



(12) 发明专利

(10) 授权公告号 CN 112002337 B

(45) 授权公告日 2024. 08. 09

(21) 申请号 202010914886.6

(22) 申请日 2016.03.02

(65) 同一申请的已公布的文献号
申请公布号 CN 112002337 A

(43) 申请公布日 2020.11.27

(30) 优先权数据
62/127,613 2015.03.03 US
62/298,905 2016.02.23 US

(62) 分案原申请数据
201680011460.1 2016.03.02

(73) 专利权人 杜比实验室特许公司
地址 美国加利福尼亚

(72) 发明人 D·S·麦克格拉斯

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

专利代理师 宿小猛

(51) Int.Cl.
G10L 19/008 (2013.01)
H04S 3/00 (2006.01)

(56) 对比文件
CN 101248483 A, 2008.08.20
CN 101133680 A, 2008.02.27

审查员 李海龙

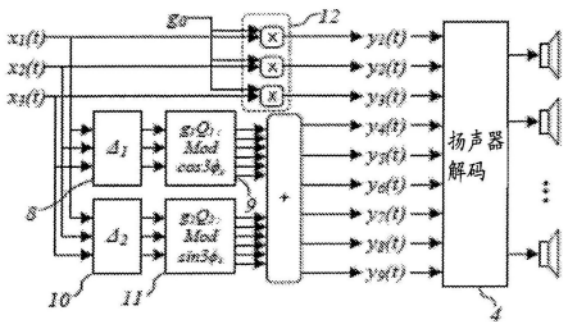
权利要求书4页 说明书19页 附图10页

(54) 发明名称

用于对音频信号进行处理的方法、装置和设备

(57) 摘要

本公开涉及用于对音频信号进行处理的方法、装置和设备。一些方法涉及接收包括N个输入音频声道的输入音频信号,输入音频信号表示具有第一声场格式分辨率的第一声场格式,N是 ≥ 2 的整数。可以将第一解相关过程应用于输入音频声道中的两个或更多个输入音频声道以产生第一组解相关的声道,第一解相关过程维持该组输入音频声道的声道间相关性。可以将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道。可以将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括0个输出音频声道的输出音频信号,0个输出音频声道表示分辨率比第一声场格式相对较高的第二声场格式,0是 ≥ 3 的整数。



1. 一种对音频信号进行处理的方法,该方法包括:

接收包括 N_r 个输入音频声道的输入音频信号,该输入音频信号表示具有第一声场格式分辨率的第一声场格式, N_r 是 ≥ 2 的整数;

将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道,第一解相关过程维持该组输入音频声道的声道间相关性;

将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道;以及

将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号, N_p 是 ≥ 3 的整数,

其特征在于, N_p 个输出音频声道包括与输出音频信号的较低分辨率分量相对应的 N_r 个未解相关的输出声道,以及与输出音频信号的较高分辨率分量相对应的解相关的且经调制的输出声道,

其中 N_r 个未解相关的输出声道中的每一个是通过将增益应用于 N_r 个输入音频声道之一而产生的。

2. 根据权利要求1所述的方法,其中,调制过程涉及将线性矩阵应用于第一组解相关的声道。

3. 根据权利要求1所述的方法,其中应用第一解相关过程涉及将相同的解相关过程应用于 N_r 个输入音频声道中的每个输入音频声道。

4. 根据权利要求1所述的方法,进一步包括:

将第二解相关过程应用于输入音频声道中的所述一组两个或更多个输入音频声道以产生第二组解相关的声道,第二解相关过程维持该组输入音频声道的声道间相关性;以及

将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道,其中组合涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及与所述未解相关的输出声道组合。

5. 根据权利要求4所述的方法,其中第一解相关过程包括第一解相关函数,并且第二解相关过程包括第二解相关函数,第二解相关函数包括具有90度或-90度相移的第一解相关函数。

6. 根据权利要求4所述的方法,其中第一调制过程包括第一调制函数,并且第二调制过程包括第二调制函数,第二调制函数包括具有90度或-90度相移的第一调制函数。

7. 根据权利要求1所述的方法,其中未解相关的输出声道通过将最小二乘格式转换器应用于 N_r 个输入音频声道而产生。

8. 根据权利要求1-7中任一项所述的方法,其中接收输入音频信号涉及接收来自音频引导逻辑过程的第一输出,该第一输出包括 N_r 个输入音频声道,进一步包括将输出音频信号的 N_p 个音频声道与来自音频引导逻辑过程的第二输出组合,该第二输出包括被引导音频数据的 N_p 个音频声道,在该被引导音频数据中,一个或多个声道的增益已经基于当前的主导声音方向而被改变。

9. 根据权利要求1-7中任一项所述的方法,其中,所述第一声场格式是B格式。

10. 一种非暂态计算机可读介质,存储有指令,所述指令在被处理器执行时使得所述处理器执行根据权利要求1-9中任一项所述的方法。

11.一种用于对音频信号进行处理的装置,包括:

接口系统;以及

控制系统,所述控制系统能够执行根据权利要求1-9中任一项所述的方法。

12.一种用于对音频信号进行处理的系统,包括:

处理器;以及

非暂态计算机可读介质,存储有指令,所述指令在被所述处理器执行时使得处理器执行根据权利要求1-9中任一项所述的方法。

13.一种用于对音频信号进行处理的系统,包括:

处理器;以及

非暂时性计算机可读介质,存储有指令,所述指令在由所述处理器执行时使得所述处理器执行包括以下的操作:

从接口接收包括 N_r 个输入音频声道的输入音频信号,该输入音频信号表示具有第一声场格式分辨率的第一声场格式, N_r 是 ≥ 2 的整数;

将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道,第一解相关过程维持该组输入音频声道的声道间相关性;

将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道;以及

将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号, N_p 是 ≥ 3 的整数,输出声道表示第二声场格式,所述第二声场格式为与第一声场格式相比具有相对较高分辨率的声场格式,该未解相关的输出声道与输出音频信号的较低分辨率分量相对应,以及解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应,

其中两个或更多个未解相关的输出声道中的每一个是通过将增益应用于 N_r 个输入音频声道之一而产生的。

14.根据权利要求13所述的系统,其中,调制过程包括将线性矩阵应用于第一组解相关的声道。

15.根据权利要求13所述的系统,其中该组合涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合。

16.根据权利要求13所述的系统,其中应用第一解相关过程包括将相同的解相关过程应用于 N_r 个输入音频声道中的每个输入音频声道。

17.根据权利要求13所述的系统,其中所述操作进一步包括:

将第二解相关过程应用于输入音频声道中的所述一组两个或更多个输入音频声道以产生第二组解相关的声道,第二解相关过程维持该组输入音频声道的声道间相关性;以及

将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道,其中组合涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及与所述两个或更多个未解相关的输出声道组合。

18.根据权利要求17所述的系统,其中第一解相关过程包括第一解相关函数,并且第二解相关过程包括第二解相关函数,第二解相关函数包括具有90度或-90度相移的第一解相关函数。

19. 根据权利要求18所述的系统, 其中第一调制过程包括第一调制函数, 并且第二调制过程包括第二调制函数, 第二调制函数包括具有90度或-90度相移的第一调制函数。

20. 一种对音频信号进行处理的方法, 包括:

通过处理器从接口系统接收包括 N_t 个输入音频声道的输入音频信号, 该输入音频信号表示具有第一声场格式分辨率的第一声场格式, N_t 是 ≥ 2 的整数;

将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道, 第一解相关过程维持该组输入音频声道的声道间相关性;

将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道; 以及

将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号, N_p 是 ≥ 3 的整数, 输出声道表示第二声场格式, 所述第二声场格式为与第一声场格式相比具有相对较高分辨率的声场格式, 该未解相关的输出声道与输出音频信号的较低分辨率分量相对应, 以及解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应,

其中两个或更多个未解相关的输出声道中的每一个是通过将增益应用于 N_t 个输入音频声道之一而产生的。

21. 根据权利要求20所述的方法, 其中, 调制过程包括将线性矩阵应用于第一组解相关的声道。

22. 根据权利要求20所述的方法, 其中该组合涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合。

23. 根据权利要求20所述的方法, 其中应用第一解相关过程包括将相同的解相关过程应用于 N_t 个输入音频声道中的每个输入音频声道。

24. 根据权利要求20所述的方法, 进一步包括:

将第二解相关过程应用于输入音频声道中的所述一组两个或更多个输入音频声道以产生第二组解相关的声道, 第二解相关过程维持该组输入音频声道的声道间相关性; 以及

将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道, 其中组合涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及与所述两个或更多个未解相关的输出声道组合。

25. 根据权利要求24所述的方法, 其中第一解相关过程包括第一解相关函数, 并且第二解相关过程包括第二解相关函数, 第二解相关函数包括具有90度或-90度相移的第一解相关函数。

26. 根据权利要求25所述的方法, 其中第一调制过程包括第一调制函数, 并且第二调制过程包括第二调制函数, 第二调制函数包括具有90度或-90度相移的第一调制函数。

27. 一种非暂时性计算机可读介质, 存储有指令, 所述指令在由处理器执行时使得所述处理器执行包括以下的操作:

从接口接收包括 N_t 个输入音频声道的输入音频信号, 该输入音频信号表示具有第一声场格式分辨率的第一声场格式, N_t 是 ≥ 2 的整数;

将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道, 第一解相关过程维持该组输入音频声道的声道间相关性;

将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道;以及

将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号, N_p 是 ≥ 3 的整数,输出声道表示第二声场格式,所述第二声场格式为与第一声场格式相比具有相对较高分辨率的声场格式,该未解相关的输出声道与输出音频信号的较低分辨率分量相对应,以及解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应,

其中两个或更多个未解相关的输出声道中的每一个是通过将增益应用于 N_r 个输入音频声道之一而产生的。

28. 根据权利要求27所述的非暂时性计算机可读介质,其中,调制过程包括将线性矩阵应用于第一组解相关的声道。

29. 根据权利要求27所述的非暂时性计算机可读介质,其中该组合涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合。

30. 根据权利要求27所述的非暂时性计算机可读介质,其中应用第一解相关过程包括将相同的解相关过程应用于 N_r 个输入音频声道中的每个输入音频声道。

31. 根据权利要求27所述的非暂时性计算机可读介质,其中所述操作进一步包括:

将第二解相关过程应用于输入音频声道中的所述一组两个或更多个输入音频声道以产生第二组解相关的声道,第二解相关过程维持该组输入音频声道的声道间相关性;以及

将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道,其中组合涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及与所述两个或更多个未解相关的输出声道组合。

32. 根据权利要求31所述的非暂时性计算机可读介质,其中第一解相关过程包括第一解相关函数,并且第二解相关过程包括第二解相关函数,第二解相关函数包括具有90度或-90度相移的第一解相关函数。

33. 一种计算机程序产品,包括指令,所述指令在计算机执行时使得计算机执行根据权利要求1-9和20-26中任一项所述的方法。

用于对音频信号进行处理的方法、装置和设备

[0001] 本申请是申请号为201680011460.1、申请日为2016年3月2日、发明名称为“通过调制解相关进行的空间音频信号增强”的发明专利申请的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求2015年3月3日提交的美国临时申请No.62/127,613和2016年2月23日提交的美国临时申请No.62/298,905的优先权,其中每个申请全文通过引用并入本文。

技术领域

[0004] 本发明涉及包括多个音频声道的音频信号的操控,具体地涉及用于从具有较低分辨率空间特性的输入音频信号创建具有高分辨率空间特性的音频信号的方法。

背景技术

[0005] 多声道音频信号用于为终端收听者存储或传输收听体验,收听体验可以包括非常复杂的声学场景的印象。多声道信号可以承载使用数个常见约定来描述声学场景的信息,常见约定包括但不限于以下:

[0006] 离散扬声器声道:音频场景可能已经被以某种方式渲染以形成扬声器声道,这些扬声器声道当在适当布置的扩音器上回放时创建期望声学场景的错觉。离散扬声器声道格式的示例包括如现今的许多声音格式中使用的立体声、5.1或7.1信号。

[0007] 音频对象:音频场景可以被表示为一个或多个对象音频声道,这些对象音频声道当被收听者回放装备渲染时可以重新创建声学场景。在一些情况下,每个音频对象将伴随有元数据(隐式的或显式的),该元数据被渲染器用于将对象平移(pan)到收听者回放环境中的适当定位。音频对象格式的示例包括Dolby Atmos,其被用在蓝光盘上的丰富声轨和其他运动图像递送格式的输送中。

[0008] 声场声道:音频场景可以用声场格式(共同包含一个或多个音频对象的一组两个或更多个音频信号)表示,其中每个对象的空间定位被以平移增益的形式编码在空间格式中。声场格式的示例包括高保真立体声(Ambisonics)和高阶高保真立体声(两者在本领域中都是众所周知的)。

[0009] 本公开涉及遵照各种空间格式的多声道音频信号的修改。

[0010] 声场格式

[0011] N声道声场格式可以由其平移函数 $P_N(\varphi)$ 限定。具体地说, $G=P_N(\varphi)$,其中G表示增益值的 $[N \times 1]$ 列矢量,并且 φ 限定对象的空间定位。

$$[0012] \quad G_N = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_N \end{pmatrix} = P_N(\varphi) \quad (1)$$

[0013] 因此,一组M个音频对象($o_1(t), o_2(t), \dots, o_M(t)$)可以根据等式2被编码为N声道空间格式信号 $X_N(t)$ (其中音频对象m定位于由 φ_m 限定的位置处):

$$[0014] \quad X_N(t) = \sum_{m=1}^M P(\phi_m) \times a_m(t) \quad (2)$$

$$[0015] \quad X_N(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{pmatrix} \quad (3)$$

发明内容

[0016] 如本文中详细描述, 在一些实现中, 一种处理音频信号的方法可以涉及接收包括 N_r 个输入音频声道的输入音频信号。 N_r 可以是 ≥ 2 的整数。在一些示例中, 输入音频信号可以表示具有第一声场格式分辨率的第一声场格式。该方法可以涉及将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道。第一解相关过程可以涉及维持该组输入音频声道的声道间相关性。该方法可以涉及将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道。

[0017] 在一些实现中, 该方法可以涉及将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号。在一些示例中, N_p 可以是 ≥ 3 的整数。根据一些实现, 输出声道可以表示第二声场格式, 该第二声场格式是分辨率比第一声场格式相对较高的声场格式。在一些示例中, 未解相关的输出声道可以与输出音频信号的较低分辨率分量相对应, 并且解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应。在一些实现中, 未解相关的输出声道可以通过将最小二乘格式转换器应用于 N_r 个输入音频声道而产生。

[0018] 在一些示例中, 调制过程可以涉及将线性矩阵应用于第一组解相关的声道。在一些实现中, 组合可以涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合。根据一些实现, 应用第一解相关过程可以涉及将相同的解相关过程应用于 N_r 个输入音频声道中的每个输入音频声道。

[0019] 在一些实现中, 该方法可以涉及将第二解相关过程应用于输入音频声道中的该组两个或更多个输入音频声道以产生第二组解相关的声道。在一些示例中, 第二解相关过程可以涉及维持该组输入音频声道的声道间相关性。该方法可以涉及将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道。在一些实现中, 组合过程可以涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及与两个或更多个未解相关的输出声道组合。

[0020] 根据一些实现, 第一解相关过程可以涉及第一解相关函数, 并且第二解相关过程可以涉及第二解相关函数。在一些情况下, 第二解相关函数可以涉及应用具有大约90度或大约-90度相移的第一解相关函数。在一些示例中, 第一调制可以涉及第一调制函数, 并且第二调制过程可以涉及第二调制函数, 第二调制函数包括具有大约90度或大约-90度相移的第一调制函数。

[0021] 在一些示例中, 解相关、调制和组合过程可以产生输出音频信号, 以使得当输出音频信号被解码并且被提供给扬声器的阵列时: a) 扬声器的阵列中的能量的空间分布与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的能量的空间分布基本

上相同;以及b)扬声器的阵列中的相邻扩音器之间的相关性与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的相关性基本上不同。

[0022] 在一些示例中,接收输入音频信号可以涉及接收来自音频引导(steering)逻辑过程的第一输出。第一输出可以包括 N_r 个输入音频声道。在一些这样的实现中,该方法可以涉及将输出音频信号的 N_p 个音频声道与来自音频引导逻辑过程的第二输出组合。在一些情况下,第二输出可以包括被引导音频数据的 N_p 个音频声道,在被引导音频数据中,一个或多个声道的增益已经基于当前的主导声音方向而被改变。

[0023] 本文中所描述的方法中的一些或全部方法可以由一个或多个设备根据存储在非暂时性介质上的指令(例如,软件)执行。这样的非暂时性介质可以包括存储器设备,诸如本文中所描述的那些存储器设备,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。例如,软件可以包括用于控制用于接收输入音频信号的一个或多个设备的指令,该输入音频信号包括 N_r 个输入音频声道。 N_r 可以是 ≥ 2 的整数。在一些示例中,输入音频信号可以表示具有第一声场格式分辨率的第一声场格式。软件可以包括用于将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道的指令。第一解相关过程可以涉及维持该组输入音频声道的声道间相关性。软件可以包括用于将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道的指令。

[0024] 在一些实现中,软件可以包括用于将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号的指令。在一些示例中, N_p 可以是 ≥ 3 的整数。根据一些实现,输出声道可以表示第二声场格式,该第二声场格式是分辨率比第一声场格式相对较高的声场格式。在一些示例中,未解相关的输出声道可以与输出音频信号的较低分辨率分量相对应,并且解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应。在一些实现中,未解相关的输出声道可以通过将最小二乘格式转换器应用于 N_r 个输入音频声道而产生。

[0025] 在一些示例中,调制过程可以涉及将线性矩阵应用于第一组解相关的声道。在一些实现中,组合可以涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合。根据一些实现,应用第一解相关过程可以涉及将相同的解相关过程应用于 N_r 个输入音频声道中的每个输入音频声道。

[0026] 在一些实现中,软件可以包括用于将第二解相关过程应用于输入音频声道中的该组两个或更多个输入音频声道以产生第二组解相关的声道的指令。在一些示例中,第二解相关过程可以涉及维持该组输入音频声道的声道间相关性。软件可以包括用于将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道的指令。在一些实现中,组合过程可以涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及两个或更多个未解相关的输出声道组合。

[0027] 根据一些实现,第一解相关过程可以涉及第一解相关函数,并且第二解相关过程可以涉及第二解相关函数。在一些情况下,第二解相关函数可以涉及应用具有大约90度或大约-90度相移的第一解相关函数。在一些示例中,第一调制可以涉及第一调制函数,并且第二调制过程可以涉及第二调制函数,第二调制函数包括具有大约90度或大约-90度相移的第一调制函数。

[0028] 在一些示例中,解相关、调制和组合过程可以产生输出音频信号,以使得当输出音频信号被解码并且被提供给扬声器的阵列时:a)扬声器的阵列中的能量的空间分布与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的能量的空间分布基本上相同;以及b)扬声器的阵列中的相邻扩音器之间的相关性与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的相关性基本上不同。

[0029] 在一些示例中,接收输入音频信号可以涉及接收来自音频引导逻辑过程的第一输出。第一输出可以包括 N_r 个输入音频声道。在一些这样的实现中,软件可以包括用于将输出音频信号的 N_p 个音频声道与来自音频引导逻辑过程的第二输出组合的指令。在一些情况下,第二输出可以包括被引导音频数据的 N_p 个音频声道,在被引导音频数据中,一个或多个声道的增益已经基于当前的主导声音方向而被改变。

[0030] 本公开的至少一些方面可以在包括接口系统和控制系统的装置中实现。控制系统可以包括以下中的至少一个: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)或其他可编程逻辑器件、分立门或晶体管逻辑、或分立硬件组件。接口系统可以包括网络接口。在一些实现中,装置可以包括存储器系统。接口系统可以包括存储器系统的至少一部分(例如,至少一个存储器设备)和控制系统之间的接口。

[0031] 控制系统可以能够经由接口系统接收包括 N_r 个输入音频声道的输入音频信号。 N_r 可以是 ≥ 2 的整数。在一些示例中,输入音频信号可以表示具有第一声场格式分辨率的第一声场格式。控制系统可以能够将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道。第一解相关过程可以涉及维持该组输入音频声道的声道间相关性。控制系统可以能够将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道。

[0032] 在一些实现中,控制系统可以能够将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号。在一些示例中, N_p 可以是 ≥ 3 的整数。根据一些实现,输出声道可以表示第二声场格式,该第二声场格式是分辨率比第一声场格式相对较高的声场格式。在一些示例中,未解相关的输出声道可以与输出音频信号的较低分辨率分量相对应,并且解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应。在一些实现中,未解相关的输出声道可以通过将最小二乘格式转换器应用于 N_r 个输入音频声道而产生。

[0033] 在一些示例中,调制过程可以涉及将线性矩阵应用于第一组解相关的声道。在一些实现中,组合可以涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合。根据一些实现,应用第一解相关过程可以涉及将相同的解相关过程应用于 N_r 个输入音频声道中的每个输入音频声道。

[0034] 在一些实现中,控制系统可以能够将第二解相关过程应用于输入音频声道中的该组两个或更多个输入音频声道以产生第二组解相关的声道。在一些示例中,第二解相关过程可以涉及维持该组输入音频声道的声道间相关性。控制系统可以能够将第二调制过程应用于第二组解相关的声道以产生第二组解相关的且经调制的输出声道。在一些实现中,组合过程可以涉及将第二组解相关的且经调制的输出声道与第一组解相关的且经调制的输出声道以及与两个或更多个未解相关的输出声道组合。

[0035] 根据一些实现,第一解相关过程可以涉及第一解相关函数,并且第二解相关过程可以涉及第二解相关函数。在一些情况下,第二解相关函数可以涉及应用具有大约90度或大约-90度相移的第一解相关函数。在一些示例中,第一调制可以涉及第一调制函数,并且第二调制过程可以涉及第二调制函数,第二调制函数包括具有大约90度或大约-90度相移的第一调制函数。

[0036] 在一些示例中,解相关、调制和组合过程可以产生输出音频信号,以使得当输出音频信号被解码并且被提供给扬声器的阵列时:a)扬声器的阵列中的能量的空间分布与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的能量的空间分布基本上相同;以及b)扬声器的阵列中的相邻扩音器之间的相关性与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的相关性基本上不同。

[0037] 在一些示例中,接收输入音频信号可以涉及接收音频引导逻辑过程的第一输出。第一输出可以包括 N_r 个输入音频声道。在一些这样的实现中,控制系统可以能够将输出音频信号的 N_p 个音频声道与来自音频引导逻辑过程的第二输出组合。在一些情况下,第二输出可以包括被引导音频数据的 N_p 个音频声道,在被引导音频数据中,一个或多个声道的增益已经基于当前的主导声音方向而被改变。

附图说明

[0038] 为了更完整地理解本公开,参照以下描述和附图,在附图中:

[0039] 图1A示出被解码到扬声器的高分辨率声场格式的示例;

[0040] 图1B示出其中低分辨率声场格式在被解码到扬声器之前被格式转换为高分辨率的系统的示例;

[0041] 图2示出3声道、低分辨率声场格式在被解码到扬声器之前被格式转换为9声道、高分辨率声场格式;

[0042] 图3示出对于两个不同的声场格式的、来自角度 ϕ 处的输入音频对象的增益被编码到声场格式中、然后被解码到 $\phi_s=0$ 处的扬声器;

[0043] 图4示出来自角度 ϕ 处的输入音频对象的增益被编码到9声道BF4h声场格式中、然后被解码到9个扬声器的阵列;

[0044] 图5示出来自角度 ϕ 处的输入音频对象的增益被编码到3声道BF1h声场格式中、然后被解码到9个扬声器的阵列;

[0045] 图6示出用于从3声道BF1h声场格式创建9声道BF4h声场格式的(现有技术的)方法;

[0046] 图7示出用于从3声道BF1h声场格式创建9声道BF4h声场格式的(现有技术的)方法,其中增益被提升以补偿损失的功率;

[0047] 图8示出用于从3声道BF1h声场格式创建9声道BF4h声场格式的替代方法的一个示例;

[0048] 图9示出来自角度 ϕ 处的输入音频对象的增益被编码到3声道BF1h声场格式中、被格式转换为9声道BF4h声场格式、并且然后被解码到定位于位置 ϕ_s 处的扬声器;

[0049] 图10示出用于从3声道BF1h声场格式创建9声道BF4h声场格式的另一替代方法;

[0050] 图11示出用于渲染具有可变大小的对象的格式转换器的示例;

- [0051] 图12示出用于对上混系统中的扩散信号路径进行处理的格式转换器的示例；
 [0052] 图13是示出能够执行本文中所描述的各种方法的装置的组件的示例的框图；以及
 [0053] 图14是示出本文中所描述的方法的示例方框的流程图。

具体实施方式

[0054] 图1A中示出了现有技术的过程，由此平移函数在平移器A[1]内部被用于产生 N_p 声道原始声场信号[5], $Y(t)$ ，该信号随后被扬声器解码器[4] ($[N_s \times N_p]$ 矩阵) 解码为一组 N_s 个扬声器信号。

[0055] 一般来说，声场格式可以在回放扬声器布置未知的情形下使用。最终收听体验的质量将取决于以下两者：(a) 声场格式的信息承载容量；以及 (b) 回放环境中所使用的扬声器的数量和布置。

[0056] 如果我们假定扬声器的数目大于或等于 N_p (所以, $N_s \geq N_p$)，则空间回放的感知质量将受 N_p 限制, N_p 是原始声场信号[5]中的声道的数目。

[0057] 通常，平移器A[1]将利用特定的平移函数族，该平移函数族被称为B格式 (在文献中也被称为球谐、高保真立体声、或高阶高保真立体声、平移规则)，并且本公开最初涉及基于B格式的平移规则的空间格式。

[0058] 图1B示出了替代平移器，平移器B[2]，其被配置为产生输入声场信号[6], N_r 声道空间格式 $x(t)$ ，该信号然后被格式转换器[3]处理以创建 N_p 声道输出声场信号[7], $y(t)$ ，其中 $N_p > N_r$ 。

[0059] 本公开描述了用于实现格式转换器[3]的方法。例如，本公开提供了可以用于构造在格式转换器[3]中所使用的线性时不变 (LTI) 滤波器的方法，以便为我们的格式转换器[3]提供 N_r 输入、 N_p 输出LTI传递函数，使得由图1B的系统提供的收听体验在感知上尽可能地接近图1A的系统的收听体验。

[0060] 示例——BF1H到BF4H

[0061] 我们从示例情景开始，其中图1A的平移器A[1]被配置为根据以下平移器函数来产生4阶水平B格式声场 (注意，术语BF4h用于指示水平4阶B格式)：

$$[0062] \quad P_A(\phi) = P_{BF4h}(\phi) = \begin{pmatrix} 1 \\ \sqrt{2}\cos\phi \\ \sqrt{2}\sin\phi \\ \sqrt{2}\cos 2\phi \\ \sqrt{2}\sin 2\phi \\ \sqrt{2}\cos 3\phi \\ \sqrt{2}\sin 3\phi \\ \sqrt{2}\cos 4\phi \\ \sqrt{2}\sin 4\phi \end{pmatrix} \quad (4)$$

[0063] 在这种情况下，变量 ϕ 表示方位角, $N_p=9$, $P_{BF4h}(\phi)$ 表示 $[9 \times 1]$ 列矢量 (因此，信号 $Y(t)$ 将包括9个音频声道)。

[0064] 现在，让我们假定图1B的平移器B[2]被配置为产生1阶B格式声场：

$$[0065] \quad P_B(\phi) = P_{BF1h}(\phi) = \begin{pmatrix} 1 \\ \sqrt{2}\cos\phi \\ \sqrt{2}\sin\phi \end{pmatrix} \quad (5)$$

[0066] 因此,在该示例中, $N_r=3$, $P_{BF1h}(\phi)$ 表示 $[3 \times 1]$ 列矢量(并且因此,图1B的信号 $X(t)$ 将包括3个音频声道)。在该示例中,我们的目标是创建图1B的9声道输出声场信号 $[7]$, $Y(t)$,以使得达到优化的收听体验,该9声道输出声场信号是通过LTI过程从 $X(t)$ 导出的,适合于解码到任何扬声器阵列。

[0067] 如图2所示,我们将将该LTI格式转换过程的传递函数称为 H 。

[0068] 扬声器解码器线性矩阵

[0069] 在图1B所示的示例中,格式转换器[3]接收作为输入的 N_r 声道输入声场信号[6],并且输出 N_p 声道输出声场信号[7]。格式转换器[3]一般将不接收关于收听者回放环境中的最终扬声器布置的信息。如果我们选择假定收听者具有足够大数目的扬声器(即前述假定, $N_s \geq N_p$),则我们可以安全地忽略扬声器布置,但是本公开中所描述的方法仍将为其回放环境具有较少扬声器的收听者产生适当的收听体验。

[0070] 话虽如此,将方便的是能够通过示出当空间格式信号 $Y(t)$ 和 $Y(t)$ 最后被解码到扩音器时的最终结果来示出本文件中所描述的格式转换器的行为。

[0071] 为了将 N_p 声道声场信号 $Y(t)$ 解码到 N_s 个扬声器,可以如下将 $[N_s \times N_p]$ 矩阵应用于声场信号:

$$[0072] \quad \text{Spkr}(t) = \text{DecodeMatrix} \times Y(t) \quad (6)$$

[0073] 如果我们将注意力集中于一个扬声器,则我们可以忽略阵列中的其他扬声器,并且看DecodeMatrix(解码矩阵)的一行。我们将这称为解码行矢量(DecodeRow Vector), $\text{Dec}_N(\phi_s)$,该解码行矢量指示DecodeMatrix的该行意图将 N 声道声场信号解码到定位于角度 ϕ_s 处的扬声器。

[0074] 对于等式4和等式5中所描述的种类的B格式信号,可以如下计算解码行矢量:

$$[0075] \quad \text{Dec}_3(\phi_s) = \frac{1}{3} P_{BF1h}(\phi)^T \quad (7)$$

$$[0076] \quad \frac{1}{3} P_{BF1h}(\phi)^T = \frac{1}{3} (1 \quad \sqrt{2}\cos\phi_s \quad \sqrt{2}\sin\phi_s) \quad (8)$$

$$[0077] \quad \text{Dec}_9(\phi_s) = \frac{1}{9} P_{BF4h}(\phi)^T \quad (9)$$

$$[0078] \quad \frac{1}{9} P_{BF4h}(\phi)^T = \frac{1}{9} (1 \quad \sqrt{2}\cos\phi_s \quad \dots \quad \sqrt{2}\cos 4\phi_s \quad \sqrt{2}\sin 4\phi_s) \quad (10)$$

[0079] 注意, $\text{Dec}_3(\phi_s)$ 在这里示出以允许我们检查3声道BF1h信号被解码到扬声器的假设性情境。然而,只有9声道扬声器解码行矢量 $\text{Dec}_9(\phi_s)$ 被用在图2中所示的系统的一些实现中。

[0080] 注意,此外,可以使用解码行矢量 $\text{Dec}_9(\phi_s)$ 的替代形式来创建具有其他期望的性质的扬声器平移曲线。本文件的意图不是定义最佳的扬声器解码器系数,并且本文中所公开的实现的值不取决于扬声器解码器系数的选择。

[0081] 从输入音频对象到扬声器的总体增益

[0082] 我们现在可以将来自图2的三个主要处理块放在一起,并且这将允许我们定义平

移到定位 φ 的输入音频对象将出现在馈送给定位于收听者回放环境中的位置 φ_s 处的扬声器的信号中的方式:

$$[0083] \quad \text{gain}_{3,9}(\varphi, \varphi_s) = \text{Dec}_9(\varphi_s) \times H \times P_3(\varphi) \quad (\text{II})$$

[0084] 在等式11中, $P_3(\varphi)$ 表示将定位 φ 处的输入音频对象平移到BF1h格式中的增益值的 $[3 \times 1]$ 矢量。

[0085] 在该示例中, H 表示执行从BF1h格式到BF4h格式的格式转换的 $[9 \times 3]$ 矩阵。

[0086] 在等式11中, $\text{Dec}_9(\varphi_s)$ 表示将BF4h信号解码到定位于收听环境中的位置 φ_s 处的扩音器的 $[1 \times 9]$ 行矢量。

[0087] 为了比较,我们还可以定义图1A中所示的(现有技术的)系统的端到端增益,该系统不包括格式转换器。

$$[0088] \quad \text{gain}_9(\varphi, \varphi_s) = \text{Dec}_9(\varphi_s) \times P_9(\varphi) \quad (12)$$

[0089] 图3中的点线示出了当定位于方位角 φ 处的音频对象被平移到BF4h声场格式中(经由增益矢量 $G_{\text{BF4h}}(\varphi)$) 并且然后被用解码行矢量 $\text{Dec}_9(0)$ 解码时从该对象到定位于 $\varphi_s = 0$ 处的扬声器的总体增益, $\text{gain}_9(\varphi, \varphi_s)$ 。

[0090] 该增益绘图示出,当原始对象定位于与扬声器相同位置处(在 $\varphi = 0$ 处)时,从该对象到扬声器的最大增益发生,并且随着该对象远离扬声器移动,增益快速地降到零(在 $\varphi = 40^\circ$ 处)。

[0091] 另外,图3中的实线示出了当对象被平移到BF1h 3声道声场格式中,并且然后被用 $\text{Dec}_3(0)$ 解码行矢量解码到扬声器阵列时的增益 $\text{gain}_3(\varphi, \varphi_s)$ 。

[0092] 低分辨率信号 $X(T)$ 中丢失了什么

[0093] 当多个扬声器被放置在围绕收听者的圆中时,图3中所示的增益曲线可以被重新绘制,以示出所有的扬声器增益。这允许我们看到扬声器如何彼此相互作用。

[0094] 例如,当9个扬声器被围绕收听者以 40° 间隔放置时,分别对于9声道情况和3声道情况,在图4和图5中示出所得的一组9个增益曲线。

[0095] 在图4和图5两者中,定位于 $\varphi_s = 0$ 处的扬声器处的增益被绘制为实线,其他扬声器则用点线绘制。

[0096] 看图4,我们可以看出,当对象定位于 $\varphi = 0$ 处时,用于该对象的音频信号将被以增益1.0呈现到前置扬声器(在 $\varphi_s = 0$ 处)。此外,来自该对象的音频信号将被以增益0.0呈现到所有的其他的扬声器。

[0097] 定性地说,基于图4的观察,我们可以说,从定位于 $\varphi = 0$ 处的对象将出现在前置扬声器中、其他8个扬声器中没有能量的意义上讲,BF4h声场格式当通过 $\text{Dec}_9(\varphi_s)$ 解码行矢量被解码时通过9个扬声器提供高质量渲染。

[0098] 不幸的是,相同的定性评估不能关于图5做出,图5示出了当BH1h声场格式被解码到9个扬声器时的结果。

[0099] 图5的增益曲线的缺陷可以关于两个不同的属性来描述:

[0100] 功率分布:当对象定位于 $\varphi = 0$ 处时,当所有功率被施加于前置扬声器(在 $\varphi_s = 0$ 处)并且零功率被施加于其他8个扬声器时,对于扩音器的最优功率分布将发生。BF1h解码器没有实现该能量分布,因为大量功率散布到其他扬声器。

[0101] 过大相关性:当定位于 $\varphi = 0$ 处的对象被用BF1h声场格式编码并且被用 $\text{Dec}_3(\varphi_s)$

解码行矢量解码时,五个前置扬声器(在 $\varphi_s = -80^\circ$ 、 -40° 、 0° 、 40° 和 80° 处)将包含相同的音频信号,导致这五个扬声器之间的相关性水平高。此外,后置的两个扬声器(在 $\varphi_s = -160^\circ$ 和 160° 处)与前置声道将是异相的(out-of-phase)。最终结果是收听者将体验到不舒服的相位感觉(phasey feeling),并且收听者的小的移动将导致显著的梳状伪像(combing artifact)。

[0102] 现有技术的方法已经尝试通过添加解相关的信号分量来解决过大相关性问题,导致功率分布问题变得更糟。

[0103] 本文中所公开的一些实现可以在保持相同的功率分布的同时降低扬声器声道之间的相关性。

[0104] 设计更好的格式转换器

[0105] 从等式4和等式5,我们可以看出,定义BF1h格式的三个平移增益值是定义BF4h格式的九个平移增益值的子集。因此,低分辨率信号 $X(t)$ 可能已经通过简单的线性投影 M_p 从高分辨率信号 $Y(t)$ 导出:

$$[0106] \quad X(t) = M_p \times Y'(t) \quad (13)$$

$$[0107] \quad M_p \times Y'(t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \times Y'(t) \quad (14)$$

[0108] 回想一下,图1中的格式转换器[3]的一个目的是重新生成新信号 $Y(t)$,该新信号为终端收听者提供与由更精确的信号 $Y(t)$ 传送的体验接近地匹配的声学体验。可以通过求取 M_p 的伪逆来计算格式转换器 H_{LS} 的操作的最小均方最优选择:

$$[0109] \quad Y_{LS}(t) = H_{LS} \times X(t) \quad (15)$$

$$[0110] \quad \text{其中, } H_{LS} = M_p^+ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (16)$$

[0111] 在等式(16)中, M_p^+ 表示Moore-Penrose伪逆,其在本领域中是众所周知的。

[0112] 这里所使用的术语意图传达最小二乘解是通过使用格式转换矩阵 H_{LS} 以产生新的9声道信号 $Y_{LS}(t)$ 来进行操作的, $Y_{LS}(t)$ 从最小二乘的意义上讲与 $Y(t)$ 尽可能接近地匹配。

[0113] 虽然最小二乘解($H_{LS} = M^+$)提供数学意义上的最佳拟合,但是收听者将发现结果的振幅太低,因为3声道BF1h声场格式与其中6个声道被扔掉的9声道BF4h格式是相同的,如图6中所示。从而,最小二乘解涉及消除声学场景的功率的2/3。

[0114] 如图7所示,一个(小的)改进可以来自于简单地放大结果。在一个这样的示例中,通过将增益 g_{LS} 施加于非零分量 $x_1(t)$ - $x_3(t)$ 来产生最小二乘解的非零分量 $y_1(t)$ - $y_3(t)$:

$$[0115] \quad H_{LS^t} = g_{LS} H_{LS} \quad (17)$$

$$[0116] \quad \text{其中, } g_{LS} = \sqrt{\frac{N_p}{N_r}} \quad (18)$$

$$[0117] \quad \sqrt{\frac{N_p}{N_r}} = \sqrt{3} \quad (19)$$

[0118] 用于解相关的调制方法

[0119] 虽然图6和图7的格式转换将为收听者提供一定程度上可接受的回放体验,但是它们可以在相邻扬声器之间产生非常大程度的相关性,如图5中的重叠曲线证明的那样。

[0120] 更好的替代方法是使用BF1h输入信号的解相关版本来将较多能量添加到BF4h信号的较高阶次的项,而不是仅提升低分辨率信号分量(如图7中所做的那样)。

[0121] 本文中所公开的一些实现涉及限定从 $X(t)$ 的一个或多个低分辨率声场分量(例如, $x_1(t)$ 、 $x_2(t)$ 和 $x_3(t)$)合成 $Y(t)$ 的一个或多个较高阶分量(例如, $y_4(t)$ 、 $y_5(t)$ 、 $y_6(t)$ 、 $y_7(t)$ 、 $y_8(t)$ 和 $y_9(t)$)的近似的方法。

[0122] 为了创建 $Y(t)$ 的较高阶分量,一些示例利用解相关器。我们将使用符号 Δ 来表示获取输入音频信号并且产生被人类收听者感知为与输入信号解相关的输出信号的操作。

[0123] 在各种出版物中关于用于实现解相关器的方法撰写了很多。为简单起见,在本文件中,我们将定义两个计算效率高的、包括256采样延迟和512采样延迟的解相关器(使用本领域技术人员熟悉的 z 变换记号):

$$[0124] \quad \Delta_1 = z^{-256} \quad (20)$$

$$[0125] \quad \Delta_2 = z^{-512} \quad (21)$$

[0126] 以上解相关器仅仅是示例。在替代实现中,其他解相关方法(诸如本领域的普通技术人员众所周知的其他解相关方法)可以用来代替或补充本文中所描述的解相关方法。

[0127] 为了创建 $Y(t)$ 的较高阶分量,一些示例涉及选择一个或多个解相关器(诸如图8的 Δ_1 和 Δ_2)以及对应的调制函数(诸如 $\text{mod}_1(\varphi_s) = \cos 3\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 3\varphi_s$)。在该示例中,我们还定义不作为(do nothing)解相关器和调制器函数, $\Delta_0 = 1$ 和 $\text{mod}_0(\varphi_s) = 1$ 。然后,对于每个调制函数,我们接着进行这些步骤:

[0128] 1. 我们被给予调制函数 $\text{mod}_k(\varphi_s)$ 。我们旨在构造 $[N_p \times N_r]$ 矩阵($[9 \times 3]$ 矩阵), Q_k 。

[0129] 2. 形成乘积:

$$[0130] \quad p = \text{mod}_k \times \text{Dec}_9(\varphi_s) \times H_{LS}$$

[0131] 乘积 p 将是行矢量($[1 \times 3]$ 矢量),其中每个元素是关于 φ_s 的 \sin 和 \cos 函数的代数表达。

[0132] 3. 求解,以找到满足以下恒等式的(唯一)矩阵 Q_k :

$$[0133] \quad p \equiv \text{Dec}_9(\varphi_s) \times Q_k$$

[0134] 注意,根据该方法,当 $k=0$ 时,在以上过程中使用不作为解相关器 $\Delta_0 = 1$ (这不是真正的解相关器)和不作为调制器函数 $\text{mod}_0(\varphi_s) = 1$ 来计算 $Q_0 = H_{LS}$ 。

[0135] 因此,与调制函数 $\text{mod}_0(\varphi_s) = 1$ 、 $\text{mod}_1(\varphi_s) = \cos 3\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 3\varphi_s$ 相对应的三个 Q 矩阵是:

$$[0136] \quad Q_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (22)$$

$$[0137] \quad Q_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (23)$$

$$[0138] \quad Q_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \quad (24)$$

[0139] 在该示例中,该方法通过将总体传递函数定义为 $[9 \times 3]$ 矩阵来实现格式转换器:

$$[0140] \quad H_{\text{mod}} = g_0 \times Q_0 + g_1 \times Q_1 \times \Delta_1 + g_2 \times Q_2 \times \Delta_2 \quad (25)$$

[0141] 注意,通过设置 $g_0=1$ 和 $g_1=g_2=0$,我们的系统在这些条件下回复到与最小二乘格式转换器相同。

[0142] 此外,通过设置 $g_0=\sqrt{3}$ 和 $g_1=g_2=0$,我们的系统在这些条件下回复到与增益提升的最小二乘格式转换器相同。

[0143] 最后,通过设置 $g_0=1$ 和 $g_1=g_2=\sqrt{2}$,我们得到其中整个格式转换器的传递函数可以被写为如下等式的实施例:

$$[0144] \quad H_{mod} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \frac{\Delta_1}{\sqrt{2}} & \frac{-\Delta_2}{\sqrt{2}} \\ 0 & \frac{\Delta_2}{\sqrt{2}} & \frac{\Delta_1}{\sqrt{2}} \\ \Delta_1 & 0 & 0 \\ \Delta_2 & 0 & 0 \\ 0 & \frac{\Delta_1}{\sqrt{2}} & \frac{-\Delta_2}{\sqrt{2}} \\ 0 & \frac{\Delta_2}{\sqrt{2}} & \frac{\Delta_1}{\sqrt{2}} \end{pmatrix} \quad (26)$$

[0145] 图8中示出了用于实现一种这样的方法的框图。注意,第一调制器[9]接收来自解相关器 Δ_1 的输出,这旨在指示在该示例中所有三个声道被用相同的解相关器修改,使得三个输出信号可以被表达为:

$$[0146] \quad x_1^{dec_1} = \Delta_1 \times x_1(t)$$

$$[0147] \quad x_2^{dec_1} = \Delta_1 \times x_2(t) \quad (27)$$

$$[0148] \quad x_3^{dec_1} = \Delta_1 \times x_3(t)$$

[0149] 在等式(27)中, $x_1(t)$ 、 $x_2(t)$ 和 $x_3(t)$ 表示到第一解相关器[8]的输入。同样地,对于图8中的第二调制器[11],我们有:

$$x_1^{dec_2} = \Delta_2 \times x_1(t)$$

$$[0150] \quad x_2^{dec_2} = \Delta_2 \times x_2(t) \quad (28)$$

$$x_3^{dec_2} = \Delta_2 \times x_3(t)$$

[0151] 为了解释该方法后面的基本原理,我们看图9中的实线曲线。该曲线示出 $gain_{2,9}^{Q_0}(0, \phi_s)$, 定位于 $\phi = 0$ 处的对象将以该增益出现在定位于 ϕ_s 处的扬声器中(如果三声道BF1h信号使用矩阵 $Q_0 = H_{LS}$ 而被转换为9声道BF4h格式)。如果数个扬声器存在于收听者回放环境中,定位于 -120° 和 $+120^\circ$ 之间的方位角处,则这些扬声器将全部包含具有正增益的对象音频信号的某个分量。因此,所有这些扬声器将包含相关信号。

[0152] 这里所示的用虚线和点线绘制的其他两个增益曲线是 $gain_{2,9}^{Q_1}(0, \phi_s)$ 和 $gain_{2,9}^{Q_2}(0, \phi_s)$ (用于 $\phi = 0$ 处的对象的增益函数,因为当格式转换分别根据 Q_1 和 Q_2 被应用时,它将出现在位置 ϕ_s 处的扬声器处)。这两个增益函数合起来将承载与实线相同的功率,但是分隔大于 40° 的两个扬声器将不会以相同的方式相关。

[0153] 一个非常期望的结果(从主观视角,根据收听者偏好)涉及这三个增益曲线的混合,其中混合系数(g_0 、 g_1 和 g_2)由收听者偏好测试确定。

[0154] 使用希尔伯特变换来形成 Δ_2

[0155] 在替代实施例中,第二解相关器可以被以下等式取代:

$$[0156] \quad \Delta_2 = -H\{\Delta_1\} \quad (29)$$

[0157] 在等式29中,H表示希尔伯特变换,该变换有效地意味着我们的第二解相关过程与我们的第一解相关过程是相同的,其中附加有 90° 的相移(希尔伯特变换)。如果我们将关于 Δ_2 的这个表达式代入图8中的第二解相关器[10]中,则我们得到图10中的新图。

[0158] 在一些这样的实现中,第一解相关过程涉及第一解相关函数,并且第二解相关过程涉及第二解相关函数。第二解相关函数可以等于具有大约90度或大约-90度相移的第一解相关函数。在一些这样的示例中,大约90度的角度可以是89度至91度范围内的角度、88度至92度范围内的角度、87度至93度范围内的角度、86度至94度范围内的角度、85度至95度范围内的角度、84度至96度范围内的角度、83度至97度范围内的角度、82度至98度范围内的角度、81度至99度范围内的角度、80度至100度范围内的角度等等。类似地,在一些这样的示例中,大约-90度的角度可以是-89度至-91度范围内的角度、-88度至-92度范围内的角度、-87度至-93度范围内的角度、-86度至-94度范围内的角度、-85度至-95度范围内的角度、-84度至-96度范围内的角度、-83度至-97度范围内的角度、-82度至-98度范围内的角度、-81度至-99度范围内的角度、-80度至-100度范围内的角度等等。在一些实现中,相移可以随着频率而变化。根据一些这样的实现,相移可以仅在某个感兴趣的频率范围上约为90度。在一些这样的示例中,感兴趣的频率范围可以包括从300Hz至2kHz的范围。其他示例可以应用其他相移和/或可以在其他频率范围上应用大约90度的相移。

[0159] 替代调制函数的使用

[0160] 在本文中所公开的各种示例中,第一调制过程涉及第一调制函数,并且第二调制过程涉及第二调制函数,第二调制函数是具有大约90度或大约-90度相移的第一调制函数。在以上参照图8描述的过程中,BF1h输入信号到BF4h输出信号的转换涉及第一调制函数 $\text{mod}_1(\varphi_s) = \cos 3\varphi_s$ 和第二调制函数 $\text{mod}_2(\varphi_s) = \sin 3\varphi_s$ 。然而,其他实现也可以使用其他调制函数来实现,在其他调制函数中,第二调制函数是具有大约90度或大约-90度相移的第一调制函数。

[0161] 例如,调制函数 $\text{mod}_1(\varphi_s) = \cos 2\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 2\varphi_s$ 的使用导致替代的Q矩阵的计算:

$$[0162] \quad Q_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (30)$$

$$[0163] \quad Q_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (31)$$

$$[0164] \quad Q_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (32)$$

[0165] 替代输出格式的使用

[0166] 在前一节中给出的示例使用替代调制函数 $\text{mod}_1(\varphi_s) = \cos 2\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 2\varphi_s$ 得出在最后两行中包含零的Q矩阵。结果,这些替代调制函数允许输出格式约简为7声道BF3h格式,其中Q矩阵约简为7行:

$$[0167] \quad Q_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (33)$$

$$[0168] \quad Q_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \quad (34)$$

$$[0169] \quad Q_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \quad (35)$$

[0170] 在替代实施例中,Q矩阵也可以约简为更少数目的行,以便减少输出格式中的声道的数目,得出以下Q矩阵:

$$[0171] \quad Q_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (36)$$

$$[0172] \quad Q_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (37)$$

$$[0173] \quad Q_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad (38)$$

[0174] 其他声场格式

[0175] 其他声场输入格式也可以根据本文中所描述的方法处理,包括:

[0176] BF1 (4声道1阶高保真立体声,也被称为WXYZ格式),其可以使用诸如 $\text{mod}_1(\varphi_s) = \cos 3\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 3\varphi_s$ 的调制函数而被格式转换为BF3 (16声道3阶高保真立体声);

[0177] BF1 (4声道1阶高保真立体声,也被称为WXYZ格式),其可以使用诸如 $\text{mod}_1(\varphi_s) = \cos 2\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 2\varphi_s$ 的调制函数而被格式转换为BF2 (9声道2阶高保真立体声);或

[0178] BF2 (9声道2阶高保真立体声,也被称为WXYZ格式),其可以使用诸如 $\text{mod}_1(\varphi_s) = \cos 4\varphi_s$ 和 $\text{mod}_2(\varphi_s) = \sin 4\varphi_s$ 的调制函数而被格式转换为BF3 (16声道6阶高保真立体声)。

[0179] 将意识到,如本文中所定义的调制方法适用于范围广泛的声场格式。

[0180] 用于渲染具有大小的对象的格式转换器

[0181] 图11示出了适合于渲染音频对象的系统,其中格式转换器[3]用于从较低分辨率BF1h信号 $x_1(t) \cdots x_3(t)$ 创建9声道BF4h信号 $y_1(t) \cdots y_9(t)$ 。

[0182] 在图11中所示的示例中,音频对象 $o_1(t)$ 被平移以形成中间9声道BF4h信号 $z_1(t) \cdots z_9(t)$ 。该高分辨率信号经由直接增益缩放器(scaler) [15]被求和到BF4h输出,从而允许音频对象 $o_1(t)$ 被以高分辨率表示在BF4h输出中(所以它将对收听者表现为紧凑的对象)。

[0183] 另外,在该实现中,BF4h信号的0阶分量和1阶分量(分别为 $z_1(t)$ 和 $z_2(t) \cdots z_3(t)$)被零阶增益缩放器[17]和一阶增益缩放器[16]修改,以形成3-声道BF1h信号 $x_1(t) \cdots x_3(t)$ 。

[0184] 在该示例中,如下,通过大小过程[14]根据与对象相关联的 $size_1$ 参数来生成三个增益控制信号:

[0185] 当 $size_1=0$ 时,增益值是:

[0186] $\{Size=0\} \{Gain_{ZerothGain}=0, Gain_{FirstGain}=0, Gain_{DirectGain}=1\}$

[0187] 当 $size_1=1/2$ 时,增益值是:

[0188] $\{size = \frac{1}{2}\} \{Gain_{ZerothGain} = 1, Gain_{FirstGain} = 1, Gain_{DirectGain} = 0\}$

[0189] 当 $size_1=1$ 时,增益值是:

[0190] $\{size = 1\} \{Gain_{ZerothGain} = \sqrt{3}, Gain_{FirstGain} = 0, Gain_{DirectGain} = 0\}$

[0191] 在该示例中,具有 $size=0$ 的音频对象对应于本质上是点源的音频对象,并且具有 $size=1$ 的音频对象对应于大小等于整个回放环境(例如,整个房间)的大小的音频对象。在一些实现中,对于在0和1之间的 $size_1$ 的值,三个增益参数的值将按照分段线性函数变化,分段线性函数可以基于这里定义的值。

[0192] 根据该实现,通过缩放BF4h信号的零阶分量和一阶分量而形成的BF1h信号被传递通过格式转换器(例如,如前述类型),以便生成格式转换的BF4h信号。直接的BF4h信号和格式转换的BF4h信号然后被组合,以便形成大小经过调整的BF4h输出信号。通过调整直接、零阶和一阶增益缩放器,平移到BF4h输出信号的对象感知大小可以在点源和非常大的源(例如,涵盖整个房间)之间变化。

[0193] 在上混器中使用的格式转换器

[0194] 上混器(诸如图12中所示的上混器)通过使用引导逻辑过程[18]进行操作,引导逻辑过程[18]采取低分辨率声场信号(例如,BF1h)作为输入。例如,引导逻辑过程[18]可以识别要被尽可能精确地引导的输入声场信号的分量(并且对这些分量进行处理以形成高分辨率输出信号 $z_1(t) \cdots z_9(t)$)。例如,引导逻辑过程[18]可以基于当前的主导声音方向来改变一个或多个声道的增益,并且可以输出被引导音频数据的 N_p 个音频声道。在图12中所示的示例中, $p=9$,因此,引导逻辑过程[18]输出被引导音频数据的9个声道。

[0195] 除了输入信号的这些被引导分量之外,在该示例中,引导逻辑过程[18]将发射残留信号 $x_1(t) \cdots x_3(t)$ 。该残留信号包含没有被引导以形成高分辨率信号 $z_1(t) \cdots z_9(t)$ 的音频分量。

[0196] 在图12中所示的示例中,该残留信号 $x_1(t) \cdots x_3(t)$ 被格式转换器[3]处理,以提供残留信号的较高分辨率版本,该版本适合于与被引导信号 $z_1(t) \cdots z_9(t)$ 组合。从而,图12示出了被引导音频数据的 N_p 个音频声道与格式转换器的输出音频信号的 N_p 个音频声道组合以便产生上混的BF4h输出信号的示例。而且,假如生成BF1h残留信号并且将格式转换器应用于该信号以生成转换的BF4h残留信号的计算复杂度低于使用引导逻辑直接将残留信号上混为BF4h格式的计算复杂度,则实现计算复杂度降低的上混。因为残留信号的感知相关性

低于主导信号,所以使用如图12中所示的上混器生成的所得的上混的BF4h输出信号在感知上将类似于由例如使用引导逻辑直接生成高精度主导和残留BF4h输出信号两者的上混器生成的BF4h输出信号,但是可以以降低的计算复杂度生成。

[0197] 图13是提供能够实现本文中所描述的各种方法的装置的组件的示例的框图。装置1300可以例如是音频数据处理系统(或者可以是音频数据处理系统的一部分)。在一些示例中,装置1300可以在另一设备的组件中实现。

[0198] 在该示例中,装置1300包括接口系统1305和控制系统1310。控制系统1310可以能够实现本文中所公开的方法中的一些或全部。控制系统1310可以例如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)或其他可编程逻辑器件、分立门或晶体管逻辑、和/或分立硬件组件。

[0199] 在该实现中,装置1300包括存储器系统1315。存储器系统1315可以包括一种或多种合适类型的非暂时性存储介质,诸如闪存、硬盘驱动器等。接口系统1305可以包括网络接口、控制系统和存储器系统之间的接口和/或外部设备接口(诸如通用串行总线(USB)接口)。尽管存储器系统1315在图13中被描绘为单独的元件,但是控制系统1310可以包括可以被认为是存储器系统的一部分的至少一些存储器。类似地,在一些实现中,存储器系统1315可以能够提供一些控制系统功能。

[0200] 在该示例中,控制系统1310能够经由接口系统1305接收音频数据和其他信息。在一些实现中,控制系统1310可以包括(或者可以实现)音频处理装置。

[0201] 在一些实现中,控制系统1310可以能够根据存储在一个或多个非暂时性介质上的软件来执行本文中所描述的方法中的至少一些。非暂时性介质可以包括与控制系统1310相关联的存储器,诸如随机存取存储器(RAM)和/或只读存储器(ROM)。非暂时性介质可以包括存储器系统1315的存储器。

[0202] 图14是示出根据一些实现的格式转换过程的示例方框的流程图。图14的方框(以及本文中提供的其他流程图的方框)可以例如由图13的控制系统1310或类似装置执行。从而,下面参照图13的一个或多个元件来描述图14的一些方框。与本文中所公开的其他方法一样,图14中概括的方法可以包括比指示的方框多或少的方框。而且,本文中所公开的方法的方框不一定按指示的次序执行。

[0203] 这里,方框1405涉及接收包括 N_t 个输入音频声道的输入音频信号。在该示例中, N_t 是 ≥ 2 的整数。根据该实现,输入音频信号表示具有第一声场格式分辨率的第一声场格式。在一些示例中,第一声场格式可以是3声道BF1h声场格式,而在其他示例中,第一声场格式可以是BF1(4声道1阶高保真立体声,也被称为WXYZ格式)、BF2(9声道2阶高保真立体声)格式或另一声场格式。

[0204] 在图14中所示的示例中,方框1410涉及将第一解相关过程应用于输入音频声道中的一组两个或更多个输入音频声道以产生第一组解相关的声道。根据该示例,第一解相关过程维持该组输入音频声道的声道间相关性。第一解相关过程可以例如与以上参照图8和图10描述的解相关器 Δ_1 的实现中的一个相对应。在这些示例中,应用第一解相关过程涉及将相同的解相关过程应用于 N_t 个输入音频声道中的每个输入音频声道。

[0205] 在该实现中,方框1415涉及将第一调制过程应用于第一组解相关的声道以产生第一组解相关的且经调制的输出声道。第一调制过程可以例如与以上参照图8描述的第一调

制器[9]的实现中的一个或与以上参照图10描述的调制器[13]的实现中的一个相对应。从而,调制过程可以涉及将线性矩阵应用于第一组解相关的声道。

[0206] 根据该示例,方框1420涉及将第一组解相关的且经调制的输出声道与两个或更多个未解相关的输出声道组合以产生包括 N_p 个输出音频声道的输出音频信号。在该示例中, N_p 是 ≥ 3 的整数。在该实现中,输出声道表示第二声场格式,该第二声场格式是分辨率比第一声场格式相对较高的声场格式。在一些这样的示例中,第二声场格式是9声道BF4h声场格式。在其他示例中,第二声场格式可以是另一声场格式,诸如7声道BF3h格式、5声道BF3h格式、BF2声场格式(9声道2阶高保真立体声)、BF3声场格式(16声道3阶高保真立体声)或另一声场格式。

[0207] 根据该实现,未解相关的输出声道与输出音频信号的较低分辨率分量相对应,解相关的且经调制的输出声道与输出音频信号的较高分辨率分量相对应。参照图8和图10,例如,输出声道 $y_1(t)$ - $y_3(t)$ 提供未解相关的输出声道的示例。从而,在这些示例中,组合涉及将第一组解相关的且经调制的输出声道与 N_r 个未解相关的输出声道组合,其中 $N_r=3$ 。在一些这样的实现中,未解相关的输出声道通过将最小二乘格式转换器应用于 N_r 个输入音频声道而产生。在图10中所示的示例中,输出声道 $y_4(t)$ - $y_9(t)$ 提供通过第一解相关过程和第一调制过程产生的解相关的且经调制的输出声道的示例。

[0208] 根据一些这样的示例,第一解相关过程涉及第一解相关函数,第二解相关过程涉及第二解相关函数,其中第二解相关函数是具有大约90度或大约-90度相移的第一解相关函数。在一些这样的实现中,第一调制过程涉及第一调制函数,第二调制过程涉及第二调制函数,其中第二调制函数是具有大约90度或大约-90度相移的第一调制函数。

[0209] 在一些示例中,解相关、调制和组合产生输出音频信号,使得当输出音频信号被解码并且被提供给扬声器的阵列时,扬声器的阵列中的能量的空间分布与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的能量的空间分布基本上相同。而且,在一些这样的实现中,扬声器的阵列中的相邻扩音器之间的相关性与将由输入音频信号经由最小二乘解码器被解码到扬声器的阵列而得到的相关性基本上不同。

[0210] 一些实现(诸如以上参照图11描述的那些)可以涉及实现用于渲染具有大小的对象的格式转换器。一些这样的实现可以涉及:接收音频对象大小的指示;确定音频对象大小大于或等于阈值大小并且将零增益值应用于该组两个或更多个输入音频声道。以上参照图11的大小过程[14]描述了一个示例。在该示例中,如果 $size_1$ 参数是1/2或更大,则 $Gain_{DirectGain}=0$ 。因此,在该示例中,直接增益缩放器[15]将零增益应用于输入声道 $z_{1-9}(t)$ 。

[0211] 一些示例(诸如以上参照图12描述的那些)可以涉及在上混器中实现格式转换器。一些这样的实现可以涉及接收来自音频引导逻辑过程的输出,该输出包括被引导音频数据的 N_p 个音频声道,其中,一个或多个声道的增益已经基于当前的主导声音方向而被改变。一些示例可以涉及将被引导音频数据的 N_p 个音频声道与输出音频信号的 N_p 个音频声道组合。

[0212] 格式转换器的其他使用

[0213] 对于本公开中所描述的实现的修改对于本领域的普通技术人员可以是容易清楚的。本文中所定义的一般原理可以应用于其他实现而不违背本公开的精神或范围。例如,将意识到,存在本文件中所描述的格式转换器将有益的许多其他应用。因此,权利要求并非意图限于本文中所示的实现,而是要被赋予与本文中所公开的本公开、原理和新颖特

征一致的最宽的范围。

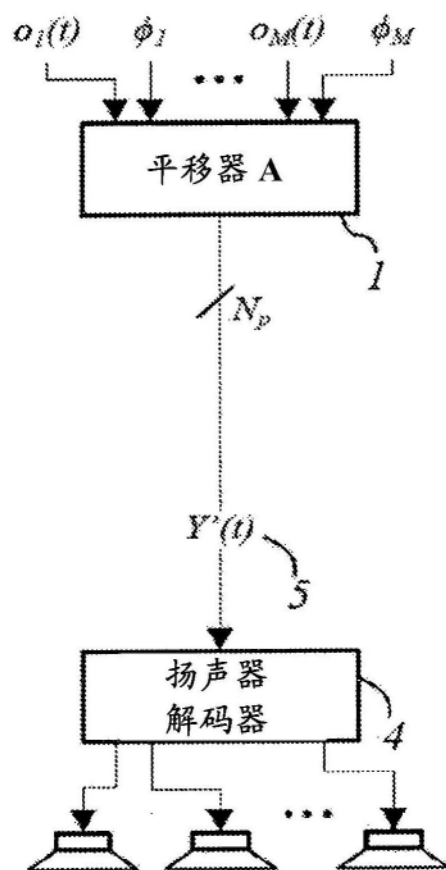


图1A

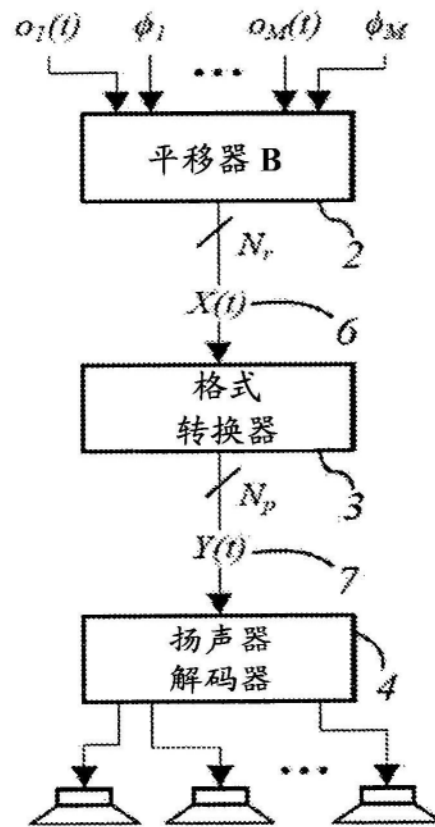


图1B

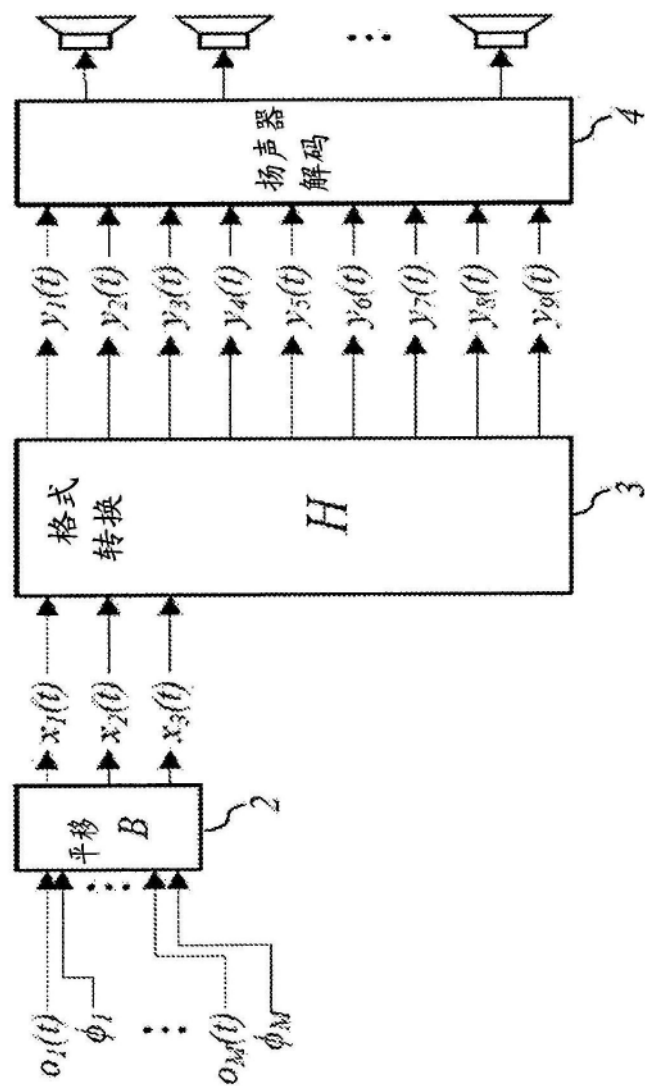


图2

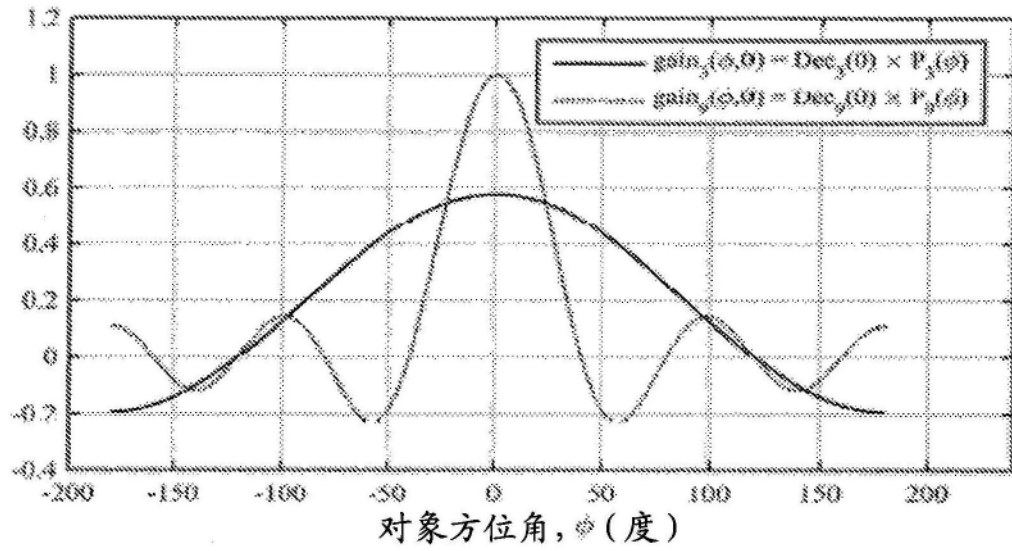


图3

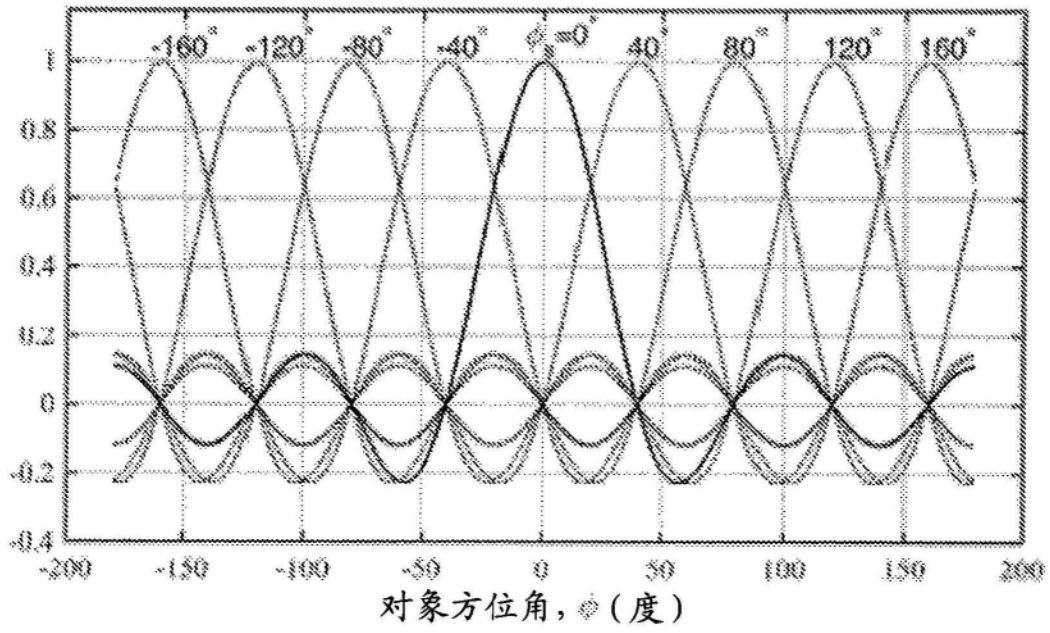


图4

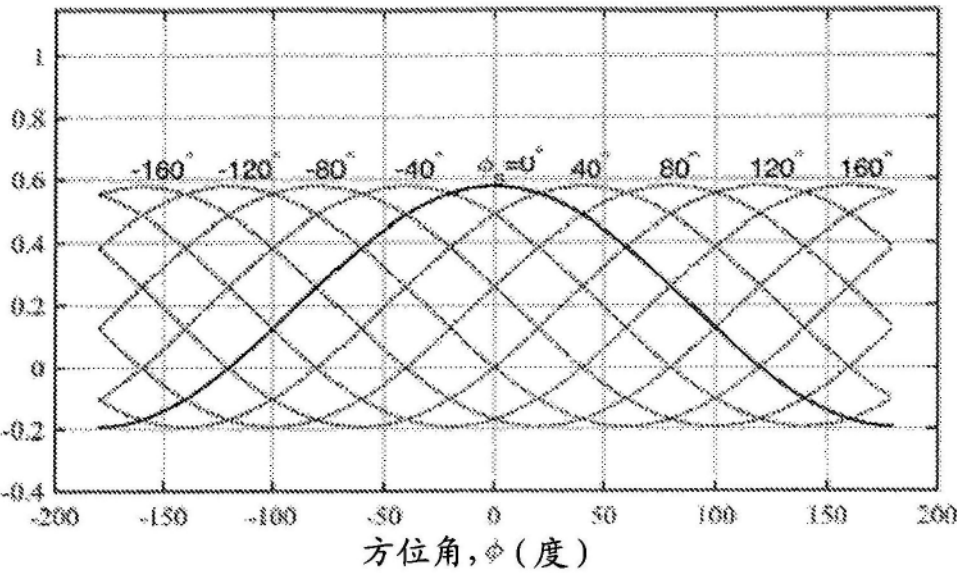


图5

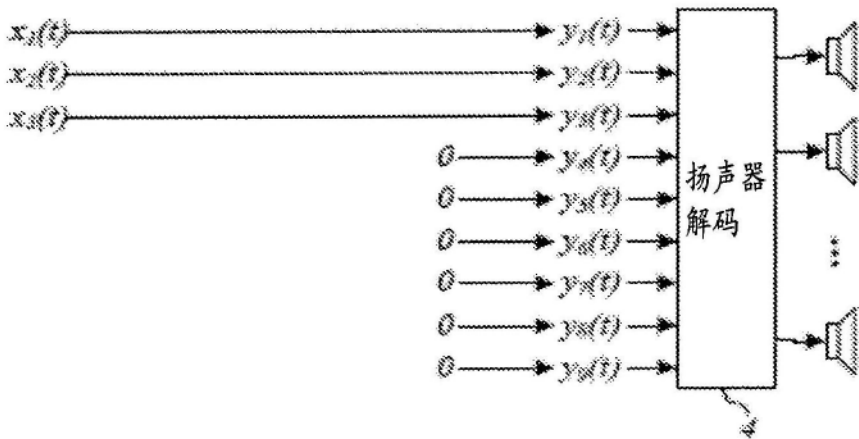


图6

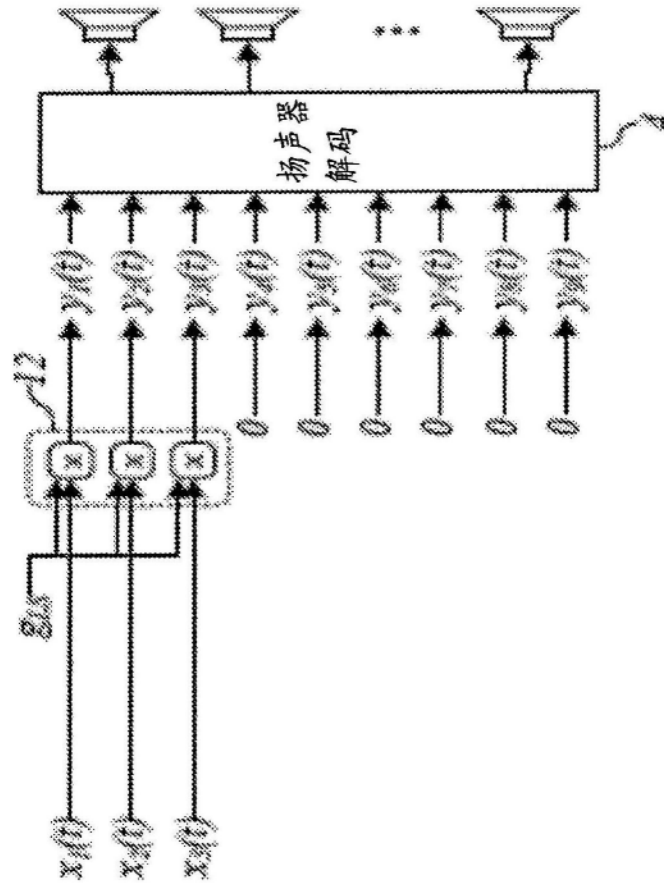


图7

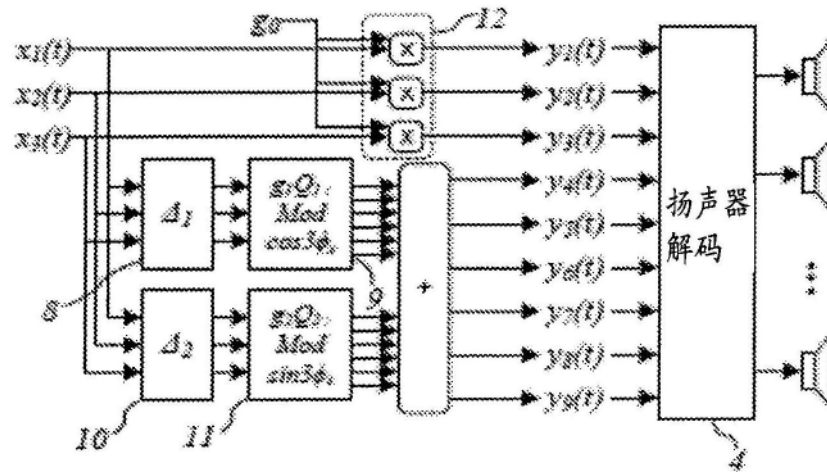


图8

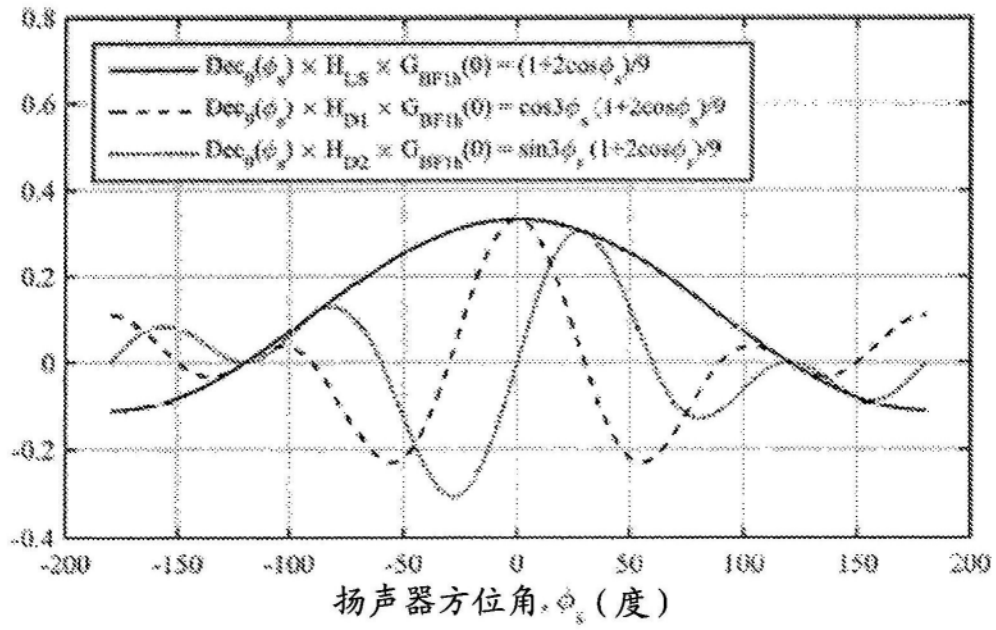


图9

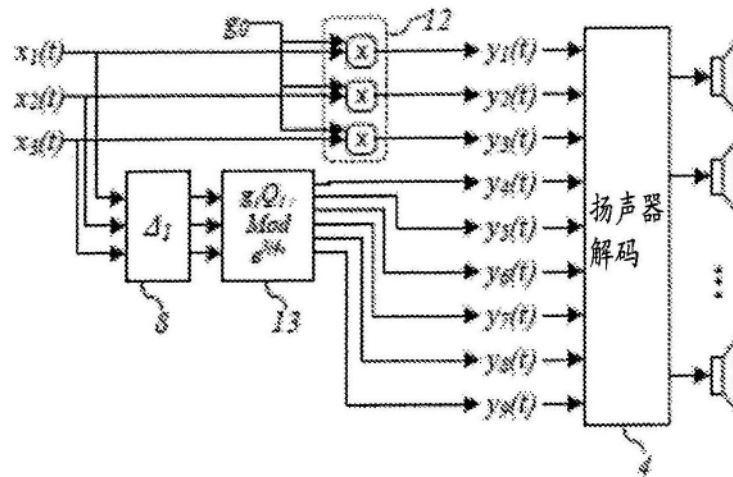


图10

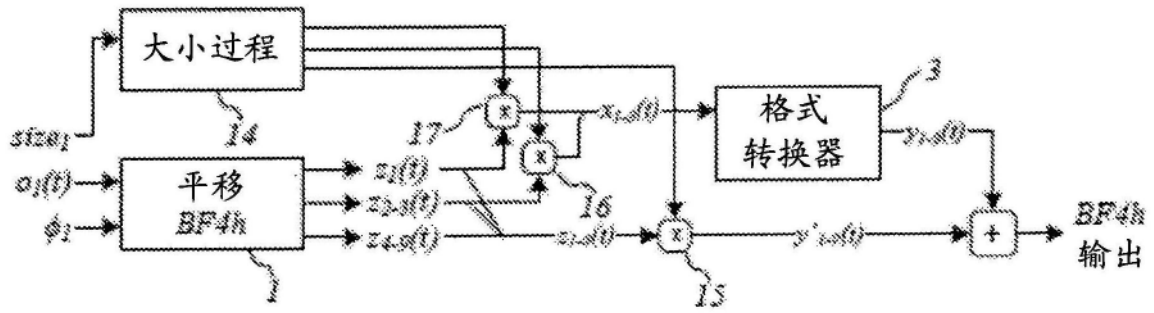


图11

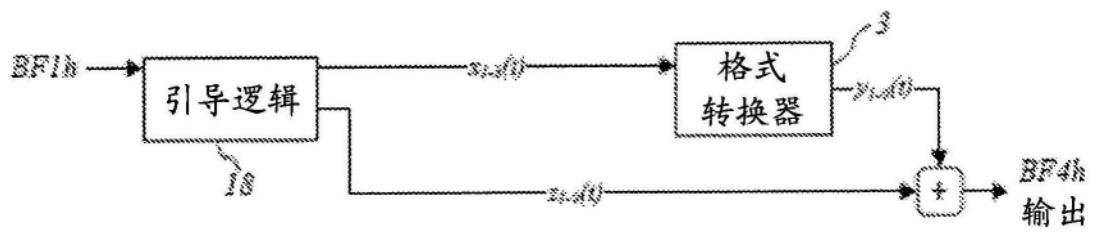


图12

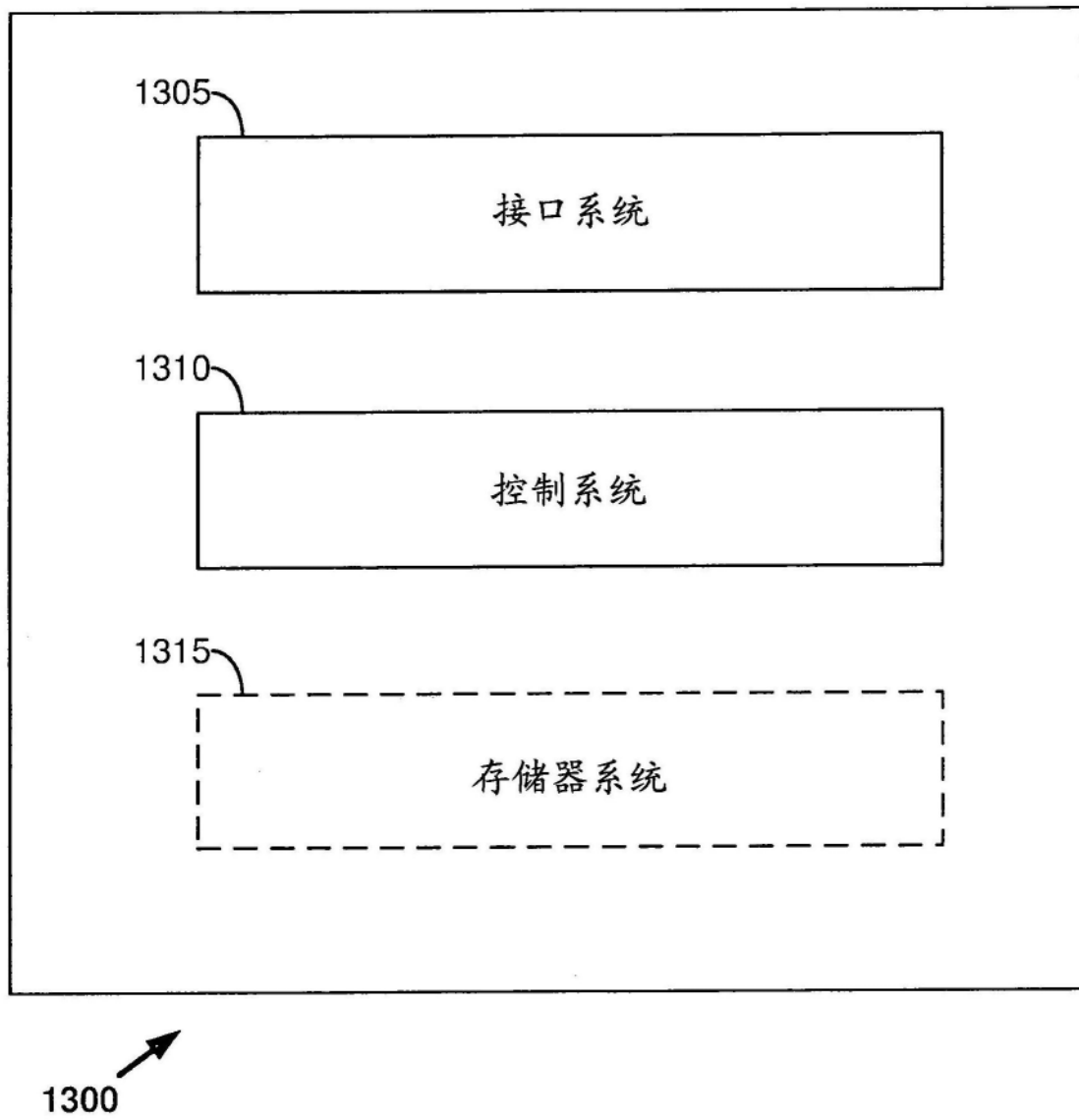


图13

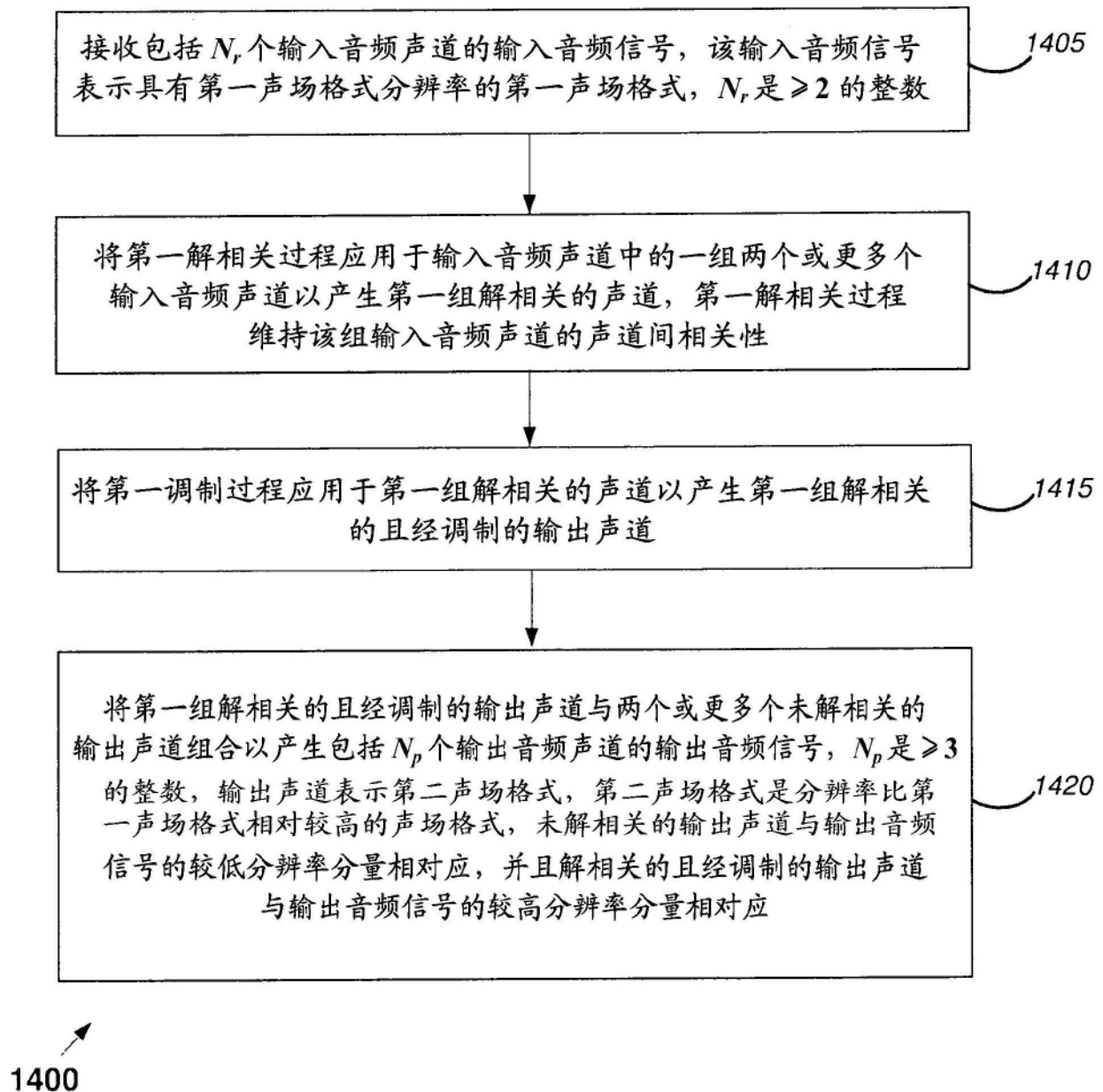


图14