SSML extensions for multi-language usage

Author: Davide Bonardo – davide.bonardo@loquendo.com
Loquendo S.p.A., Vocal Technology and Services, Italy

Abstract

The document describes two proposals: the first regarding a new extension of the `<say-as>` element and the second regarding the introduction of a new element to control in a more detailed manner a synthesis processor in case of multi-language services.

Proposal 1

The synthesis processors are generally able to synthesize every input text, and try to have the best result as acoustic output apply rules over text normalization, word pronunciation, prosody, reading style and so on. Often a specific control is necessary to help the processors during the different phases of synthesis process.
A markup insert in the text assures the synchronization of the command in the point where it is needed.
When a vocal service is developed, usually the context of the dialog is known and the developer wants to prepare the messages with all the controls for the speech engine. Many dialog contexts today are well-known, so it is important to have a easy way to indicate the text type and all the information for the correct text pronunciation.
For example, we can consider a service for the reading of SMS-messages, or news, or special messages for rescue operations and emergency and so on. This would involve the expansion of many acronyms and abbreviations (that normally are not present in the language, or have a special context depending expansion) into the correct words. Moreover, it would also involve the activation of algorithms for prosodic phrasing, because the punctuation marks could be avoided – for instance the Short Message Service allowing text messages of up to 160 characters. Last but not least, it would involve a different reading style. Not to mention that this knowledge on acronyms and algorithms are language dependent.

An easy solution would be to unify all these instructions, using the `<say-as>` element, extending the `interpret-as` attribute with new values, as SMS, NEWS, etc.
The language could be the active language or could be specified as new attribute.

Ex.: `<say-as interpret-as="sms">I call you asap.</say-as>`

Proposal 2

In the SSML it is possible to specify the language using `xml:lang` attribute, that can be specified in many elements.
The speech processor interpretation of this attribute depends on the nature of the element. For example, when the language is specified inside the element `<voice>` the command for the processor is to find a voice that meets the request. But when the language is specified in the `<p>` or `<s>` element, the behavior could be different: the processor could use the same voice but activates the modules of the new language for text analysis and phonetic transcription.
However the part of the text where the change of the language is necessary could be smaller than a paragraph; in some cases you could need to change the language for a phrase or for single words:

```
<speak version="1.0" xml:lang="en">
   …
   The title is: "La vita è bella".
   …
</speak>
```

A possible solution for this problem could be to introduce a new element, `<token>`, with the attribute `xml:lang`.
In this case:

```
<speak version="1.0" xml:lang="en">
   …
   The title is: <token xml:lang="it">"La vita è bella".</token>
   …
</speak>
```

This solution is open to other purposes: the element `<token>` could be used to insert information for the segmentation where needed (useful in some languages where a word terminator is not present). Another possibility could be to have other attributes to specify something about the token (for instance to define the part of speech).

References

[1] Speech Synthesis Markup Language (SSML) Version 1.0, W3C Recommendation, 7 September 2004 (http://www.w3.org/TR/2004/REC-speech-synthesis-20040907/)

[2] SSML 1.0 say-as attribute values, W3C Working Group Note, 26 May 2005 (http://www.w3.org/TR/2005/NOTE-ssml-sayas-20050526/)