Adding recognition metadata

Beta

This feature is in a pre-release state and might change or have limited support. For more information, see the product launch stages (https://cloud.google.com/products/#product-launch-stages).

This page describes how to add additional details about the source audio included in a speech recognition request to Speech-to-Text.

Speech-to-Text has several machine learning models to use for converting recorded audio into text. Each of these models has been trained based upon specific characteristics of audio input, including the type of audio file, the original recording device, the distance of the speaker from the recording device, the number of speakers on the audio file, and other factors.

When you send a transcription request to Cloud Speech-to-Text, you can include these additional details about the audio data as *recognition metadata* that you send. Cloud Speech-to-Text can use these details to more accurately transcribe your audio data.

Google also analyzes and aggregates the most common use cases for the Cloud Speech-to-Text by collecting this metadata. Google can then prioritize the most prominent use cases for improvements to Speech-to-Text.

Note: Google does not collect the actual data of the audio file that you send, only the metadata, unless you have opted-in for data logging (https://cloud.google.com/speech-to-text/docs/data-logging).

Available metadata fields

You can provide any of the fields in the following list in the metadata of a transcription request.

Field	Type

Field	Type
`interactionType`	[InteractionType](https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#InteractionType)
`industryNaicsCodeOfAudio`	number
`microphoneDistance`	[MicrophoneDistance](https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#MicrophoneDistanc
`originalMediaType`	[OriginalMediaType](https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#OriginalMediaType)
`recordingDeviceType`	[RecordingDeviceType](https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#RecordingDeviceTyp
`recordingDeviceName`	string
`originalMimeType`	string
`obfuscatedId`	string
`audioTopic`	string

See the `RecognitionMetadata` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#RecognitionMetadata) reference documentation for more information about these fields.

Use recognition metadata

To add recognition metadata to a speech recognition request to the Speech-to-Text API, set the `metadata` field of the speech recognition request to a `RecognitionMetadata` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#RecognitionMetadata) object. The Speech-to-Text API supports recognition metadata for all speech recognition methods: `speech:recognize` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/speech/recognize), `speech:longrunningrecognize` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/speech/longrunningrecognize), and `StreamingRecognizeRequest` (https://cloud.google.com/speech-to-text/docs/reference/rpc/google.cloud.speech.v1p1beta1#google.cloud.speech.v1p1beta1.StreamingRecognizeRequest). See the `RecognitionMetadata` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig#RecognitionMetadata) reference documentation for more information on the types of metadata that you can include with your request.

The following code demonstrate how to specify additional metadata fields in a transcription request.

PROTOCOL JAVA NODE.JS PYTHON

Refer to the `speech:recognize` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/speech/recognize) API endpoint for complete details.

To perform synchronous speech recognition, make a `POST` request and provide the appropriate request body. The following shows an example of a `POST` request using `curl`. The example uses the access token for a service account set up for the project using the Google Cloud Cloud SDK (https://cloud.google.com/sdk). For instructions on installing the Cloud SDK, setting up a project with a

service account, and obtaining an access token, see the quickstart (https://cloud.google.com/speech-to-text/docs/quickstart-protocol).

```
curl -s -H "Content-Type: application/json" \
    -H "Authorization: Bearer "$(gcloud auth print-access-token) \
    https://speech.googleapis.com/v1p1beta1/speech:recognize \
    --data '{
    "config": {
        "encoding": "FLAC",
        "sampleRateHertz": 16000,
        "languageCode": "en-US",
        "enableWordTimeOffsets":  false,
        "metadata": {
            "interactionType": "VOICE_SEARCH",
            "industryNaicsCodeOfAudio": 23810,
            "microphoneDistance": "NEARFIELD",
            "originalMediaType": "AUDIO",
            "recordingDeviceType": "OTHER_INDOOR_DEVICE",
            "recordingDeviceName": "Polycom SoundStation IP 6000",
            "originalMimeType": "audio/mp3",
            "obfuscatedId": "11235813",
            "audioTopic": "questions about landmarks in NYC"
        }
    },
    "audio": {
        "uri":"gs://cloud-samples-tests/speech/brooklyn.flac"
    }
}
```

See the `RecognitionConfig` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig) reference documentation for more information on configuring the request body.

If the request is successful, the server returns a `200 OK` HTTP status code and the response in JSON format:

```
{
  "results": [
    {
      "alternatives": [
        {
          "transcript": "how old is the Brooklyn Bridge",
          "confidence": 0.98360395
        }
      ]
    }
```

```
  ]
}
```


Last updated January 21, 2020.