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 /ई/ 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

[7] Archana Balyan, S. S. Agrawal, Amita Dev, ***"Development of Hindi Speech Synthesizer for Metro Rail Information System"***, 2nd International Conference of Electrical and Control Engineering, ***IEEE conference publication***, DOI: 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 VIIIth 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 22nd 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 , p-p 1-4 , India ,November 25-27, 2013.

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

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

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