

基于 Speech SDK 的语音应用程序实现

Implement of Speech Application Program Based on Speech SDK

高敬惠, 姜子敬, 胡金铭

Gao Jinghui, Jiang Zijing, Hu Jinming

(河北工业大学廊坊校区计算机系, 河北廊坊 065000)

(Dept. of Comp., Langfang Borough, Hebei Univ. of Tech., Langfang, Hebei, 065000, China)

摘要: 利用 Microsoft Speech SDK 的 API for Text-to-Speech 和 API for Speech Recognition, 采用 Visual Basic 6.0 语言, 建立文本语音转换应用程序和实现语音识别程序, 简单地实现了语音识别的功能, 识别出来的内容即可保存为文件, 也可作为命令使用, 让计算机执行某项操作。

关键词: 应用程序 文本语音转换 语音识别 Microsoft Speech SDK

中图分类号: TN 912.34 **文献标识码:** A **文章编号:** 1002-7378(2005)03-0169-04

Abstract: A text-to-speech application program and speech recognition procedure are developed using Visual Basic 6.0 and API for Text-to-Speech and API for speech recognition of Microsoft Speech SDK. This text realizes the function of speech recognition briefly. The content recognized can be saved as a file and be used as a order in a coming operation.

Key words: application program, text-to-speech, speech recognition, Microsoft Speech SDK

早在 20 世纪 70 年代, 国外就开始了语音技术的研究, 语音合成技术包括文本-语音转换(Text-To-Speech, 简称 TTS)和语音识别(Speech Recognition, 简称 SR)。TTS 系统有文本分析模块、韵律分析模块、语音生成模块 3 部分组成。尽管目前的 TTS 技术在人性化方面, 还存在需要提高的地方, 但是 TTS 的开发和应用已逐步走进人们的生活, 如应用 TTS 实现以语音方式的人机交互、计算机训练、帮助残疾人生活、信息服务等等诸多方面。SR 技术是指用人的语言去指挥和控制机器, 让机器“听懂”人的语言, 并根据其指令去完成各种各样的任务。如 e-talk 技术的语音识别系统可以识别多种语音, 能够为用户提供诸如国家代码、地区代码、资费、时差等内容的全天 24h 不间断的咨询服务。只需通过对话便可以获得亲切而有效的服务, 不管你运用的是英语还是汉语、日语等。

即使语音识别引擎中使用的语音模型是经过大量训练得到的, 但是当用户使用时, 用于个人的发音

特点, 往往与使用的语音模型系统之间的匹配效果不能达到最佳状态, 为了使识别系统能适应每个用户的独特性, 用户在使用之前还需要使用识别系统提供的训练系统进行语音输入的训练, 训练过程是不断地调整系统模型的参数, 使系统的某种最佳状态, 如对语音库中的所有语音有最好的识别率等。Microsoft Speech SDK 提供了语音技术支持, 本文利用其中的 API for Text-to-Speech 和 API for Speech Recognition 来建立文本语音转换应用程序和实现语音识别程序, 实现了语音识别的功能。

1 Microsoft Speech SDK 的语音技术^[1]

Microsoft Speech SDK 是微软提供的软件开发包, 提供的 Speech API (SAPI) 主要包含 API for Text-to-Speech 和 API for Speech Recognition, 其中 API for Text-to-Speech 是微软 TTS 引擎的接口, 通过它可以很容易地建立功能强大的文本语音程序, 金山词霸的单词朗读功就是使用了这个功能, 目前几乎所有的文本朗读工具都使用这个 SDK 开发。API for Speech Recognition 是与 TTS 相对应的语音识别, 通过语音识别可以使计算机听懂人的说话, 但是目前语音识别技术准确度和识别速度还

收稿日期: 2005-01-19

修回日期: 2005-03-15

作者简介: 高敬惠(1970-), 女, 讲师, 主要从事多媒体技术、数据库研究。

不太理想,未达到广泛应用的要求。

实现中文发音或中文语音识别必须先安装微软的 Speech Application SDK (SA SDK), 它的最新版本是 SAPI 5.1 它能够识别中、日、英三种语言, 你可以在微软的网站免费下载, 需要下载安装 2 个文件 Speech SDK 5.1 和 5.1 Language Pack, 其中 5.1 Language Pack 是附加的语言包支持中文。

2 文本语音转换的实现

2.1 简单语音朗读的实现

安装 Speech SDK 5.1 后, 最直观和方便的实现语音朗读的方法就是使用其中的控件“Microsoft voice text”^[1], Visual Basic 6.0 中的实现过程如下:

建立一个新工程, 选择“工程”菜单中的“部件”选项, 在控件列表框中选中“Microsoft Voice Text”, 则“TextToSpeech”控件出现在工具箱中。在窗体中建立一个文本框 Text1、一个命令按钮 Command1 和一个“TextToSpeech”控件 TextToSpeech1, 如图 1 所示。

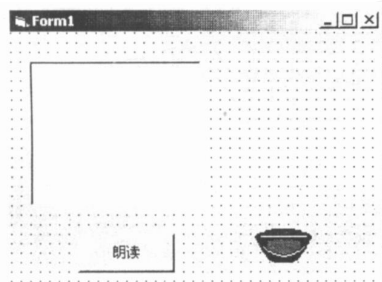


图 1 设计界面

编写 Command1 的时间代码:

```
Private Sub Command1_Click()
    If Text1.Text <> "" Then
        TextToSpeech1.Speak Text1.Text
    End If
End Sub
```

运行程序, 当在文本框中输入内容后, 单击“朗读”, 即可听到文本朗读。

上面的程序虽然实现了语音的朗读, 但是声音不能改变, 下面给出可以选择英文男声和女声及中文朗读, 并可以调节音量和音速的程序实现方法。

2.2 参数可调的语音朗读的实现

新建一个工程, 选择“工程”菜单中的“引用”, 在引用下拉列表框中选择“Microsoft Speech Object Library”, 使下面的程序可以引用语音对象库的对象。在窗体中添加一个文本框 Text1, 一个命令按钮 Command1, 一个组合框 Combo1, 两个水平滚动条

HScroll1 和 HScroll2, 如图 2 所示。

编写事件代码。

“通用”事件代码:

```
Option Explicit
Private V As SpeechLib.SpVoice
Private T As SpeechLib.ISpeechObjectToken
```

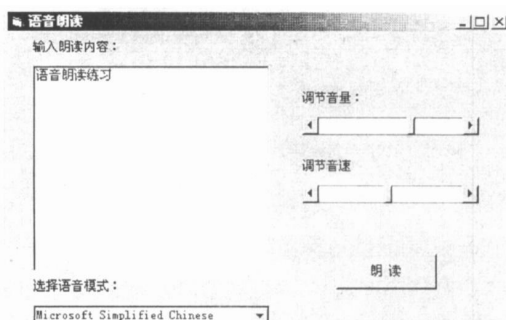


图 2 语音朗读程序界面

其中, Ispeech 是实现语法分析和朗读文本字符和文本文件的接口, IspeechObjectToken 是 SpObjectToken 对象的集合, SpObjectToken 代表了能被 SAPI 5.1 speech recognition (SR) 和 text-to-speech (TTS) 使用的各类资源, 如不同类型的声音 (如男声、女声、中文等)、词汇及声音输入设备等^[2]。

Form_Load 事件代码:

```
Dim strVoice As String
Set V = New SpVoice
```

GetVoices 返回当前使用的声音对象的标记, 如果没有目前没有声音使用返回默认的声音。Get each token in the collection returned by GetVoices

For Each T In V.GetVoices ' 通过 GetVoices 得到声音集合中的所有对象标记 strVoice = T.GetDescription ' 得到声音标记的名称

Combo1.AddItem strVoice ' 将得到的名称加入组合框, 如图 3 所示。

```
Next
HScroll2.Min = 0
HScroll2.Max = 100
HScroll2.Value = V.Volume
HScroll1.Min = -10
HScroll1.Max = 10
```

HScroll1.Value = V.Rate

水平滚动条 1 的 HScroll1_Change 代码:

V.Rate = HScroll1.Value ' 音速为滚动条的值

水平滚动条 2 的 HScroll2_Change 代码:

V. Volume = HScroll2 Value ' 音量为滚动条的值

命令按钮 Command1_Click 代码:

```
If Combo1.ListIndex > - 1 Then
    ' 设置声音对象为组合框中选中的名称
    Set V. Voice = V. GetVoices(). Item (Combo1.ListIndex)
    V. Speak Text1.Text
Else
    MsgBox "请在组合框中选择声音"
End If
```

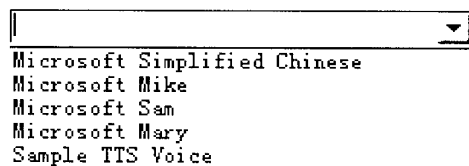


图3 组合框中声音名称

运行时, 通过在组合框中选择不同的声音可朗读文本框中输入的中文文本或英文文本, 还可以选择男声或者女声。

以上程序实现了VB 中的语音朗读功能的开发, TTS 为课件开发中的英语课件、学生的学习课件等的发音功能提供了很大的方便。

3 语音识别的实现

新建一个工程, 打开“工程”菜单中的“引用”选择下拉列表框中的“Microsoft Speech Object Library”。这样即可引用语音对象库中的对象。在窗体中添加一个列表框List1, 用于显示识别的内容, 再添加一个命令按钮“开始”(见图4), 编写事件程序主要接口代码如下。

(1) SpSharedRecoContext 接口^[3]。语音识别引擎有: 独占和共享, 独占的引擎对象只能由创建的应用程序使用, 共享的引擎可以供多个应用程序共同使用。SpSharedRecoContext 主要用于接受和发送共享引擎中与语音识别消息相关的事件消息, 装载和卸载识别语法资源。

(2) IspeechRecoGrammar 接口^[3]。通过这个接口, 应用程序可以载入、激活语法规则, 而语法规则里则定义着期望是别的单词、短语和句子。通常有两种语法规则: 听写语法和命令控制语法。听写语法用于连续语音识别, 可以识别出引擎词典中大量的词汇, 例如, 可以识别一段话、一个单词、一篇文章等, 也就是说, 可以用语音代替键盘进行文字输入。命令

控制语法主要用于识别用户在语法文件里自定义的一些特定的命令词汇和句子。

(3) IspeechPhrase 接口。用于获取识别的结果, 包括识别的文字, 识别了哪一条语法规则等。

FormLoad 事件代码:

```
Option Explicit
' 定义语音识别事件
```

```
Dim WithEvents context As SpSharedRecoContext
```

```
' 定义 IspeechRecoGrammar 类型的变量
```

```
Dim Grammar As IspeechRecoGrammar
```

“开始”按钮事件代码:

```
If (context Is Nothing) Then
```

```
Set context = New SpSharedReco-
```

```
Context
```

'CreateGrammar 方法是产生一个基于 IspeechRecoGrammar 的对象

```
Set Grammar = context CreateGrammar(1)
```

'DictationLoad 方法是装载识别语音内容的文法规则

```
Grammar.DictationLoad
```

```
End If
```

```
' 设置口述内容的接收状态为开启
```

```
Grammar.DictationSetState SGDSActive
```

语音转化为文字需要用到 SpSharedRecoContext 控件的 Recognition 事件, 这个事件的代码如下:

```
Private Sub context_recognition (ByVal stream-
```

```
number As Long, ByVal streamposition As Vari-
```

```
ant, ByVal recognitiontype As SpeechRecogn-
```

```
itionType, ByVal result As IspeechRecoResult)
```

```
' 在列表框中显示识别的结果
```

```
List1.AddItem result.PhraseInfo.GetText
```

```
End Sub
```

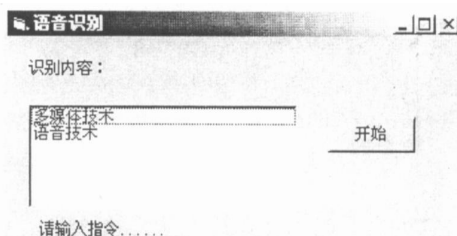


图4 窗口设计

插上麦克风, 运行程序, 开始时稍微放慢语速, 最初词汇识别效果并不理想, 主要原因是 Speech SDK 需要一个训练过程, 以便能区分语音和周围的噪音, 并且通过训练系统能逐渐适应用户的发音特

点,从而比较准确的识别出用户所读出的词汇。

安装完 Microsoft Speech SDK 后,在“控制面板”中就会出现一个“语音”图标,通过语音属性,如图 5 所示,可以对计算机朗读时使用的语言和语音识别时使用的配置文件等进行设置,还可以通过“训练配置文件”激励形成用户的语音模型,提高语音识别的准确率。训练过程如下:打开“控制面板”中的“语音”,在“语音属性”中选择“训练配置文件”,如图 6 所示,在声音训练框中选择一个“微软语音识别系

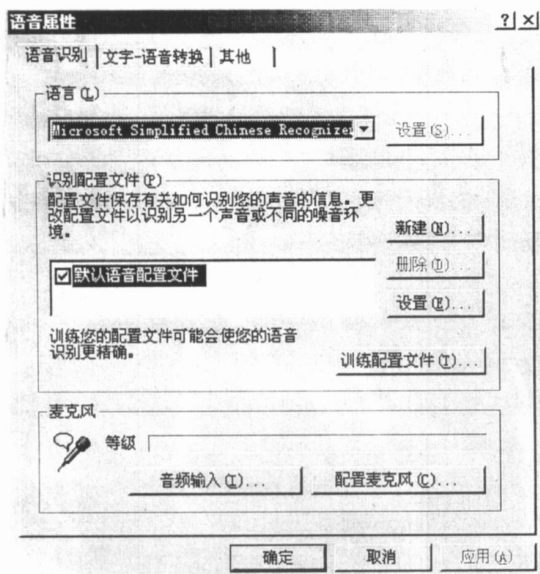


图 5 语音属性对话框



图 6 声音训练

统简介”,随后按照提供的文本内容以稍慢、清晰的语调朗读文本,训练完毕后,再运行识别程序时,准确率会得到较大的提高,如果想达到更高的识别率,可以选择更多的训练内容进行训练。

4 结束语

本文提出的方法简单地实现了语音的识别功能,识别出来的内容可以保存为文件,也可以作为命令使用,让计算机去执行某项操作。

如果要更好的执行语音命令可以建立常用语音命令集,从而实现更准确的语音命令识别和执行^[4],本文不再展开讨论。从以上的讨论中可以看到,在教学课件中引入语音识别的功能,可以实现人-机对话,通过用户口述命令,计算机就可完成用户的命令。

参考文献:

- [1] Microsoft Inc Microsoft speech SDK 帮助文档[EB/OL]. Http://www.microsoft.com/speech/download/sdk51, 2004-10-23
- [2] 李禹才,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):114-117.
- [3] Dixon M, Maclin O. Visual basic for behavioral psychologists[M]. NV: Context Press, 2003
- [4] Cabello F Barnes-Holmes, D O'Hara, Stewart D. Using Visual Basic in the experimental analysis of human behavior: A brief introduction[J]. Experimental Analysis of Human Behavior Bulletin, 2002, 20: 17-20
- [5] Staunton C, Barnes-Holmes D, Whelan R, et al. Priming and event related potentials as measures of derived relational responding[C]. Paper presented at the 29th annual convention of the association for Behavior Analysis, San Francisco, U.S.A. 2003

(责任编辑:黎贞崇)

能诱导糖尿病的蛋白

脂肪组织中的胰岛素作用遭受影响是 II 型糖尿病的一个主要病因,现在一个以前未知的,也许能够促使这一过程发生的机制已被发现。曾有报道说,糖尿病患者血清中维生素 A 的一种结合蛋白 RBP4 的含量增加,但没有怀疑到二者之间有因果关系。“全面基因表达分析”方法被用来识别脂肪组织中 GLUT4 葡萄糖转移因子功能受损(肥胖症和糖尿病的特征)的小鼠体内其表达发生改变的基因,结果显示 RBP4 含量增加。另外,增加正常小鼠血清中 RBP4 含量会引起胰岛素抗性,而用增强胰岛素敏感性的药物降低该蛋白在肥胖小鼠血清中的含量会减轻胰岛素抗性,说明降低 RBP4 也许是治疗 II 型糖尿病的一种可行方案。

据《科学时报》