文章编号:1004-1478(2012)03-0034-04

Smart Voice 系统的设计与实现

徐俊芳, 田素贞

(商丘职业技术学院 计算机系,河南 商丘 476100)

摘要: 针对目前我国语音合成系统合成的语音机器味较浓、语音质量较差的问题,设计和实现了一种中文 TTS 软件系统 Smart Voice,并给出了 Smart Voice 的调用实现. 该系统使用层次化、模块化的程序结构,完成了标记分析、语法分析、延音处理、变调处理等功能. 尤其在延音处理器中,创新地采用波形拼接技术创制了波形内部循环的延音模块,使得声音在延音时具有很好的自然度. 系统运行表明,Smart Voice 发音自然,能为网络、游戏等各种计算机应用提供高速灵活的中文语音支持.

关键词: TTS; Voice XML; 延音处理器; Smart Voice 系统

中图分类号: TP391.1 **文献标志码:** A

Design and implementation of Smart Voice system

XU Jun-fang, TIAN Su-zhen

(*Dept. of Comp., Shangqiu Vocational and Tech. College, Shangqiu* 476100, *China*)

Abstract: Aiming at the problem that the current speech synthesis systems has strong taste of speech machine and poor voice quality, Smart Voice, a kind of Chinese TTS system was designed, and the Smart Voice call implementation was given. The system used a hierarchical, modular program structure, completing the marker analysis, syntax analysis, sustain treatment, tone sandhi processing function. Especially in the sustain processor, waveform concatenation technology innovation was used, creating waveform internal circulation damper module to make a good degree of natural sustain. System operation results showed that Smart Voice sounded natural, and could provide high-speed flexible Chinese speech for the network, games and other computer applications.

Key words: TTS; Voice XML; sustain processor; Smart Voice system

0 引言

TTS(text to speech)即语音合成,又称文语转换技术,它可以在任何时候将任意文本转换成具有高自然度的语音,是中文信息处理领域的一项前沿技术. 早期的语音合成技术主要采用参数合成方法. 1980年代末期提出了基音同步叠加(PSOLA)方法,大大提高了基于时域波形拼接方法合成的语音的音色和自然度. 近些年来,在国家"863"计划智能计算机主题的支持下,汉语 TTS 技术有了长足的进步,从参数合成到拼接合成,最后到两者的有机融合,虽然目前语音合成技术已走向实用,但这些系统合成的句子及篇章语音机器味较浓. 汉语是一种非常复杂的语言,现有的汉语语音学规则很难完善

收稿日期:2011-11-28

作者简介:徐俊芳(1968—),女,河南省商丘市人,商丘职业技术学院副教授,主要研究方向为计算机应用、数据库、图形图像.

和精细，因此以此为基础开发出来的 TTS 系统的语音质量与实际应用要求距离较远[1].

目前，TTS 系统除了需要提高合成语音的自然度之外，还需降低语音合成技术的复杂度. TTS 系统一般需要几兆到几十兆字节，甚至几百兆字节的存储容量，对于像 HPC，PDA，无线通信手机及商务通等资源有限的设备就无法承受，所以需要减小语音库容量[2-4]. 为了解决以上问题，本文拟设计与实现一套中文 TTS 软件开发系统 Smart Voice，克服以往语音软件的诸多缺点，并提供如实时变调效果功能，以期为网络、游戏等各种计算机应用提供高速灵活的中文语音支持.

1　Smart Voice 的系统结构

Smart Voice 是一套主要针对网络和电脑游戏等交互式应用的程序接口和工具集，核心是汉字语音智能合成引擎，其内部逻辑结构如图 1 所示. 使用者通过调用系统所提供的函数输入句子，Smart Voice 经过分析合成等一系列工作后将输入句子的语音输出.

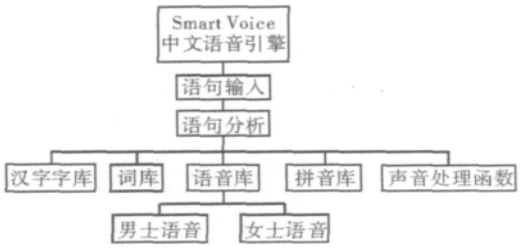

图 1　Smart Voice 中文语音引擎的内部逻辑结构示意图

2　模块分析

Smart Voice 使用层次化、模块化的程序组织，具有输入、分析、合成、输出 4 个层次，如图 2 所示. 模块之间相互依存又相互独立，因此在开发时具有很好的分立性，在保持输入输出接口不变的情况下可以分别进行设计、修改.

2.1　输入输出接口

Smart Voice 本质上是一个工具，是提供给软件开发商使用的，因此需要有一个良好的软件接口. Smart Voice 的文本输入接口有 3 个，即 Visual Basic 的动态链接库、Visual C + + 的动态链接库和 Internet Explorer 的网页接口.

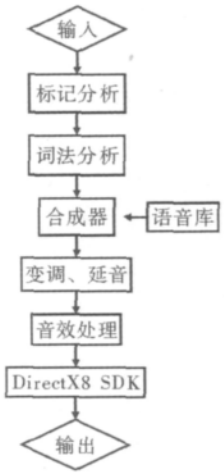

图 2　Smart Voice 的软件层次结构

2.2　标记分析器和词法分析器

Smart Voice 将输入的语音超文本标记进行分析，抽取出语音控制参数传递给下一个模块. 词法分析器则将字词之间的间隔进行划分，同标记分析器的参数输出一起送到合成参数输出器中.

2.3　合成参数输出器

该模块将标记分析器和词法分析器传递过来的声音控制参数进行综合分析，同时将文字同语音索引对应起来，并将参数控制和索引序列传递给下一个环节.

2.4　语音库

语音库包含了大约 400 个语音样本和约 1 200 种变音模式. 合成参数输出器的输出索引在这里被用于提取语音样本，传送给下一个环节通过处理. 目前绝大多数的语音合成引擎到这一步就把声音输出了. 而 Smart Voice 创新地将输出通过 DSP 进行延音和变调处理.

2.5　延音处理器

传统的波形拼接技术的延音处理能力很差，要么就没有延音，原始波形是什么就回放出什么，要么就是采用元音、复音拼接过渡的方法进行合成[5-7]. 而同一个字的发音由不同的元音、复音组合而成，因此传统的延音处理过渡很不自然. Smart Voice 的延音模块采用波形内部循环的方法合成延音.

以“王”字的发音为例. “Wang”的音色样本波形如图 3 所示. 将阴影部分的波形放大可以看到图 4 所示的波形最后部分. 再进行放大可以看出这段

波形是由一段简单的波形循环而成的,称为 loop,如图 5 所示.

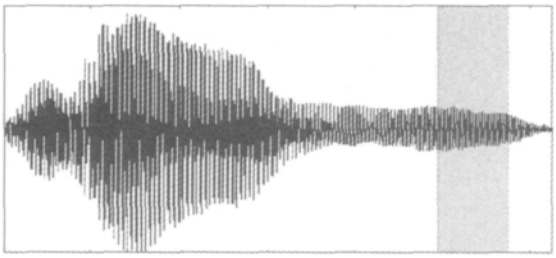

图 3 "王"字的波形样本

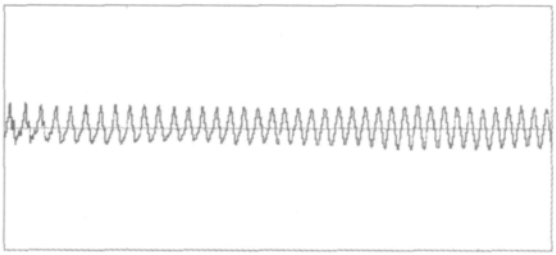

图 4 "王"字的波形的最后部分

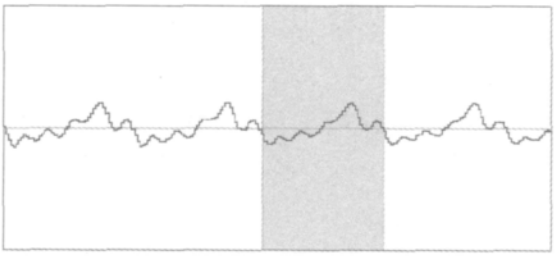

图 5 loop 的波形

显然,如果对每一个波形的 loop 进行标记,当需要延音时,在波形回放结束之前适时地进行 loop 循环回放并加入幅值衰减的包络线,则会产生逼真的延音效果. 由于语音由同一个波形合成,因此声音具有很好的自然度.

2.6 变调处理器

传统的语音合成软件利用多存放几个波形库的方法进行合成,因此只能以特定的某几个音调进行朗读. 而 Smart Voice 的变调处理器对波形进行 DSP 处理,具备无极变调功能,灵活性较高.

将波形进行降调,本质上是使得合成该波形的各个正弦波的周期成倍扩大. Smart Voice 的变调处理器模块也是基于这个原理,将上一个模块输出的数据根据合成参数进行实时拉伸或压缩,达到无极变调的效果. 以"王"字的发音为例,降调后的波形如图 6 所示.

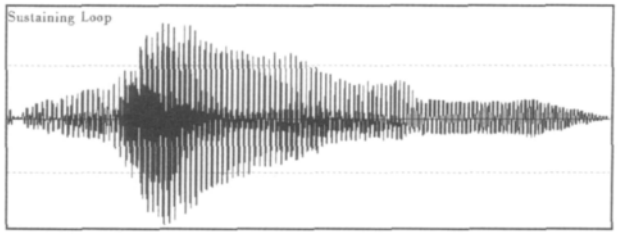

图 6 "王"字降调后的波形

2.7 音效处理器

音效处理器是语音输出前最后一个处理环节,它将波形加入混响、合声等的效果,使得前几个模块对波形处理后产生的毛刺、剧烈的过渡变得平滑自然. 如果没有这个模块,合成出来的声音就会发干,没有丰润感.

3 Smart Voice 的调用实现

Smart Voice 编程采用 Voice XML,声音内核为 DirectX 的 Direct Sound. Smart Voice 安装后会存在以下几个核心文件: SVBank. dat(语音库),SVCentre. exe(后台运行程序),IEHelper. dll(嵌入到 IE 中分析 VXML 的动态连接库),SmartVoice. dll(核心 DLL),ReadVXML. dll(方便调用后台程序朗读的 DLL). Smart Voice 有如下 2 种工作方式.

1) 直接调用 Smart Voice. dll 进行朗读. 进入 VB 的菜单 Project→References,选择 SmartVoiceWithDifWords 部件. 接下来声明 Smart Voice 对象

```
Dim sv As New Smart Voice With Dif Words. SVobject
```

在用户的程序中首先要初始化

```
sv. SetWorkDir( dir as stirng)    /设定工作文件夹,默认是 sys 目录,如果没有改变,则不用调用这个方法
sv. ObjectInit ( Voice_REVERB)    /其中可以跟 4 个参数,分别是 3D、混响、单声、立体效果
```

初始化完成后就可以使用了. 以 Vol 的音量,Tone 的音调,持续 LastTime ms 读一个字: InWord

```
sv. ReadOneWord( InWord as stirng,Vol as byte,Tone as byte,LastTime as integer)
```

直接以默认的参数朗读一句话

```
sv. ReadSentence( Sent as string)
```

设定默认参数的值

```
sv. SetReadParameter( …)
```

获得一个参数的值,如 Tone,Vol 等

```
sv. Get…( )
```

设定等待时间

sv,Delay(integer)

最后使用完要销毁

sv. SV_Desdroy( )

2) 调用 ReadVXML. dll 朗读 VXML. 前提是后台监控程序 SVCentre. exe 正在运行.

在工程中引入 ReadVXML. dll 的接口.

声名变量

Dim svr As New SVReadVXML. ReadVXML

函数也只有一个

svr. ReadV XML(str as string)　/要朗读的标记过的句子

4　结论

本文设计和实现了一种中文 TTS 软件系统 Smart Voice. 该系统主要由标记分析器、智能词法分析器、合成参数输出模块、延音处理器、变调处理器、音效处理器等模块组成. Smart Voice 采用波形拼接技术创制了波形内部循环的延音模块; 对波形进行 DSP 处理,使得变调处理器具备无极变调功能,灵活性较高. 另外 Smart Voice 内核基于 DirectX 编制,对计算机硬件进行直接存取操作,速度较快,并采用 Voice XML 语言,通过对输入语言的标记来进行对发音的语气、音调、重音等参数的控制,使发音更自然、灵活、多变. 系统运行表明,Smart Voice 能为网络、游戏等各种计算机应用提供高速灵活的中文语音支持.

参考文献:

[1] 刘浩杰,杜利民. 语音合成技术的发展与展望 [J] . 微计算机应用,2007,28(7) : 726.

[2] 陶建,蔡莲红. 语音合成的应用系统设计 [J] . 计算机世界,2001(12) : 52.

[3] 姚淑珍,陆文秀. TTS 中文语音合成技术的研究与实现 [J] . 天津科技大学学报,2004,19(1) : 65.

[4] 刘涛,叶振兴,蔡莲红. 嵌入式汉语 TTS 系统的设计与实现 [J] . 中文信息学报,2004,18(3) : 36.

[5] 陶建华,蔡莲红. 汉语 TTS 系统中可训练韵律模型的研究 [J] . 声学学报,2001(1) : 67.

[6] 杨鸿武,蔡莲红,陶建华. 屏幕文本的语音合成 [J] . 计算机应用,2002(5) : 132.

[7] 苏庄銮. 情感语音合成 [D] . 合肥: 中国科学技术大学,2006.