

---

# Internationalizing Speech Synthesis

Daniel C. Burnett, Vexeo

Zhiwei Shuang, IBM

Co-chairs of W3C SSML 1.1 Subgroup

# Speech Synthesis Markup Language 1.0

---

- The Speech Synthesis Markup Language 1.0 ([SSML 1.0](#)), a W3C Recommendation since 2004, is designed to provide a rich, XML-based markup language for assisting the generation of synthetic speech in Web and other applications.
- The essential role of the markup language is to provide authors of synthesizable content a standard way to control aspects of speech such as pronunciation, volume, pitch, rate, etc. across different synthesis-capable platforms.

# SSML needs to support more languages!

---

- It is estimated that within 3 years the World Wide Web will contain significantly more content from currently under-represented languages, e.g. Mandarin Chinese and Hindi.
- In order to make SSML more useful in current and emerging markets, more enhancements for handling those languages are required.
- Many other languages would also benefit from a new "international" version of SSML, and it would help spread the Web to places where it is not so readily accessible.

# From SSML 1.0 to SSML 1.1

---

- SSML 1.1 improves on W3C's SSML 1.0 Recommendation by adding support for more conventions and practices of the world's languages.
  - Helps to disambiguate "word boundaries" in languages that do not use whitespace as a word boundary, including Chinese, Thai, and Japanese.
  - SSML 1.1 allows references to language-specific pronunciation alphabets.
  - It provides finer-grained control over lexicon activation and entry usage.
  - It clarifies the relationship between the author's specified speaking voice and the language being spoken.
  - Provides features to better integrate with existing and upcoming Speech Interface Framework specifications.

# Word Element

**In SSML 1.0: “p”-paragraph, “s”-sentence**

**In SSML 1.1: “w” element added to disambiguate "word boundaries"**

```
<?xml version="1.0"?>
<speak version="1.1" xmlns="http://www.w3.org/2001/10/synthesis"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
      http://www.w3.org/TR/speech-synthesis/synthesis.xsd"
      xml:lang="zh-CN">

  <!-- The Nanjing Changjiang River Bridge -->
  <w>南京市</w><w>长江大桥</w>
  <!-- The mayor of Nanjin city, Jiang Daqiao -->
  南京市长<w>江大桥</w>
  <!-- Shanghai is a metropoli -->
  上海是个<w>大都会</w>
  <!-- Most Shanghainese will say something like that -->
  上海人<w>大都</w>会那么说
</speak>
```

# Language-specific pronunciation alphabets

---

- In SSML 1.0:
  - The only valid values for the “alphabet” attribute are **"ipa"** (see the next paragraph) and vendor-defined strings of the form **"x-organization"** or **"x-organization-alphabet"**.
  - However, “ipa” is not convenient or popular for Chinese & many other languages.
- In SSML 1.1
  - We are planning to create a registry which includes pronunciation alphabets to be used with "alphabet" attribute of "phoneme" in SSML.

# Finer-grained control over lexicon

---

- In SSML 1.0:
  - Reference one or more external pronunciation lexicon documents for whole SSML document.
- In SSML 1.1
  - “lookup” element to specify the lexicon

```
<lexicon uri="http://www.example.com/lexicon.pls"
  xml:id="pls"/>
<lexicon uri="http://www.example.com/strange-words.file"
  xml:id="sw"
  type="media-type"/>
<lookup ref="pls">
  tokens here are looked up in lexicon.pls
  <lookup ref="sw">
    tokens here are looked up first in strange-words.file and then, if not found, in lexicon.pls
  </lookup>
  tokens here are looked up in lexicon.pls
</lookup>
tokens here are not looked up in lexicon documents
```



# Voice Element

---

- In SSML 1.0
  - “voice” element is twisted with “xml:lang”, the change of language will probably change the “voice”.
  - The output of Mix Language Text is uncontrollable.
- In SSML 1.1
  - Clarifies the relationship between the author's specified speaking voice and the language being spoken.
  - Change of “xml:lang” should not change the voice!

Example of SSML 1.1:

```
<voice languages="en:zh,zh:zh">
```

```
  I come from <lang xml:lang="zh-CN"> 中国 </lang>.
```

```
</voice>
```



# SSML 1.1 Participants

---

- Editors:
  - Voxeo
  - IBM
- Authors:
  - Nuance Communications
  - Loquendo
  - France Telecom
  - Tellme
  - Chinese Academy of Sciences
  - Panasonic
  - Toshiba
  - HP
  - Chinese University of Hong Kong
  - iFLYTEK
- Other Contributors:
  - InterVoice
  - Related W3C working groups

# SSML Workshop & Meetings

---

- **Three Workshop to gather requirements:**
  - Beijing Workshop, 2-3 November, 2005
  - Greece Workshop, 30-31 May, 2006
  - India Workshop, 13-14 January, 2007
- **Nine Face to Face Meetings**
  - Most recent one: 9th f2f Meeting in Beijing, 22-24 April, 2008
  - Other 8 f2f Meeting from April 2006- 2008

# SSML 1.1 Working Draft Status

---

- First Public Working Draft (10 January, 2007)
- Second Working Draft (11 June, 2007)
- Third Working Draft (4 September, 2007)
- Fourth Working Draft (12 December, 2007)
- **Fifth Working Draft (17 March, 2008)**

---

Thank you!