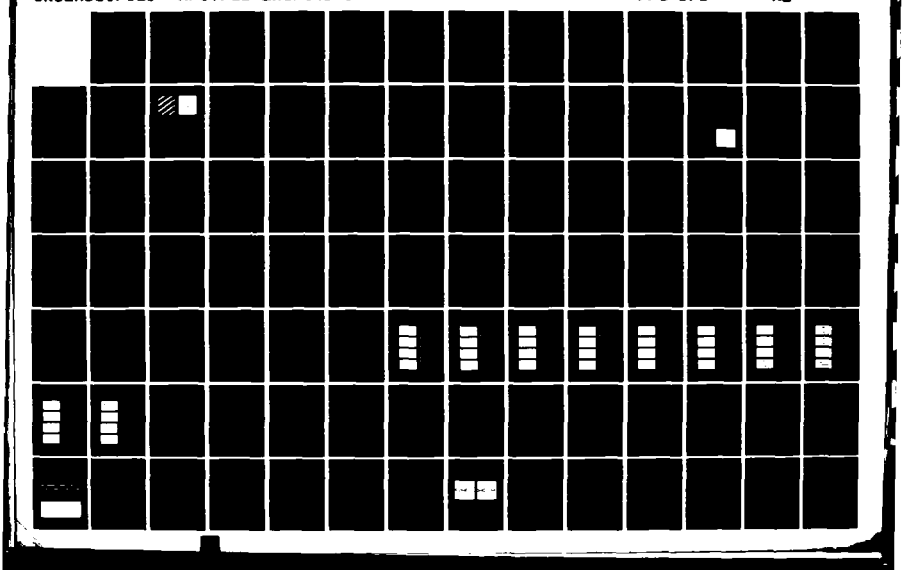
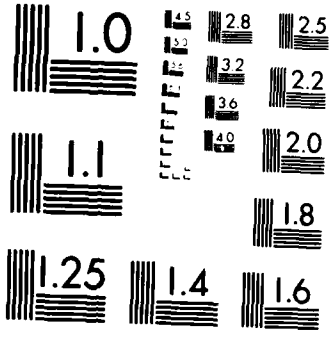
AD-A151 898 PROCESSING SPEECH FOR ANALYSIS USING OPTICAL FOURIER TECHNIQUES(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING D L JONES DEC 84 1/2

UNCLASSIFIED AFIT/GE/ENG/84D-37 F/G 9/3 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DTIC

AD-A151 898

PROCESSING SPEECH FOR ANALYSIS USING OPTICAL FOURIER TECHNIQUES

THESIS

Duane L. Jones
First Lieutenant, USAF

AFIT/GE/ENG/84D-37

DTIC FILE COPY

This document has been approved for public release and sale; its distribution is unlimited.

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

85 03 13 144

REPRODUCED AT GOVERNMENT EXPENSE

REPRODUCED AT GOVERNMENT EXPENSE

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AFIT/GE/ENG/84D-37

Copy available to DTIC does not permit fully legible reproduction

PROCESSING SPEECH FOR ANALYSIS USING OPTICAL FOURIER TECHNIQUES

THESIS

Duane L. Jones
First Lieutenant, USAF

AFIT/GE/ENG/84D-37

Approved for public release; distribution unlimited

E
2 1985
D
A

AFIT/GE/ENG/84D-37

PROCESSING SPEECH FOR ANALYSIS USING OPTICAL FOURIER METHODS

THESIS

Presented to the Faculty of the School of Engineering

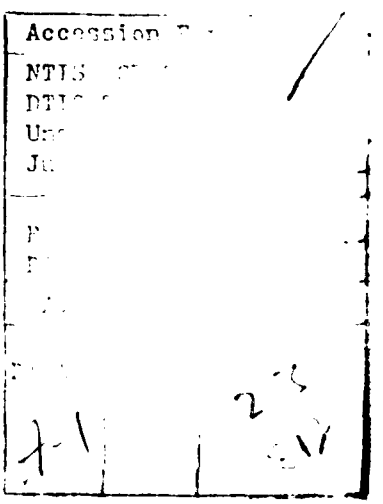
of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science in Electrical Engineering



Duane L. Jones, B.S.

First Lieutenant, USAF

December 1984

Approved for public release; distribution unlimited

Preface

This work looks at speech recognition in a new domain, making use of some recent advances in optical devices. It is not the first time that speech has been made into pictures, but is yet another attempt to get more information faster out of the frequency domain, wherein it seems the answer to speech recognition lies. This project is sponsored by the Air Force Wright Aeronautical Laboratory, and in particular Lt Col Bruce R. Altschuler, who first proposed that optical techniques be used in speech analysis.

I wish to express my appreciation to my advisor, Dr. Matthew Kabrisky, who brought this project to life and has supported me throughout the effort. I also wish to thank Mr. Dan Zambon of the AFIT Laboratory and Systems Support office for his help with the computer equipment, and Mr. Douglass J. Sauer of the Aero-Medical Research Lab for his help in editing and assembling the video data. Finally, I wish to thank my my wife Colleen for her understanding and support, and the sacrifices she has made to make my work easier.

Duane L. Jones

Table of Contents

	Page
Preface	ii
List of Figures	v
List of Tables	viii
Abstract	ix
I. Introduction	1
Background	1
Problem	4
Scope	4
Approach	6
Phoneme Nomenclature	6
II. Theory	8
Conversion of Audio to Video	8
Processing Techniques	10
Display of Baseband Audio	12
Creating Two Dimensional Images Using Modulation	13
Frequency Selection	22
Prediction of Vowel Sound Transform Images	26
Fricatives and Stop Sounds	29
III. Equipment	30
System Overview	30
Preprocessing Circuitry	32
Modulation Circuitry	35
Video Mixing Circuitry	35
Construction Techniques	39
Computer Equipmennt	40
IV. Procedures	41
Initial Recording	41
Video Processing	41
Digitizing Video Signals	43
Numerical Processing	44
Display of Transform Data	46

Page

V. Results 48

Static Data 48
Dynamic Data 70

VI. Conclusions and Recommendations 73

A New Tool for Speech Analysis 73
Recomendations for Further Development 74

Appendix A: List of Equipment 76

Appendix B: Equipment Set-up and Adjustment 77

Appendix C: Listings of Computer Programs 80

QFT 81
PLOT 84
IOF 86
UNPACK 87
REPACK 87

Bibliography 88

Vita 89

List of Figures

Figure		Page
1.	Sine Wave Grating Pattern	3
2.	Two Dimensional Fourier Transform Image	3
3.	Power Distribution Spectrum in Speech Showing Roll-off Above 500 hz	8
4.	Typical Video Signal	9
5.	Audio Signal With Sync Inserted	10
6.	Combined Audio-Video Signal	11
7.	Baseband Video Image for the Sound AH	12
8.	Two Dimensional Fourier Transform Image	12
9.	Formation of Video Pattern Angles for n=2	14
10.	Absolute Value of Pattern Angles for Frequencies up to 100 Khz	17
11.	Spatial Frequency Dependence on Pattern Angle	18
12.	Corresponding Transform Image Pattern	18
13.	Balanced Modulator Output Spectrum	21
14.	Values of Formants for Several Vowel Sounds	22
15.	First Formant Pattern Angles	24
16.	Second Formant Pattern Angles	24
17.	Spatial Filter Pass Bands	26
18.	Theoretical Transform Image Peak Locations for Vowel Sounds	28
19.	System Block Diagram	31
20.	Schematic Diagram for Preprocessing Circuitry	33
21.	Combined Pre-emphasis / Low-pass Filter Response	34

Figure		Page
22.	Schematic Diagram for Modulation Circuitry	36
23.	Schematic Diagram for Video Mixing Circuitry . . .	37
24.	Timing Diagram for Video Mixing Circuitry	38
25.	Data Processing Flow Diagram	42
26.	Spatial Filter Pass Band Points	45
27.	Transform Images for the Sound OO	49
28.	Transform Images for the Sound U	50
29.	Transform Images for the Sound OH	51
30.	Transform Images for the Sound UH	52
31.	Transform Images for the Sound AH	53
32.	Transform Images for the Sound A	54
33.	Transform Images for the Sound E	55
34.	Transform Images for the Sound EH	56
35.	Transform Images for the Sound I	57
36.	Transform Images for the Sound EE	58
37.	Baseband Pixel Intensity Plots for SH	60
38.	Baseband Pixel Intensity Plots for SS	61
39.	Baseband Pixel Intensity Plots for TH	62
40.	Baseband Pixel Intensity Plots for FF	63
41.	Baseband Pixel Intensity Plots for KK	64
42.	Baseband Pixel Intensity Plots for JJ	65
43.	Baseband Pixel Intensity Plots for ZZ	66
44.	Baseband Pixel Intensity Plots for TT	67
45.	Baseband Pixel Intensity Plots for VV	68
46.	Baseband Pixel Intensity Plots for GG	69

Figure Page

47. Time Domain Image of "EE" in "zero" 71

48. Transform Image of "EE" in "zero" 71

49. Modulator Output with Improper Balanced 78

50. Modulator Output with Proper Balanced 78

List of Tables

Table		Page
I.	Phoneme Codes and Symbols	7
II.	Video Scan Rate Multiples	23
III.	Carrier Frequency Selection	25
IV.	Theoretical Values for Pattern Angles of Vowel Sounds Based on Values from Figure 14	27

Abstract

A system for displaying speech as a two dimensional video image is presented. The speech is pre-processed by compressing its dynamic range and filtering to emphasize frequencies above 500 hz. Blanking and sync pulses are inserted to put the signal in standard video format, and every other field is blanked to prevent interference between fields in the interlaced display.

Two dimensional variation is achieved by modulating the baseband audio signal up in the spectrum near a multiple of the video scan rate. The relationship between input frequency and pattern angle of the display is derived, and it is shown that the set of frequencies near a multiple of the video scan rate have points in the spatial frequency domain which lie in a straight line at a distance from the origin proportional to the scan rate multiple.

Two modulation frequencies are selected to display in the spatial frequency domain the location of the first and second formant peaks. The two modulated signals are mixed with the baseband audio and displayed simultaneously in a single image. The images are digitized and an optical Fourier transform is simulated on the computer by creating the image which would appear in the Fourier transform plane. Entire words are processed by assembling individual frames on video tape.

The system shows the capability of processing multiple high resolution bands of frequency information for a given signal, and demonstrates the feasibility of using optical processes in the analysis of speech signals.

PROCESSING SPEECH FOR ANALYSIS USING OPTICAL FOURIER METHODS

I. Introduction

Background

The first task in the process of phoneme based speech recognition is analyzing the speech signal to identify the individual acoustic events. Once identified, these pieces are mapped into a string of words which will make up a meaningful message. Except for linear predictive coding (LPC), which models the vocal tract, most contemporary schemes for phoneme identification use frequency domain information to descriminate one sound from another, so a form of information extraction from this domain is required. This is usually achieved at some cost in information, processing time, or system complexity.

The most popular methods of examining the frequency content of a signal generally fall into two categories, each of which has its limits as to how fast information can be obtained. Those systems which pass the signal through a number of filters to detect the power in a given frequency band are limited by system complexity: the more filters built, the more information obtained, and by resolution: small bandwidths at high frequencies are difficult to achieve. On the other hand, systems

which digitize the signal, then numerically transform it to the frequency domain, cost in processing time. Naturally, the smaller the increments used and the greater the number, the more time it takes to carry out the computations. Information is also lost due to errors in digitizing and manipulating the data. The effect is the same in either case: a limited rate of obtaining information from the frequency domain. There is inevitably some trade off between the amount of information needed and system size and speed. Optical processes may provide an answer to this problem.

Advances in optics have produced devices such as the liquid crystal light valve which enable real time Fourier processing of two dimensional images, giving a representation of the image in the spatial frequency domain. Such devices do not require the image to be produced by a coherent light source, which means the picture from a video display could be used.

A sine wave can be represented in two dimensions as a grating pattern made up of sinusoidally varying bars with their spacing and width dependent on the frequency of the wave. If this pattern extended infinitely in all directions, its optical Fourier transform would be a single pair of points corresponding to the negative and positive spatial frequencies of the sine wave pattern. Their actual position in the plane would depend on the the spatial frequency and orientation of the bars in the pattern. Such a sine wave grating pattern and its resulting two dimensional Fourier transform image appear in Figures 1 and 2.

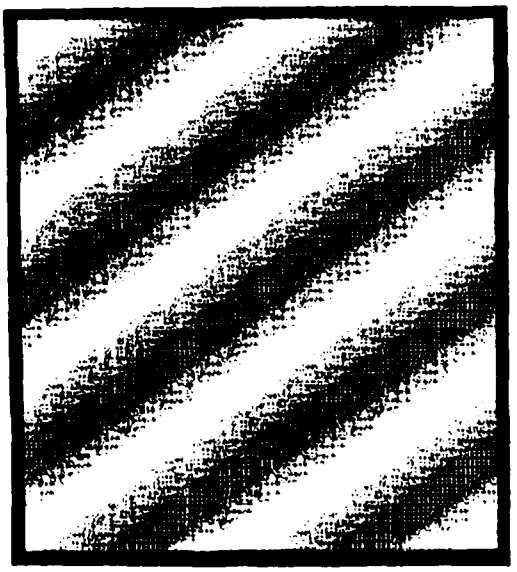


Figure 1. Sine Wave Grating Pattern

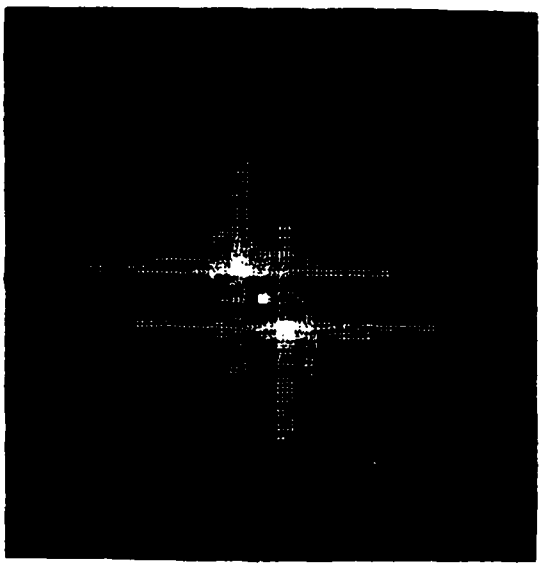


Figure 2. Two Dimensional Fourier Transform Image

This optical Fourier transform offers some very desirable characteristics. It is "instantaneous"; operating at the speed of light, it is continuous; sacrificing very little information in the process, and its resolution is limited only by the resolution capabilities of the optical devices. These result in an enormous rate of information transfer from one domain to the other.

Because of the unique qualities of the optical Fourier transform, it has been proposed that optical techniques be used in the analysis of speech signals. An image representing speech could be processed using optical techniques to obtain a second image representing the speech image in the spatial frequency domain. This new picture could then be

digitized for numerical analysis, analyzed using templates, optical detector arrays, or possibly examined and read "as is" by the human eye. The prospects of such valuable results certainly invite the exploration of using optical techniques in speech analysis and recognition systems.

Problem

Using optical Fourier techniques on a time domain speech signal requires that the signal be presented as a two dimensional picture. The appearance of the picture is unimportant as long as its optical transform image displays information about the spectral content of the speech in some readily useable form. At present, such pictures have yet to be created and the nature of their resulting Fourier transform images is unknown.

The purpose of this project is to explore the feasibility of using optical Fourier techniques in the analysis of speech signals by building a system which will display speech as a two dimensional picture, and observing the images which result when a two dimensional Fourier transform is performed.

Scope

There are numerous ways in which speech might be displayed visually. This work will deal only with raster-scan devices such as a common video monitor. The project will consist of converting the electrical speech signal from a microphone into a standard composite video signal suitable for input to a monitor, digitizer, or other video equipment.

A speech signal put directly into a video display is inherently a one dimensional phenomenon, since most of its energy lies in frequencies well below that of the scan rate of the TV. A method of making the display two dimensional must be found. The use of modulation will be examined as a possible solution to this problem.

The optical transform itself will not be treated, although its effects will be considered in the design of the system and it will be used to test the end product. Transform images will be created for various sounds of speech and compared, in the case of the vowel sounds, with theoretical results, or in the case of the fricative sounds, with measurements made using a spectrum analyzer.

Some sounds in speech are not static events at all, but a transition from one sound to another, such as the glides "Y" and "W". Other sounds require periods of silence (stop sounds). These cannot be described by a single image, but require a sequence of images to describe them. Therefore this study will also cover the processing of entire words in order to observe the dynamic characteristics of the transform images.

It is not the purpose of this work to make an exhaustive study of the transform images for every speech sound spoken by large numbers of people, but rather to take a first look at speech in a domain where it has not been examined before. Knowledge gained from this study should lead to optimization of the system to give the most useful output to a subsequent phoneme discriminator.

Approach

The speech signal will be conditioned using automatic gain control and filtered to emphasize the higher frequencies. The appropriate blanking and sync pulses must also be added to give the signal a standard video form. The audio signal may be left in the baseband or modulated up to a higher frequency in an attempt to give the picture two dimensional qualities. One or more of these bands may be combined to give a final picture.

The output picture will then be processed using optical Fourier techniques. If the optical equipment is unavailable, this will be simulated using digital signal processing techniques. This will be done for several vowel sounds and fricatives, and then for some complete words, such as the digits "zero" through "nine".

Phoneme Nomenclature

Throughout this report, different sounds will be identified using a one or two letter code which facilitates use in data processing routines on the computer. Since they may not match the traditional symbols associated with the phonemes, a listing of the codes, the sounds they represent, and their corresponding symbols are contained in Table I.

Table I

Phoneme Codes and Symbols

Vowels

Code	Symbol	Sound
OO	u	b<u>oo</u>t
U	ʊ	f<u>oo</u>t
OH	O	<u>o</u>bey
UH	ʌ	<u>u</u>p
AH	a	s<u>a</u>w
A	æ	b<u>a</u>t
E	ɛ	m<u>e</u>t
EH	e	h<u>a</u>te
I	I	<u>i</u>t
EE	i	<u>e</u>ve

Fricatives and Stops

Code	Symbol	Sound
SH	ʃ	<u>sh</u>e
JJ	ʒ	a<u>z</u>ure
SS	s	<u>s</u>ee
ZZ	z	<u>z</u>oo
TH	θ	<u>th</u>in
TT	ð	<u>th</u>en
FF	f	<u>f</u>or
VV	v	<u>v</u>ote
KK	k	<u>k</u>ey
GG	g	<u>g</u>et

II. THEORY

Conversion of Audio to Video

Displaying speech on a raster-scan device requires some special processing techniques in order to turn audio signals into video signals. Most of the processing of the speech signal will to be to resolve the differences between the two types of signals.

Audio Signals. The typical speech signal is continuous with amplitude proportional to the dynamic pressure created by the speaker and detected by a microphone. It is assumed that the majority of important frequency content of the signal is below ten kilohertz, with most of the energy in the lowest three kilohertz. This is due to the fact that above 500 hertz there

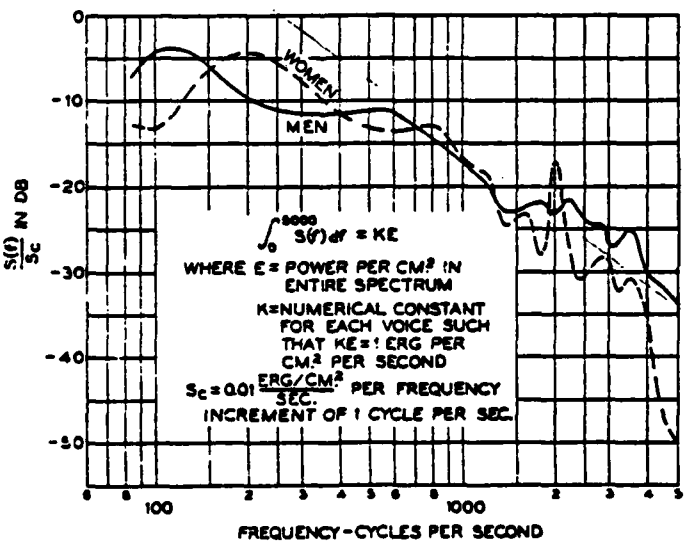


Figure 3. Power Distribution Spectrum in Speech Showing Roll-off Above 500 hz [8:651]

is a drop in the energy of speech of about eight decibels per octave, as shown in Figure 3 [8:651; 1:163]. The dynamic range of the amplitude of speech is very large, with voiced sounds having much greater energy than the fricatives. Transitions between events are relatively slow, limited by how fast the tongue and mouth can be moved. The maximum rate is thought to be about 20 hz.

Video Signals. Compared with speech signals, everything in a video signal is happening at a much faster rate. Typical video bandwidths run from three to five megahertz. The standard horizantal scan frequency is 15.75 kilohertz, or one line every 63.7 microseconds. Information is displayed discontinuously, stopping at the end of each scan period for a sync pulse and blanking during retrace. Two sets of 262.5 lines are displayed simultaneously at a field rate of 60 hertz, which means the viewer is actually seeing two events separated in time by 1/60 of a second.

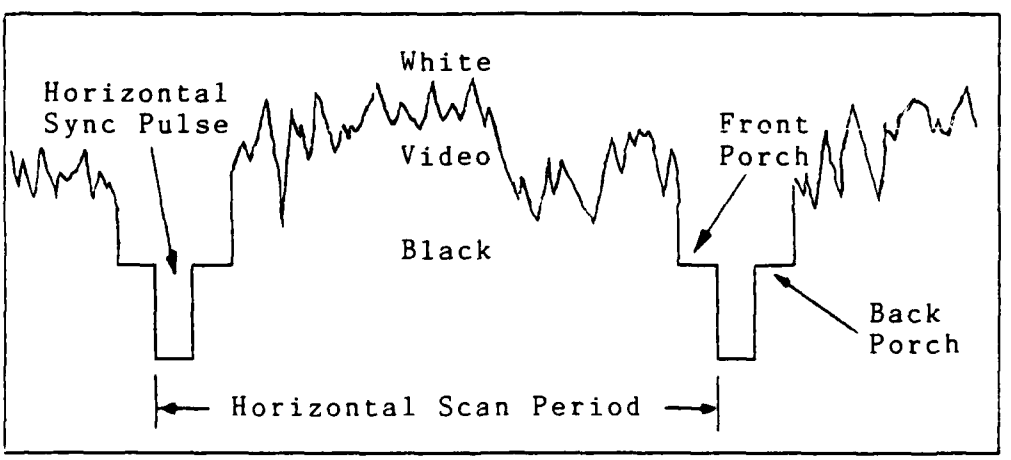


Figure 4. Typical Video Signal

The amplitude of a video signal corresponds to the brightness of the picture. Its dynamic range is comparatively small, with less than a volt between "black" and "white" levels.

Processing Techniques

The first step in turning audio into video is to compress the dynamic range of the speech amplitude into acceptable video levels. This can be done using a compressor or automatic gain circuit, which will also help eliminate differences due to variations in the volume of the speaker. In addition to compression, pre-processing of the speech signal should also include pre-emphasis filtering to compensate for the roll-off in energy of the higher frequencies [3].

The next step consists of interrupting the speech signal at appropriate intervals to insert blanking and sync pulses. The fact that blank spots have been placed in the signal creates a set of windows, each about 55 microseconds wide.

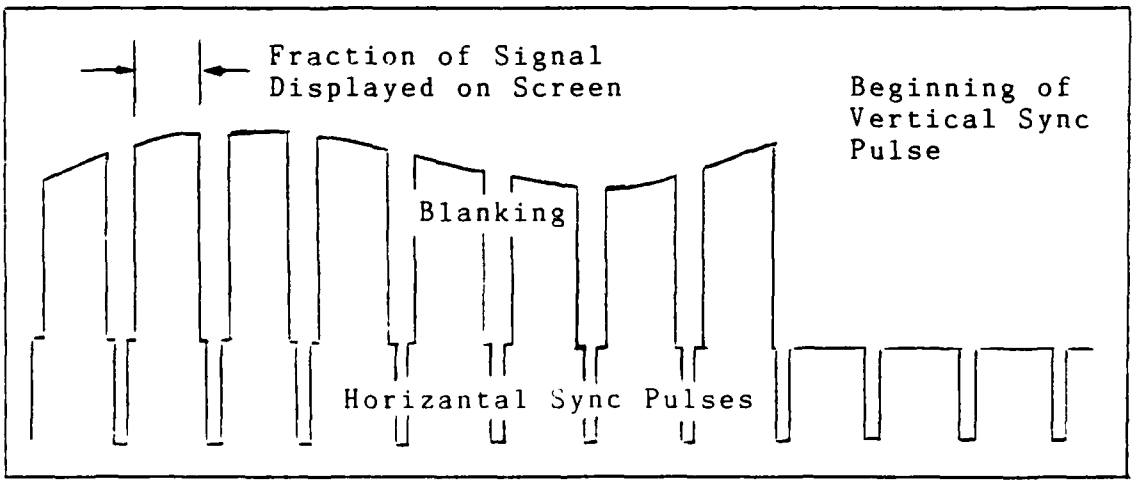


Figure 5. Audio Signal with Sync Inserted

The interlaced scan of a standard TV picture also presents a problem for displaying speech. As mentioned previously, a TV picture is actually two events taking place at different times. A picture displaying speech in such a manner would likewise display two segments of speech occurring at different times. Since the 1/60 of a second separation does not necessarily correspond to an exact multiple of the speech period, the resulting composite picture would have the effect of adding two signals of arbitrary phase. In order to prevent this, every other field should be blank. This further windows the speech, cutting the actual "sample" time to less than 50%. The resulting waveform as shown in Figure 6, would give a picture with every other line blank and the intensity or brightness of each pixel in the active part of the screen dependent on the audio signal level.

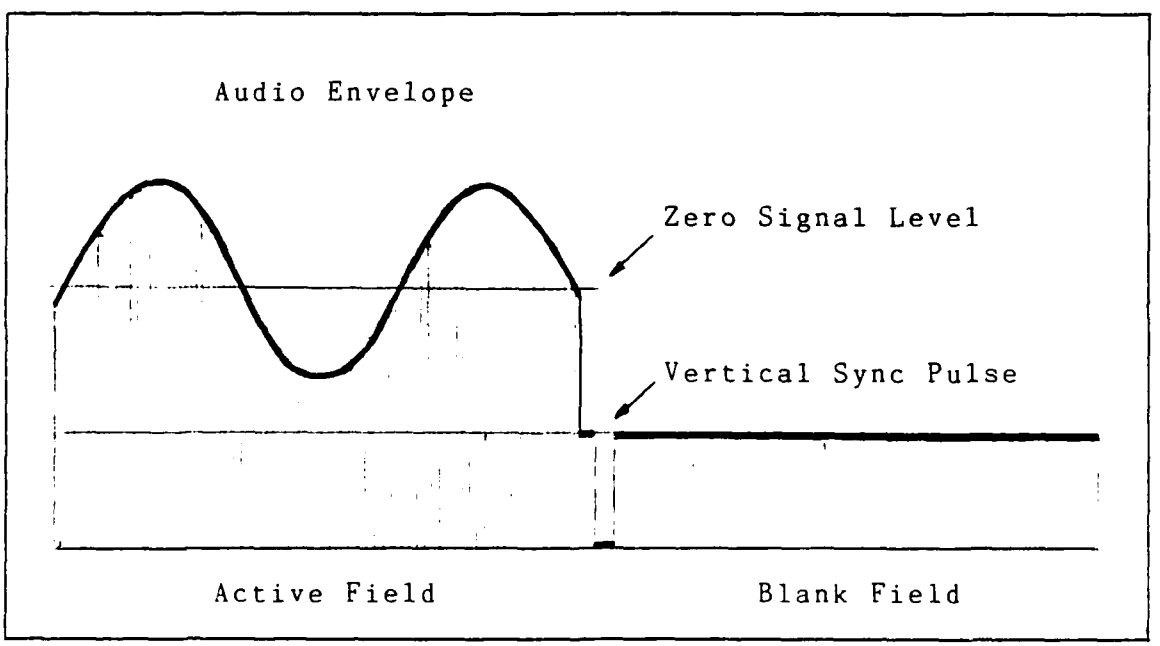


Figure 6. Combined Audio-Video Signal

Display of Baseband Audio

Processing an audio signal as described in the previous discussion results in a display composed of predominantly horizontal bars. This is because virtually all the energy of the speech signal lies in the frequencies well below the 15.75 Khz scan frequency of the video display. This creates essentially a one dimensional display, where time is the axis from top to bottom on the screen, and intensity is proportional to the amplitude of the speech signal.

For a typical male speaker with a glotal pitch around 130 hertz, a little more than two periods of speech are displayed on the screen at one time. A typical image is shown in Figure 7. The two dimensional fourier transform of this image yields a set of dots in a vertical line due to the 90 degree rotation of the transform. Shown in Figure 8, the transform is symmetrical, showing both positive and negative frequencies, with lower frequencies toward the center and higher frequencies toward the edges.

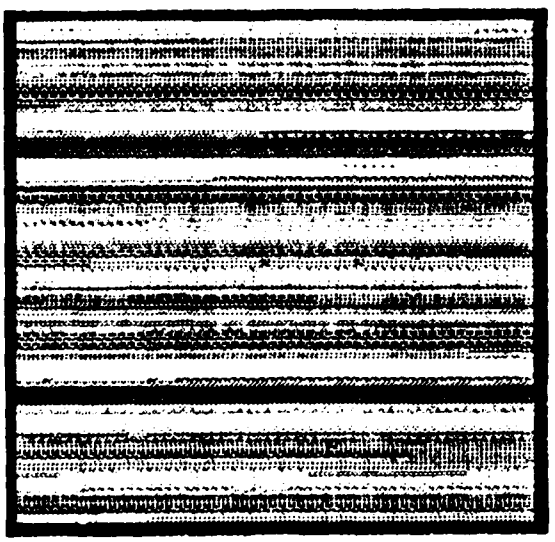


Figure 7. Baseband Video Image for the Sound "AH"

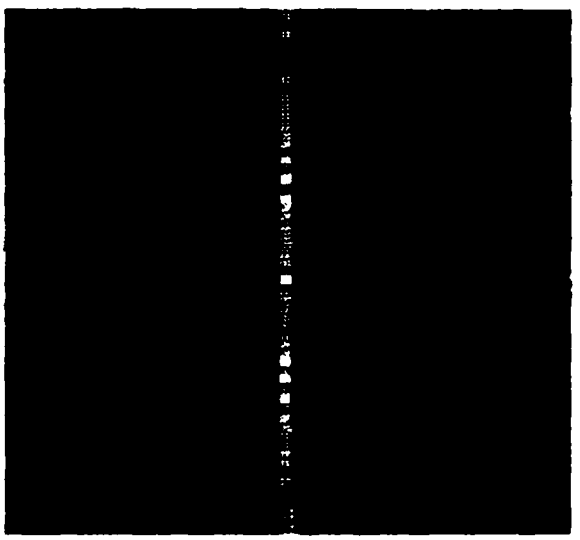


Figure 8. Two Dimensional Fourier Transform Image

The time domain pictures created from baseband audio vary greatly from speaker to speaker, such that given two sets of pictures corresponding to the vowel sounds spoken by two different speakers, an observer is unable to identify which pictures correspond to the same sounds. This indicats that the images must be made more repeatable for a given sound made by different speakers. Further, the display is essentially a one dimensional function. Some way of using the second axis to make the sounds more separable is desired.

Creating Two Dimensional Images Using Modulation

The only way to make vertical patterns on a TV display is to input frequencies which correspond roughly to a multiple of the horizontal scan frequency. This can be achieved by modulating the audio baseband up in the frequency spectrum. The resulting images should have vertical as well as horizontal components.

Pattern Angles. A typical video display is shown in Figure 9, with physical height H and width W of the screen indicated. Added to these dimensions are the horizontal and vertical retrace periods, showing what dimensions the screen would have if retrace were instantaneous. Z_h and Z_v are horizontal and vertical scan efficiencies described by

$$Z = \frac{\text{total scan period - retrace period}}{\text{total scan period}} \qquad (1)$$

Z_v is typically 90%, while Z_h is typically 80%. W/Z_h then represents the spatial period of the horizantal scan frequency F_s, and Z_v represents the spatial period of the vertical scan frequency.

The spatial period X on the screen for a given input frequency F can then be given by

$$X = \frac{F_s \; W}{F \quad Z_h} \tag{2}$$

If the frequency is an exact multiple n of F_h, then the period is W/Z_h n , and exactly n cycles are displayed during each scan period, although a portion of the pattern is blanked during retrace. The result is a pattern where the beginning of each period is directly beneath the beginning of a period on the previous line. The angle of the pattern is therefore vertical,

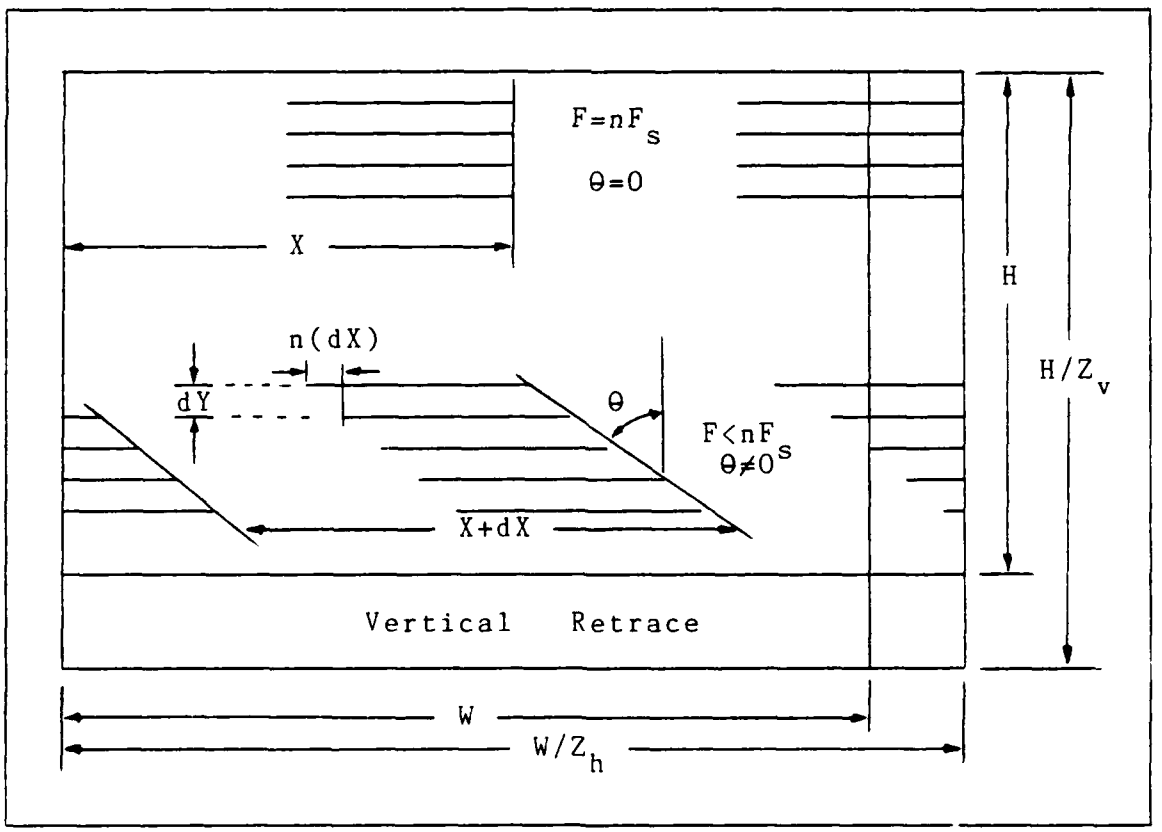


Figure 9. Formation of Video Pattern Angles for n = 2

with an angle of 0 degrees. This pattern angle θ describes the orientation of the pattern with respect to vertical, where positive angles slant left, and negative angles slant right.

Consider the angle resulting from an input frequency F which differs from a scan rate multiple (SRM) nF_s. The change in the spatial period would be

$$dX = \frac{F_s\ W}{F\ Z_h} - \frac{F_s\ W}{nF_s\ Z_h} \tag{3a}$$

which reduces to

$$dX = \frac{W\ (nF_s - F)}{Z_h\ nF} \tag{3b}$$

Since F has about n periods per scan, the beginning of each period is shifted over a distance n(dX) from the beginning of a period on the line above. At the same time, the trace has moved vertically a distance

$$dY = \frac{H}{Z_v\ N_s} \tag{4}$$

where N_s is the number of lines in a frame. The angle of the pattern from one line to the next can then be given by

$$\theta = \arctan\left(\frac{n\ dX}{dY}\right) \tag{5a}$$

$$\theta = \arctan\left(\frac{W\ Z_v\ (nF_s - F)\ N_s}{Z_h\ H\ F}\right) \tag{5b}$$

Here, nF_s is the SRM closest to F. For frequencies higher than the SRM, the angle is negative and the pattern slants right. Conversely, frequencies lower than the SRM result in patterns slanting to the left. The angle of the pattern is also dependant on the aspect ratio of the video display. For a common TV or monitor

$$\frac{W}{H}\frac{Z_v}{Z_h} \simeq .68 \tag{6a}$$

$$N_s = 262.5 \tag{6b}$$

$$F_s = 15750 \text{ hz} \tag{6c}$$

This reduces equation (5b) to

$$\theta = \arctan\left(\frac{178.5\ (15750n - F)}{F}\right) \tag{7}$$

A plot of the absolute value of pattern angles for frequencies up to 100 Khz is shown in Figure 10. The fiqure shows that the bands of frequencies where the pattern angles are not close to 90 degrees are narrow, with a large angular change for a relatively small change in frequency.

Pattern Angle Effects on Spatial Frequency. The spatial period of the pattern varies as a function of the pattern angle. If X is the horizantal dimension of the spatial period of the pattern, then the true spatial period of the pattern X_p can be described by

$$X_p = X \cos \theta \tag{8}$$

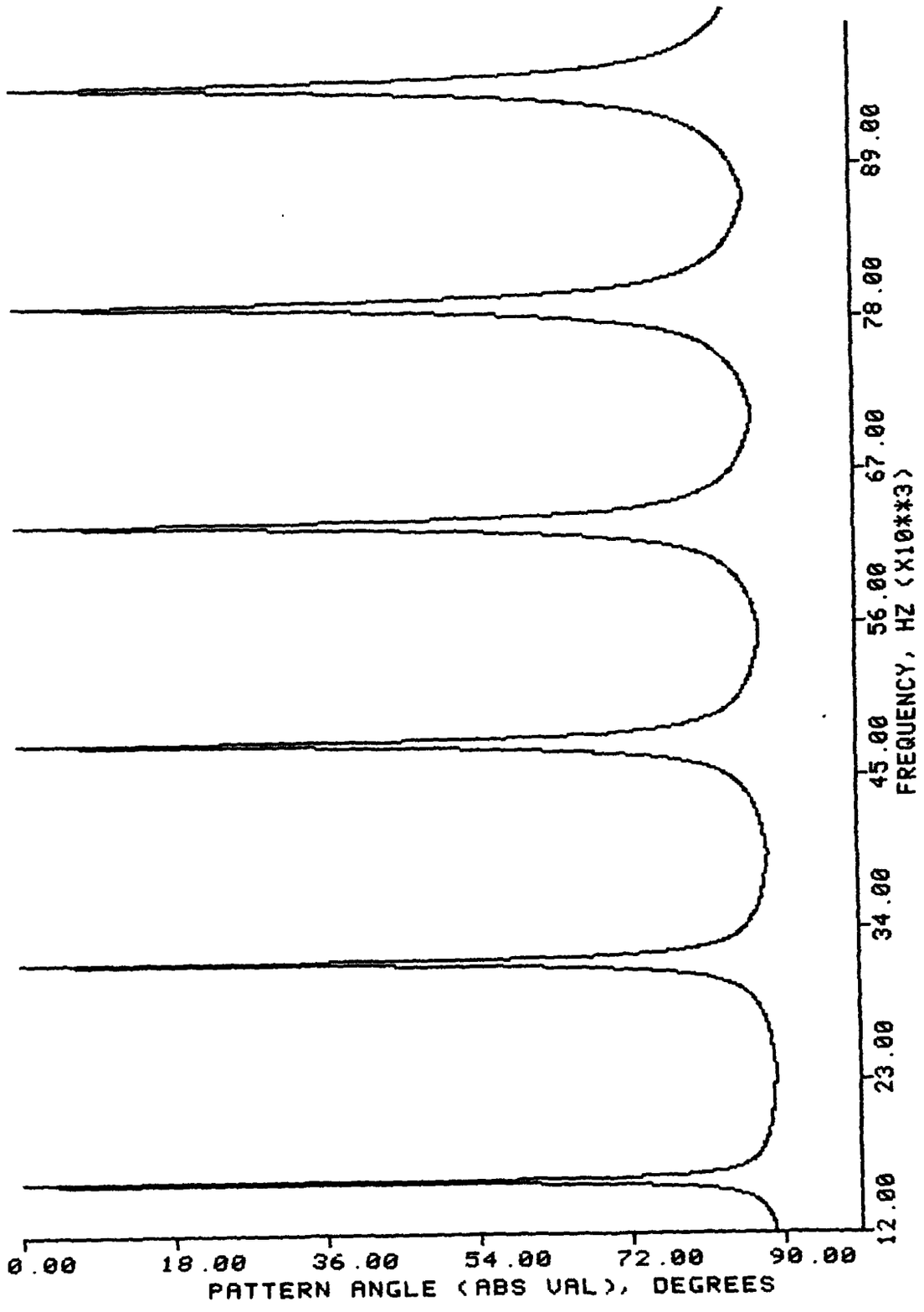


Figure 10. Absolute Value of Pattern Angles for Frequencies up to 100 Khz

This is shown in Figure 11. The relationship between the pattern's spatial frequency and its horizantal component is then

$$F_p = \frac{1}{\cos \theta} F_h \qquad (9)$$

The two dimensional Fourier transform of the sine grating pattern of spatial frequency F_p is a set of points corresponding to the negative and positive frequencies of the pattern. Their orientation will depend on the angle of the grating pattern. The image in the transform plane is rotated 90 degrees and reversed like a mirror image. The transform of the pattern in Figure 11 is shown in Figure 12.

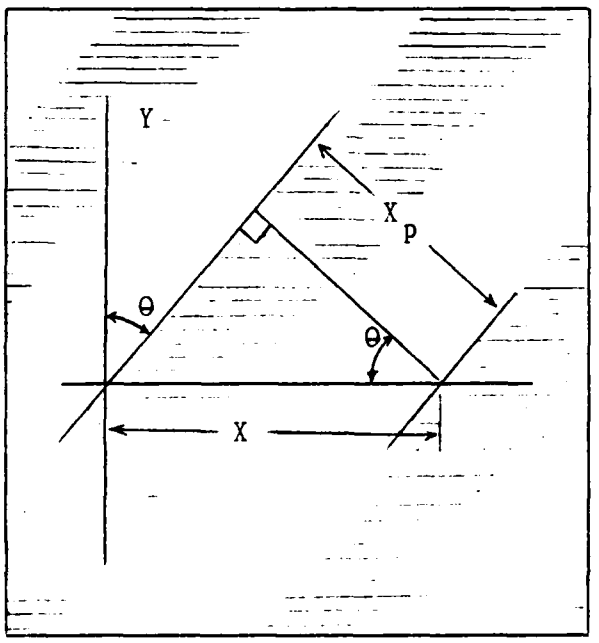


Figure 11. Spatial Frequency Dependence on Pattern Angle

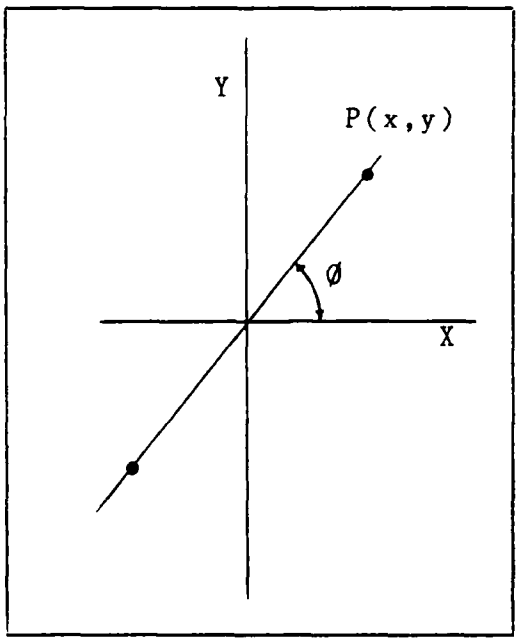


Figure 12. Corresponding Transform Image Pattern

The angle of the transform ϕ is simply the negative of the pattern angle θ, referenced to the horizantal to account for the 90 degree rotation of the transform.

$$\phi = -\theta \tag{10}$$

Consider the position of the transform points as the angle θ is varied. When $\theta = 0$, the position of one of the points $P(x,y)$ is $(C_1,0)$ where C_1 is some constant dependant on the transform process. The coordinates of a transform point $P(x,y)$ can be given as

$$x = C_1 F_p \cos \phi \tag{11a}$$

$$y = C_1 F_p \sin \phi \tag{11b}$$

Using equations (9) and (10), the coordinates of a transform point for a given input frequency F are

$$x = C F (1/\cos \theta) \cos (-\phi) \tag{12a}$$

$$y = C F (1/\cos \theta) \cos (-\phi) \tag{12b}$$

C, again, is a constant dependant on the transform. Since $\sin(-\theta) = -\sin(\theta)$ and $\cos(-\theta) = \cos(\theta)$, this gives

$$x = C F \tag{13a}$$

$$y = -C F \tan \theta \tag{13b}$$

Using θ as described in equation (5) and rewriting F as $nF_s + (F - nF_s)$,

$$x = C nF_s + C (F - nF_s) \tag{14a}$$

$$y = C \frac{W Z_v}{H Z_h} N_s (F - nF_s) \tag{14b}$$

The difference $F - nF_s$ describes the "distance" of the input frequency from the SRM, which will be called F_d:

$$F_d = F - nF_s \tag{15}$$

Examining the change in position of the transform points for a given change in F_d gives

$$\frac{dX}{dF_d} = C \tag{16a}$$

$$\frac{dY}{dF_d} = C \frac{W Z_v}{H Z_h} N_s \tag{16b}$$

Using the numbers in equations (6a), (6b), and (6c), it is evident that the rate of change in the y direction is well over 100 times greater than that in the x direction. Therefor, for all F near nF_s,

$$x = C\, nF_s \tag{17a}$$

$$y = C \frac{W Z_v}{H Z_h} N_s (F - nF_s) \tag{17b}$$

These two equations demonstrate that the set of points in the transform image corresponding to input frequencies near a SRM frequency form a pair of straight lines at a distance from the origin proportional to the SRM number n. Each input frequency near the SRM will have a corresponding set of points within the lines which form a unique angle.

Modulation. In order to move the baseband audio signal up in the frequency spectrum so that the desired pattern angles can be achieved, the signal will be modulated using a balanced modulator which has an output containing the sum and difference of the audio input frequency and the carrier frequency. If the audio signal is $E_a \cos wt$ and the carrier signal is $E_c \cos pt$ the output of the modulator will be

$$E_o = KE_aE_c [\cos (w+p)t + \cos (w-p)t] \qquad (18)$$

where K is a constant dependant on the characteristics of the modulator [10:392]. This could also be achieved using amplitude modulation, but the carrier frequency would not be suppressed.

By modulating the audio signal with a carrier slightly below a SRM frequency, the spectrum of the output signal can be placed so that the SRM falls within the upper or lower sideband, as shown in Figure 13.

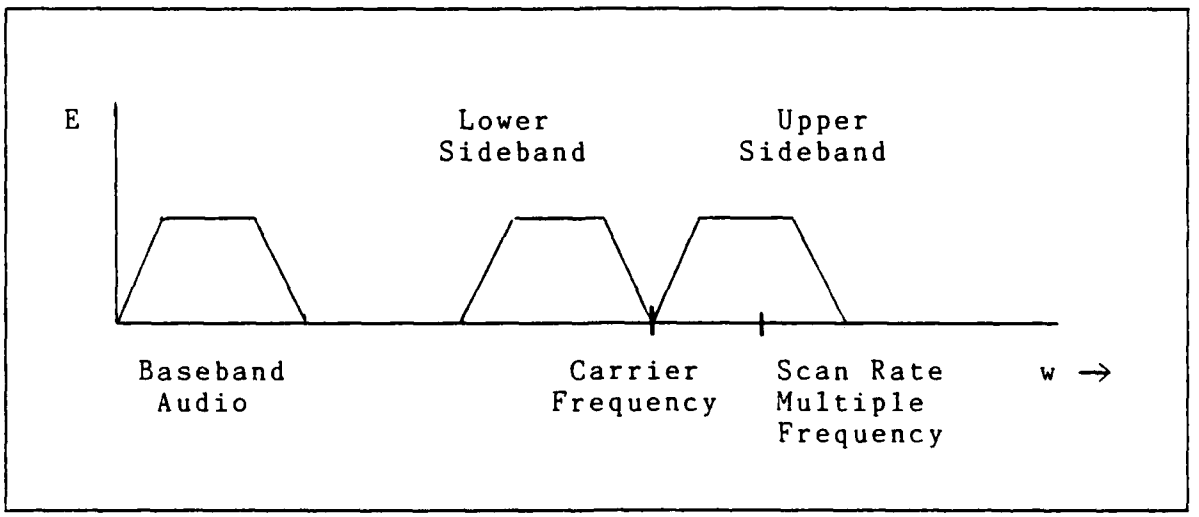


Figure 13. Balanced Modulator Output Spectrum

Frequency Selection

Vowel sounds have resonant frequencies called formants. For each of the vowel sounds, these formant frequencies take on different values. The first formant lies between 190 and 800 hz, while the second may fall somewhere between 780 and 2400 hz. A plot of these first two formants shows their values for several of the vowel sounds [1:154; 2:63; 4:102; 6:60; 9:166,175]. Though exact values may vary, all sources are in general agreement on the relative locations of the formant peaks.

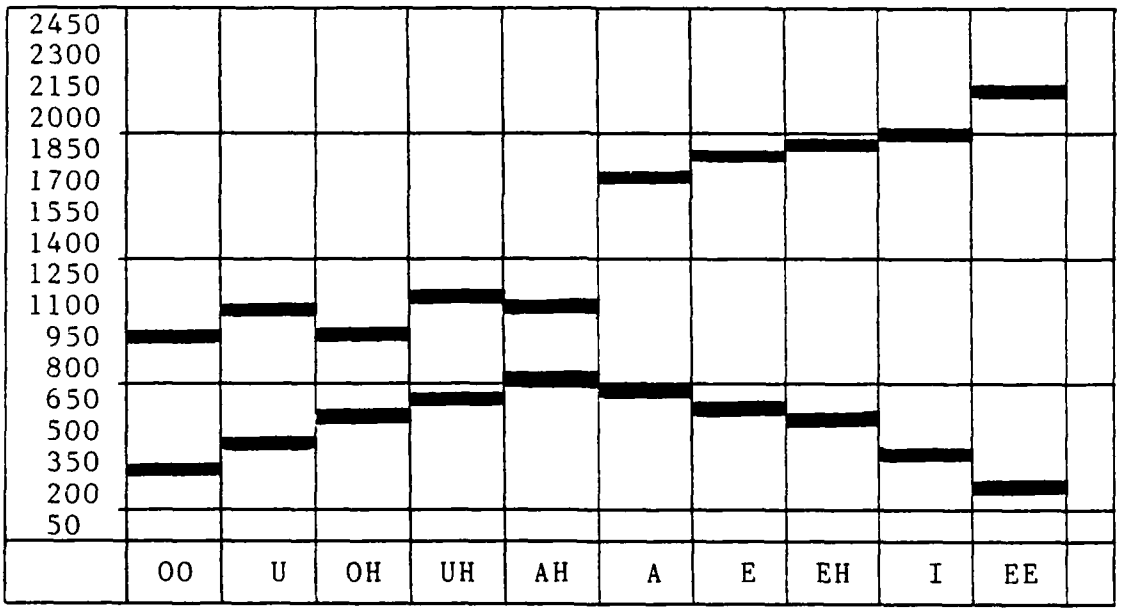


Figure 14. Values of Formants for Several Vowel Sounds

Each Formant band will be modulated so that the "center" of the band will correspond to an exact SRM. For band 1 (first formant) this is about 580 hz. For band 2 (second formant) it is around 1370 hz.

Scan Rate Multiples. The standard video scan rate is 15750 hz. At frequencies above 100 Khz processing becomes more difficult because of the bandwidth limitaions of the equipment, so the first six multiples are available for the display.

Table II

Video Scan Rate Multiples

Multiple	Frequency
1	15750
2	31500
3	47250
4	63000
5	78750
6	94500

The multiples used in this project are 47.25 Khz (n=3) and 94.5 Khz (n=6), which were chosen because they gave approximately 2 and 4 cycles per screen. This gave the greatest separation in the Fourier transform plane image. In order to find the carrier frequency required to place the SRM at the center of the upper side band of the modulated output, the following equation is used:

$$F_c = nF_s - F_b \tag{19}$$

where F_c is the carrier frequency and F_b is the frequency at the center of the formant range. The frequency input to

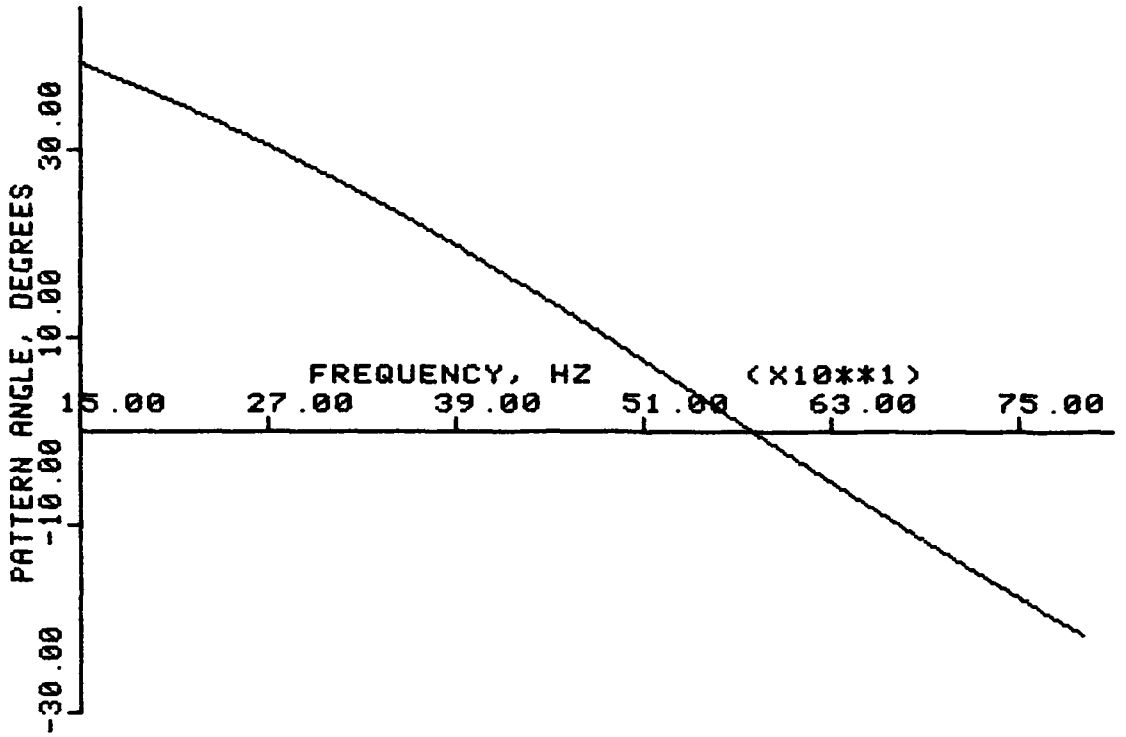


Figure 15. First Formant Pattern Angles

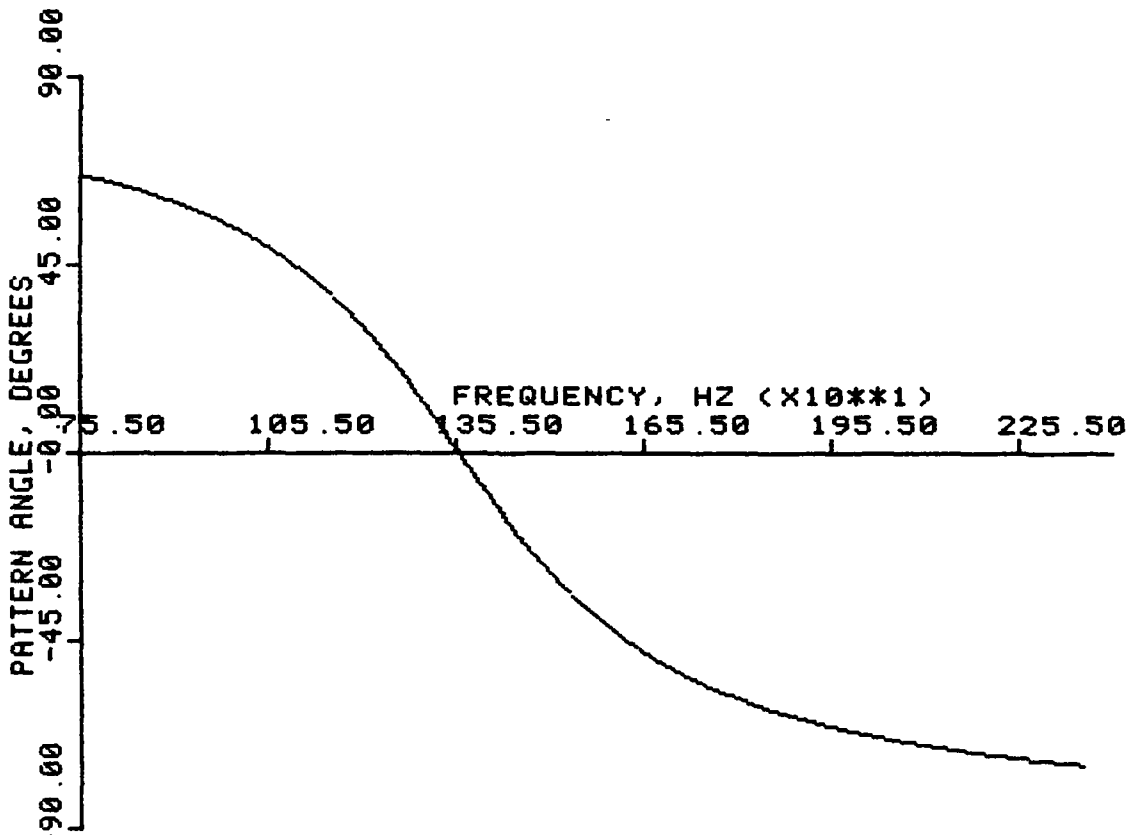


Figure 16. Second Formant Pattern Angles

Table III

Carrier Frequency Selection

Formant	SRM n	Frequency nF_s	Band Center F_b	Carrier F_c
1	6	94500	580	93920
2	3	47250	1370	45890

the video display is then the sum (and difference) of the carrier frequency and the audio frequencies. The pattern angles for both bands are plotted in Figures 15 and 16. Table III shows the values selected for modulating the first and second formant bands.

Spatial Filtering. Unwanted frequencies including the lower sideband and any carrier which leaks through may be filtered out spatially in the Fourier transform plane. This would be virtually impossible to do electrically. For example, consider the spread of less than 400 hz between the upper and lower side bands of the modulated signal for the first formant. At 93 Khz, a filter capable of blocking out the lower side band while leaving the upper one intact would require a "Q" of several hundred. This can be done spatially, however, by simply masking out any points beyond a certain angle. Figure 17 shows the spatial pass bands projected for both formant bands. The angles were taken from Figures 15 and 16. Note that the 90 degree rotation of the transform is not shown.

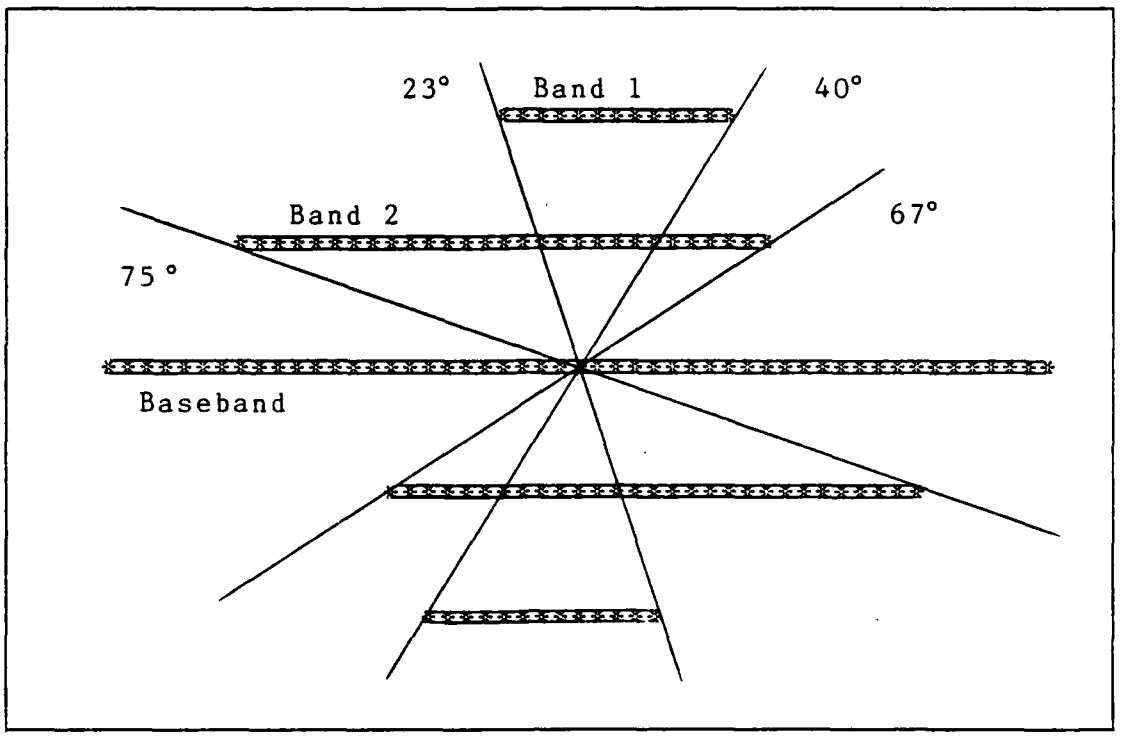


Figure 17. Spatial Filter Pass Bands

Prediction of Vowel Sound Transform Images

Using the frequency values from Figure 14 and their corresponding angles as given by equations 7 and 19, the peak locations in the transform images for the vowel sounds can be predicted. The values are listed in Table IV, and the images are shown in Figure 18. Each dot represents the position of a bright spot in the Fourier transform image. The center dot is simply a reference point to identify the center of the transform image. For convenience, the 90 degree rotation of the transform has been removed, making the transform angles the mirror image (negative) of the video image pattern angles.

Table IV

Theoretical Values for Pattern Angles
of Vowel Sounds Based on Values from Figure 14

Vowel	Formant Frequencies (hz)	Pattern Angles (degrees)
OO	300 870	27.9 61.9
U	440 1020	14.8 52.3
OH	580 930	0.0 58.6
UH	640 1190	-6.5 32.8
AH	730 1090	-15.8 45.7
A	660 1720	-8.6 -53.5
E	530 1840	5.4 -60.9
EH	510 1950	7.5 -66.9
I	390 1990	19.8 -66.9
EE	270 2290	30.4 -73.8

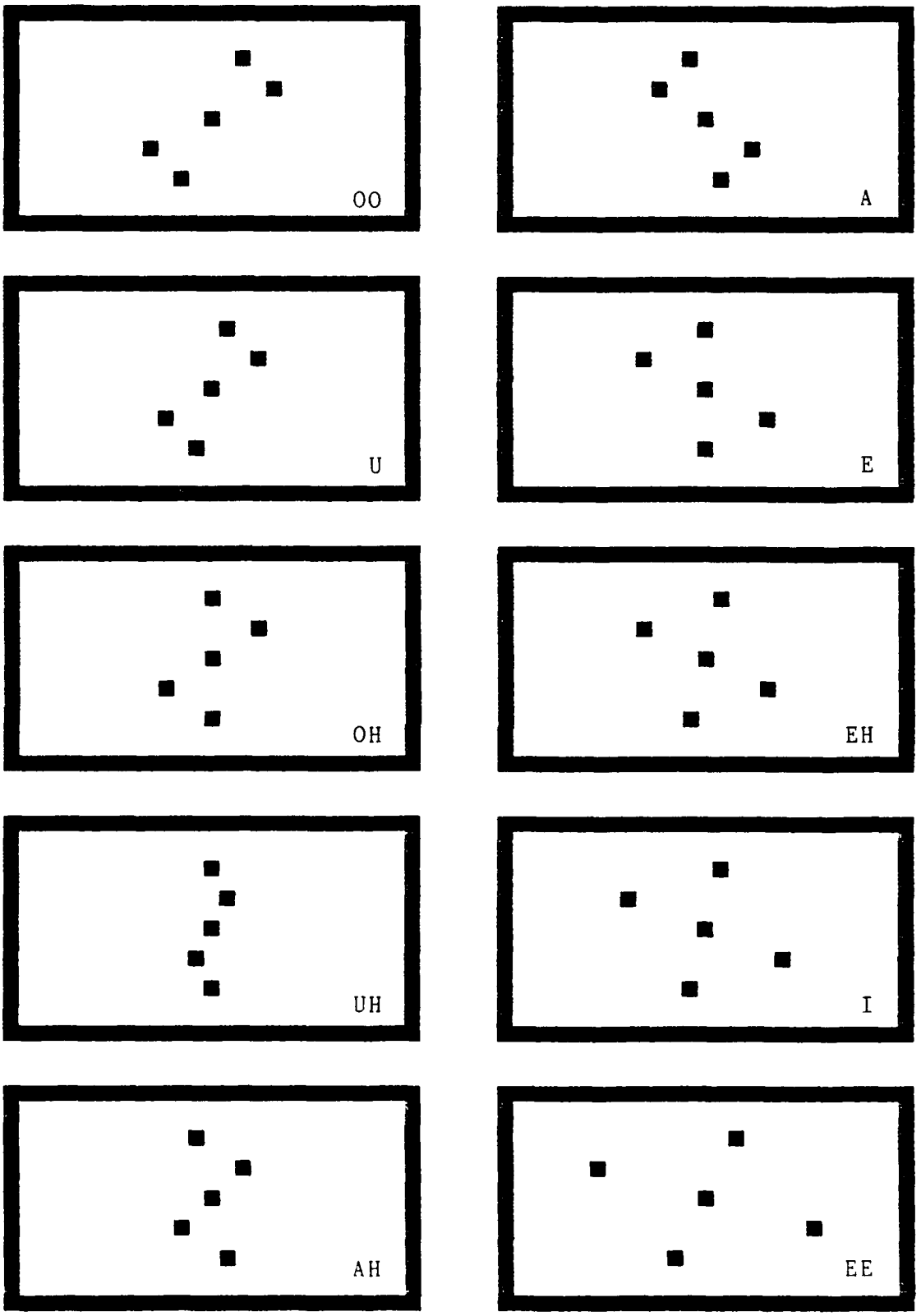


Figure 18. Theoretical Transform Image Peak Locations for Vowel Sounds

Fricative and Stop Sounds

The fricative and stop sounds are generally low power and broadband in nature. Those sounds that are unvoiced will give very little information in the formant bands, and so the base-band audio signal must also be processed if these sounds are to be identifiable.

The processed image signal will therefore consist of three parts: the direct audio baseband and the two modulated bands bands corresponding to the first and second formant frequencies. These three signals will be mixed and displayed simultaneously in a single image.

III. EQUIPMENT

Original plans for this project included performing optical transforms and filtering to test the output of the system. The apparatus required to do the optical Fourier transform is still unavailable, so this operation will be simulated using digital image processing techniques. In order to maintain the integrity of the experiment, no operations should be carried out numerically which cannot be duplicated optically.

System Overview

The speech signal is picked up by the microphone and passed through an audio amplifier. At this point, it is either recorded on magnetic tape or fed directly into the processing equipment.

Pre-processing circuitry consists of a pre-emphasis (high-pass) filter, a low-pass filter, and an automatic gain control. From here the signal is modulated or connected directly to the video mixing circuitry.

The video mixing circuitry inserts blanking into the signal and adds the appropriate sync pulses to form a composite video signal. This signal may either be recorded by the video recorder or directly fed to the video digitizer.

The video digitizer "grabs" one frame of the video picture and stores it on disk in the computer system. The computers are used to simulate the two dimensional Fourier transform and spatial

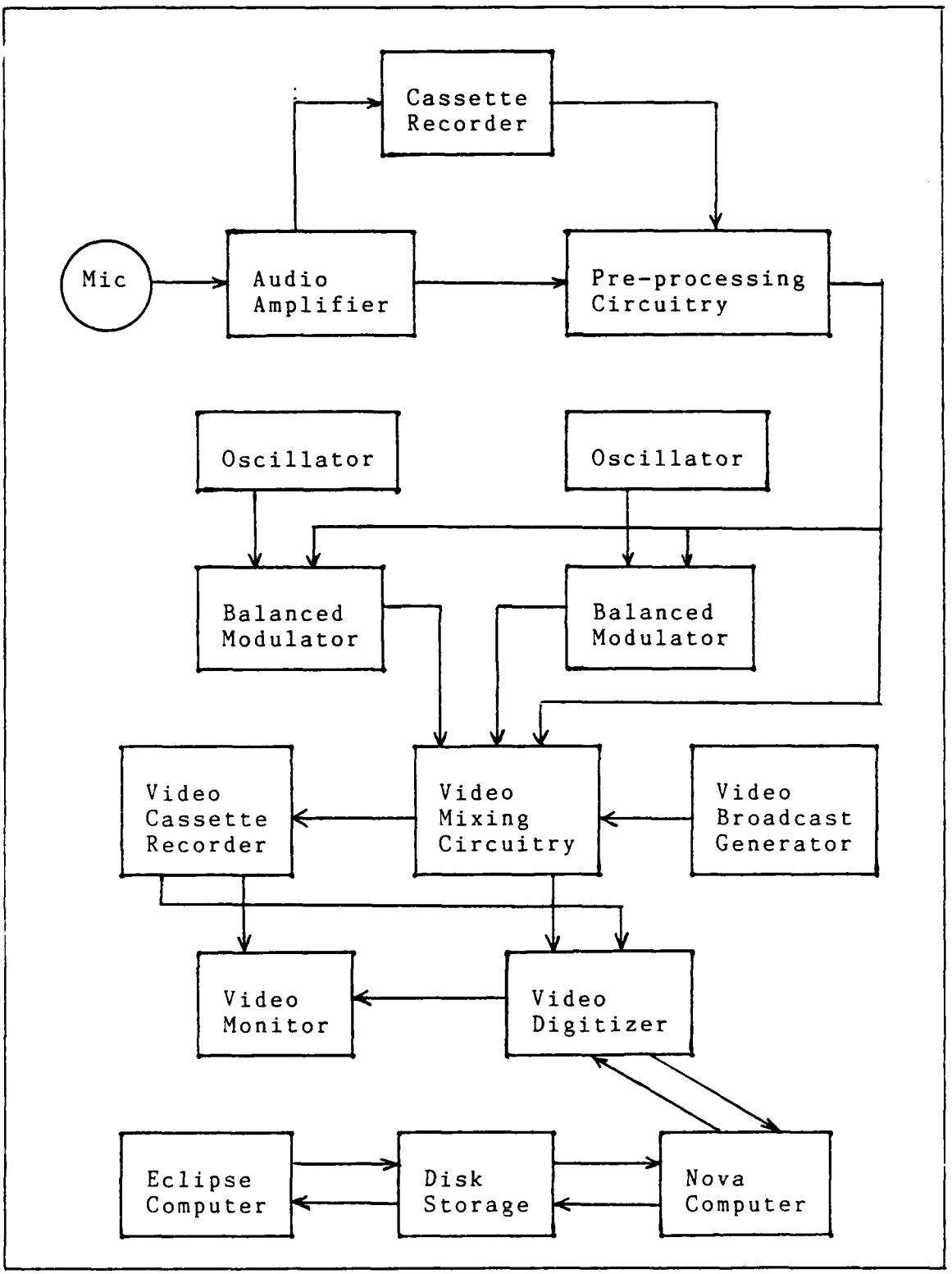


Figure 19. System Block Diagram

filtering that would otherwise be carried out optically. The transform image can then be displayed on the monitor and recorded by the video recorder. A system block diagram is shown in Figure 19 and a list of commercial equipment used is in Appendix A.

Pre-processing Circuitry

Front end circuitry shown in Figure 20 consists of three major parts including the pre-emphasis filter, low-pass filter, and automatic gain control circuits. The circuit design with only minor modification is taken from the work of Hussain [3:8,11,13].

Pre-emphasis Filter. IC-1 and associated components make up a high-pass filter with about 6db gain per octave above 500 hz. The "Balance" potentiometer controls the D.C. offset of the output for all three parts of the circuit.

Low-pass Filter. IC-2 is the active element for the low-pass filter, which limits the baseband to around 10 Khz. The combined response of both filters is shown in Figure 21.

Automatic Gain Control. IC-3 and the two transistors form the automatic gain control circuit. The active elements in the feedback loop of the op-amp vary its gain inversely proportional to the amplitude of the signal. The circuit provides about 60 db of compression. IC-4 is an output buffer and also restores the original polarity of the signal. The "Direct" and "Modulator" potentiometers provide separate attenuation for each of the output channels.

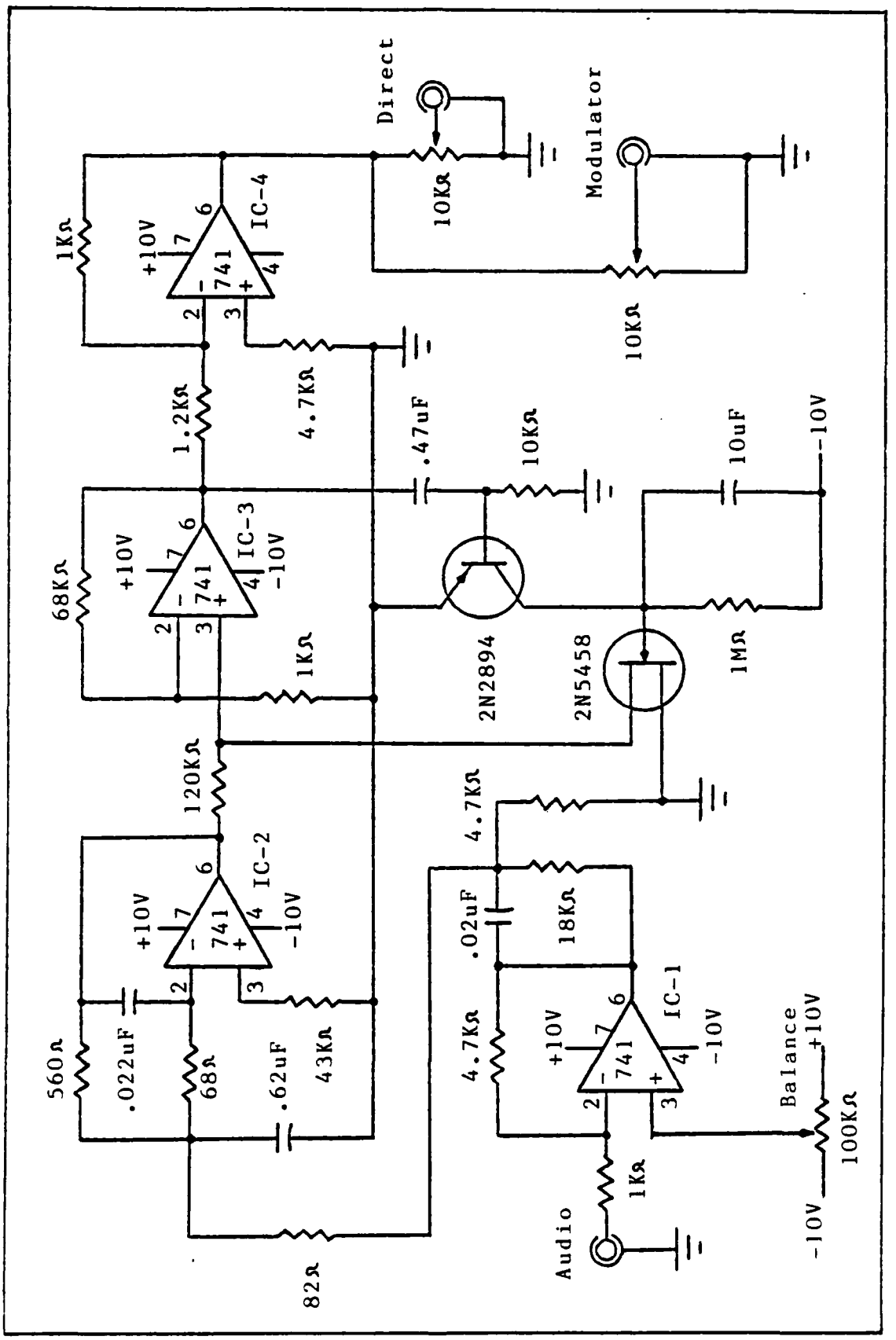


Figure 20. Schematic Diagram for Pre-processing Circuitry [3:13]

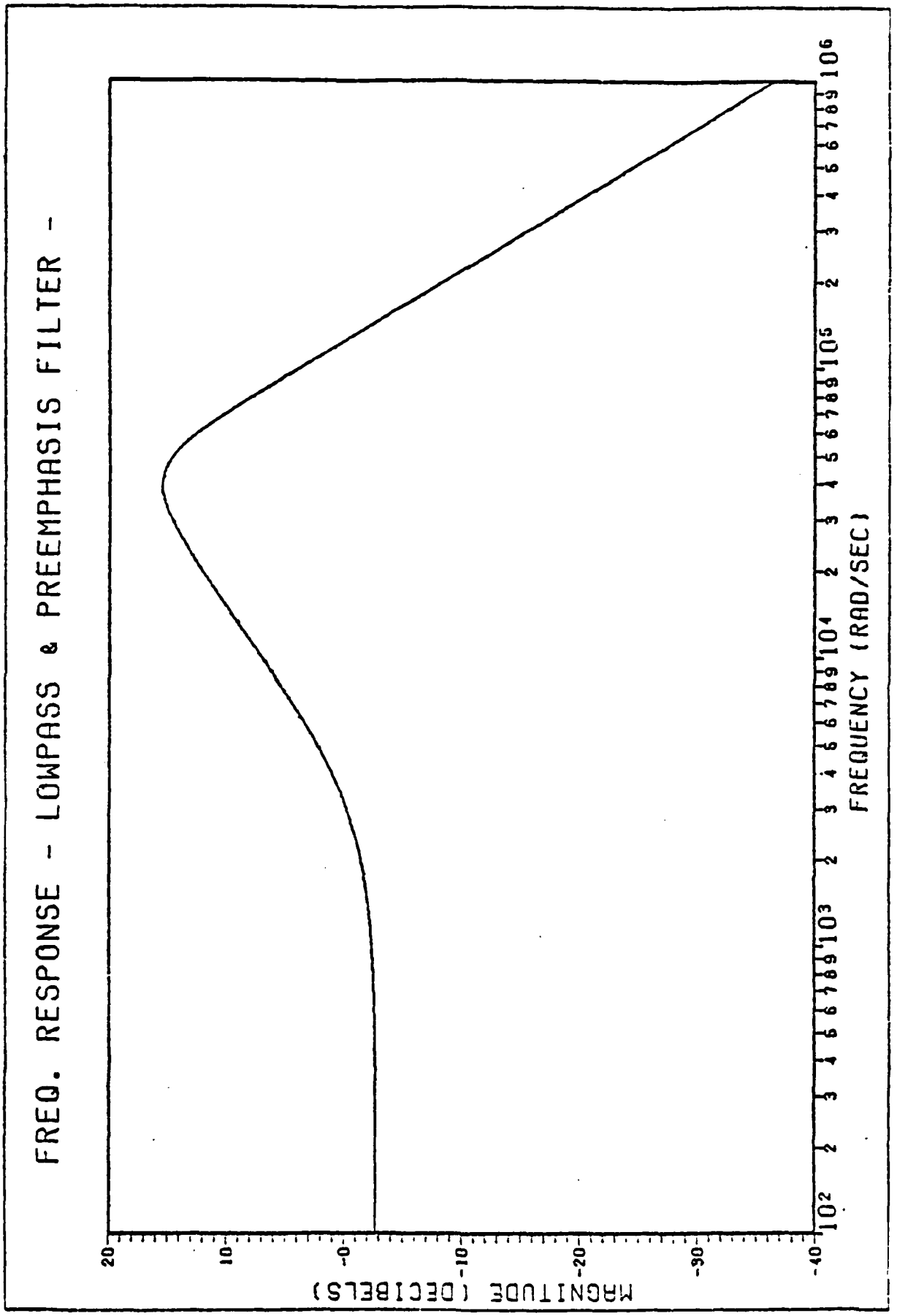


Figure 21. Combined Pre-emphasis / Low-pass Filter Response [3:39]

Modulation Circuitry

The audio signal is modulated using a balanced modulator circuit, then band-pass filtered to help control noise and stray oscillations in the video mixing circuitry. The circuitry shown in Figure 22 is duplicated for each of the two modulated band channels.

The modulation circuit is based on an LM1496 integrated circuit and operates in the suppressed carrier mode. The circuit is based on National Semiconductor's application circuit [5:10-102]. A sinusoidal carrier is supplied by a commercial signal generator, and the output is taken between the positive side of the balanced output and ground. The "Offset" and "Carrier Null" potentiometers control the symmetry of the modulated waveform.

Video Mixing Circuitry

The final processing circuitry is required to blank the signal during retrace periods and insert the appropriate sync pulses. The inputs from the video generator are negative pulses of about eight volts magnitude. The two transistor input stages are used to convert these to TTL compatible pulses.

The first 7474 flip-flop and the 74123 one-shot devices form a circuit which detects the longer duration pulses of the vertical sync. This toggles the second 7474 flip-flop which enables or overides the blanking pulse stream through the 7432 OR gate, which effectively blanks every other field of the video output.

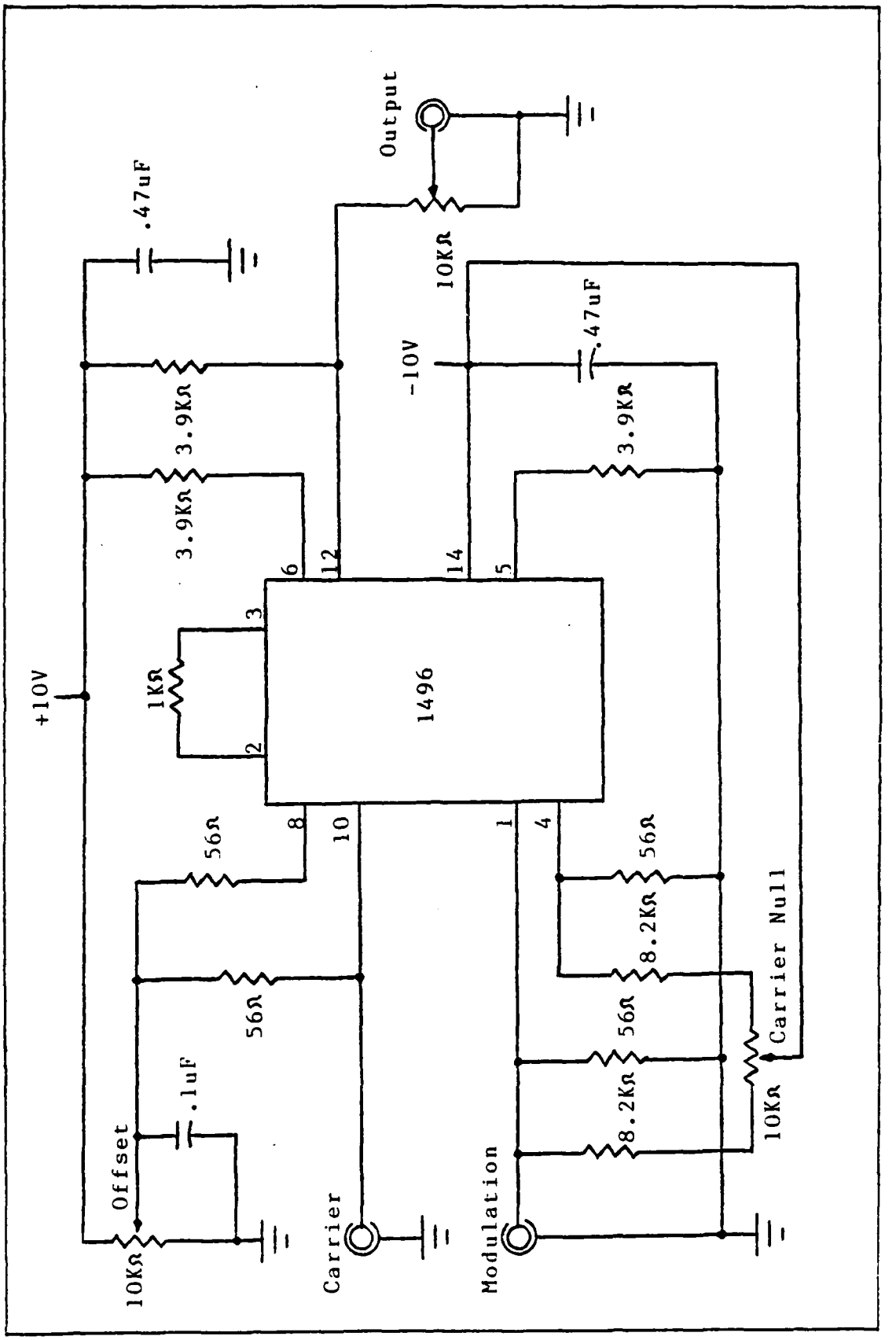


Figure 22. Schematic Diagram for Modulation Circuitry [5:10-102]

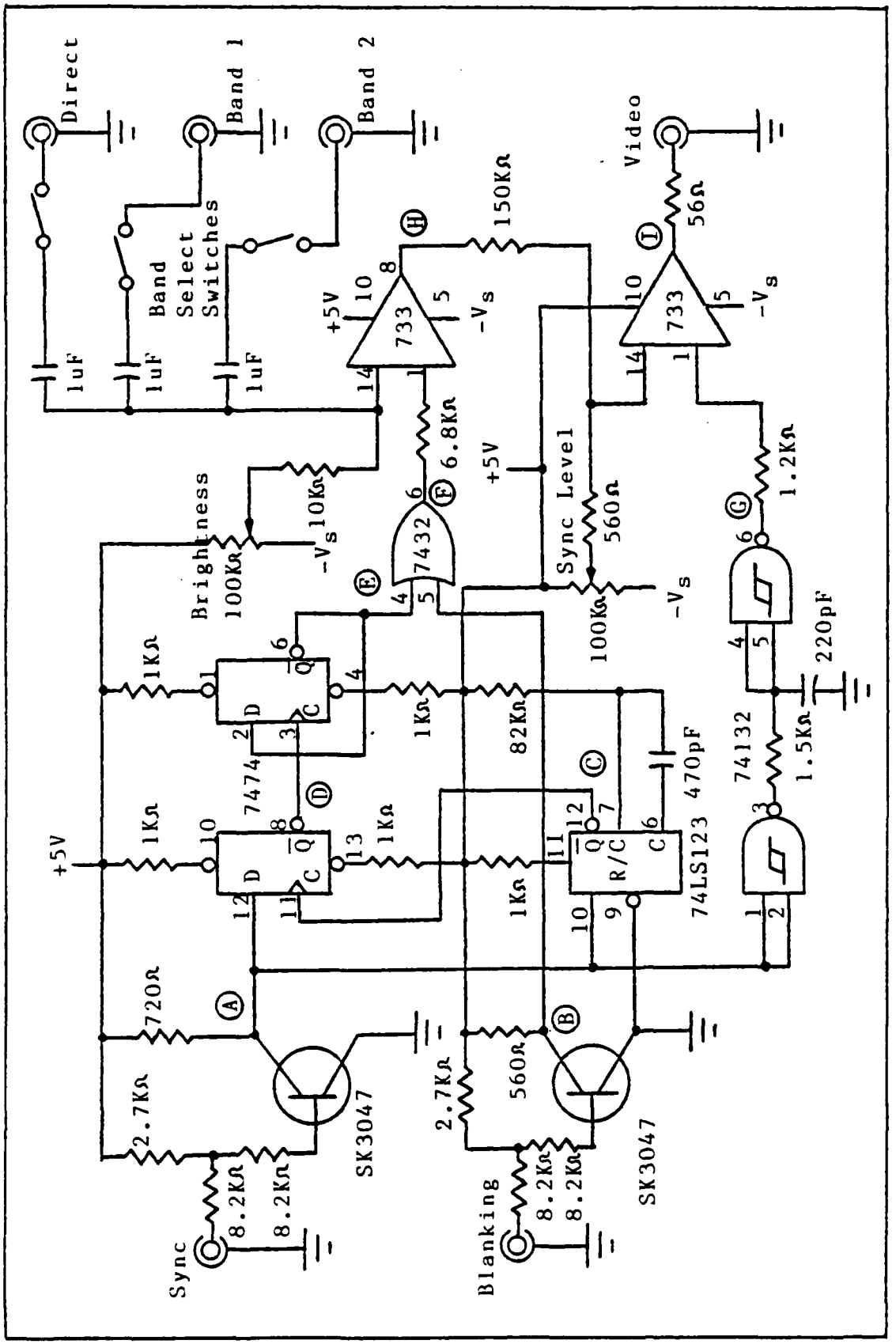


Figure 23. Schematic Diagram for Video Mixing Circuitry

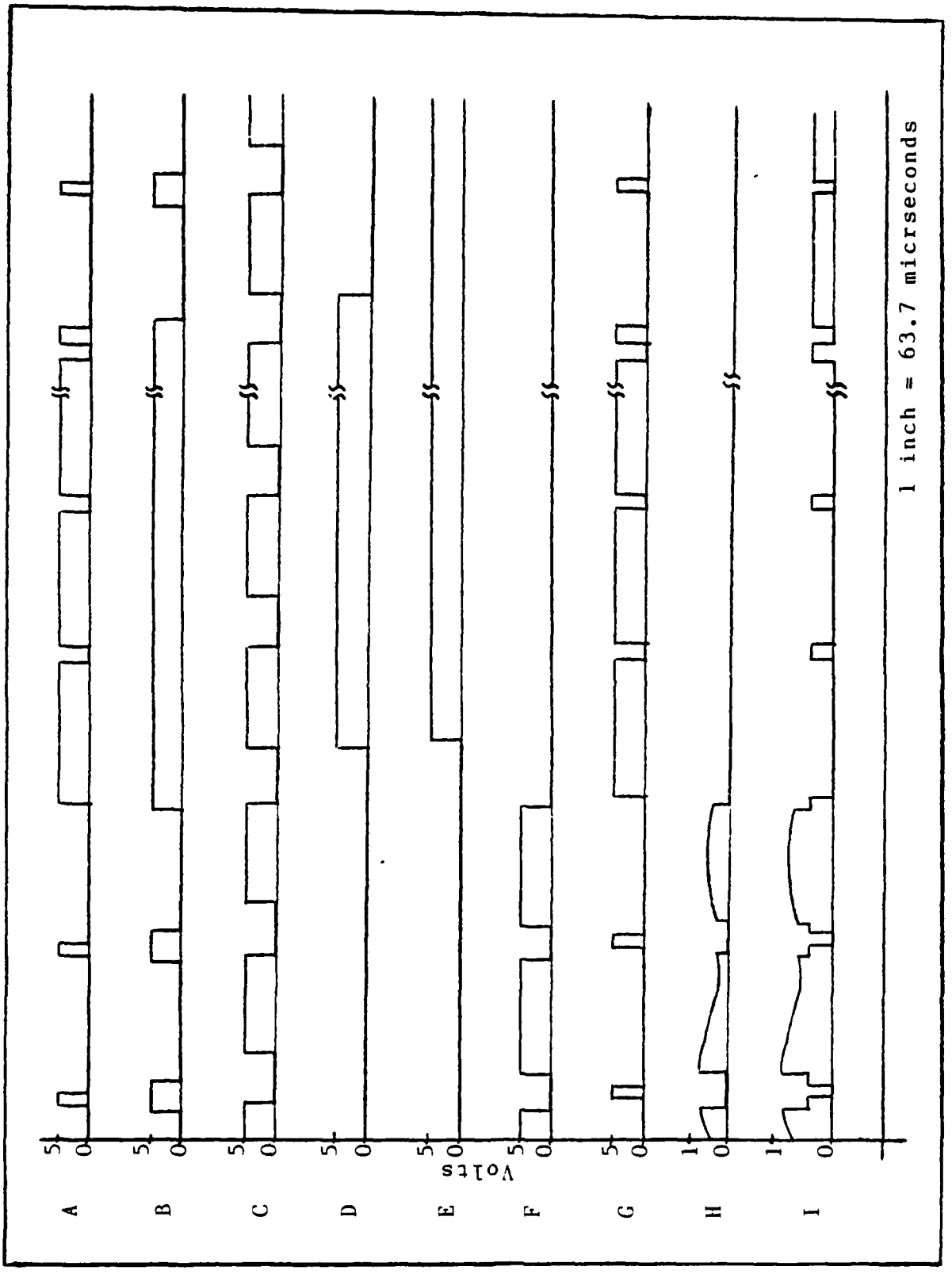


Figure 24. Timing Diagram for Video Mixing Circuitry

The baseband or modulated audio signal is summed with a D.C. offset and sent to the inverting input of the first 733 wide band op-amp. The blanking pulse stream is sent to the non-inverting input. This saturates the output to its lowest level whenever the blanking signal is high. When the blanking signal is low, the output is proportional to the audio input plus the D. C. offset. This signal is then fed to the second 733 op-amp.

The processing of the blanking pulses introduces a time delay in this signal due to the propogation delay through the digital devices. In order to maintain the integrity of the front porch of the horizantal sync pulse, the sync signal must also be delayed. This is accomplished using a pair of Schmitt trigger NAND gates and an R-C differentiator.

The delayed sync pulse is then added to the blanked audio signal in the second 733 op-amp. The minimum value of the video output signal is dependant on the supply voltages, and the negative supply voltage V_s will be adjusted to determine the final D.C. level of the signal. Allignment procedures are further discussed in Appendix B. Timing diagrams for various points in the circuit are shown in Figure 24. They show the interaction of the various blanking and sync pulses in the formation of the composite video output.

Construction Techniques

All circuitry was originally bread-boarded, but proved to be noisy and prone to stray oscillations. Final construction was on

printed circuit boards using point to point wiring. The printed circuit boards were mounted in an aluminum chasis and BNC connectors were provided for external connections to oscillators, the video broadcast generator, and audio or video equipment.

Computer Equipment

The Octek video digitizer works under control of the Nova computer. The Nova and Eclipse computers share a hard disk memory system, which allows video signals to be digitized and stored by the Nova system, then operated on by the Eclipse system, and finally output again through the Nova system. This gives the advantage of using the greater storage and faster processing capabilities of the Eclipse system.

IV. Procedures

Two sets of data are taken and processed. The first is a set of steady state sounds including ten vowels and ten fricatives for four different speakers; three male and one female. The purpose of this first set is to check the repeatability and separability of the transform images. The second data set includes complete words; the numbers zero through nine. The purpose of this data set is to simulate the output of a real time processing system. Except for some additional recording, both sets of data are processed in the same manner.

Initial Recording

Raw speech is recorded onto audio cassette tape. This step is not necessary for regular operation of the system, but is done for convenience. It also provides the opportunity for the same speech sample to be processed in a number of different ways. Speech samples are recorded with common laboratory background noise: there is no effort made to make them "noise free".

Video Processing

Either pre-recorded or live speech signals are turned into video signals using the circuitry described in the previous chapter. These video signals are then sent to the Octek board for digitizing.

Before processing data, the equipment is turned on and allowed five minutes to warm up. Once initially alligned as

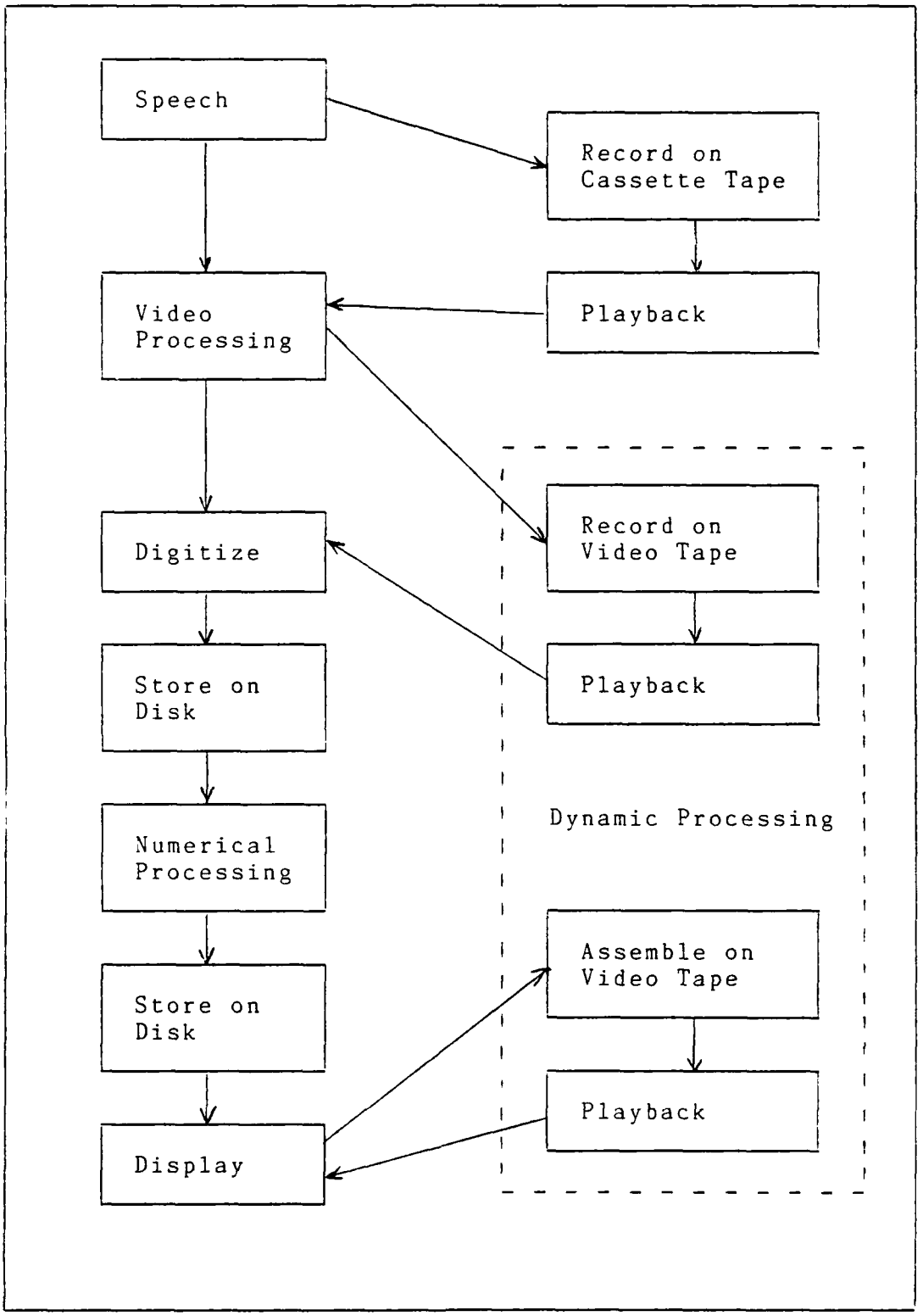


Figure 25. Data Processing Flow Diagram

described in Appendix B, the only adjustments that commonly need to be made are to the frequencies of the oscillators and the "Balance" control in the pre-processing circuitry. This is usually required only once at the beginning of each data collection session.

Frequencies of the oscillators are checked using a frequency counter before and after taking data. A deviation of plus or minus 5 hz is considered acceptable. Even at the point of greatest angular change (θ near 0 degrees) this introduces a change in pattern angle of only about one degree.

Digitizing Video Signals

Under control of the Nova computer, the Octek board "grabs" one field of the video input signal, resolves it into 16 gray levels, and stores it in memory. This can then be stored on disk and later processed by the Eclipse computer.

Steady State Processing. The term "steady state" would indicate an unchanging phenomena, which is not necessarily true of the "steady state" speech samples taken. Some are less than a second in duration. This requires quick reaction time on the part of the operator, since the return key on the computer terminal must be pressed to start the digitizing process, and must be done just as the speech event takes place. It often takes several tries, and for this reason the pre-recorded speech is more convenient to work with than a live subject.

Dynamic Processing. For the processing of complete words, the output of the video processing circuitry is recorded using a video tape recorder. The moving pictures are then played back

into the Octek board using the "Pause" and "Frame Advance" features of the recorder, which allows the Octek board to "grab" and digitize the first field of each frame.

System Modification. The blanking of every other field is necessary only when working with optical devices. Both the Octek board and the video tape recorder work with every other field of the video signal, which makes this even more unnecessary. In fact, it may be an annoyance, since the chances of grabbing the blank field are 50/50. The every-other-field blanking may be disabled by removing the wire from pin 6 of the 7474 flip-flop in the video mixing circuitry and tying it to ground.

Numerical Processing

The purpose of the numerical processing carried out in this project is to simulate the operations that would be done optically in a real time system. This includes two major tasks: performing a two dimensional Fourier transform and spatially filtering the frequency domain data. Both of these can be done in a single operation by computing only those points in the transform plane which lie in the pass bands of the spatial filter.

Spatial Filtering. The third and sixth SRM frequencies give patterns with two and four cycles in the time domain picture, respectively. This is due to the scan efficiency of the video monitor and the fact that the Octek board does not digitize the entire screen, which further reduces the scan efficiency. The set of points for the third SRM band lie in a line

at a distance corresponding to the second harmonic of the screen dimension, and the set of points for the sixth SRM lie in a line at a distance corresponding to the fourth harmonic of the screen dimension. The screen dimension in this case is that of the digitized portion of the screen.

The transform algorithm computes a value for whole multiples of the spatial frequency whose period is the screen dimension; one point for each harmonic. The points that lie in the pass bands of the filter are those in the lines distance two and four points from the origin, and within the angles described in Figure 17. These points are shown in Figure 26.

Baseband Compression. The baseband points lie in a line which passes through the origin. The first 90 harmonics are computed, but in order to make the display more convenient, the 90 points are compressed into 30 by averaging each set of three adjacent points and displaying the value in a single point. The spatial scaling of the baseband is then one third of the modulated bands. The distance from the origin to the outermost

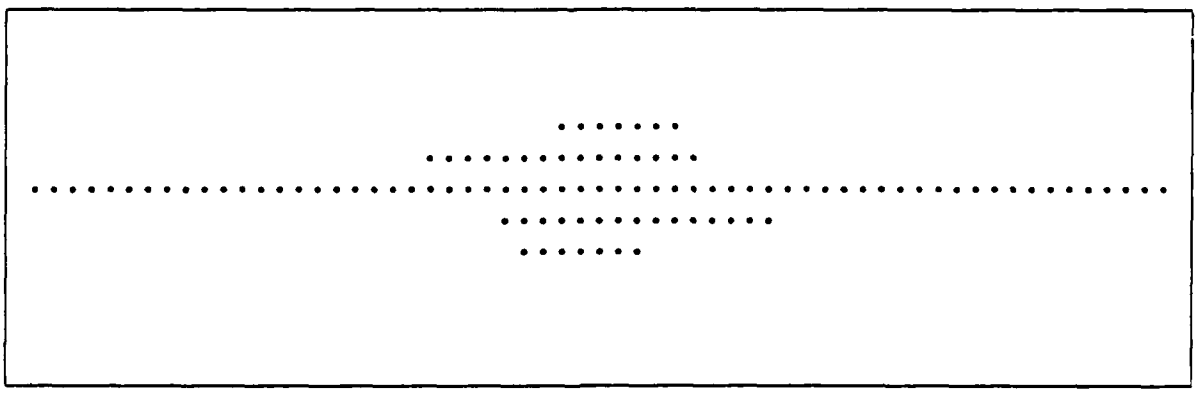


Figure 26. Spatial Filter Pass Band Points

point corresponds to roughly 6700 hz (60 fields per second x 90 harmonics / .8 scan efficiency).

Display of Transform Information

Once the points in the Fourier transform plane are calculated, a video data file is created to display them via the Octek board. Each "point" is displayed as a block of 4 x 4 pixels, so that the image, 61 points wide, nearly fills the screen.

Display of Static Transform Data. The two sets of static data are processed with some variations to facilitate comparisons with the expected results. The vowel sounds are displayed showing only the peaks in the modulated bands. To achieve this, the baseband computations are omitted and each of the modulated bands is scaled separately with a linear scaling. The image is binarized at the highest gray scale level (white) and then negated (gray scale reversed). This leaves an image made up of only the highest energy points in each band.

In order to help distinguish the fricative sounds, the baseband data is displayed separately with the pixel intensity values plotted as a function of their distance from the origin. This brings out small variations which may be hard to see by simply observing the transform images.

Display of Dynamic Transform Data. The transform images are scaled using a square root scaling with respect to an absolute value for each band. The purpose of the square root scaling is to compress the large dynamic range of the transform output into the 16 gray levels available on the Octek display.

Once transform images are created for each time domain image in the word, they are reassembled on video. This is done by recording them one at a time onto tape using video editing equipment. When played back at normal speed, the transform image will change at real time speeds.

V. Results

Static Data

Four sets of data are taken, including three male and one female. Each set consists of ten vowel sounds and ten fricative sounds.

Vowel Sounds. The vowel sounds are presented in Figures 27 through 36. Only the two modulated bands are shown, and the 90 degree rotation of the transform has been removed. Each figure has nine parts, or blocks. The top right-hand block shows the theoretical location of the band peaks as calculated in Chapter II. Below it, in order, are the four pair of experimental values with the transform image on the left and its corresponding peaks on the right. Data set number 4 (bottom) is that of the female.

Experimental values are generally in agreement with theoretical predictions, although there were few "perfect" matches. This is expected, since variations between speakers are unavoidable. More important, the images show that the processing technique is indeed capable of showing the locations of these peaks for various speakers, and that there are definite similarities between the images for the same sound spoken by different speakers.

Adjacent sounds are not all separable; for example, one person's OH may look like another person's UH. This may be attributed in part to low spatial resolution of the system.

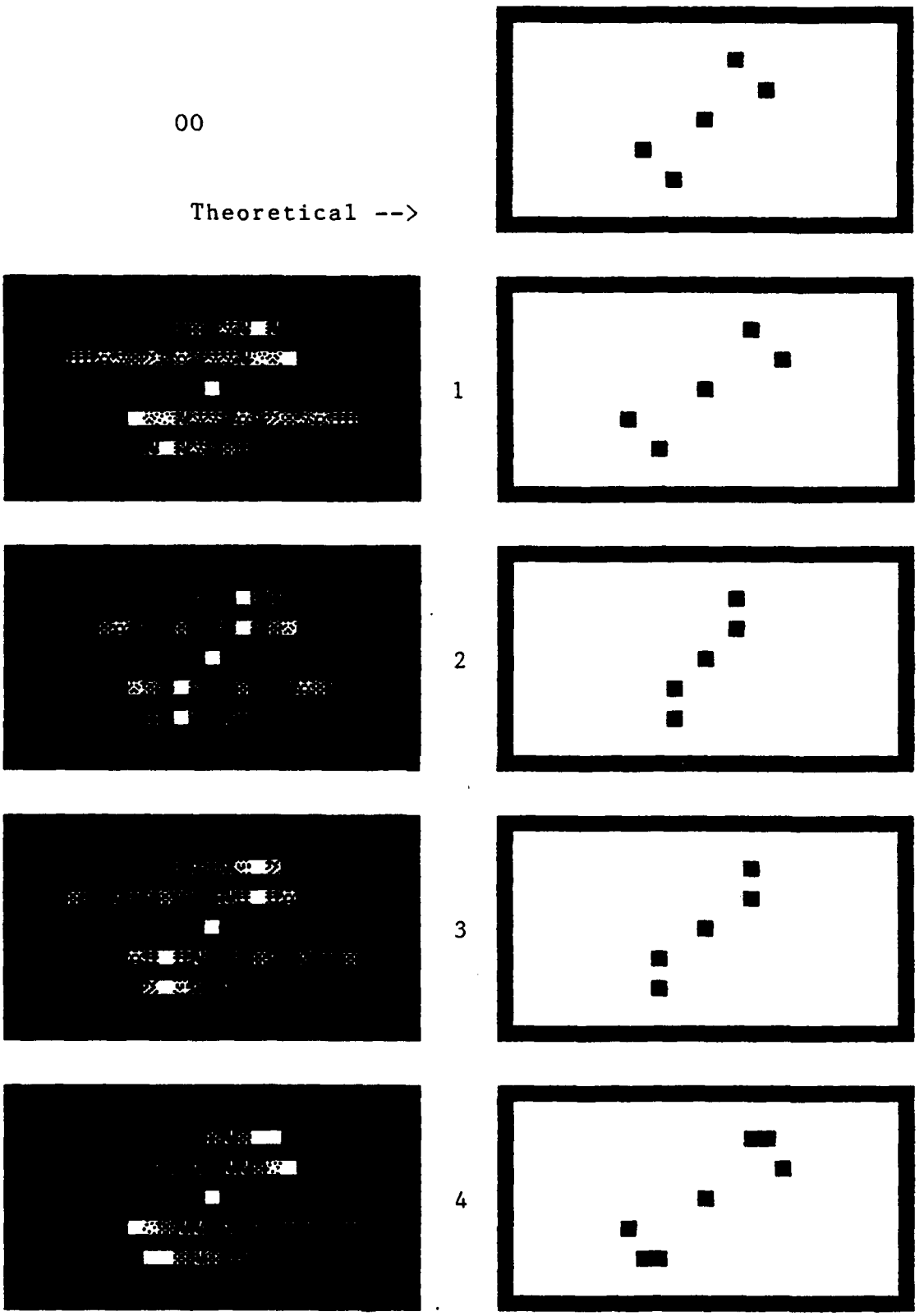


Figure 27. Trnasform Images for the Sound 00

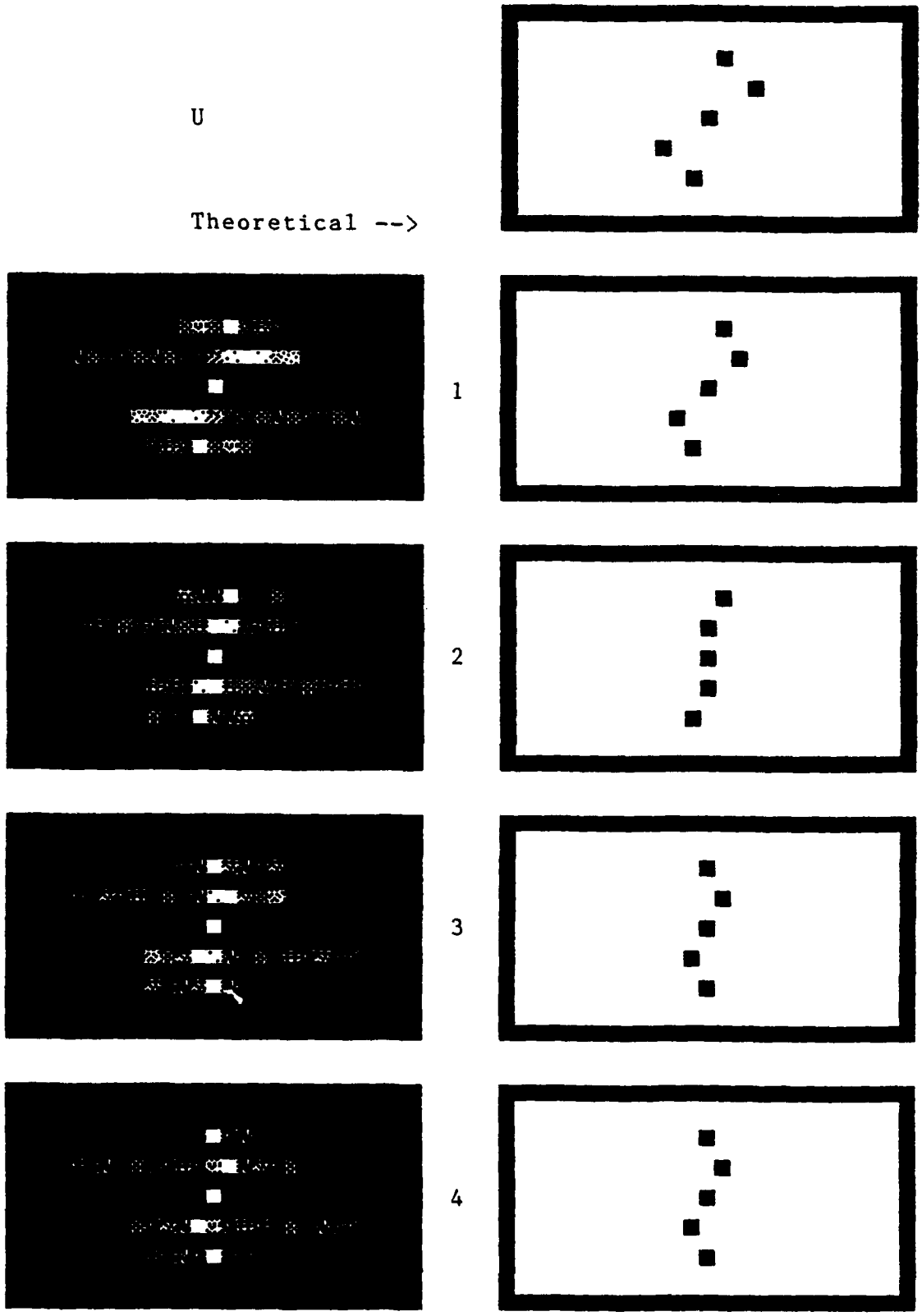


Figure 28. Transform Images for the Sound U

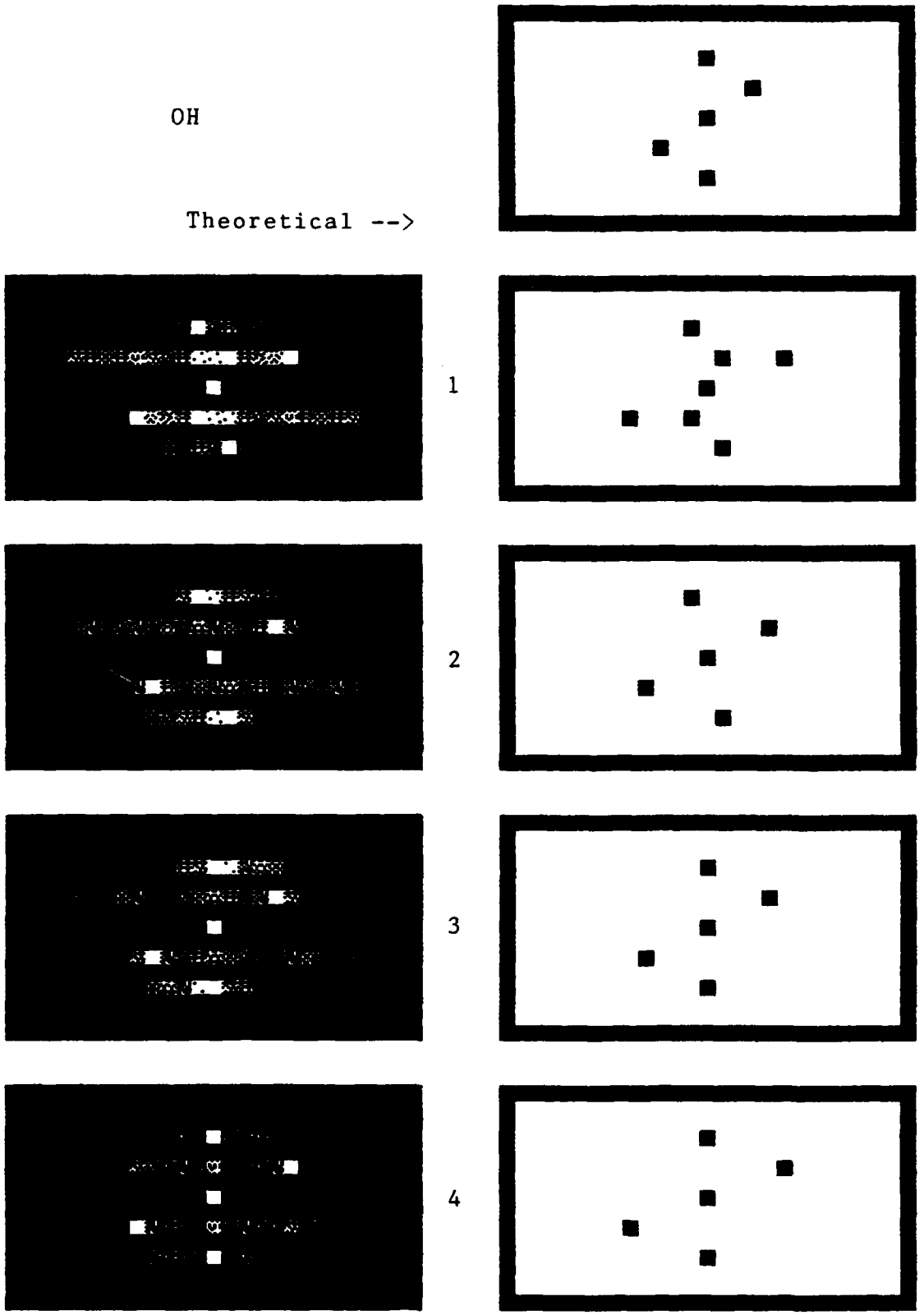


Figure 29. Transform Images for the Sound OH

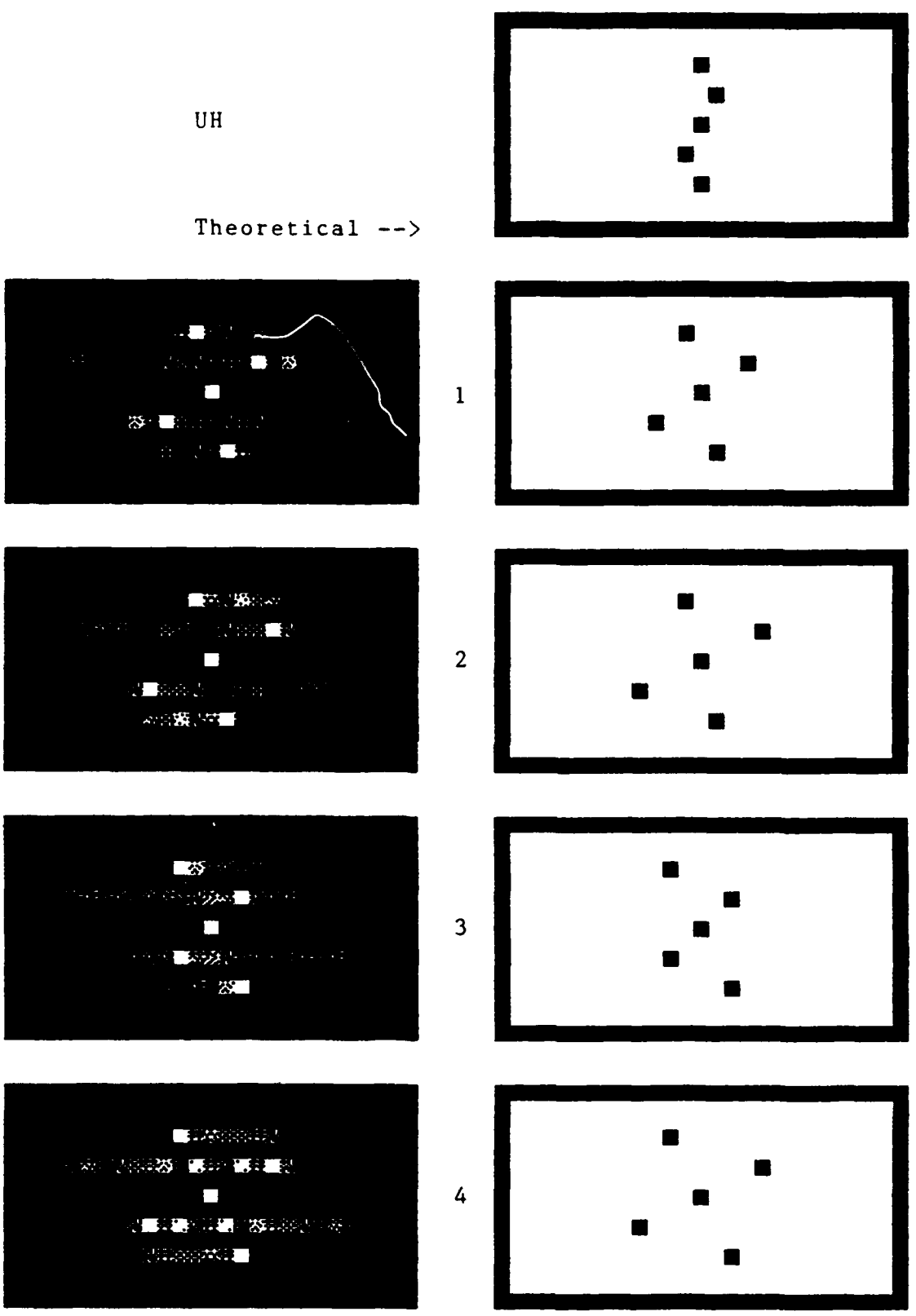


Figure 30. Transform Images for the Sound UH

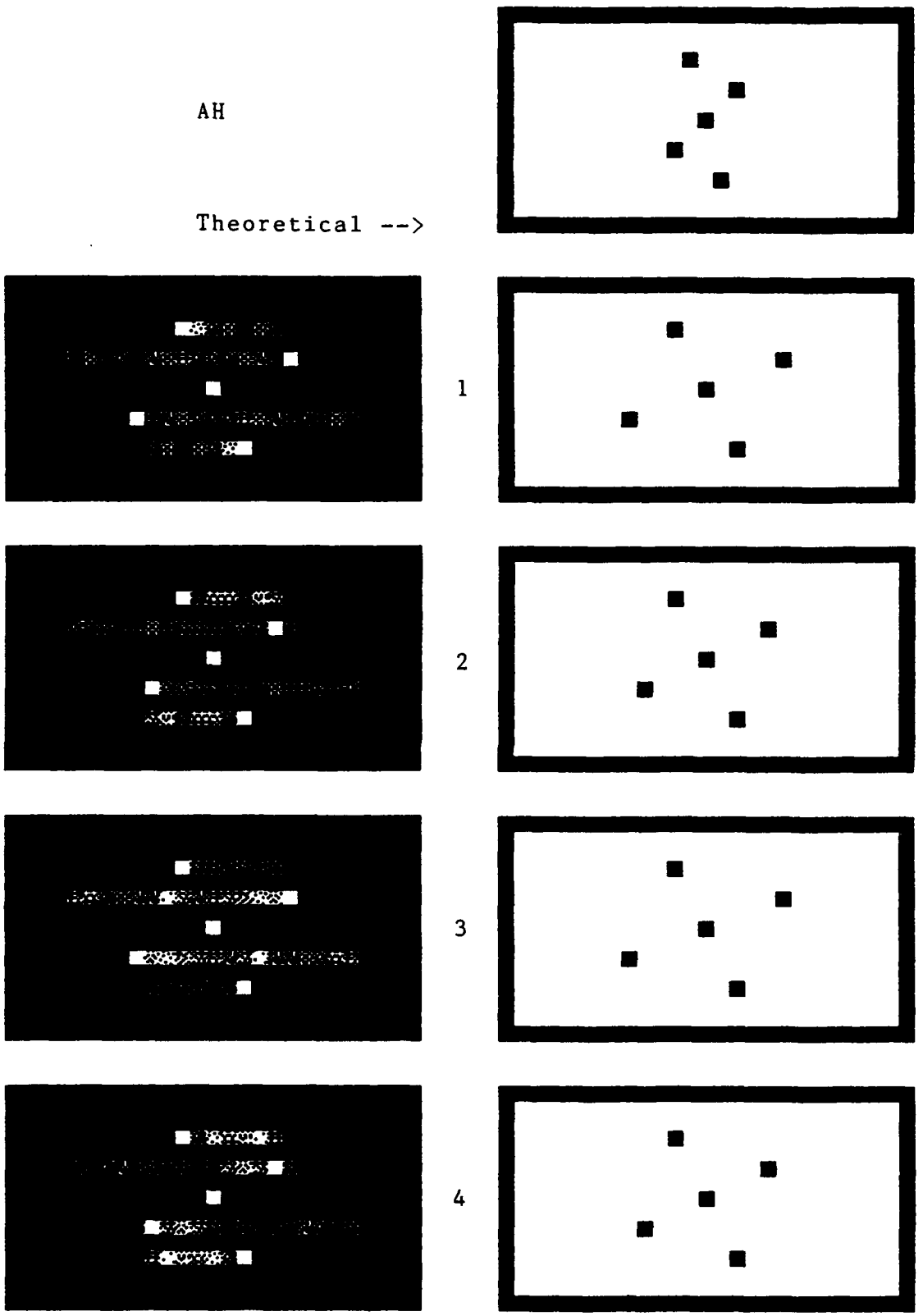


Figure 31. Transform Images for the Sound AH

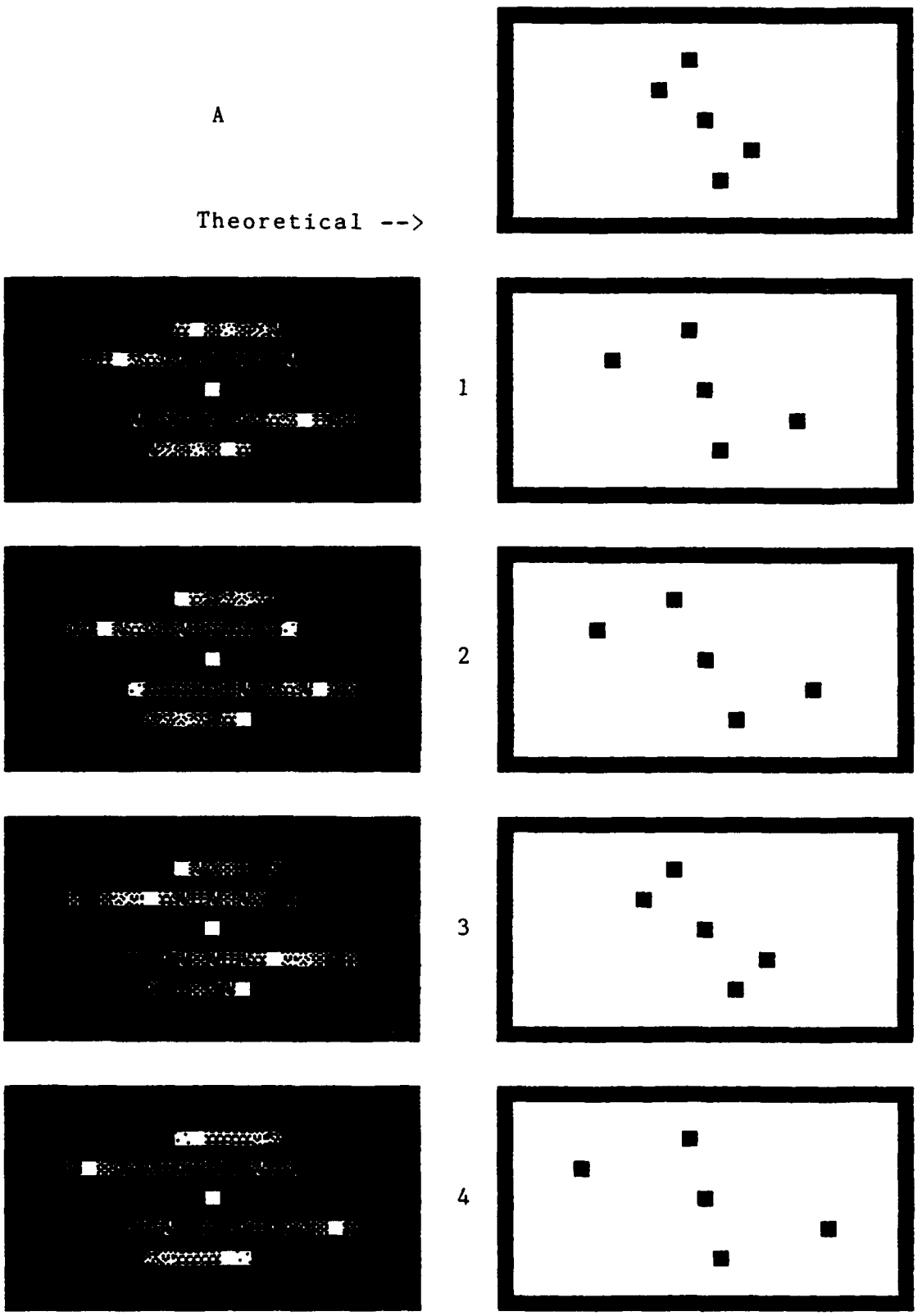


Figure 32. Transform Images for the Sound A

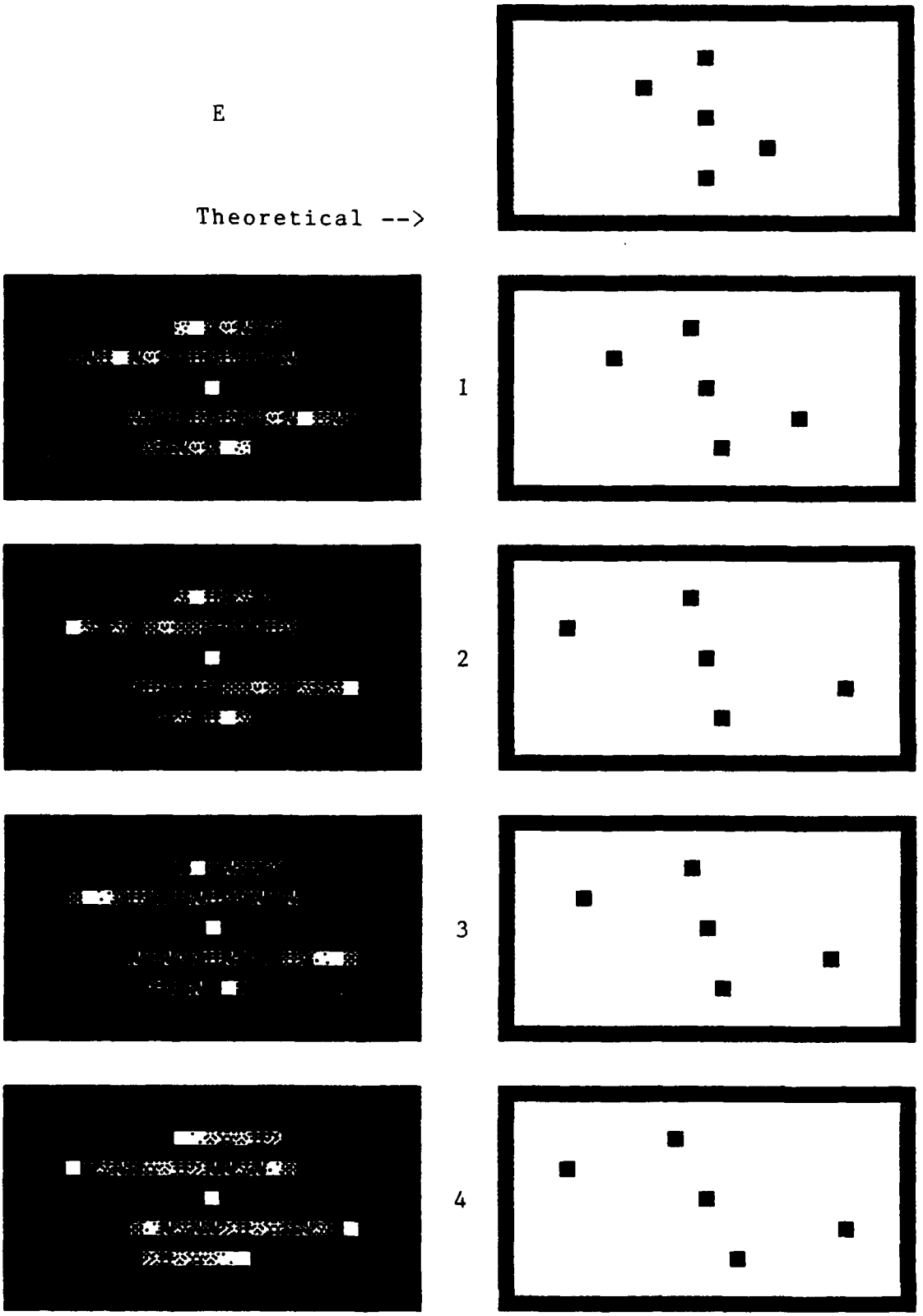


Figure 33. Transform Images for the Sound E

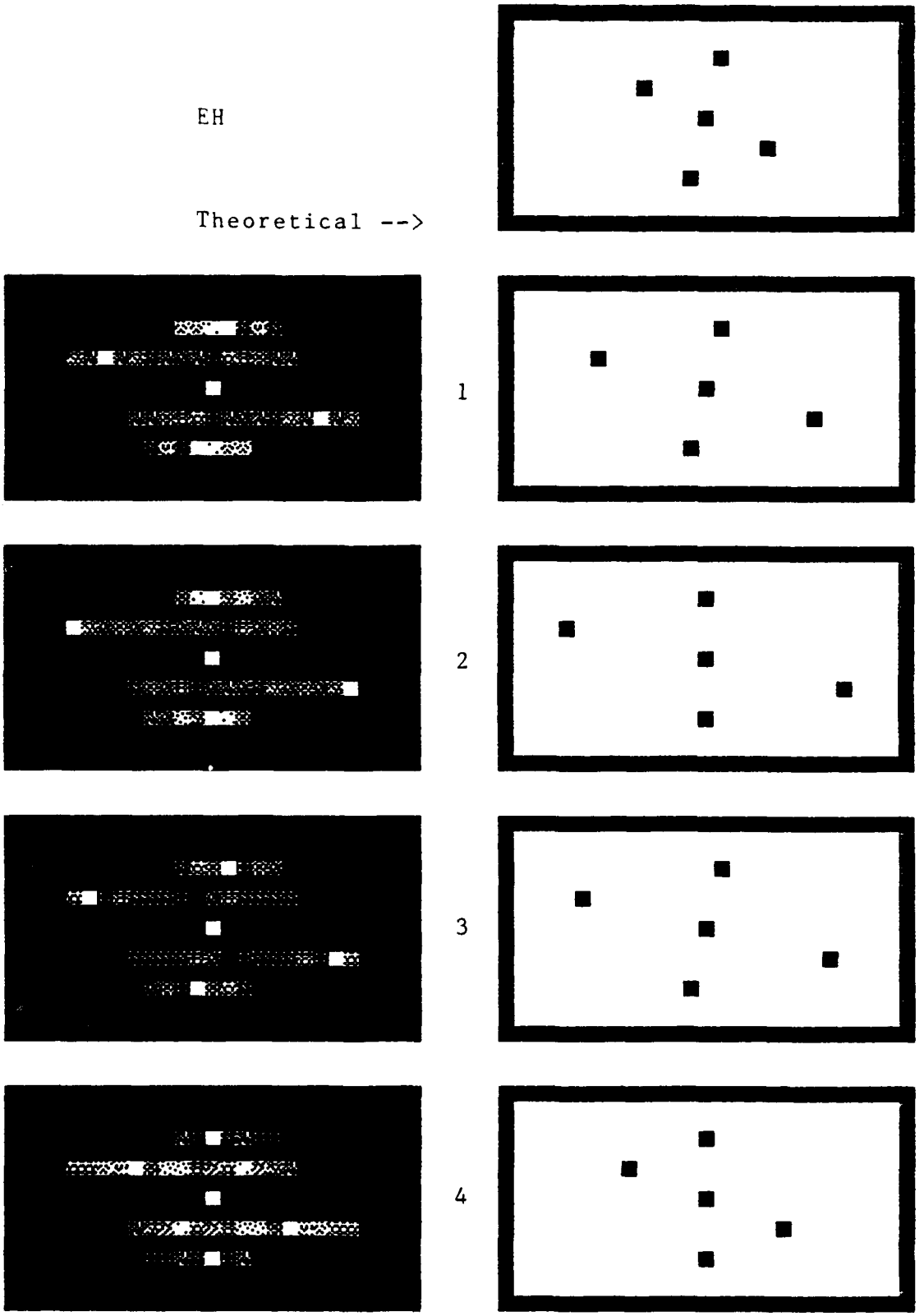


Figure 34. Transform Images for the Sound EH

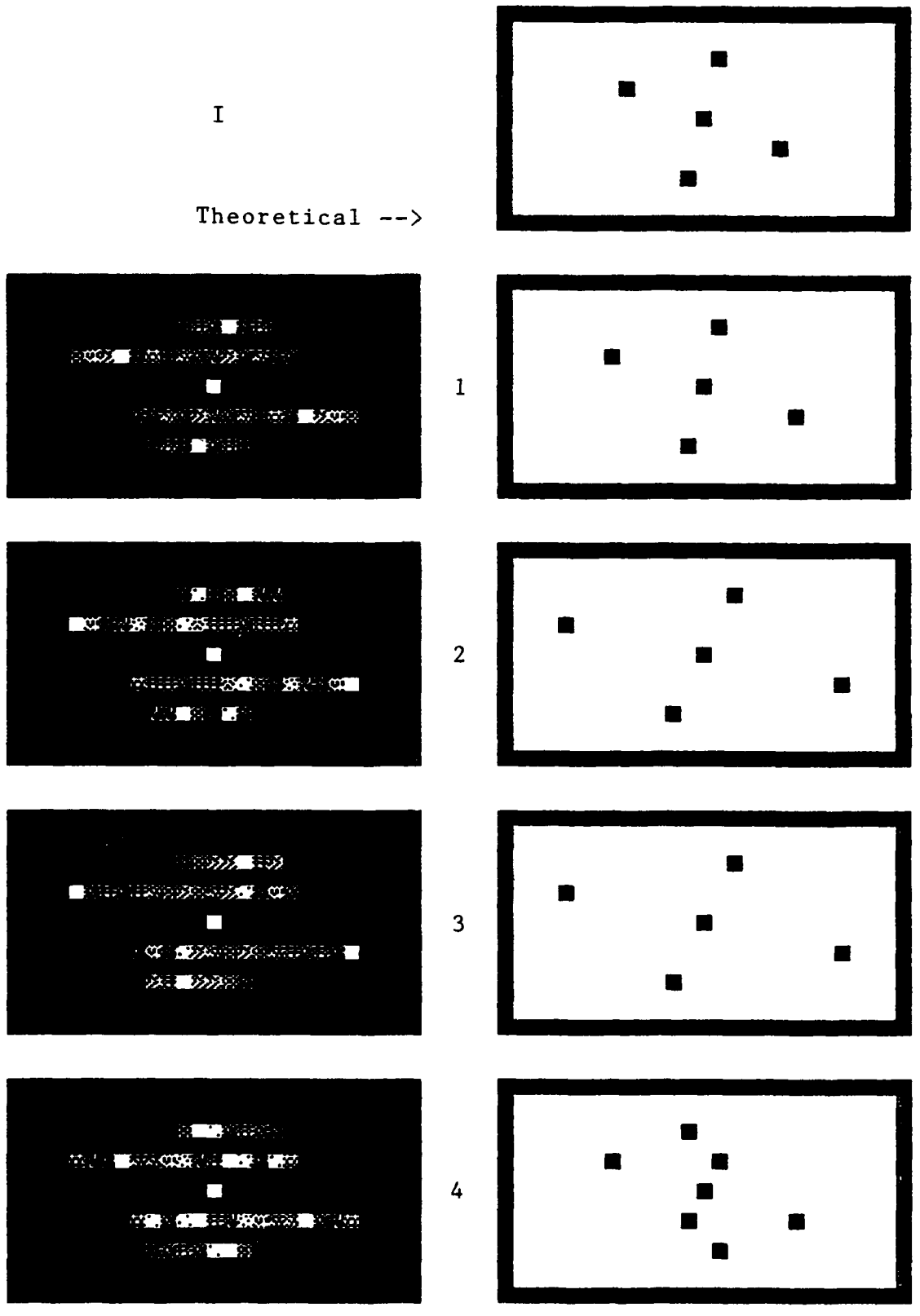


Figure 35. Transform Images for the Sound I

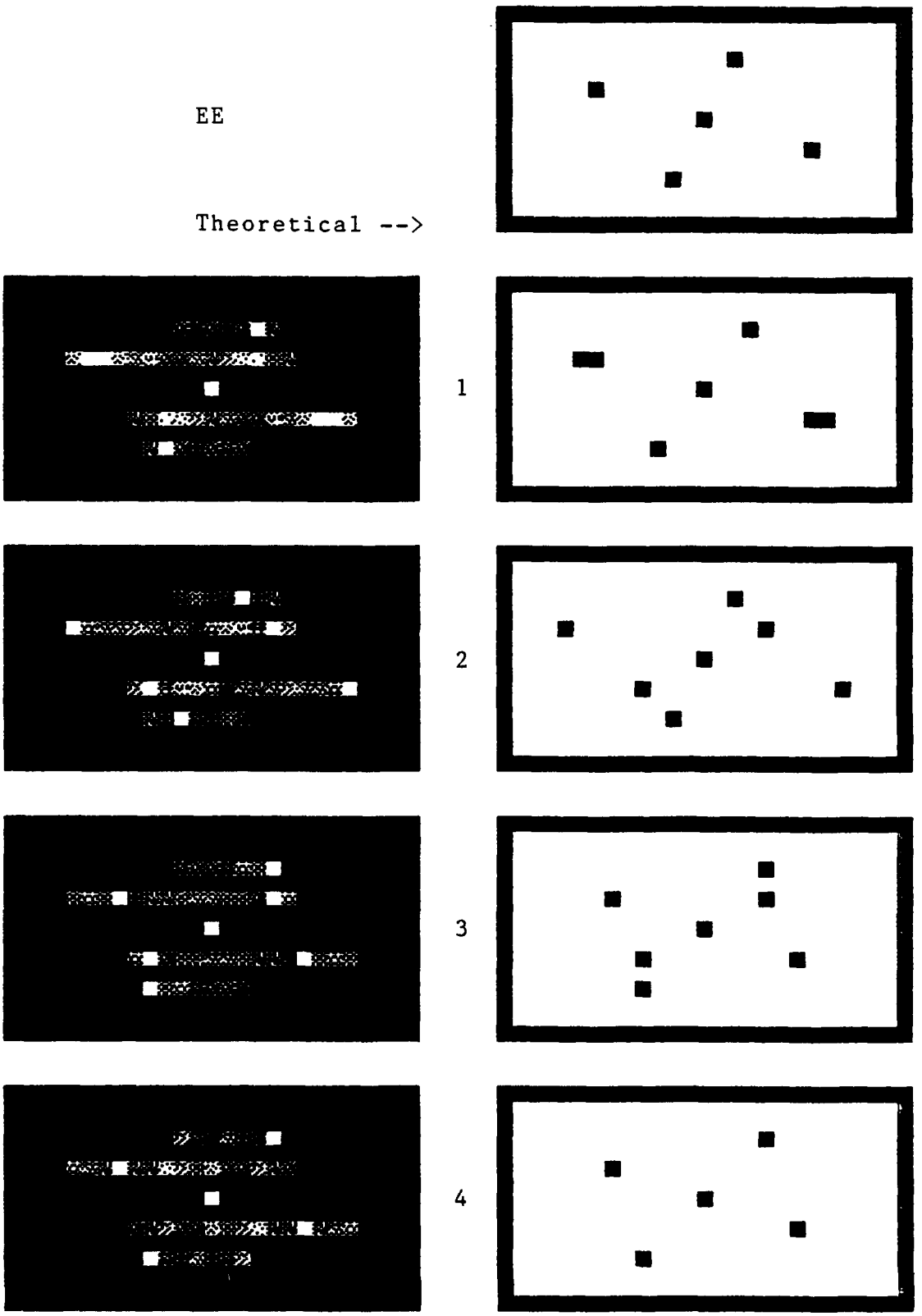


Figure 36. Transform Images for the Sound EE

Since only seven points are computed in band 1 and fifteen points in band 2, each point represents a discrete sample of the spectrum about 85 hertz apart for band 1 and 100 hertz apart for band 2. Any difference in the peak values between sounds less than these would not be resolvable in the present system. In an optical system, this resolution problem would not exist, since the optical transform is continuous.

Another undesireable characteristic of the transform images is that some bands have more than one pair of points which are at the "peak" level. This problem arises from the fact that the transform point values had to be compressed into the sixteen gray levels available on the Octek board. This may be overcome to some extent in an optical system; however, the dynamic range of the detectors may cause a similar problem dependinng on the type of detection scheme used.

Fricative and Stop Sounds. These sounds are shown in Figures 37 though 46. Each figure contains five plots. The top plot shows the expected "shape" for that sound, and is derived from observations using a spectrum analyzer. Fricative sounds are broadband in nature and it is difficult to predict any more than a general shape. The four lower plots are the experimental values, in order, of the four speakers (female at the bottom). Each plot shows the pixel intensity value (0 through 15) for pixels in the baseband as a function of their position with respect to the origin. Adjacent pixels represent samples of the frequency spectrum about 225 hertz apart.

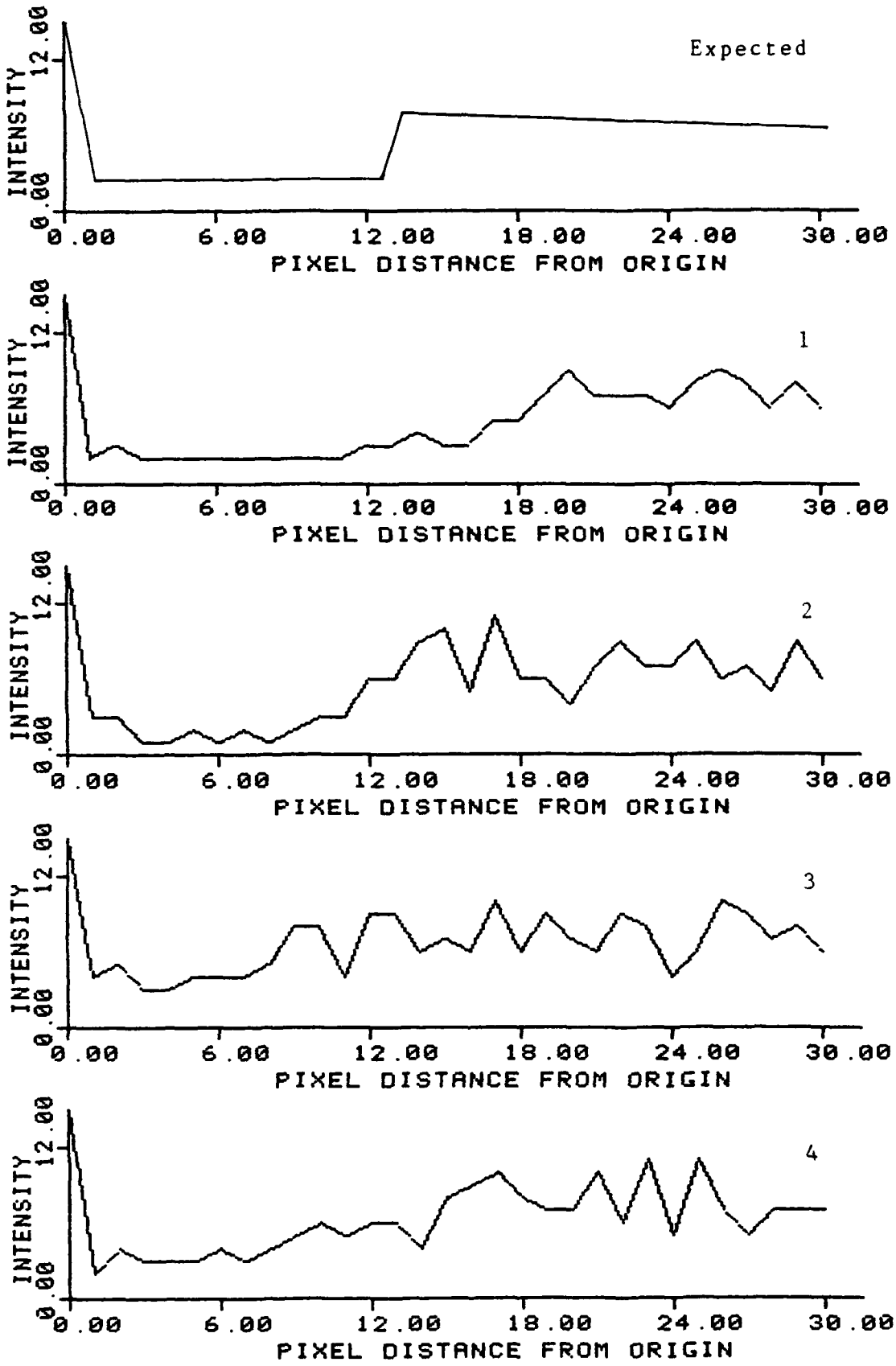


Figure 37. Baseband Pixel Intensity Plots for SH

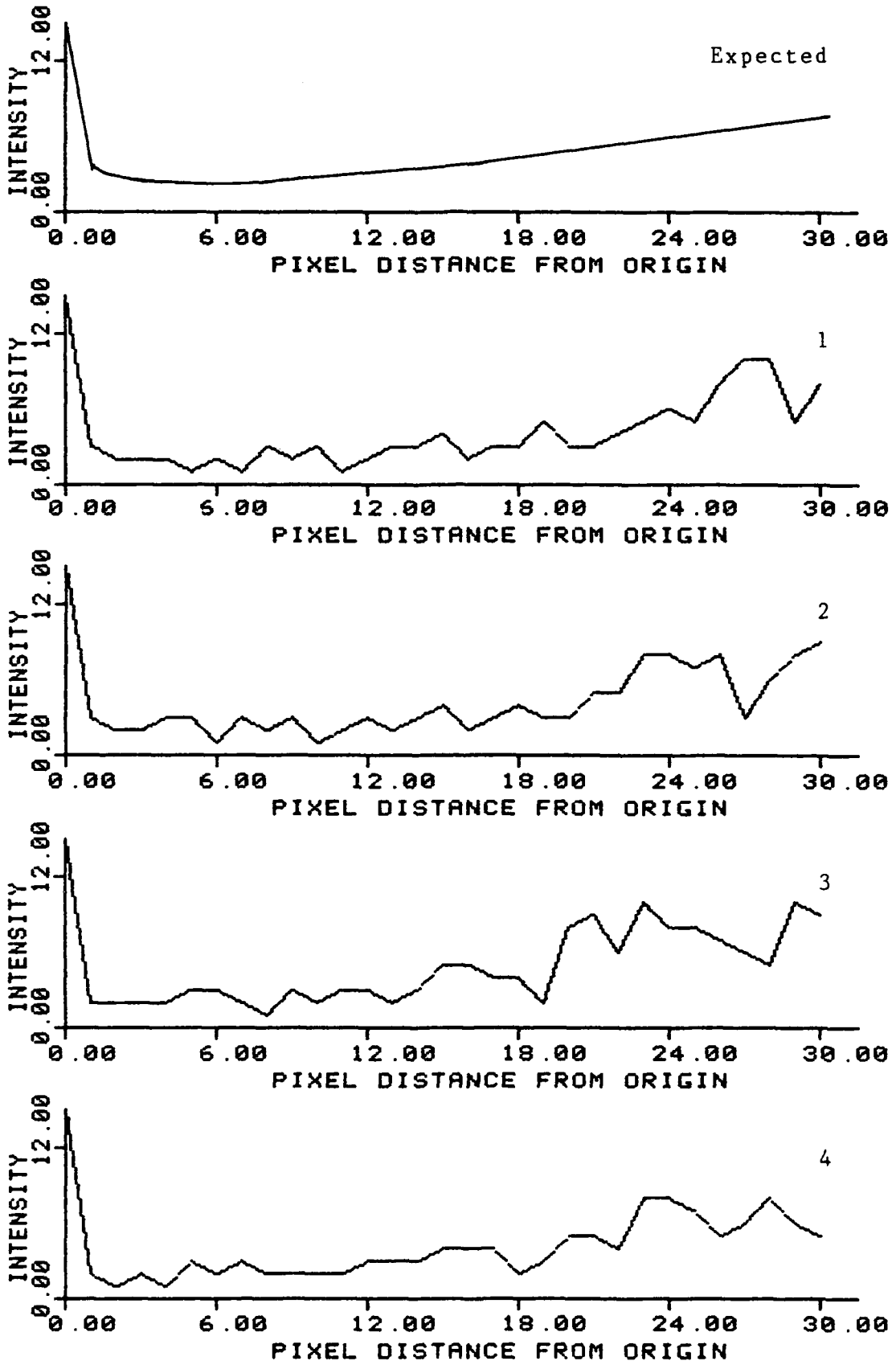


Figure 38. Baseband Pixel Intensity Plots for SS

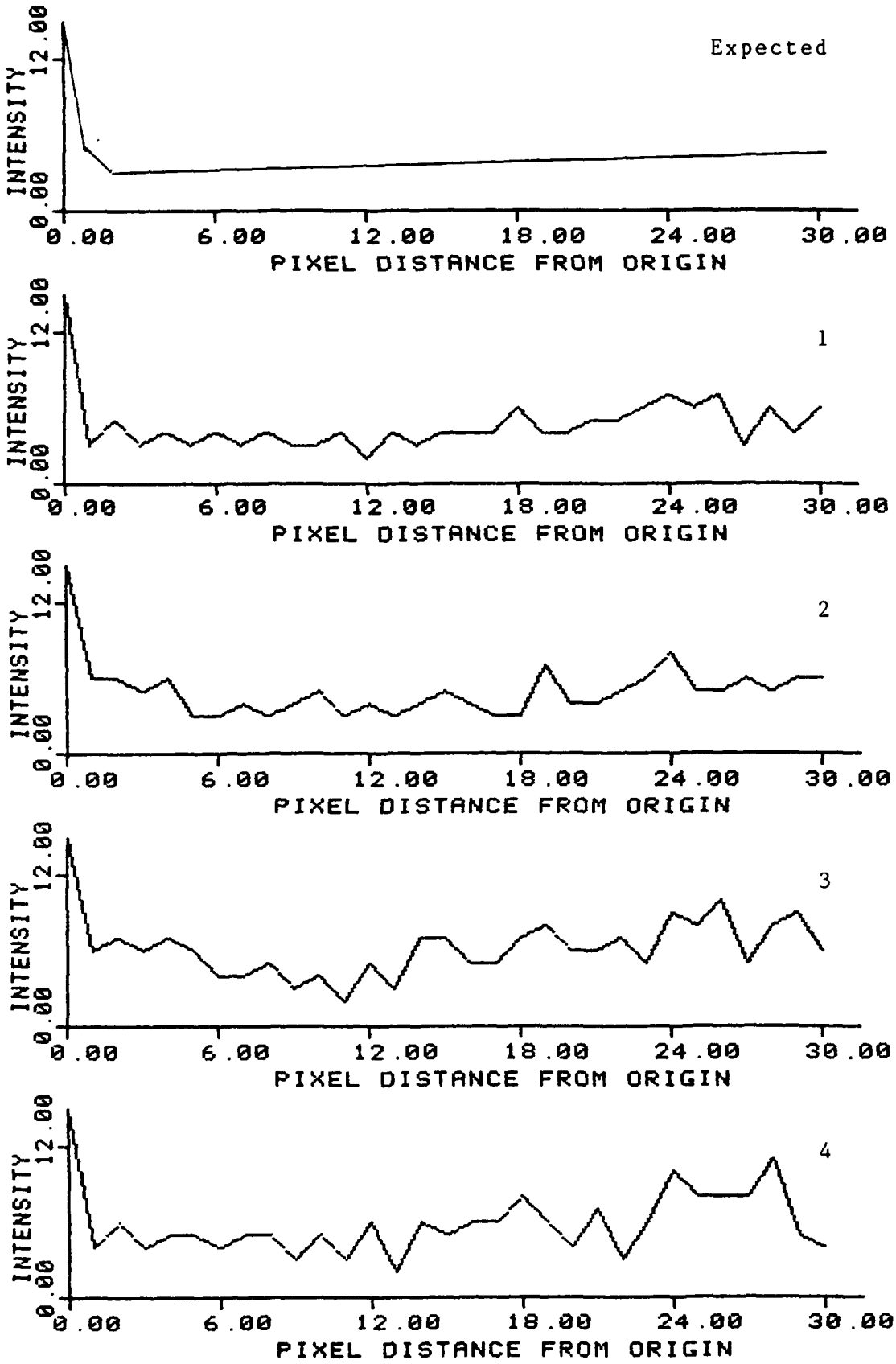


Figure 39. Baseband Pixel Intensity Plots for TH

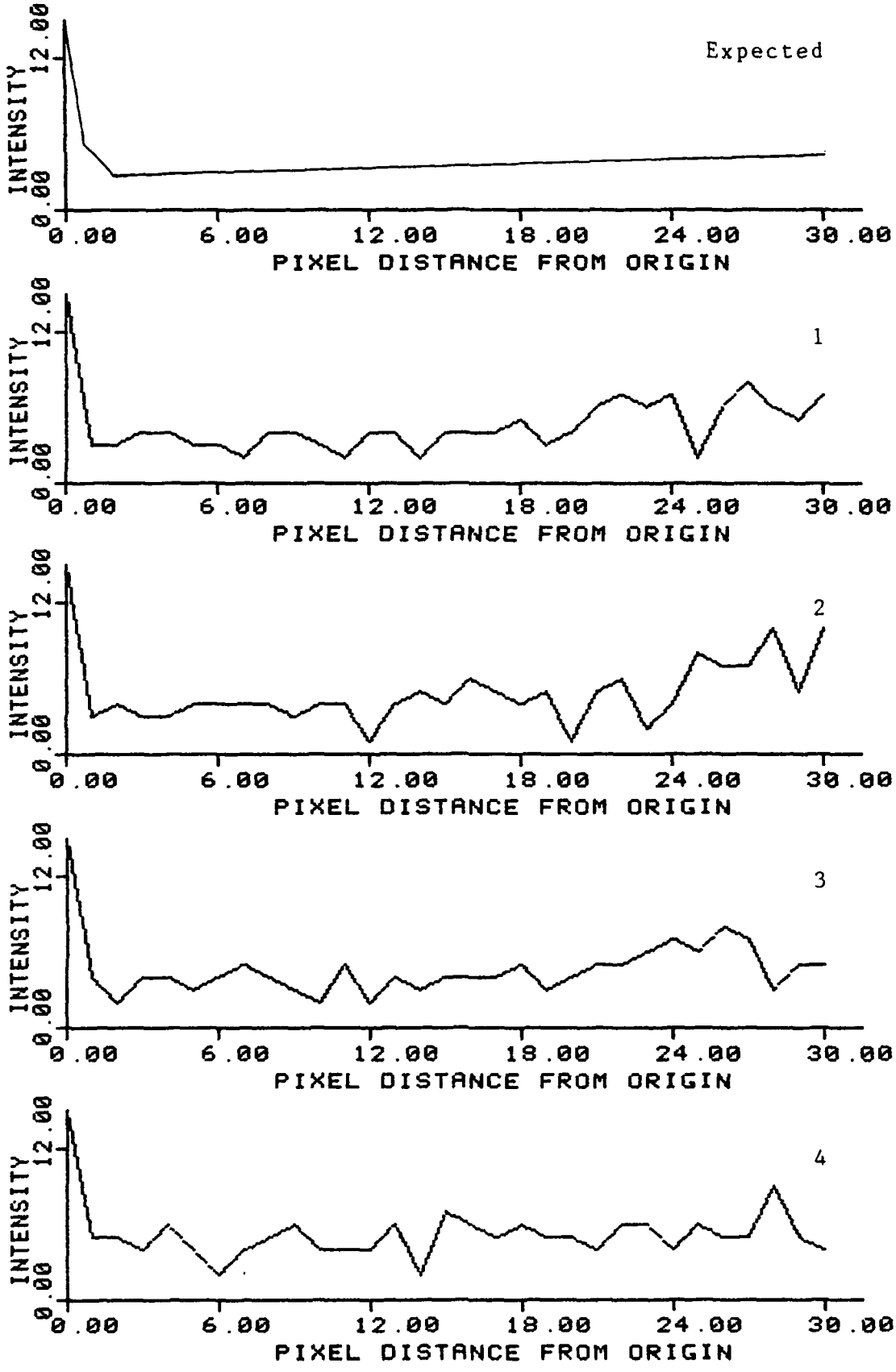


Figure 40. Baseband Pixel Intensity Plots for FF

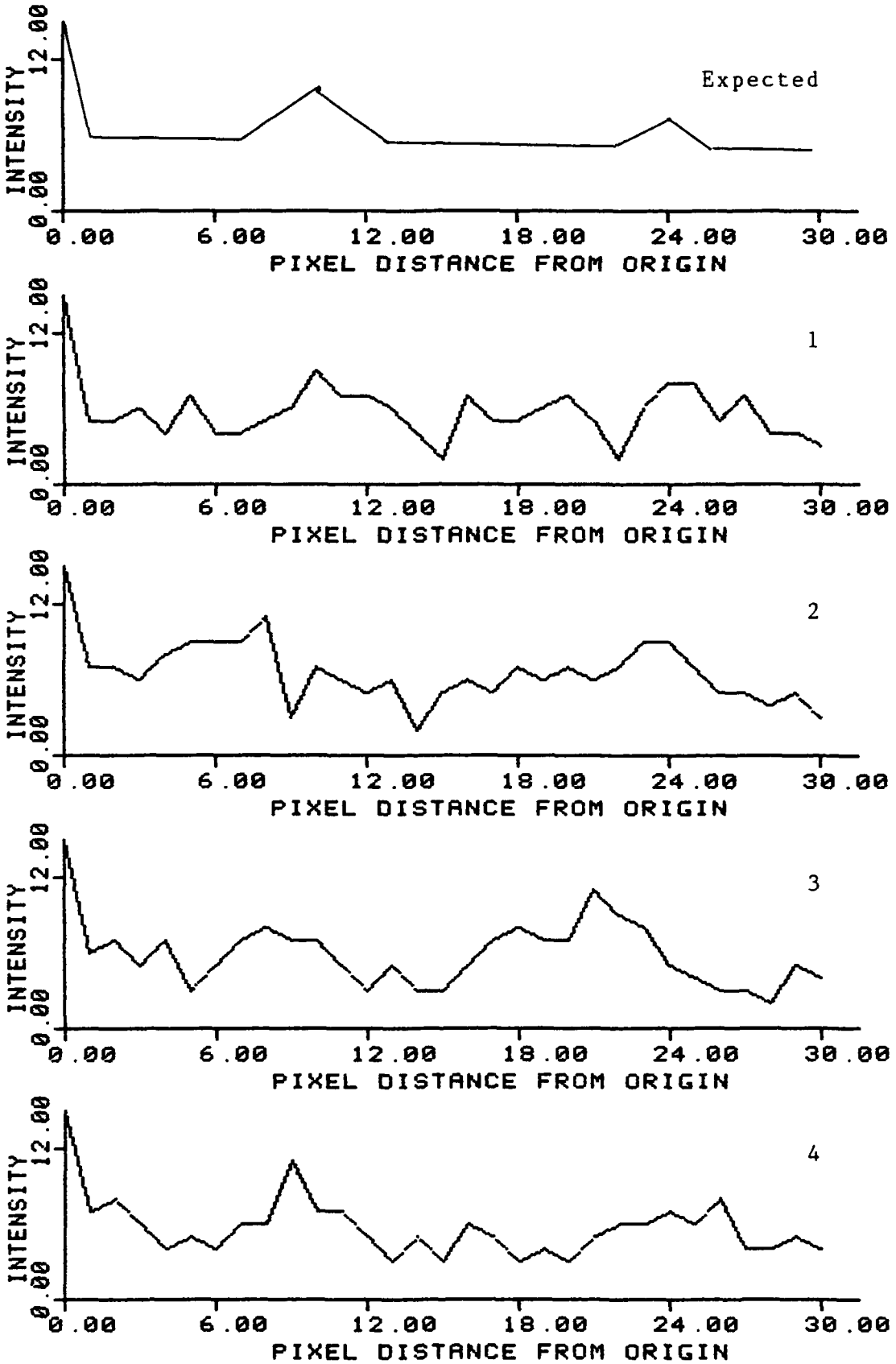


Figure 41. Baseband Pixel Intensity Plots for KK

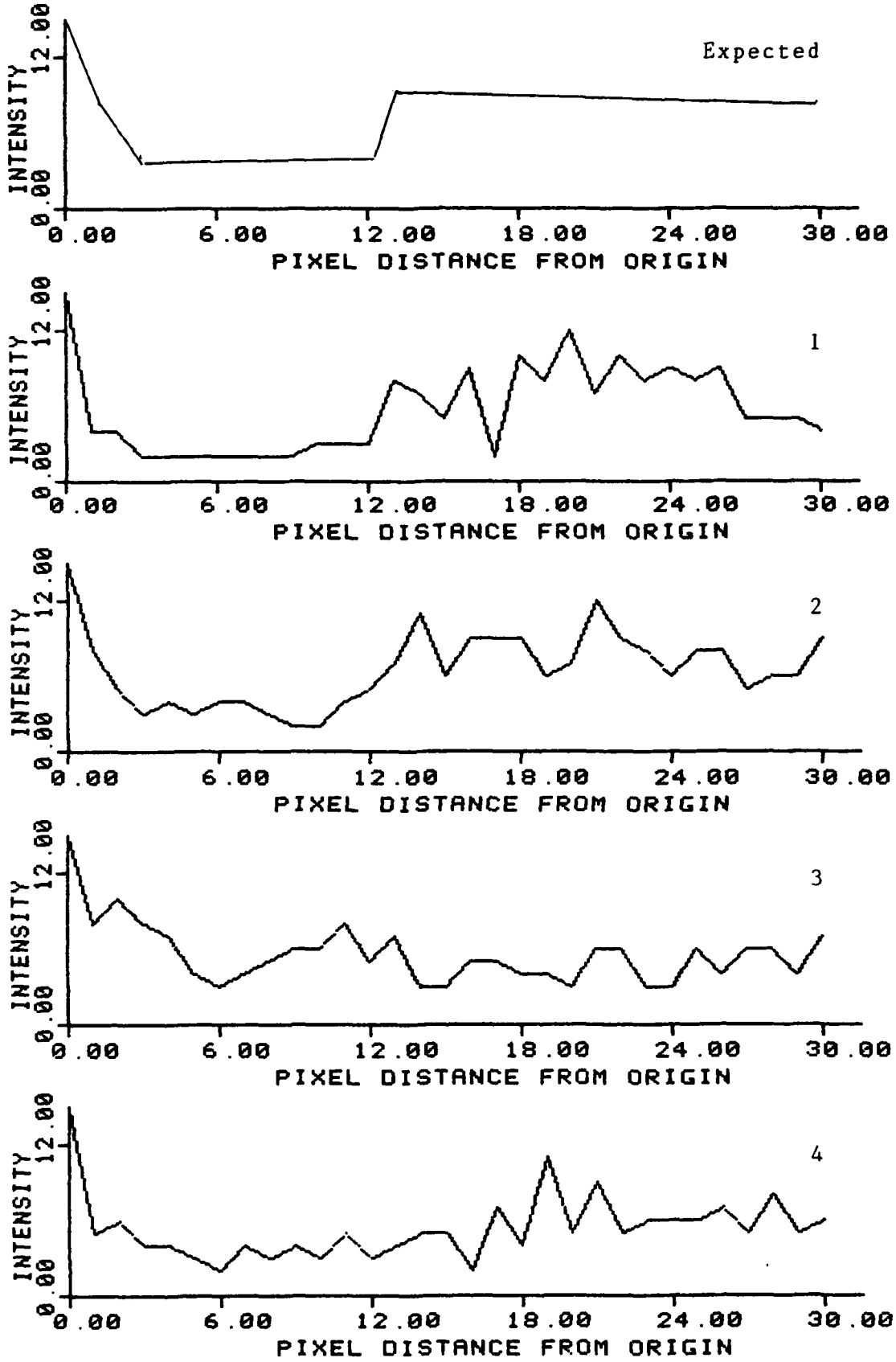


Figure 42. Baseband Pixel Intensity Plots for JJ

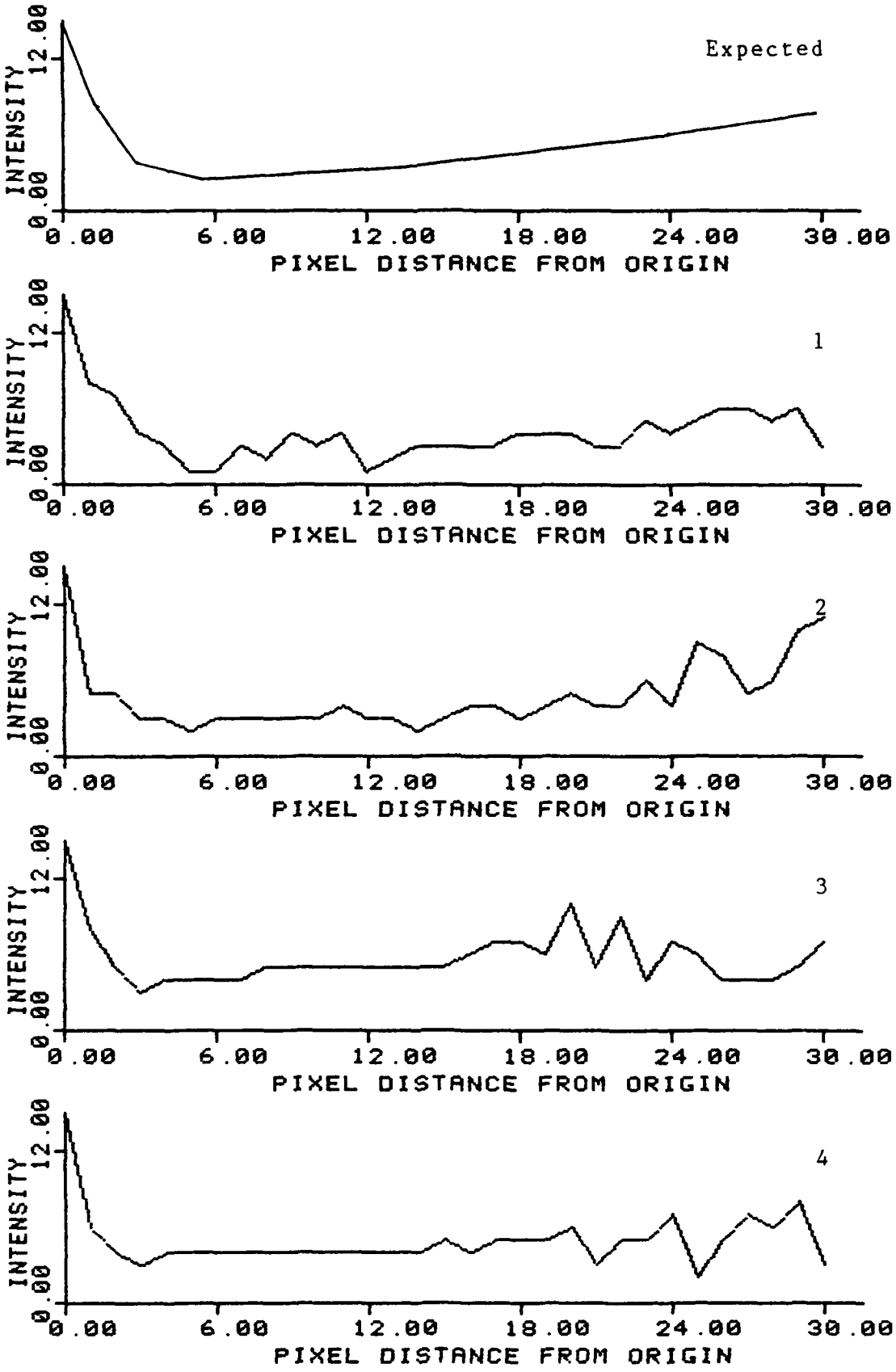


Figure 43. Baseband Pixel Intensity Plots for ZZ

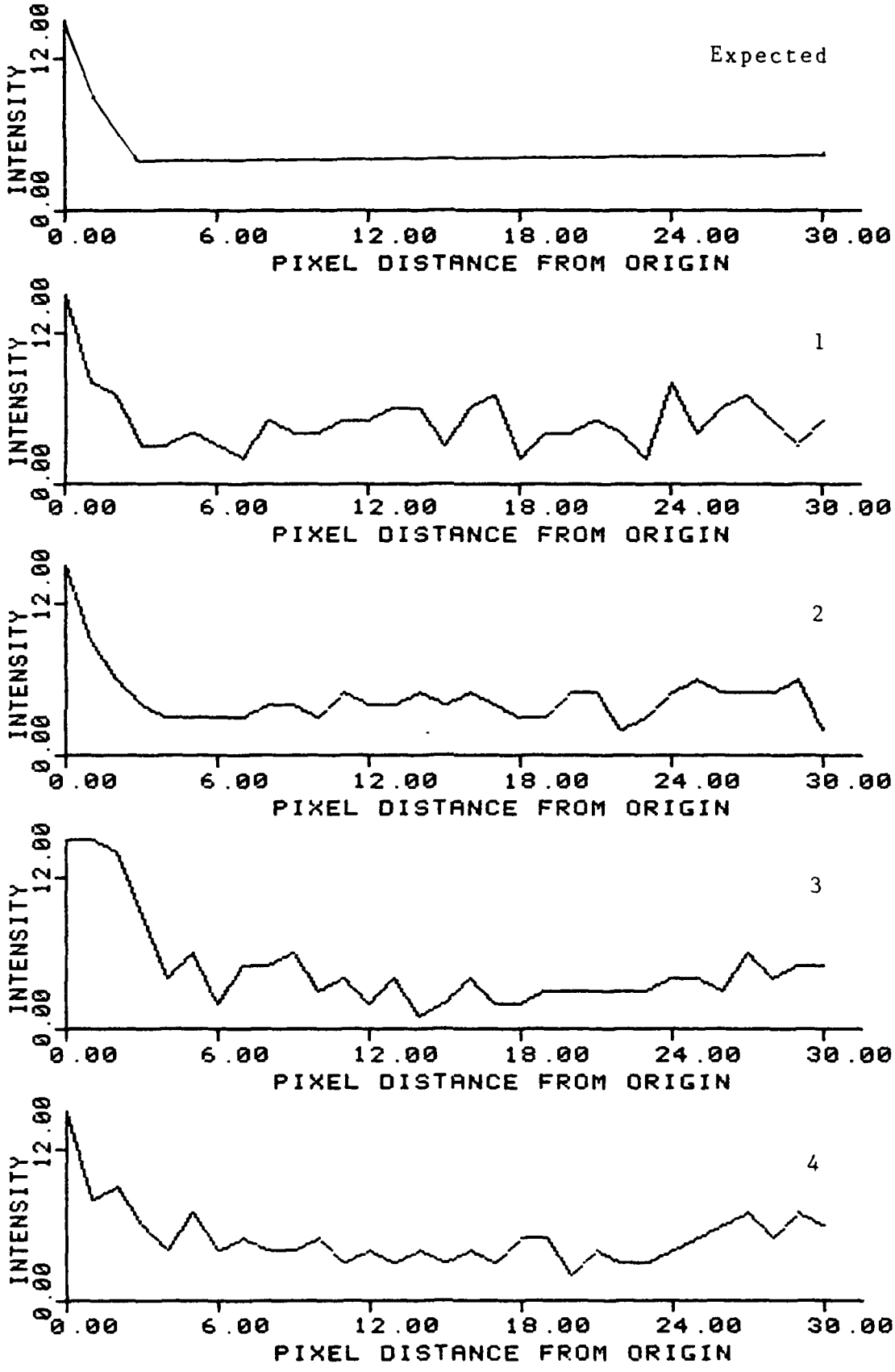


Figure 44. Baseband Pixel Intensity Plots for TT

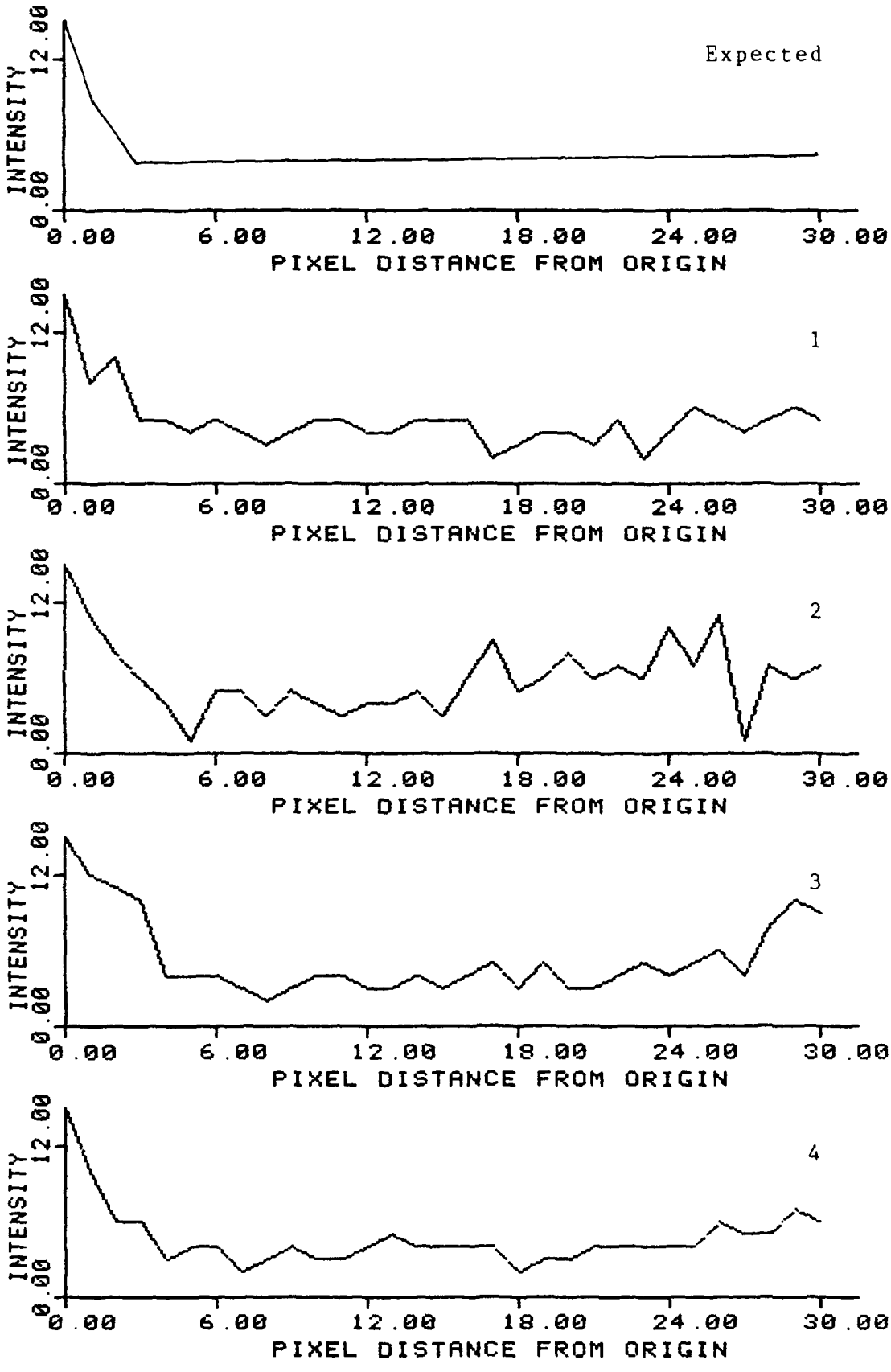


Figure 45. Baseband Pixel Intensity Plots for VV

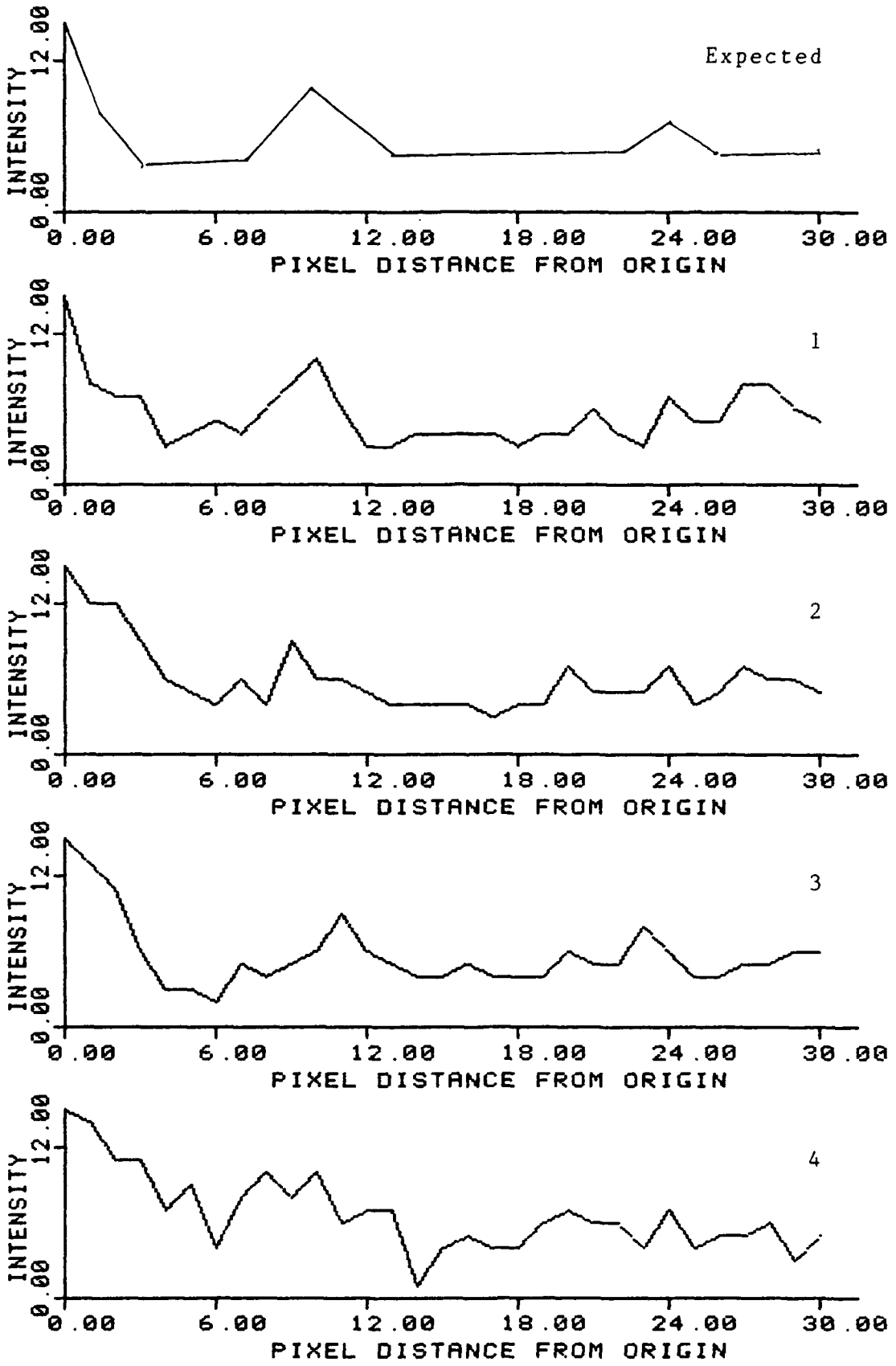


Figure 46. Baseband Pixel Intensity Plots for GG

The voiced sounds (JJ, ZZ, TT, VV, and GG) are easily distinguished from the unvoiced sounds (SH, SS, TH, FF, and KK) by observing the pixel closest to the origin. Voiced sounds consistently show higher values in this pixel indicating the presence of a glotal pitch. The sounds SH, SS, and KK and their voiced counterparts JJ, ZZ, and GG all show distinctive qualities and are fairly easily separated. The sounds TH, FF and their voiced counterparts TT and VV prove to be quite similar and it is almost impossible to distinguish between them.

Dynamic Data

One set of transform images for the words "zero" through "nine" was reassembled on video tape to give a real time picture. From this one set it was evident that the transform images changed too fast to be detected by the unaided eye. Because of the difficulty in assembling single frames, and the evidence that real time speeds were too fast for easy observation, three sets of word transforms were assembled at 1/4 speed (four frames for each image).

The 1/4 speed data also proves to be too fast for sounds to be recognized by simply viewing the images. When the moving pictures are slowed down and viewed one frame at a time, various sounds can be identified; for example, the "EE" and "OH" in "zero" can be seen as well as the dipthong transition from "AH" to "EE" in "five". The stops in "six" and "eight" are clearly distinguished by a series of dark images indicating a drop in power.

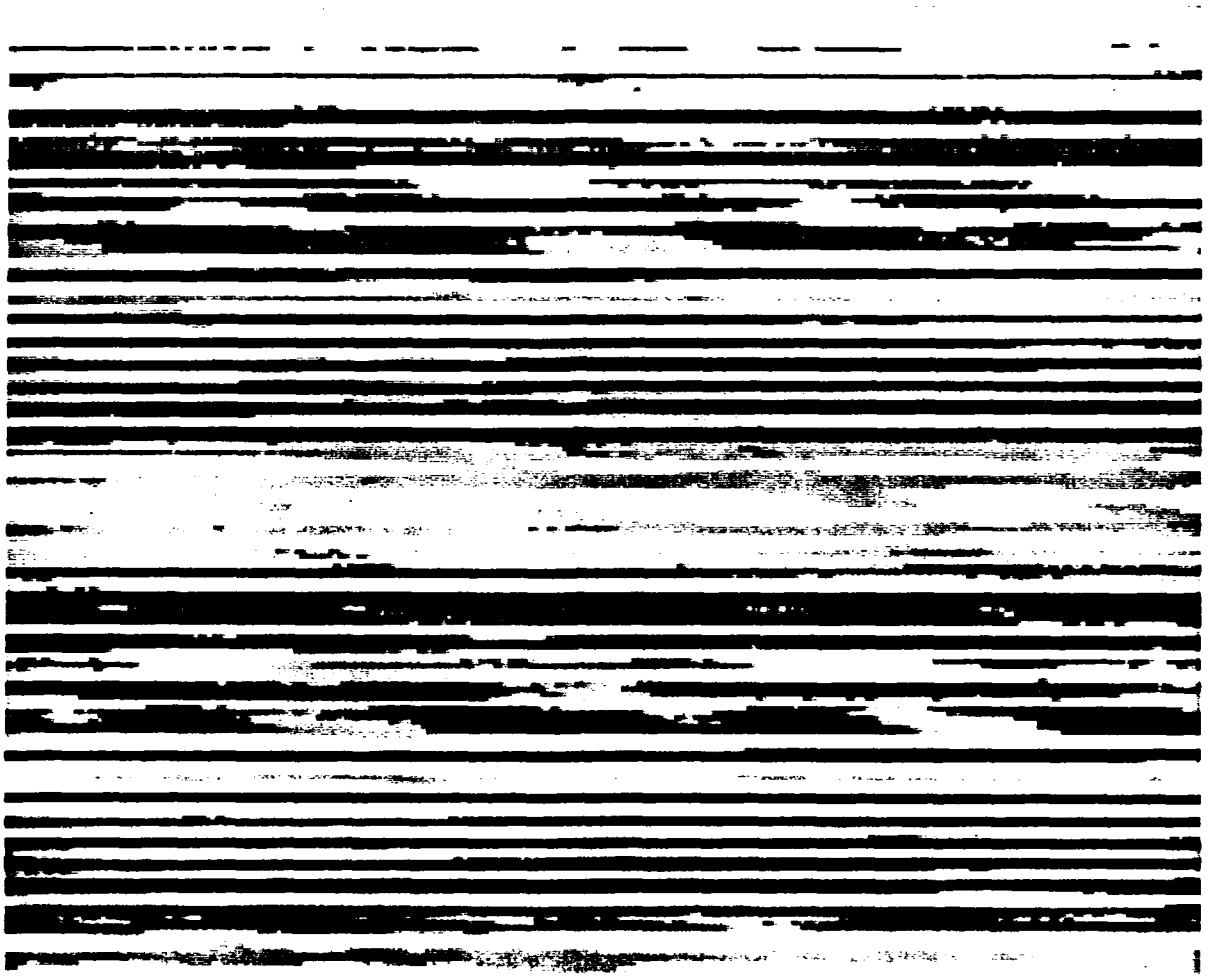


Figure 47. Time Domain Image of "EE" in "zero"

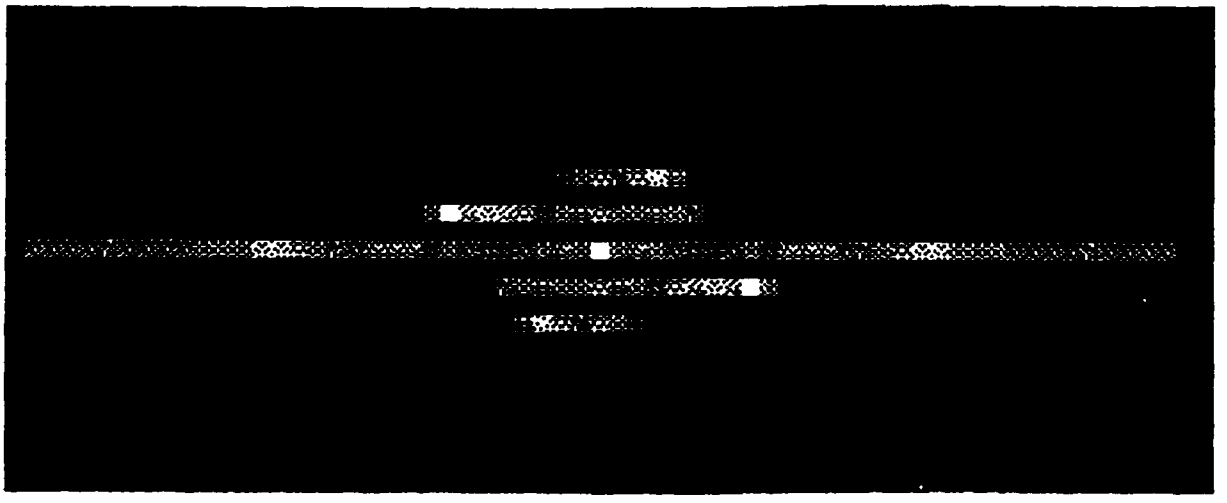


Figure 48. Transform Image of "EE" in"zero"

Similar to the static data, some images have more than one set of peaks in each band, and it is often difficult to decide which is dominant. It appears that more contrast in the images would help make peak locations easier to pick out and follow.

Data indicates that the desired information is contained in the transform images if a fast enough method of extracting it can be found. Since the display changes only once every thirtieth of a second, a large number of points could be sampled for each frame, making machine recognition a definite possibility.

VI. Conslusions and Recommendations

A New Tool for Speech Analysis

A system has been built which displays speech signals as a two dimensional picture in standard video format using a common video display. When acted on by a simulated optical Fourier transform, the resulting image portrayed a broadband look at the spectral content of the speech signal as well as a more detailed look at two smaller portions of the spectrum. Spatial filtering has allowed unwanted information to be removed from the image.

Perhaps the greatest advantage of this system was its ability to display in detail a small portion of the frequency spectrum of the speech signal, as demonstrated by the two modulated bands covering the frequency spread of the first and second formants. The use of modulation allowed portions of the spectrum to be expanded spatially in the transform image, giving greater resolution in the area of interest.

The demonstation of multiband capability is important since it indicated that many bands may be simultaneously processed. By using construction techniques better suited for high frequencies, the maximum number of modulated bands possible would be limited only by the bandwidth of the video display.

The general success of the project has shown the feasibility of using optical techniques in the analysis of speech signals. The use of such techniques offers the advantages of

operation at real time speeds and a resolution capabilty which is ultimately limited only by the optical devices themselves. Although real time images move too fast to be identified by the unaided eye, the data indicates that the transform images could be detected by machine and used as an input to a subsequent phoneme discriminator.

Recommendations for Further Development

This first look at optical processing techniques has pointed out a great number of areas in which further development is needed. First of all, this project needs to be repeated using the actual optical devices in order to verify the results found in this work.

Once a true optical system is in operation, the method of interfacing the output of the system to a subsequent phoneme discriminator must be developed. Such an interface would inevitably involve the use of detectors to convert light back to an electrical signal, which would subsequently be converted to digital form.

The system used for this project included two modulated bands used to display the frequency ranges of the first two formant frequencies of the voice. This is not necessarily the optimal use of the modulated bands, and certainly not the maximum number of bands possible. Further study is needed to optimize the placement and number of modulated bands in the system to provide the most useful display of frequency information in the transform image.

With greater resolution possible and the ability to be more selective about which parts of the spectrum are examined, the possibility of a speaker independant recognition system becomes more promising. Prior to achieving this, however, a substantial database of transform images must be collected which includes all the sounds of speech (at least those used in English), as spoken by large numbers of people. Only then can decisions be made about what traits separate two sounds or make them the same.

Appendix A: List of Equipment

1. Microphone, Shure model SM54
2. Audio Amplifier, Digital Sound Corp. model 240
3. Cassette Recorder, Tascam model 122
4. Waveform Generator, Wavetek model 148
5. Video Broadcast Generator, Telemation model TSG-3000GL
6. Video Cassette Recorder, RCA model VKP-900
7. Video Monitor, Electrohome model EVM 1710R
8. Video Digitizer, Octek model 2000
9. Digital Computers, Data General Corp.
    a. Nova 2
    b. Eclipse S/250
10. Regulated Power Supply, Hewlett-Packard model 6236B
11. Spectrum Analyzer, Hewlet Packard model 3580A
12. Video Tape Recorder. Sony model VO-5850
13. Automatic Editing Control Unit, Sony model RM-440
14. Waveform Analyzer, various
15. Oscilloscope, various
16. Frequency Counter, various

Appendix B: Equipment Set-up and Adjustment

The circuitry requires an initial alignment and should be checked periodically thereafter. A warm up period is recommended to allow oscillators to stabalize. All voltage values are peak to peak.

Pre-processing Circuitry

There are three controls which must be adjusted in the pre-processing circuitry. These are the balance control and the attenuation for the outputs to the direct (baseband) and modulation circuitry. These are initially set, but may require re-adjustment later.

Step 1. A 400 hz, 1.0 volt peak to peak signal is connected to the "Audio" input of the circuitry. Observing the output at pinn 6 of the buffer amplifier, the "Balance" control is adjusted to give zero D.C. offset in the output.

Step 2. The "Direct" output is set for .5 volts. The "Modulator" output is set for .6 volts.

Modulation Circuitry

Repeat all three steps for both modulation circuits.

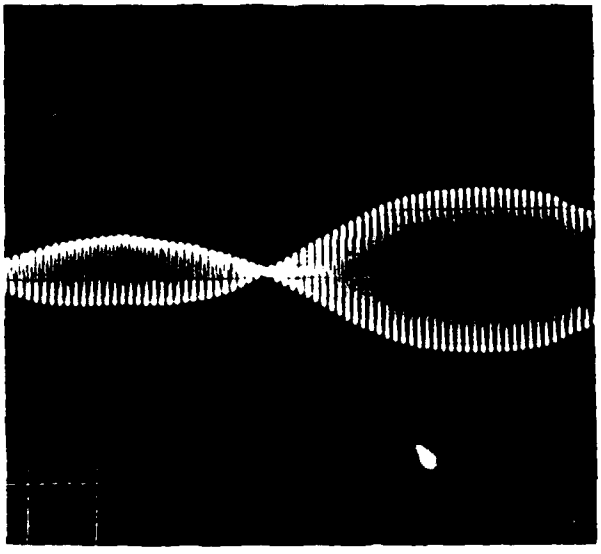
Step 1. Wavetek generators used to provide carrier inputs are set up for sine wave output, .3 volts amplitude. Output frequency is checked using a frequency counter. Tolerence is plus or minus 5 hz.

Step 2. While observing the waveform at pin 6 of the modulator I.C., the "Offset" control is adjusted to

lobe is probably greater in amplitude, as shown in Figure 49.

Step 3. Adjust the "Carrier Null" to give all lobes equal amplitude, as shown in Figure 50.

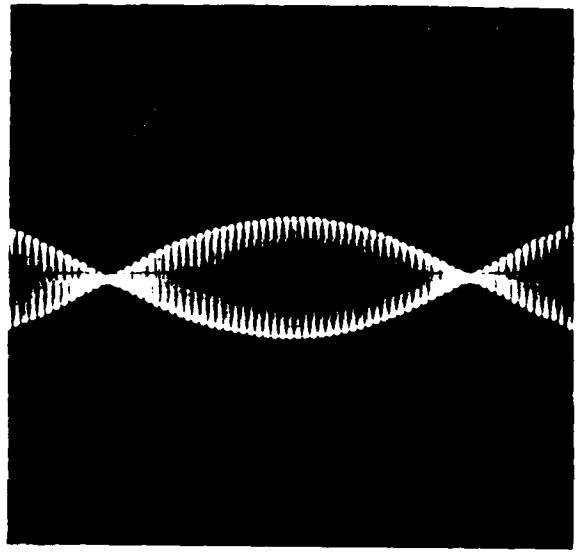
Step 4. The output of the modulator is attenuated before being sent to the video mixing circuitry. Adjust the amplitude at the output to .5 volts.



Vertical - .2 volts/div.

Horizontal - .5 milisec/div.

Figure 49. Modulator Output with Improper Balance



Vertical - .2 volts/div.

Horizontal - .5 millisec/div.

Figure 50. Modulator Output with Proper Balance

Video Mixing Circuitry

Step 1. Connect a 400 hz, 1 volt sine wave to the "Audio" input and select the baseband only using the band select switches. Observing the "Video" output, the 733 negative supply voltage is adjusted so that the bottom of the waveform is at zero volts D.C. The negative supply voltage value should be about -6 volts.

Step 2. The "Sync Level" should be adjusted so the sync level of the video output is +.5 volts.

Step 3. The "Brightness" control is adjusted so the most negative part of the 400 hz sine wave envelope is just above the sync level. This controls the "no-signal" grey level of the output.

Step 4. Disconnect the 400 hz sine wave generator and connect the cassette recorder or microphone and amplifier to the "Audio" input. Select band 1 only using the band select switches. Observing the noise level for no input signal, adjust the "Balance" control in the pre-processing circuitry to minimize this. Repeat for band 2 and then all three bands combined. There is a specific point at which the noise will drop to a minimum.

The relative amplitudes of the baseband and the two modulated bands can be changed using the attenuation controls in both the pre-processing circuitry and the modulation circuits. The values chosen for the video levels were determined experimentally and were those which worked best with the Octek digitizer. They will vary from system to system.

Appendix C: Listings of Computer Programs

Program "QFT" was used to generate the two dimensional Fourier transform images for both static and dynamic data sets. Program "PLOT" produced the baseband pixel value plots for the static data set. Both programs are implemented in Fortran V and run on the Eclipse computer. Not listed are the CALCOMP plotting subroutines and the Octek software (run on the Nova system) used to aquire data. Source listings for these programs are available at the AFIT Signal Processing Laboratory.

```
C     PROGRAM QFT - DG FORTRAN 5 - BY LT D L JONES - 23 NOV 84
C
C     CALL: QFT(/B) (output file name) (input file name)
C
C     SWITCH /B OMITS BASEBAND AND GIVES INDIVIDUAL LINEAR SCALING
C
C     USES SUBROUTINES IOF, REPACK, AND UNPACK
C
C     THIS PROGRAM COMPUTES THE FOURIER TRANSFORM OF A VIDEO FILE AND SPATIALLY
C     FILTERS THE DATA.  THE BASE BAND IS COMPUTED BY AVERAGING ALL PIXELS
C     HORIZONTALLY AND COMPUTING THE FOURIER TRANSFORM VERTICALLY.  THE TWO
C     MODULATED BANDS ARE COMPUTED BY REDUCING THE IMAGE TO A 64 X 64 ARRAY
C     AND COMPUTING THE TWO DIMENSIONAL FOURIER TRANSFORM POINTS IN THE PASS
C     BANDS.  ALL THREE BANDS ARE SCALED INDIVIDUALLY WITH A SQUAREROOT SCALING.
C
C     MAJOR VARIABLES
C
C         WORK - ARRAY CONTAINING COMPRESSED (64x64) PIXEL VALUES OF
C                ORIGINAL INPUT VIDEO FILE
C         BB   - ARRAY CONTAINING ROW AVERAGE PIXEL VALUES OF ORIGINAL
C                INPUT VIDEO FILE USED FOR BASEBAND COMPUTATION
C         BLK  - ARRAY CONTAINING FOURIER TRANSFORM POINTS USED TO
C                CREATE OUTPUT VIDEO FILE
C         TEMP, IO  - TEMPORARY ARRAYS USED IN UNPACKING AND REPACKING
C                     VIDEO FILES
C
      INTEGER MAIN(7),F3(7),MS(2),S1(2),S2(2),S3(2)
      INTEGER TEMP(256),IO(1024),R,C,IFN(7),OFN(7)
      REAL WORK(64,64),BLK(64,64),BB(240),MAX,RNG(2)

C     GET FILE NAMES
      CALL IOF(2,MAIN,OFN,IFN,F3,MS,S1,S2,S3)

C     OPEN OUTPUT FILE
      OPEN 5,OFN,ATT="RO"
      WRITE(10,1)OFN(1)
1     FORMAT(S14)

C     OPEN INPUT FILE
      CALL OPEN(3,IFN,1,IER)
      CALL CHECK(IER)

C     SET CONSTANTS
      P=3.14159265
      RNG0=100.                    ;   BASEBAND RANGE
      RNG(2)=400.                  ;   BAND 1 RANGE
      RNG(1)=200.                  ;   BAND 2 RANGE

C     LOAD BLOCK ARRAY WITH ZEROS

      DO 3 J=1,64
         DO 3 K=1,64
3           BLK(J,K)=0.0
```

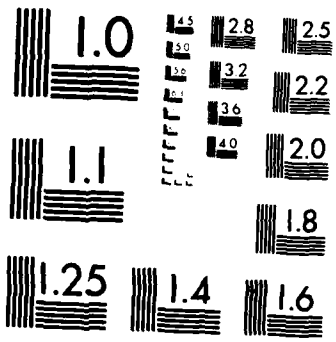
```
C       CHECK SWITCH /B
        IF(MS(1).EQ.16384)GOTO 45

C       LOAD BASEBAND WORKING ARRAY

        DO 41 I=0,59
           CALL RDBLK(3,I,TEMP,1,IER)
           CALL CHECK(IER)
           CALL UNPACK(256,TEMP,IO)
           DO 41 J=0,3
             BB(I*4+J+1)=0.0
             DO 41 K=1,240
41              BB(I*4+J+1)=BB(I*4+J+1)+IO(J*256+K)/240.0

C       COMPUTE BASEBAND TRANSFORM  /  COMPRESS BY 3  /  FIND MAX

          MAX=0.0
          DO 43 J=3,90,3
             K=INT(J/3)
             DO 42 JJ=0,2
                DO 42 M=1,240
                   A=FLOAT((1-M)*(J-JJ))*P/120.      ; COMPUTE TRIG ARGUMENT
                   BLK(30,33-K)=BB(M)/3.0*COS(A)+BLK(30,33-K)
42                 BLK(30,33+K)=BB(M)/3.0*SIN(A)+BLK(30,33+K)
             BLK(30,33+K)=SQRT(BLK(30,33+K)**2+BLK(30,33-K)**2)
43           IF(MAX.LT.BLK(30,33+K))MAX=BLK(30,33+K)

          J=0                                  ; DUMMY PRINT VARIABLE
          TYPE J,"   MAX = ",MAX

C         SCALE BASEBAND DATA

          DO 44 C=1,30
             IF(BLK(30,C+33).EQ.0.0)GO TO 44
             BLK(30,C+33)=SQRT(255.0*BLK(30,C+33)/RNGO)
             IF(BLK(30,C+33).GT.15.0)BLK(30,C+33)=15.0
44        BLK(30,33-C)=BLK(30,33+C)

C       PLACE ARTIFICIAL CENTER POINT
45        BLK(30,33)=15.0

C       LOAD PICTURE AND COMPRESS TO 1/4 ORIGINAL SIZE

        DO 10 I=0,59
C          LOAD FOUR ROWS FROM VIDEO FILE INTO IO BUFFER
           CALL RDBLK(3,I,TEMP,1,IER)
           CALL CHECK(IER)
           CALL UNPACK(256,TEMP,IO)
C          AVERAGE 16 POINTS AND LOAD INTO WORKING ARRAY
           R=I+1
           DO 10 C=1,64
              WORK(R,C)=0.0
              DO 5 J=0,3
```

```
                K=J*256+(C-1)*4              ; PIXEL INDEX IN IO
5               WORK(R,C)=WORK(R,C)+FLOAT(IO(K+1)+IO(K+2)+IO(K+3)+IO(K+4))
10          WORK(R,C)=WORK(R,C)/16.0

C     COMPUTE 2 DIMENSIONAL FOURIER TRANSFORM

C      SELECT FILTER BANDS
      DO 39 IB=2,1,-1
         MAX=0.0
         J=2*IB
         DO 25 K=IB-6,9-(IB-1)*7
            DO 20 M=1,60
               DO 20 L=1,60
                  A=FLOAT((1-M)*K+(1-L)*J)*P/30.0      ; COMPUTE TRIG ARGUMENT
                  BLK(30-J,33-K)=WORK(M,L)*COS(A)+BLK(30-J,33-K)
20                BLK(30+J,33+K)=WORK(M,L)*SIN(A)+BLK(30+J,33+K)
            BLK(30-J,33-K)=SQRT(BLK(30-J,33-K)**2+BLK(30+J,33+K)**2)
25          IF(BLK(30-J,33-K).GT.MAX)MAX=BLK(30-J,33-K)
            R=3-IB                                   ; DUMMY PRINT VARIABLE
            TYPE R,"    MAX = ",MAX

C     SCALE BAND 1 AND BAND 2 DATA

            J=IB*2
            DO 39 K=-10,10
               IF(BLK(30-J,33-K).EQ.0.0)GO TO 39
               IF(MS(1).EQ.16384)GOTO 33
                 BLK(30-J,33-K)=SQRT(255.0*BLK(30-J,33-K)/RNG(IB))
                 GOTO 35
33             BLK(30-J,33-K)=15.0*BLK(30-J,33-K)/MAX
35             IF(BLK(30-J,33-K).GT.15.0)BLK(30-J,33-K)=15.0
39             BLK(30+J,33+K)=BLK(30-J,33-K)

C     WRITE DATA TO OUTPUT FILE

       DO 55 I=0,63
          DO 53 J=1,1024
               K=INT((J-INT((J-1)/256)*256-1)/4)+1
53             IO(J)=ANINT(BLK(I+1,K))
          CALL REPACK(256,IO,TEMP)
          CALL WRBLK(5,I,TEMP,1,IER)
55        CALL CHECK(IER)

      CALL RESET
      STOP"<7><7><7><7>GFT"
      END
```

```
C     PROGRAM PLOT - DG FORTRAN 5 - BY LT D L JONES - 5 NOV 84
C
C     CALL: PLOT (input file #1) (input file #2) ... (input file #4)
C
C     USES CALCOMP SUBROUTINES AND UNPACK
C
C     TAKES FOUR VIDEO FILES AND COMPUTES THE BASEBAND TRANSFORM
C     THEN PLOTS ONE BLANK AXIS PLUS THE FOUR TRANSFORM DATA
C     SETS ON THE PRINTER; INTENSITY OF EACH TRANSFORM POINT VS
C     ITS DISTANCE FROM THE ORIGIN.
C
C     MAJOR VARIABLES
C         BB  - ARRAY CONTAINING THE ROW AVERAGE PIXEL VALUES OF
C               THE INPUT VIDEO FILES
C         BLK - ARRAY CONTAINING THE FOURIER TRANSFORM POINTS TO
C               BE PLOTTED
C         Y   - ARRAY USED TO TRANSFER DATA TO PLOTTING ROUTINES
C         TEMP, IO - TEMPORARY ARRAYS USED TO UNPACK THE INPUT
C                    VIDEO FILES

      INTEGER TEMP(256),IO(1024),R,C,IFN(7)
      DIMENSION BLK(64),BB(240),X(31),Y(31)

C     SET UP FOR COMMAND LINE INPUT
      CALL GROUND(IW)
      IF(IW.EQ.0)OPEN 1,"COM.CM"
      IF(IW.EQ.1)OPEN 1,"FCOM.CM"
      CALL COMARG(1,OFN,ISW,IER)
      CALL CHECK(IER)

C     SET CONSTANTS
      P=3.14159265
      RNG=100.0

C     LOAD BLOCK ARRAY WITH ZEROS
      DO 3 J=1,64
3     BLK(J)=0.0

C     OPEN INPUT FILE
      DO 70 IP=1,5
      IF(IP.EQ.1)GOTO 55
          CALL COMARG(1,IFN,ISW,IER)
          CALL CHECK(IER)
          CALL OPEN(2,IFN,1,IER)
          CALL CHECK(IER)
          WRITE(10,4)IFN(1)
4         FORMAT(S14)

C     LOAD BASEBAND WORKING ARRAY
      DO 41 I=0,59
         CALL RDBLK(2,I,TEMP,1,IER)
         CALL CHECK(IER)
         CALL UNPACK(256,TEMP,IO)
```

AD-A151 898 PROCESSING SPEECH FOR ANALYSIS USING OPTICAL FOURIER TECHNIQUES(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING D L JONES DEC 84 2/2

UNCLASSIFIED AFIT/GE/ENG/84D-37 F/G 9/3 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

```
          DO 41 J=0,3
            BB(I*4+J+1)=0.0
            DO 41 K=1,240
41             BB(I*4+J+1)=BB(I*4+J+1)+IO(J*256+K)/240.0

C       COMPUTE BASEBAND TRANSFORM  /  COMPRESS BY 3
          DO 43 J=3,90,3
             K=INT(J/3)
             DO 42 JJ=0,2
                DO 42 M=1,240
                   A=FLOAT((1-M)*(J-JJ))*P/120.      ; COMPUTE TRIG ARGUMENT
                   BLK(33-K)=BB(M)/3.0*COS(A)+BLK(33-K)
42                 BLK(33+K)=BB(M)/3.0*SIN(A)+BLK(33+K)
43           BLK(33+K)=SQRT(BLK(33+K)**2+BLK(33-K)**2)

C         SCALE BASEBAND DATA
          DO 44 J=1,30
             IF(BLK(J+33).EQ.0.0)GO TO 44
             BLK(J+33)=SQRT(255.0*BLK(J+33)/RNG)
             IF(BLK(J+33).GT.15.0)BLK(J+33)=15.0
44        BLK(33-J)=BLK(33+J)

C       PLACE ARTIFICIAL CENTER POINT
          Y(1)=15.0

C       LOAD DATA ARRAY
          DO 45 I=2,31
45        Y(I)=ANINT(BLK(I+32))
          CALL CLOSE(2,IER)
50        CALL CHECK(IER)

C       PLOT DATA ON PRINTER
55        DO 60 I=1,31
60        X(I)=FLOAT(I)
          IF(IP.NE.1)GO TO 62
             CALL PLOTS(0,0,6)
             XO=1.25
             YO=7.75
             GO TO 63
62        YO=-1.8
          XO=0.0
63        CALL PLOT(XO,YO,-3)
          CALL AXIS(0.0,0.0,"PIXEL DISTANCE FROM ORIGIN",-26,5.25,0.0,0.0,6.0)
          CALL AXIS(0.0,0.0,"INTENSITY",9,1.25,90.0,0.0,12.0)
          IF(IP.NE.1)CALL ALINE(X,Y,31,1.0,1,1.0,6.0,0.0,12.0)
70      CONTINUE
          CALL PLOT(0.0,0.0,999)
          WRITE(12,113)
113       FORMAT("    ")

        CALL RESET
        STOP"<7><7><7><7>PLOT"
        END
```

```
      SUBROUTINE IOF(N,MAIN,F1,F2,F3,MS,S1,S2,S3)
C
C     Written by Lt. Simmons          10 Sep 1981
C     Version 2
C
C     This FORTRAN 5 subroutine will read from the file
C     COM.CM (FCOM.CM in the foreground) the program name,
C     any global switches, and up to three local file
C     names and corresponding local switches.
C
C     Calling arguments:
C
C     N is the number of local files and switches to be
C     read from (F)COM.CM.  N must be 1, 2, or 3.
C
C     MAIN is an ASCII array for the main program file name.
C
C     F1, F2, and F3 are the three ASCII arrays to return
C     the local file names.
C
C     MS is a two-word integer array that holds any global
C     switches.
C
C     S1, S2, and S3 are two-word integer arrays that
C     hold the local switches corresponding to F1 through
C     F3 respectively.
C
      DIMENSION MAIN(7),MS(2)
      INTEGER F1(7),F2(7),F3(7),S1(2),S2(2),S3(2)
C
C     Check the bounds on N.
C
      IF(N.LT.1.OR.N.GT.3)STOP "N out of bounds in IOF."
C
C     Process the data in (F)COM.CM
C
      CALL GROUND(I)  ;Find out which ground program is in
      IF(I.EQ.0)OPEN 0,"COM.CM"        ;Open ch. 0 to COM.CM
      IF(I.EQ.1)OPEN 0,"FCOM.CM"       ;Open ch. 0 to FCOM.CM
      CALL COMARG(0,MAIN,MS,IER)       ;Read from (F)COM.CM
      IF(IER.NE.1)TYPE" COMARG error:",IER
      WRITE(10,1)MAIN(1)               ;Type program name
1     FORMAT(' Program ',S13,'running.')
      CALL COMARG(0,F1,S1,JER)         ;Read from (F)COM.CM
      IF(JER.NE.1)TYPE" COMARG error (F1):",JER
      IF(N.EQ.1)GO TO 2                ;Test N
      CALL COMARG(0,F2,S2,KER)         ;Read from (F)COM.CM
      IF(KER.NE.1)TYPE" COMARG error (F2):",KER
      IF(N.EQ.2)GO TO 2                ;Test N
      CALL COMARG(0,F3,S3,LER)         ;Read from (F)COM.CM
      IF(LER.NE.1)TYPE" COMARG error (F3):",LER
2     CLOSE 0
      RETURN
      END
```

[7:91]

```
      SUBROUTINE UNPACK(N,PIXWORD,PIXELS)
C
C     Written by Lt. Simmons          Version 2
C
C     This subroutine will unpack four 4-bit integers from a
C     16-bit integer word.  The pixels in a video file have to
C     be unpacked if each pixel is to be operated on separately.
C
      INTEGER PIXWORD(N),PIXELS(4,N)       ;Four pixels per word
      DO 1 I=1,N                          ;'N' allows higher-order
      DO 1 J=1,4                          ;arrays to be passed.
      PIXELS((5-J),I)=15.AND.PIXWORD(I)    ;Pick off right pixel
    1 PIXWORD(I)=ISHFT(PIXWORD(I),-4)     ;Shift word 4 bits right
      RETURN                              ;to pick off next pixel.
      END



      SUBROUTINE REPACK(N,PIXELS,PXWD)
C
C     Written by Lt. Simmons          Version 2
C
C     This subroutine will repack four 4-bit integer pixels
C     into one 16-bit word for use by CHOPS.  Parameter N
C     allows more than one 4-bit to 1-word repacking
C     operation in each call to REPACK.
C
      INTEGER PIXELS(4,N),PXWD(N)
      DO 1 J=1,N                      ;Loop N times
      PXWD(J)=0
      DO 1 I=1,4
      PXWD(J)=ISHFT(PXWD(J),4)        ;Shift pixel left in word
    1 PXWD(J)=PIXELS(I,J)+PXWD(J)     ;then add next pixel on right
      RETURN
      END
```

[7:98,99]

Bibliography

1. Flanagan, James L. Speech Analysis Synthesis and Perception. Springer-Verlag, New York, 1972.

2. Fletcher, Harvey. Speech and Hearing in Communication. D. Van Nostrand Co., New York, 1953.

3. Hussain, Ajmal. Limited Continuous Speech Recognition by Phoneme Analysis. MS Thesis, Wright Patterson AFB, Ohio: School of Engineering, Air Force Institute of Technology, December 1983.

4. Ladefoged, Peter. Elements of Acoustic Phonetics. University of Chicago Press, Chicago, 1962.

5. Linear Databook. National Semiconductor Corp., Santa Clara, CA, 1982.

6. Potter, Ralph K. et al. Visible Speech. D. Van Nostrand Co., New York, 1947.

7. Simmons, Robin. Machine Segmentation of Unformatted Characters. MS Theis, Wright Patterson AFB, Ohio: School of Engineering, Air Force Institute of Technology, December 1981.

8. Sivian, L. J. "Speech Power and Its Measurement," The Bell System Technical Journal. American Telephone and Telegraph Co., New York, 1929.

9. Witten, I. H. Principles of Computer Speech. Acedemic Press, New York, 1982.

10. Zeines, Ben. Electronic Communications Systems. Prentice-Hall, New Jersey, 1970.

VITA

Lieutenant Duane L. Jones was born 12 January 1957 in San Mateo, California. He graduated from high school in 1974 and moved to Provo, Utah, where he began attending Brigham Young University. In December of 1980 he recieved a Bachelor of Science degree in electrical engineering and was commissioned in the Air Force Reserve through the ROTC program. He entered active duty at Hill AFB, Utah as a test control officer and eventually served as Chief of the Munitions Test Unit and Range Instrumentation Unit, 2849th Munitions Test Squadron. He entered the School of Engineering at the Air Force Institute of Technology in June of 1983.

Permenant address: 3698 North 820 East
Provo, UT 84604

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): AFIT/GE/ENG/82D-35	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: School of Engineering; 6b. OFFICE SYMBOL (If applicable): AFIT/ENG	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code): Air Force Institute of Technology, Wright-Patterson AFB, Ohio 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. — PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification): See item 19.	
12. PERSONAL AUTHOR(S): Jones, Duane L., B.S.E.E., 1Lt, USAF	
13a. TYPE OF REPORT: MS Thesis; 13b. TIME COVERED: FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day): 1984, December; 15. PAGE COUNT: 98
16. SUPPLEMENTARY NOTATION	Approved for public release: IAW AFR 190-17. LYNN E. WOLAVER, Dean for Research and Professional Development, Air Force Institute of Technology (AIC)
17. COSATI CODES — FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Speech Recognition, Phoneme Recognition, Fourier Optics, Optical Fourier Transform, Spatial Filtering

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: Processing Speech for Analysis Using Optical Fourier Techniques

Thesis Chairman: Dr. Matthew Kabrisky, Professor of Electrical Engineering, Air Force Institute of Technology

Abstract:

A system for displaying speech as a two dimensional video image is presented. The speech is pre-processed by compressing its dynamic range and filtering to emphasize frequencies above 500 hz. Blanking and sync pulses are inserted to put the signal in standard video format, and every other field is blanked to prevent interference between fields in the interlaced display. (con't)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ OTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION: Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Matthew Kabrisky	22b. TELEPHONE NUMBER (Include Area Code): 513-255-5276; 22c. OFFICE SYMBOL: AFIT/ENG

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

(Item 19 con't)

Two dimensional variation is achieved by modulating the baseband audio up in the spectrum near a multiple of the video scan rate. The relationship between input frequency and pattern angle of the display is derived, and it is shown that the set of frequencies near a multiple of the video scan rate have points in the spatial frequency domain which lie in a straight line at a distance from the origin proportional to the scan rate multiple.

Two modulation frequencies are selected to display in the spatial frequency domain the location of the first and second formant peaks. The two modulated signals are mixed with the baseband audio and displayed simultaneously in a single image. The images are digitized and an optical Fourier transform is simulated on the computer by creating the image which would appear in the Fourier transform plane. Entire words are processed by assembling individual frames on video tape.

The system shows the capability of processing multiple high resolution bands of frequency information for a given signal, and demonstrates the feasibility of using optical processes in the analysis of speech signals.

END

FILMED

5-85

DTIC