SSML Extensions for Text to Speech Systems in Indian languages

Kishore Prahallad[*] and Nixon Patel[+]
[+]Bhrigus Inc. Hyderabad, India
[*]IIIT Hyderabad, India

A text to speech system often requires simple information such as language of the input text; voice (male/female) to be used, pronunciation of a telephone number as isolated digits etc. A raw input text could be embedded with such information using XML like tags often referred to as Speech Synthesis Markup Language (SSML) which aims to produce a better content by a TTS in various contexts. In this positional paper, we argue and discuss some of the possible SSML extensions keeping in the view of Indian language scripts and the corresponding TTS systems.

1. INTRODUCTION

Bhrigus Inc. in collaboration with IIIT Hyderabad is actively involved in developing TTS and ASR for Indian languages, and is currently developing unit selection voice for Telugu. The goal is to build high quality voices and speech recognition for many of the Indian language and interface them with computer-telephony applications. Some of these applications include verticals such as entertainment, health care, financial in the context of India. In this paper, we describe the nature of the Indian language scripts, describe their convergence and divergence of these scripts and discuss our proposal where we feel the requirements of some more SSML elements to improve the rendering of Indian language scripts.

2. INDIAN LANGUAGE SCRIPTS

The scripts of Indian language have originated from the ancient Brahmi script. The basic units of the writing system are referred to as “Aksharas”. The following are the properties of Aksharas.

- An Akshara is an orthographic representation of a speech sound in an Indian language.
- Aksharas are syllabic in nature
- The typical forms of Akshara are V, CV, CCV and CCCV, thus have a generalized form of C*V
- An Akshara always ends with a vowel.
- White space is used as word boundary thus separating Aksharas present in two successive words.
- The scripts are written from left to right.

- Roman digits (0...9) are used as numerals. Some of the languages have their own numeric symbols but are rarely used.
- English Punctuations marks such as comma, full stops are mostly used in writing... Languages such as Hindi have a set of their own punctuation marks which are often used.

2.1 Convergence and Divergence of Indian Language Scripts

India is a multi-lingual nation with 17 languages recognized as official languages. These languages are: 1) Assamese 2) Tamil 3) Malayalam 4) Gujarati 5) Telugu 6) Oriya 7) Urdu 8) Bengali 9) Sanskrit 10) Kashmiri 11) Sindhi 12) Punjabi 13) Konkani 14) Marathi 15) Manipuri 16) Kannada and 17) Nepali. Except Urdu and English, all of the remaining official languages have a common phonetic base, i.e., they share a common set of speech sounds across all of these languages, and the characters of these language have almost one to one correspondence with the sounds of the language.

While all of these languages share a common phonetic base, some of the languages such as Hindi, Marathi and Nepali also share a common script known as Devanagari. But languages such as Telugu, Kannada and Tamil have their own script. Each of these languages has their own scripts. The property that makes these languages separate could be attributed to Phonotactics that takes place in each of these languages rather than the scripts and the speech sounds. Phonotactics deals with the permissible combinations of phones that can co-occur in a language.

Many of these Indian languages have a great influence of English. Colloquial English words such as “please, bank, sorry, morning, evening,” are freely used in Indian languages along with native words. Each of these languages also has its own dialects.

3 PROPOSED EXTENSIONS

3.1 Syllable Element <syllable>

Indian language characters are syllabic in nature, and hence when a word has to be spoken as character by character, they have to be rendered as syllable by syllable to make it more meaningful to the native speakers. Thus it is essential to have syllable element for rendering of Indian language scripts.

For example if the Telugu word “naatoo” has to be spoken character by character, and the use of phoneme tag would split it as below.

```
<phoneme alphabet="itrans-3" ph="n aa t oo"> naatoo </phoneme>
```

However, for a native Telugu speaker there is no sound called “n” exists. For him/her the sound “n” always exists with a vowel which is a syllable (note characters in Indian languages are syllables). Hence it makes sense to have a syllable tag which would split the word “naatoo” as follows and which is more meaningful for the native speaker.

```
<syllable alphabet="itrans-3" syl="naa too"> naatoo </syallable>
```

Itrans-3 is a transliteration scheme which is used to write Indian language scripts using QWERTY keyboard. Itrans-3 is a phonetic alphabet proposed by IISc Bangalore and Carnegie Mellon University and is widely adapted among universities and industrial labs.

3.2 Loan-Word Element <alien>

Indian languages have a heavy influence of non-Indian languages such as English in their scripts and one of the informal experiments suggested that the 33% of the TTS errors were purely due to the system trying to speak loan words from English. It is important to note that these loan (alien) words not only occur together with native language words but are also written in native scripts. For example, consider a Telugu sentence, $\frac{baank\ -ku}{bank\ -to}\ \frac{veldaama}{go}$. The first word “bank” occurs with other Telugu words and is also written in Telugu script. By taking this actual written form, the TTS system would pronounce it as “/B/ /AH/ /N/ /K/” as apposed to “/B/ /AE/ /N/ /K/”/. Such words occur typically in political, finance, health-care and sports domain which needs to be identified automatically and pronounced correctly. In this case, the word “bank (written as baank)” has to be identified and pronounced correctly. Since these words occur as sub-elements of a sentence, a word-level tag indicating the presence of a loan word may improve the quality of TTS systems for Indian languages significantly.

By having a word-level tag which specifies the presence of loan word in the input text the TTS system would plan to use a separate letter to sound rules, stress module etc for the loan words. The syllabic structure of the Indian languages allows detecting the loan words directly from the raw text. There are a few set of simple rules based on the syllabic structure of Indian languages which allows to detect the loan words from the Indian language scripts automatically.

3.3 Dialect Element <dialect>

It may often happen that it is not only the language, but the TTS system should be able to speak out a specific dialect. A regional dialect affects both the script form and the sounds of the language. Thus it should be possible to use a dialect tag to speak the content while remaining in the same language domain and probably the same voice. For example Telugu has at least three dialects: Andhra, Telegana and Rayalseema. It should be

possible to switch to a different dialect without changing/re-loading the language resources.

```
<?xml version="1.0"?>
<speak version="1.0" xml:lang="tel-in">
<voice gender="female">
  <dialect name = “andhra”> yekkad’iki vel’laali </dialect>
  <dialect name = “telengana” pro = “yaad’iki poovaale”> yekkad’iki
vel’laali </dialect>
</voice>
</speak>
```

The dialect should be able to support attributes such as alternate pronunciation (pro) of the words.

4. Conclusions:

In this paper, we discussed the position of Bhrigus Inc. Hyd and IIIT Hyderabad in building the TTS systems for Indian languages. We described the nature of scripts of languages. Keeping in view of the issues with the scripts of these languages, we discussed the importance of syllable, alien and dialect elements and their behaviour to improve the quality of TTS system in the context of Indian languages and other similar languages in the Asia-Pacific.