

# Hidden Markov Model based Speech Synthesis: A Review

Sangramsing Kayte<sup>1</sup>  
Research Scholar  
Department of Computer  
Science & IT  
Dr. Babasaheb Ambedkar  
Marathwada University,  
Aurangabad

Monica Mundada<sup>1</sup>  
Research Scholar  
Department of Computer  
Science & IT  
Dr. Babasaheb Ambedkar  
Marathwada University,  
Aurangabad

Jayesh Gujrathi, PhD  
Department of chemistry,  
Pratap College, Amalner

## ABSTRACT

A Text-to-speech (TTS) synthesis system is the artificial production of human system. This paper reviews recent research advances in field of speech synthesis with related to statistical parametric approach to speech synthesis based on HMM. In this approach, Hidden Markov Model based Text to speech synthesis (HTS) is reviewed in brief. The HTS is based on the generation of an optimal parameter sequence from subword HMMs. The quality of HTS framework relies on the accurate description of the phoneset. The most attractive part of HTS system is the prosodic characteristics of the voice can be modified by simply varying the HMM parameters, thus reducing the large storage requirement.

## Keywords

TTS, speech corpus, Marathi phonemes.

## 1. INTRODUCTION

The primary task of Text-to-speech (TTS) synthesis is to translate input text into intelligible and natural sounding speech. The TTS components involves the two phases [1], the front end- which analyses the text, creates possible pronunciations for each word in the context with grapheme to phoneme conversion. The back end generates the speech waveform along with the prosody of the sentence to be spoken. The evaluation of TTS system is based on 3 attributes: Accuracy, Intelligibility and naturalness. The HTS system provides the frequency spectrum (Vocal tract), fundamental frequency (vocal source) and duration (Prosody) of speech, which are commonly modeled simultaneously by HMMs. Speech waveforms are generated from HMMs themselves based on maximum likelihood criterion [2]. The figure 1 shows architecture of TTS system [3].

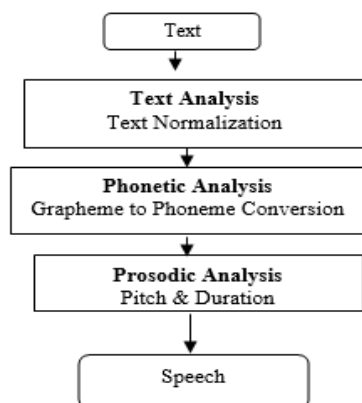


Figure1. Architecture of TTS system

The architecture of TTS system is divided into four modules as follows:

1. Text analysis: The normalization of the text wherein the numbers and symbols become words and an abbreviation are replaced by their whole words or phrases etc. The linguistic analysis which means syntactic and semantic analysis and aims at understanding the context of the text is also performed in this step. The statistical methods are used to find the most probable meaning of the utterances. This is significant because the pronunciation of a word may depend on its meaning and on the context.
2. Phonetic Analysis: This converts the orthographical symbols into phonological ones using a phonetic alphabet.
3. Prosodic Analysis: Prosody contains the rhythm of speech, stress patterns and intonation. The naturalness in speech is attributed to certain properties of the speech signal related to audible changes in pitch, loudness and syllabic length, collectively called prosody. Acoustically, these changes correspond to the variations in the fundamental frequency (F0), amplitude and duration of speech units [1].
4. Speech Synthesis: This block finally generates the speech signal. This can be achieved either based on parametric representation, in which phoneme realizations are produced by machine, or by selecting speech units from a database. The resulting short units of speech are joined together to produce the final speech signal.

## 2. LITERATURE SURVEY

All Paul Taylor explained the basic concepts of speech and signal processing in “Text-to-Speech Synthesis” [5]. Basics of speech synthesis and speech synthesis methods are discussed in “An Introduction to Text-to-Speech Synthesis” by Thierry Dutoit [1]. Text to speech system organization, functions of each module and conversion of text which is given as input in to speech is clearly explained in this book. Different speech synthesis methods, which are used for development of synthesis system, are explained in “Review of methods of Speech Synthesis” [6]. The process of text normalization is understood from “Normalization of non standard words” by Christopher Richards. In this book, conversion of non-standard words (NSWs) in to standard words is explained clearly with examples. Non-standard words are the words which are not found in the dictionary. NSWs are tokens that need to be expanded into an appropriate orthographic form before the text-to-phoneme module [7]. The concept of prosody is explained briefly in “Text to speech synthesis with prosody feature” by M. B. Chandak. This book explains how

the prosody is predicted from the text which is given as input. In linguistics, prosody includes the intonation, rhythm and lexical stress in speech. The prosodic features of a unit of speech can be grouped into syllable, word, phrase or clause level features are largely determined by the naturalness of the prosody generated during synthesis. The correct prosody also has an important role in the intelligibility of synthetic speech. Prosody also conveys paralinguistic information to the user such as joy or anger. In speech synthesis system, intonation and other prosodic aspects must be generated from the plain textual input. Paper by Ramani Boothalingam (2013) presented the comparison of the performance of Unit Selection based Synthesis (USS) and HMM based speech synthesizer. Difference between the two major speech synthesis techniques namely unit selection based synthesis and HMM based speech synthesis is explained clearly in this paper [9]. Unit selection systems usually select from a finite set of units in the speech database and try to find the best path through the given set of units. When there are no examples of units that would be relatively close to the target units, the

situation can be viewed either as lacking in the database coverage or that desired sentence to be synthesized is not in the domain of the TTS system. To achieve good quality synthesis, the speech unit database should have good unit coverage. To obtain various voice characteristics in TTS systems based on the selection and concatenation of acoustical units, a large amount of speech data is needed. It is very difficult to collect and segment large amount of speech data for different languages.. These features are manifested as duration, F0 and intensity [8]. Prosodic units do not need to necessarily correspond to grammatical units. In HMM based speech synthesis system, the parameters of speech are modeled simultaneously by HMMs. During the actual synthesis, speech waveforms are generated from HMMs themselves based on maximum likelihood criteria. The advantage of using HMM based speech synthesis technique for developing TTS system is discussed briefly in “An Overview of Nitech HMM-based Speech Synthesis System for Blizzard Challenge 2005” by Heiga Zen and Tomoki Toda. The concept of accuracy measurement is outlined. The parameters on which the accuracy of a TTS system depends are discussed [10]. The concept of HMM’s explained clearly in “Text-to-speech synthesis” by Paul Taylor. This book explains the concepts of Hidden Markov Model and how they are used in synthesis process. HMM themselves are quite general models and although developed for speech recognition have been used for many tasks in speech and language technology and are now in fact one of the fundamental techniques in speech technology applications.

Text to Speech(TTS) is a system in which sequence of words are taken as input and converts them to speech. In conversion process of speech synthesis method, vowels and consonants are most important in marathi language [1]. Each phonemes are made of combination of consonants and vowels. There are different concatenation methods like unit selection,diaphone or domain specific method for speech synthesis. In all these methods the voices are sampled from real recorded speech and speech synthesis is handled by computers. Here researchers main focus is to designing a database of phonemes in such a way that it speedup searching and retrieving process in marathi TTS.

### 3. HIDDEN MARKOV MODEL (HMM) BASED SPEECH SYNTHESIS

In the early 1970s, Lenny Baum of Princeton University invented a mathematical approach to recognize speech called Hidden markov model (HMM). The Hidden markov model (HMM) [11] [12] [13] is a doubly stochastic process that produces a sequence of operations. Table 1 compares the HMM based speech synthesis system and Unit selection.

**Table1. Compares the HMM based speech synthesis and Unit selection**

HMM based	Unit selection
Statistics based	Multi template based
Clustering( Use of HMM)	Clustering( possible use of HMM)
Multiple tree	Single tree
Advantage: Smooth , stable	Advantage: High quality at waveform level
Disadvantage: Vocoded speech (buzzy)	Disadvantage: Discontinuity or miss
Small run time data	Large run time data
Various voices	Fixed voices

#### 3.1 Speech Synthesis and Development for Indian languages

India is a multilingual society with 1652 dialects/native languages. Speech technologies can play a very important role in development of applications for common people. Prior 1990s, Indian speech synthesizers were research synthesizers, generating small segments of speech in non-real time and the progress was very slow. Speech synthesizers were not developed for commercial purpose. In the 90s, Government of India had funded Indian language projects generously, through Technology Development for Indian Languages (TDIL) and other schemes [42].

#### 3.2 Current Research projects for speech synthesis in India:

Some of the institutions in India are engaged in speech synthesis. The IIT Madras has worked on a novel scheme where the “unit” is a character of written “text”. The Tata Institute of Fundamental Research (TIFR), Mumbai has reported unlimited continuous speech synthesizer using formant synthesis technique. Whereas TIFR [14] and Central Electronics Engineering Research Institute (CEERI) [15] worked with formant synthesis, ISI, Kolkata [16], Indian Institute of Information Technology (IIIT), Hyderabad [17], and center for Development of Advanced Computing (CDAC), Pune and Kolkata developed concatenation-based synthesizers. Between the concatenation and formant synthesizers, the quality obtained so far is comparable. Speech synthesizers based on Festival has been developed in languages including Hindi, Bangla, Kannada, Marathi and Tamil.

### 3.3 Speech Corpora developed by the LDC-IL

Linguistic Data Consortium for Indian Languages (LDCIL) is the Consortium responsible to create the database and shall provide forum for the researchers all over the world to develop speech application using the collected data in various domains. The LDC-IL has collected Speech databases in various Indian languages, the details are described in [18]. The research that has been carried out is mostly for text to speech synthesis which uses phoneme/syllables concatenation on isolated words and is either based either on concatenative or formant synthesis techniques. The need of the hour is to work on the continuous speech and apply latest techniques such as Hidden Markov Models for development of T-T-S for general purpose or limited domain to achieve true application potentials of speech synthesis. Although Indian language speech synthesis has come up a long way, the amount of work for Indian languages in speech domain has not yet reached to a critical level to be used as real communication tool, as that in other languages of developed countries.

### 3.4 Current development of HMM –based speech synthesis system for Marathi (HTS)

Marathi language consists of 33 consonants and 12 vowels. The monophone HMM's designed to build the phoneset for the language [19]. The lexicon describes the set of words known by the system and their pronunciation. In HMM, based speech synthesis, the speech parameters of a speech unit such as spectrum, fundamental frequency ( $f_0$ ) and phoneme duration are statistically generated by using HMM based on maximum likelihood criterion. The most attractive part of HTS system is that its voice characteristics, speaking styles, or emotions can easily be modified by transforming HMM parameters using various techniques such as adaptation, interpolation, Eigen voice, Or multiple Regression [20]. HTS i.e. HMM based text to speech synthesis system is an open source tool which provides a research and development platform for statistical parametric speech synthesis [21]. The HMM-based speech synthesis system (HTS) has been developed by the HTS working group as an extension of the HMM toolkit (HTK). The source code of HTS is released as a patch for HTK. The first version 1.0 HTS was first released in December 2002. After an interval of three years, HTS version 2.0 was released in December 2006 with major update and inclusion of number of new features, such as introduction of global mean and variance calculation tool, for large databases the previous version often suffered from numerical errors. HTS version 2.0.1 was a bug –fixed version and the latest version, HTS version 2.1, was released in July 2008. This version includes important features; Hidden semi-markov models (HSMMs) [22][23], the speech parameter generation algorithm considering global variance (GV) [24], advanced adaptation techniques [25], and stable version of run time synthesis engine API. The HTS version 2.1, with the STRAIGHT analysis/synthesis techniques [26], provides the ability to construct the state-of-art HMM based speech synthesis systems developed for the past Blizzard Challenge events [27][28].

### 3.5 Architecture of typical HMM based speech synthesis system

The main advantages of the referred HMM –based synthesis techniques when compared with unit selection and concatenation method is the fact that the voice alteration can be performed without large databases, being at par with

quality with unit selection and concatenation ones. The figure 2 shows the architecture of HMM based speech synthesis system [29]. In the training part, spectrum and excitation parameters are extracted from speech database and modeled by context dependent HMMs. In the synthesis part, context dependent HMMs are concatenated according to the text to be synthesized. Then spectrum and excitation parameters are generated from the HMM by using a speech parameter generation algorithm. Finally, the excitation generation module and synthesis filter module synthesize speech waveform using the generated excitation and spectrum parameters. The training part performs the maximum likelihood estimation by using the Expectation Maximization (EM) algorithm [30]. In this process, spectrum (e.g., mel-cepstral coefficients) [31] and their delta and delta-delta coefficients and excitation (e.g., log  $F_0$  and its dynamic features) parameters are extracted from a database of natural speech and modeled by a set of multi-stream [32] context-dependent HMMs (phonetic, linguistic, and prosodic contexts being taken into account).

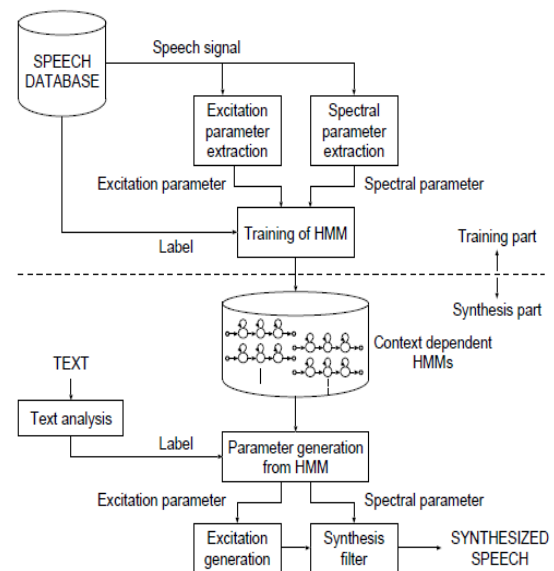


Figure2. Typical architecture of HMM based speech synthesis system

In the temporal structure of speech, each HMM has its state-duration distribution namely, the Gaussian distribution [33] and the Gamma distribution [34]. They are estimated from statistical variables obtained at the last iteration of the forward-backward algorithm. As they have their own context dependency, each of spectrum, excitation, and duration is clustered individually by using phonetic decision trees [35]. Hence, the system can model the spectrum, excitation, and duration in a unified framework. In the synthesis part, a given word sequence is converted into a context dependent label sequence, and then the utterance HMM is constructed by concatenating the context-dependent HMMs according to the label sequence. Then, various kinds of speech parameter generation algorithm [36] [37] have been used to generate the spectrum and excitation parameters HMM. Finally, the excitation generation module and synthesis filter module filter, such as Mel log spectrum approximation (MLSA) filter [38] synthesize speech waveform using the generated excitation and spectrum parameters.

## **4. CONS AND PROS OF HMM BASED SPEECH SYNTHESIS SYSTEM**

### **4.1 Advantages:**

The main advantage of statistical parametric synthesis is that it can synthesize speech with various voice characteristics such as speaker individualities, speaking styles, and emotions etc. The combination of unit-selection and voice-conversion (VC) techniques [39] can alleviate this problem but high-quality voice conversion is still difficult. However, we can easily change voice characteristics, speaking styles, and emotions in statistical parametric synthesis by transforming its model parameters. There are three major techniques to achieve this, namely adaptation, interpolation, and Eigen voices.

### **4.2 Disadvantages:**

Although the operation and advantages of statistical parameter speech synthesis is impressive, a few disadvantages are associated with it. First, the parameters must be automatically derivable from databases of natural speech. Second the parameters must give rise to high quality synthesis; finally, the parameters must be predictable from text; the synthesis quality is intelligible but nowhere close to natural speech.

## **5. CONCLUSION**

Synthetic speech is in progress from last few decades. The study presented an overview of speech synthesis-past progress and current trends. The three basic methods for synthesis are the formant, concatenative, and articulatory synthesis. The formant synthesis is based on the modeling of the resonances in the vocal tract and is perhaps the most commonly used during last decades. However, the concatenative synthesis which is based on playing prerecorded samples from natural speech is more popular. In theory, the most accurate method is articulatory synthesis which models the human speech production system directly, but it is also the most difficult approach. Currently, the statistical parametric speech synthesis has been the most rigorously studied approach for speech synthesis. We can see that statistical parametric synthesis offers a wide range of techniques to improve spoken output. Its more complex models, when compared to unit-selection synthesis, allow for general solutions, without necessarily requiring recorded speech in any phonetic or prosodic contexts. The unit-selection synthesis requires very large databases to cover examples of all required prosodic, phonetic, and stylistic variations which are difficult to collect and store. In contrast, statistical parametric synthesis enables models to be combined and adapted and thus does not require instances of any possible combinations of contexts. Additionally, T-T-S systems are limited by several factors that present new challenges to researchers. They are 1) The available speech data are not perfectly clean 2) The recording conditions are not consistent & 3) Phonetic balance of material is not ideal. Means to rapidly adapt the system using as little data as a few sentences would appear to be an interesting research direction. It is seen that synthesis quality of statistical parametric speech synthesis is fully understandable but has "processed quality" to it [43]. Control over voice quality (naturalness, intelligibility) is important for speech synthesis applications and is a challenge to the researchers. As described in this review, unit selection and statistical parametric synthesis approaches have their own advantages and drawbacks. However, by proper combination of the two approaches, a third approach could be generated which can retain the advantages of the HMM based and corpus based synthesis with an objective to generate synthetic

speech very close to the natural speech. It is suggested that a more detailed evaluation and analysis, plus integration of HMM based segmentation and labeling for building database and HMM based search for selecting best suitable units shall aid in using the better features of the two methods.

## **6. REFERENCES**

- [1] T. Dutoit, "An Introduction to Text-to-Speech Synthesis", Kluwer Academic Publishers, 1997.
- [2] Black, A. Zen, H., Tokuda, K. "Statistical Parametric Synthesis", in proc. ICASSP, Honolulu, USA, 2007.
- [3] X.Huang, A.Acero, H.-W. Hon, "Spoken Language Processing", Prentice Hall PTR, 2001.
- [4] D. Jurafsky and J.H. Martin, "Speech and Language Processing", Pearson Education, 2000.
- [5] Paul Taylor, "Text to Speech Synthesis", University of Cambridge, pp.442-446.
- [6] Newton, "Review of methods of Speech Synthesis", M.Tech Credit Seminar Report, Electronic Systems Group, November, 2011, pp. 1-15
- [7] Christopher Richards, "Normalization of non-standard words". *Computer Speech and Language* (2001), pp.287-333
- [8] M.B.Chandak, Dr.R.V.Dharaskar and Dr.V.M.Thakre,"Text to Speech with Prosody Feature: Implementation of Emotion in Speech Output using Forward Parsing", *International Journal of Computer science and Security*, Volume (4), Issue (3)
- [9] Ramani Boothalingam,V Sherlin Solomi, Anushiya Rachel Gladston,S Lilly Christina, "Development and Evaluation of Unit Selection and HMM-Based Speech Synthesis Systems for Tamil", 978-1-4673-5952-8/13, IEEE 2013 National Conference
- [10] Heiga Zen, Tomoki Toda and Keiichi Tokuda. "The Nitech-NAIST HMM-based speech synthesis system for the Blizzard Challenge 2006", *INTERSPEECH* 2005.
- [11] J. Ferguson, Ed., "Hidden Markov Models for speech" IDA, Princeton, NJ, 1980
- [12] L.R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition" *Proc. IEEE*, 77(2), pp.257-286, 1989
- [13] L.R.Rabiner and B.H. Juang, "Fundamentals of speech recognition", Prentice-Hall, Englewood Cliff, New Jersey, 1993.
- [14] Furtado X A & Sen A, "Synthesis of unlimited speech in Indian Languages using formant-based rules" *Sadhana*, 1996, pp 345-362 .
- [15] Agrawal S S & Stevens K, "Towards synthesis of Hindi consonants using KLSYN88", *Proc ICSLP92*, Canada, 1992, pp.177-180 .
- [16] Dan T K, Datta A K & Mukherjee, B, "Speech synthesis using signal concatenation", *J ASI*, vol. XVIII (3&4), 1995, pp 141-145 .
- [17] Kishore S. P., Kumar R & Sanghal R, "A data driven synthesis approach for Indian language using syllable as basic unit", *Proc ICON 2002*, Mumbai, 2002 .

- [18] Agrawal S. S. 2010, "Recent Developments in Speech Corpora in Indian Languages: Country Report of India", O-COCOSDA, Nepal.
- [19] B. Ramani, S.Lilly Christina, G Anushiya Rachel, V Sherlin Solomi, Mahesh Kumar Nandwana, Anusha Prakash., Aswin Shanmugam S, Raghava Krishnan, S P Kishore, K Samudravijaya, P Vijayalakshmi, T Nagarajan and Hema A Murthy." A Common Attribute based Unified HTS framework for Speech Synthesis in Indian Languages". 8th ISCA Speech Synthesis Workshop. August 31 – September 2, 2013, Barcelona, Spain
- [20] Zen HeigaNose, Takashi Yamagishi, Junichi Sako, Shinji Masuko, Takashi, Black, Alan W." The HMM-based speech synthesis system (HTS) version 2.0". 6th ISCA Workshop on Speech Synthesis, Bonn, Germany, August 22-24, 2007.
- [21] K. Tokuda , H. Zen, J. Yamagishi, T. Masuko, S. Sako, T. Toda, A.W. Black, T. Nose , and K. Oura, "The HMM based synthesis system(HTS)" <http://hts.sp.nitech.ac.jp/>.
- [22] H. Zen, K. Tokuda, T. Masuko, T. Kobayashi and T. Kitamura, "A hidden semi-Markov model-based speech synthesis system." *IEICE Trans. Inf.Syst.*, E90-D (5):825–834, 2007.
- [23] J. Yamagishi and T. Kobayashi. Average-voice based speech synthesis using HSMM-based speaker adaptation and adaptive training. *IEICE Trans. Inf. Syst.*, E90-D (2):533–543, 2007.
- [24] J. Yamagishi, T. Kobayashi, Y. Nakano, K. Ogata, and J. Isogai, "Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm", *IEEE Trans. Audio Speech Lang. Process.*, 17(1), pp.66–83, 2009.
- [25] H. Kawahara, I. Masuda-Katsuse, and A.de Cheveigné, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds", *Speech Comm.*, 27:187–207, 1999.
- [26] H. Zen, T. Toda, M. Nakamura, and K. Tokuda, "Details of Nitech HMM based speech synthesis system for the Blizzard Challenge 2005. *IEICE Trans. Inf. Syst.*, E90-D(1):325–333, Jan. 2007.
- [27] H. Zen, T. Toda, and K. Tokuda, "The Nitech-NAIST HMM-based speech synthesis system for the Blizzard Challenge 2006", In *Blizzard Challenge Workshop*, 2006.
- [28] J. Yamagishi, T. Nose, H. Zen, Z.-H. Ling, T. Toda, K. Tokuda, S. King, and S. Renals, "A robust speaker-adaptive HMM-based text-to-speech synthesis", *IEEE Trans. Audio Speech Lang. Process.*, 2009. (accept for publication).
- [29] T.Yoshimura, K.Tokuda, T. Masuko, T. Kobayashi and T. Kitamura, "Simultaneous Modeling of Spectrum, Pitch and Duration in HMM-Based Speech Synthesis" In *Proc. of ICASSP 2000*, vol 3, pp.1315-1318, June 2000.
- [30] Dempster, A., Laird, N., Rubin, D., 1977, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of Royal Statistics Society* 39, 1–38.
- [31] Fukada, T., Tokuda, K., Kobayashi, T., Imai, S., 1992, "An adaptive algorithm for mel-cepstral analysis of speech", In *Proc. ICASSP*. pp. 137–140.
- [32] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X.-Y., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P., 2006, "The Hidden Markov Model Toolkit (HTK) version 3.4. <http://htk.eng.cam.ac.uk/>.
- [33] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T. 1998, "Duration modeling for HMM-based speech synthesis", In *Proc. ICSLP*. pp. 29–32.
- [34] Ishimatsu, Y., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T., 2001, "Investigation of state duration model based on gamma distribution for HMM based speech synthesis", In *Tech. Rep. of IEICE*. vol. 101 of SP 2001-81. pp. 57–62, (In Japanese).
- [35] Odell, J., 1995, "The use of context in large vocabulary speech recognition", Ph.D. thesis, University of Cambridge.
- [36] Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T., Kitamura, T., 2000, "Speech parameter generation algorithms for HMM-based speech synthesis" In *Proc. ICASSP*. pp. 1315–1318.
- [37] Tachiwa, W., Furu, S., "A study of speech synthesis using HMMs" In: *Proc. Spring Meeting of ASJ*. pp. 239–240, (In Japanese), 1999.
- [38] Imai, S., Sumita, K., Furuichi, C., "Mel log spectrum approximation (MLSA) filter for speech synthesis", *Electronics and Communications in Japan* 66 (2), 10–18, 1983 .
- [39] Stylianou, Y., Cappe, O., Moulines, E., 1998, "Continuous probabilistic transform for voice conversion", *IEEE Trans. Speech Audio Process.* 6 (2), 131–142.
- [40] Sangramsing Kayte , Kavita Waghmare , Dr. Bharti Gawali "Marathi Speech Synthesis: A review" *International Journal on Recent and Innovation Trends in Computing and Communication* ISSN: 2321-8169 Volume: 3 Issue: 6
- [41] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" Monica Mundada et al, / (IJCSIT) *International Journal of Computer Science and Information Technologies*, Vol. 5 (5) , 2014, 6764-6767
- [42] Monica Mundada, Sangramsing Kayte, Dr. Bharti Gawali "Classification of Fluent and Dysfluent Speech Using KNN Classifier" *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 4, Issue 9, September 2014
- [43] Sangramsing Kayte, Monica Mundada "Study of Marathi Phones for Synthesis of Marathi Speech from Text" *International Journal of Emerging Research in Management & Technology* ISSN: 2278-9359 (Volume-4, Issue-10) October 2015.