

TVS Device SDK iOS 版本接入说明

拟制：zacardfang

日期：2017 年 10 月

审核：

日期：

深圳腾讯计算机系统有限公司

版权所有 不得复制

版本修订记录

[illegible]

目录

TVS Device SDK iOS 版本接入说明	1
1. 简介.....	4
1.1. 名词解释.....	4
2. 入门.....	4
2.1. SDK 使用步骤	4
2.1.1. 引入 AISDK.framework.....	4
2.1.2. 引入头文件	5
2.1.3. SDK 初始化.....	5
2.2. 输入音频要求.....	5
2.3. 语音识别.....	6
2.3.1. 功能.....	6
2.4. 语义分析.....	8
2.4.1. 功能.....	8
2.5. 语音合成.....	9
2.5.1. 功能.....	9

1. 简介

1.1. 名词解释

TVS DEVICE SDK：提供腾讯语音唤醒、语音识别、语音合成、语义理解、语义服务技术解决方案的软件开发工具包。

语音唤醒：智能硬件/应用在休眠状态下通过个性化语音唤醒词被唤醒。

语音识别：将语音转变为对应的语句文本。

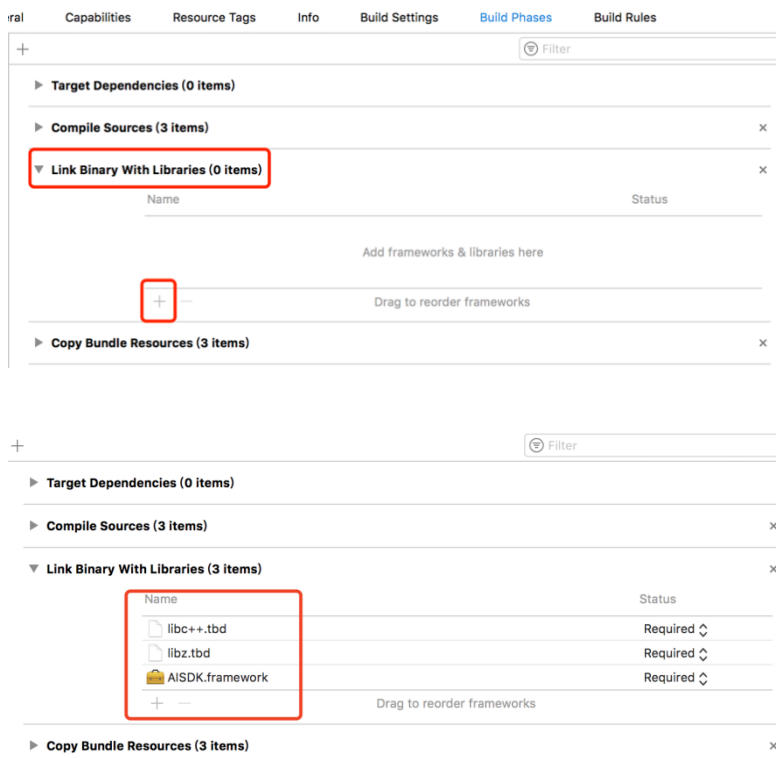
语义理解：将语句文本转变成结构化的实体、领域、意图、服务数据。

语音合成：将语句文本转变为流利的语音。

2. 入门

2.1. SDK 使用步骤

2.1.1. 引入 AISDK.framework



添加 AISDK.framework 一般放在工程根目录下。

添加 libc++、libz 依赖库。

2.1.2. 引入头文件

根据需要引入对应的头文件
#import <AISDK/AISDK.h>

2.1.3. SDK 初始化

使用 SDK 任何接口前，必须完成 SDK 初始化。

配置 info.plist 必须字段

AI_APPKEY	◇	String	ad445c3e-913e-488a-88ee-81da446078e4
AI_ACCESSTOKEN	◇	String	b58b050ac0004d58b0001cd42140dc7f
Vendor	◇	String	honor
ProductName	◇	String	dingdang
DeviceName	◇	String	PHONE

AI_APPKEY 填入应用的 appkey
AI_ACCESSTOKEN 填入应用的 accesstoken
DeviceName 固定，填入 CAR 或者 TV 或者 SPEAKER 或者 PHONE
ProductName 自定义, 英文产品名, 不要有空格
Vendor: 厂商英文名

//初始化函数
SpeechEngine *engine = [SpeechEngine sharedInstance];

- 其他重要方法
- 1、调用 addSession，将 VoiceSession 或者 SemanticSession 或者 TtsSession 注册到 SpeechEngine 的运行环境。
 - 2、调用 setConfig，调整配置。具体参照配置说明文档。

2.2. 输入音频要求

SDK 的唤醒识别和在线语音识别对输入音频的要求是一致的，要求如下表所示：

项目	要求
音频格式	PCM 格式
采样精度	16 位
采样率	16000Hz
声道	1 声道（单声道）
字节序	小端

2.3.语音识别

2.3.1. 功能

语音识别的接口，可以识别音频流中的文字，并返回到上层，方式为流式识别。例如用户还在说话中，就返回一部分文本。

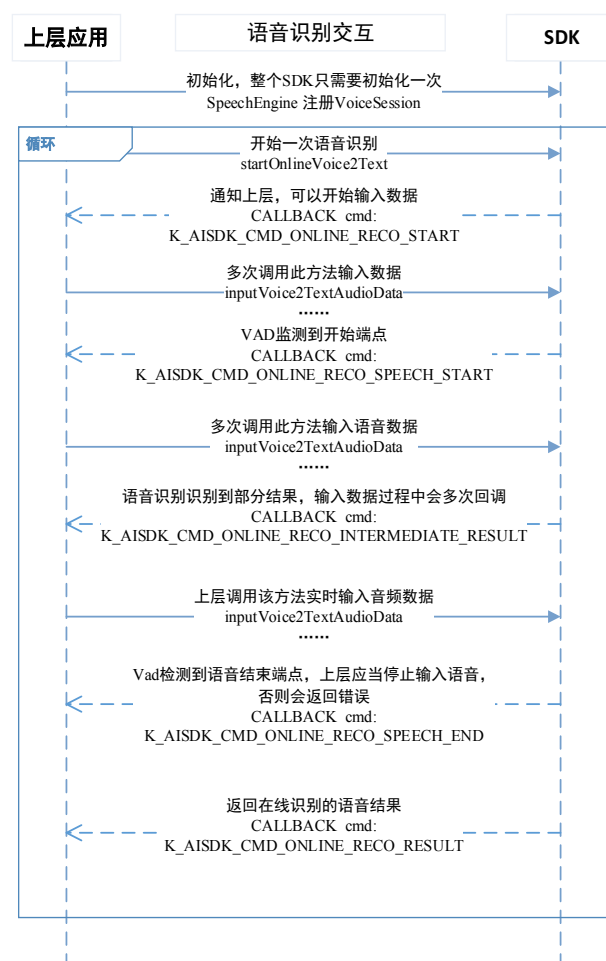
语音识别能力由 VoiceSession 提供，调用者需要实现 VoiceSessionDelegate 回调方法。

```
-(void)onOnlineVocieCallback:(NSInteger)cmd code:(NSInteger)code data:(NSString *)data userData:(id)userData;  
-(void)onOnlineVocieError:(NSInteger)cmd code:(NSInteger)code message:(NSString *)message userData:(id)userData;
```

语音识别支持自动模式和手动模式。

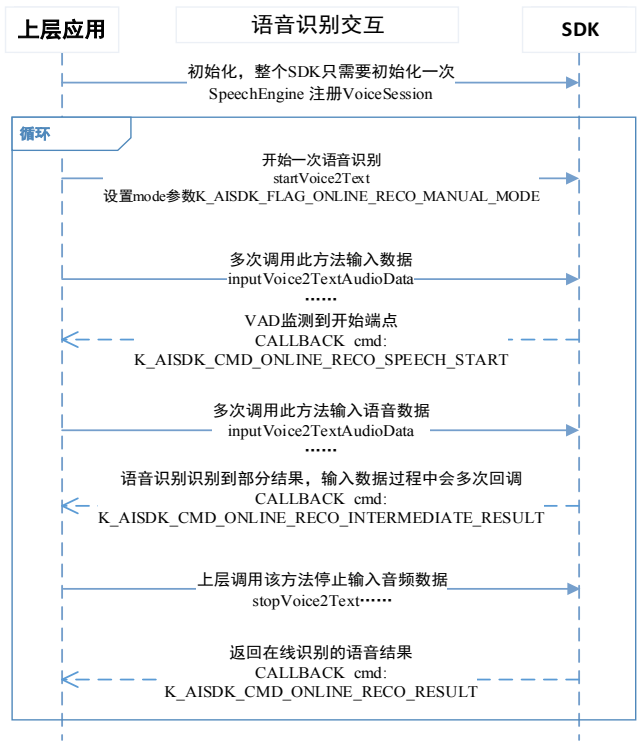
自动模式下， SDK 自动检测语音结束，并返回语音识别结果。

自动模式



手动模式下，调用 `stopVoice2Text` 控制语音识别的结束，此时会返回识别结果。
注意：`cancel` 开头的系列方法与 `stopVoice2Text` 的区别。调用 `cancel` 取消识别动作，无识别结果返回，但是会响应取消对应的回调命令。

手动模式



2.4.语义分析

2.4.1. 功能

语义的接口，可以将文本识别为的 domain 和 intent、语义实体，并返回对应的服务数据。例如把“我想听周杰伦的歌曲”识别为 domain 为 music、intent 为 play，带有的语义实体是歌手名字为“周杰伦”，服务数据为周杰伦的歌曲列表。

语义分析能力由 SemanticSession 提供，语义结果为 JSON 格式，具体请参考《SDK 语义格式》调用者需要实现 SemanticSessionDelegate 回调方法。



2.5.语音合成

2.5.1. 功能

语音合成的接口，可以将传入的文本，转换对应为音频数据。音频数据为单声道,采样率为 16000,位深度为 16bit 的 pcm 数据，对于长文本，可能分多次返回音频数据。

语音合成能力由 TtsSession 提供。TtsSession 返回的语义结果为 JSON 格式，具体请参考《SDK 语义格式》，调用者需要实现 TtsSessionDelegate 回调方法。

