



**REVERIE**

Language Technologies

## API REFERENCE DOCUMENT

For

**REVERIE ASR API**



# Abstract

Reverie ASR API Reference Document

You can use our API to convert the speech into Indic language scripts. This document describes the available variables, commands, and interfaces that make up the Reverie ASR API.

This document will help the developers to implement Automated Voice Recognition (ASR) in their applications. The API is platform agnostic, which means any device that can record the speech can use this API.

The API is organized around Websocket. All requests should be made over WSS. All requests and responses, including errors, are delivered in JSON format.

# Table of Contents

- 1. OVERVIEW1
  - 1.1. LANGUAGES SUPPORTED1
  - 1.2. KEY FEATURES1
- 2. CONTINUOUS SPEECH RECOGNITION API2
  - 2.1. PREREQUISITES2
  - 2.2. GETTING STARTED2
  - 2.3. VERSIONING2
  - 2.4. HOW IT WORKS3
  - 2.5. SAMPLE CODE4
  - 2.6. API REFERENCE4
    - 2.6.1 Secured WebSocket Request URL4
    - 2.6.2 Parameter4
    - 2.6.3 Request Post Body5
    - 2.6.4 Response5
    - 2.6.5 HANDLING ERRORS6
    - 2.6.6 API Final Output Cause6
- 3. AUDIO FILE TRANSCRIPTION API7
  - 3.1. PREREQUISTES7
  - 3.2. GETTING STARTED7
  - 3.3. VERSIONING7
  - 3.4. SAMPLE CODE7
  - 3.5. API REFERENCE8
    - 3.5.1 HTTP Request URL8
    - 3.5.2 Parameter8
    - 3.5.3 Request Post Body8
    - 3.5.4 Response8
    - 3.5.5 HANDLING ERRORS8
    - 3.5.6 Messages8

## 1. Overview

The Reverie ASR solution is an Indic language speech recognition API that transcribes - your speech in real-time and the recorded audio file, and return in the text format

The AI-powered speech to text API is a fully managed and continually trained solution, which leverages machine learning to combine knowledge of grammar, language structure, and the composition of audio and voice signals to accurately transcribe the human voice. The solution understands and transcribes in 11 mostly used Indian languages and Indian English.

### 1.1. Languages Supported

Reverie ASR API is the only solution that it understands regional accents, bi-lingual nature of Indians, and is dialect-agnostic. Currently, the solution can understand two languages:

Language	ISO Language Code used in API
1. Hindi	hi
2. Indian English	en

The languages for which we are for which are planning to release shortly are:

Language	ISO Language Code used in API
1. Gujarati	gu
2. Telugu	te
3. Bengali	bn
4. Marathi	mr

The languages on which we are working continuously and included in our pipeline arewhich are in our pipeline are:

1. Tamil	2. Kannada	3. Malayalam	4. Telugu
5. Odia	6. Punjabi	7. Assamese	

### 1.2. Key Features

Reverie ASR PI has some powerful features that help you to serve your consumers effectively in their native language:

- **Accurate & Live Streaming Transcription**

Reverie ASR is a real-time solution that transcribes the sentences spoken over the microphone. It supports multiple Indic languages and the Indian English language. It can spot specified keywords in real-time with high accuracy and confidence, even from lower quality audio file.

- **Noise Vitality**

The solution handles moderate noisy audio data recorded in various environments without requiring additional noise cancellation.

- **Customizable Vocabulary Model**

Tailor the speech recognition vocabulary to generate accurate transcriptions specific to your use-case, such as product names, domain-specific terminology, or names of individuals.

- **One Solution for Multiple Use-cases**

You can build any speech application on any platform powered by a single, trusted speech recognition engine. Easily integrates with all the applications developed on multiple platforms. You can obtain a transcription of customer service calls, interact with voice bots, and generate subtitles on audio files.

## 2. Continuous Speech Recognition API

### 2.1. Prerequisites

The prerequisites to set-up and use Continuous Speech Recognition API solution is given below:

Continuous Streaming Transcription:	
Supported Audio Format:	WAV
Audio Sampling Rate:	Signed 16 bit, 16,000 or 16K Hz
WSS Port	default
No. of channels	1
Speech Recognition Type	Continuous Speech Recognition
Time Limit	You can pass a maximum of 10-second audio data to the service with a single request.

### 2.2. Getting Started

Reverie ASR API establishes the connection using the **Secure WebSocket** interface to convert your speech into text. It uses standard HTTP response codes, and verbs. For each account, the default settings like default language code, domain, etc.

### 2.3. Versioning

When we make backward-incompatible changes to the API, we release new dated versions. The current version is **v1.0**.

All requests will use your account API settings unless you override the API version. Events generated by API requests will always be structured according to your account API version.

## 2.4. How it works

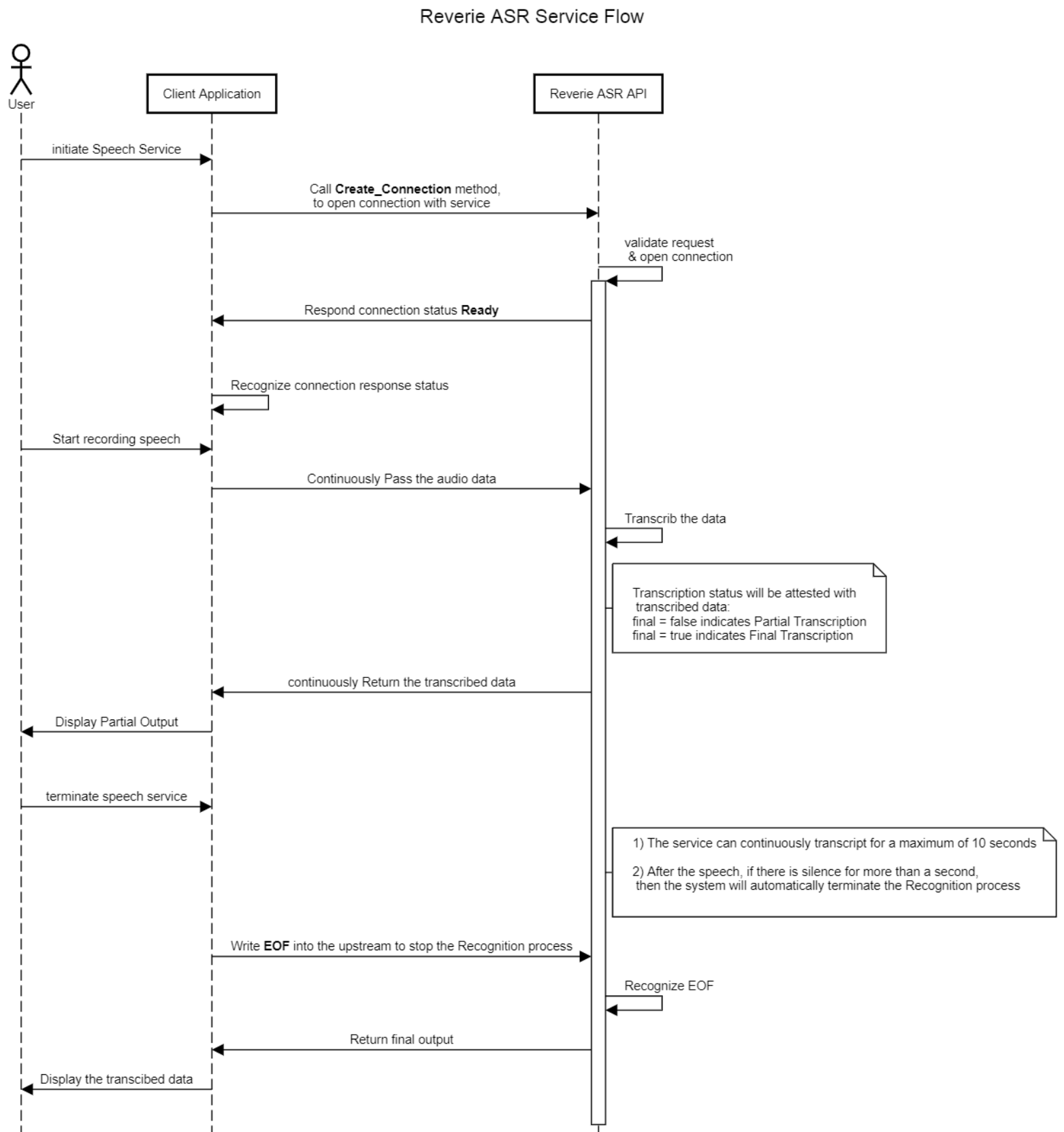


Figure 1: Reverie ASR Sequential Process Flow

The process flow to transcribe the audio file is:

1. Open the connection with the service using the **Create\_Connection** method
2. In the API response, if the key value pair **cause = Ready**, then the connection is successfully established
3. Write the speech data into the upstream and continuously receive the transcribed data

**Note:** In the response, if the **final = false**, then this denotes that the file is partially transcribed and the service is still processing the input speech

4. Write **EOF** into the upstream, to stop the recognition process

**Note:** The format to write EOF is **--EOF--**

**Note:** If you fail to write EOF into the upstream, then the ASR API will automatically terminate the recognition process.

Below are the scenarios when the API will terminate the recognition process:

1. After 10 seconds of recording time
2. After starting recording, if you maintain silence for more than 1 second

5. In the API response, if the **final = true**, then the text received should be considered as the final transcript

## 2.5. Sample Code

### Initiating Speech Service:

```
wss://revasr.reverieinc.com/stream/hi
```

### On successfully validating and establishing the connection:

```
{"id": "bb261bd789af4ba487a2667f8d942d4d7e0195fd1c8e4073", "success": true, "text": "", "final": false, "confidence": 1.0, "cause": "ready"}
```

### In-between an utterance, the partial result:

```
{"id": "bb261bd789af4ba487a2667f8d942d4d7e0195fd1c8e4073", "success": true, "text": "वन टू थ्री", "final": false, "confidence": 0.797274, "cause": "partial"}
```

### The final successful response:

```
{"id": "bb261bd789af4ba487a2667f8d942d4d7e0195fd1c8e4073", "success": true, "text": "वन टू थ्री को कॉल लगाओ", "final": true, "confidence": 0.743304, "cause": "EOF received"}
```

### The final error response:

```
{"id": "c5ba32954e094ae2abd2da0f8808c125b9e1490faa904f38", "success": false, "text": "", "final": true, "confidence": 1.0, "cause": "invalid api_key"}
```

## 2.6. API Reference

### 2.6.1 Secured WebSocket Request URL

#### URL Elements:

#### Example:

```
wss://revasr.reverieinc.com/stream/hi
```

### 2.6.2 Parameter

Parameter	Type	Is Mandatory?	Description
-----------	------	---------------	-------------

lang	string	Yes	The language code to which the audio has to be transcribed. For more information on supported languages and respective language code, refer section <a href="#">1.1 Languages Supported</a> <a href="#">above</a>
------	--------	-----	--

### 2.6.3 Request Post Body

Element	Type	Is Mandatory?	Description
Streaming Audio	Binary	Yes	The audio spoken streamed from the input device.

### 2.6.4 Response

Element	Type	Description
id	string	A unique Identity number auto-assigned by the API for each request
success	boolean	Will indicate the functional status of the API: <ul style="list-style-type: none"> <li>If the success is True, then the API is functioning and ready to generate output</li> <li>If the success is False, then the API is not functional and has some errors</li> </ul>
final	boolean	Will indicate whether the output text received is the final output or the partial: <ul style="list-style-type: none"> <li>If the final is True, then that is the final output</li> <li>If the final is False, then that is the partial output</li> </ul>
text	string	The audio sample transcribed into text format in the requested language. Empty in case of error or on connect.
confidence	float	The score lies ranges from 0 to 1. The higher the confidence score indicates the accuracy level of the output.



Element	Type	Description
cause	string	<p>The reason for the final output generated by the API. This is displayed for both successful and failed requests.</p> <p>Refer section 2.6.6 API Final Output Cause below to view the list of causes and description for the cause</p>

### 2.6.5 HANDLING ERRORS

Our API raises exceptions for many reasons, such as a failed connection, invalid parameters, authentication errors, and network unavailability. We provide more specific machine-readable messages with an error response so that users can react to errors more effectively. In the API response, if the **success=false**, then the **cause** will display the reason for the failure occurred. Refer section 2.6.6 API Final Output Cause below to view the list of causes and description for the cause

### 2.6.6 API Final Output Cause

Cause	Description
Ready	The connection between the client system and server is established, and you can pass the audio input.
EOF received	The input is provided and then stopped respectively by the user successfully.
Partial	The output received by the API is partial
Silence detected	After starting recording, the silence was recognized for more than 1 second, and the API has terminated the connection
Too many samples	The data input received length exceed 10 seconds
Timeout	The connection was established and left idle for 10 seconds
CSR limit exhausted	After the completion of the subscribed transcription limit
Language not supported	The entered language code is incorrect

### 3. Audio File Transcription API

#### 3.1. Prerequisites

Supported Audio Format:	WAV, MP3 and M4A file formats
Audio Sampling Rate:	Signed 16 bit, 16,000 or 16K Hz
Speech Recognition Type	Monologue / Single Speaker Speech

#### 3.2. Getting Started

Our Transliteration API is organized around REST and uses standard HTTP response codes, authentication, and verbs.

#### 3.3. Versioning

When we make backward-incompatible changes to the API, we release new dated versions. The current version is **v1.0**.

All requests will use your account API settings unless you override the API version. Events generated by API requests will always be structured according to your account API version.

#### 3.4. Sample Code

Initiating Speech Service:

```
https://revasr.reverieinc.com/api/file/hi
```

The final successful response:

```
{
  "status": "ok",
  "result": "रेवरी लॅंग्वेज टेक्नोलॉजीस की स्थापना 2009 में हुई थी। कंपनी मुख्यालय बेंगलोर में स्थित है। ",
  "message": "Success"
}
```

The final error response:

```
{
  "status": "error",
  "result": "",
  "message": "File type png is not allowed. Upload a wav,mp3,m4a or mp4 file"
}
```

## 3.5. API Reference

### 3.5.1 HTTP Request URL

#### URL Elements:

#### Example:

`https://revasr.reverieinc.com/api/file/hi`

### 3.5.2 Parameter

Parameter	Type	Is Mandatory?	Description
lang	string	Yes	The language code to which the audio has to be transcribed. For more information on supported languages and respective language code, refer section <a href="#">1.1 Languages Supported above</a>

### 3.5.3 Request Post Body

Element	Type	Is Mandatory?	Description
file	file	Yes	The file that should be converted into the text format  <i><a href="#">Note:</a> The file format supported are: WAV, MP3 or M4A format</i> <i><a href="#">Note:</a> This is a “form” POST request, encoding type should be ‘multipart/form-data’</i>

### 3.5.4 Response

Element	Type	Description
status	string	Will indicate the functional status of the API: <ul style="list-style-type: none"><li>If the status is <b>ok</b>, then the API is functioning and ready to generate output</li><li>If the status is <b>error</b>, then the API is not functional and has some errors</li></ul>
result	string	The audio sample transcribed into text format in the requested language. Empty in case of error or on connect.
message	string	The reason for the error or successfully completing the transcription. Refer section 3.5.6 Messages below to view the error message list
confidence	float	The score lies ranges from 0 to 1. The higher the confidence score indicates the accuracy level of the output.

### 3.5.5 HANDLING ERRORS

Our API raises exceptions for many reasons, such as a failed connection, invalid parameters, authentication errors, and network unavailability. We provide more specific machine-readable messages with an error response so that users can react to errors more effectively. In the API response, if the **success=false**, then the **cause** will display the reason for the failure occurred. Refer section 3.5.6 Messages below to view the list of causes and description for the cause

### 3.5.6 Messages

Cause	Description
Unsupported language code	The entered language code is incorrect
file parameter not found	The file is not uploaded

Cause	Description
File type XXX is not allowed. Upload a WAV,mp3,m4a or mp4 file	The uploaded file is not supported by the API
Uploading this file will exhaust the limit, please upload a file smaller file	The uploaded file duration is greater than the time subscribed for transcription
Success	The file transcribed successfully.