

**MULTI-SIGNAL PROCESSING FOR VOICE
RECOGNITION IN NOISY ENVIRONMENTS**

by
Taysir H. Nayfeh

Thesis submitted to the Faculty of
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE
in
Industrial and Systems Engineering

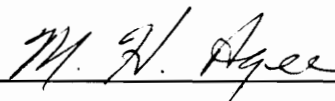
APPROVED :



O. K. Eyada, Chairman



M. P. Deisenroth



M. H. Agee

November 1991

Blacksburg, Virginia

c.2

LD
3055
V855
1991
N295
c.2

**MULTI-SIGNAL PROCESSING FOR VOICE
RECOGNITION IN NOISY ENVIRONMENTS**

by

Taysir H. Nayfeh

O. K. Eyada, Ph.D., Chairman

Industrial and Systems Engineering

(ABSTRACT)

One of the main input devices to computerized systems is Voice Recognition Systems (VRS). VRS is best suited for applications where job functions require more than two hands to be performed. The performance of VRS is highly dependent on the environment's noise. The higher the noise level the lower the recognition capability. Automatic lip reading through vision systems have been utilized to improve the recognition capability in noisy environments. However, this approach is costly and time consuming.

The objective of this thesis was to investigate the utilization of an Infrared sensor for automatic lip reading to improve the recognition performance of VRS. The developed system is cheaper and faster than other automatic lip readers. The test results of fifty words and eleven digits indicated that the method has good repeatability, and good character recognition, while not dependent on or sensitive to the ambient light level. Although speaker independence was tested, the results are inconclusive.

TABLE OF CONTENTS

1.0 INTRODUCTION	1
1.1 VOICE RECOGNITION SYSTEMS	1
1.1.1 VRS HARDWARE	2
1.1.2 ALGORITHMS AND METHODS	3
1.1.3 VRS PERFORMANCE AND CHARACTERISTICS	4
1.1.4 APPLICATIONS	5
1.1.5 LIMITATIONS	6
1.2 PROBLEM DEFINITION	7
1.3 NEED FOR VRS IN NOISY ENVIRONMENTS	8
1.4 RESEARCH OBJECTIVES	8
2.0 LITERATURE SURVEY	10
2.1 SPEECH	10
2.2 LIP READING	11
2.2.1 MANUAL LIP READING	11
2.2.2 AUTOMATIC LIP READING	12
2.2.3 VIDEO IMAGES	13
3.0 THE PROPOSED VRS SYSTEM	16
3.1 HARDWARE DESCRIPTION	19
3.2 DATA ACQUISITION AND CONDITIONING	21
4.0 DISCUSSION AND RESULTS	26
4.1 SYSTEM TESTS	26
4.1.1 SYSTEM VALIDATION	26
4.1.2 CALIBRATION	27
4.1.3 CHOICE OF SUBJECTS	27

4.1.4 DATA FORMATS	27
4.2 TEST RESULTS	28
4.3 SUMMERY OF RESULTS	43
5.0 CONCLUSIONS	46
5.1 FUTURE DEVELOPMENTS	46
5.2 APPLICATIONS	47
6.0 REFERENCES	48
7.0 APPENDICES	

LIST OF FIGURES

Figure 2.1	Improved recognition due to visual augmentation	14
Figure 3.1	An example of noisy optical signal	17
Figure 3.2	Example of raw optical and acoustic signals	18
Figure 3.3	System block diagram	20
Figure 3.4	Noisy baseline (acoustic signal)	23
Figure 3.5	Clean baseline (acoustic signal)	24
Figure 3.6	Isolated optical signal	25
Figure 4.1	Words with most distinct character	30
Figure 4.2	Words with least distinct character	31
Figure 4.3	Low dynamic range words	32
Figure 4.4	Example of similar sounding words	34
Figure 4.5	Example of similar character words	35
Figure 4.6	Example of complex character words	36
Figure 4.7	Difficult to distinguish words	37
Figure 4.8	Similar character, different aspect ratio words	38
Figure 4.9	Example of poor repeatability for speaker #3	40
Figure 4.10	Example of light level change	42

LIST OF TABLES

Table 4.1 Vocabulary test set	44
Table 4.2 Poor correlation words among speakers	45
Table 4.3 Good correlation words among speakers	45

1.0 INTRODUCTION

The predominate input method of information and commands into machines has traditionally been via human touch. This form of input uses a keyboard, a mouse or some other pointing device to enter sequences of single characters, or predefined function keys that are used to form words and commands which machines are programmed to understand and execute. Although this method of data entry is precise, it is slow except for the most skilled typist, and is limited to the number of function keys that are easily located. This method of input in most cases restricts the use of the hands for other functions.

Another form of man-machine communication is voice data input; this is a faster method since most people can speak faster than they can type. Unlike the keyboard, this method is not restricted to streams of single character entry or restricted by a limited set of function keys. In addition, the voice data input method does not require the full attention of the hands. It is unlikely that the voice input method will completely replace the other forms of input, however, this methods utility for many situations is far more superior to the other. Typical applications of VRS is for the physically disabled persons and in situations where the use of the hands is restricted, such as in advanced fighter cockpits, and robot manipulations for hazardous material handling.

1.1 VOICE RECOGNITION SYSTEMS

Voice commands are translated into a form that is understandable

by machines via Automatic Voice or Speech Recognition Systems (VRSs). The speech wave form is analyzed in the time-domain, the frequency-domain, or a combination of the two. In either domain, the features of time or frequency of a set of known words is determined and stored in templates. The features of incoming speech are compared with the stored templates to determine the closest match (i.e., recognition).

1.1.1 VRS HARDWARE

There are five main hardware components in a typical VRS. These are:

1. **Input Mechanism**: This consists of a microphone with some type of noise cancelling system to reduce background noise.
2. **Pre-processor**: This usually consists of a spectrum analyzer made up of 16 band pass filters, followed by logarithmic compression and ratification, and a spectrum shaper for normalization.
3. **Feature Extractor**: The feature extractor generates 32 binary signals, 16 of which are phoneme features and the other 16 are spectral features.
4. **Classifier**: Typically, a microprocessor is used in this section to perform the tasks of storing vocabulary templates, time normalization and speaker training. A decision algorithm is required to match the incoming utterances to the ones in storage.
5. **Feedback**: A video display or some other means of output is used to echo back the results of failed recognition attempts.

1.1.2 ALGORITHMS AND METHODS

The complexity and variety of the spoken language makes it extremely difficult to develop a single, fast, accurate, and inexpensive VRS. As a result, several algorithms have been developed and optimized with various processing methods to custom fit the utility and the application environment. The following are the main methods which are being used [1]:

(1) The best performing [1] and the most expensive VRS uses Dynamic Programming (DP), which include non-linear time wrapping for pattern matching, and Linear Predictive Coding (LPC) for acoustic analysis. DP has the highest accuracy rates because none of the data is discarded. This insures the best possible fit for the unknown signal. A drawback to this method is that the fixed sample intervals tend to distort the fine temporal features in speech. Also, this type of an algorithm requires high computational speeds and large amounts of memory.

(2) A second method for extracting features is zero-crossing analysis. The speech signal is usually divided into two frequency regions, the zero-crossings are then determined for each region, and the wave form is then integrated. This method is not as accurate nor as efficient as LPC, but additional divisions of the frequency-spectrum would improve it. The zero-crossing method works well with clear speech signals and is cost effective, since an inexpensive pre-processor can be used.

(3) Linear time compression can be used as an alternate method to nonlinear time wrapping. The common approach is to divide the signal into a set of equal parts. A representative point is stored in each spacing as part of the template. To avoid affecting the length of the template, a

spectral histogram is used to compensate for temporal differences. A variance is calculated for each data point to account for random variations. Linear time compression is a fast method, but involves large data reduction.

1.1.3 VRS PERFORMANCE AND CHARACTERISTICS

VRS can be classified into four major categories based on their performance and characteristics [1]:

1. **Discrete versus continuous speech**: VRS recognizes words by matching features of an isolated incoming utterance to ones stored. Locating the pauses between words results in word isolation (i.e., discrete speech). Because continuous speech has no pauses or very short ones, the task of isolating words becomes much more difficult and more time consuming. Continuous speech isolation is less accurate and, on the average, takes ten times longer than isolating discrete speech.

2. **Speaker dependence or independence**: There are large variations between individual speakers, such as accents, pronunciation, etc. Therefore, features of speech for one person are not necessarily representative of those of another person. Personalized speaker dependent systems are designed to the individual's own voice features. On the other hand, speaker independent systems must contain a statistical representation of features of several speakers. Because of the specialization in speaker dependent systems, the cost, the search time, and the accuracy is much better than speaker independent systems.

3. **Vocabulary size**: The vocabulary size of a system directly impacts the response time and accuracy. Large vocabulary requires a system with fast processing and more elaborate algorithms to achieve

acceptable response time and accuracy.

4. **Noisy environment:** The amount of signal processing is directly dependent on the level and type of noise. The standard designation of the quality of acoustic signals is the signal to noise ratio. High signal to noise ratio indicates a fairly clean speech signal, while low signal to noise ratio indicates a speech signal that is highly corrupted by noise. Although some types of noise are easier to filter than others, as a rule, when the signal to noise ratio deteriorates, filtering becomes more costly and more difficult, and for very low signal to noise ratio the speech signal is completely lost.

The performance measurements of voice recognition systems are determined by cost, speed, accuracy, vocabulary size and the versatility of the application environment. The overall quality of a system is a direct function of the features that the system possesses. With the current technology available, it is impossible to develop a system that contains all the above features. It is not uncommon to customize systems for specific applications since most applications do not require all the features listed above. By limiting the features, the system can deliver good performance.

1.1.4 APPLICATIONS

VRS is the ideal method of interfacing between man and machine. It can deliver good results despite the complexity of the human voice and the spoken language. Discrete word recognition systems are relatively common in friendly environments [2]. Some systems have large vocabularies reaching 20,000 words or more, however, most of these systems are speaker dependent. Speaker independent systems are not as common and typically the vocabulary is limited and discrete. Continuous

speech recognizers are not as common and most require a pause between words.

There is an enormous potential for applications of speech activated controls in the military and manufacturing settings. It is becoming exceedingly difficult for people with two hands to handle complex functions, such as weapon systems in jet fighters and tanks. Manufacturing tasks such as robot manipulation, order entry, material handling, assembly line operations, and machine controls are areas that can be improved by voice communications.

Another major application of VRS is for people with hand impairments. VRS gives the hand impaired the ability to communicate and control machines with voice commands. This application would contribute to the quality of life by giving the user self esteem by allowing him or her to contribute more to society.

1.1.5 LIMITATIONS

Although voice recognition systems have been successful in friendly office environments, they have not met with the same success in manufacturing and military environments. Manufacturing and military applications have been difficult because of the high noise levels present in these settings. The factory and military settings are not as forgiving of errors as the office setting. High degree of certainty and fast computational speeds are a must for real time reliable control. Weapon systems and flight functions in a fighter aircraft must be handled with the utmost care. Some of the manufacturing functions also require the same level of diligence as military applications.

Improved algorithms and faster computers are making these

applications more feasible. Sophisticated filtering and processing can improve the signal to noise ratio and thus the recognition rate. Unfortunately, as the signal to noise level degrades so does the confidence in the recognition. In the absence of additional information about speech, the results become suspect and, therefore, not suitable for application in crucial situations. Although in the military and manufacturing, the voice recognizer is usually expected to respond to a limited set of discrete commands, high degree of certainty and real time recognition is essential. Slowness and wrong recognition can lead to grave results.

1.2 PROBLEM DEFINITION

The standard methods of human communication evolved around the voice's acoustic signals, and so has the traditional automatic speech recognizers for non-contact human machine interaction. In situations where speech intelligibility is reduced due to high noise levels or loss of hearing, humans supplement or replace voice signals with facial expressions. Lip reading is the most disciplined method for this form of communication; 40-50% success ratio is achieved by trained experts in lip reading [3]. Methods to automate lip reading and combine this secondary source of speech information with the acoustic signals are being investigated for applications in noisy environments and for the hearing impaired.

Researchers have shown that images of facial expressions can be fused with the voice signals to improve speech recognition. This requires the acquisition and the analysis of video images of the mouth during speech and combining them with the acoustic signal for real time analysis.

Hardware requirements for these systems, which may include parallel architecture computing machines, neural networks, and the video equipment for acquiring the facial images coupled with the intense computational requirements for the analysis of the video images makes it costly and cumbersome [4]. Although the method yields improved results in noisy environments, the additional hardware and computational intensity makes it infeasible for real time application in some military and manufacturing areas.

1.3 NEED FOR VRS IN NOISY ENVIRONMENTS

The existing needs in noisy manufacturing and military applications of a real time, redundant, and cost effective VRS has not been satisfied. Current systems that are based solely on acoustic analysis are slow, mostly expensive, not redundant, and cannot deliver an acceptable level of accuracy and certainty in these settings. Fusing video images of the mouth and lips with the acoustic information improves the recognition rate in noisy environment [4], however, the size of the video data is large and of a descriptive form which requires a large amount of processing. This prevents this form of processing from delivering real-time results, and the additional hardware makes it costly and cumbersome.

The enormous potential of VRS application in the above settings, and the current limitations makes it imperative to pursue improvements in this field.

1.4 RESEARCH OBJECTIVES

The objective of this research work is to develop a low cost and fast automatic lip reading device. The intent is that the output of this device

can be easily fused with the acoustic speech information; the purpose is to improve automatic speech recognition in noisy environments. To simplify the fusion of the two sensors, the output format of the device has to be a time varying DC voltage i.e., the same format as the acoustic signal. A qualitative prototype system which meets the above requirements has been developed and tested. The test results indicate that this method of speech representation is viable and feasible. The detailed system design and test findings are presented in this work.

2.0 LITERATURE SURVEY

Applicability of VRS in the manufacturing and military environments is limited due to their stringent requirements. Noisy manufacturing and military environments require real time response, backup redundancy and very high accuracy. In addition, speaker independent systems are highly preferred. Although the required vocabulary size in these environments is usually limited, still, systems with large vocabularies are preferred for wider applicability. Currently, it is unlikely that a system meeting these requirements can be developed, and in addition, if such a system is developed, it will be extremely costly and slow.

The state of the art of VRS is a collection of application oriented systems. Since the aim of this research is to develop an automated lip reading device to improve voice recognition in noisy environments, the literature survey will be limited to this area along with some pertinent aspects.

2.1 SPEECH

Verbal sounds, the primary form of communication between humans, is extenuated with secondary ones such as facial and body expressions. The tongue and lip movements modulate sounds produced by the vocal tracts. The shortest distinguishable sounds are called phonemes. Phonemes that depend on the position of the lips are highly visible to the viewer. Other phonemes that depend on the tongue are slightly visible, and homophones are phonemes that are visually similar. Although homophones are not distinguishable by themselves, words

containing them will most of the time have enough differences to be distinguished.

During early human development vocal sounds were chosen as the primary means of communication rather than visual means. The fact that one can hear messages from others and hear events that may or may not be insight increases the utility of this form of communication. The visual and auditory receptors are related to different sense modalities with separate ways of signal processing and different types of pattern recognition methods [5]. Each pattern recognition scheme has to be optimized for the type of information present. In the evolution of speech, sounds were optimized for the auditory system and it is obvious that duplication of sounds were avoided. The movement of lips, tongue, and facial bones evolved in a strictly supportive role to help produce the various sounds, therefore, care was not taken to avoid duplication of the lip movements for the production of some sounds, indeed, it was not necessary. While this is an unfortunate turn of events for the hearing impaired, we are extremely lucky that lip movements is unique for most of the spoken words.

2.2 LIP READING

2.2.1 MANUAL LIP READING

Lip reading for the hearing impaired has been present for a long time and has evolved into a science with considerable success. A 40-50% accuracy rate is reported for trained experts in manual lip reading [3]. Although this is the upper accuracy limit for manual methods, it is not an upper limit for automated methods. The ability of humans to resolve the

visual phonemes is limited to 8-10 sounds per second, whereas speech is produced at the rate of 13 sounds per second [6]. Therefore, some amount of speech information is automatically lost by lip readers. The loss of parts of the visual information and the required correlation of rapidly varying images of the mouth imposes an upper limit on the performance of manual lip reading. Automated systems that are capable of running at high frame rates, can capture all the visual information present in speech, and of course, computers can handle the processing of this type of information much faster and simpler than manual methods.

2.2.2 AUTOMATIC LIP READING

The literature survey shows that the level of research in this area is sparse at best. The first serious attempt at automating lip reading did not occur until the mid sixties. Dr. Ernie Nassimbene of IBM informed Petajan [6] during a telephone conversation, that the first attempt at automating lip reading was accomplished by him. He reported that he was issued a patent for an automatic optical lip reading device in 1965. The device used a course array of photocells to capture light reflected off the teeth during speech. Petajan reported that Dr. Nassimbene's device worked fairly well, but was not used because the speaker was required to wear a mask to reduce ambient light noise. This made the device cumbersome and undesirable.

Hesselmann [5] reported in 1983, that his efforts at coding phonetic information into visible patterns were not successful. He attributed this to the fact that humans have not developed an efficient method of visual pattern recognition during evolution. Hesselmann attempted to simplify the visual recognition process by using Fourier descriptors to analyze lip-

contours of isolated spoken vowels. These contours were to be used for matching the patterns. He concluded that only a small number of significant structural parameters can be extracted to describe the visual information related to lip-contours during the articulation of speech. The literature search did not show any additional progress in this area.

2.2.3 VIDEO IMAGES

In 1985, Petajan [6 , 7] developed an automatic lip reading system that captures video images of the mouth. The system uses a code book of images to translate the incoming images into corresponding symbols. The symbols are compared to stored symbols for word recognition. The system worked very well for a limited vocabulary, but the computational intensity and the large amounts of data prevented this system from real time applications. Furthermore, the difficulty of obtaining consistent video images under ambient lighting conditions along with the hardware required makes it operationally difficult.

Yuhas, et al., [4] showed (see Figure 2.1) in 1989 that fusing visual images of the mouth with the acoustic signals yields improved voice recognition in noisy environments. Yuhas claimed that his method is superior to past methods because he used the analog video images rather than the digitized ones, and he fused the visual images with the acoustic signals before the classification process was performed. Yuhas notes that digitizing the video images reduces their resolution and the classification process tends to distort the original information. In Yuhas' [4] work, fusing the two signals was achieved by using a neural network and parallel processing architecture. The images used were static images of vowels, and with these, they demonstrated that the network was able to extract

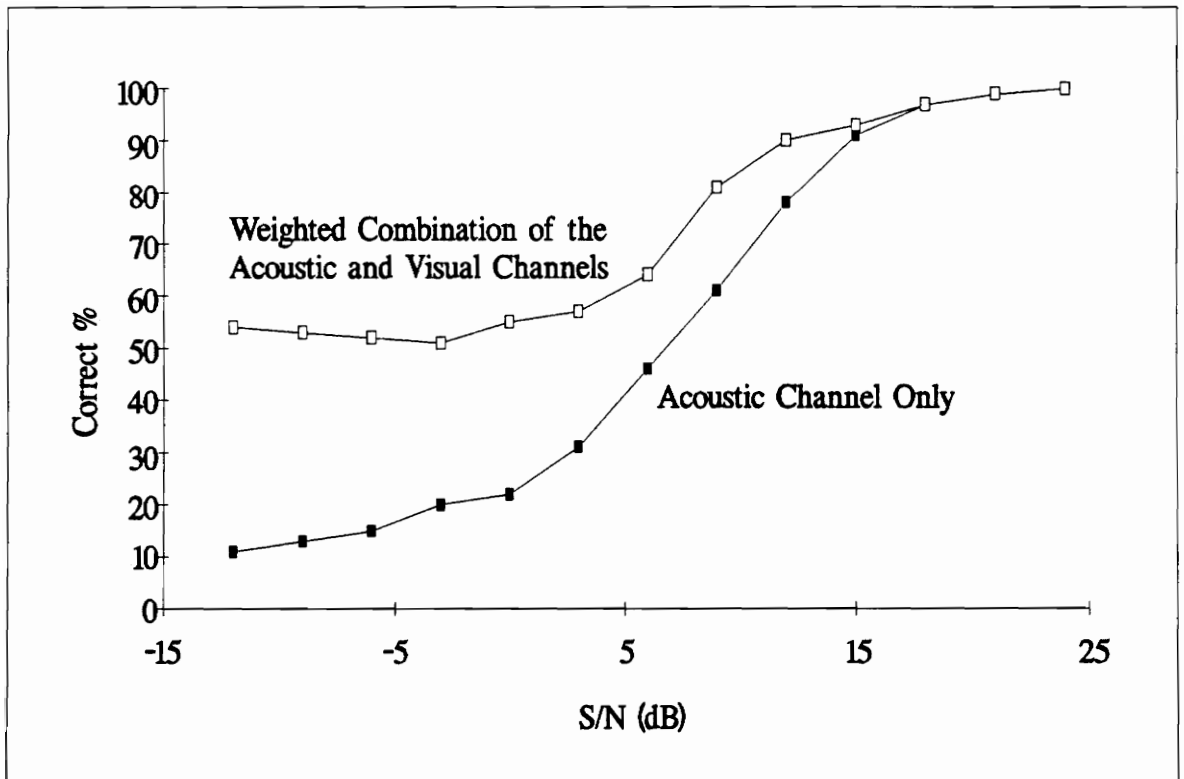


Figure 2.1 Improved recognition due to visual augmentation of noise-degraded speech

speech information from the visual images and were able to improve the recognition rate in the presence of noise.

The literature survey [4] showed (Figure 2.1) that combining the video images and the acoustic signals of speech improves automatic speech recognition in noisy environments, unfortunately, this method requires the acquisition and the analysis of large amounts of descriptive data along with the acoustic wave forms. The fusing of these different types of data is difficult, expensive, and will require sophisticated signal processing. The ultimate form of optical speech sensor is one that's inexpensive, fast, and has an output that's compatible with acoustic wave forms. The objective of this research was to develop such a sensor.

3.0 THE PROPOSED VRS SYSTEM

The automatic lip reading device proposed in this work monitors the pattern of motion of the lips during speech. The device uses an infrared light source modulated at 30kHz and is directed at the mouth. An optical detector receives the reflected infrared light along with ambient light. A frequency demodulator separates the 30kHz modulated infrared reflection from the ambient noise, thereby providing a clean signal that represents the motion of the mouth and lips. This method has the advantage of capturing more speech information and does not require the use of a light mask as utilized by Dr. Nassimbene [6].

The physical concept of the measurement is based on the fact that the motion of the lips is highly correlated to speech during the generation of most words, and that the human tissue is highly reflective to light in the infrared region of the spectra. The amount of infrared light reflected by the lips is measured and converted into a time varying DC voltage. The patterns generated are compared to previously stored patterns of the user for speech recognition.

The motion of the mouth and lips is not unique to speech. Breathing, twitching and general movements cause false signals as shown in Figure 3.1. Movements associated with speech can be isolated from noise by triggering the optical signal with the acoustic signal as shown in Figure 3.2. Facial expressions assume the same position when repeating words, otherwise the sound of words would be different each time a word is spoken. The length of the optical signal is increased by one fifth of a second to capture the lip movements that preceded and followed the actual production of sound.

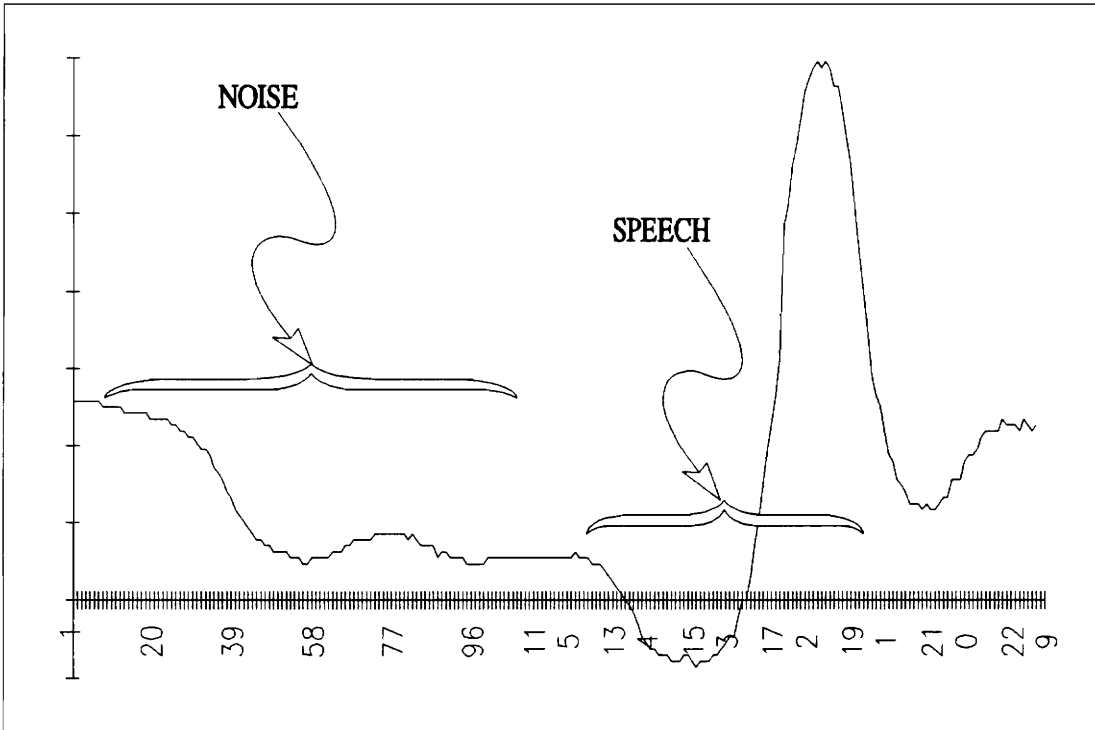


Figure 31 An example of noisy optical signal

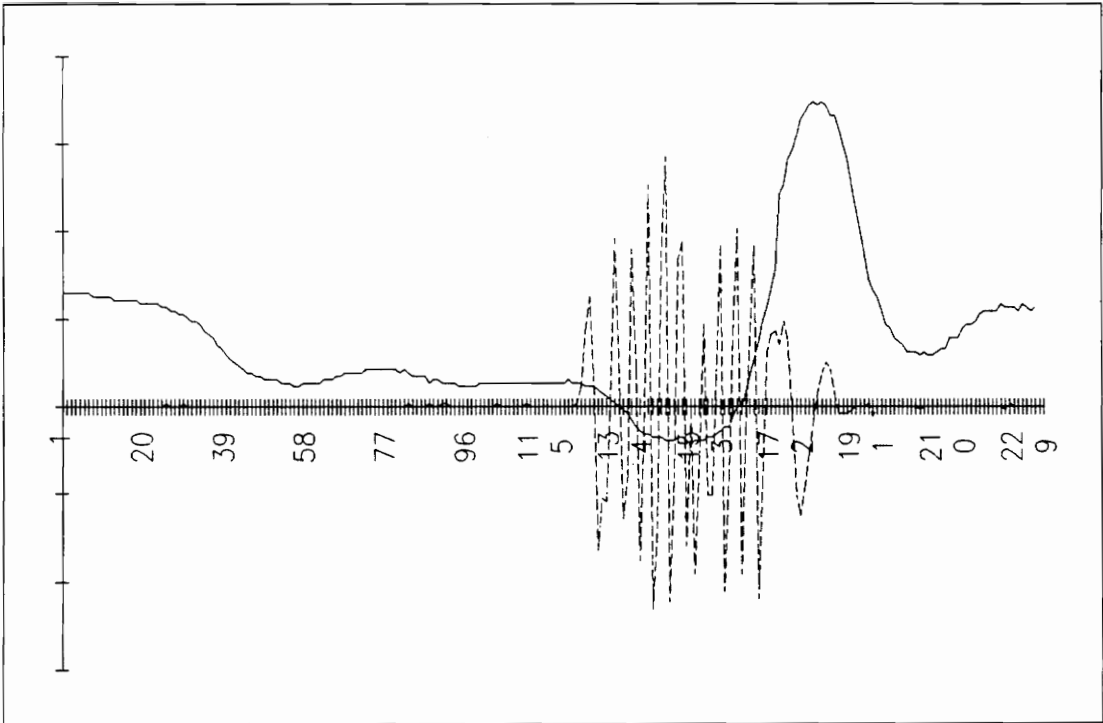


Figure 3.2 Example of raw optical and acoustic signals

3.1 HARDWARE DESCRIPTION

The hardware is divided into five operational blocks as outlined in Figure 3.3. The major components of the system are:

- (1) Sensor section (emitter, detector, microphone)
- (2) Emitter driver
- (3) Front end digital signal conditioning
- (4) Analog signal processing
- (5) Data acquisition

The following is a brief functional description of these blocks:

(1) **Emitter/Detector**: The emitter is a miniature infrared source with peak emission at 960nm and has a maximum power of 25mw. The detector is a miniature, low dark current planar photovoltaic diode. The pair are mounted on the mouth piece of a headphone facing the speaker, and are spaced approximately 1 1/2". The audio section consists of a high sensitivity miniature microphone, an audio amplifier and a frequency equalizer. Noise cancelation features are not included because the signal is only used as a marker.

(2) **Emitter driver**: The driver board provides the power to the emitter through a regulated power supply, pulsed at 30khz. Synchronization of the emitter with the detector is accomplished by using the same 30khz clock for both sections. A one kohm attenuator controls the output intensity of the emitter.

(3) **Front end digital signal conditioning**: This section contains a 30khz clock for modulating the output signal and demodulating the

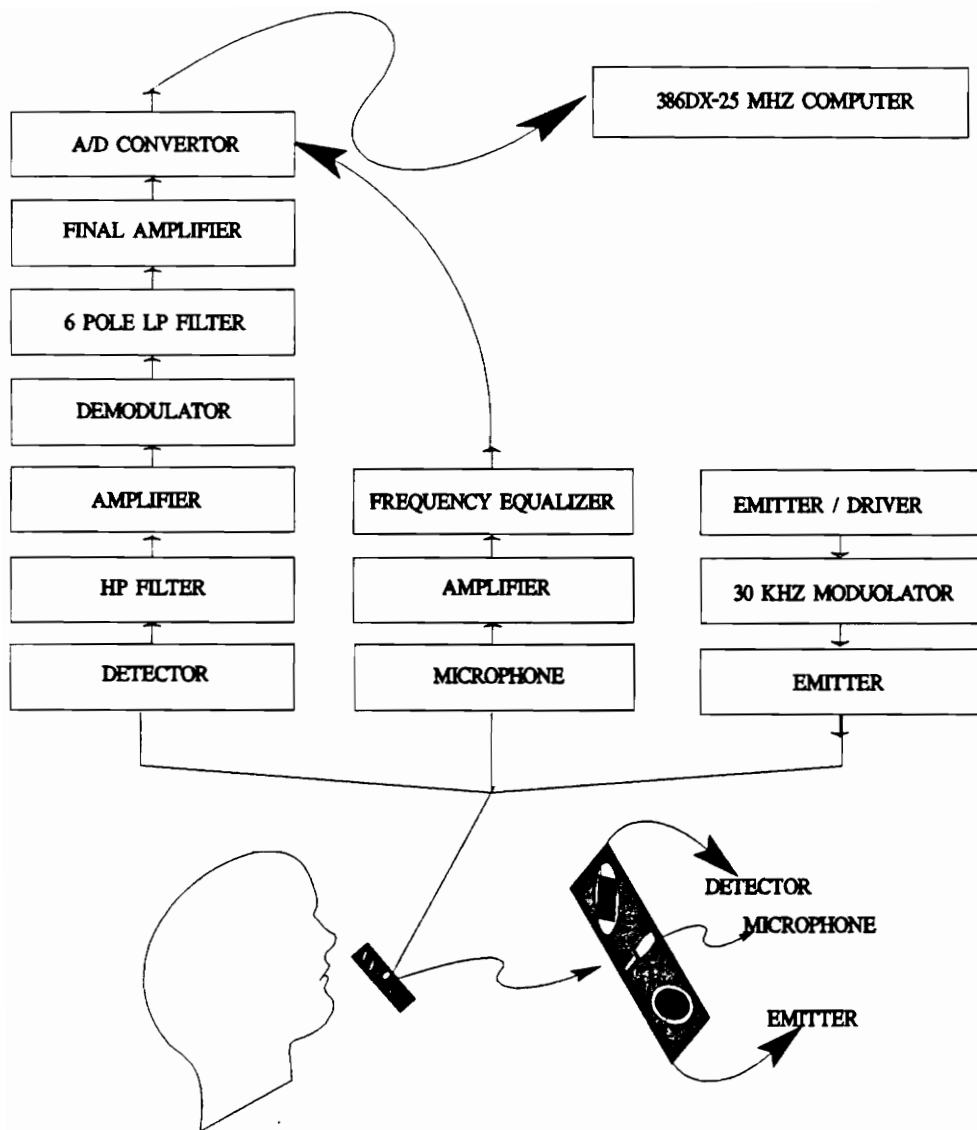


Figure 3.3 System Block Diagram

signal received. The received signal consists of an amplitude modulated 30khz component along with background noise. The signal is preamplified and passed through a high pass filter to remove any DC offset and low frequency noise. The demodulation of the signal is achieved by combining the output of an inverting and a noninverting amplifiers at the clock rate. The 30khz component of the signal generates a DC voltage and when combined with the other signal components results in a null signal.

(4) **Analog signal processing:** A six pole low pass filter is incorporated in this section to remove any Radio Frequency (RF) remnants, and is followed by a high gain amplifier to set the signal level.

(5) **Data Acquisition:** This section consists of an AT Bus architecture 386DX micro computer running at 25Mhz, and an 8 bit analog/digital converter board with a 20 micro second conversion time. The output of the final stage of the analog signal conditioning circuit and the acoustic signal are digitized at a sample rate of 100hz. The digital values are stored in an array in the RAM. The acoustic signal is used only to trigger the optical signal, and therefore, there is no need for higher sampling rates.

3.2 DATA ACQUISITION AND CONDITIONING

The data acquisition software is written in Basic. A copy of the program is included in appendix A. The acoustic and optical sensors' outputs are sampled and digitized sequentially at a sampling rate of 100 hz. The data conditioning and isolation of the optical speech information is performed digitally in six steps:

(1) The mean of the first 10 digital samples of the acoustic signal

is subtracted from the channel to normalize the signal base line.

(2) Small variations in the base line are filtered by setting a sensitivity threshold of .3 volts. This eliminates false signal detection (see Figures 3.4 and 3.5).

(3) The maximum and minimum values are determined for the acoustic signal.

(4) The sample indexes corresponding to 20% of the maximum and minimum of the acoustic signal's amplitude is determined. This is the likely beginning and end of speech.

(5) Twenty digital samples (1/5th of a second) are added to the beginning and end of the signal. This accounts for the mouth's movements that precedes and follows the actual speech.

(6) The acoustic signal indexes are used to mark the beginning and end of the optical speech information which is shown in Figure 3.6.

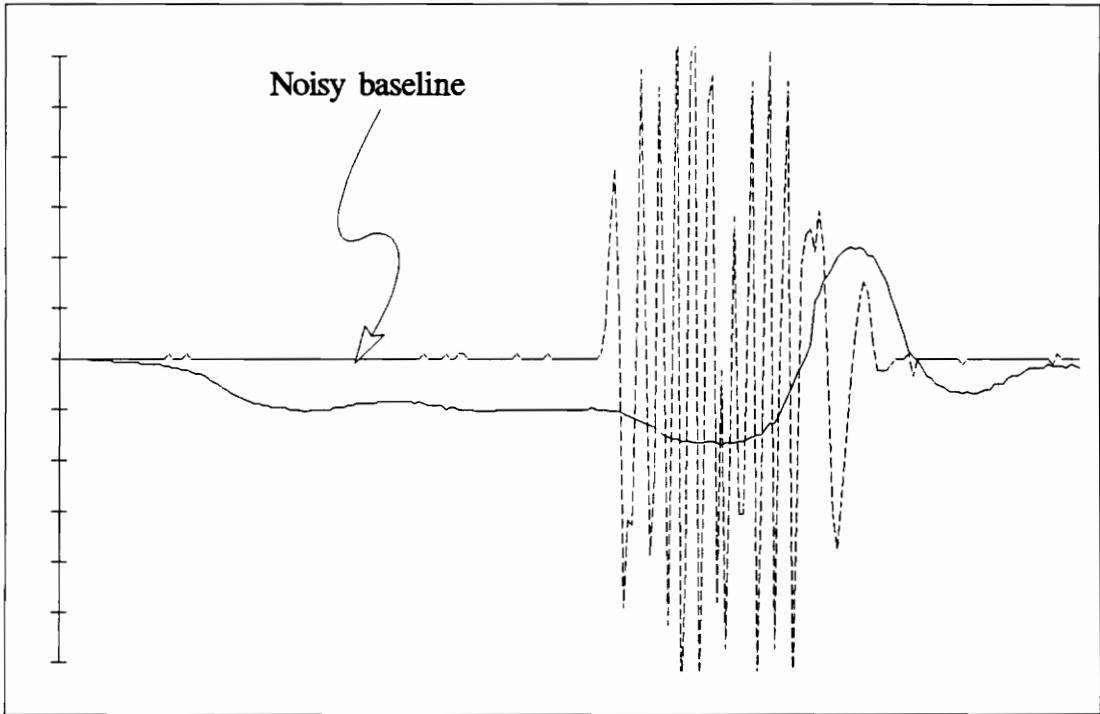


Figure 3.4 Noisy baseline (acoustic signal)

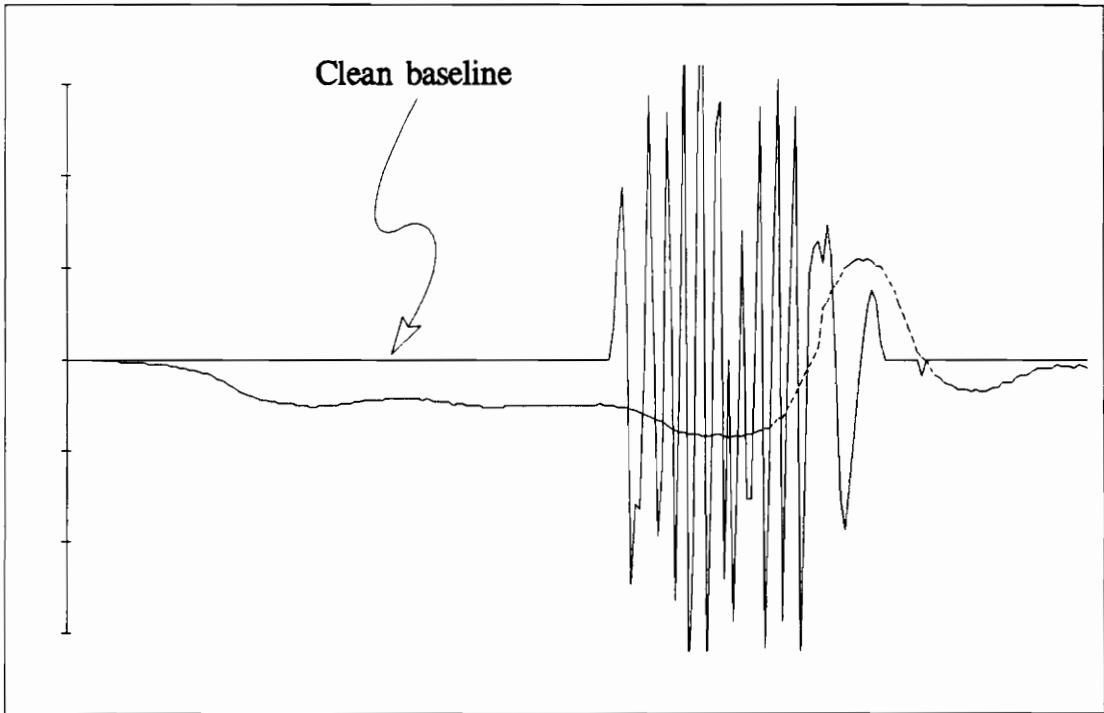


Figure 3.5 Clean baseline (acoustic signal)

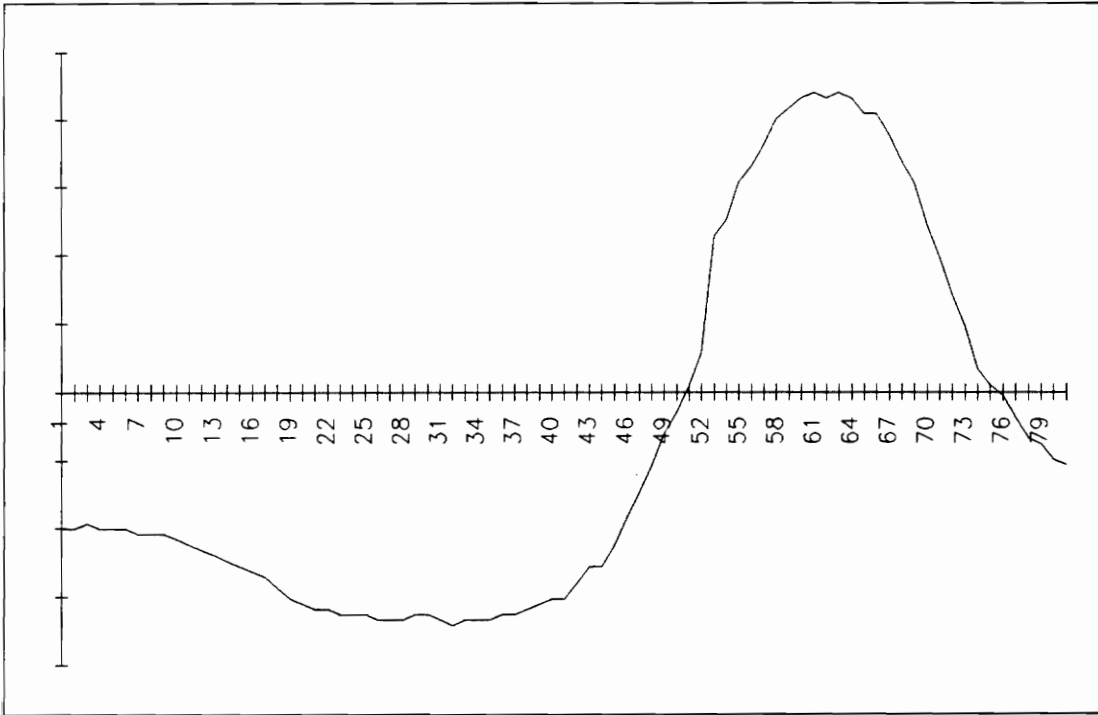


Figure 3.6 Isolated Optical Signal

4.0 DISCUSSION AND RESULTS

4.1 SYSTEM TESTS

4.1.1 SYSTEM VALIDATION

The main objective of this pilot study was to determine the viability and feasibility of the proposed method and whether the system warrants further development. The cost and effort associated with developing a fully operational experimental Voice Recognition System which fuses the optical and acoustic signals is beyond the scope of this pilot study. Therefore, the validation tests that were selected are qualitative and limited in scope because of the current system's capabilities. The current system does not have automatic signal gain nor automatic time compensation capabilities. For this reason it is not possible to conduct quantitative data analysis. Instead, the following qualitative criterion were used for the analysis: The number and order of the peaks and valleys, the relative width and amplitude of the waveform, and the aspect ratio. This set of qualitative criterion is sufficient to visually characterize the test data. The tests are:

1: **Feature characteristics**: 50 common words in manufacturing and the numbers 0-10, listed in Table 4.1 (page 44), will be used to test the feature uniqueness for all speakers.

2: **Repeatability**: The previous experiment will be repeated three times for the three speakers.

3: **Speaker dependence or independence**: The patterns generated by different speakers will be compared.

4: **Lighting conditions:** The fifty word test for the individual speaker will be repeated in ambient bright light and compared to tests that were conducted with normal office lighting.

SUMMARY of TESTS

CHARACTER	THREE SPEAKERS
REPEATABILITY	THREE SPEAKERS
LIGHT SENSITIVITY	ONE SPEAKER
SPEAKER INDEPENDENCE	ALL DATA

4.1.2 CALIBRATION

The position of the detector/emitter pair was qualitatively located by matching the characteristic of the waveform for the word MAIM for the various speakers without regard to the signal level generated. This worked fairly well for the various speakers and the different testing sessions. However, the eight bit resolution of the A/D convertor made it difficult to acquire good signals for words that has slight lip movements.

4.1.3 CHOICE OF SUBJECTS

The subjects were chosen such that a variety of pronunciation could be tested. A native English speaking female, a male with excellent command of English pronunciation and a male with a moustache and heavy accent were chosen to represent a wide range of speakers.

4.1.4 DATA FORMATS

The three vocabulary repeat curves for each subject are displayed

in Appendix A. Only one full vocabulary test for one speaker was required for the light sensitivity study. The set of light sensitivity curves is displayed along with the other three repeats for the first speaker. The individual vocabulary curves for all the speakers (10 per word) are combined together and displayed in Appendix A under the heading (ALL). This allows quick comparison of all the available data and simplifies the speaker dependence or independence portion of the study.

The temporal differences between the repeats and various speakers, in addition to the inaccuracy of the current acoustic marker for optical speech isolation, both caused the optical curves to be poorly stacked. The individual curves were manually shifted to improve the visual analysis. In most cases, the curves were slightly shifted (2-5 samples), few cases required larger shifts (10-15 samples).

The amplitude scale for the curves of the individual speakers is fixed for all the cases. The fixed scale enabled spotting relatively small differences in similar words. The amplitude scale for the curves listed under ALL in Appendix A was allowed to vary to highlight the relative characteristics of the individual curves.

4.2 TEST RESULTS

1: **Feature Characteristics:** The waveform patterns for fifty words along with the digits zero through ten (see table 1, page 44) are presented in Appendix A. The most distinct wave forms were for words that require large lip movements such as MAIM, ROBOT, WRISTUP, etc. as shown in Figure 4.1. Words that require small lip movements such as CUT, LATHE, SAW, etc., have the least character as illustrated by Figure (4.2). Because of the low resolution of the A/D convertor and without the

benefit of an automatic gain control circuit the least character (dynamic range) words tend to generate choppy waveforms as shown in Figure 4.3, but the overall character is moderately visible.

Similar sounding words, such as MAIM and MAIN, are distinct.

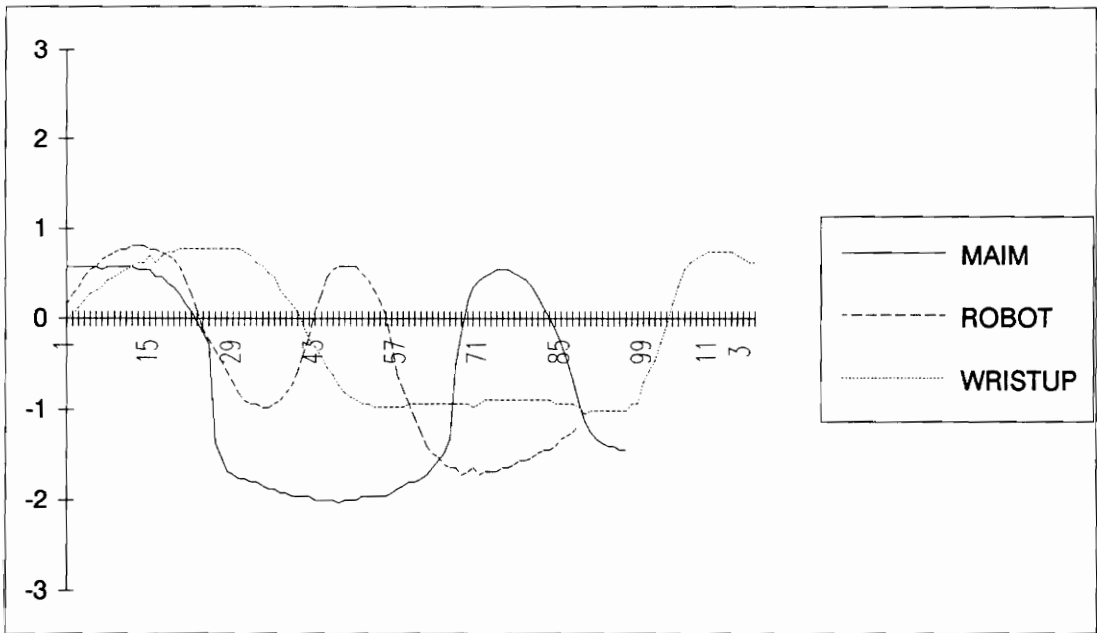


Figure 4.1 Words with most distinct character

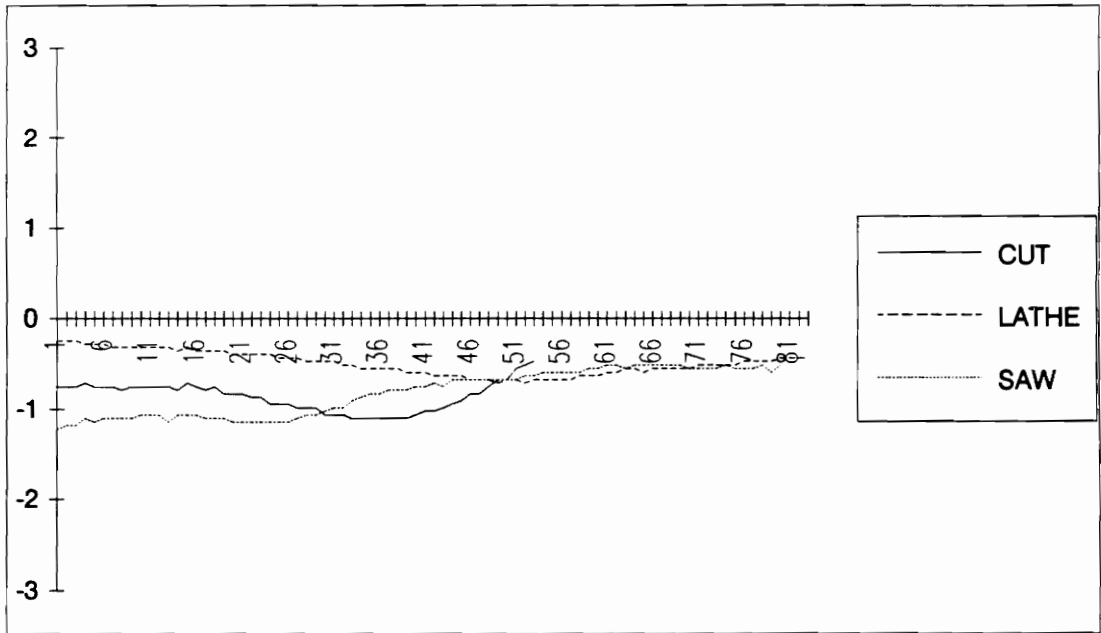


Figure 4.2 Words with least distinct character

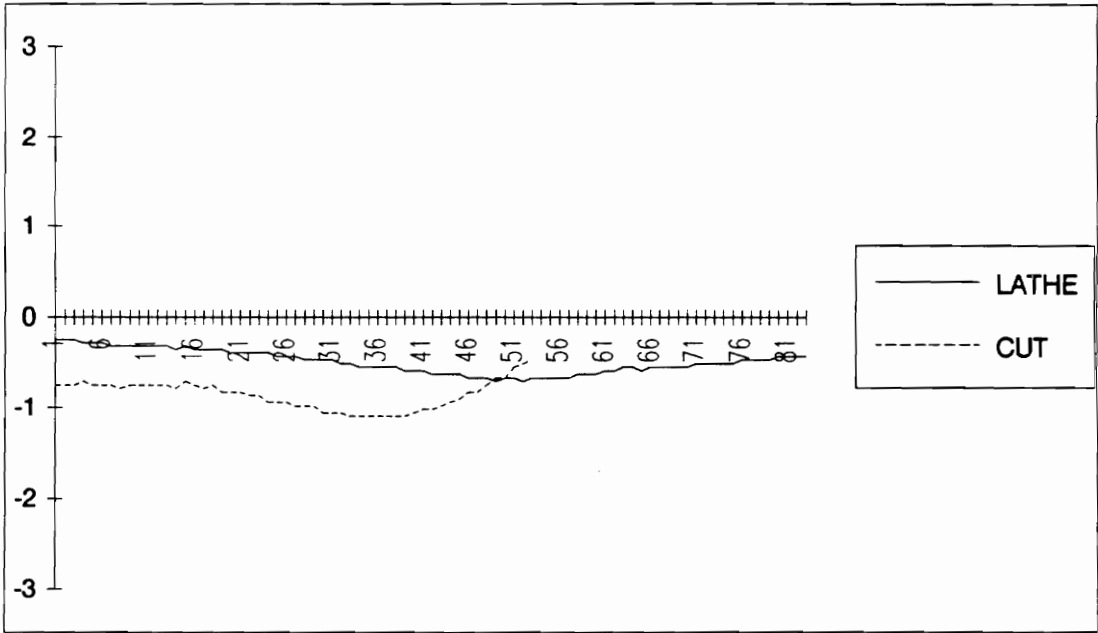


Figure 4.3 Low dynamic range words

because the letter M generates a peak whereas the letter N generates a valley (see Figure 4.4). The beginning of the waveform generated from the words NAME, DAME, and SAME is similar in that they start out with a valley and the M sound at the end is the same as shown in Figure 4.5. The difference between them is moderately visible from the depth and width of the beginning valley. However, words such as LIFT and LEFT are difficult to distinguish (Figure 4.7). Complex words such as WRISTUP, WRISTDOWN, ROTATE, and ROBOT, are highly distinct and bear much more characteristics that are easily definable as seen in Figure 4.6.

The curves for the colors RED, WHITE, BLUE, GREEN and GRAY are similar in character but are distinguishable as illustrated in Figure 4.8. The aspect ratio and the signal level generated provides very good clues about the individual words.

The waveforms for most of the numbers ZERO through TEN have good definition for speakers one and two. The overall characters of their waveforms contain several distinguishing features making them feasible to distinguish. The numbers TWO and SIX for speaker number three has low dynamic range with moderate character definition (see Figure 4.9). The features are not as distinct as the ones of the other digits for the same speaker.

Over all, the character of the waveforms of most of the words and the digits tested is distinct either in the overall features, such as the number and order of the peaks and valleys or in the qualitative analysis of the dynamic range and the aspect ratio of the signals.

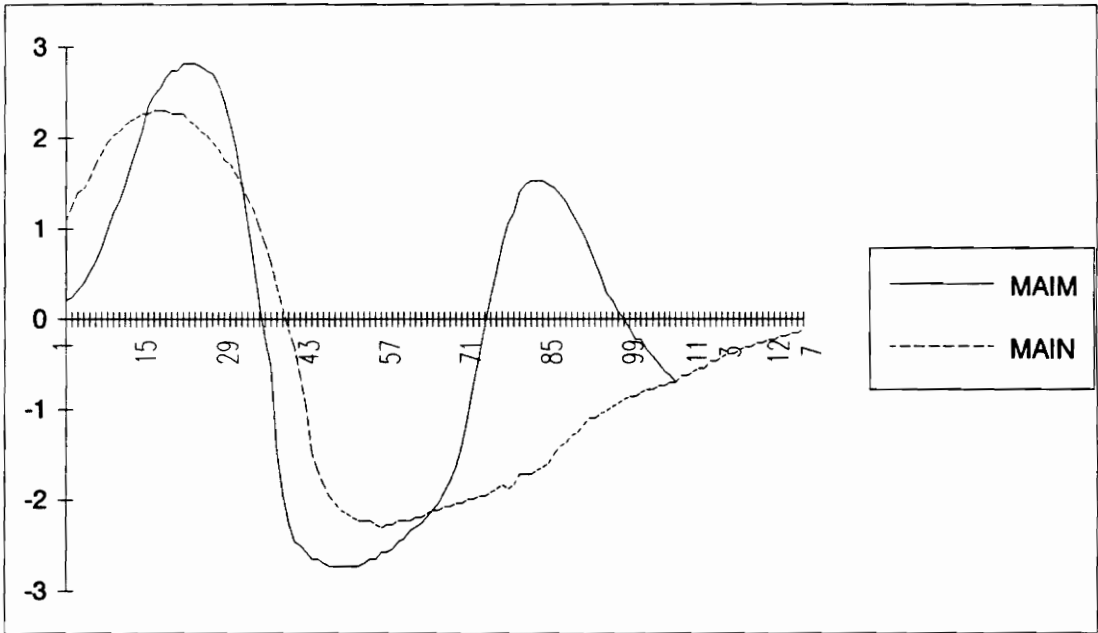


Figure 4.4 Example of similar sounding words

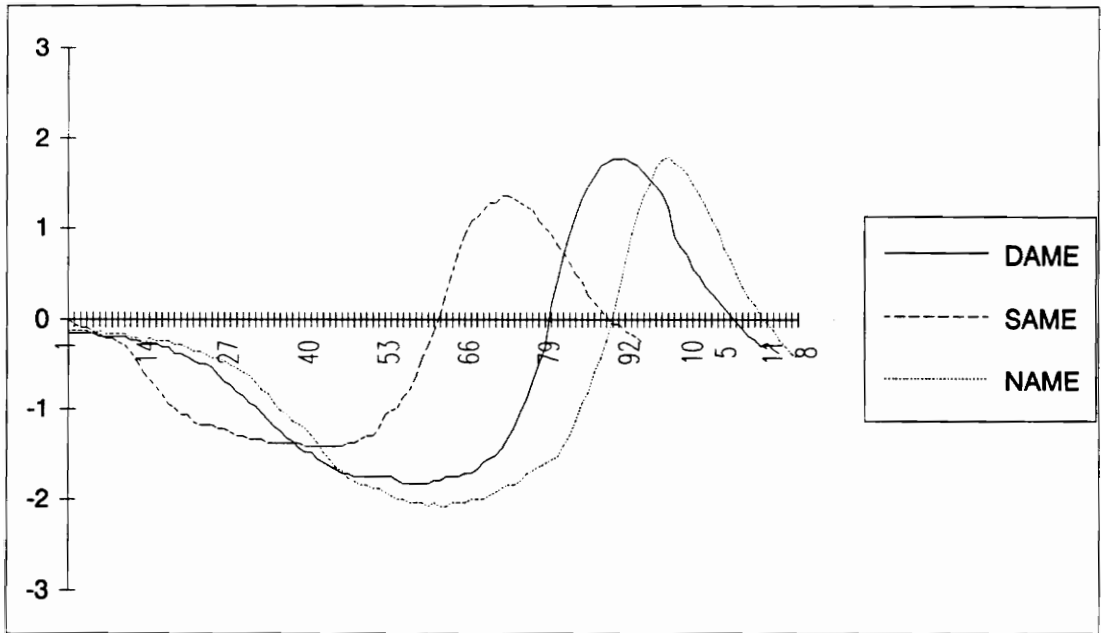


Figure 4.5 Example of similar character words

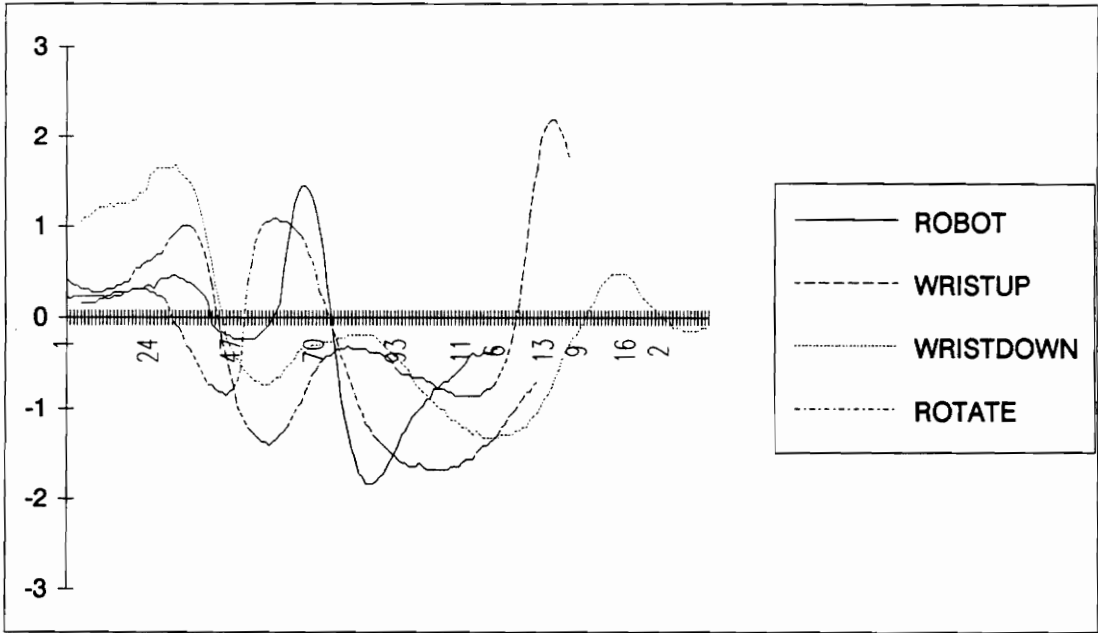


Figure 4.6 Example of complex character words

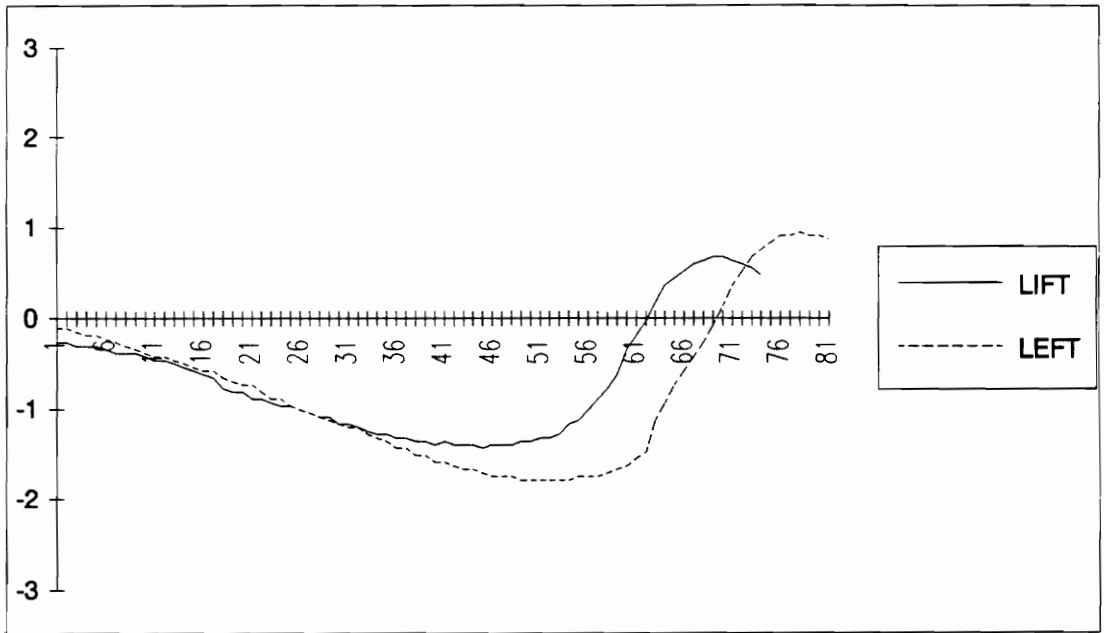


Figure 4.7 Difficult to distinguish words

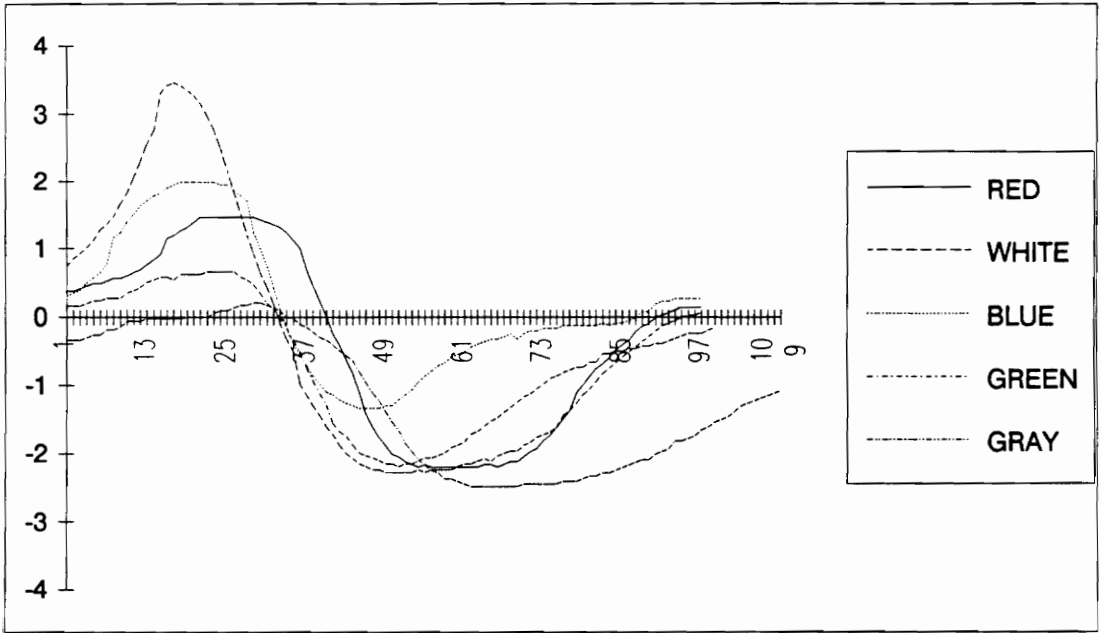


Figure 4.8 Similar character, different aspect ratio words

For speakers two and three, there were a total of five poor character words: LATHE, SAW, CUT, LIFT, and SIX for speaker number two, LATHE, SAW, GRAY, AHEAD, and SIX for speaker number three. In all these cases, the low dynamic range of the signal is the most likely cause. This could be due to poor emitter/detector positioning. The full vocabulary for speaker number one has good character definition.

A percentage score of the results is:

100% Speaker #1

91.8% Speaker #2

91.8% Speaker #3

2: **Repeatability**: This set of tests consists of three repeats per word. The overall signal level variations were not compensated for between the individual repeats. Despite this, repeatability was excellent for the overall character of the waveforms. The poor repeatability words are MIX , and WHITE for Speaker number one. MILL, LIFT, AHEAD, and LEFT for speaker number two. SAW, CARRY, ZERO, and SIX for speaker number three (see Appendix A). The percentage scores for this section are:

96.7% Speaker #1

93.4% Speaker #2

95% Speaker #3

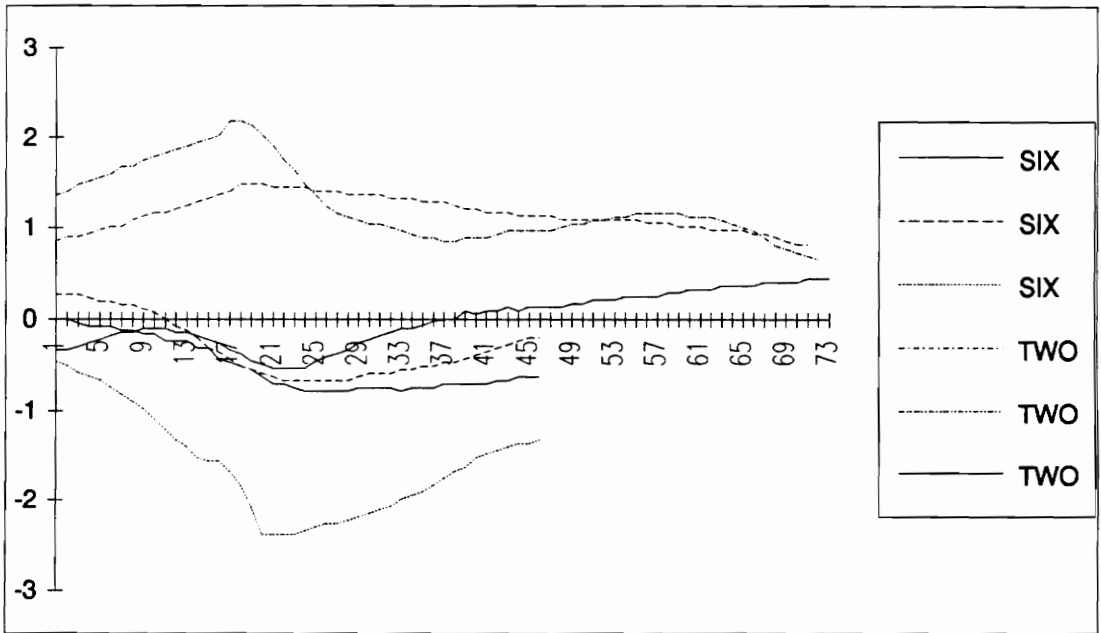


Figure 4.9 Example of poor repeatability for speaker #3

3: **Light Sensitivity**: The light intensity compensation capabilities of the device were evaluated by acquiring one full repeat of the vocabulary for speaker number one in bright ambient light with an approximate intensity of 1300 Foot Candle, normal office lighting being in the range of 100-130 Foot Candles. The results for this test are displayed along with the three repeat curves for speaker number one in Appendix A.

The overall character of the bright light curves are identical to the ones in ordinary lighting. However, the amplitude and the width of the peaks and valleys were slightly different. A slightly higher dynamic range is observed for most of the bright light curves as seen in Figure 4.10. Therefore, one deduces that the light compensation circuit is adequate for this application.

4: **Speaker Dependence or Independence**: All the repeat curves for all the speakers were merged for each word and displayed in Appendix A under the heading of All. No definite conclusion can be extracted from this test. Roughly 78.6% of the vocabulary words for the three speakers had close matching characteristics. The rest had medium to poor resemblance. The low signal level generated by some words along with the difference in the dynamic range of the speakers, are the most likely reason. Table 4.2 (page 45) lists the words that have good correlations and Table 4.3 (page 45) lists the suspect words.

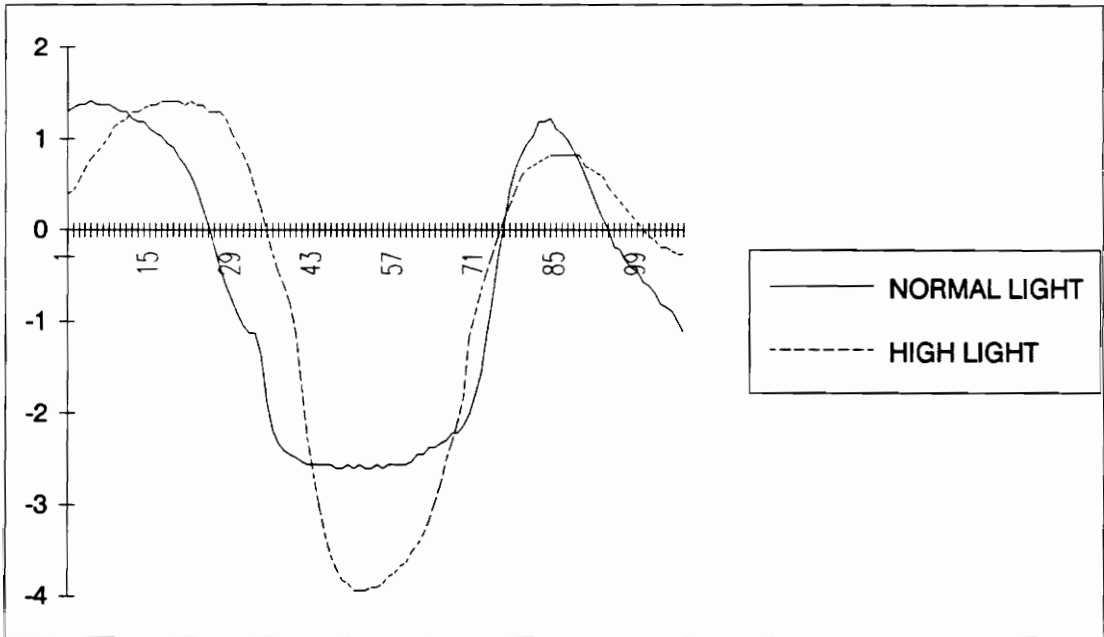


Figure 4.10 Example of light level change

4.3 SUMMARY OF RESULTS

The test results show that for the limited set of vocabulary (Table 1, page 44), the proposed method of automatic lip reading generates recognizable waveform character 98% of the time. These waveform characters can be interpreted and correlated to speech. Furthermore, these patterns are shown to be repeatable even among three diverse speakers at an average of 95% of time. The system functions equally under normal lighting or bright light conditions for 100% of the vocabulary. The issue of speaker independence or dependance was not conclusive. The 76.6% score for this category is below the usual 95% success ratio expected by experts for automatic Voice Recognition Systems.

Although the processing, electronics, and optics of this system are extremely limited, the results of the tests show great promise for future development.

TABLE 4.1 VOCABULARY TEST SET

MAIM NAME MAIN DAME SAME

LATHE MILL SAW DRILL ROBOT

CUT WELD PAINT FORGE PIERCE

COPY PRINT TYPE WRITE READ

MIX CARRY TAPE LIFT DRIVE

ENGINE IGNITION SWITCH VALVE SENSOR

RED WHITE BLUE GREEN GRAY

STOP AHEAD BEHIND RIGHT LEFT

ARMUP ARMDOWN WRISTUP WRISTDOWN ROTATE

PINCH CREEP SLOW FAST MEDIUM

0 1 2 3 4 5 6 7 8 9 10

TABLE 4.2 POOR CORRELATION WORDS (BETWEEN SPEAKERS)

LATHE SAW CUT FORGE CARRY
IGNITION AHEAD LEFT ROTATE ZERO
TWO SIX NINE TEN EIGHT

TABLE 4.3 GOOD CORRELATION WORDS (BETWEEN SPEAKERS)

MAIM NAME MAIN DAME SAME
MILL DRILL ROBOT WELD PAINT
PIERCE COPY PRINT TYPE WRITE
READ MIX TAPE LIFT DRIVE ENGINE
SWITCH VALVE SENSOR RED WHITE
BLUE GREEN GRAY STOP BEHIND
RIGHT ARMUP ARMDOWN WRISTUP
WRISTDOWN PINCH CREEP SLOW
FAST MEDIUM ONE THREE FOUR
FIVE SEVEN

5.0 CONCLUSIONS

Visual speech information fused with the acoustic speech information has been shown to improve automatic speech recognition in noisy environments. The major problems encountered with acquiring and fusing the visual images with the acoustic signals is the amount and format of the visual data. This requires neural networks and large amounts of processing in addition to the video equipment needed for image acquisition, thereby making the process costly and slow.

The results of the pilot study of the proposed alternate automatic lip reading method has shown it to be promising and warrants further development. As noted earlier, this alternate method has the advantage of low cost, and low amounts of data and processing requirements. The format of data allows for simpler, and faster fusing with the acoustic signal, thereby making it feasible to use real time.

5.1 FUTURE DEVELOPMENTS

Several improvements of the current experimental system are required for further development. The improvements are:

1: **Electronics**: An automatic gain system to produce the same signal base level for the various speakers will be needed. The resolution of the lower level optical signals can be improved by replacing the current 8 bit A/D convertor with a 12 bit or more convertor.

2: **Optics**: Two or three optical sensor arrangements needs to be investigated to allow for better signal stability and acquisition of the visual information. A more robust method and/or device for positioning the emitter/detector pair is needed to improve repeatability and speaker

independence.

3: **Signal processing:** High sample rate acoustic trigger is needed for better optical word isolation. Dynamic time wrapping is needed to smooth the temporal differences between various repeats and speakers.

5.2 APPLICATIONS

Although this method may not become a stand alone Speech Recognition System, it offers many potential benefits for aiding Automatic Speech Recognizers in noisy environments. The potential for real time application with a second source of information redundancy makes this method attractive for application in crucial environments, such as in some military and manufacturing applications. In addition, it offers a fast and low cost second source of clues about speech that can be used to resolve conflicts and speed up recognition in normal environments.

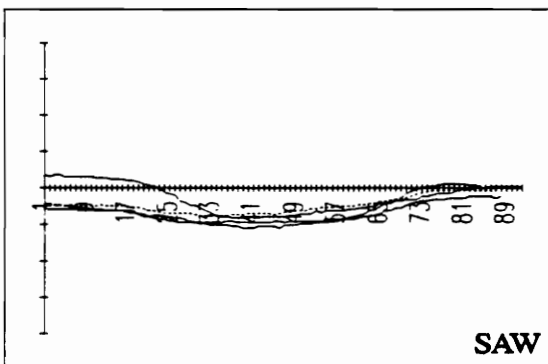
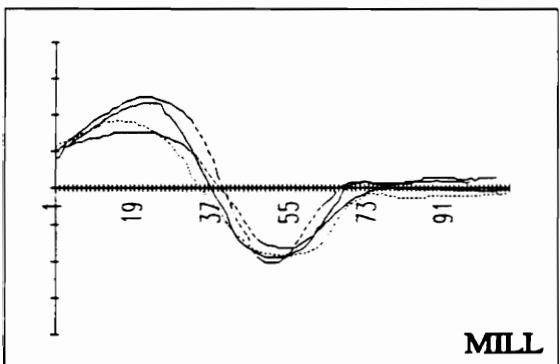
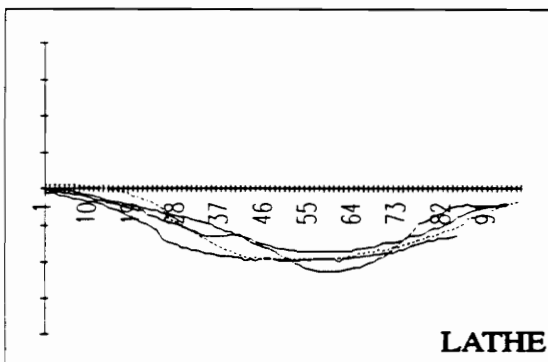
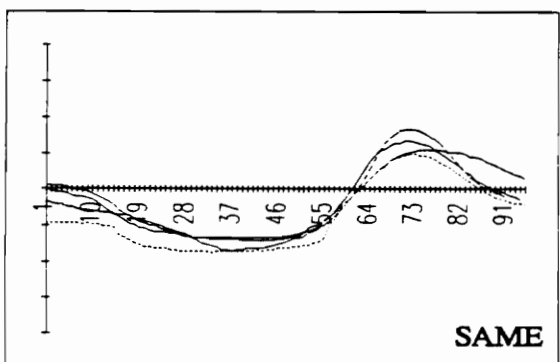
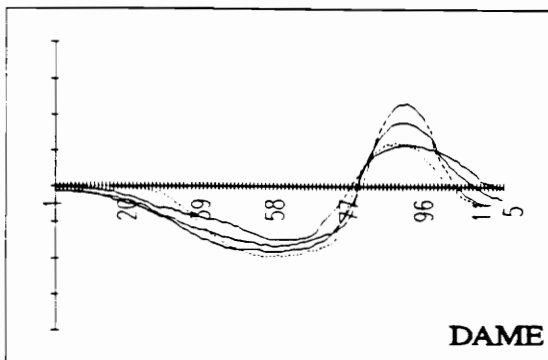
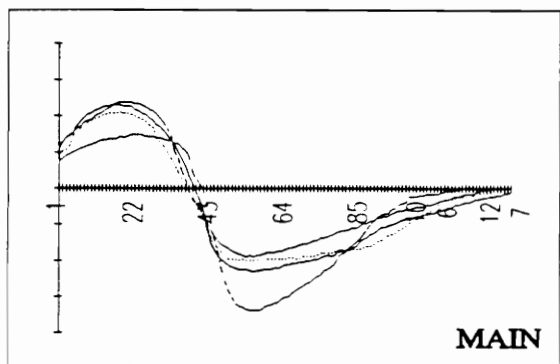
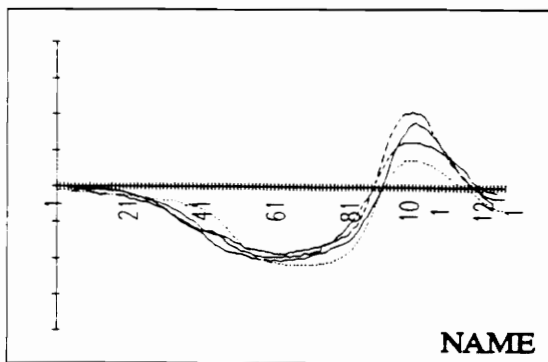
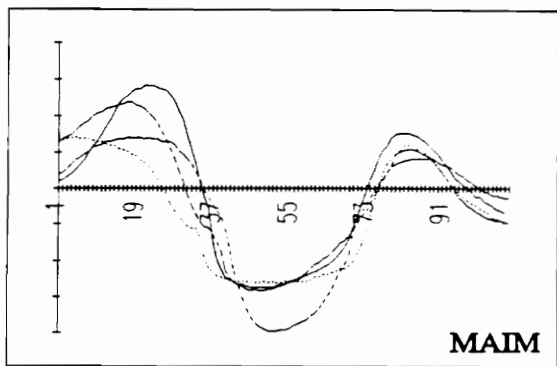
6.0 REFERENCES

1. Scott, B. I. "Voice Recognition Systems and Strategies." Computer Design. Vol. 22. No. 1. Jan 1983: 67-70.
2. Baker, J. K. "A Second-Generation Large Vocabulary System." Speech Technology. Vol. 4. No. 4. April-May 1989: 20-24.
3. Church, K. W. Phonological Parsing in Speech Recognition. Boston: Kluwer, 1987.
4. Yuhas, B. P. "Integration of Acoustic and Visual Speech Signals Using Neural Networks." IEEE Communications Magazine. Nov. 1989: 65-71.
5. Hesselmann, N. L. "Structural Analysis of Lip-Contours for Isolated Spoken Vowels Using Fourier Descriptors." Speech Communication. Vol. 2. 1983: 327-340.
6. Petajan, E. D. "Automatic lip reading to Enhance Speech Recognition." IEEE ComSoc Global Telecom. Conf. 26-29 Nov. 1984: 265-272.
7. Petajan, E. D. "An Improved Automatic Lip reading System to Enhance Speech Recognition." Bell Labs. Tech. Report No. 11251-871012-111TM. Nov. 1987.

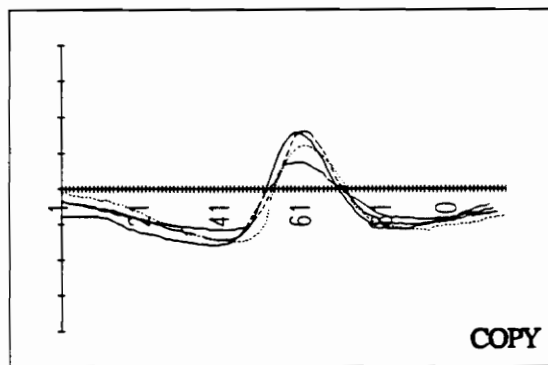
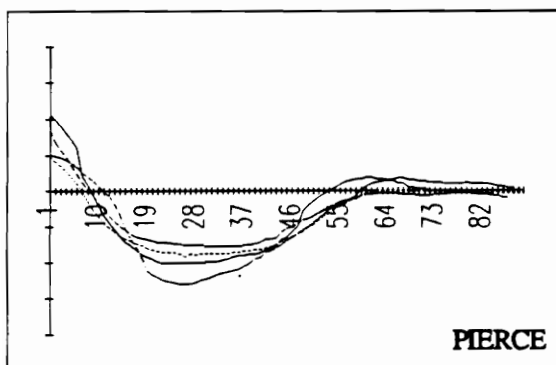
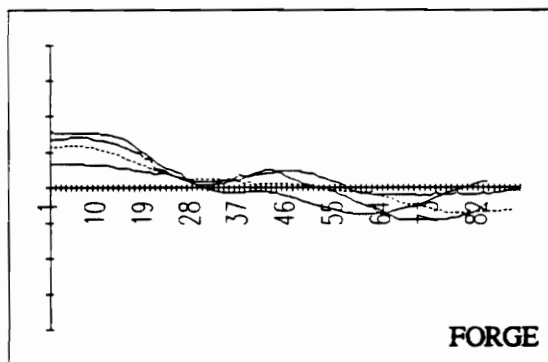
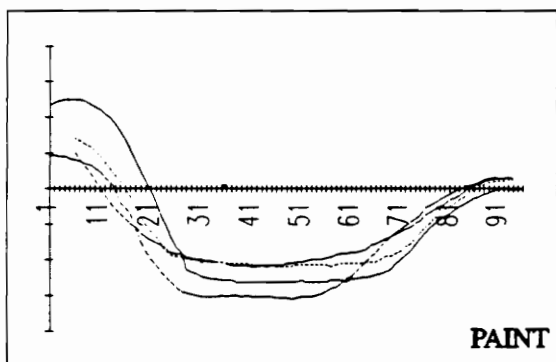
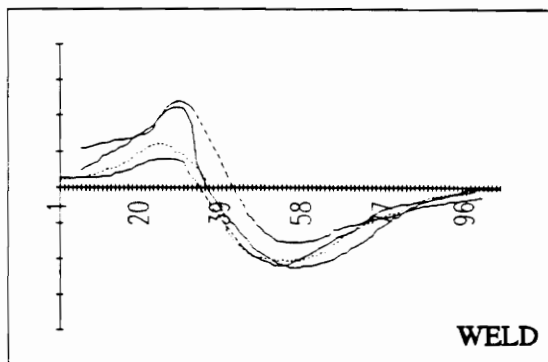
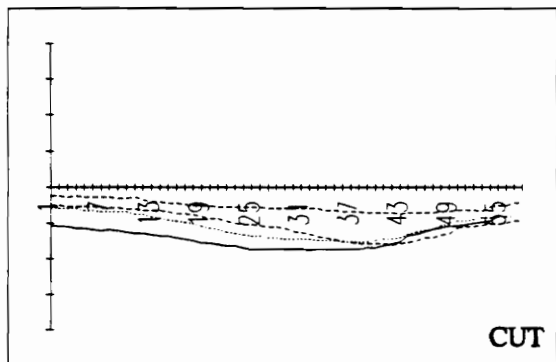
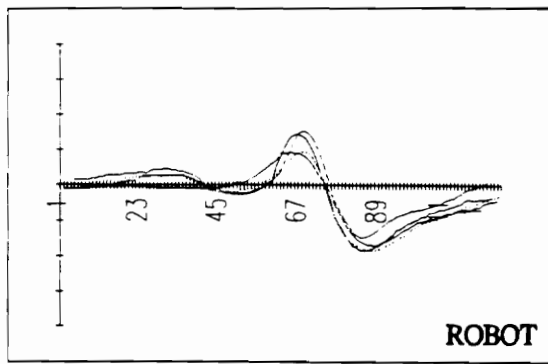
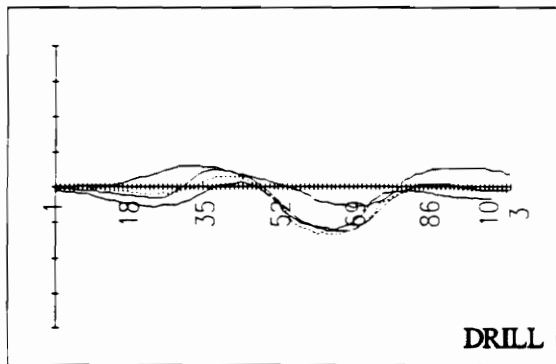
7.0 APPENDICES

APPENDIX A TEST RESULTS

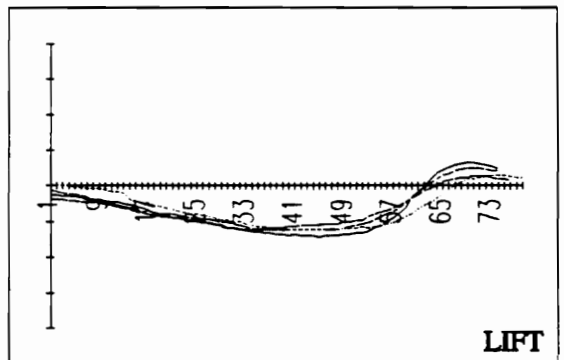
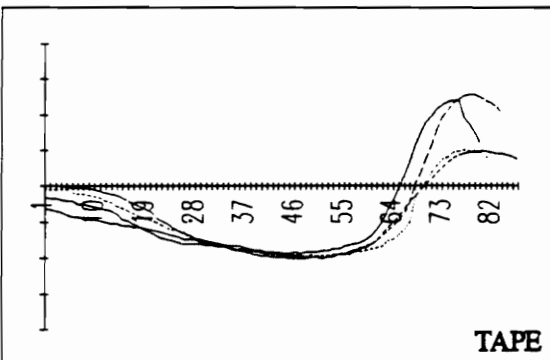
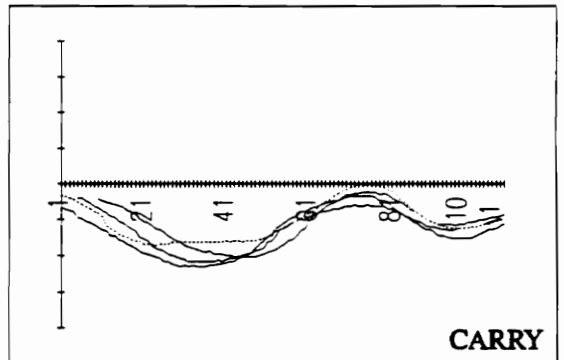
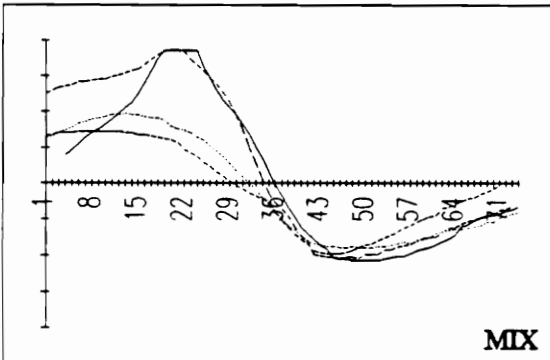
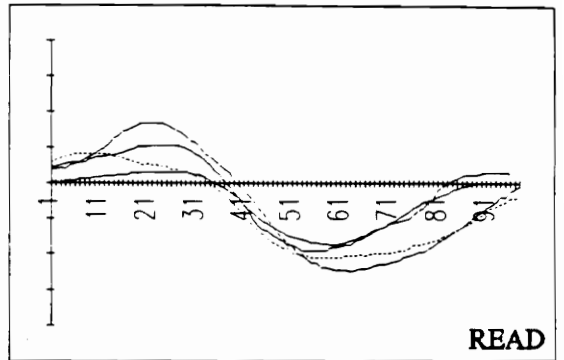
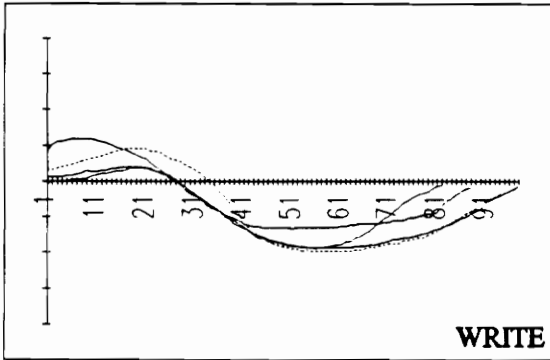
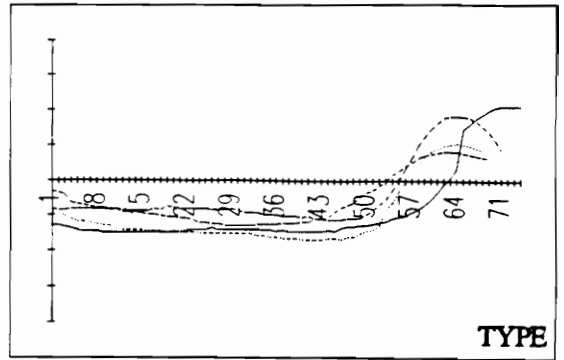
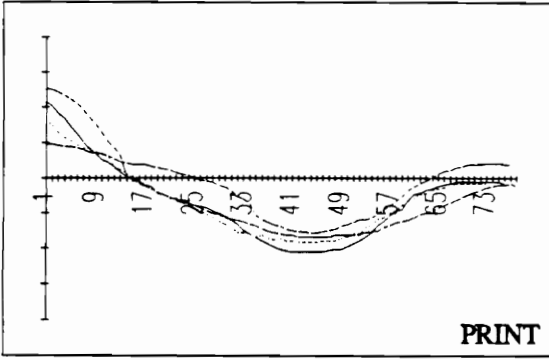
SPEAKER #1



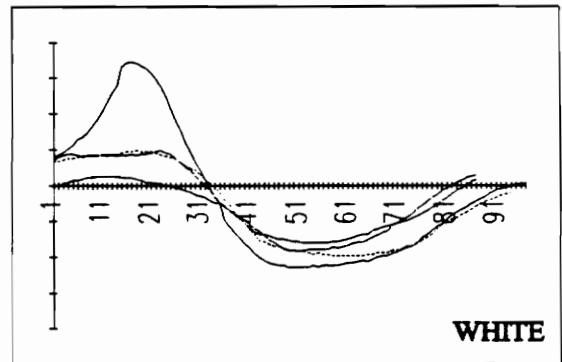
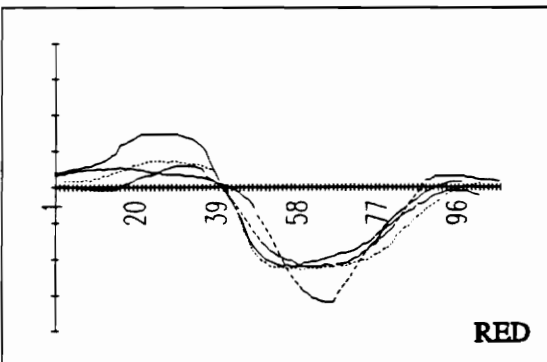
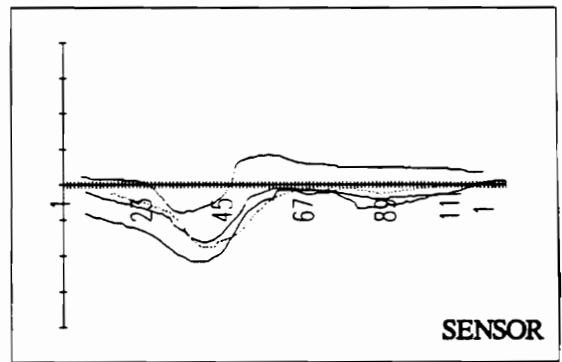
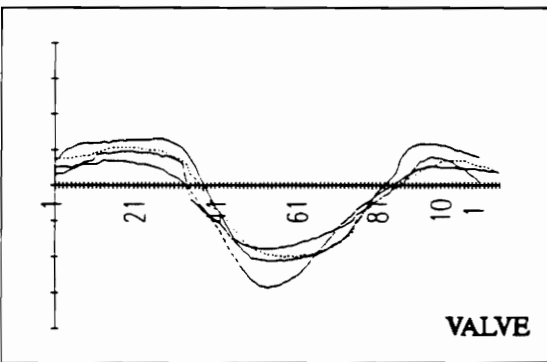
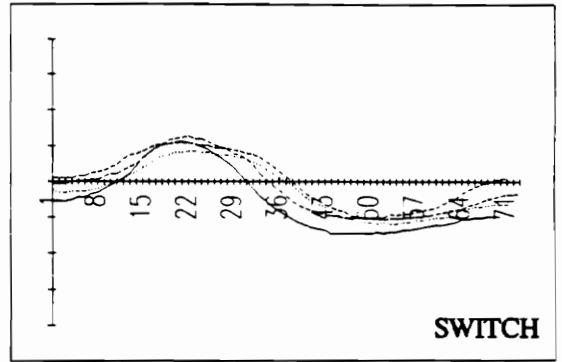
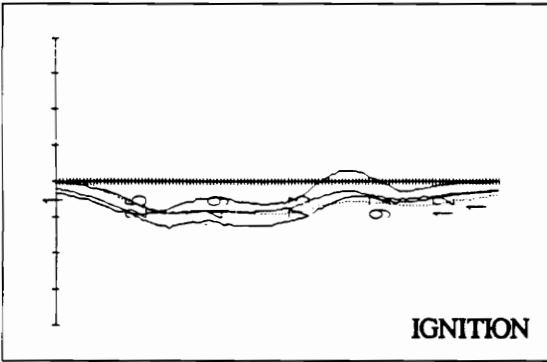
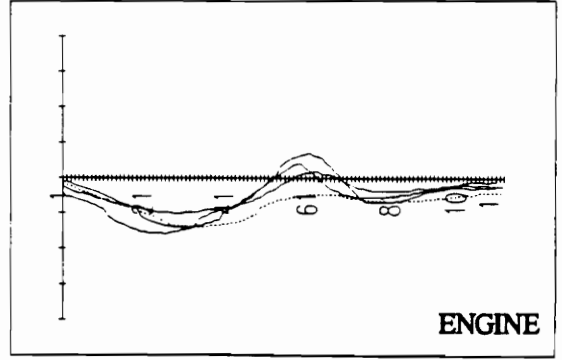
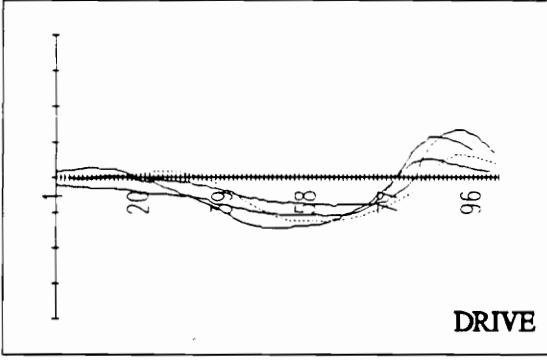
SPEAKER #1



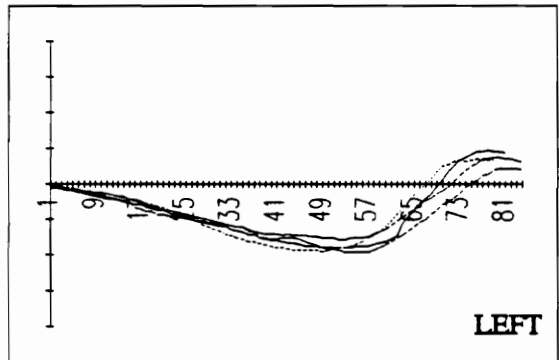
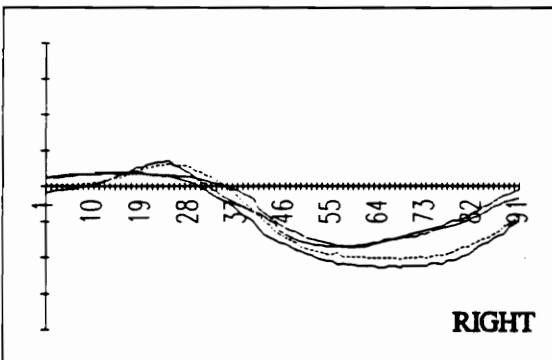
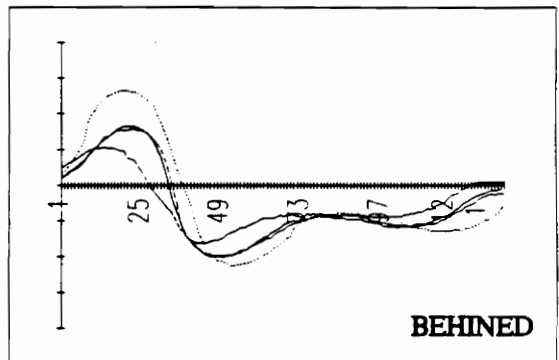
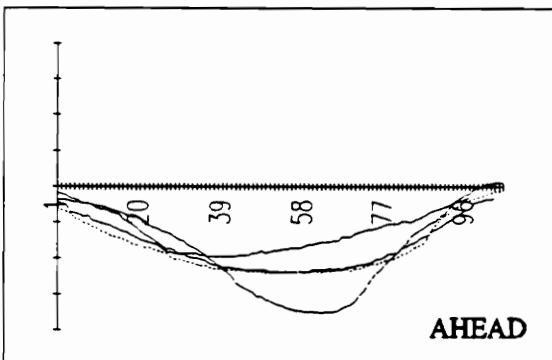
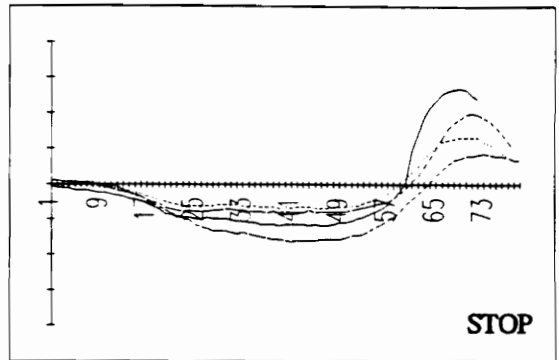
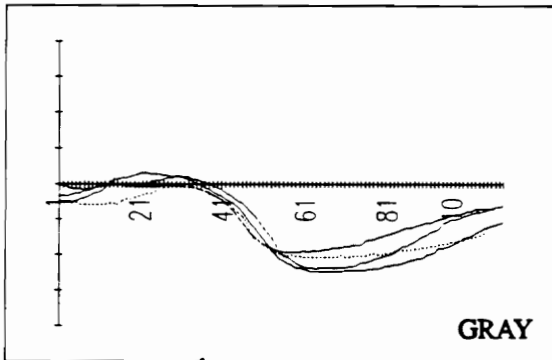
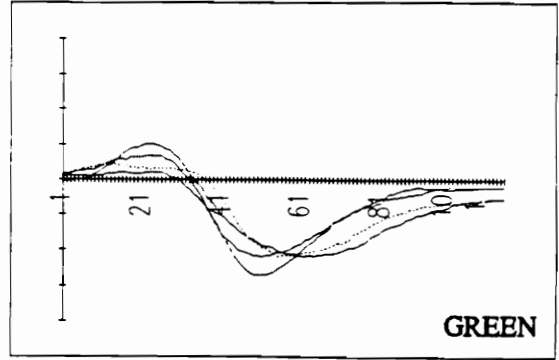
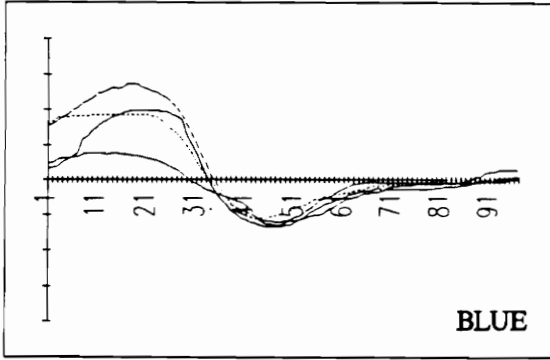
SPEAKER #1



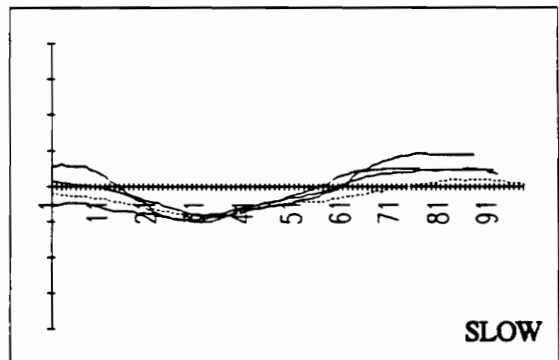
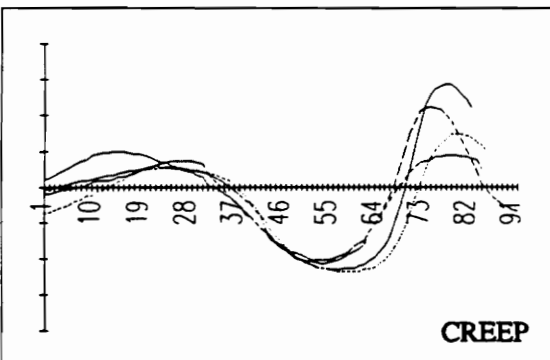
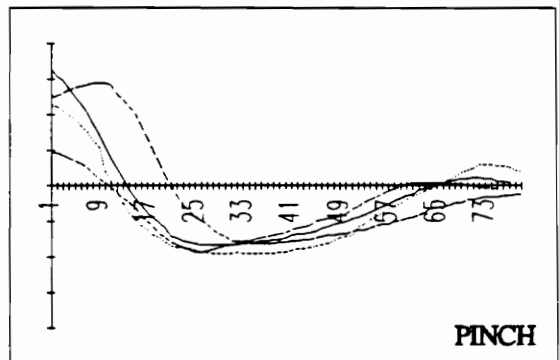
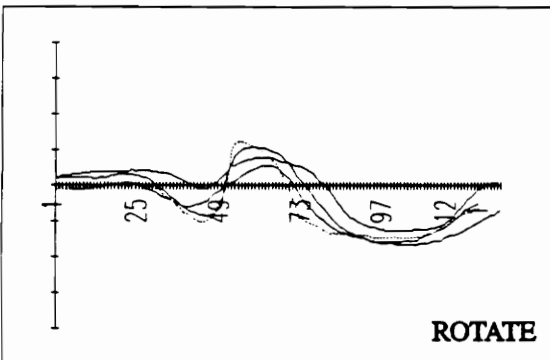
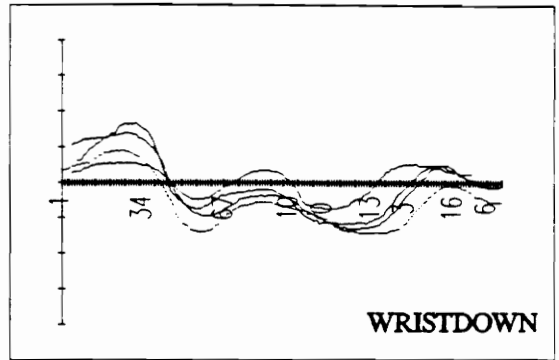
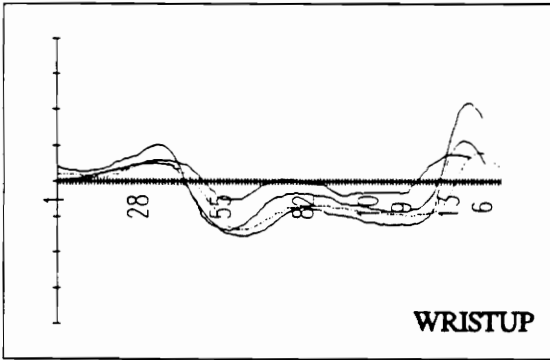
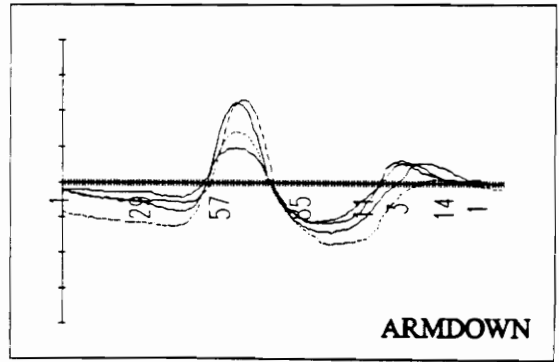
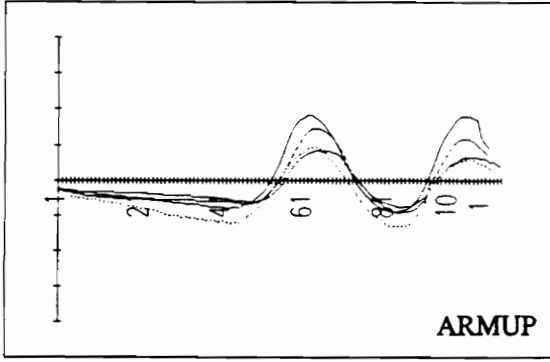
SPEAKER #1



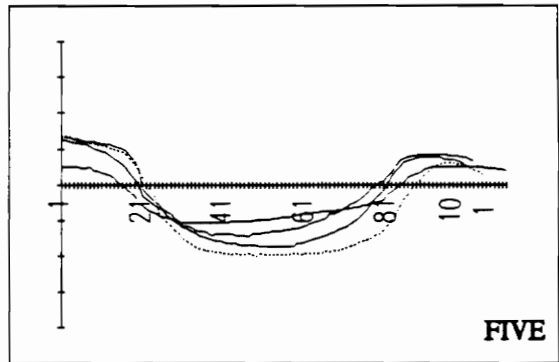
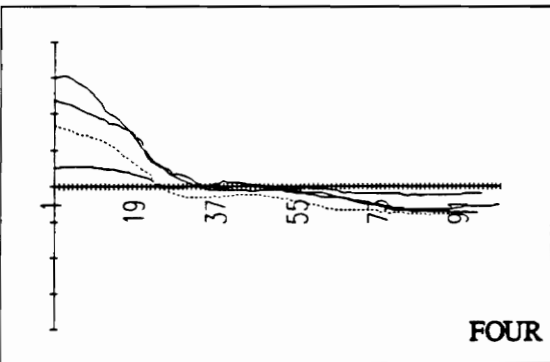
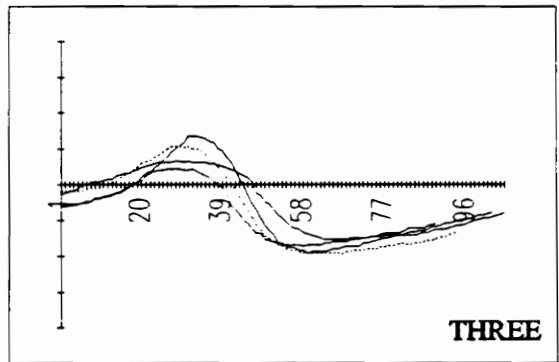
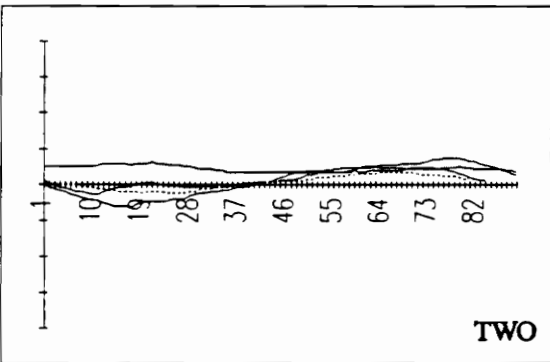
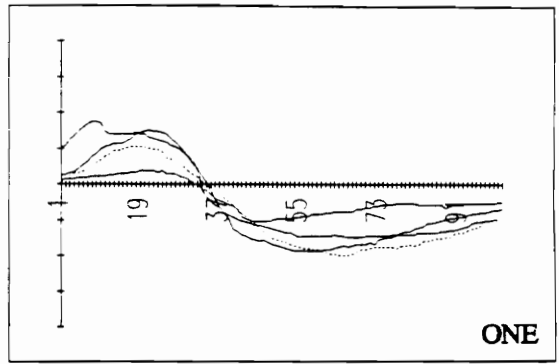
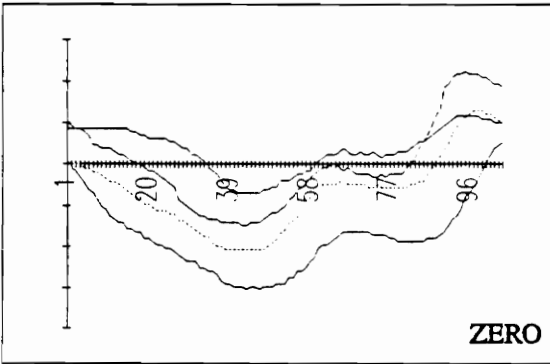
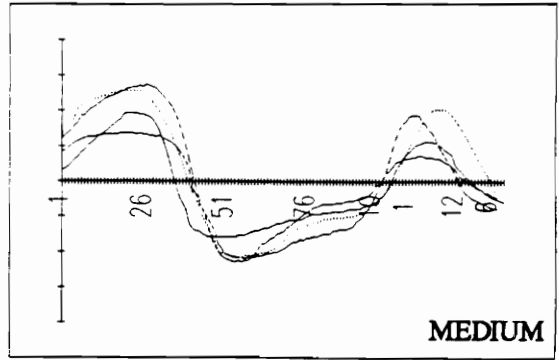
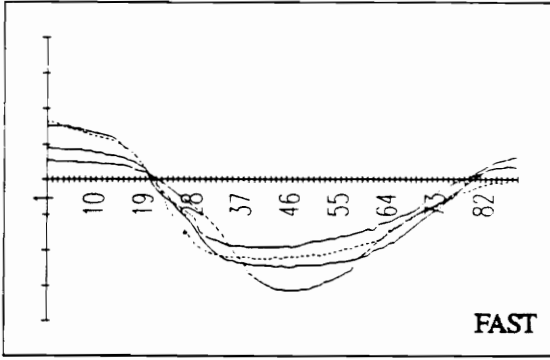
SPEAKER #1



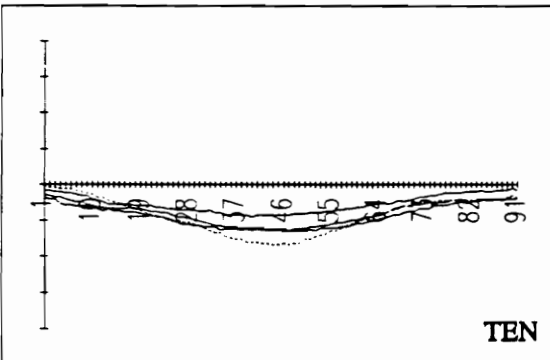
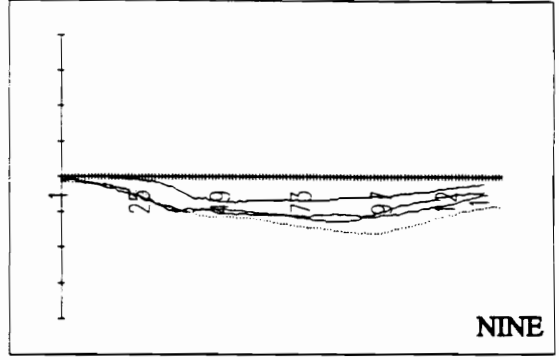
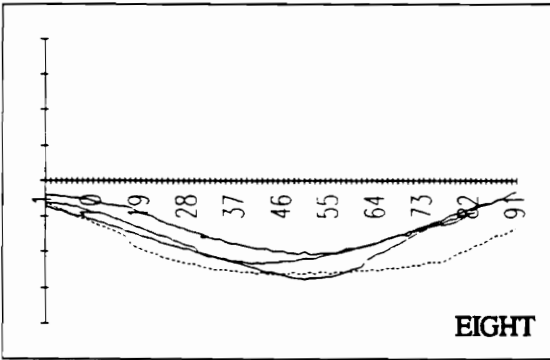
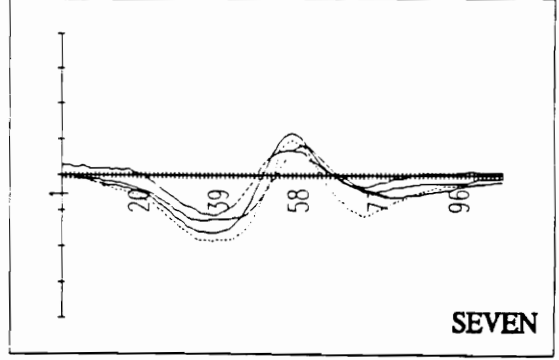
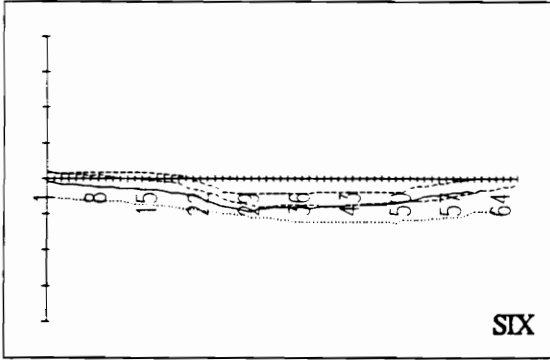
SPEAKER #1



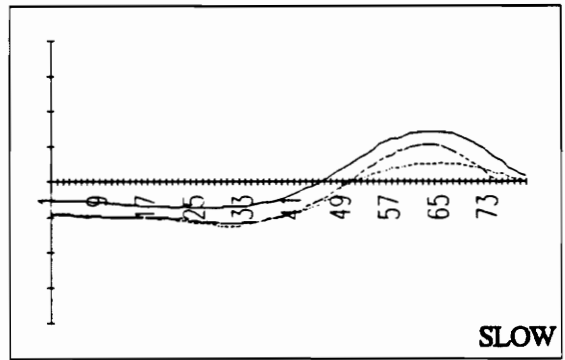
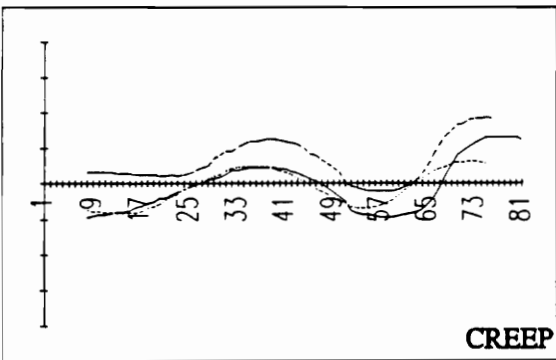
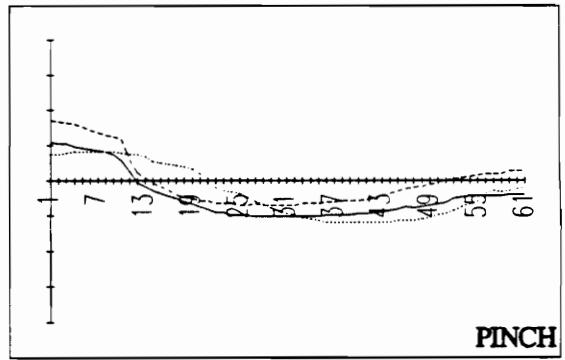
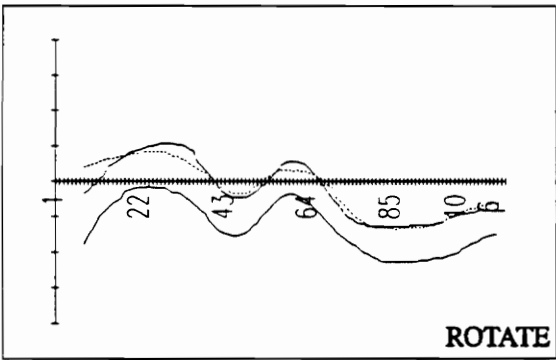
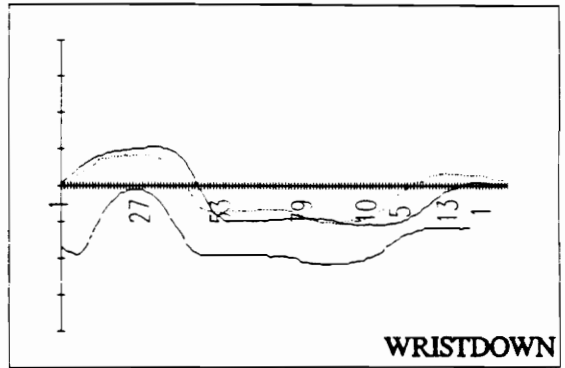
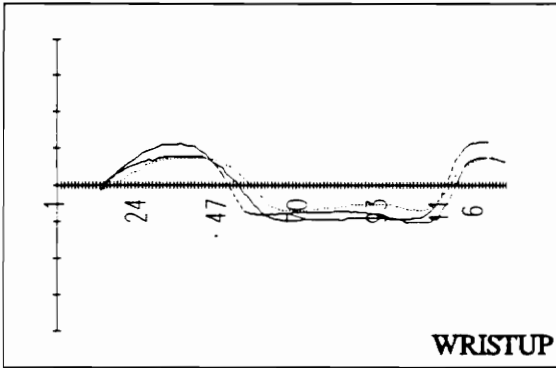
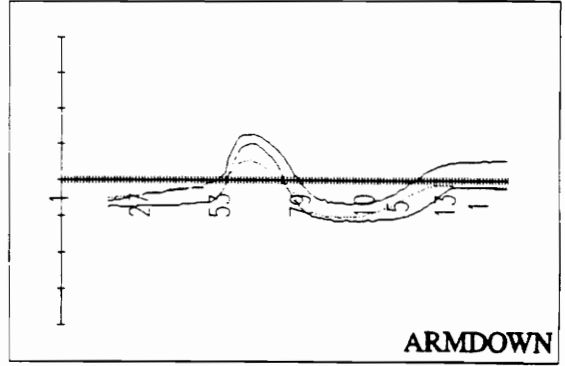
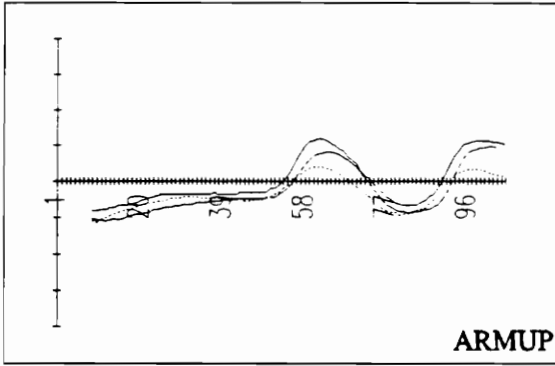
SPEAKER #1



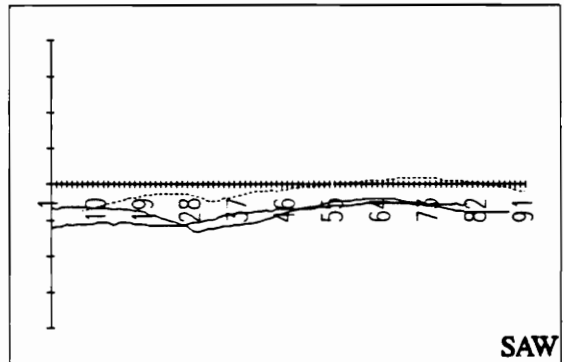
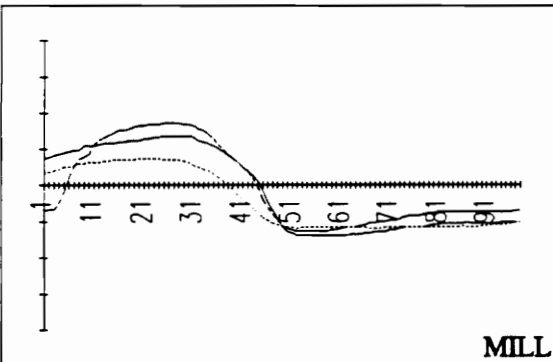
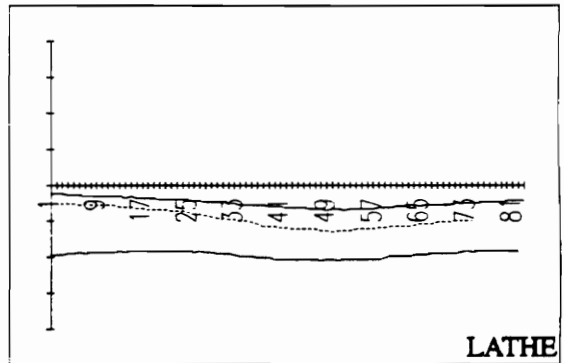
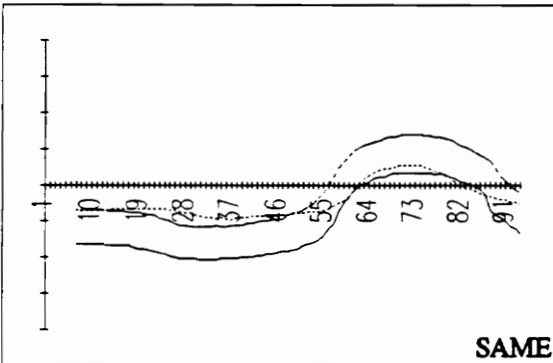
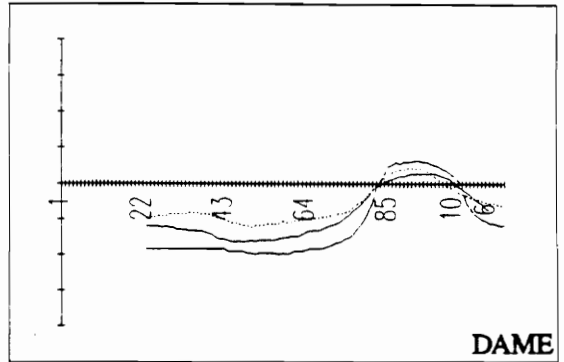
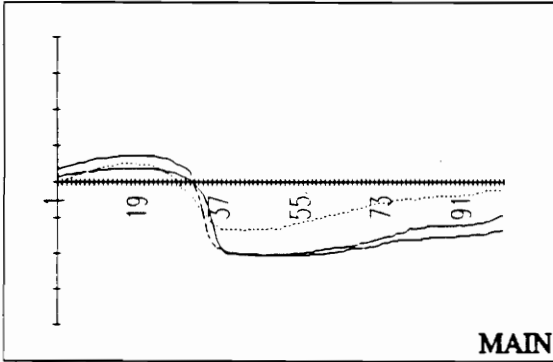
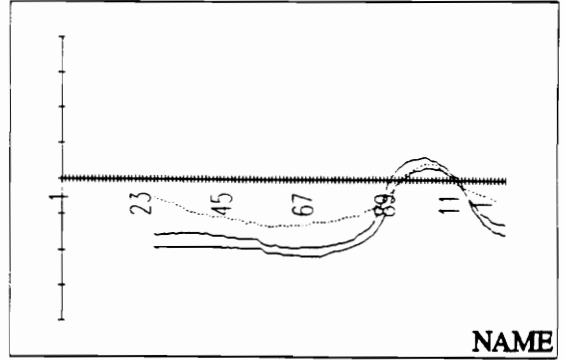
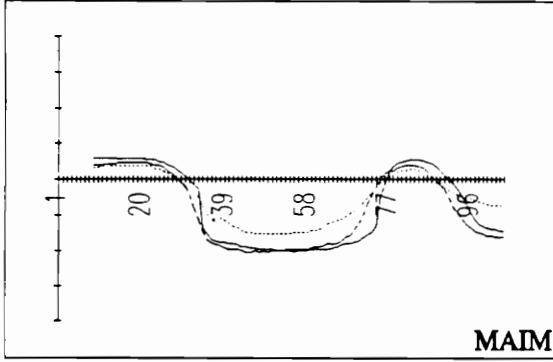
SPEAKER #1



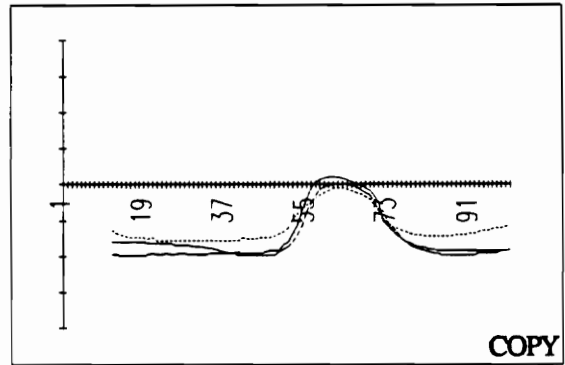
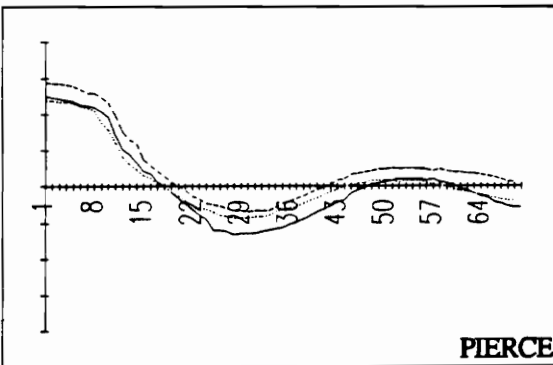
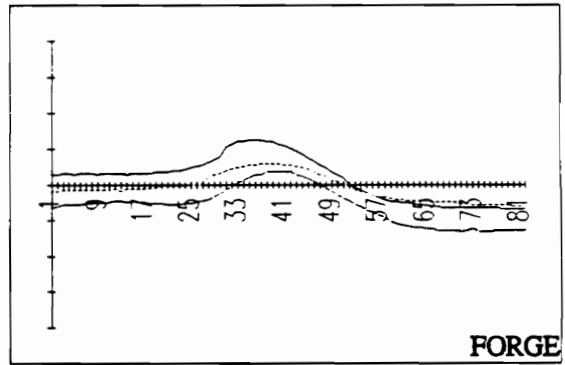
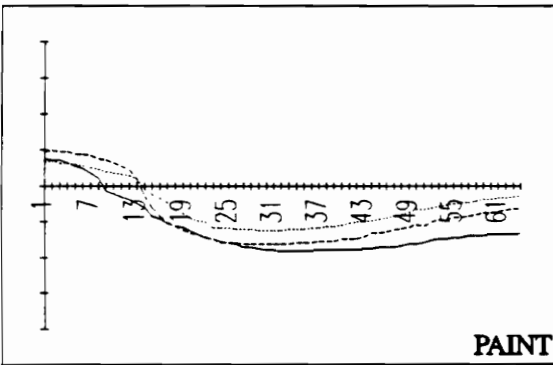
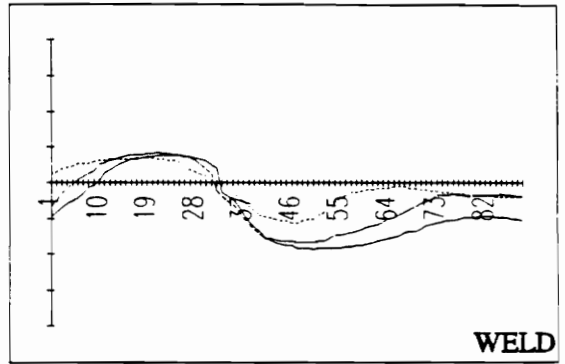
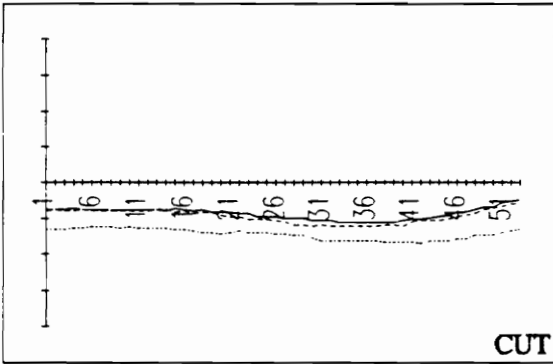
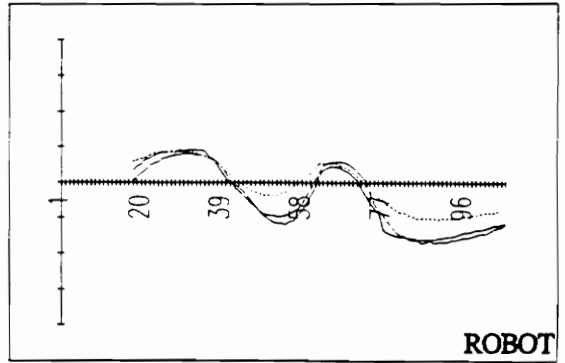
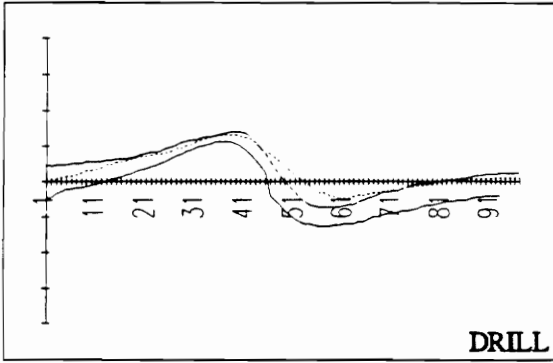
SPEAKER #2



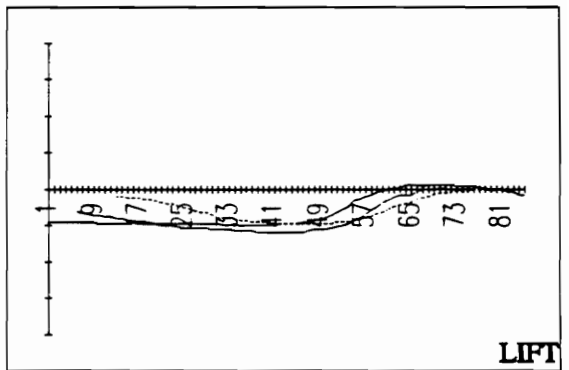
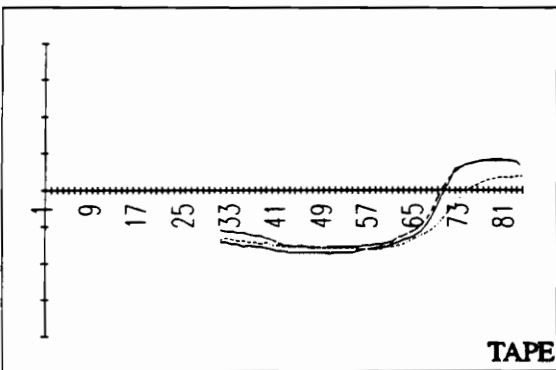
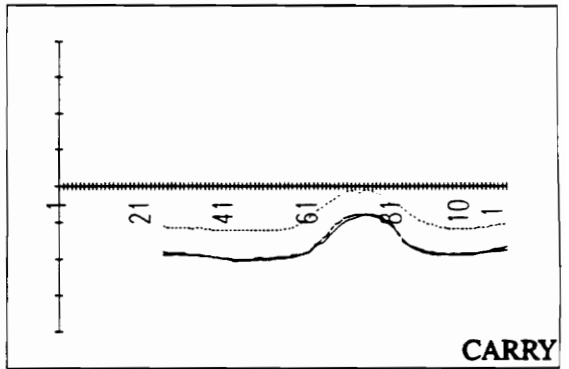
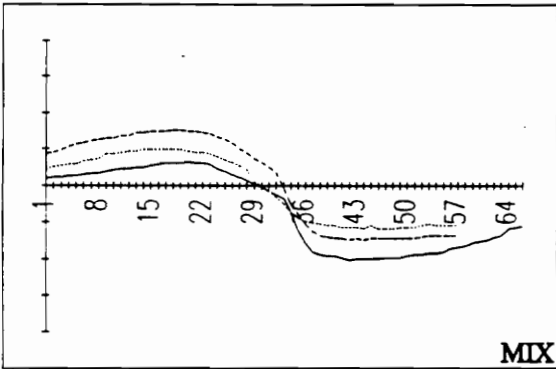
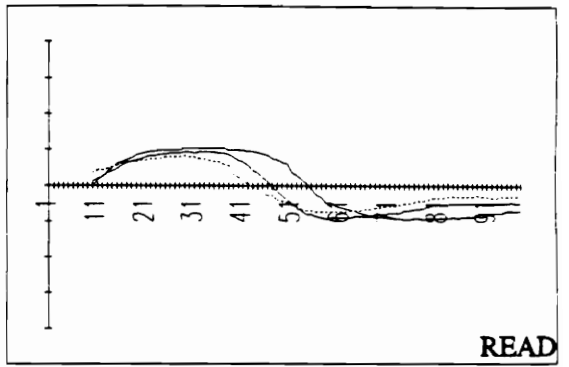
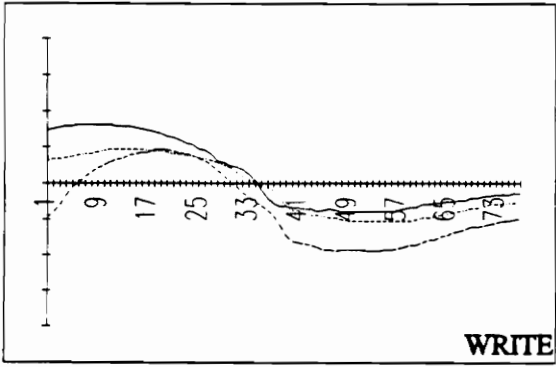
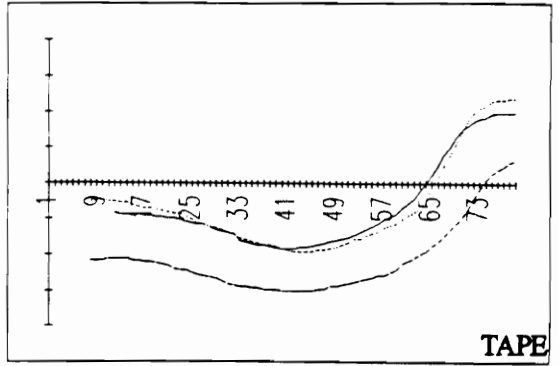
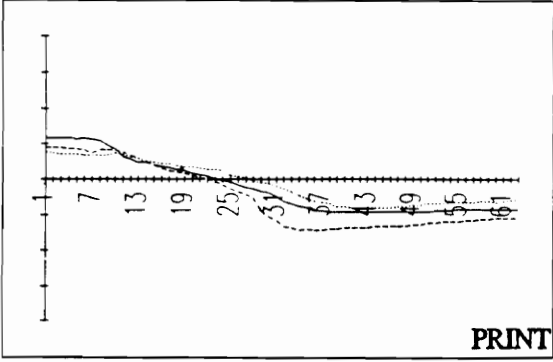
SPEAKER #2



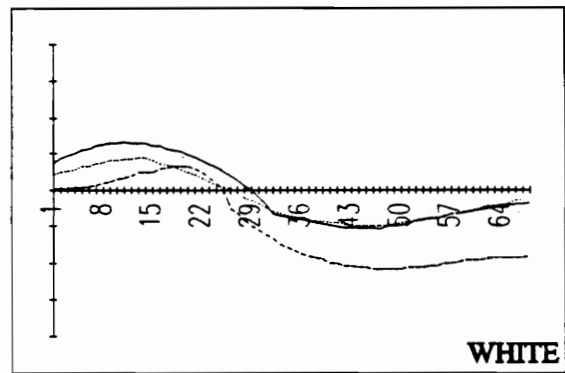
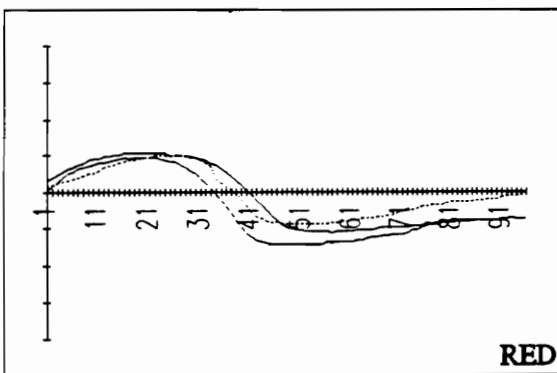
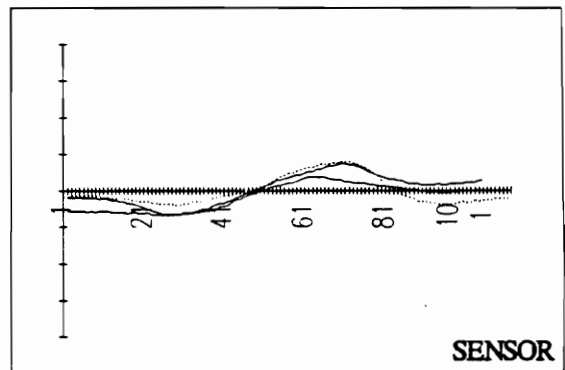
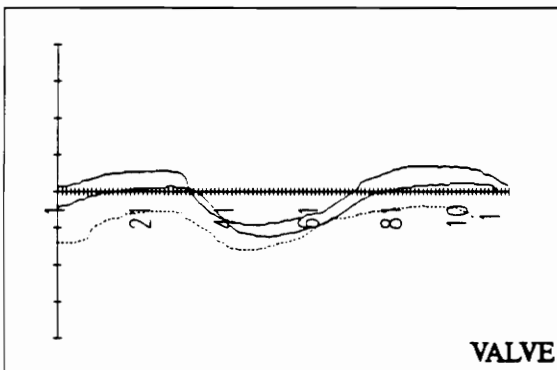
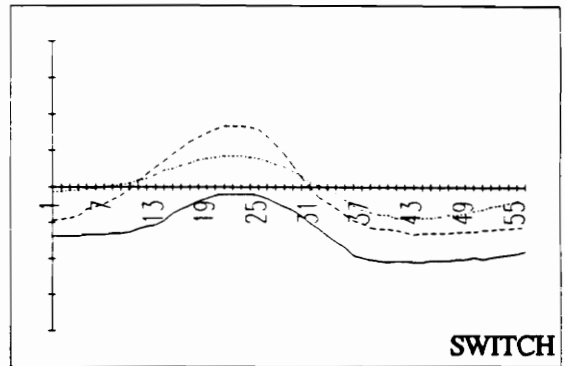
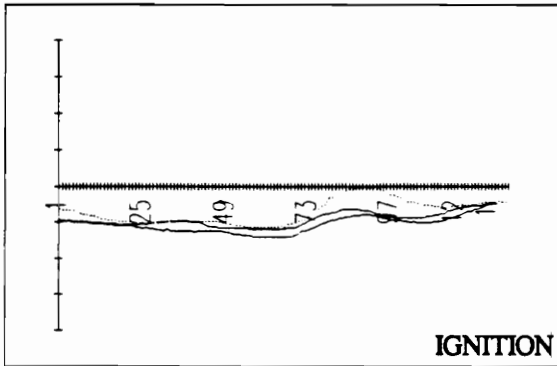
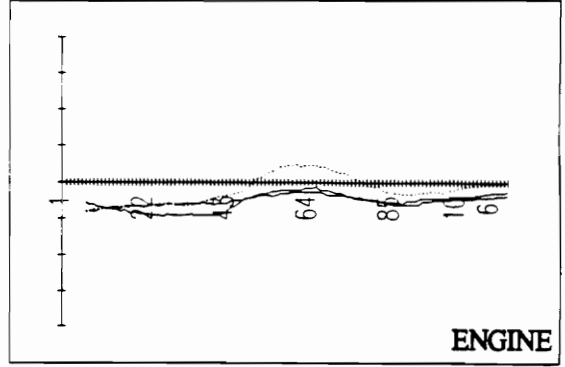
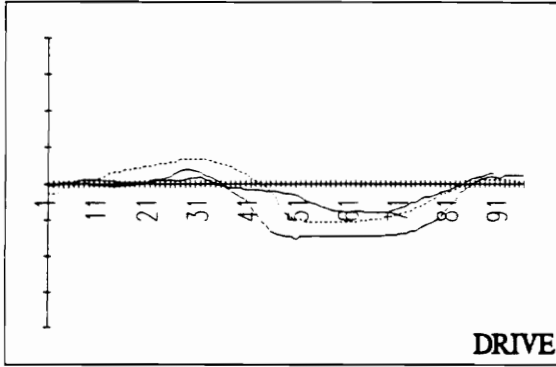
SPEAKER #2



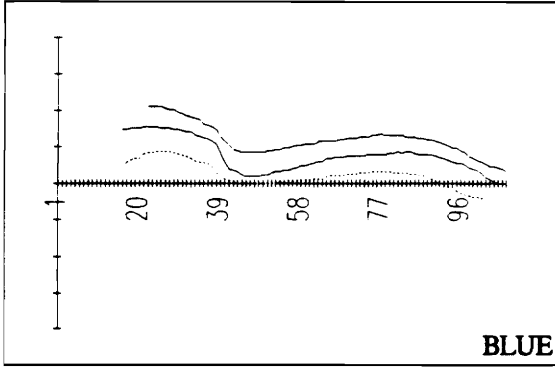
SPEAKER #2



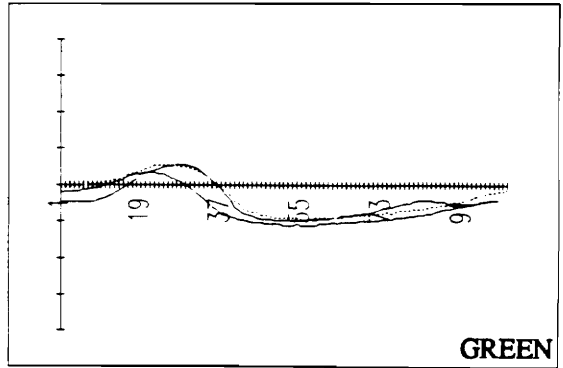
SPEAKER #2



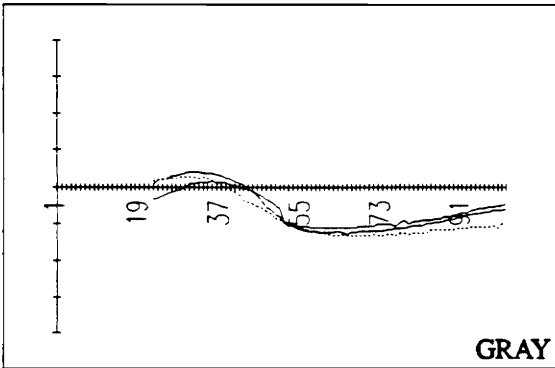
SPEAKER #2



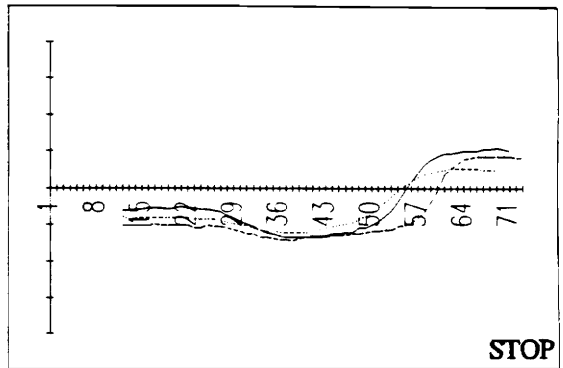
BLUE



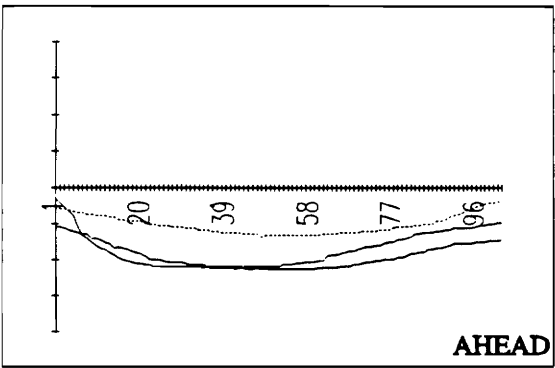
GREEN



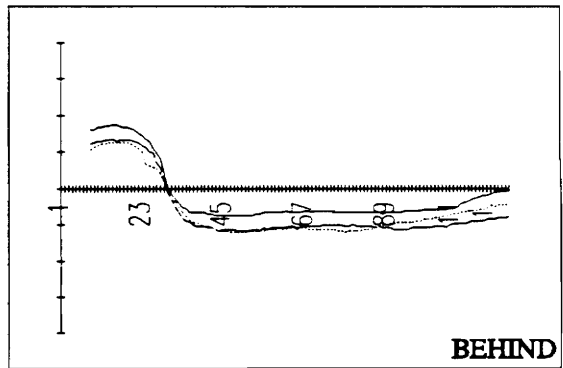
GRAY



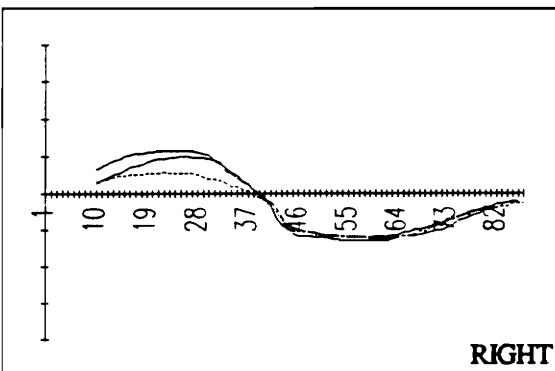
STOP



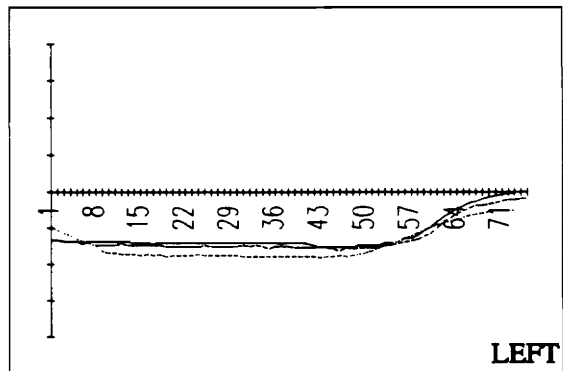
AHEAD



BEHIND

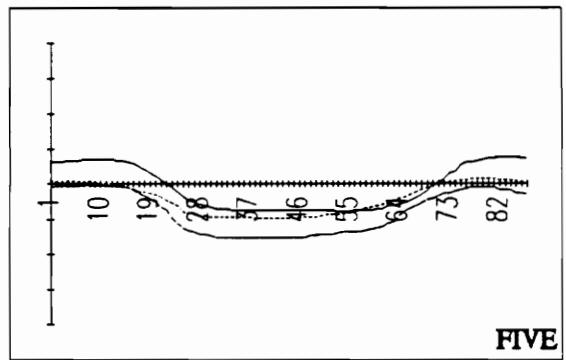
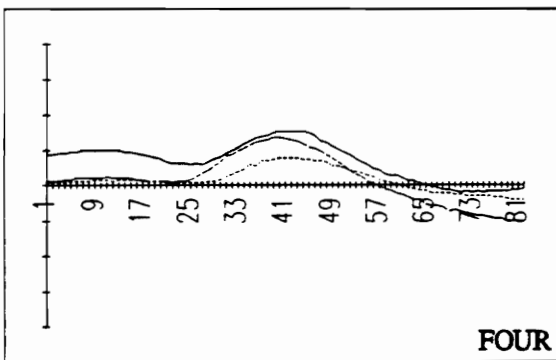
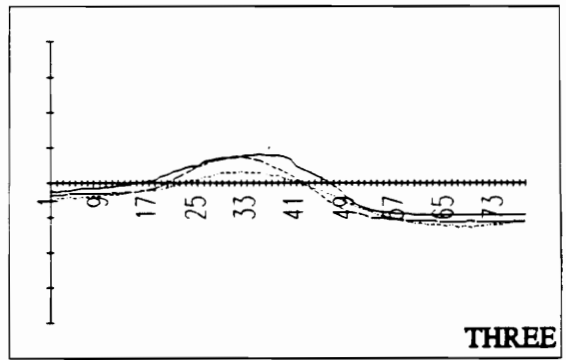
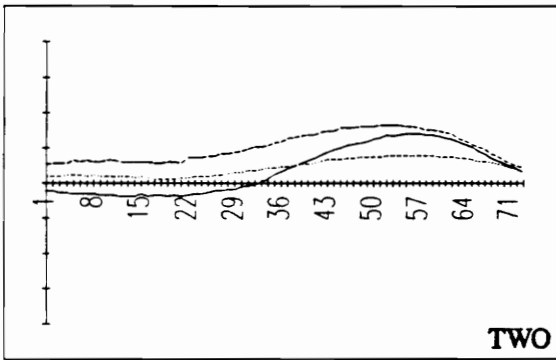
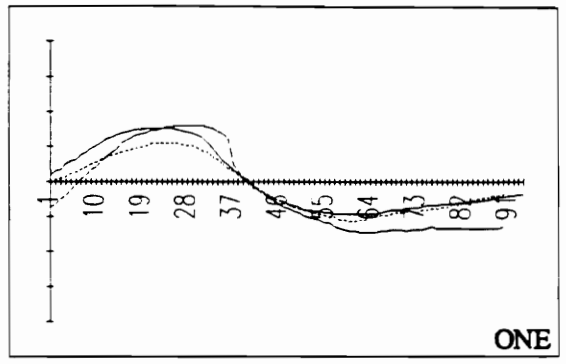
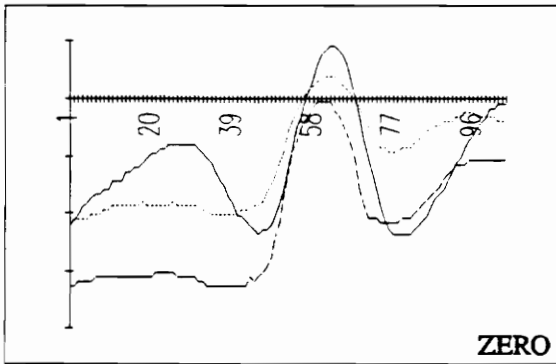
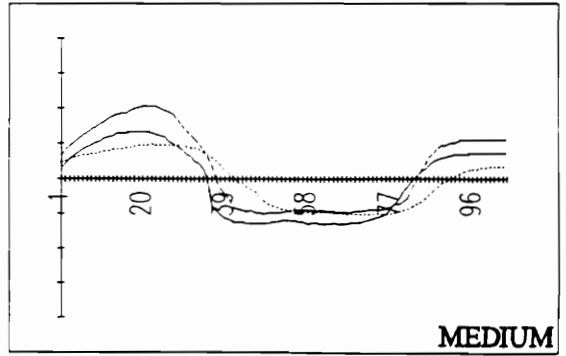
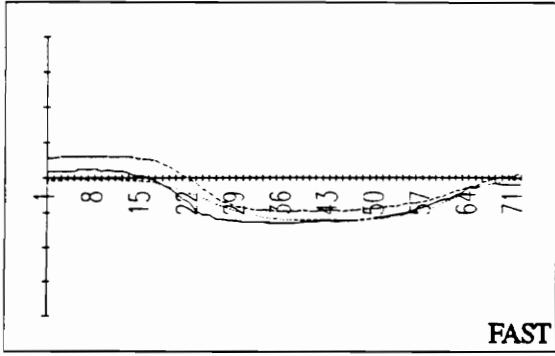


RIGHT

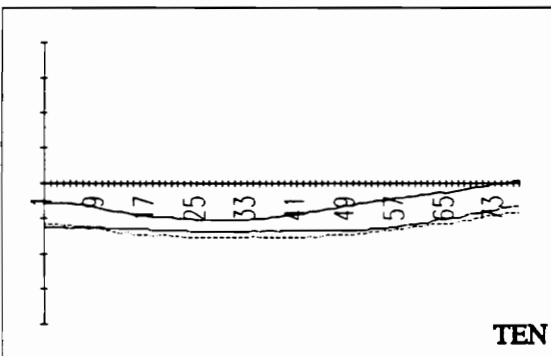
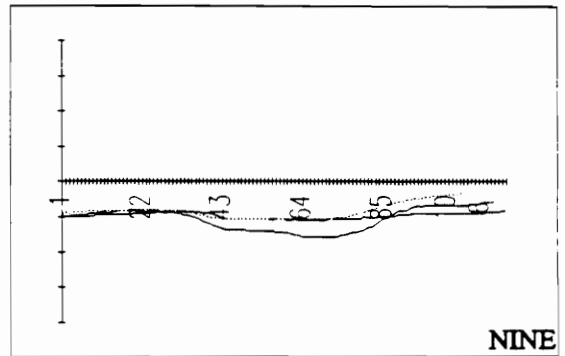
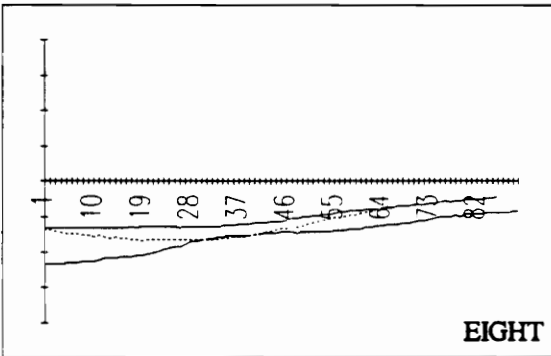
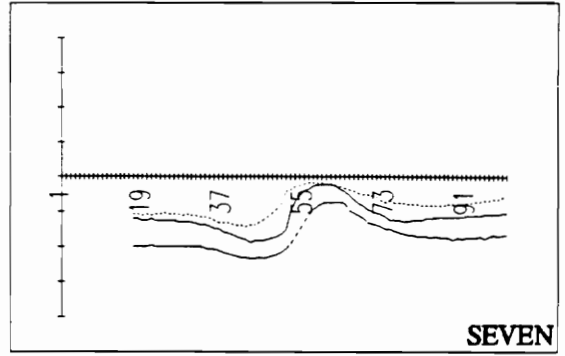
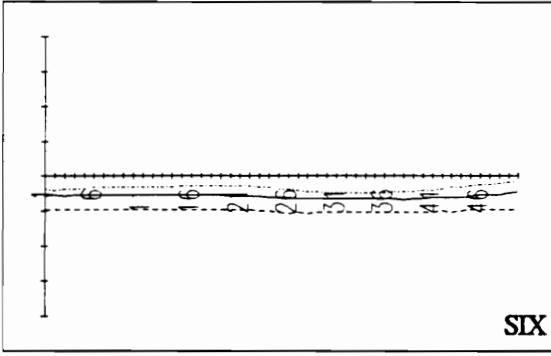


LEFT

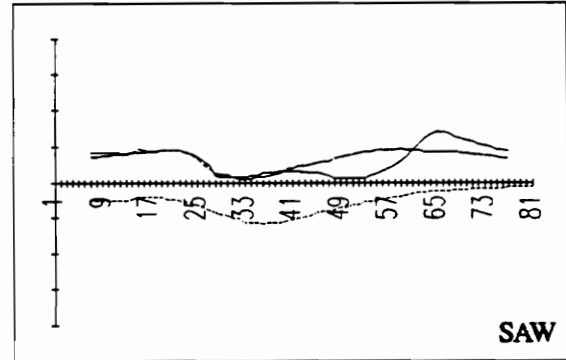
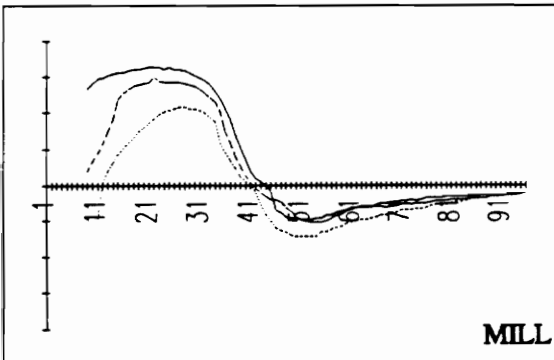
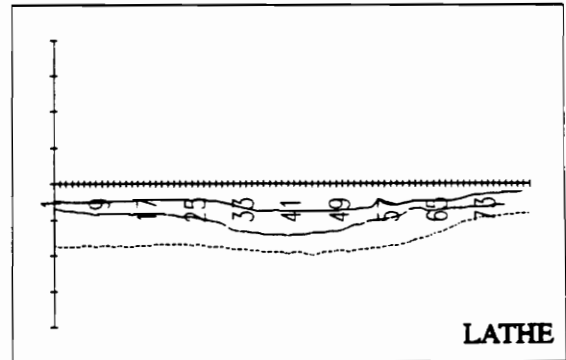
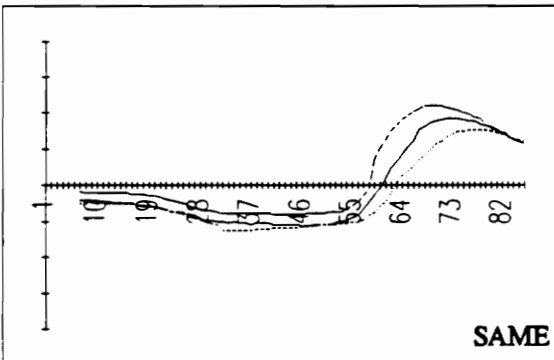
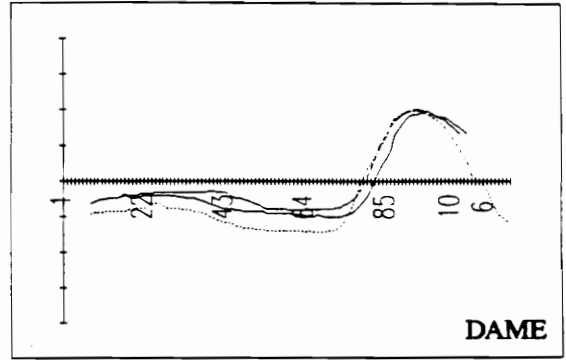
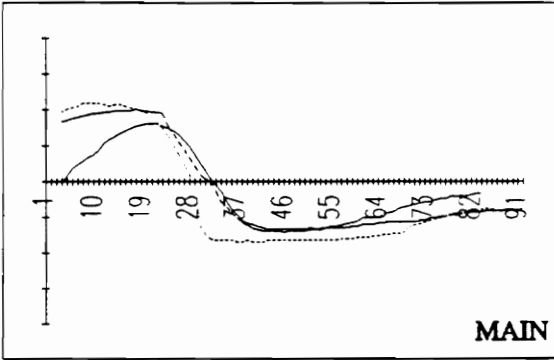
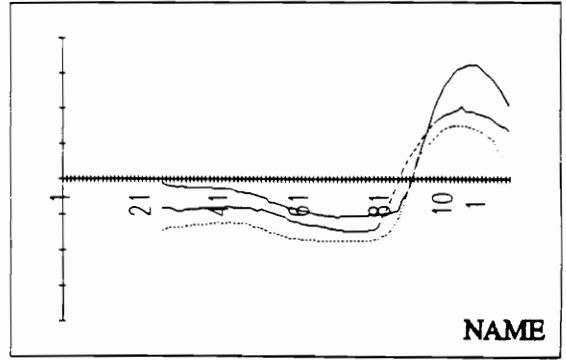
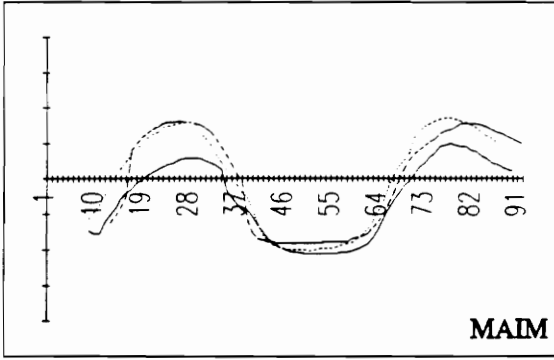
SPEAKER #2



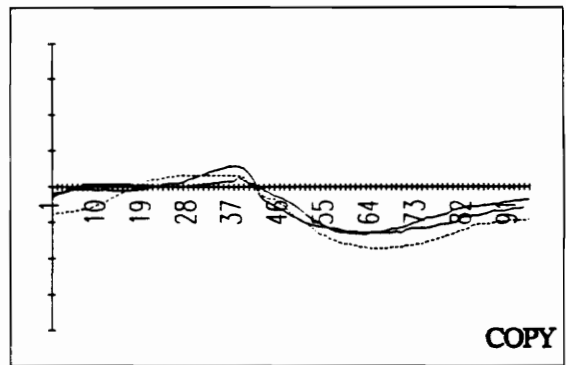
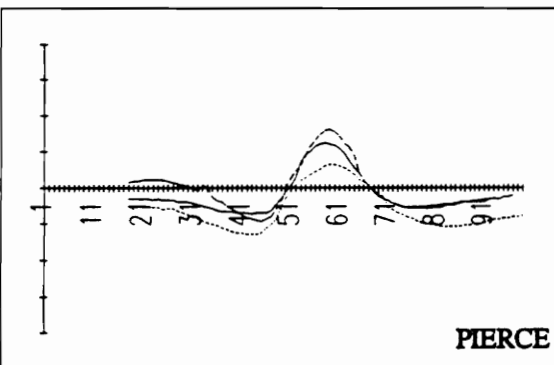
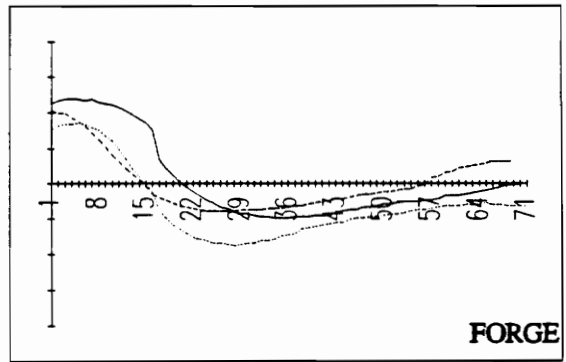
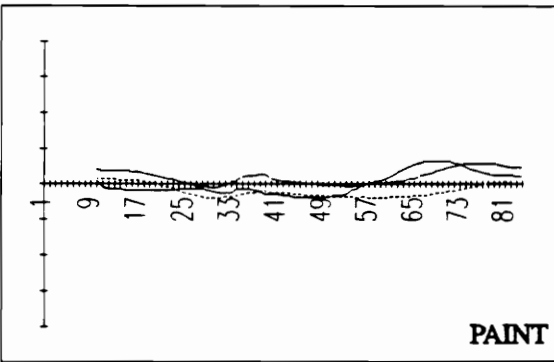
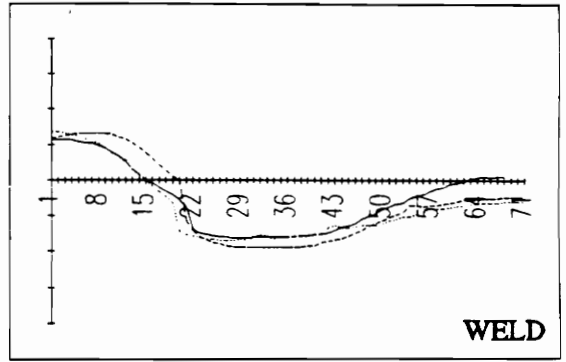
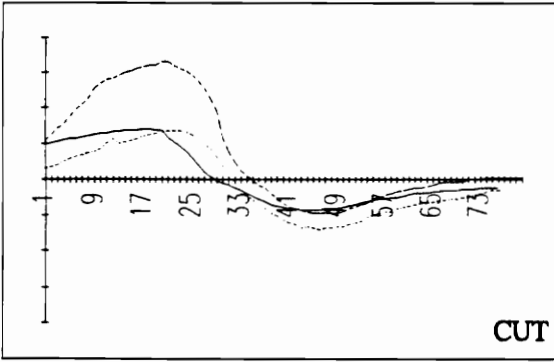
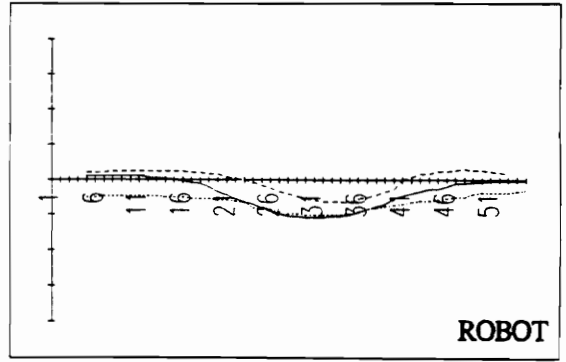
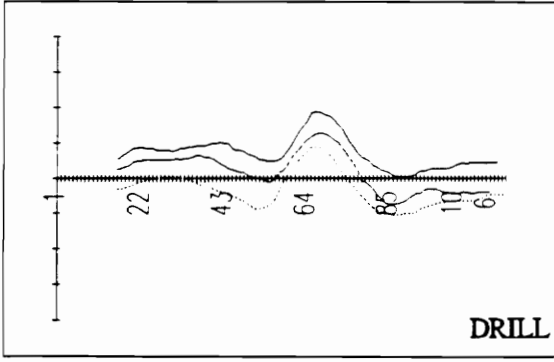
SPEAKER #2



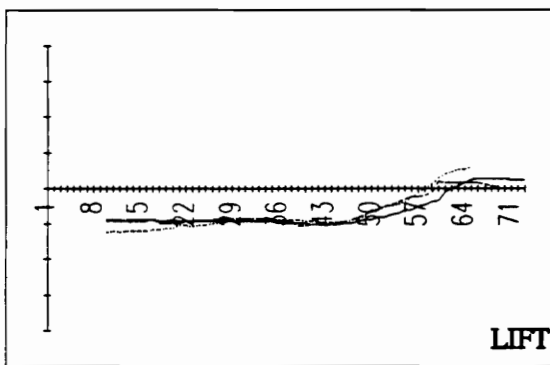
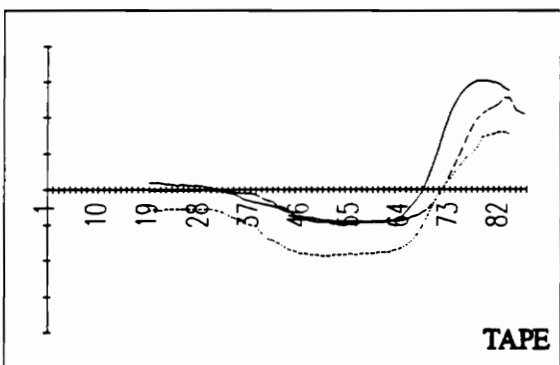
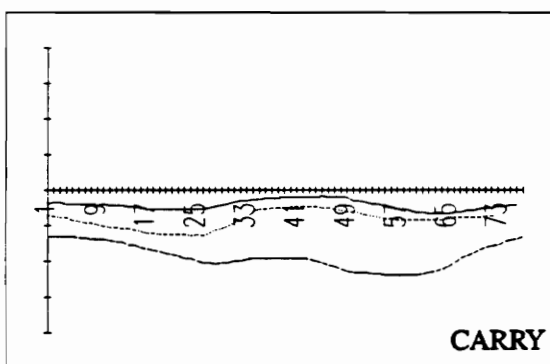
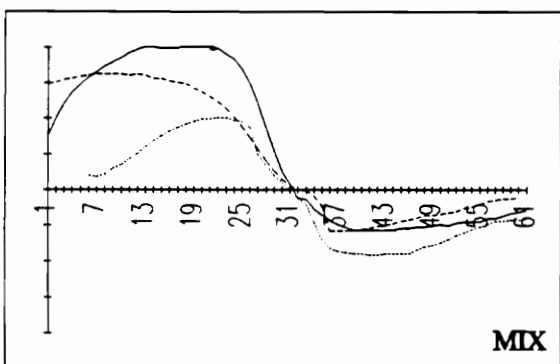
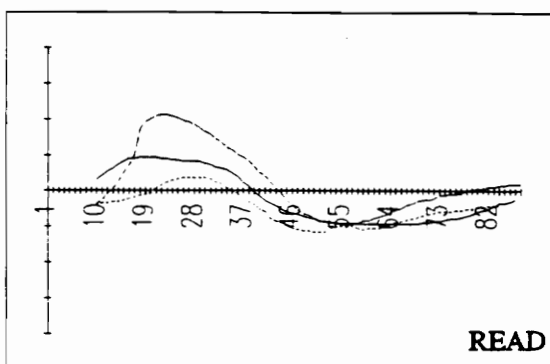
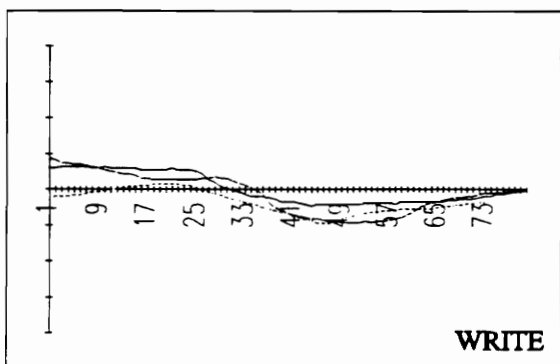
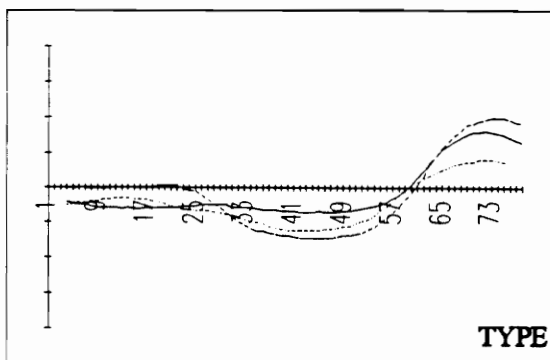
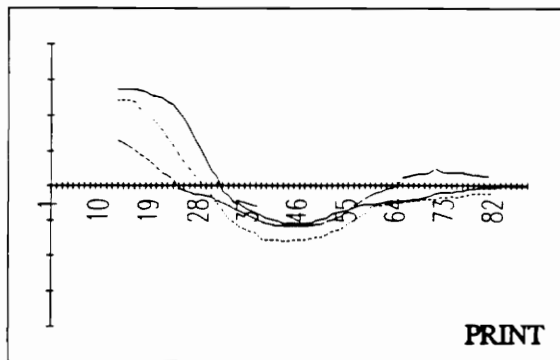
SPEAKER #3



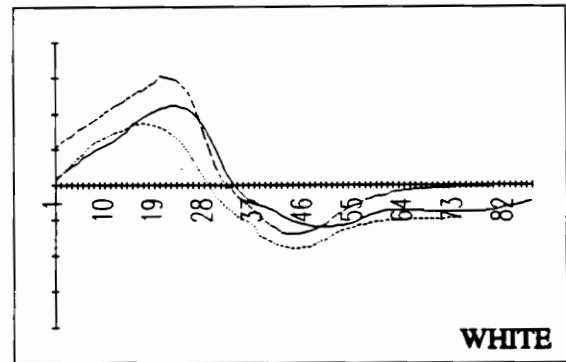
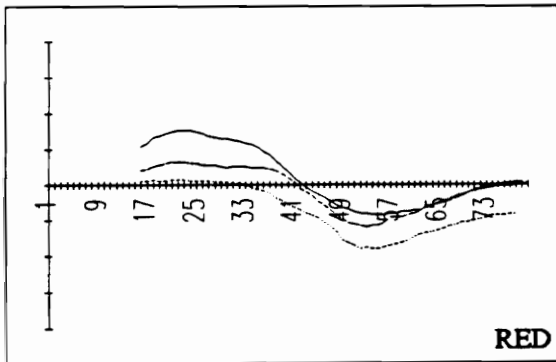
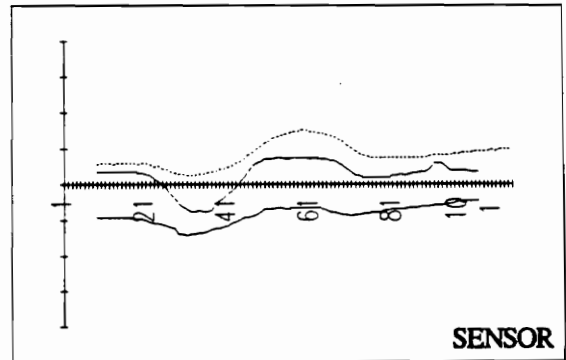
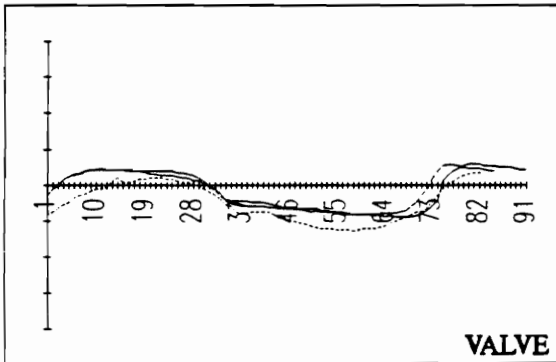
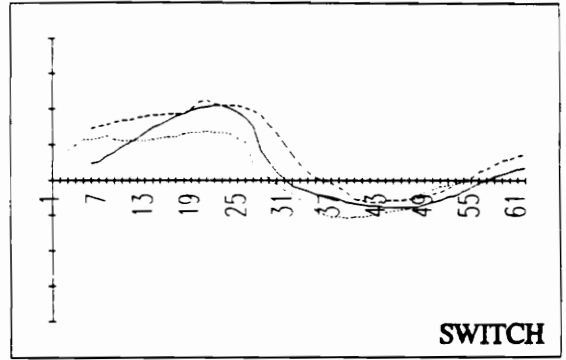
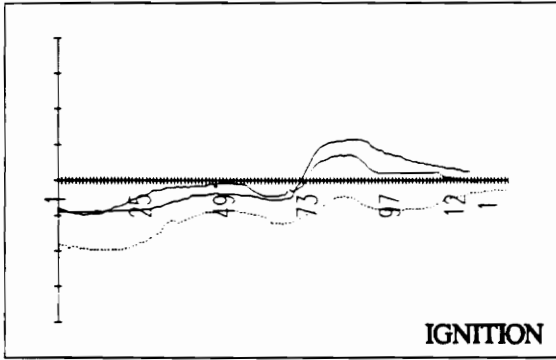
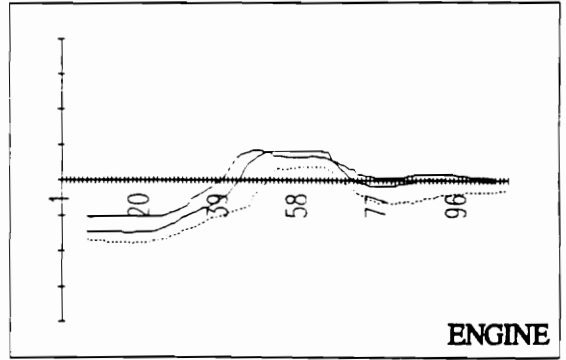
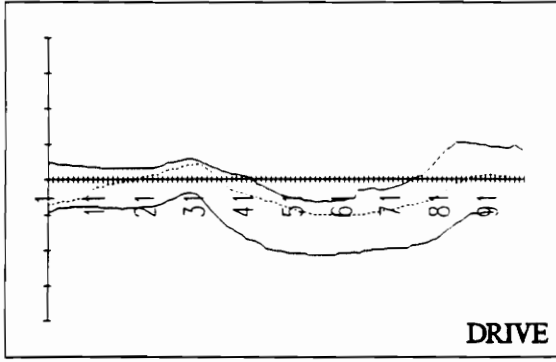
SPEAKER #3



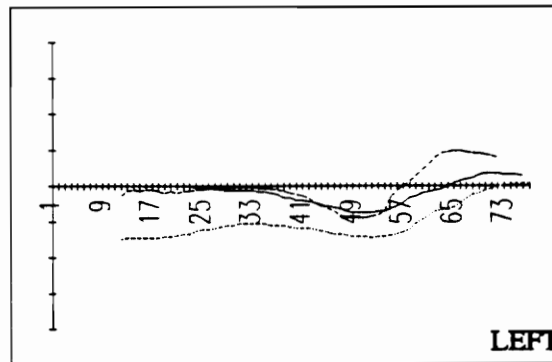
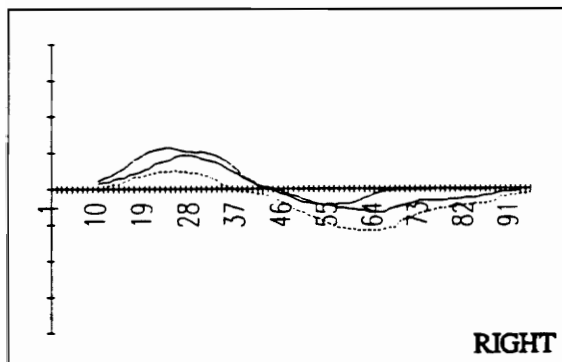
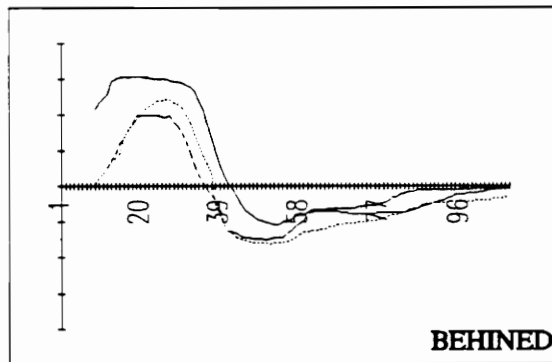
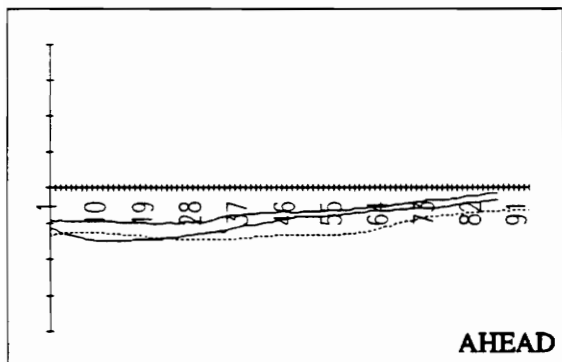
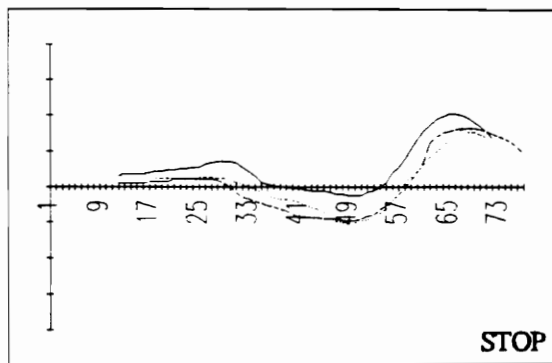
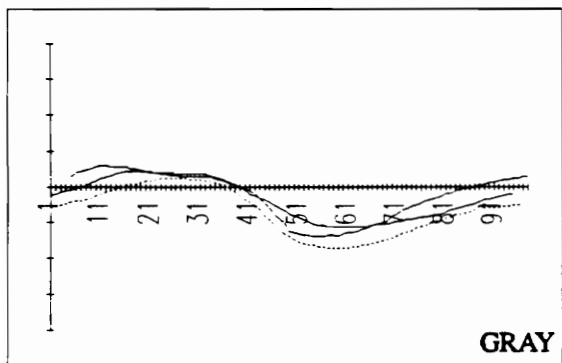
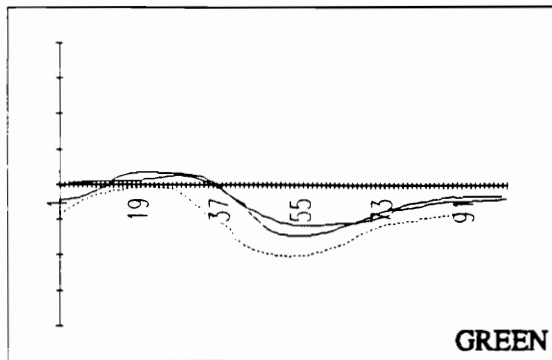
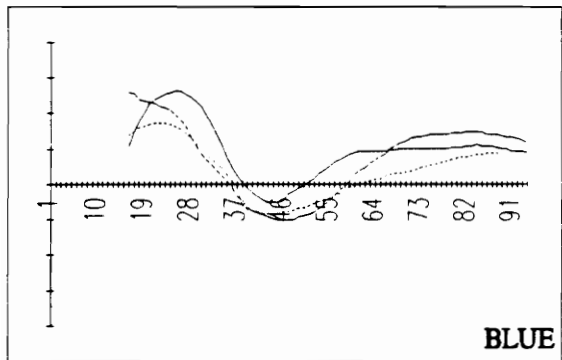
SPEAKER #3



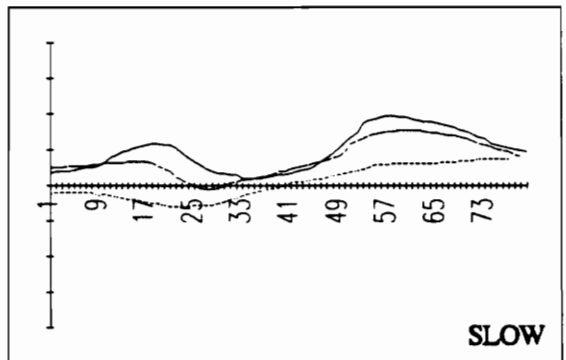
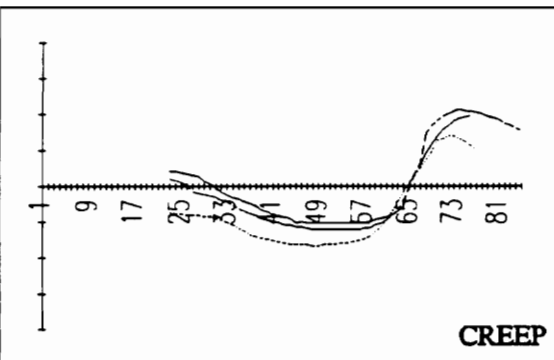
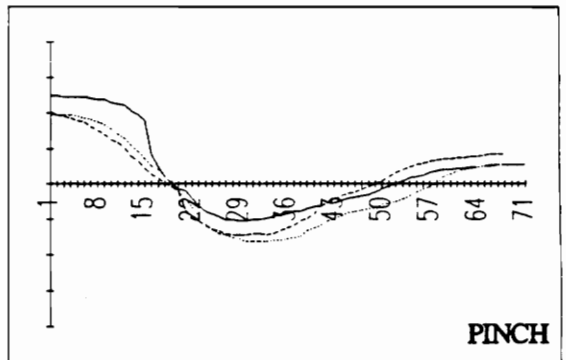
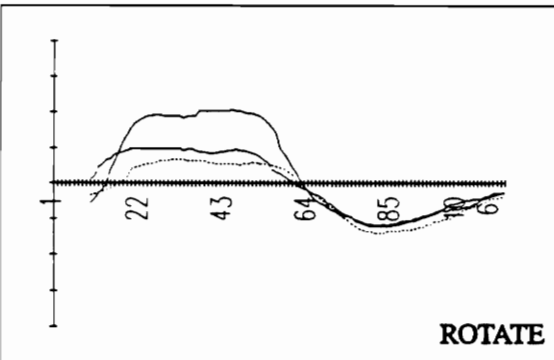
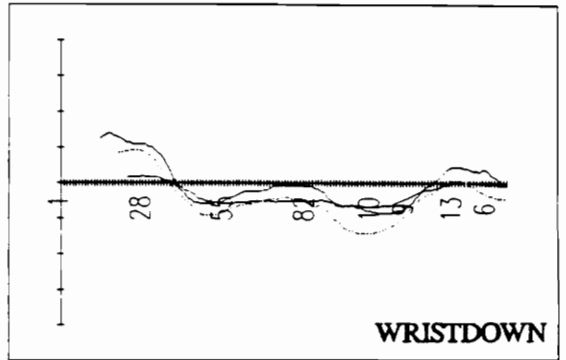
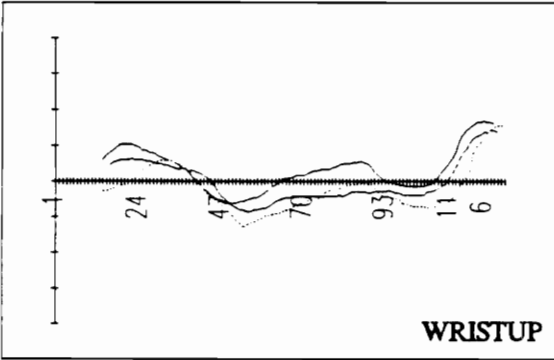
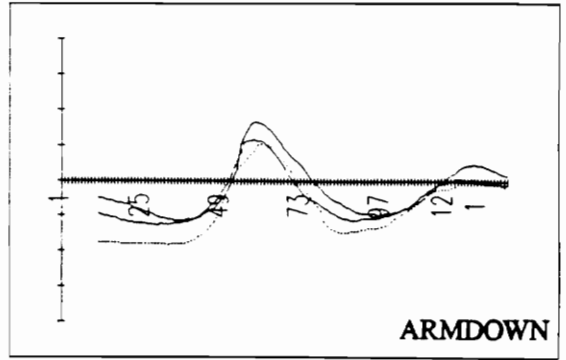
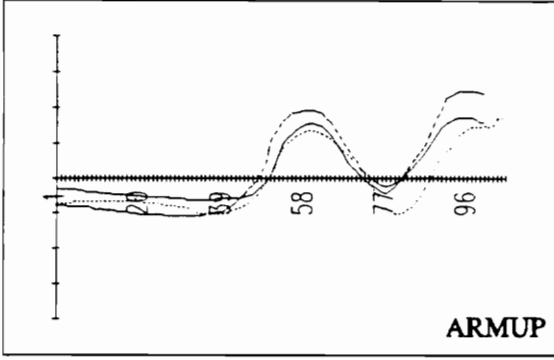
SPEAKER #3



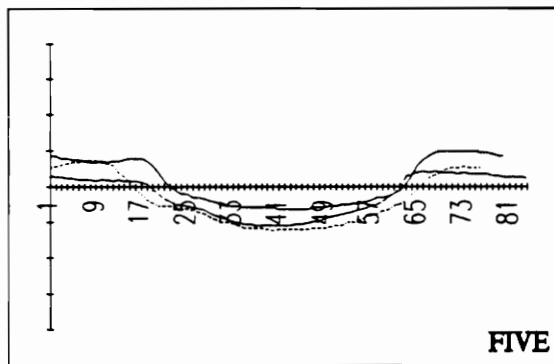
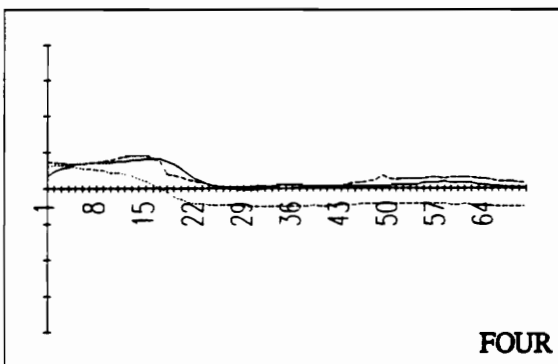
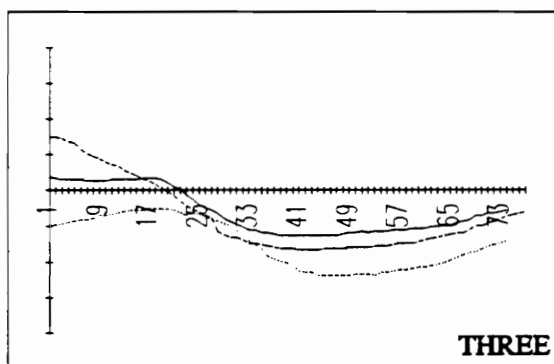
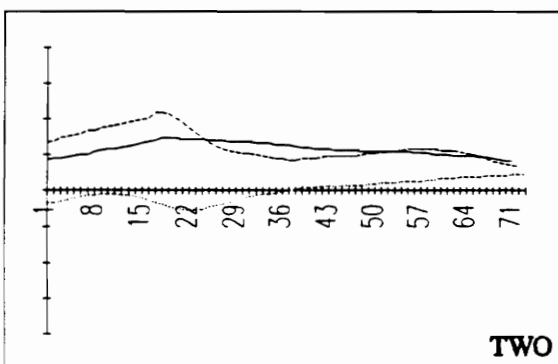
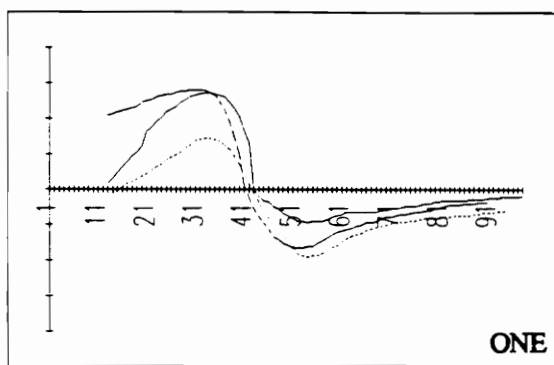
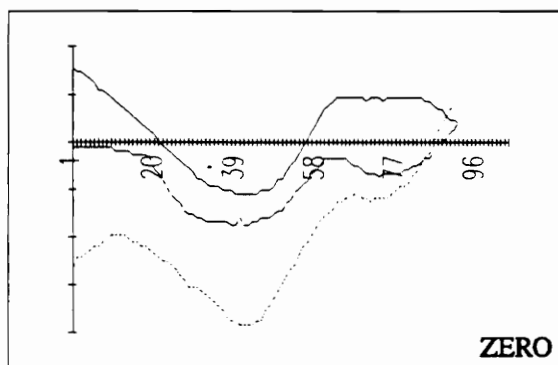
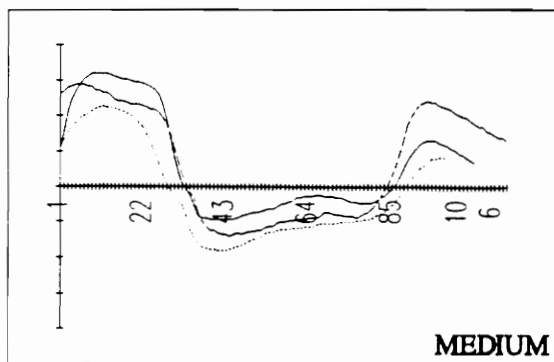
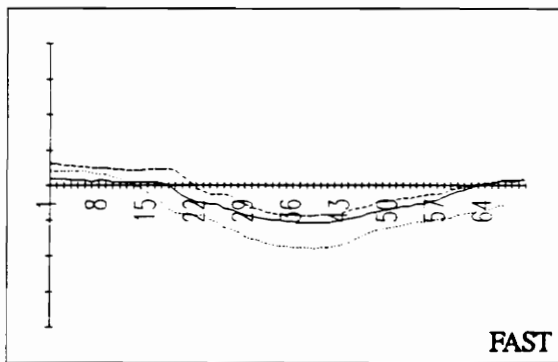
SPEAKER #3



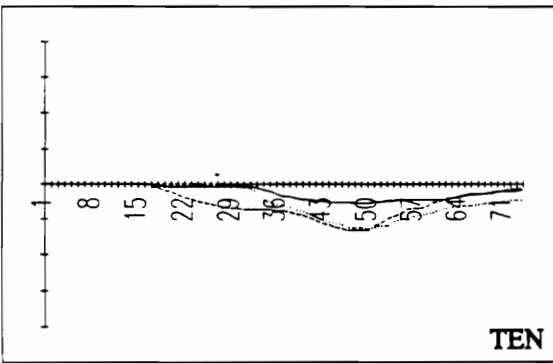
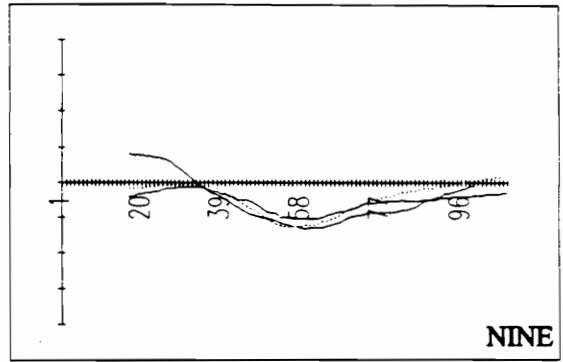
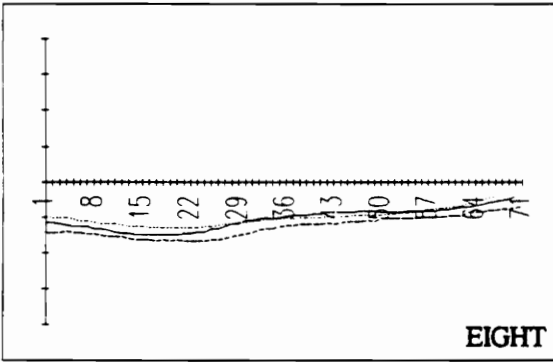
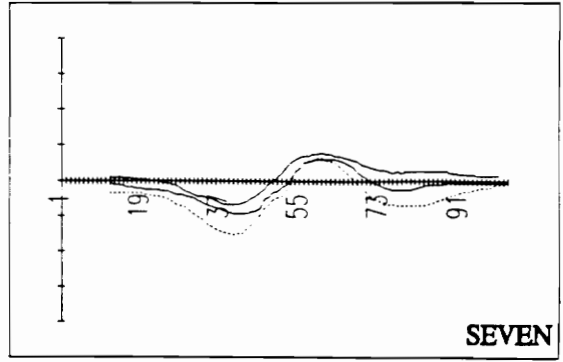
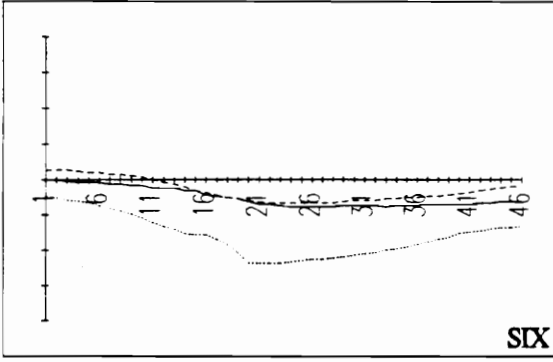
SPEAKER #3



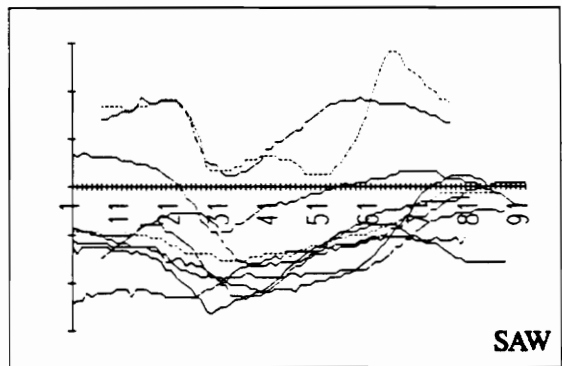
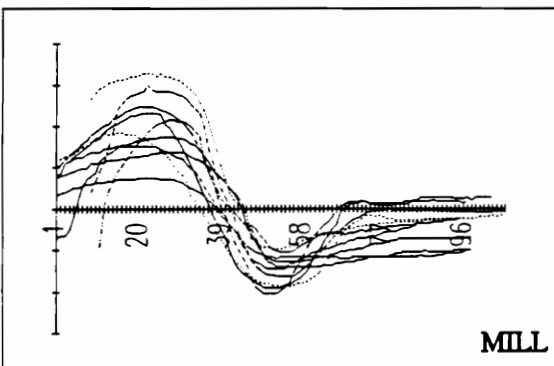
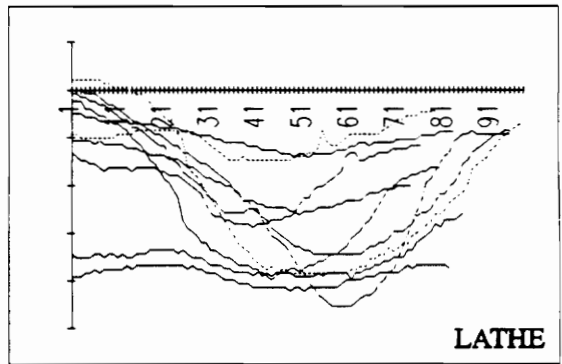
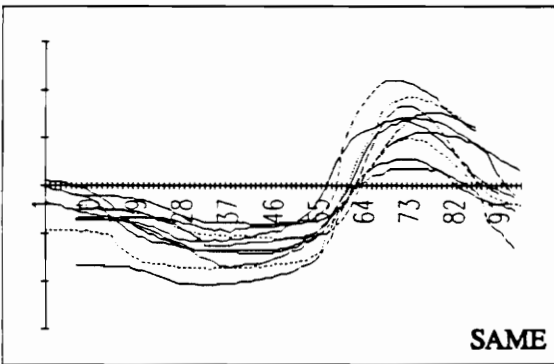
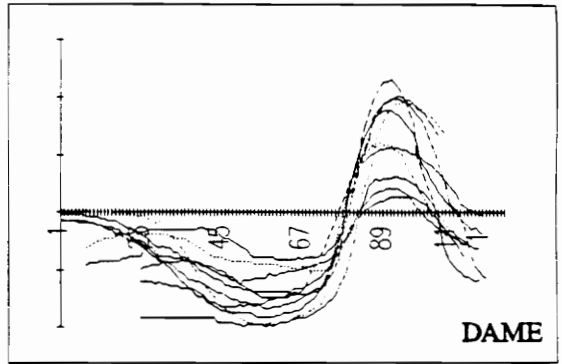
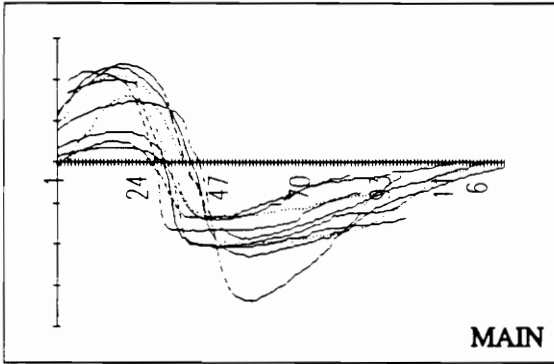
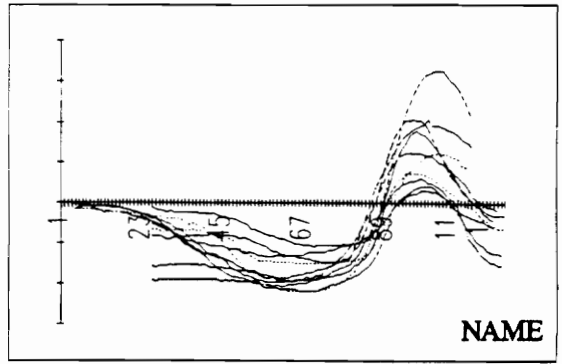
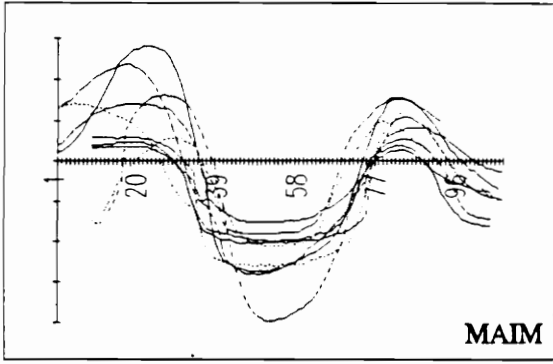
SPEAKER #3



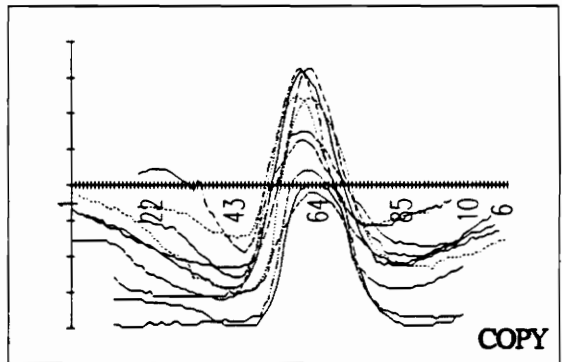
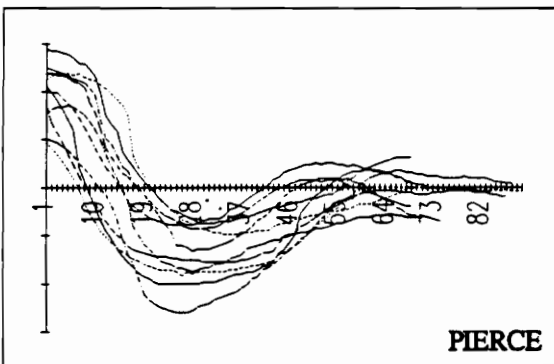
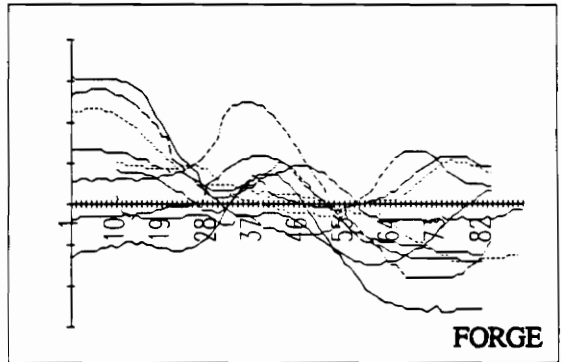
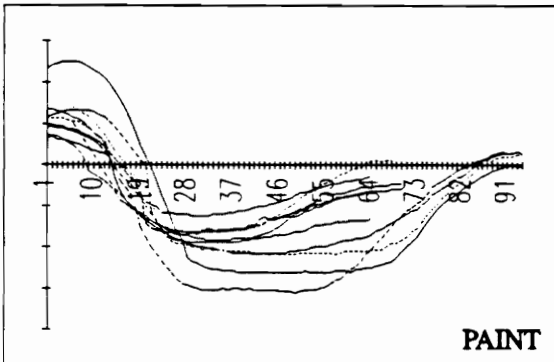
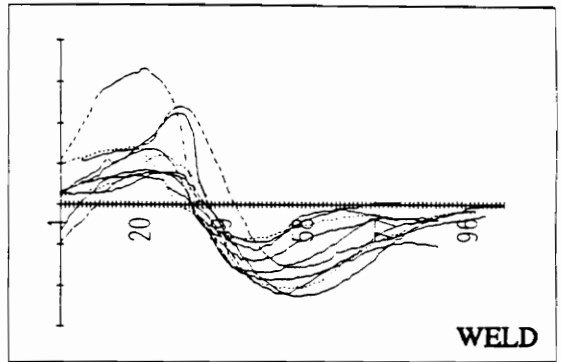
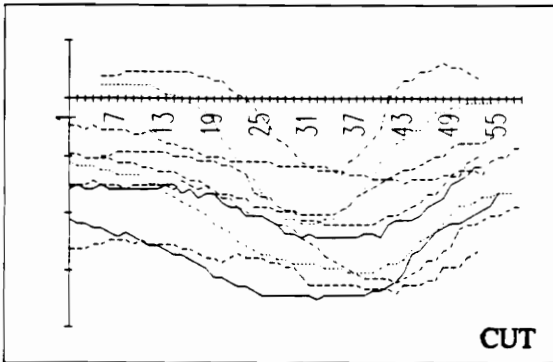
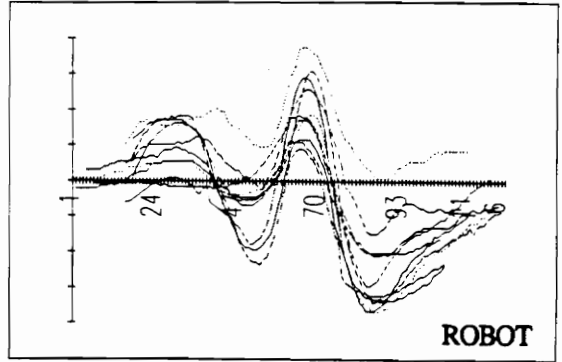
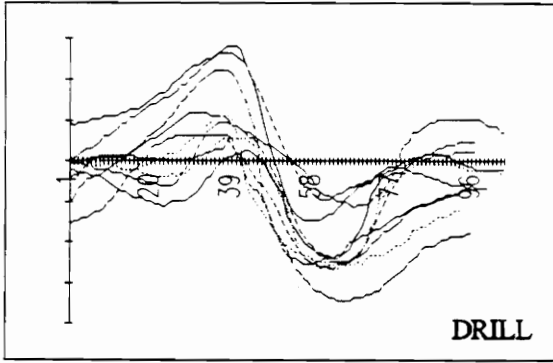
SPEAKER #3



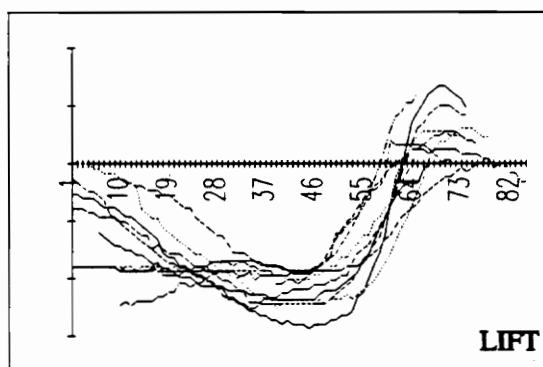
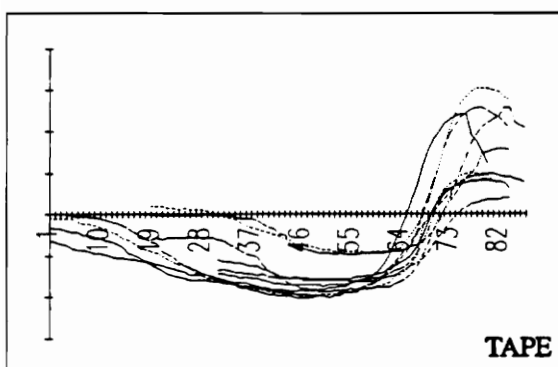
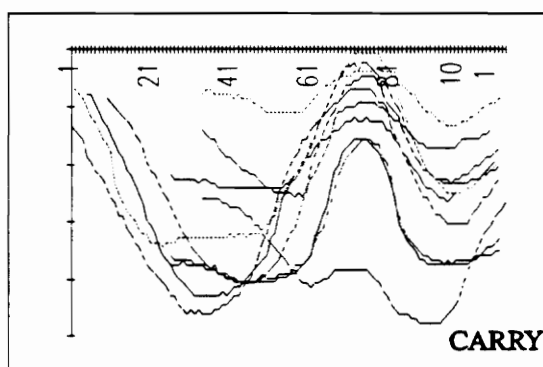
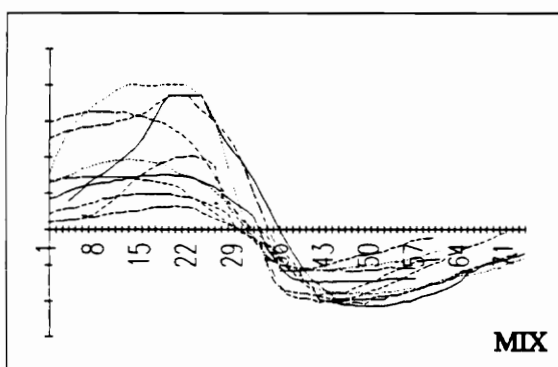
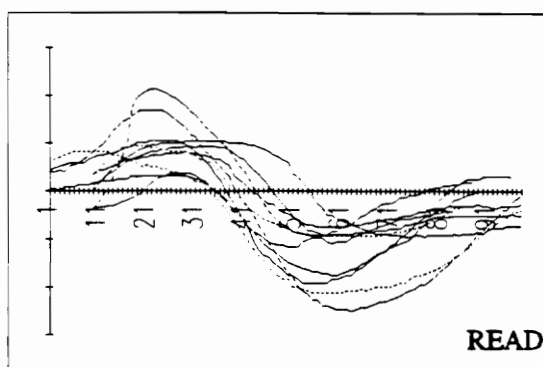
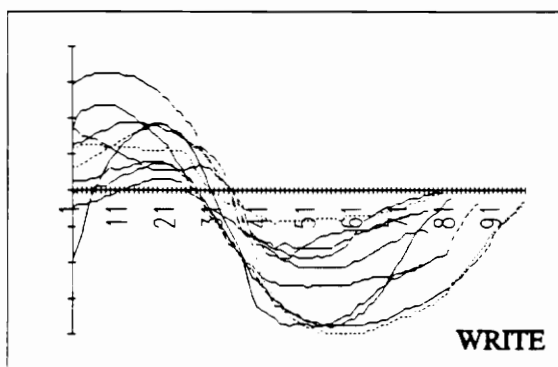
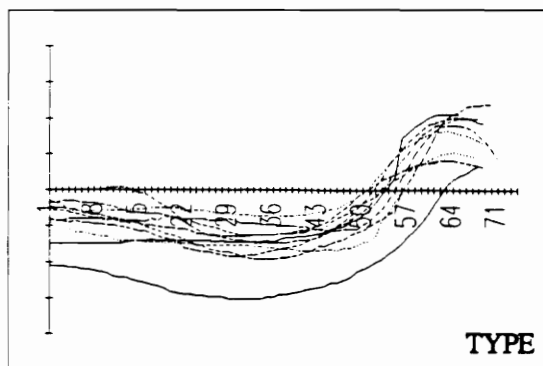
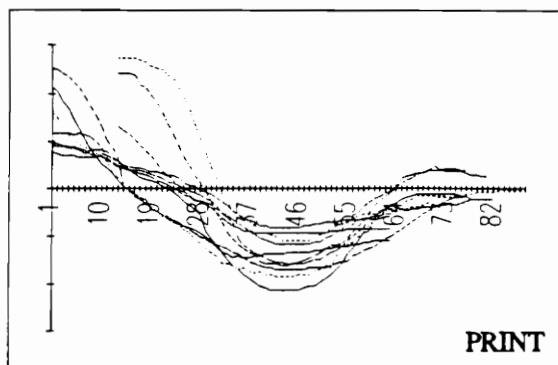
ALL SPEAKERS



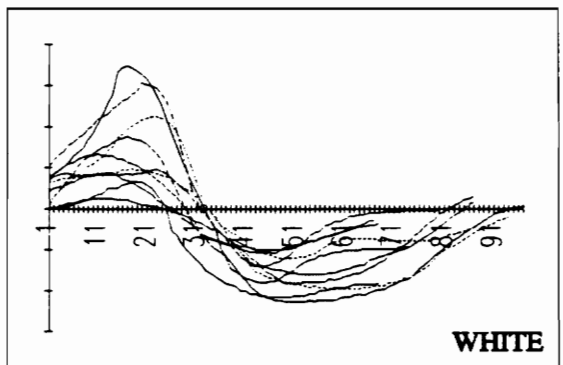
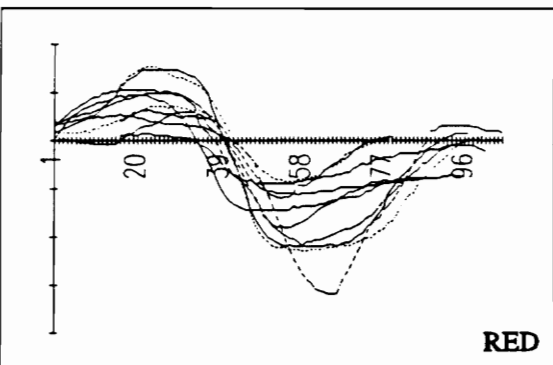
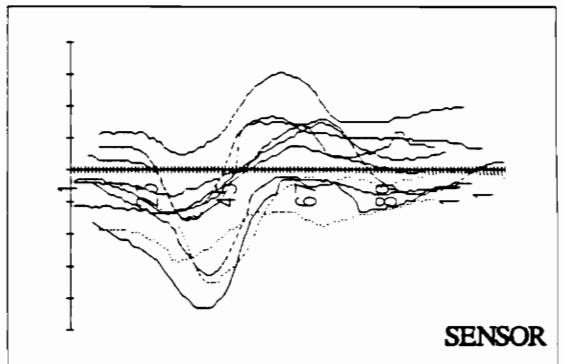
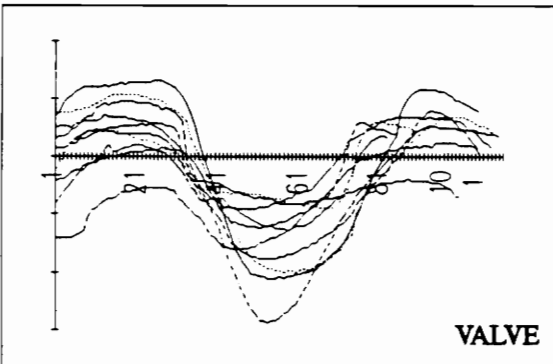
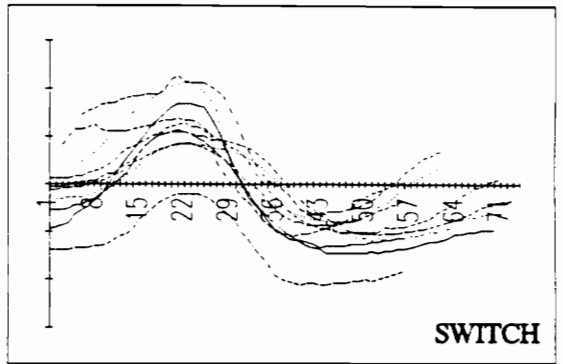
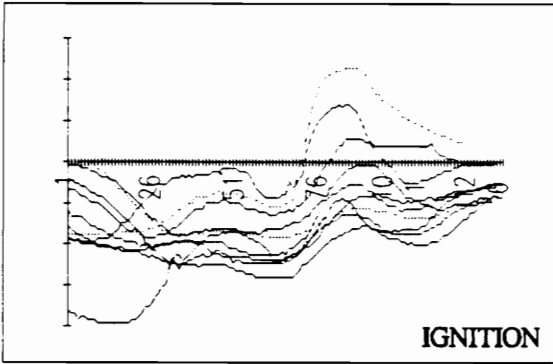
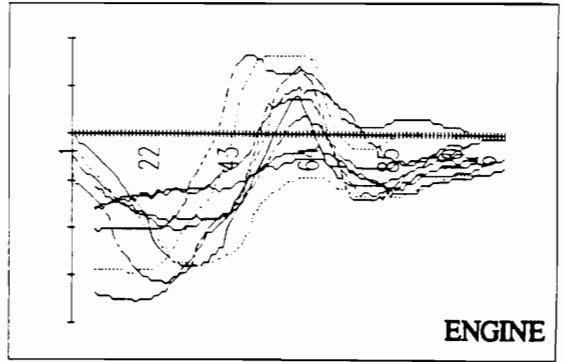
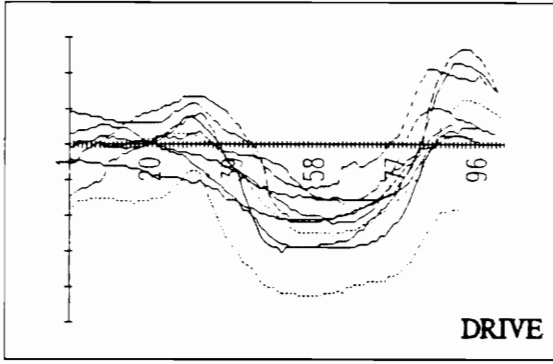
ALL SPEAKERS



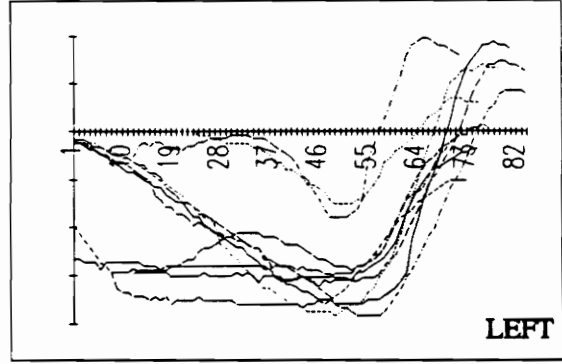
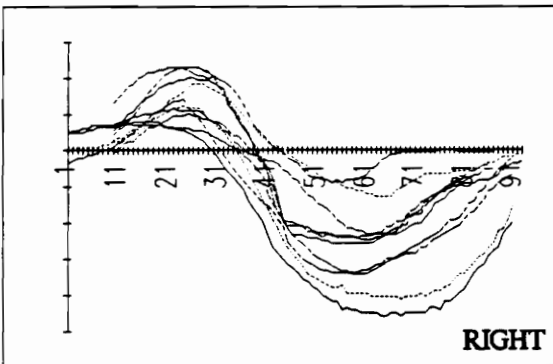
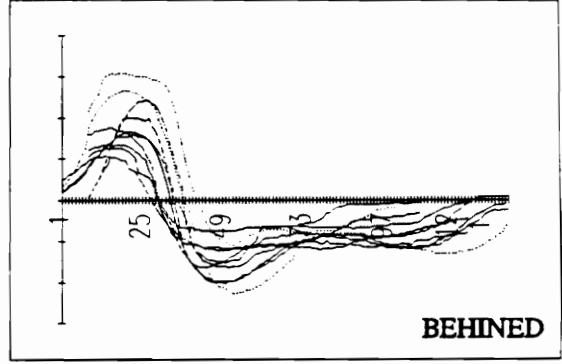
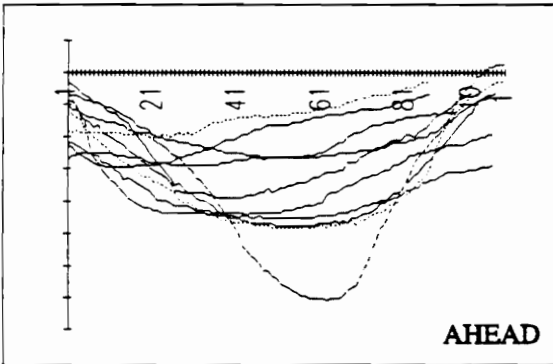
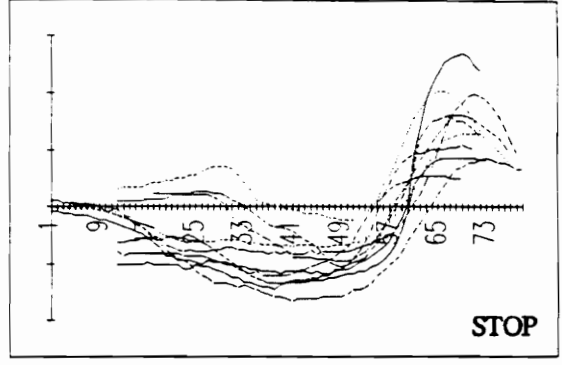
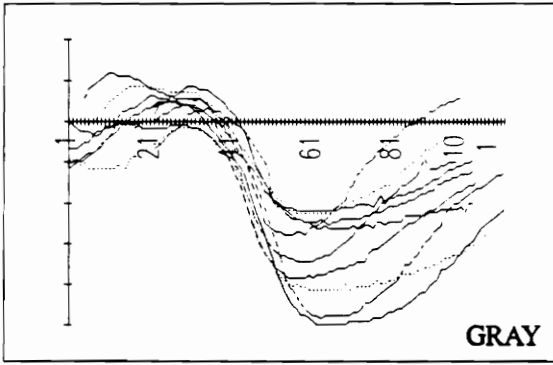
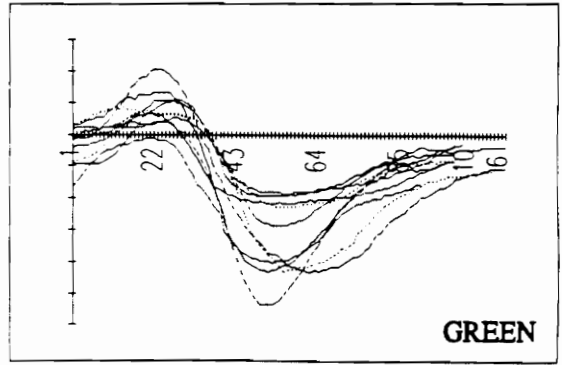
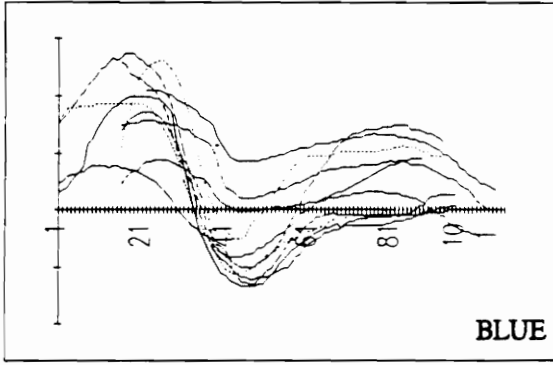
ALL SPEAKERS



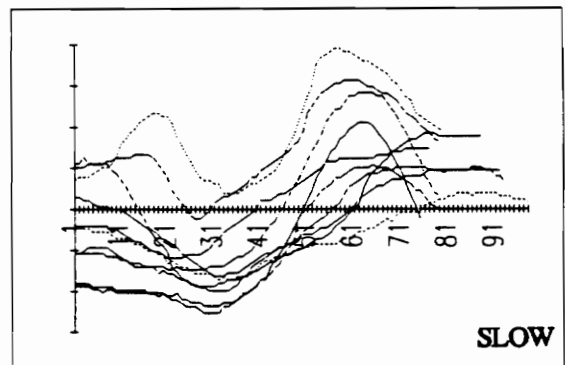
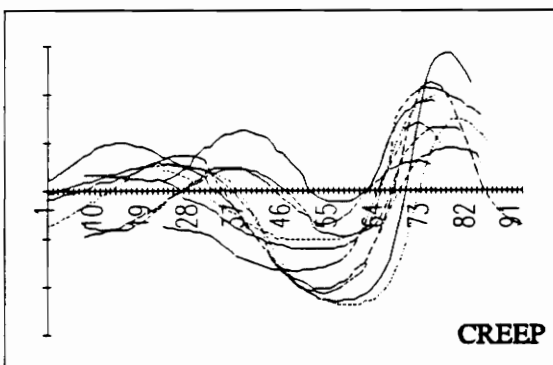
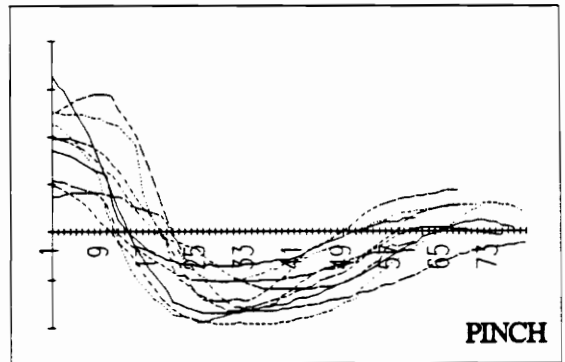
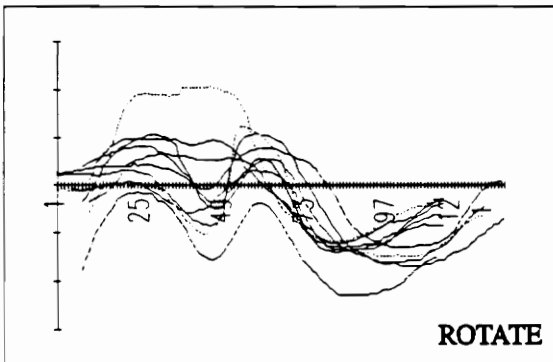
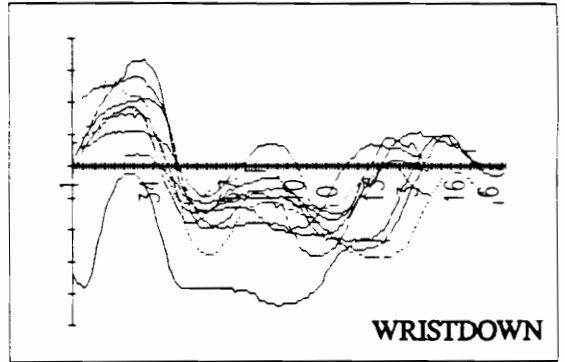
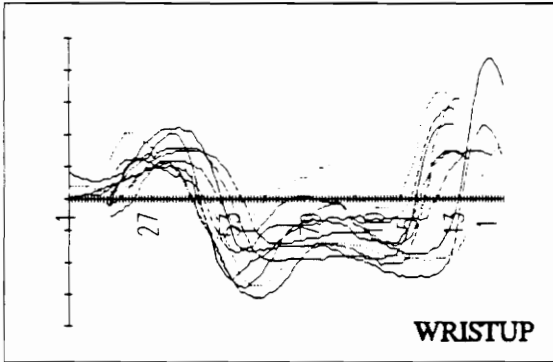
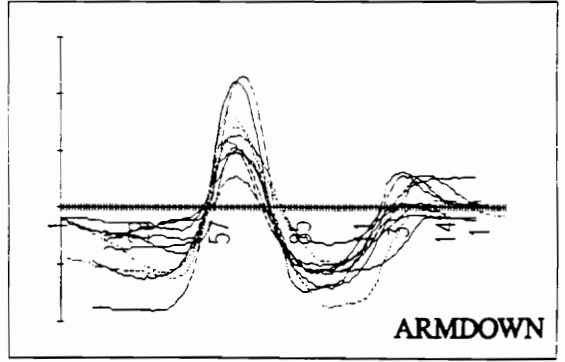
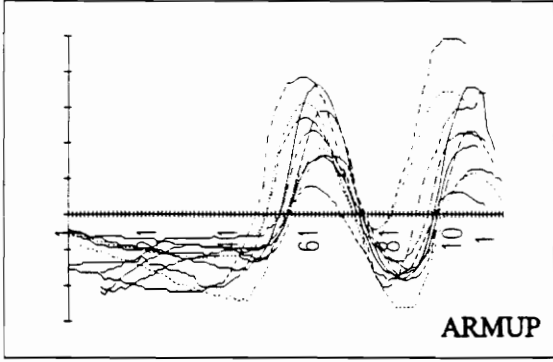
ALL SPEAKERS



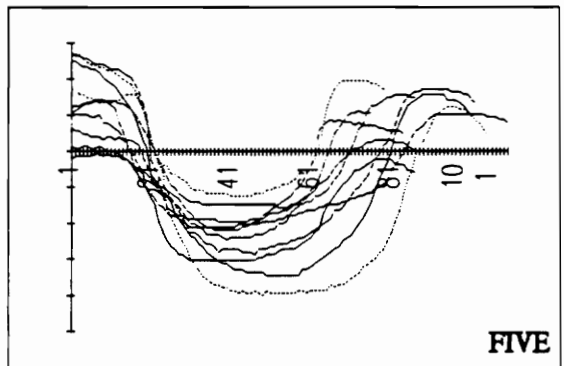
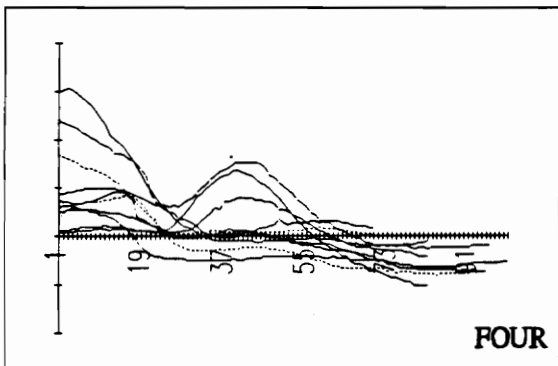
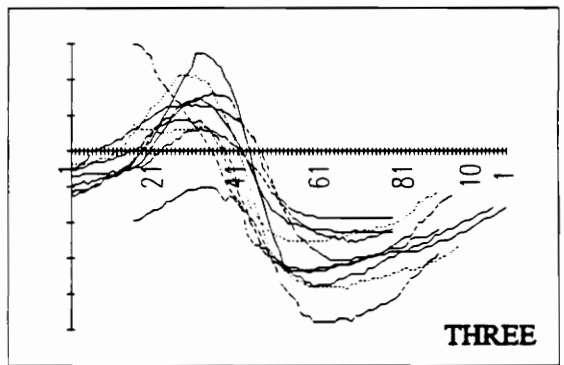
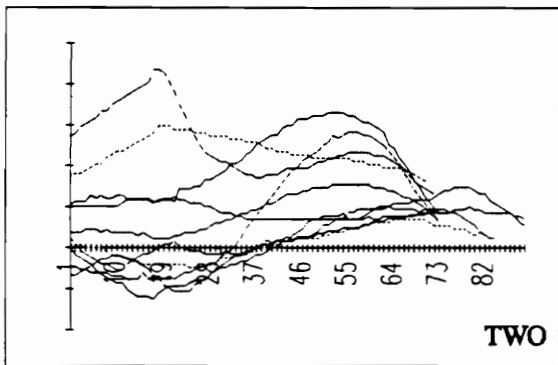
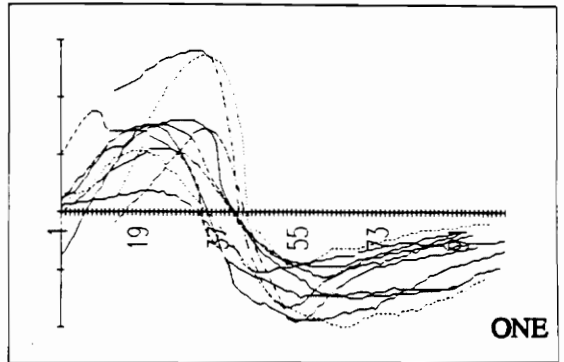
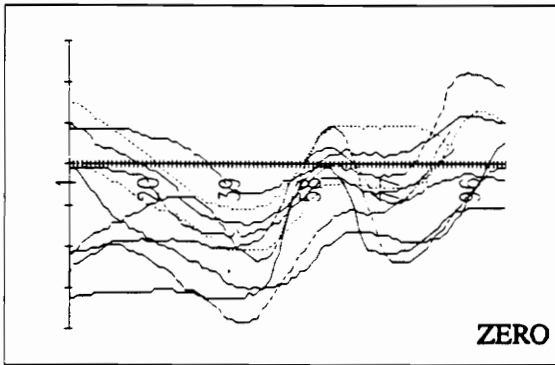
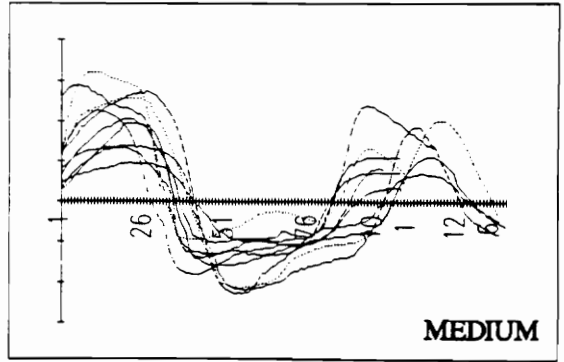
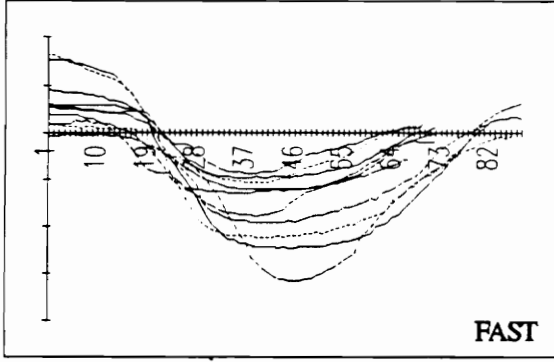
ALL SPEAKERS



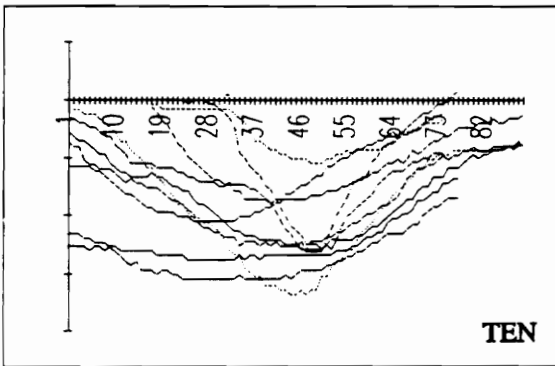
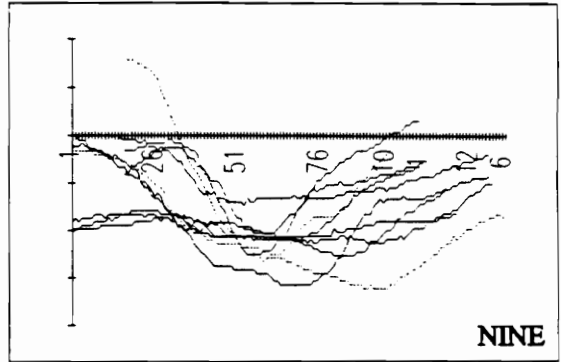
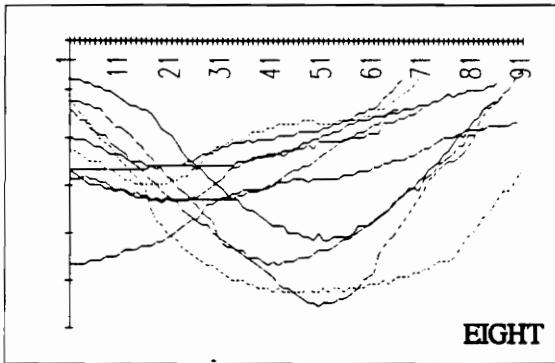
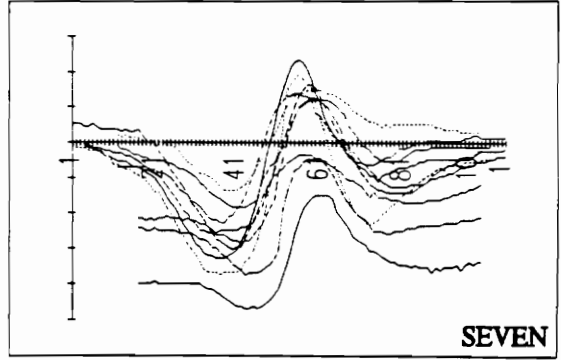
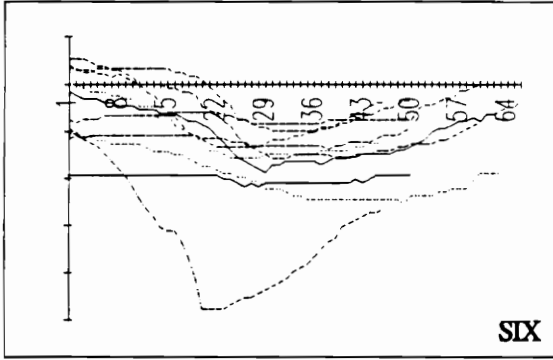
ALL SPEAKERS



ALL SPEAKERS



ALL SPEAKERS



APPENDIX B COMPUTER PROGRAM LISTING

```

10 '
11 '
12 '
13 '
14 '
15 '
151 DIM V(500)
152 DIM W(500)
160 SCREEN 0,0,0 : WIDTH 80 : KEY OFF : CLS
180 'Load DAS4.BIN @ segment Hex 6000 - any empty memory area will do
190 DEF SEG = &H6000 'remove lines 190 & 200 for compiled BASIC programs
200 BLOAD "DAS4.BIN",0 'driver routine loaded at 6000:0000
210 '
220 'Initialize using mode 0
230 DIM D%(3) 'declare 4 element integer data transfer array
240 MD% = 0 'assign mode number (0 to initialize)
250 DAS4 = 0 'set call offset to zero
260 D%(0) = &H300 'base I/O address (should correspond to dipswitch)
270 D%(1) = 2 'interrupt level (not activated unless mode 5 used)
280 FLAG% = 99 'declare error variable, value irrelevant
290 CALL DAS4 (MD%, D%(0), FLAG%) 'initialize
300 IF FLAG% <> 0 THEN GOSUB 10000 'check FLAG% for any errors
320 MD%=0
330 D%(0)=&H300:D%(1)=2
340 GOSUB 10000 'INITIALIZE
350 MD%=1
360 D%(0)=0:D%(1)=1
370 GOSUB 10000 'SET SCAN LIMITS
371 INPUT "name";A$
375 OPEN "i",#2,"vocabulary"
380 OPEN A$ FOR OUTPUT AS #1
381 IF EOF(2) THEN PRINT,"end of vocabulary";GOTO 490
390 INPUT #2,B$
391 PRINT"press S if you wish to terminate, otherwise press C to continue"
392 INPUT; BB$
393 IF BB$="s" OR BB$="S" THEN 490
394 IF BB$="c" OR BB$="C" THEN PRINT,"READ THE WORD AND PRESS THE SPACE BAR WHEN
DONE" ELSE 391
396 ' IF BB$ <> "c" OR BB$ <> "C" THEN 391
398 CLS
400 PRINT,B$
405 PRINT#1,B$
410 MD%=2 'begin A/D conversion
420 FOR I%=1 TO 10000
430 CALL DAS4(MD%,D%(0),FLAG%)
431 X%=D%(0) 'read acoustic channel
432 CALL DAS4(MD%,D%(0),FLAG%) 'read optical channel
433 PRINT#1,X%*5/128,D%(0)*5/128 'print both channels to file
441 X=TIMER*2/3 'beginning of delay loop to
442 X=TIMER*2/3 'achive a sample rate of
443 X=TIMER*2/3 ,100 HZ
444 X=TIMER*2/3
445 X=TIMER*2/3
446 X=TIMER*2/3
447 X=TIMER*2/3
448 X=TIMER*2/3
449 X=TIMER*2/3
450 X=TIMER*2/3
451 X=TIMER*2/3 'end of delay loop
452 IF INKEY$ <> "" THEN PRINT#1,"999","999":GOTO 381

```

```

470 NEXT I
490 CLOSE#1
491 CLOSE#2
600 OPEN"i",#1,A$ 'begin data processing
610 INPUT"OPTICAL OUTPUT DATA";C$
620 INPUT"ACOUSTIC OUTPUT DATA";D$
630 OPEN"O",#2,C$
640 OPEN"O",#3,D$
700 FOR I=1 TO 500
710 V(I)=0;W(I)=0
720 NEXT I
800 IF EOF(1)THEN 9999
801 INPUT#1,B$
805 FOR I=1 TO 500
820 INPUT#1,V(I),W(I)
830 IF V(I)>500 THEN 860
840 NEXT I
841 PRINT,K,I
860 FOR K=1 TO I
861 IF W(K)>100 THEN W(K)=0
862 IF V(K)>100 THEN V(K)=0
870 W(K)=W(K)*3!
880 NEXT K
900 SUMV=0
910 SUMW=0
920 FOR K=1 TO 10
930 SUMV=SUMV+V(K)
940 SUMW=SUMW+W(K)
950 NEXT K
960 A1=SUMV/10
970 A2=SUMW/10
975 FOR K=1 TO I
980 V(K)=V(K)-A1;W(K)=W(K)-A2
985 IF ABS(V(K))<.02 THEN V(K)=0
990 IF ABS(W(K))<.3 THEN W(K)=0
995 NEXT K
1100 WMIN=0
1110 WMAX=0
1120 FOR K=1 TO I
1130 IF W(K)>WMAX THEN WMAX=W(K)
1140 IF W(K)<WMIN THEN WMIN=W(K)
1150 NEXT K
1160 AMAX=.2*WMAX
1170 AMIN=.2*WMIN
1180 FOR K=1 TO I
1190 IF W(K)<AMIN OR W(K)>AMAX THEN 1200
1195 NEXT K
1200 KB=K
1205 S1=SGN(V(K))
1210 KE=K
1220 FOR K=KB+1 TO I
1230 IF W(K)<AMIN OR W(K)>AMAX THEN KE=K
1240 NEXT K
1250 KB=KB-20
1260 KE=KE+10
1391 PRINT#2,B$;PRINT#3,B$
1400 FOR K=KB TO KE
1410 PRINT#2,V(K)
1411 PRINT#3,W(K)
1415 NEXT K

```

```
1490 GOTO 800
9999 END
10000 CALL DAS4(MD%,D%(0),FLAG%)
10010 IF FLAG%<>0 THEN PRINT"ERROR#";FLAG%;"IN MODE";MD%:STOP
10020 RETURN
```

VITA

The author was born in Jordan in 1951, immigrated to the U.S. in 1969 and became U.S. citizen in 1980.

Received a B.S. degree in Physics from Virginia Polytechnic Institute and State University. Employed by Schlumberger Well Services from 1980-1989. While with Schlumberger developed experience and interest in Application and Manufacturing Engineering, with special interest in Sensor Physics and Manufacturing Automation.

Returned to VA Tech in 1989 for graduate work in Manufacturing Engineering. Completed an M.S. degree in Manufacturing Engineering in November 1991 and is continuing in the Doctorate Program.

Taysir H. Nayfeh

