

全程中文语音技术 及其在网络多播通信中的应用

鄢 笠 李 祥 (贵州大学计算机软件与理论研究所 550025)

摘要: 语音应用(VAP)技术是基于语音进行处理的技术,微软近来发布的Speech SDK5.0是基于COM的全程语音开发工具;多播通信是一种重要的网络通信技术;本文讨论与多播通信相关的TCP/IP协议,研究了全程中文语音技术在多播通信中的应用,设计并实现了一个中文语音多播通信实例,本实例可用于多播语音聊天、多播会议等领域。

关键词: VAP SpeechSDK SAPI5.0 WINSOCK 多播 广播 TCP/IP

随着语音技术的不断完善和Internet的发展,网络通信的最新技术和语音技术的接合应用已成为一大热点。本文在概述了语音技术和多播通信技术的基础上,具体设计了一个用多播实现的中文语音聊天室实例。一方面利用多播通信特别适合高带宽的特点,高效的在网络中传输字符数据,一方面利用语音技术在通信的各个端点的主机上,将接收的文本数据转化为声音。

1 全程中文语音技术

1.1 语音技术概述

语音应用(VAP)技术主要是基于语音进行处理的技术,主要包括:语音识别技术和语音合成技术等等。在语音识别技术中,要求根据输入的语音,识别出其代表的具体语义;语音合成技术,则要求根据文字资料等内容,拼读成语音信号,使人们通过“听”就可以明白信息的内容。例如,通过语音识别技术,计算机可以“听懂人类的语言,就像区别出您想买的是“香蕉”还是“苹果”,然后为您继续服务。语音合成技术,则使计算机具有了“说”的能力,能够将信息“读”给人类听,这样,您就可以让计算机给您“读”小说了,而让眼睛休息

一下。

我国全程语音技术的研究最早开始于二十世纪八十年代。经过二十多年的发展。全程语音技术已发展到较成熟的阶段。并且已经被运用到相关的应用领域。国外的一些大的公司,例如IBM、APPLE、AT&T、SIEMENS, Microsoft, 贝尔等。在全程中文语音技术方面已有相关的产品和开发工具。

1.2 语音开发工具

Microsoft Speech SDK 是微软公司提供的全程语音开发工具包。其为开发 Windows 环境的语音应用程序提供了极大的方便。SDK主要包括了WIN32语音编程接口(SAPI),连续的语音识别引擎,文本转换为语音的引擎(TTS---又称为语音合成引擎)等。SDK目前的版本是5.0。SDK5.0与SDK4.0的最主要的区别是,除了SDK5.0在功能和使用等方面较之以前的版本更强大和方便外,最为重要一点就是SDK5.0提供了中文,日文的全程语音引擎,从而支持中文和日文。

Microsoft Speech SDK 是基于COM技术(组件对象模型技术),SDK利用COM技术封装了复杂的实现过程。为开发人员提供了一系列功能实现接口。其中有音频接口,事件

接口,语法编辑接口,词典接口,资源接口,语音识别接口,语音合成接口。除了上述的方法接口外,SDK还提供了一系列属性,开发人员可以用其设定相应的属性,例如,音量的大小,发音的速度,声音的频率等。在使用SDK时,大致上分为以下几个步骤,首先初始化COM库,其次建立TTS对象,进行发声动作,然后处理SAPI事件,设定或改变相关属性。

2 单播与广播的概念

2.1 单播

单播是将数据报发送到网络中的指定的主机,既单播IP数据报只能有目的IP地址指定的主机接收。TCP,UDP都可以实现单播

2.2 TCP/IP网络广播(Broadcast)及其危害

广播是将数据报发送到网络中的所有主机(通常是本地相连的网络)。广播包括四种类型:受限的广播,指向网络的广播,指向子网的广播,指向所有子网的广播。在网络中,过度使用广播技术,极易造成网络带宽的大幅被占用。从而影响整个网络的通信效率。

3 多播技术分析

3.1 多播的含义

多播也称为多点传送 (Multicasting), 是一种让数据从一个成员送出, 然后复制给其他的多个成员的技术。是与广播 (Broadcasting) 相对应的。多播则正好弥补了广播的不足。具体表现在: 对一个网络中的各个工作站而言只有上面运行的进程表示自己“有兴趣”, 多播数据才会复制给它们, 这是通过网卡及设备驱动程序对收到的一个帧进行帧过滤加以实现的。并且进程可以随时加入或撤出多播通信过程。因此, 多播可以有效地减轻网络通信的负担, 避免资源的无谓浪费。

并非所有的协议都支持多播通信。对WIN32平台而言只有两种可以从WINSOCK内访问的协议(IP, ATM)提供了对多播通信的支持。目前, Windows CE2.1, Windows 98, Windows 2000, Windows NT 都提供了对IP多播的支持, 而只有Windows 98, Windows 2000才提供了对ATM多播的支持。

3.2 多播组地址

多播组地址: 属于D类IP地址, 不像其他三类IP地址(A、B、和C)。多播组地址, 它包括为1110的最高4Bit和由剩余的低28位Bit构成的多播组号。它们通常可表示为点分十进制数, 范围从224.0.0.0到239.255.255.255 但是其中有一些地址是为特殊用途而保留的, 用户不要使用。例如: 224.0.0.1代表“该子网内的所有系统组”, 224.0.0.2代表“该子网内的所有路由器组”等。

3.3 多播组地址到以太网地址的转换

IANA (Internet 号分配机构) 拥有一个以太网地址块, 即高位24Bit为00: 00: 5E(十六进制表示)。这意味着

该地址块所拥有的地址范围从00: 00: 5E:00:00:00到00: 00: 5e:FF:FF:FF (注: 以太网地址用48Bit表示)。IANA 将其中的一半分配为多播地址。为了指明一个多播地址, 任何一个以太网地址的首字节必须是01, 这意味着IP多播相对应的以太网地址范围从01: 00:5E:00:00:00到01: 00: 5E:FF:FF:FF。这种地址分配将使以太网多播地址中的23Bit与IP多播组号对应起来, 通过将多播组号中低位23Bit映射到以太网地址中低位23Bit实现。

由于多播组号中的最高5Bit映射过程被忽略, 因此每个以太网多播地址对应的多播组号不是唯一的。32个不同的多播组号被映射为一个以太网地址。因此, 在驱动程序或IP层就必须对数据报进行过滤。因为网卡可能接收到主机不想接收的多播数据帧, 另外如果网卡不提供足够的多播数据帧过滤功能, 设备驱动程序就必须接收所有的多播数据帧, 然后对它们进行过滤。

3.4 多播的特征

多播通信具有两个层面的重要特征: 控制层面和数据层面。其中“控制层面”(Control Plane)定义了组成员的组织方式; 而“数据层面”(Data Plane)决定了在不同的成员之间, 数据如何传送。这两方面的特征即可以是“有根的”(Rooted), 也可以是“无根的”(NoRooted)。在一个“有根的”控制层面内, 存在着一个特殊的多播组成员, 叫做C_Root(控制根或根接点)。而剩下的每个组成员都叫做C_Leaf(控制叶或叶接点)。对“无根的”控制层面而言, 只存在叶接点。“有根的”控制层面和数据层面的行为特征是: 根接点, 负责多播组的建立, 涉及同任意数量叶接点的连接。叶节点可以在某一时刻申请加入一个特定

的多播组。数据传送只能在根结点和叶节点之间进行。根节点将数据一次性多播传给每个叶节点, 叶节点只能将数据传向根节点。ATM多播具有上述特征。“无根的”控制层面和数据层面的行为特征是: 只存在叶节点, 它们可以任意加入一个多播组。从一个叶节点发送的数据会蔓延到每一个叶节点(包括它自己), 无论最开始是谁发出的数据。IP多播具有上述特征。

4 基于多播的中文语音聊天室应用程序的设计说明

4.1 软件功能

该应用程序利用多播通信实现了网上传输字符, 当接收方的主机接收到传输过来的字符, 利用语音合成技术。将字符转化为声音, 从而在安装该应用程序的计算机之间, 实现中文语音聊天室。

4.2 软件设计说明

该程序中利用了一个辅助线程来实现多播数据的接收, 当接收到字符数据后, 建立TTS对象, 进行发声动作, 然后处理SAPI事件, 设定或改变相关属性。代码详见源程序。

4.3 软件设计要点

该程序是利用MFC实现的并且是基于单文档的应用程序, 为了方便实现多播, 在该程序里定义了一个多播类CmulticastSocket。应注意的是在使用TTS时, 应该把单字节的字符类型转化多字节类型;

4.4 软件运行环境说明

该程序在VC++6.0中编译, 调试通过。运行环境是Windows 2000; ■