Transcribing long audio files

This page demonstrates how to transcribe long audio files (longer than 1 minute) to text using asynchronous speech recognition.

Asynchronous speech recognition starts a long running audio processing operation. Use asynchronous speech recognition to recognize audio that is longer than a minute. For shorter audio, Synchronous Speech Recognition (https://cloud.google.com/speech-to-text/docs/sync-recognize) is faster and simpler.

You can retrieve the results of the operation via the google.longrunning.Operations (https://cloud.google.com/speech-to-text/docs/reference/rest/v1/operations#resource-operation) interface. Results remain available for retrieval for 5 days (120 hours). Audio content can be sent directly to Speech-to-Text or it can process audio content that already resides in Google Cloud Storage. See also the audio limits (https://cloud.google.com/speech-to-text/quotas) for asynchronous speech recognition requests.

The Cloud Speech-to-Text v1 is officially released and is generally available from the **`https://speech.googleapis.com/v1/speech`** endpoint. The Client Libraries (https://cloud.google.com/speech-to-text/docs/reference/libraries) are released as Alpha and will likely be changed in backward-incompatible ways. The client libraries are currently not recommended for production use.

These samples require that you have set up **`gcloud`** and have created and activated a service account. For information about setting up **`gcloud`**, and also creating and activating a service account, see Quickstart (https://cloud.google.com/speech-to-text/docs/quickstart).

The samples also use a Cloud Storage bucket (https://cloud.google.com/storage/docs/key-terms#buckets) to store the raw audio input for the long-running transcription process.

PROTOCOL GCLOUD COMMAND MORE

Refer to the **`speech:longrunningrecognize`** (https://cloud.google.com/speech-to-text/docs/reference/rest/v1/speech/longrunningrecognize) 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 -X POST \
     -H "Authorization: Bearer "$(gcloud auth application-default print-access-toke
     -H "Content-Type: application/json; charset=utf-8" \
     --data "{
  'config': {
    'language_code': 'en-US'
  },
  'audio':{
    'uri':'gs://gcs-test-data/vr.flac'
  }
}" "https://speech.googleapis.com/v1/speech:longrunningrecognize"
```

See the RecognitionConfig (https://cloud.google.com/speech-to-text/docs/reference/rest/v1/RecognitionConfig) and RecognitionAudio (https://cloud.google.com/speech-to-text/docs/reference/rest/v1/RecognitionAudio) 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:

```
{
  "name": "7612202767953098924"
}
```

where `name` is the name of the long running operation created for the request.

Wait for processing to complete. Processing time differs depending on your source audio. In most cases, you will get results in half the length of the source audio. You can get the status of your long-running operation by making a `GET` request to the `https://speech.googleapis.com/v1/operations/` endpoint. Replace ***`your-operation-name`*** with the `name` returned from your `longrunningrecognize` request. You can get the estimated progress of the request from the `progressPercent` (https://cloud.google.com/speech-to-text/docs/reference/rest/v1/LongRunningRecognizeMetadata) field.

```
curl -H "Authorization: Bearer "$(gcloud auth application-default print-access-toke
     -H "Content-Type: application/json; charset=utf-8" \
     "https://speech.googleapis.com/v1/operations/your-operation-name"
```

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


```
{
  "name": "7612202767953098924",
  "metadata": {
    "@type": "type.googleapis.com/google.cloud.speech.v1.LongRunningRecognizeMetada
    "progressPercent": 100,
    "startTime": "2017-07-20T16:36:55.033650Z",
    "lastUpdateTime": "2017-07-20T16:37:17.158630Z"
  },
  "done": true,
  "response": {
    "@type": "type.googleapis.com/google.cloud.speech.v1.LongRunningRecognizeRespon
    "results": [
      {
        "alternatives": [
          {
            "transcript": "okay so what am I doing here...(etc)...",
            "confidence": 0.96096134,
          }
        ]
      },
      {
        "alternatives": [
          {
            ...
          }
        ]
      }
    ]
  }
}
```


If the operation has not completed, you can poll the endpoint by repeatedly making the `GET` request until the `done` property of the response is `true`.
