Development of Text-to-Speech (TTS) Synthesizer for Gujarati Language

Mr. Pravin Patel[1] Dr. Priti Sajja[2]
[1]Assistant Professor [2]Professor
[2]Department of Computer Science
[1]Anand Institute of Information Science, Anand, India [2]Sardar Patel University, India

Abstract— A Text-to-Speech Synthesizer is an application that takes input as a Gujarati text and produces its equivalent sound. To convert text into speech it uses natural language processing and speech synthesis. This paper describes existing TTS for Indian languages, objective of Gujarati TTS, Architecture of Gujarati TTS and Applications of Gujarati TTS. One approach to the generation of natural-sounding synthesized speech waveforms is to select and concatenate units from a speech database. The system used the Syllabication procedure, Phoneme and Diphone.

Key words: Gujarati Text-to-Speech (TTS) Synthesis, Concatenation Speech Synthesis, Diphone

I. INTRODUCTION

Speech is most natural way of communication in our daily life. Speech synthesis can help machines and humans to communicate in more natural way. This involves the mixing of speech technology and language technology. Speech synthesis is that the automatic generation of artificial speech signal by the Computer [5].

In the previous couple of years, this technology has been widely obtainable for many languages for various platforms starting from computer to cell phones. But for Gujarati language in Gujarat there is no successful system. The mandatory of human computer interactions through a Text To Speech (TTS) facilitate to beat the attainment barrier of common mass. It may also empower the visually impaired population and increase the probabilities of improved man-machine interaction through on-line newspaper reading from web and enhancing different data system.

The function of Text-To-Speech (TTS) system is to convert the given text to a spoken waveform. The TTS system is based on the concatenation of diphone speech units. The input text is transform into its spoken equivalent by a series of modules.

II. LITERATURE SURVEY

In India, to assist the visually impaired, vocally disabled and day to day increasing applications of speech synthesis has necessitated the development of more and more innovative text-to-speech (TTS) system. Some of the already developed TTS area unit described below.

Dhvani is a Text-to-Speech System specially designed for Indian languages. It uses diphone concatenation algorithm. In this system each language requires a Unicode parser. This speech engine has not made any attemptt to do prosody on the output. It simply concatenates basic sound units at pitch periods and plays them out [10]. Shruti is a Text-to-Speech system, which has been developed using a concatenation speech synthesis technique. This is the first text-to-speech system built specifically for two of the Indian languages, namely Bengali and Hindi [11]. HP Labs India developed a Hindi TTS system based on the open source TTS framework, Festival. This effort is a part of the Local Language Speech Technology Initiative (LLSTI), which facilitates collaboration between motivated groups around the world, by enabling sharing of tools, expertise, support and training for TTS development in local languages[12]. Vani is an Indian Language text to speech synthesizer developed at IIT Bombay, India. Generally, all existing TTS system allows user to specify what is to be spoken but does not give any control on how it has to be spoken. In Vani, a new encoding scheme has been introduced called vTrans. A vTrans file makes a person to encode what text he wants to be spoken and also the way that text to be spoken [13].

MBROLA is a high-quality, diphone-based speech synthesizer that is available in public domain. It is provided by the TCTS Lab of the Faculte Polytechnique de Mons (Belgium) which aims to obtain a set of speech synthesizers for as many languages as possible [14]. The Festival TTS synthesizer was developed in CSTR at the University of Edinburgh by Alan Black and Paul Taylor and in co-operation with CHATR, Japan. It is a freely available complete diphone concatenation and unit selection TTS synthesizer. Festival is the most complete freeware synthesis system and it includes a comprehensive manual. Festival offers a general framework for building speech synthesis systems [15]. Flite (festival-lite) is a small, fast run-time speech synthesis engine developed at CMU and primarily designed for small embedded machines and/or large servers. Flite is designed as an alternative synthesis engine to Festival for voices built using the FestVox suite of voice building tools [1].

At present, with inadequate prosodic models in place, the quality of synthetic speech generated by the synthesizers is poor. So efforts can be done for the development of prosodic models. The further work can be done to improve the naturalness and intelligibility of TTS.

III. OBJECTIVES OF PROPOSED RESEARCH WORK

The main objectives of Text-to-Speech (TTS) for Gujarati Language are

- To design architecture of TTS that supports Gujarati language.
- To study existing solution and identify limitation in order to develop a better solution.
- To develop natural and intelligible speech
- To design diphone based speech unit suitable for concatenation speech synthesis.
- To design a user interface having keyboard of Guajarati characters so that the end user can easily type the text in Gujarati.
- To make a system in which the end user can listen the text by click of a button.

IV. ARCHITECTURE

To meet above objective the system synthesizes Gujarati speech in two steps: letter to phonetic symbol conversion with prosody and Gujarati speech synthesis. The letter to phonetic symbol conversion transcribes Gujarati written strings of characters to a collection of specific features leading to phonetic symbols. Gujarati speech synthesis is a process of changing a collection of phonetic symbols to sounds. This research focuses on a tendency to develop Gujarati TTS, based on rule-based strategy to research Gujarati diphone in strings and generate sounds by concatenation technique as shown in Figure 1; we have a tendency to style Gujarati language unit analysis rules to manage with the Gujarati written strings. These rules represent the language unit structures from possible written strings. Their functions are to segment diphone unit and identify elements of each syllable [3] [4].

The text analyzer part consists of a pre-processing, a morphological analysis, a contextual analysis and a syntactic-prosodic analysis. The letter to phonetic symbol conversion process proceeds by converting abbreviations, numbers and acronyms into full text. It also separate input string into groups of words. A morphological analysis defines all possible part of speech categories for each word taken individually, on the basis of their spelling. Compound words are decomposed into their basic units in this module. The contextual analysis rule considers words in their context, which allows it to reduce the list of their possible part of speech categories to a very restricted number of highly probable hypotheses, given the corresponding possible parts of speech of neighboring words. Syntactic-prosodic rule defines the search space and finds the text structure (i.e., its organization into clause and phrase-like constituents) which more closely relates to its expected prosodic realization [1][8].

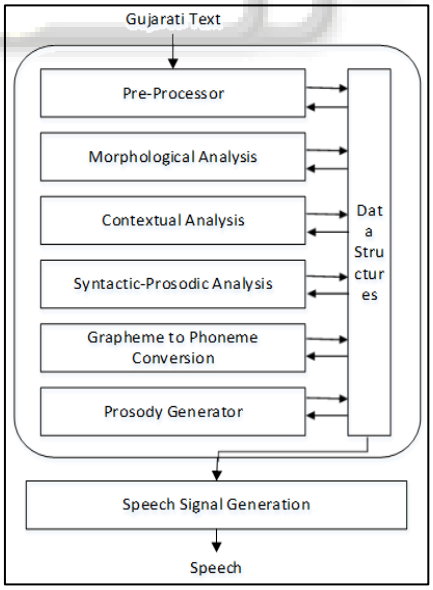

Fig. 1: TTS Framework for Gujarati Language

A Letter-To-Sound module is employed for transcription of incoming text. This transcription is on the far side a dictionary look- up operation. This can be as results of most words have completely different phonetic transcriptions depending on context. Also, pronunciation dictionaries don't account for morphological variations in words. Additionally, pronunciations of words in sentences dissent from pronunciation of these same words once they square measure isolated. Moreover, not all words are present in a phonetic dictionary. As a result, phonetization is dictionary-based or rule-based (based on a set of letter-to-sound rules). [8][16]

Prosody refers to rhythm, stress and intonation of speech. Prosody directs focus to specific components of a sentence, like stress arranged on a selected linguistic unit, therefore attributing special importance or distinction to it a part of the sentence. Speech options conjointly facilitate to phase sentences into chunks comprising of teams of words and syllables and conjointly to spot the relationships between such chunks. The prosody generator is liable for prosody generation. Generation of a natural-sounding prosody is one in all the largest challenges round-faced within the style of Text-To-Speech systems. [7][9]

The concatenation technique is one of the important aspects in speech synthesis that produces natural synthetic speech. However, concatenation synthesizers are usually limited to one speaker or one. In concatenation synthesis, recording of a human speaker is concatenated in order to generate the synthetic speech. To select correct unit length, there is usually a trade off between longer and shorter units. With a longer unit, the synthesizer can generate high naturalness, fewer concatenation points and good control of co-articulation. However, the required memory capacity increases. With shorter units, the synthesizer requires less memory capacity but the collection of speech samples become more complicated and difficult. In a general system, the units of sounds have various lengths, ordering from long to short, word, syllable, demisyllables, triphones, diphones, and phonemes. Words and syllables may be considered the most natural units for synthesis. Diphone are arguably the most commonly used units in general speech synthesis because they are used for the linguistic presentation of speech [6] [8]. Phoneme and diphone are chosen as a unit for concatenation.

V. INPUT & OUTPUT

This system takes input Gujarati text and produces sound equivalent to the Gujarati text. It facilitates making audio file of the speech for the given Gujarati text.

VI. BENEFITS AND APPLICATION

TTS applications are well known as assistive aides for people who have reading challenges, or visual impairment. There may be an obvious need that leads to the introduction of TTS into an individual's lifestyle. However, if this is not the case, you may consider many other valuable uses for this application that range from reducing eye strain from reading (digital or paper formats), reducing paper use due to printing digital text, or promoting listening skills [2].

A. Applications of TTS

- News Reader
- Talking books and toys
- Education and Sports
- Aid to Handicapped
- Telecommunication and Multimedia
- Voice enabled e-mail
- Man-Machine Communication

VII. CONCLUSIONS

The described speech synthesis system is initial TTS system for the Gujarati language. Improvement of intelligibility and naturalness depend on proper works in different context. The Gujarati TTS system is extremely a lot of vital for the visually impaired individuals who knows Gujarati. It is also typically necessities for normal people that wish to read on-line newspaper, articles and journals. So our system will not solely facilitate these visually impaired individuals but also facilitate mass people.

REFERENCES

[1] Thakur, Benoy Kumar, Bhusan Chettri, and Krishna Bikram Shah. "Current Trends, Frameworks and Techniques Used in Speech Synthesis–A Survey." International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, 2012.
[2] Shruthi, Gupta. "Comparative study of text to speech system for Indian language." International Journal of Advances in Computing and Information Technology (2012).
[3] Narupiyakul, Lalita, et al. "A stochastic knowledge-based Thai text-to-speech system." Mathematical and computer modelling 42.1 (2005): 1-16.
[4] Gupta, Shruti, and Parteek Kumar. "Hindi Text To Speech System." (2012).
[5] Sangeetha, J., et al. "Text to speech synthesis system for tamil." International Journal of Emerging Technology and Advanced Engineering 3 (2013).
[6] Rao, M. Nageshwara, et al. "Text-to-Speech Synthesis using syllable-like units." Proceedings of National Conference on Communications, IIT, India. 2005.
[7] Taylor, Paul. Text-to-speech synthesis. Cambridge University Press, 2009.
[8] Romsdorfer, Harald. Polyglot text-to-speech synthesis. Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 18210, 2009, 2009.
[9] Alam, Firoj. Kotha: the first to speech synthesis for Bangla language. Diss. School of Engineering and Computer Science (SECS), BRAC University, 2006.
[10] Hariharan, R. [Online]. Available: http://dhvani.sourceforge.net/.
[11] Mukhopadhyay, A., Chakraborty, S., Choudhury, M., Lahiri, A., Dey, S., Basu, A., 2006. "Shruti- an Embedded Text-to-speech System for Indian Languages", IEEE Proceedings on Software Engineering,153, Issue 2, pp. 75–79.
[12] Ramakrishnan, A.G., Bali, K., Talukdar, P. P., Krishna, N.S., 2004. Tools for the Development of a Hindi Speech Synthesis System, in 5th ISCA Speech Synthesis Workshop, Pittsburgh, pp. 109-114.
[13] Jain, H [Online]. Available: http://www.cse.iitb.ac.in/vani/.
[14] MBROLA, "Project homepage", 1998. Online: http://tcts.fpms.ac.be/synthesis/mbrola.html/
[15] Black A, Taylor P, Caley R (2001) The Festival speech synthesis system: system documentation. University of Edinburgh. Online: http://www.cstr.ed.ac.uk/projects/festival/
[16] Onaolapo, J. O., et al. "A Simplified Overview of Text-To-Speech Synthesis." Proceedings of the World Congress on Engineering. Vol. 1. 2014.

