



专栏：智慧家庭

智能语音机顶盒的软硬件实现方案

施唯佳, 王琳, 贾立鼎

(中国电信股份有限公司上海研究院, 上海 200122)

摘 要: 在语音交互技术发展的大背景下, 智能音箱产品在提供内容娱乐、生活服务的同时, 更多地接入家庭生活场景中, 成为智能家居的控制中心。从智慧家庭下最新的业务需求出发, 详细分析为了适应新的业务需求和智能语音机顶盒需增加的功能要求, 提出了满足新业务功能的智能语音机顶盒的软硬件实现方案。

关键词: 智慧家庭; 远场拾音; 语音控制; 语音机顶盒

中图分类号: TN915

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2017292

Software and hardware implementation of smart hands-free set-top box

SHI Weijia, WANG Lin, JIA Liding

Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China

Abstract: In the background of the development of voice interactive technology, smart speaker becomes a smart home control center, which products in the provision of content entertainment, life services at the same time, and more access to family life scenes. From the smart home under the latest business needs, the increased functional requirements to meet the new business requirements and the requirements of the smart hands-free set-top box were detailed analyzed, the hardware and software implementation scheme of the smart hands-free set-top box to satisfy the new business function was put forward.

Key words: smart home, hands-free speaker, voice control, hands-free set-top box

1 引言

随着移动互联网技术的发展, 人机交互的要求不断深入。从对操作使用的个性化、方便快捷的智能化和人性化角度看, 无论是键盘还是触控, 都远远不能与语音相比, 语音才是人类沟通和获取信息最自然的便捷手段和方式。对家居设备的操控从普通的按键式遥控器, 到蓝牙语音遥控器,

发展到现在支持拾音功能的智能语音控制, 语音技术将解放人类的眼睛和双手, 成为最佳人机交互模式。智慧家庭是一种生活方式, 是让家中设备感知人的需求, 更好地为人服务, 成为家庭成员的亲情和纽带。真正的智慧家庭绝不是把不同的家电设备简单地连接起来或实现远程控制, 而是打通了家庭中的不同场景, 人们可以通过自然语音控制智能设备满足自己需求。本文着力于探

收稿日期: 2017-09-26; 修回日期: 2017-10-13

讨在智慧家庭需求下, 智能语音机顶盒的新增功能特性和软硬件实现方式, 并对智能语音机顶盒未来的发展方向进行了展望。

2 智能语音机顶盒的业务功能分析

工业和信息化部数据显示, 截至 2017 年 5 月末, 国内 IPTV 用户规模已破亿, 其中, 中国电信占据 8 成左右的市场份额。中国电信作为固网第一运营商, 对 IPTV 业务也给予了足够的重视, 一直在探索如何让机顶盒的使用更便捷, 寻求提升机顶盒用户体验的方法。

目前用户在使用机顶盒时, 对视频节目的操作主要还是依靠红外遥控器。但是 IPTV 提供的内容非常丰富, 直播频道超过 200 个, 点播内容超过 50 000 h, EPG 页面层级过深, 使得用户在搜索内容时并不容易。随着智能语音控制技术的发展, 中国电信也探索过使用蓝牙语音遥控器的方式进行语音搜索, 但是经过验证发现, 近场语音可以提升用户体验, 但还不是交互体验最佳的, 例如, 需要凑近遥控器 MIC 说话, 说话时需要一直按住语音键; 蓝牙语音遥控器需要经常充电, 当蓝牙语音遥控器电量不足时也会影响语音识别准确率。因此, 如何让机顶盒听得更远、更清楚、更明白是中国电信一直在考虑的问题。从亚马逊的 Echo 到 2017 年初的拉斯维加斯 CES 展可以发现, 远场拾音产品能彻底解放双手, 成为最佳的人机交互方式。

因此有必要开发一款具备远场拾音功能的智能语音机顶盒, 用自然语言控制智能设备, 打破距离瓶颈, 解放双手操作, 提升用户体验。同时结合智慧家居, 在智慧家庭的环境下, 使用智能语音机顶盒可以设计出以下业务。

(1) 融合通信类业务

- 用户通过智能语音机顶盒查询实时交通状况: 用户在出门前, 可通过语音查询交通信息, 进入路况查询, 了解自己将要通行

的路况。例如, 如果常走的路线拥堵, 则改成另外稍微长些的路线, 但是路况畅通, 更能节省时间。

- 用户通过智能语音机顶盒查询实时天气状况: 用户在出门前, 可通过语音查询天气信息, 了解当天的温度、风力、降雨概率等信息。例如, 降雨概率大于 70%, 提醒用户携带雨具; 天冷适当加衣。
- 用户通过智能语音机顶盒实现家人之间快速方便的视频电话需求: 用户只需对智能语音机顶盒说“我要跟谁视频通话”。机顶盒就能打开电视, 调出视频通话客户端, 在电话簿中选择对方号码, 呼出视频通话。

(2) 高清视音频娱乐类业务

各类音/视频节目, 能轻松获取。用户想看“成龙的电影”, 只需说出“成龙的电影”, 智能语音机顶盒就能够迅速搜索到“成龙主演的相关电影”列表供用户选择观看; 想要切换电视频道, 也可以直接说“我要看东方卫视”“番茄台”等说法即可; 错过了之前的节目想要回看, 现在也变得很简单, 如 9 月 20 日的《新闻联播》, 用户可以直接说“我要看 9 月 20 日的新闻联播”。

(3) 生活提醒类业务

用户需要早起赶火车, 只需说出“给我设一个明天早晨 6 点的闹钟”, 智能语音机顶盒就能与用户确认闹钟定时, 并在第二天 6 点开启闹钟功能; 用户也可设置多个提醒, 例如, 交水、电、煤等生活账单的日期、还信用卡的日期等。

(4) 智能家居控制类业务

用户可以语音控制电视机的开启和关闭, 语音及时开启电灯、窗帘等智能家居设备或语音设置开启时间或开启条件。

为了实现以上 4 类业务, 作为智慧家庭新入口的智能语音机顶盒必须要支持以下功能。

(1) 远场拾音

智能语音机顶盒通过 HDMI 线与电视机连



接，放置在电视机下方，用户可在客厅 5 m 距离内，各角度通过语音对智能语音机顶盒发送语音指令。用户语音唤醒机顶盒时，机顶盒能自动对当前视频做静音处理，等待用户语音指令，语音反馈信息通过语音机顶盒本身携带的音箱进行发音输出。当用户语音指令完毕后，机顶盒能自动对当前视频做恢复声音处理。当用户关闭电视机后，智能语音机顶盒可作为智能音箱独立使用。

（2）即唤即用

用户开通智慧家庭语音产品业务后，将智慧家庭智能语音机顶盒接入家庭网络环境、连接电视机后，无需通过语音遥控器，在距离智慧家庭语音产品 5 m 之内，即可通过语音控制智慧家庭语音产品观看 IPTV 业务、视频通话、智能家居控制等，享受智慧家庭业务。

智慧家庭语音产品的即唤即用需对智慧家庭智能语音机顶盒进行语音唤醒，语音唤醒词具体说明如下。

- 唤醒词需支持 2~4 个汉字。
- 语音唤醒词建议语音覆盖尽量多，长度最少为两个音节。唤醒词要规避相邻音节，字要发音清晰、响度大，末字节发音要避免闭口音。例如，天翼天翼、小天小天。
- 语音唤醒词要避免与家庭成员小名重合，减少误唤醒。
- 允许用户自定义个性化唤醒词。
- 智能语音机顶盒收到用户的唤醒词指令后，即从休眠状态被唤醒，并给出指定响应。

（3）连续交互

连续交互是指用户语音唤醒智能语音机顶盒后，可以连续多次与智能语音机顶盒进行语音交互，无需再携带唤醒词。前后语音交互超过规定时间需对智慧家庭语音产品进行再次唤醒。

（4）娱乐及控制中心

智能语音机顶盒中预制的语音交互客户端需

完成 IPTV 业务平台、中间件等组件的适配和对接，以满足机顶盒智能语音交互业务功能，包括以下功能。

- 电视直播：通过语音进行直播频道切换，可通过频道名、频道别名、“上/下一频道”、频道号、特殊频道叫法等切换直播频道，支持甄别不存在的频道并给出友好的反馈，支持语音控制视频播放的进度和播放时间点等。
- 电视回看：通过说出频道名称、时间、片名等内容，完成相应频道节目单查询及内容的播放。支持语音控制视频播放的进度、支持回看最近一期及往期的节目、支持查看具体频道的节目单并能够直接定位到具体时间、支持语音控制视频播放的播放时间点。
- 点播内容搜索及播放：通过说出影片片名、类型、演员、年代、地区、频道及其组合等内容，完成影视资源搜索，并可通过语音操控其播放。
- 页面导航：通过语音进行页面导航，支持语音输入栏目名称可跳转至相应的栏目。
- 系统控制：通过语音进行播放控制、音量调节。
- 播放控制：通过语音进行播放暂停、快进快退、重新播放等控制。

随着智慧家庭需求的改变，智能语音机顶盒的能力也将不断地发展和完善。整体而言，智能语音机顶盒作为智能家庭新入口，已经准备好开启智慧家庭崭新的一天。

3 智能语音机顶盒的软硬件实现方案

对于电信运营商而言，智能语音机顶盒将是智慧家庭的新入口，是实现所有智慧家庭业务能力的控制终端，为了满足智慧家庭用户的各类业务需求，现提出智能语音机顶盒的软硬件实现方案。

3.1 硬件方案

普通智能机顶盒基本硬件系统主要包括核心控制单元、媒体处理单元、网络通信接口、媒体播放接口、控制接口及各类外设和扩展接口。普通智能机顶盒主要实现以下基本功能：媒体处理单元实现音/视频解码和基本图形显示，网络通信接口一般为有线网络接口，媒体播放接口为音/视频输出接口，控制接口为红外接口。为适应智慧家庭业务需求，智能语音机顶盒在音频输入、语音处理等方面都有所增加。新的实现方案示意如图1所示。

新增加的硬件模块包括：拾音设备、语音处理模块、全频段扬声器和摄像头。其中，拾音设

备建议是4 MIC及以上麦克风阵列，可以是环型或者线型；支持5 m的拾音距离、360°声源定位和全双工交互。

当机顶盒的CPU没有集成多路I2S时，需要增加语音处理模块。语音处理模块的主要功能是将麦克风阵列收到的多路语音流，经过语音处理模块汇聚后重采样成一路语音流传给机顶盒的CPU。要求麦克风信号和I2S信号时延小于9 ms，支持动态调整语音信号时延，满足回声抵消的要求。语音处理模块示意如图2所示。

如果智能语音机顶盒的CPU支持多路I2S，则无需额外增加语音处理模块。

具备一个全频段扬声器，要求低失真、音质

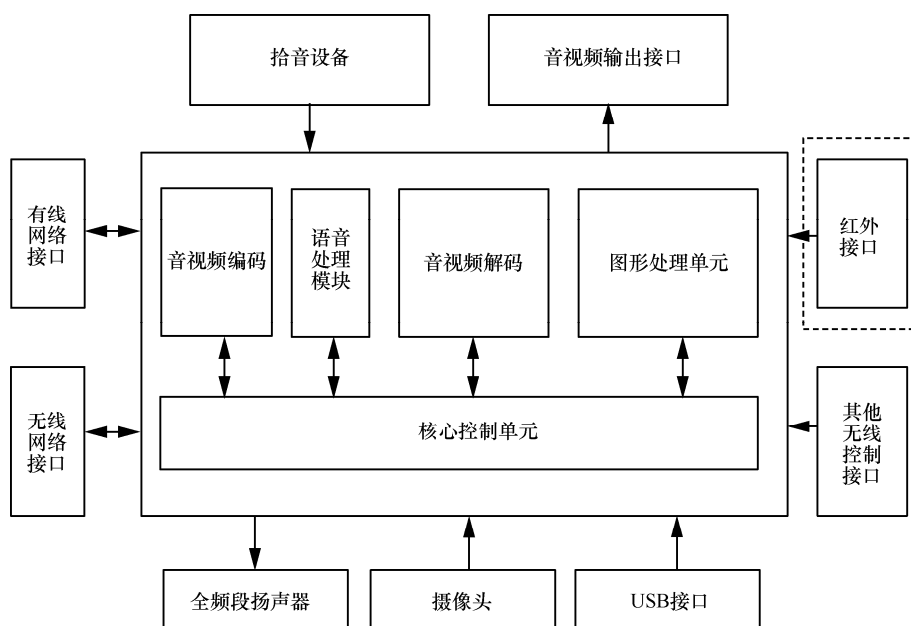


图1 智能语音机顶盒硬件结构示意图

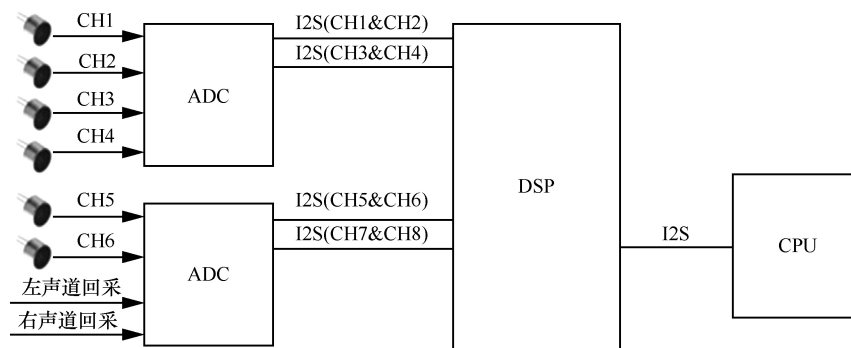


图2 语音处理模块示意



清晰。即使用户在不开电视机时，智能语音机顶盒也可通过自带的全频段扬声器进行发声，与用户进行语音交互。

摄像头要求图像传感器大于或等于 1/4 英寸，图像有效范围为 0.3~8 m，水平视角范围不小于 50°，视频分辨率达到 HD 1080P (1 920 dpi×1 080 dpi) 以上，视频帧率达到 30 f/s 以上，图像迟滞不超过 0.2 s，最低照度小于或等于 30 Lux，信噪比大于或等于 39 dB (max)，接口为 USB 2.0 以上，音视频编码支持 MJPEG 视频编码格式和 PCM 音频编码格式。

其他基本硬件的变化：目前主流机顶盒的 CPU 主频为四核 1.5 GHz、内存为 1 GB，在智慧家庭智能语音机顶盒上建议硬件配置应至少达到 CPU 主频四核 2.0 GHz，内存 2 GB 以上。

媒体处理的变化：原有的媒体处理主要指音视频解码、普通的图形处理功能，针对智能语音机顶盒的视频通信和支持超高清视频业务的开展，智能语音机顶盒必须支持 4KP60 解码能力；支持至少两路高清的多视频画面，支持 PIP 功能。

网络传输接口的变化：由于家庭房间大小和布局结构的变化，智能语音机顶盒需要支持无线网络接口，例如，支持 IEEE 802.11ac/b/g/n 2.4 GHz/5 GHz 双频、天线 2×2 的 MIMO。

控制接口的变化：用户可以通过语音直接控制语音机顶盒进行各项操作，因为红外接口基本

上可以省略；由于可同时支持近场语音，可适当增加蓝牙等无线接口方式。

操作设备的变化：为提升用户的操作体验，智能语音机顶盒的操作可以直接由人声远场控制，传统的遥控器可被语音替代。由于可同时支持近场语音，也可增加蓝牙语音遥控器，或者手机语音客户端进行辅助交互。

3.2 软件实现方案

作为智慧家庭新入口的智能语音机顶盒的软件整体架构基本不变，如图 3 所示，包括操作系统及硬件适配层、服务能力层、应用程序层，与普通智能机顶盒类似。

根据具体业务能力支持，智能语音机顶盒的软件能力也发生了相应变化。从操作系统到应用软件，智能语音机顶盒的能力和选择范围普遍得到了提升。

操作系统：普通智能机顶盒采用 Android4.4 版本，在智能语音机顶盒上建议 Android7.0 版本，可实现真正的多窗口、多任务，例如，在观看视频的同时进行视频通话，可同时多个直播频道同屏播放等。

服务能力：为支持智能语音业务的开展，机顶盒的服务能力也将相应提升。传统机顶盒基本支持媒体播放、浏览器等基本服务能力；面对智慧家庭业务，还需支持语音引擎、TTS 引擎等服务能力，

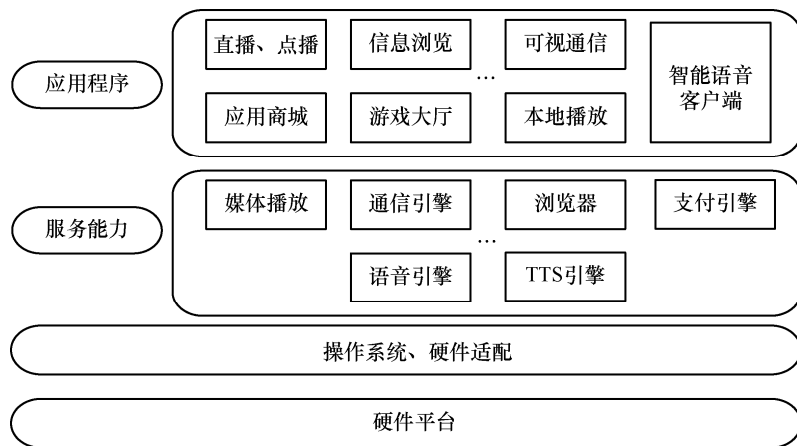


图3 智能语音机顶盒软件架构示意

通过语音算法完成噪声抑制、混响消除、回声抵消、语音模型匹配等处理,实现智能语音交互。

应用程序:智能语音机顶盒比普通智能机顶盒多了一款智能语音客户端,负责接收拾音设备传来的语音数据,调用IPTV语音识别服务完成语音识别,将用户语音操作指令转换为各IPTV业务组件可识别的程序指令。语音客户端架构如图4所示。

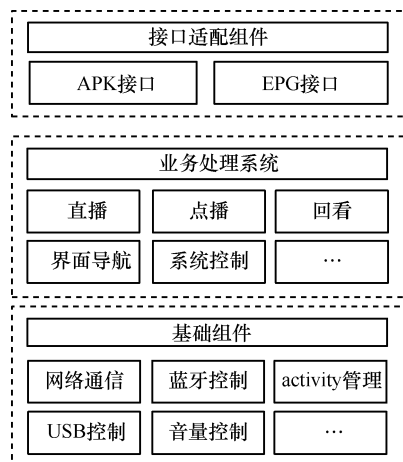


图4 智能语音客户端架构

根据上述智能语音机顶盒的软硬件实现方案,中国电信目前已完成智能语音机顶盒的原型开发。通过智能语音机顶盒顶部的环型麦克风阵列可进行远场拾音,语音识别成功率可达到85%以上。经过智能语音平台与IPTV平台的对接调试,已支持远场语音控制频道切换、语音搜索点播节目等功能;目前仍在与中国电信的多个增值业务平台对接,例如爱音乐、爱游戏、爱动漫、天翼阅读、天翼视讯等,争取实现更多电信增值业务的语音控制。智能语音机顶盒的原型如图5所示。

4 结束语

智能语音机顶盒被称作机顶盒,但它不仅仅是机顶盒。智能语音机顶盒在提供视频观看业务的同时,更被赋予了“听”和“说”的能力。智能语音机顶盒是智能机顶盒形态的扩展,是对现有智能机顶盒形态的补充,不会完全取代目前的机顶盒。针对不同电信套餐的用户,智能语音机

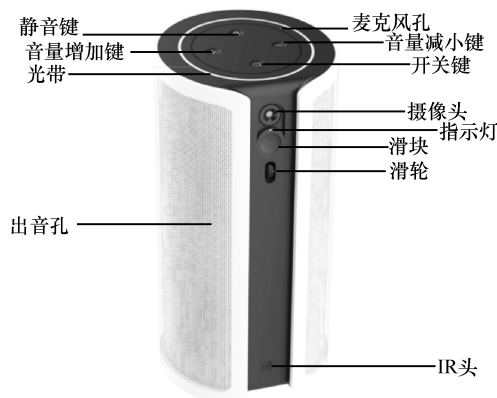


图5 智能语音机顶盒的原型

智能顶盒根据产品形态和主要功能可以多样化发展,可分为简洁型(具备智能机顶盒、麦克风阵列及中频段音箱的产品)、基本型(具备智能机顶盒、麦克风阵列及全频段音箱的产品)和增强型(具备智能机顶盒、麦克风阵列、摄像头及全频段音箱的产品)。

在家居环境中,自然语言是比手机触摸屏、红外遥控器更方便、更自然的交互方式。语音交互将使人机交互以人类最天然、熟悉的方式进行,这是未来智能生活的必经之路。依托于语音技术的智能语音机顶盒,将作为智慧家庭新入口,成为智慧家庭的新终端。

参考文献:

- [1] 杨震,徐敏捷,刘璋峰,等.语音大数据信息处理架构及关键技术研究[J].电信科学,2013,29(11):1-5.
YANG Z, XU M J, LIU Z F, et al. Study of audio frequency big data processing architecture and key technology[J]. Telecommunications Science, 2013, 29(11): 1-5.
- [2] 王琳,肖晴,马宇飞.一种实现智能电视语音搜索的方案[J].电信科学,2014,30(5):40-45.
WANG L, XIAO Q, MA Y F. A total solution of voice search on smart STB[J]. Telecommunications Science, 2014, 30(5): 40-45.
- [3] HERMAN SKY H, MORGAN N. RASTA processing of speech[J]. IEEE Transactions on Speech & Audio Processing, 1994, 2(4): 578-589.
- [4] 张敏,宋杰,刘晓峰.电信运营商面对OTT的战略选择[J].电信科学,2014,30(2):142-146,151.
ZHANG M, SONG J, LIU X F. Strategic selection of telecom operators to counter OTT[J]. Telecommunications Science, 2014, 30(2): 142-146, 151.

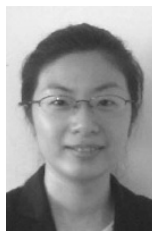


- [5] 施唯佳, 蒋力, 贾立鼎. OTT TV 和 IPTV 的技术比较分析[J]. 电信科学, 2014, 30(5): 15-19, 26.
SHI W J, JIANG L, JIA L D. Technique comparative analysis of OTT TV and IPTV[J]. Telecommunications Science, 2014, 30(5): 15-19, 26.

[作者简介]



施唯佳（1978-），男，中国电信股份有限公司上海研究院智慧家庭专业领域一级专家、高级产品经理，主要研究方向为与智慧家庭、IPTV、OTT TV 等方面相关的技术及产品研发。



王琳（1979-），女，中国电信股份有限公司上海研究院产品开发工程师，主要研究方向为 IPTV、终端技术。



贾立鼎（1970-），女，中国电信股份有限公司上海研究院产品开发高级工程师，主要研究方向为 IPTV、多媒体终端技术。