

# SSML 1.1

Daniel C. Burnett  
Nuance Communications  
January 13, 2007

# Overview

- SSML 1.1 Charter
- Goals of SSML 1.1
- Development process
- Requirements (so far)
- Specification changes (so far)

# SSML 1.1 Charter

- Extend SSML 1.0 to
  - Provide enhanced language support
  - Fix incompatibilities with other VBWG specs
- Out of scope are
  - VCR-like audio controls
  - <say-as> changes (but not requirements)

*(See Requirements Sec. 1.2 for details)*

# My goals for SSML 1.1

- Satisfy the charter
  - Making only minimal changes to SSML 1.0
  - While satisfying the subgroup members
- Defer out- of- scope changes/ requests to SSML 2.0 (potentially a re- write)

# Development process

- Review topics from Workshops
- Categories & Categorization
- Write concise problem statements
- Develop requirements from problem statements
- Agree on major points of proposals to address requirements
- Write proposals

# Topic review

- Review all the topics from the workshops
- Assign each to a category:
  - Short-term (will work on these now)
  - Long-term (will revisit when short-term is done)
  - Experts needed (we will only work on this if experts in the related languages join the working group)
  - Other SSML work (say-as, SSML 2.0 items, etc.)

# Categories and categorization

Short- term (group agrees to work on this)	Long- term (after short- term work will revisit to determine if belongs in group)	Experts needed (in order to make decision to work on this in this subgroup)	Other SSML work (SSML 2.0 or later, < say- as> Note, etc.
Token/ word boundaries	Expand Part- Of- Speech support	Providing number, case, gender agreement info	Background sound (may be handled best by VoiceXML3 work)
Ruby	Tones	Syllable markup	Tone sandhi
Verify that RFC3066 language categories are complete enough that we do not need anything new beyond xml:lang to identify languages and dialects	Text with multiple languages (changing xml:lang without changing voice; separately specifying language of content and language to speak)	Diacritics, SMS text, simplified/ alternate text	Enhance prosody rate to include "speech units per time unit" where speech units would be syllable, mora, phoneme, foot, etc. and time unit would be seconds, ms, minutes, etc.(would address mora/ sec request)
Chinese names (say- as requirements)		Sub- word unit demarcation and annotation	Special words
Phonetic alphabets			Expressive elements
			Sentence structure

# Problem statements

- Statement for “xml:lang” topic
  - The xml:lang attribute in SSML is the only way to identify the language. It represents both the natural (human) language of the text content and the natural (human) language the synthesis processor is to produce. For languages whose scripts are ideographs rather than pronunciation-related, we are not sure that the permitted values for xml:lang, as specified by RFC3066, are detailed enough to distinguish among languages (and their dialects) that use the same ideographs.



# Requirements

- Requirements for “xml:lang” problem statement:
  - SSML 1.1 must ensure the use of a version of xml:lang that uses the successor specification to [[RFC3066](#)] (for example, [[BCP47](#)]).
  - SSML 1.1 must clearly state that the 'xml:lang' attribute identifies the language of the content.
  - SSML 1.1 must clearly state that processors are expected to determine how to render the content based on the value of the 'xml:lang' attribute and must document expected rendering behavior for the xml:lang values they support.
  - SSML 1.1 must specify that selection of xml:lang and voice are independent. It is the responsibility of the TTS vendor to decide and document which languages are supported by which voices and in what way.

# Requirements

- Requirements for “xml:lang” problem statement:
  - SSML 1.1 must ensure the use of a version of xml:lang that uses the successor specification to [RFC3066] (for example, [BCP47]).
  - SSML 1.1 must clearly state that the 'xml:lang' attribute identifies the language of the content.
  - SSML 1.1 must clearly state that processors are expected to determine how to render the content based on the value of the 'xml:lang' attribute and must document expected rendering behavior for the xml:lang values they support.
  - SSML 1.1 must specify that selection of xml:lang and voice are independent. It is the responsibility of the TTS vendor to decide and document which languages are supported by which voices and in what way.

# Major points

- Approach
  - We will modify the descriptions of `xml:lang` and `<voice>` to clarify that language and voice values may be set separately. Synthesis processors should document supported combinations of voice and language and the behavior for unsupported combinations.
- Proposal 1
  - create/ use a new element that can be used to set `xml:lang` for text sizes between sentence and word. Its sole function is language annotation. Modify existing voice element description to clearly separate language setting and voice setting and require one of name, age, gender, or variant to be set.
  - ```
<s xml:lang="zh-cmn-HK"> foo  
  <lang xml:lang="en-GB">  
    bar  
  </lang>  
</s>
```
  - Major points:
    - In addition to what it says in the approach, update voice selection algorithm appropriately.
  - Volunteer: Paul has lead. Jerry Carter and Dan Burnett can help.
- Proposal 2
  - modify existing voice element description to clearly separate language setting and voice setting but still permit both to occur with this element.
  - ```
<s xml:lang="zh-cmn-HK"> foo  
  <voice xml:lang="en-GB"> bar </voice>  
</s>
```
  - Major points:
    - In addition to what it says in the approach, update voice selection algorithm appropriately.
  - Volunteer: Lou Xiaoyan

# Proposals

- Detailed text proposals are folded into the specification and reviewed.
- Outstanding issues are noted in the specification for later discussion.

# Requirements so far

- General requirements
- Speech Interface Framework requirements
  - Requirements needed to interoperate with other Voice Browser Working Group specifications
- Language-related requirements
  - These have been the primary focus of the subgroup

# Speech Interface Framework requirements

- Caching
- error messages
- type attribute on <audio>
- VoiceXML VCR control support
- Lexicon activation control
- prefetching support
- external reference to <p> , <s> , <w>

# Language- related requirements

- Token/ word boundary requirements
  - Mechanism to disambiguate word boundaries
  - Mechanism to indicate the language of the word
  - Mechanism to indicate lexicon entry to use for the word
- Phonetic Alphabet and Pronunciation Script Requirements
  - Registry for alternative pronunciation alphabets
- Language Category requirements
  - Successor to RFC3066
  - xml:lang requirements
- Name/ proper noun id requirements (future)
  - Identify content as proper noun
  - Identify content as name
  - Identify name sub- content as surname

# Specification changes so far

- `<w>` element
- `<lang>` element and `lang-voice` attribute
- pronunciation alphabet registry
- `<lookup>` element
- `<role>` attribute