A Review on Multilingual Text to Speech Synthesis by Syllabifying the Words of Devanagari and Roman

Mr. S. B. Chaudhari
Research Scholar
Computer Engineering
JJT University, Rajasthan.
sbchaudhari@yahoo.com

Dr. S. K. Yadav
Internal Guide
Computer Engineering
JJT University, Rajasthan
drskyadav@hotmail.com

Abstract: Speech synthesis is process of spoken language as an input text and converted into speech waveforms. This paper describes the text to speech system for Devanagari scripted language and Roman Language. There are many earliest TTS systems are available but for Devanagari and Roman scripts are not available.

Keywords: Unicode, Scripting Language, Optimization and Validation.

_____*****_____

Introduction: General concept of Text to Speech synthesis technology is that it takes the text as input and it converts it into speech. In everyday life for communication we use speech. There are many people in India that they can't read but they can hear the sound. For them text to speech system is very useful, Such TTS system is very useful for learning new languages. Now days there are many applications in which we require such type of tool. Presently there are many systems are available in the market, some of them are free to use and some of them are not free to use. These systems are having their pros and cons. Majority of system is available for single language like English or Hindi or other language. Basically the Multilingual Text to speech synthesis by syllabifying the words of Devanagari is a proposed system which will take input as typed text of Marathi, English, Hindi, Sanskrit and all languages that use Devanagari script and that text will be converted to speech as output. Text processing and generation of speech are the main steps involved in this system. Designing such type of system it is necessary to consider the input and the expected output.

There are many TTS systems are available in the market that reads the text of a particular language. Here it is necessary to design a system that will read Indian languages and English. In globalization the English is the language of the world, Hindi is the official language of India and Marathi is the language of Maharashtra state. Now a day's people use combination of different languages like Marathi, Hindi and English .To type the text of these languages in most of the available text to speech system different local fonts are used. In the proposed system we are going to use i18n concept for typing Devanagari and Roman script using UTF-8, Unicode. So input will be like "**What day is it today, आज Sunday आहे. आज हम क्या करेंगे. आपण आज university ला visit देवू**". So such type of

input will be converted into speech. To use combination such languages in written and spoken there is need of transliteration and syllabification. There are many applications in which speech enabled interface is required. But there no such system available that reads the multilingual text with Devanagari and Roman script.

There are mainly three approaches are used in speech synthesis Concatenative speech synthesis, Formant synthesis and Articulatory synthesis. Each approach has its own strengths and weakness. Comparatively Concatenative speech synthesis is widely used. Presently unlimited words can be generated by concatenating the phonemes, but we do not get the natural quality speech using this method. When we use phoneme unit for concatenation the concatenation cost increases, also speed and we do not get quality of speech. Instead of selecting unit as phoneme if we select Syllabified units of words (all-phones) as segment we can generate unlimited, natural quality speech. When we use one writing system like we can use ASCII character, but this code table is not sufficient for multiple languages. The proposed system will work with high accuracy and has a great potential to be used for variety of applications.

Need of TTS: In the application field of Text to Speech is expanding fast while the quality and use of TTS system is also increasing steadily. These synthesis systems are more affordable for common customer and also more suitable for everyday use.

Identification of Problems:

In the present Text to Speech System different journals, Conferences and Books have been considered from year 1990 to 2014

Problem Identification:
At different level different problems are identified.

- Different systems have used different local fonts due to this problem arises when font is not available. Here we should use unique encoding.
- Algorithm is not available for normalization of multiple languages. So language detection problem arises. Different languages have separate algorithm.
- Di-phone Unit selection approach increases the concatenation cost and degrade the speech quality.
- When English TTS converts Text to speech systems it uses their own speech Engine. Hindi TTS uses the different Speech Engine. Here multiple languages have multiple speech databases so the cost of the TTS system increases.

Objective of the present research work: Main goal is to design a system that would read the text written using combination of Roman and Devanagari script. To design a subsystem for text processing that will convert of Roman script to Devanagari using *transliteration.*

To build *unique syllable speech segment database* for all languages by using speech Digital Signal processing and with the help of speech tool.

- To design a rule-based algorithm for *word syllabification* of the words of all languages which are using Devanagari and Roman script.
- To design a parallel segment selection algorithm that will search syllable speech segments with position start, middle or end syllable in word and *concatenate* these speech segments which produces quality speech.
- The *concatenation synthesis approach* will be the best approach for creation of unlimited quality speech. There is need of designing a player that will play that audio file.

Methodology:

At different level different methodologies can be used to obtain natural quality and unlimited speech.

- Unicode encoding system can be used to solve the problem of local font.
- For normalization of multiple languages rule-based transliteration and translation is the best solution.
- Poly-syllables decreases the concatenation cost and it increases speech quality as well as it covers the words of multiple languages.
- Transliteration is the best solution to convert the source language script to target language. Normalization of multiple language text is possible. Unique speech database is the best solution for multiple languages. It decreases the total cost of the system.

- Using Digital signal processing prosody can be manipulated. So that features like pitch, Loudness etc. Speech smoothing at concatenation points removes unwanted glitches that increases intelligibility of speech. Optimal coupling is the best solution for segment smoothing.
- Software simulation: Text to speech system requires an editor for typing Roman and Devanagari script. To design a new system that will generate natural speech with unlimited performance for Indian languages like Marathi, Hindi and English. Need to implement using C/ C++, Java, MATLAB. Our aim is to build a system which will take abnormal text file encoded in Unicode UTF-8 which is used for all languages.

Scope and Significance:

- System will work with high efficiency it will cover text of Devanagari and Roman writing system. It can be extended for other writing system.

Optimization and Validation:

- When we consider larger unit size of segments of text such as words, phrases, sentences we get good quality of speech and its concatenation cost is less, it is faster, but we can't cover the whole language because it is very difficult to create a speech database., But it is not possible to create words, phrases and sentences speech units. So this solution is not feasible.
- When consider unit selection syllable concatenation cost is more but we can cover all words. So this solution is optimized solution. The optimized solution is to select unit as syllables rule based method. The speech quality will be good and cost of the speech is less as compare to other unit selection methods.

Results:

When we consider larger unit size of segments of text such as words, phrases, sentences we get good quality of speech and its concatenation cost is less, it is faster, but we can't cover the whole language because it is very difficult to create a words, phrases and sentences speech unit database., So this solution is not feasible.

When considering unit selection syllable concatenation cost is more but it covers all words of different languages. So this solution is feasible and optimized solution. Using this proposed system has achieved good quality speech and its covers all three languages.
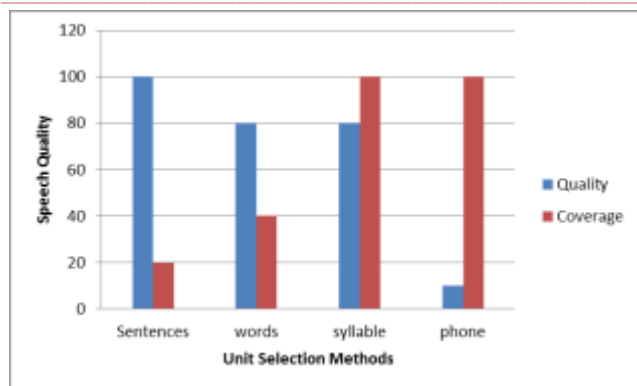
Fig: Comparison of Unit selection methods

Conclusion:

The proposed system is tested by taking different 5 peoples reviews that knows Marathi, Hindi and English language.

In this system a set of words directly recorded and a set of words generated by system are played and listen by 5 listeners. Desktop speakers are used to listen these set of words. The result in table shows, each row in this table indicates the evaluation result of listeners. Result indicates listeners are in favour of syllable based proposed system; we cannot cover all languages by creating speech segment database of larger unit selection. But by using syllable unit selection speech quality is good and we can cover all languages that are uses Devanagari and roman script.

Sr No	Preferences of Word Directly Recorded	Preferences of the Syllable by Proposed System
1	8	7
2	9	9
3	8	7
4	8	8
5	10	9
	86 %	82 %

Table: Preferences of Word and Syllables

Thus the proposed system is working with high accuracy and has a great potential to be used for variety of applications.

References:

[1] Amanpreet kaur (2008-09) "automatic syllable segmentation for indian languages" rimt-iet, mandi gobindgarh, bgiet, sangrur-2008-09.

[2] Bhuvana Rarasimhan, (2004) " schwa- deletion in hindi text -to-speech synthesis " international journal of speech technology 7, 319–333, 2004 kluwer academic publishers

[3] Catalin Ungurean, Hybrid syllabification and letter to phone conversation for tts synthesis. U.p.b. sci. Bull., series c, vol. 73, issue 3, 2011 issn 1454-234x.

[4] Diemo schwarz(2000) "a system for data - driven concatenative sound synthesis" proceedings of the cost g-6 conference on digital audio effects (dafx-00), verona, italy, december 7-9,2000

[5] D.T. Chappell (2002) "spectral smoothing for speech segment concatenation" speech communication, volume 36, issues 3-4, march 2002, pages 343-373.

[6] E.veera Raghavendray, (2008) "a multilingual screen reader in indian languages" international institute of information technology - hyderabad, india. Language technologies institute, carnegie mellon university, pittsburgh, pa 15213, usa.

[7] Hafiz Musa, (2010) "syllabification algorithm based on syllable rules matching for malay language" recent researches in applied computer and applied computational science 2010.

[8] Kalyan d. Bamane (2011) "marathi speech synthesized using unit selection algorithm" computer engineering and intelligent systems issn 2222-1719 (paper) issn 2222-2863 (online). 2011.

[9] Massimo Moneglia, (2006)"challenging the perceptual relevance of prosodic breaks in multilingual spontaneous speech corpora: c-oral-brasil / c-oral-rom" lablita, university of florence, italy. 2006.

[10] Mohamed ali kammoun (2009) "unit selection algorithm using bi-grams model for corpus-based speech synthesis" international journal of information and communication engineering 5:2 2009.

[11] Manik Dhore (2012) Hindi to English Machine Transliteration of Named Entities using Conditional Random Fields International Journal of Computer Applications (0975 – 8887) Volume 48– No.23, June 2012

[12] G. V. Garje (2013) Machine Translation In India International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, October2013

[13] G. V. Garje (2013) Transmuter: An Approach to Rule based English to Marathi Machine Translation. International Journal of Computer Applications (0975–8887) Volume 9 8–No.21, July 2014