# SUPPLEMENT 8 – SPECIFICATION OF THE SPEECH MATERIAL ANNOTATION (SMA) iXML Keyword.

The Speech Material Annotation (SMA) defines a keyword in the iXML specification for embedded metadata in production media files (Cf. <http://www.ixml.info>).

The SMA iXML object is designed to hold data about a sound file that are relevant for use in speech perception testing, such as spellings, phonetic transcriptions, and sound levels. Each spelling, phonetic transcription or sound level can be associated with a segment of the sound file, in which the SMA object is stored. All data within the SMA object should be stored in the UTF-8 format, and use points [.] as decimal delimiters.

Within an iXML structure, the SMA keyword is introduced with an <SMA> tag. The <SMA> markup should always specify the SMA version used when writing the file by including as content the <SMA\_VERSION> tag. The current SMA version is 1.0.

The SMA specification defines four segmentation levels on which the segmentation data can be stored, namely: channel, sentence, word and phone. The segmentation levels are introduced by the tags <CHANNEL>, <SENTENCE>, <WORD> and <PHONE>, respectively. No segmentation levels are mandatory in a SMA object, but a <PHONE> tag must always be contained within a <WORD> tag, which in turn must be contained within a <SENTENCE> tag. The <SENTENCE> tag must always be contained within a <CHANNEL> tag, which is finally always a subtag of the <SMA> tag. On each segmentation level, any tags presented in table A1 may occur.

Table A1

*Tags occurring on all segmentation levels in the SMA iXML object.*

|  |  |
| --- | --- |
| Markup | Content |
| <ORTHOGRAPHIC\_FORM> | An orthographic representation of the segment. |
| <PHONETIC\_FORM> | A phonetic representation of the segment. |
| <START\_SAMPLE> | The (zero-based) index of the first audio sample in the segment. |
| <LENGTH> | The number of audio samples in the segment. |
| <UNWEIGHTED\_LEVEL> | The unweighted average sound level of the current segment. |
| <UNWEIGHTED\_PEAKLEVEL> | The unweighted peak sound level of the current segment. |
| <WEIGHTED\_LEVEL> | The weighted average sound level of the current segment. |
| <FREQUENCY\_WEIGHTING> | The frequency weighting used for the weighted level value. Predefined values are: Z, A, B, C, RLB and K. |
| <TIME\_WEIGHTING> | If higher than zero, the given value indicates that the weighted level tag designates a peak level calculated with the given value (in seconds) as temporal integration time. A value of zero indicates that the weighted level tag designates the average sound level of the whole segment. |

In addition to the tags defined in table A1, the sentence and word levels may have the tag <START\_TIME> giving the time in seconds relative to a user defined point in time. The sentence level may also have the tag <INITIAL\_PEAK> for which the value should indicate the initial absolute peak amplitude, i.e. the absolute value of the most negative or the positive sample value, of the segment. The initial peak value should only be stored once for each sentence, directly after sentence segmentation, but prior to any gain modification. As such, the initial peak value may be utilized to calculate any gain changes applied to the sound section associated with a specific sentence since the initial recording.

An open source software library written in Visual Basic .NET that enables reading and writing of wave files with SMA iXML chunk objects is available at <https://github.com/witteerik/SHT.Audio>.

Below is an example of a full SMA object of the single Swedish word <lax>.

<SMA>

<SMA\_VERSION>1.0</SMA\_VERSION>

<CHANNEL>

<START\_SAMPLE>0</START\_SAMPLE>

<LENGTH>86221</LENGTH>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

<SENTENCE>

<START\_SAMPLE>24000</START\_SAMPLE>

<LENGTH>38221</LENGTH>

<UNWEIGHTED\_LEVEL>-30.6053004411253</UNWEIGHTED\_LEVEL>

<UNWEIGHTED\_PEAKLEVEL>-13.8178799758273</UNWEIGHTED\_PEAKLEVEL>

<WEIGHTED\_LEVEL>-22.9999999893213</WEIGHTED\_LEVEL>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

<INITIAL\_PEAK>0.391756653785706</INITIAL\_PEAK>

<START\_TIME>0.5</START\_TIME>

<WORD>

<ORTHOGRAPHIC\_FORM>lax</ORTHOGRAPHIC\_FORM>

<PHONETIC\_FORM>l ˈ a kː s</PHONETIC\_FORM>

<START\_SAMPLE>24000</START\_SAMPLE>

<LENGTH>38221</LENGTH>

<UNWEIGHTED\_LEVEL>-30.6053004411253</UNWEIGHTED\_LEVEL>

<UNWEIGHTED\_PEAKLEVEL>-13.8178799758273</UNWEIGHTED\_PEAKLEVEL>

<WEIGHTED\_LEVEL>-22.9999999893213</WEIGHTED\_LEVEL>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

<START\_TIME>0.5</START\_TIME>

<PHONE>

<PHONETIC\_FORM>l</PHONETIC\_FORM>

<START\_SAMPLE>24000</START\_SAMPLE>

<LENGTH>5678</LENGTH>

<UNWEIGHTED\_LEVEL>-30.7662606426989</UNWEIGHTED\_LEVEL>

<UNWEIGHTED\_PEAKLEVEL>-20.6455503079323</UNWEIGHTED\_PEAKLEVEL>

<WEIGHTED\_LEVEL>-30.38761575187</WEIGHTED\_LEVEL>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

</PHONE>

<PHONE>

<PHONETIC\_FORM>a</PHONETIC\_FORM>

<START\_SAMPLE>29651</START\_SAMPLE>

<LENGTH>5802</LENGTH>

<UNWEIGHTED\_LEVEL>-23.3782651711565</UNWEIGHTED\_LEVEL>

<UNWEIGHTED\_PEAKLEVEL>-13.8178799758273</UNWEIGHTED\_PEAKLEVEL>

<WEIGHTED\_LEVEL>-22.9194342793721</WEIGHTED\_LEVEL>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

</PHONE>

<PHONE>

<PHONETIC\_FORM>kː</PHONETIC\_FORM>

<START\_SAMPLE>35481</START\_SAMPLE>

<LENGTH>11401</LENGTH>

<UNWEIGHTED\_LEVEL>-44.384599267382</UNWEIGHTED\_LEVEL>

<UNWEIGHTED\_PEAKLEVEL>-25.7247176788488</UNWEIGHTED\_PEAKLEVEL>

<WEIGHTED\_LEVEL>-44.1038734479661</WEIGHTED\_LEVEL>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

</PHONE>

<PHONE>

<PHONETIC\_FORM>s</PHONETIC\_FORM>

<START\_SAMPLE>46877</START\_SAMPLE>

<LENGTH>15423</LENGTH>

<UNWEIGHTED\_LEVEL>-40.163549624675</UNWEIGHTED\_LEVEL>

<UNWEIGHTED\_PEAKLEVEL>-25.895113327978</UNWEIGHTED\_PEAKLEVEL>

<WEIGHTED\_LEVEL>-39.8658784341023</WEIGHTED\_LEVEL>

<FREQUENCY\_WEIGHTING>C</FREQUENCY\_WEIGHTING>

<TIME\_WEIGHTING>0.1</TIME\_WEIGHTING>

</PHONE>

</WORD>

</SENTENCE>

</CHANNEL>

</SMA>