基于 Speech SDK 的语音识别技术在三维仿真中的应用

林鸣霄
(同济大学 电子与信息技术学院,上海 201804)

摘 要:随着三维仿真技术的不断发展,简单的人机交互方式已经不能满足人们对仿真环境真实感和沉浸感的要求。针对于此,提出了将基于 Speech SDK5.1 的语音识别技术应用到三维仿真平台的构想,分析了 Speech SDK5.1 的工作原理,着重研究了其语音识别接口,对将语音识别应用到三维仿真程序中的可能性和关键技术进行了研究。提出了一种实现动态词汇识别的方法,并通过一个简单的实例展示了实现这类技术的框架和方法,对设计有语音识别功能的三维仿真程序有一定的参考价值。

关键词:语音识别;三维仿真;Speech SDK;COM;语音控制
中图分类号:TP39 **文献标识码**:A **文章编号**:1673-629X(2011)11-0160-03

Application of Speech Recognition Technology in 3D Simulation Based on Speech SDK

LIN Ming-xiao
(Electronics and Information Technology Institute of Tongji University, Shanghai 201804, China)

Abstract: With the continuous development of 3D simulation technology, simply man-machine interactive way cannot satisfy people of simulation environment realism and immersed sense requirements. Based on this, put forward the idea of applying speech recognition technology which based on Speech SDK5.1 to the 3D simulation platform, analyse the working principle of SDK5.1 and mainly focus on the speech recognition interface and the ability and key technology of that. Then present a dynamic vocabulary identification method and show the main framework and technology with a simple instance which has certain reference value for designing a 3D simulation program with speech recognition function.

Key words: speech recognition; 3D simulation; Speech SDK; COM; voice control

0 引 言

随着计算机三维仿真技术的不断发展和应用领域的不断扩大,与虚拟世界的交互方式也越来越丰富,简单的肢体交互已经不能满足人们对真实感和沉浸感的要求,人们开始寻找与真实世界中更为相似的沟通方法[1]。而语言是人类交流信息最简单、最自然的手段。所以,很自然地想到把人类语言和三维仿真技术结合起来。与此同时,语音识别和控制技术的不断发展给了人们一个这样的机会和可能,也提供了一种更真实的与仿真世界交互的方法[2]。

目前,国内基于语音识别技术的研究尚处在起步阶段,其中大多是利用了微软 Speech SDK 开发包进行的开发,如武汉大学开发的"基于语音识别的火车票查询系统"[3],上海交大的"Call Center"系统[4],林茜等人设计的"语音关键字检出系统"[5]等。但是语音识别技术在三维仿真中的研究还较少。

文中分析了微软 Speech SDK 5.1 里语音应用程序接口(SAPI)的结构和工作原理,对语音识别技术应用于三维仿真系统的可能性和技术难点进行了研究,提出了语音控制在三维仿真中应用程序设计的方法,研究了这类系统的主框架和关键技术,通过与仿真平台的结合,设计和实现了简单的具有语音控制功能的三维仿真程序,对语音控制的三维仿真程序有普遍的参考价值。

1 Speech SDK 简介

Microsoft Speech SDK 5.1 是一套语音应用程序开发的软件开发资源包,它完全基于 COM 标准开发,底

收稿日期:2011-04-07;**修回日期**:2011-07-13
基金项目:国家863计划重点项目(2010AA122200);上海市科委国际合作项目(10510712500)
作者简介:林鸣霄(1987-),男,硕士研究生,研究方向为计算机辅助设计分析与仿真、虚拟现实、图形图像技术。

层协议以 COM 组件的形式完全独立于应用程序层,开发人员可以方便使用资源包中的资源开发语音识别和语音合成的应用程序,而不必纠结于复杂的语音技术,而且 Microsoft Speech SDK 完全支持简体中文语音系统,是一个理想的开发工具。其中,语音识别由识别引擎(Speech Recognition,简称 SR)管理,语音合成由语音合成引擎(Text To Speech,简称 TTS)负责;程序员只需专注于自己的应用,调用相关的语音应用程序接口(SAPI)来实现语音功能[6,7]。结构图如图 1 所示。

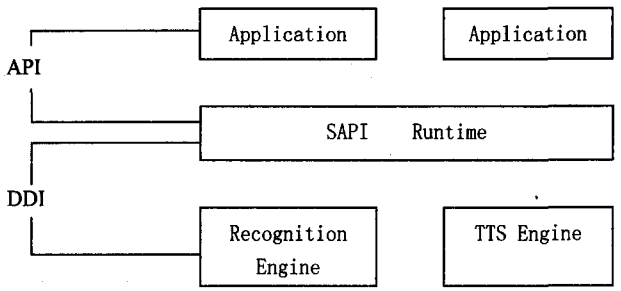

图 1　Speech SDK 5.1 结构图

其中,TTS 主要实现的是由文本到语音的转换,属于语音合成功能,现在大多数的包含发音朗读功能的软件(如金山词霸等)都是通过这个接口实现的,这里不再赘述。文中重点关注语音识别引擎的工作原理和调用方法。语音识别功能工作流程概括如下:

· 首先,初始化所需要的 COM 端口;
· 创建识别引擎,创建上下文接口;
· 设置识别消息,设置感兴趣的事件;
· 创建并激活语法规则进行识别;
· 获取识别消息,进行处理;
· 最后,释放创建的引擎,识别上下文对象、语法等。

2　Speech SDK 应用于三维仿真中的关键技术研究

三维仿真中的语音识别和控制应用核心在于用户使用语音与三维仿真环境的完美交互,其关键技术如下:

(1)语音识别技术与三维仿真平台的结合及框架搭建。

基于 Speech SDK 的语音识别技术与三维仿真平台结合的框架搭建可以应用于仿真训练、仿真控制、仿真指挥、仿真传输等各个方面。该框架定义了与其他部分的接口和交互,用户要知道控制仿真系统的基本指令和语句,通过语音识别系统将控制指令输入应用框架之中,语音识别和分析引擎通过分析得到符合预定义的指令关键字,再通过语音识别模块和三维仿真平台的接口控制三维仿真程序,从而实现用户的控制或者用户预想的功能[8]。其基本框架如图 2 所示。

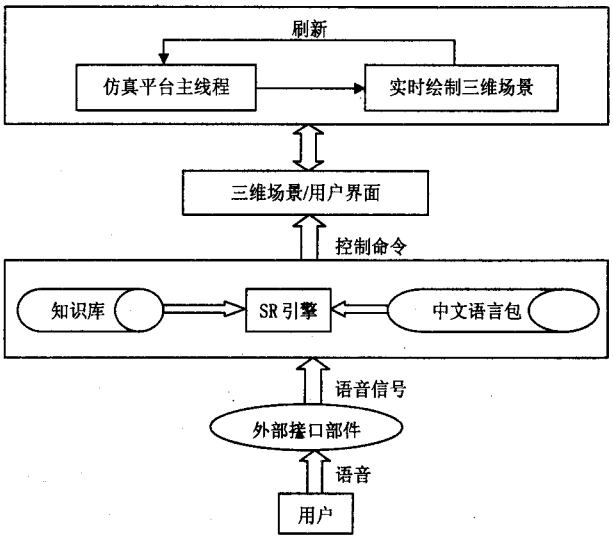

图 2　语音识别与三维仿真平台框架

(2)应用于三维仿真中的语音识别技术关注更多的是非特定人有限词汇的识别,这种识别特征正好符合命令控制语法(Command and Control Grammar)的特点[9],而对于连续语音识别的听写语法则要求不高。而 Speech SDK 虽然对连续语音识别识别程度不高,但是对命令控制的语音识别却相当精确且识别率高,完全可以满足要求。经过大量的测试统计,在忽略口音,语速正常,发音清晰,噪音较小的情况下大约能达到 95% 的正确识别[10,11]。

(3)用于三维仿真的语音识别和控制,要求识别的指令通常是静态指令(Static Content)和动态指令(Dynamic Content)的结合。这就需要把动态的指令从一段指令中分离出来。由于在 SAPI 的命令控制语法(Command and Control Grammar)之下输入的语音必须在语法下定义的范围内才能被识别,因此为了准确识别出指令就必须在语音识别模块中装载一个预设的词汇库,只有当静态指令和动态指令都在预设的词汇库中时,该段语音才能被正确识别,成为正确的指令。SAPI 的语音识别语法和词汇库一般预先设置在 XML 文件中,正确有效的 XML 文件是识别的关键。

下面简单举例说明动态指令识别方法:

举一个简单的语音语句"发送邮件给王二"为例,这个语句中"发送邮件"是静态的动作,很容易被引擎识别。而"王二"则是一个不确定的人,也可能是"张三"、"李四"、"王五",因此,XML 语法需要分离出"王二"这个目标人名,在预先设定好的词汇库中查找,若找到"王二"是预先设定的人名之一,则继续;否则,中断识别或给出提示。识别是在运行时实时(runtime)实现的[12]。示例语法如下:

```
<GRAMMAR LANGID="804">
  <RULE NAME="E-MAIL" TOPLEVEL="INACTIVE">
```

```
        <PHRASE>发送邮件给</P>
        <RULEREF NAME="ADDRESS_BOOK" PROPNAME
="NAME"/>
      </RULE>
      <RULE NAME="ADDRESS_BOOK" DYNAMIC="
TRUE">
        <PHRASE>placeholder</PHRASE>
      </RULE>
    </GRAMMAR>
```

其中,"GRAMMAR LANGID"的值是 804 表示接受的识别语言是中文,"RULEREF NAME"表示用"ADDRESS_BOOK"装载预先设定词汇库,"ADDRESS_BOOK"的属性"DYNAMIC"的值是"TRUE"表示它是一个动态词库。该规则中的 placeholder 并不是目标词汇,只是一个占位符,之后 placeholder 将被清除,替换它的是预设词汇库中的所有词汇。

以下过程用于动态设置规则"ADDRESS_BOOK":

```
HRESULT hr = S_OK;
hr = cpRecoGrammar->GetRule(L"ADDRESS_BOOK",
NULL,SPRAF_Dynamic, FALSE, &hRule);
hr = cpRecoGrammar->ClearRule(hRule);
hr = cpRecoGrammar->AddWordTransition(hRule, NULL,
L"王二", NULL, SPWT_LEXICAL, 1, NULL);
hr = cpRecoGrammar->AddWordTransition(hRule, NULL,
L"张三", NULL, SPWT_LEXICAL, 1, NULL);
……//添加所有预设人名至 ADDRESS_BOOK
hr = cpRecoGrammar->Commit(NULL);
hr = cpRecoGrammar->SetGrammarState(SPGS_ENABLED);
```

这里,首先用 GetRule 函数获取 ADDRESS_BOOK 的初始状态 hRule,再用 ClearRule 函数清除该状态的所有信息,由此清除了 placeholder 占位符,然后用 AddWordTransition 函数将需要添加的人名一个个加入预设的 ADDRESS_BOOK,再用 Commit(NULL)提交这种修改,最后用 SetGrammarState(SPGS_ENABLED)使修改生效[5]。由于实际应用中需要经常更新词汇库,因而在改变目标关键词后,需要重新执行一次检出语法生成过程实现检出语法的实时更新,以生成新的控制命令语音识别时采用的语言规则。

3 在三维仿真中的应用实例

下面将通过一个完整的应用实例介绍如何使用 SAPI 把语音识别功能嵌入三维仿真程序当中。其中三维仿真平台选择较为通用的 Vega Prime 仿真平台,其他仿真平台方法类似。仿真平台已经搭建,场景为高低起伏的平原和红色小车一辆。此应用实例功能是:利用语音口令控制红色小车的运动。预设词汇库包括"启动"、"加速"、"减速"、"停止"四项,控制命令为"小车启动"、"小车加速"、"小车减速"、"小车停止"。

核心代码如下:

· 编写语法文件(command.xml),定义需要识别的命令。

其中,"COMMAND_LIST"如下定义:

```
CpRecoGrammar->AddWordTransition(hRule, NULL, L"启
动", NULL, SPWT_LEXICAL, 1, NULL);
CpRecoGrammar->AddWordTransition(hRule, NULL, L"加
速", NULL, SPWT_LEXICAL, 1, NULL);
CpRecoGrammar->AddWordTransition(hRule, NULL, L"减
速", NULL, SPWT_LEXICAL, 1, NULL);
CpRecoGrammar->AddWordTransition(hRule, NULL, L"停
止", NULL, SPWT_LEXICAL, 1, NULL);
```

· 在主程序的 WinMain 函数中,仿真平台帧循环开始之前,初始化 COM。创建语法规则并装载激活语法。

· 在主仿真程序的 wndProc 函数里,加入一条消息处理。消息识别核心代码如下:

```
// 从语音中获得需要的对象
hr = cpRecoResult->GetPhrase(&pPhrase);
if (SUCCEEDED(hr) && pPhrase)
{
  //如果"小车"被识别
    if (0 == wcscmp(L"OBJECT", pPhrase->Rule.
pszName)
//判断接下来的语句是否是属于 COMMAND 的指令
    if (0 == wcscmp(L"COMMAND", pPhrase->pProper-
ties->pszName))
{
    //把收到的命令保存在一个数组里
hr = pPhrase->GetText(pPhrase->pProperties->ulFirstEle-
ment, pPhrase->pProperties->ulCountOfElements, FALSE,
&pwszCommand, NULL);
Cstring strResult;
StrResult = W2T(pwszCommand);
if(! strResult.CompareNoCase("启动"))
{//给三维仿真引擎发送消息,控制小车开始运动}
……//其他的命令识别类似
}}
      CoTaskMemFree(pPhrase); //释放对象
最后,在程序退出时卸载 COM,即
CoUninitialize();
```

限于篇幅,文中只展示了语音识别模块及与三维仿真平台的通信代码,仿真平台之下的小车运动控制不做赘述。

4 结束语

随着语音识别和控制技术的不断发展,语音技术

(下转第 166 页)

该文中，将功能相似的页面部分抽象出来，编成一个自定义控件，需要用到此功能时只需调用此控件即可。一次编程，多次使用。

4 结束语

采用 MVC 设计模式的三层架构将表示层的展现逻辑和控制逻辑分离开来，View 中只含有页面展示代码，无任何控制逻辑和应用逻辑，提高了程序的可测试性和展现效率。将用户显示（视图）从动作（控制器）中分离出来，提高了代码的可重用性。

此外一个模型可以为多个视图提供数据，这样一个模型一次编写可以被多个视图重用，从而避免了代码的重复编写。

界面设计者和程序员可以更高层次地分工，灵活地分配工作内容，如网页人员，美工能独自参与这些 Web 页面的开发和维护。

参考文献：

[1] 胡迎松，彭利文，池楚兵. 基于. NET 的 Web 应用三层结构设计技术[J]. 计算机工程，2003，29(8)：173-175.

[2] 任中方，张 华. MVC 模式研究的综述[J]. 计算机应用研究，2004，21(10)：1-4.

[3] 黎永良，崔杜武. MVC 设计模式的改进和应用研究[J]. 计算机工程，2005，31(9)：96-97.

[4] Sanderson S. Pro asp. net mvc framework[M]. [s. l.]：après，2009.

[5] 柴哲丽，林佳齐，朱金平，等. 基于贝叶斯的软件可靠性评估改进研究[J]. 计算机工程，2010，36(2)：73-74.

[6] 张 宇，王映辉. 一种 Web 应用框架的研究与实现[J]. 计算机研究与发展，2009，19(5)：99-106.

[7] 李成严，冯慧灵. 基于开源技术的 Web 应用架构研究[J]. 计算机技术与发展，2009，19(8)：27-29.

[8] Guttorm S. The REBOOT Approach to Software Reuse[J]. System Software，1995，30：201-212.

[9] 赵会群，孙 晶，王国仁，等. 软件体系结构研究性能评价研究[J]. 计算机科学，2003，30(2)：144-146.

[10] 郭广义，李代平，梅小虎. Z 语言与软件体系结构风格的形式化[J]. 计算机技术与发展，2009，19(5)：140-142.

[11] IEEE. IEEE recommended practice for architectural description of software-intensive systems[S]. [s. l.]：IEEE，2000.

[12] Kruchten P，Obbink H，Stafford J. The past，present，and future of software architecture[J]. IEEE Software，2006，23(2)：22-30.

[13] 李绍平，彭志平. S2Sh：一种 Web 应用框架及其实现[J]. 计算机技术与发展，2009，19(8)：117-123.

（上接第 162 页）

的应用范围会越来越宽广。目前，语音识别技术中控制和命令（Command and Control）工作方式的识别率非常高，能够达到仿真控制系统中的使用要求，这为用语音代替键盘、鼠标来控制计算机提供了技术保障。而 Speech SDK 提供的中文语音接口也可以使汉语使用者以更方便更自然的方式实现语音控制。另一方面，语音识别控制结合原有的键盘鼠标等肢体控制方法为三维仿真提供了多通道的交互方法，给予体验更好的沉浸感和真实感，由此相信语音识别和三维仿真的结合肯定会越来越多。

文中展示了这类系统的框架和关键技术，并将语音识别技术应用到三维仿真程序当中。这种开发方法对于不同三维仿真平台可以推广开来，具有一定的参考价值。

参考文献：

[1] 何好义. 计算机语音识别技术及其应用[J]. 大众科技，2005(6)：36-37.

[2] 石现峰，张学智，张 峰. 基于 HTK 的语音识别系统设计[J]. 计算机技术与发展，2006，16(10)：12-14.

[3] 吴 萍，胡瑞敏，艾浩军. 火车票查询系统中语音识别的研究及实现[J]. 计算机工程与应用，2003，39(33)：227-229.

[4] 朱 杰，张申生. 基于 COM 技术的语音应用程序的设计和实现[J]. 计算机工程，2001，27(11)：143-145.

[5] 林 茜，欧剑林，蔡 骏. 基于 Microsoft Speech SDK 的语音关键字检出系统的设计和实现[J]. 心智与计算，2007，4(1)：433-441.

[6] 李禹才，左友东，郑秀清，等. 基于 Speech SDK 的语音控制应用的设计与实现[J]. 计算机应用，2004，24(6)：31-35.

[7] 高敬惠，姜子敬，胡金铭. 基于 Speech SDK 的语音应用程序实现[J]. 广西科学院学报，2005，21(3)：169-172.

[8] Kotelly B. The art and business of speech recognition[M]. USA：Pearson，2003.

[9] Marin R，Vila P，Sanz P J，et al. Automatic speech recognition to teleoperate a robot[C]//IEEE/RSJ International Conference on Intelligent Robots and System. [s. l.]：[s. n.]，2002：1278-1283.

[10] Wilpon J G，Rabiner L R，Lee C H，et al. Automatic recognition of keywords in unconstrained speech using hidden Markov models[J]. IEEE Transactions on Acoustics，Speech and Signal Processing，1990，38(11)：1870-1878.

[11] 刘雅琴，智爱娟. 几种语音识别特征参数的研究[J]. 计算机技术与发展，2009，19(12)：19-21.

[12] Microsoft Speech SDK 5.1 Help[EB/OL]. 2001-08-08. http://www. microsoft. com.