

一、前言

二、语音识别原理

三、科大讯飞语音识别引擎

1.概述

2.离线SDK指南

一、前言

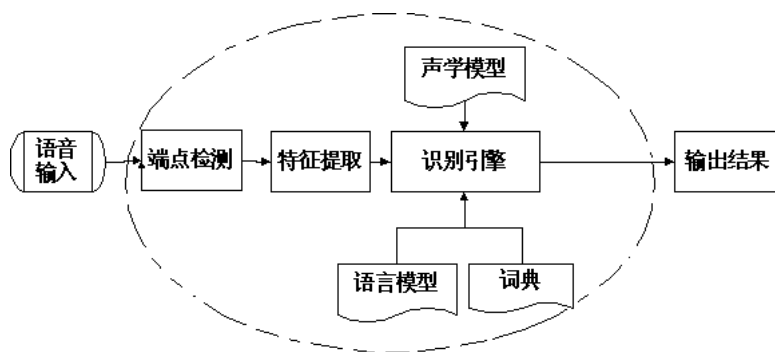


语音识别是指从语音到文本的转换，即让计算机能够把人发出的有意义的话音变成书面语言。通俗地说就是让机器能够听懂人说的话。所谓听懂，有两层意思，一是指把用户所说的话逐词逐句转换成文本；二是指正确理解语音中所包含的要求，而不要求所有词都转换正确。

语音识别技术(ASR)是以语音为研究对象，涉及到生理学、心理学、语言学、计算机科学以及信号处理等诸多领域，甚至还涉及到人的体态语言（如人在说话时的表情、手势等行为动作可帮助对方理解），目前常见的识别技术一般采用隐马尔可夫模型HMM(Hidden Markov Model)来建模。它的基本原理是系统首先对大量的真实世界中的人的语音进行分析，并建立模型。在识别时，系统同样先对输入的语音进行分析，提取特征，然后通过一定的算法并使用已建立的模型，从而识别出用户所说的话。

随着语音识别技术的逐渐成熟，语音识别技术开始得到广泛的应用，涉及日常生活的各个方面如电信、金融、新闻、公共事业等各个行业，通过采用语音识别技术，可以极大的简化这些领域的业务流程以及操作；提高系统的应用效率。

二、语音识别原理



语音识别原理图

从上图中看出，输入的语音首先要通过端点检测（也称静音检测模块），检测得到输入语音的有效语音数据部分（也就是去除了静音数据、噪声数据等）；然后进入特征提取模块，这里需要提取语音信号最本质的信息，一般采用MFCC（梅尔倒谱系数：Mel-scale Frequency Cepstral Coefficients）特征，它是采用复杂的数字信号处理技术来完成的；语音信号的特征信息被送入识别引擎，在声学模型和语言模型的指导下进行识别，最终得到识别结果，这里常常需要采用非常复杂的搜索算法，简单地说，它需要根据语言模型考虑各种词的组合（句子）及其发生概率，对于每一种可能的句子，都给出一个得分，选择得分最大者作为识别结果。但是由于可能的搜索空间非常大，所以需要利用剪枝技术，有效地去除大量冗余空间，加快搜索速度，满足实际系统的要求。

拓展：

在语音识别（Speech Recognition）和话者识别（Speaker Recognition）方面，最常用到的语音特征就是梅尔倒谱系数（Mel-scale Frequency Cepstral Coefficients，简称MFCC）。根据人耳听觉机理的研究发现，人耳对不同频率的声波有不同的听觉敏感度。从200Hz到5000Hz的语音信号对语音的清晰度影响较大。两个响度不等的声音作用于人耳时，则响度较高的频率成分的存在会影响到对响度较低的频率成分的感受，使其变得不易察觉，这种现象称为掩蔽效应。由于频率较低的声音在内耳蜗基底膜上行波传递的距离大于频率较高的声音，故一般来说，低音容易掩蔽高音，而高音掩蔽低音较困难。在低频处的声音掩蔽的临界带宽较高频要小。所以，人们从低频到高频这一段频带内按临界带宽的大小由密到疏安排一组带通滤波器，对输入信号进行滤波。将每个带通滤波器输出的信号能量作为信号的基本特征，对此特征经过进一步处理就可以作为语音的输入特征。由于这种特征不依赖于信号的性质，对输入信号不做任何的假设和限制，又利用了听觉模型的研究成果。因此，这种参数比基于声道模型的LPCC相比具有更好的鲁棒性，更符合人耳的听觉特性，而且当信噪比降低时仍然具有较好的识别性能。

### 三、科大讯飞语音识别引擎



## 1. 概述

网址: <https://www.xfyun.cn/>

网址: <https://www.xfyun.cn/services/commandWord>

# 离线命令词识别

用户对设备（手机、玩具、家电等）说出操作指令（即“命令词”），设备即作出相应的反馈，开启语音交互

免费试用

服务管理



## { 产品优势 }



### 业界先进的高识别率

讯飞开放平台拥有先进的语音识别技术，核心技术达到先进水平，语音识别准确率已经达到实用标准



### 离线命令词识别

考虑到在线命令词识别对网络的依赖，在平台上我们还以装机量授权的方式提供离线命令词识别服务。基于嵌入式离线识别引擎，零流量实时响应，实现快速稳定的本地化语音服务



### 支持用户个性化定制

命令词识别可支持约两百条语法的词汇量，并且开发者可以自定义命令词，满足应用个性化需求。还能适应不同年龄、不同终端和不同噪声的应用环境

## { 应用场景 }



### 智慧驾驶

手握方向盘，哪有精力分神去选择音乐、打电话、看新闻.....但有了离线命令词识别，仅仅动动口说出您的命令即可



### 智能家居

手机App和智能电视、智能家居实现互动，只需要集成离线命令词识别能力，便可根据需要语音操控家居、家电



### 智能硬件

语音交互为穿戴设备等硬件提供了新的使用方式，离线命令词识别更是让其可以听从指令，还无需网络

## { 开发者资源 }



iOS

语音SDK文档



Windows

语音SDK文档



Linux

语音SDK文档



Android

语音SDK文档

{ 产品价格 }

套餐	体验版	基础包	中级包	高级包	豪华包
价格	免费	¥ 8000.00	¥ 17500.00	¥ 30000.00	¥ 125000.00
装机量/个	10 (90天试用)	2000	5000	10000	50000
授权内容	离线命令词识别	离线命令词识别	离线命令词识别	离线命令词识别	离线命令词识别
单机成本	¥ 0.00	¥ 4.00	¥ 3.50	¥ 3.00	¥ 2.50
使用产品	SDK下载	<a href="#">立即购买</a>	<a href="#">立即购买</a>	<a href="#">立即购买</a>	<a href="#">立即购买</a>

2. 离线SDK指南

[https://www.xfvun.cn/doc/asr/commandWord/Linux-SDK.html#\\_1%E3%80%81%E7%AE%80%E4%BB%8B](https://www.xfvun.cn/doc/asr/commandWord/Linux-SDK.html#_1%E3%80%81%E7%AE%80%E4%BB%8B)

当语音识别引擎将识别到的结果，以xml内容格式进行返回，例子如下：

```
1 <?xml version='1.0' encoding='utf-8' standalone='yes' ?><nlp>
2 <version>1.1</version>
3 <rawtext>关门</rawtext>
4 <confidence>100</confidence>
5 <engine>local</engine>
6 <result>
7 <focus>cmd</focus>
8 <confidence>100</confidence>
9 <object>
10 <cmd id="5">关门</cmd>
11 </object>
12 </result>
13 </nlp>
```

开发者需要关注置信度（confidence）字段，来判断该次命令词（关门）是否生效。