

文章编号:1674-2869(2012)05-0060-04

分布式环境实时语音通讯系统的设计与实现

司聿宣¹, 苏远兴², 杨正芳³

(1. 烟台职业学院艺术设计系, 山东 烟台 264001; 2. 92857 部队, 北京 100161; 3. 91557 部队, 浙江 舟山 316000)

摘 要:在分布式模拟训练系统中, 语音系统的设计通常采用专用硬件的手段实现, 该方法不仅增加了系统设计的费用和复杂度, 也增加了系统硬件组成和布局的难度. 设计了一种分布式环境下的基于软件化的实时语音通讯系统, 利用 DirectX 对语音数据进行采集和回放, 并以网络通信的方式对数据进行打包、传输和解包. 同时, 针对噪声等周围环境和无人说话时产生的冗余数据造成网络通信量增加的问题, 采用离散傅里叶变换对语音信号在频域进行过滤, 有效降低了系统的网络负担. 该系统在某型模拟器中得到了较好的应用.

关键词:分布式仿真; 语音通讯; 网络; 离散傅里叶变换

中图分类号:TP391

文献标识码:A

doi:10.3969/j.issn.1674-2869.2012.05.015

0 引言

随着多媒体技术和网络技术的飞速发展, 基于软件化的手段可实现分布式环境下多点间的语音通信^[1-2]. 但在大型模拟器系统中, 不同位置的训练员和指挥员间的语音通信通常采用专用硬件设备的方法, 不仅增加了系统的设计费用、复杂度和维护的难度, 更增加了系统的硬件组成和布局难度, 尤其对于空间有限的模拟训练系统, 其缺点更为突出. 针对该问题, 笔者综合运用声音处理和网络通信技术, 基于客户机/服务器 (Client/Server, 以下简称: C/S) 的架构构建了一套语音通讯系统, 利用微软推出的 DirectX 提供的与网络音频应用相关的系统开发工具包 (System Develop Kit, 简称: SDK) 实现仿真模拟系统操作人员之间的语音捕捉与回放, 利用网络通信技术实现了语

音的打包、传输与解包, 同时, 针对冗余的语音数据将造成网络负担的问题, 采用离散傅里叶变换 (Discrete Fourier Transform, 以下简称: DFT) 算法在频域上对语音数据进行了过滤. 该方法以软件化的手段替代了传统模拟器中复杂的硬件设计, 在实际工程项目中取得了较好的效果.

1 系统设计

1.1 系统结构

网络语音通讯系统结构如图 1 所示, 系统采用 C/S 架构, 使用了一个高效的服务器程序, 运行在一台具有独立 IP 地址的计算机上, 用来发送并接收语音数据和与客户端进行交互的控制信息, 维护客户端信息及客户端连接信息等功能. 客户端程序主要实现语音数据的录取、过滤和压缩, 发送和接收与服务器交互的控制信令和音频编码数

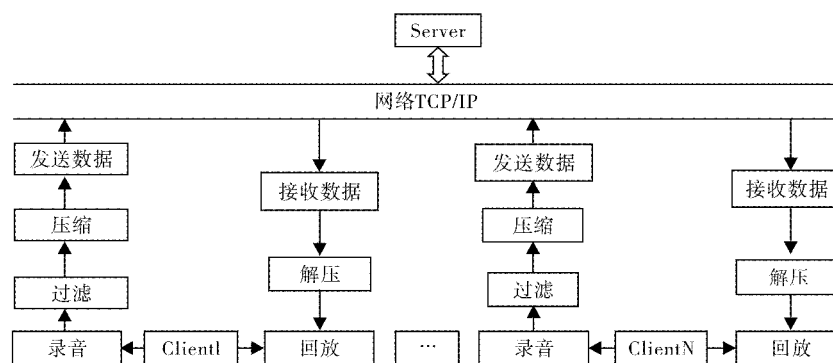


图 1 系统结构

Fig. 1 Structure of system

收稿日期:2012-03-06

作者简介:司聿宣(1981-),女,山东烟台人,助教,硕士.研究方向:语音通信技术.

据,解压语音数据并进行播放.服务器和客户端之间所有的控制信息均采用 TCP/IP 协议传输^[3-4].

每一个客户端独立完成语音数据的采集,并对语音数据进行过滤和压缩,然后将数据发往服务器端并经服务器转发给其他客户端.同时客户端服务端接收来自服务器的语音数据,经过解压后播放.

1.2 声音的捕捉和回放

Microsoft DirectX 软件中的 DirectSound 提供了一个系统捕捉来自输入设备的声音和播放那些贯穿各种各样的使用先进的 3-D 位置特效的回放设备的声音,并且过滤回声、扭曲、反射和其它特效.其优点是速度快、延时短、可控制性强.

1.2.1 声音的捕捉 DirectSound 允许从输入到声卡的设备捕获声音,数据可被捕获为脉冲编码调制或压缩格式,为了立即回放或在文件里存储,DirectSound 提供创建声音捕捉对象 IDirectSoundCapture 的 COM 接口,通过捕捉对象设置录音的格式、建立用来捕捉声音样本的缓冲区对象 IDirectSoundCaptureBuffer、设定声音捕捉长度响应事件 IDirectSoundNotify.在录音缓冲区中充满设定长度的数据后,IDirectSoundNotify 事件响应,这时可以锁定缓冲区并对缓冲区内捕获的声音进行拷贝.声音捕捉流程如图 2 中(a)所示.

1.2.2 声音的回放 DirectSound 提供建立 IDirectSoundBuffer 对象的 COM 接口并设置 Buffer 的大小及播放声音的格式.为了实现音频的连续播放,IDirectSoundBuffer 使用多个次缓冲区,通过设定播放声音长度的响应事件,在每一个次缓冲区均设置一个通知点,每当一个缓冲区的内容播放完后,DirectSound 一方面激活 Notify 机制通知填充缓冲区数据,另一方面同时无间隙播放下一个缓冲区数据,实现音频的连续播放.声音回放流程如图 2 中(b)所示.

2 基于 DFT 的声音过滤

语音通讯的目的是为了传递说话人的信息,但在实际的语音交谈中,有很多时间内可能没有人在说话,此时传输的语音数据只是没有意义的环境音.为了减少系统的网络负担,需要对录取的语音数据进行过滤,避免将只有环境噪音的数据传到网络.

2.1 语音信号 DFT 变换

由于计算机只能处理有限长度的信号,因此对原始语音信号 $x(t)$ 要以一定的采样周期 T 进行截断,即有限化,然后对采样信号 $x(n)$ ($n=0,1,$

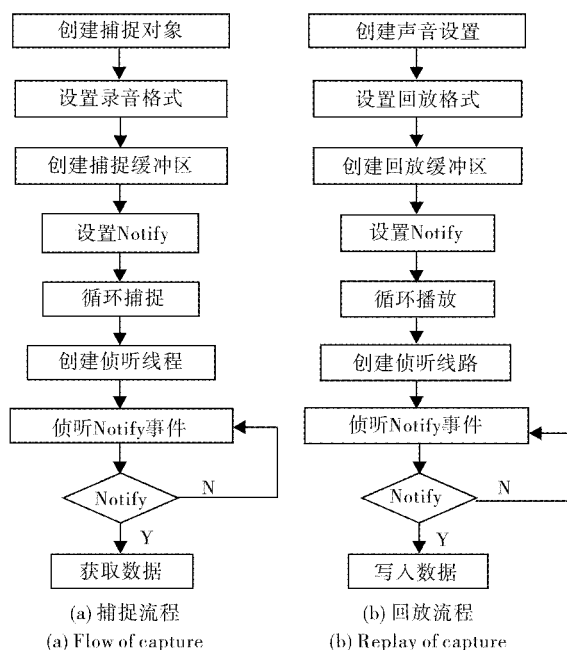


图2 声音捕捉和回放

Fig.2 Catch and replay

$2, \dots, N-1$) 做进一步处理.信号的有限化也称加窗处理,当窗函数在边缘处将信号突然截断,窗外时域信息全部消失.将长序列截短变成有限长度的短序列会出现频率分辨率问题和频谱泄漏问题.频率分辨率问题是由于频谱主瓣过宽、边瓣过大造成的,频谱泄漏是由于边瓣的影响造成的.为了解决这些问题,一般要对信号进行加窗处理,使信号截断的锐角钝化,从而使频谱的扩散减到最小.

经过处理,语音信号就已经被分割成一帧一帧的加过窗函数的短时信号,然后再把每一个短时语音帧看成平稳的随机信号,利用数字信号处理技术来提取语音特征参数.在进行处理时,按帧从数据区中取出数据,处理完后取下一帧.最后得到由每一帧参数组成的语音特征参数的时间序列.

对一帧时间内的语音信号进行 DFT 变换为^[5-7]:

$$X(k) = \sum_{n=0}^{N-1} x(n) W(n) e^{-j2\pi nk/N} \quad (k=0,1,2,\dots, N-1) \quad (1)$$

式中 $W(n)$ 为矩形窗, N 为采样点数. $|X(k)|$ 便是 $x(n)$ 的短时幅度谱估计.把 k 当作频率变量, $20 \log_{10} |X(k)|$ 就是以分贝表示的频谱.为了提高计算 $|X(k)|$ 的速度,可以采用快速傅里叶变换.这要求 N 是 2 的整数次幂,如果 N 不满足要求,可以在数据 $x(n)$ 后补零.

2.2 信号过滤

加窗的信号段经过 DFT 变换后, $|X(k)|^2$ 表示信号在频率点 k/TN 处的功率值,令:

$$P_{\text{ave}} = \frac{2}{N} \sum_{k=0}^{N/2-1} |X(k)|^2 \quad (2)$$

为 $N/2$ 频率采样点的平均功率, 则各个频点功率偏离均值的绝对值之和为:

$$P_{\sigma} = \sum_{k=0}^{N/2-1} ||X(k)|^2 - P_{\text{ave}}| \quad (3)$$

通常电话语音的频率在 $60 \sim 3400\text{Hz}$, 这个频率之外的声音会被滤除. 因此, 可以认为频率不在 $60 \sim 3400\text{Hz}$ 之间的信号为语音通讯系统的噪音. 在计算 P_{ave} 、 P_{σ} 时不用计算这些频点.

在一个特定的环境中, 环境噪声的 P_{σ} 上限值设定为 P_{NMax} , 当 $P_{\sigma} > P_{\text{NMax}}$ 时可以判定有语音信号.

2.3 仿真结果

在实验室录制一段长度为 1.6s 的单通道语音段, 信号采样频率为 8000Hz , 每个采样点的信号幅度用 8 比特表示. 将这段数据分为 50 帧, 每帧数据长度为 256 个采样点, 其时域波形如图 3 中 (a) 所示, 方框内 1 和 2 中的波形为两个频率分别为 30Hz 和 3700Hz 的信号和环境噪声的叠加; 计算每帧采样数据的 P_{σ} , 结果仿真如图 3 中 (b) 所示; 设置语音信号频率范围 $60 \sim 4000\text{Hz}$, 过滤后的 P_{σ} 结果仿真如图 3 中 (c) 所示. 根据对环境噪声的采样计算, 估计 $P_{\text{NMax}} = 30$, 利用式 $P_{\sigma} > P_{\text{NMax}}$, 过滤后的信号波形如图 3 中 (d) 所示.

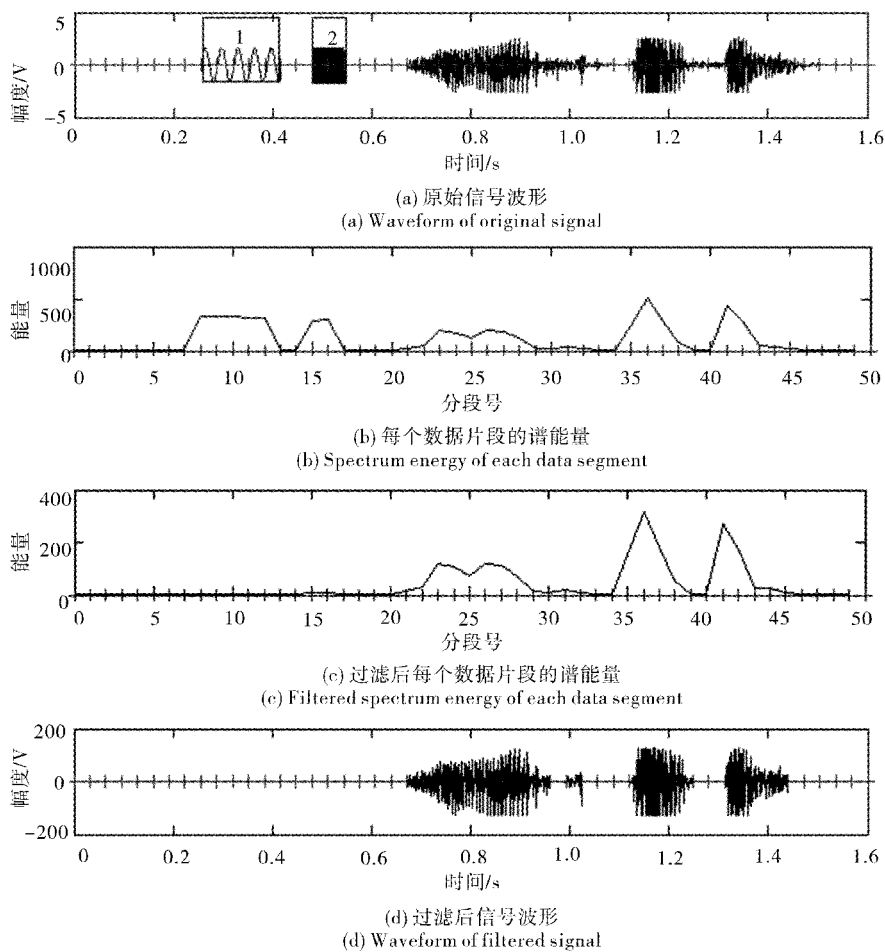


图 3 声音过滤仿真

Fig. 3 Simulation of audio filter

3 结 语

笔者采用网络语音通信技术, 利用 DirectSound 对声音进行采集和回放, 设计了网络语音通讯软件, 用以取代在以往仿真系统中使用硬件实现语音通讯的方法. 同时, 采用离散傅里叶变换对语音信号的频谱进行分析, 利用语音信号的频谱特征对通信过程中的无用信号帧进行过

滤, 只将满足判定条件的信号帧送往服务器进行转发, 从而有效减小了仿真系统中网络的数据传输负担. 该软件已在某训练模拟器中得到很好的应用, 满足训练员通话的需求.

参考文献:

- [1] 虞益诚. 网络技术和应用[M]. 南京: 东南大学出版社, 2005.
- [2] 李瑞. 一种网络语音聊天系统的设计与运用[J].

- 绿色大世界·绿色科技, 2009(7): 101-102.
- [3] 朱浩, 吕明. 用 UDP 协议实现网络语音广播[J]. 电信工程技术和标准化, 2005: 53-55.
- [4] 陈婷婷, 张彦铎. 机器人足球仿真比赛平台中网络通信问题研究[J]. 武汉工程大学学报, 2009, 31(3): 70-73.
- [5] 平殿发, 刘贤忠, 赵培洪. 短时分数阶傅里叶变换的基本性[J]. 数据采集与处理, 2009, 24(S): 39-42.
- [6] 高蕊, 史二娜. 基于 FFT 加窗与插值算法的接地电阻测量新方法[J]. 宝鸡文理学院学报, 2010, 30(3): 66-68.
- [7] 吴杰康, 龙军, 王辑祥. 基于数字微分算法的系统频率快速准确测量[J]. 电工技术学报, 2004, 19(4): 93-97.

Design and realization of real-time audio communication system based on distributed environment

SI Yu-xuan¹, SU Yuan-xing², YANG Zheng-fang³

(1. Department of Art, Yantai Vocational College, Yantai 264001, China; 2. The 92857st Unit of PLA, Beijing 100161, China; 3. The 91557st Unit of PLA, Zhoushan 316000, China)

Abstract: Audio communication using special hardware in distributed simulation not only increases the expense and complexity of system design, but also increases the difficulties of hardware composing and layout. Real-time audio communication was designed based on software in distributed environment. Audio data were collected and replayed by adopting DirectX and packed, transmitted and unpacked on the network. Aiming at redundant data produced in nobody speaking and noisy environment increasing the burden, audio signals were filtered in the frequency domain based on discrete Fourier transform (DFT), the burden of the network was reduced effectively. The system has good effects in a certain project.

Key words: distributed simulation; audio communication; network; discrete Fourier transform (DFT)

本文编辑: 苗 变



(上接第 22 页)

Synthesis of Lenalidomide intermediates in water

ZHANG Geng-zhen

(Zhejiang Better Pharmaceuticals Company Limited, Shaoxing 312071, China)

Abstract: N-bromosuccinimide (NBS) side-chain bromination reaction of alkyl arenes was traditionally conducted by lighting-initiated azobisisobutyronitrile (AIBN) generating free radicals in CCl₄ or other chlorinated solvents. In the water medium, by heating to 80 ~ 85 °C with the stirring rate of 20 ~ 30 r/min, the generated free radicals initiated α-bromination to synthesize Lenalidomide intermediates: 2-Bromomethyl-3-nitrobenzoic acid methyl ester; methyl 2-methyl-3-nitrobenzoate became molten state and two-phases were formatted in water; the yield of the aimed product is up to 80% with purity of 93%. Without lighting and toxic solvents, 2-Bromomethyl-3-nitrobenzoic acid methyl ester is easily industrialized in variable speed mixing and anti-acid equipment.

Key words: lenalidomide intermediates; bromination; green medium; yield

本文编辑: 苗 变