

# The draft specification for Sable version 0.2

## Mark Attribute

In addition to the specified attributes, every SABLE tag allows for a MARK attribute, as defined below:

MARK	Character-string identifier for this tag.
------	---

This attribute can be used to set an arbitrary mark at a given place in the text, so that, for example, an engine can report back to the calling application that it has reached the given location.

---

## Sable Version 0.2 Tagset

The following tags are defined for Sable Version 0.2:

---

## **EMPH**

*Description:* Set the emphasis of the contained text.

*Attributes:*

LEVEL	Defines the level of emphasis. Either a floating point number or integer greater than or equal to 0.0; or a descriptive term, with the following numerical interpretations:	
	Strong	2.0
	Moderate	1.0
	None	0.5
	Reduced	0.0
	Default is <i>Moderate</i> .	

*Properties:*

- **Container** element.
- Nestable.

*Example:*

"The leaders of <EMPH>Denmark</EMPH> and <EMPH>India</EMPH> meet on Friday."

---

## **BREAK**

*Description:* Sets an *intrasentential, prosodic* break at current position. (Contrasts with *DIV* which sets a text-structural break.)

*Attributes:*

LEVEL	Defines the break level. Either a floating point number or integer greater than or equal to 0.0; or a descriptive term, with the following numerical interpretations: <table border="1"><tr><td>Large</td><td>3.0</td></tr><tr><td>Medium</td><td>2.0</td></tr><tr><td>Small</td><td>1.0</td></tr><tr><td>None</td><td>0.0</td></tr></table> Default is <i>Medium</i> .	Large	3.0	Medium	2.0	Small	1.0	None	0.0
Large	3.0								
Medium	2.0								
Small	1.0								
None	0.0								
MSEC	A floating point number or integer greater than or equal to zero defining the length of the pause associated with this break. Default is that appropriate for a break of the defined LEVEL.								
TYPE	A punctuation symbol that represents (roughly) the kind of intonation contour to be associated with the utterance preceding the BREAK: currently proposed values are "?" ("sounds like a question"); "!" ("sounds like an exclamation"); "." ("sounds like a statement"); "," ("sounds as if there is more coming").								

*Properties:*

- **Empty** element.
- LEVEL of NONE (0.0) corresponds to a word boundary.
- Negative break LEVELs may be used to signify boundaries smaller than word boundaries (e.g. boundaries between elements of a clitic group).
- MSEC setting overrides whatever default silence may be associated with a BREAK of a particular LEVEL.

*Example:*

"Without style, <BREAK LEVEL="large"> Grace and I are in trouble."

---

## PITCH

*Description:* Sets properties associated with pitch of the enclosed region.

*Attributes:*

BASE	<p>Sets the bottom, or "base" line of the intonation. A specification in one of the following formats:</p> <ul style="list-style-type: none"><li>• A positive floating-point number representing an absolute Hz value</li><li>• A percentage value higher or lower than the current. Thus, for N a floating point number, the following are legal specifications:</li></ul> <table border="1" data-bbox="678 817 1077 963"><tr><td>N%</td><td>N percent above current</td></tr><tr><td>+N%</td><td>N percent above current</td></tr><tr><td>-N%</td><td>N percent below current</td></tr></table> <ul style="list-style-type: none"><li>• A descriptive term:</li></ul> <table border="1" data-bbox="494 1086 1260 1377"><tr><td>highest</td><td>highest possible value for engine/speaker/user</td></tr><tr><td>high</td><td>reasonable high value for engine/speaker/user</td></tr><tr><td>medium</td><td>reasonable medium value for engine/speaker/user</td></tr><tr><td>low</td><td>reasonable base value for engine/speaker/user</td></tr><tr><td>lowest</td><td>lowest possible value for engine/speaker/user</td></tr><tr><td>default</td><td>reset to default value for engine/speaker/user</td></tr></table> <p>Default is 0% (no change).</p>	N%	N percent above current	+N%	N percent above current	-N%	N percent below current	highest	highest possible value for engine/speaker/user	high	reasonable high value for engine/speaker/user	medium	reasonable medium value for engine/speaker/user	low	reasonable base value for engine/speaker/user	lowest	lowest possible value for engine/speaker/user	default	reset to default value for engine/speaker/user
N%	N percent above current																		
+N%	N percent above current																		
-N%	N percent below current																		
highest	highest possible value for engine/speaker/user																		
high	reasonable high value for engine/speaker/user																		
medium	reasonable medium value for engine/speaker/user																		
low	reasonable base value for engine/speaker/user																		
lowest	lowest possible value for engine/speaker/user																		
default	reset to default value for engine/speaker/user																		
MIDDLE	<p>Sets the middle, or "reference" line of the intonation. A specification in one of the following formats:</p> <ul style="list-style-type: none"><li>• A positive floating-point number representing an absolute Hz value</li><li>• A percentage value higher or lower than the current. Thus, for N a floating point number, the following are legal specifications:</li></ul> <table border="1" data-bbox="678 1825 1077 1971"><tr><td>N%</td><td>N percent above current</td></tr><tr><td>+N%</td><td>N percent above current</td></tr><tr><td>-N%</td><td>N percent below current</td></tr></table>	N%	N percent above current	+N%	N percent above current	-N%	N percent below current												
N%	N percent above current																		
+N%	N percent above current																		
-N%	N percent below current																		

	<ul style="list-style-type: none"> <li>A descriptive term: <table border="1" data-bbox="491 327 1262 622"> <tr><td>highest</td><td>highest available value for engine/speaker/user</td></tr> <tr><td>high</td><td>reasonable high value for engine/speaker/user</td></tr> <tr><td>medium</td><td>reasonable medium value for engine/speaker/user</td></tr> <tr><td>low</td><td>reasonable base value for engine/speaker/user</td></tr> <tr><td>lowest</td><td>lowest available value for engine/speaker/user</td></tr> <tr><td>default</td><td>reset to default value for engine/speaker/user</td></tr> </table> <p>Default is 0% (no change).</p> </li> </ul>	highest	highest available value for engine/speaker/user	high	reasonable high value for engine/speaker/user	medium	reasonable medium value for engine/speaker/user	low	reasonable base value for engine/speaker/user	lowest	lowest available value for engine/speaker/user	default	reset to default value for engine/speaker/user						
highest	highest available value for engine/speaker/user																		
high	reasonable high value for engine/speaker/user																		
medium	reasonable medium value for engine/speaker/user																		
low	reasonable base value for engine/speaker/user																		
lowest	lowest available value for engine/speaker/user																		
default	reset to default value for engine/speaker/user																		
RANGE	<p>Sets the "pitch range" of the intonation. A specification in one of the following formats:</p> <ul style="list-style-type: none"> <li>A positive floating-point number representing an absolute Hz value</li> <li>A percentage value higher or lower than the current. Thus, for N a floating point number, the following are legal specifications: <table border="1" data-bbox="671 990 1080 1142"> <tr><td>N%</td><td>N percent above current</td></tr> <tr><td>+N%</td><td>N percent above current</td></tr> <tr><td>-N%</td><td>N percent below current</td></tr> </table> </li> </ul> <ul style="list-style-type: none"> <li>A descriptive term: <table border="1" data-bbox="491 1256 1262 1552"> <tr><td>largest</td><td>largest available value for engine/speaker/user</td></tr> <tr><td>large</td><td>reasonable large value for engine/speaker/user</td></tr> <tr><td>medium</td><td>reasonable medium value for engine/speaker/user</td></tr> <tr><td>small</td><td>reasonable small value for engine/speaker/user</td></tr> <tr><td>smallest</td><td>smallest available value for engine/speaker/user</td></tr> <tr><td>default</td><td>reset to default value for engine/speaker/user</td></tr> </table> <p>Default is 0% (no change).</p> <p><i>Properties:</i></p> <ul style="list-style-type: none"> <li><b>Container</b> element.</li> <li>Nestable.</li> <li>BASE is interpreted as the baseline for the duration of the contained element. Thus in theories that have a declining baseline/topline, this would be interpreted as the base value for the <i>end</i> of the region.</li> </ul> <p>Note that unlike some other markup schemes, there is no explicit tag for</p> </li> </ul>	N%	N percent above current	+N%	N percent above current	-N%	N percent below current	largest	largest available value for engine/speaker/user	large	reasonable large value for engine/speaker/user	medium	reasonable medium value for engine/speaker/user	small	reasonable small value for engine/speaker/user	smallest	smallest available value for engine/speaker/user	default	reset to default value for engine/speaker/user
N%	N percent above current																		
+N%	N percent above current																		
-N%	N percent below current																		
largest	largest available value for engine/speaker/user																		
large	reasonable large value for engine/speaker/user																		
medium	reasonable medium value for engine/speaker/user																		
small	reasonable small value for engine/speaker/user																		
smallest	smallest available value for engine/speaker/user																		
default	reset to default value for engine/speaker/user																		

*downstep* or *declination*. These can, however, be implemented by appropriate resettings of BASE, MIDDLE and RANGE.

*Example:*

"Without his penguin, <PITCH BASE="-20%"> which he left at home, </PITCH> he could not enter the restaurant."

## **RATE**

*Description:* Set the speech rate of the contained text.

*Attributes:*

SPEED	Sets the speed. A specification in one of the following formats:															
	<ul style="list-style-type: none"><li>• A positive floating-point number representing an absolute rate in <i>words per minute</i></li><li>• A percentage value higher or lower than the current. Thus, for N a floating point number, the following are legal specifications:</li></ul> <table border="1"><tr><td>N%</td><td>N percent above current</td></tr><tr><td>+N%</td><td>N percent above current</td></tr><tr><td>-N%</td><td>N percent below current</td></tr></table> <ul style="list-style-type: none"><li>• Default is +0% (no change).</li><li>• A descriptive term:</li></ul> <table border="1"><tr><td>fastest</td><td>fastest available rate for engine/speaker/user</td></tr><tr><td>fast</td><td>reasonable fast rate for engine/speaker/user</td></tr><tr><td>medium</td><td>reasonable default rate for engine/speaker/user</td></tr><tr><td>slow</td><td>reasonable slow rate for engine/speaker/user</td></tr><tr><td>slowest</td><td>slowest available rate for engine/speaker/user</td></tr></table>	N%	N percent above current	+N%	N percent above current	-N%	N percent below current	fastest	fastest available rate for engine/speaker/user	fast	reasonable fast rate for engine/speaker/user	medium	reasonable default rate for engine/speaker/user	slow	reasonable slow rate for engine/speaker/user	slowest
N%	N percent above current															
+N%	N percent above current															
-N%	N percent below current															
fastest	fastest available rate for engine/speaker/user															
fast	reasonable fast rate for engine/speaker/user															
medium	reasonable default rate for engine/speaker/user															
slow	reasonable slow rate for engine/speaker/user															
slowest	slowest available rate for engine/speaker/user															

*Properties:*

- **Container** element.
- Nestable.

The term *words per minute* is to be understood rather loosely, and is probably language-dependent in its interpretation. For English speakers who are used to thinking in terms of orthographic words (i.e., words that are defined by surrounding whitespace in a text), the normal notion of words-per-minute (commonly used to

define rate of speech or rate of timing), should apply. For Japanese, where non-linguists are not used to thinking in terms of words, then another measure might be more appropriate: e.g. "bunsetsu-per-minute". The tag should really be interpreted as something like *LANGUAGE-APPROPRIATE-MINIMAL-INDEPENDENT-UNIT-PER-MINUTE*. For reasons that should be obvious this would, alas, not make a very good tag name.

*Example:*

"The address is <RATE SPEED="-20%"> 10 Main Street </RATE>."

## **VOLUME**

*Description:* Set the volume of the contained text.

*Attributes:*

LEVEL	<p>Defines the amplitude level. A specification in one of the following formats:</p> <ul style="list-style-type: none"> <li>• A floating-point number between 0 (= silence) and 1 (maximum volume for the engine)</li> <li>• A percentage value higher or lower than the current. Thus, for N a floating point number, the following are legal specifications:</li> </ul> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>N%</td> <td>N percent above current</td> </tr> <tr> <td>+N%</td> <td>N percent above current</td> </tr> <tr> <td>-N%</td> <td>N percent below current</td> </tr> </table> <ul style="list-style-type: none"> <li>• A descriptive term:</li> </ul> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>loudest</td> <td>loudest available volume for engine</td> </tr> <tr> <td>loud</td> <td>reasonable loud volume for engine</td> </tr> <tr> <td>medium</td> <td>reasonable medium volume for engine</td> </tr> <tr> <td>quiet</td> <td>quietest audible volume for engine</td> </tr> </table> <p>Default is <i>medium</i>.</p>	N%	N percent above current	+N%	N percent above current	-N%	N percent below current	loudest	loudest available volume for engine	loud	reasonable loud volume for engine	medium	reasonable medium volume for engine	quiet	quietest audible volume for engine
N%	N percent above current														
+N%	N percent above current														
-N%	N percent below current														
loudest	loudest available volume for engine														
loud	reasonable loud volume for engine														
medium	reasonable medium volume for engine														
quiet	quietest audible volume for engine														

*Properties:*

- **Container** element.
- Nestable.

This tag sets *only* the volume. Associated phonation changes are *not* implied. Thus *quiet* is not a whisper.

*Example:*

"Please speak more <VOLUME LEVEL="loud">loudly</VOLUME>."

---

## **AUDIO**

*Description:* Load and play an audio URL starting at the given point

*Attributes:*

SRC	URL of a document with an appropriate mime-type				
MODE	Either of the following: <table border="1"><tr><td>background</td><td>play as background to speech from this point on</td></tr><tr><td>insertion</td><td>play at this point, and when finished resume speaking</td></tr></table> Default is <i>insertion</i> .	background	play as background to speech from this point on	insertion	play at this point, and when finished resume speaking
background	play as background to speech from this point on				
insertion	play at this point, and when finished resume speaking				
LEVEL	A floating point number above 0.0. 1.0 is the same level as the original audio, 0.0 is silent. If not specified, the engine should scale the SRC's amplitude to be approximately that of the surrounding speech.				

*Properties:*

- **Empty** element.
- AUDIO instances that don't specify a SRC are ignored.

AUDIO is not a *required* tag in a SABLE-conformant system: it is recognized that not all engines/systems may be able to support it. Furthermore, it is acceptable if a system supports some audio types (e.g. *.au*, *.aiff*), but not others (e.g. *.wav*, *real audio*).

*Example:*

"Beethoven <AUDIO SRC="5th.au"> and Tchaikovsky <AUDIO SRC="1812.wav"> wrote good music!"



---

## **ENGINE**

*Description:* Substitute the DATA for the contained text if the system happens to be using the engine specified by ENGINE

*Attributes:*

ID	Identifier for the specific TTS engine.
DATA	Any character string to be substituted for the contained text.

*Properties:*

- **Container** element.
- Nestable.

The ENGINE tag allows one to select a specific text to be substituted for the contained text for a given synthesizer, if one happens to be using that synthesizer to read the given SABLE document. It also serves as a way to pass engine-specific controls to a given engine: this can be implemented by using the ENGINE tag to enclose *empty* text, and having the DATA be the control string. Engines other than the one specified by ID are free to ignore this tag, or may attempt to interpret it if they think they are able to.

*Example:*

```
"The <ENGINE ID="acme synth" DATA="wonderful, fantastic acme synthesizer">
Acme synthesizer</ENGINE>."
```

On an Acme system it says "wonderful, fantastic acme synthesizer". On other systems, it says just "Acme synthesizer".

---

## **MARKER**

*Description:* serves as an anchor point for a MARK that is not otherwise associated with another tag.

*Attributes:*

MARK	Character-string identifier for this tag.
------	---

*Properties:*

- **Empty** element.

MARK is an attribute of any SABLE tag. However, there may be instances where one wants to set a MARK, but where no specific tag is appropriate. MARKER should be used in such instances.

*Example:*

"Move the <MARKER MARK="mouse"> mouse to the top."

---

## **SABLE**

*Description:* Identifies the current document as being a SABLE document.

*No Attributes.*

*Properties:*

- **Container** element.
- *Not nestable.*

*Example:*

```
<SABLE>  
The text to be spoken goes here.  
It might include special tags.  
</SABLE>
```

---

## **PRON**

*Description:* Substitute the given pronunciation for the pronunciation that would normally be computed for the contained text.

*Attributes:*

IPA	Character string in <a href="#">Unicode IPA</a> describing the pronunciation to be used for the contained text.
SUB	Character string representing an attempt at "phonetic" spelling (in the language of the enclosing text) for the contained text.
ORIGIN	Identifier for the language of origin of the enclosed text, following the <a href="#">iso639</a> scheme.

*Properties:*

- **Container** element.
- *Not* nestable.

The IPA attribute is provided to allow for a precise phonetic rendering for the contained text. Recognizing that many developers may not be experienced with IPA, or other formal phonetic transcription schemes, SABLE provides an alternative method for specifying the desired pronunciation, using the SUB tag. Using this tag, an application may substitute for the contained text an attempt at phonetically spelling the text. Thus, one might want to specify the British pronunciation of "tomato" as follows:

```
<PRON SUB="tomahto">tomato</PRON>
```

Needless to say, it will depend upon the engine being used whether this will actually result in an appropriate pronunciation. This is unavoidable, and developers who desire both precision and portability should make the effort to learn IPA.

For languages that have a conventional, or semi-conventional "phonetic" writing scheme, in addition to, or part of their normal orthography, the SUB attribute is an appropriate way to include intended pronunciations transcribed in that scheme. For example, if one wants to specify the exact pronunciation of a Japanese personal name that is normally written in kanji (Chinese characters), one could specify its pronunciation with the SUB attribute of PRON, using a transcription in kana. Similarly for Korean, if one is using older-style mixed Korean text with Chinese characters, one might specify the pronunciation using the SUB attribute using hankul. For Mandarin Chinese, one might use either *pinyin* or *zhuyin fuhao* (Mandarin phonetic symbol set) in that field, though it is likely to be engine specific which one of these will be supported.

ORIGIN may be used to specify that the enclosed text comes from a particular language, and may be pronounced accordingly by the engine:

```
This is all rather <PRON ORIGIN=fr>passe</PRON>
```

If both IPA and SUB are specified, IPA takes precedence. PRON instances that specify no attribute are ignored.

## SAYAS

*Description:* Defines a way in which the contained region is to be said.

*Attributes:*

MODE	<p>Mode in which to say the contained text. Currently supported values are:</p> <table border="1"> <tr> <td>literal</td> <td>Literally read the string of characters in contained text, as appropriate to the language. E.g. "spelling out" in English, or character-by-character descriptions in Chinese.</td> </tr> <tr> <td>date</td> <td>Contained region is to be read as a date.</td> </tr> <tr> <td>time</td> <td>Contained region is to be read as a time.</td> </tr> <tr> <td>phone</td> <td>Contained region is to be read as a phone number.</td> </tr> <tr> <td>net</td> <td>Contained region is an internet address or handle (URL or e-mail address)</td> </tr> <tr> <td>postal</td> <td>Contained region is a postal address.</td> </tr> <tr> <td>currency</td> <td>Contained region is a currency amount.</td> </tr> <tr> <td>math</td> <td>Contained region is a mathematical expression.</td> </tr> <tr> <td>fraction</td> <td>Contained region is a fraction.</td> </tr> <tr> <td>measure</td> <td>Contained region is a measurement (e.g. 1km).</td> </tr> <tr> <td>ordinal</td> <td>Contained region is an ordinal number.</td> </tr> <tr> <td>cardinal</td> <td>Contained region is a cardinal number.</td> </tr> <tr> <td>name</td> <td>Contained region is a proper name.</td> </tr> </table>	literal	Literally read the string of characters in contained text, as appropriate to the language. E.g. "spelling out" in English, or character-by-character descriptions in Chinese.	date	Contained region is to be read as a date.	time	Contained region is to be read as a time.	phone	Contained region is to be read as a phone number.	net	Contained region is an internet address or handle (URL or e-mail address)	postal	Contained region is a postal address.	currency	Contained region is a currency amount.	math	Contained region is a mathematical expression.	fraction	Contained region is a fraction.	measure	Contained region is a measurement (e.g. 1km).	ordinal	Contained region is an ordinal number.	cardinal	Contained region is a cardinal number.	name	Contained region is a proper name.
literal	Literally read the string of characters in contained text, as appropriate to the language. E.g. "spelling out" in English, or character-by-character descriptions in Chinese.																										
date	Contained region is to be read as a date.																										
time	Contained region is to be read as a time.																										
phone	Contained region is to be read as a phone number.																										
net	Contained region is an internet address or handle (URL or e-mail address)																										
postal	Contained region is a postal address.																										
currency	Contained region is a currency amount.																										
math	Contained region is a mathematical expression.																										
fraction	Contained region is a fraction.																										
measure	Contained region is a measurement (e.g. 1km).																										
ordinal	Contained region is an ordinal number.																										
cardinal	Contained region is a cardinal number.																										
name	Contained region is a proper name.																										
MODETYPE	<p>Secondary specification further qualifying MODE. The following values are defined for the given MODE value:</p> <table border="1"> <tr> <td>DMY</td> <td><i>date</i> is in Day-Month-Year format.</td> </tr> <tr> <td>MDY</td> <td><i>date</i> is in Month-Day-Year format.</td> </tr> <tr> <td>YMD</td> <td><i>date</i> is in Year-Month-Day format.</td> </tr> <tr> <td>YM</td> <td><i>date</i> is in Year-Month format.</td> </tr> <tr> <td>MY</td> <td><i>date</i> is in Month-Year format.</td> </tr> <tr> <td>MD</td> <td><i>date</i> is in Month-Day format.</td> </tr> <tr> <td>HM</td> <td><i>time</i> is in Hour-Minute format.</td> </tr> <tr> <td>HMS</td> <td><i>time</i> is in Hour-Minute-Second format.</td> </tr> <tr> <td>EMAIL</td> <td><i>net</i> is an e-mail address.</td> </tr> <tr> <td>URL</td> <td><i>net</i> is a URL.</td> </tr> </table>	DMY	<i>date</i> is in Day-Month-Year format.	MDY	<i>date</i> is in Month-Day-Year format.	YMD	<i>date</i> is in Year-Month-Day format.	YM	<i>date</i> is in Year-Month format.	MY	<i>date</i> is in Month-Year format.	MD	<i>date</i> is in Month-Day format.	HM	<i>time</i> is in Hour-Minute format.	HMS	<i>time</i> is in Hour-Minute-Second format.	EMAIL	<i>net</i> is an e-mail address.	URL	<i>net</i> is a URL.						
DMY	<i>date</i> is in Day-Month-Year format.																										
MDY	<i>date</i> is in Month-Day-Year format.																										
YMD	<i>date</i> is in Year-Month-Day format.																										
YM	<i>date</i> is in Year-Month format.																										
MY	<i>date</i> is in Month-Year format.																										
MD	<i>date</i> is in Month-Day format.																										
HM	<i>time</i> is in Hour-Minute format.																										
HMS	<i>time</i> is in Hour-Minute-Second format.																										
EMAIL	<i>net</i> is an e-mail address.																										
URL	<i>net</i> is a URL.																										

*Properties:*

- **Container** element.
- Not nestable.

SAYAS instances with no specified MODE are ignored. A *date* without a specified MODETYPE interpreted as best as the engine can, possibly taking account of the locale (e.g. MDY in USA, DMY in most other countries).

*Example:*

At <SAYAS MODE="time">2pm</SAYAS> on <SAYAS MODE="date" MODETYPE="YM"> 98/3</SAYAS> Mike will send <SAYAS MODE="currency">\$4000</SAYAS> to <SAYAS MODE="net" MODETYPE="email">me@acme.com</SAYAS>.

---

## **LANGUAGE**

*Description:* Specify the language of the contained text.

*Attributes:*

ID	Identifier for the desired language, following the <a href="#">iso639</a> scheme, or dialect following the <a href="#">RFC1766</a> .
----	--

*Properties:*

- **Container** element.
- Nestable.

LANGUAGE can tag a region of any size. However, for most applications, and for most TTS systems, it will not be desirable to switch languages *within* a sentence.

Unless the SPEAKER is also specified, changing to a new language will result in using the default speaker for that language.

LANGUAGE instances without an associated ID specification will be ignored.

*Example:*

<LANGUAGE ID="en">Some text in English.</LANGUAGE>  
<LANGUAGE ID="de">Eine deutsche Satz.</LANGUAGE>

---

## **SPEAKER**

*Description:* Specify the speaker to use for the contained text.

*Attributes:*

GENDER	Gender for the desired speaker. Values are: <table border="1" data-bbox="616 562 930 656"><tr><td>male</td><td>Male speaker</td></tr><tr><td>female</td><td>Female speaker</td></tr></table> Default is the default gender for the engine.	male	Male speaker	female	Female speaker						
male	Male speaker										
female	Female speaker										
AGE	Description of the age of the desired speaker. Values are: <table border="1" data-bbox="572 808 970 1050"><tr><td>older</td><td>Older speaker</td></tr><tr><td>middle</td><td>Middle aged speaker</td></tr><tr><td>younger</td><td>Young adult</td></tr><tr><td>teen</td><td>Teenager</td></tr><tr><td>child</td><td>Child</td></tr></table> Default is the default "age" for the engine.	older	Older speaker	middle	Middle aged speaker	younger	Young adult	teen	Teenager	child	Child
older	Older speaker										
middle	Middle aged speaker										
younger	Young adult										
teen	Teenager										
child	Child										
NAME	Name of a speaker if a particular engine is being used.										

*Properties:*

- **Container** element.
- Nestable.

If NAME is specified, then it (may) override the other specifications of AGE and GENDER. If the system does not have a speaker with the given NAME, then the GENDER and AGE specifications (or their defaults) are used.

*Example:*

```
<SPEAKER GENDER="male" AGE="child">I'm a young boy!</SPEAKER>
```

---

## **DIV**

*Description:* Classifies the contained region as a division of type TYPE.

*Attributes:*

TYPE	Type of the division. Currently allowed values are:			
	<table border="1"><tr><td>sentence</td><td>Sentence</td></tr><tr><td>paragraph</td><td>Paragraph</td></tr></table>	sentence	Sentence	paragraph
sentence	Sentence			
paragraph	Paragraph			

*Properties:*

- **Container** element.
- Nestable.

Currently recommended types are only SENTENCE and PARAGRAPH. However, it is intended that DIV be used to support any reasonable division within a text. For instance, in the relatively unlikely event of a TTS system reading poetry, `<DIV TYPE=line>` and `<DIV TYPE=stanza>` might be reasonable. Similarly, in a transcription of a dialogue, DIV tags marking turn-taking might also be desirable. Indeed, such extensions to the values for TYPE are legal SABLE, though one cannot of course expect portability unless the TYPE is explicitly defined for the standard.

DIV instances with no specified TYPE are ignored.

*Example:*

```
<DIV TYPE="paragraph">  
<DIV TYPE="sentence" >  
Yesterday, Denmark and India announced an agreement of cultural exchange. </DIV>  
<DIV TYPE="sentence">  
Further talks will take place next month.  
</DIV>  
</DIV>
```



---

## Non-Standard Extensions to SABLE

SABLE is designed to function as a well-defined standard in which the same text will be handled consistently by multiple synthesizers. SABLE is also intended to function as a tool for research on speech synthesis and as a tool for innovation. As such, it is expected that research systems will support tags, attributes and attribute values not defined in the SABLE specification, and that SABLE text will be generated for specific systems which includes those tags and attributes. Where such extensions prove useful and become generally supported they can be proposed as an addition to the standard specification.

To clearly distinguish tags, attributes and attribute values that are non-standard they should include an "X-" prefix and optionally an engine identifier.

### Non-Standard Tag

A non-standard tag for providing an engine-specific pronunciation string could look like:

```
<X-ME-PRON PHON="i" DUR="120">
```

where ME is "My Engine" and the X-ME-PRON element inserts an "i" phoneme with a duration of 120msec understood by "My Engine". (Because the PHON and DUR attributes are embedded in a non-standard element, they are implicitly non-standard attributes.)

### Non-Standard Attribute

A non-standard attribute of a standard tag might look like:

```
<PRON X-ME-PHONES="ka:t">cat</X-ME-PRON>
```

or

```
<EMPH LEVEL="strong" X-PITCHACCENT="H*+L">word</EMPH>
```

The first example provides the pronunciation for "cat" in a format that is understood by "My Engine". Other synthesizers will ignore the attribute.

The second example includes both a standard attribute -- LEVEL -- and a non-standard attribute -- X-PITCHACCENT. A system that understands the non-standard attribute will apply the "H\*+L" accent when producing string emphasis on "word".

### Non-Standard Attribute Value

A non-standard attribute value might look like:

```
<DIV TYPE="x-dialog-close">...</DIV>
```

The "x-dialog-close" is a non-standard value of the standard TYPE attribute which is currently specified as being either "sentence" or "paragraph". This non-standard value could indicate that the contents of the element are the end of a dialog turn.

## Notes

If an engine gets a non-standard tag, attribute or attribute value in its input text that it does not know, it simply ignores it. For example, in the X-ME-PHONES example, a synthesizer that ignores the tag will try to say the word "cat".

Wherever possible, non-standard tags and elements should be designed so that output is not substantially impacted if ignored.

---

## CHANGES

### From Version 0.1

- LANGUAGE tag can now be used to tag a region of any size. (Just don't expect good results if you tag anything smaller than a sentence.)
- ID for language can specify dialect.
- ORIGIN is an attribute on PRON.
- Added a section on non-standard extensions to SABLE.
- Allow negative BREAK levels.
- "User" added to the specification of descriptive terms for attributes of PITCH.
- MED attribute of PITCH changed to "MIDDLE", in keeping with the fact that most other attribute names are full words.

### Unresolved from V0.1

- Should rate in words-per-minute be changed to something else?
- Should we add an attribute to EMPH allowing one to specify a particular list of specific syllables to be emphasized?
- Default for RATE (if the attribute SPEED is not specified) is +0% of the current RATE. Should this be something else?
- Various kinds of "dialogue" boundaries could be added as attributes of DIV. But does anyone have any concrete proposals as to what these should be? Otherwise it seems better for now to leave such divisions, if someone wants them, as a [non-standard attribute value](#).