Expressive Speech CORE

BARATINOO tags

Reference manual

Core version: 8.1

Reference	:	VOX32_Baratinoo_tags_reference_manual_8.1_1.0_EN			
Doc. version	:	1.0	**Status**	:	Release
Date	:	07/04/2017	**Diffusion**	:	restricted

Document review

Version	Date	Author	Verified by	Modification
1.0	07/04/2017	ER	PYJ	Creation for ES CORE 8.1. Phonemes are now in VOX349 doc.

Summary

1 Presentation

1.1 About this document

This document describes the syntax of proprietary tags supported by Voxygen Expressive Speech Core. Tags are commands that can be embedded into the textual input in order to alter the normal behaviour of the TTS system.

1.2 Terminology

Abbreviation	Description
TTS	*Text To Speech*
Baratinoo	*Abbreviation for the Voxygen Expressive Speech Core (Voxygen TTS engine)*

1.3 Reference documents

Reference	Document name
V0X349	Phonemes and visemes reference manual

2 Syntax

A tag begins with a "\", is followed by a keyword and ends with an optional parameter in curly brackets "{…}" or a "space" (normal space, tabulation or new line). The recognized keywords and their associated command are:

none	Comment
vox	Change current voice by name or rank
voice	Change current voice by features
lang	Change current language
phoneme	Phonetic input
break	Insert a pause
emph<	Start emphasis
emph>	Stop emphasis
sayas<	Start say-as
sayas>	End scope of say-as
spell	Spelling *(not implemented)*
pitch	Change pitch parameters
rate	Change rate parameter
timbre	Change timbre parameter
volume	Change volume
flush	Force processing
diacr	Diacritic processing parameter
audio	Insert a recorded audio file
audiomix	Mix speech synthesis with a recorded audio file
mark	Insert a marker
raw	Insert an item
computedpitch	Computed / intrinsic pitch contour
computedduration	Computed / intrinsic phoneme duration
token, w	Control word boundaries and grammatical tag
version	Synthesize Baratinoo version

There must not be any white spaces between "\" and the keyword or between the keyword and "{". Multiple parameters in "{…}" are separated by white spaces.
Unknown keywords followed by "{…}" are ignored, they are treated as text.

\

Comment. Text is ignored.

- `\{<comment>}`

\vox

Change current voice by name or rank.

- `\vox` : Reset to default voice.
- `\vox{<name>}` : Set current voice to that which is called <*name*>.
- `\vox{<number>}` : Set current voice to that which is ranked <*number*>. The first voice is ranked as number 0.

Input error occurs if no voice matches <*name*> or <*number*>.
The current modifications to pitch, rate, timbre and volume parameters are preserved across voice changes (the modifications are transposed to the new voice).

\voice

Change current voice by features.

- `\voice{<gender> <variant> <age> <name>}` : Set current voice to one that satisfies the requested features and that can speak the current language of the text. One or more attributes may be specified.

Accepted values for <*gender*> are:

`male`	a voice characterised as “male”
`female`	a voice characterised as “female”
`neutral`	a voice characterised as “neutral”

The value of the <*variant*> parameter must be a positive integer (greater than zero), and <*age*> must be a positive integer or zero.
Input error occurs if no voice matches the requested features for the current language.
The current modifications to pitch, rate, timbre and volume parameters are preserved across voice changes (the modifications are transposed to the new voice).

\lang

Change current language.

- `\lang{<language>}` : Set current language to <*language*>.

Valid values of the <*language*> attribute are IETF BCP47 language tags. If the current voice cannot speech the specified language, a new voice is selected.
Input error occurs if no available voice can speak the given <*language*>.

\phoneme

Phonetic input.

- `\phoneme{<phonetic input>}` : Insert a phonetic sequence.
- `\phoneme{<phonetic input> <textual input>}` : Insert a phonetic sequence and its corresponding orthographic string.

Phonetic input is a list of phonetic characters (underscore separated). See the VOX340 document for the phonetic alphabets of each language. Characters must be uppercase for a French or Spanish voice and lowercase for an English voice.

\break

Insert a pause.

- `\break` : insert a `medium` level silence.
- `\break{`<*size*>`}` : Insert a silence defined by <*size*>.

Accepted values for <*size*> are:

`x-weak`	≅ 50ms
`weak`	≅ 100ms
`medium`	≅ 500ms
`Strong`	≅ 1s
`x-strong`	≅ 2s
`none`	No silence
<*number*>	silence in millisecond
<*number*>`ms`	silence in millisecond
<*number*>`s`	silence in second

where *number* must be a signed or unsigned positive value or zero.
Break duration has an upper limit of 60 seconds.

\emph<

Start emphasis.

- `\emph<` : Start emphasis at `moderate` level
- `\emph<{`<*mode*>`}` : Start emphasis at <*mode*> level

Accepted values for <*mode*> are:

`reduced`
`moderate`
`strong`
	say the text with the designated emphasis

`none`	prevent the speech synthesis processor from emphasizing words that might otherwise be emphasized

\emph>

Stop emphasis.

- `\emph>`

\sayas<

Start say-as.

- `\sayas<{`<*interpret-as*> <*format*> <*detail*>`}` : Set start say-as mode.

The <*interpret-as*> attribute is required. The <*format*> and <*detail*> attributes are optional strings. Accepted values for <*interpret-as*> are:

`date`
`time`
`telephone`
`characters`
`cardinal`
`ordinal`
	Indicates the type of text construct contained within the say-as scope, as described in the W3C Note 26 May 2005 on say-as attribute values: http://www.w3.org/TR/ssml-sayas/

\sayas>

End scope of say-as.

- \sayas>

\spell

Spelling.

- \spell{<*word*>} : <*word*> is spelt out.

This command is currently not implemented (command is ignored).

\pitch

Change pitch parameters.

- \pitch : Reset pitch parameters for current voice.
- \pitch{<*baseline*>} : Set pitch baseline to <*baseline*>.
- \pitch{<*baseline*> <*range*>} : Set pitch baseline to <*baseline*> and pitch range to <*range*>.

More than 99% of pitch values are in the interval [<*baseline*>;<*baseline*>+<*range*>]. <*baseline*> is the lower bound of the pitch in Hertz (limited to [30;300]), range is the degree of additional pitch in Hertz (limited to [0;300]).

Accepted values for <*baseline*> and <range> are:

x-low	50% of default
low	75% of default
medium	100% of default
default	initial baseline/range for current voice
high	133% of default
x-high	200% of default
+/-<*number*>%	relative percentage
+/-<*number*>st	relative change in semitones
+/-<*number*>Hz	relative change in Hertz
<*number*>Hz	absolute value in Hertz

The current modification to pitch is preserved across voice changes (the modification is transposed to the new voice).

\rate

Change rate parameter.

- \rate : Reset average articulation and pause rate for current voice.
- \rate{<*rate*> <*rate_subject*>} : Set <rate_subject> rate to <*rate*>

Accepted values for <*rate*> are:

x-slow	50% of default
slow	75% of default
medium	100% of default
default	initial rate for current voice
fast	125% of default
x-fast	150% of default
+/-<*number*>%	relative percentage
+/-<*number*>	relative change

<*number*> <*number*>%	multiplier (positive, not zero) of `default`

Accepted values for <rate_subject> are :

`articulation`	Rate is applied to speech
`pause`	Rate is applied to pauses originated from the TTS engine (\break values are not affected)
`all`	Rate is applied to both speech and pauses (default value)

The current modification to rate is preserved across voice changes (the modification is transposed to the new voice).

`\timbre`

Change timbre parameter.

- `\timbre` : Reset timbre for current voice.
- `\timbre{`<*timbre*>`}` : Set timbre coefficient to <*timbre*>.

Accepted values for <*timbre*> are:

+<*number*>%	relative increment in %
-<*number*>%	relative decrement in %
<*number*>	multiplier (positive, not zero) of initial value

The current modification to timbre is preserved across voice changes (the modification is transposed to the new voice).

`\volume`

Change volume.

- `\volume` : Reset volume for current voice.
- `\volume{`<*volume*>`}:` Set current volume to <*volume*>.

Accepted values for <*volume*> are:

`silent`	-∞dB
`x-soft`	-12dB relative to `default`
`soft`	-6dB relative to `default`
`medium`	+0dB relative to `default`
`default`	initial volume for current voice
`loud`	+6dB relative to `default`
`x-loud`	+12dB relative to `default`
<*number*>	absolute value in interval [0;100]
+/-<*number*>dB	relative change in dB
+/-<*number*>%	relative percentage on linear scale
+/-<*number*>	relative change on linear scale

The current modification to volume is preserved across voice changes (the modification is transposed to the new voice).

\flush

Force processing.

- `\flush`

This command forces any input buffered up to this point to be processed as if an end-of-file had been reached.

\diacr

Diacritic processing parameter.

- `\diacr{acc}` : text is to be processed as if diacritics are present, if any.
- `\diacr{non}` : text is to be processed as if diacritics may have been removed.
- `\diacr{default}` : use the system's default behaviour (voice specific) for text processing with respect to the presence or absence of diacritics.

\audio

Insert a recorded audio file.

- `\audio{`*<src> <soundLevel> <fadein> <fadeout> <fadelevel> <clipBegin> <clipEnd> <speed> <repeatCount> <repeatDur> <tempo> <fadeinAttack> <fadeinRelease> <fadeoutAttack> <fadeoutRelease>*`}` : Play the audio file located at *<src>*. One or more attributes may be specified.

A section of the signal source may be selected by the *<clipBegin>* and *<clipEnd>* attributes. It may then by modified with <soundLevel>dB of gain between points specified by *<fadein>* milliseconds or seconds from the media start and *<fadeout>* milliseconds or seconds from the media end, and with gain *<fadelevel>* elsewhere. The `tempo` attribute can be used to speed up or slow down the rate of the audio file without changing the pitch level.
The default duration (20ms) of the transitional periods at either side of the fade-in and the fade-out periods may be changed. The *<fadeinAttack>* and *<fadeinRelease>* attributes are respectively for the beginning side and the end side of the fade-in period. The same with the *<fadeoutAttack>* and *<fadeoutRelease>* attributes for the fade-out period. The sum of the attack and release values for a period must not be greater than the duration of the period.
The tag is ignored if the file does not exist or its media type is not supported.
URI schemes other than `file:`, `http:`, and `ftp:` in the value of the *<src>* attribute are not supported.
Only audio/x-wav files may be mono or stereo. Other media types must contain only one signal channel (mono).

Accepted values for *<fadelevel>* and *<soundLevel>* are:

+*<number>*`dB`	Gain in dB. Maximum value is +12dB.
-*<number>*`dB`	Attenuation in dB. Minimum value is -90dB.

Accepted values for *<fadein>*, *<fadeout>*, *<fadeinAttack>, <fadeinRelease>, <fadeoutAttack>, <fadeoutRelease>, <clipBegin>*, *<clipEnd>* and *<repeatDur>* are:

<number>`s`	Time in seconds.
<number>`ms`	Time in milliseconds.

Accepted values for *<speed>* are:

<number>`%`	Percentage of the speed of the original waveform.

Accepted values for <*repeatCount*> are:

<*number*>	Number of iterations of the media to render. A fractional value describes a portion of the rendered media.

Accepted values for <*tempo*> are:

<*number*>%	Percentage of the speed of the original waveform.

\audiomix

Mix speech synthesis with a recorded audio file.

- `\audiomix{`<*src*> <*soundLevel*> <*fadein*> <*fadeout*> <*fadelevel*> <*clipBegin*> <*clipEnd*> <*speed*> <*tempo*> <*fadeinAttack*> <*fadeinRelease*> <*fadeoutAttack*> <*fadeoutRelease*>`}` : Mix synthesis with the audio file located at <*src*>. Zero or more attributes may be specified.

The mix stops when another `\audiomix` tag (empty or not) is encountered.
If the audio signal is longer than the synthesized text, then the audio file is truncated. If the audio signal is shorter than the synthesized text, then the system repeatedly reads the file.
Attributes of the `\audiomix` tag have the same meaning and restrictions as those of the `\audio` tag, except that the default fade attack and release durations is 480ms.

\mark

Insert a marker.

- `\mark :` Insert a marker tagged with an integer. The integer is increased at each new empty mark. The first nameless marker is tagged with number 1.
- `\mark{`<*name*>`}` or `\mark{`<*name*>` sync}` : Insert a synchronisation marker tagged with <*name*>.
- `\mark{`<*name*>` wait} :` Insert a wait marker tagged with <*name*>.

A wait marker allows rendering of the audio signal to be deferred until the duration of the immediately following content has been determined. The end of the content whose duration is to be determined is marked by either `\flush` or a `\mark`, of any type, that bears the same name (case-sensitive). For example:

"Text before…\mark{foo wait} piece of text \mark{foo} text after…"

When Baratinoo processes the above markup, notification is first made by a 'WAITMARKER' event with the name 'foo' and the duration in samples of the rendered content "piece of text". Then the signal for the "piece of text" is sent, and finally, notification is made by a 'MARKER' event with the name 'foo', signalling the end of the marked sequence.
It is possible to set another `\mark`, of any type, before the end of the deferred content is encountered. Examples are:

"\mark{foo wait} piece of text \mark{another} containing another marker \mark{foo}"

"\mark{foo wait} piece of text \mark{another wait} with another embedded \mark{another} wait marker sequence \mark{foo}"

"\mark{foo wait} piece of text \mark{another wait} with interleaved \mark{foo} wait marker sequences \mark{another}"

\raw

Insert an item.

- `\raw{<class number>}`
- `\raw{<class number> <data>}`

<*class number*>: class of the item (integer value)
<*data*>: a string of byte values. Each byte must be written in hexadecimal as two characters and there must be a space character between bytes. Example: 00 0A BB 0D is the string for the data sequence 0x00, 0x0A, 0xBB, 0x0D. <*data*> may be omitted.

\computedpitch

Computed / intrinsic pitch contour.

- `\computedpitch{on}`: apply pitch contour computed by system
- `\computedpitch{off}`: intrinsic pitch contour
- `\computedpitch` or `\computedpitch{default}` : reset to default behaviour of voice

Pitch contour mode is reset if there is a voice change.

\computedduration

Computed / intrinsic phoneme duration.

- `\computedduration{on}` : apply phoneme duration computed by system
- `\computedduration{off}` : intrinsic phoneme duration
- `\computedduration` or `\computedduration{default}` : reset to default behaviour of voice

Phoneme duration mode is reset if there is a voice change.

\token, \w

Control word boundaries and grammatical tag.

- `\token{`<*role*> <*text*>`}` : associate grammatical tag <*role*> with the given <*text*> (token).

Tag “w” is an alias for “token”

The token element can be used to:

- indicate its content is a token and to eliminate token (word) segmentation ambiguities of the synthesis processor.

- set a specific grammatical tag for the content text.

\version

Synthesize Baratinoo version.

- `\version :` the Baratinoo version is synthesized.