Dedicated

To

My Parents

Smt Venkata Lakshmamma

Sri Balakrishna Murthy. Mannepalli

&

Family

DECLARATION

I do hereby declare that the thesis entitled "ANALYSIS AND IDENTIFICATION OF EMOTIONS AND ACCENTS OF TELUGU SPEECH" does not constitute any part of thesis/ dissertation/ monograph submitted by/ or any other person for the award of any Degree/ Diploma to this or any other university/ Institution.

Signature

Kasiprasad. Mannepalli

Reg No:13304004

Research Scholar

CERTIFICATE

This is to certify that the thesis entitled "ANALYSIS AND IDENTIFICATION OF EMOTIONS AND ACCENTS OF TELUGU SPEECH" by Mr Kasiprasad. Mannepalli to the K L University, Vaddeswaram, in fulfillment of the requirement for the award of the degree of Doctor of Philosophy in Electronics and Communication Engineering is a bonafide record of work carried-out by him under our supervision and guidance.

Co-Supervisor

Dr. M. Suman
B.Tech, M.Tech, Ph.D.,
Professor,
Department of ECM,
KONERU LAKSHMAIAH EDUCATION FOUNDATION
Vijayawada, Andhra Pradesh.
INDIA

Supervisor

Dr. P. Narahari Sastry
PGDMM, MBA, B.Tech, M.E., Ph.d.,
Professor,
Department of ECE
CBIT, Hyderabad
Telangana.
INDIA

ACKNOWLEDGEMENTS

I sincerely thank my supervisor **Dr. P. Narahari Sastry,** Professor, CBIT, Hyderabad, Telangana for his guidance and support throughout my research work. He has been an inexhaustible source of ideas and always has taken time to discuss patiently every detail of this research work. His contagious enthusiasm and open door policy for his students greatly helped to complete this research work.

I am extremely grateful to **Dr. Suman. Maloji,** Professor, Department of Electronics and computer Engineering, K L University for supervising my research work. I am greatly obliged to his constant encouragement, timely suggestions and keen interest for publishing research papers.

My sincere gratitude to Sri K. Satyanarayana, President KLEF for providing excellent R&D facilities in the campus. My sincere thanks to Dr. M. Ramamoorty, Chancellor, K L University. I would like to convey special thanks to Dr. L.S.S. Reddy, Hon'ble Vice chancellor, KL University for his helping hand to researchers in the KL University.

I thank Dr. M. Venugopala Rao, Dr. G. V. Subba Rao and Dr P.V.V. Kishore for their valuable comments during the doctoral committee meeting. These suggestions have helped me to complete my research work and present this thesis report.

My sincere thanks to Dr.A.S.C.S. Sastry, HOD, Department of ECE for his support. I also thank Dr B. L. Prakash alternate HOD for his support during my Research. I thank Dr. K. Koteswara Rao, RPAC Chairman, ECE Department, K L University for his guidance regarding procedures involved in completion of research. I would like to thank Dr. Habibulla Khan, Dean Student Affairs, and KL University for his support. I thank Dr V. Rajesh Principal, ASC, KLU for his support and suggestions during the course of my work.

I thank Dr. K. Ch. Srikavya, Associate Dean, R&D, KLU and Dr K. L. Narayana, Dean R&D for their support during my research work.

I would like to express my gratitude to Sri V. Tarun Kumar, chairman, Prof. Sunitha, Principal I/c and all other staff members of Hasvita Group their co-operation and suggestions during my research work.

I thank Dr. Vally Maya w/o Dr. P. Narahari Sastry for her encouragement and support during my research.

I thank all my friends who supported me in recording Database and information.

I would like to thank my parents *Smt. Venkata lakshmamma* and *Balakrishna Murthy*, who are always gave confidence and supported to complete my research in time. I thank my wife ***Smt. Swetha*** for her inspiration, encouragement, infinite support and patience during my research work.

I thank my siblings *Smt kasi vidyu latha, Anjani priyadarsini* and *Durga Prasad* for their constant co-operation throughout my research work. I thank brother- in-law S. Eeswara Prasad and my niece *Asrita Venka Hiranmayi* for their support.

I thank my parents- in-law *Smt Danthala Sarada, Vidyanadh* and other family members for their timely support and encouragement during my research.

I Thank God for giving me strength and many supporting people in my life.

Signature
Kasiprasad. Mannepalli
Reg No: 13304004
Research Scholar

CONTENTS

Abstract

Speech is the fundamental mode of communication among humans. The information is contained in the speech which is needed for both the listener and speaker. It also contains message including speaker's characteristics like emotion, information regarding the language and his physiological characteristics. The speech signal contains large amount of information which is complex and is in a coded form, but due to the intelligence of humans, these can be easily decoded. The research for automation of human machine interaction is a well-known area, for understanding the production of speech and extraction of information contained in the speech. The various applications of this technology are audio indexing and retrieval, control using voice command, field of dictation etc. The speaker recognition becomes more important technological development for providing a comfortable and natural form of communication between speaker and personal computer. The main purpose of speech processing is designing a machine that mimics human behavior, particularly the capability of speaking naturally and responding properly to any spoken language.

There are different accents for any given language, which are due to different speaking styles and also for ethnic differences geographically. Native and non-native speakers can be differentiated due to the difference in acoustic space spanned by phonemes. The intonation, duration, rhythm, voiced stop release time play vital role in recognizing the accent. The accent identification will increase the performance of the speech recognition system.

The features belonging to segmental and supra segmental levels can be used for recognition of accent of any language. In general 20-30 milliseconds time period speech segment is used for extracting features belonging to segmental level. These are also called as spectral features. The prosodic features have a time period of more than 100 milliseconds and are also called supra-segmental features. The duration of speech segment is less than 3 milliseconds for sub-segmental features. There are many accents of Telugu language which are spoken in different regions of Telugu states i.e. Andhra Pradesh and Telangana. However, broadly there are three main accents namely Coastal Andhra, Rayalaseema and Telangana which are considered in this proposed research work.

Apart from different accents, emotions (state of mind of an individual) of the speaker also play an important role in the area of speech and speaker recognition. Emotion contained in a speech is very important for any communication between the speaker and listener, since the meaning completely reverses, when the same sentence is spoken in 'Angry' or 'Happy' emotion. Similarly this happens in the case of 'Bore' and 'Neutral' emotions. The symbolic representation of any speech helps to identify the emotional coefficient. In general emotional recognition is a methodology of synthesizing the individual's inherent behavior by considering a fragment of speech sample uttered by the individual. Also it can be interpreted as the science of reading the individual mind.

India is multi lingual country and has about 1652 dialects from native languages. Research in the area of speech and speaker recognition is at nascent stage for most of the Indian languages. Hence a separate database for Indian languages, based on the requirement should be developed in the laboratory environment.

This research work focuses on automatically identifying the dialect or accent of the speaker when a sample speech is given. Since there is no standard database for Telugu speeches, laboratory environment is used to develop the training set and also the testing set. In this work apart from accent based speech and speaker recognition, emotions are also considered for identifying the speech and speaker.

The main objective of this proposed research work is "to design and develop algorithms for accent based recognition system for Telugu speech". The second objective of this research is "to design and develop algorithms for emotion based recognition system of Telugu speech samples".

The main task for developing the accent recognition system is building the accent speech database for Telugu speech for different regions i.e. Coastal Andhra, Rayalaseema and Telangana. For experimental evaluation, thirteen speakers from each region were selected. The speech recorded was a text dependent Telugu speech "Evaro annam tinnaru nenu evarini Choodaledhu". All the speeches were collected in acoustic maintained laboratory. The distance between the microphone and the speaker was also maintained constant to avoid variation of the speech collected.

Feature extraction plays an important role for both emotion and speaker recognition system. The various features for every speech sample are extracted using "Colea" software. The set of features selected in the Prosodic-NNC method are formants F1, F2, F3, pitch, energy and power spectral density. In the next step "Praat" tool is used to extract prosodic features which are added to the above set of features selected in the proposed algorithm. The classifier used in this method is Nearest Neighborhood Classifier

(NNC) also popularly called as Euclidian distance.

This algorithm could successfully recognize twenty test samples pertaining to Coastal Andhra out of twenty six speech samples considered for testing. Therefore the percentage recognition accuracy for the speech from the Coastal Andhra region is (20/26) **77 %**. Similarly for the Rayalaseema region, twenty four test samples out of twenty six test samples were successfully recognized by the proposed algorithm. Therefore the percentage recognition accuracy for the speech from the Rayalaseema region is (24/26) **92%** and for Telangana region seventeen test samples were successfully recognized out of twenty six test samples. Hence the percentage recognition accuracy for the speech from Telangana region is (17/26) **65%**. On the whole, this proposed algorithm could successfully recognize **61** speech samples out of **78** speech samples and hence the **overall recognition accuracy** of this proposed system becomes **78%**.

The second algorithm uses MFCC (Mel Frequency Cepstral Coefficients) features and GMM (Gaussian Mixture Model) for accent identification. For both the training and testing samples, MFCC features were extracted. The same Telugu sentence was spoken for five times by every speaker. The MFCC features (Thirteen in number) extracted for the first three speeches of each speaker, were used for training. Further, the MFCC features of the next two speeches of the same speaker were used as testing database. This procedure is repeated for all the three regions i.e. Coastal Andhra, Rayalaseema and Telangana. The GMM training models are built for each of the three region speeches separately. The total number of Gaussian mixtures in the proposed algorithm is sixteen in number. In the testing phase, the testing samples from each of the three region’s accents i.e. Coastal Andhra, Rayalaseema and Telangana have been taken and the MFCC

feature set for each of the test speech signal were obtained as already described for training set. Further, GMM testing was conducted for each of the speech signal. Out of the 26 speech samples used for testing pertaining to Coastal Andhra, 23 were successfully identified to be of Coastal Andhra, whereas 1 test speech sample was misclassified as Rayalaseema and 2 other speech samples were misclassified as Telangana accent. Hence the recognition accuracy for Coastal Andhra accent becomes 88%. Similarly 26 speech samples of Rayalaseema accent were tested, 24 samples were correctly classified. Therefore the recognition for Rayalaseema accent becomes 92%. Further, 24 out of 26 samples of Telangana region speech samples were correctly classified. Therefore the recognition accuracy for Telangana region accent becomes 92%. The overall recognition accuracy of the MFCC-GMM system is 91%.

The third algorithm for the identification of Telugu accent uses Spectral flux, pitch chroma, Tonal power ratio, MFCC features as feature set and Deep Belief Networks (DBN) as classifier. The number of hidden layers in DBN was tried with 4 layers, 5 layers, 6 layers and 7 layers. The recognition accuracy for 7 layers DBN model gave encouraging and best results. Hence, 7 layers are used as hidden layers in this proposed method. The Recognition accuracy for "Coastal Andhra accent" test samples obtained is **94%**. For the Rayalaseema test samples **92%** of the samples are correctly recognized, and 8% of them are wrongly recognized as Telangana Samples. For the Telangana accent test samples **94%** of the samples are correctly recognized, and 6% of them are wrongly recognized as Coastal Andhra. The Overall recognition accuracy found is **93%** in this proposed system.

The proposed emotion recognition system of Telugu speeches uses prosodic features and Nearest Neighborhood Classifier for

classification. A database is prepared by recording speeches from eight people. The sentence recorded is "**entraa ila vachhavu?**". The emotions considered in this work, are **Happy, Boredom, and Neutral**. Each speaker spoke the sentence in all the three emotions for **twenty** iterations. The various features like Minimum pitch, maximum pitch, average pitch, standard deviation of pitch, range of pitch, Formant F1, Bandwidth of F1, Formant F2, Bandwidth of F2, Formant F3, Bandwidth of F3, Formant F4 and Bandwidth of F4 were extracted for these speech samples. The total number of samples under test for 'Neutral' emotion is forty in number. Thirty three speech samples with 'Neutral' emotion were correctly identified in the proposed emotion recognition system. Hence the percentage of recognition accuracy for 'Neutral' emotion is **82.5%**. In the case of 'Happy' emotion thirty four speech samples were correctly identified, when forty samples of 'Happy' emotion were tested. Hence the percentage of recognition accuracy for 'Happy' emotion is **85%**. Similarly for the emotion 'Bore' twenty eight samples were correctly recognized out of forty samples tested and hence the recognition accuracy for 'Bore' emotion becomes **70%**. The overall recognition accuracy on an average for these three emotions found to be **79%**.

The recognition accuracy in the published literature is **81%**, for the accent recognition of Hindi language. In the proposed method it is **93%** using DBN method. Similarly the recognition accuracy for the emotions of Hindi speech in the literature is found to be **78%**, which include five emotions namely- Anger, Happy, Neutral, Bore and sad. The proposed method for Telugu speech gave recognition accuracy of **79%**, for three emotions namely Happy, Bore and Neutral, which is in line with published results.

In this work algorithms were developed to identify the various accents and also different emotions which include 'Happy',

‘Neutral’ and ‘Bore’. This research work proved that both the accent and emotion of a speaker can be used to increase the recognition accuracy for identifying the region to which the speaker belongs i.e. Coastal Andhra, Rayalaseema and Telangana.

List of Publications

International Journals

1. Kasiprasad Mannepalli , P. Narahari Sastry, V. Rajesh, "Modeling and analysis of Accent based Recognition and speaker identification system", ARPN Journal of Engineering and applied Sciences, Vol 9, No. 12, pp:2807-2815, December 2014. **(Scopus indexed).**

2. Kasiprasad Mannepalli, P. Narahari Sastry, V. Rajesh, "Accent Detection of Telugu Speech using Supra-Segmental Features", Journal of Soft Computing, Vol.10, No.5, pp:287-292, 2015. **(Scopus indexed).**

3. Kasiprasad Mannepalli, P. Narahari Sastry, Maloji. Suman, MFCC-GMM Based Accent Recontion System for Telugu Speech", Springer International Journal of Speech Technology, Vol 19, issue 1, pp. 87-96, March 2016.(**Scopus Indexed, Thomson Reuters ESCI indexed).**

4. Kasiprasad Mannepalli, P. Narahari Sastry, Maloji. Suman, "A Novel Adaptive Deep Belief Network for Speaker Emotion Recogntion", Alexandria Engineering Journal (Elsevier). (**Scopus Indexed, Thomson Reuters ESCI indexed). "Article in press"**

International Conference

1. Kasiprasad Mannepalli, P.Narahari Sastry,V.Rajesh "Analysis and design of Speaker Identification System using NNC" ICACM, Elsevier digital edition, pp. 381-387, 2013.

2. Kasiprasad Mannepalli, P.Narahari Sastry,V.Rajesh "Analysis and design of Speaker Identification System using Neural networks" NCASPA-KLU, pp.149-155, 2013.

3. Kasiprasad Mannepalli, P. Narahari Sastry, V. Rajesh, "Accent Detection of Telugu Speech Using Prosodic and Formant Features", IEEE-SPACES, pp 318-322, 2015. **(Scopus indexed).**

4. Kasiprasad Mannepalli, Panyam Narahari Sastry, Maloji Suman, "Analysis of Emotion recognition system For Telugu using prosodic and formant features" in CSI - 2015; CSI - 50th Golden Jubilee Annual Convention, New Delhi during 02nd - 05th December, 2015, published by Springer under AISC Series, 2015. **(Scopus indexed).**

5. Kasiprasad Mannepalli, Panyam Narahari Sastry, Maloji Suman, "Accent recognition system using deep belief networks for Telugu speech signals" in FICTA-2016, Bhubaneswar, published under Springer AISC series, 2016. **(Scopus indexed).**

List of Figures

List of Tables

Nomenclature

RA: Recognition Accuracy

ASR: Automatic Speech Recognition systems

HTC: High Tech Computer Corporation

MOS: Mean Opinion Score

ACF: Autocorrelation function

LPC: Linear prediction coding

NNC: Nearest Neighborhood classifier

CA: Coastal Andhra

RS: Rayalaseema

TG: Telangana

PSD: Power Spectral Density

dB: Decibles

Hz: Hertz

MFCC: Mel Frequency Spectral Coefficients

AANN: Auto Associative Neural Network

SVM: Support Vector Machine

GMM: Gaussian mixture model

DBN: Deep Belief Networks

FFT: Fast Fourier Transform

SER: Speech Emotion Recognition

TTS: Text To Speech

CFCC: Cochlear Filter Cepstral Coefficients

UBM: Universal Background Model

HMM: Hidden Markov Model

SMO: Successive Minimal Optimization

LPCC: Liner Predictive Cepstral Coding

EER: Equal Error Rate