IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)
Volume 5, Issue 6, Ver. I (Nov -Dec. 2015), PP 13-19
e-ISSN: 2319 – 4200, p-ISSN No. : 2319 – 4197
www.iosrjournals.org

Performance Calculation of Speech Synthesis Methods for Hindi language

Sangramsing Kayte[1], Monica Mundada[1], [2]Dr. Charansing Kayte
Department of Computer Science and Information Technology,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad
[2]Department of Digital and Cyber Forensic, Aurangabad Maharashtra, India

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: *TTS, MOS, HMM, Unit Selection, Mean, Variance, MSE, PSNR.*

I. Introduction

A speech synthesis system is computer-based systems 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 are 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 speechand 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.

II. Concatenate Synthesis

It is dictionary-based approach. Concatenative synthesis simply plays back the waveform with matching phone string [8]. The text to speech synthesizer developed at IIIT, Hyderabad used diphonemes as their fundamental unit, the end point of the splices are in the steady region of speech, so that transitions are not missed. The basic advantage of the concatenative method is its simplicity. The units are taken from original speech. Transitions like C and V are directly captured from the speech data and the rules to concatenate are elementary. However the use of a database restricts the type of speech that can be generated. It includes Unit Selection Synthesis &Diphone Synthesis. Unit selection module is responsible for selecting the best unit realization sequence from many possible unit realization sequences from the database. Diphone Synthesis is given particular words; tokenize character to generate particular speech. It contains the transition between two phones that has been chosen as the synthesis unit for concatenative synthesizers [9][10].

III. Hidden Markov Model Based Speech Synthesis

HMM synthesis provides a means to automatically train the specification-to-parameter module, thus by passing the problems associated with hand-written rules. HMM-based synthesis is a synthesis method based on hidden Markov models also called Statistical Parametric Synthesis [13][14]. In this system, the frequency spectrum (vocal tract), fundamental frequency (vocal source) and duration (prosody) of speech are modeled simultaneously by HMM. Speech waveforms are generated from HMMs themselves based on the maximum likelihood criterion [16].

IV. 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].

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.

V. Speech Database

A database is an organized collection of data. The general purpose of a database is to store data and a software system designed to allow the definition, creation, querying, update and administration of database. A database can be sorted and/or filtered. Many databases have application software that accesses the database on behalf of end-users. Database designers and database administrators interact with the DBMS through dedicated interfaces to build and maintain the application databases, thus need some more knowledge and understanding about how DBMSs operate. There are several types of data and use relationships to pull the data into the useful businesses or tracking expenses etc. A database schema is a way to logically group objects such as tables, views and stored procedures etc. Schema can be created and altered in a database and user can be granted access to a schema. Database design uses three domains: Phonetic dictionaries, continuous speech and text [51]. A phonetic dictionary is a dictionary that allows locating the word by the way it sounds. These dictionaries are useful when the spelling of a word is unknown. Phonetic dictionaries are special dictionaries that list words by how they sound instead of how they are spelled. It is particularly helpful to people just learning a language, because phonetic dictionaries represent words the way they're meant to be pronounced using the sounds and characters of any language. Phonetics is a branch of linguistic study that concerns itself with how words sound. Phonetics can help people better understand a language through hearing and better speak a language by providing help with pronunciation. 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.

VI. Experimental Analysis

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	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10
1	5	5	5	5	4	4	5	4	4	5
2	5	5	4	5	5	4	5	4	4	5
3	4	4	5	4	3	3	4	2	5	4
4	5	4	4	5	4	4	5	5	5	5
5	5	5	5	5	4	4	5	3	3	5
6	5	4	5	5	5	4	5	4	4	5
7	4	5	4	4	4	4	4	4	4	4
8	4	4	5	4	4	5	4	5	5	4
9	5	3	5	5	3	4	5	3	5	5
10	5	5	4	5	4	4	5	4	4	5

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

Subject / Sentence	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Sub10
1	4	5	5	5	5	4	5	4	5	3
2	3	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	3
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	1	4	2	5	4	2	3
9	4	3	3	3	5	3	4	5	2	4
10	5	4	4	1	2	2	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	hin_0001	hin_0001	7.94	3.30
2	hin_0002	hin_0002	4.57	6.72
3	hin_0003	hin_0003	1.02	3.21
4	hin_0004	hin_0004	3.70	4.20
5	hin_0005	hin_0005	7.61	2.57
6	hin_0006	hin_0006	5.32	1.26
7	hin_0007	hin_0007	8.06	7.56
8	hin_0008	hin_0008	7.20	1.29
9	hin_0009	hin_0009	9.25	3.24
10	hin_0010	hin_0010	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	hin_0001	hin_0001	9.15	7.315
2	hin_0002	hin_0002	8.38	6.24
3	hin_0003	hin_0003	13.5	5.25
4	hin_0004	hin_0004	10.4	8.40
5	hin_0005	hin_0005	9.26	8.76
6	hin_0006	hin_0006	9.38	9.42
7	hin_0007	hin_0007	10.10	9.05
8	hin_0008	hin_0008	9.63	6.56
9	hin_0009	hin_0009	10.42	8.49
10	hin_0010	hin_0010	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	85	85	85	90.73	93.31
2	Unit Selection	95	90	85	95.83	96.26

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

MFCCs plus log energy and pitch values [50]. The cepstrum is the inverse Fourier transforms of the log-spectrum. The Cepstrum computed after a non-linear frequency wrapping onto a perceptual frequency scale, is called Mel-frequency scale. Since it is an inverse Fourier transform, the resulting coefficients are called Mel frequency Cepstrum coefficients (MFCC) 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	आपकेहिन्दीपसन्दकरनेपरखुशीहुई	hin_0001	hin_0001
2	शरीरमेंपित्तअग्निकाप्रतिनिधिहै	hin_0002	hin_0002
3	तरीमीतंवसकळदेहीं	hin_0003	hin_0003
4	होइहिकाजुमोहिहरषबिसेषी	hin_0004	hin_0004
5	ऑस्ट्रेलियाकेखेलकेलिएखोजें	hin_0005	hin_0005
6	ऊर्जाकेस्रोतकेलिएखोजें	hin_0006	hin_0006
7	कमलभारतवर्षकाराष्ट्रीयपुष्पहॅ	hin_0007	hin_0007
8	गुरुसिखदेइरायपहिंगयउ	hin_0008	hin_0008
9	गगनकाअर्थहैआकाश	hin_0009	hin_0009
10	इनकाआकर्षणहैखास	hin_0010	hin_0010

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		hin_0001	hin_0002	hin_0003	hin_0004	hin_0005	hin_0006	hin_0007	hin_0008	hin_0009	hin_0010
	hin_0001	0.120	3.242	2.31	1.392	3.42	3.008	2.983	5.094	7.01	1.234
	hin_0002	1.281	0.009	2.111	2.453	7.632	1.90	4.02	1.223	8.01	3.04
	hin_0003	1.453	3.21	0.080	3.25	1.99	2.843	3.921	2.963	2.093	6.70
	hin_0004	3.02	1.230	2.564	0.899	2.786	5.453	1.672	1.981	2.67	3.45
	hin_0005	2.40	1.450	5.432	2.932	0.021	3.921	6.05	4.675	7.00	3.674
	hin_0006	3.896	8.09	3.983	2.732	3.674	0.673	2.843	5.03	3.894	4.92
	hin_0007	1.893	1.563	2.03	2.100	4.92	3.67	1.460	1.273	3.521	7.38
	hin_0008	2.932	3.674	2.732	3.721	3.567	2.732	3.876	0.783	2.673	3.643
	hin_0009	1.776	2.732	4.332	5.893	2.783	3.874	2.743	4.87	1.091	3.021
	hin_0010	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

Synthesized Speech											
Original Speech Signal		hin _0001	hin _0002	hin _0003	hin _0004	hin _0005	hin _0006	hin _0007	hin _0008	hin _0009	hin _0010
	hin_0001	0.456	1.200	1.892	3.902	3.872	2.893	5.783	3.872	4.500	2.673
	hin_0002	1.231	0.632	2.762	2.090	1.988	2.763	1.235	1.6532	4.673	6.011
	hin_0003	5.632	3.982	1.050	1.928	2.782	2.782	1.892	1.292	3.020	5.873
	hin_0004	3.092	2.093	4.092	1.837	4.932	3.091	5.781	3.982	2.983	2.983
	hin_0005	1.882	1.0291	3.0281	3.091	2.872	1.092	4.092	3.982	3.982	5.021
	hin_0006	3.091	4.873	3.982	2.983	1.022	0.932	1.829	2.893	4.092	4.093
	hin_0007	2.983	3.092	2.993	4.984	2.831	8.011	1.920	1.892	3.001	3.092
	hin_0008	5.011	3.921	4.984	5.092	2.931	4.982	1.092	0.982	1.778	4.832
	hin_0009	2.938	5.011	2.932	6.091	2.983	1.921	2.932	4.632	1.092	1.920
	hin_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 11,12respectively.

Table 10: Sentences and label used hidden Markov modelspeech synthesis

Sr. No	The Original Sentence	Label Used for Original Speech File	Label Used for Synthesized Speech File
1	श्रीकृष्णपाण्डवसभीगयेभीष्मकेपास	hin_0001	hin_0001
2	हिमालयपरबसायहदेशचीन	hin_0002	hin_0002
3	प्रथमसामान्यएवंद्वितीयविशेष	hin_0003	hin_0003
4	कृतिकाअर्थहोताहैनिर्माण	hin_0004	hin_0004
5	सिक्खधर्मपंजाबकामुख्यधर्महै	hin_0005	hin_0005
6	गढ़वालकासाहित्यतथासंस्कृतिबहुतसमृद्धहैं	hin_0006	hin_0006
7	काफीसंख्यामेंयहाँप्रतिष्ठितलोगमौजूदथे	hin_0007	hin_0007
8	दोएकहीलिंगकेलोग	hin_0008	hin_0008
9	हिंदीकेविद्वानसुप्रसिद्धलेखककवकविथे	hin_0009	hin_0009
10	फिरहिंतेकाहेनहोहिंदुखारी।	hin_0010	hin_0010

Table 11: The performance of MFCC-Mean based hidden Markov modelspeech synthesis

Synthesized Speech											
Original Speech Signal		hin _0001	hin _0002	hin _0003	hin _0004	hin _0005	hin _0006	hin _0007	hin _0008	hin _0009	hin _0010
	hin_0001	0.234	3.781	5.155	6.280	2.662	5.442	3.601	5.432	3.970	11.950
	hin_0002	5.227	0.200	8.700	6.191	1.7100	5.327	5.465	8.932	2.242	6.126
	hin_0003	1.900	3.815	0.210	9.044	3.123	1.090	2.120	3.030	5.445	9.580
	hin_0004	5.559	0.934	1.980	0.936	1.2315	1.780	2.090	2.050	6.318	12.272
	hin_0005	2.980	3.800	3.178	2.153	0.119	2.130	2.150	3.092	2.339	23.011
	hin_0006	2.051	9.1400	0.221	1.050	1.781	3.873	1.363	1.030	3.335	16.09
	hin_0007	2.463	4.990	5.900	2.191	2.130	2.172	0.181	1.192	2.991	8.700
	hin_0008	4.839	6.566	2.550	2.781	1.630	1.152	0.800	0.500	5.344	9.811
	hin_0009	3.992	7.502	3.300	2.050	1.980	1.050	1.262	2.111	0.300	7.020
	hin_0010	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

Synthesized Speech											
Original Speech Signal		hin _0001	hin _0002	hin _0003	hin _0004	hin _0005	hin _0006	hin _0007	hin _0008	hin _0009	hin _0010
	hin_0001	0.110	1.221	2.119	1.290	1.890	1.233	2.030	3.01	1.178	5.020
	hin_0002	2.120	1.20	1.900	3.402	7.680	8.900	11.02	2.678	2.343	7.890
	hin_0003	0.900	1.123	2.342	3.784	3.134	4.030	2.178	9.01	2.05	5.030
	hin_0004	2.564	3.100	2.870	0.900	1.760	2.345	2.403	3.050	2.870	3.435
	hin_0005	1.890	1.450	2.123	0.200	5.40	2.656	1.934	2.999	7.030	9.01
	hin_0006	0.890	1.543	2.212	3.210	3.521	0.321	2.986	1.776	2.832	3.02

	hin_0007	2.320	1.564	2.220	5.040	3.442	5.021	0.210	3.450	3.887	7.00
	hin_0008	1.747	2.030	5.020	2.456	3.022	0.884	3.007	4.302	2.345	1.284
	hin_0009	2.336	3.020	3.040	3.121	3.998	7.060	4.990	3.2020	1.02	1.998
	hin_0010	1.987	2.030	2.0440	2.312	3.220	1.228	6.070	3.009	4.006	0.320

VII. 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.

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, Kavitawaghmare ,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]. 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 "-ETHODs FOr SUBJECTIV DETERMINATIO 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]. 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) nternational 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