

# 利用 TTS 技术实现文本文件的语音合成

毕晓君, 静广宇, 徐先锋

(哈尔滨工程大学, 黑龙江 哈尔滨 150001)

**摘 要:**本文基于语音合成的代表性技术 TTS,利用 Microsoft Speech SDK 语音开发包、TTS 引擎和 MFC 微软基础类库,开发了一个在 VC 集成环境下的文语转换应用程序,实现了从文本文件到语音的自动转换功能。

**关键词:**语音合成; TTS 引擎; 微软语音开发包; 文语转换系统

**中图分类号:** TP182 **文献标识码:** A **文章编号:** 1003-7241(2004)09-0049-03

## The Speech Synthesis from Text to Speech Based on the TTS Technique

BI Xiao - jun , JING Guang - yu , XU Xian - feng

(Harbin Engineering University, Harbin 150001, China)

**Abstract:** Based on the representative technique of speech synthesis TTS, the Microsoft Speech SDK, the TTS engine and MFC, a program under integrated environment of VC has been compiled in this paper. The quickly transformation from text to speech including English and Chinese has been realized.

**Key words:** Speech synthesis technique; TTS engine; Microsoft Speech SDK; Text - To - Speech conversion system

## 1 引言

语音合成技术是信息处理领域的一个重要分支,实现计算机文本文件语音合成,就是让计算机开口说话,这也是人工智能的一个重要研究方向。语音合成是一门跨学科的技术,涉及到自然语言理解、语音学、信号处理、心理学、声学等。TTS 技术(Text To Speech)是当前语音合成技术的代表性研究内容,它解决的主要问题是如何将文本状态的文字信息转化为可听的声音信息。这一技术在人机对话、电话咨询、自动播音、助讲助读、语音教学等方面有着广阔的应用前景<sup>[1]</sup>。

## 2 TTS 技术支持下从文本到语音的实现方法

TTS 技术本身原理十分复杂,但是微软最新提供的 Microsoft Speech SDK 开发包已经将主要的 TTS 技术包含在内,尤其是提供了 TTS 引擎接口,所以在开发的整个过程中,可以将这些技术以“类”和“函数”的方式灵活地应用到编制的程序中。

为实现这一思想,还需要选择适当的编程语言,本文选择了 Visual C++ 开发 Windows 窗口下进行中文发音的应用程序,通过编制相应的源代码,实现了从文档到语音的快速转换功能,并对中英文进行自动识别朗读,其设计原理框图如图 1 所示。

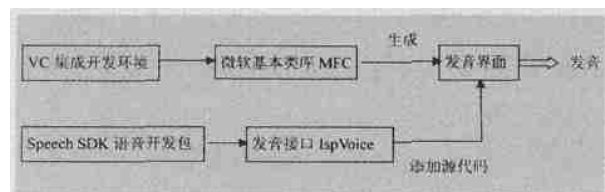


图 1 文本文件语音合成系统设计原理框图

Microsoft Speech SDK 是微软提供的一个软件开发包,它主要为实现计算机的语音功能而设计,可以用来建立语音引擎和微

软 Windows 应用程序。SDK 包含与微软 Win - 32 兼容的语音应用程序编程接口 Speech API (SAPI), 微软连续语音识别引擎以及微软连接语音合成引擎 (或称文语转换引擎), 它包括基于语音导向的用来编译源代码和执行命令的开发工具包, 它是语音与引擎技术的典型应用, 也是语音识别和语音合成的典型应用指南。

Speech API (SAPI) 主要包含以下两个部分<sup>[2]</sup>:

- (1) API for Text - to - Speech;
- (2) API for Speech Recognition.

其中 API for Text - to - Speech 就是微软 TTS 引擎的接口, 通过它可以建立功能较强的文本语音程序。

Visual C++ 是微软公司为开发 Windows 应用程序而创建的一种集成开发环境, 特别是它提供的微软基本类库——MFC (Microsoft Foundation Class Library) 更是为用户提供大量预先编好的类及支持代码, 这些类封装了大量标准的 Windows 编程中要使用的处理函数和数据, 从而大大简化了用户的 Windows 编程工作<sup>[3]</sup>。本文利用微软基本类库实现了文本文件语音合成的应用程序编程工作。

首先将 Microsoft Speech SDK 安装到计算机上, 其 5.1 版本支持中文发音, 但是要实现中文发音还需要同时下载中文发音包 SmpChinese Speech Package。在 VC 集成环境下, 调用 Speech SDK 语音开发包, 开发文语转换应用程序<sup>[4]</sup>。本文的主要工作步骤如下:

- (1) 生成程序框架;
- (2) 设计对话框: 编辑对话框资源, 并添加所需控件;
- (3) 调用对话框, 添加代码: 设定变量类型, 向消息处理函数中添加代码;
- (4) 编译运行应用程序。

### 3 从文本到语音转换的实现

为实现计算机文本文档语音合成, 首先必须利用 Visual C++ 设计一个系统对话框界面, 通过这个界面来实现发音控制, 并且能够自动导入一段中文或英文, 以便实现发音示例, 而且应该可以选择发音者, 并且还可以设定音量和朗读的节奏。

本文开发的发音界面示意图如图 2 所示。

其中各个控件实现的功能为: 单击“导入英文”可以在文本编辑框中自动导入自己设定的一段英文, 并将发音者设定为英文女声 Microsoft Mary; 单击“导入中文”可以在文本编辑框中自

动导入自己设定的一段中文, 并将发音者设定为中文男声发音; “朗读”和“停止”控件可以随意开始或停止文本编辑框中文字的发音, 另外还可以控制发音的语速和音量, 以及选择 Speech SDK 提供的不同发音者。

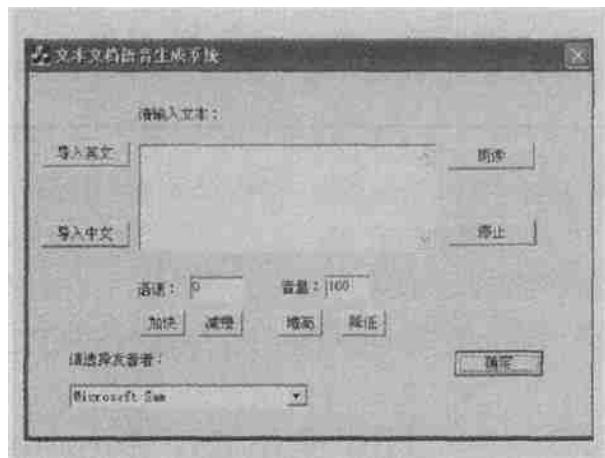


图 2 发音界面示意图

在 VC 环境下实现发音的关键源代码为:

```
#include <sapi.h>

#pragma comment(lib, "ole32.lib") //CoInitialize CoCreateInstance
#pragma comment(lib, "sapi.lib") //sapi.lib 在 SDK 的 lib 目录 int
main(int argc, char * argv[])
{
    ISpVoice * pVoice = NULL;
    //COM 初始化:
    if (FAILED(::CoInitialize(NULL)))
        return FALSE;
    //获取 ISpVoice 接口:
    HRESULT hr = CoCreateInstance(CLSID_SpVoice, NULL, CLSCTX_
        ALL, IID_ISpVoice, (void **) &pVoice);
    if (SUCCEEDED(hr))
    {
        hr = pVoice -> Speak(L "Hello world", 0, NULL);
        pVoice -> Release();
        pVoice = NULL;
    }
    ::CoUninitialize();
    return TRUE;
}
```

发音程序的流程是获取 ISpVoice 接口, 然后用 ISpVoice::Speak() 把文本输出为语音。其它函数如下:

```
HRESULT Speak(const WCHAR *pwcs, DWORD dwFlags, UL
```

ONG \*pulStreamNumber);

本函数的功能为朗读

HRESULT Pause ( void );HRESULT Resume ( void );

功能:暂停和重新开始。

HRESULT SetRate(long RateAdjust );HRESULT GetRate(long \*pRateAdjust);

功能:设置/获取播放速度,范围:-10 to 10

HRESULT SetVolume (USHORT usVolume);HRESULT GetVolume(USHORT \*pusVolume);

功能:设置/获取播放音量,范围:0 to 100

HRESULT SetSyncSpeakTimeout (ULONG msTimeout);

HRESULT GetSyncSpeakTimeout (ULONG \*pmsTimeout);

功能:设置/获取同步超时时间。由于在同步模式中,电泳 Speak 后程序就会进入阻塞状态等待 Speak 返回,为免程序长时间没相应,应该设置超时时间,msTimeout 单位为毫秒。

HRESULT SetOutput (IUnknown \*pUnkOutput ,BOOL fAllowFormatChanges);

功能:设置输出。

这些函数的返回类型都是 HRESULT,如果成功,则返回 S\_OK;如果出现错误,则有各自不同的错误码。

## 4 结束语

本文开发了一个在 VC 集成开发环境下的文语转换应用程序。它利用微软 Microsoft Speech SDK 语音开发包和提供的 TTS 引擎,根据 MFC 微软基础类库,编辑文语转换系统界面,并且通过插入开发的源代码,实现了从文本文件到语音的快速转换功能。实验结果表明,利用本文开发的系统对话框界面,用户可输入任意汉字点击朗读即可进行自动语音输出,汉字识别速度较快,语音合成效果较好。同时用户可根据系统界面任意选择男女发声、语音语速等,从而实现了计算机自动朗读文本文件。该技术可用于开发自动阅读机等产品。

## 5 参考文献:

- [1] 王凯. TTS 技术应用探讨[J]. 贵州大学学报(自然科学版), 2000, (4):304 - 306
- [2] 曾志军. 如何将最新的 TTS 引擎技术应用到 VC 程序中[J]. 计算机与现代化, 2002, (4):11 - 12
- [3] 黄维通. Visual C++ 面向对象与可视化程序设计习题解析与编程实例[M]. 北京:清华大学出版社, 2000 年
- [4] 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, (1):54 - 58

作者简介:毕晓君(1964-),女,硕士,副教授,主要研究方向:数字图像处理与语音识别。

(上接第 44 页)

## 5 结语

由于卷取区域采用新型的控制系统,具有以下优点:采用快速稳定的 ISBus 总线网络代替模拟/数字 I/O 连接,降低故障点数,提高传动系统的实时响应和稳定性;增加了助卷辊 Jump 控制功能,有效的改善了卷型;采用 TOOLBox 的控制软件,有效实现了传动的高级控制功能,系统的操作、配置和调试简单方便。各种报警和故障信息详细清楚,有助于事故的迅速判断和及时处理,提高系统的可靠性。新系统的投入到现在,系统运行稳定,创造了良好的经济效益。所以基于控制软件的计算机网络控制是复杂传动控制系统发展的必然趋势。

## 6 参考文献:

- [1] 李好文,郑岗,余健明,冯鑫,等. 计算机多级网络控制技术[J]. 重型机械, 2002, (2):16 - 18.
- [2] 童朝南,彭开香. 轧制过程计算机控制系统[J]. PLC&FA, 2004, (1):95 - 99.
- [3] System Manual: For the Innovation Series Control[EB/OL]. <http://www.geindustrial.com>, 2003 - 12 - 29.
- [4] Control System TOOLBox: For an Innovation Controller[EB/OL]. <http://www.geindustrial.com>, 2003 - 12 - 29.
- [5] Control System TOOLBox: For an Innovation Drive[EB/OL]. <http://www.geindustrial.com>, 2003 - 12 - 29.

作者简介:刘庚(1978-),男,安徽阜阳人,软件工程专业硕士研究生,研究方向为工业控制软件应用和开发。