ACOUSTIC FEATURES FOR SPEECH SYNTHESIS

NEHA RAWAT

Register No. M2K14

A Dissertation submitted in part fulfillment for the
final year M.Sc, (Speech and Hearing)
University of Mysore, Mysore

All India Institute of Speech and Hearing
Manasagangothri
Mysore - 570 006

MAY-2002

Dedicated to....

Dr. Savithri,

Vibha, Rekha, Shivam,

Vivek and Cthe

CERTIFICATE

This is to certify that the dissertation entitled "ACOUSTIC FEATURES FOR SPEECH SYNTHESIS" is the bonafide work in part fulfillment for the degree of Master of Science (Speech and Hearing) of the student with Register No. M2K14.

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

Mysore
May 2002

CERTIFICATE

This is to certify that the dissertation entitled "ACOUSTIC FEATURES FOR SPEECH SYNTHESIS" 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.

Dr. S.R. Savithri,
GUIDE
Reader and Head I/C
Department of Speech- Language Sciences
All India Institute of
Speech and Hearing
Mysore - 570 006

Mysore
May 2002

DECLARATION

I hereby declare that this dissertation entitled "ACOUSTIC FEATURES FOR SPEECH SYNTHESIS" is the result of my own study under the guidance of **Dr.** S. **R. Savithri,** Reader and Head I/C Department of Speech - Language Sciences, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier at any other University for the award of any Diploma or Degree.

Mysore
May 2002 **Reg. No. M2K14**

ACKNOWLEDGEMENT

I express my sincere thanks and gratitude to my guide, Dr. S.R. Savithri, Reader and Head I/C, Department of Speech-Language Sciences, All India Institute of Speech and Hearing, Mysore, for her constant support and guidance.

I would like to thank Dr. M. Jayaram, Director, All India Institute of Speech and Hearing, Mysore, for allowing me to carry out this study.

I'm indebted to the staff members of the Department of Speech Sciences especially Sridevi Madam and Santhosh for all their help during analysis.

I thank all my subjects for their cooperation.

Maa, Paa, Rekha, Vibha, Shivam and Star….. you make the best family on this earth. Thanks for being their with me through out. Love you lots.

Dear Seetha, you are my best friend forever. I am glad …. You found a friend in me.

Dearest Scratchy girls ……. time flew so fast with you girls around …… I can give anything to bring back the days with you. You will always remain in my heart.

Dear, Prasi and Mathew…….. thanks for lending your helping hand whenever I was in need ……. I will never forget you guys. Wish all the best…….. with love and prayers.

Vivek…. you have been very supportive throughout. I acknowledge your contribution with love and gratitude.

I am appreciative of the contribution of Pamo and Ananthi who did not let me feel lonely while doing my analysis ..... Thanks a lot.

Dear Mukunthan for being such a good postingmate. Thanks for your support.

Sai Ram, you proved to be a very good brother, your thoughtful comments are always appreciated.

Dwivedi uncle, aunt and Tanu..... staying in Mysore for these two years is more memorable just because of you people .... Thanks for all your love, appreciation and support.

All the first years especially Sheethal, Himani, Suma, Deepashri, Rehana, Shalani, Anitha, Mallika, Anjali, Alfina, Shivani, Priya, Bela, Pintu, Somesh, Nambi and Vinay..... thanks a lot for your contribution. All the very best in your future endeavours.

Dear Rohini, Bhuvana, Manika, Meenakshi, Deema, Amy, Divya, Devi, Patiala and Aditi.... you girls are just too good ..... I appreciate you girls for your direct and indirect contribution in finishing this dissertation.

Poonam, Shivani (sisten-in-law), Amita, Shweta, Shereen, Sandeep..... I acknowledge you for making my stay at AIISH a memorable one..... thanks for all your love and support.

All my classmates for their love, encouragement and support.

I acknowledge Ms Rathna for good typing. Rathna, without your diligent efforts this piece of work would not have been possible.

Thanks to Mr. Shivappa for neat xeroxing and binding.

TABLE OF CONTENTS

	Page No.
INTRODUCTION	1-2
REVIEW OF LITERATURE	3-16
METHODOLOGY	17-20
RESULTS AND DISCUSSION	21-39
SUMMARY AND CONCLUSION	40-41
BIBLIOGRAPHY	

LIST OF TABLES

Table No.	Title	Page No.
1.	Word List	18
2.	The segmental features (spectral and temporal) for male subject	23-27
3.	The segmental features (spectral and temporal) for female subject	28-32
4.	Range of values for the plosives for different temporal measures	22
5.	Range of values of the temporal measures for the affricates	22
6.	Range of values of temporal measures for the fricatives	33
7.	Range of values of temporal measures for the nasal continuants	33
8.	Range of values of temporal measures for trill, lateral and semivowel	33
9.	Average values of the F_1, F_2 and F_3 and duration for the vowels of the male subject	34

10.	Average values of the F_1, F_2 and F_3 and duration for the vowels of the female subject	34
11.	Average values for voice onset time total duration, burst duration, aspiration duration, frication duration, and terminal frequencies for the phonemes spoken by the male subject	35-36
12.	Average values for voice onset time total duration, burst duration, aspiration duration, frication duration, and terminal frequencies for the phonemes spoken by the female subject	37-38

LIST OF FIGURES

Figure No.	Title	Page No.
1.	Block diagram illustrating creation of database	4
2.	System for acoustic feature extraction	7
3.	Spectograph showing terminal frequency transition duration, study state duration of the formant, burst duration, vowel duration and voice onset time	19
4.	Spectograph showing frication noise duration, aspiration duration and closure duration	20

CHAPTER-I

INTRODUCTION

In recent times fifth generation computers are being extensively used to enhance speech communication. It includes automatic speech recognition and speech synthesis for which a database is needed.

Database is a collection of data used for a specific purpose. The data base can be phonetic, acoustic phonetic, prosodic or all of them. It can be for a limited vocabulary or unlimited vocabulary. Database provides information on normal segmental and suprasegmental features and thus helps in better understanding of the speech production system. It helps in developing synthesizers meant for text to speech synthesis and it aids in automatic speech recognition. World wide there have been many databases which have been developed in various languages for the purposes mentioned above. Some of them are specifically made for telephone use. The databases - developed in European languages are EUROM - a multilingual European speech database and IRIS - speech database etc.

Databases which are developed in English languages are many. For example Oxford acoustic phonetic database, ATIS - 2, MADCOW Speech Corpus etc.

For telephone use also these databases are used.Few of these kinds of databases are Australian English Telephone speech database, King speech

database, Annion database etc. Database for the purpose of speaker identification are also available like King Speech Data base.

Each language has its segmental and suprasegmental characteristics based on which database is made available which is then used for text to speech conversion in that language. The database of one language cannot be used for another language as the segmental and suprasegmental features vary across languages. For example, English is based on phonemic unit, but few of the Indo-European and Dravidian languages are based on syllabic unit. Hence database available in English language cannot be used for these languages.

In Indian context, accountable number of databases are available like in Kannada (Savithri, 1994) and Hindi (Zero to nine by Agrawal, 1992). But most of the time only limited units of speech have been concentrated for the data. In this context, the present study was planned. The aim of the study was to develop dynamic acoustic segmental database in Hindi language. Specifically the database will be made by acoustically analyzing one hundred and five meaningful Hindi words. Two spectral and nine temporal parameters will be extracted using the VSS software.

CHAPTER II

REVIEW OF LITERATURE

A data base is a collection of data used for a specific purpose. The data base can be a) phonetic b) acoustic phonetic c)prosodic or all of them. The database can be speaker dependent or speaker independent. The speech recognition systems developed from database can be for limited vocabulary / small vocabulary or large vocabulary.

A data base follows the segmental and supra segmental features of the language and hence a data base of one language cannot always be used for another language. To develop detailed phonetic knowledge, speech data base are accessed according to linguistically meaningful principles. Various methods have been developed like the one used by Carlson and Granstrom (1985) who focussed on generative rules which formed the integral part of the data base management.

The first step to create the data base is to record and label speech. Speech data is stored in sentence sized files. Carlson, Granstrom and Hunnicutt,1982 used text to speech system to create a phonetic transcription of the utterances. This transcription is edited to match the pronunciation as good as possible. A phoneme can be transcribed by one or two character Stress level is marked by special signs. Additional markers like syntactic boundaries can be added if needed.

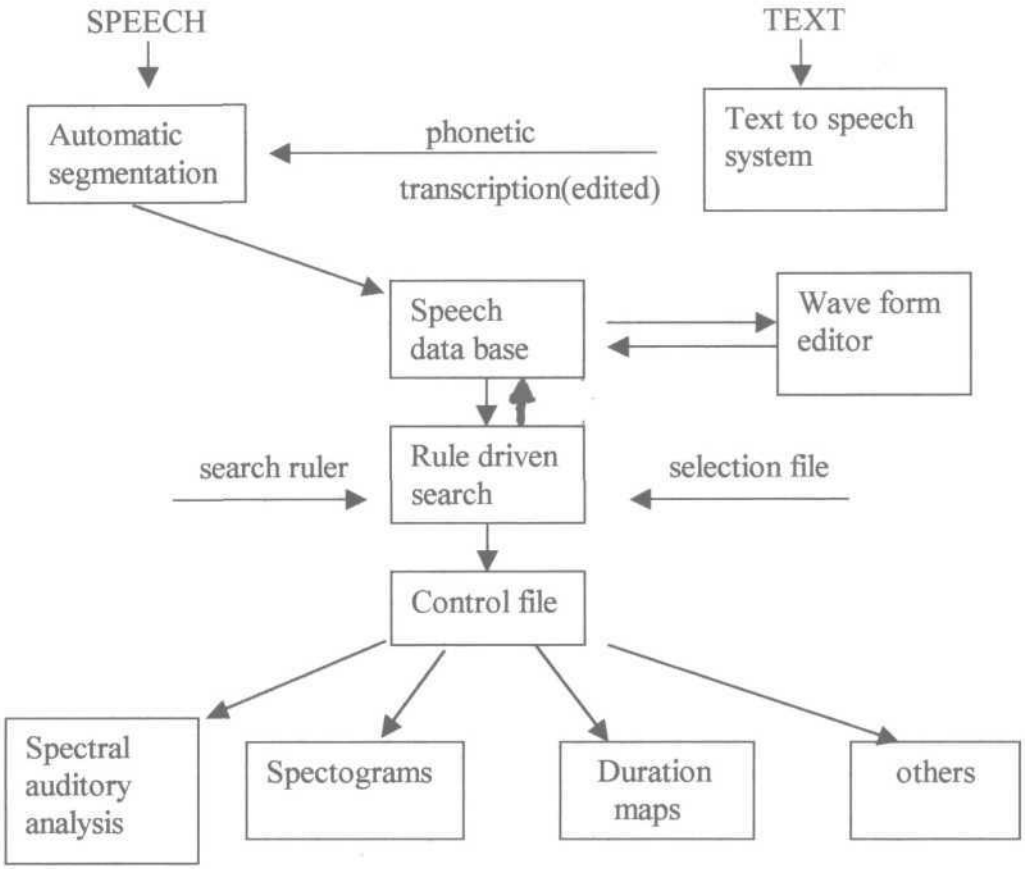

Figure 1 :Block diagram illustrating creation of database

The phonetic transcription is used by an automatic segmentation program (Blomberg and Elenius,1985) to distribute the phonetic labels along the wave form. The segmentation program gives an estimate of time positions of each phoneme. When a detailed analysis should be done, the labels have to be checked and corrected. This is done with by means of a wave form editor program, which is a general purpose program for labeling and editing sample files. It is able to handle multiple channels. Example, means of the joystick sample can be labeled and labels can be changed. These can be used to edit the waveform, to include waveforms from other recordings, to delete parts of it, and to do scalings of amplitude.

The data base is accessed by means of rules. By a brief rule statement, speech segments meeting the specified contextual conditions can be identified. The rule structure is similar to the notation used in generative phonology and is also used in the text to speech system inventory and the features that are associated to each phonemes. A selection file specifies which label files form the data base. The label files are read into the program one by one. The control file contains the phonetic transcriptions with the star positions marked (*). Each star is then decoded into a control line including the wave form file name, the specified parameters and the name of the phonetic symbol that follows the star.

Spectral analysis:

A spectral analysis is done and the histogram is calculated for the energy distribution at each point along the frequency axis. These histograms are connected into a contour spectrum, each line representing 10 percent of the distribution. The spectrum is more condensed at low compared to high frequency indicating that the glottal source has a steeper slope for low effort whereas the amplitude of the voice is relatively independent *ol* effort. We can observe how constant each phoneme is realized, and coarticulation effects can be searched for.

In the rule system, a classifier is used to group duration data on vowels depending on the stress level, phonological length, and whether the vowel is followed by an unvoiced stop or not.

Duration relations and prediction of durations :

Normally, when duration is studied, relation between phonemes is given more interest than the absolute duration. The inherent duration of each consonant category plays an important role.

The popular and successful method used to provide the database for the sounds so as to enable synthesis and text to speech conversion is spectrography. To use the database, phonemes which are of interest to be synthesised are selected. The environment (preceding and the following phonemes) in which the phoneme has to be synthesised is noted and the experiments needs to specify whether the phonemes have to be synthesized for male or female subjects. Depending on this ,the values of the parameters are selected. We can select any parameter depending on the kind of synthesizer we have to effectively synthesise a speech sound.

One of the major tasks in the text to speech synthesis is to provide a voice input / output capability. In order to achieve this ,acoustic (segmental and suprasegmental) data in different languages is essential. To utilize the computers for Indian languages, acoustic data for the same is required which would facilitate automatic speech recognition, speaker recognition and speech synthesis. The acoustic data is extracted by recording speech and using softwares.

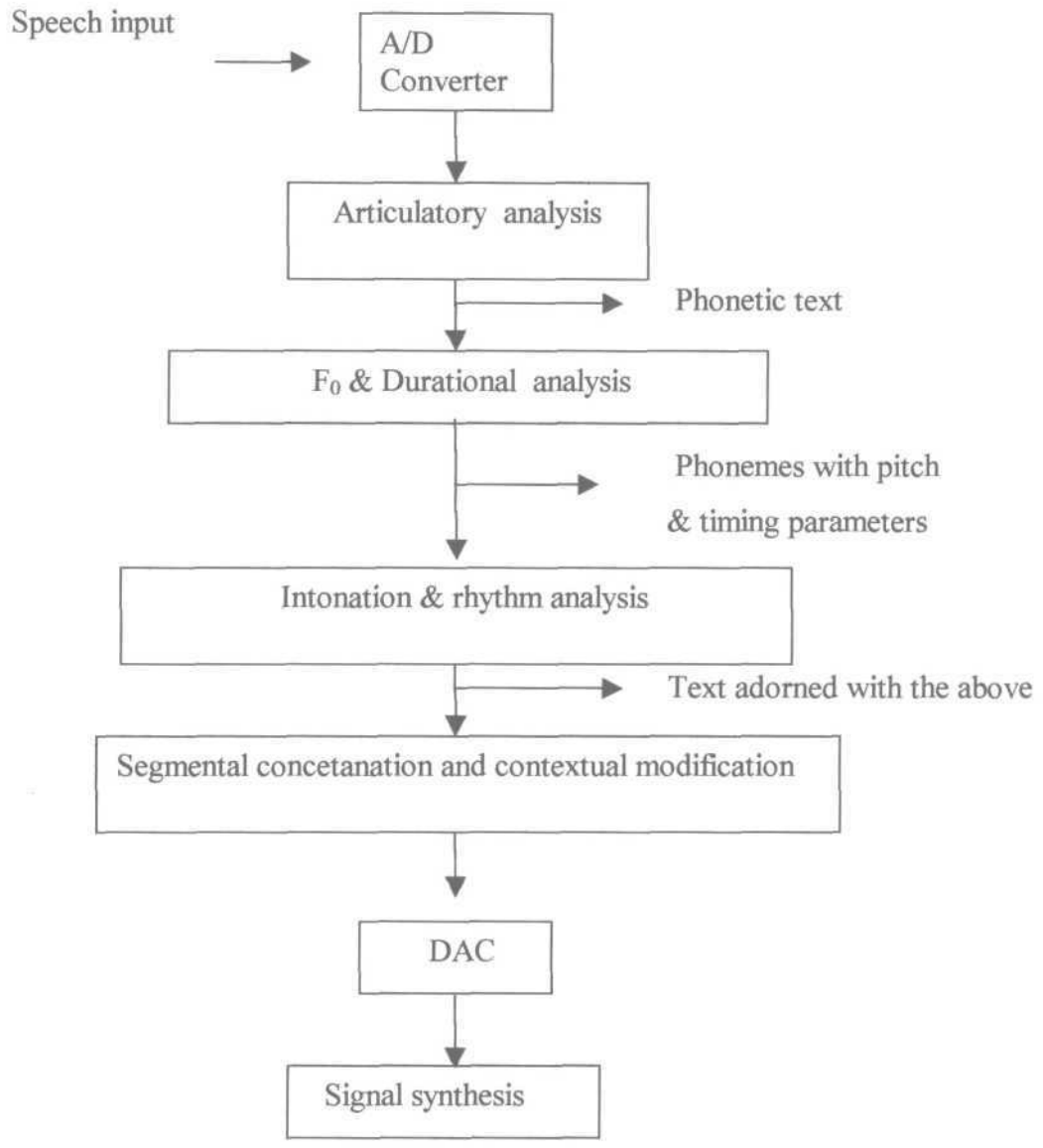

Fig 2: System for acoustic feature extraction.

Figure 2 provides a system for acoustic feature extraction. The obvious way to provide speech output from computers is to select the basic acoustic units to be used, record them, and generate utterances by concatenating together appropriate segments from this pre-stored inventory. The basic units can be whole sentences, words, syllables or phonemes.

If real speech is connected, there are few gaps between words. Coarticulation, where sounds are affected by those on either side, naturally operates across word boundaries. The time constants of coarticulation are associated with the mechanics of the vocal tract and hence measure tens or hundreds of milliseconds. Thus the effects straddle several pitch periods 100Hz (pitch has 10msec period) and cannot be simulated by simple interpolation of the speech wave form. Prosodic features, notably pitch and rhythm, span much longer stretches of speech than single words. Word to word coarticulation and the essential features of rhythm and intonation can be incorporated if the stored words are coded in source filter form.

Witten (1982) noted that the ultimate in economy of inventory size, is to use phonemes as the basic unit. This makes the most crucial part of the task interpolation between units, rather than their construction or recording. With only about forty phonemes in English, each one can be examined in many different contexts to ascertain the best data to store. It is most appropriate that data be abstracted from several different realizations rather than an exact record made of any one.

If information is stored about phonological units of speech, phonemes, the difficult task of phonological-to-phonetic conversion must necessarily be performed automatically.

AVAILABLE SPEECH DATA BASE :

1) ESTONIAN SPEECH DATA BASE :

Estonian speech database was developed in the framework of the EUCOPERNICUS project. "BABEL - A multi - language database" in 1995-1998. The database includes a short description of the Estonian sound system, systematic text corpus consisting of several parts (consonant - vowel - consonant units, numbers, isolated sentences, 5 sentence passages) and the recorded speech signals. In recording the database, 70 speakers (35 male and 35 female) of different age groups participated whose pronunciation was in accordance with the norm of standard Estonian. Thus the database represents the Estonian pronunciation norm of the end of the 20^{th} century. The database includes 12 hours of high quality speech material on three 3D-ROMs. The database will be used for the phonetic and phonological research of Estonian as well as for developing and testing the system of speech synthesis and recognition.

2) OXFORD ACOUSTIC PHONETIC DATABASE :

It contains data on vowel - consonant and consonant vowel combinations in both stressed and unstressed locations. The language covered include French, German, Hungarian, Italian, Japanese, British English, Spanish and English. It is available on compact disc.

3) EUROM - A MULTILINGUAL EUROPEAN SPEECH DATABASE :

It is the first European really multilingual speech database. It has equivalent corpora for each of the languages; same number of speakers selected

SAM format compatible - databases, as the French BD SONS and EUROM 1 French part, and the ten other parts of EUROM 1 could be included as well.

4) THE EMU SPEECH DATA BASE SYSTEM :

EMU is a collection of software tools for the creation, manipulation and analysis of speech databases. At the core of EMU is a database search engine which allows the researcher to find various speech segments based on the sequential and hierarchical structure of the utterances in which they occur. EMU includes an interactive labeler which can display spectrograms and other speech wave forms, and which allows the creation of hierarchical, as well as sequential, labels for a speech utterance.

5) README FOR THE ATIS - 2 MADCOW SPEECH CORPUS :

The ATIS -2 corpus is a 4 - cdrom (1993) set containing recordings of spontaneous speech from 453 speakers, collected at six different research laboratories around the United States.

The text data includes the labels of airline travel information and associated material used to build the relational data base, which was used to provide answers to the verbal request of the speakers.

The documentation (contained in the "doc" directory on this disc) includes the specifications for the directory organization of the corpus, and for the internal format of the wave form files and varies annotation files. It also provides information on the speakers, the techniques for data collection and evaluation.

in same way and recorded in the same conditions with common file formats. Initially eight European countries have made recording: Italy, United Kingdom, Germany, Netherlands, Denmark, Sweden, Norway, France. Additional recording have been then completed in Greece, Spain and Portugal.

The recording was done with the sampling rate of 20kHz and codage of 16 bits. All recordings were done in an anechoic room and according to common recording protocol all files in SAM format.

Corpora:

- 40 passages (five thematically linked sentences each)
- 50 patching sentences (in order to cover some language specificities)
- "Logatomes" : 82 CVC (combination of consonant + vowel + consonant) variation of initial and final consonant with vowels (a, i, u). these "logatomes" are pronounced with and without context (5 different context).
- Words (isolated), from the five context
- Numbers : 5 x 20 numbers (between 0 and 9999) covering all phonotactic possibilities if the number system.

Speakers:

30 male and 30 female (b/w 20 and 60 years old) data from the French part of EUROM 1 are integrated in the multibase multilingual speech data base management system GERSONS, which is a PC - software based on KMAN RDMS (knowledge man from DBMS company). This software manages **all**

SAM format compatible - databases, as the French BD SONS and EUROM 1 French part, and the ten other parts of EUROM 1 could be included as well.

4) THE EMU SPEECH DATA BASE SYSTEM :

EMU is a collection of software tools for the creation, manipulation and analysis of speech databases. At the core of EMU is a database search engine which allows the researcher to find various speech segments based on the sequential and hierarchical structure of the utterances in which they occur. EMU includes an interactive labeler which can display spectrograms and other speech wave forms, and which allows the creation of hierarchical, as well as sequential, labels for a speech utterance.

5) README FOR THE ATIS - 2 MADCOW SPEECH CORPUS :

The ATIS -2 corpus is a 4 - cdrom (1993) set containing recordings of spontaneous speech from 453 speakers, collected at six different research laboratories around the United States.

The text data includes the labels of airline travel information and associated material used to build the relational data base, which was used to provide answers to the verbal request of the speakers.

The documentation (contained in the "doc" directory on this disc) includes the specifications for the directory organization of the corpus, and for the internal format of the wave form files and varies annotation files. It also provides information on the speakers, the techniques for data collection and evaluation.

6) ENTROPIC LATINO 40 SPEECH DATABASE BY BERNSTEIN(1994) :

It was developed by Grundy, Bernstein, Rosenfeld, Najmi & Mankoski (1994). Entropic research laboratory designed the Latino 40 database to provide a set of recording for training speaker independent system that recognize Latin - American Spanish. The data base comprises about 5000 utterance files. These files include about 125 utterances from each of 40 different speakers, 20 males and 20 females. The recordings were all made with a high quality head mounted microphone (Shure SM 10A) in an office environment, and the utterances were digitized in 16 bit samples at 16kHz.

The material consisted of 13000 sentences that include a number of anomalous or ambiguous forms. The sentences are all shorter than 80 characters, and are not grouped into larger constituents like paragraphs or stories.

The speakers were adults, ranged from 18-59 years of age and were native speakers of Latin American Spanish.

7) **THE IRIS SPEECH DATA BASE** :

The IRIS database contains speech samples from approximately 100 languages and from variants of Swedish spoken by several ethnic minorities residing in the country. This data base is meant to provide an easily accessible reference material for cross language studies in phonetic typology and the phonetics of Swedish an a second language.

8) CASTRO SPEECH DATA BASE :

It is a database containing the full text of English translations of speakers, interviews and press conferences by Castro. These records are in the public domain.

9) LINGUISTIC DATA CONSORTIUM (LDC) :

It provides a very wide range of speech and text data to research and commercial users.

10) JAPANESE SPEECH AND TEXT CORPORA (1997) :

It is a continuous speech corpus for research in 7 CD-ROMs. It consists of ATR PB sentences, 9600 sentences and read speech.

There are other databases like **BUPT SPOKEN DIGIT DATABASE IN CHINESE UNIVERSITY and VICTORIA PHONETIC DATA BASE.**

TELEPHONE DATABASES

1) AUSTRALIAN ENGLISH TELEPHONE SPEECH DATA BASE:

It has been offered by Call data base Ltd. It is a 1000 speaker fixed network telephone speech database for Australian English recordings from 1000 male and female speakers from across Australia include a series of items suitable for development of a broad range of telephone ASR services. The speaker demographics reflect the geographical distribution and ethnic diversity of Australian citizens.

The data base is in speech data (II) format and comprises over 46,000 A-law encoded speech files recorded via ISDN lines, orthographic annotation for each utterance, and a SAMPA encoded pronounciation lexion. It has been validated by SPEX.

1) ANNION DATA BASE :

The speech data base was collected over the fixed telephone network for the purpose of automated tele services. It consists of typical application words and sentences.

2) **KING SPEECH DATA BASE** :

The KING corpus was created for research in the area of speaker identification. There are 26 Diego speakers (and 25 New Jersey speakers). All speakers are male. Each session was recorded in both a wide band and a narrow band channel. The narrow band channel represents speech that was passed through a standard telephone handset, transmitted through a local telephone

exchange to a long distance service and back to the local exchange, then recorded from an analog telephone patch. The wide band channel represents the same utterance, recorded using a high quality microphone that was mounted on the telephone handset.

Both channels are digitized at 8kHz with 16 bit linear samples.Each speaker/session wave form file consists of about thirty seconds of actual speech of the person speaking on one of the following assigned topics -

1. construction toy task
2. describing odd shapes
3. topic of speakers choice
4. road rally task
5. describing photographs
6. describing cartoon strips

Nataraj, Savithri, Venkatesh and Rohini (1990) took fifteen subjects and analysed the spoken form of the Hindi to collect information on acoustic features of Hindi sounds, which acted as input to the designing a speech to text system for isolated utterances of characters of Hindi alphabet. The speech recognition system/speech to text system was developed to provide therapy and teaching speech and language to the hearing handicapped.

The parameters measures were vowel duration, word duration, burst duration voice onset time, closure duration, range of burst frequency, fundamental frequency, formant frequencies (1-3), energy levels of formants,

bandwidths of formants, speed of transaction, transition duration and range of transition.

Savithri (1994) developed database for plosives in Kannada. Twenty nine meaningful kannada words with all the plosives - /k/, /k^h/, /g/, /g^h/, /t/, /t^h\ /d/, /d^h/, *ItI*, /t^h/, /d/, /d^h/, /p/, /p^h/, /b/ and /b^h/ were selected. Three adult males, one adult female and one child uttered these words four times. All these words were audio recorded and seven selected parameters and six temporal parameters were measured using spectography and wave forms. Totally 1508 tokens were measured for each subject. The parameters chosen were closure duration, burst duration, aspiration duration, voice onset time, transition duration for all three formants, speed of transition of the three formants, duration of the stop consonant, preceding vowel duration, terminal frequency (Fl - F3), fundamental frequency changes from vowel to plosive.

Agrawal (1992) developed the database for the Hindi language for the digits from zero to nine.

The review indicates wide data base for non-Indian languages mostly European and English. The Indo-European and Dravidian languages are different from the European languages. If one wants to have a speech recognition system or a text to speech system, it is essential that acoustic phonetic and prosodic data base are available in these languages. Though an attempt was made by **Nataraja et al (1990)** to prepare a database for Hindi it is incomplete in its nature. In this context the present study aims to make an acoustic phonetic data base for Hindi.

CHAPTER III

METHOD

Material:

It consisted of one hundred and five meaningful Hindi words with ten vowels, fifteen plosives, four affricates, three nasal continuants, four fricatives, one lateral, one trill and one semi vowel in the initial, medial and final positions. The word list is shown in table 1.

Subjects:

Two native Hindi normal speakers (one male and one female) of 22 years of age participated in the study.

Procedure:

The words as written one each on a card were visually presented to the subjects. They were instructed to read the word on to a microphone (H Legend) kept at a distance of 10 cms from the mouth. The words were audio recorded on to a cassette (Meltrack 60) using Sony stereo cassette deck TC-FX 170.

Acoustic analysis :

Using the SSL software (voice and speech systems, Bangalore), the words were digitized at a sampling rate of 16000 Hz using a 12 bit A/D converter and stored into the computer memory. Using the waveform display and wide band spectograph of the SSL software, the following parameters were extracted.

Vowel	Description	Word initial	Word medial	Word final
/a/	Low mid vowel(short)	ana:r	-	-
/a:/	Low mid vowel(long)	a:m	ana:r	pakɔda:
/ i /	Front high vowel(short)	imli	dalija	-
/ i: /	Front high vowel(long)	I:k^{h}	mani:ʃ	-
/u/	Back high vowel(short)	ullu:	anudʒ	tanu
/u:/	back high vowel(long)	u:n	tau:dʒi	ullu
/e/	Front mid vowel(short)	ek	anek	pake
/ɤ/	Front mid vowel(long)	ɔe nek	-	-
/o/	Back mid vowel(short)	ok^{h}ali:	manod	karo
/ɔ/	Back mid vowel(long)	ɔrat	pakɔda	lɔ
Plosives				
/k/	velar unvoiced unaspirated plosive	kul	maka:n	tapak
/k^{h}/	velar unvoiced aspirated plosive	kha:na	tʃark^{h}a	i:k^{h}
/g/	velar voiced unaspirated plosive	ga:di	magar	dʒag
/g^{h}/	Velar voiced aspirated plosive	g^{h}ar	meg^{h}a:	meg^{h}
/ʈ/	Retroflex unvoiced unaspirated plosive	ʈa:l	ʈamaʈar	palaʈ
/ʈ^{h}/	retroflex unvoiced aspirated plosive	ʈ^{h}aʈ^{h}era	ʈ^{h}aʈ^{h}era	Maʈ^{h}
/ɖ/	Retroflex voiced unaspirated phoneme	ɖa:l	paɳɖit	-
/ɖ^{h}/	Retroflex voiced aspirated phoneme	d^{h}akkan	-	-
/t/	Dental unvoiced unaspirated plosive	ta:l	bata:ʃa	kaja:mat
/t^{h}/	Dental unvoiced aspirated plosive	t^{h}a:li	kat^{h}a:	rat^{h}
/d/	Dental voiced unaspirated plosive	dava:t	ma:dak	kad
/d^{h}/	Dental voiced aspirated plosive	d^{h}anija:	ad^{h}arm	du:d^{h}
/p/	Bilabial unvoiced unaspirated plosive	pal	kapati	ta:p
/b/	Bilabial voiced unaspirated plosive	bal	kabaddi	kab
/b^{h}/	Bilabial voiced aspirated plosive	b^{h}a:lu	kab^{h}i	ʃub^{h}
Affricates				
/tʃ/	Palatal unvoiced unaspirated affricate	tʃal	katʃada:	tʃammatʃ
/tʃ^{h}/	Palatal unvoiced aspirated affricate	tʃ^{h}atri	matʃ^{h}ali	Kutʃ^{h}
/dʒ/	Palatal voiced unaspirated affricate	dʒa:l	ka:dʒ^{h}al	gadʒ
/dʒ^{h}/	Palatal voiced aspirated affricate	dʒ^{h}arana:	dʒ^{h}andʒ^{h}ar	bodʒ^{h}
Fricatives				
/h/	Glottal voiced fricative	hal	va:han	vah
/ʃ/	Palatal unvoiced fricative	ʃalagam	kiʃore	avaka:ʃ
/s̃/	Palatal unvoiced fricative	s̃at^{h}kon	kos̃ʈ^{h}ak	d^{h}anus̃
/s/	Palatal unvoiced fricatve	sahan	p^{h}asal	pados
/f/	Labiodental unvoiced fricative	fal	safal	sa:f
Nasals				
/ɳ/	Retroflex nasal	-	-	kaɳ
/n/	Dental nasal	neem	mana:na	m n
/m/	Bilabial nasal	ma:n	somavar	kalam
Lateral				
/l/	Lateral	la:l	tʃalana	la:l
Trill				
/r/	Post alveolar trill	rel	keral	ka:r
Semi vowel				
/v/		va:han	avaka:	na:v

Table 1 : Word list

Frequency domain parameters:

1. Formant frequency : The frequency of the formant as measured in the steady state (F_1 F_2, F_3).

2. Terminal frequency : It is the frequency of the formant of the preceding vowel at the end of the transition or the following vowel at the beginning of the transition.

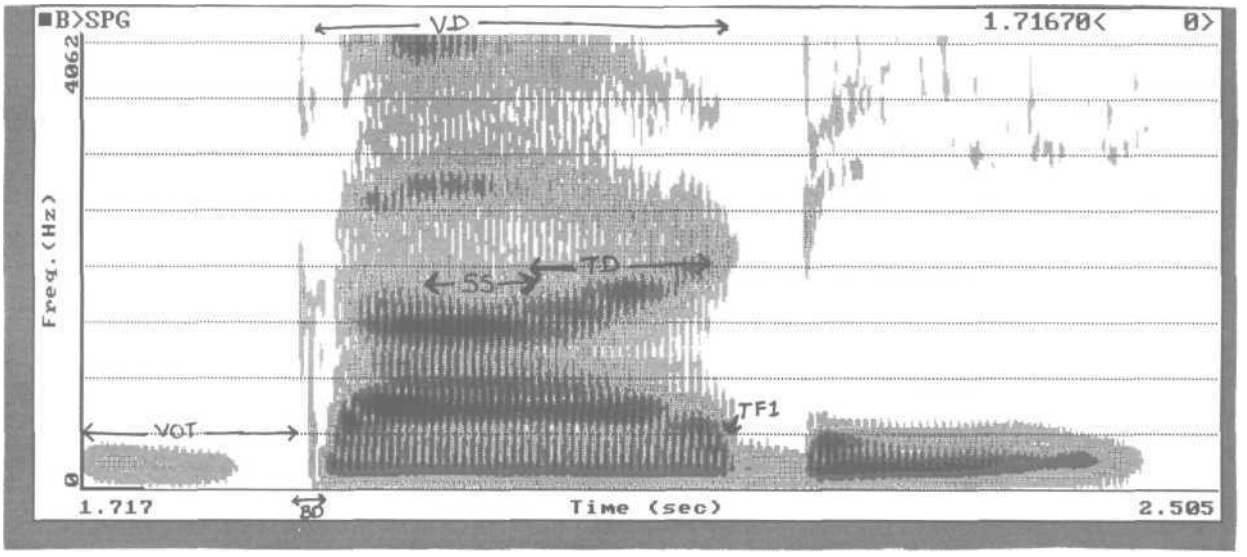

Figure 3 : Spectograph showing terminal frequency transition duration, steady state duration of the formant, burst duration, vowel duration and voice onset time.

Temporal domain parameters:

1. Formant steady state duration : Duration for which the formant is in the steady state.

2. Transition duration : The transition duration is the time difference between the onset and offset of the transition.

3. Closure duration :The difference between the offset of the preceding vowel and the onset of the burst for the plosive.

4. Burst duration : Time difference between the onset of irregular vertical striation depicting the articulatory release and the offset of the same.

5. Aspiration duration : Duration of the aspiration depicted as irregular vertical striations in the low frequency region between the burst and the onset of resonance for the felling one.

6. Voice onset time : Time difference between the onset of the burst depicted as quasirandom wave form and the onset of vocal fold vibration depicted as quasiperiodic wave form.

7. Total duration : Time difference between the onset and offset of the phonemes.

8. Vowel duration: Time difference between onset and offset of the phoneme.

9. Frication duration : Time duration for which fixation existed, marked as fills in high frequency region of spectogram.

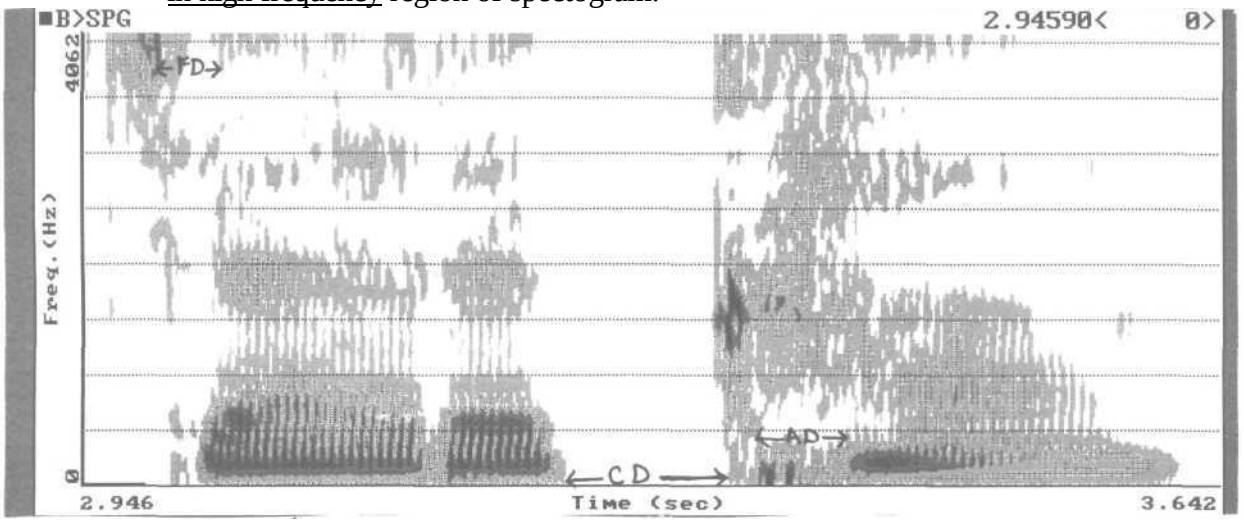

Figure 4 : Spectograph showing frication noise duration, aspiration duration and closure duration

CHAPTER-IV

RESULTS AND DISCUSSION

Various temporal and spectral values of the parameters measured for different speech sounds for male and female subjects are provided in tables (Table 2 and 3 respectively)

The abbreviations used in the tables are as follows :

1. F_1 - First formant frequency
2. F_2 - Second formant frequency
3. F_3 - Third formant frequency
4. Fl SS - Steady states of Fl
5. F2 SS - Steady states of F2
6. F3 SS- Steady states of F3
7. Fl TD-Transition duration of Fl
8. TF1 - Terminal frequency of Fl
9. F2TD - Transition duration of F2
10. TF2-Terminal frequency of F2
11. F3 TD - Transition duration of F3
12. TF3 - Terminal frequency of F3
13. CD - Closure duration
14. BD - Burst duration
15. ToD - Total duration
16. AD - Aspiration duration

17. FD - Frication duration

18. VD - Vowel duration

19. VOT - Voice onset time

20. NTS - No transition seen

Table 4 and 5 show values of various parameters for plosives and affricates respectively.

Parameter	Male		Female	
	Voiced	Unvoiced	Voiced	Unvoiced
VOT	-15 to-86	22-98	-28 to -98	15-105
BD	4-26	7-22	5-30	6-26
AD	54-90	33-168	24-121	70-173
ToD	13-74	16-278	50-246	25-310

Table 4 : Range of values for the plosives for different temporal measures (msec)

It was observed that the aspiration duration and total duration were larger for unvoiced plosives. Lead VOT was observe was observed for unvoiced plosives.

Parameter	Male		Female	
	Voiced	Unvoiced	Voiced	Unvoiced
ToD	43-201	65-252	70-300	85-420
FD	35-140	39-142	52-210	75-200
AD	50	34	87	41-75
BD	5-75	6-12	Not seen	5-6

Table 5 : Range of values of the temporal measures for the affricates (msec)

It was observed that the total duration of unvoiced affricates was longer than that of voiced affricates.

17. FD - Frication duration

18. VD - Vowel duration

19. VOT - Voice onset time

20. NTS - No transition seen

Table 4 and 5 show values of various parameters for plosives and affricates respectively.

Parameter	Male		Female	
	Voiced	Unvoiced	Voiced	Unvoiced
VOT	-15 to-86	22-98	-28 to -98	15-105
BD	4-26	7-22	5-30	6-26
AD	54-90	33-168	24-121	70-173
ToD	13-74	16-278	50-246	25-310

Table 4 : Range of values for the plosives for different temporal measures (msec)

It was observed that the aspiration duration and total duration were larger for unvoiced plosives. Lead VOT was observed voiced plosives and lag VOT was observed for unvoiced plosives.

Parameter	Male		Female	
	Voiced	Unvoiced	Voiced	Unvoiced
ToD	43-201	65-252	70-300	85-420
FD	35-140	39-142	52-210	75-200
AD	50	34	87	41-75
BD	5-75	6-12	Not seen	5-6

Table 5 : Range of values of the temporal measures for the affricates (msec)

It was observed that the total duration of unvoiced affricates was longer than that of voiced affricates.

Vowel														
Word	Context	VD	F_1	F_2	F_3	F1SS	F2SS	F3SS	F1TD	TF1	F2TD	TF2	F3TD	TF3
ana:r	a-n	43	870	1687	3311	372	38	27	NTS	870	16	1513	27	3357
a:m	a:-m	381	212	714	1200	319	195	158	44	1012	48	698	59	1043
ana:r	n-a:	232	768	1578	2968	183	81	140	20	698	16	1278	32	2894
	a:-r	232	768	1578	2968	183	81	140	NTS	-	72	2337	81	3183
pakɔḍa	ḍ-a:	254	745	1372	-	131	131	-	61	62	63	1247	-	-
imli	i-m	81	227	2298	2705	81	27	27	NTS	-	22	1200	-	-
i:k^h	i:-k^h	261	415	2486	3114	261	211	211	NTS	-	22	2329	25	2752
mani:ʃ	n-i:	199	352	2502	3003	81	27	63	NTS	-	20	2376	25	2847
	i:-ʃ	199	352	2502	3003	81	27	63	NTS	-	54	2368	NTS	-
mari:	r- i:	216	415	2000	3009	194	84	74	NTS	-	46	2141	NTS	-
ullu	u-l	70	494	1200	2345	70	70	57	NTS	-	73	1529	19	1749
anudʒ	n-u	102	698	1498	2549	30	15	35	18	651	30	1309	-	-
	u-dʒ	102	698	1498	2549	30	15	35	50	557	45	1843	-	-
tanu	n-u	210	352	1472	2800	210	103	-	NTS	-	47	1200	-	-
u:n	u:-n	314	290	8235	-	314	104	-	NTS	-	63	1247	-	-
tau:dʒi	a-u:	150	463	964	2564	82	22	75	45	588	37	1122	37	2313
	u:-dʒ	150	463	964	2564	82	22	75	30	494	52	1372	75	2094
ok^{h}li	o-k^h	157	478	854	-	157	135	-	NTS	-	18	745	-	-
manodʒ	n-o	223	360	651	996	195	157	60	NTS	-	34	635	75	1263
	o-dʒ	223	360	651	996	195	157	60	45	290	30	572	75	1357
karo	r-o	220	541	1043	2376	220	120	100	NTS	-	82	1262	65	1906
ɔrat	ɔ -r	230	667	1109	2204	230	145	154	NTS	-	85	1420	68	1592
pakɔḍa	k-ɔ	168	525	996	2188	155	90	45	NTS	-	-	-	15	2157
	ɔ -ḍ	168	525	996	2188	155	90	45	10	494	65	1184	65	1529
lɔ	l-ɔ	305	400	1482	1984	137	106	-	11	603	45	1137	-	-
ek	e-k	300	430	2156	2611	254	276	-	34	362	21	2023	-	-
ɑnek	n-e	247	445	2125	2732	100	121	94	13	682	13	1968	9	2596
	e-k	247	445	2125	2732	100	121	94	27	651	63	2094	75	2533
pake	k-e	207	494	2125	2533	139	184	184	18	452	22	2313	13	2439
ɤnek	ɤ-n	101	360	-	2231	101	-	-	10	568	16	1447	11	2514

Plosive													
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	VOT	CD	BD	ToD	AD	VD
kul	k-u	324	745	2125	NTS	45	45	30	-	16	16	-	110
maka:n	a-k	678	1215	2109	NTS	32	NTS	-	73	15	89	-	80
	k-a:	713	1372	2470	25	55	30	-	73	11	17	-	175
ṭapak	a-k	525	1231	2235	32	32	14	-	-	-	-	-	117
kʰa:na:	kʰ-a:	760	1294	2407	NTS	NTS	39	74		13	89	94	146
ʧarakʰa:	kʰ-a:	713	1262	2125	NTS	11	NTS	-	90	12	161	63	201
i:kʰ	i:kʰ	250	2203	2203	25	33	76	-	-	21	273	168	270
ga:ḍi	g-a:	556	1380	2392	23	22	33	63	-	-	63	-	234
magar	a-g	619	1184	2251	16	21	18	-	66	10	76	-	200
	g-a	462	1310	2439	19	15	NTS	-	66	10	76	-	200
ʤag	a-g	550	1219	2029	NTS	NTS	10	-	-	-	120	-	98
gʰar	gʰ-a	494	1278	2223	27	19	19	-	-	12	71	59	174
megʰa:	e-gʰ	509	1969	2565	56	17	53	29	35	26	61	50	273
	gʰ-a:	541	1482	2486	17	30	17	-	35	12	61	50	262
megʰ	e-gʰ	402	1984	2471	31	43	62	-	-	12	89	70	221
ṭa:l	ṭ-a:	666	1480	2520	18	18	33	45	-	9	69	-	209
ṭama:ṭar	a:-ṭ	635	1573	2568	71	91	68	-	62	11	78	-	189
	ṭ-a	525	1670	2392	57	57	60	-	62	11	78	-	128
ṭʰaṭʰera:	ṭʰ-a	494	1623	2141	14	NTS	12	94	-	9	60	51	94
ṭʰaṭʰera:	a-ṭʰ	556	1858	2078	46	46	20	-	51	10	92	49	96
	ṭʰ-e	462	2000	2596	40	20	15	-	51	10	92	49	174
maṭʰ	a-ṭʰ	619	1702	1952	11	34	17	-	-	21	157	55	102
ḍa:l	ḍ-a:	619	1529	2423	11	49	NTS	68	-	4	124	-	219
paɳḍit	ḍ -i	380	1840	2502	10	39	39	-	68	6	70	-	99
ḍʰakkan	ḍʰ-a	620	1576	2533	14	15	NTS	29	-	9	108	90	138
ta:l	t-a:	682	1137	2329	21	21	36	33	-	7	40	-	227
bata:ʃa	a-t	509	1380	2611	30	11	13	-	80	15	107	-	67
	t-a:	541	1341	2627	17	9	11	-	80	15	107	-	139

Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	VOT	CD	BD	ToD	AD	VD
kajamat	a-t	619	1310	2000	16	11	26	-	-	9	118	-	105
t^ha:li	t^h-a:	745	1247	-	NTS	10	-	60	-	5	130	65	153
kat^{h}a:	a-t^h	463	1247	2690	12	11	20	-	60	12	42	33	89
	t^h-a:	745	1359	2471	NTS	20	15	-	60	12	42	33	224
rat^h	a-t^h	541	1482	2580	18	16	60	-	-	12	90	78	128
dava:t	d-a	598	1286	2525	35	6	5	15	-	8	23	-	105
ma:dak	a:-d	651	1341	2078	26	11	NTS	-	15	6	93	-	160
	d-a	490	1431	2686	16	NTS	NTS	-	15	6	93	-	80
kad	a-d	557	1560	2596	26	35	15	-	-	16	82	-	135
d^{h}anija	d^h-a	509	1655	2590	20	NTS	NTS	46	-	7	107	54	93
ad^{h}arm	a-d^h	490	1272	2643	15	14	18	-	46	10	136	80	62
	d^h-a	454	1373	2566	25	29	80	-	46	10	136	80	182
du:d^h	u:-d^h	462	1084	-	NTS	38	-	-	-	12	120	108	240
pal	p-a	635	1074	2267	NTS	71	NTS	10	-	7	17	-	108
kapaṭi	a-p	462	1027	2062	24	26	20	-	112	7	119	-	86
ṭa:p	ɑ:-p	598	1105	2156	19	19	34	-	-	-	-	-	180
bal	b-a	447	1090	2345	35	72	20	95	-	7	103	-	110
kabaḍḍi	a-b	447	996	1986	13	33	11	-	95	7	102	-	80
	b-a	619	1106	2298	NTS	74	18	-	95	7	102	-	81
kab	a-b	494	1012	2204	21	33	16	-	-	-	86	-	93
b^{h}a:lu	b^h-a:	717	1184	2376	10	18	24	-	-	-	76	76	180
kab^{h}i	a-b^h	478	1153	1986	10	18	24	-	31	8	240	201	65
	b^h -i	431	2031	2423	NTS	37	52	-	31	8	240	201	223
ʃub^h	u-b^h	414	886	2210	11	11	9	-	-	6	180	65	108

Affricates														
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	VOT	CD	BD	ToD	AD	FD	VD
tʃal	tʃ-a	462	1608	2643	17	37	34	65	-	6	65	-	50	104
katʃaḍa:	a-tʃ	525	1671	2738	24	27	25	-	31	12	80	-	62	92
	tʃ-a	572	1829	2502	0	32	22	-	31	12	80	-	62	87
tʃammatʃ	a-tʃ	572	1593	2722	18	78	52	-	-	-	194	-	116	112
tʃʰatri	tʃʰ-a	509	1786	2784	24	20	34	58	-	-	85	34	39	116
matʃʰali	a-tʃʰ	494	1607	2769	30	21	0	-	57	-	148	-	79	103
	tʃʰ-a	603	1263	2141	0	30	41	-	57	-	148	-	79	142
kutʃʰ	u-tʃʰ	390	1372	3035	21	55	39	-	-	-	252	-	142	98
dʒal	dʒa	478	1905	2643	43	31	29	129	-	6	180	-	51	101
ka:dʒal	a-dʒ	588	1717	1874	50	76	69	-	129	5	169	-	35	195
	dʒ-a	447	1921	2862	40	30	30	-	129	5	169	-	35	100
gadʒ	a-dʒ	494	1874	2752	47	56	49	-	-	7	201	-	125	159
dʒʰarna	dʒʰ-a	588	1952	2752	41	39	26	59	-	-	100	-	37	142
dʒʰandʒʰar	dʒʰ-a	651	1686	2549	86	69	69	-	109	-	152	-	43	148
bodʒʰ	o-dʒʰ	478	1451	2690	51	69	52	-	-	-	190	50	140	199

Fricatives										
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	FD	VD
hal	h-a	745	1256	2408	NTS	35	35	72	75	98
va:han	a:-h	666	1269	2565	NTS	17	NTS	55	52	192
	h-a	603	1451	1984	20	10	NTS	55	52	65
vah	a-h	572	1639	2396	35	33	12	67	65	95
ʃalgam	ʃ-a	447	1592	2376	16	14	18	135	101	65
kiʃore	i-ʃ	243	2063	2627	23	NTS	8	116	108	54
	ʃ-o	447	1655	2000	23	80	10	116	108	255
avaka:ʃ	a:-ʃ	572	1581	2150	45	35	32	150	150	235
ʃ̃aṭʰkoŋ	ʃ̃- a	510	1655	2555	29	29	29	107	92	97
koʃ̃ṭʰak	o-ʃ̃	463	1451	1812	26	76	35	77	75	150
dʰanuʃ̃	u-ʃ̃	572	1529	2047	15	30	15	175	170	134
sahan	s-a	463	1529	2566	34	64	NTS	131	131	116
fasal	a-s	510	1357	-	24	18	-	122	99	89
	s-a	494	1561	2439	5	20	10	122	99	75
paḍos	o-s	431	1216	2298	24	54	26	192	192	198
fal	f-a	603	1215	2188	NTS	28	27	72	70	85
safal	a-f	588	1215	2439	15	38	18	100	91	64
	f-a	486	1090	2223	28	51	52	100	91	96
sa:f	a:-f	572	1074	2567	40	60	36	320	320	240

Nasals									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
kaṇ	a-ṇ	713	1372	1733	36	24	26	40	118
ni:m	n-i:	450	2430	3002	79	70	49	140	250
mana:na:	a-n	680	809	2641	17	65	20	95	150
	n-a	520	1489	-	20	42	-	95	201
mɔn	ɔ-n	474	1097	-	23	54	-	87	163
ma:n	a:- n	854	1138	-	43	43	-	120	236
somava:r	o-m	322	854	2063	30	88	NTS	90	280
	m-a	510	1490	2282	60	80	40	90	220
kalam	a-m	607	1169	2094	29	37	40	137	72

Lateral									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
la:l	l-a	651	1451	2471	19	NTS	30	119	215
tʃalana	a-l	604	1663	2533	25	NTS	NTS	69	85
la:l	a:-l	635	1451	2455	27	62	8	125	215

Trill									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
rel	r-e	541	2047	2753	8	8	22	39	291
ƙeral	e-r	447	1765	2345	15	66	71	36	177
	r-a:	447	1514	2361	35	15	NTS	36	103
ka:r	a:-r	620	1310	-	24	31	-	28	22

Semivowel									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
va:han	v-a:	478	886	2361	63	75	25	82	205
avaka:ʃ	a-v	494	918	2220	40	58	NTS	54	87
na:v	a:-v	478	1043	2390	90	69	40	106	253

Table 2 : The segmental features (spectral and temporal) for the male subject

Vowel														
Word	Context	VD	F_1	F_2	F_3	F1SS	F2SS	F3SS	TDF1	TF1	TDF2	TF2	TDF3	TF3
ana:r	a-n	50	917	1765	-	60	95	-	60	823	15	1890	-	-
a:m	a:-m	277	879	1294	-	252	277	-	20	831	NTS	-	-	-
ana:r	n-a:	314	906	1545	3020	182	90	314	56	819	70	1890	NTS	-
	a:-r	314	906	1545	3020	182	90	314	58	972	80	1853	24	28
pakɔ̣da:	ḍ-a:	264	808	1592	-	162	260	-	52	870	NTS	-	-	-
imli	i-m	80	478	2737	3207	38	28	28	31	431	35	2423	35	2957
dalija	l-ɾ	170	447	2627	3616	170	112	112	NTS	-	40	2592	45	3456
	i-j	170	447	2627	3616	170	112	112	40	322	NTS	-	NTS	-
i:k^h	i- k^h	315	416	2957	3725	195	240	195	90	541	105	2690	45	3364
mani:ʃ	n-i:	225	463	3055	3596	166	135	187	NTS	-	38	277	37	3471
	i:-ʃ	225	463	3055	3596	166	135	187	33	557	47	2880	NTS	-
mari:	r-i:	330	433	2941	3569	200	187	187	62	557	87	2376	62	3192
ullu	u-ɭ	77	510	918	-	78	70	-	NTS	-	42	1310	-	-
anu:dʒ	n-u:	129	729	1011	1639	52	91	100	18	651	22	902	NTS	-
	u:-dʒ	129	726	1011	1639	52	91	100	31	839	36	1121	NTS	-
tanu	n-u	25	855	-	-	25	-	-	NTS	-	-	-	-	-
u:n	u:-n	300	494	807	-	297	200	-	NTS	-	50	1035	-	-
tau:dʒj	a-u:	189	431	996	-	45	189	-	45	588	45	902	-	-
	u:-dʒ	189	431	996	-	45	189	-	NTS	-	45	1309	-	-
ullu	l-u	312	604	918	-	250	131	-	NTS	-	86	1608	-	-
ok^{h}li	o-k^h	210	416	965	-	210	172	-	NTS	-	30	886	-	-
manodʒ	n-o	260	447	933	-	206	78	-	NTS	-	88	1561	-	-
	o-dʒ	260	447	933	-	206	78	-	46	541	76	1420	-	-
karo	r-o	277	557	933	-	221	153	-	15	494	90	1686	-	-
ɔrat	ɔ-r	232	776	1278	2690	232	150	230	NTS	-	90	1655	NTS	-
pakɔ̣da	k-ɔ	240	651	1310	2659	240	130	100	NTS	-	NTS	-	NTS	-
	ɔ-ḍ	240	651	1310	2659	240	130	100	NTS	-	110	1561	130	1843
Iɔ	I-ɔ	352	620	1121	-	352	217	-	NTS	-	60	1479	NTS	-
ek	e-k	320	447	2910	3380	260	320	320	20	557	NTS	-	NTS	-
anek	n-e	292	478	2925	3553	292	182	194	NTS	-	52	2580	26	3364
	e-k	292	478	2925	3553	292	182	194	NTS	-	39	3145	58.5	3364
pake	k-e	318	447	2800	3145	310	310	310	NTS	-	NTS	-	NTS	-
ɤnek	ɤ-n	185	809	2298	3082	150	171	185	30	745	21	2172	NTS	-

Plosives													
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	VOT	CD	BD	ToD	AD	VD
kul	k-u	329	1145	4251	14	10	9	22	-	12	37	-	102
maka:n	a-k	674	1522	2745	9	14	14	-	128	20	141	-	60
	k-a:	831	1702	2651	NTS	18	42	-	128	20	141	-	292
tapak	a-k	706	1553	2839	NTS	NTS	10	-	-	15	15	-	179
kʰa:na:	kʰ-a:	1080	1584	2823	18	18	30	-	-	16	114	70	198
ʧarakʰa:	kʰ-a:	800	1615	2620	NTS	30	28	70	115	20	205	85	280
i:kʰ	i:-kʰ	651	3082	3410	NTS	25	50	-	-	10	310	173	135
ga:di	g-a:	698	1576	2517	20	108	70	28	-	16	44	-	350
magar	a-g	604	1420	-	30	NTS	NTS	-	28	16	54	-	200
	g-a	549	1647	2635	NTS	58	NTS	-	28	16	54	-	202
ʤag	a-g	635	1623	2502	17	NTS	10	-	-	-	47	-	190
gʰar	gʰ-a	760	1702	-	22	45	-	97	-	21	246	121	216
megʰa	e-gʰ	384	2894	3302	15	NTS	45	-	105	15	165	75	225
	gʰ-a:	886	1765	2643	-	60	60	-	105	15	165	75	315
megʰ	e-gʰ	510	2756	3365	NTS	45	NTS	-	-	20	241	115	282
ṭa:I	ṭ-a:	635	1780	2667	32	95	NTS	33	-	20	50	-	350
ṭama:ṭar	a:-ṭ	761	1765	2324	17	64	44	-	150	15	173	-	201
	ṭ-a	509	1892	2611	26	40	52	-	150	15	173	-	238
palaṭ	a-ṭ	714	2000	2690	NTS	19	20	-	-	20	-	-	97
ṭʰaṭʰera:	ṭʰ-a	1072	2031	2251	28	30	32	10	-	14	120	90	110
ṭʰaṭʰera:	a-ṭʰ	619	2188	2408	45	50	NTS	-	110	12	130	90	110
	ṭʰ-e	194	2627	3255	NTS	50	NTS	-	110	12	130	90	252
maṭʰ	a-ṭʰ	714	1812	2408	17	43	44	-	-	12	110	90	96
ḍa:I	ḍ-a:	564	1511	2180	35	108	144	96	-	10	175	-	304
paɳḍit	ḍ-e	500	2392	2110	NTS	NTS	14	-	44	5	50	-	114
ḍʰakkan	ḍʰ-a	98635	2118	3058	66	68	NTS	98	-	12	160	24	180
ta:I	t-a	508	1819	2722	20	50	NTS	15	-	10	25	-	255
bata:ʃa:	a-t	557	1969	2816	18	60	NTS	-	126	8	136	-	66
	t-a:	627	1922	-	40	12	-	-	126	8	136	-	170
kajamat	a-t	698	1689	-	34	32	-	-	-	-	-	-	82

word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	VOT	CD	BD	ToD	AD	VD
t^ha:li	t^h-a:	809	1529	2816	10	NTS	NTS	105	-	18	108	90	200
kat^{h}a:	a-t^h	672	1922	3004	21	40	52	-	134	14	200	62	62
	t^h-a:	1925	1725	2784	24	24	6	-	134	14	200	55	204
rat^h	a-t^h	682	1780	3161	36	40	66	-	-	12	150	32	126
dava:t	d-a	588	1717	3004	48	NTS	40	-	-	11	160	-	136
ma:dak	a:-d	714	1749	2784	44	88	97	144	92	9	116	-	230
	d-a	713	1843	2533	35	NTS	36	-	92	9	115	-	88
kad	a-d	620	1874	3003	19	NTS	18	-	-	-	-	-	103
d^{h}anija	d^h-a	651	1906	-	30	NTS	-	111	-	14	142	30	112
ad^{h}arm	a-d^h	494	1986	3090	30	20	25	-	87	15	112	25	83
	d^h-a	280	1968	3192	NTS	30	67	-	87	15	112	25	105
du:d^h	u:-d^h	494	1184	-	NTS	40	-	-	-	15	130	-	290
pal	p-a	745	1435	2627	NTS	57	30	10		12	30	-	142
kapaṭi	a-p	635	1325	2329	NTS	19	15	-	112	6	127	-	56
	p-a	478	1906	-	15	NTS	-	-	112	5	127	-	33
ta:p	a:-p	949	1576	2659	NTS	50	NTS	-	-	26	184	-	292
bal	b-a	608	1496	2604	20	36	14	62	-	60	250	-	130
kabɑḍḍi	a-b	510	1341	2627	NTS	17	11	-	102	7	107	-	56
	b-a	651	1498	2533	12	28	19	-	102	7	107	-	94
kab	a-b	557	1404	2753	38	30	NTS	-	-	14	142	-	108
b^{h}a:lu	b^h-a	808	1435	2722	NTS	NTS	NTS	76	-	30	120	76	210
kab^{h}i	a-b^h	599	1623	2690	15	20	NTS	-	90	5	138	49	90
	b^h-i	368	2502	3129	20	30	35	-	90	5	138	49	250
ʃub^h	u-b^h	478	1184	2910	22	40	NTS	-	-	5	276	49	105

Affricates										
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	BD	AD	FD
tʃal	tʃ-a	839	1906	-	42	45	-	-	-	103
katʃaḍa:	a-tʃ	620	2000	2989	35	52	26	5	-	99
	tʃ-a	557	1811	-	28	NTS	-	5	-	-
tʃammatʃ	a-tʃ	946	1984	3478	NTS	92	60	-	-	200
tʃʰatri	tʃʰ-a	643	2204	2941	NTS	26	NTS	6	42	75
matʃʰali	a-tʃʰ	713	2125	2753	24	75	22	-	28	112
kutʃʰ	u-tʃʰ	510	1827	3302	13	60	25	-	75	200
dʒal	dʒ-a	525	2188	2941	52	30	NTS	-	-	52
ka:dʒal	a:-dʒ	620	1790	2973	70	100	70	-	-	90
	dʒ- a	525	2157	3035	40	60	30	-	-	90
gadʒ	a-dʒ	556	2081	2874	15	55	52	-	-	210
dʒʰ arn a	dʒʰ-a	604	2031	3098	12	40	41	-	-	105
dʒʰ ndʒʰ ar	dʒʰ-a	542	1733	-	50	NTS	-	-	-	70
bodʒʰ	o-dʒ	572	1435	3835	NTS	70	87	-	87	183

Fricatives										
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	FD	VD
hal	h-a	839	1655	2721	NTS	30	15	103	103	90
va:han	a:-h	964	1325	3098	NTS	34	70	60	60	240
	h-a	745	1498	3129	40	50	NTS	60	60	129
vah	a-h	698	1420	2705	NTS	30	NTS	75	75	80
ʃalgam	ʃ-a	588	1718	2722	25	50	25	181	180	137
kiʃore	i-ʃ	290	2784	3098	12	NTS	NTS	180	165	75
	ʃ-o	431	1435	3078	NTS	75	NTS	180	165	330
avaka:ʃ	a:-ʃ	494	1937	2784	75	97	NTS	232	292	232
ʂaṭʰkoŋ	ʂ-a	620	1749	2941	NTS	45	NTS	117	117	90
kʊ̃ʂṭʰak	o-ʂ̃	305	1278	-	40	100	-	160	160	180
dʰanuʂ̃	u-ʂ̃	416	1764	-	37	42	-	270	270	135
sahan	s-a	604	2017	2972	40	NTS	NTS	168	168	100
fasal	a-s	604	1733	2769	25	22	NTS	140	140	98
	s-a	557	1874	2957	40	22	NTS	140	140	104
paḍos	o-s	274	1184	-	40	45	-	285	270	285
fal	f-a	682	1503	2502	NTS	67	45	157	145	105
safal	a-f	619	1906	2753	22	37	40	157	142	105
	f-a	588	1623	2627	15	52	45	157	142	120
sa:f	a:-f	651	1466	2753	30	NTS	NTS	240	290	372

Nasals									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
kaɳ	a-ɳ	760	1576	2110	37	NTS	30	97	172
ni:m	n-i:	557	2737	3502	84	86	56	148	260
mana:na:	a-n	714	1182	2941	17	68	24	97	144
	n-a	620	1718	-	25	40	-	97	200
mɔn	ɔ-n	557	1247	-	23	55	-	50	160
ma:n	a:- n	933	1216	3004	NTS	22	30	153	300
somava:r	o-m	541	871	-	NTS	58	-	90	170
	m-a	494	1122	-	20	NTS	-	90	110
kalam	a-m	734	1592	2816	22.5	45	15	172	90

Lateral									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
la:l	l-a	745	1874	2972	10	90	60	170	350
tʃalana	a-l	68	2031	2878	NTS	30	NTS	88	105
	l-a	525	1906	2784	21	22	15	88	75
la:l	a:-l	651	1906	2972	40	110	90	170	350

Trill									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
rel	r-e	431	2322	3035	NTS	92	NTS	134	300
keral	e-r	5255	2220	2941	NTS	67.5	NTS	28	220
	r-a:	525	1812	2816	NTS	30	45	28	150
ka:r	a:-r	635	1702	2502	34	76	34	90	400

Semivowel									
Word	Context	TF1	TF2	TF3	TDF1	TDF2	TDF3	ToD	VD
va:han	v-a:	500	1058	2596	37	37	37	75	195
avaka:ʃ	a-v	588	1435	2721	23	22	NTS	75	85
	v-a	494	1247	2784	15	NTS	27	75	75
na:v	a:-v	345	1310	2910	45	NTS	53	85	345

Table 3 : The segmental features (spectral and temporal) for the female subject.

Table 6 and 7 show the values of various parameters for fricatives and nasal continuants.

Parameter	Male		Female	
	Voiced	Unvoiced	Voiced	Unvoiced
ToD	34-82	55-175	75-85	60-270
FD	30-106	72-192	70-85	60-330

Table 6 : Range of values of temporal measures for the fricatives (msec)

It was noticed that the unvoiced fricatives were longer than the voiced fricatives.

Parameter	Male	Female
ToD	10-140	8-172

Table7: Range of values of the temporal measures for the nasal continuants (msec)

It was observed that the total duration was longer in the female subject than the male subject. Table 8 shows the total duration for trill, lateral and semi vowel

Parameter	Male			Female		
	Trill	Lateral	Semivowel	Trill	Lateral	Semivowel
ToD	28-39	67-125	34-106	28-90	88-170	75-85

Table 8 : Range of values of the temporal measures for the trill, lateral and semivowel(m sec)-

It was observed that the lateral /*l*/ was longer than the trill /r/ and the semivowel *Ivl.* Table 9 and 10 shows the F_1 F_2, F_3 and duration of vowels.

Vowel	F_1	F_2	F_3	Duration
/a/	870	1687	3371	43
/a:/	575	1753	2084	289
/i/	227	2298	2705	81
/i:/	394	2664	3042	225
/u/	515	1486	2674	127
/u:/	373	1116	2564	236
/o/	462	849	1686	200
/ ɔ/	530	1196	2125	234
/e/	368	2231	-	251
/æ/	456	1921	2439	101

Table 9 : Average values of the Fi, F2, F_3 and duration for the vowels of the •male subject...

Vowel	F_1	F_2	F_3	Duration
/a/	917	1765	-	50
/a:/	864	1477	3020	285
/i/	462	2682	3411	125
/i:/	437	2984	3630	290
/u/	698	964	-	152
/u:/	509	907	-	267
/o/	473	943	-	249
/ɔ/	682	1236	2675	274
/e/	452	1978	3359	310
/æ/	809	2298	3082	185

Table 10 : average values of the FI, F2, F3 and duration for the vowels of the female sjbject

Table 11 and 12 shows various values for phonemes

Plosives:

	VOT	CD	ToD	BD	AD	TF1	TF2	TF3
/k/	30	73	52	15	-	583	1136	2235
/k^{h}/	74	90	176	15.3	101	577	1586	2245
/g/	-63	66	86	10	-	546	1275	2271
/g^{h}	-29	35	74	17	60	486	1678	2436
/ṭ/	45	62	73.5	10	-	609	1574	2493
/ṭ^{h}/	94	110	103	13	53	533	1796	2192
/ḍ/	-68	68	124	4	-	540	1690	2292
/ḍ^{h}/	-29	119	108	9	90	620	1570	2533
/t/	33	80	88	10	-	588	1292	2392
/t^{h}/	60	60	129	10	59	623	1334	2580
/d/	-15	15	42	10	-	574	1404	2471
/d^{h}/	-46	46	123	10	67	478	1346	2660
/p/	10	112	119	7	-	565	1069	2162
/b/	-95	95	102	7	-	502	1051	2209
/b^{h}/	-31	31	152	7	114	510	1113	2249

Affricates:

	VOT	CD	BD	ToD	AD	FD	TF1	TF2	TF3
/t∫/	65	31	9	113	-	76	533	1675	2651
/t∫^{h}/	-	-	-	162	34	87	499	1507	2682
/dʒ/	-129	129	6	160	8	70	502	1854	2533
/ dʒ^{h}/	-59	109	-	183	50	73	572	1696	2664

Fricatives:

	ToD	FD	TF1	TF2	TF3
/h/	62	60	646	1404	2338
/ʃ/	125	129	427	1723	2288
/s̃/	120	112	515	1545	2138
/s/	97	141	474	1416	2434
/f/	164	160	562	1148	2354

Nasal Continuants :

	ToD	TF1	TF2	TF3
/n̰/	40	713	1372	1733
/n/	117.5	531	1206	2821
/m/	116	573	1163	2147

Lateral :

	ToD	TF1	TF2	TF3
/l/	104	630	1522	2486

Trill :

	ToD	TF1	TF2	TF3
/r/	34	514	1659	2483

Semivowel :

	ToD	TF1	TF2	TF3
/v/	74	483	949	2324

Table 11: Average values for VOT, total duration, burst duration, aspiration duration, frication duration, and terminal frequencies (TFl, TF2, TF3) for the phonemes spoken by male subject .

Plosive

	VOT	CD	ToD	BD	AD	TF1	TF2	TF3
/k/	22	128	144	16	-	645	1481	3121.5
/k^h/	70	115	210	15	110	844	-	2094
/g/	-28	28	44	16	-	622	1567	2551
/g^h/	-97	105	217	19	104	635	-	2278
/ṭ/	33	150	111	12	-	654	1859	2573
/$ṭ^h$/	10	110	120	13	90	725	2164	2580
/ḍ/	-96	44	112	75	-	532	2101	2595
/$ḍ^h$/	-98	-	160	12	24	635	2118	3053
/t/	5	126	136	9	-	618	1850	2769
/t^h/	105	134	154	16	61	1022	1739	2941
/d/	-144	92	130	10	-	659	1796	2831
/d^h/	-111	87	128	15	28	480	1761	3145
/p/	10	112	114	15	-	102	1560	2538
/b/	-152	102	166	9	-	581	1435	2629
/b^h/	-76	90	178	11	58	563	1686	2863

Affricates

	VOT	CD	BD	ToD	AD	FD	TF1	TF2	TF3
/tʃ/	-	89	260	5	-	134	740	1925	3133.5
/$tʃ^h$/	45	52	189	6	48	129	612	2052	2999
/dʒ/	-45	70	212	-	-	117	556	2057	2956
/$dʒ^h$/	-105	30	183	-	87	119	572	1733	3467

Fricatives

	ToD	FD	TF1	TF2	TF3
/h/	80	80	811	1474	2913
/ʃ/	198	212	451	1968	2868
/s̃/	182	182	447	1597	2941
/s/	198	192	510	1710	2899
/f/	185	176	635	1624	2659

Nasal continuants

	ToD	TF1	TF2	TF3
/ɳ/	97	760	1576	2110
/n/	98	612	1721	3221
/m/	138	675	1200	2910

Lateral

	ToD	TF1	TF2	TF3
/l/	143	650	1929	2901

Trill

	ToD	TF1	TF2	TF3
/r/	84	529	2014	2823

Semivowel

	ToD	TF1	TF2	TF3
/v/	78	482	1262	2753

Table 12 : Average values for voice onset time, total duration, burs, duration, aspiration duration, frication duration and terminal frequencies (TF1, TF2, TF3 forall the phonemes spoken by female subject.

How to use the data base ?

The phoneme to be synthesized should be selected. The environment (preceding and the following phoneme) in which the phoneme has to be synthesized and whether the synthesis is for male or female subject has to be noted. The parameters which have been measured for that phoneme should be selected depending on the kind of synthesizer. Using this data base vowels, plosives, affricates, fricatives, nasals, lateral, trill and semivowel of Hindi can be synthesized.

The Klatt synthesizer is basically prototyped for male vocal tract of 17.5 cm. However the male voice can be converted to female voice by using the scale factor 1:1.2 as obtained from this database.

CHAPTER - V

SUMMARY AND CONCLUSIONS

Database is needed for the purpose of speech synthesis and speech recognition. Various databases are available in English and European languages. In India, the work in this area is scanty. As the database developed in one language cannot be effectively used for other language due to language specific segmental and suprasegmental characteristic, a need was felt to develop an acoustic database in Hindi. The aim of the study was to develop dynamic acoustic segmental database in Hindi language.

One hundred and five meaningful Hindi words with ten vowels, fifteen plosives and four affricates, three nasal continuants, five fricatives, one trill, one lateral and one semivowel in initial, medial and final position constituted the material. These words as uttered by a twenty two year old male and twenty two year old female native Hindi speakers were audio recorded. All these words were digitized using a 12 bit A/D converter at 8000 Hz sampling frequency and were stored in the computer memory. Using the wave form and spectrogram of the SSL software (Voice and Speech systems, B'lore, India) two spectral and nine temporal parameters were extracted. The spectral parameters constituted of formant frequency (F_1, F_2, F_3) and terminal Frequency (TF1, TF2, TF3) and the temporal parameters were vowel duration, total duration, voice onset time, closure duration, burst duration, aspiration duration, affrication duration, frication duration and transition duration and formant steady state duration.

These parameters formed the acoustic database. The acoustic data base provides spectral and temporal parameters of each speech sound which can be used for further synthesis.

The acoustic database will have the following implications :

1. It provides acoustic data on speech communication to be incorporated for speech recognition.

2. It provides information on normal segmental features which would lead to therapeutic model for the handicapped.

3. The information helps in developing aids/synthesizer for the non verbally handicapped.

4. It also aids in better understanding of speech production system.

5. The data aids in automatic speech recognition.

It is suggested that the database for suprasegmental features for Hindi be made.

CHAPTER - VI

BIBILOGRAPHY

Agrawal, S.S. (1992). Acoustic - Phonetics for developing voice I/O systems. *Proceedings of Tutorial on Speech Technology*, IIT Madras. Dec. 6-8, 93-103.

Bernstein, J. Entropic Latino40 Speech Database. Available: www.cs.columbia.edu/~bgrundy/latino40.htra

Blomberg, M., & Elinius, K. (1985). Automatic time alignment of speech with a phonetic transcription. *STL-QPSRA.* 37-45.

Carlson, R., & **Granstrom.** B. (1985). Rule controlled Database search. *STL-QPSR.* 4. 29-33.

Calson, R., Granstrom, B., & **Hunnicutt, S. (1982).** A multilanguage text-to-speech module. *Proc. ICASSP*. 82(3). 1604-1607.

Grundy, B., Bernstein, J., Rosenfild, E., **Najmi, A.,** & **Mankoski, P.** Available : Entropic Latino40 Speech Database. Available: www.cs.columbia.edu/~bgrundy/latino40.htm.

Nataraj, N., **Savithri, S., Venkatesh, C.S., & Rohini. H. (1990).** Clinical application of speech oriented learning systems for Handicapped : An aid for communication with hearing impaired persons (SOLSH). Unpublished Project report, All India Institute of Speech and Hearing. India.

Savithri, S.R. (1994). Database for Kannada speech sounds. Unpublished Project report, All India Institute of Speech and Hearing. India.

Witten, I.H. (1982). Principles of computer speech London : Academic press.

The following speech database were accessed from the Internet.

Estonian speech database
http://www.esis.ee/ist2000/IT/ioc/langdb.html.

Oxford Acoustic Phonetic database
http://svr-www.eng.cam.ac .uk/comp. speech/Section 1 /Data/oxford.html.

EUROM - A multilingual European speech database
www. icp. grenet. fr/Relator/multiling/eurom 1 .html.

The EMU speech database system
http://www.deall.ohio-state.edu/grads//ee. 1460/c694/Emu2.html

Readme for the ATTS - 2 MADCOW speech Corpus
http://www.ldc.upenn.edu/catalo g/readme_files/atis2.readme.html.

Entropic Latino speech database by Bernstein
www.cs.columbia.edu/~bgrundy/latino40.htm.

The IRIS speech database
http://www.ling.su.se/staff/olle/1987d.html.

Castro speech database
http://lanic.uteras.edu/la/cb/cuba/castro.html.

Linguistic database consortium

http://www. idc. upenn. edu/.

Japanese speech and text corpora

www.milab.is.tsukuba.ac.jp/corpus/corpora-c.html.

BUPT spoken digit database in Chinese University

http://www.speech.cs.cmu.edu/comp. speech/Sectionl /Data/bupt.html.

Victoria phonetic database

http://www.speech.cs.cmu.edu/comp.speech/Sectionl/Data/victoria.html.

Australian english telephone speech database

www.callbase.com

Annion speech database

www.google.com.

King speech database

http://www.ldc.upenn.edu/Catalog/readme_files/king.readme.html.

Linguistic database consortium

http://www.idc.upenn.edu/.

Japanese speech and text corpora

www.milab.is.tsukuba.ac.jp/corpus/corpora-c.html.

BUPT spoken digit database in Chinese University

http://www.speech.cs.cmu.edu/comp.speech/Sectionl/Data/bupt.html.

Victoria phonetic database

http://www.speech.cs.cmu.edu/comp. speech/Sectionl /Data/victoria.html.

Australian english telephone speech database

www.callbase.com

Annion speech database

www.google.com.

King speech database

http://www.ldc.upenaedu/Catalog/readme_files/king.reajne.html.