

Performance Evaluation of Speech Synthesis Techniques for Marathi Language

Sangramsing Kayte
Research Scholar
Department of Computer
Science & IT
Dr. Babasaheb Ambedkar
Marathwada University,
Aurangabad.

Monica Mundada
Research Scholar
Department of Computer
Science & IT
Dr. Babasaheb Ambedkar
Marathwada University,
Aurangabad.

Charansing Kayte, PhD
Assistant Professor of Digital
and Cyber Forensic,
Maharashtra.

ABSTRACT

Text to speech synthesis (TTS) is the production of artificial speech by a machine for the given text as input. The speech synthesis can be achieved by concatenation and Hidden Markov Model techniques. The voice synthesized by these techniques should be evaluated for quality. The study extends towards the comparative analysis for quality of speech synthesis using hidden markov model and unit selection approach. The quality of synthesized speech is analyzed for subjective measurement using mean opinion score and objective measurement based on mean square score and peak signal-to-noise ratio (PSNR). The quality is also accessed by Mel-frequency cepstral coefficient features for synthesized speech. The experimental analysis shows that unit selection method results in better synthesized voice than hidden markov model.

Keywords

Keyword TTS, MOS, HMM, Unit Selection, Mean, Variance.

1. INTRODUCTION

A speech synthesis system is a computer-based system that produce speech automatically, through a grapheme-to-phoneme transcription of the sentences and prosodic features to utter. The synthetic speech is generated with the available phones and prosodic features from training speech database [1, 2]. The speech units is classified into phonemes, diaphones and syllables. The output of speech synthesis system differs in the size of the stored speech units and output is generated with execution of different methods. A text-to-speech system is composed of two parts: a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent words. This process is often called text normalization, preprocessing, or tokenization. Second task is to assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units like phrases, clauses, and sentences. Although text-to-speech systems have improved over the past few years, some challenges still exist. The back end phase produces the synthesis of the particular speech with the use of output provided from the front end. The symbolic representations from first step are converted into sound speech and the pitch contour, phoneme durations and prosody are incorporated into the synthesized speech.

The paper is structured in five sections. The techniques of speech synthesis are described in section 2. Database for synthesis system is explained in section 3. Section 4 explains speech quality measurement. Section 5 is dedicated with experimental analysis followed by conclusion.

2. CONCATENATE SYNTHESIS

Concatenate speech synthesis is a method where speech is generated by concatenating speech units one after the other as per the requirement. There are three different types of concatenate speech synthesis they are Domain specific synthesis, Di-phone synthesis and Unit selection synthesis [2]. The focus of the paper is Unit selection Synthesis.

In this method the database is built up with all phones present in the particular language. The design of such database includes well labeled phones with high quality utterances. The synthesized speech output signal generated with the concatenated parts from the database [2] [16][17]. The synthesized speech naturally sounds equal to the recorded utterances with respect to intonation, emotion and style. Nevertheless unit selection synthesis has shown itself to be capable of producing high quality natural sounding synthetic speech when constructed from large databases of well-ordered and well-labeled speech. The USS(Unit Selection Synthesis) extracts the prosodic and spectral part from input speech signal during the training part. In synthesis part, the analysis of text is done and prosody is incorporated with the use of algorithm and artificial speech is produced [2] [16][17]. In USS, initially the text is converted into phones of the particular segment. Then the phones are assigned the labels like vowels, semivowels and consonants. With the help of acoustic trees the ID's are generated for the given input. At the final step, the speech is synthesized with the USS algorithm with the incorporation of the needed prosody elements.

3. HIDDEN MARKOV MODEL BASED SPEECH SYNTHESIS

Please Hidden Markov Model synthesis is also called statistical parametric synthesis of speech. It is similar to the unit selection synthesis, trained with natural speech, however due to the parametric model approach; it allows a better modeling of variation. The work can be divided into two main parts which represent the topology of the synthesizer itself. For the synthesizer parametric models have to be trained first in order to estimate parameters from the models. The HTS system simultaneously models excitation and duration of speech using context dependent hidden markov model 's and thus generate speech waveforms from HMM's themselves [6]. The main advantages of the referred hidden markov model – based synthesis techniques is that the voice alteration can be performed without large databases, being at par with quality[2][16][17].

In the training part, spectrum and excitation parameters are extracted from speech database and modeled by context dependent hidden markov model. In the synthesis part,

context dependent hidden markov model are concatenated according to the text to be synthesized. Then spectrum and excitation parameters are generated from the hidden markov model by using a speech parameter generation algorithm. Both the Techniques are implemented using Festival framework [7][20].

4. SPEECH QUALITY MEASUREMENT

In this section, a brief overview of subjective and objective speech quality measurement methods is presented.

4.1 Subjective Quality Measure

Speech quality measure is the result of a subjective perception-and-judgment process. In this method a listener compares the perceptual event (speech signal heard) to an internal reference of what is judged to be of good quality. Subjective assessment plays a significant role in characterizing the quality of synthesis speech, as it attempts to quantify the end user's experience with the system under test. In the subjective quality measurement mean opinion score (MOS) technique was used. The mean opinion score (MOS) test is used in which listeners are asked to rate the quality of a speech signal on a 5-point scale, with 1 corresponding to unsatisfactory speech quality and 5 corresponding to excellent speech quality [8],[9][17][20].

4.2 Objective Quality Measure

Objective speech quality measurement involves the listener with the computational algorithm, thus facilitating automated real-time quality measurement. Real-time quality monitoring and control on a network-wide scale is achieved only with the objective speech quality measurement. Objective measurement methods aim to deliver quality estimates that are highly correlated with those obtained from subjective listening experiments. In the objective quality measure mean square error (MSE) and peak signal-to-noise ratio (PSNR) techniques were used.

a) Mean Square Error (MSE)

The mean squared error (MSE) measures the average of the squares of the errors, that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimation of speech synthesis [10][16].

b) Peak Signal to Noise Ratio(PSNR)

Peak signal-to-noise ratio (PSNR) is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the quality of its representation. PSNR is usually expressed in terms of the logarithmic decibel scale. PSNR is most commonly used to measure the quality of reconstruction of signal and image. The signal in this case is the original data, and the noise is the error introduced by synthesis [11].

4.3 Signal based Quality Measure

In the signal based quality measure Perceptual Evaluation of Speech Quality (PESQ) is robust and dynamic technique [12, 13, 14]. ITU-T Recommendation P.862 (better known as perceptual evaluation of speech quality, PESQ) is the current state-of-the-art standard measurement algorithm [15]. For this experiment we proposed MFCC Features for the signal based quality measure. Mel Frequency Cepstral Coefficients (MFCC) technique is robust and dynamic technique for

speech feature extraction [16]. The fundamental frequency, prosodic, energy variation in the syllable and many other features are studied with MFCC feature set. For the quality measure we extracted 13 features from synthesized speech and original speech file.

5. SPEECH DATABASE

The speech database collected for this experiment includes the sentences from philosophy and short stories. The sentences were recorded by male and female speaker. Male speaker was with south Indian accent and female voice was with normal accent. The male and female both were from academic field and practiced the session. The recording was done in noise free environment. The speech signal was sampled at 16 KHz. The set of 30 sentences were synthesized using unit selection and hidden Markov model. Noise free lab environment with multimedia laptop speaker was used to play these utterances to the post graduate students. The students were of age group 22 to 25, with no speech synthesis experience [18].

6. EXPERIMENTAL ANALYSIS

6.1 Analysis of Mean Opinion Score (MOS)

MOS is calculated for subjective quality measurement. It is calculated for the synthesized speech using the Unit selection synthesis and HMM approach. It was counseled to the listeners that they have to score between 01 to 05 (Excellent – 05 Very good – 04 Good – 03 Satisfactory – 02 Not understandable-01) for understandable. The mean of the scores given by each individual subject for ten sentences of the Unit selection approach is shown in table 1. The detail MOS score obtained from HMM speech synthesis method for ten sentences are shown in table 2.

The mean and variance of the score obtained according to the subject using unit selection and HMM based speech synthesis approach is shown in table 3.

It is observed that from table 3 and table 4 mean scores increases with the increase in the syllable coverage

Table 1. Unit selection speech synthesis of the scores given by each subject for each synthesis system

Subject Sentence	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9	Sub 10
1	5	5	5	5	5	5	5	4	5	5
2	5	5	4	5	5	5	5	4	5	5
3	5	5	5	4	3	5	5	2	5	4
4	5	5	5	5	5	5	5	5	5	5
5	5	5	5	5	5	5	5	3	3	5
6	5	5	5	5	5	5	5	5	5	5
7	5	5	4	5	5	5	5	5	5	5
8	5	4	5	4	5	5	4	5	5	5
9	5	3	5	5	3	5	5	3	5	5
10	5	5	5	5	5	5	5	5	5	5

Table 2. HMM-based speech synthesis of the scores given by each subject for each synthesis system

Sub Sentence	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9	Sub 10
1	5	5	5	5	5	4	5	4	5	3
2	4	5	4	4	3	4	3	3	3	4
3	5	4	4	4	4	3	4	4	3	5
4	5	4	4	4	3	4	4	5	4	5
5	3	4	5	5	5	3	4	4	5	4
6	2	4	4	3	4	2	4	5	4	4
7	3	5	5	2	5	1	5	3	3	5
8	4	4	4	3	4	2	5	4	2	3
9	4	3	3	3	5	3	4	5	2	4
10	5	4	4	5	2	4	4	4	4	5

Table 3. Mean and variance of the scores obtained across the subjects from unit selection and HMM approach

Subject	Unit Selection Method		HMM synthesis approach	
	Mean Score	Variance	Mean Score	Variance
Sub 1	4.56	0.25	3.90	1.19
Sub 2	4.23	0.52	2.73	1.71
Sub 3	4.03	0.79	2.56	1.35
Sub 4	4.56	0.25	2.80	1.61
Sub 5	4.10	0.43	2.46	0.947
Sub 6	4.03	0.37	3.10	1.05
Sub 7	4.56	0.25	2.80	1.68
Sub 8	3.96	0.72	2.33	1.26
Sub 9	4.16	0.62	2.73	1.37
Sub 10	4.63	0.24	2.63	1.48

a) PSNR and MSE Quality Measure

The PSNR and MSE method was used for subjective quality measure of speech synthesis based on hidden Markov model and unit selection approach. Table 4 represents the MSE and PSNR values for unit selection based speech synthesis. HMM based speech synthesis using MSSE and PSNR is shown in table 5.

Table 4: MSE and PSNR values for unit selection based speech synthesis

Sr.No	Original Speech File	Synthesized File	M.S.E	P.S.N.R
1	mar_001	mar_001	7.94	3.30
2	mar_002	mar_002	4.57	6.72
3	mar_003	mar_003	1.02	3.21
4	mar_004	mar_004	3.70	4.20
5	mar_005	mar_005	7.61	2.57
6	mar_006	mar_006	5.32	1.26
7	mar_007	mar_007	8.06	7.56

8	mar_008	mar_008	7.20	1.29
9	mar_009	mar_009	9.25	3.24
10	mar_010	mar_010	7.01	4.08
Average			5.168	4.743
Quality (100-Average)			94.83	96.26

Table 5: MSE and PSNR values for Hidden Markov Model speech synthesis

Sr.No	Original Speech File	Synthesized File	M.S.E	P.S.N.R
1	mar_01	mar_01	9.15	7.315
2	mar_02	mar_02	8.38	6.24
3	mar_03	mar_03	13.5	5.25
4	mar_04	mar_04	10.4	8.40
5	mar_05	mar_05	9.26	8.76
6	mar_06	mar_06	9.38	9.42
7	mar_07	mar_07	10.10	9.05
8	mar_08	mar_08	9.63	6.56
9	mar_09	mar_09	10.42	8.49
10	mar_010	mar_010	12.40	7.44
Average			10.26	7.69
Quality (100-Average)			82.73	92.31

The table below shows the comparative performance of both Unit and HMM for accent recognition using MFCC, MSE and PSNR techniques.

Table 6. Comparative result of Unit and HMM speech synthesis

Sr.No	Approach of Synthesis	MFCC Mean (%)	MFCC Std (%)	MFCC Var(%)	MSE (%)	PSNR (%)
1	HMM	87	84	84	94.73	94.31
2	Unit Selection	97	91	86	97.83	97.26

From the table 6, it is observed that the unit selection based accent identification gives a better performance than HMM based speech synthesis.

b) Mel Frequency Cepstral Coefficients

The signal based synthesis quality measure is experimented for unit selection and Hidden Markov Model based speech synthesis. For the performance variation, we calculated the mean, standard deviation and variance as a statistical measure. The detail sentences and label used for the unit selection based speech synthesis is described in table 7.

Table 7: Sentences and label used for unit selection based speech synthesis

Sr.No	The Original Sentence	Label Used for Original Speech File	Label Used for Synthesis Speech File
1	" कारण आपल्याकडे ती पद्धत नाही	mar_001	mar_001
2	त्याने पहिल्या व दुसऱ्या महायुद्धात	mar_002	mar_002
3	तुम्हाला त्यातून बराच अंदाज येईल	mar_003	mar_003

4	सिंह रास सिंह नावाची खगोलीय रास	mar_004	mar_004
5	अशा प्रकारे अमेरिकेचा या युद्धात प्रवेश झाला	mar_005	mar_005
6	रत्नागिरी जिल्ह्याच्या माहितीसाठी येथे टिचकी द्या	mar_006	mar_006
7	पुणे शहरातील एक मध्यवर्ती ठिकाण	mar_007	mar_007
8	जलंधर शहरातील क्रिकेटचे मैदान आहे	mar_008	mar_008

9	आपण चर्चा करताना कृपया चार वापरून सही करावी	mar_009	mar_009
10	केवळ प्रबंधक ही पाने बदलू शकतात काय "	mar_010	mar_010

The MFCC-mean based performance of unit selection based synthesis is shown in table 8. Table 9 represents the detail of standard deviation of MFCC for unit selection speech synthesis.

Table 8: The performance of MFCC-Mean based unit selection speech synthesis

Synthesized Speech											
Original Speech Signal		mar_0001	mar_0002	mar_0003	mar_0004	mar_0005	mar_0006	mar_0007	mar_0008	mar_0009	mar_0010
mar_001	0.120	3.242	2.31	1.392	3.42	3.008	2.983	5.094	7.01	1.234	
mar_002	1.281	0.009	2.111	2.453	7.632	1.90	4.02	1.223	8.01	3.04	
mar_003	1.453	3.21	0.080	3.25	1.99	2.843	3.921	2.963	2.093	6.70	
mar_004	3.02	1.230	2.564	0.899	2.786	5.453	1.672	1.981	2.67	3.45	
mar_005	2.40	1.450	5.432	2.932	0.021	3.921	6.05	4.675	7.00	3.674	
mar_006	3.896	8.09	3.983	2.732	3.674	0.673	2.843	5.03	3.894	4.92	
mar_007	1.893	1.563	2.03	2.100	4.92	3.67	1.460	1.273	3.521	7.38	
mar_008	2.932	3.674	2.732	3.721	3.567	2.732	3.876	0.783	2.673	3.643	
mar_009	1.776	2.732	4.332	5.893	2.783	3.874	2.743	4.87	1.091	3.021	
mar_010	2.873	2.983	1.873	1.90	2.763	1.563	4.02	1.788	2.032	4.328	

Table 9: The performance of MFCC-STD based unit selection speech synthesis[19]

Synthesized Speech											
Original Speech Signal		mar_0001	mar_0002	mar_0003	mar_0004	mar_0005	mar_0006	mar_0007	mar_0008	mar_0009	mar_0010
mar_0001	0.456	1.200	1.892	3.902	3.872	2.893	5.783	3.872	4.500	2.673	
mar_0002	1.231	0.632	2.762	2.090	1.988	2.763	1.235	1.6532	4.673	6.011	
mar_0003	5.632	3.982	1.050	1.928	2.782	2.782	1.892	1.292	3.020	5.873	
mar_0004	3.092	2.093	4.092	1.837	4.932	3.091	5.781	3.982	2.983	2.983	
mar_0005	1.882	1.0291	3.0281	3.091	2.872	1.092	4.092	3.982	3.982	5.021	
mar_0006	3.091	4.873	3.982	2.983	1.022	0.932	1.829	2.893	4.092	4.093	
mar_0007	2.983	3.092	2.993	4.984	2.831	8.011	1.920	1.892	3.001	3.092	
mar_0008	5.011	3.921	4.984	5.092	2.931	4.982	1.092	0.982	1.778	4.832	
mar_0009	2.938	5.011	2.932	6.091	2.983	1.921	2.932	4.632	1.092	1.920	
mar_0010	3.092	1.821	1.9082	3.921	4.921	3.842	4.983	4.530	2.321	0.210	

The sentences used for hidden Markov model based synthesis using MFCC based method is shown in table 10. The detail performance of MFCC-mean and standard deviation for hidden Markov model based speech synthesis are shown in table 11,12 respectively.

Table 10: Sentences and label used hidden Markov model speech synthesis

Sr. No	The Original Sentence	Label Used for Original Speech File	Label Used for Synthesized Speech File
1	प्राचीन भारतीय अर्थव्यवस्था अतिशय सुंदर होती	mar_01	mar_01
2	पुस्तक मागणी लेखी करणे आवश्यक	mar_02	mar_02
3	कारण मराठीत या नावाचे अशा पद्धतीने लेखन रूढ झाले आहे.	mar_03	mar_03
4	पुणे शहरातील एक मध्यवर्ती ठिकाण	mar_04	mar_04
5	मी योग्य तो बदल करीन	mar_05	mar_05

6	ऊर्जा ही सदैव स्थिर असते	mar_06	mar_06
7	औरंगाबाद शहर हे ह्या विभागाचे मुख्यालय आहे	mar_07	mar_07
8	हा मुख्यत्वे राजस्थान मध्यप्रदेश महाराष्ट्र व कर्नाटकात आढळतो	mar_08	mar_08
9	तसे करणे आवश्यक होते असे समर्थन पुढे केले गेले	mar_09	mar_09
10	हे कागदपत्र आहे साचा देश माहिती पोलंड विषयी	mar_010	mar_010

Table 11: The performance of MFCC-Mean based hidden Markov model speech synthesis[19]

		Synthesized Speech									
Original Speech Signal		mar_01	mar_02	mar_03	mar_04	mar_05	mar_06	mar_07	mar_08	mar_09	mar_010
	mar_01	0.234	3.781	5.155	6.280	2.662	5.442	3.601	5.432	3.970	11.950
	mar_02	5.227	0.200	8.700	6.191	1.7100	5.327	5.465	8.932	2.242	6.126
	mar_03	1.900	3.815	0.210	9.044	3.123	1.090	2.120	3.030	5.445	9.580
	mar_04	5.559	0.934	1.980	0.936	1.2315	1.780	2.090	2.050	6.318	12.272
	mar_05	2.980	3.800	3.178	2.153	0.119	2.130	2.150	3.092	2.339	23.011
	mar_06	2.051	9.1400	0.221	1.050	1.781	3.873	1.363	1.030	3.335	16.09
	mar_07	2.463	4.990	5.900	2.191	2.130	2.172	0.181	1.192	2.991	8.700
	mar_08	4.839	6.566	2.550	2.781	1.630	1.152	0.800	0.500	5.344	9.811
	mar_09	3.992	7.502	3.300	2.050	1.980	1.050	1.262	2.111	0.300	7.020
	mar_010	7.793	6.176	3.528	2.128	6.512	7.900	3.512	1.393	0.810	1.23

Table 12: The performance of MFCC-std based hidden Markov model speech synthesis[19]

		Synthesized Speech									
Original Speech Signal		mar_01	mar_02	mar_03	mar_04	mar_05	mar_06	mar_07	mar_08	mar_09	mar_010
	mar_01	0.110	1.221	2.119	1.290	1.890	1.233	2.030	3.01	1.178	5.020
	mar_02	2.120	1.20	1.900	3.402	7.680	8.900	11.02	2.678	2.343	7.890
	mar_03	0.900	1.123	2.342	3.784	3.134	4.030	2.178	9.01	2.05	5.030
	mar_04	2.564	3.100	2.870	0.900	1.760	2.345	2.403	3.050	2.870	3.435
	mar_05	1.890	1.450	2.123	0.200	5.40	2.656	1.934	2.999	7.030	9.01
	mar_06	0.890	1.543	2.212	3.210	3.521	0.321	2.986	1.776	2.832	3.02
	mar_07	2.320	1.564	2.220	5.040	3.442	5.021	0.210	3.450	3.887	7.00
	mar_08	1.747	2.030	5.020	2.456	3.022	0.884	3.007	4.302	2.345	1.284
	mar_09	2.336	3.020	3.040	3.121	3.998	7.060	4.990	3.2020	1.02	1.998
	mar_010	1.987	2.030	2.0440	2.312	3.220	1.228	6.070	3.009	4.006	0.320

7. CONCLUSION

The quality of speech synthesis is experimented using MOS score, MSE, PSNR, MFCC based techniques for hidden Markov model and unit selection approach. The MFCC based method is evaluated using the mean, standard deviation and variance. For all the estimated methods the unit selection method gives a better performance than hidden Markov model techniques as the database is small.

8. REFERENCES

- [1] Mohammed Waseem, C.N Sujatha, "Speech Synthesis System for Indian Accent using Festvox", International journal of Scientific Engineering and Technology Research, ISSN 2319-8885 Vol.03,Issue.34 November-2014, Pages:6903-6911
- [2] 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 3708 – 3711.

- [3] S. Martincic- Ipsic and I. Ipsic, "Croatian HMM Based Speech Synthesis," 28th Int. Conf. Information Technology Interfaces ITI 2006, pp.19-22, 2006, Cavtat, Croatia.
- [4] Alex Acero, "Formant Analysis and Synthesis using Hidden Markov Models". Proceedings of Eurospeech conference. September 1999
- [5] R.sproat, J. Hirschberg, and D. Yarowsky, "A corpus-based synthesizer", Proc. ICSLP, pp.563-566, 1992.
- [6] 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.

- [7] A. Black, P. Taylor, and R. Caley, "The Festival Speech Synthesis System System documentation Edition 1.4, for Festival Version 1.4.3 27th December 2002.
- [8] Series P: Telephone Transmission Quality "Methods for objective and subjective assessment of quality "- Methods for Subjective Determination of Transmission Quality ITU-T Recommendation P.800.
- [9] ITU-T P.830, Subjective performance assessment of telephone-band and wideband digital codecs
- [10] Lehmann, E. L.; Casella, George. "Theory of Point Estimation (2nd ed.). New York: Springer. ISBN 0-387-98502-6. MR 1639875
- [11] Huynh-Thu, Q.; Ghanbari, M. (2008). "Scope of validity of PSNR in image/video quality assessment". *Electronics Letters* 44 (13): 800. doi:10.1049/el:20080522
- [12] SR Quackenbush, TP Barnwell, MA Clements, Objective Measures of Speech Quality (Prentice-Hall, New York, NY, USA, 1988)
- [13] AW Rix, MP Hollier, AP Hekstra, JG Beerends, PESQ, the new ITU standard for objective measurement of perceived speech quality—part 1: time alignment. *Journal of the Audio Engineering Society* 50, 755–764 (2002)
- [14] JG Beerends, AP Hekstra, AW Rix, MP Hollier, PESQ, the new ITU standard for objective measurement of perceived speech quality—part II: perceptual model. *Journal of the Audio Engineering Society* 50, 765–778 (2002)
- [15] ITU-T P.862, Perceptual evaluation of speech quality: an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs 2001
- [16] 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
- [17] 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
- [18] Sangramsing N.kayte "Marathi Isolated-Word Automatic Speech Recognition System based on Vector Quantization (VQ) approach" 101th Indian Science Congress Jammu University 03th Feb to 07 Feb 2014.
- [19] 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
- [20] 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