



(12) 发明专利

(10) 授权公告号 CN 103635964 B

(45) 授权公告日 2016. 05. 04

(21) 申请号 201280032460. 1

(22) 申请日 2012. 06. 15

(30) 优先权数据

11305845. 7 2011. 06. 30 EP

(85) PCT国际申请进入国家阶段日

2013. 12. 30

(86) PCT国际申请的申请数据

PCT/EP2012/061477 2012. 06. 15

(87) PCT国际申请的公布数据

W02013/000740 EN 2013. 01. 03

(73) 专利权人 汤姆逊许可公司

地址 法国伊西莱穆利诺

(72) 发明人 P. 贾克斯 J-M. 巴特克

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 吕晓章

(51) Int. Cl.

G10L 21/00(2013. 01)

H04S 3/00(2006. 01)

(56) 对比文件

CN 1589127 A, 2005. 03. 02,

CN 1527284 A, 2004. 09. 08,

WO 2008146466 A1, 2008. 12. 04,

WO 2006006809 A1, 2006. 01. 19,

US 6201175 B1, 2001. 03. 13,

审查员 张鑫

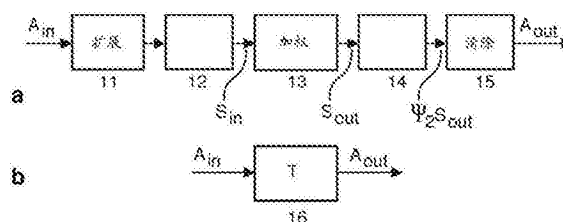
权利要求书2页 说明书11页 附图3页

(54) 发明名称

改变包含在高阶高保真度立体声响复制表示
中声音对象相对位置的方法以及装置

(57) 摘要

高阶高保真度立体声响复制 HOA 是有助于以优越的空间分辨率捕捉、处理、记录、传输和回放复杂的音频场景 2D 和 3D 的空间声场表示。在空间中的参考点处或附近的声场用傅里叶 - 贝塞尔级数近似。本发明使用空间翘曲 (12, 13, 14 ; 16) 来修改捕获的或生成为较高阶的高保真度立体声响复制表示的声场信息的空间内容和 / 或再现。不同的翘曲特性对 2D 和 3D 声场是可行的。翘曲在不执行场景分析或分解的情况下在空间域中进行。给定阶数的输入 HOA 系数被解码成常规放置的(虚拟)扬声器的权重或输入信号。



1. 一种用于改变包含在音频场景的二维或三维高阶高保真度立体声响复制HOA表示中的声音对象的相对位置的方法,其中维度为 O_{in} 的输入向量 A_{in} 确定输入信号傅里叶级数的系数,维度为 O_{out} 的输出向量 A_{out} 确定对应改变后的输出信号傅里叶级数的系数,所述方法包括下列步骤:

-使用模态矩阵 Ψ_1 的逆 Ψ_1^{-1} 通过计算 $s_{in}=\Psi_1^{-1}A_{in}$,将输入HOA系数的所述输入向量 A_{in} 解码(12)为空间域中与常规放置的扬声器位置对应的输入信号 s_{in} ;

-通过计算 $A_{out}=\Psi_2s_{in}$ 在空间域中将所述输入信号 s_{in} 翘曲并编码(14)为适应输出HOA系数的所述输出向量 A_{out} ,其中根据翘曲函数 $f(\phi)$ 修改模态矩阵 Ψ_2 的模态向量,原始扬声器位置的角度 (ϕ_{in},θ_{in}) 通过翘曲函数 $f(\phi)$ 一对一地映射到所述输出向量 A_{out} 中目标扬声器位置的目标角度 $(\phi_{out},\theta_{out})$ 。

2. 如权利要求1所述的方法,其中在所述翘曲和编码(14)之前使用增益函数 $g(\phi)$ 或 $g(\theta,\phi)$ 加权(13)所述空间域输入信号 s_{in} 。

3. 如权利要求2所述的方法,其中对于二维高保真度立体声响复制,所述增益函数是

$g(\phi)=\frac{df_{\phi}(\phi)}{d\phi}$,以及对于三维高保真度立体声响复制,所述增益函数是 ϕ 方向和 θ 方向上的

$g(\theta,\phi)=\frac{df_{\theta}(\theta)}{d\theta}\cdot\frac{\arccos((\cos f_{\theta}(\theta_{in}))^2+(\sin f_{\theta}(\theta_{in}))^2\cos\phi_{\epsilon})}{\arccos((\cos\theta_{in})^2+(\sin\theta_{in})^2\cos\phi_{\epsilon})}$,其中 ϕ 是方位角, θ 是倾角, ϕ_{ϵ} 是

较小的方位角。

4. 如权利要求1或2所述的方法,其中在虚拟扬声器的数量或维度 O_{warp} 等于或大于HOA系数的个数或维度 O_{in} 时,在所述解码(12)之前,通过添加(11)零系数来扩展(11)所述输入向量 A_{in} 的阶数或维度来得到更高的阶数。

5. 如权利要求2所述的方法,其中在HOA系数的阶数或维度低于所述模态矩阵 Ψ_2 的阶数或维度时,使用包括用于最高阶的零系数的窗口向量 w 进一步加权(15)所述翘曲处理过并编码以及可能加权(13)的信号 Ψ_2s_{in} ,以清除(15)翘曲处理过的系数的一部分以提供所述输出向量 A_{out} 。

6. 如权利要求2或5所述的方法,其中所述解码(12)、加权(13)和翘曲/解码(14)通常通过使用大小为 $O_{warp}\times O_{warp}$ 的变换矩阵 $T=\text{diag}(w)\Psi_2\text{diag}(g)\Psi_1^{-1}$ 执行,其中 $\text{diag}(w)$ 表示对角矩阵,所述窗口向量 w 的值作为其主对角线分量, $\text{diag}(g)$ 表示对角矩阵,所述增益函数 g 的值作为其主对角线分量。

7. 如权利要求6所述的方法,其中为了形成所述变换矩阵 T ,获得大小 $O_{out}\times O_{in}$,除去所述变换矩阵 T 的对应列和/或行,以执行空间翘曲操作 $A_{out}=TA_{in}$ 。

8. 一种用于改变包含在音频场景的二维或三维高阶高保真度立体声响复制HOA表示中的声音对象的相对位置的装置,其中维度为 O_{in} 的输入向量 A_{in} 确定输入信号傅里叶级数的系数,维度为 O_{out} 的输出向量 A_{out} 确定对应改变后的输出信号傅里叶级数的系数,所述装置包括:

-适用于使用模态矩阵 Ψ_1 的逆 Ψ_1^{-1} 通过计算 $s_{in}=\Psi_1^{-1}A_{in}$ 将输入HOA系数的所述输入向量 A_{in} 解码为空间域中与常规放置的扬声器位置对应的输入信号 s_{in} 的部件(12);

-适用于通过计算 $A_{out} = \Psi_2 S_{in}$ 在空间域中将所述输入信号 S_{in} 翘曲并编码为适应输出HOA系数的所述输出向量 A_{out} 的部件(14),其中根据翘曲函数 $f(\phi)$ 修改模态矩阵 Ψ_2 的模态向量,原始扬声器位置的角度 (ϕ_{in}, θ_{in}) 通过翘曲函数 $f(\phi)$ 一对一地映射到所述输出向量 A_{out} 中目标扬声器位置的目标角度 $(\phi_{out}, \theta_{out})$ 。

9.如权利要求8所述的装置,包括适用于在所述翘曲和编码(14)之前使用增益函数 $g(\phi)$ 或 $g(\theta, \phi)$ 加权所述空间域输入信号 S_{in} 的部件(13)。

10.如权利要求9所述的装置,其中对于二维高保真度立体声响复制,所述增益函数是 $g(\phi) = \frac{df_{\phi}(\phi)}{d\phi}$,以及对于三维高保真度立体声响复制,所述增益函数是 ϕ 方向和 θ 方向上的

$g(\theta, \phi) = \frac{df_{\theta}(\theta)}{d\theta} \cdot \frac{\arccos((\cos f_{\theta}(\theta_{in}))^2 + (\sin f_{\theta}(\theta_{in}))^2 \cos \phi_{\epsilon})}{\arccos((\cos \theta_{in})^2 + (\sin \theta_{in})^2 \cos \phi_{\epsilon})}$,其中 ϕ 是方位角, θ 是倾角, ϕ_{ϵ} 是较小的方位角。

11.如权利要求8或9所述的装置,包括适用于在虚拟扬声器的数量或维度 0_{warp} 等于或大于HOA系数的个数或维度 0_{in} 时,在所述解码(12)之前,通过添加零系数来扩展所述输入向量 A_{in} 的阶数或维度来得到更高的阶数的部件(11)。

12.如权利要求9所述的装置,包括适用于使用包括用于最高阶的零系数的窗口向量 w 进一步加权所述翘曲处理过并编码以及可能加权的信号 $\Psi_2 S_{in}$,并适用于清除翘曲处理过的系数的一部分以提供所述输出向量 A_{out} 的部件(15)。

13.如权利要求9或12所述的装置,包括适用于通常通过使用大小为 $0_{warp} \times 0_{warp}$ 的变换矩阵 $T = \text{diag}(w) \Psi_2 \text{diag}(g) \Psi_1^{-1}$ 执行所述解码、加权和翘曲/解码的部件(12,13,14,15),其中 $\text{diag}(w)$ 表示对角矩阵,所述窗口向量 w 的值作为其主对角线分量, $\text{diag}(g)$ 表示对角矩阵,所述增益函数 g 的值作为其主对角线分量。

14.如权利要求13所述的装置,其中为了形成所述变换矩阵 T ,获得大小 $0_{out} \times 0_{in}$,在所述适用于通常执行所述解码、加权和翘曲/解码的部件(12,13,14,15)中,除去所述变换矩阵 T 的对应列和/或行,以执行空间翘曲操作 $A_{out} = T A_{in}$ 。

改变包含在高阶高保真度立体声响复制表示中声音对象相对位置的方法以及装置

技术领域

[0001] 本发明涉及一种用于改变包含在音频场景的二维或三维高保真度立体声响复制(Ambisonics)表示中的声音对象的相对位置的方法和装置。

背景技术

[0002] 高阶高保真度立体声响复制(HOA)是有助于以优越的空间分辨率捕捉、处理、记录、传输和回放复杂的音频场景(2D和3D)的空间声场表示。声场是空间中参考点处或周围的傅立叶-贝塞尔级数近似。

[0003] 目前只有有限数量的用于处理使用HOA技术捕获的音频场景的空间排列的技术。在原理上,有两种方法:

[0004] A)将音频场景分解为分离的声音对象和相关联的位置信息,例如:通过狄拉克(DirAC),并使用处理过的位置参数构成新的场景。缺点是复杂且容易出错的场景分解是必需的。

[0005] B)HOA表示的内容可以通过HOA向量的线性变换来进行修改。这里,目前仅提出旋转、镜像以及强调前/后方向。所有这些已知的基于变换的修改技术都保持场景内对象相对位置的固定。

[0006] 为了处理或修改场景的内容,提出了空间翘曲(space warping),包括HOA声场的旋转和镜像,以及修改特定方向上的优势度(dominance):

[0007] G.J.Barton、M.A.Gerzon《Ambisonic Decoders for HDTV》,AES大会,1992年;

[0008] J.Daniel,《Representation de champs acoustiques,application a la transmission et a la reproduction de scenes sonores complexes dans un contexte multimedia》,博士论文,巴黎第六大学,2001年,巴黎,法国;

[0009] M.Chapman、Ph.Cotterell,《Towards a Comprehensive Account of Valid Ambisonic Transformations》,高保真度立体声响复制研讨会,2009年,格拉茨,奥地利。

发明内容

[0010] 本发明要解决的问题是帮助改变包含在基于HOA的音频场景中声音对象的相对位置,而不需要分析场景的构成。该问题使用权利要求1公开的方法解决。权利要求2中公开利用该方法的装置。

[0011] 本发明使用空间翘曲来修改捕获的或生成为高阶高保真度立体声响复制表示的声场信息的空间内容和/或再现。HOA域中的空间翘曲表示多步骤的方法或计算效率更高的单步骤线性矩阵乘法。不同的翘曲特性对2D和3D声场是可行的。

[0012] 翘曲在不执行场景分析或分解的情况下在空间域中进行。给定阶数的输入HOA系数被解码成常规放置的(虚拟)扬声器的权重或输入信号。

[0013] 本发明的空间翘曲处理具有以下几个优点:

[0014] -因为参数化过程中的多个自由度,因此它非常灵活;

[0015] -可以以非常高效的方式(即,以相对较低的复杂度)实现;

[0016] -不需要任何场景分析或分解。

[0017] 在原理上,本发明的方法适用于改变包含在音频场景的二维或三维高阶高保真度立体声响复制HOA表示中的声音对象的相对位置,其中维度为 O_{in} 的输入向量 A_{in} 确定输入信号傅里叶级数的系数,维度为 O_{out} 的输出向量 A_{out} 确定对应改变后的输出信号傅里叶级数的系数,所述方法包括下列步骤:

[0018] -使用模态矩阵 Ψ_1 的逆 Ψ_1^{-1} 通过计算 $s_{in} = \Psi_1^{-1} A_{in}$,将输入HOA系数的所述输入向量 A_{in} 解码为空间域中与常规放置的扬声器位置对应的输入信号 s_{in} ;

[0019] -通过计算 $A_{out} = \Psi_2 s_{in}$ 在空间域中将所述输入信号 s_{in} 翘曲并编码为适应输出HOA系数的所述输出向量 A_{out} ,其中根据翘曲函数 $f(\phi)$ 修改模态矩阵 Ψ_2 的模态向量,原始扬声器位置的角度通过翘曲函数 $f(\phi)$ 一对一地映射到所述输出向量 A_{out} 中目标扬声器位置的目标角度。

[0020] 在原理上,本发明的装置适用于改变包含在音频场景的二维或三维高阶高保真度立体声响复制HOA表示中的声音对象的相对位置,其中维度为 O_{in} 的输入向量 A_{in} 确定输入信号傅里叶级数的系数,维度为 O_{out} 的输出向量 A_{out} 确定对应改变后的输出信号傅里叶级数的系数,所述装置包括:

[0021] -适用于使用模态矩阵 Ψ_1 的逆 Ψ_1^{-1} 通过计算 $s_{in} = \Psi_1^{-1} A_{in}$ 将输入HOA系数的所述输入向量 A_{in} 解码为空间域中与常规放置的扬声器位置对应的输入信号 s_{in} 的部件;

[0022] -适用于通过计算 $A_{out} = \Psi_2 s_{in}$ 在空间域中将所述输入信号 s_{in} 翘曲并编码为适应输出HOA系数的所述输出向量 A_{out} 的部件,其中根据翘曲函数 $f(\phi)$ 修改模态矩阵 Ψ_2 的模态向量,原始扬声器位置的角度通过翘曲函数 $f(\phi)$ 一对一地映射到所述输出向量 A_{out} 中目标扬声器位置的目标角度。

[0023] 本发明更多的有利实施例在相应的从属权利要求中公开。

附图说明

[0024] 参照附图描述本发明的示例性实施例,其中:

[0025] 图1示出了空间域中的翘曲的原理;

[0026] 图2示出了 $N_{in} = 3, N_{out} = 12$ 以及翘曲函数 $f(\phi) = \phi + 2a \tan(\frac{a \sin \phi}{1 - a \cos \phi})$, $a = -0.4$ 的空间翘曲的示例;

[0027] 图3示出了不同的翘曲函数和“内部”阶数 N_{warp} 的矩阵失真。

具体实施方式

[0028] 在下面,为便于理解,应用本发明的空间翘曲在二维设置的情形中描述,HOA表示依赖于圆形谐波,并假设表示的声场仅包括平面声波。然后,描述扩展到基于球面谐波的三维情形。

[0029] 符号系统

[0030] 在高保真度立体声响复制理论中,在空间中的特定点处或附近的声场用截断的傅里叶-贝塞尔级数描述。一般情况下,假设参考点在所选坐标系的原点处。

[0031] 对于使用球面坐标的三维应用,所有定义的索引 $n=0,1,\dots,N$ 和 $m=-n,\dots,n$ 的系数 A_n^m 的傅立叶级数描述声场在方位角 ϕ ,倾角 θ 和距离原点 r 处的压力:

$$p(r, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n C_n^m j_n(kr) Y_n^m(\theta, \phi) \quad (1)$$

[0033] 其中 k 是波数, $j_n(kr) Y_n^m(\theta, \phi)$ 是与 θ 和 ϕ 定义的方向的球面谐波严格相关的傅里叶-贝塞尔级数的核函数。为方便起见,下面HOA系数 A_n^m 在 $A_n^m = C_n^m j_n(kr)$ 的定义下使用。对于特定的阶数 N ,傅里叶-贝塞尔级数中系数的个数为 $O=(N+1)^2$ 。

[0034] 对于使用圆形坐标的二维应用,核函数仅取决于方位角 ϕ 。所有 $m \neq n$ 的系数的值都为零,并可以忽略。因此,HOA系数的个数减少到仅 $O=2N+1$ 个。并且,倾角 $\theta=\pi/2$ 是固定的。值得注意的是,对于2D情形,以及对于声音对象在圆形上完全常规分布的情形,即 $\phi_i = i \frac{2\pi}{O}$, Ψ 中的模态向量与已知的离散傅里叶变换DFT的核函数相同。

[0035] 对于定义也导致高保真度立体声响复制系数 A_n^m 的不同定义的核函数,存在不同的约定。然而,精确定义对在此应用中描述的空间翘曲技术的基本规定和特性不起作用。

[0036] HOA“信号”包括每一时刻的高保真度立体声响复制系数的向量 A 。对于二维,即圆形场景,系数向量的典型构成和阶数是

$$A_{2D} = (A_N^{-N}, A_{N-1}^{-N+1}, \dots, A_1^{-1}, A_0^0, A_1^1, \dots, A_N^N)^T \quad (2)$$

[0038] 对于三维的球形场景,系数的常见阶数是不同的:

$$A_{3D} = (A_0^0, A_1^{-1}, A_0^1, A_1^1, A_2^{-2}, \dots, A_N^N)^T \quad (3)$$

[0040] HOA表示的编码以线性方式进行,因此可以将多个、不同的声音对象的HOA系数相加以推导得到所获得声场的HOA系数。

[0041] 简单编码(Plain encoding)

[0042] 对来自几个方向的多个声音对象的简单编码可以在向量代数中直接完成。“编码”的意思是从单个声音对象($i=0 \dots M-1$)在时刻 l 的压力分布 $s_i(k, l)$ 上的信息推导出在同一时刻 l 和波数 k 的HOA系数向量 $A(k, l)$,加上方向 ϕ_i 和 θ_i ,声波从该方向 ϕ_i 和 θ_i 到达坐标系的原点

$$A(k, l) = \Psi \cdot s(k, l) \quad (4)$$

[0044] 如果假设是二维设置,并且HOA向量按照等式(2)定义构成,那么模态矩阵 Ψ 由模态向量 $\mathbf{Y}(\phi) = (Y_N^{-N}, \dots, Y_0^0, \dots, Y_N^N)^T$ 构建。 Ψ 的第 i 列包含根据第 i 个声音对象的方向 ϕ_i 的模态向量

$$\Psi = (Y(\phi_0), Y(\phi_1), \dots, Y(\phi_{M-1})) \quad (5)$$

[0046] 如上定义,由于输入信号(声音对象)是空间分布的,因此对HOA表示的编码可以解读为空间-频率变换。只有当声音对象的数量与HOA系数的数量相同时,即如果 $M=O$,并且如果方向 ϕ_i 合理地散布在单位圆周围,该使用矩阵 Ψ 的变换才可以不造成信息损失地可逆。用数学语言描述,可逆条件是模态矩阵 Ψ 必须是方阵($O \times O$)并且是可逆的。

[0047] 普通解码

[0048] 通过解码,推导得到必须被施加以准确回放输入HOA系数所描述的理想声场的真实或虚拟扬声器的驱动信号。这样的解码依赖于扬声器的数量M和位置。以下三种重要的情形必须区分(注意:这些情形是在它们通过“扬声器的数量”定义的意义简化的情形,假设这些情形都以几何合理的方式设置。更准确地说,应该通过目标扬声器设置的模态矩阵的秩来进行定义)。在下面示出的示例性解码规则中,应用模态匹配解码原则,但也可以利用其他的可以导致三个场景不同的解码规则的解码原则。

[0049] • 超定情形:扬声器的数量大于HOA系数的个数,即 $M > 0$ 。在此情形中,解码问题不存在唯一解,但存在位于所有可能的解的M维空间的M-0维子空间中的一系列允许解。通常地,使用特定扬声器设置的模态矩阵 Ψ 的伪逆矩阵,以确定扬声器信号 $s, s = \Psi^T(\Psi\Psi^T)^{-1}A$ 。(6)

[0050] 这个解以最小的总回放能量 $s^T s$ 传递扬声器信号(参见例如L.L.Scharf的《Statistical Signal Processing.Detection,Estimation,and Time Series Analysis》,Addison-Wesley出版公司,读物,马萨诸塞州,1990年)。对于扬声器的常规设置(这在2D情形中很容易实现),矩阵运算 $(\Psi\Psi^T)^{-1}$ 得到单位矩阵,根据等式(6)的解码规则简化为 $s = \Psi^T A$ 。

[0051] • 恰定情形:扬声器的数量等于HOA系数的个数。解码问题恰好存在一个唯一解,通过模态矩阵 Ψ 的逆 Ψ^{-1} 来定义: $s = \Psi^{-1}A$ 。(7)

[0052] • 欠定情形:扬声器的数量M小于HOA系数的个数0。因此,声场解码的数学问题是欠定的,不存在唯一的精确解。取而代之的是,必须使用数值优化来确定可以与理想声场最佳匹配的扬声器信号。

[0053] 为了得到稳定的解,可以应用正则化,例如通过公式

$$s = \Psi^T(\Psi\Psi^T + \lambda I)^{-1}A \quad (8)$$

[0055] 其中I表示单位矩阵,缩放因子 λ 定义正则化量。作为示例, λ 可以设定为 $\Psi\Psi^T$ 特征值的平均值。

[0056] 得到的音束图可能是次优的,因为一般情况下使用这种方法得到的音束图的方向过于集中,很多声音信息不能被充分表示。

[0057] 对上面描述的所有的解码器示例,都做出扬声器发出平面波的假设。真实世界的扬声器具有不同的回放特性,解码规则应该注意这些特征。

[0058] 基本翘曲

[0059] 图1a示出了本发明的空间翘曲的原理。翘曲在空间域中进行。因此,首先,在步骤/阶段12将阶数为 N_{in} ,维度为 O_{in} 的输入HOA系数 A_{in} 解码为与常规放置的(虚拟)扬声器对应的权重或输入信号 s_{in} 。对于此解码步骤,有利地应用确定的解码器,即用于虚拟扬声器的数量 O_{warp} 等于或大于HOA系数的个数 O_{in} 的解码器。对后一种情形(扬声器数量比HOA系数个数更多),可以通过在步骤/阶段11中添加零系数容易地扩展HOA系数向量 A_{in} 的阶数或维度来得到更高的阶数。目标向量 s_{in} 的维度将在下面用 O_{warp} 表示。

$$s_{in} = \Psi_1^{-1} A_{in} \quad (9)$$

[0061] 扬声器信号的虚拟位置应该是常规的,例如,对二维情形 $\phi_i = i \cdot 2\pi/O_{warp}$ 。因此保

证很好地调整模态矩阵 Ψ_1 以确定解码矩阵 Ψ_1^{-1} 。

[0062] 下面,根据期望的翘曲特性在“翘曲”处理中对虚拟扬声器的位置进行修改。步骤/阶段14中的翘曲处理结合使用模态矩阵 Ψ_2 编码目标向量 S_{in} (或相应地 S_{out}), 得到维度为 O_{warp} 或在下面描述的进一步的处理步骤之后的维度为 O_{out} 的翘曲处理后的HOA系数向量 A_{out} 。在原理上说,翘曲特性可以通过将源角度一对一地映射到目标角度来完全定义,即对每一个源角度 $\phi_{in}=0 \dots 2\pi$, 以及可能的 $\theta_{in}=0 \dots 2\pi$, 定义目标角度, 因此对于2D情形

$$[0063] \quad \phi_{out} = f(\phi_{in}) \quad (10)$$

[0064] 以及对于3D情形

$$[0065] \quad \phi_{out} = f_{\phi}(\phi_{in}, \theta_{in}) \quad (11)$$

$$[0066] \quad \theta_{out} = f_{\theta}(\phi_{in}, \theta_{in}) \quad (12)$$

[0067] 为了便于理解,该(虚拟)重新定向可以比作将扬声器物理移动到新的位置。

[0068] 这个过程产生的一个问题是,以一定角度相邻的扬声器之间的距离会根据翘曲函数 $f(\phi)$ 的梯度改变(下面将描述2D的情形):如果 $f(\phi)$ 的梯度大于1,那么翘曲处理后的声场中的同一角度空间将被少于原始声场中的“扬声器”占据,反之亦然。换言之,扬声器的密度 D_s 的分布遵循

$$[0069] \quad D_s(\phi) = \frac{1}{\frac{df(\phi)}{d\phi}} \quad (13)$$

[0070] 反过来,这意味着空间翘曲对听者周围的声音平衡进行修改。扬声器密度增加(即, $D_s(\phi) > 1$) 的区域将变得更加优势,而 $D_s(\phi) < 1$ 的区域将变得不那么优势。

[0071] 作为一种选择,随着应用要求的不同而定,对扬声器密度的前述修改可以通过在加权步骤/阶段13中将增益函数 $g(\phi)$ 应用于虚拟扬声器输出信号 S_{out} 抵消得到信号 S_{out} 。在原理上说,可以指定任何加权函数 $g(\phi)$ 。已经通过经验确定的一个特定的有利变型是与翘曲函数 $f(\phi)$ 的导数成比例:

$$[0072] \quad g(\phi) = \frac{1}{D_s(\phi)} = \frac{df(\phi)}{d\phi} \quad (14)$$

[0073] 使用该特定的加权函数,在合适的较高的内部阶数和输出阶数的假设下(参见下面的小节如何设定HOA阶数),调节函数(panning function)在特定翘曲角度 $f(\phi)$ 的幅度保持等于原始调节函数在原始角度 ϕ 的幅度。由此获得每个辐射角度(opening angle)的同质声音平衡(幅度)。

[0074] 除了上述示例的加权函数,还可以使用其他的加权函数,例如以获得每个辐射角度的相等的能量。

[0075] 最后,在步骤/阶段14中,使用模态矩阵 Ψ_2 通过执行 $\Psi_2 S_{out}$ 对加权的虚拟扬声器信号再次进行翘曲和编码。根据翘曲函数 $f(\phi)$, Ψ_2 包括不同于 Ψ_1 的模态向量。结果是翘曲处理后的声场的 O_{warp} 维HOA表示。

[0076] 如果目标HOA表示的阶数或维度低于编码器 Ψ_2 的阶数(参见下面的小节如何设定HOA阶数),那么在步骤/阶段15中必须除去(清除(strip))翘曲处理过的系数中的一些(即,一部分)。通常情况下,该清除操作可以用窗口操作描述:编码向量 $\Psi_2 S_{out}$ 与包括应该除去

的用于最高阶数的零系数的窗口向量 w 相乘,该乘法可以视为表示进一步加权。在最简单的情形中,可以应用矩形窗口,但也可以应用如在M.A.Poletti的《A Unified Theory of Horizontal Holographic Sound Systems》(Journal of the Audio Engineering Society,48(12),第1155-1182页,2000年)的第3小节中描述的更复杂的窗口或在J.Daniel的上述博士论文的3.3.2小节中的“in-phase”或“max.re”窗口。

[0077] 3D翘曲函数

[0078] 上面已经描述了二维情形的翘曲函数 $f(\phi)$ 的概念和相关联的加权函数 $g(\phi)$ 。下面扩展到由于更高维度和必须应用球面几何而因此更复杂的三维情形。引入两个简化场景,这两者都允许通过一维翘曲函数 $f(\phi)$ 或 $g(\phi)$ 规定想要的空间翘曲。

[0079] 在沿经度的空间翘曲中,空间翘曲仅作为方位角 ϕ 的函数进行。此情形与上面介绍的二维情形非常类似。翘曲函数由以下完全定义:

$$[0080] \quad \theta_{\text{out}} = f_{\theta}(\theta_{\text{in}}, \phi_{\text{in}}) = \theta_{\text{in}} \quad (15)$$

$$[0081] \quad \phi_{\text{out}} = f_{\phi}(\theta_{\text{in}}, \phi_{\text{in}}) = f_{\phi}(\phi_{\text{in}}) \quad (16)$$

[0082] 由此可以应用与用于二维情形类似的翘曲函数。空间翘曲对赤道上的声音对象影响最大,而对球的两极的声音对象影响最小。

[0083] (翘曲后的)声音对象在球上的密度只与方位角有关。因此,恒定密度的加权函数为 $g(\theta) = \frac{df_{\phi}(\phi)}{d\phi}$ 。

$$(17)$$

[0084] 通过在应用翘曲之前(虚拟)旋转球面和之后反向旋转,在空间中自由定向特定的翘曲特性是可行的。

[0085] 在沿纬度的空间翘曲中,允许空间翘曲只沿子午线。翘曲函数由以下定义:

$$[0086] \quad \theta_{\text{out}} = f_{\theta}(\theta_{\text{in}}, \phi_{\text{in}}) = f_{\theta}(\theta_{\text{in}}) \quad (15)$$

$$[0087] \quad \phi_{\text{out}} = f_{\phi}(\theta_{\text{in}}, \phi_{\text{in}}) = \phi_{\text{in}} \quad (16)$$

[0088] 该翘曲函数在球面上的重要特性是,尽管方位角保持恒定,但方位角方向上两点的角距离可以由于修改倾角而很好地变化。原因是两条子午线之间的角距离在赤道处最大,但在两极处变为零。这个事实必须通过加权函数进行解释。

[0089] 两个点A和B之间的角距离 c 可以通过球面几何的余弦定理来确定,参见等式(3.188c)I.N.Bronstein、K.A.Semendjajew、G.Musiol、H.Milhlig的《Taschenbuch der Mathematik》(Verlag Harri Deutsch, Thun, Frankfurt/Main, 第5版, 2000年):

$$[0090] \quad \cos c = \cos \theta_A \cos \theta_B + \sin \theta_A \sin \theta_B \cos \phi_{AB} \quad (20)$$

[0091] 其中 ϕ_{AB} 表示两点A和B之间的方位角。考虑到相同倾角 θ 处的两点之间的角距离,该等式简化为:

$$[0092] \quad c = \arccos[(\cos \theta_A)^2 + (\sin \theta_A)^2 \cos \phi_{\epsilon}] \quad (21)$$

[0093] 可以应用该公式推导得到空间中某一点与相距较小的方位角 ϕ_{ϵ} 的另一点之间的角距离。“较小”的意思是在实际应用中尽量可行的小,但不是零,在理论上,极限值 $\phi_{\epsilon} \rightarrow 0$ 。这些角距离在翘曲之前和之后之间的比率给出声音对象的密度在 ϕ 方向上变化的因子:

$$[0094] \quad \frac{c_{\text{out}}}{c_{\text{in}}} = \frac{\arccos((\cos \theta_{\text{out}})^2 + (\sin \theta_{\text{out}})^2 \cos \phi_s)}{\arccos((\cos \theta_{\text{in}})^2 + (\sin \theta_{\text{in}})^2 \cos \phi_s)} \quad (22)$$

[0095] 最后,加权函数是 ϕ 方向和 θ 方向上两个加权函数的乘积:

$$[0096] \quad g(\theta, \phi) = \frac{df_{\theta}(\theta)}{d\theta} \cdot \frac{\arccos((\cos f_{\theta}(\theta_{\text{in}}))^2 + (\sin f_{\theta}(\theta_{\text{in}}))^2 \cos \phi_s)}{\arccos((\cos \theta_{\text{in}})^2 + (\sin \theta_{\text{in}})^2 \cos \phi_s)} \quad (23)$$

[0097] 并且,如在前面的场景中描述的,通过旋转在空间中自由定向特定的翘曲特性是可行的。

[0098] 单步骤处理

[0099] 结合图1a引入的步骤(即,扩展阶数、解码、加权、翘曲+编码和清除阶数)在本质上都是线性操作。因此,这一连续操作可以用如图1b所示的步骤/阶段16中输入HOA参数与单个矩阵相乘来替换。忽略扩展和清除操作,全 $0_{\text{warp}} \times 0_{\text{warp}}$ 变换矩阵 T 确定为

$$[0100] \quad \mathbf{T} = \text{diag}(\mathbf{w}) \Psi_2 \text{diag}(\mathbf{g}) \Psi_1^{-1} \quad (24)$$

[0101] 其中 $\text{diag}(\cdot)$ 表示对角矩阵,它的向量参数(vector argument)值是主对角线分量, g 是加权函数, w 是用于准备进行上述清除的窗口向量,即对于在步骤/阶段15中执行的准备清除和系数自身清除的加权的两个函数,等式(24)中的窗口向量 w 仅用于加权。

[0102] 多步骤方法中的两次修改阶数(即,在解码器之前扩展阶数和在编码之后清除HOA系数)也可以通过除去对应的列和/或行整合成变换矩阵 T 。由此得到是可以直接应用于输入HOA向量的大小为 $0_{\text{out}} \times 0_{\text{in}}$ 的矩阵。那么,空间翘曲操作变为 $A_{\text{out}} = T A_{\text{in}}$ 。 (25)

[0103] 有利地,由于变换矩阵 T 的维度从 $0_{\text{warp}} \times 0_{\text{warp}}$ 有效降低为 $0_{\text{out}} \times 0_{\text{in}}$,因此执行根据图1b的单步骤处理所需的计算复杂度显著低于图1a的多步骤方法所需的计算复杂度,尽管单步骤处理传递完全相同的结果。尤其地,它可以避免在使用其中间信号的较低阶数 N_{warp} 执行多步骤处理时可能出现的失真(详细信息请参见下面的部分如何设定HOA阶数)。

[0104] 技术现状:旋转和镜像

[0105] 声场的旋转和镜像可以视为是空间翘曲的“简单”子类。这些变换的特殊特性是不改变声音对象相对于彼此的相对位置。这意味着,已经在原始声音场景中放置在例如另一声音对象右边 30° 处的声音对象将在旋转后的声音场景中保持在同一声音对象右边的 30° 处。对于镜像,只有符号改变,而角距离保持不变。声场信息的旋转和镜像的算法和应用已经在例如上面提到的Barton/Gerzon和J.Daniel的文章中,以及在M.Noisternig、A.Sontacchi、Th.Musil、R.Holdrich的《A 3D Ambisonic Based Binaural Sound Reproduction System》(第24届AES多声道音频国际会议会议记录,班夫,加拿大,2003年),以及在H.Pomberger、F.Zotter的《An Ambisonics Format for Flexible Playback Layouts》(第一次高保真度立体声响复制研讨会,格拉茨,奥地利,2009年)中进行了探索和描述。

[0106] 这些方法都基于对旋转的解析表达式。例如,圆形声场(2D情形)旋转任意角度 α 可以通过乘以其中只有系数的子集是非零的翘曲矩阵 T_α 来实现:

$$[0107] \quad T_{\alpha}(\mu, \nu) = \begin{cases} \cos(-\alpha(\mu - (O+1))/2); & \nu = \mu \\ \sin(-\alpha(\mu - (O+1))/2); & \nu = N - \mu + 1 \\ 0; & \text{其他} \end{cases} \quad (26)$$

[0108] 如在此示例中所示,所有用于旋转和/或镜像操作的翘曲矩阵都具有只有相同阶数 n 的系数才相互影响的特殊特性。因此,这些翘曲矩阵都是非常稀疏填充的矩阵,输出的 N_{out} 可以等于输入阶数 N_{in} ,而不丢失任何空间信息。

[0109] 有很多有趣的应用,要求对声场信息进行旋转或镜像操作。一个示例是使用头部跟踪系统通过头戴式耳机回放声场。取代根据头部的(多个)旋转角度内插HRTF(头部相关传递函数),有利地对于实际回放,根据头部的位置预先旋转声场,并使用固定的HRTF。该过程已经在上面提到的Noisternig/Sontacchi/Musil/Holdrich的文章中有描述。

[0110] 另一示例已经在上面提到的Pomberger/Zotter的文章中在声场信息编码的情景中描述。HOA向量描述的空间区域可以被限制为圆形(2D情形)或球面的特定部分。由于这些限制,HOA向量的某些部分将会变为零。这篇文章中提出的想法正是利用这种冗余度减少性质(redundancy-reducing property),对声场信息进行混合阶数编码(mixed-order coding)。由于上述限制只能针对空间中非常特定的区域获得,因此一般都需要进行旋转操作将传输的部分信息移动到空间中所希望的区域。

[0111] 示例

[0112] 图2示出了二维(圆形)情形中空间翘曲的一个示例。翘曲函数已经选择为

$$f(\phi) = \phi + 2 \operatorname{atan}\left(\frac{a \sin \phi}{1 - a \cos \phi}\right), \text{ 其中 } a = -0.4, \quad (27) \text{ 它与具有单个实值参数的离散时间全}$$

通滤波器的相位响应相似,参见M.Kappelan的《Eigenschaften von Allpass-Ketten und ihre Anwendung bei der nicht-aquidistanten spektralen Analyse und Synthese》(博士论文,亚琛工业大学(RWTH),亚琛,德国,1998年)。

[0113] 图2a示出了翘曲函数。选择该特定的翘曲函数 $f(\phi)$ 是因为它保证是 2π 周期的翘曲函数,同时允许用单个参数 a 修改空间失真量。图2b示出了对应的加权函数 $g(\phi)$ 对该特定翘曲函数的确定性结果。

[0114] 图2c示出了 7×25 单步骤变换翘曲矩阵 T 。该矩阵各个系数的对数绝对值根据附带的灰度或阴影条用灰度或阴影类型表示。该示例矩阵针对输入HOA阶数 $N_{in} = 3$ 和输出阶数 $N_{out} = 12$ 设计。为了捕捉从低阶系数变换到高阶系数散布的大部分信息,要求较高的输出阶数。如果输出阶数进一步降低,那么翘曲操作的精度将会变差,因为全翘曲矩阵的非零系数将会被忽略(详细讨论参见下面的小节如何设定HOA阶数)。

[0115] 该特定翘曲矩阵的一个非常有用的特性是它的大部分都是零。这允许在执行该操作时节省很多计算能力,但单步骤变换矩阵的某些部分是零并不是一般规则。

[0116] 图2d和图2e示出了某些平面波产生的音束图示例的翘曲特性。这两张图都来自相同的七个 ϕ 位置 $0, 2/7\pi, 4/7\pi, 6/7\pi, 8/7\pi, 10/7\pi$ 和 $12/7\pi$ 的输入平面波,都具有相同的为1的幅度,并示出七个角度幅度分布,即下列超定、常规解码操作的结果向量 s

$$[0117] \quad s = \Psi^{-1}A \quad (28)$$

[0118] 其中HOA向量 A 是原始或翘曲后的一组平面波变量。圆形外面的数字代表角度 ϕ 。

虚拟扬声器的数量(例如,360)明显多于HOA参数的个数。来自前方方向的平面波的幅度分布或音束图位于 $\phi = 0$ 处。

[0119] 图2d示出了原始HOA表示的幅度分布。全部七个分布的形状都相似,并具有相同宽度的主瓣(main lobe)。如所预计的那样,主瓣的最大值都在原始的七个声音对象的角度 $\phi = (0, 2/7\pi, \dots)$ 处。主瓣具有对应于原始HOA向量的有限阶数 $N_{in} = 3$ 的宽度。

[0120] 图2e示出了相同声音对象的幅度分布,但是在执行翘曲操作之后。一般地说,这些对象已经朝0度的前方方向移动,并已经修改音束图:前方方向 $\phi = 0$ 周围的主瓣已经变得更窄,并更加集中,而180度附近的后方方向上的主瓣则变得明显更宽。在两侧,最大的影响在90度和270度处,音束图由于图2b中加权函数 $g(\phi)$ 对这些角度较大的梯度而变成非对称。翘曲后HOA向量的更高的阶数 $N_{out} = 12$ 使得音束图的这些明显的改变(变窄和形成新的形状)成为可能。理论上说,前方方向上主瓣的分辨率增加因子2.33,而后方方向上的分辨率降低因子 $1/2.33$ 。建立了局部阶数在整个空间上变化的混合阶数信号。可以假设为了以合理的精度表示翘曲后的HOA系数要求最小的输出阶数为 $2.33 \cdot N_{in} \approx 7$ 。在下面的如何设定HOA阶数的小节中,详细描述了有关内在、局部阶数的讨论。

[0121] 特性

[0122] 上面引入的翘曲步骤都是非常通用和灵活的。至少可以完成下列基本操作:沿任意轴和/或平面旋转和/或镜像,使用连续翘曲函数的空间失真,以及对特定方向加权(空间音束成形)。

[0123] 在下面的更小的小节中,对本发明的空间翘曲的若干特性进行了重点描述,这些详细描述提供了有关能够实现什么和不能实现什么的指导。另外,还描述了一些设计规则。从原理上说,可以以一定的自由度调节下列参数,以获得想要的翘曲特性:

- [0124] • 翘曲函数 $f(\theta, \phi)$;
- [0125] • 加权函数 $g(\theta, \phi)$;
- [0126] • 内部阶数 N_{warp} ;
- [0127] • 输出阶数 N_{out} ;
- [0128] • 使用向量 w 对输出系数进行的窗口操作。

[0129] 线性

[0130] 根据定义,多步骤处理中的基本变换步骤是线性的。在中间发生声音源非线性映射到新的位置对编码矩阵的定义有影响,但编码矩阵本身也是线性的。因此,组合空间翘曲操作和 T 的矩阵乘法也是线性的,即

[0131] $T A_1 + T A_2 = T(A_1 + A_2)$ (29) 该性质是重要的,因为它允许处理包括同时来自不同声音源的贡献的复杂的声场信息。

[0132] 空间不变性

[0133] 根据定义(除非翘曲函数完全以斜率1或-1线性变化),空间翘曲变换不是空间不变的。这意味着操作对原始位于半球上不同位置处的声音对象的表现是不同的。用数学语言表达,该性质是翘曲函数 $f(\phi)$ 非线性的结果,即对至少一些任意角度 $\alpha \in [0 \dots 2\pi]$, $f(\phi + \alpha) \neq f(\phi) + \alpha$ (30)

[0134] 可逆性

[0135] 通常情况下,变换矩阵 T 不能简单地通过数学反演来求逆。一个明显的原因是 T 一

般不是方阵。即使方阵空间翘曲矩阵也会不可逆,因为通常从低阶系数散布到高级系数的信息将会丢失(对比小节如何设定HOA阶数和小节示例中的示例),而在操作中丢失信息意味着该操作是不可逆的。

[0136] 因此,必须找到用于至少近似逆转空间翘曲操作的另一种方法。可以通过翘曲函数 $f(\cdot)$ 的逆函数 $f_{\text{rev}}(\cdot)$ 设计逆转翘曲变换 T_{rev} ,其中

$$[0137] \quad f_{\text{rev}}(f(\Phi)) = \Phi \quad (31)$$

[0138] 根据选择不同的HOA阶数,该处理近似逆转变换。

[0139] 如何设定HOA阶数

[0140] 在设计空间翘曲变换时要考虑的一个重要方面是HOA阶数。而通常情况下,输入向量 A_{in} 的阶数 N_{in} 通过外部限制预先确定,输出向量 A_{out} 的阶数 N_{out} 和实际非线性翘曲操作的“内部阶数” N_{warp} 可以或多或少任意分配。然而,必须小心选择阶数 N_{in} 和 N_{warp} ,这将在下面说明。

[0141] “内部”阶数 N_{warp} :

[0142] “内部”阶数 N_{warp} 限定上述多步骤的空间翘曲处理中的实际解码、翘曲和编码步骤的精度。通常情况下,阶数 N_{warp} 应该明显大于输入阶数 N_{in} 和输出阶数 N_{out} 。该要求的原因是若非如此,由于翘曲操作通常是非线性操作,会因此产生失真和瑕疵。

[0143] 为了解释该事实,图3示出了与图2示例所用的相同的翘曲函数的全翘曲矩阵的一个示例。图3a、3c和3e分别描述了翘曲函数 $f_1(\Phi)$, $f_2(\Phi)$ 和 $f_3(\Phi)$ 。图3b、3d和3f分别描述了翘曲矩阵 $T_1(\text{dB})$, $T_2(\text{dB})$ 和 $T_3(\text{dB})$ 。出于示例的原因,为了确定特定输入阶数 N_{in} 或输出阶数 N_{out} 的翘曲矩阵,没有对这些翘曲矩阵进行裁剪。作为替代,图3b、3d和3f中的虚线的中心框描述最终得到的即裁剪后的转换矩阵的目标大小 $N_{\text{out}} \times N_{\text{in}}$ 。使用这种方式,非线性失真对翘曲矩阵的影响是清晰可见的。在此示例中,目标阶数被任意设定为 $N_{\text{in}} = 30$ 和 $N_{\text{out}} = 100$ 。

[0144] 基本挑战可以在图3b中看出:显然,由于空间域中的非线性处理,翘曲矩阵中的系数散布在主对角线周围——距离矩阵中心越远就越大。在距离中心非常远的距离处,在此示例中,在大约 $|y| \geq 90$ 处, y 是垂直轴,系数散布到达全矩阵的边界,在边界处,它看起来像是要“弹回”。这造成特别类型的延伸到翘曲矩阵中较大部分的失真。在实验评估中,已经观察到这些失真会明显损害变换性能,只要失真产物位于矩阵的目标区域之内(在图中用虚线框标记)。

[0145] 对于图3b中的第一示例,所有一切都工作良好,因为处理的“内部”阶数被选择为明显高于输出阶数 $N_{\text{out}} = 100$ 的 $N_{\text{warp}} = 200$ 。失真的区域没有延伸到虚线框之中。

[0146] 另一个场景在图3d中示出。内部阶数被规定为等于输出阶数,即 $N_{\text{warp}} = N_{\text{out}} = 100$ 。该图表明失真的延伸随内部阶数线性伸缩。结果是变换输出的更高阶数的系数被失真产物污染。这种伸缩性质的优点是它似乎可以通过相应地增加内部阶数 N_{warp} 来避免这些类型的非线性失真。

[0147] 图3f示出了使用更进取的、具有更大系数 $a = 0.7$ 的翘曲函数的示例。由于使用了该更进取的翘曲函数,现在失真延伸进入目标矩阵区域,即使对于内部阶数 $N_{\text{warp}} = 200$ 。对此情形,如在前面的段落中推导的,内部阶数应该进一步增加以甚至更多预留空间(over-provisioning)。该翘曲函数的实验表明增加内部阶数到例如 $N = 400$ 可以除去这些非线性失真。

[0148] 总而言之,翘曲函数越进取,内部阶数 N_{warp} 应该越高。目前还不存在最小内部阶数的正式推导。然而,如果存疑,“内部”阶数预留空间是有益的,因为非线性效应随着全翘曲矩阵的大小线性伸缩。在原理上,“内部”阶数可以任意高。尤其地,如果要推导出单步骤变换矩阵,那么内部阶数对最终翘曲操作的复杂度不起任何作用。

[0149] 输出阶数 N_{out} :

[0150] 为了规定翘曲变换的输出阶数 N_{out} ,要考虑下面两个方面:

[0151] 通常情况下,输出阶数必须大于输入阶数 N_{in} ,以保留散布到不同阶数的系数的所有信息。实际需要的大小还取决于翘曲函数的特性。作为一个经验法则,翘曲函数 $f(\phi)$ 的“宽带”越小,需要的输出阶数越低。在某些情形中,翘曲函数可以看作是低通滤波器,以限制所需输出阶数 N_{out} 。

[0152] 图3b可以观察一个示例。对于此特定翘曲函数,输出阶数 $N_{\text{out}}=100$ (如虚线框所示)足够避免信息损失。如果输出阶数显著降低,例如降低到 $N_{\text{out}}=50$,那么变换矩阵的某些非零系数将排除在外,并可以预测对应的信息丢失。

[0153] 在一些情形中,输出HOA系数将被用于处理或仅能够处理有限阶数的设备。例如,目标可能是有限数量扬声器的扬声器设置。在这样的应用中,输出阶数应该根据目标系统的能力指定。如果 N_{out} 足够小,那么翘曲变换可以有效地减少空间信息。

[0154] 将内部阶数 N_{warp} 降低到输出阶数 N_{out} 可以仅通过丢弃较高阶数的系数来实现。这与应用矩形窗口于HOA输出向量对应。替代性地,可以应用更复杂的降低带宽技术,如在上面提到的M.A.Poletti的文章或在上面对应的J.Daniel的文章中讨论的那些技术。因此,即使可能比使用矩形窗口丢失更多的信息,但可以得到更优的方向图。

[0155] 本发明可以在音频处理链条的不同部分中使用,例如,记录、后期制作、传输、回放。

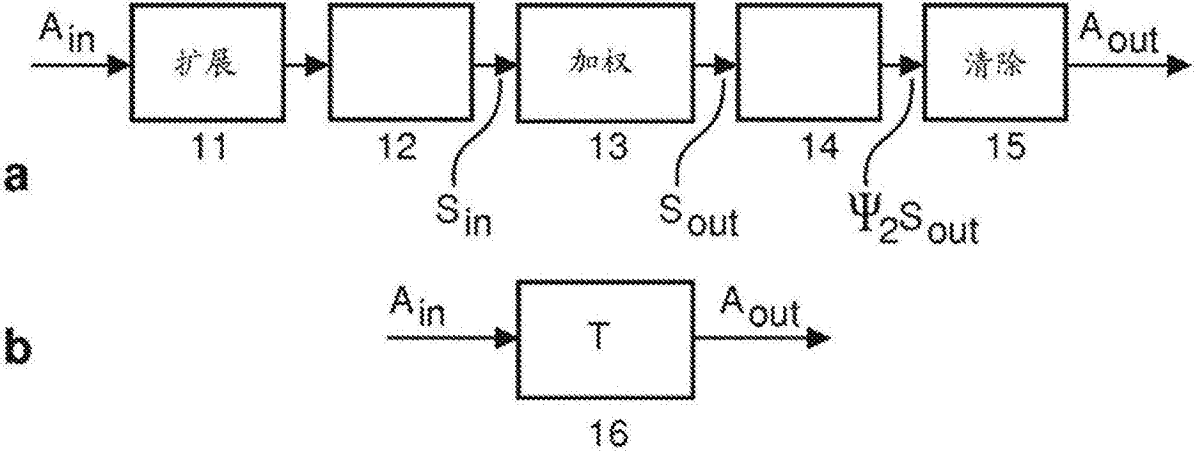


图1

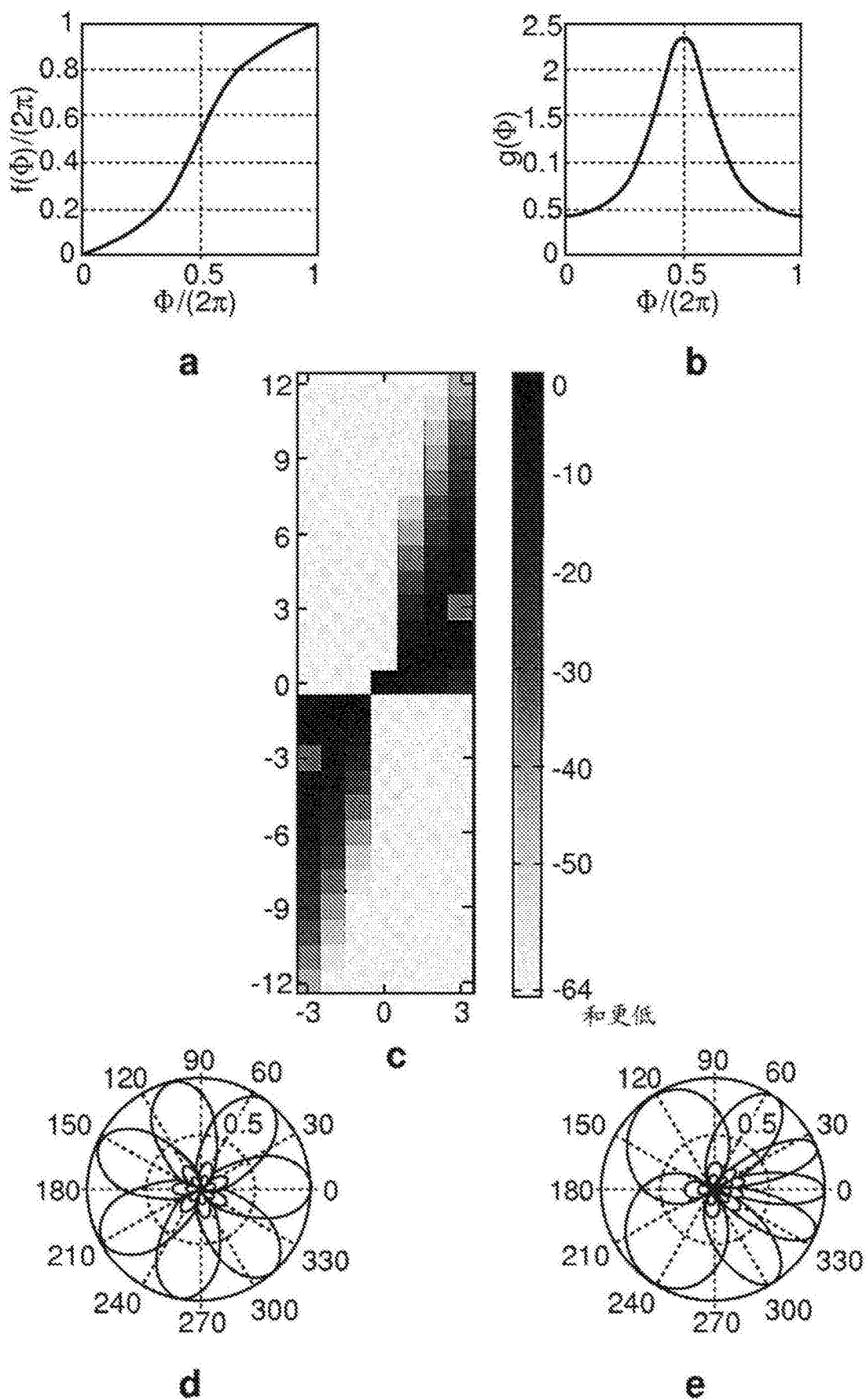


图2

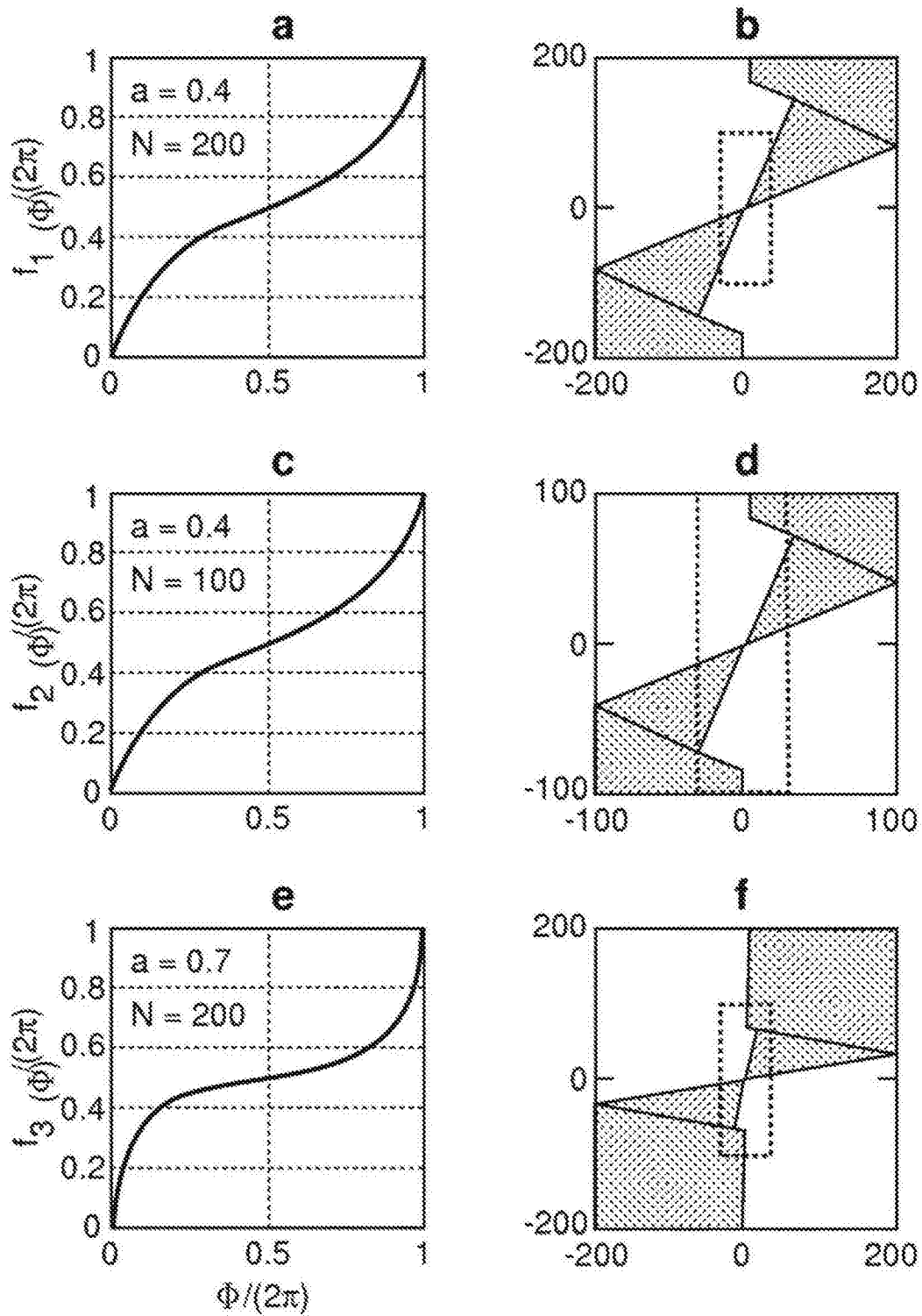


图3