Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 7, Issue 07, July -2020

Text-to-Speech Mobile Application for Maithili Language

Amit Kumar Jha[1], Piyush Pratap Singh[2]

Center for Information and Language Engineering
Mahatma Gandhi Antarrashtriya Hindi Vishwvidyalaya,
Wardha, Maharashtra 442001, India

Abstract — *This paper discusses development of Text-to-Speech (TTS) mobile application for Maithili Language. Maithili is one of the twenty-two languages included in the Eighth Schedule to the Constitution of India. This TTS mobile application is an Android based application which is able to convert Maithili text to its correspondence speech form. This application is developed on Android studio 3.0. This application supports Devanagari script as inputting the text. The nature of Maithili language is syllabic similar to other Indo-Aryan languages so syllable is taken as the basic unit for concatenative method. For enhancing the naturalness of the synthesized voice, user can control the pitch and speed using Seek Bar at the time of voice generation. Subjective and Objective evaluation conducted by 100 native speakers. A questionnaire has been created to evaluate the Maithili text-to-speech system. Naturalness test, Comprehensibility test, Productivity test conducted for testing the application and summary of results are good. For this evaluation Mean Opinion Score (MOS) is used. This is the first TTS mobile application for Maithili Language.*

Keywords- *Text-to-Speech System, Android application, Maithili, Speech processing, Natural Language Processing*

I. INTRODUCTION

A Text-to-Speech System converts raw text into its correspondence spoken form. It is a useful tool and it can be of significant aid to communication for visually impaired people other than its role in telecommunication, industrial and educational applications. It is one of the application areas of Natural Language Processing. This research paper discuss about Text-to-Speech mobile application for Maithili language. An android based text-to-speech system for Maithili language has been developed in this research, which is able to convert Maithili text to its correspondence speech form. Research on the text-to-speech system has been in place for the past several decades. But earlier this type of research was mainly limited to English or other major European languages. In Indian languages also, such research is mainly concentrate in major languages like Hindi, Bengali, Tamil, Telugu and Punjabi. There isn’t any such system has yet been developed in Maithili language. This system is the first system for Maithili language.

Maithili (EGIDS 0-4) is a language of an Indo-Aryan language family. It is primarily spoken in the state of Bihar, and also in some other parts of India and Eastern Nepal approximately by more than 30 million people in total [1]. Maithili is one of the twenty-two languages included in the Eighth Schedule to the Constitution of India. In India, it is mainly spoken in Bihar and Jharkhand. It has got the status of the second official language in Jharkhand on 2008. Apart from India, this language is mainly spoken in Nepal. Maithili speaking people live in the Terai region of Nepal. In terms of population, this language is the second most spoken language in Nepal after Nepali. Maithili is written mainly in four scripts - Mithilachar/Tirhuta, Kaithi, Nevari and Devanagari. It is primarily written in Devanagari script. In ancient times Mithilachar/Tirhuta script was used to write Maithili. It is the oldest script of Maithili. The Kaithi script was mainly used by the scribe in official work. Even today, you can see this script in Khesra, an old account of land. The Newari script was mainly used by the Newar tribes in Nepal to write Maithili. After being incorporated in 8th schedule in the year 2002, it slowly is picking up to be used in government, education and other official communicative contexts [2].

According to market research firm techARC, India had 502.2 million smartphone users as of December 2019, which means over 77 per cent of Indians are now accessing wireless broadband through smartphones. Samsung with 34 per cent led the smartphone installed base in the calendar year 2019, followed by Xiaomi at 20 per cent, Vivo at 11 per cent and OPPO at 9 per cent, respectively [3]. From this data it is clear that the smart phones user in India is increasing rapidly and approx 75 percent people used android based phone from these three companies only. Mobile Operating System Market Share in India – June 2020 is shown in Figure No. 1 through which it is clear that 95.07 percent user use Android mobile application [4]. So Android Operating system based mobile application is developed first as a Text-to-Speech mobile application for Maithili Language.

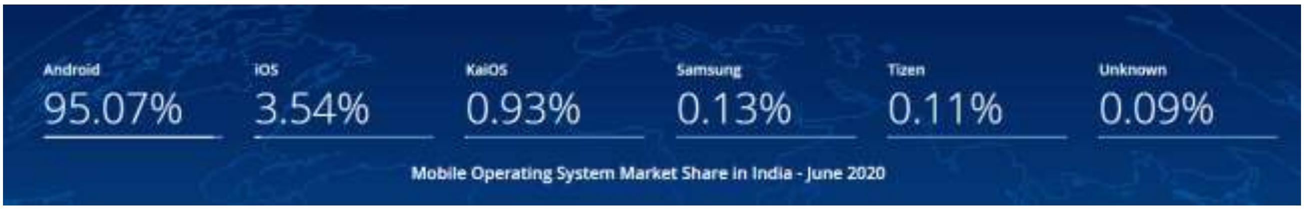


Figure 1. Mobile Operating System User statistics

This paper is divided into six sections. First section is about Introduction to the application and Maithili language. Section 2 is about method used for developing the application. User interface and working procedure of Maithili TTS application is explained in section 4. Section 5 is about evaluation of the application. At last section future scopes and future planning of the research are discussed.

II. METHOD

There are mainly three types of speech synthesis - Formant synthesis, Articulatory synthesis and Concatenate synthesis. Formant synthesis is based on the source-filter model of speech. It has two basic structures - parallel and cascaded, but some of its composite forms are used for good performance. Articulatory synthesis tries to model the human vocal organs as fully as possible, so it is probably the most satisfactory method to produce high quality synthesized speech. On the other hand, it is one of the most difficult methods to implement and the computational load is much higher than other common methods. Thus, it receives less attention than other synthesis methods and has not yet achieved the same level of success. In the last few decades, the trend of manufacturing speech synthesizers using the concatenate method has increased. Unit selection synthesis is most commonly used in concatenate synthesis. The speech produced by the text-to-speech system created by unit selection synthesis has a greater naturalness. Because it uses a small amount of digital signal processing on recorded speech. Digital signal processing often makes speech less natural, although some systems use small amounts of signal processing to smooth the wave form at the connection point. For developing this application Concatenate method is used. Under concatenate method unit selection method is used. Syllable is used as a unit for this system. For most frequent words of Maithili language words is used as a unit of concatenation.

Several organizations are working in the field of the development of TTS. They use different types of methodologies and techniques. Some milestone in the field of TTS development is shown in the figure 2. TTS systems, for different Indian languages, have been developed using various approaches such as articulatory synthesis [5], formant synthesis [6], unit selection synthesis (USS) [7], [8] and HMM based speech synthesis (HTS) [9]. Text-to-speech synthesis system for mobile applications is developed for Tamil language [10].

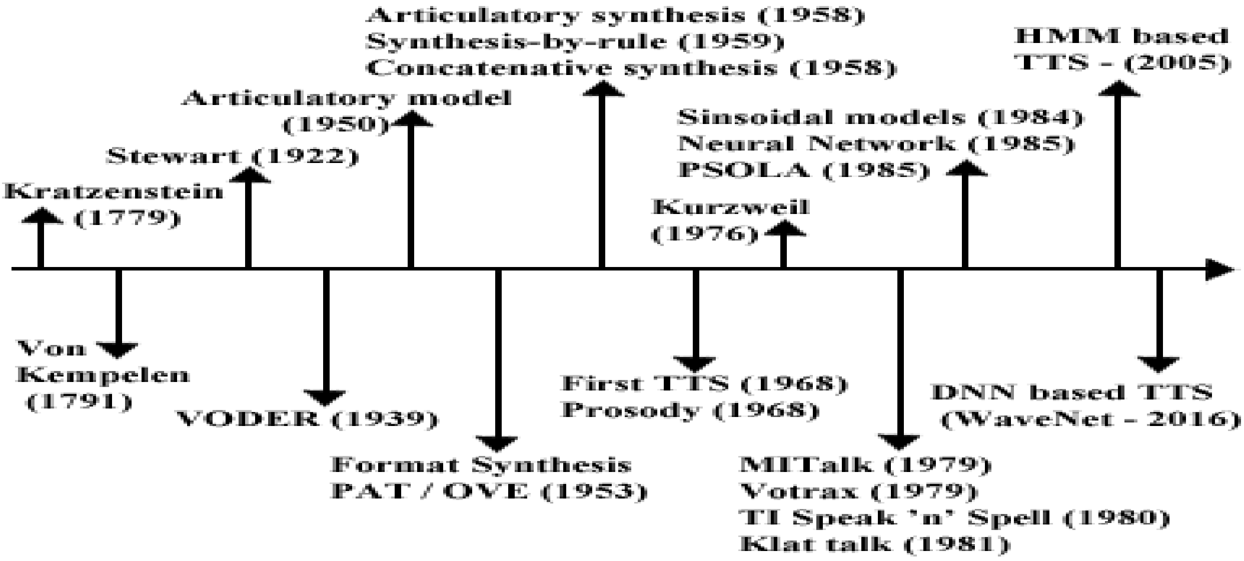
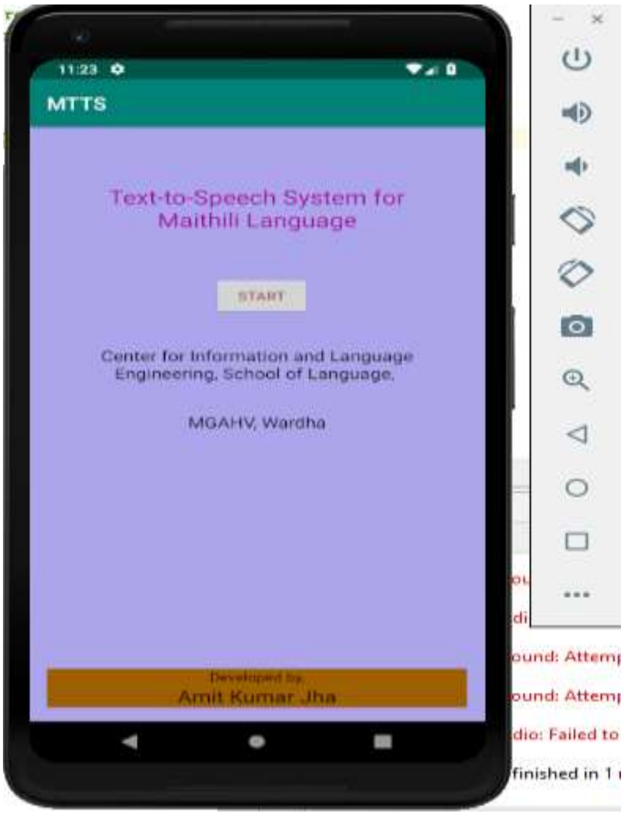


Figure 2. Milestone in the field of TTS

III. USER INTERFACE AND WORKING PROCEDURE OF MAITHILI TTS APPLICATION

This system is developed in Android Studio version 3.3. This system consist two Graphical User Interface (GUI) forms as a user interface. The first form is the introduction form, which details the introduction related to the system (see in Figure no. 3). The second form is the main form of system, In this form we can see two text boxes, and two seek bars and three buttons (see in Figure no. 4). This android based mobile system has been developed for Maithili language. It synthesizes speech by the concatenate synthesis method. Maithili text is inputted in the first text box. The normalized text is displayed in the second text box. The first seek bar is for tone/pitch and the second seek bar is for speed. The user can increase or decrease both of these as his wish or requirement. The three buttons of the system is used for different purposes. The first button is use for normalize input text, the second button is for playing the speech and the third button can empty both text boxes. The working procedure of the system is that first input Maithili text in the first text box. The input text should be in Unicode font and in Devanagari script. After this, the input text is normalized by the system when the user clicks on the normalization button. In order to normalize, first the non-standard word (which needs to be normalized) is identified from the input text. These words are then converted into standard words with the help of given rules and dictionaries. After normalizing the text, the text is first segmented into the sentence level. After that we break that text at the word level. Then that word is searched in a database made of the words; if that word is found in that database then the sound file for that word is added to a list called Temp. But if that word is not found in that database, then that word is segmented into syllable based on the syllabification rule of Maithili language and based on the location in the word of that syllable, that syllable is searched in a database made of syllables. When that syllable is found in that database, the sound file of the sentence is created by adding words and all the words by adding files of that syllable. Then by clicking on the other button say it to play that sound file from the system, the system plays the sound file. In this system, two things have been added for prosodic feature - sound and speed. Both of these can be reduced or increased as needed. Seek bars have been used to both increase and decrease this.


11:23

MTTS

Text-to-Speech System for Maithili Language

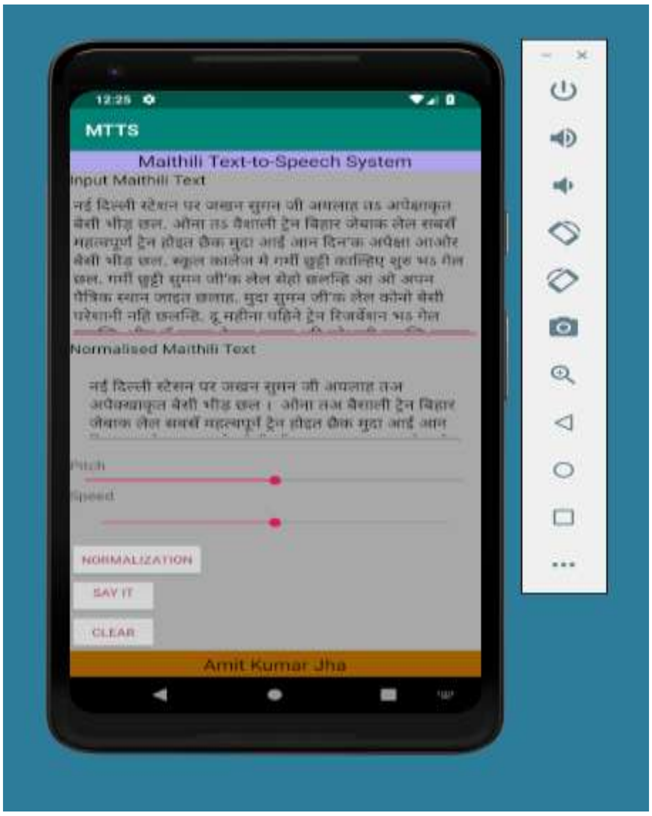
START

Center for Information and Language Engineering, School of Language,

MGAHV, Wardha

Developed by
Amit Kumar Jha

Figure 3. User Interface (First form)


12:25

MTTS

Maithili Text-to-Speech System

Input Maithili Text

नई दिल्ली स्टेशन पर जखन सुमन जी अयलाह तऽ अपेक्षाकृत बेसी भीड़ छल. ओना तऽ वैशाली ट्रेन बिहार जेबाक लेल सबसँ महत्वपूर्ण ट्रेन होइत छैक मुदा आई आन दिनक अपेक्षा आओर बेसी भीड़ छल. स्कूल कालेज मे गर्मी छुट्टी कालिहए शुरु भऽ गेल छल. गर्मी छुट्टी सुमन जी'क लेल सेहो छलन्हि आ ओ अपन पैत्रिक स्थान जाइत छलाह. मुदा सुमन जी'क लेल कोनो बेसी परेशानी नहि छलन्हि. दू महीना पहिने ट्रेन रिजर्वेशन भऽ गेल

Normalised Maithili Text

नई दिल्ली स्टेसन पर जखन सुमन जी अयलाह तअ अपेक्षाकृत बेसी भीड़ छल । ओना तअ वैशाली ट्रेन बिहार जेबाक लेल सबसँ महत्वपूर्ण ट्रेन होइत छैक मुदा आई आन

Pitch

Speed

NORMALIZATION

SAY IT

CLEAR

Amit Kumar Jha

Figure 4. User Interface Main form

IV. CORPUS DEVELOPMENT

Text and speech corpus is an important and necessary step for building a text-to-speech system. However, its developing process may differ from one language to another and it is also based on the availability of resources in that language. Database creation involves collecting text for the purpose of recording. Text database creation begins with the collection of phonologically balanced text data. The collected data is recorded in the studio by the native speaker of the language (who has a good command on that language). To build this system, text corpus has been collected from various fields like - literature, science, tourism, politics, story, contemporary, and drama. Books, newspapers, magazines, web pages and blogs have been used mainly as a source for collecting text corpus. Among these sources, books published in Maithili by NCERT, Bihar, Patna of Maithili language for class 1-8 were very helpful. Syllable combinations which could not be covered from these sources were taken from Maithili Dictionaries. Kalyani shabdkosh proved to be extremely useful among others [11].
The recording is happen at a sampling frequency of 16 KHz / 16 bits. Recording speech was segmented at the level of sentence, word and letter with the help of Praat.

V. EVALUATION OF THE APPLICATION

How much efficiently any text-to-speech system is working, it is primarily investigated by three components - productivity testing, naturalness testing and intelligibility testing. For this, both subjective and objective type evaluation have been done. Objective evaluation does not depend on the individual sentiment of the evaluators. A questionnaire has been created to evaluate the Maithili text-to-speech application. This questionnaire evaluates all the aspect of the application. In the questionnaire, at first all personal information like name, age, gender, address and occupation, mother tongue etc. has been asked to be filled. Thereafter, it is asked about the knowledge level about Maithili. After this, some sentences have been given from different domains of Maithili, on the basis of which the evaluator was asked to evaluate the system from 1 to 5 levels. Where 1 means very bad, 2 means bad, 3 means normal, 4 means good and 5 means very good. Using user feedback Mean opinion score (MOS) is calculated. At the time of choosing the evaluator, special attention was given to the fact that he should not have any problem related to hearing and didn't any hearing loss in past. The MOS score of the application is calculated 8.4. The quality of synthesized speech in terms of intelligibility and naturalness is evaluated to be approximately 84 percent. After that end user evaluation is also conducted. About 100 people were included in this process of evaluation, in which the ratio of men and women was kept almost equal. About 78.6 percent of the people consider the productivity of this system to be very good and 11.4 percent people consider it good (Fig 5).

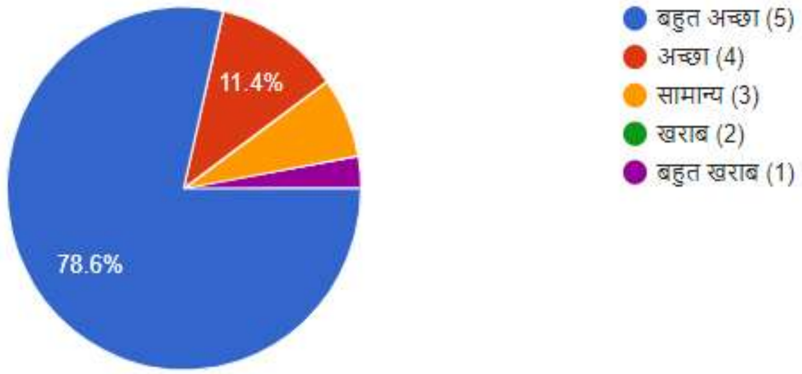


Figure 5. Productivity of the Speech of the Application

The level of speech intelligibility produced by this system is considered very good by 71.4 percent of people and good by 20 percent of people (Fig 6).

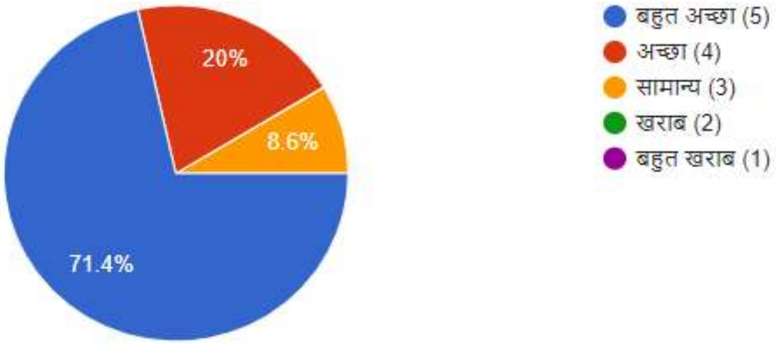


Figure 6. Intangibility of the speech generated by Application

The level of naturalness of speech produced by this system is considered very good by 74.3 percent of people and good by 14.3 percent people (Fig 7).

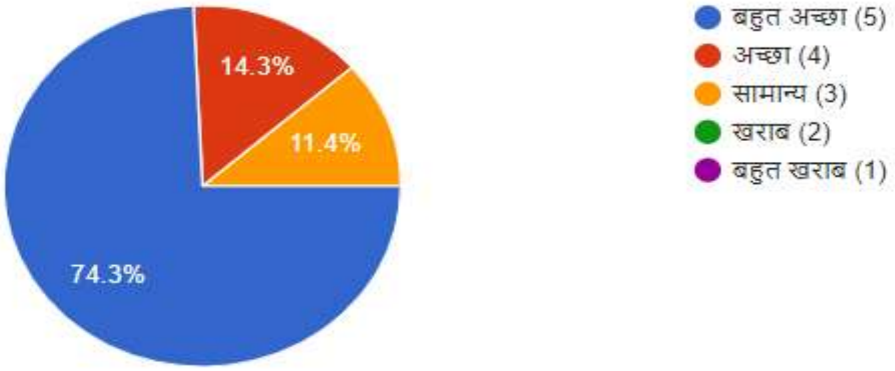


Figure 7. Naturalness of the speech of Application

When the evaluator was asked at what level does this system work best? 58.6 percent of the people consider at the paragraph level, 24.3 percent of the people at the word level and 17.1 percent at the sentence level (Fig 8).

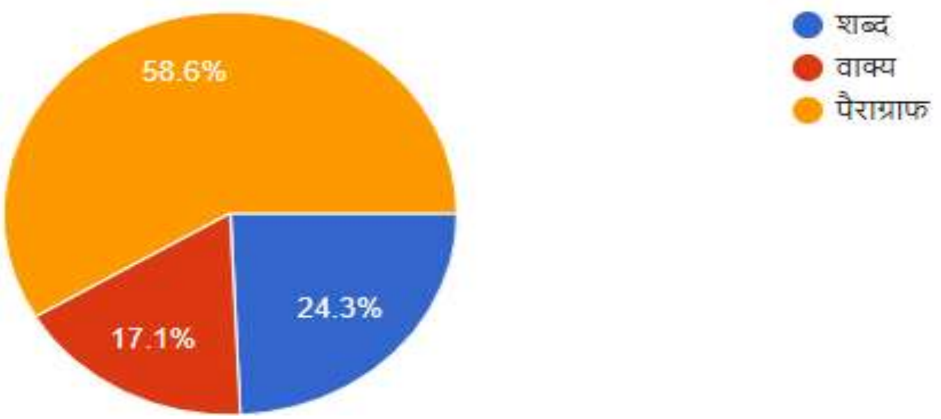


Figure 8. Level where System works better

VI. FUTURE SCOPE AND FUTURE PLANNING

Text-to-Speech Mobile Application for Maithili Language is now developed for Android operating system. The future planning is to developed same application for other operating system such as iOS and windows. To add more Indian language in this application is one of the other scopes of this application. In future we can also add Optical character reader (OCR) with this application. If Speech-to-Text application for Maithili language will be developed in future then both the systems together play an important role in Man-machine interaction.

REFERENCES

[1] https://www.ethnologue.com/language/mai , Access on: 17-02-19, 16:52.

[2] A. K. Jha, P. P. Singh and P. Dwivedi, "Maithili Text-to-Speech System," 2019 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2019, pp. 1-6, doi: 10.1109/CONECCT47791.2019.9012903.

[3] “Smartphone Users in India Crossed 500 Million in 2019, States Report”, Access on https://www.news18.com/news/tech/smartphone-users-in-india-crossed-500-million-in-2019-states-report-2479529.html on 24/05/2020, 11:40.

[4] https://gs.statcounter.com/os-market-share/mobile/india Access on: 20/06/2020 11:45.

[5] J. Benesty, M. M. Sondhi, and Y. A. Huang, Springer Handbook of Speech Processing. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007

[6] W. Lawrence, “The synthesis of speech from signals which have a low information rate,” Communication Theory, pp. 460–469, 1953.

[7] D. H. Klatt, “Review of text to speech conversion for english,” The Journal of the Acoustical Society of America, vol. 82, no. 3, pp. 737–793, 1987. [Online]. Available: http://scitation.aip.org/content/asa/journal/jasa/82/3/10.1121/1.395275

[8] A. Hunt and A. Black, “Unit selection in a concatenative speech synthesis system using a large speech database,” in Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, vol. 1, May 1996, pp. 373–376 vol. 1.

[9] H. Zen, K. Tokuda, and A. W. Black, “Statistical parametric speech synthesis,” Speech Communication, vol. 51(11), pp. 1039–1064, 2009

[10] K. P. Sarathy, and A. G. Ramakrishnan. "Text to speech synthesis system for mobile applications." In Proc. Workshop in Image and Signal Processing (WISP-2007), IIT Guwahati, 2007, pp. 74-77.

[11] G. Jha, Eds., “Kalyani Kosh A Maithili-English Dictionary”, Maharajadhiraj Kameshwar singh Kalyani foundation, Darbhanga, 1999