基于 Speex语音引擎的 VoIP系统设计与实现*

谢晓钢, 蔡 骏, 陈奇川, 欧建林
(厦门大学 计算机科学系, 福建 厦门 361005)

摘 要: 论述了一套基于 Speex语音引擎和 RTP的 VoIP系统设计和开发,介绍了该系统服务器端和客户机端的软件实现。该系统具有点对点通信、算法延时小、丢包补偿和延时补偿性能好等特点,并具有多方通话功能。性能对比实验表明,该系统的通话质量优于几套流行的开源 VoIP软件,能满足实际应用的要求。

关键词: 基于 IP网络的语音传输; Speex; 实时传输协议; 多方通话
中图分类号: TP393 **文献标志码**: A **文章编号**: 1001-3695(2007)12-0320-04

Design and implementation of VoIP system based on Speex codec

XIE Xiao-gang, CAI Jun, CHEN Qi-chuan, OU Jian-lin
(*Dept. of Computer Science, Xiamen University, Xiamen Fujian* 361005, *China*)

Abstract: The paper proposed the design and software implementation of a VoIP system, both on the server side and the client side. The VoIP system was written with Speex and RTP, and deployed for Windows platform. It featured peer-to-peer communication, low algorithmic latency, low packet loss, and effective compensation for jitter. It had a full range of practical functions, such as conference calling. Speech quality assessment has been done by comparing and contrasting the performance of the system to those of several other VoIP systems. Testing results show that the system has a satisfactory voice quality for practical use.

Key words: VoIP(voice over Internet protocol); Speex; RTP; voice conferencing

网络电话 VoIP[1]是基于 IP网络的语音传输技术,它将语音的模拟信号转换成数字信号并在 Internet上进行传输。简单地说,它是首先通过一连串的 A/D转换、编码、压缩、打包等程序来处理语音信号,并将语音数据在 IP网络上传输到目的端,然后再经由相反的一连串程序,将语音数据还原成原来的语音信号播放给接听者。VoIP目前已被广泛地应用于全球 IP互联的 Internet环境中。它与模拟语音通信系统相比具有抗干扰性强、保密性好、易于集成、成本低廉等特点,并可开发出更多的增值业务。然而, IP网络的语音传输质量成为制约 VoIP发展的瓶颈。基于分组交换的 IP网络使得 VoIP系统存在分组延迟、延迟抖动、丢包等问题,使得用户听到的话音会出现不连贯甚至中断的现象。现有的 VoIP系统还难以实现高质量的实时语音通信。如何提高语音通信质量是近年来 VoIP技术研究的一个重要课题。

Speex[2,3]是近年来开发出的一套功能强大的语音引擎,能够实现高质量和低比特率的编码。它不仅提供了基于码激励线性预测 (CELP)算法的编/解码模块,而且在其最新发布的版本中还提供了声音预处理和声学回声消除模块,为保障 IP网络中的语音通信质量提供了技术手段。此外, Speex还具有压缩后的比特率低 (2～44 kbps)的特点,并支持多种比特率。这些特点使得 Speex特别适合 VoIP的系统。

伴随着 VoIP应用的热潮,目前国内外出现了很多关于 VoIP的系统实现,但是大多数系统设计均是基于直接使用 IP地址进行呼叫。这就存在一些弊端,如呼叫方必须事先知道对方的 IP地址,而这无形中降低了系统的可用性。本文在 Speex和 RTP(实时传输协议)[4]的基础上,论述了一套基于客户机/服务器模式的 VoIP系统的设计与开发。该系统克服了上述弊端,提高了 VoIP系统的友好性和可推广性。经过实验测试,本系统即使在网络条件较差的情况下,也能达到较好的语音通信效果,而且支持 3～5人的多方通话。

概述

Speex是近年来开发出的一种基于码激励线性预测算法的开源软件语音引擎。它主要面向 Internet上的语音通信。其主要设计目标是为了提供高质量和低比特率的语音编码。Speex编码支持多种比特率,如 8 kHz采样的低比特率 (窄带 2.15～24.6 kbps)、16 kHz采样的中比特率 (宽带 3.95～42.2 kbps)以及 32 kHz采样的高比特率 (ultra-wideband)。Speex提供了大多数别的编/解码器所不具备的技术性能,主要包括:可以在同一个比特流中对语音信号实现窄带 (8 kHz)、宽带 (16 kHz)和超宽带 (32 kHz)的压缩;支持声音强度的立体声编码;具有丢包补偿能力;具有可变比特率 (variable bitrate, VBR)特性,编/解码器可以在任意时刻动态地改变语音的比特率;能实

收稿日期: 2006-10-06; **修返日期**: 2006-12-10 **基金项目**: 厦门大学"985工程"二期信息创新平台资助项目 (0000-X07204)

作者简介:谢晓钢 (1982-),男,硕士研究生,主要研究方向为计算机多媒体通信;蔡骏 (1966-),男,副教授,硕导,博士,主要研究方向为语音识别、计算机多媒体通信等 (mikecai@xmu.edu.cn);陈奇川 (1982-),男,硕士研究生,主要研究方向为计算机多媒体通信、语音识别;欧建林 (1983-),男,硕士研究生,主要研究方向为人工智能、自然语言处理.

现语音活动检测(voice activity detection, VAD);能实现声音的 DTX(discontinuous transmission,不连续传输),当背景噪声稳定时,可以完全停止声音数据包的传送;具有语音处理的定点数计算功能(正在开发中);具有声学回声消除功能。

Speex除了编/解码模块外,还包括噪声消除、静音抑制和自动增益控制等预处理模块以及回声消除模块。正是由于其完备的功能和优良的性能,Speex受到了许多 VoIP开发者的关注。2006年 9月发布的最新版本 Speex 1.2 beta1在声音处理性能上得到了进一步的改善。语音编码和解码的质量有了明显的提高。软件运行时的内存使用有了大幅度的降低,定点计算情况下窄带编/解码的内存开销下降为不到原来的一半,只需占用不到 6 KB的 RAM空间,而且 CPU的占用率也有了下降,使 Speex在诸如 PDA等便携式语音通信设备中的应用成为可能。

Speex提供了一系列调用其各个功能模块的 API,使得开发者可以很方便地使用这些功能模块来进行 VoIP系统的编程开发。各模块的 API在文献[3]中有详细描述。在新版本的 Speex中,回声消除模块的 API有了改进,开发者能更方便地调用这个模块。此外,抖动缓冲区(jitter buffer)模块不再专门针对 Speex编码,从而可以应用于其他编码格式的系统。这些特点都为 Speex在 VoIP系统中的应用奠定了基础。

系统的功能模块结构

本文所述的 VoIP系统采用 Speex编码。整个系统由服务器端和客户端两部分构成。系统总体框架如图 1所示。

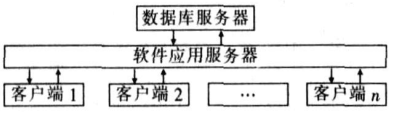

图 1　系统总体框架

数据库服务器上保存了所有已注册用户的资料。每个用户首先要通过客户端登录到服务器上,并获取在线好友的列表,然后才能对在线的好友进行语音呼叫。客户端之间的通话采用点对点的方式进行。这就避免了由于语音包经过服务器中转所造成的过大延迟。

. 服务器端的设计与实现

服务器端除了需要维护一个数据库以保存用户信息外,还要对各个客户端进行指挥和管理。

服务器一旦启动服务,则开始侦听用户请求。服务器收到客户端发来的消息时,首先,发回确认信息;然后,建立一个独立线程来处理接收到的数据。在此线程中,按照接收到数据的类别进行相应的处理。如有需要,服务器会向用户发送处理的结果,成功或失败的消息。处理完毕,此线程结束。这样,可以实时接收每个用户的请求,不会因为处理某一个用户的请求,而忽略了其他用户。服务器端的工作流程如图 2所示。

. 客户端的设计与实现

客户端的功能可以分为两部分:一部分是与服务器端进行交互,从服务器上获取相关信息;另一部分是完成不同客户端之间的点对点通话。其中第二部分是 VoIP系统的核心。

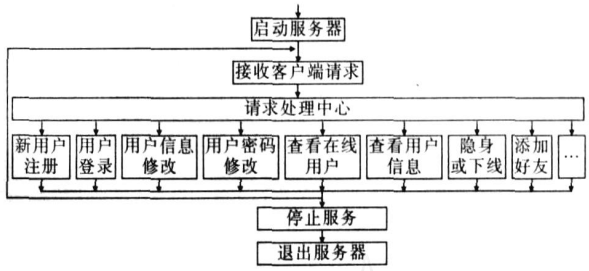

图 2　服务器工作原理

客户端程序启动后,若有本地用户信息,则加载本地用户信息,并显示登录窗口;若没有,则显示用户注册窗口(在登录窗口中,也可以选择用户注册)。登录时,可选择是否隐身,并发送账号和密码到服务器进行验证,如正确则成功登录;否则失败。当成功登录进入系统后,服务器将发送用户的全部好友信息以及在线好友信息(包含好友的 IP和端口号)。在好友列表中,在线者将以高亮度显示,并处在列表的顶部;不在线者将以灰色显示。如果有好友上线,系统会立即通知用户;好友下线,用户也可以在好友列表中看到。用户可以接收到好友发来的语音通话的请求,同时根据用户的操作,客户端可以向好友给出回应,接听或拒绝。系统还具有查看好友信息、查看在线用户、查找用户和修改个人信息等功能。若用户选择注册,在填写注册信息后,信息被发送到服务器;服务器登记用户注册信息后则会发送回由系统自动分配的账号;然后用户就可以成功登录系统。客户端的工作流程如图 3所示。

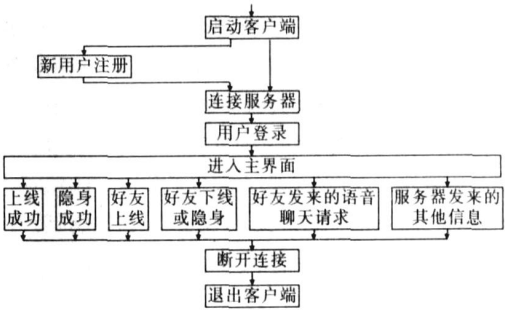

图 3　客户端工作流程

点对点通话模型

最基本的 VoIP系统应该包含录音模块、放音模块、编码模块、解码模块、接收模块和发送模块。本系统除了这些基本模块之外,还在编码之前增加了 Speex的回声消除和预处理模块。

本系统的运行期结构(run-time architecture)由六个线程组成,如图 4所示。其中录音线程负责从声卡采集声音数据并将其放到编码缓冲区中等待编码模块使用。编码线程每次从编码缓冲区中取出一帧数据进行 Speex的回声消除、预处理以及编码,并将编码后的数据放入发送缓冲区。发送线程每次从发送缓冲区中取出编码后的数据进行打包(RTP包),并且发送给通话的另一方。接收线程负责从网络上接收对方发来的 RTP语音包。由于每个 RTP包头中有包的序号信息,接收线程依据包的序号把声音数据依次放入抖动缓冲区中,以达到正

确排序和消除抖动的目的。解码线程负责从抖动缓冲区中取出接收到的语音数据进行 Speex的解码运算,并将解码后的语音数据存入播放缓冲区。播放线程从播放缓冲区中取出语音数据,送至放音设备播放。

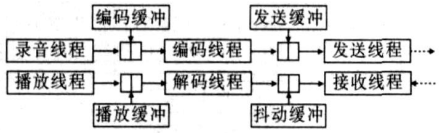

图 4 六个线程运行期结构

上述六线程的运行期结构能够使系统各模块具有高度的并行性。这对于 VoIP这种实时性要求很高的应用来说十分重要。例如,当接收线程收到一个包后,它把解码的任务交给解码线程,然后又可以去等待或处理接收其他的数据包。因此对数据包的接收响应较快。

上述通话模型的实现采用了 RTP和多方通话机制[5,6]。RTP是由 IETF的 AVT(Audio Video Transmission)小组开发的,1996年成为 RFC正式文档,为 IP网上语音、图像、传真等数据等多种需实施传输的媒体数据提供点到点和点到多点的端到端的传输功能。RTP实际上包含两个相关的协议 ——RTP和 RTCP。前者用于传送实时数据;后者实现实时传输控制。RTP本身不提供任何保证实时传送数据和服务质量的能力,而只是提供负荷类型指示、序列号、时戳、数据源标志等信息,接收端根据这些信息重新恢复正确的数据。RTCP是用来提供 RTP数据传输质量反馈的,同时可以在会议业务中传送与会者的信息。

本系统采用了点对点的多方通话模型,即为每一个通话者都建立一个 RTP会话,在语音通话前,要通过一条主 RTP线路建立连接,它的端口被记录到服务器上。用户上线后,客户端可从服务器上获得各在线好友的 RTP端口号。建立连接的过程主要是协调好双方将要在哪个端口上创建 RTP会话进行通话。断开时,通过各自 RTP的 BYEPacket结束会话。建立连接的过程如图 5所示;多方通话模型如图 6所示。

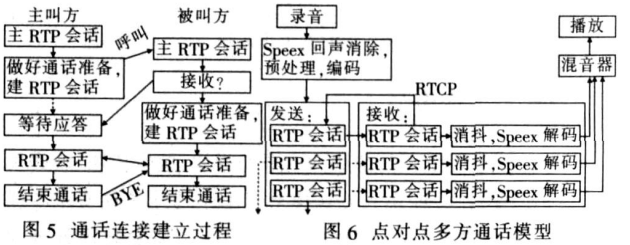

图 5 通话连接建立过程 图 6 点对点多方通话模型

图 5中,主叫方先做好通话准备,并创建一个将用来与对方通话的 RTP会话。把这个端口号加到呼叫包中周期性地发送出去,直到对方应答。被叫方收到呼叫包后,如果接收呼叫,就创建一个用来通话的 RTP会话。同时从接收到的呼叫包中得到对方用来通话的 RTP端口;然后把对方 IP和这个端口所表示的 RTP地址添加到组播地址(在单方通话时实际上是进行单播),并开始通话。主叫方用来通话的 RTP会话第一次收到对方发来的 RTP包时,也把对方的用来通话的 RTP地址加到组播地址,开始通话。当用户要结束通话时,客户端只要发送 RTP提供的 BYEPacket包即可结束通话。

在图 6所示的多方通话模型中,录音线程每录完一帧后,就把这帧发给编码线程。在编码线程中,对语音帧进行 Speex回声消除以及噪声消除、自动增益、静音抑制等预处理,并进行编码;然后将编码后的数据传送给发送线程。在发送线程中,有多个 RTP会话;线程调用每个会话的 SendPacket()函数,向每个 RTP会话方发送语音包。在接收线程中,每个 RTP会话从各自的对话方接收语音包放入抖动缓冲区,经过 Speex解码运算后,计时器周期性地从每个 RTP会话的抖动缓冲区中取出一帧进行混音,然后播放。

RTP的接收反馈 RTCP包内包含了近期内发送的数据包的丢失率。接收方在收到 RTCP包后,通过丢包率就可以知道对方的接收情况,据此可以采取相应的丢包补偿措施。本系统能够根据丢包率适当地发送一些语音冗余包,保证了语音质量。

系统测试

目前已有不少 VoIP系统在 Internet上获得了应用。Skype[7]最具代表性,并且是目前公认最好的商用 VoIP系统。iHU[8]和 NetTalk[9]则是具有代表性的开源 VoIP系统。本系统在 PC上进行了性能测试,并与上述几套系统进行了通话话音质量的对比实验。测试机器配置是 Intel P4 1.8 GHz CPU,512 MB内存,操作系统为 Windows XP。网络环境有两种:一种是在 CERNET内进行通信;另一种是在 CERNET与 CHINANET间进行通信。测试结果如表 1所示。

表 1 系统测试与性能对比

VoIP系统	编码方式	回声消除	服务器端	多方通话	通话质量	
					CERNET-CERNET	CERNET-CHINANET
Skype	iLBC	有	有	有	音质很好,偶尔有断断续续的回声	声音偶有中断,但是不影响通话,偶尔有回声
NetTalk	G.729A	无	无	无	声音流畅,但是增益不够,回声明显	声音抖动严重,无法正常通话,回声严重
iHU	Speex	无	无	无	声音流畅,但是有一定程度的失真,回声明显	声音抖动严重,无法正常通话,回声严重
本系统	Speex	有	有	有	音质较好,声音流畅,几乎没有回声	声音偶有中断,但是不影响通话,出现较明显回声

注:这里的回声是指使用音箱时的声学回声。如果使用头戴式耳机,则本系统没有回声出现。

从表 1中可以看出,在网络互联性能较好的情况下(在 CERNET内通信),上述四种 VoIP系统都能达到较好的通话效果。本系统的通话质量与 Skype相比差距不大,只是音质上略逊于 Skype。主要原因在于 Skype采用了 iLBC[10]编码方式,而 iLBC的编码后比特率是 13.3～15.2 kbps,明显高于 Speex在 8 kHz采样下的比特率。因此,Skype音质略优于本系统是以占用更多的带宽为代价的。本系统与 iHU和 NetTalk相比,优势相当明显,主要表现在本系统增加了回声消除和多方通话功能;另外,声音抖动的去除也做得相当好。在网络条件相对较差的情况下(在 CERNET与 CHINANET间进行通信),四种 VoIP系统的通话质量均出现不同程度的下降。但是本系统与 iHU和 NetTalk相比,在声音质量上要好得多,而与 Skype相比主要的不足表现在回声消除方面。本系统采用的是 Speex语

音引擎提供的回声消除模块,在网络延迟较小的情况下,具有比较明显的回声消除效果。但是实验表明,一旦网络延迟较大,Speex的回音消除效果往往不能令人满意。提高回声消除模块的性能是本系统需要进一步改进的地方。

结束语

目前,基于 IP网络的多媒体通信得到了迅猛的发展,特别是 VoIP系统已日益成为 Internet用户进行语音通信的主要手段。资费便宜以及 IP网络的广泛性和可伸缩性,使得 VoIP成为一种可以替代传统的 PSTN电话的通信方式。Speex是近年来出现的功能强大的开源语音引擎;RTP是协议是网络多媒体的核心协议之一。本文论述了一个基于 Speex语音引擎和 RTP的、客户机 /服务器结构模式的 VoIP系统的设计与开发。该系统采用客户机 /服务器模式克服了现有的多数 VoIP系统只能直接用 IP地址进行呼叫的缺陷,提高了 VoIP系统的友好性和可推广性。另外,Speex提供的高质量、低比特率的编码算法以及声音预处理和回声消除功能使得本系统的通话质量优于其他的一些开源 VoIP系统。在不同网络互联环境中对系统进行的测试表明,在网络延迟较小时,本系统回声消除效果明显,但是当网络延迟较大时,声学回声消除的效果还不能令人满意。进一步改善回声消除功能是本系统后续开发的主要任务。

参考文献:

[1] GOODE B. Voice over Internet protocol (VoIP) [J]. Proceedings of the IEEE, 2002, 90 (9): 1495.

[2] XIPH Open Souce Community. Speex: a free codec for free speech [EB/OL]. [2006-09-27]. http://www.speex.org

[3] VALIN J M. The Speex codec manual (version 1.2-beta1) [EB/OL]. [2006-09-27]. http://www.speex.org/docs/manual/speex-manual.pdf

[4] HENNING S, STEPHEN L C, RON F, *et al*. RFC 3550 RTP, a transport protocol for real-time applications[S].

[5] HERSENT O, PETIT J P, GURLE D. Beyond VoIP protocols: understanding voice technology and networking techniques for IP telephony[M]. Chichester: Wiley, 2005: 206.

[6] HERSENT O, PETIT J P, GURLE D. IP telephony: deploying voice-over-IP protocols[M]. Chichester: Wiley, 2005: 108.

[7] 张晋江. Skype再次掀起 VoIP热潮[J]. 通信世界, 2005 (12): 20.

[8] TROTTA M. IhearU (IHU) project homepage[EB/OL]. [2006-10-15]. http://ihu.sourceforge.net/index.html

[9] LUSSIER G A, MORENCY F C. NetTalk project Web page[EB/OL]. [2006-10-05]. http://ntalk.sourceforge.net

[10] The iLBCfreeware.org Project. iLBCfreeware.org project homepage [EB/OL]. [2006-10-05]. http://www.ilbcfreeware.org/index.html

(上接第 295页)综合污染损害率评价模型所得的评价结果与李祚泳的 RPL评价法、倍斜率聚类法即 TSC评价法、模糊综合法即 F法所得的评价结果基本一致,只有在监测点 4略有差异。本文在监测点 4计算的结果 R为 1.29略高于 2级和 3级的分界线,因此此点判定为 2级或 3级均合理。

. 实例

本文根据长春市环境监测中心站提供的资料,以大气中的主要污染物:PM_{10} (可吸入颗粒物)、SO_2、NO_2 三种污染物的浓度值为主要监测对象。由于中国北方的春季具有风沙大、气候变化显著的特点,在此期间大气污染现象较为明显。因此本文选取 2002年 3月 1日 ～15日间实际监测的数据作为数据样本,运用大气质量综合污染损害指数评价模型所得到的评价结果如表 6所示。

表 6　大气质量综合污染损害指数评价模型评价结果

时间	PM_{10}/ mg·m^{-3}	SO_2/ mg·m^{-3}	NO_2/ mg·m^{-3}	损害指数 I	本文评价	实际级别
3.1	0.09	0.014	0.018	0.255	良	良
3.2	0.07	0.013	0.016	0.233	良	良
3.3	0.071	0.016	0.018	0.239	良	良
3.4	0.089	0.018	0.019	0.257	良	良
3.5	0.15	0.022	0.023	0.320	良	良
3.6	0.09	0.007	0.02	0.260	良	良
3.7	0.096	0.014	0.029	0.292	良	良
3.8	0.086	0.011	0.021	0.259	良	良
3.9	0.294	0.018	0.023	0.744	轻微良	轻微
3.10	0.118	0.012	0.023	0.297	良	良
3.11	0.069	0.015	0.025	0.256	轻微	良
3.12	0.183	0.044	0.034	0.4110	良	轻微
3.13	0.089	0.021	0.031	0.268	良	良
3.14	0.142	0.041	0.025	0.336	良	良
3.15	0.116	0.021	0.025	0.273		良

表 6的评价结果表明:采用本文的大气质量综合污染损害指数评价模型所得评价结果与实际评价结果基本吻合。因此该评价模型用于城市大气污染评价是科学可行的,具有一定的应用前景。

结束语

本文采用一种基于群体的随机全局优化工具即 PSO算法对大气污染损害率的计算公式中的参数进行优化,优化后得到的大气污染损害分指数公式形式简单,计算简便;当污染物种类较多时,其优越性更为明显;本文将大气环境质量的好坏与其所受的损害程度直接相联系,因此使本文的大气质量综合污染损害指数评价法具有更明确的物理意义,原理更加直观。评价级别是以多种大气污染物对大气的损害程度作为划分等级的主要标准,故使本文的评价结果更合理、更准确。

参考文献:

[1] 李祚泳,王钰,刘国东. 大气污染损失率评价模型参数的 GA优化[J]. 环境科学研究, 2001, 14 (2): 7-10.

[2] 钱莲文,吴承祯,宏伟,等. 大气质量评价的污染危害指数法的改进[J]. 福建林学院学报, 2003, 23 (3): 249-252.

[3] 李祚泳,张欣莉,丁晶. 水质污染损害指数评价的普适公式[J]. 水科学进展, 2001, 12 (2): 161-163.

[4] 何斌,谭界忠. 大气环境质量综合评价的污染损失率法[J]. 环境保护, 1998 (9): 21-24.

[5] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proc IEEE Int Conf on Neural Networks. Piscataway, NJ: IEEE Service Center, 1995: 1942-1948.

[6] EBERHART R, KENNEDY J. A new optimizer using particle swarm theory[C]//Proc of the 6th Int Symposium on Micro Machine and Human Science. Nagoya: IEEE Press, 1995: 39-43.

[7] 李爱国,覃征,鲍复民,等. 粒子群优化算法[J]. 计算机工程与应用, 2002, 38 (21): 1-3.