COTOHA Speech Recognition

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

- Speech Recognition from File
  API for transcribing short audio files.
- Speech Recognition from Stream
  API for transcribing streaming audio, such as long audio files or input from a microphone.
- Update Speech Recognition Dictionary (Japanese models only)
  API for adding words to the default dictionary.
  The dictionary data is reflected every hour on the hour. It takes a certain amount of time for the data to actually be reflected.
- Delete Speech Recognition Dictionary(Japanese models only)
  API for deleting registered dictionary.
  The dictionary data is reflected every hour on the hour. It takes a certain amount of time for the data to actually be reflected.

Speech Recognition from File

An 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 for each `Part Type`.
`Part Type` in the request 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 the following binary format and use the converted binary for this part.

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

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"
  }
} ]
```

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 flow for 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 speech recognition parameter settings as follows.

HTTP Request

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

Request Header

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

Request Body

Key	Nested Key	Description	Valid Range
msg	msgname	message type	start
param	baseParam.samplingRate	sample rate	more than Model's Rate (8000 or 16000)
	recognizeParameter.domainId	[ASR Domain ID]	8 alphanumeric characters
	recognizeParameter.enableContinuous	enable continuous recognition	true

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>",
        "recognizeParameter.enableContinuous": true
    }
}
```

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` .

HTTP Request

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

Request Header

Add the cookie returned in the previous response header to the request header.

Key Name	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.）

Refer to Speech Recognition Ready Response for details about the cookie.

Request Body

Convert the audio data into the following binary format and use the converted binary for this part.

Format	Sample Rate[Hz]	Quantization[bit]	Channel	Byte Order	Audio length

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

The size of one request body is `3,840` bytes for `8kHz` models and `7,680` bytes for `16kHz` models for an audio length of `240ms`.
※However, the last request to send speech can be less than `240` milliseconds.

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/octet-stream
Authorization: Bearer <Access Token>
Cookie: <the Cookie returned in the 'Request To Start Speech Recognition'
response header>
```

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 Name	Description
Connection	Keep-Alive
Content-Type	application/json; charset=UTF-8
Unique-Id	Unique-Id returned in the response header of Start Speech Recognition

Key Name	Description
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	Nested Key	Description	Valid Range
msg	msgname	message type	stop

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/json; charset=UTF-8
Authorization: Bearer <Access Token>
Cookie: <the Cookie returned in the previous response header>
```

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 Name	Description
Connection	Keep-Alive
Content-Type	application/json; charset=UTF-8
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	Nested Key	Description	Valid Range
msg	msgname	message type	cancel

Sample Request

HTTP Header

```
Connection: Keep-Alive
Content-Type: application/json; charset=UTF-8
Authorization: Bearer <Access Token>
Cookie: <the Cookie returned in the previous response header>
```

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

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

Response Body

Key	Nested 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	Nested Key	Description	Remarks
msg	msgname	message type	speechEndDetected
	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	endDetectTime	end detect time[ms]	time from beginning of audio to speech end

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.）

Response Body

Key	Nested Key 1	Nested Key 2	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	1. Detection Speech End 2. Request to Stop Speech Recognition
	sentence	surface	result text	text with spaces between words
		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

Response Sample

```
[
    {
      "msg" : {
        "msgname" : "recognized",
        "uniqueId" : "a49b39de-101f-4a58-b7a6-4b3ffbfb58bb"
      },
      "result" : {
        "type" : 1,
        "sentence" : [ {
          "surface" : "これ は テスト 用 の 音声 ファイル です",
          "score" : 0.830472,
          "startTime" : 0.0,
          "endTime" : 3.2
        } ]
     }
    }
]
```

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 Name	Description
Content-Type	application/json; charset=UTF-8

Key Name	Description
Set-Cookie	token（A unique string used to identify the order of requests. ）

Response Body

Key	Nested 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
errorinfo	code	error code	
	message	error message	
	level	error level	one of the following -WARN -ERROR -FATAL
	detail	more information about the error	

Response Body

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

Update Speech Recognition Dictionary

Japanese model only

For adding words to the default dictionary.

A dictionary must be registered for each model.

Note that formerly 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 Name	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>
```

項目	Description	Range	Required
[HYOKI]	Notation	non-empty character	Required
[YOMI]	Reading	Full-width Katakana	Required

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
```

Sample Request(cURL)

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

Response

Returns notations, readings, "M" and scores.
"M" and scores are responses from the server and are not of any importance for now.

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       -3.0
COTOHA  コトハ   M       -3.0

--<Boundary Delimiter>--
```

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(cURL)

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

Response

Response Sample

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

code : 200
message : OK
detail : success

--<Boundary Delimiter>--
```

List of Models

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

- Difference between Short&Formal and Talk&Free
  - 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.）

    - 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.)
- Difference between 8kHz and 16kHz
    - 8kHz: Recommended for speech over a telephone line.
    - 16kHz: Recommended for other speech.

List of Error Code

Speech Recognition

Please review your request

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 `Subscriber inquiries`.

Error Code	Message	Description and Solution
410	Invalid Parameter	Check the parameters.
411	Invalid State	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	Wait for the previous response before making a request.
551	Recognition Timeout	Check if the speech is correct.
552	Network Error	Check if you have made a request within the specified time after receiving the response.
600	Internal Error	Check if you have made a request within the specified time after receiving the response.
651	Session Timeout	Check the order and interval of API calls.
652	Excess Of Max Voice Length	Divide audio into less than 3,000 seconds.
690	External Command Execute Failed	Check the parameters.

Please review and try your request again later

If the following error occurs, please take the actions indicated and try your request again later. 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 `Subscriber inquiries`.

Error Code	Message	Description and Solution
550	No Resource	Check that ASR Model id is correct.

Please contact us

If the following error occurs, contact `Subscriber inquiries`.

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

Update Speech Recognition Dictionary

Please review your request

If the following error occurs, please try your request again after taking the actions indicated.

code	message	detail	Solution
410	Invalid Parameter	List is null	Write and check the notation and reading.
410	Invalid Parameter	List Exceed 5000 lines	Specify no more than 5,000 additional words.
410	Invalid Parameter	Required HYOKI	Write the notation.
410	Invalid Parameter	Required YOMI	Write the reading.
410	Invalid Parameter	Invalid domainid	Check that ASR Domain id is correct.

code	message	detail	Solution
410	Invalid Parameter	Invalid Model Name	Check that ASR Model id is correct.
410	Invalid Parameter	List validation failed. Unknown word weight value X	Set Weight to M.
410	Invalid Parameter	ユーザ辞書追加リストの単語名が長すぎる	Specify HYOKI in less than 251 bytes.
410	Invalid Parameter	ユーザ辞書リストの読みが長すぎる	Specify YOMI in less than 255 bytes.
410	Invalid Parameter	単語追加の設定失敗	Check that the notation and reading are specified correctly.

Please contact us

If the following error occurs, contact `Subscriber inquiries`.

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