



(12) 发明专利

(10) 授权公告号 CN 101675623 B

(45) 授权公告日 2013. 03. 27

(21) 申请号 200880014096. X

代理人 宋鹤 南霆

(22) 申请日 2008. 04. 28

(51) Int. Cl.

(30) 优先权数据

H04L 12/18(2006. 01)

11/799, 019 2007. 04. 30 US

(56) 对比文件

(85) PCT申请进入国家阶段日

CN 1849824 A, 2006. 10. 18,

2009. 10. 29

US 5436896 A, 1995. 07. 25,

(86) PCT申请的申请数据

EP 1720283 A2, 2006. 11. 08,

PCT/US2008/061711 2008. 04. 28

US 2003/0063573 A1, 2003. 04. 03,

(87) PCT申请的公布数据

审查员 陈晨

W02008/137373 EN 2008. 11. 13

(73) 专利权人 思科技术公司

地址 美国加利福尼亚州

(72) 发明人 威加亚拉戈哈凡·R·崔普利卡恩

莫西·V·艾特玛酷睿

(74) 专利代理机构 北京东方亿思知识产权代理

有限责任公司 11258

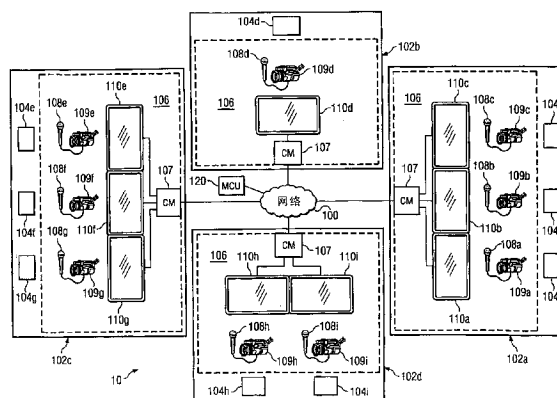
权利要求书 2 页 说明书 9 页 附图 2 页

(54) 发明名称

多点会议中的媒体检测和分组分发

(57) 摘要

本发明公开了一种用于在多点会议中分发媒体分组的方法,包括:接收多个音频信号。这多个音频信号中的每个音频信号包括音频分组,其中,来自这多个音频信号中的每个音频信号的一个或多个音频分组被利用测听值进行编码,测听值包括来自会议站点的声学度量。该方法还包括从这多个音频信号中的每个音频信号的一个或多个音频分组中提取一个或多个测听值并且基于所提取出的一个或多个测听值来选择一个或多个活动音频信号。此外,该方法还包括确定活动音频信号中的改变,并且响应于确定活动音频信号中的改变而更新媒体转发表,该媒体转发表包括用于对这多个音频信号中的一个或多个音频信号进行路由的目录。该方法还包括根据媒体转发表向一个或多个会议站点分发音频分组。



1. 一种用于在多点会议中分发媒体分组的方法,包括:

接收多个音频信号,所述多个音频信号中的每个音频信号包括音频分组,其中,来自所述多个音频信号中的每个音频信号的一个或多个音频分组被利用一测听值进行编码,所述测听值包括来自会议站点的声学度量;

从所述多个音频信号中的每个音频信号的一个或多个音频分组中提取出一个或多个测听值;

基于所提取出的所述一个或多个测听值来选择一个或多个活动音频信号;

确定所述活动音频信号中的改变;

响应于确定所述活动音频信号中的改变而更新媒体转发表,所述媒体转发表包括用于对所述多个音频信号中的一个或多个音频信号进行路由的目录;以及

根据所述媒体转发表向一个或多个会议站点分发音频分组。

2. 根据权利要求1所述的方法,还包括:

接收多个视频信号,所述多个视频信号中的每个视频信号与所述多个音频信号中的一个或多个音频信号相关联;

基于所述一个或多个活动音频信号来选择活动视频信号;以及

根据所述媒体转发表来分发所述视频信号中的一个或多个视频信号,其中,所述媒体转发表包括用于对所述多个视频信号中的一个或多个视频信号进行路由的目录。

3. 根据权利要求2所述的方法,还包括:

确定所述活动视频信号中的改变;以及

响应于确定所述活动视频信号中的改变而更新所述媒体转发表。

4. 根据权利要求2所述的方法,其中,基于一个或多个活动音频信号来选择活动视频信号包括基于档案文件来选择活动视频信号,所述档案文件维护一个或多个活动音频信号的记录。

5. 根据权利要求2所述的方法,还包括在会议站点处显示所述活动视频信号。

6. 根据权利要求1所述的方法,其中,基于所提取出的所述一个或多个测听值来选择一个或多个活动音频信号包括比较来自所述多个音频信号中的每个音频信号的多个测听值,其中,所述多个测听值是从在给定时间间隔中生成的音频分组中提取出的。

7. 根据权利要求1所述的方法,还包括:

在远程会议站点处广播所述一个或多个活动音频信号。

8. 根据权利要求1所述的方法,其中,所述测听值包括会议站点处的分贝度量。

9. 根据权利要求1所述的方法,其中,所述测听值基于会议站点处的多个测听度量被加权。

10. 根据权利要求1所述的方法,还包括:

根据所述所提取的测听值来排列所述多个音频信号。

11. 根据权利要求1所述的方法,还包括响应于确定所述活动音频信号中的改变,在根据所述媒体转发表向一个或多个会议站点分发音频分组之前接收即时解码器刷新帧。

12. 一种用于在多点会议中分发媒体分组的多点控制单元,包括:

用于接收多个音频信号的装置,所述多个音频信号中的每个音频信号包括音频分组,其中,来自所述多个音频信号中的每个音频信号的一个或多个音频分组被利用测听值进行

编码,所述测听值包括来自会议站点的声学度量;

用于从所述多个音频信号中的每个音频信号的一个或多个音频分组中提取一个或多个测听值的装置;

用于基于所提取出的所述一个或多个测听值来选择一个或多个活动音频信号的装置;

用于确定所述活动音频信号中的改变的装置;

用于响应于确定所述活动音频信号中的改变而更新媒体转发表的装置,所述媒体转发表包括用于对所述多个音频信号中的一个或多个音频信号进行路由的目录;以及

用于根据所述媒体转发表向一个或多个会议站点分发音频分组的装置。

13. 根据权利要求 12 所述的多点控制单元,还包括:

用于接收多个视频信号的装置,所述多个视频信号中的每个视频信号与所述多个音频信号中的一个或多个音频信号相关联;

用于基于所述一个或多个活动音频信号来选择活动视频信号的装置;以及

用于根据所述媒体转发表来分发所述视频信号中的一个或多个视频信号的装置,其中,所述媒体转发表包括用于对所述多个视频信号中的一个或多个视频信号进行路由的目录。

14. 根据权利要求 13 所述的多点控制单元,还包括:

用于确定所述活动视频信号中的改变的装置;以及

用于响应于确定所述活动视频信号中的改变而更新所述媒体转发表的装置。

15. 根据权利要求 13 所述的多点控制单元,还包括用于通过基于档案文件选择活动视频信号来基于一个或多个活动音频信号选择活动视频信号的装置,所述档案文件维护一个或多个活动音频信号的记录。

16. 根据权利要求 13 所述的多点控制单元,还包括用于在会议站点处显示所述活动视频信号的装置。

17. 根据权利要求 12 所述的多点控制单元,还包括用于通过比较来自所述多个音频信号中的每个音频信号的多个测听值来基于所述一个或多个所提取出的测听值来选择一个或多个音频信号的装置,其中,所述多个测听值是从在给定时间间隔中生成的音频分组中提取的。

18. 根据权利要求 12 所述的多点控制单元,还包括用于在远程会议站点处广播所述一个或多个活动音频信号的装置。

19. 根据权利要求 12 所述的多点控制单元,其中,所述测听值包括在会议站点处的分贝度量。

20. 根据权利要求 12 所述的多点控制单元,其中,所述测听值基于在会议站点处的多个测听度量被加权。

21. 根据权利要求 12 所述的多点控制单元,还包括用于根据所述所提取的测听值来排列所述多个音频信号的装置。

22. 根据权利要求 12 所述的多点控制单元,还包括用于响应于确定所述活动音频信号中的改变,在根据所述媒体转发表向一个或多个会议站点分发音频分组之前接收即时解码器刷新帧的装置。

多点会议中的媒体检测和分组分发

技术领域

[0001] 本发明一般地涉及通信系统,并且更具体地涉及多点会议中的媒体检测和分组分发。

背景技术

[0002] 存在很多方法允许位置遍布世界的多组个人参与到会议中。这样的方法一般涉及从位于一个会议站点的通信装备向位于一个或多个其它位置的通信装备发送信息和其它数据。多点控制单元(MCU)(有时称为多点会议单元)可以用来耦合在各个会议站点处的通信装备,从而允许来自分布式地理位置的用户参加电信会议(teleconference)。

[0003] 关于视频会议,MCU可以从多个会议站点接收和向多个会议站点分发多个音频和视频信号。在某些情况中,会议站点可能没有足够的装备来广播或显示由参加视频会议的远程会议站点生成的信号中的每个信号。因此,有必要在本地会议站点处所广播的音频和/或视频信号之间进行切换。

发明内容

[0004] 本发明提供了一种用于在多点会议中分发媒体分组的方法和多点控制单元,其基本上消除了或大大减少了与之前的方法和系统相关联的至少某些缺点和问题。

[0005] 根据一个具体实施例,一种用于在多点会议中分发媒体分组的方法,包括接收多个音频信号。这多个音频信号中的每个音频信号包括音频分组,其中,来自这多个音频信号中的每个音频信号的一个或多个音频分组被利用测听值(audiometric)进行编码,测听值包括来自会议站点的声学度量。该方法还包括,针对这多个音频信号中的每个音频信号,从一个或多个音频分组中提取一个或多个测听值,并且基于所提取的测听值来选择活动音频信号。此外,该方法包括确定活动音频信号中的改变,并且响应于确定活动音频信号中的改变而更新媒体转发表,媒体转发表包括用于对这多个音频信号中的一个或多个音频信号进行路由的目录。该方法还包括根据媒体转发表向一个或多个会议站点分发音频分组。

[0006] 在某些实施例中,该方法还可以包括接收多个视频信号,其中这多个视频信号中的每个视频信号与这多个音频信号中的一个或多个音频信号相关联。活动视频信号可以是基于这一个或多个活动音频信号选择的。该方法还可以包括根据媒体转发表来分发这些视频信号中的一个或多个视频信号。

[0007] 还提供了一种用于在多点会议中分发媒体分组的多点控制单元,其包括可操作用于接收多个音频信号的接口。这多个音频信号中的每个音频信号包括音频分组,其中,来自这多个音频信号中的每个音频信号的一个或多个音频分组被利用测听值进行编码,测听值包括来自会议站点的声学度量;会议控制单元还包括会议控制处理器,该会议控制处理器可操作用于从这多个音频信号中的每个音频信号的一个或多个音频分组中提取一个或多个测听值,并且基于所提取出的一个或多个测听值来选择一个或多个活动音频信号。会议控制处理器还可操作用于确定活动音频信号中的改变,并且响应于确定活动音频信号中的

改变而更新媒体转发表,媒体转发表包括用于对这多个音频信号中的一个或多个音频信号进行路由的目录。该会议控制处理器还包括根据媒体转发表向一个或多个会议站点分发音频分组。本发明的某些实施例可以提供一个或多个技术优势。本发明的一个实施例的一个技术优势是允许基于信号特性中的改变来路由信号的动态媒体转发表。本发明一个实施例的一个技术优势是基于所关联的音频信号来分发视频信号的能力。

[0008] 从以下图、描述和权利要求中,其它技术优势对于本领域技术人员将是显而易见的。并且,尽管以上已经列举了特定优势,但是各种实施例可以包括所列举的优势中的全部、某些优势或不包括这些优势。

附图说明

[0009] 为了更完整地理解本发明及其特征和优点,现在结合附图,参考以下描述,在附图中:

[0010] 图 1 是图示出根据某些实施例的用于执行多点会议的系统的框图;

[0011] 图 2 是图示出根据某些实施例的多点控制单元的框图;以及

[0012] 图 3 是图示出根据某些实施例的用于分发多点会议中的媒体分组的方法的流程图。

具体实施方式

[0013] 图 1 是用于执行多个远程位置之间的会议的通信系统 10。所示出的实施例包括通信网络 100,通信网络 100 可以支持:使用会议装备 106 在远程站点 102 之间进行会议。站点 102 可以包括可参加多个视频会议的任意合适数目的用户 104。还示出多点控制单元(MCU) 120,其在参加会议时辅助站点 102 之间的音频和/或视频信号的传送。这里使用的“会议”可以包括使用任何音频和/或视频手段发送的、多个用户之间的任何通信会话,包括通过语音和/或视频装置、文本聊天和即时消息传送发送的信号、数据或消息。

[0014] 通信网络 100 代表通信装备,包括用于将元件耦合到通信网络 100 的硬件和任何恰当的控制逻辑。一般而言,通信网络 100 可以是能够传送包括通过文本聊天和即时消息传送和电子邮件传送的信号、数据或消息的音频和/或视频电信信号、数据和/或消息的任何网络。因此,通信网络 100 可以包括以下网络中的全部或部分或它们的任何组合:无线电接入网络、公用交换电话网络(PSTN)、局域网(LAN)、城域网(MAN)、广域网(WAN)、诸如因特网之类的本地、区域或全球通信或计算机网络、有线或无线网络、企业内部网。为了辅助上述通信能力,通信网络 100 可以以任何合适的形式或布置包括路由器、集线器、交换机、网关、呼叫控制器和/或任何其它合适的组件。此外,站点 102 可以代表被配置为以分组、信元、帧、分段或其它数据部分的形式来传送信息的任何硬件和/或软件。尽管通信网络 100 被示出为单个网络,但是通信网络 100 可以包括任何数目或配置的网络。此外,通信系统 10 可以包括任何数目或配置的通信网络 100。

[0015] 用户 104 代表可能出席视频会议的一个或多个个人或一组或多组个人。用户 104 使用任何合适的装置和/或组件来参与视频会议,所述装置和/或组件例如是音频因特网协议(IP)电话、视频电话用具、基于视频电话的个人计算机(PC)和流传送客户端。在视频会议期间,用户 104 可以作为发言者参加会话或作为不发言者参与。

[0016] MCU 120 在多点会议期间充当中介。在操作中,MCU 充当互连来自各个会议站点的数据信号的桥梁。具体而言,MCU 120 可以收集由会议参加者通过它们的端点发送的音频和 / 或视频信号,并且向远程站点 102 处的多点会议的其它参加者分发这些信号。在操作中,MCU 可以向站点 102 处的具体的监视器 110 或扬声器分派具体的音频和 / 或视频信号。此外,MCU 120 可以被配置为支持在任何数目的会议上进行通信的任何数目的会议端点。MCU 120 可以包括为了支持多点会议 (包括视频会议) 所使用的任何桥接或交换装置。在各种实施例中,MCU 120 例如可以包括硬件、软件和 / 或嵌入式逻辑,例如一个或多个编解码器。此外,MCU 可以具有客户提供装备 (CPE,例如,超出网络接口) 的形式或可以被嵌入到诸如通信网络 100 之类的网络中。

[0017] 如图 1 中所示,站点 102 包括辅助用户 104 之间的会议的会议装备 106。会议装备 106 可以包括用来建立并且辅助视频会议的任何合适的元件。例如,会议装备 106 可以包括扬声器、用户接口、控制器或喇叭扩音器。在所示实施例中,会议装备 106 包括会议管理器 107、麦克风 108、摄像机 109 和监视器 110。尽管未被示出,但是会议装备 106 可以包括用于在远程位置之间进行视频会议的一个或多个网络接口、存储器、处理器、编解码器或任何其它合适的硬件或软件。根据具体的实施例,会议装备 106 可以包括任何合适的专用会议装置。在操作中,会议装备 106 可以使用任何合适的技术和 / 或协议 (例如,会话发起协议 (SIP) 或 H. 323) 来建立视频会议会话。此外,会议装备 106 可以支持诸如 H. 261、H. 263 和 / 或 H. 264 之类的其它标准的其它视频系统并且与它们互操作。

[0018] 会议管理器 (“CM”) 107 可以与通信网络 100 和会议站点 102 之间传送信息和信号。CM 107 可以包括用于管理会议的任何合适的硬件或软件。具体而言,CM 107 可以包括一个或多个处理器、存储器、接口或编解码器。在操作中,CM 107 可以向站点 102 发送或从其接收包含会议数据的信号。在具体实施例中,所发送的信号可以是除了携带音频数据之外还携带视频数据的音频 - 视频 (A/V) 信号。A/V 信号可以是模拟或数字信号并且可以是经过压缩的或未经过压缩的。在某些实施例中,A/V 信号是包括使用实时传输协议 (RTP) 发送的媒体 (音频和视频) 分组的信号。RTP 是用于通过因特网来传送音频和视频分组的标准化分组格式。尽管每个 CM 107 被描述为驻留在站点 102 处,但是 CM 107 可以位于系统 10 内的任何地方。

[0019] 麦克风 108 可以是可操作用于将声音转换成电信号的声电变换器或传感器。为了通信系统 10 的目的,麦克风 108 可以捕获本地站点 102 处的用户的语音并且将其变换成音频信号以发送给远程站点 102。尽管在所示实施例中,对于每个用户 104 有一个麦克风 108,但是具体的站点 102 可以具有比用户 104 更多或更少的麦克风。此外,在某些实施例中,麦克风 108 可以与会议装备 106 的任何其它组件 (例如摄像机 109) 相结合。

[0020] 如图 1 中所示,麦克风 108 和 / 或 CM 107 可以操作用于利用测听值来对音频信号进行编码。为了本说明书的目的,测听值是可以用来确定活动信号 (active signal) 的置信值或声学度量。活动信号是与当前正在发言的会议参加者 (即,活动发言者) 相对应的信号。可以基于具体语音的相对响度 (即,分贝水平) 来测量和 / 或计算测听值。还可以基于由具体站点 10 处的多个麦克风 108 收集的声学数据来定义测听值。例如,可以基于会议站点处的分贝轮廓 (profile) 来对具体音频信号的测听值进行加权。为了说明,如果用户 104a 当前正在发言,则麦克风 108a-108c 可以全都拾取与用户 104a 的语音相关联的声波。

由于用户 104a 最接近麦克风 108a, 所以分贝水平将在麦克风 108a 处为最高, 麦克风 108b 处较低, 并且在麦克风 108c 处为最低。如果各个麦克风 108a-108c 要向各自产生的音频信号单独地指派测听值, 则关于麦克风 108c 处的低分贝水平是由以柔和的语音发言的用户 104c 引起的还是由麦克风 108c 正在拾取来自其它发言者的残余声波而引起的存在不确定性。然而, 在给定会议站点的分贝轮廓 (即, 各个麦克风处测量出的分布水平) 的情况下, 用户 104a 是活动发言者或者可能是唯一的活动发言者的置信度增大。因此, 当针对由麦克风 108a-108c 生成的音频信号定义各自的测听值时, 针对由麦克风 108c 生成的信号进行编码的测听值可以被加权来说明该站点处的分贝轮廓。应当注意, 具体信号的测听值是动态的, 因此, 在用户 104a 停止发言或另一用户 104 开始发言之后, 针对各个信号而在各自的音频分组中被编码的测听值可以被相应地调整。

[0021] 摄像机 109 可以包括用来辅助捕获用户 104 和周围区域的图像的任何合适的硬件和 / 或软件。在某些实施例中, 摄像机 109 可以捕获并且发送用户 104 的图像作为视频信号。根据该实施例, 所发送的视频信号可以包括分离的信号 (例如, 各个摄像机 109 发送自己的信号) 或合成的信号 (例如, 来自多个源的信号被合成到一个视频信号中)。

[0022] 监视器 110 可以包括用来辅助接收视频信号并且向本地会议站点处的用户 104 显示远程用户 104 的图像的任何合适的硬件和 / 或软件。例如, 监视器 110 可以包括笔记本 PC、安装在墙上的监视器、安装在地板上的监视器或自由直立式监视器。监视器 110 可以使用提供真实图像的任何合适的技术, 例如高清晰、高功率压缩硬件和有效的编码 / 解码标准。

[0023] 在通信系统 10 的组件的操作的示例实施例中, 站点 102a 和 102d 处的用户参加会议。当用户 104 加入会议时, 针对每个摄像机 109 生成视频信号并且该视频信号被分派给监视器 110。在会议的持续时间可以维持该分派。因此, 远程用户可以总是被显示在相同的本地监视器上。这使得本地用户容易识别远程用户是谁并且位于何处。为了说明, 摄像机 109a 可以被分派给右监视器 110i, 中间的摄像机 109b 可以被指派给左监视器 110h 并且顶上的摄像机 109c 可以被指派给左监视器 110h。由于左监视器 110h 具有分派给它的中间的摄像机 109b 和顶上的摄像机 109c 两者, 所以, 该监视器可以基于哪个用户最后发言或哪个用户当前发言最大声而在摄像机 109b 和 108c 之间进行切换。因此, 在各个用户 104 在会议期间发言时, 各个监视器 110 上所显示的视频信号可以改变以显示最后的发言者的图像。

[0024] 可以对系统 10 进行修改、添加或省略。例如, 系统 10 可以包括任何合适数目的站点 102 并且可以辅助任何合适数目的站点 102 之间的视频会议。作为另一示例, 站点 102 可以包括用于辅助视频会议的任何数目的麦克风 108、摄像机 109 和显示器 110。作为另一示例, 站点 102 之间的视频会议可以是点对点会议或多点会议。对于点对点会议, 本地站点 102 处的显示器 110 的数目少于远程站点 102 处的摄像机 109 的数目。对于多点会议, 远程站点 102 处的摄像机 109 的合计数目大于本地站点 102 处的显示器 110 的数目。并且可以通过更多、更少或其它组件来执行系统 10 的操作。此外, 可以使用任何合适的逻辑来执行系统 10 的操作。

[0025] 图 2 图示出根据具体实施例的 MCU 220 的组件和操作。如图 2 中所表示的, MCU 220 包括接口 230、会议控制处理器 (CCP) 240 和存储器 260。MCU 220 可以与图 1 中所示的

MCU 120 类似。在图 2 中还图示出 A/V 输入信号 210 和 A/V 输出信号 212。

[0026] 接口 230 能够向诸如图 1 的通信网络 100 之类的通信网络传送信息和信号并且从其接收信息和信号。如图所示,接口 230 能够从参加会议的一个或多个站点接收一个或多个 A/V 输入信号 210 并且向参加会议的一个或多个其它站点发送一个或多个 A/V 输出信号 212。应当注意,A/V 输入信号 210 可以与 A/V 输出信号 212 基本相似。接口 230 代表任何真实的或虚拟的端口或连接,包括允许 MCU 230 与通信系统中的其它装置交换信息和信号的任何合适的硬件和 / 或软件。因此,接口 230 可以是或可以包括:以太网驱动器、通用串行总线 (USB) 驱动、网卡和 / 或防火墙。

[0027] 存储器 260 可以存储 CCP 指令和 / 或由 MCU 220 使用的任何其它信息。存储器 260 可以包括适于存储数据的易失性或非易失性、本地或远程装置的任何集合和布置。存储器 260 的示例可以包括 (但不限于) 随机存取存储器 (RAM) 装置、动态随机存取存储器 (DRAM)、只读存储器 (ROM) 装置、磁存储装置、光存储装置、闪存或任何合适的数据存储装置。

[0028] CCP 240 控制 MCU 220 的操作。具体地,CCP 240 处理从参加会议的站点处的摄像机或其它会议装备接收到的信息和信号。CCP 240 可以包括操作用来控制和处理信号的任何合适的硬件、软件或这两者。此外,CCP240 可以包括被布置在协议栈中的多个处理层,它们执行与媒体信号的处理相关联的各种任务。例如,如图所示,CCP 240 包括媒体层 242,切换层 244 和呼叫控制层 246。如将详细描述,这些层中的每个层可以操作用于执行一个或多个信号处理功能。尽管所示出的协议栈包括 3 层,但是 CCP 240 可以包括任何数目的处理层。此外,这些处理层中的每层可以包括用来执行所述功能的单独的处理器、存储器、硬件或软件。CCP 240 的示例包括 (但不限于) 专用集成电路 (ASIC)、现场可编程门阵列 (FPGA)、数字信号处理器 (DSP) 和任何其它合适的专用或通用处理器。

[0029] 媒体层 242 可以是接收一个或多个 A/V 信号并且提取用于较高层处理的相关信息的低级处理层。更具体而言,媒体层 242 可以从参加具体会议的一个或多个站点检测 A/V 信号并且从媒体信号中的音频分组中提取测听值。如前所述,测听值可以是用来确定活动发言者的置信值。在图 2 中所表示的 CCP 240 的实施例中,媒体层 242 与切换层 244 相接口。因此,媒体层 242 可以向切换层 244 转发所提取的测听值以用于进一步的处理。

[0030] 如图 2 中所示,媒体层 242 维护媒体转发表 243。媒体转发表 243 可以是用于将 A/V 信号路由到会议站点的目录、列表或其它索引。例如,关于图 1,媒体转发表 243 可以指示来自站点 102c 的 A/V 信号应当被导向站点 102a 和 102b。媒体转发表 243 还可以指示与具体用户 104 相关联的 A/V 信号应当被导向站点 102a 和 102b 处的具体监视器 110 和 / 或扬声器。在某些实施例中,媒体转发表可以针对音频信号及其相关联的视频信号维护单独的路由选择列表。这可以使得本地站点处的用户可以在发言者的图像未出现在一个或多个本地监视器上的情况下听到远程站点处的发言者的语音。此外,媒体转发表 243 可以是动态的。因此,其可以响应于 (一个或多个) 活动发言者的更改和 / 或根据任何合适的用户偏好而被修改或更新。因此,例如,当例如站点 102c 处的用户 104g 的会议参加者开始发言,则媒体转发表 243 可以被更新以使得与用户 104g 相关联的音频和视频信号在站点 102b 处的监视器 110d 上被广播。尽管图 2 图示出媒体转发表 243 作为媒体层 242 的组件,但是媒体转发表可以被存储或驻留在 MCU 220 内的任何位置或经由与通信系统 (例如,图 1 中

的通信系统 10) 中的其它组件的通信被 MCU 220 访问。

[0031] 如图 2 中所表示的, 切换层 244 是可操作用于分析由处理层 242 转发的测听数据的更高级处理层。具体地, 切换层 244 可以基于来自与具体会议相关联的各种音频信号的测听值来确定活动发言者。基于活动发言者, 切换层 244 可以确定在参加会议的各个站点处广播多个信号中的哪些信号。为了本说明书的目的, 被选择用于广播的 (一个或多个) 音频和视频信号可以分别称为活动音频和活动视频信号。在确定活动音频和 / 或活动视频信号之后, 切换层 244 可以更新媒体转发表 243。这可以通过向媒体层 242 传送包含相关信息的状态消息而被执行。这样的信息可以包括有关活动音频和活动视频信号的更新、更改或状态确认。响应于该状态消息, 媒体层 242 可以修改媒体转发表 243 使得与活动发言者相关联的音频和视频信号被恰当地路由。

[0032] 如图 2 中所示的呼叫控制层 246 是用于管理去往或来自会议站点的通信的处理层。具体地, 呼叫控制层可以对地址信息进行解码并且将通信从一个会议站点路由至另一会议站点。因此, 当站点拨入或以其他方式连接到会议时, 呼叫控制层 246 可以将该站点连接到一个或多个远程站点以用于会议。

[0033] 在一个实施例中, MCU 220 可以在接口 230 处从多个会议站点接收 A/V 输入信号。如所提到的, A/V 输入信号 210 可以是包括在本地站点生成以用于在远程站点处广播的音频和视频数据的媒体分组的流。音频数据可以包括测听值, 测听值被从音频分组提取出以用于提供可以用来确定活动发言者的置信度量。在接收到 A/V 输入信号 210 之后, 接口 230 可以将信号转发给 CCP 240 以用于处理。媒体层 242 之后可以检测 A/V 信号是否与具体会议相关联。在检测之后, 媒体层 242 可以从音频信号中的音频分组中提取 (一个或多个) 测听值。在提取测听值之后, 媒体层 242 可以将测听值转发给切换层 244。切换层 244 之后可以基于一个或多个测听值来确定 (一个或多个) 活动信号并且更新或修改媒体转发表 243 使得这 (一个或多个) 活动信号可以在远程会议站点处被广播。响应于该更新, 媒体层 242 可以根据媒体转发表转发与 A/V 输入信号 210 相关联的音频和 / 或视频分组, 使得它们被分发给这些会议站点。这些分组之后可以通过接口 230 被分发 (作为 A/V 输出信号 212)。

[0034] 根据 MCU 220 的具体实施例, 关于 CCP 240 描述的信号处理和转发功能可以通过接口 230 来实现。具体地, 接口 230 可以利用 Netfilter 软件来维护 Linux 内核。Netfilter 是在 Linux 内核中操作的开放源代码分组滤波架构。使用 Netfilter 钩子 (hook), 接口 230 可以在与具体会议相关联的 A/V 分组进入 CCP 240 的处理层之前检测并且截取它们。然后, Linux 内核可以提取音频分组中所编码的测听值, 并且与媒体层 242 类似地向切换层 244 呈现测听值。如前所述, 切换层 244 可以做出相应的交换判决。Linux 内核也可以维护与媒体转发表 243 类似的媒体转发表, 以用于将活动音频和活动视频信号路由到会议站点。在 RTP 被用来传送音频和视频数据的一个具体实施例中, Linux 内核可以分离 RTP 数据分组和 RTP 控制协议 (RTCP) 分组。RTCP 分组与 RTP 合作来发送和接收多媒体数据, 然而, 它们本身不传送任何数据。Linux 内核可以将 RTCP 分组转发给 CCP240 以用于由应用进行处理。由于 A/V 分组在到达 CCP 240 之前被截取, 所以在接口 230 处执行信号处理和转发可以减少通信等待时间和抖动。

[0035] 在某些实施例中, 可以以保存 MCU 220 的媒体处理的方式来实现交换判决。为了限制流量, 从不活动的会议站点向 MCU 220 发送的媒体数据可以被压缩或被限制于音频信

号。可替换地,媒体层 242 可以认识出,某些媒体分组与不活动的站点相关联并且拒绝处理该信息。因此,在具体实施例中,当信号被新指定为活动的时,媒体层 242 和 / 或接口 230 可以向用于该信号相关联的会议站点处的编解码器发送请求以发送即时解码器刷新 (IDR) 帧。IDR 帧可以包含对于 MCU 处的编解码器启动处理并且显示来自站点的音频和 / 或视频信号必要的信息。在接收到该帧之后,MCU 220 可以启动对该信号的处理,从而根据媒体转发表 243 发送该信号。因此,在信号被指定为活动的时间段到 IDR 帧被接收到的时间期间,可以由 MCU 220 发送旧的 (即,之前活动的信号)。尽管这可能增加交换时间,但是 MCU 资源可以被保存,因为可能更少的媒体处理是必要的。

[0036] 对活动信号的选择 (即,确定活动发言者) 可以以相似的方式被执行,而不论信号处理是由接口 230 还是由 CCP 240 执行。可以基于与各种音频信号的分组相关联的测听值来确定活动发言者。如关于图 1 所讨论的,每个麦克风 107 可以生成由与测听值一起编码的音频数据的分组组成的音频信号。切换层 244 和 / 或接口 230 可以通过比较与每个信号的分组相关联的测听值来确定活动发言者。在实施例中,从中提取出测听值的各个信号中的每一个信号中的分组具有大约相同的时间戳。这可以确保用来选择活动信号的测听值来自在大约相同的时间生成的分组。从比较与每个信号相关联的测听值,可以选择与活动发言者相对应的活动音频信号。例如,包括具有 (一个或多个) 最高测听值的 (一个或多个) 分组的音频信号可以被选作与活动发言者相对应的信号。在一个具体实施例中,切换层 244 和 / 或接口 230 可以按照信号的测听值来排列信号。因此,如果会议站点具有多个监视器和 / 或扬声器,则与最有可能活动的发言者相关联的信号可以被广播。

[0037] 显然,活动发言者可以在会议期间更改任何次数。因此,切换层 244 和 / 或接口 230 可以不断地监视与会议相关联的信号的测听值。因为视频信号可能由大量分组组成 (每个分组利用测听值被编码),所以,可以逐分组地执行对活动发言者的确定。然而,根据来自多个音频信号的具体一组分组中的数据来交换 / 更新活动视频和活动音频信号可能无法提供最佳的用户体验。这是因为诸如打喷嚏、咳嗽或电话铃之类的某些事件可以产生利用可以指示活动发言者的测听值编码的分组。因此,逐分组活动发言者确定的灵敏度可能引起会议参加者被错误地指定为活动发言者,这可能导致与该参加者相关联的音频和视频信号被指定为活动音频和活动视频信号。由于该活动音频和 / 或活动视频指定可能只是瞬时的,所以诸如打喷嚏之类的事件可能造成远程会议站点处的监视器或扬声器的闪烁。

[0038] 为了解决具有闪烁的潜在问题,根据具体实施例,可以基于通过 200 毫秒间隔或另一指定或编程的时间间隔生成的音频分组来确定活动发言者。200 毫秒用作阻尼 (damping) 时间段,其用来保证具体信号不会因为与该信号相关联的 (一个或多个) 测听值的突然上升而被指定为活动。因此,如果来自与会议参加者相关联的音频信号的测听值指示该参加者已经保持活动达 200 毫秒,则该参加者可以被指定为活动发言者。类似地,切换层 244 和 / 或接口 230 可以采用用于确定在给定的阻尼时间段上的活动发言者的任何合适的算法。作为一个示例,在 200 毫秒的间隔上具有最高平均测听值的 (一个或多个) 信号可以被选择作为 (一个或多个) 活动信号。这可以说明产生利用测听值编码出的音频分组的语音中的停顿或其它暂停。尽管已经使用 200 毫秒阻尼时间段描述了切换层 244 和 / 或接口 230 的前述 (一个或多个) 操作,但是可以实现任何长度的阻尼时间段。

[0039] 在一个实施例中,音频和视频信号可以被分别指定为活动。具体而言,可以针对音

频和视频信号采用不同的阻尼间隔。例如,如所讨论的,由与会议参加者相关联的麦克风生成的音频信号可以被指定为活动的而由与参加者相关联的摄像机生成的相应视频信号可以被指定为不活动的。继续 200 毫秒阻尼间隔,音频信号可以每 200 毫秒被指定为活动的。相比之下,相关联的视频信号可以每 2 秒被指定为活动的。因此,站点处的参加者可以在参加者的图像出现在本地监视器上之前听到远程参加者的语音。针对音频和视频信号采用不同的阻尼间隔可以通过在允许会议参加者听到远程站点处的发言者的同时限制监视器上的闪烁来增强用户体验。此外,相对于视频信号,针对音频信号维持较短的阻尼时间段可以防止本地参加者错过来自远程会议参与者的通信。

[0040] 由于视频信号可以与相关联的音频信号相分离地被指定为活动的,所以,相对于活动音频信号,切换层 244 和 / 或接口 230 可以采用不同的协议来确定活动视频信号。例如,切换层 244 可以与存储器 260 相结合,维护活动音频信号的档案文件。从而,该档案文件可以被用来选择活动视频信号。为了说明,该档案文件可以记录每次活动音频信号中发生改变的情况。可替换地,该档案文件可以在每个音频阻尼时间段之后记录活动音频信号。因此,如果音频阻尼时间段是 200 毫秒并且视频信号的阻尼时间段是 2 秒时,则活动视频可以是基于前十个档案文件条目的。在一个实施例中,如果档案文件指示具体音频信号对于整个视频阻尼时间段都是活动的,则与该音频信号相关联的视频信号可以被选作活动视频信号。在另一实施例中,被指定为活动的视频信号可以是与对于大半视频阻尼时间段是活动的音频信号相关联的视频信号。应当注意,对于音频信号,多于单个视频信号可以被指定为活动的。此外,尽管已经描述了用于选择活动视频信号的特定方法,但是可以实现基于(一个或多个)活动音频信号来选择活动视频信号的各种其它方法。

[0041] 现在参考图 3,提供了图示出根据一个具体实施例的 MCU 220 的示例操作的流程图。该方法从步骤 300 开始,在步骤 300,接收多个音频和视频信号。这些信号可以通过诸如接口 230 之类的接口来接收。这些音频信号可以由传送由远程站点处的一个或多个麦克风生成以用于在本地会议站点处广播的音频数据的多个分组组成。这些音频信号中的每一个中的一个或多个分组可以是利用提供用来确定活动发言者的置信值的测听值来编码的。

[0042] 接着,在步骤 302,这些分组被转发给 CCP 240 以用于处理。然后,在步骤 304,从这些信号中的每一个信号中的分组中提取测听值。可由媒体层 242 来执行该提取步骤。

[0043] 在步骤 306,可以选择活动音频信号。该选择可以由切换层 244 进行并且可以是基于所提取出的测听值的比较。该比较可以包括从具有具体时间戳的分组中提取的测听值。因此,活动音频信号的确定可以是基于对于具体时间戳具有最高测听值或在一系列时间戳上具有最高测听值的信号。此外,切换层 244 可以基于信号的测听值来选择是活动的多个信号或排列信号。在步骤 308 处,切换层还可以基于对活动音频信号的选择来选择活动视频信号。

[0044] 在(一个或多个)活动信号被选择之后,在步骤 310,切换层 244 可以判断是否更新媒体转发表 243。该判断可以是基于需要在参加会议的站点之间对音频和 / 或视频信号进行重新路由的(一个或多个)活动音频信号中的改变的。如果媒体转发表要被更新,则切换层 244 可以向媒体层 242 传送更新,从而该更新可以在步骤 312 处修改媒体转发表。

[0045] 不论媒体转发表 243 是否被更新,在步骤 314,媒体层 242 都向参加会议的站点中的一个或多个站点分发与活动音频和 / 或活动视频信号相关联的分组。这些分组可以是基

于媒体转发表 243 中的路由参数被分发的。在某些实施例中,媒体层 242 可以基于该媒体转发表来将分组分发 / 路由至会议站点处的具体监视器和 / 或扬声器。

[0046] 可以对图 3 中所描述的方法进行各种修改、添加或省略。在某些实施例中。该方法可以包括更多、更少或其它步骤。例如,媒体层 242 可以在分组被接收到之后,根据媒体转发表 243 分发这些分组。当对媒体转发表的更新发生时,媒体层 240 可以通过根据更新后的表格来路由分组而继续。因此,可以以任何合适的顺序来执行步骤而不偏离本发明的范围。

[0047] 尽管已经参考了具体实施例详细描述了本发明,但是,本领域技术人员可以确知各种改变、替换、变化、更改和修改,并且希望本发明包括落在所附权利要求的精神和范围之内的所有这样的改变、替换、变化、更改和修改。

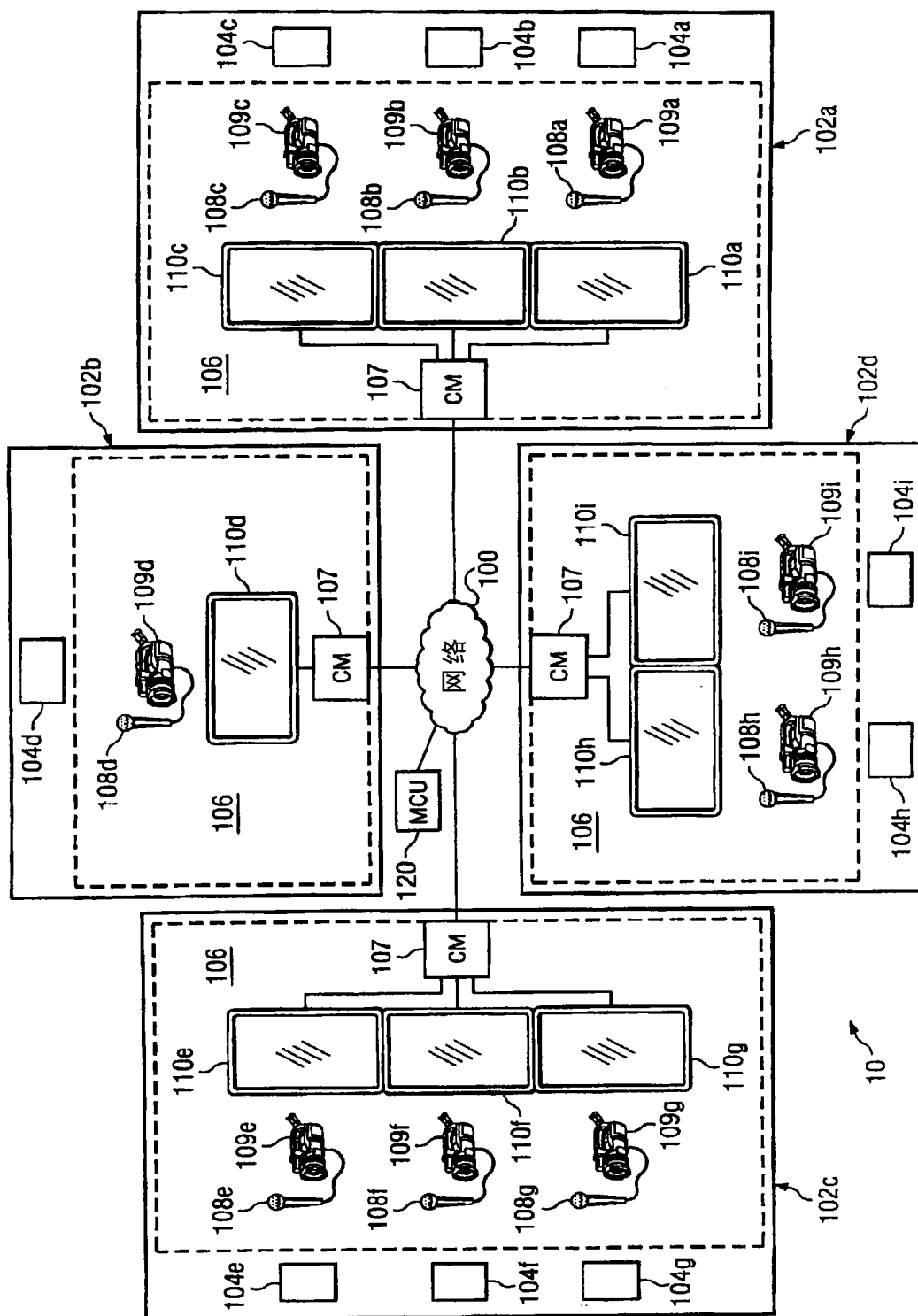


图 1

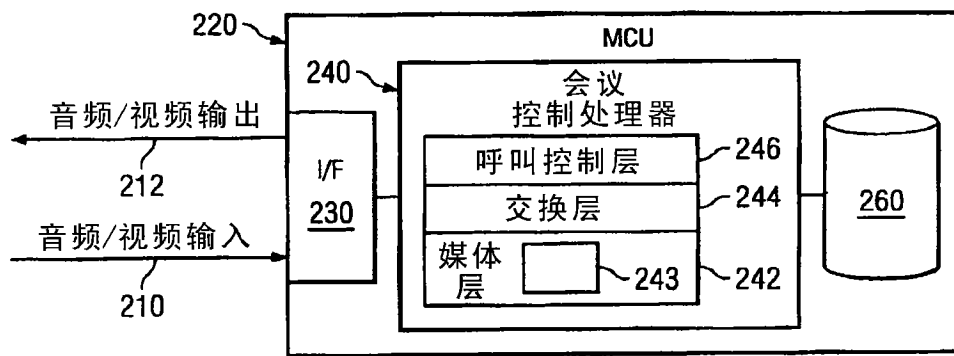


图 2

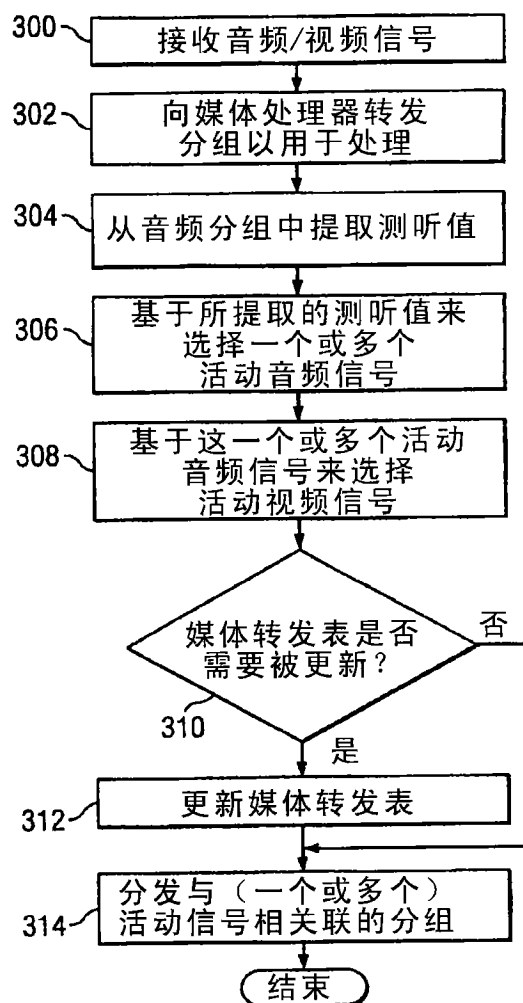


图 3