SSML 1.1: The Internationalization of SSML

Daniel C. Burnett
August 9, 2006
SSML 1.0

• Widely used
• Convenient for many languages

• However, . . .
Chinese tones

• Mandarin is syllable-based, with tone movement a distinguishing feature of the syllable
  妈(mā) 麻(má) 马(mǎ) 骂(mà)

• IPA is cumbersome when only the tone needs to be corrected
  – Eg., correcting Tone Sandhi
    你好 ni3 hao3 → ni2 hao3
Chinese word boundaries

• Word boundaries are not given in typical writing

• 這一晚會如常舉行
  - 這一 晚會 如常 舉行 means “This banquet is held as usual”
  - 這一 晚 會 如常 舉行 means “Tonight will be held as usual”
Chinese names

• Chinese characters are pronounced differently (in a consistent manner) in names, particularly family names

• Cantonese example: 單明明
  單 /daan1/ → /sin6/ (surname)
  明明 /ming4 ming4/ → /ming4 ming2/ (given name)
Japanese Ruby

- Ruby is a typesetter’s annotation used in everyday print media. It
  - disambiguates Kanji text (Chinese characters)
  - does this by giving the pronunciation

- Every Japanese person knows how to read it

- Why not use it for pronunciation?
Mixed languages

- “Tonight’s movie is ‘La vita è bella’.”
- Japanese and Chinese use the same characters, but often with very different meanings
- How should mixed-language text be annotated?
- How do you change the language without changing the voice?
  - What does this question really mean?
“Oh, and one more thing . . .”

- Korean/Hungarian need for PoS
- Sub-word level prosody annotation (e.g., contrastive stress at syllable level in Hungarian)
- Text with missing diacritics (e.g., Polish SMS text)
- Other simplified/non-traditional text
- Better support for highly-agglutinative languages
SSML 1.1

- Two workshops to solicit such examples
- SSML subgroup of W3C Voice Browser Working Group
  - Has met twice
  - Expects to release requirements later this year
SSML subgroup “charter”

“. . . For Mandarin, Cantonese, Hindi*, Arabic*, Russian*, Korean, and Japanese, we will identify and address language phenomena that must be addressed to enable support for the language. Where possible we will address these phenomena in a way that is most broadly useful across many languages. We have chosen these languages because of their economic impact and expected group expertise and contribution. . . .”

* provided there is sufficient group expertise and contribution for these languages
Some possible requirements

- Pronunciation scripts
- Word boundary
- Name identification
- Language indication
- Lexicon activation
Pronunciation scripts

• `<phoneme alphabet=“whatever” …/>`
• Today, values other than IPA are permitted but not standardized
• New requirement might be:
  – to establish registry (eg., at IANA) for standardizing values for
    • Pinyin
    • Jyutping
    • Ruby
    • etc.
Word boundary

• New requirement might be
  – to provide mechanism to eliminate word segmentation ambiguities
• Note that white space is insufficient because
  – some languages (such as Vietnamese) use white space for syllable segmentation
  – some languages (such as Urdu) use white space for other purposes
Name identification

• New requirement might be
  – to provide a mechanism to identify content as a proper noun or a name
Language indication

- xml:lang is used in all XML languages to mean the language of the content
- Successor to RFC3266 clarifies region and dialect encoding
  
  language – script – region – variant – extension – private_use
  
  “zh-Hans-CN”

- New requirements might be
  
  - To clarify that xml:lang only indicates the language of the content
  - To specify that selection of voice and language are independent and that TTS vendors must document supported combinations of language and voice
Lexicon activation

• Today, implicit lexicon activation in SSML based on
  – Language
  – Document order

• New requirement might be
  – Support explicit author control over which lexicons are used for which portions of the SSML document
Get involved

• W3C Voice Browser Working Group
  – Responsible for VoiceXML, SSML, SRGS, and many other speech-related standards

• SSML subgroup
  – Seeking experts in Russian, Hebrew, and Arabic

• Visit http://www.w3.org/Voice for more info
Empowering Your customers and employees with speech technologies

SpeechTEK! 2006 Empower
The Voice Solutions Showcase

Thank You