


**LENGTH OF UTTERANCE FOR RELIABLE
SPEAKER IDENTIFICATION**

Register No. L0380001

**A Dissertation Submitted in Part Fulfillment of
Final Year MSc (Speech Language Pathology)
University of Mysore
Mysore**

**ALL INDIA INSTITUTE OF SPEECH AND HEARING,
NAIMISHAM CAMPUS, MANASAGANGOTTHRI
MYSORE – 570 006**

MAY - 2005



*Dedicated
to
AIISH
with all gratitude
for giving me
wonderful opportunities
rich experiences
and
beautiful memories...*

Certificate

This is to certify that the dissertation entitled "**Length of Utterance for Reliable Speaker Identification**" is a bonafide work done in part fulfillment for the degree of Master Science (Speech Language Pathology) of the student with Register No. L03 80001. This has been carried out under the guidance of a faculty of this Institute and has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Mysore,
May 2005.



DIRECTOR

All India Institute of Speech & Hearing,
Naimisham Campus, Manasagangothri,
Mysore - 570006

Certificate

This is to certify that this dissertation entitled "**Length of Utterance for Reliable Speaker Identification**" has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier in any other University for the award of any diploma or degree.



PROF. M. JAYARAM,

Director,

All India Institute of Speech & Hearing,
Naimisham Campus Manasagangothri,
Mysore.

Mysore,
May, 2005

Declaration

This dissertation entitled “**Length of Utterance for Reliable Speaker Identification**” is the result of my own study under the guidance of Prof. M. Jayaram Director, All India Institute of Speech and Hearing, Mysore and not been submitted in any other University for the award of any degree or diploma.

Mysore,
May, 2005

Reg. No. L0380001

ACKNOWLEDGEMENTS

*I express my sincere thanks to my guide, **Prof. M. Jayaram**, Director, All India Institute of Speech & Hearing, for his valuable guidance. Thank you Sir, for always sparing time and for being patient in explaining my doubts, in spite of your busy schedule. Thank you for making my first research worthwhile and for making me realize what is “perfection!”*

*I thank **Dr. S. R. Savithri**, Reader & Head, Dept. of Speech–Language Sciences, for permitting me to use the equipments for this study. I also thank her for her valuable suggestions and help. Ma’am you have always been a biiiiiig inspiration for all of us!*

*Special thanks to **Yeshoda madam** for her advice, suggestions, support and encouragement. Thank You ma’am, without you I cannot imagine how I would have kept my work going. Your constant monitoring kept me on my toes and it did mean a lot. Thank you very much.*

*I thank **Prema ma’am** for kindling research interest in me, for being my mentor and a guiding light. Thank you ma’am for patiently listening to all my grievances and providing solutions... Your understanding and help was a constant source of inspiration.*

*I wish to thank **Ajish Sir** for his support, guidance and good wishes.*

*Thanks to **Goswami Sir**, for always helping me see the lighter side of life...for the timely advice and moral support...for the practical suggestions...for all the laughter and fun...Thank you Sir!*

*Thanks to **Sreedevi ma’am, Pushpa ma’am and Shyamala ma’am** for the concern!*

*I wish to thank all my **teachers** at AIISH (Animesh sir, Swapna ma’am, Asha ma’am, Manjula ma’am(s), Vijayshree ma’am, Sujatha ma’am, Geetha ma’am and all the others) who have helped me all through these 5 years to grow up as a Speech & Hearing professional!*

*I sincerely thank all the staff and research scholars of the Dept. of Speech – Language Sciences for all their help and support- **Ms. Rohini, Ms. Seema, Ms. Ananthi, Mr. Santosh, Mr. Sairam, Ms. Jayashree, Ms. Kalaiselvi, and Mr. Arun B. T.***

*I can’t find words to express my gratitude to **Kalai ma’am and Arun Sir** for their patience, co-operation, suggestions, motivation, timely help and encouragement.*

*I’m very grateful to **Sairam**, for the ready help that he rendered during my topic selection and during my Research Proposal.*

*Thanks to **Jayashree** for being instrumental in my data collection, and also for all the words of encouragement and moral support. I wish you all the best!*

*I thank all the **participants** of the study, both speakers and listeners for their enthusiastic participation, co-operation and patience.*

*I wish to thank **Shubha ma'am** for being so very kind and helpful in arranging for appointments for me to meet Sir.*

*I sincerely thank **Vasantha Lakshmi ma'am and Venkatesan Sir** for helping me with statistical analysis.*

*I wish to thank the librarians for all their help-**Mahadeva sir, Lokesh sir, Chandrashekar sir** and of course, **Raju...***

*Thanks to **Parimala ma'am** for helping me with the typing.*

*Big thanks to **Mr. Pavan, Mr. Satish** for providing with the computer peripherals, and also the staff of Electronics dept. for being kind enough in helping with the installation. Thank you, it was of grrrrrrrr8 help!*

*Thanks to **Mr. Shivappa & Co. (Xerox shop)** for their help and co- operation...for all the troubles they took to make our work worth it!*

***Chaya-** I'm indebted to you for introducing me to Speech & Hearing and for always motivating me. Due thanks to Mrs. Subhadra (Chaya's mother & my dear teacher) for all the encouragement, support and help.*

*My dearest **Mom & Dad-** Thank you for everything! Thanks Mom, for your sense of humor (PJs), your wake- up calls, and your ability to make my **BIIIIG** problems look very **SMALL**...Thanks Dad for the **NO TENSION** talks....Your support and understanding meant a lot...My dearest **sister**, you have been my strength and inspiration throughout... For all that you have been and for all that you are to me, **THANK YOU** is a very small word!!!!...**Karun**, my dear brother-in- law, the best complement to our family! ... Thank you for your support and care...My Loving Family- God's best gift to me...what would I do without u all?*

*Thanks to all my **seniors** at AIISH for the wonderful "Interaction" (????!!) and for being such great guiding stars! Sapna ma'am, Beula ma'am, Naveen sir, Purnima ma'am, Jayaradha ma'am, Prachi ma'am, Komal ma'am, Aruna ma'am, Mathew sir, Kiru, Gowri, Seetha, Mili chechi, Vimi mummy, Ramya, Namrata, Jessy chechi, Komal, Mammu, Ammu, Rakhi, Vani, Sneha, Banu, Annie mol,...thank you!!!*

***Vandy**...thanks for all those mad times...yes doctor's daughter, "laughter is the best medicine!!!"*

*Cartloads of thanks to all my terrific **classmates (BSc and MSc)**, "the trendsetters", each one uniquely special! Will cherish the unparalleled wild fun that we had together- the crazy times, co-operative work, and friendship- that kept us going through all the hard times... Yes, we gave "adversity" an "identity crisis" when we laughed through the tough times!!!*

*My dearest room-mates: **Rohini, Suji and Tanu**...Life was beautiful with you guys around.....the long chats, the fights with insects, the "mango tang" and all the fun that we had!!!! Thanks for all the love and care. It was the most beautiful part of my life at AIISH.*

Thank you **Nuzha & Amy**, my madly sweet room-mates!!! Cheers to us, the “crazy trio”. Thanks Nuz, for being there during those last moments of tension!!! We sure had a “time” together!!! All the best for everything in life!

My dear fellow grihalakshmis- **Namita & Deema**- “Home Sweet Home”...Wasn't it a wonderful experience? Indeed, it made us discover our versatility!!! Thank you Namita, for your musical inspiration while I was deciding on the topic and for all the advice, support and care at the right times.

Mili, Pooja, Rajani, Deema & Amy - Thanks for being there through all my good and bad times. Thank you for the support... for the prayers... for the fun filled moments... for the smiles... for the reassurance... for everything...U guys are grrrrrr8! Rajani, my research partner! I enjoyed all the ‘dyslexic’ times with you...Thank you for making me believe in myself...Your support was my strength, always!... Deemz, my dearest, thank you for being there for me...your close-up smile has always lighted up my days!... Pooja, thanks for unfailingly lending a helping hand always... God bless U... Mili, the girl with strong will power, you have always enriched me with your prayers, I could count upon you anytime...Thank you! Amy dear, the naughty girl, I'll never forget your words: “you think you know me?”

Sailu...we had a great time with our skits!!! U sure have a future designing costumes for Kabile!

Thank you **Sree** (aishu), **Mini** (vermagupta_australia), and **Ashly** (konji angel), for the good times!

Devi...All the very best for everything in life...Keep up the belief that “everything happens for the best!”

Thanks **Dheepu** (class fund, not again), **SoNi** (baby), **Geethu** (regular lib user) **Aditee** (wow June 18th!), and **Divya** (MLT specialist)...and **Suba** (pachai nirame), **Raje** (jingli), **Lisha** (achcha, achcha), and **Manika** (dancer) for all the crazy fun- filled times that we had in the hostel.

The Panchatantra- Sudhakar, Venu, Radha, KD and JK...also Amit (chachu), Venkat (Bhenkat), Saugat (Rocky), Srikant (Style bhai), Hari (Pragas), Kartik (Kartiku), Sandeep (Sandy)...and Kushum (KKV), Dhananjay (bhaiya), Sujeet (baby)... it was nice with you guys around as classmates.

My dearest **juniors**... “Summer rains are so pleasant”....So are my memories with you all. Thanks for all the fun, naughtiness, innocence, love, care, help and co- operation. Taking care of, and being taken care by, was sweet!!! You guys made life at AIISH memorable!!!...Rahana, Priya, Suma, Purushoth, Nitish, Yatin...thank you for the unending support! Suppi, Sachin, Alfi, Neeti, Saoji, Himani, Deepa, Sheetal, Ganju, Svetha, Fellow-Pisceans, Prasi, Janani, SimSom, ‘Trimurthis’, Balaji... Thanks for those wonderful times... Kamala, Sruthy, Priya Mathew...Keep up the good work... Gurdeep (paithyakaara), Ismail, Gnanu, Arun...Wishing you the best! Mohan, Achu, Pravin, Darshan, Priyanka, Ramya, Navitha, Priya(s), Minakshi(s), Rima, Sumita, Mani, Rachna, Anu, Maharani, Powlin, Veda, Pradyumn, Ramu, Noor... and the whole lot of u! All the very best!!!

*Due thanks to all my **teachers and sisters** at St. Joseph's School for bringing me up morally and spiritually and for making me what I am today!...Specially to Lathu, for being a grrrrrrr8 teacher and a wonderful friend, and Ramesh Sir for all the inspiration and encouragement. Also Sr. Grace, Ms. Leema, Ms. Sujatha, Ms. Nalini, Ms. Mary Pearl, Ms. Loyola, Ms. Mehfarid, Ms. Shanti, Ms. Sashikala, Ms. Pushpa, Ms. Sundari, Ms. Selvi Jesintha, Ms. Shenbagavalli...I owe you a lot!*

***My friends** Nikila, Elizabeth, Geetha, Chetana, Sandra...Together through ups & downs...Together in dreaming & discovering...we had fun growing up Together, didn't we? Ramesh, Kamal and Ganesh...will cherish the times spent with you! Venky, thanks for giving me company on Yahoo...Babu & Sam, thanks for your concern, thanks for keeping me awake with your sms!!! Thank you for all the lovely times!*

***Sandeep**, thanks for being my confidant, my best friend and a wonderful brother.*

"Together we moved on and on...

Days passed and years rolled,

Still we go,

Perpetually we will!" ...I remember these lines of yours...Rain or Sun, thanks for always being there for me...

*Thanks to the **Almighty** for guiding me and taking me safely through my journey of "post graduation".*

CONTENTS

S. No.	CHAPTER	Page No.
1	Chapter 1 Introduction.....	1
2	Chapter 2 Review of Literature.....	9
3	Chapter 3 Method.....	22
4	Chapter 4 Results.....	26
5	Chapter 5 Discussion.....	31
6	Chapter 6 Summary and Conclusions.....	34
7	References.....	37

CHAPTER 1

INTRODUCTION

Just as an artifact carries traces of its production – a carving, the marks of the chisel, or a painting, the brush strokes, and both of them the style of the artist – so a sample of speech carries the imprint of its originator (Nolan, 1997). It is true of a person's speech signal also. Indeed, as Corsi (1982) opines, a person's voice is a complex acoustic signal which encodes various kinds of information, among them, some of the anatomy and physiology of the speaker.

The notion that an individual has “a voice” by which he can be recognized is a natural one. This is based on our day-to-day experience in successfully recognizing people by their speech alone – typically over the telephone. The process of recognition seems to be so natural that the notion was adopted by many speech scientists without fundamental scrutiny; with the result that the usual question posed was not whether individuals could be uniquely recognized from their voices, but how this recognition could be most effectively and reliably carried out in an objective way (Nolan, 1983).

The kind of activity covered by the term speaker recognition is conceptually straight forward, and definition abound. Hecker (1971) suggests that speaker recognition is any decision making process that uses speaker dependent features of the speech signal. Speaker recognition, according to Atal (1976), is any decision making process that uses some features of the speech signal to determine if a particular person is the speaker of a given utterance.

Speaker recognition (or voice recognition) is a general concept which subsumes “speaker identification” and “speaker verification.” Basically, it relates to the overall process of recognizing a person from his/her speech, and/or voice, and doing so, by assessment of these factors alone.

Speaker verification through a comparison of a test sample of speech with a reference sample from just one speaker requires a preset similarity threshold, and usually yields one of four kinds of decisions: correct acceptance, correct rejection, false acceptance, false rejection (although a “no decision” response may also be permitted). The assumption underlying speaker verification tasks is that both test and reference samples are from cooperative speakers. The speech samples employed are under the operator’s strict control. The verification trials are always “closed” (i.e., the speaker is a member of the group).

In speaker identification (and elimination), an utterance from an unknown speaker has to be attributed, or not, to one of the members of the population of known speakers for whom reference samples are available. Here, the number of decisions increase with the size of the reference population (Nolan, 1983).

Under the overall heading of speaker recognition, it is necessary to distinguish a number of distinct fields of study. Bricker and Pruzansky (1976) recognize three major methods: speaker recognition by listening; by machine; and by visual inspection of spectrograms. Speaker recognition by listening involves the study of how human listeners achieve the task of associating a particular voice with a particular individual and indeed to what extent such a task could be performed.

Under speaker identification, three types of recognition tests can be carried out: closed tests, open tests and discrimination tests (Tosi, 1979). In a closed test, it is known that the speaker to be identified is among the population of reference speakers, whilst in an open test, the speaker to be identified may or may not be included in that population. Thus, in the closed test, only an error of false identification may occur, whilst in open tests, there is an additional possibility of incorrectly eliminating all the members of the reference population, when in reality, it included the test speaker. In a discrimination test, the decision procedure has to ascertain whether or not two samples of speech are similar enough to have been spoken by the same speaker; errors of false identification and false elimination are possible (Nolan, 1983).

Experiments assessing the value of the particular parameters for speaker recognition have most frequently adopted the closed- set design. The reason for this is not that this design best approximates real life applications – it is in fact the one least likely to occur in forensic cases – but rather that it gives the most straight forward comparison of parameters.

There are subjective as well as objective methods of voice identification as shown in Figures 1.1 and 1.2. The subjective procedures are based on either audio or visual comparisons of signals, while in the objective procedures, a computer usually compares the visual representation of an audio signal from one or more speakers. In any measurement or comparison, it is generally believed that objective procedures yield more valid results, and the area of speaker identification is no exception to this. However, this need not be correct because any process of speaker or speech identification must also relate to human experience.

METHODS OF VOICE IDENTIFICATION

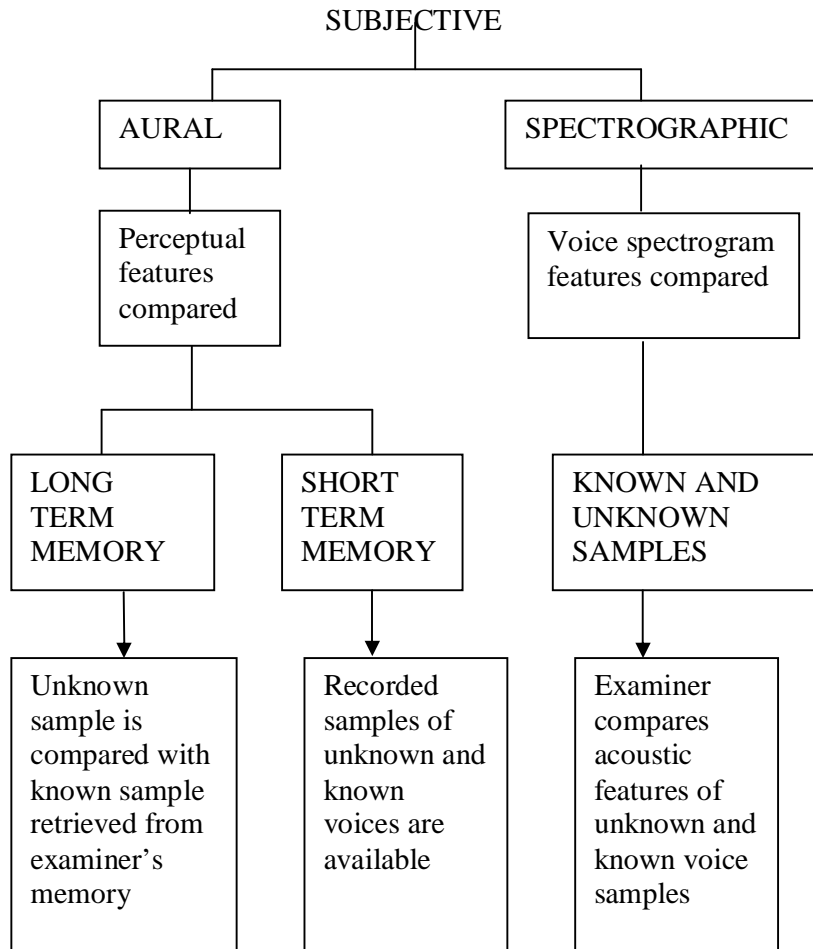


Figure 1.1: Subjective methods of voice identification

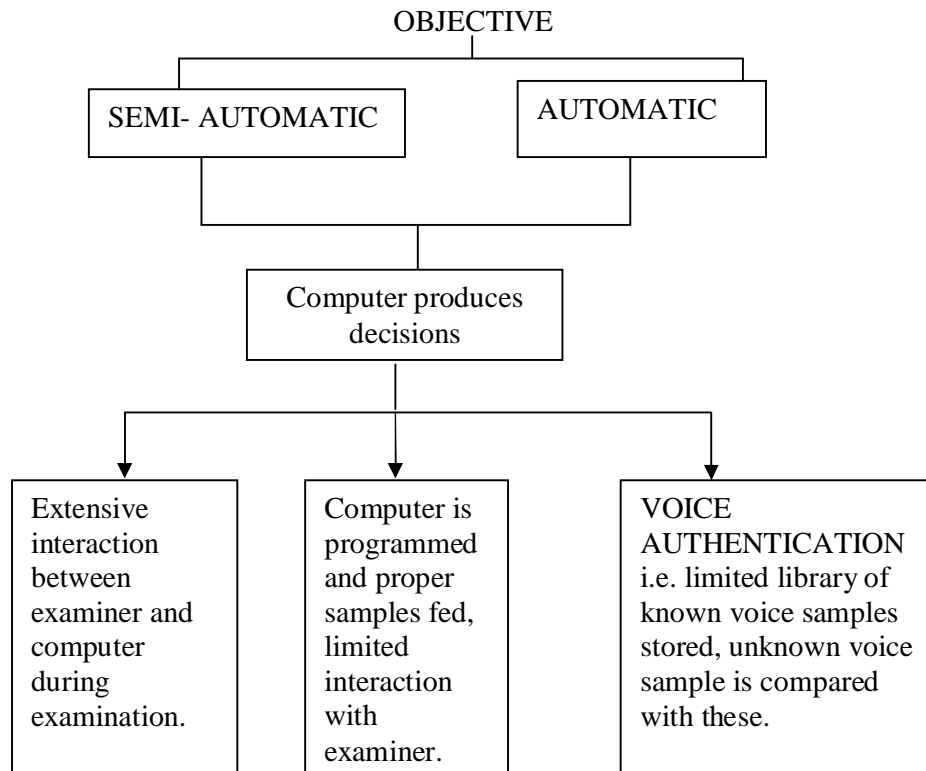


Figure 1.2: Objective methods of speaker identification

Aural examination of recorded voices is a subjective method of talker identification. A listener may use the long term memory process or the short term memory process to identify or eliminate an unknown talker as being the same as a particular known one. These two memory processes are used according to the particular situation as follows:

- The long term memory process is utilized when the voice to be identified is familiar to the listener.
- The short term memory process is used when the unknown and known voices to be compared are not familiar to the listener, but they are continually and permanently available through audio- tape recordings.
- The long term memory process can be used by any witness in a court of law. The short term memory process demands assistance of an expert witness.

The success of aural recognition based on long term memory depends, among others, on such factors as the remembrance or the familiarity of the speaker to the listener with the known voice, the time lapsed since it was last heard by the listener, the homogeneity of the talkers involved, and the discriminating ability of the listener.

In the past, numerous studies have been carried out on different factors related to speaker identification: McGehee (1937) studied memory decay for voices and found that decay in correct identifications occurred over time. She also attempted to determine such things as, whether men or women were best at recognizing voices, and how other factors (speakers with foreign dialects, voice disguise etc.) affected the recognition process. She reported that male auditors can be expected to perform at levels better than those for women. On the other hand, Bull and Clifford (1984) reported that females performed better than males in a task of speaker identification. DeJong (1998) and Koster (1981) recognized the fact that distinctive voices were easy to identify; that is, the idiosyncratic characteristics that a speaker possesses can make the speaker easily identifiable.

The size or duration of the sample required for correctly identifying a speaker has been the subject of only two studies, but the results are far different to make any meaningful decision.

Pollack et al. (1954) reported that identification accuracy can be improved by increasing speech sample, but that the increase in accuracy will only occur for periods of up to about 1200 ms. Beyond this, accuracy did not seem to be related to duration, but rather to the speaker's phonemic repertoire. However, Pollack et al. (1954) did not define a threshold, or did not indicate the accuracy level of identification that they were seeking.

In another study, Kunzel (1995) indicated that, in German, a sample of 30 seconds was necessary to attempt any type of speaker identification, but again indications on a preset similarity threshold in this study are not available. Apparently, these two studies talk in terms of the duration of speech sample which is speaker dependent (speaking rate). It is possible, that two speakers can utter, depending upon their speaking rate, widely varying number of syllables in a given unit of time.

It can, in general, be said that the greater the opportunity one has to listen to a particular speaker, the 'greater the accuracy of identification will be' (Yarmey, 1995). However, Yarmey (1995) warns that false positives often will increase in parallel with rise in correct identification.

STATEMENT OF THE PROBLEM

Therefore, the purpose of this study was to determine the minimum length of speech sequence (in terms of syllables) required for correct identification of speakers from a closed, but unfamiliar set of speakers.

NEED FOR THE STUDY

The results of the studies in this area are equivocal. In fact, the report of Kunzel (1995) is more of an observation, than the result of a controlled study. Furthermore, as said earlier, indicating the length of the speech sample in terms of time is relative in the sense that, in a speech sample of unit time, there may be less or more number of syllables, depending on the speaking or reading rate of speakers. Expressing the minimum length

of speech in syllables is a more valid indicator. Besides, no accuracy threshold (percent of correct identification) seem to have been set in either of these two studies, which is an important precondition in identification tasks as these. Therefore, further studies are required in this area.

OBJECTIVES OF THE STUDY

The objectives of the study are to

- a) determine the minimum length of utterance, defined in terms of syllables, required for correct speaker identification, from a closed set of speakers,
- b) investigate whether or not the gender of both speakers and listeners is a variable in such recognition tasks, and
- c) to study whether language is a factor to be considered in determining adequacy (length) of sample required for speaker identification.

IMPLICATIONS

Speaker identification, as applicable in forensic practice, has to be made in varying conditions. The available speech sample may or may not be adequate for correct speaker identification. Therefore, the results of the present study will be helpful in validating conditions under which identifications are done, and their results. Similarly, the results of the present study would be a guide for all future experiments on speaker identification on the length of speech to be selected for analysis.

CHAPTER 2

REVIEW OF LITERATURE

A person's voice is a complex acoustic signal which reflects certain aspects of the anatomy and functioning of the mechanism generating it. As the structure of the voice mechanism is different in different persons, with regard to size, volume and other physical aspects, it is likely that all voices are different. It means that each speaker is bestowed with a uniqueness. Thus rose the notion that different speakers could be identified based on their voice alone. However, as research continued in this direction, it became evident that other aspects of speech like articulation and prosody are as important as voice in speaker identification. As the field of speech/ voice identification or speaker identification has tremendous implications in forensic medical practice, computer operations, and voice physiology, it is quite natural that this area has been a fertile field for research. Speech pathologists, acoustic engineers, psychologists, communication engineers, physiologists, and more recently speech scientists have carried out extensive research in this area.

Speaker recognition (voice recognition) is a general concept which subsumes 'speaker identification' and 'speaker verification'. Basically, it reflects the overall process of recognizing a person from his/ her speech or voice. Speaker recognition is any decision making process that uses some speaker- dependent features of the speech signal (Hecker, 1971; Atal, 1976). Bricker and Pruzansky (1976) recognize three different methods of speaker recognition, namely, speaker recognition by listening, by machine, and by visual inspection of spectrograms. The present study concerns itself with the first of the three dimensions.

Perhaps the first significant experiment in the area of aural examination, using the long term memory process, was performed by McGehee (1937, 1944). She used a total of 31 male and 18 female talkers, reading a paragraph of 56 words. A total of 740 undergraduate students with no special training were employed as listeners in this experiment in which live voices were used. Listeners were divided into 15 panels, each panel participating in at least two sessions. They listened to a talker behind a screen reading a paragraph in the first session. Five talkers, including the one from the first session, read the same paragraph in the second session. Each listener had to identify the one whom they had previously heard. The interval between the two sessions ranged from one day to five months, and differed for different set of listeners. The average percentage of correct identification varied from 83% to 13%, according to the time elapsed; the higher percentage corresponds to a one day lapse, and the lower percentage corresponds to a five month lapse between the first and the second listening sessions.

McGehee also investigated the effects of disguising the voice by changing the pitch which drastically reduced the percentage of correct identifications. Other findings of this early study were that male and female voices were equally identifiable and that increasing the number of known talkers increased the percentage of correct identification. It should be noted that all tests of identification used in this experiment were the closed type using long term memory, and that no recordings were employed.

Speech/ voice being very unique to speakers, it is quite obvious that a large number of factors influence speaker identification. They are speaker- dependent factors (glottal) characteristics, resonance characteristics (familiarity), listener related factors

(age, familiarity with the speakers, gender, training received, profession, etc.), type of identification task (closed set, open set, discrimination test), mode of identification (visual examination and auditory perception) etc.

Speaker dependent speech/ voice characteristics

Delong (1998) and Koster (1981) recognized the fact that distinctive voices were easy to identify, that is, the idiosyncratic characteristics that a speaker possesses can make the speaker easily identifiable. Tartter (1991) studied the effect of whisper register on speech perception and reported 82% identification accuracy. Further, he stated that, independent of the register, there are acoustic cues, specific to a speaker's identity.

Coleman (1973) used 2 male and 10 female talkers and 28 listeners to perform a study of voice identification using short term memory and match / no-match discrimination tests. The influence of the speaker's glottal source was eliminated by using an artificial larynx vibrating at a fundamental frequency of 85 Hz to record speech samples for all talkers. Samples consisted of a 5 second segment of ongoing speech. The average percentage of correct identification was 90% for listeners with no special training who were forced to give positive decision in each trial. The study suggested that the resonances of the vocal tract are the clues for voice identification, rather than the glottal characteristics of the talker, including pitch.

Van Dommelen (1990) conducted identification tests on familiar voices using reiterant "ma" syllables and investigated whether the following cues were useful in speaker identification: F0 height, F0 contour and speech rhythm. He found that F0 height was a highly relevant cue in speaker identification. He also stated that the cues for recognition of familiar voices are not hierarchically fixed, but depend on speaker-specific voice characteristics.

Some authors took a different approach to studying aural identification, trying to determine significant perceptual attributes of talkers' voices in order to create reliable classification scales. These scales might help an examiner to perform aural discriminations on a systematic and consistent basis. Voiers (1964) isolated four significant perceptual scales - clarity, roughness, magnitude, and animation - to use as a means to discriminate among speakers. Holmgren (1967) found that pitch, intensity, quality, and speech rate scales helped better classify the uniqueness of each particular voice.

Aural examination Vs. Visual examination

One of the few studies in aural examination using both open and closed trials with the short term memory process was performed by Stevens et al (1968). In this study, the authors attempted to compare results obtained from aural examination with those from visual examination of spectrograms, using the same materials and the same examiners. They employed 24 talkers who were highly homogeneous from the point of view of perceptual attributes of speech. All of these talkers recorded a reading list of nine isolated words and two short sentences, all repeated 10 times. They recorded these materials twice, one week apart. These materials were loaded onto magnetic tape loops of 4.5 seconds duration, each loop containing two utterances of a short sentence. Spectrograms of these materials were also subsequently prepared. Six examiners performed open and closed tests of talker identification and elimination with these materials, using aural and visual examinations separately. In all the open and closed tests, the percentages of correct responses were significantly higher for aural examination, than for visual examination. For the closed tests, mean errors of false identification yielded by aural examination

ranged from 18% to 6%. Mean errors of false identification yielded by visual examination of spectrograms of the same materials ranged from 28% to 21%. For the open tests results were as follows: aural method: 8% to 6% error of false identification and 12% to 8% of false elimination; visual method: 47% to 31% error of false identification and 20% to 10% error of false elimination.

Tosi and Greenwald (1978) conducted a voice identification experiment employing 25 male and 25 female talkers. Four sentences (approximately 2.4 second duration each) were recorded twice through commercial telephone lines. A second recording session was held 6 months later and the same material was recorded twice, once in quiet and once in the presence of environmental noise. Spectrograms and aural materials for the experiment were prepared. Three types of voice identification tests were carried out: (1) voice identification by visual examination of a talker's spectrogram; (2) voice identification by aural examination of a talker's voice; (3) voice identification by combined aural and spectrographic examination of a talker's samples. Examiners were of two categories: (1) students of audiology and speech sciences who received approximately 1 week of training in spectrography prior to starting the experiment; (2) professional examiners certified by the International Association of Voice Identification. The results of the study suggested that (1) training of examiners is crucial for validity of results of a subjective method of voice identification based on aural and spectrographic examination of talkers' samples; (2) 6 months time elapsed between known and unknown talker samples do not produce significant errors of voice identification provided that the listener is a professionally trained person; (3) voice samples distorted by noise yielded a larger percentage of errors of voice elimination and voice identification; and (4) untrained examiners produced a wide range of errors.

Gender of the subjects

The question of whether male or female listeners identify speakers better has attracted some attention. Yarmey & Matthys (1992), Clifford (1980), Hollien & Schwartz (2000), Thompson (1985), and Yarmey (1995) have reported that, other things being equal, the gender of listeners do not appear to differ a great deal with respect to accuracy of speaker identification. As reported earlier, McGehee (1937), found that male auditors can be expected to perform at levels better than those for women. On the contrary, Bull and Clifford (1984) have reported that females perform better than males in tasks of speaker identification.

Disguise of speech and speaker identification

Speaker identification through one's natural voice/ speech is one thing, and speaker recognition when the speaker consciously modifies his speech/ voice through external means is altogether a different proposition. Logically, the disguised speech should make the task of speaker identification much more difficult because the listeners lose speaker dependent cues. This also obviates the importance of the factor of familiarity of speakers.

Reich, Moll and Curtis (1976) studied 40 adult male subjects in the age range of 21 to 42 years with the purpose of determining the effects of selected vocal disguises upon spectrograms and speaker identification. The subjects were instructed to utter a set of 4 sentences and a set of 3 sentences with 9 clue words in 2 separate sessions. The recordings were done directly onto a tape recorder, through a telephone line in a quiet environment and through a telephone line in a noisy environment. The subjects were

asked to utter the sentences in six different ways: (1) Normal speech; (2) Disguised like the speech of 70-80 years old persons; (3) Simulating severe hoarse voice; (4) Simulation of severe hyper nasal voice; (5) Slow rate; (6) Freely disguise. The spectrograms of session 2 undisguised speech were matched with disguised and undisguised speech of session 1. Four examiners compared the clue words in randomly ordered sentence pairs in terms of vowel formant frequencies, relative spacing of vowel formant frequencies, amplitude relationships between vowel formants, vowel formant bandwidths, stops of VC and CV formant transitions, frequency position and bandwidth of nasal resonance, location of spectral zeroes, spectrum and spacing of vertical striations, vowel and consonant duration, stop-gap duration, characteristic burst transients and patterns of fricative noise energy. The examiners were asked to rate the speech on a five point scale of decision certainty. They concluded that undisguised speech had significantly higher percentage of correct identification than other speech task, except slow rate speech. In general, nasal and slow rate were the least effective disguise, while free- disguise was the most effective. It was apparent that slow rate had less effect on the frequency of formants.

Reich and Duke (1979) studied the effects of selected vocal disguises upon speaker identification by listening. The experiment consisted of 360 pair discriminations presented in a fixed sequence mode. The listeners were asked to decide whether two sentences were uttered by the same or different speakers as well as to rate their degree of confidence in each decision. The speakers produced two sentence sets utilizing their normal speaking mode and five selected disguises. One member of each stimulus pair in the listening task was always an undisguised speech sample; the other member was either disguised or undisguised. Two listener groups were trained for the task: a naïve group of 24 undergraduate students, and a sophisticated group of three doctoral students and three

professors of speech and hearing sciences. Both groups of listeners were able to discriminate speakers with a moderately high degree of accuracy (92% correct when both members of the stimulus pair were undisguised. The inclusion of a disguised speech sample in the stimulus pair significantly interfered with listener performance (59% to 81% correct depending upon the particular disguise).

Hollien, Majewski and Doherty (1982) studied the perceptual identification of voices under normal, stress and disguise speaking conditions. The study attempted to assess the importance of the listeners being acquainted with the talkers. Speakers were 10 adult males who recorded speech samples under three types of conditions: (a) normal; (b) stress; (c) disguise. Three classes of listeners were utilized: (a) a group of individuals who knew the talkers; (b) a group of individuals that did not know the talkers but were trained to identify them; (c) a group that neither knew the talkers nor understood the language spoken. The analyses indicated that the performance between the groups was significantly different. Listeners who knew the talkers performed best while the non-English speaking listeners produced the lowest level of correct identification. The “middle” group, that is, the English speaking listeners was divided into two subgroups by the method of extremes. However, even in this case, the most competent of the subgroups still was significantly less able to identify the talkers than were the listeners who knew them; the least competent subgroup performed at about the same level as the listeners that did not speak English. Finally the analysis of the three types of speech revealed that the normal and stress conditions were not statistically different relative to the identification task whereas the disguised productions produced fewer correct identifications.

Other factors

Shirt (1984) reported that phoneticians can be expected to do somewhat better in speaker identification than the lay public, and that, training in phonetics resulted in only a minor advantage. Koster (1981) reported that phoneticians performed better speaker identification than the controls (students) in this study. Schiller & Koster (1998) reported that phoneticians correctly identified the target voices 98% of the time, whereas the controls only achieved a level of 92% correct; the difference between these means were statistically significant.

Goggin, Thompson, Strube & Simental (1991) investigated the role of language familiarity in voice identification. They conducted four experiments. In experiment 1, monolingual English listeners identified bilinguals' voices better when they spoke German. The opposite outcome was found in experiment 2, in which listeners were monolingual in German. In experiment 3, monolingual English listeners also showed better voice identification when bilinguals spoke a familiar language (English) than when they spoke an unfamiliar one (Spanish). However, English- Spanish bilinguals hearing the same voices showed a different pattern, with the English-Spanish difference being statistically eliminated. Finally, experiment 4 demonstrated that for English- dominant listeners, voice recognition deteriorates systematically as the passage being spoken is made less similar to English by rearranging words, rearranging syllables, and reversing normal text. Taken together, the four experiments confirm that language familiarity plays an important role in voice identification.

The degree to which speech and/or speech samples are non-contemporary is considered important to the speaker identification process. Hollien & Schwartz (2000) conducted speaker identification tests on non-contemporary speech. There are two dimensions to the problem; the first relates to the listener and, especially, to ear witness lineups. Here, the subject or witness is asked to make identifications at various times after having heard (but not having seen, of course), the speaker. It has been found that a person's memory for a voice decays over time. In the second case, it is the samples of the speaker's utterances which are temporally displaced. Non-contemporary speech samples pose just as difficult a challenge to the speaker identification process as does the decaying memory of a witness. Hollien and Schwartz (2000) found that the overall drop in correct identification over latencies from four weeks to six years was only about 15-25 per cent. Substantial amount of drop (of up to 31 per cent) occurred when the latency was about twenty years. So, they concluded that a listener's competency in identifying non-contemporary speech samples will show only modest decay over rather substantial periods of time and, hence, this factor should have only a minimal negative effect on the speaker identification process.

Duration of speech sample required for correct speaker identification

The size or duration of samples required for correctly identifying a speaker has also been studied, but the results of these studies are far too different to make any meaningful decision. It can, in general, be said that the greater the opportunity one has to listen to a particular speaker, the 'greater the accuracy of identification will be' (Yarmey, 1995). However, Yarmey (1995) warns that false positives often will increase in parallel with rise in correct identification.

Schweinberger, Herholz & Sommer (1997) measured the effects of increasing stimulus duration on the listener's ability to recognize famous voices. They also studied the influence of different types of cues (second voice sample, occupation, initials of the celebrity). Results indicated that voice recognition improved with stimulus duration (0.25 seconds to 2 seconds). They stated that voice naming is contingent on previous activation of person-specific semantic information.

Bricker and Pruzansky (1966) studied the effects of stimulus content and duration on aural voice identification. They used 16 examiners and 10 talkers with whom the examiners were familiar. The examiners listened to the voices through a loudspeaker. The best examiner was able to obtain 100% correct identification, when listening to sentences with a mean duration of 2.4 seconds containing about 15 phonemes. The worst examiner for the same tests obtained 92% correct responses. These percentages dropped to 56% correct for samples with duration of 0.12 seconds containing only one phoneme. The authors also ran tests based on short term memory, including 2 known subjects, A and B to be compared with one unknown subject X. The listeners were not familiar with the talkers in these tests. Average results of correct identification in these closed tests using short term memory reached the 75% level.

Pollack, Pickett and Sumby (1954) performed an experiment on aural recognition based on long term memory. All 16 talkers used in this experiment were familiar to the listeners who performed the "speaker naming tests" for groups varying from 2 to 8 talkers. Speech samples used in this experiment were tape recorded. The authors investigated the effect of three variables on the percentage of correct identification- duration of speech sample, filtering and whispering. The findings were as follows-

- a) using normal speech samples that are longer than 1 second does not significantly improve the percentage of correct identification which reached a figure close to 95% for this interval of time
- b) whispered speech reduces to approximately 30% of the percentage of correct identification as obtained with normal speech
- c) For low pass and high pass filtering, the authors concluded that “over a rather wide frequency range, identification performance is resistant to selective frequency of this type”. However, filtering above 500 Hz and below 2000 Hz decreased the percentage of correct identification.

The authors also reported that identification accuracy can be improved by increasing the length of the speech sample but that these increases will only occur for periods of up to 1200ms. Beyond this, accuracy does not seem to be related to duration, but rather to the speaker’s phonemic repertoire.

In another study, Kunzel (1995) indicated that, in German, a sample of 30 seconds is necessary to attempt any type of speaker identification, but indications on a preset similarity threshold in this study are not available. In fact, it is correct to classify the report of Kunzel (1995) as an observation rather than as a result of a controlled study.

The results of the studies in this area are equivocal. In fact, the report of Kunzel (1995) is more of an observation, than the result of a controlled study. Furthermore, as said earlier, indicating the length of the speech sample in terms of time is relative in the sense that, in a speech sample of unit time, there may be less or more number of syllables, depending on the speaking or reading rate of speakers. Expressing the minimum length

of speech in syllables is a more valid indicator. Besides, no accuracy threshold (percent of correct identification) seem to have been set in any of these studies, which is an important precondition in identification tasks as these.

The present study was undertaken on the length of speech required, both in terms of number of syllables and duration in seconds, for correct identification of speakers.

CHAPTER 3

METHOD

The chief objective of the study was to determine the minimum length of speech sample required for correct identification of speakers. The present study adopted a closed test design in which a set of listeners identified a given speech sample as that of one particular speaker belonging to a closed set.

Subjects

Ten bilingual speakers (of Kannada and English), 5 males and 5 females and in the age range of 20 to 30 years, provided speech samples. Twenty subjects, 10 males and 10 females in the same age range participated as listeners and identified the speakers. None of the subjects selected, either speakers or listeners, had any speech or hearing problems, and they were proficient in their use of Kannada and English. None of the listeners knew any of the speakers.

Material

Speech samples from the 10 speakers were collected in an interview situation. The subjects were asked to speak on matters of contemporary relevance in sports, politics and issues of national importance. Each subject gave 2 samples, one in English and one in Kannada. Thus there were 20 samples, 10 in English and 10 in Kannada. Sample of Kannada and English were collected from each of the speakers, on one particular topic. From this, a sample of speech of 150 seconds was selected in each language (Kannada and English). This sample, referred to hereinafter as Sample A, served as the familiarization material. Each of the speakers was then asked to speak on another topic

(distinct from the topic from which sample A was selected). This sample, referred to hereinafter as Sample B, served as the test material. Thus sample A contained 20 sets of speech samples (10 in Kannada and 10 in English). Sample A had 20 speech samples each of 150 seconds. Sample B had 20 sets (10 in Kannada and 10 in English) of speech samples, each of 60 seconds in duration.

Instruments used

The samples were recorded using a digital mini disc recorder (Sony MZR-30). During the identification task, the test samples (sample B) were loaded onto CSL 4500 to monitor the exact point at which a listener identifies a speaker.

Procedure

Two experiments were conducted. Experiment 1 for familiarization of speech samples in which each listener listened to 20 sets of speech in sample A and made a mental note of speaker 1, speaker 2(English or Kannada), and so on. In experiment 2, the listeners heard a speech set from sample B and identified it as belonging to speaker 1 or speaker 2 in the familiarization experiment.

Experiment 1: Familiarization

Twenty sets in speech sample A, each of 150 seconds duration, were presented through headphones to the listeners for familiarization till he/she was confident that they could participate in the identification task. The subjects were asked to internally categorize the 20 sets of samples as speech sample 1 (speaker 1), speech sample 2 (speaker 2) and so on. The subjects were asked to note down the perceptual cues relevant

to the speakers, for ease of remembrance if they liked. The 20 sets of speech samples in Sample A were presented randomly. The subjects were encouraged to listen to sample A as many times as required for correct internal categorization.

Experiment 2: Listening / Identification

Twenty sets in speech sample B were loaded onto CSL 4500 (Kay Elemetrics) not only to present speech samples to listeners, but also to get a visual feedback of the subjects' response. Each set of 60 seconds duration, in sample B, was transcribed using the International Phonetic Alphabet (IPA). The identification experiment was carried out, in general, 30 minutes after the familiarization experiment. Each set in sample B was presented to the listeners through a headphone connected to the external module of CSL 4500. All samples were presented in a random order. The subjects were instructed to listen to each set attentively and identify it as one belonging to a set in sample A. They were asked to press a key, the moment they were absolutely certain that they have identified the speaker. This left a mark on the CSL screen on the basis of which the experimenter counted the number of syllables (based on the transcription in IPA) as well as measured the duration of the speech sample (in seconds). The listeners were also asked to record their response in writing as to which set in sample A, the test set (in sample B) belonged to. The subjects were presented each test sentence only once and had been informed of this before the start of the experiment. The listeners were not informed of the correctness or otherwise of their identification. Figure 3.1 is the schematic diagram of the sequence of experiment

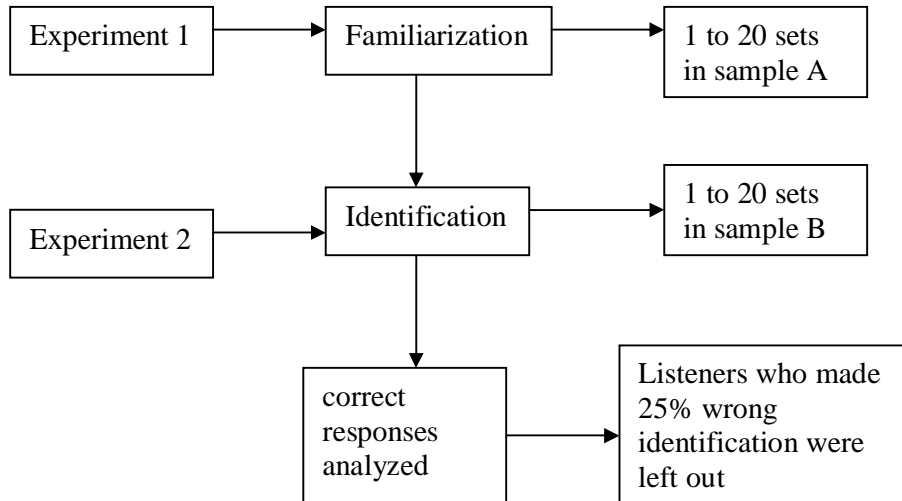


Figure 3.1: Schematic diagram of the experiments carried out

The following analyses were made from the responses given by the listeners

- a) Correctness of identification
- b) Length of sentences, both in terms of number of syllables and duration of sample in seconds, required for correct identification of speaker
- c) Gender differences of both speakers and listeners and language differences to see if they are variables influencing correct identification

CHAPTER 4

RESULTS

The objectives of this study were to determine the minimum length of utterance, defined in terms of syllables, required for correct speaker identification, from a closed set of speakers, but not familiar to the listeners. Whether or not gender (of both speakers and listeners) and language are variables to be considered in speech identification tasks was also investigated.

The subjects of this study were 10 native speakers of Kannada (5males and 5 females) and 20 listeners whose native language was also Kannada (10 males and 10 females) in the age range of 20-30 years. Each listener had to perform the identification task for 20 speech samples (10 in English and 10 in Kannada). From the responses given by the listeners, the following analyses were made:

1. Correctness of identification.
2. Length of test sentences, both in terms of syllables and duration of sample.
3. Gender differences and language differences.

Listener	Correct number of identifications	Incorrect number of identifications	Average length in syllables	SD	Average duration in seconds	SD
1	14	6	9.14	3.59	1.912	0.814
2	20	0	13.3	8.58	3.122	2.164
3	12	8	9.92	3.09	2.446	0.989
4	10	10	18.10	7.42	3.986	1.841
5	14	6	17.57	4.03	4.235	0.926
6	16	4	9.13	8.06	4.689	3.125
7	20	0	15.20	8.96	3.821	2.611
8	16	4	19.25	8.77	4.813	2.120
9	14	6	19.5	8.23	5.452	2.506
10	20	0	20.30	8.55	5.966	2.760
11	18	2	18.06	8.38	5.104	2.573
12	16	4	15.00	5.06	3.969	1.452
13	16	4	20.31	8.11	5.905	2.578
14	14	6	18.57	5072	4.917	1.529
15	16	4	19.19	7.21	5.512	1.529
16	17	3	19.53	6.19	5.305	2.198
17	16	4	21.31	6.88	5.666	2.121
18	18	2	23.67	7.82	5.788	2.498
19	15	5	27.67	6.95	7.618	2.553
20	16	4	16.69	4.88	4.572	1.855

Table 4.1: Number of correct and incorrect identifications, average number of syllables and the average duration (in seconds) required by each listener for identification. S.D: Standard Deviation.

Correctness of identification

The number of syllables required by each listener to correctly identify a speaker was noted down. The number of correct as well as incorrect identifications, mean number of syllables and mean duration of speech (in seconds) required by listeners for identifications were computed and are shown in Table 4.1. Average of minimum and maximum number of syllables, as well as minimum and maximum duration of speech (in seconds) were also computed. All those responses of listeners who made more than 25% error in identification were not considered for further analysis. Thus six listeners (1, 3, 4,

5, 9 & 14) were eliminated from the analysis. Furthermore, in tune with the objectives of the study, only the correct identification scores were considered in all analysis. Table 4.2 shows the percentage of correct identification by the 14 listeners whose scores were considered for analysis. As can be seen from Table 4.2, the listeners were able to identify speakers from an unknown set 85.71% of the time, on an average, by listening to their speech.

No. of Listeners	Total no. of identifications	No. of correct identifications	Percentage of correct identifications	Minimum percentage of correct identification	Maximum percentage of correct identifications
14	280 (14 * 20)	240	85.71	75	100

Table 4.2: Percent correct identification by 14 listeners whose responses were considered for analysis.

	Mean	95% Confidence Interval	Mean	95% Confidence Interval
Mean	18.47	15.84 - 20.09	5.132	4.490 - 5.774
Minimum	7.71	5.93 - 9.50	1.727	1.373 - 2.081
Maximum	33.93	30.96 - 36.89	10.134	9.122 - 11.147

Table 4.3: Mean and confidence intervals of the length and duration of sample required for correct identification

Table 4.3 shows that the average number of syllables required for correct identification was 18.47 and the mean duration of speech was 5.132 seconds.

Gender of speakers and listeners

Further analysis was done to find out if gender of speakers as well as listeners is a factor to be considered in identification tasks. Again only those listeners who gave 75% or more correct identifications were considered for this analysis. A “Pearson’s chi square test for independence of attributes” was performed to check for the association between the judgement (identification) and gender (of both speakers and listeners). Table 4.4 shows the relationship between results on listener gender and correct identification while Table 4.5 shows the results of Chi square test for the relationship between speaker gender and correct identification.

	JUDGEMENT		TOTAL
	YES	NO	
MALE	82	13	96
FEMALE	67	5	72
TOTAL	149	18	168

Table 4.4: Listener Gender and correct identification.

There was no association ($\chi^2 (1) = 0.096$; $p > 0.05$) between listener gender and correct identification. That is, both males and females performed statistically similar in a task on speaker identification.

	JUDGEMENT		TOTAL
	YES	NO	
MALE	25	3	28
FEMALE	124	16	140
TOTAL	149	18	168

Table 4.5: Speaker gender and correct identification

There was no association ($\chi^2 (1) = 0.067$; $p > 0.05$) between speaker gender and correct identification which means both male and female voices were equally identifiable in this task of speaker identification.

Language and correct identification

The association between language and correct identification was then analyzed using the “Pearson’s chi square test for association of attributes”. Only those listeners who gave 75% or more correct identifications were considered for this analysis.

	JUDGEMENT		TOTAL
	YES	NO	
ENGLISH	71	13	84
KANNADA	78	5	84
TOTAL	149	18	168

Table 4.6: Language and correct identification

There was no significant association ($\chi^2 (1) = 0.087$; $p > 0.05$) between language and correct identification. This implies that the language of the speaker did not influence the task of speaker identification.

A qualitative analysis of the factors which served as a cue for the listeners, as reported by the latter, revealed that listeners depended on such factors as pitch, intonation, pauses, nasality of voice, and articulation to identify the speakers.

CHAPTER 5

DISCUSSION

The results showed that on an average, listeners require 18.47 syllables to recognize speakers from an unknown set. This results in correct identification of speakers 85% of the time. In terms of the duration of speech (in seconds) required for correct identification, the study revealed that, on an average, speech sample of the duration of 5.13 seconds results in correct identification 85% of the time. Coleman (1973) reported 90% accuracy in a speaker identification task. However, he employed match-no match discrimination task between two samples.

These results on the duration of speech (in seconds) are different from those reported by Pollack et. al (1954) as well as Kunzel (1995). Pollack et. al (1954) found that a sample of speech of 1.2 seconds is adequate for identification of a speaker. These two studies and the present study are completely different methodologically, and therefore, the results should not be compared. In the present study, the subjects were asked to identify a speaker from a closed set, but which had speakers who were completely unfamiliar to the listeners. The subjects in the Pollack et. al (1954) study were asked to identify speakers speaking two sentences with normal unaltered voice. As has been said earlier, Kunzel's (1995) report should be considered more of an observation than the result of a controlled study. Bricker and Pruzansky (1966) reported that the best of listeners in their study required only 2.4 seconds of speech for correct identification. The present study also reports a similar figure (Listener 3, Table 4.1).

However, talking of the average number of syllables required for correct identification is not appropriate. The 'average' simply means that there are listeners who require more than the 'average'. Therefore, finding the range of maximum number of syllables required for correct identification would be more appropriate. In this study, the maximum number of syllables required for correct identification ranged from 30.96 to 36.89 syllables (Table 4.3). Therefore, it is more appropriate to say that 36.89 syllables are required for correct identification of the speaker. In other words, in forensic practice, if speech samples of the length of 37 syllables (rounded off) are available then speaker identification can be close to 95% accuracy.

The listeners were also encouraged to write down the factors which helped them to identify speakers from listening to their speech. The listeners listed the following factors, not necessarily in their order of importance: pitch, intonation, pauses, nasality in voice, and articulation. Holmgren (1967) reported that his subjects found pitch, intensity, quality and speech rate helped to identify the uniqueness of each speaker's speech. The factor of pitch seems to be the only common factor in its importance for speaker identification in the two studies. This observation from the present points to the need for controlling these factors, or to consider these factors, in future experiments in this area, or speech identification tasks.

The following are the other secondary results of this study:

- a) The gender of the speaker is no factor in speech identification. In other words, whether the speaker is a male or female, they are likely to be identified equally well by a set of listeners.
- b) The gender of the listener is also not important in speaker identification tasks, that is, both males and females are statistically equal in their ability to identify

speakers from listening to their speech. This result agrees with that reported by Yarmey and Matthys (1992), Clifford (1980), Hollien and Schwartz (2000), Thompson (1985) and Yarmey (1995). However, it must be realized that this seemingly similar result has come from studies with different methodologies.

- c) The language of the speaker also plays no role in the listeners' ability to identify speakers. We had speakers speaking in Kannada (native language) and English (a foreign language) to them. It appears that listeners did not pay attention to the language of the speaker, but rather paid attention to the speakers' pitch, articulation and prosody, among others. This result is in agreement with that reported by Goggin et. al (1991). Goggin et. al reported that monolingual English listeners showed better voice identification when bilinguals spoke a familiar language (English) than when they spoke an unfamiliar one (Spanish). However, this difference was not evident when the speakers and listeners were both bilinguals of English and Spanish. In other words, language is an important factor in identification for monolingual speakers whereas it is not so for bilinguals when they have to identify speakers speaking either of the two languages. The listeners in the present study were all bilinguals of Kannada and English as were the speakers. Therefore, the language of the speaker did not make any difference to the listeners in speaker identification.

The importance of these findings must be seen in from the correct perspective. These findings coming as they are from a closed set identification tasks wherein listeners and speakers were completely unfamiliar to each other must be considered internally valid. Tosi (1979) is of the opinion that in a closed set identification, errors of false identification may occur. In this study, statistical methods have revealed that such false identifications are well within statistical limits.

CHAPTER 6

SUMMARY AND CONCLUSIONS

The main objective of this study was to determine the minimum length of speech sample required for correct speaker identification from a closed, but unfamiliar set of speakers. Whether or not gender (of both speakers and listeners) and language are variables to be considered in speech identification tasks was also investigated.

The study adopted a closed test design in which a set of listeners identified a given speech sample as that of one particular speaker belonging to a closed set. Ten bilingual speakers of Kannada and English, 5 males and 5 females and in the age range of 20 to 30 years, provided speech samples. Twenty subjects, 10 males and 10 females in the same age range participated as listeners and identified the speakers. Audio recorded samples from these ten speakers in two languages (Kannada and English) served as the material for the test. The experiment was conducted in two phases: familiarization and identification. The listeners were allowed to listen to the familiarization material till they were confident that they could participate in the identification task.

Each listener had to perform the identification task for 20 speech samples (10 in English and 10 in Kannada). From the responses given by the listeners, the following analyses were made

1. Correctness of identification.
2. Length of test sentences, both in terms of syllables and duration of sample.
3. Gender differences and language differences.

The responses of listeners who made more than 25% error identification were not considered for analysis. Thus, 6 listeners were eliminated from analysis. Also, only the correct identification scores were considered for the analyses. It was found that the listeners were able to identify speakers from an unknown set 85.71% of the time, on an average, by listening to their speech. The results also showed that on an average listeners required 18.47 syllables for correct speaker identification. However, considering the average number of syllables may not be appropriate and hence the maximum number of syllables was computed. It was found that 30.96 to 36.89 syllables were required for correct speaker identification. With respect to the duration of the sample, it was found that an average of 5.132 seconds and a maximum of 9.122 to 11.147 seconds of speech sample would lead to correct identification 85.71% of the time.

Further the results of this study also revealed that gender of speakers as well as listeners do not play any role in speaker identification tasks. This means to say that both males and females perform similarly on speaker identification tasks. Also, both male and female voices are equally identifiable. The results also showed that language of the speaker does not influence speaker identification. Overall, it suggests that in speaker identification tasks, gender and language are not variables to be considered or controlled. In this study, listeners made use of the speakers' pitch, intonation, pauses, nasality in voice, and articulation as perceptual cues to identify the speakers (as reported by listeners).

This study considered normal voice for the identification task, using a closed, but unfamiliar set of speakers. Further research could be carried out

- to find the minimum length of speech sample required for correct speaker identification when the speech sample is disguised, distorted, mimicked or tapped through a telephone.
- to compare aural identification of voices and visual examination of spectrograms with respect to the length of sample to be considered for identification.
- to compare the performance of native and non- native speakers in identification tasks, that is, whether non- native speakers as a group vary in their performance compared to the native speakers, when the language is unfamiliar to them.
- to compare the performance of naïve and trained listeners in such identification tasks.

It may be recalled that identification experiment was carried out just 30 minutes after the familiarization experiment. It means that the listeners had to depend on their short term memory. This may be the reason that a high percentage of identification (85%) was recorded in this study. Therefore, experiments are warranted which would make the listener to employ long term memory for identification by giving larger intervals between familiarization and identification experiments.

REFERENCES

- Atal, B. (1972). Automatic Speaker Recognition based on pitch contour. *Journal of the Acoustical Society of America*, 52, 1687-1697.
- Bricker, P. & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America*, 40, 1441-1450.
- Bull, R., & Clifford, B.R. (1984). Ear witness voice recognition accuracy. In G.L. Wells and E. F. Loftus, (eds.), *Eyewitness testimony: Psychological perspectives*. Cambridge: Cambridge University Press.
- Clifford, B.R. (1980). Voice identification by human listeners: On Ear witness reliability. *Law of Human Behaviour*, 4, 373-394.
- Coleman, R. (1973). Speaker Identification in the absence of intersubject differences in glottal source characteristics. *Journal of the Acoustical Society of America*, 53, 1741-1743.
- DeJong, G. (1998). Ear witness characteristics and speaker identification accuracy. Doctoral thesis, University of Florida.
- Goggin, J.P., Thompson, C.P., Strube, G., & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, 19, 448-458.
- Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress and disguise speaking conditions. *Journal of Phonetics*, 10, 139-148.
- Hollien, H., & Schwartz, R. (2000). Aural-perceptual speaker identification: Problems with non-contemporary samples. *Forensic Linguistics*, 7, 199-211.
- Hollien, H. (2000). Forensic voice identification. San Diego: Academic Press.
- Holmgren, G. (1967). Physical and psychological correlates of Speaker Recognition. *Journal of Speech and Hearing Research*, 10, 57-66.

- Koster, J.P. (1981). Auditive sprecherkennug bei experten und naiven. In *Festschrift Wangler*, Helmut Buske, AG, 52, 171-180.
- Kunzel, H. (1995). Field procedures in forensic speaker recognition. In J. Lewis, (Ed.), *Festschrift for J.D. O'Connor* (pp 68- 84). London: Routledge.
- McGehee, F. (1937). The reliability of the identification of the human voice, *Journal of General Psychology*, 17, 249-271.
- McGehee, F. (1944). An experimental study in voice recognition. *Journal of General Psychology*, 31, 53-65.
- Nolan, F. (1983). *The phonetic bases of speaker recognition*. London: Cambridge University Press.
- Nolan, F. (1994). Auditory and acoustic analysis in speaker recognition. In J. Gibbons (ed.), *Language and the Law*. London: Longman.
- Nolan, F. (1997). Speaker recognition and forensic phonetics. In W. J. Hardcastle and J. Laver (eds.), *A Handbook of Phonetic Sciences*. Oxford: Blackwell.
- Pollack, I., Pickett, J.M., & Sumbly, W.H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America*, 26, 403-412.
- Reich, A. R., & Duke, J. E. (1979). Effects of selected vocal disguises upon speaker identification by listening. *Journal of the Acoustical Society of America*, 66, 1023-1028.
- Reich, A. R., Moll, K. L. & Curtis, J. F. (1976). Effects of selected vocal disguises upon spectrographic speaker identification. *Journal of the Acoustical Society of America*, 60, 919-925.

- Schweinberger, S.R., Herholz, A., & Sommer, W. (1997). Recognizing famous: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech and Hearing Research*, 40, 453-463.
- Stevens, K.N., Williams, C.E., Carbonell, J.R. & Woods, B. (1968). Speaker authentication and identification: A comparison of spectrographic and auditory presentation of speech material. *Journal of the Acoustical Society of America*, 44, 1596-1607.
- Tartter, V.C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perceptual Psychophysics*, 49, 365-372.
- Thomson, C. (1985). Voice identification: Speaker Identifiability and correction of the record regarding sex effects. *Human Learning*, 4, 19-27.
- Tosi, O & Greenwald, M. (1978). Voice identification by subjective methods of minority group voices. Paper presented at the 7th Meeting of the International Association of Voice Identification, New Orleans, LA.
- Van Dommelen. (1990). Acoustic parameters in human speaker recognition. *Language and Speech*, 33, 259-72.
- Voiers, W. (1964). Perceptual bases of speaker identity. *Journal of the Acoustical Society of America*, 36, 1065-1073.
- Yarmey, A.D. (1995). Earwitness speaker identification. *Psychol Public Policy Law*, 1, 792-816.
- Yarmey, A.D., & Mathys, E. (1992). Voice identification of an abductor. *Applied Cognitive Psychology*, 6, 367-377.