浅析语音识别原理

施超群[1] 陈坚刚[2]
（浙江工商职业技术学院，浙江 宁波 315012）

摘要：多媒体应用时代的到来，各项新技术应运而生，一直在被研究的语音识别技术也被更加广泛地应用在实际生活中。本文简单介绍和分析了语音识别技术原理，并且通过应用 windows7 的语音识别系统，使读者对语音识别技术有进一步的认识了解。

关键词：语音识别；信号分析；特征提取；隐马尔可夫模型
中图分类号：TP391.43 **文献标志码**：A **文章编号**：1671-9565(2011)03-094-03

Fundamentals of Speech Recognition

SHI Chao-qun CHEN Jian-gang
(*Zhejiang Business Technology Institute, Ningbo 315012, China*)

Abstract: With the arrival of the era of multimedia application, various new techniques have come into existence. The speech recognition technology is being widely used in real life. The paper simply introduces and analyzes the fundamentals of speech recognition,and use the speech recognition system of windows 7 to make readers further understand speech recognition technology.

Key words: speech recognition; signal analysis; feature extraction; hidden markov model

语言交流是人类一种天然的沟通模式。从儿童时代开始我们对语言的相关学习都是自发的，语言交流始终贯穿于我们的生活。它是那么的自然以至于我们根本没有发现这是一种多么复杂的现象。人类的声道和发音器官，是具有非线性特征的生物器官，不仅仅运行在有意识的控制下，而且受到性别及其成长因素情绪状态的影响。因此，声音会因为他们的口音、发音、清晰度、体积、速度等有着大幅的变动。人类希望能与机器进一步沟通，从而方便生产与生活，而在语音信号的传输过程中，我们不规则的语言行为方式会被背景噪声和回声，以及电特性(如话筒等电子设备)进一步扭曲。这一切可变性的声音来源使语音识别更加繁琐复杂。

语音识别是一个多层模式识别的任务。声音信号经过考察，结构分为一个层次词根单位(例如，音素)、词、短语、句子。每一层可提供额外的时间限制，例如，已经被认知的单词发音或法律上的单词序列，可以弥补错误或把不确定性降到较低水平。限制的最好方法是在所有较低的层次中利用概率相结合的决策，而只在最高层次中使用离散决策。

一个标准的语音识别系统如下图所示。

图 1 标准的语音识别系统

1 语音信号的预处理与分析

1.1 原始语音

实际的物理波形信号通过传声器(话筒)采集而形成

收稿日期：2011-04-15
作者简介：施超群(1981-)，男，浙江慈溪人，浙江工商职业技术学院教师，主要从事数字化多媒体及背景音乐系统工程方面研究；
陈坚刚(1982-)，男，浙江宁波人，浙江工商职业技术学院教师，主要从事楼宇智能化及节能技术方面研究。

994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

的随时间连续变化的模拟信号,在语音信号被采集的同时背景噪音也随之而来。

1.2 信号预处理

要将采集进来的模拟信号转换为计算机认识的信号,这就需要数/模转换(计算机中的声卡就有此功能)。其中最常见的方式称为脉冲编码调制 (Pulse Code Modulation),可分为采样、量化、编码三个过程。

1.2.1 采样。就是将模拟音频电信号转换成二进制码。耳朵可以听到的频率为 20Hz~20kHz, 根据采样定理:当采样频率大于信号的两倍带宽时,即采样频率大于 40kHz 时,采样过程就不会丢失信息。由于考虑到设备兼容性问题, 国际上将采样频率标准定为 44.1kHz。尽管从物理上可以证明大于 40kHz 的采样率可精确重构原始波形,但人的主管听音试验表明,采样频率越大音质越有保证。

1.2.2 量化。就是将模拟信号的采样的幅度值以特定单位进行度量,用它整数倍的数字进行表示的过程。先将整个幅度值划分成有限的区域,所有采样信号的幅度值分布在幅度值最相近的区域并用量化精度(比特数)表示。量化决定了声音动态范围, 因为都是 2 的整数倍, 如 4bit、8bit、16bit 等等。而且量化过程没有最低标准,精度一般越大越好。

经过时间上的采样和幅度上的量化后,连续的模拟信号就变成了离散信号,这样就转化为了数字信号。其中把量化好的采样值表示成二进制代码的这一过程就是编码。量化和编码基本上是同时实现的。

1.3 信号分析

为了简化后续过程必须先将数字信号中的语音信号进过转化和压缩,其中有许多技术能将其中的特征提取出来和压缩数据的时候尽量减少重要数据的丢失。

1.3.1 短时傅里叶变换。语音信号有短时的稳定,长时变化,瞬时稳态的特点,即信号的频率随时间的变化。传统傅里叶变换缺少时域定位功能而不适用于时变信号。

1.3.2 线性预测编码。这是广泛和最有效的语音分析技术,可以用过去的若干个语音信号样值的线性组合来预测以后的样值。

1.3.3 倒谱系数。语音识别中较为常用的特征参数是线性预测倒谱系数 LPCC (Linear Predictive Cepstral Coding) 和 Mel 频率倒谱系数 (Mel Frequency Cepstrum Coefficient)。前者体现人的声道特性容易受环境噪声干扰,后者构造了人的听觉模型,强调语音低频信息,有利于噪音屏蔽干扰。

2 语音识别的主要技术方法

在经过语音信号的预处理和分析,提取出来的特征参数要与模型库中的模板参数去匹配,而这个就是语言识别系统的核心。目前有代表性的语音识别方法主要有特征参数匹配法、隐马尔可夫法和神经网络法。

2.1 特征参数匹配法

这是一种传统的模式识别方法,其技术要点是:在训练过程中将预处理过的信号通过一次变换, 去掉冗余部分,而把代表语音本质的特征参数抽出来,成为相对应语音的模版;然后,在识别过程中,用同样的方法从待识别语音信号中提取出语音参数;最后,应用某种不变的测度,寻求语音参数与模版参数两者之间的相似性,用似然函数进行判决。

2.2 隐马尔可夫模型(Hidden Markov Model,HMM)

这是一种研究透彻,算法成熟,效率高,效果好,易于训练的方法,已成为识别的主流技术,目前大多数词汇量、连续语音的非特定语音识别系统, 都是基于 HMM 的模型。隐马尔可夫模型是马尔可夫链的一种,它的模式库不是预先存储好的模式样本,而是通过反复训练形成的与训练后输出信号相似合概率最大的最优模型参数。而且它的识别过程中待识别语音序列与 HMM 参数达到最佳匹配状态作为识别输出,是较为理想的识别模型。

2.3 神经网络法(ANN)

ANN 本质上是一个自适应非线性动力学系统, 模拟了人类神经元活动的原理,具有自适应性、并行性、鲁棒性、容错性和学习特性。基于神经网络的语音识别系统通常由网络结构、训练算法和神经元构成。由于具有良好的动态时变性能和结构,良好的动态时间关联特性,使其具有很大的发展潜力。

3 微软 windows 7 语音识别的应用

微软 windows 7 操作系统已经发布有段时间了,其中的语音识别功能可以比较好的控制电脑的很多功能了。首先我们准备一台装有 windows 7 操作系统的电脑,当然语音输入设备必不可少(如:话筒)。

3.1 设备调整

首先找到语音识别程序, 在开始-控制面板-所有控制面板项-语音识别 (如图 2 所示)。点击 “启动语音识别”,如果是首次运行就会出现关于语音识别系统的一些介绍,然后就会跳入麦克风设置向导(如图 3 所示),根据提示设置麦克风。这一步就是语音输入设备(传声器)的选择调整。

994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图 2　windows 7 语音识别系统

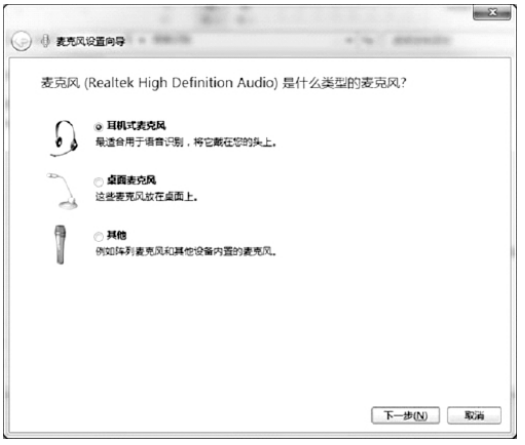

图 3　输入设备设置

3.2 学习教程

教程(如图 4 所示)的学习可以根据不同人的需要来学习,其中基础部分就展示了如何打开或者关闭语音识别系统以及一些常用到的最基础的信息。听写部分就是用声音输入文字,然后如何编辑声音输入后的文字(如:选择、删除等)其中包括一些常用编辑命令。命令部分就是允许通过声音控制窗口,启动程序,选择菜单等等,可以和屏幕上的项目交互。使用 windows 部分让人学习浏览窗口,控制窗口,启动切换程序等。完成了学习教程意味着您学会了如何使用这个语音识别的功能了。因为语音识别是一项前沿技术,有时候计算机不一定能完全听懂您说的内容所以您需要不断地训练它。

图 4　系统自带语音识别教程

3.3 训练计算机

就在刚才的学习教程的过程中,语音识别系统已经进行了由模拟信号到数字信号的转换,语音信号的预处理和分析,然后经过了特征参数的提取,和模式库的建立匹配,最后识别执行识别后的命令,这一系列的过程。为了更好的识别控制着的语音,用自然平稳的语调向计算机阅读文本,提高计算机理解语音的能力。通过反复训练(如图 5 所示)形成的与训练后输出信号相似合概率最大的最优模型参数,使识别更加准确。

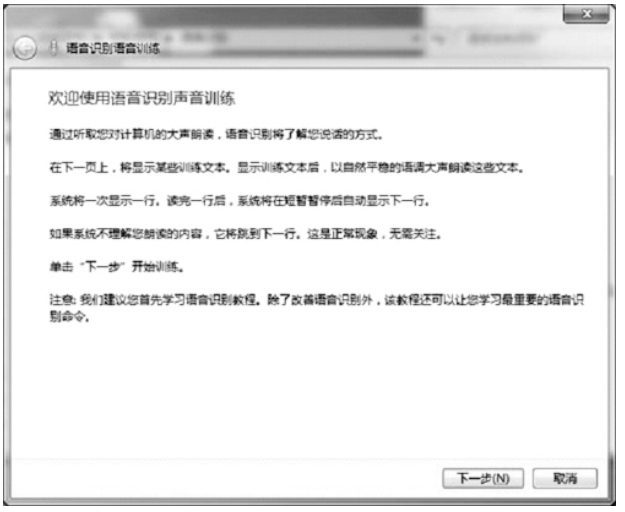

图 5 语音识别语音训练

虽然目前的语音识别技术还不完善,但是它已经应用到了日常生活的各个领域中,如数字通信、教育培训、医学医疗等方面。要使机器能理解人类的自然语言,能和人类交流,这是一个长远的长期的研究课题,但也是一个具有划时代意义的重大工程。

参考文献:

[1]胡泽,雷伟. 计算机数字音频工作站[M]. 北京:中国广播电视出版社,2005.
[2]柏逢明. 音频检测技术与仪器[M]. 北京:国防工业出版社,2006.
[3]刘幺和,宋庭新. 语音识别与控制应用技术[M]. 北京:科学出版社,2008.
[4](德)布劳尔特(Blauert,J.). 通信声学[M]. 李昌立,李双田译.北京:科学出版社,2009.

【责任编辑:熊荣生】

994-2017 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net