

**项目开发计划规范**

学 院 计算机与信息技术学院

专 业 物联网工程

年 级 2017级

姓 名 20175105011 郑玉祥

2019 年 11 月 14 日

**前言**

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。 它是一门交叉学科， 正逐步成为信息技术中人机接口的关键技术。语音识别 技术与语音合成技术结合使人们能够甩掉键盘，通过语音命令进行操作。 语音技术的应用已经成为一个具有竞争性的新兴高技术产业。

**背景**

语音识别起源于 20 世纪 50 年代， 1952年 Bell实验室开发的 Audry 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，苏联的 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，这一提法在以后的识别中不可或缺。

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、韩国以及 IBM 、Apple、AT&T 、Microsoft 等公司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 ViaVoice 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。基于语音识别芯片的嵌入式产品也越来越多，如 Sensory公司的 RSC 系列语音识别芯片、Infineon 公司的 Unispeech和 Unilite 语音芯片等，这些芯片在嵌入式硬件开发中得到了广泛的应用。在软件上，目前比较成功的语音识别软件有 :Nuance、IBM 的Viavoice 和 Microsoft 的 SAPI 以及开源软件 HTK，这些软件都是面向非特定人、大词汇量的连续语音识别系统。

**原理**

模拟的语音信号进行采样得到波形数据之后，首先要送人到特征提取模块，提取出 合适的声学特征参数供后续声学模型训练使用。好的声学特征应当考虑以下三个方面 的因素。第一，应当具有比较优秀的区分特性．以使声学模型不同的建模单元可以方便准确的建模。其次，特征提取也可以认为是语音信息的压缩编码过程，既需要将信道、说话人的因素消除保留与内容相关的信息，又需要在不损失过多有用信息的情况下使用尽量低的参数维度，便于高效准确的进行模型的训练。最后，需要考虑鲁棒性，即对环境噪声的抗干扰能力。

以汉语为例，汉语按音素的发音特征分类分为辅音、单元音、复元音、复鼻尾音四种，按音节结构分类为声母和韵母。并且由音素构成声母或韵母。有时，将含有声调的韵母称为调母。由单个调母或由声母与调母拼音成为音节。汉语的一个音节就是汉语一个字的音，即音节字。由音节字构成词，最后再由词构成句子。汉语声母共有22个，其中包括零声母，韵母共有38个。按音素分类，汉语辅音共有22个，单元音13个，复元音13个，复鼻尾音16个。目前常用的声学模型基元为声韵母、音节或词，根据实现目的不同来选取不同的基元。汉语加上语气词共有412个音节，包括轻音字，共有1282个有调音节字，所以当在小词汇表孤立词语音识别时常选用词作为基元，在大词汇表语音识别时常采用音节或声韵母建模，而在连续语音识别时，由于协同发音的影响，常采用声韵母建模。

基于统计的语音识别模型常用的就是HMM模型λ(N,M,π,A,B)，涉及到HMM模型的相关理论包括模型的结构选取、模型的初始化、模型参数的重估以及相应的识别算法等。

**未来前景**

语音识别是涉及心理学、生理学、声学、语言学、信息理论、信号处理、计算机科学、模式识别等多个学科的交叉学科，具有广阔的应用前景，如语音检索、命令控制、自动客户服务、机器自动翻译等。当今信息社会的高速发展迫切需要性能优越的，能满足各种不同需求的自动语音识别技术。但是，这样的目标面I临着诸的困难，如：语音信号会受 到上下文的影响而发生变化；发音人以及口音的不同会导致语音特征在参数空间分布的不同；同一发音人心理和生理变化带来的语音变化；不同的发音方式和习惯引起的省略、连读等多变的语音现象；⑤环境和信道等因素造成的语音信号失真问题。