



(12)发明专利

(10)授权公告号 CN 105794226 B

(45)授权公告日 2020.01.03

(21)申请号 201480066176.5

(22)申请日 2014.11.28

(65)同一申请的已公布的文献号  
申请公布号 CN 105794226 A

(43)申请公布日 2016.07.20

(30)优先权数据  
14/098,283 2013.12.05 US

(85)PCT国际申请进入国家阶段日  
2016.06.03

(86)PCT国际申请的申请数据  
PCT/US2014/067803 2014.11.28

(87)PCT国际申请的公布数据  
W02015/084683 EN 2015.06.11

(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 D·弗洛伦西澳 A·耶莱佩迪

(74)专利代理机构 永新专利商标代理有限公司  
72002  
代理人 刘瑜 王英

(51)Int.Cl.  
H04R 3/00(2006.01)  
H04S 7/00(2006.01)  
H04M 3/00(2006.01)  
H04M 9/08(2006.01)

审查员 张彩霞

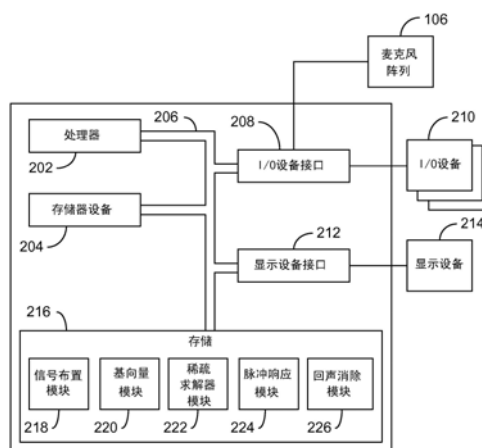
权利要求书3页 说明书8页 附图7页

(54)发明名称

用于声学回声消除的房间脉冲响应估计

(57)摘要

描述了用于估计音频源与麦克风阵列之间的房间脉冲响应的各种方法和系统。在一个示例中,方法包括接收麦克风阵列中的麦克风处的音频信号。所述音频信号对应于所述麦克风中的每个麦克风。所述方法还包括:基于所接收的音频信号,确定与所述麦克风阵列的几何结构兼容的子空间中的房间脉冲响应。



1. 一种用于估计音频源和麦克风阵列之间的房间脉冲响应的方法,包括:  
接收麦克风阵列中的每个麦克风处的音频信号,所述音频信号基于校准信号;  
估计所述麦克风阵列的几何结构;以及  
基于所估计的所述麦克风阵列的几何结构和所接收的音频信号来确定子空间中的房间脉冲响应,所接收的音频信号是要至少部分地基于基向量来建模的,所述基向量包括采样间隔,从所述音频源到所述麦克风阵列的中心的传播延迟,信道长度,所述麦克风阵列中的单独的麦克风,以及从所接收的音频信号的到达角。
2. 根据权利要求1所述的方法,包括使用所估计的房间脉冲响应来对声学回声消除进行初始化。
3. 根据权利要求1所述的方法,包括使用稀疏求解器来确定满足重构误差的权重,并且其中,利用凸优化技术来对所述权重进行优化,并且其中,所述重构误差对应于建模误差加噪声。
4. 根据权利要求1所述的方法,其中,所述脉冲响应是使用依赖于快速傅里叶变换算法的技术来进行估计的。
5. 根据权利要求1所述的方法,包括计算与麦克风阵列几何结构兼容的子空间中的脉冲响应。
6. 根据权利要求1所述的方法,其中,所接收的音频信号是由基向量来表示的;其中,所述基向量被配置为表示将由所述麦克风阵列从未衰减源接收的实像信号或虚像信号。
7. 一种用于实施在音频源和麦克风阵列的布置中所经历的可听反馈的声学回声消除的系统,包括:  
处理器,所述处理器用于执行处理器可执行代码;  
麦克风阵列;以及  
存储设备,所述存储设备存储处理器可执行代码,其中,所述处理器可执行代码在被所述处理器执行时,使得所述处理器:  
接收麦克风阵列中的每个麦克风处的音频信号,所述音频信号基于校准信号;  
估计所述麦克风阵列的几何结构;  
将单源信号组定义为将同时影响所述麦克风阵列中的所有麦克风的单源信号,所述单源信号中的每个单源信号对应于单实源或单虚源的不同位置,其中,虚源代表反射;  
找到所接收的音频信号的近似作为所述单源信号组的加权组合;  
估计从所述音频源到所述麦克风阵列中的所述麦克风中的每个麦克风的房间脉冲响应,所述脉冲响应对应于所述加权组合并且所述房间脉冲响应是要至少部分地基于所述麦克风阵列的几何结构和基向量的,所述基向量包括采样间隔,从所述音频源到所述麦克风阵列的中心的传播延迟,信道长度,所述麦克风阵列中的单独的麦克风,以及从所接收的音频信号的到达角;以及  
使用所述脉冲响应,从所接收的音频信号中消除声学回声。
8. 根据权利要求7所述的系统,其中,所述处理器通过基向量表示所接收的音频信号;其中,所述基向量被配置为表示将由所述麦克风阵列从未衰减源接收的实像信号或虚像信号。
9. 根据权利要求7所述的系统,其中,所述单源信号组的所述加权组合是利用凸优化技

术来进行优化的。

10. 根据权利要求7所述的系统,其中,所述脉冲响应是使用依赖于快速傅里叶变换算法的技术来进行估计的。

11. 根据权利要求7所述的系统,其中,所接收的音频信号的结构是通过以下基向量来进行建模的:

$$b_{T,\theta} = \begin{bmatrix} l(0 - T - \tau(z_1, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_1, \theta)) \\ \vdots \\ l(0 - T - \tau(z_K, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_K, \theta)) \end{bmatrix} \cdot \text{其中, } l(t) \text{ 是所述音频源信号的连续时间版本的函数; } L_x$$

表示在每个麦克风处所接收的校准信号的长度;  $T_s$  是采样间隔;  $T$  是从所述源到所述麦克风阵列的中心的传播延迟;  $T$  是  $z_k$  和  $\theta$  的函数; 其中,  $z_k$  是所述麦克风阵列中的单独的麦克风; 并且  $\theta$  是从所接收的信号的到达角。

12. 根据权利要求7所述的系统,其中,所述脉冲响应是通过以下向量来进行建模的:

$$h_{T,\theta} = \begin{bmatrix} \text{sinc}\left(0 - \frac{T + \tau(z_1, \theta)}{T_s}\right) \\ \vdots \\ \text{sinc}\left(N - 1 - \frac{T + \tau(z_1, \theta)}{T_s}\right) \\ \vdots \\ \text{sinc}\left(0 - \frac{T + \tau(z_K, \theta)}{T_s}\right) \\ \vdots \\ \text{sinc}\left(N - 1 - \frac{T + \tau(z_K, \theta)}{T_s}\right) \end{bmatrix} \cdot \text{其中, } T_s \text{ 是采样间隔; } T \text{ 是从所述源到所述麦克风阵列}$$

的中心的传播延迟;  $N$  表示信道长度;  $T$  是  $z_k$  和  $\theta$  的函数; 其中,  $z_k$  是所述麦克风阵列中的单独的麦克风; 并且  $\theta$  是从所接收的信号的到达角。

13. 根据权利要求1所述的方法,其中,所接收的信号的结构是通过以下基向量来进行建模的:

$$b_{T,\theta} = \begin{bmatrix} l(0 - T - \tau(z_1, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_1, \theta)) \\ \vdots \\ l(0 - T - \tau(z_K, \theta)) \\ \vdots \\ l((L_x - 1)T_s - T - \tau(z_K, \theta)) \end{bmatrix} \cdot \text{其中, } l(t) \text{ 是所述音频源信号的连续时间版本的函数;}$$

$L_x$  表示在每个麦克风处所接收的校准信号的长度;  $T_s$  是采样间隔;  $T$  是从所述源到所述麦克风阵列的中心的传播延迟;  $T$  是  $z_k$  和  $\theta$  的函数; 其中,  $z_k$  是所述麦克风阵列中的单独的麦克风; 并且  $\theta$  是从所接收的信号的到达角。

14. 根据权利要求1所述的方法,其中,所述脉冲响应是通过以下向量来进行建模的:

$$h_{T,\theta} = \begin{bmatrix} \operatorname{sinc}\left(0 - \frac{T+\tau(z_1,\theta)}{T_s}\right) \\ \vdots \\ \operatorname{sinc}\left(N-1 - \frac{T+\tau(z_1,\theta)}{T_s}\right) \\ \vdots \\ \operatorname{sinc}\left(0 - \frac{T+\tau(z_K,\theta)}{T_s}\right) \\ \vdots \\ \operatorname{sinc}\left(N-1 - \frac{T+\tau(z_K,\theta)}{T_s}\right) \end{bmatrix};$$

其中,  $T_s$  是采样间隔;  $T$  是从所述源到所述麦克风阵列

的中心的传播延迟;  $N$  表示信道长度;  $T$  是  $z_k$  和  $\theta$  的函数; 其中,  $z_k$  是所述麦克风阵列中的单独的麦克风; 并且  $\theta$  是从所接收的信号的到达角。

## 用于声学回声消除的房间脉冲响应估计

### 背景技术

[0001] 声学回声消除 (AEC) 是从由麦克风采集的信号中消去由邻近该麦克风放置的扬声器辐射的信号的问题。AEC 已经是包括电话会议系统、免提电话等双向通信的重要方面。标准问题场景包括一个扬声器和一个麦克风, 并且传统解决方案是训练自适应滤波器以去除扬声器反馈的指示。由于问题的重要性, 增大收敛速度、检测收敛状态、降低计算复杂性等是研究继续推进的领域。

[0002] 第二场景最近已更为重要, 即, 用于家庭和汽车娱乐系统的语音控制。这些系统通常具有多信道音频再现, 例如, 在相当高的再现水平下操作的立体声或 5.1 环绕系统。用户 (即, 期望的声源) 可能离麦克风几米远, 并且噪声水平可能是显著的。

### 发明内容

[0003] 下文呈现了简化的发明内容以提供对本文所描述的一些方面的基本理解。该发明内容不是所要求保护的主题的广泛概述。该发明内容不旨在确定所要求保护的主题的关键或重要元素, 也不旨在划定所要求保护的主题的范围。该发明内容的唯一目的是以简化的形式呈现所要求保护的主题的一些概念来作为稍后呈现的更具体的描述的前奏。

[0004] 实施例提供了用于估计音频源与麦克风阵列之间的房间脉冲响应的方法。该方法包括接收麦克风阵列中的麦克风处的音频信号, 音频信号与麦克风中的每个麦克风相对应。该方法还包括基于所接收的音频信号来确定与麦克风阵列的几何结构兼容的子空间中的房间脉冲响应。

[0005] 另一实施例提供了用于估计音频源与麦克风阵列之间的房间脉冲响应的方法。该方法包括接收麦克风阵列中的麦克风处的音频信号, 音频信号与麦克风中的每个麦克风相对应。该方法还包括将单源信号组定义为将同时影响麦克风阵列中的所有麦克风的单源信号, 单源信号中的每个与单实源或单虚源的不同位置相对应。此外, 该方法包括找到所接收的音频信号的近似作为单源信号组的加权组合。该方法包括估计从音频源到麦克风阵列中的麦克风中的每个麦克风的房间脉冲响应, 脉冲响应与该加权组合相对应。该方法通过使用脉冲响应从所接收的音频信号中消除声学回声而结束。

[0006] 另一实施例提供了用于实施音频源和麦克风阵列的分布中所经历的可听反馈的声学回声消除的系统。该系统包括用于执行处理器可执行代码的处理器, 麦克风阵列以及存储处理器可执行代码的存储设备。当处理器可执行代码由系统处理器执行时, 其使得处理器接收麦克风阵列中的麦克风处的音频信号, 音频信号与麦克风中的每个麦克风相对应。处理器被配置为将单源信号组定义为将同时影响麦克风阵列中的所有麦克风的单源信号, 单源信号中的每个与单实源或单虚源的不同位置相对应。使得处理器找到所接收的音频信号的近似作为单源信号组的加权组合。处理器被配置为估计从音频源到麦克风阵列中的麦克风中的每个麦克风的房间脉冲响应, 脉冲响应与加权组合相对应。最终使得处理器使用脉冲响应从所接收的音频信号中消除声学回声。

## 附图说明

[0007] 通过参考附图,将更好地理解以下具体实施方式,附图包括所公开主题的许多特征的特定示例。

[0008] 图1是可听信号反射在其中建模为虚像源的房间的示意图。

[0009] 图2是可以用于使用稀疏阵列信号处理的AEC的计算设备的示例的框图。

[0010] 图3表示麦克风阵列和扬声器配置的系统模型。

[0011] 图4是到达平面1-d阵列处的可听信号的示意图。

[0012] 图5是用于接收音频信号并且估计脉冲响应的示例性方法的过程流程图。

[0013] 图6是用于使用稀疏阵列信号处理来消除声学回声的示例性方法的过程流程图。

[0014] 图7是示出可以存储用于使用稀疏阵列信号处理来消除声学回声的指令的有形计算机可读存储介质的框图。

## 具体实施方式

[0015] 考虑了在具有麦克风阵列和一个或多个扬声器的回响环境中的声学回声消除的问题。消除回声需要学习从多个扬声器到麦克风阵列中的每个麦克风的脉冲响应。这已经常规地针对每个扬声器在每个麦克风处单独地完成。然而,到达阵列处的信号共享共同结构,这可以被利用以提供脉冲响应估计。

[0016] 本文所呈现的实施例目标在于基于校准信号的初始房间传输函数(RTF)估计。本文所描述的技术可以被延伸到自适应AEC滤波器的更复杂问题,然而,大多数当前设置的麦克风阵列使用校准信号来初始化滤波器,并且稍后调整滤波器系数的子集。现有实施例涉及但是不限于初始RTF估计的更相关的问题。RTF的稀疏性是公知的,并且已经用于回声消除,参见J.Benesty, et al. "Adaptive algorithms for the identification of sparse impulse responses," Selected methods for acoustic echo and noise control, vol. 5, pp. 125-153, 2006. 还考虑了波束形成(或空间滤波)与回声消除之间的交互。参见W.Herbordt, et al. "Joint optimization of acoustic echo cancellation and adaptive beamforming," Topics in acoustic echo and noise control, pp. 19-50, 2006. 然而,在这些情况下,回声消除滤波估计本身不从阵列信息中获益。先前技术分别计算每个AEC滤波,并且不利用本公开内容中详述的阵列结构。

[0017] 作为正文前的图文,附图中的一些图描述了一个或多个结构部件的上下文中的概念,这些部件被称作功能、模块、特征、元素等。可以例如通过软件、硬件(例如,分立的逻辑部件等)、固件等或这些实施方式的任何组合,以任何方式来实施附图中所示的各部件。在一些实施例中,各部件可以反映对应部件在实际实施方式中的使用。在其它实施例中,可以通过多个实际部件来实施附图中图示的任何单个部件。附图中的任何两个或更多个单独部件的描绘可以反映由单个实际部件执行的不同功能。

[0018] 其它附图以流程图形式描述了概念。在该形式中,将某些操作描述为构成以某一顺序执行的分立框。这样的实施方式是示例性而非限制性的。本文所描述的特定框可以集合在一起并且在单个操作中执行,某些框可以被拆分为多个组成框,并且某些块可以以不同于本文所图示的顺序的顺序来执行,包括执行框的并行方式。流程图中所示的框可以通过软件、硬件、固件、手工处理等或这些实施方式的任何组合来实施。如本文所使用,硬件可

以包括计算机系统、分立逻辑部件(例如,专用集成电路(ASIC))等以及其组合。

[0019] 至于术语,短语“被配置为”包含任何种类的结构部件可以被构建为执行所标识的操作的任何方式。结构部件可以被配置为使用软件、硬件、固件等或其任何组合来执行操作。

[0020] 术语“逻辑”包含用于执行任务的任何功能。例如,流程图中图示的每个操作对应于用于执行该操作的逻辑。可以使用软件、硬件、固件等或其任何组合来执行操作。

[0021] 如本文所使用,术语“部件”、“系统”、“客户端”等旨在指代计算机相关的实体,或者硬件、软件(例如,执行中)和/或固件,或其任何组合。例如,部件可以是运行在处理器上的进程、对象、可执行文件、程序、功能、库、子例程和/或计算机或软件和硬件的组合。通过例示的方式,运行在服务器上的应用和服务器两者可以是部件。一个或多个部件可以驻留于进程内,并且部件可以本地化到一个计算机上和/或分布于两个或更多个计算机之间。

[0022] 此外,所要求保护的主体可以被实施为方法、装置、或制品,其使用标准编程和/或工程技术来生产软件、固件、硬件或其任何组合以控制计算机实施所公开的主题。本文所使用的“制品”旨在包含从任意有形计算机可读设备或介质可访问的计算机程序。

[0023] 计算机可读的存储介质可以包括但不限于磁存储设备(例如,硬盘、软盘、和磁条等)、光盘(例如,压缩盘(CD)和数字通用盘(DVD)等)、智能卡以及闪存存储器设备(例如,卡、棒、以及键驱动等)。相比而言,计算机可读介质通常(即,不是存储介质)可以另外地包括诸如用于无线信号的传输介质等通信介质。

[0024] 本文所描述的技术的实施例涉及但是不限于房间传输函数估计。本实施例考虑将阵列信息并入回声消除滤波器的估计中以实现声学回声消除的期望结果。

[0025] 图1是可听信号反射110在其中建模为虚像源108的房间100的示意图。扬声器102在所有方向上发出在房间100内回响的声波104,声波104在在麦克风阵列106处被接收之前由墙和其它表明反射。每次反射110对应于在本文所描述的示例中用于仿真房间100内的声响的像源108。在麦克风阵列106处接收的信号为沿着发射信号的扬声器102与麦克风阵列106处的接收器之间的多个路径所接收的信号的叠加。

[0026] 本技术的实施例合并了利用麦克风阵列结构106以及到达阵列处的反射110的稀疏性的算法,以便于形成对每个扬声器102和麦克风106之间的脉冲响应的估计。该算法被示出为关于合成数据和真实数据两者相对于匹配的滤波器算法提高了性能。实施例使用已知麦克风阵列几何结构来改善在每个麦克风处的房间传输函数(RTF)估计。这通过开发RTF中出现的规律来实现,这是由于所有的麦克风位于相同房间、从相同扬声器接收信号并且从相同障碍物反射的事实。尽管我们假设麦克风阵列几何结构是已知的,但是应该注意麦克风阵列106的几何结构不需要在收集本文所描述的信号之前已知。换言之,不必预先知道阵列的配置,这是因为可以使用所接收的信号来估计它。熟悉本领域的技术人员将知道如何使用相关和多个接收信号来估计阵列几何结构。

[0027] 图2是可以用于使用稀疏阵列信号处理的AEC的计算设备200的示例的框图。计算系统200可以是例如结合 Microsoft® 的 Xbox 360® 或 Xbox One® 平台使用的 Xbox Kinect®, 或类似的音频采集设备。在一些实施例中,计算系统200可以是例如由观察者使用以提供反馈的台式计算机。在另一实施例中,计算系统200可以并入许多较新的汽车中的仪表盘显示器和用户界面内。计算系统200可以包括适于执行所存储指令的处理器202,以

及存储可以由处理器202执行的指令的存储器设备204。处理器202可以是单核处理器、多核处理器、计算集群或任何数量的其它配置。存储器设备204可以包括随机访问存储器(例如, SRAM、DRAM、零电容器RAM、SONOS、eDRAM、EDO RAM、DDR RAM、RRAM、PRAM等)、只读存储器(例如,掩膜ROM、PROM、EPROM、EEPROM等)、闪速存储器或任何其他适当的存储器系统。由处理器202执行的指令可以用于基于稀疏地表示的阵列使用信号处理的声学回声消除。

[0028] 处理器202可以通过系统总线206(例如,专用总线、PCI、ISA、PCI-Express、HyperTransport®、NuBus等)连接到适于将计算系统200连接到一个或多个I/O设备210的输入/输出(I/O)设备接口208。I/O设备210可以包括例如照相机、手势识别输入设备、键盘、指向设备、语音识别设备以及网络接口等。指向设备可以包括触摸板或触摸屏等。在本公开内容中,重要的I/O设备为连接到计算系统200的麦克风阵列106。I/O设备210可以是计算系统200的内置部件,或者可以是外部地连接到计算系统200的设备。

[0029] 处理器202还可以通过系统总线206链接到适于将计算系统200连接到显示设备214的显示设备接口212。显示设备214可以包括作为计算系统200的内置部件的显示屏。显示设备214还可以包括外部连接到计算系统200的计算机监控器、电视机或投影仪等。

[0030] 存储216可以通过总线206耦合到处理器202。存储216可以包括硬盘驱动器、固态驱动器、光盘驱动器、USB闪速驱动器、驱动器阵列或其任何组合。存储216可以包括被配置为使用本文所描述的稀疏阵列信号处理来实施声学回声消除的多个模块。例如,存储216可以包括被配置为将在麦克风阵列106的麦克风中的所有麦克风处接收的信号110布置为单个大向量的信号布置模块218。

[0031] 存储216还可以包括基向量模块220,用于计算合并了从所接收信号到麦克风阵列106的源距离以及所接收信号到麦克风阵列106的角的基向量。对于每个可能的角和每个可能的延迟时间(其可以计算为距离),如果基向量已经从特定角和延迟到达阵列的中心,基向量可以被定义为扬声器信号,其对于麦克风阵列106中的每个麦克风累计。在实施例中,在可以生成脉冲响应之前,所有可能的角和延迟被离散化并且经历稀疏求解器,其消除麦克风阵列106处经历的声学回声。因此,所接收的信号成为基向量中的少许基向量的加权组合(少许是因为角和延迟的可能数量比在麦克风阵列106处实际接收的那些角和延迟的数量大得多)。

[0032] 存储106中还可以包括稀疏求解器模块222。因为有数百万个角和延迟,并且因为它们中的每个需要由基向量表示,所以基矩阵的总大小是巨大的。每个可能的位置将被表示,无论由基向量表示的这些信号是实还是虚。通过凸优化过程,稀疏求解器模块222使用加权的基向量并且确定满足特定重构误差的最小值的权重。优化步骤可以例如是基追踪降噪(BPDN)、用于解决具有稀疏约束的近似表示问题的公知的方法。稀疏求解器还可以包括诸如SPGL求解器、用于大规模1-范数正则化最小二乘法的Matlab™求解器(其依赖于矩阵-向量操作)之类的求解器。

[0033] 存储106中还可以包括脉冲响应模块224。脉冲响应模块224被配置为利用从稀疏求解器模块222中确定的基权重并且估计从扬声器102到麦克风阵列106中的麦克风中的每个麦克风的脉冲响应。可以使用与由基向量模块220计算的基向量相对应的基向量来估计脉冲响应。还可以通过脉冲响应模块224来配置对应的基矩阵,并且来自该基矩阵的向量可以包括按麦克风阵列106中的每个麦克风的顺序累计的脉冲响应估计。存储216还可以包括

回声消除模块226,其使用脉冲响应估计来消除在麦克风阵列106处接收的声学回声。回声消除模块226对每个麦克风处接收的信号进行过滤并且求解AEC问题。

[0034] 将理解图2的框图不旨在指示计算系统200包括图2中所示的部件中的所有部件。相反,计算系统200可以包括较少部件或图2中未图示的另外的部件,例如,另外的应用、另外的模块、另外的存储器设备、另外的网络接口(未示出)等。此外,计算系统200不限于被示为用于实施可以被实施的这些功能的代码的任何组合的模块。

[0035] 图3表示麦克风阵列和扬声器配置的系统模型。在校准阶段期间,训练信号 $l(t)$ 在扬声器302处播放。我们使用离散时间记号并且使用 $n$ 作为我们的时间索引。因此,我们将离散时间版本的 $l(t)$ 表示为 $l[n]$ 。示例性实施例利用多个扬声器,并且该步骤必须分别对于每个扬声器重复。从扬声器302到第 $k$ 个麦克风304的RTF被表示为 $h_k[n]$  306,并且被假设为时间不变的,但是可以执行自适应来追踪随时间变化的特征。因此,通过等式(1)来表示在麦克风 $k$  304处接收的信号如下:

$$[0036] \quad x_k[n] = \sum_{m=0}^{N-1} h_k[m]l[n-m] + d_k[n] + v_k[n] \quad (1)$$

[0037] 在等式(1)中, $m$ 为用于计算总和的虚设变量, $d_k[n]$ 是期望信号, $n$ 是时间索引的传统记号,并且 $v_k[n]$ 为干涉噪声。注意, $d_k[n]$ 假设在校准阶段期间为零。如何RTF  $h_k[n]$  306是已知的,则回声可以容易地从所接收的信号中减去。相反,消除滤波器 $h_k[n]$ 是通过计算由下式给出的输出信号 $y_k[n]$ 来计算的近似并且用于(部分地)从 $x_k[n]$ 去除回声:

$$[0038] \quad y_k[n] = x_k[n] - \sum_{m=0}^{N-1} \hat{h}_k[m]l[n-m]. \quad (2)$$

[0039] 在等式(2)中,相同的变量与等式(1)中所定义的变量相同。

[0040] 当只存在一个麦克风和一个扬声器时,RTF的最小均方误差估计已知为“匹配滤波器”,其可以计算为:

$$[0041] \quad h_k^{MF}[n] = \frac{\sum_m \tilde{l}[m]\tilde{x}_k[n+m]}{\sum_m (\tilde{l}[m])^2} \quad (3)$$

[0042] 在等式(3)中, $\tilde{l}[n]$ 指代 $l[n]$ 的白化版本,并且 $\tilde{x}_k[n]$ 指代在 $x_k[n]$ 由用于白化扬声器信号 $l[n]$ 的相同滤波器滤波后的 $x_k[n]$ 。满意的白化过程有利地转换数据使得它们具有同一协方差矩阵,其中,所有的采样为统计上独立的。等式(3)的消除滤波器相对于其来比较当前回声消除技术的基线计算。此外,通过开发麦克风阵列中的多个麦克风而在估计器308处获得的每个RTF的估计可能好于以上优化。

[0043] 图4是到达平面阵列400处的可听信号的示意图。沿着任何一个路径,信号具有到阵列中心的传播延迟,传播延迟取决于路径的长度以及(可能依赖于频率的)增益,增益取决于所用的反射110的属性。假设反射110作为平面波到达阵列106,图4中指示了沿着特定路径的所接收信号的结构。为了记号的简单,假设麦克风的线性阵列,并且麦克风沿着 $z$ 轴对齐,如图4所示,并且位于位置 $(z_1, z_2, \dots, z_k)$  106。对于影响线性阵列106的平面波,如果原点402处的持续时间信号为 $s(t)$ ,那么位于 $z_k$ 处的麦克风处的信号为 $s(t-T(z_k, \theta))$ ,其中, $T(z_k, \theta) = -z_k \cos(\theta)/c$ ,并且 $c$ 为声速。换言之,在每个麦克风106处由平面波产生的信号是相同的,除了延迟 $(T)$ , $(T)$ 只取决于阵列几何结构和到达的方向 $(\theta)$  404。

[0044] 每个反射110可以对应于虚像源108,如图1所指示。下面的等式(4)首先假设阵列106的维度足够小,使得波前可以近似为平的,还假设每个反射器为平坦衰落的,并且最后假设源(真实或像)的P数量足以提供对到达信号的近似。然后,在这些假设下,在麦克风阵列106的麦克风k处接收的信号可以写为:

$$[0045] \quad x_k[n] = \sum_{p=1}^P \alpha_p l \left( nT_s - T_p - \frac{z_k}{c} \cos \theta_p \right) + \tilde{v}_k[n] \quad (4)$$

[0046] 在等式(4)中, $T_p$ 和 $\alpha_p$ 分别为由波前p(包括传播损失以及麦克风和扬声器的指向性)经受的延迟和衰减。此外,c是声速, $\tilde{v}_k[n]$ 包括实际噪声 $v_k[n]$ 以及可能任何未建模的信号组成两者,并且 $\theta_p$ 为沿着第p路径到达的信号与阵列轴所成的角。 $T_s$ 是采样间隔,并且 $z_k$ 是z轴中的第k个麦克风的位置,如之前所定义。单源信号组可以被定义为将同时影响麦克风阵列中的所有麦克风的单源信号,单源信号中的每个信号对应于单实源或单虚源的不同位置。

[0047] 等式(4)和图4假设麦克风阵列106是均匀并且线性对齐的。然而,该配置仅提供了示例并不要求实施本公开内容。实际上,该阵列不需要总体上是线性的。此外,等式(4)隐式地假设增益是不依赖于频率的,这也不是被要求的,而是在本文所描述的技术的实施方式中被假设的。

[0048] 图5是用于接收音频信号并且估计脉冲响应的示例性方法的过程流程图。方法500始于框502,此时,音频信号在一个或多扬声器处播放,并且对应信号在麦克风阵列中的麦克风处被接收,可能被环境噪声污染。该方法包括框504处的步骤,在框504处估计可能作为扬声器处播放的声音的结果而到达麦克风处的信号族。该方法还包括在框506处基于所接收的音频信号确定与麦克风阵列的几何结构兼容的子空间中的房间脉冲响应。

[0049] 图5的过程流程图不旨在指示方法500的步骤将以任何特定顺序执行,或者方法500的步骤中的所有步骤将包括在每个案例中。此外,方法500内可以包括任何数量的另外步骤,这取决于具体应用。

[0050] 图6是用于使用稀疏阵列信号处理来实施声学回声消除的方法500的示例的过程流程图。可以通过关于图2所描述的计算设备200来实施该方法。方法600开始于框602,在框602处,音频信号在麦克风阵列处被接收。然后在框604处布置所接收的信号,并且将其表示为单个大的数学向量。在框606处,基向量被定义为包括由扬声器针对与许多虚源相对应的多个听觉角度以及时间延迟而发送的信号。虚源或像源可以是可在在麦克风阵列处被接收之前在房间内回响的巨大数量的可能信号反射的表示。不是所有的可能向量将这样被分析,并且可以施加加权构造以提供更易于计算的数据集。

[0051] 因此,方法在框608处继续,在框608处,所接收的音频信号(无论虚或实)被解释为多个基向量的组合的加权函数。可以对基向量和对应的权重进行优化以产生所观察的向量在一些期望的回声空间中的最好表示。可以通过以下来对该优化进行建模:

$$[0052] \quad \hat{w} = \underset{w}{\operatorname{arg\,min}} \|w\|_1, \quad \text{s.t.} \quad \|Bw - x\|_2 \leq \sigma. \quad (5)$$

[0053] 在等式(5), $\hat{w}$ 是所计算的最好权重向量(即,使(5)中的表达式最小化的向量),B是描述期望的回声空间的大矩阵,并且通过构成所有单独的基向量来形成;x是所观察的向量;并且 $\sigma$ 是所允许的重构误差。在示例性实施例中, $\sigma$ 被设定为稍大于未建模成分的标准差

的值。更具体而言,尽管 $\sigma$ 的任何值将产生结果,但是在该示例中,对于某些 $\beta > 1$ , $\sigma$ 被设定为 $\sigma = \beta\sigma_0$ ,其中, $\sigma_0$ 是噪声标准差,并且 $\beta$ 是正整数。在另一实施例中, $\beta$ 可以被设定为1.5。

[0054] 由于可以表示可听信号的大量数据,矩阵B很可能为巨大的,并且很可能构成信号空间的超完备基。因此,找到解决(5)的高效方式并且找到处理超完备性的高效方式变得必要。这可以通过施加稀疏约束(嵌入在(5)中指示的范数-1测度中)并且使用稀疏求解器来完成。这允许B中的数据被存储并且更容易地被访问。在框610处,稀疏求解器用于确定满足特定重构误差的权重。如所指示,该重构误差可以理想地稍大于未建模成分的标准差。重构误差可以例如是接近于建模误差加噪声的数字。还可以实施清除过程以进一步提高性能。该清除通过使 $\|x - \sum \theta_{t,w_{\theta,t}} b_{\theta,t}\|_2$ 的值最小化来对(5)中非零(或在特定阈值之上)的所有系数 $w_{\theta,t}$ 进行再计算。其中, $b_{\theta,t}$ 是与具有延迟 $t$ 的以角 $\theta$ 到达的反射相对应的基向量,并且 $w_{\theta,t}$ 是与该基向量(以及在(5)中为非零的那些)相关联的权重。

[0055] 在框612处,估计从扬声器到麦克风阵列中的麦克风中的每个麦克风的脉冲响应。脉冲响应对应于由扬声器生成的对于声学回声的估计,并且在其在麦克风阵列处被接收时高效地用于消除掉这样的回声。脉冲响应估计可以被计算为例如以下向量:

$$[0056] \quad \tilde{h} = \sum_{\theta,t} \tilde{w}_{\theta,t} h_{\theta,t} \quad (6)$$

[0057] 在框614处,等式(6)的脉冲响应估计向量可以用于从在麦克风阵列的所有麦克风处接收到的音频信号的表示中去除声学回声。

[0058] 图6的过程流程图不旨在指示方法600的步骤将以任何特定的顺序执行,或者方法600中的步骤中的所有步骤将包括在任何案例中。此外,任何数量的额外步骤可以包括在方法600中,这取决于具体应用。

[0059] 图7是示出了用于使用稀疏阵列信号处理来执行AEC的有形计算机可读存储介质700的框图。有形计算机可读存储介质700可以由处理器702通过计算机总线704访问。此外,有形计算机可读存储介质700可以包括代码以指导处理器702来执行本文所公开技术的步骤。

[0060] 有形计算机可读存储介质700可以包括代码706,代码706被配置为指导处理器702来接收和处理到达麦克风阵列处的音频信号。此外,有形计算机可读存储介质700可以包括代码708,代码708被配置为指导处理器702来便于向量布置。例如,代码可以指示处理器702来定义所接收音频信号的可能表示的单个大向量。另一代码块710可以指示处理器来定义可以是虚源信号和实源信号的组合的基向量。代码块712然后可以指示处理器702来创建在基向量值上的加权函数,以便于将代表性数据的量降低为更可管理的大小。有形计算机可读存储介质700还可以包括代码714,代码714指示处理器702实施稀疏求解器技术以对由加权基向量所定义的依旧非常大量的数据施加稀疏约束。716处的代码块然后可以指导处理器702来计算麦克风阵列中的每个麦克风处的脉冲响应估计。然后通过代码块718来利用脉冲响应估计以消除在具有寄存器延迟的反射的音频信号的麦克风的音频系统中典型生成的声学回声。

[0061] 将理解,有形计算机可读存储介质700中可以包括图7中未示出的任何数量的另外的软件部件,这取决于具体应用。尽管已经用对结构特征和/或方法特定的语言来描述了本主题,但是将理解所附权利要求中定义的主题不必限于以上所描述的特定结构特征或方

法。相反,以上所描述的特定结构特征或方法被公开为实施权利要求的示例性形式。

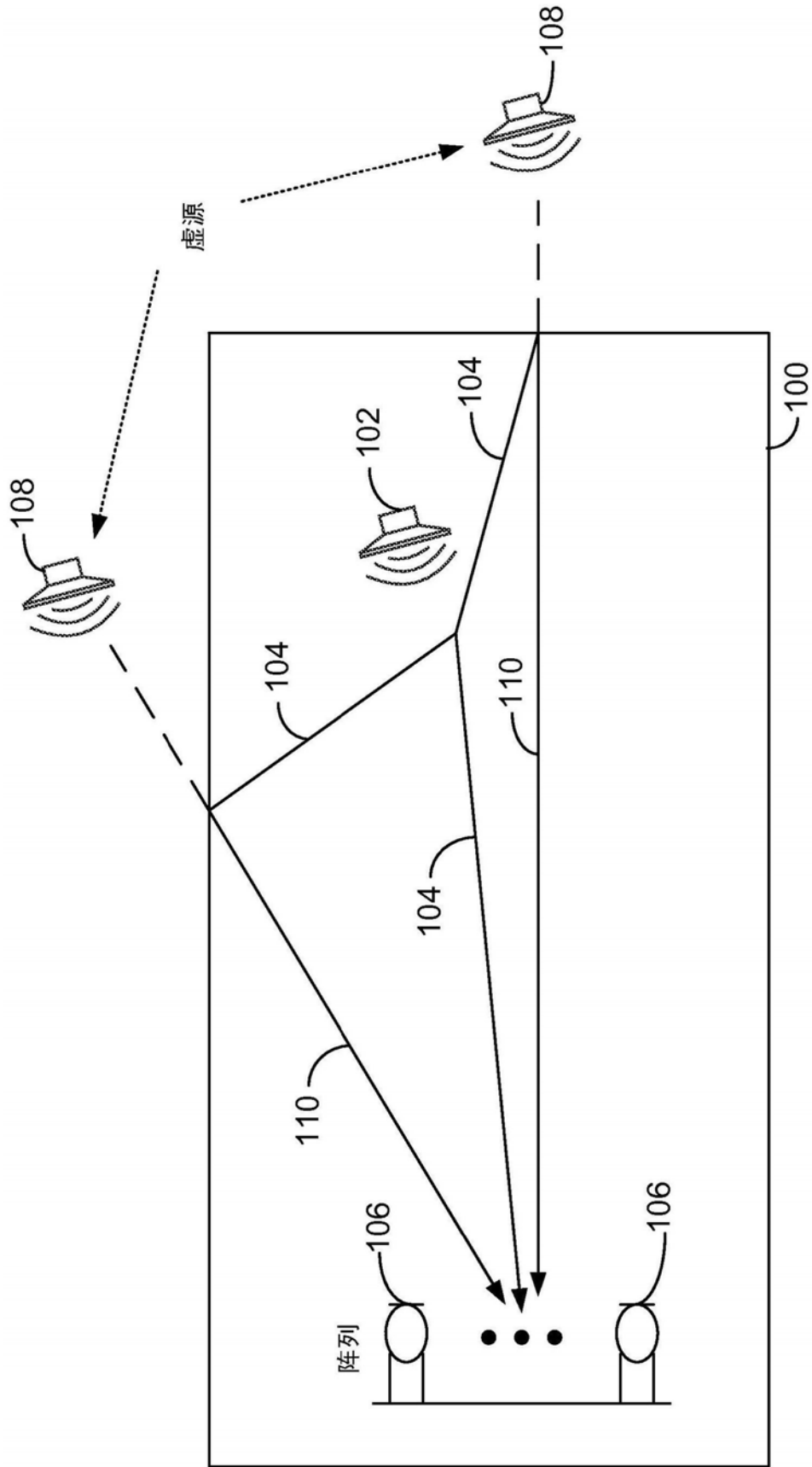
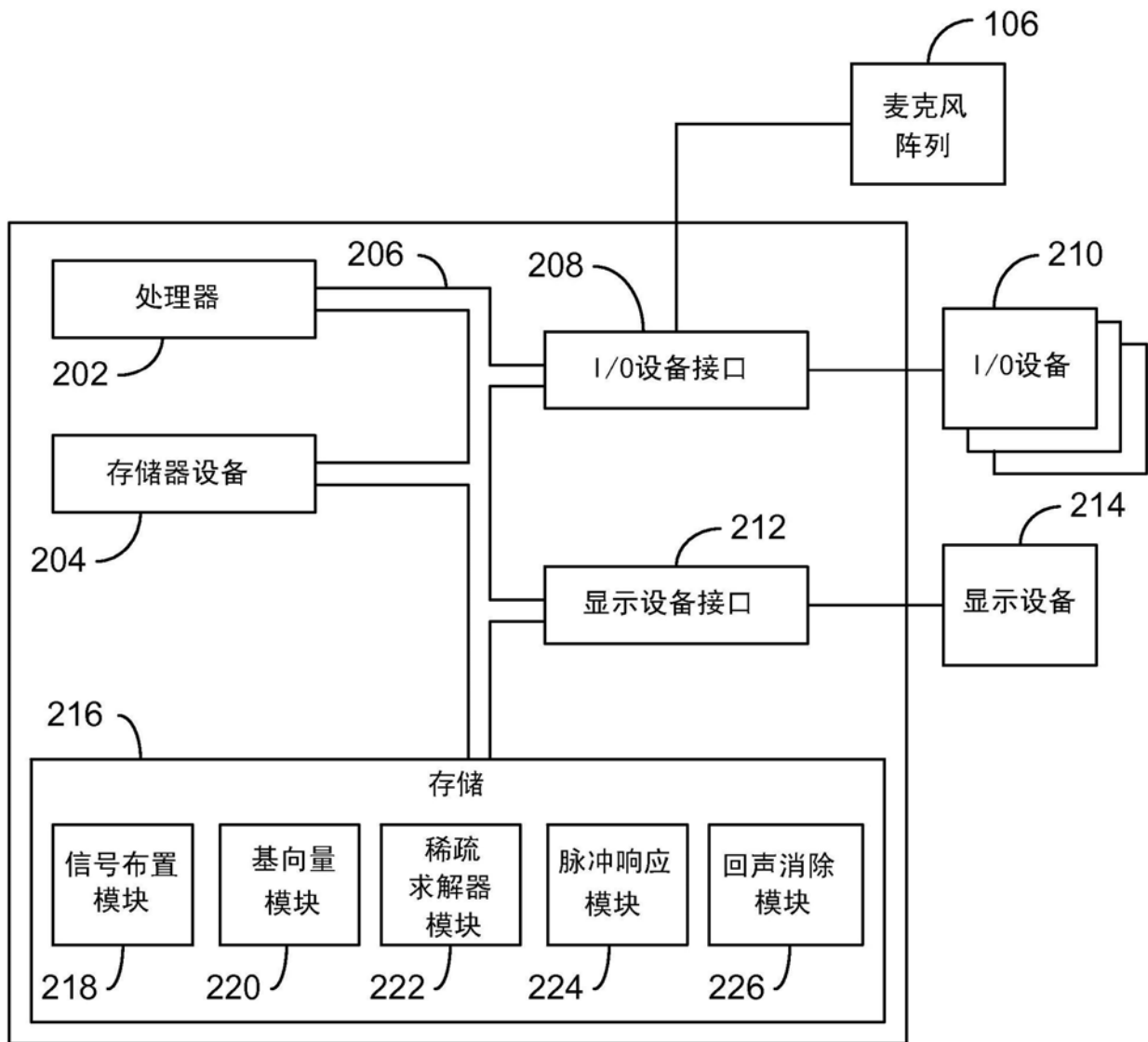


图1



200

图2

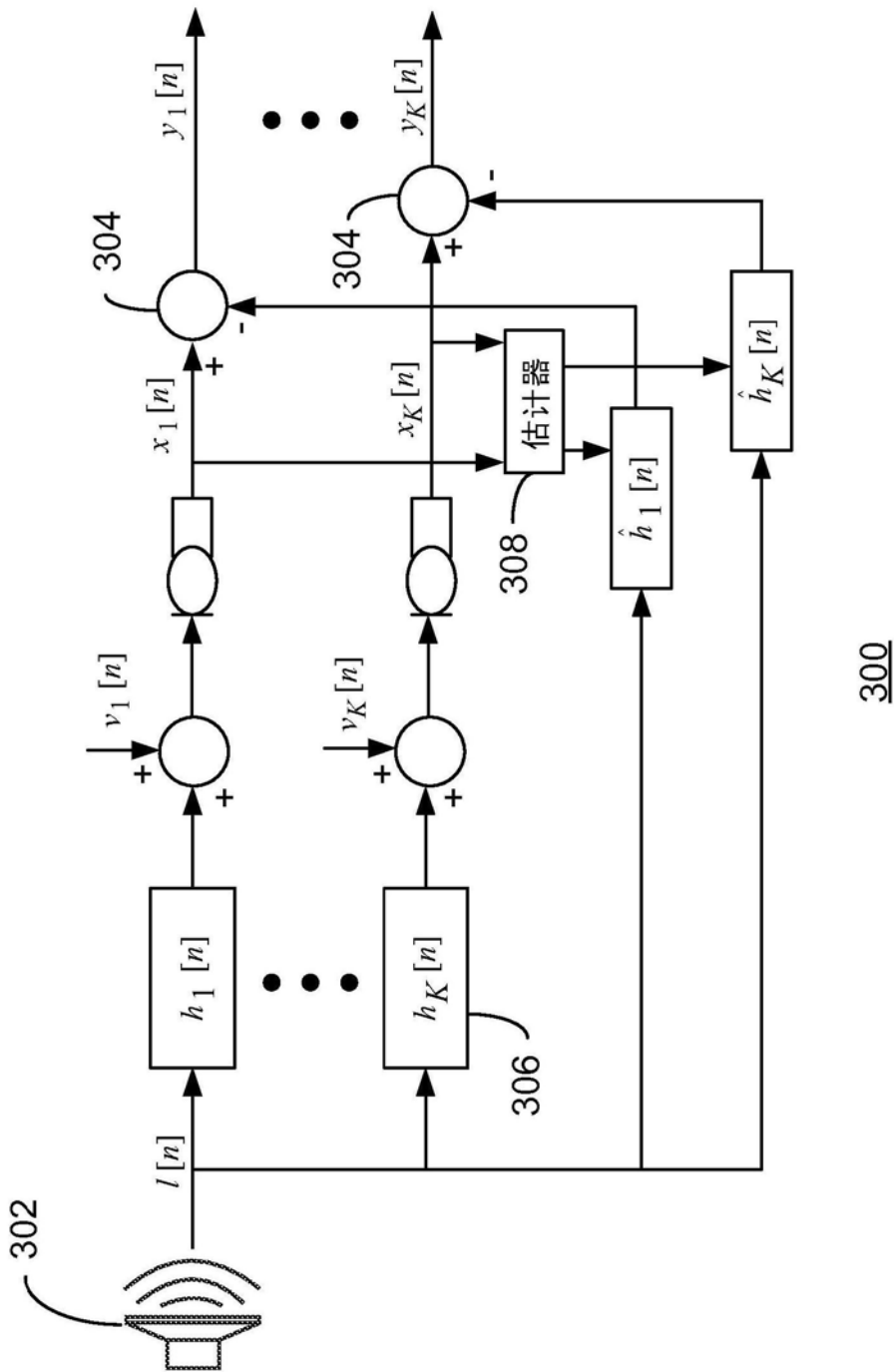
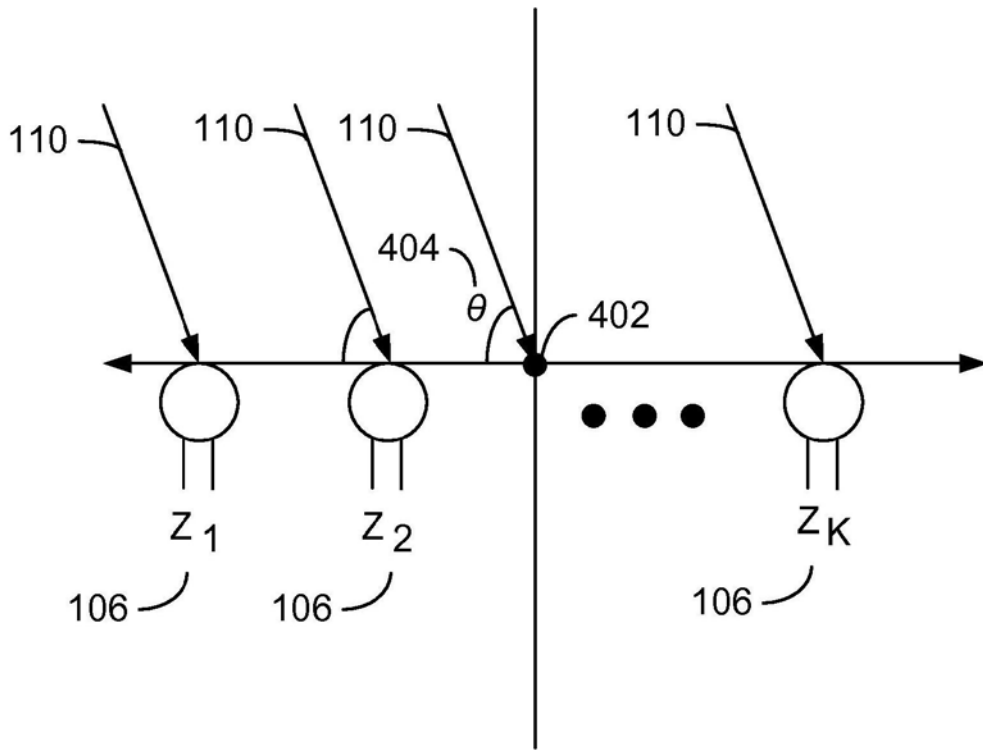
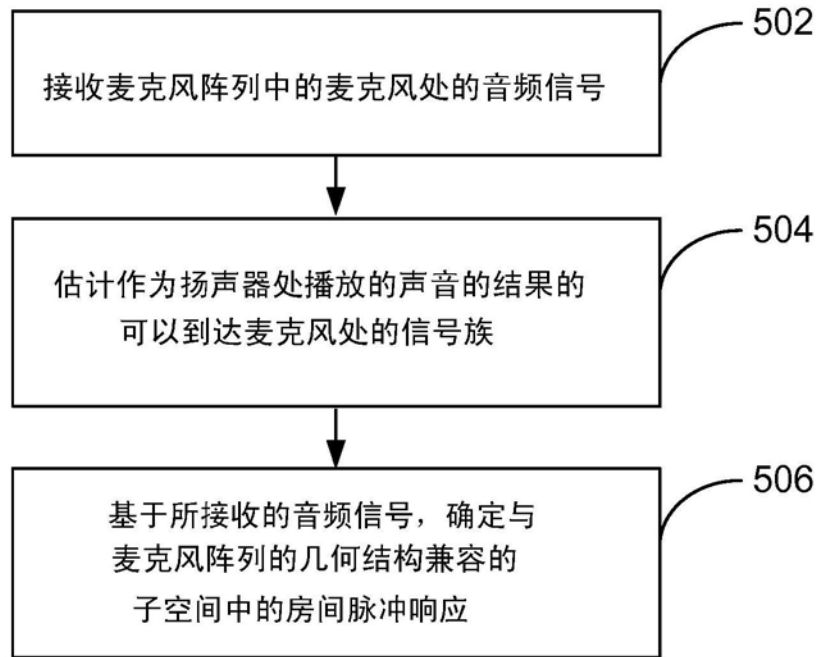


图3



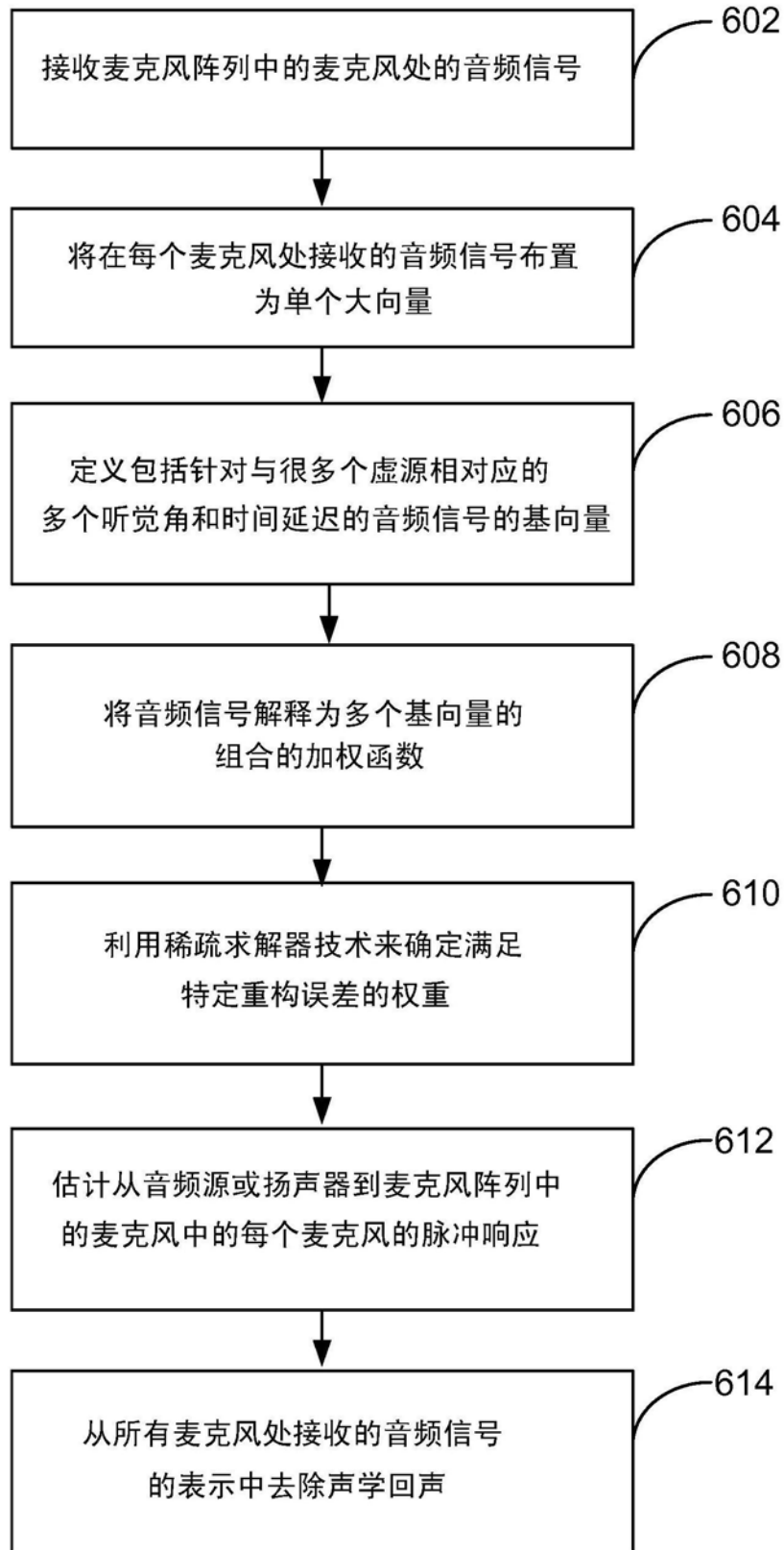
400

图4



500

图5



600

图6

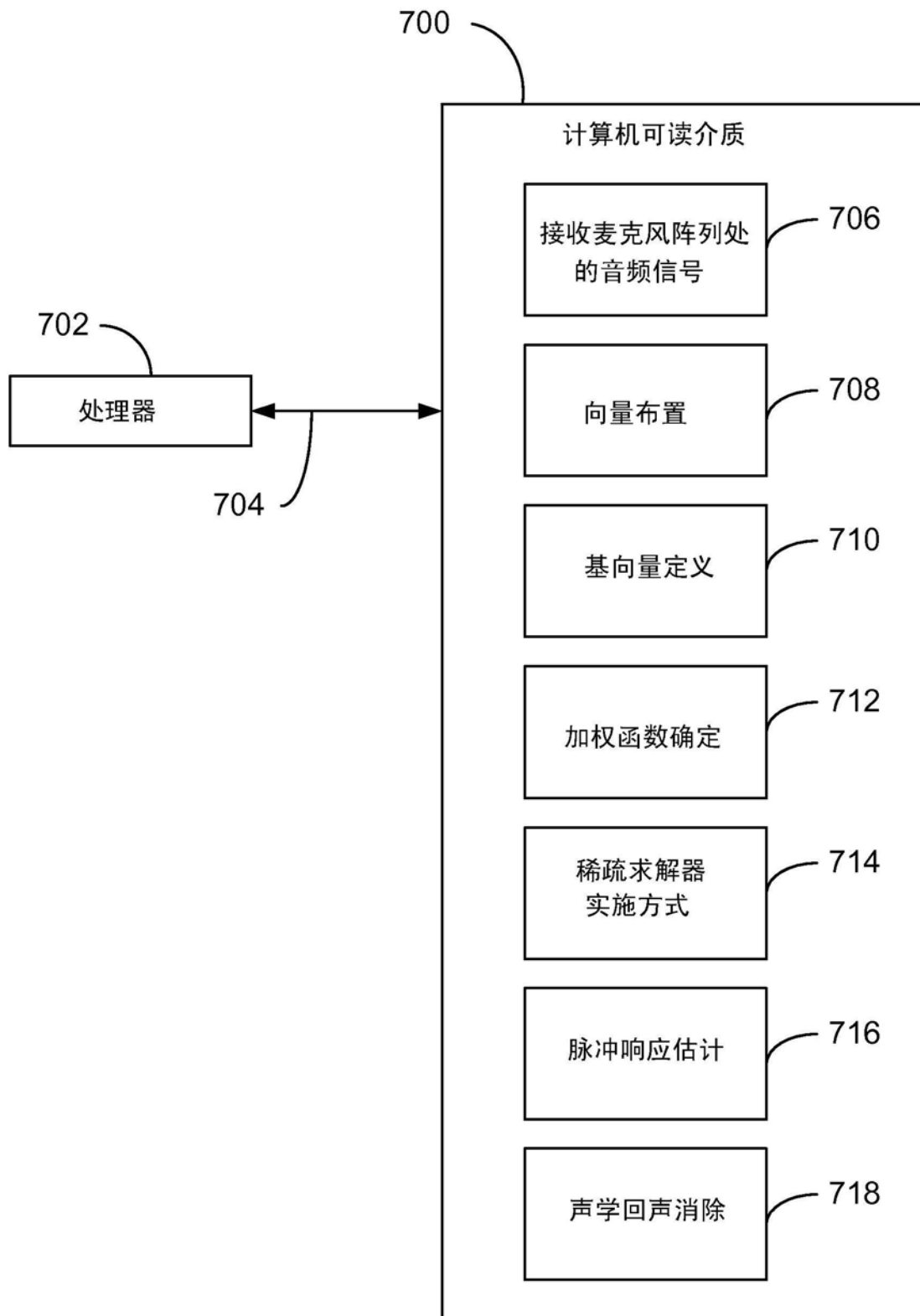


图7