

# **CREATING AN UNLIMITED VOICE RESPONSE SYSTEM IN HINDI**

**By**

**ARCHANA BALYAN**



**UNIVERSITY SCHOOL OF ENGINEERING AND TECHNOLOGY**

**GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY**

**NEW DELHI**

**2016**

2016      ARCHANA BALYAN      Ph.D

# **CREATING AN UNLIMITED VOICE RESPONSE SYSTEM IN HINDI**

**BY**

**ARCHANA BALYAN**

**UNIVERSITY SCHOOL OF ENGINEERING AND  
TECHNOLOGY**

Submitted

In fulfilment of the requirement of the degree of Doctor of  
Philosophy

to the



**Guru Gobind Singh Indraprastha University**

**Dwarka, Delhi**

# CERTIFICATE

The work embodied in the thesis entitled “**Creating an Unlimited Voice Response in Hindi**” is original and has been carried out by the author and it has not been submitted in full or in part for any other diploma or degree of this or any other university.

**Dr. Amita Dev**

(Supervisor)

Bhai Parmanand Institute of Business Studies

Government of Delhi, Delhi

**Dr. S. S. Agrawal**

(Co-Supervisor)

Director General

KIIT College of Engineering, Gurgaon &

Advisor, CDAC Noida

# ACKNOWLEDGEMENTS

Completion of this doctoral thesis was possible with the support of several people. I would like to express my sincere gratitude to all of them. First of all, I am extremely grateful to my research guide, **Dr. Amita Dev**, for her valuable guidance, scholarly inputs and consistent encouragement, I received throughout the research work.

I would like to thank my co-supervisor **Dr. S. S. Agrawal**, for his help and encouragement throughout my time as a PhD scholar. His expert advice and detailed knowledge of the field was invaluable. I thank him for all his guidance and help he has extended over the days.

This feat was possible only because of the unconditional support provided by **Dr. B.V.R. Reddy, Dean, USET, Guru Gobind Singh Indraprastha University, Delhi**.

I owe a lot to my Parents, who encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true.

I am very much indebted to my family, husband Mr. Sanjay Mann, daughter Himanshi Mann and Son Ayush Mann, who supported me in every possible way to see the completion of this work.

Above all, I owe it all to Almighty God for granting me the wisdom, health and strength to undertake this research task and enabling me to its completion.

Signature of Candidate

# ABSTRACT

*Keywords: Text-to-speech synthesis, Concatenative speech synthesis, Unit selection based speech synthesis, Hidden Markov Model, Speech segmentation, Phoneme, Syllable*

Speech is the most natural, convenient, and useful means of communication. Moreover, speech can convey other information such as emotion, attitude, and speaker individuality. Therefore, it is important to realize a man-machine interface to facilitate communication between people and computers.

The objective of the research is to develop high quality speech synthesizer, primarily in Hindi language based on corpus speech synthesis using concatenation technique. This technique is used as the naturalness obtained is highest as compared to other techniques. A major problem associated with implementation of speech synthesizers for Hindi is lack of availability of database. For development of efficient speech synthesizer, it is mandatory to have phonetically rich sentences, consistent recording conditions and availability of properly annotated database tagged with prosodic features. Literature survey has revealed availability of such database for English and other European languages but they are not available in Indian context. Hence, it poses a major challenge to develop synthesizer for Hindi language. A sincere effort has been made to address this crucial and time consuming issue i.e. Development of an algorithm for automatically creating appropriate synthesis units.

Since larger domain synthesizer cannot be created using syllable as the basic unit the phoneme has been chosen as the basic unit. In the proposed technique, the base line Hidden Markov Model (HMM) has been used for segmentation of speech signals. It is applied on single speaker

segmentation task, using Hindi speech database. The automatic phoneme segmentation framework evolved imitates the human phoneme segmentation process. A set of 42 Hindi phonemes were chosen for the segmentation experiment, wherein, Continuous Density Hidden Markov Model (CDHMM) with a mixture of Gaussian distribution were used. The left –to-right topology with no skip states has been selected as it is effective in speech recognition due to its consistency with the natural way of articulating the spoken words. This system accepts speech utterances along with their orthographic “transcriptions” and generates segmentation information of the speech. The system was trained using numerous sentences relevant to task specific database for travel domain. In a novel manner, HMM has been used for phoneme segmentation and labelling. The modelling of HMMs has been implemented using Microsoft Visual Studio 2005 (C++) and the system is designed to work on Windows operating system. The segmentation was performed initially using feature vector comprising of 12 MFCCs. Then, the delta and delta-delta coefficients of the 12 static parameters above were appended resulting in 36 dimension feature vector. This improved the segmentation accuracy from 28.69% to 73.68% in (0-20msec) tolerance range. The achieved segmentation accuracy of 73.68 % is sufficient for producing high quality synthesized output. The system was validated against a few manually segmented speech utterances. The evaluation of the experiments was generated using combination of 2 Gaussians mixtures and 5 HMM states. The proposed automated segmentation method assured less error in concatenation points.

The public address system installed in Metro rail coaches provides routine information, however, special announcements related to certain technical snags, accidents, or any such eventuality requires certain specific approaches to handle the communication between the Metro Rail operators and passengers. An effort has been made to address the above problem by proposing a

TTS system for use in the transport system, specifically, in Metro Rail by creating database relevant to metro rail passenger information system. The front-end of the TTS has been developed using concatenation approach for limited domain. Unit proposed for our speech corpus is phoneme.

The objective evaluation using spectrogram reading has been performed. Acoustic features such as pitch and duration of the generated synthetic speech have been evaluated. It was observed that the formant frequencies of phonemes such as /स/ (fricative), /व/ (approximants), /म/ (nasal), /न/ (nasal) and , /ह/ (fricatives) remains consistent and similar values occur in synthesized speech.

The percentage change in formant frequencies with respect to the original speech was is 1.4%, 7.04%, 4.97%, 17.45% and 5.88% respectively. However, phonemes such as /झ/ showed substantial change of 70.3681%. The reason attributed for this change is due to articulation effect of the preceding phoneme on the target phoneme which results in “problem phones” in the synthesized speech. However change observed in the duration of the phoneme in the synthesized were of acceptable limits.

During synthesis, selection of units is based on phonetic context of its (previous or succeeding) unit. The proposed approach minimizes co-articulation effects. The subjective evaluation is done by conducting perception tests. The intelligibility was found to be 79.36%. The naturalness of the synthesized speech demonstrates the appropriateness of the approach adopted for building TTS.



# LIST OF PUBLICATIONS

## International Journals

- [1] Archana Balyan , S. S. Agrawal, Amita Dev, “*Automatic phonetic segmentation of Hindi speech using Hidden Markov Model*”, “Journal of Artificial Intelligence & Society” of Springer Verlag, London, United Kingdom, DOI: 10.1007/s00146-012-0386-2, May 2012, Vol. 2, Issue 27, p-p 543-549 ,2012
- [2] Archana Balyan , S. S. Agrawal, Amita Dev, “*A Hindi speech synthesizer using phoneme concatenation approach* ”, International Journal of Graphics and Image processing , ISSN number : 2249 –5452, Vol. 2, Issue 1, p-p 12-17, 2012.
- [3] Archana Balyan, S. S. Agrawal, Amita Dev, “*Speech Synthesis: A Review*”, International Journal of Engineering research and Technology (IJERT), ISSN number: 2278-0181, Vol. 2, Issue 6, June 2013.

## Proceedings of International conference

- [4] Archana Balyan, S. S. Agrawal, Amita Dev, “*Building Syllable Dominated Hindi Speech Corpora for Metro Rail Information System*” O-COCOSDA-2008, p-p 135-140, Kyoto, Japan, Nov 25-27, 2008.
- [5] Archana Balyan, S. S. Agrawal, Amita Dev, “*Development of database for speech synthesizer in Hindi language using Festvox*”, Joint Conference of the Association for Computational Linguistics and Natural Language Processing of the Asian Federation of Natural Language Processing, ISBN 978-1-932432-61-9 / 1-932432-61-2, Suntec city, Singapore, Aug 2-5, 2009.

- [6] Archana Balyan, S. S. Agrawal, Amita Dev, “*Phonetic Segmentation based on HMM of Hindi Speech*”, O-COCOSDA 2010, p-p 23-25, Nov. 2010, Nepal. Available at : [www.desceco.org/o-COCOSDA 2010/proceedings /paper 51.pdf](http://www.desceco.org/o-COCOSDA 2010/proceedings /paper 51.pdf)
- [7] Archana Balyan, S. S. Agrawal, Amita Dev, “*Development of Hindi Speech Synthesizer for Metro Rail Information System*”, 2<sup>nd</sup> International Conference of Electrical and Control Engineering, *IEEE conference publication*, DOI: [10.1109/ICECENG.2011.6057268](https://doi.org/10.1109/ICECENG.2011.6057268) , p-p 2166-2170, Yichang, China, September 16-18, 2011.
- [8] Archana Balyan, S. S. Agrawal, Amita Dev, “*Development of Female Voice Speech Synthesizer for Hindi Language*”, O-COCOSDA 2011, Oct 26-28, Hsinchu, Taiwan, 2011.
- [9] Archana Balyan, S. S. Agrawal, Amita Dev “*Development of Database for Synthesizer in Hindi Language*”, “The VIII<sup>th</sup> International conference on Electrical, Computer, Electronics and Communication Engineering”, organized by World Academy of Science Engineering and Technology, ISSN 2010-376X, p-p 582-586, Issue 61, p-p 582-586, Zurich, Switzerland , 2012.
- [10] Archana Balyan, S.S. Agrawal, Amita Dev, “*A Unit Concatenation Based Speech Synthesizer as Applied to Hindi Language*”, The 22<sup>nd</sup> Annual International Symposium Frontiers of Research in Speech and Music-2012, p-p 157-161, India,January16-18, 2012.
- [11] Archana Balyan, S.S. Agrawal, Amita Dev, “*Semi- Automatic Syllable- like Segmentation for Hindi*”, Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation, *IEEE conference publications*, DOI:[10.1109/ICSDA.2013.6709854](https://doi.org/10.1109/ICSDA.2013.6709854) , p-p 1-4 , India ,November 25-27, 2013.

# LIST OF TABLES

2	Comparison between speech synthesis techniques	29
3	The consonant and vowel classification in Hindi	35
3.1	List of Dutch stops	42
3.1(a)	Voice onset time in msec (Dutch single speaker)	42
3.2	List of Tamil stops	42
3.2(a)	Voice onset time VOT in msec (Tamil single speaker)	42
3.3	List of English stops	43
3.3 (a)	Voice onset time VOT in msec (English four speakers)	43
3.4	List of Armenian phoneme	44
3.4(a)	Voice onset time in msec: Eastern Armenian (1 speaker)	44
3.5	List of Thai phonemes	44
3.5(a)	Voice onset time in msec: Thai (1 speaker)	45
3.6	List of Korean phonemes	45
3.6(a)	Voice onset time in msec: Koren (1 speaker)	45
3.7	List of Hindi phonemes	46
3.7(a)	Categories of phoneme in Hindi	46
3.7(b)	Voice onset time in msec: Hindi (I speaker)	47
3.8	List of Marathi phonemes	47
3.8(a)	Voice onset time in msec: Marathi (1 speaker)	48
3.9	Reported VOT values for velar stops in English compared to reported VOT values for velar stops in Hindi	49

3.10	Confusion matrix (in percent) for Hindi stops after the vowel /आ/ (ohala&ohala2001)	53
3.11	Confusion matrix (in percent) for Hindi stops after the vowel /ऊ/ (ohala&ohala 2001)	53
3.12	Confusion matrix (in percent) for Hindi stops after the vowel /ई/ (ohala&ohala 2001)	54
3.13	Comparison of Burst frequencies of Dental and Retroflex sounds in /a/ /आ/ context	58
3.14	Comparison of Burst frequencies of Dental and Retroflex sounds in /i/ /ई/ context	58
3.15	Comparison of Burst frequencies of Dental and Retroflex sounds in /u/ /ऊ/ context	58
3.16	Comparison of Burst frequencies of Dental and Retroflex sounds in /e/ /ए/ context	59
3.17	Comparison of Burst frequencies of Dental and Retroflex sounds in /o/ /औ/ context	59
4	Devanāgarī CV aksharas	61
4.1	Anusvāra forms in Hindi	62
4.2	Devanāgarī vowels in their full and diacritic forms	62
6	Various types of syllables in Hindi	75
6.1	Analysis of the recorded speech corpus	76
6.2	Few examples of high frequency words	77
6.3	Syllables as the basic speech unit, with some corresponding examples	78
6.4	Distribution of various syllables in database	78

6.5	An example illustrating the output of text syllabification	94
6.9	Overall segmentation performance of continuous speech	98
7	Statistics of the recorded speech	112
7.1	Example showing list of phonemes	116
7.2	Example showing phoneme boundaries obtained for <Wfile002>	120
7.3	The segmentation accuracy results	121
7.4	Category-wise phoneme segmentation results	122
7.5	Comparison of segmentation accuracy between Hindi and Croatian Language	123
8	Analysis for phonetic richness of sentences	126
8.1	Number of Tokens generated for News Bulletin	130
8.2	Number of Tokens generated for Travel Domain	131
8.3	Formant frequencies of the synthesized speech signal (upper half) and original signal	137
8.4	Percentage change in formant frequencies	138
8.5	MOS for overall quality of speech	140

# LIST OF FIGURES

1.0	Main organs of human body involved in production of sounds	2
1.1	Closed glottis	3
1.2	Open glottis	3
1.3	The Larynx	3
1.4	The oral and nasal cavity	4
1.5	Midsagittal view of the vocal tract in producing Hindi vowels	7
1.5 (a)	आ /a/ top- left	7
1.5 (b)	ई /i/ lower- left	7
1.5 (c)	ऊ /u/ top- right	7
1.5 (d)	/अ/ ँ lower- right	7
1.6	A view of the outer, middle and inner ear	9
1.7	Principle of speech communication	11
1.8	Spectral representation of s(n)	15
1.9	LPC analysis model	15
1.10	Block diagram of extraction of feature vector (MFCC)	18
2	A simple functional block diagram of TTS	23
3	On the left, mean durations of the 8 vowels before all consonants; on the right, before the 10 stops/affricates differing in voicing and the overall mean duration before the same 10 phonemes	37
3.1	Mean durations of / i u a / /ई ऊ अ/ following (from left to right) voiced, voiceless, aspirated, Voiced Aspirated and breathy voiced stops; in the case of latter two categories, measurements are presented	38

in two ways: with the distinctive release (R) and without it. Dashed line A estimates the vowels intrinsic durations, C underestimates them, and B gives the true duration

3.2	Mean durations of the first vowel in the indicated minimal pairs which contain medial consonants differing in aspiration	39
3.3	Mean durations of the medial consonant, geminate or singleton, in the indicated minimal pairs	40
3.4	Spectra of phoneme / t. / [ɽ]	55
3.5	Spectra of phoneme /t/ [t̪]	56
5.	Architecture of the TTS system	72
6.	Labelling of word, syllable and phoneme occurring in speech waveform	75
6.1	Pie-chart representation of distribution of syllables in database	79
6.2	Block diagram for creating Annotated speech corpora	79
6.3	Example of Annotation of sentence using Speech Filing System	80
6.4	Resolving power of group delay spectrum: z-plane, magnitude spectrum and group delay spectrum (I) a pole inside the unit circle at $(0.8, \pi/8)$ (II) a pole inside the unit circle at $(0.8, \pi/4)$ and (III) a pole at $(0.8, \pi/8)$ and another pole at $(0.8, \pi/4)$ , inside the unit circle	86
6.5	Group delay property of different types of signals: the z-plane, the	87

	magnitude spectrum, the phase spectrum, and the group delay spectrum for (I) minimum phase, (II) non-minimum phase—type (1) and (III) non-minimum phase—type (2) systems	
6.6	Block diagram showing steps involved in finding syllable boundaries	92
6.7	Manually segmented speech signal at syllable level using PRAAT	95
6.8	Screen shot of Group - delay based segmentation in the order (top to bottom); speech signal, syllables in the corresponding text and group-delay of speech signal	96
6.9	Screen shot of labelled text file	97
7	Typical HMM structure	101
7 (a)	3 state ergodic model	101
7 (b)	3 state left-to-right Model	101
7.1	Output distributions	103
7.1 (a)	Gaussian PDF	103
7.1 (b)	Multi Mixture PDF	103
7.1 (c)	Multi Stream PDF	103
7.2	Extraction of input vector frames by use of overlapping window functions on speech signals	109
7.3	Block diagram of base- line HMM model used	112



7.4	Flow chart of implemented phoneme parser	115
7.5	Example of HMM file generated for phoneme occurring in database	117
7.6	An example showing annotation of speech signal using PRAAT	118
7.7	Screen shot of HMM for phoneme	119
8	Block diagram of TTS Screen shot of front-end of the TTS System developed	127
8.1	Screen shot of database generated for phoneme ' <i>a. text</i>	129
8.2	An illustrative example of selection of phoneme 's' in a word	132
8.3	Front end of Speech synthesizer	133
8.4	Screenshot of the GUI of the TTS	135
8.5 (a)	Sound wave and its spectrogram, pitch, intensity and duration of the original speech for text 001	136
8.5 (b)	Sound wave and its spectrogram, pitch, intensity and duration of the synthesized speech for text 001	136
8.6	Variation of Formant frequencies of phonemes (synthesized versus original)	139

# ACRONYMS

<b>RAVI</b>	Reading Aids for Visually Impaired
<b>LPC</b>	Linear Predictive Coding
<b>PARCOR</b>	Partial Coefficient of Reflection
<b>BM</b>	Basal Membrane
<b>MFCC</b>	Mel Frequency Cepstral Coefficients
<b>DCT</b>	Discrete Cosine Transform
<b>ASR</b>	Automatic Speech Recognition
<b>PCA</b>	Principle Components Analysis
<b>NLP</b>	Natural Language Processing Module
<b>DSP</b>	Digital Signal Processing
<b>TDIL</b>	Technology Development for Indian Languages
<b>TIFR</b>	Tata Institute of Fundamental research
<b>TTS</b>	Text to Speech
<b>HMM</b>	Hidden Markov Models
<b>POA</b>	Place of Articulation
<b>MOA</b>	Manner of Articulation
<b>CDAC</b>	Centre for Development of Advanced Computing
<b>IIT</b>	Indian Institute of Information Technology
<b>CEERI</b>	Central Research Electronics Institute
<b>SFS</b>	Speech Filing System
<b>HTK</b>	Hidden Markov Model Kit
<b>ISI</b>	Indian Statistical Institute

<b>FFT</b>	Fast Fourier Transform
<b>CD-HMM</b>	Continuous Distribution HMM
<b>GMM</b>	Gaussian Mixture Model
<b>VOT</b>	Voice Onset Time
<b>EM</b>	Expectation Maximization
<b>DP</b>	Dynamic Programming
<b>ML</b>	Maximum Likelihood
<b>PDF</b>	Probability Density Function
<b>APS</b>	Automatic Phonetic Segmentation
<b>INSROT</b>	Indian Script Roman Transliteration
<b>ASCII</b>	American Standard Code for International Information
<b>ZCR</b>	Zero Crossing Rate
<b>CMN</b>	Cepstral Mean Normalization
<b>VUAs</b>	Voiced Unaspirated
<b>VAs</b>	Voiced Aspirated
<b>UVAs</b>	Voiced Unaspirated
<b>UVUAs</b>	Unvoiced Unaspirated
<b>UTF-8</b>	8-bit Unicode Transformation format
<b>MOS</b>	Mean Opinion Score
<b>IPA</b>	International Phonetic Alphabet

# ORGANIZATION OF THESIS

**In chapter 1**, the basic mechanism of human production and perception mechanism, and speech analysis techniques are described.

**In chapter 2**, a detailed introduction to various existing TTS synthesis techniques and in particular, the system architecture of the concatenative speech synthesis from the perspectives of developing synthesizers which can generate high quality continuous and unlimited synthetic speech is discussed.

**In chapter 3**, study of general phonetic and signal properties of different sounds in Hindi language and get a feel of the time domain, frequency domain and time –frequency representation of sounds description of specific phonetic and signal properties of different sounds in Hindi language for speech technology is undertaken and the results are analyzed.

**In chapter 4**, study of general linguistic and phonological aspects of Hindi language is presented.

**In chapter 5**, the plan of thesis specifying the objectives and methodology adopted for implementation is discussed.

**In Chapter 6**, emphasis is on design and development of syllable dominated database for development of TTS. This chapter outlines semi-automatic generation of syllable speech unit based on the group delay segmentation algorithm. The performance of the algorithm has been analyzed and compared against manually segmented speech.

**In Chapter 7**, the issue of creating database by automatic segmentation and labelling of speech units is addressed. An automatic phoneme segmentation technique using Hidden Markov models for continuous speech is proposed and implemented.

**In chapter 8**, details of development and implementation of phoneme based TTS based on unit selection is described.

**In chapter 9**, analysis of results obtained is presented and future scope of the work is discussed.

# CONTENTS

Thesis Certificate	(ii)
Acknowledgements	(iii)
Abstract	(iv)
List of Publications	(vii)
List of Tables	(ix)
List of Figures	(xii)
Acronyms	(xiv)
Organization of Thesis	(xviii)
<b>Chapter 1: Introduction</b>	
1.1 Text-to-speech synthesis	1
1.2 Human speech communication – Production, Perception and Analysis	1
1.3 Process of speech production	2
1.3.1 Initiation and oro-nasal process	3
1.3.2. Articulation	4
1.3.2.1 Relation between the geometry of the vocal tract and the resonance structure of vowels	5
1.3.3 Phonation	7
1.4 Speech Perception Mechanisms	9
1.4.1 Hearing Organs	9
1.5 Relationship between Speech Production and Perception Mechanisms	11

1.5.1 The Speech Chain	11
1.5.2. Sound Production	12
1.5.3. Sound Perception	12
1.6 Introduction to speech analysis	13
1.6.1 Short term windows analysis	13
1.6.2 Short- term Average Zero –Crossing Rate	14
1.7 Speech Analysis Methods Based on Production Mechanisms	14
1.7.1 Linear predictive coding	14
1.7.2 Linear predictive coding analysis	16
1.7.2.1 Durbins Algorithm	16
1.7.2.2 LPC Parameter conversion to Cepstral Coefficients	17
1.8 Speech Analysis Methods Based on Perception Mechanisms	17
1.8.1 Filter Bank Analysis	17
1.9 Auditory-Motivated Speech Analysis	18
1.9.1. Mel Frequency Cepstral Coefficients	18
1.10 Conclusions	22
<b>Chapter 2: Speech synthesis: A Review</b>	
2.1 Introduction	23
2.2. Recent techniques of speech synthesis	24
2.2.1 Articulatory synthesis	24
2.2.2 Formant speech synthesis	24
2.2.3 Concatenative synthesis	25
2.2.3.1 Diphone concatenation synthesis	25
2.2.3.2 Segmentation and labelling trends	25

2.3.3.3 Corpus –based speech synthesis	27
2.3 Development of TTS systems at International level	29
2.4 Speech synthesis and Development in Indian Scenario	30
2.5 Human-machine communication	31
2.6 Conclusions	32
<b>Chapter 3: Specific Features of Hindi Sounds</b>	
3.1 Introduction	33
3.2 Specific features and Acoustics characteristics of speech sound	33
3.2.1 Specific features of Hindi speech sound	33
3.2.2 Acoustic characteristics of Hindi speech	34
3.3 Study of Voicing of Initial stops of different languages	41
3.3.1 Comparison of distribution of stop categories between Hindi and English	48
3.3.2 Analysis of measurement of VOT of different languages	49
3.4 Study of Acoustic cues and perceptual properties of retroflex consonants	51
3.4.1 Cues of retroflex consonants	51
3.5 Results and analysis	52
3.6 Spectral representation of dental and retroflex consonants	55
3.7 Comparative analysis of retroflex and dental CV Syllables	57
3.8 Results and analysis	57
3.9 Conclusions	59
<b>Chapter 4: Study of Linguistic and Phonological Aspects of Hindi</b>	
4.1 Introduction	61
4.2 Phonological Aspects of the Devanāgarī scripts	61



4.2.1 Two consonant clusters	63
4.3 Hindi Phonology and Orthography	64
4.3.1 Schwa deletion in Hindi	64
4.4 Conclusions	65
<b>Chapter 5: Speech Synthesis for Hindi</b>	
5.1 Introduction	66
5.1.1 Improvement of naturalness of synthetic speech	67
5.2 Previous works	68
5.3 Motivation and Objectives	69
5.4 Methodology adopted	70
5.4.1 Phone Definition	70
5.4.2 Construction of Acoustic Inventory	71
5.5 Designing of TTS	72
5.6 Major Contributions of Thesis	73
<b>Chapter 6: Creation of Syllable Dominated speech Database</b>	
6.1 Introduction	74
6.2 Hindi Syllable types and their Distribution	74
6.3 Database Design and Collection Methodology	76
6.3.1 Selection of Sentences	76
6.3.2 Recording of Speech corpora	79
6.3.3 Annotation of speech signal	79
6.3.4 Structural storage of speech corpora	81
6.4 Results	82
6.5 Semi –Automatic Segmentation of speech for syllable Units	83

6.6 Segmentation using Minimum Phase Group Delay Function	83
6.6.1 Group Delay Function and its Properties	83
6.6.2 Minimum Phase Signal	83
6.6.3 Properties of the Group Delay Function	84
6.6.4 Properties of Minimum Phase Group Delay Function	86
6.7 The root Cepstrum Approach to Segment Continuous Speech	88
6.7.1 Derivation of a Minimum Phase Signal from the Magnitude Spectrum	89
6.7.2 The Minimum Phase Property of the Root Cepstrum	89
6.7.3 Minimum Phase Group Delay based Speech Segmentation Algorithm	91
6.8 Text Segmentation Procedure	93
6.9 Database Used	95
6.10 Results and Analysis	95
6.11 Conclusions	98
<b>Chapter 7: Automatic Phonetic Segmentation using Hidden Markov Model</b>	
7.1 Introduction	100
7.2 Framework of the Hidden Markov Model	101
7.2.1 Output Probability Specification	102
7.2.2 Probability Evaluation	103
7.2.3 Optimal State Sequence	105
7.2.4 Parameter Estimation	107
7.2.5 Auxiliary Function Q	107
7.3 HMMs as Acoustic models	108
7.3.1 Speech Input for HMM Systems	109
7.3.2 Recognition Units	110

7.3.3 Training	111
7.4 Preparation of Speech Corpus	111
7.5 Speech Segmentation using HMM	112
7.5.1 Feature Extraction Process	113
7.5.2 HMM Structure and Estimation of HMM Parameters	113
7.5.3 Text Transcription	114
7.5.4 The Phoneme Parser	114
7.5.5 The HMM Models of Phonemes	116
7.5.6 Phonetic Labelling of Recording	117
7.6 Results and Analysis	119
7.7 Conclusions	123
<b>Chapter 8: Development and Implementation of TTS</b>	
8.1 Introduction	124
8.2 Methodology of Construction of TTS	124
8.3 Speech Database Design	125
8.3.1 Corpus Development for Travel Domain	125
8.3.2 Corpus Development for News Domain	125
8.4 Implementation of TTS	127
8.4.1 Text Analysis Module	128
8.4.2 Segmentation and Design of speech database	128
8.4.2.1 Volume of Database prepared	129
8.4.3 Speech Unit selection	132
8.4.3.1 Speech synthesis using phoneme as speech unit	132
8.4.4 Synthesis Module	133
8.5 Graphic User Interface of TTS	184

8.6 Results and Analysis	136
8.7 Conclusions	143
<b>Chapter 9: Conclusions and Future work</b>	144
<b>References</b>	
<b>Appendix 1:</b> Hindi alphabets and their Roman Transliteration	
<b>Appendix 2:</b> Definition of Hindi Phones	
<b>Appendix 3:</b> CD containing framework of the developed TTS	
<b>Appendix 4:</b> Transliterated sentences collected for Travel Domain	
<b>Appendix 5:</b> Transliterated sentences collected from newspaper	
<b>Curriculum Vitae</b>	