(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106444472 A
(43)申请公布日 2017.02.22

(21)申请号 201610553996.8

(22)申请日 2016.07.07

(71)申请人 广东技术师范学院
地址 510665 广东省广州市天河区中山大道293号

(72)发明人 郑振兴 梁鹏 余经烈 林泽芳 吴玉婷

(51)Int.Cl.
G05B 19/042(2006.01)
G10L 15/22(2006.01)

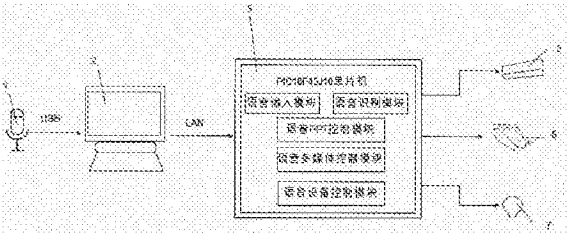
权利要求书1页 说明书4页 附图3页

(54)发明名称

一种语音数码控制中间件装置及系统

(57)摘要

本发明公开了一种语音数码控制中间件装置及系统，该装置包括语音输入设备、上位计算机、语音控制中间件、控制应用单元。所述语音输入设备通过音频接口或者USB接口与所述上位计算机相连接，所述上位计算机通过网口与所述语音控制中间件相连接，所述语音控制中间件通过串口与所述控制应用单元相连接。所述的语音控制中间件包括PIC18F45J10集成单片机、嵌入在单片机内的语音识别程序、语音消息控制程序，其中，所述PIC18F45J10集成单片机接收来自上位计算机从网口发送的语音信号，将之输入语音识别程序，语音识别程序将语音信号识别为语音消息输入语音消息控制程序，语音消息控制程序根据消息的种类，实现不同的逻辑控制。本发明采用语音识别技术，以语音控制代替传统的手动控制，从而增加教师教学的灵活性和控制的便捷性。



CN 106444472 A

1.一种语音数码控制中间件，其特征在于，该装置包括用于接收语音信号的语音输入模块；用于对输入的语音信号进行识别，得到语音消息的语音识别模块；用于根据语音消息控制PowerPoint开启关闭等功能的语音PowerPoint控制模块；用于根据语音消息控制多媒体开启、关闭、快进、后退等功能的语音多媒体控制模块；用于根据语音消息控制开启关闭等功能的语音设备控制模块。

2.根据权利要求1所述的语音输入模块，其特征在于，所述语音输入模块由语音输入设备通过音频接口或者USB接口与上位计算机相连接。

3.根据权利要求1所述的语音识别模块，其特征在于，用于对所述输入的语音信号进行识别，得到语音消息，并根据语音消息的内容将之输入到对应的语音控制模块。

4.根据权利要求1所述的语音PowerPoint控制模块，其特征在于，当语音识别程序识别特定的命令时，语音识别程序将其输入到语音消息控制程序，语音消息控制程序根据消息的不同，从而进行相关PowerPoint控制功能，具体包括开启、关闭、翻页、全屏、隐藏、计时等功能。

5.根据权利要求1所述的语音多媒体控制模块，其特征在于，语音消息控制程序接收来着语音识别程序识别的语音控制多媒体信息，对多媒体进行开启、关闭、快进、后退等控制操作。

6.根据权利要求4所述的语音设备控制模块，其特征在于，语音消息控制程序根据所述的语音消息控制设备，具体的功能包括设备开启、关闭、增大控制量、减少控制量等功能。

一种语音数码控制中间件装置及系统

技术领域

[0001] 本发明涉及语音识别控制技术，具体涉及多媒体教学中利用语音控制命令控制多媒体教室的硬件设备和教学辅助软件。

背景技术

[0002] 如今的教学方式主流是多媒体教学，而在我们日常的教学当中最为常用的就是PowerPoint这一辅助软件。在很多时候，教师授课时需要脱离电教平台，如在教师在黑板板书，这时教师若想要对幻灯片进行操作的话，通常情况下得回到计算机旁点击鼠标或者键盘来操作PowerPoint，这就造成授课的不便。本发明专利实现了语音控制PowerPoint的操作功能，使教师可以脱离电教平台进行授课，同时可以用语音控制教学视频或者音频，从而达到教师教学的最佳效果。除此之外，还可以通过语音中间件对其他串口设备进行控制，例如教室LED灯光、空调、电动窗帘、投影仪等设备，实现语音控制设备智能化、一体化的目的。

[0003] 专利“语音控制的PPT翻页器，申请号20140670519.6”发明了一种语音控制PPT翻页器，教师可以通过语音对PPT进行翻页操作，便于教师在上课环境中通过语音对PPT进行翻页操作，但本发明功能教简单，只能实现“上一页”和“下一页”的翻页操作功能，无法实现全屏播放、首页，末页等更具体的功能；此外，该语音控制系统不具有控制设备的功能。

[0004] 专利“语音控制视频播放器的系统及其方法，申请号201410016388.4”通过将视频播放器和语音程序进行绑定，从而实现用语音控制视频播放器快进、快退、音量控制及播放时间定位等功能。但该发明涉及的是具有语音程序功能的电视机，应用范围较小。且功能设计较简单，无法实现其他电子设备的智能语音控制。

[0005] 专利“一种用扩音器的语音输入控制PPT演示软件的方法，申请号201310181711.9”采用带有语音识别、训练、语音存储功能的专用中央处理器、编制及调制无线发射模块、无线接收解调模块、USB接口控制模块实现远程语音无线控制PPT演示软件。但该发明中，实现语音无线控制PPT演示软件需要依赖扩音器这一设备，若扩音器设备出现故障，则无法实现语音控制PPT演示软件这一功能。且该发明不能做到直接用语音随时控制PPT演示软件，交互体验较差。

[0006] 上述发明中，语音控制装置功能较单一，只能实现教学设备或者教学软件中某一项特定的语音控制，无法实现更多实用、复杂的功能。同时，部分发明涉及的对象是具有语音程序功能的电视机，无法在多媒体教学实现语音智能控制。

发明内容

[0007] 本发明目的在于克服现有技术的不足，解决教师在多媒体课室进行教学活动时，受限于手动控制各种设备的问题。本发明专利通过网络远程监控，利用语音命令控制多媒体教室的硬件设备(空调、电灯、投影屏幕、学生主机等)和教学辅助软件(PowerPoint、播放器、Author Ware等)。

[0008] 为解决上述技术问题，本发明采用如下的技术方案：

[0009] 一种语音数码控制中间件，其特征在于，语音输入设备通过音频接口或者USB接口与上位计算机相连接，上位计算机通过网口与语音控制中间件相连接，语音控制中间件通过串口与控制应用单元：空调、电动窗帘、灯光等相连接。语音控制中间件包括用于接收语音信号的语音输入模块；用于对输入的语音信号进行识别，得到语音消息的语音识别模块；用于根据语音消息控制PowerPoint开启关闭等功能的语音PowerPoint控制模块；用于根据语音消息控制多媒体开启、关闭、快进、后退等功能的语音多媒体控制模块；用于根据语音消息控制开启关闭等功能的语音设备控制模块。

[0010] 所述的语音输入模块，其特征在于，由麦克风对来自用户发送的语音指令进行采样，得到语音信号，将所述语音信号输入语音中间件模块。

[0011] 所述的语音识别模块，其特征在于，用于对所述输入的语音信号进行识别，得到语音消息，并根据语音消息的内容将之输入到对应的语音控制模块。它包括以下几个部分：

[0012] (1)PIC18F45J10集成单片机。本模块通过串口通信技术从而实现上位计算机与PIC18F45J10单片机进行通信。

[0013] (2)语音识别程序。在本发明专利中，语音识别程序基于Microsoft Speech SDK进行设计和开发，包括以下主要过程：

[0014] 初始化COM端口。调用CoInitializeEx函数进行COM初始化；

[0015] 创建识别引擎；

[0016] 创建识别上下文接口。调用ISpRecognizer::CreateRecoContext函数创建识别上下文接口(ISpRecoContext)；

[0017] 设置识别消息。调用SetNotifyWindowMessage函数；

[0018] 设置感兴趣的事件；

[0019] 创建语法规则。先调用ISpRecoContext::CreateGrammar函数创建语法对象，然后加载不同的语法规则，然后调用ISpRecoGrammar::LoadCmdxxx加载语法；

[0020] 获取识别消息；

[0021] 释放创建的引擎、识别上下文对象、语法等。调用Release函数

[0022] (3)匹配数据字典XML命令实现。由于SDK提供的接口方法存在每次只能识别一次语音命令的局限，而本系统需要在长时间监听用户的语音命令并且需要在监听命令的同时进行动作判断，单线程的程序只能专注于语音命令监听或相应动作标记的检测，所以在本设计中使用多线程与计时器相结合的方式解决系统的需求。

[0023] 所述的语音PowerPoint控制模块，其特征在于，本设计通过添加Microsoft PowerPoint类型库msppt.olb，从而实现程序自动化操作PowerPoint，其中提供的可控制动作包括启动\结束PowerPoint放映、翻到上一页\下一页、跳到上一步\下一步、翻到首页\末页等。具体实施过程为：当语音识别程序识别特定的命令时，语音识别程序将其输入到语音消息控制程序，语音消息控制程序根据消息的不同，从而进行相关PowerPoint控制功能，具体包括开启、关闭、翻页、全屏、隐藏、计时等功能。

[0024] 所述的语音多媒体控制模块，具体实现如下：

[0025] (1)在线程监听中获取用户指令；

[0026] (2)实现具体功能，本设计采取抓取播放器窗口句柄，发生虚拟按键进行控制，对多媒体进行开启、关闭、快进、后退等控制操作。

[0027] 所述的语音设备控制模块，其特征在于，本模块在Visual C++6.0集成开发环境中利用MSComm控件开发串口通信程序，实现计算机通过串口智能控制灯泡、电动窗帘、空调等功能。具体的功能包括设备开启、关闭、增大控制量、减少控制量等功能。

附图说明

[0028] 图1是本发明中语音数码控制中间件的系统结构示意图。

[0029] 图2是本发明中语音数码控制中间件各模块实现的流程图。

[0030] 图3是PIC18F45J10集成单片机与PC机的连接示意图。

[0031] 图4是RS232串口接线图。

[0032] 图5是基于Speech SDK涉及的语音识别流程图。

[0033] 图6是多媒体播放器设计示意图。

具体实施方式

[0034] 下面结合附图及具体实施例对本发明进行更加详细与完整的说明。可以理解的是，此处所描述的具体实施例仅用于解释本发明，而非对本发明的限定。

[0035] 参见图1，所述的语音数码控制中间件装置及系统包括语音输入设备、上位计算机、语音控制中间件、控制应用单元。

[0036] 参见图1，所述语音输入设备1通过音频接口或者USB接口与上位计算机2相连接，所述上位计算机2通过网口与语音控制中间件3相连接，所述语音控制中间件3通过串口与控制应用单元：空调5、电动窗帘6、灯光7等相连接。

[0037] 参见图1，所述的语音控制中间件包括嵌入在PIC18F45J10集成单片机内的语音输入模块、语音识别模块、语音PPT控制模块、语音多媒体控制模块、语音设备控制模块，其中：

[0038] 参照图2，所述语音输入模块接收来自上位计算机从网口发送的语音信号，将之输入语音识别模块，语音识别模块将语音信号识别为语音消息。当语音消息为PPT控制消息时，转发至语音PPT控制模块，实现对PPT开启、关闭、翻页、全屏、隐藏、计时等控制；当语音消息为多媒体控制消息时，转发至语音语音多媒体控制模块，实现对多媒体播放器进行开启、关闭、快进、后退等控制；当语音消息为设备控制消息时，转发至语音设备控制模块，实现对灯泡、电动窗帘、空调等设备的开启、关闭、增大控制量、减少控制量等控制。

[0039] 所述的语音识别模块用于对所述输入的语音信号进行识别，得到语音消息，并根据语音消息的内容将之输入到对应的语音控制模块。

[0040] 参照图3，本模块通过串口通信技术从而实现上位计算机与PIC18F45J10单片机进行通信。具体实现如下：

[0041] (1)PIC单片机连接线为三线控制接发；

[0042] (2)参照图3、4，PIC单片机通过RXD与TXD连接，GND与GND连接，将单片机和PC机相连；

[0043] (3)利用单片机烧录器将写好的语音识别程序烧录在PIC单片机上。

[0044] 参照图5，在本发明专利中，语音识别程序基于Microsoft Speech SDK进行设计和开发，具体实现如下：

[0045] (1)在Microsoft Speech SDK中利用命令识别模式，数字字典技术进行命令的编

写；

[0046] (2)Microsoft Speech SDK开发步骤：

[0047] 初始化COM端口。一般在CWinApp的子类中，调用CoInitializeEx函数进行COM初始化；

[0048] 创建识别引擎。微软Speech SDK 5.1支持两种模式的：共享(Share)和独享(InProc)。一般情况下可以使用共享型，大的服务型程序使用InProc；

[0049] 创建识别上下文接口。调用ISpRecognizer::CreateRecoContext创建识别上下文接口(ISpRecoContext)；

[0050] 设置识别消息。调用SetNotifyWindowMessage告诉Windows哪个是我们的识别消息，需要进行处理；

[0051] 设置感兴趣的事件；

[0052] 创建语法规则。首先利用ISpRecoContext::CreateGrammar创建语法对象，然后加载不同的语法规则，然后利用ISpRecoGrammar::LoadCmdxxx加载语法；

[0053] 获取识别消息，进行处理。截获识别消息(WM_RECOEVENT)，然后处理。识别的结果放在CSpEvent的ISpRecoResult中。

[0054] 释放创建的引擎、识别上下文对象、语法等。调用相应的Release函数即可。

[0055] (3)匹配数据字典XML命令实现。由于SDK提供的接口方法存在每次只能识别一次语音命令的局限，而本系统需要在长时间监听用户的语音命令并且需要在监听命令的同时进行动作判断，单线程的程序只能专注于语音命令监听或相应动作标记的检测，所以在本设计中使用多线程与计时器相结合的方式解决系统的需求。具体的设计如下：

[0056] 线程函数：把初始化组件、创建语音引擎、加载数据字典、和监听用户语音口令的代码写进线程函数中，当用户的口令与speech sdk后缀名为xml语法库里面定义的语法相同时，则改变相应的动作标记(动作标记为work)。

[0057] 计时器函数：由于该软件需要每时每刻的检测用户输入的命令是否与语法里面相应的语法是否对应，所以在本设计的主线程中定义一个计时函数，定义相隔0.1s进行一次动作标记的检测从而调用相应的函数，进行相应的动作。

[0058] 参照图6，用于语音控制多媒体的是一个辅助的小软件，不同于其他一般的运用程序，运用于教学时ppt播放、视频播放等。当页面显示1时，表示打开受控的PowerPoint演示文稿；当页面显示2时，显示当前是监听状态；当页面显示3时，显示当前接受到的语音命令；当页面显示4时，表示在放映演示文稿时可进行计时或使用语音命令切换显示当前系统时间；当页面显示5时，显示当前系统音量，也可通过语音命令或手动方式调节。

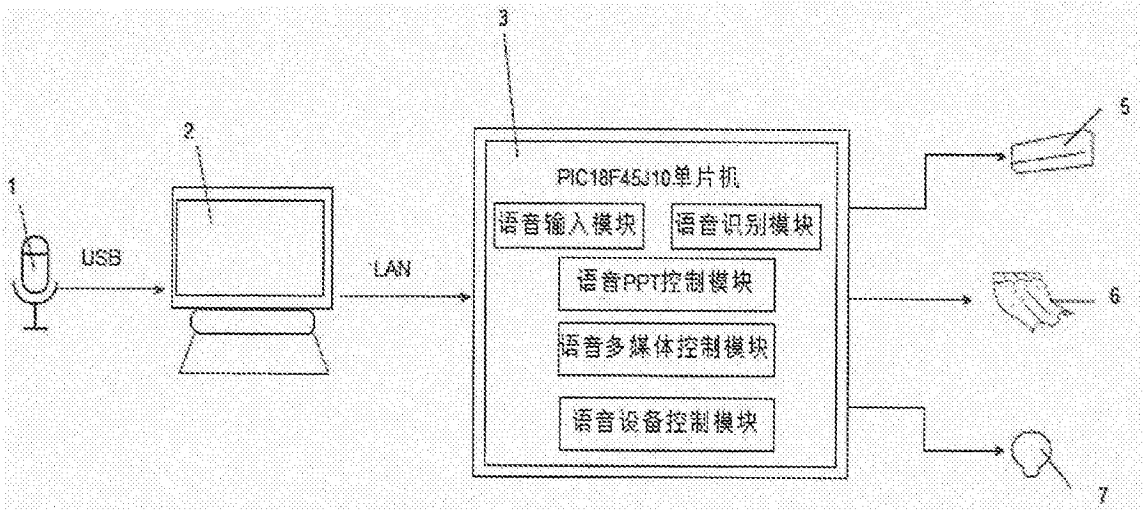

1
2
3
USB
LAN
PIC18F45J10单片机
语音输入模块
语音识别模块
语音PPT控制模块
语音多媒体控制模块
语音设备控制模块
5
6
7

图1

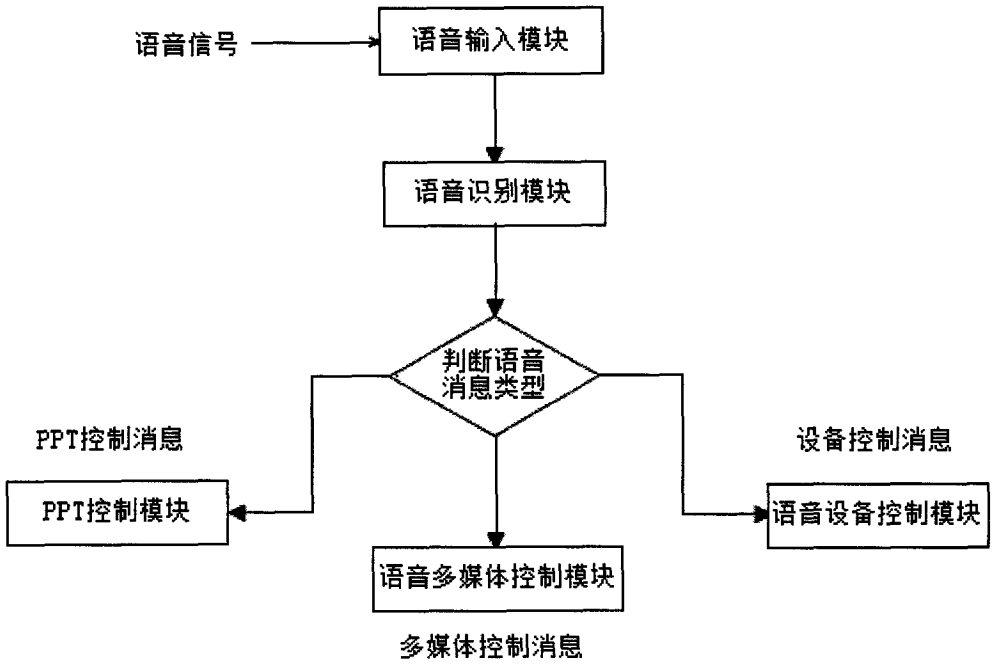

语音信号
语音输入模块
语音识别模块
判断语音消息类型
PPT控制消息
PPT控制模块
设备控制消息
语音设备控制模块
语音多媒体控制模块
多媒体控制消息

图2

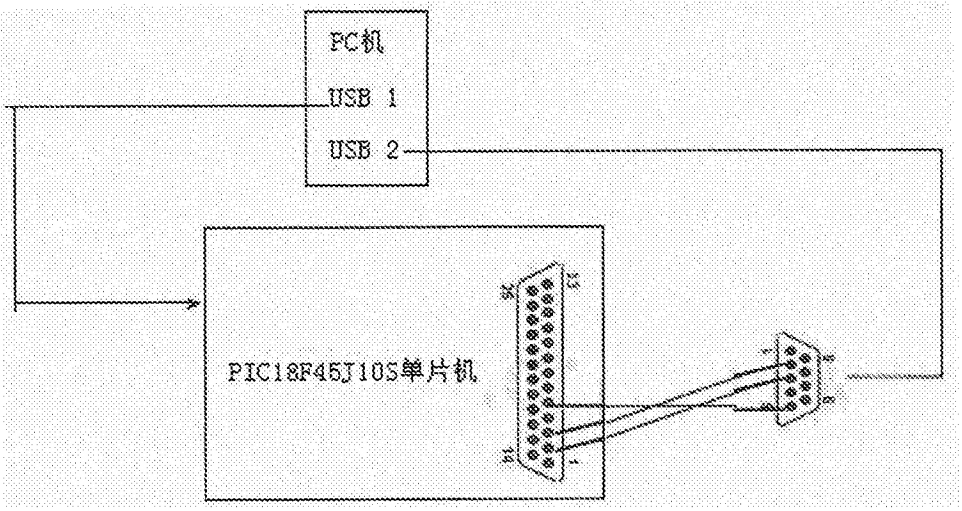


图3

RS232串口接线图

9针串口（DB9）			25针串口（DB25）		
针号	功能说明	缩写	针号	功能说明	缩写
1	数据载波检测	DCD	8	数据载波检测	DCD
2	接收数据	RXD	3	接收数据	RXD
3	发送数据	TXD	2	发送数据	TXD
4	数据终端准备	DTR	20	数据终端准备	DTR
5	信号地	GND	7	信号地	GND
6	数据设备准备好	DSR	6	数据准备好	DSR
7	请求发送	RTS	4	请求发送	RTS
8	清除发送	CTS	5	清除发送	CTS
9	振铃指示	DELL	22	振铃指示	DELL

图4

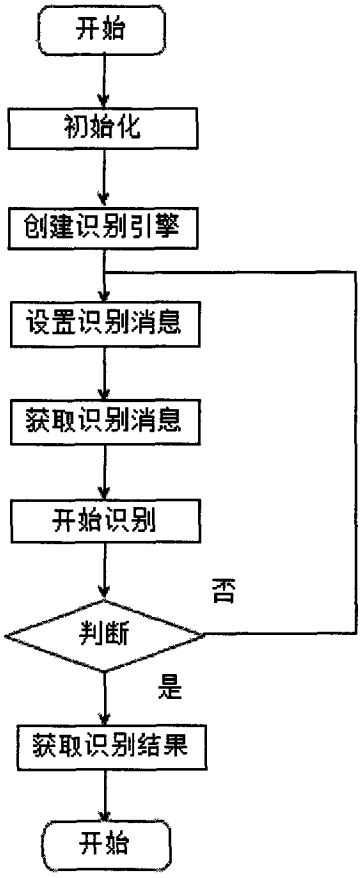


图5

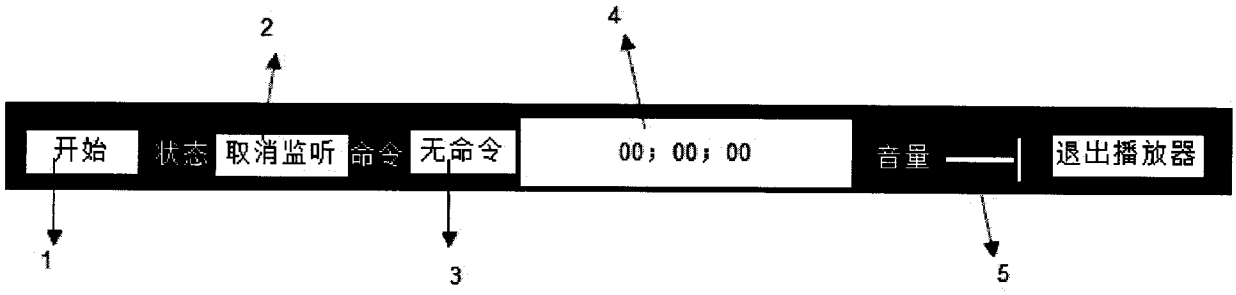


图6