

[AI & Machine Learning Products](https://cloud.google.com/products/machine-learning/) (<https://cloud.google.com/products/machine-learning/>)

[Cloud Speech-to-Text](https://cloud.google.com/speech-to-text/) (<https://cloud.google.com/speech-to-text/>)

[Documentation](https://cloud.google.com/speech-to-text/docs/) (<https://cloud.google.com/speech-to-text/docs/>) [Guides](#)

Transcribing phone audio with enhanced models

This tutorial shows how to transcribe the audio recorded from a phone using Speech-to-Text.

Audio files can come from many different sources. Audio data can come from a phone (like voicemail) or a soundtrack included in a video file.

Cloud Speech-to-Text can use one of several machine learning *models* (<https://cloud.google.com/speech-to-text/docs/basics#select-model>) to transcribe your audio file, to best match the original source of the audio. You can get better results from your speech transcription by specifying the source of the original audio. This allows the Cloud Speech-to-Text to process your audio files using a machine learning model trained for data similar to your audio file.

Caution: This tutorial uses the [enhanced speech recognition model](https://cloud.google.com/speech-to-text/docs/enhanced-models) (<https://cloud.google.com/speech-to-text/docs/enhanced-models>). Cloud Speech-to-Text has two types of phone call model that you can use for speech recognition, a standard model and an enhanced model. The enhanced model can provide better results at a higher price (although you can reduce the price by opting into [data logging](https://cloud.google.com/speech-to-text/docs/data-logging) (<https://cloud.google.com/speech-to-text/docs/data-logging>)).

Objectives

- Send a audio transcription request for audio recorded from a phone (like voicemail) to Cloud Speech-to-Text.
- Specify an [enhanced speech recognition model](https://cloud.google.com/speech-to-text/docs/enhanced-models) (<https://cloud.google.com/speech-to-text/docs/enhanced-models>) for an audio transcription request.

Costs

This tutorial uses billable components of Cloud Platform, including:

- Speech-to-Text

Use the [Pricing Calculator](https://cloud.google.com/products/calculator/#id=411d8ca1-210f-4f2c-babd-34c6af2b5538)

(<https://cloud.google.com/products/calculator/#id=411d8ca1-210f-4f2c-babd-34c6af2b5538>) to generate a cost estimate based on your projected usage. New Cloud Platform users might be eligible for a [free trial](https://cloud.google.com/free-trial) (<https://cloud.google.com/free-trial>).

Before you begin

This tutorial has several prerequisites:

- You've [set up a Speech-to-Text project](https://cloud.google.com/speech-to-text/docs/quickstart-client-libraries) (<https://cloud.google.com/speech-to-text/docs/quickstart-client-libraries>) in the Google Cloud Console.
- You've set up your environment using [Application Default Credentials](https://cloud.google.com/speech-to-text/docs/quickstart-client-libraries) (<https://cloud.google.com/speech-to-text/docs/quickstart-client-libraries>) in the Google Cloud Console.
- You have set up the development environment for your chosen programming language.
- You've installed the [Google Cloud Client Library](https://cloud.google.com/speech-to-text/docs/reference/libraries) (<https://cloud.google.com/speech-to-text/docs/reference/libraries>) for your chosen programming language.

Sending a request

To best transcribe audio captured on a phone, like a phone call or voicemail, you can set the `model` field in your `RecognitionConfig`

(<https://cloud.google.com/speech-to-text/docs/reference/rest/v1/RecognitionConfig>) payload to `phone_model`. The `model` field tells Speech-to-Text API which speech recognition model to use for the transcription request.

Note: The Speech-to-Text API supports the `phone_call` model for the en-US language only. For other languages, use the `video` or `default` models.

You can improve the results of phone audio transcription by using an [enhanced model](https://cloud.google.com/speech-to-text/docs/enhanced-models) (<https://cloud.google.com/speech-to-text/docs/enhanced-models>). To use an enhanced model, you set the `useEnhanced` field to `true` in your [RecognitionConfig](https://cloud.google.com/speech-to-text/docs/reference/rest/v1/RecognitionConfig) (<https://cloud.google.com/speech-to-text/docs/reference/rest/v1/RecognitionConfig>) payload.

Note: If you do not want to use the enhanced model in this tutorial, you can remove the `useEnhanced` field from your request to the Cloud Speech-to-Text.

The following code samples demonstrate how to select a specific transcription model when calling Cloud Speech-to-Text.

PROTOCOL

C#

GO

MORE ▾

Refer to the [speech:recognize](https://cloud.google.com/speech-to-text/docs/reference/rest/v1/speech/recognize)

(<https://cloud.google.com/speech-to-text/docs/reference/rest/v1/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) (<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) (<https://cloud.google.com/speech-to-text/docs/quickstart-protocol>).

```
curl -s -H "Content-Type: application/json" \  
  -H "Authorization: Bearer "$(gcloud auth application-default print-access-token \  
  https://speech.googleapis.com/v1/speech:recognize \  
  --data "{  
  'config': {  
    'encoding': 'LINEAR16',  
    'languageCode': 'en-US',  
    'enableWordTimeOffsets': false,  
    'enableAutomaticPunctuation': true,  
    'model': 'phone_call',  
    'useEnhanced': true  
  },  
  'audio': {  
    'uri': 'gs://cloud-samples-tests/speech/commercial_mono.wav'  
  }  
}"
```

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

(<https://cloud.google.com/speech-to-text/docs/reference/rest/v1/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": "Hi, I'd like to buy a Chromecast. I was wondering whether",
          "confidence": 0.8930228
        }
      ],
      "resultEndTime": "5.640s"
    },
    {
      "alternatives": [
        {
          "transcript": " Certainly, which color would you like? We are blue black",
          "confidence": 0.9101991
        }
      ],
      "resultEndTime": "10.220s"
    },
    {
      "alternatives": [
        {
          "transcript": " Let's go with the black one.",
          "confidence": 0.8818244
        }
      ],
      "resultEndTime": "13.870s"
    },
    {
      "alternatives": [
        {
          "transcript": " Would you like the new Chromecast Ultra model or the regu",
          "confidence": 0.94733626
        }
      ],
      "resultEndTime": "18.460s"
    },
    {
      "alternatives": [
```

```
{
  "transcript": " Regular Chromecast is fine. Thank you. Okay. Sure. Would",
  "confidence": 0.9519095
},
{
  "alternatives": [
    {
      "transcript": " Express, please.",
      "confidence": 0.9101229
    }
  ],
  "resultEndTime": "28.260s"
},
{
  "alternatives": [
    {
      "transcript": " Terrific. It's on the way. Thank you. Thank you very much",
      "confidence": 0.9321616
    }
  ],
  "resultEndTime": "34.150s"
}
]
```

Caution: Enhanced models cost more than standard models. Review the [pricing](https://cloud.google.com/speech-to-text/pricing) (<https://cloud.google.com/speech-to-text/pricing>) for more details.

Cleaning up

To avoid incurring charges to your Google Cloud Platform account for the resources used in this tutorial:

Deleting the project

The easiest way to eliminate billing is to delete the project that you created for the tutorial.

To delete the project:


Caution: Deleting a project has the following effects:

- **Everything in the project is deleted.** If you used an existing project for this tutorial, when you delete it, you also delete any other work you've done in the project.
- **Custom project IDs are lost.** When you created this project, you might have created a custom project ID that you want to use in the future. To preserve the URLs that use the project ID, such as an **appspot.com** URL, delete selected resources inside the project instead of deleting the whole project.

If you plan to explore multiple tutorials and quickstarts, reusing projects can help you avoid exceeding project quota limits.

1. In the Cloud Console, go to the **Manage resources** page.

[GO TO THE MANAGE RESOURCES PAGE \(HTTPS://CONSOLE.CLOUD.GOOGLE.COM/IAM-ADMIN/PROJ...](https://console.cloud.google.com/iam-admin/projects)

2. In the project list, select the project you want to delete and click **Delete** .

3. In the dialog, type the project ID, and then click **Shut down** to delete the project.


Deleting instances

To delete a Compute Engine instance:

1. In the Cloud Console, go to the **VM Instances** page.

[GO TO THE VM INSTANCES PAGE \(HTTPS://CONSOLE.CLOUD.GOOGLE.COM/COMPUTE/INSTANCES...](https://console.cloud.google.com/compute/instances)

2. Click the checkbox for the instance you want to delete.


3. Click **Delete**  to delete the instance.

Deleting firewall rules for the default network

To delete a firewall rule:

1. In the Cloud Console, go to the **Firewall Rules** page.

[GO TO THE FIREWALL RULES PAGE \(HTTPS://CONSOLE.CLOUD.GOOGLE.COM/NETWORKING/FIREV...](https://console.cloud.google.com/networking/firewall-rules)

2. Click the checkbox for the firewall rule you want to delete.
3. Click **Delete**  to delete the firewall rule.

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (https://creativecommons.org/licenses/by/4.0/), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0) (https://www.apache.org/licenses/LICENSE-2.0). For details, see our [Site Policies](https://developers.google.com/terms/site-policies) (https://developers.google.com/terms/site-policies). Java is a registered trademark of Oracle and/or its affiliates.

Last updated January 21, 2020.