

文章编号: 1672 - 058X(2009)03 - 0247 - 05

智能语音教学系统的研究和实现

王 行 荣

(重庆工商大学 计算机科学与信息工程学院, 重庆 400067)

摘 要: 介绍了语音合成 (TTS) 技术的最新进展; 分析了现有语音合成技术用于课堂教学的几种方法; 研究了 TTS 用于教学的关键技术, 解决了语音合成技术用于课堂教学时讲解语句显示与声音同步的问题, 并在此技术基础之上, 用 VB. NET 实现了一个通用的智能语音教学系统。

关键词: 语音合成技术; 智能; 语音教学系统

中图分类号: TP311

文献标识码: A

人们一直希望实现完全意义的人机对话, 即计算机能“说话”, 同时能听懂人们的说话。计算机“话”相对容易。经过不断研究, 目前并联共振峰合成、串联共振峰合成^[1,2]、线性预测合成等技术, 特别是基音同步叠加技术 (PSOLA)^[3-5] 的应用, 通过对人发出的原声进行波形拼接^[6,7], 把文本转换为相应语音输出, 其声音基本上和普通人说话一样。

计算机语音合成即 TTS (Text to Speech), 指计算机将随机产生的或外部输入的文字信息转变为人可以听懂的、较流利的口语输出。语音合成是计算机应用的重要领域, 其研究涉及计算机科学、语言学、语音学、人工智能、心理学等多种学科^[8]。

实现计算机语音输出一般有 2 种方法: 一种是录音/重放, 另一种是文字-语音转换。录音/重放方法是原语音的重现, 可获得高质量的声音, 但它不能满足实时修改发音内容的需要, 而且所需的存储容量会随发音时间而线性增长。文字-语音转换方法可以根据给定文字即时发出语音, 文字转换的参数库不会随发音时间增长而加大, 但是要寻找合适的参数和生成算法, 并合成极似人类发音的、自然流畅的语音难度非常大。

一般来说, 录音/重放技术常应用于一些不需要语言理解和随时更新的系统中, 如航班信息发布、语音报时、汽车报站、电话语音服务等系统。而文字-语音转换系统是基于语音合成技术实现的文字到语音的即时转换, 能适应各种需求, 应用领域极为广泛。

1 计算机辅助教学 CAI 中的语音合成

随着现代教育技术水平的提高, 以及多媒体教学设备的大量使用, 许多教师都采用了 CAI 技术和多媒体教学手段, 其中不可避免地要碰到输出语音的问题, 比如在利用 PowerPoint 电子演示文稿软件放映幻灯片时, 有时就需要输出语音。利用 PowerPoint 自带的“录制旁白”的方法, 将教师的语音保存在计算机中, 放映

收稿日期: 2009 - 04 - 02; 修回日期: 2009 - 05 - 20。

作者简介: 王行荣 (1963 -), 男, 重庆市开县人, 硕士, 讲师, 从事人工智能与计算机应用技术研究。

时同步播出,这样做虽然语音质量较好,但是一是录制旁白比较麻烦,花费时间较长,对录音时的环境要求较高,二是需要较大的存储空间,不利于课件的传送,三是内容不容易更新。因而,将计算机语音合成技术应用到多媒体教学中,让计算机自己按照教师的要求即时“读”出语音,是非常有意义的。

2 用于 CAI 中的语音合成技术

要在 CAI 中实现语音合成,可以采取直接调用 TTS 软件合成语音的方法,也可以在 CAI 课件中嵌入语音。TTS 软件可以将输入文本直接转换成语音输出,而且可以控制语音的语速、音量、暂停、继续播放等,操作简便。在课件中嵌入语音,可以利用 Microsoft Office XP 自带的语音合成引擎,经过简单设置后实现指定内容的语音输出。

2.1 利用 TTS 软件直接输出语音

自 20 世纪 60 年代首个英语 TTS 系统研制成功以来,迄今世界上已推出多种语言的 TTS 系统。近年来我国也研制出许多优秀的汉语 TTS 软件,都较好地实现了输入语篇后,及时转换成语音的任务,而且这些系统合成的汉语语音的可懂度、清晰度都达到了较高的水平。

在这样的软件环境下,教师可以直接利用现有的 TTS 软件实现语音输出。在 Microsoft Speech SDK 5.1 下载页面上,包括了几个独立的供下载的文件。其中 SpeechSDK51.exe 文件提供了英语语音引擎,SpeechSDK51LangPack.exe 文件提供了汉语语音引擎,下载并安装后,运行该软件 TTS App,在 Enter text you wish spoken here (在此输入你想朗读的文本)文本输入框中,输入需要语音输出的文本内容,单击右边的“Speak”按钮,就可以听到计算机合成的语音输出,详细操作参见文献 [9]。

利用 Microsoft Speech SDK 这个 TTS 软件来进行计算机语音合成,因为需要在教师机上下载相应的软件并进行安装,另外输入的文本篇幅要求小且不能保存,所以在使用上有局限性。

2.2 SAPI 5.1 SDK 浅析

SAPI SDK 是微软公司免费提供的语音应用开发工具包,这个 SDK 中包含了语音应用设计接口 (SAPI)、微软的连续语音识别引擎 (MCSR) 以及微软的语音合成 (TTS) 引擎等。SAPI 中包括对于低层控制和高度适应性的直接语音管理、训练向导、事件、语法编译、资源、语音识别 (SR) 管理以及 TTS 管理等强大的设计接口。其结构如图 1:

语音引擎则通过 DDI 层 (设备驱动接口) 和 SAPI (SpeechAPI) 进行交互,应用程序通过 API 层和 SAPI 通信。通过使用这些 API,用户可以快速开发在语音识别或语音合成方面应用程序。

2.3 导入 COM 对象到 .Net

SAPI 5.1 是基于 Windows 平台的,通过 COM 接口进行调用。在 .Net 平台下要应用 SAPI 5.1,可以利用



图 1 SAPI 5.1 结构图

.Net Framework 自带的强大工具 TlbImp.exe 来把 SAPI SDK 的 COM 对象导入到 .Net 中。TlbImp.exe 产生一个管制的包装类,管理客户端可以使用它,包装类管理实际的 COM 对象的参数。当包装类当作收集的垃圾时,包装类释放掉它包装的 COM 对象。当然,也可以在 VS NET 环境中通过从项目参考对话框选择 COM 对

象,实现 COM对象的导入,这个过程也是通过 Tlb Imp. exe来完成的。

3 智能语音教学系统实现

3.1 系统的基本功能

智能语音教学系统能够根据课堂内容的设置,对教学内容进行自动的语音讲解,大大减轻教师上课时候的工作负担,将教师上课时候的教学实施工作转换为以教学组织工作为重点的课堂教学模式上来。

系统使用 VB. NET进行开发,用 NeoSpeech Lily语音引擎对讲解内容进行自动合成。系统讲解语音清楚、语言较为流畅,能够智能显示讲解的内容,根据不同要求可以方便地对语速、音量、语句停顿时间、语音角色等进行调整。系统包含备课子系统和语音讲解子系统。

3.1.1 备课子系统

目前系统讲稿内容支持 txt和 doc两种常见文件格式,可以使用写字板、记事本、Office Word等编辑工具进行讲稿的编辑工作,同时也可以直接使用系统自带的编辑工具编辑和制作讲稿。

一般 TTS系统虽然也可以取得较好的语音效果,但是对教学中的专业词汇有时还是不能够生成完全正确的语音内容,特别是对多音字的读音还无法做到完全正确,在备课子系统中提供多音字的“标音”功能,可以根据需要选择恰当的方式进行语音合成,得到完全满意的语音讲解。

3.1.2 讲解子系统

系统可以以窗口方式和全屏方式进行讲解。在窗口方式进行讲解时,教师可以灵活地进行控制,包括实时进行语音速度调整、音量大小调整等操作;在全屏方式讲解时效果类似与 PowerPoint的演示效果,不同的是自动进行讲解,讲解时能自动同步高量显示讲解的语句内容。

3.2 系统中的 TTS实现方法

智能语音教学系统的核心是 TTS,语音合成的效果起着最为重要的作用。目前实用的语音合成系统中,发音效果好的有安徽科大讯飞信息科技股份有限公司的 InterPhonic 6.0和美国的 NeoSpeech Lily语音引擎,它们的发音都接近自然人的效果。由于 NeoSpeech Lily的语音效果好,并且有更好的通用访问接口,系统采用了 NeoSpeech Lily语音引擎。

3.2.1 语句语音朗读

系统按照教师的备课内容自动进行讲解,也就是要求系统按照选择的讲解内容实时进行语音合成并发音,系统通过函数 txtToSound实现:

```
private void txtToSound (String outTxt) {  
    try{  
        SpeechVoiceSpeakFlags SpFlags =  
            DW_SpeechVoiceSpeakFlags_SV_SFlagsA_sync;  
        SpVoice Voice = new SpVoice();  
        Voice.Speak(outTxt, SpFlags);  
    }  
    catch (Exception er) {  
        MessageBox.Show("An Error Occured!", "SpeechApp", _
```

```

    MessageBoxButtons OK,MessageBox Icon Error);
}
}

```

3.2.2 语句自动定位

语音教学系统中,不但能够对教学内容进行讲解,还能够用高亮颜色显示当前讲解的语句的内容,保证讲解和输出显示中内容的同步性。

系统中,语句内容的定位由语句的起始点 `beginSeat`和终止点 `endSeat`确定,每次从上一个语句后开始查找下一个语句,查找的结束条件是语句停顿符号,包括标点符号、空格、换行符、文件结束符,系统中通过 `findEndSeat()`查找语句的结束位置。

语句高亮度显示设置通过下面 3 个语句实现:

```

viewTextBox.SelectionStart = beginseat + 1
viewTextBox.SelectionLength = endSeat - beginSeat
viewTextBox.SelectionColor = Color Blue

```

3.3 多音字标音处理

由于汉字中有大量的多音字,占汉字总数的 25.68%^[10],系统对多音字提供标音处理接口,教师在备课时对专业术语中的多音字进行标音处理。

首先,建立多音字表;之后,当程序运行时,先判断是否为多音字;如果不是多音字,按照正常 TTS 转换输出语音,如果是多音字,按照标音规则将语句改为无多音字的语句,再调用语音输出函数进行处理。系统中,用鼠标选择多音字,按“标音”按钮,可以选择该字在指定点处的发音方法,系统在讲解时就按照指定的方式发音。

3.4 语言控制参数设置

设 `objVoice`为 SAPI SpVoice对象,通过 `CreateObject`创建。则其成员:

`objVoice.rate = velocity(0)` 速度的取值范围是: -10 ~ 10

`objVoice.Volume = soundVolume(0)` 声音音量取值范围是: 0 ~ 100

`objVoice.Voice = objVoice.GetVoices(Role).Item(0)` 设置读音者,可选择的范围与系统中第三方 TTS 软件的安装有关。

其中设置 `Role = "Name = VW Lily"`,则可以得到 `Neo Speech`的 Lily 读音,若设置 `Role = "Name = Microsoft Simplified Chinese"`,则可以获得 `Microsoft`里的中文男声读音。

系统中通过菜单中的“设置”选择读音者,通过 `Trackbar`的拖动设置语音速度、音量大小和语句停顿时间,系统运行过程如图 2。

4 结 语

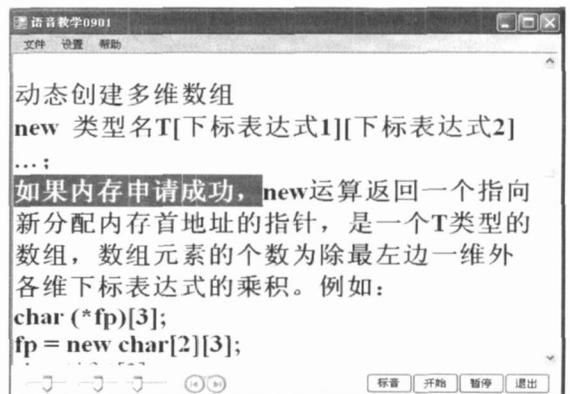


图 2 语音教学过程

计算机语音合成技术在计算机辅助教学 CAI 中具有

很大的作用。教师可以通过专门的 TTS 软件实现语音输出,也可以在 Microsoft Office XP 中实现语音输出。利用所述方法,教师只需通过简单的设置,对课件中专业术语的多音字进行语音标注,即可完成 CAI 课件中的语音合成并用系统进行自动讲解。

无论用何种方法实现语音输出,都存在合成语音的质量问题。合成语音的质量取决于语音合成引擎的质量。计算机语音合成的质量不够高、自然度和流畅度不够好等,是语音合成研究领域仍在努力攻关的难题。随着计算机语音合成技术的发展,合成语音的质量、自然度和情感表现力将会不断地提高。相信在不久的将来,合成语音能像人类发音一样自然流畅,随着计算机技术的不断发展,一个拥有流畅语音的智能教学系统必将会出现。

参考文献:

- [1] 张家驷. 论语音技术的发展. 声学学报 [J]. 2004, 29(3): 193-199
- [2] 张蒙,陶建华. 基于共振峰增强训练算法的 HTS 研究与改进 [J]. NCMMSC, 2007(4): 315-350
- [3] 李锐. 基于 PSOLA 语音合成技术在雷达模拟机上的应用 [J]. 中国民航飞行学院学报, 2004(2): 11-14
- [4] CHARPENTIER E, MOULNES E. Pitch - synchronous wave form processing techniques for text - to - speech synthesis using diaphones [J]. Speech Communication, 1990(9): 453-467
- [5] MOULNES E, LAROCHE J. Non-parametric Techniques for Pitch-Scale and Time-Scale Modification of Speech [J]. Speech Communication, 1995, 16: 175-205
- [6] 郭兰英. 汉语语音拼接技术的研究 [J]. 计算机应用与软件, 2005(11): 49-50
- [7] 刘浩杰. 语音合成技术的发展与展望 [J]. 微计算机应用, 2007(7): 726-730
- [8] 廖正和. 浅析语音合成技术 [J]. 科技情报开发与经济, 2006, 16: 216-217
- [9] 吴稟雅. 计算机语音合成在 CAI 中的应用 [J]. Computer Era, 2007, 20: 76-78
- [10] 周海涛. 语音合成中多音字识别的实现 [J]. 科技资讯, 2008(11): 241

Research and realization of intelligent voice teaching system

WANG Xing-rong

(School of Computer Science and Information Engineering,
Chongqing Technology and Business University, Chongqing 40067, China)

Abstract: This paper introduces the brand-new progress in voice combination technology, analyzes several methods to use present voice combination technology in classroom teaching, studies key technology of TTS in teaching, solves the problem of synchronism between sentence display and sounds in TTS, on the basis of this, uses VB.NET to realize a general intelligent voice teaching system.

Key words: TTS; intelligence; voice teaching system

责任编辑:田 静