语音识别技术人工智能手机助手现状

王华民

郑州大学软件学院

## 一：前沿

随着数据处理技术的进步以及移动互联网的快速普及,计算机技术被广泛地运用到了社会的各个领域,随之而来的则是海量数据的产生.其中,语音数据受到了人们越来越多的重视.语音识别是一门交叉学科.近二十年来,语音识别技术取得显著进步,开始从实验室走向市场.人们预计,未来10年内,语音识别技术将进入工业、家电、通信、汽车电子、医疗、家庭服务、消费电子产品等各个领城.语音识别听写机在一些领域的应用被美国新闻界评为1997年计算机发展十件大事之一.很多专家都认为语音识别技术是2000年至2010年间信息技术领域十大重要的科技发展技术之一.语音识别技术所涉及的领域包括:信号处理、模式识别、概率论和信息论、发声机理和听觉机理、人工智能等等.[1.]

## 二：语音识别技术概述

语音识别技术，也被称为自动语音识别Automatic Speech Recognition，(ASR)，其目标是将人类的语音中的词汇内容转换为计算机可读的输入，例如按键、二进制编码或者字符序列。与说话人识别及说话人确认不同，后者尝试识别或确认发出语音的说话人而非其中所包含的词汇内容。语音识别技术的应用包括语音拨号、语音导航、室内设备控制、语音文档检索、简单的听写数据录入等。语音识别技术与其他自然语言处理技术如机器翻译及语音合成技术相结合，可以构建出更加复杂的应用，例如语音到语音的翻译。语音识别技术所涉及的领域包括：信号处理、模式识别、概率论和信息论、发声机理和听觉机理、人工智能等等。[2.] 语音识别是解决机器“听懂”人类语言的一项技术。作为智能计算机研究的主导方向和人机语音通信的关键技术，语音识别技术一直受到各国科学界的广泛关注。如今，随着语音识别技术研究的突破，其对计算机发展和社会生活的重要性日益凸现出来。说话人识别是用来进行身份确认的一种认证技术,通常也被称为声纹识别。它通过不同说话人语音中的个性特征来识别说话人身份,具有操作方便,设备成本低等优势,并被应用于证券银行、军事国防以及公安司法等领域。深度学习是近年来发展比较迅速的一种深层机器学习模型,其本质上是多层非线性的神经网络,可以对复杂的数据关系进行建模。将深度学习应用在说话人识别中,优化说话人语音个性特征,提高系统性能[3.]

## 三．语音识别的研究历史

语音识别的研究工作始于20世纪50年代，1952年Bell实验室开发的Au dry系统是第一个可以识别10个英文数字的语音识别系统。1959年，Rorgie和Forge采用数字计算机识别英文元音和孤立词，从此开始了计算机语音识别。60年代，苏联Matin等提出了语音结束点的端点检测，使语音识别水平明显上升；Vintsyuk提出了动态编程。 60年代末、70年代初的重要成果是提出了信号线性预测编码(LPC)技术和动态时间规整(DTW)技术，有效地解决了语音信号的特征提取和不等长语音匹配问题；同时提出了矢量量化(VQ)和隐马尔可夫模型(HMM)理论。80年代语音识别研究进一步走向深入：HMM模型和人工神经网络(ANN)在语音识别中成功应用。1988年，FULEE Kai等用VQ/I-IMM方法实现了997个词汇的非特定人连续语音识别系统SPHINX。这是世界上第1个高性能的非特定人、大词汇量、连续语音识别系统。进入90年代后，语音识别技术进一步成熟，并开始向市场提供产品。许多发达国家如美国、日本等都为语音识别系统的实用化开发研究投以巨资。同时汉语语音识别也越来越受到重视。IBM开发的Via Voice和Microsoft开发的中文识别引擎都具有了相当高的汉语语音识别水平。进入21世纪，随着消费类电子产品的普及，嵌入式语音处理技术发展迅速[2]。基于语音识别芯片的嵌入式产品也越来越多。

## 四．手机助手使用情况

语音助手是一款智能型的手机应用,通过智能对话與即时问答的智能交互,实现帮助用户解决问题,其主要是帮助用户解决生活类问题。三年前苹果公司发布iPhone 4S,引人注目的智能手机语音助手Siri第一次将语音识别技术引入到移动终端的人机交互中,引发语音识别技术的发展浪潮,成为智能语音交互时代的开端。语音交互作为一种最自然、最和谐、最有效的交互手段,开始受到了人们普遍重视。移动互联网市场瞬息万变,移动产品竞争越来越激烈。移动化时代的来临改变了用户需求,手机屏幕越来越大改变了用户的操作习惯,移动产品功能越来越复杂增加了用户的操作难度,这些改变都影响产品的交互效率和用户体验。语音助手类产品几乎都面临着功能的同质化、交互效率低、用户粘度不够、语音识别准确率在复杂条件下距离实用化尚有距离等问题。[4.]在刚问世的时候，智能助手被赞为“下一个伟大事物”，但结果却是，买账的人并不多。Siri、Google Now和微软小娜都未能获得大众的亲睐。实际上，Siri很快从让人们感到惊奇的东西变成了大家嘲笑的对象。

为了更好地促进智能助手的接受程度，App界面设计的启示性侧重于人脑的认知研究，挖掘物理世界事物的固有属性，用人们熟悉的元素在虚拟界面中构筑新的概念，从而优化数字产品的性能。从交互界面的可用性视角出发，遵循“以用户为中心”的设计理念，剖析影响用户认知的思维惯性、操作环境等关键要素，结合移动端特性，寻求APP界面设计中启示性的应用内涵。启示性善用隐喻对事物的固有特征进行强调，它是用户了解产品的“心灵钥匙”。[5.]尽管智能助手在不断完善，但要真正为大部分人所用——尤其是年龄大的消费者，仍有很长的路要走。

关键词：人工智能 语言识别 智能手机 手机助手

1. 林圣晔. 语音识别技术[D] 上海:华东理工大学,2019
2. 郭亚娜. 基于深度神经网络的蒙古文语音识别系统[D] 内蒙古:内蒙古大学,2016
3. 刘俊坤. 基于深度学习和矢量量化的说话人识别研究[D] 南京:南京邮电大学,2018
4. 方婷.基于交互效率的智能手机语音助手的研究与设计[D] 湖南:湖南大学 2014
5. ding jianpeng The Affordance Research on App Interface Design of Smart Phone[D] 北京:中国地质大学(北京) 2017