最新语音识别技术

〔美〕Michael Koerner 编著　　李逸波　郭天杰　王华驹　等译

- 语音识别的用途
- 它是如何工作的
- 它的有效使用
- 裁剪定制语音输入听写系统(VTD)
- 七种语言的识别

IBM国际技术支持组织(ITSO)

PTR PH

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
URL: http: // www.phei.com.cn

最新语音识别技术

[美] Michael Koerner

李逸波　郭天杰　王华驹　等译

電子工業出版社
Publishing House of Electronics Industry

内 容 简 介

自从计算机诞生以来，人们就期盼着能用口述语言与计算机对话。几年以前，这种期盼还被当做是未来的梦，而如今，梦想与期盼变成现实——THE FUTURE NOW!

本书是 ITSO 的“REDBOK”之一，介绍了最新的计算机语音识别技术和由 IBM 公司提供的语音识别产品，这些产品适用于各种个人计算机及操作系统。有意思的是，本书内容正是使用这些语音识别产品听写制作而成的。

本书内容丰富，图文并茂，适用于专业技术人员及计算机爱好者使用。

书　　名：最新语音识别技术
著　　者：[美]Michael Koerner
译　　者：李逸波　郭天杰　王华驹等
责任编辑：焦桐顺
特约编辑：德姝
排版制作：电子工业出版社计算机排版室
印 刷 者：北京大中印刷厂
出版发行：电子工业出版社出版、发行　URL:http://www.phei.com.cn
北京市海淀区万寿路 173 信箱　邮编 100036　发行部电话 68214070
经　　销：各地新华书店经销
开　　本：787×1092 1/16 印张:13.25　字数：339 千字
版　　次：1998 年 4 月第 1 版　1998 年 4 月第 1 次印刷
书　　号：ISBN 7-5053-4593-1/TP·2174
定　　价：25.00 元
著作权合同登记号　图字:01-98-0056

译 者 序

这本名为《最新语音识别技术》的书是 ITSO 的 REDBOOK(红书)之一,它是介绍计算机语音识别技术的一本新书。使用语音与计算机对话,这是多年来普通计算机用户的希望。在语音识别产品已进入中国计算机市场的今天,我们翻译这本书是想为推动语音识别技术的应用作一些工作。本书翻译章节情况如下:王华驹译前言、第一章和第二章,郭天杰译第四章、第六章,李逸波译第三章、第五章、第七章和附录。全书由李逸波统稿,王翌、齐玉东、李连、刘英、刘建成、张功、张红波等同志作了大量协助工作。由于译者水平有限,书中难免有不当之外,敬请读者指正。

在本书的翻译过程中,得到张景生同志的大力支持和帮助,在此特别表示感谢。

目　录

前　言

让计算机理解常规的语言是人们多年来研究的一个目标。今天,语音识别在几乎是任意的环境中都可以加速信息的处理。

这本书涉及语音识别技术和 IBM 提供的适合不同的个人计算机硬件和操作系统平台的语音识别产品,还提供了如何使这些产品适合你的目标环境的细节。本书描述的制作听写产品的实例,实际上已被用来口述了本书,包括必需的接口。

此资料是用来为用户、经销商、系统工程师和顾问而写的,他们应具有个人计算机硬件和操作系统的一些知识。

本书是如何编排的?

本书是这样安排的:

• 第一章　语音识别介绍

这一章介绍语音识别。讲述了用户接口从键盘和鼠标到 GUI 和语音的发展。叙述了已经或可能运用语音识别的各种不同环境。

• 第二章　语音输入技术

这一章讲述了语音识别技术是如何工作的。

• 第三章　产品

这一章讲述了 IBM 对不同的个人计算机硬件和操作系统平台提供的各种语音识别产品。

• 第四章　系统选择、安装和注册

这一章讲述了什么样的系统能使你充分得到语音识别的好处。此外,还说明了在不同的个人计算机硬件和操作系统平台上安装和确定问题的过程。

• 第五章　使用语音识别

这一章提供了如何最好地使用语音识别技术的详细信息。这一章中心放在一个分为两部分的方案上,它说明了 IBM 语言口授产品的能力和性能。

• 第六章　民族语言支持状况

民族语言支持(NLS)是使用任何语音识别产品的关键。

——为什么 NLS 对语音识别产品特别重要?

——NLS 和语音识别的困难是什么? 如何解决这些困难?

• 第七章　加工制作语音听写产品

这一章解释如何制作语音识别产品以适应你的特殊环境。

• 附录 A　IBM VTD(VoiceType Dictation)产品订货说明

这个附录提供了从 Windows, Windows 95 和 OS/2 可以得到的不同的 IBM VTD 产品和 IBM VTD 词汇产品。

• 附录 B　IBM VTD 指南和文件结构

这个附录提供了 IBM VTD 产品的指南结构。

• 附录 C　错误代码

这个附录提供了 IBM VTD 适配器卡可能产生的错误代码。

• 附录 D　IBM VTD PCMCIA 兼容表

提供了 IBM VTD 1.32 和 IBM VTD PCMCIA 驱动器测试的便携式系统一览表

• 附录 E　CD-ROM 说明

这个附录提供了补充此书“语音识别 – The Future Now!”的 CD-ROM 的内容。

第一章　语音识别介绍

用户接口技术经历了从穿孔卡到命令行和到今天的图形用户接口(GUI)的发展过程。GUI是比较容易使用的。研究表明,大多数用户希望这些工具有更大的适应性。用户根据所要完成任务的要求及自己的爱好,来确定如何用计算机使工作做得更好。提高工作效率的关键是通过扩展计算机的接口能力使计算机具有更大的适应性,使人们能自由地选择让他们感到舒适的接口。

直到目前为止,由于没有实际可行的替代方法,人们只好调整自己的行为去适应技术条件的限制。因此,工作效率经常是很低的。但是随着新接口的出现,通过自动化装置,人们将获得工作的高效率。随着以下接口技术的发展,IBM正在促进该技术的发展:

1. 提供一个宽范围的工作媒体(从字母数字文本到全运动电视)。

2. 让人更多的感官参加计算机操作,比如语音。

每天都有更多的证据表明计算机越来越成为我们生活中的重要工具和焦点。计算机可以用来写信,玩游戏,听音乐,看电视,通过电视会议进行通讯等等。其用处是无止境的。例如,已在Internet网上盛行起来的WWW(World Wide Web),打开了信息共享的新大门。想象一下以下的景象:你坐在计算机旁,想知道当前IBM股票的价格,你进入WWW的家庭页面(home page),就能得到最新股市行情表。现在你想购物,就连接到购物网络上,发现了一套精美的服装。你用信用卡订购,就实现了购物。现在你又想给在南非的朋友打电话,通过applet拨打朋友的计算机,用你的带麦克风和扬声器的计算机和朋友通话。

计算机正在改变我们的通讯、工作和玩耍的方式。正象你在前面的例子中所看到的那样,计算机有许多的工作方法。

用户用他们最自然的方式工作时,会工作得更有成效。说一声"打开文件"不是比用鼠标在"文件"上单击一下,然后再在"打开"上单击一下更容易吗?用户头一次学习使用鼠标时,指针通常在几分之一秒的时间内从屏幕的一处跑到下一处。用户调节鼠标之后,将鼠标放在选定的图标上,但是当他们按下鼠标的按钮时,指针又离开了图标。这是很令人失望的。鼠标的使用需要一个适应过程。到目前为止,用户必须适应计算机。那么,现在是使计算机适应人的时候了。

语音识别是自动抄录用户说的话。计算机用语音识别软件听取指令,就像用键盘敲入和鼠标输入数据那样。计算机分析并解释这些指令。以前要求打印的一长串命令或要求鼠标频繁移动的复杂任务,现在口述一个命令就能完成。例如说"此信结束"(end letter),屏幕上就显示出信的最后一段并有结束语和你的签名。

通常,提高效率的方法是在软件程序上增加更多的功能并增加打印速度。用语音识别是我们正在经历的一次改变。我们不再寻找改善软件应用的方法,而是正在寻找和软件交互的新方法。

1.1　计算机与语音

新的交互方法是语音识别。这是与以往不同的方法,是你在桌面上用声音命令控制或操

纵计算机。你可能口述文本并把它转换为支持剪贴功能的任何应用程序(cut-and-paste function)。你可以口述文件、报告和邮件。

任何人都可以从语音识别获益。语音识别可以在几乎任意的环境中加速信息的处理。当人们利用他们的眼睛和手处理其他事务时,可以用语音识别进行口述信息。

经理们可以不用打印而利用语音识别来产生文件。他们可以继续打印他们的大多数材料,但机要文件可以迅速有效地口述到计算机中去。常有这种情况,经理能讲几种不同的语言,而秘书却不能。在这种情况下这个经理可以用他的第二种语言来使用语音识别器。

教授利用语音识别,可以口述任意选择的专门领域的单词和短语。例如,在急救室里一个病人被诊断为"**hematoma on the occipital skull**"(在颅骨上的血肿),另一个得了"**edema and ecchymosis under the right cheek**"(在右面颊下有水肿和瘀斑)。医生可以口述这些名词和短语到病人的记录里,并有很高的精确度。

用此方法,办公室的工作人员可以集中精力在他们的实际工作中而不是在打印材料上,这样就可以保持他们的工作思路。

1.2 IBM 的技术领导地位

早在 50 年代,IBM 就开始了语音识别的研究。IBM 研究语音识别的方法几乎与任何其他人都不同。IBM 采取统计的方法而不是基于规则的方法。较早的基于规则的方法起源于人工智能,光谱图说明了声音频率对时间变化的强度。人类专家推出基于光谱的规则。他们需要对原始数据(声音的频率图象)进行鉴定和分类,然后把分类的数据和已知的声音联系起来。当变换成计算机程序时,这个模式模仿人类解码语言的方法。然而,我们实际上并不知道人脑是如何处理信息的。这种方法的另一个问题是原始数据的变化。用这么少的数据得到的光谱图不是一张清楚的图形。对给定的声音光谱图 不能清楚地说出它的特性。因此,在推导规则时,专家们并没有立即得到反馈告诉他们推出的规则是否正确。

统计方法利用单词随时间变化的声音图象的几率,通过用大量数据训练系统和使用各种各样的统计分析程序。语音识别的统计方法有了一个快速的、准确的基础。这个方法成为语音识别的主要方法。有关此方法的更多的信息,参阅第二章"语音输入技术"。

直到 70 年代,IBM 才研究出相应的工具和技术,形成了语音识别的基础。70 年代末,IBM 成功地展示了识别自然英语语句发音的统计方法。

在展示实时语音识别系统之前,IBM 想知道什么样的语音听写设备用户才能接受。两名程序设计员在 Apollo 计算机上编制了一个实验程序来模拟语音识别程序的功效。Apollo 计算机是基于 Motorola 68000 处理器,运行在 Aegis 操作系统下(与 UNIX 类似)的精密图形计算机。该计算机被戏称为洗衣机。因为其尺寸和小的洗衣机相仿。一个人坐在 Apollo 计算机前用一个小麦克风口述一个记录,使得隔壁屋里的一个打字快手能听到这个人的口述。此打字员将口述的内容打到 IBM 个人计算机内。IBM 机将它发送到 Apollo 计算机,Apollo 以精美的格式把口述的内容显示出来。口述时,有意加了一定比例的错误,使其效果和语音识别程序一样。通过这个实验,IBM 得到了一些关于什么样的语音口述设备才能被用户接受的感性知识。使研究人员可以进行不同错误率影响和不同错误纠正策略的实验。

在 1984 年,IBM 首次展示了有 5,000 个词汇的语音识别系统。该系统适用于使用孤立的单词的自然文本的听写。这次展示在三个 6 英尺高的阵列处理器和一个 4341 主计算机上进

行,并带有用户与 Apollo 计算机的接口。通过口述让机器进行语音识别几乎可以作到了。

IBM 的研究人员认识到需要用个人计算机来取代这一庞大的设备。到 1986 年,主机系统已经减少到在辅助底盘上带 12 个适配卡的个人计算机。这头一个基于个人计算机的语音识别系统,能识别 5,000 个单词,使语音识别器独立起来了,它被恰如其分的叫做 Tangora.。

列入吉尼斯世界记录的 Apollo Tangora 保持着持续打字的最快记录。1942 年它创下这个记录时,人们仍然使用手工打字机。但在竞赛中打字员们使用了专门设计的比赛设备。除了作为一个出众的打字员外,Apollo Tangora 还有一个专门技术是送纸,叫做快速移动换纸。

当 IBM 推出了 Tangora 以后。其他实验室的研究人员看到了统计方法比基于规则的方法优越的地方,1987 年研制出两万个词汇的产品。两年以后,其硬件减少为一个单卡。

在 1992 年,IBM 推出它的第一个产品叫做"用于词处理和听写应用的 IBM 语音服务器系列(IBM Speech Server Series)。它可在多种配置下运行:

·独立的运行 AIX 的 RISC 系统 6000

·在局部网环境下的 AIX/6000 服务器和 OS/2 及 AIX/6000 用户

IBM 语音服务器系列有 2 万个词汇,另外可扩展 2,000 词汇以满足专门需要。用户利用它可用语言产生文本,因此获得高效益。对某些行业,如医学和法律使用的语言和词汇,有现成的 IBM 语音服务器系列。从一开始,IBM 的口述产品就同步地提供几种语言,或在提供两种语言之间有一些不长的延迟。这些语言是:

·美国英语

·英国英语

·阿拉伯语(只有对 OS/2 的 VTD1.1 版)

·法语

·德语

·意大利语

·西班牙语

IBM 语音服务器系列依赖于口述者,并要求分离或孤立地口述。依赖于口述者意味着用户必需先训练系统,使系统知道他们是如何说话的,然后才能使用系统。分离或孤立地口述意味着在每个词之间,用户必须作简短的暂停。训练系统的过程叫做登记注册(enrollment)。训练过程包括讲一组系统规定的句子,对分离语音系统,在讲这组规定的句子时,在每个词之间也必须作短暂的暂停。

在 1993 年 10 月,IBM 推出了它的下一个产品——IBM 个人口述系统。它也是依赖于口述者并工作在孤立方式。功能与 1.x 产品相同。IBM 个人口述系统是在个人计算机上运行的头一个产品。IBM 个人口述系统运行在 PS/2、ValuePoint 及和 IBM 个人机兼容的机器的 OS/2 2.1 版本下。IBM 个人口述系统提供了形形色色的附加词汇和专门行业的语言模式。

一年以后的 1994 年末,IBM 推出了 IBM 个人口述系统的新版本,命名为 VTD(VoiceType Dictation)1.1 版。IBM VTD 的平台扩展到 Windows 3.1 和 Windows 3.11。1.32 版是 1995 年推出的基于 Windows 的第一个 32 位的 VT(语音打印)版本。行业专业语言模式与这些版本兼容,也和 VTD 3.0 以下的版本兼容。

在 1996 年推出了用于 Windows 95 的 VTD 3.0 版。IBM VTD 产品现在包含了许多以前认为是只有将来才能实现的特点。IBM VTD 3.0 版包括以下四种口述方法:

·听写内容插入应用程序(快速口述)

·在应用中听写(直接语音打印)

·在 Microsoft Word 中听写(直接口述)

·语音控制口述和格式(VoicePad)

以上内容在第五章"使用语音识别"中详细解释。

OS/2 Warp4 包括 IBM VTD,VT 的所有优点结合在 OS/2 的这个版本中。将导航命令直接编入操作系统中,充分显示了易于使用的优越性。口授一个应用,OS/2 窗口和 WIN - OS/2 窗口都不需要训练。

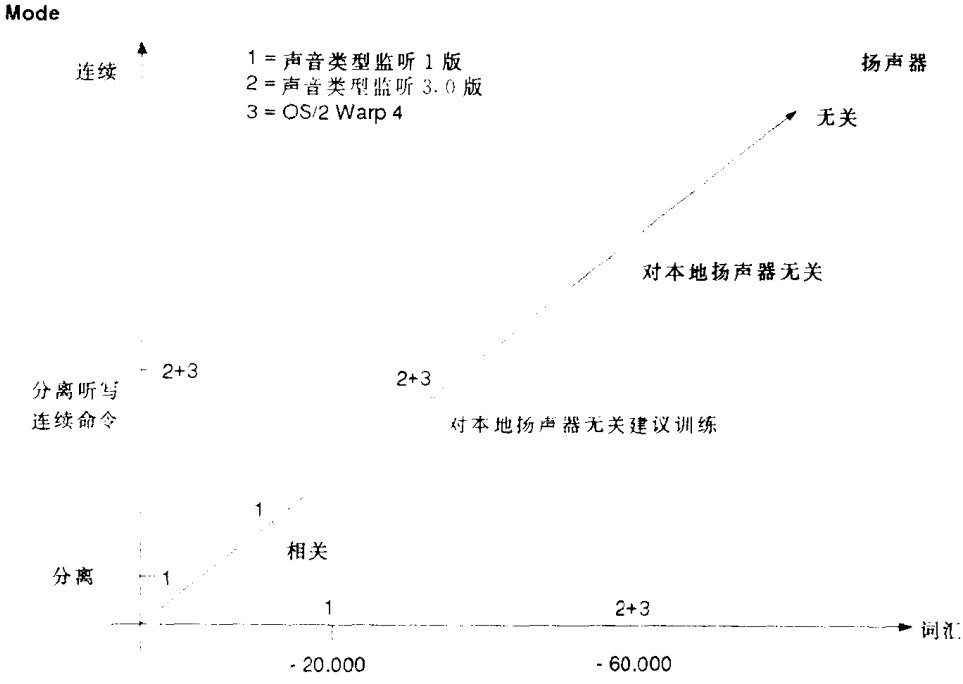

图 1　语音识别技术趋势图

图 1 说明了语音识别的趋势。当前的产品使用分离语音和连续语音两种方法。分离语音要求口述时在每个词之间作短暂停顿。VTD 的导航功能接受连续语音。IBM VTD 要求用分离语音讲话,然后稍经实践,听写的速度仍能达到很高。典型的可以达到每分钟 80 个字。

直到 VTD 3.0 版为止,语音口述仍依赖于口授者。这意味着用户必须先训练系统然后才能使用。用户使用 IBM VTD 3.0 版可以在安装该产品后即开始口述,但如果用户使用不同于原来的语调说话,系统要经过短期的训练。

实际词汇的数量十分重要。你拼出的字越少,使用 IBM VTD 越有效。最初的词汇表有 5,000单词,现在的词汇表可达 65,000 个单词。IBM VTD 3.0 版在 22,000 个单词(美国英语)到 42,000 个单词(西班牙语)之间,可以扩展到 65,000 个单词。

你可以看到,IBM 过去是并且继续是语音识别的领导者。表 1 是 IBM 与其竞争对手的产品的分析比较。

表 1　竞争分析

	IBM	Dragon	Kurzweil	Philips
一般特性				
孤立词	是	是	是	否
与说话者无关	是	否	是	是
注册	是	是	否	是
眼睛解脱	是	否	否	是
声音回放	是	否	否	是
外语支持	是	是	否	是
字宽支持	是	否	否	是
专用 DSP 辅助	否	否	是	是
导航	是	是	是	是
精确度				
启动	95%	75%	75%	90%
稳态	97%	85%	85%	95%
三字组语言模式	是	否	否	否
行业专用词汇表	是	否	是	否
邮件口授校正	是	否	否	是
性能				
通过速度(每秒字数)	70 - 100	45 - 55	45 - 55	连续的
应用				
允许动态说话	是	是	否	否

由于 30 多年研究的结果,在当今市场上,IBM 的语音识别装置有最高的准确度。IBM 最大的竞争优势之一是在识别过程中使用三字组(trigram)。一个三字组是三个字的序列。一个两字组(birram)是两个字的序列。其他的听写产品使用两字组,IBM 使用三字组的统计,如果一个字在三个字中更相似则优先选择。三字组统计对专门词汇作单独的推测,这使得 VTD 的准确度达到 90%以上。

第二章　语音输入技术

在这一章里，我们讨论语音识别技术，这是 IBM VTD 系列产品的核心。这些技术大部分是由 IBM 研究开发小组首创，并得到了世界范围的仿效。例如应用马尔科夫隐式模式（Hidden Markov modeling）来描述语言和语言的固有特性，被认为是最好的策略。感谢许多国家的 IBM 科学中心提供的民族语言支持。有七种语言的 VT 产品：美国英语、英国英语、阿拉伯语（只有适合 OS/2 的 VTD1.1 版）、法语、德语、意大利语和西班牙语。

我们从早期发表 BYTE 杂志的一篇文章讲起（1992 年 3 月 BYTE 杂志，版权属 McGaw - Hill Companies inc. 保留所有版权）。这篇文章简明地描述了语音识别所涉及的基本问题并构想出处理这些问题的技术。它也勾画了某些应用情景并展示了该领域的光辉前景。

2.1　说话

普遍存在的对语音自动识别技术的期望，令人们心动了 20 年。想象一下你的语音命令被理解执行，你的讲话被立即自动记录的情景。

今天，自动语音识别系统使任务完成得容易并且快速。处理少于几百个字的系统十分普遍，因为它的设计容易。以下介绍自动语音识别的简单配置，并引用一个例子说明你使用这一技术可以做什么。在探究了语音识别的某些技术细节以后，我们通过预示某些将来的可能性而作出结论。

2.1.1　让自动语音识别进入使用状态

图 2 展示了自动语音识别的配置。I/O 接口由通常的接口装置组成（如监视器、键盘和鼠标），为进行语音输入还应有一个麦克风。

当你对语音识别系统讲话时，它分析你的讲话，确定讲的是什么，然后在监视器上显示出分析结果。

除了直接进行语音记录外，你还可以用语音输入来控制或执行命令功能。当写入记录的内容时，你可以用声音发出命令产生一个新的段落。类似地，你也可以利用声音命令让屏幕的内容上下滚动或移动光标。

理想的自动语音识别系统在正常环境下工作时，应不出差错。但自动语音识别技术处理词汇的多少、识别准确度和其他因素受到一定限制。尽管如此，在目前技术所达到的范围内的许多应用还是给人以深刻印象。

考虑一个基于自动语音识别技术的医学报告系统。X 光医生检查 X 光片后，通常必须迅速地为委托医生和保险单携带者准备一个书面报告。通常的作法是由医生口述，然后由使用字处理器的打字员录制，这是相当耗时的，有时要几周的时间。因为懂医学专门术语的打字员的工资很高，所以花费也很多。使用自动语音识别系统，X 光医生自己就可以作出书面报告。

利用为特殊任务设计的通过口述建立起来的词汇模块，自动语音识别技术可以使医务工作者输入专门的医学术语。这种技术的另一应用是用于手忙眼忙的情况。例如在汽车里用语

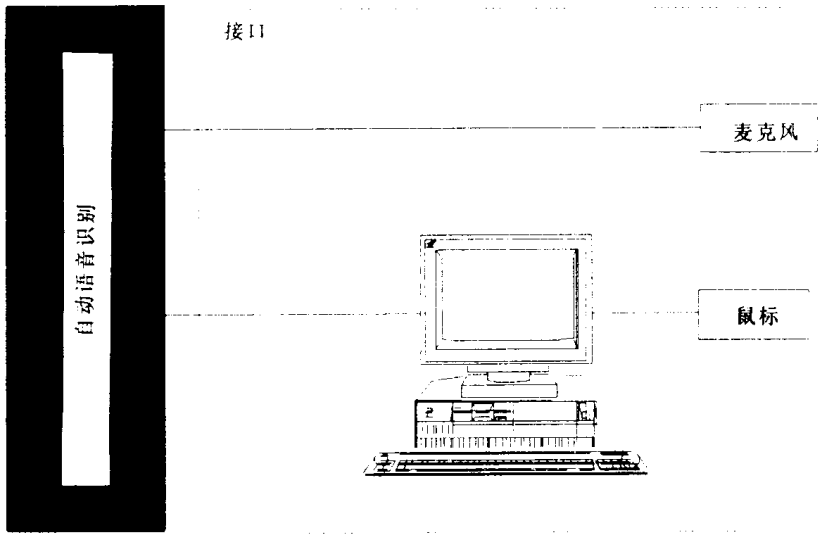

图 2　简单的语音识别配置

音拨打电话。语音输入也用来核准信用卡,查询数据库。例如查询股票行情或在一个机构中存取各种业务信息。这种询问可用电话远距离操作或在语音综合器能够使计算机响应的地方以会话方法操作。可以通过使用录制输入的语音而不需要用键盘来产生文本文件。自动语音识别也有希望使听力受损害的人能用普通电话会话。

"旅行者"是会话系统的一个有趣的试验品。它是由 MIT 小组作为高级防御研究计划而研制的讲话的系统程序。"旅行者"可以交互回答一些关于宾馆、医院、饭店及本地其他一些机构的问题。如果你问"MIT 有多远?",在回答你的问题之前,它要用语言问你当前所处的位置。

Tangora 系统是办公用的语音操作的打字机的实验样机。它是由在纽约 Yorktown Heights 的 IBM Thomas J. Watson 研究中心的由 Frederick Jelinek 领导的研究小组的开拓性工作研制出来的。它的词汇包含办公通信中最普遍使用的 2 万个单词。它的专用的 VLSI 硬卡插入 IBM AT,PS/2 或 RISC 系统 6000 的底板上。自 1987 年以后,在许多场合公开表演过。这是头一个高精度实时运行的大词汇自动语音识别系统。使用这个系统时,有一个限制,你必须在两个词的中间作短暂的暂停。这个技术表征为孤立或分离语音输入。如果允许在速度上和性能上作些牺牲,这个技术可以扩大为大词汇量或正常的连续的语音输入方法。

2.1.2　自动语音识别系统的内部结构

为了说明自动语音识别系统的构成技术,我们选择一个能处理几千个字的典型系统。这个系统的组成如图 3 所示。这些部件的功能都概括在表 2 中。以后我们再研究另一种配置。

表 2　自动语音识别系统部件词汇表

部　　件	说　　明
声学模型	将模型的小单位比如音素结合在一起,建立单字声学特点的模型
译码器	借助声学的语言模型分析矢量数字转换器的标号流以确定讲的是什么
接口	含输入输出设备,包括用户的自动语音识别系统输入用的麦克风

续表

部　件	说　明
语言模型	根据上下文制造出可能的语言
信息处理器	把语音转换成在有规则的间隔里的一组特性数字，典型的间隔是百分之一秒
矢量数字转换器	将信息处理器的输出转换成识别各种声音的标号流

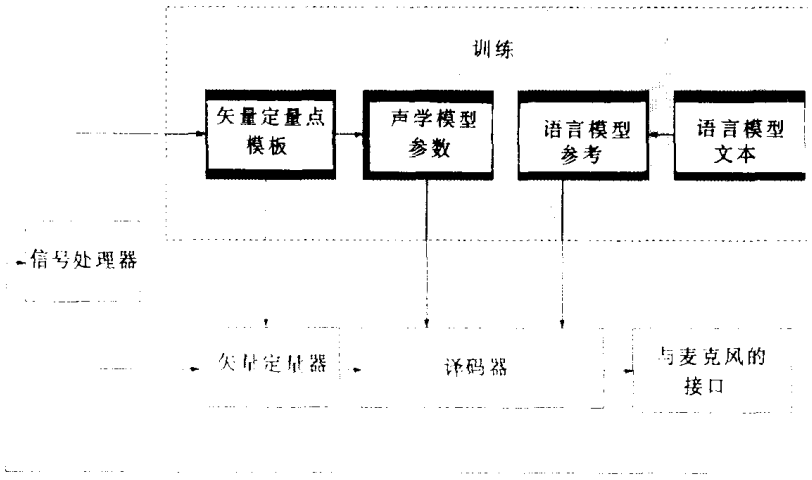

图3　自动语音识别系统的组成

在典型的自动语音识别系统里，由麦克风接受的语音输入由信息处理器分析。分析产生一组数据或特征值。用紧凑的形式表示千分之一秒的语音段。这些特征抓住了语音的重要特点(例如，不同频带的能量与你用立体声音响的均衡器控制的能量相似)。它们每千分之一秒产生出来，组成了语音参数矢量。

为了理解捕捉的特性，自动语音识别系统使用矢量数字转换器和语音模型以及通过设计和训练得到的语音模型参数。矢量数字转换器参数是不同语音的样板。这些样板大致与所讲语言的音素对应(比如 CAT 的 \ k \，\ ae \ 和 \ t \)。自动语音识别的设计者们通过研究他们想要并入系统的语音的所有的声音类型的语言参数的几千个矢量才获得这些样板。

将输入矢量与存贮的确定每千分之一秒的声音类型的样板比较，矢量转换器的输出反映了语音输入。于是，当你对语音识别系统说“CAT”时，可能产生若干个 \ k \ 标号跟随一些 \ ae \ 和 \ t \ 标号的情况。大致与你发这些声音所用的时间对应。但是因为没有人知道如何去设计一个理想的矢量数字转换器，典型矢量数字转换器的输出通常包含有与 \ k \，\ ae \，\ t \ 交错的类似的声音的标号。于是，你可能发现出现了一些 \ eh \ (像在词 pet 中)和 \ uh \ (像在词 cut 中)的标号，它们的发音与 \ ae \ 类似。此外，语音过度部分(如声音从一个音素变化到另一个音素)往往发生错误标记，因为中间语音参数矢量与前后的声音样板配不好。

更复杂的是，实际上人决不会用同样的方法对一个词或一个句子说两次。通常由于人的听觉器官的补偿，人们感觉不到这种差异。因为这个补偿装置不好理解，这些变化通常认为是听觉系统的噪音。到目前为止，处理此类噪音的最好的办法是植根于信息原理与统计方法的隐式 Markov 模型原理。隐式 Markov 模型由隐式状态(S)、状态间的隐式过渡及与状态有关的可观察输出组成(L)。在自动语音识别中，认为每个状态是由一个或多个标号输出(L)引起