Architecture and Implementation of Punjabi Text to Speech System Using Transcriptions Concept

Er. Sheilly Padda[1], Ms. Rupinderdeep Kaur[2], Er. Nidhi[3]

[1]Pursuing M.Tech, Department of Computer Science, Swami Vivekanand Institute of Engineering & Technology, Banur, Punjab
[2]Lecturer, Department of Computer Science, Thapar University, Patiala, Punjab
[3]Lecturer, Department of Computer Science, Swami Vivekanand Institute of Engineering & Technology, Banur, Punjab

Abstract—This paper presents the approach towards converting text to speech using new methodology. The text to speech conversion system enables user to enter text in Punjabi and as output it gets sound. The paper presents the steps followed for converting text to speech for Punjabi (gurmukhi) language and the algorithm used for it. The focus of this paper is based on the tokenisation process and the orthographic representation of the text that shows the mapping of letter to sound using the description of language's phonetics. Here the main focus is on the text to IPA transcription concept. It is in fact, a system that translates text to IPA transcription which is the primary stage for text to speech conversion. The whole procedure for converting text to speech involves a great deal of time as it's not an easy task and requires efforts.

Keywords—IPA, Orthographic representation, phonetic, tokenisation, transcription.

I. INTRODUCTION

In today's life everyone wants that the present computer systems should behave like humans and proved to be user friendly. Even many of the great research scholars had dreamed of involving the machines in every facet of human life. With the growth of the power of computing machines their applications in modern day to day life is also rising [2]. This paper depicts one of the modern technologies that is the technology of text to speech convertor. This paper presents the frame work of text to speech conversion methodology. The text to speech system has many applications that are useful in our life such as for video conferencing, helping people with reading difficulty, for visual aids process, Elementary and Applied Research, Telecommunications and Multimedia, Educational State, Screen reader programs for the visually impaired [3].

In this paper we have presented the linguistic features of Punjabi text stating its phonetic representation. Through the knowledge of language's phonological orthographic we can develop a more promising text to speech conversion [4]. The first step in our text to speech conversion task is having the text tokenisation and to determine the token to sound concept.

II. OVERVIEW OF TEXT TO SPEECH SYSTEM

The text to speech conversion is carried out with the processing of text normalisation (text preprocessing), text analysis, syllabification, prosody generation and finally the speech production [10]. The process of text analysis, text normalisation and prosody generation all come under natural language processing while the speech generation is a part of digital processing system. A general view of the system is depicted in the Fig.1
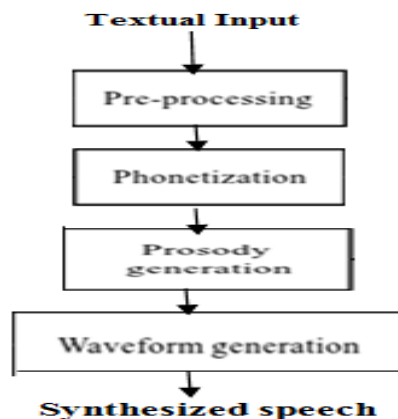

Fig 1: Overview of TTS

The text to speech conversion involves a number of steps that are to be followed. The first step includes Tokenisation/Text preprocessing where the inputted text is broken down into smallest unit i.e. tokens. Tokenisation also includes the expansion of abbreviations.

The second step involves Syllabification process includes the combination of the small units to represent them as phonemes that identify the sound associated with each unit and we also called them as syllables.

In Punjabi seven types of syllables are recognized [8].These syllable types are: V, VC, CV, VCC, CVC, CCVC and CVCC; where V and C represent vowel and consonant respectively. Out of these seven types, occurrence of last two syllable types having sound clusters, is very rare in Punjabi.

The third step is Speech Synthesizer which involves the generation of speech as output. There are various techniques that can be used for synthesizing speech [7]. The speech can be produced by concatenating different pieces of recorded speech from the database. Also, speech can be produced from concatenation of phonemes, diphones, unit selection techniques and by Formant synthesis where the synthesized speech output is created using additive synthesis and an acoustic model (physical modelling synthesis).Parameters such as fundamental frequency, voicing, and noise levels are varied over time to create a waveform of artificial speech.
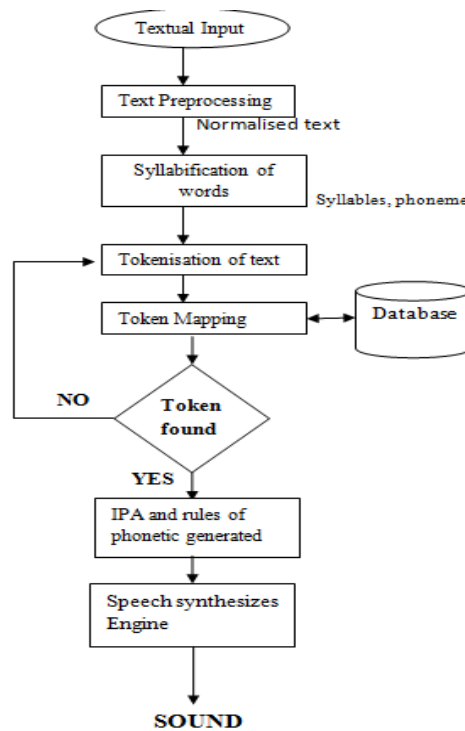
III.   ARCHITECTURE AND   IMPLEMENTATION

Fig-2: **Architecture of TTS**

The entire process of converting text to speech includes analysis and synthesis [6]. Although analysis part is similar for every techniques used for text to speech conversion, the difference is only in the synthesis part. The analysis part deals with determining the orthographic description of a particular language.

The implementation of text to speech conversion goes through various steps as depicted in Fig-2. The first step identifies the text for the processing. The text normalisation is carried out. After the text has been normalised, the syllabification process is carried out where the input text is broken down into smallest unit called syllables.

A)   Text to IPA transcription

The phonetization is done which is the process of representing vocal sounds by signs and written characters. There are number of phonetic sounds that represent a letter of particular language. These phonetic sounds along with their written representation can be combined to generate sound using various speech synthesis tools.

The orthographic representation of text enables to represent the syllables with phonetic representation [1]. The orthographic set for a language consists of letter to sound rules. These rules define the function mapping of sequence of

letter to sound segments.Specifically, the orthographic representation determined for a language, specifies the initial stages for text to speech conversion [5].

The phonetic notation can be represented for Punjabi language using an IPA chart (International Phonetic Alphabet).The IPA chart defines is an alphabetic system of phonetic notation based on some Latin alphabets [9]. They can be used for representation of sounds of spoken language.

The principal vowels for Punjabi can be depicted from Fig-3. The principal vowels are symmetrically distributed on a standard vowel chart: three front vowels, two central vowels, and three back vowels. The three back vowels are rounded[9].
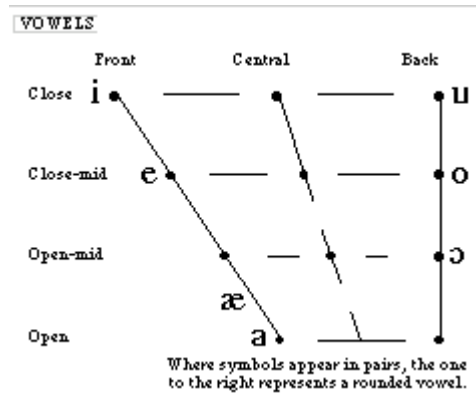

Fig-3: **Vowel chart**

B) **Consonants IPA chart**

Representations of consonant sounds outside of the core set are created by adding diacritics to letters with similar sound values.Fig-4 represent the consonants and their IPA symbols[9].

Fig-4: **IPA chart for consonants**

The methodology followed here translates the text to IPA transcription using some defined rules. These rules are used for determining the letter orthographic features. These orthographic features can further be used by the speech synthesis tool to generate sound.

10

IV.   CONCLUSION

In this paper, we have described how the phonological attributes of Punjabi orthography may be used to develop a text-to-speech system. The phonetic nature of the language was surveyed, and the set of rules for deriving these sounds using the orthography were made explicit and carefully deconstructed. Although the work done is not complete as it has been done until determining the linguistic features of Punjabi text while the final signal processing aspect of the task are un explored. In the near future we would love to complete the text-to-speech system, the text to speech system that is able to be useful to the people in need.

REFERENCES

[1].   Prof. Ian Maddieson,"Applied Phonetics: Portuguese Text-to-Speech", University of California, Berkeley Linguistics 110: May 16, 2003.

[2].   Ruvan Weerasinghe, Asanka Wasala, Viraj Welgama and Kumudu Gamage, "Festival-si: A Sinhala Text-to-Speech System", Language Technology Research Laboratory, University of Colombo School of Computing, Colombo, Sri Lanka.

[3].   A. Chauhan, Vineet Chauhan, Surendra P. Singh, Ajay K. Tomar and Himanshu Chauhan , "A Text to Speech System for Hindi using English Language" , IJCST Vol. 2, Issue 3, September 2011

[4].   Kalika Bali, Partha Pratim Talukdar, N. Sridhar Krishna and A.G. Ramakrishnan, "TOOLS FOR THE DEVELOPMENT OF A HINDI SPEECH SYNTHESIS SYSTEM" , 5th ISCA Speech Synthesis Workshop – Pittsburgh,2004.

[5].   M. H. Mateus, E. d'Andrade, "The Phonology of Portuguese: The phonology of the world's languages", Oxford University Press Inc., New York, 2000.

[6].   Dhvani-TTS System for Indian Languages,(http://dhvani.sourceforge.net), 2001.

[7].   *Vincent Pollet*, *Andrew Breen,* "Synthesis by Generation and Concatenation of Multiform Segments", Text-To-Speech Research, Nuance Communications, Merelbeke, Belgium, 2008.

[8].   Parminder Singh and Gurpreet Singh Lehal, "Text-To-Speech Synthesis System for Punjabi Language".

[9].   International Phonetic Association (IPA), Handbook.

[10].   Simon King, "An introduction to statistical parametric speech synthesis", Indian Academy of Sciences, October 2011, pp. 837–852.