International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 04 | Apr-2016 www.irjet.net p-ISSN: 2395-0072

An Improved System for Converting Text into Speech for Punjabi Language using eSpeak

Ramanpreet Kaur[1], Dharamveer Sharma[2]

[1]*Dept. of Computer Science, Punjabi University, Punjab, India*

[2] *Associate Professor, Dept. of Computer Science, Punjabi University, Punjab, India*

----------***----------

Abstract - *A large number of text-to-speech (tts) softwares are available for speech synthesis. But it is a challenging task to provide a single generalized system for many languages. eSpeak provides support for several languages including Punjabi. It is an open source application that provides rules and phoneme files for more than 30 languages. This paper discusses some improvements in this formant based text to speech synthesis system for Punjabi text input. After analysis of eSpeak for Punjabi input some faults are identified and corrected by using eSpeakedit.*

Key Words: *text-to-speech,* tts, Punjabi Language, eSpeak, eSpeakedit, Phoneme.

1. INTRODUCTION

Text- to -Speech system is basically used to generate sound waves as output from the given text input. This is also known as speech synthesis, and a complement of speech recognition (Speech to Text). Text-to-speech systems have an enormous range of applications. They are widely used by people having speech or reading disabilities. Every speech synthesis system has two main features naturalness and intelligibility. The intelligibility of tts can be measured by using

the units of speech such as phonemes, syllables, words, phrases etc. It is used to define the understandability of produced output speech. Naturalness also known as pleasantness. It presents how much the output sound look like a real human voice. According to the type of synthesis technique the Text-to-speech synthesizers can be classified into three types:

1.1 Articulation based synthesis

This technique generates speech sounds by providing mathematical modeling of human speech organs such as lips, tongue, palate and jaws. It is also classified as a rule based speech synthesis method. Articulation based speech synthesis is a complicated technique but it is the best method to provide high quality synthetic speech. First articulatory synthesizer was introduced in 1958 by George Rosen at the Massachusetts Institute of Technology [1].

1.2 Concatenation based synthesis

It uses recorded speech feature segments (word, phoneme, sub-phoneme, etc.) and generates the speech by concatenating appropriate units from the speech database based on the phonemic units in the text. The speech generated by this technique inherently possesses natural quality; however, the quality depends on the size of the speech database. A concatenation based speech synthesis system with a small database may create audio

discontinuity due to the unavailability of the speech unit in its database [2].The concatenation based speech synthesis can be divided into three sub types; Unit selection synthesis, Diphone synthesis and Domain specific synthesis.

1.3 Formant based synthesis

The rule based speech synthesis technique does not require a pre-recorded speech database; instead, it produces speech by generating signals with rules which mimics the formant structure and other spectral properties of natural speech as closely as possible. The output speech in formant synthesis is created using additive synthesis and an acoustic model which uses parameters like, fundamental frequency, voicing, and noise levels that varied over time and creates a waveform of artificial speech. However, the rule based techniques produces artificial or robotic speech. Also, specifying the rules for timing of the source and the dynamic values of all filter parameters is relatively difficult for even simple words [2]. Some other methods for speech synthesis are also available, Hybrid synthesis, HMM (Hidden Markov Model) based synthesis, Sinusoidal Model based Synthesis and Linear predictive methods [3].

Next discussion in this paper is based on a formant based text-to-speech system. Speech waves are created by defining various features of them. Section 2 represents the details about the platform used. A brief introduction of eSpeak software is given. The section 3 discusses about the problems that are identified after analysis. These problems are classified into groups. The solution for each group is presented in this section. The last section of the paper presents the conclusion.

2 eSpeak SOFTWARE

eSpeak is a compact open source software speech synthesizer for English and other languages, for Linux and Windows. eSpeak uses a "formant synthesis" method. This allows many languages to be provided in a small size. The speech is clear, and can be used at high speeds, but is not as natural or smooth as larger synthesizers which are based on human speech recordings.

The eSpeak speech synthesizer supports several languages; however in many cases these are initial drafts and need more work to improve them. eSpeak does text to speech synthesis for the following languages, some better than others.

Afrikaans, Albanian, Aragonese, Armenian, Bulgarian, Cantonese, Catalan, Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Farsi, Finnish, French, Georgian, German, Greek, Hindi, Hungarian, Icelandic, Indonesian, Irish, Italian, Kannada, Kurdish, Latvian, Lithuanian, Lojban, Macedonian, Malaysian, Malayalam, Mandarin, Nepalese, Norwegian, Polish, Portuguese, Punjabi, Romanian, Russian, Serbian, Slovak, Spanish, Swahili, Swedish, Tamil, Turkish, Vietnamese, Welsh [4].

Punjabi language was added in espeak version 1.47.01 on 18.03.2013. Many languages were improved after their initial addition in this software. Punjabi language also needs improvements; an analysis is performed as being a native speaker of it. Punjabi language is being spoken by about 104 million peoples in India, Pakistan and other countries with Punjabi migrants. The language is being written in Gurmukhi script in Indian Punjab, whereas in Shahmukhi script in Pakistani Punjab [5].

3 Analysis and Improvements

During analysis of eSpeak Text-to-Speech system for Punjabi text some problems were noticed. These problems are divided into following groups.

1. Incorrect pronunciation of some words.
2. No sound output is produced for "speak character name" option for most of the characters.
3. It produces speech output in English language for some symbols when Punjabi voice option is selected.
4. Pronunciation for some symbols is not available in Punjabi.
5. Incorrect pronunciation of some characters when they come with some symbols.

These are the problems that were identified and corrected. In order to make improvements, different files such as ph_punjabi, pa_rules, pa_list are required. ph_punjabi is a phoneme definition file. This contains phoneme definitions for the vowels and consonants for Punjabi language. pa_rules contains the spelling-to-phoneme translation rules. pa_list contains pronunciations for numbers, letter and symbol names, and words with exceptional pronunciations. For improvements, these files are needed to be updated. The results after improvements are explained below.

Group 1

There is a problem of incorrect pronunciation of some words for Punjabi text for example when we input word "ਤਿੰਨ" and digit 3 then it generate different speech output for both. It searches pa_list file and find the different rules for 3 and word "ਤਿੰਨ". For digit 3 is t'I~n and for corresponding Punjabi word "ਤਿੰਨ" is t: 'Vnn .This is solved by making changes in pa_rules file. After this it produces correct output for both.

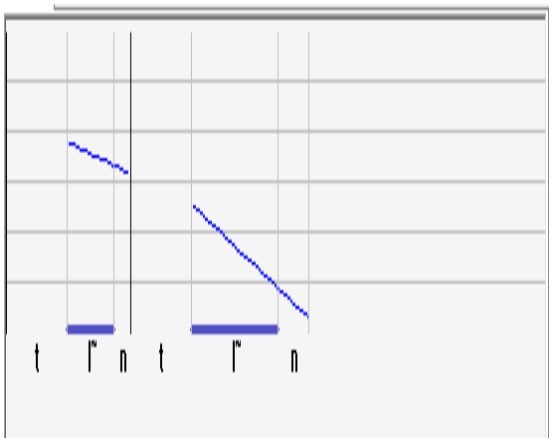

Fig-1: Prosody for word ਤਿੰਨ and 3

Similarly there was a problem with words with sihari and tippi as "ਮਿੰਨਤ" ,"ਸਿੰਨ "etc is solved.

Group 2

There is a speak character name option in eSpeakedit. But after analysis it is found that it does not speak the name of all Punjabi characters as it speak those for English language. So this problem is solved by adding rules for each character name in pa_list file. After updating the pa_list file is compiled using eSpeakedit.

Table -1: Character names in pa_list file

ੳ	ਅ	ੲ	ਸ	ਹ	
'Ur.a_!	'Er.a_!	'Ir.i_!	s'Vsa_!	h'aha_!	
ਕ	ਖ	ਗ	ਘ	ਙ	
k'Vka_!	k#'Vk#a_!	g'Vga_!	g#'Vga_!	n'Vna_!	
ਚ	ਛ	ਜ	ਝ	ਞ	
c'Vca_!	c#'Vc#a_!	J'VJa_!	J#'VJa_!	n^@'i;a_!	
ਟ	ਠ	ਡ	ਢ	ਣ	
t.'E~ka_!	t.#'Vt.#:a_!	d.'Vd.:a_!	d.#'Vd.:a_!	n.'an.a_!	
ਤ	ਥ	ਦ	ਧ	ਨ	
t'Vt:a_!	t#'Vt#:a_!	d'Vd:a_!	t'V+d:a_!	n'Vnna_!	
ਪ	ਫ	ਬ	ਭ	ਮ	
p'@p:a_!	f'Vffa_!	b'Vb:a_!	p'V+b:a_!	m'Vmma_!	
ਯ	ਰ	ਲ	ਵ	ੜ	
j@'i;a_!	r'a*a_!	l'Vlla_!	v'ava_!	r.'ar.a_!	
ਸ਼	ਖ਼	ਗ਼	ਜ਼	ਫ਼	ਲ਼
S'VSSa_!	x'Vxa_!	Q'VQa_!	z'Vza_!	f'Vfa_!	l.'Vl.a_!

Group 3

It does not produce sound for some symbols in Punjabi. For example if ( ) comes in an expression than it produces output in English speech. For symbol (it searches in en_list file and find a rule lEftpa#rEn and for) raItpa#rEn But now new rules for (and) are added and file is compiled using eSpeakedit. After this it produces output in Punjabi speech. As

(k#'Vb:i b@*'Ek@t.

) s'VJ:i b@*'Ek@t.

Similarly for other symbols like *, /, etc rules are added in rule file. For "ੴ "it was not producing correct speech output before correction.

Group 4

For Punjabi voice, if input symbol is - it does not find any rule so no voice is produced. Now during improvement, rule is added for - symbol. It produces output in Punjabi language.

For example input expression is 3-4, then the output will be

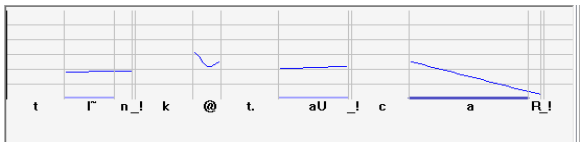

Fig -2: Prosody for input 3-4

Group 5

It produces incorrect speech for some words for e.g in words ਸਤਿ ਮਤਿ ,.Words where 'sihari' comes before ' ਤ .the output sound was not correct ,'

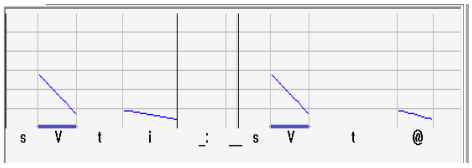

Fig -4: Prosody for correct result for input ਸਤੀ and ਸਤਿ

3. CONCLUSIONS

It is a formant based speech synthesis system and a universal system to support several languages. So it is a difficult task to produce a natural speech output. As some of the problems are identified for Punjabi language but there may exist more. These can be identified and corrected. The improvements are quite significant so they can be incorporated within the software with a new version, which has been one of the goals of this paper.

REFERENCES

[1] Pardeep Gera, "Text-to-speech synthesis for Punjabi language", M.Tech, Thesis, Thapar Institute of Engineering and Technology, Patiala, pp 10-20, 2006.

[2] eSpeak, Available: http://espeak.sourceforge.net, Accessed: 15.01.2016

[3] Soumya Priyadarsini Panda and Ajit Kumar Nayak, "An efficient model for text-to-speech synthesis in Indian languages", Int J Speech Technol , pp 305–315, February 2015

[4] Wikipedia, Available: https://en.wikipedia.org/wiki/Speech_synthesis, Accessed: 16.02.2016

[5] Parminder Singh and Gurpreet Singh Lehal, "Text-to-speech synthesis system for Punjabi language", Punjabi University, Patiala, 2005

[6] Othman. O. Khalifa, M.Z. Obaid, A.W. Naji and Jamal I. Daoud, "A rule-based Arabic text-to-speech system based on hybrid synthesis technique", Australian Journal of Basic and Applied Sciences, pp 342-354, 2011

[7] Mentor Hamiti and Ramiz Kastrati, "Adapting eSpeak for converting text into speech in Albanian", International Journal of Computer Science Issues, Vol. 11, Issue 4, No 2, pp 21-26, July 2014

[8] Paul Taylor , "Text-to-Speech-System", University of Cambridge, 1st Edition