

# 智能音箱在电信运营商业务的 应用研究

## Research on the Application of Smart Speakers in Telecom Business

李奥<sup>1</sup>,梁博<sup>1</sup>,苏雨聃<sup>1</sup>,刘含宇<sup>2</sup>(1. 中国联通网络技术研究院,北京 100048;2. 中国联通系统集成有限公司,北京 100032)

Li Ao<sup>1</sup>,Liang Bo<sup>1</sup>,Su Yudan<sup>1</sup>,Liu Hanyu<sup>2</sup>(1. China Unicom Network Technology Research Institute,Beijing 100048,China;2. China Unicom System Integration Co.,Ltd.,Beijing 100032,China)

### 摘要:

近年来,智能家居行业发展迅速,尤其以人工智能和多场景应用为主线的发展最为明显。行业内普遍认为AI入口之争将会成为未来的竞争焦点。作为智能语音技术的应用形式之一,智能音箱承载着智能家居交互入口的观点被一再提及。电信运营商正积极研究智能音箱相应业务,增强产业竞争力与服务能力。分析了智能音箱的现状与发展趋势,分析智能音箱在电信运营商的应用实践,提出电信运营商在智能语音市场发展的策略建议。

### Abstract:

In recent years, the smart home industry has developed rapidly, especially the development of artificial intelligence and multi-scene applications. It is widely believed in the industry that the AI entry dispute will become the focus of future competition. As one of the application forms of intelligent voice technology, the idea that smart speakers carry smart home interactive portals is repeatedly mentioned. Telecom operators are actively studying the corresponding business of smart speakers, enhancing industry competitiveness and service capability. The current situation and development trend of smart speakers is analyzed, as well as the application of smart speakers in telecom operators, and development strategies for telecom operators in the intelligent voice market.

### Keywords:

Artificial intelligence; Smart speaker; Speech recognition; Natural language processing; Speech synthesis

引用格式:李奥,梁博,苏雨聃,等. 智能音箱在电信运营商业务的应用研究[J]. 邮电设计技术,2019(2):52-56.

## 0 引言

随着人工智能相关技术飞速发展以及人们对生活品质要求不断提高,人工智能技术逐渐被引进智能家居产品中。语言是人类最重要最便捷的交流方式。通过语音交互技术实现智能音箱对智能家居产品的控制,是目前智能家居交互入口的目标。本文从技术和产业层面分析智能音箱的发展现状及趋势,研究智能音箱在电信运营商业务中的难点、对应解决方案和

实际应用部署情况,展望电信运营商在智能音箱产业的发展策略。

## 1 智能音箱发展现状及趋势

### 1.1 智能音箱概述

智能音箱是一种由人工智能与传统音箱结合的产品,拥有了解天气、打电话、点歌等基础功能。借助稳定的网络连接和良好的收音效果唤醒音箱,最终实现人机对话。目前市场上智能音箱的核心功能有4个,如表1所示。

随着家居控制进入信息化、智能化时代,智能音

收稿日期:2019-01-10

表1 智能音箱的核心功能及其具体内容

核心功能	具体内容
语音交互	包括语音识别、语音理解等技术,保证音箱快速识别用户需求,提供更为自然的人机交互模式
内容服务	提供丰富的音乐、有声读物等流媒体内容,保证用户的音频内容享受
互联网服务	涵盖外卖、信息查询、出行、商务等多种服务,接入家居中多样生活场景
智能家居控制	通过智能音箱实现对照明、安防产品等其他智能家居产品的控制,并能场景化地定义操控指令,使智能音箱成为智能家居的控制中心

箱逐渐成为市场热点,其作为智能家居的入口,也将成为电信运营商们竞争的焦点。

### 1.2 智能音箱核心技术

智能音箱拥有从听(语音识别)到理解思考(语义理解)再到说(语音合成)的完整闭环能力。基本的智能音箱语音交互流程如图1所示。

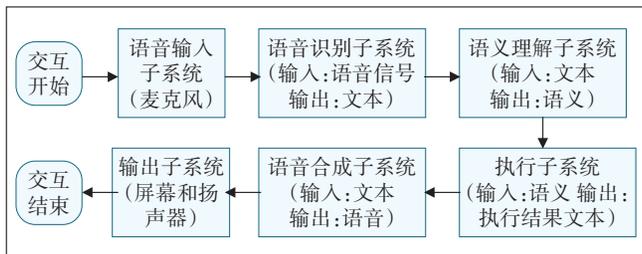


图1 智能音箱语音交互流程图<sup>[9]</sup>

#### 1.2.1 语音识别技术

语音识别的原理图如图2所示。

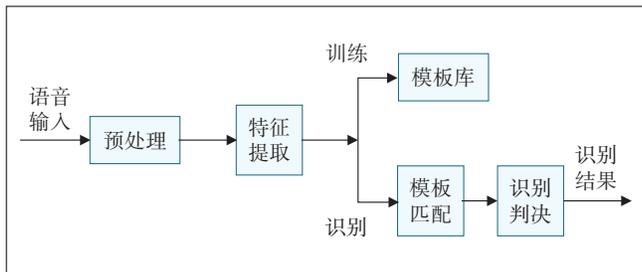


图2 语音识别原理图<sup>[6]</sup>

预处理模块:包含预滤波、采样和量化、分帧、加窗、预加重、端点检测等。其中,端点检测(VAD)是语音唤醒的重要环节,它标记语音信号的开始和结束,准确性直接影响语音识别的效果。智能音箱多采用基于隐马尔可夫模型的语音唤醒模型。隐马尔可夫模型(HMM)解决的3个问题如表2所示。

特征提取模块:提取语音信号中反映语音特征的若干参数。常用的有线性预测系数(LPC)、线性预测倒谱系数(LPCC)、MEL频率倒谱系数(MFCC)。LPCC

表2 隐马尔可夫模型要解决的3个问题及其解决方法和应用

问题	对应算法	应用
给定一串观察符号序列,计算HMM生成该序列的概率	前向算法	语音唤醒,通过计算一段语音信号序列对应的生成概率,可以根据概率大小判断是否需要唤醒
给定一串观察符号序列,找出观测值最可能从属的HMM状态序列	维特比算法	语音识别,通过对各个单词建模,可以将语音序列转换为最可能的单词序列——语句
给定一个观察符号序列,计算出HMM的参数	Baum-Welch算法	最大似然估计,通过模型训练来获取模型参数

从发音模型角度出发,运算复杂度低,但是高频干扰多、系统可靠性差;MFCC从听觉模型出发,与LPCC相比,MFCC有更好的系统识别性能。

训练阶段:用户输入训练的语音,经过预处理及特征提取后,每个词条取得一个特征矢量参数,并作为模板保存,建立特征模板库,参与训练的人越多,识别的鲁棒性越好。

识别阶段:将输入的语音信号经过特征矢量参数提取后,生成测试模板,与参考模板进行相似性比较,并根据一定的搜索和匹配策略找出一系列最优的、与输入语音信号匹配的模板,最后查表得出识别结果。

目前语音识别最常用的算法有:动态时间规整(DTW)算法(解决说话速度不均匀造成的识别困难)、矢量量化(VQ)、隐马尔可夫模型(HMM)法(研究热点)和人工神经网络(ANN)法(语音识别常用的神经网络有反向传播(BP)神经网络、径向基函数网络(RGF)神经网络等,虽然可以处理一些环境复杂、知识背景不清楚、推理原则不明确的问题,但是不能很好地描述语音信号的时间动态性,所以多与传统识别方法结合使用)。

#### 1.2.2 语义理解技术

语义理解就是将自然语言文本转化为用户的意图。基于深度学习的自然语言处理(NLP)技术,可支持多个垂直领域的精准语义解析,可以进行新领域的快速扩展。建立基于感知机的中文分词、词性标注和命名实体识别框架,可以实现分词、词性标注和命名实体识别,还可以进行“在线学习”(或“增量训练”)。最后将识别结果匹配到具体的操作指令中。

#### 1.2.3 语音合成技术

语音合成技术使智能音箱能根据文本内容发出人类的声音。该技术涵盖了语言学、声学、信号数字化处理等多种技术,具体流程如图3。

语音合成的步骤如下:

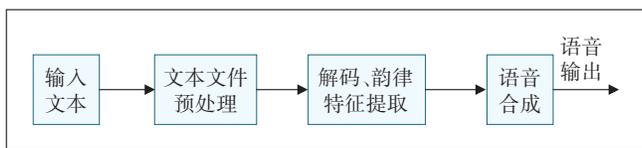


图3 语音合成流程图<sup>[10]</sup>

a) 文本整理、词汇切分、语法意义分析,使智能音箱对文本信息完全理解。

b) 解码和韵律特征提取,为语音合成设计和规划音频特征(音长、音强、音调等)。

最后合成的语音可以正确表达语义,得到音频输出。

### 1.3 智能音箱产业

智能语音技术刚刚落地,还需要获取大量用户的数据资源,通过用户数据不断学习训练,智能音箱才能更加智能。市场占有率和销量会影响到智能音箱的体验,也是智能音箱产品的重要影响因素。全球智能音箱的市场份额见图4。

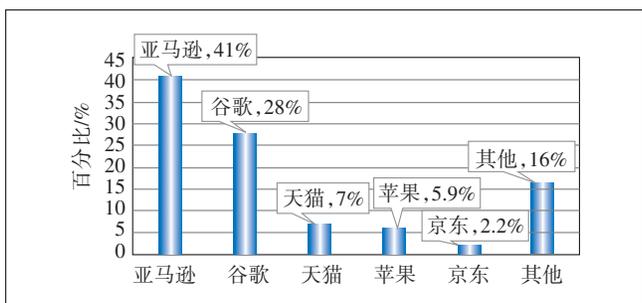


图4 全球智能音箱的市场份额

如图4所示,2014年11月首发的亚马逊智能音箱(Echo)凭借其先发优势和丰富的产品,占据了市场41%的份额;2016年5月首发的谷歌智能音箱(google home)通过其低、中、高产品矩阵和用户基础,占据了市场28%的份额,2017年7月首发的阿里巴巴智能音箱(天猫精灵)占据7%的份额,2018年2月首发的苹果智能音箱(HomePod)占据5.9%的份额,2015年5月首发的京东智能音箱(叮咚)占据2.2%的份额,其他品牌共占据16%的份额。

### 1.4 智能音箱发展趋势分析

#### 1.4.1 技术发展趋势

语音识别系统已进入商品化阶段,但是机器与人自然交流的终极目标目前仍未实现。需要继续研究解决的技术难点见表3。

语义理解技术是自然语言理解领域研究的热点,目前需要继续研究解决的技术难点见表4。

表3 语音识别技术难点及其对应解决方案

难点	解决方案
连续语音中音素、音节或单词之间调音结合引起的音变使基元模型之间的边界不明确	进一步完善声学模型和语言模型
语法语义规则的建立与理解困难	设计自然语音数据库,深入研究人与人之间语音对话和自然口语
方言、口音对语音识别的正确率影响严重	开展多语种混合识别,包括方言和外语
单个字母及单个词发音受上下文环境的影响,音调、重音、音素等都可能存在差异	通过抗噪声和去口音技术提高语音识别的鲁棒性
外界环境噪声和干扰	
嵌入式系统可利用资源少,而语音识别对计算能力要求高	利用云端识别计算能力强、识别正确率高的优势,同时结合离线识别不受网络状况影响的特点来提高用户体验

表4 语义理解技术难点及其对应解决方案

难点	解决方案
单词的边界定义;口语中的词间是连续的,需选择最符合上下文且语法无误的单词组合	建造词汇知识库,使用概率的方法,在概率上下文无关语法中引入词汇信息,采用词汇中心概率表示法,降低“词汇主义”倾向影响
词义的消歧	建立大规模语料库,运用更多的语言学知识解决
句法的模糊性	
有瑕疵的或不规范的输入:外国口音或地方口音,文本处理中拼写、语法、光学字符(OCR)的错误	运用统计学方法,同时结合机器学习的方法来获取语言知识
语言行为与计划:句子存在字面以上的含义	浅层语义理解与深层语义理解并重,统计与规则方法并重,形成混合式的系统

语音合成技术未来还需要提升自然度(从句子到篇章这一级的自然度),丰富合成语音的表现力(增加年龄特征、性别特征及语气语速特征);降低语音合成技术的复杂度,减小音库容量;增加多语种语音合成,容纳不同地域的方言。

#### 1.4.2 产业发展趋势

智能音箱是一个潜在而又庞大的市场,也是家庭娱乐的延伸领域。国外市场中如亚马逊、谷歌等互联网企业多采用“走量”策略,通过大量设备抢占市场份额。未来智能音箱市场不但重视硬件销量,还重视挖掘新盈利模式,如口播广告、内容订阅服务(音乐订阅、高级会员服务和企业解决方案)等。相比2018年美国智能音箱市场亚马逊、谷歌和苹果“三足鼎立”的态势,国内市场则更复杂,京东、阿里都有相应产品,在2017年底,百度、腾讯也参与进来。

国内智能音箱产业在未来仍需要深入研究解决以下问题。

a) 如何提高智能家居普及率以促进智能音箱在家居控制中的循环发展。

b) 中西方生活方式差异对智能音箱的使用率有何影响,如国外的应用场景主要在客厅和厨房,国内则是客厅和卧室。

c) 厂商在开拓语音交互技术过程中如何提升用户体验。

d) 智能音箱可能会收集越来越多的敏感数据,如通话记录、电子邮件信息、生活习惯等,如何保障用户隐私安全也是问题之一。

e) 智能等级不清晰,外界不能以量化的方式合理评估智能音箱能力。

## 2 电信运营商智能音箱应用研究

### 2.1 电信运营商使用智能音箱发展业务的难点

国内电信运营商正在寻找新的收入增长领域,目前,市场上主流智能家居产品均需联网,对于擅长做“管道”的运营商而言,进入智能家居的门槛并不高,特别是对于技术、硬件和内容整合都不占优势的国内电信运营商而言,聚焦智能音箱以打通智能家居交互入口是一个好的选择,但是目前仍面临以下难点,其中难点d)和e)为国内电信运营商特有,其他为国内外运营商共有。

a) 价格影响。互联网公司相对更早进入智能音箱市场,拥有更成熟的技术、更丰富的配套智能家居产品,多采用低价“走量”策略。在低价策略下,电信运营商很难与互联网公司推出的智能音箱抢占市场,难以形成一定的规模,降低成本难度较大。

b) 盈利策略。电信运营商的盈利模式是通过智能音箱控制智能家居产品以实现流量变现,这促使电信运营商将智能音箱与IPTV类产品进行融合。但是音箱类产品完全不同于传统电视视频,其仍需要大量数据建模分析及运营优化工作来满足广泛的用户需求,以此增强用户黏性,进而达到开拓用户、汇聚流量的目的,最终实现流量变现,为电信运营商带来更大的盈利。

c) 互联网生态。与互联网公司相比,电信运营商缺乏丰富的生态内容和强大的消费硬件物联网。

d) 捆绑补贴政策。国内电信运营商对智能音箱的销售主要采取与宽带业务捆绑补贴的政策,保证用户能以更低的价格买到高端音箱。补贴有助于国内电信运营商的智能音箱普及,但是如果不能快速带来

相应的增量收入,则会陷入成本泥潭。

e) “运营”策略。如何平衡国内电信运营商和智能音箱硬件厂商的利益仍需要考虑,这关乎到双方的盈利策略。国内电信运营商不愿仅仅替硬件厂商销售设备,变成智能音箱硬件厂商的零售渠道商,而智能音箱硬件厂商更愿意把产品拿到互联网公司的电商平台上销售。

f) 标准规范。各电信运营商都在做自己的智能音箱规范,但是目前并没有一个统一的规范,而各电信运营商自己的规范可能不全面,或多或少地缺乏完善的智能等级评测、安全条例、防广告规范。

g) 技术水平。相较于互联网企业,擅长做“管道”的电信运营商的优势在于其拥有成熟的通信网络技术,劣势在于其涉足人工智能相关技术较晚且技术能力较弱,即便拥有自研的人工智能平台,运营商在语音识别、自然语言处理等模型构建、训练方面仍然与互联网企业存在一定差距。

### 2.2 电信运营商使用智能音箱发展业务的建议方案

根据上文提及的电信运营商难点,提出了如下的解决方案。

为了避免陷入成本泥潭,国内电信运营商可以加大运营力度,提升流量变现能力;控制智能设备的补贴水平(如对高值用户和战略业务可以适当加大补贴力度),这既体现了经营的向导性,也是一种客户经营的差异化手段,同时也会提升用户的感知度,尽量避免赠送后用户感知度低、使用率低的局面。

为了缩小与互联网企业的价格差距,同时扩展规模,电信运营商可以采用与光猫、机顶盒类似的手段,统一标准,一起降低成本。

国内电信运营商可以向德国电信学习,与智能音箱硬件厂商合作,通过类似于收取商场或者商家租金的模式,从设备商的销售收入中获得分成。

电信运营商可以充分利用产业链资源,提出相关的标准规范。运营商可以利用自己的品牌优势,作为安全保障者或第三方机构为智能音箱产品做测评、评级、认证,与互联网企业联合起草评估规范(评估规范包括语音识别、语音交互、语音合成、安全服务等测试),共同完成对音箱整体智能化能力的评估,为用户的选择提供量化依据。

电信运营商可以与互联网企业合作,利用互联网企业丰富的技术资源、生态内容等,共同打造一款智能音箱,并集全国之力进行线上线下的销售,保持一

致性、稳定性的销售政策,从而可以汇聚流量,实现流量变现,达到互利共赢的目标。

电信运营商需要发掘智能音箱与传统IPTV业务在内容处理和推荐业务逻辑上的差异,短期内对用户使用数据进行深度分析,并依次优化内容运营工作,从而尽可能满足更多用户的需求,为电信运营商汇聚更多流量,实现流量变现,带来更大的盈利。

### 2.3 电信运营商基于智能音箱的应用

目前国内外电信运营商在智能音箱产业上均部署了相应的应用。

中国联通自主研发的智能音箱可以由语音控制IPTV机顶盒实现视频点播、频道切换、系统控制等功能,智能家居的其他智能设备控制、VoWiFi等能力将被进一步整合融入,从而打造中国联通智慧家庭智能语音生态。同时中国联通也在积极参加智能音箱评测标准的制定工作。

中国电信与科大讯飞合作,运用科大讯飞语音识别技术研发出智能音箱小翼。这款音箱是基于中国电信光纤宽带网络能力,可进行语音操控的智能化终端。该音箱集4K机顶盒、高清摄像头和蓝牙音箱于一体。

中国移动自主研发的智能AI音箱(咪咕Home音箱)具备独特的多轮语音操控功能。另外音箱还与移动通信业务相结合,通过Andlink快速连接智能家居,成为智能家居的控制入口。中国移动也在参加智能音箱评测标准的制定工作。

德国电信发布了一款具备语音助理功能的智能音箱Magenta。该智能音箱的定位是连接智能家庭、电视和其他在线服务的枢纽。该智能音箱可以支持EntertainTV及Magenta SmartHome智能家居服务,该音箱还能与亚马逊的智能语音助理Alexa服务互联。

## 3 电信运营商智能音箱业务发展策略

电信运营商可以通过加大运营力度来提升流量变现的能力,控制智能设备的补贴水平和力度,提升客户感知度,进而避免陷入成本泥潭。另外,互联网公司的智能音箱多以线上销售为主,电信运营商的智能音箱销售服务政策需要保持一定的稳定性和一致性,集全国之力聚焦某几款重点销售,在形成规模后,进而降低成本,汇聚流量,实现流量变现。电信运营商可以建立标准规范,增强对智能音箱的量化评级力度,最终才能够真正实现“以智能音箱为统一入口,以

智能家居产品群为核心载体,以智慧化运营为运营逻辑和流量变现手段,以智能家居产品的合作伙伴群为生态圈”。

## 4 结束语

电信运营商应该抓住人工智能行业快速发展的契机,学习互联网企业的商业模式,大力推动以智能音箱为首的智能家居产品的应用,进而抢占智能家居市场,变现流量,提高收益。同时电信运营商也应该积极关注智能音箱的安全隐患,增大对系统安全的投入,并推动建立相关行业标准(如智能音箱接口、智能度测评等)。

### 参考文献:

- [1] 陈白杨. 基于语音交互的智能家居控制系统[D]. 兰州: 兰州大学, 2018.
- [2] 李剑平. 服务机器人自然语言理解与任务规划的研究及应用[D]. 广州: 广东工业大学, 2013.
- [3] 徐亮. 智能音箱: 电信运营商“仰望星空”也在“脚踏实地”[J]. 通信世界, 2018(20): 26-27
- [4] 陈志刚. 运营商如何在智能家居市场进行差异化竞争?[J]. 中国电信业, 2018(6): 64-65
- [5] 崔悦, 宋齐军. 智能语音技术发展趋势及电信运营商应用浅析[J]. 邮电设计技术, 2016(12): 6-11
- [6] 张凤美. 基于语音识别的移动机器人远程控制系统研究[D]. 武汉: 武汉科技大学, 2013.
- [7] 刘荣辉. 基于智能家居控制的嵌入式语音识别系统研究[D]. 广州: 广东工业大学, 2013.
- [8] 甘玉珏, 苏军根, 林健, 等. 智能语音产品现状及展望[J]. 广东通信技术, 2017(12): 66-68
- [9] 袁彬, 肖波, 侯玉华, 等. 移动智能终端语音交互技术现状及发展趋势[J]. 信息通信技术, 2014(2): 39-43
- [10] 张越. 智能语音交互系统的开发与应用[D]. 哈尔滨: 黑龙江大学, 2018.
- [11] 李忠楨, 唐立臣. 基于ZigBee的语音管控智能家居研究[J]. 数字通信世界, 2017(5): 93-94
- [12] WILSON C, HARGREAVES T, Hauxwell-Baldwin R. Benefits and risks of smart home technologies[J]. Energy Policy, 2017(103): 72-83.
- [13] MARIANA Z. 2018 Trends to Watch: AI Home Assistants [R]. Ovum, 2018.

### 作者简介:

李奥, 毕业于曼彻斯特大学, 工程师, 硕士, 主要从事人工智能在通信网络中的应用研究工作; 梁博, 工程师, 硕士, 主要从事视频技术及智能语音系统研究工作; 苏雨聃, 毕业于复旦大学, 工程师, 博士, 主要从事接入网质量优化、创新相关工作; 刘含宇, 毕业于兰卡斯特大学, 工程师, 硕士, 主要从事物联网应用研究工作。