

文章编号: 1005—0523(2009)02—0086—05

基于 TTS 和 SR 技术的英语自学系统设计

李广丽, 张红斌

(华东交通大学 信息工程学院, 江西 南昌 330013)

摘要: 文语转换技术和语音识别技术是实现英语自学系统的关键技术。介绍了英语自学系统的工作原理和软件结构, 并在 .Net 环境下导入了语音应用程序编程接口 The Microsoft Speech API 利用该接口编程实现了朗读训练、听写训练和阅读训练等功能。系统综合运用了声音和动画的多媒体手段, 获得了良好的人机交互效果。

关键词: 英语自学系统; 文语转换; 语音识别; the Microsoft Speech API; .Net

中图分类号: TP391

文献标识码: A

目前, 很多英语学习者开始使用软件辅助英语学习。这些软件借助计算机技术和多媒体技术使得学习过程更加规律、科学、轻松、寓教于乐, 更能激发学习者的学习热情。文语转换 (Text To Speech 简称 TTS) 技术和语音识别 (Speech Recognition 简称 SR) 技术是实现英语自学系统的技术保障。

众所周知, 英语学习强调“听”、“说”、“写”三个方面的技能训练, 而近年英语口语学习更是成为英语学习的热点话题, 在传统的英语自学系统中口语训练几乎为空白, 因此, 文中的英语自学系统综合运用 TTS 和 SR 技术在完成听写、阅读等传统功能的基础之上, 以朗读训练为突破, 全面实现“听”、“说”、“写”等英语技能训练。

软件开发环境 Microsoft Visual Studio 2003 (Microsoft .Net Framework SDK 1.1), 数据库 SQL Server 2000, 开发平台 Windows XP/2000。语音开发组件选择 Microsoft Speech SDK 5.1。

1 英语自学系统的工作原理及软件结构

声卡是实现文语转换和语音识别的主要硬件设备。语音编程 API 则是实现文语转换和语音识别的软件接口^[1]。微软开发了一套基于 Windows 环境的语音编程专用组件 Microsoft Speech SDK 5.1, 该组件包含了语音识别和语音合成 (Speech Synthesis 简称 SS 即文语转换) 两大技术, 它的应用程序编程接口称为 The Microsoft Speech API (简称 SAPI), SAPI 屏蔽了底层的硬件操作细节, 提供了面向不同程序开发环境的 SS 和 SR 编程接口, 实时控制 SS 或 SR 语音引擎完成语音编程^[2]。因此, SAPI 的应用减少了程序员的工作量, 提高了程序开发的效率。图 1 是 SAPI 组件结构。

收稿日期: 2009—01—16

基金项目: 江西省教育厅基金 (赣教技字 [2006]177 号); 华东交通大学科研基金 (08XX05)

作者简介: 李广丽 (1978—), 女, 讲师, 硕士, 研究方向为 WEB 数据挖掘、多媒体技术。

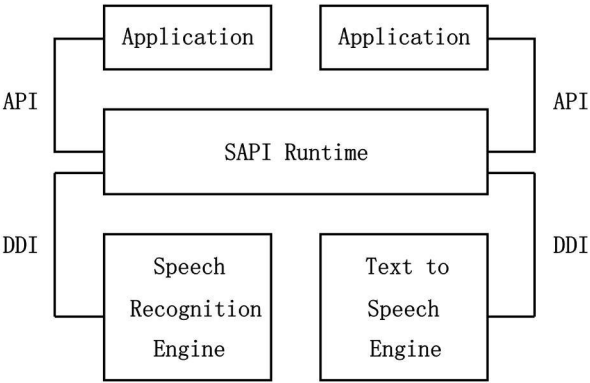


图 1 SAPI组件结构

英语自学系统的软件结构如图 2所示.

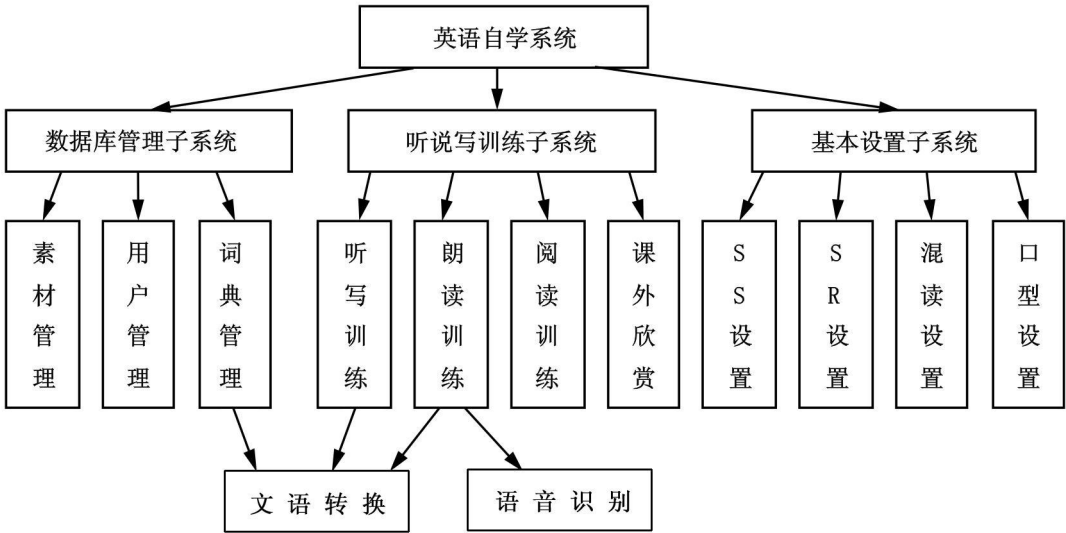


图 2 英语自学系统的软件结构图

英语自学系统收录了各个类别的英文单词、句子、文章,并组织成词典数据库。此外,系统还收录了经典电影对白、演讲片断、广播播音等音视频素材文件。听写训练过程中转换听写内容为语音信息输出,供训练者听写;朗读训练过程中接收训练者的语音信息,并判断其是否正确,然后转换语音信息为文本信息输出;阅读训练提供阅读、翻译和作文等训练。基本设置子系统设置不同的朗读者、朗读语速、朗读音量和朗读口型等信息。

2 英语自学系统的实现

2.1 SAPI接口调用

SAPI基于 Windows平台,通过 COM (Component Object Model)接口进行调用^[3]。在 .Net平台下要应用 SAPI可利用 .Net Framework自带的 TlbImp.exe工具,把 SAPI的 COM对象导入到 .Net中。导入 SAPI后,只要在引用中的 COM 页面添加 Microsoft Speech Object Library,程序员就可进行语音编程了^[4]。

若要在 .Net中引用 SAPI接口,必须声明一个调用该接口的对象: public static SpVoiceClass voice=new SpVoiceClass(),然后操作该对象完成语音编程^[5]

表 1 SAPI引用对象 voice的主要结构

接口	属性 函数	描述
Volume	属性	文 语转换时的语音音量大小
Rate	属性	文 语转换时的语音音速
Voice	属性	文 语转换时的朗读者
V iseme	属性	文 语转换时的朗读口型
Speak()	函数	完成文 语转换的函数
GetVoices()	函数	获取文 语转换时的朗读者
Resume()	函数	继续文 语转换
Pause()	函数	暂停文 语转换

2.2 朗读训练的实现

朗读训练可以纠正和规范训练者的英语发音。在训练过程中,训练者选择训练单元后,系统首先给出标准发音,然后训练者通过麦克风将语音信息输入系统,系统对语音信息进行识别,若发音准确,系统识别输出正确文本信息,否则,系统不能识别,并提示训练者发音不准确。实现过程:设置两个定时器,一个定时器顺序地完成文字信息的读取、文本信息的语音转换和标准语音信息的输出。而另一个定时器则顺序地完成语音信息的输入、识别和文字信息的输出。其中双语混读、朗读者口型和语音识别是系统实现的关键。

(1) 双语混读的实现

某些文章中既包含中文信息,又包含英文信息,这就需要双语混读功能。双语混读的设计思想:根据 ASCII码判断下一个单词是否为中文^[6]。若为英文字符,系统自动切换当前 SS引擎为英文文语转换引擎,相反,则切换为中文文语转换引擎。具体实现:把要阅读的字符串全部存储到数组中,然后从第一个字符开始判断,如果它的 ASCII码小于 256则可判断为英文字符,逐一继续判断,直到判断出 ASCII码大于 256的字符,此时便将当前字符之前的英文字符串一次读出。同理也可读出中文字符。具体伪代码如下:

```
private void hybridSpeak( string ceSpeak)
{For(从字符串的第一个字符开始处理直到最后一个字符)
{ If(已经到了字符串的最后一个字符)
{ If(如果该字符串为中文字符串)
    中文发音并且对中文字符进行累加
Else
    英文发音并且对英文字符进行累加
}
Else/如果没有到最后一个字符
{读取字符的 ASCII码
  If(该字符为中文字符)
  {
    If(当前英文单词累加数为 0)
      中文字符个数累加
    Else
      {读出之前所有的英文字符串;
        英文字符串个数复位
        中文字符个数累加
      }
    }
```

```
Else/该字符为英文字符
处理与处理中文字符刚好相反,不再赘述
}}
```

(2) 朗读者口型的实现

朗读者口型在 SAPI中被称为 Viseme。当训练者在朗读英文时,英语自学系统提供一个模拟真人朗读口型的动画,以纠正不正确的朗读口型。软件设计思想:

通过一组静态图片模拟朗读者口型的变化,在文 语转换的程序代码处设置 Visime的事件驱动函数,该事件驱动函数顺序地完成口型图片的更新。文 语转换时系统自动调用驱动函数完成口型动画演示。具体实现:Visime由不同口型图片组合而成,放在 ImageList控件中。每调用一次 Visime驱动函数,系统就从 ImageList中取出合适的口型图片并显示出来。微软设计了一组较合理的口型图片顺序,如表 2所示。代码实现分步完成:

Step1:创建一个 ImageList控件,按照表 2中的顺序,把口型图片添加到 ImageList控件中,并指定当前口型图片为 1号口型图片。

```
pictureBox4. Image = imageList1. Images[0];
Step2:设置事件驱动函数。
voice Visime+=new SpeechVoiceEventsVisimeEventHandler( voice Visime);
Step3:在驱动函数中实现朗读者口型的动画显示。
private void voiceVisime( int StreamNumber, object StreamPosition, int Duration, SpeechVisimeType
NextVisimeId, SpeechVisimeFeature Feature, SpeechVisimeType CurrentVisimeId)
{
    int i=int Parse(CurrentVisimeId.ToString().Replace("SVP-", "")); //提取口型图片
    pictureBox4. Image = imageList1. Images[ i]; //显示口型图片
    pictureBox10. Image = imageList1. Images[ i];
}
```

(3) 语音识别的实现

语音识别功能实时收集训练者的语音信息,然后分析其正确性并以文本形式显示。它是纠正训练者英语口语音的关键。通过设计语音识别事件驱动函数完成语音识别。伪代码实现分步完成:

- Step1:定义语音识别文本类型及语音识别语法规则。
- Step2:为语音识别指定事件驱动函数和语法规则。
- Step3:在收集测试者语音信息的过程中,设置语法规则可用。
- Step4:设计事件驱动函数 Reco_Event将收集到的语音信息转化为文本信息。

2.3 听写训练和阅读训练的实现

系统实时判断训练者所拼写的单词,并记录错误单词。设计思想:英文中每两个单词间是通过空格加以分隔,所以用空格来“联想”新的英文单词。实现过程:设置两个数组 gAlW ord和 gErW ord 分别记录听写过的单词和拼写出错的单词;设置一个定时器获取训练者输入的文本信息,然后对这些信息进行单词识别,把错误单词添加到 gErW ord中;读取 gErW ord数组,标示出所有出错单词。听写训练中,训练者还可设置不同的朗读者、语速、音量等。为代码实现分步完成:

Step1:从数据库中获取训练内容,以单词训练为例。

Step2: 错误单词的判断及记录。

Step3: 定时事件被触发, 检测测试者输入的单词。

Step4: 单词分析, 把空格作为联想单词的依据。

阅读训练从阅读、翻译和写作三个方面分别对训练者进行训练。阅读训练的设计思想: 用定时器来实时获取训练者输入的文本信息, 并对错误单词进行标识。而翻译训练和写作训练的设计思想较相似, 即从数据库中随机抽取训练内容后, 测试者开始倒计时训练, 训练完成后, 给出范文或参考答案, 代码不再赘述。

3 结束语

随着计算机技术和多媒体技术的迅速发展, 英语自学系统也需要不断创新以适应日益变化的应用需求。文中在 Windows 环境下实现了包含“听、说、写”等功能的英语自学系统。该系统可在一定程度上辅助英语学习者, 完成英语自主学习。为了完善系统, 笔者还会在不同语音转换引擎的移植、语音识别的正确率等方面做进一步的研究, 期望能取得更有价值的研究成果。

参考文献:

- [1] 舒 后. 文语转换技术的应用研究 [J]. 北京印刷学院学报, 2003, 11(2): 12—15.
- [2] 姚涵珍, 陆文秀. TTS 中文语音合成技术的研究与实践 [J]. 天津科技大学学报, 2002, 19(1): 65—67.
- [3] 赵峻松. 利用 TTS 技术开发计算机语音功能 [J]. 五邑大学学报 (自然科学版), 2002, 16(1): 54—58.
- [4] 李禹材, 左友东, 郑秀清, 等. 基于 Speech SDK 的语音控制应用程序的设计与实现 [J]. 计算机应用, 2004, 24(6): 114—116.
- [5] Microsoft speech SDK 5.1 Help [EB/OL]. <http://www.microsoft.com>. 2001—08—15.
- [6] 袁 军, 张思民. 嵌入式中文 TTS 系统的研究与实现 [J]. 电脑知识与技术, 2008(6): 1 345—1 347.

Design of an English Self-study System Based on TTS and SR

LI Guang-li ZHANG Hong-bin

(School of Information Engineering East China Jiaotong University Nanchang 330013 China)

Abstract: Transformation from text to speech and speech recognition are two key technologies to design an English self-study system. The paper firstly introduces the working theory and software structure of the English self-study system. And then it puts forward how to input the Microsoft Speech API in the Net environment. With the help of this program design API, the system implements several functions such as the listening training function and the reading training function. Moreover, the system gains better interaction between human and computer by using several multimedia means of voice and animation.

Key words: English self-study system; text to speech; speech recognition; the Microsoft Speech API; .Net

(责任编辑: 王建华)