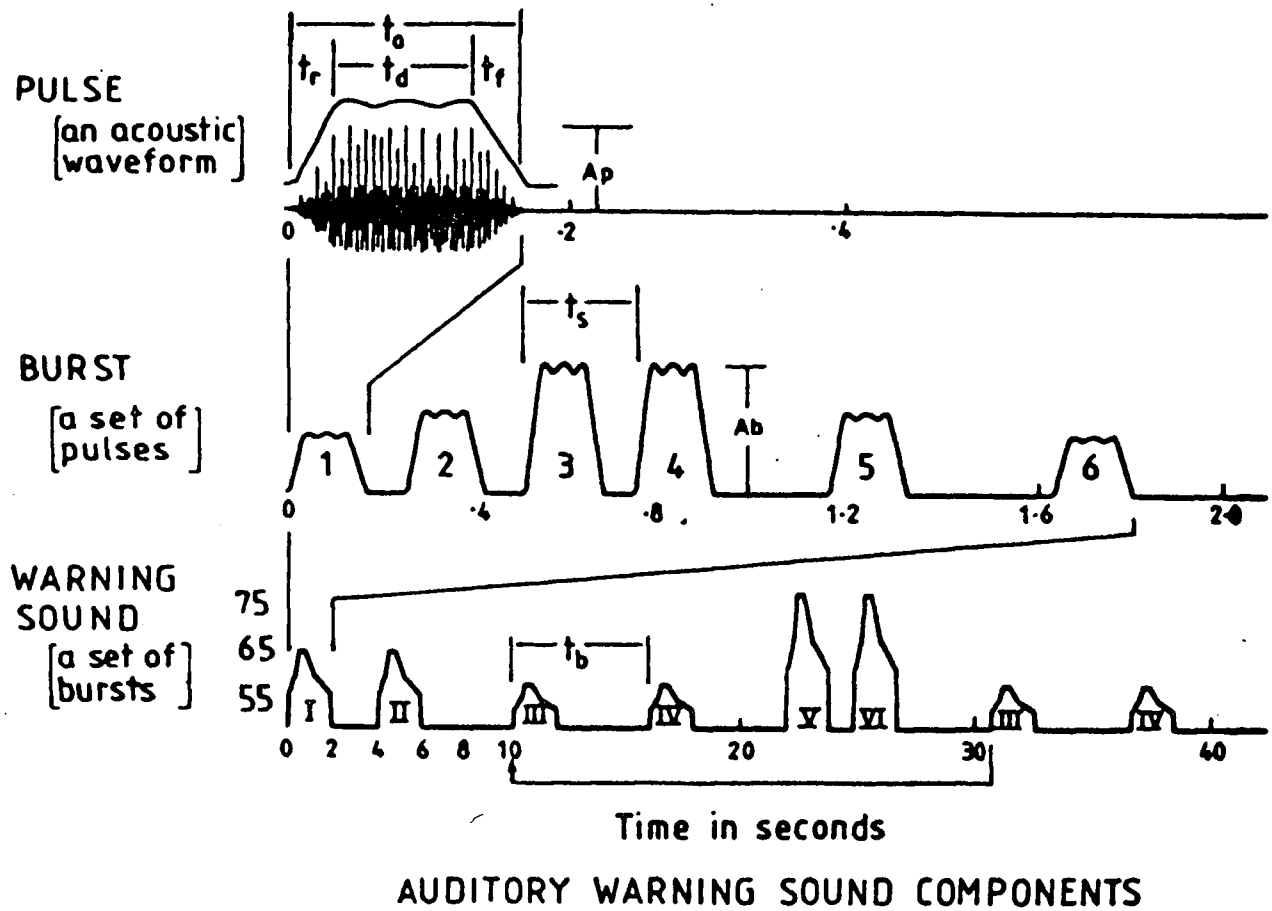
In Patterson's work, the pulse design starts with measurement of the ambient noise spectrum. Then, the warning signal spectrum is determined by setting all its components 15 – 25 dB above the corresponding ambient noise spectral values. In order to avoid excessive peak factors in the signal waveform, sine or cosine phase is assigned to the spectrum. Consequently, the pulses are generated by applying the Inverse Fast Fourier Transform. These pulses vary in duration from 75 msec to 200 msec in accordance with the guidelines set down by Patterson [17,23]. Also, the pulses are gated with sinusoidal ramps at both ends in order to avoid uncontrollable transients. At this stage, by varying the fundamental frequency, and the relative weight of high and low frequencies of the pulses, any degree of perceived urgency can be designed. Usually, greater urgency is signalled by higher fundamentals, and by relatively more high frequency energy.

A burst is produced by assembling three-to-nine copies of the basic pulse. By changing the elapsed time between the start of one pulse, and the start of the next, distinct pitch and temporal patterns may be created. By varying the amplitude of

the pulses different loudness patterns may be obtained. The perceived urgency is generated by changing the overall pitch, the speed and the loudness pattern of the pulses. In general, a burst with a high pulse rate will convey greater urgency than a burst with a low pulse rate. A rising pitch-contour can produce a more urgent burst than a falling pitch-contour. Additionally, an urgent burst will remain at, or near, the maximum loudness while a less urgent burst will decrease in loudness towards the end of the burst.

Such bursts serve as templates from which warning sounds may be synthesized. The amplitude variations and spacing of the bursts are determined experimentally. The criterion is that the resulting warning sound should effectively convey the desired specific warning message to personnel in the vicinity without activating their startling reflex.

Patterson successfully implemented this scheme on warning systems of commercial aircrafts and military helicopters [17]. A slight modification of this scheme was also adopted for medical equipment used in intensive-care units and operating theatres of hospitals in the U.K. [22,23].



AUDITORY WARNING SOUND COMPONENTS

Figure 2.1: Auditory Warning Sound Components [17,21,23].

Chapter 3

Measurement and Analysis of Timing & Spectral Characteristics

As we have seen it in Chapter 2, the literature on warning sounds yields little useful information on their timing and short-time spectral characteristics. Since it is the purpose of this work to apply timing and short-time spectral analysis techniques to systematically extract the unique identifying characteristics of these warning sounds in real-life environments, such information is essential for us. Specifically, the detailed knowledge of warning sound characteristics provides the basis for the exploration of different signal recognition schemes.

3.1 Timing Characteristics

The objective of this part of our work was to derive useful information on the timing of warning sounds from measurements of signal waveforms. For this purpose we used telephone rings, siren sounds, and smoke alarm sounds. Telephone rings were generated by both electro-mechanical and electronic ringers; siren sounds were produced by an electronic siren driver; and the smoke alarm sounds were obtained from a commercial smoke alarm.

3.1.1 A PC-Based Data Acquisition System

To obtain quantitative data, a PC-based data acquisition system was designed and constructed. This system accepts the instantaneous absolute amplitude waveform of the signal, and transforms it into the short-time average absolute amplitude (STAAA)

waveforms. Then, the transformed waveforms are stored for plotting. The instantaneous amplitude and the short-time average variations in absolute amplitudes of the signal are given in Table 3.3, where $x(n)$ represents the discrete instantaneous signal amplitudes, and N denotes the number of samples accumulated.

Table 3.3: Instantaneous and short-time signal amplitudes

	signal amplitudes	absolute signal amplitudes
instantaneous	$x(n)$	$\|x(n)\|$
short-time average	$\frac{1}{N}\sum_{n=1}^{N} x(n)$	$\frac{1}{N}\sum_{n=1}^{N} \|x(n)\|$

The instantaneous absolute signal amplitudes are generated by hardware, and the derivation of the short-time average absolute signal amplitudes, and storage of these derived samples is accomplished by software.

Fig. 3.2 shows the block diagram of the method used to generate the discrete instantaneous absolute signal amplitudes. Basically, sounds are collected by a suitable microphone, are pre-amplified by a low-noise voltage amplifier, and are low-pass filtered prior to input to a full-wave rectifier. The output from the full-wave rectifier gives the instantaneous amplitude of the waveform. Then, an 8-bit A/D converter samples this waveform at 10 kHz. Consequently, the digitized sample is stored temporarily in an output buffer until the 8-bit microprocessor (INTEL 8088) is ready to accept the data via a bi-directional bus. In addition, a LED bar graph is used to display the variations in the instantaneous absolute amplitudes of the signal waveforms.

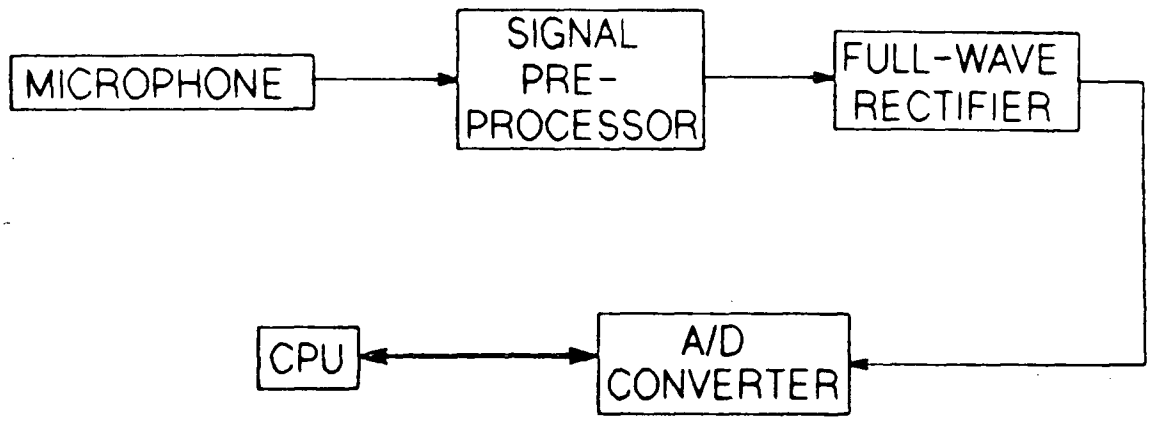


Figure 3.2: Signal acquisition and derivation of instantaneous absolute signal amplitudes

In this implementation, the short-time average absolute signal amplitudes are derived from 12.8 msec accumulation of the instantaneous absolute signal amplitude samples (A/D converted data). With these instantaneous signals sampled at 10 kHz, a sample of the short-time average absolute signal amplitudes can be obtained by summing 128 of the instantaneous signal samples. In order to avoid the problem of overflow during the accumulation process, a 16-bit register is used to accumulate this sum. Consequently, a sample of the short-time average absolute signal amplitudes is obtained by dividing the 16-bit register content by the total number of accumulated samples (i.e 128 in this case). The resulting quotient is then rounded to eight bits to provide the short-time average absolute signal amplitude sample which is transferred to a designated file. This file stores 1000 bytes. These data manipulation and transfer procedures are repeated until the data file is completely filled with 1000 samples (equivalent to

12.8 sec of the signal waveform). The program to handle this data manipulation and transfer in real-time was written in INTEL 8088/8086 assembly language. A flowchart of these operations is shown in Fig. 3.3.

3.1.2 Data Collection

With this data acquisition system, we collected data on the absolute amplitudes of warning sounds in the normal acoustic environment of our laboratory. Fig. 3.4 shows the experimental set-up. The siren horn produced siren sounds; and a radio cassette player provided the pre-recorded telephone rings and smoke alarm sounds. A sound pressure level (SPL) meter placed aside the microphone measured the SPL variations of the environment throughout the data collection process.

The SPL meter was set to "C" weighting and "SLOW" response, because the "C" weighting network of the SPL meter has a flat frequency response similar to that of the signal processing circuit of the data acquisition system; and the "SLOW" response provides an average of 1.0 sec of the acoustic energy variations of the environment.

Based on the SPL measurements in the absence and during the presence of warning sounds, the signal-to-noise ratio (SNR) could be deduced. SNR, in this work, is defined as the ratio of peak signal power to peak noise power. Noises, in this thesis, are defined as all sounds other than warning sounds. Such unwanted sounds may include steady and transient random noises, radio broadcasts, or surrounding conversations. A detailed derivation of the relationship between the SNR and SPL measurements is given in Appendix A.

Data on absolute amplitudes of warning sounds were collected in two different background environments. The first set of data were collected in a steady random noise background which originated from a ventilation fan of a PC-computer. Such noise is typical for office environments. A value of 60-62 dBC was recorded throughout the data

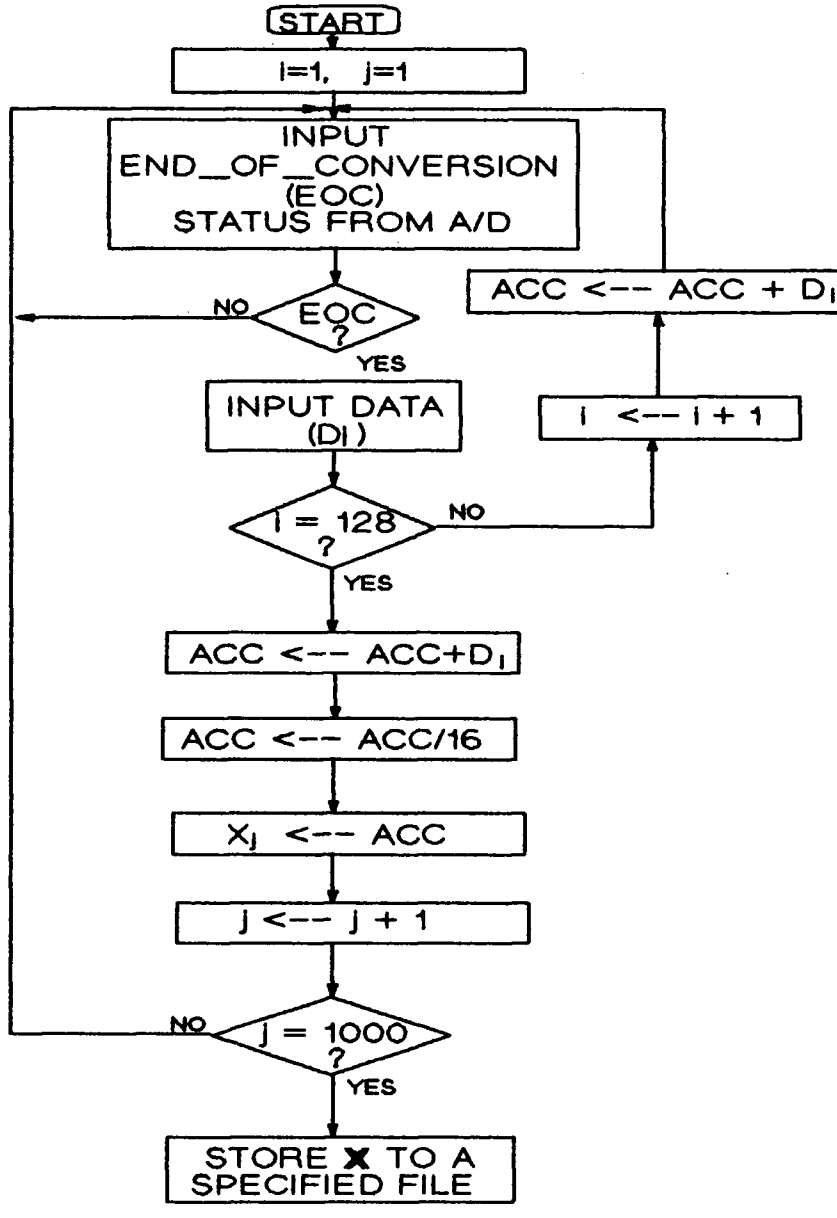


Figure 3.3: Flowchart of procedure to accumulate and store 1000 samples

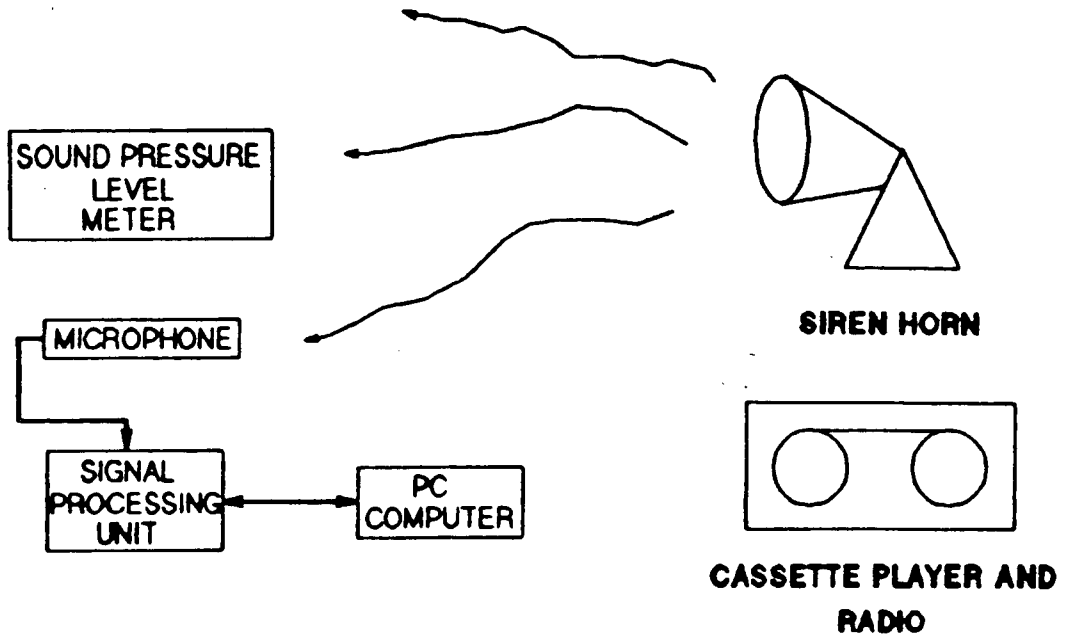


Figure 3.4: Experimental set-up for data collection

collection process. This set of data were referred to as "clean", because the SNR was maintained at least over 20 dB. To study the effect of more complex background noises on warning sounds, the second set of data were collected at a SNR of 10 dB. The background noise sources consisted of both steady random noise, radio music broadcast, and speech.

To establish the short-time average absolute amplitude profiles of the various noise sounds (without warning sounds present), a third set of data was also collected. This included all the noise sources used above, and the noise SPL was the same as that used in the SNR measurements.

3.1.3 Timing Features of Different Warning Sounds

The plots of the first set of data are shown in Fig. 3.5, Fig. 3.6, Fig. 3.7 and Fig. 3.8. Since the purpose of these measurements is to establish the time variations of the short-time average absolute signal amplitudes, the actual value of these amplitudes is of no particular interest. Therefore, the vertical axes show a relative scale without units.

The following observations may be drawn from these figures:

1. Fig. 3.5, Fig. 3.6(b) & (d) (siren sounds), and Fig. 3.7(a) & (b) (telephone rings) show on-off type repetitive patterns of warning signal bursts; Fig. 3.6(a) & (c) (siren sounds), and Fig. 3.7(c) (smoke alarm sound) display the steady sounds;

2. Fig. 3.5 (a) and (b) show devices which produces sounds with very similar temporal structures, but with different repetition rates;

3. Fig. 3.5 (d) is a two-tone siren sound, and its amplitude contour can be characterized by i) a transition from background level amplitudes, and ii) a repetitive

on-off pattern representing two tones of different intensities (for other siren sounds or telephone rings, the off-patterns represent the background noise levels);

4. The width of the bursts of these waveforms varies from 102.4 msec to 3.24 sec;

5. The repetition period of on-off patterns ranges from 140 msec to 5.86 sec;

6. Steady sounds are characterized by signal level transition to higher steady amplitude level; and

7. Contours of the average of short-time absolute signal amplitude of radio broadcasts (Fig. 3.8) consist of random, nonrepetitive sequences of signal bursts.

The plots of the second set of data are shown in Fig. 3.9 and Fig. 3.10. Comparative examination of these plots yields the following observations:

1. For short-burst, such as (a), and (b) in Fig. 3.9, and (d) in Fig. 3.10, the introduction of radio broadcast background alters the baseline levels, and smooths out the weak peaks of the "clean" signals; however it produces no significant change in relative timing between consecutive amplitude peaks of the waveforms;

2. For signals with long silence intervals($\geq$ 400 msec) such as (c), and (d) in Fig. 3.9, and (b) in Fig. 3.10, spurious small peaks appear randomly during these intervals; and

3. The repetition rate of the on-off patterns of burst-type sounds is unchanged by variations in background noise.

In summary, we can conclude from these measurements that the short-time average absolute amplitude contours provide unique timing information on both steady and burst-type sounds.

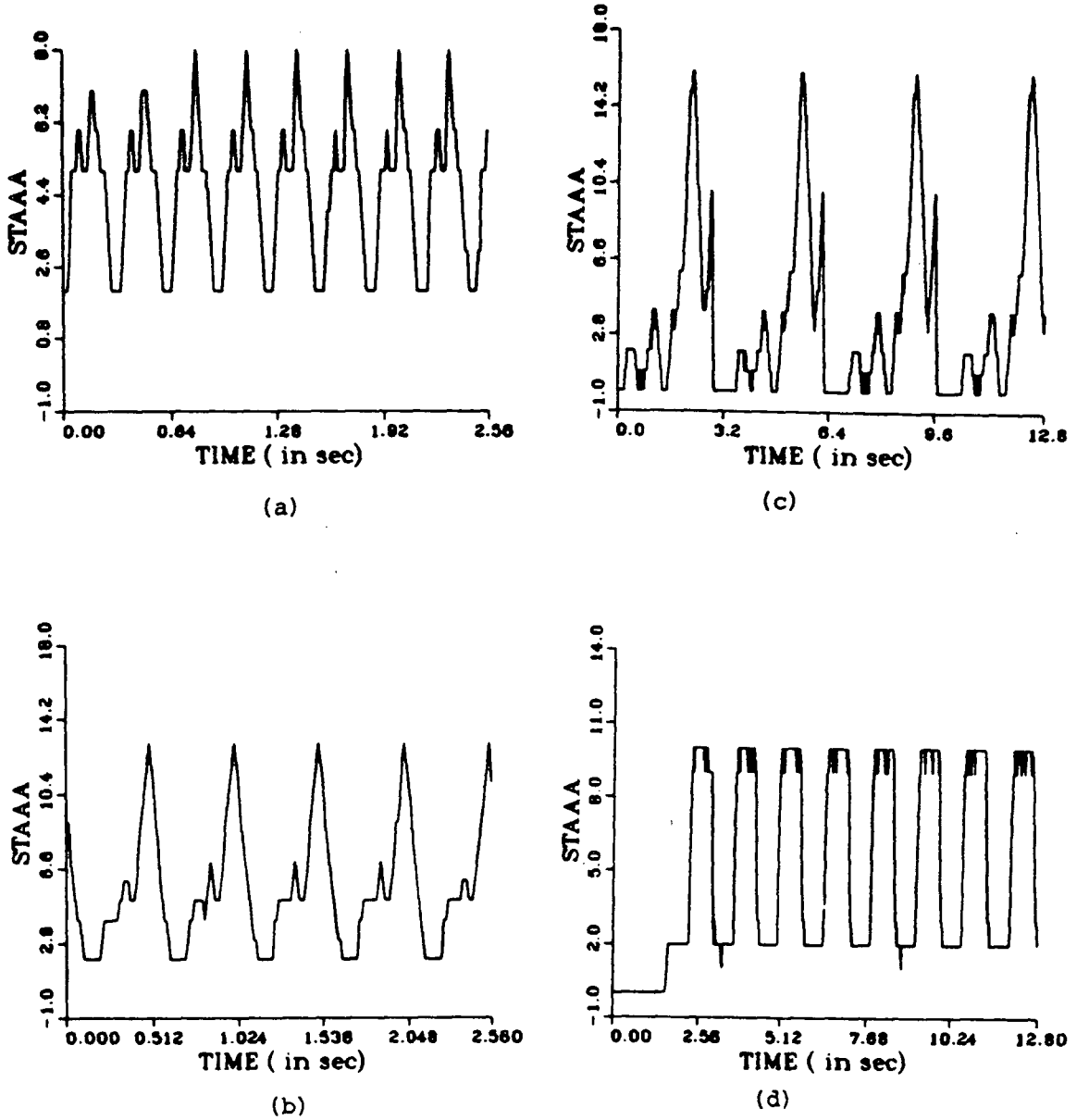


Figure 3.5: Short-time average absolute amplitudes (STAAA) of siren sounds: a) J1 : Burglar alarm (JDS-100); b) J2 : MPI-11; c) J3 : JDS-100 I; and d) J4 : HI-LO

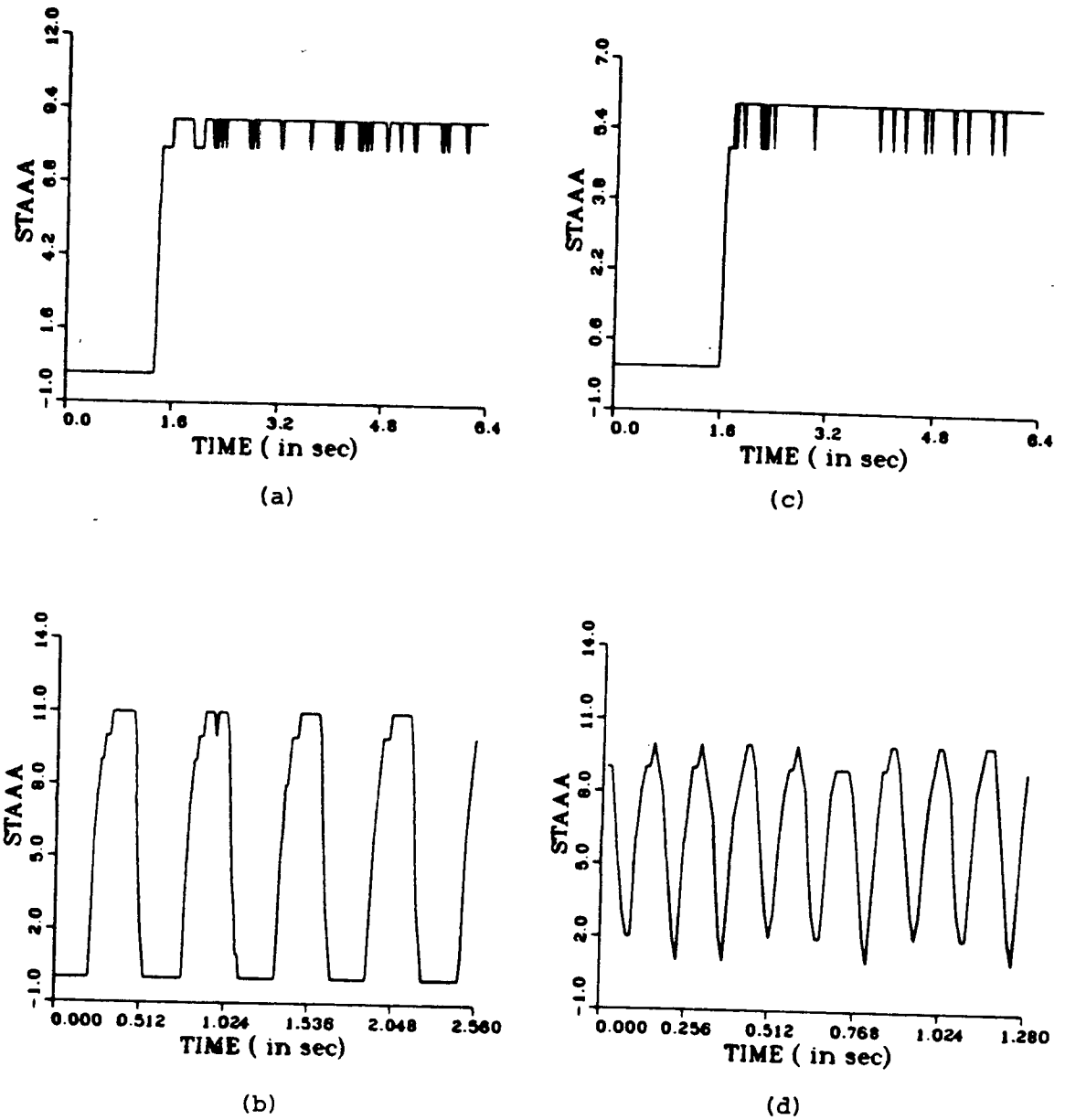


Figure 3.6: Short-time average absolute amplitudes (STAAA) of siren sounds: a) J5 : High steady sound; b) J6 : Pulser; c) J7 : Steady horn; and d) J8 : Electronic Synthesized Bell sound

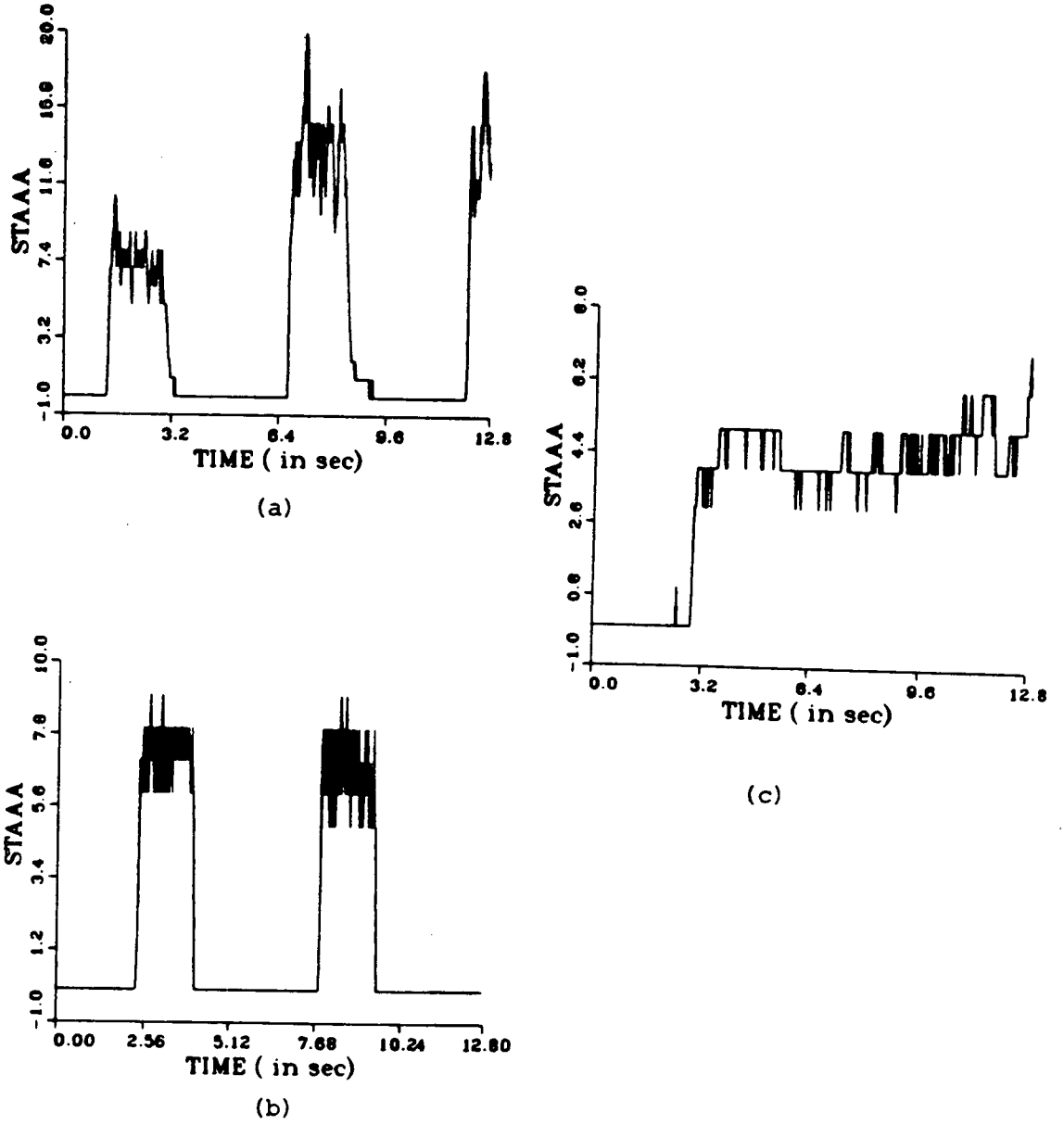


Figure 3.7: Short-time average absolute amplitudes (STAAA) of telephone rings and smoke alarm sound: a) Electro-mechanical Ringer; b) Electronic Ringer; and c) Smoke alarm sound

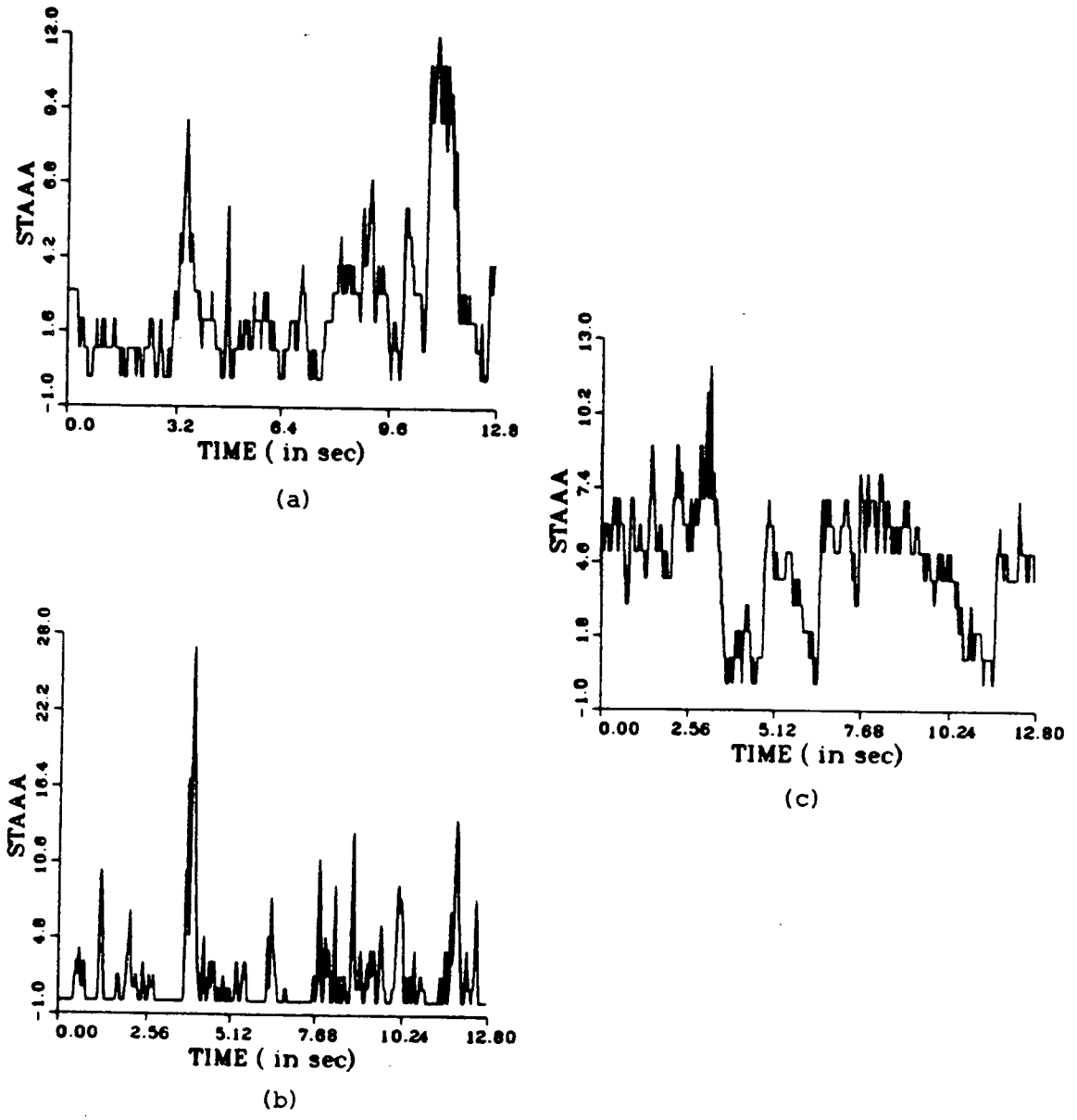


Figure 3.8: Short-time average absolute amplitudes (STAAA) of radio broadcasts a) Pop music; b) Speech; and c) Rock music

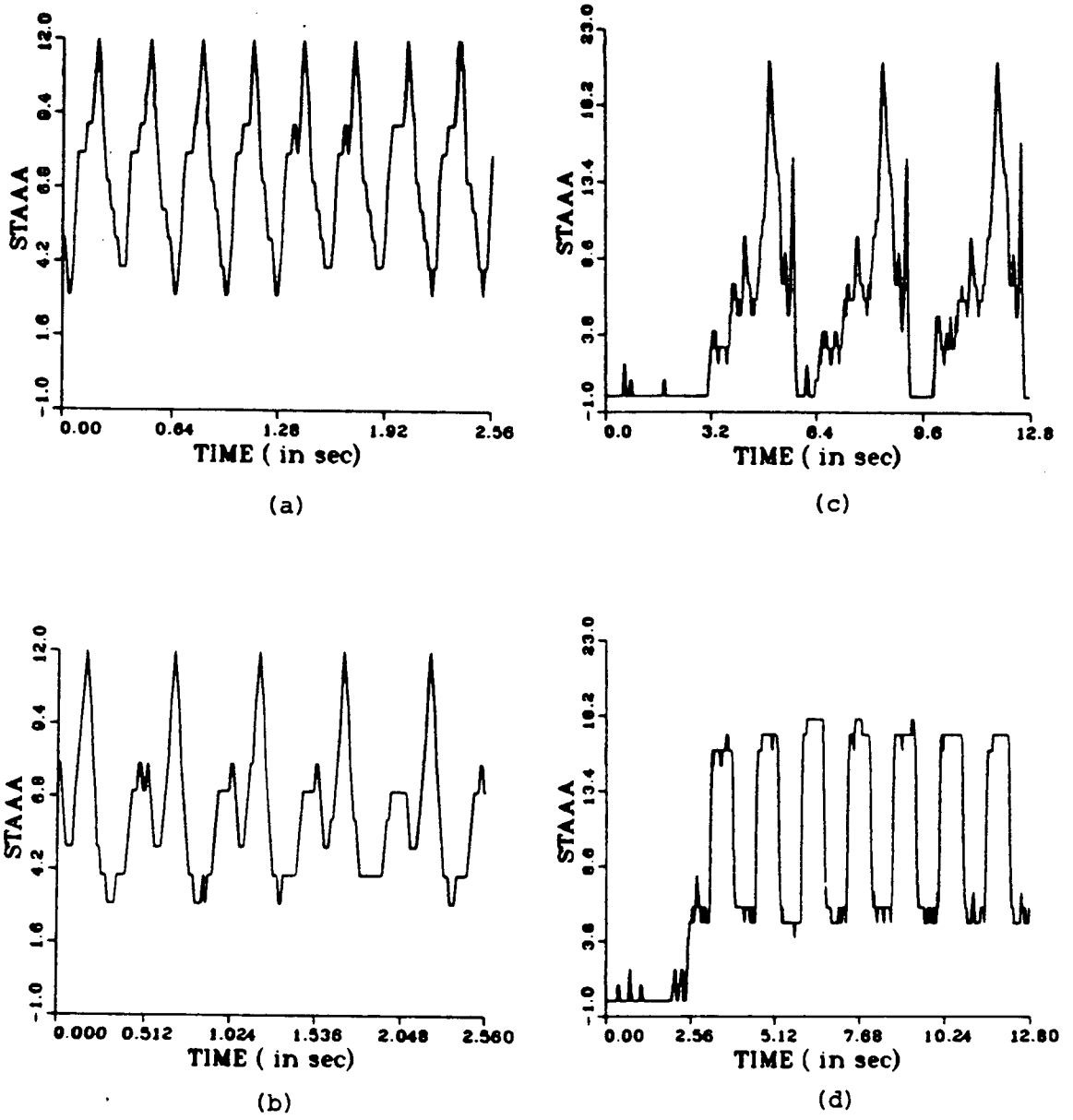


Figure 3.9: Short-time average absolute amplitudes (STAAA) of siren sounds with radio-broadcast as background: a) J1; b) J2; c) J3; and d) J4

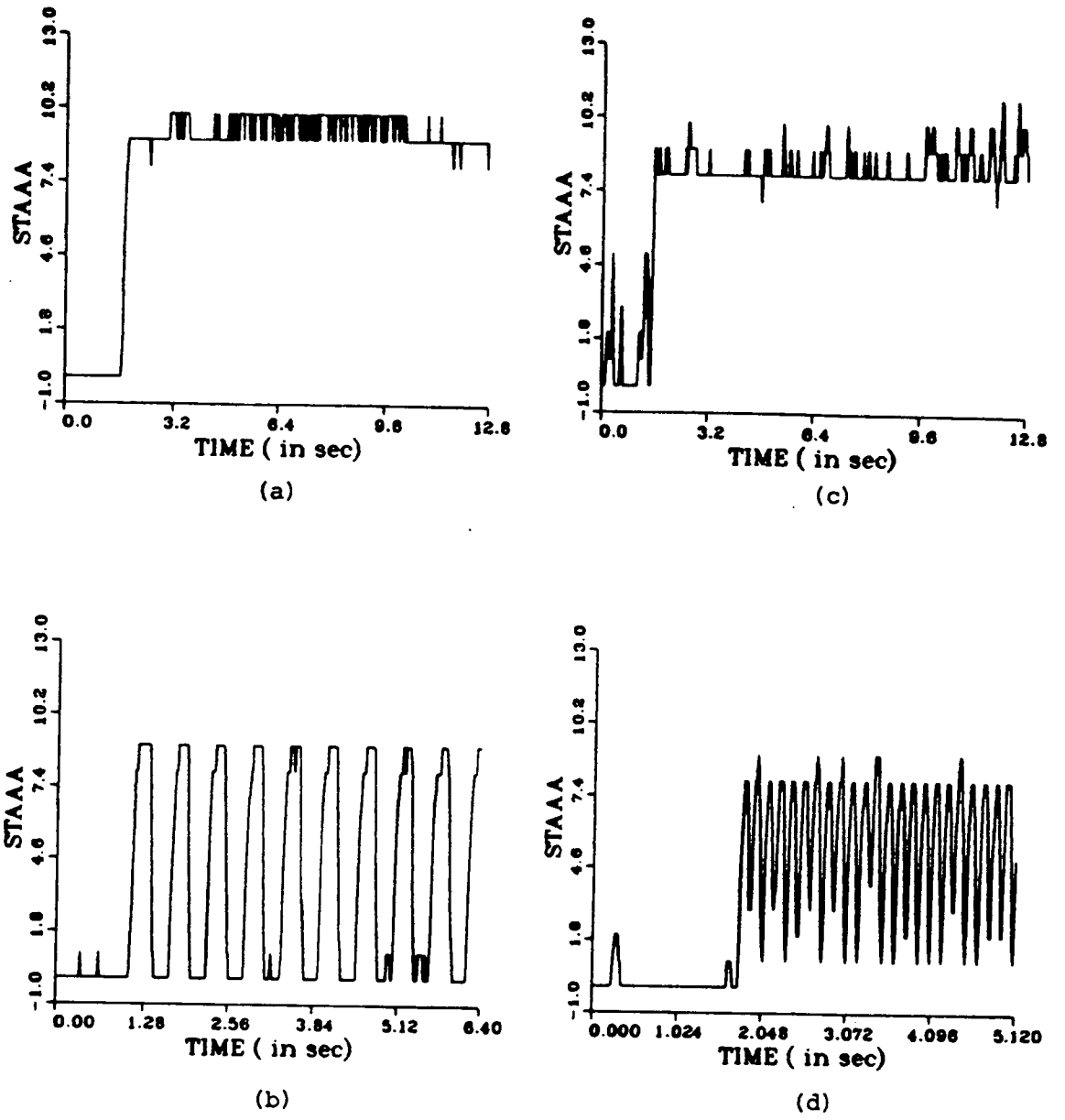


Figure 3.10: Short-time average absolute amplitudes (STAAA) of different siren sounds with same background noise: a) J5; b) J6; c) J7 ; and d) J8

3.2 Spectral Characteristics

Based on the assumption that the short data records deduced from the observed time sequences are ergodic, and that their estimated spectra are slowly time-varying, spectral estimation techniques provide an insight into the frequency contents carried by the observed time sequences. Generally, spectral estimation methods use either the parametric, or the nonparametric approach. A detailed exposition of many different algorithms used for obtaining waveform spectra was given by Kay and Marple [24].

In general, parametric spectral analysis involves three steps. The first step is to select a time series model, with assumed model order, for the observed data record. Time series models such as the autoregressive model(AR), the moving-average model(MA), or the autoregressive-moving average model(ARMA), are the most common choices for practical applications. For example, the linear prediction coding (LPC), or AR model with model order of 10-16, has been proven to be a very suitable choice for speech analysis and synthesis [25,26].

The second step is to estimate the model parameters using the available data samples [24]. Depending on the specific time series model selected, different algorithms may be applied for such parameter extraction. The third step is to compute the estimated spectra by substituting the specific parameter values derived in the second step into the theoretical power spectral density function of the model used.

The nonparametric spectral estimation approach assumes that the observed data record is produced from a set of sinusoidal components governed by the Fourier Series model of signals. Two popular and conventional spectral estimation techniques are the Blackman-Tukey [27] and the Welch's periodogram [28] methods. Both of these techniques employ the computationally efficient Fast Fourier Transform (FFT). A new, unified, FFT-based spectral estimation method, capable of producing more statistically

stable spectra with better frequency resolution than the conventional methods, has been proposed by Nuttall and Carter [29].

3.2.1 Comparison of Parametric and Nonparametric Spectral Estimation Methods

With relatively short data sequences recorded under high signal-to-noise (SNR) conditions, the parametric technique can produce smoother and finer frequency resolution spectra. Unfortunately, the parametric spectral estimation approach is susceptible to noise interference. Such degradation in performance of the AR model has been extensively investigated by Lim [30] and Kay [31].

The nonparametric spectral estimation approach is implemented in practice by the Discrete Fourier Transform (DFT). Since the DFT considers every data sequence to be periodic, such periodic extensions of the original data sequence exhibit discontinuities at the boundaries of the observed time interval. In the subsequent numerical analysis, these boundary discontinuities result in spectral leakage over the entire frequency spectrum. Harris [32] discussed the application of using various windows with nonuniform weighting to reduce this spectral leakage. This can be accomplished only at the expense of frequency resolution in the spectrum. Finally, to obtain a statistically stable spectrum, spectrum averaging of short-time spectra is definitely required [28].

In general, the frequency resolution of spectra obtained by the nonparametric spectral estimation approach is limited by the data duration, and is independent of the SNR of the signals. Theoretically, the frequency resolution of spectra is inversely proportional to the duration of the original data sequence. Since zero-padding of the data sequence before transformation effectively increases the signal duration, it has been a misconception that such a zero-padding procedure will improve the frequency resolution of the resultant spectra. As demonstrated in [24], zero padding is useful only for

1) smoothing the appearance of the resultant spectra via interpolation, 2) resolving potential ambiguities of computed spectra, and 3) reducing the "quantization" error in the accuracy of estimating the frequencies of spectral peaks. It is common procedure to apply windowing prior to the zero-padding of the data sequence.

3.2.2 Welch's Non-overlapping Spectral Estimation Method

For this work, we selected the conventional Welch's non-overlapping spectral estimation approach to investigate different warning sounds. The rationale behind this choice has four aspects.

First, most warning sounds usually maintain a regular rhythm, and continuous, long data records can be obtained. This allows spectral averaging, and results in the statistical stability of the computed spectra. Secondly, by Welch's spectral estimation technique is robust with respect to noise corruption of the signals, because the frequency resolution and the stability of the computed spectra are independent of the SNR. Thirdly, no a priori knowledge of a signal model for various warning sounds is needed. Finally, limitations inherent in Welch's spectral estimation method have been thoroughly studied, and techniques used to reduce discrepancies have been well explored [32].

Welch's non-overlapped spectral estimation technique may be described in four steps:

1. Consider a data sequence, $x(n)$ of length N, where $n \in [0, N-1]$, and divide N into K non-overlapped segments, each of which has an integral length of N/K, say M, and is denoted as $x_k(m)$, where $m \in [0, M-1]$, and $k \in [0, K-1]$.

2. Select an "DFT-even" window sequence [1], $w(m)$, with length identical to $x_k(m)$, and multiply this window sequence onto $x_k(m)$, giving $\hat{x}_k(m)$ as follows,

$$\hat{x}_k(m) = x_k(m)w(m) \tag{3.1}$$

3. Take the magnitude square of the windowed sequence to obtain the k^{th} segment discrete Fourier spectrum (often called modified periodogram) denoted as $S_k(l)$,

$$\begin{aligned} S_k(l) &= \frac{1}{MU}\left|\sum_{m=0}^{M-1} x_k(m)w(m)e^{-j\frac{2\pi ml}{M}}\right|^2 \\ &= \frac{1}{MU}\left|\sum_{m=0}^{M-1} \hat{x}_k(m)e^{-j\frac{2\pi ml}{M}}\right|^2 \end{aligned} \tag{3.2}$$

where U = window average power given by,

$$U = \frac{1}{M}\sum_{m=0}^{M-1}|w(m)|^2 \tag{3.3}$$

4. Compute $S_k(l)$ for $k \in [0, K-1]$, and obtain the average spectrum, $S_{avg}(l)$,

$$S_{ave}(l) = \frac{1}{K}\left\{\sum_{k=0}^{M-1} S_k(l)\right\} \tag{3.4}$$

[1] DFT-even window is a conventional even window sequence with the right-end point missing. [32]

Welch demonstrated that the variance of spectral estimates can be reduced by dividing a long original data sequence into finer segments. However, he also cautioned that the statistical bias generated by the estimation process increases linearly with increasing number of segments [28]. Therefore, the trade-off between the size of data segments and the amount of spectral variance reduction is to be determined by the user.

3.2.3 Implementation of Welch's Method

Since the DFT can accept complex input quantities, we may make use of this feature to establish an efficient scheme for the computation of average spectrum from two real data sequences. Such a scheme is implemented by the use of the FFT algorithm, and involves only a single pass of the FFT computation. The three steps of calculations are summarized as follows.

The first step is to substitute the real and imaginary parts of a complex input data sequence by two non-overlapped real data segments. Then, we take the DFT of this complex sequence, and after further calculations we can obtain the average spectra of the two non-overlapped data segments. The detailed mathematical derivations are given in [33], with the major steps summarized below:

1. Consider now $g(m)$ being a complex input data sequence whose real and imaginary parts are substituted by the two non-overlapped real data segments $x_1(m)$ and $x_2(m)$. Then, $g(m)$ can be expressed by,

$$g(m) = x_1(m) + jx_2(m) \tag{3.5}$$

where $m \in [0, M-1]$.

2. The DFT of $g(m)$ which is denoted as $G(k)$ is expressed by,

$$G(k) = \sum_{m=0}^{M-1} g(m)w(m)e^{-j\frac{2\pi mk}{M}} \tag{3.6}$$

$$= G_R(k) + jG_I(k)$$

where G_R = real part of DFT of $G(k)$

G_I = imaginary part of DFT of $G(k)$

$w(m)$ = "DFT-even" window sequence

$k \in [0, M-1]$

3. Now we take into consideration that given two real data sequences, $x_1(m)$, and $x_2(m)$, and a DFT-even window sequence, $w(m)$, for $m \in [0, M-1]$, the DFT of these windowed data sequences denoted as $X_1(k)$, and $X_2(k)$, respectively, can be represented by their real and imaginary parts given below:

$$X_1(k) = X_{1R}(k) + j\, X_{1I}(k) \tag{3.7}$$

$$X_2(k) = X_{2R}(k) + j\, X_{2I}(k) \tag{3.8}$$

where $X_{1R}(k)$ = real part of the DFT of $x_1(m)$

$X_{1I}(k)$ = imaginary part of the DFT of $x_1(m)$

$X_{2R}(k)$ = real part of the DFT of $x_2(m)$

$X_{2I}(k)$ = imaginary part of the DFT of $x_2(m)$

$k \in [0, M-1]$

It can be shown that,

$$X_{1R}(M-k) = X_{1R}(k) \tag{3.9}$$

$$X_{2R}(M-k) = X_{2R}(k) \tag{3.10}$$

$$X_{1I}(M-k) = -X_{1I}(k) \tag{3.11}$$

$$X_{2I}(M-k) = -X_{2I}(k) \tag{3.12}$$

Using the expression 3.5 for $g(m)$ in Eq. (3.6), we can express $G_R(k)$ and $G_I(k)$ in terms of the real and imaginary parts of $X_1(k)$ and $X_2(k)$:

$$G_R(k) = X_{1R}(k) - X_{2I}(k) \tag{3.13}$$

$$G_I(k) = X_{1I}(k) + X_{2R}(k) \tag{3.14}$$

If we substitute k by (M-k) into Eq.(3.13-3.14) and utilize the results obtained from Eq. (3.9-3.12), we obtain,

$$G_R(M-k) = X_{1R}(k) + X_{2I}(k) \tag{3.15}$$

$$G_I(M-k) = -X_{1I}(k) + X_{2R}(k) \tag{3.16}$$

4. The average spectrum, $P_{avg}(k)$ for $x_1(n)$ and $x_2(n)$ is given by,

$$\begin{aligned} P_{avg}(k) &= \frac{1}{2MU}\left\{|X_1(k)|^2 + |X_2(k)|^2\right\} \\ &= \frac{1}{2MU}\left\{|X_{1R}(k)|^2 + |X_{1I}(k)|^2 + |X_{2R}(k)|^2 + |X_{2I}(k)|^2\right\} \end{aligned} \tag{3.17}$$

where

$$U = \frac{1}{M}\left\{\sum_{m=0}^{M-1}|w(m)|^2\right\} \tag{3.18}$$

Therefore, by making use of the results obtained from Eq. (3.13-3.16) to solve for $X_{1R}(k)$, $X_{1I}(k)$, $X_{2R}(k)$, and $X_{2I}(k)$ in terms of $G_R(k)$, $G_R(M-k)$, $G_I(k)$, and $G_I(M-k)$, we can, subsequently, derive $P_{avg}(k)$ from the real and imaginary parts of $G(k)$. Thus, we can show that,

$$P_{avg}(k) = \frac{1}{4MU}\left\{G_R^2(k) + G_R^2(M-k) + G_I^2(k) + G_I^2(M-k)\right\} \tag{3.19}$$

In this work, warning signals were sampled at a rate of 20 kHz with 12 bit resolution. The non-overlapped data segment length was a multiple of 12.8 msec, or of 256 data samples. With regard to the specific window used to reduce spectral leakage, the minimum 4-sample Blackman-Harris window (Fig. 3.11), with – 92 dB highest sidelobe level, – 6 dB/octave sidelobe fall-off rate, and two frequency bins [2] of the equivalent noise bandwidth [32], was used to multiply onto each non-overlapped data segment. The actual spectral calculations were performed on a VAX 750 general computer. The flowchart of the program is given in Fig. 3.12.

3.2.4 Data Collection

In order to explore the variations of warning sound spectral characteristics, the sounds emitted by 1.) electromechanical ringers of five rotary dial phones, 2.) a multiple-line

[2] A bin is a basis frequency for a spectrum and is derived from the ratio of the signal sampling frequency to the total number of data points used in the spectrum.

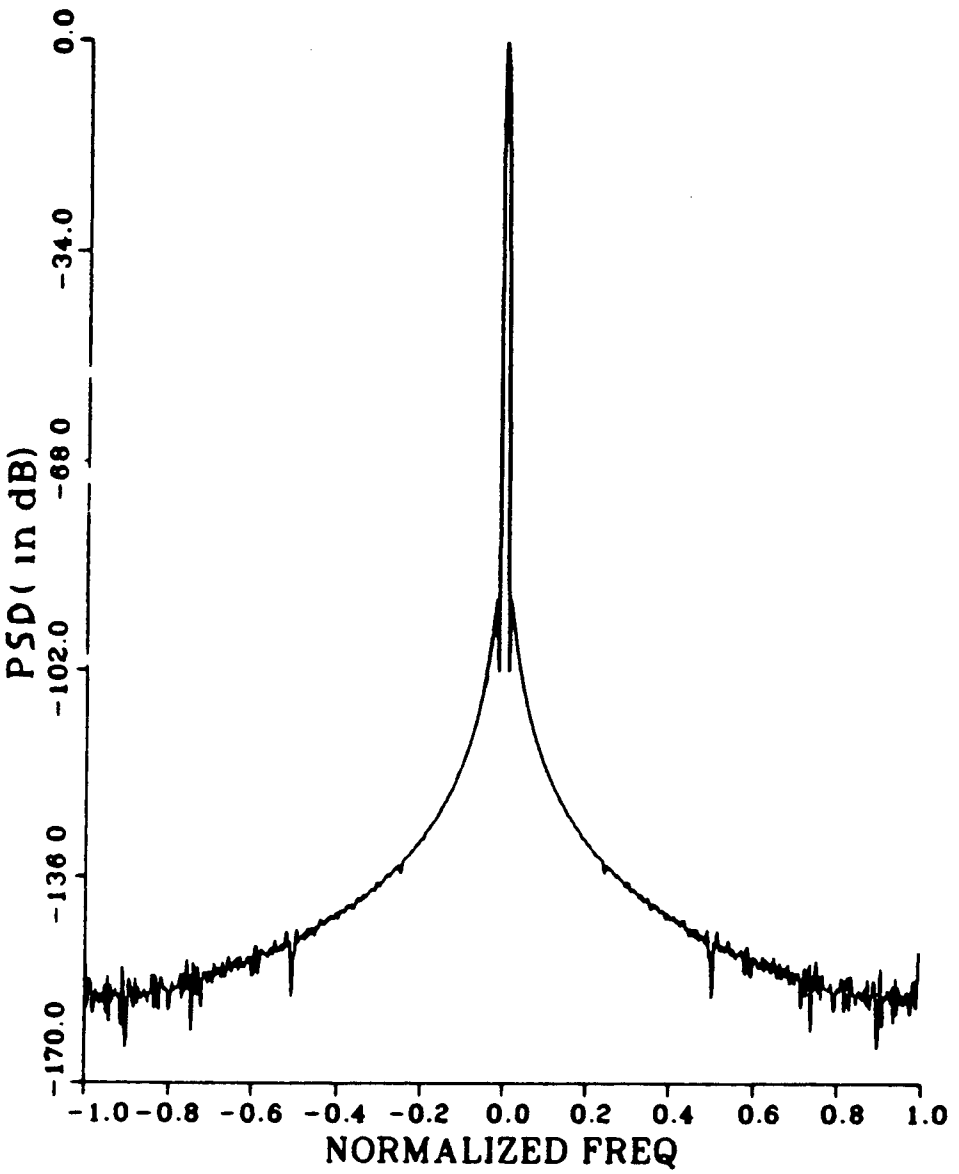


Figure 3.11: Spectrogram of the minimum 4-sample Blackman-Harris window, where PSD denotes power spectral density

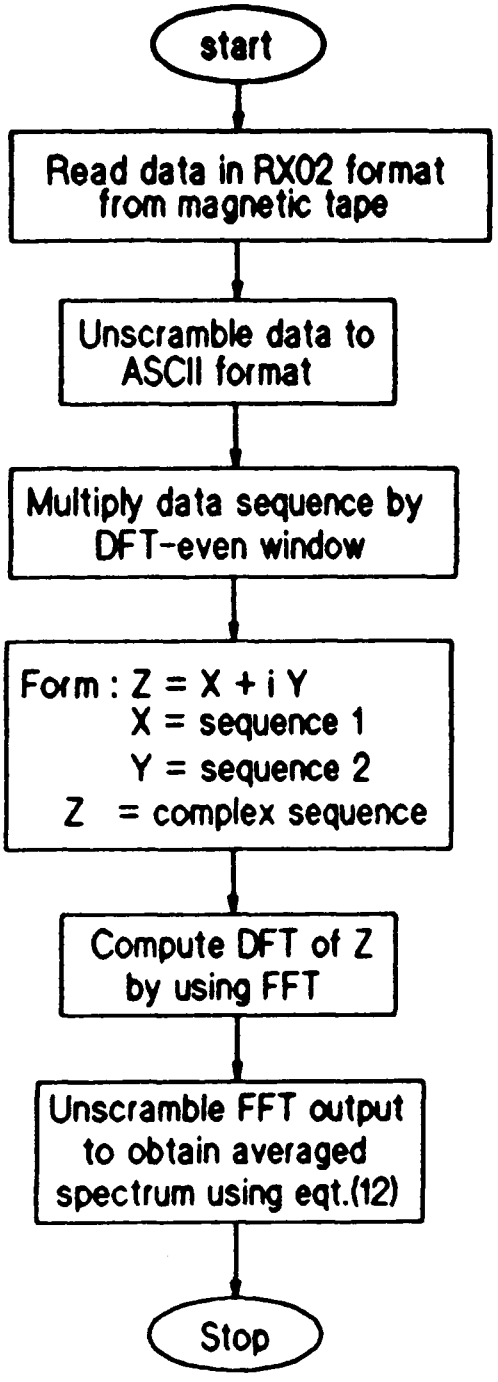


Figure 3.12: Flowchart of the spectral analysis program

push-button telephone, 3.) an electronic ringer of a touch-tone telephone, and 4.) an electronic siren driver (used in timing feature measurement) were used. These sounds were recorded on a tape recorder in various ambient noisy environments in order to investigate the effects of background noises on warning sound spectra.

The recorded warning sounds were fed to an A/D conversion system, and the digitized samples were stored onto a magnetic tape for storage and for further processing. To suppress the aliasing effect of the sampling process, a Kronhite electronic filter was used to remove the spectral components of the analog signals beyond the 10 kHz frequency bandwidth. Then, the filtered signal was fed to a 12-bit MINC/DECC AB-23 A/D converter with selectable data sampling frequency under the master control of a PDP-11 computer. In our work, the sampling frequency was set to 20 kHz. Consequently, each 6.5 seconds of the digitized sound record was transferred from a PDP-11 computer to a VAX-750 general computer for spectral analysis.

3.2.5 Spectra of Warning Sounds Generated by various Warning Devices

Unless otherwise stated, most of the short-time spectra were obtained by averaging four consecutive 25.6 msec segments of the spectrum. We assume that within this 102.4 msec the signals are slowly-varying, and that the average spectrum provides a statistically stable representation of the frequency content of the signals.

3.2.5.1 Spectra of Telephone Rings generated by Electro-mechanical Ringers

Although frequency specification on telephone rings are provided by various standard associations, the acceptable variations of the short-time spectra of telephone rings have not been published. In addition, there is no information on the effect on spectral

variations of the different loudness level adjustments that can be made on electro-mechanical ringers equipped with loudness controls. Similarly, there is no mention in the standards (or in the literature) of the effect of the pitch setting of electronic ringers on the spectra of emitted sounds. The measurements reported here were made to obtain this missing information. Five different aspects were examined:

Short-time averaged spectra of an electro-mechanical ringer

Fig. 3.13 gives a typical example of short-time spectra of telephone rings with the loudness level set to one. (The loudness adjustment control is found at the bottom panel of some rotary dial telephones.) These rings were recorded in an ordinary office environment. Two regions of spectra are identified: the transient, and the steady-state regions. During the beginning 600 msec of the ringing period (transient) these short-time spectra are very similar, and are rich in harmonic content (dominated by three to five major spectral peaks in the 10 kHz frequency bandwidth). Following this is the steady-state of the ringing period with only two or three dominant peaks retained.

Long-time averaged spectra of an electro-mechanical ringer at seven different loudness levels

The next two figures show how telephone ring spectra vary with respect to changes in loudness level adjustments. The same telephone was used as in the previous measurement. These spectra were obtained by averaging 256 25.6 msec long record segments (6.55 sec). Fig. 3.14 (a) shows that two major peaks always occur in the spectra at each of the seven loudness settings. However, for another rotary dial telephone, Fig. 3.14 (b) shows the dramatic changes in spectral characteristics when the loudness adjustment is

altered from level two to three. The disappearance of these dominant peaks is caused by some change in the internal ringing mechanism. These figures clearly illustrate the unpredictability of the effect of varying loudness settings on telephone ring spectral characteristics.

Long-time averaged spectra of five electro-mechanical ringers

Spectra from five rotary dial phones of the same model were used in this measurement. To provide a general view of their spectral variations, Fig. 3.15 gives an example of long-time averaged spectra of five electromechanical ringers with a preset loudness level. In Fig. 3.15, the dominant spectral peaks produced by phone samples 1, 2 and 3, do not appear in the spectra generated by phone samples 4 and 5. This indicates that phone rings generated from telephones of same model do not produce similar spectral characteristics.

Short-time averaged spectra of a multiple-line telephone

Fig. 3.16 depicts another set of short-time spectra for a multiple-line push-button telephone. Since this telephone is not equipped with a loudness adjustment control, our study on the effect of varying loudness adjustment on short-time spectra was not performed. Compared to other telephone ring spectra, Fig. 3.16 consists of spectral peaks at different frequency locations: 1.6 kHz, 3.2 kHz, 5.9 kHz and 9.2 kHz.

Short-time spectra of an electro-mechanical ringer in steady noise background

To demonstrate how steady fan noise affected the short-time spectra of telephone rings, we used the same phone as in the first two measurements. These telephone rings were recorded inside a computer room where an air-ventilation system was operating. Compared to Fig. 3.13, in Fig. 3.17 the amplitudes of dominant peaks decreased, the number of dominant peaks was reduced, and the transient regions of the spectra have largely disappeared. This may be caused by the effect of spectral flattening of the background noise. However, two of the dominant peaks of successive spectra are still retained.

Conclusions

Spectra of telephone rings produced by electro-mechanical ringers consist of a) two distinct regions (transient and steady) of short-time spectra, and b) spectral peaks are always located in the 1.6 – 2.5 kHz and 4.7 – 6.2 kHz bands. Details of the spectral characteristics vary with loudness, with the model, and with individual units of the same model.

In general, it is difficult to predict the spectral distortion caused by background noise because such distortion is highly dependent on the characteristics of the noise. Such characteristics are both time and spatial variant. Since real environmental noise situations are very variable, there is very little practical value in further study on the effect of noise on the spectra.

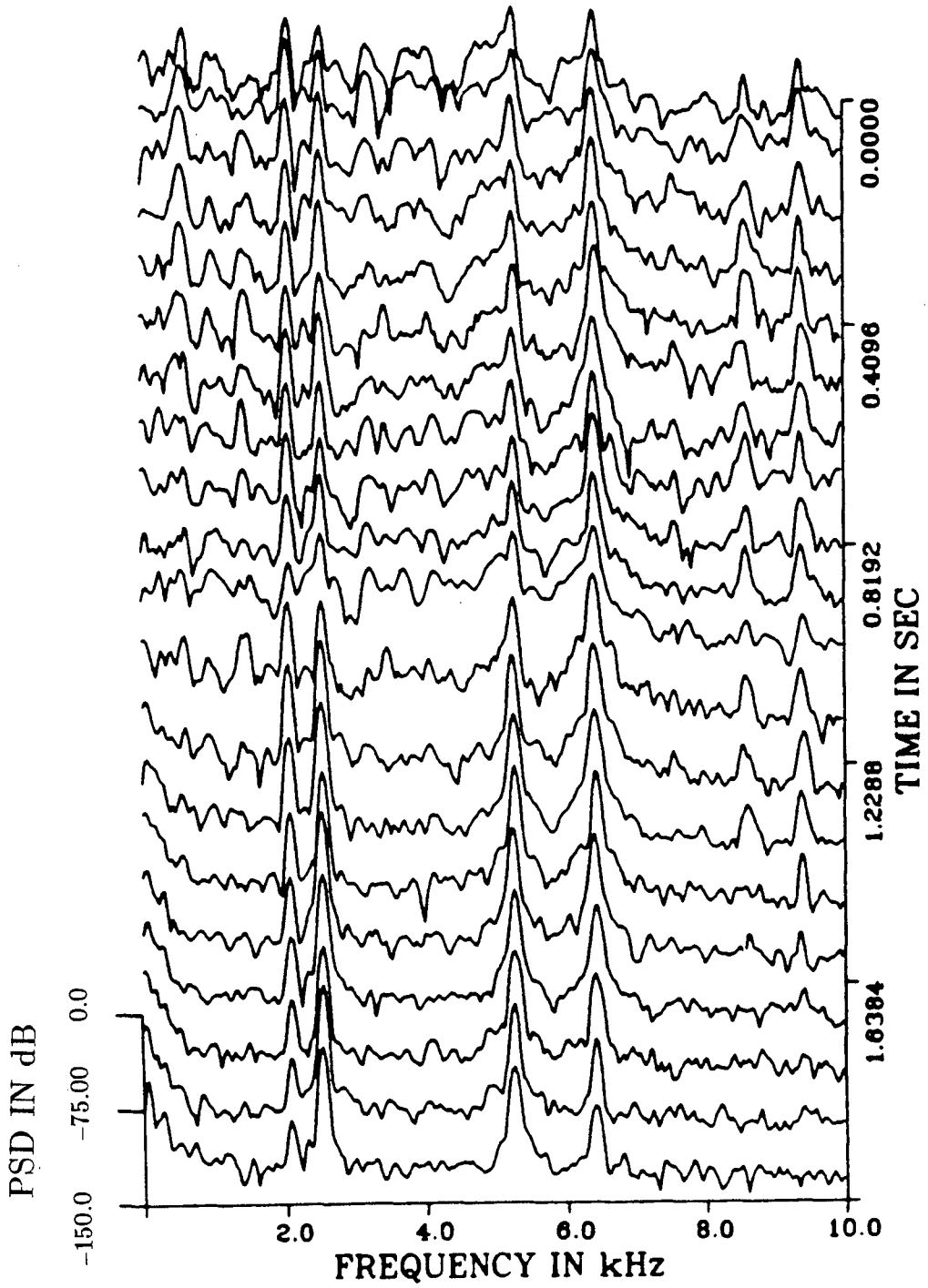


Figure 3.13: Short-time spectra of an electromechanical ringer

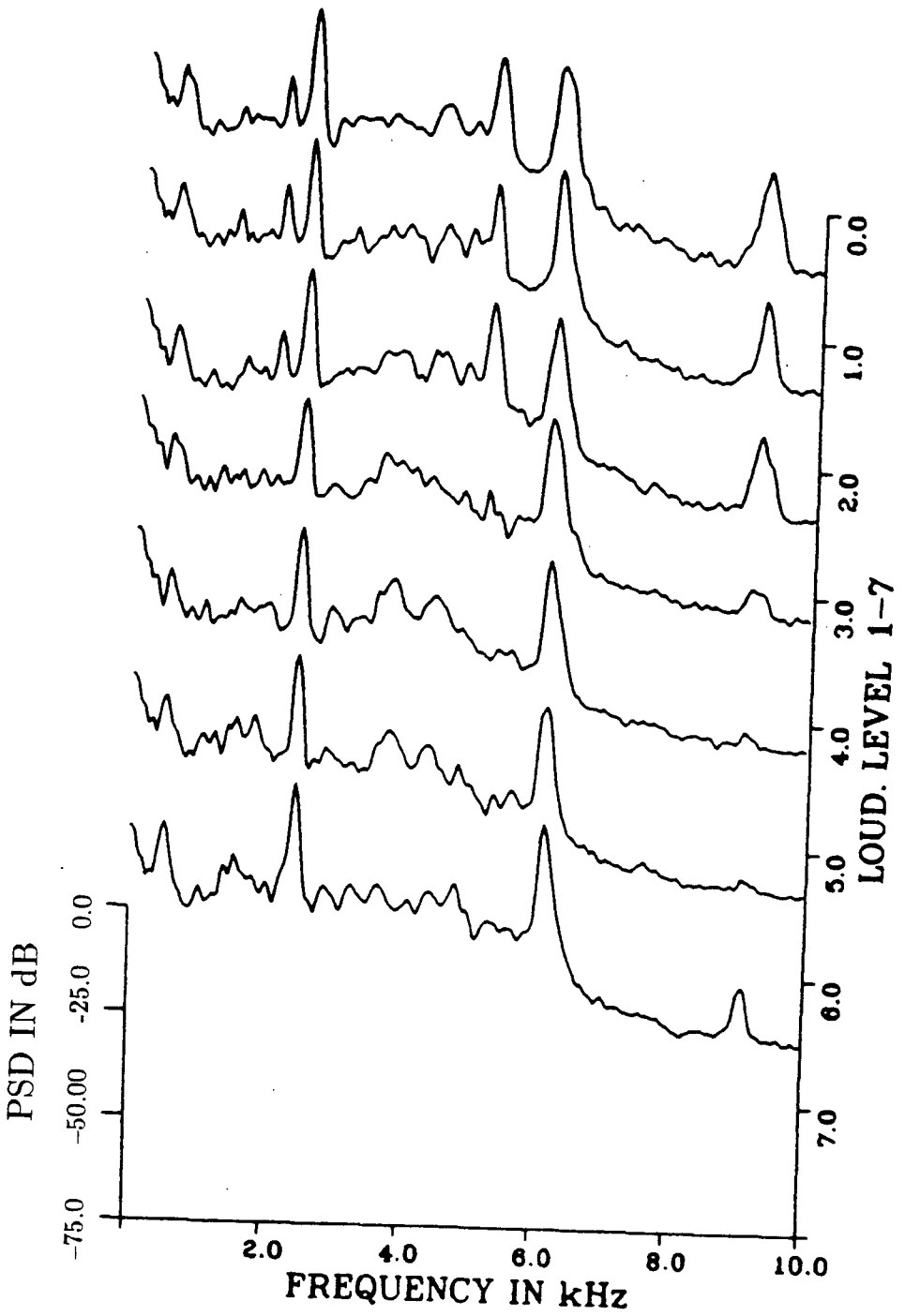


Figure 3.14 (a) : Spectra of an electromechanical ringer with seven loudness settings

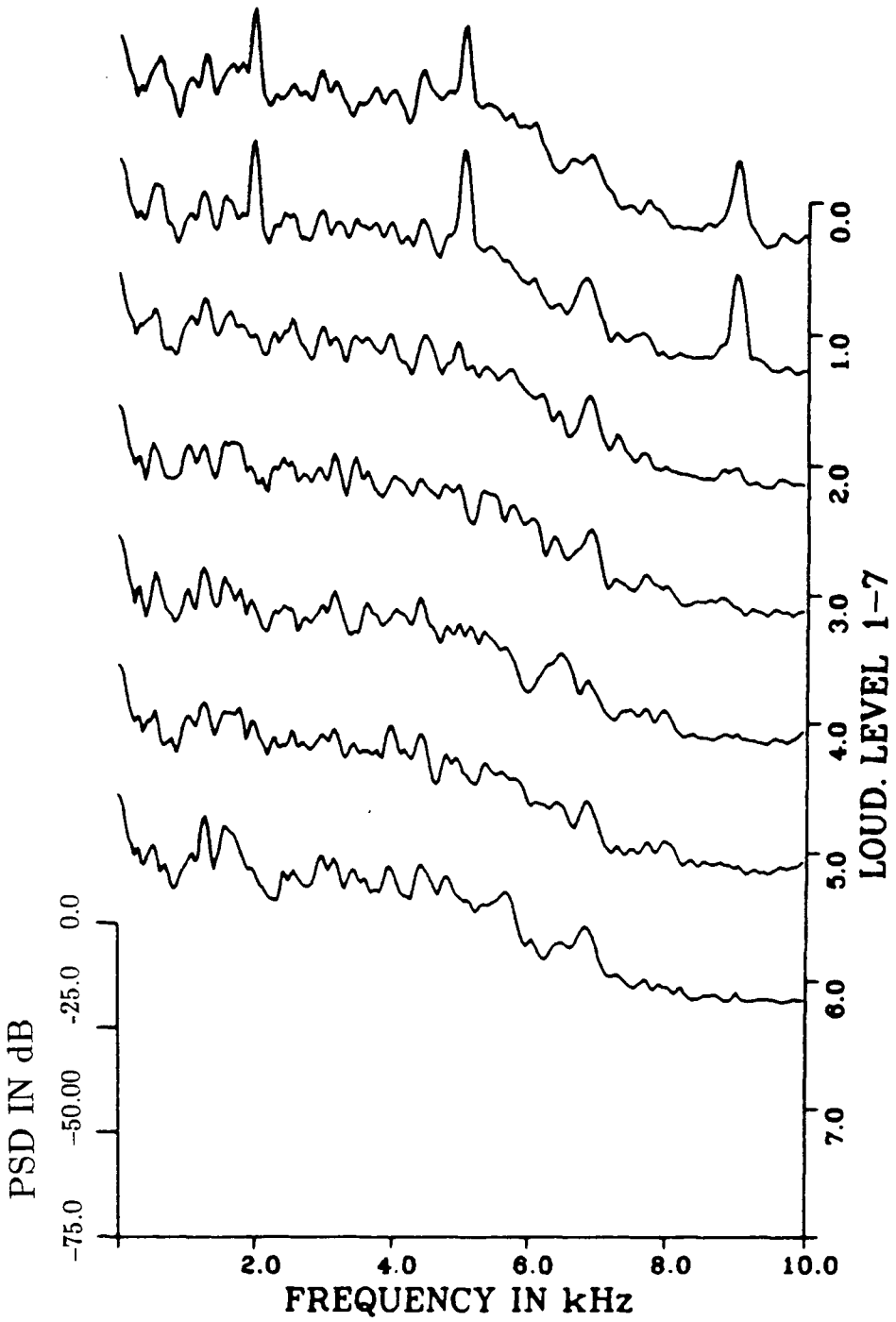


Figure 3.14 (b) : Spectra of another electromechanical ringer with seven loudness settings

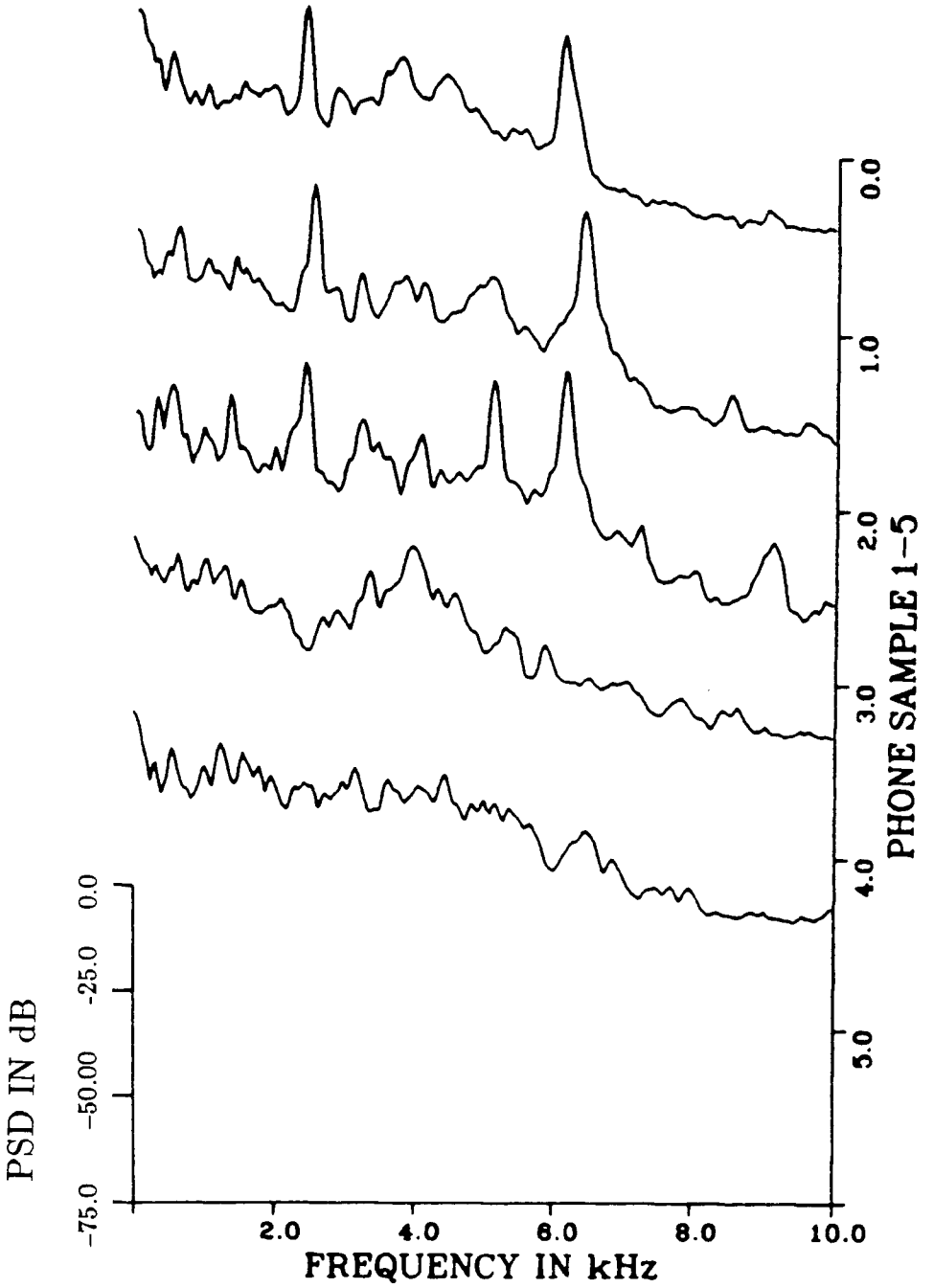


Figure 3.15: Long-time averaged spectra of five electromechanical ringers

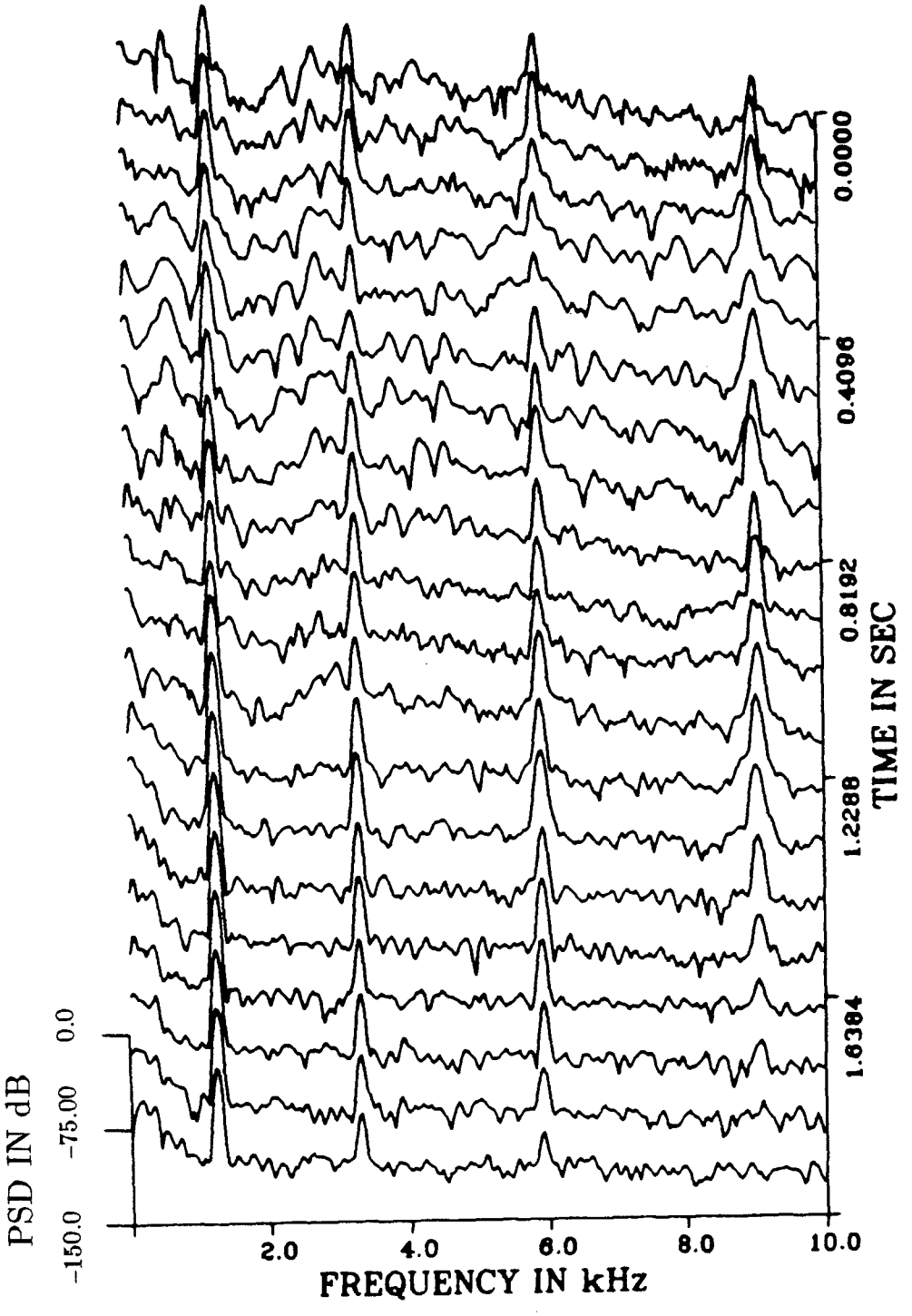


Figure 3.16: Short-time averaged spectra of a multiple-line telephone

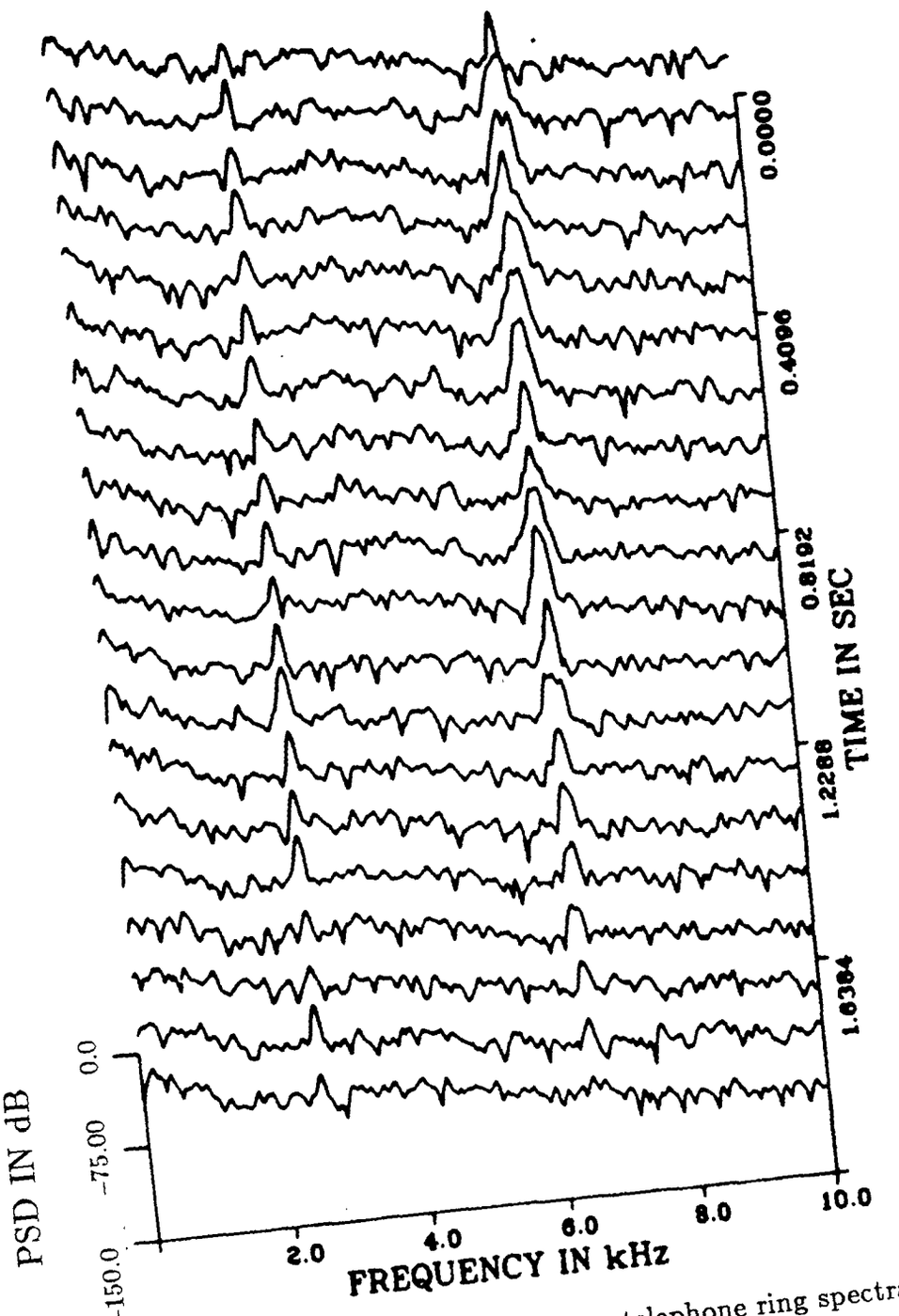


Figure 3.17: Effects of steady fan noise on telephone ring spectra

3.2.5.2 Spectra of telephone rings generated by an electronic ringer

Since solid state transducers are manufactured to close tolerance, and the control circuits generate very consistent tone frequencies, electronic ringers of the same type will produce sounds with very similar features. In addition, the different types all conform to applicable standards. Therefore, only one electronic ringer unit was examined in detail. Since the telephone we examined was equipped with pitch adjustment controls, the effects of different pitch settings on the spectra were also studied.

Each of the short-time spectra was obtained by averaging two consecutive 102.4 msec long spectra. The reason for selecting 102.4 msec segments was to provide a frequency resolution of 19.6 Hz for the separation of the two dominant tones generated by the electronic tone ringer. Fig. 3.18 (a), (b), (c), and (d) show that the change of pitch setting results in more high energy peaks appearing. Although it is difficult to see in these plots, the pitch setting also results in the shifting of the dominant lowest frequency peaks. The tabulated numbers indicate that for this particular electronic ringer, one tone frequency varies from 468 Hz to 546 Hz, and the other varies from 546 Hz to 683 Hz.

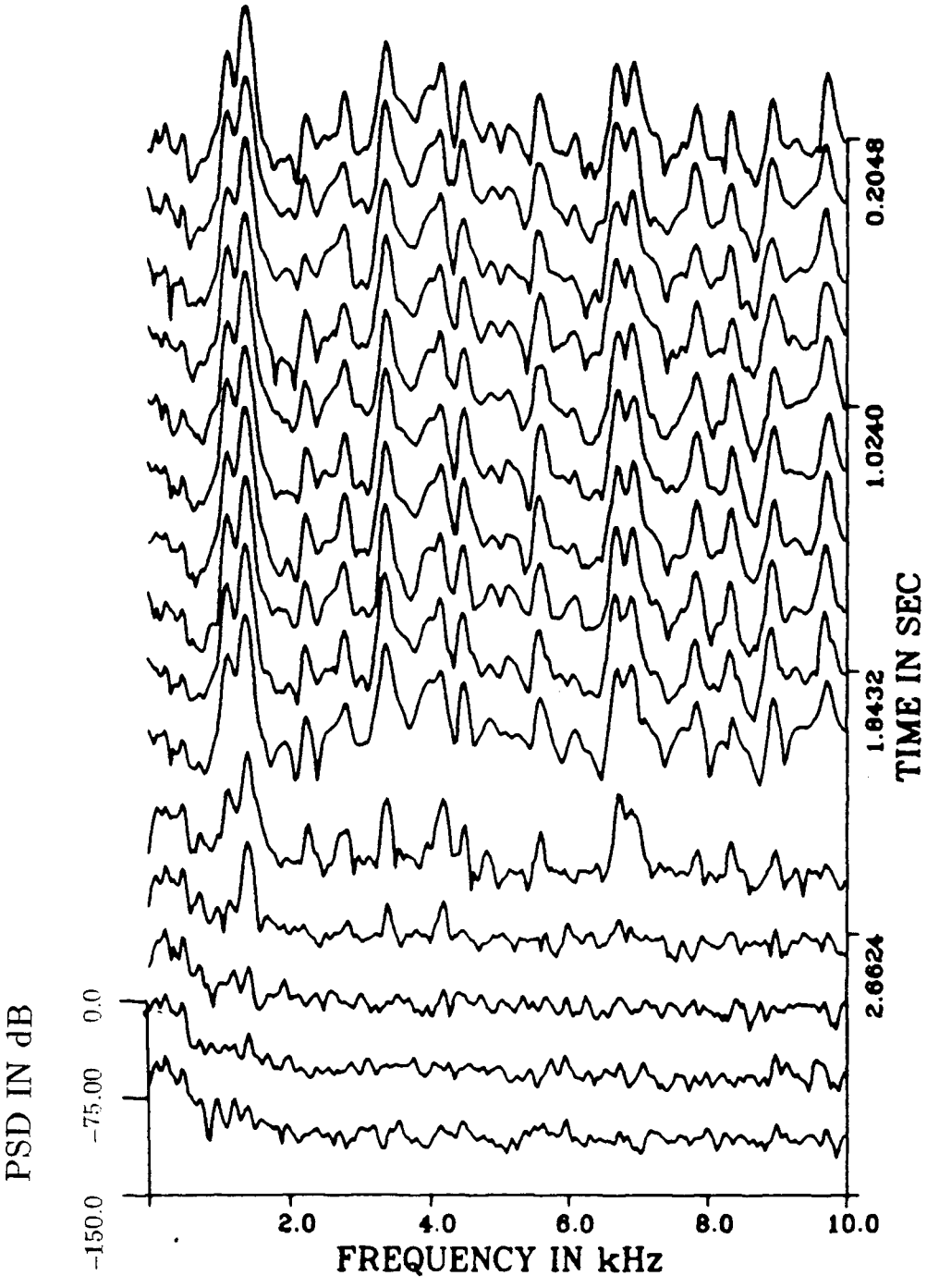


Figure 3.18 (a): Short-time spectra of electronic rings with pitch set at position one

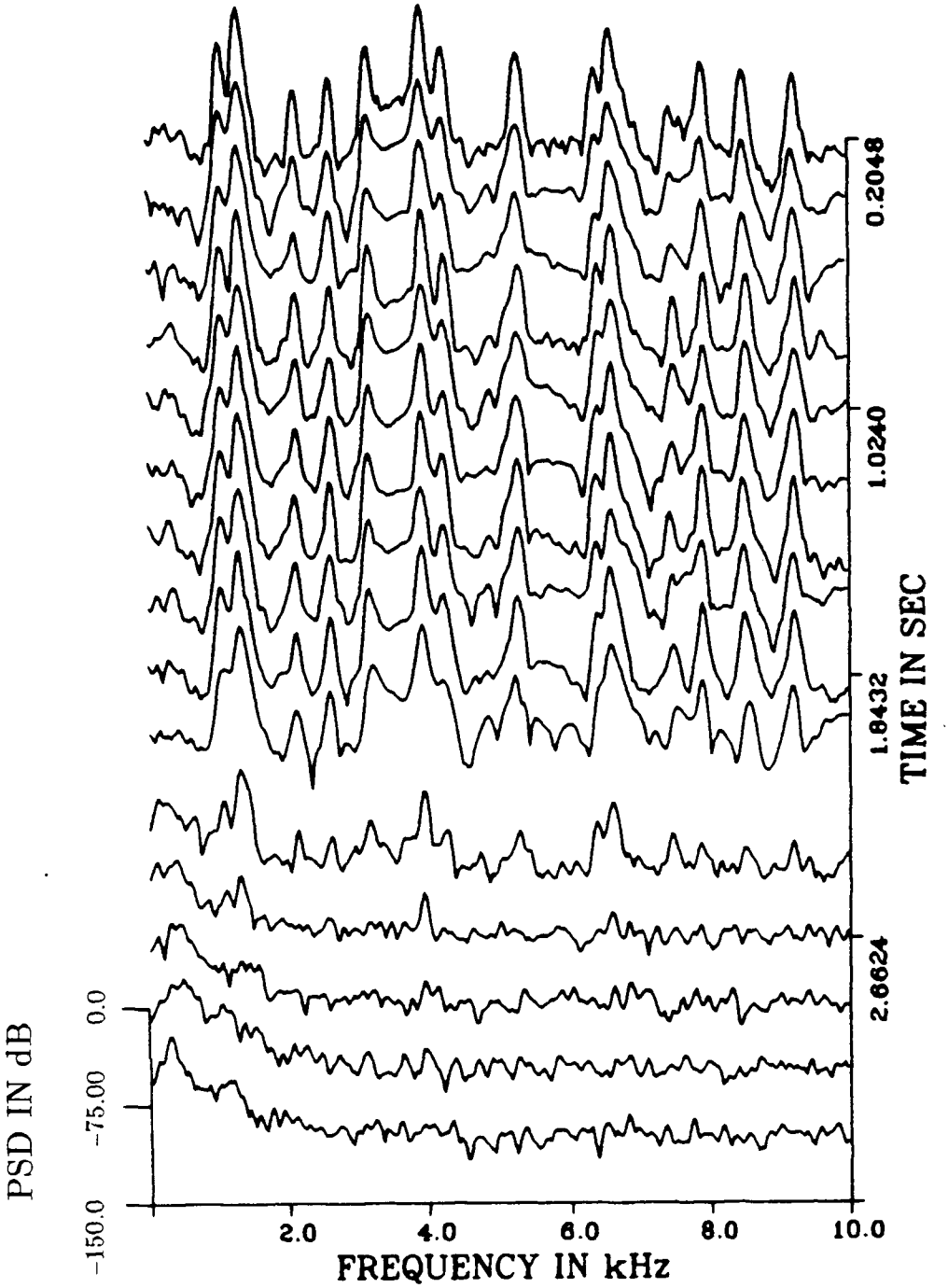


Figure 3.18 (b): Short-time spectra of electronic rings with pitch set at position two

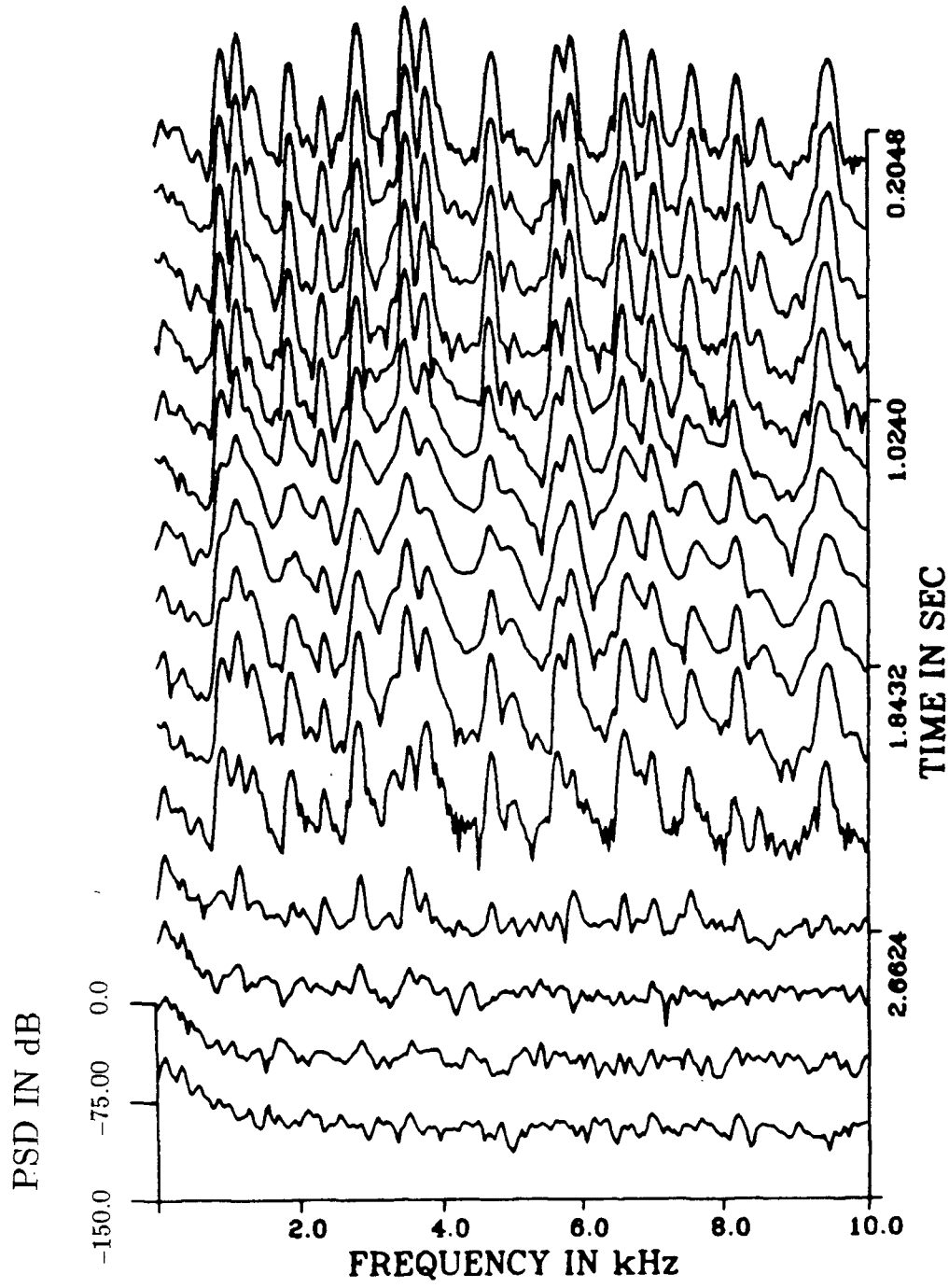


Figure 3.18 (c) : Short-time spectra of electronic rings with pitch set at position three

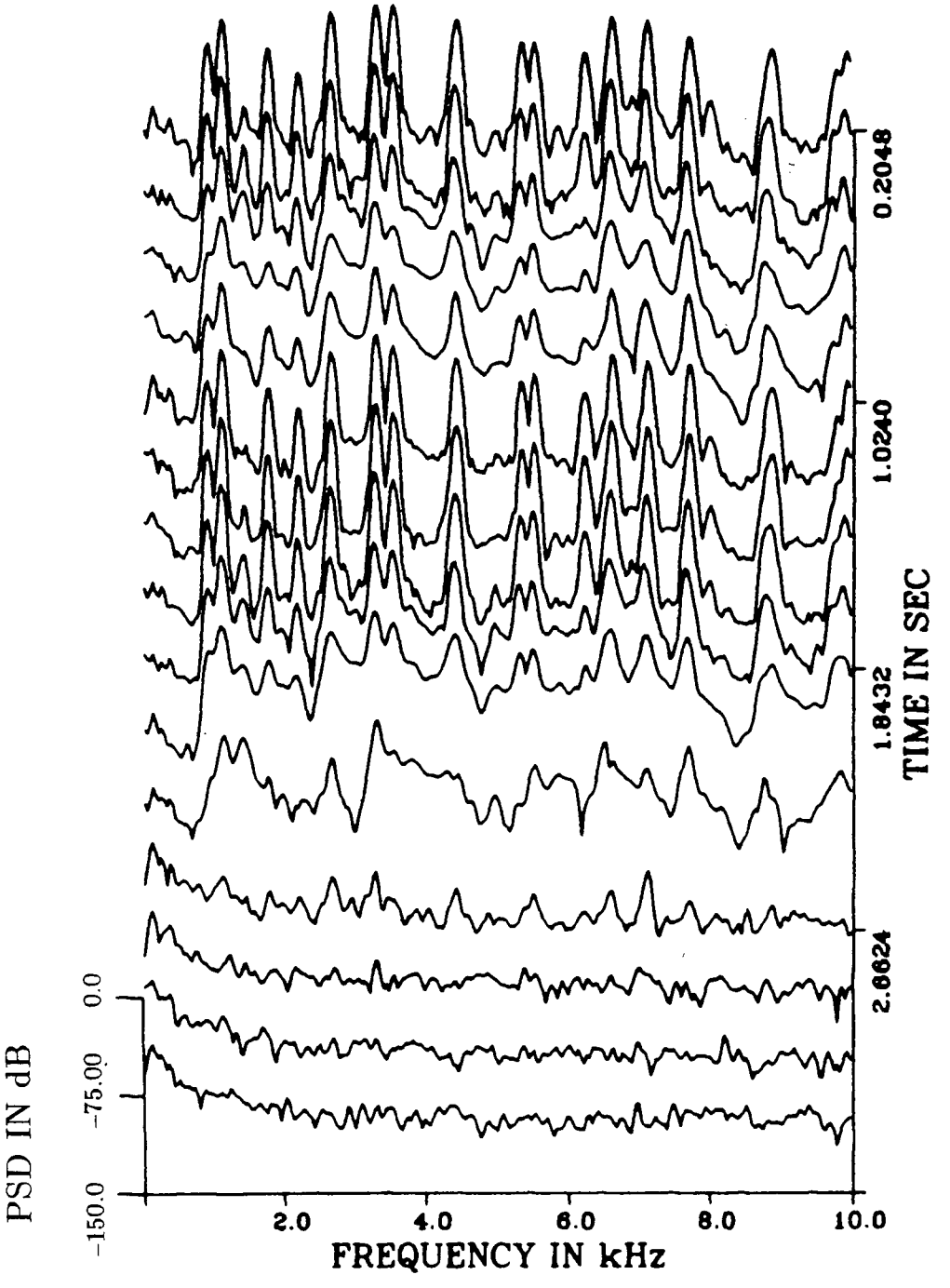


Figure 3.18 (d) : Short-time spectra of electronic rings with pitch set at position four

3.2.5.3 Spectra of Siren Sounds

Lastly, we studied the spectral characteristics of different warning siren sounds produced by an electronic siren driver. Eight different siren sounds can be produced with this device. In all of these spectra, note that the peaks located at 7.0 kHz are produced by ambient noise monitored independently with the sound pressure level meter.

Rapid-Yelp

The short-time spectra of this sound consist of a band of frequencies varying from 1400 Hz – 3000 Hz (Fig. 3.19).

Conventional Yelp

Fig. 3.20 shows the variation of short-time spectra of this sound which consists of a varying band of frequencies ranged from 666 Hz – 1333 Hz.

Low-high Sweep

Fig. 3.21 shows a very interesting 'chirp-signal' type of short-time spectra. The spectra consist of peaks varying from 820 Hz to 4.0 kHz.

European Hi-low

Fig. 3.22 shows spectra which consist of fundamental spectral component at 1093 Hz, along with its harmonics at 1640 Hz and 3164 Hz.

Hi-frequency Steady

Fig. 3.23 gives the spectra, which consist of a fundamental spectral component at 833 Hz, together with its harmonics at 1640 Hz and 3200 Hz.

Pulsating Siren

Fig. 3.24 gives the spectra of a 'Pulsating Horn' siren sound, which consists of a poorly defined peak at 1600 Hz and a distinct peak at 2400 Hz.

Steady Horn

Fig. 3.25 shows spectra, which consist of two major bands of frequencies at 500 – 700 Hz and 1200 – 1400 Hz.

Electronic Synthesized Bell

Fig. 3.26 shows the spectra of a bell sound, which consists of four peaks at 700 Hz, 1406 Hz, 2070 Hz, and 2812 Hz.

3.2.6 Summary

Summing up the spectral analysis results, we reached the following conclusions [34]:

- dominant spectral features of warning signals generally appear within the frequency range between 300 Hz to 5.0 kHz,
- warning signal spectra may consist of a single spectral peak, or regular clusters of spectral peaks and valleys and,
- in general, the spectral features of warning signals are simpler than those of speech signals with regards to:

1. absence of nonstationary segment of short-time spectrum (while isolated speech utterance may consist of nonstationary short-time spectra caused by weak fricatives at the utterance boundaries) and,
2. repeatability of spectral features of warning sounds (while due to variable utterance rate of a word, nonlinear time distortion in spectral features occurs).

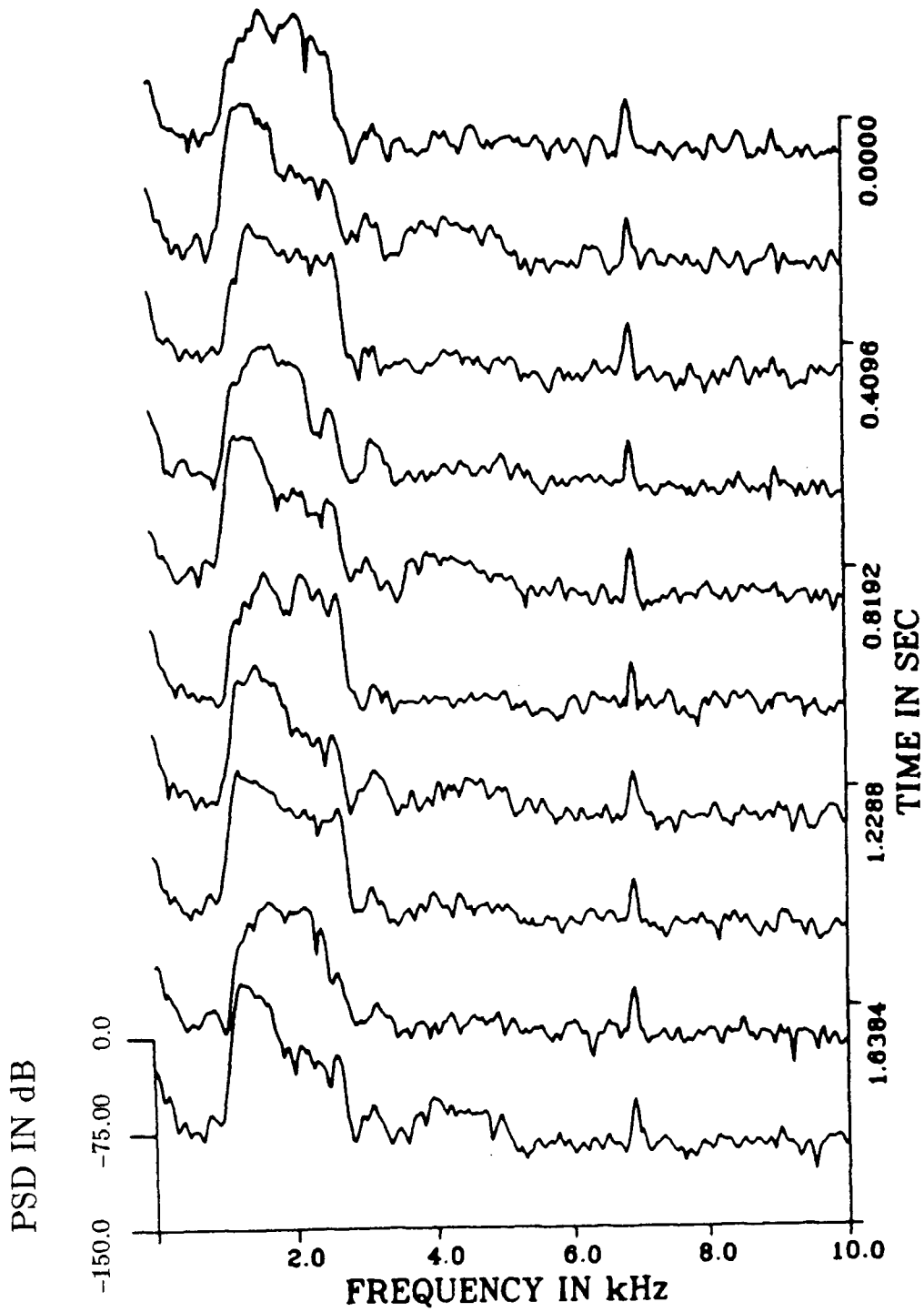


Figure 3.19: Spectra of Rapid Yelp sound

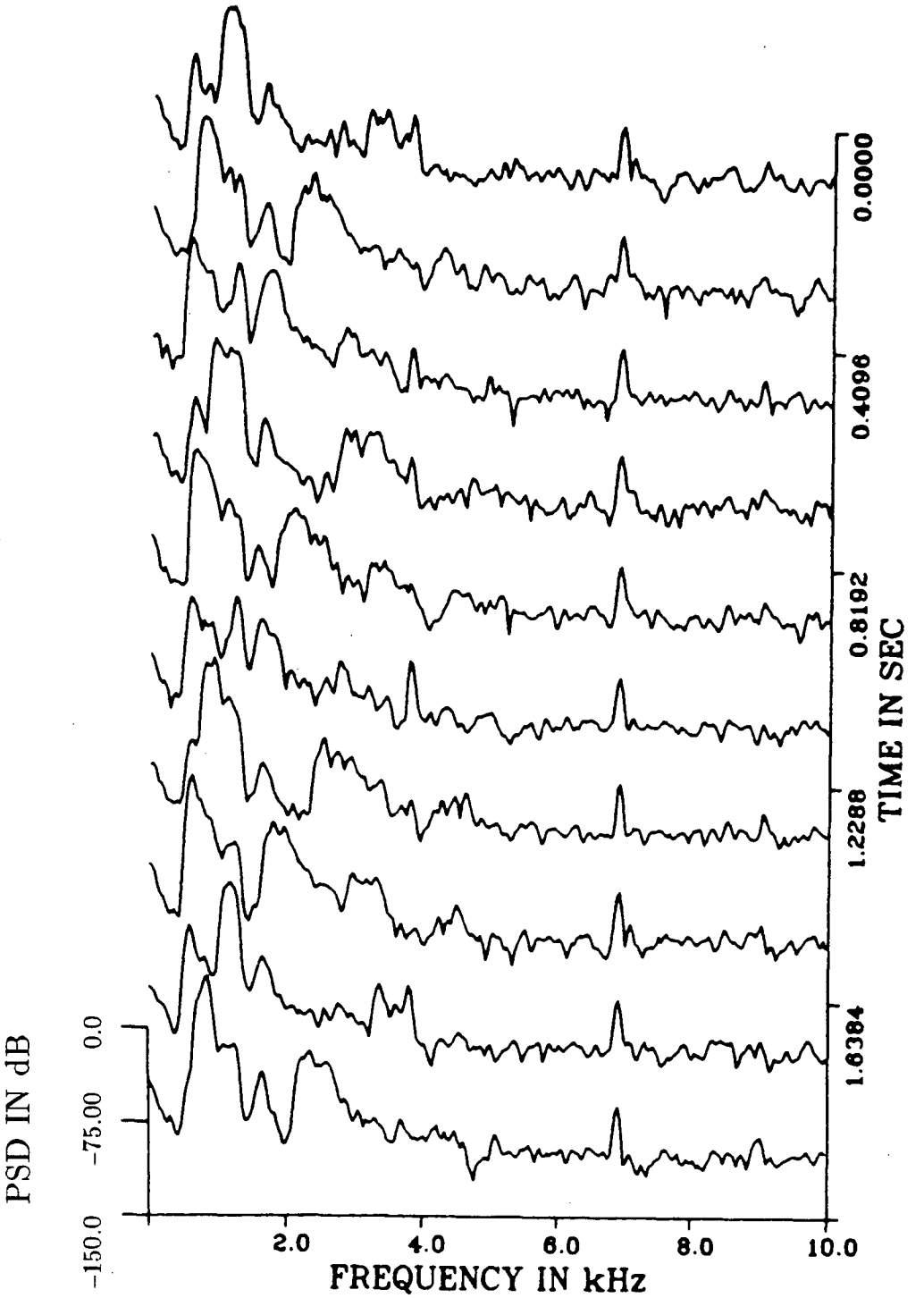


Figure 3.20: Spectra of Conventional Yelp sound

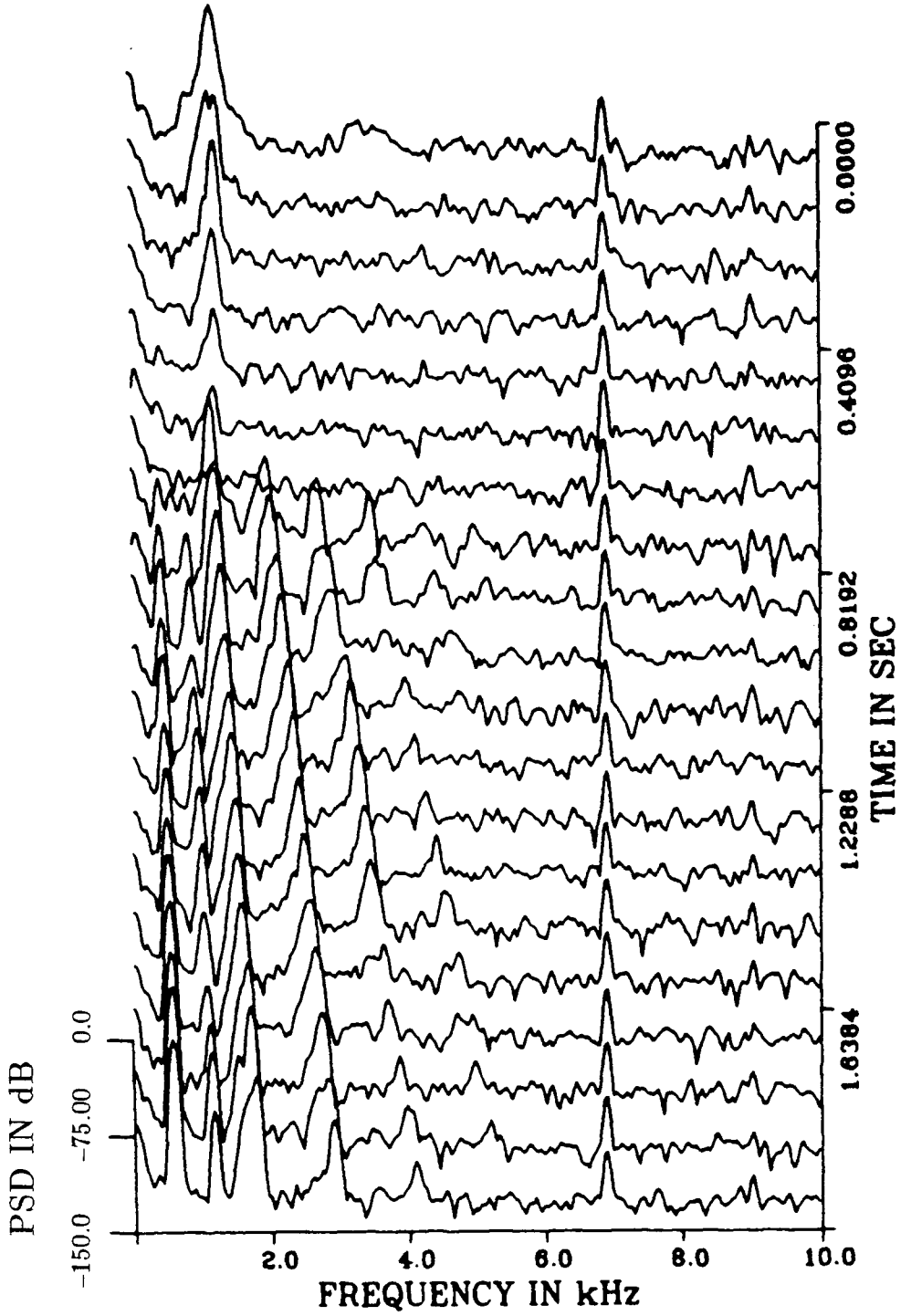


Figure 3.21: Spectra of Low-Hi sweep sound

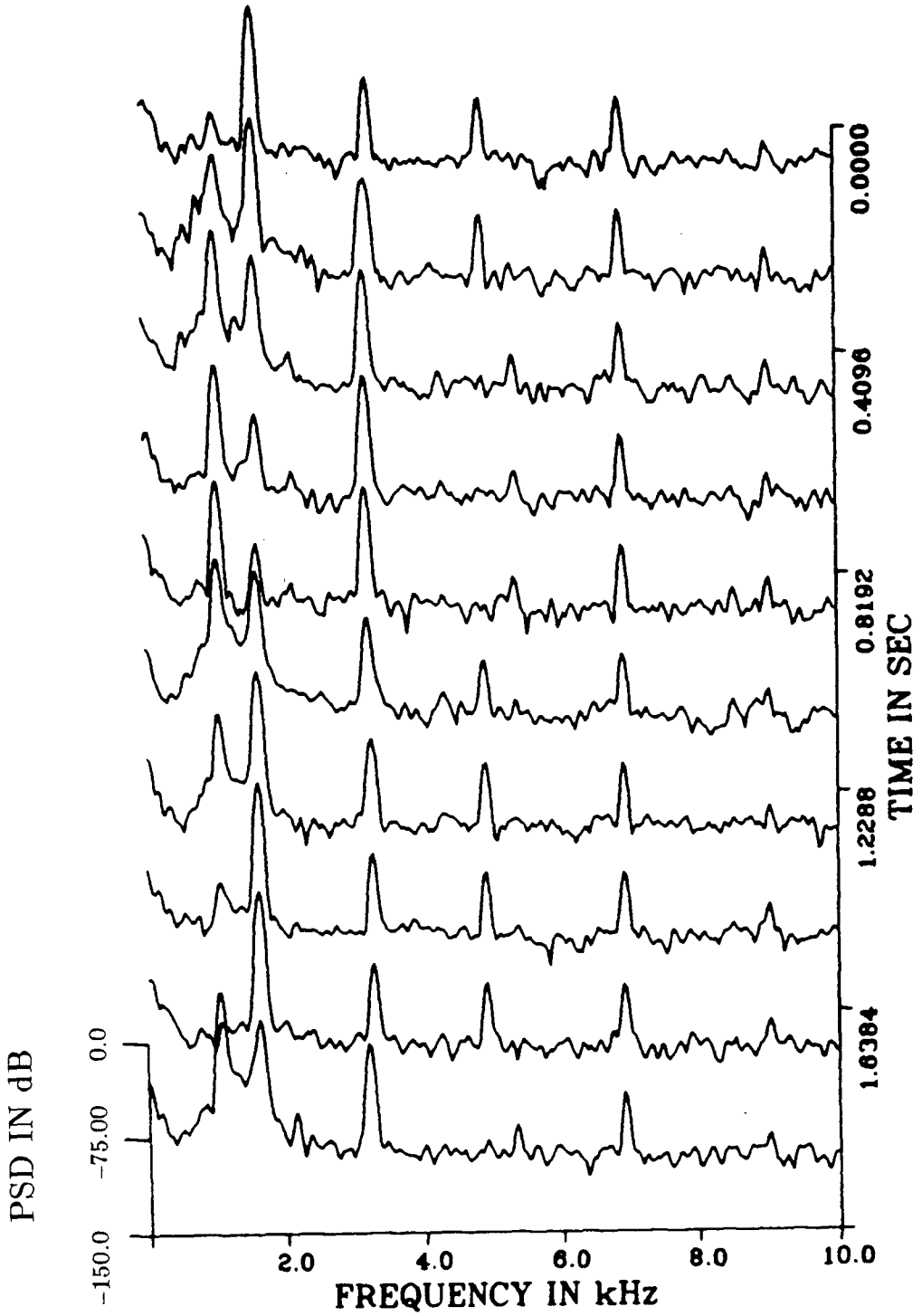


Figure 3.22: Spectra of European Hi-Low sound

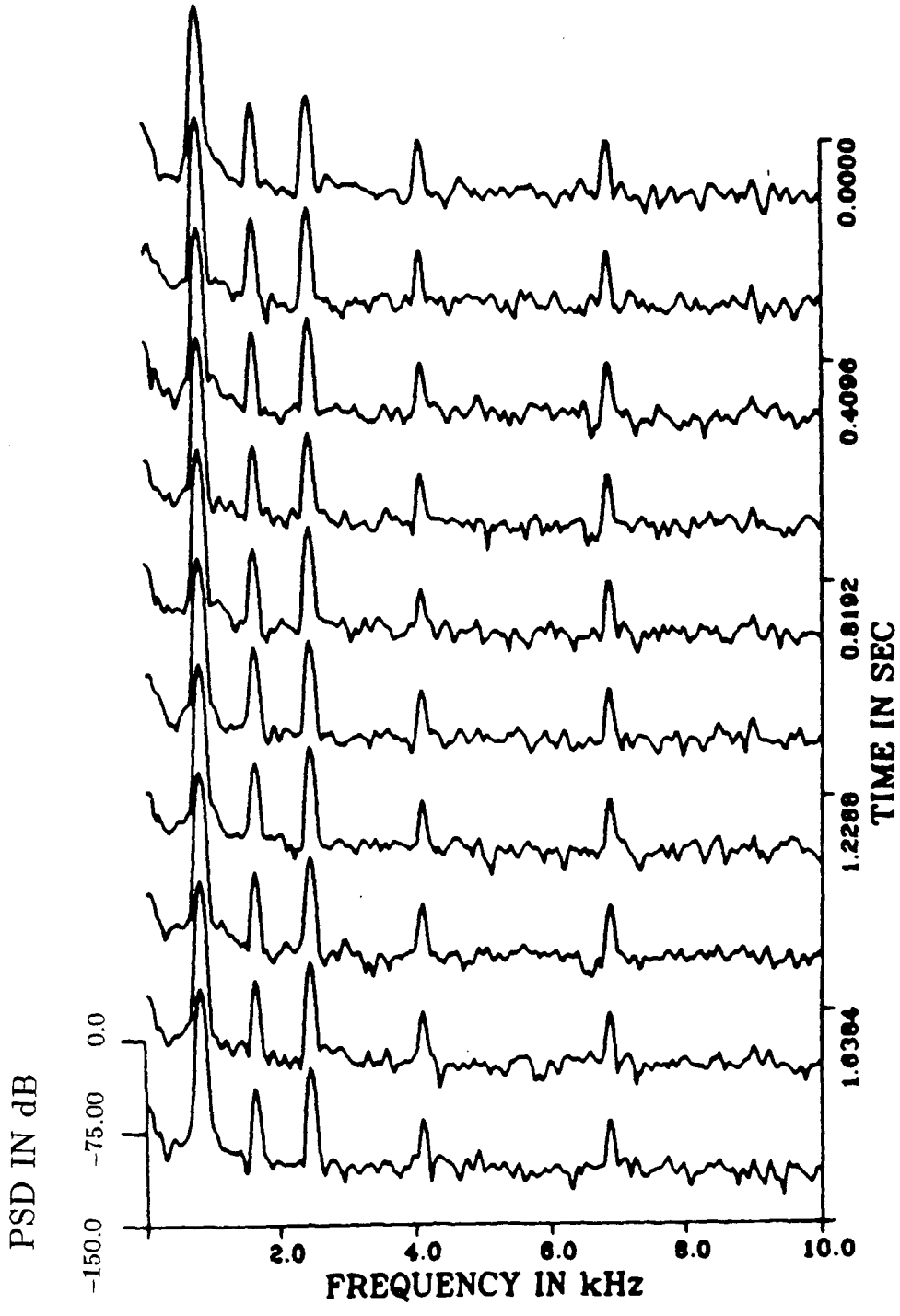


Figure 3.23: Spectra of Hi-Frequency Steady sound

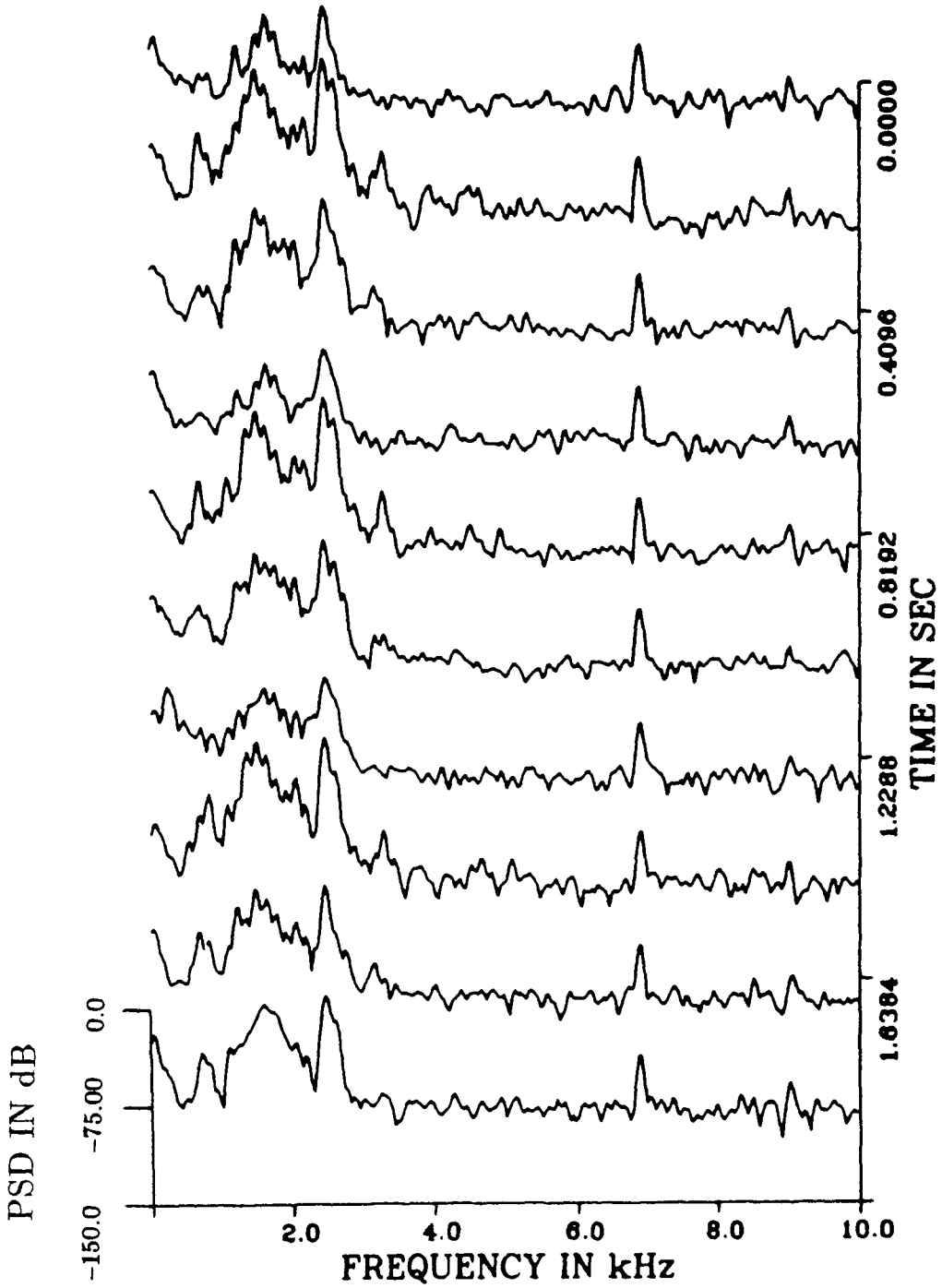


Figure 3.24: Spectra of Pulsating Horn sound

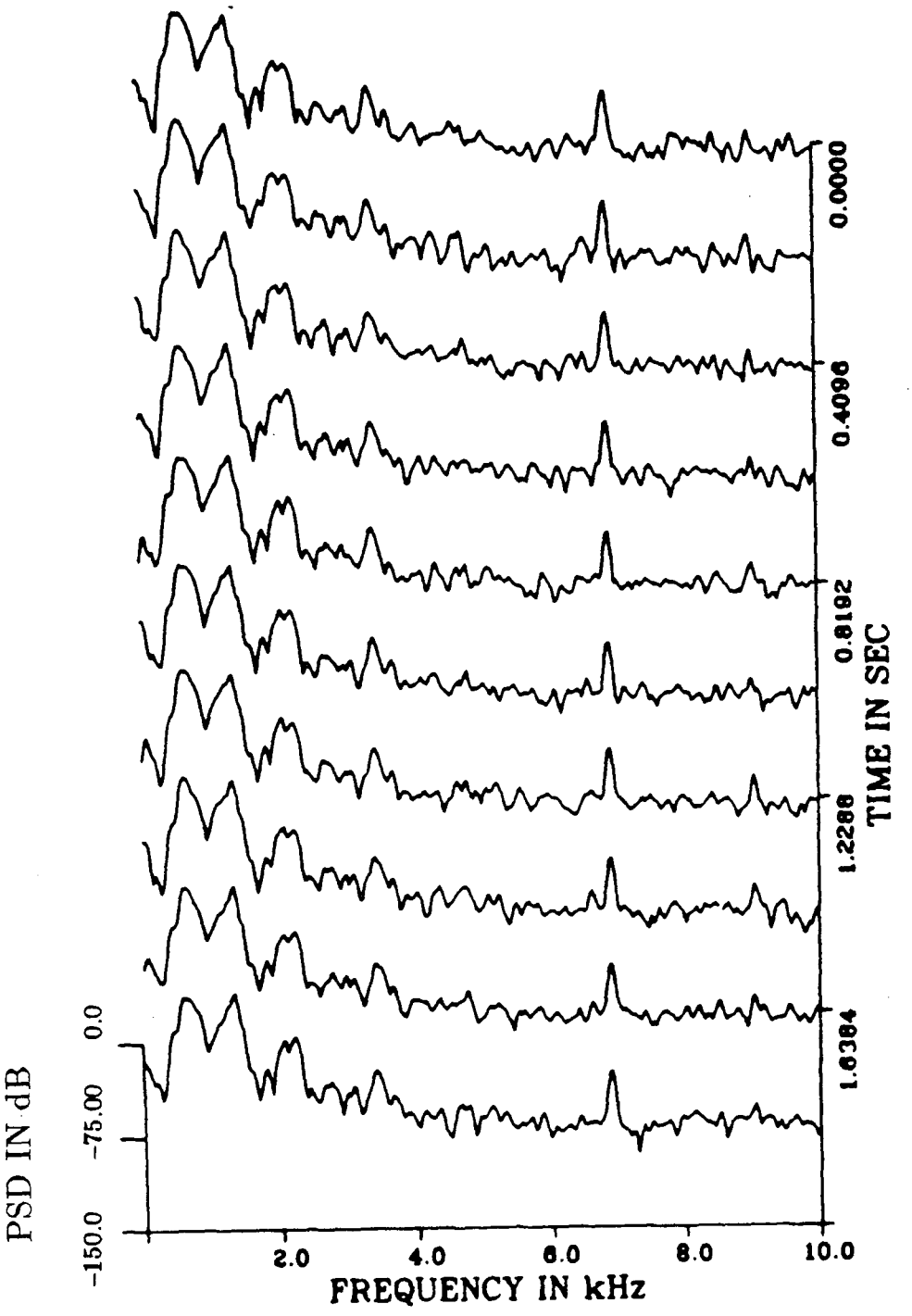


Figure 3.25: Spectra of Steady Horn sound

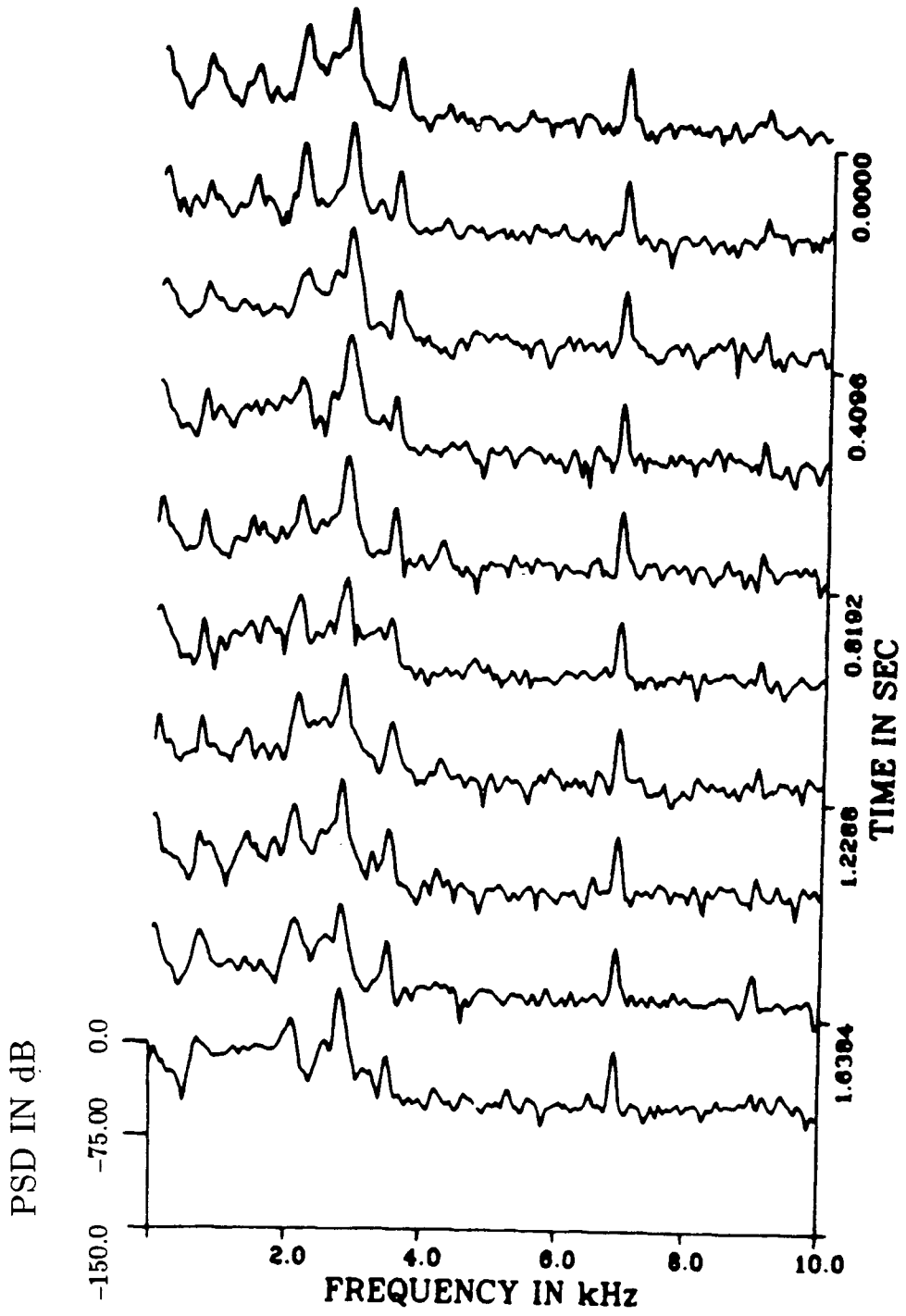


Figure 3.26: Spectra of Electronic Synthesized Bell sound

Chapter 4

Solutions to the Recognition Problem

4.1 Pattern-Recognition Model for Signal Identification

The classic pattern-recognition scheme for signal identification is shown in Fig. 4.27. This scheme consists of feature extraction, pattern matching (similarity tests), and decision making blocks. It forms the basis of many applications, because it places no restrictions on the use of different feature sets, similarity algorithms, and decision rules, and it is possible to implement it in a wide range of circumstances [37].

The function of the feature extraction stage is to convert the signal into parameters or feature sets. This results in the reduction, and sometimes elimination, of redundancies that exist in the original signal. Such signal reduction procedures provide a manageable number of signal features, making practical machine recognition feasible. Extractable signal features include timing information, short-time spectra, Linear Prediction Coding (LPC) parameters, LPC-derived cepstral coefficients, or statistical parameters derived from the Hidden Markov Model (HMM).

For pattern comparison, the signal features must be either known a priori, or the system must "learn" them. Such learning may be accomplished by training the system with the signal(s). This involves the extraction of features, and their consequent storage in template memory. The signal feature sets are obtained from consecutive short-time segments of the signals. To recognize a specific signal, the features of the unknown signal are compared with the different sets of pre-stored reference signal features.

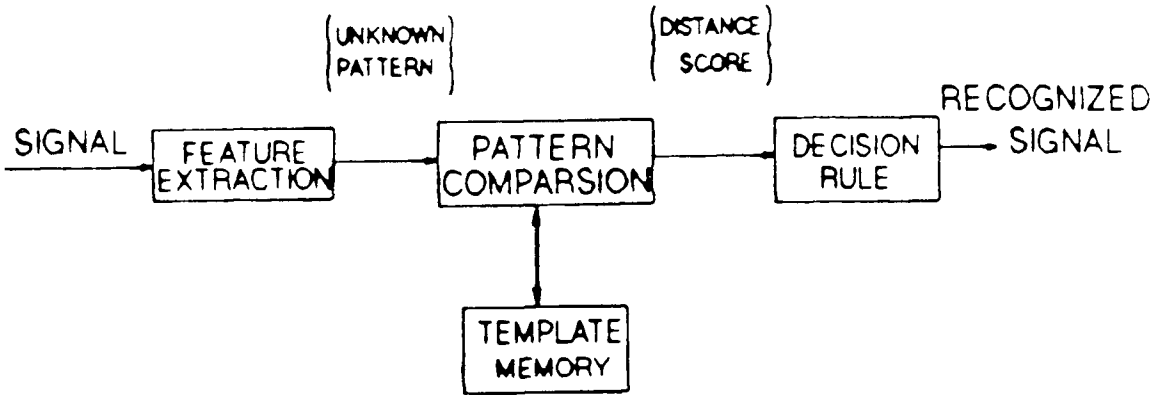


Figure 4.27: Classic Signal Recognition Scheme [37,38]

The matching of the unknown signal features to the templates is generally complicated by the non-linear time mis-alignment of the short-time feature segments of the unknown signal and of the reference templates. To solve this matching problem, the well-known dynamic time warping (DTW) algorithm is employed [39]. Based on this algorithm, for each reference template an optimum match between the unknown signal and the reference features is sought. In these pattern comparisons, distance calculations are performed on the short-time segments of signal feature sets in order to provide a measure of similarity between the unknown signal and the reference templates. The literature offers several distance measures [40,41,42].

One of two decision rules are used in most practical systems: the nearest neighbor rule (NN rule), and the K-nearest neighbor rule (KNN rule). The NN rule is applied when there is a unique reference template for each possible signal. In comparing an unknown signal with the reference templates, the pre-stored template which is the

smallest distance from the signal, is recognized to be the unknown signal. The KNN rule is applied when multiple reference templates are learned from each possible signal, giving several template sets representing different signals. The unknown signal is associated with the template set for which the minimum average distance is computed.

4.2 Review & Evaluation of Signal Recognition Techniques

In the following sections, an overview of previous research in signal recognition is presented. Emphasis is focused on speech signal (isolated utterance) recognition techniques, because 1) these recognition schemes fit well to the pattern-recognition model, 2) recognition performance of each applicable technique has been reported [35,36], and 3) warning sounds have acoustic features (i.e., pitch and formant) similar to speech signals. Based on this survey, and on our signal analysis results, the most suitable recognition method will be selected for the WARNSIS.

4.2.1 Analyzing & Utilizing Timing Features

Timing information may be extracted from signals using autocorrelation coefficients, zero-crossing measurements, energy waveform analysis, and peak detection. Such information has been used in signal recognition in a variety of ways.

4.2.1.1 Auto-correlation coefficients

Purton [43] used speech signal autocorrelation coefficients in his speaker-dependent recognition experiments. Specifically, these autocorrelation coefficients were derived from the outputs of two bandpass filters used to capture the formants of speech signals. He achieved an average of 90 % recognition accuracy for a vocabulary size of 10 words.

Sondhi [44] applied autocorrelation analysis to the speech signals which were preprocessed by a center-clipping technique which removed the formant structure. The

signal pitch was then extracted from the autocorrelation function. The combined formant structure removal of signals and autocorrelation analysis provided a robust pitch estimation method. The effects of different degrees of formant structure removal prior to autocorrelation analysis on pitch estimation of speech signals was given by Rabiner [45]. A real-time hardware implementation of a pitch estimation scheme based on a combination of center-clipping and peak-clipping methods, followed by autocorrelation analysis, was reported by Dubnowski [46].

To use the correlator-bank approach similar to that of Purton [43] in our work, the number of correlators used and the number of terms (autocorrelation function coefficients) retained to formulate a signal feature need to be determined. This may be achieved by spectral analysis, and for our signals a multiple-band correlator would be needed. In addition, if the autocorrelation function coefficients generated with zero-delay is used, this is equivalent to utilizing the short-time signal spectral information. Hence, such correlator-based recognizer produces a large feature set, making difficult and uneconomical to design and implement a real-time recognizer based on this concept.

4.2.1.2 Zero-crossing

Rabiner and Sambur [47] analyzed energy and zero-crossing measurements of pre-recorded speech signals in determining the endpoint locations of isolated utterances. First, the energy contour of an utterance was generated and studied to provide a crude boundary. To refine this utterance boundary, zero-crossing measurements were used. At an SNR of 30 dB or better, this endpoint detection algorithm worked very well over all tested conditions.

To develop a low-cost, microprocessor-based speaker dependent recognizer, Whitaker and Angus [48] employed zero-crossing measurements to track two formant variations of speech sounds. The zero-crossing counts were obtained from the outputs of two filters

(one of which was a low-pass filter with a cut-off frequency at 800 Hz, and another was a high-pass filter with 3 dB corner frequency at 1000 Hz). In order to optimize storage, they used the variable rate encoding technique to reduce redundancies in signal features. With a vocabulary size of 10 - 20 words, they attained an averaged recognition accuracy of 95 % - 99 %, depending upon the formant structure of the utterances.

The use of zero-crossing detectors for warning signal recognition is attractive. However, the accuracy of zero-crossing measurements depends on the relative amplitude of the dominant frequency compared to other frequency components within each frequency band, and also on the spectral spacing of the components [48]. In addition, zero-crossing analysis is very prone to noise interference. Although zero-crossing detectors can be implemented easily and economically, the inconsistency of their operational performance in noisy environments makes this approach unsuitable for WARNSIS.

4.2.1.3 Energy Waveform

To counter the effect of nonstationary background noise added to the signal during transmission over telephone lines, Lamel [49] et. al developed a hybrid endpoint detection scheme for isolated utterances. This detector derives one or more endpoint pair estimates from the energy contours of the utterances. In order to determine the best endpoint pair, word recognition is performed using each possible set of endpoint pairs. The selection of the best pair is based on the best match achieved by the recognition process. The authors call this detector "hybrid" because 1) sets of possible endpoint pairs are obtained, and 2) decision to select the best endpoint pair depends on feedback from the recognition scores. Using the best endpoint pairs corresponding to different utterances, the hybrid endpoint detector produces recognition results close to that obtained from hand-edited endpoints. A real-time implementation of this endpoint location scheme was given in [50].

It should be noted that energy contours are easily derived in practice. Since the energy contour "waveform" contains information on energy level changes occurring in time, it is potentially useful in our application.

4.2.1.4 Peak Detection

Gold and Rabiner [51] analyzed the relative timing relationships between the peaks of low-pass filtered speech signals, and reported a reliable pitch estimation method for speech signals of pitch frequency less than 300 Hz, even in a high level of white-noise background. An extension of this technique was developed to detect periodic and nonperiodic signals [52].

This method is especially susceptible to transient noise, such as those commonly occurring in the everyday acoustic environment. Therefore, this approach is not suitable for us.

4.2.2 Feature Extraction by Filter Banks

Conceptually, the simplest way to extract spectral information from a signal is to pass it through a set of parallel bandpass filters tuned to different mid-frequencies. These mid-frequencies, and the filter bandwidth, would be selected to cover the frequency range of interest. The output of the filter is a measure of the average spectral intensity within the filter band.

White and Neely [53] implemented their broadband speech signal recognizer using a bank of 20 one-third octave bandpass filters. These overlapping filters spanned the frequency range from 100 Hz to 10 kHz. Using a list of multisyllabic words from a North American dictionary, they achieved a recognition accuracy of 99.6 % in their experiments. Another filter-bank based speech recognizer was developed by Kwok, Tai and Fung [54] for the identification of the monophonemic Cantonese digits zero to ten.

With 12 eight-pole overlapping filters, this recognizer provided an average recognition accuracy of 96.8 %.

In industry, NEC has developed its integrated filter-bank based isolated word recognition LSI chip set [55]. The feature extraction processor of this chip set consists of eight biquad digital bandpass filters spanning the frequency range from 100 Hz to 5.0 kHz. This chip set employs a specific data compression algorithm to remove redundancy in signal spectral features, and is "firm-wared" with dynamic programming algorithm for dynamic time warping calculations for signal recognition. A recognition accuracy of more than 98 % was reported.

Miyazaki and Ishida [14] developed a traffic alarm sound monitor for aurally handicapped drivers. This traffic alarm sound monitor consists of seven bandpass filters followed by seven line spectrum detectors. In order to reduce the false-alarm triggering due to the squeaking noises of brakes, tires, engine-noise at high revolutions, wind noise at high-speed driving, human voice, and music, an error suppression circuit was designed to detect the sudden rise of the SPL of the input signal. The successful detection of traffic alarm sounds depends on both the outputs from the seven line spectrum detectors, and the error suppression circuit. During field tests of this monitor on moderately crowded downtown roads in Tokyo, on the average one false-alarm per three minutes was observed.

For our application the filter bank approach offers the advantages of robustness, noise-resistance, and straightforward implementation at a low cost. These will be discussed in more detail in Section 4.5.

4.2.3 The LPC/AR Model

The LPC/AR model assumes that signals can be parametrically modeled as the outputs of a linear, time-varying system excited by either quasi-periodic pulse trains, or random noise. The LPC/AR signal analysis technique has been widely applied to seismic and speech signal processing. To discriminate between earthquakes and underground nuclear explosions, Tjostheim [56] employed a third-order autoregressive model to analyze short period seismic events. The extracted AR parameters produced two discernible clusters characterizing earthquakes and explosions, respectively.

So far LPC/AR parameters have been proven to give the most effective characterization of speech signals. These LPC/AR coefficients represent the combined information about the formant frequencies, their bandwidth, and the glottal waveforms [57]. Therefore, during the past decade, considerable effort was directed at the study of the performance of LPC/AR-based isolated word recognizers. Ackenhusen and Oh [58] implemented an eighth-order LPC-based isolated word recognizer using an AT&T DSP-20 processor. This recognizer has also been used in research for 1) statistically clustered templates for speaker-independent word recognition, 2) recognition based vector quantization, and 3) recognition based hidden Markov Modeling (HMM) of speech signals. Dautrich et al. [59] demonstrated that in high SNR environments and for signals transmitted via telephone lines, LPC-based recognizers can perform several percentages better than filter-bank based recognizers.

In considering an LPC/AR approach for WARNSIS, we must deal with two problems inherent to this technique. First, the order, 'p' , of the LPC/AR signal analysis has to be estimated. Different criteria exist for estimating 'p' for the LPC/AR analysis, but these criteria are signal dependent [24]. Second, the LPC/AR signal analysis is very vulnerable to noise interference [60]. Since the LPC/AR model tends to fit

spectral peaks more accurately than the valleys [26], it is logical to compensate those spectral peaks caused by noise interference by increasing 'p' in noisy environments. Unfortunately, for a practical recognizer, 'p' must always be fixed and independent of the varying unknown signals received. Tierney [61] showed that noise reduction should be applied prior to the analysis to ensure the best LPC/AR based recognition system performance in noisy backgrounds. To compensate the LPC/AR parameter variations due to different noise sources, the derived LPC-cepstral coefficients with different weighting factors were adopted as signal features. Improvement in system recognition performance was reported in [63,64].

To implement a real-time LPC/AR based recognizer with "intelligent" noise pre-filtering for our application, a complex multiple-processor based system would be required. Such complexity makes this approach undesirable for WARNSIS.

4.2.4 LPC-derived Cepstral Coefficients

Pioneer work of investigating the effectiveness of using different speech parameters for speaker identification and verification was done by Atal [62]. He concluded that LPC-derived cepstrum coefficients provided better identification performance than either LPC coefficients, or signal autocorrelation coefficients, or signal impulse response filter coefficients of an all-pole filter derived from the estimated LPC/AR coefficients.

Recently, the use of LPC-derived coefficients for speech signal recognition has been reconsidered by Juang et al. [63] who applied bandpass liftering in speech recognition. He showed that bandpass liftering of the LPC-derived cepstral coefficients (equivalent to applying a smoothing window) tends to reduce undesirable spectral sensitivity by smoothing the spectral peaks without distorting the fundamental formant structure. Such undesirable spectral sensitivity may be caused by the presence of spectral notches or zeros in the signal spectrum, introduced during signal transmission, by filtering,

or by improper preemphasis. Smoothing transforms the original LPC-derived cepstral coefficients into more reliable parameters. Juang's recognition results showed that the bandpass liftering process produced one percent less error than a process using standard cepstral coefficients.

Hanson and Wakita [64] used "root-power sums" or weighted cepstral coefficients as spectral distortion measures for speaker-dependent isolated word recognition in different noise environments. They showed that for white noise interference, a gain of 16 % in recognition accuracy may be achieved by using weighted rather than standard cepstral coefficients.

This method suffers from the same limitations of complexity and computational requirements as the LPC/AR approach. Therefore, it is equally unsuitable for our application.

4.2.5 The Hidden Markov Model (HMM) Approach

One application of HMM for signal recognition is speaker-independent isolated word recognition. The left-to-right topology of HMM is generally adopted in practice. Such a HMM model has N states and each state corresponds to a set of temporal events in the speech signals. The HMM is characterized by a state transition matrix, and a statistical characterization of the acoustic vectors within the state. A detailed exposition on the application of HMM to speech recognition is given in [65]. Rabiner et al. [66] showed that the HMM based recognizer requires ten times less storage, and about 17 times less computation for recognizing a test utterance than does an equivalent recognizer using LPC coding and DTW. This is at the expense of a slight increase in error rate, and of extensive computation while training the model with a reasonable large ensemble of utterance samples. The improvement of the HMM performance in different noisy environments has received considerable attention in the last few years [67].

Considering HMM for our application, we must concern ourselves with the topology of the model. Based on such a topology, the Baum-Welch algorithm could be employed to extract the statistical parameters of the model [65]. To evaluate these probabilistic model parameters, scaling of temporary results must be performed with great care to avoid underflow problems which occur even when mainframes are used [66]. Therefore, HMM appears to be unattractive for hardware implementation using integer arithmetic amenable to real-time operation.

4.3 Overview of the Recognition Scheme for WARNSIS

In the selection of the recognition scheme for WARNSIS the following criteria must be considered:

1. reliability and robust recognition performance in different noise environments;
2. real-time operation;
3. portability; and
4. reasonable cost.

Our preliminary experiments have shown that neither timing nor short-time spectral information is sufficient on its own for reliable recognition performance (see Chapter 6 for performance results). Since both timing and spectral information contributes unique identifiers, a "hybrid" recognition scheme, utilizing both timing and short-time spectral information was designed for WARNSIS (Fig. 4.28). In particular, our design uses timing features as "tokens" to assign sounds to various groups (steady, on-off, variable, etc). Spectral analysis is then used to correlate the spectra of the unknown sound with the spectra of the warning sounds belonging to that group. We have

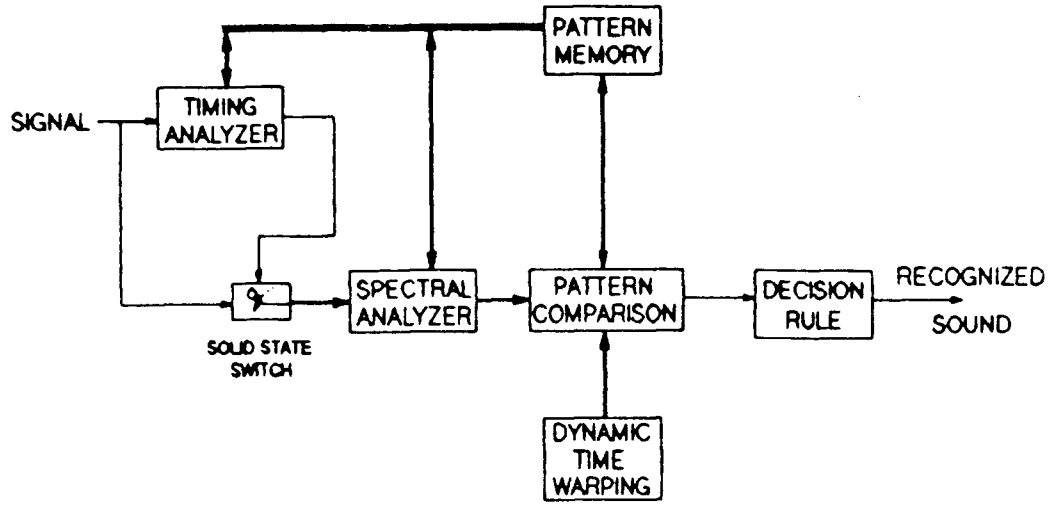


Figure 4.28: The 'hybrid' recognition scheme for WARNSIS

designed a unique analyzer which produces timing information and obtains spectra using the filter bank approach.

Operationally, the system works as follows. In the training stage, the warning sounds of interest are analyzed, and relevant timing information is derived and stored in the timing pattern memory. Consequently, short-time spectra of these sounds are generated by the spectral analyzer. The short-time spectra of warning sounds are classified and stored in the spectral pattern memory according to the group classification determined earlier by timing analysis.

In the recognition stage, two types of pattern comparisons are performed sequentially, before a decision is reached to declare a successful recognition for a specific warning sound. The first stage involves the timing pattern comparison between the timing features of an unknown signal and the pre-stored timing patterns. If the matching criteria are not satisfied for any of these patterns, no spectral analysis is performed

on the incoming signal, and the timing analysis resumes for the next sample.

If a match is found with one of the timing patterns, the signal is assigned to the corresponding "group", and spectral extraction and pattern comparisons are performed on it. Based on the minimum distance score computed for the pre-stored templates, the unknown signal is recognized as the corresponding warning sound. The details of the design are given in Sections 4.4 and 4.5.

Since pattern comparisons involve the most intensive computations in producing a set of distance measures (similarity measures), any possible reduction in number of comparisons between the unknown signal and the pre-stored templates enhances the real-time performance of recognizers. In our recognition scheme this is achieved by making use of the timing features to group warning sounds. An additional use of timing information is to prevent unnecessary spectral and pattern analysis work when only noise is present.

4.4 Extracting & Classifying Timing Information

One or two signal processing steps may be needed to extract timing features from steady or burst-type warning sounds (Fig. 4.29). The first step classifies warning sounds according to the features derived from signal waveforms. For steady sounds, timing feature extraction terminates after this processing; for burst-type sounds, the processing proceeds to the next step, which estimates the repetition period.

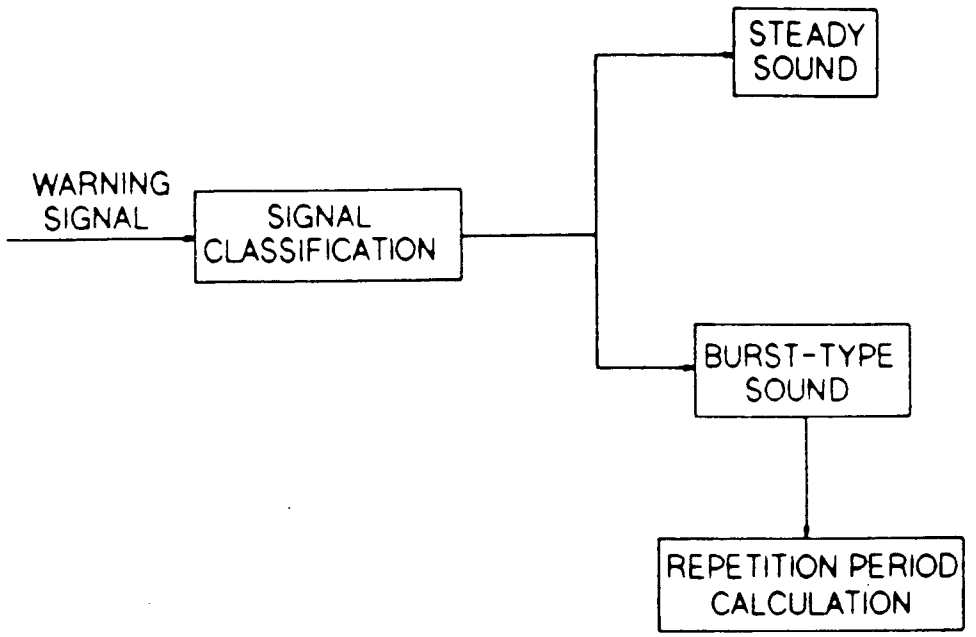


Figure 4.29: Block diagram of the Timing Feature Extractor

In real-life, warning sounds are modified acoustically by the environment, and the addition of unwanted sounds. These background sounds may be either continuous, or transient. In addition, what a microphone receives from a source depends on the paths between the two, their orientation with respect to each other, and the sound modification characteristics of the environment.

Extracting timing features from distorted and noisy signals has not been addressed by other workers in the literature. Compelled by the demands of real-life circumstances, we developed the algorithm presented here to deal with this problem. This development was inspired by the work of Gold and Rabiner [51], and Lamel [49].

4.4.1 A Scheme to Extract Timing Features

We have demonstrated in Chapter 3 that the contour characteristics of the short-time average absolute amplitudes (STAAA) of warning sounds are distinctively defined for steady and burst-type sounds. Working with the short-time average absolute amplitude is more attractive for us than the average energy used in Lamel's work because the short-time average absolute amplitude: 1) is a simple measurement which preserves the essential features of the corresponding energy contours, 2) requires no multiplication operations, and 3) has a smaller dynamic range which can be coded in 8 bits. The relationships between the short-time average absolute amplitudes and the average energy of a discrete sequence $x(n)$ are shown in Fig. 4.30.

Since the short-time average absolute amplitude is obtained from an 8-bit A/D conversion, and is coded in integer arithmetic, its dynamic variations are limited to 256 levels. The value of the short-time average absolute amplitudes is zero when the environmental noise level falls below the threshold value of the A/D conversion system. In order to compress the dynamic variations of the short-time average absolute amplitudes for plotting purposes, we adopted a logarithmic measure to readjust these short-time average absolute amplitude values. This logarithmic measure is:

$$\widehat{STAAA} = 10 \log_{10} (STAAA + 1) \tag{4.20}$$

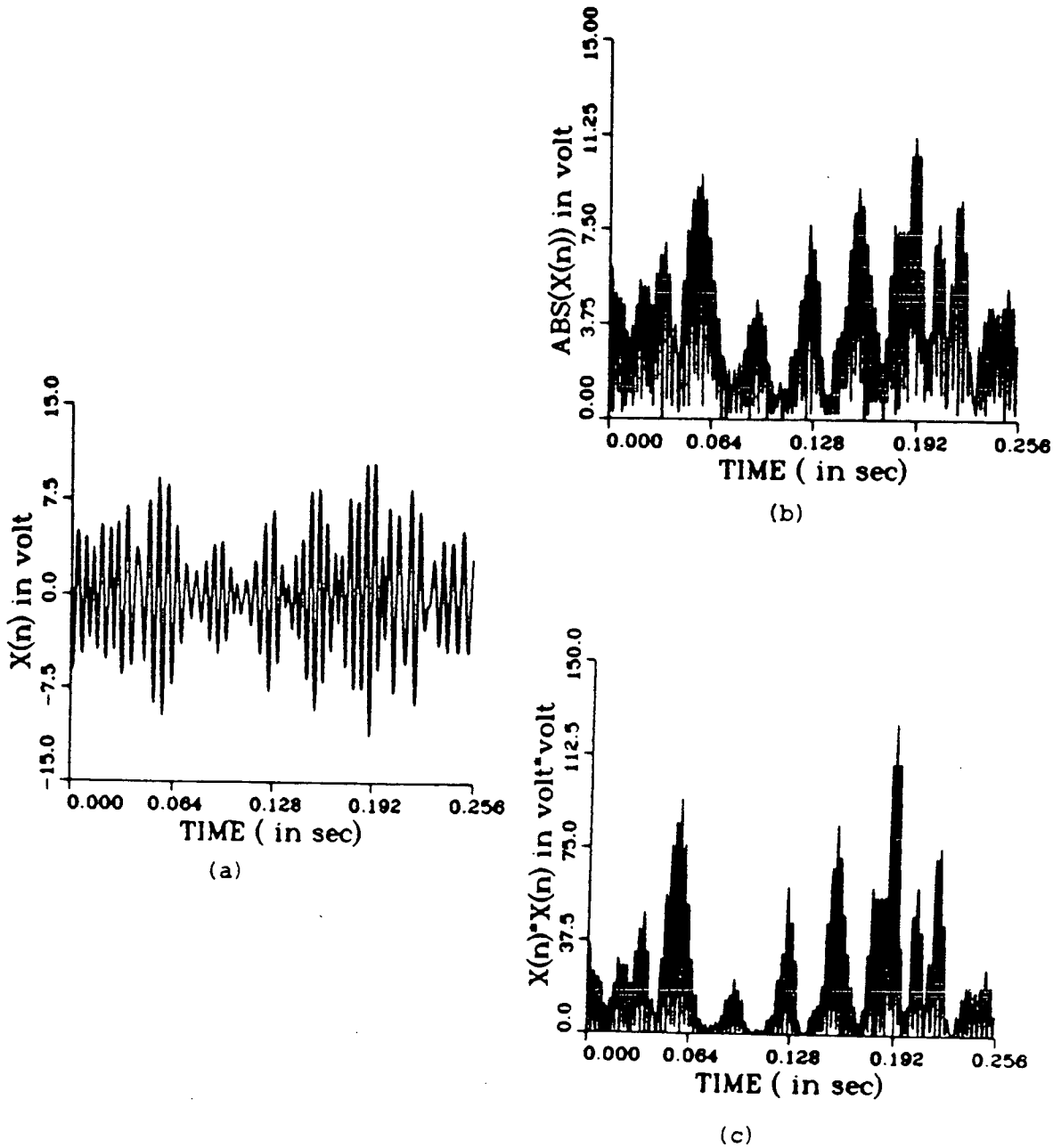


Figure 4.30: Relationships between the instantaneous energy and the instantaneous absolute amplitudes of a sequence, x(n). (a) : the plot of x(n); (b): the plot of $|x(n)|$; and (c): the plot of $x^2(n)$

Note the value of the short-time average absolute amplitude is incremented by one to prevent the argument of the logarithm to take on the value of zero. The error introduced by this is not relevant since the essential features of the contour are not affected.

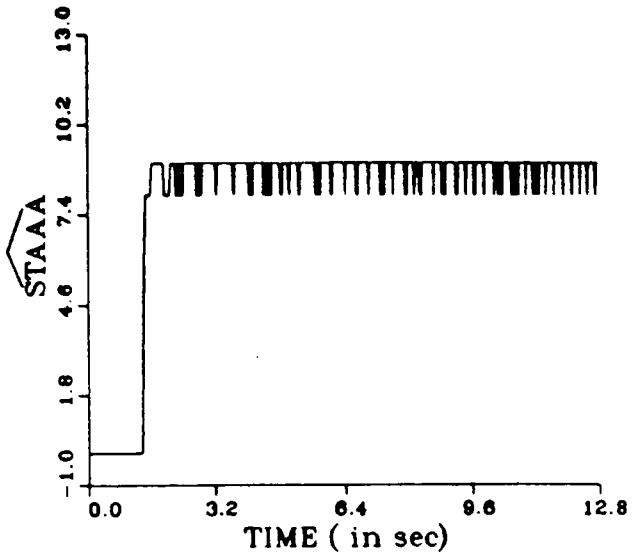
From the $\widehat{STAAA}$ contours of warning sounds, the break-points or transitions (rising and falling) in these waveforms are located. Timing features of warning sound are thus derived from the timing relationship between these transitions similarly to the method of Gold and Rabiner. Fig. 4.31 (a) gives the $\widehat{STAAA}$ contour of a steady sound, whereas Fig. 4.31 (b) shows the $\widehat{STAAA}$ contour of a burst-type sound.

With reference to Fig. 4.31 (a), a steady sound is identified if a rising transition of the waveform of short-time average absolute signal amplitude is detected, and a new value of short-time average absolute signal amplitude is then maintained for at least four seconds. For burst-type sounds two rising and falling transitions must be detected (T_1, T_3, and T_2, T_4, respectively are shown in Fig. 4.31 (b)). The repetition period (RP) and the average width of signal bursts (AWSB) can then be obtained according to the following equations:

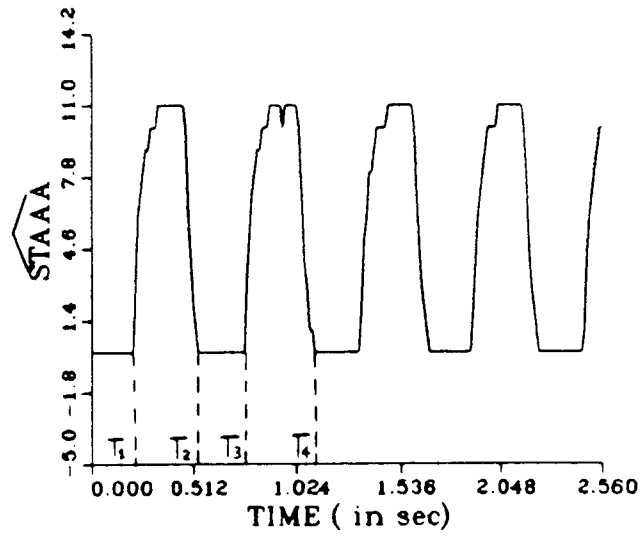
$$RP = \frac{(T_3 - T_1) + (T_4 - T_2)}{2} \tag{4.21}$$

$$AWSB = \frac{(T_4 - T_3) + (T_2 - T_1)}{2} \tag{4.22}$$

To detect these transitions, a signal amplitude threshold is derived from the short-time average absolute amplitude of the acoustic background. This short-time average absolute amplitude is dynamically updated every 12.8 msec to accommodate the acoustic energy variations of the environment. This dynamic amplitude threshold (DAT) provides the baseline level of the background, and is used for transition (rising and falling) detection.



(a)



(b)

Figure 4.31: (a): The $\widehat{STAAA}$ contour of a steady sound; (b): The $\widehat{STAAA}$ contour of a burst-type sound

When the detection scheme starts, the dynamic amplitude threshold is assigned the maximum value. Then, the incoming short-time average absolute amplitude is compared to the dynamic amplitude threshold. If the incoming short-time average absolute amplitude is less than the dynamic amplitude threshold, the dynamic amplitude threshold is updated by averaging the short-time average absolute amplitude and the dynamic amplitude threshold:

$$DAT(updated) = \left\{\frac{DAT + STAAA}{2}\right\} \tag{4.23}$$

Updating ensures that the dynamic amplitude threshold follows the amplitude level changes due to background noise. This method continuously adjusts the dynamic amplitude threshold downwards until a rising transition is detected. Such a transition may be either due to a warning signal, or due to a sudden increase in background noise. If no rising transition is detected for a period of four seconds, the dynamic amplitude threshold is reset to its initial value, and the search for a rising transition resumes. Fig. 4.32 shows an example how the dynamic amplitude threshold adapts to acoustic energy variations in the environment.

Since the dynamic amplitude threshold and short-time average absolute amplitudes are expressed in integer arithmetic, the value of the minimum detectable difference between them is one. To avoid the false detection of a rising transition due to random noise disturbance, we set the value of the threshold for detecting this transition as two. If the short-time average absolute amplitude is larger than the dynamic amplitude threshold by this preset threshold, a rising transition is detected and a reference time marker (T_1) is set. A corresponding falling transition will be detected and marked (T_2) as soon as an incoming short-time average absolute amplitude falls below the dynamic amplitude threshold. However, if no falling transition is detected in a period of four seconds (maximum allowable burst width), this sound may be a steady sound. To

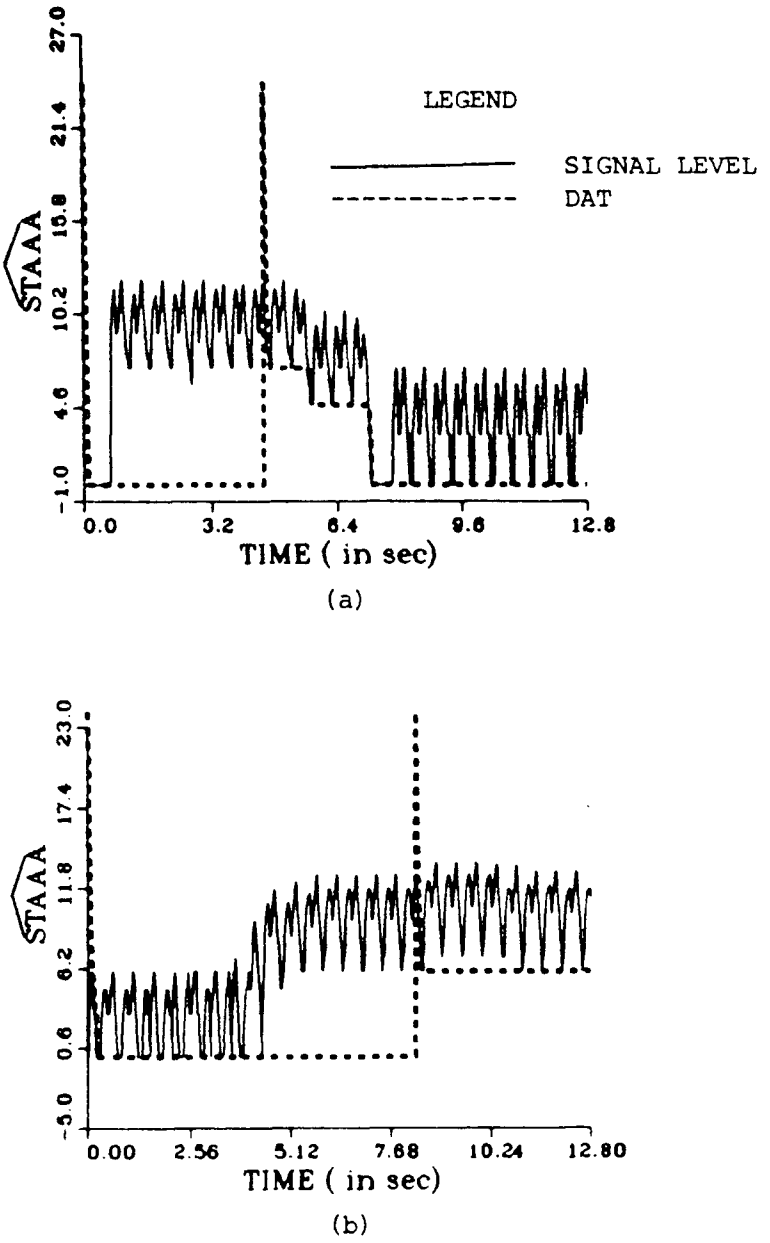


Figure 4.32: Two typical examples of how the dynamic amplitude threshold adapts to acoustic energy variations of the environment. (a): sudden decrease in signal levels; (b): sudden increase in signal levels

confirm this, the dynamic amplitude threshold is reset to its initial value, and if no rising transition is detected in one second period following, the sound is declared to be a steady sound, and the timing feature extraction process terminates.

If a rising transition is detected within one second, the search for its corresponding falling transition continues, and the hypothesis of a steady sound is rejected. Assuming a burst-type signal this detection process continues until a second transition pair set is detected and marked with T_3 and T_4 for rising and falling transitions, respectively. Consequently the RP and AWSB are computed and the timing feature extraction process terminates.

A typical example of the detection of a siren sound is illustrated in Fig. 4.33 (a), and Fig. 4.33 (b) demonstrates how the steady sound detection scheme rejects non-steady sounds.

This scheme works well for warning sounds in backgrounds with steady noises. To deal with nonstationary noises such as radio broadcasts, and transient sounds due to door slamming or movement of chairs, additional parameters and conditional tests are included in the scheme. These are: 1) the minimum burst duration (MBD), and 2) the maximum inter-arrival time (MIAT) between two consecutive signal bursts. As shown in Fig. 4.34, any signal with duration less than the MBD is declared as an unwanted transient. Furthermore, if the signal shows pulsative variations that last longer than MIAT, the hypothesis of a burst-type sound is rejected.

These conditional tests were incorporated into the basic scheme as follows. When any signal burst is detected, its width is calculated and compared to the MBD. If the computed width is less than the MBD, the detected burst is treated as transient noise, and the search continues. If the burst is longer than the MBD, the system waits until a second transition is detected. The time difference between following transitions

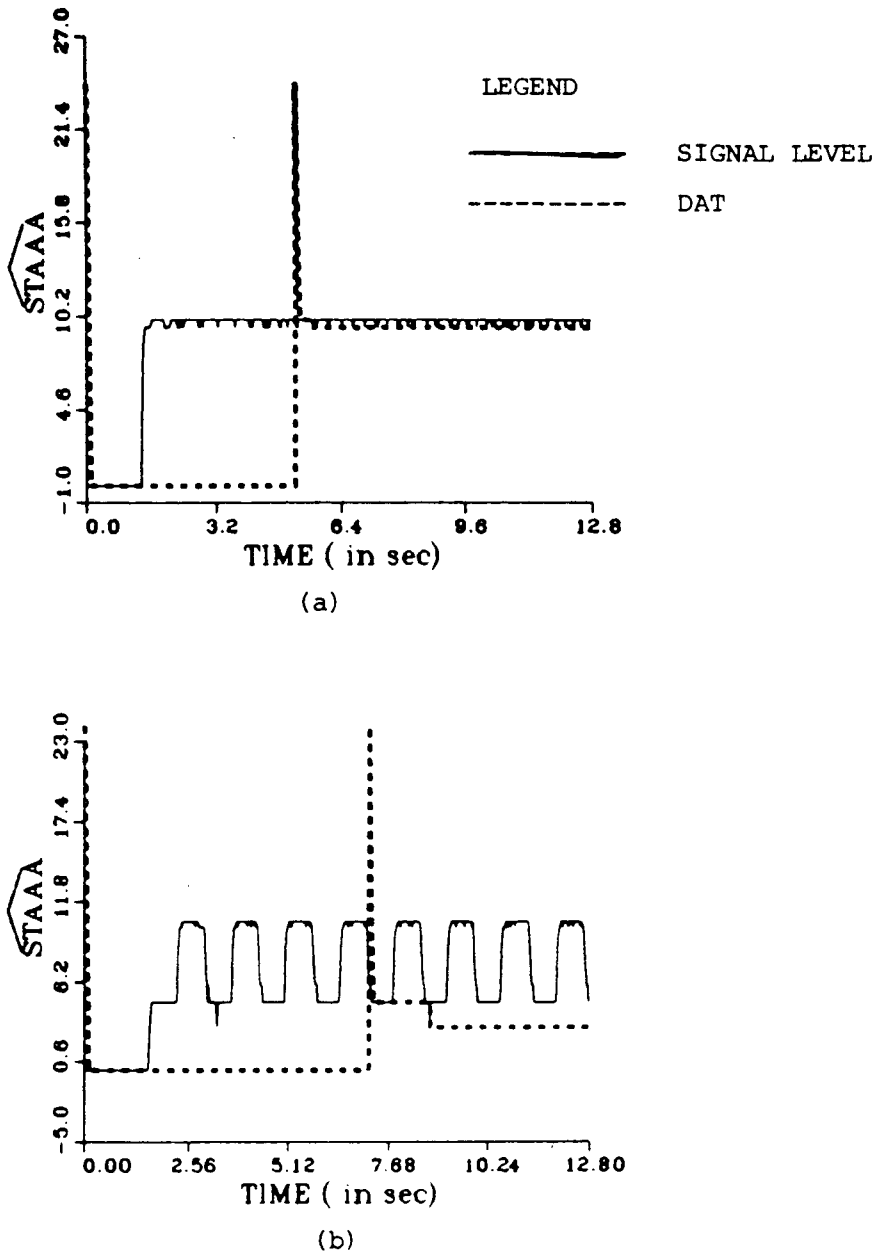


Figure 4.33: (a) : Detection of a steady sound; (b): An illustration of how the scheme rejects a non-steady sound

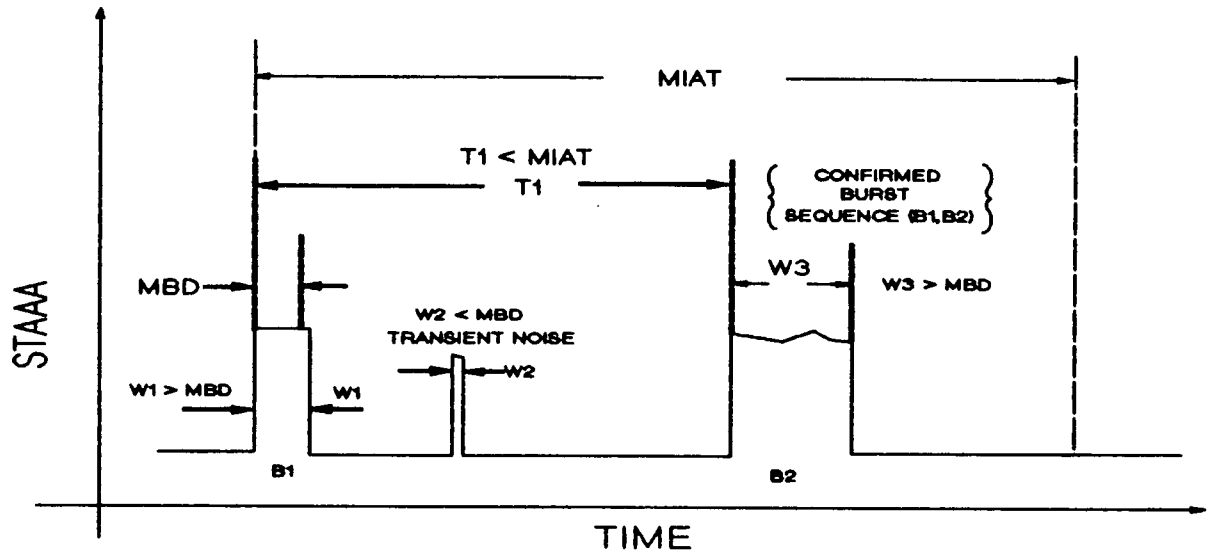


Figure 4.34: A demonstration of the use of the MBD and MIAT to refine the basic warning sound analysis scheme

is computed, and compared to the MIAT. If this time is longer than the MIAT, the hypothesis of a burst-type sound is rejected, dynamic amplitude threshold is reset, and the timing feature extraction process is reset and restarted.

A flowchart of the complete scheme for timing feature extraction is shown in Fig. 4.35. The program was written in INTEL 8088/8086 assembly language for real-time operation. The hardware developed in Chapter 3 for timing parameter measurement is employed here to generate the instantaneous absolute amplitudes of warning sounds.

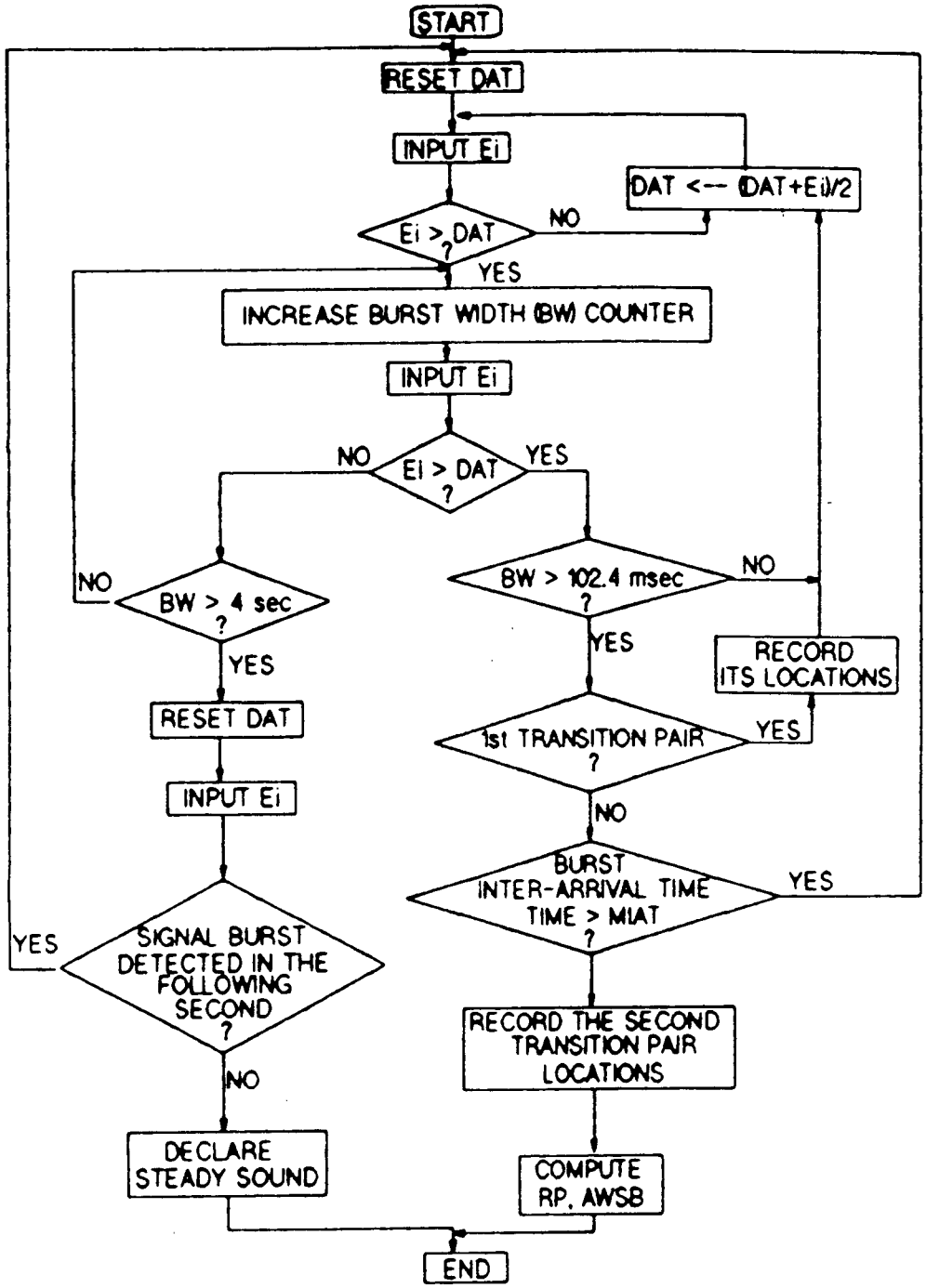


Figure 4.35: Flowchart of the Timing Feature Extraction Scheme

4.5 Extracting Spectral Information

As shown in Fig. 4.28, timing analysis is followed by spectral analysis. The latter is initiated only if the timing analysis indicates the possibility of the presence of one of the recognizable warning sounds. Since timing analysis of warning sounds gives the time markers for the rising and falling transition of sound bursts, it is equivalent to the end-point detection of isolated utterances [49]. Thus, the timing analyzer conveniently provides the on/off control for the spectral analyzer.

4.5.1 Feature Extraction

In our review of methods of obtaining spectral information from signals in real-time we have already indicated our preference for the filter-bank approach. Firstly, the filter-bank method works well for simple speech signals, and the warning signal spectra are simpler than the spectra of speech. In particular, Dautrich et. al. [59] demonstrated that for spoken digits the performance of a filter-bank recognizer was equal to the performance of the more complicated LPC recognizer. Secondly, as shown by Lim [60], in noisy environments filter-bank recognizers are less error prone than the LPC-based recognizers. This a very important criterion for us, since our specific goal is to recognize warning signals in low SNR situations. Thirdly, filter-bank recognizers are fast, are relatively simple, and are commercially available at a reasonable cost.

Fig. 4.36 gives the block diagram of our spectral analyzer which uses a filter-bank. Signals pass through a bank of eight bandpass filters covering frequency bands from 100 Hz to 5.0 kHz. The output of each bandpass filter is passed through a full-wave rectifier, and low-pass filtered to give a value related to the energy of the incoming warning sounds in each band. The outputs of bandpass filters are sampled (typical rate 50 – 100 Hz) to give a segment of a feature set. At a time index k, a segment of

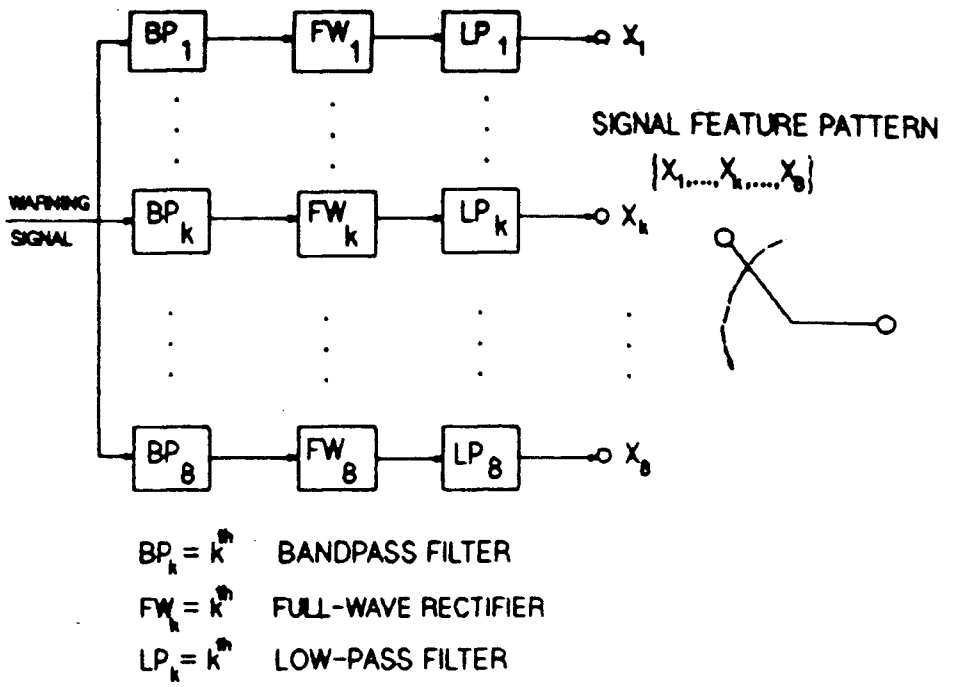


Figure 4.36: Filter-bank analysis of Warning sounds

parallel outputs $\{x_1(k), x_2(k), \ldots, x_8(k)\}$ defines a 8^{th} order feature vector $\mathbf{X(k)}$ as,

$$\mathbf{X(k)} = \{x_1(k), x_2(k), \ldots, x_8(k)\} \tag{4.24}$$

A complete spectral pattern of a warning sound is given as,

$$\mathbf{R} = \{\mathbf{X(1)}, \mathbf{X(2)}, \ldots, \mathbf{X(k)}, \ldots, \mathbf{X(N)}\} \tag{4.25}$$

In the recognition stage these reference patterns are compared to the spectral pattern $\mathbf{T}$, of an unknown signal. Dynamic time warping is employed to provide a quantitative similarity measure between reference and unknown patterns.

4.5.2 Dynamic Time Warping (DTW)

The basic idea of DTW is to provide an optimum similarity measure between two patterns of different time durations. DTW can compensate for the nonlinear time misalignment of patterns which may be caused by noise giving rise to errors in the detection of endpoints.

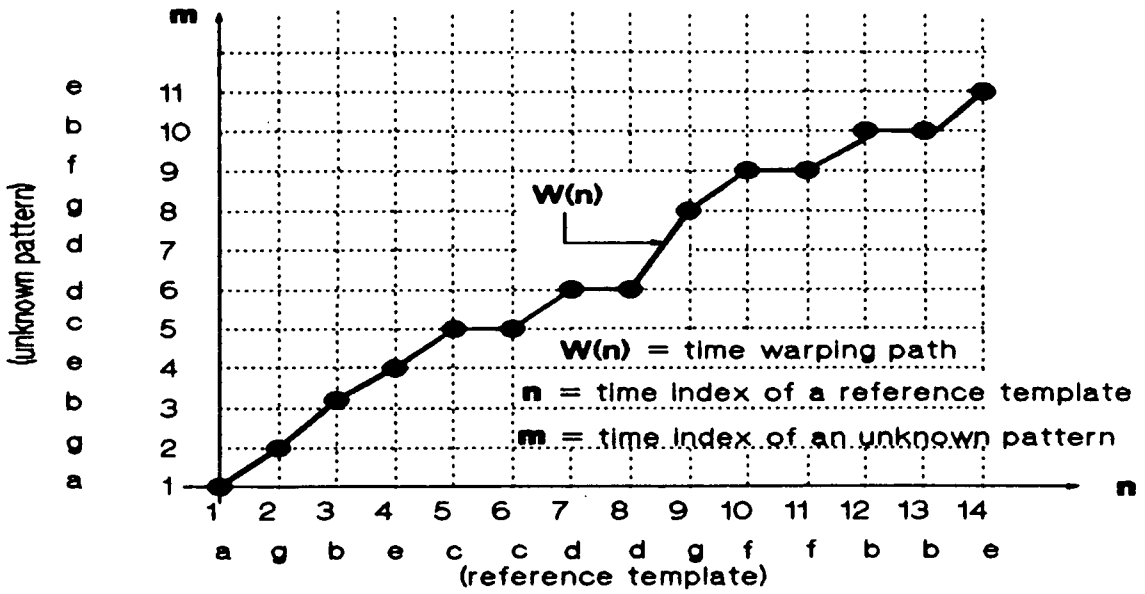
Conceptually, matching between these patterns involves the search for a time warping function for which the segment-to-segment comparison is optimal according to some distance criteria. Fig. 4.37 gives an example of the optimum match between a reference template and an unknown pattern whose feature sets consist of letter alphabets.

Mathematically, the problem can be stated in the following manner. Consider $R(n), T(m) \;\; \forall\, n \;\in [1, N], \;\; m \;\in [1, M]$ where $N \neq M$ (in general), and $R(n), T(m)$ are the reference and the test pattern at time indices n, m, respectively. DTW is to find an optimum time warping function $w(n)$ to minimize the accumulated distance, (D_A^*) between these two patterns with D_A^* given by

$$D_A^* \;=\; \underbrace{\min}_{\{w(n)\}} \sum_{n=1}^{N} \mathbf{d}\,[\, R(n), T(w(n))\,] \tag{4.26}$$

where $\mathbf{d}\,[\, R(n), T(w(n))\,]$ is defined as the frame-by-frame (segment-by-segment) distance measure. Several possible distance measures can be used, depending on the form of the feature sets [37]. In this discussion, the absolute magnitude difference is used as a distance measure. Thus, $\mathbf{d}\,[\, R(n), T(w(n)\,]$ is expressed by,

$$\mathbf{d}\,[\, R(n), T(w(n))\,] \;=\; \sum_{k=1}^{L} |X_n^R(k) - X_{w(n)}^T(k)| \tag{4.27}$$



n	W(n)
1	1
2	2
3	3
4	4
5	5
6	5
7	6
8	6
9	8
10	9
11	9
12	10
13	10
14	11

Figure 4.37: An example of pattern matching between a reference template and an unknown pattern

where

$X_n^R(k)$ = the k^{th} bandpass filter output at time index n of a reference spectral pattern,

$X_{w(n)}^T(k)$ = the k^{th} bandpass filter output at time index $w(n)$ of a test spectral pattern, and

L = the total number of bandpass filters of the filter-bank used.

Since one would expect the optimum warping path to be close to a straight line, most of the computations at the beginning and the end of this path can be reduced by establishing boundary conditions for the search. In general, the optimum warping path function can be obtained by Dynamic Programming [39,53,69].

Rewriting the original path searching equation, a recursive accumulated distance function, denoted as $D_A(n,m)$, is defined as

$$D_A(n,m) = \mathbf{d}\,[\,R(n),T(m)\,] + \underbrace{\min}_{l \le m}\,[\,D_A(n-1,l)\,] \tag{4.28}$$

The above equation defines the minimum accumulated distance to grid point (n,m), and consists of the local distance between feature set $R(n)$ and $T(m)$, plus the minimum accumulated distance to the grid point $(n-1,l)$ where l are the possible values of m constrained by a given set of local paths. As an example, Fig. 4.38 shows one of the possible sets consisting of three paths leading to the grid point (n,m): $(n-1,m)$, $(n-1,m-1)$, and $(n-1,m-2)$. To ensure that the time warping function is monotonically increasing, an additional path constraint is applied. Specifically, if the best path to grid point $(n-1,m)$ came from grid point $(n-2,m)$, then no path can lead from the grid point $(n-1,m)$.

Formulating these path constraints mathematically, we obtain

$$\begin{aligned} w(n) - w(n-1) &= 0,1,2 \quad if \quad w(n-1) \neq w(n-2) \\ &= 1,2, \quad if \quad w(n-1) = w(n-2) \end{aligned} \tag{4.29}$$

Therefore, substituting the above constraint equations into Eq. 4.28, we have the DP recursive solution to the DTW,

$$\begin{aligned} D_A(n,m) &= \mathbf{d}\,[\, R(n), T(m)\,] + \\ &\quad \mathbf{min}\,\{D_A(n-1,m)\; g(n-1,m), \\ &\quad D_A(n-1,m-1), D_A(n-1,m-2))\,\} \end{aligned} \tag{4.30}$$

where

$$\begin{aligned} g(n-1,m) &= 1 \quad if \quad w(n-1) \neq w(n-2) \\ &= \infty \quad if \quad w(n-1) = w(n-2) \end{aligned} \tag{4.31}$$

with boundary conditions governed by,

$$w(1) = 1 \tag{4.32}$$

$$w(N) = M \tag{4.33}$$

and continuity criterion for $w(n)$ expressed by,

$$w(n) \geq w(n-1) \tag{4.34}$$

This iteration is carried out over all valid m, for each n sequentially from $n = 1$ to N. The constraint of Eq.(4.33) means that the last segment of the template and test signal must coincide and the distance function is $D_A(N,M)$. When the last segment is reached, the warping path $w(n)$ is completely defined.

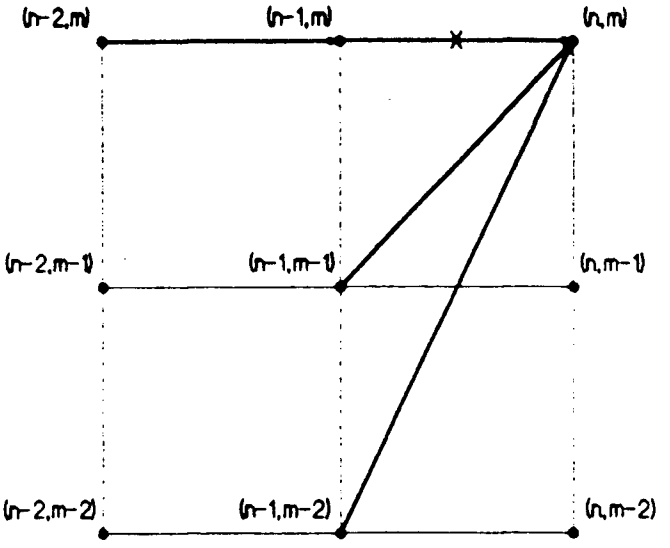


Figure 4.38: Local path constraints for DTW

The complexity involved in DTW implementation depends on the boundary conditions, the local path constraints, and on the distance measure. Both Sakoe and Chiba [39], and Myers [70] have investigated the effects of varying these factors on both speed and performance of the DTW algorithm in speech-recognition systems. They have shown that only small differences are found in performance for a fairly wide range of variations of these parameters.

If the reference and test patterns are dissimilar, the distance measures will be consistently large. Therefore an accumulative distance limit must be established to stop unnecessary computation. Whenever an accumulated minimum distance is obtained, it is compared to the distance limit. If it is larger than the limit value, the matching process between this reference and the test pattern terminates, and another reference pattern is used to compare to the test pattern.

Chapter 5

Design & Implementation

Utilizing the methodologies discussed in the previous Chapters, we designed and implemented a WARNSIS prototype. Fig. 5.39 shows the four main hardware building blocks of our device: the microphone, the signal conditioner, the control & timing processor (CTP), and the spectral recognizer (SR).

5.1 Timing Analyzer

5.1.1 Microphone

A microphone is used as the transducer that receives environmental sounds and produces the electrical input for the WARNSIS. The characteristics of the microphone play a crucial role in determining the quality of the signal that is fed to the analog signal conditioner. We selected a SONY model directional microphone which has a frequency response of 100 – 15000 Hz, and a sensitivity of –70 $\pm$ 3 dB (with reference to 0 $dB = 1V/\mu bar$) at 1000 Hz. It is an electret-condenser microphone with two selectable angles (90° and 120°) of reception. A microphone with a narrower angle of reception may provide better spatial separation between the signal and the background noise when the sources are separated, and the microphone is oriented at the direction of the signal source. On the other hand, when such a microphone is not oriented in direction of the signal source, the signal quality may be degraded substantially.

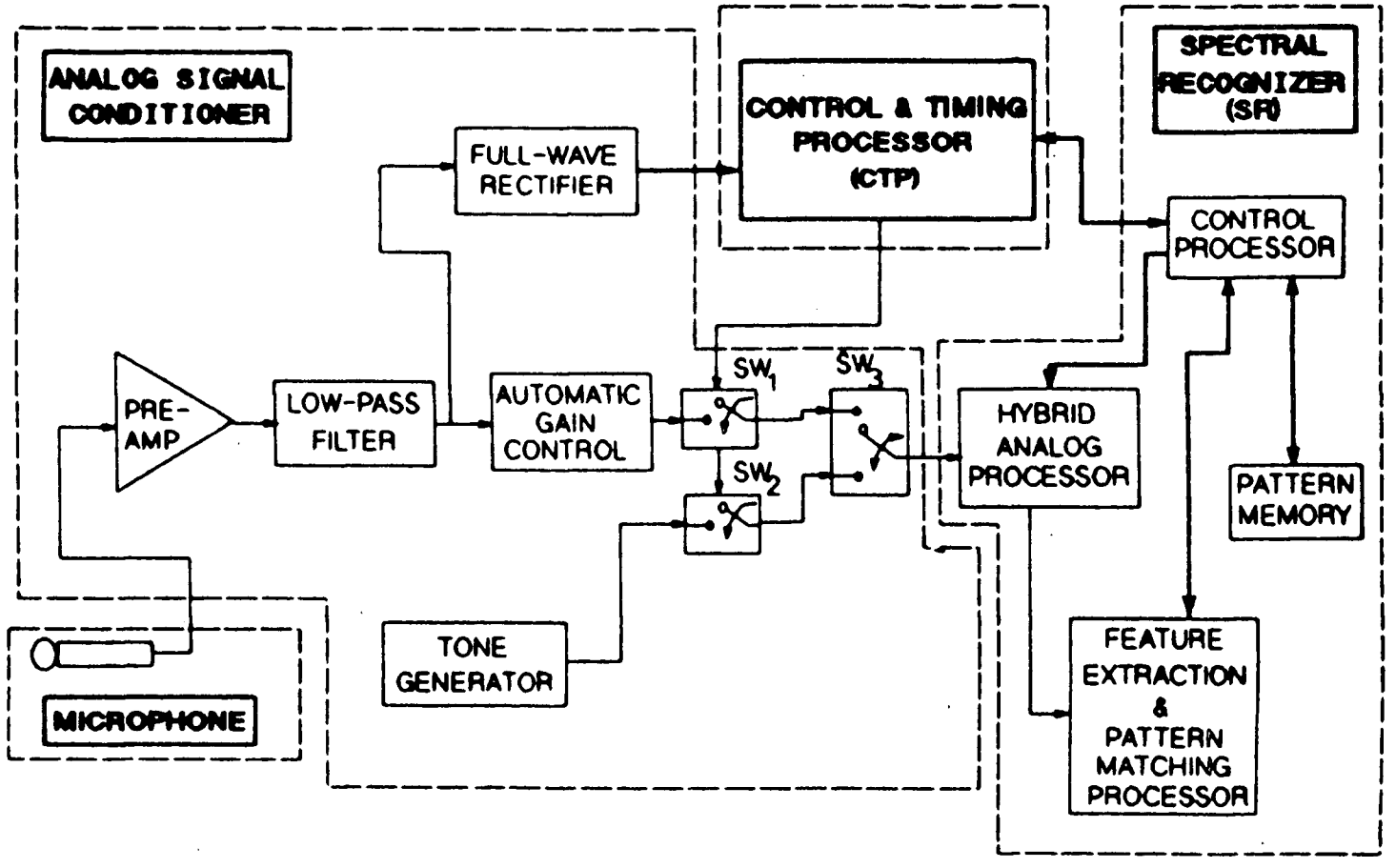


Figure 5.39: The building blocks of WARNSIS

5.1.2 Analog Signal Conditioner

The function of the analog signal conditioner is to: 1) pre-process the microphone output to generate an analog input for the spectral recognizer, and 2) to calculate the instantaneous amplitudes of the signal for the use of this information by the control & timing processor. Correspondingly, the signal conditioner consists of an audio pre-amplifier, a low-pass filter, an automatic gain controller (AGC), two solid-state analog switches, an SPDT manual switch, a full-wave rectifier, and a 1 kHz calibrating tone generator.

The voltage produced by the directional microphone is fed to an audio-preamplifier. Since the noise characteristics of an audio pre-amplifier system depend primarily on the noise generated by its first stage, we used a low-noise audio operational amplifier (with noise characteristic of 9 nV^2/Hz). This pre-amplifier provides a voltage gain of 58.3 dB at a 100 Hz – 8.0 kHz bandwidth.

To reduce the unwanted high frequency content of the signal, the pre-amplified signal is fed to a 6^{th} order Chebyshev low-pass filter, with a cut-off frequency at 6.4 kHz. This 6^{th} order filter was constructed from three cascaded second order filters. The overall voltage gain of the filter chain is 11.2 dB. The filtered signal is consequently branched into two signal processing modules: the full-wave rectifier and the AGC. We used the same full-wave rectifier module as the one described in Chapter 3.

The AGC is employed to maintain the signal level at values that prevent signal clipping. This AGC limits output signal amplitude variations to 3 dB when the incoming signal varies by 60 dB.

The analog switch, SW_1, provides a windowed segment of the signal from the AGC output. This switch is controlled by the control & timing processor, and the gating window duration is set to 470 msec. This gating duration can easily be altered by an

external timing resistance. The control & timing processor will activate SW_1 according to the timing information extracted from the instantaneous amplitudes of the signal. The output of the AGC module is then fed to an SPDT manual switch (SW_3).

The 1 kHz calibrating tone has a peak-to-peak voltage of three volts. The tone generator is connected to another analog switch (SW_2) whose output is tied to the second input of the SW_3. The function of the 1.0 kHz tone is to calibrate the input signal level of the hybrid analog processor of the spectral recognizer during the initialization of the WARNSIS. In this prototype, the user has to manually flip the switch to determine which one of the two signals (the processed signal from the microphone, or the calibration 1 kHz tone) is fed to the hybrid analog processor.

5.1.3 Control & Timing Processor (CTP)

The control & timing processor consists of decoding circuits, a software programmable port, and a microprocessor. The port (INTEL 8255, software programmable) allows parallel communication between the microprocessor and the spectral recognizer to monitor the step-by-step operation of the recognizer logic, and is the gateway for the control signal that operates the switch in the analog signal conditioner. The microprocessor is an INTEL 8088, housed in a personal computer.

The first function of the control & timing processor is to perform 'real-time' timing analysis as described in Chapter 4. Its second function is to initiate the spectral recognition process.

5.2 Spectral Recognizer (SR)

The spectral recognizer hardware consists of an NEC LSI speech chip set. This set has three processors as shown in Fig. 5.39: 1) the hybrid analog processor (MC4760),

2) the feature extraction and pattern matching processor (μPD7761), and 3) the control processor (μPD7762) [55]. We selected this speech recognition chip set since it has the features required by our method:

- filter-bank based recognizer;
- signal frequency bandwidth of 100 Hz to 5.0 kHz;
- allowable windowed signal duration from 0.2 sec to 2.0 sec;
- supports a maximum storage of 512 signal templates;
- uses syntax number in grouping signal templates;
- pattern comparison using DTW via "firmwared" DP method;
- simple set of twelve macro commands to operate the chip set; and
- average recognition time of 0.5 sec.

This chip set, coupled with external memory for signal template storage, constitutes the spectral recognizer of our WARNSIS.

5.2.1 The Hybrid Analog Processor (MC4760)

The hybrid analog processor performs signal equalization and digital sampling of input signals. Fig. 5.40 gives a simplified block diagram of MC4760. Signal is accepted to the equalization amplifier whose voltage gain can be altered by varying an external resistance. Since sufficient voltage gain is provided from the signal conditioner, the voltage gain of the equalization amplifier is set to the possible minimum gain (0.59 dB). The gain of the input signal can further be adjusted by a digital programmable attenuator under the control of the control processor. For speech application, this

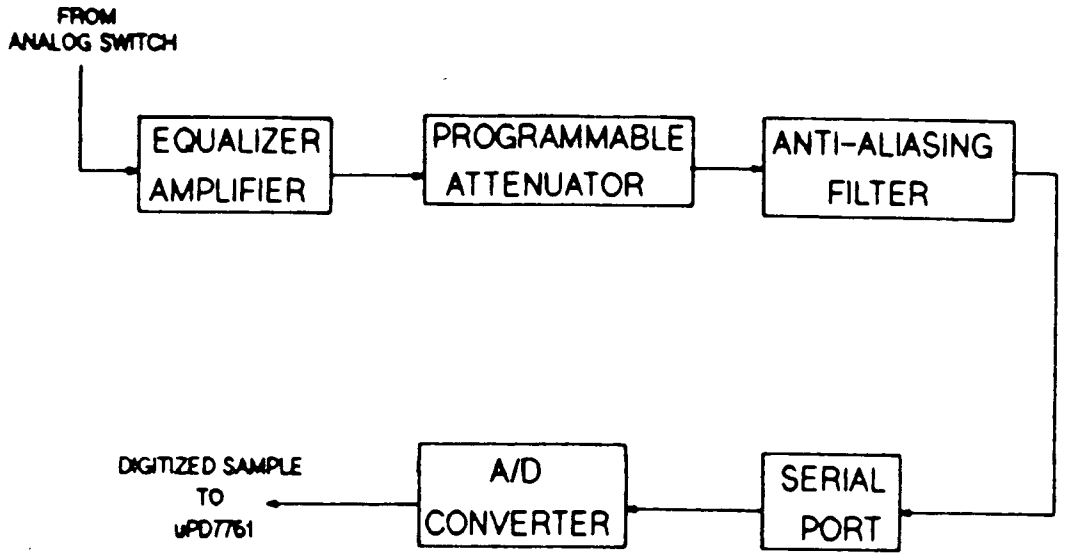


Figure 5.40: Block diagram of MC4760

attenuator compensates for signal level variations due to microphone position. However, in our application signal level equalization is performed by an external AGC circuit, and thus, the attenuator gain is permanently set to unity.

The attenuated signal is then low-pass filtered by an anti-aliasing filter (5 kHz bandwidth), and is input to a built-in 8-bit A/D converter. The converter samples the signal at a rate of 10 kHz, and the sampled data are converted into inverted μ-law PCM codes. Subsequently, this output is serially transmitted to a dedicated serial input port of the feature extraction processor at a 2 MHz clock rate.

5.2.2 Feature Extraction and Pattern Matching Processor (μPD7761)

The μPD7761 is an NMOS device optimized for single instruction cycle arithmetic operation. It runs at a clock rate of 8 MHz, and operates in either of two modes (analysis or pattern matching) as selected by the control processor (μPD7762). A block diagram of the functional operation of the μPD 7761 is shown in Fig. 5.41.

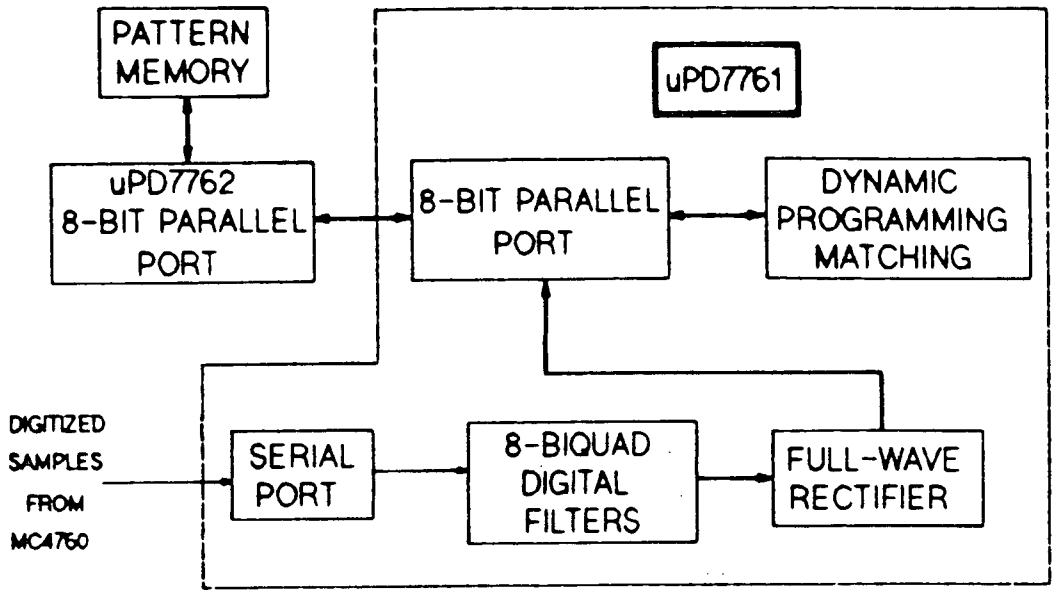


Figure 5.41: Block diagram of the functional operation of μPD7761

In the analysis mode, the μPD7761 accepts digitized data samples from the MC4760 via a dedicated built-in serial port. Data transfer timing is controlled by an input clock at 2 MHz, which is the rate at which data is fed from the MC4760. These samples are analyzed by a 8-channel biquad filter bank firmwared into the on-chip ROM memory. This filter bank spans the frequency spectrum from 100 Hz to 5.0 kHz. Each output of the bandpass filter is full-wave rectified. The rectified outputs are sampled at a frame rate of 12 msec, and sent to the control processor via a 8-bit parallel port. This process is repeated for successive frames until the entire windowed segment of the signal is analyzed.

In the pattern matching mode, the μPD7761 compares the features of the unknown signal with the pre-stored signal templates using the DTW approach. The algorithm

is firmwared onto the chip to perform the computationally intensive distance calculations. Each comparison with a pre-stored template takes an average of 5 ms. Upon completion, the recognition result is transferred to the control processor and subsequent templates are compared, until all templates have been checked.

5.2.3 The Control Processor (μPD7762)

The control processor provides the only communication link between the control & timing processor and the spectral recognizer. In addition, it performs two important functional operations. First, it serves as a system controller for the MC4760 and μPD7761 by providing the necessary control signals to synchronize all operations. Such control signals include the communication protocols with the control & timing processor, the memory selection, read and write signals, reset signal for the MC4760 and μPD7761, and specific command code to initiate the feature extraction and pattern operations of the μPD7761. Secondly, it functions as a spectral feature compressor, by retaining only one of a set of vectors whose values are close to each other [55]. Pattern compression is important because it allows a significant amount of reference memory to be saved, and it speeds up the calculations involved in pattern matching.

When a specific operation code is sent from the control & timing processor to the spectral recognizer, decoding is performed by the μPD7762, providing the necessary control signals for execution. The μPD7762 also reports the result(s) obtained from the execution of the code to the control & timing processor. For example, if a training command code is received by the μPD7762, the following series of events occur:

- the μPD7762 decodes the command;
- it activates the μPD7761 to extract spectral contents from the digitized input signal samples fed from MC4760;

- the spectral information is sent to the μPD7762 for feature compression;
- the compressed spectral features are stored into the external pattern memory; and
- a successful training status flag is sent to the control & timing processor when all training procedures are completed. Otherwise, an error status is reported to the control & timing processor.

5.2.4 Pattern Memory

The chip set can maximally allow 64 kbyte of pattern memory, which stores 512 signal templates. This pattern memory is divided into four banks, each of which consists of 16 kbyte of memory, and can be randomly selected by the spectral recognizer in the training and recognition stages. In our prototype we used 32 kbyte of static RAM.

5.3 Software Program

The software program co-ordinates the functional operations of the timing analyzer and the spectral recognizer. Basically, it consists of different program modules which are responsible for various operational stages of the system. Such stages include the initialization of the system (the timing analyzer and the spectral recognizer), the signal timing analysis, and the signal training and recognition. The program module for the timing analysis is a direct implementation of the algorithm developed in Chapter 4, and the program module for the signal training and recognition was developed by using the specific set of commands provided by the chip-set manufacturer.

We start the detailed description of the software with a summary of the most important commands of the spectral recognizer control language. Then we present the

three major modules of the program. These modules correspond to the three modes of operation of the system: initialization, training, and recognition.

5.3.1 The Command Set of the Spectral Recognizer

Twelve commands are provided to operate the spectral recognizer. These commands are sent to the control & timing processor to initiate specific operations. Each command consists of a command code (8-bits), the required parameter(s), and a termination code marking the end of each command character string. Upon completion of the execution of the command, the status of the operation is reported to the timing & control processor from the μPD7762. A detailed description of the format of each command is given in Appendix B.

One of the special features of the spectral recognizer is the use of syntax numbers to group the reference signal templates. Such syntax numbers can be specified in the training and recognition stages. A valid syntax number can range from 0 – 127 [55]. If none of the syntax numbers is specified, the default value of zero is assumed. When the spectral recognizer learns the spectral features of a warning sound, this reference template will be assigned to the group of templates which have the same syntax number. Similarly, in the recognition stage, one or more syntax number(s) will be assigned to the unknown signal. To minimize useless comparisons, the spectral recognizer will use only the reference templates which have the same syntax number(s) as the unknown signal being examined.

In this work the syntax number is derived from the timing features of warning signals. From the timing analyzer the repetition period of the burst-type signal is obtained. Then, the syntax number of this warning signal is evaluated by dividing its repetition period by eight, in order to assure that the computed syntax number is bound within the allowable range. However, steady sounds have no repetition period.

Therefore, the syntax number of 110 is assigned arbitrarily to this group of signal templates. Furthermore, since telephone rings have by far the longest repetition period of all warning signals considered, any sound with a repetition period of about six seconds will be given the syntax number of the the telephone group (101).

5.3.2 Initialization Stage

In the initialization stage the parallel port (INTEL 8255) is reset and configured to mode 0 operation (i.e. port A = bidirectional port, port B is set to output port for this implementation, four pins of the port C are for handshaking signals and two other pins are for output control signals). Then, the three processors of the spectral recognizer are also reset, and the pattern memory is tested. If any I/O hardware interfacing problem occurs during the memory testing process, a failure status from the μPD7762 will be reported to the control & timing processor. Consequently, the 1 kHz tone is fed to the MC4760 for signal level adjustment. After level adjustment, the experimentally determined distance threshold is set to constrain the distance calculations between an unknown signal and the reference patterns. Then the user is prompted for any prestored template(s) to be transferred from permanent storage to the active pattern memory.

Table 5.4 shows the parameters used in the timing analysis and their initial values.

Table 5.4: Parameters used for the Timing Analyzer

Timing Analysis Parameter	Designated Values
Minimum burst duration	102.4 msec
Maximum burst duration	4000.0 msec
Minimum detectable transition level	2
Starting DAT	255
Duration to average the absolute signal amplitudes	12.8 msec

5.3.3 Training Stage

In the training stage we employ the "training-by-recognition" strategy to learn the characteristics of warning sounds. In brief, this strategy is achieved by three steps: 1) learning the timing features of warning sounds, 2) extracting their spectral features, and 3) verifying the learned spectral features. First, the timing information of warning sounds is provided by the timing feature extraction program (cf. Section 4.4). With this information, warning sounds are classified into two groups: steady and burst-type sounds.

Following the timing analysis, the spectral recognizer will learn the spectral patterns of these sounds. For steady sounds, the spectral recognizer immediately learns the spectral features and subsequently stores them in the pattern memory under syntax number 110.

For burst-type sounds, spectral extraction process must be synchronized with the rising transition of the burst. As shown in Fig. 5.42, if the spectral recognizer idling time is known, this synchronization can be accomplished by activating the spectral recognizer prior to the expected beginning of the burst. With the learned timing information (i.e., repetition period and average signal burst width) of a burst-type warning sound, the idling time is obtained by subtracting the average signal burst width from the repetition period. Consequently, the spectral patterns are stored in the pattern memory under the syntax number derived from the detected repetition period.

To verify the learned spectral patterns of warning sounds, the process described above is repeated. If the results of the two sets of recognition procedures are identical, the training procedure is completed. Otherwise, the training procedure repeats until the sound is "learned". If the spectral recognizer cannot successfully learn the spectral features of the signal, the user can interrupt the spectral recognizer, and restart the

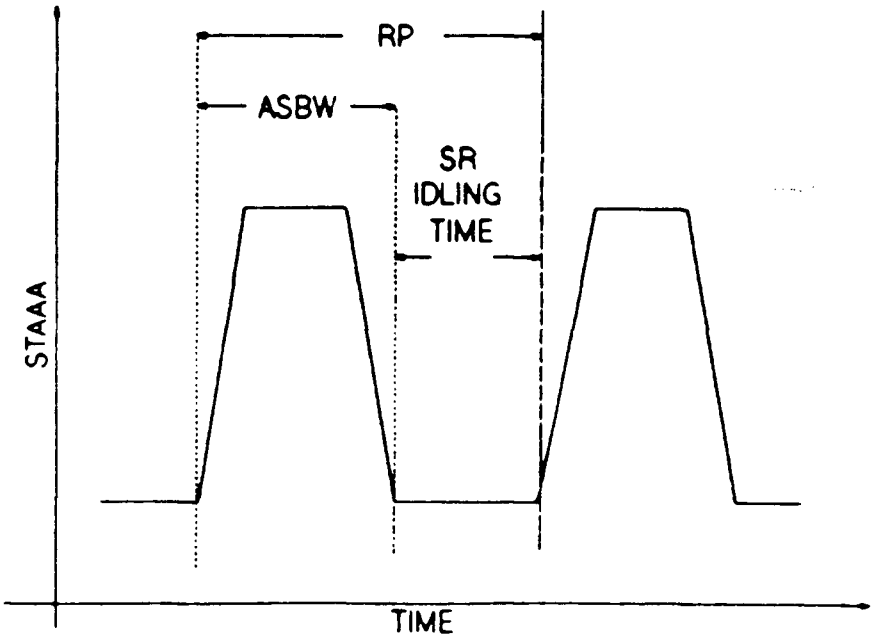


Figure 5.42: Timing relationships associated with the synchronization of the spectral recognizer to burst-type warning signals, where STAAA is the short-time average absolute amplitude of signal; RP is the repetition period; ASBW is the average signal burst width, and SR is the spectral recognizer

training procedure. Fig. 5.43, and Fig. 5.44 show the flowcharts of the training procedures for steady and, burst-type warning sounds, respectively.

Specific information relevant to each warning signal is stored for identification. This information includes the syntax number, the pattern registration number which is automatically generated for each warning sound, the signal type (steady or burst-type), and an identifier (name) of the warning sound assigned by the user during training.

5.3.4 Recognition Stage

Signal recognition consists of two stages: 1) warning signal detection by the timing analyzer, and 2) signal recognition by the spectral recognizer. The system continuously monitors the variations of the short-time average absolute amplitude of sound in the environments. If a steady sound is detected, the spectral recognizer identifies the sound twice. If the two recognition results identify the presence of a known warning sound, the unknown sound is declared to be that warning sound. If a potential burst-type sound is detected, its repetition period, burst width, and syntax number are derived. Based on these measurements, the spectral recognizer attempts to recognize the warning sound at the rising transition of the signal burst. If any spectral reference template can be matched to the unknown signal, the warning signal is identified with the known warning sound associated with that template. A flowchart of this recognition scheme is given in Fig. 5.45.

Upon completion of the recognition process, a summary of signal timing analysis and recognition results is displayed on the screen. These results include the syntax number, the signal type, the sound identifier, and the distance score from the matching calculations.

A system operating manual has been written for users (Appendix C).

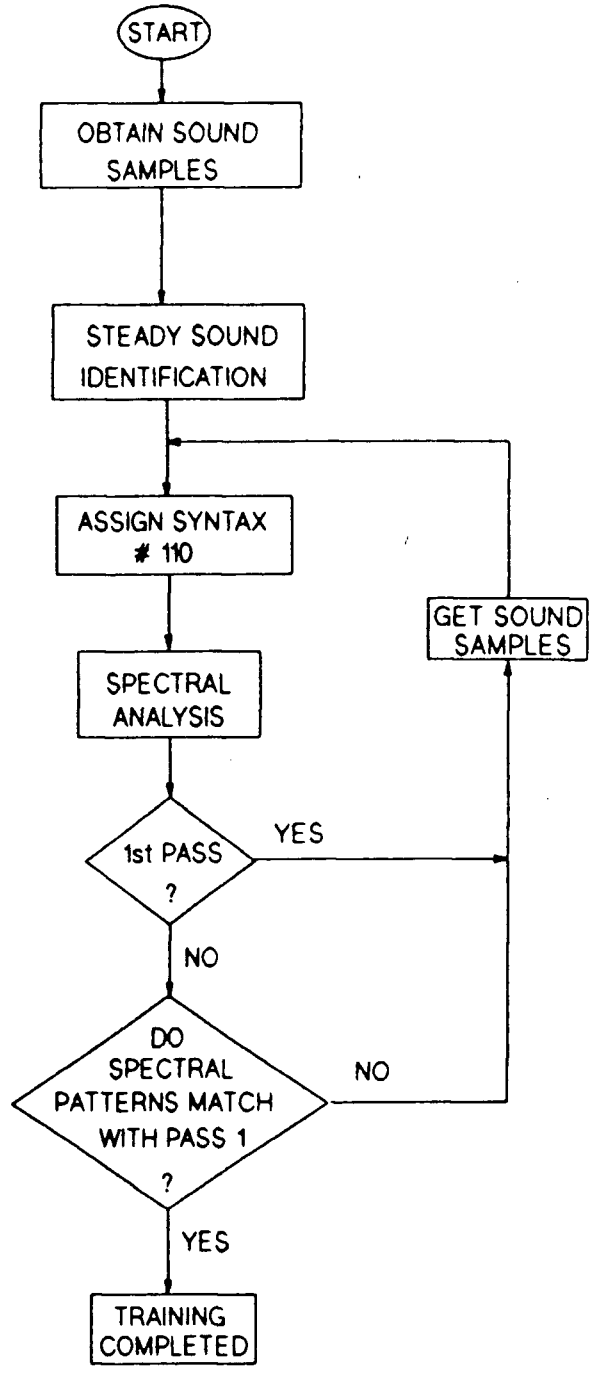


Figure 5.43: Flowchart of the training scheme for steady sounds

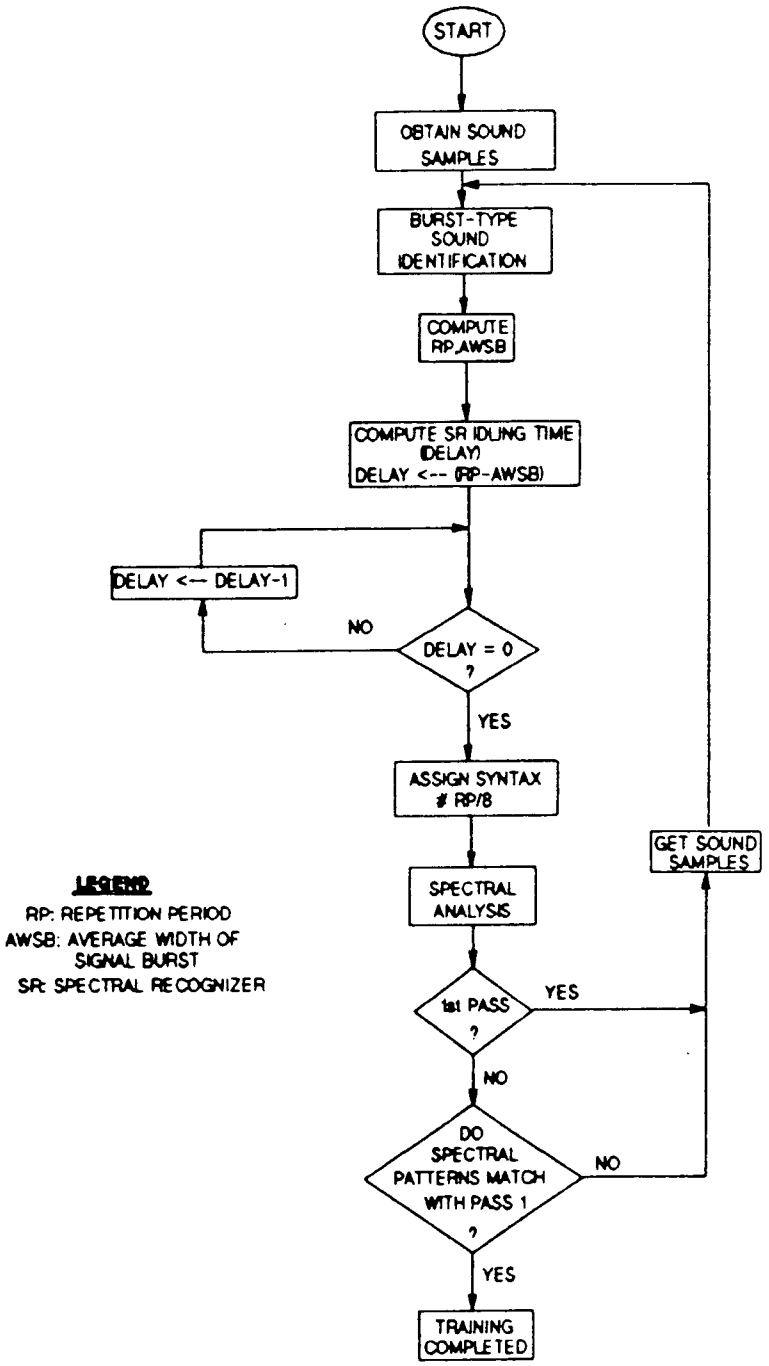


Figure 5.44: Flowchart of training procedures for burst-type warning sounds

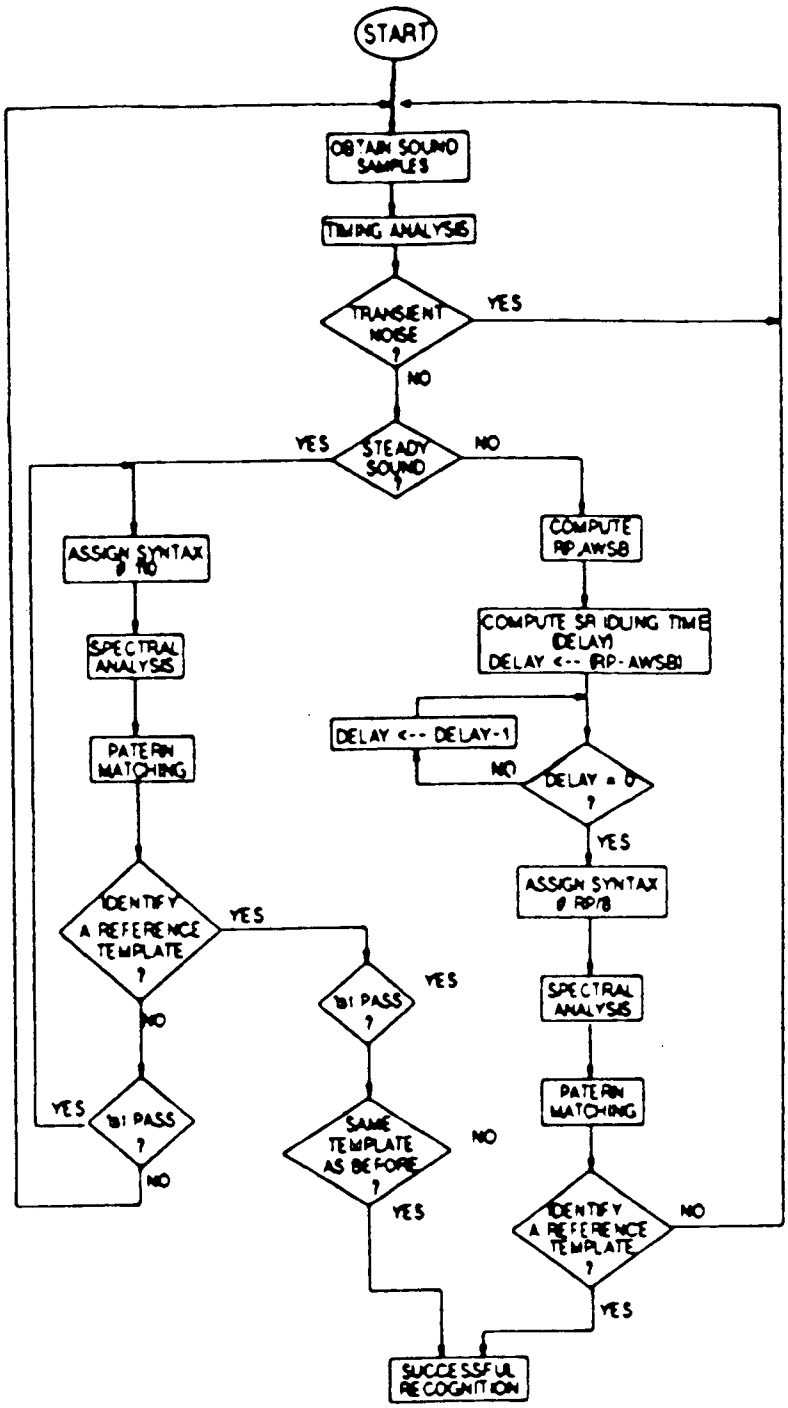


Figure 5.45: Flowchart of the recognition procedure

Chapter 6

Evaluation

Experiments were conducted to evaluate the performance of the WARNSIS under different noisy situations. Performance criteria were the average recognition rate and the false-alarm rate. Three noise backgrounds were used: 1) steady fan noise, 2) fan noise plus FM-radio broadcasts, and 3) fan noise plus AM-radio broadcasts. In view of the variations of spectra with loudness and noise contamination (cf. Section 3.2.5), three templates were prepared for the spectral recognizer at different SNRs (i.e. 10dB, 20 dB, and 30 dB) with the steady fan as a noise source.

Peterson demonstrated that in order to hear sounds reliably in the presence of noise, their spectral components have to be 15 dB to 25 dB above the background SPL level [17,18]. Furthermore, current standards demand the audible warning devices used in private residences must produce a minimum 10 dBA SPL above the average ambient level [11]. Therefore, we took the stricter criteria which was to maintain the average SPL of the noisy background at a minimum of 10 dBC below the SPL of the warning sounds.

Throughout the experiments, a value of 62 dBC SPL was measured for steady noise. When radio-broadcast was introduced into the steady noise background, the variations in SPL of the environment was monitored for five minutes in order to provide the average SPL estimate of the noisy background. This estimate was obtained by averaging the SPL variations within the observed time interval. More specifically, this value was maintained approximately at 65 dBC. Note that the three dBC SPL increase

was caused by acoustically adding two signals of equal strength (i.e. steady noise and radio-broadcast signal). Then, we activated an auditory warning device, and adjusted the loudness of the emitted sound so that the SPL reading was on the average 10 dBC above the noisy background.

The set-up for these experiments was similar to the one used for the measurement of the average short-time absolute amplitude of warning sounds in Chapter 3. Siren sounds were emitted from a siren horn; the pre-recorded telephone rings and smoke alarm sounds were produced by a tape recorder; and the radio-broadcasts originated from a radio-cassette player.

To explore the contribution of the timing and spectral recognizer parts to the performance of the WANRSIS, we also evaluated the recognition rate and the false-alarm rate using these subsystems separately. Specifically, for the timing analyzer part alone, the repetition period was our prime feature for warning sound recognition. Since steady sounds have no repetition period, their recognition accuracy rate cannot be found under these circumstances. In the training stage, the timing analyzer learned the repetition periods from the warning sounds. To recognize a warning sound, the repetition period of an unknown sound was extracted and compared to the values of the pre-stored repetition periods. If the absolute difference was less than 10 % of the pre-stored repetition period used in the comparison, the unknown sound was assigned to the corresponding reference warning sound.

For the spectral recognizer part alone, the environmental sounds were continuously monitored. Under the steady noise background, the spectral recognizer learned the signal templates using the 'training-by-recognition' scheme. For signal recognition, only spectral features were used without utilization of any timing information.

6.1 Average Recognition Accuracies

For each warning sound the recognition rate was derived by dividing the number of times the correct sound was identified by the total number of times the sound was present. The average accuracy for each of the three types of warning sounds is the average of the recognition rates calculated for all sounds belonging to the type. The detailed calculations may be found in Appendix D.

Table 6.5 shows the summary of recognition results for the complete WARNSIS, the timing analyzer part alone, and the spectral recognizer part alone. The first column gives the three types of noisy backgrounds in which the experiments were conducted; the second column shows the types of warning sounds used: 1) 'burst', denoting burst-type sounds, 2) 'steady', denoting steady sounds, and 3) 'phone', denoting telephone rings; the third, fourth, and fifth columns give the average recognition accuracies (ARA) achieved by the complete WARNSIS, the timing analyzer part alone, and the spectral recognizer part alone, respectively. The recognition results for the spectral recognizer part alone in a steady noise background were reported in [71].

In steady noise background, the complete WARNSIS produced 100 % average recognition accuracy for all three types of warning sounds. The timing analyzer part alone yielded perfect recognition scores for burst-type sounds and phone rings; and the spectral recognizer part alone gave more than 95 % average recognition accuracy in all cases. As mentioned previously, the timing analyzer can detect the presence of steady sounds, but cannot distinguish any particular steady sound. Therefore, we cannot find the average recognition accuracy for the steady sound in the column for the timing analyzer part alone.

With the addition of FM broadcast to the steady noise, the complete WARNSIS could still reliably identify burst-type, and steady sounds. As shown in Table 6.5,

Table 6.5: A summary of recognition results with MBD set to 0.1024 sec

Background Noises	Type of Warning Sound	Complete WARNSIS $ARA^{\dagger}$ (%)	Timing Analyzer Alone ARA (%)	Spectral Recognizer Alone ARA (%)
Steady Noise	Burst	100.0	100.0	100.0
	Steady	100.0	N/A	97.6
	Phone	100.0	100.0	95.8
FM + Steady Noise	Burst	98.0	97.7	65.6
	Steady	100.0	N/A	91.1
	Phone	0.0	0.0	70.0
AM + Steady Noise	Burst	99.3	98.3	67.2
	Steady	100.0	N/A	91.1
	Phone	0.0	0.0	69.2

$ARA^{\dagger}$: Average Recognition Accuracy in %
Minimum Burst Duration (MBD) : 0.1024 sec
N/A : Not Applicable

the recognition accuracies were measured as 98.0 % for burst-type sounds, and 100 % for steady sounds. But, the complete WARNSIS failed to recognize the telephone rings. Under the same noisy conditions the timing analyzer could recognize burst-type sounds with a 97.7 % average recognition accuracy, but failed to detect the presence of telephone rings. For the spectral recognizer part alone the average accuracy dropped from 100 % to 65.6 % for burst-type sounds, and was reduced from 95.8 % to 70 % for phone rings. However, this subsystem could still achieve a 91.1 % average recognition accuracy for steady sounds.

These results indicate that the complete WARNSIS consistently obtains higher recognition accuracy rates for burst-type and steady sounds than those of its subsystems separately. In close examination the complete WARNSIS gives a 0.3 % recognition accuracy better than that of the timing analyzer part for burst-type sounds with the background of FM broadcast plus steady fan noise. In the same situations, the complete WARNSIS outperforms the spectral recognizer by 24.4 % in identifying burst-type sounds, and by 8.9 % in correctly recognizing different steady sounds.

Similar results were also obtained when AM-radio broadcast and steady noise was used as background.

With the background of radio broadcast, both the complete WARNSIS and the timing analyzer failed to detect the presence of phone rings. Analysis showed that this is due to the value of the minimum burst duration (MBD) selected. It is possible to set MBD to provide greatly improved phone ring recognition (1.024 sec). Table 6.6 gives the recognition results with this MBD value.

Over 92 % recognition accuracy for phone rings is achieved by the complete WARNSIS, and the timing analyzer can always correctly identify the presence of phone rings in radio-broadcast backgrounds. According to the timing analysis algorithm, the modification of the minimum burst duration has no effect on the performance of the complete

Table 6.6: A summary of recognition results with MBD set to 1.024 sec

Background Noises	Type of Warning Sound	Complete WARNSIS $ARA^{\dagger}$ (%)	Timing Analyzer Alone ARA (%)	Spectral Recognizer Alone ARA (%)
Steady Noise	Burst	0	0	100.0
	Steady	100.0	N/A	97.6
	Phone	100.0	100.0	95.8
FM + Steady Noise	Burst	0	0	65.6
	Steady	100.0	N/A	91.1
	Phone	92.5	100.0	70.0
AM + Steady Noise	Burst	0	0	67.2
	Steady	100.0	N/A	91.1
	Phone	94.2	100.0	69.2

$ARA^{\dagger}$: Average Recognition Accuracy in %
MBD : 1.024 sec
N/A : Not Applicable

WARNSIS in steady sound recognition, and of the spectral recognizer alone in all noise situations. Therefore, we reproduced those average recognition accuracies from Table 6.5 in Table 6.6.

The effect of different MBD's on the performance of the WARNSIS is discussed in detail in Section 6.3.1.

6.2 False-alarm Rates

Since the occurrence of warning sounds in real-life environments is quite infrequent, it is essential for the WARNSIS not only to achieve an acceptable recognition accuracy for various sounds, but also to operate with a low false-alarm rate.

With the same experimental set-up as used before, we recorded the number of false-alarms over long period of time. The false-alarm rates for the complete WARNSIS, the timing analyzer part alone, and the spectral recognizer part alone were determined.

Table 6.7 shows that in steady noise situations WARNSIS produces no false-alarms. With radio-broadcast background the false alarm rate maybe as high as 2.33 per hour. Interestingly, phone ring false alarms are never produced.

For the timing analyzer alone the 'worst' false-alarm rate is 144.59 mis-recognitions per hour, 113 of which belongs to burst-type, 31 to steady, and 0.59 to phone ring sounds, respectively. In the two radio-broadcast backgrounds, over 99 % of mis-recognitions are classified into burst-type and steady sounds.

For the spectral recognizer alone, the 'worst' false-alarm rate is 1848 mis-recognitions per hour, 21 of which belongs to burst-type, 200 to steady, and 1627 to phone ring sounds, respectively. In two noisy conditions, over 80 % of mis-recognitions are classified into phone rings.

With the MBD set to 1.024 sec, the WARNSIS gave no false phone indications no matter what the noise conditions were (Table 6.8). Since the different MBD's have no effect on the performance of the spectral recognizer, the false-alarm rates for the spectral recognizer in Table 6.7 are reproduced in Table 6.8.

Although it is very difficult to quantify, experience has shown that the false alarm rate is highly dependent on the type of music played.

6.3 Discussion

6.3.1 Average Recognition Accuracies

Table 6.5 shows that the combined use of timing and spectral characteristics of warning sounds gives better recognition scheme for burst-type and steady sounds than any

Table 6.7: Results of the false-alarm test with MBD set to 0.1024

Background Noises	Mis-recognized As	Complete WARNSIS $FAR^{\dagger}$ (#/hour)	Timing Analyzer Alone FAR (#/hour)	Spectral Recognizer Alone FAR (#/hour)
Steady Noise	Burst	0	0	0
	Steady	0	0	0
	Phone	0	0	0
	Total	0	0	0
FM + Steady Noise	Burst	1.33	49	21
	Steady	1.0	35	200
	Phone	0	0.76	1627
	Total	2.33	84.76	1848
AM + Steady Noise	Burst	0.5	113	153
	Steady	0.5	31	296
	Phone	0	0.59	1270
	Total	1.0	144.59	1719

$FAR^{\dagger}$: False-alarm Rate
MBD : 0.1024 sec

Table 6.8: Results of false-alarm test with MBD set to 1.024 sec

Background Noises	Mis-recognized As	Complete WARNSIS $FAR^{\dagger}$ (#/hour)	Timing Analyzer Alone FAR (#/hour)	Spectral Recognizer Alone FAR (#/hour)
Steady Noise	Burst	0	0	0
	Steady	0	0	0
	Phone	0	0	0
	Total	0	0	0
FM + Steady Noise	Burst	0	2.67	21
	Steady	1.0	36	200
	Phone	0	4	1627
	Total	1.0	42.67	1848
AM + Steady Noise	Burst	0	4.67	153
	Steady	0.5	26	296
	Phone	0	9.33	1270
	Total	0.5	40.0	1719

$FAR^{\dagger}$: False-alarm Rate
MBD : 1.024 sec

scheme using only one of them. In particular, for these two types of warning sounds in radio broadcast backgrounds, the complete WARNSIS gives at least 0.3 % better average recognition accuracy than that of the timing analyzer alone, and provides minimally 8 % better average recognition accuracy rate than that of the spectral recognizer part alone.

The explanation for the failure of the complete WARNSIS and the timing analyzer to recognize phone ring is as follows. Fig. 6.46 gives an example of a phone ring sequence added with nonstationary background noise. The phone ring sequence is comprised of two 2 seconds bursts (B_1, and B_3), and of 4 seconds of silence. After the first phone ring, the burst, B_1, is detected by the timing analyzer, and the time markers for both rising and falling transitions are located. Without storing the detected burst waveform, the timing analyzer continues to monitor the environmental sounds. During the silence interval, B_2, which may be caused by radio music/conversation, is also detected by the timing analyzer. Unfortunately, the two criteria for a successful detection of a potential repetitive burst sequence are satisfied (i.e. $W_2 \geq MBD$, and the burst interarrival time $\geq MIAT$). Therefore, the repetition period for these bursts is calculated, and compared to the prestored template values. Mis-recognition to one of the warning sounds occurs, if this value matches to any one of the prestored values. Otherwise, the timing analyzer considers this burst sequence is caused by random noise, and their time markers are cleared as it restarts to search for another potential burst sequence. Similarly, the timing analyzer decides either mis-recognition or random noise rejection for the following phone bursts (i.e. B_3 in Fig. 6.46).

As a result, the timing analyzer fails to detect the presence of phone rings. If the timing analyzer cannot provide the timing information on phone ring sequence, the WARNSIS cannot utilize this timing analysis result, and eventually, it also cannot identify the presence of phone rings.

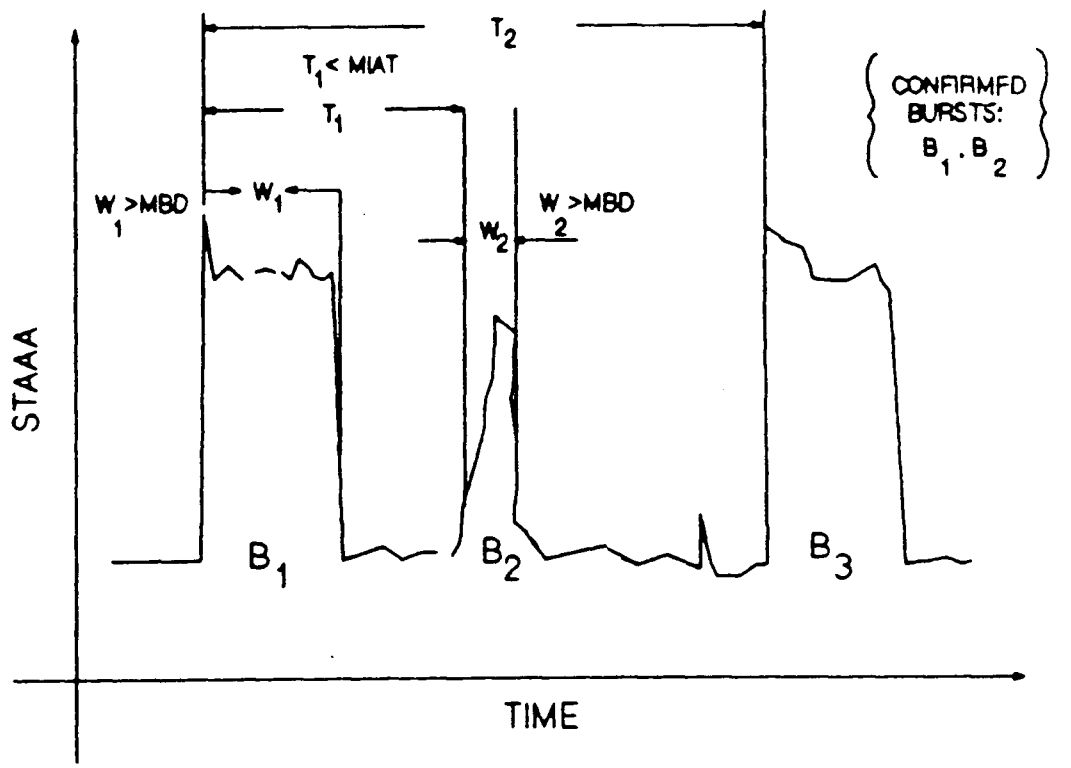


Figure 6.46: An example of a phone ring sequence added with nonstationary background noise

In Table 6.6 we find that the timing analyzer performs better than the complete WARNSIS in phone ring recognition. An explanation for this observation is as follows. For the timing analyzer part alone the repetition period is the only feature used to to detect the presence of phone rings. Since the repetition periods of the phone ring sequences used are approximately six seconds, the timing analyzer, therefore, cannot identify the sounds emitted from a specific telephone ringer. However, based on the timing information derived from a phone ring sequence, the complete WARNSIS then examines the spectral content of a phone ring and compares it to the pre-stored spectral patterns belonging to the group of telephone rings. Thus, the complete WARNSIS not only identifies the sound as a phone ring, but also provides additional information on the specific ringer. As deduced from Table D.25 and D.27 in Appendix D (a complete set of evaluation results), the decreased recognition rate occurs even though it identifies the correct ringer, as it chooses the incorrect loudness or pitch template. This is because of the similar spectral characteristics between templates with adjacent settings (cf. Section 3.2). In a practical system, however, this would not matter as long as the "phone is ringing" event is detected.

The repetition period of burst-type sounds ranges from 140 msec to 3.2 sec. With the value of the minimum burst duration changed from 0.1024 sec to 1.024 sec, the timing analyzer is prevented from extracting timing features of those burst-type sounds with repetition periods less than 1.024 sec. However, the modification has no effect on the steady sound recognition performance of the complete WARNSIS because steady sounds require a minimum burst duration of four seconds.

6.3.2 False-alarm Rates

The results of the false-alarm rate indicate that the combined use of timing and spectral features to characterize warning sounds provides an effective scheme to eliminate false

recognitions triggered by environmental noise. For random noise there are no false alarms. In the presence of FM broadcasts, the complete WARNSIS gives a false-alarm rate of about 2.33 false recognitions per hour, which we consider to be unacceptably high for a practical recognition system operating in real-life environments. It should be remembered, however, that the measurements presented here represent the 'worst-case' false-alarm recognition performance of the WARNSIS. Real life performance should be better, since SNR's are usually higher than the 10 dB used in our measurements. Evaluation of performance in use will require field testing beyond the scope of this work.

The specifications for the WARNSIS are given in Appendix E.

Chapter 7

Conclusions and Recommendations

7.1 Summary & Conclusions

This work was divided into two major parts: 1) the analysis of warning sounds, and 2) the design of a prototype recognition device based on (1). An extensive search for existing warning sound characteristics yielded only a limited amount of timing and spectral information. Therefore, we used various timing and spectral analysis techniques to study the warning sounds emitted by telephones, smoke alarms, and electronic siren drivers.

First, the short-time average absolute amplitudes of warning sounds were analyzed to provide timing features. Results show that warning sounds can be categorized into either steady or burst-type sounds.

Secondly, Welch's nonoverlapping spectral estimation method was used to analyze the short-time spectra of warning sounds. Our findings indicate that the spectra of telephone rings produced from electromechanical ringers of dial phones of the same model may vary significantly. These spectral characteristics also depend on the setting of the loudness adjustments provided. Typically, the short-time spectra of a two second telephone ring consist of two discernible parts: the transient region and the steady-state regions. Analyses were also performed on telephone rings emitted from an electronic ringer. Results indicate that by varying the pitch setting, the two tones generated from

the ringer change accordingly. For siren sounds, the short-time spectra can be divided into two groups: 1) spectra with rich harmonics and, 2) spectra with frequency clusters.

Based on the timing and spectral analysis results, a 'hybrid' prototype recognition device (WARNSIS) was developed and constructed using commercially available components. This device utilizes a combination of timing and spectral features of warning sounds as signal patterns. A 'real-time' algorithm is used to extract timing features in noisy environments. According to the relative timing characteristics of these features, warning sounds are classified. Then, the incoming signals are passed on for spectral analysis.

A filter-bank approach is employed to analyse the short-time spectra of warning sounds. To categorize these spectral patterns, the timing information of warning sounds is used to group these patterns with sounds of similar timing features. This grouping technique greatly reduces the amount of computation involved in the recognition stage.

The real-time program to extract timing features was written in assembler language. The spectral recognizer was constructed with commercial electronic components. A software operating system was developed to co-ordinate the timing analyzer and the spectral recognizer. Our device consists of 79 chips, and the software program is comprised of 2490 lines of assembler source codes.

Experiments were conducted to investigate the performance of the WARNSIS in noisy environments. For burst-type and steady sounds, the WARNSIS provides average recognition accuracies over 98 %. With regard to the false-alarm rates, the complete WARNSIS gives much lower values than the false-alarm rates of its separate timing and spectral subsystems.

In this work, we designed, constructed, and evaluated a warning sound recognition system. The evaluation results indicate that the WARNSIS operates satisfactorily in real environments, where it can be taught to learn new sounds and to recognize them.

This system will reliably recognize warning sounds in random noise with no false alarms. In very loud music and conversation the recognition is still good, although more false alarms are created. Considering that our evaluation criteria have been very stringent, the performance of the system in real-life situations is expected to be satisfactory.

7.2 Recommendations for Future Directions of Research

To improve the performance of the complete WARNSIS in noisy environments with SNR of lower than 10 dB, future work should be directed towards the following:

1. The improvement of the transition or break-point detection scheme and implementation: In the present design none of the short-time average amplitudes are stored for analysis. It is feasible to store these amplitude values, and then use a fast CPU to analyze the stored signal amplitude samples. Faster CPU than the one presently employed will permit more elaborate analysis of these amplitude samples, so that the timing analyzer becomes more intelligent in rejecting unwanted transient noises. A possible extension of this work is to use the shape of the amplitude contours of burst-type sounds to provide additional signal features.

2. Exploration of the adaptive noise cancellation (ANC) technique: Since noise in this work consists of music, speech signals and transient noises, cancellation of these noises in real-life environments leads us into unexplored territory. Then we need to find a suitable ANC algorithm and explore its implementation for optimum performance. For real-time operation, a compromise may exist between the SNR improvement and the complexity of the algorithm.

3. Use of microphone array to provide better spatial separation between warning sound source and background noise: A microphone array can provide a much sharper directional beam to obtain better quality warning sound than a single directional microphone. Research in this area should involve the selection the microphone array structure, its orientation, and a signal processing algorithm to analyze the outputs from the microphone array to yield the desired output.

   A possible extension is the combined use of adaptive noise cancellation and multi-microphone array system for sound tracking capability and noise removal enhancement. Research in this area will require a multiple digital signal processing (DSP) system to facilitate the real-time operation in nonstationary noise environments.

References

[1] J. E. Harkins and C. J. Jensema, *Focus-group discussions with deaf and severely hard of hearing people on needs for sensory devices*, Gallaudet Research Institute, Technology Assessment Program, Washington D. C., 1987.

[2] J. Hurvitz and R. Carmen, *Special Devices for Hard of Hearing, Deaf, and Deaf-Blind Persons*, Little, Brown and Company, Boston, 1981.

[3] T. Hustak, *Directory of Technical Aids Available to Hearing Impaired Persons*, Services for Hearing Impaired Persons, Inc., Regina, Sakatchewan 1984.

[4] J. E. Harkin and C. J. Jensema and H. Ryland, "Toward Emergency Vehicle Detection: Systemic Considerations", Proceedings of ICARRT at Montreal, pp.228-229, 1988.

[5] Underwriters Laboratories Inc. Standard for Safety UL217 : "Single and Multiple Station Smoke Detectors", Oct., 1985.

[6] Underwriters Laboratories Inc. Standard for Safety UL985 : "Household Fire Warning System Units", June, 1985.

[7] Underwriters Laboratories Inc. Standard for Safety UL904 : "Vehicle Alarm Systems and Units", July, 1982.

[8] Canadian Standards Association, National Standard of Canada, CAN/CSA-T510-M87, "Performance and Compatibility Requirements for Telephone Sets", March, 1987.

[9] Electronic Industries Association, EIA-470-A, "Telephone Instruments with Loop Signalling for Voiceband Applications", 1988.

[10] Bell System Voice Communications Technical Reference, PUB 48005, "Functional Product Class Criteria : Telephones", Jan., 1980

[11] National Fire Protection Association, NFPA 72G, "Guide for the Installation, Maintenance and Use of Notification Appliances for Protective Signalling Systems", 1985.

[12] National Fire Protection Association, NFPA 72A, "Standard for Installation, Maintenance and Use of Local Protective Signalling Systems for Guards's Tour, Fire Alarm and Supervisory Service", 1985.

[13] R. E. Halliwell and M. A. Sultan, "Attenuation of Smoke Detector Alarm Signals in Residental Buildings", National Research Council Canada, Institute for Research in Construction, NRCC 25897.

[14] S. Miyaaki and A. Ishida, "Traffic-alarm Sound Monitor for Aurally Handicapped Drivers", J. of Medical & Computer, Vol.25, pp.68-74, Jan., 1987.

[15] Installation and Service Instructions for Model MCS-1 Motor Signal, Federal Signal Corporation.

[16] Installation Manual for Electronic Siren, Model SA 400-63, Southern Vehicle Products, Inc.

[17] R. D. Patterson, CAA Paper 82017, Civil Aviation Authority, London, U.K., 1982.

[18] J. Edworthy and R. D. Patterson, "Ergonomic Factors in Auditory Warnings", Ergonomics International 85, edited by I. D. Brown, R. Goldsmith, K. Coombes and M. A. Sinclair, pp.232-235, 1985.

[19] Lower and Wheeler, "Design of Auditory Warnings for Aircraft, Industry and Hospitals", Ergonomics International 85, edited by I. D. Brown, R. Goldsmith, K. Coombes and M. A. Sinclair, pp.226-228, 1985.

[20] G. M. Rood, J. A. Chillery and J. B. Collister, "Requirements and Application of Auditory Warnings to Military Helicopters", Ergonomics International 85, edited by I. D. Brown, R. Goldsmith, K. Coombes and M. A. Sinclair, pp.169-170, 1985.

[21] M. J. Shailer and R.D. Patterson, "Pulse generation for Auditory Warning Systems", Ergonomics International 85, edited by I. D. Brown, R. Goldsmith, K. Coombes and M. A. Sinclair, pp.229-231, 1985.

[22] J.H. Kerr, "Warning Devices", Br. J. Anaesth., **57**, pp.696-708, 1985.

[23] R. D. Patterson, J. Edworthy and M. J. Shailer, "Alarm sounds for Medical Equipment in Intensive Care Areas and Operation Theatres", Institute of Sound and Vibration Research Paper AC598, 1986.

[24] S. M. Kay and S. L. Marple ,Jr., "Spectral Analysis: A Modern Perspective", Proceedings of IEEE, Vol.69, No.11, pp.1380-1419, Nov., 1981.

[25] B. S. Atal and M. R. Schroeder, "Linear Prediction Analysis of Speech based on a Pole-zero Representation", Journal of Acoust. Soc. of Amer., Vol.64, No.5, pp.1310-1318, Nov., 1978.

[26] J. Makoul, "Linear Prediction: A tutorial Review", Proceeding of IEEE, Vol.63, pp.561-580, Apr., 1975.

[27] R. B. Blackman and J. W. Tukey, "The Measurement of Power Spectra from the point of view of Communication Engineering", New York, Dover, 1959.

[28] P. D. Welch, "The Use of fast Fourier transform for the estimation of Power Spectra: A method based on Time Averaging over Short, Modified Periodograms", IEEE Trans. on Audio Electroacoust., Vol.AU-15, pp.70-73, June, 1967.

[29] G. C. Carter and A. H. Nuttall, "On the Weighted Overlapped Segment Averaging Method for Power Spectral Estimation", Proc. of the IEEE, Vol.68, No.10, pp.1352-1353, Oct., 1980.

[30] J. S. Lim, "All Pole Modelling of Degraded Speech", IEEE Trans. on ASSP, Vol.ASSP-26, pp.197-209, June, 1978.

[31] S. M. Kay, "The Effects of Noise on the Autoregressive Spectral Estimator", IEEE Trans. on ASSP, Vol.ASSP-27, pp.478-485, Oct., 1979.

[32] F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier transform", Proceedings of IEEE, Vol.66, No.1, pp.51-83, Jan., 1978.

[33] D. N. Romalo, "An Interference Monitor for a Radio Observatory", M.A.Sc. Thesis, Dept. of Electrical Engineering, University of British Columbia, pp.42-44, April, 1988.

[34] Simon Chau and Charles Laszlo, "Spectra of Telephone Rings and Annunciating Signals used in an Aid for Hearing Impaired", Proceedings of the 13^{th} CMBEC, pp.147-148, Halifax, June, 1987.

[35] B. S. Atal and L. R. Rabiner, "Speech Research Directions", AT&T Technical Journal, Vol.62, No.5, Sept/Oct., pp.75-88, 1986.

[36] S. E. Levinson, "Structural Methods in Automatic Speech Recognition", Proceedings of IEEE, Vol.73, No.11, Nov., pp.1625-1650, 1985.

[37] L. R. Rabiner and S. E. Levinson, "Isolated and Connected Word Recognition - Theory and Selected Applications", IEEE Trans. on Communications, Vol.COM-29, No.5, pp.621-659, May, 1981.

[38] D. O'Shaughnessy, "Speech Recognition", IEEE ASSP Magazine, pp.4-17, Oct., 1986.

[39] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. on ASSP, Vol.ASSP-26, No.1, pp.43-49, Feb., 1978.

[40] A. H. Gray, Jr. and J. D. Markel, "Distance Measures for Speech Processing", IEEE Trans. on ASSP, Vol.ASSP-24, No.5, pp.380-391, Oct., 1976.

[41] N. Nocerino, F. K. Soony, L. R. Rabiner and D. H. Klatt, "Comparative study of Several Distortion Measures for Speech Recognition", Proc. ICASSP, pp.25-28, 1985.

[42] H. Matsumoto and H. Iami, "Comparative Study of Variable Spectrum Matching Measures on Noise Robustness", Proc. ICASSP, pp.769-772, 1986.

[43] R. F. Purton, "Speech Recognition Using Autocorrelation Analysis", IEEE Trans. on Audio and Electroacoustics, Vol.AU-16, No.2, pp.235-239, June, 1968.

[44] M. M. Sondhi, "New Methods for Pitch Detection", IEEE Trans. on Audio-Electro., Vol.AU-16, pp.262-266, June, 1968.

[45] L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection", IEEE Trans. on ASSP, Vol.ASSP-25, No.1, pp.24-33, Feb., 1977.

[46] J. J. Dubnowski, R. W. Schafer and L. R. Rabiner, "Real-time digital Hardware pitch detector", IEEE Trans on ASSP, Vol.ASSP-24, pp.2-8, Feb., 1976.

[47] L. R. Rabiner and M. R. Sambur, "An Algorithm for determining the Endpoints of Isolated Utterances", The Bell System Technical Journal, Vol.54, Vol.2, pp.297-315, Feb., 1975.

[48] M. T. Whitaker and J. A. S. Angus, "A Low Cost Continuous Word Speech Recognizer", International Conf. on Speech Input/Output Techniques and Applications, IEE Conf. Publication # 258, pp.119-123, March, 1986.

[49] L. F. Lamel, L. R. Rabiner, A. E. Rosenberg and J. G. Wilpon, "An Improved Endpoint Detector for Isolated Word Recognition", IEEE Trans. on ASSP, Vol.ASSP-29, No.4, August, pp.777-785, 1981.

[50] J. G. Ackenhusen and L. R. Rabiner, "Microprocessor implementation of an LPC-based isolated word recognizer", in Proc. 1980 BTL/WE Microprocessor Symp., Sept., pp.35-42, 1980.

[51] B. Gold and L. R. Rabiner, "Parallel Processing Technique for Estimation Pitch Periods of Speech in the Time Domain", J. Acoust. Soc. Amer., Vol.46, pp.442-448, Aug., 1969.

[52] B. Gold, "Note on buzz-hiss detection", J. Acoust. Soc. Amer., Vol.36, pp.1659-1661, 1964.

[53] G. M. White and R. B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering and Dynamic Programming", IEEE Trans. on ASSP, Vol.ASSP-24, No.2, pp.183-188, April, 1976.

[54] H. L. Kwok, L. C. Tai, and Y. M. Fung, "Machine Recognition of the Cantonese Digits Using Bandpass Filters", IEEE Trans. on ASSP, Vol.ASSP-31, No.1, pp.220-222, Feb., 1983.

[55] NEC Speech Recognition LSI Set Manual, June, 1985.

[56] D. Tjostheim, "Recognition of Waveforms Using Autoregressive Feature Extraction", IEEE Trans. on Computer, Vol.C-26, No.3, pp.268-270, March, 1977.

[57] B. S. Atal and M. R. Schroeder, "Adaptive Predictive Coding of Speech Signals", Bell System Tech. Journal, Vol.49, pp.1973-1986, 1971.

[58] J. G. Ackenhusen and Y. H. Oh, "Single-chip Implementation of Feature Measurement for LPC-based Speech Recognition", AT&T Technical Journal, Vol.64, No.8, pp.1787-1805, Oct., 1985.

[59] B. A. Dautrich, L. R. Rabiner and T. B. Martin, "On the Effects of Varying Filter Bank Parameters in Isolated Word Recognition", IEEE Trans. on ASSP, Vol.ASSP-31, No.4, pp.793-806, August, 1983.

[60] J. S. Lim, "Estimation of LPC coefficients from speech waveforms degraded by additive random noise", Proc ICASSP 78, pp.599-601.

[61] J. Tierney, "A Study of LPC Analysis of Speech in Additive Noise", IEEE Trans. on ASSP, Vol.ASSP-28, No.4, pp.389-397, August, 1980.

[62] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification", J. Acoust. Soc. Am., Vol.55, No.6, pp.1304-1312, June, 1974.

[63] B. H. Juang, L. R. Rabiner and J. G. Wilpon, "On the use of Bandpass Liftering in Speech Recognition", IEEE Trans. on ASSP, Vol.ASSP-35, No.7, pp.947-953, July, 1987.

[64] B. A. Hanson and H. Wakita, "Spectral Slope Distance Measures with Linear Prediction Analysis for Word Recognition in Noise", IEEE Trans. on ASSP, Vol.ASSP-35, No.7, pp.968-973, July, 1987.

[65] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", The Bell System Technical Journal, Vol.62, pp.1035-1074, April, 1983.

[66] L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition", The Bell System Technical Journal, Vol.62, No.4, pp.1075-1105, April, 1983.

[67] A. Varga, R. Moore, J. Bridle, K. Ponting, and M. Russell, "Noise Compensation Algorithms for use with Hidden Markov Model based Speech Recognition", Proc. of IEEE Conf., pp.481-484, 1988.

[68] R. W. Schafer and L. R. Rabiner, "Digital Representation of Speech Signals", Proceedings of IEEE, Vol.63, No.4, pp.662-677, April,1975.

[69] L. R. Rabiner, A. E. Rosenberg and S. E. Levinson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition", IEEE Trans. on ASSP, Vol.ASSP-26, No.6, pp.575-582, Dec., 1978.

[70] C. Myers, L. R. Rabiner and A. E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition", IEEE Trans. on ASSP, Vol.ASSP-28, No.6, pp.623-635, Dec., 1980.

[71] Simon Chau and Charles Laszlo, "A Warning Signal Identification System (WARNSIS) for Hard of Hearing Individuals", Proceedings of the 14^{th} CMBEC, pp.145-146, Montreal, June, 1988.

Appendix A

Formulation of Relationship between SNR and SPL measurements

In this work SNR is defined as the ratio of the peak power of the signal to peak power of the background noise. To calculate the SNR directly, we need to obtain both signal and noise power. From the SPL measurement of acoustic background (noise), the noise power can be derived. We found, however, that the measurement of the SPL of the warning sound alone in any real acoustic environment is impossible, since there is always some background noise present. Here we will show the relationship of noise SPL, and the warning sound plus noise SPL to the SNR.

The following notation will be used:

I_{ref} = reference sound intensity

I_a = peak acoustic intensity of background noise

I_s = peak warning sound intensity

I_{a+s} = peak acoustic intensity of a warning sound plus background noise

P_a = peak SPL of background noise

P_{a+s} = peak SPL of a warning sound plus background noise

$SI_a = I_a$ in dB

$SI_s = I_s$ in dB

SI_{a+s} = I_{a+s} expressed in dB

SPL_a = SPL measurement of background noise

SPL_{a+s} = SPL measurement of a warning sound plus background noise

P_{ref} = the reference sound pressure level (20 μ Pa)

From the definition of SNR,

$$SNR = \frac{I_s}{I_a} \tag{A.35}$$

and

$$\widehat{SNR}(dB) = 10 \quad \log_{10}\left\{\frac{I_s}{I_a}\right\} \tag{A.36}$$

Also,

$$SI_{a+s} = 10 \quad \log_{10}\left\{\frac{I_{a+s}}{I_{ref}}\right\} \tag{A.37}$$

$$SI_a = 10 \quad \log_{10}\left\{\frac{I_a}{I_{ref}}\right\} \tag{A.38}$$

But $(SI_{a+s} - SI_a)$ = difference in sound intensity level in dB, and

using equations (A.37) and (A.38), it gives

$$\begin{aligned}(SI_{a+s} - SI_a) &= 10 \quad \log_{10}\left\{\frac{I_{a+s}}{I_{ref}}\right\} - 10 \quad \log_{10}\left\{\frac{I_a}{I_{ref}}\right\} \\ &= 10 \quad \log_{10}\left\{\frac{I_{a+s}}{I_a}\right\}\end{aligned} \tag{A.39}$$

Since $I_{a+s} = I_s + I_a$ (without resonance), we have

$$\begin{aligned}(SI_{a+s} - SI_a) &= 10 \quad \log_{10}\left\{\frac{(I_a + I_s)}{I_a}\right\} \\ &= 10 \quad \log_{10}\left\{1 + \frac{I_s}{I_a}\right\}\end{aligned} \tag{A.40}$$

To find $(SI_{a+s} - SI_a)$ by measurement, consider

$$\begin{aligned} I_a &= \kappa P_a^2 \\ I_{a+s} &= \kappa P_{a+s}^2 \end{aligned} \tag{A.41}$$

where $\kappa = constant$

Then we can express $(SI_{a+s} - SI_a)$ in terms of P_a and P_{a+s} which can be measured by a commercially available SPL meter.

$$\begin{aligned} (SI_{a+s} - SI_a) &= 10 \quad \log_{10}\left\{\frac{\kappa P_{a+s}^2}{\kappa P_a^2}\right\} \\ &= 20 \quad \log_{10}\left\{\frac{P_{a+s}}{P_a}\right\} \end{aligned} \tag{A.42}$$

Rewriting equation (A.42) using P_{ref} gives,

$$\begin{aligned} (SI_{a+s} - SI_a) &= 20 \quad \log_{10}\frac{P_{a+s}}{P_{ref}} - 20 \quad \log_{10}\frac{P_a}{P_{ref}} \\ &= SPL_{a+S} - SPL_a \\ &= 10 \quad \log_{10}(1 + \frac{I_s}{I_a}) \end{aligned} \tag{A.43}$$

Equation A.43 indicates that the difference in sound intensity can be expressed in terms of two measurable physical quantities — the difference in SPL measurements in the absence and during the presence of a warning sound. Hence, we have

$$\begin{aligned} (SPL_{a+s} - SPL_a) &= 10 \quad \log_{10}(1 + \frac{I_s}{I_a}) \\ &= 10 \quad \log_{10}(1 + SNR) \end{aligned} \tag{A.44}$$

Hence

$$\begin{aligned} SNR &= \frac{I_s}{I_a} \\ &= \left\{ anti\ \log_{10}\left(\frac{(SPL_{a+s} - SPL_a)}{10}\right)\right\} - 1 \end{aligned} \tag{A.45}$$

or

$$\widehat{SNR}(dB) = 10 \quad \log_{10}(SNR) \tag{A.46}$$

When the difference in SPL readings is more than 10 dB, SNR in dB is very close to the SPL difference in dB (Table A.9).

Table A.9: Tabulation of SPL reading difference and SNR

$(SPL_{a+s} - SPL_a) in (dB)$	$\widehat{SNR}$ (in dB)	SNR
0.5	-9.14	0.12
1.0	-5.9	0.26
1.5	-3.8	0.41
2.0	-2.43	0.59
2.5	-1.1	0.78
3.0	0.0	1.0
3.5	0.9	1.2
4.0	1.8	1.5
4.5	2.6	1.8
5.0	3.3	2.2
5.5	4.1	2.6
6.0	4.7	3.0
6.5	5.4	3.5
7.0	6.0	4.0
8.0	7.2	5.3
9.0	8.4	6.9
10.0	9.5	9.0
11.0	10.6	11.6
12.0	11.7	14.8
13.0	12.8	19.0
14.0	13.8	24.1
15.0	14.8	30.6
16.0	15.9	38.8
17.0	16.9	49.1
18.0	17.9	62.0
19.0	18.9	78.4
20.0	20.0	99.0
21.0	21.0	125
22.0	22.0	158
23.0	23.0	199

Appendix B

Format of the command set of the SR

The format of the twelve commands used to control the operation of the SR is given in Table B.10.

Correspondingly, Table B.11 shows the legal values for the memory bank, the bank rejected value, the signal rejected value, the syntax #, and the registration #.

In response to a specified command one or more of the following status output codes is(are) reported from the μPD7762 to the control & timing processor. The interpretations of these status output codes are given in Table B.12.

Table B.10: Format of command set of SR

Command Code	Format
1. Initialize (2 byte code)	$\underbrace{00}_{code}\ \underbrace{H}_{hex}$, $0FFH$ (termination code)
2. Level_adjust (3 – 6 bytes)	01H, [memory bank], [memory bank], [memory bank], [memory bank], 0FFH
3. Recognition (2 – 32 bytes)	003H, [syntax # (S)], [..., S ...], 0FFH
4. Training (3 – 5 bytes)	002H, registration #, [syntax #], [signal rejected value], 0FFH
5. Second Decision (2 bytes)	004H, 0FFH
6. Hot start (2 bytes)	005H, 0FFH
7. Down load (3 bytes)	006H, # of patterns, 0FFH
8. Up load (2 bytes)	007H, 0FFH
9. Change memory reject value (3 bytes)	008H, bank reject value, 0FFH
10. Memory test (2 bytes)	009H, 0FFH
11. Select memory bank (3 bytes)	00AH, bank #, 0FFH
12. Change signal reject value (3 bytes)	00CH, registration #, signal reject value, 0FFH

Table B.11: Legal Values for parameters of the command set

Parameters	Legal Value
1) Memory bank value (B)	$0 \leq B \leq 03$
2) Bank reject value (BRV)	$0 \leq BRV \leq$ 0FEH
3) signal reject value (SRV)	$0 \leq SRV \leq$ 080H
4) pattern registration value (PRV)	$0 \leq PRV \leq$ 080H

Table B.12: Interpretation of status output codes from μPD7762

Code	Interpretation
000H	normal completion of a command
001H	Input signal level too high
002H	Input signal level too low
003H	Input signal longer than 2.0 sec
004H	Request signal level adjustment
005H	Specified syntax # non-existing
006H	Registered pattern does not exist
007H	the distance value is greater than BRV
008H	Specified memory bank does not exist
009H	Command format error
00AH	The distance is greater than PRV, but less than BRV
00BH	Signal duration is less than 200 msec
00CH	Memory test error or hardware I/O error

Appendix C

Software Operating Manual of The WARNSIS

C.1 Program Files

This manual provides a guidance for the user to follow the operation procedure developed for the signal recognition software. The software was designed to provide an interactive dialogue between the user and the device. Messages will constantly display on the monitor to enquire the user to input the requested parameter values, and to indicate the status of the device. In this manual, such messages are shown in bold-face.

The software was saved on a PC-computer, and was located at the sub-directory called \simon\nec\. To enter this sub-directory, the user needs to type the following statements:

type : **cd simon**

displayed on the monitor: **d:\simon**

type : **cd nec**

displayed on the monitor : **d:\simon\nec**

Once the user has entered the sub-directory of \simon\nec\, he/she can find the programs necessary to run this software. These programs are :

- **nec.asm** : the source program of the system operating software in assembly language,
- **nec.exe** : the executable file of nec.asm,
- **nec_dat.asm** : the data file consisting of constants and variables for nec.exe and,
- **enec.bat** : the batch file used to automatically assemble **nec.asm** to produce its object codes, to link its object code file (nec.obj) to yield the executable file (nec.exe), and to delete the redundant object file to optimize memory storage on the hard-disk. This batch file is activated only when modification(s) has been made to the **nec.asm**. Execution of this batch job is accomplished by typing **ENEC**.

A signal template file was stored at the directory of \simon\nec\temp\. This data file is called as **50_warn.dat**, and consists of 50 templates of various warning sounds. Such warnings include siren sounds emitted from an electronic siren driver, telephone rings and smoke alarm sounds.

C.2 Interactive Operations

To execute the system software, the user types **NEC**. By executing the nec.exe, the user enters the interactive operation mode, and is prompted to answer a number of questions. There are two stages in this mode of operation, namely, the initiatization stage, and the training/recognition stage.

C.2.1 Initialization Stage

Once the program is executed, the following events occur. They are:

1. **System Initialization in Progress**

2. **System Hardware Checking:** if everything is OK, these statements are displayed on the monitor:

   - **MEMORY CHECK OK !!**
   - **MEMORY CHECK OK !!**

   Otherwise, error statements are reported, and they are :

   - **Invalid Command,** or
   - **MEMORY error or HARDWARE I/O error !**

   Under such circumstances, the user must exit the program by pressing CTRL-C, and shut off the power supply for 20 seconds, turn on the power supply, and re-run the program.

3. the user is prompted to flip a manual switch before the system begins the process of signal level adjustment.

- **Please, flip the switch to LEVEL_ADJUST,**
- **If ready, Please press ENTER key.**

After the ENTER key is pressed, the system starts the signal level adjustment.

- **Level adjustment in PROGRESS**

Upon completion of the level adjustment, the system requests if the user wants to transfer any pre-stored signal template(s) to the template memory of the device.

- **Do you want to download signal templates from host CPU? (y/n)**

If the answer is '**y**', then the user needs to provide the template file name and the value of the total # of the prestored templates.

- **Please, input the file name consisting of the templates – *.dat. (d:\simon\nec\temp*.dat)** , and a file opening statement is shown on the monitor.
- **SUCCESSFUL open data file !!**
- **Please, input # of templates for downloading** After this number is entered, data transfer begins to take place. Upon completion of the data transfer, these statements are shown on the monitor;
- **Signal file HAS BEEN CLOSED !!**
- **SUCCESSFUL data downloading !!**
- **Do you want another downloading? (y/n)** If '**y**' is entered, the preceding steps repeat. Otherwise, the user enters the second stage of this software.

C.2.2 Training/Recognition Stage

Once the user stays in this stage, he/she has to flip the manual switch to training/recognition position.

- **Please, flip the switch to signal TRAINING/RECOGNITION**

Training Procedure

Then, the user is prompted if he/she wants the system to learn a new sound.

- **Do you want to train the system to learn a new sound? (y/n)**

If the answer is 'n', the user proceeds to the recognition stage. If the answer is 'y', he/she needs to provide an identification for the new sound, and then presses the ENTER key to start the training procedure. The interactive statements on the monitor are :

- **Please, specify an identification for input signal =,**
- **template # =** whose value is automatically generated by the system software,
- **Please, input SIGNAL for Training.**
- **If ready, Please press ENTER key.**
- **Signal template training in PROGRESS**

For successful training, a summary of the template information is shown:

- **SUCCESSFUL TRAINING**
- **Burst signal !!** (for burst signal), or **Steady sound** (for continuous , steady sound)!!
- **SYNTAX # =**
- **Template # =**
- **Signal template identification =**

Subsequently, the user is prompted if he/she wants the device to learn a new sound, or to recognize another new sound. If the training mode is selected, the affore-mentioned training steps repeat. If the recognition mode is selected, the user enters the recognition stage.

Recognition Procedure

The statement displaying on the monitor is

- **Do you want the system to recognize the signal ? (y/n)**

If the answer is '**n**', the statement to enquire the signal template uploading is displayed on the monitor. But, if the user wants the device to recognize the signal, then the monitor shows the following statements, and the signal recognition process starts.

- **Start to recognize the input signal !**

- **Signal recognition in PROGRESS**

For a successful recognition, a summary of the recognition results appears on the monitor:

- **SUCCESSFUL RECOGNITION**
- **The closest distance measured =**
- **Burst sound, or Steady sound**
- **SYNTAX # =**
- **Template # =**
- **Signal template identification =**

Consequently, the user is prompted for another signal recognition. If the response is '**y**', the preceding recognition steps repeat. If the response is '**n**', he/she is enquired if the user wants to perform a signal template uploading process.

- **Do you want to save memory templates ?? (y/n)**

If the response is not '**y**', the user needs to select one of the following options.

- **What do you want to do next? (please, select one of the following choices)**
- **r : another signal recognition**

- **d : another template file downloading**
- **t : another signal training**
- **e : exit the program**

Otherwise, for the memory uploading, the user provides a template file name for the identification of the stored signal templates. Then, the process of data transfer is performed transparently. The interactive statements are:

- **# of template for uploading =**
- **Please, enter the file name**
- **Successful open file**
- **Successful uploading**
- **File closed**
- **Do you want another signal memory uploading? (y/n)**

If the answer is 'y', the uploading steps repeat. Otherwise, the user has to select one of the previously mentioned options (r; d; t; e).

Appendix D

Evaluation Results

In this work, confusion matrices are used to present the recognition results produced by the complete WARNSIS, the timing analyzer part alone, and the spectral recognizer part alone. To simplify the notation for the confusion matrices given in the following sections, different warning sounds are assigned a "number" as shown in Table D.13. Each assigned number in the first horizontal row indicated the specific warning sound which was identified by a recognition system; and each assigned number in the first vertical column indicated the warning sound which was present in the environments. Each element of the confusion matrix yielded the number of times that a warning sound was identified as the emitted sound in the environments. Based on these results, the recognition rates for each warning sound are derived.

Otherwise stated, the results presented here assumes that the MBD value is set to 0.1024 sec.

TE(L1) represents telephone rings generated from electromechanical ringer with loudness level set at one. ETE(P1) represents telephone rings produced by electronic ringer with pitching adjustment preset at a specific position.

Table D.13: "Numbers" assigned for different warning sounds

Type of Sound	Assigned Number
J1 (B)	1
J2 (B)	2
J3 (B)	3
J4 (B)	4
J5 (S)	5
J6 (B)	6
J7 (S)	7
J8 (B)	8
smoke alarm (S)	9
TE(L1) (PH)	10
TE(L3) (PH)	11
TE(L5) (PH)	12
TE(L7) (PH)	13
ETE(P1) (PH)	14
ETE(P2) (PH)	15
ETE(P3) (PH)	16
ETE(P4) (PH)	17

B: Burst-type Sound
S: Steady Sound
PH: Phone Ring

D.1 The Complete WARNSIS

D.1.1 Recognition Results with Background Steady Noise

Table D.14: Confusion matrix for recognition results generated by the complete WARNSIS in the presence of steady noise

	1	2	3	4	5	6	7	8	9
1	30	.	.	.	.	.	.	.	.
2	.	30	.	.	.	.	.	.	.
3	.	.	30	.	.	.	.	.	.
4	.	.	.	30	.	.	.	.	.
5	.	.	.	.	30	.	.	.	.
6	.	.	.	.	.	30	.	.	.
7	.	.	.	.	.	.	30	.	.
8	.	.	.	.	.	.	.	30	.
9	.	.	.	.	.	.	.	.	30

Table D.15: Recognition rates of burst-type sounds under steady noise condition

Burst-type Sound	Assigned Value	Recognition Rate (%)
J1	1	100
J2	2	100
J3	3	100
J4	4	100
J6	6	100
J8	8	100
Average	-	100

Table D.16: Recognition rates of steady sounds generated by the complete WARNSIS under steady noise condition

Steady Sound	Assigned Value	Recognition Rate (%)
J5	5	100
J7	7	100
smoke alarm	9	100
Average	-	100

Table D.17: Confusion matrix for phone ring recognition generated by the complete WARNSIS under steady noise condition

	10	11	12	13	14	15	16	17
10	30	.	.	.	.	.	.	.
11	.	30	.	.	.	.	.	.
12	.	.	30	.	.	.	.	.
13	.	.	.	30	.	.	.	.
14	.	.	.	.	30	.	.	.
15	.	.	.	.	.	30	.	.
16	.	.	.	.	.	.	30	.
17	.	.	.	.	.	.	.	30

Table D.18: Recognition rates of phone ring generated by the complete WARNSIS under steady noise condition

Phone Ring	Assigned Value	Recognition Rate (%)
TE(L1)	10	100
TE(L3)	11	100
TE(L5)	12	100
TE(L7)	13	100
ETE(P1)	14	100
ETE(P2)	15	100
ETE(P3)	16	100
ETE(P4)	17	100
Average	-	100

D.1.2 Recognition Results with Background of FM Broadcast plus Steady Noise

Table D.19: Confusion matrix for recognition results generated by the complete WARNSIS in the presence of FM broadcast plus steady noise

	1	2	4	5	6	7	8	9
1	30	.	.	.	.	.	.	.
2	.	30	.	.	.	.	.	.
4	.	.	29	.	.	.	1	.
5	.	.	.	30	.	.	.	.
6	.	2	.	.	28	.	.	.
7	.	.	.	.	.	30	.	.
8	.	.	.	.	.	.	30	.
9	.	.	.	.	.	.	.	30

Table D.20: Recognition rates of burst-type sounds produced by the complete WARNSIS under FM broadcast plus steady noise condition

Burst-type Sound	Assigned Value	Recognition Rate (%)
J1	1	100
J2	2	100
J4	4	96.7
J6	6	93.3
J8	8	100
Average	-	98.0

Table D.21: Recognition rates of steady sounds generated by the complete WARNSIS under FM broadcast plus steady noise condition

Steady Sound	Assigned Value	Recognition Rate (%)
J5	5	100
J7	7	100
smoke alarm	9	100
Average	-	100

D.1.3 Recognition Results with Background of AM Broadcast plus Steady Noise

Table D.22: Confusion matrix for recognition results generated by the complete WARNSIS in AM broadcast plus steady noise background

	1	2	4	5	6	7	8	9
1	30	.	.	.	.	.	.	.
2	.	30	.	.	.	.	.	.
4	.	.	30	.	.	.	.	.
5	.	.	.	30	.	.	.	.
6	.	.	.	.	29	.	1	.
7	.	.	.	.	.	30	.	.
8	.	.	.	.	.	.	30	.
9	.	.	.	.	.	.	.	30

Table D.23: Recognition rates of burst-type sounds generated by the complete WARNSIS in AM broadcast plus steady noise environment

Burst-type Sound	Assigned Value	Recognition Rate (%)
J1	1	100
J2	2	100
J4	4	96.7
J6	6	100
J8	8	100
Average	-	99.3

Table D.24: Recognition rates of steady sounds generated by the complete WARNSIS in AM broadcast plus steady noise background

Steady Sound	Assigned Value	Recognition Rate (%)
J5	5	100
J7	7	100
smoke alarm	9	100
Average	-	100

D.1.4 Results of phone ring recognition with minimum burst duration (MBD) set to 1.024 sec

Table D.25: Confusion matrix for phone ring recognition generated by the complete WARNSIS under the condition of FM broadcast and the steady noise with MBD set to 1.024 sec

	10	11	12	13	14	15	16	17
10	28	2	.	.	.	.	.	.
11	.	29	1	.	.	.	.	.
12	.	.	26	4	.	.	.	.
13	.	.	1	29	.	.	.	.
14	.	.	.	.	26	4	.	.
15	.	.	.	.	3	27	.	.
16	.	.	.	.	.	1	29	.
17	.	.	.	.	.	.	2	28

Table D.26: Results of recognition rates of phone rings generated by the complete WARNSIS in FM broadcast plus the steady noise background

Phone Ring	Assigned Value	Recognition Rate (%)
TE(L1)	10	93.3
TE(L3)	11	96.7
TE(L5)	12	86.7
TE(L7)	13	96.7
ETE(P1)	14	86.7
ETE(P2)	15	90.0
ETE(P3)	16	96.7
ETE(P4)	17	93.3
Average	-	92.5

Table D.27: Confusion matrix for the results of phone ring recognition generated by the complete WARNSIS in the presence of AM broadcast plus the steady noise with MBD set to 1.024 sec

	10	11	12	13	14	15	16	17
10	29	1	.	.	.	.	.	.
11	2	28	.	.	.	.	.	.
12	.	.	27	3	.	.	.	.
13	.	.	1	29	.	.	.	.
14	.	.	.	.	27	3	.	.
15	.	.	.	.	2	28	.	.
16	.	.	.	.	.	.	29	1
17	.	.	.	.	.	.	1	29

Table D.28: Results of phone ring recognition rates generated by the complete WARNSIS in the presence of AM broadcast plus steady noise with MBD set to 1.024 sec

Phone Ring	Assigned Value	Recognition Rate (%)
TE(L1)	10	96.7
TE(L3)	11	93.3
TE(L5)	12	90.0
TE(L7)	13	96.7
ETE(P1)	14	90.0
ETE(P2)	15	93.3
ETE(P3)	16	96.7
ETE(P4)	17	96.7
Average	-	94.2

D.1.5 Results of the False-alarm Tests for the complete WARNSIS

Table D.29: Results of the false-alarm tests for the complete WARNSIS with MBD set to 0.1024 sec

Mis-recognized as	FM			AM		
	Heavy Rock	Pop Music	Soft Music	Speech	Soft Music	Soft Rock
J1	1	0	1	1	0	0
J2	0	0	1	0	0	0
J3	0	0	0	0	0	0
J4	0	0	0	0	0	0
J5	2	2	2	1	2	0
J6	0	0	0	0	0	0
J7	0	2	0	0	0	0
J8	2	0	1	0	1	1
Smoke Alarm	0	0	0	0	0	0
TE(L1)	0	0	0	0	0	0
TE(L3)	0	0	0	0	0	0
TE(L5)	0	0	0	0	0	0
TE(L7)	0	0	0	0	0	0
ETE(P1)	0	0	0	0	0	0
ETE(P2)	0	0	0	0	0	0
ETE(P3)	0	0	0	0	0	0
ETE(P4)	0	0	0	0	0	0
Total # of recognitions	5	4	5	2	3	1
Duration (hours)	2	2	2	2	2	2

Table D.30: Results of the false-alarm tests for the complete WARNSIS with MBD set to 1.024 sec

Mis-recognized as	FM			AM		
	Heavy Rock	Pop Music	Soft Music	Speech	Soft Music	Soft Rock
J1	0	0	0	0	0	0
J2	0	0	0	0	0	0
J3	0	0	0	0	0	0
J4	0	0	0	0	0	0
J5	1	0	0	1	1	0
J6	0	0	0	0	0	0
J7	0	0	0	0	0	0
J8	0	0	0	0	0	0
Smoke Alarm	1	1	1	0	0	1
TE(L1)	0	0	0	0	0	0
TE(L3)	0	0	0	0	0	0
TE(L5)	0	0	0	0	0	0
TE(L7)	0	0	0	0	0	0
ETE(P1)	0	0	0	0	0	0
ETE(P2)	0	0	0	0	0	0
ETE(P3)	0	0	0	0	0	0
ETE(P4)	0	0	0	0	0	0
Total # of recognitions	2	1	1	1	1	1
Duration (hours)	2	1	1	0.5	0.5	0.5

D.2 Timing Analyzer Part Alone

D.2.1 Recognition Results with Background Steady Noise

Table D.31: Confusion matrix for warning sound recognition generated by the timing analyzer alone in the presence of steady noise

Type of Sound	J1	J2	J3	J4	J6	J8	Phone Ring
J1	30	.	.	.	.	.	.
J2	.	30	.	.	.	.	.
J3	.	.	30	.	.	.	.
J4	.	.	.	30	.	.	.
J6	.	.	.	.	30	.	.
J8	.	.	.	.	.	30	.
Phone Ring	.	.	.	.	.	.	30

Table D.32: Recognition rates of the timing analyzer part alone in the presence of steady noise

Burst-type Sound	Assigned Value	Recognition Rate (%)
J1	1	100
J2	2	100
J3	3	100
J4	4	100
J6	6	100
J8	8	100
Average	-	100
Phone Ring	-	100

D.2.2 Recognition Results with Background of FM Broadcast Plus Steady Noise

Table D.33: Confusion matrix for warning sound recognition generated by the timing analyzer part alone in the presence of FM broadcast plus steady noise

Type of Sound	J1	J2	J3	J4	J6	J8	Phone Ring
J1	30	.	.	.	.	.	.
J2	.	30	.	.	.	.	.
J3	4	.	26	.	.	.	.
J4	.	.	.	30	.	.	.
J6	.	.	.	.	30	.	.
J8	.	.	.	.	.	30	.
Phone Ring	12	3	.	.	5	10	0

Table D.34: Recognition rates of the timing analyzer part alone in the presence of FM broadcast plus steady noise

Burst-type Sound	Assigned Value	Recognition Rate (%)
J1	1	100
J2	2	100
J3	3	86.6
J4	4	100
J6	6	100
J8	8	100
Average	-	97.7
Phone Ring	-	0

D.2.3 Recognition Results with Background of AM Broadcast Plus Steady Noise

Table D.35: Confusion matrix for warning sound recognition generated by the timing analyzer part alone in the presence of AM broadcast plus steady noise

Type of Sound	J1	J2	J3	J4	J6	J8	Phone Ring
J1	30	.	.	.	.	.	.
J2	.	30	.	.	.	.	.
J3	3	.	27	.	.	.	.
J4	.	.	.	30	.	.	.
J6	.	.	.	.	30	.	.
J8	.	.	.	.	.	30	.
Phone Ring	6	5	.	7	.	12	0

Table D.36: Recognition rates of the timing analyzer part alone in the presence of AM broadcast plus steady noise

Burst-type Sound	Assigned Value	Recognition Rate (%)
J1	1	100
J2	2	100
J3	3	90
J4	4	100
J6	5	100
J8	8	100
Average	-	98.3
Phone	-	0

D.3 False-alarm Results for the Timing Analyzer Alone

Table D.37: False-alarm test results of the timing analyzer part alone with MBD set to 0.1024 sec

mis-recognized as	FM			AM		
	Pop Music	Rock Music	Classical	Speech + Music	Pop Music	Speech
J1	16	19	13	34	16	40
J2	16	13	10	18	26	25
J3	0	0	0	0	0	0
J4	0	2	1	0	2	1
J6	12	19	7	9	5	16
J8	0	1	0	0	0	0
Steady Sound	23	23	45	3	50	0
Phone	0	1	1	0	0	1
Total # of Mis-recognitions	67	78	77	64	99	83
Duration (minutes)	56	46	56	19	59	24

Table D.38: False-alarm test results of the timing analyzer part alone with MBD set to 1.024 sec

mis-recognized as	FM			AM		
	Pop Music	Rock Music	Classical	Speech + Music	Pop Music	Speech
J1	0	0	0	0	0	0
J2	0	0	0	0	0	
J3	1	2	1	2	3	2
J4	0	0	0	0	0	0
J6	0	0	0	0	0	0
J8	0	0	0	0	0	0
Steady Sound	20	10	24	15	10	14
Phone	3	2	1	6	6	2
Total # of Mis-recognitions	24	14	26	23	19	18
Duration (minutes)	30	30	30	30	30	30

D.4 Spectral Recognizer Part Alone

D.4.1 Recognition Results with Background Steady Noise

Table D.39: Confusion matrix for warning sound recognition generated by the spectral recognizer part alone in the presence of steady noise

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2	.	30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
3	.	.	30	.	.	.	.	.	.	.	.	.	.	.	.	.	.
4	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.	.	.
5	.	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.	.
6	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.
7	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.
9	.	.	.	.	.	.	.	.	28	1	1	.	.	.	.	.	.
10	.	.	.	.	.	.	.	.	2	27	1	.	.	.	.	.	.
11	.	.	.	.	.	.	.	.	1	1	28	.	.	.	.	.	.
12	.	.	.	.	.	.	.	.	.	.	2	28	.	.	.	.	.
13	.	.	.	.	.	.	.	.	.	.	.	.	30	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	.	.	.	30	.	.	.
15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	30	.	.
16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	29	.
17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	28

Table D.40: Results of steady sound recognition rate generated by the spectral recognizer part alone in steady noise background

Steady Sound	Assigned Value	Recognition Rate (RR) in %
J5	5	100
J7	7	100
smoke alarm	9	93.3
Average	-	97.6

Table D.41: Results of burst-type sound recognition rates produced by the spectral recognizer part alone in steady noise background

Burst-type Sound	Assigned Value	Recognition Rate in (%)
J1	1	100
J2	2	100
J3	3	100
J4	4	100
J6	6	100
J8	8	100
Average	-	100

Table D.42: Results of phone ring recognition rate produced by the spectral recognizer part alone in steady noise background

Phone Ring	Assigned Value	Recognition Rate (%)
TE(L1)	10	90.0
TE(L3)	11	93.3
TE(L5)	12	93.3
TE(L7)	13	100
ETE(P1)	14	100
ETE(P2)	15	100
ETE(P3)	16	96.7
ETE(P4)	17	93.3
Average	-	95.8

D.4.2 Recognition Results with Background of FM Broadcast plus Steady Noise

Table D.43: Confusion matrix for the results of warning sound recognition generated by the spectral recognizer part alone in FM broadcast and steady noise background

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
2	.	22	.	.	.	.	.	.	.	.	1	7	.	.	.	.	.
3	.	5	6	.	.	.	1	.	.	2	5	6	1	1	1	1	2
4	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.	.	.
5	.	8	.	.	22	.	.	.	.	.	.	.	.	.	.	.	.
6	.	.	2	.	.	1	.	.	14	10	.	.	.	.	.	1	1
7	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.
9	.	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.
10	.	1	.	.	.	2	.	1	.	24	1	1	.	.	.	.	.
11	.	.	1	.	4	4	.	.	.	.	21	.	.	.	.	.	.
12	.	.	1	.	1	1	.	.	3	.	.	22	1	1	.	.	.
13	.	.	1	.	.	3	2	.	1	2	.	.	20	1	.	.	.
14	.	.	2	.	.	.	3	.	.	3	.	.	2	19	1	.	.
15	.	.	.	.	.	.	1	.	.	.	5	.	.	.	22	2	.
16	.	.	2	.	.	.	2	.	.	.	.	.	.	1	4	21	.
17	.	.	1	.	.	.	1	.	.	.	3	.	.	.	.	5	20

Table D.44: Results of steady sound recognition rate produced by the spectral recognizer part alone in FM broadcast plus steady noise background

Steady Sound	Assigned Value	Recognition Rate (%)
J5	5	73.3
J7	7	100
smoke alarm	9	100.0
Average	-	91.1

Table D.45: Results of burst-type sound recognition rates produced by the spectral recognizer part alone in FM broadcast plus steady noise background

Burst-type Sound	Assigned Value	Recognition Rate(RR) (%)
J1	1	96.7
J2	2	73.3
J3	3	20.0
J4	4	100
J6	6	3.3
J8	8	100
Average	-	65.6

Table D.46: Results of phone ring recognition rates produced by the spectral recognizer part alone under FM broadcast plus steady noise condition

Phone Ring	Assigned Value	Recognition Rate (%)
TE(L1)	10	80.0
TE(L3)	11	70.0
TE(L5)	12	73.3
TE(L7)	13	66.7
ETE(P1)	14	63.3
ETE(P2)	15	73.3
ETE(P3)	16	70.0
ETE(P4)	17	63.3
Average	-	70.0

D.4.3 Recognition Results with Background of AM Broadcast plus Steady Noise

Table D.47: Confusion matrix for warning sound recognition generated by the spectral recognizer part alone under AM broadcast plus steady noise condition

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	29	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
2	.	23	.	.	.	.	.	.	6	1	.	.	.	.	.	.	.
3	.	4	7	.	.	.	1	.	.	.	.	.	5	5	5	3	.
4	.	.	.	30	.	.	.	.	.	.	.	.	.	.	.	.	.
5	.	.	.	.	24	.	.	.	.	.	.	6	.	.	.	.	.
6	.	4	.	.	.	2	.	.	.	1	.	6	16	1	.	.	.
7	.	.	.	.	.	.	29	.	.	.	.	.	.	.	1	.	.
8	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.	.
9	.	.	.	.	.	.	.	.	30	.	.	.	.	.	.	.	.
10	.	1	.	.	.	.	.	.	.	25	4	.	.	.	.	.	.
11	.	1	.	.	.	1	.	.	3	.	20	5	.	.	.	.	.
12	.	.	1	.	.	3	.	.	3	.	2	19	3	.	.	.	.
13	.	2	.	.	4	3	.	.	.	.	.	4	17	.	.	.	.
14	.	1	.	.	.	1	.	.	.	.	.	4	.	22	.	.	.
15	.	2	.	.	3	.	.	.	1	.	.	.	.	3	21	.	.
16	.	3	.	.	1	.	.	.	3	.	.	.	.	.	.	20	.
17	.	2	.	.	1	1	.	1	1	.	.	.	.	.	.	2	22

Table D.48: Results of steady sound recognition rates produced by the spectral recognizer part alone under AM broadcast plus steady noise condition

Steady Sound	Assigned Value	Recognition Rate (%)
J5	5	80.0
J7	7	93.3
smoke alarm	9	100.0
Average	-	91.1

Table D.49: Results of burst-type sound recognition rate produced by the spectral recognizer part alone in the presence of AM broadcast plus steady noise

Burst-type Sound	Assigned Value	Recognition Rate (RR) (%)
J1	1	96.7
J2	2	76.7
J3	3	23.3
J4	4	100
J6	6	6.7
J8	8	100
Average	-	67.2

Table D.50: Results of phone ring recognition rate produced by the spectral recognizer part alone in the presence of AM broadcast plus steady noise

Phone Ring	Assigned Value	Recognition Rate (%)
TE(L1)	10	83.3
TE(L3)	11	66.7
TE(L5)	12	63.3
TE(L7)	13	56.7
ETE(P1)	14	73.3
ETE(P2)	15	70.0
ETE(P3)	16	66.7
ETE(P4)	17	73.3
Average	-	69.2

D.4.4 Results of false-alarm tests for the spectral recognizer part alone

Table D.51: False-alarm tests for the spectral analyzer part alone

mis-recognized as	FM			AM		
	Heavy Rock	Pop Music	Soft Music	Speech	Soft Music	Soft Rock
J1	0	0	0	0	0	0
J2	0	0	0	0	0	0
J3	0	1	1	6	1	1
J4	0	0	0	0	0	0
J5	0	0	0	0	0	0
J6	0	0	1	24	0	0
J7	1	30	4	0	17	4
J8	0	0	1	0	0	0
Smoke Alarm	0	0	4	31	10	0
TE(L1)	28	10	30	2	3	7
TE(L3)	30	7	24	9	0	36
TE(L5)	9	29	9	48	79	23
TE(L7)	50	43	40	0	10	46
ETE(P1)	1	0	4	0	0	0
ETE(P2)	0	0	0	0	0	0
ETE(P3)	0	0	0	0	0	1
ETE(P4)	1	0	2	0	0	2
Total # of Mis-recognitions	120	120	120	120	120	120
Duration (minutes)	3.42	4	4.27	4.8	4.2	3.57

Appendix E

Specifications

1. Power Supply :

   - + 5 V : 700 mA
   - + 12 V : 64.3 mA
   - − 12 V : 51.6 mA

2. Signal Features: Timing and short-time spectral patterns

3. The WARNSIS: a 'hybrid' system consisting of the parts of the timing analyzer and the spectral analyzer

4. Timing Analyzer Part Alone:

   - Function : the classification of warning sounds based on the absolute short-time average signal amplitudes
   - short-time duration : 12.8 msec
   - Timing Features : the repetition period and the average signal burst width for burst-type sounds; whereas a rising signal amplitude transition and a new signal amplitude for steady sounds

5. Spectral Recognizer Part Alone:

   - Function : extraction of short-time spectral features from warning sounds by the filter-bank approach,
   - Short-time Duration : 12 msec
   - # of filters : 8
   - Type of Filter : digital biquad
   - Frequency Span : 100 Hz to 5.0 kHz
   - Implementation : software
   - Pattern Matching : Dynamic Time Warping

6. Modes of Operations:

   - burst-type and steady warning sound recognition
   - phone ring recognition

7. Recognition Accuracy :

   - 98 % for steady and burst-type warning sounds for a SNR of over 10 dB
   - 93 % for phone rings for a SNR of over 10 dB or better

8. False-alarm Rate:

   - one false recognition per 90 minutes (worst-case) for burst-type and steady sounds

- no false ring indications

9. Recognition Time : 0.5 sec to 10 sec depending on the type of warning sounds