

Text-to-Speech - RemoteSpeech interface

Overview

[SmartBody](#) is responsible for acquiring the audio that is associated with a character utterance. This audio is then preferably played back by the game engine, but it can also be played back by SmartBody itself. To acquire the audio, SmartBody can either:

- read in an existing prerecorded speech audio file
- send a request to a text-to-speech engine

This document focuses on the latter.

SmartBody will send a [RemoteSpeechCmd](#) message to the [TtsRelay](#) module, requesting for a line of text to be converted into audio. The message contains what voice to use and where to put the generated file. TtsRelay will send back a [RemoteSpeechCmd](#) message, containing the exact file location, a viseme schedule with detailed timing information for lip-synching and word boundary timing information for synchronization of nonverbal behavior as specified through BML.

TTS Engines

Rhetorical (RVoiceRelay)

Voice Codes:

```
set character doctor voice remote M021 <- Saso Doctor's voice
set character elder voice remote M009 <- Saso Elder's voice
```

Cerevoice (CerevoiceRelay)

Voice Codes:

```
set character doctor voice remote star
set character doctor voice remote katherine
set character doctor voice remote starconv
```

Cepstral (CepstralRelay)

MSSpeech (MSSpeechRelay)

Voice Codes:

```
set character doctor voice remote BradVoice
```

Festival (FestivalRelay)

Voice Codes:

```
set character doctor voice remote BradVoice
```

RemoteSpeech Interface

To trigger a TTS call:

```
sbm bml char doctor speech "Hello world.  Testing Text to Speech"
```

Sent by Smartbody to TTS Engine:

```
RemoteSpeechCmd speak doctor 1 M021 ../../data/cache/audio/utt_20110528_175743_doctor_1.aiff
<?xml version="1.0" encoding="UTF-8"?>
<speech type="text/plain">
  Hello world.  Testing Text to Speech
</speech>
```

RVoiceRelay Example:

Actual message sent to Rhetorical:

```
<?xml version="1.0" encoding="UTF-8"?>
<speech type="text/plain">Hello world.  Testing Text to Speech</speech>
```

Sent by TTS Engine:

```
RemoteSpeechReply doctor 2 OK:
<?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\saso\core\beavin\...\data\cache\audio\utt_20110528_180148_doctor_2.aiff"/>
  <viseme start="0.0" type="_"/>
  <word end="0.4049886621315193" start="0.049977324263038546">
    <viseme start="0.049977324263038546" type="Ih"/>
    <viseme start="0.14498866213151929" type="Ih"/>
    <viseme start="0.2" type="D"/>
    <viseme start="0.2549659863945578" type="OW"/>
  </word>
  <word end="0.8099773242630386" start="0.4049886621315193">
    <viseme start="0.4049886621315193" type="OO"/>
    <viseme start="0.5199546485260771" type="Er"/>
    <viseme start="0.5849886621315192" type="R"/>
    <viseme start="0.6649886621315193" type="D"/>
    <viseme start="0.7699773242630386" type="D"/>
  </word>
  <viseme start="0.8099773242630386" type="_"/>
  <viseme start="0.860498866213152" type="_"/>
  <viseme start="1.060498866213152" type="_"/>
  <word end="1.5854875283446712" start="1.1104761904761904">
    <viseme start="1.1104761904761904" type="D"/>
    <viseme start="1.1574603174603175" type="Ih"/>
    <viseme start="1.2354648526077097" type="Z"/>
    <viseme start="1.3304761904761904" type="D"/>
    <viseme start="1.3824943310657596" type="Ih"/>
    <viseme start="1.4374603174603175" type="NG"/>
  </word>
  <word end="1.8724716553287981" start="1.5854875283446712">
    <viseme start="1.5854875283446712" type="D"/>
    <viseme start="1.6424943310657596" type="Ih"/>
    <viseme start="1.7174603174603174" type="KG"/>
    <viseme start="1.7674829931972789" type="Z"/>
    <viseme start="1.8374603174603175" type="D"/>
  </word>
  <word end="1.927482993197279" start="1.8724716553287981">
    <viseme start="1.8724716553287981" type="D"/>
    <viseme start="1.9024943310657596" type="Ih"/>
  </word>
  <word end="2.408480725623583" start="1.927482993197279">
    <viseme start="1.927482993197279" type="Z"/>
    <viseme start="2.0224943310657597" type="BMP"/>
    <viseme start="2.1174603174603175" type="EE"/>
    <viseme start="2.207482993197279" type="j"/>
  </word>
  <viseme start="2.408480725623583" type="_"/>
  <viseme start="2.4584580498866213" type="_"/>
</speak>
```

MSSpeechRelay Example:

Actual message sent to MSSpeech:

```
<speak version="1.0" xml:lang="en-US">Hello world.  Testing Text to Speech .</speak>
```

(note the added period at the end)

Sent by TTS Engine:

```

RemoteSpeechReply doctor 1 OK:
<?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\vhtoolkit\data\cache\audio\utt_20110528_180527_doctor_1.wav"/>
  <viseme start="0" type="_"/>
  <viseme start="0.003" type="Oh"/>
  <viseme start="0.047" type="Ih"/>
  <viseme start="0.098" type="D"/>
  <viseme start="0.258" type="Oh"/>
  <viseme start="0.418" type="Oh"/>
  <viseme start="0.479" type="Er"/>
  <viseme start="0.54" type="R"/>
  <viseme start="0.601" type="D"/>
  <viseme start="0.695" type="D"/>
  <viseme start="0.745" type="_"/>
  <viseme start="1.367" type="_"/>
  <viseme start="1.37" type="D"/>
  <viseme start="1.461" type="Ih"/>
  <viseme start="1.546" type="Z"/>
  <viseme start="1.6" type="D"/>
  <viseme start="1.654" type="Ih"/>
  <viseme start="1.729" type="KG"/>
  <viseme start="1.804" type="D"/>
  <viseme start="1.9" type="Ih"/>
  <viseme start="2.022" type="KG"/>
  <viseme start="2.087" type="Z"/>
  <viseme start="2.16" type="D"/>
  <viseme start="2.233" type="D"/>
  <viseme start="2.297" type="Oh"/>
  <viseme start="2.341" type="Z"/>
  <viseme start="2.425" type="BMP"/>
  <viseme start="2.509" type="Ih"/>
  <viseme start="2.606" type="j"/>
  <viseme start="2.73" type="_"/>
</speak>

```

CerevoiceRelay Example:

Actual text sent to cerevoice engine:

```

<?xml version="1.0" encoding="UTF-8"?>
<speech type="text/plain">Hello world.  Testing Text to Speech </speech>

```

(note the space, also note that cerevoicerelay removes punctuation because of an apparent bug in cerevoice)

Sent by TTS Engine (CerevoiceRelay Example) (hand-formatted):

RemoteSpeechReply doctor 1 OK:

```
<?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\saso\data\cache\audio\utt_20110621_192933_doctor_1.wav"/>
  <viseme start="0.000000" type="_"/>
  <mark name="spl:T0" time="0.010975"/>
  <mark name="spl:T1" time="0.010975"/>
  <word end="2.468209" start="0.010975">
    <viseme start="0.010975" type="Ih"/>
    <viseme start="0.090975" type="Ih"/>
    <viseme start="0.120952" type="D"/>
    <viseme start="0.231157" type="Oh"/>
    <viseme start="0.430088" type="OO"/>
    <viseme start="0.527008" type="Er"/>
    <viseme start="0.663673" type="D"/>
    <viseme start="0.723719" type="D"/>
    <viseme start="0.768662" type="D"/>
    <viseme start="0.848662" type="Ih"/>
    <viseme start="0.948662" type="Z"/>
    <viseme start="1.113696" type="D"/>
    <viseme start="1.173651" type="Ih"/>
    <viseme start="1.223510" type="NG"/>
    <viseme start="1.357624" type="D"/>
    <viseme start="1.431655" type="Ih"/>
    <viseme start="1.511610" type="KG"/>
    <viseme start="1.566621" type="Z"/>
    <viseme start="1.636644" type="D"/>
    <viseme start="1.696644" type="Oh"/>
    <viseme start="1.833379" type="Z"/>
    <viseme start="1.958231" type="BMP"/>
    <viseme start="2.028209" type="EE"/>
    <viseme start="2.188209" type="j"/>
  </word>
  <mark name="spl:T2" time="2.468209"/>
  <mark name="spl:T3" time="2.468209"/>
  <viseme start="2.468209" type="_"/>
</speak>
```

new output 11/7/11:

RemoteSpeechReply doctor 1 OK:

```
<?xml version="1.0" encoding="UTF-8"?>
<speak>
<soundFile name="d:\edwork\saso\core\TtsSpeechRelay\bin\data\cache\audio\utt_20110528_175743_doctor_1.wav.wav"/>
  <viseme start="0.000000" type="_"/>
  <mark name="spl:T0" time="0.010975"/>
  <mark name="spl:T1" time="0.010975"/>
  <word end="0.353100" start="0.010975">
    <viseme start="0.010975" type="Ih"/>
    <viseme start="0.099709" type="Ih"/>
    <viseme start="0.126943" type="D"/>
    <viseme start="0.252789" type="Oh"/>
  </word>
  <mark name="spl:T2" time="0.353100"/>
  <mark name="spl:T3" time="0.353100"/>
  <word end="0.762222" start="0.353100">
    <viseme start="0.353100" type="OO"/>
    <viseme start="0.446472" type="Er"/>
    <viseme start="0.532245" type="D"/>
    <viseme start="0.602222" type="D"/>
  </word>
  <mark name="spl:T4" time="0.762222"/>
  <mark name="spl:T5" time="0.762222"/>
  <viseme start="0.762222" type="_"/>
  <mark name="spl:T6" time="0.000000"/>
  <mark name="spl:T7" time="0.962222"/>
  <viseme start="0.000000" type="_"/>
  <mark name="spl:T8" time="1.162222"/>
  <mark name="spl:T9" time="1.162222"/>
  <word end="1.617595" start="1.162222">
    <viseme start="1.162222" type="D"/>
    <viseme start="1.254784" type="Ih"/>
    <viseme start="1.340280" type="Z"/>
    <viseme start="1.419229" type="D"/>
    <viseme start="1.479229" type="Ih"/>
    <viseme start="1.509215" type="NG"/>
  </word>
  <mark name="spl:T10" time="1.617595"/>
  <mark name="spl:T11" time="1.617595"/>
  <word end="2.077460" start="1.617595">
    <viseme start="1.617595" type="D"/>
    <viseme start="1.747483" type="Ih"/>
    <viseme start="1.827483" type="KG"/>
    <viseme start="1.927438" type="Z"/>
    <viseme start="2.037460" type="D"/>
  </word>
  <mark name="spl:T12" time="2.077460"/>
  <mark name="spl:T13" time="2.077460"/>
  <word end="2.227483" start="2.077460">
    <viseme start="2.077460" type="D"/>
    <viseme start="2.197460" type="Ih"/>
  </word>
  <mark name="spl:T14" time="2.227483"/>
  <mark name="spl:T15" time="2.227483"/>
  <word end="2.847438" start="2.227483">
    <viseme start="2.227483" type="Z"/>
    <viseme start="2.347483" type="BMP"/>
    <viseme start="2.427438" type="EE"/>
    <viseme start="2.587438" type="j"/>
  </word>
  <mark name="spl:T16" time="2.847438"/>
  <mark name="spl:T17" time="2.847438"/>
  <viseme start="2.847438" type="_"/>
</speak>
```

FestivalRelay example:

Actual text sent to Festival:

```
<?xml version="1.0" encoding="UTF-8"?>
<speech type="text/plain">Hello world.  Testing Text to Speech </speech>
```

(note that this gets edited by FestivalRelay and eventually gets sent out as 'Helloworld.TestingTexttoSpeech')

Sent by TTS Engine (FestivalRelay Example) (hand-formatted):

```
RemoteSpeechReply doctor 7 OK:
<?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\vh toolkit\bin\FestivalRelay\data\cache\festival\utt_20110722_185051_doctor_7.wav"
/>
  <viseme start="0.000000" type="_" />
  <mark name="T0" time="0.080000"/>
  <word end="0.640000" start="0.080000" >
    <viseme start="0.080000" type="Ih" />
    <viseme start="0.160000" type="Ih" />
    <viseme start="0.240000" type="D" />
    <viseme start="0.320000" type="Oh" />
    <viseme start="0.400000" type="Er" />
    <viseme start="0.440000" type="R" />
    <mark name="T1" time="0.480000"/>
  </word>
  <mark name="T2" time="0.080000"/>
  <word end="0.640000" start="0.080000" >
    <viseme start="0.480000" type="D" />
    <viseme start="0.560000" type="D" />
    <mark name="T3" time="0.640000"/>
  </word>
  <mark name="T4" time="0.640000"/>
  <word end="0.880000" start="0.640000" >
    <viseme start="0.640000" type="D" />
    <viseme start="0.720000" type="Ao" />
    <viseme start="0.800000" type="D" />
    <mark name="T5" time="0.880000"/>
  </word>
  <mark name="T6" time="0.880000"/>
  <word end="2.160000" start="0.880000" >
    <viseme start="0.880000" type="D" />
    <viseme start="0.960000" type="Ih" />
    <viseme start="1.040000" type="Z" />
    <viseme start="1.120000" type="D" />
    <viseme start="1.200000" type="Ih" />
    <viseme start="1.280000" type="NG" />
    <viseme start="1.360000" type="D" />
    <viseme start="1.440000" type="Ih" />
    <viseme start="1.520000" type="KG" />
    <viseme start="1.600000" type="Z" />
    <viseme start="1.680000" type="D" />
    <viseme start="1.760000" type="Ao" />
    <viseme start="1.840000" type="Z" />
    <viseme start="1.920000" type="BMP" />
    <viseme start="2.000000" type="EE" />
    <viseme start="2.080000" type="j" />
    <mark name="T7" time="2.160000"/>
  </word>
  <viseme start="2.160000" type="_" />
</speak>
```

new output 11/7/11:

```
RemoteSpeechReply doctor 1 OK: <?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\saso\core\TtsSpeechRelay\bin\data\cache\audio\utt_20110528_175743_doctor_1.wav"/>
<mark name="T0" time="0.210000"/>
<word end="0.795159" start="0.210000" >
  <viseme start="0.367043" type="D" />
  <viseme start="0.704177" type="D" />
  <viseme start="0.756153" type="D" />
  <mark name="T1" time="0.795159"/>
</word>
<mark name="T2" time="0.795159"/>
<word end="1.013328" start="0.795159" >
  <viseme start="0.795159" type="D" />
  <viseme start="0.953081" type="D" />
  <mark name="T3" time="1.013328"/>
</word>
<mark name="T4" time="1.013328"/>
<word end="2.455301" start="1.013328" >
  <viseme start="1.013328" type="D" />
  <viseme start="1.210314" type="Z" />
  <viseme start="1.282180" type="D" />
  <viseme start="1.358164" type="Ih" />
  <viseme start="1.394886" type="NG" />
  <viseme start="1.452691" type="D" />
  <viseme start="1.608044" type="KG" />
  <viseme start="1.690684" type="Z" />
  <viseme start="1.788436" type="D" />
  <viseme start="1.962315" type="Z" />
  <viseme start="2.065681" type="BMP" />
  <viseme start="2.312202" type="j" />
  <mark name="T5" time="2.455301"/>
</word>
</speak>
```

NPCEditor/NVBG Example

Utterance #20 in Toolkit

RemoteSpeechCmd sent by SBM

```

RemoteSpeechCmd speak brad 1 BradVoiceFestival ../../data/cache/audio/utt_20110809_151922_brad_1.aiff
<?xml version="1.0" encoding="utf-16"?>
<speech id="spl" ref="tech_sapiTTS" type="application/ssml+xml">
  <mark name="T0" />SAPI
  <mark name="T1" /><mark name="T2" />is
  <mark name="T3" /><mark name="T4" />a
  <mark name="T5" /><mark name="T6" />speech
  <mark name="T7" /><mark name="T8" />and
  <mark name="T9" /><mark name="T10" />text
  <mark name="T11" /><mark name="T12" />to
  <mark name="T13" /><mark name="T14" />speech
  <mark name="T15" /><mark name="T16" />interface
  <mark name="T17" /><mark name="T18" />by
  <mark name="T19" /><mark name="T20" />Microsoft.
  <mark name="T21" /><mark name="T22" />I
  <mark name="T23" /><mark name="T24" />use
  <mark name="T25" /><mark name="T26" />it
  <mark name="T27" /><mark name="T28" />to
  <mark name="T29" /><mark name="T30" />be
  <mark name="T31" /><mark name="T32" />able
  <mark name="T33" /><mark name="T34" />to
  <mark name="T35" /><mark name="T36" />talk
  <mark name="T37" /><mark name="T38" />to
  <mark name="T39" /><mark name="T40" />you.
  <mark name="T41" />
</speech>

```

Festival example

```

RemoteSpeechReply brad 2 OK: <?xml version="1.0" encoding="UTF-8"?>
<speech>
  <soundFile name="d:\edwork\vhtoolkit\bin\FestivalRelay\data\cache\festival\utt_20110809_152521_brad_2.wav"/>
  <viseme start="0.000000" type="_" />
  <mark name="T0" time="0.080000"/>
  <word end="0.400000" start="0.080000" >
    <viseme start="0.080000" type="Z" />
    <viseme start="0.160000" type="Ao" />
    <viseme start="0.240000" type="BMP" />
    <viseme start="0.320000" type="EE" />
    <mark name="T1" time="0.400000"/>
  </word>
  <mark name="T2" time="0.400000"/>
  <word end="0.560000" start="0.400000" >
    <viseme start="0.400000" type="Ih" />
    <viseme start="0.480000" type="Z" />
    <mark name="T3" time="0.560000"/>
  </word>
  <mark name="T4" time="0.560000"/>
  <word end="0.640000" start="0.560000" >
    <viseme start="0.560000" type="Ih" />
    <mark name="T5" time="0.640000"/>
  </word>
  <mark name="T6" time="0.640000"/>
  <word end="0.960000" start="0.640000" >
    <viseme start="0.640000" type="Z" />
    <viseme start="0.720000" type="BMP" />
    <viseme start="0.800000" type="EE" />
    <viseme start="0.880000" type="j" />
    <viseme start="0.960000" type="_" />
    <mark name="T7" time="1.040000"/>
  </word>
  <mark name="T8" time="1.040000"/>
  <word end="1.280000" start="1.040000" >
    <viseme start="1.040000" type="Ih" />
    <viseme start="1.120000" type="NG" />
    <viseme start="1.200000" type="D" />
    <mark name="T9" time="1.280000"/>
  </word>

```



```

</word>
<mark name="T10" time="1.280000"/>
<word end="1.680000" start="1.280000" >
  <viseme start="1.280000" type="D" />
  <viseme start="1.360000" type="Ih" />
  <viseme start="1.440000" type="KG" />
  <viseme start="1.520000" type="Z" />
  <viseme start="1.600000" type="D" />
  <mark name="T11" time="1.680000"/>
</word>
<mark name="T12" time="1.680000"/>
<word end="1.840000" start="1.680000" >
  <viseme start="1.680000" type="D" />
  <viseme start="1.760000" type="Ih" />
  <mark name="T13" time="1.840000"/>
</word>
<mark name="T14" time="1.840000"/>
<word end="2.160000" start="1.840000" >
  <viseme start="1.840000" type="Z" />
  <viseme start="1.920000" type="BMP" />
  <viseme start="2.000000" type="EE" />
  <viseme start="2.080000" type="j" />
  <mark name="T15" time="2.160000"/>
</word>
<mark name="T16" time="2.160000"/>
<word end="2.720000" start="2.160000" >
  <viseme start="2.160000" type="Ih" />
  <viseme start="2.240000" type="NG" />
  <viseme start="2.320000" type="D" />
  <viseme start="2.400000" type="Er" />
  <viseme start="2.440000" type="R" />
  <mark name="T17" time="2.480000"/>
</word>
<mark name="T18" time="2.160000"/>
<word end="2.720000" start="2.160000" >
  <viseme start="2.480000" type="F" />
  <viseme start="2.560000" type="Ih" />
  <viseme start="2.640000" type="Z" />
  <mark name="T19" time="2.720000"/>
</word>
<mark name="T20" time="2.720000"/>
<word end="2.880000" start="2.720000" >
  <viseme start="2.720000" type="BMP" />
  <viseme start="2.800000" type="Ih" />
  <mark name="T21" time="2.880000"/>
</word>
<mark name="T22" time="2.880000"/>
<word end="3.599999" start="2.880000" >
  <viseme start="2.880000" type="BMP" />
  <viseme start="2.960000" type="Ih" />
  <viseme start="3.039999" type="KG" />
  <viseme start="3.119999" type="R" />
  <viseme start="3.199999" type="Oh" />
  <viseme start="3.279999" type="Z" />
  <viseme start="3.359999" type="Ao" />
  <viseme start="3.439999" type="F" />
  <viseme start="3.519999" type="D" />
  <viseme start="3.599999" type="_" />
  <mark name="T23" time="3.679999"/>
</word>
<mark name="T24" time="3.679999"/>
<word end="3.759999" start="3.679999" >
  <viseme start="3.679999" type="Ih" />
  <mark name="T25" time="3.759999"/>
</word>
<mark name="T26" time="3.759999"/>
<word end="3.999999" start="3.759999" >
  <viseme start="3.759999" type="OO" />
  <viseme start="3.839999" type="Oh" />
  <viseme start="3.919999" type="Z" />
  <mark name="T27" time="3.999999"/>

```

```

</word>
<mark name="T28" time="3.999999"/>
<word end="4.159998" start="3.999999" >
  <viseme start="3.999999" type="Ih" />
  <viseme start="4.079998" type="D" />
  <mark name="T29" time="4.159998"/>
</word>
<mark name="T30" time="4.159998"/>
<word end="4.319998" start="4.159998" >
  <viseme start="4.159998" type="D" />
  <viseme start="4.239998" type="Ih" />
  <mark name="T31" time="4.319998"/>
</word>
<mark name="T32" time="4.319998"/>
<word end="4.479998" start="4.319998" >
  <viseme start="4.319998" type="BMP" />
  <viseme start="4.399998" type="EE" />
  <mark name="T33" time="4.479998"/>
</word>
<mark name="T34" time="4.479998"/>
<word end="4.799998" start="4.479998" >
  <viseme start="4.479998" type="Ih" />
  <viseme start="4.559998" type="BMP" />
  <viseme start="4.639998" type="Ih" />
  <viseme start="4.719998" type="D" />
  <viseme start="4.799998" type="_" />
  <mark name="T35" time="4.879998"/>
</word>
<mark name="T36" time="4.879998"/>
<word end="5.039998" start="4.879998" >
  <viseme start="4.879998" type="D" />
  <viseme start="4.959998" type="Ih" />
  <mark name="T37" time="5.039998"/>
</word>
<mark name="T38" time="5.039998"/>
<word end="5.279997" start="5.039998" >
  <viseme start="5.039998" type="D" />
  <viseme start="5.119998" type="Ao" />
  <viseme start="5.199997" type="KG" />
  <mark name="T39" time="5.279997"/>
</word>
<mark name="T40" time="5.279997"/>
<word end="5.439997" start="5.279997" >
  <viseme start="5.279997" type="D" />
  <viseme start="5.359997" type="Ih" />
  <mark name="T41" time="5.439997"/>
</word>
<mark name="T42" time="5.439997"/>
<word end="5.599997" start="5.439997" >
  <viseme start="5.439997" type="OO" />
  <viseme start="5.519997" type="Oh" />
  <mark name="T43" time="5.599997"/>
</word>
<viseme start="5.599997" type="_" />
</speak>

```

MSSpeechRelay Example

Text sent to MSSpeech:

```

<speak version="1.0" xml:lang="en-US">
<mark name="spl:T0" />SAPI
<mark name="spl:T1" />
<mark name="spl:T2" />is
<mark name="spl:T3" />
<mark name="spl:T4" />a
<mark name="spl:T5" />
<mark name="spl:T6" />speak
<mark name="spl:T7" />
<mark name="spl:T8" />and
<mark name="spl:T9" />
<mark name="spl:T10" />text
<mark name="spl:T11" />
<mark name="spl:T12" />to
<mark name="spl:T13" />
<mark name="spl:T14" />speak
<mark name="spl:T15" />
<mark name="spl:T16" />interface
<mark name="spl:T17" />
<mark name="spl:T18" />by
<mark name="spl:T19" />
<mark name="spl:T20" />Microsoft.
<mark name="spl:T21" />
<mark name="spl:T22" />I
<mark name="spl:T23" />
<mark name="spl:T24" />use
<mark name="spl:T25" />
<mark name="spl:T26" />it
<mark name="spl:T27" />
<mark name="spl:T28" />to
<mark name="spl:T29" />
<mark name="spl:T30" />be
<mark name="spl:T31" />
<mark name="spl:T32" />able
<mark name="spl:T33" />
<mark name="spl:T34" />to
<mark name="spl:T35" />
<mark name="spl:T36" />talk
<mark name="spl:T37" />
<mark name="spl:T38" />to
<mark name="spl:T39" />
<mark name="spl:T40" />you.
<mark name="spl:T41" />.
</speak>

```

Reply:

```

RemoteSpeechReply brad 4 OK:
<?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\vhtoolkit\data\cache\audio\utt_20110809_154741_brad_4.wav"/>
  <viseme start="0" type="_"/>
  <mark name="T0" time="0.003"/>
  <word end="0.347" start="0.003">
    <viseme start="0.003" type="Z"/>
    <viseme start="0.099" type="Ih"/>
    <viseme start="0.196" type="BMP"/>
    <viseme start="0.259" type="Ih"/>
    <mark name="T1" time="0.347"/>
  </word>
  <mark name="T2" time="0.347"/>
  <word end="0.465" start="0.347">
    <viseme start="0.347" type="Ih"/>
    <viseme start="0.416" type="Z"/>
    <mark name="T3" time="0.465"/>
  </word>
  <mark name="T4" time="0.465"/>
  <word end="0.527" start="0.465">

```

```
<viseme start="0.465" type="Ih"/>
<mark name="T5" time="0.527"/>
</word>
<mark name="T6" time="0.527"/>
<word end="0.874" start="0.527">
  <viseme start="0.527" type="Z"/>
  <viseme start="0.605" type="BMP"/>
  <viseme start="0.683" type="Ih"/>
  <viseme start="0.795" type="KG"/>
  <mark name="T7" time="0.874"/>
</word>
<mark name="T8" time="0.874"/>
<word end="1.053" start="0.874">
  <viseme start="0.874" type="Ih"/>
  <viseme start="0.957" type="D"/>
  <viseme start="1.04" type="D"/>
  <mark name="T9" time="1.053"/>
</word>
<mark name="T10" time="1.053"/>
<word end="1.401" start="1.053">
  <viseme start="1.053" type="D"/>
  <viseme start="1.119" type="Ih"/>
  <viseme start="1.238" type="KG"/>
  <viseme start="1.295" type="Z"/>
  <viseme start="1.348" type="D"/>
  <mark name="T11" time="1.401"/>
</word>
<mark name="T12" time="1.401"/>
<word end="1.47" start="1.401">
  <viseme start="1.401" type="D"/>
  <viseme start="1.442" type="Oh"/>
  <mark name="T13" time="1.47"/>
</word>
<mark name="T14" time="1.47"/>
<word end="1.878" start="1.47">
  <viseme start="1.47" type="Z"/>
  <viseme start="1.547" type="BMP"/>
  <viseme start="1.624" type="Ih"/>
  <viseme start="1.736" type="KG"/>
  <mark name="T15" time="1.878"/>
</word>
<mark name="T16" time="1.878"/>
<word end="2.523" start="1.878">
  <viseme start="1.878" type="Ih"/>
  <viseme start="1.955" type="D"/>
  <viseme start="2.032" type="D"/>
  <viseme start="2.075" type="Ih"/>
  <viseme start="2.11" type="R"/>
  <viseme start="2.145" type="F"/>
  <viseme start="2.257" type="Ih"/>
  <viseme start="2.399" type="Z"/>
  <mark name="T17" time="2.523"/>
</word>
<mark name="T18" time="2.523"/>
<word end="2.665" start="2.523">
  <viseme start="2.523" type="D"/>
  <viseme start="2.554" type="Ih"/>
  <mark name="T19" time="2.665"/>
</word>
<mark name="T20" time="2.665"/>
<word end="3.931" start="2.665">
  <viseme start="2.665" type="BMP"/>
  <viseme start="2.753" type="Ih"/>
  <viseme start="2.841" type="KG"/>
  <viseme start="2.913" type="R"/>
  <viseme start="2.943" type="Ih"/>
  <viseme start="2.973" type="Z"/>
  <viseme start="3.067" type="Ao"/>
  <viseme start="3.202" type="F"/>
  <viseme start="3.255" type="D"/>
  <viseme start="3.308" type="_"/>
```

```
<viseme start="3.928" type="_"/>
<mark name="T21" time="3.931"/>
</word>
<mark name="T22" time="3.931"/>
<word end="4.067" start="3.931">
  <viseme start="3.931" type="Ih"/>
  <mark name="T23" time="4.067"/>
</word>
<mark name="T24" time="4.067"/>
<word end="4.336" start="4.067">
  <viseme start="4.067" type="Ih"/>
  <viseme start="4.17" type="Oh"/>
  <viseme start="4.273" type="Z"/>
  <mark name="T25" time="4.336"/>
</word>
<mark name="T26" time="4.336"/>
<word end="4.474" start="4.336">
  <viseme start="4.336" type="Ih"/>
  <viseme start="4.403" type="D"/>
  <mark name="T27" time="4.474"/>
</word>
<mark name="T28" time="4.474"/>
<word end="4.54" start="4.474">
  <viseme start="4.474" type="D"/>
  <viseme start="4.515" type="Oh"/>
  <mark name="T29" time="4.54"/>
</word>
<mark name="T30" time="4.54"/>
<word end="4.691" start="4.54">
  <viseme start="4.54" type="D"/>
  <viseme start="4.588" type="Ih"/>
  <mark name="T31" time="4.691"/>
</word>
<mark name="T32" time="4.691"/>
<word end="5.051" start="4.691">
  <viseme start="4.691" type="Ih"/>
  <viseme start="4.847" type="D"/>
  <viseme start="4.901" type="Ih"/>
  <viseme start="4.976" type="D"/>
  <mark name="T33" time="5.051"/>
</word>
<mark name="T34" time="5.051"/>
<word end="5.15" start="5.051">
  <viseme start="5.051" type="D"/>
  <viseme start="5.095" type="Oh"/>
  <mark name="T35" time="5.15"/>
</word>
<mark name="T36" time="5.15"/>
<word end="5.469" start="5.15">
  <viseme start="5.15" type="D"/>
  <viseme start="5.244" type="Ao"/>
  <viseme start="5.404" type="KG"/>
  <mark name="T37" time="5.469"/>
</word>
<mark name="T38" time="5.469"/>
<word end="5.64" start="5.469">
  <viseme start="5.469" type="D"/>
  <viseme start="5.584" type="Oh"/>
  <mark name="T39" time="5.64"/>
</word>
<mark name="T40" time="5.64"/>
<word end="6.558" start="5.64">
  <viseme start="5.64" type="Ih"/>
  <viseme start="5.784" type="Oh"/>
  <viseme start="5.928" type="_"/>
  <viseme start="6.553" type="_"/>
  <mark name="T41" time="6.558"/>
</word>
<viseme start="6.558" type="_"/>
</speak>
```

Saso Agent Example

Start Saso - sbm, nvbg, nlu, Fake Recognizer, Agent 1. Click "hello gentlemen".

```
RemoteSpeechCmd speak doctor-perez 1 M021 ../../data/cache/audio/utt_20110809_193606_doctor-perez_1.aiff
<?xml version="1.0" encoding="UTF-8"?>
<speech id="sp1" ref="" type="application/ssml+xml">
  <mark name="T0" />hello
  <mark name="T1" />
  <mark name="T2" />captain
  <mark name="T3" />
</speech>
```

RvoiceRelay Example

Text sent to Rvoice:

```
<?xml version="1.0" encoding="UTF-8"?>
<speech id="sp1" ref="" type="application/ssml+xml">
  <mark name="T0" />hello
  <mark name="T1" />
  <mark name="T2" />captain
  <mark name="T3" />
</speech>
```

Reply:

```
RemoteSpeechReply doctor-perez 1 OK:
<?xml version="1.0" encoding="UTF-8"?>
<speak>
  <soundFile name="d:\edwork\saso\core\beavin\..\..\data\cache\audio\utt_20110809_193606_doctor-perez_1.aiff"/>
  <viseme start="0.0" type="_"/>
  <viseme start="0.0" type="_"/>
  <mark name="T0" time="0.049977324263038546"/>
  <word end="0.33696145124716553" start="0.049977324263038546">
    <viseme start="0.049977324263038546" type="Ih"/>
    <viseme start="0.14498866213151929" type="Ih"/>
    <viseme start="0.2" type="D"/>
    <viseme start="0.24997732426303854" type="OW"/>
  </word>
  <mark name="T2" time="0.33696145124716553"/>
  <mark name="T1" time="0.33696145124716553"/>
  <word end="0.8029931972789116" start="0.33696145124716553">
    <viseme start="0.33696145124716553" type="KG"/>
    <viseme start="0.39696145124716553" type="Ih"/>
    <viseme start="0.4819954648526077" type="BMP"/>
    <viseme start="0.5419954648526077" type="D"/>
    <viseme start="0.6399546485260771" type="Ih"/>
    <viseme start="0.7029931972789115" type="NG"/>
  </word>
  <mark name="T3" time="0.8029931972789116"/>
  <viseme start="0.8029931972789116" type="_"/>
  <viseme start="0.8529705215419501" type="_"/>
</speak>
```