

Issues of SSML in Japanese  
(The position of the JEITA Speech Group)

Wataru IMATAKE (ANIMO LIMITED)  
Makoto AKABANE (Sony Computer Entertainment Inc.)  
Kazuyo Tanaka (Tsukuba University)  
[imatake@animo.co.jp](mailto:imatake@animo.co.jp)  
[Makoto.Akabane@jp.sony.com](mailto:Makoto.Akabane@jp.sony.com)  
[ktanaka@slis.tsukuba.ac.jp](mailto:ktanaka@slis.tsukuba.ac.jp)

JEITA Technical Standardization Group on Speech Input/Output Systems

About JEITA Technical Standardization Group on Speech Input/Output Systems:

JEITA[1] Technical Standardization Group on Speech Input/Output Systems was organized for dealing standardization issues on speech technologies in 2001, following activities of former JEIDA Speech Committee. The Group is organized by members from major companies in electronic and information technologies industries in Japan: NTT, Hitachi, Sony, Toshiba, Matsushita, Mitsubishi, CANON, Sharp, Fujitsu, OKI, SANYO, IBM, NEC, OMRON, Ricoh, ANIMO, National Institute of AIST, etc.

Recent activities of the JEITA Speech Group:

The Group published “Standard for Speech Synthesis Systems Performance Evaluation Method, JEITA IT-4001” in 2002[2], “Symbols for Japanese Text-to-Speech Synthesizer, JEITA IT-4002”, and “Symbols for Japanese Speech Recognizer, JEITA IT-4003” both in March, 2005. The JEITA IT-4002 is a revised version of JEIDA-62-2000 (Standard of Symbols for Japanese Text-to-Speech Synthesizer[3]) issued in 2000, which included control tags for Japanese speech Synthesizer defined using XML. (However the control tags are removed in the IT-4002 for several reasons.) The Group is also continuously investigating and surveying speech technologies, devices and speech application systems.

Basic Position of the JEITA Speech Group:

We recently decided that the JEITA Speech Group recommends SSML for the standard of control tags for Japanese speech synthesizers. Therefore, we are now studying and examining details of SSML.

Comments on Specific Issues:

1. We established JEITA IT-4002 as a character set for speech synthesis descriptions. When refer to this standard in a phoneme element of SSML, we suggest that we describe it with "x-JEITA-4002-kana",

"x-JEITA-4002-ipa", "x-JEITA-4002-sampa" as alphabet attribute.

2. When using a break element in Japanese SSML document, we suggest the expansion that can specify by a mora unit as time attribute.
3. We think that the expansion that can specify by a mora/second is necessary for specification of speaking speed by a rate attribute in a prosody element in Japanese.
4. Specification of a ruby is necessary in Japanese.
5. We think that a designated method of how to read Japanese peculiarity is necessary for a say-as element. (ex: Japanese language readings of a number)
6. We want to identify usage of a voice element and a lexicon element.

#### References

[1] <http://www.jeita.or.jp/english/>

[2] S. Itahashi, et al, "Standard for Japanese Speech Synthesizer Performance Evaluation," Paper presented at Workshop Oriental COCOSDA 2004, New Delhi.

[3] K. Tanaka, et al, "JEIDA Standard of Symbols for Japanese Text-to-Speech Synthesizers," Paper presented at Workshop Oriental COCOSDA 2000, Beijing.