Text-To-Speech Synthesis of Marathi Numerals

G. D. Ramteke, R. J. Ramteke

Abstract— **The paper proposes Text-To-Speech (TTS) synthesis system for Marathi numerals. Marathi TTS-Synthesis system is processing of Marathi text written in Devnagari script and converting into a spoken form. This speech synthesis system is based on rule-based approach. The speech corpus of Marathi numerals is recorded through mike by different genders with various ages. The speech signals are normalized through the PRAAT tool. In this paper, any number can be synthesized such that a person would be able to understand the number by listening speech signals. The features of generated speech signals are extracted by pitch detection algorithm. The extracted feature assists to find out length of signals too. Mean and Standard Deviation were calculated for pitch detection of each voice samples. The results for generated sound signals or waveform are found to be satisfactory.**

Index Terms—**Text-To-Speech Synthesis, Rule-based, Pitch Detection, Prosody.**

I. Introduction

Text or speech plays pivotal role in communication amongst the people [1]. There are two ways of the communication, first one is imaginary application like E-mail, Text chatting and another one is physical equipment such as the mobile phone and many other devices. The present paper is concentrated on Text-to-Speech synthesis system which is an application domain of speech processing. In last decade, a great deal of TTS-Synthesis system has done much work in various languages as well as different synthesis techniques such as Unit-selection, Formant, Hidden Markov Model and Domain-Specific by researchers [2, 3]. Recently, progress in Text-To-Speech (TTS) has produced synthesizers with very high intelligibility. Text-To-Speech synthesis system is playing a vital role of information technology for 6-standard Indian languages [4-6]. Marathi is one of 22-constitutional languages of India and local language for state of Maharashtra in India. While writing, Marathi language is written by Devnagari script and its writing style is from left to right. It includes vowels, consonants and numbers. Mainly, present paper focuses on numerals Text-To-Speech synthesis system for Marathi language. TTS-system requires prosody for designing and implementation approach of TTS-system which is to generate waveform signals and convert Marathi text written into spoken form. There are two main aspects of speech synthesizer; first one is Text normalization of Marathi Number which is process of transforming Marathi text into a single numeric form and another one is speech generation [7]. The output range of few phones or speech samples is stored in the system [8-11]. The storage of entire numeral permits to use for high-quality speech output. All speech signals are normalized through PRAAT tools. The generated speech signals in speech synthesis domain are extracted the features by pitch detection algorithm with the help of Digital Signal Processing (DSP). DSP is a subfield of speech synthesis. Mostly, DSP is used for extracting the features and generating the waveform. Autocorrelation pitch detection is such a popular technique in the area of speech processing [13, 14, 24]. This method provides a way for estimation of pitch. The pitch is the part of prosodic variation in TTS system [8, 15]. Autocorrelation pitch detection method is applied on every speech signals and compared with a standard existing tool. After getting the pitch reading, Mean and Standard Deviation were calculated. The results for intelligible of speech synthesis program allows to visually challenged persons or students for listening as well as reading on a home computer or educational organization or private institute without any assistant [17-19, 25].

The initial objective of TTS is to convert Marathi numerals text into natural spoken waveform. These generated noise-free signals further proceeds for analyzing of prosody and compared with existing PRAAT tools.

This paper is summarized as follows: The coming soon section gives the brief information about history of text-to-speech model. The section-3 describes the Marathi numerals in Devnagari Script. The section-4 explains two subsections: first subsection is implementation and process of Text-to-Speech synthesis system; another one is description of analysis of speech signals. The section-5 discusses on the data acquisition of Marathi spoken numerals. The section-6 deals with experimental work. The section-7 gives the explanation of result and discussion. Ultimately, paper concludes in the section-8.

II. History of Text-to-Speech Model

Text-To-Speech system is an artificial speech generation system like something out of science fiction. A lot of researchers have been discovering the novelty in the research domain of speech synthesis since last few centuries. In 1769, articulatory speech synthesis which has been used for the bagpipe components to produce crude noises similar to a mankind voice. The talking mechanical machine originated to a revolution by Wolfgang von Kempelen. The mechanical version of the human vocal system on 1770 has been built using modified organ pipes by Danish scientist that the machine can speak the isolated vowels {'A', 'E', 'I', 'O', 'U'}. In 1837, the musical sound and instruments was rediscovered and popularized an improved version of the von Kempelen speaking machine by English physicist and prolific. Homer W. Dudley on 1928 developed an electronic speech synthesizer system which is operated through a

G. D. Ramteke, Research Fellow, School of Computer Sciences, North Maharashtra University, Jalgaon (MH), India. (e-mail: shyam.ramteke@gmail.com).

R. J. Ramteke, Professor & Head, Department of Information Technology, School of Computer Sciences, North Maharashtra University, Jalgaon (MH), India. (e-mail: rakeshj.ramteke@gmail.com).

keyboard. In 1940, Frank cooper of Haskins laboratories implemented a system called pattern playback that can be generated the voice signals. Walter Lawrence who was an American scientist in 1953 made Parametric Artificial Talker (PAT); the initial formant synthesizer produced speech sounds. George Rosen of MIT scientist on 1958 discovered a new thing of development such as pioneering articulatory synthesizer. Cecil Coker at Bell laboratories in 1960 developed the method of articulatory synthesis. Joseph P. Olive on 1970 executed concatenative synthesis. In 1978, a handheld electronic toy has been launched such as Texas instruments which was used crude formant speech synthesis for example a teaching aid. In 1984, Radiohead's Fitter Happier and Paranoid Android introduced the built-in MacInTalk speech synthesizer of Apple Macintosh widely used in popular songs. A concatenative speech synthesizer for natural-sounding has been revolutionized based on a huge corpus of recorded sound samples which have been collected from various persons. For online applications, the model of speech synthesizer can be utilized such as websites can read emails aloud. In the modern age, many reputed government and private organizations in India and out of India such as Google, TDIL, C-DAC and so on are working on a nice research domain of speech synthesis system in various languages which already exists in all over world. [3, 4, 11, 26-28]

III. Marathi Numerals in Devnagari Script

Marathi is the main language in Maharashtra, India. It is an officially language in most of the government and private sector among 22 regions all over India. Marathi numerals are written in Devnagari script, which is also used to write Hindi [24]. Marathi is similar to Hindi in structure and grammar. Marathi numbers are very essential because that structure is used in daily routine for their conversation.

Table 1 : Isolated Marathi Numerals in Devnagari Script

English Digit	Marathi Digit	Prononciation of Marathi Language
1	१	एक(Ek)
2	२	दोन (Don)
3	३	तीन (Teen)
4	४	चार (Chaar)
5	५	पाच (Paach)
6	६	सहा (Sahaa)
7	७	सात (Saat)
8	८	आठ (Aath)
9	९	नऊ (Nau)
10	१०	दहा (Dahaa)

There are two types of Marathi numbers as cardinal and ordinal. 1st-10 Marathi numerals have collected some samples in written form as shown in Table 1. English Digits and Marathi Digits are shown in above table for 1st-10 Numerals and How to pronounce in Marathi Language. The synthesized speech signal is sufficiently intelligible. Numerals can produce varying pronunciation depending on the way of various input text data [12].

IV. Implementation and TTS Process

A TTS process is to determine the pronunciation of a word based on its spelling which is often called Grapheme-to-phoneme (G2P). Grapheme is written form in specific language. Phoneme is sound form in particular language [8]. G2P conversion by rule is perhaps the classic application of tradition knowledge based rules in TTS [20].

A. TTS-engine

TTS-engine uses two basic approaches: rule-based and dictionary-based approaches which have been utilized for implementation. Rule-based approach was composed for grapheme-to-phoneme conversion using few rules. When the numbers are fetched from the corpus of text, the received text is converted into the pronunciation form which is available in voice corpus. Rule-based approach and dictionary-based approach were used for increasing the level of intelligibility and natural sound as possible respectively. Dictionary-based approach is proposed for storing the text and voice signals when it has a huge amount of text as well as speech signals. Both approaches are utilized for development of the TTS-engine which was supported for a very small amount of the corpus due to it is built up the level of intelligibility and natural voice.

These approaches are assisted for parsing of the numbers and symbols, which shows the following forms:

$$number ::= \text{int}\,eger \tag{1}$$

$$\text{int}\,eger ::= [sign]\{digit\} + [decmial - po\text{int}] \tag{2}$$

$$sign ::=, | / - \tag{3}$$

$$decimal - po\text{int} ::= . \tag{4}$$

$$digit ::= 0|1|2|3|4|5|6|7|8|9 \tag{5}$$

In general, the token is a potential number if it satisfies the syntax for numbers specified. A number marker is a letter. Whether a letter may be treated as a number marker depends on context. It contains at least on digit and should be begun with a digit. An optional part can end with a sign.

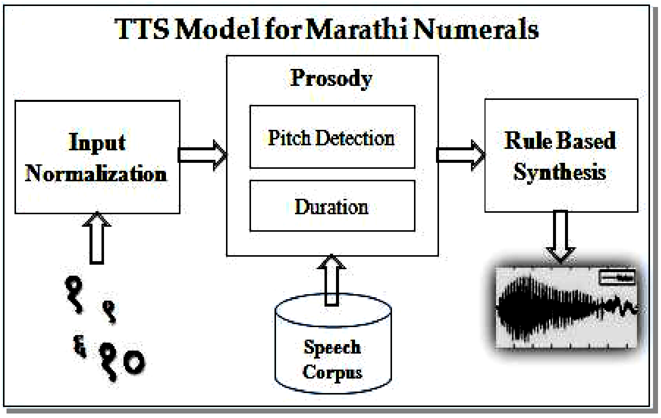

Fig. 1. TTS Model for Marathi Numerals

Speech Synthesis and pre-processing is implemented by using MATLAB. The MATLAB program uses modified Phonetic received text for waveform generation as per seen in

Fig. 1. Text Normalization is converting the Non-Standard Word (NSW) into standard pronounceable/readable numerals [11]. The non-natural language text is to convert into words and this process-id often called verbalization. Cardinal and Ordinal are different components in numerals system. Cardinal numbers which translates a number into a word representing the type of no. as 1 is "One" in English and **"एक"** in Marathi. Ordinal converts a number into the position of something in a list as 1 is "First" in English and **"पहिला"** in Marathi. The input for numerals processing will be a number. This is analyzed the number and find it. All speech signals are to be used from the database. All noise-free speech signals have been provided to system. Prosody includes duration of speech signals, intensity of signals and pitch related to frequency of vibration of the vocal cords [23]. But the present system designs and implements how to find out duration of speech signals and pitch detection of voice signals. Actually, Pitch is calculated fundamental frequency (F0) of speech signals [22]. How to detect the pitch signals in details next section.

B. Analysis of Speech Signals

i. Autocorrelation Pitch Detection Technique

Pitch detection is used to detect the fundamental frequency in speech signals [20]. The frequency of speech signals is measured in Hertz (cycles per second). The determination of pitch is difficult in implementing the reliable statistical model. This algorithm is used for speech synthesis [6, 22]. Because of many pitch detection algorithm fail to detect the correct pitch values in unvoiced or voiceless signals, Autocorrelation pitch detection technique is used in speech signal processing for analyzing the series of values, such time domain signals [22].

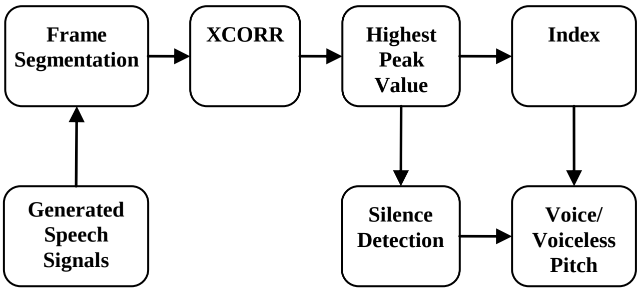

Fig. 2. Block Diagram of Autocorrelation Pitch Detection Technique

The MATLAB function xcorr is used to indicate the highest value/auto peak value of the articulated speech signal [23]. Fig. 2 shows block diagram of autocorrelation pitch detection technique. Actually, the generated speech signals as input speech is used the short-term analysis.

The autocorrelation function of a signal is as follows:

$$R[n] = \sum_{m=-\infty}^{\infty} y[m]y[n-m] \tag{6}$$

Where R[n] represents the radiation, y[m] shows the actual speech signal. This is to calculate the error from ∞ to -∞. Window, which has the effect of setting all values outside the range 0 <= n < N to 0. However, for P samples after the window, the error will not be 0 as it will still be influenced by the last few samples at the end of the region.

$$\varphi(j,k) = \sum_{-\infty}^{\infty} y[n-k]y[n-k] \tag{7}$$

$$= \sum_{n=0}^{N-1+P} y[n-j]y[n-k] \tag{8}$$

$$= \sum_{n=0}^{N-1-(j-k)} y[n]y[n+j-k] \tag{9}$$

This is a function of one independent variable j-k rather than the two of above equation.

$$\varphi(j,k) = \sum_{n=0}^{N-1-k} y[n]y[n+k] \tag{10}$$

$$R(j-k) = \varphi(j,k) \tag{11}$$

This is strongly related to periodicity in the waveform in the form of time domain. In prosodic analysis, F0 is seen as the direct expression of intonations and often intonation is defined as the linguistic use of F0 [20]. Speakers use a variety of pitch ranges to be more expressive. Ranges of pitch vary for a number of reasons. When one "raises one's voice" in anger one is using an increasing pitch range [20-24]. It factors also have a role to play in longer utterances.

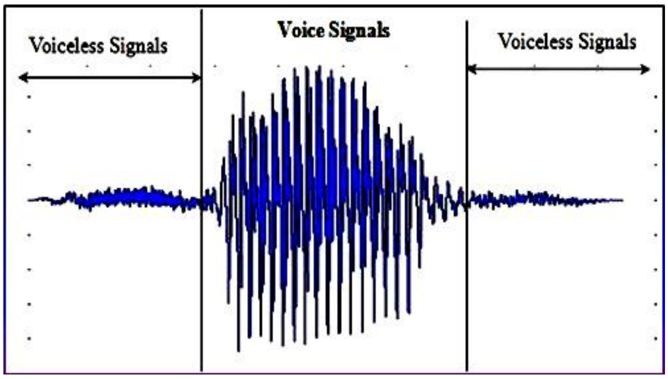

Fig. 3. Time Waveform as Marathi Sample Numeral of "७" (Seven) without Noise Signals

The time waveform represents a pattern the variation of information over time as per depicted Fig. 3. Signals can be classified periodic and those signals which repeat themselves over time are aperiodic. Identification of periodic is exactly no speech sound in original speech [13].

In normal, human voice range is about 500Hz to 2 KHz. Hearing range is about 16Hz–18 KHz. It is assumed that cepstrum maximum fundamental frequency among the value corresponds to the frequency band [50Hz-500Hz]. The speech signals of Marathi spoken numbers for male usual range of frequency values (F0) is [100Hz-150Hz] and female as well as children frequency range value is [250Hz-350Hz] respectively [20-23].

ii. Length of Speech Signals

This following equation is to find out the length of speech signals to calculate the duration of voice signals with noise or noise-free.

$$T = \frac{1}{F} \tag{12}$$

Where, T is total duration of speech signals in seconds, 1 is sampled data, F is sampling frequency of speech signals in Hertz.

iii. Statistical Methods

Statistical methods are required calculating means and standard deviations for data analysis as below.

$$\mu = \frac{\sum X}{N} \tag{13}$$

Where, μ is mean of signal,
X is individual sampled data,
N is number of sampled data of speech signals.

$$\sigma = \frac{\sum (X-\mu)^2}{n-1} \tag{14}$$

Where, σ = standard deviation of signal,
X = Individual sampled data of signal,
μ = Mean of all sampled data,
n = Number of samples.

V. Data Acquisition

The data acquisition of Marathi numerals have written in Devnagari script. The technical characteristics of the recorded data for Marathi Numerals speech signals are as follows: Microsoft wave format (.wav), Sampling Frequency (20KHz), One channel (mono) [5, 11, 20]. The Marathi spoken speech samples have been collected by unprofessional speakers with various genders and different ages. These Marathi numerals are spoken by 17-Females and 14-Males from School of Computer Sciences, North Maharashtra University, Jalgaon. These speeches are acquired through standard PRAAT tool but noise free speech signals by using spectral subtraction technique. Because Speech synthesis system needs quality and clear pronunciation speech signals. The recorded Marathi numerals speech database size is 310. These Marathi Numerals speech signal frequency are analyzed the prosodic by pitch detection algorithm and also, to find out length of speech signals. Autocorrelation pitch detection is popular technique for estimation of pitch [20].

VI. Experimental Work

In this experiment, The Marathi number speech synthesizer is able to synthesize any number. The synthesized speech signal is sufficiently intelligible. Numbers can produce varying pronunciation depending on the way of input text information. The Marathi synthesized numerals is generated speech signals as well as waveform. The generated speech signals as Marathi synthesize numbers are extracted the features by using autocorrelation pitch detection method and length of speeches. This method is used for pitch detection which is found out the pitch tracking reading and rather than closed to reading of Praat tool but various undefined signal is closest to zero hertz of pitch reading. Total speech samples are 310 and used for the implementation.

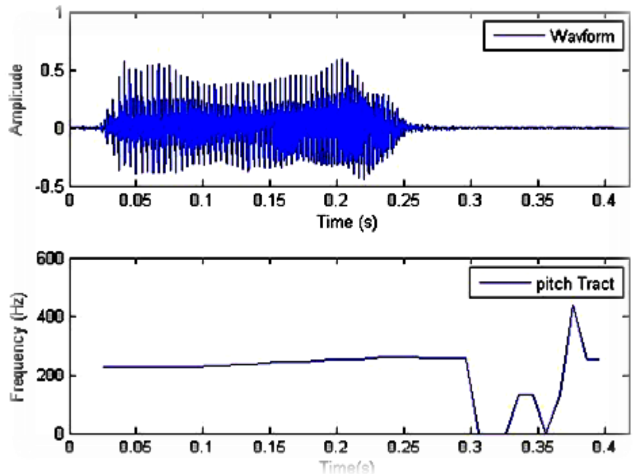

Fig. 4. Pitch Detection Waveform without Noise Signals of Marathi Sample Number "१" (One)

The output generated waveform is delivered to pitch frequency estimation from noise free speech signals which are accepted through autocorrelation pitch detection technique and also, to find out pitch tracking as shown in Fig. 4. Pitch estimation of generated speech signals synthesized Marathi number through autocorrelation pitch tracking method is applied to all generated utterances.

VII. Result and Discussion

The evaluation of the result is done for the intelligibility and natural as possible. The evaluated results have been focused on the intelligibility which is decided on speaking rate. The speaking rate of various speakers is achieved to compare with a standard tool. The Marathi numbers are discovered in the several results which are discussed on the following tables. The speech signals were collected from various speakers with different ages and estimated by clean speech signal waveform. The synthetic speech signals are reflecting to calculate the pitch detection. Estimated pitch frequency from generated speech signals such as १ ते १० in Marathi (1 to 10 in English) numbers are showing mean and standard deviation as shown in Table 2 and Table 3. Table 4 : Mean Pitch Detection Reading of Praat Tools as PT and Autocorrelation Pitch Technique as AP can be shown in detail. "१" Marathi number standard PRAAT tool as PT reading of Male-1 is 157.58Hz and "१" as Marathi number autocorrelation reading as AP of Male-1 is 121.58Hz as per Table 4. It is easy way to find out undefined speech signal in autocorrelation method. Another one example of "४" Marathi number standard PRAAT tool reading of Female-3 is 109.34Hz and "४" as Marathi number autocorrelation reading of Female-3 is 120.78Hz. Table 5: Pitch Tracking Reading in Praat Tools as PT and Autocorrelation Pitch Technique as AP in Standard Deviation. One is "१" in Marathi numeral that

PRAAT tool as PT reading of Male-1 is 10.13Hz and same Marathi numeral is 13.15Hz in autocorrelation pitch tracking technique as AP. Another one is "४" as Marathi number that PRAAT tool reading of Female-3 is 6.45 Hz and similar Marathi numeral of Female-3 is 8.45 Hz in autocorrelation pitch tracking technique.

Table 2 : Dissimilar Rate of Mean Pitch Tracking Reading with Duration of Speech Signals

Marathi Numbers	Duration of Speech in Seconds	Mean of Pitch Reading		Dissimilar Rate
		Praat Tool	Auto. Tech.	
१	0.37	198.11	191.94	6.16
२	0.24	179.52	177.83	1.69
३	0.57	177.23	176.09	1.13
४	0.44	176.86	171.35	5.51
५	0.40	172.01	171.12	0.89
६	0.50	179.94	175.79	4.14
७	0.47	187.10	182.34	4.76
८	0.41	165.40	16027	5.12
९	0.50	172.84	169.35	3.48
१०	0.44	183.14	178.43	4.71

Table 3 : Difference Rate of Pitch Detection Reading in Standard Deviation with Length of Generated Speech

Marathi Numbers	Length of Speech in Seconds	Pitch Tracking Reading in SD		Difference Rate
		Praat Tool	Auto. Tech.	
१	0.37	9.38	9.28	0.10
२	0.24	12.13	11.52	0.61
३	0.57	13.26	13.04	0.22
४	0.44	9.58	9.55	0.03
५	0.40	11.10	11.39	-0.29
६	0.50	11.63	11.83	-0.20
७	0.47	14.18	14.10	0.08
८	0.41	10.45	10.61	-0.16
९	0.50	10.56	10.45	0.11
१०	0.44	12.09	11.56	0.53
Mean		**11.43**	**11.33**	**0.10**

VIII. Conclusion

This paper presents techniques for Marathi numerals and data driven preparation. A typical TTS system has been developed to verify the performance of converting input number into synthesized speech. A text-to-speech (TTS) system has been demonstrated small applications. The generated speech signals are applied autocorrelation pitch detection algorithm for feature extraction purpose. Estimation of fundamental frequency (F0) which has been calculated statistical manner through Mean and Standard Deviation. The statistical results of pitch detection which have been appropriate matched with existing tools and to find out duration of speech signals as output of TTS system by providing Marathi numerals. Especially, intelligible text-to-speech programs allow playing an important role for visually disabled people and it is betterment for society.

Future Scope

In future work, an attempt will be made to estimate the Text-To-Speech of Devnagari vowels spoken by speakers with noise-free environments and will be applied various statistical techniques.

Acknowledgment

The authors would like to thank the Rajiv Gandhi Science and Technology Commission, NMU Centre, Govt. of Maharashtra, India for funding the project (Code No. 7-II-DP/2014) and a G. H. Raisoni Doctoral fellowship, North Maharashtra University, Jalgoan (MH-India).

References

[1] Mustafa Zeki, Othman O. Khalifa, A. W. Naji, "Development of An Arabic Text-To-Speech System", *International Conference on Computer and Communication Engineering (ICCCE)*, 2010, pp. 1-5.

[2] Aimilios Chalamandaris, Sotiris Karabetsos, "A Unit Selection Text-to-Speech Synthesis System Optimized for Use with Screen Readers", *IEEE Transactions on Consumer Electronics*, 56 (3), 2010, pp.1890-1897.

[3] Priyanka Jose, Govindaru V., "Malayalam Text-to-Speech", *International Journal of Engineering and Technical Research (IJETR)*, Vol. 1,Issue-3, 2013, pp. 13-16.

[4] S. D. Shirbahadurkar, D. S. Bormane, "Speech Synthesizer Using Concatenative Synthesis Strategy for Marathi language (Spoken in Maharashtra, India)", *International Journal of Recent Trends in Engineering*, 2 (4), 2009, pp. 80-82.

[5] Pamela Chaudhur, K Vinod Kumar, "Vowel classification based approach for Telugu Text-to-Speech System using symbol concatenation", *International Conference [ACCTA-2010], IJCCT*, 2010, pp. 183-187.

[6] Parminder Singh, Gurpreet Singh Lehal, "Text-To-Speech Synthesis System for Punjabi Language", *Internal Conference Information Systems for Indian Languages, Patiala, India*, 139, 2011, pp. 302-303.

[7] Mumhammad Masud Rashid, Md. Akter Hussain, M. Shahidur Rahman, "Text Normalization and Diphone Preparation for Bangla Speech Synthesis", Journal of Multimedia, 5 (6), 2010, pp. 51-55.

[8] Langla and Narunat Wongsawat, "A Dictionary-Based Approach for Thai Text to Speech (TTTS)", *Third International Conference on Measuring Technology and Mechatronics Automation*, 2011, pp. 40–43.

[9] D. J. Ravi and Sudarshan Patilkulkarni, "A Novel Approach to Develop Speech Database for Kannada Text-to-Speech System", *Internal Journal on Recent Trends in Engineering & Technology*, 5 (1), 2011, pp. 119-122.

[10] Krzysztof Popowski, Edward Szpilewski, "Optimization of allophone database compression with wavelets for polish speech synthesis TTS systems", *SPECOM'2006*, St. Petersburg, 2006, pp. 439-444.

[11] Sunil S. Nimbhore, Ghanshyam D. Ramteke, Rakesh J. Ramteke, "Implementation of English-Text to Marathi-Speech (ETMS) Synthesizer", *IOSR Journal of Computer Engineering (IOSR-JCE)*, 17 (1), 2015, pp. 34-43.

[12] Firoj Alam, S.M. Murtoza Habib and Mumit Khan, "Bangla Text to Speech using Festival", *Conference on Human Language Technology for Development (HLTD 2011), Alexandria, Egypt*, 2011.

[13] S. Saraswathi and R. Vishalakshy, "Design of Multilingual Speech Synthesis System", *Intelligent Information Management*, 2 (1), 2010, pp. 58-64, *Published Online (http://www.scirp.org/journal/iim).*

[14] S. Martinic- Ipsic and I. Ipsic, "Croation HMM-based Speech Synthesis", *28th IEEE International Conference on Information Technology Interface*, 4, 2006, pp. 251-256.

[15] Masanori Morise, CheapTrick, "A Spectral Envelope Estimator for High-Quality Speech Synthesis", *Speech Communication*, 67, 2015, pp. 1-7.

[16] Marc Schroder, Jurgen Trouvain, "The German Text-to-Speech synthesis System MARY: A tool for Research Development and Teaching, *International Journal of Speech Technology*, 6, pp. 365-377.

[17] Thierry Dutoit, Milos Cernak, "TTSBOX: A MATLAB Toolbox for teaching Text-to-Speech Synthesis" *IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (ICASSP '05)*, 5, 2005, pp. 537-540.

[18] Theophile K. Dagba, Charbel Boco, "A Text to Speech System for Fon Language using Multisyn Algorithm", *18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems – KES2014*, 35, 2014, pp. 447-455.

[19] Giedrius Norkevicius, Gailius Raskinis, Asta Kazlauskrience, "Knowledge-based Grapheme-to-Phoneme Conversion of Lithuanian Words", *SPECOM 2005, 10th International Conference SPEECH and COMPUTER, [Patras, Greece]*, 2005, pp. 235–238.

[20] Fu-Chiang Chou, Chiu-Yu Tseng, and Lin-Shan Lee, "A Set of Corpus-Based Text-to-Speech Synthesis Technologies for Mandarin Chinese", *IEEE Transactions on Speech and Audio Processing*, 10 (7), 2002, pp. 481 - 494.

[21] Yoshifunmi Chisaki, Hidetoshi Nakashima, Shuuichi Shiroshita, Tusyoshi Usagawa and Masanao Ebata, "A Pitch Detection method based on continuous wavelet transform harmonic signal Subtraction Based Noise Reduction Method", *AST Japan Science and Technology Information Aggregator, Electronic*, 24 (1), 2003, pp. 7-16.

[22] Jong-soon Jung, Jeong-jin Kim and Myung-jin Bae, "Pitch Alternation technique in speech synthesis system", *IEEE Transaction on acoustics, speech and signal processing*, 2001.

[23] S. S. Nimbhore, G. D. Ramteke and R. J. Ramteke, "Pitch Estimation of Devnagari Vowels using Cepstral and Autocorrelation Techniques for Original Speech Signal", *International Journal of Computer Applications*, 55 (17), 2012, pp. 38-43.

[24] Mukta Gahlawat, Amita Malik, Poonam Bansal, "Natural Speech Synthesizer for Blind Persons using Hybrid Apprach", *5th Annual International Conference on Biologically Inspired Cognitive Architectures*, 41, 2014, pp. 83-88.

[25] Guangcai Yin, Jianzhong Luo, Xu Wang, Bin Yuan, "Utilizing Text-to-Speech to Assist Bilingual Teaching", *Second International Workshop on Education Technology and Computer Science*, 2010, pp. 594-597.

[26] "Wolfgang von Kempelen's speaking machine and its successors", Available at : http://www2.ling.su.se/staff/hartmut/kemplne.htm

[27] "History and Development of Speech synthesis", Available online at: http://research.spa.aalto.fi/publications/theses/lemmetty_mst/chap2.html

[28] "Speech synthesizers: Who invented speech synthesis?", Available at: http://www.explainthatstuff.com/how-speech-synthesis-works.html

G. D. Ramteke received the MCA degree in Computer Science from North Maharashtra University, Jalgaon in 2010. He has worked on several positions such as software support engg., software trainee and project fellow in IT industry as well as government organization during 2010-2014. He has been pursing Ph. D. in Computer Science at School of Computer Sciences, North Maharashtra University, Jalgaon (MS), India since May, 2014. He is interested in Speech Enhancement, Speech Recognition, Speech Cryptography and Text-To-Speech Synthesis.

R. J. Ramteke received the degree of Ph. D., M. Tech., M. Sc. (Comp. Sci.) from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India. He awarded UGC teacher research fellowship during 2004-2006. He got the best student paper award at IEEE International Conference on Cybernetics and Intelligent System (CIS-2006), Bangkok, Thailand. His experience of the professional and research is above 17 years and 9 years respectively. He is Professor & Head of Department of Information Technology at School of Computer Sciences in North Maharashtra University, Jalgaon. He has published more than 67 research papers in journals and conferences. His areas of interest are Image Processing, Pattern Recognition, Document Image Analysis etc.

Table 4 : Mean Pitch Detection Reading in Hertz of Praat Tool as PT and Autocorrelation Pitch Technique as AP

Marathi Numbers / Speakers	Mean	१	२	३	४	५	६	७	८	९	१०
Male 1	PT	157.58	116.35	115.87	109.90	109.96	112.52	140.60	106.09	120.27	128.15
	AP	121.58	117.43	116.88	109.47	112.64	117.20	137.18	139.37	117.84	126.78
Male 2	PT	122.07	145.52	149.34	150.83	148.86	145.72	107.27	103.07	153.13	106.68
	AP	149.80	142.03	148.70	146.72	148.32	140.81	113.51	105.18	147.62	105.44
Female 3	PT	296.83	142.45	151.09	109.34	108.58	103.32	125.74	149.35	114.23	284.48
	AP	216.36	140.14	147.32	120.78	107.32	106.57	126.22	146.65	113.78	271.37
Male 4	PT	138.71	166.99	178.01	152.87	144.04	273.32	221.20	121.38	148.54	147.00
	AP	135.97	158.53	169.16	149.15	141.43	261.67	205.10	122.60	145.11	144.52
Male 5	PT	164.69	269.45	124.23	152.56	167.22	145.00	272.73	221.15	159.64	157.63
	AP	152.02	257.16	121.56	151.49	156.53	140.02	261.71	209.66	158.92	160.01
Male 6	PT	126.15	121.72	250.16	123.66	123.77	235.16	113.98	233.59	126.04	120.00
	AP	123.09	119.33	239.82	120.78	122.08	226.64	111.68	226.74	124.38	117.96
Female 7	PT	266.72	138.34	283.61	227.64	222.90	134.86	265.55	260.47	222.42	231.37
	AP	258.70	133.04	267.46	213.39	209.51	132.89	273.50	245.29	243.64	221.53
Male 8	PT	118.42	240.82	118.72	240.77	261.51	245.15	262.35	119.66	261.78	220.04
	AP	144.29	227.78	122.58	233.17	280.68	250.06	258.88	122.62	251.01	220.41
Female 9	PT	150.71	367.02	146.35	111.17	108.80	228.18	134.25	134.48	116.83	107.39
	AP	146.14	350.56	139.32	109.83	111.69	219.97	131.25	131.65	115.88	105.98
Male 10	PT	139.76	114.22	265.97	140.77	136.71	116.42	111.47	212.60	136.13	241.17
	AP	134.65	114.29	255.72	140.27	132.50	124.40	112.47	229.47	134.30	229.80
Male 11	PT	275.60	107.59	113.65	250.78	241.42	128.82	124.12	104.81	386.18	245.39
	AP	265.58	109.85	111.95	240.58	232.48	131.24	124.96	109.64	363.68	239.44
Female 12	PT	255.81	140.38	111.27	359.00	362.68	145.76	152.41	112.21	106.91	117.47
	AP	277.34	139.04	110.02	338.23	336.90	145.62	150.58	112.97	108.36	115.54
Male 13	PT	208.31	152.28	150.37	113.08	109.77	306.46	280.59	129.03	113.34	218.53
	AP	206.00	154.22	143.71	121.65	139.21	298.61	263.69	126.82	112.64	210.07
Male 14	PT	169.09	181.36	152.77	149.24	102.76	233.50	238.83	163.57	133.95	371.20
	AP	142.49	178.14	148.71	109.98	119.31	221.86	234.85	162.91	130.12	348.84
Male 15	PT	161.77	171.18	292.94	277.01	129.34	174.18	185.09	280.08	157.03	102.74
	AP	168.70	173.81	283.09	215.65	124.61	168.89	177.77	177.87	153.24	110.45
Male 16	PT	135.90	269.63	166.93	231.48	150.36	141.08	141.88	138.27	291.02	109.44
	AP	142.49	258.79	162.31	224.72	151.59	142.34	142.26	138.29	278.83	112.93
Male 17	PT	339.67	115.67	134.55	140.87	276.73	150.94	149.89	130.52	179.07	122.31
	AP	325.35	132.42	141.95	179.23	268.00	144.35	140.97	126.44	176.53	119.54
Female 18	PT	266.90	239.53	318.53	154.50	237.71	137.62	273.92	268.27	152.22	151.61
	AP	255.50	241.29	305.22	167.43	223.31	137.39	273.50	244.10	152.90	147.83
Male 19	PT	262.99	117.37	117.21	129.61	181.83	260.49	116.48	117.12	146.46	278.49
	AP	254.89	116.90	123.67	130.25	181.61	250.68	118.98	115.99	143.43	265.70
Female 20	PT	128.28	249.12	262.71	278.69	135.69	109.20	251.84	248.10	284.20	220.03
	AP	134.23	235.62	250.90	257.11	139.21	109.44	231.38	247.68	266.62	209.39
Female 21	PT	274.42	203.09	117.59	110.36	151.74	251.08	259.01	119.67	120.28	164.97
	AP	275.67	234.17	187.95	118.47	154.60	221.14	238.74	123.85	117.72	163.54

Table 5 : Pitch Tracking Reading in Hertz Praat Tool as PT and Autocorrelation Pitch Technique as AP in Standard Deviation

Marathi Numbers / Speakers	Pitch Tracking	१	२	३	४	५	६	७	८	९	१०
Male 1	PT	10.13	5.57	8.26	6.35	4.55	7.77	7.70	8.01	10.82	8.27
	AP	13.15	3.83	9.34	5.48	9.15	15.34	19.01	10.15	9.85	8.4
Male 2	PT	2.29	8.51	6.90	11.99	7.80	9.91	4.40	5.33	10.85	3.87
	AP	3.21	7.70	12.42	12.95	16.45	12.18	7.58	14.16	8.82	4.2
Female 3	PT	8.81	6.39	4.44	6.45	4.43	4.58	5.43	2.66	6.07	22.28
	AP	7.36	6.49	6.86	8.45	4.13	12.19	9.64	2.24	8.31	20.23
Male 4	PT	5.20	19.26	17.40	4.71	3.26	14.71	11.71	7.82	6.25	8.92
	AP	4.76	25.22	20.77	3.67	3.5	12.43	7.4	13.76	18.11	8.44
Male 5	PT	14.28	16.20	5.63	11.75	5.13	4.65	16.59	9.63	15.04	18.39
	AP	19.53	13.18	4.92	9.69	19.92	4.45	18.23	8.84	16.74	17.09
Male 6	PT	5.51	8.04	18.78	7.00	8.73	13.48	8.47	3.54	4.43	6.79
	AP	5.20	5.22	20.34	4.79	9.73	10.69	15.8	2.75	5.49	5.72
Female 7	PT	7.24	8.80	17.93	12.83	17.75	6.02	21.90	8.22	14.21	15.45
	AP	10.86	11.78	14.7	20.57	20.23	6.28	19.28	7.77	4.6	17.89
Male 8	PT	2.91	13.48	5.33	14.83	2.90	7.68	19.56	7.09	6.11	10.34
	AP	1.89	12.29	14.11	12.41	3.4	8.51	18.09	6.71	2.64	11.23
Female 9	PT	5.20	16.80	14.19	2.22	3.25	13.35	4.22	4.27	2.40	5.06
	AP	3.49	15.15	8.83	2.27	11.56	5.61	10.47	4.84	2.19	5.58
Male 10	PT	11.23	10.53	14.25	9.95	6.68	6.96	5.82	9.42	6.60	18.00
	AP	8.17	19.40	12.92	12.55	11.03	7.51	7.89	18.06	5.93	15.32
Male 11	PT	8.31	4.60	15.50	9.17	3.36	15.35	6.77	6.69	16.53	15.94
	AP	6.34	6.81	5.91	7.09	7.6	16.3	8.45	10.88	16.66	10.01
Female 12	PT	16.36	4.84	7.09	9.94	17.57	11.65	8.35	6.12	5.83	9.81
	AP	22.68	7.75	10.51	9.08	17.28	9.72	7.61	3.36	7.48	8.54
Male 13	PT	9.66	11.99	7.02	5.49	4.84	19.11	10.27	7.73	8.34	17.90
	AP	7.37	14.05	18.74	9.35	10.66	12.66	10.68	8.06	17.28	15.27
Male 14	PT	11.31	9.48	16.08	17.58	6.50	7.57	14.92	12.44	6.99	11.49
	AP	7.66	7.74	12.65	21.34	6.60	13.01	18.79	4.66	4.49	15.02
Male 15	PT	14.50	15.18	12.87	5.77	7.45	18.66	5.92	9.03	9.53	19.13
	AP	19.5	23.30	14.04	9.21	4.8	20.03	23.48	9.91	18	11.62
Male 16	PT	2.94	5.94	12.67	9.45	10.14	12.46	12.46	16.37	11.93	12.51
	AP	2.49	3.65	12.63	6.35	8.3	8.42	4.43	4.24	19.51	14.42
Male 17	PT	16.56	7.16	8.74	14.96	14.87	13.64	12.26	11.69	11.00	13.43
	AP	10.79	7.76	23.71	12.95	11.09	12.07	15.86	14.73	11.94	11.24
Female 18	PT	4.78	18.04	18.07	16.15	16.04	11.02	8.32	17.91	9.51	6.36
	AP	4.73	15.25	19.42	9.95	16.65	15.71	20.23	12.21	3.37	11.75
Male 19	PT	18.86	7.78	9.29	6.91	6.10	10.12	6.74	8.91	13.62	14.61
	AP	20.30	5.81	8.57	7.00	10.32	9.81	18.07	16.26	12.71	7.87
Female 20	PT	9.14	19.04	19.04	9.56	12.65	7.77	4.44	13.01	16.27	15.76
	AP	10.12	18.29	10.66	11.23	10.66	6.21	22.84	13.75	8.44	13.98
Female 21	PT	11.93	19.47	7.11	3.25	14.38	10.35	9.58	7.91	12.34	12.87
	AP	5.89	18.47	12.05	4.52	19.9	19.35	12.49	2.94	9.91	11.62