



链滴

# Google Speech-to-Text API 初体验

作者: [flowaters](#)

原文链接: <https://ld246.com/article/1540180287743>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

# 背景

体验一下Google的语音转文本服务。

## 简介

Google Speech-to-Text服务，利用机器学习技术将语音转换为文本，无论音频长短均可实现。

## 初体验

官方提供三种工具：客户端库，gcloud工具和命令行。

使用gcloud可以无痛上手。

## gcloud准备工作

1. 开通账户：需要验证信用卡，具体略
2. 创建一个项目
3. 下载安装云SDK

## 发出请求

### 音频转录请求

请求

```
gcloud ml speech recognize gs://cloud-samples-tests/speech/brooklyn.flac --language-code en-US
```

结果

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

## 转录短音频文件

使用同步语音识别，将短音频文件(不到1分钟)转录为文字。

## 本地文件

### 指令

```
gcloud ml speech recognize PATH-TO-LOCAL-FILE --language-code='en-US'
```

## 限制

- 只支持单声道 (mono) audio

## 体验

```
gcloud ml speech recognize nls-sample-16k.wav --language-code='zh-CN'
```

结果

```
{
  "results": [{
    "alternatives": [{
      "confidence": 0.97500247,
      "transcript": "北京的天气"
    }]
  }]
}
```

## 远程文件

最上面第一个初体验正是远程文件。

## 文件格式

| 编解码器 | 名称 | 无损 | 使用说明 |

| **FLAC** | 免费无损音频编解码器 | 是 | 信息流要求使用 16 位或 24 位的位深 |

| **LINEAR16** | 线性 PCM | 是 | |

| **MULAW** |  $\mu$  律 | 否 | |

| **AMR** | 自适应多速率窄带 | 否 | 采样率必须为 8000 Hz |

| **AMR\_WB** | 自适应多速率宽带 | 否 | 采样率必须为 16000 Hz |

| **OGG\_OPUS** | Ogg 容器中的 Opus 编码音频帧 | 否 | 采样率必须为 8000 Hz、12000 Hz、16000 Hz、24000 Hz 或 48000 Hz 之一 |

| **SPEEX\_WITH\_HEADER\_BYTE** | Speex 宽带 | 否 | 采样率必须为 16000 Hz |

## 转录长音频文件

使用异步语音识别，将长音频文件（时长超过 1 分钟）转录为文字

长音频文件推荐使用 GCS URI，即将音频存储在 gs 上。

如果是本地音频，则有限制：时长 < 8 分钟，大小 <

## 远程文件初体验

## 提交请求

输入

```
gcloud ml speech recognize-long-running 'gs://cloud-samples-tests/speech/brooklyn.flac' --language-code='en-US' --async
```

响应

Check operation [875262839786112xxxx] for status.

```
{
  "name": "875262839786112xxxx"
}
```

## 查询结果

请求

```
gcloud ml speech operations describe 875262839786112xxxx
```

响应

```
{
  "done": true,
  "metadata": {
    "@type": "type.googleapis.com/google.cloud.speech.v1.LongRunningRecognizeMetadat",
    "lastUpdateTime": "2018-10-19T09:50:44.091939Z",
    "progressPercent": 100,
    "startTime": "2018-10-19T09:50:41.952946Z"
  },
  "name": "875262839786112xxxx",
  "response": {
    "@type": "type.googleapis.com/google.cloud.speech.v1.LongRunningRecognizeRespons",
    "results": [{
      "alternatives": [{
        "confidence": 0.98360395,
        "transcript": "how old is the Brooklyn Bridge"
      }]
    }]
  }
}
```

## 持续查询结果

查询

```
gcloud ml speech operations wait 875262839786112xxxx
```

响应

Waiting for operation [875262839786112xxxx] to complete...done.

```
{
  "@type": "type.googleapis.com/google.cloud.speech.v1.LongRunningRecognizeResponse",
  "results": [{
    "alternatives": [{
      "confidence": 0.98360395,
      "transcript": "how old is the Brooklyn Bridge"
    }]
  }]
}
```

## 本地文件初体验

### 提交

```
gcloud ml speech recognize-long-running "news-20170607-8k-50s.wav" --language-code='mn-Hans-CN' --async
```

结果

Check operation [30093786222362xxxx] for status.

```
{
  "name": "30093786222362xxxx"
}
```

### 查询

```
gcloud ml speech operations wait 30093786222362xxxx
```

结果

```
Waiting for operation [30093786222362xxxx] to complete...done.
.....
```

## 从流式输入转录音频

必须使用客户端库，暂且略过。

## 附录

### 测试中文音频

- [新闻联播音频](#)

## 参考

- [Cloud Speech-to-Text](#)
- [Google quickstart-gcloud](#)
- [音频编码简介](#)