分
计算机世界/２００１年/０５月/２１日/第Ｂ０３版/
专题报道

语音技术的拓展与展望

清华大学计算机系人机交互与媒体集成研究所　蔡莲红　吴志勇　王玮　陶建华　王志明

研究现状

１．语音识别获得应用

伴随着语音识别技术的不断发展，诞生了全球首套多语种交谈式语音识别系统Ｅ－ｔａｌｋ。这是全球惟一拥有中英混合语言的识别系统，能听能讲普通话、广东话和英语，还可以高度适应不同的口音，因而可以广泛适用于不同文化背景的使用者，尤其是中国地区语言差别较大的广大用户。由于Ｅ－ｔａｌｋ可以大大提高工作效率，降低运营成本，并为用户提供更便捷的增值服务，我们相信它必将成为电信、证券、金融、旅游等重视客户服务的行业争相引用的电子商务应用系统，并成为电子商务发展的新趋势，为整个信息产业带来无限商机。

目前，飞利浦推出的语音识别自然会话平台ＳｐｅｅｃｈＰｅａｒｌ和ＳｐｅｅｃｈＭａｎｉａ已成功地应用于国内呼叫中心，ＳｐｅｅｃｈＰｅａｒｌ中的每个识别引擎可提供高达２０万字的超大容量词库，尤其在具有大词汇量、识别准确性和灵活性等要求的各种电信增值服务中有着广泛的应用。

２．语音合成信息服务被用户接受

语音合成技术把可视的文本信息转化为可听的声音信息，其应用的经济效益和社会效益前景良好。尤其对汉语语音合成技术的应用而言，全球有十几亿人使用中文，其市场需求、应用前景和经济效益等可见一斑。

语音技术已逐渐在电信声讯信息服务领域智能电话查询系统中展开应用，并迅速推广。在电话高度普及的今天，如果打电话就能查询到所需信息，无疑将给人们的日常生活带来极大方便。汉语语音合成技术应用到声讯服务领域内，对现有的电话查询系统将产生革命性的影响。

语音技术与互联网已成功地结合。电话Ｉｎｔｅｒｎｅｔ网关是一种用于实现电话网和Ｉｎｔｅｒｎｅｔ网之间信息互访的系统。简而言之，就是让电话用户能够轻松地通过电话访问Ｉｎｔｅｒｎｅｔ网。系统的功能主要体现在两个方面。一方面，让用户通过电话、手机或传真机随时随地访问Ｉｎｔｅｒｎｅｔ上的各种信息，如新闻、电子邮件等，大大扩展了Ｉｎｔｅｒｎｅｔ信息的用户群和地域范围，同时大大降低了用户参与到Ｉｎｔｅｒｎｅｔ的技术难度；另一方面，能够将电话终端上信息流或控制指令发送到Ｉｎｔｅｒｎｅｔ上，例如用户可以通过电话方便地发送电子邮件和类似的留言信息，不仅具有传统的语音信箱功能，还可以将用户语音以ＩＰ的方式廉价地发送到全球任何一个电脑或电话终端上，大大

类号：ＢＪ８２

降低了信息交流的成本。利用语音合成技术的信息服务得到了用户的广泛接纳，给用户生活提供了极大的方便。

3．面向对象的语音编码

长期以来，在通信网的发展中，解决信息传输效率是一个关键问题，极其重要。目前科研人员已通过两个途径研究这一课题，其一是研究新的调制方法与技术，来提高信道传输信息的比特率，指标是每赫兹带宽所传送的比特数；其二是压缩信源编码的比特率，例如标准PCM编码，对3．4kHz频带信号需用64Kbps编码比特率传送，而压缩这一比特率显然可以提高信道传送的话路数。这对任何频率资源有限的传输环境来说，无疑是极为重要的，尤其是在无线通信技术决定今后通信发展命运的今天更显得重要。实际上，压缩语音编码比特率与话音存储、语音识别及语音合成等技术都直接相关。

语音编码技术的进展对通信新业务的发展有极为明显的影响，例如IP电话业务、实时长途翻译业务、交换机的人工智能接口等。因此，国际电报电话咨询委员会(CCITT)第15组提出了许多急需制订的话音编码标准建议，以推动通信网的发展。由于VLSI的发展，实现这一技术的代价已从在昂贵的信道中采用，发展到一般信道中都可接受的水平，因此，编码技术日益受到重视。当前，数字移动通信和个人通信(PCN)是深受人们重视的通信手段，其重要问题之一是压缩语音编码速率，形成面向对象的语音编码技术。

数字语音编码技术从1938年提出PCM开始，其编码方法已有了很大的发展，如1968年提出的线性预测编码技术(LPC)、20世纪70年代末出现的隐马科夫技术(HMM)以及矢量量化(VQ)等。

当前，语音编码技术不仅受到研究部门、应用部门的重视，而且推动了标准的制订，因为标准是工业生产的一个重要前提，对通信体制的确定有很大影响。目前，关于低速率语音编码的算法发展较快，它可应用的范围也相当广泛，人们将从中获得极大的效益。这些对推动各种通信标准及网络的建设都十分重要。

4．口语机器翻译受到重视

口语翻译的一个重要目的就是帮助聋哑人与正常人交流，近来越来越受到人们的重视。首先，聋哑人要戴上一副特制的手套，计算机根据他打出的手语进行识别，然后，通过语音合成系统就可以把图像信息翻译成语言信息。同时，系统还能够完成将正常人的语言翻译成聋哑人的手语，只要将正常人说的话键入计算机，经程序分析处理之后，翻译成有表情、有动作的三维图像，从而最终达到聋哑人与正常人之间通过翻译机进行交流的目的。口语翻译的研究在其他很多方面都有重要价值，如用手势控制计算机，甚至用手势导航等。

语音合成的最新进展

1．神经网络用于训练韵律模型

由于人工神经网络具备良好的自学习和自适应能力，将其应用于语音合成系统中的韵律模型研究具有很重要的意义。将神经网络模型与已有的文语转换系统有机结合，可以改变传统的文语转换系统的韵律模型，具有更强的适应性和可训练性，使合成语音的自然度得到显著提高，增加了系统的灵活性和风格的多样性。

2．数据挖掘用于发现语音知识

数据挖掘作为一种在大量数据库中发现隐藏新知识的计算技术方法，通过语音定性模型的建立，将数据分析和挖掘结果转化为逻辑规则或用可视化的形式进行表达。因此，将数据挖掘和人机交互接口紧密地联系在一起，将对计算机语音信号处理的研究工作产生巨大的推动力，为语音信号处理提供了一条崭新的研究途径。

3．文本-可视语音转换系统研制成功

文本-可视语音转换技术的出现是多媒体技术迅速发展的产物，也迎合了社会发展的需求。它给人们的生活增添了新的色彩，使计算机更加人性化，人们与计算机的交流变得更为简单。相信在不久的将来，它会在众多的技术、商业和娱乐领域得到广泛的应用，并逐步进入我们每个人的生活。

拓展语音计算

1．韵律研究与感知相结合

韵律是语音信号的自身属性，它反映了一个人说话时的语调高低和时间长短信息，同时反映了说话人说话时的语境信息。韵律模块也是语音合成系统中的重要组成模块，韵律参数研究的成功与否直接影响合成系统的输出。感知信息主要体现说话人对一句话中某些部分的强调和语句重音信息，语句重音也会对系统的合成输出产生很大的影响，因此，要想得到较好的语音合成效果，需要对韵律和感知进行深入的研究。

2．从语法、语义层面探索语音计算的理论和方法

语音计算中包含对语言语法、语义的理解，语音合成系统的输出不仅仅取决于语音数据音质的好坏，同时在很大程度上受到所处理文本的语法及语义现象的制约，如果没有正确的语法描述、合理地体现语义信息，就不可能产生很好的合成效果。而获得这种相互关系只有通过对大量的语言现象进行分析总结，形成规则描述。为了更加客观地进行描述，可以借助于人工智能领域里的数据挖掘方法，因此，语音计算的关键技术是挖掘语法、语义和语音之间的相互关系，采用规则描述，将这种关系结合到实际合成语音系统中，提高语音合成输出的自然度。

3．建设海量语音数据资源

语音计算的成功与否在很大程度上取决于语音资源的积累。目前，在比较先进的语音处理方法中，无一例外都提到了采用基于数据的驱动方式，然而这种方式首先就需要大量的语料数据，没有大语料，数据的驱动就无从谈起。因此，为了尽可能地覆盖各种语言现象，需要长期积累各种语音资源，同时对于语音信号的处理也需要大量的语音处理软件。这些都是日积月累的过程。

语音技术的研究方向

1．连续自然语音的识别与理解

自然语音识别与理解研究的是计算机如何理解人类的语言，其目的就是让计算机能够理解人说的话，当我们使用计算机时，只要告诉它应该做什么，它就能按照所理解的去执行。虽然现在自然语音识别与理解的理论研究得到了进一步完善，同时，计算机的功能、容量和

速度都有了很大的提高，但研究仍局限在对孤立音节的识别与理解上。人类流畅的自然发音不是孤立音节发音的简单组合，它是在一定时间范围内输出的一种连续语流，因此，需要对连续语音进行处理。连续语音识别与理解技术中需要解决的难点很多，对它的研究是语音技术今后的目标之一。

2．高自然度、具有表现力的合成语音

提高合成语音的自然度仍然是高性能文语转换的当务之急。就汉语语音合成来说，目前在单字和词组级上，合成语音的可懂度和自然度已基本解决，但是对于句子乃至篇章级，其自然度问题就比较大。未来的文语转换系统的发展趋势是采用基于语境相关的合成思想进行设计，能够将发音人的原始发音特征最大限度地保留下来，辅助以先进的层次化语言韵律模型，通过分散统计的模型方法来涵盖语义语音之间的内在联系，使系统能够输出具有高自然度和表现力的合成语音。但是，在目前的合成系统中，普遍存在合成输出语音的机器味比较浓、语境的知识层次模型研究不完善等问题。因此，获得高自然度、具有表现力的合成语音也是今后语音技术的研究目标之一。

3．语音技术与多媒体技术的结合

伴随着现代语音技术的不断发展，人类对语音信号的需要已经不仅仅停留在可懂性和正确性上，语音合成技术的研究方向已是合成语音的美感并同时输出辅助的视频特征，实现虚拟主持人的效果，通过将视觉效果包括人的头部建模、唇形同步技术和表情因素等视频信息的加入，可以更好地体现语音合成系统的表现力和感染力。因此，我们完全有理由相信，语音技术和多媒体技术的有机结合将使合成系统展现出广阔的应用前景。

4．语音技术与网络技术的结合

目前，语音技术已逐渐应用于电信的声讯信息服务领域和互联网消息收发方面。随着电话网与互联网的融合、网络信息项目的增多和时效性要求逐步提高，建立适合于股票交易、航班动态查询、电话自动报税等业务的语音系统成为可能，电话用户可以通过传统的语音、传真获取互联网上无穷无尽的信息。这些业务将彻底解决传统数字录音回放技术所无法解决的海量信息库和动态变化信息的实时生成与存储的难题，因此，将语音技术与网络进行完美的结合具有强大的生命力。

5．多语种

语言是人们交流的工具，不同民族有自己不同的语言，不同语言之间的交流在今天开放的信息社会和网络时代显得十分重要，因此，多语种的文语合成有着独特的应用价值。例如在自动电话翻译、有声电子邮件等应用中都提出了多语种语音合成的需求，即使是对汉语合成也有多方言文语转换问题。理想的多语种合成系统最好是各种语言共用一种合成算法或语音合成器，但现有的语音合成系统大多是针对某一种语言或若干种语言开发出来的，所采用的算法及规则都是与某种语言密切相关的，因此很难推广到其他语种。如汉语和西方语言之间存在着很大的差异，而目前国内的系统都是做汉语文语转换的，其韵律控制规则完全不适合于英语，而且它们主要是合成汉语普通话的，即使推广到广东话和上海话都有相当的难度。可见要真正解决多语种的文语合成，从文本处理到语音合成都必须有新的思路，因此，研制

多语种语音合成转换系统具有重要的理论和现实意义。