Marathi Speech Synthesis: A review

Sangramsing Kayte[1]
Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India
bsangramsing@gmail.com

Kavita Waghmare[1]
Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India
kavitawaghmare22@gmail.com

Dr. Bharti Gawali[2]
Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India
bharti_rokade@yahoo.co.in

Abstract**—** This paper seeks to reveal the various aspects of Marathi Speech synthesis. This paper has reviewed research development in the International languages as well as Indian languages and then centering on the development in Marathi languages with regard to other Indian languages. It is anticipated that this work will serve to explore more in Marathi language.

Keywords—*speech synthesis; festival; articulatory synthesis; formant synthesis;concatenative synthesis, Hidden Markov Model.*

*****

I. INTRODUCTION

Speech synthesis is a procedure of automatic generation of spoken language by computer [1]. It is also referred as text-to-voice communication. In this process a text in normal speech is changed into voice communication. The goal of speech synthesis is to produce a machine having an understanding and natural sounding voice for communicating [2][3]. A general architecture of a text-to speech system is shown in figure 1 consist of various components such as text analysis, text normalization. The text-to-speech synthesis procedure consists of mainly two phases. The foremost one is text analysis, where the input text is transliterated into a phonetic or some other linguistic representations, and the second one is the generation of speech waveforms, where the acoustic output is created from the phonetic and prosodic information. These two forms are broadly mentioned as high and low level synthesis [4]. A data processor system utilized for this determination is addressed a speech synthesizer and can be enforced as a software or a hardware

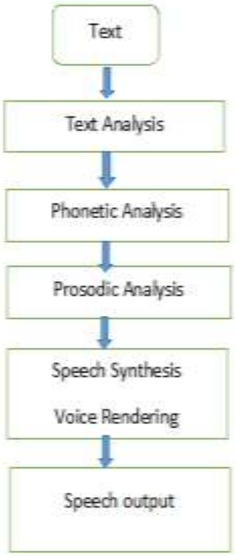

Figure 1 shows Architecture of Text-To-Speech system

Text-to-speech system have an enormous range of applications. Speech synthesis has been widely researched in the last four decades. The quality and intelligibility of the synthetic speech produced using the latest methods have been unusually well for most of the applications [5] [6]. Speech synthesis can be habituated to understand a written text aloud, an email, SMS, newspaper, talking books and lots more. This paper is presented in seven sections in which section II the techniques used in speech synthesis, section III applications of a speech synthesis system, section IV speech synthesis in national and internal scenario, section V tools of speech synthesis, section VI conclusion.

II. TECHNIQUES OF SPEECH SYNTHESIS

There are basically four techniques of speech synthesis namely Articulatory, Formant, Hidden Markov Model and Concatenative synthesis.

A Articulatory synthesis

Articulatory synthesis directly models the physical articulators such lips, jaws, tongue, soft palate, and so on [7]. In this human speech production system is modeled. It involves simulating the acoustic parts of vocal tract and its dynamic movement. The command parameters are sub-glottal pressure, vocal cord tension, and the relative location of the different articulatory organs. It is really hard to obtain accurate three-dimensional vocal tract representations and modeling system with a special set of parameters.

B. Formant synthesis

Formant synthesis is a rule based synthesis, which describes the resonant frequencies of the vocal tract. This method uses source-filter model of language output. The parameters controlling the frequency response of the vocal tract filter-and those controlling the source signal-are updated at each phoneme. Excitation produced by the root passes through the filter, is qualified by the resonance characterizes of the filter to create language. It was the most used method in the last decade.

This method applies the parameters such as fundamental frequency, voicing and noise levels over a point of time to make a waveform of artificial speech.

C. Hidden Markov Model

The Hidden Markov Model (HMMs) is widely-used statistical models to characterize the sequence of speech spectra and have successfully been applied to speech recognition and speech synthesis systems. This system simultaneously models, spectrum, excitation, and continuance of speech using content-dependent HMMs and generates speech waveforms. HMM creates stochastic models from known utterances and compares the probability that the unknown utterance can be engendered by each model. It also includes various methods for text to voice communication such as Dynamic Features (Delta and Delta-Delta parameters of Speech). HTK (Hidden Markov Model Toolkit) is a software toolkit for handling HMMs applications that employ HMMs for speech synthesis. This technique uses very less memory, but consumes large CPU resources. This approach gives good prosody features with natural sounding language.

D. Concatenative Synthesis

Concatenative synthesis is the most uncomplicated method to synthesize the speech which is got by concatenating the different sentences, words, syllables, phones, iPhones, and triphone. These are already stored to get the desired output language. It requires large databases, sometimes it is quite impossible to store. More natural sounding speech is produced by this technique. Sometimes, the difference between natural variations and the nature of the automated techniques for segmenting the waveforms may result in audible glitches in the production. There are three sub-types of Concatenative synthesis:

1) Unit selection Synthesis or Corpus based synthesis: Unit selection synthesis is also referred as corpus based synthesis. It uses large database. During database creation, each recorded utterance is segmented into some individual phones, syllables, morphemes, words, phrases, and sentences. An index of the units in the speech database is then made based on the segmentation and acoustic parameters such as fundamental frequency, pitch, duration, the status of the syllable and previous and next phones. This method provides naturalness in output speech as compared to other techniques.

2) Diphone Synthesis: Diphone synthesis techniques require less database as compared to the unit selection synthesis. It uses two adjacent phones to make the speech waveform. But this techniques suffers through the problem of coarticulation.

3) Domain specific synthesis: Domain specific synthesis is associated to the particular field. In this database consists of language related to that particular line of business and that are concatenated to create the end product.

III. Applications of Speech Synthesis

The text-to-speech system have an enormous range of applications. Some of those are discussed below:

In telecommunication service: Textual information over the phone can be synthesized and used for calls that require less connectivity.

In e-governance service: In various e-administration services like polling center information, land record information, application tracking and monitoring.

Aid to disabilities: T-T-S can give invaluable support to voice handicapped individuals with the avail of an especially design keyboard and fast sentence assembling program, also helpful for visually handicapped.

Voice browsing: T-T-S is the backbone of voice browsers, which can be manipulated by voice instead of mouse and keyboard, thus allowing hands-free and eye free browsing.

Vocal monitoring: At times oral information is said to be more effective than its counterpart. Hence the thought of incorporating speech synthesizers in the measurement and control systems like cockpits to prevent pilots from being overwhelmed with visual data.

Complex interactive voice response systems: With the reinforcement of good quality speech recognizers, speech synthesis systems are capable to create complex interactive voice response systems a reality.

Multimedia, man-machine communication: Multimedia is the first, but a promising move in the direction and it includes talking books and toys, mail and document readers.

IV. Speech Synthesis in International and National Scenario

There are various foreign languages in which work has been done or going on such as American English, Japanese, European Portuguese, Arabic, Polish, Korean, German, Turkish, Mongolian, and Greek. While focusing towards Indian languages there are total 22 official languages out of which Hindi, Malayalam, Kannada, Bengali, Oriya, Punjabi, Gujarati, Telugu, and Marathi are being in focused. Different systems have been developed in these languages such as Dhvani, Shruti, HP Lab, Vani. Various institutions has been working on speech synthesis such as IIIT-H, CDAC-Mumbai, CDAC-Pune, and IIT-Madras in different languages. They have built several applications such as e-speak, a-speak, I-speak, Sandesh Pathak but in Hindi, Telugu and other languages. Marathi is an Indo-Aryan language. It is the co-official language in Maharashtra and Goa states of Wereten India and is one of the 23 official languages of India. The basic unit of Marathi writing system are the Aksharas which are an orthographic representations of speech sounds. An Aksharas are the combination of consonants and vowels. As it seen very less work has been done in Marathi

[6] [7] [8]. C-DAC is the institution which has been working in the area of speech synthesis since 25 years. This institute has developed text-to-voice communication system in Hindi, Malayalam, Bangla, Mizo and Nepali. They have developed ESNOLA based Bangla synthesis techniques. It is called as "BANGLA VAANI". Other systems are Mizo Text reading system, Mozhy TTS. [8] [9] [10] [11] [12][13][14].

A. Applications available in International languages

Classic text to speech engine: It is a Text To Speech Engine from SVOX, in combination with 40+ male/female voices in more than 25 languages that allows to read aloud texts from e-book, navigation, translation and other apps. The languages in which it is available are Arabic (male), Australian English (female), Brazilian Portuguese (female), Canadian French (male/female), Cantonese (female), Czech (female), Danish (female), Dutch (male/female), Finnish (female), French (male/female), German (male/female), Greek (female), Hungarian (female), Italian (male/female), Japanese (female), Korean (female), Mandarin (female), Mexican Spanish (male/female), Norwegian (female), Polish (female), Portuguese (male/female), Russian (male/female), Slovak (female), Spanish (male/female), Swedish (female), Thai (female), Turkish (male/female), UK English (male/female), US English (male/female).

a) IVONA: It is a TTS system available in 17 languages. It gives natural sounding and more accurate voices. It is compatible with Windows, Unix, Android, Tizen, iOS based systems. The compatible languages are American, Australian, British, Welsh, German,French,Castilian,Icelandic,Italian,Canadia,Dutch,European,Brazilian,Polish,Romanian,Russian,Danish.

b) CereVoice Engine: It gives support to 9 languages. It can be easily deployed with any variety of English voices. It is available in English, French, Spanish, Italian, German, Portuguese, Japanese, Dutch, and Catalan.

c) eSpeak: It supports 51 languages over the world. It lets in different voices, whose features can be changed. It partially supports SSML (Speech Synthesis Markup Language).The languages supported by eSpeak are Afrikaans, Albanian, Aragonese, Bulgarian, Danish, Dutch, Cantonese, Catalan, Croatian, English, Esperanto, Estonain, Farsi, Finnish, French, Georgian, German, Greek, Hindi, Hungarian, Icelandic, Indonesian, Irish, Italian, Kannada, Kurdish, Latvian, Lojban, Malayalam, Malaysian, Nepalese, Norwegian, Polish, Portuguese,Punjabi,Romanian,Russian,Serbain,Slovak,Spanish, Swahili,Swedish,Tamil,Turkish,Vietnamese.

d) Google Text-to-Speech: This TTS system supports 15 languages. It is a low level speech synthesis. The languages supported by this TTS are Dutch, English (India), English (United Kingdom), English (United States), French, German, Hindi, Italian, Japanese, Korean, Polish, Portuguese, Russian, Spanish.

B. Applications available in Indian languages:

aSpeak: It is an application available in 2 Indian languages such as Telugu and Hindi. The speech produced is intelligible, natural, and clear and can be used at different speeds.

Sandesh Pathak: It is an application which supports 5 Indian languages namely Hindi, Marathi, Tamil, Telugu, and Gujarati. It is mainly used in agricultural based application. The text can be heard at various speeds.

Shruti: It is a TTS system developed using Concatenative speech synthesis for two languages namely Hindi and Bengali. It is designed in such a way that it can be extended in any other languages.

HP Labs: It is a TTS system developed in Hindi language.

Vani: It is a system to be developed in Hindi language

Dhvani: It is a TTS designed for 11 Indian languages such as Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Oriya, Punjabi, Tamil, Telugu, and Pashto. The development of this system is still in progress. It is based on diphone concatenation algorithm.

V. TOOLS FOR SPEECH SYNTHESIS

The festival offers a general framework for building speech synthesis systems. It was developed at CSTR (Center for Speech Technology Research), at the University of Edinburgh by Alan Black and Paul Tayor and in co-operation with CHATR, Japan. The festival is a multi-lingual framework freely available software. It is compatible to work with all types of voices and in different platforms. It is provided with Scheme-based scripting language which means that it is fully controllable at run time without re-compiling the system. The system is written in C++ and supports residual excited LPC and PSOLA methods and MBROLA databases. It has general utterance representation which provides easy and efficient way for writing functions. It supports many waveforms formats. This framework also contains various tools such as Viterbi decoder-gram support, regular expression, matching, linear regression support, CART support, weighted finite state transducers, and stochastic context free grammars. It is also provided with Server-Client model. The system is developed for three different aspects. For the people who want to use the system just for text-to-speech, for people who are developing language systems and wish to include synthesis output, such as different voices, specific phrasing, dialog types and so on, and for those who are developing and testing new synthesis methods. It provides all the tools and its documentation to build new voices. The festival is primarily designed as a component of a larger speech application. It can be used to simply synthesize text file.

VI. CONCLUSION

Speech synthesis is a continuously developing technology since many years. This paper has presented various aspects of speech synthesis such as methods, tools, techniques, applications currently available in various languages. It is seen that the work in the area of Marathi speech synthesis is very less as compared to other Indian languages and it can be explored more.

REFERENCES

[1] Shruti Gupta, Prateek Kumar, "Comparative study of text to speech system for Indian Language", International Journal of Advances in Computing and Information Technology ISSN 2277-9140 April 2012.

[2] Monica Mundada, Bharti Gawali, Sangramsing Kayte "Recognition and classification of speech and its related fluency disorders" Monica Mundada et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6764-6767

[3] 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

[4] Prof. Preeti S. Rao, "Review of methods of Speech Synthesis" M-Tech Credit Seminar Report, Electronic Systems Group, EE Dept., IIT Bombay, November, 2011.

[5] Archana Balyan, S. S. Agrawal, Amita Dev,"Speech Synthesis: A review", IJERT, vol.2 Issue 6, June 20013.

[6] A. Indumathi, Dr. E. Chandra," Survey on speech synthesis", Signal Processing: An International Journal (SPIJ), Volume (6): Issue (5): 2012.

[7] D.Sasirekha, E.Chandra ,"Text To Speech: A Simple Tutorial".

[8] Kalyan D.Bamane, Kishor N.Honwadkar,"Marathi speech Synthesized Using Unit selection Algorithm",Computer Engineering and Intelligent Systems ISSN 2222-1719(Paper)

[9] Mr.S.D.Shirbahadurkar, Dr.D.S.Bormane,"Speech synthesizer Using Concatenative Synthesis Strategy for Marathi Language",International Journal of Recent Trends in Engineering,Vol 2,No.4,November 2009.

[10] https://www.ivona.com/

[11] https://www.cereproc.com/en/products/sdk

[12] http://espeak.sourceforge.net/

[13] https://apps.mgov.gov.in/descp.do?appid=527

[14] http://sourceforge.net/projects/as-peak/

.