How IPA is Used in SSML and PLS

Paolo Baggia, Loquendo
Wed. August 9th, 2006
Agenda

- Modifying Pronunciation:
  - Why? Where? When?

- From Written to Spoken:
  - Pronunciation Alphabets
  - IPA – International Phonetic Alphabet

- Adding Pronunciations for Speech Applications:
  - For TTS
  - For ASR (if time)
Why Change Pronunciation?

- The need to customize, improve, adapt pronunciation in your speech application
  - Improve reading of prompts by TTS
  - Improve ASR performance in atypical cases

- To create applications with prompts and grammars that mix languages, i.e. movies, news, etc.

- To create applications for many languages, such as large portals, CRMs, etc.
Where?

- VoiceXML 2.0/2.1 applications are the reference scenarios
  - Prompts are based on SSML 1.0
  - Recognition grammars are based on SRGS 1.0

- The specification of Pronunciation Lexicon Specification (PLS) is ongoing
  - Container for customized pronunciations to be used in both SSML and SRGS
When?

Two options:

• Fix mispronunciations in the specific prompt or grammar

• Create resources to be used in specific context, e.g. a pronunciation lexicon for movie titles another one for restaurants, and so on
From Written to Spoken

➢ For text to be read or understood:
  • Written in a given language (e.g. English, Chinese, Arabic)
  • Each word is composed of letters in a written alphabet (26 in English, x in Chinese, y in Arabic)

➢ When pronunciation is different from spelling:
  • For many other languages where the pronunciation follows strict rules
How to Modify Pronunciations?

➢ Need to use a *phonetic alphabet*!

• Proprietary phonetic alphabets:
  – Each speech technology provider supports several phonetic alphabets

• General/standard phonetic alphabets:
  – SAMPA
    ASCII based (simple to write)
  – IPA – *International Phonetic Alphabet*
    “universally agreed system of notation for sounds of languages”
    ▪ Covers all languages
    ▪ Requires UNICODE to write it

• Other well-known phonetic alphabets:
  – Chinese Mandarin: Pinyin
  – Japanese: JEITA
Adding Pronunciations to SSML 1.0

- `<phoneme>` element to customize pronunciation:
  
  `<phoneme alphabet="ipa" ph="">baggia</phoneme>`
  `<phoneme alphabet="x-loquendo" ph=""></phoneme>`

- Requested extensions for Eastern languages:
  - Pinyin: widely used system of romanization of Standard Chinese
    `<phoneme alphabet="pinyin" ph="ma1">媽</phoneme>`
  - JEITA: for Japanese language
    `<phoneme alphabet="jeita" ph="??">日本語</phoneme>`

- `<sub>` element to expand acronyms, short forms:
  
  `<sub alias="World Wide Web">W3C</sub>`
  `<sub alias="later">l8r</sub>`
An SSML 1.0 Example Document

- This is a simple SSML 1.0 document:
  ```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<speak version="1.0"
  xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  The title of the movie is: "La vita è bella" (Life is beautiful),
  which is directed by Roberto Benigni.
</speak>
  ```

- This is an enhancement of the same example:
  ```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<speak version="1.0"
  xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">
  The title of the movie is: <phoneme alphabet="ipa" ph="ˈla ˈviːə ˈʔeɾ ˈbeːle">"La vita è bella"</phoneme>
  (Life is beautiful),
  which is directed by
  <phoneme alphabet="ipa" ph="ɛrˈərmənu bɛˈniːnji">Roberto Benigni.</phoneme>
</speak>
  ```
A PLS document is a container of several <lexeme>s

Each <lexeme> contains:
- One or more spellings
- One or more pronunciations or substitutions

SSML 1.0 and SRGS 1.0 document can reference one or more PLS document
An SSML 1.0 Example with PLS

The same SSML 1.0 document with reference to PLS:

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<speak version="1.0"
   xmlns="http://www.w3.org/2001/10/synthesis" xml:lang="en-US">

  <lexicon uri="www.example.com/mylexicon.pls"/>

  The title of the movie is: "La vita è bella" (Life is beautiful),
  which is directed by Roberto Benigni.
</speak>
```

Here is the referenced PLS document:

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<lexicon version="1.0" xmlns="http://www.w3.org/2005/pronunciation-lexicon"
   alphabet="ipa" xml:lang="en-US">

  <!-- to be finished -->
</lexicon>
```
International Phonetic Alphabet

- Created by International Phonetic Association (active since 1896), collaborative effort by all the major phoneticians around the world


- IPA – International Phonetic Alphabet is largely used by phoneticians, by dictionaries and phonetic resources

  ➔ W3C chose to normatively reference IPA in SSML and PLS specifications

For more details on IPA: [http://www.arts.gla.ac.uk/IPA/](http://www.arts.gla.ac.uk/IPA/) and to listen to sounds from languages: [http://phonetics.ucla.edu/index/sounds.html](http://phonetics.ucla.edu/index/sounds.html)
IPA Full Chart

- Describes the phonemes that cover all the world languages:
  - Consonants
  - Vowels
  - Other Symbols
  - Diacritics
  - Suprasegmental
  - Tones and Word Accent

- IPA is used by phoneticians for broad and narrow transcriptions

- IPA is used in many dictionaries

Acknowledgements to the International Phonetic Association
Consonants (some)

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Postalveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pharyngeal</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p b</td>
<td></td>
<td>t</td>
<td>d</td>
<td>t d c j k g q g</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m ñ</td>
<td></td>
<td>ñ</td>
<td>ñ</td>
<td>ñ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>B r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap or Flap</td>
<td></td>
<td>V̄</td>
<td>r̅</td>
<td></td>
<td></td>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>ɸ β f v</td>
<td>θ ð s z j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>l ɹ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>u j</td>
<td></td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>l ɹ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

- All these are possible Pulmonic Consonants
- The columns are “places of articulation”
- The rows are “manner of articulation”
- The gray areas are considered to be impossible to articulate
Vowels

- Vowels:
  - A speech sound created by the relatively free passage of breath through the larynx and oral cavity, usually forming the most prominent and central sound of a syllable.
  - Vowels are distinguished on the basis of “Height” and “Backness”.
  - The IPA diagram resembles the place of articulation of the vowels.
Diacritics

- Small marks that can be added to a symbol to modify its value
- Used to differentiate allophones of a phoneme

They are very important for narrow transcriptions, which shows more phonetic details

Diarycritics may be placed above a symbol with a descender, e.g., ħ

<table>
<thead>
<tr>
<th>Voiceless</th>
<th>Voiced</th>
<th>Breathy voiced</th>
<th>Dental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opaque</td>
<td>Open</td>
<td>Close</td>
<td>Consonant</td>
</tr>
<tr>
<td>Aspirated</td>
<td>Inflated</td>
<td>Creaky voiced</td>
<td>Apical</td>
</tr>
<tr>
<td>More rounded</td>
<td>Labialized</td>
<td>Nasalized</td>
<td></td>
</tr>
<tr>
<td>Less rounded</td>
<td>Palatalized</td>
<td>Nasal release</td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>Velarized</td>
<td>Lateral release</td>
<td></td>
</tr>
<tr>
<td>Retracted</td>
<td>Pharyngealized</td>
<td>No audible release</td>
<td></td>
</tr>
<tr>
<td>Centralized</td>
<td>Velarized or pharyngealized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-centralized</td>
<td>Raised</td>
<td>Voice of alveolar fricative</td>
<td></td>
</tr>
<tr>
<td>Syllabic</td>
<td>Lowered</td>
<td>Voice of bilabial approximant</td>
<td></td>
</tr>
<tr>
<td>Non-syllabic</td>
<td>Advanced Tongue Root</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhotic</td>
<td>Retracted Tongue Root</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Loquendo

SpeechTEK®
Suprasegmentals and Tones

Suprasegmentals
- Aspects of speech that involve more than single phonemes
- The principal features are stress, length, tone and intonation

Tones and Word Accents
- Pitch variations that affect the meaning of word
  - i.e. /ma/ in Chinese Mandarin may mean "mother", "hemp", "horse", or "scold", by changing tone from "high level", "low level", "rising", and "going"
Conclusions

➢ In many speech applications, there is the need to modify pronunciation

➢ SSML 1.0 and SRGS 1.0 with the help of PLS 1.0 are giving support to address it

➢ IPA is a candidate to describe the pronunciation of many languages

➢ In some cases like Chinese Mandarin or Japanese other phonetic alphabet might be adopted
Empowering Your customers and employees with speech technologies

SpeechTEKI 2006 Empower
The Voice Solutions Showcase

Thank You