

SSML for Urdu Speech Synthesis

*Sarmad Hussain, Professor and Head, Center for Research in Urdu Language Processing,
National University of Computer and Emerging Sciences. Lahore, Pakistan.*

sarmad.hussain@nu.edu.pk

www.crupl.org

The comments reported here are based on experience in the development of an Urdu Text-to-Speech synthesis system over last 3 years¹. A complete system has been developed which contains a Natural Language Processor (NLP) and a Speech Synthesizer (SS). NLP is divided into text normalization and phonological processing unit (latter containing letter to sound rules, syllabification and stress assignment). SS takes the annotated output of NLP and converts it into speech, based on concatenative synthesis. The entire pipeline of NLP and SS has been designed engineered and developed in the project.

Existing SSML standard covers much of the requirements for Urdu voice browsing. However, there are some issues which are being highlighted in this document which may need to be further addressed.

Multilingual text

It is becoming increasingly common for Urdu web pages to contain English text. It can be processed in three different ways: One way to read such words is to switch the language for English text. However, another option is to transliterate the text and read it as Urdu text. Finally English text can be spelled out character by character. Either of these policies may be needed to be defined at document level, with word level definition over-riding it. However, no document level tags exist to handle multilingual text.

Digits

It is also becoming increasingly common in web documents to write date, time, etc. using “English” digits and western formats. However, they represent date, time, etc. in Urdu and should be read in Urdu. Thus, both “2” and “۲” should be read as /do/ and not as /tu/ and /do/ respectively. But, there can be cases where “2” should be read as /tu/ in these documents. Thus, there should be an attribute in Urdu documents which can indicate how to treat “English” digits.

Date Format

The date formats defined in the “Say-as” do not leave provision to include Sanah sign (Unicode), which is an essential for writing dates in Urdu.

Also, there is no provision to specify the type of calendar. Same sequence of digits in a date can represent the lunar or solar calendars. There must be some tags to specify the calendar being used at document level and means to over-ride that at each instance of date.

Diacritics

There needs to be some document level definitions on how to treat the diacritics which are given in the text and those which are missing. For example, for optional diacritics, a given diacritic could over-ride that given in the lexicon. Alternatively, the parameters can be set to only follow the diacritics in the lexicon. The latter option is required because there are significant diacritic

¹ The project has been funded by E-Government Directorate of Ministry of IT, Government of Pakistan.

errors, if they are typed because correct pronunciation of words is not sometimes known. However, this option will not be able to handle new words not in the lexicon. In such cases, the former option would work.

If diacritics are not given, the character sequence may map into multiple words in the lexicon. The choice of pronunciation then can depend on multiple strategies, simplest one based on frequency of different variations. However, more advanced analysis at syntactic and semantic levels are also possible (same problem as for “perfect” in English, in which the stressed syllable is determined by the POS). Deeper analyses take much longer time to process. To control the performance of the system, it may be necessary to allow the tags to determine how deep the analysis should be done.

Space between words

Though space is required between words for processing, it is optionally type in Urdu. One can have redundant space within words like یونیورسٹی and sometimes may not have space between two words. Again, policy is needed to be defined for the document (using a tag and its attribute) to determine whether the processor at the back should try to say the text as is, or whether it should try to process it and correct space-related issues. This also has implications on performance and accuracy.