Powered by Universal Speech Solutions LLC

UniMRCP

Bing SS Plugin

Usage Guide

Revision: 7

Created: November 11, 2017

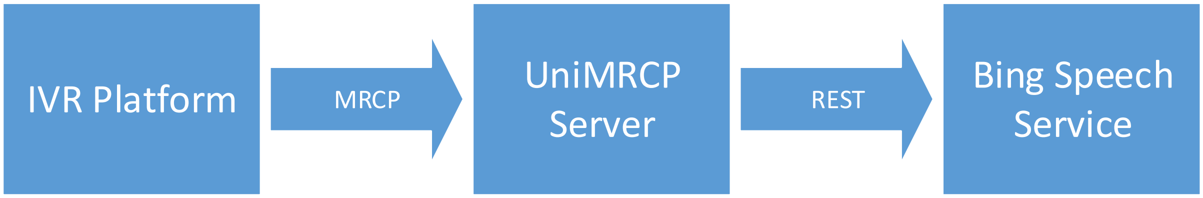
Last updated: January 29, 2019

Author: Arsen Chaloyan

Table of Contents

1 Overview

This guide describes how to configure and use the Microsoft Bing Speech Synthesis (BingSS) plugin to the UniMRCP server. The document is intended for users having a certain knowledge of Microsoft Bing Speech API and UniMRCP.



1.1 Installation

For installation instructions, use one of the guides below.

- RPM Package Installation (Red Hat / Cent OS)
- Deb Package Installation (Debian / Ubuntu)

1.2 Applicable Versions

Instructions provided in this guide are applicable to the following versions.

UniMRCP 1.5.0 and above

UniMRCP BingSS Plugin 1.0.0 and above

2 Supported Features

This is a brief check list of the features currently supported by the UniMRCP server running with the BingSS plugin.

2.1 MRCP Methods

- ✓ SPEAK
- ✓ STOP
- ✓ PAUSE
- ✓ RESUME
- ✓ BARGE-IN-OCCURRED
- ✓ SET-PARAMS
- ✓ GET-PARAMS

2.2 MRCP Events

- ✓ SPEECH-MARKER
- ✓ SPEAK-COMPLETE

2.3 MRCP Header Fields

- ✓ Kill-On-Barge-In
- ✓ Completion-Cause
- ✓ Voice-Gender
- ✓ Voice-Name
- ✓ Prosody-Rate
- ✓ Prosody-Volume
- ✓ Speech-Language

2.4 Speech Data

- ✓ Plain text (text/plain)
- ✓ SSML (application/ssml+xml or application/synthesis+ssml)

3 Supported Voices

All the supported voices are stored in the configuration file *umsbingvoices.xml* located in the directory */opt/unimrcp/conf*. The configuration file is synched with the actual set of voices supported by Microsoft Bing Speech API listed in the following page:

Supported locales and voice fonts

4 Configuration Format

The configuration file of the BingSS plugin is located in */opt/unimrcp/conf/umsbingss.xml*. The configuration file is written in XML.

4.1 Document

The root element of the XML document must be <*umsbingss*>.

Attributes

Name	Unit	Description
license-file	File path	Specifies the license file. File name may include patterns containing '*' sign. If multiple files match the pattern, the most recent one gets used.
subscription-key-file	File path	Specifies the Microsoft subscription key file to use. File name may include patterns containing '*' sign. If multiple files match the pattern, the most recent one gets used.

Parent

None.

Children

Name	Unit	Description
<synth-settings>	String	Specifies synthesis parameters.
<waveform-manager>	String	Specifies parameters of the waveform manager. Available since BingSS 1.4.0.
<sdr-manager>	String	Specifies parameters of the Synthesis Details Record (SDR) manager. Available since BingSS 1.4.0.
<monitoring-agent>	String	Specifies parameters of the monitoring manager.
<license-server>	String	Specifies parameters used to connect to the license server. The use of the license server is optional.

Example

This is an example of a bare document.

```
< umsbingss license-file="umsbingss_*.lic" subscription-key-
file="cognitive.subscription.key">
</ umsbingss>
```

4.2 Synthesis Settings

This element specifies synthesis parameters.

Attributes

Name	Unit	Description
language	String	Specifies the default language to use, if not set by the client.
voice-name	String	Specifies the default voice name. Can be overridden by client. Available since BingSS 1.3.0.
voice-gender	String	Specifies the default voice gender. Can be overridden by client. Available since BingSS 1.3.0.
bypass-ssml	Boolean	Specifies whether transparently bypass or normalize received SSML content before sending it to Bing speech service for synthesis.
auth-validation-period	Integer	Specifies a period in seconds used to re-validate access token based on subscription key. The lifetime of retrieved access token is set to 10 min by Microsoft.

Parent

<umsbingss>

Children

None.

Example

This is an example of synthesis parameters.

```
<synth-settings
```

```
  language="en-US"
  bypass-ssml="false"
  auth-validation-period="480"
/>
```

4.3 Waveform Manager

This element specifies parameters of the waveform manager.

Availability

>= BingSS 1.4.0.

Attributes

Name	Unit	Description
save-waveforms	Boolean	Specifies whether to save waveforms or not.
purge-existing	Boolean	Specifies whether to delete existing records on start-up.
max-file-age	Time interval [min]	Specifies a time interval in minutes after expiration of which a waveform is deleted. Set 0 for infinite.
max-file-count	Integer	Specifies the max number of waveforms to store. If reached, the oldest waveform is deleted. Set 0 for infinite.
waveform-folder	Dir path	Specifies a folder the waveforms should be stored in.

Parent

<umsbingss>

Children

None.

Example

The example below defines a typical utterance manager having the default parameters set.

```
<waveform-manager
  save-waveforms="false"
```

```
      purge-existing="false"
      max-file-age="60"
      max-file-count="100"
      waveform-folder=""
    />
```

4.4 SDR Manager

This element specifies parameters of the Synthesis Details Record (SDR) manager.

Availability

>= BingSS 1.4.0.

Attributes

Name	Unit	Description
save-records	Boolean	Specifies whether to save recognition details records or not.
purge-existing	Boolean	Specifies whether to delete existing records on start-up.
max-file-age	Time interval [min]	Specifies a time interval in minutes after expiration of which a record is deleted. Set 0 for infinite.
max-file-count	Integer	Specifies the max number of records to store. If reached, the oldest record is deleted. Set 0 for infinite.
record-folder	Dir path	Specifies a folder to store recognition details records in. Defaults to ${UniMRCPInstallDir}/var.

Parent

<umsbingss>

Children

None.

Example

The example below defines a typical utterance manager having the default parameters set.

```
<sdr-manager
  save-records="false"
  purge-existing="false"
  max-file-age="60"
  max-file-count="100"
  waveform-folder=""
/>
```

4.5 Monitoring Agent

This element specifies parameters of the monitoring agent.

Attributes

Name	Unit	Description
refresh-period	Time interval [sec]	Specifies a time interval in seconds used to periodically refresh usage details. See <usage-refresh-handler>.

Parent

<umsbingss>

Children

<usage-change-handler>
<usage-refresh-handler>

Example

The example below defines a monitoring agent with usage change and refresh handlers.

```
<monitoring-agent refresh-period="60">

  <usage-change-handler>
    <log-usage enable="true" priority="NOTICE"/>
  </usage-change-handler>

  <usage-refresh-handler>
    <dump-channels enable="true" status-file="umsbingss-channels.status"/>
  </usage-refresh-handler >

</monitoring-agent>
```

4.6 Usage Change Handler

This element specifies an event handler called on every usage change.

Attributes

None.

Parent

<monitoring-agent>

Children

<log-usage>
<update-usage>
<dump-channels>

Example

This is an example of the usage change event handler.

```
<usage-change-handler>
  <log-usage enable="true" priority="NOTICE"/>
  <update-usage enable="false" status-file="umsbingss-usage.status"/>
  <dump-channels enable="false" status-file="umsbingss-channels.status"/>
</usage-change-handler>
```

4.7 Usage Refresh Handler

This element specifies an event handler called periodically to update usage details.

Attributes

None.

Parent

<monitoring-agent>

Children

<log-usage>
<update-usage>
<dump-channels>

Example

This is an example of the usage change event handler.

```
<usage-refresh-handler>
  <log-usage enable="true" priority="NOTICE"/>
  <update-usage enable="false" status-file="umsbingss-usage.status"/>
  <dump-channels enable="false" status-file="umsbingss-channels.status"/>
</usage-refresh-handler>
```

4.8 License Server

This element specifies parameters used to connect to the license server.

Attributes

Name	Unit	Description
enable	Boolean	Specifies whether the use of license server is enabled or not. If enabled, the license-file attribute is not honored.
server-address	String	Specifies the IP address or host name of the license server.
certificate-file	File path	Specifies the client certificate used to connect to the license server. File name may include patterns containing a '*' sign. If multiple files match the pattern, the most recent one gets used.
ca-file	File path	Specifies the certificate authority used to validate the license server.
channel-count	Integer	Specifies the number of channels to check out from the license server. If not specified or set to 0, either all available channels or a pool of channels will be checked based on the configuration of the license server.

Parent

<umsbingss>

Children

None.

Example

The example below defines a typical configuration which can be used to connect to a license server located, for example, at 10.0.0.1.

```
<license-server
  enable="true"
  server-address="10.0.0.1"
  certificate-file="unilic_client_*.crt"
  ca-file="unilic_ca.crt"
/>
```

For further reference to the license server, visit

```
http://unimrcp.org/licserver
```

5 Configuration Steps

This section outlines common configuration steps.

5.1 Using Default Configuration

The default configuration should be sufficient for the general use.

5.2 Specifying Synthesis Language

Synthesis language can be specified by the client per MRCP session by means of the header field *Speech-Language* set in a *SET-PARAMS* or *SPEAK* request, or inline in the SSML data. Otherwise, the parameter *language* set in the configuration file *umsbingss.xml* is used. The parameter defaults to *en-US*.
For supported languages and their corresponding codes, visit the following link.

> https://docs.microsoft.com/en-us/azure/cognitive-services/speech/api-reference-rest/supportedlanguages

5.3 Specifying Sampling Rate

Sampling rate is determined based on the SDP negotiation. Refer to the configuration guide of the UniMRCP server on how to specify supported encodings and sampling rates to be used in communication between the client and server. Either 8 or 16 kHz can be used by Microsoft Bing Speech API for synthesis.

5.4 Specifying Voice Parameters

Global Settings

The default voice name and gender can be specified from the configuration file *umsbingss.xml* using the *voice-name* and *voice-gender* attributes of the *synth-settings* element. This functionality is available since BingSS 1.3.0 release.

MRCP Header Fields

The voice name and gender can be specified by the MRCP client in *SET-PARAMS* and *SPEAK* requests.

- Voice-Name

  This is an optional parameter indicating the name of the voice to use for synthesis.

- Voice-Gender

  This is an optional parameter indicating the preferred gender of the voice to use for synthesis, which can be set to either *male* or *female*.

5.5 Specifying Prosody Parameters

The following prosody parameters can be specified by the MRCP client in *SET-PARAMS* and *SPEAK* requests.

- Prosody-Rate

  This is an optional parameter indicating the speaking rate, which can be set to one of the following labels: *x-slow*, *slow*, *medium*, *fast*, *x-fast*, *default*.

- Prosody-Volume

  This is an optional parameter indicating the speaking volume, which can be set to one of the following labels: *silent*, *x-soft*, *soft*, *medium*, *loud*, *x-loud*, *default*.

5.6 Specifying Speech Data

Speech data can be specified by the MRCP client in *SPEAK* requests using one of the following content types:

- plain/text
- application/ssml+xml (or application/synthesis+ssml)

5.7 Maintaining Waveforms

Collection of waveforms is not required for regular operation and is disabled by default. However, enabling this functionality allows to save synthesized speech received from the Microsoft Bing Speech service and later listen to them offline.

The relevant settings can be specified via the element *waveform-manager*.

- save-waveforms

  Utterances can optionally be recorded and stored if the configuration parameter *save-waveforms* is set to true.

- purge-existing

  This parameter specifies whether to delete existing waveforms on start-up.

- max-file-age

  This parameter specifies a time interval in minutes after expiration of which a waveform is deleted. If set to 0, there is no expiration time specified.

- max-file-count

  This parameter specifies the maximum number of waveforms to store. If the specified number is reached, the oldest waveform is deleted. If set to 0, there is no limit specified.

- waveform-folder

This parameter specifies a path to the directory used to store waveforms in. The directory defaults to *${UniMRCPInstallDir}/var*.

5.8 Maintaining Synthesis Details Records

Collection of synthesis details records (SDR) is not required for regular operation and is disabled by default. However, enabling this functionality allows to store details of each synthesis attempt in a separate file and analyze them later offline. The SDRs ate stored in the JSON format.

The relevant settings can be specified via the element *sdr-manager*.

- save-records

This parameter specifies whether to save synthesis details records or not.

- purge-existing

This parameter specifies whether to delete existing records on start-up.

- max-file-age

This parameter specifies a time interval in minutes after expiration of which a record is deleted. If set to 0, there is no expiration time specified.

- max-file-count

This parameter specifies the maximum number of records to store. If the specified number is reached, the oldest record is deleted. If set to 0, there is no limit specified.

- record-folder

This parameter specifies a path to the directory used to store records in. The directory defaults to *${UniMRCPInstallDir}/var*.

The following is the content of a sample SDR.

```
{"synth-details-record": {
  "datetime": "2019-01-19 12:57:44",
  "language": "en-US",
  "voice-name": "",
  "sampling-rate": "8000 Hz",
  "codec": "PCMU",
  "data-length": "18150 bytes",
  "start-of-streaming-ts": "282 ms"
  "completion-ts": "2550 ms"
  "completion-cause": "normal",
}}
```

where the stored attributes are:

- datetime

This attribute denotes the date and time captured when the corresponding MRCP SPEAK request is received.

- language

This attribute denotes the speech language used with the request.

- voice-name

This attribute denotes the voice name used with the request.

- sampling-rate

This attribute denotes the sampling rate used with the request.

- codec

This attribute denotes the codec of synthesized speech received from the service.

- data-length

This attribute denotes the number of bytes of synthesized speech received from the service.

- start-of-streaming-ts

This attribute denotes a time interval in milliseconds, elapsed since initiation of the request, captured when streaming to the MRCP client is started. This attribute also denotes the HTTP response time of the service.

- completion-ts

This attribute denotes a time interval in milliseconds, elapsed since initiation of the request, captured upon completion of the request (SPEAK-COMPLETE is sent).

- completion-cause

This attribute denotes the completion cause of the request.

6 Monitoring Usage Details

The number of in-use and total licensed channels can be monitored in several alternate ways. There is a set of actions which can take place on certain events. The behavior is configurable via the element *monitoring-agent*, which contains two event handlers: *usage-change-handler* and *usage-refresh-handler*.

While the *usage-change-handler* is invoked on every acquisition and release of a licensed channel, the *usage-refresh-handler* is invoked periodically on expiration of a timeout specified by the attribute *refresh-period*.

The following actions can be specified for either of the two handlers.

6.1 Log Usage

The action *log-usage* logs the following data in the order specified.

- The number of currently in-use channels.
- The maximum number of channels used concurrently. Available since BingSS 1.4.0.
- The total number of licensed channels.

The following is a sample log statement, indicating 0 in-use, 0 max-used and 2 total channels.

```
[NOTICE] BINGSS Usage: 0/0/2
```

6.2 Update Usage

The action *update-usage* writes the following data to a status file *umsbingss-usage.status*, located by default in the directory *${UniMRCPInstallDir}/var/status*.

- The number of currently in-use channels.
- The maximum number of channels used concurrently. Available since BingSS 1.4.0.
- The total number of licensed channels.
- The current status of the license permit.

The following is a sample content of the status file.

```
in-use channels: 0
max used channels: 0
total channels: 2
license permit: true
```

6.3 Dump Channels

The action *dump-channel* writes the identifiers of in-use channels to a status file *umsbingss-channels.status*, located by default in the directory *${UniMRCPInstallDir}/var/status*.

7 Usage Examples

7.1 SSML

This examples demonstrates how to perform speech synthesis by using a SPEAK request with an SSML content.

C->S:

```
MRCP/2.0 309 SPEAK 1
Channel-Identifier: 4dde51f37d1a9546@speechsynth
Content-Type: application/ssml+xml
Voice-Age: 28
Content-Length: 163

<?xml version="1.0"?>
<speak version="1.0" xml:lang="en-US" xmlns="http://www.w3.org/2001/10/synthesis">
  <p>
    <s>Welcome to Uni MRCP.</s>
  </p>
</speak>
```

S->C:

```
MRCP/2.0 83 1 200 IN-PROGRESS
Channel-Identifier: 4dde51f37d1a9546@speechsynth
```

S->C:

```
MRCP/2.0 122 SPEAK-COMPLETE 1 COMPLETE
Channel-Identifier: 4dde51f37d1a9546@speechsynth
Completion-Cause: 000 normal
```

7.2 Plain Text

This examples demonstrates how to perform speech synthesis by using a SPEAK request with a plain text content.

C->S:

```
MRCP/2.0 155 SPEAK 1
Channel-Identifier: 85667d0efbf95345@speechsynth
Content-Type: text/plain
Voice-Age: 28
Content-Length: 20

Welcome to Uni MRCP.
```

S->C:

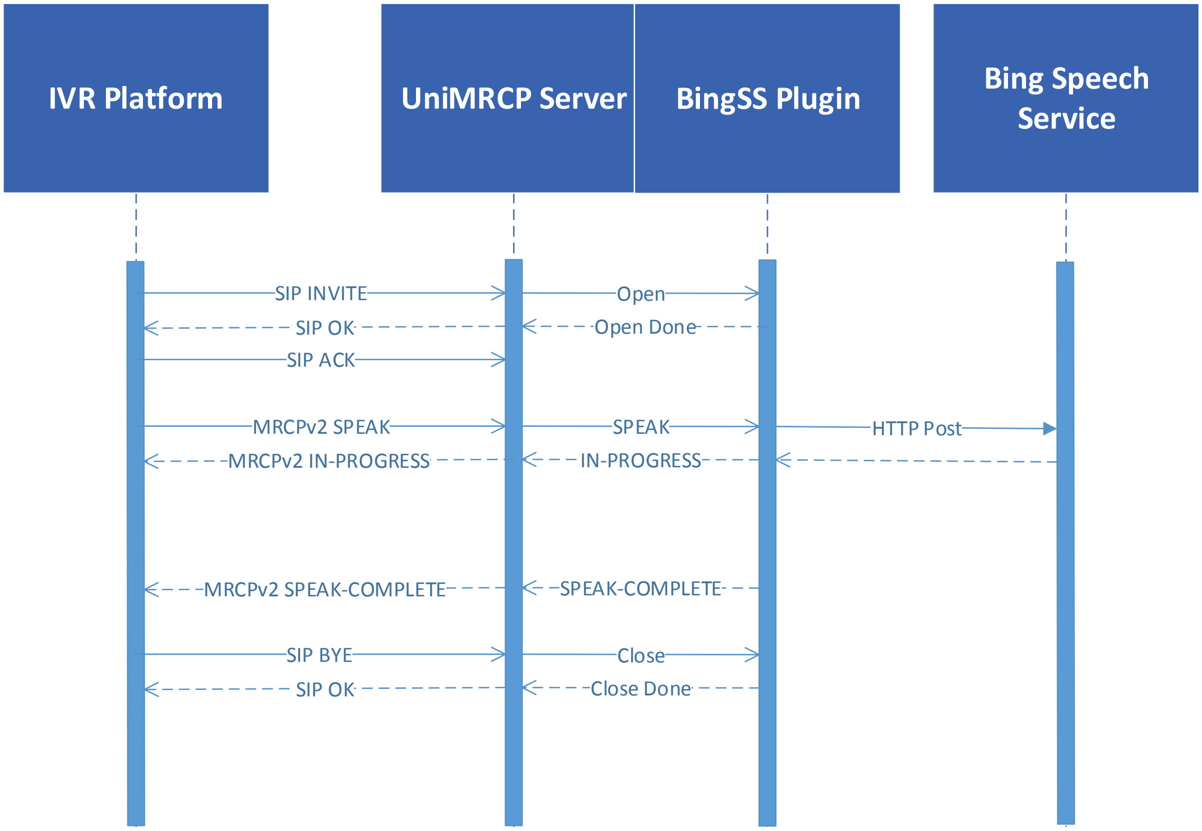
```
MRCP/2.0 83 1 200 IN-PROGRESS
Channel-Identifier: 85667d0efbf95345@speechsynth
```

S->C:

```
MRCP/2.0 122 SPEAK-COMPLETE 1 COMPLETE
Channel-Identifier: 85667d0efbf95345@speechsynth
Completion-Cause: 000 normal
```

8 Sequence Diagram

The following sequence diagram outlines common interactions between all the main components involved in a typical synthesis session performed over MRCPv2.



9 References

9.1 Microsoft Azure

- Text to Speech API
- Authentication

9.2 Specifications

- Speech Synthesizer Resource
- SSML