SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348 – 8379 / https://doi.org/10.14445/23488379/IJEEE-V9I9P101

Volume 9 Issue 9, 1-11, September 2022
© 2022 Seventh Sense Research Group®

Original Article

Implementing Concatenative Text-To-Speech Synthesis System for Marathi Language using Python

Vinayak K. Bairagi[1], Sarang L. Joshi[2], Vastav Bharambe[3]

[1]*Department of Electronics & Telecommunication, AISSMS IOIT, Pune, India.*
[2]*AISSMS IOIT, Pune, India and School of Mechatronics, Symbiosis Skills and Professional University, Pune, India.*
[3]*School of Mechatronics, Symbiosis Skills and Professional University, Pune, India.*

[2]*Corresponding Author : jsarang70@gmail.com*

Received: 12 July 2022 Revised: 01 September 2022 Accepted: 12 September 2022 Published: 30 September 2022

Abstract *- A Text To Speech (TTS) synthesiser is a computer-based system which converts arbitrary input text into speech. A TTS system is helpful not only for speech or visually impaired people but also for educationally backward and underprivileged. Many TTS systems exist for English but still many people worldwide are not literate and comfortable in speaking, writing and reading English. A local language interface needs to be developed for such people. Considering this need, we have attempted to develop a TTS system for the Marathi language using python. Marathi is the fourth largest spoken language in India and an official language of the Indian state of Maharashtra and Goa. Marathi is known to and spoken by over 100 million people not only from India but also from Mauritius and Israel. Developing a Marathi TTS system will be useful for people in Maharashtra and several migrants coming to the state in search of jobs, business or education.*

Keywords *- TTS, Speech Synthesis, Natural Language Processing, Text processing.*

1. Introduction

Speech is the oldest way of communication among human beings and one of the premier forms of everyday communication. Text-to-speech conversion has great potential in Human-computer interaction, education for rural communities, and interaction with visually impaired or speech impaired persons. Speech intelligibility and naturalness are the key factors for the user acceptability of synthesised speech [1]. For languages like English; which is acceptable worldwide, the user interfaces for IT services have grown rapidly, but, in a country like India, where local languages are used for communication, the interfaces in local language need to be developed to access IT applications, information and services [2].

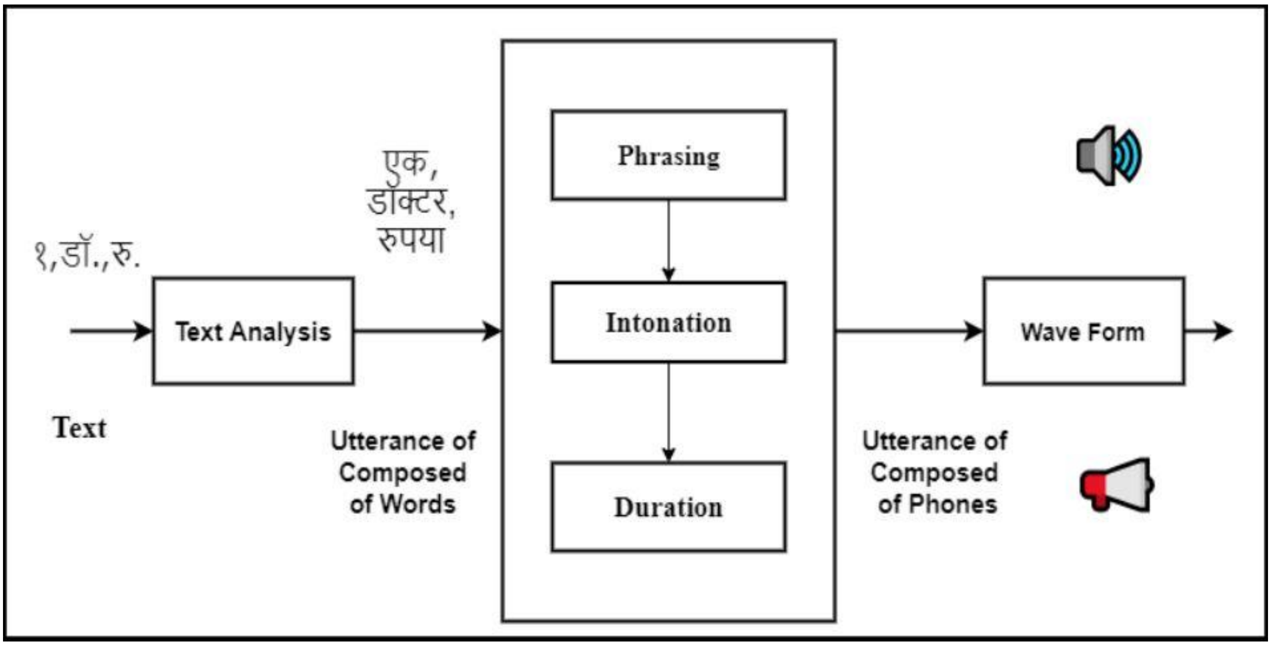


Fig. 1 Generalised block diagram of Text to Speech Synthesis

A generalised block diagram of text-to-speech synthesis techniques is shown in Fig. 1. It has input in the form of written text, a combination of alphabets, numbers, or special characters coded electronically (ISCII/ASCII/UNICODE). The text analysis involves segmentation, i.e. the long train of characters broken into small blocks easier to handle, e.g. converting sentences to words and syllables. Here, accurate segmentation is a vital task. It is difficult to handle the sentences as in some cases, where a dot in the statement may represent the full stop or the short form end or decimal point to represent fractional numbers [1][3][4]. Normalisation is required to generate meaningful pronounceable words, which may differ in different contexts. E.g. 05/06/1996 can be treated with mathematic language where '/' represent the division and can also be treated with the date.

1.1. Prosody

One of the important aspects of the speech signal is its Prosody. Prosody maintains expressiveness and intelligibility in speech. Prosody depends on the meaning of sentences. Short breaks or breath pauses in the wrong places in the sentences will change the statement's meaning [3][4]. Prosody is also dependent on language. The Prosody features like stress, duration, and pitch depend on age, gender, the speaker's attitude and physical and emotional state[5]. Written text hardly contains any information about these features. A stressed syllable can be identified by a rise or fall in fundamental frequency[1]. The speaker's physical and emotional state, attitude, and gender will affect the pitch contour. According to the meaning of the sentence, the pitch contour will be modified, e.g. the pitch slightly rises toward the end of the sentence for question-marked sentences, whereas and lower amplitude to the end of the sentence when it is in a normal form [6].

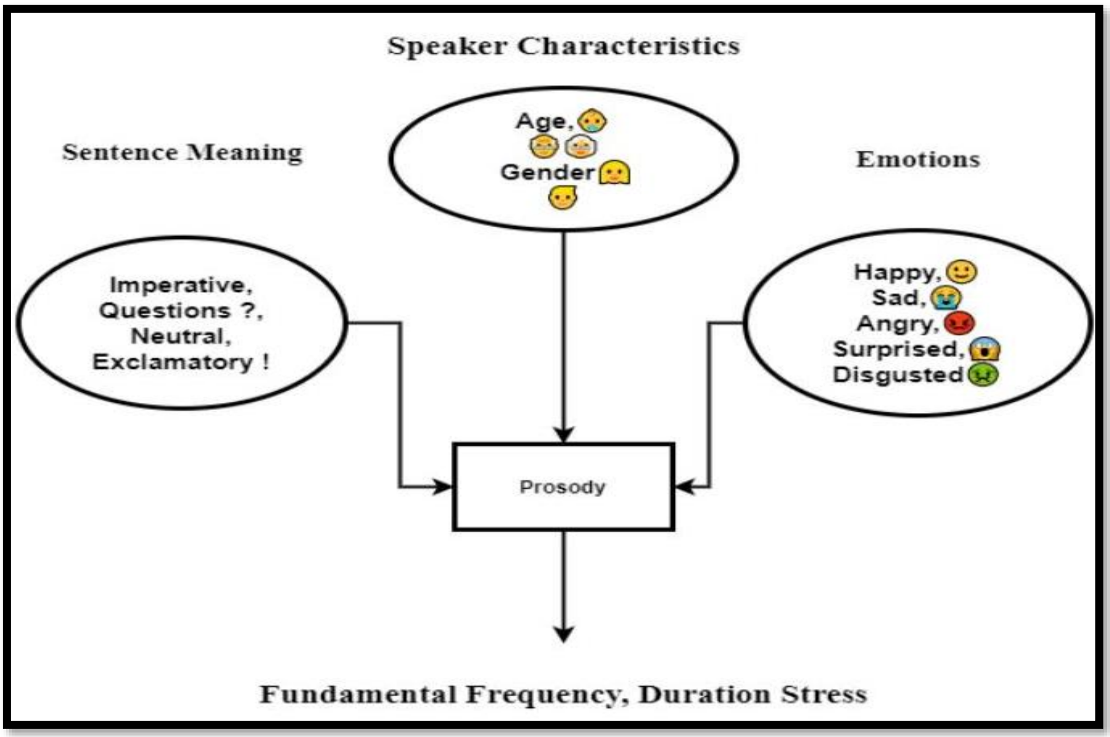


Fig. 2 Prosody dependencies

To achieve high-quality speech in a TTS system, one needs to derive as much relevant information from the input text as possible. (e.g. stress, duration, correct intonation)[7].

2. Database

Speech corpora (database collection) is the most important in speech processing. Collecting the speech database is essential to study the acoustic and linguistic properties of the speech. The corpus should be rich phonetically and prosodically. One can get the text sentences from news bulletins, interviews, everyday conversations etc. The selected text should be read and recorded by a native speaker. The duration of the recording is from several minutes to hours.

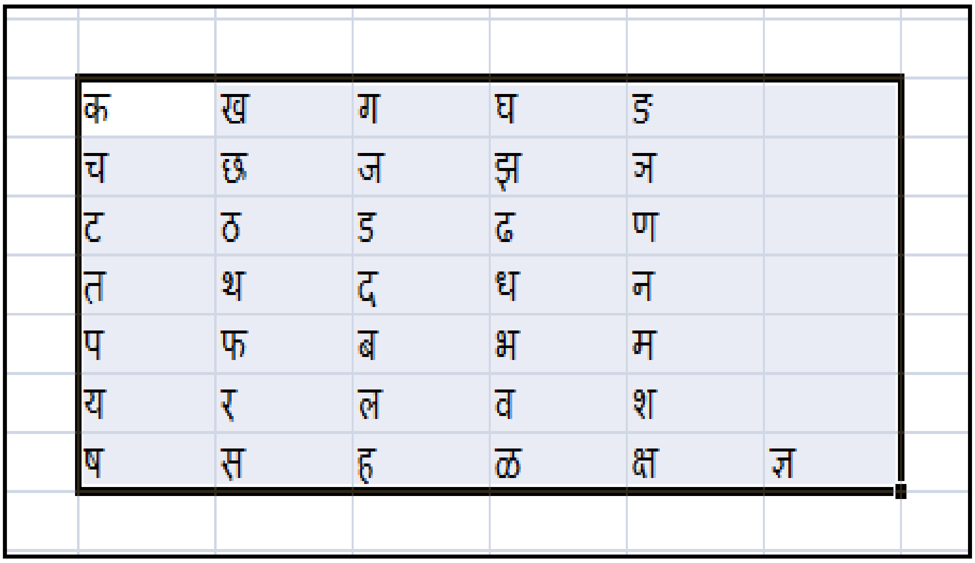
Following are the few desirable characteristics of the database:

- Simple, short, easy-to-read sentences.
- Grammatically correct sentences.
- Sentences from diverse sources.
- Meaningful and natural sentences.

The quality of the synthesised output depends heavily on the quality of the recorded speech. Hence, studio recording by professional male/female speakers is necessary with proper care and measures to ensure the same speech quality during the multiple recording sessions. The system's performance depends on the database to which it refers. Creating the database is limited because of the high cost of the recording process.

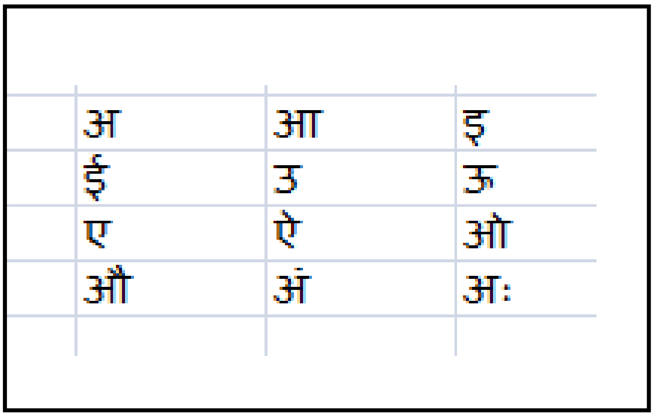
2.1. Text database

Marathi script contains a total of 48 alphabets which includes 36 consonants and 12 vowels[8]. Out of the 36 consonants, based on their pronunciation, the first 25 consonants are divided into 5 groups of 5 letters each. The Marathi consonants and vowels are shown in Fig. 3 and Fig. 4, respectively.



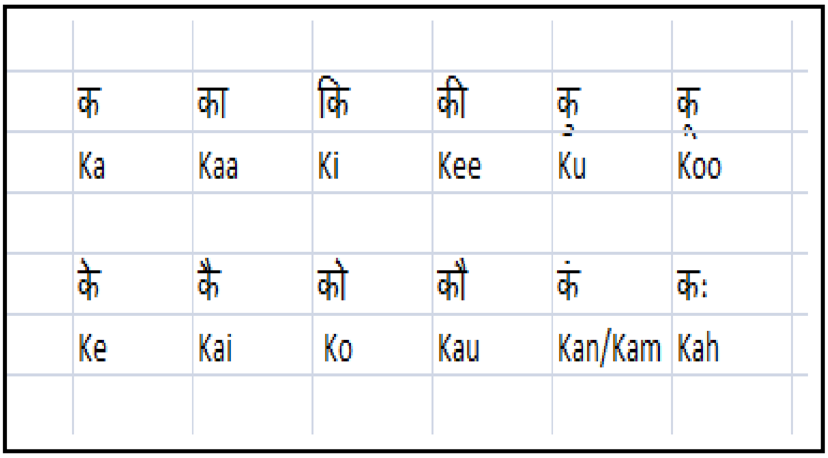
क	ख	ग	घ	ङ	
च	छ	ज	झ	ञ	
ट	ठ	ड	ढ	ण	
त	थ	द	ध	न	
प	फ	ब	भ	म	
य	र	ल	व	श	
ष	स	ह	ळ	क्ष	ज्ञ

Fig. 3 Marathi Consonants



अ	आ	इ
ई	उ	ऊ
ए	ऐ	ओ
औ	अं	अः

Fig. 4 Marathi Vowels



क	का	कि	की	कु	कू
Ka	Kaa	Ki	Kee	Ku	Koo
के	कै	को	कौ	कं	कः
Ke	Kai	Ko	Kau	Kan/Kam	Kah

Fig. 5 Marathi Vowel-Consonant combination

The database of Marathi words for numerical or countable is also created.

We have collected Marathi text and speech corpus from C-DAC consisting of syllables, most frequently used words, sentences and prosody-rich sentences from various fields/domains like Agriculture, Geography, History, Literature, Religion, Science & Technology, Tourism and Economy. The database consists of: Conjunct words: 588; frequently used words: 969; Marathi Barakhadi: 423; Most frequent sentences: 522; Dynamo Article: 272, Prosody rich sentences: 516; Vocabulary sentences: 101; Other Vocabulary: 236. The text database also contains all possibilities of Marathi vowels and consonant combinations.

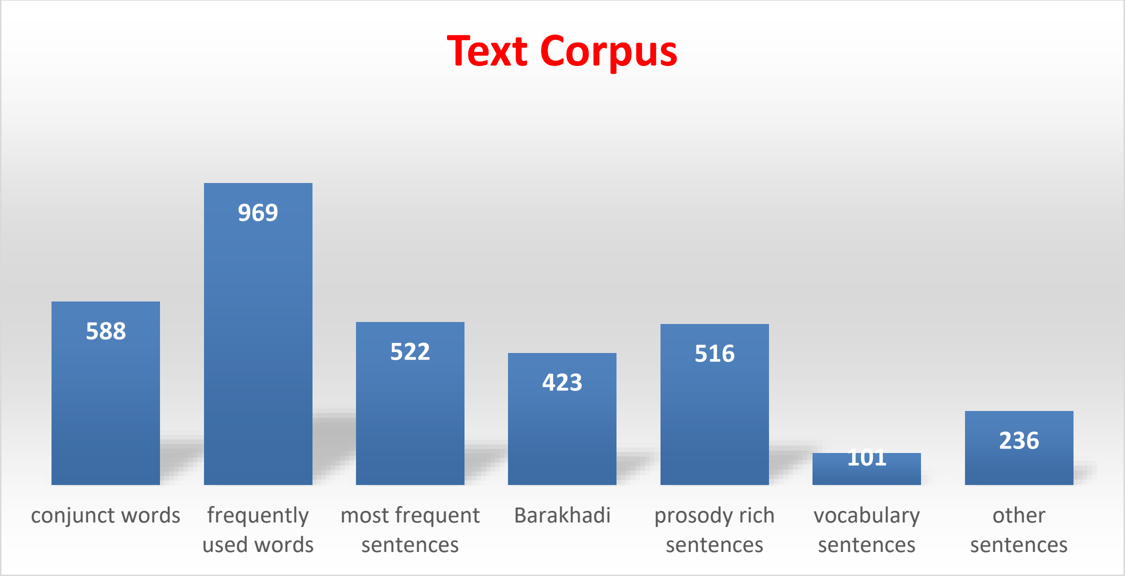


Fig. 6 Marathi text corpus

Table 1 represents some of the Prosody-rich Marathi statements from the database.

Table 1. Prosody-rich Marathi statements

Sr No	Statement	Prosody type/ Punctuation
1	तुझे नाव काय आहे?	Question mark, space
2	तुम्ही ब्रह्मदेव आहात आणि ना महात्मा.	Jodakshar
3	कबूतर ताशी 100 मैल वेगानं उडू शकत नाही.	Numbers + statement
4	उषाकडून जोरात पळलं जात नाही.	Anusvara
5	तो माझं बोलणं ऐकून दुःखी झाला नाही.	Visarg (:)
6	परतताना इकडे येणार ना !	Exclamation
7	न्यूयॉर्कला आल्यावर नाही, लंडन सोडण्यापूर्वीच रोकलं.	Comma (,) (pause), Anusvara

2.2. Setup used for Recording

The text database is recorded in the voice of native Marathi male and female speakers in the professional (noise-free) studio. The microphone used for recording was Behringer C-IU, USB studio condenser, USB studio condenser, and cardioid pickup pattern.

2.2.1. Technical Details

1. Software used for recording: Cool Edit Pro, Wave Surfer 1.8.8, Audacity1.3.6.
2. Resolution: 16-bit PCM, Mono.
3. Sampling Frequency used: 48 KHz
4. Text format for recording: UTF-8

3. Open Source Text to Speech Synthesis Engines

Table 2 represents the comparison of the text-to-speech synthesis engines.

Table 2. Open source speech synthesis engines[11-13]

Sr No	Engine	Programming Language	Technique	Comment
1	MARY TTS	JAVA	USS & HMM	Marathi not supported
2	eSpeak	C	Formant Synthesis	Less Natural speech output
3	GNUSPEECH	(LINUX OS)	Articulatory Synthesis	Rule-based approach
4	FESTIVAL	C++	Concatenative Synthesis	Slow
5	FLITE	C	Concatenative Synthesis	Lighter, Faster, For Embedded system Not R&D platform
6	Free TTS	JAVA	-	Based on "Flite"

The 'Festival' system was (and still is) primarily developed under Unix (Linux, FreeBSD and Solaris). However, full support for Windows is not yet as mature or stable as the versions under Unix. 'Flite' is unsuitable and intended for researchers in research and development platforms in the case of Marathi speech synthesis.

4. Methodology

A pronunciation in Marathi is mainly based on syllables. Neighboring sound elements are not influencing the naturalness of the syllables. Syllables can be the best unit for Marathi TTS (Text to speech) synthesis as it is acoustically and perceptually more stable units than phones [3][5][14][15].

The Marathi language is the input to the system in the form of text. The database of commonly used words and all combinations of consonants and vowels is created. Corresponding Marathi pronunciation of this entire database is created in the recording studio through professional Marathi native speakers.

The input Marathi text is first normalised. It will first check for any abbreviation, acronyms or symbol in the sentences. So there are processes of normalisation of the text data, e.g. 'Rs' (abbreviation), '₹' (symbol). Both are to be transformed into words such as 'rupees'. The detailed working methodology in the form of the flowchart is represented in Fig. 7.

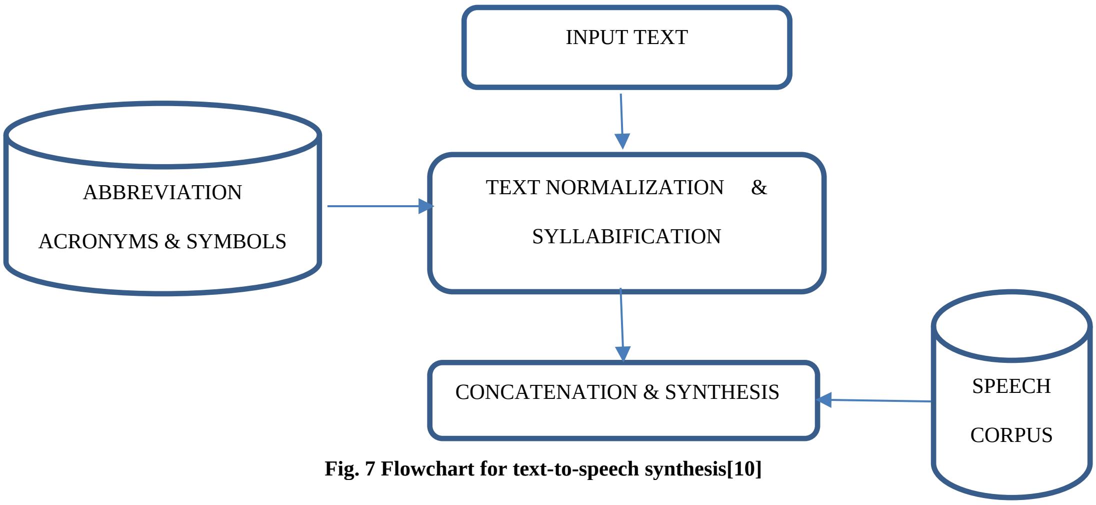


Fig. 7 Flowchart for text-to-speech synthesis[10]

The audio database consists of recorded audio files of consonants, vowels, frequently used words and statements in ".wav" format. The text database consists of the text files corresponding to audio files in the audio database. Syllables are the unit of spoken words. It is a phonological unit consisting of one or more sounds, including a vowel sound[9][10]. For the Marathi language, the syllables combine vowels and consonants to form meaningful sounds and their pronunciation. It contains only one vowel at a time.

The speech corpus stores the Marathi speech database, pre-recorded by professional native Marathi speakers in a studio environment. The databases consist of both male and female voices. The algorithm will match the text with its respective syllable for its voice from the corpus. The final output consists of concatenating the voice from the corpus for respective syllables. Few initial experiments are done with the above methodology for observation and further adaptation.

We have finalised the Natural Language Tool Kit (NLTK) python library for Marathi text processing[22]. The text is encoded in Unicode format. We split the sentences into words and words into letters and identified the punctuations present in the sentences.

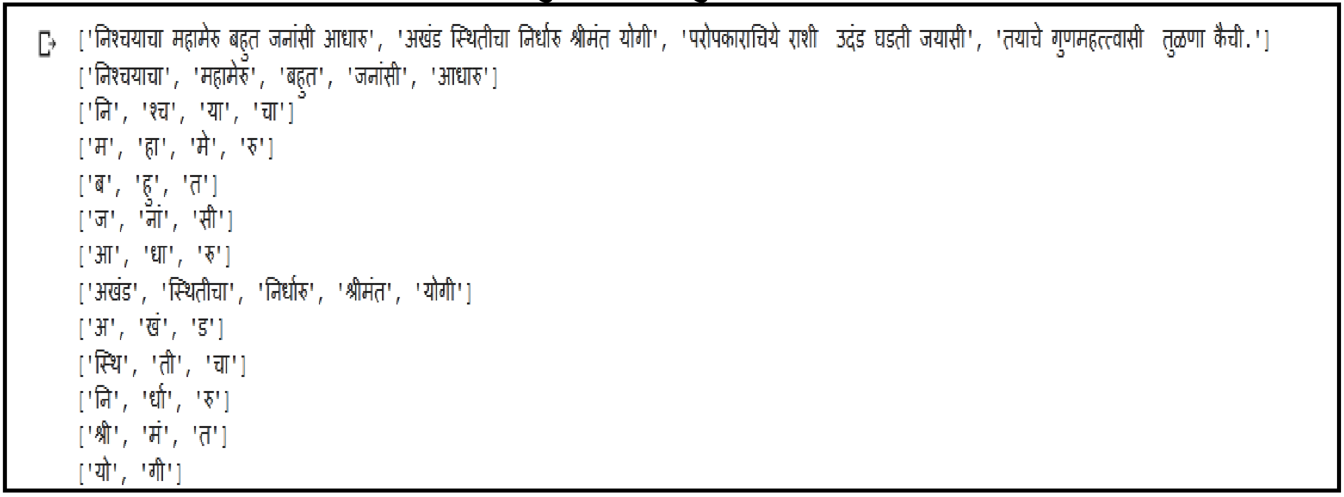
4.1. Working with Complex Statements

A complex statement is taken from Marathi's' abhang' – a kind of poem. The complex statement contains the joint word. This complex statement is given as the input to the system.

The input text paragraph is split into sentences by identifying punctuation marks, sentences are further split into words by identifying white space, and words are split into syllables (CV structures). Fig. 8 shows the result of python programming of splitting action done on the complex statement.

4.1.1. Marathi Input Text

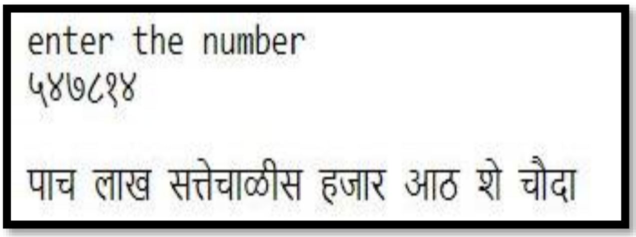
"निश्चयाचा महामेरु बहुत जनांसी आधारु. अखंड स्थितीचा निर्धारु श्रीमंत योगी. परोपकाराचिये राशी उदंड घडती जयासी. त याचे गुणमहत्त्वासी तुळणा कैची."



```
['निश्चयाचा महामेरु बहुत जनांसी आधारु', 'अखंड स्थितीचा निर्धारु श्रीमंत योगी', 'परोपकाराचिये राशी  उदंड घडती जयासी', 'तयाचे गुणमहत्त्वासी  तुळणा कैची.']
['निश्चयाचा', 'महामेरु', 'बहुत', 'जनांसी', 'आधारु']
['नि', 'श्च', 'या', 'चा']
['म', 'हा', 'मे', 'रु']
['ब', 'हु', 'त']
['ज', 'नां', 'सी']
['आ', 'धा', 'रु']
['अखंड', 'स्थितीचा', 'निर्धारु', 'श्रीमंत', 'योगी']
['अ', 'खं', 'ड']
['स्थि', 'ती', 'चा']
['नि', 'र्धा', 'रु']
['श्री', 'मं', 'त']
['यो', 'गी']
```

Fig. 8 Screenshot of breaking paragraph (multiple sentences) to words and words to syllables

4.2. Digit Processing



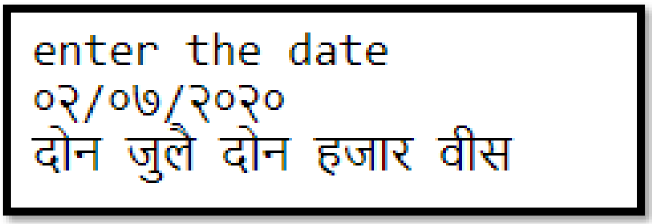
```
enter the number
५४७८१४

पाच लाख सत्तेचाळीस हजार आठ शे चौदा
```

Fig. 9 Text normalisation- Converting six-digit numbers to equivalent Marathi text

4.3. Date Processing

The corresponding English text of entered date below figure: 02/07/2020

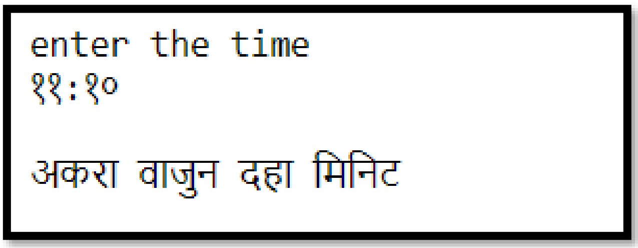


```
enter the date
०२/०७/२०२०
दोन जुलै दोन हजार वीस
```

Fig. 10 Text normalisation- Converting date to equivalent Marathi text

4.4. Time Processing

The corresponding English text of entered time in below figure: 11:10

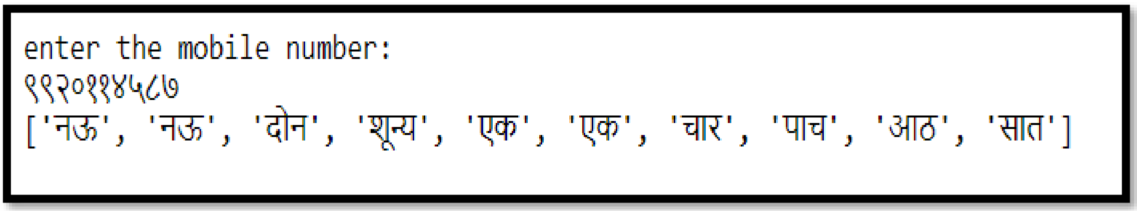


```
enter the time
११:१०

अकरा वाजुन दहा मिनिट
```

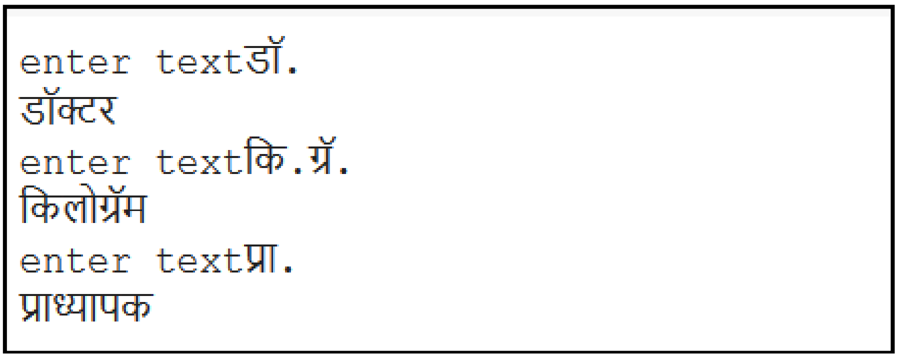
Fig. 11 Text normalisation- Converting time to equivalent Marathi text

The corresponding English text of entered mobile number below figure: 9920114587



```
enter the mobile number:
९९२०११४५८७
['नऊ', 'नऊ', 'दोन', 'शून्य', 'एक', 'एक', 'चार', 'पाच', 'आठ', 'सात']
```

Fig. 12 Text normalisation- Converting Mobile phone number to equivalent Marathi text

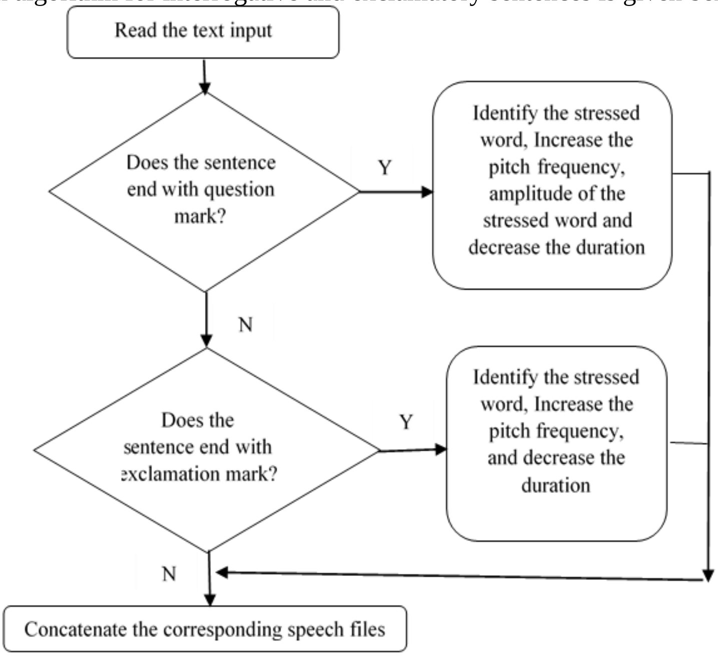


```
enter textडॉ.
डॉक्टर
enter textकि.ग्रॅ.
किलोग्रॅम
enter textप्रा.
प्राध्यापक
```

Fig. 13 Converting abbreviations into equivalent Marathi words

4.5. Prosody Conversions

The prosody conversion algorithm for interrogative and exclamatory sentences is given below.



Read the text input
Does the sentence end with question mark?
Y
Identify the stressed word, Increase the pitch frequency, amplitude of the stressed word and decrease the duration
N
Does the sentence end with exclamation mark?
Y
Identify the stressed word, Increase the pitch frequency, and decrease the duration
N
Concatenate the corresponding speech files

Fig. 14 Prosody conversion

5. Measuring the Quality of TTS

Quality measures are used to judge the output's quality and the speech's naturalness. Since the subject deals with the naturalness of the generated output speech, subjective quality measures are used.

We used the Mean Opinion Score (MOS) to analyse the system performance. The Mean Opinion Score is a subjective measurement used to test the listener's perception of speech quality[25]. It is calculated as the arithmetic mean of ratings given by human subjects (listeners).

$$MOS = \frac{\sum_{n=0}^{N} R_n}{N} \quad \text{.................................eq (1)}$$

R: individual ratings are given by N Subject (listeners)
MOS was calculated for Intelligibility, Speed and Naturalness.

5.1. Intelligibility

How easily the output can be understood (considering syllables, words etc.)

5.2. Naturalness

How much the output sounds like the speech of a real person

6. Experiments and results

Sentence 1: पाच वाजायला दहा मिनीट आहेत.

The input Marathi text is spitted into syllables as per the method discussed. The generated output, which is the concentration of syllables, is analysed in Praat software (version 6.1.16) [19]. The output of the Praat software tool for time domain analysis and frequency domain analysis is shown in Fig. 15. The upper part of the figure represents the time domain representation of the generated output. In contrast, the lower waveform in the following figure represents the frequency domain representation of the generated output. Both graphs combinely represent spectro-temporal representation of the sound signal called spectrogram of the signal. The horizontal axis represents time; it is common for both waveforms, i.e. start with the same point with the same scale. The y-axis for the time domain waveform (upper waveform) represents amplitude variation in the signal. The y-axis of the frequency domain waveform represents frequency starting from 0 Hz to 5000 Hz.

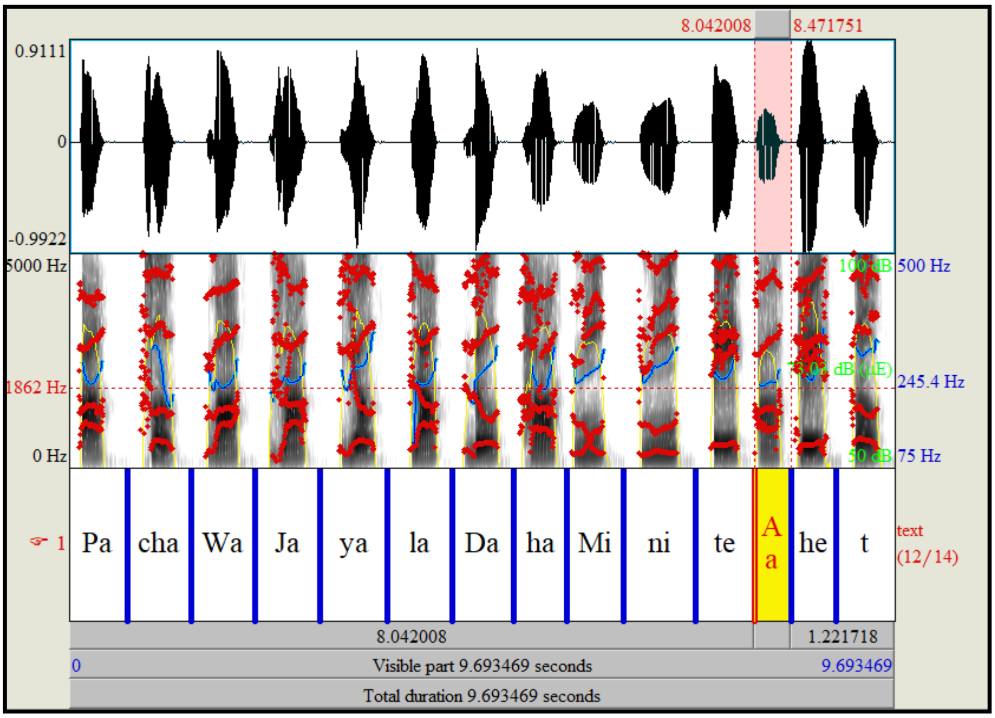


Fig. 15 Spectrogram of the synthesised waveform for sentences 1.

The text annotation is added at the lower side of the spectrogram for better correlation of words and waveform. Darker parts of the spectrogram represent higher energy densities, whereas the lighter parts of the waveform represent lower energy densities. The blue dots represent pitch contour in the signal, and the y-axis on the right hand represents its scale in blue colour. The red lines on the spectrogram represent the formant in the speech signal. The yellow line indicates the intensity contour in the signal. From the waveform in Fig. 15, it is noticed that we are getting the discrete output. Discrete in the sense that the word पाच has to be pronounced combined, whereas the system-generated output is 'पा___च'. It is an unnatural sound for the listeners. It is because; we are processing each letter separately. A database of commonly used words is created and stored to overcome this difficulty. The algorithm is modified to make the database of most frequently used sentences and common words in Marathi language communication.

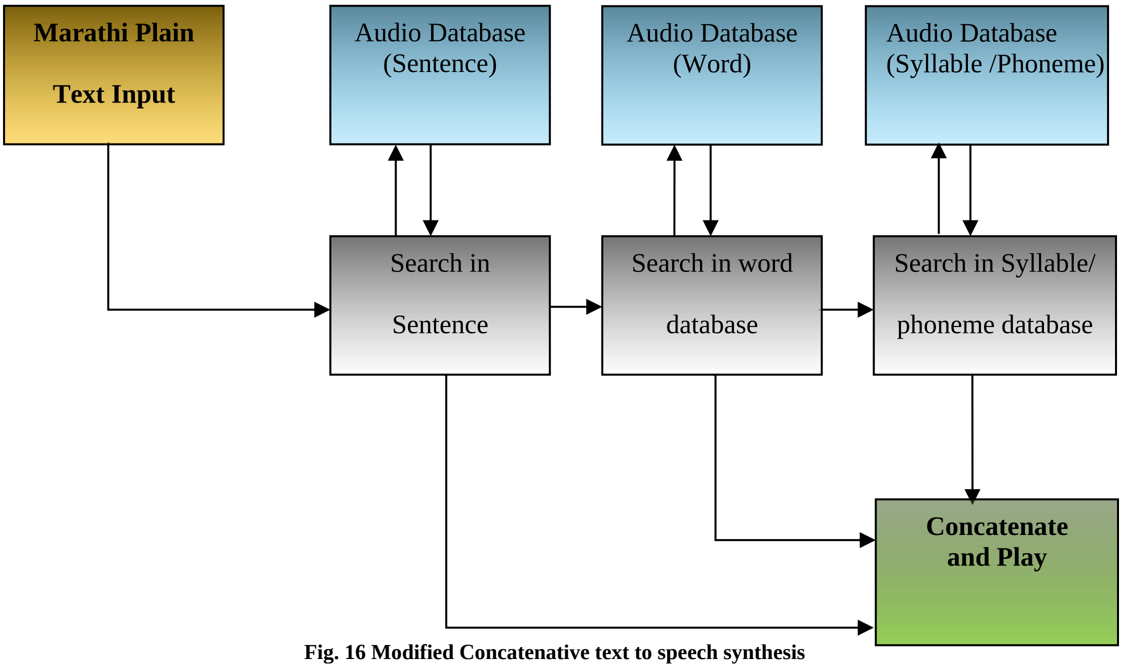


Fig. 16 Modified Concatenative text to speech synthesis

The same sentence is given as input to the revised algorithm. The spectrogram of the same is shown in Fig. 17. We can easily compare the two spectrograms shown in Fig. 15 and Fig. 17 for one sentence. The time gets reduced from 9.69 sec to 2.86 sec.

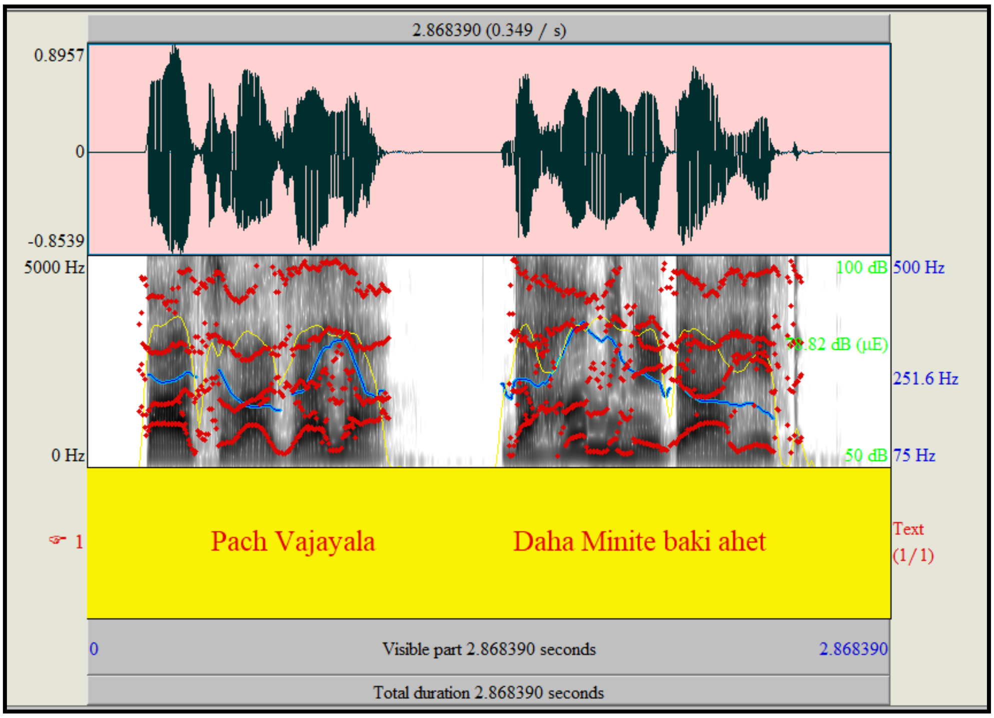


Fig. 17 Spectrogram of studio recorded waveform

Ten native Marathi male and female participants in the age group of 18 to 60 years were asked to rate the quality of output on a scale of 1 to 5 where, 5- Excellent, 4- Good, 3- fair, 2- Poor, 1-Bad

Table 3. MOS of Naturalness and Intelligibility for Marathi sentences

Sentence	Naturalness	Intelligibility
पाच वाजायला दहा मिनीट आहेत.	5	5
(phone number) :8985165767	4.9	4.9
कबूतर ताशी 100 मैल वेगानं उडू शकत नाही	4.7	4.6
{negative sentence} राम वेडा नाही.	4.9	4.9
{negative sentence} तो जाणार नाही.	4.8	4.9
{Interrogative sentence} तू काम करशील ना?	4.8	4.8
(Time) : १०:५० (10:50)	4.8	4.9
(Date) : १०/१२/२०१६ (10/12/2016)	4.9	4.8
(countable number):१६,४३,५४८ (16,43,548)	4.8	4.9
(Jodakshar): आज दोन मार्च आहे.	4.8	4.8

7. Discussion and Findings

From the above MOS table, it is seen that the developed system generates a satisfactory level of output. At the beginning of the research, we split the sentence into words and letters. It is observed that if we concatenate letters, then the naturalness of sound and pronunciation is not up to the mark. The problem is fixed up by using words directly rather than using letters. The database of such commonly used words is created and preserved. If the word is present in the database, it will be counted as full, and a corresponding sound signal is generated. If the word is not present in the stored database, then the word division takes place to form letters. The algorithm is tested for combinations of letters, numbers, digits, dates, time and joint letters. The speed variation is also tested with the proposed algorithm.

The system developed by TDIL (Technology Development for Indian Languages, MEITY) is available on their website. The system generates Marathi speech only one sentence at a time. One more system which TDIL suggests is IndianTTS, which is web-based software. Few other software/app downloaded from websites are also evaluated for their performance for selected 10 sentences. The selected sentences contain the date, time, joint letters, numbers and countable.

It is observed that for Text to Speech (TTS), the pronunciation of "मार्च" is not proper. It is a joint letter. In TextToSpeech application it generates a somewhat Hindi accent; also, the pronunciation of "मार्च" is not proper. There should be prosody conversions, but any of them hardly supports it. The option of speed variation is available in the algorithms mentioned above, but the naturalness is disturbed when the speed varies.

Conclusion

In this paper, we have developed a python based text to speech system to convert raw Marathi text data into easily processable sentences and words. The developed system assigns phonetic transcriptions to each word from the library database to create a natural effect in synthetic speech. The developed system is tested for its performance for speaking speed variation.

Funding

The authors are thankful to the School of Education, Banaras Hindu University, for financially supporting the research work by providing a grant.

References

[1] Repe Madhavi R., S. D. Shirbahadurkar, and Smita Desai, "Prosody Model for Marathi Language TTS Synthesis With Unit Search and Selection Speech Database," In *International Conference on Recent Trends In Information, Telecommunication and Computing (ITC). IEEE,* pp.362-364, 2010.

[2] Barhate, Sanket, Shrutikshirsagar, Niramaysanghvi, Kaminisabu, Preetirao, and Nandini Bondale, "Prosodic Features of Marathi News Reading Style," *Region 10 Conference (TENCON), IEEE,* pp.2215-2218, 2016.

[3] Kiruthiga, S., and K. Krishnamoorthy, "Design Issues In Developing Speech Corpus for Indian Languages - A Survey," In *International Conference on Computer Communication and Informatics (ICCCI), IEEE,* pp.1-4, 2012.

[4] S. L. Joshi, V. K. Bairagi, "Recent Trends in Text to Speech Synthesis of Indian Languages," *International Journal of Helix,* vol.9 , no.3, pp 4931- 4936, 2019.

[5] Kishore, S. P., Rohit Kumar, and Rajeev Sangal, "A Data-Driven Synthesis Approach for Indian Languages Using Syllable as Basic Unit," In *Proceedings of Intl. Conf. on NLP (ICON),* pp.311-316, 2022.

[6] Panda, Soumya Priyadarsini, Ajit Kumar Nayak, and Srikanta Patnaik , "Text-to-Speech Synthesis With an Indian Language Perspective," In *International Journal of Grid and Utility Computing,* vol.6, no.3-4, pp.170-178, 2015.

[7] Oloko-Oba Mustapha O, Ibiyemi T.S, Osagie Samuel E, "Text-to-Speech Synthesis Using Concatenative Approach," *In International Journal of Trend in Research and Development,* vol.3, no.5, 2016

[8] Sangramsing Kayte, Kavita Waghmare, Dr. Bharti Gawali, “Marathi Speech Synthesis: A Review,” *In International Journal on Recent and Innovation Trends in Computing and Communication,* vol.3, no.6, 2015.

[9] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte, “Di-Phone-Based Concatenative Speech Synthesis Systems for Marathi Language,” *In IOSR Journal of VLSI and Signal Processing,* vol.5, no.5, 2015.

[10] Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte, “A Corpus-Based Concatenative Speech Synthesis System for Marathi,” *In IOSR Journal of VLSI and Signal Processing,* vol.5, no.5, 2015.

[11] http://Tcts.Fpms.Ac.Be/Synthesis/Mbrola.html

[12] http://Espeak.Sourceforge.Net/

[13] https://www.Cstr.Ed.Ac.Uk/Projects/Festival/

[14] Murthy, Hema A., Ashwin Bellur, Vinodh Viswanath, Badri Narayanan, Anila Susan, G. Kasthuri, K. Sreenivasa Rao, “Building Unit Selection Speech Synthesis in Indian Languages: an Initiative by an Indian Consortium,” *In Proceedings of COCOSDA,* pp 358-361, 2010.

[15] Pradhan, Abhijit, Anusha Prakash, S. Aswin Shanmugam, G. R. Kasthuri, Raghava Krishnan, and Hema A. Murthy, “Building Speech Synthesis Systems for Indian Languages,” *In Twenty-First National Conference on Communications (NCC),IEEE,* pp.1-6, 2015.

[16] Tabet, Youcef, and Mohamed Boughazi, “Speech Synthesis Techniques-A Survey,” *In 7th International Workshop on Systems, Signal Processing and Their Applications (WOSSPA), IEEE,* pp.67-70, 2011.

[17] https://Cdac.In/Index.Aspx?Id=Mc_St_Speech_Technology

[18] http://Tdil-Dc.In/Index.Php?Option=Com_Vertical&Parentid=85&Lang=En

[19] https://www.Fon.Hum.Uva.Nl/Praat/

[20] http://Ivr.Indiantts.Co.In/En/Home

[21] https://Play.Google.Com/Store/Apps/Details?Id=Com.Sinwho.Tts

[22] www.Nltk.Org

[23] K.Sureshkumar and Dr.P.Thatchinamoorthy, “Speech and Spectral Landscapes Using Mel-Frequency Cepstral Coefficients Signal Processing,” *SSRG International Journal of VLSI & Signal Processing,* vol.3, no.1, pp.5-8, 2016. *Crossref,* https://doi.org/10.14445/23942584/IJVSP-V3I1P102

[24] ZENG Runhua, ZHANG Shuqun, “Improving Speech Emotion Recognition Method of Convolutional Neural Network,” *International Journal of Recent Engineering Science*, vol.5, no.3, pp.1-7, 2018.
Crossref, https://doi.org/10.14445/23497157/IJRES-V5I3P101.

[25] Petra Wagner, Jonas Beskow, Simon Betz , Jens Edlund , Joakim Gustafson , Gustav Eje Henter , Sébastien Le Maguer , Zofia Malisz , Éva Székely , Christina Tånnander , Jana Voße, “Speech Synthesis Evaluation — State-of-the-Art Assessment and Suggestion for A novel Research Program,” *In Proceedings of the 10th Speech Synthesis Workshop (SSW10),* 2019.

[26] Smita S. Hande, “A Review of Concatenative Text to Speech Synthesis,” In *International Journal of Latest Technology in Engineering, Management & Applied Science,* 2014.

[27] Abitha A and Lincy K, “A Faster RCNN Based Image Text Detection and Text to Speech Conversion,” *SSRG International Journal of Electronics and Communication Engineering,* vol.5, no.5, pp.11-14, 2018.
Crossref, https://doi.org/10.14445/23488549/IJECE-V5I5P103.

[28] Anamika Baradiya and Vinay Jain, “Speech and Speaker Recognition Technology Using MFCC and SVM,” *SSRG International Journal of Electronics and Communication Engineering,* vol.2, no.5, pp.6-9, 2015.
Crossref, https://doi.org/10.14445/23488549/IJECE-V2I5P105.

[29] *Yin Zhigang,* “An Overview of Speech Synthesis Technology,” *In Eighth International Conference on Instrumentation and Measurement, Computer, Communication and Control, IEEE,* 2018.