(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 107430858 B
(45)授权公告日 2020.11.03

(21)申请号 201680017187.3

(22)申请日 2016.03.18

(65)同一申请的已公布的文献号
申请公布号 CN 107430858 A

(43)申请公布日 2017.12.01

(30)优先权数据
14/664,047 2015.03.20 US

(85)PCT国际申请进入国家阶段日
2017.09.20

(86)PCT国际申请的申请数据
PCT/US2016/023000 2016.03.18

(87)PCT国际申请的公布数据
WO2016/153943 EN 2016.09.29

(73)专利权人 微软技术许可有限责任公司
地址 美国华盛顿州

(72)发明人 G·卡施坦 B·施莱辛格
H·菲特斯

(74)专利代理机构 北京市金杜律师事务所 11256
代理人 王茂华 黄捷

(51)Int.Cl.
G10L 17/22(2013.01) (续)

(56)对比文件
CN 101314081 A,2008.12.03
CN 104112449 A,2014.10.22
US 8902274 B2,2014.12.02
US 7305078 B2,2007.12.04
US 2013294594 A1,2013.11.07
US 2011060591 A1,2011.03.10
US 2012163576 A1,2012.06.28
CN 104424955 A,2015.03.18
CN 102404545 A,2012.04.04
CN 102713935 A,2012.10.03
CN 103314389 A,2013.09.18
CN 101669324 A,2010.03.10
CN 1805489 A,2006.07.19
US 6853716 B1,2005.02.08
US 2014211929 A1,2014.07.31
JP 2011191542 A,2011.09.29
WO 2014154262 A1,2014.10.02
US 2015025888 A1,2015.01.22 (续)

审查员 叶双清

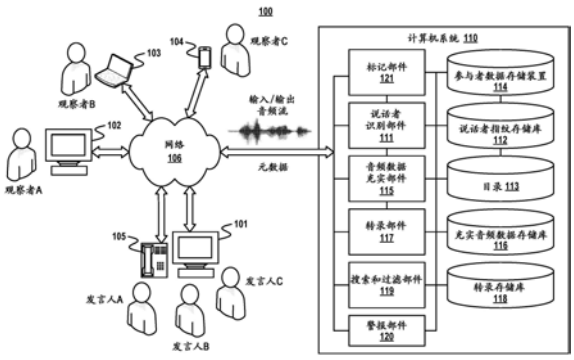
权利要求书4页 说明书31页 附图11页

(54)发明名称

传送标识当前说话者的元数据

(57)摘要

一种计算机系统可以传送标识当前说话者的元数据。该计算机系统可以接收表示当前说话者的语音的音频数据，基于该音频数据来生成当前说话者的音频指纹，以及通过将当前说话者的音频指纹与包含在说话者指纹存储库中的已存储音频指纹进行比较来执行自动说话者识别。计算机系统可以向观察者的客户端设备传送指示当前说话者未被识别的数据，并且从观察者的客户端设备接收标识当前说话者的标记信息。计算机系统可以将当前说话者的音频指纹和标识当前说话者的元数据存储在说话者指纹存储库中，并且向观察者的客户端设备或者不同观察者的客户端设备中的至少一个客户端设备传送标识当前说话者的元数据。



CN 107430858 B

[转续页]

[接上页]

(51)Int.Cl.

G10L 17/04(2013.01)

H04M 3/56(2006.01)

(56)对比文件

US 8731935 B2,2014.05.20

US 2005135583 A1,2005.06.23

US 8416937 B2,2013.04.09

ROSA GONZALEZ HAUTAMAKI."Merging human and automatic system decisions to improve speaker recognition performance". 《PROC. INTERSPEECH 2013》.2013,

1.一种用于传送标识当前说话者的元数据的计算机系统,所述计算机系统包括:

处理器,被配置成执行计算机可执行指令;以及

存储器,存储一个或多个计算机可执行指令,所述计算机可执行指令当被所述处理器执行时,执行以下操作,包括:

接收表示所述当前说话者的语音的音频数据;

基于所述音频数据,来生成所述当前说话者的音频指纹;

执行自动说话者识别,包括:将所述当前说话者的所述音频指纹与被包含在说话者指纹存储库中的一个或多个已存储音频指纹进行比较;

向观察者的第一客户端设备传送指示所述当前说话者未被识别的数据;

从所述观察者的所述第一客户端设备接收标识所述当前说话者的标记信息;

将所述当前说话者的所述音频指纹和标识所述当前说话者的元数据存储在所述说话者指纹存储库中,所述元数据至少部分基于所述标记信息;

向所述观察者的所述第一客户端设备或不同观察者的第二客户端设备中的至少一个客户端设备,传送标识所述当前说话者的所述元数据;

从所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备接收标识特定说话者的请求;以及

在所述特定说话者当前正在说话时,向所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备传送警报。

2.根据权利要求1所述的计算机系统,其中所述存储器还存储一个或多个计算机可执行指令,所述计算机可执行指令当被所述处理器执行时,执行以下操作,包括:

从至少一个另外的观察者的至少一个另外的客户端设备处接收附加的标记信息,所述附加的标记信息标识所述当前说话者;以及

通过基于由大多数观察者所提供的身份而标识所述当前说话者,来解决标识所述当前说话者的所述标记信息以及标识所述当前说话者的附加的标记信息之间的冲突。

3.根据权利要求1所述的计算机系统,其中所述存储器还存储一个或多个计算机可执行指令,所述计算机可执行指令当被所述处理器执行时,执行以下操作,包括:

接收所述当前说话者已经被正确地标识的确认。

4.根据权利要求1所述的计算机系统,其中所述存储器还存储一个或多个计算机可执行指令,所述计算机可执行指令当被所述处理器执行时,执行以下操作,包括:

从信息源取回用于所述当前说话者的附加信息;以及

传送在标识所述当前说话者的所述元数据中的所述附加信息。

5.根据权利要求4所述的计算机系统,其中所述附加信息包括以下一项或多项:所述当前说话者的公司、所述当前说话者的部门、所述当前说话者的职位、或者用于所述当前说话者的联系信息。

6.根据权利要求1所述的计算机系统,其中所述存储器还存储一个或多个计算机可执行指令,所述计算机可执行指令当被所述处理器执行时,执行以下操作,包括:

生成增强音频数据,所述增强音频数据包括表示所述当前说话者的语音的所述音频数据和标识所述当前说话者的所述元数据。

7.根据权利要求6所述的计算机系统,其中经由所述增强音频数据,标识所述当前说话

者的所述元数据被传送到所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备。

8.根据权利要求6所述的计算机系统，其中所述存储器还存储一个或多个计算机可执行指令，所述计算机可执行指令当被所述处理器执行时，执行以下操作，包括：

存储所述增强音频数据；

接收指示所识别的说话者的查询；

对所述增强音频数据进行搜索标识所述所识别的说话者的元数据；以及

输出所述增强音频数据的部分，所述增强音频数据的所述部分表示所述所识别的说话者的语音。

9.根据权利要求1所述的计算机系统，其中所述存储器还存储一个或多个计算机可执行指令，所述计算机可执行指令当被所述处理器执行时，执行以下操作，包括：

生成具有多个说话者的会话的转录，其中由所识别的说话者所说出的语音的文本与所述所识别的说话者的标识符相关联；

存储所述转录；

接收指示所述所识别的说话者的查询；

从所述转录中搜索所述所识别的说话者的所述标识符；以及

输出所述转录的部分，所述转录的所述部分包括由所述所识别的说话者说出的语音文本。

10.根据权利要求1所述的计算机系统，其中所述存储器还存储一个或多个计算机可执行指令，所述计算机可执行指令当被所述处理器执行时，执行以下操作，包括：

接收表示所述当前说话者的语音的后续音频数据；

基于所述后续音频数据来生成所述当前说话者的新音频指纹；

通过将所述当前说话者的所述新音频指纹与所述当前说话者的所述已存储音频指纹比较，来执行说话者识别；以及

向所述观察者的客户端设备或所述不同观察者的客户端设备，传送标识所述当前说话者的所述元数据。

11.根据权利要求1所述的计算机系统，其中所述存储器还存储一个或多个计算机可执行指令，所述计算机可执行指令当被所述处理器执行时，执行以下操作，包括：

对被包括在被实时传送的增强音频数据中的元数据执行操作，以确定所述特定说话者当前正在说话；以及

其中在所述特定说话者当前正在说话时，向所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备传送警报包括：向所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备发送数据，所述数据用于每当所述特定说话者发言时，生成可听的或可见的警报中的至少一个。

12.根据权利要求1所述的计算机系统，其中所述存储器还存储一个或多个计算机可执行指令，所述计算机可执行指令当被所述处理器执行时，执行以下操作，包括：

为参与者提供在线会议；

从所述参与者的至少一个客户端设备接收至少一个参与者的音频指纹；以及

将至少一个参与者的所述音频指纹和标识至少一个参与者的元数据存储在所述说话者指纹存储库中。

13.根据权利要求1所述的计算机系统，其中所述存储器还存储一个或多个计算机可执行指令，所述计算机可执行指令当被所述处理器执行时，执行以下操作，包括：

向所述观察者的所述第一客户端设备传送所述当前说话者的所述音频指纹。

14.一种由包括一个或多个计算设备的计算机系统执行的计算机实现的方法，所述计算机实现的方法用于传送标识当前说话者的元数据，所述方法包括：

基于表示所述当前说话者的语音的音频数据来生成所述当前说话者的音频指纹；

基于所述当前说话者的所述音频指纹和已存储的一个或多个音频指纹来执行自动说话者识别；

当所述当前说话者未被识别时，从观察者的第一客户端设备接收标识所述当前说话者的标记信息；

存储所述当前说话者的所述音频指纹和标识所述当前说话者的元数据，所述元数据至少部分基于所述标记信息；

向所述观察者的所述第一客户端设备或不同观察者的第二客户端设备中的至少一个客户端设备，传送标识所述当前说话者的所述元数据；

从所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备接收标识特定说话者的请求；以及

在所述特定说话者当前正在说话时，向所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备传送警报。

15.根据权利要求14所述的计算机实现的方法，还包括：

向所述观察者的所述客户端设备传送指示所述当前说话者未被识别的数据。

16.根据权利要求14所述的计算机实现的方法，还包括：

从至少一个另外的观察者的至少一个另外的客户端设备来接收附加的标记信息，所述附加的标记信息标识所述当前说话者；以及

通过基于由大多数观察者所提供的身份而标识所述当前说话者，来解决标识所述当前说话者的所述标记信息以及标识所述当前说话者的附加的标记信息之间的冲突。

17.根据权利要求14所述的计算机实现的方法，还包括：

基于表示所述当前说话者的语音的后续音频数据，来生成所述当前说话者的新音频指纹；以及

基于所述当前说话者的所述新音频指纹和所述当前说话者的所述已存储音频指纹，来执行说话者识别。

18.一种存储有计算机可执行指令的计算机可读存储介质，所述计算机可执行指令当被计算设备执行时，使得所述计算设备实现：

说话者识别组件，其被配置为基于表示当前说话者的音频数据，来生成所述当前说话者的音频指纹，并且通过将所述当前说话者的所述音频指纹与已存储的音频指纹进行比较，来执行自动说话者识别；

标记组件，其被配置为当所述自动说话者识别不成功时，从观察者的第一客户端设备接收标识所述当前说话者的标记信息，并且将所述当前说话者的所述音频指纹与所述已存

储的音频指纹一起存储；

音频数据充实组件，其被配置为向所述观察者的所述第一客户端设备或不同观察者的第二客户端设备传送标识所述当前说话者的元数据，所述元数据至少部分基于所述标记信息；以及

警报组件，其被配置为从所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备接收标识特定说话者的请求，并且在所述特定说话者当前正在说话时，向所述观察者的所述第一客户端设备或所述不同观察者的所述第二客户端设备中的至少一个客户端设备传送警报。

19.根据权利要求18所述的计算机可读存储介质，其中所述标记组件还被配置为从至少一个另外的观察者的至少一个另外的客户端设备来接收附加的标记信息，所述附加的标记信息标识所述当前说话者，并且通过基于由大多数观察者所提供的身份而标识所述当前说话者，来解决标识所述当前说话者的所述标记信息以及标识所述当前说话者的附加的标记信息之间的冲突。

20.根据权利要求18所述的计算机可读存储介质，其中所述音频数据充实组件被配置为传送表示所述当前说话者的语音的所述音频数据以及标识所述当前说话者的所述元数据，作为音频数据和元数据的同步流。

传送标识当前说话者的元数据

背景技术

[0001] 基于网络的会议服务可以包括诸如互联网语音协议(VoIP)音频会议、视频会议、即时消息传送和桌面共享之类的特征，以允许在线会议的参与者实时地进行通信并且同时查看在通信会话期间呈现的文档和/或在通信会话期间呈现的文档上工作。当参加在线会议时，会议发起者或被邀请方可以使用个人计算机、移动设备和/或座机电话连接至基于web的会议服务，并且可以被提示提供帐户信息或身份，以及在一些情况下，会议标识符。在线会议的参与者可以在不同时间充当发言人或参加者，并且可以通过说话、收听、聊天、呈现共享文档和/或查看共享文档进行交流和协作。

发明内容

[0002] 提供以下发明内容以简化形式介绍下文在具体实施方式中进一步描述的一些概念的选择。本发明内容并非旨在标识所要求保护的主题的关键特征或基本特征，也不旨在用于限制所要求保护的主题的范围。

[0003] 在各种实现方式中，一种计算机系统可以传送标识当前说话者的元数据。该计算机系统可以接收表示当前说话者的语音的音频数据，基于该音频数据来生成当前说话者的音频指纹，以及通过将当前说话者的音频指纹与包含在说话者指纹存储库中的、已存储音频指纹进行比较来执行自动的说话者识别。计算机系统可以向观察者的客户端设备传送指示当前说话者未被识别的数据，并且从观察者的客户端设备接收标识当前说话者的标记信息。计算机系统可以将当前说话者的音频指纹和标识当前说话者的元数据存储在说话者指纹存储库中，并且向观察者的客户端设备中的至少一个客户端设备或者不同观察者的客户端设备传送标识当前说话者的元数据。

[0004] 通过阅读以下具体实施方式和对附图的观察，这些和其他特征和优点将变得明显。应当理解，前面的发明内容、以下的具体实施方式和附图仅是说明性的，而非对所要求保护的各个方面的限制。

附图说明

[0005] 图1图示了可以实现所描述的主题的各方面的示例性操作环境的实施例。

[0006] 图2A至图2D图示了根据所描述的主题的各方面的示例性用户界面的实施例。

[0007] 图3图示了根据所描述的主题的各方面的示例性操作环境的实施例。

[0008] 图4图示了根据所描述的主题的各方面的示例性过程的实施例。

[0009] 图5图示了可以实现所描述的主题的各方面的示例性操作环境的实施例。

[0010] 图6图示了可以实现所描述的主题的各方面的示例性计算机系统的实施例。

[0011] 图7图示了可以实现所描述的主题的各方面的示例性移动计算设备的实施例。

[0012] 图8图示了可以实现所描述的主题的各方面的示例性计算环境的实施例。

具体实施方式

[0013] 下文结合附图提供的详细描述旨在作为示例的描述，而不旨在表示当前示例可以被构造或使用的唯一形式。该描述阐述了示例的功能以及用于构建和操作示例的步骤的顺序。然而，相同或等同的功能和顺序可以通过不同的示例来实现。

[0014] 对于“一个实施例”、“实施例”、“示例实施例”、“一个实现方式”、“实现方式”、“一个示例”、“示例”等的引用指示所描述的实施例、实现方式或示例可以包括特定特征、结构或特性，但是每个实施例、实现方式或示例可以不一定包括特定特征、结构或特性。而且，这样的短语不一定指代相同的实施例、实现方式或示例。进一步地，当结合实施例、实现方式或示例描述特定特征、结构或特性时，应当领会，可以结合其他实施例、实现方式或示例来实现这样的特征、结构或特性，而不管是否明确地描述。

[0015] 阐述许多具体细节以便提供对所描述的主题的一个或多个方面的透彻理解。然而，应当领会，可以在没有这些具体细节的情况下实践这些方面。尽管某些部件以框图形式示出以描述一个或多个方面，但是应当理解，由单个部件执行的功能性可以由多个部件执行。类似地，单个部件可以被配置成执行被描述为由多个部件执行的功能性。

[0016] 现在参考附图更详细地描述本发明公开的各个方面，其中自始至终，同样的附图标记通常指代同样的或对应的元件。附图和具体实施方式不旨在将所要求保护的主题限制于所描述的特定形式。相反，意图是涵盖落入所要求保护的主题的精神和范围内的所有变型、等同物和备选物。

[0017] 图1图示了作为可以实现所描述的主题的各方面的示例性操作环境的实施例的操作环境100。应当领会，所描述的主题的各方面可以通过各种类型的操作环境、计算机网络、平台、框架、计算机体系结构和/或计算设备来实现。

[0018] 操作环境100的实现方式可以在被配置成根据所描述的主题的各方面来执行各种步骤、方法和/或功能性的计算设备和/或计算机系统的上下文中进行描述。应当领会，计算机系统可以由一个或多个计算设备来实现。还可以在“计算机可执行指令”的上下文中描述操作环境100的实现方式，该计算机可执行指令被执行以根据所描述的主题的各方面执行各种步骤、方法和/或功能性。

[0019] 通常，计算设备和/或计算机系统可以包括一个或多个处理器和存储设备(例如，存储器和磁盘驱动器)以及各种输入设备、输出设备、通信接口和/或其他类型的设备。计算设备和/或计算机系统还可以包括硬件和软件的组合。应当领会，各种类型的计算机可读存储介质可以是计算设备和/或计算机系统的一部分。如本文中所使用的，术语“计算机可读存储介质”和“计算机可读存储媒体”不意味且明确地排除传播的信号、经调制的数据信号、载波或任何其他类型的暂态计算机可读介质。在各种实现方式中，计算设备和/或计算机系统可以包括被配置成执行计算机可执行指令的处理器以及存储计算机可执行指令的计算机可读存储介质(例如，存储器和/或附加硬件存储装置)，该计算机可执行指令被配置成执行根据所描述的主题的各方面的各种步骤、方法和/或功能性。

[0020] 计算机可执行指令可以以诸如通过计算机程序(例如，客户端程序和/或服务器程序)、软件应用程序(例如，客户端应用和/或服务器应用)、软件代码、应用代码、源代码、可执行文件、可执行部件、程序模块、例程、应用编程接口(API)、功能、方法、对象、属性、数据结构、数据类型等之类的各种方式来体现和/或实现。计算机可执行指令可以存储在一个或

多个计算机可读存储介质上，并且可以由一个或多个处理器、计算设备和/或计算机系统执行以执行特定任务或者根据所描述的主题的各方面实现特定数据类型。

[0021] 如所示出的，操作环境100可以包括客户端设备101至105，其例如由适合于执行根据所描述的主题的各方面的操作的各种类型的计算设备实现。在各种实现方式中，客户端设备101至105可以通过网络106彼此和/或与计算机系统110进行通信。

[0022] 网络106可以由任何类型的网络或网络组合来实现，包括但不限于：诸如因特网的广域网(WAN)、局域网(LAN)、对等(P2P)网络、电话网络、私有网络、公共网络、分组网络、电路交换网络、有线网络和/或无线网络。客户端设备101至105和计算机系统110可以使用各种通信协议(例如，因特网通信协议、WAN通信协议、LAN通信协议、P2P协议、电话协议和/或其他网络通信协议)、各种认证协议(例如，Kerberos认证、NT LAN管理器(NTLM)认证、摘要认证和/或其他认证协议)和/或各种数据类型(基于web的数据类型、音频数据类型、视频数据类型、图像数据类型、消息传送数据类型、信令数据类型和/或其他数据类型)经由网络106进行通信。

[0023] 计算机系统110可以由一个或多个计算设备实现，诸如服务器计算机，其被配置成根据所描述的主题的各方面提供各种类型的服务和/或数据存储。示例性服务器计算机可以包括但不限于web服务器、前端服务器、应用服务器、数据库服务器(例如，SQL服务器)、域控制器、域名服务器、目录服务器和/或其他合适的计算机。

[0024] 计算机系统110可以被实现为分布式计算系统，其中部件位于通过网络(例如，有线和/或无线)和/或其他形式的直接连接和/或间接连接彼此连接的不同计算设备上。计算机系统110的部件可以由软件、硬件、固件或其组合来实现。例如，计算机系统110可以包括通过存储在一个或多个计算机可读存储介质上并且被执行以根据所描述的主题的各方面执行各种步骤、方法和/或功能性的计算机可执行指令实现的部件。

[0025] 在一些实现方式中，计算机系统110可以使用冗余和地理上分散的数据中心来提供托管的和/或基于云的服务，其中每个数据中心包括物理服务器的基础设施。例如，计算机系统110可以由数据中心的物理服务器实现，其提供共享的计算和存储资源并且托管具有用于执行不同任务连同提供基于云的服务的具有各种角色的虚拟机。示例性虚拟机角色可以包括但不限于web服务器、前端服务器、应用服务器、数据库服务器(例如，SQL服务器)、域控制器、域名服务器、目录服务器和/或其他合适的机器角色。

[0026] 在利用用户相关数据的实现方式中，为了用户隐私和信息保护起见，这种用户相关数据的提供商(例如，客户端设备101至105、应用等)和消费者(例如，计算机系统110、web服务、基于云的服务等)可以采用多种机制。这样的机制可以包括但不限于需要授权监控、收集或报告数据；使得用户能够选择参与和选择不参与数据监控、收集和报告；采用隐私规则来防止某些数据被监控、收集或报告；提供用于匿名化、截断或混淆敏感数据的功能性，该敏感数据先前被准许监控、收集或报告；采用数据保留政策来保护和清除数据；和/或用于保护用户隐私的其他合适机制。

[0027] 多方通信会议

[0028] 根据所描述的主题的各方面，客户端设备101至105中的一个或多个客户端设备和/或计算机系统110可以执行涉及传送在多方通信会话的上下文中标识当前说话者的元数据的各种操作。

[0029] 在一个实现方式中，客户端设备101至105可以由参与者用于发起、加入和/或参与多方通信会话(例如，音频、视频和/或web会议、在线会议等)用于经由音频会议(例如，VoIP音频会议、电话会议)、视频会议、web会议、即时消息传送和/或桌面共享中的一个或多个来实时通信。在多方通信会话期间的不同时间，参与者可以使用客户端设备101至105通过说话、收听、聊天(例如，即时消息传送、文本消息传送等)、呈现共享内容(例如，文档、演示、图纸、图形、图像、视频、应用、文本、注释等)、和/或查看共享内容来彼此进行交流和协作。

[0030] 在图1所示的示例性场景中，多方通信会话可以包括位于同一地点并且共享相同客户端设备的多个参与者。例如，客户端设备101可以是诸如台式计算机之类的固定计算设备，该台式计算机正在被位于同一地点的参与者使用和共享，这些参与者包括发言人A、发言人B和发言人C。在多方通信会话期间，位于同一地点的参与者发言人A至发言人C可以轮流说话和/或向多个远程参与者呈现共享内容，这些参与者包括观察者A、观察者B和观察者C。应当理解，发言人A至C和观察者A至C表示不同的人类参与者，并且客户端设备101至105和参与者(例如，发言人/演讲者和观察者/听众)的数目、类型和角色是为了说明的目的而被提供的。

[0031] 如所示出的，客户端设备102可以是诸如台式计算机之类的固定客户端设备，其由观察者A使用，以呈现多方通信会话的视觉或视听表示。同样如所示出的，参与者可以使用无线和/或移动客户端设备来参与多方通信会话。例如，客户端设备103可以是观察者B使用的膝上型计算机，并且客户端设备104可以是观察者C使用的智能手机。为了方便位于同一地点的参与者，发言人A至发言人C中的一个发言人可以附加地使用客户端设备105(例如，座机电话、会议免提电话等)拨入多方通信会话并且传送多方通信会话的音频。如上文所提及的，网络106可以表示网络的组合，其可以包括公共交换电话网(PSTN)或其他合适的电话网络。

[0032] 在相同音频流上存在多个说话者的情形下，远程参与者通常难以仅基于音频流来标识在任何给定时刻正在说话的人。这更是个问题，并且当远程参与者对说话者不太熟悉时尤其常见。在一些情况下，可以基于参与者在加入多方通信会话时提供的帐户信息来提供正在说话的参与者的视觉指示。例如，当参与者提供帐户信息以加入在线会议时，可以显示包括参与者的名称和/或化身(avatar)的指示符。然而，基于帐户信息来显示参与者的指示符仅在每个参与者单独地加入时有效，当多个说话者位于同一地点并且共享单个客户端设备时无效。经常存在这种情况，当位于同一地点的参与者共享客户端设备时，只有一个位于同一地点的参与者提供帐户信息以加入多方通信会话。在这种情况下，对于那些没有分别加入在线会议的共同参与者，无法提供基于被提供以加入在线会议的信息所显示的名称和/或化身。

[0033] 在参与者使用电话(例如，座机电话、会议免提电话、移动电话等)拨打多方通信会话的情形下，可以显示电话号码。然而，显示电话号码可能不足以标识正在与其他参与者说话的人。此外，当使用固定电话或会议免提电话作为附加设备来加入多方通信会话时，单个参与者可以使得两个不同的指示符被显示，其可以看起来像是存在意外的或未受邀的听众。

[0034] 当加入和/或参与多方通信会话时，多个(或全部)客户端设备101至105可以通过网络106与计算机系统110进行通信。在各种实现方式中，计算机系统110可以被配置成提供

和/或支持多方通信会话。例如，计算机系统110可以实现基于web的和/或基于云的会议服务，其被配置成提供以下各项中的一项或多项：音频会议(例如，VoIP音频会议、电话会议)、视频会议、web会议、即时消息传送和/或桌面共享。在一些实现方式中，计算机系统110可以被实现为公司内部网的一部分。

[0035] 在多方通信会话期间，客户端设备101至104可以呈现各种用户界面和/或计算机系统110可以提供各种用户界面，用于允许参与者参与多方通信会话用户界面。用户界面可以由客户端设备101至104经由web浏览应用或为多方通信会话提供用户界面的其他合适类型的应用、应用程序和/或app来呈现。在各种场景下，可以在启动允许用户加入并且参与多方通信会话的应用(例如，web浏览器、在线会议app、消息传送app等)之后呈现用户界面。

[0036] 客户端设备101至104可以被配置成接收和响应各种类型的用户输入，诸如语音输入、键盘输入、鼠标输入、触摸板输入、触摸输入、手势输入等等。客户端设备101至104可以包括硬件(例如，麦克风、扬声器、摄像机、显示屏等)以及用于捕获语音、传送的音频和/或视频数据以及输出多方通信会话的音频、视频和/或视听表示的软件。客户端设备101至104还可以包括用于提供客户侧语音和/或说话者识别的硬件和软件。当被采用时，客户端设备105(例如，座机电话、会议免提电话等)可以仅使得仅能够传送音频。

[0037] 计算机系统110可以被配置成执行涉及接收和处理表示当前说话者的语音的音频数据(例如，输入音频流)并且传送标识当前说话者的元数据的各种操作。计算机系统110可以以各种方式接收表示当前说话者的语音的音频数据。在一个实现方式中，客户端设备101至105之间的通信可以通过计算机系统110进行，并且从客户端设备101至105中的任一个输出的音频数据可以流入计算机系统110并且可以被实时地处理。在另一实现方式中，计算机系统110可以自动地或者响应于来自参与者的请求来支持多方通信会话，并且可以接收和分析客户端设备101至105之间的音频通信。

[0038] 计算机系统110可以包括说话者识别部件111，其被配置成处理表示当前说话者的语音的音频数据。从客户端设备101至105中的任一个输出的音频数据可以由说话者识别部件111接收和/或送往说话者识别部件111以供处理。在各种实现方式中，说话者识别部件111可以处理表示当前说话者的语音的音频数据，以生成当前说话者的音频指纹。例如，说话者识别部件111可以处理音频数据以确定和/或提取可以用于表示当前说话者的话音的一个或多个语音特征，并且可以基于音频数据的一个或多个语音特征来生成当前说话者的音频指纹。示例性语音特征可以包括但不限于音高、能量、强度、持续时间、过零率、节奏、韵律、音调、音色、共振峰位置、共振属性、短时频谱特征、线性预测系数、梅尔频率倒频谱系数(MFCC)、语音学、语义学、发音、方言、口音和/或其他用于区别的语音特性。当前说话者的音频指纹可以实现为以下各项中的一项或多项：声纹、话音生物特征、说话者模板、特征向量、特征向量的集合或序列、诸如特征向量分布之类的说话者模型、诸如通过对所提取的特征向量的序列执行向量量化所产生的代表性编码向量或质心的集合之类的说话者码本、和/或其他合适的说话者指纹。

[0039] 说话者识别部件111可以执行自动说话者识别，以利用当前说话者的音频指纹识别或尝试识别当前说话者。在各种实现方式中，说话者识别部件111可以将当前说话者的音频指纹与已经被计算机系统110和/或说话者识别部件111先前识别的个人的已存储音频指纹进行比较。例如，说话者识别部件111可以确定当前说话者的音频指纹的特征和已存储音

频指纹的特征之间的距离，诸如欧几里德距离。说话者识别部件111可以基于已存储音频指纹来识别当前说话者，该指纹具有最小的距离，只要这种存储的音频指纹足够接近，足以做出肯定标识。示例性特征匹配技术可以包括但不限于：隐马尔可夫模型(HMM)技术、动态时间扭曲(DTW)技术、向量量化技术、神经网络技术和/或其他合适的比较技术。

[0040] 计算机系统110可以包括说话者指纹存储库112，其被配置成存储先前由计算机系统110和/或说话者识别部件111识别的个人的音频指纹。说话者指纹存储库112可以被实现为存储音频指纹的公共存储库，并且可以将每个存储的音频指纹与诸如个人的名称或其他合适身份之类的元数据相关联。说话者指纹存储库112可以存储针对各种人群的音频指纹。例如，说话者指纹存储库112可以与组织机构相关联并且被配置成存储组织机构的雇员、成员和/或会员的音频指纹。

[0041] 计算机系统110可以包括目录113，其被配置成维持与特定组织机构或群组相关联的个人有关的各种类型的个人和/或职业信息。例如，目录113可以包含个人的组织信息，诸如名称、公司、部门、职位名称、联系信息、化身(例如，简档图片)、电子名片和/或其他类型的简档信息。在各种实现方式中，目录113和说话者指纹存储库112可以彼此关联。例如，存储在目录113中的个人的个人或职业信息可以引用或包括所存储的针对个人的音频指纹。可替代地或附加地，与说话者指纹存储库112中存储的音频指纹相关联的元数据可以引用或包括来自目录113的各种类型的信息。虽然被示出为单独的存储设施，但是在一些部署中，目录113和说话者指纹存储库112可以被一体化。

[0042] 在一些实现方式中，计算机系统110可以包括参与者数据存储装置114，其被配置成在多方通信会话期间临时存储用于快速访问和使用的参与者信息。例如，当参与者(例如，发言人A和观察者A至观察者C)提供账户信息用于单独加入在线会议时，可以从目录113访问可用于这些参与者的个人或职业信息，并且临时存储在参与者数据存储装置114中，用于在呈现在线会议的用户界面中显示名称、化身和/或其他组织信息。类似地，可以从说话者指纹存储库112访问可用于这些参与者的已存储音频指纹，并且临时存储在参与者数据存储装置114中以便于说话者识别。当参与者数据存储装置114包含参与者的一个或多个存储的音频指纹时，说话者识别部件111可以通过首先将生成的当前说话者的音频指纹与参与者数据存储装置114中存储的音频指纹进行比较来执行自动说话者识别，并且如果自动说话者识别不成功，则将当前说话者的音频指纹与说话者指纹存储库112中存储的音频指纹进行比较。尽管被示出为单独的存储设施，但是在一些部署中，参与者数据存储装置114和说话者指纹存储库112可以被实现为在一些存储设施内的单独区域。

[0043] 在一些实现方式中，客户端设备101至104中的一个或多个客户端设备可以包括语音和/或说话者识别功能性，并且可以与计算机系统110共享参与者的本地音频指纹。由客户端设备101至104中的一个或多个客户端设备提供的参与者的本地音频指纹可以临时存储在参与者数据存储装置114中以便在当前的多方通信会话期间进行访问，并且还可以被持久地存储在说话者指纹存储库112和/或目录113中，以用于随后的多方通信会议。在一些情况下，由客户端设备101至104中的一个或多个客户端设备提供的参与者的本地音频指纹可以用于增强、更新和/或替换由计算机系统110维持的现有的参与者的存储的音频指纹。

[0044] 计算机系统110和/或说话者识别部件111可以通过将当前说话者的所生成的音频指纹与可能包含在说话者指纹存储库112、目录113和/或参与者数据存储装置114中的一个

或多个中的已存储音频指纹进行比较来执行自动说话者识别。如果基于已存储音频指纹成功地识别了当前的说话者，则计算机系统110和/或说话者识别部件111可以取回与已存储音频指纹相关联的个人的名称或其他合适的身份，并且传送标识当前说话者的元数据。在各种实现方式中，标识当前说话者的元数据可以通过网络106传送到客户端设备101至104。响应于接收到这样的元数据，客户端设备101至104可以在呈现多方通信会话的用户界面内显示标识当前说话者的指示符。可替代地，为了避免打断或分散当前说话者的注意力，标识当前说话者的指示符可以仅由除了当前说话者之外的参与者正在使用的客户端设备显示，和/或标识当前说话者的元数据可以仅被传送到除了当前说话者之外的参与者使用的客户端设备。

[0045] 在一些实现方式中，标识当前说话者的元数据可以包括附加信息，例如以下各项中的一项或多项：当前说话者的公司、当前说话者的部门、当前发言人的职务名称、当前说话者的联系信息和/或当前说话者的其他简档信息。这样的附加信息可以例如从目录113取回并且包括在由计算系统110传送的元数据中。接收这种元数据的客户端设备101至104中的任一客户端设备可以在标识当前说话者的指示符内显示附加信息，或在呈现多方通信会话的用户界面内的其他地方显示附加信息。可替代地，元数据可以仅包括当前说话者的名称或身份，其可以由接收这样的元数据的客户端设备101至104中的任一客户端设备用于查询目录113以获得附加信息并且使用附加信息填充标识当前说话者的指示符和/或用户界面。

[0046] 标识当前说话者的指示符还可以包括确认按钮或用于请求参与者确认当前说话者已经被正确识别的其他合适的图形显示元件。在从一个或多个参与者接收当前说话者的身份的确认之后，可以使用由说话者识别部件111生成的当前说话者的音频指纹和/或表示当前说话者的语音的其他音频数据来进一步增强由说话者指纹存储库112维持的当前说话者的已存储音频指纹。

[0047] 计算机系统110可以包括音频数据充实部件115，该音频数据充实部件115被配置成利用标识当前说话者的元数据来增强表示当前说话者的语音的音频数据。例如，当说话者识别部件111成功地识别当前说话者时，输入音频数据(例如，输入音频流)可以被导向音频数据充实部件115以供处理。在各种实现方式中，音频数据充实部件115可以通过基于时间来生成与音频流数据同步的元数据流来处理音频流。元数据流可以被实现为标识所识别的说话者的元数据序列，并且包括每个已识别的说话者的语音的开始时间和结束时间。例如，当检测到新识别的说话者时，可以创建用于先前已经识别的说话者的结束时间和新识别的说话者的开始时间。音频数据充实部件115还可以从目录113或其他信息源取回所识别的说话者的附加个人或职业信息(例如，名称、公司、部门、职位名称、联系信息和/或其他简档信息)，并且将这样的附加信息包括在元数据流内。增强的音频流可以作为同步的音频和元数据流或作为与元数据分组集成的音频数据分组来传送。包括标识所识别的说话者的元数据的增强的音频数据可以由计算机系统110实时输出，并且可以由客户端设备101至104渲染以标识当前在多方通信会话中正在说话的每个所识别的说话者。

[0048] 在一些实现方式中，客户端设备101至104中的一个或多个客户端设备可以包括语音和/或说话者识别功能性，并且正在说话的参与者的客户端设备可以与计算机系统110共享标识正在说话的参与者的本地元数据。由客户端设备101至104中的一个或多个客户端设

备提供的本地元数据可以由计算机系统110接收并且传送到客户端设备101至104中的每个客户端设备，或者仅传送到正在除了当前说话者之外的参与者使用的客户端设备。由客户端设备101至104中的一个或多个客户端设备提供的本地元数据也可以由计算机系统110接收并且使用标识当前说话者的元数据来增强表示当前说话者的语音的音频数据。

[0049] 计算机系统110可以包括充实音频数据存储库116，其被配置成存储已经用标识一个或多个已识别的说话者的元数据来增强的充实音频数据。在各种实现方式中，音频数据充实部件115可以在多方通信会话发生时创建包括多个说话者的会话记录。所记录的对话可以包括与音频数据同步并且标识所识别的说话者的元数据。例如，在由客户端设备101至104中的一个或多个客户端设备回放所记录的会话期间，当与所识别的说话者相关联的音频被播放时，可以渲染元数据以显示所识别的说话者的名称、身份和/或附加个人或职业信息。

[0050] 计算机系统110可以包括转录部件117，其被配置成生成多方通信会话的文本转录。转录部件117可以基于由音频数据充实部件提供的增强音频数据来实时生成文本转录。例如，当增强音频数据开始存储在充实音频数据存储库116中时，转录部件117可以同时生成增强音频数据的文本转录。可替代地，转录部件117可以基于已经存储在充实音频数据存储库116中的增强音频数据来在稍后的时间生成文本转录。在各种实现方式中，转录部件117可以转录包括一个或多个所识别的说话者的会话，并且文本转录可以包括与所识别的说话者所说出的语音文本相关联的用于所识别的说话者的标识符。

[0051] 计算机系统110可以包括转录存储库118，其被配置成存储多方通信会话的文本转录。转录库118可以维持与当前多方通信会话相关联的一个或多个已存储转录以及与其他多方通信会话相关联的已存储转录。在各种实现方式中，转录部件117可以在多方通信会话结束时和/或在参与者请求的某个其他时间点向客户端设备101至104提供文本转录。

[0052] 计算机系统110可以包括搜索和过滤部件119，其被配置成使用标识所识别的说话者的元数据执行搜索和/或过滤操作。例如，搜索和过滤部件119可以接收指示所识别的说话者的查询，搜索标识所识别的说话者的元数据，以及执行各种动作和/或提供各种类型的输出。在一些情况下，搜索和过滤部件119可以从增强音频数据搜索标识所识别的说话者的元数据，并且提供表示所识别的说话者的语音的增强音频数据的部分。搜索和过滤部件119还可以从一个或多个文本转录搜索所识别的说话者的标识符，并且提供包括所识别的说话者所说出的语音的文本的转录的部分。

[0053] 在各种实现方式中，搜索和过滤部件119可以对正在被实时传送的增强音频数据执行操作。例如，在多方通信会话期间，可以获得根据上下文与特定的所识别的说话者相关联的附加文档或其他类型的内容，并且可以向多方通信会话的参与者提供。参与者还可以请求表示特定的所识别的说话者(例如，主发言人)的语音的音频数据相对于在一些情况下可能与特定的所识别的说话者共享相同的客户端设备的其他说话者被渲染得更大声。

[0054] 计算机系统110可以包括警报部件120，其被配置成当特定的所识别的说话者当前正在说话时生成警报。例如，警报部件120可以接收来自参与者的标识特定的所识别的说话者的的请求，对包括在正在实时传送的增强音频数据中的元数据进行操作，并且向参与者的客户端设备传送数据以在每当特定的所识别的说话者说话时生成可听见的和/或视觉警报。

[0055] 计算机系统110可以包括标记部件121，其被配置成在计算机系统110无法识别当前说话者的情况下，请求和接收指定当前说话者的身份的标记信息。例如，可能存在说话者指纹储存库112不包含当前说话者的存储的音频指纹和/或说话者识别部件111不能将当前说话者的生成的音频指纹与存储的音频指纹(该音频指纹足够近，足以做出肯定识别)相匹配的情形。

[0056] 在基于已存储音频指纹未能成功地识别当前说话者的情形下，计算机系统110和/或标记部件121可以传送数据以呈现说话者未被识别的指示(例如，消息、图形等)。计算机系统110和/或标记部件121还可以请求一个或多个参与者以提供标记信息，诸如未被识别的说话者的名称或其他合适的身份。在各种实现方式中，可以通过网络106向客户端设备101至104传送用于标记信息的请求。响应于接收到用于标记信息的请求，客户端设备101至104中的每个客户端设备可以显示对话框，其提示每个参与者在呈现多方通信会话的用户界面内标记未被识别的说话者。为了避免中断或分散当前说话者的注意力，该对话框可以仅由除当前说话者之外的参与者使用的客户端设备显示，和/或用于标记信息的请求可以仅传送到正在被除了当前说话者以外的参加者使用的客户端设备。

[0057] 客户端设备101至104中的一个或多个客户端设备可以响应于参与者手动输入未被识别的说话者的名称或身份、参与者从目录113选择用户简档信息、参与者选择为呈现多方通信会话的用户界面中的未被识别的当前说话者所显示的化身、和/或其他参与者标识未被识别的说话者的输入，来提供未被识别的说话者的标记信息。在一些情况下，标记部件121可以从一个或多个参与者接收标识未被识别的说话者的标记信息。在不同的参与者为当前说话者提供冲突的标记信息(例如，不同的名称)的情况下，标记部件121可以基于由大多数参与者提供的身份或其他合适的启发式方法标识当前说话者来解决该冲突。

[0058] 标记部件121接收到标识未被识别的说话者的标记信息之后，可以通过网络106向客户端设备101至104传送标识当前说话者的元数据。响应于接收到这样的元数据，客户端设备101至104中的每个客户端可以在呈现多方通信会话的用户界面内显示标识当前说话者的指示符。可以从包括在标识当前说话者的元数据中的目录113或其他信息源中取回用于当前说话者的组织信息(例如，名称、公司、部门、职位名称、联系信息和/或其他简档信息)，并显示在标识当前说话者的指示符中，或显示在呈现多方通信会话的用户界面内的其他地方内。为了避免中断或分散当前说话者的注意力，标识当前说话者的指示符可以仅由除了当前说话者之外的参与者正在使用的客户端设备显示，和/或标识当前说话者的元数据可以仅传送给正在被除了当前演讲者之外的参与者使用的客户端设备。

[0059] 标识当前说话者的指示符还可以包括确认按钮或用于请求参与者确认当前说话者已经被正确识别的其他合适的图形显示元件。在从一个或多个参与者接收到对当前说话者的身份的确认之后，由说话者识别部件111生成的当前说话者的音频指纹可以与标识当前说话者的元数据相关联，并且存储在说话者指纹存储库112、目录113或参与者数据存储装置114中的一个或多个中。当前用户的已存储音频指纹可以用于在当前多方通信会话或随后的多方通信会话期间执行说话者识别。例如，计算机系统110可以接收表示当前说话者的语音的后续音频数据，基于随后的音频数据来生成当前说话者的新音频指纹，并且通过将当前说话者的新音频指纹与当前说话者的存储的音频指纹进行比较来执行自动说话者识别。

[0060] 示例性用户界面

[0061] 图2A至图2D图示了作为可以实现所描述的主题的各方面的示例性用户界面的实施例的用户界面200。应当领会，所描述的主题的各方面可以由各种类型的用户界面来实现，其可以由客户端设备101至104或其他合适的计算设备呈现和/或由计算机系统110或其他合适的计算机系统提供。

[0062] 参考图2A，继续参考前述附图，用户界面200可以由客户端设备101至104中的一个或多个客户端设备显示，以呈现在线会议的视觉表示。用户界面200可以在示例性场景期间呈现，其中发言人A和观察者A至观察者C分别使用客户端设备101至104以加入在线会议，并且单独地向提供和/或支持在线会议的计算机系统110提供帐户信息。在该示例性场景中，发言人A至发言人C位于同一地点(例如，在同一会议室中)并且使用客户端设备101来呈现在线会议的视觉或视听表示并且与观察者A至观察者C共享内容。在该示例性场景中，发言人A还经由客户端设备105(例如，座机电话、会议免提电话等)拨入在线会议，然后该客户端设备105可以由发言人A至发言人C用来传送在线会议的音频。

[0063] 如上文所描述的，计算机系统110可以包括说话者指纹存储库112，其被配置成存储发言人A至发言人C和观察者A至观察者C的音频指纹；并且可以包括先前已经被计算机系统110识别的任何参与者的已存储音频指纹和相关联的元数据(例如，名称、身份等)。计算机系统110可以包括目录113，其被配置成维持与发言人A至发言人C和观察者A至观察者C中的一个或多个有关的各种类型的个人和/或职业信息(例如，名称、公司、部门、职务名称、联系信息、化身(例如，简档图片)、电子名片和/或其他类型的简档信息)。目录113可以引用或包括用于个人的存储的音频指纹和/或与说话者指纹存储库112中存储的音频指纹相关联的元数据，说话者指纹存储库112可以引用或包括来自目录113的各种类型的信息。

[0064] 在一些实现方式中，计算机系统110可以包括参与者数据存储装置114，其被配置成当发言人A和观察者A至观察者C加入在线会议时，临时存储参与者信息(例如，来自目录113的简档信息、已存储音频指纹和来自说话者指纹存储库112的元数据等)。当发言人A和观察者A至观察者C加入在线会议时，计算机系统110还可以从包括语音和/或说话者识别功能性的任一个客户端设备101至104请求和接收本地音频指纹和/或本地元数据。如果提供了本地音频指纹，则可以存储在参与者数据存储装置114、说话者指纹存储库或目录113中的一个或多个中，以执行自动化说话者识别和/或可以用于增强、更新和/或替换由计算机系统110维持的现有存储的参与者的音频指纹。

[0065] 用户界面200可以指示在在线会议中检测到的参与者的数目，并且可以包括被配置成列出在线会议的参与者的参与者框201。如所示出的，用户界面200可以基于客户端设备101至105来指示五个检测到的参与者，并且参与者框201可以基于由参与者在加入在线会议时提供的帐户信息来显示发言人A、观察者A、观察者B以及观察者C的名称。参与者名称可以从目录112和/或参与者数据存储装置114获得。参与者框201还可以显示客户端设备105的电话号码以及有访客在线会议中的存在的指示。参与者框201可以指示每个参与者是发言人还是参加者以及发言人和参加者的数目。在该示例中，参与者名称将基于帐户信息进行显示，而不提供未单独提供帐户信息以加入在线会议的发言人B和发言人C。

[0066] 用户界面200可以呈现被配置成显示传送给在线会议的所有参与者的即时消息的消息框202。消息框202可以允许参与者实时聊天并且显示包括在线会议期间提交的所有即

时消息的即时消息对话。消息框202可以包括消息撰写区域203，用于撰写要向在线会议的所有参与者显示的新即时消息。

[0067] 用户界面200可以向在线会议的参与者显示演示204。演示204可以表示可以与在线会议的参与者共享的各种类型的内容(例如，文档、演示、图纸、图形、图像、视频、应用、文本、注释等)。例如，演示204可以是幻灯片放映或由发言人A共享并且由客户端设备101至104显示的其他类型的文档。

[0068] 用户界面200可以显示参与者的名称和化身。如所示出的，用户界面200可以基于当加入在线会议时由参与者提供的帐户信息来显示发言人A名称和化身205、观察者A名称和化身206、观察者B名称和化身207以及观察者C名称和化身208。可以从目录112和/或参与者数据存储装置114获得参与者名称和化身205至208(例如，简档图片)。在该示例中，参与者名称和化身205至208基于帐户信息进行显示，而不提供未提供账户信息以加入在线会议的发言人B和发言人C。用户界面200还可以响应于发言人A使用客户端设备105拨入在线会议来显示客户端设备105的电话号码和通用化身209。

[0069] 参考图2B，继续参考前述附图，可以在发言人A正在说话的示例性情形下显示用户界面200，并且通过客户端设备105(例如，座机电话、会议免提电话等)向计算机系统110传送表示发言人A的语音的音频数据。在该示例性情形下，用户界面200可以由客户端设备101至104中的每个客户端设备显示，或仅由正在被不是发言人A的参与者使用的客户端设备102至104显示。

[0070] 计算机系统110可以接收和处理表示发言人A的语音的音频数据，以生成发言人A的音频指纹。计算机系统110可以通过将Participant A的音频指纹与包含在说话者指纹存储库112、目录113和/或参与者数据存储装置114中的一个或多个中的存储的音频指纹进行比较来执行自动说话者识别。在该示例性情形下，计算机系统110基于为发言人A创建的音频指纹和存储的音频指纹成功地识别发言人A，并且从与已存储音频指纹相关联的元数据取回名称或身份(例如，发言人A的名称或用户名)。

[0071] 计算机系统110还从目录113或其他信息源取回组织信息，其包括名称、公司、部门、职位名称、联系信息、以及化身或者对于发言人A的化身的引用。计算机系统110使用包括发言人A的名称或身份以及组织信息的元数据来增强表示发言人A的语音的音频数据(例如，输入音频流)，并且实时地向客户端设备101至104或仅向客户端设备102至104传送增强音频数据(例如，同步音频数据和元数据流)。

[0072] 在从计算机系统110接收到增强音频数据时，客户端设备101至104中的一个或多个客户端设备可以渲染音频以听到发言人A的语音，同时显示标识发言人A并且包括发言人A的组织信息的元数据。如所图示的，用户界面200可以显示呈现当前说话者名称和化身211的当前说话者框210，其在这种情况下呈现发言人A的名称和化身。当前说话者名称和化身211可以被包括在元数据中和/或在元数据中被引用，或者可以使用包括在元数据中的名称或身份从目录113或其他信息源取回。当前说话者框210还从元数据中呈现组织信息212，在这种情况下，该元数据包括发言人A的名称、公司、部门、职务名称以及联系信息。

[0073] 当前说话者框210显示确认按钮213，其可以由参与者点击或触摸以确认当前说话者已经被正确识别。在接收到发言人A已经被正确标识为当前说话者的确认之后，可以使用发言人A的所生成的音频指纹和/或表示发言人A的语音的其他音频数据来增强所存储的发

言人A的音频指纹。当前说话者框210还显示标记按钮214，其可以被参与者点击或触摸，以为已经被错误地标识的当前说话者提供经校正的名称或身份。

[0074] 如所示出的，当前说话者框210呈现警报按钮215，其可以由参与者部件120点击或触摸以在每当发言人A当前正在说话时接收可听见的和/或视觉警报。例如，如果发言人A停止说话并且另一参与者说话，则当发言人A恢复说话时，可能会生成警报。

[0075] 当前说话者框210还显示“更多”按钮216，其可以由参与者部件120点击或触摸以呈现进一步的信息和/或执行附加的动作。例如，该“更多”按钮216可以提供根据上下文与发言人A有关的文档或其他类型的内容的访问和/或可以使得表示发言人A的语音的音频数据能够相对于其他说话者被渲染得更大声。

[0076] 包括标识发言人A的元数据的增强音频数据可以用于将发言人A标识为当前说话者，并且还可以由计算机系统110存储和/或用于生成在线会议的文本转录。可以搜索和/或过滤由计算机系统110维持的增强音频数据和文本转录，以提供表示发言人A的语音的部分。

[0077] 附加地，在该示例中，计算机系统110基于从客户端设备105(例如，固定电话、会议免提电话等)接收的音频数据来识别发言人A，并且修改和/或传送数据以修改用户界面200的显示。如所示出的，可以修改用户界面200，以指示包括四个检测到的参与者的在线会议基于计算机系统110确定发言人A正在使用客户端设备101和客户端设备105两者。参与者框201也可以被改变以去除客户端设备105的电话号码和访客存在的指示。客户端设备105的电话号码和通用化身209也可以从用户界面200移除。在图2B中，用户界面200仍然不提供尚未说话并且没有单独提供帐户信息以加入在线会议的发言人B和发言人C的名称或化身。

[0078] 参考图2C，继续参考前述附图，用户界面200可以在演示者B正在说话的示例性情形下显示，并且表示发言人B的语音的音频数据被实时地传送到计算机系统110。发言人B可以使用被实现为台式计算机或其他合适设备的客户端设备101，或者可以使用由发言人A至发言人C共用的客户端设备105(例如，座机电话、会议免提电话等)。在该示例性情形下，用户界面200可以由客户端设备101至104中的每个客户端设备显示，或者仅由正在被除了发言人B之外的参与者使用的客户端设备102至104显示。

[0079] 计算机系统110可以接收和处理表示发言人B的语音的音频数据，以生成发言人B的音频指纹。计算机系统110可以通过将参与者B的音频指纹与包含在说话者指纹存储库112、目录113和/或参与者数据存储装置114中的一个或多个中的存储的音频指纹进行比较来执行自动说话者识别。在该示例性情形下，计算机系统未维持所存储的发言人B的音频指纹，并且未识别当前说话者。

[0080] 计算机系统110修改和/或传送数据以修改用户界面200的显示，以指示当前说话者未被识别。用户界面200可以显示当前说话者框210，其呈现图形217和指示当前说话者未被识别的消息218。当前说话者框210还显示标记按钮214，其可以由参与者点击或触摸以提供用于未被识别的说话者的名称或身份。

[0081] 同样如所示出的，可以修改用户界面200，以指示包括五个检测到的参与者的在线会议基于计算机系统110检测新的说话者。参与者框201可以被改变以向参与者和发言人的列表添加未知的说话者，和/或可以在用户界面200中添加未知的说话者化身219。

[0082] 参考图2D，继续参考前述附图，可以在发言人B正在说话并且计算机系统110从客

户端设备102至104接收到冲突的标记信息的示例性情形下显示用户界面200。例如，观察者A和观察者B可能提供标识发言人B的标记信息，而观察者C提供标识发言人C的标记信息。在该示例性情形下，用户界面200可以由客户端设备101至104中的每个客户端设备显示，或者仅由除了发言人B之外的参与者使用的客户端设备102至104显示。

[0083] 计算机系统110基于少数服从多数规则或其他合适的用于解决冲突的标记信息的启发式方法来将当前说话者识别为发言人B。计算机系统110使用标记信息(例如，发言人B的名称或用户名)来识别发言人B，并且从目录113或其他信息源取回组织信息，包括名称、公司、部门、职务名称、联系信息、以及发言人B的化身或对其化身的引用。计算机系统110利用包含发言人B的名称或身份以及组织信息的元数据来增强表示发言人B的语音的音频数据(例如，输入音频流)，并且实时地向客户端设备101至104或仅向客户端设备102至104传送增强音频数据(例如，同步的音频数据和元数据流)。

[0084] 在从计算机系统110接收到增强音频数据时，客户端设备101至104中的一个或多个客户端设备可以渲染音频以听到发言人B的语音，同时显示标识发言人B并且包括发言人B的组织信息的元数据。如所图示的，用户界面200可以显示当前说话者框210，其呈现当前说话者名称和化身211，在这一情况下呈现发言人B的名称和化身。当前说话者名称和化身211可以被包括在元数据中和/或在元数据中被引用，或者可以使用包含在元数据中的名称或身份从目录113或其他信息源取回。当前说话者框210还呈现来自元数据的组织信息212，其在这种情况下包括发言人B的名称、公司、部门、职务名称和联系信息。

[0085] 当前说话者框210显示确认按钮213，其可以由参与者点击或触摸以确认当前说话者已经被正确识别。在收到发言人B被正确标识为当前说话者的确认之后，所生成的发言人B的音频指纹可以存储在说话者指纹存储库112、目录113和/或参与者数据存储装置114中的一个或多个中。当前说话者框210还显示标记按钮214，其可以由参与者点击或触摸，以为已经被错误地标识的当前说话者提供经校正的名称或身份。

[0086] 如所示出的，当前说话者框210呈现警报按钮215，其可以由参与者部件120点击或触摸以在每当发言人B正在说话时接收可听和/或视觉警报。例如，如果发言人B停止说话并且另一参与者说话，当发言人B恢复讲话时可能会生成警报。

[0087] 当前说话者框210还显示“更多”按钮216，其可以由参与者部件120点击或触摸以呈现进一步的信息和/或执行附加的动作。例如，附加按钮216可以提供根据上下文与发言人B有关的文档或其他类型的内容的访问和/或可以使得表示发言人B的语音的音频数据能够相对于其他说话者被渲染得更大声。

[0088] 包括标识发言人B的元数据的增强音频数据可以用于将发言人B标识为当前说话者，并且还可以由计算机系统110存储和/或用于生成在线会议的文本转录。可以搜索和/或过滤由计算机系统110维护的增强音频数据和文本转录，以提供表示由计算机系统110先前识别的发言人B和/或发言人A的语音的部分。

[0089] 附加地，在该示例中，计算机系统110修改或传送数据以修改用户界面200的显示。如所示出的，参与者框201可以被改变，以基于计算机系统110识别出发言人B将发言人B的名称添加到参与者和发言人的列表中。还有，发言人B的名称和化身220可以代替用户界面200中的未知说话者化身219。

[0090] 在计算机系统110存储发言人B的音频指纹之后，图2D所示的用户界面200也可以

响应于计算机系统110接收表示发言人A或发言人B的语音的后续音频数据而被显示。例如，如果发言人B停止说话而发言人A讲话，则计算机系统110可以接收标识发言人A的语音的后续音频数据，基于后续音频数据来生成发言人A的新音频指纹，并且通过将发言人A的新音频指纹与发言人A的存储的音频指纹来成功地执行自动说话者识别。在发言人A正在讲话时，图2D所示的用户界面可以显示有当前说话者名称和化身211，其示出了发言人A的名称和化身；以及显示有组织信息212，其示出了发言人A的名称、公司、部门、职务名称和联系信息。

[0091] 同样地，当发言人B恢复说话时，计算机系统110可以接收表示发言人B的语音的后续音频数据，基于后续音频数据来生成发言人B的新音频指纹，并且通过将发言人B的新音频指纹与发言人B的存储的音频指纹进行比较来成功执行自动说话者识别。当发言人B恢复说话时，图2D所示的用户界面可以显示有当前说话者名称和化身211，其示出了发言人B的名称和化身；以及显示有组织信息212，其示出了发言人B的名称、公司、部门、职务名称和联系信息。

[0092] 还可以领会，图2D所示的用户界面200还没有提供即没有发言也没有单独提供帐户信息以加入在线会议的发言人C的名称或化身。当发言人C最后说话时，计算机系统110可以以与上文关于位于同一地点的发言人B所描述的方式类似的方式来识别发言人C，传送标识发言人C的元数据和/或修改或传送数据以修改用户界面200以便标识发言人C。

[0093] 填充指纹存储库

[0094] 计算机系统110可以以各种方式使用存储的音频指纹来填充说话者指纹存储库112。在一些实现方式中，计算机系统110可以通过采用离线或在线训练过程来使用音频指纹填充说话者指纹存储库112。例如，计算机系统110可以请求参与者说出预先定义的句子，基于表示所说出的句子的音频数据来为参与者创建音频指纹，并且将用于参与者的音频指纹存储在说话者指纹存储库112中以供在一个或多个多方通信会议期间使用。计算机系统110可以请求参与者执行训练过程以在多方通信会话之前离线创建音频指纹，或者在多方通信会话期间在线创建音频指纹。例如，计算机系统110可以确定说话者指纹存储库112是否包含加入多方通信的每个参与者的存储的音频指纹，并且可以请求没有存储的音频指纹的参与者创建音频指纹。

[0095] 可替代地或附加地，计算机系统110可以通过采用众包过程使用存储的音频指纹填充说话者指纹存储库112。例如，在说话者指纹存储库112不包含用于当前说话者的已存储音频指纹的情况下，计算机系统110可以采用众包过程来通过创建用于当前说话者的音频指纹并且请求客户端设备101至104中的一个或多个客户端设备来提供标记信息以使用存储的音频指纹填充说话者指纹存储库112。当接收到标识当前说话者和/或确认当前说话者的身份的标记信息时，计算机系统110可以将用于当前说话者的音频指纹和标识当前说话者的元数据存储在说话者指纹存储库112、目录113或参与者数据存储装置114中的一个或多个中，以在当前多方通信会话或后续多方通信会话期间使用。标记部件121可以促进众包过程，并且即使在说话者指纹存储库112中没有存储当前说话者的音频指纹的情况下也允许说话者识别。

[0096] 计算机系统110还可以采用众包过程来通过请求客户端设备101至104为参与者提供本地音频指纹来使用已存储音频指纹填充说话者指纹存储库112。当参与者加入多方通

信会话时，例如，计算机系统110可以向参与者请求许可，以访问对应的客户端设备101至104或从对应的客户端设备101至104接收本地音频指纹、本地元数据或其他信息，以便于说话者识别。由计算机系统110接收的参与者的本地音频指纹可以与参与者的元数据一起存储在说话者指纹存储库112、目录113或参与者数据存储装置114中的一个或多个中，以供在当前多方通信会话或后续多方通信会话期间使用。

[0097] 音频/视频说话者识别

[0098] 图3图示了作为可以实现所描述的主题的各方面的示例性操作环境的实施例的操作环境300。应当领会，所描述的主题的各方面可以通过各种类型的操作环境、计算机网络、平台、框架、计算机体系结构和/或计算设备来实现。

[0099] 操作环境300可以由一个或多个计算设备、一个或多个计算机系统和/或计算机可执行指令来实现，该计算机可执行指令被配置成根据所描述的主题的各方面来执行各种步骤、方法和/或功能性。如所示出的，操作环境100可以包括例如通过适合于根据所描述的主题的各方面来执行操作的各种类型的计算设备来实现的客户端设备301至305。在各种实现方式中，客户端设备301至305可以通过网络106彼此通信和/或与计算机系统110进行通信。网络306可以由上文所描述的任何类型的网络或网络组合来实现。

[0100] 计算机系统310可以由诸如服务器计算机311至313(例如，web服务器、前端服务器、应用服务器、数据库服务器(例如，SQL服务器)、域控制器、域名服务器、目录服务器等)的一个或多个计算设备实现，其被配置成根据所描述的主题提供各种类型的服务和/或数据存储装置。计算机系统310可以包括数据存储装置314至316(例如，数据库、云存储装置、表存储装置、二进制大型对象(blob)存储装置、文件存储装置、队列存储装置等)，其可由服务器计算机311至313访问并且被配置成根据所描述的主题存储各种类型的数据。

[0101] 计算机系统310可以被实现为分布式计算系统和/或可以提供托管和/或基于云的服务。计算机系统310的部件可以由软件、硬件、固件或其组合来实现。例如，计算机系统310可以包括由计算机可执行指令实现的部件，该计算机可执行指令存储在一个或多个计算机可读存储介质上，并且被执行以根据所描述的主题的各方面来执行各种步骤、方法和/或功能性。在各种实现方式中，计算机系统310的服务器计算机311至313和数据存储装置314至316可以提供上面所描述的关于图1所示的计算机系统110的部件中的部分或全部部件。

[0102] 在利用用户相关数据的实现方式中，为了用户隐私和信息保护起见，这样的用户相关数据的提供者(例如，客户端设备301至305、应用等)和消费者(例如，计算机系统310、web服务、基于云的服务等)可以采用多种机构。这样的机构可以包括但不限于：需要授权以监控、收集或报告数据；使得用户能够选择参加或不参加数据监控、收集和报告；采用隐私规则来防止某些数据被监控、收集或报告；提供用于匿名化、截断或混淆被准许监控、收集或报告的敏感数据的功能性；采用数据保留政策来保护和清除数据；和/或用于保护用户隐私的其他合适机制。

[0103] 根据所描述的主题的各方面，客户端设备301至305和/或计算机系统310中的一个或多个可以执行涉及分析音频或视频数据并且传送标识当前说话者的元数据的各种操作。

[0104] 在各种实现方式中，观察者可以使用客户端设备301至305中的一个或多个客户端设备来渲染表示当前说话者的音频和/或视频内容。示例性类型的音频和/或视频内容可以包括但不限于：实时音频和/或视频内容、预先记录的音频和/或视频内容、音频和/或视频

流、电视内容、无线电内容、播客内容、web内容和/或表示当前说话者的其他音频和/或视频数据。

[0105] 客户端设备301至305可以呈现和/或计算机系统310可以提供用于允许观察者与所渲染的音频和/或视频内容交互的各种用户界面。用户界面可以由客户端设备301至305经由web浏览应用或提供用于渲染音频和/或视频内容的用户界面的其他合适类型的应用、应用程序和/或app来呈现。客户端设备301至305可以接收并且响应于诸如语音输入、触摸输入、手势输入、遥控输入、按钮输入等之类各种类型的用户输入。

[0106] 客户端设备301至305中的一个或多个客户端设备可以包括用于提供或支持说话者识别的硬件和软件。例如,在一个或多个客户端设备301至305上运行的应用可以被配置成将各种类型的输入解释为标识当前说话者的请求。这样的应用可以被实现为呈现内容并且提供说话者识别功能性的音频和/或视频应用,和/或通过与呈现内容的音频和/或视频应用一起运行的单独的说话者识别应用来实现。说话者识别功能性可以包括以下各项中的一项或多项:接收用于标识当前说话者的请求,捕获音频和/或视频内容语音,向计算机系统310提供音频和/或视频数据,提供有助于说话者识别的描述性元数据,传送标识当前说话者的元数据,和/或接收、存储和使用用于当前说话者的音频指纹。

[0107] 如图3所示,客户端设备301可以是正在被观察者用来呈现和查看视频内容的平板设备。当视频内容示出当前说话者时,观察者可以在不同时间与视频内容进行交互。例如,观察者可以点击触敏显示屏幕,点击正在显示当前说话者的用户界面的特定区域,点击图标或菜单项,按下按钮等等。客户端设备301和/或在客户端设备301上运行的应用可以接收和解释来自观察者的触摸输入,作为标识当前说话者的请求。可替代地或附加地,观察者可以说出可以被客户端设备301和/或在客户端设备301上运行的应用解释的语音命令(例如,“谁在说话”、“标识说话者”等),作为标识当前说话者的请求。

[0108] 客户端设备302和客户端303可以分别是观察者用来呈现和查看视频内容的电视和媒体设备(例如,媒体和/或游戏控制台、机顶盒等)。观察者可以通过使用手势、遥控器和/或其他合适的输入设备以移动光标、选择正在显示当前说话者的用户界面的特定区域、选择图标或菜单项、选择按钮等等来与示出当前说话者的视频内容进行交互。客户端设备302、客户端设备303和/或在客户端设备302和/或客户端设备303上运行的应用可以接收和解释来自观察者的输入,作为标识当前说话者的请求。可替代地或附加地,观察者可以说出可以由客户端设备302、客户端设备303和/或在客户端设备302和/或客户端设备303上运行的应用解释的语音命令,作为标识当前说话者的请求。

[0109] 客户端设备304可以被实现为由观察者用来收听音频内容的无线电。观察者可以使用各种类型的输入(例如,触摸输入、按钮输入、语音输入等)与客户端设备304和/或在客户端设备304上运行的应用来交互,以请求标识当前说话者。可以领会,客户端设备304可以由其他类型的音频设备来实现,诸如免提电话、便携式媒体播放器、用于具有视觉障碍的个人使用的音频设备和/或其他合适的音频设备。

[0110] 客户端设备305可以被实现为正在由观察者用来提供音频和/或视频内容的智能手机。在一些实现方式中,当音频和/或视频内容由客户端设备305渲染时,观察者可以使用各种类型的输入(例如,触摸输入、按钮输入、语音输入等)与客户端设备305和/或运行在客户端设备305上的应用进行交互,以请求标识当前说话者。可替代地,当音频和/或视频内容

由诸如客户端设备302或客户端设备304之类的不同的客户端设备渲染时，观察者可以与客户端设备305和/或在客户端设备305上运行的应用进行交互，以请求标识当前说话者。例如，可以在客户端设备305上启动说话者识别应用，以标识在由客户端设备305检测到的外部音频内容中呈现的当前说话者。

[0111] 响应于接收到标识当前说话者的请求，客户端设备301至305中的任一客户端设备和/或在客户端设备301中的任一客户端设备上运行的应用可以捕获并且向计算机系统310传送表示当前说话者的音频和/或视频数据的样本以供处理。在一些实现方式中，客户端设备301至305中的一个或多个客户端设备和/或在客户端设备301至305中的一个或多个客户端设备上运行的应用还可以检测、生成和/或传送促进说话者识别的描述性元数据，诸如视频内容的标题、广播的日期、当前说话者正在被示出的时间戳、当前说话者的图像或图像特征、和/或其他描述性信息。

[0112] 在接收到音频和/或视频数据的样本之后，计算机系统310可以生成用于当前说话者的音频指纹，并且将当前说话者的音频指纹与由计算机系统310维持的已存储音频指纹进行比较。如果描述性元数据被计算机系统310接收到，则其可以与已存储音频指纹相关联的元数据进行比较，以通过将已存储音频指纹标识为候选者和/或将已存储音频指纹与当前说话者的音频指纹进行匹配来促进说话者识别。

[0113] 如果计算机系统310基于已存储音频指纹来成功地识别当前说话者，则计算机系统310可以从与已存储音频指纹相关联的元数据中取回当前说话者的名称或身份。计算机系统310还可以从一个或多个信息源取回用于当前说话者的附加信息(例如，个人和/或职业信息、组织信息、目录信息、简档信息等)。计算机系统310可以传送标识当前说话者的元数据，其包括用于由一个或多个客户端设备301至305在呈现音频和/或视频内容的同时显示的附加信息。在一些实现方式中，计算机系统310可以提供(例如，广播、流式传输等)音频和/或视频内容以供客户端设备301至105中的一个或多个客户端设备显示，并且可以使用元数据增强音频和/或视频数据，并且向客户端设备301至105中的一个或多个客户端设备实时传送增强音频和/或视频数据作为同步流。

[0114] 在从计算机系统310接收到元数据和/或增强音频和/或视频数据时，客户端设备301至105中的一个或多个客户端设备可以渲染音频和/或视频内容，同时显示标识当前说话者和/或包括当前说话者的附加信息的元数据。在一些实施方式中，计算机系统310可以向客户端设备301至105中的一个或多个客户端设备传送当前说话者的音频指纹，以供由这些客户端设备存储和使用，以标识后续音频和/或视频内容中的当前说话者。从计算机系统310接收的当前说话者的音频指纹可以由客户端设备301至305之中的观察者共享和/或与其他用户共享。

[0115] 标识当前说话者的元数据和/或当前说话者的附加信息可以被客户端设备301至305中的一个或多个客户端设备显示为渲染音频和/或视频内容的用户界面内的指示符，显示为单独的用户界面，和/或以各种其他方式显示。指示符可以呈现按钮或其他显示元件以确认当前说话者的身份，每当检测到当前的说话者时请求警报，呈现与当前说话者有关的进一步内容和/或执行附加动作。

[0116] 在计算机系统310不维持所存储的当前说话者的音频指纹和/或说话者识别不成功的情形下，计算机系统310可以传送候选者或可能说话者的列表，并且可以请求观察者以

通过选择可能的说话者和/或输入未被识别的说话者的名称或其他合适的身份来提供标记信息。例如，计算机系统310可以基于从客户端设备301至305中的一个或多个客户端设备和/或在客户端设备301至305中的一个或多个客户端设备上运行的应用接收到的描述性元数据来生成并且传送可能说话者的列表。可替代地或附加地，计算机系统310可以向一个或多个用户(例如，用户的群体、用户的子集等)传送用于标记信息的请求，这些用户在一些情况下可能正在同时渲染音频和/或视频内容。如果提供了用于当前说话者的冲突的标记信息(例如，不同的名称)，则计算机系统310可以基于由大多数用户提供的身份或其他合适的用于解决冲突的启发式方法来标识当前说话者。

[0117] 在接收到标识当前说话者的标记信息之后，计算机系统310可以将标识当前说话者的元数据与所生成的当前说话者的音频指纹相关联。计算机系统310可以存储所生成的用于当前说话者的音频指纹以及相关联的元数据，以供用于后续说话者识别和/或可以向一个或多个客户端设备301至305提供所生成的音频指纹和相关联的元数据，以供这些客户端设备存储和使用以标识后续音频和/或视频内容中的当前说话者。

[0118] 示例性过程

[0119] 继续参考前述附图，下文描述示例性过程以进一步说明所描述的主题的各方面。应当理解，以下示例性过程不旨在将所描述的主题限制于特定实现方式。

[0120] 图4图示了作为根据所描述的主题的各方面的示例性过程的实施例的计算机实现方法400。在各种实施例中，计算机实现方法400可以通过计算机系统110、计算机系统310和/或包括一个或多个计算设备在内的其他合适的计算机系统来执行。应当领会，计算机实现的方法400或其部分可以由各种计算设备、计算机系统、部件和/或存储在一个或多个计算机可读存储介质上的计算机可执行指令执行。

[0121] 在410处，计算机系统可以基于表示当前说话者的语音的音频数据来生成当前说话者的音频指纹。例如，计算机系统110和/或计算机系统310可以接收各种类型的音频和/或视频数据(例如，实时音频和/或视频数据、预先记录的音频和/或视频数据、音频和/或视频流、电视内容、无线电内容、播客内容、web内容和/或表示当前说话者的其他类型的音频和/或视频数据)。计算机系统110和/或计算机系统310可以基于该音频和/或视频数据的特征来生成音频指纹(例如，声纹、语音生物特征、说话者模板、特征向量、特征向量的集合或序列、说话者模型、说话者码本和/或其他合适的说话者指纹)。

[0122] 在420处，计算机系统可以执行自动说话者识别。例如，计算机系统110和/或计算机系统310可以将所生成的当前说话者的音频指纹与包含在一个或多个存储位置中的存储的音频指纹进行比较。可以比较各种语音特征以匹配音频指纹。如果基于已存储音频指纹成功地识别了当前说话者，则计算机系统110和/或计算机系统310可以取回与已存储音频指纹相关联的个人的名称或其他合适的身份，并且传送标识当前说话者的元数据。

[0123] 在430处，计算机系统可以向观察者的客户端设备传送指示当前说话者未被识别的数据。例如，当自动说话者识别不成功时，计算机系统110和/或计算机系统310可以传送数据以呈现指示说话者未被识别的消息和/或图形。计算机系统110和/或计算机系统310还可以请求远程参与者或观察者提供标记信息，诸如未被识别的说话者的名称或其他合适的身份。在一些实现方式中，可以向远程参与者或观察者呈现可能的说话者的列表以供选择。

[0124] 在440处，计算机系统可以从观察者的客户端装置接收标识当前说话者的标记信

息。例如，计算机系统110和/或计算机系统310可以响应于一个或多个远程参与者或观察者手动输入当前说话者的名称或身份和/或选择用于当前演讲者的名称、用户简档和/或化身来接收标记信息。如果接收到用于当前说话者的冲突的标记信息(例如，不同的名称)，则计算机系统110和/或计算机系统310可以基于由大多数远程参与者或观察者提供的身份来识别当前的说话者。

[0125] 在450处，计算机系统可以存储当前说话者的音频指纹和标识当前说话者的元数据。例如，响应于接收标记信息，计算机系统110和/或计算机系统310可以将所生成的当前说话者的音频指纹和标识当前说话者的元数据存储在一个或多个存储位置中。在一些实现方式中，可以在接收到从一个或多个远程参与者或观察者正确地标识当前说话者的确认之后，存储当前说话者的音频指纹和标识当前说话者的元数据。当接收到表示当前说话者的语音的后续音频数据时，计算机系统110和/或计算机系统310可以使用所存储的当前说话者的音频指纹来执行说话者识别。

[0126] 在460处，计算机系统可以向观察者的客户端设备传送标识当前说话者的元数据。例如，计算机系统110和/或计算机系统310可以传送仅标识当前说话者的元数据以及各种类型的附加信息(例如，个人和/或职业信息、组织信息、目录信息、简档信息等)，和/或作为增强音频数据的一部分(例如，同步的音频数据和元数据流)。标识当前说话者和/或包括附加信息的元数据可以由远程参与者或观察者的客户端设备渲染，使用标识当前说话者的指示符和/或用户界面。

[0127] 在460之后，计算机系统可以重复一个或多个操作。例如，计算机系统110和/或计算机系统310可以基于表示当前说话者的语音的后续音频数据来生成当前说话者的新音频指纹，并且执行自动说话者识别。计算机系统110和/或计算机系统310可以在将当前说话者的新音频指纹与所存储的当前说话者的音频指纹进行比较时，可以成功地识别当前说话者，并且可以向观察者的客户端设备传送标识当前说话者的元数据。

[0128] 如上文所描述的，所描述的主题的各方面可以提供各种伴随的和/或技术的优点。作为说明而非限制，当经由音频和/或视频会议和/或以其他方式渲染音频和/或视频内容进行实时通信时，执行自动说话者识别和传送标识当前说话者的元数据允许多方通信会话(例如，音频、视频和/或web会议、在线会议等)的远程参与者和/或观察者在任何给定时刻标识当前说话者。

[0129] 当多个说话者位于同一地点并且共享单个客户端设备并且对于不分开加入在线会议的位于同一地点的参与者不提供基于被提供以加入在线会议的信息来显示的名称和/或化身时，执行自动说话者识别并且传送标识当前说话者的元数据允许多方通信会话的参与者标识当前说话者。

[0130] 执行自动说话者识别并且传送标识当前说话者的元数据允许多方通信会话的参与者标识使用电话(例如，座机电话、会议免提电话、移动电话等)拨入多方通信会话和/或作为附加设备加入多方通信会话的当前说话者。在这样的上下文中，执行自动说话者识别还允许呈现多方通信会话的用户界面指示单个参与者正在使用多个客户端设备进行通信。

[0131] 传送元数据允许远程参与者或观察者标识当前说话者，并且从数据源实时接收附加信息(例如，个人和/或职业信息、组织信息、目录信息、简档信息等)以在进行在线对话时提供更多价值和/或以其他方式增强用户体验。

[0132] 使用元数据增强音频流允许当特定所识别的说话者正在说话时生成警报，使得能够从实时或存储的音频流中搜索所识别的说话者的音频内容和/或促进创建和搜索具有多个说话者的对话的文本转录。

[0133] 即使在没有当前说话者的存储的音频指纹的情况下，向观察者的客户端设备传送指示当前说话者未被识别的数据并且从观察者的客户端设备接收标识当前说话者的标记信息也允许标识当前说话者。

[0134] 向观察者的客户端设备传送指示当前说话者未被识别的数据，从观察者的客户端设备接收标识当前说话者的标记信息，以及存储当前说话者的音频指纹和标识当前说话者的元数据允许指纹存储库经由众包过程来填充。这种众包过程允许以与说话者无关和/或语音无关的方式执行自动说话者识别，并且消除或减少对计算机系统执行训练说话者模型的计算方面昂贵的过程的需要。

[0135] 示例性操作环境

[0136] 所描述的主题的各方面可以针对和/或通过各种操作环境、计算机网络、平台、框架、计算机体系结构和/或计算设备来实现。所描述的主题的各方面可以由计算机可执行指令来实现，这些计算机可执行指令可以由一个或多个计算设备、计算机系统和/或处理器执行。

[0137] 在其最基本的配置中，计算设备和/或计算机系统可以包括至少一个处理单元(例如，单处理器单元、多处理器单元、单核单元和/或多核单元)和存储器。根据计算机系统或计算设备的确切配置和类型，由计算设备和/或计算机系统实现的存储器可以是易失性的(例如，随机存取存储器(RAM))、非易失性存储器(例如，只读存储器(ROM)、闪存存储器等)或其组合。

[0138] 计算设备和/或计算机系统可以具有附加特征和/或功能性。例如，计算设备和/或计算机系统可以包括诸如附加存储(例如，可移除和/或不可移除)之类的硬件，包括但不限于：固态、磁盘、光盘或磁带。

[0139] 计算设备和/或计算机系统通常可以包括或可以访问多种计算机可读介质。例如，计算机可读介质可以包含计算机可执行指令以供计算设备和/或计算机系统执行。计算机可读介质可以是任何可用介质，其可以由计算设备和/或计算机系统访问，并且包括易失性和非易失性介质，以及可移除和不可移除介质。如本文中所使用的，术语“计算机可读介质”包括计算机可读存储介质和通信介质。

[0140] 本文中所使用的术语“计算机可读存储介质”包括用于存储诸如计算机可执行指令、数据结构、程序模块或其他数据之类的信息的易失性和非易失性、可移除和不可移除介质。计算机可读存储介质的示例包括但不限于：存储器存储设备，诸如RAM、ROM、电可擦除程序只读存储器(EEPROM)、半导体存储器、动态存储器(例如，动态随机存取存储器(DRAM)、同步动态随机存取存储器(SDRAM)、双倍数据速率同步动态随机存取存储器(DDR SDRAM)等)、集成电路、固态驱动器、闪存(例如，基于NAN的闪存)、存储器芯片、存储器卡、记忆棒、拇指驱动器等；光存储介质，诸如蓝光光盘、数字视频光盘(DVD)、光盘(CD)、CD-ROM、光盘盒等；磁存储介质，其包括硬盘驱动器、软盘、软磁盘、磁带盒、磁带等；以及其他类型的计算机可读存储设备。可以领会，各种类型的计算机可读存储介质(例如，存储器和附加硬件存储装置)可以是计算设备和/或计算机系统的一部分。如本文中所使用的，术语“计算机可读存储

介质”和“计算机可读存储介质”不意味且明确地排除传播的信号、经调制的数据信号、载波或任何其他类型的暂态计算机可读介质。

[0141] 通信介质通常在诸如载波或其他传送机制的经调制的数据信号中包含计算机可执行指令、数据结构、程序模块或其他数据，并且包括任何信息递送介质。术语“经调制的数据信号”是指具有以对信号中的信息进行编码的方式设置或改变其特性中的一个或多个特性的信号。作为示例而非限制，通信介质包括诸如有线网络或直接有线连接的有线介质；以及诸如声学、射频、红外和其他无线介质的无线介质。

[0142] 在各种实施例中，所描述的主题的各方面可以通过存储在一个或多个计算机可读存储介质上的计算机可执行指令来实现。可以使用各种类型的任何合适的编程和/或标记语言来实现计算机可执行指令，诸如：可扩展应用标记语言(XAML)、XML、XBL HTML、XHTML、XSLT、XMLHttpRequestObject、CSS、文档对象模型(DOM)、Java®、JavaScript、JavaScript对象符号(JSON)、Jscript、ECMAScript、Ajax、Flash®、Silverlight™、VisualBasic®(VB)、VBScript、PHP、ASP、Shockwave®、Python、Perl®、C、Objective-C、C++、C#/.net和/或其他。

[0143] 计算设备和/或计算机系统可以包括各种输入设备、输出设备、通信接口和/或其他类型的设备。示例性输入设备包括但不限于：用户界面、键盘/小键盘、触摸屏、触摸板、笔、鼠标、轨迹球、遥控器、游戏控制器、相机、条形码读取器、麦克风或其他话音输入设备、视频输入设备、激光测距仪、运动感测设备、手势检测设备和/或其他类型的输入机构和/或设备。计算设备可以提供自然用户接口(NUI)，其使得用户能够以“自然”的方式与计算设备进行交互，而不受诸如鼠标、键盘、遥控器等之类的输入设备的人为约束。NUI技术的示例包括但不限于：话音和/或语音识别，触摸和/或触控笔识别，使用加速度计、陀螺仪和/或深度相机(例如，立体视觉或飞行时间相机系统、红外相机系统、RGB相机系统和/或其组合)在屏幕上和屏幕附近的运动和/或手势识别，头部和眼睛跟踪，注视跟踪，面部识别，3D显示器，沉浸式增强现实和虚拟现实系统，用于使用电场感测电极(EEG和相关方法)感测脑活动的技术，意图和/或目标理解以及机器智能。

[0144] 计算设备可以被配置成根据实现方式来以各种方式接收和响应输入。响应可以以各种形式呈现，其包括例如呈现用户界面；输出诸如图像之类的对象、视频、多媒体对象、文档和/或其他类型的对象；输出文本响应；提供与响应内容相关联的链接；输出计算机生成的话音响应或其他音频；或响应的其他类型的视觉和/或音频呈现。示例性输出设备包括但不限于显示器、投影仪、扬声器、打印机和/或其他类型的输出机构和/或设备。

[0145] 计算设备和/或计算机系统可以包括一个或多个通信接口，其允许在其他计算设备和/或计算机系统之间和之中进行通信。通信接口可以在各种计算设备和/或计算机系统之间和之中的网络通信的上下文中使用。通信接口可以允许计算设备和/或计算机系统与其他设备、其他计算机系统、web服务(例如，附属web服务、第三方web服务、远程web服务等)、web服务应用和/或信息源(例如，附属信息源、第三方信息源、远程信息源等)通信。如此，通信接口可以在访问各种类型的资源、从各种类型的资源获取数据和/或与各种类型的资源合作的上下文中使用。

[0146] 通信接口还可以在通过网络或网络组合分发计算机可执行指令的上下文中使用。

例如，可以利用远程计算机和存储设备来组合或分发计算机可执行指令。本地或终端计算机可以访问远程计算机或远程存储设备并且下载计算机程序或计算机程序的一个或多个部分以供执行。还可以领会，可以通过在本地终端执行一些指令并且在远程计算机上执行一些指令来分发计算机可执行指令的执行。

[0147] 计算设备可以由诸如以下各项的移动计算设备来实现：移动电话(例如，蜂窝电话，智能手机，诸如Microsoft Windows Phone、Apple iPhone、BlackBerry®电话、实现Google® Android ™操作系统的电话、实现Linux®操作系统的电话或实现移动操作系统的其他类型的电话)、平板电脑(例如，Microsoft® Surface®设备、Apple iPad™、三星Galaxy Note® Pro或其他类型的平板电脑)、膝上型电脑、笔记本电脑、上网本电脑、个人数字助理(PDA)、便携式媒体播放器、手持式游戏控制台、可穿戴式计算设备(例如，智能手表、包括诸如Google® Glass™之类的智能眼镜的头戴式设备、可穿戴式显示器等)、个人导航设备、车辆计算机(例如，车载导航系统)、相机或其他类型的移动设备。

[0148] 计算设备可以由固定计算设备实现，诸如：台式计算机、个人计算机、服务器计算机、娱乐系统设备、媒体播放器、媒体系统或控制台、视频游戏系统或控制台、多用途系统或控制台(例如，组合的多媒体和视频游戏系统或控制台，诸如Microsoft® Xbox®系统或控制台、Sony® PlayStation®系统或控制台、Nintendo®系统或控制台或其他类型多用途游戏系统或控制台)、机顶盒、电器(例如，电视机、冰箱、烹饪器具等)或其他类型的固定计算设备。

[0149] 计算设备还可以由其他类型的基于处理器的计算设备来实现，其包括数字信号处理器、现场可编程门阵列(FPGA)、程序专用集成电路和应用专用集成电路(PASIC/ASIC)、程序专用标准产品和/应用专用标准产品(PSSP/ASSP)、片上系统(SoC)、复杂可编程逻辑器件(CPLD)等。

[0150] 计算设备可以包括和/或运行例如由计算设备的软件、固件、硬件、逻辑和/或电路实现的一个或多个计算机程序。计算机程序可以以各种方式分发到和/或安装在计算设备上。例如，计算机程序可以由原始设备制造商(OEM)预先安装在计算设备上、作为另一计算机程序的安装的一部分安装在计算设备上、从应用商店下载并且安装在计算设备上、使用企业网络管理工具由系统管理员分发和/或安装，并且根据实现方式以各种其他方式分发和/或安装。

[0151] 由计算设备实现的计算机程序可以包括一个或多个操作系统。示例性操作系统包括但不限于：Microsoft®操作系统(例如，Microsoft® Windows®操作系统)、Google®操作系统(例如，Google® Chrome OS™操作系统或Google® Android™操作系统)、苹果操作系统(例如，MacOS®或Apple iOS™操作系统)、开源操作系统或适合在基于移动、固定和/或基于处理器的计算设备上运行的任何其他操作系统。

[0152] 由计算设备实现的计算机程序可以包括一个或多个客户端应用。示例性客户端应

用包括但不限于:web浏览应用、通信应用(例如,电话应用、电子邮件应用、文本消息传送应用、即时消息传送应用、Web会议应用等)、媒体应用(例如,视频应用、电影服务应用、电视服务应用、音乐服务应用、电子书应用、照片应用等)、日历应用、文件共享应用、个人助理或其他类型的对话应用、游戏应用、图形应用、购物应用、支付应用、社交媒体应用、社交联网应用、新闻应用、运动应用、天气应用、地图应用、导航应用、旅行应用、餐馆应用、娱乐应用、医疗保健应用、生活方式应用、参考应用、财务应用、业务应用、教育应用、生产力应用(例如,文字处理应用、电子表格应用、幻灯片放映演示应用、记笔记应用等)、安全应用、工具应用、实用程序应用和/或适用于在移动、固定和/或基于处理器的计算设备上运行的任何其他类型的应用、应用程序和/或app。

[0153] 由计算设备实现的计算机程序可以包括一个或多个服务器应用。示例性服务器应用包括但不限于:与上文所描述的各种类型的示例性客户端应用中的任一种相关联的一个或多个服务器托管的、基于云的、和/或在线的应用;上文所描述的各种类型的示例性客户端应用中的任一种的一个或多个服务器托管的、基于云的和/或在线的版本;被配置成提供web服务、网站、网页、web内容等的一个或多个应用;被配置成提供和/或访问信息源、数据存储装置、数据库、存储库等的一个或多个应用;和/或适用于在服务器计算机上运行的其他类型的应用、应用程序和/或app。

[0154] 计算机系统可以由诸如服务器计算机之类的计算设备或被配置成实现其中一个或多个适当配置的计算设备可以执行一个或多个处理步骤的服务的多个计算设备来实现。计算机系统可以被实现为分布式计算系统,其中部件位于通过网络(例如,有线和/或无线)和/或其他形式的直接和/或间接连接彼此连接的不同计算设备上。还可以经由基于云的架构(例如,公共、私有或其组合)来实现计算机系统,其中通过共享数据中心来传递服务。计算机系统的一些部件可以被设置在云内,而其他部件被设置在云之外。

[0155] 图5图示了作为可以实现所描述的主题的各方面的示例性操作环境的实施例的操作环境500。应当领会,在各种实施例中,操作环境500可以由客户端-服务器模型和/或体系架构以及其他操作环境模型和/或体系架构来实现。

[0156] 操作环境500可以包括计算设备510,其可以实现所描述的主题的各方面。计算设备510可以包括处理器511和存储器512。计算设备510还可以包括附加的硬件存储装置513。应当理解,计算机可读存储介质包括存储器512和硬件存储装置513。

[0157] 计算设备510可以包括输入设备514和输出设备515。输入设备314可以包括上文所描述的示例性输入设备中的一个或多个输入设备和/或其他类型的输入机构和/或设备。输出设备515可以包括上文所描述的示例性输出设备中的一个或多个输出设备和/或其他类型的输出机构和/或设备。

[0158] 计算设备510可以包含一个或多个通信接口516,其允许计算设备510与其他计算设备和/或计算机系统通信。通信接口516还可以在分布式计算机可执行指令的上下文中使用。

[0159] 计算设备510可以包括和/或运行例如由计算设备510的软件、固件、硬件、逻辑和/或电路实现的一个或多个计算机程序517。计算机程序517可以包括操作系统518,其例如由上文所描述的一个或多个示例性操作系统和/或适于在计算设备510上运行的其他类型的操作系统实现。计算机程序517可以包括一个或多个应用519,其例如由上文所描述的一个

或多个示例性应用和/或适用于在计算设备510上运行的其他类型的应用实现。

[0160] 计算机程序517可以经由一个或多个合适的接口(例如,API或其他数据连接)被配置成与一个或多个资源进行通信和/或协作。资源的示例包括计算设备510的本地计算资源和/或诸如服务器托管的资源、基于云的资源、在线资源、远程数据存储装置、远程数据库、远程存储库、web服务、网站、网页、web内容和/或其他类型的远程资源之类的远程计算资源。

[0161] 计算机程序517可以实现例如存储在诸如存储器512或硬件存储装置513的计算机可读存储介质中的计算机可执行指令。由计算机程序517实现的计算机可执行指令可以被配置成与操作系统518和应用程序519中的一个或多个一起工作,支持和/或增强它们。由计算机程序517实现的计算机可执行指令还可以被配置成提供一个或多个单独的和/或独立的服务。

[0162] 计算设备510和/或计算机程序517可以实现和/或执行所描述的主题的各个方面。如所示出的,计算设备510和/或计算机程序517可以包括说话者识别代码520。在各种实施例中,说话者识别代码520可以包括计算机可执行指令,其被存储在计算机可读存储介质上并且被配置成实现所描述的主题的一个或多个方面。作为示例而非限制,说话者识别代码520可以由计算设备510来实现,该计算设备510又可以表示客户端设备101至104和/或客户端设备301至305中的一个或多个客户端设备。作为进一步的示例而不是限制,说话者识别代码520可以被配置成呈现用户界面200。

[0163] 操作环境500可以包括计算机系统530,其可以实现所描述的主题的各方面。计算机系统530可以由一个或多个计算设备(诸如一个或多个服务器计算机)来实现。计算机系统530可以包括处理器531和存储器532。计算机系统530还可以包括附加的硬件存储装置533。应当理解,计算机可读存储介质包括存储器532和硬件存储装置533。计算机系统530可以包括输入设备534和输出设备535。输入设备534可以包括上文所描述的示例性输入设备中的一个或多个示例性输入设备和/或其他类型的输入机构和/或设备。输出设备535可以包括上文所描述的示例性输出设备中的一个或多个示例性输出设备和/或其他类型的输出机构和/或设备。

[0164] 计算机系统530可以包含一个或多个通信接口536,其允许计算机系统530与各种计算设备(例如,计算设备510)和/或其他计算机系统通信。通信接口536还可以在分布式计算机可执行指令的上下文中使用。

[0165] 计算机系统530可以包括和/或运行例如由计算机系统530的软件、固件、硬件、逻辑和/或电路实现的一个或多个计算机程序537。计算机程序537可以包括操作系统538,其例如通过上文所描述的一个或多个示例性操作系统和/或适于在计算机系统530上运行的其他类型的操作系统来实现。计算机程序537可以包括一个或多个应用539,其例如由上文所描述的一个或多个示例性实施例和/或适用于在计算机系统530上运行的其他类型的应用来实现。

[0166] 计算机程序537可以经由一个或多个合适的接口(例如,API或其他数据连接)被配置成与一个或多个资源通信和/或协作。资源的示例包括计算机系统530的本地计算资源和/或诸如服务器托管的资源、基于云的资源、在线资源、远程数据存储装置、远程数据库、远程存储库、web服务、网站、网页、web内容和/或其他类型的远程资源之类的远程计算资

源。

[0167] 计算机程序537可以实现计算机可执行指令，其被存储在诸如存储器532或硬件存储装置533之类的计算机可读存储介质中。由计算机程序537实现的计算机可执行指令可以被配置成与操作系统538和应用539中的一个或多个一起工作，支持和/或增强它们。由计算机程序537实现的计算机可执行指令还可以被配置成提供一个或多个单独的和/或独立的服务。

[0168] 计算系统530和/或计算机程序537可以实现和/或执行所描述的主题的各个方面。如所示出的，计算机系统530和/或计算机程序537可以包括说话者识别代码540。在各种实施例中，说话者识别代码540可以包括计算机可执行指令，其被存储在计算机可读存储介质上并且被配置成实现所描述的主题的一个或多个方面。作为示例而非限制，说话者识别代码540可以由计算机系统530来实现，该计算机系统530又可以实现计算机系统110和/或计算机系统310。作为进一步的示例而非限制，说话者识别代码540可以实现计算机实现的方法400的一个或多个方面。

[0169] 计算设备510和计算机系统530可以通过网络550进行通信，其可以由适合于在计算设备510和计算机系统530之间提供通信的任何类型的网络或网络组合来实现。网络550可以包括例如和不限于：诸如因特网之类的WAN、LAN、电话网络、专用网络、公共网络、分组网络、电路交换网络、有线网络和/或无线网络。计算设备510和计算机系统530可以使用各种通信协议和/或数据类型通过网络550进行通信。计算设备510的一个或多个通信接口516和计算机系统530的一个或多个通信接口536可以在通过网络550进行通信的上下文中使用。

[0170] 计算设备510和/或计算机系统530可以通过网络550与存储系统560通信。可替代地或附加地，存储系统560可以与计算设备510和/或计算机系统530集成。存储系统560可以代表根据所描述的主题的各种类型的存储装置。例如，存储系统560可以根据所描述的主题实现以下各项中的一项或多项：说话者指纹存储库112、目录113、参与者数据存储装置114、充实音频数据存储库116、转录存储库118和/或其他数据存储设施。存储系统560可以使用数据库存储装置、云存储装置、表存储装置、blob存储装置、文件存储装置、队列存储装置和/或其他合适类型的存储机构来为关系型(例如，SQL)和/或非关系型(例如，NO-SQL)数据提供任何合适类型的数据存储装置。存储系统560可以由一个或多个计算设备(诸如数据中心中的计算机集群)来实现、由虚拟机来实现和/或作为基于云的存储服务而被提供。

[0171] 图6图示了作为可以实现所描述的主题的各方面的示例性计算机系统的实施例的计算机系统600。在各种实现方式中，计算机系统600的部署和/或其多个部署可以提供用于在一个物理主机服务器计算机上同时地运行多个虚拟服务器实例的服务器虚拟化和/或用于在同一物理网络上同时地运行多个虚拟网络基础设施的网络虚拟化。

[0172] 计算机系统600可以由各种计算设备来实现，诸如提供可以包括处理器611、存储器612和通信接口613的硬件层610的一个或多个物理服务器计算机。计算机系统600可以实现管理程序620，其被配置成管理、控制和/或仲裁对硬件层610的访问。在各种实现方式中，管理程序620可以管理硬件资源以提供隔离的执行环境或分区，诸如父(根)分区和一个或多个子分区。父分区可以操作来创建一个或多个子分区。每个分区可以被实现为抽象容器或逻辑单元，用于隔离由管理程序620管理的处理器和存储器资源，并且可以分配一组硬件

资源和虚拟资源。逻辑系统可以映射到分区，并且逻辑设备可以映射到分区内的虚拟设备。

[0173] 父分区和子分区可以实现虚拟机，诸如例如，虚拟机630，640和650。每个虚拟机可以以软件实现方式来仿真物理计算设备或计算机系统，像物理机器那样执行程序。每个虚拟机可以具有一个或多个虚拟处理器，并且可以提供用于执行操作系统的虚拟系统平台(例如，Microsoft操作系统、Google操作系统、来自Apple®的操作系统、Linux操作系统、开源操作系统等)。如所示出的，父分区中的虚拟机630可以运行管理操作系统631，并且子分区中的虚拟机640，650可以托管客户操作系统641，651，每个客户操作系统641，651被实现为例如全功能操作系统或专用核。客户操作系统641，651中的每个客户操作系统可以调度线程以在一个或多个虚拟处理器上执行并且分别实现应用642，652的实例。

[0174] 父分区中的虚拟机630可以经由设备驱动器632和/或其他合适的接口访问硬件层610。然而，子分区中的虚拟机640，650通常不能访问硬件层610。相反，虚拟机640，650被呈现硬件资源的虚拟视图，并且由通过父分区中的虚拟机630提供的虚拟化服务支持。父分区中的虚拟机630可以托管提供虚拟化管理功能性的虚拟化堆栈633，该虚拟化管理功能性包括经由设备驱动器632访问硬件层610。虚拟化堆栈633可以实现和/或操作为虚拟化服务提供商(VSP)，以处理来自虚拟化服务客户端(VSC)的请求并且向其提供各种虚拟化服务，该VSC由在子分区中操作的虚拟机640，650中的一个或多个虚拟化堆栈643，653实现。

[0175] 计算机系统600可以实现和/或执行所描述的主题的各个方面。作为示例而非限制，一个或多个虚拟机640，650可以实现具有说话者识别功能性的web服务和/或基于云的服务。作为进一步的示例，但不限于，一个或多个虚拟机640，650可以实现计算机系统110、计算机系统310和/或计算机实现方法400的一个或多个方面。另外，硬件层610可以由计算机系统110、计算机系统310和/或计算机系统530的一个或多个计算设备来实现。

[0176] 图7图示了作为可以实现所描述的主题的各方面的示例性移动计算设备的实施例的移动计算设备700。在各种实现方式中，移动计算设备700可以是以下各项中的一项或多项的示例：客户端设备101至104、客户端设备301至305和/或计算设备510。

[0177] 如所示出的，移动计算设备700包括可以彼此通信的多种硬件和软件部件。移动计算设备700可以表示本文中所描述的各种类型的移动计算设备中的任一种，并且可以允许通过诸如一个或多个移动通信网络(例如，蜂窝和/或卫星网络)、LAN和/或WAN之类的网络进行无线双向通信。

[0178] 移动计算设备700可以包括操作系统702和各种类型的移动应用704。在一些实现方式中，移动应用704可以包括一个或多个客户端应用和/或说话者识别代码520的部件。

[0179] 移动计算设备700可以包括用于执行诸如以下各项的任务的处理器706(例如，信号处理器、微处理器、ASIC或其他控制和处理逻辑电路)：信号编码、数据处理、输入/输出处理、功率控制和/或其他功能。

[0180] 移动计算设备700可以包括存储器708，其被实现为不可移除存储器710和/或可移除存储器712。不可移除存储器710可以包括RAM、ROM、闪存、硬盘或其他存储器设备。可移除存储器712可以包括闪存、用户身份模块(SIM)卡、“智能卡”和/或其他存储器设备。

[0181] 存储器708可以用于存储用于运行操作系统702和/或移动应用704的数据和/或代码。示例数据可以包括网页、文本、图像、声音文件、视频数据或其他数据，以经由一个或多个有线和/或无线网络被发送到一个或多个网络服务器或其他设备和/或从其接收。存储器

708可以用于存储诸如国际移动用户标识(IMSI)之类的用户标识符以及诸如国际移动设备标识符(IMEI)之类的设备标识符。这样的标识符可以被传送到网络服务器以标识用户和设备。

[0182] 移动计算设备700可以包括和/或支持一个或多个输入设备714，诸如触摸屏715、麦克风716、相机717、键盘718、轨迹球719和其他类型的输入设备(例如，自然用户界面(NUI)设备等)。触摸屏715可以例如使用电容式触摸屏和/或光学传感器来实现以检测触摸输入。移动计算设备700可以包括和/或支持一个或多个输出设备720，诸如扬声器721、显示器722和/或其他类型的输出设备(例如，压电或其他触觉输出设备)。在一些实现方式中，触摸屏715和显示器722可以组合在单个输入/输出设备中。

[0183] 移动计算设备700可以包括无线调制解调器724，其可以耦合至天线(未示出)并且可以支持处理器706和外部设备之间的双向通信。无线调制解调器724可以包括用于与移动通信网络和/或其他基于无线电的调制解调器(例如，Wi-Fi 726和/或蓝牙727)进行通信的蜂窝调制解调器725。通常，无线调制解调器724中的至少一个无线调制解调器被配置成：与一个或多个蜂窝网络(诸如GSM网络)进行通信以在单个蜂窝网络内进行数据和语音通信；蜂窝网络之间的通信；或移动计算设备700与公共交换电话网(PSTN)之间的通信。

[0184] 移动计算设备700还可以包括至少一个输入/输出端口728、电源730、加速度计732、物理连接器734(例如，USB端口、IEEE 1394(FireWire)端口、RS-232端口等)和/或全球定位系统(GPS)接收器736或其他类型的卫星导航系统接收器。可以领会，移动计算设备700的所图示的部件不是必需的或并不包括全部的，因为可以省略各种部件并且可以在各种实施例中包括其他部件。

[0185] 在各种实现方式中，移动计算设备700的部件可以被配置成执行结合客户端设备101至104和/或客户端设备301至304中的一个或多个客户端设备描述的各种操作。用于执行这种操作的计算机可执行指令可以存储在计算机可读存储介质(诸如例如，存储器708)中，并且可以由处理器706执行。

[0186] 图8图示了作为可以实现所描述的主题的各方面的示例性计算环境的实施例的计算环境800。如所示出的，计算环境800包括具有计算机810形式的通用计算设备。在各种实现方式中，计算机810可以是以下各项中的一项或多项的示例：客户端设备101至104、计算机系统110的计算设备、客户端设备301至305、计算机系统310的计算设备、计算设备510、计算机系统530的计算设备、计算机系统600的计算设备和/或移动计算设备700。

[0187] 计算机810可以包括各种部件，其包括但不限于：处理单元820(例如，一个或多个处理器或处理单元的类型)、系统存储器830和系统总线821，其将包括系统存储器830的各种系统部件耦合至处理单元820。

[0188] 系统总线821可以是多种类型的总线结构中的任一种总线结构，其包括使用多种总线体系架构中的任一种的存储器总线或存储器控制器、外围总线以及本地总线。作为示例而非限制，这种体系架构包括工业标准体系架构(ISA)总线、微通道体系架构(MCA)总线、增强型ISA(EISA)总线、视频电子标准协会(VESA)本地总线和还被称为Mezzanine总线的外围部件互连(PCI)总线。

[0189] 系统存储器830包括诸如ROM 831和RAM 832之类的具有易失性和/或非易失性存储器形式的计算机存储介质。基本输入/输出系统(BIOS)833包含有助于诸如在启动期间在

计算机810内部的元件之间传送信息的基本例程，通常存储在ROM 831中。RAM 832通常包含数据和/或程序模块，这些数据和/或程序模块可由处理单元820直接访问和/或正在其上操作。作为示例而非限制，示出了操作系统834、应用程序835、其他程序模块836和程序数据837。

[0190] 计算机810还可以包括其他可移除/不可移除和/或易失性/非易失性计算机存储介质。仅作为示例，图8图示了从不可移除非易失性磁性介质读取或写入的硬盘驱动器841，从可移除非易失性磁盘852读取或写入的磁盘驱动器851以及从可移除非易失性光盘856读取或写入的光盘驱动器855(诸如CD ROM或其他光学介质)。可以在示例性操作环境中使用的其他可移除/不可移除、易失性/非易失性计算机存储介质包括但不限于磁带盒、闪存卡、数字通用盘、数字录像带、固态RAM、固态ROM等。硬盘驱动器841通常通过诸如接口840之类的不可移除存储器接口连接至系统总线821，并且磁盘驱动器851和光盘驱动器855通常通过诸如接口850之类的可移除存储器接口连接至系统总线821。

[0191] 可替代地或另外，本文中所描述的功能性可以至少部分地由一个或多个硬件逻辑部件来执行。例如但不限于，例如可以使用的说明性类型的硬件逻辑部件包括FPGA、PASIC/ASIC、PSSP/ASSP、SoC和CPLD。

[0192] 上文所讨论的并且在图8中说明的驱动器及其相关联的计算机存储介质提供存储计算机可读指令、数据结构、程序模块和计算机810的其他数据。例如，硬盘驱动器841被图示为存储操作系统844、应用程序845、其他程序模块846和程序数据847。注意，这些部件可以与操作系统834、应用程序835、其他程序模块836和程序数据837相同或不同。操作系统844、应用程序845、其他程序模块846和程序数据847在这里给出不同的数字来说明在最低限度上，它们是不同的副本。

[0193] 用户可以通过诸如键盘862、麦克风863和诸如鼠标、轨迹球或触摸板之类的指示设备861之类的输入设备将命令和信息录入到计算机810中。其他输入设备(未示出)可以包括触摸屏操纵杆、游戏垫、卫星天线、扫描仪等。这些和其他输入设备通常通过耦合到系统总线的用户输入接口860连接至处理单元820，但是可以通过其他接口和总线结构(诸如并行端口、游戏端口或通用串行总线(USB))连接。

[0194] 视觉显示器891或其他类型的显示设备还经由诸如视频接口890之类的接口连接至系统总线821。除了监控器之外，计算机还可以包括其他外围输出设备，诸如扬声器897和打印机896，其可以通过输出外围接口895连接。

[0195] 计算机810使用到一个或多个远程计算机(诸如远程计算机880)的逻辑连接在联网环境中操作。远程计算机880可以是个人计算机、手持式设备、服务器、路由器、网络PC、对等设备或其他公共网络节点，并且通常包括上文涉及计算机810描述的元件中的许多或所有元件。所描绘的逻辑连接包括局域网(LAN)871和广域网(WAN)873，而且可能包括其他网络。这样的网络环境在办公室、企业范围的计算机网络、内联网和因特网中是常见的。

[0196] 当在LAN联网环境中使用时，计算机810通过网络接口或适配器870连接至LAN871。当在WAN联网环境中使用时，计算机810通常包括调制解调器872或用于通过诸如互联网之类的WAN873建立通信的其他器件。可以是内部或外部的调制解调器872可以经由用户输入接口860或其他适当的机制连接到系统总线821。在网络环境中，涉及计算机810描绘的程序模块或其部分可存储在远程存储器存储设备中。作为示例而非限制，如所示出的，远程

应用程序885驻留在远程计算机880上。应当领会，所示出的网络连接是示例性的，并且可以使用在计算机之间建立通信链路的其他器件。

[0197] 支持方面

[0198] 上文结合附图提供的具体实施方式明确地描述并且支持根据所描述的主题的各个方面。作为说明而不是限制，支持方面包括用于传送标识当前说话者的元数据的计算机系统，该计算机系统包括处理器，其被配置成执行计算机可执行指令；以及存储器，其存储计算机可执行指令，这些计算机可执行指令被配置成接收表示当前说话者的语音的音频数据；基于音频数据来生成当前说话者的音频指纹；通过将当前说话者的音频指纹与包含在说话者指纹存储库中的存储的音频指纹进行比较来执行自动说话者识别；向观察者的客户端设备传送指示当前说话者未被识别的数据；从观察者的客户端设备接收标识当前说话者的标记信息；将当前说话者的音频指纹和标识当前说话者的元数据存储在说话者指纹存储库中；以及向观察者的客户端设备或不同观察者的客户端设备中的至少一个客户端设备传送标识当前说话者的元数据。

[0199] 支持方面包括前述的计算系统，其中存储器还存储计算机可执行指令，该计算机可执行指令被配置成通过基于由大多数观察者提供的身份来标识当前说话者来解决冲突的标记信息。

[0200] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成接收当前说话者已经被正确识别的确认。

[0201] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成从信息源取回当前说话者的附加信息；以及传送标识当前说话者的元数据中的附加信息。

[0202] 支持方面包括前述计算系统中的任一计算系统，其中附加信息包括以下各项中的一项或多项：当前说话者的公司、当前说话者的部门、当前说话者的职务名称或者当前说话者的联系信息。

[0203] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成生成增强音频数据，该增强音频数据包括表示当前说话者的语音的音频数据和标识当前说话者的元数据。

[0204] 支持方面包括前述计算系统中的任一计算系统，其中标识当前说话者的元数据经由增强音频数据被传送到观察者的客户端设备或不同观察者的客户端设备。

[0205] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成存储增强音频数据；接收指示所识别的说话者的查询；从增强音频数据中搜索标识所识别的说话者的元数据；以及输出表示所识别的说话者的语音的增强音频数据的部分。

[0206] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成生成具有多个说话者的会话的转录，其中由所识别的说话者说出的语音文本与用于所识别的说话者的标识符相关联；存储转录；接收指示所识别的说话者的查询；从转录中搜索所识别的说话者的标识符；以及输出包括所识别的说话者所说出的语音文本的转录的部分。

[0207] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行

指令，其被配置成接收表示当前说话者的语音的后续音频数据；基于后续音频数据来生成当前说话者的新音频指纹；通过将当前说话者的新音频指纹与所存储的当前说话者的音频指纹进行比较来执行说话者识别；以及向观察者的客户端设备或不同观察者的客户端设备传送标识当前说话者的元数据。

[0208] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成从观察者的客户端设备接收标识特定识别的说话者的请求；以及当特定识别的说话者当前正在说话时，向观察者的客户端设备传送警报。

[0209] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成为参与者提供在线会议；从参与者的客户端设备接收参与者的音频指纹；以及将参与者的音频指纹和标识参与者的元数据存储在说话者指纹存储库中。

[0210] 支持方面包括前述计算系统中的任一计算系统，其中存储器还存储计算机可执行指令，其被配置成向观察者的客户端设备传送当前说话者的音频指纹。

[0211] 支持方面还包括一种装置、计算机可读存储介质、计算机实现的方法和/或用于实现前述计算机系统或其部分中的任一个的器件。

[0212] 支持方面包括一种由包括一个或多个计算设备的计算机系统执行的计算机实现的方法，用于传送标识当前说话者的元数据，该计算机实现的方法包括：基于表示当前说话者的语音的音频数据来生成当前说话者的音频指纹；基于当前说话者的音频指纹和已存储音频指纹来执行自动说话者识别；当当前说话者未被识别时，从观察者的客户端设备接收标识当前说话者的标记信息；存储当前说话者的音频指纹和标识当前说话者的元数据；以及向观察者的客户端设备或不同观察者的客户端设备中的至少一个客户端设备传送标识当前说话者的元数据。

[0213] 支持方面包括前述计算机实现方法，其还包括：向观察者的客户端设备传送指示当前说话者未被识别的数据。

[0214] 支持方面包括前述计算机实现方法中的任一方法，其还包括：通过基于由大多数观察者提供的身份来标识当前说话者以解决冲突的标记信息。

[0215] 支持方面包括前述计算机实现方法中的任一方法，其还包括：基于表示当前说话者的语音的后续音频数据来生成当前说话者的新音频指纹；以及基于当前说话者的新音频指纹和所存储的当前说话者的音频指纹来执行说话者识别。

[0216] 支持方面还包括一种系统、装置、计算机可读存储介质和/或用于实现和/或执行前述计算机实现的方法或其部分中的任一个的器件。

[0217] 支持方面包括一种存储计算机可执行指令的计算机可读存储介质，当由计算设备执行时，该计算机可执行指令使得计算设备实现：说话者识别部件，其被配置成基于表示当前说话者的语音的音频数据来生成当前说话者的音频指纹，并且通过将当前说话者的音频指纹与已存储音频指纹进行比较来执行自动说话者识别；标记部件，其被配置成当自动说话者识别不成功时，从观察者的客户端设备接收标识当前说话者的标记信息，并且将当前说话者的音频指纹与存储的音频指纹一起存储；以及音频数据充实部件，其被配置成向观察者的客户端设备或不同观察者的客户端设备传送标识当前说话者的元数据。

[0218] 支持方面包括前述计算机可读存储介质，其中标记部件还被配置成通过基于由大多数观察者提供的身份来标识当前说话者以解决冲突的标记信息。

[0219] 支持方面包括前述计算机可读存储介质中的任一计算机可读存储介质，其中音频数据充实部件还被配置成传送表示当前说话者的语音的音频数据和标识当前说话者的元数据作为同步的音频数据和元数据流。

[0220] 根据关于功率消耗、存储器、处理器周期和/或其他计算方面昂贵的资源的效率改善和/或节省，支持方面可以提供各种伴随的和/或技术的优点。

[0221] 上文结合附图提供的具体实施方式旨在作为示例的描述，而不旨在表示可以构造或利用本示例的唯一形式。

[0222] 应当理解，本文中所描述的配置和/或途径本质上是示例性的，并且因为许多变化是可能的，所以所描述的实施例、实现方式和/或示例不被认为是限制性的。本文中所描述的具体过程或方法可以表示任何数目个处理策略中的一个或多个处理策略。如此，可以以所图示和/描述的顺序、以其他顺序、并行地或省略地来执行所图示和/或描述的各种操作。类似地，可以改变上文所描述的过程的顺序。

[0223] 尽管主题已经以结构特征和/或方法动作特有的语言进行了描述，但是应当理解，所附权利要求中限定的主题不一定限于上文所描述的具体特征或动作。相反，上文所描述的具体特征和动作被呈现为实现权利要求的示例形式。

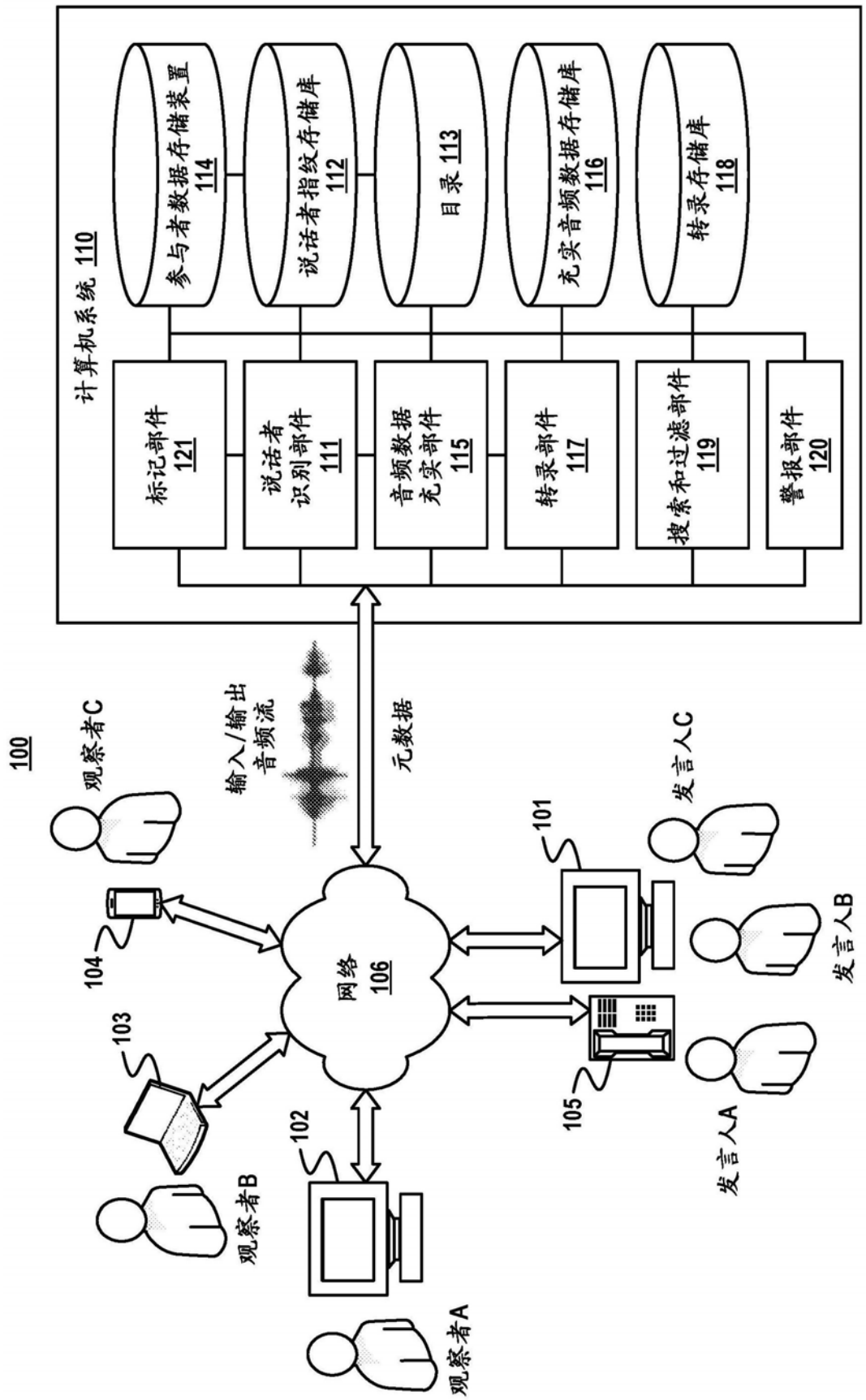

100
计算机系统 110
参与者数据存储装置 114
说话者指纹存储库 112
目录 113
充实音频数据存储库 116
转录存储库 118
标记部件 121
说话者识别部件 111
音频数据充实部件 115
转录部件 117
搜索和过滤部件 119
警报部件 120
输入/输出音频流
元数据
网络 106
观察者C
观察者B
观察者A
发言人C
发言人B
发言人A
101
102
103
104
105

图1

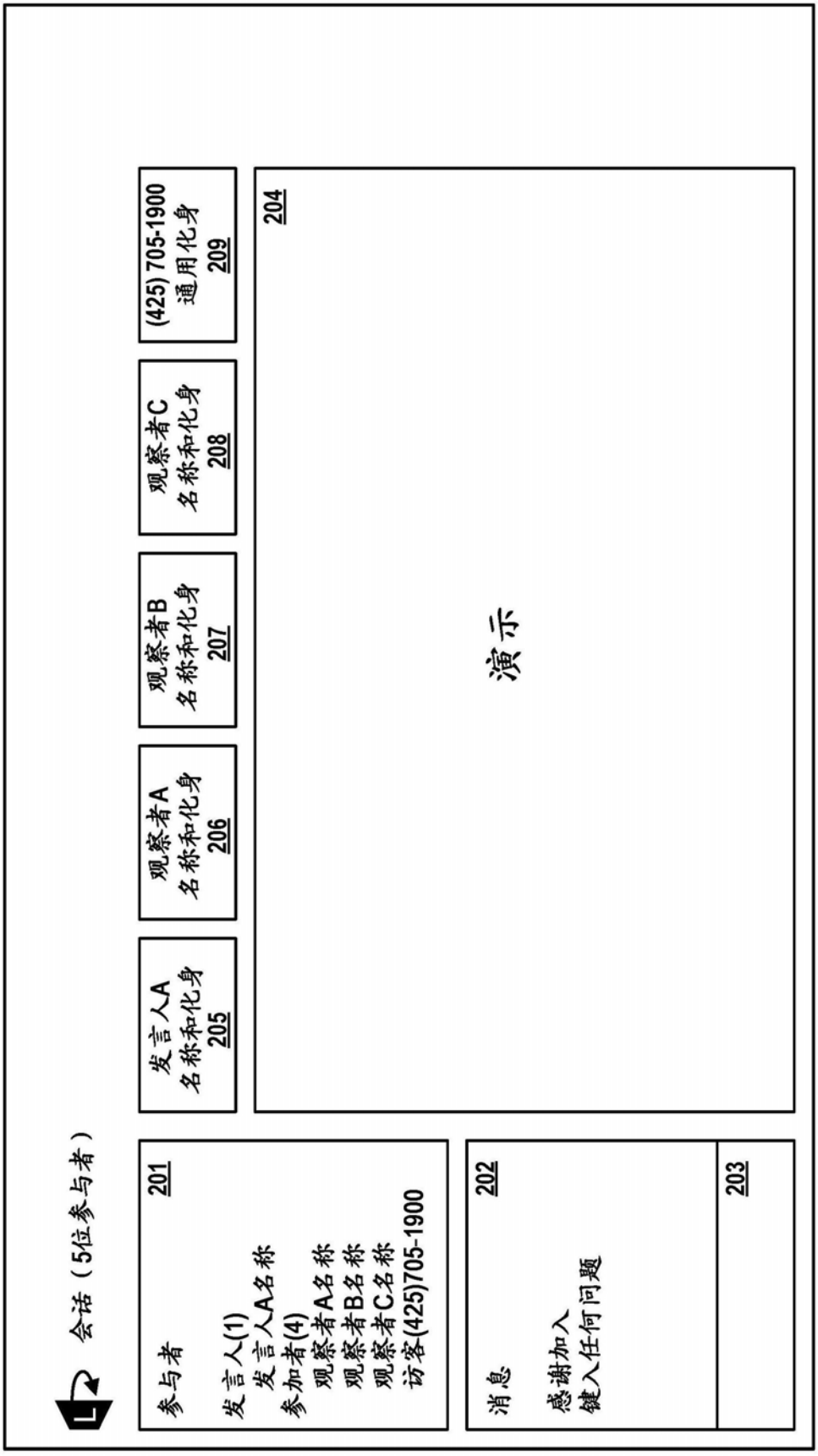

200
会话（5位参与者）
参与者 201
发言人(1)
发言人A名称
参加者(4)
观察者A名称
观察者B名称
观察者C名称
访客(425)705-1900
消息 202
感谢加入
键入任何问题
203
发言人A
名称和化身
205
观察者A
名称和化身
206
观察者B
名称和化身
207
观察者C
名称和化身
208
(425) 705-1900
通用化身
209
演示
204

图2A

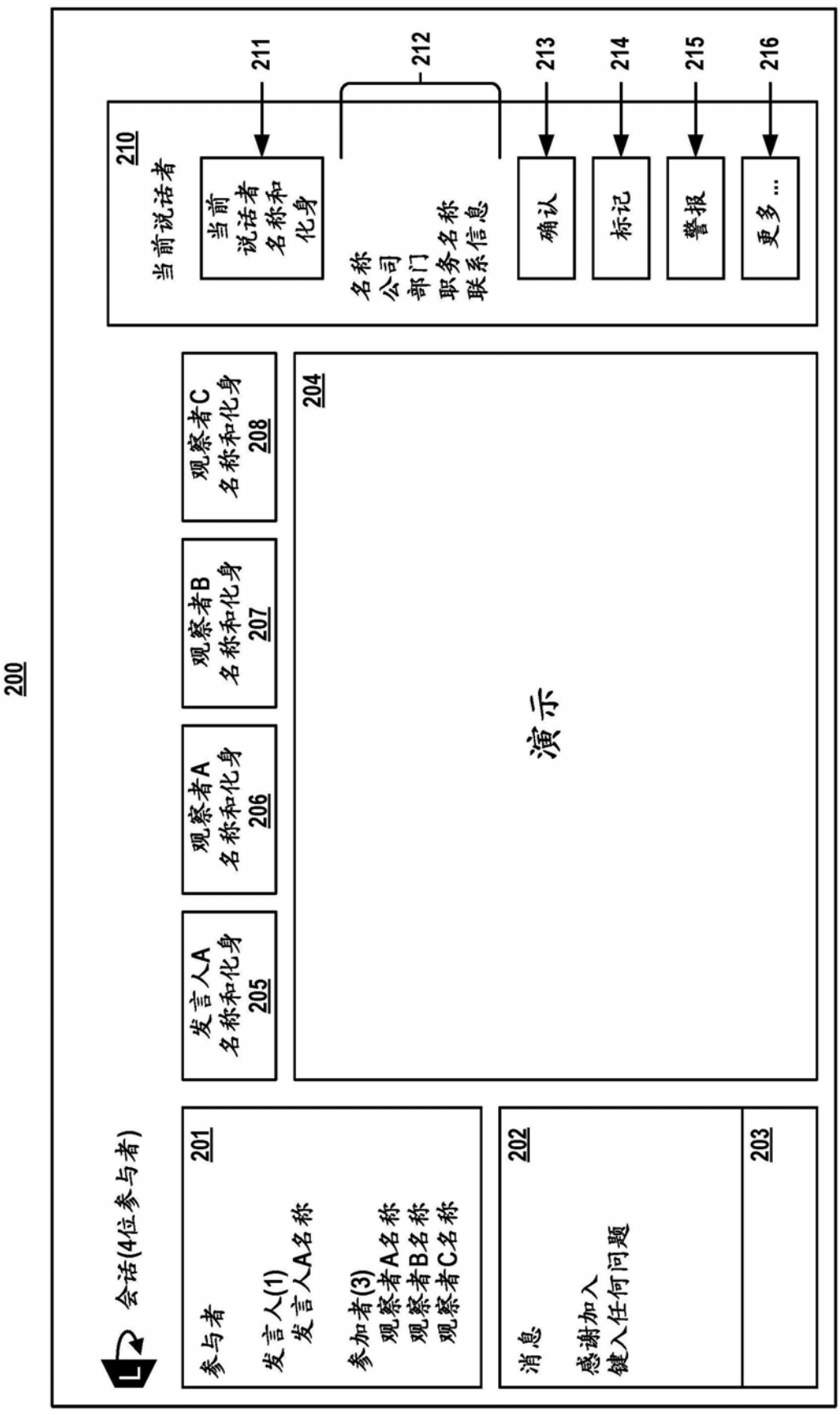

200
会话(4位参与者)
参与者 201
发言人(1)
发言人A名称
参加者(3)
观察者A名称
观察者B名称
观察者C名称
消息 202
感谢加入
键入任何问题
203
发言人A
名称和化身
205
观察者A
名称和化身
206
观察者B
名称和化身
207
观察者C
名称和化身
208
演示
204
210
当前说话者
当前
说话者
名称和
化身
211
名称
公司
部门
职务名称
联系信息
212
确认
213
标记
214
警报
215
更多...
216

图2B

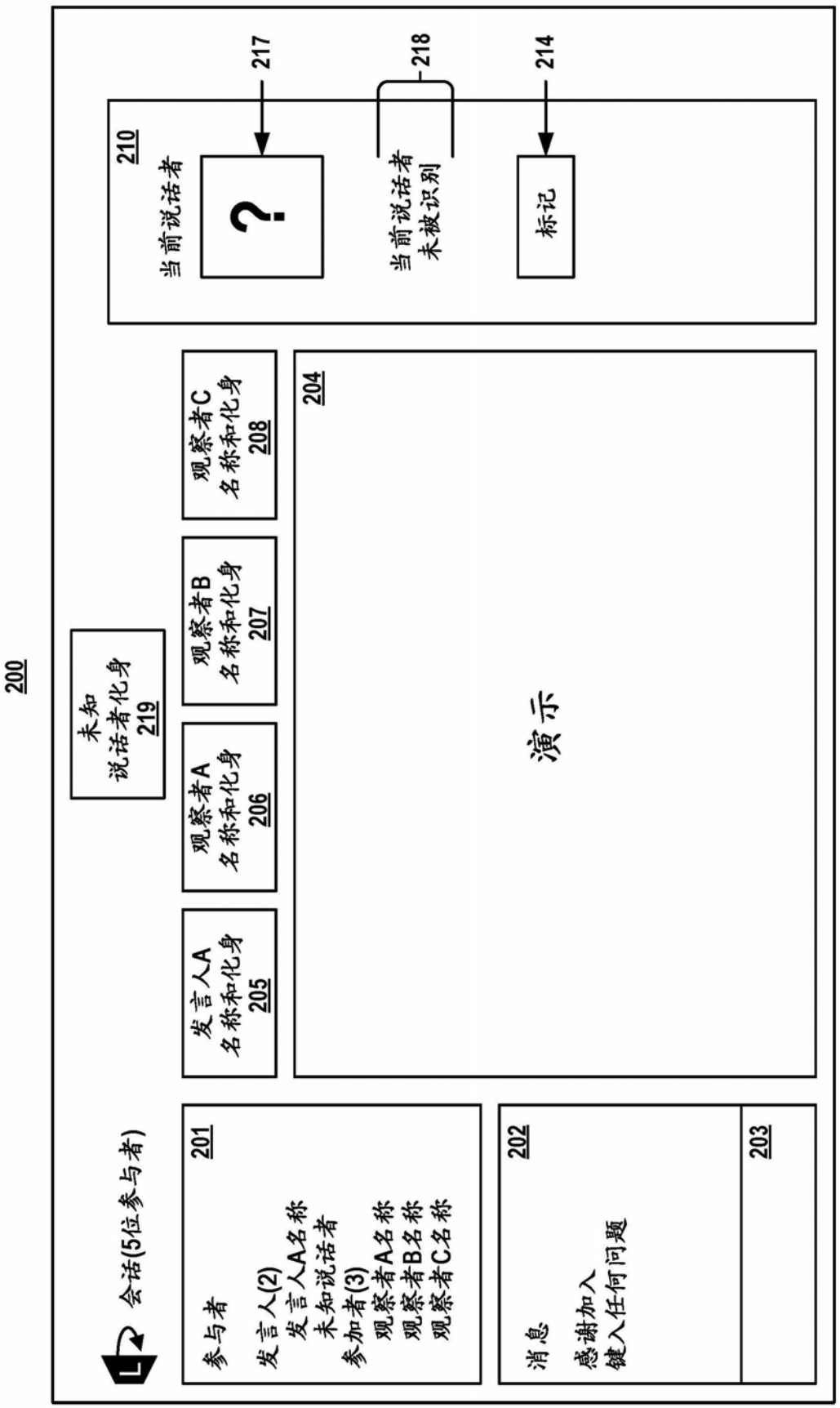

200
会话(5位参与者)
参与者 201
发言人(2)
发言人A名称
未知说话者
参加者(3)
观察者A名称
观察者B名称
观察者C名称
消息 202
感谢加入
键入任何问题
203
发言人A
名称和化身
205
观察者A
名称和化身
206
未知
说话者化身
219
观察者B
名称和化身
207
观察者C
名称和化身
208
演示
204
210
当前说话者
?
217
当前说话者
未被识别
218
标记
214

图2C

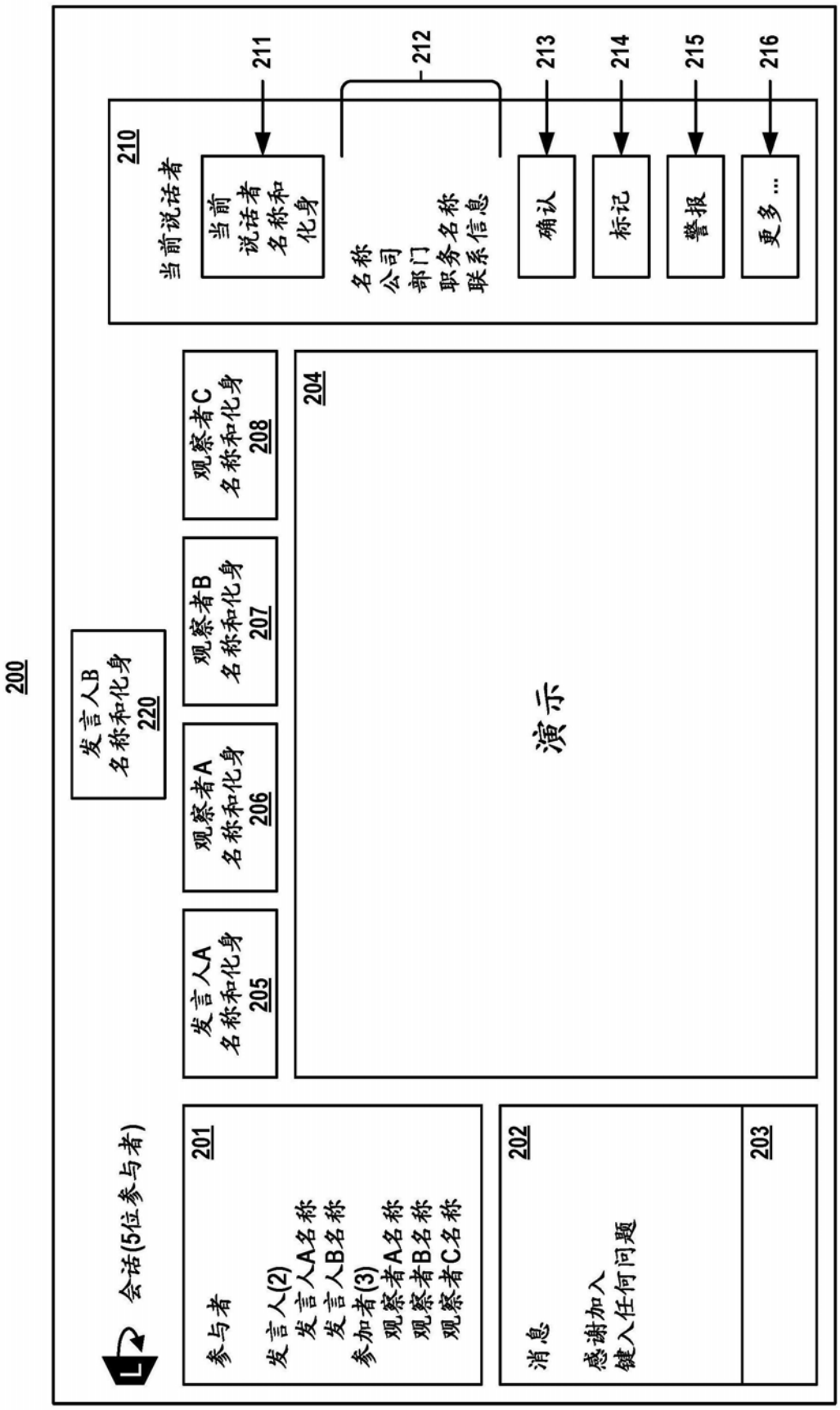

200
会话(5位参与者)
201
参与者
发言人(2)
发言人A名称
发言人B名称
参加者(3)
观察者A名称
观察者B名称
观察者C名称
202
消息
感谢加入
键入任何问题
203
发言人A
名称和化身
205
观察者A
名称和化身
206
发言人B
名称和化身
220
观察者B
名称和化身
207
观察者C
名称和化身
208
204
演示
210
当前说话者
当前
说话者
名称和
化身
211
名称
公司
部门
职务名称
联系信息
212
确认
213
标记
214
警报
215
更多...
216

图2D

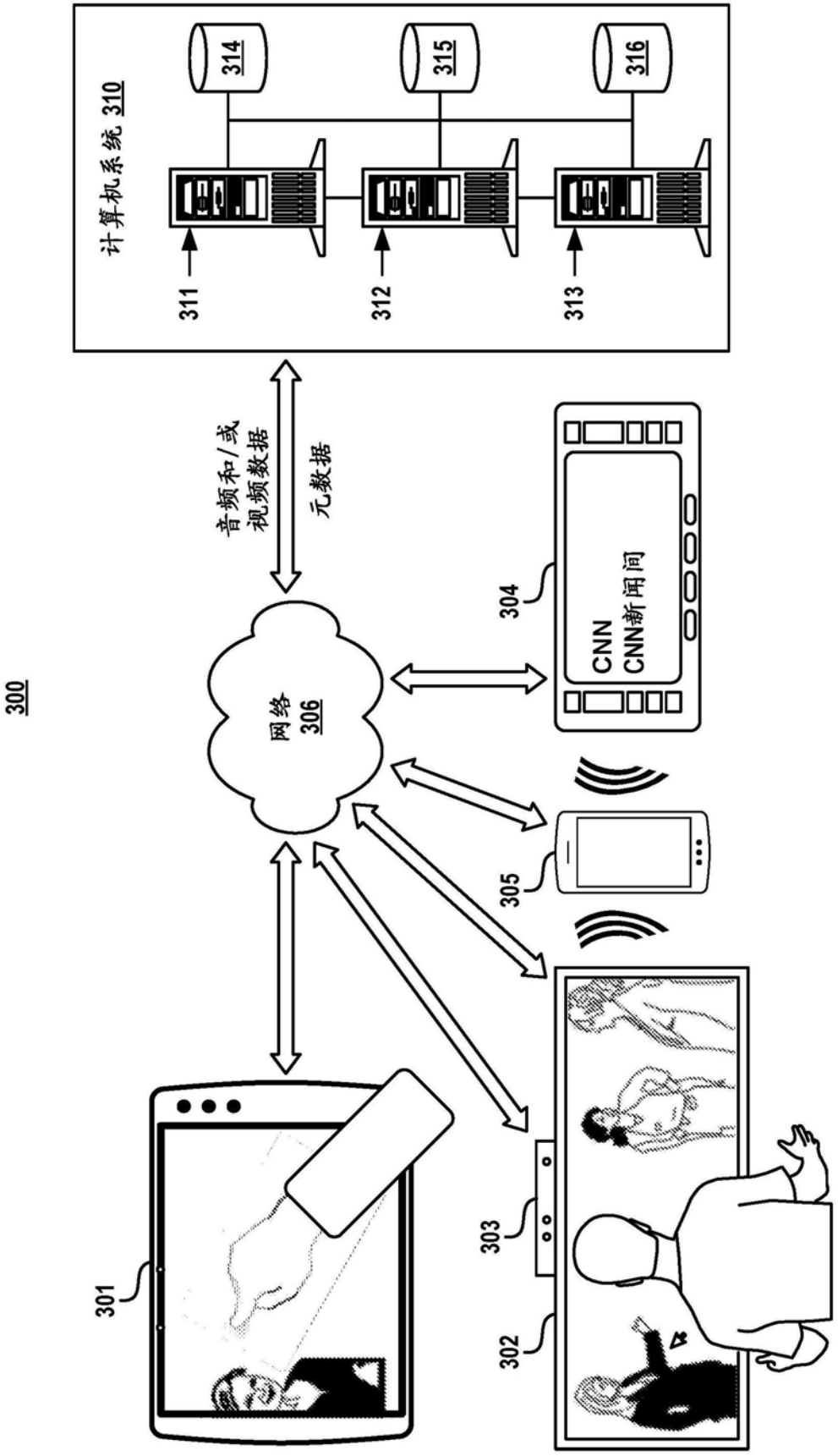

300
计算机系统 310
314
315
316
311
312
313
音频和/或
视频数据
元数据
网络
306
304
CNN
CNN新闻间
305
301
303
302

图3

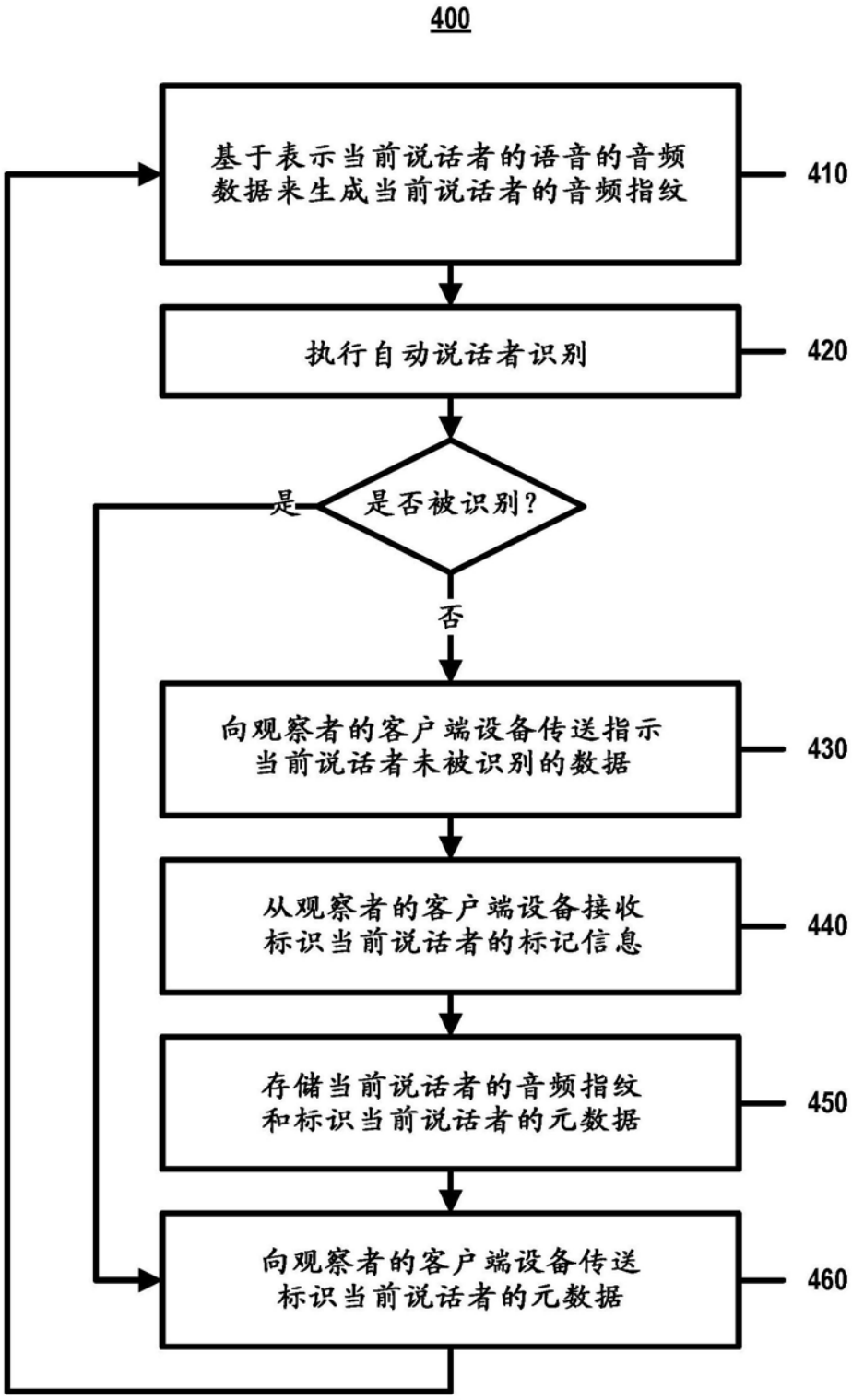

400
基于表示当前说话者的语音的音频数据来生成当前说话者的音频指纹
410
执行自动说话者识别
420
是否被识别?
是
否
向观察者的客户端设备传送指示当前说话者未被识别的数据
430
从观察者的客户端设备接收标识当前说话者的标记信息
440
存储当前说话者的音频指纹和标识当前说话者的元数据
450
向观察者的客户端设备传送标识当前说话者的元数据
460

图4

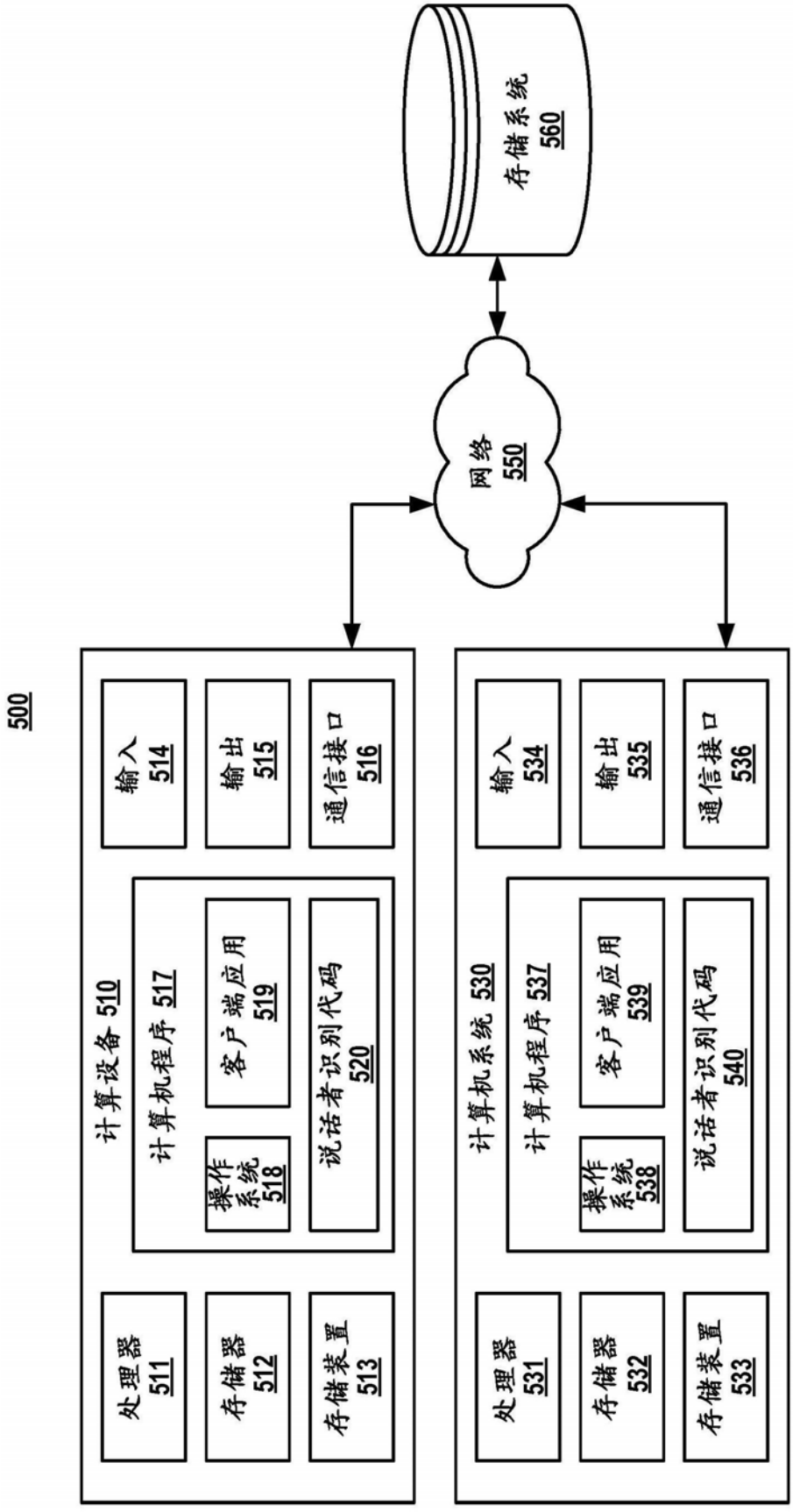

500
存储系统 560
网络 550
计算设备 510
处理器 511
存储器 512
存储装置 513
计算机程序 517
操作系统 518
客户端应用 519
说话者识别代码 520
输入 514
输出 515
通信接口 516
计算机系统 530
处理器 531
存储器 532
存储装置 533
计算机程序 537
操作系统 538
客户端应用 539
说话者识别代码 540
输入 534
输出 535
通信接口 536

图5

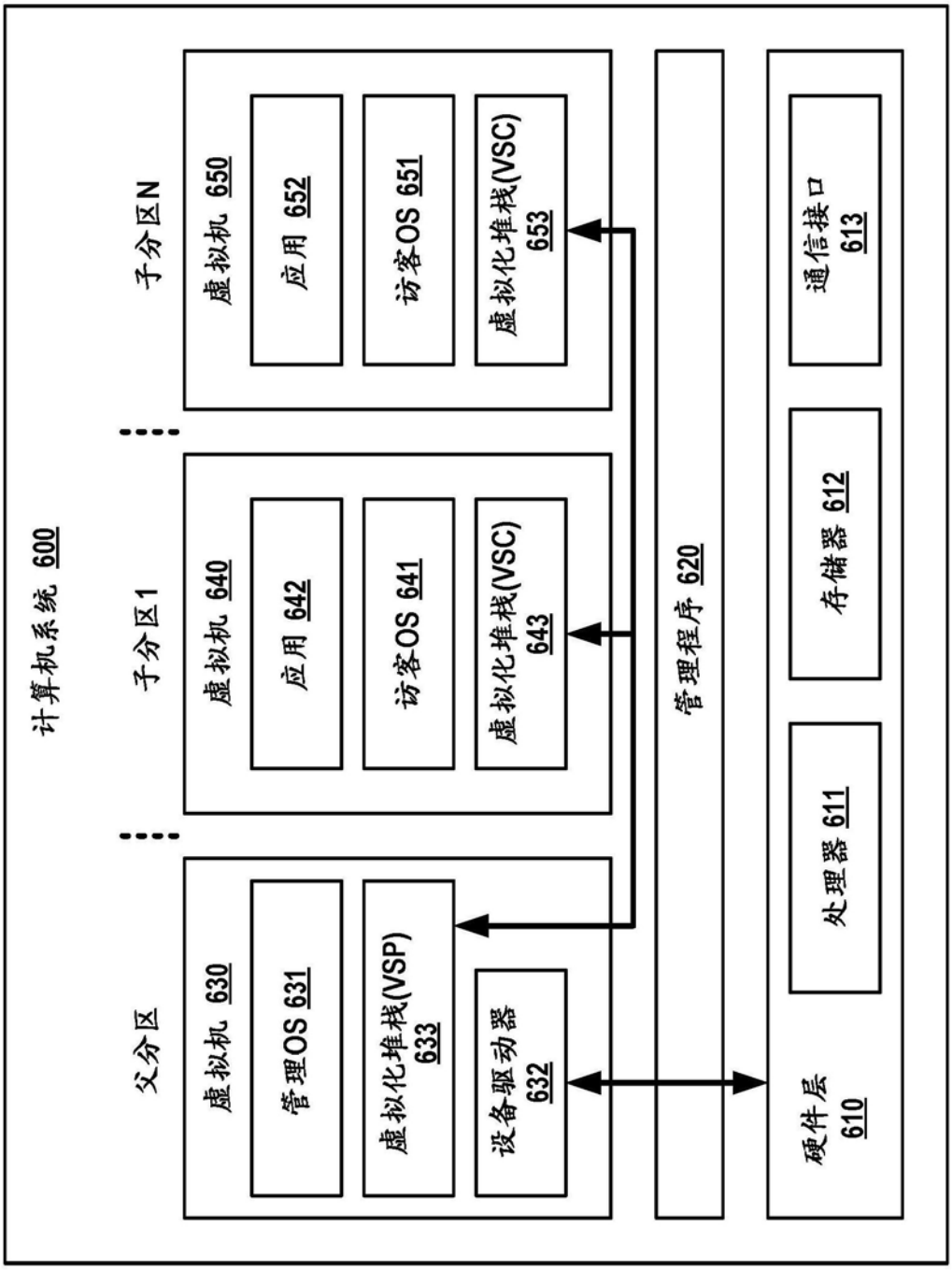

计算机系统 600
父分区
虚拟机 630
管理OS 631
虚拟化堆栈(VSP) 633
设备驱动器 632
子分区1
虚拟机 640
应用 642
访客OS 641
虚拟化堆栈(VSC) 643
子分区N
虚拟机 650
应用 652
访客OS 651
虚拟化堆栈(VSC) 653
管理程序 620
硬件层 610
处理器 611
存储器 612
通信接口 613

图6


移动计算设备
700
操作系统
702
应用
704
I/O端口
728
存储器
708
不可移除存储器
710
可移除存储器
712
电源
730
加速度计
732
输入设备
714
触摸屏
715
麦克风
716
相机
717
键盘
718
轨迹球
719
处理器
706
输出设备
720
扬声器
721
显示器
722
GPS接收器
736
物理连接器
734
无线调制
解调器
724
蜂窝
725
WI-FI
726
蓝牙
727

图7

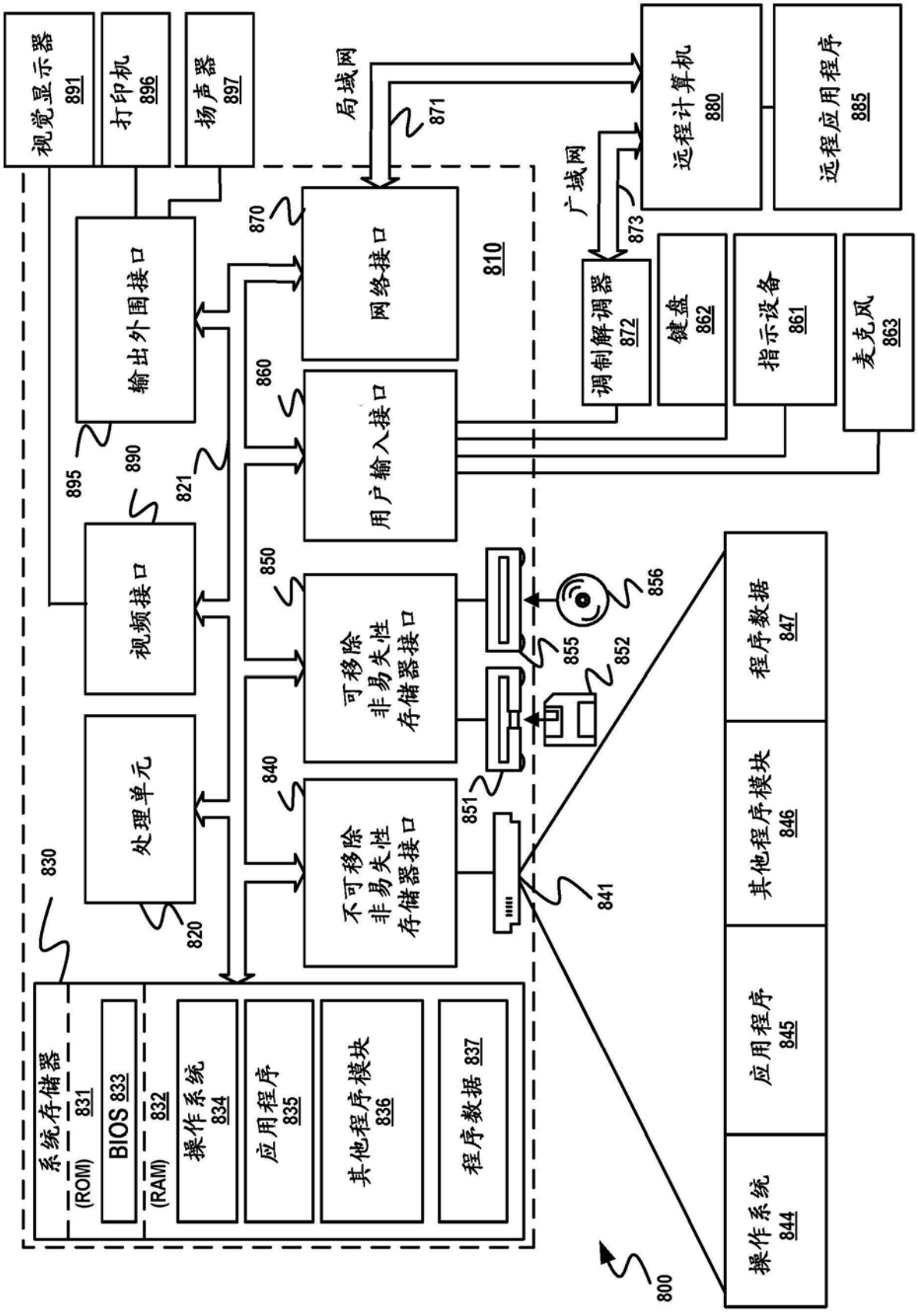

视觉显示器 891
打印机 896
扬声器 897
局域网
871
远程计算机 880
远程应用程序 885
广域网
873
870
网络接口
输出外围接口
895
890
821
860
用户输入接口
调制解调器 872
键盘 862
指示设备 861
麦克风 863
810
视频接口
850
可移除 非易失性 存储器接口
856
855
852
处理单元
840
不可移除 非易失性 存储器接口
851
841
程序数据 847
其他程序模块 846
应用程序 845
操作系统 844
830
820
系统存储器
(ROM) 831
BIOS 833
(RAM) 832
操作系统 834
应用程序 835
其他程序模块 836
程序数据 837
800

图8