



(12) 发明专利申请

(10) 申请公布号 CN 105247611 A

(43) 申请公布日 2016. 01. 13

(21) 申请号 201480030011. 2

(51) Int. Cl.

(22) 申请日 2014. 05. 23

G10L 19/008(2006. 01)

(30) 优先权数据

61/827, 246 2013. 05. 24 US

(85) PCT国际申请进入国家阶段日

2015. 11. 24

(86) PCT国际申请的申请数据

PCT/EP2014/060727 2014. 05. 23

(87) PCT国际申请的公布数据

W02014/187986 EN 2014. 11. 27

(71) 申请人 杜比国际公司

地址 荷兰阿姆斯特丹

(72) 发明人 海科·普尔哈根 拉尔斯·维尔默斯

利夫·约纳什·萨穆埃尔松

托尼·希尔沃宁

(74) 专利代理机构 北京集佳知识产权代理有限

公司 11227

代理人 杜诚 杨铁成

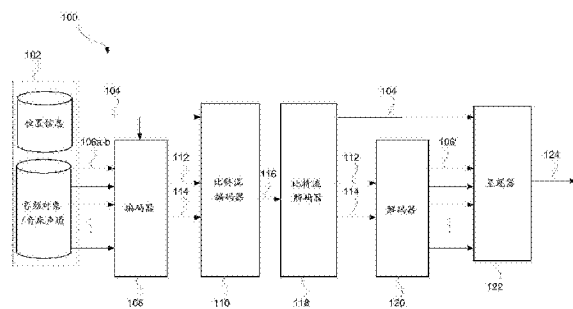
权利要求书3页 说明书11页 附图7页

(54) 发明名称

对音频场景的编码

(57) 摘要

示例实施例提供了编码方法和解码方法,以及相关的编码器和解码器,以对至少包括一个或更多个音频对象(106a)的音频场景进行编码和解码。编码器(108、110)生成包括下混信号(112)和边信息的比特流(116),边信息包括能使得能够在解码器(120)中重构一个或更多个音频对象(106a)的重构矩阵的各个矩阵元素(114)。



1. 一种对至少包括 N 个音频对象的音频场景的时频块进行编码的方法,所述方法包括:

接收所述 N 个音频对象;

基于至少所述 N 个音频对象生成 M 个下混信号;

用矩阵元素生成重构矩阵,所述重构矩阵使得能够根据所述 M 个下混信号重构至少所述 N 个音频对象;以及

生成比特流,所述比特流包括所述 M 个下混信号和所述重构矩阵的所述矩阵元素中的至少一些矩阵元素。

2. 根据权利要求 1 所述的方法,其中,使用第一格式将所述 M 个下混信号布置在所述比特流的第一字段中,并且使用第二格式将所述矩阵元素布置在所述比特流的第二字段中,从而允许仅支持所述第一格式的解码器解码和重放所述第一字段中的所述 M 个下混信号,并且丢弃所述第二字段中的所述矩阵元素。

3. 根据前述权利要求中的任一权利要求所述的方法,还包括步骤:接收对应于所述 N 个音频对象中的每个音频对象的位置数据,其中,基于所述位置数据生成所述 M 个下混信号。

4. 根据前述权利要求中的任一权利要求所述的方法,其中,所述重构矩阵的所述矩阵元素是时变的和频变的。

5. 根据前述权利要求中的任一权利要求所述的方法,其中,所述音频场景还包括多个音床声道,其中,基于至少所述 N 个音频对象和所述多个音床声道生成所述 M 个下混信号。

6. 根据权利要求 5 所述的方法,其中,所述重构矩阵包括使得能够根据所述 M 个下混信号重构所述音床声道的矩阵元素。

7. 根据前述权利要求中的任一权利要求所述的方法,其中,所述音频场景初始包括 K 个音频对象,其中 $K > N$,所述方法还包括步骤:接收所述 K 个音频对象,并且通过将所述 K 个音频对象聚类成 N 个聚类并将每个聚类用一个音频对象代表,来将所述 K 个音频对象减少到所述 N 个音频对象。

8. 根据权利要求 7 所述的方法,还包括步骤:接收对应于所述 K 个音频对象中的每个音频对象的位置数据,其中,将所述 K 个对象聚类成 N 个聚类基于由所述 K 个音频对象的所述位置数据给出的所述 K 个对象之间的位置距离。

9. 根据前述权利要求中的任一权利要求所述的方法,其中,下混信号的所述数目 M 大于 2。

10. 根据前述权利要求中的任一权利要求所述的方法,还包括:

由所述 N 个音频对象形成 L 个辅助信号;

将使得能够根据所述 M 个下混信号和所述 L 个辅助信号重构至少所述 N 个音频对象的矩阵元素包括在所述重构矩阵中;以及

将所述 L 个辅助信号包括在所述比特流中。

11. 根据权利要求 10 所述的方法,其中,所述 L 个辅助信号的至少之一与所述 N 个音频对象之一相同。

12. 根据权利要求 10 至 11 中的任一权利要求所述的方法,其中,所述 L 个辅助信号的至少之一被形成为所述 N 个音频对象中的至少两个音频对象的组合。

13. 根据权利要求 10 至 12 中的任一权利要求所述的方法,其中,所述 M 个下混信号跨超平面,并且其中,所述多个辅助信号的至少之一不位于被所述 M 个下混信号所跨的所述超平面中。

14. 根据权利要求 13 所述的方法,其中,所述多个辅助信号中的所述至少之一与被所述 M 个下混信号所跨的所述超平面正交。

15. 一种计算机可读介质,其包括当在具有处理能力的装置上运行时适于执行根据权利要求 1 至 14 中的任一权利要求所述的方法的计算机代码指令。

16. 一种对至少包括 N 个音频对象的音频场景的时频块进行编码的编码器,所述编码器包括:

接收部件,其被配置成接收所述 N 个音频对象;

下混生成部件,其被配置成接收来自所述接收部件的所述 N 个音频对象,以及基于至少所述 N 个音频对象生成 M 个下混信号;

分析部件,其被配置成用矩阵元素生成重构矩阵,所述重构矩阵使得能够根据所述 M 个下混信号重构至少所述 N 个音频对象;以及

比特流生成部件,其被配置成接收来自所述下混生成部件的所述 M 个下混信号和来自所述分析部件的所述重构矩阵,以及生成包括所述 M 个下混信号和所述重构矩阵的所述矩阵元素中的至少一些矩阵元素的比特流。

17. 一种对至少包括 N 个音频对象的音频场景的时频块进行解码的方法,所述方法包括步骤:

接收包括 M 个下混信号和重构矩阵的至少一些矩阵元素的比特流;

使用所述矩阵元素生成所述重构矩阵;以及

使用所述重构矩阵根据所述 M 个下混信号重构所述 N 个音频对象。

18. 根据权利要求 17 所述的方法,其中,所述 M 个下混信号被使用第一格式布置在所述比特流的第一字段中,并且所述矩阵元素被使用第二格式布置在所述比特流的第二字段中,从而允许仅支持所述第一格式的解码器解码和重放所述第一字段中的所述 M 个下混信号,并且丢弃所述第二字段中的所述矩阵元素。

19. 根据权利要求 17 至 18 中的任一权利要求所述的方法,其中,所述重构矩阵的所述矩阵元素是时变的和频变的。

20. 根据权利要求 17 至 19 中的任一权利要求所述的方法,其中,所述音频场景还包括多个音床声道,所述方法还包括使用所述重构矩阵根据所述 M 个下混信号重构所述音床声道。

21. 根据权利要求 17 至 20 中的任一权利要求所述的方法,其中,下混信号的数目 M 大于 2。

22. 根据权利要求 17 至 21 中的任一权利要求所述的方法,还包括:

接收由所述 N 个音频对象形成的 L 个辅助信号;

使用所述重构矩阵根据所述 M 个下混信号和所述 L 个辅助信号重构所述 N 个音频对象,其中,所述重构矩阵包括使得能够根据所述 M 个下混信号和所述 L 个辅助信号重构至少所述 N 个音频对象的矩阵元素。

23. 根据权利要求 22 所述的方法,其中,所述 L 个辅助信号的至少之一与所述 N 个音频

对象之一相同。

24. 根据权利要求 22 至 23 中的任一权利要求所述的方法,其中,所述 L 个辅助信号的至少之一是所述 N 个音频对象的组合。

25. 根据权利要求 22 至 24 中的任一权利要求所述的方法,其中,所述 M 个下混信号跨超平面,并且其中,所述多个辅助信号的至少之一不位于被所述 M 个下混信号所跨的所述超平面中。

26. 根据权利要求 25 所述的方法,其中,不位于所述超平面中的所述多个辅助信号中的所述至少之一与被所述 M 个下混信号所跨的所述超平面正交。

27. 根据权利要求 17 至 26 中的任一权利要求所述的方法,其中,关于第一频域表示所述 M 个下混信号,并且其中,关于第二频域表示所述重构矩阵,所述第一频域和所述第二频域是相同的频域。

28. 根据权利要求 27 所述的方法,其中,所述第一频域和所述第二频域是改进离散余弦变换 (MDCT) 域。

29. 根据权利要求 17 至 28 中的任一权利要求所述的方法,还包括:接收对应于所述 N 个音频对象的位置数据,以及

使用所述位置数据呈现所述 N 个音频对象以创建至少一个输出音频声道。

30. 根据权利要求 29 所述的方法,其中,关于对应于第二滤波器组的第二频域表示所述重构矩阵,并且在对应于第三滤波器组的第三频域中执行所述呈现,其中,所述第二滤波器组和所述第三滤波器组是至少部分地相同的滤波器组。

31. 根据权利要求 30 所述的方法,其中,所述第二滤波器组和所述第三滤波器组包括正交镜像滤波器 (QMF) 滤波器组。

32. 一种计算机可读介质,其包括当在具有处理能力的装置上运行时适于执行根据权利要求 17 至 31 中的任一权利要求所述的方法的计算机代码指令。

33. 一种对至少包括 N 个音频对象的音频场景的时频块进行解码的解码器,所述解码器包括:

接收部件,其被配置成接收包括 M 个下混信号和重构矩阵的矩阵元素中的至少一些矩阵元素的比特流;

重构矩阵生成部件,其被配置成接收来自所述接收部件的所述矩阵元素,并且基于所述矩阵元素生成所述重构矩阵;以及

重构部件,其被配置成接收来自所述重构矩阵生成部件的所述重构矩阵,并且使用所述重构矩阵根据所述 M 个下混信号重构所述 N 个音频对象。

对音频场景的编码

[0001] 相关申请的交叉引用

[0002] 本申请要求于 2013 年 5 月 24 日提交的美国临时专利申请第 61/827,246 号的优先权,通过引用将该申请整体地合并到本文中。

技术领域

[0003] 本文所公开的发明总体上涉及音频编码和解码领域。特别地,本发明涉及对包括音频对象的音频场景的编码和解码。

背景技术

[0004] 存在用于参数空间音频编码的音频编码系统。例如, MPEG Surround 描述了一种用于多声道音频的参数空间编码的系统。MPEG SAOC(空间音频对象编码)描述了一种用于音频对象的参数编码的系统。

[0005] 在编码器侧,这些系统通常将声道/对象下混成下混,下混通常为单声道(一个声道)或立体声(两个声道)下混,并且提取通过如电平差和互相关来描述声道/对象的性质的边信息。然后对下混和边信息进行编码并且将其发送解码器侧。在解码器侧,在边信息的参数的控制下根据下混来重构即近似估计声道/对象。

[0006] 这些系统的缺点在于重构通常在数学上是复杂的并且经常需要依赖于对由作为边信息发送的参数未明确描述的音频内容的性质的假设。这种假设例如可以是:除非发送了互相关参数,否则声道/对象被认为是不相关的;或者以特定方式生成声道/对象的下混。此外,当下混的声道的数目增加时,数学复杂度和对额外的假设的需要会显著增加。

[0007] 此外,在应用在解码器侧的处理的算法细节中内在地反映出所需要的假设。这意味着在解码器侧必须包括相当多的智能。这是个缺点,因为当解码器被设置在例如很难或甚至不可能升级的消费者装置中时,很难升级和改进算法。

附图说明

[0008] 在下文中,将参考附图并且更加详细地描述示例实施例,其中:

[0009] 图 1 是根据示例实施例的音频编码/解码系统的示意图;

[0010] 图 2 是根据示例实施例的具有遗留解码器的音频编码/解码系统的示意图;

[0011] 图 3 是根据示例实施例的音频编码/解码系统的编码侧的示意图;

[0012] 图 4 是根据示例实施例的编码方法的流程图;

[0013] 图 5 是根据示例实施例的编码器的示意图;

[0014] 图 6 是根据示例实施例的音频编码/解码系统的解码器侧的示意图;

[0015] 图 7 是根据示例实施例的解码方法的流程图;

[0016] 图 8 是根据示例实施例的音频编码/解码系统的解码器侧的示意图;以及

[0017] 图 9 是在根据示例实施例的音频编码/解码系统的解码器侧执行的时频变换的示意图。

[0018] 所有附图都是示意性的,并且一般仅示出为阐明本发明所必须的部分,而可以省略或仅暗示其它部分。除非另有说明,否则相同附图标记在不同附图中的指示相同部件。

具体实施方式

[0019] 考虑到上述内容,目的是提供编码器和解码器,以及提供音频对象的较不复杂的且更灵活的重构的相关方法。

[0020] I. 概述——编码器

[0021] 根据第一方面,示例实施例提出了编码方法、编码器以及用于编码的计算机程序产品。所提出的方法、编码器和计算机程序产品一般可以具有相同特征和优势。

[0022] 根据示例实施例,提供了一种对至少包括 N 个音频对象的音频场景的时频块进行编码的方法。该方法包括:接收 N 个音频对象;基于至少 N 个音频对象生成 M 个下混信号;用矩阵元素生成重构矩阵,重构矩阵使得能够根据 M 个下混信号重构至少 N 个音频对象;以及生成包括 M 个下混信号以及重构矩阵的矩阵元素中的至少一些矩阵元素的比特流。

[0023] 音频对象的数目 N 可以等于或大于 1。下混信号的数目 M 可以等于或大于 1。

[0024] 通过该方法,从而生成了比特流,该比特流包括作为边信息的重构矩阵的矩阵元素中的至少一些矩阵元素以及 M 个下混信号。通过将重构矩阵的各个矩阵元素包括在比特流中,在解码器侧需要非常少的智能。例如,在解码器侧不需要基于所传输的对象参数和额外的假设对重构矩阵进行复杂计算。因此,显著降低了解码器侧的数学复杂度。此外,因为该方法的复杂度不依赖于所使用的下混信号的数目,所以与现有技术方法相比,增加了关于下混信号的数目的灵活性。

[0025] 如本文中所使用的,音频场景一般指如下三维音频环境:其包括与可以被呈现以在音频系统上回放的三维空间中的位置相关联的音频单元。

[0026] 如本文中所使用的,音频对象指音频场景的单元。音频对象通常包括音频信号以及诸如对象在三位空间中的位置的附加信息。附加信息通常被用于在给定的回放系统上最优地呈现音频对象。

[0027] 如本文中所使用的,下混信号指是作为至少 N 个音频对象的组合的信号。诸如音床声道(将在下文中描述)的音频场景的其它信号也可以被组合到下混信号中。例如, M 个下混信号可以对应于对给定扬声器配置,例如标准 5.1 配置的音频场景的呈现。在本文中由 M 表示的下混信号的数目通常(但不必须地)少于音频对象和音床声道的数目之和,这解释了为什么 M 个下混信号称为下混。

[0028] 音频编码/解码系统通常例如通过将适合的滤波器组应用于输入音频信号而将时频空间划分成时频块。时频块的一般意思是对应于时间间隔和频率子带的时频空间的一部分。时间间隔可以通常对应于用在音频编码/解码系统中的时间帧的持续时间。频率子带可以通常对应于由用在编码/解码系统中的滤波器组所定义的一个或若干相邻频率子带。在频率子带对应于由滤波器组定义的若干相邻频率子带的情形下,这允许在音频信号的解码过程中存在不均匀的频率子带,例如,更宽的频率子带用于音频信号的较高频率。在音频编码/解码系统对整个频率范围进行操作的宽波段的情形下,时频块的频率子带可以对应于整个频率范围。上述方法公开了用于在一个这样的时频块期间对音频场景进行编码的编码步骤。然而,要理解的是,可以针对音频编码/解码系统的每个时频块重复该方法。

并且,还要理解的是,可以同时对于若干时频块进行编码。通常,相邻的时频块可以在时间和/或频率上稍稍重叠。例如,时间上的重叠可以相当于重构矩阵的元素在时间上,即从一个时间间隔到下一个时间间隔的线性插值。然而,本公开内容的目标在于编码/解码系统的其它部件,而相邻的时频块之间的时间和/或频率上的任何重叠留给本领域技术人员去实现。

[0029] 根据示例实施例,使用第一格式将 M 个下混信号布置在比特流的第一字段中,并且使用第二格式将矩阵元素布置在比特流的第二字段中,从而允许仅支持第一格式的解码器解码和回放第一字段中的 M 个下混信号并且丢弃第二子段中的矩阵元素。这样做的优势在于比特流中的 M 个下混信号与不用于实现音频对象重构的遗留解码器后向兼容。换言之,遗留解码器仍然可以例如通过将每个下混信号映射到解码器的声道输出来解码和回放比特流的 M 个下混信号。

[0030] 根据示例实施例,该方法还可以包括步骤:接收对应于 N 个音频对象中的每个音频对象的位置数据,其中,基于位置数据生成 M 个下混信号。位置数据通常将每个音频对象与三位空间中的位置相关联。音频对象的位置可以随时间而变化。通过在对音频对象进行下混时使用位置数据,将通过以下方式将音频对象混合到 M 个下混信号中:例如如果在具有 M 个输出声道的系统上听 M 个下混信号,则音频对象听起来就像它们近似地位于其各自的位置。这例如在 M 个下混信号要与遗留解码器后向兼容的情况下是有利的。

[0031] 根据示例实施例,重构矩阵的矩阵元素是时变的和频变的。换言之,重构矩阵的矩阵元素可以对于不同的时频块而不同。以这样的方式,实现了音频对象的重构的极好的灵活性。

[0032] 根据示例实施例,音频场景还包括多个音床声道。这例如在音频内容除了包括音频对象以外还包括音床声道的影院音频应用中是常见的。在这种情形下,可以基于至少 N 个音频对象和多个音床声道生成 M 个下混信号。音床声道的一般意思是对应于三维空间中的固定位置的音频信号。例如,音床声道可以对应于音频编码/解码系统的输出声道之一。这样,音床声道可以被解释为具有三维空间中与音频编码/解码系统的输出扬声器之一的位置相同的相关位置。因此,音床声道可以与仅指示相应输出扬声器的位置的标签相关联。

[0033] 当音频场景包括音床声道时,重构矩阵可以包括使得能够根据 M 个下混信号重构音床声道的矩阵元素。

[0034] 在某些情况下,音频场景可以包括大量的对象。为了降低表现音频场景所需要的复杂度和数据量,可以通过减少音频对象的数量来简化音频场景。因此,如果音频场景初始包括 K 个音频对象,其中 $K > N$,则该方法还可以包括步骤:接收 K 个音频对象,并且通过将 K 个音频对象聚类成 N 个聚类并将每个聚类用一个音频对象表示,来将 K 个音频对象减少到 N 个音频对象。

[0035] 为了简化场景,该方法还可以包括步骤:接收对应于 K 个音频对象中的每个音频对象的位置数据,其中,将 K 个对象聚类成 N 个聚类基于由 K 个音频对象的位置数据所给出的 K 个对象之间的位置距离。例如,三维空间中位置彼此靠近的音频对象可以被聚类在一起。

[0036] 如上所述,该方法的示例实施例在所使用的下混信号的数目方面是灵活的。具体地,当存在多于两个下混信号时,即当 M 大于二时,可以有利地使用该方法。例如,可以使用

对应于常规的 5.1 或 7.1 音频设置的五个或七个下混信号。这么做是有利的,因为与现有技术系统相反,无论使用的下混信号的数目为多少,所提出的编码原则的数学复杂度保持相同。

[0037] 为了能够进一步改进 N 个音频对象的重构,该方法还可以包括:根据 N 个音频对象形成 L 个辅助信号;将矩阵元素包括在使得能够根据 M 个下混信号和 L 个辅助信号来重构至少 N 个音频对象的重构矩阵中;以及将 L 个辅助信号包括在比特流中。因此,辅助信号充当帮助信号,其例如可以捕获很难根据下混信号重构的音频对象的方面。辅助信号还可以基于音床声道。辅助信号的数目可以等于或大于 1。

[0038] 根据一个示例实施例,辅助信号可以对应于特别重要的音频对象,诸如表示对话的音频对象。因此,L 个辅助信号中的至少之一可以与 N 个音频对象之一相同。这使得与必须仅根据 M 个下混声道进行重构的情况相比以更高质量呈现重要的对象。实际上,音频内容提供者可能已经优先化和 / 或标注了音频对象中的一些音频对象作为优选地单独作为辅助对象而被包括的音频对象。此外,这使得呈现之前对这些对象的修改 / 处理较不容易发生伪影。作为比特率和质量之间的折中,也可以发送两个或更多个音频对象的混合以作为辅助信号。换言之,L 个辅助信号中的至少之一可以被形成为 N 个音频对象中的至少两个音频对象的组合。

[0039] 根据一个示例实施例,辅助信号表示在生成 M 个下混信号的过程中丢失的音频对象的信号维度,该丢失例如由于独立对象的数目通常多于下混声道的数目,或者由于两个对象所关联的位置使得该两个对象被混合到同一下混信号中。后一种情形的示例是两个对象仅在纵向上分离而在投影到水平平面上时共享同一位置的情况,这意味着该两个对象通常将被呈现成标准 5.1 环绕扬声器设置的相同下混声道,在标准 5.1 环绕扬声器设置中所有的扬声器都在同一水平平面上。具体地,M 个下混信号跨信号空间中的超平面。通过形成 M 个下混信号的线性组合,仅可以重构位于超平面中的音频信号。为了改进重构,可以包括不位于超平面中的辅助信号,从而也能够重构不位于超平面中的信号。换言之,根据示例实施例,多个辅助信号中的至少之一不位于被 M 个下混信号所跨的超平面中。例如,多个辅助信号中的至少之一可以与被 M 个下混信号所跨的超平面正交。

[0040] 根据示例实施例,提供了一种包括当在具有处理能力的装置上运行时适于执行第一方面的任何方法的计算机代码指令的计算机可读介质。

[0041] 根据示例实施例,提供了一种对至少包括 N 个音频对象的音频场景的时频块进行编码的编码器,该编码器包括:接收部件,被配置成接收 N 个音频对象;下混生成部件,被配置成接收来自接收部件的 N 个音频对象并且基于至少 N 个音频对象生成 M 个下混信号;分析部件,被配置成用矩阵元素生成重构矩阵,重构矩阵使得能够根据 M 个下混信号重构至少 N 个音频对象;以及比特流生成部件,被配置成接收来自下混生成部件的 M 个下混信号以及来自分析部件的重构矩阵,并且生成包括 M 个下混信号和重构矩阵的矩阵元素中的至少一些矩阵元素的比特流。

[0042] II. 概述——解码器

[0043] 根据第二方面,示例实施例提出了解码方法、解码装置和用于解码的计算机程序产品。所提出的方法、装置和计算机程序产品一般可以具有相同特征和优势。

[0044] 与在上述编码器的概述中呈现的特征和设置有关的优势可以一般对解码器的相

应特征和设置有效。

[0045] 根据示例实施例,提供了一种对至少包括 N 个音频对象的音频场景的时频块进行解码的方法,该方法包括步骤:接收包括 M 个下混信号和重构矩阵的矩阵元素中的至少一些矩阵元素的比特流;使用矩阵元素生成重构矩阵;以及使用重构矩阵根据 M 个下混信号重构 N 个音频对象。

[0046] 根据示例实施例,使用第一格式将 M 个下混信号布置在比特流的第一字段中,并且使用第二格式将矩阵元素布置在比特流的第二子段中,从而允许仅支持第一格式的解码器解码和回放第一字段中的 M 个下混信号并且丢弃第二子段中的矩阵元素。

[0047] 根据示例实施例,重构矩阵的矩阵元素是时变的和频变的。

[0048] 根据示例实施例,音频场景还包括多个音床声道,该方法还包括使用重构矩阵根据 M 个下混信号来重构音床声道。

[0049] 根据示例实施例,下混信号的数目 M 大于 2。

[0050] 根据示例实施例,该方法还包括:接收由 N 个音频对象形成的 L 个辅助信号;使用重构矩阵根据 M 个下混信号和 L 个辅助信号重构 N 个音频对象,其中,重构矩阵包括使得能够根据 M 个下混信号和 L 个辅助信号重构至少 N 个音频对象的矩阵元素。

[0051] 根据示例实施例, L 个辅助信号的至少之一与 N 个音频对象之一相同。

[0052] 根据示例实施例, L 个辅助信号的至少之一是 N 个音频对象的组合。

[0053] 根据示例实施例, M 个下混信号跨超平面,并且其中多个辅助信号的至少之一不位于被 M 个下混信号所跨的超平面中。

[0054] 根据示例实施例,不位于超平面中的多个辅助信号的至少之一正交于被 M 个下混信号所跨的超平面。

[0055] 如上所述,音频编码/解码系统通常在频域中工作。因此,音频编码/解码系统使用滤波器组执行音频信号的时频变换。可以使用不同类型的时频变换。例如,可以关于第一频域来表示 M 个下混信号并且可以关于第二频域来表示重构矩阵。为了减少解码器的计算负担,以聪明的方式选择第一频域和第二频域是有利的。例如,第一频域和第二频域可以被选择成相同的频域,诸如改进离散余弦变换(MDCF)域。以这种方式,可以避免在解码器中将 M 个下混信号从第一频域变换到时域然后变换到第二频域。可替代地,能够通过以下方式选择第一频域和第二频域:可以共同实现从第一频域到第二频域的变换,使得在第一频域与第二频域之间没有必要通过时域。

[0056] 该方法还可以包括接收对应于 N 个音频对象的位置数据,并且使用位置数据呈现 N 个音频对象以创建至少一个输出音频声道。以这种方式,基于重构的 N 个音频对象在三维空间中的位置将其映射到音频编码器/解码器系统的输出声道上。

[0057] 优选地在频域中执行呈现。为了减少解码器的计算负担,优选地以聪明的方式关于重构音频对象的频域来选择呈现的频域。例如,如果关于对应于第二滤波器组的第二频域表示重构矩阵,并且在对应于第三滤波器组的第三频域中执行呈现,则优选地将第二滤波器组和第三滤波器组选择成至少部分地为相同的滤波器组。例如,第二滤波器组和第三滤波器组可以包括正交镜像滤波器(QMF)域。可替代地,第二频域和第三频域可以包括 MDCT 滤波器组。根据示例实施例,第三滤波器组可以由一系列滤波器组组成,诸如 QMF 滤波器组,后接奈奎斯特滤波器组。如果这样,则序列的滤波器组中至少之一(序列的第一滤波

器组)与第二滤波器组相同。以这种方式,可以说第二滤波器组和第三滤波器组至少部分地为相同的滤波器组。

[0058] 根据示例实施例,提供了包括当在具有处理能力的装置上运行时适于执行第二方面的任一方法的计算机代码指令的计算机可读介质。

[0059] 根据示例实施例,提供了一种对至少包括N个音频对象的音频场景的时频块进行解码的解码器,该解码器包括:接收部件,被配置成接收包括M个下混信号和重构矩阵的矩阵元素中的至少一些矩阵元素的比特流;重构矩阵生成部件,被配置成接收来自接收部件的矩阵元素,并且基于矩阵元素生成重构矩阵;以及重构部件,被配置成接收来自重构矩阵生成部件的重构矩阵,并且使用重构矩阵根据M个下混信号重构N个音频对象。

[0060] III. 示例实施例

[0061] 图1图示出对音频场景102进行编码/解码的编码/解码系统100。编码/解码系统100包括编码器108、比特流生成部件110、比特流解码部件118、解码器120以及呈现器122。

[0062] 音频场景102由一个或更多个音频对象106a(诸如N个音频对象)即音频信号来表示。音频场景102还可以包括一个或更多个音床声道106b,即直接对应于呈现器122的输出声道之一的信号。音频场景102还由包括位置信息104的元数据来表示。在呈现音频场景102时例如由呈现器122使用位置信息104。位置信息104可以将音频对象106a以及可能还有音床声道106b与三维空间中的空间位置关联起来以作为时间的函数。元数据还可以包括对于呈现音频场景102有用的其它类型的数据。

[0063] 系统100的编码部分包括编码器108和比特流生成部件110。编码器108接收音频对象106a、音床声道106b(如果存在),以及包括位置信息104的元数据。基于此,编码器108生成一个或更多个下混信号112,诸如M个下混信号。举例来说,下混信号112可以对应于5.1音频系统的声道[Lf Rf Cf Ls Rs LFE]。(“L”代表左,“R”代表右,“C”代表中央,“f”代表前,“s”代表环绕,并且“LFE”代表低频效果)。

[0064] 编码器108还生成边信息。边信息包括重构矩阵。重构矩阵包括使得能够根据下混信号112重构至少音频对象106a的矩阵元素114。重构矩阵还可以使得能够重构音床声道106b。

[0065] 编码器108将M个下混信号112以及矩阵元素114中的至少一些矩阵元素传输到比特流生成部件110。比特流生成部件110通过执行量化和编码来生成包括M个下混信号112和矩阵元素114中的至少一些矩阵元素的比特流116。比特流生成部件110还接收包括位置信息104的元数据,以包括在比特流116中。

[0066] 系统的解码部分包括比特流解码部件118和解码器120。比特流解码部件118接收比特流116,并且执行解码和去量化(dequantization)以提取M个下混信号112和包括重构矩阵的至少一些矩阵元素114的边信息。M个下混信号112和矩阵元素114随后被输入到解码器120,解码器120基于下混信号112和矩阵元素114生成N个音频对象106a以及很可能还有音床声道106b的重构106'。因此,N个音频对象的重构106'是N个音频对象106a以及很可能还有音床声道106b的近似。

[0067] 举例来说,如果下混信号112对应于5.1配置的声道[Lf Rf Cf Ls Rs LFE],则解码器120可以仅使用全波段声道[Lf Rf Cf Ls Rs]来重构对象106',从而忽略LFE。这同

样适用于其它声道配置。可以将下混 112 的 LFE 声道（基本未修改）发送到呈现器 122。

[0068] 重构的音频对象 106' 以及位置信息 104 随后被输入到呈现器 122。基于重构的音频对象 106' 和位置信息 104, 呈现器 122 呈现具有适合于在期望的扬声器或耳机配置上回放的格式的输出信号 124。典型的输出格式是标准 5.1 环绕设置 (3 个前扬声器、2 个环绕扬声器以及 1 个低频效果 (LFE) 扬声器) 或者 7.1+4 设置 (3 个前扬声器、4 个环绕扬声器、1 个 LFE 扬声器以及 4 个高架扬声器)。

[0069] 在一些实施例中, 原始音频场景可以包括大量的音频对象。对大量的音频对象进行处理的代价是高计算复杂度。并且, 要嵌入到比特流 116 中的边信息量 (位置信息 104 和重构矩阵元素 114) 取决于音频对象的数目。通常, 边信息量随音频对象的数目线性地增长。因此, 为了节省计算复杂度和 / 或为了降低对音频场景进行编码所需要的比特率, 在编码之前减少音频对象的数目是有利的。为此, 音频编码器 / 解码器系统 100 还可以包括设置在编码器 108 上游的场景简化模块 (未示出)。场景简化模块将原始音频对象以及很可能还有音床声道作为输入, 并且执行处理以输出音频对象 106a。场景简化模块通过执行聚类而将原始音频对象的数目例如 K 减少到音频对象 106a 的更合适的数目 N。更确切地, 场景简化模块将 K 个原始音频对象以及很可能还有音床声道组织成 N 个聚类。通常, 基于 K 个原始音频对象 / 音床声道在音频场景中的空间接近度来定义聚类。为了确定空间接近度, 场景简化模块可以将原始音频对象 / 音床声道的位置信息作为输入。当场景简化模块已经形成了 N 个聚类时, 其继续执行以将每个聚类用一个音频对象代表。例如, 代表聚类的音频对象可以被形成形成聚类的一部分的音频对象 / 音床声道之和。更具体地, 可以添加音频对象 / 音床声道的音频内容以生成代表性音频对象的音频内容。此外, 可以对聚类中音频对象 / 音床声道的位置取平均, 以给出代表性音频对象的位置。场景简化模块将代表性音频对象的位置包括在位置数据 104 中。此外, 场景简化模块输出构成图 1 的 N 个音频对象 106a 的代表性音频对象。

[0070] 可以使用第一格式将 M 个下混信号 112 布置在比特流 116 的第一字段中。可以使用第二格式将矩阵元素 114 布置在比特流 116 的第二字段中。以这种方式, 仅支持第一格式的解码器能够解码和回放第一字段中的 M 个下混信号 112, 并且丢弃第二字段中的矩阵元素 114。

[0071] 图 1 的音频编码器 / 解码器系统 100 支持第一格式和第二格式。更确切地, 解码器 120 被配置成解译第一格式和第二格式, 这意味着其能够基于 M 个下混信号 112 和矩阵元素 114 来重构对象 106'。

[0072] 图 2 图示出音频编码器 / 解码器系统 200。系统 200 的编码部分 108、110 对应于图 1 的编码部分。然而, 音频编码器 / 解码器系统 200 的解码部分与图 1 的音频编码器 / 解码器系统 100 的解码部分不同。音频编码器 / 解码器系统 200 包括支持第一格式但不支持第二格式的遗留解码器 230。因此, 音频编码器 / 解码器系统 200 的遗留解码器 230 不能够重构音频对象 / 音床声道 106a 到 106b。然而, 因为遗留解码器 230 支持第一格式, 所以其仍可以对 M 个下混信号 112 进行解码以生成输出 224, 输出 224 是适合于通过相应的多声道扬声器设置实现直接回放的基于声道的表示, 诸如 5.1 表示。下混信号的这个性质称为后向兼容, 后向兼容意味着不支持第二格式, 即不能够解译包括矩阵元素 114 的边信息的遗留解码器也可以解码和回放 M 个下混信号 112。

[0073] 现在将参考图 3 和图 4 的流程图更详细地描述音频编码 / 解码系统 100 的编码器侧的操作。

[0074] 图 4 更详细地图示出图 1 的编码器 108 和比特流生成部件 110。编码器 108 具有接收部件 (未示出)、下混生成部件 318 和分析部件 328。

[0075] 在步骤 E02 中, 编码器 108 的接收部件接收 N 个音频对象 106a 和音床声道 106b (如果存在)。编码器 108 还可以接收位置数据 104。使用向量标记, N 个音频对象可以由向量 $S = [S_1 S_2 \dots S_N]^T$ 表示, 并且音床声道由向量 B 表示。N 个音频对象和音床声道可以一起由向量 $A = [B^T S^T]^T$ 表示。

[0076] 在步骤 E04 中, 下混生成部件 318 根据 N 个音频对象 106a 和音床声道 106b (如果存在) 生成 M 个下混信号 112。使用向量标记, M 个下混信号可以由包括 M 个下混信号的向量 $D = [D_1 D_2 \dots D_M]^T$ 表示。一般多个信号的下混是信号的组合, 诸如信号的线性组合。举例来说, M 个下混信号可以对应于特定的扬声器配置, 诸如 5.1 扬声器配置中的扬声器 [Lf Rf Cf Ls Rs LFE] 的配置。

[0077] 下混生成部件 318 在生成 M 个下混信号时可以使用位置信息 104, 使得基于各对象在三维空间中的位置将这些对象组合成不同的下混信号。当 M 个下混信号本身如同上述示例中那样对应于特定扬声器配置时, 这是特别相关的。举例来说, 下混生成部件 318 可以基于位置信息得出表示矩阵 Pd (对应于应用在图 1 的呈现器 122 中的表示矩阵), 并且使用该表示矩阵根据 $D = pd * [B^T S^T]^T$ 生成下混。

[0078] N 个音频对象 106a 和音床声道 106b (如果存在) 也被输入到分析部件 328。分析部件 328 通常对输入音频信号 106a、106b 的时频块进行操作。为此, 可以将 N 个音频对象 106a 和音床声道 106b 馈送过对输入音频信号 106a、106b 执行时间到频率变换的滤波器组, 即 QMF 组。特别地, 滤波器组 338 与多个频率子带相关联。时频块的频率分辨率对应于这些频率子带中的一个或更多个。时频块的频率分辨率可以是不均匀的, 即其可以对频率变化。例如, 低频率分辨率可以用于高频, 这意味着高频范围内的时频块可以对应于由滤波器组 338 定义的若干频率子带。

[0079] 在步骤 E06 中, 分析部件 328 生成在本文中由 R1 表示的重构矩阵。生成的重构矩阵由多个矩阵元素组成。重构的矩阵 R1 使得能够在解码器中根据 M 个下混信号 112 重构 (近似) N 个音频对象 106a 以及很可能还有音床声道 106b。

[0080] 分析部件 328 可以采取不同的方法来生成重构矩阵。例如, 可以使用将 N 个音频对象 106a / 音床声道 106b 以及 M 个下混信号 112 作为输入的最小均方误差 (MMSE) 预测方法。可以将该方法描述成旨在得出能够最小化重构的音频对象 / 音床声道的均方误差的重构矩阵的方法。特别地, 该方法使用候选重构矩阵来重构 N 个音频对象 / 音床声道, 并且关于均方误差将音频对象 / 音床声道与输入音频对象 106a / 音床声道 106b 进行比较。将最小化均方误差的候选重构矩阵选作重构矩阵, 并且其矩阵元素 114 是分析部件 328 的输出。

[0081] MMSE 方法需要对 N 个音频对象 106a / 音床声道 106b 以及 M 个下混信号 112 的相关矩阵和协方差矩阵进行估计。根据上述方法, 基于 N 个音频对象 106a / 音床声道 106b 和 M 个下混信号 112 来测量这些相关矩阵和协方差矩阵。在替选的基于模型的方法中, 分析部件 328 将位置数据 104 而不是 M 个下混信号 112 作为输入。通过做出某些假设, 例如假设 N 个音频对象互不相关, 并且使用该假设并结合应用在下混生成部件 318 中的下混规则, 分

析部件 328 可以计算出执行上述 MMSE 方法所需要的所需相关矩阵和协方差矩阵。

[0082] 重构矩阵的元素 114 和 M 个下混信号 112 随后输入到比特流生成部件 110。在步骤 E108 中,比特流生成部件 110 对 M 个下混信号 112 和重构矩阵的至少一些矩阵元素 114 进行量化和编码,并且将它们布置在比特流 116 中。特别地,比特流生成部件 110 可以使用第一格式将 M 个下混信号 112 布置在比特流 116 的第一字段中。此外,比特流生成部件 110 可以使用第二格式将矩阵元素 114 布置在比特流 116 的第二字段中。如前面参考图 2 所描述的,这允许仅支持第一格式的遗留解码器解码和回放 M 个下混信号 112 并且丢弃第二字段中的矩阵元素 114。

[0083] 图 5 图示出编码器 108 的替选实施例。与图 3 中示出的编码器相比,图 5 的编码器 508 还使得一个或更多个辅助信号能够被包括在比特流 116 中。为此,编码器 508 包括辅助信号生成部件 548。辅助信号生成部件 548 接收音频对象 106a/ 音床声道 106b,并且基于音频对象 106a/ 音床声道 106b 生成一个或更多个辅助信号 512。辅助信号生成部件 548 例如可以生成辅助信号 512 以作为音频对象 106a/ 音床声道 106b 的组合。用向量 $C = [C_1 \ C_2 \ \dots \ C_L]^T$ 来表示辅助信号,辅助信号可以被生成为 $C = Q * [B^T \ S^T]^T$,其中, Q 为可以是时变和频变的矩阵。这包括辅助信号等于音频对象中的一个或更多个音频对象的情形以及辅助信号是音频对象的线性组合的情形。例如,辅助信号可以代表一个特别重要的对象,诸如对话。

[0084] 辅助信号 512 的作用是改善解码器中音频对象 106a/ 音床声道 106b 的重构。更具体地,在解码器侧,可以基于 M 个下混信号 112 以及 L 个辅助信号 512 来重构音频对象 106a/ 音床声道 106b。因此,重构矩阵将包括能够根据 M 个下混信号 112 以及 L 个辅助信号重构音频对象 / 音床声道的矩阵元素 114。

[0085] 因此, L 个辅助信号 512 可以被输入到分析部件 328,使得在生成重构矩阵时考虑到 L 个辅助信号 512。分析部件 328 也可以将控制信号发送至辅助信号生成部件 548。例如,分析部件 328 可以控制哪些音频对象 / 音床声道包括在辅助信号中以及它们是如何被包括的。特别地,分析部件 328 可以控制 Q 矩阵的选择。该控制例如可以基于上述 MMSE 方法,使得可以选择辅助信号以使得重构的音频对象 / 音床声道与音频对象 106a/ 音床声道 106b 尽可能地接近。

[0086] 现在将参考图 6 和图 7 的流程图更加详细地描述音频编码 / 解码系统 100 的解码器侧的操作。

[0087] 图 6 更具体地图示出图 1 的比特流解码部件 118 和解码器 120。解码器 120 包括重构矩阵生成部件 622 和重构部件 624。

[0088] 在步骤 D02 中,比特流解码部件 118 接收比特流 116。比特流解码部件 118 对比特流 116 中的信息进行解码和去量化,以提取 M 个下混信号 112 以及重构矩阵中的至少一些矩阵元素 114。

[0089] 重构矩阵生成部件 622 接收矩阵元素 114 并且在步骤 D04 中继续进行以生成重构矩阵 614。重构矩阵生成部件 622 通过将矩阵元素 114 布置在矩阵中的适当位置来生成重构矩阵 614。如果没有接收到重构矩阵的全部矩阵元素,重构矩阵生成部件 622 例如可以插入零来代替缺少的元素。

[0090] 重构矩阵 614 和 M 个下混信号随后被输入到重构部件 624。重构部件 624 随后在

步骤 D06 中重构 N 个音频对象,并且如果可以,重构音床声道。换言之,重构部件 624 生成 N 个音频对象 106a/音床声道 106b 的近似 106'。

[0091] 举例来说,M 个下混信号可以对应于特定的扬声器配置,诸如 5.1 扬声器配置中的扬声器 [Lf Rf Cf Ls Rs LFE] 的配置。如果这样,重构部件 624 可以使得对象 106' 的重构仅基于对应于扬声器配置的全波段声道的下混信号。如上文所解释的,带限信号(低频 LFE 信号)可以基本未修改地被发送到呈现器。

[0092] 重构部件 624 通常在频域中工作。更确切地,重构部件 624 对输入信号的各个时频块进行操作。因此,在输入到重构部件 624 之前,M 个下混信号 112 通常经受时间到频率变换 623。时间到频率变换 623 通常与在编码器侧应用的变换 338 相同或相似。例如,时间到频率变换 623 可以是 QMF 变换。

[0093] 为了重构音频对象/音床声道 106',重构部件 624 应用矩阵操作。更具体地,使用先前引入的标记,重构部件 624 可以将音频对象/音床声道的近似 A' 生成为 $A' = R1 * D$ 。重构矩阵 R1 可以根据时间和频率变化。因此,重构矩阵在由重构部件 624 处理的不同的时频块之间可以不同。

[0094] 在从解码器 120 输出之前,重构的音频对象/音床声道 106' 通常被变换回时域 625。

[0095] 图 8 图示出当比特流 116 额外地包括辅助信号时的情况。与图 7 的实施例相比,比特流解码部件 118 现在额外地对来自比特流 116 的一个或多个辅助信号 512 进行解码。辅助信号 512 被输入到重构部件 624,辅助信号 512 在重构部件 624 处被包括在音频对象/音床声道的重构中。更具体地,重构部件 624 通过应用矩阵运算 $A' = R1 * [D^T \ C^T]^T$ 生成音频对象/音床声道。

[0096] 图 9 图示出在图 1 的音频编码/解码系统 100 的解码器侧使用的不同的时频变换。比特流解码部件 118 接收比特流 116。解码和去量化部件 918 对比特流 116 进行解码和去量化,以提取位置信息 104、M 个下混信号 112 和重构矩阵的矩阵元素 114。

[0097] 在该阶段,通常在第一频域中表示 M 个下混信号 112,第一频域对应于在本文中由 T/F_c 和 F/T_c 表示以分别用于从时域到第一频域的变换和从第一频域到时域的变换的第一组时频滤波器组。通常,对应于第一频域的滤波器组可以实现重叠窗变换,诸如 MDCT 和反 MDCT。比特流解码部件 118 可以包括通过使用滤波器组 F/T_c 将 M 个下混信号 112 变换到时域的变换部件 901。

[0098] 解码器 120,尤其是重构部件 624 通常关于第二频域处理信号。第二频域对应于在本文中由 T/F_0 和 F/T_0 表示的分别用于从时域到第二频域的变换和从第二频域到时域的变换的第二组时频滤波器组。因此,解码器 120 可以包括通过使用滤波器组 T/F_0 将在时域中表示的 M 个下混信号 112 变换到第二频域的变换部件 903。当重构部件 624 已经通过在第二频域中执行处理而基于 M 个下混信号重构对象 106' 时,变换部件 905 可以通过使用滤波器组 F/T_0 将重构对象 106' 变换回时域。

[0099] 呈现器 122 通常关于第三频域处理信号。第三频域对应于在本文中由 T/F_R 和 F/T_R 表示的分别用于从时域到第三频域的变换以及从第三频域到时域的变换的第三组时频滤波器组。因此,呈现器 122 可以包括通过使用滤波器组 T/F_R 将重构的音频对象 106' 从时域变换到第三频域的变换部件 907。当呈现器 122 通过呈现部件 922 已经呈现输出声道

124 时,可以由变换部件 909 通过使用滤波器组 F/T_R 将输出声道变换到时域。

[0100] 从以上描述显而易见,音频编码/解码系统的解码器侧包括许多时频变换步骤。然而,如果以一定方式选择第一频域、第二频域和第三频域,则时频变换步骤中的一些步骤会变得冗余。

[0101] 例如,可以将第一频域、第二频域和第三频域中的一些选择成为一样的,或者可以共同地实现为从一个频域直接到另一频域而不通过它们之间的时域。后者一个示例是以下情形:第二频域和第三频域的不同仅在于呈现器 122 中的变换部件 907 除了使用两个变换部件 905 和 907 共同的 QMF 滤波器组以外还使用奈奎斯特滤波器组以提高低频处的频率分辨率。在这种情形下,可以以奈奎斯特滤波器组的形式共同实现变换部件 905 和 907,从而节省计算复杂度。

[0102] 在另一示例中,第二频域和第三频域是相同的。例如,第二频域和第三频域可以都是 QMF 频域。在这种情形下,变换部件 905 和 907 是冗余的并且可以被去除,从而节省计算复杂度。

[0103] 根据另一示例,第一频域和第二频域可以是相同的。例如,第一频域和第二频域可以都是 MDCT 域。在这种情形下,可以去除第一变换部件 901 和第二变换部件 903,从而节省计算复杂度。

[0104] 等同物、扩展、替选方案以及其它

[0105] 本领域技术人员在研究以上描述之后将会明白本公开内容的其它实施例。虽然本说明书和附图公开了实施例和示例,但本公开内容不限于这些具体示例。在不偏离由所附权利要求所定义的本公开内容的范围的情况下可以做出许多修改和变形。在权利要求中出现的任何附图标记不被理解为限制它们的范围。

[0106] 另外,根据对附图、公开内容和所附权利要求的研究,本领域技术人员在实践本公开内容时可以理解并实现对所公开的实施例的变型。在权利要求书中,词语“包括”不排除其它元件或步骤,并且不定冠词“一”不排除复数形式。在相互不同的从属权利要求中引述某些措施的事实不指示不可以使用这些措施的组合来获利。

[0107] 在上文中公开的系统和方法可以实现为软件、固件、硬件或者它们的组合。在硬件实现中,以上描述中提到的功能单元之间的任务划分不一定对应于实体单元的划分;相反地,一个物理部件可以具有多个功能,并且可以由若干物理部件共同执行一个任务。某些部件或全部部件可以被实现为由数字信号处理器或微处理器执行的软件,或者可以被实现为硬件或专用集成电路。这样的软件可以分布在可以包括计算机存储介质(或非暂态介质)和通信介质(或暂态介质)的计算机可读介质上。如为本领域技术人员所熟知的,术语计算机存储介质包括以任何方法或用于存储诸如计算机可读指令、数据结构、程序模块或其它数据的信息的技术实现的易失性介质和非易失性介质,可移除介质和不可移除介质。计算机存储介质包括但不限于 RAM、ROM、EEPROM、闪速存储器或其它存储器技术、CD-ROM、数字多功能盘(DVD)或其它光盘存储装置、磁盒、磁带、磁盘存储器或其它磁存储装置、或者可用于存储期望信息并且可被计算机访问的任何其它介质。此外,技术人员熟知的是,通信介质通常包含计算机可读指令、数据结构、程序模块或诸如载波的调制数据信号中的其它数据,或者其它传输机制,并且包括任何信息传输介质。

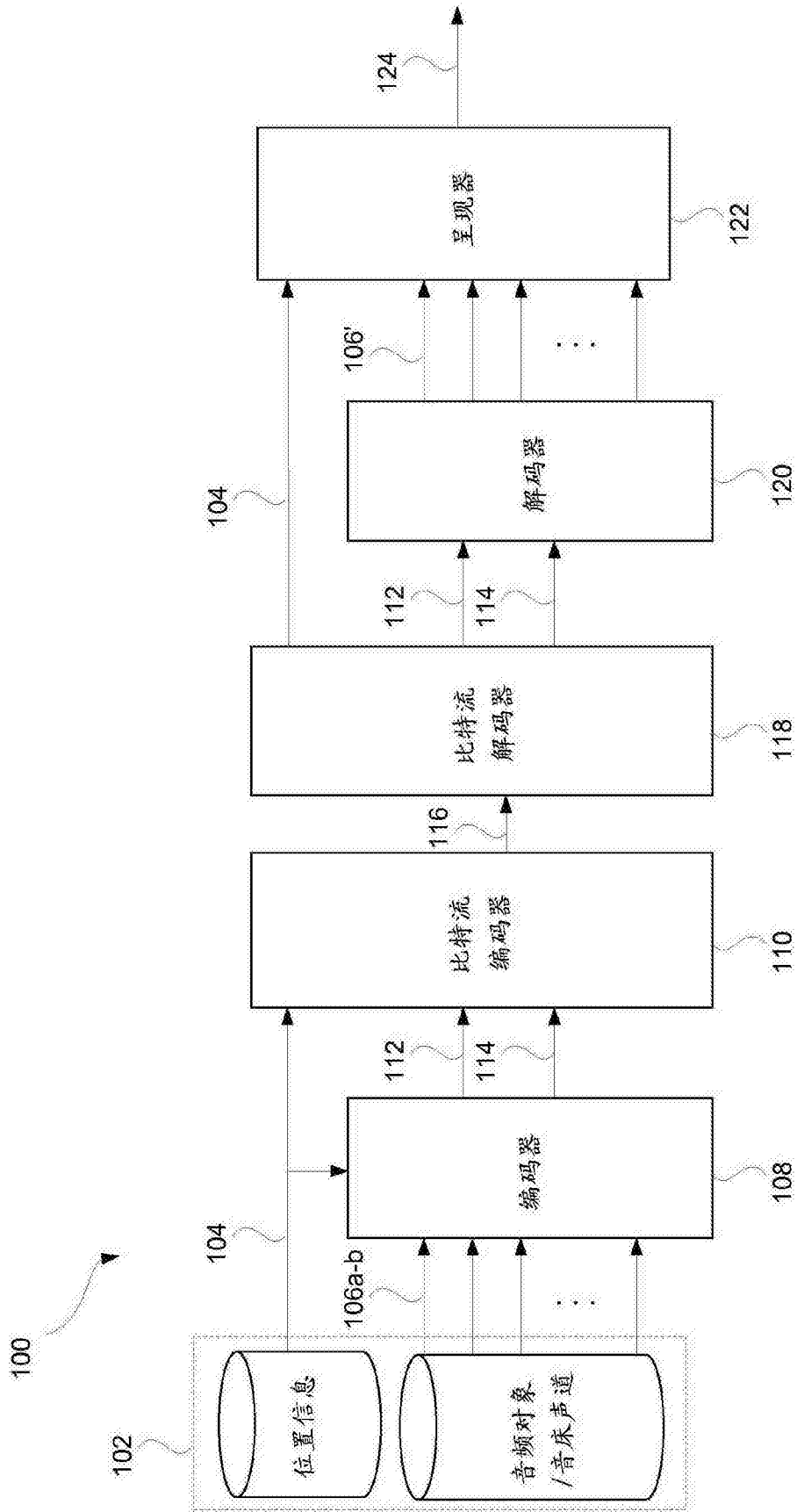


图 1

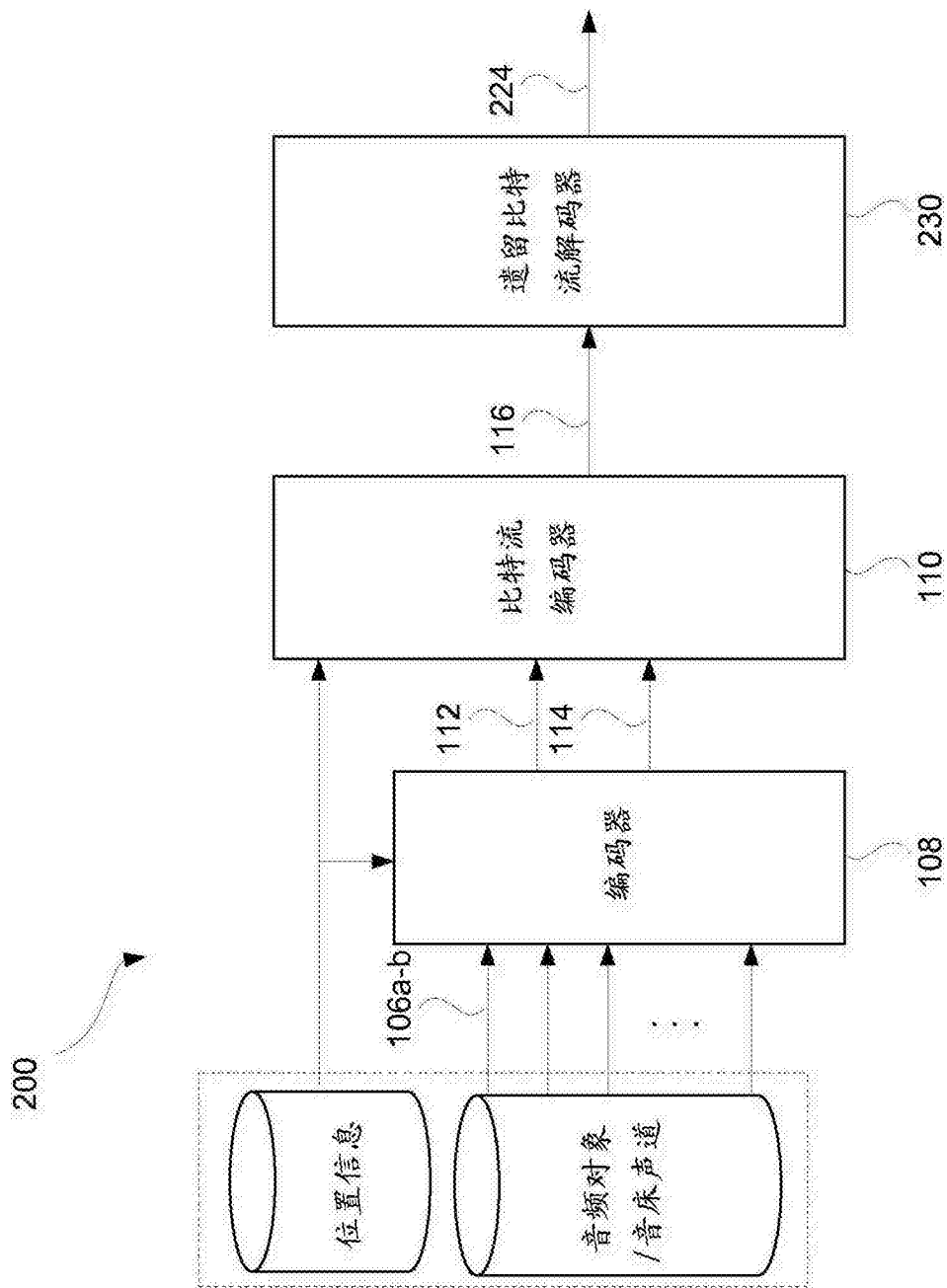


图 2

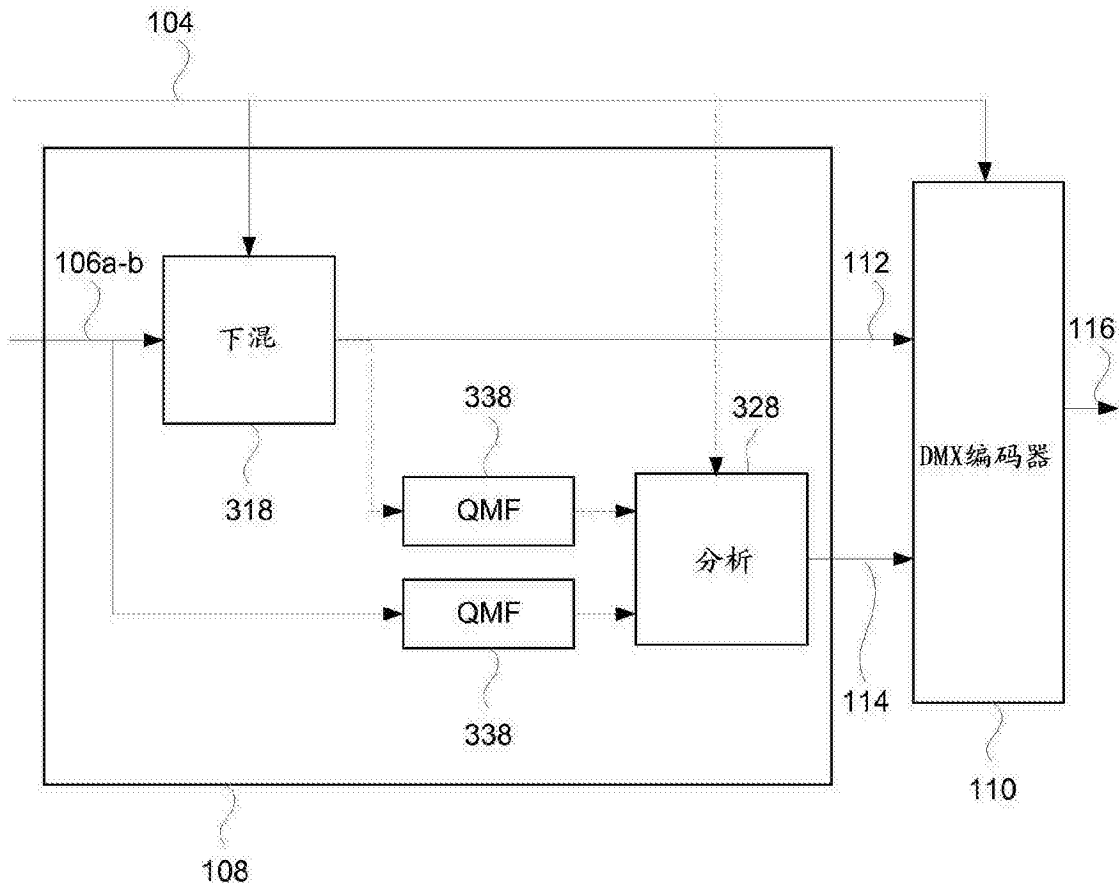


图 3

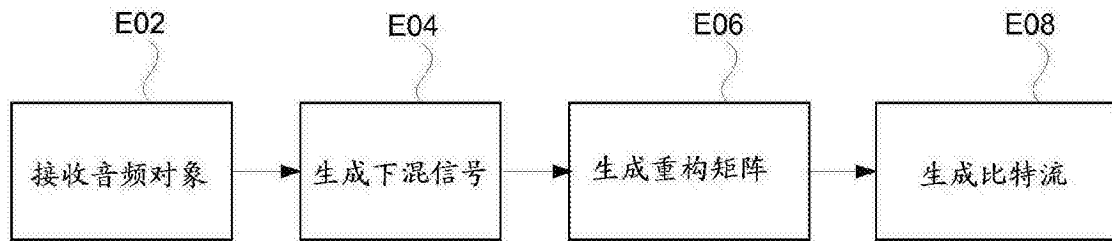


图 4

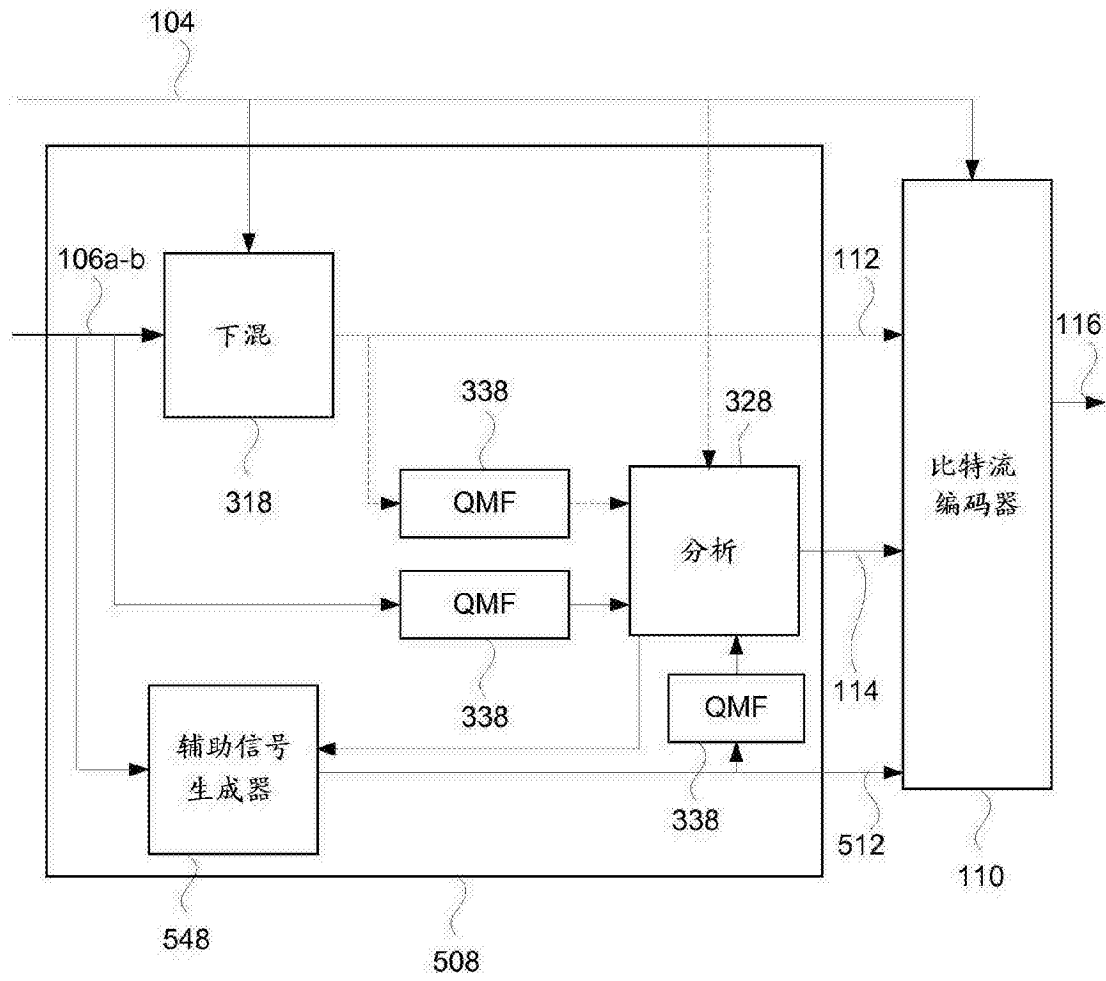


图 5

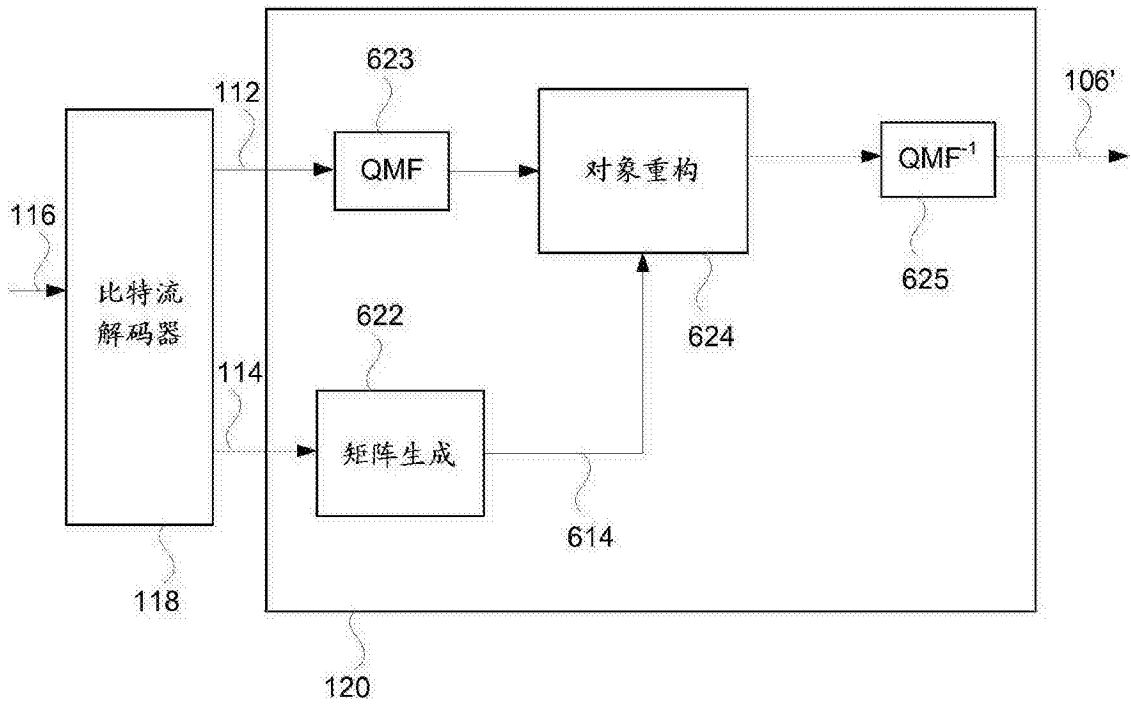


图 6

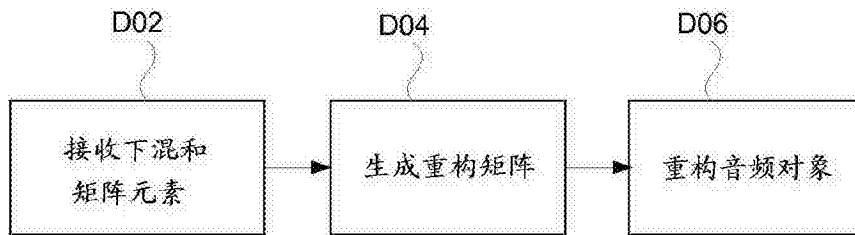


图 7

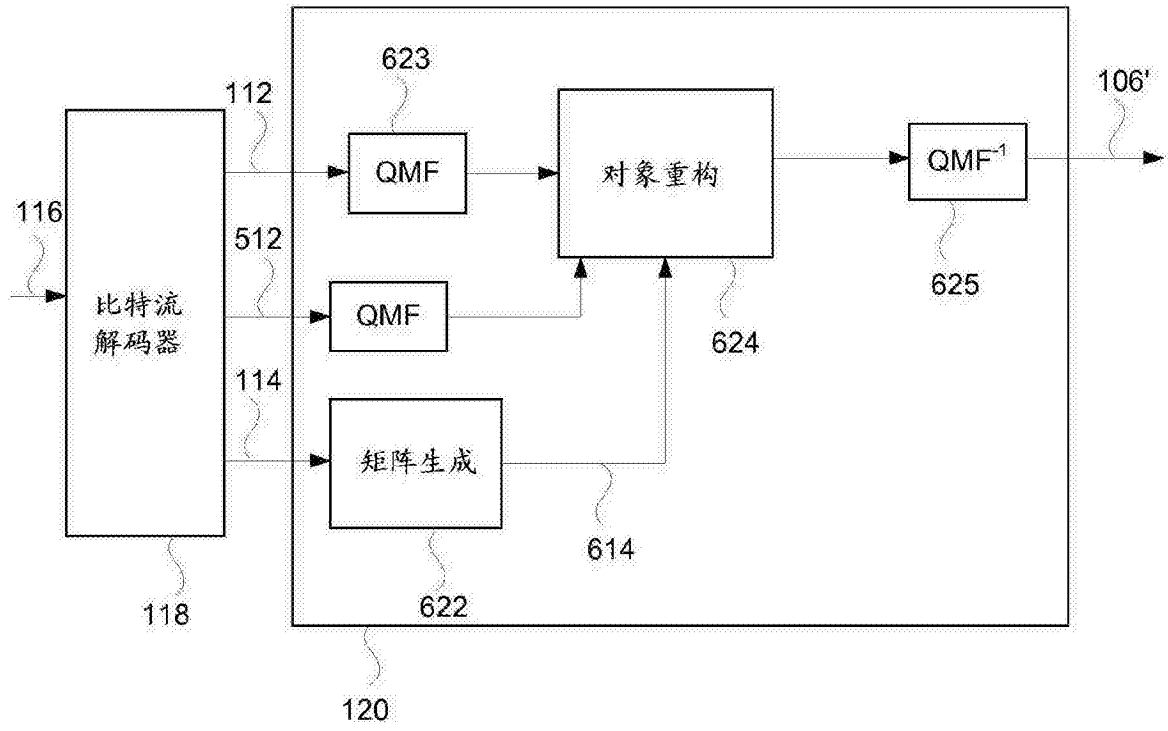


图 8

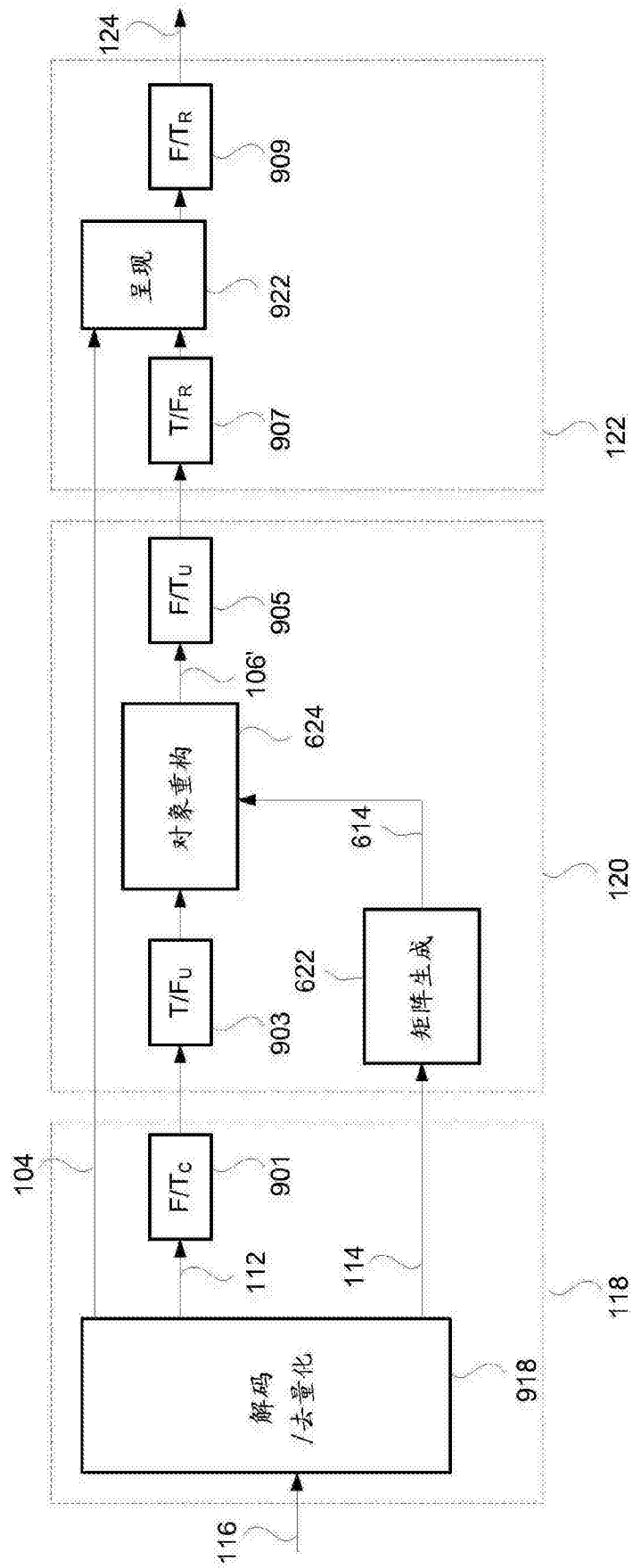


图 9