

汉语文-语转换系统的研究与实现*

蔡莲红 魏华武

(清华大学计算机系 北京 100084)

1993年8月5日收到

本文重点介绍一种基于语音学的分词算法和语音韵律规律的研究,同时还介绍了我们研制的无限词汇的文字到语音的转换系统,该系统可把计算机内的文本转换成语音输出。

Abstract

This paper presents our research results on a word-cutting algorithm based on phonetics and on some prosodic rules. Besides, the paper describes our unlimited vocabulary text-to-speech system which can convert computer-stored text into speech.

一、前言

随着计算机和多媒体技术的发展和 应用,计算机语音输出的需求愈加迫切,应用愈加广泛。

一般来讲,实现计算机语音输出有两种方法:一是录音/回放,二是文-语转换。若采用第一种方法,首先把模拟语音信号转换成数字序列,编码后,暂存在存储设备中(录音),需要时,再经解码,重建语音信号(回放)。录音/回放的音质好,能够保留个人的音色。但存储量随发音时间线性增长。为了减少存储容量,已研究出多种压缩编码算法。国际标准化组织(CCITT、ISO)也制定了一系列编码标准。录音/回放语音的方法应用极其广泛:数字电话、语音电子邮件、视频图象的配音和配乐、玩具、报时钟等等。

文-语转换是一种高级的语音输出系统,它应能把文本转换成连续自然的语流。若采用这种方法,预先建立语音参数数据库,发音规则库等。需要输出语音时,只要输入待发音的字符,

系统便能按语言规则,输出语声流。文-语转换的参数库不随发音时间增长而加大;但规则库却随语音质量的要求而增大。文-语转换系统可用于信息发布系统、文稿校对、盲人阅读、静态图象配音等等。

经过多年的努力,我们在微机平台上实现了汉语文-语转换系统 TH-Speech。它可将计算机内的文本,转换成声音,实时输出连续自然的语流。汉语具有复杂的音调系统,为此我们对汉字发音规则进行重点的研究和实验,还设计了基于语音学的分词算法。

二、TH-Speech 汉语文-语转换系统

TH-Speech 系统原理框图如图 1。输入的文本材料经过语言学处理、语音学处理,得到语流控制参数,然后读取语音数据库,经语音信号处理,输出连续语流^[1]。

语言学处理指的是分段、断句、分词。我们分析了文章的语言学特点,归纳出一系列处理

* 清华大学智能技术与系统国家实验室基金资助课题

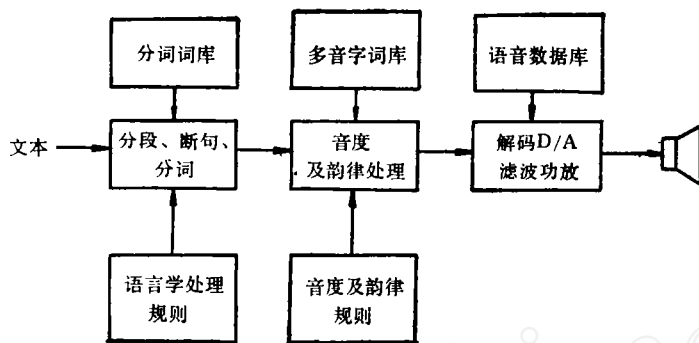


图1 TH-Speech 系统原理框图

规则,建立了相应的规则库。对文章进行分割,其目的有二:一是在输出语流中插入停顿,听起来自然,有节奏感。二是按语义修正各词发音的强弱、长短、语调的高低等。总之是让听者能正确理解其文章。

目前,TH-Speech 中,重点从语音学角度处理了停顿和发音特点问题,未进行语义理解处理。例如建立了停顿规则,停顿分三类:

1. 文本流中多于两个空格,判为标题或段结束,插入一类停顿(时间最长)
2. 文本流中遇标点符号、表格的界线,判为句结束,插入二类停顿(时间中等)
3. 词与词之间,插入三类停顿(时间最短)
4. 词内各字之间的拼接关系也可稍加调整。

有关分词的问题,下节专做介绍。

语音学处理指的是确定每个字该发什么音的处理。包括文本替换、多音字处理、变音、变调等。文本替换就是把文本用它的汉字发音字来替换,如把“=”换成“等于”,把“246”换成“贰百肆十陆”。多音字处理的目的是使一字多音的字在相应的词中得以正确发音。如“行”在“银行”一词中读“háng”,在“行走”中读“xíng”。我们的作法是建立多音字词典。分词后,到多音字词典查出相应的发音。有关其它的语音学处理请看本文第四节。

语音信号处理^[2]包括语音合成(语音解码)、数模转换、滤波、功放等。本系统语音合成基元是音节。首先我们建立了语音数据库,以

音节为单位,用 ADPCM 压缩编码。需要输出语音时,从语音数据库中取出压缩码,进行解压,再经拼接,平滑等处理。最后按句子输出连续语流。实际上,上述处理是实时完成的。

三、语音分词算法设计及实现

自动分词^[3]是汉语自然理解的基础。它把不带词边界标志的语句,自动切分成符合语言含义的词串。依照分词的目的,分词可划分为两类:一类是语法分词,应用于自然语言理解。另一类是语音分词,应用于文字-语音转换。

语音分词的目的是根据人们在表达上的习惯和语流的停顿及强弱变化,在每个词之间插入长度不等的空语音符号(停顿)。提高语流的节奏和自然度,以利于听者的理解。

建立一个好的自动分词系统,有两个关键:一是词库,二是分词算法。对于一个实时文-语转换系统来说,要求语音分词速度快,能处理歧义词串和未登录词。词库不宜过大。本系统的基本分词词库中的词条约 8 万。包括二字词、三字词和四字词。为加快检索速度,给词库建立了基本的字索引,按汉字内码排序。除基本分词词库外,还建立了几个信息表:(1)符号表:包括中英文标点和数字符号,主要用于文本结构分析(断句)和文本到语音的替换。(2)汉字拼音表:包括国标一、二级汉字的所有汉字的拼音,分为两张——非多音字读音表和多音字读音表。表内存有音调、音变等语音学信

息。(3) 语音数据库的拼音索引表: 用于由拼音序列获得语音数据。(4) 多音字词典和索引表: 包括一、二级汉字中所有多音字的非第一读音词。用于自动标注多音字读音。

基于语音分词的目的, 我们确定了匹配策略及歧义处理算法。研究表明, 最大匹配法优于最小匹配法、正向扫描划分优于逆向扫描划分。本系统采用了正向扫描逆向极大匹配分词法(FSBM)。先从句首向后扫描匹配, 将句子分隔成词语段。再处理长度大于四的歧义词串和

分词时, 未登录词常被分成单字词。若就把它当做单字词处理, 简单易行, 但对语流的自然度有影响, 考虑到汉语语流节奏性强的特点, 本系统按“二三原则”把连续单字组词。如把阿、尔、巴、尼、亚五个单字组成“阿尔”“巴尼亚”两个词, 便取得较好的听音效果。(4) 歧义词串的极大匹配最大词规则。如词语段“化合成分子”, 每相邻两字都可组成词: 化合、合成、成分、分子, 与词库匹配的最长词是“化合成”。因此这个歧义词串划分为: “化合成”、“分子”两个词。

本分词系统的分隔速度快, 每秒可分几千个词。在文-语转换系统中, 以句子作为语流输出单位。每次仅对一个句分词, 故时间短暂, 完全不影响语流输出的实时性。

四、语音韵律规则

在文-语转换系统中, 音节的易懂和自然很重要, 但更重要的是语音的韵律规则^[4]。汉语节奏性强, 音节在语流中会发生音变、调变、弱化、儿化。总结发音规律, 建立语音韵律规则库, 是改善语声自然度的关键。本系统进行的研究如下:

1 声调规则: 阴平、阳平、上声、去声为基本声调, 轻声为附加声调。在词组中, 由于音节所处位置不同, 声调发生变化, 音长、音强也会改变。因此系统中每个音节具备不止一组数据, 输出时, 根据音节在词组中的位置, 选用不同数据和参数, 使语流自然流畅。

2 变调规则: 每个音节有确定的音调。但在词组中, 由于前后音节协同发音, 从而发生了有规律的“变调”。如:

(1) 上声相连时, 前面的上声近似阳平, 三个上声相连时, 前两个上声近似阳平。如友好、管理组。

(2) 两个阴平字、阳平字、去声字相连时, 前面的音节分别变为半阴、半阳、半去。如期刊、熟食、扩大。

(3) 上声在非上声之前变为半上声。例如: 友谊, 五一, 友协。

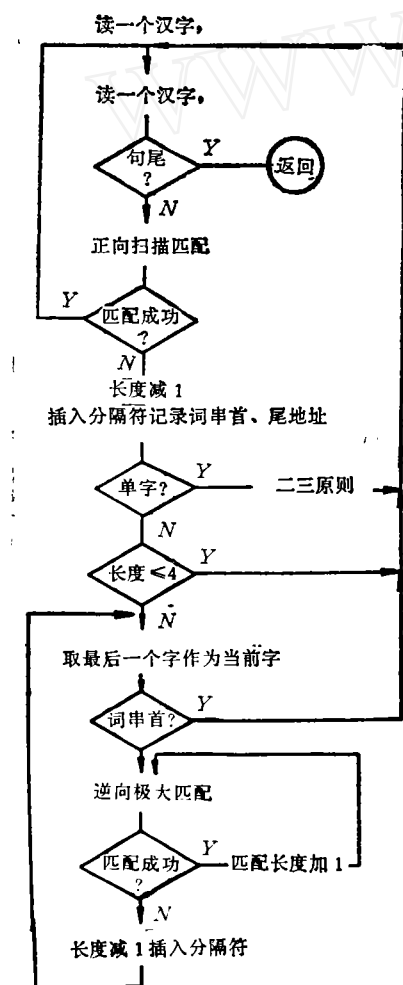


图2 分词算法流程

连续单字词。算法流程框图如图(2)。在分词处理中, 遵循一些规则: (1) 坚持将句子分成一个个词。(2) 字粘法规则, 如数词优先与其后的量词连接成词。(3) 连续单字词的“二三原则”。

(4) 单音节形容词重叠, 后一个音节形容词变为阴平。例如: “好好”读作“hǎo hǎo”; 多音节形容词重叠, 第一个音节重叠部分轻读, 后面音节变为阴平。如: “马马虎虎”读作“mǎ ma hū hū”。

(5) “一”、“七”、“八”、“不”的音调: 在去声前, 它们都读阳平; “一”、“不”在非去声前读去声, 在词语中读轻声。例如: 七上八下, 不对, 一个, 一天, 不好, 想一想, 好不好。

3 轻声规则: 轻声是语流音变中的弱化现象, 主要决定于音长, 跟语法、语义、语汇也有一定的联系。

(1) 构成助词的音节。如: 你的、笑着、走了、吃吧、行啊、轻轻地。

(2) 表示虚化语素的音节。如: 杯子、木头、尾巴。

(3) 构成方位词的音节。如: 眼里、手上、乡下、上面。

(4) 构成趋向动词的音节。如: 吃饭去、站起来。

(5) 重叠动词的后一个音节。如: 读读、看看、笑笑、听听。

(6) 重叠名词的后一个音节。如: 爸爸、妈妈、姐姐、哥哥、弟弟、嫂嫂。

(7) 大多数双音节单纯词的后一个音节。如: 萝卜、葡萄。

(8) 做宾语的人称代词。如: 打你、拉他。

(9) 某些量词和部分双音节词的后一个音节。如: 一些。

对轻声的处理办法: 上声后的轻声用阴平替代, 其它的作为独立的音节加入语音库中。

4 儿化韵规则: 一些音节词中带有“儿尾”辅助成份, 产生了几化韵。其特征是“儿”在词尾。

(1) 原韵母不变, 加卷舌“R”。

(2) 改变原韵母、韵尾脱落, 加卷舌“R”。

有时, 也可把儿化韵作为单独的音节来对待。即在四声之外增加一种调音。

五、系统的功能与特点

综合了上述研究成果, 我们实现了一个汉语文-语转换系统 TH-Speech。该系统的功能和特点如下:

1. TH-Speech 综合了语言学、语音学和语音信息处理技术。它能把计算机内的文本, 经过分词、多音字处理、音变处理, 输出高自然度的连续汉语语流。

2. 以语音方式输出包括一、二级汉字的句子或文章、英语字母。

3. 数字可以读成数码(123——读壹贰叁)或数据(123——读壹百贰十叁)。

4. 可以读各种表格。

5. 带有软中断接口, 方便地与 C.PASCAL、dBASE 等软件相连。

6. 任选男声、女声输出。

7. 利用键盘上的控制键, 随时休眠(不读)或激活发音程序。

8. 根据机器速度, 可调节音调或语流输出速度。

9. 硬件环境: PC AT 或更高档微机, DOS 3.3 以上。增加一块语音输出卡或 Sound Blaster 卡。

六、结 束 语

文-语转换系统的研究与实现是现有语音合成技术的延伸与进步。我们研究设计了基于语音学的汉语自动分词算法, 实施了音变规则, 大大提高了文-语转换语句的自然度。采用中科院声学所制定的清晰度试验词表和句表进行测试, 词清晰度为 88.9%, 句子可懂度为 94.4%。但目前该系统还没有句子语调变化功能。计算机的语音输出与人的言语过程还相差甚远, 有更多的研究课题摆在我们面前有待解决。

参 考 文 献

[1] 魏华武, 蔡莲红, “汉语普通话全音词汇智能合成系统”

DFT 用于检测未知线谱信号的应用研究

刘春跃 孙 予

(中船总七五〇试验场 昆明 650051)

1993 年 9 月 17 日收到

本文介绍了 DFT 谱估计线谱检测器的原理和 C25-DSP 系统的硬件、软件构成,分析了该系统的性能,并给出了部分实验室测试 results 和湖上试验结果。

Abstract

This paper explains the principles of a line spectrum detector using DFT spectrum estimation method and describes the hardware and software of our C25-DSP system. It also analyzes the performance of the system and presents some results obtained in laboratory and from lake test.

一、前 言

线谱检测是声纳信号处理中常用技术之一。在白噪声背景中,为检测已知频率的线谱信号,匹配滤波器(或副本相关器)加似然比检测可以构成最佳检测系统;检测未知线谱信号的最佳信号处理方式则是窄带滤波加能量检测^[1]。利用离散傅里叶变换(DFT)谱估计技术构成的线谱检测器能够实现窄带滤波加能量检测的功能,所以是在白噪声背景中检测未知线谱信号的最佳检测器,而且能完成线谱的频率、振幅的最大似然估计^[2-3]。我们设计了以

TMS320C25 为 CPU 的实时 DSP 系统,运用自适应滤波器(包括 ANC, ALE 和 ALECA)和 DFT 谱估计器技术,开展了宽带强噪声背景中未知线谱信号的检测应用研究。本文介绍 DFT 谱估计线谱检测器的原理和 DSP 系统的硬件、软件构成,分析了该系统的性能,并给出了部分试验结果。

二、原 理

DFT 谱估计是根据离散傅里叶变换的输出幅值平方求得的,见图 1 的 DFT 谱估计线谱检测器原理图。 x_i 为输入信号, X_m 为 x_i 的

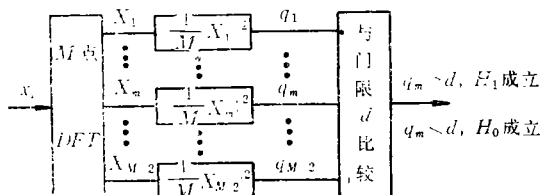


图 1 DFT 谱估计线谱检测器原理图