

Powered by Universal Speech Solutions LLC



Yandex SS Plugin

Usage Guide

Revision: 4

Created: November 19, 2018

Last updated: August 28, 2020

Author: Arsen Chaloyan

Table of Contents

1	Overview	4
1.1	Installation	4
1.2	Applicable Versions	4
2	Supported Features	5
2.1	MRCP Methods	5
2.2	MRCP Events	5
2.3	MRCP Header Fields	5
2.4	Speech Data	5
3	Supported Voices	6
4	Configuration Format	7
4.1	Document	7
4.2	Synthesis Settings	8
4.3	Waveform Manager	9
4.4	SDR Manager	10
4.5	Monitoring Agent	11
4.6	Usage Change Handler	12
4.7	Usage Refresh Handler	13
4.8	License Server	13
5	Configuration Steps	16
5.1	Using Default Configuration	16
5.2	Specifying Folder ID	16
5.3	Specifying Synthesis Language	16
5.4	Specifying Sampling Rate	16
5.5	Specifying Voice Parameters	16
5.6	Specifying Prosody Parameters	17
5.7	Specifying Vendor-Specific Parameters	17
5.8	Specifying Speech Data	17
5.9	Maintaining Waveforms	17
5.10	Maintaining Synthesis Details Records	18
5.11	Using Cache	19
6	Monitoring Usage Details	20
6.1	Log Usage	20
6.2	Update Usage	20
6.3	Dump Channels	21
7	Usage Examples	22
7.1	Plain Text	22

8 Sequence Diagram 23

9 References 24

 9.1 Yandex SpeechKit 24

 9.2 Specifications 24

1 Overview

This guide describes how to configure and use the Yandex Speech Synthesis (SS) plugin to the UniMRCP server. The document is intended for users having a certain knowledge of Yandex SpeechKit Text-to-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 1.6.0 and above

UniMRCP Yandex SS 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 Yandex SS 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
- ✓ Logging-Tag
- ✓ Cache-Control

2.4 Speech Data

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

3 Supported Voices

All the voices supported by Yandex Text-to-Speech API are listed in the following page:

https://cloud.yandex.ru/docs/speechkit/tts/request#body_params

4 Configuration Format

The configuration file of the Yandex SS plugin is located in `/opt/unimrcp/conf/umsyandexss.xml`. The configuration file is written in XML.

4.1 Document

The root element of the XML document must be `<umsyandexss>`.

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 Yandex SpeechKit 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
<code><synth-settings></code>	String	Specifies synthesis parameters.
<code><waveform-manager></code>	String	Specifies parameters of the waveform manager.
<code><sdr-manager></code>	String	Specifies parameters of the Synthesis Details Record (SDR) manager.
<code><monitoring-agent></code>	String	Specifies parameters of the monitoring manager.
<code><license-server></code>	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.

```
<umsyandexss license-file="umsyandexss_*.lic"  
              subscription-key-file="yandex.subscription.key">  
</ umsyandexss>
```

4.2 Synthesis Settings

This element specifies synthesis parameters.

Attributes

Name	Unit	Description
folder-id	String	Specifies the Yandex SpeechKit folder identifier. Required.
language	String	Specifies the default language to use, if not set by the client.
bypass-ssml	Boolean	Specifies whether to transparently bypass or parse received content in order to determine voice parameters set in SSML. Available since YSS 1.3.0.
normalize-ssml	Boolean	Specifies whether to normalize SSML. The parameter is observed only when <i>bypass-ssml</i> is set to <i>false</i> . Available since YSS 1.3.0.
voice-name	String	Specifies the default voice name. Can be overridden by client.
prosody-rate	String	Specifies the default prosody rate. Use either labels (x-slow, slow, medium, fast, x-fast) or relative changes from the default rate as a percentage (+10% or -20%) or numeric values as a multiplier where 1 indicates no change. Can be overridden by client. Available since YSS 1.4.0.
emotion	String	Specifies an emotion, which can be set to either "good", "evil" or "neutral".
auth-validation-period	Integer	Specifies a period in seconds used to re-validate access token based on credentials.
http-proxy	String	Specifies the URI of HTTP proxy, if used. Available since YSS 1.2.0.
caching	Boolean	Specifies whether to enable caching of synthesized

		waveforms. Available since YSS 1.4.0.
--	--	---------------------------------------

Parent

<umsyandexss>

Children

None.

Example

This is an example of synthesis parameters.

```
<synth-settings
  folder-id="abcd1234edfg"
  language="en-US"
  voice-name=""
  emotion=" "
  auth-validation-period="3600"
/>
```

4.3 Waveform Manager

This element specifies parameters of the waveform manager.

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.

file-prefix	String	Specifies a prefix used to compose the name of the file to be stored. Defaults to 'umsyandexss-', if not specified.
use-logging-tag	Boolean	Specifies whether to use the MRCP header field Logging-Tag, if present, to compose the name of the file to be stored. Available since YSS 1.4.0.

Parent

< umsyandexss >

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.

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 <code>\${UniMRCPIInstallDir}/var</code> .
file-prefix	String	Specifies a prefix used to compose the name of the file to be stored. Defaults to 'umsyandexss-', if not specified.
use-logging-tag	Boolean	Specifies whether to use the MRCP header field Logging-Tag, if present, to compose the name of the file to be stored. Available since YSS 1.6.0.

Parent

< umsyandexss >

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

<umyandexss>

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="umyandexss-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="umsyandexss-usage.status"/>
  <dump-channels enable="false" status-file="umsyandexss-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="umsyandexss-usage.status"/>
  <dump-channels enable="false" status-file="umsyandexss-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.
http-proxy-address	String	Specifies the IP address or host name of the HTTP proxy server, if used. Available since YSS 1.6.0.
http-proxy-port	Integer	Specifies the port number of the HTTP proxy server, if used. Available since YSS 1.6.0.

Parent

<umsyandexss>

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_*.cert"
  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 Folder ID

The Yandex SpeechKit folder identifier must be set in the configuration file *umsyandexss.xml*.

5.3 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. Otherwise, the parameter *language* set in the configuration file *umsyandexss.xml* is used. The parameter defaults to *en-US*.

5.4 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.

5.5 Specifying Voice Parameters

Global Settings

The default voice name can be specified from the configuration file *umsyandexss.xml* using the *voice-name* attribute of the *synth-settings* element.

MRCP Header Fields

The voice name 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.

SSML Content

The voice name can also be specified using the corresponding attribute of the *voice* element in SSML content. In order to parse and determine the parameters and pass them forward to Yandex Text-to-Speech API accordingly, the *bypass-ssml* attribute of the *synth-settings* element must be set to *false* in the configuration file *umsyandexss.xml*. This functionality is available since YSS 1.3.0 release.

Since YSS 1.3.0 release, if the *bypass-ssml* attribute is set to *false* and the *normalize-ssml* attribute is set to *true*, then the *voice* element, if present, is stripped off from the SSML content passed to the service in order to conform to the subset of SSML supported by Yandex Text-to-Speech API.

5.6 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*.

5.7 Specifying Vendor-Specific Parameters

The following parameters can optionally be specified by the MRCP client in *SET-PARAMS* and *SPEAK* requests via the MRCP header field *Vendor-Specific-Parameters*.

Name	Unit	Description
folder-id	String	Specifies the Yandex SpeechKit folder identifier.
emotion	String	Specifies an emotion, which can be set to either "good", "evil" or "neutral".

5.8 Specifying Speech Data

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

- plain/text
- application/ssml+xml (or application/synthesis+ssml), available since YSS 1.3.0 release

5.9 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 Yandex SpeechKit 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*.

5.10 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 are 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*.

5.11 Using Cache

Since YSS 1.4.0, synthesized waveforms can be stored and re-used for consecutive speech synthesis requests, when applicable. In order to use this functionality, the attribute *caching* of the element *synth-settings* must be set to *true*. The attribute defaults to *false*.

The lifetime and size of cached records are controlled by the attributes *max-file-age* and *max-file-count* of the element *waveform-manager*.

The cached records are persistent and populated on initial loading, unless the attribute *purge-existing* of the element *waveform-manager* is set to *true*.

The following speech synthesis parameters are observed while searching for a cached record.

- language
- voice-name
- voice-gender
- sampling-rate
- prosody-pitch
- prosody-contour
- prosody-range
- prosody-rate
- prosody-duration
- prosody-volume
- content

The following cache control directives are observed while searching for a cached record.

- max-age
- min-fresh

The cache control directives can be specified by the client per individual speech synthesis request via the MRCP header field *Cache-Control*. By default, no cache control directives are applied.

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.
- 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] YandexSS Usage: 0/0/2
```

6.2 Update Usage

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

- The number of currently in-use channels.
- The maximum number of channels used concurrently.
- The total number of licensed channels.
- The current status of the license permit.
- The license server alarm. Set to *on*, if the license server is not available for more than one hour; otherwise, set to *off*. This parameter is maintained only if the license server is used. Available since YSS 1.2.0.

The following is a sample content of the status file.

```
in-use channels: 0
```

```
max used channels: 0
total channels: 2
total channels: 2
licserver alarm: off
```

6.3 Dump Channels

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

7 Usage Examples

7.1 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:

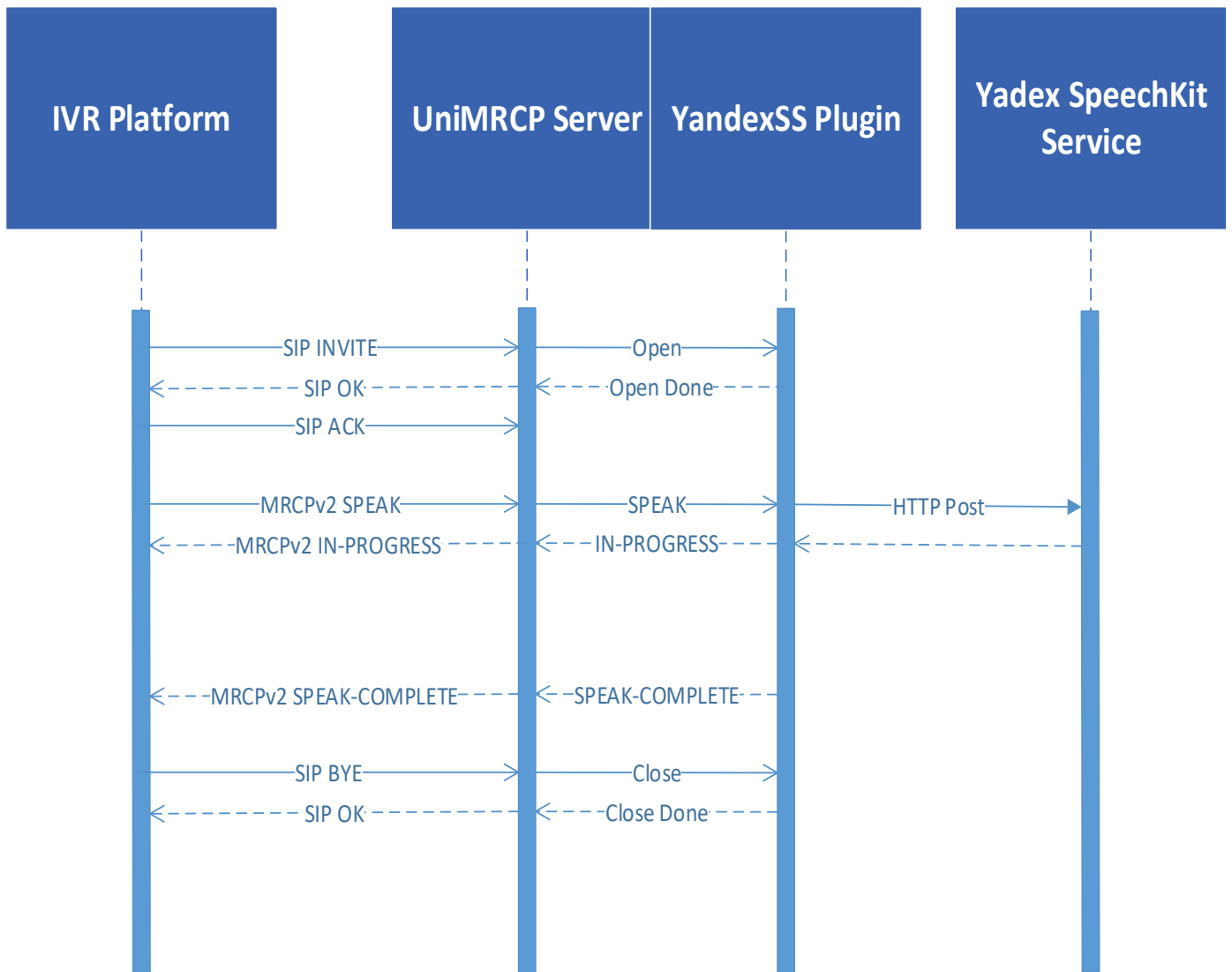
```
MRCP/2.0 83 1 200 IN-PROGRESS
Channel-Identifier: 85667d0efbf95345@speechsynth
```

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 Yandex SpeechKit

- [Text to Speech API](#)
- [Authentication](#)

9.2 Specifications

- [Speech Synthesizer Resource](#)