南京航空学院

研究生硕士学位论文

研究生姓名 胡 嘉
专 业 计算机应用
研究方向 微机系统结构、应用
指导教师 夏振华 高级工程师

一九八三年一月

VIMS语音输入微机系统的设计与实现

摘要

本文介绍了一种与说者有关的孤立单词语音识别系统。该系统用带通滤波和线性时间规正的方法提取语音特征，用改进最近邻法进行识别。实验结果表明，该系统已达到了平均97%的识别率。文中，介绍了识别系统主要部分的原理及实现方法；着重分析了最近邻法引起的分类性能及其在实际应用中的限制，在此基础上提出了一种有效的学习方法；此外，还详细地讨论并用多门限法解决了在确定语音首尾时改进引起的问题。最后，本文给出了一个应用实例——在16K扩展BASIC上实现的语音控制BASIC。

DESIGN AND REALIZATION

OF

A VOICE INPUT MICROCOMPUTER SYSTEM (VIMS)

Abstract

In this paper, a speaker-dependent isolated word recognition system VIMS is described. This system extracts the speech feature by bandpass filtering and linear time normalizing, and recognizes the input utterances, using the modified nearest neighbor rules. The experimental result shows that the system has reached an average accuracy of 97%. The principle and method for accomplishing every part of the system are introduced. Especially analysed are the performances of the nearest neighbor rules and some of its limitations in practical use, on the basis of which an effective learning method is proposed. Besides, the problems encountered in determining the endpoints of utterances are discussed in detail and solved with the help of the so-called Multi-Threshold method. Finally, an application example---Voice-controlled BASIC realized on 16K EXTENDED BASIC, is given.

目录

页子

目录(续)

第一章 引言

语音识别技术是人工智能的一个分支。自从1952年第一台语音识别装置问世到现在的三十多年间，许多研究人员在这方面作出了极大的努力，也正是因为他们的辛勤劳动，在语音识别这个领域内，才出现了不少可喜的成果：涌现出了多种识别的新技术。

语音识别系统无论用在什么场合，都将显示出它那独特的优点：

(1) 语言，是人们通讯的最自然、最有效的方式，因此，用户勿需接受什么特殊的训练就将熟练地使用语音识别装置。

在我国，电子计算机的应用正在全面普及，形势要求越来越多的人学会使用计算机，——从机关到地方、从大学到中学、从工厂到研究所，都是如此。对计算机专业人员来说，摆在我们面前的任务是如何使计算机的操作更加简便。我国的用户，可以说，大多数大体得其文，这更给他们操作计算机带来了困难。因此，在我国大力开发语音识别技术就更显得有必要了。

(2) 语音识别装置直接与用户打交道的只有一只话筒，就以它的使用可以不受什么客观条件的限制：既灵活又方便。人们还可以用它，通过无线收发装置，对远距离的设备进行遥控。例如，人们可以用语音对正在从事危险工作的机组人发出指令，让他们来完成我们预想的动

又无法做到的习惯；又如，在狭小的飞机座舱里，人们无法再设置计算机终端的键盘和显示器，飞行员也只能用到的手段与空中设备或地面控制设备进行通讯，在这到的手段中，语音识别是很有力的一种。

(3) 使用语音识别装置时，只要把麦克风固定在使用者的嘴边(例如使用头盔式麦克风)，因此，可以把使用者的手、眼释放出来，同时进行其它的工作。有时，人的手眼已忙于一项重要的工作，而同时又想口授一些命令来控制其它机器运行，这时，语音识别更能发挥它的作用。

此外，语音识别还允许用户同时控制多台机器；同时与机器和其它的人进行通讯。

语音识别的优点很多，应用范围也很广；特别是发达研制机器人和第五代计算机的需要，它更引起了人们的关注，研究者也越来越多。最近，在日本召开的一次有关第五代计算机的学术会议上，欧洲共同体提出的发展"外接口"的设想，其中就包括语音识别这个研究方向。

语音识别系统有孤立单词识别系统和连续语音识别系统两大类，每一类又可分为说者有关(Speaker-dependent)与说者无关(Speaker-independent)两种。目前，虽然有很多人在研究连续语音识别系统；然而，据我们所知，这些系统还只处在连续数字的阶段，虽说也能识别非数字语音，但效果不好。说者有关系统要求每个用户在使用前先对系统进行训练，而说者无关系统

别句需这样做。但话者有关系统而识别内容比较灵活；话者无关系统由于必先要对各种人而大量而发音样本进行处理，故识别内容不易更动。

语音识别是模式识别理论而一个应用，一个语音识别系统就是一个完整而模式识别系统，它主要包括三个部分：信息四、特征提取四、分类四。在语音识别这个领域中讨论得最多而是特征提取四，即抽取特征和压缩数据而方法。1952年，Bell电话实验室而Davis等人研制出了第一台孤立单词语音识别四(话者有关)，它能识别0～9十个数字。该识别四使用900Hz上下而两个通道(带通滤波四)来分析语音信号，分别提取两个通道而过零率，或过零次数（axis－crossing）作为特征，然后使用称谓而"最佳匹配法"(best－match)来作出最后而分类决策[1]。数年之后，Dudley和Balashek开发了一个称为"Audrey"而识别系统，用十个频率通道来提取语音而频谱特征，把所得而特征向量与存贮而模式进行比较，最后求得分类结果[2]。应用计算机来进行语音识别大概出现在1959～1960年之间，那时，Denes和Mathews提出了时间规正(time normaligation)而概念，并把它应用于实践，取得了成功[3]。

此后，各种特征提取而方法层出不穷——各种系统，有孤立单词而，也有连续语音而，有话者有关而，也有话者无关而，数目之多令人眼花缭乱。文献[4]对此作了评

细的介绍。这种方法的编程技巧不相同，比较起来也各有利弊；但它中间，人们比较重视的方法可分为带通滤波四组法和 LPC 法 [3]（Linear Predictive Coding —— 线性预测编码）。比较这两种方法，LPC 法的特征要比带通滤四法的特征准确，但是，LPC 的运算量很大，单靠程序实现在目前的条件下还难以做到实时。

国内，这方面的工作开始得比较晚，但经过努力也取得了不少成果。据我们所知，中科院声学所 [5]、[6]、清华大学 [7]、沈阳自动化所，使用带通滤波四的方法，先后在频谱分析仪上实现了语音有关的孤立单词识别系统。另外，西工大在单板机上实现了这样的系统；沈阳自动化所还对语音无关系统进行了实验。此外，其它的方法也正在被大力开发。

目前，我系正在着手进行人工智能和机器人的研究工作，鉴于语音识别的重要性，并且结合教研室的科研计划，我们选择了这一课题。考虑到客观条件的限制以及科研任务的要求，设计本课题的宗旨是：在 Cromemco 系统实现有的硬件条件下，使用相对简单的电路来进行实验，并且要求识别系统具有较好的实时性。因此，我们选择了通道滤波四的方法，不用频谱分析仪，只用 *孤立元件设计了四个低频通道来抽取元音的特征，一个高频通道来检测辅音。高频通道的引入使系统检测辅音的能力大大提高，克服了国类系统难以检测弱摩擦音的缺点；*

但引入通道后引入会影响系统的抗干扰性能，为此，本文分析了语音信号和干扰噪音的特点及二者的区别，在此基础上使用了多次判头尾的双门限和自适应门限结合的方法，采用循环取模的存储结构，较好地解决了这个问题，提高了系统的抗干扰性。

从各种不同的识别系统可以看出，人们对特征提取的研究比较多，而对分类的，即识别方法的研究却很少。然而，特征提取是分类算法的基础，特征提取的好坏会直接影响到整个识别系统的性能；然而，我们也不能忽视分类方法的重要性。我们知道，即使采用相同的特征提取方法，但分类方法不同，系统的性能也不会相同。本文分析了最近邻法的性质，说明了某些系统在使用最近邻法时的不足，提出了一种识别过程中的学习的方法。最后，作为语音识别系统的一个应用，本文在 Cromemco 16K 扩展 BASIC 的基础上实现了语音控制 BASIC —— VBASIC (Voice BASIC)，并介绍了 VBASIC 解释程序的结构，使用方法，和实现过程中所遇到的问题及解决的方法。

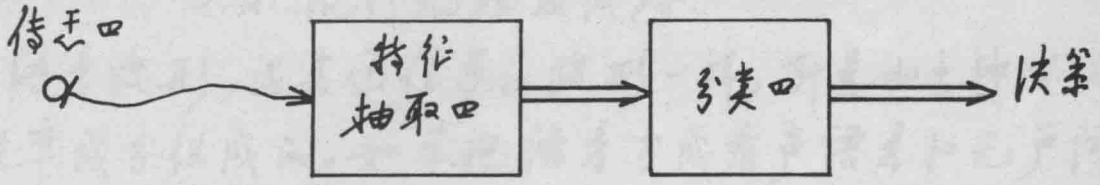
本文的第二章详细介绍了系统的设计原理、整个识别系统的结构及系统的识别过程，另外，还给出了本文所采用的放大、滤波的原理图。第三章讨论怎样确定语音信号首尾的问题，分析了语音与背景噪音的共同性和二者的区别，最后，根据语音与噪音的各自特点，给出了判别算法。第四章给出了参数规正和数据压缩的概念和方法。

第四章详细地分析了最近邻法的分类性能，以及使用最近邻法时存在的问题，并给出了解决的方法。最后，在本文的第三章，我们给出了一个应用实例——VBASIC，并且介绍了修改 BASIC 解释程序的方法及修改程序。

第二章 设计原理与系统结构

2.1. 概述

语音识别的过程也是一个模式识别的过程。从理论上说，这个过程可以分为两个步骤，一个是特征抽取，另一个是模式分类。但是，为了抽取各种模式的特征，还必须要有一个待定口。这样，我们就可以得到模式识别系统的一个抽象模型（分类模型），它包括三个部分：待定口、特征抽取口、分类口，它们的联接关系如图[1]所示。



图[1]：模式识别系统的一个抽象模型

待定口的作用是检测、录入，同时把录入信号转换成适于机口处理的形式；特征抽取口（也称接收口、特征滤波口、原性检测口、预处理口等）是紧接着待定口的一个环节，它从待定口送来的数据中抽取有关的信息，作为反映该模式的特征。每类模式都存在着能反映出该类模式本身固有的特征，它也是一模式区别于它模式的特殊标志。最后，分类口使用这些标志，根据一定的策略作出最后的决策，即把对象的模式分到它应属的类别中去。

语音识别中，待定口就是话筒（麦克风），它把语音变为电信号送给后面的部分——特征提取口和分类口。从理论上讲，特征提取口和分类口之间没有明确的分界线。

一个理想的特征抽取四可以使分类四的工作变得简单，同样，一个全线的分类四也将无求于特征抽取四；然而，我们将会看到，在实际问题中，这种区分还是有必要的，这是因为特征抽取问题比分类问题更依赖于实际问题本身，区分后可以使设计工作更为单一。我们知道，一个适合于各次识别的特征抽取四会对语音识别毫无用处。

本文根据课题的要求，使用通道滤波四组来提取语音信号的特征。

2.2. 设计原理及线路

语音波形，与其它信号的波形一样，都是由多种不同的频率成分组成的。如果把语音分成有声语音和无声语音两大类的话，从频谱图中可以看到（附录[1]中图[1]～[4]），有声语音的低频段内容比较丰富，波形具有明显的周期性（这点可从附录[2]的波形图中看到）；从频谱图中还可以看到在频率取某几个值时谱线具有明显的峰值，这几个频率值就是声道的谐振点，通常称之为共振峰。一个元音（其它浊辅音也有相似的性质）有好几个共振峰，它们分布的频率范围大致在 300～4000 Hz 之间。一般说来，前几个——2或3个——共振峰对区分元音比较有用。元音不同，共振峰也不同；元音相同，共振峰也不致相同。例如，附录图[1]、[2]和[3]、[4]的频谱图共振峰看到较大，而图[1]和[2]、图[3]和[4]就很相似（注意纵座标的不同）。目前已经公认，共振峰——频谱包络的一系列峰值，是区别不同元音的主要

特征量[8]。下面给出汉语(普通话)十个元音的前两个共振峰可部助说明这一性质。

表：汉语元音的共振峰(引自[8])

元音		a	o	e	i	u	ü	ê	er	ɿ	ʅ
F_1(Hz)	男	905	631	598	337	377	341	587	662	448	426
	女	1061	772	853	394	491	415	670	797	511	522
F_2(Hz)	男	1236	1113	1134	2380	641	2132	2122	1578	1993	1498
	女	1464	1131	1389	3098	836	2660	2428	1763	2381	1874

那些无声辅音，它们没有明显的共振峰，谱线较平，而且覆盖的频带很宽，高频分量(5K以上)很丰富，见附录[1]各[5]。

以上的分析是设计带通滤波组——频谱分析的依据，工作原理可大致表述如下：

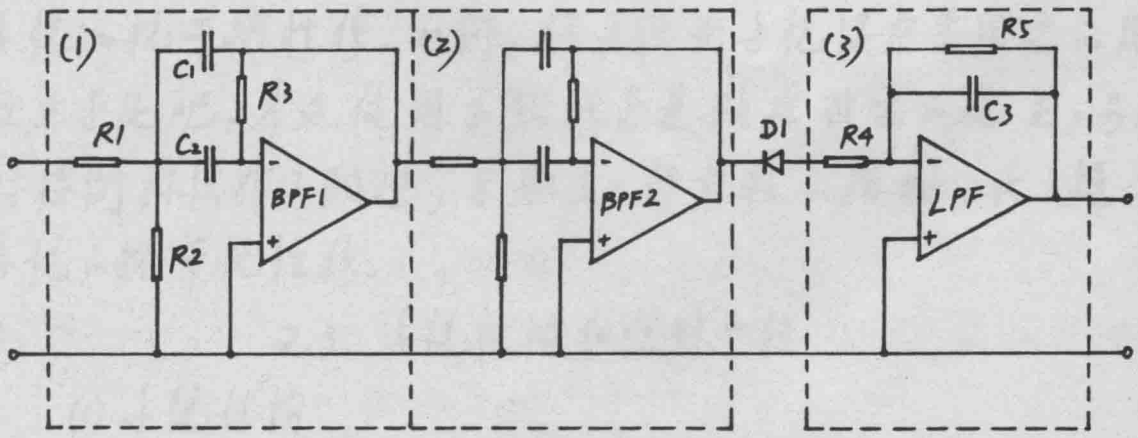
语音信号取入后进入由一组带通滤波电路组成的频谱分析，滤波的数目和频率划分可有许不同，一般可由10～16个滤波覆盖主种不同语音信号的频率范围。通过带通滤波对语音信号进行频率分析以后，再经过一个整流和低通滤波电路(截止频率 $f_c = 25$ Hz)加以平滑，在分析的取出端便得到了语音信号的频谱包络。

A/D转换对取入的频谱包络进行采样，将得到的数字量送入计算机进行处理(数据压缩和分类)。

清华大学用一台数字式动态频谱分析仪完成了上述功能[7]，在200～4000Hz这个频率范围内使用了14个通道

来抽取语谱仪，取得了较好的效果。在选择方案时，考虑到客观的条件，我们设计了一组带通滤波电路和整流、低通滤波电路来代替动态频谱分析仪。

通道数目的选择没有唯一的标准，从理论上讲，通道越多越好，越多，抽取的特征越准确；但是，在实际的应用中，考虑到诸如造价、复杂性等因素，通道的数目不可能做到很大。文献[7]由于使用了现成的分析仪，所以选择了14个通道，但在本文中，一共设计了5个通道，4个低频通道和一个高频通道。设计时，主要考虑让四个低频通道覆盖元音的前两个共振峰(主要的共振峰)，高频通道检测辅音。每个通道都由两个二阶带通滤波电路和一个低通滤波电路组成，见参[2]。参中



参[2]：通道线路参

参中，虚线框(1)、(2)内的电路及元件参数完全相同，是两个二阶无限增益多路反馈带通滤波电路[9]。两个滤波电路级联是为了使幅度响应曲线更加陡峭。二极管具有单向导电性，作整流用。框(3)的电路是低通滤波电路，它和二极

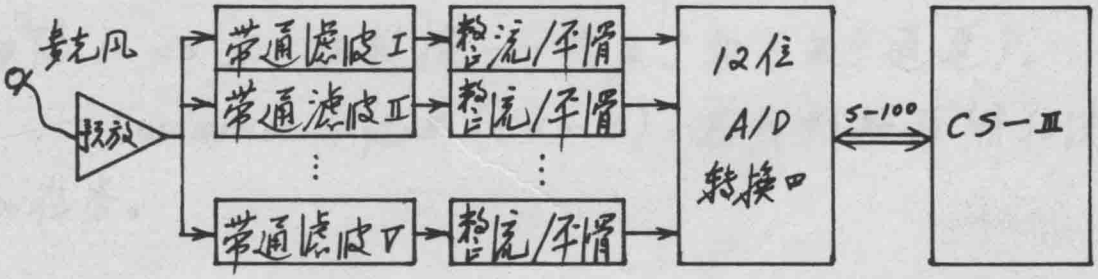
管一起作抽取频谱包络之用。

五个通道的中心频率分别为225、450、900、1800、7200 Hz，品质因素(Q值)为3。五个通道中，每个的放大倍数各不相同，滤波器的中心频率愈高，放大倍数也愈大，这是为了补偿人的声道特征，因为它在高频范围内产生的线号要比低频的小；所以，在确定放大倍数时应尽量使各种不同的频率成分有同样数量级的幅度。

值得一提的是，高频通道的使用使系统线的较好地检测到辅音"t"，尤其是摩擦音"s"，这使本系统在检测摩擦音的能力上大"超过了没有不使用高频通道的系统，但是，也正是由于高频通道的使用，使系统对外新"噪音更为敏感，因为外新"噪声大都具有较高的频率分量，这严重地影响了系统的抗干扰性线。也许，许多识别系统没有采用它的原因就在于此吧。本文使用高频线号是放虑到它的优点；另外，对检测环境作了改进，采取了一些有效的措施，大"提高了系统的抗干扰性线。

2.3. 系统结构和识别过程

(1) 系统结构

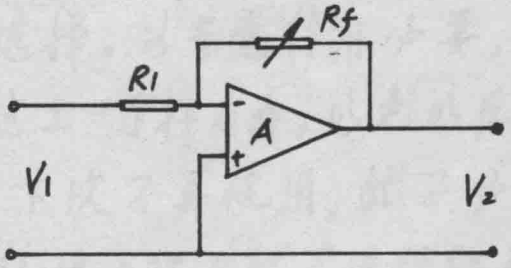


图[3]：语音识别系统的硬件框图

图[3]所示是整个语音识别系统的硬件框图，它由以下几部分组成：

•麦克风：它的作用是把语音信号变为电信号。这里采用的是灵敏度较高的驻极体话筒。

•预放大器：麦克风取出的信号很微弱，又采用灵敏度较低的动圈式话筒，信号就更为弱小。带通滤波器虽然有一定的放大倍数，但不能满足要求；因此，话筒的取出信号在进入带通滤波器之前，还必须经过预放大。图[4]是它的原理图。Rf是负反馈电阻，阻值大小可调，Rf改变，放大倍数也随之改变，这样可以使系统适应不同灵敏度的麦克风。



图[4]：预放原理图

•滤波、整流、平滑：得到谱包络。详见前节。

•A/D转换器：使用的模/数转换器是 Cromemco 的 AIM-12（12 BIT ANALOG INPUT MODULE），它通过 S-100 总线与 Cromemco 系统Ⅲ相联，把输入的模拟电压转换成12位的数字量送给主机。采样速率为100 Hz，也就是每隔10 ms采一组数据——共五千数（五千通道）。

•Cromemco 系统Ⅲ（CS-Ⅲ）：完成数据压缩和决策的任务。