

谈谈汉语文语转换系统的发展

陈晓勤

摘 要 本文就汉语文语转换系统的概念、实现过程和方法作了简单的介绍,着重分析比较了波形合成、参数合成和按规则合成三种方法的优缺点以及国内一些研究单位在这一领域已经取得的研究成果,指出影响汉语文语转换系统性能的关键技术以及今后的重点发展方向。

关键词 文语转换 语音合成 波形拼接 参数合成 合成规则 韵律规则

文语转换(Text-to-Speech)是指把文本文件或文字串通过一定的软硬件转换后由计算机或电话语音系统等输出语音的过程。近年来,随着计算机多媒体技术的发展和信息产业的蓬勃兴起,文语转换系统已初步显示了其巨大的应用前景,因而也逐渐成为一个活跃的研究课题。

文语转换的基本步骤有两个,一个是文本分析即对文本进行理解,另一个是语音合成。前者是从文本中提取各种控制信息,后者是在语音信息的控制下发出语音。文本分析是要对文本进行语义分析、句法分析和词法分析,从文本中提取的信息包含有语音信息和韵律信息,它们大致相当于音标和声调符号。一般情况下,在这种分析中,需要有发音词典和一些语法规则,除此之外还要使用发音规则对语音符号略加修改,以适合由词到句子的发音,这就需要用韵律规则去临时组织语音段持续时间的长短、音调的高低和声调(基音频率)。文本中提取的控制参数用来控制语音波形的产生,在这过程中,使用语音合成器和一套合成规则把发音符号变成语音,语音段的持续时间、基音频率等这类韵律数据用来确定临时性的语音结构,并用来控制声道模型发音。所以一个文语转换系统有两个部分组成,一部分是发音器,主要是指语音合成器,它相当于人的发音系统,另一部分是发声的驱动器,它的输入是要发声的文本串或其它语言信息,而它的输出用来驱动发声器发音。

从上面的分析可以看出,文语转换系统实际上是一个人工智能系统,它既包含有很高级的信息处理,又包含发音器官复杂的生理控制,牵涉到语音学、语言学、数字信号处理、电子工程技术、生理学、计算机科学等多个领域的内容,因而要实现这种文

语转换需要用到多学科的知识和技术,难度也比较大。迄今为止,即使是对于研究历史较长的英语来说,也未能开发出一套相当满意的文语转换系统。人们正在从各个领域(特别是语音学)去探索。

一、语音合成的基本方法

根据语音生成原理,现在的语音合成方法大致可分为三种类型。

1. 波形编码合成法(Waveform Coding Synthesis)

汉语国标二级字库共有 6768 个汉字,但其对应的单音节则只有 1281 个。用这种方法可以把国标二级字库对应的 1281 个单音节录下来存放在存储器中,若有一段话或文章要让计算机读出来,只要根据内容把一个个音节连接产生语音输出。这样就实现了汉语无限词汇合成的目标,而且合成语音的清晰度可以达到比较高的水平。但是由于是一个字一个字念出来的,听起来比较呆板,即自然度差。为了克服音节拼接单调呆板的缺点,人们自然会想到如果以词为合成单元,合成效果应该会好一些。以单音节为合成单元的语音合成,其质量之所以不好,是因为语音质量受单元之间连接处的声学特性所影响,连接处的声学特性包括谱包络、幅度、基频及速率。以词为合成单元甚至以部分句子为合成单元,则可以使文本的连接处大为减少,合成语音的可懂度和自然度能大大提高。但即使这样也还有波形连接处的声学特性问题。而且由于合成单元数量较大,往往要付出比较大的存储空间。

2. 参数合成法(Parametric Analysis Synthesis)

这种方法多以音节、半音节或音素为合成单元,也有以词或词组为合成单元的。按照语音理论,

对所有合成单元的语音进行分析,一帧一帧地提取有关语音参数,这些参数经编码后组成一个合成语音库。输出时,根据待合成的语音的信息,从语音库中取出相应的合成参数,经编辑和连接,顺序送入语音合成器。在合成器中,在合成参数的控制下,再一帧一帧地重新还原语音波形。合成的主要参数有:控制音强的幅度、控制音高的基频和控制音色的共振峰参数。这种方法系统结构会复杂一些,而且合成语音的自然度稍差,但由于存储的是词或短语的特性参数,所以可以大大降低存储容量的要求。此外,单元连接处的语音特性可以通过控制特性参数来改善。这种方法存储的语音单元不是简单的原始语音,而是对词或短语进行压缩,存储的是特性参数,因此从这个观点看,参数合成法可以认为是波形编码方法的一种高级形式。

3. 按规则合成法(Synthesis-by-Rule)

按规则合成法是根据音标/音节系列,或字母系列产生任意词或句子的一种方法。在这种方法中,语音基本单元的特性参数,如音节、音素或一个基音周期的语音存储在存储器中,在生成语音时按规则把它们连接起来,与此同时,语音的韵律特性,如基音和幅度也由规则去控制。语音基本单元的质量以及控制规则,如控制信息和控制机理,在控制语音中起极其重要的作用,而且它们必须要以各国自然语言的语音和语言特性为依据。为产生自然的清晰的语音,基音、重音和频谱的过渡必须平滑,其它特性如暂停的位置和持续的时间也必须恰当。

与分析合成方式相比,规则合成方式的语音库的存储量更小,这是以牺牲音质为代价的。这种方法牵涉到许多语言学和语音学模型,系统结构复杂。目前按规则合成还不完善,合成音质一般较差。

这三种方法的基本原理可以用图 1 来表示。

这三种方法的合成特性比较如表 1。

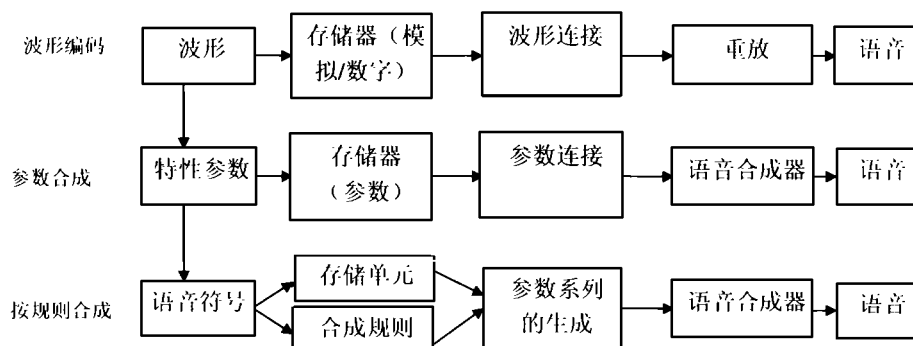


图 1 三种语音合成方法的基本原理

表 1 三种语音合成方法特性比较

特性	波形编码	参数合成	按规则合成
可懂度	高	高	中
自然度	高	中	低
词汇量	小	大	无限
位速率	24 ~ 64kb/s	2.4 ~ 9.6kb/s	50 ~ 75b/s
1Mb 内存语音长度	15 ~ 40s	100s ~ 7min	无限
存储单元	句子、词、音节	句子、词、音节	音素、音节
复杂程度	低	中	高
主要硬件	存储器	存储器、处理器	存储器、处理器

二、文本—音素转换

文本—音素的转换,首先要对文本进行理解。文本分析的目的是抽取语音信息和韵律信息。为了从文本中产生音标和每个词的音调,必须有发音词典。但发音词典不可能列出所有的数字、缩写、专有名词、新名词、复合词等,因此还需有字母—声音的转换规则,这些规则可用来对不知道的词进行发音。此外由于有多音字等问题的存在,在文本到音标和韵律的转换过程中,还需要作语法分析、句法分析、词法分析以及音调分析。

文本分析首要的工作是词的分割和词法分析。目前通常采用的分词方法有正向最大匹配法、逆向最大匹配法、最小匹配法和逐词遍历匹配法等,以前两种最为通用。为了提高分词精度,汉语分词系统都辅之以区分歧义字段的规则,这些规则要用到相应的词法、句法和语义知识。其次,要确定句子的停顿、语法轻重音和语调,必须进行句法分析,它同时能解决某些发音的歧义性。语义分析与句法分析是紧密相关的,某些句子的正确的句法分析必须依据语义知识,而语义分析又必须以句法分析为基础。

语音的韵律分量包括诸如节奏、音调和特定词的加重和去加重。韵律规则在语言学中已做了大量的研究,它似乎是语言学中最复杂的部分之一,因而也是文语转换过程中最困难的内容,它是对自然

度影响最大的因素。如何更好地把韵律规则应用到文语转换系统中,仍是一个值得进一步探讨和提高的技术关键。

三、国内文语转换系统研究现状

汉语语音合成的研究始于 60 年代,最初发

展比较缓慢,到了 70 年代后期 80 年代初期,随着计算机科学的发展,才有了较快的发展。一些专家和研究单位对汉语的文本语音转换做了大量的研究工作,取得了很大的成绩。国内对波形编码合成技术研究得比较成功的单位主要是清华大学和中国科学院声学研究所。该方法采用语音编码技术,存储适当的语音基元,合成时,经解码、波形编辑、平滑等处理输出语音。它的特点是语音基元是从自然语音中切分出来的,保留了自然语音的韵律特性,因而自然度较好。基元拼接过程中的平滑是必不可少的,它直接决定合成的音质好坏。汉语有 22 个辅音、38 个元音共 60 个音素,有学者曾以汉语音素实现文语转换。清华大学研制的 TH-SPEECH 系统采用 1200 多个音节为合成基元,标准语音经过 12 位 A/D 转换成数字量,再经 ADPCM 编码压缩成 4 位/抽样。该系统建立了语音韵律规则库,实现语流的变音、变调、轻声、儿化。对语言学处理,该系统采用正向扫描逆向极大匹配分词法(FSBM),将句子分割成词或词组串,若有歧义词串再用逆向最大词匹配法切分成词。该系统的不足是因为采用单音节为合成基元,音节之间的平滑工作还不能满足自然度的要求。80 年代末期基音同步叠加技术(PSOLA)日趋成熟,日、法、德都有相应的文语转换系统问世。PSOLA 既能保持原始发音的主要音段特征,又能在拼接时灵活调整其基频、时长和强度等超音段特征,因而成为一种比较好的波形拼接技术。中国科学院声学研究所研制成基于 PSOLA 技术的汉语文语转换系统。该系统建立了包括 1278 个单音节、350 个轻声音节、40 个儿化音节的音库,并且建立了韵律复杂库。它的合成过程是对从音库中取出的音节波形段进行基音同步调节使其达到韵律规则库确定的该音节所应具有音高、音长和音强,将调节后的音节波形拼接成输出语音。该系统优点有二:首先,系统的音库取自自然语音,保持了自然语音的音段特征,因此系统输出的清晰度高;其次,系统具有一个较完整的韵律规则,而 PSOLA 技术又使系统能根据韵律规则去改变取自音库中的音节的基频、音长、音强等韵律特征,因而在自然度上较好。该系统所需的附加硬件是一块声卡和一个音箱。但不足之处是在调整韵律时要把韵律符号置于文本文件中。

四、汉语文语转换技术的展望

1. 语音合成系统的质量指标

1992 年国际语音资料库和语音输入/输出系统

质量评测协调委员会(International Coordinating Committee of Speech Databases and Speech I/O Systems Assessment,简称 COCOSDA)成立,1993 年 12 月中国成立汉语言语输入/输出评价方法协调委员会,简称 Chinese COCOSDA,并于 1994 年和 1995 年组织了两次全国语音识别、语音合成系统性能评测。

作为文语转换系统,最重要的三个质量指标应该是准确、清晰、自然。所谓准确是指在多音字和姓氏的处理上准确无误,在切分词组上不产生歧义,使听者能完全听懂,不会产生误解。所谓清晰是指发音清楚,无干扰杂音,不会使听者听错。所谓自然是指和自然语音要接近,也就是听起来“顺耳”。1995 年 12 月国内专家对文语转换系统评测时用的质量指标是五个:音节清晰度(S)、单词可懂度(W)、单句可懂度(J)、言语自然度(MOS)和文本处理能力。前面三项按满分 100 分评估,作为清晰度测试。言语自然度满分是 5 分。文本处理能力满分为 100%。

在 1995 年这次评测中,中国科学院声学研究所的 GK TALK 汉语文语转换系统在自然度上效果最佳。该系统以语流音节为合成单元,按照中性语调模式作韵律特征调节,采用音节拼接技术,实现文语转换。基本词库有 20000 高频词和国标一、二级字库的 6768 个汉字,词库具有良好的开放性。在韵律规则上采用中性语调模式。该系统在语调的研究上取得了一定的成果。

2. 影响汉语文语转换系统质量的关键因素

在汉语文语转换系统中,自然度和清晰度已成为两个最重要的指标,也是影响转换质量的关键因素。目前各种合成语音系统输出的语音音质跟自然语言的语音音质相比还有较大差距。对于单音节的声学语言学表现与音节在词语中有什么不同,尚未找到普遍的规律。因而为了提高汉语文语转换系统自然度和清晰度,今后将在声学语言学上多作些研究。文语转换系统要涉及语句的语言学问题,不仅有句法问题,还有语义和语用问题。只有全面考虑才能对句子的轻重位置、各成分的时长分布,各种节奏停顿,即较为全面的韵律特征给出定量的结果,这样才能把包括文字串和符号串的一篇文章转换成抑扬顿挫、流畅自然的语音。

GK TALK 系统作为一个波形拼接系统,它在系统的清晰度和自然度上的效果是比较好的,而且该系统纯粹用软件来实现。在韵律规则上还可以进一步优化。作为参数合成系统,其最大好处是节省存储空间,但目前合成效果上还略逊于波形合

计算机网络规划 设计及实施方案

刘清玉

摘 要 本文介绍了铁道部第十一工程局计算机网络方案的规划、设计及实施,包括网络结构、网络设备、系统软件以及网络安全等各个方面。

关键词 网络 服务器 网络软件 互联 协议 安全

现代企业计算机网络的概念与目前各种典型的局域网络概念不同,大多数网络都是自下而上地进行网络规划和设计,而现代企业计算机网络则必须按照自上而下的网络结构进行网络整体规划与设计。因此本方案将网络作为一个整体系统进行规划和设计,使网络具有先进性、开放性、经济性、可靠性、稳定性、安全性和可扩充性。

一、基本原则

铁道部第十一工程局组建的计算机网络系统,将机关各部处室现有的微机和局属各单位局域网以及工作站进行有效的相互联接,组成一个高性能的企业级网络,由网络主机进行统一管理和调控,以实现资源共享,使企业内各类信息资源更有效地为决策者、管理人员和科研人员提供服务,实现企业内外的信息传输网络化、办公自动化、通信现代化。

我们这个网络的规划和建设本着“实用、先进、升级简便、扩充性好、开放性好”五项基本原则进行。

二、网络系统方案设计

1. 拓扑结构

网络结构见图 1。

在服务控制中心室内,以 DEC Alpha 小型机和 Intel Express 交换机构成一个虚拟网,传输速率为 200M,能够最大程度地发挥小型机速度快、功能多的特点,确保在与下面的十多个 HUB 同时通信时,不会发生阻塞的现象。同时也确保了整个系统的可靠性,即使有某个环的数据通信出现问题,系统也可自动切换至单环方式工作。

局计算机网络采用综合布线系统,主配线架与各分配线架之间采用五类无屏蔽双绞线或光纤连接,分配线架与各信息口之间采用五类双绞线连接。

在局计算机房内设光纤及五类双绞线综合配线架,交换机连接到综合配线架上,通过五类双绞线及光纤与各分配线架间的 HUB 相连,再通到各信息口上。所有的配线架都可以跳线,可根据用户需要灵活地改变跳线口,简化了管理工作。

2. 网络设备

(1) 网络服务器

网络服务器方案有两种。

小型机服务器方案。

根据国内外现代化综合信息系统建设的经验,基于性能价格比的考虑和目前铁路计算机应用现状和应用产品的实际情况,本设计方案采用 DEC 公司的 Digital Alpha Server 服务器

成。声学所的李彤、莫福源等人把文本预处理、合成规则和参数合成在芯片上得到了实现,其中参数合成用的芯片就是 Texas Instruments 公司的 LPC 参数合成芯片。而能在芯片上集成的前提是参数合成的存储量小,存储量大的波形拼接合成系统在目前的技术条件下就难以办到。这对实现技术产业化是比较有意义的。

3. 文语转换系统的发展方向

针对以上情况分析,国内对汉语文语转换系统已经做了大量的工作,而且也取得了很大的成果,尤其是中科院声学所、清华大学、中科院自动化所等单位在这一领域处于国内领先地位。文语转换这一技术在自然度和清晰度的提高上以及技术产业

化上,还有许多工作可做。估计今后文语转换系统的发展方向主要有以下三个:

波形合成尤其是 PSOLA 技术因为其在合成的自然度效果上目前处于领先地位,因而将继续得到发展。重点应该是扩大语音库中词组库的规模和完善其韵律规则库,以使自然度和清晰度进一步提高。

参数合成法因为它在存储量上的巨大优势,下一阶段将重点在提高自然度上有所发展。如果这种合成方法在自然度和清晰度上达到较高程度,应该是最具应用前景的一种方法。

文语转换技术已基本成熟,进一步发展的一个趋势是产品化。目前有一些单位已经开始作了一些工作,但仅仅是开了个头,这方面工作的潜力很大。