



## (12)发明专利

(10)授权公告号 CN 106471578 B

(45)授权公告日 2020.03.31

(21)申请号 201580027072.8

(22)申请日 2015.05.15

(65)同一申请的已公布的文献号  
申请公布号 CN 106471578 A

(43)申请公布日 2017.03.01

(30)优先权数据

61/994,763 2014.05.16 US

62/004,076 2014.05.28 US

62/118,434 2015.02.19 US

14/712,854 2015.05.14 US

(85)PCT国际申请进入国家阶段日  
2016.11.11(86)PCT国际申请的申请数据  
PCT/US2015/031195 2015.05.15(87)PCT国际申请的公布数据  
W02015/176005 EN 2015.11.19(73)专利权人 高通股份有限公司  
地址 美国加利福尼亚州

(72)发明人 金墨永 尼尔斯·京特·彼得斯

(74)专利代理机构 北京律盟知识产权代理有限公司 11287

代理人 宋献涛

(51)Int.Cl.

G10L 19/008(2013.01)

(56)对比文件

US 2014016786 A1, 2014.01.16,

US 2012053710 A1, 2012.03.01,

US 2005157894 A1, 2005.07.21,

CN 101578865 A, 2009.11.11,

CN 103384900 A, 2013.11.06,

CN 102971789 A, 2013.03.13,

Johannes Boehm et al.《Scalable

Decoding Mode for MPEG-H 3D Audio HOA》.

《108. MPEG MEETING》.2014,

审查员 吕新强

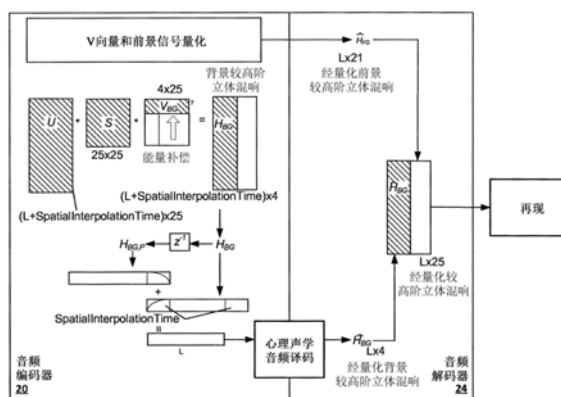
权利要求书3页 说明书26页 附图14页

## (54)发明名称

用于较高阶立体混响信号之间的交叉淡化的方法和装置

## (57)摘要

本发明大体上描述用于交叉淡化球谐系数的集合的技术。包括存储器和处理器的音频编码装置或音频解码装置可经配置以执行所述技术。所述存储器可经配置以存储球谐系数SHC的第一集合和SHC的第二集合。SHC的所述第一集合描述第一声场。SHC的所述第二集合描述第二声场。所述处理器可经配置以在SHC的所述第一集合与SHC的第二集合之间进行交叉淡化以获得经交叉淡化SHC的第一集合。



1. 一种用于较高阶立体混响信号之间的交叉淡化的方法,其包括:

由音频编码器获得对应于球谐系数SHC的第一集合的SHC的分解,所述分解包括表示环境声场的空间特性的向量的第一集合和表示所述环境声场的时间和能量特性的向量的第二集合;

由所述音频编码器对向量的所述第一集合执行能量补偿以获得经能量补偿的向量的集合;

由所述音频编码器将向量的所述第二集合乘以经能量补偿的向量的所述集合以获得经能量补偿环境SHC的第一集合;以及

由所述音频编码器在经能量补偿环境球谐系数SHC的所述第一集合与经能量补偿环境SHC的第二集合之间进行交叉淡化以获得经交叉淡化经能量补偿环境SHC的第一集合;

其中经能量补偿环境SHC的所述第一集合对应于当前帧,且其中经能量补偿环境SHC的所述第二集合对应于前一帧。

2. 根据权利要求1所述的方法,

其中SHC的所述第一集合包含对应于具有大于一的阶的基底函数的SHC,且

其中SHC的所述第二集合包含对应于具有大于一的阶的基底函数的SHC。

3. 根据权利要求1所述的方法,其中执行所述能量补偿包括使用至少部分地随着指示帧长度的一或多个位而变获得的开窗函数执行所述能量补偿。

4. 根据权利要求1所述的方法,其中交叉淡化包括基于经能量补偿环境SHC的所述第二集合的一部分修改经能量补偿环境SHC的所述第一集合的一部分。

5. 根据权利要求1所述的方法,其中所述方法进一步包括由耦合到所述音频编码器的麦克风俘获表示环境SHC的第一集合和环境SHC的第二集合的音频数据。

6. 一种音频解码装置,其包括:

存储器,其经配置以存储表示前景声场的空间特性的向量的第一集合、表示所述前景声场的时间和能量特性的向量的第二集合、经能量补偿环境球谐系数SHC的第一集合和经能量补偿环境SHC的第二集合,其中经能量补偿环境SHC的所述第一集合描述第一环境声场且经能量补偿环境SHC的所述第二集合描述第二环境声场,以及

一或多个处理器,其耦合到所述存储器且经配置以:

在经能量补偿环境SHC的所述第一集合与经能量补偿环境SHC的所述第二集合之间进行交叉淡化以获得经交叉淡化经能量补偿环境SHC的第一集合;以及

基于向量的所述第一集合、向量的所述第二集合、以及经交叉淡化经能量补偿环境SHC的所述第一集合再现一或多个扬声器馈送。

7. 根据权利要求6所述的音频解码装置,

其中SHC的所述第一集合包含对应于具有大于一的阶的基底函数的SHC,且

其中SHC的所述第二集合包含对应于具有大于一的阶的基底函数的SHC。

8. 根据权利要求6所述的音频解码装置,

其中经能量补偿环境SHC的所述第一集合对应于当前帧,且

其中经能量补偿环境SHC的所述第二集合对应于前一帧。

9. 根据权利要求6所述的音频解码装置,其中所述一或多个处理器经配置以通过至少基于经能量补偿环境SHC的所述第二集合的一部分修改经能量补偿环境SHC的所述第一集

合的一部分而进行交叉淡化。

10. 根据权利要求6所述的音频解码装置,其进一步包括扬声器,所述扬声器经配置以基于扬声器馈送而再生声场。

11. 一种音频编码装置,其包括:

一或多个处理器,其经配置以:

获得对应于球谐系数SHC的第一集合的SHC的分解,所述分解包括表示环境声场的空间特性的向量的第一集合和表示所述环境声场的时间和能量特性的向量的第二集合;

对向量的所述第一集合执行能量补偿以获得经能量补偿的向量的集合;

将向量的所述第二集合乘以经能量补偿的向量的所述集合以获得经能量补偿环境SHC的第一集合;以及

存储器,其耦合到所述一或多个处理器且经配置以存储经能量补偿环境球谐系数SHC的所述第一集合和经能量补偿环境SHC的第二集合,以及

其中所述一或多个处理器经配置以在经能量补偿环境SHC的所述第一集合与经能量补偿环境SHC的所述第二集合之间进行交叉淡化以获得经交叉淡化经能量补偿环境SHC的第一集合;

其中经能量补偿环境SHC的所述第一集合对应于当前帧,且

其中经能量补偿环境SHC的所述第二集合对应于前一帧。

12. 根据权利要求11所述的音频编码装置,

其中SHC的所述第一集合包含对应于具有大于一的阶的基底函数的SHC,且

其中SHC的所述第二集合包含对应于具有大于一的阶的基底函数的SHC。

13. 根据权利要求11所述的音频编码装置,其中所述一或多个处理器经配置以使用至少部分地随着指示帧长度的一或多个位而变获得的开窗函数执行所述能量补偿。

14. 根据权利要求11所述的音频编码装置,其中所述一或多个处理器经配置以通过至少基于经能量补偿环境SHC的所述第二集合的一部分修改经能量补偿环境SHC的所述第一集合的一部分而进行交叉淡化。

15. 根据权利要求11所述的音频编码装置,其进一步包括麦克风,所述麦克风经配置以捕获指示SHC的所述第一和第二集合的音频数据。

16. 一种用于较高阶立体混响信号之间的交叉淡化的方法,其包括:

由音频解码器获得表示前景声场的空间特性的向量的第一集合和表示所述前景声场的时间和能量特性的向量的第二集合;

由所述音频解码器获得经能量补偿环境球谐系数SHC的第一集合和经能量补偿环境SHC的第二集合,其中经能量补偿环境SHC的所述第一集合描述第一环境声场且经能量补偿环境SHC的所述第二集合描述第二环境声场;

由所述音频解码器在经能量补偿环境SHC的所述第一集合与经能量补偿环境SHC的所述第二集合之间进行交叉淡化以获得经交叉淡化经能量补偿环境SHC的第一集合;以及

由所述音频解码器基于向量的所述第一集合、向量的所述第二集合、以及经交叉淡化经能量补偿环境SHC的所述第一集合再现一或多个扬声器馈送。

17. 根据权利要求16所述的方法,

其中SHC的所述第一集合包含对应于具有大于一的阶的基底函数的SHC,且

其中SHC的所述第二集合包含对应于具有大于一的阶的基底函数的SHC。

18. 根据权利要求16所述的方法，

其中经能量补偿环境SHC的所述第一集合对应于当前帧，且

其中经能量补偿环境SHC的所述第二集合对应于前一帧。

19. 根据权利要求16所述的方法，其中在经能量补偿环境SHC的所述第一集合与经能量补偿环境SHC的所述第二集合之间进行交叉淡化包括通过至少基于经能量补偿环境SHC的所述第二集合的一部分修改经能量补偿环境SHC的所述第一集合的一部分而进行交叉淡化。

20. 根据权利要求16所述的方法，其进一步包括由一或多个扬声器且基于所述一或多个扬声器馈送而再生声场。

21. 根据权利要求16所述的方法，其中获得经能量补偿环境球谐系数SHC的第一集合和经能量补偿环境SHC的第二集合包括获得位流，所述位流包含所述经交叉淡化经能量补偿环境SHC的表示和对应于所述经交叉淡化经能量补偿环境SHC的经交叉淡化前景SHC的表示。

22. 根据权利要求16所述的方法，其中向量的所述第一集合和向量的所述第二集合表示经交叉淡化前景SHC，且其中获得向量的所述第一集合和向量的所述第二集合包括获得包含所述经交叉淡化前景SHC的表示的位流。

## 用于较高阶立体混响信号之间的交叉淡化的方法和装置

[0001] 本申请案主张以下各美国临时申请案的权利：

[0002] 2014年5月16日申请的标题为“较高阶立体混响信号之间的交叉淡化 (CROSSFADING BETWEEN HIGHER ORDER AMBISONIC SIGNALS)”的第61/994,763号美国临时申请案；

[0003] 2014年5月28日申请的标题为“较高阶立体混响信号之间的交叉淡化”的第62/004,076号美国临时申请案；以及

[0004] 2015年2月19日申请的标题为“较高阶立体混响信号之间的交叉淡化”的第62/118,434号美国临时申请案，

[0005] 前述所列各美国临时申请案中的每一者以引用的方式并入本文中，如同在其相应全文中所阐述般。

### 技术领域

[0006] 本发明涉及音频数据，且更确切地说，涉及高阶立体混响音频数据的译码。

### 背景技术

[0007] 较高阶立体混响 (HOA) 信号 (常由多个球谐系数 (SHC) 或其它阶层元素表示) 是声场的三维表示。此HOA或SHC表示可以独立于用以重放从此SHC信号再现的多信道音频信号的局部扬声器几何布置的方式来表示此声场。此SHC信号还可促进向后相容性，因为可将此SHC信号再现为众所周知的且被广泛采用的多信道格式 (例如，5.1音频信道格式或7.1音频信道格式)。SHC表示因此可实现对声场的更好表示，其也适应向后兼容性。

### 发明内容

[0008] 大体上描述用于在环境HOA系数之间进行交叉淡化的技术。举例来说，描述用于在能量补偿域中在环境HOA系数的当前集合与环境HOA系数的前一集合之间进行交叉淡化的技术。以此方式，本发明的技术可使环境HOA系数的前一集合与环境HOA系数的当前集合之间的转变平滑。

[0009] 在一个方面中，一种方法包含：由装置在环境球谐系数 (SHC) 的第一集合与环境SHC的第二集合之间进行交叉淡化以获得经交叉淡化环境SHC的第一集合，其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场。

[0010] 在另一方面，一种装置包含：一或多个处理器；以及至少一个模块，其可由所述一或多个处理器执行以在环境SHC的第一集合与环境SHC的第二集合之间进行交叉淡化以获得经交叉淡化环境SHC的第一集合，其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场。

[0011] 在另一方面，一种装置包含：用于获得环境SHC的第一集合的装置，其中SHC的所述第一集合描述第一声场；用于获得环境SHC的第二集合的装置，其中SHC的所述第二集合描述第二声场；以及用于在环境SHC的第一集合与环境SHC的第二集合之间进行交叉淡化以获

得经交叉淡化环境SHC的第一集合的装置。

[0012] 在另一方面,一种计算机可读存储媒体存储指令,所述指令当执行时致使装置的一或多个处理器在环境SHC的第一集合与环境SHC的第二集合之间进行交叉淡化以获得经交叉淡化环境SHC的第一集合,其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场。

[0013] 在另一方面,一种方法包括:由装置在球谐系数(SHC)的第一集合与SHC的第二集合之间进行交叉淡化以获得经交叉淡化SHC的第一集合,其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场。

[0014] 在另一方面,一种音频解码装置包括存储器,其经配置以存储球谐系数(SHC)的第一集合和SHC的第二集合,其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场。所述音频解码装置进一步包括一或多个处理器,其经配置以在SHC的所述第一集合与SHC的第二集合之间进行交叉淡化以获得经交叉淡化环境SHC的第一集合。

[0015] 在另一方面,一种音频编码装置包括存储器,其经配置以存储球谐系数(SHC)的第一集合和SHC的第二集合,其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场。所述音频编码装置还包括一或多个处理器,其经配置以在SHC的所述第一集合与SHC的第二集合之间进行交叉淡化以获得经交叉淡化SHC的第一集合。

[0016] 在另一方面,一种设备包括:用于存储球谐系数(SHC)的第一集合和SHC的第二集合的装置,其中SHC的所述第一集合描述第一声场且SHC的所述第二集合描述第二声场;以及用于在SHC的所述第一集合与SHC的第二集合之间进行交叉淡化以获得经交叉淡化SHC的第一集合的装置。

[0017] 在附图和以下描述中陈述所述技术的一或多个方面的细节。这些技术的其它特征、目标及优点将从所述描述及图式以及所附权利要求书中显而易见。

## 附图说明

[0018] 图1是说明具有各种阶数及子阶数的球谐基底函数的图。

[0019] 图2是说明可执行本发明中所描述的技术的各种方面的系统的图。

[0020] 图3为更详细说明可执行本发明中描述的技术的各种方面的图2的实例中所示的音频编码装置的一个实例的框图。

[0021] 图4是更详细说明图2的音频解码装置的框图。

[0022] 图5是说明音频编码装置执行本发明中所描述的基于向量的合成技术的各种方面的示范性操作的流程图。

[0023] 图6是说明音频解码装置在执行本发明中描述的技术的各种方面中的示范性操作的流程图。

[0024] 图7和8是更详细说明可指定压缩空间分量的位流的图。

[0025] 图9是更详细说明可指定压缩空间分量的位流的一部分的图。

[0026] 图10说明用于获得如本文所描述的空间-时间内插的技术的表示。

[0027] 图11是说明根据本文所述的技术的用于多维信号的依序SVD块的人工US矩阵( $US_1$ 及 $US_2$ )的框图。

[0028] 图12是说明根据本发明中所描述的技术使用奇异值分解及空间时间分量的平滑

来分解较高阶立体混响 (HOA) 信号的后续帧的框图。

[0029] 图13是说明经配置以执行本发明中描述的一或多个技术的一或多个音频编码器和音频解码器的图。

[0030] 图14是更详细说明图3的实例中所示的音频编码装置的交叉淡化单元的框图。

## 具体实施方式

[0031] 环绕声的演进现今已使得许多输出格式可用于娱乐。此等消费型环绕声格式的实例大部分为“信道”式的,这是因为其以某些几何坐标隐含地指定到扩音器的馈入。消费型环绕声格式包含普遍的5.1格式(其包含以下六个信道:左前 (FL)、右前 (FR)、中心或前中心、左后或左环绕、右后或右环绕,及低频效应 (LFE))、发展中的7.1格式、包含高度扬声器的各种格式,例如7.1.4格式及22.2格式(例如,用于供超清晰电视标准使用)。非消费型格式可涵括任何数目的扬声器(成对称和非对称几何形状),其常常称为“环绕阵列”。此种阵列的一个实例包含定位在截角二十面体的拐角上的坐标处的32个扩音器。

[0032] 到未来MPEG编码器的输入视情况为三个可能格式中的一者:(i)传统的基于信道的音频(如上文所论述),其意图由处于预先指定的位置处的扩音器重放;(ii)基于对象的音频,其涉及用于单个音频对象的具有含有其位置坐标(以及其它信息)的相关联元数据的离散脉码调制 (PCM) 数据;及 (iii) 基于场景的音频,其涉及使用球面谐波基底函数的系数(也称为“球谐系数”或SHC、“较高阶立体混响”或HOA及“HOA系数”)来表示声场。此未来MPEG编码器更详细地描述于国际标准化组织/国际电工委员会 (ISO) / (IEC) JTC1/SC29/WG11/N13411的标题为“要求对于3D音频的提议 (Call for Proposals for 3D Audio)”的文献中,该文献于2013年1月在瑞士日内瓦发布,且可在<http://mpeg.chiariglione.org/sites/default/files/files/standards/parts/docs/wl3411.zip>获得。

[0033] 在市场中存在各种“环绕声”基于信道的格式。它们的范围(例如)是从5.1家庭影院系统(其在使起居室享有立体声方面已获得最大成功)到NHK(日本广播协会或日本广播公司)所开发的22.2系统。内容创建者(例如,好莱坞工作室)将希望产生电影的声轨一次,而不花费精力来针对每一扬声器配置对其进行重混。近来,标准开发组织 (Standards Developing Organizations) 一直在考虑如下方式:提供到标准化位流中的编码,以及后续解码,其可调适且不知晓重放位置(涉及再现器)处的扬声器几何形状(和数目)以及声学条件。

[0034] 为向内容创建者提供此种灵活性,可使用分层要素集合来表示声场。所述阶层元素集合可指其中元素经排序以使得较低阶元素的基础集合提供模型化声场的完整表示的元素集合。在所述集合扩展以包含高阶元素时,所述表示变得更详细,从而增加分辨率。

[0035] 阶层元素集合的一个实例为球谐系数 (SHC) 的集合。以下表达式示范使用SHC对声场的描述或表示:

$$[0036] \quad p_i(t, r_r, \theta_r, \varphi_r) = \sum_{\omega=0}^{\infty} \left[ 4\pi \sum_{n=0}^{\infty} j_n(kr_r) \sum_{m=-n}^n A_n^m(k) Y_n^m(\theta_r, \varphi_r) \right] e^{j\omega t},$$

[0037] 所述表达式展示了在时间t处在声场的任一点 $\{r_r, \theta_r, \varphi_r\}$ 处的压力 $p_i$ 可以通过SHC,  $A_n^m(k)$ 唯一地表示。此处,  $k = \frac{\omega}{c}$ , c是声速( $\sim 343\text{m/s}$ ),  $\{r_r, \theta_r, \varphi_r\}$ 是参考点(或观察点),  $j_n$

( $\cdot$ ) 是阶数  $n$  的球面贝塞尔函数, 且  $Y_n^m(\theta_r, \varphi_r)$  是阶数  $n$  且子阶数  $m$  的球面谐波基底函数。可以认识到, 在方括号中的项是信号的频域表示 (即,  $S(\omega, r_r, \theta_r, \varphi_r)$ ), 其可由各种时频变换近似表示, 例如离散傅里叶变换 (DFT)、离散余弦变换 (DCT) 或小波变换。阶层集合的其它实例包含小波变换系数的集合和多分辨率基底函数的系数的其它集合。

[0038] 图1是说明从零阶 ( $n=0$ ) 到四阶 ( $n=4$ ) 的球谐基底函数的图。如可看出, 对于每一阶, 存在子阶  $m$  的扩展, 出于易于说明的目的, 在图1的实例中展示所述子阶但未明确注释。

[0039] SHC  $A_n^m(k)$  可由各种麦克风阵列配置物理获取 (例如, 记录), 或者, 其可从声场的基于信道或基于对象的描述导出。SHC 表示基于场景的音频, 其中 SHC 可输入到音频编码器以获得经编码 SHC, 所述经编码 SHC 可促成更有效的发射或存储。举例来说, 可使用涉及  $(1+4)^2$  个 (25, 且因此为第四阶) 系数的四阶表示。

[0040] 如上文所指出, 可使用麦克风阵列从麦克风记录导出 SHC。可如何从麦克风阵列导出 SHC 的各种实例描述于波莱蒂 • M (Poletti, M) 的 “基于球面谐波的三维环绕声系统 (Three-Dimensional Surround Sound Systems Based on Spherical Harmonics)” (听觉工程学协会会刊 (J. Audio Eng. Soc.), 第53卷, 第11期, 2005年11月, 第1004-1025页) 中。

[0041] 为说明可如何从基于对象的描述导出 SHC, 考虑以下等式。用于对应于个别音频对象的声场的系数  $A_n^m(k)$  可以表达为:

$$[0042] \quad A_n^m(k) = g(\omega)(-4\pi i k) h_n^{(2)}(kr_s) Y_n^{m*}(\theta_s, \varphi_s),$$

[0043] 其中  $i$  是  $\sqrt{-1}$ ,  $h_n^{(2)}(\cdot)$  是阶数  $n$  的球面汉克尔函数 (第二种), 且  $\{r_s, \theta_s, \varphi_s\}$  是对象的位置。知道对象源能量  $g(\omega)$  为频率的函数 (例如, 使用时频分析技术, 例如对 PCM 流执行快速傅立叶变换) 允许我们将每一 PCM 对象和对应位置转换为 SHC  $A_n^m(k)$ 。此外可展示 (因为以上是线性且正交分解) 用于每一对象的  $A_n^m(k)$  系数是加性的。以此方式, 大量 PCM 对象可以由  $A_n^m(k)$  系数表示 (例如, 作为用于个别对象的系数向量的总和)。基本上, 所述系数含有关于声场的信息 (随着 3D 坐标而变的压力), 且以上表示在观测点  $\{r_r, \theta_r, \varphi_r\}$  附近从个别对象到总体声场的表示的变换。下文在基于对象和基于 SHC 的音频译码的上下文中描述其余各图。

[0044] 图2是说明可执行本发明中所描述的技术的各种方面的系统 10 的图。如图2的实例中所示, 系统 10 包含内容创建者装置 12 和内容消费者装置 14。尽管在内容创建者装置 12 和内容消费者装置 14 的上下文中描述, 但可在其中声场的 SHC (其也可称为 HOA 系数) 或任何其它阶层表示经编码以形成表示音频数据的位流的任何上下文中实施所述技术。此外, 内容创建者装置 12 可表示能够实施本发明中所描述的技术的任何形式的计算装置, 包含手持机 (或蜂窝式电话)、平板计算机、智能电话或桌上型计算机 (提供几个实例)。同样, 内容消费者装置 14 可表示能够实施本发明中所描述的技术的任何形式的计算装置, 包含手持机 (或蜂窝式电话)、平板计算机、智能电话、机顶盒, 或桌上型计算机 (提供几个实例)。

[0045] 内容创建者装置 12 可由电影工作室或可产生多信道音频内容以供内容消费者装置 (例如, 内容消费者装置 14) 的操作者消耗的其它实体来操作。在一些实例中, 内容创建者装置 12 可由将希望压缩 HOA 系数 11 的个别用户操作。内容创建者通常产生音频内容与视频



内容。内容消费者装置14可由个人操作。内容消费者装置14可包含音频重放系统16,其可能能够再现SHC以供作为多信道音频内容重放的任何形式的音频重放系统。

[0046] 内容创建者装置12包含音频编辑系统18。内容创建者装置12获得各种格式(包含直接作为HOA系数)的实况记录7和音频对象9,内容创建者装置12可使用音频编辑系统18对其进行编辑。麦克风5可俘获实况记录7。内容创建者可在编辑过程期间再现来自音频对象9的HOA系数11,从而收听所再现的扬声器馈送以试图识别需要进一步编辑的声场的各个方面。内容创建者装置12可接着编辑HOA系数11(可能经由操纵以上文所描述的方式从中导出源HOA系数的音频对象9中的不同者而间接地编辑)。内容创建者装置12可采用音频编辑系统18来产生HOA系数11。音频编辑系统18表示能够编辑音频数据且输出所述音频数据作为一或多个源球谐系数的任何系统。

[0047] 当编辑过程完成时,内容创建者装置12可基于HOA系数11产生位流21。也就是说,内容创建者装置12包含音频编码装置20,其表示经配置以根据本发明中描述的技术的各个方面编码或以其它方式压缩HOA系数11以产生位流21的装置。音频编码装置20可产生位流21以供(作为一个实例)跨越发射信道发射,所述发射信道可为有线或无线信道、数据存储装置等。位流21可表示HOA系数11的经编码版本,且可包含主要位流和另一旁侧位流(其可称为旁侧信道信息)。

[0048] 虽然在图2中展示为直接发射到内容消费者装置14,但内容创建者装置12可将位流21输出到位于内容创建者装置12与内容消费者装置14之间的中间装置。中间装置可存储位流21以供稍后传递到可请求所述位流的内容消费者装置14。所述中间装置可包括文件服务器、网络服务器、桌上型计算机、膝上型计算机、平板计算机、移动电话、智能电话,或能够存储位流21以供音频解码器稍后检索的任何其它装置。中间装置可驻留在能够将位流21(且可能结合发射对应视频数据位流)串流到请求位流21的订户(例如,内容消费者装置14)的内容传递网络中。

[0049] 或者,内容创建者装置12可将位流21存储到存储媒体,例如压缩光盘、数字视频光盘、高清视频光盘或其它存储媒体,其中大多数能够由计算机读取且因此可称为计算机可读存储媒体或非暂时性计算机可读存储媒体。在此上下文中,发射信道可指代借以发射存储到媒体的内容的信道(且可包含零售商店及其它基于商店的传递机构)。在任何情况下,本发明的技术因此就此而言不应限于图2的实例。

[0050] 如图2的实例中进一步所示,内容消费者装置14包含音频重放系统16。音频重放系统16可表示能够重放多信道音频数据的任何音频重放系统。音频重放系统16可包含若干不同再现器22。再现器22可各自提供不同形式的再现,其中所述不同形式的再现可包含执行基于向量的振幅平移(VBAP)的各种方式中的一或多者,和/或执行声场合成的各种方式中的一或多者。如本文所使用,“A和/或B”意味着“A或B”,或“A和B”两者。

[0051] 音频重放系统16可进一步包含音频解码装置24。音频解码装置24可表示经配置以解码来自位流21的HOA系数11'的装置,其中HOA系数11'可类似于HOA系数11但归因于有损操作(例如,量化)和/或经由发射信道的发射而不同。音频重放系统16可随后解码位流21以获得HOA系数11'且再现HOA系数11'以输出扩音器馈送25。扩音器馈送25可驱动一或多个扩音器(其为便于说明的目的在图2的实例中未图示)。

[0052] 为了选择适当再现器或在一些情况下产生适当再现器,音频重放系统16可获得指

示扩音器的数目和/或扩音器的空间几何布置的扩音器信息13。在一些情况下,音频重放系统16可使用参考麦克风获得扩音器信息13且以动态地确定扩音器信息13的方式驱动所述扩音器。在其它情况下或结合动态确定扩音器信息13,音频重放系统16可提示用户与音频重放系统16介接并输入扩音器信息13。

[0053] 音频重放系统16可随后基于扩音器信息13选择音频再现器22中的一者。在一些情况下,在音频再现器22中无一者处于至扩音器信息中所指定的扩音器几何结构的某一阈值相似性量度(就扩音器几何结构而言)内时,音频重放系统16可基于扩音器信息13产生音频再现器22中的一者。音频重放系统16可在一些情况下基于扩音器信息13产生音频再现器22中的一者,而不首先尝试选择音频再现器22中的现有一者。一或多个扬声器3接着可重放经再现的扩音器馈送25。

[0054] 图3为更详细说明可执行本发明中描述的技术的各种方面的图2的实例中所示的音频编码装置20的一个实例的框图。音频编码装置20包含内容分析单元26、基于向量的分解单元27和基于方向的分解单元28。尽管下文简要描述,但关于音频编码装置20及压缩或以其它方式编码HOA系数的各种方面的更多信息可在2014年5月29日申请的标题为“用于声场的经分解表示的内插 (INTERPOLATION FOR DECOMPOSED REPRESENTATIONS OF A SOUND FIELD)”的第W0 2014/194099号国际专利申请公开案中获得。

[0055] 内容分析单元26表示经配置以分析HOA系数11的内容以识别HOA系数11表示从实况记录产生的内容还是从音频对象产生的内容的单元。内容分析单元26可确定HOA系数11是从实际声场的记录产生还是从人工音频对象产生。在一些情况下,当帧式HOA系数11是从记录产生时,内容分析单元26将HOA系数11传递到基于向量的分解单元27。在一些情况下,当帧式HOA系数11是从合成音频对象产生时,内容分析单元26将HOA系数11传递到基于方向的合成单元28。基于方向的合成单元28可表示经配置以执行HOA系数11的基于方向的合成以产生基于方向的位流21的单元。

[0056] 如在图3的实例中所展示,基于向量的分解单元27可包含线性可逆变换 (LIT) 单元30、参数计算单元32、重排序单元34、前景选择单元36、能量补偿单元38、心理声学音频译码器单元40、位流产生单元42、声场分析单元44、系数减少单元46、背景 (BG) 选择单元48、空间-时间内插单元50以及量化单元52。

[0057] 线性可逆变换 (LIT) 单元30接收呈HOA信道形式的HOA1111,每一信道表示与球面基底函数的给定阶数、子阶数相关联的系数的块或帧(其可表示为HOA[k],其中k可表示样本的当前帧或块)。HOA系数11的矩阵可具有维度 $D:M \times (N+1)^2$ 。

[0058] LIT单元30可表示经配置以执行被称作奇异值分解的分析形式的单元。虽然相对于SVD加以描述,但可相对于提供数组线性不相关的能量密集输出的任何类似变换或分解执行本发明中所描述的技术。此外,在本发明中对“集合”的参考一般希望指代“非零”集合(除非特定地相反陈述),且并不希望指代包含所谓的“空集合”的集合的经典数学定义。替代变换可包括常常被称作“PCA”的主分量分析。取决于上下文,PCA可由若干不同名称指代,例如离散卡亨南-洛维变换、霍特林变换、恰当正交分解 (POD) 和本征值分解 (EVD),仅举几例。有利于压缩音频数据的基本目标的此种操作的特性为多信道音频数据的“能量压缩”和“去相关”。

[0059] 在任何情况下,出于实例的目的,假定LIT单元30执行奇异值分解(其再次可被称

作“SVD”), LIT单元30可将HOA系数11变换成两个或更多个经变换HOA系数的集合。经变换HOA系数的“集合”可包含经变换HOA系数的向量。在图3的实例中, LIT单元30可相对于HOA系数11执行SVD以产生所谓的V矩阵、S矩阵及U矩阵。在线性代数中, SVD可按如下形式表示 $y$ 乘 $z$ 实数或复数矩阵 $X$  (其中 $X$ 可表示多信道音频数据, 例如HOA系数11) 的因子分解:

[0060]  $X=USV^*$

[0061]  $U$ 可表示 $y$ 乘 $y$ 实数或复数单位矩阵, 其中 $U$ 的 $y$ 列被称为多信道音频数据的左奇异向量。 $S$ 可表示在对角线上具有非负实数的 $y$ 乘 $z$ 矩形对角线矩阵, 其中 $S$ 的对角线值被称为多信道音频数据的奇异值。 $V^*$  (其可表示 $V$ 的共轭转置) 可表示 $z$ 乘 $z$ 实数或复数单位矩阵, 其中 $V^*$ 的 $z$ 列被称为多信道音频数据的右奇异向量。

[0062] 在一些实例中, 以上提及的SVD数学表达式中的 $V^*$ 矩阵表示为 $V$ 矩阵的共轭转置以反映SVD可应用于包括复数的矩阵。当应用于仅包括实数的矩阵时,  $V$ 矩阵的复数共轭 (或换句话说,  $V^*$ 矩阵) 可被视为 $V$ 矩阵的转置。下文中为容易说明的目的, 假定HOA系数11包括实数, 结果是经由SVD而非 $V^*$ 矩阵输出 $V$ 矩阵。此外, 尽管在本发明中表示为 $V$ 矩阵, 但对 $V$ 矩阵的提及应理解为在适当的情况下涉及 $V$ 矩阵的转置。尽管假定为 $V$ 矩阵, 但所述技术可以类似方式应用于具有复数系数的HOA系数11, 其中SVD的输出为 $V^*$ 矩阵。因此, 就此而言, 所述技术不应限于仅提供应用SVD以产生 $V$ 矩阵, 而是可包含将SVD应用于具有复数分量的HOA系数11以产生 $V^*$ 矩阵。

[0063] 以此方式, LIT单元30可相对于HOA系数11执行SVD以输出具有维度 $D:M \times (N+1)^2$ 的 $US[k]$ 向量33 (其可表示 $S$ 向量及 $U$ 向量的组合版本) 及具有维度 $D:(N+1)^2 \times (N+1)^2$ 的 $V[k]$ 向量35。 $US[k]$ 矩阵中的个别向量元素也可以被称为 $X_{PS}(k)$ , 而 $V[k]$ 矩阵的个别向量也可以被称为 $v(k)$ 。

[0064]  $U$ 、 $S$ 及 $V$ 矩阵的分析可显示, 这些矩阵携带或表示上文由 $X$ 表示的基本声场的空间及时间特征。 $U$  (长度为 $M$ 个样本) 中的 $N$ 个向量中的每一者可表示依据时间 (对于由 $M$ 个样本表示的时间周期) 的经正规化的分离音频信号, 其彼此正交且已与任何空间特性 (其也可被称作方向信息) 解耦。表示空间形状和位置 $(r, \theta, \phi)$ 的空间特性可实际上由 $V$ 矩阵 (长度 $(N+1)^2$ 的每一者) 中的个别第 $i$ 向量 $v^{(i)}(k)$ 表示。 $v^{(i)}(k)$ 向量中的每一者的个别元素可表示描述相关联音频对象的声场的形状 (包含宽度) 和位置的HOA系数。 $U$ 矩阵和 $V$ 矩阵中的向量均经正规化而使得其均方根能量等于单位。 $U$ 中的音频信号的能量因而由 $S$ 中的对角线元素表示。将 $U$ 与 $S$ 相乘以形成 $US[k]$  (具有个别向量元素 $X_{PS}(k)$ ), 因此表示具有能量的音频信号。SVD分解使音频时间信号 ( $U$ 中)、其能量 ( $S$ 中) 与其空间特性 ( $V$ 中) 解耦的能力可支持本发明中所描述的技术的各个方面。另外, 通过 $US[k]$ 与 $V[k]$ 的向量乘法合成基本HOA $[k]$ 系数 $X$ 的模型给出贯穿此文献使用的术语“基于向量的分解”。

[0065] 尽管描述为直接相对于HOA系数11执行, 但LIT单元30可将线性可逆变换应用到HOA系数11的导出项。举例来说, LIT单元30可相对于从HOA系数11导出的功率谱密度矩阵应用SVD。通过相对于HOA系数的功率谱密度 (PSD) 而非系数本身执行SVD, LIT单元30可在处理器循环及存储空间的一或多者方面可能地降低执行SVD的计算复杂性, 同时实现相同的源音频编码效率, 如同SVD是直接应用于HOA系数一般。

[0066] 参数计算单元32表示经配置以计算各种参数的单元, 所述参数例如相关性参数 ( $R$ )、方向性质参数 $(\theta, \phi, r)$ 及能量性质 ( $e$ )。用于当前帧的参数中的每一者可表示为 $R[k]$ 、 $\theta$

$[k]$ 、 $\phi[k]$ 、 $r[k]$ 及 $e[k]$ 。参数计算单元32可相对于 $US[k]$ 向量33执行能量分析和/或相关(或所谓的交叉相关)以识别这些参数。参数计算单元32还可确定先前帧的参数,其中先前帧参数可基于具有 $US[k-1]$ 向量及 $V[k-1]$ 向量的先前帧表示为 $R[k-1]$ 、 $\theta[k-1]$ 、 $\phi[k-1]$ 、 $r[k-1]$ 及 $e[k-1]$ 。参数计算单元32可将当前参数37和先前参数39输出到重排序单元34。

[0067] 由参数计算单元32计算的参数可供重排序单元34用以将音频对象重排序以表示其自然评估或随时间推移的连续性。重排序单元34可逐轮地将来自第一 $US[k]$ 向量33的参数37中的每一者与第二 $US[k-1]$ 向量33的参数39中的每一者相比较。重排序单元34可基于当前参数37及先前参数39对 $US[k]$ 矩阵33及 $V[k]$ 矩阵35内的各种向量重排序(作为一个实例,使用匈牙利算法)以输出经重排序的 $US[k]$ 矩阵33'(其可数学表示为 $\overline{US[k]}$ )及经重排序的 $V[k]$ 矩阵35'(其可数学表示为 $\overline{V[k]}$ )至前景声音(或优势的声音-PS)选择单元36(“前景选择单元36”)及能量补偿单元38。

[0068] 声场分析单元44可表示经配置以相对于H0A系数11执行声场分析以便有可能实现目标位速率41的单元。声场分析单元44可基于所述分析和/或基于所接收目标位速率41,确定心理声学译码器示例的总数目(其可为环境或背景信道的总数目( $BG_{TOT}$ )的函数)和前景信道(或换句话说,占优势信道)的数目。心理声学译码器示例的总数可表示为numHOATransportChannels。

[0069] 再次为了潜在地实现目标位速率41,声场分析单元44还可确定前景信道的总数( $nFG$ )45、背景(或换句话说,环境)声场的最小阶数( $N_{BG}$ 或替代地, $MinAmbH0Aorder$ )、表示背景声场的最小阶数的实际信道的对应数目( $nBGa = (MinAmbH0Aorder+1)^2$ ),及要发送的额外BG H0A信道的索引( $i$ )(其在图3的实例中可共同地表示为背景信道信息43)。背景信道信息42也可被称作环境信道信息43。保持来自numHOATransportChannels- $nBGa$ 的信道中的每一者可为“额外背景/环境信道”、“作用中的基于向量的占优势信道”、“作用中的基于方向的占优势信号”或“完全非作用”。在一个方面中,可通过两个位将信道类型指示为(如“ChannelType”)语法元素(例如,00:基于方向的信号;01:基于向量的占优势信号;10:额外环境信号;11:非作用信号)。可由 $(MinAmbH0Aorder+1)^2$ +索引10(在以上实例中)在用于所述帧的位流中作为信道类型出现的次数给出背景或环境信号的总数目 $nBGa$ 。

[0070] 声场分析单元44可基于目标位速率41选择背景(或换句话说,环境)信道的数目及前景(或换句话说,占优势)信道的数目,从而在目标位速率41相对较高时(例如,在目标位速率41等于或大于512Kbps时)选择更多背景及/或前景信道。在一个方面中,在位流的标头部分中,numHOATransportChannels可设定为8,而 $MinAmbH0Aorder$ 可设定为1。在此情境下,在每个帧处,四个信道可专用于表示声场的背景或环境部分,而其它4个信道可逐帧地在信道类型上变化--例如,用作额外背景/环境信道或前景/占优势信道。前景/占优势信号可为基于向量或基于方向的信号中的一者,如上文所描述。

[0071] 在一些情况下,用于帧的基于向量的占优势信号的总数目可通过ChannelType索引在所述帧的位流中是01的次数给出。在上述方面中,对于每个额外背景/环境信道(例如,对应于ChannelType 10),可在所述信道中表示可能的H0A系数中的哪一者的对应信息(超出前四)。对于四阶H0A内容,所述信息可为指示H0A系数5至25的索引。可在 $minAmbH0Aorder$ 设定为1时始终发送前四个环境H0A系数1至4,因此,音频编码装置可能仅需要指示额外环

境H0A系数中具有索引5至25中的一者。因而可使用5位语法元素(对于四阶内容)发送所述信息,其可表示为“CodedAmbCoeffIdx”。在任何情况下,声场分析单元44将背景信道信息43、US[k]向量33和V[k]向量35输出到基于向量的合成单元27B的一或多个其它组件,例如BG选择单元48B。

[0072] 背景选择单元48可表示经配置以基于背景信道信息(例如,背景声场( $N_{BG}$ )及要发送的额外BG H0A信道的数目( $n_{BGa}$ )及索引(i))确定背景或环境 $V_{BG}[k]$ 向量35 $_{BG}$ 的单元。举例来说,当 $N_{BG}$ 等于一时,背景选择单元48可将用于具有等于或小于一的阶的音频帧的每一样本的V[k]向量35选择为 $V_{BG}[k]$ 向量35 $_{BG}$ 。在此实例中,背景选择单元48可接着选择具有由索引(i)中的一者识别的索引的V[k]向量35作为额外 $V_{BG}[k]$ 向量35 $_{BG}$ ,其中将待于位流21中指定的 $n_{BGa}$ 提供到位流产生单元42以便使得音频解码装置(例如,图4的实例中所展示的音频解码装置24)能够从位流21剖析背景H0A系数47。背景选择单元48接着可将 $V_{BG}[k]$ 向量35 $_{BG}$ 输出到交叉淡化单元66的一或多个其它组件,例如能量补偿单元38。 $V_{BG}[k]$ 向量35 $_{BG}$ 可具有维度 $D: [(N_{BG}+1)^2 + n_{BGa}] \times (N+1)^2$ 。在一些实例中,背景选择单元48还可将US[k]向量33输出到交叉淡化单元66的一或多个其它组件,例如能量补偿单元38。

[0073] 能量补偿单元38可表示经配置以相对于 $V_{BG}[k]$ 向量35 $_{BG}$ 执行能量补偿以补偿由于背景选择单元48对V[k]向量35中的各种向量的移除所致的能量损失的单元。能量补偿单元38可相对于经重新排序的US[k]矩阵33’、经重新排序的V[k]矩阵35’、nFG信号49、前景V[k]向量51 $_k$ 和 $V_{BG}[k]$ 向量35 $_{BG}$ 中的一或多个者执行能量分析,且接着基于此能量分析执行能量补偿以产生经能量补偿的 $V_{BG}[k]$ 向量35 $_{BG}$ ’。能量补偿单元38可将经能量补偿 $V_{BG}[k]$ 向量35 $_{BG}$ ’输出到基于向量的合成单元27的一或多个其它组件,例如矩阵数学单元64。在一些实例中,能量补偿单元38还可将US[k]向量33输出到交叉淡化单元66的一或多个其它组件,例如矩阵数学单元64。

[0074] 矩阵数学单元64可表示经配置以对一或多个矩阵执行任何多种运算的单元。在图3的实例中,矩阵数学单元64可经配置以将US[k]向量33乘以经能量补偿 $V_{BG}[k]$ 向量35 $_{BG}$ ’以获得经能量补偿环境H0A系数47’。矩阵数学单元64可将所确定的经能量补偿环境H0A系数47’提供到基于向量的合成单元27的一或多个其它组件,例如交叉淡化单元66。经能量补偿环境H0A系数47’可具有维度 $D: M \times [(N_{BG}+1)^2 + n_{BGa}]$ 。

[0075] 交叉淡化单元66可表示经配置以执行信号之间的交叉淡化的单元。例如,交叉淡化单元66可在帧k的经能量补偿环境H0A系数47’与前一帧k-1的经能量补偿环境H0A系数47’之间进行交叉淡化以确定帧k的经交叉淡化的经能量补偿环境H0A系数47’。交叉淡化单元66可将确定的帧k的经交叉淡化的经能量补偿环境H0A系数47’输出到基于向量的合成单元27的一或多个其它组件,例如心理声学音频译码器单元40。

[0076] 在一些实例中,交叉淡化单元66可通过基于帧k-1的经能量补偿环境H0A系数47’的一部分修改帧k的经能量补偿环境H0A系数47’的一部分而在帧k的经能量补偿环境H0A系数47’与前一帧k-1的经能量补偿环境H0A系数47’之间进行交叉淡化。在一些实例中,交叉淡化单元66可在确定经交叉淡化的经能量补偿环境H0A系数47’时移除所述系数的一部分。下文参考图14提供交叉淡化单元66的额外细节。

[0077] 前景选择单元36可表示经配置以基于nFG 45(其可表示识别前景向量的一或多个索引)选择表示声场的前景或不同分量的经重排序的US[k]矩阵33’及经重排序的V[k]矩阵

35'的单元。前景选择单元36可将nFG信号49(其可表示为经重排序的 $US[k]_1, \dots, nFG49$ 、 $FG_1, \dots, nFG[k]49$ 或 $X_{PS}^{(1..nFG)}(k)49$ )输出到心理声学音频译码器单元40,其中nFG信号49可具有维度 $D:M \times nFG$ 且各自表示单音频对象。前景选择单元36还可将对应于声场的前景分量的经重排序的 $V[k]$ 矩阵35'(或 $v^{(1..nFG)}(k)35'$ )输出到空间-时间内插单元50,其中经重排序的 $V[k]$ 矩阵35'中的对应于前景分量的子集可表示为具有维度 $D:(N+1)^2 \times nFG$ 的前景 $V[k]$ 矩阵51<sub>k</sub>(其可在数学上表示为 $\bar{V}_{1, \dots, nFG}[k]$ )。

[0078] 空间-时间内插单元50可表示经配置以接收第k帧的前景 $V[k]$ 向量51<sub>k</sub>和前一帧(因此为k-1记法)的前景 $V[k-1]$ 向量51<sub>k-1</sub>且执行空间-时间内插以产生经内插前景 $V[k]$ 向量的单元。空间-时间内插单元50可将nFG信号49与前景 $V[k]$ 向量51<sub>k</sub>重新组合以恢复经重排序的前景HOA系数。空间-时间内插单元50可接着将经重排序的前景HOA系数除以经内插的 $V[k]$ 向量以产生经内插的nFG信号49'。空间-时间内插单元50还可输出前景 $V[k]$ 向量51<sub>k</sub>的用于产生经内插前景 $V[k]$ 向量的那些向量使得例如音频解码装置24等音频解码装置可产生经内插前景 $V[k]$ 向量且借此恢复前景 $V[k]$ 向量51<sub>k</sub>。将前景 $V[k]$ 向量51<sub>k</sub>的用以产生经内插的前景 $V[k]$ 向量的那些前景 $V[k]$ 向量51<sub>k</sub>表示为剩余前景 $V[k]$ 向量53。为了确保在编码器和解码器处使用相同的 $V[k]$ 和 $V[k-1]$ (以创建经内插的向量 $V[k]$ ) ,可在编码器和解码器处使用这些的经量化/经解量化的版本。

[0079] 就此而言,空间-时间内插单元50可表示从第一音频帧的一些其它部分和第二时间上在后或在前的音频帧内插第一音频帧的第一部分的单元。在一些实例中,所述部分可表示为子帧。在其它实例中,空间-时间内插单元50可相对于前一帧的某一最后数目的样本和后续帧的某一第一数目的样本操作。空间-时间内插单元50可在执行此内插时减小位流21中需要指定的前景 $V[k]$ 向量51<sub>k</sub>的样本的数目,因为前景 $V[k]$ 向量51<sub>k</sub>的仅那些用于产生经内插 $V[k]$ 向量的前景 $V[k]$ 向量51<sub>k</sub>表示前景 $V[k]$ 向量51<sub>k</sub>的子集。也就是说,为了潜在地使HOA系数11的压缩更有效(通过减小在位流21中指定的前景 $V[k]$ 向量51<sub>k</sub>的数目),本发明中所描述的技术的各个方面可提供第一音频帧的一或多个部分的内插,其中所述部分中的每一者可表示HOA系数11的经分解版本。

[0080] 空间-时间内插可导致数个益处。首先,归因于执行SVD或其它LIT的逐块性质,nFG信号49可不从帧到帧为连续的。换句话说,在LIT单元30逐帧应用SVD的条件下,所产生的经变换HOA系数中可存在特定不连续性,如例如 $US[k]$ 矩阵33和 $V[k]$ 矩阵35的无序性质所证明。通过执行此内插,在内插可具有潜在地减少归因于帧边界(或换句话说,HOA系数11分段为帧)而引入的任何假象的平滑效果的条件可减小不连续。使用前景 $V[k]$ 向量51<sub>k</sub>执行此内插且接着基于经内插前景 $V[k]$ 向量51<sub>k</sub>从所恢复的经重排序HOA系数产生经内插nFG信号49'可使归因于逐帧运算以及归因于对nFG信号49重排序的至少一些效果平滑化。

[0081] 在操作中,空间-时间内插单元50可内插来自包含于第一帧中的第一多个HOA系数11的一部分的第一分解(例如,前景 $V[k]$ 向量51<sub>k</sub>)和包含于第二帧中的第二多个HOA系数11的一部分的第二分解(例如,前景 $V[k]$ 向量51<sub>k-1</sub>)的第一音频帧的一或多个子帧以产生用于所述一或多个子帧的经分解的经内插球谐系数。

[0082] 在一些实例中,第一分解包括表示HOA系数11的所述部分的右奇异向量的第一前景 $V[k]$ 向量51<sub>k</sub>。同样,在一些实例中,第二分解包括表示HOA系数11的所述部分的右奇异向

量的第二前景 $V[k]$ 向量 $51_k$ 。

[0083] 换句话说,就球面上的正交基底函数而言,基于球谐的3D音频可为3D压力场的参数表示。所述表示的阶数 $N$ 越高,空间分辨率潜在地越高,且常常球谐(SH)系数的数目越大(总共 $(N+1)^2$ 个系数)。对于许多应用,可能需要系数的带宽压缩能够有效地发射且存储所述系数。在本发明中所针对的此技术可提供使用奇异值分解(SVD)的基于帧的维度减少过程。SVD分析可将系数的每一帧分解成三个矩阵 $U$ 、 $S$ 及 $V$ 。在一些实例中,所述技术可将 $US[k]$ 矩阵中的向量中的一些作为基础声场的方向分量处置。然而,当以此方式处置时,这些向量(在 $US[k]$ 矩阵中)在帧间是不连续的--即使其表示同一相异音频分量。当通过变换音频译码器馈入所述分量时,这些不连续性可导致显著假象。

[0084] 本发明中所描述的技术可解决此不连续性。即,所述技术可基于以下观测结果: $V$ 矩阵可经解译为球谐域中的正交空间轴。 $U[k]$ 矩阵可表示球谐(HOA)数据根据那些基底函数的投影,其中不连续性可归因于正交的空间轴( $V[k]$ ),所述空间轴线每帧改变且因此自身为不连续的。这不同于例如傅立叶变换的类似分解,其中基底函数在一些实例中在帧间将为常数。在这些术语中,SVD可认为是匹配追求算法。本发明中所描述的技术可使内插单元50能够通过在其间进行内插而在帧间维持基底函数( $V[k]$ )之间的连续性。

[0085] 如上文所指出,可相对于样本执行内插。当子帧包括样本的单个集合时,此情况在以上描述中得以一般化。在经由样本和经由子帧的内插的两个情况中,内插运算可呈以下等式的形式:

$$[0086] \quad \overline{v(l)} = w(l)v(k) + (1 - w(l))v(k - 1).$$

[0087] 在此上述等式中,内插可以相对于单个 $V$ 向量 $v(k)$ 从单个 $V$ 向量 $v(k-1)$ 执行,所述 $V$ 向量在一个实施例中可表示来自邻近帧 $k$ 和 $k-1$ 的 $V$ 向量。在上述等式中, $l$ 表示执行内插所针对的分辨率,其中 $l$ 可指示整数样本且 $l=1, \dots, T$ (其中 $T$ 为样本的长度,在所述长度内执行内插且在所述长度内需要输出的经内插的向量 $\overline{v(l)}$ 且所述长度还指示此过程的输出产生这些向量的 $l$ )。替代地, $l$ 可指示由多个样本组成的子帧。当(例如)将帧划分成四个子帧时, $l$ 可包括用于所述子帧中的每一者的值1、2、3和4。可经由位流将 $l$ 的值作为被称为“CodedSpatialInterpolationTime”的字段用信号表示,使得可在解码器中重复内插运算。 $w(l)$ 可包括内插权数的值。当内插为线性的时, $w(l)$ 可随着 $l$ 而变在0与1之间线性地且单调地变化。在其它实例中, $w(l)$ 可随着 $l$ 而变以非线性但单调方式(例如,升余弦的四分之一周期)在0与1之间变化。函数 $w(l)$ 可在几个不同的函数可能性之间标引且在位流中用信号表示为被称为“SpatialInterpolationMethod”的字段以使得解码器可重复相同的内插运算。当 $w(l)$ 是接近于0的值时,输出 $\overline{v(l)}$ 可由 $v(k-1)$ 较高地加权或影响。而当 $w(l)$ 是接近于1的值时,其确保输出 $\overline{v(l)}$ 由 $v(k-1)$ 较高地加权或影响。

[0088] 系数减少单元46可表示经配置以基于背景信道信息43相对于剩余前景 $V[k]$ 向量53执行系数减少以将经减少的前景 $V[k]$ 向量55输出到量化单元52的单元。经减少的前景 $V[k]$ 向量55可具有维度 $D: [(N+1)^2 - (N_{BG}+1)^2 - BG_{TOT}] \times n_{FG}$ 。系数减少单元46可在这方面中表示经配置以减少其余前景 $V[k]$ 向量53中的系数数目的单元。换句话说,系数减少单元46可表示经配置以消除(形成其余前景 $V[k]$ 向量53的)前景 $V[k]$ 向量中的具有极少至不具有方向信息的系数的单元。在一些实例中,相异或(换句话说)前景 $V[k]$ 向量的对应于一阶及零

阶基底函数的系数(其可表示为 $N_{BG}$ )提供极少方向信息,且因此可从前景V向量移除(经由可被称作“系数减少”的过程)。在此实例中,可提供较大灵活性以不仅从集合 $[(N_{BG}+1)^2+1, (N+1)^2]$ 识别对应于 $N_{BG}$ 的系数而且识别额外HOA信道(其可由变量TotalOfAddAmbHOAChan表示)。

[0089] 量化单元52可表示经配置以执行任何形式的量化以压缩减少的前景V[k]向量55以产生经译码前景V[k]向量57,从而将经译码前景V[k]向量57输出到位流产生单元42的单元。在操作中,量化单元52可表示经配置以压缩声场的空间分量(即,在此实例中经减少的前景V[k]向量55中的一或多者)的单元。量化单元52可执行如通过表示为“NbtsQ”的量化模式语法元素指示的以下12种量化模式中的任一者:

[0090]	NbtsQ值	量化模式的类型
[0091]	0-3:	保留
[0092]	4:	向量量化
[0093]	5:	无霍夫曼译码的标量量化
[0094]	6:	具有霍夫曼译码的6位标量量化
[0095]	7:	具有霍夫曼译码的7位标量量化
[0096]	8:	具有霍夫曼译码的8位标量量化
[0097]	...	...

[0098] 16:具有霍夫曼译码的16位标量量化

[0099] 量化单元52还可执行前述类型的量化模式中的任一者的预测版本,其中确定先前帧的V向量的元素(或执行向量量化时的权数)与当前帧的V向量的元素(或执行向量量化时的权数)之间的差。量化单元52可接着将当前帧与先前帧的元素或权数之间的差而非当前帧本身的V向量的元素的值量化。

[0100] 量化单元52可相对于减少的前景V[k]向量55中的每一者执行多种形式的量化,以获得减少的前景V[k]向量55的多个经译码版本。量化单元52可选择减少的前景V[k]向量55的经译码版本中的一者作为经译码前景V[k]向量57。换句话说,量化单元52可基于本发明中论述的标准的任何组合选择未经预测的经向量量化的V向量、经预测的经向量量化的V向量、未经霍夫曼译码的经标量量化的V向量及经霍夫曼译码的经标量量化的V向量中的一者,以用作经输出变换量化的V向量。在一些实例中,量化单元52可从包含向量量化模式及一或多个标量量化模式的一组量化模式中选择量化模式,且基于(或根据)所述选定模式量化输入V向量。量化单元52可接着将以下各者中的选定者提供到位流产生单元52以用作经译码前景V[k]向量57:未经预测的经向量量化的V向量(例如,就权数值或指示权数值的位而言)、经预测的经向量量化的V向量(例如,就误差值或指示误差值的位而言)、未经霍夫曼译码的经标量量化的V向量及经霍夫曼译码的经标量量化的V向量。量化单元52还可提供指示量化模式的语法元素(例如,NbtsQ语法元素)及用于将V向量解量化或以其它方式重构V向量的任何其它语法元素。

[0101] 音频编码装置20内包含的心理声学音频译码器单元40可表示心理声学音频译码器的多个例子,其每一者用于编码经能量补偿环境HOA系数47'和经内插nFG信号49'中的每一者的不同音频对象或HOA信道以产生经编码环境HOA系数59和经编码nFG信号61。心理声学音频译码器单元40可将经编码环境HOA系数59和经编码nFG信号61输出到位流产生单元



42。

[0102] 包括于音频编码装置20内的位流产生单元42表示将数据格式化以符合已知格式(其可指代解码装置已知的格式)借此产生基于向量的位流21的单元。换句话说,位流21可表示以上文所描述的方式编码的经编码音频数据。位流产生单元42在一些实例中可表示多路复用器,其可接收经译码前景 $V[k]$ 向量57、经编码环境HOA系数59、经编码nFG信号61,和背景信道信息43。位流产生单元42可接着基于经译码前景 $V[k]$ 向量57、经编码环境HOA系数59、经编码nFG信号61及背景信道信息43产生位流21。以此方式,位流产生单元42可由此在位流21中指定向量57以获得如下文关于图7的实例更详细描述有位流21。位流21可包含主要或主位流和一或多个旁侧信道位流。

[0103] 尽管在图3的实例中未展示,但音频编码装置20还可包含位流输出单元,所述位流输出单元基于当前帧将使用基于方向的合成还是基于向量的合成编码而切换从音频编码装置20输出的位流(例如,在基于方向的位流21与基于向量的位流21之间切换)。位流输出单元可基于由内容分析单元26输出的指示执行基于方向的合成(作为检测到HOA系数11是从合成音频对象产生的结果)还是执行基于向量的合成(作为检测到HOA系数经记录的结果)的语法元素执行所述切换。位流输出单元可指定正确的标头语法以指示用于当前帧以及位流21中的相应一者的切换或当前编码。

[0104] 此外,如上文所提及,声场分析单元44可识别 $BG_{TOT}$ 环境HOA系数47,所述 $BG_{TOT}$ 环境HOA系数可逐帧改变(但时常 $BG_{TOT}$ 可跨越两个或更多个邻近(在时间上)帧保持恒定或相同)。 $BG_{TOT}$ 的改变可导致在减少的前景 $V[k]$ 向量55中表达的系数的改变。 $BG_{TOT}$ 的改变可导致背景HOA系数(其也可被称作“环境HOA系数”),所述背景HOA系数逐帧改变(但再次,时常 $BG_{TOT}$ 可跨越两个或更多个邻近(在时间上)帧保持恒定或相同)。所述改变常常导致声场的各方面的能量改变,所述声场通过额外环境HOA系数的添加或移除及系数从减少的前景 $V[k]$ 向量55的对应移除或系数至减少的前景 $V[k]$ 向量55的添加表示。

[0105] 因此,声场分析单元44可进一步确定环境HOA系数何时从帧至帧改变且产生指示环境HOA系数的改变的旗标或其它语法元素(就用于表示声场的环境分量而言)(其中所述改变也可被称作环境HOA系数的“转变”或环境HOA系数的“转变”)。具体而言,系数减少单元46可产生旗标(其可表示为AmbCoeffTransition旗标或AmbCoeffIdxTransition旗标),从而将所述旗标提供到位流产生单元42,使得可将所述旗标包含在位流21中(有可能作为旁侧信道信息的一部分)。

[0106] 除指定环境系数转变旗标之外,系数减少单元46也可修改产生减少的前景 $V[k]$ 向量55的方式。在一个实例中,在确定环境HOA环境系数中的一者在当前帧期间处于转变中时,系数减少单元46可指定用于减少的前景 $V[k]$ 向量55的 $V$ 向量中的每一者的向量系数(其也可被称作“向量元素”或“元素”),其对应于处于转变中的环境HOA系数。同样,处于转变中的环境HOA系数可添加至背景系数的 $BG_{TOT}$ 总数目或从背景系数的 $BG_{TOT}$ 总数目中移除。因此,背景系数的总数目的所得改变影响环境HOA系数是否包含于位流中,及在上文所描述的第二及第三配置模式中是否针对位流中所指定的 $V$ 向量包含 $V$ 向量的对应元素。关于系数减少单元46可如何指定减少的前景 $V[k]$ 向量55以克服能量的改变的更多信息提供于2015年1月12日申请的标题为“环境较高阶立体混响系数的转变(TRANSITIONING OF AMBIENT HIGHER\_ORDER AMBISONIC COEFFICIENTS)”的第14/594,533号美国申请案中。

[0107] 图14是更详细说明图3的实例中所示的音频编码装置20的交叉淡化单元66的框图。交叉淡化单元66可包含混合器单元70、成帧单元71和延迟单元72。图14说明交叉淡化单元66的仅一个实例,且其它配置是可能的。例如,成帧单元71可定位于混合器单元70之前以使得在经能量补偿环境HOA系数47'由混合器单元70接收之前移除第三部分75。

[0108] 混合器单元70可表示经配置以将多个信号组合为单个信号的单元。举例来说,混合器单元70可组合第一信号与第二信号以产生经修改信号。混合器单元70可通过使第一信号淡入同时使第二信号淡出来组合第一信号与第二信号。混合器单元70可应用任何多种函数来使所述部分淡入和淡出。作为一个实例,混合器单元70可应用线性函数以使第一信号淡入且应用线性函数以使第二信号淡出。作为另一实例,混合器单元70可应用指数函数以使第一信号淡入且应用指数函数以使第二信号淡出。在一些实例中,混合器单元70可将不同函数应用于信号。举例来说,混合器单元70可应用线性函数以使第一信号淡入且应用指数函数以使第二信号淡出。在一些实例中,混合器单元70可通过使信号的一部分淡入或淡出来使信号淡入或淡出。在任何情况下,混合器单元可将经修改信号输出到交叉淡化单元66的一或多个其它组件,例如成帧单元71。

[0109] 成帧单元71可表示经配置以使输入信号成帧以配合一或多个特定尺寸的单元。在例如其中输入信号的尺寸中的一或多者大于特定尺寸中的一或多者的一些实例中,成帧单元71可通过移除输入信号的一部分而产生成帧输出信号,例如超过特定尺寸的部分。举例来说,在特定尺寸是1024乘4且输入信号具有1280乘4的尺寸的情况下,成帧单元71可通过移除输入信号的256乘4部分而产生成帧输出信号。在一些实例中,成帧单元71可将成帧输出信号输出到音频编码装置20的一或多个其它组件,例如图3的心理声学音频译码器单元40。在一些实例中,成帧单元71可将输入信号的经移除部分输出到交叉淡化单元66的一或多个其它组件,例如延迟单元72。

[0110] 延迟单元72可表示经配置以存储信号用于稍后使用的单元。举例来说,延迟单元72可经配置以在第一时间存储第一信号且在第二稍后时间输出所述第一信号。以此方式,延迟单元72可作为先进先出(FIFO)缓冲器操作。延迟单元72可在所述第二稍后时间将所述第一信号输出到交叉淡化单元66的一或多个其它组件,例如混合器单元70。

[0111] 如上文所论述,交叉淡化单元66可接收当前帧(例如,帧k)的经能量补偿环境HOA系数47',使当前帧的经能量补偿环境HOA系数47'与前一帧的经能量补偿环境HOA系数47'进行交叉淡化,且输出经交叉淡化的经能量补偿环境HOA系数47'。如图14中所说明,经能量补偿环境HOA系数47'可包含第一部分73、第二部分74和第三部分75。

[0112] 根据本发明的一或多种技术,交叉淡化单元66的混合器单元可组合(例如,在其间交叉淡化)当前帧的经能量补偿环境HOA系数47'的第一部分73与前一帧的经能量补偿环境HOA系数47'的第三部分76以产生中间经交叉淡化的经能量补偿环境HOA系数77。混合器单元70可将所产生中间经交叉淡化的经能量补偿环境HOA系数77输出到成帧单元71。由于在此实例中混合器单元70利用前一帧的经能量补偿环境HOA系数47'的第三部分76,因此可假定交叉淡化单元66在处理当前帧之前是在操作中。因此,与分开地使当前帧的US矩阵与前一帧的US矩阵进行交叉淡化且使当前帧的V矩阵与前一帧的V矩阵进行交叉淡化相反,混合器单元70可在能量补偿域中进行交叉淡化。以此方式,根据本发明的技术可减少交叉淡化单元66的计算负载、电力消耗和/或复杂性。

[0113] 成帧单元71可通过在中间经交叉淡化的能量补偿环境HOA系数77的尺寸超过当前帧的尺寸的情况下从中间经交叉淡化的经能量补偿环境HOA系数77移除第三部分75而确定经交叉淡化的经能量补偿环境HOA系数47”。举例来说,在当前帧的尺寸是1024乘4且中间经交叉淡化的经能量补偿环境HOA系数77的尺寸是1280乘4的情况下,成帧单元71可通过从中间经交叉淡化的经能量补偿环境HOA系数77移除第三部分75(例如,256乘4部分)而确定经交叉淡化的经能量补偿环境HOA系数47”。成帧单元71可将第三部分75输出到延迟单元72用于未来使用(例如,由混合器单元70在交叉淡化后续帧的经能量补偿环境HOA系数47’时)。成帧单元71可将所确定经交叉淡化的经能量补偿环境HOA系数47”输出到图3的心理声学音频译码器单元40。以此方式,交叉淡化单元66可使前一帧与当前帧之间的转变平滑。

[0114] 在一些实例中,交叉淡化单元66可在HOA系数任何两个集合之间进行交叉淡化。作为一个实例,交叉淡化单元66可在HOA系数的第一集合与HOA系数的第二集合之间进行交叉淡化。作为另一实例,交叉淡化单元66可在HOA系数的当前集合与HOA系数的先前集合之间进行交叉淡化。

[0115] 图4为更详细说明图2的音频解码装置24的框图。如在图4的实例中所展示,音频解码装置24可包含提取单元72、基于方向性的重构单元90及基于向量的重构单元92。尽管下文进行描述,但关于音频解码装置24及解压缩或以其它方式解码HOA系数的各种方面的更多信息可在2014年5月29日申请的标题为“用于声场的经分解表示的内插 (INTERPOLATION FOR DECOMPOSED REPRESENTATIONS OF A SOUND FIELD)”的第WO 2014/194099号国际专利申请公开案中获取。

[0116] 提取单元72可表示经配置以接收位流21和提取HOA系数11的各种经编码版本(例如,基于方向的经编码版本或基于向量的经编码版本)的单元。提取单元72可根据以上所述确定指示HOA系数11是否经由各种基于方向的或基于向量的版本而经编码的语法元素。当执行基于方向的编码时,提取单元72可提取HOA系数11的基于方向的版本及与所述经编码版本相关联的语法元素(其在图4的实例中表示为基于方向的信息91),从而将所述基于方向的信息91传递到基于方向的重构单元90。基于方向的重构单元90可表示经配置以基于所述基于方向的信息91重构呈HOA系数11’的形式的HOA系数的单元。

[0117] 当语法元素指示HOA系数11使用基于向量的合成进行编码时,提取单元72可提取经译码前景 $V[k]$ 向量57(其可包含经译码权数及/或索引63或经标量量化的 $V$ 向量)、经编码环境HOA系数59及对应音频对象61(其也可被称作经编码nFG信号61)。音频对象61各自对应于向量57中的一者。提取单元72可将经译码前景 $V[k]$ 向量57传递至 $V$ 向量重构单元74,且将经编码环境HOA系数59以及经编码nFG信号61提供到心理声学解码单元80。

[0118]  $V$ 向量重构单元74可表示经配置以从经编码前景 $V[k]$ 向量57重构 $V$ 向量的单元。 $V$ 向量重构单元74可以与量化单元52互逆的方式操作。

[0119] 心理声学解码单元80可以与图3的实例中所展示的心理声学音频译码器单元40互逆的方式操作以便对经编码环境HOA系数59及经编码nFG信号61进行解码且由此产生经能量补偿的环境HOA系数47’及经内插的nFG信号49’(其也可被称作经内插的nFG音频对象49’)。心理声学解码单元80可将能量补偿环境HOA系数47’传递到淡化单元770且将nFG信号49’传递到前景制订单元78。

[0120] 空间-时间内插单元76可以类似于上文相对于空间-时间内插单元50所描述方式

的方式操作。空间-时间内插单元76可接收减少的前景 $V[k]$ 向量 $55_k$ 且相对于前景 $V[k]$ 向量 $55_k$ 及减少的前景 $V[k-1]$ 向量 $55_{k-1}$ 执行空间-时间内插以产生经内插的前景 $V[k]$ 向量 $55_k$ ”。空间-时间内插单元76可将经内插的前景 $V[k]$ 向量 $55_k$ ”转发到淡化单元770。

[0121] 提取单元72还可将指示环境HOA系数中的一者何时处于转变中的信号757输出到淡化单元770,所述淡化单元可接着确定 $SHC_{BG} 47'$  (其中 $SHC_{BG} 47'$ 还可表示为“环境HOA信道 $47'$ ”或“环境HOA系数 $47'$ ”)及经内插的前景 $V[k]$ 向量 $55_k$ ”的元素中的哪一者将淡入或淡出。在一些实例中,淡化单元770可相对于环境HOA系数 $47'$ 及经内插的前景 $V[k]$ 向量 $55_k$ ”的元素中的每一者相反地操作。也就是说,淡化单元770可相对于环境HOA系数 $47'$ 中的对应环境HOA系数执行淡入或淡出或执行淡入或淡出两者,同时相对于经内插的前景 $V[k]$ 向量 $55_k$ ”的元素中的对应一个元素执行淡入或淡出或执行淡入与淡出两者。淡化单元770可将经调整的环境HOA系数 $47'$ 输出到HOA系数制订单元82且将经调整的前景 $V[k]$ 向量 $55_k$ ”输出到前景制订单元78。在这方面中,淡化单元770表示经配置以相对于HOA系数或其衍生物(例如,呈环境HOA系数 $47'$ 以及经内插前景 $V[k]$ 向量 $55_k$ ”的元素的形式)的各种方面执行淡化操作的单元。

[0122] 前景制订单元78可表示经配置以相对于经调整的前景 $V[k]$ 向量 $55_k$ ”及经内插的nFG信号 $49'$ 执行矩阵乘法以产生前景HOA系数65的单元。在这方面中,前景制订单元78可组合音频对象 $49'$  (其为借以表示经内插的nFG信号 $49'$ 的另一种方式)与向量 $55_k$ ”以重构HOA系数 $11'$ 的前景(或换句话说,优势)方面。前景制订单元78可执行经内插的nFG信号 $49'$ 乘以经调整的前景 $V[k]$ 向量 $55_k$ ”的矩阵乘法。

[0123] HOA系数制订单元82可表示经配置以将前景HOA系数65组合到经调整的环境HOA系数 $47'$ 以便获得HOA系数 $11'$ 的单元。撇号记法反映HOA系数 $11'$ 可类似于HOA系数11但与HOA系数11不相同。HOA系数11与 $11'$ 之间的差可起因于由于有损发射媒体上的发射、量化或其它有损操作而产生的损失。

[0124] 图5是说明音频编码装置(例如图3的实例中所示的音频编码装置20)执行本发明中所描述的基于向量的合成技术的各个方面的示范性操作的流程图。最初,音频编码装置20接收HOA系数11 (106)。音频编码装置20可调用LIT单元30,其可相对于HOA系数应用LIT以输出经变换HOA系数(例如,在SVD的情况下,经变换HOA系数可包括 $US[k]$ 向量33和 $V[k]$ 向量35) (107)。

[0125] 音频编码装置20接下来可调用参数计算单元32以按上文所描述的方式相对于 $US[k]$ 向量33、 $US[k-1]$ 向量33、 $V[k]$ 和/或 $V[k-1]$ 向量35的任何组合执行上文所描述的分析以识别各种参数。也就是说,参数计算单元32可基于经变换的HOA系数33/35的分析确定至少一个参数(108)。

[0126] 音频编码装置20可接着调用重排序单元34,重排序单元34基于参数将经变换的HOA系数(再次在SVD的上下文中,其可指 $US[k]$ 向量33及 $V[k]$ 向量35)重排序以产生经重排序的经变换的HOA系数 $33'/35'$  (或,换句话说, $US[k]$ 向量 $33'$ 及 $V[k]$ 向量 $35'$ ),如上文所描述(109)。音频编码装置20可在以上操作或后续操作中的任一者期间还调用声场分析单元44。声场分析单元44可如上文所描述相对于HOA系数11和/或经变换HOA系数33/35执行声场分析以确定前景信道(nFG) 45的总数、背景声场( $N_{BG}$ )的次序以及待发送的额外BG HOA信道的数目( $nBGa$ )和指数( $i$ ) (其可在图3的实例中统一表示为背景信道信息43) (109)。

[0127] 音频编码装置20还可调用背景选择单元48。背景选择单元48可基于背景信道信息43确定背景或环境H0A系数47(110)。音频编码装置20可进一步调用前景选择单元36,前景选择单元36可基于nFG 45(其可表示识别前景向量的一或多个索引)选择表示声场的前景或相异分量的经重排序的US[k]向量33'及经重排序的V[k]向量35'(112)。

[0128] 音频编码装置20可调用能量补偿单元38。能量补偿单元38可相对于环境H0A系数47执行能量补偿以补偿归因于由背景选择单元48移除H0A系数的各者而导致的能量损失,且以上文所描述的方式交叉淡化经能量补偿的环境H0A系数47'(114)。

[0129] 音频编码装置20也可调用空间-时间内插单元50。空间-时间内插单元50可相对于经重排序的经变换H0A系数33'/35'执行空间-时间内插以获得经内插前景信号49'(其也可被称作“经内插nFG信号49'”)及剩余前景方向信息53(其也可被称作“V[k]向量53’”(116)。音频编码装置20可接着调用系数减少单元46。系数减少单元46可基于背景信道信息43执行相对于剩余前景V[k]向量53的系数减少,以获得减少的前景方向信息55(其也可被称作减少的前景V[k]向量55)(118)。

[0130] 音频编码装置20可接着调用量化单元52以按上文所描述的方式压缩经缩减前景V[k]向量55且产生经译码前景V[k]向量57(120)。

[0131] 音频编码装置20还可调用心理声学音频译码器单元40。心理声学音频译码器单元40可对经能量补偿的环境H0A系数47'和经内插nFG信号49'的每一向量进行心理声学译码以产生经编码环境H0A系数59和经编码nFG信号61。音频编码装置接着可调用位流产生单元42。位流产生单元42可基于经译码前景方向信息57、经译码环境H0A系数59、经译码nFG信号61和背景信道信息43而产生位流21。

[0132] 图6是说明音频解码装置(例如图4的实例中所示的音频解码装置24)在执行本发明中描述的技术的各个方面中的示范性操作的流程图。最初,音频解码装置24可接收位流21(130)。在接收到位流后,音频解码装置24可即刻调用提取单元72。出于论述的目的假定位流21指示将执行基于向量的重构,提取单元72可剖析位流以检索上文所提及的信息,将此信息传递到基于向量的重构单元92。

[0133] 换句话说,提取单元72可按上文所描述的方式从位流21中提取经译码前景方向信息57(再次,其也可被称作经译码前景V[k]向量57)、经译码环境H0A系数59和经译码前景信号(其也可被称作经译码前景nFG信号59或经译码前景音频对象59)(132)。

[0134] 音频解码装置24可进一步调用解量化单元74。解量化单元74可对经译码前景方向信息57进行熵解码及解量化以获得经减少的前景方向信息55<sub>k</sub>(136)。音频解码装置24还可调用心理声学解码单元80。心理声学音频译码单元80可解码经编码环境H0A系数59及经编码前景信号61以获得经能量补偿的环境H0A系数47'及经内插的前景信号49'(138)。心理声学解码单元80可将能量补偿环境H0A系数47'传递到淡化单元770且将nFG信号49'传递到前景制订单元78。

[0135] 音频解码装置24接下来可调用空间-时间内插单元76。空间-时间内插单元76可接收经重排序的前景方向信息55<sub>k</sub>'且相对于经减少的前景方向信息55<sub>k</sub>/55<sub>k-1</sub>执行空间-时间内插以产生经内插的前景方向信息55<sub>k</sub>”(140)。空间-时间内插单元76可将经内插的前景V[k]向量55<sub>k</sub>”转发到淡化单元770。

[0136] 音频解码装置24可调用淡化单元770。淡化单元770可(例如,从提取单元72)接收

或以其它方式获得指示经能量补偿环境HOA系数47'何时在转变中的语法元素(例如,AmbCoeffTransition语法元素)。淡化单元770可基于转变语法元素和维持的转变状态信息而淡入或淡出能量补偿环境HOA系数47',从而将经调整环境HOA系数47"输出到HOA系数制订单元82。淡化单元770还可基于语法元素和维持的转变状态信息而淡出或淡入经内插前景V[k]向量55k"的对应一或多个元素,从而将经调整前景V[k]向量55k"输出到前景制订单元78(142)。

[0137] 音频解码装置24可调用前景制订单元78。前景制订单元78可执行nFG信号49'与经调整前景方向信息55k"的矩阵乘法以获得前景HOA系数65(144)。音频解码装置24还可调用HOA系数制订单元82。HOA系数制订单元82可将前景HOA系数65与经调整环境HOA系数47"相加以便获得HOA系数11'(146)。

[0138] 图7是说明图2到4的实例中所示的位流21的部分250的图。图7的实例中所示的部分250可被称为位流21的HOAConfig部分250,且包含HOAOrder字段、MinAmbHoaOrder字段、方向信息字段253、CodedSpatialInterpolationTime字段254、SpatialInterpolationMethod字段255、CodedVVecLength字段256以及增益信息字段257。如图7的实例中所示,CodedSpatialInterpolationTime字段254可包括三位字段,SpatialInterpolationMethod字段255可包括一位字段,且CodedVVecLength字段256可包括两位字段。

[0139] 部分250还包含SingleLayer字段240和FrameLengthFactor字段242。SingleLayer字段240可表示指示是否使用多个层来表示HOA系数的经译码版本或者是否使用单个层来表示HOA coefficients的经译码版本的一或多个位。FramelengthFactor字段242表示指示帧长度因数的一或多个位,其在下文相对于图12更详细论述。

[0140] 图8是说明根据本发明中所描述的技术的各种方面指定的实例帧249S及249T的图。在图8的实例中,帧249S和249T各自包含四个输送信道275A到275D。输送信道275A包含指示ChannelSideInfoData 154A和HOAGainCorrectionData的标头位。输送信道275A还包含指示VVectorData 156A的有效负载位。输送信道275B包含指示ChannelSideInfoData 154B和HOAGainCorrectionData的标头位。输送信道275B还包含指示VVectorData 156B的有效负载位。输送信道275C和275D不用于帧249S。帧275T在输送信道275A到275D方面大体上类似于帧249S。

[0141] 图9是说明根据本文所描述的技术的至少一个位流的一或多个信道的实例帧的图。位流450包含帧810A到810H,其可各自包含一或多个信道。位流450可为图9的实例中所示的位流21的一个实例。在图9的实例中,音频解码装置24维持状态信息,更新所述状态信息以确定如何解码当前帧k。音频解码装置24可利用来自配置814的状态信息以及帧810B到810D。

[0142] 换句话说,音频编码装置20可在位流产生单元42内包含例如维持用于对帧810A到810E中的每一者进行编码的状态信息的状态机402,因为位流产生单元42可基于状态机402指定用于帧810A到810E中的每一者的语法元素。

[0143] 音频解码装置24可同样在位流提取单元72内包含例如类似的状态机402,其基于状态机402而输出语法元素(其中一些未在位流21中显式地指定)。音频解码装置24的状态机402可以类似于音频编码装置20的状态机402的方式操作。因此,音频解码装置24的状态

机402可维持状态信息,基于配置814更新所述状态信息,以及在图9的实例中帧810B到810D的解码。基于所述状态信息,位流提取单元72可基于由状态机402维持的状态信息而提取帧810E。所述状态信息可提供若干隐式语法元素,当对帧810E的各种输送信道进行解码时音频编码装置20可利用所述语法元素。

[0144] 图10说明用于获得如本文所描述的空间-时间内插的技术的表示。图3的实例中所示的音频编码装置20的空间-时间内插单元50可执行下文更详细地描述的空间-时间内插。空间-时间内插可包含在空间和时间维度两者中获得较高分辨率空间分量。空间分量可基于由较高阶立体混响(HOA)系数(或HOA系数也可以称为“球谐系数”)构成的多维信号的正交分解。

[0145] 在所说明的曲线图中,向量 $V_1$ 和 $V_2$ 表示多维信号的两个不同空间分量的对应向量。空间分量可通过多维信号的逐块分解而获得。在一些实例中,空间分量是通过相对于较高阶立体混响(HOA)音频数据(其中此立体混响音频数据包含块、样本或任何其它形式的多信道音频数据)的每一块(其可指代帧)执行逐块形式的SVD而得到。变量 $M$ 可用以表示音频帧的长度(以样本数计)。

[0146] 因此, $V_1$ 和 $V_2$ 可表示用于HOA系数11的顺序块的前景 $V[k]$ 向量 $51_k$ 和前景 $V[k-1]$ 向量 $51_{k-1}$ 的对应向量。 $V_1$ 可例如表示第一帧( $k-1$ )的前景 $V[k-1]$ 向量 $51_{k-1}$ 的第一向量,而 $V_2$ 可表示第二且后续帧( $k$ )的前景 $V[k]$ 向量 $51_k$ 的第一向量。 $V_1$ 及 $V_2$ 可表示包含在多维信号中的单个音频对象的空间分量。

[0147] 用于每一 $x$ 的经内插向量 $V_x$ 是通过根据多维信号(经内插向量 $V_x$ 可应用于所述多维信号以平滑化时间(且因此,在一些情况下,空间)分量)的时间分量的时间片段或“时间样本”的数目 $x$ 对 $V_1$ 及 $V_2$ 进行加权而获得。如上文所描述,采用SVD组成,可通过对每一时间样本向量(例如,HOA系数11的样本)与对应经内插 $V_x$ 进行向量除法来获得nFG信号49的平滑。即, $US[n] = HOA[n] * V_x[n]^{-1}$ ,其中此表示行向量乘以列向量,因而产生US的标量元素。 $V_x[n]^{-1}$ 可作为 $V_x[n]$ 的伪逆而获得。

[0148] 相对于 $V_1$ 及 $V_2$ 的加权,归因于在时间上在 $V_1$ 之后出现的 $V_2$ , $V_1$ 的权数沿着时间维度按比例较低。即,尽管前景 $V[k-1]$ 向量 $51_{k-1}$ 为分解的空间分量,但时间上连续的前景 $V[k]$ 向量 $51_k$ 随时间推移表示空间分量的不同值。因此, $V_1$ 的权数减小,而 $V_2$ 的权数随着 $x$ 沿着 $t$ 增大而增长。此处, $d_1$ 及 $d_2$ 表示权数。

[0149] 图11是说明根据本文所述的技术的用于多维信号的依序SVD块的人工US矩阵( $US_1$ 及 $US_2$ )的框图。经内插 $V$ 向量可应用于人工US矩阵的行向量以恢复原始多维信号。更具体来说,空间-时间内插单元50可将经内插前景 $V[k]$ 向量 $53$ 的伪逆乘以nFG信号49与前景 $V[k]$ 向量 $51_k$ (其可表示为前景HOA系数)的相乘结果以获得 $K/2$ 经内插样本,其可代替nFG信号的 $K/2$ 样本用作第一 $K/2$ 样本,如 $U_2$ 矩阵的图11的实例中所示。

[0150] 图12是说明根据本发明中所描述的技术使用奇异值分解及空间时间分量的平滑来分解较高阶立体混响(HOA)信号的后续帧的框图。帧 $n-1$ 及帧 $n$ (其也可以表示为帧 $n$ 及帧 $n+1$ )表示时间上连续的帧,其中每一帧包括1024个时间片段且具有HOA阶数4,从而得出 $(4+1)^2=25$ 个系数。可通过如所说明应用经内插 $V$ 向量而获得为帧 $n-1$ 及帧 $n$ 处的经人工平滑的U矩阵的US矩阵。每一灰色行或列向量表示一个音频对象。

[0151] 计算作用中基于向量的信号的HOA表示

[0152] 通过取表示于XVEck中的基于向量的信号中的每一者且将其与其对应(经解量化)空间向量VVEck相乘而产生瞬时CVEck。每一VVEck表示于MVEck中。因而,对于N阶HOA信号及M个基于向量的信号,将存在M个基于向量的信号,其中的每一者将具有由帧长度P给出的维度。这些信号可因此表示为:XVEck<sub>mn</sub>,  $n=0, \dots, P-1$ ;  $m=0, \dots, M-1$ 。对应地,将存在M个空间向量,维度 $(N+1)^2$ 的VVEck。这些可表示为MVEck<sub>ml</sub>,  $l=0, \dots, (N+1)^2-1$ ;  $m=0, \dots, M-1$ 。每一基于向量的信号的HOA表示CVEck<sub>m</sub>是如下给出的矩阵向量乘法:

[0153]  $\text{CVEck}_m = (\text{XVEck}_m (\text{MVEck}_m)^T)^T$

[0154] 其产生 $(N+1)^2$ 乘P的矩阵。通过将每一基于向量的信号的贡献求和而给出完整HOA表示如下:

[0155]  $\text{CVEck} = \sum_{m=0}^{M-1} \text{CVEck}_m$

[0156] V向量的空间-时间内插

[0157] 然而,为了维持平滑的空间-时间连续性,仅对于帧长度的部分P-B进行以上计算。改为通过使用从当前MVEck<sub>m</sub>及先前值MVEck<sub>m-1</sub>导出的经内插集合MVEck<sub>ml</sub> ( $m=0, \dots, M-1$ ;  $l=0, \dots, (N+1)^2$ ) 进行HOA矩阵的前B个样本。此导致较高时间密度空间向量,因为我们如下对于每一时间样本p导出一向量:

[0158]  $\text{MVEck}_{mp} = pB-1\text{MVEck}_m + B-1-pB-1\text{MVEck}_{m-1}$ ,  $p=0, \dots, B-1$ 。

[0159] 对于每一时间样本p,具有 $(N+1)^2$ 个维度的新HOA向量计算为:

[0160]  $\text{CVEck}_p = (\text{XVEck}_{mp}) \text{MVEck}_{mp}$ ,  $p=0, \dots, B-1$

[0161] 通过先前区段的P-B样本增强这些前B个样本以导致第m基于向量的信号的完整HOA表示CVEck<sub>m</sub>。

[0162] 在解码器(例如,图5的实例中所示的音频解码装置24)处,对于某些相异、前景或基于向量的优势声音,可使用线性(或非线性)内插来自前一帧的V向量及来自当前帧的V向量以产生特定时间片段内的较高分辨率(在时间上)经内插V向量。空间时间内插单元76可执行此内插,其中空间-时间内插单元76可接着将当前帧中的US向量与较高分辨率经内插V向量相乘以产生所述特定时间片段内的HOA矩阵。

[0163] 或者,空间-时间内插单元76可将US向量与当前帧的V向量相乘以产生第一HOA矩阵。此外,解码器可将US向量与来自前一帧的V向量相乘以产生第二HOA矩阵。空间-时间内插单元76可接着将线性(或非线性)内插应用于特定时间片段内的第一HOA矩阵及第二HOA矩阵。假定共同输入矩阵/向量,此内插的输出可匹配US向量与经内插V向量的乘法的输出。

[0164] 在一些实例中,对于其执行内插的时间片段的大小可随帧长度而变。换句话说,音频编码装置20可经配置以相对于某一帧长度操作或可配置以相对于若干不同帧长度操作。音频编码装置20可支持的实例帧长度包含768、1024、2048和4096。不同帧长度可导致可能的时间片段长度的不同集合(其中可在样本数目方面指定时间片段)。以下表指定随帧长度(可由变量L表示)而变的可能的时间片段长度的不同集合。



[0165]

	CodedSpatialInterpolationTime							
<i>L</i>	0	1	2	3	4	5	6	7
768	0	32	64	128	256	384	512	768
1024	0	64	128	256	384	512	768	1024
2048	0	128	256	512	768	1024	1536	2048
4096	0	256	512	1024	1536	2048	3072	4096

[0166] 在前述表中,语法元素“CodedSpatialInterpolationTime”表示指示空间内插时间的一或多个位。如上所述,变量*L*表示帧长度。对于768的帧长度,可能的时间片段长度在此实例中由0、32、64、128、256、384、512和768的集合界定。用于当前帧的一个值由CodedSpatialInterpolationTime语法元素的值指定,其中零的值指示0的时间片段长度,一的值指示32等等的时间片段长度。对于1024的帧长度,可能的时间片段长度在此实例中由0、64、128、256、384、512、768和1024的集合界定。用于当前帧的一个值由CodedSpatialInterpolationTime语法元素的值指定,其中零的值指示0的时间片段长度,一的值指示64等等的时间片段长度。对于2048的帧长度,可能的时间片段长度由0、128、256、512、768、1024、1536和2048的集合界定。用于当前帧的一个值由CodedSpatialInterpolationTime语法元素的值指定,其中零的值指示0的时间片段长度,一的值指示128等等的时间片段长度。对于4096的帧长度,可能的时间片段长度在此实例中由0、256、512、1024、1536、2048、3072和4096的集合界定。用于当前帧的一个值由CodedSpatialInterpolationTime语法元素的值指定,其中零的值指示0的时间片段长度,一的值指示256等等的时间片段长度。

[0167] 音频编码装置20的空间-时间内插单元50可相对于选自由帧长度*L*识别的对应集合的若干不同时间片段执行内插。空间-时间内插单元50可选择使跨越帧边界的转变充分平滑(例如,在信噪比方面)且需要最少数目样本(假定内插可为在功率、复杂性、操作等方面相对昂贵的操作)的时间片段。

[0168] 空间-时间内插单元50可以任何数目的不同方式获得帧长度*L*。在一些实例中,音频编码装置20以预设帧速率配置(其可经硬译码,或换句话说,经静态地配置或手动地被配置作为配置音频编码装置20以对H0A系数11进行编码的部分)。在一些实例中,音频编码装置20可基于心理声学音频译码器单元40的核心译码器帧长度指定帧长度。关于标题为“信息技术-MPEG音频技术-部分3:统一语音和音频译码”的ISO/IEC23003-3:2012中的“coreCoderFrameLength”的论述可找到关于核心译码器帧长度的更多信息。

[0169] 当基于核心译码器帧长度确定时,音频编码装置20可参考下表:

[0170] 表一FrameLengthFactor定义

[0171]

核心译码器帧长度	FrameLengthFactor			
	00	01	10	11
	<i>C</i> 帧长度			
768	1	1	1	保留
1024	1	1	1	
2048	1	1/2	1/2	
4096	1	1/2	1/4	

[0172] 在前述表中,音频编码装置20可设定一或多个位(由语法元素

“FrameLengthFactor”表示),其指示将乘以上表的第一列中指定的核心译码器帧长度的因数。音频编码装置20可基于各种译码准则而选择1、1/2和1/4的帧长度因数中的一者,或可基于以各种因数中的每一者对帧进行译码的尝试而选择因数中的一者。音频编码装置20可例如确定核心译码器帧长度是4096且选择1、1/2或1/4的帧长度因数。音频编码装置20可在位流21的HOAConfig部分(如上相对于图7的实例所述)中用信号表示帧长度因数,其中00(二进制)的值指示1的帧长度因数,01(二进制)的值指示1/2的帧长度因数,且10(二进制)的值指示1/4的帧长度因数。音频编码装置20还可将帧长度L确定为核心译码器帧长度乘以帧长度因数(例如,1、1/2或1/4)。

[0173] 在这方面中,音频编码装置20可至少部分地基于指示帧长度(L)的一或多个位和指示空间-时间内插时间的一或多个位(例如,codedSpatioInterpolationTime语法元素)而获得时间片段。音频编码装置20还可通过至少部分地相对于第一多个球谐系数的第一分解和第二多个球谐系数的第二分解执行内插而获得时间片段的经分解的经内插球谐系数。

[0174] 音频解码装置24可执行与上文相对于音频编码装置20描述的那些操作大体上类似的操作。确切地说,音频解码装置24的空间-时间内插单元76可获得随指示帧长度因数的一或多个位(例如,frameLengthFactor语法元素)和核心译码器帧长度(其也可以由心理声学音频编码单元40在位流21中指定)而变的帧长度。空间-时间内插单元76还可获得指示空间-时间内插时间的一或多个位(例如,CodedSpatialInterpolationTime语法元素)。空间-时间内插单元76可使用帧长度L和codedSpatialInterpolationTim语法元素作为识别时间片段长度的键而在上文提到的表中执行查找。音频解码装置24接着可针对所获得时间片段以上文所描述的方式执行内插。

[0175] 在这方面中,音频解码装置24可至少部分地基于指示帧长度(L)的一或多个位和指示空间-时间内插时间的一或多个位(例如,codedSpatioInterpolationTime语法元素)而获得时间片段。音频解码装置24还可通过至少部分地相对于第一多个球谐系数的第一分解和第二多个球谐系数的第二分解执行内插而获得时间片段的经分解的经内插球谐系数。

[0176] 图13是说明经配置以执行本发明中描述的一或多个技术的一或多个音频编码器和音频解码器的图。如上文所论述,SVD可用作HOA信号压缩系统的基础。在一些实例中,HOA信号H可经分解为 $USV'$ ( $'$ 是矩阵的转置)。在一些实例中,US和V矩阵的第一少数行可经界定为背景信号(例如,环境信号),且US和V矩阵的第一少数列可经界定为前景信号。在一些实例中,背景和前景信号可以类似方式交叉淡化。然而,以类似方式交叉淡化背景和前景信号可导致执行冗余计算。为了减少执行的计算且改善系统的其它方面,本发明描述用于背景信号的新交叉淡化算法。

[0177] 在一些系统中,US矩阵和V矩阵分别单独地经交叉淡化为US\_C矩阵(例如,经交叉淡化US矩阵)和V\_C矩阵(例如,经交叉淡化V矩阵)。随后,经交叉淡化HOA信号H\_C可经重构为 $US_C * V_C'$ 。根据本发明的一或多种技术,原始HOA信号H可经重构为 $USV'$ (例如,在交叉淡化之前)。接着可如贯穿本发明所描述在HOA域中执行交叉淡化。

[0178] 如上所述,帧的长度(或换句话说,样本的数目)可变化(例如,随核心译码器帧长度而变)。帧长度的差连同空间-时间内插时间的不同集合如上文所描述可影响交叉淡化。一般来说,由CodedSpatialInterpolationTime语法元素识别的空间-时间内插时间和帧长度L可指定将交叉淡化的样本的数目。如图13的实例中所示,U矩阵的大小是(L+

SpatialInterpolationTime)\*25,其中SpatialInterpolationTime变量表示使用上文相对于图12所论述的表随CodedSpatialInterpolationTime语法元素和L而变获得的空间内插时间。当L等于1024且CodedSpatialInterpolationTime语法元素的值等于三时SpatialInterpolationTime的实例值可为256。当L等于2048且CodedSpatialInterpolationTime语法元素的值等于三时将用于下文说明的目的的SpatialInterpolationTime的另一实例值可为512。在此说明性实例下,L+SpatialInterpolationTime等于2048+512或2560。

[0179] 在任何情况下,背景HOA系数在此实例中具有尺寸2560\*4。交叉淡化因此在前一帧的SptailInterpolationTime数目的样本(例如,512样本)与当前帧的第一SptailInterpolationTime数目的样本(例如,512样本)之间发生。输出因此是L个样本,其经AAC或USAC译码。因此,用于空间时间内插V向量的SpatialInterpolationTime还可识别对其执行交叉淡化的样本的数目。以此方式,指示FrameLength的一个或多个位和指示空间-时间内插时间的一或多个位可影响交叉淡化持续时间。

[0180] 此外,能量补偿单元38可通过将开窗函数应用于 $V_{BG}[k]$ 向量35<sub>BG</sub>以产生经能量补偿 $V_{BG}[k]$ 向量35<sub>BG</sub>'而执行能量补偿以产生环境HOA系数47'。所述开窗函数可包括具有等于帧长度L的长度的开窗函数。在这方面中,能量补偿单元38可至少部分地在指示帧长度因数的一或多个位(例如,FrameLengthFactor语法元素)上使用获得的用于能量补偿的相同帧长度L。

[0181] 交叉淡化单元66的混合器单元70可组合(例如,在其间交叉淡化)当前帧的经能量补偿环境HOA系数47'的第一部分73与前一帧的经能量补偿环境HOA系数47'的第三部分76以产生中间经交叉淡化的经能量补偿环境HOA系数77。混合器单元70可将所产生中间经交叉淡化的经能量补偿环境HOA系数77输出到成帧单元71。由于在此实例中混合器单元70利用前一帧的经能量补偿环境HOA系数47'的第三部分76,因此可假定交叉淡化单元66在处理当前帧之前是在操作中。因此,与分开地使当前帧的US矩阵与前一帧的US矩阵进行交叉淡化且使当前帧的V矩阵与前一帧的V矩阵进行交叉淡化相反,混合器单元70可在能量补偿域中进行交叉淡化。以此方式,根据本发明的技术可减少交叉淡化单元66的计算负载、电力消耗和/或复杂性。

[0182] 可相对于任何数目个不同情形及音频生态系统执行前述技术。下文描述多个实例情形,但所述技术不应限于所述实例情形。一个实例音频生态系统可包含音频内容、电影工作室、音乐工作室、游戏音频工作室、基于信道的音频内容、译码引擎、游戏音频原声、游戏音频译码/再现引擎,以及递送系统。

[0183] 电影工作室、音乐工作室及游戏音频工作室可接收音频内容。在一些实例中,音频内容可表示获取的输出。电影工作室可例如通过使用数字音频工作站(DAW)输出基于信道的音频内容(例如,呈2.0、5.1及7.1)。音乐工作室可例如通过使用DAW输出基于信道的音频内容(例如,呈2.0及5.1)。在任一情况下,译码引擎可基于一或多个编解码器(例如,AAC、AC3、Dolby True HD、Dolby Digital Plus及DTS MasterAudio)接收并编码基于信道的音频内容以供传递系统输出。游戏音频工作室可例如通过使用DAW输出一或多个游戏音频原声。游戏音频译码/再现引擎可译码音频原声及或将音频原声再现成基于信道的音频内容以供传递系统输出。可执行所述技术的另一实例情形包括音频生态系统,其可包含广播记

录音频对象、专业音频系统、消费型装置上捕获、HOA音频格式、装置上再现、消费型音频、TV及附件及汽车音频系统。

[0184] 广播记录音频对象、专业音频系统及消费型装置上捕获都可使用HOA音频格式对其输出进行译码。以此方式,可使用HOA音频格式将音频内容译码成单一表示,可使用装置上再现、消费型音频、TV及附件及汽车音频系统重放所述单一表示。换句话说,可在通用音频重放系统(即,与需要例如5.1、7.1等的特定配置相比)(例如,音频重放系统16)处重放音频内容的单一表示。

[0185] 可执行所述技术的情形的其它实例包含可包含获取元件及重放元件的音频生态系统。获取元件可包含有线及/或无线获取装置(例如,本征麦克风)、装置上环绕声捕获及移动装置(例如,智能电话及平板计算机)。在一些实例中,有线及/或无线获取装置可经由有线及/或无线通信信道耦合到移动装置。

[0186] 根据本发明的一或多个技术,移动装置可用于获取声场。举例来说,移动装置可经由有线及/或无线获取装置及/或装置上环绕声捕获(例如,集成到移动装置中的多个麦克风)获取声场。移动装置可接着将所获取声场译码成HOA系数以用于由重放元件中的一或多者重放。举例来说,移动装置的用户可记录实况事件(例如,集会、会议、比赛、音乐会等)(获取实况事件的声场),且将记录译码至HOA系数中。

[0187] 移动装置还可利用重放元件中的一或多者来重放经HOA译码声场。举例来说,移动装置可对经HOA译码的声场进行解码,且将致使重放元件中的一或多者重新产生声场的信号输出到重放元件中的一或多者。作为一个实例,移动装置可利用无线及/或无线通信信道将信号输出到一或多个扬声器(例如,扬声器阵列、声棒等)。作为另一实例,移动装置可利用对接解决方案将信号输出到一或多个对接站及/或一或多个对接扬声器(例如,智能汽车及/或家庭中的声音系统)。作为另一实例,移动装置可利用头戴式耳机再现将信号输出到一组头戴式耳机(例如)以产生实际的立体声声音。

[0188] 在一些实例中,特定移动装置可获取3D声场并且在稍后时间重放同一3D声场。在一些实例中,移动装置可获取3D声场,将3D声场编码为HOA且将经编码3D声场发射到一或多个其它装置(例如,其它移动装置及/或其它非移动装置)以用于重放。

[0189] 可执行所述技术的又一情形包含音频生态系统,其可包含音频内容、游戏工作室、经译码音频内容、再现引擎及传递系统。在一些实例中,游戏工作室可包含可支持HOA信号的编辑的一或多个DAW。举例来说,所述一或多个DAW可包含可经配置以与一或多个游戏音频系统一起操作(例如,工作)的HOA插件及/或工具。在一些实例中,游戏工作室可输出支持HOA的新原声格式。在任何情况下,游戏工作室可将经译码音频内容输出到再现引擎,所述再现引擎可再现声场以供传递系统重放。

[0190] 也可相对于示范性音频获取装置执行所述技术。举例来说,可相对于可包含共同地经配置以记录3D声场的多个麦克风的本征麦克风执行所述技术。在一些实例中,本征麦克风的所述多个麦克风可位于具有大约4cm的半径的实质上球面球的表面上。在一些实例中,音频编码装置20可集成到本征麦克风中以便直接从麦克风输出位流21。

[0191] 另一示范性音频获取情形可包含可经配置以从一或多个麦克风(例如,一或多个本征麦克风)接收信号的制作车。制作车还可包含音频编码器,例如图3的音频编码器20。

[0192] 在一些情况下,移动装置还可包含共同地经配置以记录3D声场的多个麦克风。换

句话说,所述多个麦克风可具有X、Y、Z分集。在一些实例中,移动装置可包含可旋转以相对于移动装置的一或多个其它麦克风提供X、Y、Z分集的麦克风。移动装置还可包含音频编码器,例如图3的音频编码器20。

[0193] 加固型视频捕获装置可进一步经配置以记录3D声场。在一些实例中,加固型视频捕获装置可附接到参与活动的用户的头盔。举例来说,加固型视频捕获装置可在用户泛舟时附接到用户的头盔。以此方式,加固型视频捕获装置可捕获表示用户周围的动作(例如,水在用户身后的撞击、另一泛舟者在用户前方说话等)的3D声场。

[0194] 还可相对于可经配置以记录3D声场的附件增强型移动装置执行所述技术。在一些实例中,移动装置可类似于上文所论述的移动装置,其中添加一或多个附件。举例来说,本征麦克风可附接到上文所提及的移动装置以形成附件增强型移动装置。以此方式,附件增强型移动装置可捕获3D声场的较高质量版本,而不是仅使用与附件增强型移动装置成一体式的声音捕获组件。

[0195] 下文进一步论述可执行本发明中所描述的技术的各种方面的实例音频重放装置。根据本发明的一或多个技术,扬声器及/或声棒在重放3D声场时可布置于任何任意配置中。此外,在一些实例中,头戴受话器重放装置可经由有线或无线连接耦合到解码器24。根据本发明的一或多个技术,可利用声场的单一通用表示来在扬声器、声棒及头戴受话器重放装置的任何组合上再现声场。

[0196] 多个不同实例音频重放环境还可适合于执行本发明中所描述的技术的各种方面。举例来说,以下环境可为用于执行本发明中所描述的技术的各种方面的合适环境:5.1扬声器重放环境、2.0(例如,立体声)扬声器重放环境、具有全高前扩音器的9.1扬声器重放环境、22.2扬声器重放环境、16.0扬声器重放环境、汽车扬声器重放环境,及具有耳机重放环境的移动装置。

[0197] 根据本发明的一或多个技术,可利用声场的单一通用表示来在所述重放环境中的任一者上再现声场。另外,本发明的技术使得再现器能够从通用表示再现声场以供在除上文所描述的环境之外的重放环境上重放。举例来说,如果设计考虑禁止扬声器根据7.1扬声器重放环境的恰当置放(例如,如果不可能置放右环绕扬声器),那么本发明的技术使得再现器能够以其它6个扬声器进行补偿,使得可在6.1扬声器重放环境上实现重放。

[0198] 此外,用户可在佩戴头戴受话器时观看运动比赛。根据本发明的一或多个技术,可获取运动比赛的3D声场(例如,可将一或多个本征麦克风放置于棒球场中及/或周围),可获得对应于3D声场的HOA系数且将所述HOA系数发射到解码器,所述解码器可基于HOA系数重构3D声场且将经重构3D声场输出到再现器,并且所述再现器可获得关于重放环境的类型(例如,头戴式耳机)的指示,且将经重构3D声场再现成致使头戴式耳机输出运动比赛的3D声场的表示的信号。

[0199] 在上述各种实例中的每一者中,应理解,音频编码装置20可执行方法,或另外包括执行音频编码装置20经配置以执行的方法的每一步骤的装置。在一些情况下,所述装置可包括一或多个处理器。在一些情况下,所述一或多个处理器可表示借助于存储到非暂时性计算机可读存储媒体的指令配置的专用处理器。换句话说,编码实例的集合中的每一者中的技术的各方面可提供其上存储有指令的非暂时性计算机可读存储媒体,所述指令在执行时致使所述一或多个处理器执行音频编码装置20已经配置以执行的方法。

[0200] 在一或多个实例中,所描述的功能可在硬件、软件、固件或其任何组合中实施。如果以软件实施,那么所述功能可以作为一或多个指令或代码在计算机可读媒体上存储或传输,并且由基于硬件的处理单元来执行。计算机可读媒体可包含计算机可读存储媒体,其对应于例如数据存储媒体等有形媒体。数据存储媒体可为可由一或多个计算机或者一或多个处理器存取以检索用于实施本发明中描述的技术的指令、代码及/或数据结构的任何可用媒体。计算机程序产品可以包含计算机可读媒体。

[0201] 同样,在上文所描述的各种情况中的每一者中,应理解,音频解码装置24可执行方法或另外包括用于执行音频解码装置24经配置以执行的方法的每一步骤的装置。在一些情况下,所述装置可包括一或多个处理器。在一些情况下,所述一或多个处理器可表示借助于存储到非暂时性计算机可读存储媒体的指令配置的专用处理器。换句话说,编码实例的集合中的每一者中的技术的各方面可提供其上存储有指令的非暂时性计算机可读存储媒体,所述指令在执行时致使所述一或多个处理器执行音频解码装置24已经配置以执行的方法。

[0202] 借助于实例而非限制,此类计算机可读存储媒体可包括RAM、ROM、EEPROM、CD-ROM或其它光盘存储装置、磁盘存储装置或其它磁性存储装置、快闪存储器或可以用来存储指令或数据结构的形式的期望程序代码并且可以由计算机存取的任何其它媒体。然而,应理解,所述计算机可读存储媒体和数据存储媒体并不包含连接、载波、信号或其它暂时性媒体,而是实际上针对非暂时性的有形存储媒体。如本文中所使用,磁盘和光盘包含压缩光盘(CD)、激光光盘、光学光盘、数字多功能光盘(DVD)、软性磁盘和蓝光光盘,其中磁盘通常以磁性方式再现数据,而光盘利用激光以光学方式再现数据。以上各项的组合也应包含在计算机可读媒体的范围内。

[0203] 可由例如一或多个数字信号处理器(DSP)、通用微处理器、专用集成电路(ASIC)、现场可编程逻辑阵列(FPGA)或其它等效集成或离散逻辑电路等一或多个处理器来执行指令。因此,如本文中所使用的术语“处理器”可指前述结构或适合于实施本文中所描述的技术的任一其它结构中的任一者。另外,在一些方面中,本文中所描述的功能性可以在经配置用于编码和解码的专用硬件和/或软件模块内提供,或者并入在组合编解码器中。而且,所述技术可完全实施于一或多个电路或逻辑元件中。

[0204] 本发明的技术可在广泛多种装置或设备中实施,包括无线手持机、集成电路(IC)或一组IC(例如,芯片组)。本发明中描述各种组件、模块或单元是为了强调经配置以执行所公开的技术的装置的功能方面,但未必需要由不同硬件单元实现。实际上,如上文所描述,各种单元可以结合合适的软件及/或固件组合在编码解码器硬件单元中,或者通过互操作硬件单元的集合来提供,所述硬件单元包含如上文所描述的一或多个处理器。

[0205] 已经描述了所述技术的各方面。所述技术的这些和其它方面在所附权利要求书的范围内。

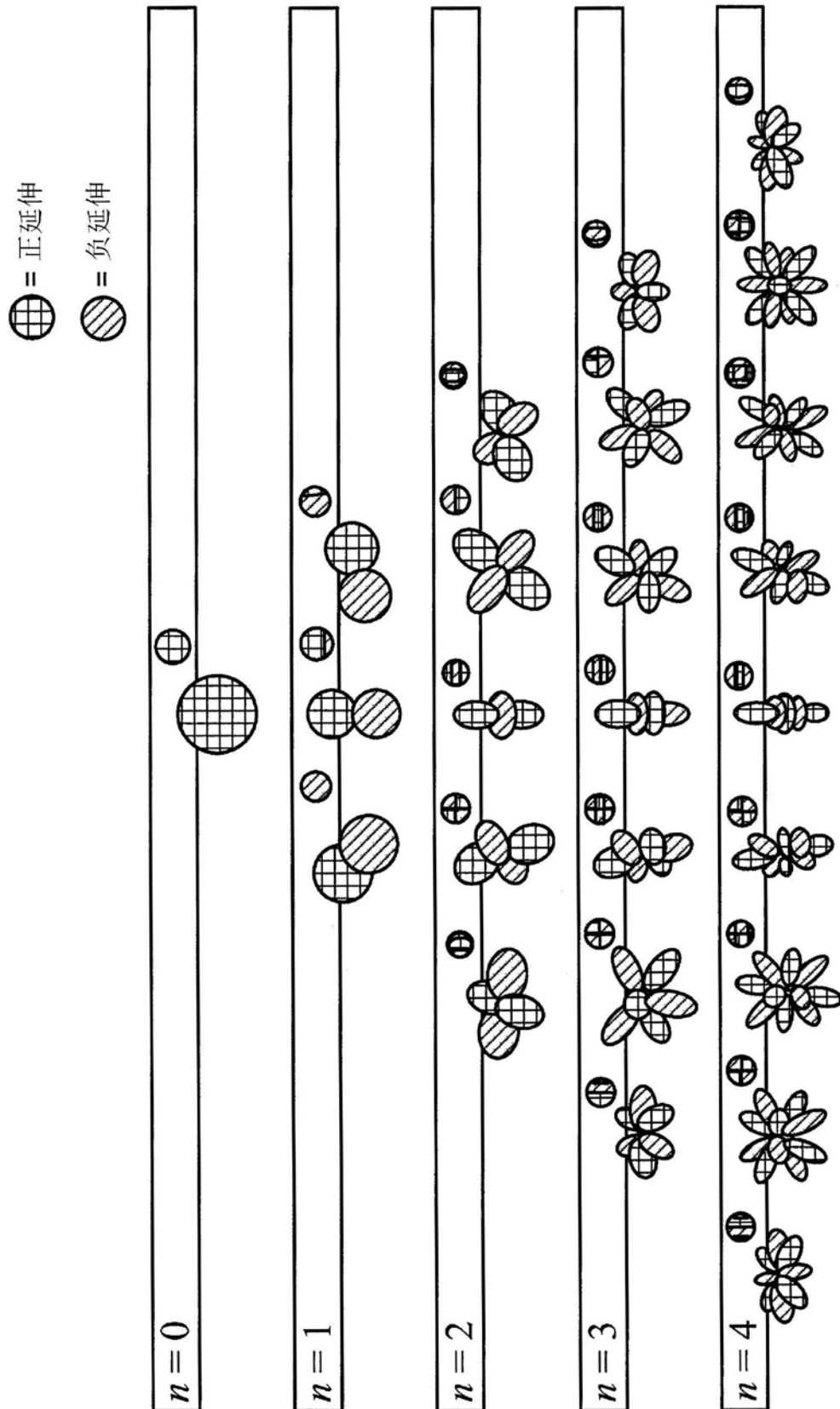


图1

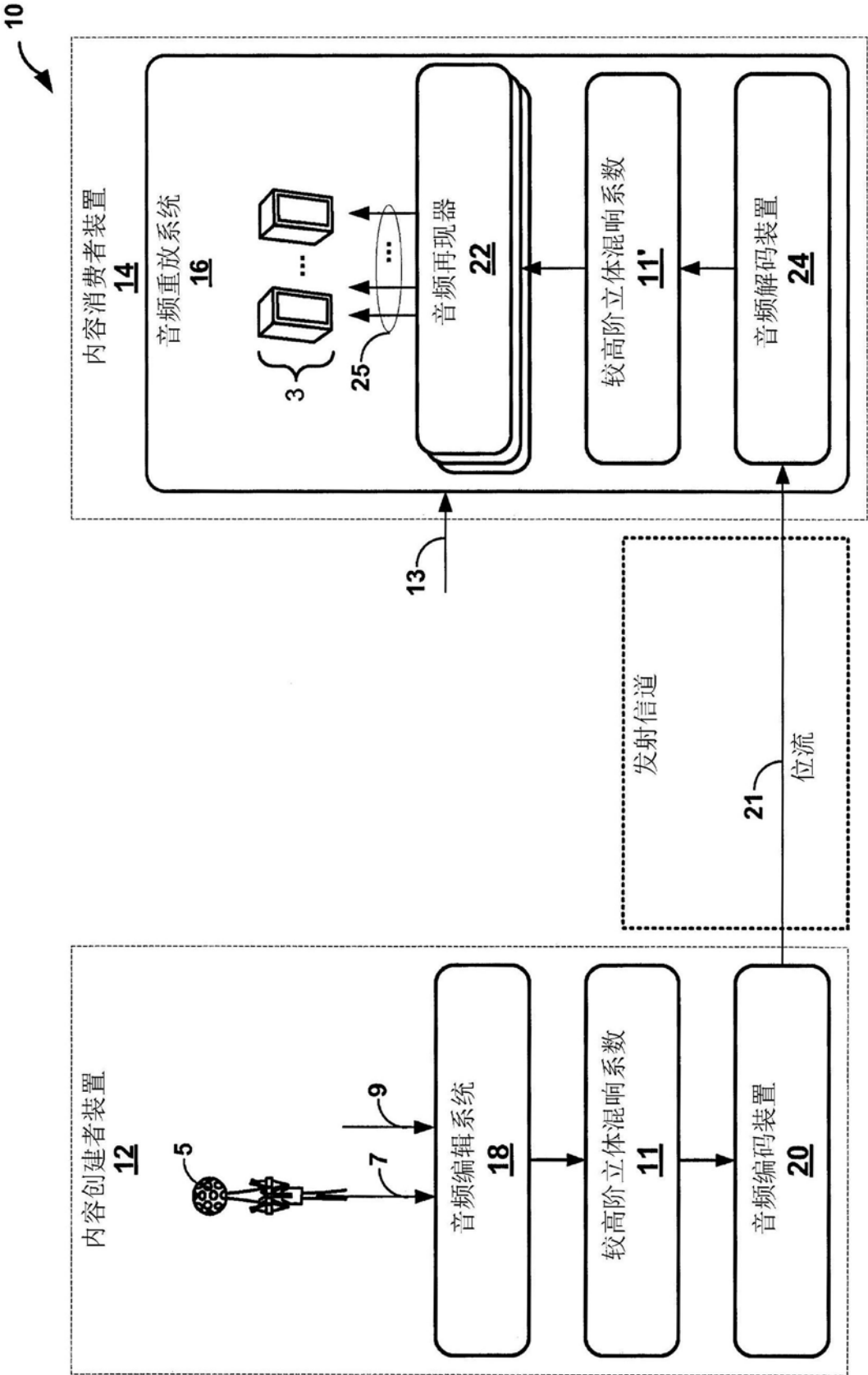


图2



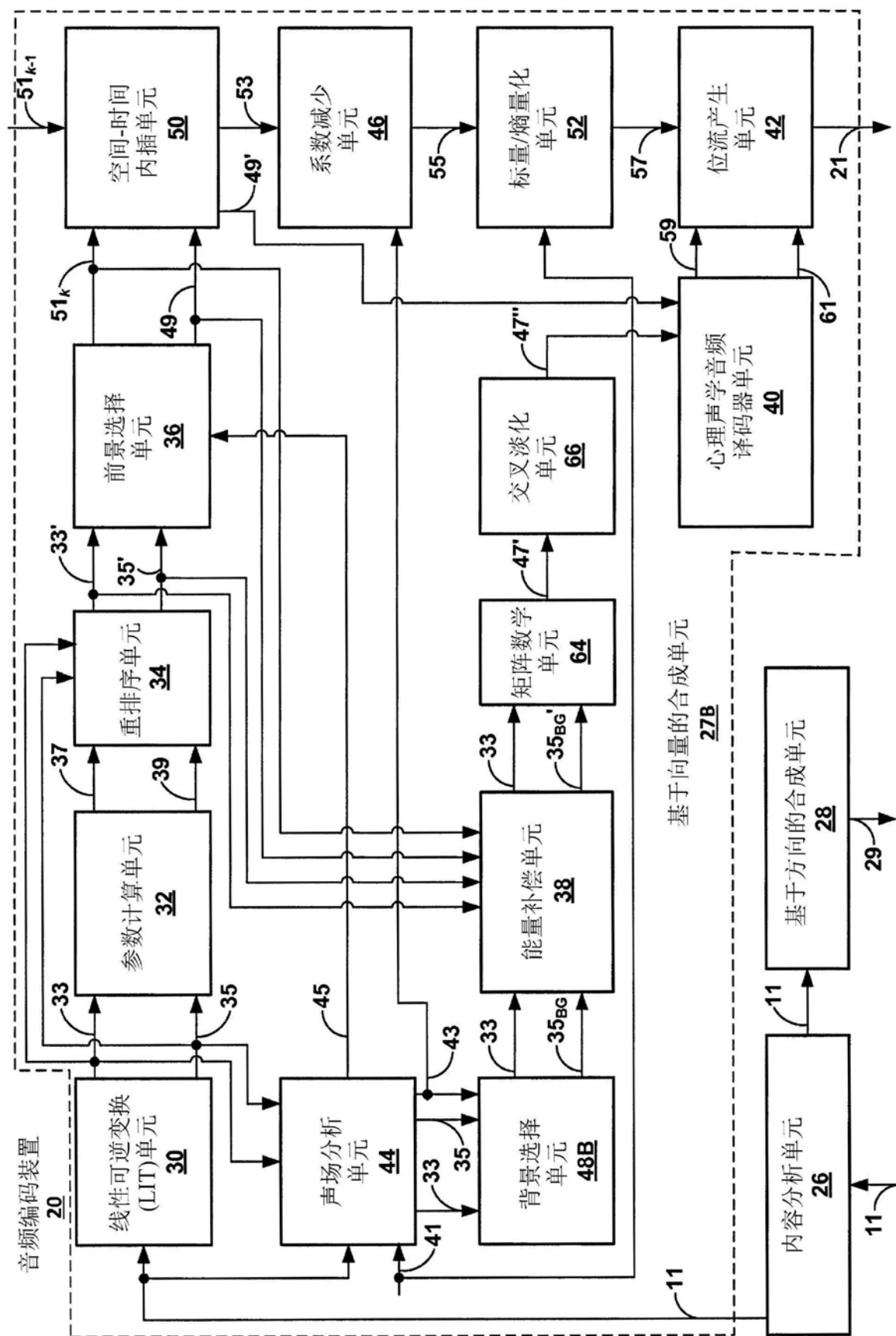


图3

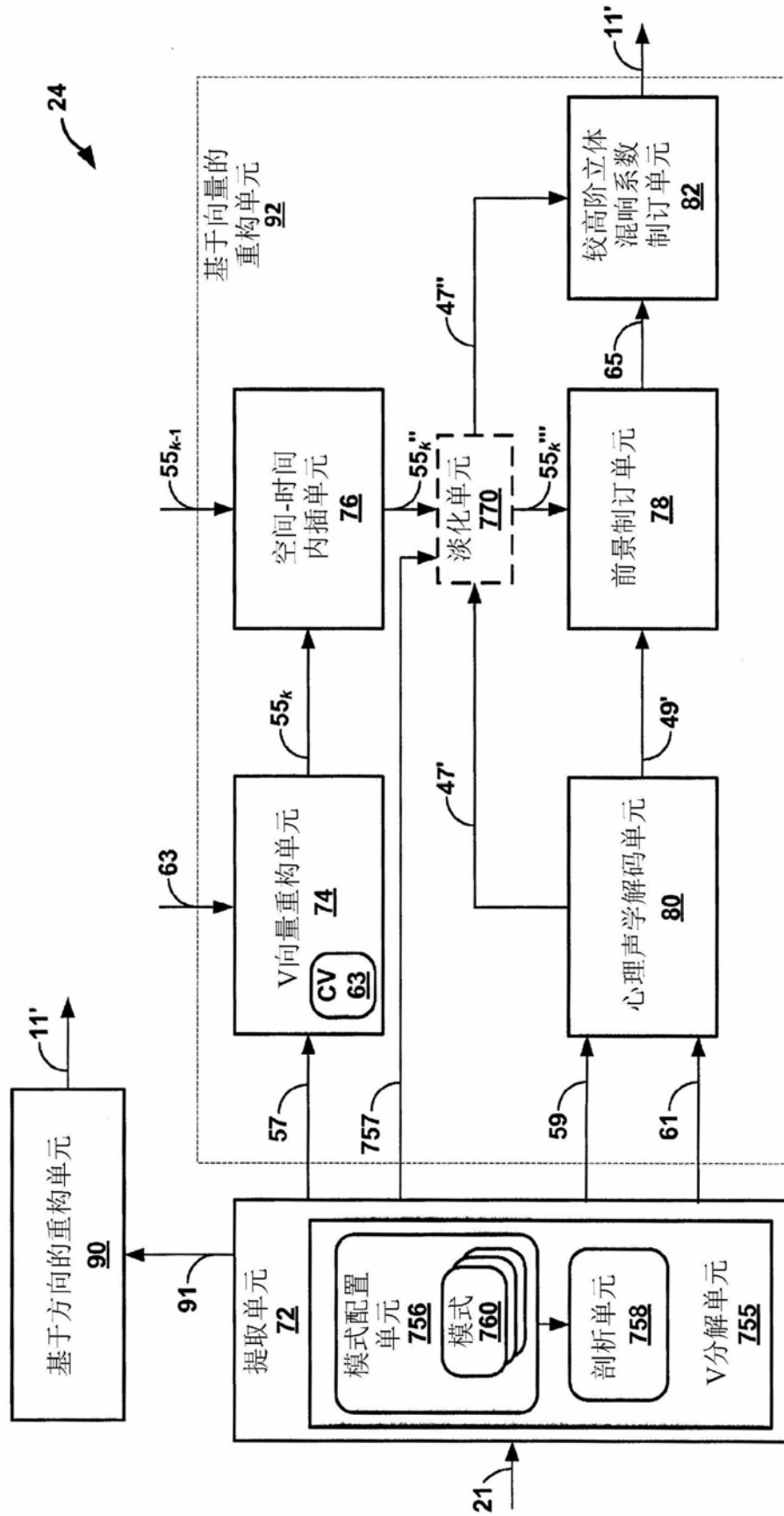


图4

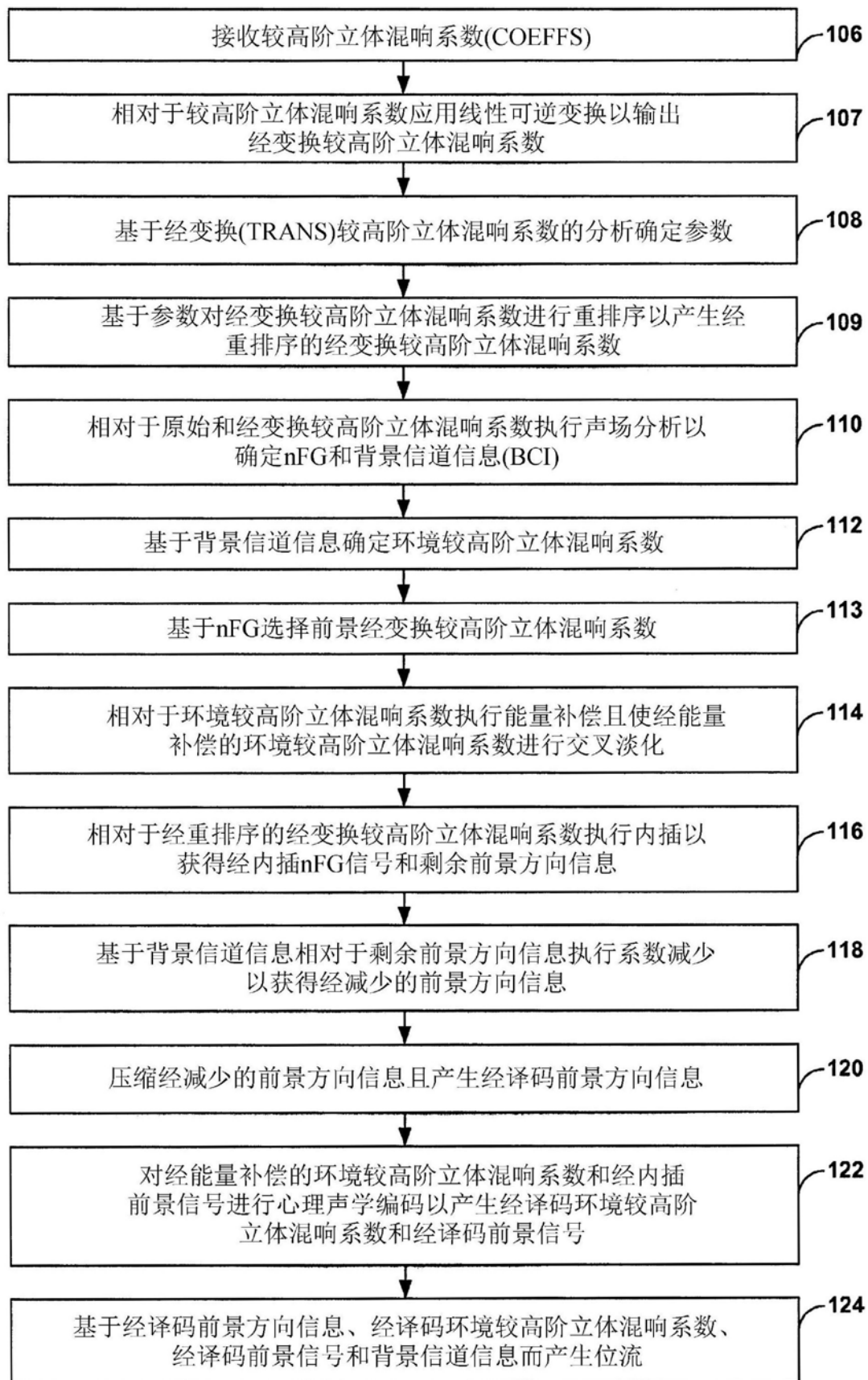


图5

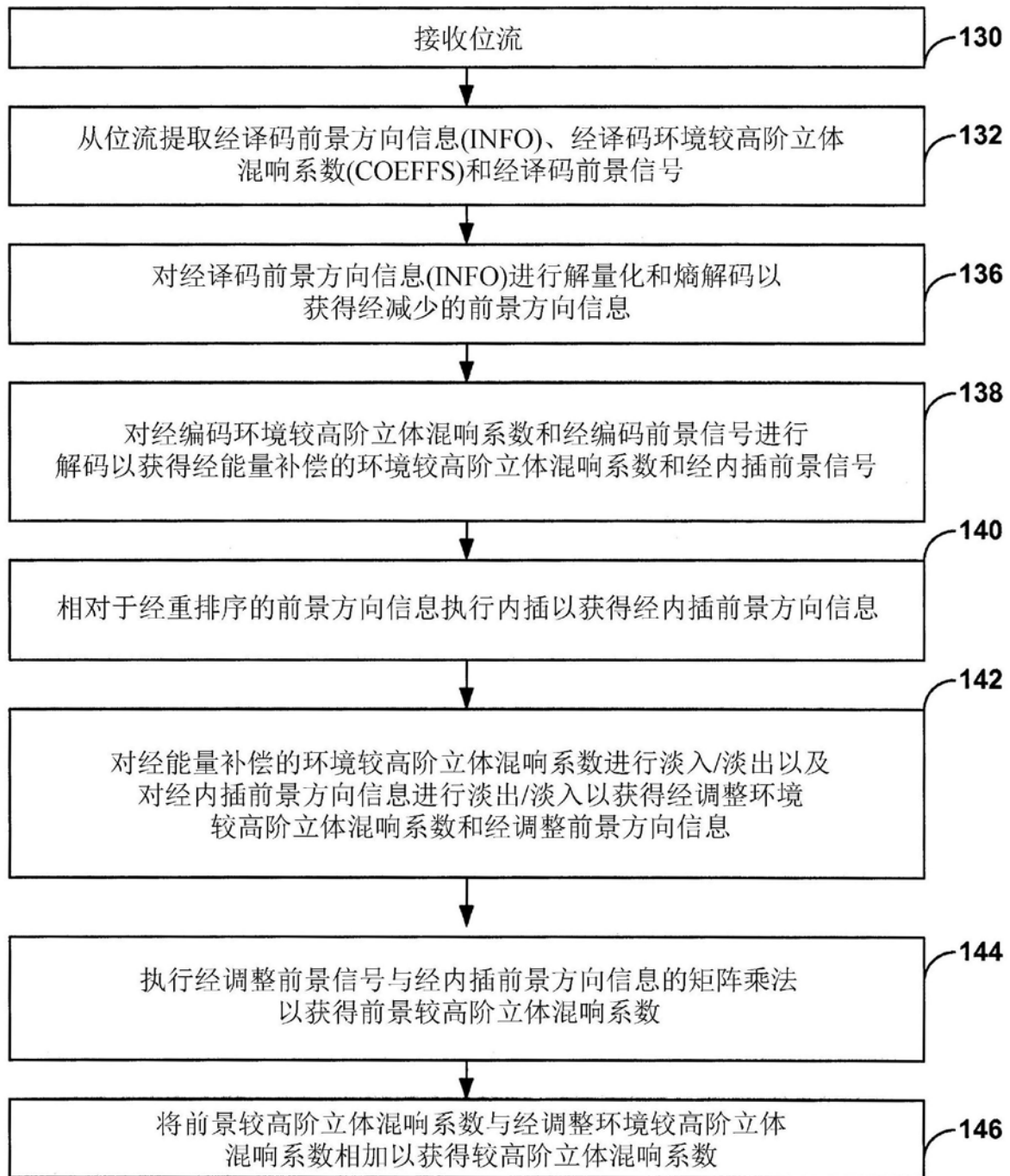


图6

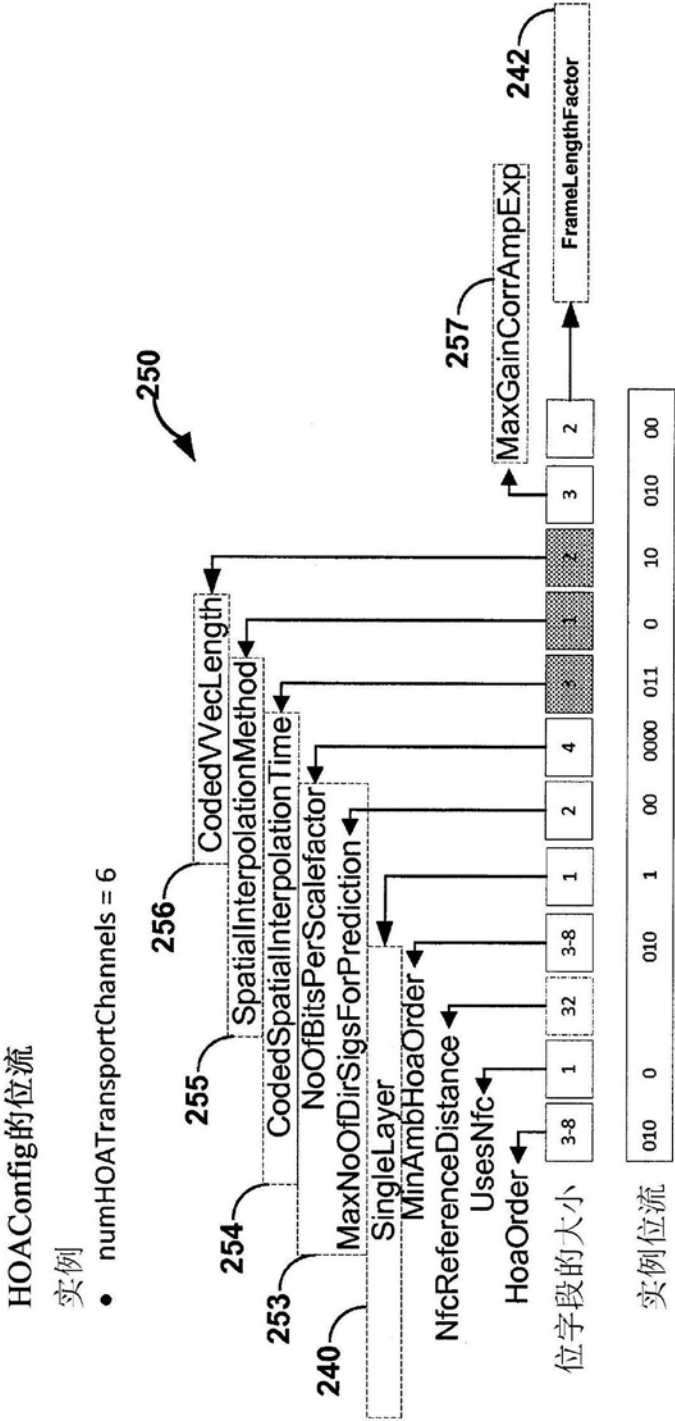


图7

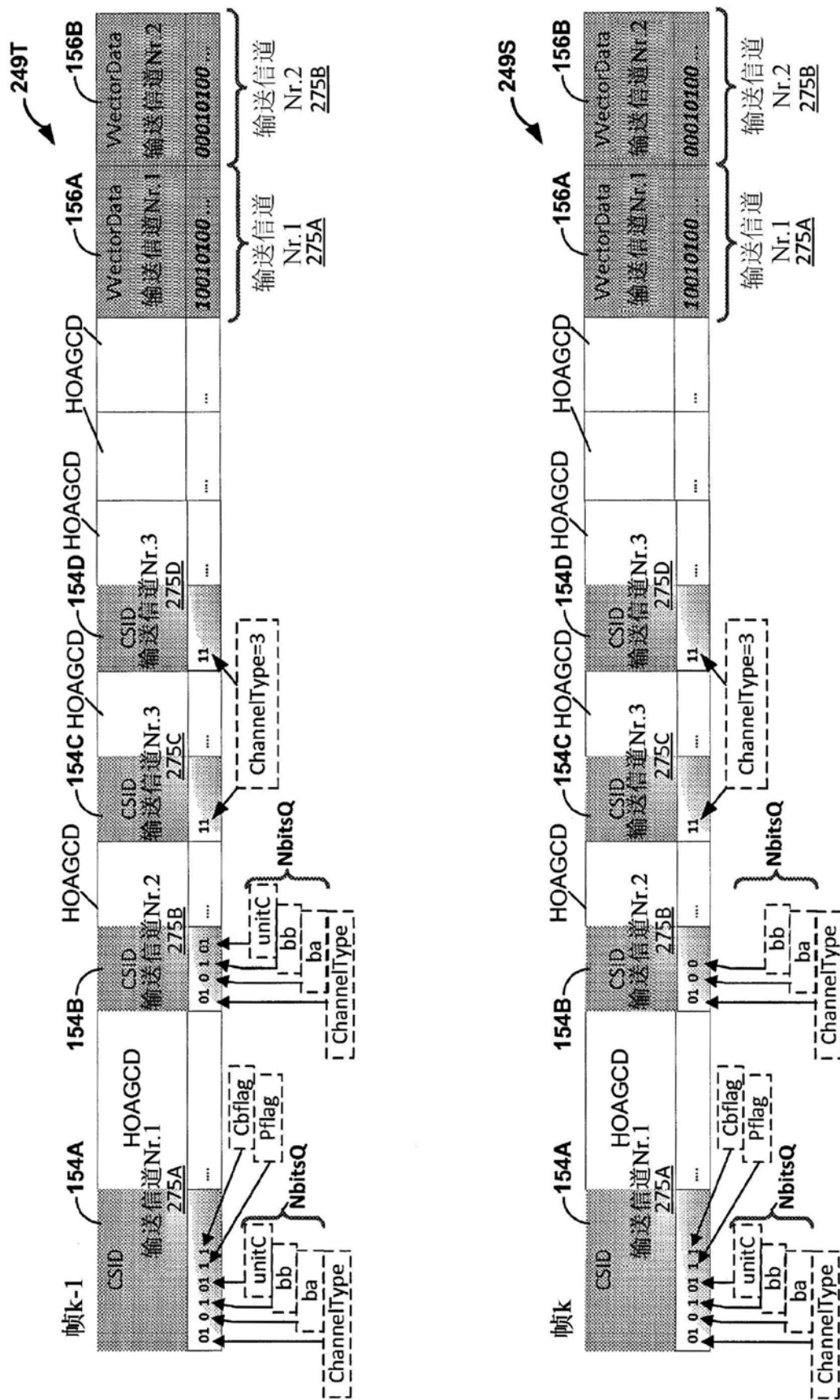


图8

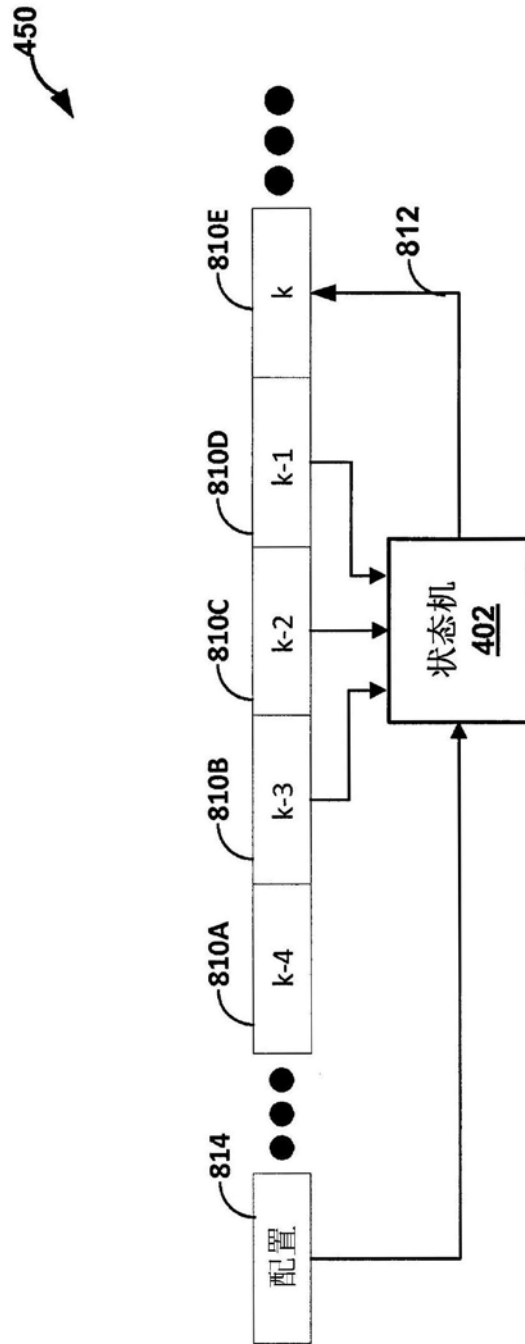


图9

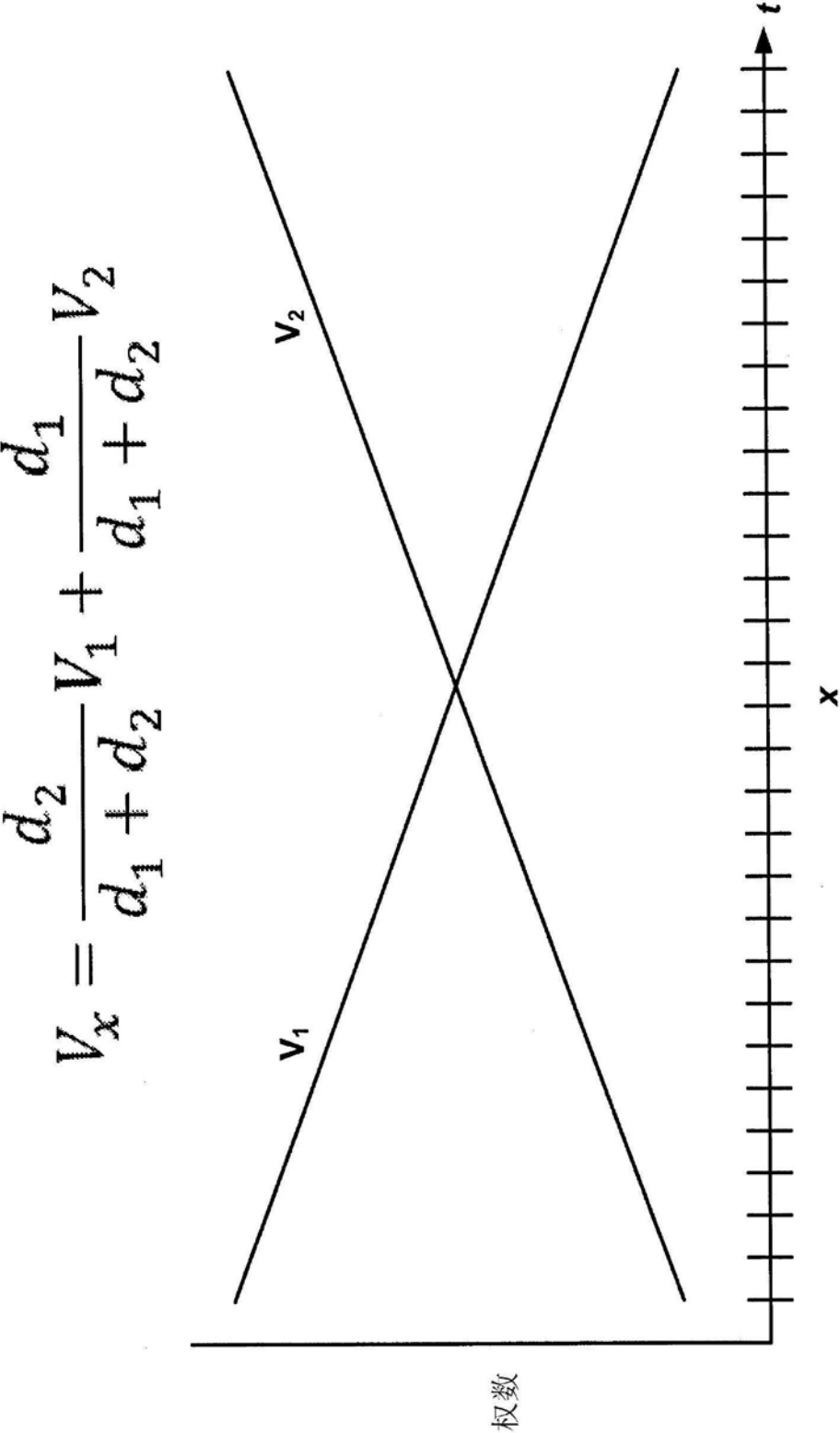


图10



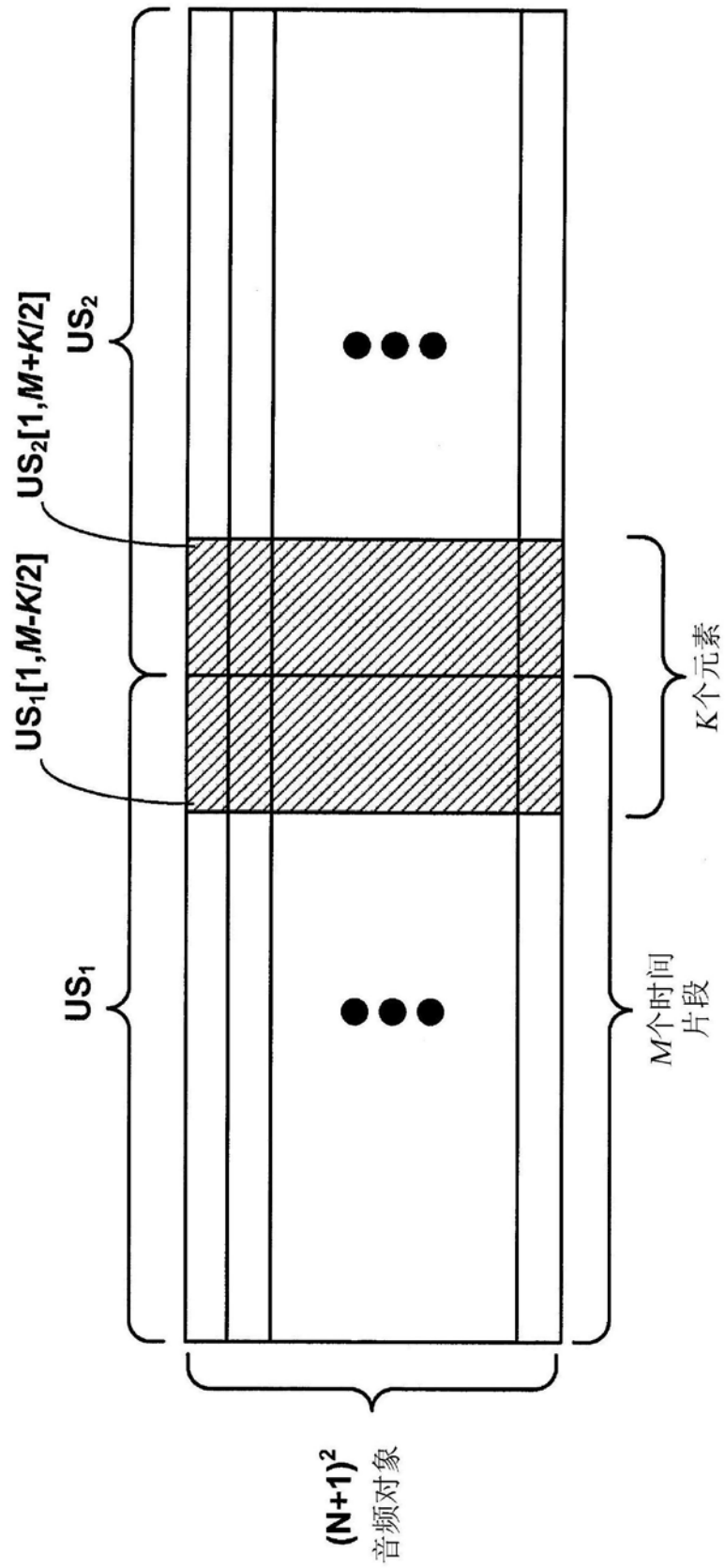


图11

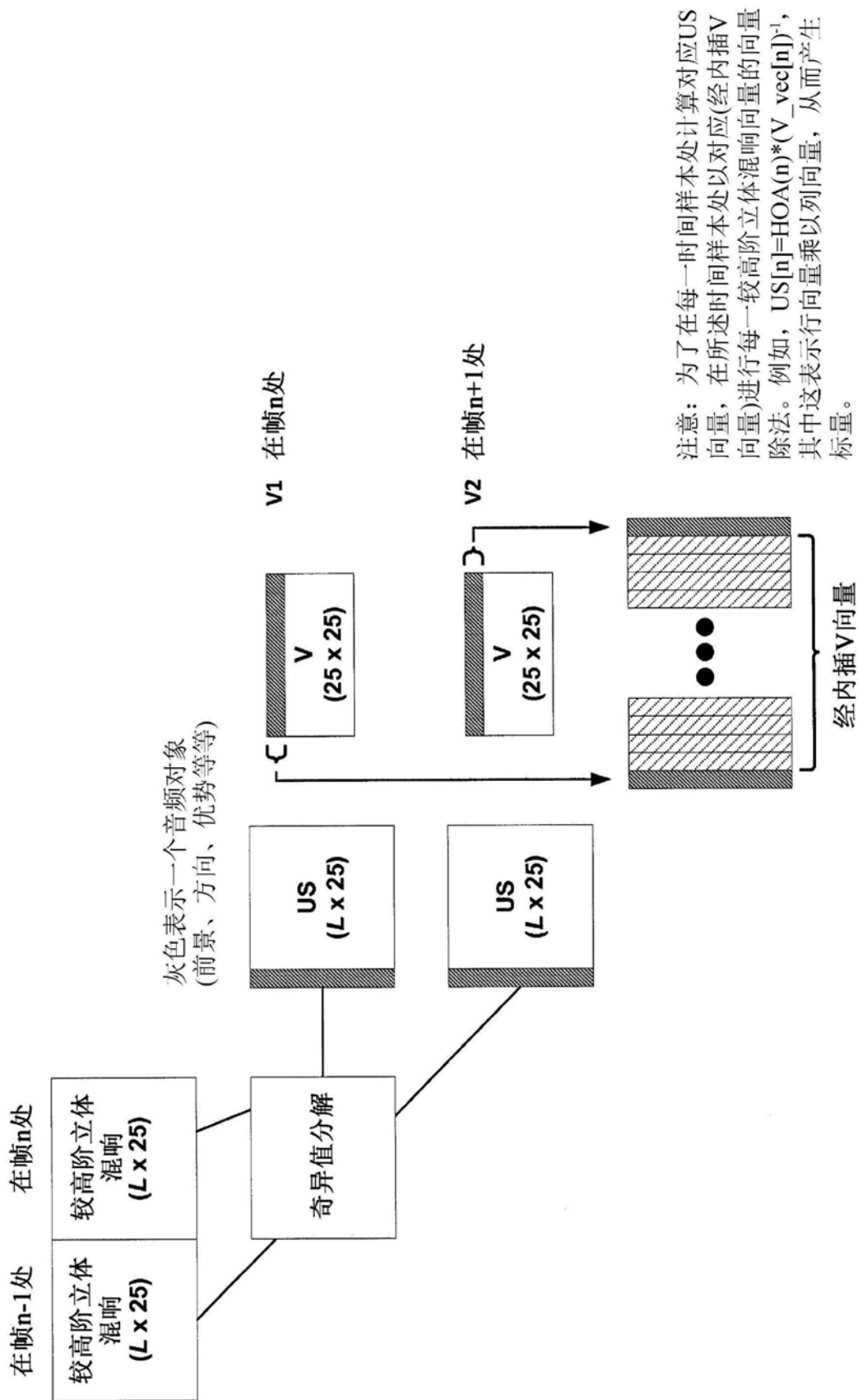


图12

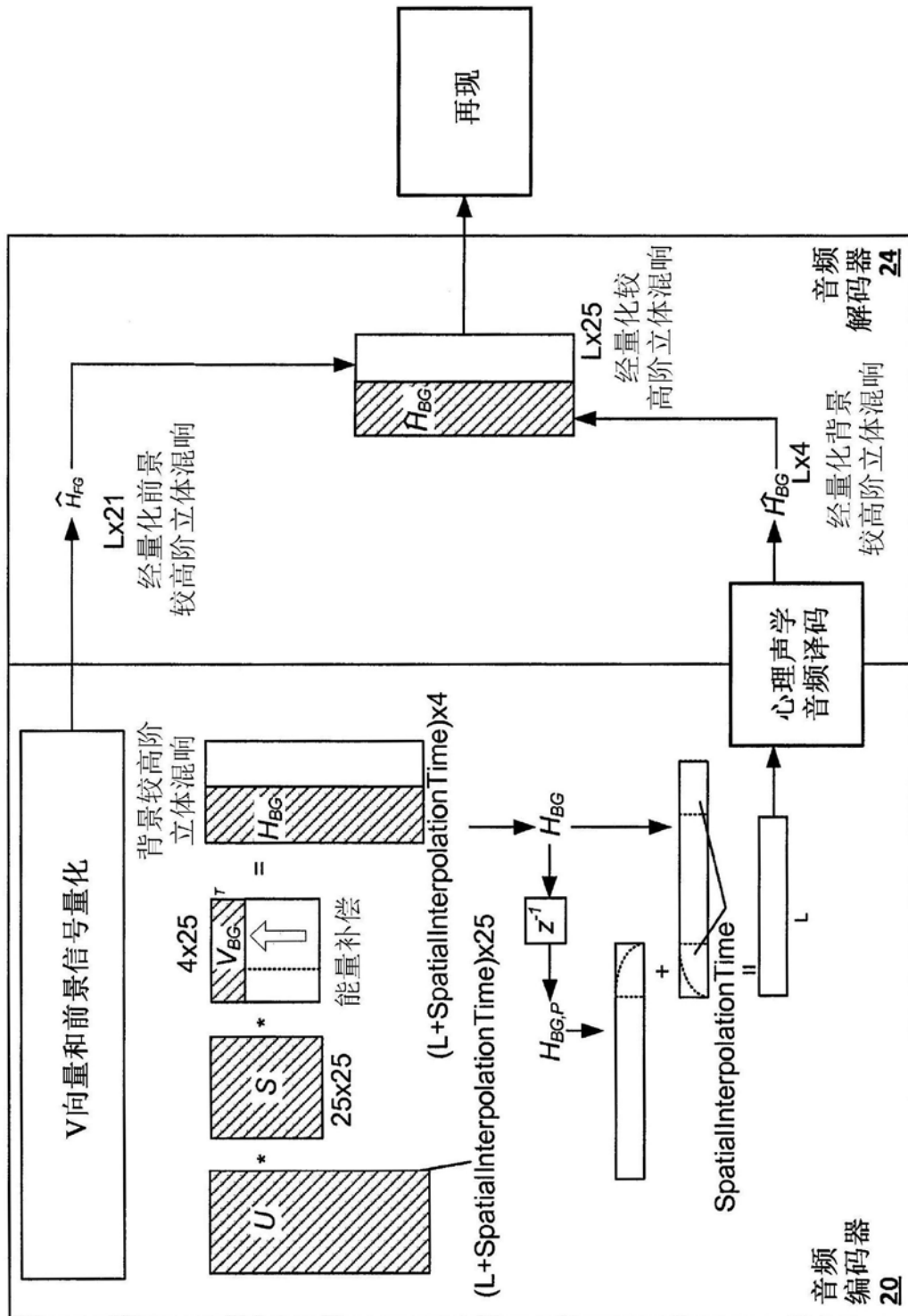


图13

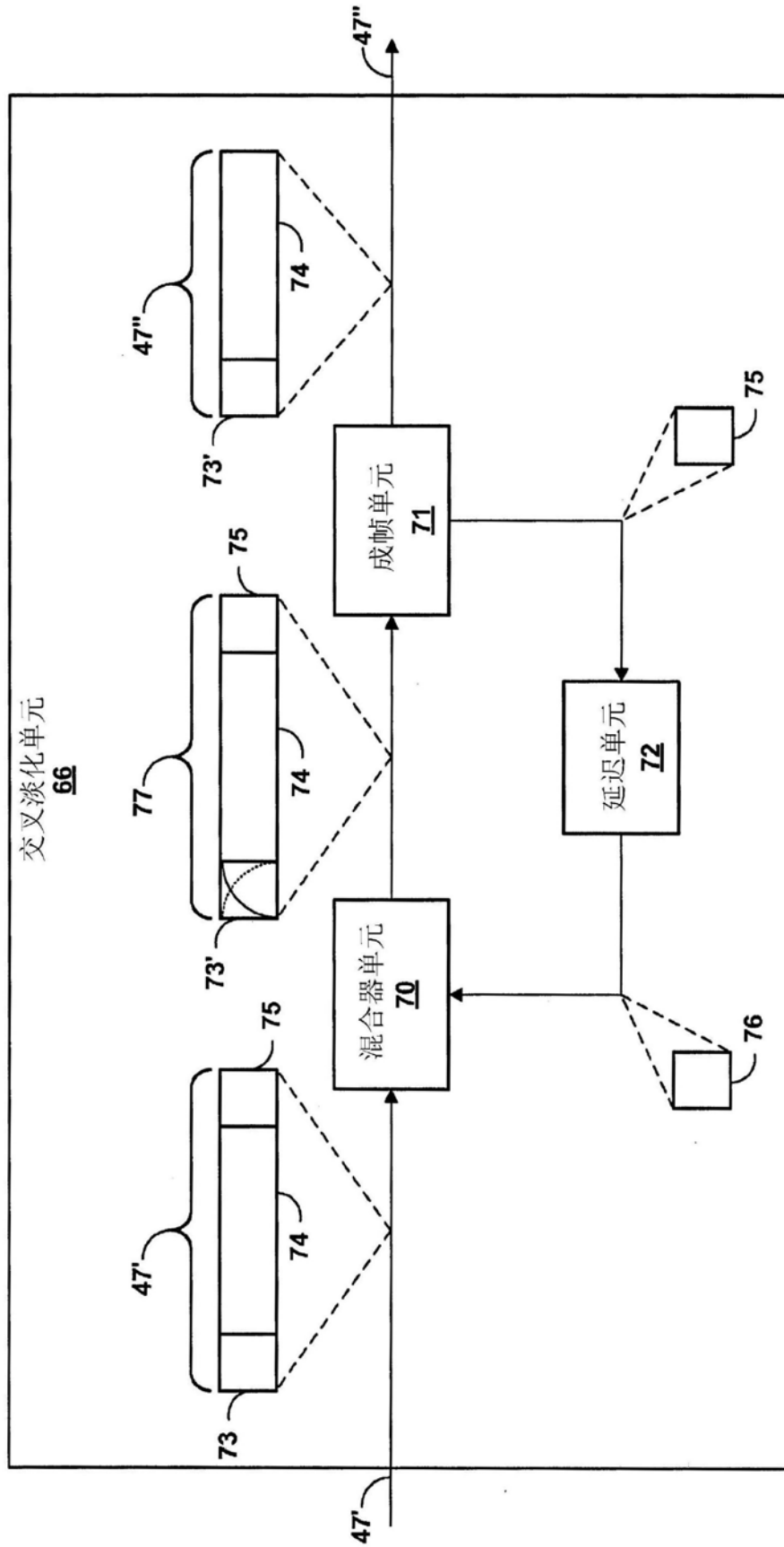


图14