

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

# Como separar locutores diferentes em uma gravação de áudio

## Beta

Esta é uma versão BetaCloud Speech API Speaker Diarization. Este recurso talvez sejam alteradas de formas incompatíveis e não estão sujeitos a nenhum SLA ou política de suspensão de uso.

Veja nesta página como criar rótulos para locutores diferentes em dados de áudio transcritos pela Cloud Speech-to-Text.

Às vezes, dados de áudio contêm amostras de mais de uma pessoa falando. Por exemplo, o áudio de uma chamada telefônica geralmente apresenta vozes de duas ou mais pessoas. O ideal é que a transcrição dessa chamada identifique quem fala em quais ocasiões.

## Diarização de locutor

Com a API Cloud Speech-to-Text, é possível reconhecer vários locutores no mesmo clipe de áudio. Ao enviar uma solicitação de transcrição de áudio para a Cloud Speech-to-Text, você pode incluir um parâmetro para que a API identifique os diferentes locutores presentes na amostra. Esse recurso, chamado de *diarização de locutor*, detecta a alternância entre os locutores e rotula por número as vozes individuais identificadas no áudio.

Quando você ativa a diarização de locutor na sua solicitação de transcrição, a Cloud Speech-to-Text tenta distinguir as diferentes vozes incluídas na amostra de áudio. O resultado da transcrição rotula cada palavra com um número atribuído a locutores individuais. Palavras ditas pelo mesmo locutor têm o mesmo número. O resultado da transcrição incluirá um número diferente para cada locutor que a Cloud Speech-to-Text puder identificar de forma exclusiva na amostra de áudio.

Com o uso da diarização de locutor, a Cloud Speech-to-Text produz um conjunto atualizado de todos os resultados fornecidos na transcrição. Cada resultado inclui as palavras do resultado anterior. Assim, você verá os resultados completos e diarizados da transcrição na matriz `words` do resultado final.

**Observação:** a Cloud Speech-to-Text oferece diarização de locutor apenas para os idiomas inglês americano (en-US), inglês indiano (en-IN) e espanhol (es-ES). Esse recurso é compatível apenas para a

transcrição de chamadas telefônicas.

## Como ativar a diarização de locutor em uma solicitação

Para ativar a diarização de locutor, é necessário definir o campo `enableSpeakerDiarization` como `true` nos parâmetros `RecognitionConfig` (<https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig>) da solicitação. Se quiser melhorar os resultados da transcrição, especifique o número de locutores presentes no clipe de áudio, definindo o campo `diarizationSpeakerCount` nos parâmetros `RecognitionConfig` (<https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/RecognitionConfig>). Caso não seja informado um valor para `diarizationSpeakerCount`, a Cloud Speech-to-Text usará um valor padrão.

A Speech-to-Text é compatível com a diarização de locutor em todos os métodos de reconhecimento de fala: `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>) e `streaming` (<https://cloud.google.com/speech-to-text/docs/reference/rpc/google.cloud.speech.v1p1beta1#google.cloud.speech.v1p1beta1.StreamingRecognizeRequest>).

Vea no snippet de código a seguir como ativar a diarização de locutor em uma solicitação de transcrição para a Cloud Speech-to-Text.

PROTOCOLO	JAVA	NODE.JS	PYTHON
<p>Consulte o endpoint <code>speech:recognize</code> (<a href="https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/speech/recognize">https://cloud.google.com/speech-to-text/docs/reference/rest/v1p1beta1/speech/recognize</a>) da API para mais detalhes.</p> <p>Para executar o reconhecimento de fala síncrono, faça uma solicitação POST e forneça o corpo apropriado a ela. Veja abaixo um exemplo de uma solicitação POST que usa <code>curl</code>. Nele, usamos o token de acesso de uma conta de serviço configurada no projeto por meio do <a href="https://cloud.google.com/sdk">SDK do Cloud</a> (<a href="https://cloud.google.com/sdk">https://cloud.google.com/sdk</a>) do Google Cloud Platform. Para instruções sobre como instalar o SDK do Cloud, configurar um projeto com uma conta de serviço e receber um token de acesso, consulte o <a href="https://cloud.google.com/speech-to-text/docs/quickstart-protocol">Guia de início rápido</a> (<a href="https://cloud.google.com/speech-to-text/docs/quickstart-protocol">https://cloud.google.com/speech-to-text/docs/quickstart-protocol</a>).</p>			

```
curl -s -H "Content-Type: application/json" \
-H "Authorization: Bearer $(gcloud auth application-default print-access-token)" \
https://speech.googleapis.com/v1p1beta1/speech:recognize \
--data '{
  "config": {
    "encoding": "LINEAR16",
    "languageCode": "en-US",
    "enableSpeakerDiarization": true,
    "diarizationSpeakerCount": 2,
    "model": "phone_call"
  },
  "audio": {
    "uri": "gs://cloud-samples-tests/speech/commercial_mono.wav"
  }
}' > speaker-diarization.txt
```

Quando a solicitação é bem-sucedida, o servidor retorna um código de status HTTP 200 OK e a resposta no formato JSON, salva em um arquivo chamado `speaker-diarization.txt`.

```
{
  "results": [
    {
      "alternatives": [
        {
          "transcript": "hi I'd like to buy a Chromecast and I was wondering",
          "confidence": 0.92142606,
          "words": [
            {
              "startTime": "0s",
              "endTime": "1.100s",
              "word": "hi",
              "speakerTag": 2
            },
            {
              "startTime": "1.100s",
              "endTime": "2s",
              "word": "I'd",
              "speakerTag": 2
            },
            {
              "startTime": "2s",
              "endTime": "2s",
              "word": "like",
              "speakerTag": 2
            },
            {
              "startTime": "2s",
              "endTime": "2.100s",

```

```
        "word": "to",
        "speakerTag": 2
    },
    ...
    {
        "startTime": "6.500s",
        "endTime": "6.900s",
        "word": "certainly",
        "speakerTag": 1
    },
    {
        "startTime": "6.900s",
        "endTime": "7.300s",
        "word": "which",
        "speakerTag": 1
    },
    {
        "startTime": "7.300s",
        "endTime": "7.500s",
        "word": "color",
        "speakerTag": 1
    },
    ...
    ]
}
],
"languageCode": "en-us"
}
]
```

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/) (<https://creativecommons.org/licenses/by/3.0/>), 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.

Última atualização a Abril 26, 2019.