

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

Práticas recomendadas

Este documento contém recomendações sobre como fornecer dados de fala à Speech-to-Text API. Estas orientações foram concebidas para proporcionar maior eficiência e precisão, além de tempos de resposta razoáveis do serviço. O uso da Speech-to-Text API é melhor quando os dados enviados ao serviço estão dentro dos parâmetros descritos neste documento.

Se você seguir estas diretrizes e não conseguir os resultados que espera da API, consulte [Solução de problemas e suporte \(https://cloud.google.com/speech-to-text/docs/support\)](https://cloud.google.com/speech-to-text/docs/support).

Para os melhores resultados...	Se possível, evite...
Capture o áudio com uma taxa de amostra de 16.000 Hz ou superior.	Baixas taxas de amostra podem reduzir a precisão. No entanto, evite a repetição da coleta de amostras. Por exemplo, na telefonia, a taxa nativa geralmente é 8.000 Hz, que é a taxa que precisa ser enviada ao serviço.
Use um codec sem perdas para gravar e transmitir áudio. É recomendado FLAC ou LINEAR16 .	Usar mp3, mp4, m4a, mu-law, a-law ou outros codecs com perdas durante a gravação ou transmissão porque isso pode reduzir a precisão. Se o áudio já estiver em uma codificação não compatível com a API, transcodifique-a para FLAC ou LINEAR16 sem perda. Se o aplicativo precisar usar um codec com perdas para economizar largura de banda, recomendamos os codecs AMR_WB , OGG_OPUS ou SPEEX_WITH_HEADER_BYTE , nessa ordem.
O reconhecedor foi projetado para ignorar ruídos e vozes de fundo sem cancelamento de ruído adicional. No entanto, para resultados ideais, posicione o microfone o mais próximo possível do usuário, particularmente quando houver ruído de fundo.	Ruído de fundo excessivo e ecos podem reduzir a precisão, especialmente se um codec com perdas também é usado.
Se você está capturando áudio de mais de uma pessoa e cada pessoa é gravada em um canal separado, envie cada canal separadamente para receber os melhores resultados de reconhecimento. No entanto, se todos os alto-falantes estiverem misturados em uma gravação em um único canal, envie a gravação como ela está.	Muitas pessoas que falam ao mesmo tempo ou em volumes diferentes podem ser interpretadas como ruído de fundo e ignoradas.

Para os melhores resultados...**Se possível, evite...**

Use dicas de palavras e frases para adicionar nomes e termos ao vocabulário e aumentar a precisão de palavras e frases específicas.

O reconhecedor tem um vocabulário muito grande, no entanto os termos e nomes próprios fora do vocabulário não serão reconhecidos.

Para consultas ou comandos curtos, use **StreamingRecognize** com **single_utterance** definido como true. Isso otimiza o reconhecimento de enunciados curtos e também minimiza a latência.

Usar **Recognize** ou **LongRunningRecognize** para usos de consultas ou comandos curtos.

Taxa de amostra

Se possível, ajuste a taxa de amostra da fonte de áudio para 16.000 Hz. Caso contrário, defina o [sample_rate_hertz](https://cloud.google.com/speech-to-text/docs/reference/rpc/google.cloud.speech.v1#recognitionconfig)

(<https://cloud.google.com/speech-to-text/docs/reference/rpc/google.cloud.speech.v1#recognitionconfig>)

para que corresponda à taxa de amostra nativa da fonte de áudio em vez de fazer uma nova amostra.

Tamanho do frame

O reconhecimento de streaming detecta áudio ao vivo porque ele é capturado de um microfone ou de outra fonte de áudio. O stream de áudio é dividido em frames e enviado em mensagens **StreamingRecognizeRequest** consecutivas. Qualquer tamanho de frame é aceitável. Os frames maiores são mais eficientes, mas aumentam a latência. Um frame com tamanho de 100 milissegundos é recomendado como um bom equilíbrio entre latência e eficiência.

Pré-processamento de áudio

É melhor fornecer áudio que seja o mais limpo possível, usando um microfone de boa qualidade e bem posicionado. No entanto, aplicar ao áudio um processamento de sinal de redução de ruídos antes de enviá-lo ao serviço geralmente reduz a precisão do reconhecimento. O serviço foi projetado para lidar com áudios com ruídos.

Para conseguir os melhores resultados:

- posicione o microfone o mais próximo possível da pessoa que está falando, especialmente quando houver ruído de fundo;
- evite recortar o áudio;
- não use controle de ganho automático (AGC);
- todo o processamento de redução de ruído deve ser desativado;
- ouça algumas amostras do áudio. Ele precisa soar claro, sem distorção ou ruído inesperado.

Configuração das solicitações

Certifique-se de descrever com precisão os dados de áudio enviados com a solicitação para a API Speech-to-Text. Garantir que o `RecognitionConfig` (<https://cloud.google.com/speech-to-text/docs/reference/rpc/google.cloud.speech.v1#recognitionconfig>) da solicitação descreva `sampleRateHertz`, `encoding` e `languageCode` corretamente faz com que a transcrição e o faturamento sejam mais precisos.

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 Março 14, 2019.