International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 8– No.7, May 2015 – www.ijais.org

A Novel Approach for Hindi Text Description to Speech and Expressive Speech Synthesis

Kamble Kaveri
Department of Computer Engineering
DYPSOET, Lohegaon
Pune, India

Ramesh Kagalkar
Department of Computer Engineering
DYPSOET, Lohegaon
Pune, India

ABSTRACT
Communication plays a very important role in every days life. With the help of communication we can share information from one person to another. Speech is the primary means of communication. A Text to Speech (TTS) synthesizer is a computer based application which is capable of given reading out to the typed text. This generally forms basic two steps, such as text processing and speech generation. Our aim is to develop software that enhances the users way of speech through correctness of pronunciation for the Hindi phonetics. Firstly the simple TTS system is to perform operation to get the output in the form of Text for Hindi language. Then Speech to Text (STT) conversion may form effectively. Additionally we have to add Expressions for Expressive Speech synthesis for Hindi Language. TTS is one of the major applications of NLP.Expressive speech synthesis deals with synthesizing speech and adding various expressions related to different emotions and speaking styles to the synthesized speech. Emotion is an important element in expressive speech synthesis.

General Terms
Expressive Speech synthesis, Emotions, Audio-Visual Speech, Natural Speech,

Keywords
Text To Speech ; Speech To Text; Boosting-Gaussian Mixture Model(GMM); Mel Frequency Cepstral Coefficient (MFCC); Prosody Conversion; Hidden Markov Model(HMM); Time Domain Pitch Synchronous Overlap Add(TD-PSOLA).

1. INTRODUCTION
New programs have been successful through the development of pronunciation of words, as text and speech synthesis techniques to provide an integrated solution for the blind and visually impaired people, making them not always associated with a person to read or write for them, and made them able to deal with the computer easily for their very known language like Hindi Language. Even many of the great researchers have dreamed of involving the machines in every face of human life. With the growth in the power of computing machines, their applications in modern daily life are also rising [19]. Speech represents the spoken form of a language for Hindi Language and also one of important means of communication [7]. This Research has resulted in important advances with many systems Being able to generate a close to a real natural speech. These advances in speech synthesis also pave the way for many speech related new applications.[1][2]. This will also help physically challenged people and even partially blind people who find it difficult to read from a monitor [4]. When we talk about speech based interfaces for computer system computer system, we refer to two basic technologies: Speech Recognition Speech Synthesis. Speech Synthesis, i.e., Text-to Speech system and Speech Recognition, i.e., Speech-to- Text system, together form a speech interface.[10] Speech is often based on concatenation of natural speech that is units, which are taken from natural speech put together to form a word or sentence [1]. A Text-To-Speech synthesizer is a computer based system capable of converting computer readable text into speech [20][21]. Text-to-Speech (TTS) for the Hindi language is the generation of synthesized given speech from text. A TTS synthesizer allows people visual impairments and reading disabilities to listen to work on a home computer[13][18][20]. The TTS can be a voice for those people who cannot speak. The TTS system can be used to read text from emails, SMSs, web pages, news, articles, etc in Hindi language.. In such reading applications, the TTS technology can reduce the eye-strain.[10].

2. TEXT TO SPEECH
A Text-To-Speech (TTS) synthesizer is computer based system that should be able to read any text aloud and output getting in the form of Speech. Here we are developing our approach for Hindi language [3]. In our system we have to get input in the form of Hindi word and output that generated in the form of Speech. Official language of India is Hindi. A very small percent of Indian use English as a means of communication. Spoken language interfaces enabled with TTS synthesis have potential to make information. Speech synthesis is the process of converting message written in text to equivalent message in spoken form [7]. TTS system has many challenges when conversion takes place.[8] The most important qualities achieved from Hindi speech synthesis system are naturalness and intelligibility.[10][12][15][16]. Text-to-Speech (TTS) systems today have formed a high level of performance and prosody and synthesis voice quality prevent them from being deployed in man machine communication for Hindi Language.

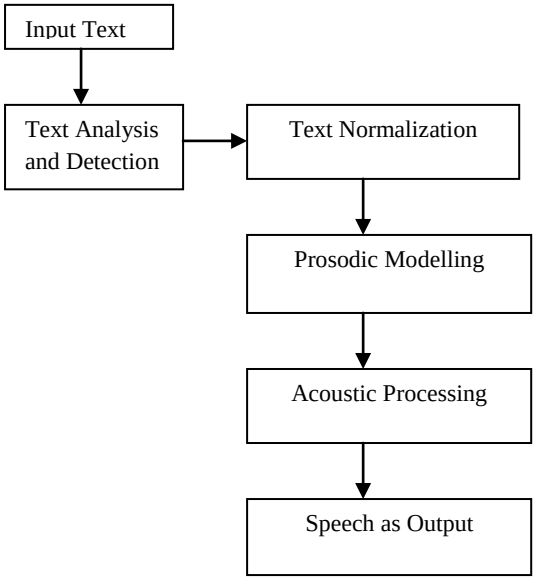


Fig.1.System overview of TTS system

3. SPEECH TO TEXT

Speech is the primary means of communication [6]. Spoken words always play very important role in communication. In this STT system we have to Speech as input to the system and output get in the form of Text.[7]. Hindi TTS Engine play very important role for converting speech to text. Voice segmentation is necessary in dividing the signal into smaller parts based on the syllable [2]. Intralingual speech to-text-conversion is a useful tool for integrating people with hearing impairments in oral communication settings, e. g. counseling interviews or conferences [3]. Speech-to-text-translation (Audio visual translation) of spoken language into written text is an upcoming field since movies on DVDs are usually sold with subtitles in various languages [4][9].

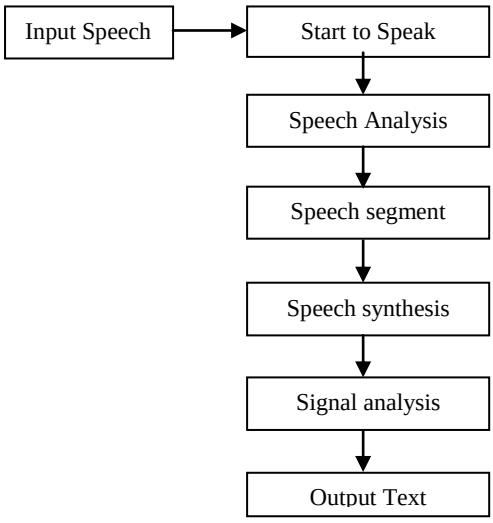


Fig.2.System overview of STT system

4. EXPRSSIVE SPEECH SYNTHESIS

Emotion which is a very important element in expressive speech synthesis.[6] Recently, more and more efforts have been made in the research for expressive speech synthesis, among which emotion is a very important element used emotional keywords and emotion trigger words to generate an emotional TTS system for Hindi language [6][7]. The emotional training and testing corpus with four degrees, strong, normal, weak, unlike among emotions, happiness, sadness, fear, and anger. So called speech is used as the reference source for conversion[7][13][14]. The GMM method is more suitable for a small training set. Work on expressive speech synthesis mostly focused on the expression of basic emotions exclusively The simplest way to produce synthetic speech is to play long prerecorded samples of natural speech, such as single words or sentences.

5. PROPOSED SYSTEM

Many Text to Speech systems are already developed for Different languages existing but they are not so user friendly. Hence, to full fill this requirement we develop our system for Automatic Translation of Text to Speech and Vice Versa for Hindi Language. The aim of this system is to develop a desktop application that can combine between three functions, converting Hindi Text to Speech, translating the Hindi Speech to Text and Expressive Audio visual speech synthesis based on Emotions. The System is very user friendly so user can use this system at any time and any place. The user also can use each function separately in order to provide more benefits and efficiency.

6. SYSTEM ARCHITECTURE

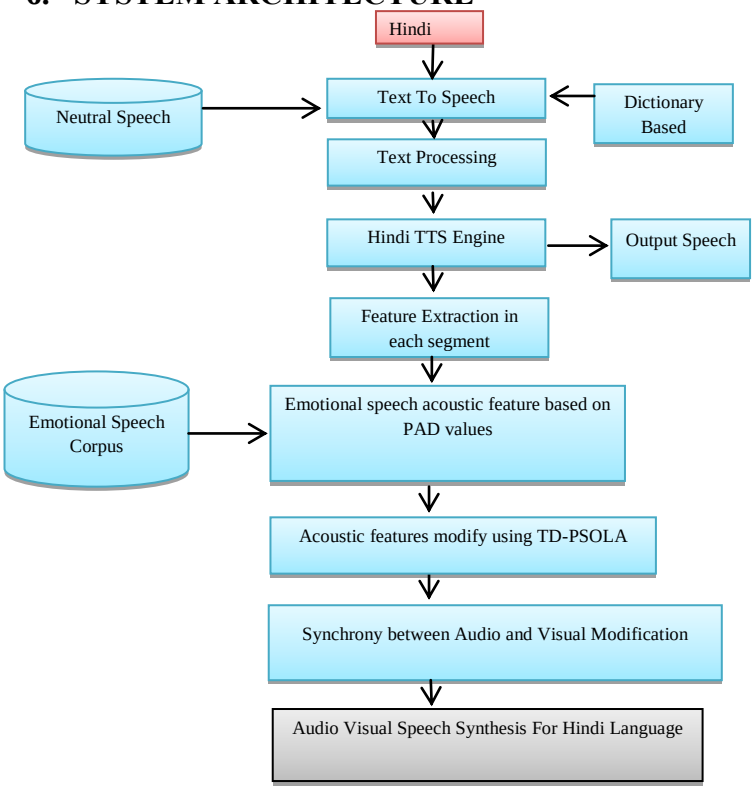


Fig.3.System Architecture of overall system

7. CHALLENGES OF SYSTEM

Speech synthesis has been developed rapidly over the recent years and it has been integrated into several new applications. Developing speech synthesis system is a

complicated process and, it includes the following important challenges. [3]

1. Development of TTS systems require knowledge about human speech production and about languages being developed.
2. The actual implementation of a fully functional system requires good software skills.
3. To detect emotions are not simple task. So to identify emotions the naturalness is very important.
4. The most important qualities of a speech synthesis system are naturalness and intelligibility. Naturalness describes how closely the output sounds like human speech, while intelligibility is the ease with which the output is understood. The ideal speech synthesizer should be both natural and intelligible.
5. In our system we have to combine three functions so each one required the basic database which is very important so database should be strongly identified.
6. In this system we proposed Audio visual speech synthesis but when we developing system various problems occur during the emotions specific area.

8. AUDIO VISUAL SPEECH SYNTHESIS

Here, in our system the mouth movement plays an important role in both speaking and expression style. We first introduce our emotional speech datasets and the quality of the datasets. Then we evaluate the emotional speech conversion models, by comparing the performance of GMM and Boosting GMM, also the segmental conversion strategy and the global conversion strategy[6][3]. Emotion is an important element in expressive speech synthesis.[7]

Here we describe some basic aspects related to our system which can be explained with the help of table.

Table 1. That describes the amount of time that required for the system. The input in the form of the Words, Sentence or Audio file. Amount of time that required for the TTS system is less as compare to STT and Expressive system. The time that taken into the milliseconds (ms). When we formed TTS sys-tem the input in the form of Text, words. But in STT system input should be in the form of Audio file.

Words or Sentence count	TTS	STT	Emotion Detection
5	1000ms	2000ms	4000ms
10	1200ms	2200ms	4200ms
15	1400ms	2500ms	4500ms

Comparative process analysis for TTS, STT, and Emotion Detection system. The Time (ms) that required for completion of one process for each one of system and the input that shows in term of Word, Sentence count input file size in (Kb) and the amount of time required in (ms).

9. PROCESS EXECUTION

9.1 Text To Speech Conversion

i Get Text from user.

ii Get English similar words related to text entered.

English Word	Related Hindi Word
aap	आप
arthapurna	अर्थपूर्ण

iii Apply Hindi text to TTS engine.

iv Extract audio from database.

For natural speech extract audio related to words given at input

v Recognize audio.

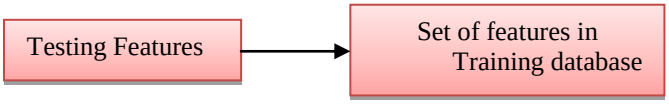
9.2 Speech to Text Conversion

i Get Audio from user.

Input = .wav file (speech audio file)

ii Extract Features of audio signals using MFCC.

iii Match Features from training data (stored into database).



iv Extract text from database according to features.

9.3 Emotion Detection from Video

i Get Audio from user.

ii Extract features of Audio and Video.

iii Apply Boosting GMM.

Boosting-GMM algorithm contains several weak prediction models. One of them is the basic prediction model and the others are assistant prediction models. The basic model is the regression model for predicting feature differences between emotional and neutral speeches using GMM.

iv Search from database matches with expressions of testing clip.

v Result type of expressions in clip.

10. ALGORITHM STRATEGY USED

Modify the acoustic features of the neutral speech using TD-PSOLA algorithm to obtain the emotional speech. Hindi TTS Engine plays very important role in TTS and STT system. In our system we basically used GMM and Boosting GMM are useful for the Emotional Speech acoustic Features. TDPSOLA Algorithm is used to convert the neutral speeches to the target emotional speeches.TD-PSOLA algorithm to modify the pitch and duration. Specifically, the Boosting-GMM is used to convert the neutral speeches to emotional speeches [3][6].

11. CONCLUSION

The text to speech conversion may seem effective and efficient to its users if it produces Natural speech and by making several modifications to it. This system is useful for deaf and dumb people to Interact with the other peoples from society. The system which is a desktop application that works primarily as a converter between text and speech, for the Hindi language in both directions. Text to speech synthesis is a critical research and application area in the field of multimedia inter-faces. The proposed work presents an algorithm for converting text to speech by using natural language processing for Hindi language. There are many text to speech systems (TTS) available in the market and also much improvisation is going on in the research area to make the speech more effective, natural with stress and emotions. At present, with inadequate prosodic models in place, the quality of synthetic speech generated by the synthesizers is poor. So efforts can be done for the development of prosodic models.

The further work can be done to improve the natural-ness and intelligibility of TTS. A Web based application can also be designed which can convert text in any Indian languages into speech.

12. ACKNOWLEDGMENTS

The authors would like to thank Chairman Groups and Management and the Director/Principal Dr. Uttam Kalwane, Colleague of the Department of Computer Engineering and Colleagues of the Department the D. Y. Patil School of Engineering and Technology, Pune Dist. Pune Maharashtra, India, for their support, suggestions and encouragement.

13. REFERENCES

[1] J. Tao ,Y. Kang,and A. Li ,Prosody Conversion From Neutral Speech to Emotional Speech, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 14, No. 4, July 2006.

[2] M. Theune, K. Meijs, D. Heylen, and R. Ordelman ,Generating Expressive Speech for storytelling Applications' , IEEE Transaction on Audio, Speech and Language Processing,Vol.14, No.4, July 2006.

[3] D. Govind , S. Mahadeva Prasanna , Expressive speech synthesis: a review , Springer Science Business Media New York 2012.

[4] B. Yegnarayana and K. Sri Rama Murty , Event-Based Instantaneous Fundamental Frequency Estimation From Speech signals, IEEE Transactions on Audio ,Speech. And Language Processing, Vol.17. No. 4, May 2009.

[5] O. Turk and M. Schroder , Evaluation of Expressive Speech Synthesis with Voice Conversion and Copy Resynthesis Techniques, IEEE Transactions on Audio, Speech and Language Processing, Vol.18, No.5 ,July 2010.

[6] J. Jia, S. Zhang, F. Meng, Y. Wang, and L. Cai, Member, IEEE, Emotional Audio-Visual Speech Synthesis Based on PAD,IEEE Transactions on Audio, Speech and on Audio, Speech and Language Processing, Vol.19, No.3 , march 2011.

[7] J.Sangeetha,S.Jothilakshmi , S. Sindhuja , V. Ramalingam, Text to Speech synthesis system for Tamil, International Conferenceon Information Systems and Computing (ICISC-2013),India.

[8] K. Kamble and R. Kagalkar,A Review:Translation of Text toSpeech Conversion for Hindi language, International Journal of Science and Research (IJSR) Volume 3 Issue 11, November, 2014.

[9] M. Singh, K. Verma , Text to Speech Synthesis for numerals into Punjabi language, International Journal of Computational Linguistics and Natural Language Processing Vol 2 Issue 7 July 2013 ISSN 2279 0756.

[10] N. Swetha, K. Anuradha ,Text-to-speech conversion, International Journal of Advanced Trends in Computer Scienc and Engineering ,Vol .2,No.6, Pages (2013).

[11] S. Ahlawat, R. Dahiya , A Novel Approach of Text to Speech Conversion Under Android Environment, (IJCSMS) International Journal of Computer Science Management Studies, Vol. 13, Issue 05, July 2013

[12] P. Shetake, A. Patil, P. Jadhav , Review Of Text To Speech Conversion MethodS, International Journal of Industrial Electronics and Electrical Engineering, ISSN: 2347-6982 Volume-2, Issue-8, Aug.-2014.

[13] S. Suryawanshi, R. Itkarkar, D. Mane , High Quality Text to Speech Synthesizer using Phonetic Integration, International Journal , Advanced Research in Electronics and Communication Engineering (IJARECE) Volume3, Issue 2, February 2014.

[14] D. Sasirekha, E. Chandra ,Text to Speech: A Simple Tutorial,International Journal of Soft Computing and Engineering (IJCSE) ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[15] S. Hertz, J. Kadin, And K. Karplus, Member, IEEE,The Delta Rule Development System for Speech Synthesis from Text,Proceedings of the IEEE ,Vol.73, No.11, November 1985.

[16] R. San-Segundo, J. Montero, R. Barra-Chicote, J. Lorenzo, Architecture for Text Normalization using Statistical Machine translation techniques, Springer-verlag Berlin Heidelberg 2011.

[17] A. Chauhan, V. Chauhan, S. Singh, A. Tomar, and H. Chauhan, A Text to Speech System for Hindi using English Language,IJCST Vol 2.Issue 3,September 2011.

[18] S. Padmavathi, K. Reddy , Conversion Of Braille To Text in English,Hindi and Tamil Languages International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.3, No.3, June 2013.

[19] S. Suryawanshi, R. Itkarkar, D. Mane , High Quality Text to Speech Synthesizer using phonetic Integration, International Journal of Advenced Research in Electronics and Communication Engineering (IJARECE) Volume 3, Issue 2,February 2014.

[20] O. Trk, O. Byk, A. Haznedaroglu, and L. Arslan, Application of conversion for cross language rap singing transformation in proc.IEEE ICASSP, Taipei, Taiwan, April 2009.

[21] Z. Zeng, P. Maja, G. Roisman, and S. Thomas ,A survey of affect recognition methods: Audio, visual and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 3958, Jan.2009.