



(12) 发明专利申请

(10) 申请公布号 CN 116830561 A

(43) 申请公布日 2023. 09. 29

(21) 申请号 202280013990.5

D · 古纳万

(22) 申请日 2022.02.07

(74) 专利代理机构 北京市汉坤律师事务所

11602

(30) 优先权数据

专利代理师 魏小微 吴丽丽

21177382.5 2021.06.02 EP

63/147,573 2021.02.09 US

63/201,939 2021.05.19 US

(51) Int.Cl.

H04M 9/08 (2006.01)

(85) PCT国际申请进入国家阶段日

2023.08.08

(86) PCT国际申请的申请数据

PCT/US2022/015529 2022.02.07

(87) PCT国际申请的公布数据

W02022/173706 EN 2022.08.18

(71) 申请人 杜比实验室特许公司

地址 美国加利福尼亚州

(72) 发明人 B · J · 索斯韦尔 C · G · 海因斯

权利要求书2页 说明书29页 附图14页

(54) 发明名称

回声参考优先级排序和选择

(57) 摘要

一些实施方式涉及获得多个回声参考,多个回声参考包括针对音频环境中的多个音频设备中的每个音频设备的至少一个回声参考,每个回声参考对应于由多个音频设备中的一个音频设备的一个或多个扩音器回放的音频数据。一些示例涉及对多个回声参考中的每个回声参考做出重要性估计。做出重要性估计可以涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻的预期贡献。一些实施方式涉及至少部分地基于重要性估计来选择一个或多个所选回声参考并将一个或多个所选回声参考提供给至少一个回声管理系统。

1. 一种用于管理音频系统的多个音频设备中的第一音频设备的回声的音频处理方法, 其中, 所述多个音频设备中的每个音频设备包括一个或多个扩音器, 其中, 所述第一音频设备进一步包括控制系统, 其中, 所述控制系统包括回声管理系统, 所述回声管理系统包括声学回声消除器(AEC)、声学回声抑制器(AES)、或者AEC和AES两者, 所述方法包括:

由所述第一音频设备的所述控制系统获得多个回声参考, 所述多个回声参考包括针对所述多个音频设备中的每个音频设备的至少一个回声参考, 每个回声参考对应于由对应音频设备的所述一个或多个扩音器回放的音频数据;

由所述控制系统对所述多个回声参考中的每个回声参考做出重要性估计, 其中, 做出所述重要性估计涉及由所述第一音频设备的所述回声管理系统确定每个回声参考对回声减轻的预期贡献;

由所述控制系统并且至少部分地基于所述重要性估计从所述多个回声参考中选择一个或多个回声参考;

由所述控制系统将所述一个或多个所选回声参考提供给所述回声管理系统; 以及

由所述第一音频设备的所述回声管理系统至少部分地基于所述一个或多个所选回声参考来抑制或消除回声。

2. 如权利要求1所述的音频处理方法, 其中, 获得所述多个回声参考涉及:

接收包括音频数据的内容流; 以及

基于所述音频数据确定所述多个回声参考中的一个或多个回声参考。

3. 如权利要求2所述的音频处理方法, 进一步包括:

由所述控制系统渲染所述音频数据以供在所述第一音频设备上再现, 以产生本地扬声器馈送信号; 以及

确定与所述本地扬声器馈送信号相对应的本地回声参考。

4. 如权利要求3所述的音频处理方法, 其中, 获得所述多个回声参考涉及基于所述音频数据确定一个或多个非本地回声参考, 所述非本地回声参考中的每一个对应于用于在所述音频环境的另一音频设备上回放的非本地扬声器馈送信号。

5. 如权利要求3所述的音频处理方法, 其中, 获得所述多个回声参考涉及接收一个或多个非本地回声参考, 所述非本地回声参考中的每一个对应于用于在所述音频环境的另一音频设备上回放的非本地扬声器馈送信号。

6. 如权利要求5所述的音频处理方法, 其中, 接收所述一个或多个非本地回声参考涉及从所述音频环境的一个或多个其他音频设备接收所述一个或多个非本地回声参考。

7. 如权利要求5所述的音频处理方法, 其中, 接收所述一个或多个非本地回声参考涉及从所述音频环境的单个其他设备接收所述一个或多个非本地回声参考中的每一个。

8. 如权利要求1至7中任一项所述的音频处理方法, 进一步包括成本确定, 所述成本确定涉及确定所述多个回声参考中的至少一个回声参考的成本, 其中, 选择所述一个或多个所选回声参考至少部分地基于所述成本确定。

9. 如权利要求8所述的音频处理方法, 其中, 所述成本确定基于用于传输所述至少一个回声参考所需的网络带宽、用于编码所述至少一个回声参考的编码计算要求、用于解码所述至少一个回声参考的解码计算要求、用于由所述回声管理系统使用所述至少一个回声参考的回声管理系统计算要求、或其组合。

10. 如权利要求8或权利要求9所述的音频处理方法,其中,所述成本确定基于所述至少一个回声参考在时域或频域中的复制品、所述至少一个回声参考的下采样版本、所述至少一个回声参考的有损压缩、所述至少一个回声参考的分段功率信息、与相对不太重要的回声参考相比对相对更重要的回声参考进行更少压缩的方法、或其组合。

11. 如权利要求8至10中任一项所述的音频处理方法,其中,所述成本确定基于与相对不太重要的回声参考相比对相对更重要的回声参考进行更少压缩的方法。

12. 如权利要求1至11中任一项所述的音频处理方法,进一步包括确定当前回声管理系统性能水平,其中,选择所述一个或多个所选回声参考至少部分地基于所述当前回声管理系统性能水平。

13. 如权利要求1至12中任一项所述的音频处理方法,其中,做出所述重要性估计涉及确定对应回声参考的重要性度量。

14. 如权利要求13所述的音频处理方法,其中,确定所述重要性度量至少部分地基于所述对应回声参考的水平、所述对应回声参考的唯一性、所述对应回声参考的时间持续性、所述对应回声参考的可听度、或其组合。

15. 如权利要求13或权利要求14所述的音频处理方法,其中,确定所述重要性度量至少部分地基于与音频设备布局相对应的元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合矩阵、扩音器激活矩阵、或其组合。

16. 如权利要求13至15中任一项所述的音频处理方法,其中,确定所述重要性度量至少部分地基于当前收听目标、当前环境噪声估计、所述回声管理系统的当前性能的估计、或其组合。

17. 如前述权利要求中任一项所述的音频处理方法,其中,所述音频系统的音频设备经由有线或无线通信网络通信地耦接,并且其中,所述多个回声参考是经由所述有线或无线通信网络获得的。

18. 一种装置,所述装置被配置为执行如权利要求1至17中任一项所述的方法。

19. 一种系统,所述系统被配置为执行如权利要求1至17中任一项所述的方法。

20. 一个或多个其上存储有软件的非暂态介质,所述软件包括用于控制一个或多个设备执行如权利要求1至17中任一项所述的方法的指令。

回声参考优先级排序和选择

[0001] 相关申请的交叉引用

[0002] 本申请要求于2021年2月9日提交的美国临时申请No.63/147,573、于2021年5月19日提交的美国临时申请No.63/201,939以及于2021年6月2日提交的欧洲申请No.21177382.5的优先权,所有这些申请均通过援引被整体并入本文。

技术领域

[0003] 本公开涉及用于实施声学回声管理的设备、系统和方法。

背景技术

[0004] 具有声学回声管理系统的音频设备已被广泛部署。声学回声管理系统可以包括声学回声消除器和/或声学回声抑制器。尽管用于声学回声管理的现有设备、系统和方法提供了益处,但改进的设备、系统和方法将仍是期望的。

[0005] 符号和术语

[0006] 贯穿本公开,包括在权利要求书中,术语“扬声器(speaker)”、“扩音器(loudspeaker)”和“音频再现换能器”同义地用于表示任何发声换能器(或一组换能器)。一套典型的耳机包括两个扬声器。扬声器可以被实施为包括多个换能器(例如,低音扬声器和高音扬声器),所述换能器可以由单个公共扬声器馈送或多个扬声器馈送驱动。在一些示例中,(多个)扬声器馈送可以在耦接到不同换能器的不同电路分支中经历不同处理。

[0007] 贯穿本公开,包括在权利要求中,在广义上使用“对信号或数据执行操作(performing an operation“on” a signal or data)”的表达(例如,对信号或数据进行滤波、缩放、变换或应用增益)来表示直接对信号或数据执行操作或对信号或数据的已处理版本(例如,在对其执行操作之前已经历了初步滤波或预处理的信号版本)执行操作。

[0008] 贯穿本公开,包括在权利要求中,在广义上使用表达“系统”来表示设备、系统或子系统。例如,实施解码器的子系统可以被称为解码器系统,并且包括这样的子系统的系统(例如,响应于多个输入而生成X个输出信号的系统,其中,所述子系统生成M个输入,而其他X-M个输入是从外部源接收的)也可以被称为解码器系统。

[0009] 贯穿本公开,包括在权利要求中,在广义上使用术语“处理器”来表示可编程或以其他方式可配置(例如,用软件或固件)为对数据(例如,音频或视频或其他图像数据)执行操作的系统或设备。处理器的示例包括现场可编程门阵列(或其他可配置集成电路或芯片组)、被编程和/或以其他方式被配置为对音频或其他声音数据执行流水线式处理的数字信号处理器、可编程通用处理器或计算机、以及可编程微处理器芯片或芯片组。

[0010] 贯穿本公开,包括在权利要求中,术语“耦接(couple)”或“耦接的(coupled)”用于意指直接或间接连接。因此,如果第一设备耦接到第二设备,则该连接可以通过直接连接或者通过经由其他设备和连接的间接连接实现。

[0011] 如本文所使用的,“智能设备”是可以在某种程度上交互地和/或自主地操作的电子设备,其通常被配置用于经由如蓝牙、Zigbee、近场通信、Wi-Fi、光保真(Li-Fi)、3G、4G、

5G等各种无线协议与一个或多个其他设备(或网络)进行通信。一些著名的智能设备类型是智能电话、智能汽车、智能恒温器、智能门铃、智能锁、智能冰箱、平板手机和平板计算机、智能手表、智能手环、智能钥匙链和智能音频设备。术语“智能设备”还可以是指展现出如人工智能等普适计算的一些性质的设备。

[0012] 在本文中,使用表达“智能音频设备”来表示智能设备,其是单一用途音频设备或多用途音频设备(例如,实施虚拟助理功能的至少一些方面的音频设备)。单一用途音频设备是包括或耦接到至少一个麦克风(并且可选地还包括或耦接到至少一个扬声器和/或至少一个相机)并且很大程度上或主要被设计为实现单一用途的设备(例如,电视(TV))。例如,尽管TV通常可以播放(并且被认为能够播放)来自节目素材的音频,但在大多数实例中,现代TV运行某种操作系统,应用程序(包括看电视的应用程序)在所述操作系统上本地运行。从这个意义上说,具有(多个)扬声器和(多个)麦克风的单一用途音频设备通常被配置为运行本地应用程序和/或服务以直接使用所述(多个)扬声器和(多个)麦克风。一些单一用途音频设备可以被配置为组合在一起以实现在一定区或用户配置区域上播放音频。

[0013] 一种常见类型的多用途音频设备是实施虚拟助理功能的至少一些方面的音频设备,尽管虚拟助理功能的其他方面可以由比如一个或多个服务器等一个或多个其他设备来实施,多用途音频设备被配置用于与所述一个或多个服务器通信。这样的多用途音频设备在本文中可以被称为“虚拟助理”。虚拟助理是包括或耦接到至少一个麦克风(并且可选地还包括或耦接到至少一个扬声器和/或至少一个相机)的设备(例如,智能扬声器或语音助理集成设备)。在一些示例中,虚拟助理可以提供将多个设备(不同于虚拟助理)用于某种意义上支持云的应用程序或以其他方式未在虚拟助理本身中或之上完全实施的应用程序的能力。换句话说,虚拟助理功能的至少一些方面(例如,言语识别功能)可以(至少部分地)由一个或多个服务器或其他设备实施,虚拟助理可以经由网络(如因特网)与所述一个或多个服务器或其他设备通信。虚拟助理有时可以一起工作,例如,以离散和有条件地定义的方式。例如,两个或更多个虚拟助理可以在其中之一(例如,最确信已经听到唤醒词的虚拟助理)对唤醒词作出响应的意义上一起工作。在一些实施方式中,连接的虚拟助理可以形成一种星座,所述星座可以由一个主应用程序管理,所述主应用程序可以是(或实施)虚拟助理。

[0014] 在本文中,“唤醒词”在广义上用于表示任何声音(例如,人类说出的词或其他声音),其中智能音频设备被配置成响应于检测到(“听到”)声音(使用包括在智能音频设备中或耦接到所述智能音频设备的至少一个麦克风,或至少一个其他麦克风)而唤醒。在这种背景下,“唤醒”表示设备进入等待(换句话说,正在收听)声音命令的状态。在一些实例中,本文中所谓的“唤醒词”可以包括多于一个词,例如,短语。

[0015] 在本文中,表达“唤醒词检测器”表示被配置成连续搜索实时声音(例如,言语)特征与训练模型之间的对齐的设备(或表示包括用于将设备配置成连续搜索实时声音特征与训练模型之间的对齐的指令的软件)。通常,每当唤醒词检测器确定检测到唤醒词的概率超过预定义阈值,就会触发唤醒词事件。例如,所述阈值可以是被调整以在错误接受率与错误拒绝率之间给出合理折衷的预定阈值。在唤醒词事件之后,设备可能会进入一种状态(可以被称为“唤醒”状态或“注意力”状态),在所述状态下设备会收听命令并且将接收到的命令传递给更大、计算更密集的认可器。

[0016] 如本文所使用的,术语“节目流”和“内容流”是指一个或多个音频信号的集合,并

且在一些实例中是指视频信号的集合,所述信号集合的至少部分是旨在要一起听到的。示例包括音乐选集、电影原声、电影、电视节目、电视节目的音频部分、播客、现场语音通话、来自智能助理的合成语音响应等。在一些实例中,内容流可以包括音频信号的至少一部分的多个版本,例如,多于一种语言的同一对话。在这样的实例中,一次旨在再现音频数据或其部分的仅一个版本(例如,与单一语言相对应的版本)。

发明内容

[0017] 本公开的至少一些方面可以经由一种或多种音频处理方法来实施。所述音频处理方法管理音频系统中的回声。所述音频系统包括音频环境中的多个音频设备。所述多个音频设备中的每个设备包括一个或多个扩音器。在一些实例中,(多种)方法可以至少部分地由控制系统和/或经由存储在所述音频系统的所述多个音频设备中的第一设备的一个或多个非暂态介质上的指令(例如,软件)来实施。所述第一设备可以包括一个或多个麦克风。一些这样的方法涉及由所述第一设备的所述控制系统获得多个回声参考。所述多个回声参考可以包括针对所述音频环境中的所述多个音频设备中的每个音频设备的至少一个回声参考。每个回声参考可以对应于由所述多个音频设备中的对应音频设备的一个或多个扩音器回放的音频数据。所述多个回声参考包括所述第一音频设备的至少一个回声参考。

[0018] 所述方法可以涉及由所述控制系统对所述多个回声参考中的每个回声参考做出重要性估计。在一些示例中,做出重要性估计可以涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻的预期贡献。(多个)回声管理系统可以例如包括声学回声消除器(AEC)和/或声学回声抑制器(AES)。

[0019] 所述方法可以涉及由所述控制系统并且至少部分地基于所述重要性估计从所述多个回声参考中选择一个或多个回声参考。所选回声参考可以是(整个)多个回声参考中的一个或多个回声参考的子集。所述方法可以涉及由所述控制系统将所述一个或多个所选回声参考提供给所述至少一个回声管理系统。在一些示例中,所述方法可以涉及使得至少一个回声管理系统至少部分地基于所述一个或多个所选回声参考来消除或抑制回声。

[0020] 根据一些示例,所述音频系统的音频设备可以经由有线或无线网络通信地耦接。所述多个回声参考(例如,不同于所述第一音频设备的其他音频设备的非本地回声参考和/或所述第一音频设备的回声参考)可以经由所述有线或无线网络获得。

[0021] 根据一些示例,获得所述多个回声参考可以涉及接收包括音频数据的内容流并基于所述音频数据来确定所述多个回声参考中的一个或多个回声参考。

[0022] 在一些实施方式中,所述控制系统可以是或者可以包括所述音频环境中的音频设备的音频设备控制系统。在一些这样的实施方式中,所述方法可以涉及由音频设备控制系统渲染音频数据以用于在音频设备上再现,从而产生本地扬声器馈送信号。在一些这样的实施方式中,所述方法可以涉及确定与所述本地扬声器馈送信号相对应的本地回声参考。

[0023] 在一些示例中,获得多个回声参考可以涉及基于音频数据来确定一个或多个非本地回声参考。在一些这样的示例中,每个非本地回声参考可以对应于用于在所述音频环境的另一音频设备上回放的非本地扬声器馈送信号。

[0024] 根据一些示例,获得多个回声参考可以涉及接收一个或多个非本地回声参考。在一些这样的示例中,每个非本地回声参考可以对应于用于在所述音频环境的另一音频设备

上回放的非本地扬声器馈送信号。在一些示例中,接收一个或多个非本地回声参考可以涉及从音频环境的一个或多个其他音频设备接收一个或多个非本地回声参考。在一些示例中,接收一个或多个非本地回声参考可以涉及从音频环境的单个其他设备接收一个或多个非本地回声参考中的每一个。

[0025] 在一些示例中,所述方法可以涉及成本确定。根据一些示例,成本确定可以涉及确定多个回声参考中的至少一个回声参考的成本。在一些示例中,选择所述一个或多个所选回声参考可以至少部分地基于所述成本确定。

[0026] 根据一些示例,所述成本确定可以基于用于传输所述至少一个回声参考所需的网络带宽、用于编码所述至少一个回声参考的编码计算要求、用于解码所述至少一个回声参考的解码计算要求、用于由所述回声管理系统使用所述至少一个回声参考的回声管理系统计算要求、或其一个或多个组合。

[0027] 在一些示例中,成本确定可以基于至少一个回声参考在时域或频域中的复制品、至少一个回声参考的下采样版本、至少一个回声参考的有损压缩、至少一个回声参考的分段功率信息、或其一个或多个组合。根据一些示例,所述成本确定可以基于与相对不太重要的回声参考相比对相对更重要的回声参考进行更少压缩的方法。

[0028] 在一些示例中,所述方法可以涉及确定当前回声管理系统性能水平。根据一些示例,选择所述一个或多个所选回声参考可以至少部分地基于所述当前回声管理系统性能水平。

[0029] 根据一些示例,做出重要性估计可以涉及确定对应回声参考的重要性度量。在一些这样的示例中,确定重要性度量可以涉及确定对应回声参考的水平、确定对应回声参考的唯一性、确定对应回声参考的时间持续性、确定对应回声参考的可听度、或其一个或多个组合。

[0030] 在一些实例中,确定所述重要性度量可以至少部分地基于与音频设备布局相对应的数据或元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合(upmixing)矩阵、扩音器激活矩阵、或其一个或多个组合。

[0031] 根据一些示例,确定所述重要性度量可以至少部分地基于当前收听目标、当前环境噪声估计、所述至少一个回声管理系统的当前性能的估计、或其一个或多个组合。

[0032] 本文描述的一些或所有操作、功能和/或方法可以由一个或多个设备根据存储在一个或多个非暂态介质上的指令(例如,软件)来执行。这样的非暂态介质可以包括如本文描述的存储器设备等存储器设备,包括但不限于随机存取存储器(RAM)设备、只读存储器(ROM)设备等。因此,本公开中描述的主题的一些创新方面可以经由在其上存储有软件的一个或多个非暂态介质来实施。

[0033] 本公开的至少一些方面可以经由装置来实施。例如,一个或多个设备可以能够至少部分地执行本文公开的方法。在一些实施方式中,装置是或包括具有接口系统和控制系统的音频处理系统。控制系统可以包括一个或多个通用单芯片或多芯片处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其他可编程逻辑设备、离散门或晶体管逻辑、离散硬件部件或其组合。

[0034] 在以下附图和说明中阐述了本说明书中所描述的主题的一个或多个实施方式的细节。从所述描述、附图和权利要求中,其他特征、方面和优点将变得显而易见。注意,以下

附图的相对尺寸可能不是按比例来绘制的。

附图说明

- [0035] 在各个附图中,相似的附图标记和名称指示相似的要素。
- [0036] 图1A是示出了能够实施本公开的各个方面的装置的部件的示例的框图。
- [0037] 图1B示出了音频环境的示例。
- [0038] 图1C和图1D示出了音频设备110A-110C可以如何接收回放声道的示例。
- [0039] 图1E示出了音频环境的另一示例。
- [0040] 图2A呈现了能够执行至少一些所公开的实施方式的音频设备的框图。
- [0041] 图2B和图2C示出了音频环境中的音频设备的附加示例。
- [0042] 图3A呈现了示出根据一个示例的音频设备的部件的框图。
- [0043] 图3B和图3C是示出预期回声管理性能与用于回声管理的回声参考的数量的示例的图表。
- [0044] 图4呈现了示出根据一个示例的回声参考编排器的部件的框图。
- [0045] 图5A是概述所公开方法的一个示例的流程图。
- [0046] 图5B是概述了所公开方法的另一示例的流程图。
- [0047] 图6是概述所公开方法的一个示例的流程图。
- [0048] 图7示出了音频环境的平面图的示例,所述音频环境在该示例中是生活空间。

具体实施方式

[0049] 图1A是示出了能够实施本公开的各个方面的装置的部件的示例的框图。与本文提供的其他图一样,图1A所示的要素的类型和数量仅作为示例提供。其他实施方式可以包括更多、更少和/或不同类型和数量的要素。根据一些示例,装置50可以被配置用于执行本文公开的方法中的至少一些方法。在一些实施方式中,装置50可以是或者可以包括音频系统的一个或多个部件。例如,在一些实施方式中,装置50可以是音频设备,如智能音频设备。在其他示例中,装置50可以是移动设备(如蜂窝电话)、膝上型计算机、平板计算机设备、电视或其他类型的设备。

[0050] 根据一些替代性实施方式,装置50可以是或者可以包括服务器。在一些这样的示例中,装置50可以是或者可以包括编码器。因此,在一些实例中,装置50可以是被配置用于在如家庭音频环境的音频环境内使用的设备,然而在其他实例中,装置50可以是被配置用于在“云”中使用的设备,例如,服务器。

[0051] 在该示例中,装置50包括接口系统55和控制系统60。在一些实施方式中,接口系统55可以被配置用于与音频环境的一个或多个其他设备进行通信。在一些示例中,音频环境可以是家庭音频环境。在其他示例中,音频环境可以是另一种类型的环境,如办公室环境、汽车环境、火车环境、街道或人行道环境、公园环境等。在一些实施方式中,接口系统55可以被配置用于与音频环境的音频设备交换控制信息和相关联的数据。在一些示例中,控制信息和相关联的数据可以与装置50正执行的一个或多个软件应用程序有关。

[0052] 在一些实施方式中,接口系统55可以被配置用于接收内容流或用于提供内容流。内容流可以包括音频数据。音频数据可以包括但可以不限于音频信号。在一些实例中,音频

数据可以包括如声道数据和/或空间元数据等空间数据。例如,元数据可以由本文中可以被称为“编码器”的设备提供。在一些示例中,内容流可以包括视频数据和与视频数据相对应的音频数据。

[0053] 接口系统55可以包括一个或多个网络接口和/或一个或多个外部设备接口(如一个或多个通用串行总线(USB)接口)。根据一些实施方式,接口系统55可以包括一个或多个无线接口。接口系统55可以包括用于实施用户接口的一个或多个设备,如一个或多个麦克风、一个或多个扬声器、显示系统、触摸传感器系统和/或手势传感器系统。在一些示例中,接口系统55可以包括控制系统60与存储器系统(如图1A中示出的可选存储器系统65)之间的一个或多个接口。然而,在一些实例中,控制系统60可以包括存储器系统。在一些实施方式中,接口系统55可以被配置用于从环境中的一个或多个麦克风接收输入。

[0054] 例如,控制系统60可以包括通用单芯片或多芯片处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其他可编程逻辑设备、离散门或晶体管逻辑和/或离散硬件部件。

[0055] 在一些实施方式中,控制系统60可以驻留在多于一个设备中。例如,在一些实施方式中,控制系统60的一部分可以驻留在本文描绘的环境之一内的设备中,并且控制系统60的另一部分可以驻留在环境之外的设备中,如服务器、移动设备(例如,智能电话或平板计算机)等。在其他示例中,控制系统60的一部分可以驻留在本文描绘的环境之一内的设备中,并且控制系统60的另一部分可以驻留在环境的一个或多个其他设备中。例如,控制系统的功能可以跨环境的多个智能音频设备分布,或者可以由编排设备(如本文中可以被称为智能家居中枢的设备)和环境的一个或多个其他设备共享。在其他示例中,控制系统60的一部分可以驻留在实施基于云的服务的设备(如服务器)中,并且控制系统60的另一部分可以驻留在实施基于云的服务的另一设备(如另一服务器、存储器设备等)中。在一些示例中,接口系统55还可以驻留在多于一个设备中。

[0056] 在一些实施方式中,控制系统60可以被配置用于至少部分地执行本文公开的方法。根据一些示例,控制系统60可以被配置为获得多个回声参考。多个回声参考可以包括针对音频环境中的多个音频设备中的每个音频设备的至少一个回声参考。每个回声参考可以例如对应于由多个音频设备中的一个音频设备的一个或多个扩音器回放的音频数据。

[0057] 在一些实施方式中,控制系统60可以被配置为对多个回声参考中的每个回声参考做出重要性估计。在一些示例中,做出重要性估计可以涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻的预期贡献。(多个)回声管理系统可以包括声学回声消除器(AEC)和/或声学回声抑制器(AES)。

[0058] 根据一些示例,控制系统60可以被配置为至少部分地基于重要性估计来选择一个或多个所选回声参考。在一些示例中,控制系统60可以被配置为将一个或多个所选回声参考提供给至少一个回声管理系统。

[0059] 本文描述的一些或所有方法可以由一个或多个设备根据存储在一个或多个非暂态介质上的指令(例如,软件)来执行。这样的非暂态介质可以包括如本文描述的存储器设备等存储器设备,包括但不限于随机存取存储器(RAM)设备、只读存储器(ROM)设备等。一个或多个非暂态介质可以例如驻留在图1A中示出的可选存储器系统65和/或控制系统60中。因此,可以在其上存储有软件的一个或多个非暂态介质中实施本公开中所描述的主题的各

个创新方面。例如,所述软件可以包括用于控制至少一个设备执行本文公开的一些或所有方法的指令。例如,软件可以由如图1A的控制系统60等控制系统的一个或多个部件执行。

[0060] 在一些示例中,装置50可以包括图1A中示出的可选麦克风系统70。可选麦克风系统70可以包括一个或多个麦克风。根据一些示例,可选麦克风系统70可以包括麦克风阵列。在一些示例中,麦克风阵列可以被配置成例如根据来自控制系统60的指令来确定到达方向(DOA)和/或到达时间(TOA)信息。在一些实例中,麦克风阵列可以被配置用于例如根据来自控制系统60的指令来进行接收侧波束成形。在一些实施方式中,一个或多个麦克风可以是另一设备(如扬声器系统的扬声器、智能音频设备等)的一部分或与其相关联。在一些示例中,装置50可以不包括麦克风系统70。然而,在一些这样的实施方式中,装置50仍然可以被配置为经由接口系统60接收音频环境中的一个或多个麦克风的麦克风数据。在一些这样的实施方式中,装置50的基于云的实施方式可以被配置成经由接口系统60从音频环境中的一个或多个麦克风接收麦克风数据或与麦克风数据相对应的数据。

[0061] 根据一些实施方式,装置50可以包括图1A中示出的可选扩音器系统75。可选扩音器系统75可以包括一个或多个扩音器,所述扩音器在本文中也可以被称为“扬声器”,或更通常地被称为“音频再现换能器”。在一些示例(例如,基于云的实施方式)中,装置50可以不包括扩音器系统75。

[0062] 在一些实施方式中,装置50可以包括图1A中示出的可选传感器系统80。可选传感器系统80可以包括一个或多个触摸传感器、手势传感器、运动检测器等。根据一些实施方式,可选传感器系统80可以包括一个或多个相机。在一些实施方式中,相机可以是独立式相机。在一些示例中,可选传感器系统80的一个或多个相机可以驻留在智能音频设备中,所述智能音频设备可以是单一用途音频设备或虚拟助理。在一些这样的示例中,可选传感器系统80的一个或多个相机可以驻留在电视、移动电话或智能扬声器中。在一些示例中,装置50可以不包括传感器系统80。然而,在一些这样的实施方式中,装置50仍然可以被配置为经由接口系统60接收音频环境中的一个或多个传感器的传感器数据。

[0063] 在一些实施方式中,装置50可以包括图1A中示出的可选显示系统85。可选显示系统85可以包括一个或多个显示器,如一个或多个发光二极管(LED)显示器。在一些实例中,可选显示系统85可以包括一个或多个有机发光二极管(OLED)显示器。在一些示例中,可选显示系统85可以包括智能音频设备的一个或多个显示器。在其他示例中,可选显示系统85可以包括电视显示器、膝上型计算机显示器、移动设备显示器、或另一种类型的显示器。在装置50包括显示系统85的一些示例中,传感器系统80可以包括接近显示系统85的一个或多个显示器的触摸传感器系统和/或手势传感器系统。根据一些这样的实施方式,控制系统60可以被配置用于控制显示系统85来呈现一个或多个图形用户界面(GUI)。

[0064] 根据一些这样的示例,装置50可以是或者可以包括智能音频设备。在一些这样的实施方式中,装置50可以是或者可以包括唤醒词检测器。例如,装置50可以是或者可以包括虚拟助理。

[0065] 对于立体声或单声道的回放媒体,传统上它是经由物理线缆连接到音频播放器(例如,CD/DVD播放器、电视(TV)等)的一对扬声器渲染到音频环境(例如,生活空间、汽车、办公空间等)中的。随着智能扬声器的流行,用户通常在其家中(或其他音频环境)中拥有多于两个能够回放音频的被配置为无线通信的音频设备(其可以包括但不限于智能扬声器或

其他智能音频设备)。

[0066] 智能扬声器通常被配置为根据语音命令进行操作。因此,这种智能扬声器通常被配置为连续收听唤醒词,唤醒词之后将通常跟着语音命令。任何连续收听任务(比如等待唤醒词或执行任何类型的“连续校准”)将优选地在内容回放(比如音乐回放、电影和电视节目的音轨回放等)以及发生设备交互时(例如,在电话通话期间)继续运行。需要在内容回放期间收听的音频设备通常需要采用某种形式的回声管理,例如回声消除和/或回声抑制,以从麦克风信号中去除“回声”(由设备播放的内容)。

[0067] 图1B示出了音频环境的示例。与本文提供的其他图一样,图1B中示出的要素的类型、数量和布置仅作为示例提供。其他实施方式可以包括更多、更少和/或不同类型、数量和/或布置的要素。

[0068] 根据该示例,音频环境100包括音频设备110A、110B和110C。在该示例中,音频设备110A-110C中的每一个是图1A的装置50的实例,并且包括麦克风系统70和扩音器系统75的实例,但这些在图1B中未示出。根据一些示例,每个音频设备110A-110C可以是智能音频设备,如智能扬声器。

[0069] 在该示例中,音频设备110A-110C在人130正在说话的同时回放音频内容。音频设备110B的麦克风不仅检测由其自身的扬声器回放的音频内容,而且还检测人130的语音声音131以及由音频设备110A和110C回放的音频内容。

[0070] 为了同时利用尽可能多的扬声器,典型的方法是让音频环境中的所有音频设备回放相同的内容,并使用某种定时机制来使回放媒体保持同步。这样做的优点是使分发变得简单,因为所有设备都会收到相同的回放媒体副本,无论是下载或流式传输到每个音频设备,还是由一个设备广播并多播到所有音频设备。

[0071] 这种方法的一个主要缺点是无法获得空间效果。空间效果可以通过添加更多回放声道(例如每个扬声器一个)来实现,例如通过上混合。在一些示例中,空间效果可以经由诸如质心振幅平移(CMAP)、灵活虚拟化(FV)、或CMAP和FV的组合等灵活渲染过程来实现。CMAP、FV及其组合的相关示例描述于国际专利公开号WO 2021/021707A1(例如,第25-41页)中,该专利特此通过引用并入。

[0072] 图1C和图1D示出了音频环境中的音频设备的附加示例。根据这些示例,音频环境100包括智能家居中枢105和音频设备110A、110B和110C。在这些示例中,智能家居中枢105和音频设备110A-110C是图1A的装置50的实例。根据这些示例,音频设备110A-110C中的每一个包括对应的一个扩音器121A、121B和121C。根据一些示例,每个音频设备110A-110C可以是智能音频设备,如智能扬声器。

[0073] 图1C和图1D示出了音频设备110A-110C可以如何接收回放声道的示例。在图1C中,已编码的音频比特流被多播到所有音频设备110A-110C。在图1D中,音频设备110A-110C中的每一个仅接收该音频设备进行回放所需的声道。比特流分发的选择可以根据单独的实施方式而变化,并且可以例如基于可用的系统带宽、所使用的音频编解码器的编解码效率、音频设备110A-110C的能力和/或其他因素。图1C和图1D中所示的音频环境的确切拓扑并不重要。然而,这些示例说明了这样一个事实:将音频声道分发给设备音频设备将产生一些成本。成本可以从所需的网络带宽、对音频声道进行编码解码所增加的计算成本等方面进行评估。

[0074] 图1E示出了音频环境的另一示例。根据该示例,音频环境100包括音频设备110A、110B、110C和110D。在该示例中,音频设备110A-110D中的每一个是图1A的装置50的实例并且包括至少一个麦克风(参见麦克风120A、120B、120C和120D)、至少一个扩音器(参见扩音器121A、121B、121C和121D)。根据一些示例,每个音频设备110A-110D可以是智能音频设备,如智能扬声器。

[0075] 在该示例中,音频设备110A-110D经由扩音器121A-121D渲染内容122A、122B、122C和122D。麦克风120A-120D中的每一个检测到与由音频设备110A-110D中的每一个回放的内容122A-122D相对应的“回声”。在该示例中,音频设备110A-110D被配置为收听来自音频环境100内的人130的语音131中的命令或唤醒词。

[0076] 图2A呈现了能够执行至少一些所公开的实施方式的音频设备的框图。与本文提供的其他图一样,图2A中示出的要素的类型、数量和布置仅作为示例提供。其他实施方式可以包括更多、更少和/或不同类型、数量和/或布置的要素。在该示例中,音频设备110A是图1E的音频设备110A的实例。在此,音频设备110A包括控制系统60a,其是图1A的控制系统60的实例。根据该实施方式,控制系统60能够在存在与由音频环境100中的每个音频设备回放的内容122A、122B、122C和122D相对应的回声的情况下收听人130的语音131。

[0077] 根据该示例,控制系统60实施渲染器201A、多声道声学回声管理系统(MC-EMS)203A和语音处理块240A。MC-EMS203A可以包括声学回声消除器(AEC)、声学回声抑制器(AES)、或者AEC和AES两者,具体取决于特定实施方式。根据该示例,语音处理块240A被配置为检测用户的唤醒词和命令。在一些实施方式中,语音处理块240A可以被配置为支持通信会话,比如电话通话。

[0078] 在该实施方式中,渲染器201A被配置为向MC-EMS203A提供本地回声参考220A。本地回声参考220A对应于(并且在该示例中等同于)提供给扩音器121A以供音频设备110A回放的扬声器馈送信号。根据该示例,渲染器201A还被配置为向MC-EMS203A提供非本地回声参考221A(对应于由音频环境100中的其他音频设备回放的内容122B、122C和122D)。

[0079] 根据一些示例,音频设备110A接收包括图1E的所有音频设备110A-110D的音频数据的组合比特流(例如,如图1C所示)。在一些这样的示例中,渲染器201A可以被配置为将本地回声参考220A与非本地回声参考221A分开,以向扩音器121A提供本地回声参考220A,并向MC-EMS203A提供本地回声参考220A和非本地回声参考221A。在一些替代示例中,音频设备110A可以接收仅旨在音频设备110A上回放的比特流,例如,如图1D所示。在一些这样的示例中,智能家居中枢105(或其他音频设备110B-D)可以向音频设备110A提供非本地回声参考221A,如图2A中附图标记221A旁边的虚线箭头所示。

[0080] 在一些实例中,本地回声参考220A和/或非本地回声参考221A可以是提供给扩音器121A-121D以供回放的扬声器馈送信号的全保真度复制品。在一些替代示例中,本地回声参考220A和/或非本地回声参考221A可以是提供给扩音器121A-121D以供回放的扬声器馈送信号的较低保真度表示。在一些这样的示例中,非本地回声参考221A可以是提供给扩音器121B-121D以供回放的扬声器馈送信号的下采样版本。根据一些示例,非本地回声参考221A可以是提供给扩音器121B-121D以供回放的扬声器馈送信号的有损压缩。在一些示例中,非本地回声参考221A可以是与提供给扩音器121B-121D以供回放的扬声器馈送信号相对应的分段功率信息(banded power information)。

[0081] 根据该实施方式,MC-EMS203A被配置为使用本地回声参考220A和非本地回声参考221A来预测并消除和/或抑制来自麦克风信号223A的回声,从而产生残差信号224A,在该残差信号中,语音回声比(SER)相对于麦克风信号223A可能已经得到改善。该残差信号224A可以使得语音处理块240A能够检测用户唤醒词和命令。在一些实施方式中,语音处理块240A可以被配置为支持通信会话,比如电话通话。

[0082] 本公开的一些方面涉及对多个回声参考中的每个回声参考(例如,对本地回声参考220A和非本地回声参考221A)做出重要性估计。做出重要性估计可以涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻(例如,音频设备110A的MC-EMS203A进行的回声减轻)的预期贡献。下文中提供了各种示例。

[0083] 在分布式和编排式设备的背景下,出于回声管理的目的,根据一些示例,除了自身的回声参考之外,每个音频设备还可以获得与音频环境中一个或多个其他音频设备回放的内容相对应的回声参考。将特定回声参考包括在本地回声管理系统或“EMS”(比如音频设备110A的MC-EMS203A)中的影响可以根据多个参数而变化,诸如正在播出的音频内容的多样性、用于传输回声参考所需的网络带宽、在传输已编码回声参考的情况下用于编码回声参考的编码计算要求、用于解码回声参考的解码计算要求、用于由回声管理系统使用回声参考的回声管理系统计算要求、音频设备的相对可听度等。

[0084] 例如,如果每个音频设备正在渲染相同的内容(换句话说,如果正在回放单声道音频),则向EMS提供附加参考几乎没有(尽管非零)益处。此外,由于实际限制(比如带宽受限的网络),可能不希望所有设备都共享其本地回声参考的复制品。因此,一些实施方式可以提供分布式和编排式EMS(DOEMS),其中,对回声参考进行优先级排序并相应地传输(或不传输)。一些这样的示例可以实施每个附加回声参考的成本(例如,所需的网络带宽和/或所需的计算开销)与效益(例如,预期的回声减轻改进,其可以根据信号回声比(SER)和/或回声损失增强(ERLE)来度量)之间的权衡。

[0085] 图2B和图2C示出了音频环境中的音频设备的附加示例。根据这些示例,音频环境100包括智能家居中枢105和音频设备110A、110B和110C。在这些示例中,智能家居中枢105和音频设备110A-110C是图1A的装置50的实例。根据这些示例,音频设备110A-110C中的每一个包括对应的一个麦克风120A、120B和120C以及对应的一个扩音器121A、121B和121C。根据一些示例,每个音频设备110A-110C可以是智能音频设备,如智能扬声器。

[0086] 在图2B中,智能家居中枢105将相同的已编码音频比特流发送到所有音频设备110A-110C。在图2C中,智能家居中枢105仅发送每个音频设备110A-110C进行回放所需的音频声道。在这两个示例中,音频声道0旨在用于在音频设备110A上回放,音频声道1旨在用于在音频设备110B上回放并且音频声道2旨在用于在音频设备110C上回放。

[0087] 图2B和图2C示出了在本地网络上共享回声参考数据的示例。在这些示例中,音频设备110A通过本地网络向音频设备110B和110C发送回声参考220A',该回声参考是与音频设备110A的扩音器回放相对应的回声参考。在这些示例中,回声参考220A'与在比特流中找到的声道0音频不同。在一些实例中,回声参考220A'可能不同于声道0音频,因为在音频设备110A上实施了回放后处理。在图2C所示的示例中,不是将组合比特流提供给所有音频设备110A-110C,因此另一设备(诸如音频设备110A或智能家居中枢105)提供回声参考220A'。在图2B中描绘的场景中,即使组合比特流被提供给所有音频设备110A-110C,在一些这样的

实例中,可能仍然需要传输回声参考220A'。

[0088] 在其他示例中,回声参考220A'可能不同于声道0音频,因为回声参考220A'可能不是在音频设备110A上回放的音频数据的全保真度复制品。在一些这样的示例中,回声参考220A'可以对应于在音频设备110A上回放的音频数据,但是可能需要比完整复制品相对较少的数据,并且因此当传输回声参考220A'时可以消耗相对较少的本地网络带宽。

[0089] 根据一些这样的示例,音频设备110A可以被配置为产生上文参考图2A描述的本地回声参考220A的下采样版本。在一些这样的示例中,回声参考220A'可以是或可以包括下采样版本。

[0090] 在一些示例中,音频设备110A可以被配置为对本地回声参考220A进行有损压缩。在这种实例中,回声参考220A'可以是控制系统60a对本地回声参考220A应用有损压缩算法的结果。

[0091] 根据一些示例,音频设备110A可以被配置为向音频设备110B和110C提供与本地回声参考220A相对应的分段功率信息。在一些这样的示例中,代替传输在音频设备110A上回放的音频数据的全保真度复制品,控制系统60a可以被配置为确定在音频设备110A上回放的音频数据的多个频带中的每个频带中的功率水平,并将对应的分段功率信息传输到音频设备110B和110C。在一些这样的示例中,回声参考220A'可以是或者可以包括分段功率信息。

[0092] 图3A呈现了示出根据一个示例的音频设备的部件的框图。与本文提供的其他图一样,图3A中示出的要素的类型、数量和布置仅作为示例提供。其他实施方式可以包括更多、更少和/或不同类型、数量和/或布置的要素。例如,一些实施方式可以被配置为发送和/或接收“原始”回声参考(其可以是在音频设备上再现的音频的完整的全保真度复制品)、在音频设备上再现的音频的低保真度版本或表示(比如下采样版本、通过有损压缩产生的版本、或与在音频设备上再现的音频相对应的分段功率信息),但不同时发送和/或接收原始版本和低保真度版本。

[0093] 在该示例中,音频设备110A是图1E的音频设备110A的实例并且包括控制系统60a,该控制系统是图1A的控制系统60的实例。根据该示例,控制系统60a被配置为实施渲染器201A、多声道声学回声管理系统(MC-EMS)203A、语音处理块240A、回声参考编排器302A、解码器303A以及噪声估计器304A。读者可以假设MC-EMS203A和语音处理块240A如上文参考图2A所描述的那样起作用,除非图3A的以下描述另有指示。在该示例中,网络接口301A是上文参考图1A描述的接口系统55的实例。

[0094] 在该示例中,图3A的要素如下:

[0095] 110A: 音频设备;

[0096] 120A: 代表性麦克风。在一些实施方式中,音频设备110A可以具有多于一个麦克风;

[0097] 121A: 代表性扩音器。在一些实施方式中,音频设备110A可以具有多于一个扩音器;

[0098] 201A: 渲染器,产生针对本地回放的参考和模拟由音频环境中的其他音频设备回放的音频的回声参考;

[0099] 203A: 多声道声学回声管理系统(MC-EMS),其可以包括声学回声消除器(AEC)和/

或声学回声抑制器(AES)；

[0100] 220A:用于回放和消除的本地回声参考；

[0101] 221A:一个或多个非本地音频设备(音频环境中的一个或多个其他音频设备)正在播放的回声参考的本地产生的副本；

[0102] 223A:多个麦克风信号；

[0103] 224A:多个残差信号(MC-EMS203A消除和/或抑制所预测的回声之后的麦克风信号)；

[0104] 240A:语音处理块,被配置用于唤醒词检测、语音命令检测和/或提供电话通信；

[0105] 301A:网络接口,被配置用于音频设备之间的通信,其也可以被配置用于经由因特网和/或经由一个或多个蜂窝网络进行通信；

[0106] 302A:回声参考编排器,被配置为对回声参考进行排名并选择一个或多个回声参考的适当集合；

[0107] 303A:音频解码器块；

[0108] 304A:噪声估计器块；

[0109] 310A:由音频设备110A从音频环境中的一个或多个其他设备接收的一个或多个已解码的回声参考；

[0110] 311A:从一个或多个其他设备(比如智能家居中枢或音频设备110B-110D中的一个或多个)通过本地网络发送回声参考的请求；

[0111] 312A:元数据,其可以是或可以包括与音频设备布局相对应的元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合矩阵、和/或扩音器激活矩阵；

[0112] 313A:由回声参考编排器302A选择的回声参考；

[0113] 314A:设备110A从一个或多个其他设备接收的回声参考；

[0114] 315A:从设备110A发送到其他设备的回声参考；

[0115] 316A:设备110A从音频环境的一个或多个其他设备接收的原始回声参考；

[0116] 317A:设备110A从音频环境的一个或多个其他设备接收的回声参考的低保真度(例如,编解码的)版本；

[0117] 318A:音频环境噪声估计；

[0118] 350A:指示MC-EMS203A的当前性能的一个或多个指标,其可以是或者可以包括自适应滤波器系数数据或其他AEC统计数据、语音回声(SER)比数据等。

[0119] 取决于特定实施方式,回声参考编排器302A可以以各种方式起作用。本文中公开了许多示例。在一些示例中,回声参考编排器302A可以被配置为对多个回声参考中的每个回声参考(例如,对本地回声参考220A和非本地回声参考221A)做出重要性估计。做出重要性估计可以涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻(例如,音频设备110A的MC-EMS203A进行的回声减轻)的预期贡献。

[0120] 做出重要性估计的一些示例可以涉及确定重要性度量。在一些这样的示例中,重要性度量可以至少部分地基于每个回声参考的一个或多个特性,比如水平、唯一性、时间持续性、可听度、或其一个或多个组合。在一些示例中,重要性度量可以至少部分地基于元数据(例如,元数据312A),诸如与音频设备布局相对应的元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合矩阵、扩音器激活矩阵、或其一个或多个组合。在一些示

例中,重要性度量可以至少部分地基于当前收听目标、当前环境噪声估计、至少一个回声管理系统的当前性能的估计、或其一个或多个组合。

[0121] 根据一些示例,回声参考编排器302A可以被配置为至少部分地基于成本确定来选择的一个或多个回声参考的集合。在一些示例中,回声参考编排器302A可以被配置为进行成本确定,而在其他示例中,控制系统60a的另一个块可以被配置为进行成本确定。在一些实例中,成本确定可以涉及确定多个回声参考中的至少一个回声参考的成本,或者在一些情况下确定多个回声参考中的每一个的成本。在一些示例中,成本确定可以基于用于传输回声参考所需的网络带宽、用于编码至少一个回声参考的编码计算要求、用于解码至少一个回声参考的解码计算要求、制作回声参考的下采样版本的下采样成本、由回声管理系统使用至少一个回声参考的回声管理系统计算要求、或其一个或多个组合。

[0122] 根据一些示例,成本确定可以基于至少一个回声参考在时域或频域中的复制品、至少一个回声参考的下采样版本、至少一个回声参考的有损压缩、至少一个回声参考的分段功率信息、或其一个或多个组合。在一些实例中,成本确定可以基于与相对不太重要的回声参考相比对相对更重要的回声参考进行更少压缩的方法。在一些实施方式中,回声参考编排器302A(或控制系统60a的另一个块)可以被配置为确定当前回声管理系统性能水平(例如,至少部分地基于(多个)指标350A)。在一些这样的示例中,选择一个或多个所选回声参考可以至少部分地基于当前回声管理系统性能水平。

[0123] 取决于分布式音频设备系统、其配置和音频会话的类型(例如,通信或收听音乐)和/或经渲染的内容的性质,估计每个回声参考的重要性的速率和评估回声参考集的速率可能不同。此外,估计重要性的速率不需要等于回声参考选择过程做出决策的速率。如果这两者不同步,则在一些示例中,重要性计算将会更加频繁。在一些实例中,回声参考选择可以是离散过程,在该过程中,做出包括或不包括特定回声参考的二元决策。

[0124] 图3B和图3C是示出预期回声管理性能与用于回声管理的回声参考的数量的示例的图表。在图3B中,可以发现,随着添加附加参考,预期的回声性能也会提升。然而,在该示例中,可以发现,只有系统可以在其上操作的几个离散点。在一些示例中,图3B中所示的点可以对应于处理每个回声参考的完整的、全保真度复制品。例如,点301可以对应于处理本地回声参考(例如,图2A或图3A的本地参考220A)的实例,并且点310可以对应于接收第一非本地回声参考的完整复制品(例如,图3A的所接收回声参考314A之一的全保真度版本,其可能已被选为最重要的非本地回声参考)并处理本地回声参考和第一非本地回声参考的完整复制品两者的实例。

[0125] 图3C图示了在图3B所示的离散操作点中的任意两个之间操作的一个示例。连接图3B中的点的线可以例如对应于一定的回声参考保真度范围,包括每个回声参考的较低保真度版本或表示。例如,点303、305和307可以对应于第一非本地回声参考的保真度水平增加的副本或表示,其中,点303对应于最低保真度表示,并且点307对应于除全保真度复制品之外的最高保真度表示。在一些示例中,点303可以对应于第一非本地回声参考的分段功率信息。根据一些示例,点305和307可以分别对应于第一非本地回声参考的相对较高有损压缩和第一非本地回声参考的相对较少有损压缩。

[0126] 回声参考的副本或表示的保真度通常与每个这样的副本或表示所需的比特数成反比。因此,回声参考的副本或表示的保真度提供了网络成本(由于传输所需的比特数不

同)和预期回声管理性能(因为性能应随着保真度增加而提高)之间的权衡的指示。注意,用于连接图3C中的点的直线仅表示许多不同的可能轨迹之一,部分原因是从一个回声参考到下一个回声参考的增量变化取决于将选择哪个回声参考作为下一个回声参考,并且部分原因是预期回声管理性能与保真度之间可能不存在线性关系。

[0127] 图4呈现了示出根据一个示例的回声参考编排器的部件的框图。与本文提供的其他图一样,图4中示出的要素的类型、数量和布置仅作为示例提供。其他实施方式可以包括更多、更少和/或不同类型、数量和/或布置的要素。例如,一些实施方式可以被配置为发送和/或接收“原始”回声参考(其可以是在音频设备上再现的音频的全保真度复制品)、在音频设备上再现的音频的低保真度版本或表示(比如下采样版本、通过有损压缩产生的版本、或与在音频设备上再现的音频相对应的分段功率信息),但不同时发送和/或接收原始版本和低保真度版本。

[0128] 在该示例中,回声参考编排器302A是图3A的回声参考编排器302A的实例并且由图3A的控制系统60a的实例实施。根据该示例,图4的要素如下:

[0129] 220A:用于回放和消除的本地回声参考;

[0130] 221A:音频环境的另一音频设备正在播放的非本地回声参考的本地产生的副本;

[0131] 302A:回声参考编排器,被配置为对一个或多个回声参考的集合进行排名和选择的模块;

[0132] 310A:由音频设备110A从音频环境中的一个或多个其他设备接收的一个或多个已解码的回声参考;

[0133] 311A:从音频环境的一个或多个其他设备通过本地网络发送回声参考的请求;

[0134] 312A:元数据,其可以是或可以包括与音频设备布局相对应的元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合矩阵、和/或扩音器激活矩阵;

[0135] 313A:在该示例中,由回声参考编排器302A选择并发送到MC-EMS203A的一个或多个回声参考的集合;

[0136] 316A:设备110A从音频环境的一个或多个其他设备接收的原始回声参考;

[0137] 317A:设备110A从音频环境的一个或多个其他设备接收的回声参考的低保真度(例如,编解码的)版本;

[0138] 318A:音频环境噪声估计;

[0139] 350A:指示MC-EMS203A的当前性能的一个或多个指标,其可以是或者可以包括自适应滤波器系数数据或其他AEC统计数据、语音回声(SER)比数据等。

[0140] 401A:回声参考重要性估计器,其被配置为估计每个回声参考的预期重要性,并且在该示例中生成对应的重要性度量420A;

[0141] 402:回声参考选择器,其被配置为在该示例中至少部分地基于当前收听目标(如421A所示)、每个回声参考的成本(如422A所示)、EMS的当前状态/性能(如350A所示)以及每个候选回声参考的估计重要性(如重要性度量420A所示)来选择回声参考集313A;

[0142] 403A:成本估计模块,其被配置为确定将回声参考包括在回声参考集313A中的(多个)成本(例如,计算和/或网络成本);

[0143] 404A:可选模块,用于确定或估计音频设备110A的当前收听目标;

[0144] 405A:被配置为实施一个或多个MC-EMS性能模型的模块,其在一些示例中可以产

生诸如图3B或图3C中所示的数据；

[0145] 420A:由回声参考重要性估计器401A生成的重要性度量420A；

[0146] 421A:指示当前收听目标的信息；

[0147] 422A:指示将回声参考包括在回声参考集313A中的(多个)成本的信息；以及

[0148] 423A:由MC-EMS性能模型405A产生的信息,其在一些示例中可以是或者包括诸如图3B或图3C中所示的数据。

[0149] 取决于特定实施方式,回声参考重要性估计器401A可以以各种方式起作用。本公开中提供了各种示例。在一些示例中,回声参考重要性估计器401A可以被配置为对多个回声参考中的每个回声参考(例如,对本地回声参考220A和非本地回声参考221A)做出重要性估计。做出重要性估计可以涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻(例如,音频设备110A的MC-EMS203A进行的回声减轻)的预期贡献。

[0150] 在该示例中,做出重要性估计涉及确定重要性度量420A。重要性度量420A可以至少部分地基于每个回声参考的一个或多个特性,比如水平、唯一性、时间持续性、可听度、或其一个或多个组合。在一些示例中,重要性度量可以至少部分地基于元数据(例如,元数据312A),该元数据可以包括与音频设备布局相对应的元数据、扩音器元数据(例如,声压级(SPL)评级、频率范围、扩音器是否是向上发声的扩音器等)、与接收到的音频数据相对应的元数据(例如,位置元数据、指示人声或其他语音的元数据等)、上混合矩阵、扩音器激活矩阵、或其一个或多个组合。在一些实例中,如虚线箭头420A所示,回声参考重要性估计器401A可以将重要性度量420A提供给MC-EMS性能模型405A。

[0151] 根据该示例,重要性度量420A至少部分地基于当前收听目标,如信息421A所指示的。如下文更详细描述,当前收听目标可以显著改变对诸如水平、唯一性、时间持续性、可听度等因素的评估方式。例如,在电话通话期间的重要性分析可能与等待唤醒词时截然不同。

[0152] 在该示例中,重要性度量420A至少部分地基于当前环境噪声估计318A、指示MC-EMS203A的当前性能的(多个)指标350A、由MC-EMS性能模型405A产生的信息423A、或其一个或多个组合。在一些实施方式中,回声参考重要性估计器401A可以确定,如果房间噪声水平相对较高(如当前环境噪声估计318A所指示的),则添加回声参考将不太可能有助于显著减轻回声。如上所述,信息423A可以对应于上文参考图3B和图3C描述的信息类型,其可以提供回声参考的使用与MC-EMS203A的预期性能增加之间的直接相关性。如下文更详细描述,EMS的性能可以部分地基于当受到音频环境中的噪声干扰时EMS的稳健性。

[0153] 根据该实施方式,回声参考选择器402至少部分地基于以下各项来选择一个或多个回声参考的集合:指示MC-EMS203A的当前性能的一个或多个指标350A、重要性度量420A、当前收听目标421A、指示将回声参考包括在回声参考集313A中的(多个)成本的信息422A、以及由MC-EMS性能模型405A产生的信息423A。下文提供了回声参考选择器402可以如何选择回声参考的一些详细示例。

[0154] 在该示例中,成本估计模块403A被配置为确定将回声参考包括在回声参考集313A中的计算和/或网络成本。计算成本可以例如包括由MC-EMS203A使用特定回声参考的附加计算成本。该计算成本进而可能取决于表示回声参考所需的比特数。在一些示例中,计算成

本可以包括有损回声参考编码过程的计算成本和/或对应回声参考解码过程的计算成本。确定网络成本可以涉及确定跨本地数据网络(例如,本地无线数据网络)发送回声参考的完整复制品或者回声参考的副本或表示所需的数据量。

[0155] 在一些实例中,回声参考选择块402A可以生成并传输使音频环境中的另一设备通过网络向其发送一个或多个回声参考的请求311A。(图3A的要素314A指示由音频设备110A接收一个或多个回声参考,其在一些实例中可能已经响应于请求311A)。在一些示例中,请求311A可以指定所请求的回声参考的保真度,例如,是否应当发送回声参考的“原始”副本(全保真度复制品)、是否应当发送回声参考的已编码版本、在应当发送回声参考的已编码版本的情况下应当将相对较多还是相对较少的有损压缩算法应用于回声参考、是否应当发送与回声参考相对应的分段功率信息等。

[0156] 人们可能注意到,对已编码回声参考的请求不仅由于发送请求和参考而引入网络成本,而且还增加了(多个)响应设备(例如,智能家居中枢105或音频设备110B-110D中的一个或多个)必须对参考进行编码的计算成本,以及音频设备110A对接收到的参考进行解码的计算成本。然而,该编码成本可能是一次性成本。因此,从一个音频设备到另一音频设备通过网络发送已编码参考的请求改变了在其他设备(例如,在音频设备402C和402D)中执行的潜在性能/成本权衡。

[0157] 在一些实施方式中,回声参考编排器302A的一个或多个块可以由编排设备(例如,智能家居中枢105或音频设备110A-110D之一)来执行。根据一些这样的实施方式,回声参考重要性估计器401A和/或回声参考选择块402A的至少一些功能可以由编排设备来执行。一些这样的实施方式可能能够考虑到音频环境中MC-EMS的所有实例的性能增强、MC-EMS的所有实例的总体计算需求、本地网络的总体需求、和/或所有编码器和解码器的总体计算需求而确定整个系统的成本/效益权衡。

[0158] 各种指标和分量的示例

[0159] 重要性度量

[0160] 简单地说,重要性度量(本文中可称为“重要性”或“I”)可以是由于包含特定回声参考而对EMS性能的预期改进的度量。在一些实施例中,重要性可以取决于EMS的当前状态,特别是取决于已经在使用的回声参考集以及它们正在以什么保真度水平被接收。取决于特定的实施方式,重要性可以在不同的时间尺度上获得。在一个极端情况下,重要性可以逐帧地实施(例如,根据每一帧的重要性信号)。在其他示例中,重要性可以被实施为针对内容片段持续时间的恒定值,或者被实施为针对使用音频设备的特定配置的时间的恒定值。音频设备的配置可以对应于音频设备位置和/或音频设备取向。

[0161] 因此,可以取决于特定的实施方式在各种时间尺度上计算重要性度量,例如:

[0162] • 实时,例如,根据音频环境中的事件(例如来电通话)等来分析当前的音频内容;

[0163] • 在较长的时间尺度上,例如,逐个音轨地,其中,音轨对应于诸如歌曲或可以例如在几分钟的时间尺度上持续的其他音乐内容片段等内容片段;或者

[0164] • 仅一次,例如,当音频系统最初配置或重新配置时。

[0165] 可以在与评估重要性度量的时间尺度相似(或更慢)的时间尺度上做出关于出于回声管理的目的而选择哪些回声参考的决策。例如,设备或系统可能每30秒估计一次重要性,并每几分钟做出关于改变所选回声参考的决策。

[0166] 根据一些示例,控制系统可以被配置为确定重要性矩阵,其可以包括当前音频设备系统的所有重要性信息。在一些这样的示例中,重要性矩阵可以具有维度 $N \times M$,包括每个音频设备的条目和每个潜在回声参考声道的条目。在一些这样的示例中, N 表示音频设备的数量,并且 M 表示潜在回声参考的数量。由于一些音频设备可能会回放多于一个声道,因此这种类型的重要性矩阵并不总是方形的。

[0167] 在一些实施方式中,重要性度量 I 可以基于以下各项中的一项或多项:

[0168] • L :回声参考的水平;

[0169] • U :回声参考的唯一性;

[0170] • P :回声参考的时间持续性,和/或

[0171] • A :渲染回声参考的设备的可听度。

[0172] 如本文所使用的,首字母缩略词“LUPA”总体上指代可以从中确定重要性度量的回声参考特性,包括但不限于 L 、 U 、 P 和/或 A 中的一项或多项。

[0173] L 或“水平”方面

[0174] 该方面描述了回声参考的水平或响度。在其他条件相同的情况下,众所周知,回放信号越响,对EMS性能的影响越大。如本文所使用的,术语“水平”是指音频信号的数字表示内的水平,而不一定是指音频信号在经由扩音器再现之后的实际声压级。在一些示例中,回声参考的单个声道的响度可以基于均方根(RMS)指标或LKFS(相对于满刻度的 k 加权响度)指标。这样的指标很容易在回声参考上实时计算,或者可以作为比特流中的元数据存在。根据一些实施方式, L 可以根据音量设置来确定,比如音频系统音量设置或媒体应用内的音量设置。

[0175] U 或“唯一性”方面

[0176] 唯一性方面旨在捕获特定回声参考所提供的有关整体音频呈现的新信息量。从统计的角度来看,多声道音频呈现通常跨声道包含冗余。例如,这种冗余的出现可能是由于乐器和其他声源在房间左右两侧的声道上被复制,或者信号被平移并因此同时在多个活动扩音器中进一步复制。尽管这种场景导致EMS需要解决超标的问题(其中,回声滤波器可能从多个回声路径来推断观察结果),但在实践中仍然可以观察到一些益处和更高的性能。

[0177] U 可以以各种方式来计算或估计。在一些示例中, U 可以至少部分地基于每个回声参考之间的相关系数。在一个这样的示例中, U 可以如下进行估计:

[0178]
$$U_r = 1 - \max_r (\sum_{m=0}^M \sum_{n=0}^N x_r[n] x_m[n])$$
,其中,下标“ r ”对应于所评估的特定回声参考, N 表示音频环境中的音频设备的总数, n 表示单个音频设备, M 表示音频环境中的潜在回声参考的总数,并且 m 表示单个回声参考。

[0179] 可替代地或附加地,在一些示例中, U 可以至少部分地基于对音频信号进行分解以寻找冗余。一些这样的示例可以涉及瞬时频率估计、基频(F_0)估计、频谱图反演和/或非负矩阵因式分解(NMF)。

[0180] 根据一些示例, U 可以至少部分地基于用于矩阵解码的数据。矩阵解码是一种音频技术,其中,少量离散音频声道(例如2个)在回放时被解码为大量声道(例如4或5个)。声道通常被布置用于由编码器传输或记录,并由解码器解码以进行回放。矩阵解码允许将多声道音频(如环绕声)编码为立体声信号,在立体声设备上作为立体声回放,并在环绕声设备

上作为环绕声回放。在一个这样的示例中,如果杜比5.1系统正在接收立体声音频数据流,则可以将静态上混合矩阵应用于立体声音频数据,以便为杜比5.1系统中的每个扩音器提供正确渲染的音频。根据一些示例,U可以至少部分地基于用于将音频分配给音频环境的每个扩音器(例如,音频设备110A-110D中的每个)的上混合或下混合(down-mixing)矩阵的系数。

[0181] 在一些示例中,U可以至少部分地基于在音频环境中使用的标准规范的扩音器布局(例如,杜比5.1、杜比7.1等)。一些这样的示例可以涉及利用传统上在这种规范的扩音器布局中混合和呈现媒体内容的方式。例如,在杜比5.1或杜比7.1系统中,艺术家通常将人声放在中央声道,而不是环绕声道。如上所述,与乐器和其他声源相对应的音频通常在房间左右两侧的声道上被复制。在一些实例中,可以经由与对应的音频数据一起接收的元数据来识别声音、对话、器乐等。

[0182] P或“持续性”方面

[0183] 持续性指标旨在捕获不同类型的回放媒体可能具有大范围的时间持续性这一方面,其中,不同类型的内容具有不同程度的无声和扩音器激活。频谱密集连续内容流(比如音乐或视频游戏控制台的音频输出)可能具有高水平的持续性,而播客可能具有较低水平的持续性。不频繁的系统通知的时间持续性水平将非常低。取决于手头的具体列表任务,与具有较低程度的持续性的媒体相对应的回声参考对于EMS来说可能不太重要。例如,偶尔的系统通知不太可能与唤醒词或插话请求发生冲突,因此管理该回声的相对重要性较低。

[0184] 以下是可以用于度量或估计持续性的指标示例:

[0185] • 最近历史窗口中回放信号高于特定数字响度阈值的时间百分比;

[0186] • 表明内容对应于音乐、广播内容、播客或系统声音的元数据标签或媒体分类指示;和/或

[0187] • 最近历史窗口期间回放信号处于人声典型频率范围(例如,100Hz至3KHz)的时间百分比。

[0188] 根据一些示例,音频内容类型可能影响L、U和/或P的估计。例如,知道音频内容是立体声音乐将允许仅使用上述声道指派来对所有回声参考进行排名。可替代地,如果控制系统不分析音频内容,而是依赖于声道指派,则知道音频内容是Atmos可以更改默认的L、U和/或P假设。

[0189] A或“可听度”方面

[0190] 可听度指标针对以下事实:音频设备具有不同的回放特性,并且在任何给定的音频环境中,音频设备之间的距离可能不同。以下是可以用于度量或估计音频设备可听度的指标示例:

[0191] • 音频设备可听度的直接测量结果;

[0192] • 指包括音频设备的一个或多个扩音器的特性的数据结构,比如额定SPL、频率响应和方向性(例如,扩音器是否是全向的、向前发声、向上发声等);

[0193] • 基于与音频设备的距离的估计;和/或

[0194] • 上述任意组合。

[0195] 可以评估其他因素以用于估计重要性,并且在一些实例中用于确定重要性度量。

[0196] 收听目标

[0197] 收听目标可以定义EMS的背景和期望的性能特性。在一些示例中,收听目标可以修改LUPA评估的参数和/或域。以下讨论将考虑收听目标发生变化的3种潜在背景。在这些不同的背景下,我们将看到概率和关键性可以如何影响LUPA。

[0198] 1. 插话(例如,检测唤醒词的实例)

[0199] 当等待插话时,没有立即的紧迫性:通常认为,用户在未来的所有时间间隔内说出唤醒词的概率是相同的。此外,唤醒词检测器可能是语音助理中最稳健的元件,并且回声泄漏的影响不那么关键。

[0200] 2. 命令

[0201] 在一个人说出唤醒词之后,这个人立即说出命令的可能性非常高。因此,在不久的将来与回声发生冲突的概率很大。此外,因为命令识别模块可能相对不如唤醒词检测器稳健,所以回声泄漏的关键性通常会很高。

[0202] 3. 交流

[0203] 在语音通话期间,任何参与者(音频环境中的(多个)人和远端的(多个)人)彼此交谈的可能性是确定的。换句话说,回声与用户语音冲突的概率本质上是1。然而,由于远端的一个或多个人是人类并且可以很好地应对背景噪声,因此关键性很小,因为他们不太可能受到回声泄漏的困扰。

[0204] 在这些不同的收听目标背景下,在一些示例中,评估LUPA的方式可能会改变。

[0205] 1. 插话

[0206] 可能没有时间上的区分,因为认为在所有未来时间间隔说出唤醒词的概率是相同的。因此,控制系统评估LUPA的时间范围可能相当长,以便获得这些参数的更好估计。在一些这样的示例中,控制系统评估LUPA的时间间隔可以被设置为着眼于相对较远的未来(例如,在几分钟的时间范围内)。

[0207] 2. 命令

[0208] 紧接在说出唤醒词之后的时间间隔很可能会说出命令。因此,在检测到唤醒词之后,在一些实施方式中,可以在比插话背景下短得多的时间尺度(例如,大约几秒)上评估LUPA。在一些示例中,由于冲突的可能性很高,所以在该时间间隔期间,在时间上稀疏且在唤醒词检测后的接下来几秒内有内容播放的参考将被认为更加重要。

[0209] 图5A是概述所公开方法的一个示例的流程图。与本文描述的其他方法一样,不必以所指示的顺序来执行方法500的框。在一些示例中,一个或多个框可以同时执行。此外,这样的方法可以包括比所示出和/或所描述的框更多或更少的框。例如,一些实施方式可以不包括框501。

[0210] 在该示例中,方法500是回声参考选择方法。方法500的框可以例如由控制系统(比如图2A或图3A的控制系统60a)执行。在一些示例中,方法500的框可以由回声参考选择器模块(比如上文参考图4描述的回声参考选择器402A)执行。

[0211] 图5A的参考选择方法是本文中可以称为“贪婪”回声参考选择方法的示例,其涉及仅在MC-EMS的当前操作点上评估成本和预期性能提升(换句话说,MC-EMS当前正在使用多少个参考,包括已选择的回声参考),并评估添加每个附加回声参考的结果,例如,按重要性降序排列。相应地,该示例涉及确定是否添加新的回声参考的过程。在一些实施方式中,在

方法500中评估的回声参考可能已经根据估计的重要性进行了排名(例如,由回声参考重要性估计器401A)。如果采用更复杂的技术(如树搜索方法),则在成本和性能方面可能会存在更优化的解决方案类型。替代示例可以涉及其他搜索和/或优化例程,包括蛮力方法。一些替代实施方式可以涉及确定是否丢掉或丢弃先前选择的回声参考。

[0212] 在该示例中,框501涉及确定EMS的当前性能水平是否大于或等于期望的性能水平。如果是,则过程终止(框510)。然而,如果确定当前性能水平低于期望性能水平,则在该示例中,过程继续到框502。根据该示例,框501的确定至少部分地基于指示EMS的当前性能的一个或多个指标,比如自适应滤波器系数数据或其他AEC统计数据、语音回声(SER)比数据等。在框501的确定由回声参考编排器302A做出的一些示例中,这一确定可以至少部分地基于来自MC-EMS203A的一个或多个指标350A。如上所述,一些实施方式可以不包括框501。

[0213] 根据该示例,框502涉及按重要性对剩余未选择的回声参考进行排名,并且估计通过包括EMS尚未使用的最重要的回声参考而获得的潜在EMS性能提升。在框502的过程由回声参考编排器302A执行的一些示例中,该过程可以至少部分地基于由MC-EMS性能模型405A产生的信息423A,在一些示例中,该信息可以是或者包括如图3B或图3C所示的数据。在一些实施方式中,上述排名和预测过程可以在方法500的较早阶段执行,例如,当评估先前的回声参考时。在一些示例中,可以在执行方法500之前执行上述排名和预测过程。在先前已经执行了上述排名和预测过程的一些实施方式中,框502可以简单地涉及选择由这样的先前过程所确定的最高排名的未选择回声参考。

[0214] 在该示例中,框503涉及比较添加在框502中选择的回声参考的性能和成本。在框503的过程由回声参考编排器302A执行的一些示例中,框503可以至少部分地基于来自成本估计模块403A的信息422A,该信息指示将回声参考包括在回声参考集313A中的(多个)成本。

[0215] 因为性能和成本可能是具有不同范围和/或域的变量,所以直接比较这些变量可能具有挑战性。因此,在一些实施方式中,可以通过将可能是变量的性能和成本映射到类似的尺度(比如预定义的最小值与最大值之间的范围)来促进框503的评估。

[0216] 在一些实施方式中,如果添加回声参考不会导致超过预定的网络带宽和/或计算成本的预算,则添加被评估的回声参考的成本可以简单地设置为零。在一些这样的示例中,如果添加回声参考将导致超过预定的网络带宽和/或计算成本的预算,则添加被评估的回声参考的成本可以被设置为无穷大。这种示例具有简单和高效的益处。以这种方式,控制系统可以简单地添加在预定的网络带宽和/或计算成本的预算所允许范围内的最大数量的回声参考。

[0217] 根据一些示例,如果与添加回声参考相对应的估计性能提升不高于预定阈值(例如,1%、2%、3%、4%、5%、6%、7%、8%、9%、10%等),则该估计性能提升可以设置为零。这样的方法可以防止由于包括仅增加微不足道的性能提升的回声参考而消耗网络带宽和/或计算开销。下文描述了成本确定的一些详细的替代示例。

[0218] 在该示例中,框504涉及在给定框503的性能/成本评估的情况下确定是否将添加新的回声参考。在一些示例中,框503和504可以组合为单个框。根据该示例,框504涉及确定添加被评估的回声参考的成本是否将小于估计由添加该回声参考引起的EMS性能提升。在该示例中,如果估计成本不小于估计性能提升,则过程继续到框511并且方法500终止。然

而,在该实施方式中,如果估计成本小于估计性能提升,则过程继续到框505。

[0219] 根据该示例,框505涉及将新的回声参考添加到所选回声参考集。在一些实例中,框505可以包括通知渲染器202输出相关回声参考。根据一些示例,框505可以涉及通过本地网络发送回声参考或者向另一设备发送命令311以通过本地网络发送回声参考。

[0220] 在方法500中评估的回声参考可以是本地回声参考或非本地回声参考,非本地回声参考可以在本地确定(例如,由如上所述的本地渲染器)或通过本地网络接收。因此,一些回声参考的成本估计可能涉及评估计算成本和网络成本两者。

[0221] 根据一些示例,为了在框505之后评估下一个回声参考,控制系统可以简单地重置所选择的和未选择的回声参考并恢复到图5A的先前框,诸如框501、框502或框503。然而,更复杂的方法还可能涉及评估已经选择的参考,例如,对已经选择的所有参考进行排名,并决定是否丢掉具有最低估计重要性的回声参考。

[0222] 替代回声参考形式

[0223] 回声参考可以以多种形式或变体来传输(或在诸如产生所有回声参考的设备等设备内本地使用),这可能更改该特定回声参考的成本/效益比。例如,如果我们将回声参考变换为分段功率形式(换句话说,确定多个频带中的每一个频带中的功率并传输关于每个频带中的功率的分段功率信息),则有可能降低通过本地网络发送回声参考的成本。然而,使用回声参考的低保真度变体的EMS可以获得的潜在改进通常也会较低。选择使得回声参考的任何特定变体可用可以被解释为使其成为潜在的选择候选。

[0224] 在一些实施方式中,回声参考可以是下文所列出的以下形式之一(其中前四种按估计性能的降序排列):

[0225] • 全保真度(原始、确切)回声参考,这将产生全部的计算成本和网络成本(如果通过网络传输的话)

[0226] • 下采样回声参考,其计算成本和网络成本将根据下采样因子按比例减少,但会产生下采样过程的计算成本;

[0227] • 经由有损编码过程产生的已编码回声参考,其网络成本可以根据编码方案的压缩比而降低,但会产生编码和解码计算成本;

[0228] • 与回声参考相对应的分段功率信息,其网络成本可以显著降低,因为频带的数量可以远低于全保真度回声参考的子带数量,并且其计算成本可以显著降低,因为实施分段AES的成本远低于实施子带AEC的成本;或者

[0229] • 降低保真度以换取一定成本(无论是计算、网络还是其他成本,例如存储器)降低的任何其他形式。

[0230] 图5B是概述了所公开方法的另一示例的流程图。与本文描述的其他方法一样,不必以所指示的顺序来执行方法550的框。在一些示例中,一个或多个框可以同时执行。此外,这样的方法可以包括比所示出和/或所描述的框更多或更少的框。

[0231] 方法550的框可以例如由控制系统(比如图2A或图3A的控制系统60a)执行。在一些示例中,方法550的框可以由回声参考选择器模块(比如上文参考图4描述的回声参考选择器402A)执行。

[0232] 方法550考虑到以下事实:回声参考不一定以全保真度形式来传输或使用,而是可以以上述替代的部分保真度形式之一来传输或使用。因此,在方法550中,性能和成本的评

估不涉及关于是否将使用或不使用全保真度形式的回声参考的二元决策。相反,方法550涉及确定是否包括回声参考的一个或多个低保真度版本,这可能涉及并且潜在地较少的EMS性能提升,但是成本较低。诸如方法550等方法为要由回声管理系统使用的潜在回声参考集提供了附加的灵活性。

[0233] 在该示例中,方法550是上文参考图5A描述的回声参考选择方法500的扩展。因此,框501(如果包括的话)、502、503、504和505可以如上参考图5A来执行,除非下文有相反说明。方法550将包括框506和507的潜在迭代循环添加到方法500。根据该示例,如果确定(此处,在框504中)添加回声参考的一个版本的估计成本将不小于估计的EMS性能提升,则在框506中确定是否存在回声参考的另一版本。在一些示例中,回声参考的全保真度版本可以在较低保真度版本(如果有任何版本可用的话)之前被评估。根据该实施方式,如果在框506中确定回声参考的另一版本可用,则在框507中将选择回声参考的另一版本(例如,不是全保真度版本的最高保真度版本)并在框503中进行评估。

[0234] 因此,方法550涉及评估回声参考的较低保真度版本,如果有任何版本可用的话。这种较低保真度版本可以包括回声参考的下采样版本、经由有损编码过程产生的回声参考的已编码版本、和/或与回声参考相对应的分段功率信息。

[0235] 成本模型

[0236] 回声参考的“成本”是指利用该参考进行回声管理所需的资源,无论是使用AEC还是AES。一些公开的实施方式可能涉及估计以下类型的成本中的一种或多种:

[0237] • 计算成本,其可以参考对音频环境中的一个或多个设备上可用的有限量的处理能力的使用来确定。计算成本可以指以下一项或多项:

[0238] ◦使用该参考在特定听音设备上执行回声管理所需的成本。这可能是指在AEC或AES中使用该参考。人们会注意到,AEC在仓(bin)或子带(其是复数)上运行,并且需要比在频带(其数量与AES使用的仓/子带相比较少,并且频带功率是实数,而不是复数)上运行的AES多得多的CPU运算;

[0239] ◦在使用编解码的参考时对回声参考进行编码或解码所需的成本;

[0240] ◦对信号进行分段所需的成本(换句话说,将信号从简单的线性频域表示变换为分段频域表示);和/或

[0241] ◦产生回声参考所需的成本(例如,通过渲染器)。

[0242] • 网络成本,是指对有限量的网络资源的使用,比如用于在设备之间共享回声参考的本地网络(例如,音频环境中的本地无线网络)中可用的带宽。

[0243] 特定回声参考集的总成本可以被确定为该集合中的每个回声参考的成本之和。一些公开的示例涉及组合网络成本和计算成本。根据一些示例,总成本 C_{total} 可以按下式确定:

$$[0244] \quad C_{total} = \max \left(\frac{1}{R_{comp}} \sum_m^M C_m^{comp}, \quad \frac{1}{R_{network}} \sum_m^M C_m^{network} \right)$$

[0245] 在上述等式中, R_{comp} 表示可用于回声管理的计算资源总量, $R_{network}$ 表示可用于回声管理的网络资源总量; C_m^{comp} 表示与使用第m个参考相关联的计算成本,并且 $C_m^{network}$ 表示与使用第m个参考相关联的网络成本(其中,EMS中总共使用了M个参考)。人们可能会注意到这一定义意味着

[0246] $0 \leq C_{\text{total}} \leq 1$,

[0247] 并且 C_{total} 仅包括与变得受系统可用资源限制的成本最接近的成本分量。

[0248] 性能

[0249] 回声管理系统(EMS)的“性能”可以指以下内容:

[0250] • 从麦克风馈送中去除的回声量,其可以用回声损失增强(ERLE)来度量,回声损失增强是以分贝为单位测量的并且是发送功率与残差信号的功率的比率。该指标可以例如根据基于应用的指标(比如支持自动语音识别(ASR)处理器执行在存在回声的情况下检测说出的特定关键词的唤醒词检测任务所需的最小ERLE)来标准化;

[0251] • 当受到房间噪声源、本地音频系统的非线性、双端通话等干扰时EMS的稳健性;

[0252] • 当使用低于全保真度的回声参考时EMS的稳健性;

[0253] • EMS跟踪系统变化的能力,包括EMS初始收敛的能力;和/或

[0254] • EMS跟踪经渲染音频场景的变化的能力。例如,这可以指回声参考协方差矩阵的移位以及EMS对非平稳非唯一性问题的稳健性。

[0255] 一些示例可能涉及确定单个性能指标P。一些这样的示例使用ERLE和根据自适应滤波器系数数据或从EMS获得的其他AEC统计数据而估计的稳健性。根据一些这样的示例,性能稳健性指标 P_{rob} 可以使用从AEC提取的“麦克风概率”来确定,例如如下:

[0256] $P_{\text{Rob}} = 1 - M_{\text{prob}}$

[0257] 在上述等式中, $0 \leq P_{\text{Rob}} \leq 1$, $0 \leq M_{\text{prob}} \leq 1$, 并且 M_{prob} 表示麦克风概率,它是AEC中产生在各自子带中不提供实质性(或任何)回声消除的不良回声预测的子带自适应滤波器的数量比例。

[0258] 唤醒词(WW)检测器的性能很大程度上取决于语音回声比(SER),该语音回声比可通过EMS的ERLE按比例提高。当SER太低时,WW检测器更有可能错误触发(误报)并漏掉用户说出的关键词(漏检测),因为回声会破坏麦克风信号并降低系统的准确度。由ASR处理器(例如,图2A的语音处理块240A)消耗的残差信号(例如,图2A的残差信号224A)的SER由EMS与EMS的ERLE成比例地提高,从而改进WW检测器的性能。

[0259] 因此,一些公开的示例涉及将期望的WW性能水平映射到标称SER水平,这进而结合系统中设备的典型回放水平的知识而允许控制系统将这种期望的WW性能水平直接映射到标称ERLE。在一些示例中,可以扩展该方法以将系统在各种SER水平的WW性能映射到ERLE。在一些这样的实施方式中,可以使用具有一定范围的SER值的输入数据来产生特定WW检测器的接收器操作特性(ROC)曲线。一些示例涉及将选择感兴趣的特定误报率(FAR)并针对该特定FAR将WW检测器的准确度作为SER的函数来作为我们的应用基础。在一些这样的示例中,

[0260] $\text{Acc}(\text{SER}_{\text{res}}) = \text{ROC}(\text{SER}_{\text{res}}, \text{FAR}_1)$

[0261] 上述等式中, $\text{Acc}(\text{SER}_{\text{res}})$ 表示WW检测器的准确度作为表示由EMS输出的残差信号的SER的 SER_{res} 的函数。 $\text{ROC}()$ 表示针对多个SER的ROC曲线的集合,并且 FAR_1 表示感兴趣的误报率,其典型值可以是每24小时3个和每10小时1个。准确度 $\text{Acc}(\text{SER}_{\text{res}})$ 可以表示为百分比或被归一化使得其在0到1的范围内,其可以表达如下:

[0262] $0 \leq \text{Acc}(\text{SER}_{\text{res}}) \leq 1$

[0263] 有了音频环境中的音频设备的回放能力的知识,就可以结合使用例如实际回声水

平的LUPA分量和目标音频环境中典型的语音水平来确定麦克风信号(例如,图2A的麦克风信号223A)中的典型SER值,例如如下:

$$[0264] \quad SER_{mic} = \frac{Speech_pwr}{Echo_pwr}$$

[0265] 在上述等式中,Speech_pwr和Echo_pwr分别表示目标音频环境的预期基线语音功率水平和回声功率水平。通过EMS,SER_{mic}可以改进为与ERLE成比例的SER_{res},例如如下:

$$[0266] \quad SER_{res}^{dB} = SER_{mic}^{dB} + ERLE^{dB}$$

[0267] 在上述等式中,上标dB指示在该示例中变量以分贝为单位。为了完整性,一些实施方式可以将EMS的ERLE定义如下:

$$[0268] \quad ERLE^{dB} = 10 \log_{10} \left(\frac{mic_echo_pwr}{residual_echo_pwr} \right)$$

[0269] 使用前述等式,一些实施方式可以定义基于WW应用的EMS性能指标,如下所示:

$$[0270] \quad P_{WW} = Acc(SER_{mic}^{dB} + ERLE^{dB}),$$

[0271] 其中,SER_{mic}^{dB}代表目标环境中的SER。在一些示例中,SER_{mic}^{dB}可以是静态默认数字,而在其他示例中,SER_{mic}^{dB}可以被估计为例如一个或多个LUPA分量的函数。一些实施方式可以涉及将净性能指标P定义为包含每个要素的向量,例如如下:

$$[0272] \quad P = [P_{ww}, P_{Rob}]$$

[0273] 在一些示例中,可以通过增加净性能向量的大小来添加一个或多个附加性能分量。在一些替代示例中,一个或多个附加性能分量可以通过对它们进行加权而组合成单个标量指标,例如如下:

$$[0274] \quad P = (1-K)P_{ww} + KP_{Rob}$$

[0275] 在上述等式中,K表示由系统设计者选择的加权因子,其用于确定每个分量对净性能的贡献程度。一些替代示例可以使用另一种方法,例如,只是对各个性能指标进行平均。然而,将各个性能指标组合成单个标量指标可能是有利的。

[0276] 成本和性能的权衡

[0277] 当比较回声参考的估计成本和估计EMS性能增强时,需要一种方法以某种方式比较通常不在同一域中的这两个参数。一种这样的方法涉及单独地评估成本估计和性能估计,并采用成本最低且满足预定义最低性能标准P_{min}的解决方案。该预定义的EMS性能标准可以例如根据特定下游应用的要求(例如,提供电话通话、音乐回放、等待WW等)来确定。

[0278] 例如,在应用是WW检测的实施方式中,性能可以与WW性能指标P_{ww}相关。在一些这样的示例中,可能存在被认为足够的某个最低水平的WW检测器准确度(例如,80%水平的WW检测器准确度、85%水平的WW检测器准确度、90%水平的WW检测器准确度、95%水平的WW检测器准确度等),这按照上一节将具有对应的ERLE^{dB}。在一些这样的示例中,可以使用EMS性能模型(例如,图4的MC-EMS性能模型405)来估计EMS的ERLE。因此,如果目标只是找到成本最低的解决方案(例如,就总成本C_{total}而言),则这样的实施方式不需要直接权衡成本和性能。

[0279] 作为满足一些最低性能指标的替代方案,一些实施方式可能涉及使用性能指标P和成本指标C。一些这样的示例可能涉及使用权衡参数λ(例如,拉格朗日乘数),并将成本/性能评估过程表述为寻求使某个量最大化的优化问题,比如在以下表达式中的变量F:

[0280] $F = P - \lambda C_{\text{total}}$

[0281] 可以观察到,在上述等式中,F的值相对较大对应于性能指标P与 λ 和总成本 C_{total} 的积之间的差相对较大。权衡参数 λ 可以(例如,由系统设计者)选择以便直接权衡成本和性能。然后可以使用优化算法来找到由EMS使用的回声参考集的解,其中,回声参考集(其可以包括所有可用的回声参考保真度水平)决定了搜索空间。

[0282] 图6是概述所公开方法的一个示例的流程图。与本文描述的其他方法一样,不必以所指示的顺序来执行方法600的框。此外,这样的方法可以包括比所示出和/或所描述的框更多或更少的框。在一些示例中,两个或更多个框可以同时执行。在该示例中,方法600是音频处理方法。

[0283] 方法600可以由如图1A中示出且上文描述的装置50的装置或系统执行。在一些示例中,方法600的框可以由音频环境内的一个或多个设备来执行,例如,由音频系统控制器(如本文中被称为智能家居中枢的设备)或由音频系统的另一个部件来执行,如智能扬声器、电视、电视控制模块、膝上型计算机、移动设备(如蜂窝电话)等。在一些实施方式中,音频环境可以包括家庭环境的一个或多个房间。在其他示例中,音频环境可以是另一种类型的环境,如办公室环境、汽车环境、火车环境、街道或人行道环境、公园环境等。然而,在替代性实施方式中,方法600的至少一些框可以由实施基于云的服务的设备(如服务器)来执行。

[0284] 在该实施方式中,框605涉及由控制系统获得多个回声参考。在该示例中,多个回声参考包括针对音频环境中的多个音频设备中的每个音频设备的至少一个回声参考。这里,每个回声参考对应于由多个音频设备中的一个音频设备的一个或多个扩音器回放的音频数据。

[0285] 在该示例中,框610涉及由控制系统对多个回声参考中的每个回声参考做出重要性估计。根据该示例,做出重要性估计涉及确定每个回声参考对由音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻的预期贡献。在该示例中,至少一个回声管理系统包括声学回声消除器(AEC)和/或声学回声抑制器(AES)。

[0286] 在该实施方式中,框615涉及由控制系统并且至少部分地基于重要性估计来选择一个或多个所选回声参考。在该示例中,框620涉及由控制系统将一个或多个所选回声参考提供给至少一个回声管理系统。在一些实施方式中,方法600可以涉及使得至少一个回声管理系统至少部分地基于一个或多个所选回声参考来消除或抑制回声。

[0287] 在一些示例中,获得多个回声参考可以涉及接收包括音频数据的内容流并基于音频数据来确定多个回声参考中的一个或多个回声参考。上文参考图2A的渲染器201A描述了一些示例。

[0288] 在一些实施方式中,控制系统可以包括音频环境中的多个音频设备中的音频设备的音频设备控制系统。在一些这样的示例中,所述方法可以涉及由音频设备控制系统渲染音频数据以用于在音频设备上再现,从而产生本地扬声器馈送信号。在一些这样的示例中,所述方法可以涉及确定与本地扬声器馈送信号相对应的本地回声参考。

[0289] 在一些示例中,获得多个回声参考可以涉及基于音频数据来确定一个或多个非本地回声参考。例如,每个非本地回声参考可以对应于用于在音频环境的另一音频设备上回放的非本地扬声器馈送信号。

[0290] 根据一些示例,获得多个回声参考可以涉及接收一个或多个非本地回声参考。例

如,每个非本地回声参考可以对应于用于在音频环境的另一音频设备上回放的非本地扬声器馈送信号。在一些示例中,接收一个或多个非本地回声参考可以涉及从音频环境的一个或多个其他音频设备接收一个或多个非本地回声参考。在一些示例中,接收一个或多个非本地回声参考可以涉及从音频环境的单个其他设备接收一个或多个非本地回声参考中的每一个。

[0291] 在一些示例中,所述方法可以涉及成本确定。根据一些这样的示例,成本确定可以涉及确定多个回声参考中的至少一个回声参考的成本。在一些这样的示例中,选择一个或多个所选回声参考可以至少部分地基于成本确定。根据一些这样的示例,成本确定可以至少部分地基于用于传输至少一个回声参考所需的网络带宽、用于编码至少一个回声参考的编码计算要求、用于解码至少一个回声参考的解码计算要求、用于由回声管理系统使用至少一个回声参考的回声管理系统计算要求、或其一个或多个组合。在一些示例中,成本确定可以至少部分地基于至少一个回声参考在时域或频域中的全保真度复制品、至少一个回声参考的下采样版本、至少一个回声参考的有损压缩、至少一个回声参考的分段功率信息、或其一个或多个组合。在一些示例中,成本确定可以至少部分地基于与相对不太重要的回声参考相比对相对更重要的回声参考进行更少压缩的方法。

[0292] 根据一些示例,所述方法可以涉及确定当前回声管理系统性能水平。在一些这样的示例中,选择一个或多个所选回声参考可以至少部分地基于当前回声管理系统性能水平。

[0293] 在一些示例中,做出重要性估计可以涉及确定对应回声参考的重要性度量。在一些示例中,确定重要性度量可以涉及确定对应回声参考的水平、确定对应回声参考的唯一性、确定对应回声参考的时间持续性、确定对应回声参考的可听度、或其一个或多个组合。根据一些示例,确定重要性度量可以至少部分地基于与音频设备布局相对应的元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合矩阵、扩音器激活矩阵、或其一个或多个组合。在一些示例中,确定重要性度量可以至少部分地基于当前收听目标、当前环境噪声估计、至少一个回声管理系统的当前性能的估计、或其一个或多个组合。

[0294] 图7示出了音频环境的平面图的示例,所述音频环境在该示例中是生活空间。与本文提供的其他图一样,图7中示出的要素的类型和数量仅作为示例提供。其他实施方式可以包括更多、更少和/或不同类型和数量的要素。

[0295] 根据该示例,环境700包括在左上方处的客厅710、在下方中央处的厨房715、以及在右下方的卧室722。跨生活空间分布的方框和圆圈表示一组扩音器705a-705h,该组扩音器中的至少一些扩音器在一些实施方式中可以是智能扬声器,放置在对空间方便的位置,但不遵循任何标准规定的布局(任意地放置)。在一些示例中,电视730可以被配置为至少部分地实施一个或多个所公开的实施例。在该示例中,环境700包括分布在整个环境中的相机711a-711e。在一些实施方式中,环境700中的一个或多个智能音频设备还可以包括一个或多个相机。所述一个或多个智能音频设备可以是单一用途音频设备或虚拟助理。在一些这样的示例中,可选传感器系统130的一个或多个相机可以驻留在电视730中或所述电视上、移动电话中或智能扬声器(如扩音器705b、705d、705e或705h中的一个或多个)中。尽管相机711a-711e没有在本公开中呈现的音频环境的每个描绘中示出,但在一些实施方式中,每个音频环境仍然可以包括一个或多个相机。

[0296] 本公开的一些方面包括一种被配置(例如,被编程)成执行所公开方法的一个或多个示例的系统或设备,以及一种存储用于实施所公开方法或其步骤的一个或多个示例的代码的有形计算机可读介质(例如,磁盘)。例如,一些公开的系统可以是或者包括可编程通用处理器、数字信号处理器或微处理器,所述可编程通用处理器、数字信号处理器或微处理器用软件或固件编程为和/或以其他方式被配置成对数据执行各种操作中的任一个,包括所公开方法或其步骤的实施例。这样的通用处理器可以是或者包括计算机系统,所述计算机系统包括输入设备、存储器和处理子系统,所述处理子系统被编程(和/或以其他方式被配置)为响应于向其断言的数据而执行所公开方法(或其步骤)的一个或多个示例。

[0297] 一些实施例可以被实施为可配置的(例如,可编程的)数字信号处理器(DSP),所述数字信号处理器被配置(例如,被编程和以其他方式被配置)为对(多个)音频信号执行需要的处理,包括对所公开方法的一个或多个示例的执行。可替代地,所公开系统(或其元件)的实施例可以被实施为通用处理器(例如,个人计算机(PC)或其他计算机系统或微处理器,其可以包括输入设备和存储器),所述通用处理器用软件或固件编程为和/或以其他方式被配置成执行各种操作中的任一个,包括所公开方法的一个或多个示例。可替代地,本发明系统的一些实施例的元件被实施为被配置(例如,被编程)成执行所公开方法的一个或多个示例的通用处理器或DSP,并且所述系统还包括其他元件(例如,一个或多个扩音器和/或一个或多个麦克风)。被配置成执行所公开方法的一个或多个示例的通用处理器可以耦接到输入设备(例如,鼠标和/或键盘)、存储器和显示设备。

[0298] 本公开的另一方面是一种计算机可读介质(例如,磁盘或其他有形存储介质),所述计算机可读介质存储用于执行所公开方法或其步骤的一个或多个示例的代码(例如,可执行以执行所公开方法或其步骤的一个或多个示例的编码器)。

[0299] 虽然在本文中已经描述了本公开的具体实施例和本公开的应用,但是对于本领域普通技术人员而言显而易见的是,在不脱离本文描述的并要求保护的本公开的范围的情况下,可以对本文描述的实施例和应用进行许多改变。应当理解,虽然已经示出和描述了本公开的某些形式,但是本公开不限于所描述和示出的具体实施例或所描述的具体方法。

[0300] 可以从以下枚举的示例实施例(EEE)中理解本发明的各个方面:

[0301] 1.一种音频处理方法,包括:

[0302] 由控制系统获得多个回声参考,所述多个回声参考包括针对音频环境中的多个音频设备中的每个音频设备的至少一个回声参考,每个回声参考对应于由所述多个音频设备中的一个音频设备的一个或多个扩音器回放的音频数据;

[0303] 由所述控制系统对所述多个回声参考中的每个回声参考做出重要性估计,其中,做出所述重要性估计涉及确定每个回声参考对由所述音频环境的至少一个音频设备的至少一个回声管理系统进行的回声减轻的预期贡献,所述至少一个回声管理系统包括声学回声消除器(AEC)、声学回声抑制器(AES)、或者AEC和AES两者;

[0304] 由所述控制系统并且至少部分地基于所述重要性估计来选择一个或多个所选回声参考;以及

[0305] 由所述控制系统将所述一个或多个所选回声参考提供给所述至少一个回声管理系统。

[0306] 2.如EEE 1所述的音频处理方法,进一步包括使得至少一个回声管理系统至少部

分地基于所述一个或多个所选回声参考来消除或抑制回声。

[0307] 3. 如EEE 1或EEE 2所述的音频处理方法,其中,获得所述多个回声参考涉及:

[0308] 接收包括音频数据的内容流;以及

[0309] 基于所述音频数据确定所述多个回声参考中的一个或多个回声参考。

[0310] 4. 如EEE 3所述的音频处理方法,其中,所述控制系统包括所述音频环境中的所述多个音频设备中的音频设备的音频设备控制系统,所述音频处理方法进一步包括:

[0311] 由所述音频设备控制系统渲染所述音频数据以供在所述音频设备上再现,以产生本地扬声器馈送信号;以及

[0312] 确定与所述本地扬声器馈送信号相对应的本地回声参考。

[0313] 5. 如EEE 4所述的音频处理方法,其中,获得所述多个回声参考涉及基于所述音频数据确定一个或多个非本地回声参考,所述非本地回声参考中的每一个对应于用于在所述音频环境的另一音频设备上回放的非本地扬声器馈送信号。

[0314] 6. 如EEE 4所述的音频处理方法,其中,获得所述多个回声参考涉及接收一个或多个非本地回声参考,所述非本地回声参考中的每一个对应于用于在所述音频环境的另一音频设备上回放的非本地扬声器馈送信号。

[0315] 7. 如EEE 6所述的音频处理方法,其中,接收所述一个或多个非本地回声参考涉及从所述音频环境的一个或多个其他音频设备接收所述一个或多个非本地回声参考。

[0316] 8. 如EEE 6所述的音频处理方法,其中,接收所述一个或多个非本地回声参考涉及从所述音频环境的单个其他设备接收所述一个或多个非本地回声参考中的每一个。

[0317] 9. 如EEE 1-8中任一项所述的音频处理方法,进一步包括成本确定,所述成本确定涉及确定所述多个回声参考中的至少一个回声参考的成本,其中,选择所述一个或多个所选回声参考至少部分地基于所述成本确定。

[0318] 10. 如EEE 9所述的音频处理方法,其中,所述成本确定基于用于传输所述至少一个回声参考所需的网络带宽、用于编码所述至少一个回声参考的编码计算要求、用于解码所述至少一个回声参考的解码计算要求、用于由所述回声管理系统使用所述至少一个回声参考的回声管理系统计算要求、或其组合。

[0319] 11. 如EEE 9或EEE 10所述的音频处理方法,其中,所述成本确定基于所述至少一个回声参考在时域或频域中的复制品、所述至少一个回声参考的下采样版本、所述至少一个回声参考的有损压缩、所述至少一个回声参考的分段功率信息、或其组合。

[0320] 12. 如EEE 9-11中任一项所述的音频处理方法,其中,所述成本确定基于与相对不太重要的回声参考相比对相对更重要的回声参考进行更少压缩的方法。

[0321] 13. 如EEE 1-12中任一项所述的音频处理方法,进一步包括确定当前回声管理系统性能水平,其中,选择所述一个或多个所选回声参考至少部分地基于所述当前回声管理系统性能水平。

[0322] 14. 如EEE 1-13中任一项所述的音频处理方法,其中,做出所述重要性估计涉及确定对应回声参考的重要性度量。

[0323] 15. 如EEE 14所述的音频处理方法,其中,确定所述重要性度量涉及确定所述对应回声参考的水平、确定所述对应回声参考的唯一性、确定所述对应回声参考的时间持续性、确定所述对应回声参考的可听度、或其组合。

[0324] 16. 如EEE 14或EEE 15所述的音频处理方法,其中,确定所述重要性度量至少部分地基于与音频设备布局相对应的元数据、扩音器元数据、与接收到的音频数据相对应的元数据、上混合矩阵、扩音器激活矩阵、或其组合。

[0325] 17. 如EEE 14-16中任一项所述的音频处理方法,其中,确定所述重要性度量至少部分地基于当前收听目标、当前环境噪声估计、所述至少一个回声管理系统的当前性能的估计、或其组合。

[0326] 18. 一种装置,所述装置被配置为执行如EEE 1-17中任一项所述的方法。

[0327] 19. 一种系统,所述系统被配置成执行如EEE 1-17中任一项所述的方法。

[0328] 20. 一个或多个其上存储有软件的非暂态介质,所述软件包括用于控制一个或多个设备执行如EEE 1-17中任一项所述的方法的指令。

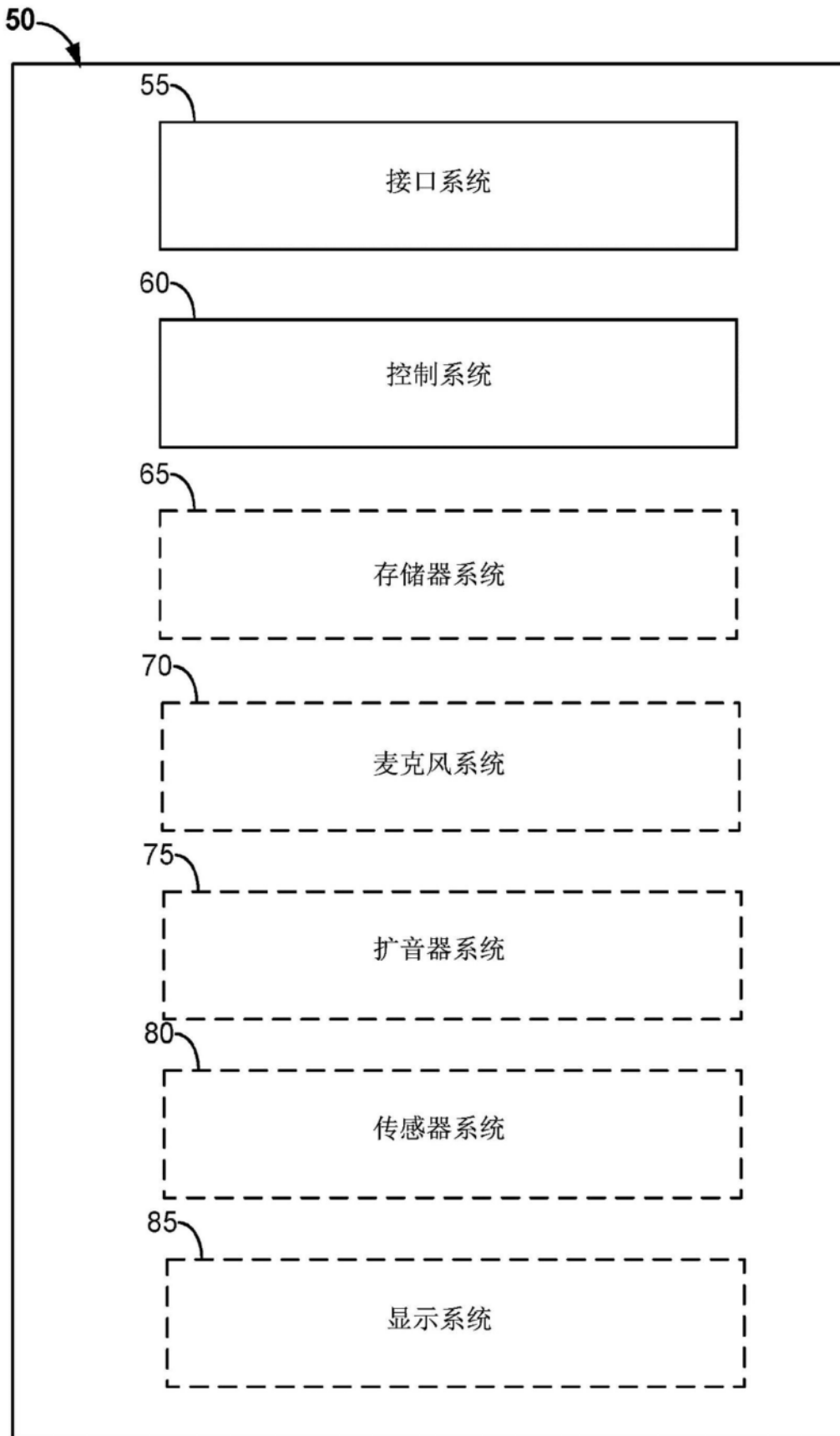


图1A

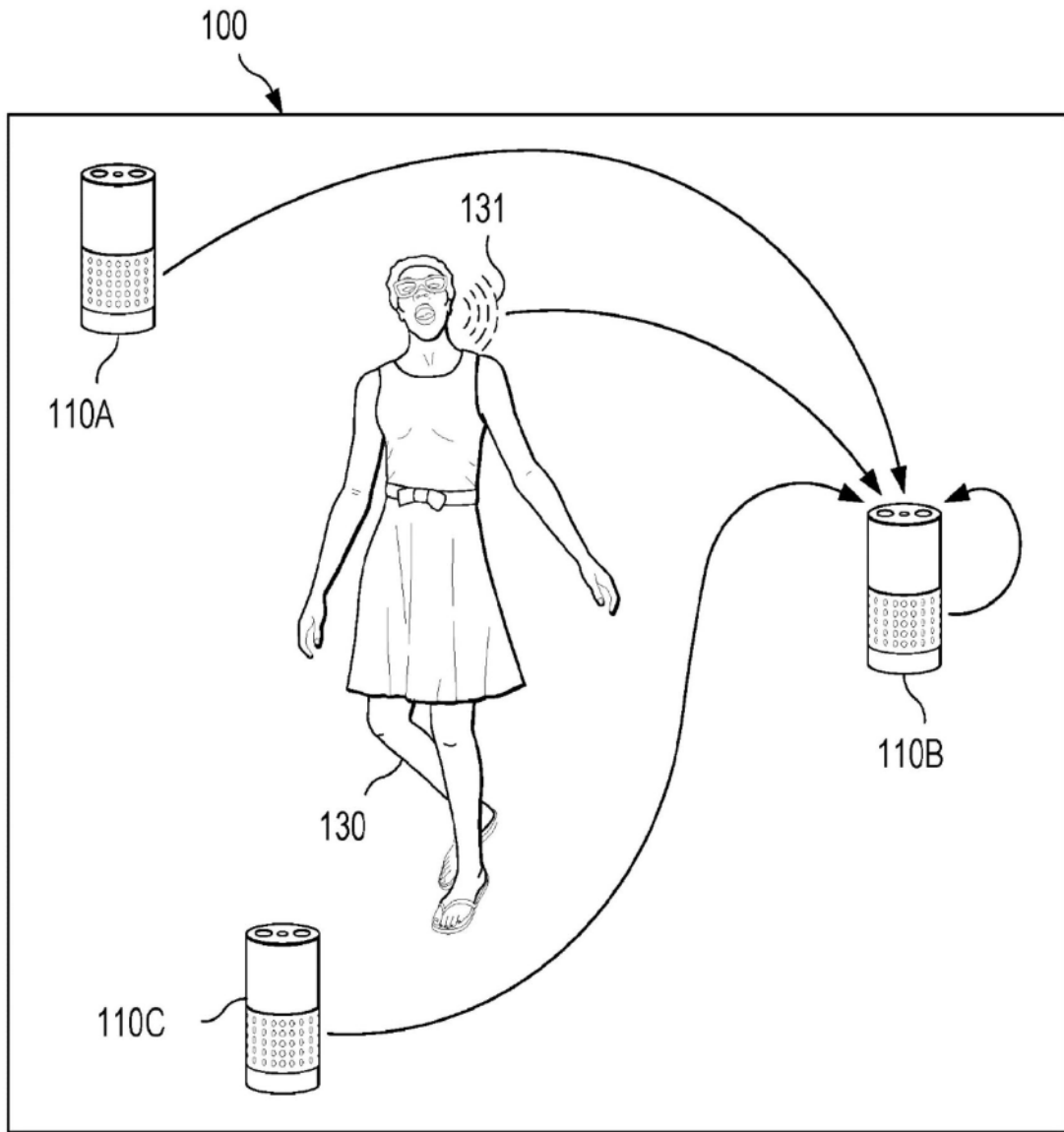


图1B

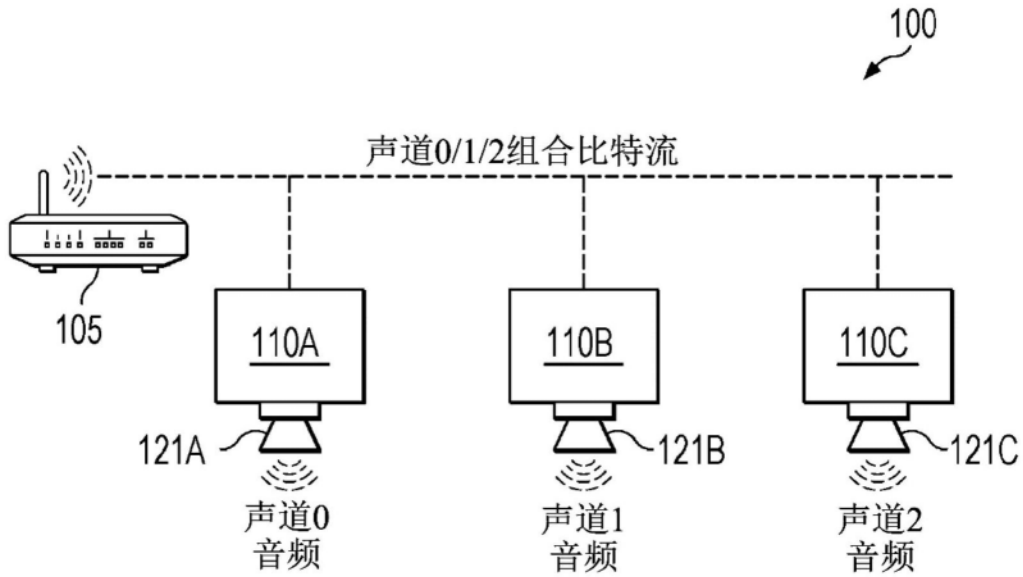


图1C

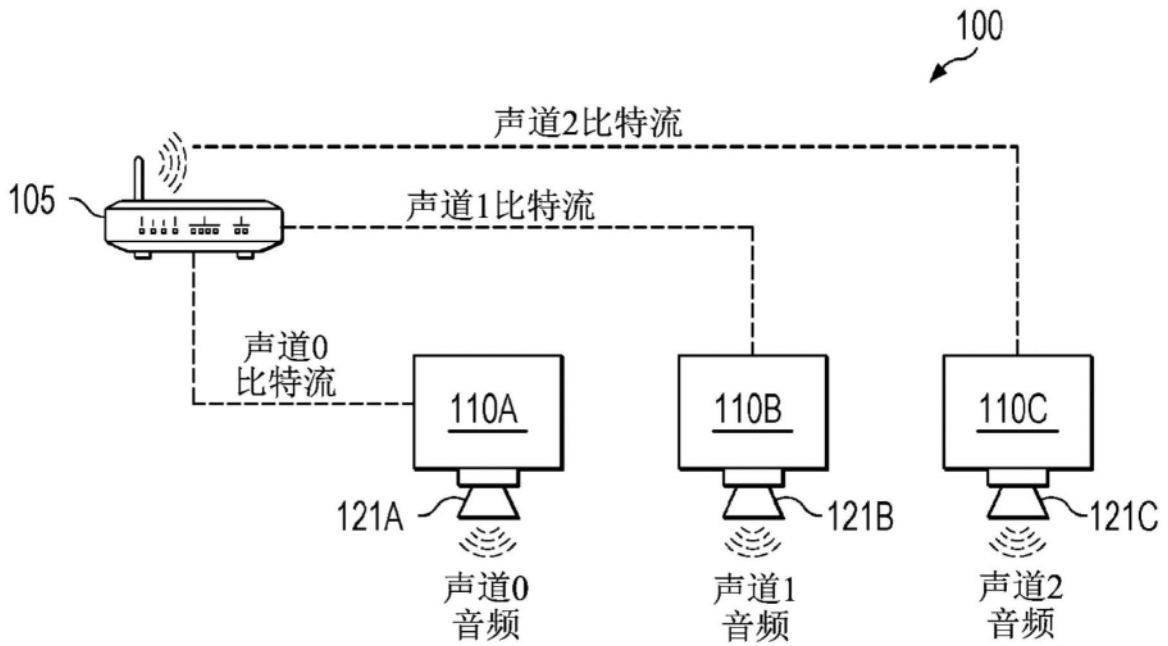


图1D

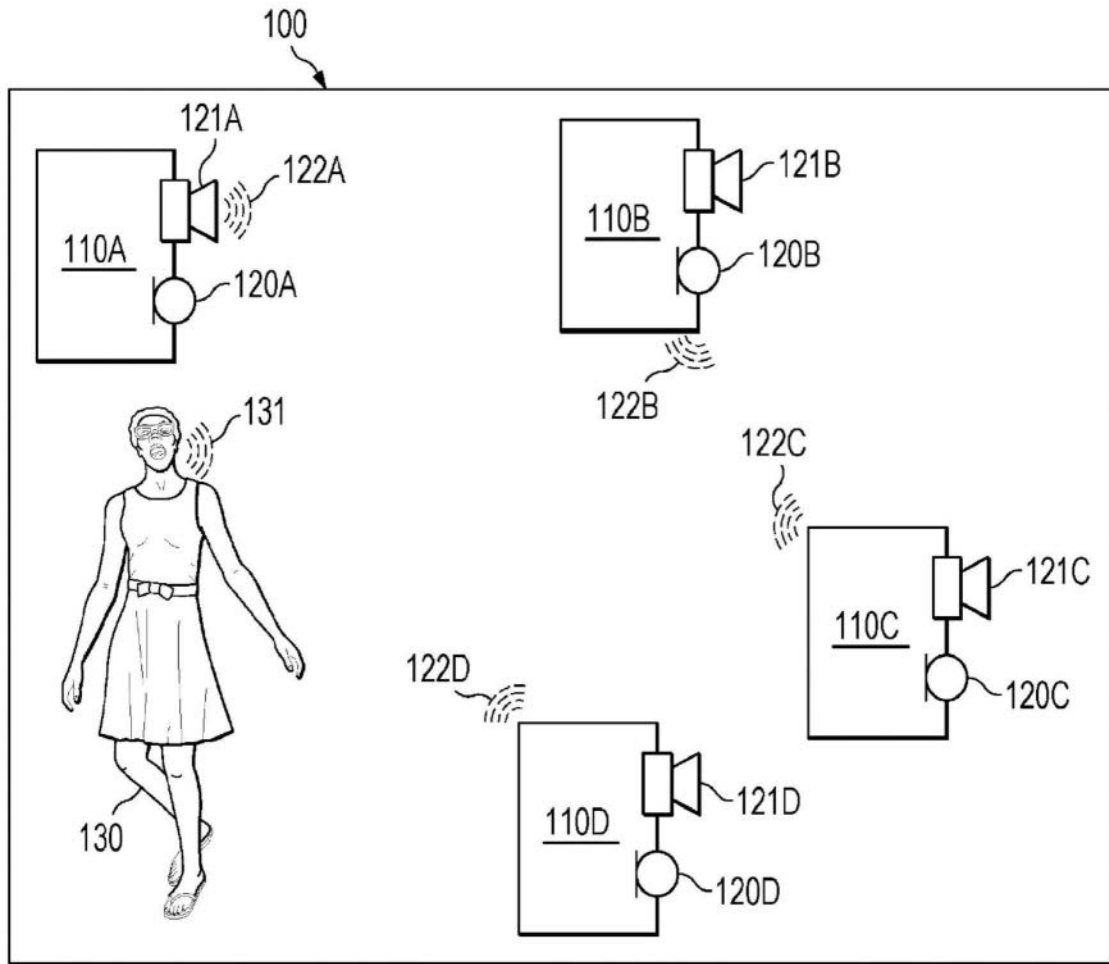


图1E

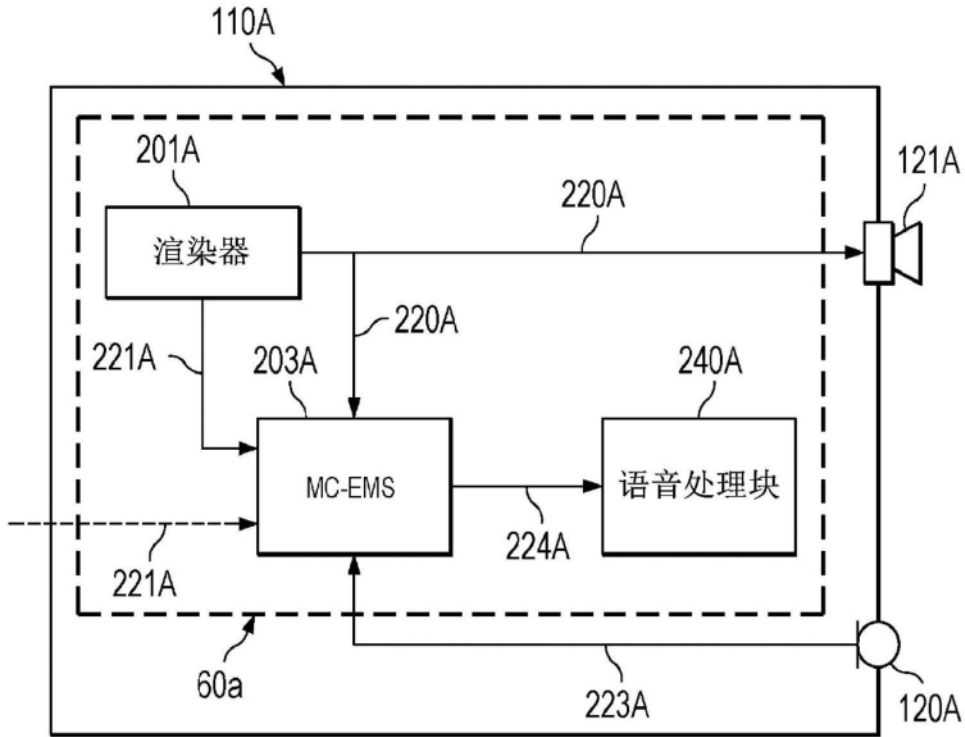


图2A

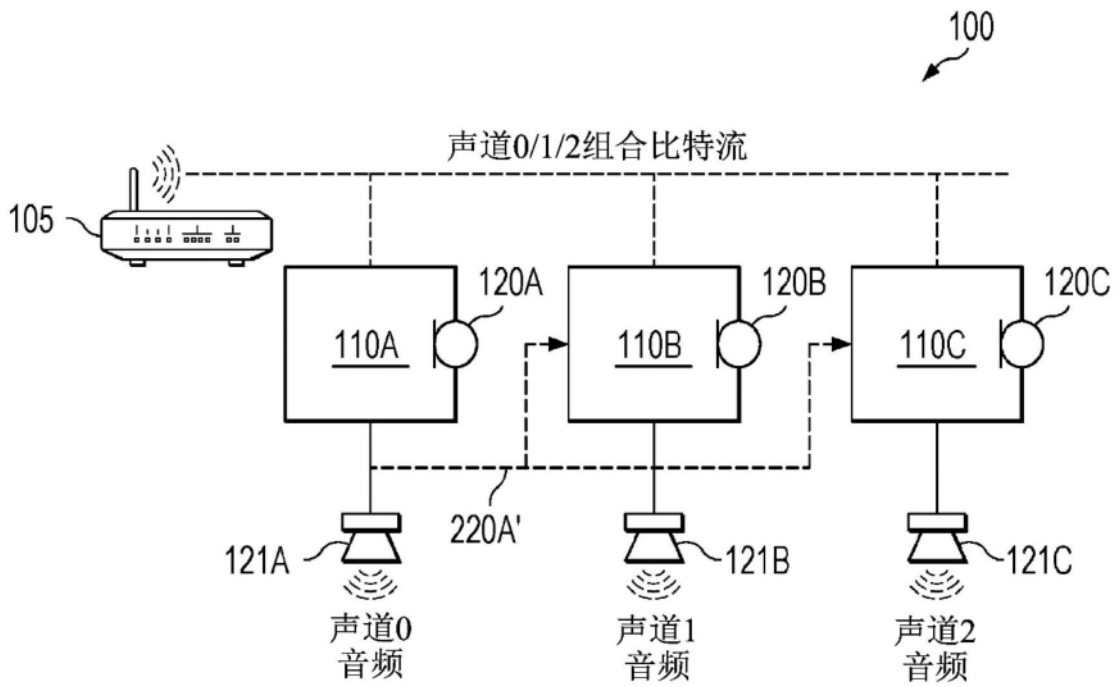


图2B

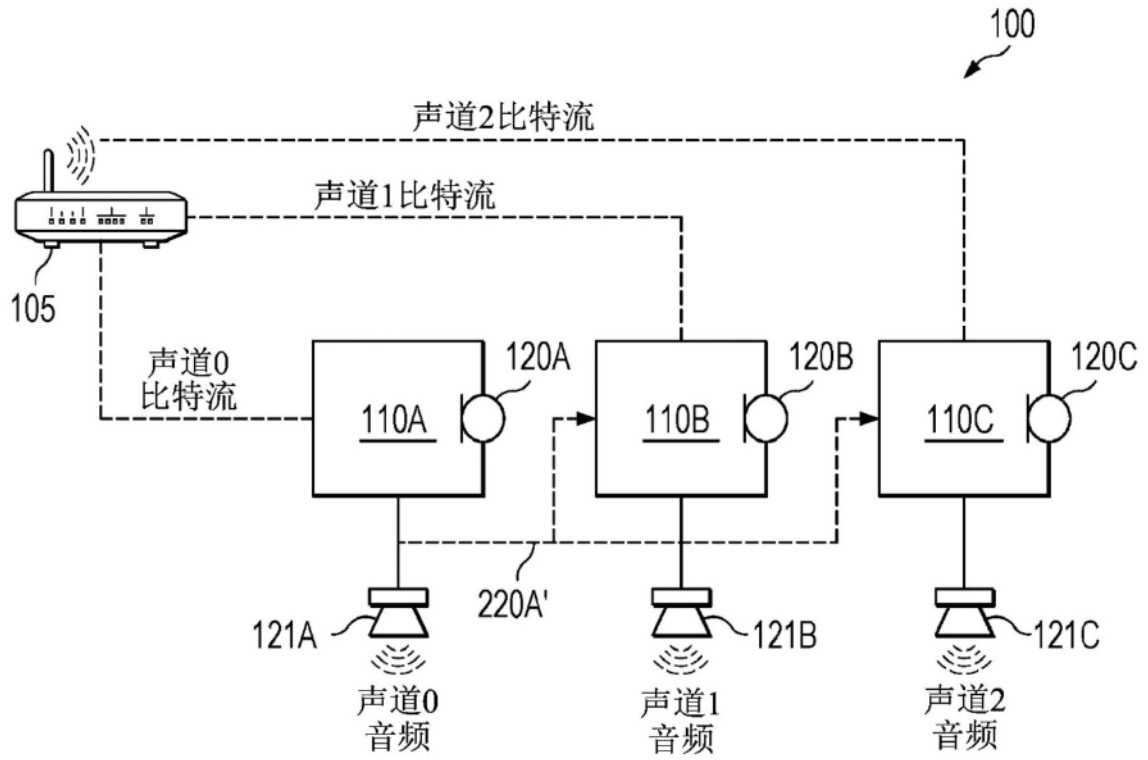


图2C

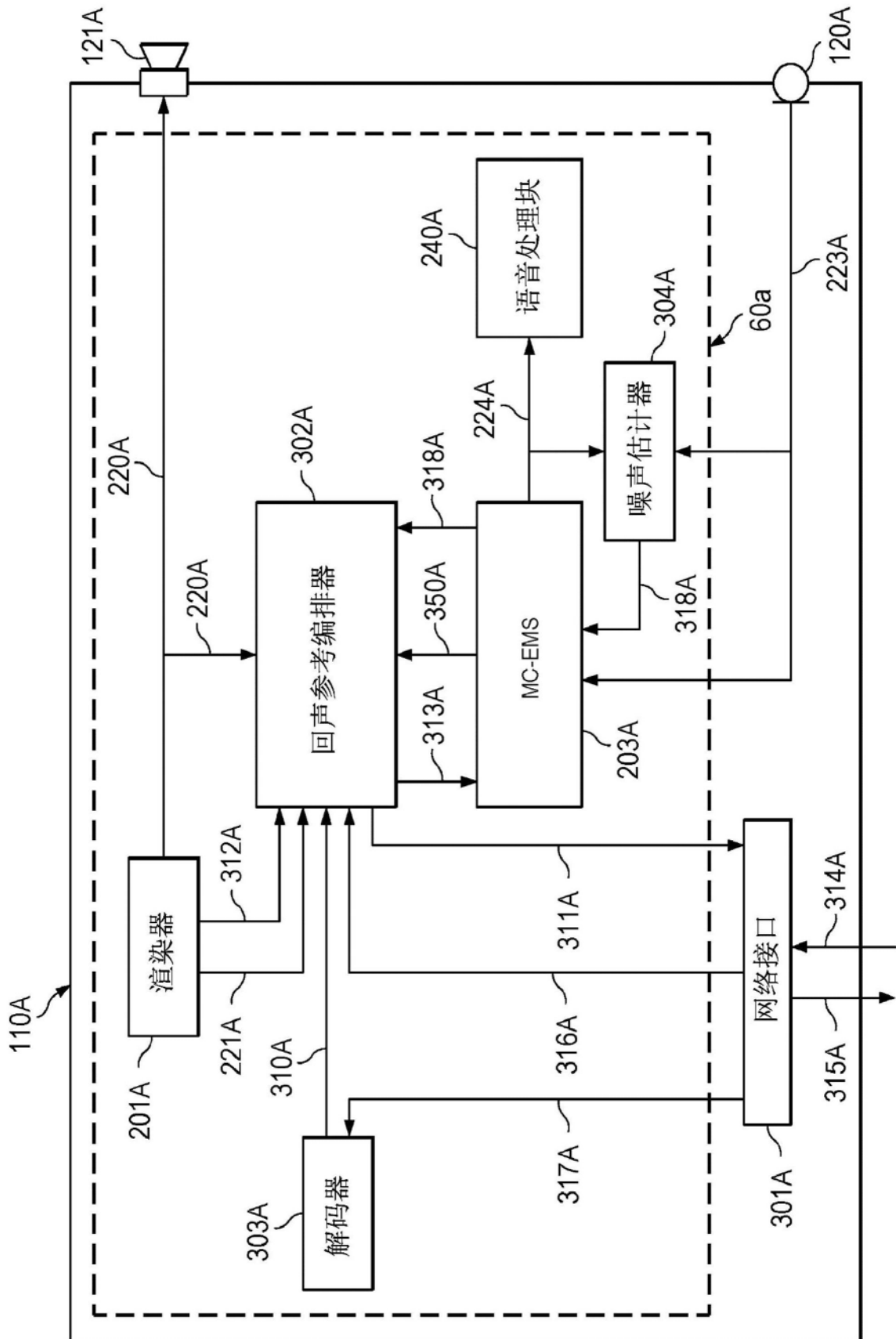


图3A

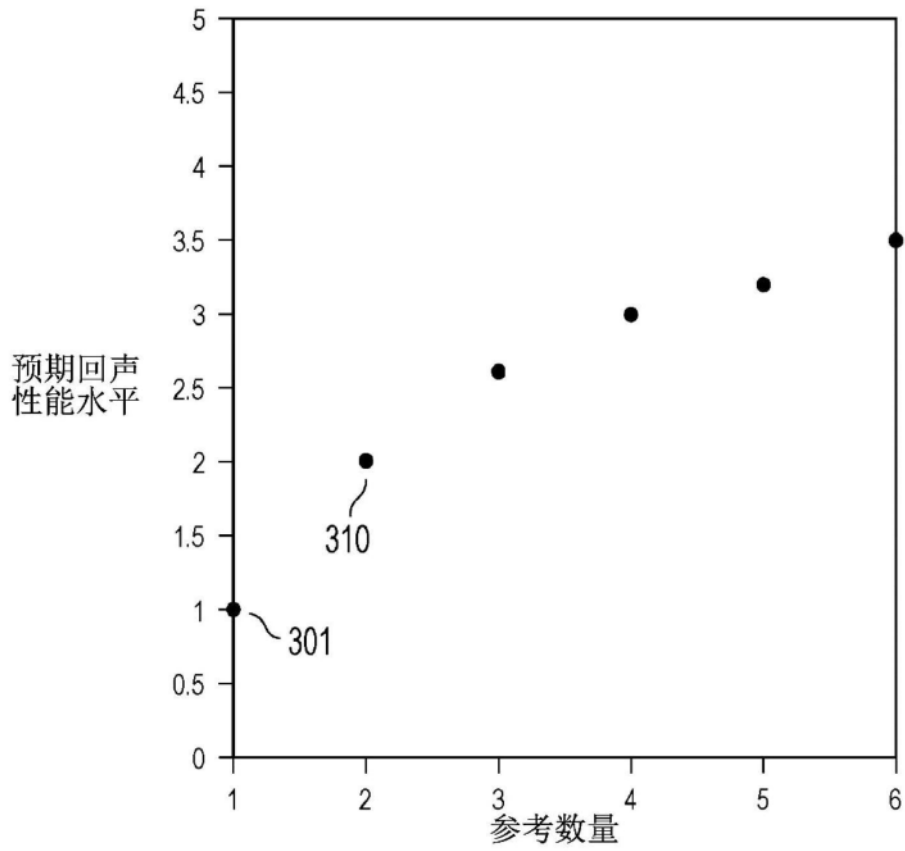


图3B

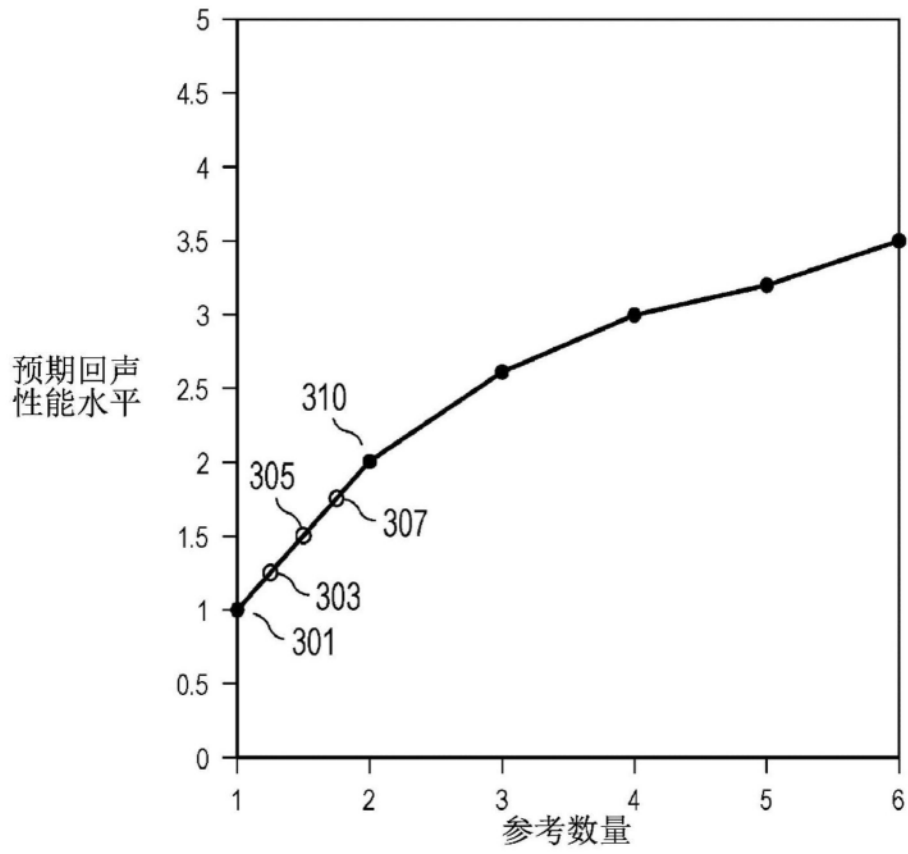


图3C

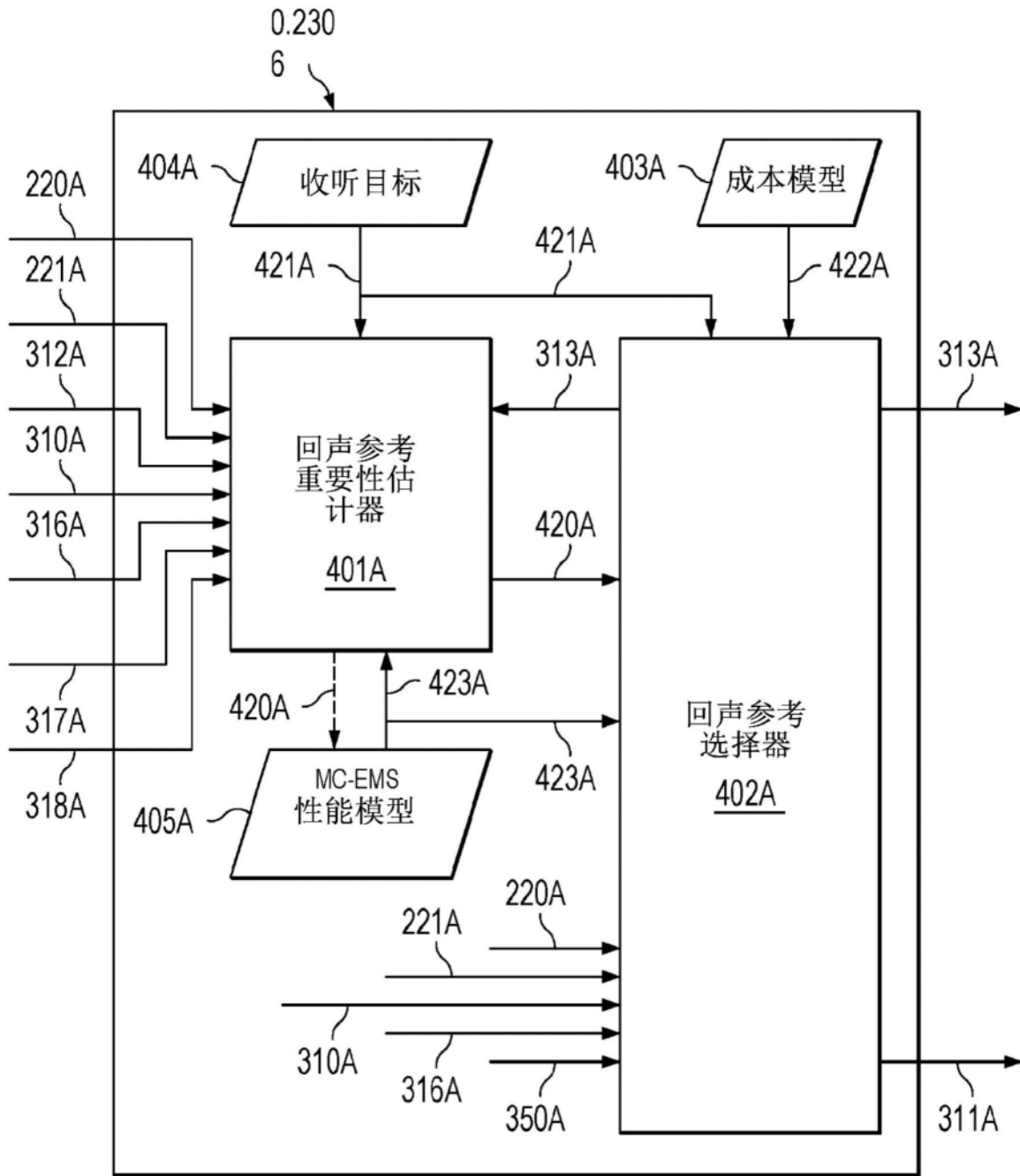


图4

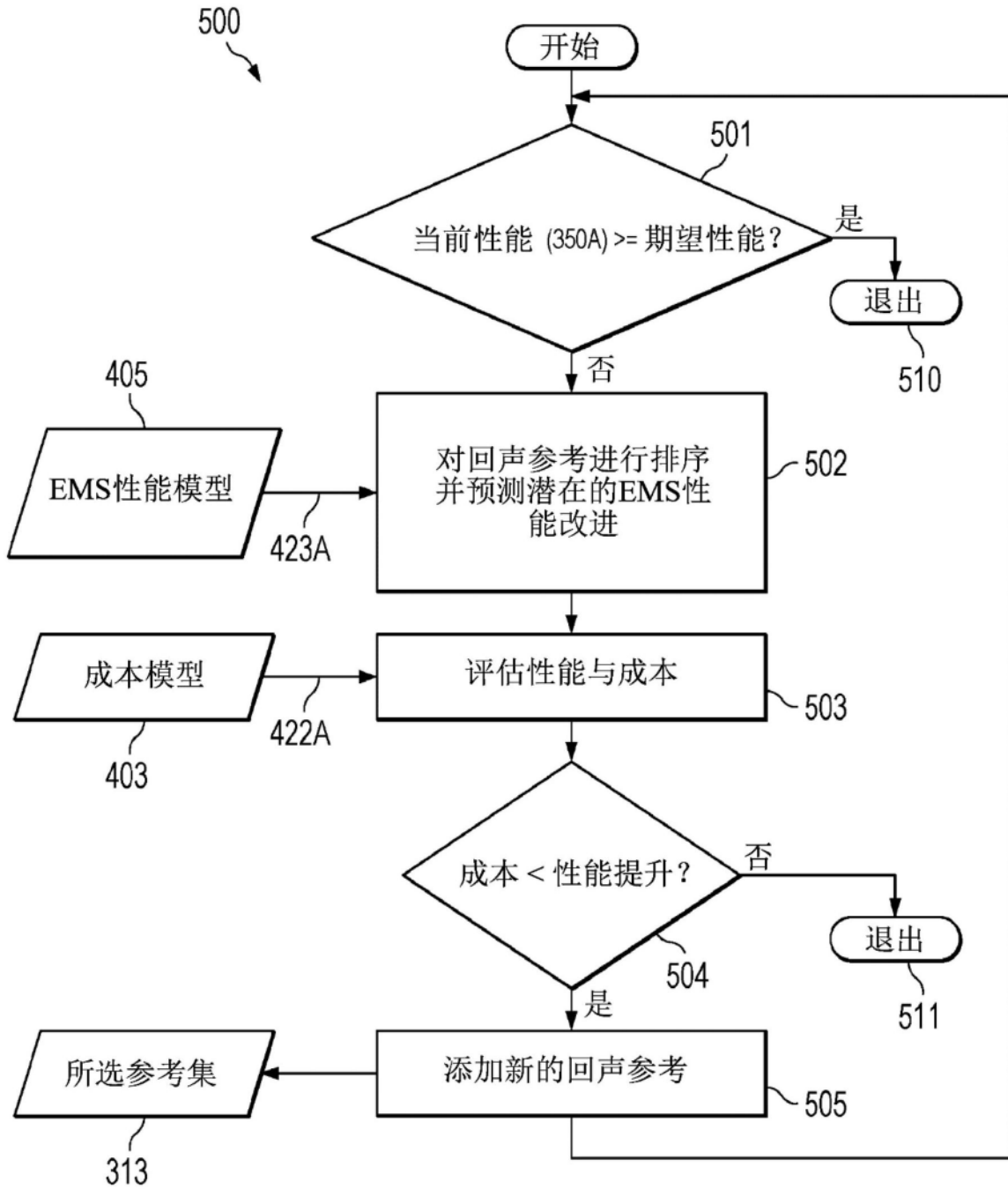


图5A

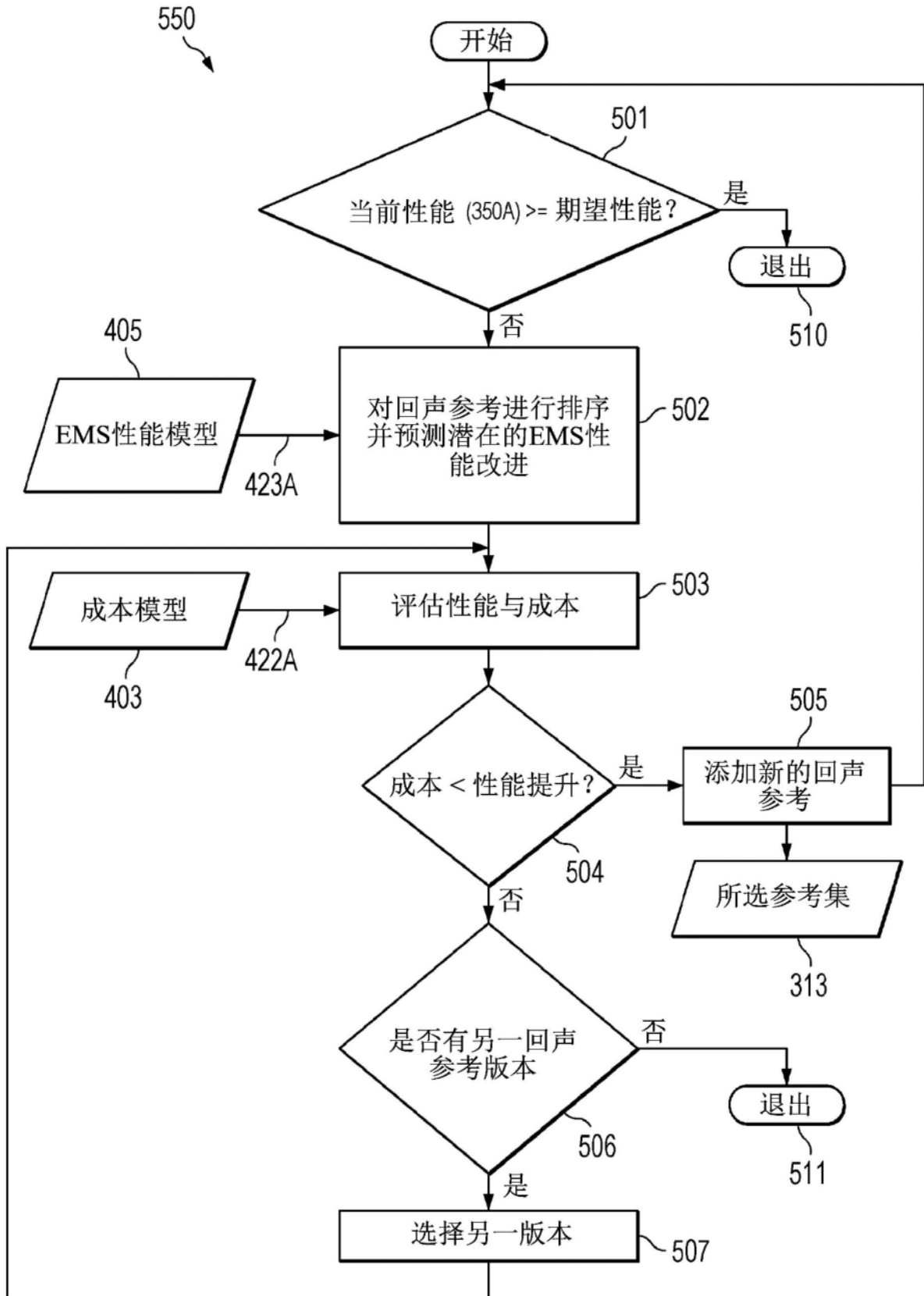


图5B

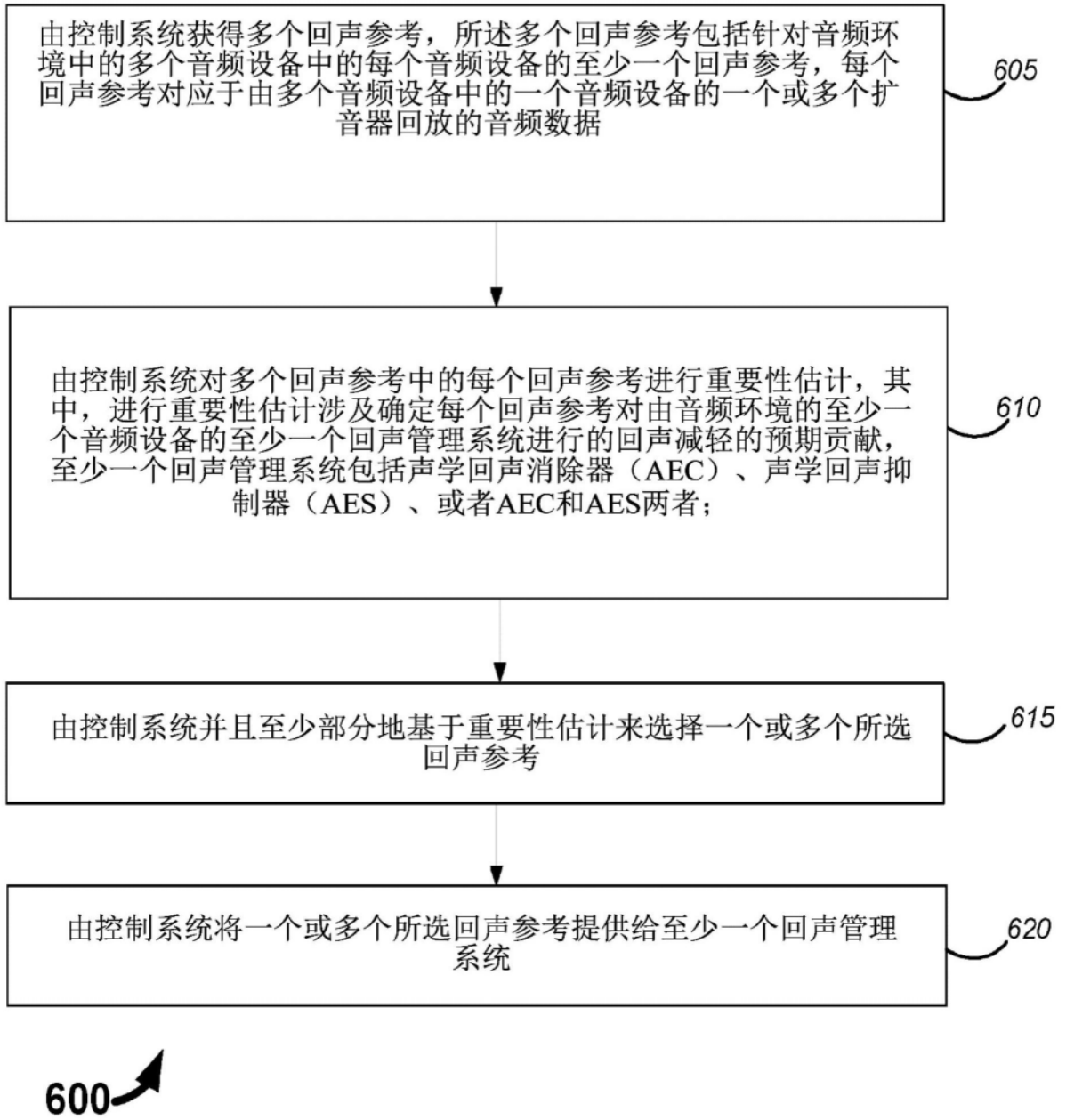


图6

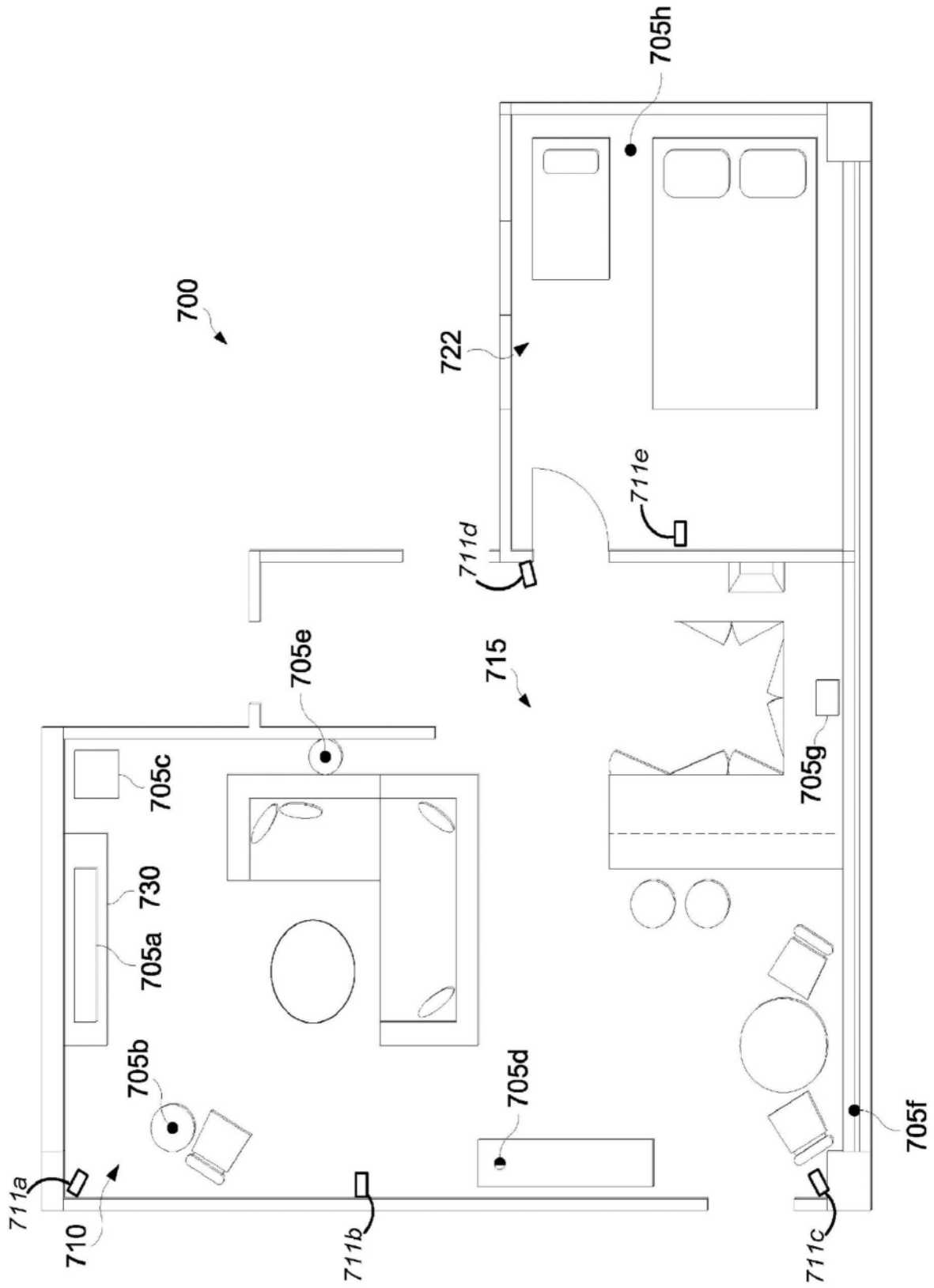


图7