

*Powered by Universal Speech Solutions LLC*



# Watson SR Plugin

## Usage Guide

---

Revision: 7

Created: July 4, 2018

Last updated: September 4, 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 Grammars.....	6
2.5 Results.....	6
3 Configuration Format.....	7
3.1 Document.....	7
3.2 Streaming Recognition .....	8
3.3 Speech Contexts.....	11
3.4 Speech Context .....	11
3.5 Phrase.....	12
3.6 Speech and DTMF Input Detector.....	13
3.7 Utterance Manager.....	15
3.8 RDR Manager .....	16
3.9 Monitoring Agent .....	17
3.10 Usage Change Handler .....	18
3.11 Usage Refresh Handler .....	18
3.12 License Server.....	19
4 Configuration Steps .....	21
4.1 Using Default Configuration.....	21
4.2 Specifying Recognition Language.....	21
4.3 Specifying Sampling Rate.....	21
4.4 Specifying Speech Input Parameters .....	21
4.5 Specifying DTMF Input Parameters.....	22
4.6 Specifying No-Input and Recognition Timeouts .....	22
4.7 Specifying Vendor-Specific Parameters .....	23
4.8 Maintaining Utterances.....	25
4.9 Maintaining Recognition Details Records .....	25
5 Recognition Grammars and Results.....	27
5.1 Using Built-in Speech Transcription .....	27
5.2 Using Built-in DTMF Grammars.....	28
5.3 Retrieving Results.....	28

6	Monitoring Usage Details .....	29
6.1	Log Usage .....	29
6.2	Update Usage .....	29
6.3	Dump Channels.....	30
7	Usage Examples.....	31
7.1	Speech Transcription .....	31
7.2	DTMF Recognition.....	32
7.3	Speech and DTMF Recognition.....	33
8	Sequence Diagram .....	35
9	References.....	36
9.1	IBM Watson.....	36
9.2	Specifications.....	36

# 1 Overview

This guide describes how to configure and use the IBM Watson Speech Recognition (SR) plugin to the UniMRCP server. The document is intended for users having a certain knowledge of IBM Watson Speech-to-Text 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 Watson SR 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 Watson SR plugin.

## 2.1 MRCP Methods

- ✓ DEFINE-GRAMMAR
- ✓ RECOGNIZE
- ✓ START-INPUT-TIMERS
- ✓ STOP
- ✓ SET-PARAMS
- ✓ GET-PARAMS

## 2.2 MRCP Events

- ✓ RECOGNITION-COMPLETE
- ✓ START-OF-INPUT

## 2.3 MRCP Header Fields

- ✓ Input-Type
- ✓ No-Input-Timeout
- ✓ Recognition-Timeout
- ✓ Speech-Complete-Timeout
- ✓ Speech-Incomplete-Timeout
- ✓ Waveform-URI
- ✓ Media-Type
- ✓ Completion-Cause
- ✓ Confidence-Threshold
- ✓ Start-Input-Timers
- ✓ DTMF-Interdigit-Timeout
- ✓ DTMF-Term-Timeout
- ✓ DTMF-Term-Char
- ✓ Save-Waveform
- ✓ Speech-Language
- ✓ Cancel-If-Queue

- ✓ Sensitivity-Level

## 2.4 Grammars

- ✓ Built-in speech transcription grammar
- ✓ Built-in/embedded DTMF grammar
- ✓ SRGS XML (limited support)

## 2.5 Results

- ✓ NLSML

# 3 Configuration Format

The configuration file of the Watson SR plugin is located in `/opt/unimrcp/conf/umswatsonsr.xml`. The configuration file is written in XML.

## 3.1 Document

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

### Attributes

Name	Unit	Description
<b>license-file</b>	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.
<b>credentials-file</b>	File path	Specifies the IBM Watson credentials 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
<b>&lt;ws-streaming-recognition&gt;</b>	String	Specifies parameters of streaming recognition employed via the WebSocket protocol.
<b>&lt;speech-contexts&gt;</b>	String	Contains a list of speech contexts. Available since WSR 1.2.0.
<b>&lt;speech-dtmf-input-detector&gt;</b>	String	Specifies parameters of the speech and DTMF input detector.
<b>&lt;utterance-manager&gt;</b>	String	Specifies parameters of the utterance manager.
<b>&lt;rdr-manager&gt;</b>	String	Specifies parameters of the Recognition Details Record (RDR) manager.
<b>&lt;monitoring-agent&gt;</b>	String	Specifies parameters of the monitoring manager.

<code>&lt;license-server&gt;</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.

```
<umswatsonsr license-file="umswatsonsr_*.lic" credentials-file="watsonsr.credentials">
</umswatsonsr>
```

## 3.2 Streaming Recognition

This element specifies parameters of streaming recognition employed via the WebSocket interface.

### Attributes

Name	Unit	Description
<b>language</b>	String	Specifies the default language to use, if not set by the client. For a list of supported languages, visit <a href="https://console.bluemix.net/docs/services/speech-to-text/input.html#models">https://console.bluemix.net/docs/services/speech-to-text/input.html#models</a>
<b>language-customization-id</b>	String	Specifies the custom language model id to be used for recognition, if any. Can be overridden by client. Available since WSR 1.5.0.
<b>acoustic-customization-id</b>	String	Specifies the custom acoustic model id to be used for recognition, if any. Can be overridden by client. Available since WSR 1.5.0.
<b>base-model-version</b>	String	Specifies the base model version to be used for recognition, if any. Can be overridden by client. Available since WSR 1.5.0.
<b>grammar-name</b>	String	Specifies the grammar name used for speech recognition in conjunction with the customization id of the custom language model for which the grammar is defined. Can be overridden by client. Available since WSR 1.6.0.
<b>customization-weight</b>	Double	Specifies the weight to give to words from the custom language model compared to those from the base model, if any. Can be overridden by client. Available since WSR



		1.5.0.
<b>max-alternatives</b>	Integer	Specifies the maximum number of speech recognition result alternatives to be returned. Can be overridden by client by means of the header field <i>N-Best-List-Length</i> .
<b>alternatives-below-threshold</b>	Boolean	Specifies whether to return speech recognition result alternatives with the confidence score below the confidence threshold. Can be overridden by client. Available since WSR 1.2.0.
<b>confidence-format</b>	String	Specifies the format of the confidence score to be returned (use "auto" for a format based on protocol version, "mrcpv2" for a float value in the range of 0..1, "mrcpv1" for an integer value in the range of 0..100). Available since WSR 1.2.0.
<b>smart-formatting</b>	Boolean	Specifies whether the service converts dates, times, numbers, currency, and similar values into more conventional representations in the final transcript.
<b>word-confidence</b>	Boolean	Specifies whether the service returns a confidence for each word. Can be overridden by client. Available since WSR 1.5.0.
<b>timestamps</b>	Boolean	Specifies whether the service returns time alignment for each word. Can be overridden by client. Available since WSR 1.5.0.
<b>speaker-labels</b>	Boolean	Specifies whether the service returns labels identifying a speaker for each word. Can be overridden by client. Available since WSR 1.5.0.
<b>redaction</b>	Boolean	Specifies whether the service returns masked numeric data having three or more consecutive digits. Can be overridden by client. Available since WSR 1.5.0.
<b>processing-metrics</b>	Boolean	Specifies whether the service returns metrics on transcription of the input audio data. Can be overridden by client. Available since WSR 1.5.0.
<b>processing-metrics-interval</b>	Time interval [sec]	Specifies an interval in seconds at which the service returns processing metrics. Can be overridden by client. Available since WSR 1.5.0.
<b>audio-metrics</b>	Boolean	Specifies whether the service returns signal characteristics of the input audio data. Can be overridden by client. Available since WSR 1.5.0.
<b>end-of-phrase-</b>	Time interval	Specifies an interval between 0.0 and 12.0 seconds at which the service splits a transcript into multiple final results if it

<b>silence-time</b>	[sec]	encounters silence. Not used, if set to -1. Can be overridden by client. Available since WSR 1.6.0.
<b>start-of-input</b>	String	Specifies the source of start of input event sent to the client (use "service-originated" to rely on service-originated first interim result and "internal" for plugin-originated event).
<b>skip-unsupported-grammars</b>	Boolean	Specifies whether to skip or raise an error while referencing a malformed or not supported grammar.
<b>transcription-grammar</b>	String	Specifies the name of the built-in speech transcription grammar. The grammar can be referenced as <i>builtin:speech/transcribe</i> or <i>builtin:grammar/transcribe</i> , where <i>transcribe</i> is the default value of this parameter.
<b>auth-validation-period</b>	Time interval [sec]	Specifies a period in seconds used to re-validate access token based on subscription key. Since WSR 1.4.0, if set to 0, the period is based on <i>expires_in</i> received in response from service.
<b>http-proxy</b>	String	Specifies the URI of HTTP proxy, if used. Available since WSR 1.5.0.
<b>inter-result-timeout</b>	Time interval [msec]	Specifies a timeout between interim results containing transcribed speech. If the timeout is elapsed, input is considered complete. The timeout defaults to 0 (disabled). Available since WSR 1. 6.0.

## Parent

<umswatsonsr>

## Children

None.

## Example

This is an example of streaming recognition element.

```
<ws-streaming-recognition
  language="en-US"
  max-alternatives="1"
  alternatives-below-threshold="false"
  confidence-format="auto"
  smart-formatting="true"
  start-of-input="internal"
  skip-unsupported-grammars="true"
  transcription-grammar="transcribe"
  auth-validation-period="0"
```

```
</>
```

### 3.3 Speech Contexts

This element specifies a list of speech contexts.

#### Availability

>= WSR 1.2.0.

#### Attributes

None.

#### Parent

<umswatsonsr>

#### Children

<speech-context>

#### Example

The example below defines a speech contexts *directory*.

```
<speech-contexts>
  <speech-context id="directory" speech-complete="true" enable="true">
    <phrase>call Steve</phrase>
    <phrase>call John</phrase>
    <phrase>dial 5</phrase>
    <phrase>dial 6</phrase>
  </speech-context>
</speech-contexts>
```

### 3.4 Speech Context

This element specifies a speech context.

#### Availability

>= WSR 1.2.0.

#### Attributes

Name	Unit	Description
id	String	Specifies a unique string identifier of the speech context to

		be referenced by the MRCP client.
<b>enable</b>	Boolean	Specifies whether the speech context is enabled or disabled.
<b>speech-complete</b>	Boolean	Specifies whether to complete input as soon as an interim result matches one of the specified phrases.
<b>language</b>	String	The language the phrases are defined for. Available since WSR 1.3.0.
<b>scope</b>	String	Specifies a scope of the speech context, which can be set to either <i>hint</i> or <i>strict</i> . Available since WSR 1.4.0.

### Parent

<speech-contexts>

### Children

<phrase>

### Example

This is an example of speech context element.

```
<speech-context id="directory" speech-complete="true" enable="true">
  <phrase>call Steve</phrase>
  <phrase>call John</phrase>
  <phrase>dial 5</phrase>
  <phrase>dial 6</phrase>
</speech-context>
```

## 3.5 Phrase

This element specifies a phrase in the speech context.

### Availability

>= WSR 1.2.0.

### Attributes

Name	Unit	Description
<b>Tag</b>	String	Specifies an optional arbitrary string identifier to be returned as an instance in the NLSML result, if the

transcription result matches the phrase.

## Parent

<speech-context>

## Children

None.

This is an example of a speech context with phrases having tags specified. Available since GSR 1.9.0.

```
<speech-context id="boolean" speech-complete="true" scope="strict" enable="true">
  <phrase tag="true">yes</phrase>
  <phrase tag="true">sure</phrase>
  <phrase tag="true">correct</phrase>
  <phrase tag="false">no</phrase>
  <phrase tag="false">not sure</phrase>
  <phrase tag="false">incorrect </phrase>
</speech-context>
```

## 3.6 Speech and DTMF Input Detector

This element specifies parameters of the speech and DTMF input detector.

### Attributes

Name	Unit	Description
<b>vad-mode</b>	Integer	Specifies an operating mode of VAD in the range of [0 ... 3]. Default is 1.
<b>speech-start-timeout</b>	Time interval [msec]	Specifies how long to wait in transition mode before triggering a start of speech input event.
<b>speech-complete-timeout</b>	Time interval [msec]	Specifies how long to wait in transition mode before triggering an end of speech input event. The complete timeout is used when there is an interim result available.
<b>speech-incomplete-timeout</b>	Time interval [msec]	Specifies how long to wait in transition mode before triggering an end of speech input event. The incomplete timeout is used as long as there is no interim result available. Afterwards, the complete timeout

		is used.
<b>noinput-timeout</b>	Time interval [msec]	Specifies how long to wait before triggering a no-input event.
<b>input-timeout</b>	Time interval [msec]	Specifies how long to wait for input to complete.
<b>dtmf-interdigit-timeout</b>	Time interval [msec]	Specifies a DTMF inter-digit timeout.
<b>dtmf-term-timeout</b>	Time interval [msec]	Specifies a DTMF input termination timeout.
<b>dtmf-term-char</b>	Character	Specifies a DTMF input termination character.
<b>speech-leading-silence</b>	Time interval [msec]	Specifies desired silence interval preceding spoken input.
<b>speech-trailing-silence</b>	Time interval [msec]	Specifies desired silence interval following spoken input.
<b>speech-output-period</b>	Time interval [msec]	Specifies an interval used to send speech frames to the recognizer.

## Parent

<umswatsonsr>

## Children

None.

## Example

The example below defines a typical speech and DTMF input detector having the default parameters set.

```
<speech-dtmf-input-detector
vad-mode="2"
speech-start-timeout="300"
speech-complete-timeout="1000"
speech-incomplete-timeout="1000"
noinput-timeout="5000"
input-timeout="10000"
dtmf-interdigit-timeout="5000"
dtmf-term-timeout="10000"
dtmf-term-char=""
speech-leading-silence="300"
speech-trailing-silence="300"
speech-output-period="200"
```

</>

## 3.7 Utterance Manager

This element specifies parameters of the utterance manager.

### Attributes

Name	Unit	Description
<b>save-waveforms</b>	Boolean	Specifies whether to save waveforms or not.
<b>purge-existing</b>	Boolean	Specifies whether to delete existing records on start-up.
<b>max-file-age</b>	Time interval [min]	Specifies a time interval in minutes after expiration of which a waveform is deleted. Set 0 for infinite.
<b>max-file-count</b>	Integer	Specifies the max number of waveforms to store. If reached, the oldest waveform is deleted. Set 0 for infinite.
<b>waveform-base-uri</b>	String	Specifies the base URI used to compose an absolute waveform URI.
<b>waveform-folder</b>	Dir path	Specifies a folder the waveforms should be stored in.
<b>file-prefix</b>	String	Specifies a prefix used to compose the name of the file to be stored. Defaults to 'umswatsons-', if not specified.
<b>use-logging-tag</b>	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 WSR 1.6.0.

### Parent

<umswatsons>

### Children

None.

### Example

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

```
<utterance-manager
  save-waveforms="false"
  purge-existing="false"
  max-file-age="60"
  max-file-count="100"
  waveform-base-uri="http://localhost/utterances/"
  waveform-folder=""
/>
```

## 3.8 RDR Manager

This element specifies parameters of the Recognition Details Record (RDR) manager.

### Attributes

Name	Unit	Description
<b>save-records</b>	Boolean	Specifies whether to save recognition details records or not.
<b>purge-existing</b>	Boolean	Specifies whether to delete existing records on start-up.
<b>max-file-age</b>	Time interval [min]	Specifies a time interval in minutes after expiration of which a record is deleted. Set 0 for infinite.
<b>max-file-count</b>	Integer	Specifies the max number of records to store. If reached, the oldest record is deleted. Set 0 for infinite.
<b>record-folder</b>	Dir path	Specifies a folder to store recognition details records in. Defaults to <code>\${UniMRCPIInstallDir}/var</code> .
<b>file-prefix</b>	String	Specifies a prefix used to compose the name of the file to be stored. Defaults to 'umswatonsr-', if not specified.
<b>use-logging-tag</b>	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 WSR 1.6.0.



## Parent

<umswatsonsr>

## Children

None.

## Example

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

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

## 3.9 Monitoring Agent

This element specifies parameters of the monitoring agent.

### Attributes

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

## Parent

<umswatsonsr>

## 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="umswatonsr-channels.status"/>
</usage-refresh-handler >

</monitoring-agent>
```

## 3.10 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="umswatonsr-usage.status"/>
  <dump-channels enable="false" status-file="umswatonsr-channels.status"/>
</usage-change-handler>
```

## 3.11 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="umswatonsr-usage.status"/>  
  <dump-channels enable="false" status-file="umswatonsr-channels.status"/>  
</usage-refresh-handler>
```

## 3.12 License Server

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

### Attributes

Name	Unit	Description
<b>enable</b>	Boolean	Specifies whether the use of license server is enabled or not. If enabled, the license-file attribute is not honored.
<b>server-address</b>	String	Specifies the IP address or host name of the license server.
<b>certificate-file</b>	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.
<b>ca-file</b>	File path	Specifies the certificate authority used to validate the license server.
<b>channel-count</b>	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.
<b>http-proxy-address</b>	String	Specifies the IP address or host name of the

		HTTP proxy server, if used. Available since WSR 1.6.0.
<b>http-proxy-port</b>	Integer	Specifies the port number of the HTTP proxy server, if used. Available since WSR 1.6.0.

### Parent

<umswatsonsr>

### 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>

# 4 Configuration Steps

This section outlines common configuration steps.

## 4.1 Using Default Configuration

The default configuration should be sufficient for the general use.

## 4.2 Specifying Recognition Language

Recognition language can be specified by the client per MRCP session by means of the header field *Speech-Language* set in a *SET-PARAMS* or *RECOGNIZE* request. Otherwise, the parameter *language* set in the configuration file *umswatsonsr.xml* is used. The parameter defaults to *en-US*.

For supported languages and their corresponding codes, visit the following link.

<https://console.ibm.com/docs/services/speech-to-text/input.html#models>

Since WSR 1.4.0, the recognition language can also be set by the attribute *xml:lang* specified in the SRGS grammar. For example:

```
<?xml version="1.0" encoding="UTF-8"?>
<grammar mode="voice" root="transcribe" version="1.0"
  xml:lang="en-AU"
  xmlns="http://www.w3.org/2001/06/grammar">
  <meta name="scope" content="builtin"/>
  <rule id="transcribe"><one-of/></rule>
</grammar>
```

## 4.3 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 IBM Watson Speech-to-Text API.

## 4.4 Specifying Speech Input Parameters

While the default parameters specified for the speech input detector are sufficient for the general use, various parameters can be adjusted to better suit a particular requirement.

- `speech-start-timeout`

This parameter is used to trigger a start of speech input. The shorter is the timeout, the sooner a *START-OF-INPUT* event is delivered to the client. However, a short timeout may also lead to a false

positive. Note that if the *start-of-input* parameter in the *ws-streaming-recognition* is set to *service-originated*, then a START-OF-INPUT event is sent to the client at a later stage, upon reception of a first interim result.

- *speech-complete-timeout*

This parameter is used to trigger an end of speech input. The shorter is the timeout, the shorter is the response time. However, a short timeout may also lead to a false positive.

- *vad-mode*

This parameter is used to specify an operating mode of the Voice Activity Detector (VAD) within an integer range of [0 ... 3]. A higher mode is more aggressive and, as a result, is more restrictive in reporting speech. The parameter can be overridden per MRCP session by setting the header field *Sensitivity-Level* in a *SET-PARAMS* or *RECOGNIZE* request. The following table shows how the *Sensitivity-Level* is mapped to the *vad-mode*.

Sensitivity-Level	Vad-Mode
[0.00 ... 0.25)	0
[0.25 ... 0.50)	1
[0.50 ... 0.75)	2
[0.75 ... 1.00]	3

## 4.5 Specifying DTMF Input Parameters

While the default parameters specified for the DTMF input detector are sufficient for the general use, various parameters can be adjusted to better suit a particular requirement.

- *dtmf-interdigit-timeout*

This parameter is used to set an inter-digit timeout on DTMF input. The parameter can be overridden per MRCP session by setting the header field *DTMF-Interdigit-Timeout* in a *SET-PARAMS* or *RECOGNIZE* request.

- *dtmf-term-timeout*

This parameter is used to set a termination timeout on DTMF input and is in effect when *dtmf-term-char* is set and there is a match for an input grammar. The parameter can be overridden per MRCP session by setting the header field *DTMF-Term-Timeout* in a *SET-PARAMS* or *RECOGNIZE* request.

- *dtmf-term-char*

This parameter is used to set a character terminating DTMF input. The parameter can be overridden per MRCP session by setting the header field *DTMF-Term-Char* in a *SET-PARAMS* or *RECOGNIZE* request.

## 4.6 Specifying No-Input and Recognition Timeouts

- *noinput-timeout*

This parameter is used to trigger a no-input event. The parameter can be overridden per MRCP session by setting the header field *No-Input-Timeout* in a *SET-PARAMS* or *RECOGNIZE* request.

- input-timeout

This parameter is used to limit input (recognition) time. The parameter can be overridden per MRCP session by setting the header field *Recognition-Timeout* in a *SET-PARAMS* or *RECOGNIZE* request.

## 4.7 Specifying Vendor-Specific Parameters

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

Name	Unit	Description
<b>start-of-input</b>	String	Specifies the source of start of input event sent to the client (use "service-originated" for an event originated based on a first-received interim result and "internal" for an event determined by plugin). Available since WSR 1.5.0.
<b>alternatives-below-threshold</b>	Boolean	Specifies whether to return speech recognition result alternatives with the confidence score below the confidence threshold. Available since WSR 1.5.0.
<b>speech-start-timeout</b>	Time interval [msec]	Specifies how long to wait in transition mode before triggering a start of speech input event. Available since WSR 1.5.0.
<b>language-customization-id</b>	String	Specifies the custom language model id to be used for recognition, if any. Available since WSR 1.5.0.
<b>acoustic-customization-id</b>	String	Specifies the custom acoustic model id to be used for recognition, if any. Available since WSR 1.5.0.
<b>base-model-version</b>	String	Specifies the base model version to be used for recognition, if any. Available since WSR 1.5.0.
<b>grammar-name</b>	String	Specifies the grammar name used for speech recognition in conjunction with the customization id of the custom language model for which the grammar is defined. Available since WSR 1.6.0.
<b>customization-weight</b>	Double	Specifies the weight to give to words from the custom language model compared to those from the base model, if any. Available since WSR 1.5.0.
<b>smart-formatting</b>	Boolean	Specifies whether the service converts dates, times, numbers, currency, and similar values into more conventional representations in the final transcript.

		Available since WSR 1.5.0.
<b>word-confidence</b>	Boolean	Specifies whether the service returns a confidence for each word. Available since WSR 1.5.0.
<b>timestamps</b>	Boolean	Specifies whether the service returns time alignment for each word. Available since WSR 1.5.0.
<b>speaker-labels</b>	Boolean	Specifies whether the service returns labels identifying a speaker for each word. Available since WSR 1.5.0.
<b>redaction</b>	Boolean	Specifies whether the service returns masked numeric data having three or more consecutive digits. Available since WSR 1.5.0.
<b>processing-metrics</b>	Boolean	Specifies whether the service returns metrics on transcription of the input audio data. Available since WSR 1.5.0.
<b>processing-metrics-interval</b>	Time interval [sec]	Specifies an interval in seconds at which the service returns processing metrics. Available since WSR 1.5.0.
<b>audio-metrics</b>	Boolean	Specifies whether the service returns signal characteristics of the input audio data. Available since WSR 1.5.0.
<b>end-of-phrase-silence-time</b>	Time interval [sec]	Specifies an interval between 0.0 and 12.0 seconds at which the service splits a transcript into multiple final results if it encounters silence. Not used, if set to -1. Available since WSR 1.6.0.
<b>interim-result-timeout</b>	Time interval [msec]	Specifies a timeout between interim results containing transcribed speech. If the timeout is elapsed, input is considered complete. Available since WSR 1.16.0.

All the vendor-specific parameters can also be specified at the grammar-level via a built-in or SRGS XML grammar.

The following example demonstrates the use of a built-in grammar with the vendor-specific parameters *alternatives-below-threshold* and *speech-start-timeout* set to *true* and *100* correspondingly.

```
builtin:speech/transcribe?alternatives-below-threshold=true;speech-start-timeout=100
```

The following example demonstrates the use of an SRGS XML grammar with the vendor-specific parameters *alternatives-below-threshold* and *speech-start-timeout* set to *true* and *100* correspondingly.



```

<grammar mode="voice" root="transcribe" version="1.0" xml:lang="en-US"
xmlns="http://www.w3.org/2001/06/grammar">
  <meta name="scope" content="builtin"/>
  <meta name="alternatives-below-threshold" content="true"/>
  <meta name="speech-start-timeout" content="100"/>
  <rule id="transcribe">
    <one-of ><item>blank</item></one-of>
  </rule>
</grammar>

```

## 4.8 Maintaining Utterances

Saving of utterances is not required for regular operation and is disabled by default. However, enabling this functionality allows to save utterances sent to the Watson Speech service and later listen to them offline.

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

- save-waveforms

Utterances can optionally be recorded and stored if the configuration parameter *save-waveforms* is set to true. The parameter can be overridden per MRCP session by setting the header field *Save-Waveforms* in a *SET-PARAMS* or *RECOGNIZE* request.

- 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-base-uri

This parameter specifies the base URI used to compose an absolute waveform URI returned in the header field *Waveform-Uri* in response to a *RECOGNIZE* request.

- waveform-folder

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

## 4.9 Maintaining Recognition Details Records

Producing of recognition details records (RDR) is not required for regular operation and is disabled by default. However, enabling this functionality allows to store details of each recognition attempt in a separate file and analyze them later offline. The RDRs are stored in the JSON format.

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

- `save-records`

This parameter specifies whether to save recognition 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 *\${UniMRCPIInstallDir}/var*.

# 5 Recognition Grammars and Results

## 5.1 Using Built-in Speech Transcription

For generic speech transcription, having no speech contexts defined, a pre-set identifier *transcribe* must be used by the MRCP client in a RECOGNIZE request as follows:

```
builtin:speech/transcribe
```

The name of the identifier *transcribe* can be changed from the configuration file *umswatsons.xml*.

Speech contexts are defined in the configuration file *umsbingsr.xml* and available since BingSR 1.5.0. A speech context is assigned a unique string identifier and holds a list of phrases.

Below is a definition of a sample speech context *directory*:

```
<speech-context id="directory" speech-complete="true">
  <phrase>call Steve</phrase>
  <phrase>call John</phrase>
  <phrase>dial 5</phrase>
  <phrase>dial 6</phrase>
</speech-context>
```

Which can be referenced in a RECOGNIZE request as follows:

```
builtin:speech/directory
```

The prefixes *builtin:speech* and *builtin:grammar* can be used interchangeably as follows:

```
builtin:grammar/directory
```

Since WSR 1.4.0, a speech context can be referenced by means metadata in SRGS XML grammar. For example, the following SRGS grammar references a built-in speech context *directory*.

```
<grammar mode="voice" root="directory" version="1.0"
  xml:lang="en-US"
  xmlns="http://www.w3.org/2001/06/grammar">
  <meta name="scope" content="builtin"/>
  <rule id="directory"><one-of/></rule>
</grammar>
```

Where the root rule name identifies a speech context.

## 5.2 Using Built-in DTMF Grammars

Pre-set built-in DTMF grammars can be referenced by the MRCP client in a RECOGNIZE request as follows:

```
builtin:dtmf/$id
```

Where *\$id* is a unique string identifier of the built-in DTMF grammar.

Note that only a DTMF grammar identifier *digits* is currently supported.

Since WSR 1.4.0, built-in DTMF digits can also be referenced by metadata in SRGS XML grammar. The following example is equivalent to the built-in grammar above.

```
<grammar mode="dtmf" root="digits" version="1.0"
  xml:lang="en-US"
  xmlns="http://www.w3.org/2001/06/grammar">
  <meta name="scope" content="builtin"/>
  <rule id="digits"><one-of/></rule>
</grammar>
```

Where the root rule name identifies a built-in DTMF grammar.

## 5.3 Retrieving Results

Results received from the speech service are transformed to the [NLSML](#) format.

# 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 WSR 1.2.0.
- 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] WSR Usage: 0/0/2
```

## 6.2 Update Usage

The action *update-usage* writes the following data to a status file *umswatsonsr-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. Available since WSR 1.2.0.
- 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  
total channels: 2
```

## 6.3 Dump Channels

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

# 7 Usage Examples

## 7.1 Speech Transcription

This examples demonstrates how to perform speech recognition by using a RECOGNIZE request.

C->S:

```
MRCP/2.0 336 RECOGNIZE 1
Channel-Identifier: 6e1a2e4e54ae11e7@speechrecog
Content-Id: request1@form-level
Content-Type: text/uri-list
Cancel-If-Queue: false
No-Input-Timeout: 5000
Recognition-Timeout: 10000
Start-Input-Timers: true
Confidence-Threshold: 0.87
Save-Waveform: true
Content-Length: 25

builtin:speech/transcribe
```

S->C:

```
MRCP/2.0 83 1 200 IN-PROGRESS
Channel-Identifier: 6e1a2e4e54ae11e7@speechrecog
```

S->C:

```
MRCP/2.0 115 START-OF-INPUT 1 IN-PROGRESS
Channel-Identifier: 6e1a2e4e54ae11e7@speechrecog
Input-Type: speech
```

S->C:

```
MRCP/2.0 498 RECOGNITION-COMPLETE 1 COMPLETE
Channel-Identifier: 6e1a2e4e54ae11e7@speechrecog
Completion-Cause: 000 success
Waveform-Uri: <http://localhost/utterances/utter-6e1a2e4e54ae11e7-
1.wav>;size=20480;duration=1280
Content-Type: application/x-nlsml
```

Content-Length: 214

```
<?xml version="1.0"?>
<result>
  <interpretation grammar="builtin:speech/transcribe" confidence="0.95">
    <instance>what's the weather like</instance>
    <input mode="speech">what's the weather like</input>
  </interpretation>
</result>
```

## 7.2 DTMF Recognition

This examples demonstrates how to reference a built-in DTMF grammar in a RECOGNIZE request.

C->S:

```
MRCP/2.0 266 RECOGNIZE 1
Channel-Identifier: d26bef74091a174c@speechrecog
Content-Type: text/uri-list
Cancel-If-Queue: false
Start-Input-Timers: true
Confidence-Threshold: 0.7
Speech-Language: en-US
Dtmf-Term-Char: #
Content-Length: 19

builtin:dtmf/digits
```

S->C:

```
MRCP/2.0 83 1 200 IN-PROGRESS
Channel-Identifier: d26bef74091a174c@speechrecog
```

S->C:

```
MRCP/2.0 113 START-OF-INPUT 1 IN-PROGRESS
Channel-Identifier: d26bef74091a174c@speechrecog
Input-Type: dtmf
```

S->C:



MRCP/2.0 382 RECOGNITION-COMPLETE 1 COMPLETE

Channel-Identifier: d26bef74091a174c@speechrecog

Completion-Cause: 000 success

Content-Type: application/x-nlsml

Content-Length: 197

```
<?xml version="1.0"?>
```

```
<result>
```

```
  <interpretation grammar="builtin:dtmf/digits" confidence="1.00">
```

```
    <input mode="dtmf">1 2 3 4</input>
```

```
    <instance>1234</instance>
```

```
  </interpretation>
```

```
</result>
```

### 7.3 Speech and DTMF Recognition

This examples demonstrates how to perform recognition by activating both speech and DTMF grammars. In this example, the user is expected to input a 4-digit pin.

C->S:

MRCP/2.0 275 RECOGNIZE 1

Channel-Identifier: 6ae0f23e1b1e3d42@speechrecog

Content-Type: text/uri-list

Cancel-If-Queue: false

Start-Input-Timers: true

Confidence-Threshold: 0.7

Speech-Language: en-US

Content-Length: 47

```
builtin:dtmf/digits?length=4
```

```
builtin:speech/pin
```

S->C:

MRCP/2.0 83 2 200 IN-PROGRESS

Channel-Identifier: 6ae0f23e1b1e3d42@speechrecog

S->C:

MRCP/2.0 115 START-OF-INPUT 2 IN-PROGRESS

Channel-Identifier: 6ae0f23e1b1e3d42@speechrecog

Input-Type: speech

S->C:

MRCP/2.0 399 RECOGNITION-COMPLETE 2 COMPLETE

Channel-Identifier: 6ae0f23e1b1e3d42@speechrecog

Completion-Cause: 000 success

Content-Type: application/x-nlsml

Content-Length: 214

```
<?xml version="1.0"?>
```

```
<result>
```

```
  <interpretation grammar=" builtin:speech/pin" confidence="1.00">
```

```
    <instance>one two three four</instance>
```

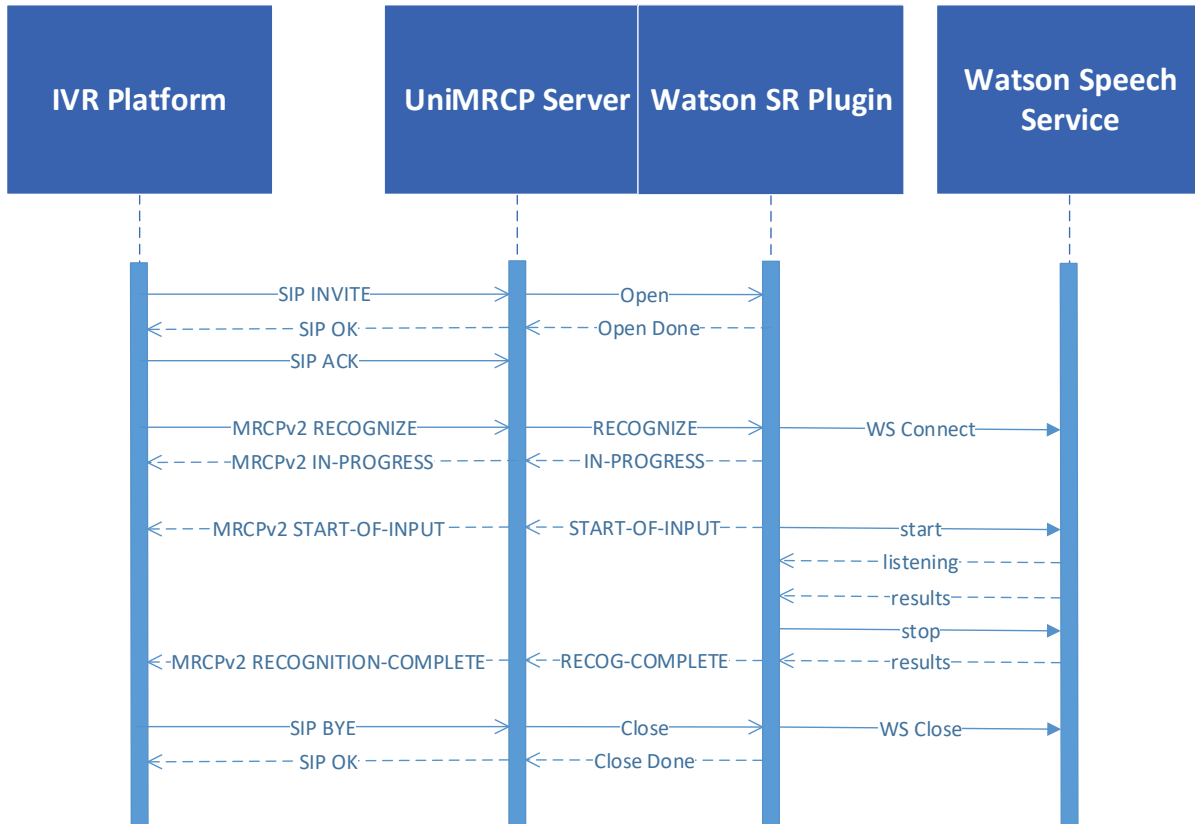
```
    <input mode="speech">one two three four</input>
```

```
  </interpretation>
```

```
</result>
```

# 8 Sequence Diagram

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



# 9 References

## 9.1 IBM Watson

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

## 9.2 Specifications

- [Speech Recognizer Resource](#)
- [NLSML Results](#)