

# SSML Extensions for Indian Languages

Samuel Thomas, Ashish Verma and Nitendra Rajput  
IBM India Research Lab, New Delhi, India

## 1. Introduction

The Speech Synthesis Markup Language specification (SSML) is a W3C markup language specification that defines directives in the form of XML tags that can be used along with Text-to-Speech synthesis systems (TTS) to control different speech parameters (e.g. pronunciation, prosody) and also provide additional information such as, language and metadata for enhancing the quality of synthetic speech output in voice based applications. This paper brings out certain aspects that would be useful in addition to the existing tags that are specified in the current SSML version, especially in the context of Indian languages.

## 2. Position

IBM has been actively working in the area of text-to-speech synthesis and speech recognition in various languages for server and embedded applications. IBM's India Research Lab (IRL) is focusing on local language speech technologies such as in Hindi and Indian English.

## 3. Proposals

We propose two tag extensions to the existing SSML specification in the context of Indian languages.

### 3.1 Transliteration tag - <transliterate>

The text input to most Indian language TTS systems is either an English transliteration of the Indian language script or is in Unicode. There are however no uniform transliteration scheme to represent different Indian language scripts. Various Indian language TTS systems assume a particular input transliteration scheme. A popular scheme in use is the IRANS package [1]. It is therefore important to have a mechanism to specify whether the input text is using Unicode or has been transliterated using a particular transliteration scheme. We propose a <transliterate> tag that has two attributes: "codepage" and "uri". This can allow the correct phone sequence to be generated for the input text stream by the speech synthesizer. An example of such a situation is given below.

```
<?xml version="1.0"?>
<speech version="1.0" xml:lang="in">
  <transliterate codepage="1137">
    मेरा नाम राम है </transliterate>
  </speech>
```

```
<?xml version="1.0"?>
<speech version="1.0" xml:lang="in">
  <transliterate
    codepage="1252" uri="http://www.example.com/trans.file">
    mera nam ram hai </transliterate>
  </speech>
```

In these instances of the `<transliterate>` tag, the `codepage` attribute specifies how the input text has to be interpreted. In the first example, the `codepage` attribute is set to the code page index for Hindi which is 1137. This indicates that the text has to be interpreted using the Hindi character set and has not been transliterated. In the second example, because the input has already been transliterated to English, the `codepage` attribute is set to 1252 (the `codepage` for English). The mapping scheme that has been used for transliteration is indicated in the `uri` attribute. This tag is also useful because it would allow the speech synthesizer to use different text parsers depending on various transliteration schemes that could be used.

### 3.2 Tag to specify lexicon for phrases or words - `<foreign>`

The current SSML specification has `<lexicon>`, `<phoneme>` tags that can be used to find the pronunciation of words or phrases. The `<lexicon>` tag is used to reference external pronunciation dictionaries that are applicable to the entire document. In the case of the `<phoneme>` tag, the pronunciation for the word or short phrase has to be specified explicitly. We propose to have a tag that can be used to indicate that a certain word or phrase needs to be pronounced using a different pronunciation scheme without having to specify its exact phone sequence. In this case, the tag would point to a lexicon which is different from the globally specified lexicon for the whole document. Such a tag would be helpful when dealing with foreign language words/phrases embedded in a given language text or even in the case of loan words. We propose a tag `<foreign>` tag that has two attributes "lang" and "uri".

```
<?xml version="1.0"?>
<speak version="1.0" xml:lang="en-US">
  I greeted her with a <foreign lang="in"
  uri="http://www.example.com/lex.file"> namaste </foreign> and
  showed her where she could get a ticket for the movie <foreign
  lang="in" uri="http://www.example.com/lex.file"> "Jaane bhi do
  yaaron"</foreign>
</speak>
```

In the above example, instances of the tag `</foreign>` tag have been used to indicate how a Hindi word – *namaste* (Indian greeting) and a Hindi phrase “*Jaane bhi do yaaron*” (the name of a Hindi movie) have been used in an English sentence without having to specifying their exact pronunciations. External lexicons are specified for these phrases using the `uri` attribute of the tags.

## 4. Conclusions

In this paper, two important extensions in the context of Indian languages have been proposed based on our experience. Speech synthesis based technologies and products have a large scope in the Indian context. We believe these extensions would be helpful in the on going internationalization efforts of SSML.

## 5. References

[1] The ITRANS specification, <http://www.aczoom.com/itrans/>