

iFLY Speech Service1.0

讯飞语音+

产品白皮书

安徽科大讯飞信息科技股份有限公司
ANHUI USTC iFLYTEK CO., LTD.

版权声明

版权所有 © 2013, 安徽科大讯飞信息科技股份有限公司, 保留所有权利。

商标说明

安徽科大讯飞信息科技股份有限公司的产品是安徽科大讯飞信息科技股份有限公司专有。在提及其他公司及其产品时将使用各自公司所拥有的商标, 这种使用的目的仅限于引用。本文档可能涉及安徽科大讯飞信息科技股份有限公司的专利 (或正在申请的专利)、商标、版权或其他知识产权, 除非得到安徽科大讯飞信息科技股份有限公司的明确书面许可协议, 本文档不授予使用这些专利 (或正在申请的专利)、商标、版权或其他知识产权的任何许可协议。

不作保证声明

安徽科大讯飞信息科技股份有限公司不在此文档中的任何内容作任何明示或暗示的陈述或保证, 而且不对特定目的的适销性及适用性或者任何间接、特殊或连带的损失承担任何责任。本手册内容若有变动, 恕不另行通知。本手册例子中所用的公司、人名和数据若非特别声明, 均属虚构。未得到安徽科大讯飞信息科技股份有限公司明确的书面许可, 不得为任何目的、以任何形式或手段 (电子的或机械的) 复制或传播手册的任何部分。

保密声明

本文档 (包括任何附件) 包含的信息是保密信息。接收人了解其获得的本文档是保密的, 除用于规定的目的外不得用于任何目的, 也不得将本文档泄露给任何第三方。本软件产品受最终用户许可协议 (EULA) 中所述条款和条件的约束, 该协议位于产品文档和/或软件产品的联机文档中, 使用本产品, 表明您已阅读并接受了EULA 的条款。

版权所有© 安徽科大讯飞信息科技股份有限公司
Copyrights © Anhui USTC iFLYTEK CO., LTD.

目 录

1. 前言.....	1
2. 产品概况.....	1
3. 应用场景.....	2
4. 功能特性.....	4
4.1 语音服务.....	5
4.1.1 语音合成.....	5
4.1.2 语音识别.....	6
4.1.3 语义理解.....	9
4.1.4 语音搜索.....	10
4.2 界面展现.....	10
4.3 升级扩展.....	12
5. 技术指标.....	13
5.1 语音服务特性.....	13
5.1.1 合成的特性.....	13
5.1.2 识别的特性.....	14
5.2 资源占用指标.....	15
6. 联系我们.....	16
7. 公司介绍.....	17

1. 前言

语音是人们获取信息最天然的方式，随着移动互联网时代的到来，语音技术在人机交互中的重要性日渐凸显。智能语音交互能够解决现有移动终端设备上的文字输入不便（键盘太小）、阅读信息不便（屏幕太小）、以及开车和步行等场景下的操作不便等交互局限，其对于各种移动互联网终端下改善用户体验具有不可替代的作用。智能语音技术已经被业界公认为移动互联网时代的关键技术之一，蕴藏着巨大的市场前景和重要的战略意义。

早在 2010 年 10 月，科大讯飞面向移动互联网率先发布了能同时提供语音合成、语音搜索、语音听写的“讯飞语音云”，并在核心技术上达到国际领先水平，在全球英文合成大赛“Blizzard Challenge”中，讯飞第七次蝉联冠军；语音识别在复杂环境下的综合准确率已经达到 90%。语音作为一种自然交互正在被越来越多的移动互联网用户接受，语音交互服务，也正在进一步深入到了更多用户的日常生活中。

为进一步方便移动互联网开发者快速接入语音能力，语音云发布面向移动互联网应用的“讯飞语音+”。“讯飞语音+”是为移动互联网应用所量身打造的语音服务产品，是语音云客户端的关键组成部分。各类应用可通过其共享使用丰富灵活的语音能力，促进移动互联网应用的不断创新和发展，让更多的终端用户体验到语音技术给生活带来的改变。

2. 产品概况

“讯飞语音+”基于语音云开放平台面向移动互联网等行业应用，开放了语音合成、语音识别、语义理解、语音对话和声纹识别等智能语音交互技术的语音服务接口，是语音云平台对外输出语音服务的载体，其在终端设备上的表现形式是一个独立的应用程序。

“讯飞语音+”能够提供丰富灵活的语音服务，为各类移动终端应用集成语音能力打开了方便快捷的通道，降低语音应用的开发门槛和周期，使开发伙伴能快速构建“能听会说”的语音应用。并针对移动终端设备的通讯、存储和计算资源受限的特性，在保证语音服务质量的同时，实现轻量级的接口访问方式，兼顾了应用和终端用户的需求。另外，讯飞语音+具备升级扩展功能，保证用户始终能够获取最新最优质的语音服务。

“讯飞语音+”可以广泛应用在智能手机、平板电脑、智能电视、车载导航、智能家居

和智能玩具等移动互联网行业和领域。由“讯飞语音+”支撑的终端应用和成功案例正在越来越多，这些应用使用到语音的场景主要包括语音拨号、语音短信播报、语音电子书、语音导航、语音输入、语音闲聊等。

讯飞语音+的架构图如下图所示。智能手机和平板电脑等设备是语音服务的最终输出对象，处在产品架构中的第一层。这些设备上的各类终端应用程序共享使用讯飞语音+作为语音服务入口，在使用语音服务过程中，讯飞语音+的播音和录音模块与应用程序关系密切。合成类的语音服务是通过讯飞语音+播音输出给应用的，而识别和声纹等类语音服务是通过应用程序录音输出音频到讯飞语音+，以便进行后续处理的。合成分为本地合成和在线合成，本地合成包括词典处理、韵律分析和语言处理等模块，在线合成包括个性设置、音频编码和合成交互等模块。与合成类似，识别等业务包括端点检测、噪音消除、音频编码、识别交互等模块。对于需要网络交互的语音服务，讯飞语音+通过与语音云平台的网络交互来实现。



3. 应用场景

讯飞语音+支持的语音功能广泛，并能够针对不同行业应用的需求，采用不同的应用配置和功能配置，因此可广泛应用在智能手机、平板电脑、车载导航、智能电视等移动设备，

即将面市使用“讯飞语音+”技术的产品和成功应用案例越来越丰富。这些产品使用的语音服务场景主要包括，语音打电话、语音短信、语音输入、短信播报、语音电子书、语音导航、语音翻译、语音闲聊。

1) 语音拨号。语音拨号是通过语音控制进行拨打电话的操作，用户只需对手机说出“拨打张三的电话”、“拨打张三的住宅电话”等形式的语音指令，即可实现拨打张三的电话，同时用语音播报提示正在拨打张三的电话作为语音控制的反馈。语音打电话通过语音的操控方式，可有效简化传统语音打电话中的查找联系人号码等操作。

讯飞语音+向终端应用开放了讯飞语音云业界领先的语音识别技术，包括用户联系人的个性化识别技术。终端应用只需通过简单的两个步骤实现语音打电话功能：首先是将用户联系人通过讯飞语音+的 API 方式输入到云端，云端将返回和该用户联系人匹配的个性化语法模型标识；然后应用启动讯飞语音+的识别功能，并传递个性化语法标识，根据语音识别的联系人结果打开电话，便实现了语音打电话的功能。

2) 语音短信播报。短信播报是通过智能语音合成技术的方式实现短信播报，能够方便、实时、准确和自然地朗读出用户的短信内容，短信联系人、未读短信数量等信息。短信播报通过智能语音技术，有效改善了手机短信功能的使用体验，满足了手机的人性化交互需求。

讯飞语音+向终端应用开放科大讯飞业内领先的嵌入式语音合成技术，可实现中文、英文、粤语等语种的语音合成，并支持四川、湖南等方言发音，能够满足广大终端应用对短信播报的需求。终端应用只需通过简便的 API 调用方式，将文本输入讯飞语音+便可以轻松实现短信播报功能。

3) 语音输入。语音输入是指在短信、记事本以及任何应用程序的文本编辑框中，用户将文本内容以语音的方式说出来，并经过智能语音识别后将文本内容回显至文本编辑框。语音输入主要涉及到智能语音技术中的连续语音识别技术，近年来，随着这项技术的不断发展，语音输入的功能已经达到实用。据统计，采用触摸屏方式的文字输入速度平均为 60 字/分钟，而语音输入则可达到 180 字/分钟，语音输入在根本解决触摸屏输入文字信息不便的同时，让人机交互回归到人类天然的信息沟通方式。

讯飞语音+开放了语音云业界领先的连续语音识别技术，识别语种支持中文、英文、粤语等，并对音频的背景噪音有成熟的降噪方案，将语音输入满足了更多终端应用的需求。同时，讯飞语音+网络交互优化、录音、界面等语音识别相关接口进行了整合和优化，终端应用只需通过讯飞语音+调用几行简略的识别 API，便可获取到连续语音识别的能力，实现语

音输入的功能。

4) 语音电子书是以语音播报的方式来辅助电子书的阅读和使用,即对于存在手机、教育电子产品、MP3/MP4、及智能阅读终端等设备里的电子小说、报刊杂志、网络文学等这些具有大篇幅的电子类阅读内容,语音电子书可以有声有色为用户朗读出来,让读书变成听书,这样可以使得用户在使用语音电子书的同时,能够处理其他事务,提高了效率。

如同短信播报一样,电子书类应用可以通过集成讯飞语音+的嵌入式语音合成功能实现语音播报。讯飞语音+支持语境的智能判断和实时切换,因而更加适合对大篇幅的电子书内容朗读。同时对终端应用集成而言,讯飞语音+已经对音频数据缓冲、播音相关的接口进行了整合,使用方式也更为简便。

5) 语音导航。语音导航是指语音的方式说出查询的地点名称,语音识别返回结果后显示导航路线,并在导航的过程中用语音来提示用户的路线选择,包含两方面的语音交互过程:地点信息的语音识别和导航路线的语音播报。地点信息的识别有别于日常用语,需要收集到地理信息资源,并在此基础上建立语言模型,导航路线的播报也有别于日常用语,其音调、语速、韵律需要满足导航时的用户需求。语音导航通过将语音服务嵌入到车载导航设备中,让用户更加方便的使用车载导航设备,保证了用户的驾驶安全

讯飞语音+提供的地点信息的语音识别,是基于业内领先的语音识别引擎,在海量地理热点信息的基础上训练而成,地点识别的准确度达到 95%,并针对导航地点信息中特色生僻的地名路名,定制了导航版的语音配置,确保导航信息播报的准确、自然、清晰和悦耳。

6) 语音闲聊类应用是基于语音识别、语音合成和语义理解等技术基础上,讯飞语音+基于讯飞语音云提供的语音合成、语音识别技术前文已有相关介绍,不做多述。讯飞语音+提供的语义理解技术支持 30 个以上闲聊场景,以及 3000 个子类,并支持应用程序对场景进行自定义配置。讯飞语音+提供的语义 API 类似识别,仅在调用时传输的参数略有不同,终端应用可通过讯飞语音+的语义 API 实现语音闲聊的功能。

4. 功能特性

为了满足更多应用场景下的语音需求,方便更多的终端应用集成,让语音技术服务于更多地用户,讯飞语音+的功能主要包括这几个方面:提供更先进的语音服务,支持灵活的界面展现和触发方式、以及产品的更新升级功能。

4.1 语音服务

讯飞语音+提供的语音服务主要包括语音合成、语音识别、语义理解和语音搜索。语音合成是指将文本转为声音，让机器能够“说话”；语音识别是指将声音转为文字，让机器能够“听话”；语义理解是指能够辨别用户的说话行为和关键信息，让机器能够“思考”；语音搜索是指能够根据用户的说话内容在互联网进行信息搜索，让机器更为方便地使用。

4.1.1 语音合成

语音合成服务主要包括基于讯飞语音云的在线语音合成服务和基于嵌入式引擎的离线语音合成服务。在线语音合成服务运行资源占用小，合成音频具有更高的清晰度、自然度和可懂度，离线语音合成服务不需要依赖网络，能够实时响应。两种均采用先进的文本分析、韵律分析算法和大语料库的合成方法，合成语音已经接近真人的自然效果。其主要功能包括：

- 1) 高质量语音，将输入文本实时转换为流畅、清晰、自然和具有表现力的语音数据；
- 2) 多语种服务，整合了多语种语音合成引擎，可提供中文、中英文混读、纯正英文、粤语、粤英文混读的语音合成服务；
- 3) 多音色服务，提供丰富、风格多样化的音色选择，如浑厚淳正的男声，温柔甜美的女声，标准地道的英语男女声等等，所有音色库均继承科大讯飞语音合成技术一贯的优良品质。用户可供根据不同应用业务需要，选择最适合应用场景的语音风格，并支持实时动态的音色切换；
- 4) 高精度文本分析技术，保证了对文本中未登录词（如地名）、多音字、特殊符号（如标点、数字）、韵律短语等智能分析和处理；
- 5) 多字符集支持，支持输入 GB2312、GBK、Big5、Unicode 和 UTF-8 等多种字符集，普通文本和带有 CSSML 标注等多种格式的文本信息；
- 5) 多种数据输出格式，支持输出多种采用率的线性 Wav，A/U 率 Wav 和 Vox 等格式的语音数据；
- 7) 提供预录音合成模板，对合成文本中符合语音模板固定成分的文本使用发音人预录语

音，非固定成分使用合成语音。这种方法有利于改进定制领域合成效果，简化定制流程，加快定制速度，同时也使得预录音的使用更自然、更灵活，满足更广泛的应用需求；

8) 语音调整功能，开发接口提供了音量、语速、音高（基频）等多种合成参数动态调整功能；

9) 配置和管理工具，合成引擎提供了统一进行配置和管理的工具，完成了全局参数配置、用户词典、用户规则、定制资源包管理等功能；

10) 效果优化，合成引擎提供了以定制资源包和 CSSML 为代表的多种针对实际应用环境进行合成效果优化的方法；

11) 一致的访问方式，能以 C/S 方式访问远程的语音合成服务，并且提供与本地调用相同的开发接口，实现了完全透明的访问；

12) 背景音和预录音，提供了背景音和预录音的功能，大大提高了合成语音的自然度和表现力。

在实际的语音应用中，合成文本有时属于某个特定领域（如银行领域），其内容和形式具有显著的特征。针对这种特性，科大讯飞公司提出了领域定制的概念，领域定制是指在特定领域应用时使用适合的定制资源包来提高合成语音效果。“定制资源包”是一组语音合成引擎在特定领域应用时使用的资源的集合，包含音库、词库、外部规则库、自定义符号库等等。使用定制资源包有助于提升特定应用领域的语音合成效果。

4.1.2 语音识别

讯飞语音+提供与说话人无关的智能语音识别技术，来源于科大讯飞国际领先的语言识别引擎，支持连续语音识别和为自助语音服务提供的关键字语音识别，具备优秀的识别率，提供全面的开发支持，丰富的工具易于使用，核心技术上达到了国际领先水平。针对语音识别应用中面临的方言口音、背景噪声等问题，科大讯飞在实际业务系统中所收集的涵盖不同方言和不同类型背景噪声的海量语音数据，通过先进的区分性训练方法进行语音建模，使语音识别在复杂应用环境下均有良好的效果表现。语音识别主要包括以下功能特性：

（1） 前端语音处理

前端语音处理指利用信号处理的方法对说话人语音进行检测、降噪等预处理，以便得到最适合识别引擎处理的语音。主要功能包括：

1) 端点检测

端点检测是对输入的音频流进行分析，确定用户说话的起始和终止的处理过程。一旦检测到用户开始说话，语音开始流向识别引擎，直到检测到用户说话结束。这种方式使识别引擎在用户在说话的同时即开始进行识别处理。

2) 噪音消除

在实际应用中，背景噪声对于语音识别应用是一个现实的挑战，即便说话人处于安静的办公室环境，在电话语音通话过程中也难以避免会有一定的噪声。InterReco 语音识别系统具备高效的噪音消除能力，以适应用户在千差万别的环境中应用的要求。

3) 智能打断

智能打断功能使用户可以在自助语音服务的提示语播放过程中随时说出自己的需求，而无需等待播放结束，系统能够自动进行判断，立即停止提示语的播放，对用户的语音指示做出响应。该功能使人机交互更加高效、快捷、自然，有助于增强客户体验。

(2) 后端识别处理

后端识别处理对说话人语音进行识别，得到最适合的结果，主要特性有：

1) 大词汇量、独立于说话人的健壮识别功能

满足大词汇量、与说话人无关的识别要求。可以支持数十万条语法规模的词汇量；并能适应不同年龄、不同地域、不同人群、不同信道、不同终端和不同噪声环境的应用环境。

2) 连续语音识别

连续语音识别是指能够把用户说的任意语音转换成对应的文字信息，是移动互联时代全新的信息录入方式，可以大大提高在移动终端上的文字输入速度。配合语义理解后，连续语音识别就能够完成“听懂人说话”的功能，连续语音识别集成了讯飞研究院多年的研究成果，提供领先的、最准确、最顺畅的连续语音识别服务。

连续语音识别支持中文和中英文混读的常见语句听写，对于日常使用的常用对话有着很高的识别准确率，包含日常用语，如短信类、生活、交通、娱乐、科技、数字数值、名人、互联网热词、新闻、金融等领域。

3) 个性化语音识别

个性化识别是指为提高用户的个性化词条识别准确率，而采取的基于每个用户个性化词

条定制的语言模型的识别方式。个性化识别需由语音应用采集对识别效果有帮助的个性化语料信息，并传递在语音云的后台实时生成语法模型，并在用户后续进行语音识别的过程动态加载。

4) 置信度输出

置信度反映了识别结果的可信程度。语音识别引擎可以在返回识别结果时会携带该识别结果的置信度，应用程序可以通过置信度的值进行分析和后续处理。

5) 多识别结果

又称多候选技术，在某些识别过程中，识别引擎可以通过置信度判决的结果向应用程序返回满足条件的多个识别结果，而不是唯一的结果。识别系统提供了可能的识别结果列表，并按置信度结果从高到低进行排列。在业务设计中，可以根据应用要求向用户提供这些结果，供用户进行二次选择。通过置信度判决和多识别结果输出技术可以开发更加灵活、更加人性化的业务流程。

6) 说话人自适应

当用户与语音识别系统进行多次会话过程中，系统能够在线提取通话的语音特征，自动调整识别参数，使识别效果得到持续优化。

7) 多槽识别

语音识别的槽 (Slot) 代表一个关键字，即在一次会话过程中可以识别说话人语音中包含的多个关键字，这可以提高语音识别应用的效率，增强用户体验。

8) 热词识别

热词识别使得语音识别应用程序能够在说话者说话的同时检测一个特定的词或短语。

当说话者说到这个短语的时候，识别引擎会把控制权交还给应用程序。在应用程序使用该功能可以使识别器能够在后台监听输入的语音，直到用户说出特定的短语进行请求时才与用户交互。

9) 智能调整识别策略

InterReco 能够自动根据系统运行情况动态调整语音识别策略，在系统较忙时 (CPU 占用较高)，采用计算量较小但具有足够精度的策略以保证系统的响应速度；在系统不忙时 (CPU 占用较低)，采用精度更高的策略以达到更优的识别效果。该功能在保障稳定运行的基础上，充分利用系统的计算资源，有利于保护客户的设备投资。

10) 语音录入

允许用户通过语音方式动态增加识别语法。该功能使用户可以更加灵活的维护语法，扩展语音识别范围，并且能够提高识别系统对用户语音的适应能力，提高识别的准确率。

11) 访问日志

语音识别的访问日志在系统中有着非常重要的作用，该日志记录输入的音频、加载的语法、识别过程的中间结果、识别模块访问过程、识别使用的各种参数、识别结果以及当时的系统环境信息。这些数据是效果分析和优化的基础和依据。

4.1.3 语义理解

语义理解技术是整个人机交互的关键技术之一，是让计算机能够按需正确理解人类的自然语言的一项关键技术，具有较高的挑战性。针对语音交互中的语义理解广泛需求，科大讯飞研发了面向业务需求的基于文法规则的语义理解、基于相似问句语义距离度量的知识问答系统，以及基于本体库自动构建及推理的自动问答系统，为各种类型的语义理解需求提供相应的技术支撑。语义理解支持的特性包括：

1) 能够处理基于文法规则的实际业务需求。为支持用户通过自然语言向机器发出某种业务需求，如打电话、发短信、查天气、查股票、查航班、订酒店等，科大讯飞研发了基于文法规则的语义理解。通过建立一套灵活、通用的语义框架，并基于 ABNF 文法构建文法规则体系及相关解释、编译工具，重点解决基于文法的语义理解技术中集外名称识别难点，并引入基于发音相似度的模糊匹配策略实现语音识别错误的主动纠错，同时针对主流业务再辅以一定规模的人工知识建设工作，最终完成基于文法的语义理解技术及针对主流业务需求理解的相关知识库；

2) 具备基于相似问句语义距离度量的智能知识问答。互联网上已有的海量问答对是人类知识的一种表示，为实现知识问题，科大讯飞通过互联网上知识问答对的自动发现、下载和清洗，建立相应的问句及其答案的检索引擎，基于同义词近义词扩展、对立词反义词惩罚、及说法转换规则等技术的相似问句间语义距离度量，以及集成内容审查环节的答案过滤排序等一系列模块的研发，最终实现知识问答系统。语音云的第三方开发者可以基于此项特性建立自己的知识问答系统；

3) 具备基于本体库自动构建及推理的智能问答。本体是指一种对特定领域之中概念及其相互之间关系的形式化表达。科大讯飞针对面向某些特定领域，建立了本体概念关系体系，

具体包括音乐知识问答、电影知识问答、旅游知识问答等。通过研发基于互联网上的结构化数据实现本体库的实例的不断填充，并基于三元组的方式实现语义的在线或离线推理，以及检索加速机制，完成面向多个概念之间深层复杂型问题的自动回答系统。

4.1.4 语音搜索

语音搜索技术能够将用户输入的语音内容识别成文字，并经过智能搜索解析，给出用户识别结果并展现最匹配的搜索结果列表，这是移动互联网搜索领域的一种全新尝试方式，回归了人类用语言询问的本能，从本质上改善了搜索体验性，有广阔的市场前景。讯飞语音搜索技术基于前一章节述及的语音识别技术和文本搜索技术。其主要功能特性有：

1) 支持超大规模的语言模型

当前讯飞语音搜索使用的语言模型已达千万级条目，涵盖的热词范围包括全国各地名、音乐视频名称和时下网络热词，对这些热词的识别率达到 93%以上，为下一步精确的文本搜索打下了扎实的基础。

2) 支持语言模型的动态更新

热词语言模型可以通过训练工具动态更新，以满足各类热词库的动态更新需要。一般 1G 左右的新语料需要耗费 1~2 天的时间完成更新。

3) 支持针对不同用户给出最匹配的搜索结果列表

能够根据不同用户的特殊性，如所在位置、年龄阶层、兴趣爱好等特性，智能分析并给出最匹配的搜索结果列表。

4.2 界面展现

为了满足更多应用集成语音的需求，讯飞语音+提供的语音服务支持丰富灵活的界面展现方式，包括对话框、置顶\底\边、全屏、桌面悬浮、冒泡弹出式、用户自定义等。

1) 对话框。对话框是最常见的界面展现方式，在用户触发语音后，屏幕顶层出现语音交互的对话框界面。对话框界面主要展现使用帮助、正在启动录音、录音时的音量显示、等待语音识别结果、出错提示等信息。对话框的大小一般是屏幕尺寸的一半左右，对话框的位置可在屏幕任意坐标，但在语音交互结束后，对话框也随之消失。



图 4.2.1 对话框图 4.2.2 置底

2) 置顶\底\边。为了增加用户点击语音服务的可能，将语音界面固定放置在应用程序的顶部、底部或左右两边处。语音交互过程中，置顶\底\边的展现方式中展现的元素和对话框类似，一般情况下其尺寸较对话框小，在语音交互结束后，展现界面回归原始状态而不消失。

3) 全屏。全屏的展现方式可以认为是对话框放大至整个桌面，展现的内容和消失方式与对话框方式一样。全屏的目的在于让用户在语音交互时，更加集中注意力，从而提升语音服务的使用效果和体验，促进用户的语音活跃度。



图 4.2.3 全屏图 4.2.4 桌面悬浮

4) 桌面悬浮。在使用手机等设备时，桌面是用户最频繁接触的。桌面悬浮是指将应用的主要数据信息或 logo 固定在桌面上的展现方式，桌面悬浮可以增加用户使用语音服务的

机会。在语音交互过程中，由于尺寸限制，桌面悬浮需要提取出关键信息展现。另外，桌面悬浮不可遮挡用户的主要应用程序，否则容易引发用户的抵触。

5) 应用内悬浮。是指打开应用时，固定悬浮在应用的顶层界面上。只要应用在前台运行，悬浮界面便一直存在。应用内悬浮的语音交互过程信息展现和桌面悬浮方式类似，需要注意的是悬浮的位置不可以遮挡应用的主要信息界面。



图 4.2.5 应用内悬浮图 4.2.6 自定义

6) 用户自定义。为了让语音交互界面更好的适配每个应用，讯飞语音+支持语音交互过程中的界面自定义功能。自定义功能包括完整的交互流程、界面设计和提示语设计，用户自定义的展现方式有利于促进移动互联网上形成个性特色的应用。

4.3 升级扩展

随着信息技术的不断发展和语音的普及，语音云平台在发展壮大的同时逐步得到进化，讯飞语音+产品也将持续更新升级，以满足更多应用的语音需求，保证广大最终用户能体验到最新最领先的语音技术。讯飞语音+的产品更新包括两个方面：

1) 语音服务效果的更新。智能语音技术具有自适应学习的功能，语音服务的效果能够在用户持续使用的过程中得到提升。语音服务效果的更新是在后台完成，主要是通过云计算服务平台获取到海量数据，并以人工智能的方式进行标注和挖掘，完成语音交互核心引擎的优化，主要是声学模型、语言模型等的优化。语音服务更新主要是基于语音云的分布式计算能力实现，部署了声学模型建模、语言模型建模等模型的训练或自适应算法，并建立了核心

效果测试集合和测试流程，通过使用更多更新的语音语言数据，最终完成语音交互核心引擎的优化工作，保证语音服务核心效果的不断优化。

2) 产品的更新。讯飞语音+既是语音服务的使用入口，也是一个独立应用，产品的更新是指讯飞语音+应用的更新，即需要用户下载新的安装包来更新产品。这种方式更新能给最终用户带来更好的语音交互体验，包括讯飞语音+应用和交互界面的优化，响应时间的优化等。另外，产品升级更新前后，其对外语音接口调用方式能具备兼容性，因此对于终端应用开发者而言，讯飞语音+的产品更新对他们进行应用开发，不会造成额外开销。

5. 技术指标

5.1 语音服务特性

5.1.1 合成的特性

下表分别列出了通过讯飞语音+所支持的嵌入式语音合成和在线语音合成主要功能特性，每个功能对应的实现方式请参考开发文档。

离线 语音合成	文本类别	普通书面文本
	字符范围	ASCII 码；CJK 及扩展 A、B 区（香港增补字符集 HKSCS）；音标；常用符号
	输入文本代码页	UTF-16；UTF08 (CP65001)；简体中文 GB18030 (CP936，兼容 GB2312-80、GBK)；繁体中文 Big5 (CP950，支持 HKSCS)；Phonetic Plain (金山词霸音标字体)
	输入文本模式	块式文本输入；流式文本输入
	发音人	汉语普通话男女声；美式英语男女声；汉语方言男女声等
	语种	英语；汉语普通话（支持中英混读）；各种特色方言
	语境智能判断	支持，根据上下文环境智能判断字符在语境中的读法
	发音人自动切换	支持，根据上下文环境自动切换不同语种发音人
	输出数据格式	PCM signed 16bit
	输出采样率	8K/11K/16K (Hz) 可选
	文本标注	支持，完整的标注体系详见标记手册
	音量调节	支持
	语速调节	支持
	音调调节	支持

	音效*	支持回声、混响等 7 种合成音效，选配（默认不开启）
	全/半角	支持阿拉伯数字、英文字母、常用符号的全/半角形式
	阿拉伯数字处理	默认根据上下文智能判断；可指定按号码或数值处理； 语 言可设置汉语普通话和英语，默认为自动判断
	特殊符号处理	根据上下文智能判断
	多音字识别	支持普通多音字、姓名、地名路名等多音字的智能处理
	进度通知	支持
	同步回显	支持
	英语音标朗读	支持，可设置
	汉语拼音朗读	支持，可设置
	预设韵律	支持
	自然度	4.34（5 分制）
在线 语音合成	支持语种	汉语普通话、英语、粤语
	支持文本控制标记	支持 TCCM 文本标记，CSSML
	输出语音数据格式	8K/16K 16bit/8bit PCM, aLaw/uLaw, ADPCM
	支持字符集	GB2312, GBK, BIG5, UNICODE, GB18030
	识别汉字数量	支持 GBK 字库，共 21003 个简繁体汉字
	系统词库词组数量	约 300000 条
	支持用户定义词库	是
	语音数据缓冲区设置	支持 64K~2M Bytes
	断句参数设置	支持回车符处理参数和最大断句长度参数
	标点符号发音设置	支持标点符号发音方式的选择和具体发音的设置
	数字串发音设置	支持数字串发音方式的选择
	发音停顿风格设置	支持发音停顿风格的选择
	节奏正确率	97%
	句子易懂度	99%
	自然度	4.5（5 分制）

5.1.2 识别的特性

下表分别列出了通过讯飞语音+所支持的智能语音识别主要功能特性，每个功能对应的实现方式请参考开发文档。

语音识别	中文连续识别	支持普通话连续语音听写
	中英文混合识别	支持中文中混读的简单英文
	数字识别	支持数字串听写

	字母识别	支持
	标点智能预测	支持
	识别结果多候选输出	支持听写结果按照词语提供 1-10 个候选
	识别结果输出格式	支持 JSON 格式的听写输出结果
	语义解析	支持对听写结果做语义理解，提出用户关心的多个槽信息
	在线自适应	支持说话人自适应
	音频输入格式	8K/16K uLaw/aLaw，16bit PCM，AMR NB/WB，Speex NB/WB
	智能打断	支持
	状态保存、恢复	暂不支持
	自定义语言包	支持
	噪声消除	支持

5.2 资源占用指标

讯飞语音+提供的服务是轻量级的，对系统资源占用要求较低。下表具体列出了语音服务的尺寸、CPU、内存、耗电量、带宽等资源使用情况。

版本类别	基本项			参数
支持离线合成	安装包尺寸			2.5MB
	CPU 占用			60MIPS
	内存占用	ROM		3.0MB
		RAM	合成	2.0MB
			识别	2.5MB
	流量	合成		在线合成 100 个字，流量低于 60KB

	识别		在线识别 3 秒音频，流量低于 10KB
	耗电量		连续使用合成或识别 1 小时，耗电量不超过 10%。
不支持离线合成	安装包尺寸		600K
	CPU 占用		40MIPS
	内存占用	ROM	800KB
		RAM	合成 2.5MB
			识别 2.5MB
	流量	合成	在线合成 100 个字，流量低于 60KB
		识别	在线识别 3 秒音频，流量低于 10KB
	耗电量		连续使用合成或识别 1 小时，耗电量不超过 10%。

6. 联系我们

如果您在安装或使用中有任何问题或者建议，请与我们联系！！

联系时对问题的描述应尽量包含以下内容：

- 系统配置（包括 CPU、内存、硬盘、操作系统及产品版本等信息）
- 问题细节（包括问题的重现过程及合成的文本内容等）
- 问题重现（包括详细的操作过程）

◆ 电话支持

请于周一～周五，北京时间 9：00～17：00 间，拨打电话 0551—5309063 获得技术支持信息。

◆ 电子邮件支持

请将问题的详细描述发至：msp_support@iflytek.com

◆ 信件支持

请将问题详细描述发至：安徽省合肥市高新区望江西路 666 号讯飞大厦

邮政编码：230088

或传真至：0551—5331801 / 5331802

7. 公司介绍

科大讯飞是领先的专业语音技术提供商,拥有自主知识产权的智能语音技术代表了世界最高水平,并始终引领和不懈推动语音产业的发展。2008 年,科大讯飞在深圳证券交易所挂牌上市,股票代码:002230。

科大讯飞在智能语音技术领域有着长期的研究积累,并在中文语音合成、语音识别和口语评测等多项技术上拥有国际领先的成果。科大讯飞是我国唯一以语音技术为产业化方向的“国家 863 计划成果产业化基地”、“国家规划布局内重点软件企业”、“国家火炬计划重点高新技术企业”、“国家高技术产业化示范工程”,并被信息产业部确定为中文语音交互技术标准工作组组长单位,牵头制定中文语音技术标准。2003 年,科大讯飞获迄今中国语音产业唯一的“国家科技进步奖”,2005 年获中国信息产业自主创新最高荣誉“信息产业重大技术发明奖”。2006 年开始连续七届在英文语音合成国际大赛(Blizzard Challenge)中蝉联大赛第一名。在 2008 年由 NIST(美国国家标准技术研究院)举办的国际说话人识别评测大赛上,科大讯飞语音实验室送评的识别系统获得综合指标第一名。

语音识别和语音合成技术让信息设备与系统“能听会说”,实现人机语音交互,使人与机器之间沟通变得像人与人沟通一样高效、自然。智能语音技术有着非常广阔的应用前景。科大讯飞是中文语音技术市场的领导者,在电信、金融、能源、交通、政府等主流行业的份额达到 80%以上。

- 国家 863 计划成果产业化基地
- 国家规划布局内重点软件企业
- 国家火炬计划重点高新技术企业
- 国家高技术产业化示范工程基地
- 国家级优秀软件企业
- 中国电子政务 IT100 强企业
- 中国中文语音交互技术标准工作组组长
- 国家博士后科研工作站

如果您想了解更多信息，敬请访问 <http://www.iflytek.com>。