API Reference Guide

AudioCodes Speech-to-Text Transcription Engine for Voice.AI

AudioCodes Speech REST Application Programming Interface (API)

Version 0.5

audiocodes

Table of Contents

Notice

Information contained in this document is believed to be accurate and reliable at the time of printing. However, due to ongoing product improvements and revisions, AudioCodes cannot guarantee accuracy of printed material after the Date Published nor can it accept responsibility for errors or omissions. Updates to this document can be downloaded from https://www.audiocodes.com/library/technical-documents.

This document is subject to change without notice.

Date Published: June-20-2023

Customer Support

Customer technical support and services are provided by AudioCodes or by an authorized AudioCodes Service Partner. For more information on how to buy technical support for AudioCodes products and for contact information, please visit our website at

https://www.audiocodes.com/services-support/maintenance-and-support.

Stay in the Loop with AudioCodes

Abbreviations and Terminology

Each abbreviation, unless widely used, is spelled out in full when first used.

Related Documentation

Document Name
LTRT-26007 AudioCodes Automatic Speech Recognition - WebSocket API (v0.49)
LTRT-26008 AudioCodes Speech – LVCSR WebSocket API (v0.59)
LTRT-26004 AudioCodes Speech –Speaker Recognition Enrollment and Segment WebSocket API (v0.11)

Document Revision Record

LTRT	Description
26003	Initial document release for Version 0.5.
26009	Added two API parameters.

Documentation Feedback

AudioCodes continually strives to produce high quality documentation. If you have any comments (suggestions or errors) regarding this document, please fill out the Documentation Feedback form on our website at https://online.audiocodes.com/documentation-feedback.

1 Introduction

This document describes how to utilize AudioCodes' Large Vocabulary Continuous Speech Recognition (LVCSR) technology and the process of speaker diarization suitable for offline Speech-To-Text (STT) in various of use cases, where speaker diarization is required. The technology operates via REST-based protocol API, which is explained below. The document provides details of API parameters that governs the operation.

2 API Reference

2.1 Offline Transcribe API

The `<Speech_Server_IP>/v1/speech:transcribe/{context}` URL when used with the POST method, provides the ability for the transcription client to send a request to the server to transcribe and diarize an audio file.

REST Resource

```
<Speech_Server_IP>/v1/speech:transcribe/{context}
```

HTTP Method

```
POST
```

Content-Type

```
application/json
```

Path Variables

Attribute	Type	Description
`Context`	String	Indicator of context to transcribe with. For default generic model use ***sttg***

Request Message Body

Fields	Description
String : audio-file String : accept-language String: content-type String: cookie Integer: diarization-gap Integer: functionality-bits Integer: save-waveform String: adhoc-glossary-id Array: adhoc-glossary (see below - optional) Object: adhoc-glossary-lexicon (see below - optional) Integer: adhoc-glossary-sensitivity Integer: adhoc-glossary-strictness	**audio-file** base64 audio file **accept-language** speech recognition language. **Content-type** mime type of the **audio-file** one of: ■ audio/l16;rate=8000;channels=1 ■ audio/l16;rate=16000;channels=1 ■ audio/PCMA;rate=8000;channels=1 ■ audio/PCMA;rate=16000;channels=1 ■ audio/PCMU;rate=8000;channels=1 ■ audio/PCMU;rate=16000;channels=1 ■ audio/l16;rate=8000;channels=2 ■ audio/l16;rate=16000;channels=2 ■ audio/PCMA;rate=8000;channels=2 ■ audio/PCMA;rate=16000;channels=2 ■ audio/PCMU;rate=8000;channels=2 ■ audio/PCMU;rate=16000;channels=2 **cookie** optional application session labeling **diarization-gap** an optional parameter that governs merge segments of same speaker , where gap is less than the parameter given in [msec]. the default value is 20000 [msec]. **functionality-bits** an optional bitwise control parameter that governs transcribe operation. **Bit 0 (LSB)** 0 – default operation, speech to text in parallel to speaker diarization with joined post processing, 1 – speech to text prior and as input to speaker diarization with joined post processing. **Bit 1** 0 – End Of Sentence (EOS) is not reported 1 – EOS is reported. **Bit 2** 0 – internal activity detection default mode 1 – disable speaker diarization internal activity detection **save-waveform** an optional parameter that enables (value 1) or disable (value 0) waveform save at server side (usually used for debugging purposes) **adhoc-glossary-id** an optional parameter that make use of pre-compiled adhoc glossaries, the id of glossary is retrieved from other session , reported in adhoc-glossary-id-tag **adhoc-glossary-strictness** an optional parameter that determines how flexible the glossary could be spoken (e.g., saying "Open Word" instead of "Open Microsoft Word for Windows" and prefix & suffix flexibility). Value range: 0 to 100. Where 100 restricts recognition to the exact full phrases and exact prefix or suffix used). The value selection policy is highly dependent on the ability to characterize the recognizer actual inputs in the service. Changing the parameter value employs glossary compilation. **adhoc-glossary-sensitivity** an optional parameter that determines the sensitivity in holding to the glossary defined phrases/words rather than generic speech. Value range: 0 to 100. Where 100 sensitivity is more attuned to the glossaries phrases (low rate of detection but higher accuracy), and 0 leads to less recognition to glossaries phrases and allows for more generic speech (high rate of detection but lower accuracy). Changing the parameter value does not employ glossary compilation.
adhoc-glossary	Array of strings
adhoc-glossary-lexicon	String: word Array: transcriptions (see below)
transcriptions	Array of strings

Reply Content-Type

```
application/json; charset=utf-8
```

Reply Message Body

Entity	Fields	Description
transcription	Array : *objectarray1 (see below)* String: cookie String: waveform-tag String: adhoc-glossary-id-tag	 **cookie** application session labeling **waveform-tag** uri path to server waveform recording. the entity appears only if save-waveform was enabled. **adhoc-glossary-id-tag** uri path to server pre-compiled glossary. the entity appears only if adhoc glossary were used in the session.
objectarray1	String: id Int: location Int: duration String: text Array: words *(see below)*	**id** diarization speaker label **location** frame index where diarized speaker segment starts in audio frames (x10msec) **duration** period of diarized speaker lasts in audio frames (x10msec) **text** speech to text words sequence recognized under the diarized segment **words** speech to text words detailed information including location, duration , confidence and text
Words	Array : *objectarray2 (see below)*	
objectarray2	String: word Int: location Int: duration Float: confidence	**word** the text representing the word under the segment **location** frame index where word starts in audio frames (x10msec) **duration** period of word lasts in audio frames (x10msec) **confidence** word level confidence score in the range 0.0 to 1.0

HTTP Response

- 200 OK

2.2 Offline Transcribe API response example

```
{
 "transcription": [
   {
     "id": "Anonymous-Speaker-1",
     "location": 250,
     "duration": 170,
     "text": "word1 word2 word3",
     "words": [
       {
```

```
            "word": "word1",
            "location": 250,
            "duration": 40,
            "confidence": 0.5854
          },
          {
            "word": "word2",
            "location": 320,
            "duration": 50,
            "confidence": 0.6854
          },
          {
            "word": "word3",
            "location": 370,
            "duration": 30,
            "confidence": 0.7854
          }
        ]
      },
      {
        "id": "Anonymous-Speaker-2",
        "location": 800,
        "duration": 250,
        "text": "word3 word4 word5",
        "words": [
          {
            "word": "word3",
            "location": 800,
            "duration": 80,
            "confidence": 0.2854
          },
          {
            "word": "word4",
            "location": 880,
            "duration": 60,
            "confidence": 0.4454
          },
          {
            "word": "word5",
            "location": 1000,
            "duration": 30,
            "confidence": 0.9854
          }
        ]
      }
   ]
"cookie" : "application defined cookie"
}
```

International Headquarters
1 Hayarden Street,
Airport City
Lod 7019900, Israel
Tel: +972-3-976-4000
Fax: +972-3-976-4040

AudioCodes Inc.
80 Kingsbridge Rd
Piscataway, NJ 08854, USA
Tel: +1-732-469-0880
Fax: +1-732-469-2298

Contact us: https://www.audiocodes.com/corporate/offices-worldwide
Website: https://www.audiocodes.com

Document #: LTRT-26009