UniMRCP Google Speech

Plugin Proposal

- Created: February 14, 2017
- Last updated: February 14, 2017
- Author: Arsen Chaloyan
- Vendor: Universal Speech Solutions, LLC

Contents

1 Overview

This document is composed in response to a request for quotation (RFQ).

The RFQ has been placed for development of a speech recognition plugin to the UniMRCP server. The plugin is supposed to extend the functionality of the server with the services provided by the Google Speech API.

The document contains deliverable components categorized as *REQUIRED* and *OPTIONAL*. For each of the components, there is a completion timeline expressed in an hourly basis.

2 Framework

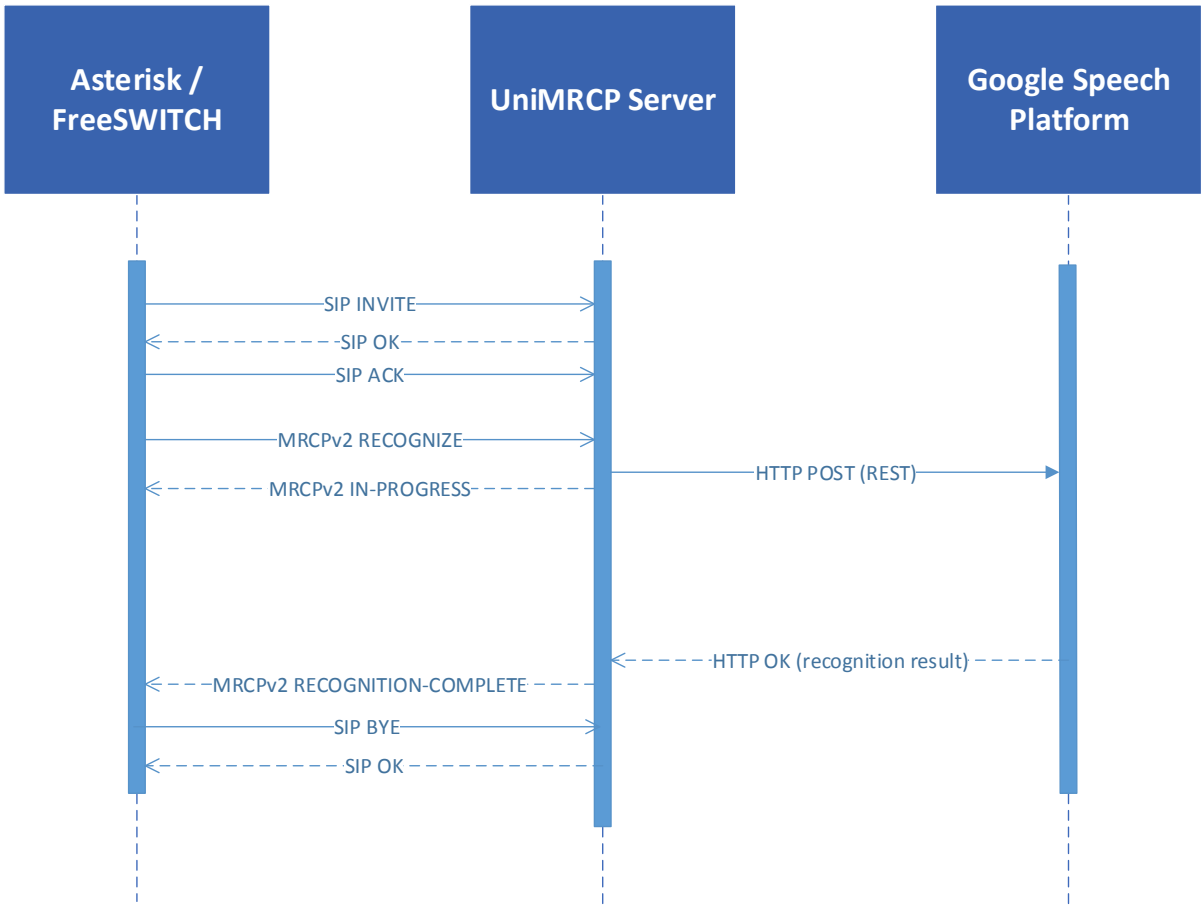
Description

This section provides a high level overview of iteraction between

- an IVR platform (Asterisk, FreeSWITCH, …)
- the UniMRCP server
- the Google Speech API.

The UniMRCP server receives an MRCP speech recognition request from the IVR platform and places a corresponding HTTP/REST or gRPC request to the Google Speech Platform. The Google Speech Platform performs recognition and returns results back to the server.

The following is a typical sequence diagram outlining the described session utilizing the HTTP/REST interface.



Deliverable

Component	Timeline (hours)	Category
ASR Plugin Framework	16	REQUIRED
HTTP REST API	16	OPTIONAL
gRPC API	24	OPTIONAL

Note: support for at least one of HTTP REST API or gRPC API is required.

3 Recognition Modes

Description

The following modes of recognition are supported by the Google Speech API.

- Synchronous Recognition (REST and gRPC)
- Asynchrounous Recognition (REST and gRPC)
- Streaming Recognition (gRPC only)

Deliverable

Component	Timeline (hours)	Category
Synchronous Recognition	8	OPTIONAL
Asynchronous Recognition	16	OPTIONAL
Streaming Recognition	16	OPTIONAL

Note: support for at least one of the recognition methods is required.

4 Audio Encodings

Description

Support for the Linear 16-bit Pulse-Code Modulation (PCM) and the MULAW encoding is included by default. In order to support additional audio encodings listed below, a 3-rd party library needs to be utilized.

Name	Codec
Free Lossless Audio Codec	FLAC
Adaptive Multi-Rate Narrowband	AMR
Adaptive Multi-Rate Wideband	AMR_WB

Deliverable

Component	Timeline (hours)	Category
Extended Audio Encodings	24	OPTIONAL

5 Sample Rates

Description

Support for the audio that samples 8 kHz is included by default. The 16 kHz sample rate can be supported in addition.

Deliverable

Component	Timeline (hours)	Category
Extended Sample Rate (16 kHz)	8	OPTIONAL

6 Audio URI Reference

Description

Support for embedding audio in speech recognition requests is included by default. Additionally, audio can be referenced by a URI.

Deliverable

Component	Timeline (hours)	Category
Audio URI Reference	8	OPTIONAL

7 Phrase Hints

Description

Any given recognition request may contain a speech context that provides information to aid in processing the given audio by improving the recognition accuracy.

The speech context might be transparently passed from the MRCP client to the server in the format supported by the Google Speech Platform or be converted from of one of the well-known recognition grammar formats such as SRGS.

Deliverable

Component	Timeline (hours)	Category
Speech Context	8	OPTIONAL
SRGS Conversion	24	OPTIONAL

Note: it may not be feasible to convert an arbitrary SRGS grammar to the speech context format supported by the Google Speech Platform.

Note: support for all the languages provided by the Google Speech Platform is included by default, no additional component is required.

8 Recognition Results

Description

Recognition results returned from the Google Speech Platform are in the JSON format. The results may need to be converted to NLSML in order to provide unified and standards-conformed results to the IVR application.

Deliverable

Component	Timeline (hours)	Category
NLSML Conversion	16	OPTIONAL

9 Installation Package

Description

Binary packages allow convenient deployment.

Deliverable

Component	Timeline (hours)	Category
Setup/RPM Installer	40	OPTIONAL

10 Documentation

Description

Basic build and installation instructions are provided by default. More complete documentation can be composed and provided in addition.

Deliverable

Component	Timeline (hours)	Category
Documentation	16	OPTIONAL

11 Summary

The following table summarizes the details of two possible projects: *basic* and *advanced*. While the *basic* project consists of the required components only, the *advanced* project includes all the required and optional components listed in the document.

Component	Timeline (hours)	Basic	Advanced
ASR Plugin Framework	16	x	x
HTTP REST API	16	x	x
gRPC API	24		x
Synchronous Recognition	8	x	x
Asynchronous Recognition	16		x
Streaming Recognition	16		x
Extended Audio Encodings	24		x
Extended Sample Rate (16 kHz)	8		x
Audio URI Reference	8		x
Speech Context	8		x
SRGS Conversion	24		x
NLSML Conversion	16		x
Setup/RPM Installer	40		x
Documentation	16		x
Total (hours)		40	240
Start Date		Beg of Apr, 2017	Beg of Apr, 2017
Completion Date		End of Apr, 2017	N/A