A New Approach of Emotion and Facial Expression Detection of Speaker with Conversion of Text To Speech and Vice Versa for Hindi Language.

KambleKaveri
Department of Computer Engineering
DYPSOET, Lohegaon
Pune, India.

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

ABSTRACT
In India, Hindi is the official language spoken by the majority of the population. Text is ahuman readable sequence of characters and the words they form that can be encoded into computer readable formats. Speech play very important role in individual and collective lives of thepeople and it represents the spoken form of a language. Expression is defined as the quality orpower of expressing an attitude, emotion, feeling, spirit, character etc. on the face in the voiceor in artistic execution. Emotion is an affective state of consciousness in which joy, sorrow,fear, hate, or the like is experienced as distinguished from cognitive and states of consciousness. The main goal of the proposed system is to be performed Text to Speech (TTS), Speechto Text (STT) and expressive speech synthesis.The proposed system database consists of huge set of words, images and audio files which help to perform training of database.The most important qualities expected from aspeech synthesis system are naturalness and intelligibility. Speech to text system may seem effective and very useful.

General Terms
Speech Synthesis, Text Preprocessing, Speech generation, Feature Extraction, Feature Matching, Expressions and Emotions.

Keywords
Text to Speech, Speech to Text, Expressive Speech Synthesis, Gaussian Mixture Model, SpeechSynthesizer and Mel Frequency Cepstral Coefficients.

1. INTRODUCTION
In India different languages are spoken, each language being the mother tongue of tens ofmillions of people. Expressive speech synthesis is associated with a wide variety of feelings,thoughts, and behavior [1]. The expression is defined as the indicator of various emotionalstates that reflect in the speech waveform [2]. Text to speech synthesizer is a computer based application that is capable of reading out typed text. This generally involves two steps, text processing and speech generation. A proposed system takes input in the form of .wav file and output generate in the form of Hindi text [3]. In order to improve naturalness in human machine interaction, expressive speech synthesishas started to attract increasing attention in recent years. Emotion is an important element in theexpressive speech synthesis. TTS is one of the major applications of Natural Language Processing (NLP)[4]. Communication plays very important role in individual and collective lives of the people.Speech is the primary means of communication. Hindi is a very popular language of India. TTS system most widely used in the audio reading devices for the deaf and dumb peoplenow days [5].Speech to text translation is useful for integrating people with hearing impairments in the oral communication process. Speech is often based on concatenation of natural speech is that the units that aretaken from natural speech put together to form a word or sentence. Expressive speech synthesis deals with the synthesizing speech and various expressions are related todifferent emotions and speaking styles to the synthesized speech.

2. TEXT TO SPEECH
A proposed a system takes input in the form Hindi text and output generate speech with spoken waveform. There are several steps performed in a TTS system such as the text preprocessing, text analysis and detection, text normalization, acoustic processing. When all these steps are performed we get output as speech with spoken waveform. In India different languages are spoken, each language being the mother tongue of tens ofmillions of people. Text to Speech (TTS) synthesizer is a computer based applicationthat is capable of reading out typed text. Speech synthesis is the process of converting message written in text to the equivalent message in the spoken form.

It is one of the major application of Natural Language processing (NLP). In proposed system we simply take the input inthe form of Hindi text and generate output as speech with spoken waveform. Preprocessing isone of the important steps performed. Next text analysis which analyses the text and displayinto a manageable list of words. In text normalization method identify the punctuation marksand manage the words into the pronounceable form. Then the last step is the acoustic processing. After all these steps are performed the lastly output in the form of speech with signal wave.

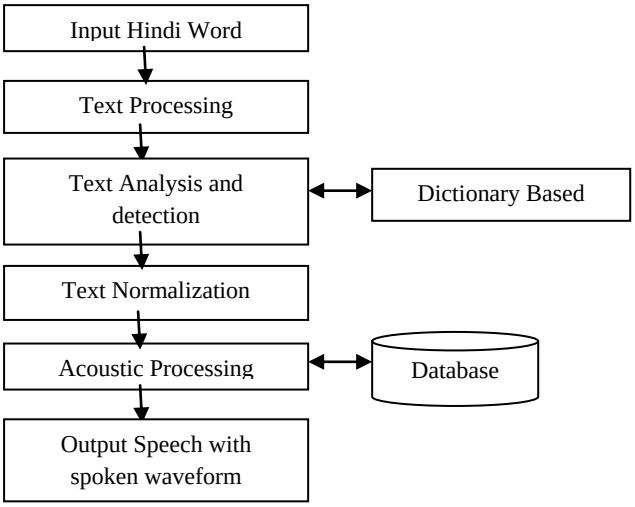


Fig.1. System Architecture of TTS System

3. SPEECH TO TEXT

Speech play very important role in individual and collective lives of the people and it represents the spoken form of a language. A proposed system takes input in the form of .wav file and output generate in the form of Hindi text. When we play audio file we have to reduce noise in that by using Double Data Source (DDS) Algorithm. Then extract the features by using Gaussian Mixture Model (GMM) and Mel Frequency Cepstral Coefficients (MFCC) algorithm and match the features with the database. Then we get output in the form of text.

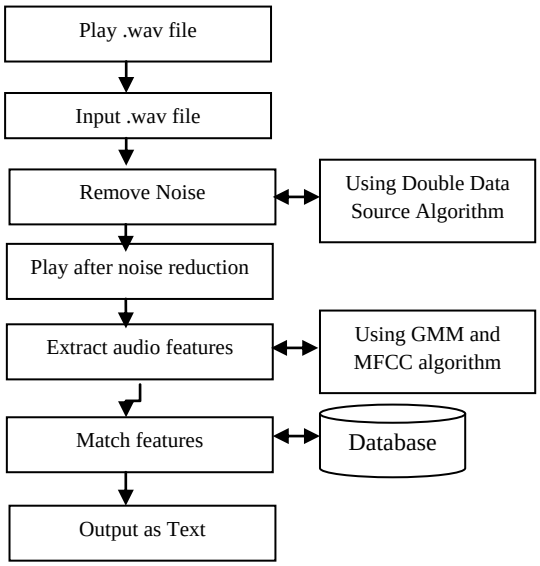


Fig.2. System Architecture of STT System

4. EXPRESSION DETECTION

Expression is defined as the quality or power of expressing an attitude, emotion, feeling, spirit, character etc. on the face in the voice or in artistic execution. Training and testing data set are basic elements in proposed system. A proposed system takes input as image with different expression and identifies the expression for extracting different features. Then display the image and next steps is preprocessing and extract the features of image by using the Scale Invariant Feature Transform (SIFT) algorithm. Then feature identification and feature matching process performed and then lastly the output that display the expression contain in the image.

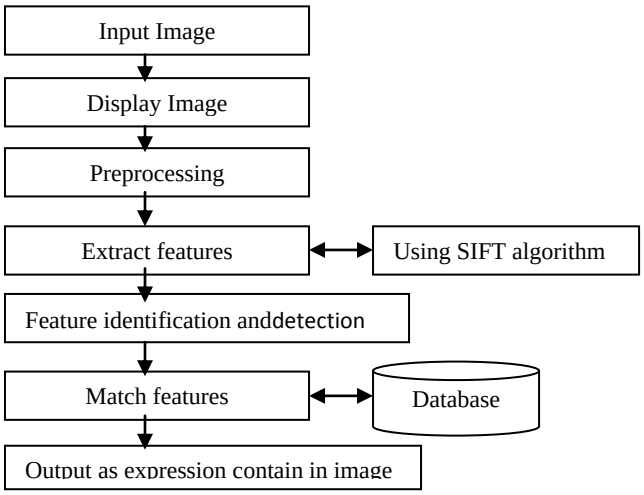


Fig.3. System Architecture of Expression Detection

5. EXPRESSION DETECTION

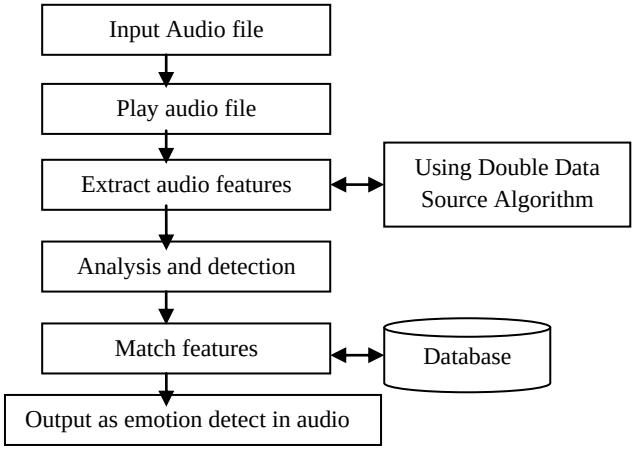


Fig.4. System Architecture of Emotion Detection

Emotion is an affective state of consciousness in which joy, sorrow, fear, hate, or the like is experienced as distinguished from cognitive and states of consciousness.A proposed system takes input as an audio file. Next to play the audio file and extract the features of audio file by using (MFCC) algorithm. Then analysis and detection steps are performed. Next match the features with the database and the output get the emotion in the audio file.We have to provide a strong database for text to speech, speech to text, expression detection and emotion detection system.

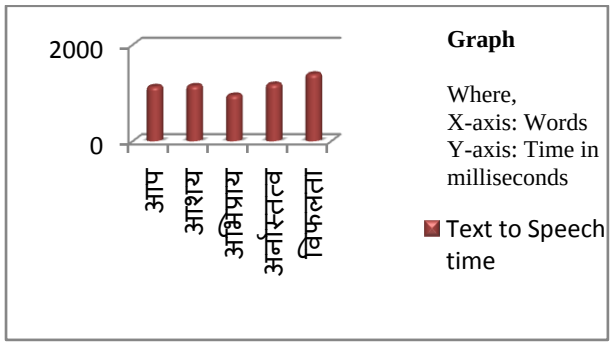
6. RESULT AND DATA TABLE DISCUSSION

In **text to speech system** table 9.1 that shows the data dependent words required theamount of time (ms). In data dependent process the system take less time as compare to independent process. In data dependent process the words which are already store into the databasesystem which get result correctly.

Table 6.1 Data dependent text to speech time for words

Sr.No	Words	Text to speech time
1	Aap	1130 ms
2	Aashay	1348 ms
3	Abhipray	951 ms
4	Arnastav	1179 ms
5	Viphalta	1390 ms

Figure 6.1 that shows the time required for TTS system for dependent process and we get inputas some Hindi words. The amount of time that required for this words can display into thegraph.

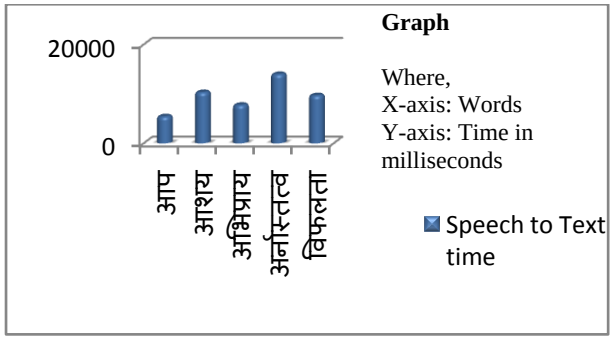


In **speech to text system** table 6.2 that shows the data dependent words required the amount of time (ms). In data dependent process the system take less time as compare to independent process. In data dependent process the words which are already store into the database system which get result correctly.

Table 6.2: Data dependent speech to text time for word

Sr No	.wav files	Speech to text time
1	Aap	1130 ms
2	Aashay	1348 ms
3	Abhipray	951 ms
4	Arnastav	1179 ms
5	Viphalta	1390 ms

Figure 6.2 that shows the time required for STT system for dependent process and we get inputas some .wav files. The amount of time that required for this words can display into the graph.

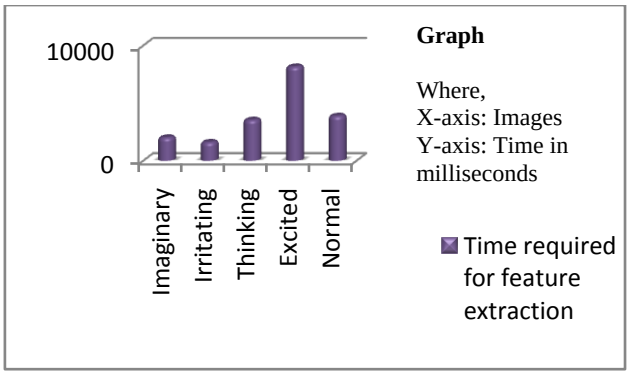


In **Expression detection system** table 6.3 that shows the input in the form ofimages. Time required for the feature extraction for image.

Table 6.3Feature extraction time required for the images

Sr No	Images	Time required for feature extraction
1	Imaginary	2025 ms
2	Irritating	1620 ms
3	Thinking	3555 ms
4	Excited	8176 ms
5	Normal	3898 ms

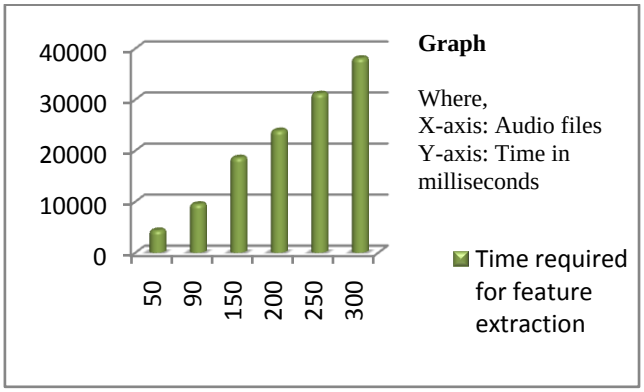
Figure 6.3 shows that the graph shows the time required for extract the features of different imagesaccording to the expression like imaginary, surprise, excited etc.



In **Emotion detection system** able 9.4 that shows the input in the form of audio files. Time required for audio file extraction in (ms).

Table 6.4Time required for the feature extraction of audio files.

Sr No	Audio file size	Time required for feature extraction
1	50	4382ms
2	90	9517 ms
3	150	18637 ms
4	200	23945 ms
5	250	31167 ms
6	300	38103 ms



7. APPLICATION SCREEN

7.1 Home Screen

In figure 6.1, we have shown the home screen of our system. In proposed system performed task such as Text to Speech, Speech to Text, Expression detection and Emotion detection. When we click on Text to Speech system this window can be open.

Figure 7.1: Home screen of system.

7.2 Words Selection

In figure 7.2, shows the data dependent words selection screen of the proposed system.Here, the words can be selected from the database that user wants to convert and get speechwith waveform. In figure 8.4, shows the select आप word in the database.

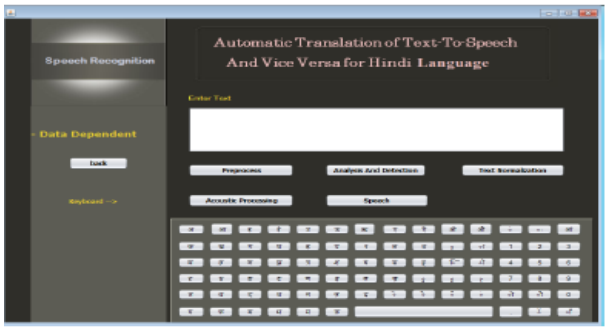


Figure 7.2: words selection for Text to speech system

7.3 Waveform for Word as आप

In figure 7.3, shows the data dependent wordsआप selection screen of the system. Here, the words can be selected from the database that user wants to convert and get speech with waveform and aap words with the speech with spoken waveform.

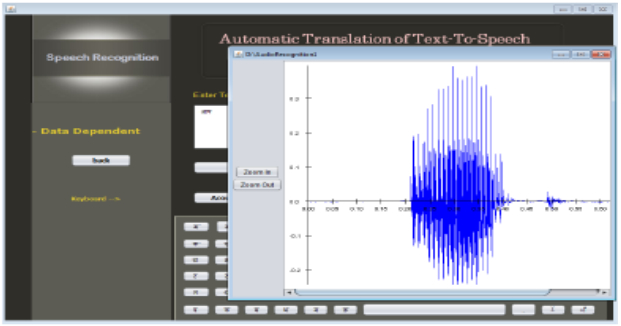


Figure 7.3: Shows waveform for input आप

7.4 .wav File Selection for Speech to Text

In figure 8.6, shows the data dependent .wav file selection screen of the system. Here, the.wav file can be selected from the database and when select we play .wav file reduce noisesuccessfully. Then we extract the features and detect text.

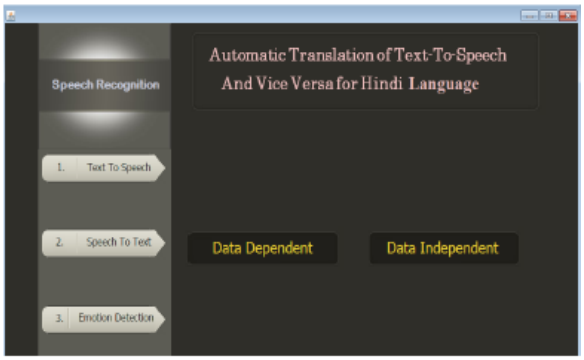


Figure 7.4: Shows data dependent selection.

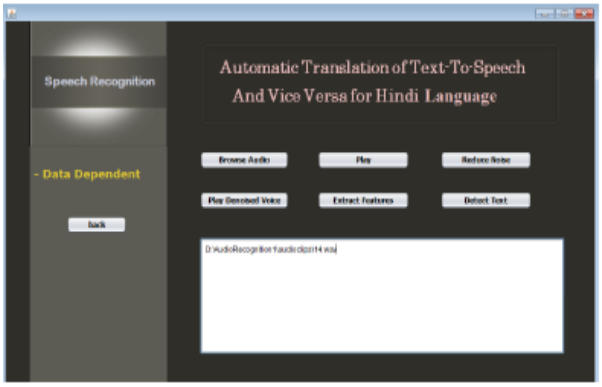


Figure 7.5: Shows .wav file selection.

7.5 Detect text for given input

When feature extraction successfully done then display the text successfully.

Figure 7.6 shows the text of proposed input for the word विफलता

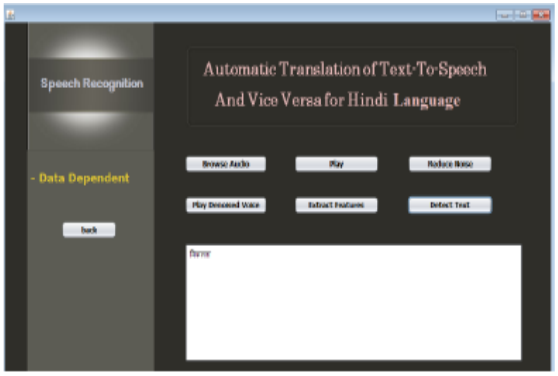


Figure 7.6: shows the text of proposed input for the word

7.6 Expression Detection System

In figure 7.6, shows expression detection system. Here we browse image and further steps which extract the relevant features successfully.

Figure 7.6: Shows the input image and feature extraction of image in system.

7.7 Display expression contain in image

In figure 8.11, shows expression contain in that image.

Figure 7.7: Shows the expression contain in given image-Surprise

7.8 Emotion Detection System

In figure 7.8, shows emotion detection system. Here we browse the audio file, play that audio file and remove noise successfully. Then we extract the features of that audio file.

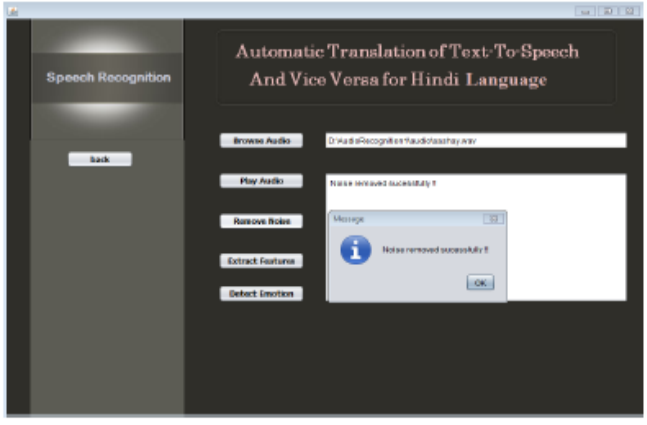


Figure 7.8: Shows browse the audio file.

In figure 7.9, shows emotion contain in the audio file.When feature extraction successfully done then display the emotion in audio file.

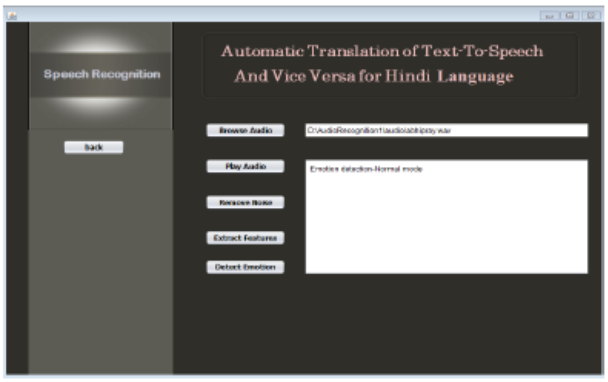


Figure 7.9: Shows the emotion contains in that audio file.

8. CONCLUSION

The proposed work may seem effective for text to speech and speech to text for HindiLanguage. User friendly interface for Hindi text to speech conversion provided. The most important qualities of a speech synthesis system are naturalness and intelligibility. Speech to text translation is useful for integrating people with hearing impairments in oral communication process. Expressive speech synthesis is a critical research and application area in the field ofmultimedia interfaces. Emotions are the spirit of expression.Recent advances in TTS will impact is wide number of disciplinesfrom education, business and entertainment applications to medical aids.

9. FUTURE SCOPE

All Indian language scripts have common phonetic base. Hence, a common userfriendly TTS can be supposed for all Indian languages which can run on various operatingsystems. The further work can be done to improve the naturalness and intelligibility of text tospeech. A Web based application can also be designed which can convert text in any Indianlanguages into speech.

10. ACKNOWLEDGMENTS

The authors would like to thank Chairman Groups and Management and the Director/Principal Dr. UttamKalwane, 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.

11. REFERENCES

[1] J. Jia, S. Zhang, F. Meng, Y.Wang, and L.Cai, “Emotional Audio-Visual Speech SynthesisBased on PAD”,IEEE transactions on audio, speech, andlanguage processing, vol. 19,no. 3,march 2011.

[2] D. Govind , S.R. MahadevaPrasanna ,“ Expressive speech synthesis: a review ”,Springer Science+BusinessMedia New York 2012.

[3] O. Turk and M. Schroder ,“ Evaluation of Expressive Speech Synthesis With Voice Conversionand Copy Resynthesis Techniques ”, IEEE Transactions On Audio, Speech, AndLanguage Processing, Vol. 18, No. 5, July 2010.

[4] B. Yegnanarayana, and K. Sri Rama Murty ,“ Event-Based Instantaneous FundamentalFrequency Estimation From Speech Signals ”, IEEE Transactions On Audio, Speech,And Language Processing, Vol. 17, No. 4, May 2009.

[5] M.Theune, K. Meijs, D. Heylen, and R. Ordelman , " Generating Expressive Speech forstorytelling Applications " , IEEE Transactions On Audio, Speech, And Language Processing,Vol. 14, No. 4, July 2006 .

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

[7]] K.Kamble and R.Kagalkar, "A Review: Translation of Text to Speech Conversion forHindi Language, International Journal of Science and Research (IJSR) Vol. 3 Issue 11,November 2014.

[8] K.Kamble and R.Kagalkar , " Audio Visual Speech Synthesis and Speech Recognition forHindi Language, International Journal of Computer Science and Information Technologies(IJCSIT) Vol. 6 Issue 2, April 2015.

[9] K.Kamble and R.Kagalkar , " A Novel Approach for Hindi Text Description to Speechand Expressive Speech Synthesis , International Journal of Applied Information Systems(IJAIS) Vol. 8 Issue 7 May 2015.

[10] S. Suryawanshi, R.Itkarkar, D.Mane , " High Quality Text to Speech Synthesizer usingPhonetic Integration " , International Journal of Advanced Research in Electronics andCommunication Engineering (IJARECE) Vol.3, Issue 2, February 2014.

[11] J.Sangeetha , S.Jothilakshmi , S.Sindhuja , V.Ramalingam , " Text to Speech synthesissystem for Tamil ", International Journal of Emerging Technology and Advanced Engineering(IJETAT)Vol.3, Special Issue 1, January 2013.

[12] M. Singh, K. Verma ," Text to Speech Synthesis for numerals into Punjabi language", International Journal of Computational Linguistics and Natural Language Processing(IJCLNLP)Vol.2,Issue 7 July 2013.

[13] N.Swetha, K.Anuradha , " Text-to-speech conversion ", International Journal of AdvancedTrends in Computer Science and Engineering,(IJATCSE) Vol.2 , Issue 6, November2013

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

[15] D.Sasirekha, E.Chandra ," Text to Speech: A Simple Tutorial, International Journal ofSoft Computing and Engineering (IJSCE) Vol.2, Issue-1, March 2012.

[16] S. Hertz, James Kadin, And Kevin J. Karplus,"The Delta Rule Development System forSpeech Synthesis from Text ", Proceedings Of The IEEE, Vol. 73, No. 11, November1985.

[17] L. Sahu and A.Dhole, " Hindi Telugu Text-to-Speech Synthesis (TTS) and inter-languagetext Conversion, International Journal of Scientific and Research Publications,(IJSRP)Volume 2, Issue 4, April 2012 .

[18] R. Segundo, J. Montero, R.Chicote, and J. Lorenzo ," Architecture for Text Normalizationusing Statistical Machine Translation techniques " , Springer-Verlag Berlin Heidelberg2011.

[19] 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.

[20] P.Shetake, A.Patil and M Jadhav, " Review of Text To Speech Conversion Methods, InternationalJournal of Industrial Electronics and Electrical Engineering,(IJIEE) ISSN:2347-6982 Volume-2, Issue-8, Aug.-2014.

[21] P.Rajput and P. Lehana," Investigations Of The Distributions Of Phonemic Durations InHindi And Dogri, International Journal on Natural Language Computing (IJNLC) Vol.2, No.1, February 2013.

[22] G. Rajadhyaksha, S.Mody and S.Venkateswar," Portable Text to Speech Convertor,International Journal of Emerging Technology and Advanced Engineering(IJETAE),Vol.3, Issue 8, August 2013.