Position Paper for

W3C Workshop on Internationalizing

the Speech Synthesis Markup Language (SSML)

2-3 November 2005

Title: **The Usage of Pos(Part Of Speech) for Resolving Multiple Pronunciations in SSML**

Source: **KT**

Author: **Myoung-Wan Koo and Du-Seong Chang**

1. Introduction

The Speech Synthesis Markup Language Specification Version 1.0 (SSML[1]) is designed to provide a rich, XML-based markup language for assisting the generation of synthetic speech in Web and other applications. However, current SSML doesn't fully support the pronunciation lexicon for agglutinative language such as Korean, Turkish etc. In this position paper, we propose the use of pos for pronunciation lexicon in SSML. In SSML, the “lexicon” element is defined for pronunciation handling, which is expanded to include the pronunciation effect for both speech recognition and speech synthesis in pronunciation lexicon specification [2].

2. Pronunciation information in SSML

```
<?xml version="1.0"?>
<!DOCTYPE speak PUBLIC "-//W3C//DTD SYNTHESIS 1.0//EN"
              "http://www.w3.org/TR/speech-
synthesis/synthesis.dtd">
<speak version="1.0"  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="en-US">
  <lexicon uri="http://www.example.com/lexicon.file"/>
  <lexicon uri="http://www.example.com/strange-words.file"
         type="media-type"/>
  ...
</speak>
```

The lexicon element in SSML contains pronunciation information for tokens that can appear in

a text to be spoken. However, there is no detail explanation about how to express the pronunciation. Any number of lexicon elements may occur as children of the speak element. The lexicon element must have an uri attribute specifying a URI that identifies the location of the pronunciation lexicon document. Fig. 1 shows an example of the usage of lexicon element. It means that the “http://www.example.com/lexicon.file” has pronunciation information for each token and that http://www.example.com/strange-words.file has also pronunciation information in media type. And the format of lexicon dictionary is not recommended since the format of lexicon dictionary can be proprietary and language specific.

The phoneme element is also defined for phonemic/phonetic pronunciation of the contained text. The example of phoneme element is shown in Fig. 2.

```
<?xml version="1.0"?>
<speak version="1.0" 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="en-US">
  <phoneme alphabet="ipa" ph="t&#x259;mei&#x325;&#x27E;ou&#x325;">
  tomato </phoneme>
  <!-- This is an example of IPA using character entities -->
</speak>
```

Fig. 2. Example of the usuage of phoneme element

Fig.2 shows how “tomato” can be pronounced in the form of IPA (International Phonetic Association). This element can be used together with lexicon element.

3. Pronunciation information in Pronunciation Lexicon Specification

W3C released the pronunciation lexicon specification (PLS) version 1.0 on Feb., 2005. This specification may be helpful for the success of speech applications since it provides most ASR(automatic speech recognition) and TTS(Text-To-Speech) engines with extensive high quality lexicons having pronunciation information for most words or phrase. The PLS is designed to allow interoperable specification of pronunciation information for either ASR and TTS engines within voice browsing applications. Fig. 3 shows the simple PLS document for the word “tomato” and its pronunciation.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<lexicon version="1.0" alphabet="ipa" xml:lang="en-US">
  <lexeme>
    <grapheme>tomato</grapheme>
    <phoneme>t&#x259;mei&#x325;&#x27E;ou&#x325;</phoneme>
    <!-- This is an example of IPA phonetic string -->
   </lexeme>
</lexicon>
```

Fig. 3. Example of the usuage of lexicon, lexeme, elements

Fig.3 shows that lexicon element can be combined with lexeme element for showing pronunication information, which is compared with phoneme element for showing the pronunciation of same word “tomato” in Fig. 2. PLS can express detail pronunciation such as multiple pronunciations for the same orthograph. The detail information can be obtained in PLS version 1.0 [2].

4. The usage of POS information for resolving multiple pronunciations

In PLS specification, multiple pronunciations can be showed with phoneme elements. And attribute “prefer” can be used to give one pronunciation high priority among many pronunciation candidates. This “prefer” attribute can be well effective in speech synthesis. However this attribute can not be effective in speech recognition engine since “prefer” attribute does not give any special information except alternative pronunciation. There has been no element and attribute for resolving multiple pronunciations in PLS specification.

According to informative note of PLS spec., contextual processing is regarded as outside scope of the current version of PLS. However, pos information can reduce the overhead of resolving multiple pronunciations in speech synthesis and speech recognition engine. The word “refuse” can have two different pronunciations depending on pos information. The noun “refuse” has a different pronunciation to the verb “refuse”. The current specification doest not say about this.

We propose that "pos" information should be included in the current version. Of course, we can use "prefer" attribute for showing alternative pronunciations. However, this one may let speech recognition and synthesis engine need more cpu time. This kind of phenomenon can always happen in a large vocabulary continuous speech recognition system. In that system, basic units in pronunciation dictionary are not words any longer. Instead of word, morpheme (psedo-morpheme) are used since morpheme can reduce the size of vocabulary expecially in agglutinative language. We need to add "pos" information to morpheme to get a proper

pronunciation in a dictionary as well as to resolve multiple pronunciation in some words. We propose two methods to insert pos information in PLS specification.

4.1 Proposal 1: Pos attribute

The "pos" can be an optional attribute which indicates the detail information for obtaining the pronunciation for speech recognition and speech synthesis.The possible values are: "verb", "noun" etc.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
  xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="en-US">
  <lexeme>
    <grapheme>refuse</grapheme>
    <phoneme pos="verb">r&#x026A;'fju:z</phoneme>
    <!-- IPA string is: "r 'fju:z" -->
  </lexeme>
  <lexeme>
    <grapheme>refuse</grapheme>
    <phoneme pos="noun">'refju:s</phoneme>
  </lexeme>
```

Fig. 4 Example of pos attribute for resolving multiple pronunciations

Fig. 4 shows an example of the usage of pos attribute for resolving multiple pronunciations. And this attribute can be very effective in agglutinative language such as Korean, Turkis etc.

Fig. 5 shows other example for resolving multiple pronunciations in Korean.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="Korean">
  <lexeme>
    <grapheme>ki</grapheme>
    <phoneme pos="verb-ending">kki</phoneme>
    <!–감기 (gam ki) : winding -→
  </lexeme>
  <lexeme>
```

```
    <grapheme>ki</grapheme>
    <phoneme pos="noun">gi</phoneme>
<!—감기 (gam gi) : cold -→
  </lexeme>
</lexicon>
```

Fig. 5 The usage of pos information for resolving multiple pronunciations

4.2 Proposal 2: Pos element

The <lexeme> element may contain optionally one or more <pos> element. Each <pos> element contains CDATA specifying the pronunciation.

```
<?xml version="1.0" encoding="UTF-8"?>
<lexicon version="1.0"
  xmlns="http://www.w3.org/2005/01/pronunciation-lexicon"
     alphabet="ipa" xml:lang="en-US">
  <lexeme>
    <grapheme>refuse</grapheme>
    <phoneme>r&#x026A;'fju:z</phoneme>
    <pos> verb </pos>
    <!-- IPA string is: "r 'fju:z" -->
  </lexeme>
</lexicon>
```

Fig. 6 The usage of pos element

5. Conclusion

In this position paper, we propose to use the pos information for resolving multiple pronunciations. This information can be used to choose one pronunciation among multiple pronunciations and can be very effective in agglutinative language such as Korean, Turkis etc. Finally, this information can reduce the search time in a large vocabulary recognition system.

References

[1] Speech Synthesis Markup Language (SSML) Version 1.0 W3C Recommendation September 2004.

[2] Pronunciation Lexicon Specification (PLS) Version 1.0 W3C Working Draft 23 Aug.2005.