



(12) 发明专利

(10) 授权公告号 CN 114025301 B

(45) 授权公告日 2024. 07. 30

(21) 申请号 202111170487.4

(22) 申请日 2017.10.11

(65) 同一申请的已公布的文献号
申请公布号 CN 114025301 A

(43) 申请公布日 2022.02.08

(30) 优先权数据
2016-211803 2016.10.28 JP

(62) 分案原申请数据
201780059396.9 2017.10.11

(73) 专利权人 松下电器(美国)知识产权公司
地址 美国加利福尼亚州

(72) 发明人 江原宏幸 吴恺 S.H.尼奥

(74) 专利代理机构 北京市柳沈律师事务所
11105

专利代理师 邸万奎

(51) Int.Cl.
H04S 7/00 (2006.01)
H04S 1/00 (2006.01)
G10L 19/008 (2013.01)

(56) 对比文件
EP 2806658 A1, 2014.11.26
WO 2015102920 A1, 2015.07.09

审查员 冯晨露

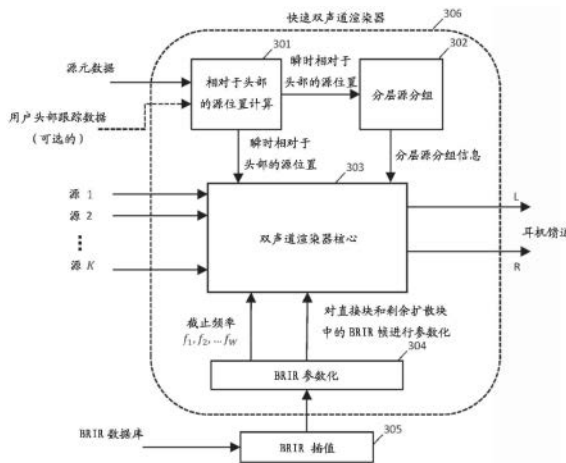
权利要求书1页 说明书9页 附图8页

(54) 发明名称

用于回放多个音频源的双声道渲染装置和方法

(57) 摘要

本公开公开了一种在给定多个音频源信号的情况下利用相关联的源元数据和双声道房间脉冲响应BRIR数据库生成双声道耳机回放信号的设备,其中相关联的源元数据指定一段时间内的源位置和移动轨迹中的至少一个,音频源信号能够是基于信道的、基于对象的、或者是两种信号的混合,设备包括:计算模块,计算相对于用户头部的方向和面向方向的、音频源的相对于头部位置;分组模块,根据音频源的相对于头部位置,对源信号以分层方式进行分组;参数化模块,对要用于渲染的BRIR进行参数化;双声道渲染器核心,将要渲染的每个源信号划分为多个块和帧;对标识有分层分组结果的、经参数化的BRIR序列求平均;以及对标识有分层分组结果的、经划分的源信号进行缩混。



1. 一种在给定多个音频源信号的情况下利用相关联的源元数据和双声道房间脉冲响应BRIR数据库生成双声道耳机回放信号的设备,其中所述相关联的源元数据指定一段时间内的源位置和移动轨迹中的至少一个,所述音频源信号能够是基于信道的、基于对象的、或者是两种信号的混合,所述设备包括:

计算模块,计算相对于用户头部的位置和面向方向的、音频源的相对于头部位置;

分组模块,根据所述音频源的所述相对于头部位置,对所述音频源信号以分层方式进行分组;

参数化模块,对要用于渲染的BRIR参数化为直接块和多个扩散块,并将所述直接块划分为多个直接块帧;

双声道渲染器核心,将要渲染的每个音频源信号划分为当前块和多个先前块,并将所述当前块划分为多个帧;

对于所述当前块的所述多个帧中的每个帧,对标识有分层分组结果的、经参数化的BRIR序列的对应直接块帧进行卷积;以及

对于所述先前块中的每一个,将标识有所述分层分组结果的、经划分的音频源信号缩混为信道信号;以及

求和器,对所述当前块的所述多个帧和经混缩的先前块求和,以生成回放信号。

2. 根据权利要求1所述的设备,其中,计算模块在给定源元数据和用户头部跟踪数据的情况下,对于所述音频源信号的每个时间帧或块,计算所述相对于头部位置。

3. 根据权利要求1所述的设备,其中,分组模块在给定对于每个帧所计算的相对的源位置的情况下,以具有不同分组分辨率的多个层来分层地执行所述分组。

4. 根据权利要求1所述的设备,其中,双声道渲染器核心使用所选择的BRIR帧,对所述音频源信号的当前块的帧执行逐帧双声道化处理,并且基于搜索最接近的已标记的BRIR帧选择每个BRIR帧,所述最接近的已标记的BRIR帧最靠近每个源的所计算的相对的位置。

5. 根据权利要求1所述的设备,其中,双声道渲染器核心通过加入源信号缩混模块来执行逐帧双声道化处理,使得能够根据所计算的源分组决策对所述音频源信号进行缩混,并且对所缩混的信号应用所述双声道化处理以降低计算复杂度。

6. 根据权利要求1所述的设备,其中,双声道渲染器核心使用BRIR的扩散块对所述音频源信号的先前块的缩混版本执行后期混响处理,并且对每个块应用不同的截止频率。

用于回放多个音频源的双声道渲染装置和方法

[0001] 本申请是申请日为2017年10月11日、申请号为:201780059396.9、发明名称为“用于回放多个音频源的双声道渲染装置和方法”的发明专利申请的分案申请。

技术领域

[0002] 本公开涉及用于耳机回放 (playback) 的数字音频信号的有效渲染 (render)。

背景技术

[0003] 空间音频是指沉浸式音频再现系统,允许听众感知高度的音频环绕。这种环绕感包括在方向和距离上对音频源的空间位置的感觉,使得听众将声音场景感知为好像他们处于自然声音环境中。

[0004] 通常有三种用于空间音频再现系统的录音格式。格式取决于音频内容制作站点使用的录制和混音方法。第一种格式是最熟知的基于信道的,其中音频信号的每个信道被指派为在再现站点的特定扬声器上回放。第二种格式称为基于对象的,其中可以由多个虚拟源(也称为对象)描述空间声音场景。每个音频对象可以由具有相关联的元数据的声音波形表示。第三种格式称为基于环绕声 (Ambisonic),其可以视为表示声场的球形扩展的系数信号。

[0005] 随着诸如移动电话、平板电脑等个人便携式设备的激增以及虚拟/增强现实的新兴应用,通过耳机渲染沉浸式空间音频变得越来越必要和有吸引力。双声道化是将输入的空间音频信号(例如,基于信道的信号、基于对象的信号或基于环绕声的信号)转换为耳机回放信号的过程。实质上,实际环境中的自然声音场景是由一对人耳感知的。这推断出,如果这些回放信号接近人类在自然环境中感知的声音,则耳机回放信号应该能够将空间声场渲染得尽可能地自然。

[0006] 双声道渲染的典型示例记录在MPEG-H 3D音频标准[参见NPL 1]中。图1示出了将基于信道的和基于对象的输入信号渲染到MPEG-H 3D音频标准中的双声道馈送的流程图。给定虚拟扬声器布局配置(例如,5.1、7.1或22.2),基于信道的信号 $1 \dots L_1$ 和基于对象的信号 $1 \dots L_2$ 首先分别经由格式转换器(101)和VBAP渲染器(102)转换为多个虚拟扬声器信号。然后,通过考虑BRIR数据库,经由双声道渲染器(103)将虚拟扬声器信号转换为双声道信号。引用列表

[0007] 非专利文献

[0008] [NPL 1] ISO/IEC DIS23008-3 “Information technology-High efficiency coding and media delivery in heterogeneous environments-Part 3:3D audio”

[0009] [NPL 2] T.Lee, H.O.Oh, J.Seo, Y.C.Park 以及 D.H.Youn, “Scalable Multiband Binaural Renderer for MPEG-H 3D Audio”, 在 IEEE Journal of Selected Topics in Signal Processing, 第9卷, 第5号, 第907-920页, 2015年8月。

发明内容

[0010] 一个非限制性和示例性实施例提供了一种用于多个移动音频源的快速双声道渲染的方法。本公开采用音频源信号(可以是基于对象的、基于信道的或两者的混合)、相关联的元数据、用户头部跟踪数据以及双声道房间脉冲响应(BRIR)数据库来生成耳机回放信号。本公开的一个非限制性和示例性实施例提供当在双声道渲染器中使用时的高空间分辨率和低计算复杂度。

[0011] 在一个一般方面,这里公开的技术的特征在于一种在给定多个音频源信号的情况下利用相关联的元数据和双声道房间脉冲响应(BRIR)数据库、有效地生成双声道耳机回放信号的方法,其中所述音频源信号可以是基于信道的、基于对象的或两种信号的混合。该方法包括以下步骤:(a)计算相对于用户头部的位置和面向方向的、音频源的瞬时相对于头部的位置,(b)根据所述音频源的所述瞬时相对于头部的位置,对源信号以分层方式进行分组,(c)对用于渲染的BRIR进行参数化(或者,将用于渲染的BRIR划分为多个块),(d)将每个要渲染的源信号划分为多个块和帧,(e)对标识有分层分组结果的经参数化的(经划分的)BRIR序列求平均,以及(f)对标识有分层分组结果的经划分的源信号进行缩混(downmixing)(平均)。

[0012] 在另一个一般方面,这里公开的技术的特征在于一种在给定多个音频源信号的情况下利用相关联的源元数据和双声道房间脉冲响应BRIR数据库生成双声道耳机回放信号的设备,其中相关联的源元数据指定一段时间内的源位置和移动轨迹中的至少一个,音频源信号能够是基于信道的、基于对象的、或者是两种信号的混合,设备包括:计算模块,计算相对于用户头部的位置和面向方向的、音频源的相对于头部位置;分组模块,根据音频源的相对于头部位置,对源信号以分层方式进行分组;参数化模块,对要用于渲染的BRIR进行参数化;双声道渲染器核心,将要渲染的每个源信号划分为多个块和帧;对标识有分层分组结果的、经参数化的BRIR序列求平均;以及对标识有分层分组结果的、经划分的源信号进行缩混。

[0013] 通过使用本公开的实施例中的方法,使用头部跟踪使能的头戴式设备来渲染快速移动的对象是有用的。

[0014] 应当注意,一般或特定实施例可以实现为系统、方法、集成电路、计算机程序、存储介质或其任何选择性组合。

[0015] 根据说明书和附图,所公开的实施例的其它益处和优点将变得显而易见。可以通过说明书和附图的各种实施例和特征单独地获得益处和/或优点,不需要提供全部以获得这些益处和/或优点中的一个或多个。

附图说明

[0016] 图1示出了MPEG-H 3D音频标准中将基于信道的和基于对象的信号渲染到双声道端的框图。

[0017] 图2示出了MPEG-H 3D音频中双声道渲染器的处理流程的框图。

[0018] 图3示出了所提出的快速双声道渲染器的框图。

[0019] 图4示出了源分组的图示。

[0020] 图5示出了将BRIR参数化为块和帧的图示。

- [0021] 图6示出了在不同的扩散块上应用不同的截止频率的图示。
- [0022] 图7示出了双声道渲染器核心的框图。
- [0023] 图8示出了基于分组的逐帧双声道化的框图。

具体实施方式

[0024] 下面将参考附图描述本公开的实施例中的配置和操作。以下实施例仅用于说明各种创造性步骤的原理。应理解,本文描述的细节的变型对于本领域其他技术人员将是显而易见的。

[0025] <形成本公开的基础的基本知识>

[0026] 作者研究了使用MPEG-H 3D音频标准来解决双声道渲染器所面临的问题的方法作为实例。

[0027] <问题1:空间分辨率受到信道/对象-信道-双声道渲染框架中的虚拟扬声器配置的限制>

[0028] 间接双声道渲染在诸如在MPEG-H 3D音频标准中的3D音频系统中被广泛采用,所述间接双声道渲染经由首先将基于信道的和基于对象的输入信号转换为虚拟扬声器信号,然后转换为双声道信号。然而,这样的框架导致空间分辨率是固定的,并且受到渲染路径中间的虚拟扬声器的配置的限制。例如,当虚拟扬声器被设置为5.1或7.1配置时,空间分辨率受到少量虚拟扬声器的约束,导致用户感知仅来自这些固定方向的声音。

[0029] 另外,双声道渲染器(103)中使用的BRIR数据库与虚拟听音室中的虚拟扬声器布局相关联。该事实偏离了BRIR应当是与生产场景相关联的BRIR(如果这样的信息能够从解码的比特流中获得)的预期情况。

[0030] 改善空间分辨率的方式包括增加扬声器的数量,例如增加到22.2配置,或者使用对象-双声道直接渲染方案。然而,当使用BRIR时,随着用于双声道化的输入信号的数量增加,这些方式可能导致高计算复杂度问题。将在以下段落中说明计算复杂度问题。

[0031] <问题2:使用BRIR的双声道渲染中的高计算复杂度>

[0032] 由于BRIR通常是长脉冲序列的事实,BRIR和信号之间的直接卷积是高计算要求的。因此,许多双声道渲染器寻求计算复杂度和空间质量之间的折衷。图2示出了MPEG-H 3D音频中双声道渲染器(103)的处理流程。这种双声道渲染器将BRIR拆分成“直接和早期回响(reflections)”和“后期混响(reverberation)”部分和处理,这两部分分开。因为“直接和早期回响”部分保留最多空间信息,因此每个BRIR的这部分在(201)中分别与信号进行卷积。

[0033] 另一方面,由于BRIR的“后期混响”部分包含较少的空间信息,所以可以将信号缩混(202)到一个信道中,使得在(203)中利用缩混的信道仅需要执行一次卷积。尽管该方法减少了后期混响处理(203)中的计算负荷,但是对于直接和早期部分处理(201),计算复杂度仍然可能非常高。这是因为在直接和早期部分处理(201)中分别处理每个源信号,并且随着源信号的数量增加,计算复杂度增加。

[0034] <问题3:不适合快速移动目标的情况或启用头部跟踪的情况>

[0035] 双声道渲染器(103)将虚拟扬声器信号视为输入信号,并且可以通过将每个虚拟扬声器信号与对应的双声道脉冲响应对进行卷积,来执行双声道渲染。头部相关的脉冲响

应 (HRIR) 和双声道房间脉冲响应 (BRIR) 通常用作脉冲响应, 其中后者由房间混响滤波器系数组成, 这使得它比HRIR长得多。

[0036] 卷积处理隐含地假设, 源位于固定位置——这对于虚拟扬声器是如此。然而, 有许多情况下音频源可以是移动的。一个示例是在虚拟现实 (VR) 应用中使用头戴式显示器 (HMD), 其中预期音频源的位置对于用户头部的任何旋转是不变的。这是通过沿相反方向旋转对象或虚拟扬声器的位置以消除用户头部旋转的效果来实现的。另一个示例是直接渲染对象, 其中这些对象可以随元数据中指定的不同位置移动。

[0037] 理论上, 由于渲染系统因为移动源而不再是线性时不变 (LTI) 系统, 因此没有直接的 (straight forward) 方法来渲染移动源。然而, 可以进行近似, 使得源被假设在短时间内是静止的, 并且在该短时间内, LTI假设是有效的。这在我们使用HRIR时是真的, 并且可以假设源在HRIR的滤波器长度内 (通常是毫秒的分数) 是静止的。因此, 源信号帧可以与对应的HRIR滤波器卷积以生成双声道馈送。然而, 当使用BRIR时, 由于滤波器长度通常更长 (例如, 0.5秒), 因此在BRIR滤波器长度时段期间不再假设源是静止的。除非利用BRIR滤波器对卷积进行附加处理, 否则源信号帧不能与BRIR滤波器直接卷积。

[0038] <问题的解决方案>

[0039] 本公开包括以下内容。首先, 它是直接将基于对象的和基于信道的信号渲染到双声道端而不通过虚拟扬声器的方法。可以解决<问题1>中的空间分辨率限制问题。其次, 它是将紧密 (close) 源分组到一个集群中的方法, 使得某些处理部分可以应用于一个集群内的源的缩混版本, 以节省<问题2>中的计算复杂度问题。将BRIR拆分成若干块并进一步将直接块 (对应于直接和早期回响) 划分成若干帧, 然后通过新的逐帧卷积方案执行双声道化滤波的方法, 所述新的逐帧卷积方案根据移动源的瞬时位置选择BRIR帧, 以解决<问题3>中的移动源问题。

[0040] <所提出的快速双声道渲染器的总体视图>

[0041] 图3示出了本公开的概览图。所提出的快速双声道渲染器 (306) 的输入包括K个音频源信号、源元数据, 所述源元数据指定一段时间内的源位置/移动轨迹以及指派的BRIR数据库。上述源信号可以是基于对象的信号、基于信道的信号 (虚拟扬声器信号) 或两者的混合, 并且源位置/移动轨迹可以是基于对象的源在一段时间上的位置串或基于信道的源的静态虚拟扬声器位置。

[0042] 此外, 输入还包括可选的用户头部跟踪数据, 该用户头部跟踪数据可以是瞬时用户头部面向方向或位置, 如果这些信息可从外部应用获得并且需要相对于用户头部旋转/移动调整所渲染的音频场景。快速双声道渲染器的输出是用于用户倾听的左右耳机馈送信号。

[0043] 为了获得输出, 快速双声道渲染器首先包括相对于头部的源位置计算模块 (301), 其通过采用瞬时源元数据和用户头部跟踪数据, 来计算相对于瞬时用户头部面向方向/位置的相对源位置数据。然后, 在分层源分组模块 (302) 中使用计算的相对于头部的源位置, 以生成分层源分组信息和双声道渲染器核心 (303), 用于根据瞬时源位置选择经参数化的BRIR。由 (302) 生成的分层信息还用在双声道渲染器核心 (303) 中, 用于降低计算复杂度的目的。分层源分组模块 (302) 的细节在<源分组>章节中描述。

[0044] 所提出的快速双声道渲染器还包括BRIR参数化模块 (304), 其将每个BRIR滤波器

拆分成若干块。它进一步将第一块划分成帧,并将每个帧与对应的BRIR目标位置标签附接。BRIR参数化模块(304)的细节在<BRIR参数化>章节中描述。

[0045] 注意,所提出的快速双声道渲染器将BRIR视为用于渲染音频源的滤波器。在BRIR数据库不足或者用户更喜欢使用高分辨率BRIR数据库的情况下,所提出的快速双声道渲染器支持外部BRIR插值模块(305),其基于附近的BRIR滤波器针对丢失的目标位置插入BRIR滤波器。然而,本文档中未指定这种外部模块。

[0046] 最后,所提出的快速双声道渲染器包括双声道渲染器核心(303),其是核心处理单元。它采用上述单独的源信号、计算的相对于头部的源位置、分层源分组信息以及经参数化的BRIR块/帧用于生成耳机馈送。在<双声道渲染器核心>章节和<基于源分组的逐帧双声道渲染>章节中描述双声道渲染器核心(303)的细节。

[0047] <源分组>

[0048] 图3中的分层源分组模块(302)将计算的瞬时相对于头部的源位置作为输入以用于基于任何两个音频源之间的相似性(例如,间距)计算音频源分组信息。这种分组决策可以用P层分层地进行,其中较高层具有较低分辨率,而较深层具有较高分辨率,以对源进行分组。第p层的第0簇被表示为:

[0049] [数学1]

[0050] $C_0^{(p)}$

[0051] 其中0是簇索引,p是层索引。图4示出了当P=2时这种分层源分组的简单示例。该图被示为顶视图,其中原点指示用户(倾听者)位置,y轴的方向指示用户面向方向,并且根据从(301)相对于用户计算的、他们的二维的相对于头部的位置绘制源。深层(第一层:p=1)将源分组为8个簇,其中第一簇 $C_1^{(1)}=\{1\}$ 包含源1,第二簇 $C_2^{(1)}=\{2,3\}$ 包含源2和3,第三簇 $C_3^{(1)}=\{4\}$ 包含源4,等等。高层(第二层:p=2)将源分为4个簇,其中源1、2和3被分组为簇1,由 $C_1^{(2)}=\{1,2,3\}$ 表示,源4和5被分组为簇2,由 $C_2^{(2)}=\{4,5\}$ 表示,以及源6被分组为簇3,由 $C_3^{(2)}=\{6\}$ 表示。

[0052] 层数P由用户根据系统复杂度要求选择,并且可以大于2。在高层上具有较低分辨率的适当层级设计可以导致较低的计算复杂度。要对源进行分组,一种简单的方式是基于将音频源存在的整个空间划分为多个小区域/区块(enclosure),如前面的示例所示。因此,基于它们所属的区域/区块对源进行分组。更专业地,可以基于一些特定的聚类算法(例如,k均值、模糊c均值算法)对音频源进行分组。这些聚类算法计算任何两个源之间的相似性度量,并将源分组为簇。

[0053] <BRIR参数化>

[0054] 本节描述图3中BRIR参数化模块(304)中的处理过程,将指派的BRIR数据库或插值的BRIR数据库作为输入。图5示出了将BRIR滤波器之一参数化为块和帧的过程。通常,由于包含房间回响,BRIR滤波器可以很长,例如在大厅中大于0.5秒。

[0055] 如上所述,如果在滤波器和源信号之间应用直接卷积,则使用这种长滤波器会导致高计算复杂度。如果音频源的数量增加,计算复杂度将增加。为了节省计算复杂度,每个BRIR滤波器被划分为直接块和扩散块,并且如在<双声道渲染器核心>章节中所述,将简化的处理应用于扩散块。可以通过每个BRIR滤波器的能量包围和成对的滤波器之间的耳间相

干性来确定将BRIR滤波器划分为块。由于能量和耳间相干性随着BRIR中时间的增加而减少,因此可以使用现有算法凭经验得出[见NPL 2]分离块的时间点。图5示出了BRIR滤波器被划分为直接块和W个扩散块的示例。直接块表示为:

[0056] [数学2]

[0057] $h_{\theta}^{(0)}(n)$

[0058] 其中n表示样本索引,上标(0)表示直接块, θ 表示该BRIR滤波器的目标位置。类似地,第w个扩散块表示为:

[0059] [数学3]

[0060] $h_{\theta}^{(w)}(n), w = 1, 2, \dots, W$

[0061] 其中w是扩散块索引。此外,如图6所示,基于BRIR的时频域中的能量分布,为每个块计算不同的截止频率 f_1, f_2, \dots, f_w ,它们是图3中的(304)的输出。在图3中的双声道渲染器核心(303)中,不处理高于截止频率 f_w 的频率(低能量部分)以便节省计算复杂度。因为扩散块包含较少的方向信息,因此它们将用于图7中的后期混响处理模块(703),所述后期混响处理模块(703)处理源信号的缩混版本以节省计算复杂度,这在<双声道渲染器核心>章节中详细描述。

[0062] 另一方面,BRIR的直接块包含重要的方向信息,并将在双声道回放信号中生成方向提示。为了满足音频源快速移动的情况,基于音频源仅在短时间段内静止的假设(即,例如在16kHz采样率时具有1024样本的时间帧)来执行渲染,并且,在图7所示的基于源分组的逐帧双声道化(701)的模块中逐帧处理双声道化。因此,直接块 $h_{\theta}^{(0)}(n)$ 被划分成帧,所述帧被表示为:

[0063] [数学4]

[0064] $h_{\theta}^{(0),m}(n)$

[0065] 其中 $m=0, \dots, M$ 表示帧索引,M是直接块中的帧总数。划分的帧还被分配位置标签 θ ,其对应于该BRIR滤波器的目标位置。

[0066] <双声道渲染器核心>

[0067] 本章节描述了如图3所示的双声道渲染器核心(303)的细节,其采用源信号、经参数化的BRIR帧/块以及计算的源分组信息用于生成耳机馈送。图7示出了双声道渲染器核心(303)的处理图,其分别处理源信号的当前块和先前块。首先,每个源信号被划分成当前块和W个先前块,其中W是在<BRIR参数化>章节中定义的扩散BRIR块的数量。第k个源信号的当前块被表示为:

[0068] [数学5]

[0069] $S_k^{(current)}(n)$

[0070] 并且先前的第w个块被表示为:

[0071] [数学6]

[0072] $S_k^{(current-w)}(n), w = 1, 2, \dots, W$

[0073] 如图7所示,使用BRIR的直接块在逐帧快速双声道化模块(701)中处理每个源的当前块。该处理表示为

[0074] [数学7]

$$[0075] \quad y^{(current)} = \beta(s_1^{(current)}(n), \dots, s_k^{(current)}(n), H^{(0)})$$

[0076] 其中 $y^{(current)}$ 表示(701)的输出,函数 $\beta(\cdot)$ 表示(701)的处理函数,其采用从图3中的(302)生成的分层源分组信息、所有源信号的当前块以及直接块中的BRIR帧作为输入, $H^{(0)}$ 表示直接块的BRIR帧的集合,其对应于当前块时间段期间的所有瞬时帧知道(frame-wise)的源位置。在<基于源分组的逐帧双声道渲染>章节中描述了这种逐帧快速双声道化模块(701)的细节。

[0077] 另一方面,源信号的先前块将在缩混模块(702)中缩混到一个信道中并传递到后期混响处理模块(703)。(703)中的后期混响处理被表示为:

[0078] [数学8]

$$[0079] \quad y^{(current-w)} = \gamma\left(\frac{1}{K} \sum_{k=1}^K s_k^{(current-w)}(n), h_{\theta_{ave}}^{(w)}(n)\right)$$

[0080] 其中 $y^{(current-w)}$ 表示(703)的输出, $\gamma(\cdot)$ 表示(703)的处理函数,其采用源信号的先前块的缩混版本、以及BRIR的扩散块作为输入。变量 θ_{ave} 表示块 $current-w$ 处所有K个源的平均位置。

[0081] 注意,可以使用卷积在时域中执行该后期混响处理。它也可以通过使用应用具有 fW 的截止频率的快速傅立叶变换(FFT)在频域中进行乘法来实现。还值得注意的是,取决于目标系统的计算复杂度,可以在扩散块上实现时域下采样。这种下采样可以减少信号样本的数量,从而减少FFT域中的乘法次数,从而降低计算复杂度。

[0082] 鉴于上述情况,最终通过以下生成双声道回放信号:

[0083] [数学9]

$$y^{(current)} + \sum_{w=1}^W y^{(current-w)}$$

[0084]

$$= y^{(current)} + \sum_{w=1}^W \gamma\left(\frac{1}{K} \sum_{k=1}^K s_k^{(current-w)}(n), h_{\theta_{ave}}^{(w)}(n)\right)$$

[0085] 如上式所示,对于每个扩散块 w ,由于对源信号应用了缩混处理 $\frac{1}{K} \sum_{k=1}^K s_k^{(current-w)}(n)$,所以仅需要执行一次后期混响处理 $\gamma(\cdot)$ 。与典型的直接卷积方法(其中必须针对K个源信号单独地执行这种处理(滤波))的情况相比,本公开降低了计算复杂度。

[0086] <基于源分组的逐帧双声道渲染>

[0087] 该章节描述了图7中基于源分组的逐帧快速双声道化模块(701)的细节,该模块处理源信号的当前块。首先,将第 k 个源信号 $s_k^{(current-w)}(n)$ 的当前块划分成帧,其中最近的帧由 $s_k^{(current),lfrm}(n)$ 表示,并且先前第 m 个帧由 $s_k^{(current),lfrm-m}(n)$ 表示。源信号的帧长度等于BRIR滤波器的直接块的帧长度。

[0088] 如图8所示,最近的帧 $s_k^{(current),lfrm}(n)$ 与包含在集合 $H^{(0)}$ 中的BRIR的直接块的第0个

帧 $h_{[\theta_k^{(current),lfrm}]_j}^{(0),0}(n)$ 卷积。通过搜索BRIR帧的已标记的位置 $[\theta_k^{(current),lfrm}]$ 来选择该BRIR帧, 该已标记的位置在最近的帧处最接近源的瞬时位置 $\theta_k^{(current),lfrm}$, 其中 $[\theta_k^{(current),lfrm}]$ 表示在BRIR数据库中找到最接近的标记值。由于BRIR的第0个帧包含最多的方向信息, 所以对每个源信号单独执行卷积以保留每个源的空间提示。可以使用频域中的乘法来执行卷积, 如图8中的(801)所示。

[0089] 对于先前帧 $s_k^{(current),lfrm-m}(n)$ 中的每一个, 其中 $m \geq 1$, 假定卷积是用包含在 $H^{(0)}$ 中的BRIR的直接块的第 m 个帧 $h_{[\theta_k^{(current),lfrm-m}]_j}^{(0),m}(n)$ 执行的, 其中 $[\theta_k^{(current),lfrm-m}]$ 表示该BRIR帧的已标记的位置, 该已标记的位置最接近帧 $lfrm-m$ 处的源位置。

[0090] 注意, 随着 m 增加, $h_{[\theta_k^{(current),lfrm-m}]_j}^{(0),m}(n)$ 中包含的方向信息减少。因此, 为了节省计算复杂度并且如(802)所示, 本公开根据分层源分组决策 $c_0^{(p)}$ (从(302)生成并在<源分组>章节中讨论) 对 $s_k^{(current),lfrm-m}(n)$, $k=1, 2, \dots, K$ (其中 $m \geq 1$) 进行缩混, 随后是与源信号帧的缩混版本的卷积。

[0091] 例如, 如果第二层源分组应用于信号帧 $s_k^{latest,frame-2}(n)$ (即, $m=2$) 并且源4和5被分组为第二簇 $c_2^{(2)}=\{4,5\}$, 可以通过将源信号平均为 $(s_4^{latest,frame-2}(n) + s_5^{latest,frame-2}(n))/2$ 来应用缩混, 并且在这个帧处在这个平均信号和具有平均源位置BRIR帧之间应用卷积。

[0092] 注意, 可以在帧上应用不同的分层。本质上, 应该考虑高分辨率分组用于BRIR的早期帧以保留空间提示, 而低分辨率分组被考虑用于BRIR的后期帧以降低计算复杂度。最后, 帧知道的处理信号被传递到混音器, 该混音器执行求和以生成(701)的输出, 即 $y^{(current)}$ 。

[0093] 根据本公开的实施例, 至少公开了如下方法。

[0094] 根据本公开的实施例的在给定多个音频源信号的情况下利用相关联的元数据和双声道房间脉冲响应BRIR数据库生成双声道耳机回放信号的方法, 其中所述音频源信号能够是基于信道的、基于对象的、或者是两种信号的混合, 所述方法包括: 计算相对于用户头部的位置和面向方向的、所述音频源的瞬时相对于头部位置; 根据所述音频源的所述瞬时相对于头部位置, 以分层方式对所述源信号进行分组; 对要用于渲染的BRIR进行参数化; 将要渲染的每个源信号划分为多个块和帧; 对标识有分层分组结果的、经参数化的BRIR序列求平均; 以及对标识有所述分层分组结果的、经划分的源信号进行缩混。

[0095] 根据本公开的实施例的方法, 其中, 在给定源元数据和用户头部跟踪数据的情况下, 对于所述源信号的每个时间帧/块, 立即计算所述相对于头部的源位置。

[0096] 根据本公开的实施例的方法, 其中, 在给定对于每个帧所计算的瞬时相对的源位置的情况下, 以具有不同分辨率的多个层来分层地执行所述分组。

[0097] 根据本公开的实施例的方法, 其中, 将所述BRIR数据库中的每个BRIR滤波器信号划分为包含多个帧的直接块以及多个扩散块, 并且使用所述BRIR滤波器信号的目标位置来标记所述帧和块。

[0098] 根据本公开的实施例的方法, 其中, 将所述源信号划分为当前块和多个先前块, 并且将所述当前块进一步划分为多个帧。

[0099] 根据本公开的实施例的方法,其中,使用所选择的BRIR帧,对所述源信号的所述当前块的所述帧执行逐帧双声道化处理,并且每个BRIR帧的选择是基于搜索最接近的已标记的BRIR帧,所述最接近的已标记的BRIR帧最靠近每个源的所计算的瞬时相对的位置。

[0100] 根据本公开的实施例的方法,其中,通过加入源信号缩混模块来执行逐帧双声道化处理,使得能够根据所计算的源分组决策对所述源信号进行缩混,并且对所缩混的信号应用所述双声道化处理以降低计算复杂度。

[0101] 根据本公开的实施例的方法,其中,使用BRIR的所述扩散块对所述源信号的所述先前块的缩混版本执行后期混响处理,并且对每个块应用不同的截止频率。

[0102] 在前述实施例中,通过上述示例,本公开配置有硬件,但是本公开还可以通过与硬件协作的软件来提供。

[0103] 另外,在实施例的描述中使用的功能块通常被实现为LSI设备,其是集成电路。功能块可以形成为单独的芯片,或者功能块的一部分或全部可以集成到单个芯片中。这里使用术语“LSI”,但是也可以使用术语“IC”、“系统LSI”、“超级LSI”或“超LSI”,这取决于集成度。

[0104] 另外,电路集成不限于LSI,并且可以通过专用电路或除LSI之外的通用处理器来实现。在制造LSI之后,可以使用可编程的现场可编程门阵列(FPGA),或者允许重新配置LSI中的电路单元的连接和设置的可重构处理器。

[0105] 如果替代LSI的电路集成技术由于半导体技术或源自该技术的其他技术的进步而出现,则可以使用这种技术来集成功能块。另一种可能性是生物技术和/或类似物的应用。

[0106] 工业适用性

[0107] 本公开可以应用于用于渲染用于耳机回放的数字音频信号的方法。

[0108] 参考标记列表

[0109] 101 格式转换器

[0110] 102 VBAP渲染器

[0111] 103 双声道渲染器

[0112] 201 直接和早期部分处理

[0113] 202 缩混

[0114] 203 后期混响部分处理

[0115] 204 混音

[0116] 301 相对于头部的源位置计算模块

[0117] 302 分层源分组模块

[0118] 303 双声道渲染器核心

[0119] 304 BRIR参数化模块

[0120] 305 外部BRIR插值模块

[0121] 306 快速双声道渲染器

[0122] 701 逐帧快速双声道化模块

[0123] 702 缩混模块

[0124] 703 后期混响处理模块

[0125] 704 求和

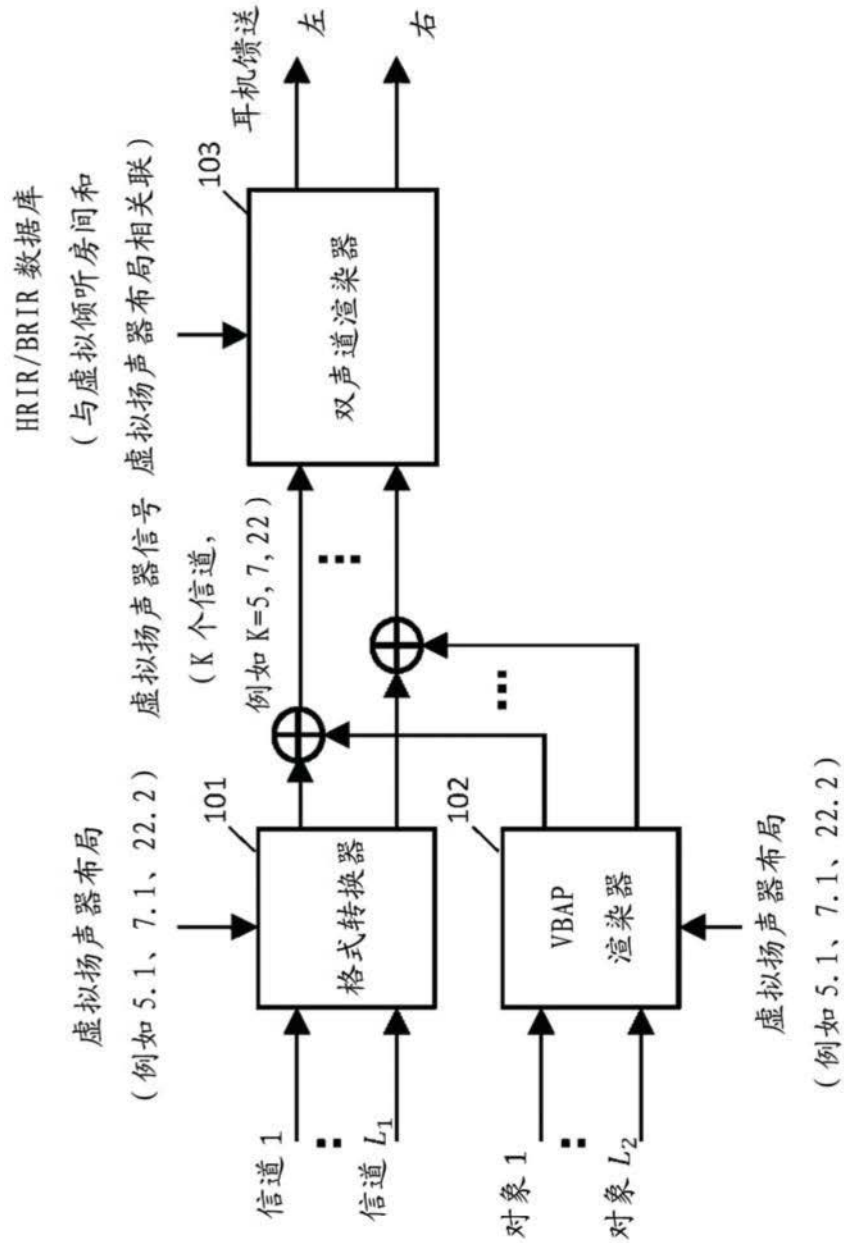


图1

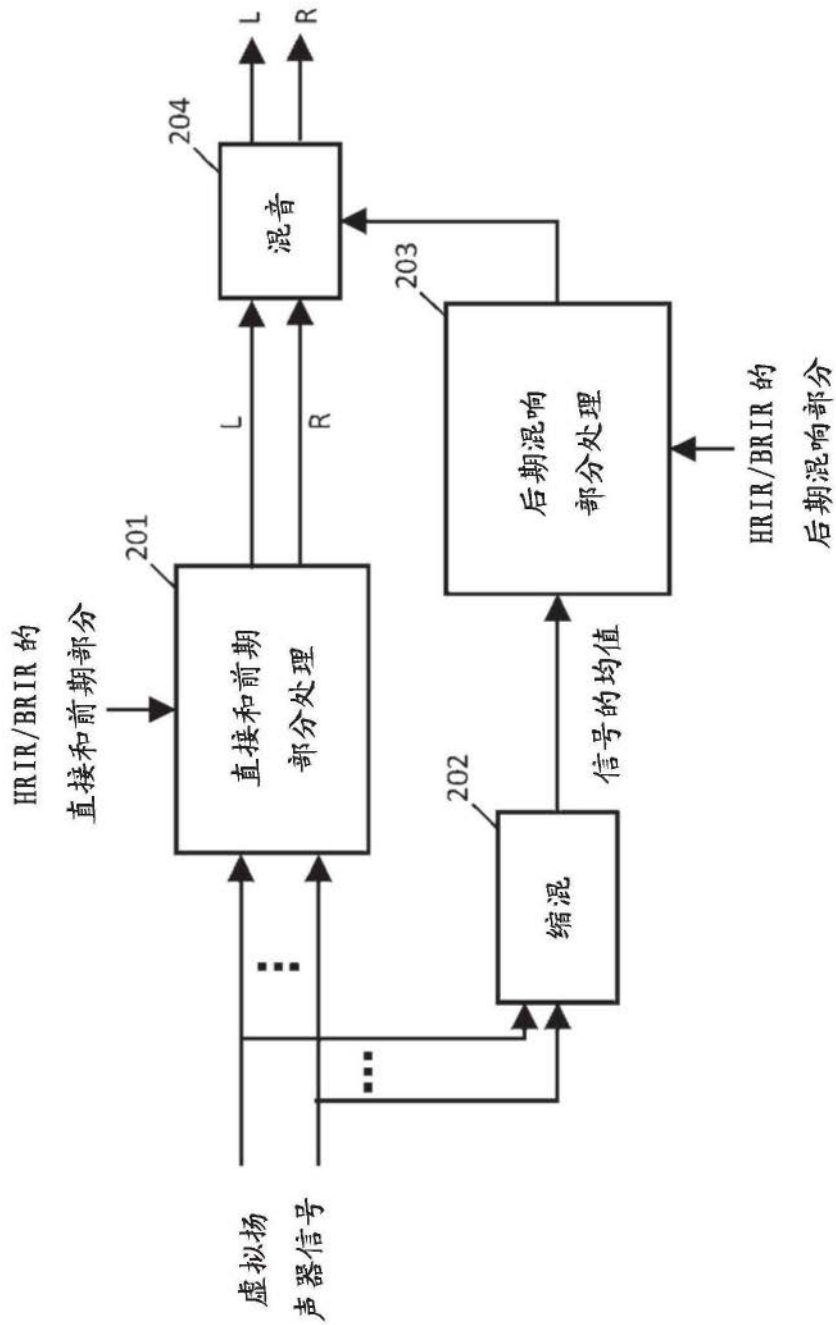


图2

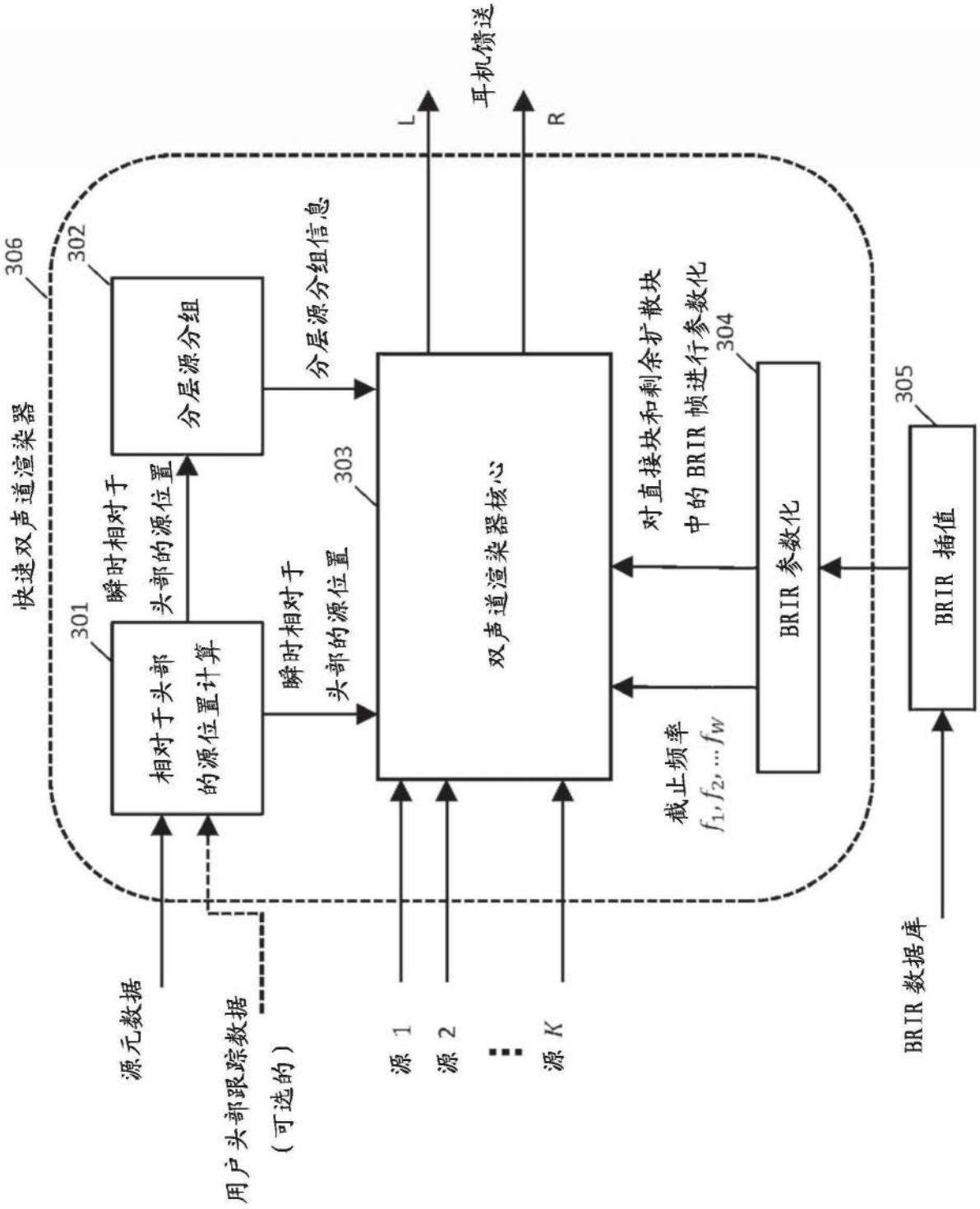


图3

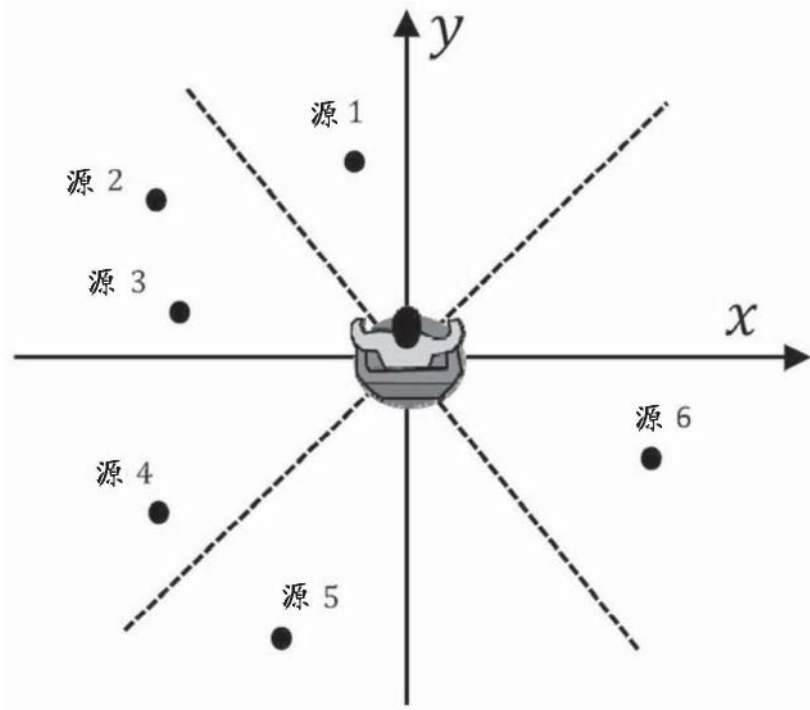


图4

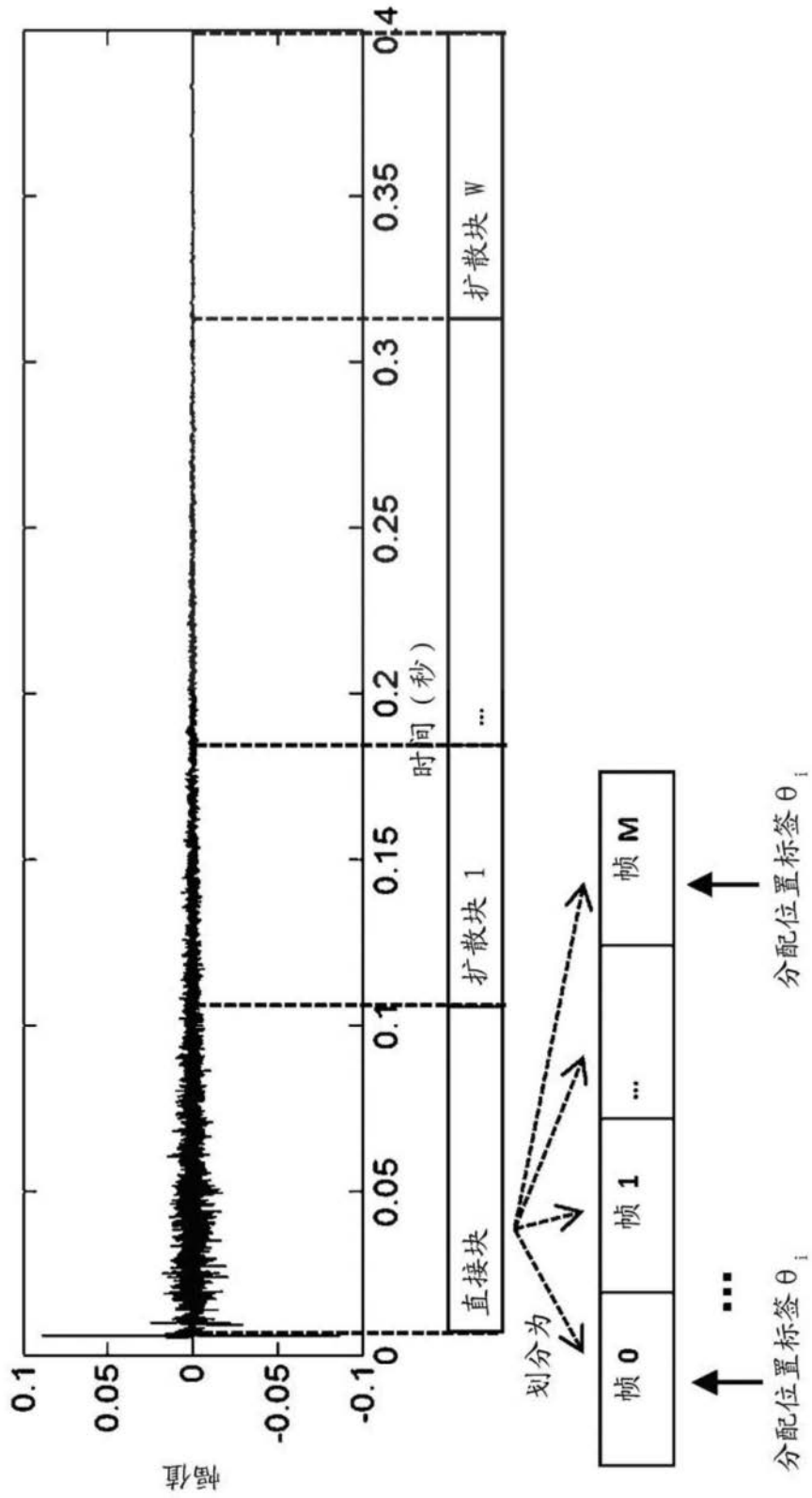


图5

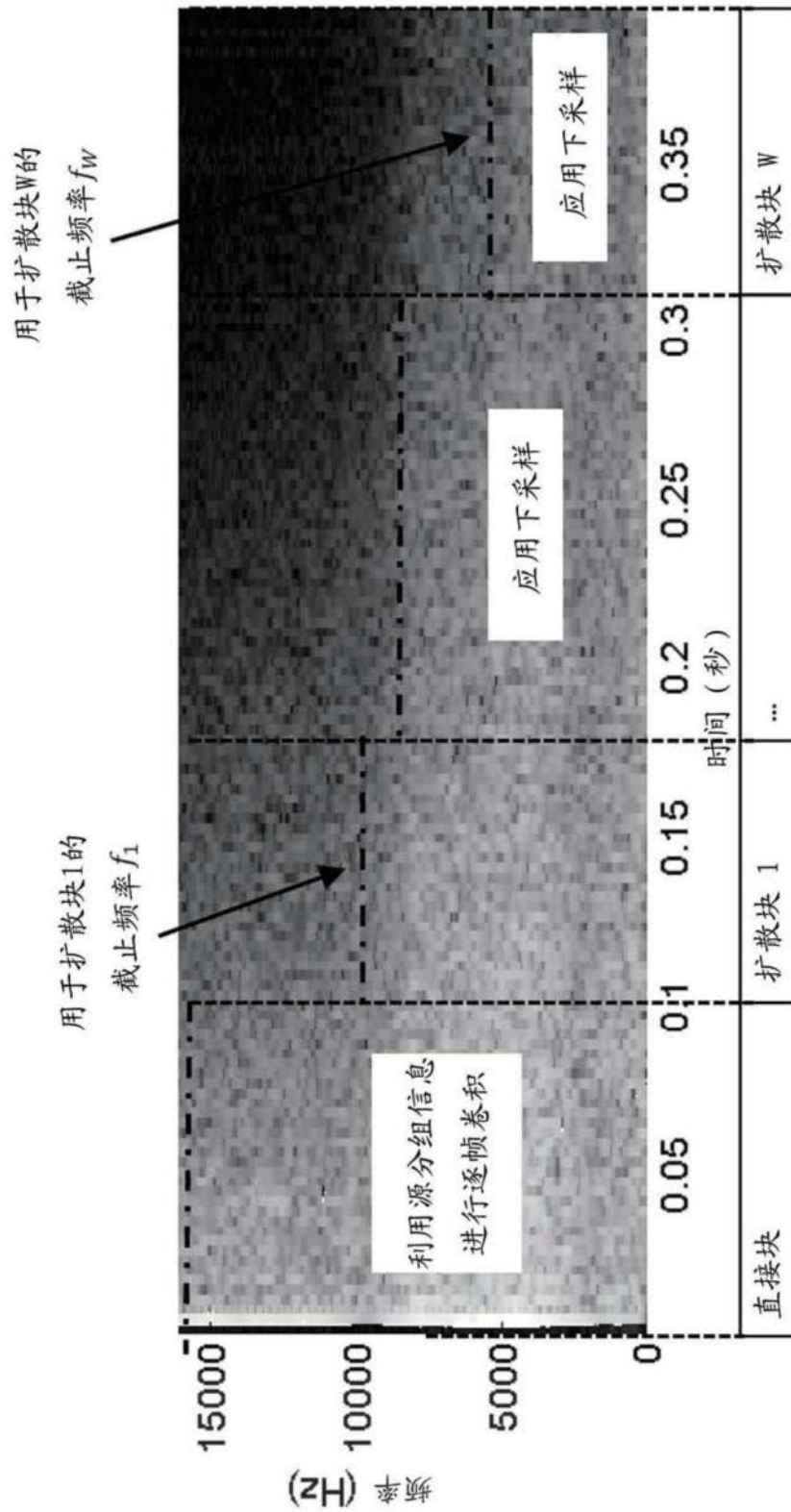


图6

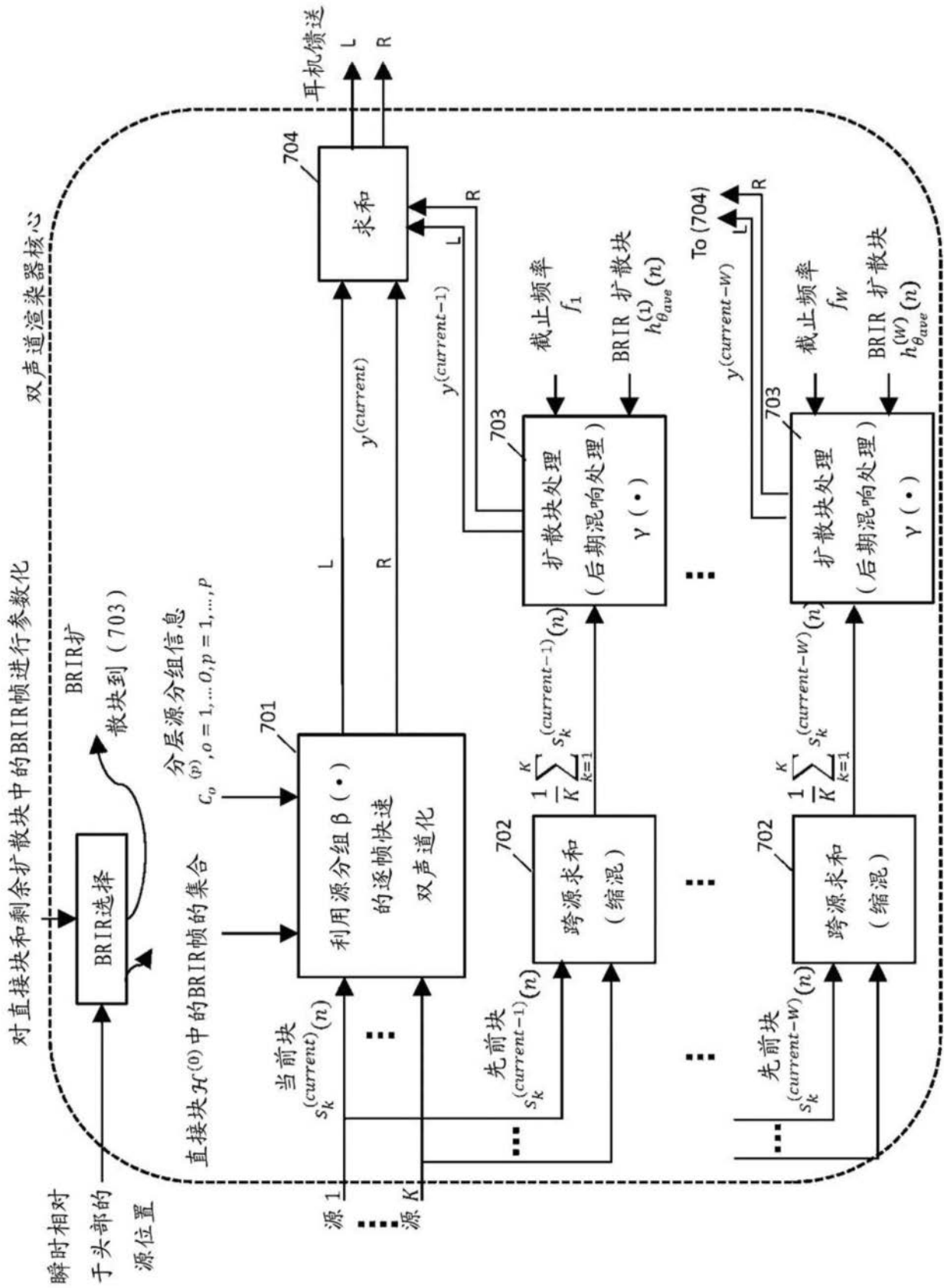


图7

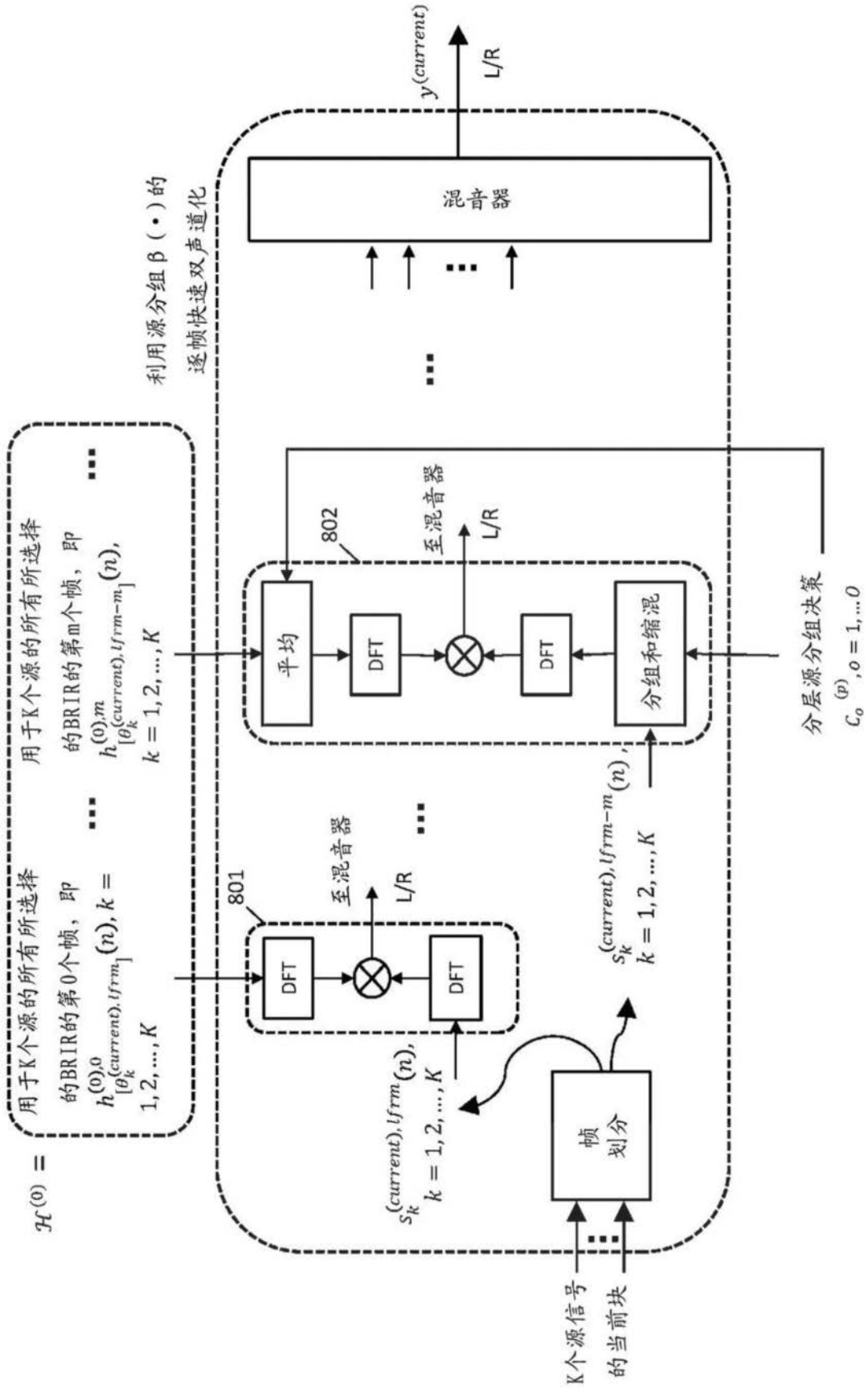


图8