Watson Text to Speech

Working with SSML

The Speech Synthesis Markup Language (SSML) is an XML-based markup language designed to provide annotations of text for speech-synthesis applications. It is a recommendation of the W3C that has been adopted as the standard. SSML provides developers of speech applications with a standard way to control aspects of the synthesis process, allowing to specify pronunciation, volume, pitch, speed, and so on via markup.

With the Text to Speech service, you can use SSML to control the synthesis of your text with all supported languages.

Introduction to SSML

SSML operates by augmenting the plain text that is passed to a synthesizer with a predefined set of tags. An XML parser first separates the plain input text from the markup specifications. The specifications are then processed and sent as a set of instructions in a form that can be understood by the synthesizer to produce the desired effects.

<prosody>

The `<prosody>` tag controls the pitch, range, speaking rate, and volume of the text. All attributes are optional, but an error occurs if no attribute is specified.

- `pitch` modifies the baseline pitch for the text within the tag. Accepted values are a relative change value in semitones (for example `+2st`) or one keyword (`default`, `x-low`, `low`, `medium`, `high` or `x-high`)
- `rate` indicates a change in the speaking rate for the text within the tag. The rate is specified in terms of words per minute; Also, relative value changes (for example, `+15`)
- `volume` modifies the volume for the text within the tag. Specify an integer or decimal value in the range of 1.0 to 100.0

Example:

```
<prosody volume="75">Modified volume is 75</prosody>
<prosody volume="88.9">Modified volume is 88.9</prosody>
<prosody rate="slow">Decrease speaking rate by 25%</prosody>
<prosody rate="50">Set speaking rate at 50 words per minute</prosody>
<prosody pitch="-2st">Lower pitch by 2 semitones from baseline</prosody>
<prosody pitch="+2st">Increase pitch by 2 semitones from baseline</prosody>
```

Example combining all the attributes:

```
   <prosody volume="90" rate="+15%" pitch="+2st">Enthusiast sentence.</prosody>
   <prosody volume="40" rate="-15%" pitch="-1st">Not very excited sentence.</prosody>
```

<sub>

The `<sub>` tag indicates that the text specified in the `alias` attribute is to replace the text enclosed within the tag when speech is synthesized.
Example:

```
<speak version="1.0">
   <sub alias="International Business Machines">IBM</sub>
</speak>
```

<break>

The `<break>` tag inserts a pause into the spoken text. It has the following attributes:

- `strength` specifies the length of the pause in terms of varying strength values: `x-weak`, `weak`, `medium`, `strong`, or `x-strong`.

Example:

```
   Different sized <break strength="weak" />   <!--weak pause-->
   Different sized <break strength="medium" />   <!--medium pause-->
```

Documentation

https://www.ibm.com/watson/developercloud/doc/text-to-speech/SSML.shtml