文章编号:1001 - 9081(2010)10 - 2825 - 03

基于概率决策的自适应跨平台多方会议方案

张历卓[1,2],贾维嘉[1,2],曹慧玲[1]
(1. 中南大学 信息科学与工程学院,长沙 410083; 2. 香港城市大学 电脑科学系,香港 九龙)
(zhanglizhuo@ gmail. com)

摘 要:针对多方会议的实际应用需求,同时兼顾 PDA 等小设备的个性特征,提出一种新颖而简单的快速实时自适应跨平台多方会议方案。该方案采用概率决策优先权的方式,即各客户端根据语音能量值和编码后帧长度计算其语音概率值,服务器由语音概率值决策出当前发言者的语音流,并使用叠加原理将选出的多路流进行混音,最后转发混音后的语音包。该方案弥补了 PDA 等小设备计算能力弱的缺陷,同时又降低了服务器进行混音操作的运算量。实验结果表明该方案具有算法复杂度低、听觉主观效果好、易在 PDA 以及手机等硬件设备上实现等特点,可广泛应用在多媒体多方会议跨平台系统的实现中。

关键词:3G;跨平台;多方会议;静音检测;混音
中图分类号: TP393 **文献标志码**:A

Probability based adaptive cross-platform multi-party conference scheme

ZHANG Li-zhuo[1,2], JIA Wei-jia[1,2], CAO Hui-ling[1]
(1. *School of Information Science and Engineering, Central South University, Changsha Hunan* 410083, *China*;
2. *Department of Computer Science, City University of Hong Kong, Kowloon Hongkong, China*)

Abstract: Based on the practical application and the characteristics of some small devices, such as PDA, a new and simple fast real-time adaptive cross-platform conference scheme was put forward. In the proposed scheme, priority was decided by probability, namely, the client calculated its audio probability according to energy value and the coded frame length, then the server decided the current speakers and mixed the streams by audio probability, finally transmitted mixed packages. The scheme skillfully offsets the weak computation of PDA and other small devices, at the same time reduces the calculation complexity of the server. The simulation results show that the proposed scheme has a low complexity and good hearing perceptibility. It can be easily implemented in hardware, such as PDA and mobile phone, and can be widely applied in cross-platform multimedia conference system.

Key words: 3G; cross-platform; multi-party conference; Voice Activity Detection (VAD); audio mixing

0 引言

在多方会议中,常常采用混音技术使多个参与者能顺畅地进行语音交流。传统的混音技术是将多个音频源的音频根据音频叠加原理[1-2]混合为一路音频输出,使音频的接收者感觉到多人会议交流的效果。

文献[3]提出了集中控制会议模式,在多点控制单元中对来自各发言者的语音信号进行混音处理,即客户端先对语音流编码,编码后的数据发送给控制单元解码,然后根据音频叠加原理进行混音。然而由于控制单元需要进行多路解码以及混音、编码,计算量和时间复杂度均较大,限制了该方案的应用。文献[4]提出了分散控制会议模式,客户端先对语音流编码,然后发送给服务器,服务器将各个客户端的语音数据发送其他所有客户端,客户端再对所有接收到的语音流进行合成。这种方式需要占用大量的网络带宽,影响语音信号的服务质量(Quality of Service, QoS),同时对客户端的计算能力有较高的要求,而且由于 3G 手机终端不可修改性,手机端混音基本不可行。

基于此,本文提出了语音自适应切换策略。该策略采用基于概率决策的自适应跨平台方案,并根据语音信息进行概率计算,然后将计算结果通过语音信息包发送到服务器,服务器根据收到的决策概率值进行切换决策。首先判断终端发送过来的语音概率值并选出大于服务器决策概率值的数据流,若选出流的数量超过一条,则使用叠加原理将选出的多路流进行叠加,最后将叠加后的音频发送到客户端,若没有或只有一条流大于服务器决策概率值,则直接将语音概率值最大的流发送给客户端。该策略计算复杂度低,容易实现,且可扩展性好,可满足大规模多媒体多方会议的跨平台应用需求。

1 应用场景

如图 1 所示的应用系统,即在“普适设备/终端/手机”上实现跨平台的多方会议技术,以便多媒体多方会议参与者在异构网络上可以跨平台地进行 QoS 通信。由于硬件设备的限制,如网关、CPU 处理能力以及其他硬件指标均低于一般服务器,因此已有的方案不能直接应用在这些设备上实现多媒体多方会议。而本文提出的基于概率决策的自适应跨平台多方会议方案正是为适应这种设备要求而设计的,其计算量不大,如图 1 所示,各参与者客户端,如 PC 机、PDA 持有者等,

收稿日期:2010 - 03 - 31;**修回日期**:2010 - 05 - 26。
基金项目:国家 973 计划项目(2003CB317003),香港城市大学基金项目(9668009)。
作者简介:张历卓(1974 -),男,湖南长沙人,讲师,博士研究生,主要研究方向:多媒体通信; 贾维嘉(1957 -),男,河北承德人,教授,博士生导师,博士,主要研究方向:组播、选播、路由、无线宽带网络、移动多媒体通信、分布式系统; 曹慧玲(1984 -),女,湖南长沙人,硕士,主要研究方向:计算机网络、路由协议设计。

将几路语音发往网关系统，手机用户则使用 3G 网络通过 3G324M 网关参与到会议中。

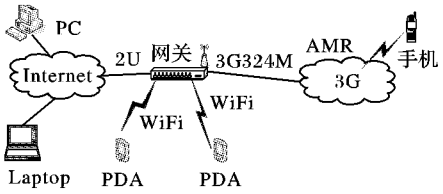

图 1 应用场景

2 自适应跨平台方案

客户端使用语音脉冲编码调制(Pulse Code Modulation, PCM)数据计算语音能量等相关参数，并利用自适应多速率(Adaptive Multi Rate, AMR)编码器编码的输出长度和历史参数信息来计算 PCM 数据中语音与噪声的比例值，根据这个比例值计算当前使用者的语音概率值。服务器利用各客户端发送的语音概率值和历史语音概率值计算决策概率值，最后决策出当前发言人和预备发言人。

2.1 客户端语音分析策略

由于语音信号的特征随时间而变化，只有在一短段时间间隔内，语音信号才保持相对稳定[5]。因此，语音信号的分析和处理一般建立在“短时性”基础上，即对语音信号流采用分段或分帧来处理。本文采用分帧来处理，采取连续方式，利用可移动的有限长度窗口对语音信号进行加权来实现。一般而言，只要信噪比不太低，语音信号的能量总是大于背景噪声的能量[6]，所以可以通过短时能量法即比较输入信号的能量与语音能量阈值的大小来判断输入信号为语音或是噪声。

假设客户端音频捕获模块每 20 ms 抓取到一帧语音信号(用 $x_n(m)$ 表示)，收集 160 个采样值后，第 n 帧语音信号 $x_n(m)$ 的短时能量值 E_n 用式(1)计算：

$$E_n = \sum_{m=1}^{160} x_n^2(m) \tag{1}$$

其中 E_n 的主要意义在于它是区分清音段与浊音段的基础。AMR 编码器(采用 MR122 模式[7])对语音 PCM 数据编码后，获得编码后的数据以及数据长度，假设数据长度用 $nSize$ 表示：当 $nSize = 1$ 时，为静音状态；当 $nSize = 6$ 时，为噪声状态；当 $nSize = 31$ 时，为语音状态[8]。通过 AMR 编码后的数据输出长度和语音能量值等历史参数，可以计算出 PCM 数据中语音与噪声的比例值。根据比例值的大小，可以计算出其语音概率值。但仅根据 AMR 编码后的数据输出长度来计算语音概率值并不准确。

由图 2 和图 3 可看出，当客户端没有发出语音时(可能有周围环境的影响)，PCM 数据能量值很小，但是 AMR 编码后的数据输出长度也有可能为 31。因此不能仅根据输出长度来判断是语音或噪声，要以语音能量等历史参数信息来辅助计算。当数据输出长度为 31 时，不能简单认为这帧数据为语音帧，要比较其能量与历史能量参数的大小。当其能量值大于历史能量参数时，才认为其为语音帧。假设客户端的历史噪声能量最大值表示为 E_{noise}，历史语音能量最大值表示为 E_{voice}，则历史能量参数 E_{refer} 可用式(2)计算：

$$E_{refer} = E_{noise} + (E_{voice} - E_{noise})/n \tag{2}$$

其中 n 为系数因子，可由实验效果调节其大小。此公式保证历史能量参数值大于历史噪声能量最大值。使用 AMR 编码后的数据输出长度和语音能量值可以计算出语音概率值，最后通过信令方式将该概率值发送至服务器端。

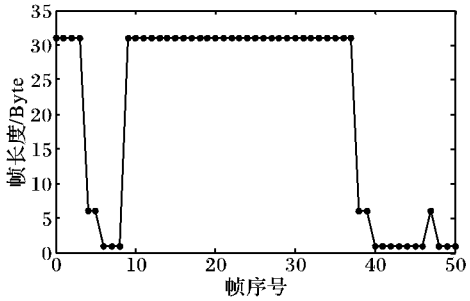

图 2 AMR 编码后帧长度变化图

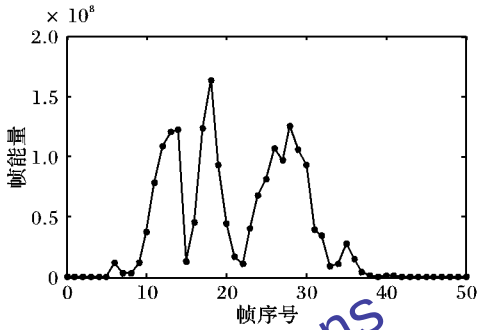

图 3 PCM 数据能量变化图

2.2 服务器端自适应切换策略

服务器端接收到客户端语音概率值后，每隔一个时间段统计其概率值，并更新其决策概率值。利用各客户端的决策概率值及其历史资料决策出需要进行混音处理的数据流。但是存在一个问题，若某客户端此刻为当前发言人，突然丢包或发生延迟，就会使决策概率值失真。为此，该方案加入一个时间变量，记录每次从客户端接收语音概率值的时间。假设当接收到三个语音概率值时统计一次，设 $t_i(i = 1,2,3)$ 时刻接收到的语音概率值为 $p_i(i = 1,2,3)$，如果 $\Delta t = t_i - t_{i-1}$ 和 $\Delta t = t_{i-1} - t_{i-2}$ 小于设定的阈值，认为没有出现丢包或延迟的现象，此时决策概率值 P 由式(3)计算：

$$P = p_i + p_{i-1} + p_{i-2} \tag{3}$$

否则决策概率值 P 由式(4)计算：

$$P = p_i + 2p_{i-1} \tag{4}$$

服务器每接收到一个语音概率值，决策概率值将重新赋值一次。目的是避免发生丢包或延迟的时间间隔较长时，决策概率值没有实时更新。

当获得每个客户端的决策概率值后，媒体交换模块根据决策结果进行相应的语音包转发，即将语音概率值大于决策概率值的客户端的语音进行叠加，并将叠加后的数据发送到每个会议参与者。

3 算法复杂度分析与性能分析

本文策略的时间和空间复杂度均较低，而且计算压力也不大。在客户端方面，只需要计算每个数据帧的能量值，以及判断每帧数据是语音、静音还是噪声，时间复杂度为 $O(n)$，空间复杂度为 $O(1)$；在服务器端方面，只需要计算每个客户端的决策概率值，时间复杂度为 $O(n)$，空间复杂度为 $O(1)$。客户端只需对使用者的语音流编码，以及对接收的语音流解码，不需要进行叠加计算；服务器端不需要编码，也不需要叠加计算。可看出该方案减少了对所有客户端数据进行混音所带来的巨大计算压力及资源消耗。

4　实验分析

本文采用视频会议系统原型进行数据收集,并通过实验数据来分析该方案是否可以实现自适应地选取语音数据并转发,以及分析语音概率值是否能反映出当前发言人的语音状态。本实验的测试环境是 PC 机、PDA 及 3G 手机,附加麦克风、话筒等硬件设备。下列数据均为模拟多方会议场景时收集的数据。

由图 4 可看出,客户端语音分析模块能够正确地识别语音帧、噪声帧和静音帧。当客户端使用者连接上服务器后,客户端语音分析模块对使用者的语音进行分析,计算出语音概率值,并把计算结果实时地发送给服务器端,服务器端根据接收到的语音概率值计算出决策概率值。图 4 是客户端说两句话时收集的数据,从图中数据可以看出,当客户端使用者发出语音时,其语音概率值能达到 100;而保持沉默时,如果周围没有噪声,其语音概率值为 0,如果周围有噪声,或者是前后语音切换的过程中,其语音概率值为一个较小的整数,这充分说明该语音分析模块能够正确地识别语音帧、噪声帧和静音帧。

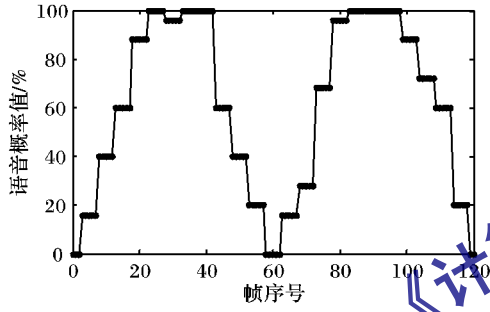

图 4　语音概率值变化

多人会议时,服务器端可以根据各客户端的语音概率值计算出决策概率值。由图 5 可看出,服务器端的混音模块能够根据决策概率值正确地判断出当前发言人。该实验假设服务器端接收到客户端三个语音概率值时,计算一次决策概率值。由图中数据可看出,当某个客户端的使用者发出语音时,服务器端能正确地计算出其决策概率值是 0 ~ 300 的整数值,而没有噪声的情况下,其他与会者的决策概率值均为 0。

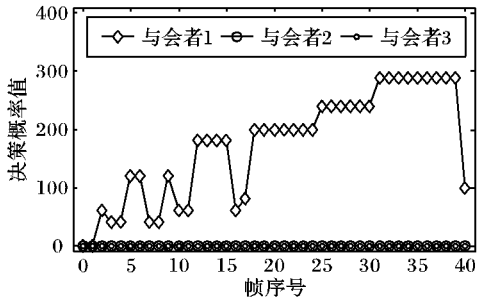

图 5　三人会议决策概率值变化

多人会议时,若与会者旁边有噪声,服务器端也可以根据客户端的决策概率值判断出谁是当前发言人,然后选取语音流。由图 6 可看出,当与会者 2 旁边有噪声时,其决策概率值不为 0,但是它是一个比较小的整数。而同时刻,当前发言人与会者 1 的决策概率值远远大于与会者 2 的决策概率值。因此根据决策概率值的大小很容易就可以决策出谁是当前发言人。

多人会议时,服务器端能根据决策概率值选取语音流,并依据决策结果转发数据。因此客户端根据接收到的语音,自适应地切换当前发言人身份。由图 7 可以看出,当与会者 1 发言时,其决策概率值最大,被判为当前发言人,其他与会者均能收到他的语音包。而预备发言人的数据包发送给当前发言人,当前发言人接到数据后,自适应切换发言人的身份。从第 10 帧到 13 帧的数据可以看出,这是一个切换当前发言人身份的过程。由与会者 1 的决策概率值最大,实时过渡到与会者 2 的决策概率值最大。当与会者 2 决策概率值最大时,被视为当前发言人。此时,服务器端把与会者 2 的语音包转发给其他所有的与会者。

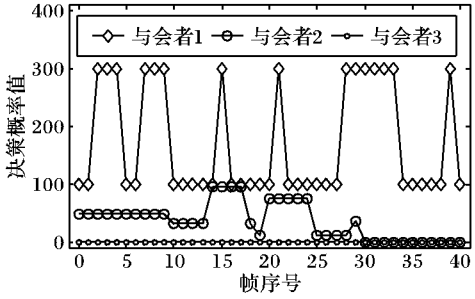

图 6　多人会议决策概率值变化

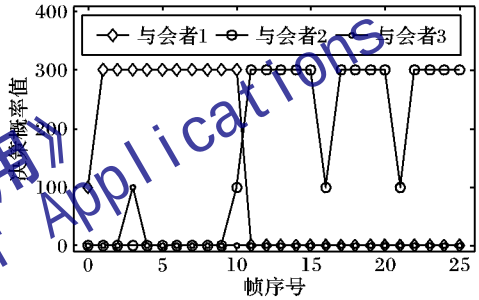

图 7　三人会议自适应语音切换

5　结语

为适应 3G 时代的应用需求,本文提出了一种新颖而简单的快速实时自适应跨平台多方会议方案。其主要创新点是采用概率分析的方法,客户端对语音流进行分析,服务器端利用接收到的语音概率值进行决策,充分利用服务器端与客户端的资源,让其共同分担计算压力;且具有算法简单、易实现、可扩展性好的特点,能满足大规模应用的需求;其应用范围广,可适应 PDA、手机等小设备的应用。

参考文献:

[1]　徐保民, 王秀玲. 一个改进的混音算法[J]. 电子与信息学报, 2003, 25(12): 1709 – 1713.

[2]　韩钰, 普杰信. 一种新的网络电话会议混音算法[J]. 计算机应用, 2010, 2(30): 564 – 566.

[3]　RANGAN P V, VIN H M, RAMANATHAN S. Communication architectures and algorithms for media mixing in multimedia conferences[J]. IEEE/ACM Transactions on Networking, 1993, 1(1): 20 – 30.

[4]　ITU-T. ITU-T Rec H.323 v4, Packet-based multimedia communication system [S]. ITU-T, 2000.

[5]　鲍长春. 数字语音编码原理[M]. 西安: 西安电子科技大学出版社, 2007.

[6]　张微, 毛敏. VOIP 电话会议系统中混音模块关键技术的研究与实现[D]. 上海: 华东师范大学, 2007.

[7]　王炳锡, 王洪. 变速率语音编码[M]. 西安: 西安电子科技大学出版社, 2004.

[8]　樊星, 顾伟康, 叶秀清. 多媒体会议中的快速实时自适应混音方案研究[J]. 软件学报, 2005, 16(1): 108 – 115.