PSYCHOACOUSTIC TESTING OF SEARLE'S MODEL BASED ON HUMAN AUDITION

GNANASIROMONI SURESH CHANDRAN

PSYCHOACOUSTIC TESTING OF SEARLE'S MODEL BASED ON HUMAN AUDITION

BY

© GNANASIROMONI SURESH CHANDRAN B.E. (Elect.)
M.E. (Elect.)

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Engineering.

Faculty of Engineering and Applied Sciences
Memorial University of Newfoundland
August 1984

St. John's Newfoundland Canada A1B 3X5

ABSTRACT

A software model was built to investigate the feasibility of reducing the dimensionality of the representation of speech for speech recognition. The model was based on a model of human audition developed by Searle in hardware, the characteristics of which closely match physiological characteristics obtained experimentally from the human auditory system. The model consisted of 16, one third octave bandpass filters followed by envelope detectors, the outputs of which were then subjected to the linear discrete cosine transform in an attempt to reduce the data to a small number of perceptually important dimensions. Several signal processing techniques were investigated to reconstruct the filtered, detected and transformed speech to permit effective intelligibility testing. Most of the work done was qualitative and the methodology concentrated on testing the model rather than the theory. A test sentence produced by a male and female speaker was transformed using the model and the transformed speech was reconstructed. Informal testing with a few experienced listeners suggests that it might be possible to recognize speech with as few as three out of the sixteen channels. This study confirms the value of realizing the models in software. Although the model is not real time it was far easier to modify than the comparable hardware models.

ACKNOWLEDGEMENT

I express my thanks to Professor Michael P. Bruce-Lockhart, who has been a source of inspiration and guidance throughout the work. I would also like to thank and acknowledge the following: Dr. Michael Zagorski of the Department of Psychology and Dr. Harold J. Paddock of the Department of Linguistics for their suggestions; Mr. Dwight Howse for his instructions and help in the use of Fourier Analyzer System; Corinne Reddle for making available the facilities and help with using the PDP 11/60 System; Dr. A. S. J. Swamidas for his encouragement when I needed it; and Valerie Blundell for her typing.

CONTENTS

PAGE

PAGE

LIST OF TABLES

LIST OF FIGURES

PAGE

LIST OF SYMBOLS

C_{ij} [C]	Covariance matrix
$G_x(k)$	k'th discrete cosine transform coefficient
k, N, M	Filter bank indices and counts
S(n)	Log spectral magnitude from n'th filter
$X_n(e^{j\omega_k})$	Short-time Fourier transform
$a_n(\omega_k)$	Real part of the output from k'th filter
$b_n(\omega_k)$	Imaginary part of the output from k'th filter
$h_k(m)$	Impulse response of the filter
$p_k(n)$	Real window sequence
x(m)	Input sequence
$Y_k(n)$	Output of the k'th bandpass filter
ω_k	Radian frequency

INTRODUCTION

Developments in digital speech processing techniques as well as in digital hardware have made it possible to implement speech recognition systems. The main goal is the transmissionor storage or real time processing of speech adequate for satisfying specified perceptual criteria. Speech recognition involves the determination of spoken words or phrases or sentences. The implementation of these systems has been concerned with syntactic and semantic analysis of utterances, closely allied with linguistic theories of speech.

Many theories for automatic recognition of speech are based on the studies of speech production by the human vocal tract. In voice recognition, time-honored mathematical abstractions and equations to break speech into fundamental processing components are used. But the accuracy of computer based perception still falls short of that attained by human ears. Hence some researchers have come to believe that a sounder approach is to emulate the time domain processing of the human auditory system (the most successful speech recognizer presently in existence). Studies over the past few years have begun to show that neural temporal patterns from the cochlea's firing nerves contain much more perceptual information than had earlier been believed. Quite obviously signal processing techniques, physiology, psychology and linguistics must all be combined to achieve a successful outcome in the area of speech recognition. The research in this thesis is directed toward the software simulation of a model based on human audition and more particularly at a reduction in the dimensionality of speech representations. The two basic objectives of the work are (1) the development of software model based on human audition that parallels the Searle hardware model and (2) the achievement of successful reconstruction of speech at the output of the model in order that an investigation of a reduction in the dimensionality of speech fragment representations may be carried out.

Since the work is based on human audition, Chapter 1 describes the physiology and psychophysics of the human auditory system. The linguistic categorization of speech and the human speech production system are also explained. The various models developed in the area of speech recognition based on human audition are also outlined.

The Searle hardware model forms the basic background for the work and hence the design considerations, implementation and the work done using the model are explained in Chapter 2. The various data reduction techniques are also outlined.

In order to give a mathematical background, the implementation of the filter detector is explained mathematically using short-time Fourier Transforms in Chapter 3.

The implementation of the model, the signal analysis techniques used and the approaches to the signal reconstruction are explained briefly in Chapter 4.

The results of the analysis are discussed in Chapter 5.

1. AUDITORY PERCEPTION

This chapter includes a brief description of the physiology of the ear and the various experimental paradigms pertaining to design considerations which must be taken into account in the modelling of the human auditory system. The chapter also reviews human speech production system, as well as the perception of speech based on its linguistic aspects and concludes by outlining a few models of human audition.

1.1 THE PHYSIOLOGY OF THE EAR

The primary acoustic transducer of man is shown in Fig. 1.1. The acousto-mechanical components of the organ are divided into three anatomical units, the outer, the middle and the inner ear.

1.1.1. The Outer Ear

The outer ear consists of a cartilagious structure on the side of the head, the pinna and the ear canal or external meatus. Pinna facilitates localization of sound sources (Batteau, 1967) while the meatus, a uniform pipe open at one end and closed at the other end, is a resonant structure with a resonant frequency of approximately 3000 Hz (Flanagan, 1972).

1.1.2. The Middle Ear

The middle ear is the region bounded by the tymphanic membrane or the ear drum and the oval window. This region consists of the three

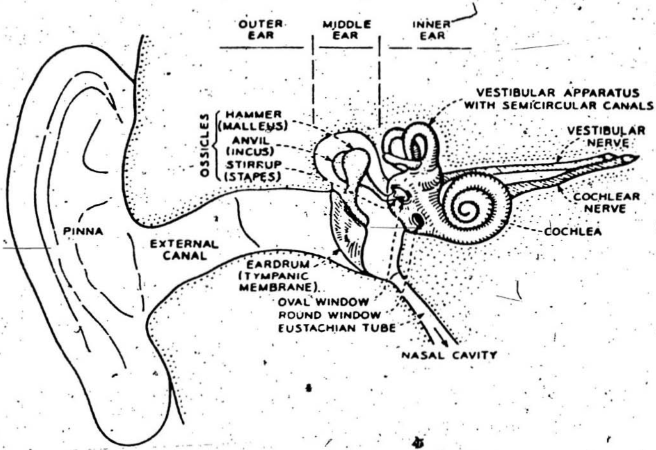


Fig. 1.1 Schematic diagram of the ear showing the outer, middle and inner regions. (from Flanagan, 1972).

ossicles, the malleus, the incus and the stapes, and the supporting musculature. Pressure fluctuations propagated through the meatus set the membrane in motion. The function of the ossicles is one of impedance transformation from the air medium of the outer ear to the liquid medium of the inner ear, the cochlea. The important characteristic of the middle ear is its transmission as a function of frequency, that is, the volume displacement of the stapes foot plate produced by a given sound pressure at the ear drum.

1.1.3. The Inner Ear

The inner ear is composed of the cochlea, the vestibular apparatus and the auditory nerve termination. The cochlea is responsible for the transduction of the mechanical motion at the stapes to neural pulses in the auditory nerve. The cochlear chamber is filled and it is divided along its whole length by a partition which itself is a channel - the basilar membrane (Fig. 1.2). The movement of the basilar membrane causes the hair cells in the organ of corti to move and this motion stimulates the hair cells which initiates the electrical activity in the auditory nerve (Bekesy 1960, Tonndorf 1970).

1.2 PHYSIOLOGICAL AND PSYCHOPHYSICAL MEASUREMENTS

The concept of basilar membrane performing a frequency analysis on incoming sound was first considered by Helmholtz (1885). Current knowledge of the membrane is due almost exclusively to the efforts of Bekesy (1960) who made measurements of basilar membrane vibration. Fig. 1.3 shows his results. These frequency response curves were quite broad, the 3 db

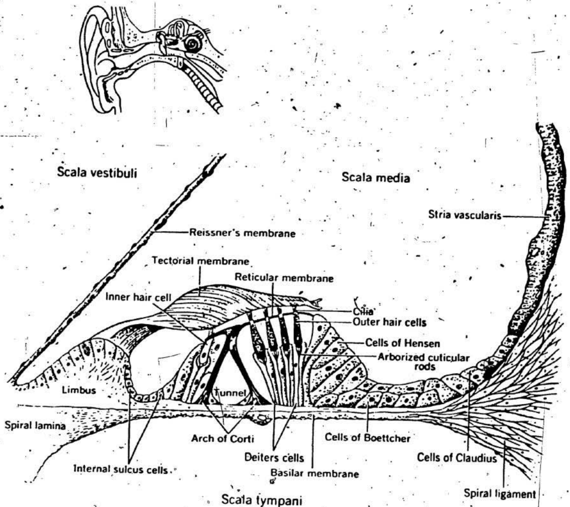


Fig. 1.2 Cross section of the scale media showing the basilar membrane and the hair cells (from Gulick, 1971).

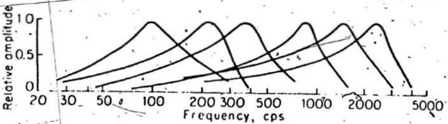


Fig. 1.3 Resonance curves for six positions along the basilar membrane (from Bekesy, 1960)

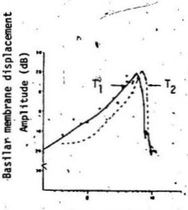


Fig. 1.4 The resonance curves for two positions along the basilar membrane of the squirrel monkey (from Rhode 1971).

response bandwidth of the membrane is approximately one octave.

Measurement of basilar membrane motion using the Mossbauer effect (Johnstone and Boyle 1967, Rhode 1971) suggests considerably sharper tuning characteristics. Fig. 1.4 illustrates the measurement made by Rhode which indicate a 3 db bandwidth of less than one half octave. Kiang et al., (1974) measured the frequency response of single auditory nerve fiber of cats to acoustic stimuli; the tuning curves thus obtained (Fig. 1.5) are also narrower than those of Bekesy.

The psychoacoustic tuning curves of human as found by Zwicker (1974) in Fig. 1.6 are almost identical to the one of Kiang et al., of Fig. 1.5. Besides this, the psychophysical auditory masking experiments used by Patterson (1976) indicate that the auditory filter shapes have a narrow bandwidth of 0.13 of the center frequency.

In summary the different experimental paradigms given above suggest that the peripheral auditory system - the ear drum, the basilar membrane and the hair cells - splits the incoming sound wave into thousands of separate channels, each with a frequency resolution of roughly one third of an octave.

1.2.1. Latency or Response Delay

Apart from the concept of the basilar membrane performing a frequency analysis, studies on the membrane indicate a response delay or latency in the membrane motion that decreases with increasing frequency (Bekesy, 1960). Flanagan (1972) illustrated this response delay using a digital computer simulation of the impulse responses for 40 points along the basilar membrane. The simulation is shown in Fig. 1.7. The figure indicates

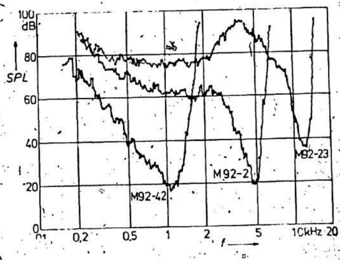


Fig. 1.5 Tuning curves in the auditory nerve of cats (from Kiang, 1973 cited in Zwicker, 1974).

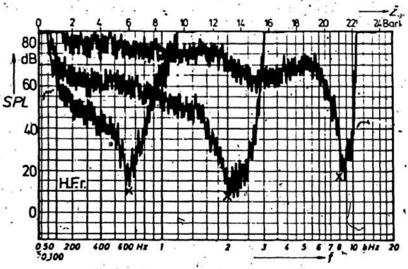


Fig. 1.6 Psychoacoustic tuning curves in humans (from Zwicker 1974).

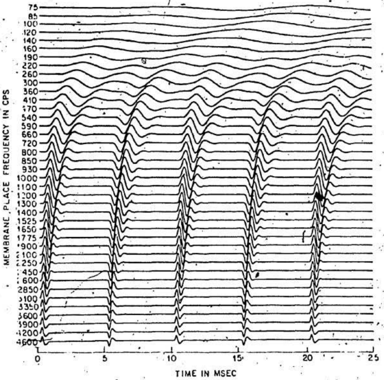


Fig. 1.7 Digital Computer Simulation of the pulse response of the basilar membrane (from Flanagan, 1972).

that the interval between pulse onset and first indication of the membrane motion increases with decrease in frequency. Kiang et al. (1965) recorded the response delay of nerve discharges and their results indicated the presence of an inverse relationship between frequency and response delay. The auditory system's limited temporal resolution cannot follow the temporal changes if the changes occur too rapidly and the psychophysical measurements made by Viemeister (1979) indicate that the temporal resolution of the auditory system is about 2.5 milli-seconds.

1.3 PSYCHOACOUSTICS - CRITICAL BANDWIDTH

The concept of critical band relates to many areas of psychoacoustic measurements (Scharf, 1970). The critical bandwidth is that bandwidth at which subjective responses rather abruptly change and it refers to a filtering process assumed to take place within the auditory system. Critical band theory treats the sound energy within the band differently from sound energy outside the band, analogous to a band pass filter having a bandwidth equal to the critical bandwidth. This analogy has been used to analyse the results of experiments involving the measurement of pitch, loudness, intelligibility of speech, etc. Fig. 1.8 summarizes most of these results and shows how the critical bandwith changes as a function of its center frequency. The bandwidth of a 1/3 octave filter shown as dashed line in the figure matches these results closely for frequency above 400 Hz. Fig. 1.9 shows a more detailed comparison of the filter characteristics using the single response curves of Kiang (1974). As reported by Searle et al., (1979) the above-resonance skirt of the neural response is much sharper than the 1/3 octave response, whereas the below-

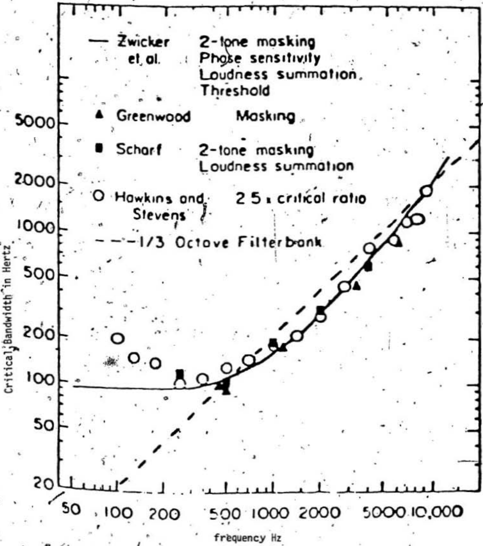


Fig. 1.8 Critical Bandwidth as a function of frequency (from Tobias, 1970).

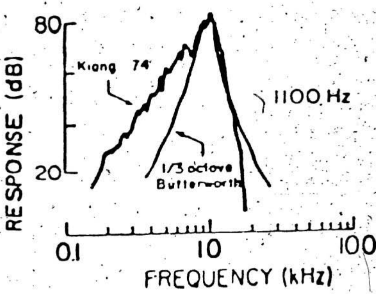


Fig. 1.9 Comparison of 1/3-octave response of auditory nerve fiber (from Kiang, 1974).

resonance skirt is less sharp, but in the critical region within 15 db of maximum response, the agreement is reasonable.

1.4 HUMAN SPEECH PRODUCTION AND PERCEPTION

Speech can be produced as well as heard by a listener and consequently the auditory processing of speech bears a close relation to the cognitive and motor activity involved in the generation of speech. It is appropriate therefore to look into some of the acoustic properties and linguistic categories of speech.

1.4.1. Speech Production

Speech represents a sub-class of audible acoustic signals - those that can be produced by the human vocal mechanism. Speech production involves an interaction between three groups of structures in the human anatomy - the lungs, the larynx (which contains the vocal chords) and the resonant cavity formed by the vocal tract. Fig. 1.10 shows the schematic diagram of the human speech production mechanism.

Speech is radiated from this system when air is expelled from the lungs and the resulting flow of air is perturbed by the vocal tract. These speech sounds can be classified into three classes according to their mode of excitation.

Voiced sounds are produced by forcing air through the opening between the vocal chords (glottis) with the tension of the vocal chords adjusted, thereby producing quasi-periodic pulses of air which excite the vocal tract.

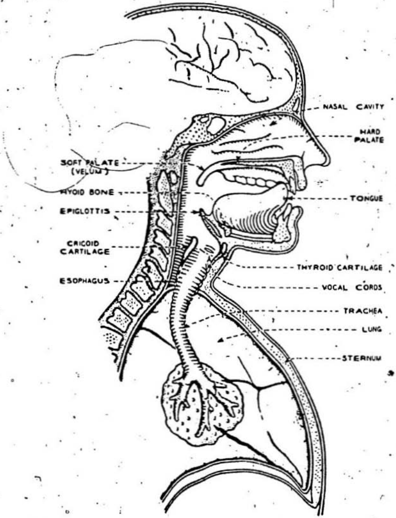


Fig. 1.10 Schematic diagram of the vocal mechanism (from Flanagan 1972).

Unvoiced sounds (fricatives) are generated by forcing air through the constriction in the vocal tract thereby causing turbulence. This creates a broad spectrum of noise source to excite the vocal tract.

Plosive sounds result from the transient excitation of the vocal tract caused by an abrupt release of pressure built up at some closure point.

Most languages can be described in terms of a set of distinctive sounds or phonemes. Each phoneme is classified as either a continuant or non-continuant sound. Continuant sounds are produced by fixed vocal tract configurations excited by the appropriate source. The class of continuants includes the vowels, the fricatives (voiced and unvoiced) and the nasals. The remaining sounds (diphthongs, semi-vowels, stops and affricates) are produced by a changing vocal tract configuration, and are classed as non-continuants.

1.4.2. <u>A Linguistic Description of Speech</u>

Speech signals are composed of a sequence of sounds. These sounds and the transition between them serve as a symbolic representation of information and can be described by segments which are characterized by attributes or features. The arrangement of these sounds is governed by the rules of the language.

About 20 to 30 features are necessary to describe the segments that occur in most languages and have been postulated by linguists to fall into several groups. For example one feature describes the basic dichotomy of vowel vs. consonant, while a set of features specify the place of articulation of the segment. Another set of features are necessary to describe the manner of articulation and type of acoustic source of vocal tract excitation. Finally certain prosodic features are used to indicate stress, length and tonal attributes of vowels.

1.4.3. Feature Detectors in the Auditory System

The most straight-forward-view of speech perception would hold that speech is composed of a series of acoustic features which correspond to a particular phoneme from the listeners' language. Phonemes would then be directly extracted in a speech wave and combined by an unknown mechanism into words and phrases which comprise speech.

Eimas and Corbit (1973) demonstrated the existence of feature detectors in the auditory system and proposed that feature extraction is accomplished by sets of detectors that are specifically tuned to the relevant information underlying feature distinctions and that yield as their output specific distinct phonetic feature values. Cole and Scott (1974) emphasized the importance of transitional, invariant and envelope cues present in the acoustic waveform monitored to perceive speech. Miller (1975) conducted experiments to provide behavioral evidence for the existence of detector mechanisms that extract information relevant for the assignment of phonetic feature values and supported certain specific assumptions containing the operation of the detector systems.

These and other authors have concluded that their results suggest the presence of feature detectors in the auditory system each of which is sensitive to one particular aspect or feature of the auditory stimuli.

1.4.4. Speech Perception

Auditory mechanisms coupled with central processes are predisposed to manipulate segments and features of the type that occur in speech and the attributes of which are anchored to the perceptual processes involved in speech. This has been emphasized by linguists and psychologists.

The general objectives of a theory of speech perception are to describe the process whereby an acoustic speech signal is decoded into linguistic units.

All acoustic units undergo some common peripheral processing. As a consequence of this, an acoustic signal is transformed into some kind of neural space time pattern and all the subsequent analysis and decision with regard to the signal are based on the transformation of this pattern. The patterns that result from the peripheral processing are related to the segments and features that constitute the lowest levels of linguistic description of the utterance. The listener presumably focuses on certain aspects of these patterns and processes them further in order to achieve an organized linguistic description.

1.5 VARIOUS MODELS BASED ON HUMAN AUDITION

Because of the great diversity and variety of techniques used to attempt recognition of speech, only those techniques closely related to the current work are described.

Christovich et al., (1974) outlined a functional model of auditory analysis based on psychophysical and neurophysical facts. The model was implemented with resonant circuits for auditory spectral analysis followed by principal transformation of signal intensity which included half wave rectification, nonlinear amplitude transformation and peripheral adaptation. The model was used to simulate certain effects taking place in the auditory analyzer.

Experimental data from the basilar membrane motion and auditory nerve fiber activity for sinusoidal and click-stimuli along with the study of sensory receptors/neurons have been used by Dolmalzan et al., (1977) to define a model which simulated one section of the basilar membrane. The model consisted of 128 sections with center frequencies decreasing from 10 KHz to 25 Hz. The model was used to reproduce the harmonic and transient behaviour of peripheral auditory system. A Russian word utterance was processed and the model was able to detect the maximum energy regions of the input signal, giving frequency and intensity information pertaining to the spectrum of speech.

In the model described by Pols (1971), the spectral analysis during the pronunciation of a word was carried out by a bank of 17 1/3 octave bandpass filters. The outputs of the filters were logarithmically amplified and the maximum amplitude of the envelope was determined and sampled every 15 msec. In this way the word was characterized by a sequence of sample points in a 17 dimensional space. By method of principal component analysis the original 17 dimensions were reduced to 3. After a linear time normalization, the 3 dimensional trace was used to identify the spoken utterance.

White and Neely (1976) described automatic speech recognition using a model of 25 1/3 octave Chebyshev filters. The output of the filter bank was passed through a rectifier integrator circuit and digitized. The features extracted from the model were subjected to data reduction stage using character string encoding and classified using linear time scaling. The unknown utterances were recognized by template match on known utterances.

Recognition of connected speech was carried out by Mariani and Lienard (1977) by constructing a model of 1/3 octave filter bank of 32 bandpass filters over the range 200 Hz to 7000 Hz. The speech spectrogram thus obtained was divided into various segments. The acoustical parameter variation within the segments were used for synthesis and recognition.

Zwicker and Terhardt (1979) realized a speech recognition model based on psycho-acoustic principles of the perception of loudness, pitch, roughness and subjective duration. The model consisted of 24 bandpass filters followed by rectification by a lowpass filter with a time constant of 1.3 msec and log amplification. From the output of the model the necessary spectral, loudness and roughness parameters were extracted. The recognition process was characterized by discrimination between speech and silent periods, detection of syllable peaks and assumption of syllable boundaries. The model was used for recognition of German vocabulary.

Linggard and Black (1981) described a model of the basilar membrane with a digital filter bank of 128 second order band pass filters for a real time speech recognition system. The short time wide band spectra from this filter bank was processed to extract the perceptually relevant acoustic features for word recognition.

2. SEARLE'S HARDWARE MODELLING OF THE AUDITORY PERIPHERY

In this chapter the process domain (that is, the domain representation that supports Searle's model), the hardware design considerations, implementation and the work done using the model are described briefly. It also includes the theory behind data reduction and the various data reduction techniques used in speech analysis including Principal Component Analysis and the Discrete Cosine Transform.

2.1 THE PROCESS DOMAIN

It can be argued that the modelling of the human auditory system involves two steps: the first being to process the incoming speech signal in a manner similar to the human peripheral system and the second to extract the necessary relevant features for recognition. In order to process the signal the relevant spectral and temporal characteristics of the speech signal should be detected by the model.

The vocal mechanism is a quasi-stationary source of sound. Its excitation and normal modes change with time. Any spectral measure applicable to the speech signal should therefore reflect temporal features of perceptual significance as well as spectral features. The standard Fourier representations that are appropriate for periodic, transient or stationary random signals are not applicable to the representation of speech whose properties change markedly as a function of time. Based on this Flanagan (1972) developed short - time Fourier Transform as a mathematical approach to speech processing (Flanagan 1972, Rabiner and Schafer, 1978).

Cerrillo (Cerrillo, 1957 cited in Rayment 1977) also reported that Fourier analysis does not provide the needed temporal characteristics of the auditory stimuli and he suggested a process domain between the time and frequency domains such that all the relevant signal features can be detected. In considering the process domain, a compromise has to be reached between spectral and temporal resolution as arbitrarily good resolution of both cannot be simultaneously obtained. In a careful review of the literature on physiological and psychophysical experiments done on the auditory system, Searle was able to define the process domain for auditory processing as a three-dimensional space of time, frequency and amplitude with a resolution of approximately 1/3 to 1/6 of an octave in the frequency dimension and 5 millisec in the time dimension at 1 KHz (Searle 1975 as reported by Rayment 1977).

2.2 SEARLE'S HARDWARE MODELLING OF THE AUDITORY PERIPHERY

The results of the different experimental paradigms that were described in the previous chapter were taken into consideration in constructing a front end acoustic analyzer to model the human ear. The hardware model was explained in detail by Dockendorff (1978).

The acoustic analyzer is shown in Fig. 2.1. It consisted of a bank of 18, 1/3 octave filters spanning the frequency range from 125 Hz to 6.3 KHz. The filter bank modelled the frequency characteristics of the basilar membrane and the 1/3 octave bandwidth was chosen for its similarity to the critical bandwidth (Fig. 1.8) and to the skirt response observed by Kiang (Fig. 1.9). The filters were realized using three pole pair,

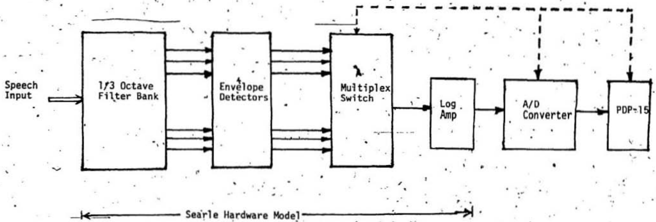


Fig. 2.1 The Searle Hardware Model.

Butterworth designs with maximally flat frequency response. Even though the ear has 30,000 redundant overlapped filter channels, for practical reasons, assuming that output of the adjacent channels vary very little, only enough 1/3 octave filters were used to span completely the band without redundancy.

To model the rectification and the envelope detection function of the hair cells, each filter in the bank was followed by a detector. The fast-click response of the auditory nerve fiber and the inverse relationship that exists between the response delay and the filter bandwidth was considered in the design of the detectors. The rise time of the 1 KHz envelope detector[1] was chosen to be 5 msec. The other detectors in the bank were chosen such that they had appropriately scaled rise times proportional to the channel center frequency. The rise time of any channel was given (in millisecs) by $5.9/f_0$ where f_0 was the center freq. in KHz. To match the critical band curve of Fig. 1.8, the detector outputs of the lower frequency channels were added together (i.e. 125 Hz + 160 Hz and 200 Hz + 250 Hz). The detector output channels were multiplexed and the resultant signal was sampled every 1.6 msec. The sampled output was passed through a logarithmic amplifier, included to match the perceived loudness of the ear. The output was digitized and stored in a PDP-15 computer for further processing.

2.3 WORK DONE USING SEARLE'S MODEL

As discussed previously the model approximated the preprocessing of the auditory peripheral system of the human ear. The various spectral

1 It has since been discovered that these were incorrectly reported. Above 400 Hz, fixed 7 ms time constants were used, slower at low frequencies.

and temporal characteristics of speech were preserved by the model. In order to visually compare the amplitudes of different channels at a given time and to view the temporal and spectral aspect of speech, the output data of the filter detector system was plotted as a succession of spectra, that is log magnitude vs log frequency with time as the parameter. Fig. 2.2 shows such a running spectrum plot of the stop consonants (Searle et al., 1979).

The previous chapter outlined briefly the psychophysical experiments that confirmed the presence of feature detectors in the auditory system sensitive to particular aspects of the auditory stimuli. Further, psychophysical experiments have also been used to determine the characteristics of these acoustic features of the speech signals. Various tape splicing experiments conducted by Cole and Scott (1974) and Schatz (1954) provide information about the recognition of different cues present in the speech waveform. In the work done by Searle et al. (1979) recognition was based on these features which were abstracted from the filter detector output, viz., the voice onset time, location and shape of the spectral peaks during bursts and during transition regions, average formant track slopes etc. Measurement of these features was explained by Searle et al. (1979). The recognition process was carried out by using Discriminant Analysis (based on statistical decision theory) to decide from the feature detection data what phoneme was uttered.

The above procedure was followed for stop consonant discrimination (Searle et al., 1979) and a discrimination accuracy of about 77% for stop consonants in the initial position was achieved with a 15 speaker data set. The same model was also used to test the discrimination of isolated

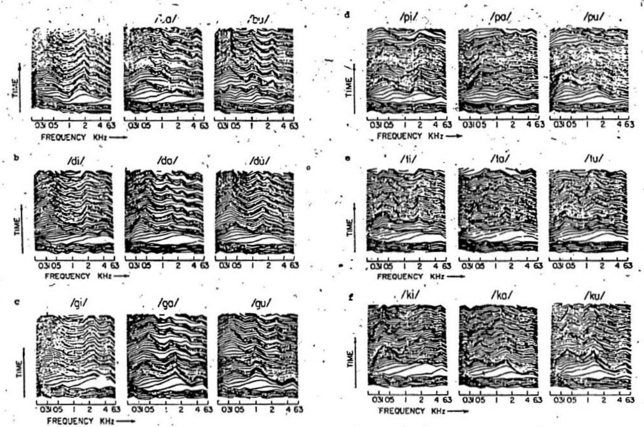


Fig. 2.2 Running spectra for stop consonants, A three-dimensional representation of the acoustical event showing the amplitude as a function of frequency and time (from Searle, 1979).

fricative consonants on a 19 speaker data set and a discrimination accuracy of 74% was achieved (Searle et al., 1979). Analysis of music on auditory perspective was also carried out using the model (Searle,1980). Searle (1982) stated that in terms of speech characteristics, the model should have narrow bandwidth filters below 400 Hz to obtain good frequency resolution and increasingly wider bandwidth filters above 400 Hz to obtain good temporal resolution of the speech signal. Good low frequency spectral resolution is ideal for measuring vowel formants and good high frequency temporal resolution is good for perceiving sharp transients as in stop consonant bursts. The above mentioned work not only substantiated the validity of the design considerations taken into account in the filter detector model of the auditory system but also identified with the properties of human speech.

2.4 DATA REDUCTION

Some of the basic functions that are accomplished by the sensory system in the analysis of speech can be viewed as follows (Searle,1982). The incoming speech signal is converted to electrical or chemical form which must be detected. The detected signal should then be systematically concentrated and some process should take place which would extract only the relevant information and discard the irrelevant. Finally from the relevant information, there must occur some decision process to pick out the salient features necessary for speech recognition.

From the discussion of the Searle's model and from the work done using the model, it is clear that the model preserves the spectral and temporal characteristics of speech that are necessary for recognition. It is perhaps less clear that the input speech signal is more or less the same as the signal summed at the output of the model. In that case some transformation has to take place at the output of the model which should reorganize the data so that the irrelevant information can be discarded.

Yilmaz (1967, 1968) developed a theory of speech perception based on evolutionary adaptive postulates and applied it to whispered speech and to vowel perception. The theory suggested that the general laws of perception apply to the ear as well as to the eye and that there was a structural homology of visual perception and vowel perception.

Yilmaz theory suggested a three dimensional vowel space with loudness, hue and saturation and three primary vowels from which all vowels should be synthesized. In his experimental set up to support his theory, a two dimensional vowel circle was displayed by deflecting the X and Y axis of an oscilloscope with sine and cosine weighted averages of the log spectral magnitude.

$$X = \sum_{n=1}^{N} S(n) \; \mathrm{Sin} \; \frac{2\pi n}{N} \qquad (1)$$

$$Y = \sum_{n=1}^{N} S(n) \; \mathrm{Cos} \; \frac{2\pi n}{N} \qquad (2)$$

S(n) is the log spectral magnitude from the n'th filter and N is the total number of filters in the filter bank. It is evident that by representing speech this way, Yilmaz stated that vowels and vowel-like sounds can be represented by fewer parameters than were used to represent the original spectrum and that for speech recognition, it was enough to pay attention to few weighted averages calculated in equation (1) and (2). Considerable experimental work was done to evaluate various algorithms for vowels and consonants and for inter-speaker variations. Searle (1982) analyzed speech spectra derived from his model using the Yilmaz vowel-circle method.

Richards (1979) resorted to the theory of generalized colourimetry to study various aspects of visual perception. He suggested that all sensory processes including auditory ones should resemble colour perception and that they should be represented in terms of a limited number of primary perceptual channels. The work of Richards supported the Yilmaz theory of speech perception.

From the above discussion it can be seen that the work of Yilmaz and Richards suggests that a smaller number of perceptual channels can be used to represent speech than is generally supposed.

2.5 DATA REDUCTION TECHNIQUES

Efficient and accurate reduction of speech is needed when processing time and memory space are at a premium. Some data reduction techniques are outlined below.

2.5.1 Dimensional Analysis of Speech Spectra

In general, any spectral analysis of speech can be considered in terms of a multi-dimensional concept. The spectral data from the filter bank can be thought of as points in a multi-dimensional space. The measurement of such a space has a dimensionality equal to the number of filters in the bank and the sequence of time samples can be considered as a moving point in space. The outputs of adjacent filters and the adjacent time samples are highly correlated; hence it is expected that all possible speech data points will be located within a restricted portion of the space. There are various possible techniques to diminish the number of dimensions in order to reduce the amount of data. The important requirement is that as much of the original information bearing or liquistic content of the signal as possible should be preserved. Pols et. al., (1972) described a few possible procedures for data reduction, such as analysis of variance, principal component analysis, discriminant analysis and maximally discriminating plane analysis. The type of data reduction technique one should prefer depends on the type of data, the computational limitations and the final goals of the research project.

2.5.2 Principal Component Analysis

The Principal Component method is a statistical procedure for finding an efficient representation of a set of correlated data. From a geometric point of view this procedure can be seen as translating and rotating the coordinate system used to measure the data. The procedure can be considered as deriving an optimal set of orthonormal basis vectors for representing the data. The statistical properties of the spectra can be represented as a convariance matrix [C] with each element given by

$$C_{ij} = \frac{1}{K} \left[\sum_{k=1}^{K} [x_{ki} - \bar{x}_i] [x_{kj} - \bar{x}_j] \right]$$

$$\text{for } i, j = 1, 2, \ldots\ldots n.$$

where K is the no of filters, x_{ki} is the i'th data sample of the k'th filter, $\overline{x}_i$ is the average over K of the i'th data sample and n is the number of data elements in each filter data block. It is known that the eigen-vectors of a convariance matrix, ie., its principal components, are a linear combination of random variables that yield a unique set of statistically independant orthonormal coordinates. Variances accompanying the principal components characterize the statistical properties of the spectral data. Transformation of the coordinates of the original data space into principal components is accomplished by a simple rotation of axes, which preserves the statistical characteristics of the original space. Data reduction is achieved by selecting m components in the eigen-vector space, where m is a number smaller than the number of filters. Furthermore it has been shown that the minimum estimation error is met if the selected components are thr first m components of the eigen-vectors of [C].

The principal component analysis methods was used by many investigators as Pols (1971), Li et.al., (1969), Zohorian and Rothenverg (1981) to study the data reduction in speech processing. Generally they adapted different liguistic methods to recognize the data after data reduction.

The theories proposed by Yilmaz and Richards were experimentally supported by the work of Pols (1972). Pols analyzed twelve Dutch vowels each pronounced by 50 male speakers, using a model of 18 1/3 octave filter bands comparable in bandwidth to the ear's critical band. Since the samples from contiguous frequency bands were not independent, a data reduction was possible. He applied Principal component analysis to the 18 point spectra obtained using the model. For this analysis the variance along each dimension and convariances between dimensions was calculated. The new dimensions obtained were a linear combination of the original dimensions. Detailed analysis verified that the maximum amount of variance was forced into the first component, the maximum of the remaining variance into the second component and so on. The same approach is used for analyzing the conversational speech by Pols (reported by Searle 1982). The log magnitude spectra for a minute of speech were generated and principal component analysis was performed. Fig. 2.3 shows the variance plot of Pols spectral data. It could be seen that the variance is more or less uniformly distributed throughout the filter channels. After subjecting the data to Principal Component Analysis, it was observed that 50% of the variance was in the first component, 29% in the second and the other 21% in decreasing amounts in the remaining 16 components. This indicated that the transform was effective in forcing the information into a few components as the variance plot suggested. In order to test if perceptually relevant information was concentrated in a likely manner or not, Pols constructed a speech synthesizer to reconstruct the transformed data with the higher channels set to zero. The perceptual experiments did not indicate the concentration of the data as the variance plot suggested but it indicated data reduction (as reported by Searle 1982).

Searle (1982) in analyzing the simplified representations of speech subjected the output of his model to principal component analysis. Fig. 2.4 shows the running spectrum representation of the conversational speech (the underlined portion of the sentence 'The watch dog gave a warning growl')

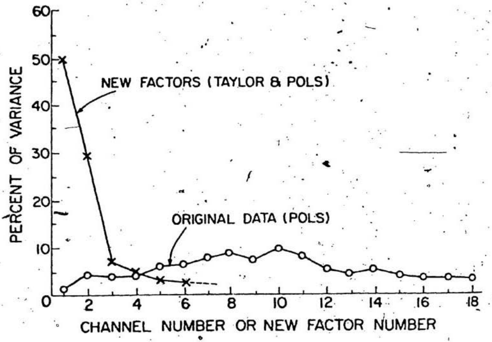


Fig. 2.3 Plot of percent of variance in various components of the original filter-bank and the new representation arising from principal component analysis of Pols spectral data (from Searle, 1982).

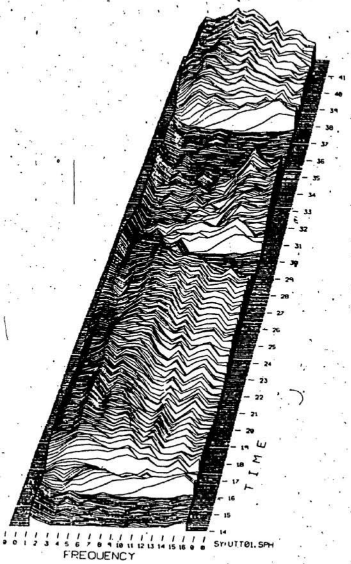


Fig. 2.4 Running spectrum plot of the output of the Searle's model for the underlined portion of the sentence "The Watch Dog gave a warning growl". (from Searle, 1982).

obtained as the output of the filter detector model. Fig. 2.5 shows the first three basis vectors generated by the principal component analysis of the above sentence. These basis vectors resembled the half cosine series. The work of Pols (1977 as reported by Searle, 1982) and Zahorian and Rothenberg (1981) also supported this resemblance. The spectral data in Fig. 2.4 was subjected to Principal Component Analysis and the new representation is shown in Fig. 2.6. The horizontal axis in the plot represents the component number. It is evident from the figure that the spectral data is packed into fewer lower channels. One possible interpretation of the variance plot is the first three components indicate the three functions of loudness, hue and saturation which is the outcome of Yilmaz theory of speech perception.

2.5.3. The Discrete Cosine Transform

Orthogonal transforms of signals have been widely used in the area of digital signal processing and they offer an effective means of data reduction. The Discrete Cosine Transform (DCT) is in the class of orthogonal transforms. For speech signals it has been demonstrated that the DCT is nearly optimal in terms of its performance (Zielinski and Noll, 1977). The DCT is also found to be well suited for speech coding (Flanagan 1979, Tribolet 1979). It has an inverse - the Inverse Discrete Cosine Transform (IDCT) - which permits testing the output of the model for speech intelligibility (the same thing can be accomplished with the principle components calculation by transposition of the principal components matrix).

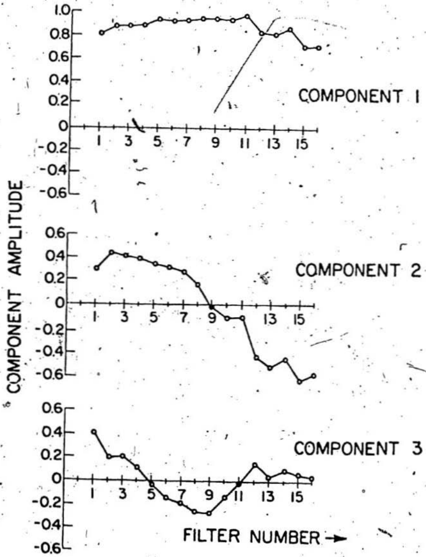


Fig. 2.5 The first three basis vectors generated by principal component analysis for the sentence "The watch dog gave a warning growl". (from Searle, 1982).

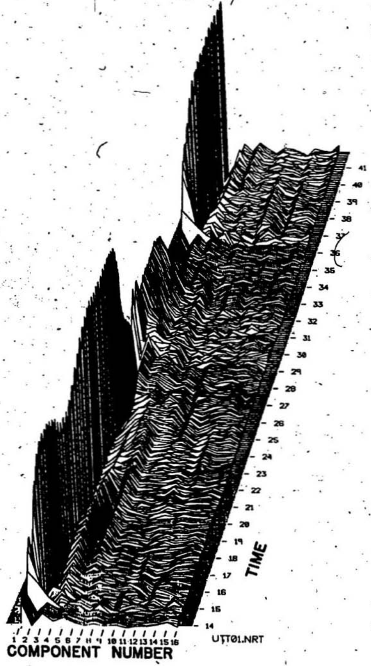


Fig. 2.6 New spectral representation obtained using principal component analysis for the underlined portion of the sentence, "The watch dog gave a warning growl". (from Searle, 1982).

The DCT has a big advantage over principal component analysis in that it is a data independent case opposed to data dependent transform - The DCT also has a fast algorithm, which permits a substantial saving in the computing time; hence the DCT is preferable in terms of practical implementation. Ahmed and Natarajan (1974) have developed an algorithm to compute the DCT using the Fast Fourier Transform. Bertocci et al. (1982) have developed a hardware implementation of the DCT for application to coding of speech. Searle (1982) in his work subjected a sentence of speech to the DCT and showed that the speech data is packed into a fewer channels. It is noted from his work that the basis vectors of Principal component analysis resemble the basis vectors of the DCT. Fig. 2.7 shows the sentence fragment "The watch dog' subjected to the DCT.

In the Searle's model, the linguistic features extracted from the output of the model were used for recognition. In the present work, we are interested in effecting a reconstruction of speech at the output of the model in order that an investigation of reduction in the dimensionality of speech may be carried out. Since the IDCT provides an effective reconstruction from the reduced component set of transformed data, the DCT is used for further analysis.

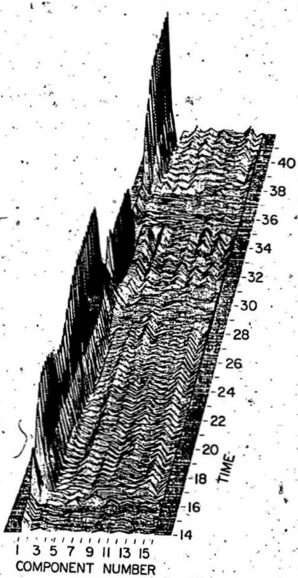


Fig. 2.7 New spectral representation of "The watch dog" derived using Discrete Cosine Transform. (from Searle, 1982).

3. A MATHEMATICAL APPROACH TO THE FILTER DETECTOR BANK USING THE SHORT-TIME FOURIER TRANSFORM

This chapter explains the mathematical approach to the filter detector bank implementation using the short-time Fourier transform as well as the data reduction technique using discrete cosine transform.

3.1. DISCRETE SHORT-TIME FOURIER TRANSFORM

Standard Fourier representations that are appropriate for periodic, transient, or stationary random signals are not directly applicable to the representation of speech signals whose properties change markedly as a function of time. In order to study the temporal and spectral characteristics of speech, the short-time Fourier transform was developed. It is also one way of describing mathematically the filtering process that occurs in the ear.

The short-time Fourier transform (discrete case) is defined as [Rabiner and Schafer, 1978],

$$X_n\,(e^{j\omega_k}) \triangleq \sum_{m=-\infty}^{\infty} x(m)\; p_k\,(n-m)\; e^{-j\omega_k m} \qquad (1)$$

where $p_k\,(n-m)$ is a real window sequence which determines the portion of the input signal that receives emphasis at a particular time index n.

Equation (1) can alternatively be expressed (by transposing m and n-m) as

$$X_n(e^{j\omega_k}) \triangleq e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} x(n-m)\; p_k(m)\; e^{j\omega_k m} \qquad (2)$$

$$= e^{-j\omega_k n}\; \overline{X}_n\,(e^{j\omega_k})$$

where $\overline{X}_n(e^{j\omega_k}) = \sum_{m=-\infty}^{\infty} x(n-m)\ p_k(m)\ e^{j\omega_k m}$ (3)

$X_n(e^{j\omega_k})$ in equation (2) can be interpreted in two different ways: (i) if the index n is assumed to be fixed, $X_n(e^{j\omega_k})$ is simply the normal Fourier transform of the sequence $p_k(n-m)\ x(m)$ and (ii) if ω_k is fixed, then $X_n(e^{j\omega_k})$ is a function of the time index, giving the time dependent or short-time Fourier transform of the sequence $p_k(n-m)\ x(m)$, thus leading to Fourier representation in terms of linear filtering.

3.2 FILTER BANK SUMMATION METHOD OF SHORT-TIME SYNTHESIS

In equation (2), $p_k(m)$ is the window used at frequency ω_k. If we set

$$h_k(m) = p_k(m)\ e^{j\omega_k m} \tag{4}$$

Equation (2) becomes

$$X_n(e^{j\omega_k}) = e^{-j\omega_k n} \sum_{m=-\infty}^{\infty} x(n-m)\ h_k(m) \tag{5}$$

$$= e^{-j\omega_k n}\ [a_n(\omega_k) + ib_n(\omega_k)] \tag{6}$$

Since the window function $p_k(m)$ has the properties of a low pass filter, Equation (5) can be interpreted as a bandpass filter with impulse response $h_k(n)$ followed by modulation with a complex exponential $e^{-j\omega_k n}$.

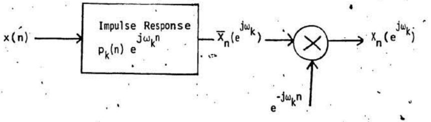


Fig. 3.1 Interpretation of short-time spectral analysis.

If we define $y_k(n) = X_n(e^{j\omega_k})\, e^{j\omega_k n}$ (7)

then from equation (5)

$$y_k(n) = \sum_{m=-\infty}^{\infty} x(n-m)\, h_k(m) \tag{8}$$

ie. $y_k(n)$ is simply the output of a bandpass filter with complex impulse response $h_k(n)$ (Fig. 3.2).

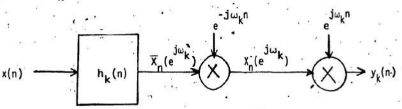


Fig. 3.2. Method of implementing the synthesis of a single channel in terms of linear filtering.

This provides the key to a practical method of reconstructing the input signal from its short-time Fourier transform. In the filter bank analysis of speech, the entire frequency range is divided into N frequency bands. Hence for a N filter bank design, the output is obtained as

$$y(n) = \sum_{k=0}^{N-1} y_k(n)$$

$$= \sum_{k=0}^{N-1} X_n(e^{j\omega_k})\, e^{j\omega_k n} \tag{9}$$

In implementing the filter bank design given by Eqn. (9) the following precautions must be observed (1) to make the filter a lowpass filter with a narrow band, the window function $p_k(m)$ in equations (2) and (4) should appear as an impulse response with respect to $X(e^{j\omega_k})$ having non zero values over a narrow band around zero frequency; and 2 the effective bandwidth of the window should be equal to the bandwidth of the individual filter. Equation 9 is graphically illustrated by Fig. 3.3 where the output signal is obtained as the sum of the signals from each band of the filter bank translated to the original center frequency of the band.

3.3 SEARLE FILTER-DETECTOR MODEL AND SHORT-TIME FOURIER TRANSFORM.

The filter detector band model implemented by Searle is given in Fig. 3.4.

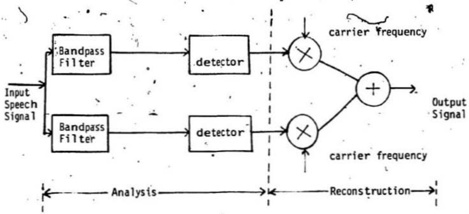


Fig. 3.4 Implementation of the filter detector model.

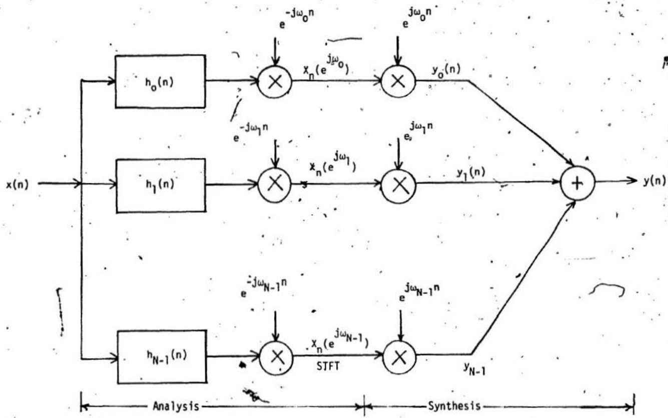


Fig. 3.3 Analysis-Synthesis Method Using Short Time Fourier Transform.

Assuming that the output of the k^{th} filter in the N filter bank model is the real valued sequence $a_n(\omega_k)$, the output of the filter detector model is (Fig. 3.5) given by [Rabiner and Schafer, 1978 and Flanagan, 1972].

$$
\begin{aligned}
y(n) &\triangleq [a_n(\omega_k)^2 + \hat{a}_n(\omega_k)^2]^{1/2} \\
&= [a_n(\omega_k)^2 + b_n(\omega_k)^2]^{1/2} = |X_n(e^{j\omega_k})| \qquad (10)
\end{aligned}
$$

where $b_n(\omega_k)$ is the complex part of the output from the k^{th} filter and $\hat{a}_n(\omega_k)$ is the Hilbert transform of the real output sequence $a_n(\omega_k)$ of the k^{th} filter. Hence the output of the filter bank model after the detector stage is the amplitude of the complex valued sequence represented by

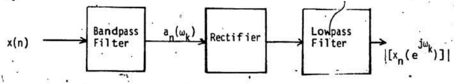


Fig. 3.5 Practical measurement of the short time spectrum by means of the bandpass filter and detector.

$$
\begin{aligned}
y(n) &= [a_n(e^{j\omega_k}) + j\, b_n(e^{j\omega_k})]\, e^{-j\omega_k n} \\
&= X_n(e^{j\omega_k}) \qquad (11)
\end{aligned}
$$

where $X_n(e^{j\omega_k})$ is the short-time Fourier transform of the input sequence $x(n)$. Hence the output of the Searle's filter bank model after the detector stage is the absolute amplitude spectrum of the short-time Fourier transform [Flanagan, 1972]. Consequently the short-time Fourier transform is obtained by the procedure given in equation (11), where $b_n(\omega_k)$ is the Hilbert-transform of the real sequence $a_n(\omega_k)$. Once the short-time Fourier transform is computed, the reconstructed signal $y(n)$ is obtained by the use of equation (9).

In the Searle's model, the bandwidth of the individual filters have been taken as 1/3 octave (variable bandwidth) reflecting the characteristics of the human auditory system. Since it is not possible to achieve both good temporal resolution and frequency resolution, narrow bandwidth filters were used below 400 Hz to achieve good frequency resolution and increasingly wider bandwidth above 400 Hz to achieve good temporal resolution. Even though it is different from some of the conditions imposed on the windows (Section 3.2) these variable bandwidth filters are found to give good results as seen in chapter 2 to characterize the speech. (Searle et. al. 1979, 1982).

3.4 DATA REDUCTION USING THE DISCRETE COSINE TRANSFORM

One of the most important applications of the orthogonal transforms is data compression. The key to securing data compression is signal representation of a given class of signals in an efficient manner. If a discrete signal is comprised of N sampled values, then it can be thought of as being a point in an N dimensional space. Each sample is then a component of the data N vector X which represents the signal in this space. For more efficient representation an orthogonal transform of X is obtained such that $Y = TX$ where Y and T denotes the transform vector and the transform matrix respectively. The main objective behind data reduction is to select a subset of M components of Y where M is substantially less than N. The (N-M) components can then be discarded without introducing objectionable error, when the signal is reconstructed using the retained M components of Y.

3.4.1. Discrete Cosine Transform

The DCT of a sequence $x(m)$, $m = 0, 1, \ldots (m-1)$ is defined as

$$G_x(0) = \frac{\sqrt{2}}{M} \sum_{m=0}^{M-1} x(m)$$

$$G_x(k) = \frac{2}{M} \sum_{m=0}^{M-1} x(m) \cos \frac{(2m+1)k\pi}{2m} \quad k = 1, 2, \ldots (m-1) \qquad (12)$$

where $G_x(k)$ is the k^{th} DCT coefficient.

Inverse discrete cosine transform is defined as

$$x(m) = \frac{1}{\sqrt{2}} G_x(0) + \sum_{k=1}^{M-1} G_x(k) \cos \frac{(2m+1)}{2m} k\pi . \qquad (13)$$

This can be written in the matrix form and Λ is the (M x M) matrix that denotes the cosine transformation, then the orthogonal property can be expressed as

$$\Lambda^T \Lambda = \frac{M}{2} [I] \qquad (14)$$

where Λ is the transpose of Λ^T and I is the (M x M) identity matrix.

4. SIGNAL ANALYSIS USING A DIGITAL IMPLEMENTATION OF THE SEARLE'S MODEL

This chapter describes the various design considerations taken into account in the design of the software model. Fig. 4.1 shows the block diagram describing the work done.

The digitized speech is passed through a filter bank, which simulates the frequency analysis of the basilar.membrane, and then through a detector bank which simulates the detection function of the hair cells. The output of the filter detectors is used to create a running spectral plot which reflects the temporal and spectral characteristics of speech. These spectra are then subjected to a linear transform - Discrete Cosine Transform - to concentrate the data to a minimum number of perceptually important channels. The Inverse Discrete Cosine Transform is done on a reduced number of channels to obtain the output spectra. The output is modulated and bandpass filtered to obtain the signal for perceptual testing of the intelligibility of speech.

4.1 DESIGN OF THE DIGITAL BANDPASS FILTERS IN THE BANK

The design considerations that were taken into account by Searle in developing the hardware model of the auditory periphery are considered in the implementation of the digital software model. The physical arrangement of the filter bank provides 18 second order Butterworth bandpass filters. The filters are logarithmically spaced with a bandwidth of 1/3 octave. The 3 db upper and lower cutoff frequencies are calculated using the relation $F_h = 1.26\ F_\ell$. With 125 Hz as the cutoff frequency of the lower channel and using the above relationship the higher and the lower cutoff frequencies of the 18 bandpass filters in the channel are calculated.

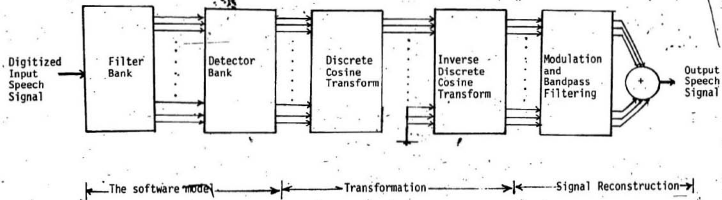


Fig. 4.1 Digital Implementation of the Searle Model and Signal Processing

4.1.1 Digital Bandpass Filter Design

Butterworth filters which have maximally flat frequency response are used in the design of the digital filters. The digital bandpass filters are derived from second order Butterworth analog filters by the bilinear transform method. Three such filters are cascaded together to implement a sixth order bandpass filter using the design equations of Ahmed and Natarajan (1983). For speech processing, the cascade form is preferred as it exhibits superior performance with respect to round-off noise, inaccuracies and stability. In addition cascade realizations are preferred for hardware implementation since they require shorter word length and are less prone to overflow problems.

The transfer function of the k'th stage second order Butterworth bandpass filter is given as $R_{2K}(S)$

$$R_{2k}(S) = \frac{W^2 s^2}{s^4 - 2W \cos\theta_k s^3 + (2\omega_o^2 + W^2)s^2 - 2W\omega_o^2 \cos\theta_k s + \omega_o^4}$$

$$\text{where } \omega_{A_i} = \tan(\omega_{D_i} T/2), \; i = 1,2$$

$$W = \omega_{A_2} - \omega_{A_1}$$

ω_{A_i} is the analog frequency variable corresponding to the critical digital frequencies ω_{D_i}

$$\omega_o^2 = \omega_{A_1} \cdot \omega_{A_2}$$

$$\theta_k = \frac{(2k + n - 1)\pi}{2n}$$

where T is the sampling frequency, n is the number of cascaded stages.

$$H_k(z) = R_{2k}(s) \Big|_{s = (z-1)/(z+1)}$$

$$= \frac{(1 - 2z^{-2} + z^{-4})\, W^2}{D_k + E_k z^{-1} + F_k z^{-2} + G_k z^{-3} + J_k z^{-4}}$$

where $D_k = (1 + \omega_0^2)\,[1 - 2W \cos\theta_k + \omega_0^2] + W^2$

$$E_k = 4(\omega_0^2 - 1)\,[1 - W \cos\theta_k + \omega_0^2]$$

$$F_k = 6\,\omega_0^4 - 4\,\omega_0^2 + (6 - 2W^2)$$

$$G_k = 4(\omega_0^2 - 1)\,[1 + W \cos\theta_k + \omega_0^2]$$

$$J_k = (1 + \omega_0^2)\,[1 + 2W \cos\theta_k + \omega_0^2] + W^2$$

Using these equations the bandpass filter coefficients can be found for all the stages. The output of the k'th stage is obtained by

$$Y_k(z) = \frac{A_k + B_k z^{-2} + C_k z^{-4}}{D_k + E_k z^{-1} + F_k z^{-2} + G_k z^{-3} + J_k z^{-4}}$$

where A_k, B_k and C_k are the numerator coefficients and D_k, E_k, F_k, G_k and J_k are the denominator coefficients which are calculated using the above mentioned equations. The input-output equations of the filter can be written as

$$y_k(I) = \frac{1}{D_k}\,[A_k\,x_k(I) + B_k\,x_k(I-2) + C_k\,x_k\,(I-4)$$

$$- E_k\,y_k(I-1) - F_k\,y_k(I-2) - G_k\,y_k(I-3) - J_k\,y_k(I-4)]$$

where k denotes the number of stages x(I-2), x(I-4) are the past input values and y(I-1), y(I-2), y(I-3), y(I-4) are past output values.

The bandpass filter algorithms are written in HP BASIC and are implemented using the HP 5451B Fourier Analyzer System.

In order to obtain the impulse response of the bandpass filter, a unit impulse input data block is created and used as the input to the input/output equation. Fig. 4.2 shows the frequency response plot of the bandpass filter in channel 4. It can be seen from the plot that the 3 db cutoff frequencies are 396 Hz and 500 Hz for the particular channel. The digital filters are implemented using convolution approach. The values of the impulse response h(n) are stored in the system memory and as the samples of the input signal x(m) enters the system, the operation of $y(n) = \sum_{m=0}^{n} x(m)h(n-m)$ is performed where y(n) is the output sequence.

A similar approach is followed in implementing the 18 bandpass filters in the channel. Fig. 4.3 shows the log magnitude vs. log frequency plot of the filters in the filter bank. It can be seen from the magnitude plot that the impulse response is maximally flat with a slope of 120 db/octave.

The two filters with f_c = 141 and 178 Hz are combined as are the two filters with f_c = 224 and 282 Hz, reducing the system to a set of 16 filters. These filters in the low frequency range are combined so that the bandwidths used are comparable with the ear's critical bandwidth.

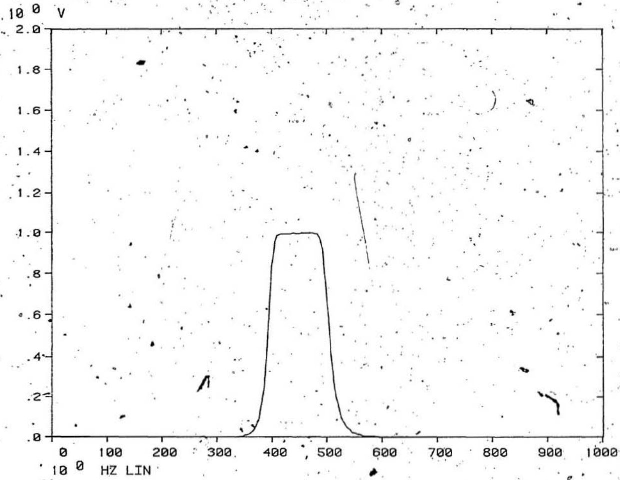


Fig. 4.2 Frequency response of the bandpass filter in channel 4.

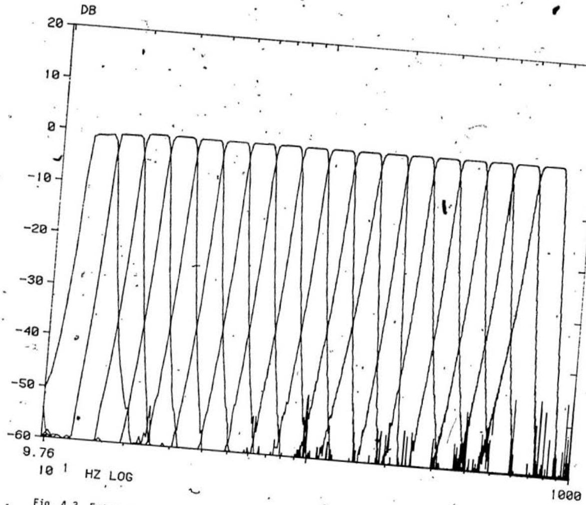


Fig. 4.3 Frequency response curves for 18, 1/3 octave bandpass filters in the filter bank.

The combination also provides a good frequency resolution in measuring the vowel formants. Fig. 4.4 shows log magnitude vs log frequency plot of the 16 filters in the filter bank.

4.2 DESIGN OF THE ENVELOPE DETECTOR

The envelope detector in the filter bank simulates the rectification detection function of the hair cells. It consists of the rectifier followed by a low pass filter. The rectifier is implemented by finding the absolute magnitude of the output of the bandpass filter.

4.2.1 The Design of the Lowpass Filter in the Detector

The detectors in the model are designed such that the time constants are optimized for the individual filters. Filters with broader bandwidths have faster rise times; detector time constants are chosen to be commensurate with filter rise times. According to Searle's definition of process domain, the rise time for the 1 KHz envelope detector is chosen to be 5 millisec and the other detectors have appropriately scaled rise times (i.e. proportional to the channel center frequency). The rise time in millisec for each channel = $5.9/f_0$ where f_0 is the channel center frequency in KHz. Cutoff frequencies of the lowpass filters are calculated as follows.

Rise time of the 1 KHz channel	= 5 millisec
Center frequency of the 1 KHz channel	= 897.3295 Hz.
Rise time - frequency product (based on 1 KHz channel) is	= 4.4866476

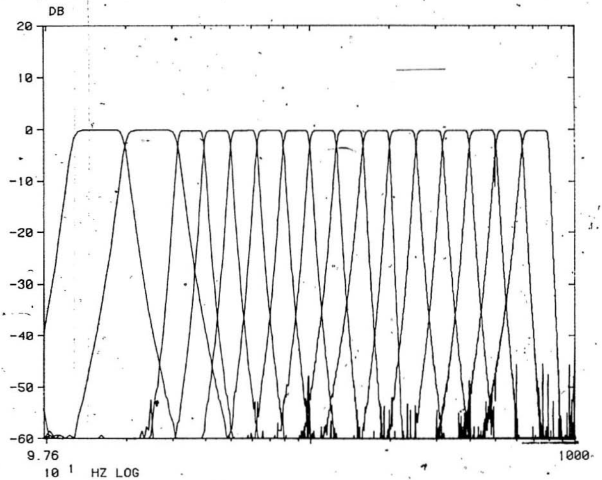


Fig. 4.4 Frequency response curves for 16, 1/3 octave bandpass filters in the filter bank (critical bandwidth)

Rise times of each envelope detector in the channel $= 4.4866476 * 1/f_c$ where f_c is the channel center frequency of the particular bandpass filter. Using this rise time the lowpass cutoff frequency of the single stage R-C lowpass filter is calculated.

Rise time of the single stage R-C lowpass filter $= 2.2/\omega_c$

$= 0.35/f_c$

where f_c is the lowpass filter cutoff frequency. Table 4.1 gives the rise time for each channel and the cutoff frequency of the lowpass filter in the channel.

The digital lowpass filter is realized using the bilinear transform technique.

The transfer function of the lowpass filter is

$$H(z) = \frac{A\,(1 + z^{-1})}{1 + B\,z^{-1}}$$

where $A = \omega_{A_1}/(\omega_{A_1} + 1)$

$B = (\omega_{A_1} - 1)/(\omega_{A_1} + 1)$

$\omega_{A_1} = \tan \pi T f_c$

Also f_c is the cut off frequency of the lowpass filter and T the sampling interval. If x(n) and y(n) are the input and the output, the following difference equation is satisfied.

$$y(n) = Ax(n) + Ax(n-1) - By(n-1)$$

Channel Number	Frequency Band of the filters in the channel (Hz)	Center frequency of the channel (Hz)	Rise time of the envelope detector (MSecs)	Low pass filter cut off frequency (Hz)
1	125 - 157.5	141.25	31.720599	11.03384
2	157.5 - 198.45	177.975	25.209576	13.883613
3	198.45 - 250.047	224.2485	20.005962	17.494785
4	250.047 - 315.059	282.553	15.879143	22.041492
5	315.059 - 396.975	356.017	12.602544	27.77217
6	396.975 - 500.188	448.5815	9.9734587	35.093142
7	500.188 - 630.237	565.2125	7.9377769	44.09295
8	630.237 - 794.099	712.168	6.3001506	55.554228
9	794.099 - 1000.56	897.3295	5.000001	69.999986
10	1000.56 - 1260.71	1130.635	3.9684398	88.19587
11	1260.71 - 1588.5	1424.605	3.149178	111.14011
12	1588.5 - 2001.5	1795	2.4995114	140.02737
13	2001.5 - 2521.9	2261.7	1.9835469	176.45159
14	2521.9 - 3177.59	2849.745	1.5743646	222.31191
15	3177.59 - 4003.76	3590.675	1.2495314	280.10501
16	4003.76 - 5044.74	4524.25	0.9915491	352.98302
17	5044.74 - 6356.57	5700.655	0.7869579	444.7506
18	6336.57 - 8009.03	7182.7	0.6245413	560.4113

Table 4.1 Cut off frequencies of the bandpass filter and the lowpass filter (detector stage) in the channel.

The impulse response is obtained by implementing the above equation. The procedure used for the implementation of the bandpass filter is followed for the lowpass filter also. The impulse response of the lowpass filter is obtained and normalized. Fig. 4.5 shows the magnitude response of the lowpass filter in channel 4 and the lowpass filter cutoff frequency can be seen to be 35 Hz.

The overall rise time of the channel is $T_o^2 = T_d^2 + T_n^2$ where T_n is the rise time of the bandpass filter and T_d is the rise time of the detector. In the Searle's model the overall rise time of the channel is $5.9/f_c$ where f_c is the center frequency of each bandpass filter in KHz. Table 4.2 gives the values calculated using the above mentioned relationship and the determined values. There is a small variation in the calculated values and the determined values.

4.3 IMPLEMENTATION OF THE SEARLE'S MODEL

The bandpass and lowpass filter impulse responses are normalized and stored. The filter detector model is implemented in the HP 5451B Fourier Analyzer System using the Fourier Keyboard Programmes. Fig. 4.6 shows the block diagram of the digital implementation of the filter detector model.

The time domain implementation of the model is realized by convolving the digitized speech block with the bandpass filter impulse response. Detection is done by taking the absolute magnitude of the filter output and by convolving the rectified signal with the lowpass filter impulse response. This operation gives the envelope of the input speech signal. The output is stored as data files in the magnetic tapes for further processing.

Channel Number	Filter detector rise time according to the relationship Rise time = $5.9/f_c$ (mSecs)	Determined values of the filter detector rise time mSecs.
1	41.769912	44.831415
2	33.150723	33.786963
3	26.310098	26.959011
4	20.881038	21.508386
5	16.572242	17.147597
6	13.152571	13.661293
7	10.438552	10.873676
8	8.2845621	8.6570073
9	6.5750652	6.930392
10	5.2183065	5.4370148
11	4.1414989	4.3160872
12	3.2869081	3.455862
13	2.6086572	2.7419348
14	2.0703607	2.2179536
15	1.6431451	1.7582692
16	1.3040835	1.1094857
17	1.0349688	1.0994682
18	0.8214181	0.835001

Table 4.2 Comparison of the filter detector rise times.

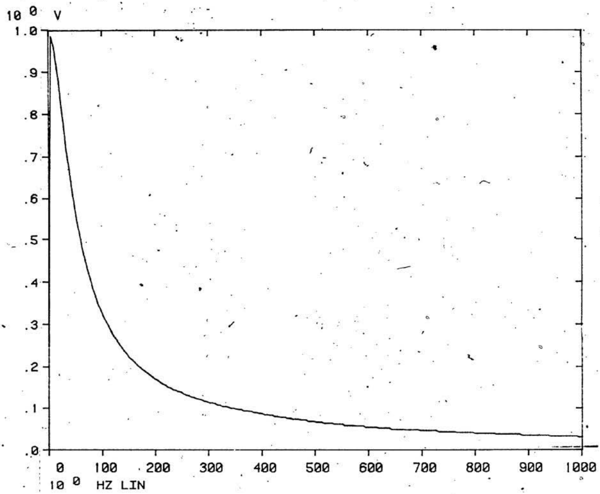


Fig. 4.5 Frequency response of the lowpass filter (detector stage) in channel 4.

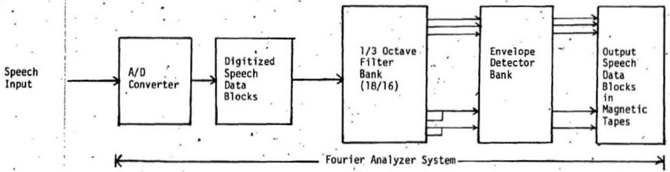


Fig. 4.6 Digital Implementation of the Searle Model.

4.3.1 Data Acquisition

The sentence spoken for analysis is:
'THE WATCH DOG GAVE A WARNING GROWL'.

Speech is recorded and played using a HP 3968A instrumentation tape recorder. The continuous speech is recorded with a normal tape speed of 7 1/2 ips, and played back at a reduced speed of 1 7/8 ips. The speech is band limited at playback using a lowpass filter with a cutoff frequency of 2 KHz.

Analog to Digital Conversion Particulars:

Input Fourier Block size: 2048

Sampling frequency : 5 KHz for the reduced speed of 1 7/8 ips. (20 KHz effective)

Since the impulse response of the filters use a block size of 4096, and to eliminate the error that occurs due to convolution the speech data blocks are converted to a block size of 4096 by overlapping.

The digitized speech data block is analyzed using the filter detector model. Fig. 4.7 shows the implementation of a single channel in the model. Fig. 4.8 shows a block of speech and its magnitude spectrum. Fig. 4.9 shows the bandpass filtered output through channel 4. The magnitude spectrum which indicates the particular band of signal can be seen from the figure. The rectified output of the bandpass filter and its magnitude spectrum is shown in Fig. 4.10. It is seen that the rectification results in a dc value, low frequency components up to approximately 80 Hz and the higher frequency components in the range of 800 Hz to 1000 Hz. These additional higher frequency components were filtered out using the lowpass filter and the lowpass filter output and the magnitude spectrum

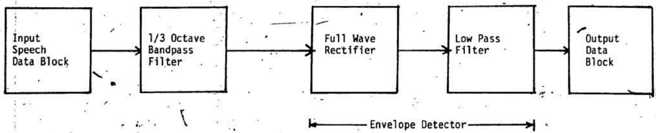


Fig. 4.7 Implementation of a single channel in the filter detector bank.

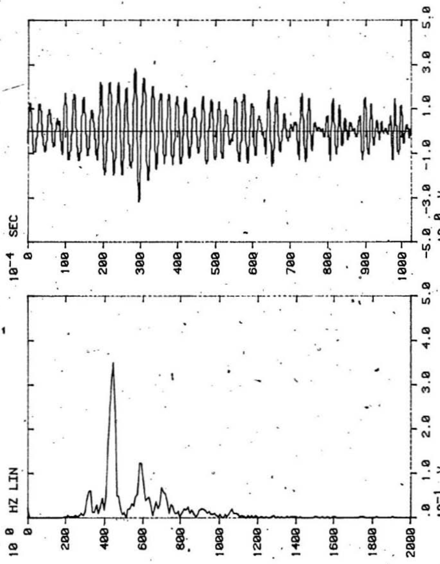


Fig. 4.8 (a) Input speech signal

(b) Magnitude spectrum

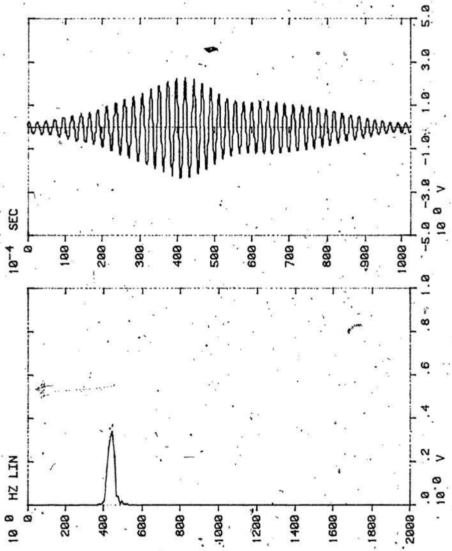


Fig. 4.9 (a) Bandpass filter output (channel 4)

(b) Magnitude spectrum

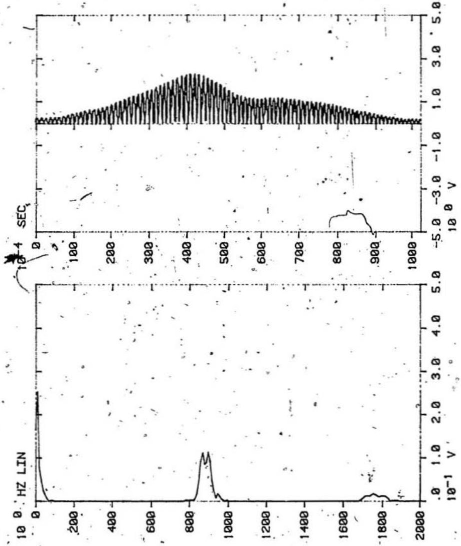


Fig. 4.10 (a) Rectifier output of the detector stage (channel 4)

(b) Magnitude spectrum

is shown in Fig. 4.11. From the figure it is evident that the output of the filter detector provides the envelope information. The rectification also results in the input signal and higher order terms. The higher order terms are filtered in the detector as explained.

In order to test if all the information in the input speech signal is preserved or not the output of the filter bank is summed together and converted to an analog signal using the digital to analog converter. Fig. 4.12 shows the filter bank summation output and the magnitude spectrum. Comparing with the input signal in Fig. 4.8, it can be seen that there is a slight distortion in the signal but the magnitude remains to be the same.

4.4 SPEECH ANALYSIS USING RUNNING SPECTRUM PLOT

In order to visually compare the amplitudes of different channels at a given time and to view the spectral and temporal characteristics of speech, the output of the filter detector system is plotted as a succession of spectra, that is magnitude vs. frequency with time as parameter. Fig. 4.13 and Fig. 4.14 show the running spectrum plot of the underlined portion of the sentence 'The watch dog gave a warning growl'.

In plotting the running spectrum the log amplitude was considered. Since the spectral and temporal characteristics were not clearly visible from this running spectrum, the linear magnitude spectrum is plotted for further analysis. Fig. 4.15 and 4.16 shows the running spectrum of the log magnitude data plotted which corresponds to the linear data plotting in Fig. 4.13 and Fig. 4.14.

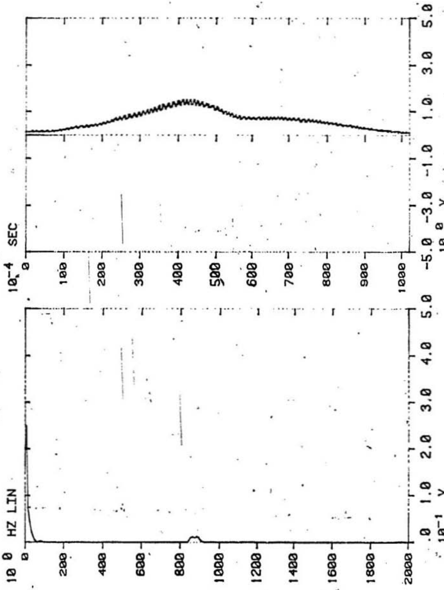


Fig. 4.11 (a) Filter-detector output (channel 4)

(b) Magnitude spectrum

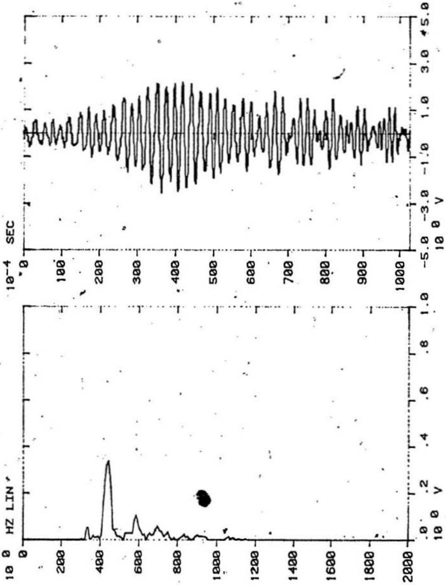


Fig. 4.12 (a) Filter bank summation output

(b) Magnitude spectrum

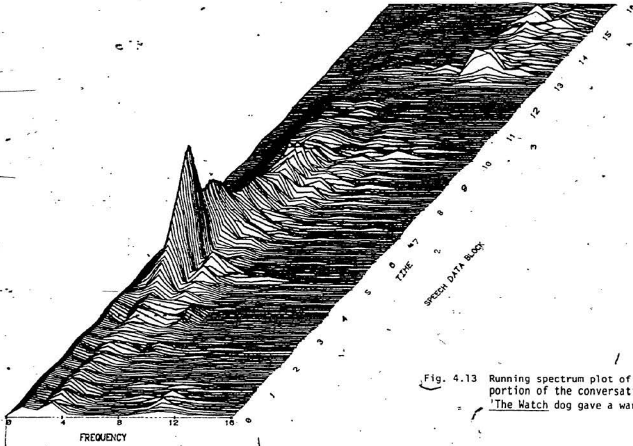


Fig. 4.13 Running spectrum plot of the underlined portion of the conversational speech 'The Watch dog gave a warning growl'.

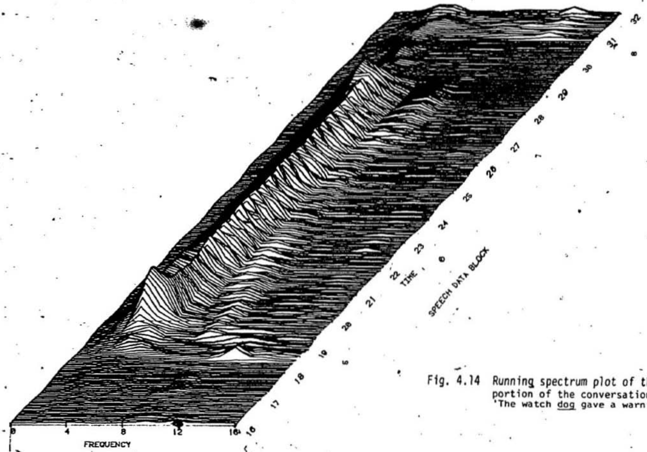


Fig. 4.14 Running spectrum plot of the underlined portion of the conversational speech 'The watch dog gave a warning growl'.

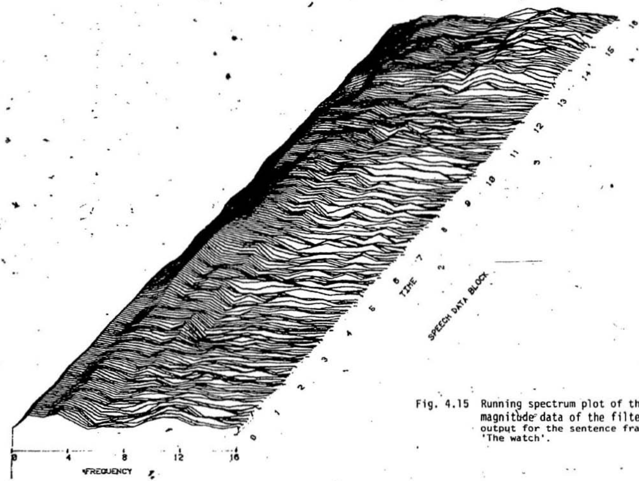


Fig. 4.15 Running spectrum plot of the log magnitude data of the filter detector output for the sentence fragment 'The watch'.

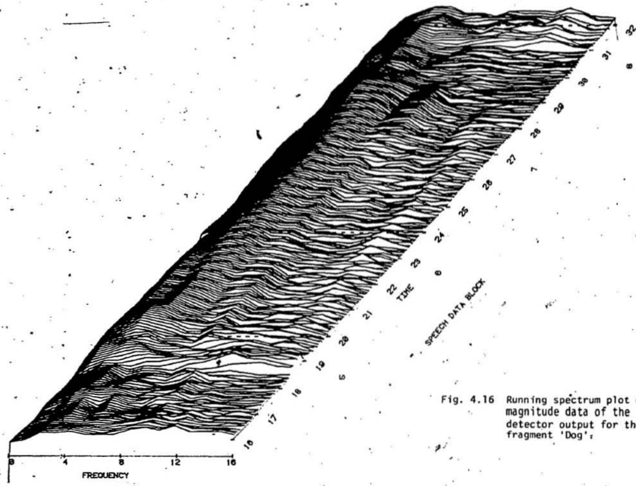
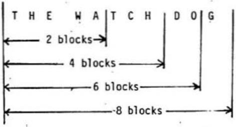


Fig. 4.16 Running spectrum plot of the log magnitude data of the filter detector output for the sentence fragment 'Dog'.

The amplitude of the output from the 16 channels at a given instant in time is plotted. This gives a cross section at that time of 16 discrete values. These 16 discrete values are joined with straight lines to obtain the short-time amplitude spectrum with 1/3 octave frequency resolution. This is repeated for successive instants in time one above the other on the page and it emphasizes a clear observation of the changes in time of the frequency spectrum. The horizontal axis shows the 1/3 octave frequency scale in steps of 4 channels (4,1/3 octave channels) between divisions. Frequency increases from left to right with the first four channels indicate frequency range from 125 Hz to 500 Hz, the second four channels from 500 Hz to 1260 Hz, the third four channels from 1260 Hz to 3177 Hz and the last four channels from 3177 Hz to 8000 Hz. The vertical displacement upwards indicates the progression in time. The time resolution between the spectra is 1.6 millisec. Time averaging can also be performed. The vertical axis upwards shows the linear amplitude. The plots are tapered down in width from top to bottom to enhance the perception of depth. This gives a three dimensional perspective to the plot.

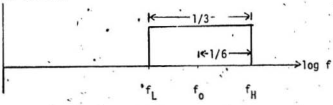
Each time spectra represents 1.6 millisec and each time marker on the right represents a block of 16 spectra giving a time scale of approximately 25.6 millisecs per division. The underlined portion of the continuous speech "The Watch Dog gave a warning growl" is plotted. The sentence fragment consists of 8 blocks of speech data equalling 256 blocks of filter detector output data. The block numbers are marked on the right. The block numbers are categorized as follows:



The corresponding blocks are reconstructed for audibility in steps of first 2, 4, 6 and 8 blocks of speech. Each plot has 4 block markers with 64 spectral lines per block each representing 0.1024 secs of speech. The first plot has 0.4096 secs of speech and the second plot has 0.4096 secs of speech which agrees with the time taken (0.8 secs) to speak the sentence in real time. The spectral and temporal characteristics of speech can be seen from the running spectrum plot and the linguistic categorization of the running spectrum plot is explained in the next chapter.

4.5 PROPOSED CHANGES IN THE MODEL

The output of the filter detector can be thought of as a lowpass representation of the bandpass signal. One of the important aspects of choosing the lowpass cut off frequency apart from the rise time consideration is that the lowpass filter should preserve all the information that is passed by the bandpass filter. In this regard the amount of information that comes out of the 1/3 octave bandpass filter can be calculated as follows.



$$\text{Log } f_h - \text{Log } f_c = \text{Log } f_c - \text{Log } f_\ell = 1/6 \text{ Log } 2$$

From whence the bandwidth is

$$f_h - f_\ell = 2^{1/6} f_c - f_c/2^{1/6}$$

$$= 0.232\ f_c$$

where f_c is the center frequency of the bandpass filter.

The above discussion implies that he cutoff frequency of the lowpass filter has to be 0.232 times the center frequency of the bandpass filter to retain all its information. In the filter detector model implemented, for the 1000 Hz channel the center frequency is 897.331 Hz and the lowpass cut off frequency is 70.0 Hz. The ratio of the lowpass cutoff frequency to the bandpass filter center frequency is 0.078. This ratio indicates that there could be an information reduction of as much as a factor of 3 (as compared to 0.232). To preserve the information potential of the signal in each channel, the lowpass filter cutoff frequency is increased to three times the original design and the filter detector model is implemented. Fig. 4.17 shows the output of the detector stage for the modified channel 4. The running spectrum of the filter detector output is also plotted. Fig. 4.18 shows the plot corresponding to the one shown in Figure 4.13 from which the slight enhancement of the spectral energy at the peaks can be visualized.

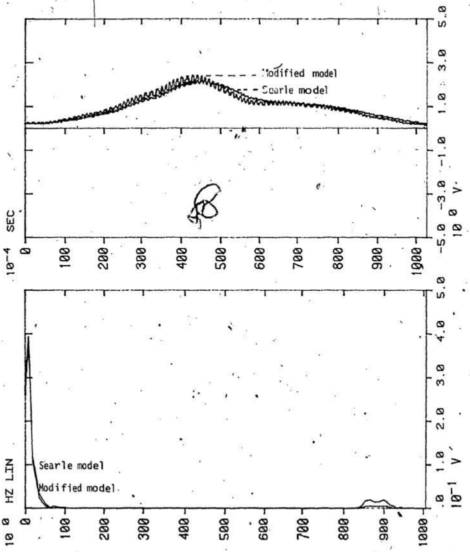


Fig. 4.17 Comparison of implementation of detector stage according to Searle and modified design considerations.

a) Filter-detector output
b) Magnitude spectrum

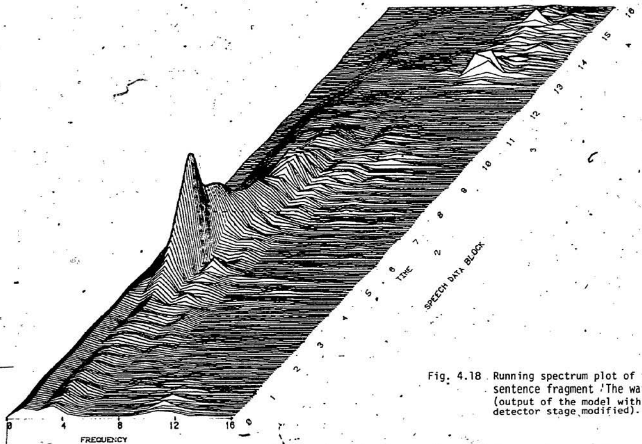


Fig. 4.18 Running spectrum plot of the sentence fragment 'The watch' (output of the model with the detector stage modified).

4.6 SIGNAL RECONSTRUCTION FOR PERCEPTUAL TESTING

In the time domain implementation of the model, the output of the filter detector provides the envelope information of the speech signal. The envelope is used to reconstruct the speech signal for perceptual testing.

The bandpass spectrum of the output of the filters in the bank is rectified and lowpass filtered in the implementation of the model. The detector stage translates the frequency content of the bandpass signal to zero frequency with loss in information which is discussed in the next chapter. The output of the detector can be thought of as a lowpass representation of a bandpass signal. Hence in order to reconstruct the speech from the detector output, the detector output is to be amplitude modulated with a carrier frequency by which the frequency contents can be shifted to the band centered around the carrier frequency. The choice of the carrier frequency is an important factor as it determines the information preserved in the band. In order to investigate the choice of the carrier frequency in the reconstruction procedure, the idealized case is considered. The reconstruction procedure for the idealized case is shown in Fig. 4.19. Fig. 4.19 (a) shows the bandpass spectrum of the signal. Fig. 4.19 (b) shows the bandpass spectrum translated to the zero frequency. Fig. 4.19 (c) shows the modulation using the center frequency of the channel and Fig. 4.19 (d) shows the modulation using the lower cut-off frequency of the channel. In order to get rid of the irrelevant information, the bandpass filter in the channel is used. The frequency response of the bandpass filter to extract the correct data is shown in

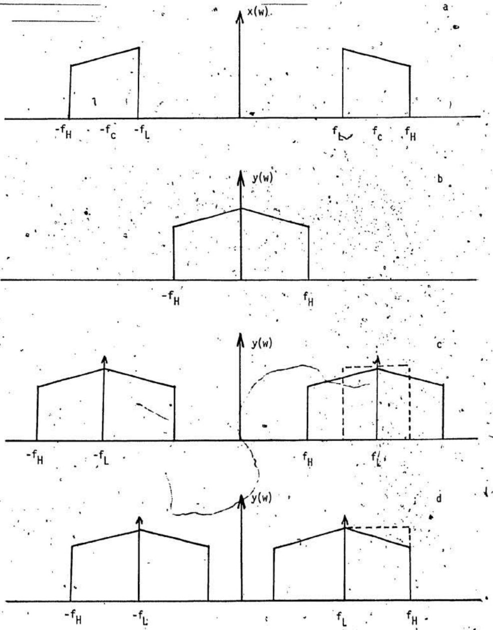


Fig. 4.19 Reconstruction procedure for idealized case.
(a) The bandpass signal (b) the lowpass equivalent
(c) Modulating with the center frequency of the channel
(d) Modulating with the lower cut off frequency of the channel.

dotted lines. From the figure it is evident that for the reconstruction procedure, the choice of the lower cut-off frequency as the carrier provides the maximum relevant information. This approach is followed in the reconstruction of the signal. The block diagram in Fig. 4.20 shows the implementation of the reconstruction procedure. The filter detector output is amplitude modulated with the lower cut-off frequency as the carrier frequency and the modulated output is bandpass filtered with the bandpass filter in each channel. The outputs are summed together and the sum is converted to an analog signal using the digital to analog converter. The converted signal is taped using the instrumentation tape recorder. The implementation of the reconstruction procedure is done using the Fourier Analyzer System. The results of reconstruction are explained in the next chapter.

4.7 SIGNAL PROCESSING - DATA REDUCTION USING DISCRETE COSINE TRANSFORM

Having outlined a method of reconstructing the signal from the output of the filter detector model, a study to investigate the reduction in the dimensionality of speech is to be carried out. The Discrete Cosine Transform (DCT) is used to study the reduction in the dimensionality representation of speech at the output of the model. The log magnitude of the filter detector output is subjected to DCT. Fig. 4.21 shows the output of the filter detector in Fig. 4.18 subjected to DCT. It can be seen from the plot that the DCT forces the data into fewer channels than the original representation. The main objective of the study is to

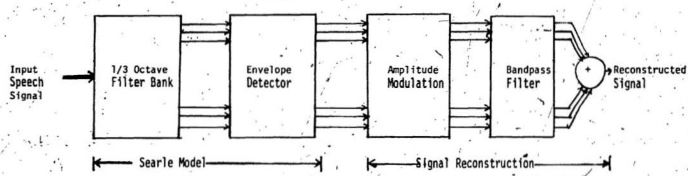


Fig. 4.20 Implementation of signal reconstruction for perceptual testing.

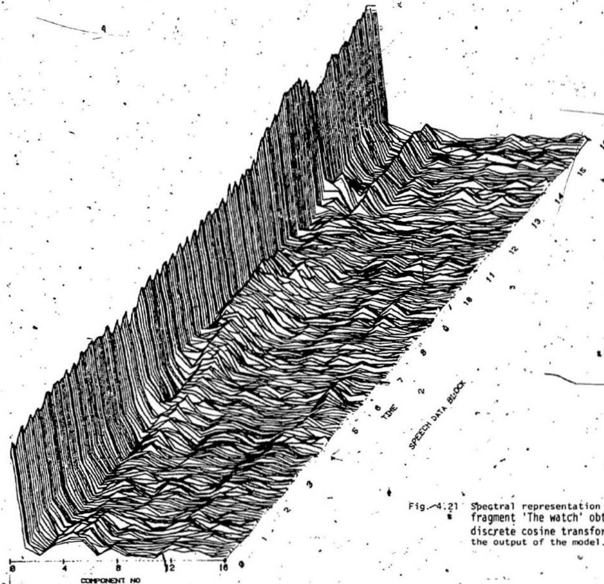


Fig. 4.21 Spectral representation of the sentence fragment 'The watch' obtained using discrete cosine transform (DCT) at the output of the model.

investigate if perceptually relevant information is concentrated in a likely manner so that an effective data reduction can be attempted by setting the higher channels of the transformed data to zero and reconstructing the signal with the minimum number of channels. The lower channels are set to zero and with the data in the remaining channels, the signal is reconstructed using the Inverse Discrete Cosine Transform (IDCT). Fig. 4.22 shows the output of the model subjected to DCT and 6 channels set to zero. Fig. 4.23 shows the signals reconstructed from the transformed data in Fig. 4.21 using the IDCT. The spectral and temporal degradation of the signal as the number of channels cut out increases are studied. The implementation of the data reduction technique is carried out in PDP 11/60 computer. The results of this signal processing are explained in detail in the next chapter.

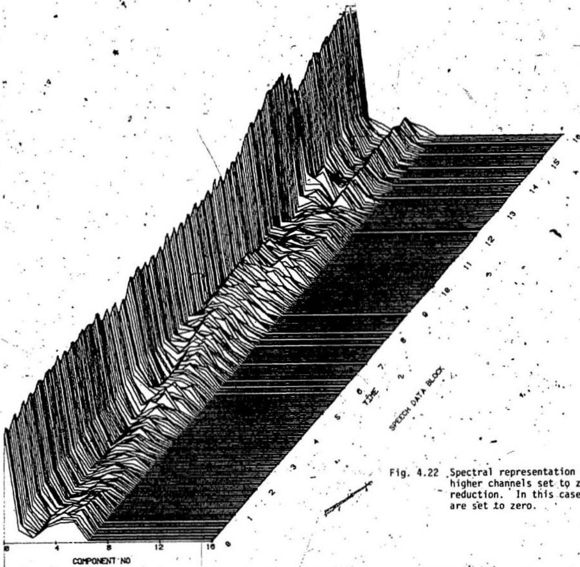


Fig. 4.22 Spectral representation showing the higher channels set to zero for data reduction. In this case 10 channels are set to zero.

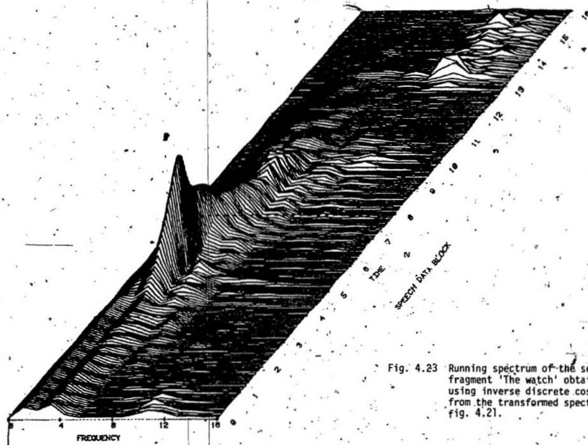


Fig. 4.23 Running spectrum of the sentence fragment 'The watch' obtained using inverse discrete cosine transform from the transformed spectrum in fig. 4.21.

5. RESULTS AND DISCUSSION

This chapter explains the various procedures followed to test the digital model and discusses the results obtained at the output of the model.

5.1 DESIGN CONSIDERATIONS IN THE FILTER DETECTOR MODEL

One of the main objectives of the work is to study an effective reconstruction of speech from the output of the filter detector model for intelligibility testing. Hence the information preserving capacity of the model is studied.

The detector that follows the filter consists of full wave rectifier followed by lowpass filter. It was argued in the previous chapter that in order to preserve the information content passed by the band pass filter, the low pass filter in the detector stage should have a cutoff frequency of three times the cutoff frequency proposed by Searle's model and based on this argument the cutoff frequencies of the low pass filter are changed for further analysis. In order to study the information preserving capacity of the detector stage, a deliberatly asymmetric ramp signal with a frequency spectrum of 0 to 1000 Hz is generated (Fig. 5.1a) and is applied as input to the fourth channel of the filter detector bank. Fig. 5.1(b) indicates the particular frequency range selected by the band pass filter in the fourth channel. The output of the bandpass filter is shifted to the origin using amplitude modulation with 397 Hz carrier frequency (the lower cutoff frequency of the channel). The modulation results in (Fig. 5.2) a lower frequency band whose bandwidth is approximately equal to the bandwidth of the channel (104 Hz) and higher frequency components in the range of 800 Hz to 900 Hz. The spike at the origin in Fig. 5.2 results from the overlap of positive and

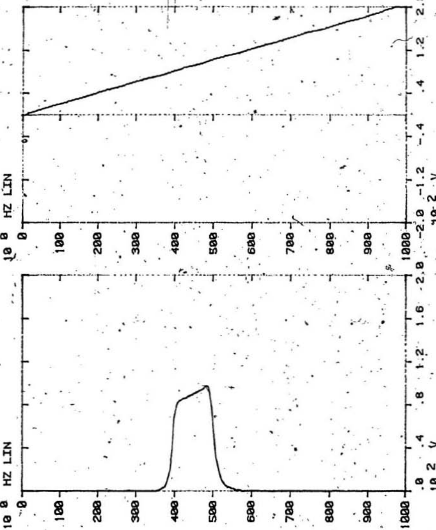


Fig. 5.1 (a) Ramp signal to test the model.
(b) Bandpass filter output in channel 4 to the ramp input.

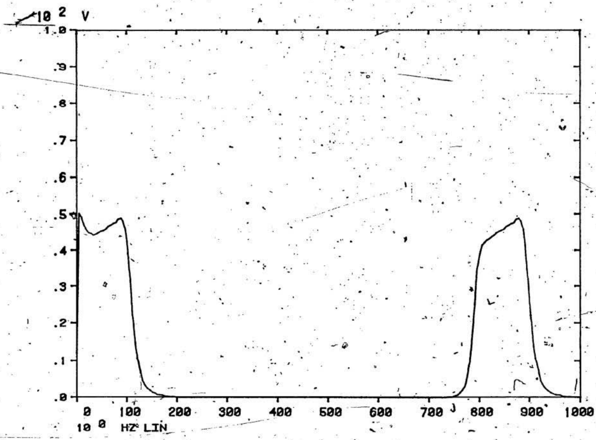


Fig. 5.2 Lowpass signal representation of the bandpass signal obtained by modulation.

negative halves of the spectrum as the carrier frequency chosen for modulation is the 3 db lower cutoff frequency. Using this a low pass representation of a bandpass signal is obtained. The lowpass filter has a cutoff frequency of 104 Hz which is approximately equal to the bandwidth of the translated signal. Fig. 5.3 shows the comparison between the lowpass filter responses for channel 4 due to Searle's model and the modified model (lowpass filter cut off frequencies are 35.093 Hz and 105.2794 Hz). In Fig. 5.4 the outputs indicating the effect of lowpass filtering of the frequency translated input signal are compared. It is noted from the figures that even though the bandwidth is preserved there is a spectral degradation in the shifted spectral shape.

The output of the bandpass filter in Fig. 5.1(b) is rectified. The rectified output is lowpass filtered using Searle's design and the modified design (Fig. 5.5). Fig. 5.5 indicates negligible enhancement in the spectral information in the modified design. It is evident that information loss occurs in the detector stage primarily due to rectification.

In the digital implementation of the model the nonlinear rectification done using the diode in the detector stage of the analog model is simulated by taking the absolute magnitude of the output of the bandpass filter. This is equivalent to full wave rectification. In order to see the effect of performing a half wave rectification in the detector stage the output of the bandpass filter is half wave rectified and Fig. 5.6 shows comparison of the output of the detector stage which includes a half wave rectification and full wave rectification. The better performance of full wave rectification against the half wave rectification, as shown in fig. 5.6 governs the choice of full wave rectification for the detector stage even though the processing done by the hair cells of the ear is essentially a half wave rectification

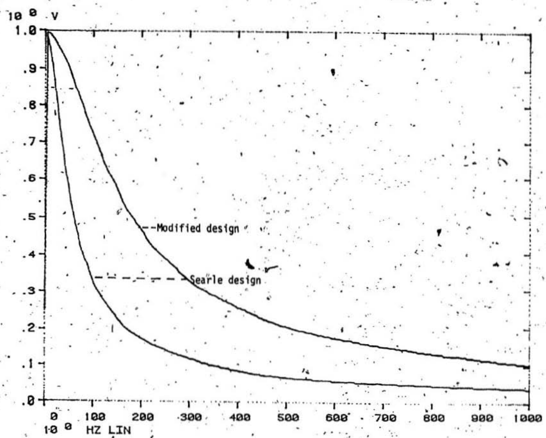


Fig. 5.3 Comparison of the lowpass filter frequency responses

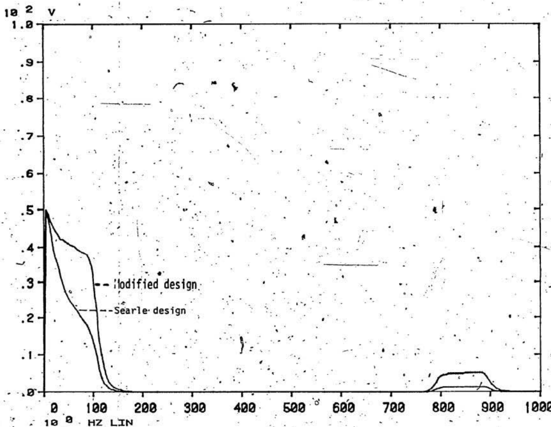


Fig. 5.4 Comparison of the output of the lowpass filter. The frequency translated signal is applied as input.

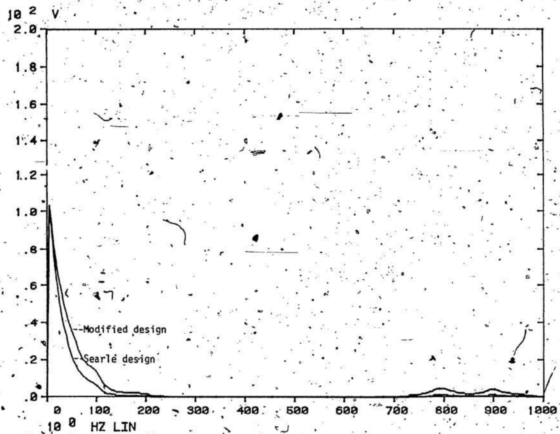
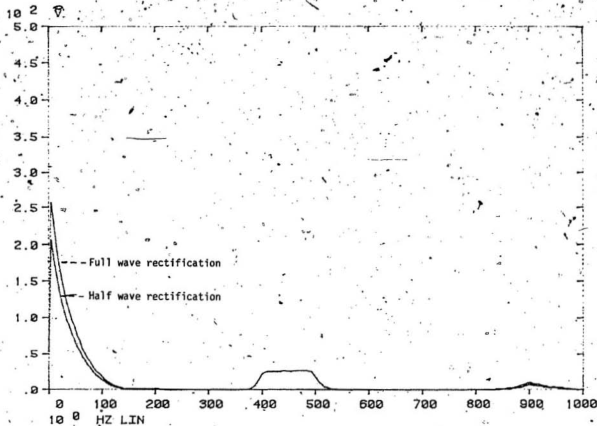


Fig. 5.5 Comparison of the output of the lowpass filter. The rectified bandpass signal is applied as input



5.6 Comparison of the output of the detector stage which includes the half wave and full wave rectification.

process. The objective at the moment is to effect the best reconstruction possible and once achieved the exact configuration of the detector stage can be looked at more closely.

5.2 SPEECH RECONSTRUCTION FROM THE OUTPUT OF THE MODEL

In the numerical modelling, 1/3 octave bandpass filters with maximally flat frequency response are used. In comparing the 1/3 octave frequency response with the response of the auditory nerve fiber in Fig. 1.9, the 1/3 octave filters in the numerical model do not have the same asymmetric frequency response skirts as shown by the psychophysical experiments done on the membrane (Kiang, 1974). Apart from this, only 16 logarithmically spaced channels are used as compared to the 30,000 channels. In order to test the validity of the model, and to test the output of the model for the speech intelligibility, the output of the model at each stage, ie., at the filter bank and at the detector bank are reconstructed.

5.2.1 Filter bank summation

The digitized speech data block is passed through the filter bank and summed at the output. The summation signal is converted to an analog signal using the digital to analog converter and is taped. The input speech signal and the filter bank summation output and its magnitude spectrum are shown in Fig. 4.8 and Fig. 4.12 respectively. The speech signal shown in Fig. 4.12 corresponds to the voiced portion |WA| of the sentence fragment 'watch'. The sudden stress in the signal followed by the periodicity in the waveform can be seen. The waveform corresponds to approximately 0.1 sec of speech. The filter bank summation signal when tested for audibility indicated that it is intelligible and the message content of the sentence is also preserved. This is clearly evident by examining the magnitude spectrum of the input and the filter bank summation signal. The slight spectral degradation of the filter bank summation signal can be seen but the spectral formant energy peaks are evident and are in close resemblence to the input signal.

In order to test the effect of phase on the filter bank summation output, the alternate channels of the filter bank are phase shifted by 180^{0} and the signal is summed. The summed signal is converted to an analog signal and is taped. When tested for audibility, the speech sounded similar to the signal obtained from the earlier filter bank summation. Fig. 5.7 shows the phase shifted filter bank summation signal and its magnitude spectrum. It can be seen that there is more spectral information but this did not have much influence on the audibility of the signal. A pulsation was also noticed in the signal which was due to the system limitations in the analog to digital conversion of the signal. These experiments indicate that the model preserves enough information for analysis.

5.2.2 Filter Detector Bank Reconstruction

The output of the filter detector model provides the envelope information of the input signal. The output of the model is to be reconstructed for intelligibility testing. It was observed earlier that there is information loss due to rectification and this makes it impossible for perfect reconstruction. It was discussed in the previous chapter that the speech is reconstructed at the output of the model by modulating the output of the filter detector bank with the carrier frequency. The choice of the carrier

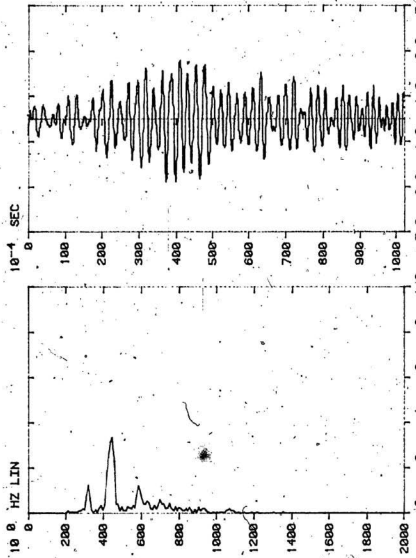


Fig. 5.7 (a) Filter bank summation output signal (alternate channels phase shifted)

(b) Magnitude spectrum

frequency is shown using the ramp signal of Fig. 5.1(b). Fig. 5.2 shows the frequency translated signal. The lower cutoff frequency and the center frequency of the channel are considered in the reconstruction. As discussed in the previous chapter, the bandpass filters in the channel are used to remove the redundant information that occurs due to modulation. In this regard the lower cutoff frequency modulation followed by the bandpass filter in the channel provides a more effective reconstruction. This can be seen from the Fig. 5.8 where in the reconstructed signal obtained using center frequency as carrier and, lower cutoff frequency as carrier are shown. Since we have used the bandpass filter in the channel to extract relevant information the reconstructed signal obtained using lower cutoff frequency resembles more closer to the input spectrum. Center frequency as carrier can be used provided we use different filter to extract the information. In addition to this reconstruction, bandpass filtered noise is also used to modulate the output of the filter. Fig. 5.9 shows the modulation obtained at the output of channel 4, (with bandpass filtered noise) wherein, it can be seen that there is more information but the spectral shape of the translated envelope is distorted. Fig. 5.10 shows the reconstructed signal using lower cutoff frequency as carrier along with its magnitude spectrum. Fig. 5.11 shows the reconstructed signal using center frequency as carrier. Fig. 5.12 shows the reconstructed signal using bandpass filtered noise. The spectral shift caused due to the choice of lower cutoff frequency as the carrier can be seen from the Fig. 5.10.

Comparison of the Fig. 5.10 and Fig. 5.11 indicates that the use of lower cutoff frequency preserves more relevant information and resembles

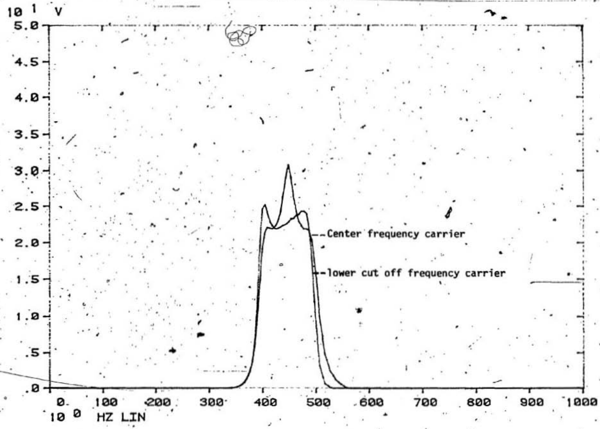


Fig. 5.8 Comparison of reconstruction using the center frequency and lower cut off frequency of the channel as the carrier.

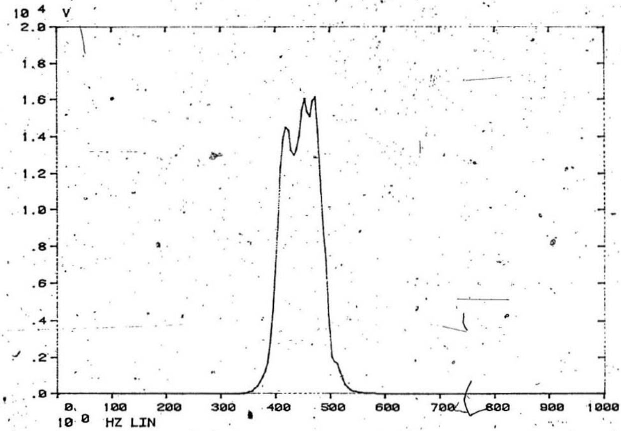


Fig. 5.9 Reconstruction using bandpass filtered noise.

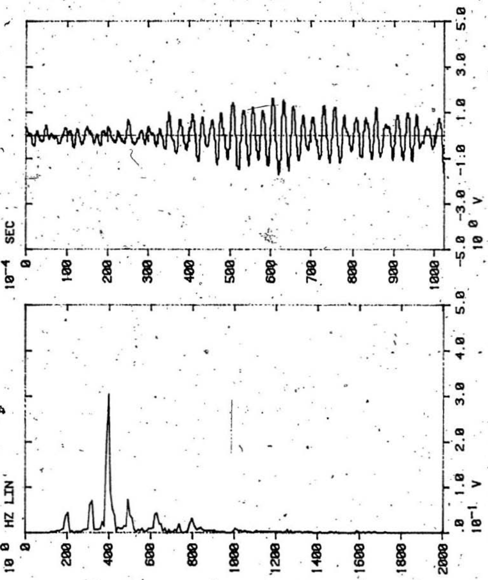


Fig. 5.10 (a) Speech signal reconstructed using the channel lower cut off frequency

(b) Magnitude spectrum

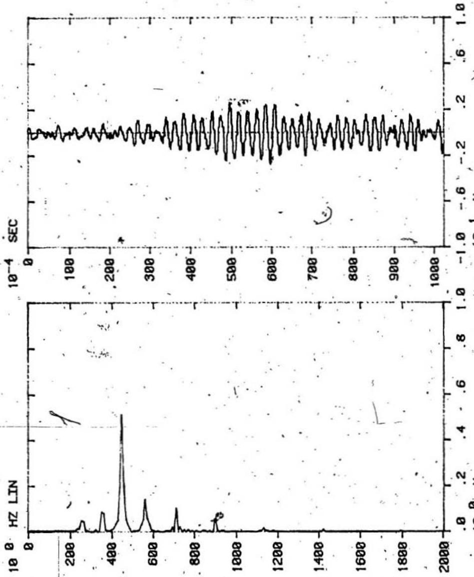


Fig. 5.11 (a) Speech signal reconstructed using the channel center frequency

(b) Magnitude spectrum

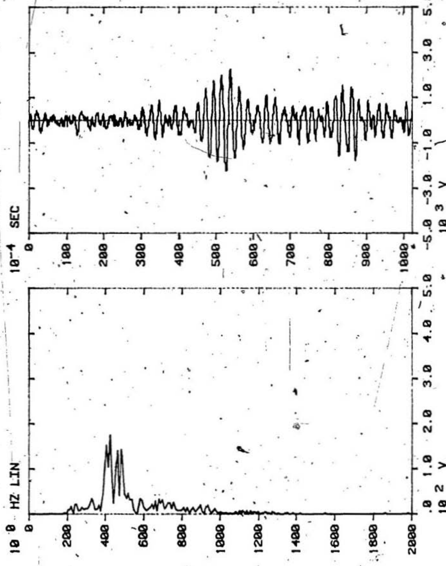


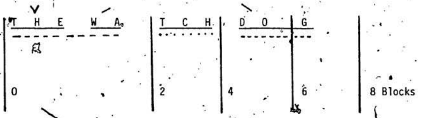
Fig. 5.12 (a) Speech signal reconstructed using the bandpass filtered noise.

(b) Magnitude spectrum.

better the input speech spectrum. The spectral peaks are clearly evident. The magnitude spectrum of the above processed speech signal indicate that the spectral information content of the original signal are present in the reconstructed signal. But when the reconstructed signal is tested for audibility the reconstruction is found to be imperfect. Due to system limitations, there is a considerable pulsation. Also a higher frequency ringing sound is heard in the signal. But it is noted that the audibility of the different segments of the spoken sentence are clear. The signal reconstructed using the bandpass filtered noise modulation sounds like a hoarse whisper but the sentence is understandable. Comparing the carrier frequency modulated signal with the noise modulated one, the carrier frequency modulated signal is much closer to the original signal as it retains the inflection and the originality of the speech spoken by the speaker. The better performance of the signal reconstructed using lower cutoff frequency is clearly seen in the signal when tested for audibility. But noise and other disturbances have to be removed if reconstructed signal has to be tested for intelligibility.

5.3 LINGUISTIC CATEGORIZATION OF THE RUNNING SPECTRUM PLOT

The running spectrum is linguistically categorized to emphasize the linguistic features in the spoken sentence. This is done to show that the linguistic features are preserved and are clearly evident in the running spectrum plot. The phonetic structure of the sentence is categorized as follows



----- Voiced segments

..... Unvoiced segments

Stress :

´ High

^ High Mid

` Low Mid

v Low

The sentence is classified as having 3 voiced segments and an unvoiced segment. The stress in each segment is also indicated. Fig. 4.13 shows the running spectrum plot of the portion 'THE WATCH' and Fig. 4.14 shows the running spectrum plot of the portion 'DOG'. It can be seen from the plots that the strongest peak occurs in speech data block 2 in Fig. 4.13. The semivowel /WA/ in the linguistic categorization is the strongly stressed portion of the sentence. So the strongest peak of amplitude coincides with the strongest stress on the vowel of 'The Watch'.

From the linguistic categorization of the sentence it can be seen that all the linguistic segments are voiced and they have a periodic source of excitation except for the /TCH/ which is unvoiced sound which has a silent closure period followed by sudden onset of energy in the higher channels.

The characteristic feature of the affricative /TCH/ can be noticed in speech data block 4 of Fig. 4.13 to have a short burst of noise followed by a prolonged noise. It can be seen that there is no fundamental frequency in the block.

In Fig. 4.14 the sudden burst of the voiced segment /D/ in 'Dog' can be seen. /D/ is a plosive which has a short burst of noise followed by a prolonged vowel /A/. The sentence has a voiced segment /A/ repeated and hence it should have some resemblence in characteristics. This can be noticed by examining the /WA/ and the /DA/ portion of the spectrum in Fig. 4.13 and Fig. 4.14 where there is a sharp rise and a fall off. The formant peaks of the spoken sentence is also seen in the plots. It can be concluded from the analysis that the identification of the linguistic segments are relatively easy and the syllables are also clearly seen separated in the plot.

Even though many of the speech characteristics are evident in the log magnitude plot (Fig. 4.15 and Fig. 4.16), the running spectrum analysis couldn't be carried out because of the high noise level present in the processed data. This was due to system limitations (explained in section 5.6) and also to the modification in the detector design which yielded faster rise times than the ones obtained using Searle's model. Comparing Fig. 2.4 with Fig. 4.15, the sharp changes in character between the voiced sounds, the unvoiced sounds (Blocks 31 to 34 of Fig. 2.4) and the silent periods evident in Fig. 2.4 are not clearly evident in Fig. 4.15. But comparing with Fig. 4.13, Fig. 4.14, the changes between voiced and unvoiced sounds are seen (Blocks 12-16 of Fig. 4.13). The abrupt attack of the /D/ (Block 37 of Fig. 2.4) is also noticeable in Fig. 4.14 (Blocks 18-19). Apart from this in Fig. 2.4, 0.7168 seconds of speech are plotted where as in Fig. 4.15 and in Fig. 4.16, 0.8192 seconds of speech are plotted. The variation in speech due to different speakers can be seen but the speech characteristics remain to be the same without much change.

In order to study the speaker variation of the same sentence, the speech of the male speaker and a female speaker is also processed using the model. Fig. 5.13 and Fig. 5.14 show the sentence 'The Watch Dog' spoken by another male speaker and Fig. 5.15 and Fig. 5.16 show the same sentence spoken by a female speaker. The previously explained linguistic categorization can be extended to these running spectra also. Comparing the male voice in Fig. 4.13 and Fig. 5.13 with the female voice in Fig. 5.15 it can be seen that the spectral formant peaks of the female voice are closer and the peaks are strong compared to male voice spectrum. There is a rapid/falloff in the male spectrum whereas there is no such fall for the female spectrum and the formant peaks are more pronounced in the female voice spectrum.

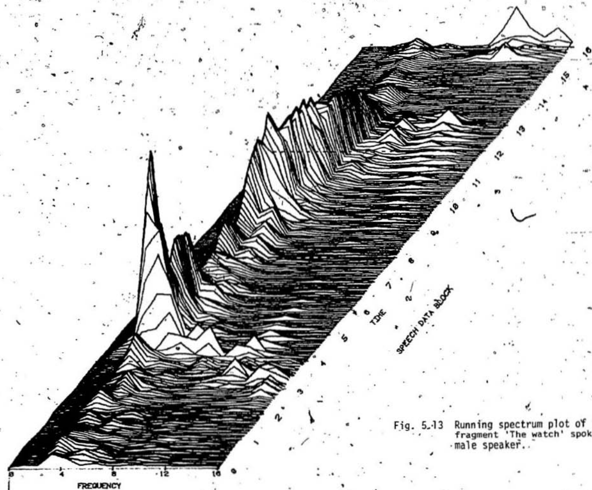


Fig. 5.13 Running spectrum plot of the speech fragment 'The watch' spoken by a male speaker.

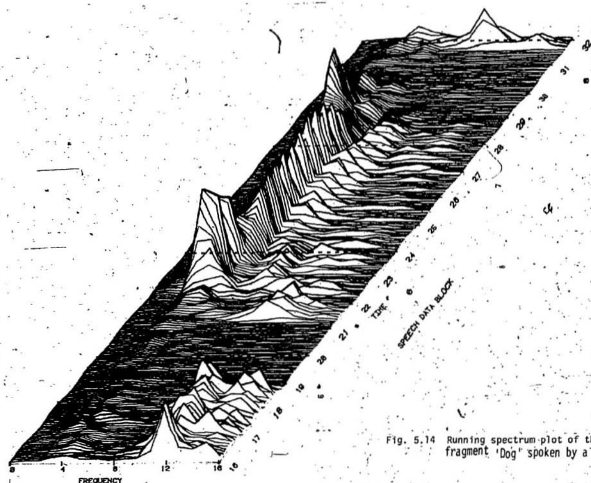


Fig. 5.14 Running spectrum plot of the sentence fragment 'Dog' spoken by a male speaker.

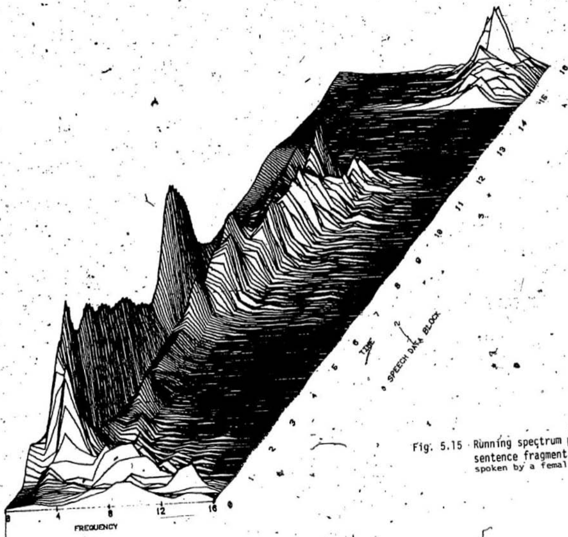


Fig. 5.15 Running spectrum plot of the sentence fragment 'The watch' spoken by a female speaker.

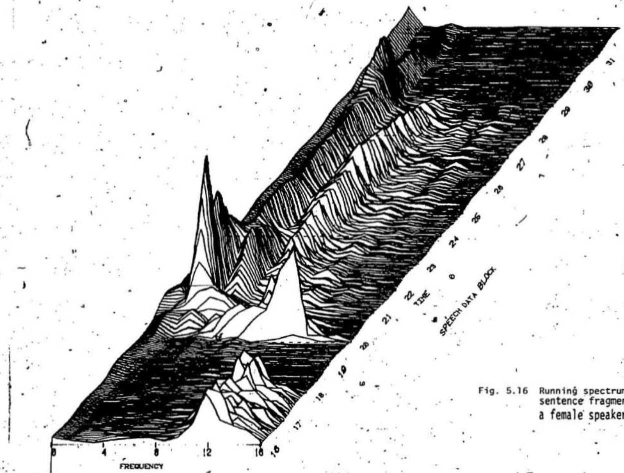


Fig. 5.16 Running spectrum plot of the sentence fragment 'Dog' spoken by a female speaker.

Thus the running spectrum provides a clear visual picture of the temporal and spectral characteristics of speech.

5.4 SIGNAL PROCESSING

In the signal processing carried out in this study, the data reduction scheme using the orthogonal transform 'Discrete Cosine Transform' (DCT) is investigated. The output of the filter detector when subjected to DCT effectively packs the data into a fewer number of lower channels. One of the objectives in signal processing using DCT is to test if perceptually relevant information is packed in a suitable manner or not so that an effective data reduction can be achieved.

The signal processing carried out is shown in the block diagram given below in Fig. 5.17.

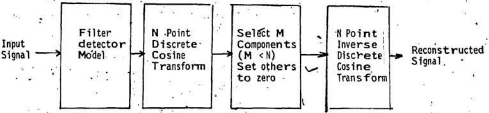


Fig 5.17 Block diagram of the signal processing done

The processing is carried out and two different studies are done while transforming the filter detector output, viz., subjecting (1) the log magnitude data and (2) the linear data of the model to discrete cosine transform; the merits and limitations of the two procedures are also reviewed. In the analysis the running spectrum plot is obtained to visualize the temporal and spectral degradation and the output of the processed speech is reconstructed for testing the intelligibility of speech.

5.4.1 Log Magnitude Data Subjected to Discrete Cosine Transform

The output of the filter detector is subjected to DCT and the signal reduction is investigated by retaining 16, 12, 10, 8, 6, 4, 3, or 2 components of DCT. In each case the remaining components are set to zero. After setting the necessary channels to zero the signal is transformed back using the IDCT. The output is reconstructed using the method explained before. The underlined portion of the sentence is used for processing viz., The Watch Dog'. The running spectrum representation for each case is shown in Fig. 5.18 to Fig. 5.24. By closely comparing the plots with the output of the filter detector in Fig. 4.13, the spectral and temporal degradation of the signal can be seen as the number of channels of data retained is reduced. The strong peak in block 2 is degraded and the spectral peaks can be seen to be flattening. The variation in shape is noticeable even though the basic form remains the same. Signal reconstruction using up to 6 channels of data indicates signal degradation although the temporal and spectral characteristics are preserved to some extent. But as the channels are reduced further the spectrum gets smoothed, with considerable loss of information.

In order to test the intelligibility of the reconstructed signal, the output of the filter detector, subjected to the DCT, is reconstructed and taped; due to system limitations (See Section 5-6) the reconstruction is far from perfect - considerable pulsation is present in the reconstructed voice. In spite of that, the spoken words could be identified by a "trained"

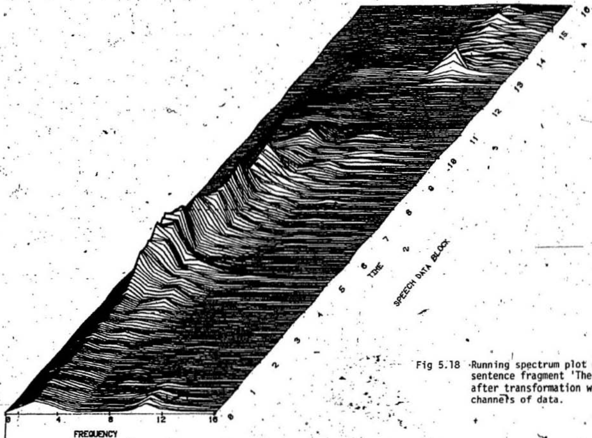


Fig 5.18 Running spectrum plot of the sentence fragment 'The watch' after transformation with 12 channels of data.

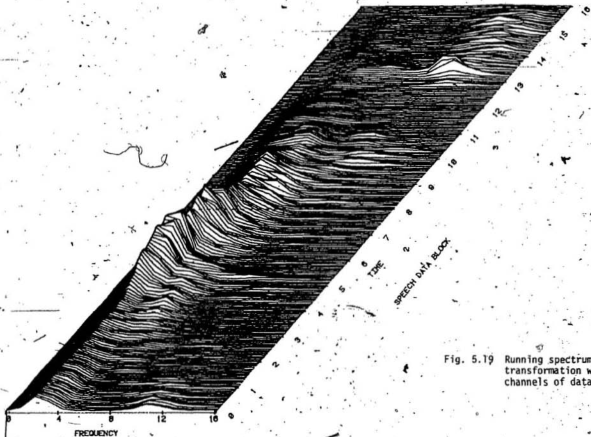


Fig. 5.19 Running spectrum plot after transformation with 10 channels of data.

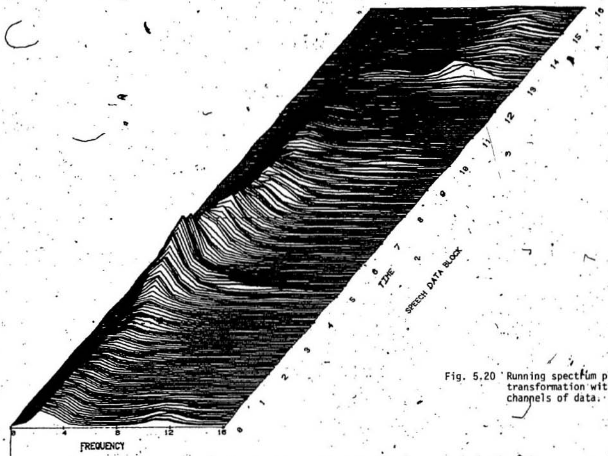


Fig. 5.20 Running spectrum plot after transformation with 8 channels of data.

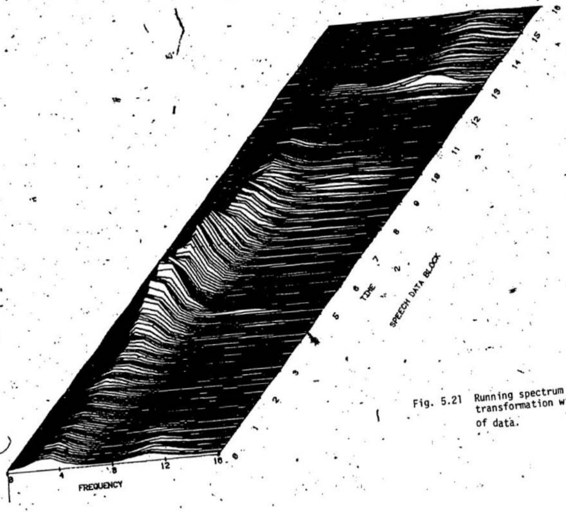


Fig. 5.21 Running spectrum plot after transformation with 6 channels of data.

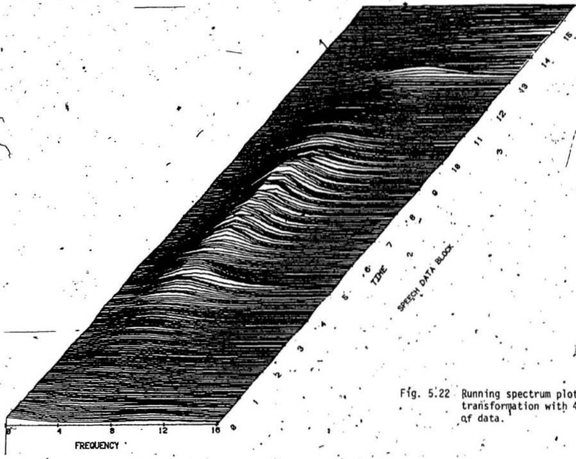


Fig. 5.22 Running spectrum plot after transformation with 4 channels of data.

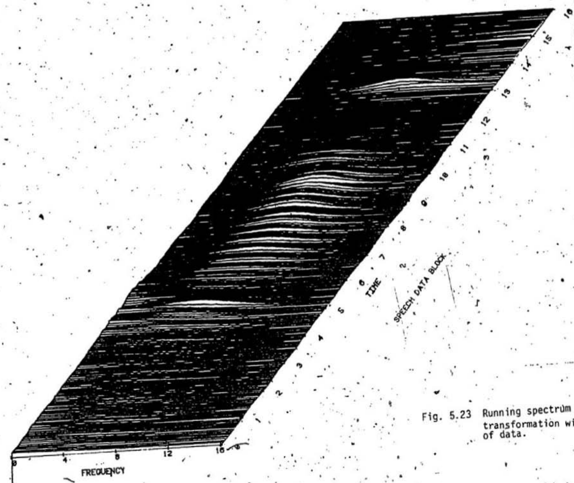


Fig. 5.23 Running spectrum plot after transformation with 3 channels of data.

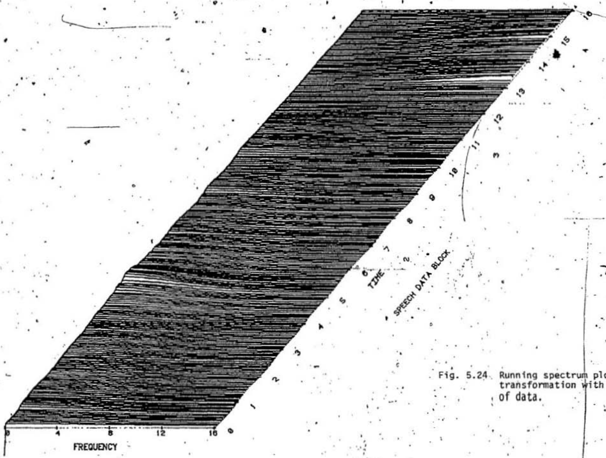


Fig. 5.24 Running spectrum plot after transformation with 2 channels of data.

listener i.e., one who already knows what is being said. This is true even when only four channels are used in the reconstruction. But when the channels are reduced further the signal cannot be identified. The noise and other disturbances in the signal have to be eliminated if the synthesized output is to be used for perceptual testing. Nevertheless, data reduction appears to be possible using the transformation for speech recognition.

The reconstructed speech signal and the magnitude spectrum for each of the reduced-channel systems described above are plotted. Fig. 5.25 to 5.30 shows the processed speech signal and its magnitude spectrum. The figures are compared with the signal reconstructed using 16 channels of data (Fig. 5.25). It can be noticed from the magnitude spectrum that the reconstructed signal is totally distorted for the reconstruction with channels less than 6. The spectral energy is considerably reduced and it can be seen that unwanted spectral information and lower frequency components are introduced. The signal degradation can also be seen from the time signal given in Fig. 5.25 to Fig. 5.30.

5.4.2 Linear Data Subjected to Discrete Cosine Transform

In the second case the linear output of the filter detector is subjected to the DCT. (Fig. 5.31). It is seen from the plot that the information is concentrated in the lower channels but noticeably more information is also available in the higher channels, as compared with the plots obtained using the logarithmic data. Thus it can be inferred that the DCT doesn't concentrate linear data nearly as effectively as logarithmic data. When the reconstruction procedure using IDCT is followed at the output of the transformed data, the exact signal is recovered.

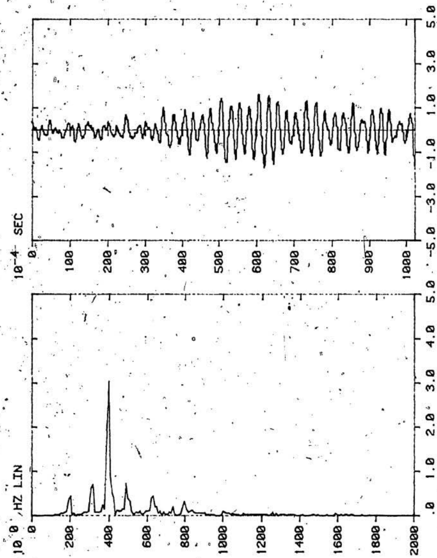


Fig. 5.25 (a) Speech signal reconstructed after transformation with 16 channels

(b) Magnitude spectrum.

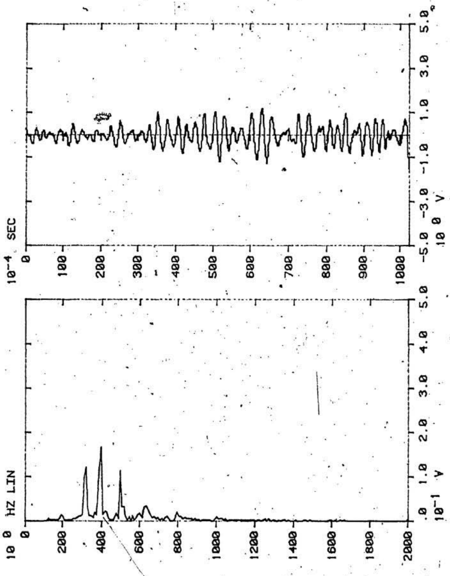


Fig. 5.26 (a) Speech signal reconstructed after transformation with 8 channels

(b) Magnitude spectrum

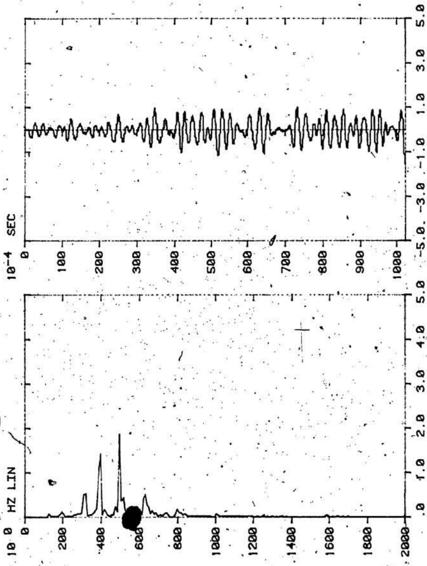


Fig. 5.27 (a) Speech signal reconstructed after transformation with 6 channels.

(b) Magnitude spectrum.

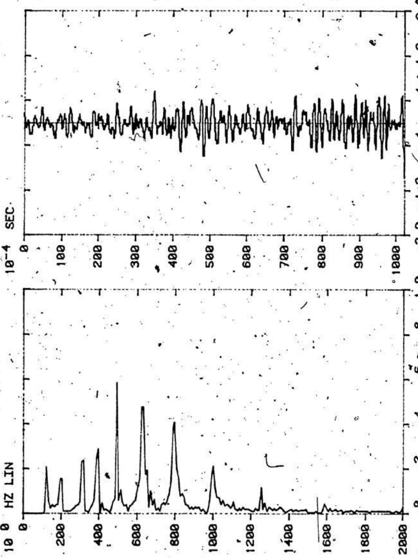


Fig. 5.28 (a) Speech signal reconstructed after transformation with 4 channels.

(b) Magnitude spectrum.

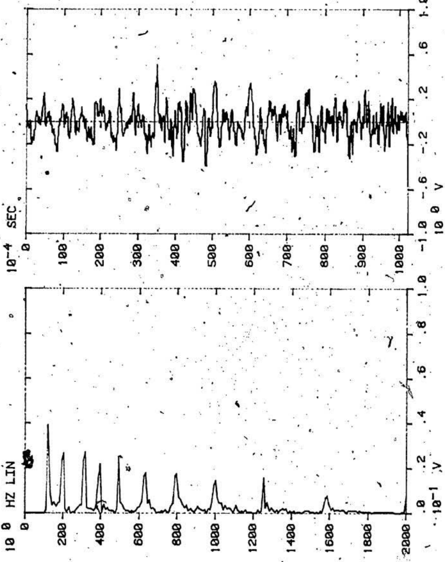


Fig. 5.29 (a) Speech signal reconstructed after transformation with 3 channels.

(b) Magnitude spectrum.

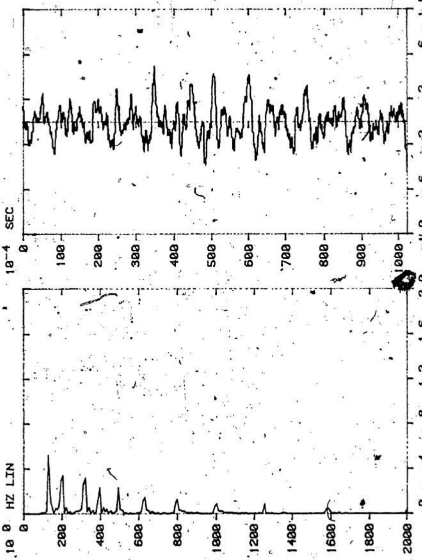


Fig. 5.30 (a) Speech signal reconstructed after transformation with 2 channels.

(b) Magnitude spectrum

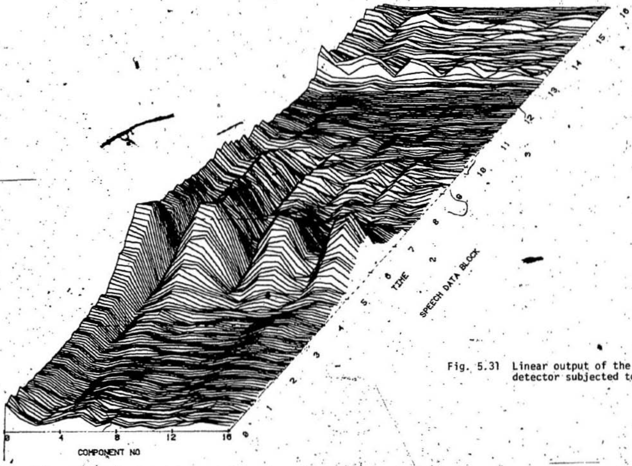


Fig. 5.31 Linear output of the filter detector subjected to transformation

(Fig. 5.32). Since logarithmic transformation evens out the time signal to a more smoother data the DCT with log transformed data concentrates the information more in the lower channels than the linear data. The same channel cutoff followed for the log magnitude data are repeated for the linear data and the signal processed. Fig. 5.33 to Fig. 5.39 show the running spectrum representation with data retained in 16, 12, 10, 6, 4, 3 or 2 channels respectively. By comparing the running spectrum plots obtained by linear and log magnitude data transformation it can be seen that data reduction with logarithmic data gives a better spectral preservation when larger number of channels are cut off but when fewer channels are cut off, the linear data gives better spectral preservation. The linear processed data is also reconstructed (with data retained in 16, 12, 10, 6 and 4 channels) and taped for intelligibility testing. The spoken words are identified with data in only 6 channels and with data in two channels the spoken words cannot be understood. The time history and the magnitude spectrum of the signal are given in Fig. 5.40 to 5.44 from which the distortion introduced in the signal due to data reduction can be seen.

Comparing the signal reconstructed for the linear and the log processed data with 4 channels (Fig. 5.28 and Fig. 5.44), the degraded spectral peaks are seen inspite of distortion in the log processed data but in the linear case the distortion is more pronounced which makes the audibility difficult.

One other result that can be discussed regarding the transform is that the transform is speaker independent. In order to test this, the same sentence 'The Watch Dog' spoken by a male and female speaker are processed using the

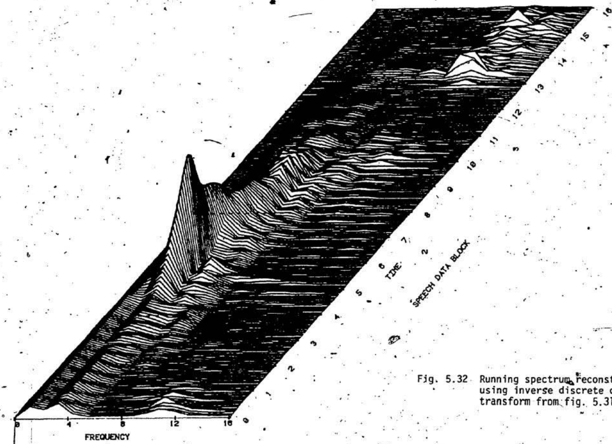


Fig. 5.32 Running spectrum reconstructed using inverse discrete cosine transform from fig. 5.31.

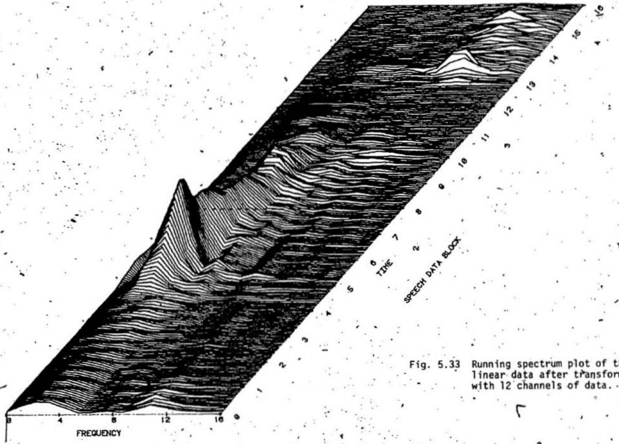


Fig. 5.33 Running spectrum plot of the linear data after transformation with 12 channels of data.

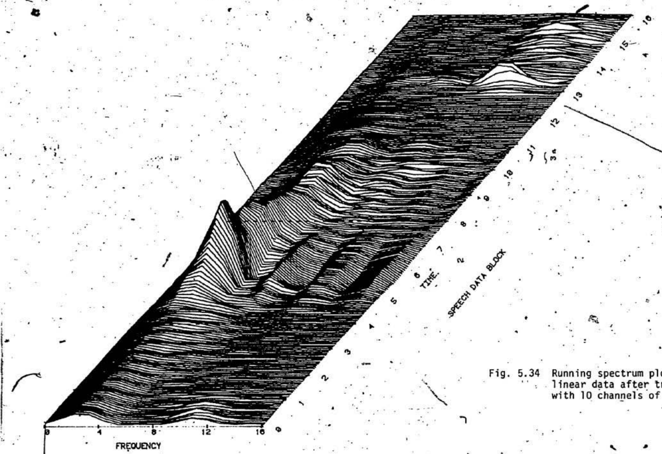


Fig. 5.34 Running spectrum plot of linear data after transformation with 10 channels of data.

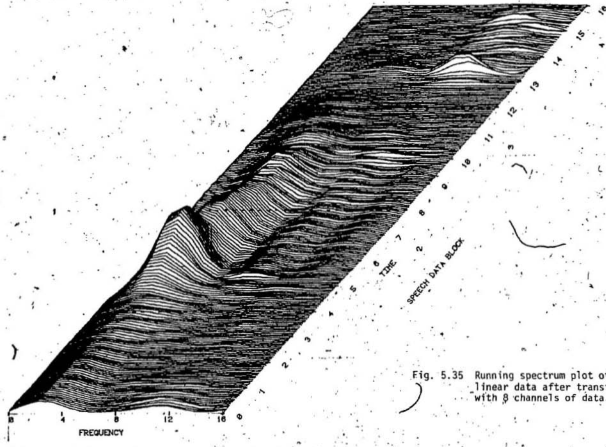


Fig. 5.35 Running spectrum plot of the linear data after transformation with 8 channels of data.

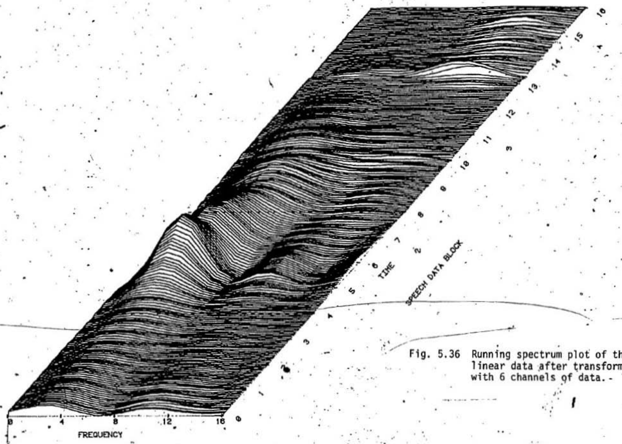


Fig. 5.36 Running spectrum plot of the linear data after transformation with 6 channels of data.

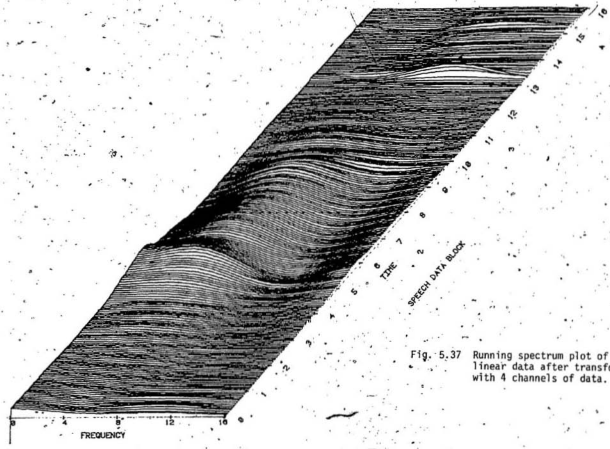


Fig. 5.37 Running spectrum plot of the linear data after transformation with 4 channels of data.

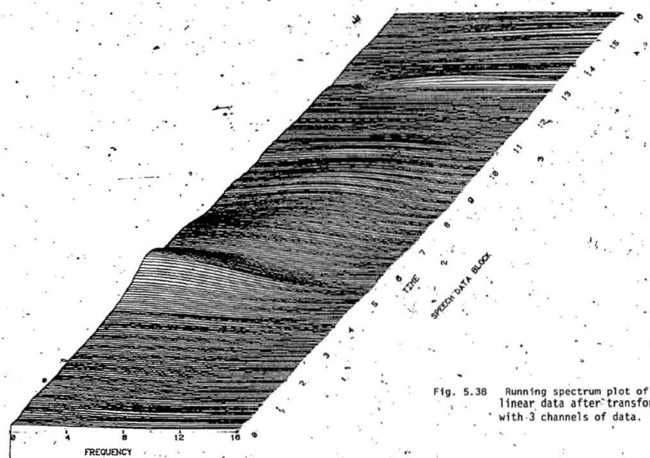


Fig. 5.38 Running spectrum plot of the linear data after transformation with 3 channels of data.

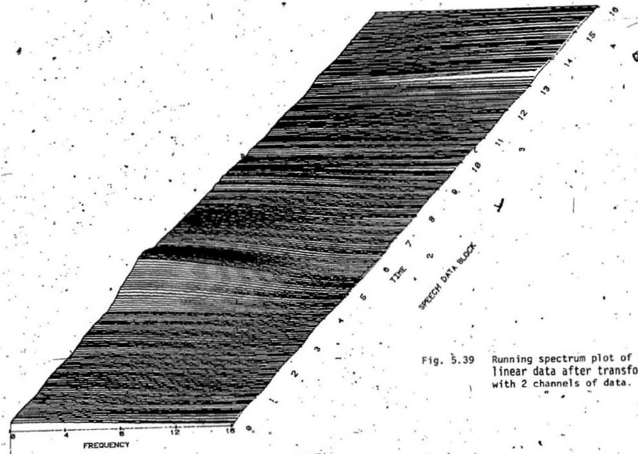


Fig. 5.39 Running spectrum plot of the linear data after transformation with 2 channels of data.

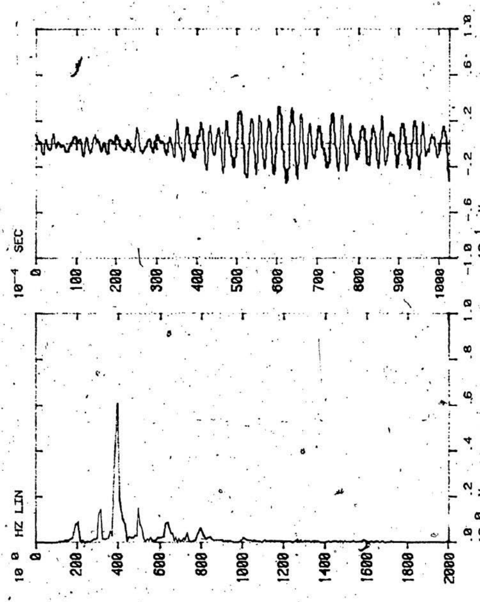
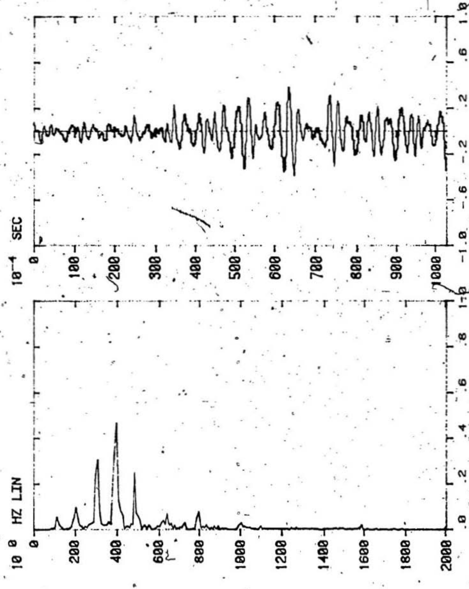


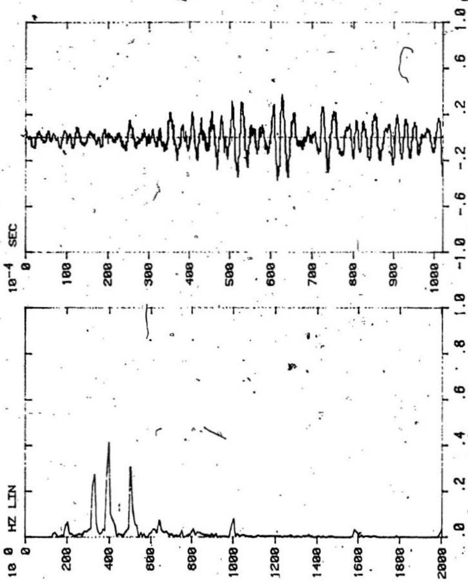
Fig. 5.40 Speech signal reconstructed (linear data) after transformation with 16 channels.

(b) Magnitude spectrum.



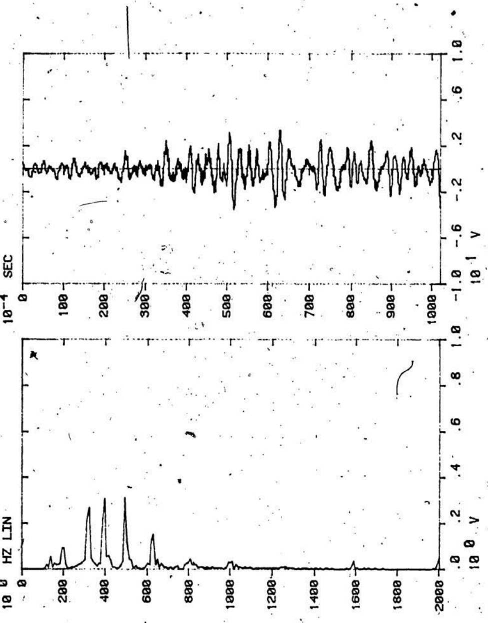
5.41 (a) Speech signal reconstructed (linear data) after transformation with 12 channels.

(b) Magnitude spectrum.



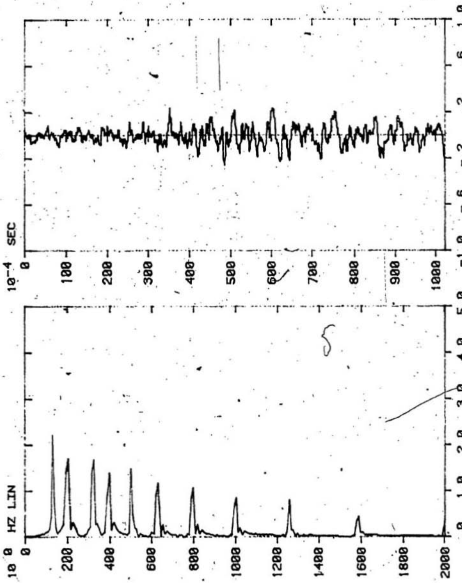
5.42 (a) Speech signal reconstructed (linear data) after transformation with 10 channels.

(b) Magnitude spectrum.



5.43 (a) Speech signal reconstructed (linear data) after transformation with 6 channels.

(b) Magnitude spectrum.



5.44 (a) Speech signal reconstructed (linear data) after transformation with 4 channels.

(b) Magnitude spectrum.

model). The running spectrum representation is already explained in Figs. 5.13 to 5.16. The filter detector outputs are subjected to DCT and Fig. 5.45 shows the detector output subjected to DCT for the male speaker and Fig. 5.46 shows the filter detector output subjected to DCT for the female speaker. By comparing the two figures, it can be noticed that irrespective of the speaker the transform forces the data into the lower channels and in both the cases 90% of the total data can be seen to be concentrated in the lower 5 to 6 channels.

5.5 COMPUTER FACILITIES USED AND PROCESSING DONE

As most of the work done related to the research was a study of the possibilities in speech recognition for real time implementation, a real time processing was not attempted at this early stage of work. The digital bandpass filter in the model were realized using the HP BASIC in HP 5451B Fourier Analyzer System. The system is equipped with a disk, magnetic tape storage, plotter and a screen display; it is a 16 bit machine with a 32K core memory, performing most of the signal processing operations. The implementation of the various stages in the model, the real time data acquisition and the reconstruction of the speech were done using the Fourier Keyboard programs.

The signal processing involved larger amounts of data. The amount of data could be estimated as follows. The continuous speech sentence that was processed is 'THE WATCH DOG GAVE A WARNING GROWL'. This sentence had 32 digitized blocks of data with 4096 data points in each data block. Since the data reduction and processing were carried out at the output of the filter detector, with 16 channels in the model, the total no of blocks that

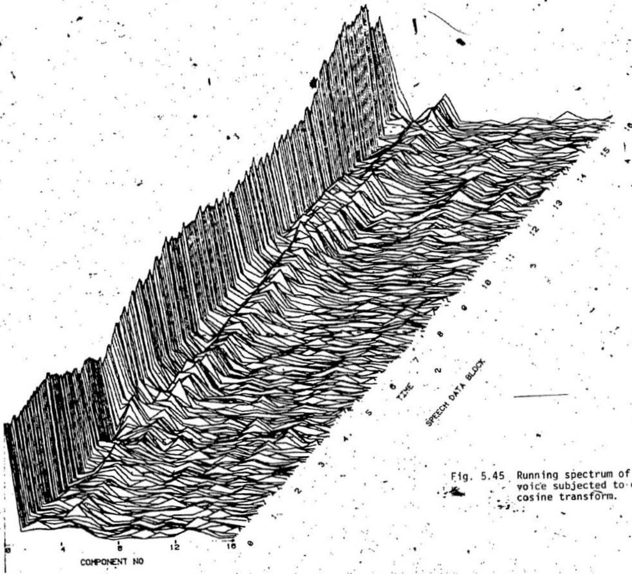


Fig. 5.45 Running spectrum of the male voice subjected to discrete cosine transform.

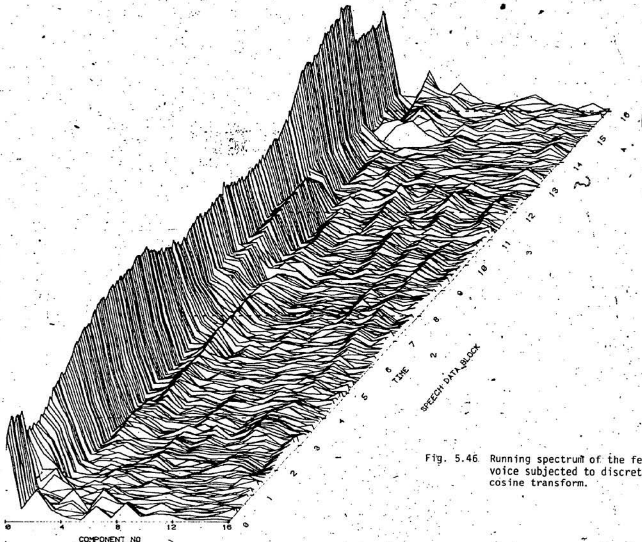


Fig. 5.46 Running spectrum of the female voice subjected to discrete cosine transform.

needed to be processed was 512 blocks with 4096 data points in each data block. This large amount of data processing couldn't be done using the Fourier Analyzer System due to memory limitations. Hence the signal processing was carried out on a PDP 11/60, a 32 bit machine with 256 K core memory. Even in this system memory limited the processing to 4 speech blocks at a time.. This procedure was repeated 8 times to completely process the sentence. The data representation of the Fourier Analyzer System and the PDP 11/60 were different and every time data is read in, it was made compatible with the PDP data format, processed and changed to Fourier format for further processing. Hence suitable tape routines to read and write the data were developed using FORTRAN-77. Each signal processing of the continuous speech sentence (4 data blocks) took approximately 8 hours in the PDP 11/60 system and 4 hours in the Fourier Analyzer System totalling 12 hours of processing.

The running spectrum plotting was obtained using the Fourier Analyzer System screen. The processed data from the Fourier was also plotted using PDP 11/60. The plotting algorithms were developed in FORTRAN-77.

5.6 POSSIBLE SOURCES OF ERROR

As discussed previously, the filter characteristics obtained using the psychophysical experiments have asymmetric frequency response characteristics whereas in the design, only maximally flat response characteristics were considered. Besides this; the 30000 redundant hair cells have been modelled using only 16 channels. As the model is implemented in software these issues can be addressed in the future relatively easily.

The detector function which Searle modelled with a diode, was modelled using the absolute value function. Since a virtually perfect reconstruction of the signal is possible before detection, the exact characteristics of the ear's detectors should be examined more closely to see if, in fact, better reconstruction may not be possible.

Due to limitations in the Fourier Analyzer System in real time conversion from digital data to analog signal, the processed data couldn't be taped properly. The DAC available only provided 10 bit conversion; moreover when the data was output to it through an internal buffer, there occured a delay of approximately 20 msecs between block transfers due to switch-over from one buffer to a second one. This introduced a pulsation in the converted analog signal. Moreover the processing was done in two different systems with different data representations. This effect also introduced noise in the processed signal. In the reconstruction, the modulation introduced a high frequency tone. Because of these noise interference the processed signal couldn't be used for psychoacoustic testing for speech recognition. But by listening to the processed sentence it was clear that the message content and the proper inflection in the signal was present which made the processed speech intelligible.

5.7 SUGGESTION FOR FUTURE WORK

In the present work most of the analysis done in the three stages viz., implementation of the model, reconstruction and signal processing; the analysis was carried out in a digital signal processing perspective. It was inferred from the linguistic categorization of the speech, the linguistic features were clearly seen at the output of the model. In data reduction, from the reconstructed output, it was evident that inspite of the spectral and temporal degradation in the signal, the perceptual information for recognition was present in the reduced data. Hence a study of the variation of the linguistic features with data reduction, could be done to derive the optimum values of the features for recognition (since most of the research in this area are carried out in recognizing the linguistic features, as it is considered to be the manner in which brain interprets the data). Use of reconstruction based on signal modulation theory, analysis based on signal processing and an extensive linguistic analysis based on psychophysical experiments may unravel facts in this research area leading to a real time speech recognition with reduced memory storage.

As such in this work, only 16,1/3 octave filters were used. But this can be increased to a larger number of filters by reducing the bandwidth to 1/6 or 1/8 octave within the speech analysis range. This may lead to a better reduced dimension for analysis.

The DCT is applied and data reduction is based on the variance criterion. The error introduced due to setting of channels to zero depends on the number of channels cut off. One measure of the error is the mean square error criterion. It is suggested by Tribolet et. al. (1979) that the mean square error criterion is not the appropriate error criterion used for testing in terms of speech perception. Hence a suitable measure can be investigated to relate the spectral degradation to perceptual information degradation.

The information reduction can further be studied using this procedure by sampling the input signal at a lower rate or by averaging the output of the filter detector.

5.8 Conclusions

The software model based on human audition was built to investigate the feasibility of reducing the dimensionality representation of speech. The model paralled the Searle's hardware model. The software model provided flexibility to change the design characteristics and the implementation of the model for research. The output of the model was subjected to a linear transform, the discrete cosine transform to reduce the data to a minimum number of perceptually important dimensions. In order to study an effective reconstruction procedure for reconstructing the speech from the filtered, detected and transformed output of the model, the model was tested in each stage of implementation for information loss. It was seen that even though there was information loss in the model, sufficient information required for recognition was available at the output of the model. Several approaches based on signal processing techniques were investigated to reconstruct the speech so that an effective intelligibility testing could be resorted. A reconstruction was implemented by modulating the output of the model with the channel lower cutoff frequency and by bandpass filtering with the channel bandpass filter for each channel in the model. The continuous speech sentence 'THE WATCH DOG GAVE A WARNING GROWL' was analyzed, processed and reconstructed. Informal testing with a few trained listeners led us to believe that it might be possible to recognize speech with a minimum of three out of sixteen channels, thus leading to a possibility of data reduction. The effect of speaker variation on the performance of the model was also tested by processing a male and a female voice and it appeared that the transform is speaker independent. But much work has to be done.

REFERENCES

Ahmed, N., and Rao, K.R., Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, New York, 1975.

Ahmed, N., Natarajan, T., and Rao, K.R., Discrete Cosine Transform, IEEE Transactions on Computers, C-25, p. 90-93, 1974.

Ahmed, N., Natarajan, T., Discrete Time Signals and Systems, Reston Publishing Company, Inc., Virginia, 1983.

Batteau, D.W., The Role of Pinna in Human Localization, Royal Society of London Proceedings Series 8, 168(B), pp. 158-180, 1967.

Bekesy, von, G., Experiments in Hearing, McGraw Hill, New York, 1960.

Bertocci, G., Schoenherr, B.W., and Messerschmitt, D.G., An approach to the implementation of a Discrete Cosine Transform, IEEE Transactions on Communications, COM-30 (4) p. 635-641, 1982.

Campenella, S.J., and Robinson, G.R., A comparison of Orthogonal Transforms for Digital Speech Processing, IEEE Transactions on Communication Technology. COM-19 (6), p. 1045-1049., 1971.

Christovich, J.A., Grostrem, M.P., Kozhevnikov, W.A., Lesagor, L.W., Shuplyakov, V.S., Taljasin, P.A., and Tjulkov, W.A., A functional model of signal processing in the peripheral auditory system, Acoustica 31, p. 349-353., 1974.

Cole, R.A., and Scott, B., Toward a theory of speech perception, Psychological Review 81, p. 348-374., 1974.

Dockendorff, D., Application of a Computer Controlled Model of the Ear to Multiband Amplitude Compression, Unpublished M.Sc Thesis, Queen's University, Kingston, Ontario., 1978.

Dolmalzan, J.M., Baslet, L., and Schupljakov, V.S., A functional model of Peripheral Auditory System in Speech Processing, IEEE Acoust. Speech Signal Process., Rec., p. 261-264., April, 1977.

Eimas, P.D., and Corbit, J.D., Selective adaptation of Linguistic feature detectors, Cognitive Psychology 4, 99., 1973.

Flanagan, J.L., Speech Analysis, Synthesis and Perception, Springer, New York., 1972.

Flanagan, J.L., Schroeder, M.R., Atal, B.S., Crochiere, R.E., Jayant, N.S., Tribolet, J.M., Speech coding, IEEE Transactions on Communication, COM-27 (4) p. 710-736., 1979.

Gulick, W.L., Hearing, Physiology and Psychophysics, Oxford University Press, New York., 1971.

Helmholtz, H., On the Sensation of Tone, Dover (1954). 1885.

Hewlett Packard: Fourier Analyzer Operating Manual., 1973.

Johnstone, B.M., and Boyle, A.J.F., Basilar Membrane Vibration Examined with the Mossbauer Effect, Science, 15[illegible] p. 389., 1967.

Kiang, N.Y.S., and Moxon, E.C., Tails of Tuning Curvers of Auditory nerve fibers, Journal of Acoustical Society of America, 5[illegible] p. 620-630. 1974.

Kiang, N.Y.S., Watanabe, T.E.C., and Clark, L.F., Discharge patterns of single nerve fibers in a cat's auditory nerve, M.I.T. Press., 1965.

Klein, W., Plomb, R., and Pols, L.C.W., Vowel Spectra, Vowel Spaces and Vowel Identification, Journal of Acoustical Society of America 48(4). p. 999-1009., 1970.

Li, K. P., Hughes, G.W., and House, A.S., Correlation Characteristics and Dimensionality of Speech Spectra, Journal of Acoustical Society of America, 46(4), p.1019-1025., 1969.

Linggard, R., and Black, N.D., High-resolution filter bank for speech analysis, Journal of Acoustical Society of America, Suppl. 1 (5103) Vol. 71, 1982.

Mariani, J.J., and Lienard, J.S., Acoustic phoneme recognition of connected speech using transient information, IEEE Acoust. Speech Signal Process. Rec., 667-670, April, 1977.

Miller, J.L., Properties of feature detectors for speech. Perception Psychophysics, 18, p.389-397., 1975.

Oppenheim, A.V., and Schafer, R.W., Digital Signal Processing, Prentice-Hall, Inc., New Jersey, 1975.

Panter, P.F., Modulation, Noise and Spectral Analysis, McGraw-Hill, New York, 1965.

Patterson, R.D., Auditory filter shapes derived with noise stimuli, Journal of Acoustical Society of America, 59, p. 640-654, 1976.

Pols, L.C.W., Tromp, H.R.C., and Plomb, R., Frequency Analysis of Dutch vowels from 50 male speakers, Journal of Acoustical Society of America, 53 (4), p. 1093-1101., 1972.

Pols, L.C.W., Real Time recognition of spoken words. IEEE Transactions on Computers, C-20, p.972-978., 1971.

Rabiner, L.R., and Schafer, R.W., Digital Processing of Speech Signals, Prentice-Hall, Inc. New Jersey, 1978.

Rayment, S.G., Phoneme Recognition based on Feature Extraction from a model of the Auditory System, M.Sc., thesis (Queen's University, Kingston, Ontario), (unpublished), 1977.

Rhode, W.S., Observation of the Basilar Membrane in Squirrel Monkeys using the Mossbauer Technique, Journal of Acoustical Society of America, 49, p.1218-1231, 1971.

Richards, W., Quantifying Sensory Channels: Generalizing Colourimetry to Orientation and Texture, Touch and Tones, Sensory Processes, 3, p.207-229, 1979.

Scharf, B., Critical Bands, In Tobias, J.V., (ed) Foundations of Modern Auditory Theory, Vol. I, Academic Press, New York, 1970.

Schatz, C., The Role of Context in Perception of Stops, Language 30, 47-56., 1954.

Schwartz, M., Information Transmission, Modulation and Noise, McGraw-Hill, New York, 1980.

Searle, C.L., Speech perception from an auditory and visual view point, Canadian Journal of Psychology, 36(3), p.402-419, 1982.

Searle, C.L., Jacobson, J.Z., and Rayment, S.G., Stop consonant discrimination based on human audition, Journal of Acoustical Society of America, 65, p.799-809, 1979.

Searle, C.L., Analysis of music from an auditory perspective, Humanities Association Review, 30, p. 93-103, 1979/80.

Searle, C.L., Jacobson, J.Z., and Rayment, S.G., A phoneme recognition system based on human audition, International Conference on Acoustic, Speech Signal Processing Record, p557-560, 1978.

Searle, C.L., and Kimberly, B.P., Automatic discrimination of fricative Consonants based on Human audition, IEEE International Conference on Acoustic, Speech Signal Processing Record, p.89-92, 1979.

Tobias, J.V., Foundations of Modern Auditory Theory, Vol., Academic Press, New York, 1970.

Tonndorf, J., Cochlea Mechanics and Hydro-dynamics, In Tobias, J.V. (ed) Foundations of Modern Auditory Theory, Vol. I., Academic Press, New York, 1970.

Tribolet, J.M., and Crochiere, R.E., Frequency domain coding of speech, IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-27(5), P. 512-529. 1979.

Viemeister, N.F., Temporal Modulation Transfer Functions Based upon Modulation Thresholds, Journal of Acoustical Society of America, 66, p.1364-1380, 1979.

White, G.M., and Neely, R.B., Speech recognition experiments with linear prediction bandpass filtering and dynamic programming, IEEE Trans. Acoust. Speech Signal Process., ASSP 24, p.183-188, 1976.

Yilmaz, H.A., A Theory of Speech Perception, Bulletin of Mathematical Biophysics 29, p. 793-824, 1967.

Yilmaz, H.A., A Theory of Speech Perception, Bulletin of Mathematical Biophysics 30, p. 455-479, 1968.

Yilmaz, H.A., Ferber, L., Kellet, H., Park, W., and Koprucu, E., Speech Perception, TR-75-265, Rome Air Development Center, 1975.

Yilmaz, H.A., Ferber, L., Park, W., and Koprucu, E., Perception of Continuous Speech Recognition, TR-74-180, Rome Air Development Center, 1974.

Zahorian, S.A., and Rothenberg, M., Principal Component Analysis for low-redundancy encoding of speech spectra, Journal of Acoustical Society of America 69(3), p. 832-845, 1981.

Zielinski, R., and Noll, P., Adaptive Transform Coding of Speech Signals, IEEE Transactions on Acoustics Speech and Signal Processing, 25, p. 299-309, 1977.

Zwicker, E., On a Psychoacoustical Equivalent of Tuning Curves, In Zwicker, E., and Terhardt, E., (ed) Facts and Models in Hearing, Springer, New York, 1974.

Zwicker, E., Terhardt, E., and Paulas, E., Automatic Speech Recognition Using Psychoacoustic Models, Journal of Acoustical Society of America, 65, p. 487-498, 1979.

APPENDIX A

HP BASIC PROGRAMMES USED FOR THE IMPLEMENTATION OF THE DIGITAL BANDPASS FILTER AND LOWPASS FILTER (DETECTOR STAGE) IN THE FILTER DETECTOR BANK MODEL.

The following programmes are implemented in HP 5451B Fourier Analyzer System.

BPF: This routine calculates the coefficients for a cascade of N, fourth order Butterworth bandpass filter and obtains the impulse response of the bandpass filter.

LPF: This routine calculates the coefficients for the first order RC low pass filter and obtains the impulse response of the lowpass filter.

```
5     CALL SCORE(7)
8     REM: PROGRAM BPF.BAS
10    REM:DIGITAL IMPLEMENTATION OF THE SEARLE'S 1/3 OCTAVE
12    REM:FILTER BANK MODEL BASED ON HUMAN AUDITION
30    REM:CALCULATION OF BANDPASS FILTER COEFFICIENTS
35    REM:ROUTINE TO CALCULATE THE COEFFICIENTS FOR A CASCADE
40    REM:OF N 4'TH ORDER BUTTERWORTH BANDPASS FILTERS.
45    REM:EQUATIONS ARE AFTER "AHMED & NATARAJAN";DISCRETE
47    REM:SIGNALS & SYSTEMS.(1983)
50    PRINT "WOULD YOU LIKE TO RUN THE PROGRAM OR EXIT? ENTER 1 TO EXIT"
60    INPUT A1
70    IF A1=1 THEN 1056
80    REM:CONSTANT VALUE DECLARATION
90    LET P1=3.14159
100   DIM P[8,128],X[8,128],Q[128],Y[8,128],Z[8,128]
110   DIM W[8,128],V[8,128]
120   REM:THE ARRAY SIZE ACCORDING TO THE SET BLOCK SIZE
130   LET D3=8
150   PRINT "PLEASE INPUT THE LOWER CUT OFF FREQUENCY"
160   INPUT F1
170   PRINT "AND NOW THE UPPER CUT OFF FREQUENCY"
180   INPUT F2
185   PRINT "PLEASE INPUT THE NO. OF BUTTERWORTHS IN CASCADE:"
190   INPUT N
200   PRINT "PLEASE INPUT THE SAMPLING FREQUENCY IN HZ.:"
210   INPUT F3
220   LET T1=1/F3
225   LET W1=TAN(P1*F1*T1)
226   PRINT "F2",F2
227   PRINT "T1",T1
230   LET W2=TAN(P1*F2*T1)
235   LET W3=(W2-W1)
240   LET W4=W1*W2
249   PRINT
250   PRINT "THE BANDPASS FILTER COEFFICIENTS"
251   PRINT
255   PRINT "THE NUMERATOR COEFFICIENTS"
256   PRINT
260   REM:A(K):CONSTANT TERM;B(K):COEF OF Z^-2;
262   REM:C(K):COEF OF Z^-4;
265   PRINT "A(K)","B(K)","C(K)"
266   PRINT
280   FOR K=1 TO N
285   LET T4=((2*(K+N)-1)*P1)/(4*N)
290   LET T5=W3*COS(T4)
295   LET D2=(1+W4)*(1-2*T5+W4)+W3**2
300   LET A[K]=(W3**2)/D2
305   LET B[K]=-(2*W3**2)/D2
310   LET C[K]=(W3**2)/D2
315   PRINT A[K],B[K],C[K]
320   PRINT
325   NEXT K
```

```
355     PRINT
360     PRINT "THE BANDPASS FILTER DENOMINATOR COEFFICIENTS"
365     PRINT
367     REM:D(K):CONST TERM,E(K):COEF OF 1/Z,F(K):COEF OF Z**(-2)
368     REM:G(K):COEF OF Z**(-3),H(K):COEF OF Z**(-4)
370     PRINT "D(K)","E(K)","F(K)","G(K)","H(K)"
372     PRINT
375     FOR K=1 TO N
380     LET T2=((2*(K+N)-1)*P1)/(4*N)
390     LET T3=W3*COS(T2)
400     LET D1=(1+W4)*(1-2*T3+W4)+W3**2
410     LET D[K]=1
420     LET E[K]=(4*(W4-1)*(1-T3+W4))/D1
430     LET F[K]=(6*W4**2-4*W4+(6-2*W3**2))/D1
440     LET G[K]=(4*(W4-1)*(1+T3+W4))/D1
450     LET H[K]=((1+W4)*(1+2*T3+W4)+W3**2)/D1
460     PRINT D[K],E[K],F[K],G[K],H[K]
465     PRINT
470     NEXT K
555     PRINT
790     REM:THE FOURIER KEYBOARD PROGRAM IS USED TO COLLECT 1024
795     REM:POINTS OF DATA. THE SAMPLING PARAMETERS ARE SET ON THE
800     REM:CONTROL PANEL. THE FOURIER KEYBOARD PROGRAM RESIDES IN
805     REM:RECORD 0 OF THE PROGRAM AREA IN THE DISK $& DATA IS STORED
810     REM:IN RECORD OF THE DATA AREA IN THE DISK.
815     PRINT
820     PRINT "FOURIER KEYBOARD PROGRAM"
825     PRINT
830     CALL FCORE(4)
833     PRINT
835     PRINT "THE BLOCK SIZE IS TAKEN TO BE 1024"
836     PRINT
840     LET A4=1024
850     LET A5=A4-1
855     REM:THE RECORD NUMBER WHERE INPUT DATA IS STORED IS : 0
860     LET A6=0
865     REM:READ THE FOURIER DATA INTO THE BASIC ARRAY IN
866     REM:INTEGER FORMAT
870     LET A7=1
880     CALL DREAD(P[1,1],K[1],A4,A6,A7)
881     REM:NOW DISPLAY THE DATA IN THE BASIC ARRAY
890     CALL DSPLY(P[1,1],K[1],0,A5,0)
900     PAUSE
910     CALL NODIS
920     REM:ROUTINE TO CONVERT THE FOURIER DATA (INTEGER FORMAT)
925     REM:IN THE BASIC ARRAY TO FLOATING POINT FORMAT FOR BASIC
930     REM:PROGRAM MANIPULATION. THE ROUTINE HANDLES BLOCK SIZE > 256
935     REM:J,K,ARE THE ARRAY VARIABLES.
940     REM:CONST VALUE DECLARATION
950     LET J=1
960     LET K=0
970     LET A8=0
980     LET A9=128
```

```
984     PRINT
985     PRINT "INPUT TO THE STAGE NO:1"
986     PRINT
990     FOR I=0 TO A5
1000    CALL DGET(P[1,1],K[1],I,B1,A8)
1010    LET B2=A0*J
1020    IF I=B2 THEN 1040
1030    GOTO 1060
1040    LET J=J+1
1050    LET K=0
1060    LEY K=K+1
1070    LET X[J,K]=B1
1090    NEXT I
1100    REM*PRINT THE VALUES IN THE INPUT DATA ARRAY
1111    LET J=1
1115    FOR K=1 TO C1
1120    PRINT J,K,X[J,K]
1125    NEXT K
1150    REM:ROUTINE TO REALIZE THE BANDPASS FILTER FUNCTION
1160    REM:THE K'TH SECTION HAS THE FOLLOWING TRANSFER FUNCTION
1162    REM:     A(K)+B(K)*Z^-2+C(K)*Z^-4
1170    REM:        - - - - - - - - - - - - - - - - - - - - - - -
1180    REM:       D(K)+E(K)*Z^-1+F(K)*Z^-2+G(K)*Z^-3+H(K)*Z^-4
1192    REM:
1195    REM:IF X(N) AND Y(N) ARE THE INPUT AND OUTPUT OF THE K'TH
1200    REM:SECTION,THE FOLLOWING DIFFERENCE EQUATION IS SATISFIED
1202    REM:
1210    REM:Y(N)=1/D(K)*(A(K)*X(N)+B(K)*X(N-2)+C(K)*X(N-4)-E(K)*Y(N-1)
1220    REM:     -F(K)*Y(N-2)-G(K)*Y(N-3)-H(K)*Y(N-4)
1222    REM:
1230    FOR K=1 TO N
1234    PRINT
1236    PRINT
1240    LET Q[1]=A[K]*X[1,1]
1250    LET Q[2]=A[K]*X[1,2]-E[K]*Q[1]
1260    LET Q[3]=A[K]*X[1,3]+B[K]*X[1,1]-E[K]*Q[2]-F[K]*Q[1]
1270    LET Q[4]=A[K]*X[1,4]+B[K]*X[1,2]-E[K]*Q[3]-F[K]*Q[2]-G[K]*Q[1]
1273    REM:
1280    FOR I=1 TO D3
1290    FOR J=5 TO 128
1300    LET M1=A[K]*X[I,J]+B[K]*X[I,J-2]+C[K]*X[I,J-4]
1310    LET N1=E[K]*Q[J-1]+F[K]*Q[J-2]+G[K]*Q[J-3]+H[K]*Q[J-4]
1320    LET Q[J]=(1/D[K])*(M1-N1)
1330    NEXT J
1333    REM:
1340    FOR L=1 TO 128
1350    LET Y[I,L]=Q[L]
1360    NEXT L
1368    IF I=D3 THEN 1403
1370    LET M2=A[K]*X[I+1,1]+B[K]*X[I,127]+C[K]*X[I,125]
1380    LET N2=E[K]*Q[128]+F[K]*Q[127]+G[K]*Q[126]+H[K]*Q[125]
1390    LET Q[1]=(1/D[K])*(M2-N2)
1393    REM:
```

```
1400    LET M3=A[K]*X[I+1,2]+B[K]*X[I,128]+C[K]*X[I,126]
1410    LET N3=E[K]*Q[1]+F[K]*Q[128]+G[K]*Q[127]+H[K]*Q[126]
1420    LET Q[2]=(1/D[K])*(M3-N3)
1423    REM:
1430    LET M4=A[K]*X[I+1,3]+B[K]*X[I+1,1]+C[K]*X[I,127]
1440    LET N4=E[K]*Q[2]+F[K]*Q[1]+G[K]*Q[128]+H[K]*Q[127]
1450    LET Q[3]=(1/D[K])*(M4-N4)
1453    REM:
1460    LET M5=A[K]*X[I+1,4]+B[K]*X[I+1,2]+C[K]*X[I,128]
1470    LET N5=E[K]*Q[3]+F[K]*Q[2]+G[K]*Q[1]+H[K]*Q[128]
1480    LET Q[4]=(1/D[K])*(M5-N5)
1483    REM:
1490    NEXT I
1493    REM:
1494    PRINT
1496    PRINT
1500    FOR I=1 TO D3
1510    FOR J=1 TO 128
1520    LET X[I,J]=Y[I,J]
1530    NEXT J
1540    NEXT I
1550    NEXT K
1650    REM:ROUTINE TO CONVERT THE BASIC ARRAY DATA TO INTEGER
1655    REM:FORMAT FOR CONVERSION TO FOURIER DATA BLOCK
1656    PRINT
1658    PRINT "OUTPUT OF THE FILTER-OUTPUT OF CASCADED STAGES:"
1659    PRINT
1660    LET J=1
1670    LET K=0
1680    FOR I=0 TO A5
1690    LET B2=A0*J
1700    IF I=B2 THEN 1720
1710    GOTO 1740
1720    LET K=0
1730    LET J=J+1
1740    LET K=K+1
1750    LET B3=V[J,K]
1760    CALL DPUT(Z[1,1],K[1],I,B3,A0)
1780    NEXT I
1781    REM:PRINT THE OUTPUT DATA IN THE BASIC ARRAY.
1782    LET J=1
1783    FOR K=1 TO C1
1785    PRINT J,K,V[J,K]
1786    NEXT K
1790    REM:NOW DISPLAY THE DATA IN THE BASIC ARRAY
1795    PRINT
1800    PRINT "THE OUTPUT OF THE FILTER IS DISPLAYED"
1805    PRINT
1810    CALL DSPLY(Z[1,1],K[1],C,A5,0)
1820    PAUSE
1830    CALL NODIS
1834    PRINT
1870    REM:
```

```
1880   REM:
1890   REM:WRITE THE OUTPUT DATA ON TO THE DATA AREA IN THE DISK
1891   PRINT
1900   PRINT "TYPE THE RECORD NO. WHERE THE DATA IS TO BE STORED"
1910   INPUT B4
1920   CALL DWRIT(Z[1,1],K[1],B4,A4,A7)
1923   PRINT
1930   PRINT "DATA IS STORED IN THE DISK DATA AREA:";B4
1935   PRINT
1940   PRINT "WHERE WOULD YOU LIKE TO STORE THE INPUT DATA?"
1945   INPUT B7
1950   PRINT
1952   PRINT
1954   CALL EXIT
1956   END
```

```
5     CALL SCORE(6)
8     REM: PROGARAM LPF.BAS
10    REM:DIGITAL IMPLEMENTATION OF THE SEARLE'S 1/3 OCTAVE
12    REM:FILTER BANK MODEL BASED ON HUMAN AUDITION.
25    REM:
30    REM:ROUTINE TO FIND THE GOEF.AND TO IMPLEMENT THE FIRST
35    REM:ORDER RC LOW PASS FILTER (DETECTION FILTER) WHICH FOLLOWS
40    REM:THE BANDPASS FILTERS IN THE 1/3 OCTAVE FILTER BANK.
42    REM:
50    PRINT "WOULD YOU LIKE TO RUN THE PROGRAM OR EXIT? ENTER 1 TO EXIT"
60    INPUT A1
70    IF A1=1 THEN 1950
80    REM:CONSTANT VALUE DECLARATION
90    LET P1=3.14159
100   DIM P[16,128],X[16,128],Q[128],Y[16,128],Z[16,128]
120   REM:THE ARRAY SIZE ACCORDING TO THE SET BLOCK SIZE
122   LET D3=16
125   PRINT "PLEASE INPUT THE LOWER CUT OFF FREQUENCY IN HZ"
128   INPUT F1
130   PRINT "AND NOW THE UPPER CUT OFF FREQUENCY"
132   INPUT F2
135   REM:TO FIND THE RISE TIME OF THE LOW PASS FILTER
138   LET F4=(F1+F2)/2
140   LET T5=1/F4
143   REM:THE CONSTANT 4.49 IS OBTAINED FROM THE DESIGN CONSIDERATION
145   LET T6=4.5*T5
148   REM:TO FIND THE CUT OFF FREQUENCY OF THE LOW PASS FILTER
150   LET F5=(.35/T6)
155   PRINT "PLEASE INPUT THE SAMPLING FREQUENCY IN HZ.:"
157   INPUT F3
200   REM:TO CALCULATE THE NUMERATOR AND DENOMINATOR COEF OF,
210   REM:LOW PASS FILTER TRANSFER FUNCTION.
211   PRINT
215   REM: A: NUMERATOR COEF, B: DENOMINATOR COEF
216   PRINT
220   LET T1=1/F3
230   LET W1=TAN(P1*F5*T1)
235   PRINT "THE NUMERATOR AND DENOMINATOR COEF"
240   LET A=W1/(W1+1)
245   LET B=(W1-1)/(W1+1)
250   PRINT
260   PRINT " A "," B "
265   PRINT
270   PRINT A,B
275   PRINT
790   REM:THE FOURIER KEYBOARD PROGRAM IS USED TO COLLECT 2048
795   REM:POINTS OF DATA. THE SAMPLING PARAMETERS ARE SET ON THE
800   REM:CONTROL PANEL. THE FOURIER KEYBOARD PROGRAM RESIDES IN
805   REM:RECORD 0 OF THE PROGRAM AREA IN THE DISK $& DATA IS STORED
810   REM:IN RECORD 1 OF THE DATA AREA IN THE DISK.
815   PRINT
820   PRINT "FOURIER KEYBOARD PROGRAM"
```

```
825    PRINT
830    CALL FCORE(5)
833    PRINT
835    PRINT "THE BLOCK SIZE IS TAKEN TO BE 2048"
838    PRINT
840    LET A4=2048
850    LET A5=A4-1
855    REM:THE RECORD NUMBER WHERE INPUT DATA IS STORED IS : 0
860    LET A6=0
865    REM:READ THE FOURIER DATA INTO THE BASIC ARRAY IN
866    REM:INTEGER FORMAT
870    LET A7=1
880    CALL DREAD(P[1,1],K[1],A6,A4,A7)
881    REM:NOW DISPLAY THE DATA IN THE BASIC ARRAY
890    CALL DSPLY(P[1,1],K[1],0,A5,0)
900    PAUSE
910    CALL NODIS
920    REM:ROUTINE TO CONVERT THE FOURIER DATA (INTEGER FORMAT)
925    REM:IN THE BASIC ARRAY TO FLOATING POINT FORMAT FOR BASIC
930    REM:PROGRAM MANIPULATION. THE ROUTINE HANDLES BLOCK SIZE > 256
935    REM:J,K,ARE THE ARRAY VARIABLES.
940    REM:CONSTANT VALUE DECLARATION
950    LET J=1
960    LET K=0
970    LET A8=0
980    LET A9=128
984    PRINT
985    PRINT "INPUT TO THE STAGE NO:1"
986    PRINT
990    FOR I=0 TO A5
1000   CALL DGET(P[1,1],K[1],I,B1,A8)
1010   LET B2=A9*J
1020   IF I=B2 THEN 1040
1030   GOTO 1060
1040   LET J=J+1
1050   LET K=0
1060   LEY K=K+1
1070   LET X[J,K]=B1
1090   NEXT I
1100   REM:PRINT THE VALUES IN THE INPUT DATA ARRAY
1111   LET J=1
1115   FOR K=1 TO C1
1120   PRINT J,K,X[J,K]
1125   NEXT K
1130   REM:
1140   REM:ROUTINE TO REALIZE THE LOWPASS FILTER FUNCTION
1145   REM:
1150   REM:                A*(1+Z^-1)
1155   REM:    H(Z)= ----------------
1160   REM:                (1+B*Z^-1)
1165   REM:
1170   REM:IF X(N) AND Y(N) ARE THE INPUT AND OUTPUT
1175   REM: THE FOLLOWING DIFFERENCE EQUATION IS SATISFIED
```

```
1180   REM:
1185   REM: Y(N)= A*X(N)+A*X(N-1)-B*Y(N-1)
1190   REM:
1200   LET Q[1]=A*X[1,1]
1210   FOR I = 1 TO D3
1220   FOR J = 2 TO 128
1230   LET Q[J]=A*X[I,J]+A*X[I,J-1]-B*Q[J-1]
1240   NEXT J
1250   FOR L=1 TO 128
1260   LET Y[I,L]=Q[L]
1270   NEXT L
1280   IF I = D3 THEN 1300
1290   LET Q[1]=A*X[I+1,1]+A*X[I,128]-B*Q[128]
1300   REM:
1310   NEXT I
1315   REM:ROUTINE TO GENERATE THE BLOCK QUALIFIER INFORMATION
1318   REM:U(1) :BS; U(2) : SF; U(3) : CALIBRATOR; U(4) : FREQ CODE
1320   REM:U(5) : COORDINATE CODE
1325   LET U(1) = 2048
1330   LET U(2) = -15
1335   LET U(3) = 32767
1340   LET U(4) = 48
1345   LET U(5) = 0
1350   CALL FIXQ(U[1])
1650   REM:ROUTINE TO CONVERT THE BASIC ARRAY DATA TO INTEGER
1655   REM:FORMAT FOR CONVERSION TO FOURIER DATA BLOCK
1656   PRINT
1658   PRINT "OUTPUT OF THE FILTER-OUTPUT OF CASCADED STAGES:"
1659   PRINT
1660   LET J=1
1670   LET K=0
1680   FOR I=0 TO A5
1690   LET B2=A8*J
1700   IF I=B2 THEN 1720
1710   GOTO 1740
1720   LET K=0
1730   LET J=J+1
1740   LET K=K+1
1750   LET B3=V[J,K]
1760   CALL DPUT(Z[1,1],K[1],I,B3,A8)
1780   NEXT I
1781   REM:PRINT THE OUTPUT DATA IN THE BASIC ARRAY.
1782   LET J=1
1783   FOR K=1 TO C1
1785   PRINT J,K,V[J,K]
1786   NEXT K
1790   REM:NOW DISPLAY THE DATA IN THE BASIC ARRAY
1795   PRINT
1800   PRINT "THE OUTPUT OF THE FILTER IS DISPLAYED"
1805   PRINT
1810   CALL DSPLY(Z[1,1],K[1],C,A5,0)
1820   PAUSE
1830   CALL NODIS
```

```
1834   PRINT
1870   REM:
1880   REM:
1890   REM:WRITE THE OUTPUT DATA ON TO THE DATA AREA IN THE DISK
1891   PRINT
1900   PRINT "TYPE THE RECORD NO. WHERE THE DATA IS TO BE STORED"
1910   INPUT B4
1920   CALL DWRIT(Z[1,1],K[1],B4,A4,A7)
1923   PRINT
1930   PRINT "DATA IS STORED IN THE DISK DATA AREA:";B4
1935   PRINT
1940   PRINT "WHERE WOULD YOU LIKE TO STORE THE INPUT DATA?"
1945   INPUT B7
1950   PRINT
1952   PRINT
1954   PRINT
1956   CALL EXIT
1958   END
```

APPENDIX B

FOURIER KEYBOARD PROGRAMMES USED IN THE DIGITAL IMPLEMENTATION OF THE SEARLE'S MODEL BASED ON HUMAN AUDITION.

These programmes are also implemented in the HP 5451B Fourier Analyzer System.

1. Real time data acquisition.
2. Generation of the impulse response for realizing the bandpass filter input/output equation.
3. Fourier program to input the impulse response.
4. Overlapping input data block.
5. Implementation of the filter detector.
6. Filter bank summation.
7. Filter bank summation with alternate channels phase shifted.
8. Extracting the data for conversion to analog signal.
9. Digital to analog signal conversion.
10. Generation of the sinusoidal carrier frequency for reconstruction.
11. Reconstruction of the output of the model.
12. Plotting programme.

THIS APPENDIX GIVES THE IMPORTANT FOURIER KEYBOARD PROGRAMMES USED IN THE DIGITAL IMPLEMENTATION OF THE SEARLE'S MODEL BASED ON HUMAN AUDITION USING THE HP 5451B FOURIER ANALYZER SYSTEM

REAL TIME DATA ACQUISITION

COMMAND USED TO ACQUIRE THE REAL TIME DIGITAL DATA. THE COMMAND IS USED AFTER SETTING THE NECESSARY SAMPLING PARAMETERS AT THE FRONT PANEL OF THE FOURIER ANALYZER SYSTEM

N : NO OF SPEECH DATA BLOCKS TO BE DIGITIZED

```
MS      22      1       N
```

GENERATION OF THE IMPULSE RESPONSE FOR REALIZING THE BANDPASS FILTER INPUT-OUTPUT EQUATION.

```
K       0       0       0
K       -4      0       46
I       0
```

-4 is the scale factor & 46 is the frequency code corresponding to the maximum frequency (20 KHz)
I is the amplitude of the impulse.

FOURIER PROGRAM TO INPUT THE IMPULSE RESPONSE

THIS PROGRAMME IS CALLED BY THE BASIC PROGRAM WHICH REALIZES THE BANDPASS FILTER. THIS PROGRAM RESIDES IN STACK N (Refer BASIC programme for further details)

```
I       L       0
5       BS      1024
9       MS      31      2
14      MS      11
18      MS      31      0
23
```

```
!     OVERLAPPING THE INPUT DATA BLOCK
!
!     THIS PROGRAMME CONVERTS THE DIGITIZED INPUT DATA BLOCK
!     OF BLOCK SIZE 2048 TO A BLOCK SIZE OF 4096 BY OVERLAPPING
!     FOR THE IMPLEMENTATION OF THE MODEL.
!     N   : No of data blocks to be processed
!     REC : record number on the tape to read/write
!
      1        L        0
      5        MS       32
      9        MS       32       REC
      14       BS       2048
      18       Y        8823     REC
      23       Y        3010     1
      28       MS       12
      32       X>       1
      36       MS       12
      40       X        1
      44       BS       4096
      48       Y        8822
      52       L        1
      56       BS       2048
      60       MS       12
      64       X        1
      68       BS       4096
      72       Y        8822
      76       Φ        1        N
      82       .
```

```
I       IMPLEMENTATION OF THE FILTER DETECTOR

I       THIS PROGRAMME INPUTS THE DIGITIZED SPEECH DATA
I       BLOCK,CONVOLVES IT WITH THE BANDPASS FILTER IMPULSE
I       RESPONSE,FINDS THE ABSOLUTE MAGNITUDE VALUE AND
I       CONVOLVES IT WITH THE LOWPASS FILTER IMPULSE
I       RESPONSE.THE OUTPUT THUS OBTAINED IS STORED
I       IN THE MAGNETIC TAPE.

I       NFLTR   : NO OF FILTERS IN THE MODEL
I       NREC    : RECORD NUMBER WHERE IMPULSE RESPONSE
I         RESPONSE IS STORED
I       REC     : RECORD NUMBER TO READ/WRITE THE DATA
I       NTIMES  : NUMBER OF SPEECH DATA BLOCKS TO BE
I         PROCESSED

        1       L       0
        5       BS      4096
        9       Y       0023    REC
        14      Y       0033    REC
        18      L       1
        22      Y       0031
        26      X>      2
        30      MS      31      NREC
        35      L       2
        39      MS      11
        43      CV      2
        47      Y       3001
        51      X>      1
        55      MS      11
        59      CV      1
        63      *       0       1000
        68      :       8       030
        73      Y       0022
        77      ◆       2       NFLTR
        83      ◆       1       NTIMES
        89      .
```

```
!       FILTER BANK SUMMATION
!
!       THIS PROGRAMME IMPLEMENTS THE FILTER BANK AND
!       SUMS THE OUTPUT FOR CONVERSION TO ANALOG SIGNAL
!
!       REC    : RECORD NUMBER ON THE TAPE TO READ/WRITE
!       NREC : RECORD NUMBER WHERE BANDPASS IMPULSE RESPONSE
!         IS STORED
!       NFLTR: NUMBER OF FILTERS IN THE MODEL
!       NTIMES: NUMBER OF DATA BLOCKS TO BE PROCESSED
!

1       L       0
5       BS      4096
9       Y       8823    REC
14      Y       8833    REC
18      L       1
22      CL      2
26      Y       8821
30      X>      1
35      MS      31      NREC
39      L       2
43      MS      11
47      CV      1
51      A+      2
55      X>      2
59      ♦       2       NFLTR
63      Y       8832
68      ♦       1       NTIMES
73      .
```

```
FILTER BANK SUMMATION WITH ALTERNATE CHANNELS
PHASE SHIFTED

REC    : RECORD NUMBER ON THE MAGNETIC  TAPE TO READ/WRITE
NTIMES: NUMBER OF SPEECH DATA BLOCKS TO BE PROCESSED


1       L       0
5       BS      4096
9       Y       0023    REC
14      Y       0033    REC
18      L       1
23      CL      1
27      L       2
31      Y       0021
35      A+      1
39      X>      1
43      Y       0021
47      *       0       -1
52      A+      1
56      X>      1
60      #       2       8
66      Y       0032
70      #       1       NTIMES
76
```

```
|     EXTRACTING THE DATA FOR CONVERSION TO
|     ANALOG SIGNAL
|
|     THIS PROGRAM EXTRACTS THE CORRECT DATA BLOCK OF
|     BLOCKSIZE OF 2048 FROM THE OVERLAPPED DATA BLOCK OF
|     BLOCKSIZE 4096 FOR DIGITAL TO ANALOG CONVERSION
|     4096 FOR DIGITAL TO ANALOG CONVERSION
|
|     NTIMES : NUMBER OF SPEECH DATA BLOCKS TO BE PROCESSED

      1       L       0
      5       BS      4096
      10      Y       8831
      15      BS      2048
      19      X<      1
      23      Y       8822
      27      ♦       0        NTIMES
      31      .
```

```
|
|     DIGITAL TO ANALOG SIGNAL CONVERSION
|
|     THIS PROGRAMME OUTPUTS THE PROCESSED SPEECH DATA
|     BLOCK THROUGH THE DIGITAL TO ANALOG CONVERTER AND THE
|     ANALOG SIGNAL IS TAPED
|
|     NTIMES : NO OF SPEECH DATA BLOCKS TO BE PROCESSED

      1       L       0
      5       Y       8821
      9       B       0        1
      14      RA      1
      18      ♦       0        NTIMES
      24      .
```

```
|
|    GENERATION OF THE SINUSOIDAL CARRIER FREQUENCY
|
|    THIS COMMANDS GENERATES THE SINUSOIDAL CARRIER FREQUENCY
|    USED FOR THE RECONSTRUCTION OF THE SIGNAL FROM THE
|    OUTPUT OF THE FILTER DETECTOR
|
|    X  =  (Fc * 2048)/10000
|    X  : PARAMETER FOR GENERATING THE DESIRED FREQUENCY
|    Fc : DESIRED CARRIER FREQUENCY
|    2048 : REAL DATA POINTS
|    10000 : NORMALIZING FACTOR

     K       0       X       X
     K       -4      4       Fs
     0       5000
     F

|    X       : CALCULATED PARAMETER
|    Fs      : SAMPLING FREQUENCY
|    5000    : NORMALIZING FACTOR
```

```
!
!       RECONSTRUCTION OF THE OUTPUT OF THE MODEL
!
!       THIS PROGRAMME MODULATES THE OUTPUT OF THE FILTER
!       DETECTOR WITH THE SINUSOIDAL CARRIER FREQUENCY
!
!       REC     : RECORD NUMBER TO READ/WRITE THE DATA
!       NREC    : RECORD NUMBER WHERE SINUSOIDAL CARRIER
!         FREQUENCIES ARE STORED
!       NFLTR : NUMBER OF FILTERS IN THE MODEL
!       NTIMES : NUMBER OF SPEECH DATA BLOCKS TO BE PROCESSED
!

1       L       0
5       BS      4096
9       Y       8823    REC
14      Y       8833    REC
19      L       1
23      MS      31      NREC
28      L       2
32      MS      11
36      X>      1
40      Y       8821
44      *       1
48      Y       8832
52      #       2       NFLTR
58      #       1       NTIMES
64      .
```

PLOTTING PROGRAMME

THIS PROGRAM PLOTS THE INPUT SIGNAL AND ITS
MAGNITUDE SPECTRUM

```
1    L  0
4    BS 4096
7    Y  5821  6
11   Y  5809  0  0  1
14   Y  5814
18   Y  5807  -1
24   Y  5805  1  500
27   Y  5804  1000  500
31   Y  5800  1
36   F
41   TP
45   Y  5810  1000  0
47   Y  5805  1  1
49   Y  5800  1
54
```

APPENDIX C

FORTRAN - 77 PROGRAMMES USED IN SIGNAL PROCESSING - DATA REDUCTION USING DISCRETE COSINE TRANSFORM AND FOR OBTAINING THE RUNNING SPECTRUM PLOT.

These programmes are used in PDP 11/60 minicomputer.

1. FTREAD: This routine reads the "Fourier Analyzer" processed data (implementation of the model) and transforms it into readable PDP 11/60 data for subjecting the data to data reduction using discrete cosine transform.

2. FTPROCESS: This routine reads the data into an array and subjects the log magnitude data to data reduction using discrete cosine transform. The reduced data is reconstructed using inverse discrete cosine transform and is stored as data files.

3. FTWRITE: This routine transfers the processed data to magnetic tape as Fourier readable data for reconstruction using Fourier Analyzer system.

4. PLOT: This routine reads the filter-detector output and plots the running spectrum to visualize the temporal and spectral characteristics of speech.

5. LPLOTIDCT: This routine reads the filter detector output and subjects the log magnitude data to data reduction using discrete cosine transform. The transformed data is reconstructed using inverse discrete cosine transform and used to plot the running spectrum to visualize the temporal and spectral characteristics of speech subjected to data reduction.

6. ORTH: This routine is written to test the orthogonal property of the discrete cosine transform.

```
!       PROGRAM FTREAD.FTN
!
!       THIS ROUTINE READS MASTAPE DATA FROM THE FOURIER ANALYZER
!       AS BS 4096 AND TRANSFORMS IT INTO READABLE PDP11/60 DATA
!       AS DIRECT ACCESS RECORDS OF RECORD LENGTH 512
!
!       FOR EACH TAPEBLOCK
!       * READ A BLOCK OF DATA (Default)
!       * REVERSE THE BYTES IN THE BLOCKWORD SIZE
!       * EXTRACT THE SCALE FACTOR & CALIBRATOR
!       * ACTUAL DATA = (FOURIER DATA*CAW)/(32767*32767)*10**SFW
!       * WRITE THE DATA TO A DIRECT ACCESS FILE
!       REPEAT N BLOCKS
!
!       INPUT: FOURIER BLOCKSIZE
!       REC NO ON THE TAPE WHERE DATA IS TO BE READ FROM
!       (Stored in data file 'FTREAD.FTN')
!       DATA FILE NAME TO STORE THE DATA.
!       NO OF DATA BLOCKS TO BE TRANSFERED TO DISK
!       STARTING RECORD NUMBER
!
!       OUTPUT: READABLE PDP11/60 DATA FILES FOR PROCESSING
!
        BYTE TEMP,BUF(8200)
        INTEGER DATA(4100)
        INTEGER BLKSIZ,  !(* BLOCKSIZE IN BYTES *)
     1          SFW,     !(* SCALE FACTOR WORD  *)
     1          CAW,     !(* BLOCK CALIBRATOR WORD *)
     1          NPTS     !(* # OF ACTUAL DATA PTS. PER BLOCK *)
        EQUIVALENCE (DATA(1),BUF(1))
        REAL ACTDAT(512)
!
!       CONSTANT VALUE DECLARATION
!
        FACT = 32767.
        IBLK = 4096
        NFLTR = 16
        NUM = 1  !(STARTING REC NO)
        NREC = 4          !( NO OF SPEECH DATA BLOCKS )
!
!       INPUT THE PARAMETERS
!
        OPEN(UNIT=3,NAME='FTREAD.DAT',TYPE='OLD')
        READ(3,*)NPEC,IDUM
        CLOSE(UNIT=3)
        TYPE*,'THE NO OF SPEECH DATA BLOCKS TO BE TRANSFERED:',NREC
        TYPE*,'THE FOURIER BLOCKSIZE IS 4096:'
        TYPE*,'THE DATA FILE IS ASSIGNED IS : SPEC.DAT'
!
!       REWIND & POSITION THE TAPE TO READ THE DATA
!
        BLKSIZ = (IBLK*2)+8
```

```
      NPTS = (BLKSIZ-6)/2
      TYPE *,'NPTS = ',NPTS
      NBLOK = 512
      NSEG = IBLK/NBLOK
      CALL MAGMAG
      CALL STCMAG
      CALL RWDMAG
      CALL SPBMAG(NPEC)
      CALL STAMAG(JREC,ISTAT)
      WRITE(6,300)
300   FORMAT(/,10X,'ROUTINE TO READ DATA FROM MAGTAPE',//)
      WRITE(6,302)
302   FORMAT(5X,'MAGTAPE REC NO :')
      WRITE(6,*)NPEC,JREC
!
!     READ THE TAPE DATA REVERSE THE BYTES & FIND THE ACTUAL DATA
!
      OPEN(UNIT=7,NAME='SPEC',ACCESS='DIRECT',STATUS='NEW',
     1        RECORDTYPE='FIXED',FORM='UNFORMATTED',RECL=512)
      DO 20 LL = 1,NREC
      DO 10 J = 1,NFLTR
      CALL RLBMAG(BUF,BLKSIZ)
        DO 12 K = 1 , BLKSIZ ,2
          TEMP = BUF(K)
          BUF(K) = BUF(K+1)
          BUF(K+1) = TEMP
12      CONTINUE
        SFW = DATA(NPTS+1)
        CAW = DATA(NPTS+2)
!     (*SHIFT SFW RIGHT 6 BITS TO GET SCALE FACTOR *)
!       SFW = IISHFT(SFW,-6)
        SFW = SFW/64
!
!     WRITE THE DATA TO DIRECT ACCESS RECORD
!
      DO 14 KK = 1,NSEG
      DO 16 I = 1,NBLOK
      KD = (NBLOK*(KK-1)+I)
      FLDAT = FLOAT(DATA(KD))
      ACTDAT(I)=((FLDAT*FLOAT(CAW))/(FACT*FACT))*10.**FLOAT(SFW)
16    CONTINUE
      WRITE(7,REC=NUM)(ACTDAT(M),M=1,NBLOK)
      NUM = NUM+1
14      CONTINUE
      TYPE*,'NO OF FILTER OUTPUTS READ IN:',J
10    CONTINUE
      NU1 = NUM-1
      TYPE*,'NO OF SPEECH DATA BLOCKS READ IN:',LL
      TYPE*,'TOTAL NO OF RECORDS TRANSFERED :',NU1
      WRITE(6,310)
310   FORMAT(/,5X,'NO OF SPEECH DATA BLOCKS READ IN :')
      WRITE(6,*)LL
      WRITE(6,312)
```

```
312     FORMAT(/,5X,'TOTAL NO OF DIRECT ACCESS RECS TRANSFERED :')
        WRITE(6,*)NU1
20      CONTINUE
        OPEN(UNIT=4,NAME='FTREAD.DAT',TYPE='NEW')
        NPEC = NPEC+64
        WRITE(4,*) NPEC,IDUM
        CLOSE (UNIT=4)
        CLOSE (UNIT=7)
        STOP
        END
```

```
!       PROGRAM FTPROCESS.FTN
!
!       THIS ROUTINE READS THE FILTER DETECTOR OUTPUT DATA FROM THE
!       DIRECT ACCESS DATA FILE AND STORES THE DATA IN THE ARRAY.THE
!       DATA IN THE ARRAY IS SUBJECTED TO DCT.THE REQUIRED NO OF
!       CHANNELS ARE SET TO ZERO AND THE DATA IS SUBJECTED TO IDCT
!       THE PROCESSED DATA IS WRITTEN ON TO A DIRECT ACCESS DATA FILE.
!
!       INPUT    : No of channels to be set to zero
!
!       OUTPUT   : Processed data (DCT/IDCT) in direct
!         access files
!
!
!       CONSTANT VALUE DECLARATION
!
        REAL A(512,16),B(16)
        NBLOK = 512
        NFLTR = 16
        PI = 3.1415927
        NSEG = 8
        IOFS = 1
        NUM = 1 !(*STARTING REC NO FOR DATA RETRIEVAL)
        NSTO = 1!(*STARTING REC NO FOR STORING)
        ZERO = 0.0
        CONST = 0.000001 !(TO AVOID LOG OF ZERO)
        NREC = 4!(NO OF SPEECH DATA BLOCKS TO BE PROCESSED)
        NZERO = 14      !(NO OF COMP SET TO ZERO)
!
!       INPUT THE PARAMETERS
!
        TYPE*,'NO OF SPEECH DATA BLOCKS TO BE PROCESSED:',NREC
        TYPE*,'THE DATA IS RETRIEVED FROM: SPEC.DAT'
        TYPE*,'& STORED IN: PROCESS.DAT'
        TYPE*,'ENTER THE NO OF CHANNELS TO BE SET TO ZERO:',NZERO

        WRITE(6,300)
300     FORMAT(/,10X,'DATA PROCESSING ROUTINE',//)
        WRITE(6,302)
302     FORMAT(5X,'NO OF CHANNELS SET TO ZERO :')
        WRITE(6,*)NZERO

        OPEN(UNIT=7,NAME='SPEC',ACCESS='DIRECT',STATUS='OLD',
     1        RECORDTYPE='FIXED',FORM='UNFORMATTED',RECL=512)
        OPEN(UNIT=8,NAME='PROCESS',ACCESS='DIRECT',STATUS='NEW',
     1        RECORDTYPE='FIXED',FORM='UNFORMATTED',RECL=512)
!
!       READ THE DATA IN TO THE ARRAY
!
        DO 110 KK = 1,NREC
        DO 100 JJ = 1,NSEG
        DO 24 K = 1,NFLTR
        READ(7,REC=NUM)(A(ID,K),ID=1,NBLOK)
```

```
      NUM = NUM+NSEG
24    CONTINUE
      TYPE*,'DATA IS READ INTO ARRAY FOR PROCESSING'
      TYPE*,'FOR SEG & SPEECH BLOCK:',JJ,KK
!
!     FIND THE LOG MAGNITUDE OF THE VALUES IN THE ARRAY
!
      DO 52 K = 1,NFLTR
      DO 52 ID = 1,NBLOK
      BDATA = A(ID,K)
      IF (BDATA.LE.ZERO)BDATA=CONST
      FDATA = 0.43429448*ALOG(BDATA)
      A(ID,K) = FDATA
52    CONTINUE
!
!     FIND DCT COEFS OF THE DATA IN THE ARRAY
!
      DO 10 ID = 1,NBLOK
      SUM1 = 0.
      DO 20 K = 1,NFLTR
      SUM1 = A(ID,K)+SUM1
20    CONTINUE
      K = 1
      B(K) = ((1.4142136/FLOAT(NFLTR))*SUM1)
      DO 30 K = 2,NFLTR
      SUM2 = 0.
      DO 40 M = 1,NFLTR
      THETA = ((2.*FLOAT(M-1)+1.)*FLOAT(K-1)*PI)/(2.*FLOAT(NFLTR))
      XA = COS(THETA)
      XK = A(ID,M)*XA
      SUM2 = XK+SUM2
40    CONTINUE
      B(K) = ((2./FLOAT(NFLTR))*SUM2)
30    CONTINUE
      DO 50 K = 1,NFLTR
      A(ID,K)=B(K)
50    CONTINUE
10    CONTINUE
!
!     SET THE DCT COMPONENTS TO ZERO FOR DATA REDUCTION
!
      NSET = NFLTR-NZERO+1
      DO 35 ID = 1,NBLOK
      DO 45 LL = NSET,NFLTR
      A(ID,LL) = 0.0
45    CONTINUE
35    CONTINUE
!
!     INVERSE DISCRETE COSINE TRANSFORM
!
      DO 60 ID = 1,NBLOK
      DO 70 M = 1,NFLTR
      K = 1
```

```
      GX0 = (1./1.4142136)*A(ID,K)
      SUM3 = 0.
      DO 80 K = 2,NFLTR
      THETA1 = ((2.*FLOAT(M-1)+1.)*FLOAT(K-1)*PI)/(2.*FLOAT(NFLTR))
      GX = A(ID,K)*COS(THETA1)
      SUM3 = GX+SUM3
80    CONTINUE
      B(M) = GX0+SUM3
70    CONTINUE
      DO 90 K = 1,NFLTR
      A(ID,K) = B(K)
90    CONTINUE
60    CONTINUE
      TYPE*,'DATA IN THE ARRAY IS PROCESSED,IDCT:'
      TYPE*,'FOR SEG & SPEECH BLOCK:',JJ,KK
!
!     CONVERSION TO LINEAR SCALE BY TAKING ANTILOG
!     (For Reconstruction)
!
      DO 57 K = 1,NFLTR
      DO 57 ID = 1,NBLOK
      EDATA = A(ID,K)
      GDATA = EXP(2.3025851*EDATA)
      A(ID,K) = GDATA
57    CONTINUE
!
!     WRITE THE DATA IN THE ARRAY TO A FILE
!
      DO 77 K = 1,NFLTR
      WRITE(8,REC=NSTO)(A(ID,K),ID=1,NBLOK)
      NSTO = NSTO+IOFS
77    CONTINUE
      NST1 = NSTO-1
      NUM = NUM-(NFLTR*NSEG)
      NUM = NUM+1
100   CONTINUE
      NUM = NUM+(NFLTR-1)*NSEG
      TYPE*,'NO OF SPEECH DATA BLOCKS PROCESSED:',KK
      TYPE*,'TOTAL NO OF RECS TRANSFERRED:',NST1
      WRITE(6,310)
310   FORMAT(/,5X,'NO OF SPEECH DATA BLOCKS PROCESSED :')
      WRITE(6,*)KK
      WRITE(6,312)
312   FORMAT(/,5X,'TOTAL NO OF RECS TRANSFERED :')
      WRITE(6,*)NST1
110   CONTINUE
      CLOSE(UNIT=7)
      CLOSE(UNIT=8)
      STOP
      END
```

```
! PROGRAM FTWRITE.FTN
!
! THIS ROUTINE READS THE PROCESSED PDP11/60 DATA FROM DIRECT
! ACCESS FILE AS RECORD LENGTH 512 & TRANSFORMS TO READABLE
! FOURIER DATA BLOCK OF BS 4096 ON TO THE MAGTAPE.
!
! FOR EACH FOURIER DATA BLOCK
! * READ THE DATA FROM THE DIRECT ACCESS FILE
! * FIND THE SCALE FACTOR & CALIBRATION FACTOR
! * FOURIER DATA = (PDP DATA*(32767*32767))/((10**SFW)*CAW)
! * REVERSE THE BYTES IN THE BLOCK WORD SIZE
! * WRITE TO THE TAPE
! REPEAT FOR N BLOCKS
!
! INPUTS: THE FOURIER BLOCK SIZE (Default)
!  REC NO ON THE TAPE WHERE DATA IS TO BE WRITTEN
! (Retrieved from FTREAD.DAT - Parameter passed by
! FTREAD.FTN)
!  DATA FILE NAME FROM WHERE DATA IS TO BE RETRIEVED
! (Default)
!  NO OF DATA BLOCKS TO BE TRANSFERRED TO TAPE
!
! OUTPUT: FOURIER READABLE DATA BLOCKS WRITTEN ON THE MAG TAPE

      BYTE TEMP,BUF(8200)
      INTEGER DATA(4100),KDATA(8)
      INTEGER BLKSIZ, !(* BLOCKSIZE IN BYTES *)
     1        SFW,    !(* SCALE FACTOR WORD  *)
     1        CAW,    !(* BLOCK CALIBRATOR WORD *)
     1        NPTS !(* # OF ACTUAL DATA PTS. PER BLOCK *)
      EQUIVALENCE (DATA(1),BUF(1))
      REAL ACTDAT(512),BDATA(8)
!
! CONSTANT VALUE DACLARATION
!
      FACT = 32767.
      AMNV = 0.1
      AMXV = 1.
      NFLTR = 16
      NBLOK = 512
      IFCODE = 48
      IBLK = 4096
      NUM = 1 !(STARTING REC NO :)
      NREC = 41(NO OF SPEECH DATA BLOCKS TO BE TRANSFERED)
!
! INPUT THE PARAMETERS
!
      OPEN(UNIT=3,NAME='FTREAD.DAT',ACCESS='SEQUENTIAL',TYPE='OLD')
      READ(3,*)IDUM,NPEC
      CLOSE(UNIT=3)
      TYPE*,'DATA IS RETRIEVED FROM :PROCESS.DAT'
      TYPE*,'NO OF SPEECH DATA BLOCKS TO BE TRANSFERED:',NREC
```

```
      TYPE*,'REC NO TO WRITE THE DATA',NPEC
!
!     REWIND AND POSITION THE TAPE TO WRITE THE DATA
!
      BLKSIZ = (IBLK*2)+6
      NPTS = (BLKSIZ-6)/2
      TYPE *,'NPTS = ',NPTS
      NSEG = IBLK/NBLOK
      CALL MAGMAG
      CALL STCMAG
      CALL RWDMAG
      CALL SPBMAG(NPEC)
      CALL STAMAG(JREC,ISTAT)
      WRITE(6,300)
300   FORMAT(/,10X,'ROUTINE TO WRITE TO TAPE',//)
      WRITE(6,302)
302   FORMAT(5X,'MAGTAPE REC NO :')
      WRITE(6,*)NPEC,JREC
!
!     TO FIND THE SCALE FACTOR WORD
!     FIND THE MAX VALUE IN THE DATA BLOCK & FIX THE SCALE FACTOR
!
      OPEN(UNIT=8,NAME='PROCESS',ACCESS='DIRECT',STATUS='OLD',
     1    RECORDTYPE='FIXED',FORM='UNFORMATTED',RECL=512)
      DO 15 MM = 1,NREC
      DO 10 J = 1,NFLTR
      VAL = 0.
      ISFACT = 0
      DO 16 K = 1,NSEG
      READ(8,REC=NUM)(ACTDAT(JJ),JJ=1,NBLOK)
      DO 11 LL = 1,NBLOK
      CDATA = ABS(ACTDAT(LL))
      IF (CDATA.GT.VAL)VAL=CDATA
11    CONTINUE
      NUM = NUM+NFLTR
16    CONTINUE
      NUM = NUM-(NFLTR*NSEG)
      IF (VAL.GT.AMXV)GOTO 66
      IF (VAL.GT.AMNV)GOTO 68
7     VAL = 10.*VAL
      ISIGN = 1
      ISFACT = 1+ISFACT
      IF (VAL.GT.AMNV.AND.VAL.LT.AMXV)GOTO 68
      GOTO 7
66    VAL = VAL/10.
      ISIGN = 2
      ISFACT = -1+ISFACT
      IF (VAL.GT.AMNV.AND.VAL.LT.AMXV)GOTO 68
      GOTO 66
68    IF (ISIGN.EQ.1)SFW = -ISFACT
      IF (ISIGN.EQ.2)SFW = -ISFACT
      CAW = FACT
```

```
      DATA(NPTS+1)=SFW*64
      DATA(NPTS+2)=CAW
      DATA(NPTS+3)=IFCODE
!
!     CONVERSION TO FIXED POINT FOURIER DATA
!
      DO 12 KK = 1,NSEG
      READ(6,REC=NUM)(ACTDAT(JJ),JJ=1,NBLOK)
      DO 78 LL = 1,NBLOK
      FLDAT=(ACTDAT(LL)*FACT*FACT)/(FLOAT(CAW)*10.**FLOAT(SFW))
      KD = NBLOK*(KK-1)+LL
      DATA(KD) = FLDAT
78    CONTINUE
      NUM = NUM+NFLTR
12    CONTINUE
!
!     REVERSE THE BYTES IN THE BLOCKWORD SIZE
!
      DO 13 K = 1,BLKSIZ,2
      TEMP = BUF(K)
      BUF(K) = BUF(K+1)
      BUF(K+1) = TEMP
13    CONTINUE
!
!     WRITE THE FOURIER DATA TO THE MAGTAPE
!
      CALL WLBMAG(BUF,BLKSIZ)
      NUM = NUM-(NSEG*NFLTR)
      NUM = NUM+1
      TYPE*,'NO OF FILTER OUTPUTS TRANSFERRED:',J
10    CONTINUE
      TYPE*,'NO OF SPEECH DATABLOCKS TRANSFERRED:',MM
      NUM = NUM+(NSEG-1)*NFLTR
      NU1 = NUM-1
      TYPE*,'TOTAL NO OF RECS TRANSFERED :',NU1
      WRITE(6,310)
310   FORMAT(/,5X,'NO OF SPEECH DATA BLOCKS TRANSFERED :')
      WRITE(6,*)MM
      WRITE(6,312)
312   FORMAT(/,5X,'NO OF RECS TRANSFERED :')
      WRITE(6,*)NU1
15    CONTINUE
      CLOSE(UNIT=8)
      OPEN(UNIT=4,NAME='FTREAD.DAT',ACCESS='SEQUENTIAL',TYPE='NEW')
      NPEC=NPEC+64
      WRITE(4,*)IDUM,NPEC
      CLOSE(UNIT=4)
      STOP
      END
```

```
!     PROGRAM PLOT.FTN
!
!     THIS ROUTINE READS THE FILTER DETECTOR OUTPUT DATA FROM THE
!     DIRECT ACCESS FILE AND PLOTS THE RUNNING SPECTRUM
!
!
!     INPUT   : SCALE FACTOR
!       X & Y AXIS TITLES
!       STARTING NUMBER FOR X AND Y AXIS
!     (All the other plotting parameters were
!     Default values)
!
!     CONSTANT VALUE DECLARATION
!
      REAL A(256,16),P(500),PMAX(500),B(16)
      BYTE ALPH(80),ISMY(20),ISMX(20),ISMB(20)
      NARR = 2
      NBLOK = 512/NARR
      IBLOK = 2048
      NFLTR = 16
      PI = 3.1415927
      NUM = 1
      NAVE = 32
      NSEG = 4!(CORROSPONDS TO NO OF SPEECH BLOK TO PLOT)
      NP = 10 !(NO OF POINTS FOR INTERPOLATION)
      DX = 0.08       !(INCREMENT FOR X AXIS )
      IY = 1  !(INCREMENT FOR Y AXIS)
      XH = 0.3!(WIDTH OF THE CHARECTOR)
      YH = 0.3!(HEIGHT OF THE CHARECTER)
      YANG = 270.0     !(ANGLE FOR Y SHIFT)
      IANG = 270
      XANG = 320.0
      NCHAR = 20
      DNH = 0.1
      DH = 0.3
      XSHFTN = 0.3
      XSHFTL = 0.8
      YSFTLX = 6.
      ISX = 7
      ISY = 6
      DBH = 0.1
      DMULT = 2.54
      YSHFTB = 4.
      YSFTBL = 7.
      DBNH = 0.3
      X1 = 1.
      NBTIC = 25
      YMAX = 25.0      !(INITIAL POSITION TO START)
      YDIS = 6.0       !(WIDTH OF THE PLOT)
      NXTIC = 16       !(NO OF TIC MARKS FOR X AXIS)
      NYTIC = 4        !(NO OF TIC MARKS FOR Y AXIS)
```

```
C
C       INPUT THE PARAMETERS
C
        TYPE*,'DATA IS RETRIEVED FROM :THEPRO.DAT'
        TYPE*,'& THE SCALE FACTOR :'
        ACCEPT*,SFACT
        TYPE*,'NBLOK :',NBLOK
        TYPE*,'ENTER TITLE OF X AXIS:'
        ACCEPT132,(ISMX(K),K=1,20)
132     FORMAT(20A1)
        TYPE*,'ENTER TITLE OF Y AXIS:'
        ACCEPT133,(ISMY(K),K=1,20)
133     FORMAT(20A1)
        TYPE*,'ENTER TITLE OF THE BLOCK:'
        ACCEPT134,(ISMB(K),K=1,20)
134     FORMAT(20A1)
        TYPE*,'INPUT THE STARTING NO FOR X AXIS:'
        ACCEPT*,NUMX
        TYPE*,'& FOR THE SPEECH DATA BLOCK:'
        ACCEPT*,NUMB
C
C
        OPEN(UNIT=7,NAME='THEPRO',ACCESS='DIRECT',STATUS='OLD',
     1       RECORDTYPE='FIXED',FORM='UNFORMATTED',RECL=64)
C
C       READ THE DATA IN TO THE ARRAY
C
        X2 = X1
        DO 24 K = 1,NFLTR
        N1 = 1
        N3 = IBLOK/NAVE
        DO 34 I = 1,NSEG
        READ(7,REC=NUM)(A(ID,K),ID=N1,N3)
        TYPE*,'NUM :',NUM
        NUM = NUM + NFLTR
        TYPE*,'NFLTR :',K
        N1 = N3+1
        N3 = N3+(IBLOK/NAVE)
34      CONTINUE
        TYPE*,'REPOSITION REC',NUM
        NUM = NUM-(NFLTR*NSEG)
        NUM = NUM+1
24      CONTINUE
C
C       RUNNING SPECTRUM PLOTTING ROUTINE
C
```

```
      TYPE*,'SET THE PLOTTER PLEASE :'
      ACCEPT*,NPR
      CALL PLOTS
      CALL FACTOR(1./DMULT)
      Y1 = 0.
      DO 200 I = 1,NBLOK
      X1 = X1+DX
      DO 200 J = 1,NFLTR
      A(I,J)=A(I,J)*SFACT+X1
200   CONTINUE
      NT = NFLTR*NP
      DO 300 I = 1,NT
      PMAX(I) = 0.
300   CONTINUE
      IYS = 0
      DO 400 I = 1,NBLOK
      DO 501 J = 1,NFLTR-1
      DO 501 K = 0,NP
      IPT = (NP*(J-1))+K+1+IYS
      P(IPT) = (A(I,(J+1))-A(I,J))/FLOAT(NP)*FLOAT(K)+A(I,J)
501   CONTINUE
      DO 550 J = 1,IPT
      IF (P(J).LE.PMAX(J))GOTO 550
      PMAX(J) = P(J)
550   CONTINUE
      DO 600 J = IYS,IPT
      Y = YMAX-YDIS*FLOAT(J)/FLOAT((NFLTR-1)*NP)
      X = PMAX(J)
      IF (J.GT.IYS)GOTO 601
      CALL XYPLOT(X,Y,3)
601   CALL XYPLOT(X,Y,2)
600   CONTINUE
      IYS = IYS+IY
      TYPE*,'IYS :',IYS
400   CONTINUE
      YDIT = Y
      YDIF = YMAX-YDIS
      CALL XYPLOT(X2,YDIF,2)
      CALL XYPLOT(X2,YMAX,2)
C
C     FIXING UP THE AXIS & LABELLING THE PLOTS
C     Y AXIS
C
      DSX = X2
      DSY = YMAX
      YSHFTL = YDIS/3.
      YSTEP = YDIS/FLOAT(NYTIC)
      DXN = DSX-XSHFTN
      DXL = DSX-XSHFTL
      DYL = DSY-YSHFTL
```

```
      DO 62 KK = 1,NYTIC+1
      CALL SYMBOL(DSX,DSY,DNH,ISY,YANG,-1)
      DSY1 = DSY+0.4
      CALL NUMBER(DXN,DSY1,DH,YANG,3.0,'(I3)',IN)
      IN = IN+4
      DSY = DSY-YSTEP
62    CONTINUE
      CALL SYMBOL(DXL,DYL,YH,ISMY,YANG,NCHAR)
!
!     X AXIS LABELLING
!
      DSX = X2
      DSY = YDIF
      YSHFTN = XSHFTN
      XDIS = (X1-X2)
      DYN = DSY-YSHFTN
      DXL = DSX+(XDIS/3.)
      DYL = DSY-YSFTLX
      YSTEP = (YDIF-YDIT)/FLOAT(NXTIC)
      XSTEP = XDIS/FLOAT(NXTIC)
      DO 64 JJ = 1,NXTIC+1
      CALL SYMBOL(DSX,DSY,DNH,ISX,XANG,-1)
      DSX1=DSX-0.5
      DYN1=DYN-0.1
      CALL NUMBER(DSX1,DYN1,DH,XANG,3.0,'(I3)',NUMX)
      DSX = DSX+XSTEP
      DSY = DSY-YSTEP
      DYN = DYN-YSTEP
      NUMX = NUMX+1
64    CONTINUE
      CALL SYMBOL(DXL,DYL,YH,ISMX,XANG,NCHAR)
      CALL DUMP
!
!     MARKING THE BOUNDARIES FOR THE PLOT
!
      TYPE*,'PLEASE CHANGE THE PEN FOR MARKING BLOCKS:'
      ACCEPT*,NMAR
      DSX = X2
      DSY = YDIF
      XDIS = X1-X2
      YSTEP = (YDIF-YDIT)/FLOAT(NSEG)
      XSTEP = XDIS/FLOAT(NSEG)
      YBSTEP = YDIS/FLOAT(NBTIC)
      DYBN = DSY-YSHFTB
      DXBL = XDIS/3.
      DYBL = DSY-YSFTBL
      DSX = DSX+XSTEP
      DSY = DSY-YSTEP
      DO 66 LL = 1,NSEG
```

```
      DO 66 KK = 1,NBTIC+1
      CALL SYMBOL(DSX,DSY,DBH,ISX,XANG,-1)
      DSY = DSY+YBSTEP
68    CONTINUE
      DSY = DSY-(YBSTEP*(NBTIC+1))
      DYBN1 = DYBN-0.0
      DSBX1 = DSX-(XSTEP/2.0)
      CALL NUMBER(DSBX1,DYBN1,DBNH,BANG,3.0,'(I3)',NUMB)
      NUMB = NUMB+1
      DSY = DSY-YSTEP
      DSX = DSX+XSTEP
      DYBN = DYBN-YSTEP
66    CONTINUE
      CALL SYMBOL(DXBL,DYBL,YH,ISMB,XANG,NCHAR)
      CALL DUMP
!
!     LABEL THE DIAGRAM
!
      TYPE*,'MOVE THE PEN TO THE POSITION & ENTER LABEL:'
      READ(5,1)ALPH
1     FORMAT(80A1)
      K = 80
3     IF (ALPH(K).NE.' ')GO TO 2
      K = K-1
      GOTO 3
2     CALL ASCALE(XH,YH)
      CALL ROTAT(IANG)
      CALL CHOUT(ALPH,K)
      CALL PLEXIT
      CLOSE(UNIT=7)
      STOP
      END
```

```
      DBNH = 0.5
      NBTIC = 25
      YMAX = 25.0
      YDIS = 8.0
      NXTIC = 16
      NYTIC = 4
      X1 = 1.0
      VAL = 0.0
      OFSET = 0.0
!
!     INPUT THE PARAMETERS
!
      TYPE*,'DATA IS RETRIEVED FROM :DWTHE11.DAT'
      TYPE*,'INPUT THE NO OF CHANNELS TO BE SET TO ZERO:'
      ACCEPT*,NZERO
      TYPE*,'& THE SCALE FACTOR :'
      ACCEPT*,SFACT
      TYPE*,'NBLOK ;',NBLOK
      TYPE*,'ENTER TITLE OF X AXIS:'
      ACCEPT132,(ISMX(K),K=1,20)
132   FORMAT(20A1)
      TYPE*,'ENTER TITLE OF Y AXIS:'
      ACCEPT133,(ISMY(K),K=1,20)
133   FORMAT(20A1)
      TYPE*,'ENTER TITLE OF THE BLOCK:'
      ACCEPT134,(ISMB(K),K=1,20)
134   FORMAT(20A1)
      TYPE*,'INPUT THE STARTING NO FOR X AXIS:'
      ACCEPT*,NUMX
      TYPE*,'& FOR THE SPEECH DATA BLOCK:'
      ACCEPT*,NUMB
!
!
      OPEN(UNIT=7,NAME='DWTHE11',ACCESS='DIRECT',STATUS='OLD',
     1      RECORDTYPE='FIXED',FORM='UNFORMATTED',RECL=64)
!
!     READ THE DATA IN TO THE ARRAY
!
      X2 = X1
      DO 24 K = 1,NFLTR
      N1 = 1
      N3 = IBLOK/NAVE
      DO 34 I = 1,NSEG
      READ(7,REC=NUM)(A(ID,K),ID=N1,N3)
      TYPE*,'NUM :',NUM
      NUM = NUM + NFLTR
      TYPE*,'NFLTR :',K
      N1 = N3+1
      N3 = N3+(IBLOK/NAVE)
34    CONTINUE
      TYPE*,'REPOSITION REC',NUM
      NUM = NUM-(NFLTR*NSEG)
      NUM = NUM+1
```

```
24      CONTINUE
!
!       FIND THE LOG MAGNITUDE OF THE VALUES IN THE ARRAY
!
        DO 50 K = 1,NFLTR
        DO 50 ID = 1,NBLOK
        BDATA = A(ID,K)
        FDATA = 0.43429448*ALOG(BDATA)
        A(ID,K) = FDATA
50      CONTINUE
!
!       DISCRETE COSINE TRANSFORM
!       TO FIND THE VALUE OF THE DCT COEF OF THE ARRAY
!
        DO 220 ID = 1,NBLOK
        SUM1 = 0.
        DO 240 K = 1,NFLTR
        SUM1 = A(ID,K)+SUM1
240     CONTINUE
        K = 1
        B(K) = ((1.4142136/FLOAT(NFLTR))*SUM1)
        DO 260 K = 2,NFLTR
        SUM2 = 0.
        DO 280 M = 1,NFLTR
        THETA=((2.*FLOAT(M-1)+1.)*FLOAT(K-1)*PI)/(2.*FLOAT(NFLTR))
        XA = COS(THETA)
        XK = A(ID,M)*XA
        SUM2 = XK+SUM2
280     CONTINUE
        B(K) = ((2./FLOAT(NFLTR))*SUM2)
260     CONTINUE
        DO 310 K = 1,NFLTR
        A(ID,K) = B(K)
310     CONTINUE
220     CONTINUE
!
!       TO SET THE COMPONENTS TO ZERO FOR DATA REDUCTION
!
        NSET = NFLTR-NZERO+1
        DO 30 ID = 1,NBLOK
        DO 40 K = NSET,NFLTR
        A(ID,K) = 0.0
40      CONTINUE
30      CONTINUE
!
!       INVERSE DISCRETE COSINE TRANSFORM
!
        DO 320 ID = 1,NBLOK
        DO 340 M = 1,NFLTR
        K = 1
        GX0 = (1./1.4142136)*A(ID,K)
        SUM3 = 0.
        DO 360 K = 2,NFLTR
```

```
      THETA1=((2.*FLOAT(M-1)+1.)*FLOAT(K-1)*PI)/(2.*FLOAT(NFLTR))
      GX = A(ID,K)*COS(THETA1)
      SUM3 = GX+SUM3
360   CONTINUE
      B(M) = GX0+SUM3
      J = J+1
340   CONTINUE
      DO 305 K = 1,NFLTR
      A(ID,K) = B(K)
305   CONTINUE
320   CONTINUE
!
!     FIND THE ARRAY FOR PLOTTING
!     Conversion to linear scale by taking the Anti Log
!
!
      DO 57 K = 1,NFLTR
      DO 57 ID = 1,NBLOK
      EDATA = A(ID,K)
      GDATA = EXP(2.3025851*EDATA)
      A(ID,K) = GDATA
57    CONTINUE
!
!     RUNNING SPECTRUM PLOTTING ROUTINE
!
!
      TYPE*,'SET THE PLOTTER PLEASE :'
      ACCEPT*,NPR
      CALL PLOTS
      CALL FACTOR(1./DMULT)
      Y1 = 0.
      DO 200 I = 1,NBLOK
      X1 = X1+DX
      DO 200 J = 1,NFLTR
      A(I,J)=A(I,J)*SFACT+X1
200   CONTINUE
      NT = NFLTR*NP
      DO 300 I = 1,NT
      PMAX(I) = 0.
300   CONTINUE
      IYS = 0
      DO 400 I = 1,NBLOK
      DO 501 J = 1,NFLTR-1
      DO 501 K = 0,NP
      IPT = (NP*(J-1))+K+1+IYS
      P(IPT) = (A(I,(J+1))-A(I,J))/FLOAT(NP)*FLOAT(K)+A(I,J)
501   CONTINUE
      DO 550 J = 1,IPT
      IF (P(J).LE.PMAX(J))GOTO 550
      PMAX(J) = P(J)
550   CONTINUE
      DO 600 J = IYS,IPT
      Y = YMAX-YDIS*FLOAT(J)/FLOAT((NFLTR-1)*NP)
```

```
      X = PMAX(J)
      IF (J.GT.IYS)GOTO 601
      CALL XYPLOT(X,Y,3)
601   CALL XYPLOT(X,Y,2)
600   CONTINUE
      IYS = IYS+IY
      TYPE*,'IYS :',IYS
400   CONTINUE
      YDIT = Y
      YDIF = YMAX-YDIS
      CALL XYPLOT(X2,YDIF,2)
      CALL XYPLOT(X2,YMAX,2)
!
!     FIXING UP THE AXIS & LABELLING THE PLOTS
!     Y AXIS
!
      DSX = X2
      DSY = YMAX
      YSHFTL = YDIS/3.
      YSTEP = YDIS/FLOAT(NYTIC)
      IN = 0
      DXN = DSX-XSHFTN
      DXL = DSX-XSHFTL
      DYL = DSY-YSHFTL
      DO 62 KK = 1,NYTIC+1
      CALL SYMBOL(DSX,DSY,DNH,ISY,YANG,-1)
      DSY1 = DSY+0.4
      CALL NUMBER(DXN,DSY1,DH,YANG,3.0,'(I3)',IN)
      IN = IN+4
      DSY = DSY-YSTEP
62    CONTINUE
      CALL SYMBOL(DXL,DYL,YH,ISMY,YANG,NCHAR)
!
!     X AXIS LABELLING
!
      DSX = X2
      DSY = YDIF
      YSHFTN = XSHFTN
      XDIS = (X1-X2)
      DYN = DSY-YSHFTN
      DXL = DSX+(XDIS/3.)
      DYL = DSY-YSFTLX
      YSTEP = (YDIF-YDIT)/FLOAT(NXTIC)
      XSTEP = XDIS/FLOAT(NXTIC)
      DO 64 JJ = 1,NXTIC+1
      CALL SYMBOL(DSX,DSY,DNH,ISX,XANG,-1)
      DSX1=DSX-0.5
      DYN1=DYN-0.1
      CALL NUMBER(DSX1,DYN1,DH,XANG,3.0,'(I3)',NUMX)
      DSX = DSX+XSTEP
      DSY = DSY-YSTEP
      DYN = DYN-YSTEP
      NUMX = NUMX+1
```

```
64      CONTINUE
        CALL SYMBOL(DXL,DYL,YH,ISMX,XANG,NCHAR)
        CALL DUMP
!
!       MARKING THE BOUNDARIES FOR THE PLOT
!
        TYPE*,'PLEASE CHANGE THE PEN FOR MARKING BLOCKS:'
        ACCEPT*,NMAR
        DSX = X2
        DSY = YDIF
        XDIS = X1-X2
        YSTEP = (YDIF-YDIT)/FLOAT(NSEG)
        XSTEP = XDIS/FLOAT(NSEG)
        YBSTEP = YDIS/FLOAT(NBTIC)
        DYBN = DSY-YSHFTB
        DXBL = XDIS/3.
        DYBL = DSY-YSFTBL
        DSX = DSX+XSTEP
        DSY = DSY-YSTEP
        DO 66 LL = 1,NSEG
        DO 68 KK = 1,NBTIC+1
        CALL SYMBOL(DSX,DSY,DBH,ISX,XANG,-1)
        DSY = DSY+YBSTEP
68      CONTINUE
        DSY = DSY-(YBSTEP*(NBTIC+1))
        DYBN1 = DYBN-0.0
        DSBX1 = DSX-(XSTEP/2.0)
        CALL NUMBER(DSBX1,DYBN1,DBNH,BANG,3.0,'(I3)',NUMB)
        NUMB = NUMB+1
        DSY = DSY-YSTEP
        DSX = DSX+XSTEP
        DYBN = DYBN-YSTEP
66      CONTINUE
        CALL SYMBOL(DXBL,DYBL,YH,ISMB,XANG,NCHAR)
        CALL DUMP
!
!       LABEL THE DIAGRAM
!
        TYPE*,'MOVE THE PEN TO THE POSITION & ENTER LABEL:'
        READ(5,1)ALPH
1       FORMAT(80A1)
        K = 80
3       IF (ALPH(K).NE.' ')GO TO 2
        K = K-1
        GOTO 3
2       CALL ASCALE(XH,YH)
        CALL ROTAT(IANG)
        CALL CHOUT(ALPH,K)
        CALL PLEXIT
        CLOSE(UNIT=7)
        STOP
        END
```

```
      PROGRAM ORTH
C
C     ROUTINE TO TEST THE ORTHOGONAL PROPERITY OF THE
C     DISCRETE COSINE TRANSFORM
C
C
C     CONSTANT VALUE DECLARATION
C
      DIMENSION A(16,16),B(16,16),C(16,16)
      PI = 3.1415927
      NCOMP = 16
      NPRINT = 16
C
C     CALCULATION OF THE COEFFICIENT MATRIX
C
      K = 1
      DO 10 ID = 1, NCOMP
      A(ID,K) = 1./(SQRT(FLAOT(NCOMP)))
10    CONTINUE
      DO 20 M = 1,NCOMP
      DO 30 K = 2,NCOMP
      THETA=((2.*FLOAT(M-1)+1.)*FLOAT(K-1)*PI)/(2.*FLOAT(NCOMP))
      XA = COS(THETA)
      A(M,K) = SQRT(2./FLOAT(NCOMP))*XA
30    CONTINUE
20    CONTINUE
      WRITE(6,12)
12    FORMAT(/,10X,'ID',5X,'K','A(ID,K)',/)
      DO 40 ID = 1,NPRINT
      DO 40 K = 1,NCOMP
      WRITE(6,14)ID,K,A(ID,K)
14    FORMAT(10X,I8,5X,F12.8)
40    CONTINUE
C
C     TRANSPOSE THE COEFFICIENT MATRIX
C
      DO 50 ID = 1,NCOMP
      DO 50 K = 1,NCOMP
      B(K,ID) = A(ID,K)
50    CONTINUE
      WRITE(6,18)
18    FORMAT(/,5X,'TRANSPOSE OF THE COEF. MATRIX:ID,K,A(ID,K)',/)
      DO 52 ID = 1,NPRINT
```

```
      DO 52 K = 1,NCOMP
      WRITE(6,16)ID,K,B(ID,K)
16    FORMAT(10X,I8,5X,I8,5X,F12.8)
52    CONTINUE
C
C     TO FIND THE PRODUCT OF THE MATRIX
C
      DO 60 I = 1,NCOMP
      DO 60 J = 1,NCOMP
      C(I,J) = 0.0
      DO 60 K = 1,NCOMP
      C(I,J) = C(I,J)+B(I,K)*A(K,J)
60    CONTINUE
      WRITE(6,24)
24    FORMAT(/,5X,'THE PRODUCT OF THE MATRIX:',/)
      DO 62 ID = 1,NPRINT
      DO 62 K = 1,NCOMP
      WRITE(6,22)ID,K,C(ID,K)
22    FORMAT(10X,I8,5X,I8,5X,I8,5X,F12.8)
62    CONTINUE
      END
```