

**PARAMETERS AFFECTING (INTERSUBJECT  
VARIABILITY AND INTRASUBJECT  
VARIABILITY IN) VOICE IDENTIFICATION**

M - 9518

Sharmila (S)

A Dissertation submitted as part fulfilment of  
final year M.Sc. (Speech and Hearing)  
to the University of Mysore,  
Mysore.

May 1997

All India Institute of Speech and Hearing.  
Mysore - 570 006  
INDIA

*Dedicated to PAPA, MAMA*  
&  
*AKKA*

## CERTIFICATE

*This is to certify that this dissertation entitled "PARAMETERS AFFECTING (INTER - SUBJECT VARIABILITY AND INTRA SUBJECT VARIABILITY IN) VOICE IDENTIFICATION" is the bonafide work in partfulfilment for the degree of "Master of Science (Speech and Hearing)" of the student with register number M9518.*

Mysore  
May, 1997

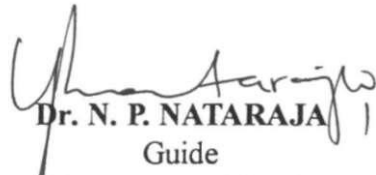
  
r. (Miss) S. NIKAM  
Director,

All India Institute of Speech and Hearing,  
Manasagangotri,  
MYSORE - 570 006.  
INDIA

## CERTIFICATE

*This is to certify that this dissertation entitled "PARAM-  
ETERS AFFECTING (INTER - SUBJECT VARIABILITY AND INTRA SUB-  
JECT VARIABILITY IN) VOICE IDENTIFICATION" has been prepared  
under my supervision and guidance.*

Mysore  
May, 1997

  
Dr. N. P. NATARAJA 14/6/97  
Guide

Professor and Head,  
Department of Speech Sciences  
All India Institute of Speech and Hearing,  
Manasagangotri,  
MYSORE - 570 006.  
INDIA

## DECLARATION

This dissertation entitled "PARAMETERS AFFECTING (INTER - SUBJECT VARIABILITY AND INTRA SUBJECT VARIABILITY) INVOICE IDENTIFICATION" is the result of my own study under the guidance of Dr. N.P. Nataraja, Professor and Head. Department of Speech Sciences, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier at any University for any other Diploma or Degree.

Mysore  
May, 1997

**M9518**

## A C K N O W L E D G E M E N T

I would like to extend my sincere gratitude to my guide and teacher **Dr. NATARAJA**, for his invaluable guidance, and support. Thank you Sir, for every thing.

I thank **Dr. (Miss) NIKAM**, Director, All India Institute of Speech & Hearing, Mysore, for having given me the permission and opportunity to undertake this dissertation.

A million thanks to my subjects, Rahul, Binu, Rajendra Swamy Sir, Gokul and Saji for being patient and co-operative throughout the testing.

Rohit, friends like you are very few. I am lucky to have you.

Sara (S.H), our frindship means a lot to me. I will always cherish it. Thank you for being there for me Knowing you has definitely enriched my life !.

Veena, thank you so much for your thoughtfulness, help and support without which this study would have remained incomplete.

Achu, Madhu, Sonu, I have enjoyed the times we have spent together. Life means so much more with friends like you to share it with.

Rakhee, Asha, Munna, Anu, Su, Jay, Megha, Divya, Raji, and Pradeep. The boisterous fun times we have had, will remain wonderful memory for always.

J.P. and Rajesh, having you'll around makes life brighter. I appreciate your thoughtfulness and concern.

Manoj, Binu, Muthu, Prarthana, Amri, Kaveri, Beula, Arushi, Prachi and Manju, It's been fun knowing you all.

Amit, your constant calls and letters were pleasant break in myu busy schedule. Thank you so much for your support and concern.

Raju, your love, support and friendship makes life a whole lot easier.

Usha and Anusha, friends who care are rare. I am glad to have you in my life.

I would like to extend my thanks to Akka for her typing work.

I am highly obliged to Spaceage Electronic Typing, for typing and printing my study. Thank you.

Thanks are also due to Mr. B.K. Venkatesh, for typing part of this dissertation.

## C O N T E N T S

			Page No.
CHAPTER	I	Introduction	1
CHAPTER	II	Review of literature	5
CHAPTER	III	Methodology	51
CHAPTER	IV	Results and discussion	57
CHAPTER	V	Summary and conclusions	105
		Bibliography	109

## INTRODUCTION

We had not dreamed these things were so  
of sorrow and of mirth.

Her speech is as a thousand eyes,  
through which we see the earth  
God wave a web of loveliness  
of clouds and stars and birds,  
But made not anything at all  
so beautiful as words.

They are as fair as bloom or air,  
They shine like any star,  
Am I rich who learned from her  
how beautiful they are.

-Anna Hempstead Branch,

Her words.

The heritage of the present day student of speech is an ancient and honourable one, for the study of speech is one of the oldest of academic disciplines. It is 2500 years since Aristotle, but 2500 years before Aristotle the study of speech occupied a place of interest in human affairs. The last few decades, however, have seen various other contributions to our understanding of speech. It may be said that more progress has been made in the past 50 years than was made in all the previous centuries..

The role of voice in speech is obvious. The majority of phonemes are voiced, including all vowels, semi vowels and nasals. Most of the remaining consonants made up of voiced, unvoiced pairs, making the total phonemes set predominantly voiced. In addition voicing carries the rhythm and melody of speech. These are patterns of pitch, loudness and duration that tie together



syllables, phrases and sentences. In 1940's a new field arose wherein a lot of attention was focussed on the process of voice identification. This dissertation deals mainly with the above mentioned topic.

In 1944, Gray and Kopp found that spectrograms could be used for speaker identification. Spectrogram portrays talker dependent features, in addition to phonetic variations. Kresta (1962) re-examined "voice prints" and stated that voice spectrograms could provide reliable means of identification other investigators tried to duplicate these favourable results, but with higher error rates than Kresta's (Stevens et al 1968; Tosi 1967, 1968). They concluded that " Kresta's method showed promise " and suggested the need for further study.

Speaker recognition can be carried out by many means of.

- a) Aural examination of voices,
- b) Visual examination of spectrograms.
- c) With the help of computers.

The initial two methods are considered as subjective methods and the last is considered as an objective method. Subjective methods are those wherein the trained personnel makes the decision as to whether the voice belongs to the talker. After evaluation of laboratory conditions and field conditions, (Tosi et al and Nash) had drawn the conclusion that a combined method of aural and visual examination of speech samples can be used in the investigation of a crime (provided certain standards are maintained). This method was called as " Voice printing".

Objective methods of voice identification are those in which the decision as to whether or not an unknown voice belongs to the same talker is

produced by a machine (computer). Until recently, only subjective methods of voice identification and elimination have been used as evidence in courts of law. Though not all courts accept this method as proof enough for indictment.

Wolf (1972) says " for mechanical recognition of speakers it is desirable to use acoustic parameters that are closely related to voice characteristics that distinguish speakers" Wolf (1972) in his study found that, there is known relation between voice signal and vocal tract shapes and gestures. Parameters which were found to be useful in speaker identification were fundamental frequency, features of vowel and nasal consonant spectrum, word duration and voice onset time. When such parameters were used no errors were made in voice identification.

The area of speaker recognition can be divided into a large number of sub areas depending on the following factors.

- 1) Nature of the task,
- 2) Nature of parameter used,
- 3) Phonetic context
- 4) Population,
- 5) Recognition rate obtained etc.

Nature of the task can be classified as:

- 1) Verification task.,
- 2) Identification task.

Majority of these studies are based on efficiency of visual examination of spectrogram for speaker identification and verification did not determine the importance of specific acoustic features (Mani Rao and Agrawal, 1984). Mani Rao and Agrawal (1984) in their study made an attempt to find the relative importance of acoustic features, which can be used in matching spectrograms. Mani Rao and Agrawal (1984) report that correct verification was possible by verifying the spectrograms to about 85% when the speakers

were male and 72% when they were female. Feature by feature comparison analysis, revealed that the acoustic features for correct verification is different than those for correct elimination. Very few studies have been done on the area of " voice identification" on the Indian population.

The present study was aimed to find out the parameters which affect or influence voice identification in normals and to gain a better perspective of the entire process of voice identification.

**Hypothesis:**

- 1) All features obtained from a speaker remain constant and are equally important.
- 2) Parameters obtained by analysis do not differ from person to person.

**Limitations:**

1. This study was conducted on a limited population of subjects only.
2. Only male subjects were considered.
3. Subjects between age range of 20 - 30 years only were considered
4. The recordings were done under laboratory situation.
5. Recordings over a long period of time were not considered.

**Implications:**

The study will aid in selecting the parameters for the process of voice identification.

## **CHAPTER - II**

### **REVIEW OF LITERATURE**

"Communication has long been recognized as one of the most fundamental component of human behaviour". (Peterson, 1958).

The ability of the human beings to use their vocal apparatus with other organs to express their feelings, describe an event and to establish communication is unique to them. It took millions of years for human beings to develop this faculty. The onset of the human era is recognized to have started with the acquisition of the ability to communicate using the vocal apparatus for social interaction. No normal person has failed to develop this faculty and no other species is known to have developed this ability.

Speech is a form of language that consists of sound produced by utilizing the flow of air from the lungs. The act of speaking is a specialized way of using the vocal mechanism. Speech once acquired becomes a constant companion to a human being. Human culture and ingenuity are based on the ability of the human beings to use symbolic language and thus establish communication.

Speech is easily produced by the human beings. The range of possible variations of speech are immense, it can be varied from soft whisper to a loud shout, on one hand the simplest form of imitation to the highest level of singing.

"The act of speaking is a very specialized way of using the vocal mechanism. The act of singing is even more so. Speaking or singing demand a combination or interaction of the mechanism of respiration, fountain, resonance and speech articulation", (Boone, 1972).

The underlying basis of speech is voice. The importance of voice in speech is very well depicted when one considers the cases of laryngectomy or even voice disorders.

"Voice plays the musical accompaniment to speech, rendering it tuneful, pleasing, audible and coherent, and is an essential feature of efficient communication by the spoken word". (Guene, 1957).

The sounds used in human speech serve for communication at many levels. Less than one percent of the speech is used for linguistic purpose, as such, the rest gives other kind of information about the specific characteristic of a speaker, which enables one to recognize the speaker's physical well being, emotional states and attitudes toward the entire context in which the speech event occurs.

It is well established that voice has both linguistic and non-linguistic functions in any language. The degree of dependence of language on these functions varies from language to language. For eg. 'tone languages rely more upon the voice more specifically, than other languages'.

Voice is the carrier of speech, it acts as musical accompaniment, variation in voice, in terms of pitch and loudness provide rhythm and also

breaks the monotony. This function of voice draws attention, when there is a disorder of voice, leading to monotonous speech.

"Voicing-presence of voice, has been found to be the major distinctive feature" in almost all languages. 'Voicing' functions as a distinctive feature and provides more phonemes and makes the language more broad. When this function is 'absent' or used 'abnormally', it would lead to speech disorders". (Peterson, 1966).

At the semantic level also voice plays an important role, specifically in tone languages. The use of different pitches, high and low, with the same string of phonemes would mean different things. This function of voice is very well demonstrated in tonal languages like "Punjabi" and Thai".

Speech prosody, the tone, the intonation and the stress or the rhythm of language is a function of vocal pitch and loudness as well as of phonetic duration. (O'Malley and Peterson, 1966).

The term 'tone' refers to a feature of syllable in a sequence and the term intonation is used to denote a sequence of tones whose function relates to a sentence or part sentences. Fry (1968) is of the opinion that all the languages make use of the same system of tones, and this may operate at two or three different levels. In some languages, tone may function at a phonological level and contrasts of tone may have effects similar to those of phonemic differences. The tone also functions at grammatical or syntactic level both in tone and non-tone languages.

The distinction between a statement and a question, between a question and a command and so on, having the same string phonemics in the same order. In many languages it is signaled by a difference tone. The other function of voice is conveying the affective state of the speaker. There is considerable interaction between the effects of these three functions. But they form a kind of hierarchy i.e., the phonological tones may get modified by the demands of grammatical intonation, and in turn it may be modified by the need for emotional expression. However, there is never a complete subordination of one level to another.

Each spoken word or sentence consists of series of stresses, just like tones. Each syllable carries some stress and a succession of these stresses make the rhythm or rhythmic partem. The stress and rhythm differences may serve to differentiate the words. Apart from this, stress and rhythm are also used for grammatical and affective functions of a language. Thus the parameters of voice pitch and loudness play a vital role in language. however, the importance of these vary from language to language.

Perkins (1971) has identified at least five non-linguistic functions of voice. Voice can reveal speaker identity i.e., voice can give information regarding the sex, age, height and weight of the speaker. Lass (1980) reports of several studies which have shown that it is possible to identify the speakers age, sex, race, socio-economic status, facial features, height and weight based on voice.

"This aspect of voice has received considerable attention and has been found to be useful in criminology. The ability of the voice to provide information regarding the speaker is from the well perfected implicit code (Voicers, 1964). This code is gaining importance, which is evident from the rapidly increasing interest in voice printing, the telecommunication analogue of finger printing". (Perkins, 1971).

It is a prevailing notion that there is a relationship between voice and personality i.e., and voice reflects the personality of an individual, there were no convincing evidences until investigations by Stark Weather (1961). Ostwald (1963), Mankel, Meisels and Hauck (1964) and Rousey and Moriarty (1965) were conducted and showed the relationship between these two. However, more studies are required in this area.

Fairbanks (1938, 1939, 1941, 1966), Pronovost (1938) and Huttar (1967) have concluded from their studies that the voice reflects the emotional conditions reliably.

Voice can also be considered to be reflecting the physiological state of an individual. For example, a very weak voice may indicate that the individual is not keeping good health, or a denasal voice may indicate that speaker has common cold. Apart from this, it is a well known fact that voice basically reflects the anatomical and physiological condition of the respiratory, phonatory and resonatory systems i.e., disturbance in any one or more of these systems may lead to voice disorders.



An attempt has been made to find out the physiological conditions based on voice analysis. A recently developed aspect in the area of early identification is infant cry analysis. It has been found by many investigators (Blinick, 1971; Fisichelliv, et.al. 1963, 1966; Illingworth, 1981; Indira, 1962). It is possible to identify abnormalities in the neonates by analyzing their cry immediately after birth or within a few hours after birth. The cry analysis has been found to be a reliable and valid predictor of the conditions of the child and it has been adopted as a routine test in many children's hospital.

Several techniques have been developed to identify voice using different information (Kresta, 1961; Dreher, 1967). Kresta (1962) has used spectrographic information to identify the speaker using voice. Dreher (1967) in one of his techniques has used computer analysis of frequencies, intensities, durations and pauses, and in another technique he has used a "Quasi-Fourier analysis" in which speech power is plotted in a circle, whirled under stroboscopic light and analysed in terms of various relationships among standing patterns that can be detailed visually. (Perkins, 1971).

Studies have shown that it is possible to identify race. (Stroud, 1956; Hibler, 1960; Dickens & Sewyer, 1962; Larson & Larson, 1966; Lass et. al. 1979) socio-economic status, (Harms, 1961, 1963) personality (Stagner, 1936; Eisler, Reese, 1967), specific identity (McGhee, 1937; Pollack, Pickett and Sumbly, 1954; Voiers, 1967; Coleman, 1973b), and facial features of the speaker (Lass, & Harvey, 1976), by analysis of voice of the speaker.

These studies have been considered to " provide very useful information in a variety of future theoretical and applied areas of investigators" (Lass, 1980).

The information from such studies will be useful in training listeners in recognizing various characteristics of speakers.

Some of the clues to speakers sex and size identity are derived from an auditory analysis, and from absolute and relative resonant frequencies (McGhee and Ladefoged, 1963, 1967).

For the purpose of speakers sex identification studies have employed voiceless fricatives, isolated spectral noise, and whispered vowels (Schwartz, 1968; Ingermann, 1968; Coleman, 1971). Studies have also been done using electrolarynx as the voice source (Lackurene, 1974). "All these studies have shown that speakers sex can be identified accurately":. (Dennis, 1980).

The results of (Dennis, Ingressano, Gray Weismer & Gordon H. Sehucker, 1980) study on sex identification in children has shown that vocal tract resonance characteristic makes the greatest contribution in the accurate perception of sex, when the information on fundamental frequency is absent.

"In a study of spontaneous speech of five and six year old children. Murray (1975) - listeners, were able to identify the speakers sex with 78 % to 71% accuracy for male and female separately". (Dennis, 1980).

Infant and child sex identification research further, has illustrated the source-filter controversy. Mothers were able to recognize the cry of their own infants but failed to identify at better than chance level. (Lanker, 1980).

Sachs (1973) found that 81% of adult listener identification were correct, further, analysis revealed that fundamental frequencies in males was significantly higher than in females. For vowels |l| and |u| lower formant frequency values were seen in male voice than in female voice. They concluded that judgement was not based on fundamental frequency, but, on the difference in formant patterns between boys and girls.

Dennis, Ingress, Gray, Weismer & Schucker (1980) in their study on sex identification have shown that vocal tract resonance characteristics makes the greatest contribution. In the absence' of fundamental frequency information.

A person's voice is a complex acoustic signal which encodes various kinds of information, among them some reflect the anatomy and physiology of the speaker due to large amount of speaker identity information in the speech signal speaker recognition can be carried by many means". (Corsi, 1982).

Voice identification can be considered to be a very old or very modern technique/process depending on the point of view from which it is analysed. Multiple methods of voice identification can be represented along a continuum that goes from very subjective to very objective. The oldest method (placed at the extreme subjective end of the continuum) would be listening to a talker

and recognizing him/her through familiarity with his/her voice. During the thousands of years, not much attention was given to this area. It was only in 1935's that scientists attempted to bring scientific insight into the modality of the process of voice identification.

**Subjective methods of talker identification and elimination :**

Aural examination of recorded voices and visual examination of speech spectrographs are considered to be subjective methods of voice identification, each within a different category of subjectivity. Aural examination of voices : A listener is asked to use long term memory or short term memory process to identify/eliminate an unknown talker as being the same as a past known one.

The first significant experiment done in the area of aural examination using the long term memory process was by McGhee (1937, 1944). She used 31 male and 18 female talkers, reading a passage of 56 words. 740 untrained listeners were used. Two sessions were conducted. During the first session, the listeners heard a talker behind a screen and during the second 5 talkers read the same passage. The listeners task was to identify the speaker of the first session. The 2nd listening session was spaced differently. Results indicated that the : a. Average percentage of correct identification varied from 83% to 13%. b. As time was increased (one day - 5 month lapse) between the 2 sessions, lower percentages were secured, c. Disguising the voice, reduced percentage of identification, d. Male and female voices were equally identifiable. Pollack, Pickett and Sumby (1954) also performed an experiment based on long-term memory. Three variables were investigated : duration of speech sample, filtering and whispering. Their findings are summarized as follows: a. Whispered speech reduced the percentage of correct identification by approximately 30%. b. Whispered speech samples must have a duration of at least 4 secs, (normal speech- 1 sec) to get correct

identification scores, c. For low pass and high pass filtering - identification performance is resistant to selective frequency; however filtering above 500 Hz and below 2000 Hz decreased the percentage of correct identification.

Coleman (1973) eliminated the influence of glottal source by using an artificial larynx with a fundamental frequency of 85 Hz. According to him, resonances of the vocal tract are the clues for voice identification rather than the glottal characteristics of the talker.

Researchers have conducted experiments using different methods of presentation of speech material. Stevens et al. (1968) presented the speech samples aurally through headphones and visually as spectrograms. Two kinds of experiments were carried out. 1. A series of closed tests in which there was a library of samples from 8 speakers and test utterances were known to be produced by one speaker. 2. A series of open tests in which the same library of 8 speakers was used, but test utterances may or may not have been produced by one of the speakers.

The results of the closed tests indicate that after 4 hours of exposure to the test situation the percent error in identification of speaker from isolated speech samples (words or phrases) was about 6% for aural presentation and about 21% for visual presentation. These scores depend upon the talker, the subject, and the phonetic content and duration of the speech material. For the open visual tests, appreciable number of false - acceptances (incorrect authentications) were made.

The results suggest the following :

1. Aural identification was more accurate than identification from spectrograms using a matching from-a-sample technique.
2. For visual identification, longer utterances increased the probability of correct identification.
3. It is easier to identify a talker when he utters a word containing a front vowel than when he utters a word containing a back vowel.
4. There are large differences in the ability of subjects to identify voices on either a visual basis or an aural basis.
5. Indirect evidence suggests that matching - from - sample technique in which the comparison items consists of several repetitions of the utterance by each talker leads to improved scores relative to the case in which only a single comparison utterance is available from each talker.
6. Authentication of voices is much poorer on a visual basis than on an aural basis.

They have suggested some variables which needed further probing.

1. The effect of more extensive training of subjects particularly for visual tests.
2. The advantages of using more than one standard utterance for each member of the ensemble of talkers.
3. The effect of using subjects working together in groups rather than individually.
4. The improvement to be achieved by combining aural and visual methods.
5. The resistance of both methods to mimicking.

Visual examination of speech spectrograms :

Speech spectrography consists of a display of the main parameters of a speech wave time, frequency and intensity. This operation was first performed for sustained vowels in 1900's using mechanical spectrographs such as the

Henrici Analyzer. In 1941, an electromechanical acoustic spectrograph project led by Ralph Potter was started at the Bell Telephone Laboratories. In 1944, Gray and Kopp found that spectrograms could be used for speaker identification. In addition to phonetic variations, spectrograms also portrays talker dependent features. Gray and Kopp coined the term 'voice printing' to designate the application of speech spectrograms to voice identification.

Kresta (1962 a) reexamined 'voice print' using spectrograms taken from 5 clue words spoken in isolation. The test was a closed typed one using contemporary spectrograms. A maximum of 12 known talkers were used in each trial. The examiners were asked to give a positive decision as to which of the known talkers were same as the known one. Training was given for one week. Results showed that the percentage of correct identification was better than 99%.

Stevens and Tosi supported these findings although they argued that error rates were higher than those reported by Kresta. Young and Campbell (1967) in their study on contextual influence on speaker identification employed five talkers uttering two words. They were used as known talkers. The closed type of test was used. 10 examiners received 2-5 hours of training prior to the examination. The words used were 'you' 'it' and were spoken in isolation. The correct identification was 78%. Then, words were extracted from sentences. The percentage of correct identification of words was 37.3%. They attributed this low score to coarticulation factor and to the difference in duration of word spoken in isolation and in context.



Kresta (1962), Prozanski (1963), Pollack, Pickett and Sumbly (1954) have shown that it was possible to obtain correct identification for the words spoken in isolation and in context. Bruce (1966) carried out an experiment similar to the above one. Six talkers were considered. The standard spectrograms consisted of ten key words spoken in isolation. One sentence containing all these ten key words was used. The observers task was to determine the speaker of the test utterance using the spectrograms. The error rate for this task was found to be 50%. Stevens et al. (1968) studied speaker identification by aural and visual examinations of spectrograms. Both the tasks were done separately. The task was a matching task which employed a closed set of eight talkers. The examiners were given a set of 8 standard spectrogram. They were then presented with unknown spectrograms to identify. The error of false identification ranged from 18% to 50% depending upon the utterance.

Hazen (1973) reported an experiment performed to determine the effects of context on speaker identification. Five words were extracted from the speaker's spontaneous speech. Seven team panels consisting of two examiners received few sessions of training before the starting of the experiment. Task to be performed was absolute identification or elimination of one unknown talker among the 50 known ones. The error ranged from 0% to 83.33%. The error was greater when the speech samples from different speech context were compared, than, when the samples of speech from the same contexts were compared.

An extensive study on speaker identification has been done by Tosi et al. (1972). The experiment was carried over a span of two years. A total of 34,996 experimental trials were performed by 29 trained examiners. This study had two stages : 1. To check out the finding reported by Kresta (1962). 2. To test the models including variables related to forensic tasks.

Each trial involved forty known voices in various conditions, with closed and open trials, fixed and random context, contemporary and non-contemporary spectrograms of 906 clue words spoken in isolation. The examiners were forced to reach a positive decision (identification/ elimination) in 15 minutes by visually examining the spectrograms. Results were graded on a four point confidence scale. The results confirmed Kresta's (1962) finding. Experimental trials correlated with forensic models (open trials, fixed and random content, non-contemporary spectrograms) yielded an error score of 6% for false identification and 13% score for false elimination. Examiners judged 60% of their wrong answers and 20% of their correct answers as uncertain which suggested that if they were allowed for 'no opinion' choice when in doubt, only 74% of the total number would have had a positive answer. A score of 2% for false identification and 5% for false elimination would be obtained.

Mani Rao and Agrawal (1984) have conducted an experiment to verify speakers identity by comparing the pair of spectrographs. Fifteen adult speakers and ten novice examiners participated in the experiment. The speech sample consisted of three English digits (one, two and zero). The examiners were required to watch the spectrograms of two speech samples in

terms of acoustic features and decide whether they belonged to the same speaker or not. Results showed that 10 novice examiners could correctly identify the speakers about 85% for male talkers and 72% for female talkers. They also carried out feature to feature analysis. Results showed, the relative importance of rank order of acoustic features for correct identification were different than those for correct elimination.

Latha (1987) studied speaker identification by verifying the spectrograms based on acoustic features and to identify the acoustic feature needed for verification. Words extracted from sentences were used. A total of 30 inter- speaker and 4 intra-speaker pairs and one pair for test- retest reliability were prepared. The three judges considered could identify the speakers correctly (95.5%). The acoustic features found to be helpful in verifying the speakers were : overall clarity, total duration of the word and duration of the individual phonemes, frequency range of burst, frequency range of noise, energy concentration, voice onset time. They suggest that by obtaining a weighting factor for each features, which the examiner can use for verification, speaker verification by spectrogram can be made more objective.

Combined aural and visual examination are more in vogue. In sum, subjective methods of voice identification might offer a reasonable degree of validity if properly applied to practical cases by trained examiners. Distortions produced by transmission and recording systems background noise, and psychological and physiological conditions of talkers will greatly decrease the percentage of cases in which a positive identification could be

reached. They will increase the percentage of no-opinion decisions or at the worst will increase the percentage of false eliminations.

Objective methods of voice identification are those in which a decision as to whether or not an unknown and a known voice belong to the same talker is produced by a machine, specifically a computer, rather than directly by a human examiner. Objective methods can be classified into two groups i.e., Semiautomatic and automatic.

Semiautomatic methods need a large and continuous interaction of an examiner with the computer. In the automatic methods human interaction is limited, usually it consists of preparing and inputting proper samples as well as interpreting output from the computer. Although some recent word recognition devices have demonstrated high success rates while giving real time operation, no systems are known which can maintain their performance in practical application. This is due in part to their inability to make allowance for human factors which become noticeable in live situations. There is a need to quantify the performance degradation due to these human factors. Other investigations related to voice identification have also been carried out the spectral and temporal properties of speech signals that distinguish phonetic categories can be substantially altered by factors such as phonetic content (Liberman et al. 1967), Stress (Klatt, 1976), vocal tract size and shape (Font, et al. 1973) and speaking rate (Miller, 1987a).

Consequently, changes in one or more of these factors can alter the perception and categorization of speech sounds. Sommer, Pisoni and Nygaard

used several different sources of stimulus variability within speech signals to see the effect on spoken word recognition. The effect of varying talker characteristics, speaking rate and overall amplitude, on identification performance was assessed by comparing spoken word recognition scores for contexts with and without variability along a specified stimulus dimension. Identification scores for word lists produced by single talkers were significantly better than for the identical items produced in multiple talker contexts. Similarly, recognition scores for words produced at a single speaking rate were significantly better than for the corresponding mixed rate condition.

Simultaneous variations in both speaking rate and talker characteristics produced greater reductions in perceptual identification scores than variability along either dimension alone. In contrast, variability in the overall amplitude of test items over a 30 dB range did not significantly alter spoken word recognition scores. The results provide evidence for one or more resource demanding normalization processes which function to maintain perceptual constancy by compensating for acoustic phonetic variability in speech signals that affect phonetic identification.

As evident from this study, basing our findings only on the results of perceptual (listening) findings would be inappropriate.

A study was done by Cole et al. where the subject spent 2000-2500 hours learning to read speech spectrographs. The subject's ability to identify the phonetic content of broad band speech spectrographs of unknown

utterances during 8 separate sessions of 4 hours each. The expert was presented with 23 spectrograms of English sentences and sequences of words and nonsense words, and 45 English words embedded in a known carrier phrase. The phonetic labels produced by the expert agreed with the phonetic labels produced by trained phoneticians (who listened to the speech) between 80% and 90% of the time, depending upon the scoring method used. When presented with words in a known carrier phrase, labelling performance was seen to improve to about 93%. A linguist presented with the phonetic transcriptions produced by the spectrograph reader was unable to identify all the words in 10 of 15 utterances and missed a single word in each the remaining five.

'Voice Prints', a technique based on traditional methods of speech spectrography is currently being used in criminal investigations and courts of law to identify speakers from recorded voice samples. Kersta (1962) argued the parallelism of spectrograms and finger prints. He demonstrated that contour spectrograms were more suited for this purpose. The contour spectrogram has amplitude and frequency dimensions like the bar spectrogram. The amplitude however is shown by seven quantized/or contour steps. The amplitude doubles with each inward progression from one contour to the next. He conducted an experiment in which high school girls were trained in spectrogram reading and then presented with spectrograms of 10 frequently occurring monosyllables. Tests were conducted in which these examiners were given a matrix of four voice prints for each speaker and they had to sort test of utterances into piles for each speaker. It yielded promising

results of 99%. When words were excerpted from the context of a cue sentence instead of spoken in isolation were used, the deterioration in error rate was merely 1%.

Young and Campbell (1967) examined the effect of taking words from sentences. They used the same words as Kersta and recorded them in isolation and in sentences. 10 observers, all familiar with spectrograms, were trained to point out visible clues like frequency, intensity, regularity of vertical striations. They found that it was difficult to identify the words in sentences, the correct identification being 37.3% and 78.4% in isolation and sentences respectively. They refuted the claim that voice was unique.

Tosi et al. (1972) conducted an extensive study over two years. 250 speaker's samples were identified by 29 trained examiners. Various conditions such as closed vs. open trials, contemporary vs. non-contemporary spectrograms, a few clue words, spoken in isolation, in a defined context and in a random control etc. were considered. Decisions were based solely on inspection of spectrograms. The examiners were asked to grade their degree of confidence in each decision on a four point scale. There was 6.4% false identification and 60% which were rated as uncertain by the examiners. They suggest that if in addition to visual comparison of spectrograms, the examiners were allowed to listen to known and unknown voices, the errors might be further reduced. For reasonable reliability, Tosi (1975) opined fulfillment of certain conditions. 1. Examiners should be qualified, with a training in phonetics and 2 year apprenticeship in field work. 2. They should avoid positive conclusion if the slightest doubt exists. 3. They should

be entitled to ask for as many samples of speech and as much time is needed.

With regard to the above mentioned, Holken (1974) speaks of : 1. Social relevancy of the problem. 2. The relations of voice prints to the larger issue of speaker identification and, 3. The differences between laboratory experiments and forensic investigations. 1) The "voice print" controversy does not exist simply as a scientific curiosity rather, it forms the basis on which an individual accused of criminal acts can be convicted or exonerated. Therefore, speaker identification must be considered with perspective of social implication involved. 2) In cases, the greater issue concerns the positive identification from their speech by means of any acoustic, temporal, perceptual etc. approach ( or combination and approach), possible in any or all situations. But many a times only a narrow approach is used. 3) There is no one single forensic model i.e. the forensic situation actually is made up of a rather large number of conditions that may vary either in presence or magnitude. A few of these complicating conditions consist of the possible non-contemporaneous aspects of speech samples, the psychological state (stress, fear etc.) of the talker, attempts at disguise noise in the transmission system, interface inadequacies within the system, system band pass and distortion, competing speech signals and so on.

Additional confusion could exist because a number of studies have been reported where one or more of the above variables have been included in



the experimental design and the data have been argued to be applicable to forensic situations.

The proponents of the "voice print" technique base justification of its use as an investigative tool on 2 arguments.

First, they contend that their procedures would be even more accurate in real life situations than in the lab. That is professional examiners would - a) be permitted all the time and samples they wished. b) be more serious about the laboratory subjects, c) have a greater degree of appropriate training than laboratory subjects. d) be permitted to have no opinions.

But in real life forensic situations, the impact of problems that face the examiner would be so profound that, in most cases, they would seriously outweigh the Osgood advantage. For e.g. : The evidence presented by 2 California "Voice Print" cases is reviewed.

In one, the defendant was accused of making a telephone call. The call was recorded on a 24 hour taperecorder with a very low signal-to-noise ratio and an apparent upper frequency limit of about 2300 Hz. spectrograms used for comparing the voices of the accused and the unknown caller were displayed. Due to distortion and lack of evidence, no decision was finalized and the case ended in a hung jury.

In another case, the defendant was accused of making a bomb threat. The court found the defendant not guilty and the evidence not reliable to this particular case due to : a) Mistakes and errors made in the preparation of spectrograms used in making the identifications b) Failure to ascertain the

existence of such errors, c) Demonstrated listening errors in court while under cross examination, d) Tentative mis-identification of the court ordered exemplar, e) Failure to maintain adequate records/logs during conducting of tests.

These two cases raise the issue of how to extrapolate from laboratory studies to forensic situations.

In 1973 Hazen's subjects received training equivalent to that received by experimenters in voice print identification training courses. Experimental tasks required subjects to identify unknown speakers from a population of 50 known speakers, by first eliminating all known speakers they were certain were not the unknown speaker and then attempting identification.

Ten attempts were made where the unknown and known speech samples of the same speaker were excerpted from the same phonetic contexts and ten attempts when they were excerpted from different contexts. Statistically significant difference in subject performance for these two contexts indicated that even the phonetic contexts cause spectral variation and should be considered as an important variable during voice identification.

The objections to the use of voice print techniques may be classified into three kinds concerning the interpretation of results from laboratory assessment, the procedures of decision making, and most fundamentally the nature of information on which those decisions have to be based.

Questions have been raised as to whether visual examination of speech samples gives more accurate results than aural examination. Young and

Campbell (1967) concluded that humans can extract more relevant information from the unprocessed acoustic signal than they do from a visual representation. Also, interpretation is very subjective.

Hollien (1974) states that if the proponents of "voice prints" are successful, a subculture would develop expressly for the judicial system, where certified professional examiners could testify in courts of law. Despite the fact that voice printing is very vulnerable to criticisms that it is subjective and unvalidated and its practitioners do not undergo objective, independent testing in realistic conditions, it is used in courts in 23 states in America and in Canada, Italy and Israel.

The present level of knowledge about personal voice characteristics, their recognition and how they change under different conditions is still rudimentary. This is a prerequisite for successful voice printing.

Experiments in the 1960s and 70's reviewed the scientific basis of speaker identification through use of speech spectrograms in connection with legal proceedings. Experimental results showed that error rates ranging from 6% to 65% false identification under various conditions were encountered in forensic situations. It was concluded that scientific information available at that time was not adequate to provide valid estimates of the degree of reliability of voice identification by elimination of spectrograms.

They suggested some experiments required to establish this technique on a scientifically solid basis. The key question - "What are the odds? What are the probabilities of correct, incorrect, or mixed identification of a person

through spectrograms? What are the probabilities under the particular set of conditions involved in forensic situations. Relevant conditions include the selection and number of persons represented by the spectrograms examined, the methods by which voice samples were recorded, the time and circumstances when the recordings were made and the confidence criteria of the examiner in making his decisions.

They wanted to see if the probabilities would qualify speech spectrograms as admissible for evidence in court.

In Tosi's and others method of analysis generally effect of five variables were seen. 1. Number of speakers in the known set. 2. Open vs. closed tests. 3. The context of speech materials (test words were either spoken in isolation or in sentences). 4. Certain characteristics of speech transmission system. 5. Contemporary vs. non-contemporary voice samples.

Identification errors are of two types : 1. Errors of false identification. The observer selects from the known set, a speaker who is not the person represented in unknown spectrogram. 2. Errors of false rejection or missed identification in open tests the observer wrongly decides that the unknown speaker is not represented in the known set.

In the forensic situation false identification could erroneously single out a particular individual as one of the suspected persons. Such errors take on special significance in that they relate to the possible conviction of an innocent person. Errors of false rejection on the other hand are important in

investigation work because they may lead to the elimination of a guilty person from consideration as a suspect.

Hazen conducted a study wherein the results of closed vs. open tests onwards in isolation words embedded in sentences were compared. Closed tests resulted in better identification when isolated words were used. On the open tests with words from conversations, false identifications were significantly less than false rejections.

Black, Lashbrook, Tosi, Nash, Oyer etc. conferred on the necessary conditions required by police department to obtain legal evidence through voice identification not present in laboratory studies are : a) A voice identification trainer must complete at least two years of supervised apprenticeship dealing with field cases, and possess academic training in audiology and speech sciences before applying for a test proficiency to become a professional examiner, b) A professional examiner in voice identification must be entitled to render five decisions after each examinations namely : positive identification, positive elimination, probability of identification possibility of elimination and no opinion, c) A professional examiner in voice identification must be entitled to use as much time and as many samples as he deems necessary to complete or examination. d) A professional examiner in voice identification must be held responsible for the positive decisions he may reach after his examination.

In order to ensure that these conditions are met in real-life cases, as well as to enforce a code of ethics, a non-profit International Association of Voice Identification was incorporated in 1971.

Authors are aware of the possible misuse of voice identification. However, they say that when evidence of voice identification/elimination is presented in a court of law, a complete information of the process used and the present limitations and restrictions of the method is given.

Endress et al. (1971) studied the changes using spectrograms, due to age, voice disguise and mimicking. The results showed, a) Shift in the frequency of formants to lower frequencies with increasing age. b) Spectrograms of text spoken in normal and disguised voice revealed strong variations in format structure, c) Result on mimicking the voice of well-known people suggested that though the imitators could vary format structure and fundamental frequency, they were not able to adopt these parameters to match those of imitated persons.

Hollien and McGlane (1976) studied disguised voice too. They employed positive decision criterion. The results indicated only 23.3% correct identification. The positive decision criterion was criticized by various investigators like Flosser (1971), Hazen (1973), Steven (1968), Young and Campbell (1967) and others.

Reich, Moll and Curtis (1976) studied the effect of selected vocal disguises on spectrographic speaker identification. Two recordings of 40 males were taken with a time gap of 40 weeks. Sentences with nine clue

words were spoken in six different modes, they are - Normal speech, old age, hoarse, hyper nasal, slow rate and free style.

Spectrograms were presented to four examiners who received 50 hours training prior to the starting of the experiment, they were not allowed for no opinion decision, and were asked to rate their confidence on a five point scale. Results indicated high percentage of correct identification when unknown and known undisguised voices were compared, than when undisguised known voices were compared with unknown voices disguised in any other mode.

The factors that may effect the speaker identification task are - 1. Different number of clue words used in speaker identification task. 2. Different number of utterances. 3. Different types of recording condition of the clue words: a) Speech samples recorded directly into the tape recorder. b) Speech samples recorded via a regular telephone line, in quiet environment or in noisy environment. 4. Different context of clue word used for speaker identification. a) Clue word spoken in isolation b) Clue word spoken in fixed context c) Words from random context. 5. Different number of known speakers included in each experimental trial i.e. 10, 20, or 40. 6. Intra-speaker variations 7. Awareness of the examiner.

Most of these factors increase the percentage of no opinion decision or possibilities of false elimination than the possibilities of false identification.

Tosi (1979) concludes .. . "considering all these variables it is difficult to develop a method that gives 100% correct identification. To insist that a

system of voice identification which should be resistant to disguises, noise or other distortions for practical use is unrealistic and unfair".

The vocal characteristics have their origin in the tone generated by the larynx include pitch and intensity. Certain phonemic voicing patterns such as duration of the voicing cue and the properties of a particular speaker's glottal waveform. Taken as a group, these characteristics are considered to make an important contribution to the identifiability of a speaker. The extent of this contribution, however has not been fully described. If, as has been suggested by Wolf, the fundamental frequency is the easiest acoustic property to modify for purposes of disguising the voice, it is important to know how much speaker identification is retained when the normal inter-subject differences in the laryngeal fundamental are eliminated. In any laryngeal tone, certain personal characteristics such as the shape of the glottal waveform which would be largely dictated by the properties of the individual's vocal folds could be presumed to remain. It might be inferred therefore that the loss of identifiability brought about by the equalization of all glottal source characteristics would represent the maximum degree of confusion that would result from attempts to disguise the voice by altering the vocal fundamental.

Houlihan (1977) performed a study on the effects of disguise on identification. She used 9 females and 5 males talkers, who read a short sentence in several conditions like undisguised, low pitch, falsetto, whispered and muffled voice. In a closed trial, 22 examiners who received training participated in the test. In each trial unknown spectrogram prepared with undisguised voice of each talker had to be matched against all other talkers



had of the same sex, in all voice conditions 100% identification for undisguised voice and 5% for whispered speech were reported.

In order to determine if speech spectrograms can be used to identify human beings, 2 questions must be studied : 1. Does the formant structure of phonemes uttered by a certain speaker change over a long interval of time, and 2. Can the formant structure be changed by disguise, or is it even possible to imitate the formant structure of another speaker?

Spectrograms of utterances produced by 7 speakers and recorded over periods of upto 29 years showed that the frequency position of formants and pitch of voiced sounds shift to lower frequencies with increasing age of test person. Speech spectrograms of texts spoken in a normal and a disguised voice revealed strong variations in formant structure. Speech spectrograms of utterances of well known people have been compared with those of imitators. The imitators succeeded in varying the formant structure and fundamental frequency of their voices, but they were not able to adopt these parameters to match or even be similar to those of initiated persons. 1. Do formant center frequencies and the mean pitch frequency of the phonemes uttered by a speaker remain constant during his life or do they depend upon his age? 2. Do formant center frequencies also remain constant if the voice is disguised? 3. Does an imitator succeed in adopting his manner of spectrogram to that of the person to be imitated so that the formant structure of his phonemes and the curve of his speech melody are similar?

Results :

Neither the formant structures of vowels and vowel like sounds nor the fundamental frequencies determined from spoken sentences are independent of age. On the contrary, it has been shown that with increasing age the points of concentration of the formants move towards lower frequencies. Moreover, the ability of controlling the pitch frequencies begin to decrease with increasing age. This allows the conclusion that human phonation may change predictably with increasing age.

There is a possibility of considerably changing formant structure of vowels and vowel like sounds as well as the mean pitch frequency by deliberate disguise of the voice. The attainable degree of such changes varies from person to person.

In the case of imitations, they try to adapt, the mean pitch frequency of their voice to that of the person to be imitated. In general, they do not succeed in striking the exact frequency position.

It has been shown that the sound of the voice and the mean pitch frequency above do not play a predominant role in the identification of the imitated speaker by other people. The following characteristics may then be of special importance, the curve of the intonation of the sentences, general habitual features such as loudness, richness of voice and speech dynamics, typical phrases and construction of sentences and dialect in which the text to be imitated is spoken. These features cause the listener to associate this imitation with the imitated person, but most of them are difficult to define and trace in speech spectrograms.

Fahnsworth took up a consonantal speeded classification task to assess the effect of talker variability across CV and VC environments. In addition, conditions were blocked by syllable type (CV or VC) or consisted of mixed sets of CV and VC syllables. Talker variability was manipulated by presenting stimuli spoken by a single talker or by many talkers. The results showed that the magnitude of the effects of talker variability was approximately the same when comparing performance across talker contexts for CV's and VC's. Also the talker variability effect was stronger under conditions where syllables were blocked. This, provides further information about the underlying mechanisms involved in processing perceptual variations in the speech signal.

Hollien et al. (1977) conducted 2 experiments in which long-term spectra were extracted from controlled speech samples in order to study the effectiveness of that technique as a cue for speaker identification.

In the first study, power spectra were computed separately for groups of male speakers under full band and pass band conditions, an n-dimensional euclidean distance technique was used to permit identifications. The procedure resulted in high levels of speaker identification for large groups, especially under the full band conditions.

In a second experiment, the same approach was employed in order to discover if it was resistant to the effects of variation in speech production (at least under lab conditions. Talkers were 25 adult males, 3 different

conditions were studied, (a) Normal speech (b) Speech during stress (c) Disguised speech

The results demonstrated high levels of correct speaker identification for normal speech, slightly reduced scores for speech during stress and markedly reduced scores for disguised speech. It would appear that the L.T.S can be utilized to identify individuals even in relatively large groups when they are speaking normally or under stress. LTS does not appear to be an effective technique when voice disguise is employed.

This approach may have some merit for use in applied situations or as one of the features in a multiple-vector approach.

An investigation led by Wolf (1972) for selecting acoustic parameters which help to distinguish speakers, motivated by known relations between the voice signal and vocal tract shapes and gestures was carried out. Only significant features of selected segments were used. A simulation of a speaker recognition system was performed by manually locating speech events within utterances and using parameters measure data these locations to classify the speakers. Useful parameters were found in fundamental frequency, features of vowel and nasal consonant spectra, estimation of glottal source, spectrum slope, word duration and voice onset time.

These parameters were tested in speaker, recognition paradigms using simple linear classification procedures. When only 17 such parameters were used, - no errors were made in identification from a set of 21 adult male

speakers. Under the same conditions, speaker verification error of the order of 2% were also obtained.

Speaker recognition and verification effectiveness of a set of 92 measurements were examined by Sambur (1973). The measurements included the formant structure of vowels, the duration of certain speech events, the dynamic behaviour of the formant contours, various aspects of the pitch contour throughout an utterance, formant band widths, glottal source "poles" and, pole and zero locations during the production of nasals and strident consonants. Linear prediction methods were employed in the analysis, and a probability of error criterion was derived to evaluate the speaker characterizing potential of the measurement. The experimental speech data were collected during 5 different recording sessions (the vast time gap being three and a half years between the original and least recording). The measurements that were found most useful were related to the nasals, certain vowel resonances, certain temporal attributes and average fundamental frequency. A speaker identification experiment using only the five best measurements resulted in only one error in the identification of 11 speakers for 320 test utterances.

Coleman carried out a study to provide information on two questions :

- (1) With what degree of accuracy can speaker identification be made in the absence of the information normally provided by inter-subject differences in the laryngeal fundamental.
- (ii) How comparable are male and female speakers under these experimental conditions.

Twenty normal spectrograph adults (10 male, 10 female) were taken. The sound source used as a substitute for normal tone was a Western Electric Company model with electrolarynx which produced a steady buff having a frequency of 85 Hz ( $\pm 3$  Hz).

The speech sample consisted of words. The samples were then paired with same or different samples. These were given to listeners to judge as same or different.

The results indicated that more than 90% correct identifications were possible.

This indicates that sufficient individuality exists in speech characteristics other than those associated with the glottal source to support speaker-pair discriminations with slightly better than 90% accuracy. This indicates that maximum reduction in speaker identification might be expected to result from attempts to disguise the voice by modifying the laryngeal tone with less than 10% accuracy). This study also says that female speakers may be expected to be more successful in disguising their voices than males. Males are said to differ more among themselves on the non-phonatory aspects of speech.

The effect of speaking rate and stress on the temporal and spectral quality of vowels in four adult male speaker was evaluated by Stark (1993). Conversational style speech was used which, four vowels in two target words were analyzed. The target words were produced in two different sentence stress conditions. Vowel durations were measured and formant values were

obtained at 1/4th, 1/2 and 3/4th points of the syllables. Rapid rate tokens were consistently shorter in duration shortening between stressed and unstressed words or vowels. Speakers were very consistent in their overall sentence compression, but word and vowel compression showed nonsystematic individual differences. Target under school (any deviation greater than one Bark from the stressed normal rate condition) was found in only one speaker for the first formant of one vowel. Formant movement from 1/4th to the 3/4th point was not affected by rate or stress in any speaker.

Ananthapadmanabha and Stevens (1991) studied the production of stop consonants. They say that the production of stop consonants produces several kinds of acoustic properties : (1) The spectrum of the initial transient and burst indicating the size of the cavity anterior to the constriction. (2) Place dependent articulatory dynamics leading to different time courses of the noise burst, onset of glottal vibrations and formant transitions. (3) Formant transitions indicating the changing vocal tract shape from the closed positions of the stop to a more open configuration of the following vowel. This study measured the relative contributions of these acoustic properties to the classification of the consonantal place of articulation using a semi-automatic procedure. The acoustic data consisted of a number of repetitions of voiceless unaspirated stops in meaningful words spoken by several female and male speakers. The spectra averaged over the stop release and at the vowel onset were used as the acoustic feature. Speaker independent and vowel independent classification was about 80% using either

the burst or vowel onset spectrum and a combined strategy led to a higher degree of accuracy.

Blumstein and Stevens (1980) attempted to determine whether just the onset of a synthetic CV syllable can provide cues to the perception of place of articulation for voiced stop consonants.

A series of listening tests with brief synthetic CV syllables was carried out to determine whether the initial part of a syllable can provide cues to place of articulation for voiced stop consonants independent of the remainder of the syllable. The data shows that stimuli as short as 10-20 msec, sampled from the onset of a CV syllable can be reliably identified for consonantal place of articulation, whether the second higher formants contain moving or straight transitions and whether or not an initial burst is present. In most instances, the brief stimuli also contain sufficient information for vowel identification.

Stimulus continua in which formant transitions ranged from values appropriate to [b] [d] [g] in various vowel environments and in which stimulus durations were 20 and 40 msec, yielded categorical labelling functions with a few exceptions.

On the basis of this study, the suggested the following hypothesis in the perception of speech. 1. In the stream of speech, abrupt onsets and offsets provide markers to indicate points in time where acoustic information relevant to consonantal place of articulation is sampled. 2. This acoustic information resides for the most part in 10- 20 msec. time interval immediately adjoint to the onset or offset. 3. The gross shape of the spectra



in this region provides the essential perceptual cues for place of articulation across vowel contexts.

Stevens et al. (1992) - several types of measurements were made to determine the acoustic characteristics of those that distinguish between voiced and voiceless fricatives in various phonetic environments. The selection of measurements was based on a theoretical analysis that indicated the acoustic and aerodynamic attributes at the boundaries between fricatives and vowels. As expected, glottal vibration extended over a longer time in the obstruent interval for voiced fricatives than for unvoiced fricatives and there were more extensive transitions of the first formant adjacent to voiced fricative than for the voiceless cognates. When two fricatives with different voicing were adjacent, there were substantial modifications of these acoustic attributes, particularly for the syllable final fricative. In some cases, these modifications lead to complete assimilation of the voicing feature. Several perceptual studies with synthetic vowel C-V stimuli and with edited natural stimuli examined the role of consonant duration, extent of location of glottal vibration and extent of formant transitions on the identification of the voicing characteristics of fricatives. The perceptual results were in general consistent with the acoustic observations and with expectations based on the theoretical model. The results suggest that listeners base their voicing judgements of intervocalic fricatives on an assessment of the time interval in the fricative during which there is no glottal vibration. This time interval must exceed about 60 msec. if the fricative is to be judged as voiceless, except that a small

correction to their threshold is applied on the extent to which the first formant transitions are truncated at the consonant boundaries.

Jonathan (1994) considered 1946 syllable initial and 2848 syllable final nasal consonants taken from continuous speech data. Relational information in the acoustic waveform is based on difference spectra, in which spectral information in the vowel is subtracted from spectral information in the murmur, and on combined spectra in which classifications are made from combinations of murmur and vowel spectra. These two kinds of relational spectra are compared with static spectra, in which single spectral slices are taken from either the murmur or the vowel (contrary to recent theoretical predictions), difference spectra are shown to perform more poorly than same kinds of static spectra. However, since classification scores from combined spectra are better than from either static or difference spectra, cues to nasal part of articulation can nevertheless be defined as rational. In the best scoring, combined spectra classification scores on open test are just under 94% correct for syllable initial nasals and just under 82% correct for syllable final nasals. The high classification scores show that there is considerable information in the acoustic waveform for identifying nasal place of articulation from continuous speech data.

Glass et al. (1984) attempted to qualify the temporal and spectral characteristics of the nasal consonants in American English. 200 words with nasal consonants in different position and clusters were taken. The analysis focused on the static characteristics of the nasal murmur, the effect of nasalization on the spectral shape of vowels, and the properties of the

transitional region between the nasal consonant and the adjacent vowel. The results suggested that :

- (i) The duration of the nasal murmur is strongly influenced by the environment in which it appears,
- (ii) For a given speaker, the spectral shape of the nasal murmur is relatively unaffected by the phonetic environment,
- (iii) For a given speaker, the spectral shapes of nasal murmur are very similar for all nasal consonants.

The results of numerous studies implied that speakers can be recognized provided inter-speaker variability is greater than intra-speaker variability. Techniques which would facilitate this would improve the reliability of speaker identification.

Su, Li et al. conducted a quantitative study of coarticulation of nasal consonants with the vowels following them in isolated the ' cvdl utterances was studied. The spectral differences between the mean spectra of nasal followed by front vowels, and those of nasals followed by back vowels are used as the acoustic measure of the coarticulation of /m/ and/n/ with the following vowel /v/. The coarticulation between /n/ and /v/ was found to be only one-third of that between /m/ and /v/. The coarticulated nasal spectrum particularly between /m/ and /v/ was found to have strongly idiosyncratic characteristics which are not likely to be modified in natural speech. A method was developed by which the coarticulation between /m/ and/v/ was taken as the acoustic clue and the speaker was identified by use of a correlation decision criteria. Coarticulation was found to give more reliable cues than the nasal spectrum alone, which had earlier been found to be one of the best acoustic cues for identifying speakers.

Based on a study by Rabiner and Wilpan (1974) it was seen that speaker trained isolated word recognizers had notable success. The training generally involved a single (or sometimes 2) repetitions of each word of the vocabulary of the talker. Word reference templates are then formed directly from the replicates. In recent work, it has been found that statistical clustering procedures provide an efficient way for determining the structure in multiple replications of a word by different talkers. Such techniques were used to provide a set of reference templates based on clustering results. It is shown that significant improvements in recognition accuracy are obtained when using templates obtained from a clustering analysis of multiple replications of a word by the designated talker.

Green, et al. (1984) - 8 observers are given training for a two month period, at the end of which they could successfully identify 50 PB words of a single speaker. Generalization tasks were carried out with different speakers and a novel set of words. High levels of accuracy was found in identifying the visual displays protocol analysis revealed that the subjects were able to extract features from the spectrograms that corresponded in many cases to well known acoustic phonetic features (visual correlates of criteria! features) even though they were not explicitly trained to do so.

Inconsistent results have been obtained from studies in which the effects of phonetic contexts on identification accuracy were investigated. Kersta compared the ability of subjects to make identifications using single words under both isolated and contextual speech conditions. Error rates between these conditions differed by less than 1% for contextual condition. It

was considered that phonetic context had negligible effect on identification accuracy.

Steven et al. investigated the ability of subjects to make speaker identification spectrographically and compared it to their ability to make identification aurally. Using spectrograms, error rates varied widely depending upon the conditions. They observed that the mean error rate decreased for approximately 33% to 18% as the duration of speech sample increased from monosyllabic words to phrases and sentences. Subjects consistently achieved lower error rates when identifying speakers aurally, rather than spectrographically. There are at least two contextual factors that may decrease one's ability to make a correct identification. 1. The shorter duration of words spoken in context as opposed to isolation provides less acoustic information. 2. The spectral characteristics of speech samples are altered by the coarticulatory forces involved in producing spectral variations that caused Kersta to conceive of a file card system.

By filing the spectrograms of two separate utterances of certain cue words for known speakers, it was hoped that the effects of contextual variation could be minimized. The two specific spectrogram of each word chosen for filing could be the two on hand that are judged visually to be most dissimilar. Supposedly, this would afford an examiner an indication of a speakers expected range of contextually caused variability for selected words. Kersta concluded that 4 or 5 samples of the same word would be sufficient to get a fairly good indication of a speaker's range of variability.

This spectrogram filling system was also conceived as a population reduction method that, when used in conjunction with a speaker classification system, might serve to reduce a large speaker population to a small number of "suspects". The aim would then be to obtain additional speech samples from the suspect speakers prior to making further identification decisions.

Because the ability of this filling system to meet these aims has not been tested, the present study was designed. Its purpose were to determine whether the system could - 1. Minimize the effects of contextually caused spectral variation. 2. Serve as an effective absolute identification tool. 3. Serve as an effective population reduction tool.

Subjects received training to identify\* unknown speakers from a population of 50 known speakers by first excluding all known speakers they were certain of, and then attempting absolute identification or elimination. Attempts were made under five experimental conditions created by combining two variables, phonetic context and inclusion of the unknown speaker in the known speaker population. The data show that the system tested does not effectively reduce the effects of contextual variation, and cannot be used for either absolute identification/elimination or population reduction. The data suggest that the value of spectrograms for speaker identification purposes is limited to use as a investigative aid and then only if speech samples are of similar context and adequate duration are compared.

The ability to identify talkers from monosyllables spoken in a context was examined. Kersta's method of visually comparing spectrograms was employed. Ten observers were trained to identify five talkers from spectrograms of two words spoken in isolation.

The experimental task required the observer to identify the some talkers from the same words spoken in different contexts. The correct rates for the training task (78.4%) could not be reproducing in the experimental task (37.3%). The results were interpreted to indicate that different contexts decrease the identification ability of observers because : a) The shorter stimulus duration of words in context decreases the amount of acoustic information available for matching, and b) The different spectrographic portrayals introduced by different phonetic contexts outweigh any intra-talker consistency.

Santen studied and gave a description of contextual factors affecting duration.

Two natural speech data bases produced by male and female speakers were analysed. Large quantity of data (50,000 manually measured segmental duration) made it possible to perform a detailed analyses of the effects of several contextual factors, including lexical stress, word accent, the identities of adjacent segments, the syllabic structure of a word and proximity to a syntactic boundary. Among the key results were the following : 1. The contextual factors accounted for upto 90% of the variance, and reduced the within vowel standard deviation by a factor of 3. 2. There were complex

interactions between factors in particular between boundary proximity and post vocalic consonant identity and between lexical stress and syllabic word structure. 3. The effects of adjacent segments were reducible to the effects of voicing and manner of production, effects of place of articulation were negligible. 4. Proximity to a boundary should be measured in terms of syllabic and segmental position, not in terms of the sum of the intrinsic duration of segments between the target and the boundary.

Klatt (1974) used broad band spectrograms and the sonorant consonants /w, r, l, y/ observed in five sentences, which were read and recorded on two separate occasions by 7 speakers. Formant frequency motions in sentence contents have been compared with data in the literature on sonorants in citation form utterances. Results indicate that in stressed syllables, prevocalic and post vocalic allophones are similar in formant target values to corresponding citation form data, though initial allophones have somewhat less extreme formant targets than previous data would imply. In unstressed syllables, sonorant segments were shorter in duration and displayed significant coarticulation in the form of substantial formant target undershoot. Several phonological recording rules influence the acoustic realization of sonorant segments in consonant cluster sequences. Speaker differences in the implementation of optional word boundary and junctural cues also have an effect on sonorant clusters.

Zue (1979) in order to assess the role of syntactic, semantic and discourse knowledge in spectrogram reading recorded three short stories and speech spectrograms were made of the individual sentences of each story.



The stories were presented one at a time to an expert spectrographic reader who is instructed to read each word story word-by-word without writing down segment labels. There were totally 370 words and 612 (91%) were correctly identified. Further analysis reveal that many common syllables were immediately recognized as complete patterns (eg. "ment", "tion") and the use of content to recognize words from partial information was evident in many cases.

Thus the review of literature shows that there are three major variables related to (1) Speaker (2) Transmission and recording (3) Procedures used in analysis and identification. Among the variables related to speaker, the temporal and spectral aspect of speech has been found to be an important variable. This varies within subject i.e., on repeated utterances show a variability) and across the subjects. Further, it has been found that majority of the workers in the field of speaker identification have used word duration, vowel duration, burst duration, closure duration, voice onset time, fundamental frequency, intensity, formant frequencies, transition formants, etc. for the purpose of speaker identification.

Therefore, it was felt that it would be useful to study the inter-subject and intra-subject variability with reference to the above parameters. The present study also provides the range of variability in terms of inter-subject and intra-subject variability.

### CHAPTER - III

#### METHODOLOGY

The study was conducted to find out the variability in the measurement of the following parameters as they have been used in the process of voice identification.

Parameters studied were word duration, vowel duration, burst duration, lag VOT, closure duration, lead VOT, frication duration, fundamental frequency, intensity, formants:  $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$  formant transitions in terms of duration extent and speed.

**Subjects:** Five male subjects with age range between 20 - 30 years were selected. The main criteria for selection being that all the subjects had normal speech, voice and language and could read English fluently. Subjects with low-pitched voice were preferred, as it would be easy to read spectrogram of voices with low pitch.

**Test Material:** Test words considered for analysis were embedded within sentences. The following were used as test sentences.

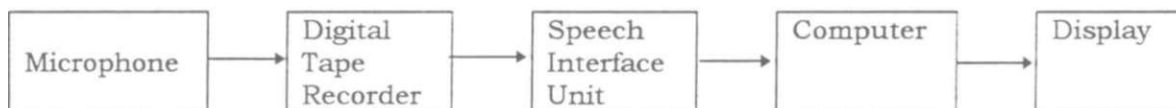
1. Knock eight times, keep the bag and go away.
2. I will call you tonight and give you further information.
3. Come and meet me outside the hospital.
4. Don't come with any one else.
5. At eight 'O' clock, come near the temple with the money
6. Bring a suitcase with rupees eight lakhs in it.
7. Give the suitcase to them

8. Put the money in a suitcase.

The following words were chosen from these sentences - 'come', "and" "the", "with" , "eight", "suitcase".

**Equipment:**

1. Digital tape recorder (Sony Digital Audio Tape Deck DTC - 59ES)
2. Microphone (33 - 992 A)
3. 3 VSS - SSL software program was used (voice and speech system - Bangalore)
4. PC - 80 - 386
5. DSP - Sonograph - Model 5500 (Kay Elemetrics)



Block diagram of the arrangement of instruments used for recording and analysis.

**Instructions:** The subjects were requested to read the sentences one after the other after the experimenter signals them to do so. The subjects were instructed to read the sentences in a natural way as far as possible.

**Procedure:** Before recording, the subjects were given the test sentences so as to familiarize themselves with it.

During the recording sessions, the distance between the speaker's mouth and the microphone was kept constant i.e., six cms. from the mouth of the speaker. The test samples were collected from three sessions First session: Each subject was asked to read out the eight test sentences. This was recorded. This constituted of one sample. The same procedure was repeated and sample two was obtained. Similarly, a third sample was also collected. Session 2: The entire procedure used during session 1 is repeated

here after an interval of two days. Session 3: the samples are recorded as mentioned for session one and session two with an interval of two days after session 2.

The recorded speech was then transferred from the digital tape recorder onto the computer through the speech interface unit (SIU) using the line-feed method. The signal from SIU was digitized at a sampling rate of 16 kHz using a twelve bit analogue - digital (A - D) and digital - analogue (D - A) convertor housed within the computer. The software program "record" provided by voice and speech systems (VSS) was used. The digitized signals were stored on the hard disk of the computer with individual file names for each sample of 8 sentences.

Using the program "display" of SSL (VSS), each sentence was displayed and the test words were segmented from the sentences. These were stored again as individual files for further analysis. These test words were selected from the middle four sentences (out of each sample of eight sentences) leaving the first two and last two sentences. The first two sentences were used as trials or carrier sentences. This is true for all subjects.

Thus, 18 samples of 6 test words (6 words x 3 trials for each subject) were collected from the recordings of the first session. The same procedure is carried out for all the subjects.

Similarly, 18 samples of six test words for each subject for each

|a:], |i:], | |u:], | $\hat{e}$  |, | | on to and three were obtained. The vowels were analyzed and measured from mentioned above.

The words were analyzed for the following parameters.

1) Word duration was defined as the time in milli-seconds between the onset and offset of the phonemes of a word. The word duration was marked from the beginning of striations to the end of the striations as depicted in the figure

**2) Vowel duration** was defined as the time in milli-seconds between the onset and of the vowel within a word. As seen in the figure - B, the vowel duration was measured from the beginning of the occurrence of regular striations to the end of regular striations indicating vocal fold vibration.

**3) Burst Duration:** Was defined as the time in milli-seconds between the onset of sudden voice bursts till its offset. The figure - C shows the duration for the sound "come" from the onset to release of the burst.

**4) Voic onset time** was defined as the time in milli-seconds between the offset of the burst of consonant to onset of vocal fold vibration. The figure - D, depicts the interval between release of the stop burst and the appearance of periodic modulation voicing) for the word "come".

**5) Closure Duration was defined** the time in milli-seconds from the offset of vocal fold vibration to the burst. As seen from figure - E, the duration from the fading away of striations to the burst is measured as the closure duration.

**6) Fricative Duration** was defined the time in milli-seconds between the onset and offset of striations on the wave form. As shown in figure - F, the vertical striations on the wave form indicate frication.

**7) Fundamental Frequency and Intensity :** As displayed on the screen when the word was fed using the Inton program.

**8) Formant green signal (F<sub>1</sub>, F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub>)** The first four formants (F<sub>1</sub>, F<sub>2</sub>, F<sub>3</sub>, F<sub>4</sub>) for each vowel were measured from the spectrogram display with sectioning on the screen of the computer. Formant frequency estimates were made by measuring the midpoint of the visible dark bands of energy

appropriate to the first four vowel resonances. The measurements were made at a comparatively steady state portion of the vowel (Refer Figure - G).

**9) Formant Transition was defined as** the change in formant pattern typically associated with phonetic boundary. These parameters were measured as shown in the figure - H.

**(a) Duration of formant transition** was defined as time in milli-seconds from the point the formant started rising or falling to the point it stopped.

**(b) Extent of formant transition** was defined as the distance for which the formant existed.

**(c) Speed of formant transition** was obtained by dividing the value obtained from formant transition excursion by the value obtained from the formant transition duration.

The reliability of testing was checked by randomly selecting five words from the samples and matching the values with the previously obtained ones.

Using the definitions presented above, the parameters : word duration, vowel duration, burst duration, voice onset time (VOT), lead VOT, closure duration, frication duration, fundamental frequency, intensity, formant frequencies ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ ) and formant transition in terms of duration, extent and speed were obtained for each subject.

The data collected was further subjected to statistical analysis. Descriptive statistical procedures were used for this purpose. The results have been discussed in the next chapter.

FIGURE : A

WORD : "COME"

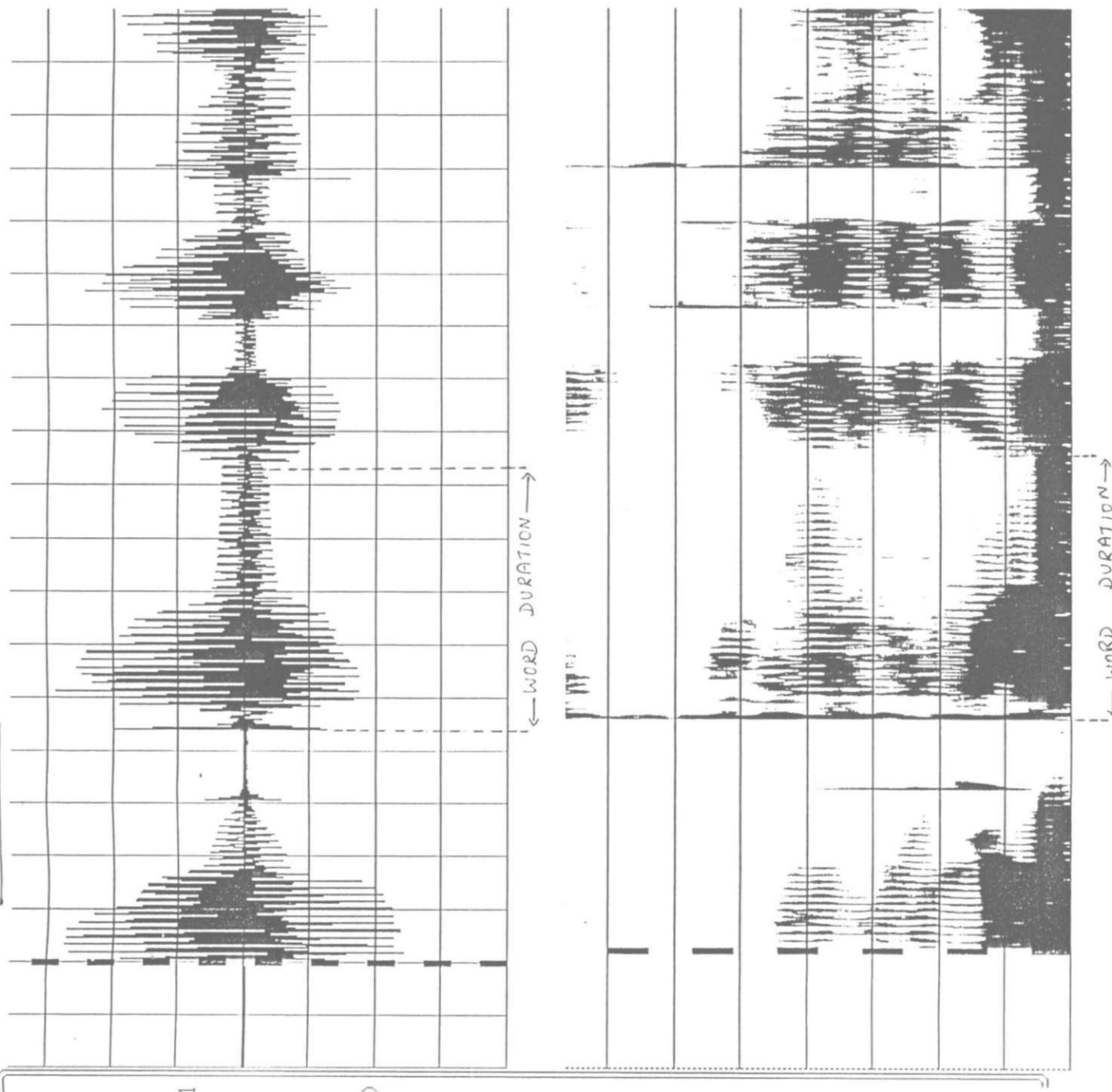
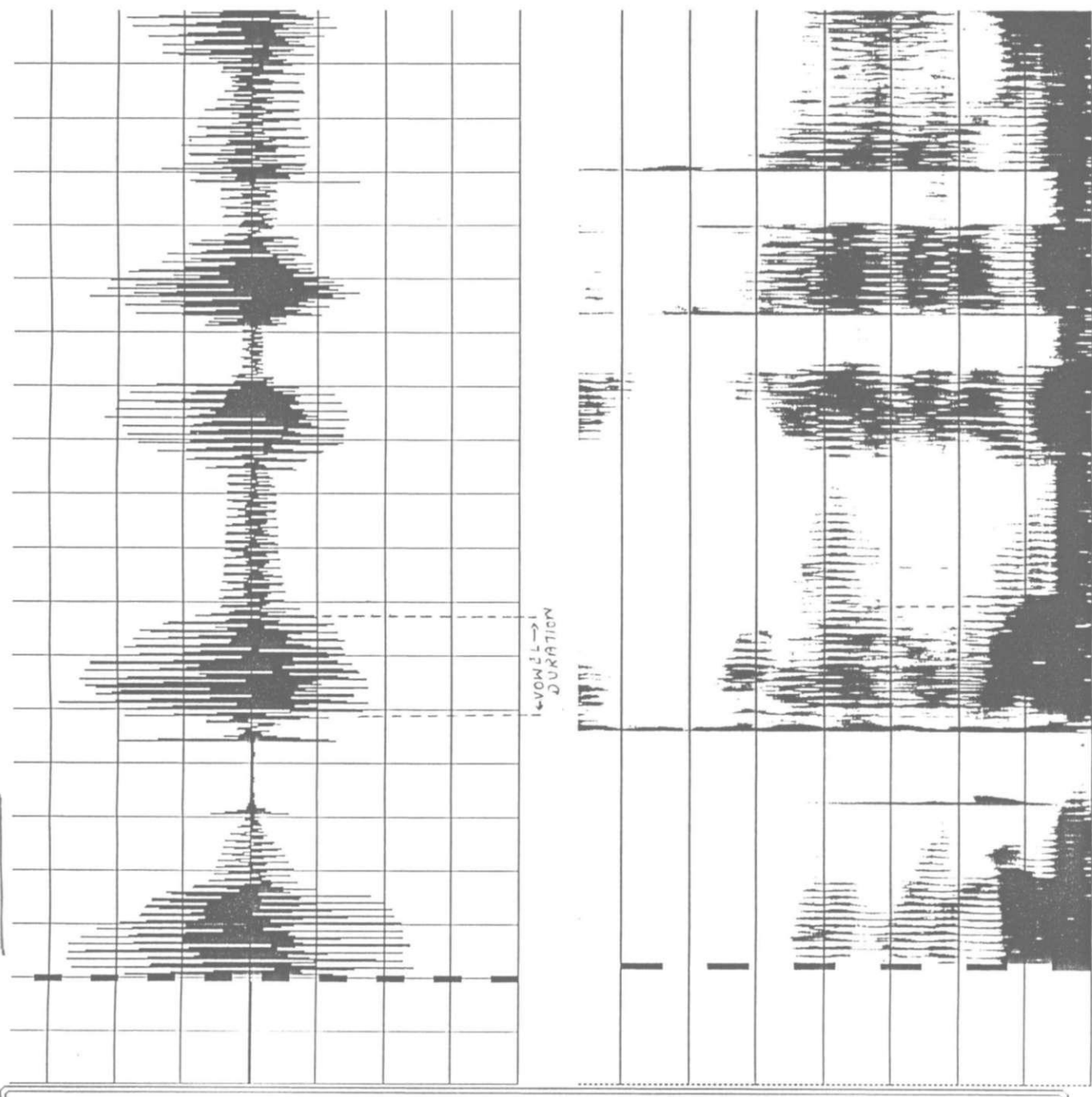


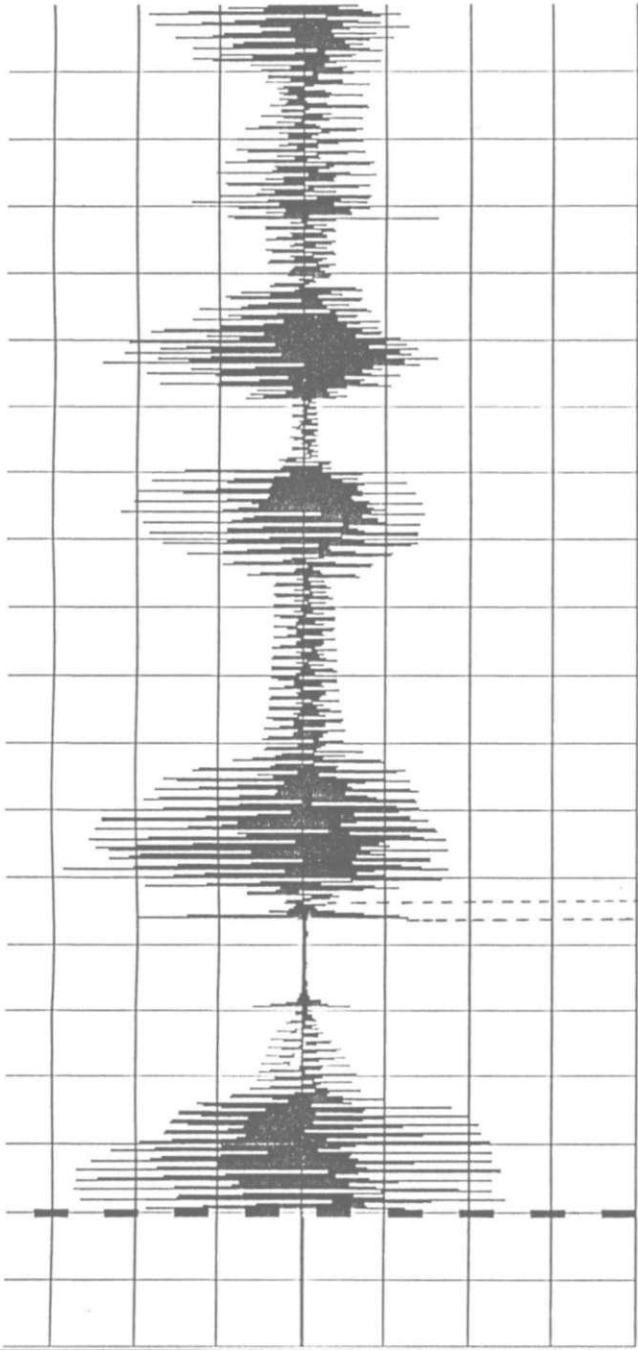
FIGURE : B

WORD : "COME"

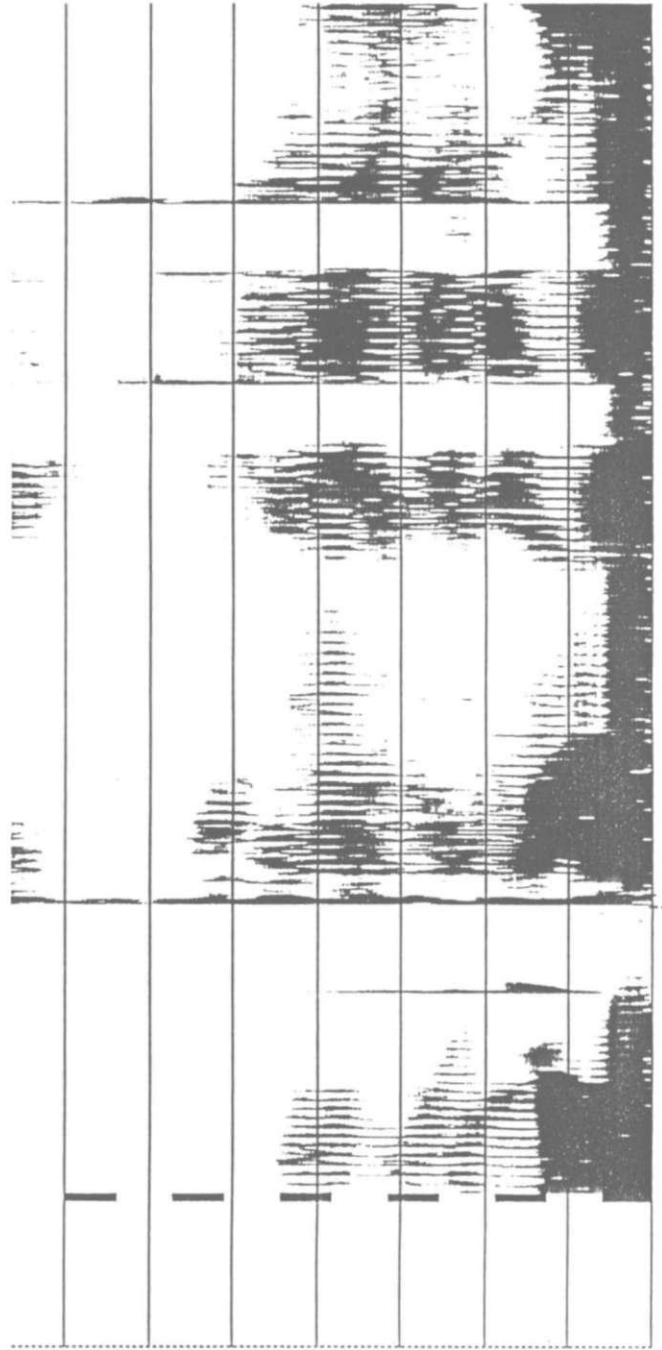




WORD : "COME"



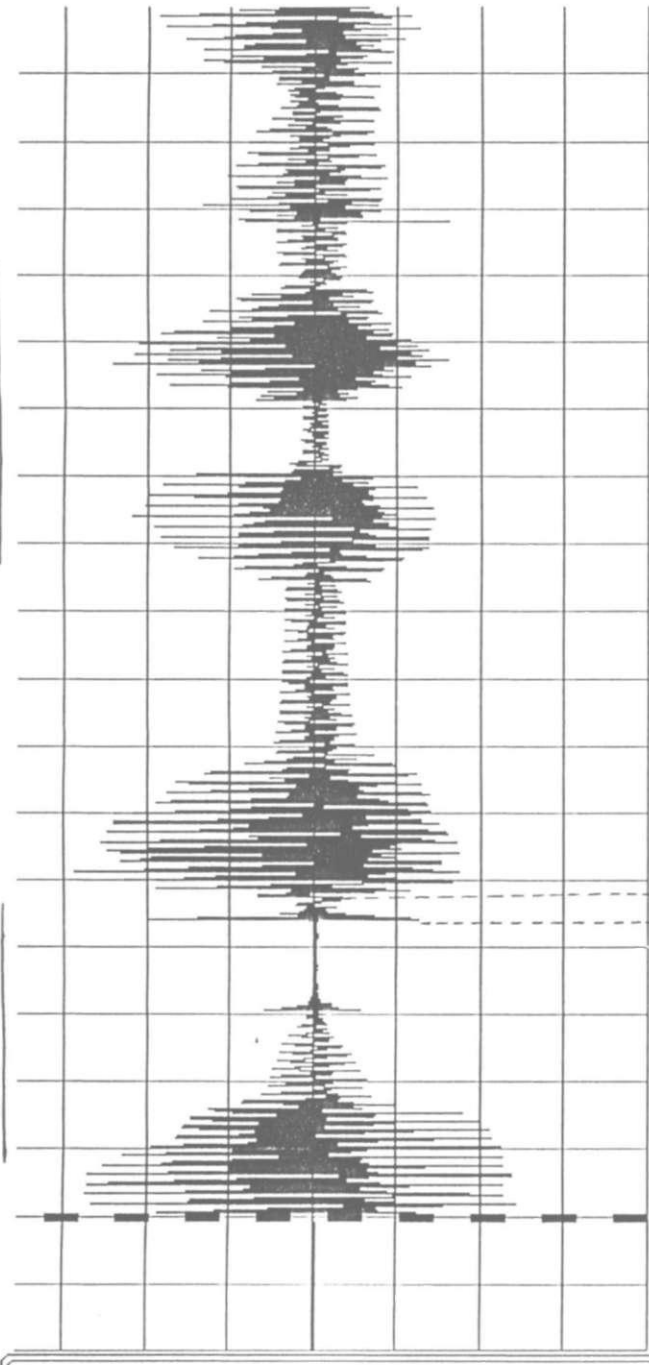
BURST DURATION



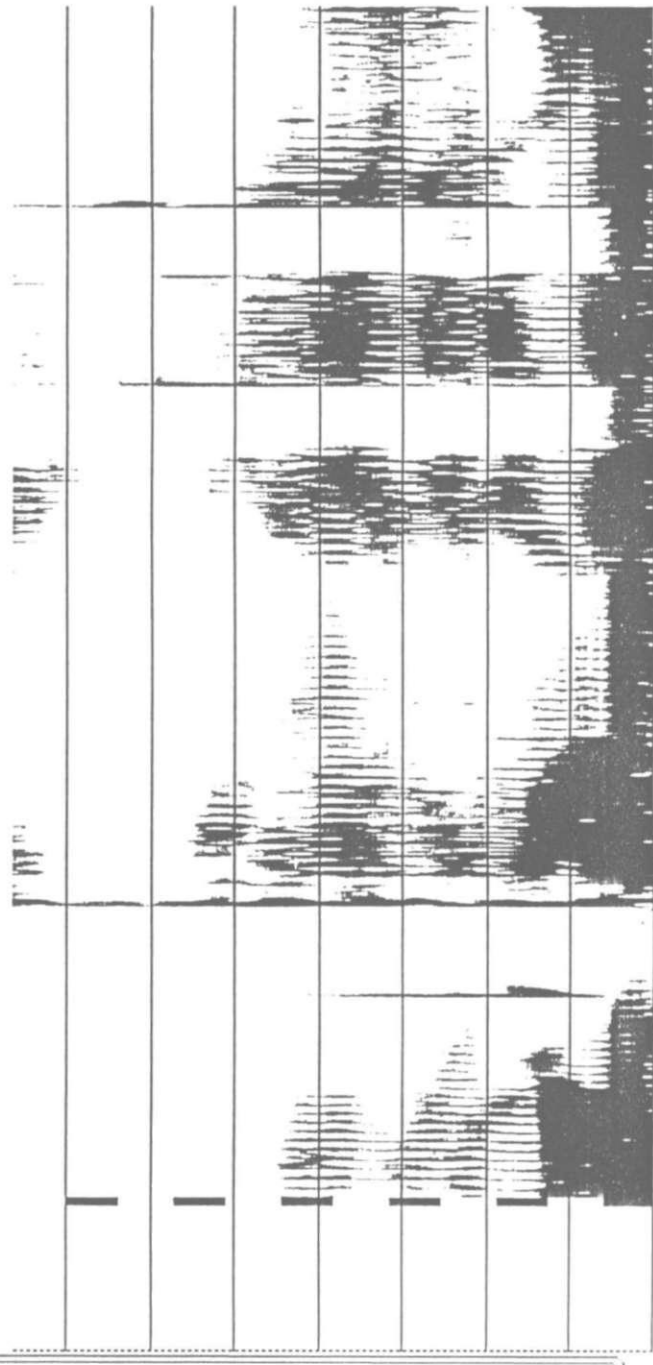
BURST DURATION

WORD: "COMÉ"

FIGURE: D



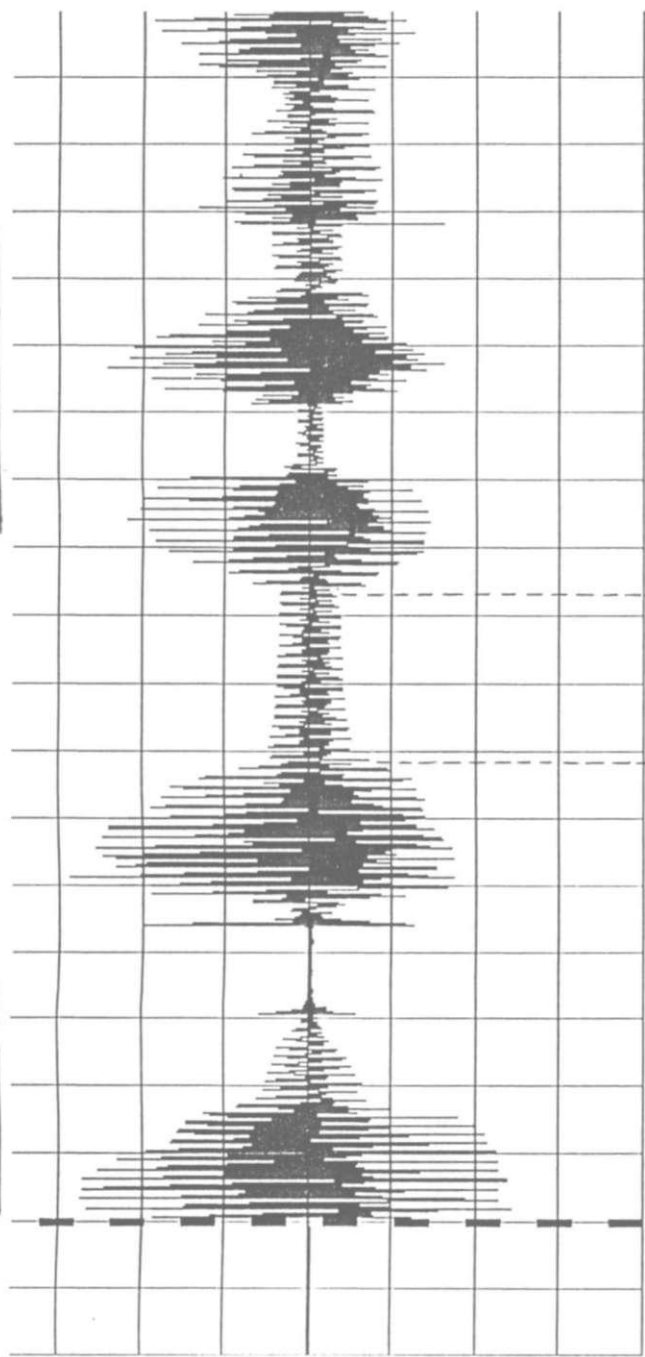
VOICE ONSET TIME



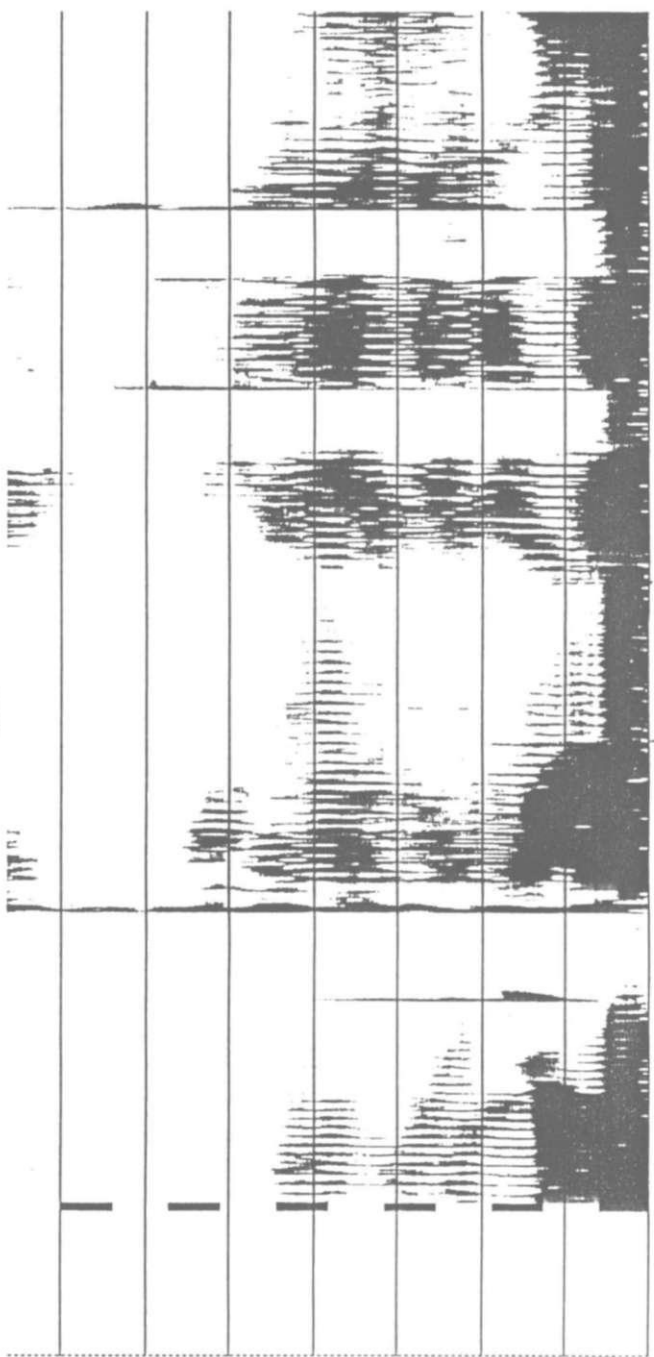
VOICE ONSET TIME

WORD: "COME"

FIGURE: E



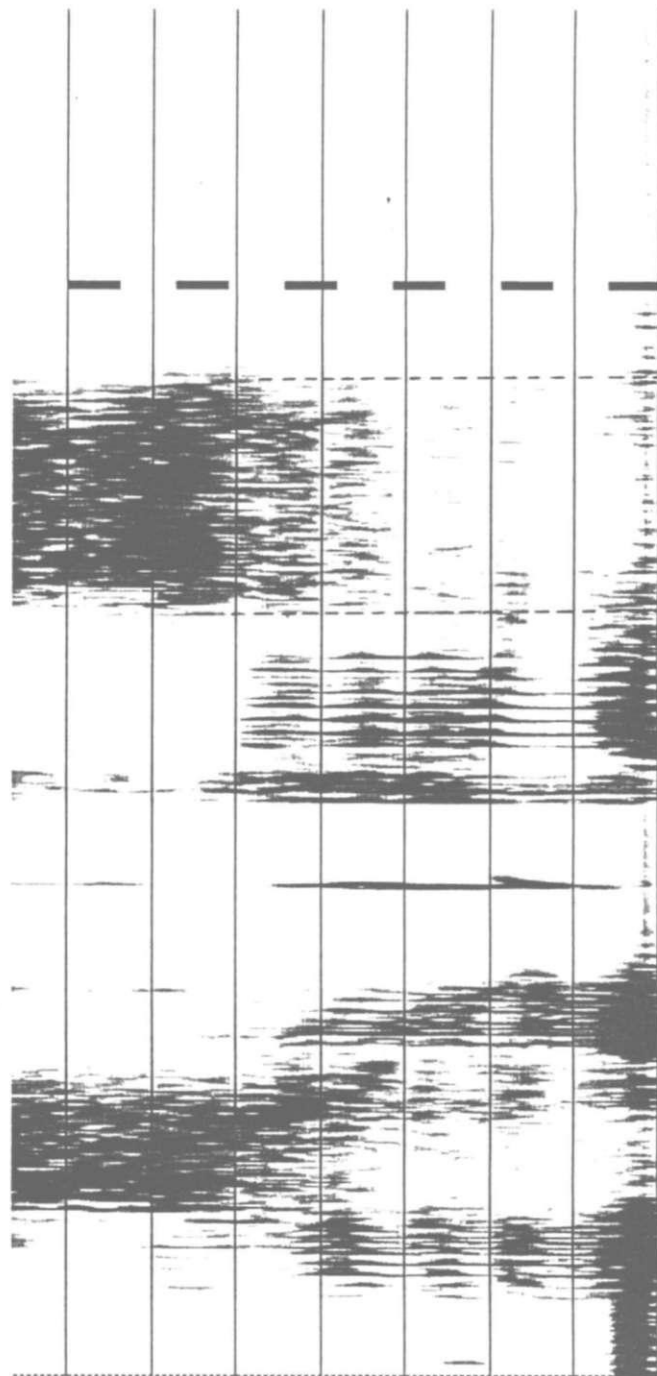
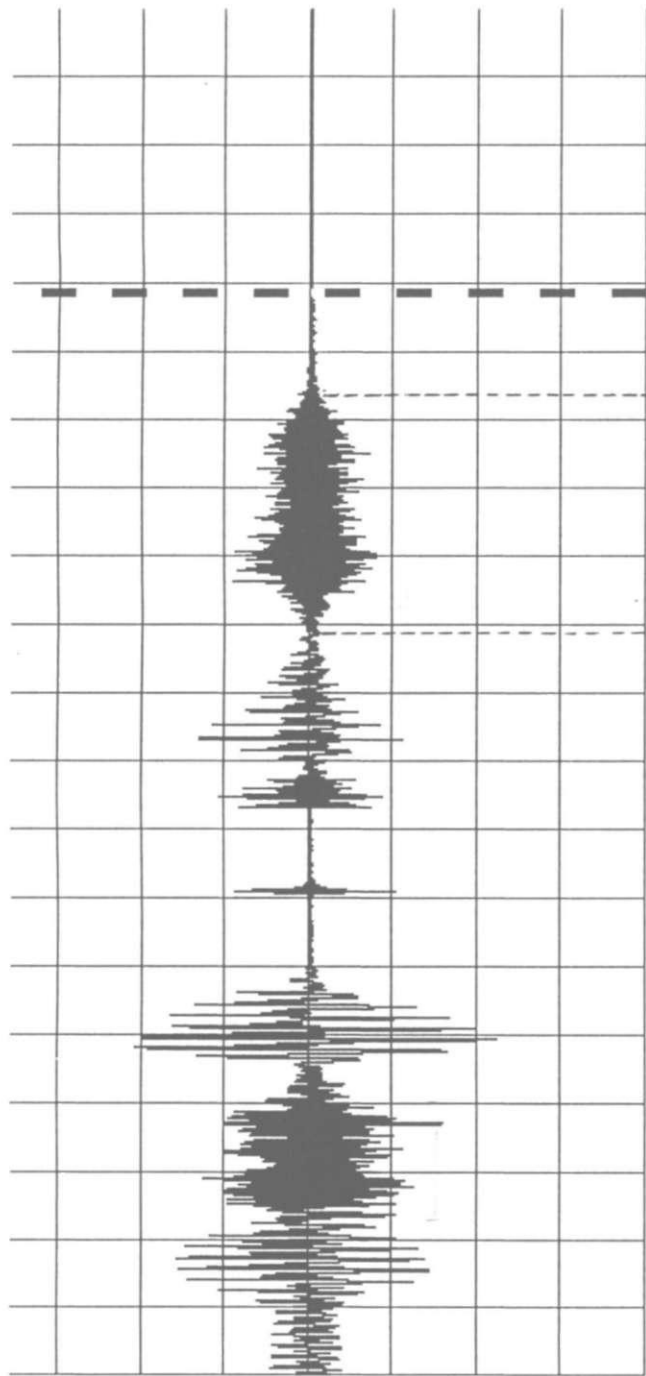
CLOSURE →  
DURATION



CLOSURE →  
DURATION

WORD: "CASE"

FIGURE: F



WORD: COME

FIGURE: 6

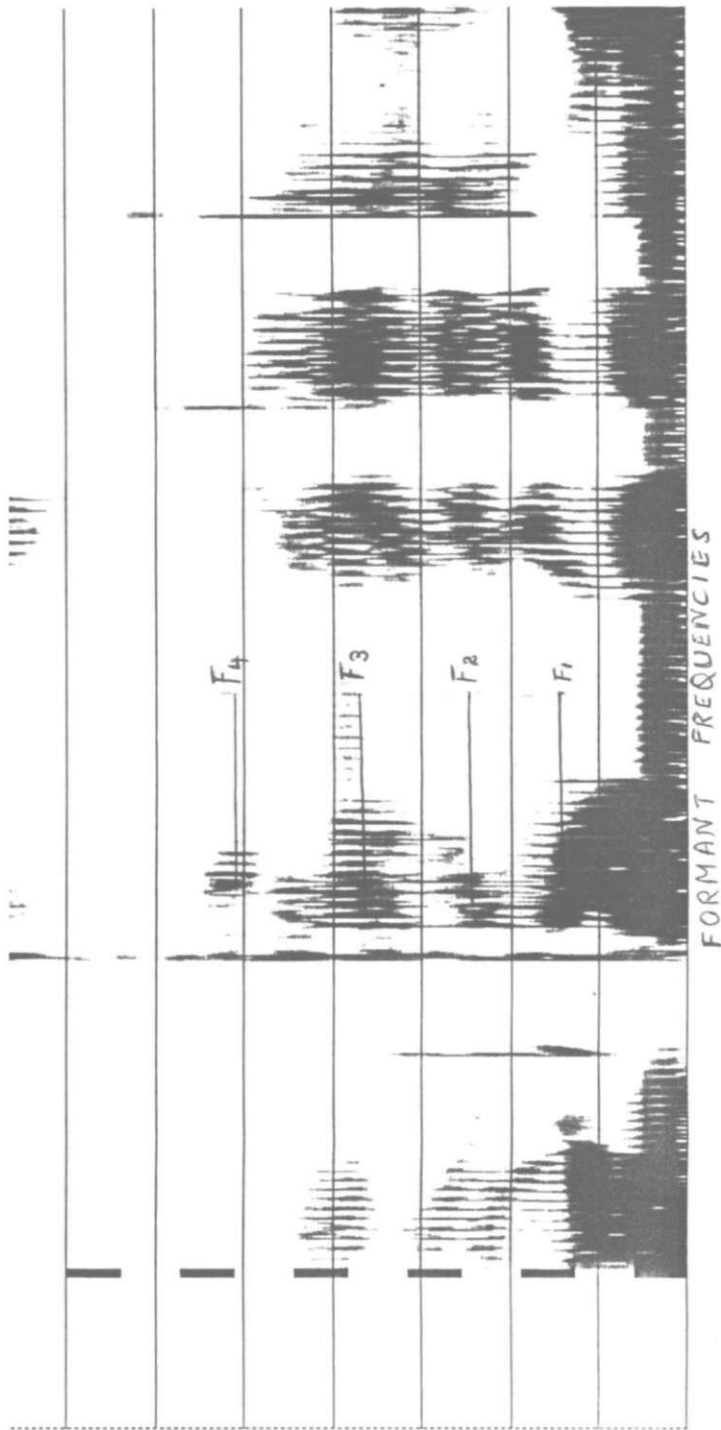
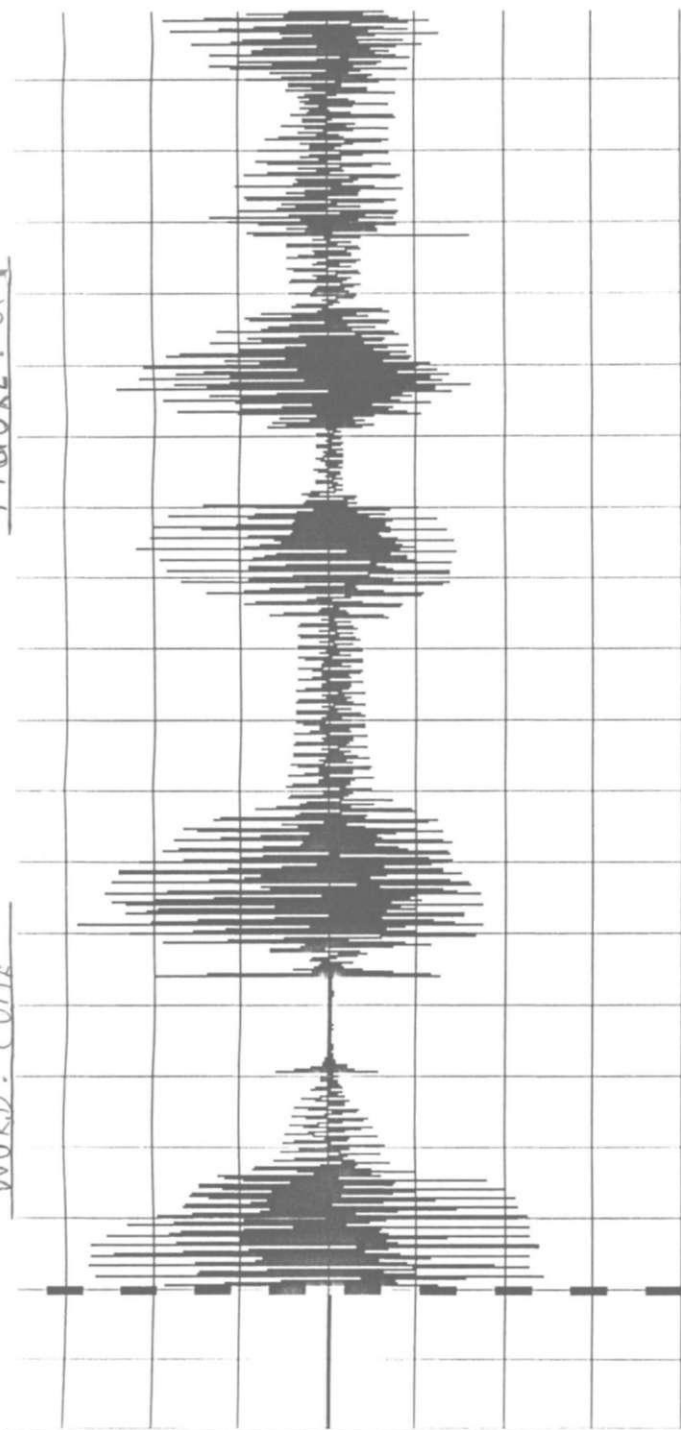
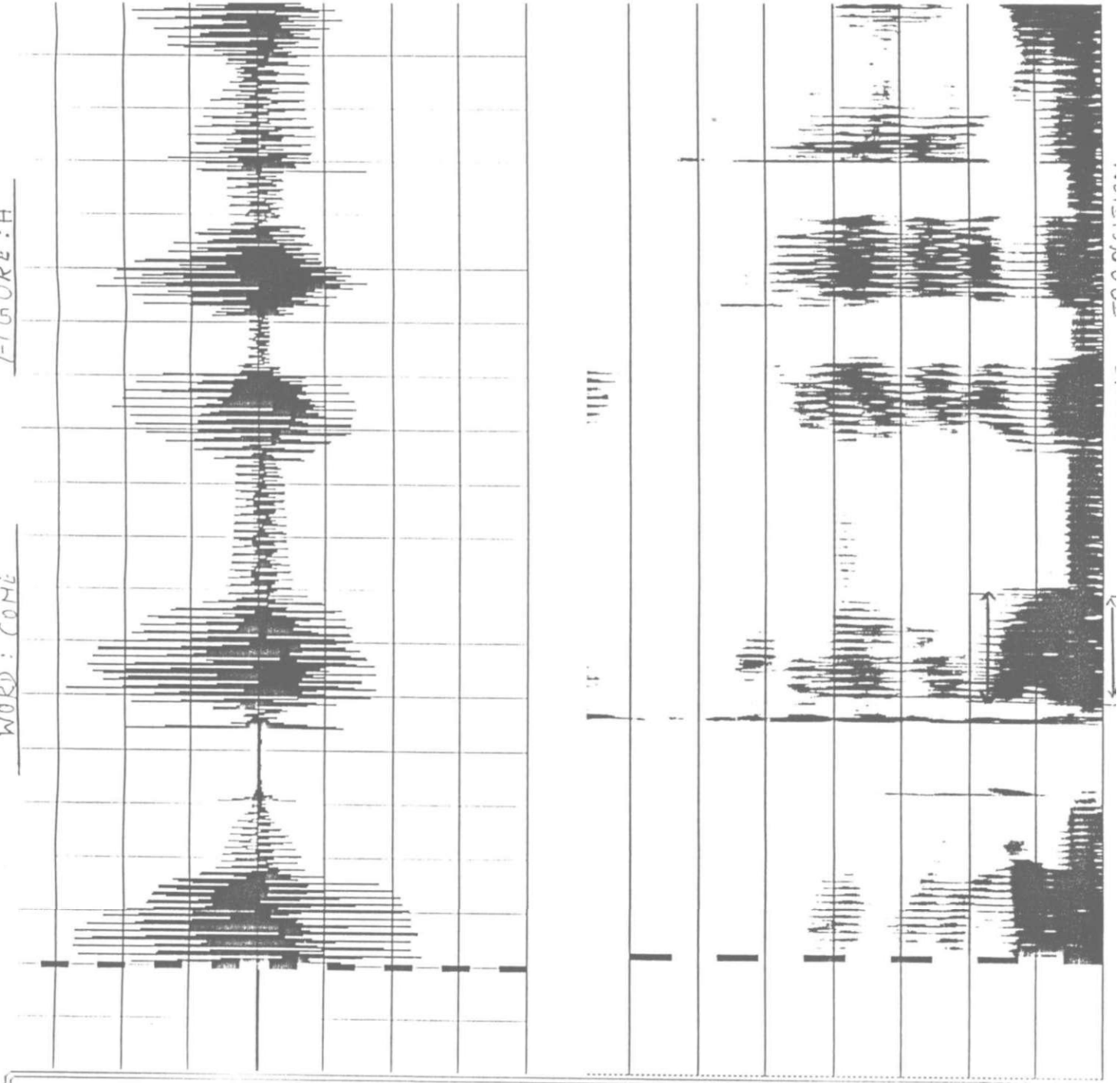


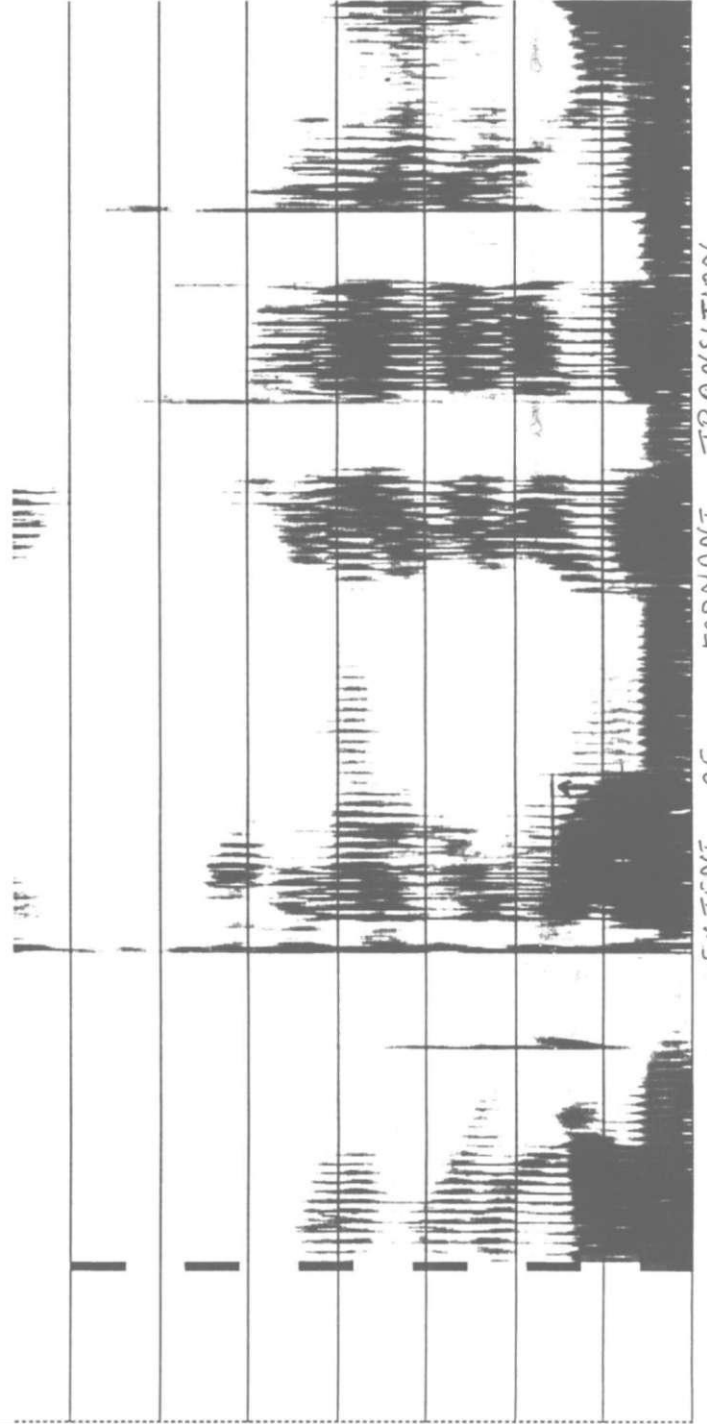
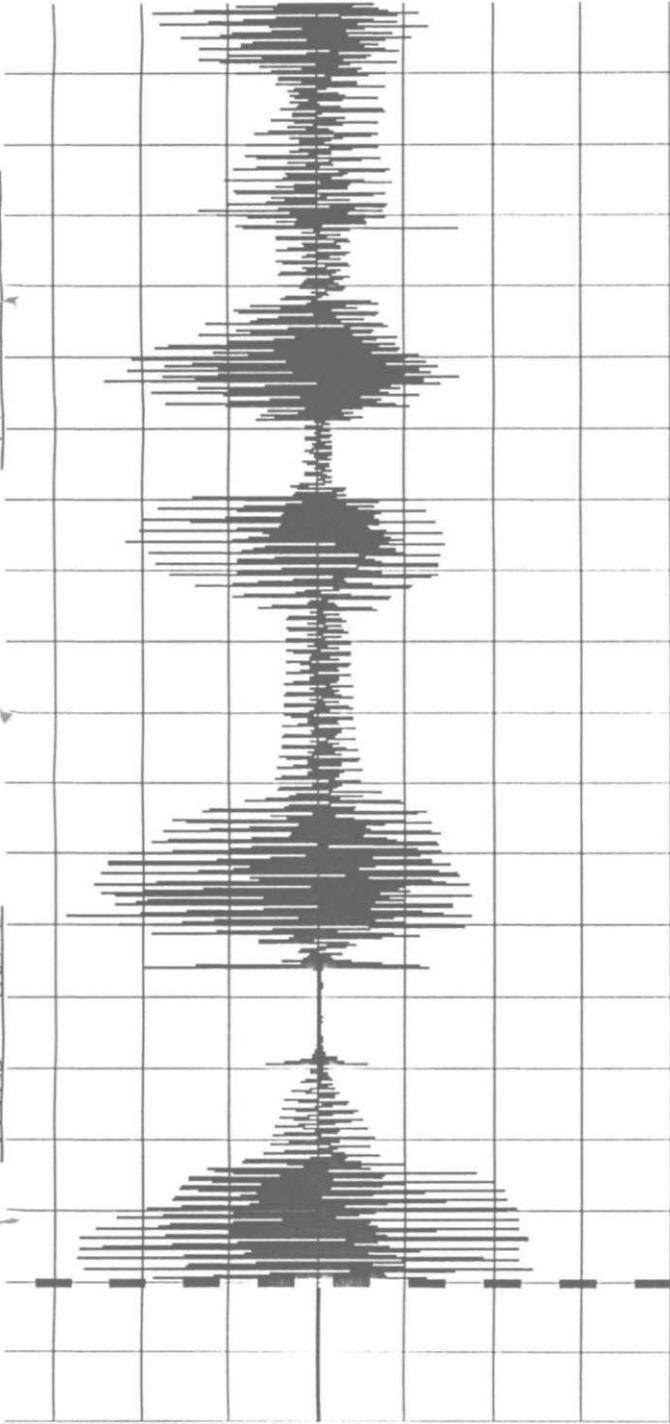
FIGURE: H

WORD: COME



WORD: COME

FIGURE: H



## CHAPTER - IV

### RESULTS AND DISCUSSION

The purpose of this study was to find out the intra- subject and inter-subject variability in terms of the following parameters.

1. Word duration
2. Vowel duration
3. Burst duration
4. Voice onset time
5. Closure duration
6. Lead VOT
7. Frication duration
8. Fundamental frequency
9. Intensity
10. Formants  $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$
11. Transition of formants: Formant transition duration, extent of formant transition, speed of formant transition.

This was done by spectrograph analysis of repeated utterances of six words by five subjects.

The following words were considered to derive the above parameters: "come", "and", "the", "with eight" and "suitcase".

Thus analysis of 270 test words in terms of 11 parameters yielded data which was further statistically analysed.



The results are presented in terms of INTRA SUBJECT VARIABILITY and INTER SUBJECT VARIABILITY.

**INTRA SUBJECT VARIABILITY :**

The intra subject variability i.e., the variation in terms a particular parameter was determined by making a comparison across the utterances of the subject under three conditions. For example for subject 1 comparisons of Condition 1 versus Condition 2, Condition 1 versus Condition 3 and Condition 3 versus Condition 2, were made. The results two or more of these comparisons were taken as the result for that particular word i.e., whenever the significance of difference was present or absent for two or more comparisons then it was considered that there was presence or absence of significance of difference in the utterances of the subject for that particular word. The results of such comparisons are presented in Tables 1 to , for each word for each of the parameters for all subjects studied.

Further the presence or absence of significance of difference among the utterances of test words ( Six words X 3 times X 3 conditions) = 54 utterances by each subject) was determined by considering the total number of presences or absences of significance of differences for each subject,(presented within the brackets with each word for each subject) which is shown for each subject as total P/A.

WORD	SUB1	SUB2	SUB3	SUB4	SUB5	RANGE
1	P (3)	P (3)	P (2)	a (2)	a (3)	31.3-703.7
2	a (3)	a (3)	P (2)	P (2)	P (3)	8.0-155
3	P (3)	P (3)	P (2)	P (2)	P (3)	6.4-80.4
4	a (3)	P (3)	P (2)	P (2)	P (3)	35.0-149
5	P (3)	P (3)	P (2)	P (2)	P (3)	7.0-85
6	P (3)	a (3)	P (2)	P (2)	P (3)	31.3-275.7
Total	P	P	P	P	P	P
P/A	(12)	(12)	(12)	(10)	(15)	(5)

**Table 1:** Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the WORD DURATION

The study of Table - 1 indicates the presence of significance of difference among the utterances in each condition and across the conditions. It can be seen that all the subjects have presence of significance of difference except on two occasions for subject 1 and two and once for subject 4. Thus it can be concluded that there is significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers. The minimum range of variation that was found for these subjects interms of word duration was 6.4 msec and the maximum was 703 msec. for all that words.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms word duration is rejected.

Vowel duration has been considered as one of important variables used in speaker identification there it was considered in the present study to note the intra subject variability. The results of comparisons across the utterances of each subject for different vowels occurring in different words are presented along with the range of variation. The study of Table -2 shows that there was significant difference across the vowels occurring in different words interms of duration, in all subjects in most of the conditions. Therefore it can be stated that the vowel duration varies within the subject's utterances.

The minimum range of variation that was found for these subjects interms of vowel duration was 1.0 msec and the maximum was 76 msec. for all the words..

<b>WORD</b>	<b>SUB1</b>	<b>SUB2</b>	<b>SUB3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P (3)	P (3)	a (3)	P (3)	a (3)	3.3-70.2
2	P (3)	a (2)	a (2)	a (3)	P (3)	9.0-56
3	P (3)	P (3)	P (3)	P (3)	P (3)	22.4-76.4
4	a (2)	P (3)	P (2)	P (2)	P (3)	4.4-12.3
5	P (2)	P (2)	P (2)	P (2)	P (3)	1.0-9
6	P (3)	P (2)	a (2)	P (2)	P (3)	1.5-10.5
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(15)</b>	<b>(14)</b>	<b>(12)</b>	<b>(12)</b>	<b>(15)</b>	<b>(5)</b>

Table 2: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the VOWEL DURATION

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms vowel duration is rejected

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB3</b>	<b>SUB 4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P (3)	a (3)	P (3)	P (3)	a (3)	2.5-11.5
2	P (3)	a (2)	a (2)	a (3)	P (3)	9.0-56
3	P (3)	P (2)	P (3)	P (2)	P (3)	2.0-8.5
4	a (3)	P (3)	a (3)	P (2)	P (3)	2.6-9.5
5	P (2)	P (3)	P (2)	P (2)	a (2)	1.0-9
6	P (2)	P (3)	P (3)	P (3)	a (2)	2-69
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(13)</b>	<b>(11)</b>	<b>(11)</b>	<b>(12)</b>	<b>(9)</b>	<b>(5)</b>

Table 3: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the BURST DURATION

The study of Table - 3 indicates the presence of significance of difference among the utterances in each condition and across the conditions for burst duration. It can be seen that all the subjects have presence of

significance of difference except on nine occasions . Thus it can be concluded that there is significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers in terms of burst duration. The minimum range of variation that was found for these subjects in terms of burst duration was 1 msec and the maximum was 69 msec for all that words.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of burst duration is rejected.

Closure duration has been considered as one of important variables used in speaker identification there it was considered in the present study to note the intra subject variability. The results of comparisons across the utterances of each subject for different consonants occurring in different words are presented along with the range of variation. The study of Table -4 shows that there was significant difference across the closure duration of consonants occurring in different words , in all subjects in most of the conditions. Therefore it can be stated that the closure duration varies within the subject's utterances.

The minimum range of variation that was found for these subjects in terms of closure duration was 1.5 msec and the maximum was 86.8 msec. for all that words for closure duration.

<b>WORD</b>	<b>SUB1</b>	<b>SUB2</b>	<b>SUB3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P (2)	P (2)	P (2)	A (2)	P (2)	14.5-86.8
2	- (0)	- (0)	- (0)	- (0)	- (0)	
3*	P (2)	a (3)	a (3)	P (3)	P (2)	22.4-76.4
4	- (0)	- (0)	- (0)	- (0)	- (0)	
5	P (2)	a (3)	P (2)	P (2)	P (3)	6.7-77.3
6	a (2)	a (2)	P (2)	a (2)	P (2)	1.5-10.5
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(15)</b>	<b>(14)</b>	<b>(12)</b>	<b>(12)</b>	<b>(15)</b>	<b>(5)</b>

Table 4: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the CLOSURE DURATION

\* the word CASE was considered for this purpose.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms closure duration is rejected

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB 3</b>	<b>SUB 4</b>	<b>SUB 5</b>	<b>RANGE</b>
1	P (2)	P (2)	a (3)	P (3)	P (3)	2.5-11.5
2	- (0)	- (0)	- (0)	- (0)	- (0)	
3	P (2)	P (2)	P (3)	P (3)	P (2)	3.5-44.3
4	a	a	P	P	P	63.5-132

	(2)	(3)	(2)	(3)	(2)	
5	-	-	-	-	-	
	(0)	(0)	(0)	(0)	(0)	
6	a	a	p	a	p	2-29.5
	(2)	(2)	(3)	(2)	(3)	
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(12)</b>	<b>(12)</b>	<b>(12)</b>	<b>(10)</b>	<b>(15)</b>	<b>(5)</b>

Table 5: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the VOICE ONSET TIME

\* Word 3 was with and hence lead VOT has been considered and word 6 was case.

The study of Table - 5 indicates the presence of significance of difference among the utterances in terms of voice onset time in each condition and across the conditions. It can be seen that all the subjects have presence of significance of difference except on six occasions out of 20 occasions for VOT.

Thus it can be concluded that there is significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers. The minimum range of variation that was found for these subjects in terms of word duration was 2msecs and the maximum was 135 msecs. for all the words considered for the measurement of VOT.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of Voice Onset Time is rejected.

The fundamental frequency has been considered as one of important variables used in speaker identification therefore it was considered in the present study to note the intra subject variability. The results of comparisons across the utterances of each subject occurring in different words are presented along with the range of variation. The study of Table -6 shows that there was significant difference across the words interms of mean fundamental frequency, in all subjects in most of the conditions. Therefore it can be stated that the fundamental frequency varies within the subject's utterances.

The minimum range of variation that was found for these subjects interms of fundamental frequency was 2 Hz and the maximum was 89 Hz for all the words.

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB 3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P (3)	a (2)	p (3)	a (2)	p (3)	14.-89
2	P (2)	P (2)	a (2)	a (2)	p (2)	11-56
3	P (2)	P (3)	P (3)	a (2)	a (2)	3.5-44.3
4	P (2)	P (2)	P (2)	P (2)	P (2)	33-82
5	a (2)	a (3)	a (2)	p (2)	p (2)	11-52
6	P (3)	P (2)	P (2)	P (2)	a (3)	2-43
7	P (2)	P (2)	P (2)	P (2)	P (2)	15-61
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(15)</b>	<b>(12)</b>	<b>(14)</b>	<b>(10)</b>	<b>(12)</b>	<b>(6)</b>



Table 6: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the FUNDAMENTAL FREQUENCY

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of Fundamental frequency is rejected.

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB 3</b>	<b>SUB 4</b>	<b>SUB 5</b>	<b>RANGE</b>
1	a (2)	P (2)	a (3)	a (3)	P (2)	9-13.2
2	P (3)	a (2)	a (2)	a (3)	P (3)	.5-16
3	P (3)	P (3)	P (2)	P (2)	P (3)	0.4-19
4	P (2)	P (2)	P (2)	P (2)	P (3)	1-2.8
5	P (3)	a (3)	P (2)	P (2)	P (2)	3-13.5
6	P (2)	P (2)	P (3)	P (2)	P (3)	5-13
7	P (2)	P (2)	P (3)	P (2)	P (2)	2-17
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(16)</b>	<b>(12)</b>	<b>(13)</b>	<b>(10)</b>	<b>(18)</b>	<b>(5)</b>

Table 7: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the INTENSITY

The study of Table - 7 indicates the presence of significance of difference among the utterances in each condition and across the conditions IN TERMS INTENSITY. It can be seen that all the subjects have presence of

significance of difference except on seven out of 35 occasions . Thus it can be concluded that there is significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers. The minimum range of variation that was found for these subjects interms of intensity was 0.4 db and the maximum was 19db for all the words.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of intensity is rejected.

The formant frequencies have been considered as one of the important variables used in speaker identification therefore it was considered in the present study to note the intra subject variability. The results of comparisons across the utterances of each subject occuring in different words are presented along with the range of variation. The study of Table -8 shows that there was significant difference across the words interms of formant frequency F1 , in all subjects in most of the conditions. Therefore it can be stated that the formant frequency varies within the subject's utterances.

The minimum range of variation that was found for these subjects interms of formant frequency F1 was 9.42Hz and the maximum was 975Hz for all the words.

<b>WORD</b>	<b>SUB 1</b>	<b>SUB 2</b>	<b>SUB 3</b>	<b>SUB 4</b>	<b>SUB 5</b>	<b>RANGE</b>
1	p (3)	a (3)	p (2)	p (3)	p (2)	9.4-975
2	p (2)	p (3)	p (2)	a (2)	a (2)	55-401

3	P	a	a	P	P	70-360.5
	(3)	(2)	(2)	(2)	(2)	
4	a	a	a	P	P	31-298
	(3)	(3)	(3)	(3)	(2)	
5	P	a	a	a	P	31-188
	(2)	(3)	(2)	(2)	(2)	
6	P	a	P	a	P	42.5-243
	(2)	(2)	(3)	(2)	(2)	
7	a	P	a	P	a	23-165
	(2)	(3)	(2)	(3)	(2)	
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(13)</b>	<b>(18)</b>	<b>(10)</b>	<b>(14)</b>	<b>(12)</b>	<b>(4)</b>

Table 8: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the FORMANT FREQUENCY F1

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of Formant frequency F1 is rejected.

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P	P	a	P	P	86-1130
	(2)	(2)	(2)	(2)	(2)	
2	P	a	P	P	P	8-660
	(3)	(3)	(3)	(2)	(3)	
3	P	P	a	a	P	251-666
	(3)	(2)	(2)	(2)	(3)	
4	P	P	P	P	P	33-612
	(3)	(2)	(2)	(2)	(2)	
5	P	P	P	P	P	118-1036

	(3)	(2)	(2)	(2)	(2)	
6	P	P	P	P	P	32-573
	(3)	(2)	(3)	(3)	(2)	
7	P	P	P	P	P	86-840
	(3)	(3)	(2)	(2)	(2)	
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	
<b>P/A</b>	<b>(20)</b>	<b>(13)</b>	<b>(14)</b>	<b>(14)</b>	<b>(14)</b>	<b>(7)</b>

Table 9: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the FORMANT FREQUENCY F2

The study of Table - 9 indicates the presence of significance of difference among the utterances in each condition and across the conditions. It can be seen that all the subjects have presence of significance of difference except on 31 out of 35 occasions . Thus it can be concluded that there is significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers. The minimum range of variation that was found for these subjects interms of intensity was 8 HZ and the maximum was 1130 Hz for all the words.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of formant frequency F2 is rejected.

The results of comparisons of formant frequency F4 across the utterances of each subject occuring in different words are presented along with the range of variation. The study of Table -10 shows that there was

significant difference across the words in terms of formant frequency F3, in all subjects in most of the conditions. Therefore it can be stated that the formant frequency varies within the subject's utterances.

The minimum range of variation that was found for these subjects in terms of formant frequency F3 was 15 Hz and the maximum was 1325 Hz for all the words.

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P (2)	P (3)	P (3)	P (2)	P (2)	469-622
2	P (3)	a (2)	P (2)	a (3)	P (3)	55-1098
3	P (3)	P (2)	P (3)	P (3)	P (2)	15-447
4	a (3)	P (2)	P (2)	P (2)	P (2)	874-1324
5	P (2)	P (3)	a (2)	P (2)	P (2)	722-1325
6	P (2)	P (3)	P (2)	P (3)	P (3)	86-840
7	P (2)	P (2)	P (2)	P (3)	P (3)	31-631
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(14)</b>	<b>(16)</b>	<b>(14)</b>	<b>(16)</b>	<b>(17)</b>	<b>(7)</b>

Table 10: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the FORMANT FREQUENCY F3

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of Formant frequency F3 is rejected.

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB3</b>	<b>SUB4</b>	<b>SUB 5</b>	<b>RANGE</b>
1	P (2)	P (3)	P (2)	a (2)	P (3)	94-1726
2	P (2)	P (3)	a (2)	P (2)	P (3)	73-76
3	a (2)	P (3)	P (2)	P (2)	P (2)	23-745
4	a (3)	P (2)	P (2)	P (3)	P (2)	94-972
5	P (2)	a (2)	P (2)	P (2)	P (3)	1435-2847
6	P (2)	a (2)	P (2)	P (2)	P (3)	232-1629
7	P (2)	a (2)	P (2)	P (2)	P (2)	31-912
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(11)</b>	<b>(14)</b>	<b>(13)</b>	<b>(14)</b>	<b>(18)</b>	<b>(7)</b>

Table 11:Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the FORMANT FREQUENCY F4

The study of Table 11 indicates the presence of significance of difference among the utterances in each condition and across the conditions. It can be seen that all the subjects have presence of significance of difference except on 28 out of 35 occasions . Thus it can be concluded that there is

significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers. The minimum range of variation that was found for these subjects in terms of intensity was 23Hz and the maximum was 2847 Hz for all the words.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of formant frequency F4 is rejected.

The DURATION OF FORMANT TRANSITION has been considered as one of important variables used in speaker identification therefore it was considered in the present study to note the intra subject variability. The results of comparisons across the utterances of each subject occurring in different words are presented along with the range of variation.

The study of Table: 12 shows that there was significant difference across the words in terms of duration of transition, in all subjects in most of the conditions. Therefore it can be stated that the duration of formant transition varies within the subject's utterances.

The minimum range of variation that was found for these subjects in terms of duration of transition was 2msec and the maximum was 118msec for all the words.

<b>WORD</b>	<b>SUB1</b>	<b>SUB2</b>	<b>SUB 3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P	P	P	P	a	12-51
	(2)	(2)	(2)	(3)	(3)	
2	a	P	P	P	P	9-50
	(2)	(2)	(3)	(3)	(3)	

3	a	P	P	P	P	20-66.3
	(3)	(2)	(2)	(3)	(2)	
4	a	a	P	P	P	2-48
	(3)	(2)	(3)	(2)	(3)	
5	a	P	a	P	P	22-102
	(2)	(2)	(2)	(3)	(3)	
6	P	P	a	P	P	73-118
	(9)	(14)	(14)	(18)	(16)	
7	P	P	P	a	P	10-70
	(2)	(2)	(3)	(2)	(3)	
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(15)</b>	<b>(12)</b>	<b>(14)</b>	<b>(10)</b>	<b>(12)</b>	<b>(7)</b>

Table 12: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the DURATION OF FORMANT FREQUENCY

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of duration of formant transition is rejected.

<b>WORD</b>	<b>SUB1</b>	<b>SUB 2</b>	<b>SUB 3</b>	<b>SUB 4</b>	<b>SUB5</b>	<b>RANGE</b>
1	P	P	P	P	P	62-486
	(2)	(2)	(3)	(3)	(2)	
2	P	a	P	a	P	126-638
	(3)	(3)	(3)	(2)	(3)	
3	P	a	P	P	P	313-702
	(2)	(3)	(2)	(3)	(3)	
4	P	P	P	P	P	47-588
	(2)	(2)	(2)	(3)	(3)	
5	a	P	a	a	P	32-498
	(2)	(3)	(2)	(2)	(3)	
6	P	P	P	a	P	169-440



	(3)	(3)	(2)	(3)	(2)	
7	P	P	P	P	P	31-431
	(2)	(2)	(3)	(2)	(2)	
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(16)</b>	<b>(12)</b>	<b>(13)</b>	<b>(10)</b>	<b>(18)</b>	<b>(5)</b>

Table 13: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the extent of formant transition

The study of Table -13 indicates the presence of significance of difference among the utterances in each condition and across the conditions in terms of extent of formant transition. It can be seen that all the subjects have presence of significance of difference except on seven out of 35 occasions. Thus it can be concluded that there is significant differences in the utterances that the subjects made. Therefore one has to be careful while making comparisons of the same utterances while identifying the speakers. The minimum range of variation that was found for these subjects in terms of intensity was 31msec and the maximum was 637.6msec for all the words.

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of intensity is rejected.

The SPEED OF FORMANT TRANSITION has been considered as one of important variables used in speaker identification therefore it was considered in the present study to note the intra subject variability. The results of comparisons across the utterances of each subject occurring in different words are presented along with the range of variation. The study of Table: 14 shows

that there was significant difference across the words in terms of speed duration of transition, in all subjects in most of the conditions. Therefore it can be stated that the speed of formant transition varies within the subject's utterances.

The minimum range of variation that was found for these subjects in terms of speed of transition was .14 msec and the maximum was 19.9 msec for all the words.

<b>WORD</b>	<b>SUB1</b>	<b>SUB2</b>	<b>SUB3</b>	<b>SUB4</b>	<b>SUB5</b>	<b>RANGE</b>
1	a	P	P	P	a	.14-9.6
	(3)	(2)	(2)	(2)	(3)	
2	P	P	a	P	P	1-11.64
	(2)	(2)	(3)	(2)	(2)	
3	P	P	a	P	a	5.6-19.9
	(3)	(2)	(3)	(2)	(2)	
4	P	P	a	P	P	.2-14.1
	(2)	(3)	(3)	(2)	(2)	
5	a	P	a	P	a	.4-8
	(3)	(3)	(3)	(2)	(2)	
6	P	P	P	P	a	5.4-12
	(2)	(3)	(2)	(2)	(2)	
7	P	P	a	P	P	.6-5.4
	(2)	(3)	(2)	(2)	(2)	
<b>Total</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>	<b>P</b>
<b>P/A</b>	<b>(11)</b>	<b>(13)</b>	<b>(5)</b>	<b>(14)</b>	<b>(9)</b>	<b>(7)</b>

Table 14: Table showing the presence or absence of significance of difference between the utterances of each word by each subject for the SPEED OF FORMANT FREQUENCY

Therefore the hypothesis stating that there is no significance of difference between the utterances of the subject (intra subject variability) in terms of SPEED of formant transition is rejected.

### **INTER-SUBJECT VARIABILITY**

The variability arising in different parameters among various subjects was calculated.

Once, the values for all the parameters were acquired the data was subjected to statistical analysis. Descriptive statistical procedure was implemented to find out the variability among subjects.

For example: For the word "come" all the parameters were first obtained. This was done for all subjects. Then, the values between the subjects were compared for each parameter in the following manner.

1. Word duration for the word "come" in subject 1 vs word duration for the word "come" in subject 2.

2. Word duration for the word "come" in subject 1 vs word duration for the word "come" in subject 3.

Similarly for the following:

Subject 1 vs subject 4,

Subject 1 vs subject 5

Subject 2 vs subject 3

Subject 2 vs subject 4

Subject 2 vs subject 5

Subject 3 vs subject 4

Subject 3 vs subject 5

Subject 4 vs subject 5

The same procedure is carried out for all the other words which were considered, i.e. 'and', 'the', 'with', 'eight', 'suit' and 'case'. Other parameters were obtained in a similar fashion.

Also, the minimum and maximum variability possible between the word for the same parameter was calculated.

As the objective of the study is to note the variability, the results have been considered in terms of range and inter-subject only. However, data regarding mean and standard deviation are also presented.

**Table I: Parameter considered - Word duration**

Word: Come	Mean	Standard deviation	Range
Subject 1	151.64	19.89	73.25
Subject 2	204.88	47.99	128.027
Subject 3	134.33	10.06	32.30
Subject 4	164.43	14.56	48.00
Subject 5	151.61	17.90	52.13
Word: And			
Subject 1	61.72	14.35	47.50
Subject 2	96.62	36.11	116.40
Subject 3	97.23	18.61	54.00
Subject 4	100.70	35.48	119.50
Subject 5	67.11	12.62	44.00
Word: The			
Subject 1	79.34	17.24	54.00
Subject 2	84.16	15.73	53.15
Subject 3	90.60	20.09	67.00
Subject 4	67.97	8.69	26.50
Subject 5	78.72	22.63	72.00
Word: With			
Subject 1	138.44	19.32	53.00
Subject 2	187.39	18.81	53.20
Subject 3	148.63	14.21	47.00
Subject 4	207.33	71.49	221.00
Subject 5	142.94	31.13	106.00
Word : eight			

Subject 1	196.63	9.90	32.00
Subject 2	169.36	20.43	65.70
Subject 3	151.21	15.19	44.00
Subject 4	170.44	13.05	35.00
Subject 5	167.21	24.60	77.60
Word: Suitcase			
Subject 1	593.02	35.03	122.00
Subject 2	607.42	74.40	200.00
Subject 3	574.64	67.99	211.65
Subject 4	586.67	65.03	204.00
Subject 5	540.71	70.77	202.40

The study of table I, shows that the lowest range (lowest value or the next to lowest value have been considered) of word duration for all the words studied was shown by subject 1 in 4 out of 5 words. Similarly, the highest variability, i.e. range was shown by subject 5.

Thus, it can be concluded that the parameter word duration seems to be a factor which varies from individual to individual. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of word duration is rejected.

**Table II: Parameter considered - Vowel duration**

Word: Come	Mean	SD	Range
Subject 1	63.93	11.16	38.00
Subject 2	79.74	17.50	47.20
Subject 3	42.57	6.04	20.00
Subject 4	60.91	13.35	35.00
Subject 5	55.95	11.93	33.00

## Word: And

Subject 1	45.80	12.73	40.00
Subject 2	59.01	10.33	30.50
Subject 3	54.60	11.27	36.00
Subject 4	67.78	16.45	53.00
Subject 5	48.91	9.93	34.00

## Word: The

Subject 1	62.22	10.49	29.00
Subject 2	61.44	14.40	47.50
Subject 3	66.82	21.12	76.40
Subject 4	54.98	8.57	23.20
Subject 5	58.22	15.67	42.00

## Word: With

Subject 1	55.87	6.62	20.00
Subject 2	62.53	32.43	107.00
Subject 3	53.54	11.55	32.80
Subject 4	61.29	14.74	48.00
Subject 5	63.22	31.59	104.00

## Word: Eight

Subject 1	133.78	8.74	21.00
Subject 2	115.71	15.70	43.00
Subject 3	98.00	16.84	58.00
Subject 4	117.11	21.63	46.00
Subject 5	99.04	26.67	94.40

## word suit

Subject 1	91.06	21.62	56.00
Subject 2	87.64	16.32	51.50
Subject 3	87.76	10.22	24.20
Subject 4	87.76	10.22	24.20
Subject 5	73.89	17.49	61.50

Word: Case

Subject 1	128.56	21.05	58.00
Subject 2	137.28	15.58	38.00
Subject 3	122.74	22.57	68.00
Subject 4	118.89	13.20	34.00
Subject 5	130.24	18.24	50.00

The study of table II, shows that the lowest range and the highest range of variability for the parameter vowel duration.

Lowest range: 20 msec

Highest range: 107 msec

Thus, it can be concluded that the parameter vowel duration seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of vowel duration is rejected.

**Table III: Parameter considered - Burst duration**

Word: Come	Mean	SD	Range
Subject 1	9.03	1.86	6.00
Subject 2	11.02	2.50	7.50
Subject 3	6.84	2.50	11.50
Subject 4	11.60	2.50	7.50
Subject 5	9.64	2.50	7.50
Word: The			
Subject 1	4.28	1.96	6.00
Subject 2	4.09	1.38	2.01



Subject 3	2.56	1.16	3.00
Subject 4	2.56	1.16	3.00
Subject 5	3.79	0.95	2.50
Word: With			
Subject 1	4.37	1.98	6.00
Subject 2	7.13	2.51	8.50
Subject 3	5.39	1.85	5.00
Subject 4	5.06	1.95	5.50
Subject 5	5.16	2.27	6.00
Word: Eight			
Subject 1	4.76	1.91	5.50
Subject 2	5.11	2.14	6.00
Subject 3	6.73	2.82	8.20
Subject 4	4.57	1.37	4.50
Subject 5	4.53	1.33	3.50
Word: Suit			
Subject 1	5.24	1.06	3.20
Subject 2	5.86	3.42	10.70
Subject 3	5.14	2.59	7.00
Subject 4	3.83	1.02	3.00
Subject 5	5.26	1.80	5.00
Word: Case			
Subject 1	11.12	6.95	19.50
Subject 2	11.12	6.65	22.50
Subject 3	6.94	4.30	12.50
Subject 4	10.67	3.78	10.00
Subject 5	8.26	3.65	10.00

The study of table III, shows that the lowest range and the highest range of variability for the parameter burst duration.

Lowest range: 2.01 msec

Highest range: 22.50 msec

Thus, it can be concluded that the parameter burst duration seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of burst duration is rejected.

**Table IV: Parameter considered - Voice onset time**

Word: Come	Mean	SD	Range
Subject 1	23.30	3.17	11.70
Subject 2	32.38	4.33	12.30
Subject 3	28.40	4.47	14.00
Subject 4	21.48	3.67	11.00
Subject 5	28.18	4.85	13.50
Word: The			
Subject 1	10.01	3.16	9.40
Subject 2	9.91	2.92	9.00
Subject 3	20.33	6.86	22.00
Subject 4	9.36	1.69	5.10
Subject 5	9.26	2.92	10.00
Word: Case			
Subject 1	29.38	4.27	11.60
Subject 2	38.17	7.95	26.70

Subject 3	32.50	2.83	10.00
Subject 4	22.13	2.70	7.50
Subject 5	29.94	5.81	19.50
<b>Parameter considered: Lead voice onset time</b>			
Word: With	Mean	SD	Range
Subject 1	56.68	7.25	24.10
Subject 2	92.66	12.92	36.70
Subject 3	67.49	16.69	56.00
Subject 4	102.51	27.64	84.60
Subject 5	61.36	8.27	26.00

The study of table IV, shows that the lowest range and the highest range of variability for the parameter voice onset time.

Lowest range: 5.10 msec

Highest range: 26.70 msec

The lead voice onset time was also included in this table. The lowest range observed for the word 'with' was 36.70 and the highest was 84.60.

Thus, it can be concluded that the parameter voice onset time seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of voice onset time is rejected.

**Table V: Parameter considered - Closure duration**

Word: Come	Mean	SD	Range
Subject 1	65.11	9.86	31.50
Subject 2	85.56	26.80	73.20
Subject 3	61.68	9.95	35.80
Subject 4	79.41	13.89	46.00
Subject 5	65.56	9.85	28.50
Word: Eight			
Subject 1	57.48	5.59	20.00
Subject 2	48.03	6.64	20.90
Subject 3	45.22	15.18	39.00
Subject 4	49.56	12.72	42.00
Subject 5	58.36	8.46	26.80
Word: Suit			
Subject 1	61.33	7.18	23.00
Subject 2	68.39	17.60	54.00
Subject 3	42.83	3.24	10.00
Subject 4	48.44	6.95	19.00
Subject 5	62.31	23.09	77.30
Word: Case			
Subject 1	67.24	7.26	24.00
Subject 2	54.39	7.78	25.00
Subject 3	62.84	11.05	31.50
Subject 4	54.83	7.46	22.70
Subject 5	62.18	20.79	52.50

The study of table V, shows that the lowest range and the highest range of variability for the parameter closure duration.

Lowest range: 10.00 msec

Highest range: 73.20 msec

Thus, it can be concluded that the parameter closure duration seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of closure duration is rejected.

**Table VI: Parameter considered - Frication duration**

Word: Suit	Mean	SD	Range
Subject 1	93.69	13.10	42.00
Subject 2	84.58	15.13	52.00
Subject 3	87.28	14.02	45.00
Subject 4	88.39	12.44	38.00
Subject 5	89.33	11.91	35.00
Word: Case			
Subject 1	103.42	10.22	31.20
Subject 2	128.61	45.87	131.50
Subject 3	125.93	25.55	70.50
Subject 4	152.51	38.72	125.00
Subject 5	87.87	31.29	108.00

The study of table VI, shows that the lowest range and the highest range of variability for the parameter frication duration.

Lowest range: 31.20 msec

Highest range: 131.50 msec

Thus, it can be concluded that the parameter frication duration seems to be a factor which varies

considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of frication duration is rejected.

**Table VII: Parameter considered - Fundamental frequency**

Word: Come	Mean	SD	Range
Subject 1	119.00	2.81	9.00
Subject 2	106.21	10.99	33.80
Subject 3	120.88	7.51	19.50
Subject 4	123.97	14.43	45.00
Subject 5	125.17	12.28	36.90
Word: And			
Subject 1	111.22	7.36	24.00
Subject 2	104.17	12.75	33.10
Subject 3	115.97	7.45	26.00
Subject 4	118.22	4.76	12.00
Subject 5	108.46	6.41	17.00
Word: The			
Subject 1	112.89	10.60	31.00
Subject 2	95.84	5.61	18.50
Subject 3	99.14	10.86	31.60
Subject 4	115.86	6.55	22.00
Subject 5	112.93	3.46	6.40
Word: With			
Subject 1	119.22	9.50	26.00
Subject 2	97.14	8.44	21.80
Subject 3	95.64	4.32	14.20
Subject 4	114.22	11.32	36.00
Subject 5	114.89	9.16	25.00
Word: Eight			
Subject 1	118.03	12.72	36.00
Subject 2	98.40	6.66	19.50

Subject 3	108.56	8.59	24.00
Subject 4	114.33	7.18	23.00
Subject 5	115.88	7.46	26.00
<hr/>			
Word: Suit			
Subject 1	112.00	8.47	22.00
Subject 2	103.67	11.89	40.50
Subject 3	111.94	12.62	44.00
Subject 4	120.11	10.46	30.00
Subject 5	113.78	11.26	34.00
<hr/>			
Word: Case			
Subject 1	106.67	14.71	36.00
Subject 2	98.17	8.84	24.60
Subject 3	107.86	16.40	53.70
Subject 4	116.44	13.31	45.00
Subject 5	120.01	7.09	20.10

The study of table VII, shows that the lowest range and the highest range of variability for the parameter fundamental frequency.

Lowest range: 6.40 Hz

Highest range: 53.70 Hz

Thus, it can be concluded that the parameter fundamental frequency seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of fundamental frequency is rejected.

**Table VIII: Parameter considered - Intensity**

Word: Come	Mean	SD	Range
Subject 1	48.96	5.48	15.94
Subject 2	48.81	2.56	7.00

Subject 3	51.03	2.46	7.93
Subject 4	49.84	4.16	11.90
Subject 5	45.53	3.53	9.50
Word: And			
Subject 1	45.60	6.14	19.00
Subject 2	45.74	3.30	10.18
Subject 3	46.43	3.82	11.60
Subject 4	49.11	2.62	7.00
Subject 5	43.63	7.57	24.40
Word: The			
Subject 1	46.79	4.90	13.60
Subject 2	46.98	3.11	8.91
Subject 3	52.88	2.86	7.50
Subject 4	50.89	3.10	9.00
Subject 5	45.98	2.67	8.20
Word: With			
Subject 1	43.82	5.44	15.30
Subject 2	47.27	3.43	10.00
Subject 3	46.48	3.24	9.70
Subject 4	47.89	2.80	8.00
Subject 5	43.60	4.98	13.00
Word: Eight			
Subject 1	49.43	3.43	12.30
Subject 2	49.30	4.48	13.20
Subject 3	48.74	4.00	12.00
Subject 4	47.96	3.35	10.0
Subject 5	46.12	2.50	8.50
Word: Suit			
Subject 1	49.10	3.86	11.00
Subject 2	48.99	4.20	13.00
Subject 3	50.22	3.74	13.00
Subject 4	49.30	3.43	12.30
Subject 5	49.22	2.24	7.00
Word: Case			



Subject 1	50.28	3.49	17.00
Subject 2	50.72	3.64	11.40
Subject 3	49.54	5.08	16.00
Subject 4	50.33	2.29	6.00
Subject 5	47.83	3.26	9.85

The study of table VIII, shows that the lowest range and the highest range of variability for the parameter intensity.

Lowest range: 6.00 dB

Highest range: 17.00 dB

Thus, it can be concluded that the parameter intensity seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of intensity is rejected.

**Table IX: Parameter considered - Formant frequency 1 (F1)**

Word: Come	Mean	SD	Range
Subject 1	655.56	54.01	157.00
Subject 2	627.58	61.82	172.60
Subject 3	636.04	48.25	172.00
Subject 4	614.67	34.04	111.00
Subject 5	645.28	24.10	70.50
Word: And			
Subject 1	624.56	49.83	157.00
Subject 2	433.87	43.96	125.00
Subject 3	626.89	77.89	252.00
Subject 4	673.00	63.76	182.00
Subject 5	498.67	37.22	94.00

Word: The			
Subject 1	539.59	36.77	123.70
Subject 2	483.18	38.39	117.70
Subject 3	592.89	55.40	196.00
Subject 4	544.67	38.05	108.00
Subject 5	640.89	69.47	258.00
Word: With			
Subject 1	471.22	28.73	94.00
Subject 2	447.12	51.72	172.60
Subject 3	446.11	67.81	251.00
Subject 4	436.78	34.10	86.00
Subject 5	469.72	81.14	251.00
Word: Eight			
Subject 1	455.10	22.69	70.00
Subject 2	436.76	55.33	141.00
Subject 3	473.56	50.49	126.00
Subject 4	433.56	54.67	188.00
Subject 5	448.96	47.02	141.00
Word: Suit			
Subject 2	371.28	56.95	188.50
Subject 3	389.67	46.86	146.00
Subject 4	371.11	40.46	140.00
Subject 5	411.33	40.20	134.00
Word: Case			
Subject 1	450.67	34.90	94.00
Subject 2	410.36	50.60	145.00
Subject 3	450.33	31.22	97.00
Subject 4	430.00	30.19	86.00
Subject 5	406.39	42.64	134.00

The study of table IX, shows that the lowest range and the highest range of variability for the parameter formant frequency 1 (F1).

Lowest range: 70.00 Hz

Highest range: 258.00 Hz

Thus, it can be concluded that the parameter formant frequency 1 (F<sub>1</sub>) seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of formant frequency 1 (F<sub>1</sub>) is rejected.

**Table X: Parameter considered: Formant frequency 2 (F<sub>2</sub>)**

Word: Come	Mean	SD	Range
Subject 1	1237.06	141.44	197.00
Subject 2	1349.26	110.70	376.00
Subject 3	1083.33	67.30	196.00
Subject 4	1055.22	64.71	197.00
Subject 5	1349.26	110.70	220.00
Word: And			
Subject 1	1550.67	183.87	663.00
Subject 2	1372.97	71.42	506.00
Subject 3	1276.89	71.42	220.00
Subject 4	1671.44	188.60	498.00
Subject 5	1671.11	134.78	352.00
Word: The			
Subject 1	1411.78	227.82	666.00
Subject 2	1497.54	79.02	290.00
Subject 3	1524.11	120.89	392.00
Subject 4	1385.56	75.96	204.00
Subject 5	1469.11	87.57	236.00
Word: With			
Subject 1	1706.56	98.96	270.00
Subject 2	1562.43	136.91	423.00
Subject 3	1779.33	189.91	691.00
Subject 4	1753.33	126.30	341.00
Subject 5	1570.44	73.72	225.00
Word: Eight			

Subject 1	1895.89	226.68	784.00
Subject 2	1819.17	207.81	691.00
Subject 3	1947.56	171.13	564.00
Subject 4	2017.78	105.72	368.00
Subject 5	2170.67	129.32	471.00
Word: Suit			
Subject 1	1237.06	141.44	502.00
Subject 2	1349.26	110.70	376.00
Subject 3	1083.33	67.30	196.00
Subject 4	1055.72	64.71	197.00
Subject 5	1139.78	85.11	220.00
Word: Case			
Subject 1	1770.33	119.63	344.00
Subject 2	1720.67	196.08	659.00
Subject 3	1742.00	94.71	267.00
Subject 4	1805.11	88.65	286.00
Subject 5	1939.11	99.01	305.00

The study of table X, shows that the lowest range and the highest range of variability for the parameter formant frequency 2 (F2).

Lowest range: 196.00 Hz

Highest range: 784.00 Hz

Thus, it can be concluded that the parameter formant frequency 2 (F2) seems to be a factor which varies considerably. Herefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of formant frequency 2 (F2) is rejected.

**Table XI: Parameter considered - Formant frequency 3 (F3)**

Word: Come	Mean	SD	Range
Subject 1	2262.11	130.47	369.00
Subject 2	2400.69	151.42	471.00

Subject 3	2127.33	94.87	282.00
Subject 4	2288.72	117.82	370.00
Subject 5	2350.22	164.14	471.00
Word: And			
Subject 1	2317.61	373.01	769.00
Subject 2	2520.29	310.17	431.00
Subject 3	2259.89	110.34	361.00
Subject 4	2435.00	208.76	532.00
Subject 5	2693.56	159.11	431.00
Word: The			
Subject 1	2467.67	102.65	329.00
Subject 2	2492.02	65.26	173.00
Subject 3	2395.44	70.56	212.00
Subject 4	3494.56	111.17	312.00
Subject 5	2568.67	128.54	321.00
Word: With			
Subject 1	2323.22	37.10	125.00
Subject 2	2380.98	85.16	288.00
Subject 3	2554.89	384.24	1255.00
Subject 4	2370.11	110.12	302.00
Subject 5	2603.89	68.57	197.00
Word: Eight			
Subject 1	2515.22	184.92	596.00
Subject 2	2486.39	126.05	408.00
Subject 3	2571.73	110.16	282.00
Subject 4	2530.00	114.91	345.00
Subject 5	2714.67	269.70	886.00
Word: Suit			
Subject 1	2163.67	177.43	659.00
Subject 2	2407.86	80.71	237.00
Subject 3	2146.11	80.96	153.00
Subject 4	2207.44	106.09	353.00
Subject 5	2377.00	166.09	462.00
Word: Case			

Subject 1	2439.56	52.13	186.00
Subject 2	2410.88	93.54	281.90
Subject 3	2320.00	48.87	153.00
Subject 4	2407.44	93.54	266.00
Subject 5	2469.78	85.28	251.00

The study of table XI, shows that the lowest range and the highest range of variability for the parameter formant frequency 3 (F<sub>3</sub>).

Lowest range: 153.00 Hz

Highest range: 886.00 Hz

Thus, it can be concluded that the parameter formant frequency 3 (F<sub>3</sub>) seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of formant frequency 3 (F<sub>3</sub>) is rejected.

**Table XII: Parameter considered - Formant frequency 4 (F<sub>4</sub>)**

Word: Come	Mean	SD	Range
Subject 1	3295.87	214.70	650.00
Subject 2	3867.53	194.34	546.80
Subject 3	3737.78	511.14	726.00
Subject 4	3279.56	298.11	859.00
Subject 5	3618.33	318.78	879.00
Word: And			
Subject 1	3501.33	202.70	627.00
Subject 2	3743.76	527.68	102.00
Subject 3	3416.11	198.60	749.00
Subject 4	3462.89	203.44	600.00
Subject 5	3456.78	297.18	847.00
Word: The			

Subject 1	3433.00	254.53	682.00
Subject 2	3671.60	141.15	345.10
Subject 3	3725.78	185.91	596.00
Subject 4	3468.44	199.64	714.00
Subject 5	3453.45	252.80	675.50
Word: With			
Subject 1	3376.67	134.79	431.00
Subject 2	3789.68	195.16	564.50
Subject 3	3852.11	196.07	680.00
Subject 4	3535.78	166.22	503.00
Subject 5	3965.00	133.91	474.00
Word: Eight			
Subject 1	3321.33	155.93	439.00
Subject 2	3715.14	297.52	412.00
Subject 3	3754.00	193.06	627.00
Subject 4	3514.89	322.55	879.00
Subject 5	3756.11	215.58	753.00
<hr/>			
Word: Suit			
Subject 1	2878.89	224.08	832.00
Subject 2	3465.39	328.10	985.00
Subject 3	3300.00	146.67	502.00
Subject 4	2892.56	100.08	337.00
Subject 5	3329.06	299.48	941.00
Word: Case			
Subject 1	3539.89	115.59	338.00
Subject 2	3839.07	287.36	847.00
Subject 3	3786.22	224.92	659.00
Subject 4	3459.44	194.97	504.00
Subject 5	3852.11	205.93	526.00

The study of table XII, shows that the lowest range and the highest range of variability for the parameter formant frequency 4 (**F4**).

Lowest range: 337.00 Hz

Highest range: 1102.00 Hz

Thus, it can be concluded that the parameter formant frequency 4 (F4) seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of formant frequency 4 (F4) is rejected.

**Table XIII: Parameter considered - Duration of formant transition**

Word: Come	Mean	SD	Range
Subject 1	45.44	10.31	29.00
Subject 2	59.00	12.04	36.00
Subject 3	37.22	5.21	15.00
Subject 4	50.22	11.38	32.00
Subject 5	45.67	7.26	22.00
Word: And			
Subject 1	39.56	7.72	20.00
Subject 2	52.13	10.23	34.00
Subject 3	46.16	9.21	28.60
Subject 4	61.56	9.49	28.00
Subject 5	46.00	6.46	22.00
Word: The			
Subject 1	44.56	11.22	33.00
Subject 2	50.20	16.76	51.20
Subject 3	47.56	8.79	25.00
Subject 4	43.56	13.11	45.00
Subject 5	55.89	8.59	30.00
Word: With			
Subject 1	46.22	7.64	23.00
Subject 2	40.78	14.42	58.00
Subject 3	40.11	10.80	33.00
Subject 4	49.89	18.32	59.00
Subject 5	38.89	6.27	22.00



Word: Eight			
Subject 1	87.22	32.70	85.00
Subject 2	83.00	23.22	72.00
Subject 3	72.11	12.64	42.00
Subject 4	80.19	20.95	62.00
Subject 5	91.89	12.46	40.00
Word: Suit			
Subject 1	49.67	12.14	31.00
Subject 2	71.56	2.01	5.00
Subject 3	64.67	10.11	31.00
Subject 4	56.78	6.18	20.00
Subject 5	52.83	11.92	32.00
Word: Case			
Subject 1	76.78	20.92	70.00
Subject 2	95.44	16.61	60.00
Subject 3	84.89	18.00	60.00
Subject 4	72.67	9.91	30.00
Subject 5	78.67	9.06	26.00

The study of table XIII, shows that the lowest range and the highest range of variability for the parameter duration of formant transition.

Lowest range: 5.00 msec

Highest range: 85.00 msec

Thus, it can be concluded that the parameter duration of formant transition seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of duration of formant transition is rejected.

**Table XTV: Parameter considered - Extent of formant transition**

Word: Come	Mean	SD	Range
Subject 1	508.11	125.57	376.00
Subject 2	503.68	128.15	376.90
Subject 3	407.33	87.55	329.00
Subject 4	400.89	74.06	217.00
Subject 5	477.66	84.16	251.50
Word: And			
Subject 1	552.72	113.00	376.00
Subject 2	510.70	142.76	322.20
Subject 3	464.67	145.02	439.00
Subject 4	730.00	105.10	329.00
Subject 5	602.11	122.99	377.00
Word: The			
Subject 1	356.11	137.09	488.00
Subject 2	499.38	128.20	377.00
Subject 3	479.11	194.85	627.00
Subject 4	382.39	71.78	217.00
Subject 5	510.11	46.93	128.00
Word: With			

Subject 1	434.87	98.34	222.00
Subject 2	569.81	92.80	304.90
Subject 3	424.39	167.62	494.00
Subject 4	526.22	100.54	331.00
Subject 5	567.19	113.58	290.00

Word: Eight

Subject 1	382.42	70.78	217.00
Subject 2	569.85	93.20	304.05
Subject 3	602.12	122.95	377.00
Subject 4	510.12	46.93	128.00
Subject 5	424.37	98.23	221.00
Subject 1	325.89	70.84	251.00

Word: Suit

Subject 2	454.20	79.72	281.00
Subject 3	340.22	67.41	211.00
Subject 4	310.11	71.17	252.00
Subject 5	338.17	51.54	157.00

Word: Case

Subject 1	401.44	121.26	314.00
Subject 2	479.50	93.09	283.00

Subject 3	376.56	116.53	368.00
Subject 4	363.44	95.76	372.00
Subject 5	413.00	44.33	125.50

The study of table XIV, shows that the lowest range and the highest range of variability for the parameter extent of formant transition.

Lowest range: 125.50 msec

Highest range: 627.00 msec

Thus, it can be concluded that the parameter extent of formant transition seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of extent of formant transition is rejected.

**Table XV: Parameter considered - Speed of formant transition**

Word: Come	Mean	SD	Range
Subject 1	11.32	1.87	5.70
Subject 2	8.73	2.34	7.20
Subject 3	11.15	2.40	7.57
Subject 4	8.11	1.15	3.80
Subject 5	10.80	3.36	10.18
Word: And			
Subject 1	14.22	2.15	5.40
Subject 2	10.00	3.00	9.52
Subject 3	10.31	3.11	9.80
Subject 4	11.82	0.95	3.00
Subject 5	13.43	2.68	8.50
Word: The			
Subject 1	8.48	2.56	8.90

Subject 2	10.43	3.19	10.85
Subject 3	11.80	5.25	16.40
Subject 4	9.39	1.50	4.20
Subject 5	10.26	2.83	9.40
Word: With			
Subject 1	9.44	1.63	4.90
Subject 2	13.72	2.80	8.90
Subject 3	14.12	8.84	6.30
Subject 4	11.24	2.91	9.20
Subject 5	14.71	3.01	7.80
Word: Eight			
Subject 1	5.07	1.89	5.00
Subject 2	7.66	2.36	5.80
Subject 3	7.23	2.74	7.90
Subject 4	5.47	1.09	3.20
Subject 5	5.94	1.30	3.40
Word: Suit			
Subject 1	7.04	1.34	3.90
Subject 2	6.04	0.32	1.00
Subject 3	5.06	0.99	3.10
Subject 4	5.70	1.44	4.00
Subject 5	6.36	1.95	5.90
Word: Case			
Subject 1	5.26	1.04	3.20
Subject 2	5.07	0.84	2.60
Subject 3	4.56	1.55	4.80
Subject 4	5.02	1.41	4.60
Subject 5	5.25	0.53	1.78

The study of table XV, shows that the lowest range and the highest range of variability for the parameter speed of formant transition.

Lowest range: 1.00 msec

Highest range: 16.40 msec

Thus, it can be concluded that the parameter speed of formant transition seems to be a factor which varies considerably. Therefore, it may be a useful parameter in speaker identification. Hence, the hypothesis that there is no significant difference across the subjects in terms of speed of formant transition is rejected.

The range of minimum and maximum variability for both inter-subject and intra-subject variability with respect to all the parameters considered for each word are summarized below.

Parameters	Intra-subject variability		Inter-subject variability	
	Min	Max	Min	Max
Word duration	6.4 msec	703 msec	26.50 msec	211.65 msec
Vowel duration	1.0 msec	76 msec	20.00 msec	107.00 msec
Burst duration	1.0 msec	6.9 msec	2.01 msec	22.50 msec
VOT	1.5 msec	87 msec	5.10 msec	26.70 msec
Closure duration	2.0 msec	135 msec	10.00 msec	73.20 msec
Frication duration	-	-	31.20 msec	137.50 msec
FO	2.0 Hz	89 Hz	6.40 Hz	53.70 Hz
Intensity	0.4 dB	19 dB	6.00 dB	17.00 dB
F1	9.4 Hz	975 Hz	70.00 Hz	258.00 Hz
F2	8.0 Hz	1130 Hz	196.00 Hz	784.00 Hz
F3	15.0 Hz	1325 Hz	153.00 Hz	886.00 Hz
F4	23.0 Hz	2847 Hz	337.00 Hz	1102.00 Hz

Duration	2.0 msec	118 msec	5.00 msec	85.00 msec
Extent	31.0 msec	637 msec	125.50 msec	627.00msec
Speed	0.14 msec	19.9 msec	1.00 msec	16.40 msec

As seen from the above table, inter-subject variability is greater than intra-subject variability for most of the parameters.

Latha, J. (1987) had conducted a similar study based on analysis of acoustic features.

Words extracted from sentences were used. A total of 30 inter-speaker and four intra-speaker pairs and one pair for test-retest reliability were prepared. Three judges considered could identify the speakers correctly (95.5%). The acoustic features found to be helpful in verifying the speakers in her study were overall clarity, total duration of the word and duration of the individual phonemes, frequency range of burst, frequency range of noise, energy concentration, voice onset time.

Sambur (1973) measured the format structure of vowels, the duration of certain speech events, the dynamic behaviour of the formant contours, various aspects of the pitch contour throughout on utterance, formant bandwidths, glottal source 'poles' and, pole and zero locations during the production of nasals and strident consonants. The measurements that were found most useful were related to the nasals, certain vowel resonances, certain temporal attributes and average fundamental frequency.

As seen from the present study the parameters: word duration, vowel duration, burst duration, voice onset time, closure duration, frication duration, fundamental frequency, intensity, formant frequencies  $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ , formant transition in terms of duration, extent and speed are useful in the process of voice identification since significant differences in terms of intra-subject and inter-subject variability for all parameters have been seen. Therefore, they should be used in the process of speaker identification.



## **CHAPTER - V**

### **SUMMARY AND CONCLUSION**

Currently, the field of speaker identification is fast gaining importance. Numerous research has been conducted, all with the aim of identifying speaker with the help of his voice.

The greatest need for this perhaps is in the field of Forensics Science. Voice print technology has not yet been accepted universally for acquitting a suspect. Identifying a speaker based just on his voice signature would indeed be a much awaited breakthrough.

Further, the use of voice identification procedures can be applied to various acts of life. Example: voice operated machinery, voice assisted cheque system at the banks, for security purposes, for production of synthetic speech, for identifying the presence of any disorder, etc.

But for all this, a thorough knowledge of the parameters involved in voice identification, namely - word duration, vowel duration, burst duration, voice onset time, closure duration, frication duration, fundamental frequency, intensity, formant frequencies F1, F2, F3, F4 formant transition in terms of duration, extent and speed, affecting it, changes seen due to age progression, differences seen between the sexes, contextual differences effects of attempts at disguising voice, etc. is a must. So, strong base regarding aspects of normal voice has to be built.

The current study is aimed at identifying the parameters which remain reasonably constant on repeated measures and across subjects.

Here, the speech samples were collected from five normal speakers. Three sessions with an interval of two days between them were considered to account for both intra- subject and inter-subject variability.

The samples were subjected to spectrographic analysis to obtain following parameters after which statistical analysis was carried out to arrive at the final values.

The parameters considered:

1. Word duration
2. Vowel duration
3. Burst duration
4. Voice onset time
5. Closure duration
6. Lead voice onset time
7. Frication duration
8. Fundamental frequency
9. Intensity
10. Formants  $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ .
11. Transition of formants: Formant transition duration, extent of formant transition, speed of formant transition

## **RESULTS**

The range of minimum and maximum variability for both inter-subject and intra-subject variability with respect to all the parameters considered for each word are summarized below.

Parameters	Intra-subject variability		Inter-subject variability	
	Min	Max	Min	Max
Word duration	6.4 msec	703 msec	26.50 msec	211.65 msec
Vowel duration	1.0 msec	76 msec	20.00 msec	107.00 msec
Burst duration	1.0 msec	6.9 msec	2.01 msec	22.50 msec
VOT	1.5 msec	87 msec	5.10 msec	26.70 msec
Closure duration	2.0 msec	135 msec	10.00 msec	73.20 msec
Frication duration	-	-	31.20 msec	137.50 msec
FO	2.0 Hz	89 Hz	6.40 Hz	53.70 Hz
Intensity	0.4 dB	19 dB	6.00 dB	17.00 dB
F1	9.4 Hz	975 Hz	70.00 Hz	258.00 Hz
F2	8.0 Hz	1130 Hz	196.00 Hz	784.00 Hz
F3	15.0 Hz	1325 Hz	153.00 Hz	886.00 Hz
F4	23.0 Hz	2847 Hz	337.00 Hz	1102.00 Hz
Duration	2.0 msec	118 msec	5.00 msec	85.00 msec
Extent	31.0 msec	637 msec	125.50 msec	627.00 msec
Speed	0.14 msec	19.9 msec	1.00 msec	16.40 msec

### CONCLUSION

There is a significant difference among the parameters when both inter- and intra-subject variability are considered.

As seen from the table, inter-subject variability is greater than intra-subject variability for most of the parameters.

Thus the hypothesis stating that the parameters involved in speaker identification do not vary is rejected.

**Recommendations**

1. The study could be conducted on a larger population.
2. Only male subjects have been considered here. Studies involving both male and female subjects, the difference and the range of variability between them could be conducted.
3. Only subjects between age range of 20-30 years were considered here.
4. Studies considering longer period of time between the initial and final recordings could be conducted.

### **BIBLIOGRAPHY**

- Ananthapadmanabha, T.V., Steven, K.N. (1991). "Acoustic properties contributing to the classification of place of articulation for stops". *Journal of Acoustical Society of America*, 91(4), 2472 (A).
- Black, J.W., Lashbrook, W., Nash, E., Dyer, H.J., Podrey, C, Tosi, O.I., Truby, H., (1973). "Reply to speaker identification by speech spectrograms : Some further observations". *Journal of Acoustical Society of America*, 54, 535-537.
- Bumstein, S.E., Stevens, K.N. (1980). "Perceptual invariance and onset spectro for stop consonants in different vowel environments". *Journal of Acoustical Society of America*, 54, 532-534.
- Bolt, H.R., Cooper, S.F., David, E.E., Denes, B.P., Pickett, M.J., Stevens, N.K. (1973). "Speaker identification by speech spectrograms : Some further observations". *Journal of Acoustical Society of America*, 54, 531-534.
- Bolt, H.R., Cooper, S.F., David, E.E., Denes, B.B., Pickett, M.J., Stevens, N.K. (1970). "Identification of speaker by speech spectrogram : A scientists view of its reliability for legal purposes". *Journal of Acoustical Society of America*, 47, 597-612.
- Bonne.D. (1971) "Voice and voice therapy" Englewood Cliffs. New Jersey; Prentice Hall.
- Cole, R.A., Rudnicki, A.I., Zue, V.M. (1979). "Performance of an expert spectrograph reader". *Journal of Acoustical Society of America*, 65 (S1), S(81) (A)
- Coleman, R. (1973). "Speaker identification in the absence of inter-subject differences in glottal source characteristics". *Journal of Acoustical Society of America*, 53, 1741-1743.
- Darby. (Ed) (1981), "Speech evaluation in medicine" Gruene Stratton Inc. New York.
- Darby. (Ed) (1978), "Speech evaluation in Psychiatry" Gruene Stratton Inc. New York.
- Endress, W., Bambach, W., Flossa, H. (1971). "Voice spectrograms as function of age, voice disguise and voice imitation". *Journal of Acoustical Society of America*, 49, 1842-1848.
- Farnsworth, L.M., Mullennix, J.W. (1995). "The effects of talker variability across CV and VC environments". *Journal of Acoustical Society of America*, 97(5), 3249 (A).
- Fairbanks, G. (1940) "Voice and Articulation Drill Book" Ed2, Harper and Row Publishers, New York.
- Glass, J.R., and Zue, V.W. (1984). "Acoustic characteristics of nasal consonants in American English". *Journal of Acoustical Society of America*, 76 (S1), S(15) (A).

- Greene, B.G., Pisoni, D.B., and Carrell, T.D. (1984). Recognition of speech spectrograms". *Journal of Acoustical Society of America*, 75, 32-43.
- Haten, T.P., (1982) "Automatic Speech Analysis and Recognition" D. Reidd Publishing Company Holland.
- Hazen, B. (1973). "Effects of context on voice print identification". *Journal of Acoustical Society of America*, 53, 354 (A).
- Hazen, B. (1973). "Effects of different phonetic contexts on spectrographs speaker identification". *Journal of Acoustical Society of America*, 54, 650-660.
- Hollien, H. (1974). "Peculiar case of voice prints". *Journal of Acoustical Society of America*, 56, 210-213.
- Hollien, H., Majeswski (1977). "Speaker identification by long term spectro under normal and distorted speech condition". *Journal of Acoustical Society of America*, 62(4), 975-980.
- Jonathan, H. (194). "The contribution of the murmur and vowel to the place of articulation distribution in nasal consonants". *Journal of Acoustical Society of America*, 55(2), 397.
- Klatt, D. (1974). "Acoustic characteristics of /w, r, l, y/ in sentence contents". *Journal of Acoustical Society of America*, 55(2), 397.
- Kresta (1962,1962a), Cited from Tosi (1979) "Voice Identification, Theory and Legal Application. University, Park Press, Baltimore.
- Latha, J. (1987). "Speaker identification by spectrograms". An unpublished Master's dissertation, University of Mysore.
- Mani Rao and Agrawal, S.S., (1984): "A method for speaker verification by comparison of spectrograms using novice examiner". *JASI*, Vol.12, No.3, 48-56.
- McGhee (1937,1944) Cited from Tosi (1979) "Voice Identification, Theory and Legal Application". University, Park Press, Baltimore.
- Papcun, G., Ladefoged, D. (1974), "Two voice print cases". *Journal of Acoustical Society of America*, 55, 463(A).
- Perkins (Ed) (1977), "Speech Pathology". The CV. Mosby Company, Saint Louis.
- Pollack (1954), Cited from Tosi (1979) "Voice Identification, Theory and Legal Application". University, Park Press, Baltimore.
- Pronovost, W., (1942) "An experimental study of methods for determining natural and habitual pitch". *Speech Monograph*-9.
- Rabiner, L.R., Wilbon, J.C. (1979). "On the use of clustering for speaker dependent isolated word recognition". *Journal of Acoustical Society of America*, 66(S1), 535(A).

Reich, A., Moll, K., Curtis, J. (1976). "Effect of selected vocal disguises upon spectrographic speaker identification". *Journal of Acoustical Society of America*, 60, 919-925.

Sambur, H.R. (1973). "Speaker recognition and verification using lined prediction analysis". *Journal of Acoustical Society of America*, 53, 354 (A).

Samuel George (1973). "A study of the fundamental frequency, voice and natural frequency of vocal tract on Indian population of different age range". Dissertation submitted to University of Mysore.

Santon, J.P.H. "Description of contextual factors affecting duration". *Journal of Acoustical Society of America*, 94(2), 1278 - 1385.

Sommer, M.S., Nygaard, L.C., and Pisoni, D.B. (1994). "Stimulus variability and spoken word recognition : Effects of variability in speaking rate and overall amplitude". *Journal of Acoustical Society of America*, 96(3), 1314-1324.

Stack, J.W. (1993). "Effects of speaking rate and stress on vowel durations and formant structures". *Journal of Acoustical Society of America*, 93 (2296).

Stevens, K.N., Blumstein, S.D., Glaksman, C, Burlon, M., Kurowshik (1992). "Acoustic and perceptual characteristics of voicing in fricative and fricative clusters". *Journal of Acoustical Society of America*, 91(5), 2979-3000.

Stevens, K.N., Williams, C.E., Carbonell, J.R., and Woods, B. (1968). "Speaker authentication and identification. A comparison of spectrographic and auditory presentation of speech material". *Journal of Acoustical Society of America*, 43, 1596-1607.

Su, L., Li, K.P., Fu, K.S. (1974). "Identification of speakers by use of nasal coarticulation". *Journal of Acoustical Society of America*, 56, 1876-1882.

Tosi, O.I., Oyer, H., Lashbrook, W., Pedrey, C, Nocil, J. , and Nash, E. (1972). "Experiments on voice identification". *Journal of Acoustical Society of America*, 51, 2030-2043.

Tosi,(1979). "Voice Identification, Theory and Legal Application". University, Park Press, Baltimore.

Wolf, C.G. (1972). "Efficient acoustic parameters for speaker recognition". *Journal of Acoustical Society of America*, 51, 2044-2056.

Young and Campbell (1967). "Effect of context on talker identification". *Journal of Acoustical Society of America*, 42, 1250 1254.

Zue, V.W. (1979). "The use of content in spectrogram reading". *Journal of Acoustical Society of America*, S(1) S(81) (A).