



一种实现智能电视语音搜索的方案

王琳¹, 肖晴², 马宇飞¹

(1. 中国电信股份有限公司上海研究院 上海 200122;

2. 中国电信股份有限公司上海分公司 上海 200120)

摘要:从“三网融合”新政下最新的业务需求出发,对智能语音搜索业务进行需求分析,并且与同类产品进行比较,之后提出了一种满足 IPTV 智能电视语音搜索业务功能要求的端到端实现方案,从智能电视语音搜索提供的业务、平台功能要求、系统架构等方面对方案进行了阐述和分析,该方案可以使用户更方便地搜索 IPTV 上的视频内容。

关键词:智能机顶盒;IPTV;语音搜索

doi: 10.3969/j.issn.1000-0801.2014.05.006

A Total Solution of Voice Search on Smart STB

Wang Lin¹, Xiao Qing², Ma Yufei¹

(1.Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China;

2.Shanghai Branch of China Telecom Co., Ltd., Shanghai 200120, China)

Abstract: Based on the new business requirements in network convergence, the demand of intelligent voice search service was analyzed. And a systematic solution of voice search on smart STB was presented, including functionality analysis, and comparison with similar products. From voice search service smart TV offered, platform functional requirements, system architecture and other aspects, the solution was described and analyzed, which could make video content search more convenient.

Key words: smart STB, IPTV, voice search

1 引言

随着“三网融合”新政下业务的不断发展,2012 年整个行业已经进入智能化时代,IPTV 正经受着来自互联网电视、移动互联网等多个方面的考验。为了保持在视讯领域的技术领先性,IPTV 业务需要引入互联网中具有用户影响力的前沿技术。人机交互技术,特别是智能语音控制技术,是目前在智能电视、互联网电视应用的新兴技术,该技术的应用不但提升了客户体验,而且有望助力电视界面成为新的用户信息获取界面。苹果、谷歌、微软等业界公司都积极地在相关的产品中引入智能语音控制技术。

为了适应业务及技术的发展潮流,进一步提升 IPTV

业务体验,需要通过语音控制技术为用户提供更好的人机交互手段,实现对传统遥控习惯的改变。本文对比了现有智能电视语音搜索方案,针对现状及需要,提出了一种新的智能电视语音搜索方案,并从智能电视语音搜索提供的业务、平台功能要求、系统架构等方面对方案进行了阐述和分析。

2 智能电视语音搜索业务的使用方式及需求

智能电视语音搜索业务是指用户在收看电视时,可以利用语音遥控器或手机客户端中安装的语音控制产品,通过标准普通话对想要收看的节目进行搜索,完成点播影

片、频道、节目等操作,使用自然语言理解控制代替遥控器控制,是视频业务的一种新业务。

已有越来越多的智能电视开发语音搜索业务,如联想、TCL、长虹、海信、三星智能电视以及 XBOX one、Dragon TV 等。智能电视语音搜索方案的构成主要包括两个方面:终端侧对语音的采集和展现、语音平台侧对语音的识别和理解。

终端侧对语音的采集和展现主要包括语音搜索的启动和结束、语音搜索的交互界面、语音搜索结果的显示等,以下将从这几个方面对已有智能电视语音产品的语音使用方式进行比较分析。

(1)语音搜索的启动与结束

智能电视语音搜索产品的语音开启与关闭方式主要分为以下 3 种:

- 通过长按遥控器上的语音按键开启语音功能,如联想智能电视;
- 通过按一下遥控器上的语音按键开启语音功能,如 TCL、长虹、海信智能电视;
- 通过纯语音方式开启语音功能,如 XBOX one、Dragon TV、三星智能电视。

3 种语音开启方式的优缺点比较见表 1。

(2)语音开启页面

语音搜索功能开启后,智能电视产品会出现一个语音功能启动页面,启动页面的展现方式主要有以下 4 种:

- 全屏显示,如 TCL 产品;

- 覆盖屏幕中间显示,如海信、长虹、康佳产品;
- 覆盖屏幕上方显示,如联想产品;
- 覆盖屏幕下方显示,如 Dragon TV、创维、三星、XBOX one、Google TV 等产品。

4 种语音开启页面显示方式的优缺点比较见表 2。

在智能电视产品中,语音功能启动时部分产品会出现卡通形象与用户进行模拟对话,如创维、康佳、长虹、海信、TCL 等产品。有无动画形象的语音页面显示的优缺点比较见表 3。

(3)语音录入后是否确认执行

智能电视语音搜索产品在语音录入结束后,主要采用以下 3 种方式启动语音搜索的执行功能:

- 语音录入结束后直接执行搜索,如 Google TV、Dragon TV、XBOX one、三星产品;
- 语音录入结束等待数秒后自动执行搜索,如康佳产品;
- 语音录入结束后,需先确认识别结果再执行搜索,如 TCL、海信、联想、长虹产品。

以上 3 种操作反馈方式的优缺点比较见表 4。

(4)语音搜索结果的显示

智能电视语音搜索产品在语音搜索的结果显示方面,主要分为以下 2 类:

- 覆盖显示,如长虹、联想、Dragon TV、三星、海信产品;
- 全屏显示,如康佳、Google TV、TCL 产品。

2 种语音搜索结果显示方式的优缺点比较见表 5。

表 1 语音开启方式比较

操作方式	优点	缺点
长按语音键开启	可以清晰地判断语音输入的结束状态	操作复杂,易误操作,物理按键容易损坏
按键即可开启	操作相对方便	
纯语音方式开启	操作的自由度和连贯性较高	语音识别技术要求较高,易受外界环境影响

表 2 语音开启页面显示方式比较

页面显示方式	优点	缺点
全屏显示	可使用户聚焦到语音模式下	页面跳转过大,操作不流畅
覆盖屏幕中间显示	页面跳转不大,操作流畅	当前页面的部分信息容易被覆盖,播放时易覆盖主题信息
覆盖屏幕上方显示		当前页面的部分信息容易被覆盖,首页时易覆盖导航信息
覆盖屏幕下方显示		当前页面的部分信息容易被覆盖,播放时易覆盖字幕等信息

表 3 语音开启页面动画形象显示比较

动画形象显示与否	优点	缺点
显示	对于品牌推广有辅助作用	卡通形象无特色,页面显示相对繁杂
不显示	界面简洁,用户更容易聚焦搜索本身;没有明确的卡通形象,使用户有一定的想象空间	语音对话时场景的真实度可能会有所欠缺



表 4 语音搜索操作反馈方式比较

操作反馈方式	优点	缺点
直接执行	搜索的流程快速简单流畅	容易出现识别错误情况
等待数秒后自动执行	给予用户修改的机会,减少错误几率	等待时间变长
确认识别结果后执行	避免出现错误	操作步骤增加,搜索时间变长

表 5 语音搜索结果的显示方式比较

语音搜索结果显示方式	优点	缺点
覆盖显示	保证当前页面不被打断	界面显示比较乱
全屏显示	显示比较清晰	当前页面被中断

从上述比较可以看出,语音采集时操作要相对方便,界面展现要能保留原先的视频业务并使语音搜索的页面不大且操作流畅,同时应减少用户的交互步骤,能相对快速地定位到想要搜索的节目,从而更能吸引用户使用语音搜索业务。

3 IPTV 智能电视语音搜索方案的流程

本文设计了 IPTV 智能电视语音搜索产品的流程,如图 1 所示,具体介绍如下。

(1)按下语音键即可开启语音功能,从使用上来说,更方便快速,误操作的成本也比较低,误操作造成的影响基本可忽略。

(2)考虑到用户启动语音录音后,可能还要思考一下再输入想要观看的节目,故预留 10 s 的时间;如超过 10 s 用户还未录入语音,则提示未收到录音,并退出语音录音界面。

(3)语音录音时,启动端点检测功能,对用户的语音进行分析,若检测到语音停止,则直接启动语音识别及搜索功能,可以减少用户的操作步骤,更符合语音产品快速简洁的定位要求。

(4)当语音搜索结果唯一时直接播放,以减少用户的操作步骤;若搜索结果不唯一,则显示结果列表,让用户对搜索结果进行选择。

IPTV 智能电视语音搜索的界面如图 2 所示,具体介绍如下。

- 用户在观看 IPTV 过程中,可以在任意页面调出语音搜索功能,语音搜索界面以半透明的形式叠加在 IPTV 界面下方,如图 2(a)所示。用户观看的节目保持原来的播放速度,不影响观看。
- 用户第一次使用语音功能时,会弹出“帮助”图示,告知用户如何使用语音功能,使用户能更快地掌握语音搜索业务的使用技巧,如图 2(b)所示。
- 当语音搜索有结果时,搜索结果以半透明的形式叠加在 IPTV 界面下方,通过海报和节目名称进行呈现,方便用户选择,如图 2(c)所示。当搜索结果较多时,用户可以按翻页键,以对更多的结果进行选择。
- 当没有搜到用户想观看的节目时,语音平台根据用户输入的关键词进行分析,给出推荐的节目列表,如图 2(d)所示。

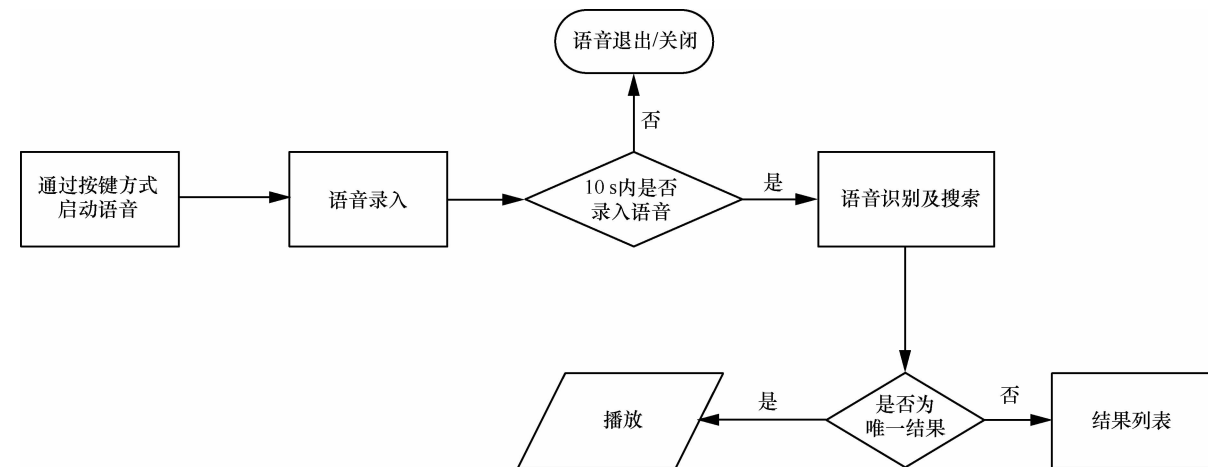


图 1 IPTV 智能电视语音搜索产品的流程

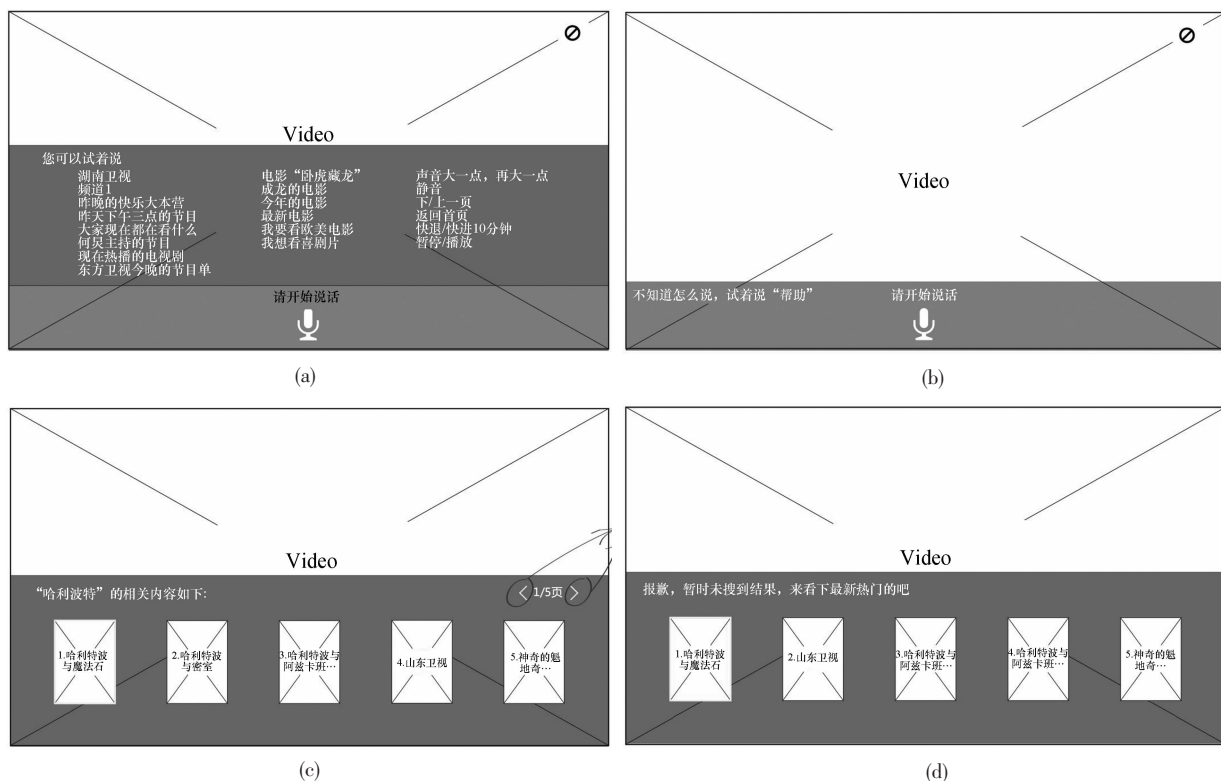


图 2 语音搜索界面

4 语音平台的功能要求

语音搜索业务到底能搜索到哪些节目,能听懂用户哪些语句,这些都需要语音平台提供支持。本方案的 IPTV 语音平台由语音识别服务器、语义理解服务器、IPTV 知识库构成。用户输入的语音音频先经 IPTV 知识库训练的语音识别服务器转换为文本,再通过语义理解服务器将文本转换为 IPTV 节目对象列表输出,输出的结果由机顶盒呈现。

(1) 语音识别服务器

语音识别服务器的原理如图 3 所示。

与 IPTV 相关的语音语料和文字语料分别通过训练形成声学模型和语言模型,语音信号经过特征提取后,与声学模型和语言模型进行比对并进行语言解码,输出音频相

对应的文字。

IPTV 用户输入的语音音频经过语音识别服务器转化为音频对应的文字。语音识别服务器可以支持全中文任意语音音频的输入,系统可以自动将之转化为对应的文字并输出到语义理解服务器。

(2) 语义理解服务器

语义理解服务器的原理如图 4 所示。

语音识别服务器输出的文本通过与语义理解系统的句式模板相匹配,对语音识别服务器输出结果进行纠错,同时提取其中的关键词,再通过 IPTV 相关的关键词匹配,输出对应的节目列表,并按照特定的排序规则,输出最终的节目列表。该方案采用语义理解服务器,在解析语音识别服务器输出结果的同时,还可以对语音识别服务器的

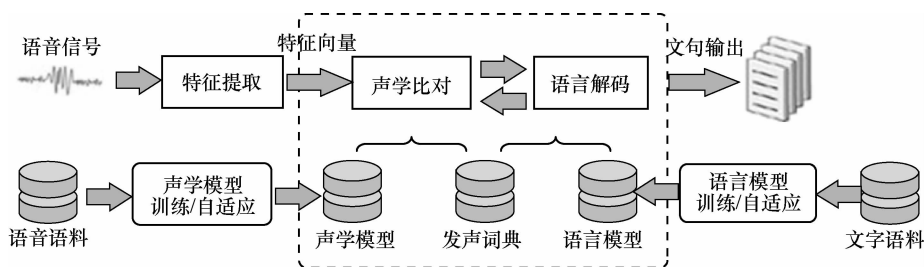


图 3 语音识别服务器的原理

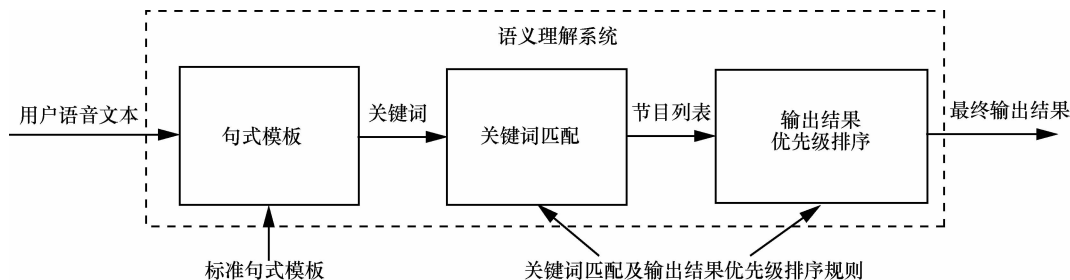


图4 语义理解服务器的原理

输出结果进行二次纠错,从而可以大大提高系统识别的准确率。

(3) IPTV 知识库

IPTV 知识库的原理如图5所示。

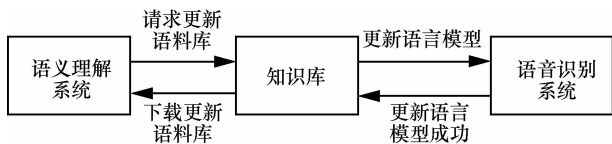


图5 IPTV 知识库的原理

IPTV 知识库主要实现以下3个功能:

- 从IPTV 节目服务器上下载节目信息并进行存储;
- 对节目信息进行解析、分类,并提供给语义理解系统进行关键词匹配和节目列表输出;
- 将节目信息提炼成文字语料,供语音识别服务器训练成语言模型。

本方案将IPTV 知识库独立于语音识别服务器和语义理解服务器,可以使它在为语义理解服务器提供IPTV 节目更新库的同时,也能为语音识别服务器提供最新的训练语料,从而大大提高整个系统的实时性。

本方案的优势主要体现在以下3个方面。

- 与市面上的智能电视语音控制系统相比,真正实现了全中文语音的任意输出以及全量节目的任意搜索。集成了全中文语言模型的语音识别引擎,可以针对任意中文语音输入进行语音识别,再配合集成丰富句式模板的语义理解引擎,可以从任意句式的文本中提取出业务信息。
- 语音识别与自然语义理解技术相结合,保证输出结果的准确性。市面上大多数语音控制电视仅用语音识别技术将音频信息转化为文本结果,但语音识别的识别准确率与音频源文件的质量关系很大,并不能保证完全识别准确。本方案利用语义理解技术对语音识别的输出文本进行二次纠错,大大提高了输

出结果的准确性。

- 将IPTV 知识库独立出来,可以同时服务于语音识别系统和语义理解系统,大大提高系统增量信息的更新效率。由于IPTV 系统的节目更新时效性非常高,会出现很多新节目,如果不及时训练到语音识别平台的语言模型中,会影响语音识别的准确率,本方案将集成在语义理解平台中的知识库独立出来,使它在为语义理解平台提供节目搜索资源的同时,也能建立语音识别平台的训练库,并对语音识别平台进行实时训练,从而大大提高IPTV 节目识别和更新的效率。

5 IPTV 智能电视语音搜索的系统架构

图6为本文提出的IPTV 语音搜索产品的总体系统架构。

(1) 用户终端

用户终端是指在用户家庭内部可以使用的智能终端,主要包括:语音遥控器,具备语音功能的IPTV 遥控器,将用户语音通过麦克风进行拾取,进行A/D 转换之后传给IPTV 机顶盒;IPTV 智能机顶盒,基于Android 操作系统,集成了中国电信的IPTV、视频通信、TV 伴侣、应用商城等应用,其中TV 伴侣是语音客户端内集成的语音控制模块和视频叠加模块,用于智能语音功能的实现和界面展现。

(2) 智能语音平台

包含语音识别平台、语义理解平台、IPTV 知识库,提供针对IPTV 应用的语音识别客户端调用能力,将语音转化为文字并识别出客户端传递的文字内容含义,从而可以被客户端用来对机顶盒进行控制。

(3) TV 伴侣平台

TV 伴侣平台是本方案所特有的,主要功能是将IPTV 内容元数据同步给智能语音平台中的语义理解服务器,同步的内容元数据包含点播内容和直播频道列表以及直播、

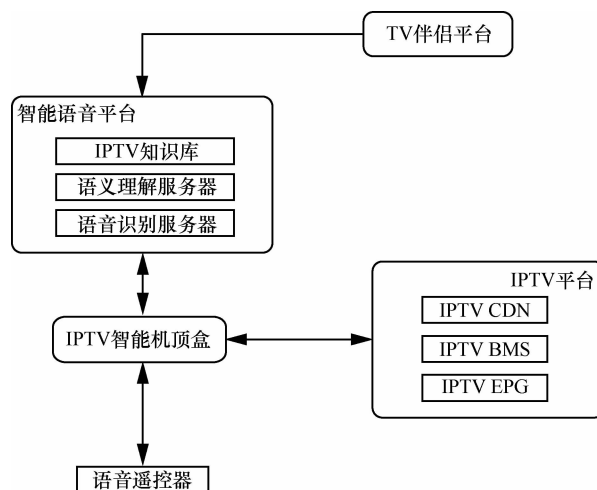


图6 IPTV 智能电视语音搜索产品架构

回看节目单等 IPTV 智能语音遥控业务所需要的相关信息数据。

IPTV 的内容每天都在实时更新,TV 平台需要实时将 IPTV 的播放内容元数据注入语义理解服务器,从而保证用户能随时搜索到最新的 IPTV 视频内容和节目列表。

(4)IPTV 平台

IPTV 业务中提供 EPG 及流媒体服务的基本能力平台具有 EPG 导航浏览、直播频道、即时时移、菜单时移等基本业务功能。业务能力平台接收来自于内容运营平台的基本业务内容及服务打包关系,当业务管理平台完成内容、服务、产品的打包后,由该平台为用户提供 EPG 浏览及视频流媒体服务。用户在使用业务过程中,通过与业务管理平台的横向接口完成基本的业务鉴权、订购等操作。

基于以上架构的 IPTV 智能电视语音搜索业务已于 2013 年完成开发,经过 3 个多月的测试和优化,2014 年 4 月初在上海电信开始试商用。

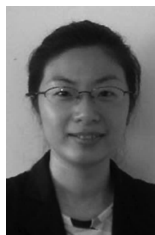
6 结束语

随着智能语音产品的全面推广,语音技术正逐渐成为智能电视的标配。语音识别技术面向语音交互控制的需求,还有许多需要完善的技术点,以下两个方面的研究需求最为急迫:第一,降噪技术,语音交互控制的场景十分复杂,依靠硬件降噪的技术在产品实际推广中会遇到问题,依靠软件进行服务端降噪及有效控制语音分离技术显得非常重要;第二,语言模型实时或准实时更新问题,虽然大词汇量连续语音识别解决了目前所需的应用问题,但语言模型本身的更新及维护在实际业务系统中还存在问题,即如

何使语音识别系统适应业务的发展,满足业务内容变动的识别需求。

伴随着语音识别技术和智能电视的发展,越来越多的用户将会体验到智能电视语音搜索带来的友好、智能的视频体验。

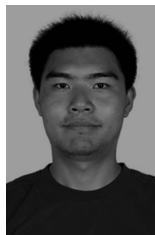
[作者简介]



王琳,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。



肖晴,女,博士,中国电信股份有限公司上海分公司教授级高级工程师,主要研究方向为 IPTV 及视频服务。



马宇飞,男,中国电信股份有限公司上海研究院产品开发工程师,主要研究方向为语音识别技术、自然语言理解技术。

(收稿日期:2014-05-04)