SPEAKER IDENTIFICATION IN DISGUISED SPEECH

Register No. MSHM0118

A
Dissertation submitted in
part fulfillment of the
Master's Degree (Speech and Hearing),
University of Mysore,
Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING
NAIMISHAM CAMPUS,
MANASAGANGOTHRI,
MYSORE- 570 006

MAY - 2003

Dedicated to

My

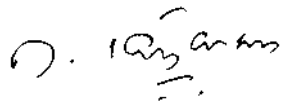
Parents,

Savi & The Nation,

Who showed me the path . . . 

C E R T I F I C A T E

This is to certify that this dissertation entitled "SPEAKER IDENTIFICATION IN DISGUISED SPEECH" is a bonafide work in part fulfillment for the degree of Master of Science (Speech and Hearing) of the student (Register No. MSHM0118).



Dr.M. Jayaram
Director
All India Institute of Speech and Hearing,
Mysore - 570 006.

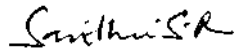
Mysore

May, 2003

C E R T I F I C A T E

This is to certify that this dissertation entitled **"SPEAKER IDENTIFICATION IN DISGUISED SPEECH"** has been prepared under my supervision and guidance. It is also certified that this dissertation has not been submitted earlier in any University for the award of any Diploma or Degree.

Guide



Dr. Savithri S.R.
Reader & HOD,
Dept. of Speech Language Sciences,
All India Institute of Speech and Hearing,
Mysore - 570 006.

Mysore

May, 2003

DECLARATION

This dissertation entitled **"SPEAKER IDENTIFICATION IN DISGUISED SPEECH"** is the result of my own study under the guidance of Dr. Savithri S.R. Reader & HOD, Dept. of Speech Language Sciences, All India Institute of Speech and Hearing, Mysore and has not been submitted earlier in any University for the award of any Diploma or Degree.

Mysore

May, 2003

Register No. MSHM0118

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

My sincere thanks to Dr. S. R. Savithri, Reader & Head Incharge, Dept. of Speech and language sciences for her patiently guiding me throughout this study.

I am thankful to Dr. M Jayaram, Director, AIISH, Mysore, for permitting me to carryout this study.

I thank Kalai & Santhosh for their help during data collection and analysis.

I thank my subjects without whose co-operation this study would not have been completed.

I am thankful to my APPA, AMMA, BROTHER and SISTER whose constant guidance and encouragement brought me to this level.

Lots of thanks to my well wisher, friend, and guide SAVI, thanks for being with me and helping me throughout.

I thank BABA, RADHA, JK & RK who understood my feelings and encouraged me to do things at time.

I thank all the GEMS & GERMS of our class. Yes, they were GEMS in interaction and they were GERMS when they infected me with their good qualities.

My thanks to Sundar, Ambeth, Kavitha, Banu, Hari & Sudhakar for indirectly helping me in this project.

I thank all the Library staffs, Mr. Shivappa (Xerox), and Mr Girish & Mrs. Shankunthala for their kind help.

Last but not least, all of my friends at AIISH, thanks a lot.

Table of Contents

LIST OF TABLES

CHAPTER I

INTRODUCTION

Speaker identification refers to any decision making process that uses the speaker-dependent features of the speech signal (Hecker, 1971). Speaker identification can be done using various methods. The classification of these methods according to Bricker & Pruzansky (1976) is as follows:

1. Speaker recognition by listening,
2. Speaker recognition by machine and
3. Speaker recognition by visual inspection of spectrograms.

In this, speaker recognition by listening involves how human listeners achieve the task of associating a particular voice with an individual or group. Speaker recognition by machines involves the matching of a given voice with the pre-recorded voice and saying whether the speaker is same or different. It is an automatic or semi-automatic strategy, standardly computer based. Therefore, this is also called as "objective" method. Speaker recognition by visual inspection of spectrograms comprises decision-making based on the identity or non-identity of voices by trained observers.

Way back in 1944, Gray & Kopp had coined the term "voice print" in a report discussing the identification of speaker by visual inspection of spectrograms and concluded that this method seemed to offer good possibilities. After the World War II got

over there was no need of voiceprints as such. In 1962, Kersta had re-examined "voice prints" and reported that spectrograms of several utterances of the same words by a given speaker always contain more similar spectral features than those produced by different speakers.

The spectrogram portrays three main parameters of speech: time (on the horizontal axis), frequency (on the vertical axis), and relative amplitude (degree of darkness of different formant frequencies). Each of the isolated phoneme, word or phrase is correlated with a characteristic spectrographic pattern.

Usually no person utters the same word twice with all characteristics being exactly the same. Laymen, for the most part, are not aware that such variations occur. Speech Scientists refer to these variations as "intra-subject variability". As yet, this is not quantified or correlated with specific acoustical parameters of the speech signal.

In contrast, differences between the same words uttered by different speakers are quite apparent to any listener; such differences are labeled as "inter-speaker variability". This variability stems mainly from anatomical differences in vocal tracts and from learned differences in the use of the speech mechanism. This is also not yet quantified or correlated with specific acoustical parameters of speech signal (Tosi, Oyer, Lashbook, Pedrey, Nicol & Nash, 1972).

In a spectrographic study, Enders, Bambach & Flosser (1971) found that in disguised speech, the individual formants were shifted to higher or lower frequencies with respect to normal voice, only first formant remaining relatively stable. Doherty and Hollien (1978) found that their identification as quoted above dropped drastically to 24 % when speakers were allowed to disguise their voices.

At the Department of Speech Language Sciences, forensic voice identifications are being done. However, it is not known as to how disguise would affect the acoustic characteristics. Hardly any literature is available on speaker identification in disguised speech. In this context, the present study was planned. The aim of the study was to determine the effects of selected vocal disguise upon the accuracy of spectrographic speaker identification. Specifically the similarity and dissimilarity of the same word uttered by inter and intra-speaker and also disguised and non-disguised speech was investigated.

CHAPTER II

REVIEW OF LITERATURE

Speaker verification 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 (Atal, 1976). Speaker verification is not an easy job. It is affected by many factors such as intra-subject variabilities and inter subject variabilities, that is within subject variabilities and between subject variabilities. Nolan (1983) proposed a model to explain these variabilities. This model explains the immense complexity of the linguistic mechanism and the human communicative ability.

As a starting point the frequently noted dichotomy between "organic" and "learned" sources of between speaker differences was taken (Garvin & Ledefoged, 1963 and Tosi, 1979). The inadequacies of this dichotomy is that the "intrinsic" component of speaker idiosyncrasy is not in the form of absolute values, but of limitations on the variation, which a speaker can induce in his vocal apparatus. Within speaker differences can also be caused by changes in intrinsic constraints, due to changes of health etc., as reported by Glenn & Kleiner (1968).

If all other sources of idiosyncrasy are grouped together under the heading of "learned", then it is evident that, at the very least, different kinds of learning are involved. The model says that, the speaker acquires by trial and error, than by learning through

direct imitation of what cannot by its nature be accessible to him, a set of implementational strategies for achieving appropriate auditory phonetic effects.

The speaker also learns, on the basis of the language use he is expected to and arguably also on the basis of innate preconceptions as to the nature of language, a complex mechanism of expression. This mechanism serves for the mapping of different aspects of the communicative intent of the speaker, and this mapping is such that many parts of the mechanism-segmental and suprasegmental, short and long-term primes and relational rules-can be affected by one aspect of communicative intent.

At each point where communicative intent is mapped there may be thought of a existing default values, which are peculiar to the speaker, though they fall within the range permitted by particular variety of the language he speaks. The point in a hyper space defined by all the speaker's default values might be thought of as constituting his extrinsic personal quality; but this point is a purely fictional obstruction, because in any utterance a speaker will be mapping communicative intent in such a way as to replace some default values by determined values. For example, a speaker may have a long term default value of non-nasalization, and a default value of /ə̈/ for /æ/, but may change these to nasalization and /æ/when communicating an attitude of irony in a social context where he is converging to a speaker with a different pronunciation.

Within speaker variability is clearly of concern to speaker recognition, but experiments based on the assumption that this variability results purely from random

intrinsic changes in time. For example, by getting subjects to read a passage several times over a few months, will not permit theoretically sound extrapolation to the real world. The way a speaker speaks on a given occasion is the result of a complex interaction between his communicative intent, the language mechanism he controls, and the context in which he is speaking. It may be that the within speaker variation that results is trivial compared with the gross acoustic similarity of utterances from the same vocal apparatus; it may be that the parameters used in " voice print" and automatic speaker identification schemes are just those which are inert to social context, attitude of the speaker, interaction management, etc., but these hypothetical states of affairs need to be demonstrated, not assumed a priori as at present, if techniques of speaker recognition are to be acceptable outside the lab. In the real world, speakers communicate rather than merely exercise their vocal apparatus.

Table 1 summarizes Sources of between-subject variabilities described by Nolan (1983) as follows:

Segmental

- Systemic differences
- Phonetic differences
- Stress
- Realization rules, allophones & co-articulation
- Realizational differences
- Long term segmental stand

Supra segmental

- Types of primes
- Phonatactic interpretation of the sequences and contrastive primes
- Differences in mean pitch, pitch range, mean loudness, speaking rate.

Interpretation rules

- Implementation rules
- Physical constraints

Mapping of communicative intent

- Phonetic representation
- Cognitive intent
- Effective Intent
- Social intent

Accentual vs personal information in speech

- Interaction management
- Self persencation

Table 1 : Summary of sources of between-speaker differences

It is impossible to cover all variations of all sorts that are apparently lacking in the literature. However, the literature review warrants study of all the variables. The objective of the present study is limited to "phonetic features variation" between and within speakers. Specifically, acoustic parameters measured from spectrography reflecting the phonetic feature differences in disguised and undisguised speech are studied here. The following section reviews the literature on disguised speech.

Reich, Moll & Curtis (1976) studied forty 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 4 sentences and 3 sentences with 9 clue words in 2 sessions respectively. The recordings were done directly in to a tape-recorder, through a telephone line in quite 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. Disguise like 70-80 years old speech, 3. Simulating severely hoarse voice, 4. Severely hypernasal voice, 5. Slow rate and 6. Freely disguised speech). The spectrograms of session 2 undisguised speech were matched with disguised and undisguised speech of session 1. Four examiners compared these clue words randomly ordered sentence pairs interms 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. In slow rate it is apparent that it has less effect on frequency of formants.

Ingmann (1968) studied the speaker's sex identification from voiceless fricatives. Five men and five women were taken and 5 series of experiments were conducted. In the first experiment, all the unvoiced fricatives in English spoken by 14 phonetically trained speakers were taken. The study concluded that /h/ was highly identifiable than the other sounds. In experiment 2 which was identical to experiment 1, one female speaker was added and the percentage of speaker identification was improved in /X/ from 61% to 72% and in /ϕ/ from 50 % to 60%. In experiment 3, 4 and 5, the subjects of experiment 1 reported that even only with the breath they could identify the sex, so each fricative was cut into 3 parts (initial, medial and decay) and presented. The identification of sex remained the same as poor for all the parts presentation. It was poor when presented as a whole. In experiment 1, / s / was the most identifiable sound but in experiment 3, 4, and 5 that was the poorly identified sound. They thought that there was some missing information in the experiment 3, 4 and 5. In experiment 6 and 7, / s / sounds were cut into 2 halves and the results indicated that the sex was identified 59% with first half and 55% with second half presentation. The scores of the better half of each / s / were added together; but this only raised the score to 66% still below the 75% of the whole.

From experiments 1 and 2, it was found that / s / & /ʃ/ had spectral peaks higher in frequency when spoken by females than by males.

The review indicated that the information on disguised speech is scanty. Therefore, the present study investigated the acoustic parameter in disguised speech.

CHAPTER III

METHOD

Subjects :

Five male Tamil-speaking subjects in the age range of 20 to 24 years participated in the study. The subjects had normal orofacial structure and function.

Material:

Consonants | k |, | b |, | g |, | n |, | K |, | ḍ |, | pp |, | ṭṭ |, | ṯ |, | m |, | l |, | ḷ |, | l̲ |, | ḏ |, | p |, | kk |, | r |, and vowels | a |, | i |, | u |, | e |, | o | were selected. Tamil meaningful words with the target phonemes in the initial, medial and final positions were selected. Using these words, three-word sentences were formed. The sentences represented either statement or command and the target word was always in the initial position. A total of 46 sentences, each written on a card, formed the material (Appendix I).

Method:

Subjects were instructed to utter the sentences with required emotion (statement and command). They also uttered the sentences in a disguised form. The disguise used was a handkerchief. All these were recorded on to the computer memory using a 12-bit A/D (Analog to Digital) converter.

Analysis:

Using the wide band spectrograms of the SSL (Voice and Speech Systems, Bangalore), F_1, F_2, F3, burst duration, closure duration, phoneme duration and voice onset time were measured. F_1, F_2, F_3 were measured as the frequency of the first, second and third formants respectively. Burst duration was measured as the time duration between the onset and offset of the burst. Closure duration was measured as the time duration between the offset of voicing for the preceding vowel and the onset of burst. Voice onset time was measured as the time duration between the onset of burst and the onset of voicing, phoneme duration was measured as the time duration between the onset and offset of a phoneme. The values for each parameter were averaged for each emotion and for original and disguised speech separately.

CHAPTER IV

RESULTS AND DISCUSSION

Burst duration in stops (voiced and unvoiced) in statement:

Table 2 shows the burst duration in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference at 0.01 level for the burst duration of disguised and undisguised speech. Burst duration in disguised speech was shorter than that in undisguised speech. A negative sign is used in the differences when burst duration in disguised speech was higher than that in undisguised speech. The difference was most evident in the stop / k / in the initial position, / pp / and / kk / in the medial position and was least evident in /t / and / d / in the initial position.

Phoneme	U	D	U-D
k-	10	15	-5
-b-	7	7	0
-g-	9	7	2
-K-	3	0	3
g-	13	9	4
d-	8	8	0
-pp-	12	7	5
-d-	5	4	1
-tt-	9	6	3
t-	7	7	0
d-	8	11	-3
p-	5	8	-3
-kk-	14	9	5
b-	7	6	1
Average	8	7	1

Table 2: Burst duration in disguised (D) and undisguised (U) speech and the differences (U-D) for stops in statement(in msec).

II. Closure duration of disguised and undisguised stops in statement.

Table 3 shows the closure duration (CD) of stops and the differences between disguised and undisguised speech. Results of t-test indicated no significant difference at 0.01 level for the closure duration in disguised and undisguised speech. CD in disguised speech was always longer compared to that in undisguised speech. The difference was more observed in / tt / and least in /kk/.

Phoneme	**U**	**D**	**U-D**
-tt-	81	90	-9
-kk-	93	96	-3
Average	87	93	6

Table 3: Closure duration in disguised (D) and undisguised (U) and the differences (U- D) for stops in statement (in msec).

III. Voice onset time (VOT) in disguised and undisguised stops (voiced and unvoiced) in statement:

Table 4 shows the VOT in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between the VOT of disguised and undisguised speech. VOT in disguised speech was longer than that in the undisguised speech for voiced stops and it was equal for unvoiced

stops. The difference was most evident in the stop / t /, / g / and least evident in / p/, / k / & / b / .

Phoneme	U	D	U-D
Unvoiced	17	21	- 4
k-			
t-	19	12	7
P-	10	14	-4
Average	15	15	0
Voiced	-62	-51	11
g-			
d-	-52	-61	-9
b-	-43	-49	-6
Average	-52	-54	-2

Table 4: VOT in disguised (D) and undisguised (U) and the differences (U-D) for stops in statement (in msec).

IV. Phoneme duration in disguised and undisguised speech for stops in statement:

Table 5 shows the phoneme duration (PD) in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between the phoneme duration of disguised and undisguised speech. The phoneme duration in disguised speech was longer than that in undisguised speech. The difference was most evident in / k / and / g / in the initial and medial position respectively and least evident in / kk / in the medial position and / b / in the initial position.

Phoneme	U	D	U-D
k-	108	126	-18
-b-	142	149	-7
-g-	40	22	18
-K-	20	23	-3
g-	75	60	15
d-	70	81	-11
-PP-	75	62	13
-d-	20	27	-7
-tt-	22	32	-10
t-	26	19	7
d-	60	72	-12
P-	15	22	-7
-kk-	30	25	5
b-	50	55	5
Average	54	55	1

Table 5: Phoneme duration in disguised (D) and undisguised (U) speech and the differences (U-D) for stops in statement (in msec).

V. Fl of consonants in disguised and undisguised speech for statement:

Table 6 shows the Fl of consonants in disguised and undisguised speech and their differences. Results of t-test indicated no significant differences between Fl in disguised and undisguised speech at 0.01 level. In disguised speech Fl was lower than in undisguised speech. The difference between the Fl of disguised and undisguised speech was most evident in /m/ in the medial position and was least evident in /n/ in the initial position.

Phoneme	U	D	U-D
-m-	799	601	199
-r-	659	644	144
-m	669	672	131
-l	675	575	121
-l-	818	696	99
-n-	706	606	80
-l	754	675	79
-n	734	687	66
-n-	614	665	-51
-l-	671	722	-50
-l-	721	700	22
m-	700	715	15
r-	719	680	10
l-	690	696	-6
n-	703	686	2
Average	709	668	41

Table 6: Fl in disguised (D) and undisguised (U) speech and the differences (U-D) for consonants in statement (in Hz).

VI. Fl in disguised and undisguised speech for vowels in statement:

Table 7 shows the Fl in disguised and undisguised speech and their differences. Results of t-test indicated no significant differences between disguised and undisguised speech for Fl of vowels at 0.01 level. It was observed that Fl in disguised speech was higher than in undisguised speech. The difference in Fl was more in / a / in the final position and least in / e / in the initial position.

Phoneme	U	D	U-D
-a	765	625	140
-i	665	656	-109
0-	553	621	-69
-u	649	689	-39
u-	651	687	-36
-i	665	656	9
a-	715	707	8
-0	615	618	-3
e-	599	600	0
Average	649	660	11

Table 7: F 1 of vowels in disguised (D) and undisguised (U) speech and the differences (U-D) in statement (in Hz).

VII. F2 in disguised and undisguised speech for consonants in statement:

Table 8 shows the F2 in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between F2 in disguised and undisguised speech at 0.01 level. On observation F2 of disguised speech was lower than that of undisguised speech. The difference was most evident in / m / in the medial position and least evident in / l / in the final position.

Phoneme	U	D	U-D
-m-	1533	1234	300
-m	1219	1078	282
-n-	1262	1215	254
-l-	1515	1325	190
m-	1174	1274	-100
-m	1219	1078	84
n-	1400	1362	66
r-	1421	1359	62
-n-	1447	1193	47
-l-	1649	1595	34
-l-	1783	1753	29
-r-	1302	1331	-28
-l	1279	1243	24
l-	1490	1481	9
-l	1790	1793	-3
Average	1438	1366	72

Table 8: F2 in disguised (D) and undisguised (U) speech and the differences (U-D) for consonants in statement (in Hz).

VIII. F2 in disguised and undisguised speech for vowels in statement:

Table 9 shows the F2 in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between F2 of vowels in disguised and undisguised speech at 0.01 level. In general F2 in disguised vowels was higher than that in undisguised vowels. On observation the difference was most evident in the vowel / u / in the initial position and least evident in / i / in the final position.

Phoneme	U	D	U-D
u-	1200	1609	359
-0	1152	1510	-354
-u	1137	1350	-212
0-	1172	1306	-134
-a	1215	1106	110
e-	1756	1687	69
a-	1087	1125	-38
i-	1496	1514	22
-i	1667	1668	-1
Average	1320	1431	-111

Table 9: F2 in disguised (D) and undisguised (U) speech and the differences (U- D) for vowels in statement (in Hz).

IX. F3 in disguised and undisguised speech for consonants in statement:

Table 10 shows the F3 of consonants in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between F3 in disguised and undisguised speech at 0.01 level. In general, F3 in disguised speech was higher than that in undisguised speech. The difference observed was highest for / n / in the initial position and was lowest for / 1 / in the medial position.

Phoneme	U	D	U-D
n-	2212	2267	455
-n-	2550	2290	260
m-	2090	2337	-259
-m	2159	2359	203
r-	2252	2096	156
-n-	2247	2400	-153
-l-	2447	2571	-125
-r-	2170	2281	-111
-m-	2318	2246	72
-l-	2290	2353	-63
-n	2563	2765	60
-l	2362	2312	50
-l-	2340	2381	-41
-l	2278	2245	32
-l-	2575	2578	-3
Average	2324	2362	38

Table 10: F3 in disguised (D) and undisguised (U) speech and the differences (U-D) for consonants in statement (in Hz).

X. F3 in disguised and undisguised speech for vowels in statement:

Table 11 shows the F3 of vowels in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between disguised and undisguised speech for F3. In general F3 of disguised speech was higher than the undisguised speech. The difference was most evident in / o / in the final position and least evident in / i / in the final position.

Phoneme	U	D	U-D
-0	2024	2274	-250
-u	2171	2345	-173
u-	2030	2222	-154
0-	2364	2268	96
a-	2324	2241	82
i-	2243	2196	47
-a	2387	2371	21
e-	2200	2384	16
-i	2242	2237	5
Average	2221	2282	61

Table 11: F3 in disguised (D) and undisguised (U) speech and the differences (U-D) for vowels in statement (in Hz).

XI. Burst duration of stops (voiced and unvoiced) in commands:

Table 12 shows the burst duration in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between burst duration in disguised and undisguised speech at 0.01 level. The burst duration in disguised speech was shorter than that in undisguised speech. The difference was most evident in the stop / g / in the medial position and least evident in / g / in the initial position and / p / in the initial position.

Phoneme	U	D	U-D
k-	10	11	-1
-b-	7	6	1
-g-	5	2	3
-K-	4	2	2
g-	8	8	0
d-	5	6	-1
-pp-	8	7	1
-d-	3	2	1
-tt-	6	5	1
t-	6	5	1
d-	6	7	-1
p-	6	6	0
-kk-	11	9	2
b-	4	5	-1
Average	6	5	3

Table 12: Burst duration in disguised (D), and undisguised (U) speech and the differences (U-D) for stops in commands (in msec).

XII. Closure duration in disguised and undisguised stops for commands:

Table 13 shows the closure duration (CD) of stops and their differences in disguised and undisguised speech. Results of t-test indicated no significant difference between closure duration of disguised and undisguised speech at 0.01 level. CD in disguised speech was shorter compared to that in undisguised speech. The difference was most evident in / tt / in the medial position compared to / kk / in the medial position.

Phoneme	U	D	U-D
-tt-	51	43	8
-kk-	81	78	3
Average	66	61	6

Table 13: Closure duration in disguised (D) and undisguised (U) speech and the differences (U-D) for stops in commands (in msec).

XIII. Voice onset time (VOT) in disguised and undisguised stops (voiced and unvoiced) for commands:

Table 14 shows the VOT in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between the disguised and undisguised speech at 0.01 level. VOT in disguised speech was shorter compared to undisguised speech in voiced stops and longer in unvoiced stops. The difference was most evidently observed in / t̬ /, / g / and least evident in / p /, / d̬ / and / b /.

Phoneme	U	D	U-D
Unvoiced			
k-	19	23	4
t-	10	24	14
p-	14	16	2
Average	14	21	7
Voiced			
g-	-106	-94	-8
d	-82	-79	-3
b-	-81	-84	-3
Average	90	86	4

Table 14: VOT in disguised (D) and undisguised (U) and the differences (U-D) for stops in command (in msec).

XIV. Phoneme duration in disguised and undisguised speech for stops in commands:

Table 15 shows the phoneme duration (PD) in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between the phoneme duration of disguised and undisguised speech at 0.01 level. Phoneme duration in disguised was longer than in undisguised speech. The difference was most evident in /ḍ/ in the medial position and least evident in /ṭ/ in the initial position.

Phoneme	U	D	U-D
k-	80	77	3
-b-	126	156	-30
-g-	59	48	9
-K-	28	40	-12
g-	114	107	7
d-	40	91	-51
-PP-	118	101	17
d-	38	35	3
-tt-	22	14	8
t-	20	21	-1
d-	87	91	-4
P-	16	30	-14
-kk-	34	30	4
b-	86	84	2
Average	62	66	4

Table 15: Phoneme duration in disguised (D), and undisguised (U) speech and the differences (U-D) for stops in commands (in msec).

XV. Fl of consonants in disguised and undisguised speech for commands:

Table 16 shows the Fl of consonants in disguised and undisguised speech and their differences. Results of t-test indicated no significant differences between Fl in disguised and undisguised speech at 0.01 level. In general Fl in consonants in disguised speech was lower than that in undisguised speech. The difference was most evident in /l/ in the medial position and least evident in / n / in the medial position.

Phoneme	U	D	U-D
-l-	659	768	-109
-l	715	609	107
-n-	734	631	103
-m	565	484	81
m-	746	681	66
-n	695	633	62
-m-	640	693	-53
-l-	693	737	-44
l-	696	653	43
-l-	715	678	37
-r-	637	609	28
r-	675	696	-22
-l	971	690	-20
n-	756	737	19
-n-	678	681	-3
Average	676	659	17

Table 16: Fl in disguised (D) and undisguised (U) speech and the differences (U-D) for consonants in commands (in Hz).

XVI. Fl in disguised and undisguised speech for vowels in commands:

Table 17 shows the Fl in disguised and undisguised speech and their differences. Results of t-test indicated a significant difference between disguised and undisguised speech for Fl at 0.01 level. In general Fl in disguised speech was higher than that in undisguised speech. The difference was most evident in / i / in the final position and least evident in / a / in the final position.

Phoneme	U	D	U-D
-i	515	615	-100
0-	575	640	-66
a-	640	706	-66
i-	600	665	-66
u-	681	743	-63
-u	631	675	-45
e-	609	643	-35
-o	656	640	16
-a	671	684	-13
Average	620	667	47

Table 17: Fl in disguised (D) and undisguised (U) speech and the differences (U-D) for vowels in commands (in Hz).

XVII. F2 in disguised and undisguised speech for consonants in commands:

Table 18 shows the F2 in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between disguised and undisguised speech for F2 at 0.01 level. In general, F2 in consonants in

disguised speech was lower than that in undisguised speech. The difference observed was more in / l / in the final positions and least evident in / l / in the medial position.

Phoneme	U	D	U-D
-l	1537	1772	-235
n-	1615	1428	187
r-	1321	1168	153
-n-	1446	1296	150
-l	1725	1578	148
-m-	1425	1284	141
m-	1245	1118	128
-l-	1206	1315	-109
-r-	1455	1350	85
-m	1167	1097	70
-n	1094	1156	-62
-l-	1409	1465	-56
-n-	1392	1418	-26
l-	1462	1472	-10
-l-	1701	1693	8
Average	1413	1374	39

Table 18: F2 in disguised (D) and undisguised (U) speech and the differences (U-D) for consonants in commands (in Hz).

XVIII. F2 in disguised and undisguised speech for vowels in commands:

Table 19 shows the F2 in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between F2 of

disguised and undisguised speech at 0.01 level. In general F2 of vowels in disguised speech was higher than that in undisguised vowels. On observation the difference was most evident in the vowel / i / in the final position and least evident in / a / in the final position.

Phoneme	U	D	U-D
-i	1353	1731	-378
-0	1384	1565	-182
i-	1458	1565	-107
0-	1198	1103	95
-u	1378	1285	92
u-	1281	1307	-26
a-	1112	1193	-8
e-	1740	1748	-8
-a	1146	1150	-4
Average	1339	1405	66

Table 19: F2 in vowels in disguised (D) and undisguised (U) speech and the differences (U-D) in commands (in Hz).

XIX. F3 in disguised and undisguised speech for consonants in commands:

Table 20 shows F3 in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between disguised and undisguised speech for F3 at 0.01 level. The F3 of consonants was higher in disguised speech than in undisguised speech. The difference was most evident in / l / in the final position and was least in / n / in the medial position.

Phoneme	U	D	U-D
-l	2277	2003	274
-r-	2350	2328	218
m-	2153	2337	-184
-l-	2346	2518	-172
-n	1810	1915	-105
l-	2387	2481	-94
-l-	2512	2424	88
-l	2177	2259	-82
-m	1852	1771	81
-n-	2384	2461	-77
-l-	2305	2347	-42
r-	2202	2237	-35
-m-	2156	2190	-34
n-	2428	2397	31
-n-	2346	2341	5
Average	2246	2303	57

Table 20: F3 in disguised (D) and undisguised (U) speech and the differences (U-D) for consonants in commands (in Hz).

XX. F3 in disguised and undisguised speech for vowels in commands:

Table 21 shows the F3 in disguised and undisguised speech and their differences. Results of t-test indicated no significant difference between disguised and undisguised speech for F3 at 0.01 level. The F3 of vowels in disguised speech was higher than that in undisguised speech. The difference was highest in / u / in the final position and lowest in the / o / in the final position.

Phoneme	U	D	U-D
-u	2121	2337	-216
e-	2191	2373	-182
u-	2228	2359	-131
-i	2198	2306	-108
i-	2100	2206	-106
-a	2249	2316	-69
0-	2165	2128	37
a-	2243	2266	-23
-o	2221	2221	0
Average	2191	2279	88

Table 21: F3 in disguised (D) and undisguised (U) speech and the differences (U-D) for vowels in commands (in Hz).

To summarize:

In disguised speech (statement) burst duration and VOT were shorter, closure and phoneme duration were longer compared to undisguised speech. In disguised speech of commands all temporal parameters except phoneme duration were shorter compared to undisguised speech.

In statement, Fl and F2 were lower in consonants and higher in vowels and F3 was lower in both consonants and vowels in disguised speech when compared to undisguised speech. In command, Fl and F2 were lower in consonants and higher in vowels and F3 was higher in both consonants and vowels in disguised speech compared to undisguised speech. However, no

CHAPTER V

SUMMARY AND CONCLUSION

Speaker verification is the process in which speaker dependent cues / features are used and verified to know whether the given speech sample is by the same speaker or not. There are many ways in which speakers can be verified. Speaker identification by visual inspection of spectrograms is one of them. There are many factors within and between speakers, which influence the speaker identification (Nolan, 1983). Disguising is a problem in speaker identification. Very few studies have been conducted on speaker identification in disguised speech. Therefore, the present study was designed to investigate the effect of vocal disguise upon the accuracy of spectrographic speaker identification, and the sounds that can be used in the spectrographic speaker identification in disguise.

The material consisted of phonemes (stops, nasals, trills, laterals and vowels) in the initial, medial and final position of meaningful Tamil words. The words with the target phoneme were placed in the initial position of the 3 -word sentences (both in statement and command). A total of 48 sentences formed the material. Five adults' males who had "Tamil" as native language served as subjects. The subjects speaking of the sentences in statement / command was recorded on to the computer using a 12 bit *AID* converter. Both disguised and undisguised speech was recorded digitally and analyzed for the parameters, burst duration, closure duration, voice onset time, phoneme

significant difference between disguised and undisguised speech was noticed. The results indicated that the consonants / n / and vowel / e / in the initial position, / l / in the medial position, / l / and / i /in the final position in statements, / n / and / n / in the medial position, / a / and / o / in the final position could be used in the speaker verification under disguised speech as they showed least difference in the parameters studied. Further research on speaker verification in disguised speech is warranted.

duration for stops and F1, F2 & F3 of nasals, trills, laterals and vowels separately for statements and commands and compared. The disguise used was a handkerchief. The results indicated no significance difference between the disguised and undisguised speech for both statements and commands.

However, some of the speech sounds did not show great variations in disguised and undisguised speech, which could be used for speaker verification under disguise (handkerchief). These phonemes included / e /, and / n / in the initial position, / l /, / 1 / and / n / in the medial position and / 1 /, / i /,/ a / and / o / in the final position. These results are restricted to Tamil language and further research on speaker verification under disguise is recommended.

B I B L I O G R A P H Y

Atal, B.S. (1976). Automatic recognition of speakers from their voices. Precedings of the IEEE, 64 (4), 460-475.

Bricker, P.D., & Pruzansky, (1966). Effects of stimulus content and duration on talker identification. J.Acoust.Soc.Amer., 40,1441-1449.

Doherty,E.T.& Hollien,H.(1978). Multiple-factor speaker identification of normal and distorted speech. Journal of Phonetics, 6, 1-8.

Endres, W., Bambach, W., & Flosser, G (1971). Voice spectrograms as a function of age, voice disguise and voice imitation, J.Acoust.Soc.Amer., 49,1842-1848.

Garvin, P.L.& Ladefoged (1963). Speaker identification and message identification in Speech recognition, Phonetica, 9,193-196.

Glenn, J.W.& Kleiner, N. (1968). Speaker identification based on nasal phonation, J.Acoust.Soc.Amer., 43, 368-372.

Gray, C. & Kopp, G. (1944). Voice print identification, Bell Telephone Lab., Inc., 1-14

Hecker, H.L (1971). Speaker recognition: Basic considerations and methodology. J.Acoust.Soc.Amer., 49 (38A).

Ingmann, F. (1968)- Identification of sex from voiceless fricatives, J.Acoust.Soc.Amer., 44 (4), 1142-1144.

Kersta, L.G. (1962). Voiceprint identification. Nature 196,1253-1257.

Nolan, F. (1983). The phonetic bases of speaker recognition. London: Cambridge University Press.

Reich, A.R., Moll, K.L.&Curtis, J.F. (1976). Effects of selected vocal disguises upon spectrographic speaker identification, J.Acoust.Soc.Amer., 60 (4), 919-925.

Tosi, O.I (1979). Voice identification: Theory and legal applications. Baltimore University Park Press.

Tosi, O.I., Oyer, H. Lashbrook, Pedrey, C, Nicol, L., & Nash, E. (1972). Experiment on voice identification, J.Acoust.Soc.Amer., 51,2030-43.

APPENDIX - I

(Material used for the study)

1. kabaḍi oru viḷaya:ṭṭu

   kabaḍi maṭṭum viḷaya:ḍu

2. nagam viraḷil irukku

   nagaṭa ippave veṭṭu

3. mo:Kan vi:ṭṭil irukira:n

   mo:Kan inge va.

4. ga:nam nalla iruku

   ga:nam ippo pa:ḍa:ḍe

5. ḍappa enkiṭṭa iruku

   ḍappa:va ki:ḷe po:ḍu

6. aḍuppu vi:ṭṭula iruku

   aḍuppa ipo ṭoḍa:ḍe

7. Oṭṭagam romba ujarama:naḍu

   Oṭṭagam kiṭṭa po:ga:ḍe

8. tamiḷ inimaija:na moḷi

   ṭamiḷ ippave kaṭuko

9. ḍaṇam iṅge irukira:!

ḍaṇam ṭirumbi po:

10. paṇam peṭṭijil iruku

paṇam enake kuḍu

11. makka! iṅgu varuvaiṅga

makka! ja:rum varaku:ḍa:ḍu

12. baṅga!a a!aga irukku

baṅga!a u!!a po:ga:ḍa

13. maram namakku avasijam

maraṭṭa ippo veṭṭa:ḍe

14. ra:man oru nallavan

ra:man vi:ṭṭuku po:

15. vanadza nalla poṇṇu

vanadza iṅge va:

16. ka!ivu ni:r keṭṭaḍu

ka!ivu ni:ra miḍika:ḍe

17. ta:lam ello:rukkum terijum

ta:lam enakku po:da:de

18. laddu mandzal nirama:nadu

laddu enakku kudu

19. pa:lam ra:dza kattijadu

pa:latu me:la nadakka:da

20. amma romba nallaval

amma inge vara:de

21. indzi udambukku nalladu

indzi adula po:da:de

22. uppu sa:pa:ttukku te:va

uppu enakku tara:de

23. ennej udambukku nalladu

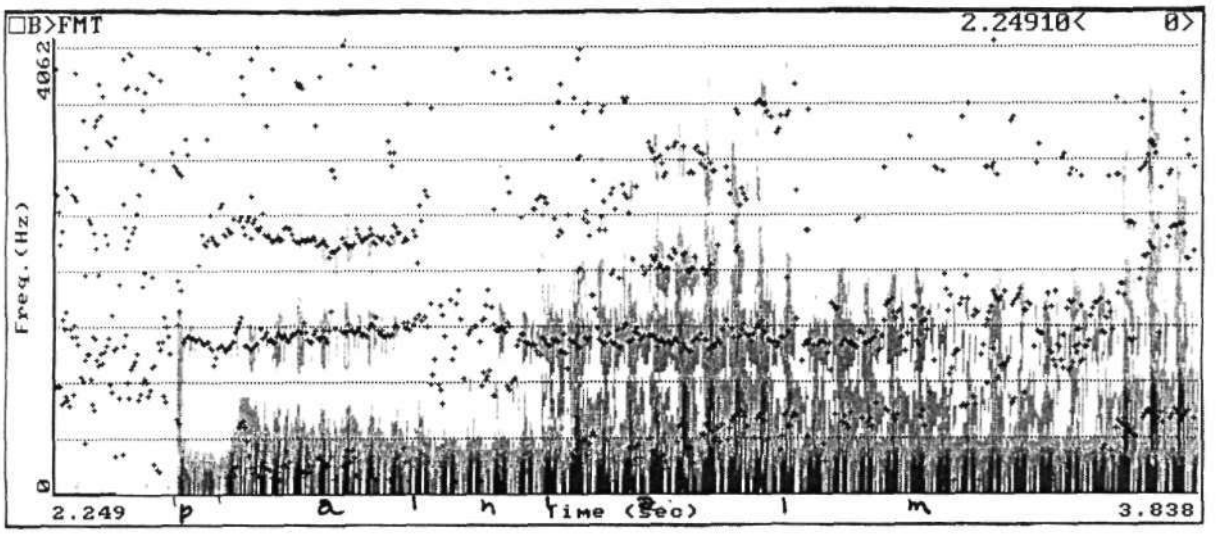
ennja ni: sa:pida:de

24. po:tto enakku edukkanum

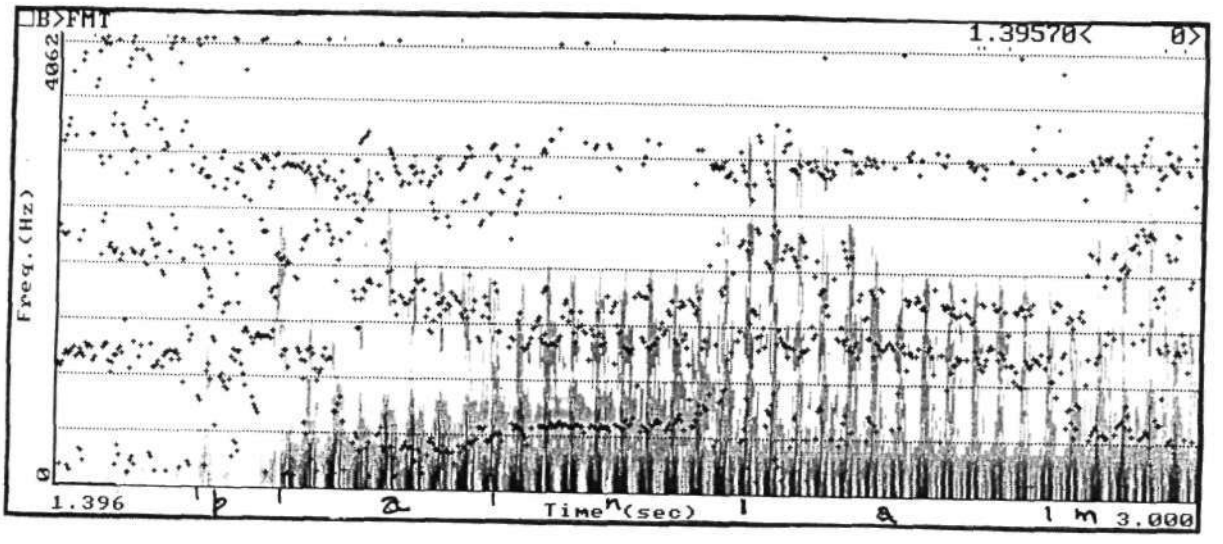
po:tto enna edukka:de

In all sentence pairs first sentence is the statement and second is the command.

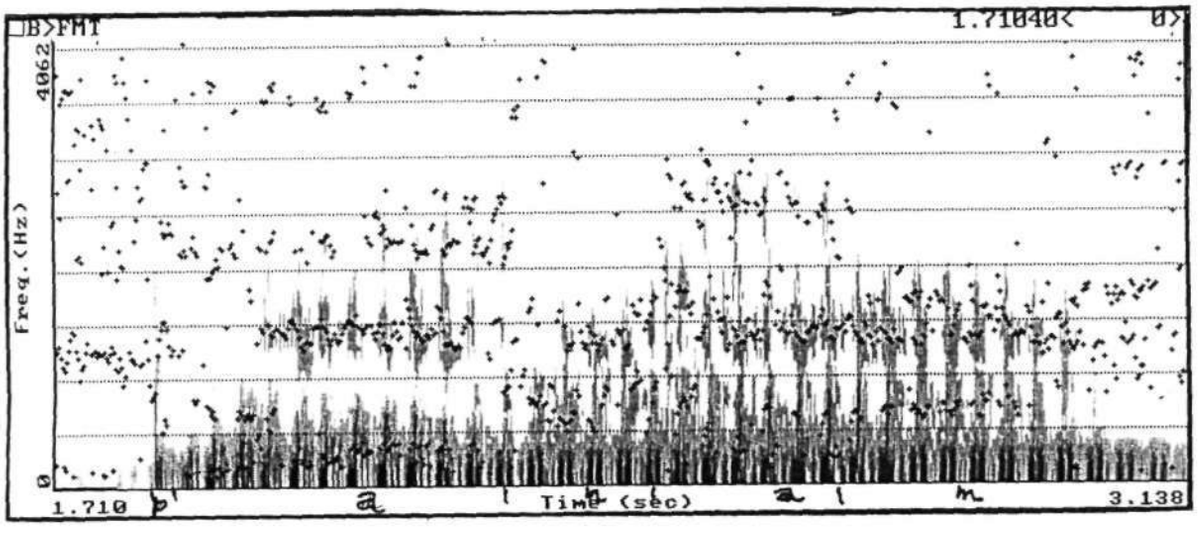
Appendix II
Spectrograms in disguised and undisguised speech



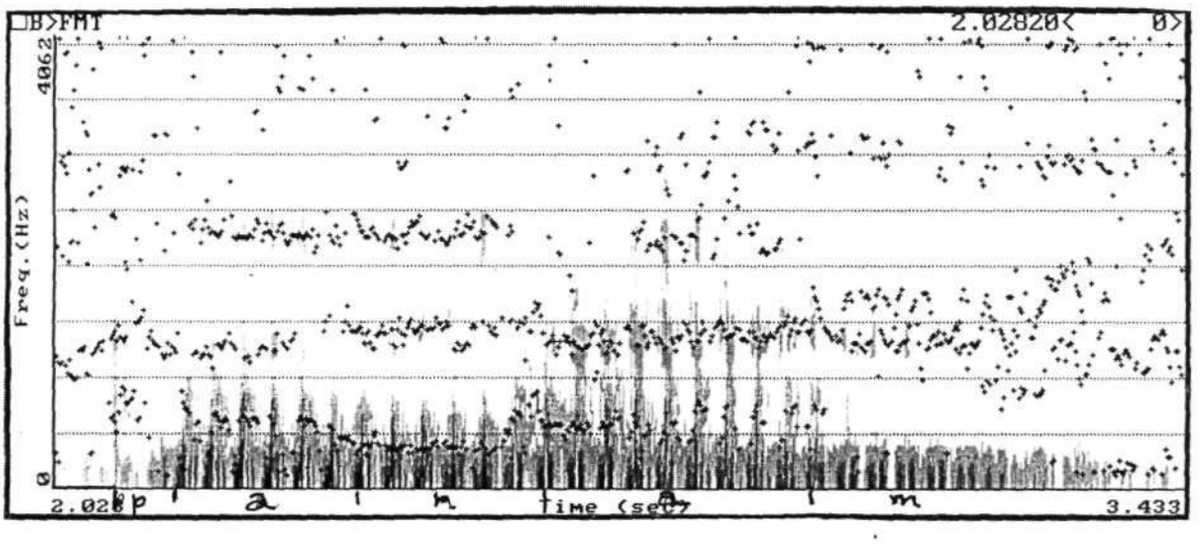
Spectrogram of the word /pan am/ in undisguised speech (command).



Spectrogram of the word /pan.am/ in disguised speech (command).



Spectrogram of the word /pan am/ in undisguised speech (Statement).



Spectrogram of the word /pan am/ in disguised speech (Statement).