



(12) 发明专利

(10) 授权公告号 CN 106716526 B

(45) 授权公告日 2021.04.13

(21) 申请号 201580047111.0

(22) 申请日 2015.08.25

(65) 同一申请的已公布的文献号
申请公布号 CN 106716526 A

(43) 申请公布日 2017.05.24

(30) 优先权数据
14306365.9 2014.09.05 EP
14306947.4 2014.12.04 EP(85) PCT国际申请进入国家阶段日
2017.03.02(86) PCT国际申请的申请数据
PCT/EP2015/069417 2015.08.25(87) PCT国际申请的公布数据
W02016/034454 EN 2016.03.10(73) 专利权人 交互数字麦迪逊专利控股公司
地址 法国巴黎(72) 发明人 光坎沃·堂 皮埃尔·贝尔泰
埃里克·扎伯拉
米歇尔·凯尔德朗瓦(74) 专利代理机构 中科专利商标代理有限责任
公司 11021
代理人 潘剑颖

(51) Int.Cl.

G10L 21/0208 (2013.01)

G10L 21/0216 (2013.01)

G10L 21/0272 (2013.01)

(56) 对比文件

CN 102265642 A, 2011.11.30

CN 102164328 A, 2011.08.24

CN 101510426 A, 2009.08.19

CN 102324237 A, 2012.01.18

CN 103282961 A, 2013.09.04

CN 102831898 A, 2012.12.19

US 2011/0307251 A1, 2011.12.15

US 2007/0150268 A1, 2007.06.28

WO 2009/048291 A1, 2009.04.16

US 2012/0327115 A1, 2012.12.27

US 2007/0021958 A1, 2007.01.25

US 2010/0241426 A1, 2010.09.23

李芳兰 等. 采用可调波束形成器的GSC麦克风
风阵列语音增强方法.《厦门大学学报(自然科学
版)》.2013,第52卷(第2期),第186-189页.Joe C. Chen et al..Source
localization and beamforming.《IEEE Signal
Processing Magazine》.2002,第30-39页.

审查员 李召卿

权利要求书2页 说明书7页 附图6页

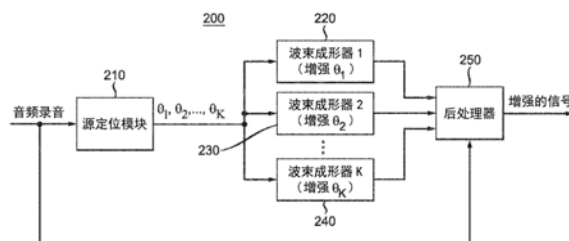
(54) 发明名称

用于增强声源的方法和装置

(57) 摘要

录音通常是来自若干声源的信号的混合。录音中的显著源的方向可以是已知的或使用源定位算法来确定。为了隔离或专注于目标源,可以使用多个波束成形器。在一个实施例中,每个波束成形器指向显著源的方向,并且来自波束成形器的输出被处理以专注于目标源。根据指向目标源的波束成形器的输出是否大于其他波束成形器的输出,可以使用参考信号或指向目标源的波束成形器的缩放的输出来确定与目标源相对应

的信号。缩放因子可以取决于指向目标源的波束成形器的输出和其他波束成形器的输出的最大值的比率。



1. 一种在音频处理装置中执行的方法,所述方法包括:

处理音频信号以生成至少两个输出信号,所述音频信号是来自至少两个音频输入的输入信号的混合,每个输出信号是通过使用指向不同空间方向的波束成形器生成的;

生成针对第一空间方向的第一增强信号,所述第一空间方向是由用于生成所述至少两个输出信号中的第一输出信号的波束成形器所指向的空间方向,当所述第一输出信号是所述至少两个输出信号中的显著输出信号时,所述第一增强信号是所述音频输入中的第一音频输入所捕获的所述输入信号中的第一输入信号,当所述第一输出信号是所述显著输出信号之外的输出信号时,所述第一增强信号是基于所述第一输出信号的。

2. 根据权利要求1所述的方法,包括:

对所述音频信号执行源定位。

3. 根据权利要求2所述的方法,其中由所述波束成形器中的至少两个波束成形器所指向的所述不同空间方向中的至少一个空间方向考虑所述源定位。

4. 根据权利要求1所述的方法,其中当所述第一输出信号是所述显著输出信号之外的输出信号时,所述第一增强信号是基于所述第一输出信号通过第一因子加权的。

5. 根据权利要求1所述的方法,其中所述波束成形器中的至少一个波束成形器的空间方向是所述音频处理装置的相机所面向的方向。

6. 根据权利要求1所述的方法,包括:

将所述第一增强信号和所述至少两个输入信号中的一个第一输入信号组合以提供一个第一组合信号并输出所述第一组合信号。

7. 根据权利要求6所述的方法,其中,所述组合包括:根据从用户界面提供的比率来混合所述第一增强信号和所述第一输入信号。

8. 根据权利要求7所述的方法,还包括:组合所述第一增强信号和所述至少两个输入信号中的一个第二输入信号以提供一个第二组合信号并输出所述第二组合信号。

9. 一种包括至少两个波束成形器和至少一个处理器的装置,所述至少一个处理器配置为:

-处理音频信号以生成至少两个输出信号,所述音频信号是来自至少两个音频输入的输入信号的混合,每个输出信号是通过使用指向不同空间方向的所述波束成形器中的一个波束成形器生成的;

-生成针对第一空间方向的第一增强信号,所述第一空间方向是由用于生成所述至少两个输出信号中的第一输出信号的波束成形器所指向的空间方向,当所述第一输出信号是所述至少两个输出信号中的显著输出信号时,所述第一增强信号是所述音频输入中的第一音频输入所捕获的所述输入信号中的第一输入信号,当所述第一输出信号是所述显著输出信号之外的输出信号时,所述第一增强信号是基于所述第一输出信号的。

10. 根据权利要求9所述的装置,包括:

源定位模块,配置为对所述音频信号执行源定位。

11. 根据权利要求10所述的装置,其中所述波束成形器中的至少两个波束成形器所指向的所述不同空间方向中的至少一个空间方向考虑所述源定位。

12. 根据权利要求9所述的装置,其中所述处理器配置为:

当所述第一输出信号是所述显著输出信号之外的输出信号时,基于所述第一输出信号

通过第一因子加权而生成所述第一增强信号。

13. 根据权利要求9所述的装置,其中所述波束成形器中的至少一个波束成形器的空间方向是所述装置的相机所面向的方向。

14. 根据权利要求9所述的装置,包括:音频捕获设备,所述音频捕获设备包括所述音频输入。

15. 根据权利要求9所述的装置,其中所述处理器配置为:

将所述第一增强信号和所述至少两个输入信号中的一个第一输入信号组合以向所述装置的输出模块(730)提供一个第一组合信号。

16. 一种计算机可读存储介质,其上存储有用于在音频输出装置中执行方法的指令,所述方法包括:

-处理音频信号以生成至少两个输出信号,所述音频信号是来自至少两个音频输入的输入信号的混合,每个输出信号是通过使用指向不同空间方向的波束成形器生成的;

-生成针对第一空间方向的第一增强信号,所述第一空间方向是由用于生成所述至少两个输出信号中的第一输出信号的波束成形器所指向的空间方向,当所述第一输出信号是所述至少两个输出信号中的显著输出信号时,所述第一增强信号是所述音频输入中的第一音频输入所捕获的所述输入信号中的第一输入信号,当所述第一输出信号是所述显著输出信号之外的输出信号时,所述第一增强信号是基于所述第一输出信号的。

用于增强声源的方法和装置

技术领域

[0001] 本发明涉及用于增强声源的方法和装置,并且更具体地涉及用于从有噪声录音中增强声源的方法和装置。

背景技术

[0002] 录音通常是几个声源(例如,目标语音或音乐、环境噪声和来自其他语音的干扰)的混合,其使收听者无法理解并专注于感兴趣的声源。在诸如但不限于音频/视频会议、语音识别、助听器和音频缩放的应用中期望有从有噪声录音中分离和专注于感兴趣的声源的能力。

发明内容

[0003] 根据本原理的实施例,提供了一种用于处理音频信号的方法,音频信号是至少来自第一音频源的第一信号和来自第二音频源的第二信号的混合,所述方法包括:使用指向第一方向的第一波束成形器来处理音频信号以生成第一输出,第一方向与第一音频源相对应;使用指向第二方向的第二波束成形器来处理音频信号以生成第二输出,第二方向与第二音频源相对应;和处理第一输出和第二输出以生成增强的第一信号,如下所述。根据本原理的另一个实施例,还提出了一种用于执行这些步骤的装置。

[0004] 根据本原理的实施例,提供了一种用于处理音频信号的方法,音频信号是至少来自第一音频源的第一信号和来自第二音频源的第二信号的混合,所述方法包括:使用指向第一方向的第一波束成形器来处理音频信号以生成第一输出,第一方向与第一音频源相对应;使用指向第二方向的第二波束成形器来处理音频信号以生成第二输出,第二方向与第二音频源相对应;确定第一输出在第一输出和所述第二输出之间是显著的;和处理第一输出和第二输出以生成增强的第一信号,其中如果确定第一输出是显著的,则生成增强的第一信号的处理基于参考信号,并且其中如果确定第一输出不是显著的,则生成增强的第一信号的处理基于通过第一因子加权的第一输出,如下所述。根据本原理的另一个实施例,还提出了一种用于执行这些步骤的装置。

[0005] 根据本原理的实施例,提出了一种其上存储有用于根据上述方法处理音频信号的指令的计算机可读存储介质,音频信号是至少来自第一音频源的第一信号和来自第二音频源的第二信号的混合。

附图说明

[0006] 图1示出了增强目标声源的示例性音频系统。

[0007] 图2示出了根据本原理的实施例的示例性音频增强系统。

[0008] 图3示出了根据本原理的实施例的用于执行音频增强的示例性方法。

[0009] 图4示出了根据本原理的实施例的示例性音频增强系统。

[0010] 图5示出了根据本原理的实施例的具有三个波束成形器的示例性音频缩放系统。

[0011] 图6示出了根据本原理的实施例的具有五个波束成形器的示例性音频缩放系统。

[0012] 图7描绘了根据本原理的实施例的可以使用音频处理器的示例性系统的框图。

具体实施方式

[0013] 图1示出了增强目标声源的示例性音频系统。例如移动电话的音频捕获设备(105)获得有噪声录音(例如,来自在方向 θ_1 的人的语音、在方向 θ_2 的播放音乐的扬声器、来自背景的噪声以及在方向 θ_k 的播放音乐的仪器的混合,其中 θ_1 、 θ_2 、……或 θ_k 表示源相对于麦克风阵列的空间方向)。音频增强模块110基于用户请求(例如,来自用户界面的专注于人的语音的请求)对所请求的源执行增强并输出增强的信号。注意,音频增强模块110可以位于与音频捕获设备105分离的设备中,或者它还可以并入作为音频捕获设备105的模块。

[0014] 存在可用于从有噪声录音中增强目标音频源的方法。例如,已知音频源分离是将多个声源与它们的混合进行分离的强大技术。在具有挑战性的情况下(例如,具有高混响,或者当源的数量是未知的并且超过传感器的数量时)分离技术仍然需要改进。此外,分离技术目前不适用于具有有限处理能力的实时应用。

[0015] 称为波束成形的另一种方法使用指向目标源的方向的空间波束以便增强目标源。波束成形通常与后滤波技术一起使用以用于进一步的扩散噪声抑制。波束成形的一个优点是,对于少量的麦克风,计算需求不昂贵,并且因此适合于实时应用。然而,当麦克风的数量较小时(例如,对于当前移动设备,2或3个麦克风),所生成的波束图案不够窄而不足以抑制背景噪声和来自不需要的源的干扰。一些现有的工作还提出将波束成形与频谱相减结合以满足移动设备中的识别和语音增强。在这些工作中,通常假设目标源方向是已知的,并且所考虑的零波束成形可能对于混响效应不够鲁棒。此外,频谱相减步骤还可能向输出信号添加伪像。

[0016] 本原理针对从有噪声录音中增强声源的方法和系统。根据本原理的新颖方面,我们提出的方法使用若干信号处理技术(例如但不限于基于指向空间中不同源方向的若干波束成形器的输出的源定位、波束成形和后处理),其可以有效地增强任何目标声源。通常,增强将提高来自目标声源的信号的质量。我们提出的方法具有轻的计算负荷,并且可以用于实时应用中(例如但不限于音频会议和音频缩放,甚至在具有有限处理能力的移动设备中)。根据本原理的另一个新颖方面,可以基于增强声源执行渐进音频缩放(0%-100%)。

[0017] 图2示出了根据本原理的实施例的示例性音频增强系统200。系统200接受音频录音作为输入,并提供增强的信号作为输出。为了执行音频增强,系统200采用若干信号处理模块,包括源定位模块210(可选)、多个波束成形器(220、230、240)和后处理器250。在下文中,我们更详细地描述每个信号处理块。

[0018] 源定位

[0019] 考虑到音频录音,源定位算法(例如,相位变换-广义互相关(GCC-PHAT))可用于估计显著源的方向(还称为到达方向DoA)(当它们未知时)。因此,可以确定不同源的DoA θ_1 、 θ_2 、…… θ_K ,其中K是显著源的总数。当提前知道DoA时,例如,当我们将智能手机指向某个方向以捕获视频时,我们知道感兴趣的源在麦克风阵列的正前方($\theta_1=90$ 度),并且我们不需要执行源定位功能来检测DoA,或者我们仅执行源定位以检测显著干扰源的DoA。

[0020] 波束成形

[0021] 考虑到显著声源的DoA,波束成形可以采用为强大的技术以增强空间中的特定声音方向,同时抑制来自其它方向的信号。在一个实施例中,我们使用指向显著源的不同方向的几个波束成形器来增强相应的声源。用 $x(n, f)$ 表示观察到的时域混合信号 $x(t)$ 的短时傅里叶变换(STFT)系数(在时频域中的信号),其中 n 是时间帧索引,并且 f 是频率窗口(bin)索引。第 j 个波束成形器(在方向 θ_j 上增强声源)的输出可以计算为

$$[0022] \quad s_j(n, f) = \mathbf{w}_j^H(n, f) \mathbf{x}(n, f) \quad (1)$$

[0023] 其中 $\mathbf{w}_j(n, f)$ 是从指向波束成形器 j 的目标方向的导引向量导出的加权向量,并且 H 表示向量共轭转置。可以针对不同类型的波束成形器以不同方式(例如,使用最小方差无失真响应(MVDR)、鲁棒MVDR、延迟和总和(DS)以及广义旁瓣消除器(GSC))计算 $\mathbf{w}_j(n, f)$ 。

[0024] 后处理

[0025] 波束成形器的输出通常不足以分离干扰,并且直接对该输出应用后处理可能导致强的信号失真。一个原因是,由于(1)波束成形中的非线性信号处理,(2)估计显著源的方向的误差,增强源通常包含大量的音乐噪声(伪像),这可能导致在高频处更多的信号失真,因为DoA误差可能导致大的相位差。因此,我们建议对几个波束成形器的输出应用后处理。在一个实施例中,后处理可以基于参考信号 x_I 和波束成形器的输出,其中参考信号可以是输入麦克风之一,例如面向目标源的智能手机中的麦克风、紧邻相机的智能手机中的麦克风或靠近嘴的蓝牙耳机中的麦克风。参考信号还可以是从多个麦克风信号生成的更复杂的信号,例如,多个麦克风信号的线性组合。此外,时频掩蔽(并且可选地频谱相减)可以用于产生增强的信号。

[0026] 在一个实施例中,例如对于源 j ,增强的信号生成如下:

$$[0027] \quad \hat{s}_j(n, f) = \begin{cases} x_I(n, f) & \text{如果 } |s_j(n, f)| > \alpha * \max\{|s_i(n, f)|, \forall i \neq j\} \\ \beta * s_j(n, f) & \text{其它} \end{cases} \quad (2)$$

[0028] 其中 $x_I(n, f)$ 是参考信号的STFT系数, α 和 β 是调谐常数,在一个示例中, $\alpha=1, 1.2$ 或 1.5 , $\beta=0.05-0.3$ 。 α 和 β 的特定值可以基于应用来调整。等式(2)中的一个基本假设是:声源在时频域中几乎不重叠,因此如果源 j 在时频点 (n, f) 上是重要的(即,波束成形器 j 的输出大于所有其他波束成形器的输出),则参考信号可以认为是目标源的良好近似。因此,我们可以将增强的信号设置为参考信号 $x_I(n, f)$ 以减少包含在 $s_j(n, f)$ 中的由波束成形引起的失真(伪像)。否则,我们假设信号是噪声或噪声和目标源的混合,并且我们可以选择通过将 $\hat{s}_j(n, f)$ 设置为小值 $\beta * s_j(n, f)$ 来抑制它。

[0029] 在另一实施例中,后处理还可以使用频谱相减、噪声抑制方法。数学上,它可以描述为:

$$[0030] \quad \hat{s}_j(n, f) = \begin{cases} \sqrt{|x_l(n, f)|^2 - \sigma_j^2(f)} * \text{phase}(x_l(n, f)) & \text{如果 } |s_j(n, f)| > \alpha * \max\{|s_i(n, f)|, \forall i \neq j\} \\ \beta * s_j(n, f) & \text{其它, 并更新噪声电平 } \sigma_j^2(f) \end{cases}$$

(3)

[0031] 其中 $\text{phase}(x_l(n, f))$ 表示信号 $x_l(n, f)$ 的相位信息,并且 $\sigma_j^2(f)$ 是影响可以连续更新的源 j 的噪声的频率相关频谱功率。在一个实施例中,如果帧被检测为有噪声的帧,则可以将噪声电平设置为该帧的信号电平,或者可以通过考虑先前噪声值的遗忘因子来平滑地更新该噪声电平。

[0032] 在另一实施例中,后处理对波束成形器的输出执行“清理”以便获得更鲁棒的波束成形器。这可以用如下的滤波器自适应地完成:

$$[0033] \quad \hat{s}_j(n, f) = \beta_j(n, f) * s_j(n, f) \quad (4)$$

[0034] 其中 β_j 因子取决于可以看作时频信号干扰比的量 $\frac{|s_j(n, f)|}{\max\{|s_i(n, f)|, \forall i \neq j\}}$ 。例如,我们可以设置 β 如下以用于进行“软”后处理“清理”:

$$[0035] \quad \beta_j(n, f) = \frac{1}{\varepsilon + \frac{\max\{|s_i(n, f)|, \forall i \neq j\}}{|s_j(n, f)|}} \quad (5)$$

[0036] 其中 ε 是小常数,例如 $\varepsilon=1$ 。因此,当 $|s_j(n, f)|$ 比每个其他的 $|s_i(n, f)|$ 高得多时,清理的输出是 $\hat{s}_j(n, f) \approx s_j(n, f)$,并且当 $s_j(n, f)$ 比另一个 $s_i(n, f)$ 小得多时,清理的输出是 $\hat{s}_j(n, f) \approx 0$ 。

[0037] 我们还可以设置 β 如下以进行“硬”(二进制)清理:

$$[0038] \quad \beta_j(n, f) = \begin{cases} 1 & \text{如果 } |s_j(n, f)| > \max\{|s_i(n, f)|, \forall i \neq j\} \\ 0 & \text{其它} \end{cases} \quad (6)$$

[0039] 还可以通过根据 $|s_j(n, f)|$ 和 $|s_i(n, f)|$ ($i \neq j$)之间的电平差调整其值而以中间(即,在“软”清理和“硬”清理之间)方式来设置 β_j 。

[0040] 上述这些技术(“软”/“硬”/中间清理)还可以扩展到对 $x_l(n, f)$ 而不是 $s_j(n, f)$ 进行滤波:

$$[0041] \quad \hat{s}_j(n, f) = \beta_j(n, f) * x_l(n, f) \quad (7)$$

[0042] 注意,在这种情况下,仍然使用波束成形器的输出 $s_j(n, f)$ (而不是原始麦克风信号)来计算 β_j 因子以利用波束成形。

[0043] 对于上述技术,我们还可以添加记忆效应以便避免增强信号中的准时错误检测或毛刺。例如,我们可以对后处理的决定中暗示的量进行平均,例如以以下总和:

$$[0044] \quad \frac{1}{M} \sum_{m=0}^{M-1} |s_j(n-m, f)| > \alpha * \max \left\{ \frac{1}{M} \sum_{m=0}^{M-1} |s_i(n-m, f)|, \forall i \neq j \right\}$$

[0045] 替换

$$[0046] \quad |s_j(n, f)| > \alpha * \max\{|s_i(n, f)|, \forall i \neq j\}$$

[0047] 其中M是为决定而考虑的帧的数量。

[0048] 此外,在如上所述的信号增强之后,可以使用其他后滤波技术来进一步抑制扩散背景噪声。

[0049] 在下文中,为了便于表示,我们将如等式(2)、(4)和(7)中描述的方法称为窗口分离,并且将如等式(3)中描述的方法称为频谱相减。

[0050] 图3示出了根据本原理的实施例的用于执行音频增强的示例性方法300。方法300在步骤305开始。在步骤310,执行初始化,例如,确定是否有必要使用源定位算法来确定显著源的方向。如果是,则选择用于源定位的算法并设置其参数。还可以例如基于用户配置来确定要使用哪个波束成形算法或波束成形器的数量。

[0051] 在步骤320,使用源定位来确定显著源的方向。注意,如果显著源的方向是已知的,则可以跳过步骤320。在步骤330,使用多个波束成形器,每个波束成形器指向不同的方向以增强相应的声音源。可以根据源定位来确定每个波束成形器的方向。如果目标源的方向已知,则还可以在360°场中对方向进行采样。例如,如果已知目标源的方向为90°,则我们可以使用90°、0°和180°来对360°场进行采样。可以使用不同的方法(例如但不限于最小方差无失真响应(MVDR)、鲁棒MVDR、延迟和总和(DS)以及广义旁瓣消除器(GSC))以用于波束成形。在步骤340,对波束成形器的输出执行后处理。后处理可以基于如等式(2)-(7)中描述的算法并且还可以结合频谱相减和/或其它后滤波技术来执行。

[0052] 图4描绘了根据本原理的实施例的可以使用音频增强的示例性系统400的框图。麦克风阵列410对需要处理的有噪声录音进行录音。麦克风可以对来自一个或多个扬声器或设备的音频进行录音。有噪声录音还可以被预先录音并存储在存储介质中。源定位模块420是可选的。当使用源定位模块420时,其可以用于确定显著源的方向。波束成形模块430应用指向不同方向的多个波束成形。基于波束成形器的输出,后处理器440例如使用等式(2)-(7)中描述的方法之一来执行后处理。在后处理之后,增强的声源可以由扬声器450播放。输出声音还可以存储在存储介质中或通过通信信道发送给接收器。

[0053] 图4中所示的不同模块可以在一个设备中实现或者分布在多个设备上。例如,所有模块可以包括在但不限于平板电脑或移动电话中。在另一个示例中,源定位模块420、波束成形模块430和后处理器440可以与计算机中或云中的其他模块分离地定位。在又一个实施例中,麦克风阵列410或扬声器450可以是独立模块。

[0054] 图5示出了其中可以使用本原理的示例性音频缩放系统500。在音频缩放应用中,用户可能仅对空间中的一个源方向感兴趣。例如,当用户将移动设备指向特定方向时,可以假设移动设备指向的特定方向是目标源的DoA。在音频-视频捕获的示例中,可以假设DoA方向是相机所面向的方向。干扰项则是范围外的源(在音频捕获设备的侧面和后面)。因此,在音频缩放应用中,由于通常可以从音频捕获设备推断DoA方向,所以源定位可以是可选的。

[0055] 在一个实施例中,主波束成形器设置为指向目标方向 θ ,而(可能)几个其他波束成形器指向其他非目标方向(例如, $\theta-90^\circ$ 、 $\theta-45^\circ$ 、 $\theta+45^\circ$ 、 $\theta+90^\circ$)以在后处理期间为用户捕获更多的噪声和干扰。

[0056] 音频系统500使用四个麦克风 m_1-m_4 (510、512、514、516)。例如使用FFT模块(520、

522、524、526) 将来自每个麦克风的信号从时域变换到时频域。波束成形器530、532和534基于时频信号执行波束成形。在一个示例中,波束成形器530、532和534可以分别指向 0° 、 90° 、 180° 的方向以对声场(360°)进行采样。后处理器540例如使用等式(2)–(7)中描述的方法之一,基于波束成形器530、532和534的输出来执行后处理。当参考信号用于后处理器时,后处理器540可以使用来自麦克风(例如, m_4)的信号作为参考信号。

[0057] 例如使用IFFT模块550将后处理器540的输出从时频域变换回时域。基于例如由用户请求通过用户界面提供的音频缩放因子 α (具有从0到1的值),混合器560和570分别生成右输出和左输出。

[0058] 音频缩放的输出是左和右麦克风信号(m_1 和 m_4)与来自IFFT模块550的增强的输出根据缩放因子 α 的线性混合。输出为具有左输出和右输出的立体声。为了保持立体声效果, α 的最大值应小于1(例如0.9)。

[0059] 除了在等式(2)–(7)中描述的方法之外,可以在后处理器中使用频率和频谱相减。可以根据窗口分离输出来计算心理声学频率掩蔽。原理是具有心理声学掩蔽之外的级别的频率窗口不用于生成频谱相减的输出。

[0060] 图6示出了可以使用本原理的另一示例性音频缩放系统600。在系统600中,使用5个波束成形器而不是3个波束成形器。特别地,波束成形器分别指向 0° 、 45° 、 90° 、 135° 和 180° 的方向。

[0061] 音频系统600还使用四个麦克风 m_1 – m_4 (610、612、614、616)。例如使用FFT模块(620、622、624、626)将来自每个麦克风的信号从时域变换到时频域。波束成形器630、632、634、636和638基于时频信号执行波束成形,并且它们分别指向 0° 、 45° 、 90° 、 135° 和 180° 的方向。后处理器640例如使用等式(2)–(7)中描述的方法之一,基于波束成形器630、632、634、636和638的输出来执行后处理。当参考信号用于后处理器时,后处理器540可以使用来自麦克风(例如, m_3)的信号作为参考信号。例如使用IFFT模块660将后处理器640的输出从时频域变换回时域。基于音频缩放因子,混合器670生成输出。

[0062] 一种或另一种后处理技术的主观质量随着麦克风的数量而变化。在一个实施例中,对于两个麦克风仅优选窗口分离,而对于4个麦克风,优选窗口分离和频谱相减。

[0063] 当存在多个麦克风时,可以应用本原理。在系统500和600中,我们假设信号来自四个麦克风。当只有两个麦克风时,如果需要,则可以在使用频谱相减的后处理中使用平均值 $(m_1+m_2)/2$ 作为 m_3 。注意,这里的参考信号可以来自更靠近目标源的一个麦克风或是麦克风信号的平均值。例如,当存在三个麦克风时,关于频谱相减的参考信号可以是 $(m_1+m_2+m_3)/3$,或者如果 m_3 面向感兴趣的源则参考信号直接为 m_3 。

[0064] 一般来说,本实施例使用多个方向上的波束成形的输出来增强目标方向上的波束成形。通过在多个方向上执行波束成形,在多个方向上对声场(360°)进行采样,并且然后可以对波束成形器的输出进行后处理以对来自目标方向的信号进行“清理”。

[0065] 音频缩放系统(例如系统500或600)还可以用于音频会议,其中来自不同位置的扬声器的语音可以被增强,并且可以很好地应用指向多个方向的多个波束成形器的使用。在音频会议中,录音设备位置通常是固定的(例如,放置在具有固定位置的桌子上),而不同的扬声器位于任意位置。在将波束成形器导向到这些源之前,可以使用源定位和跟踪(例如,用于跟踪移动的扬声器)来获知源的位置。为了提高源定位和波束成形的精度,可以使用解

混响技术来预处理输入混合信号以便减少混响效应。

[0066] 图7示出了其中可以使用本原理的音频系统700。对系统700的输入可以是音频流(例如mp3文件)或音频-视频流(例如mp4文件)或来自不同输入的信号。输入还可以来自存储设备或者从通信信道接收。如果音频信号被压缩,则在被增强之前对其进行解码。音频处理器720例如使用方法300或系统500或600执行音频增强。对音频缩放请求可以与对视频缩放的请求分离或包括在对视频缩放的请求中。

[0067] 基于来自用户接口740的用户请求,系统700可以接收音频缩放因子,音频缩放因子可以控制麦克风信号和增强信号的混合比例。在一个实施例中,音频缩放因子还可以用于调整 β_j 的加权值以便控制后处理之后剩余的噪声量。随后,音频处理器720可以混合增强的音频信号和麦克风信号以生成输出。输出模块730可播放音频、存储音频或将音频发送给接收器。

[0068] 可以例如用方法或过程、装置、软件程序、数据流或信号来实现本文所描述的实现。虽然仅在单个实现形式的上下文中进行讨论(例如,仅作为方法讨论),但是所讨论特征的实现还可以以其他形式来实现(例如装置或程序)。装置可以例如以适合硬件、软件和固件来实现。方法可以例如在诸如处理器的装置中实现,处理器一般称为处理设备,包括例如计算机、微处理器、集成电路或可编程逻辑器件。处理器还包括通信设备(诸如计算机、蜂窝电话、便携/个人数字助理("PDA"))以及便于终端用户之间的信息通信的其他设备。

[0069] 对本原理的“一个实施例”或“实施例”或“一个实现”或“实现”以及它的其它变形的提及意味着结合实施例描述的具体特征、结构、特性等包括在本原理的至少一个实施例中。因此,出现在整个说明书中的各个地方的短语“在一个实施例中”或“在实施例中”或“在一个实现中”或“在实现中”以及任何其他变体的出现不一定都是指同一实施例。

[0070] 此外,本申请或其权利要求可能提及“确定”各种信息。确定信息可以包括以下一项或多项,例如,估计信息、计算信息、预测信息或从存储器检索信息。

[0071] 此外,本申请或其权利要求可能提及“访问”各种信息。访问信息可以包括以下一项或多项,例如,接收信息、检索信息(如从存储器)、存储信息、处理信息、发送信息、移动信息、拷贝信息、擦除信息、计算信息、确定信息、预测信息或估计信息。

[0072] 此外,本申请或其权利要求可能提及“接收”各种信息。接收与“访问”一样,意在作为通过第一因子加权而广义术语。接收信息可以包括以下一项或多项,例如,访问信息或检索信息(如从存储器)。此外,“接收”通常以一种方式或另一种方式包括在诸如以下操作的操作期间:存储信息、处理信息、发送信息、移动信息、拷贝信息、擦除信息、计算信息、确定信息、预测信息或估计信息。

[0073] 如对于本领域技术人员明显的是,实现可以产生被格式化为携带可以例如被存储或发送的信息的各种信号。信息可以包括例如执行方法的指令或通过描述的实现之一所产生的数据。例如,信号可以被格式化为携带所述实施例的比特流。这种信号可被格式化为例如电磁波(例如使用频谱的射频部分)或基带信号。格式化可包括例如对数据流进行编码和用所编码的数据流调制载波。信号携带的信息可以是例如模拟或数字信息。可以通过如已知的多种不同有线或无线链路来发送信号。信号可存储在处理器可读介质上。

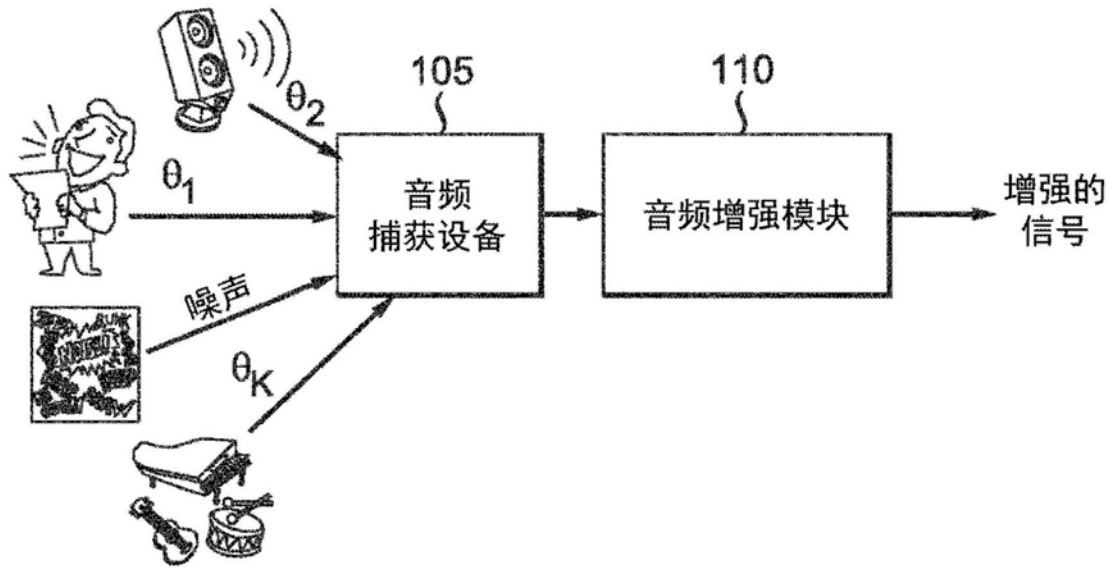


图1

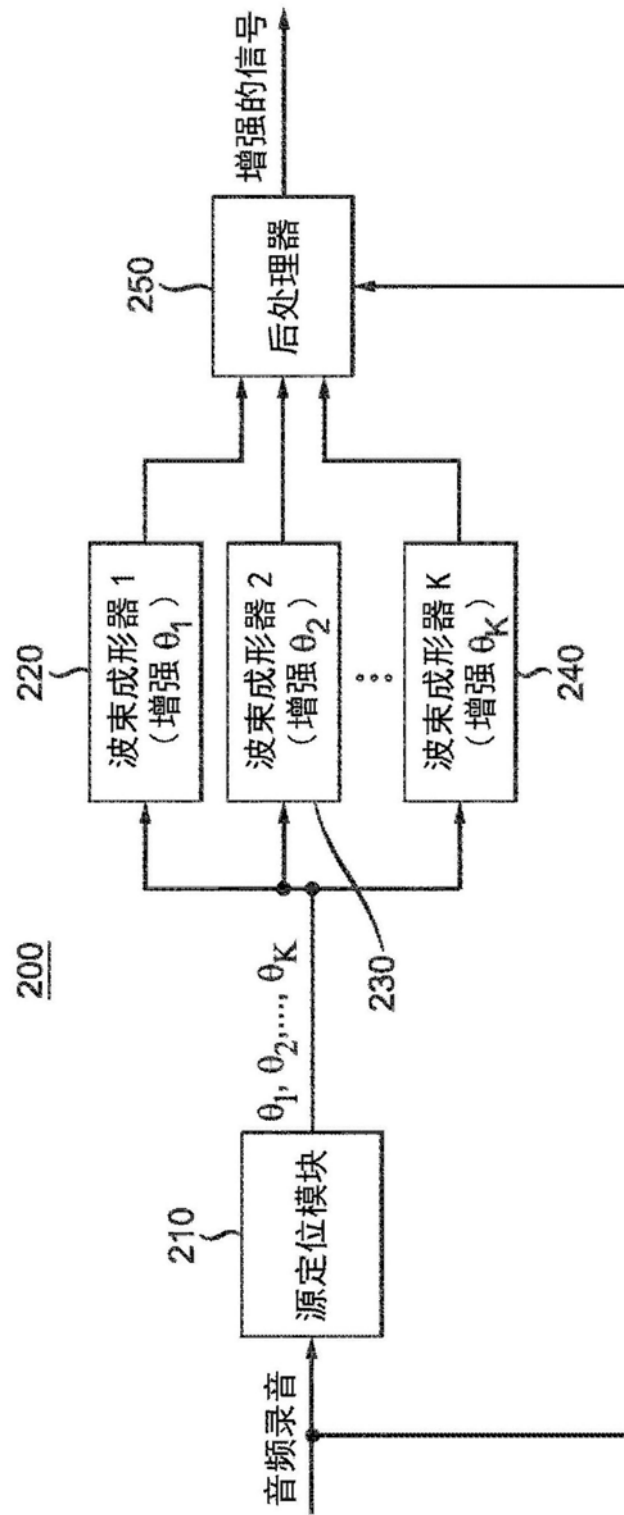


图2

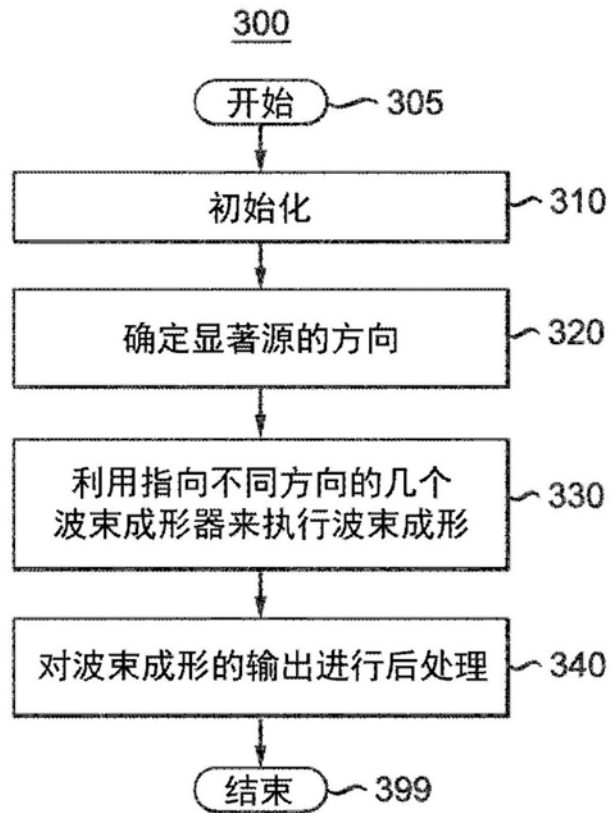


图3

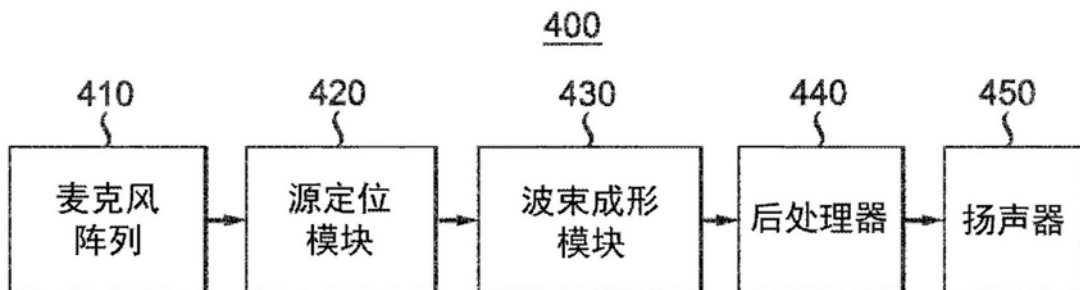


图4

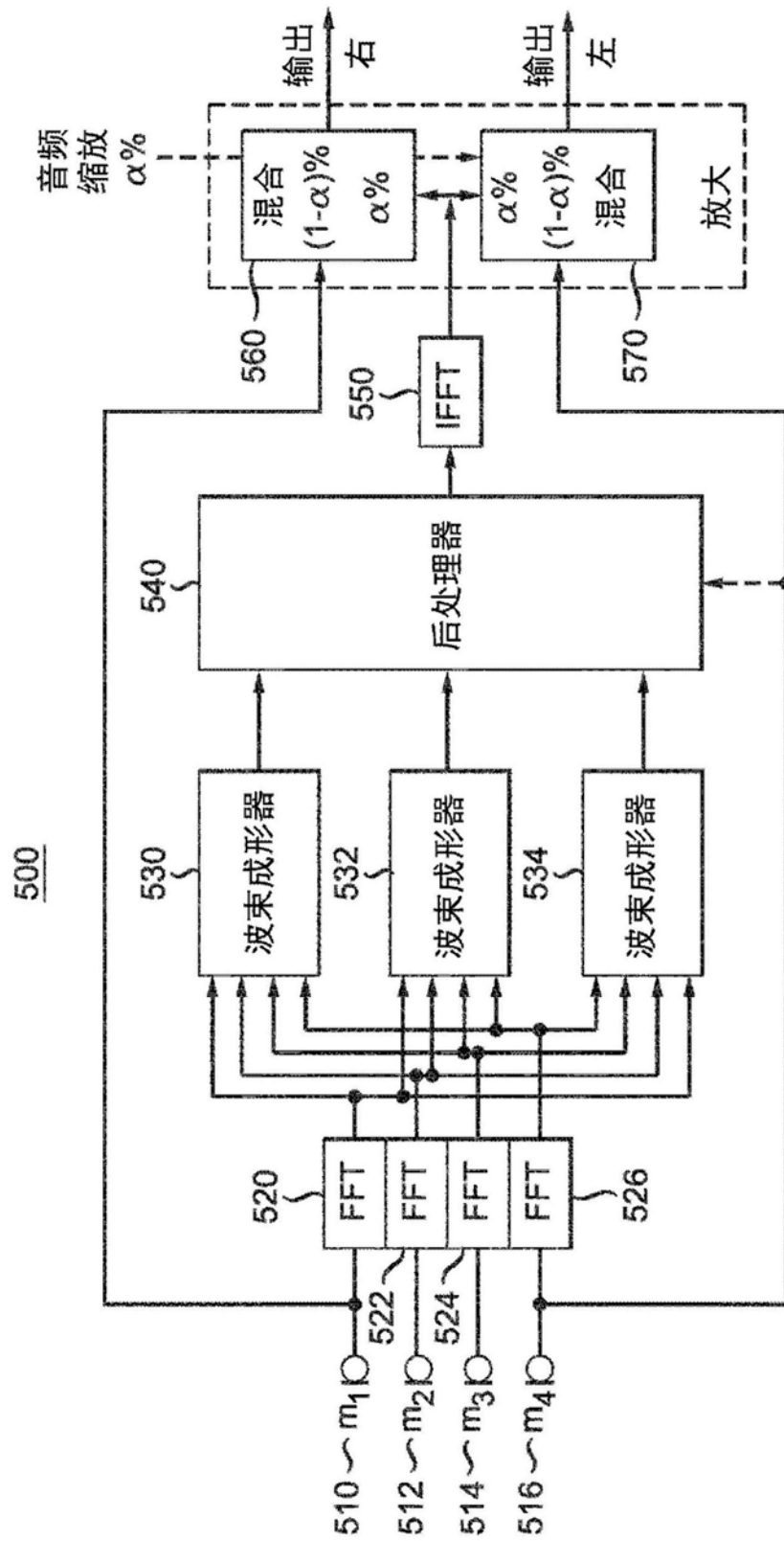


图5

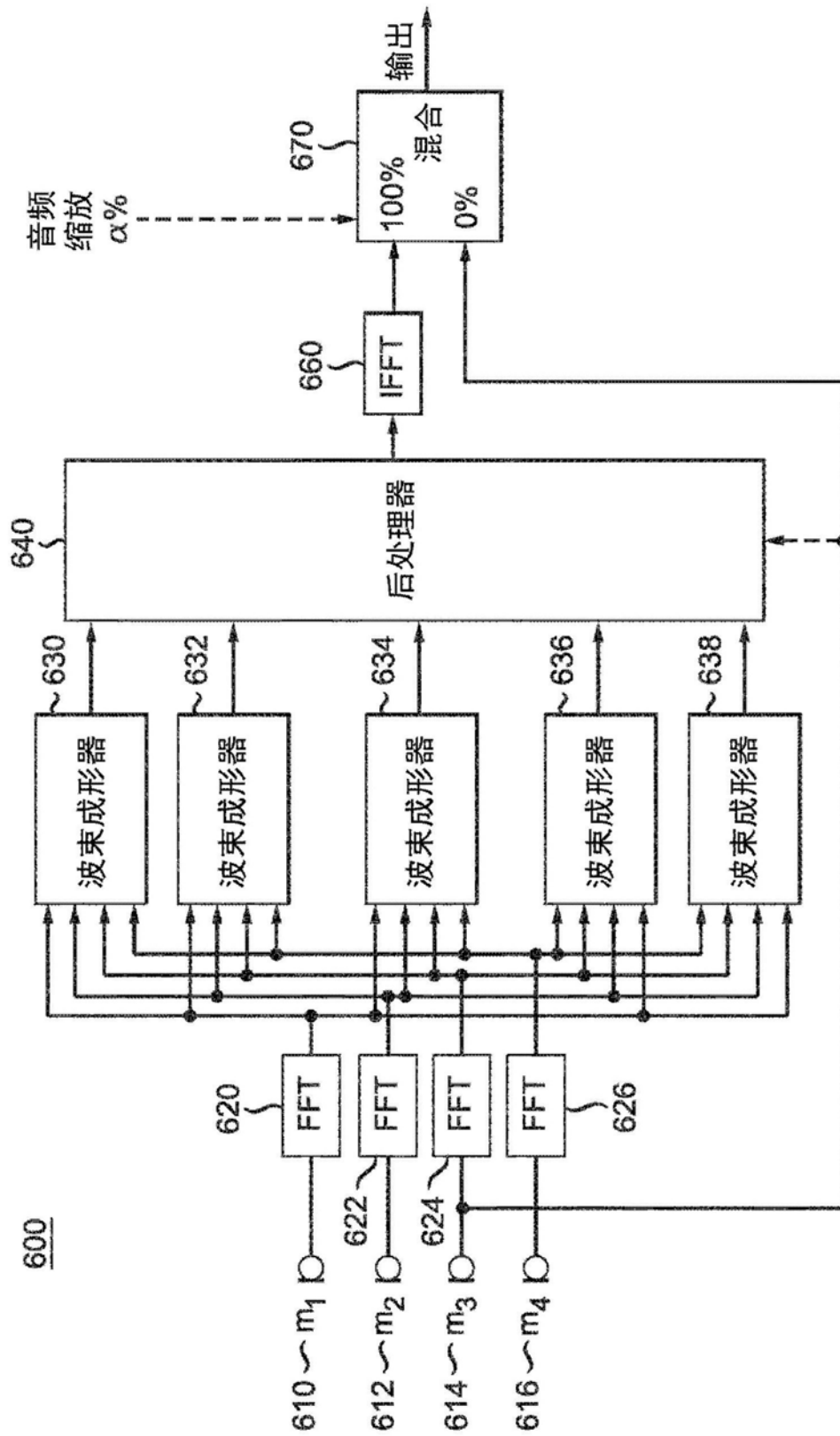


图6

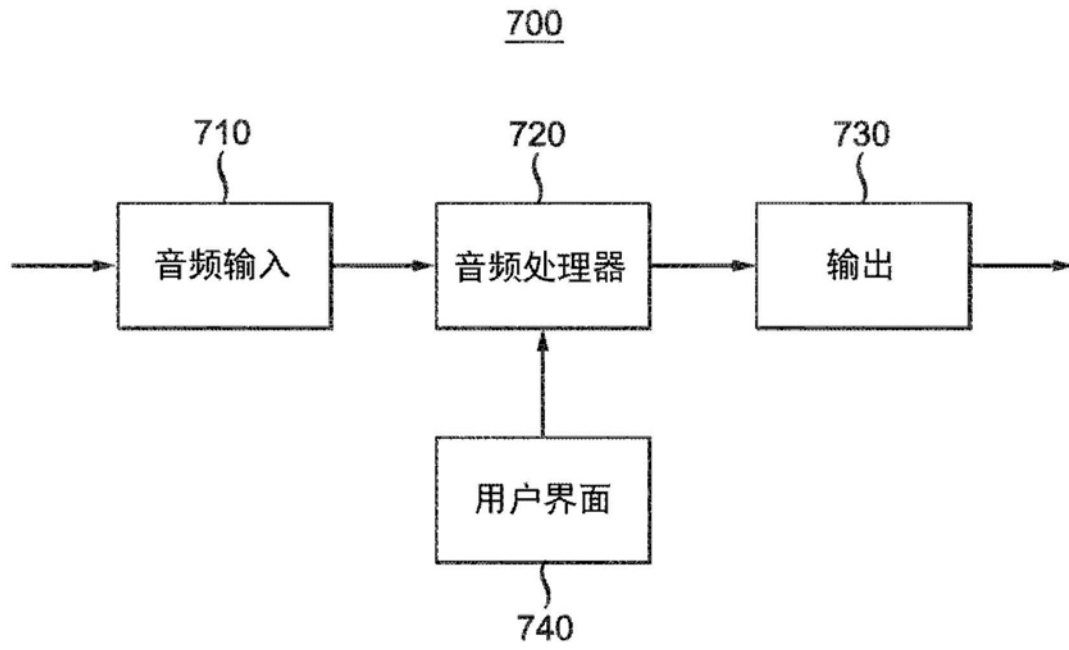


图7