COTOHA Speech Recognition

Please note that this API is not available for `developer` accounts.
COTOHA Speech Recognition API consists of the following six API methods:

1. [Speech Recognition from File](#)

   - API for transcribing short audio files.
2. [Speech Recognition from Stream](#)

   - API for transcribing streaming audio, such as long audio files or input from a microphone.
3. [Update Speech Recognition Dictionary](#)※

   - API for adding words to the default dictionary.
   - The dictionary data is reflected every hour on the `0th` minute (e.g 0800, 1100, 2300). It takes a certain amount of time for the data to actually be reflected.
4. [Delete Speech Recognition Dictionary](#)※

   - API for deleting registered dictionary.
   - The dictionary data is reflected every hour on the `0th` minute (e.g 0800, 1100, 2300). It takes a certain amount of time for the data to actually be reflected.
5. [Get Speech Recognition Dictionary State](#)※

   - API for checking if the dictionary data is reflected.
6. [Download speech Recognition Dictionary](#)※

   - API for downloading the last uploaded dictionary.

※Japanese models only

GitHub

Following are sample codes released on GitHub by the Speech Recognition Team.
[GitHub](#)

- [Python](#)
- [Node.js](#)
- [JavaScript](#)
- [Reference](#)

Use cases

- Speech Recognition from File

  - If the audio is longer than `60` seconds, please use [Speech Recognition from Stream](#).
  - Otherwise, [Speech Recognition from File](#).
- Speech Recognition in Real Time

  - Please use [Speech Recognition from Stream](#).
- Speech Recognition from WAV Files

  - If the audio format is wav, please remove the header properly and convert to Linear PCM or μ-Law format.

- Please refer to this [Sample Code](#), with wav file input.
- Register a Dictionary
  - If you want to register words in advance, please use [Update Speech Recognition Dictionary](#).
  - If you want to change the dictionary for each speech recognition sequence, please use [Temporary Dictionary](#).

Speech Recognition from File

API for transcribing short audio files.
The length of the audio is limited up to `60` seconds.
Please use [Speech Recognition from Stream](#) if audio is longer than `60` seconds.

HTTP Request

```
POST <API Base URL>/asr/v1/speech_recognition/<ASR Model id>
```

`ASR Model id` is the ID that identifies the model used for speech recognition.
Refer to [List of Models](#) to select the appropriate model.

Request Header

You need to send a multipart form. Define a boundary delimiter and indicate it as "boundary" in your request header.

Key Name	Description
Content-Type	multipart/form-data; boundary= <Boundary Delimiter>
Authorization	Bearer [Access Token]

Request Body

Three parts are required for the request body.

1. Parameter Part
2. Audio Data Part
3. Command Part

The data structure to be set for each part of the multipart is specified in each `Part Type`.
`Part Type` is set with the `Content-Disposition` name parameter.
Body part must be set **in the order** of Parameter Part, Audio Data Part then Command Part.

Part Type	Required	Name	Content-Type
Parameter Part	Required	parameter	application/json; charset=UTF-8
Audio Data Part	Required	audio	application/octet-stream
Command Part	Required	command	application/json; charset=UTF-8

Parameter Part

Refer to [Request to Start Speech Recognition](#) for details.

Audio Data Part

Convert the audio data to one of the following formats and use the converted binary for this part.

※If the audio format is wav, please remove the header property and convert it to Linear PCM or µ-Law format.

Format	Sample Rate[Hz]	Quantization[bit]	Channel	Byte Order
Linear PCM	more than Model's rate (8000 or 16000)※	16	1	Little Endian
µ-Law	8000	8	1	Little Endian

※ Please note that we cannot guarantee correct operation for audio other than 8000, 11025, 16000, 22050, 32000, 44100, 48000, 88200, 96000[Hz]

Command Part

See section [Request to Stop Speech Recognition](#) for details.

Sample Request

HTTP Header

```
Content-Type: multipart/form-data; boundary=<Boundary Delimiter>
Authorization: Bearer <Access Token>
```

HTTP Body

```
--<Boundary Delimiter>
Content-Disposition: form-data; name="parameter"
Content-Type: application/json

{
  "msg":
  {
    "msgname": "start"
  },
  "param":
  {
    "baseParam.samplingRate": 16000,
    "recognizeParameter.domainId": "<ASR Domain id>"
    "recognizeParameter.enableContinuous": true
  }
}
--<Boundary Delimiter>
Content-Disposition: form-data; name="audio"
Content-Type: application/octet-stream
```

```
<Binary Audio Data>
--<Boundary Delimiter>
Content-Disposition: form-data; name="command"
Content-Type: application/json

{
  "msg": {
    "msgname": "stop"
  }
}
--<Boundary Delimiter>--
```

Response

Response Sample

```
[ {
  "msg" : {
    "msgname" : "started",
    "uniqueId" : "4d97031a-cfa9-4d66-968a-be708644a893"
  }
}, {
  "msg" : {
    "msgname" : "speechStartDetected",
    "uniqueId" : "4d97031a-cfa9-4d66-968a-be708644a893"
  },
  "timeinfo" : {
    "startDetectTime" : 0
  }
}, {
  "msg" : {
    "msgname" : "speechEndDetected",
    "uniqueId" : "4d97031a-cfa9-4d66-968a-be708644a893"
  },
  "timeinfo" : {
    "endDetectTime" : 3460
  }
}, {
  "msg" : {
    "msgname" : "recognized",
    "uniqueId" : "4d97031a-cfa9-4d66-968a-be708644a893"
  },
  "result" : {
    "type" : 1,
    "sentence" : [ {
      "surface" : "これ は テスト 用 の 音声ファイル です",
      "score" : 0.975848,
      "startTime" : 0.0,
      "endTime" : 3.46
    } ]
  }
}, {
  "msg" : {
    "msgname" : "recognized",
    "uniqueId" : "4d97031a-cfa9-4d66-968a-be708644a893"
  },
  "result" : {
```

```
      "type" : 2,
      "sentence" : [ ]
    }
  }, {
    "msg" : {
      "msgname" : "completed",
      "uniqueId" : "4d97031a-cfa9-4d66-968a-be708644a893",
      "cause" : "STOP"
    }
  } ]
```

Error Code

Troubleshooting for Request sent

If the following error occurs, please try your request again after taking the rectifying using the described solutions.
If the error persists, please contact us using [Contact Us](#).

code	message	detail	Description and Solution
410	Invalid Parameter	msg.msgname is invalid -> XXX.	Check msgname.
410	Invalid Parameter	param.baseParam.samplingRate is invalid -> XXX and over.	Check that the sampling rate is higher than the model bandwidth.
410	Invalid Parameter	recognizeParameter.enableProgress must be false for file speech recognition.	Sequential recognition result function cannot be used with `Speech Recognition from File`.

code	message	detail	Description and Solution
410	Invalid Parameter	param.baseParam.samplingRate is invalid -> The sampling rate must be 8000 Hz.	Check that the sampling rate is 8000 Hz.
411	Invalid State	Invalid Unique ID	Check the order in which the speech recognition API is called.
413	Invalid Data	Data is not available	If the audio data length exceeds the maximum audio length, perform the recognition multiple times. Also make sure that the sampling rate, audio codec, and model specifications are appropriate.
550	No Resource	VRG_RESPONSE_SERVICE_UNAVAILABLE	Check that ASR Domain id is correct. If the ASR Domain ID is correct, API access may be congested. Please wait a while and try again.
651	Session Timeout		Check the order and interval of API calls.
652	Excess Of Max Voice Length	Excess Of Max Voice Length -> 60 [s]	If the audio data length exceeds the maximum audio length, perform the recognition multiple times. Also make sure that the sampling rate, audio codec, and model specifications are appropriate.

Troubleshooting for `temporary dictionary`

If the following error occurs, please try your request again after taking the actions indicated. If the error persists, please contact us using Contact Us.

code	message	detail	Solution
410	Invalid Parameter	Invalid Parameter -> List is null	The notation and the reading is missing
410	Invalid Parameter	Cascade words exceeded 1000	Reduce the additional words in the temporary dictionary to 1000 lines or less.
410	Invalid Parameter	Surface empty	Write the notation.
410	Invalid Parameter	Reading empty	Write the reading.
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.
410	Invalid Parameter	Prob illegal	Specify the weight as an integer between 1 and 100, inclusive.
410	Invalid Parameter	Surface illegal	Check that the character string can be set as a notation.
410	Invalid Parameter	Reading illegal	Check that the character string can be set as a reading.
410	Invalid Parameter	Surface too long	Please ensure that the total of size of the notation and reading in in 251 bytes or less.

Please contact us

If the following error occurs, please contact us using Contact Us.

Error Code	Message
500	Internal Error
510	Out Of Memory
551	Recognition Timeout
552	Network Error
553	Network Timeout
601	Recognition Converter Error
610	Out Of Memory
611	Invalid License
612	Invalid Config
650	No Resource
690	External Command Execute Failed
691	External Command Fatal
692	External Command Error
693	External Command Warn

Troubleshooting for Request sent

If there is an invalid parameter in the `Request to Start Speech Recognition`, the following response may be returned with HTTP status code 500.

```
{
    "fault": {
        "faultstring": "NullPointerException",
        "detail": {
            "errorcode": "Internal Server Error"
        }
    }
}
```

Speech Recognition from Stream

An API for transcribing streaming audio, such as long audio files or input from a microphone. A length of audio can be up to `3,000` seconds. If your audio exceeds `3,000` seconds, divide it into parts less than `3,000` seconds and use this method for each separate audio.

Request Type

Speech Recognition for Streaming uses four possible types of requests.

- Request to Start Speech Recognition

- Send Audio Data
- Request to Stop Speech Recognition
- Request to Cancel Speech Recognition

The standard sequence for a `Streaming` request is as follows.

1. Request to Start Speech Recognition
2. Send Audio Data（one or more than one）
3. Request to Stop Speech Recognition

Request to Start Speech Recognition

This is a request to start speech recognition from the client to the API server.
When using speech recognition, you must first send this request.
This request requires parameter settings as follows.

HTTP Request

```
POST <API Base URL>/asr/v1/speech_recognition/<ASR Model id>
```

`ASR Model id` is the ID that identifies the model used for speech recognition.
Refer to List of Models to select the appropriate model.

Request Header

Key	Value
Connection	Keep-Alive
Content-Type	application/json; charset=UTF-8
Authorization	Bearer [Access Token]

Request Body

Key	Required	Description	Valid Range
msg			
msgname	Required	message type	start
param			
baseParam.samplingRate	Required	sample rate	more than Model's Rate (8000 or 16000)※1、8000 for mulaw
recognizeParameter.domainId	Required	[ASR Domain ID]	8 alphanumeric characters
recognizeParameter.enableContinuous	Optional※2	enable continuous recognition function	true（default: true）
recognizeParameter.enableProgress	Optional	sequential recognition result function※3	true, false（default: false）
recognizeParameter.maxResults	Optional	getting N-Best	1~30（default: 1）
baseParam.audioCodec	Optional	audio codec	pcm, mulaw（default: pcm）
baseParam.filler	Optional	getting filler	true, false（default: false）
baseParam.punctuation	Optional	getting punctuation	true, false（default: false）
baseParam.reading	Optional	getting reading	true, false（default: false）
baseParam.delimiter	Optional	getting delimiter（space）	true, false（default: true）
words	Optional	temporary dictionary	1000 words or less

※1 Please note that we cannot guarantee correct operation for audio other than 8000, 11025, 16000, 22050, 32000, 44100, 48000, 88200, 96000[Hz]

※2 This is required until 2020/3/23, but it is now an optional parameter.

※3 Please note that this cannot be used with `Speech Recognition from File`.

Audio Codec

Linear PCM (pcm) and μ-Law (mulaw) can be specified.
However, μ-Law speech recognition is available only for 8kHz models(`ja-gen_tf-08`, `ja-mdl1_nrw-08`, `ja-mdl2_nrw-08`).
If the audio format is wav, please remove the header properly and convert to Linear PCM or μ-Law format.

Getting Filler

By setting this parameter to `true`, it is possible to output the filler results(utterances such as `あー` and `えー`) in the response, where they are wrapped with `[]`.

※ This is valid only for the models (`ja-gen_tf-16`, `ja-gen_sf-16`, `ja-gen_tf-08`, `ja-mdl1_nrw-08`, `ja-mdl2_nrw-08`).

`baseParam.filler=false`(default)

```
これ は テスト 用 の 音声ファイル です
```

`baseParam.filler=true`

```
[えーと] これ は テスト 用 の 音声ファイル です
```

Getting Punctuation

By setting this parameter to `true`, speech recognition result is output with punctuation.

※ This is valid only for the models (`ja-gen_tf-16`, `ja-gen_tf-08`, `ja-mdl1_nrw-08`, `ja-mdl2_nrw-08`).

※ There is no guarantee that a sentence will end with a `。`.

`baseParam.punctuation=false`(default)

```
こんにちは よろしくお願いします
```

`baseParam.punctuation=true`

```
こんにちは 。 よろしくお願いします
```

Getting Reading

※ By setting this parameter to `true`, you can also get the reading in the response.

※ This is valid only for (`ja-gen_tf-16`, `ja-gen_sf-16`, `ja-gen_tf-08`, `ja-mdl1_nrw-08`, `ja-mdl2_nrw-08`).

```
'reading': 'コレ ワ テスト ヨウ ノ オンセイファイル デス'
```

Please refer to [Response Sample](#) for output responses that include `reading`.

Getting Delimiter

By setting this parameter to `false`, it is possible to remove space normally inserted between words.

※　When using the English model (`en_en-gen_sf-16`), the setting `true` is recommended.

`baseParam.delimiter=true` (default)

> これ は テスト 用 の 音声ファイル です

`baseParam.delimiter=false`

> これはテスト用の音声ファイルです

Sequential Recognition Result Function

By setting this parameter to `true`, it is possible to sequentially obtain the results during a recognition that have not ended.
Recognition results are returned as recognition result type `type = 0`.
Note that `reading` will not be output and `filler, punctuation, delimiter` will be in the output for the results during recognition, regardless of the parameters.

Getting N-Best

Returns multiple speech recognition results up to the number specified in `recognizeParameter.maxResults`.
However, do note that the specified number is the upper limit, and it is not guaranteed that the specified number of results will always be returned.
For the output format when `recognizeParameter.maxResults` is set to 2 or more, refer to Response sample.

Temporary Dictionary

`Temporary Dictionary` that can be set for each speech recognition sequence.

It can be used in conjunction with Update Speech Recognition Dictionary.
`Temporary Dictionary` is discarded when the request sequence ends, and the user dictionary will not be updated.

Key	Description	Range	Required
surface	surface※1	Character string that does not contain `|,　/,` `:,　;,　[,　],　\t,　\"` and is not empty	Required
reading	reading※1	Full-width katakana（Annexed Table）	Required
prob	weight of word※2	Integer value from 1 to 100（default: 30）	Optional

※1 Please ensure that the total of size of the notation and reading in in 251 bytes or less.

※2 Use this when you want to add weight to registered words.

By setting a value greater than 30, you will be setting a weight greater than the default user dictionary.
By setting a value smaller than 30, you will be setting a weight lower than the default user dictionary.

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/json; charset=UTF-8
Authorization: Bearer <Access Token>
```

HTTP Body

```
{
  "msg": {
        "msgname": "start"
    },
    "param": {
        "baseParam.samplingRate": 16000,
        "recognizeParameter.domainId": "<ASR Domain ID>",
        "baseParam.audioCodec": "pcm",
        "baseParam.filler": true,
        "baseParam.reading": true,
        "baseParam.delimiter": false,
        "baseParam.punctuation": true,
        "recognizeParameter.enableProgress": true,
        "recognizeParameter.maxResults": 2
    },
    "words": [
        {
            "surface": "エヌ・ティ・ティ・コミュニケーションズ株式会社",
            "reading": "エヌティティコミュニケーションズカブシキガイシャ",
            "prob": "20"
        },
        {
            "surface": "COTOHA",
            "reading": "コトハ",
            "prob": "40"
        }
    ]
}
```

Send Audio Data

This is a request to send speech from the client to the API server.
Each request interval for submitting data is `240` milliseconds※.
Please ensure that the response contains no errors and before submitting the next audio request.
The first `Send Audio Data` must be performed within 1 second of receiving the response to `Request to Start Speech Recognition`.

※If the next request cannot be sent within the specified time due to network delay, please send it as soon as possible.

HTTP Request

```
POST <API Base URL>/asr/v1/speech_recognition/<ASR Model ID>
```

Request Header

Key	Description
Connection	Keep-Alive
Content-Type	application/octet-stream
Unique-Id	Unique-Id returned in the response header of Start Speech Recognition
Authorization	Bearer [Access Token]
Cookie	- token（A unique string used to identify the order of requests. Set as the token received in last response.） - GCLB（A unique set of strings for each speech recognition sequence. Set as the GCLB received in Speech Recognition Ready Response.）

Request Body

Audio Data Part

Convert the audio data to one of the following formats and use the converted binary for this part.
※If the audio format is wav, please remove the header property and convert to Linear PCM or μ-Law format.

Format	Sample Rate[Hz]	Quantization[bit]	Channel	Byte Order	Audio length
Linear PCM	more than Model's rate (8000 or 16000)※1	16	1	Little Endian	240※2
μ-Law	8000	8	1	Little Endian	240※2

※1 Please note that we cannot guarantee correct operation for audio other than 8000, 11025, 16000, 22050, 32000, 44100, 48000, 88200, 96000[Hz]

※2 The time interval when sending the last speech request can be less than `240` milliseconds.

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/octet-stream
Authorization: Bearer <Access Token>
Cookie: GCLB=<GCLB that received in Speech Recognition Ready Response>;token=
<token that received in last response>
```

HTTP Body

```
<Binary Audio Data>
```

Request to Stop Speech Recognition

This is the request to stop speech recognition from the client to the API server.
The API server returns `200(OK)` after all speech recognition is complete.
`Request to Stop Speech Recognition` must be performed within 1 second of receiving the
response to the last `Send Audio Data` .

HTTP Request

```
POST <API Base URL>/asr/v1/speech_recognition/<ASR Model ID>
```

Request Header

Key	Description
Connection	Keep-Alive
Content-Type	application/octet-stream
Unique-Id	Unique-Id returned in the response header of Start Speech Recognition
Authorization	Bearer [Access Token]
Cookie	- token（A unique string used to identify the order of requests. Set as the token received in last response.） - GCLB（A unique set of strings for each speech recognition sequence. Set as the GCLB received in Speech Recognition Ready Response.）

Request Body

Key		Descriptioin	Valid Range
msg			
	msgname	message type	stop

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/octet-stream
Authorization: Bearer <Access Token>
Cookie: GCLB=<GCLB that received in Speech Recognition Ready Response>;token=
<token that received in last response>
```

HTTP Body

```
{
    "msg": {
        "msgname": "stop"
    }
}
```

Request to Cancel Speech Recognition

This is the request to cancel speech recognition to the server.
The server will cancel speech recognition and return `200(OK)`.
`Request to Cancel Speech Recognition` must be performed within 1 second of receiving the response to the last `Send Audio Data`.

HTTP Request

```
POST <API Base URL>/asr/v1/speech_recognition/<ASR Model id>
```

Request Header

Key	Description
Connection	Keep-Alive
Content-Type	application/octet-stream
Unique-Id	Unique-Id returned in the response header of Start Speech Recognition
Authorization	Bearer [Access Token]
Cookie	- token（A unique string used to identify the order of requests. Set as the token received in last response.） - GCLB（A unique set of strings for each speech recognition sequence. Set as the GCLB received in Speech Recognition Ready Response.）

Request Body

Key	Descriptioin	Valid Range
msg		
	msgname	message type	cancel

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/octet-stream
Authorization: Bearer <Access Token>
Cookie: GCLB=<GCLB that received in Speech Recognition Ready Response>;token=
<token that received in last response>
```

HTTP Body

```
{
    "msg": {
        "msgname": "cancel"
    }
}
```

Response

The HTTP response consists of a combination of the following server responses.
If there is more than one server response, an HTTP response containing JSON is returned with HTTP status code `200(OK)`.
If there are **0** server responses, an HTTP status code of `204(No Content)` is returned.

Server Response Type	msgname	Description
Speech Recognition Ready Response	started	response to Speech Recognition Start Request
Detect Speech Start Response	speechStartDetected	response to request after detection of speech start
Detect Speech End Response	speechEndDetected	response to request after detection of speech end
Speech Recognition Result Response	recognized	speech recognition results
Speech Recognition End Response	completed	message when speech recognition ends or error occurs during speech recognition processing

All server responses include `msgname` and `uniqueId`.

`msgname` represents the type of server response, and `uniqueId` is a unique identifier for each speech recognition.

Speech Recognition Ready Response

In response to Speech Recognition Start Request, this message is sent from the API server to the client to notify that the server is ready to receive audio data.

Response Header

Key Name	Description
Content-Type	application/json; charset=UTF-8
Set-Cookie	- token（A unique string used to identify the order of requests.） - GCLB（A unique set of strings for each speech recognition sequence.）

Response Body

Key		Description	Remarks
msg			
	msgname	message type	started
	uniqueId	unique identifier for each speech recognition	uniqueId is required for "Send Audio Data", "Request to Stop Speech Recognition" and "Request to Cancel Speech Recognition

Response Sample

```
[
    {
        "msg": {
            "msgname": "started",
            "uniqueId": "3bfbe5de-eee7-4824-a661-3750d8cb9328"
        }
    }
]
```

Detect Speech Start Response

This message is returned when the start of an utterance is detected from sent audio.

Response Header

Key	Description
Content-Type	application/json; charset=UTF-8
Set-Cookie	token（A unique string used to identify the order of requests.）

Response Body

Key	Description	Remarks
msg		
msgname	message type	speechStartDetected
uniqueId	unique identifier for each speech recognition	uniqueId is required for "Send Audio Data", "Request to Stop Speech Recognition" and "Request to Cancel Speech Recognition"
timeinfo		
startDetectTime	start detect time[ms]	time from beginning of audio to speech start

Response Sample

```
[ {
  "msg" : {
    "msgname" : "speechStartDetected",
    "uniqueId" : "3bfbe5de-eee7-4824-a661-3750d8cb9328"
  },
  "timeinfo" : {
    "startDetectTime" : 0
  }
} ]
```

Detect Speech End Response

This message is returned when the end of an utterance is detected from sent audio.

Response Header

Key	Description
Content-Type	application/json; charset=UTF-8
Set-Cookie	token（A unique string used to identify the order of requests.）

Response Body

Key		Description	Remarks
msg			
	msgname	message type	speechStartDetected
	uniqueId	unique identifier for each speech recognition	uniqueId is required for "Send Audio Data", "Request to Stop Speech Recognition" and "Request to Cancel Speech Recognition"
timeinfo			
	startDetectTime	start detect time[ms]	time from beginning of audio to speech start

Response Sample

```
[ {
  "msg" : {
    "msgname" : "speechEndDetected",
    "uniqueId" : "4b96875f-2137-48ed-8b49-1e20483a7c86"
  },
  "timeinfo" : {
    "endDetectTime" : 3200
  }
} ]
```

Speech Recognition Result Response

This message is the speech recognition result.

Response Header

Key Name	Description
Content-Type	application/json; charset=UTF-8
Set-Cookie	token（A unique string used to identify the order of requests.）

Resonse Body

Key			Description	Remarks
msg				
	msgname		message type	recognized
	uniqueId		unique identifier for each speech recognition	uniqueId is required for "Send Audio Data", "Request to Stop Speech Recognition" and "Request to Cancel Speech Recognition"
result				
	type		Types of Recognition Result	• Sequential Recognition Result • Detection Speech End • Request to Stop Speech Recognition
	sentence			
		surface	result text	
		score	score	confidence of result (0-1 scale)
		startTime	start time[s]	time from beginning of audio to speech start
		endTime	end time[s]	time from beginning of audio to speech end
		reading	reading of results	`baseParam.reading=true` only

Response Sample

Paramters Paramters Specified at `Request to Start Speech Recognition`

```
baseParam.filler: false(default)
baseParam.reading: false(default)
baseParam.delimiter: true(default)
baseParam.punctuation: false(default)
recognizeParameter.maxResults: 1(default)
```

Response

```
{
    "msg": {
        "msgname": "recognized",
        "uniqueId": "49bbc297-6a8a-4354-9679-2866cc27385a"
    },
    "result": {
        "type": 1,
        "sentence": [
            {
                "surface": "これ は テスト 用 の 音声ファイル です",
                "score": 0.851436,
                "startTime": 0,
                "endTime": 2.99
            }
        ]
```

```
        }
    }
```

Response Sample

`filler=true, reading=true, delimiter=false, punctuation=true, maxResults=2`

Paramters Paramters Specified at `Request to Start Speech Recognition`

```
baseParam.filler: true
baseParam.reading: true
baseParam.delimiter: false
baseParam.punctuation: true
recognizeParameter.maxResults: 2
```

Response

```
{
    "msg": {
        "msgname": "recognized",
        "uniqueId": "a6342b20-6b0d-41b4-9c71-86af4f123369"
    },
    "result": {
        "type": 1,
        "sentence": [
            {
                "surface": "[えーと]、これはテスト用の音声ファイルです",
                "score": 0.851436,
                "startTime": 0,
                "endTime": 2.99,
                "reading": "エット、コレワテストヨウノオンセイファイルデス"
            },
            {
                "surface": "[えーと]。これはテスト用の音声ファイルです",
                "score": 0.532213,
                "startTime": 0,
                "endTime": 2.99,
                "reading": "エット。コレワテストヨウノオンセイファイルデス"
            }
        ]
    }
}
```

Speech Recognition End Response

This message is returned when speech recognition is completed.
This message is also returned if an error occurs during the speech recognition process.

Response Header

Key	Description
Content-Type	application/json; charset=UTF-8
Set-Cookie	token（A unique string used to identify the order of requests.）

Response Body

Key	Description	Remarks
msg		
msgname	message type	completed
uniqueId	unique identifier for each speech recognition	uniqueId is required for "Send Audio Data", "Request to Stop Speech Recognition" and "Request to Cancel Speech Recognition"
cause	reason for stop	one of the following • STOP • CANCEL • ERROR • END DETECTION※
errorinfo		
code	error code	
message	error message	
level	error level	one of the following • WARN • ERROR • FATAL • END DETECTION※
detail	more information about the error	

※This is a stop factor that does not normally occur. When it occurs, please confirm that `recognizeParameter.enableContinuous` is set properly at the time of `Request to Start Speech Recognition`.

Response Sample

```
[ {
  "msg" : {
    "msgname" : "completed",
    "uniqueId" : "4b96875f-2137-48ed-8b49-1e20483a7c86",
    "cause" : "STOP"
  }
} ]
```

Error Code

Troubleshooting for Request sent

If the following error occurs, please try your request again after taking the actions indicated. Note that in the case of errors in Speech Recognition from Stream API, restart from `Request to Start Speech Recognition` . You don't need to `Request to Stop Speech Recognition` or `Request to Cancel Speech Recognition` . If the error persists, please contact us from [Contact Us](#).

code	message	detail	Description and Solution
410	Invalid Parameter	msg.msgname is invalid -> XXX.	Check msgname.

code	message	detail	Description and Solution
410	Invalid Parameter	param.baseParam.samplingRate is invalid -> XXX and over.	Check that the sampling rate is higher than the model bandwidth.
410	Invalid Parameter	param.baseParam.samplingRate is invalid -> The sampling rate must be 8000 Hz.	Check that the sampling rate is 8000 Hz.
411	Invalid State	Invalid Unique ID	Check the order in which the speech recognition API is called.
412	Interval Too Brief		Audio transmission interval is too short. Set the transmission interval of audio correctly.
450	Invalid Token	token Error	Cookie `token` (not the Access Token) is invalid. Please make a request after receiving a response.
550	No Resource	VRG_RESPONSE_SERVICE_UNAVAILABLE	Check that ASR Domain id is correct. If the ASR Domain ID is correct, API access may be congested. Please wait a while and try again.
552	Network Error	Network	Check that the request has been made within the specified time after receiving the response.
600	Internal Error		Check that the request has been made within the specified time after receiving the response.
651	Session Timeout		Check the API call order and call interval.

Troubleshooting for `temporary dictionary`

If the following error occurs, please try your request again after taking the actions indicated.
Note that in the case of errors in Speech Recognition from Stream API, restart from `Request to Start Speech Recognition` .
You don't need to `Request to Stop Speech Recognition` or `Request to Cancel Speech Recognition` .
If the error persists, please contact us using Contact Us.

code	message	detail	Solution
410	Invalid Parameter	Invalid Parameter -> List is null	The notation and the reading is missing
410	Invalid Parameter	Cascade words exceeded 1000	Reduce the additional words in the temporary dictionary to 1000 lines or less.
410	Invalid Parameter	Surface empty	Write the notation.
410	Invalid Parameter	Reading empty	Write the reading.
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.
410	Invalid Parameter	Prob illegal	Specify the weight as an integer between 1 and 100, inclusive.
410	Invalid Parameter	Surface illegal	Check that the character string can be set as a notation.
410	Invalid Parameter	Reading illegal	Check that the character string can be set as a reading.
410	Invalid Parameter	Surface too long	Please ensure that the total of size of the notation and reading in in 251 bytes or less.

Please contact us

If the following error occurs, please contact us using Contact Us.

Error Code	Message
500	Internal Error
510	Out Of Memory
551	Recognition Timeout
553	Network Timeout
601	Recognition Converter Error
610	Out Of Memory
611	Invalid License
612	Invalid Config
650	No Resource
690	External Command Execute Failed
691	External Command Fatal
692	External Command Error
693	External Command Warn

Update Speech Recognition Dictionary

Japanese model only

For adding words to the default dictionary.
A dictionary must be registered for each model.

Note that previously added dictionary will be overwritten by this call.

HTTP Request

```
POST <API Base URL>/asr/v1/speech_words/<ASR Model ID>/upload?domainid=<ASR Domain ID>
```

Request Header

You need to send a multipart form.
Define a boundary delimiter and indicate it as `boundary` in your request header.

Key	Description
Content-Type	multipart/form-data; boundary=<Boundary Delimiter>
Authorization	Bearer [Access Token]

Request Body

Include `Speech Recognition Dictionary` (Below) in the request body.

Part Type	Required	Name	Content-Type
Dictionary	Required	cascadeword	text/plain; charset=UTF-8

Speech Recognition Dictionary

Write Notation, Horizontal Tab and Reading on each line.

```
<HYOKI><HT><YOMI>
<HYOKI><HT><YOMI>
...
<HYOKI><HT><YOMI>
```

When specifying the weight of word※1, describe it as follows. Weight can be specified for each word.

```
<HYOKI><HT><YOMI><HT><PROB>
<HYOKI><HT><YOMI>
...
<HYOKI><HT><YOMI><HT><PROB>
```

Key	Description	Range	Required
HYOKI	Notation※2	Character string that does not contain `|`, `/`, `:`, `;`, `[`, `]`, `\t`, `\"` and is not empty	Required
YOMI	Reading※2	Full-width katakana（Annexed Table）	Required
PROB	weight of word※1	Integer value from 1 to 100（default: 30）	Optional

※1 Please ensure that the total of size of the notation and reading in in 251 bytes or less.

※2 Use this when you want to add weight to registered words.

By setting a value greater than 30, you will be setting a weight greater than the default user dictionary.
By setting a value smaller than 30, you will be setting a weight lower than the default user dictionary.

Speech Recognition Dictionary Sample

```
エヌ・ティ・ティ・コミュニケーションズ株式会社  エヌティティコミュニケーションズカブシキガイシ
ャ
COTOHA  コトハ  20
```

Sample Request

HTTP Header

```
Content-Type: multipart/form-data; boundary=<Boundary Delimiter>
Authorization: Bearer <Access Token>
```

HTTP Body

```
--<Boundary Delimiter>
Content-Disposition: form-data; name="cascadeword"; filename="sample.tsv"
Content-Type: text/plain

エヌ・ティ・ティ・コミュニケーションズ株式会社  エヌティティコミュニケーションズカブシキガイシ
ャ
COTOHA　コトハ　20


--<Boundary Delimiter>--
```

cURL

```
curl -H "Authorization:Bearer  <Access Token>" -X POST -F
cascadeword=@sampple.tsv <API Base URL>/asr/v1/speech_words/<ASR Model
ID>/upload?domainid=<ASR Domain ID>
```

Response

The specified `notation`, the specified `reading`, `M` (fixed), and `weight of word` are returned.
`M` is automatically assigned by the server, so you usually do not need to be concerned about it. If
`weight of word` is not specified, the default value of 30 will be returned.

Response Header

Key	Description
Content-Type	multipart/form-data; boundary=

Response Sample

```
--<Boundary Delimiter>
Content-Type: text/plain
Content-Disposition: form-data; name="status"

code : 200
message : OK
detail : success

--<Boundary Delimiter>
Content-Type: text/plain
Content-Disposition: form-data; name="cascadeword"
```

```
エヌ・ティ・ティコミュニケーションズ株式会社　エヌティティコミュニケーションズカブシキガイシャ
M　30
COTOHA　コトハ M　20


--<Boundary Delimiter>--
```

Error Code

Please review your request

If the following error occurs, please try your request again after taking the actions indicated. If the error persists, please contact us from [Contact Us](#).

code	message	detail	Solution
410	Invalid Parameter	Invalid Parameter -> List is null	Write the notation and the reading.
410	Invalid Parameter	Invalid Parameter -> List Exceed 5000 lines	In the speech recognition dictionary, specify additional words in 1000 lines or less.
410	Invalid Parameter	Surface empty	Write the notation.

code	message	detail	Solution
410	Invalid Parameter	Reading empty	Write the reading.
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.
410	Invalid Parameter	Prob illegal	Specify the weight as an integer between 1 and 100, inclusive.
410	Invalid Parameter	Surface illegal	Check that the character string can be set as the notation.
410	Invalid Parameter	Reading illegal	Check that the character string can be set as the reading.
410	Invalid Parameter	Surface too long	Please specify the total of notation and reading in 251 bytes or less.

Please contact us

If the following error occurs, please contact us using [Contact Us](#).

code	message	detail
410	Invalid Parameter	Upload Error
410	Invalid Parameter	Download Error
600	Internal Error	-

Delete Speech Recognition Dictionary

Japanese models only

For deleting registered dictionary.

The registered dictionaries should be deleted individually for each model.

HTTP Request

```
GET <API Base URL>/asr/v1/speech_words/<ASR Model ID>/clear?domainid=<ASR Domain
ID>
```

Request Header

Key Name	Description
Authorization	Bearer [Access Token]

Sample Request

HTTP Header

```
Authorization: Bearer <Access Token>
```

cURL

```
curl -H "Authorization:Bearer  <Access Token>" <API Base
URL>/asr/v1/speech_words/<ASR Model ID>/clear?domainid=<ASR Domain ID>
```

Response

Response Header

Key	Description
Content-Type	multipart/form-data; boundary=<Boundary Delimiter>

Response Sample

```
--<Boundary Delimiter>
Content-Type: text/plain
Content-Disposition: form-data; name="status"

code : 200
message : OK
detail : success


--<Boundary Delimiter>--
```

Error Code

Please review your request

If the following error occurs, please try your request again after taking the actions indicated. If the error persists, please contact us from Contact Us.

code	message	detail	Solution
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.

code	message	detail
410	Invalid Parameter	Upload Error
410	Invalid Parameter	Download Error
600	Internal Error	-

Please contact us

If the following error occurs, please contact us from [Contact Us](#).

code	message	detail
410	Invalid Parameter	Upload Error
410	Invalid Parameter	Download Error
600	Internal Error	-

Get Speech Recognition Dictionary State

Japanese models only

API for checking if the dictionary data is reflected. For each model, it is possible to obtain whether the user dictionary has been updated.

HTTP Request

```
GET <API Base URL>/asr/v1/speech_words/<ASR Model ID>/isset?domainid=<ASR Domain ID>
```

Sample Request

HTTP Header

```
Authorization: Bearer <Access Token>
```

cURL

```
curl -H "Authorization:Bearer  <Access Token>" <API Base URL>/asr/v1/speech_words/<ASR Model ID>/isset?domainid=<ASR Domain ID>
```

Response

If the user dictionary has been updated, true will be returned; if not completed, false will be returned.
Note that the response `Content-Type` differs depending on whether an error has occurred.

Response Header

Key	Description
Content-Type	application/json

Response Body

Key	Range
isSet	true, false

Response Sample

```
{
  "isSet" : true
}
```

Error Response

Response Header

Key	Description
Content-Type	multipart/form-data; boundary=<Boundary Delimiter>

Response Sample

```
--<Boundary Delimiter>
Content-Type: text/plain
Content-Disposition: form-data; name="status"

code : 410
message : Invalid Parameter
detail : Invalid domainid


--<Boundary Delimiter>--
```

Error Code

Please review your request

If the following error occurs, please try your request again after taking the actions indicated. If the error persists, please contact us using Contact Us.

code	message	detail	Solution
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.

Please contact us

If the following error occurs, please contact us using Contact Us.

code	message	detail
410	Invalid Parameter	Upload Error
410	Invalid Parameter	Download Error
600	Internal Error	-

Download speech Recognition Dictionary

Japanese models only

API for downloading the last uploaded dictionary. It is possible to download the last uploaded dictionary for each model.

HTTP Request

```
GET <API Base URL>/asr/v1/speech_words/<ASR Model ID>/download?domainid=<ASR Domain ID>
```

Request Header

Key	Description
Authorization	Bearer [Access Token]

Sample Request

HTTP Header

```
Authorization: Bearer <Access Token>
```

cURL

```
curl -H "Authorization:Bearer  <Access Token>" <API Base URL>/asr/v1/speech_words/<ASR Model ID>/download?domainid=<ASR Domain ID>
```

Response

The words in the last uploaded dictionary will be returned. Each line contains `notation`, `reading`, and `weight of word`.

Response Header

Key	Description

Key	Description
Content-Type	multipart/form-data; boundary=<Boundary Delimiter>

Response Sample

```
--<Boundary Delimiter>
Content-Type: text/plain
Content-Disposition: form-data; name="status"

code : 200
message : OK
detail : success

--<Boundary Delimiter>
Content-Type: text/plain
Content-Disposition: form-data; name="cascadeword"

エヌ・ティ・ティコミュニケーションズ株式会社　エヌティティコミュニケーションズカブシキガイシャ
30
COTOHA　コトハ 30

--<Boundary Delimiter>--
```

Error Code

Please review your request

If the following error occurs, please try your request again after taking the actions indicated.
If the error persists, please contact us using Contact Us.

code	message	detail	Solution
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.

Please contact us

If the following error occurs, please contact us using Contact Us.

code	message	detail
410	Invalid Parameter	Upload Error
410	Invalid Parameter	Download Error
600	Internal Error	-

List of Models

Model Name	ASR Model id
Japanese General Talk&Free(16kHz)	ja-gen_tf-16
Japanese General Short&Formal(16kHz)	ja-gen_sf-16
Japanese General Talk&Free(8kHz)	ja-gen_tf-08
Japanese Telecommunications(8kHz)	ja-mdl1_nrw-08
Japanese Insurance(8kHz)	ja-mdl2_nrw-08
English General Native Short&Formal(16kHz)	en_en-gen_sf-16

- Difference between　Talk&Free　and Short&Formal
  - Talk&Free: This model is suitable for speech recognition that is freestyle, where you do not have an exact idea
    of what to say beforehand, where natural speech elements like hesitation and mispronunciations often occur.
    （e.g. meeting, chat, call center and customer service, etc.）
  - Short&Formal: This model is suitable for speech recognition of audio with a single utterance in which you
    have an idea of what to say and is able to speak relatively clearly.（
    e.g. search queries, interactions with question and answer systems, etc.）
- Difference between 8kHz and 16kHz

  - 8kHz: Recommended for speech over a telephone line.
  - 16kHz: Recommended for other speech.