



## (12) 发明专利申请

(10) 申请公布号 CN 113646836 A

(43) 申请公布日 2021.11.12

(21) 申请号 202080024441.9

(22) 申请日 2020.03.19

(30) 优先权数据

1904261.3 2019.03.27 GB

(85) PCT国际申请进入国家阶段日

2021.09.26

(86) PCT国际申请的申请数据

PCT/FI2020/050174 2020.03.19

(87) PCT国际申请的公布数据

W02020/193852 EN 2020.10.01

(71) 申请人 诺基亚技术有限公司

地址 芬兰埃斯波

(72) 发明人 M-V·莱蒂南 J·维卡莫

L·拉克索南

(74) 专利代理机构 北京市中咨律师事务所  
11247

代理人 杨晓光

(51) Int.Cl.

G10L 19/22 (2006.01)

H03M 7/00 (2006.01)

G10L 19/008 (2006.01)

G10L 19/16 (2006.01)

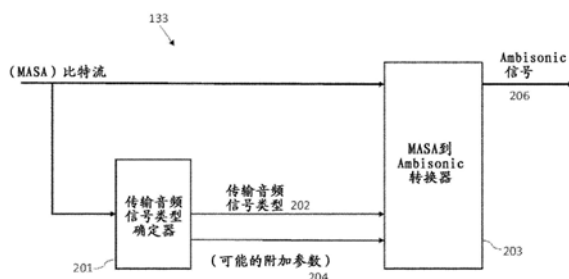
权利要求书3页 说明书24页 附图17页

(54) 发明名称

声场相关渲染

(57) 摘要

一种装置,包括被配置为执行以下操作的部件:获得至少两个音频信号;确定至少两个音频信号的类型;基于至少两个音频信号的所确定的类型,处理至少两个音频信号以被配置为被渲染。



1. 一种装置,包括被配置为执行以下操作的部件:  
获得至少两个音频信号;  
确定所述至少两个音频信号的类型;  
基于所述至少两个音频信号的所确定的类型,处理所述至少两个音频信号以被配置为被渲染。
2. 根据权利要求1所述的装置,其中,所述至少两个音频信号是以下中的一个:  
传输音频信号;以及  
先前处理的音频信号。
3. 根据权利要求1或2所述的装置,其中,所述部件被配置为:获得与所述至少两个音频信号相关联的至少一个参数。
4. 根据权利要求3所述的装置,其中,所述部件被配置为:基于与所述至少两个音频信号相关联的所述至少一个参数,确定所述至少两个音频信号的所述类型。
5. 根据权利要求4所述的装置,其中,被配置为基于所述至少一个参数来确定所述至少两个音频信号的所述类型的所述部件被配置为执行以下中的一个:  
从所述至少一个参数中提取并解码至少一个类型信号;以及  
当所述至少一个参数表示与所述至少两个音频信号相关联的空间音频方面时,分析所述至少一个参数以确定所述至少两个音频信号的所述类型。
6. 根据权利要求5所述的装置,其中,被配置为分析所述至少一个参数以确定所述至少两个音频信号的所述类型的所述部件被配置为:  
基于所述至少两个音频信号,确定宽带左或右通道对总能量比;  
基于所述至少两个音频信号,确定高频左或右通道对总能量比;  
基于所述至少两个音频信号,确定总和对总能量比;  
基于所述至少两个音频信号,确定差值对目标能量比;以及  
基于以下中的至少一个,确定所述至少两个音频信号的所述类型:所述宽带左或右通道对总能量比;基于所述至少两个音频信号的所述高频左或右通道对总能量比;基于所述至少两个音频信号的所述总和对总能量比;以及所述差值对目标能量比。
7. 根据权利要求1至6中任一项所述的装置,其中,所述部件被配置为:确定与所述至少一个音频信号的所述类型相关联的至少一个类型参数。
8. 根据权利要求7所述的装置,其中,被配置为基于所述至少两个音频信号的所确定的类型来处理所述至少两个音频信号以被配置为被渲染的所述部件被配置为:基于与所述至少两个音频信号的所述类型相关联的所述至少一个类型参数,转换所述至少两个音频信号。
9. 根据权利要求1至8中任一项所述的装置,其中,所述至少两个音频信号的所述类型包括以下中的至少一个:  
捕获麦克风布置;  
捕获麦克风分隔距离;  
捕获麦克风参数;  
传输通道标识符;  
间隔音频信号类型;

下混合音频信号类型；  
重合音频信号类型；以及  
传输通道布置。

10. 根据权利要求1至9中任一项所述的装置，其中，被配置为处理所述至少两个音频信号的所述部件被配置为执行以下中的一个：

将所述至少两个音频信号转换成全景环绕声音频信号表示；  
将所述至少两个音频信号转换成多通道音频信号表示；以及  
将所述至少两个音频信号下混合成更少的音频信号。

11. 根据权利要求1至10中任一项所述的装置，其中，被配置为处理所述至少两个音频信号的所述部件被配置为：基于所述至少两个音频信号和所述至少两个音频信号的所述类型，生成至少一个原型信号。

12. 一种方法，包括：

获得至少两个音频信号；

确定所述至少两个音频信号的类型；

基于所述至少两个音频信号的所确定的类型，处理所述至少两个音频信号以被配置为被渲染。

13. 根据权利要求12所述的方法，其中，所述至少两个音频信号是以下中的一个：

传输音频信号；以及

先前处理的音频信号。

14. 根据权利要求12或13所述的方法，还包括：获得与所述至少两个音频信号相关联的至少一个参数。

15. 根据权利要求14所述的方法，其中，确定所述至少两个音频信号的所述类型包括：基于与所述至少两个音频信号相关联的所述至少一个参数，确定所述至少两个音频信号的所述类型。

16. 一种装置，包括至少一个处理器和包括计算机程序代码的至少一个存储器，所述至少一个存储器和所述计算机程序代码被配置为与所述至少一个处理器一起使所述装置至少：

获得至少两个音频信号；

确定所述至少两个音频信号的类型；

基于所述至少两个音频信号的所确定的类型，处理所述至少两个音频信号以被配置为被渲染。

17. 根据权利要求16所述的装置，其中，所述至少两个音频信号是以下中的一个：

传输音频信号；以及

先前处理的音频信号。

18. 根据权利要求16或17所述的装置，其中，所述装置被使得：获得与所述至少两个音频信号相关联的至少一个参数。

19. 根据权利要求18所述的装置，其中，所述装置被使得：基于与所述至少两个音频信号相关联的所述至少一个参数，确定所述至少两个音频信号的所述类型。

20. 根据权利要求19所述的装置，其中，被使得基于所述至少一个参数来确定所述至少

两个音频信号的所述类型的所述装置进一步被使得执行以下中的一个：

从所述至少一个参数中提取并解码至少一个类型信号；以及

当所述至少一个参数表示与所述至少两个音频信号相关联的空间音频方面时，分析所述至少一个参数以确定所述至少两个音频信号的所述类型。

21. 根据权利要求20所述的装置，其中，被使得分析所述至少一个参数以确定所述至少两个音频信号的所述类型的所述装置进一步被使得：

基于所述至少两个音频信号，确定宽带左或右通道对总能量比；

基于所述至少两个音频信号，确定高频左或右通道对总能量比；

基于所述至少两个音频信号，确定总和对总能量比；

基于所述至少两个音频信号，确定差值对目标能量比；以及

基于以下中的至少一个，确定所述至少两个音频信号的所述类型：所述宽带左或右通道对总能量比；基于所述至少两个音频信号的所述高频左或右通道对总能量比；基于所述至少两个音频信号的所述总和对总能量比；以及所述差值对目标能量比。

22. 根据权利要求16至21中任一项所述的装置，其中，所述装置被使得：确定与所述至少一个音频信号的所述类型相关联的至少一个类型参数。

23. 根据权利要求16至22中任一项所述的装置，其中，被使得处理所述至少两个音频信号的所述装置进一步被使得执行以下中的一个：

将所述至少两个音频信号转换成全景环绕声音频信号表示；

将所述至少两个音频信号转换成多通道音频信号表示；以及

将所述至少两个音频信号下混合成更少的音频信号。

24. 根据权利要求16至23中任一项所述的装置，其中，被使得处理所述至少两个音频信号的所述装置进一步被使得：基于所述至少两个音频信号和所述至少两个音频信号的所述类型，生成至少一个原型信号。

25. 根据权利要求16至24中任一项所述的装置，其中，被使得处理所述至少两个音频信号以被渲染的所述装置进一步被使得：基于与所述至少两个音频信号的所述类型相关联的所述至少一个类型参数，转换所述至少两个音频信号。

## 声场相关渲染

### 技术领域

[0001] 本申请涉及用于与声场相关的音频表示和渲染的装置和方法,但非排他地涉及用于音频解码器的音频表示的装置和方法。

### 背景技术

[0002] 沉浸式音频编解码器正被实现,以支持范围从低比特率操作到透明性的大量操作点。这种编解码器的示例是沉浸式语音和音频服务 (IVAS) 编解码器,其被设计为适合于在诸如3GPP 4G/5G网络之类的通信网络上使用,包括在诸如例如用于虚拟现实 (VR) 的沉浸式语音和音频之类的沉浸式服务中使用。该音频编解码器被预期处理语音、音乐和通用音频的编码、解码和渲染。此外还被预期支持基于通道的音频和基于场景的音频输入,包括关于声场和声源的空间信息。编解码器还被预期以低延迟进行操作,以使能会话服务并在各种传输条件下支持高差错鲁棒性。

[0003] 输入信号可以以所支持的多个格式之一 (以及以一些所允许的格式组合) 被呈现给IVAS编码器。例如,单声道音频信号 (没有元数据) 可以使用增强型语音服务 (EVS) 编码器来编码。其他输入格式可以使用IVAS编码工具。至少一些输入可以使用元数据辅助空间音频 (MASA) 工具或任何合适的基于空间元数据的方案。这是适用于空间音频处理的参数空间音频格式。参数空间音频处理是使用一组参数来描述声音 (或声音场景) 的空间方面的音频信号处理领域。例如,在从麦克风阵列进行参数化空间音频捕获时,从麦克风阵列信号估计一组参数是一种典型且有效的选择,该组参数诸如是频带中声音的方向、以及频带中被捕获声音的方向性与非方向性部分的比率。众所周知,这些参数很好地描述了在麦克风阵列的位置处的被捕获声音的感知空间特性。这些参数可以相应地在空间声音的合成中使用,以用于双耳式耳机、扬声器、或诸如全景环绕声 (Ambisonics) 之类的其他格式。

[0004] 例如,可以存在两个通道 (立体声) 的音频信号和空间元数据。此外,空间元数据可以定义参数,诸如:方向索引 (Direction index),描述了声音按时频参数间隔的到达方向;直接对总能量比 (Direct-to-total energy ratio),描述了针对方向索引 (即,时频子帧) 的能量比;扩展相干性 (Spread coherence),描述了针对方向索引 (即,时频子帧) 的能量扩展;扩散对总能量比 (Diffuse-to-total energy ratio),描述了非方向性声音在周围方向上的能量比;环绕相干性 (Surround coherence),描述了非方向性声音在在周围方向上的相干性;剩余对总能量比 (Remainder-to-total energy ratio),描述了剩余部分 (诸如麦克风噪声) 声能的能量比,以满足能量比之和为1的要求;以及距离 (Distance),以对数尺度描述了以米为单位的源自方向索引 (即,时频子帧) 的声音的距离。

[0005] IVAS流可以被解码并被渲染为各种输出格式,包括双声道输出、多通道输出、以及 Ambisonic (FOA/HOA) 输出。另外,可以存在用于外部渲染的接口,其中输出格式可以对应于例如输入格式。

[0006] 由于空间 (例如MASA) 元数据以与输出格式无关的方式描绘期望的空间音频感知,因此,任何具有空间元数据的流可以被灵活地渲染成任何一种上述的输出格式。然而,由于

MASA流可以源自各种输入,因此,解码器接收的传输音频信号可具有不同的特性。因此,解码器必须考虑这些方面,以便能够产生最佳的音频质量。

## 发明内容

[0007] 根据第一方面,提供了一种装置,其包括被配置为执行以下操作的部件:获得至少两个音频信号;确定至少两个音频信号的类型;基于至少两个音频信号的所确定的类型,处理至少两个音频信号以被配置为被渲染。

[0008] 至少两个音频信号可以是以下中的一个:传输音频信号;以及先前处理的音频信号。

[0009] 该部件可以被配置为获得与至少两个音频信号相关联的至少一个参数。

[0010] 被配置为确定至少两个音频信号的类型该部件可以被配置为基于与至少两个音频信号相关联的至少一个参数来确定至少两个音频信号的类型。

[0011] 被配置为基于至少一个参数来确定至少两个音频信号的类型该部件可以被配置为执行以下中的一个:从至少一个参数中提取并解码至少一个类型信号;以及当至少一个参数表示与至少两个音频信号相关联的空间音频方面时,分析至少一个参数以确定至少两个音频信号的类型。

[0012] 被配置为分析至少一个参数以确定至少两个音频信号的类型该部件可以被配置为:基于至少两个音频信号,确定宽带左或右通道对总能量比(broadband left or right channel to total energy ratio);基于至少两个音频信号,确定高频左或右通道对总能量比(higher frequency left or right channel to total energy ratio);基于至少两个音频信号,确定总和对总能量比(sum to total energy ratio);基于至少两个音频信号,确定差值对目标能量比(subtract to target energy ratio);以及基于以下中的至少一个来确定至少两个音频信号的类型:宽带左或右通道对总能量比;基于至少两个音频信号的高频左或右通道对总能量比;基于至少两个音频信号的总和对总能量比;以及差值对目标能量比。

[0013] 该部件可以被配置为确定与至少一个音频信号的类型相关联的至少一个类型参数。

[0014] 被配置为基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的该部件可以被配置为基于与至少两个音频信号的类型相关联的至少一个类型参数来转换至少两个音频信号。

[0015] 至少两个音频信号的类型包括以下中的至少一个:捕获麦克风布置;捕获麦克风分隔距离;捕获麦克风参数;传输通道标识符;间隔音频信号类型;下混合音频信号类型;重合音频信号类型;以及传输通道布置。

[0016] 被配置为基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的该部件可以被配置为:将至少两个音频信号转换成ambisonic音频信号表示;将至少两个音频信号转换成多通道音频信号表示;以及将至少两个音频信号下混合成更少的音频信号。

[0017] 被配置为基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的该部件可以被配置为基于至少两个音频信号和至少两个音频信号的类型

来生成至少一个原型信号。

[0018] 根据第二方面,提供了一种方法,其包括:获得至少两个音频信号;确定至少两个音频信号的类型;基于至少两个音频信号的所确定的类型,处理至少两个音频信号以被配置为被渲染。

[0019] 至少两个音频信号可以是以下中的一个:传输音频信号;以及先前处理的音频信号。

[0020] 该方法还可以包括获得与至少两个音频信号相关联的至少一个参数。

[0021] 确定至少两个音频信号的类型可以包括基于与至少两个音频信号相关联的至少一个参数来确定至少两个音频信号的类型。

[0022] 基于至少一个参数来确定至少两个音频信号的类型可以包括以下中的一个:从至少一个参数中提取并解码至少一个类型信号;以及当至少一个参数表示与至少两个音频信号相关联的空间音频方面时,分析至少一个参数以确定至少两个音频信号的类型。

[0023] 分析至少一个参数以确定至少两个音频信号的类型可以包括:基于至少两个音频信号,确定宽带左或右通道对总能量比;基于至少两个音频信号,确定高频左或右通道对总能量比;基于至少两个音频信号,确定总和对总能量比;基于至少两个音频信号,确定差值对目标能量比;以及基于以下中的至少一个来确定至少两个音频信号的类型:宽带左或右通道对总能量比;基于至少两个音频信号的高频左或右通道对总能量比;基于至少两个音频信号的总和对总能量比;以及差值对目标能量比。

[0024] 该方法还可以包括确定与至少一个音频信号的类型相关联的至少一个类型参数。

[0025] 基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染还可以包括基于与至少两个音频信号的类型相关联的至少一个类型参数来转换至少两个音频信号。

[0026] 至少两个音频信号的类型可以包括以下中的至少一个:捕获麦克风布置;捕获麦克风分隔距离;捕获麦克风参数;传输通道标识符;间隔音频信号类型;下混合音频信号类型;重合音频信号类型;以及传输通道布置。

[0027] 基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染可以包括以下中的一个:将至少两个音频信号转换成ambisonic音频信号表示;将至少两个音频信号转换成多通道音频信号表示;以及将至少两个音频信号下混合成更少的音频信号。

[0028] 基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染可以包括基于至少两个音频信号和至少两个音频信号的类型来生成至少一个原型信号。

[0029] 根据第三方面,提供了一种装置,其包括至少一个处理器和包括计算机程序代码的至少一个存储器,该至少一个存储器和计算机程序代码被配置为与至少一个处理器一起使用该装置至少:获得至少两个音频信号;确定至少两个音频信号的类型;基于至少两个音频信号的所确定的类型,处理至少两个音频信号以被配置为被渲染。

[0030] 至少两个音频信号可以是以下中的一个:传输音频信号;以及先前处理的音频信号。

[0031] 该装置可以被配置为获得与至少两个音频信号相关联的至少一个参数。

[0032] 被使得确定至少两个音频信号的类型该装置可以被使得基于与至少两个音频信号相关联的至少一个参数来确定至少两个音频信号的类型。

[0033] 被使得基于至少一个参数来确定至少两个音频信号的类型该装置可以被使得执行以下中的一个：从至少一个参数中提取并解码至少一个类型信号；以及当至少一个参数表示与至少两个音频信号相关联的空间音频方面时，分析至少一个参数以确定至少两个音频信号的类型。

[0034] 被使得分析至少一个参数以确定至少两个音频信号的类型该装置可以被使得：基于至少两个音频信号，确定宽带左或右通道对总能量比；基于至少两个音频信号，确定高频左或右通道对总能量比；基于至少两个音频信号，确定总和对总能量比；基于至少两个音频信号，确定差值对目标能量比；以及基于以下中的至少一个来确定至少两个音频信号的类型：宽带左或右通道对总能量比；基于至少两个音频信号的高频左或右通道对总能量比；基于至少两个音频信号的总和对总能量比；以及差值对目标能量比。

[0035] 该装置可以被使得确定与至少一个音频信号的类型相关联的至少一个类型参数。

[0036] 被使得基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的该装置可以被使得基于与至少两个音频信号的类型相关联的至少一个类型参数来转换至少两个音频信号。

[0037] 至少两个音频信号的类型包括以下中的至少一个：捕获麦克风布置；捕获麦克风分隔距离；捕获麦克风参数；传输通道标识符；间隔音频信号类型；下混合音频信号类型；重合音频信号类型；以及传输通道布置。

[0038] 被使得基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的该装置可以被使得：将至少两个音频信号转换成ambisonic音频信号表示；将至少两个音频信号转换成多通道音频信号表示；以及将至少两个音频信号下混合成更少的音频信号。

[0039] 被使得基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的该装置可以被使得基于至少两个音频信号和至少两个音频信号的类型来生成至少一个原型信号。

[0040] 根据第四方面，提供了一种装置，其包括：获得电路，被配置为获得至少两个音频信号；确定电路，被配置为确定至少两个音频信号的类型；处理电路，被配置为基于至少两个音频信号的所确定的类型，处理至少两个音频信号以被配置为被渲染。

[0041] 根据第五方面，提供了一种包括指令（或包括程序指令的计算机可读介质）的计算机程序，这些指令用于使装置至少执行以下操作：获得至少两个音频信号；确定至少两个音频信号的类型；基于至少两个音频信号的所确定的类型，处理至少两个音频信号以被配置为被渲染。

[0042] 根据第六方面，提供了一种非暂时性计算机可读介质，其包括用于使装置至少执行以下操作的程序指令：获得至少两个音频信号；确定至少两个音频信号的类型；基于至少两个音频信号的所确定的类型，处理至少两个音频信号以被配置为被渲染。

[0043] 根据第七方面，提供了一种装置，其包括：用于获得至少两个音频信号的部件；用于确定至少两个音频信号的类型部件；用于基于至少两个音频信号的所确定的类型来处理至少两个音频信号以被配置为被渲染的部件。



[0044] 根据第八方面,提供了一种计算机可读介质,其包括用于使装置至少执行以下操作的程序指令:获得至少两个音频信号;确定至少两个音频信号的类型;基于至少两个音频信号的所确定的类型,处理至少两个音频信号以被配置为被渲染。

[0045] 一种装置,包括用于执行如上所述的方法的动作的部件。

[0046] 一种装置,被配置为执行如上所述的方法的动作。

[0047] 一种计算机程序,包括用于使计算机执行如上所述的方法的程序指令。

[0048] 一种被存储在介质上的计算机程序产品可以使装置执行本文所述的方法。

[0049] 一种电子设备可以包括如本文所述的装置。

[0050] 一种芯片组可以包括如本文所述的装置。

[0051] 本申请的实施例旨在解决与现有技术相关联的问题。

## 附图说明

[0052] 为了更好地理解本申请,现在将通过示例的方式参考附图,其中:

[0053] 图1示意性地示出适于实现一些实施例的装置的系统;

[0054] 图2示意性地示出根据一些实施例的示例解码器/渲染器;

[0055] 图3示出根据一些实施例的示例解码器/渲染器的操作的流程图;

[0056] 图4示意性地示出根据一些实施例的如图2中所示的示例传输音频信号类型确定器;

[0057] 图5示意性地示出根据一些实施例的如图2中所示的第二示例传输音频信号类型确定器;

[0058] 图6示出根据一些实施例的第二示例传输音频信号类型确定器的操作的流程图;

[0059] 图7示意性地示出根据一些实施例的如图2中所示的示例元数据辅助空间音频信号到ambisonics格式转换器;

[0060] 图8示出根据一些实施例的示例元数据辅助空间音频信号到Ambisonics格式转换器的操作的流程图;

[0061] 图9示意性地示出根据一些实施例的第二示例解码器/渲染器;

[0062] 图10示出根据一些实施例的另一示例解码器/渲染器的操作的流程图;

[0063] 图11示意性地示出根据一些实施例的如图9中所示的示例元数据辅助空间音频信号到多通道音频信号格式转换器;

[0064] 图12示出根据一些实施例的示例元数据辅助空间音频信号到多通道音频信号格式转换器的操作的流程图;

[0065] 图13示意性地示出根据一些实施例的第三示例解码器/渲染器;

[0066] 图14示出根据一些实施例的第三示例解码器/渲染器的操作的流程图;

[0067] 图15示意性地示出根据一些实施例的如图13中所示的示例元数据辅助空间音频信号下混合器;

[0068] 图16示出根据一些实施例的示例元数据辅助空间音频信号下混合器的操作的流程图;

[0069] 图17示出适于实现如图1、2、4、5、7、9、11、13和15中所示的装置的示例设备。

## 具体实施方式

[0070] 下面进一步详细描述了用于提供空间元数据辅助音频信号的有效渲染的合适装置和可能的机制。

[0071] 关于图1,示出了一种用于实现音频捕获和渲染的示例装置和系统。系统100被示出为具有“分析”部分121和“解复用器/解码器/合成器”部分133。“分析”部分121是从接收多通道扬声器信号到元数据和传输信号的编码的部分,而“解复用器/解码器/合成器”部分133是从编码元数据和传输信号的解码到重新生成信号的呈现(例如,以多声道扬声器的形式)的部分。

[0072] 到系统100和“分析”部分121的输入是多通道信号102。在以下示例中描述了麦克风通道信号输入,然而,在其他实施例中,可以实现任何合适的输入(或合成多通道)格式。例如,在一些实施例中,空间分析器和空间分析可以在编码器外部实现。例如,在一些实施例中,与音频信号相关联的空间元数据可以作为单独的比特流被提供给编码器。在一些实施例中,空间元数据可以被提供为一组空间(方向)索引值。

[0073] 多通道信号被传递给传输信号生成器103和分析处理器105。

[0074] 在一些实施例中,传输信号生成器103被配置为接收多通道信号,生成包括确定数量的通道的合适的传输信号,并输出传输信号104。例如,传输信号生成器103可以被配置为生成多通道信号的2-音频通道下混合。确定数量的通道可以是任何合适数量的通道。在一些实施例中,传输信号生成器被配置为例如通过波束成形技术来以其他方式选择或组合输入音频信号到确定数量的通道并将其输出为传输信号。

[0075] 在一些实施例中,传输信号生成器103是可选的,并且多通道信号以与此示例中的传输信号相同的方式在未经处理的情况下被传递给“编码器/MUX”块107。

[0076] 在一些实施例中,分析处理器105还被配置为接收多通道信号,并分析该信号以产生与多通道信号相关联并因此与传输信号104相关联的元数据106。分析处理器105可以被配置为生成元数据,该元数据对于每个时频分析间隔可以包括方向参数108和能量比率参数110(其示例是扩散度参数)和相干性参数112。在一些实施例中,方向、能量比和相干性参数可以被认为是空间音频参数。换句话说,空间音频参数包括旨在表征由多通道信号(或通常是两个或更多个回放音频信号)创建的声场的参数。

[0077] 在一些实施例中,所生成的参数可在频带与频带之间不同。因此,例如,在频带X中,所有参数被生成和发送,而在频带Y中,仅其中一个参数被生成和发送,此外,在频带Z中,没有参数被生成或发送。这种情况的实际示例可以是对于诸如最高频带的一些频带,出于感知方面原因而不需要一些参数。传输信号104和元数据106可以被传递给“编码器/MUX”块107。

[0078] 在一些实施例中,空间音频参数可以被分组或分离成方向性和非方向性(例如,扩散)参数。

[0079] “编码器/MUX”块107可以被配置为接收传输(例如,下混合)信号104,并生成这些音频信号的合适编码。。在一些实施例中,“编码器/MUX”可以是计算机(运行存储在存储器上和至少一个处理器上的合适的软件),或者可替代地可以是例如使用FPGA或ASIC的特定设备。编码可以使用任何合适的方案来实现。此外,“编码器/MUX”块107可以被配置为接收元数据,并生成编码或压缩形式的信息。在一些实施例中,在图1中由虚线所示的传输或存

储之前,“编码器/MUX”块107可以进一步交织、复用到单个数据流111,或者将元数据嵌入编码的下混合信号内。复用可以使用任何合适的方案来实现。

[0080] 在解码器侧,所接收或获取的数据(流)可以由“解复用器/解码器/合成器”133接收。“解复用器/解码器/合成器”133可以对编码流进行解复用,并对音频信号进行解码以获得传输信号。类似地,“解复用器/解码器/合成器”133可以被配置为接收编码的元数据并对其进行解码。在一些实施例中,“解复用器/解码器/合成器”133可以是计算机(运行存储在存储器和至少一个处理器上的合适的软件),或者可替代地是例如使用FPGA或ASIC的特定设备。

[0081] 系统100的“解复用器/解码器/合成器”部分133可以进一步被配置为基于传输信号和元数据来以任何合适的格式重新创建采用多通道信号110形式的合成空间音频(其可以是多通道扬声器格式,或者在一些实施例中是任何合适的输出格式,诸如用于耳机收听的双耳信号或Ambisonics信号,具体取决于使用实例)。

[0082] 因此,总而言之,首先,该系统(分析部分)被配置为接收多通道音频信号。

[0083] 进而,该系统(分析部分)被配置为生成合适的传输音频信号(例如,通过选择或下混合一些音频信号通道)。

[0084] 该系统进而被配置为对传输信号和元数据进行编码以用于存储/传输。

[0085] 此后,该系统可以存储/发送编码的传输和元数据。

[0086] 该系统可以获取/接收编码的传输和元数据。

[0087] 进而,该系统被配置为从编码的传输和元数据参数中提取传输和元数据,例如,对编码的传输和元数据参数进行解复用和解码。

[0088] 该系统(合成部分)被配置为基于所提取的传输音频信号和元数据来合成输出多通道音频信号。关于解码器(合成部分),它被配置为接收空间元数据和传输音频信号,其例如可以是5.1信号的下混合(的可能的预处理版本)、来自移动设备的两个间隔麦克风信号、或来自重合麦克风阵列的两个波束模式。

[0089] 解码器可以被配置为从空间元数据和传输音频信号来渲染空间音频(诸如Ambisonics)。这通常通过使用如下用于从这种输入来渲染空间音频的两种方法之一来实现:线性渲染和参数化渲染。

[0090] 假定在频带中进行处理,线性渲染是指使用一些静态混合权重来生成所期望的输出。参数化渲染是指基于空间元数据来修改传输音频信号以生成所期望的输出。

[0091] 已经提出了从各种输入来生成Ambisonics的方法:

[0092] 在来自5.1信号的传输音频信号和空间元数据的情况下,参数化处理可用于渲染Ambisonics;

[0093] 在来自间隔麦克风的传输音频信号和空间元数据的情况下,还可以使用线性和参数化处理的组合;

[0094] 在来自重合麦克风的传输音频信号和空间元数据的情况下,可以使用线性和参数化处理的组合。

[0095] 因此,存在多种用于从各种输入来渲染Ambisonics的方法。然而,所有这些Ambisonic渲染方法都假定某种输入。如下文所讨论的一些实施例示出了防止如下问题发生的装置和方法。

[0096] 使用线性渲染,Y信号,其是Ambisonics中的左右定向的一阶(八字形)信号,可以通过 $Y(f) = S_0(f) - S_1(f)$ 从两个重合的相对心形被创建,其中,f是频率。作为另一个示例,Y信号可以通过 $Y(f) = -i(S_0(f) - S_1(f))g_{eq}(f)$ 从间隔麦克风被创建,其中, $g_{eq}(f)$ 是频率相关均衡器(其取决于麦克风距离), $i$ 是虚数单位。对间隔麦克风的处理(包含-90度相移和频率相关均衡)与对重合麦克风的处理不同,并且使用错误的处理技术可能会导致音频质量下降。

[0097] 在某些渲染方案中使用参数化渲染需要使用线性手段来生成“原型”信号。进而,基于空间元数据,在时频域中自适应地修改这些原型信号。优化地,原型信号应当尽可能多地跟随目标信号,以使最小化对参数化处理的需求,从而最小化来自参数化处理的可能的伪影。例如,原型信号应当在足够程度上包含与对应的输出通道相关的所有信号分量。

[0098] 例如,当渲染全向信号W时(其他Ambisonic信号也存在类似的效果),可以例如用两种直接的方法来从立体声传输音频信号创建原型:

[0099] 选择一个通道(例如,左通道);或

[0100] 对这两个通道求和。

[0101] 其选择在很大程度上取决于传输音频信号类型。如果传输信号源自5.1信号,则通常左侧信号是仅左传输音频信号,而右侧信号是仅右传输音频信号(当使用常规下混合矩阵时)。因此,使用一个通道用于原型会丢失另一个通道的信号内容,从而导致生成清晰的伪影(例如,在最坏的情况下,在一个所选择的通道上根本没有信号存在)。因此,在这种情况下,W原型最好被构造为这两个通道之和。另一方面,如果传输信号源自间隔麦克风,则使用传输音频信号之和作为W信号的原型会导致严重的梳状滤波(因为在信号之间存在时间延迟)。这会导致与上述类似的伪影。在这种情况下,最好只选择这两个通道之一作为W原型,至少在高频范围内。

[0102] 因此,没有适合于所有传输音频信号类型的好的选择。

[0103] 因此,利用线性和参数化方法,将针对某个传输音频信号类型而设计的空间音频处理应用于另一个传输音频信号类型,预计将会产生明显的音频质量劣化。

[0104] 关于以下实施例和示例进一步详细讨论的概念涉及音频编码和解码,其中,解码器从编码器接收至少两个传输音频信号。此外,实施例可以是其中传输音频信号可以是至少两个类型,例如5.1信号的下混合、间隔麦克风信号、或重合麦克风信号。此外,在一些实施例中,装置和方法实现了一种提高传输音频信号的处理的质量并提供所确定的输出(例如,Ambisonics、5.1、单声道)的解决方案。通过确定传输音频信号的类型并基于所确定的传输音频信号类型来执行音频处理,可以提高质量。

[0105] 在如本文中进一步详细讨论的一些实施例中,传输音频信号类型通过以下任意一个操作来确定:

[0106] 获得说明传输音频信号类型的元数据,或

[0107] 基于传输音频信号(以及可能的空间元数据,如果可用的话)它们本身,确定传输音频信号类型。

[0108] 说明传输音频信号类型的元数据可以包括例如以下条件:

[0109] 间隔麦克风(可能伴随有麦克风的位置);

[0110] 重合麦克风或与重合麦克风有效类似的波束(可能伴随有麦克风的方向性模式);

[0111] 从多通道音频信号(诸如5.1)的下混合。

[0112] 基于对传输音频信号它们本身的分析来确定传输音频信号类型可以是基于将频带或组合(以不同方式)的频谱效果与预期的频谱效果(部分地基于空间元数据,如果可用的话)相比较。

[0113] 此外,在一些实施例中,音频信号的处理可以包括:

[0114] 渲染Ambisonic信号;

[0115] 渲染多通道音频信号(例如,5.1);以及

[0116] 将传输音频信号下混合成更少数量的音频信号。

[0117] 图2示出了适于实现一些实施例的示例解码器的示意图。示例实施例可以例如在“解复用器/解码器/合成器”块133内实现。在此示例中,输入是包含两个音频通道和空间元数据的元数据辅助空间音频(MASA)流。然而,如本文所讨论的,输入格式可以是任何合适的元数据辅助空间音频格式。

[0118] (MASA)比特流被转发到传输音频信号类型确定器201。传输音频信号类型确定器201被配置为基于比特流来确定传输音频信号类型202和可能的一些附加参数204(诸如麦克风距离)。所确定的参数被转发到MASA到Ambisonic信号转换器203。

[0119] MASA到Ambisonics信号转换器203被配置为接收比特流和传输音频信号类型202(以及可能的一些附加参数204),并被配置为基于所确定的传输音频信号类型202(和可能的附加参数204)来将MASA流转换成Ambisonic信号。

[0120] 在图3中所示的流程图中概述了该示例的操作。

[0121] 如图3中的步骤301所示,第一操作是接收或获得比特流(MASA流)。

[0122] 如图3中的步骤303所示,随后的操作是基于该比特流来确定传输音频信号类型(以及生成类型信号或指示符和可能的其他附加参数)。

[0123] 如图3中的步骤305所示,在已确定传输音频信号类型之后,下一操作是基于所确定的传输音频信号类型来将比特流(MASA流)转换成Ambisonic信号。

[0124] 图4示出了示例传输音频信号类型确定器201的示意图。在此示例中,示例传输音频信号类型确定器适用于其中传输音频信号类型在MASA流中可用的情况。

[0125] 在此示例中,示例传输音频信号类型确定器201包括传输音频信号类型提取器401。传输音频信号类型提取器401被配置为接收比特(MASA)流,并从MASA流中提取(即,读取和/或解码)类型指示符。这种信息例如可以在MASA流的“通道音频格式”字段中获得。此外,如果附加参数可用,那么它们也被提取。此信息从传输音频信号类型提取器401被输出。在一些实施例中,传输音频信号类型可以包括“间隔(spaced)”、“下混合(downmix)”、“重合(coincident)”。在一些其他实施例中,传输音频信号类型可以包括任何合适的值。

[0126] 图5示出了另一示例传输音频信号类型确定器201的示意图。在此示例中,传输音频信号类型不能直接从MASA流中提取或解码而获得。因此,此示例从对MASA流的分析来估计或确定传输音频信号类型。在一些实施例中,此确定是基于使用揭示不同的传输音频信号类型的某些频谱效果的一组估计器/能量比较。

[0127] 在一些实施例中,传输音频信号类型确定器201包括传输音频信号和空间元数据提取器/解码器501。传输音频信号和空间元数据提取器/解码器501被配置为接收MASA流,并从MASA流中提取和/或解码传输音频信号和空间元数据。所得到的传输音频信号502可以

被转发到时间/频率变换器503。此外,所得到的空间元数据522可以被转发到差值与目标能量比较器511。

[0128] 在一些实施例中,传输音频信号类型确定器201包括时间/频率变换器503。时间/频率变换器503被配置为接收传输音频信号502,并将它们转换到时频域。合适的变换例如包括短时傅立叶变换(STFT)和复调制正交镜像滤波器组(QMF)。所得到的信号被标示为 $S_i(b,n)$ ,其中, $i$ 是信道索引, $b$ 是频率点(frequency bin)索引, $n$ 是时间索引。在传输音频信号(从提取器和/或解码器的输出)已经在时频域中的情况下,这可以被省略,或者可替代地,可以包含从一个时频域表示到另一个时频域表示的变换。T/F域传输音频信号504可以被转发到比较器。

[0129] 在一些实施例中,传输音频信号类型确定器201包括宽带L/R对总能量比较器505。宽带L/R对总能量比较器505被配置为接收T/F域传输音频信号504,并输出宽带L/R对总能量比参数。

[0130] 在宽带L/R对总能量比较器505内,计算宽带左、右和总能量:

$$[0131] \quad E_{\text{left,bb}}(n) = \sum_{b=0}^{B-1} |S_0(b,n)|^2$$

$$[0132] \quad E_{\text{right,bb}}(n) = \sum_{b=0}^{B-1} |S_1(b,n)|^2$$

$$[0133] \quad E_{\text{total,bb}}(n) = E_{\text{left,bb}}(n) + E_{\text{right,bb}}(n)$$

[0134] 其中, $B$ 是频率点的数量。这些能量例如通过下式进行平滑化:

$$[0135] \quad E'_{x,bb}(n) = a_1 E_{x,bb}(n) + b_1 E'_{x,bb}(n-1)$$

[0136] 其中, $a_1$ 和 $b_1$ 是平滑系数(例如, $a_1=0.01$ , $b_1=1-a_1$ )。进而,宽带L/R对总能量比较器505被配置为选择并缩放最小的左右能量:

$$[0137] \quad E'_{\text{lr,bb}}(n) = 2 \min(E'_{\text{left,bb}}(n), E'_{\text{right,bb}}(n))$$

[0138] 其中,乘数2用于相对于两个通道之和 $E'_{\text{total,bb}}(n)$ 来归一化能量。

[0139] 进而,宽带L/R对总能量比较器505可以生成宽带L/R对总能量比506为:

$$[0140] \quad \beta(n) = 10 \log_{10} \frac{E'_{\text{lr,bb}}(n)}{E'_{\text{total,bb}}(n)}$$

[0141] 其进而被输出为比率506。

[0142] 在一些实施例中,传输音频信号类型确定器201包括高频L/R对总能量比较器507。高频L/R对总能量比较器507被配置为接收T/F域传输音频信号504,并输出高频L/R对总能量比参数。

[0143] 在宽带L/R对总能量比较器507内,计算高频带左、右和总能量:

$$[0144] \quad E_{\text{left,hi}}(n) = \sum_{b=B_1}^{B-1} |S_0(b,n)|^2$$

$$[0145] \quad E_{\text{right,hi}}(n) = \sum_{b=B_1}^{B-1} |S_1(b, n)|^2$$

$$[0146] \quad E_{\text{total,hi}}(n) = E_{\text{left,hi}}(n) + E_{\text{right,hi}}(n)$$

[0147] 其中,  $B_1$  是其中定义高频区以开始的第一点 (first bin) (其值取决于所应用的 T/F 变换, 例如, 它可以对应于 6kHz)。这些能量例如通过下式进行平滑化:

$$[0148] \quad E'_{x,hi}(n) = a_2 E_{x,hi}(n) + b_2 E'_{x,hi}(n-1)$$

[0149] 其中,  $a_2$  和  $b_2$  是平滑系数。在高频下, 能量差异可以以更高的频率发生, 因此, 平滑系数可以被设置为提供更少的平滑 (例如,  $a_2 = 0.1$ ,  $b_2 = 1 - a_2$ )。

[0150] 进而, 高频 L/R 对总能量比较器 507 可以被配置为从左右能量中选择更小的, 并且结果被乘以 2:

$$[0151] \quad E'_{\text{lr,hi}}(n) = 2 \min(E'_{\text{left,hi}}(n), E'_{\text{right,hi}}(n))$$

[0152] 进而, 高频 L/R 对总能量比较器 507 可以生成高频 L/R 对总能量比 508 为:

$$[0153] \quad \eta(n) = 10 \log_{10} \frac{E'_{\text{lr,hi}}(n)}{E'_{\text{total,hi}}(n)}$$

[0154] 其进而被输出。

[0155] 在一些实施例中, 传输音频信号类型确定器 201 包括总和对总能量比较器 509。总和对总能量比较器 509 被配置为接收 T/F 域传输音频信号 504, 并输出总和对总能量比参数。总和对总能量比较器 509 被配置为检测在一些频率下两个通道异相的情况, 特别地, 这是间隔麦克风记录的典型现象。

[0156] 总和对总能量比较器 509 被配置为针对每个频率点来计算和信号的能量以及总能量:

$$[0157] \quad E_{\text{sum}}(b, n) = |S_0(b, n) + S_1(b, n)|^2$$

$$[0158] \quad E_{\text{total}}(b, n) = |S_0(b, n)|^2 + |S_1(b, n)|^2$$

[0159] 这些能量例如可以通过下式进行平滑化:

$$[0160] \quad E'_x(b, n) = a_3 E_x(b, n) + b_3 E'_x(b, n-1)$$

[0161] 其中,  $a_3$  和  $b_3$  是平滑系数 (例如,  $a_3 = 0.01$ ,  $b_3 = 1 - a_3$ )。

[0162] 进而, 总和对总能量比较器 509 被配置为计算最小总和对总能量比 510 为:

$$[0163] \quad \chi(n) = 10 \log_{10} \min_b \frac{E'_{\text{sum}}(b, n)}{E'_{\text{total}}(b, n)}, 0 \leq b \leq B_2$$

[0164] 其中,  $B_2$  是其中执行此计算的频率区的最高点 (highest bin) (其值取决于所使用的 T/F 变换, 例如, 它可以对应于 10kHz)。

[0165] 进而, 总和对总能量比较器 509 被配置为输出比率  $\chi(n)$  510。

[0166] 在一些实施例中, 传输音频信号类型确定器 201 包括差值对目标能量比较器 511。差值对目标能量比较器 511 被配置为接收 T/F 域传输音频信号 504 和空间元数据 522, 并输出差值对目标能量比参数 512。

[0167] 差值对目标能量比较器 511 被配置为计算左和右通道的差异能量:

$$[0168] \quad E_{\text{sub}}(b, n) = |S_0(b, n) - S_1(b, n)|^2$$

[0169] 这可以被认为对于至少某些输入信号类型, Ambisonics的Y信号的“原型”(Y信号具有偶极子的方向性模式, 其中左侧为正瓣, 右侧为负瓣)。

[0170] 进而, 差值对目标能量比较器511可以被配置为针对Y信号计算目标能量 $E_{\text{target}}(b, n)$ 。这是基于估计应当如何基于空间元数据而在球面谐波之间分布总能量。例如, 在一些实施例中, 差值对目标能量比较器511被配置为基于空间元数据和能量估计来构建目标协方差矩阵(信道能量和互相关)。然而, 在一些实施例中, 估计仅Y信号的能量, 其是目标协方差矩阵的一个条目。因此, 作为针对Y的目标能量 $E_{\text{target}}(b, n)$ 由两部分组成:

$$[0171] \quad E_{\text{target}}(b, n) = E_{\text{target,amb}}(b, n) + E_{\text{target,dir}}(b, n)$$

[0172] 其中,  $E_{\text{target,amb}}(b, n)$  是目标能量的环境/非方向性部分, 其由下式被定义:

$$[0173] \quad E_{\text{target,amb}}(b, n) = \frac{E_{\text{total}}(b, n)}{3} (1 - r(b, n))(1 - c_{\text{sur}}(b, n))$$

[0174] 其中,  $r(b, n)$  是空间元数据的在0与1之间的直接对总能量比参数,  $c_{\text{sur}}(b, n)$  是空间元数据的在0与1之间的环绕相干性参数(环绕相干声音不会被Y偶极子捕获, 因为在这种情况下, 正瓣和负瓣相互抵消)。除以3是因为假定用于Ambisonic输出的SN3D归一化方案, 并且在这种情况下, Y分量的环境能量是总全向能量的三分之一。

[0175] 应当注意, 空间元数据的频率和/或时间分辨率对于每个 $b, n$ 可以更低, 从而对于若干频率或时间索引, 参数可以相同。

[0176]  $E_{\text{target,dir}}(b, n)$  是更方向性部分的能量。在其构造中, 需要定义作为空间元数据中在0与1之间的扩展相干性 $c_{\text{spread}}(b, n)$  参数的函数的扩展相干分布向量:

$$[0177] \quad \mathbf{v}_{\text{DISTR},3}(b, n) = \begin{cases} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{(2 - 2c_{\text{spread}}(b, n))^2 + 2}} & , \text{当 } c_{\text{spread}}(b, n) > 0.5 \\ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} (1 - c_{\text{spread}}(b, n)) + \begin{bmatrix} \sqrt{1/3} \\ \sqrt{1/3} \\ \sqrt{1/3} \end{bmatrix} c_{\text{spread}}(b, n) & , \text{否则} \end{cases}$$

[0178] 差值对目标能量比较器511还可以被配置为确定方位角值的向量:

$$[0179] \quad \theta(b, n) = \theta(b, n) + \begin{bmatrix} 0 \\ \pi/6 \\ -\pi/6 \end{bmatrix}$$

[0180] 其中,  $\theta(b, n)$  是以弧度为单位的空间元数据的方位角值。假定基于向量条目的 $\sin()$ 操作, 则直接部分目标能量为:

$$[0181] \quad E_{\text{target,dir}}(b, n) = \sin(\theta(b, n))^T \mathbf{v}_{\text{DISTR},3}(b, n) E_{\text{total}}(b, n) r(b, n)$$

[0182] 因此, 获得 $E_{\text{target}}(b, n)$ 。在一些实施例中, 这些能量例如可以通过下式进行平滑化:

$$[0183] \quad E'_x(b, n) = a_4 E_x(b, n) + b_4 E'_x(b, n-1)$$

[0184] 其中,  $a_4$  和  $b_4$  是平滑系数(例如,  $a_4 = 0.0004$ ,  $b_4 = 1 - a_4$ )。

[0185] 此外, 差值对目标能量比较器511被配置为使用在最低频率点的能量来计算差值对目标比512为:



$$[0186] \quad v(n) = 10 \log_{10} \frac{E'_{\text{sub}}(0, n)}{E'_{\text{target}}(0, n)}$$

[0187] 其进而被输出。

[0188] 在一些实施例中,传输音频信号类型确定器201包括传输音频信号类型(基于所估计的度量)确定器513。传输音频信号类型确定器513被配置为接收宽带L/R对总能量比506、高频L/R对总能量比508、最小总和总能量比510和差值对目标比512,并基于这些所接收的所估计度量来确定传输音频信号类型。

[0189] 决定可以通过各种方式来完成,并且实际实现可以在许多方面有所不同,诸如所使用的T/F变换。非限制性形式的示例可以是传输音频信号类型(基于所估计的度量)确定器513首先计算间隔度量的变化:

$$[0190] \quad \Xi_s(n) = \frac{(-v(n) - 3)}{3} + \frac{\max(-\chi(n), 0)}{6} + \frac{\beta(n)}{6}$$

[0191] 如果 $v(n) < -3$ , 否则 $\Xi_s(n) = 0$

[0192] 进而,传输音频信号类型(基于所估计的度量)确定器513可以被配置为计算下混合度量的变化:

$$[0193] \quad \Xi_{d1}(n) = \frac{v(n)}{3} + \frac{\chi(n) + 1}{6} + \frac{-\beta(n)}{6}$$

[0194] 如果 $v(n) > 0$ , 否则 $\Xi_{d1}(n) = 0$

$$[0195] \quad \Xi_{d2}(n) = \frac{(v(n) + 4)}{3} + \frac{\chi(n)}{6} + \frac{(-\eta(n) - 12)}{3}$$

[0196] 如果 $\eta(n) < -12$ , 否则 $\Xi_{d2}(n) = 0$

[0197] 进而,传输音频信号类型(基于所估计的度量)确定器513可以基于这些度量来决定传输音频信号是源自间隔麦克风还是它们是来自环绕声信号(诸如5.1)的下混合。例如,其中,

[0198] 如果 $\Xi_s(n) > 1$ ,  $T(n) = \text{"spaced"}$

[0199] 否则如果 $\Xi_{d1}(n) > 1 \vee \Xi_{d2}(n) > 1$ ,  $T(n) = \text{"downmix"}$

[0200] 否则 $T(n) = T(n-1)$

[0201] 在此示例中,传输音频信号类型(基于所估计的度量)确定器513不检测重合麦克风类型。然而,在实践中,根据 $T(n) = \text{"downmix"}$ 类型的处理通常也可以在重合捕获的情况下(例如,朝向左和右定向的心形)产生良好的音频。

[0202] 进而,传输音频信号类型(基于所估计的度量)确定器513可以被配置为输出传输音频信号类型 $T(n)$ 作为传输音频信号类型202。在一些实施例中,可以输出其他参数204。

[0203] 图6概述了图5中所示的装置的操作。因此,在一些实施例中,如图6中步骤601所示,第一操作是从MASA流(或比特流)中提取和/或解码传输音频信号和元数据。

[0204] 如图6中步骤603所示,下一操作可以是对传输音频信号进行时频域变换。

[0205] 然后可以进行一系列比较。例如,如图6中步骤605所示,通过将宽带L/R能量值与总能量值进行比较,可以生成宽带L/R对总能量比。

[0206] 例如,如图6中步骤607所示,通过将高频L/R能量值与总能量值进行比较,可以生

成高频L/R对总能量比。

[0207] 如图6中步骤609所示,通过将和能量值与总能量值进行比较,可以生成总和与总能量比。

[0208] 此外,如图6中步骤611所示,可以生成差值对目标能量比。

[0209] 在已确定这些度量之后,如图6中步骤613所示,该方法进而可以通过分析这些度量比率来确定传输音频信号类型。

[0210] 图7更详细地示出了示例MASA到Ambisonic转换器203。MASA到Ambisonic转换器203被配置为接收MASA流(比特流)和传输音频信号类型202以及可能的附加参数204,并被配置为基于所确定的传输音频信号类型来将MASA流转换成Ambisonic信号。

[0211] MASA到Ambisonic转换器203包括传输音频信号和空间元数据提取器/解码器501。其被配置为以与在图5中示出并在其中讨论的传输音频信号类型确定器中发现的相同的方式来接收MASA流并输出传输音频信号502和空间元数据522。在一些实施例中,提取器/解码器501是来自传输音频信号类型确定器的提取器/解码器。所得到的传输音频信号502可以被转发到时间/频率变换器503。此外,所得到的空间元数据522可以被转发到信号混合器705。

[0212] 在一些实施例中,MASA到Ambisonic转换器203包括时间/频率变换器503。时间/频率变换器503被配置为接收传输音频信号502,并将它们转换到时频域。合适的变换例如包括短时傅立叶变换(STFT)和复调制正交镜像滤波器组(QMF)。所得到的信号被标示为 $S_i(b, n)$ ,其中, $i$ 是信道索引, $b$ 是频率点索引, $n$ 是时间索引。如果音频提取和/或解码的输出已经在时频域中,则可以省略该块,或者可替代地,它可以包含从一种时频域表示到另一种时频域表示的变换。T/F域传输音频信号504可以被转发到原型信号创建器701。在一些实施例中,时间/频率变换器503是来自传输音频信号类型确定器的相同的时间/频率变换器。

[0213] 在一些实施例中,MASA到Ambisonic转换器203包括原型信号创建器701。原型信号创建器701被配置为接收T/F域传输音频信号504、传输音频信号类型202和可能的附加参数204。进而,T/F原型信号702可以被输出到信号混合器705和去相关器703。

[0214] 在一些实施例中,MASA到Ambisonic转换器203包括去相关器703。去相关器703被配置为接收T/F原型信号702,应用去相关,并将去相关的T/F原型信号704输出到信号混合器705。在一些实施例中,去相关器703是可选的。

[0215] 在一些实施例中,MASA到Ambisonic转换器203包括信号混合器705。信号混合器705被配置为接收T/F原型信号702、去相关的T/F原型信号和空间元数据522。

[0216] 原型信号创建器701被配置为基于传输音频信号类型来针对球面谐波Ambisonics(F0A/H0A)中的每一个生成原型信号。

[0217] 在一些实施例中,原型信号创建器701被配置为进行操作以使得:

[0218] 如果 $T(n) = \text{"spaced"}$ ,则W信号的原型可以被创建如下:

$$[0219] \quad W_{\text{proto}}(b, n) = \frac{S_0(b, n) + S_1(b, n)}{2}, \quad 0 \leq b \leq B_3$$

$$[0220] \quad W_{\text{proto}}(b, n) = S_0(b, n), \quad b > B_3$$

[0221] 在实践中, $W_{\text{proto}}(b, n)$ 可以被创建为在低频下的传输音频信号的平均(其中信号大致同相并且没有梳状滤波发生),并且通过选择在高频下的通道之一。 $B_3$ 的值取决于T/F变

换和麦克风之间的距离。如果距离未知,则可以使用一些默认值(例如,对应于1kHz的值)。

[0222] 如果 $T(n) = \text{"downmix"}$ 或 $T(n) = \text{"coincident"}$ ,W信号的原型可以被创建如下:

[0223]  $W_{\text{proto}}(b, n) = S_0(b, n) + S_1(b, n)$

[0224]  $W_{\text{proto}}(b, n)$  是通过传输音频信号求和而创建的,因为可以假定通常在具有这些信号类型的原始音频信号之间没有显著的延迟。

[0225] 关于Y原型信号

[0226] 如果 $T(n) = \text{"spaced"}$ ,则Y信号的原型可以被创建如下:

[0227] 
$$Y_{\text{proto}}(b, n) = \frac{S_0(b, n) + S_1(b, n)}{2}, \quad 0 \leq b \leq B_4$$

[0228]  $Y_{\text{proto}}(b, n) = -i(S_0(b, n) - S_1(b, n))g_{\text{eq}}(b), B_4 < b \leq B_5$

[0229]  $Y_{\text{proto}}(b, n) = S_0(b, n), b > B_5$

[0230] 在中频下( $B_4$ 与 $B_5$ 之间),可以通过减去传输信号、相移-90度、以及进行均衡来创建偶极子信号。因此,它可以作为Y信号的良好原型,特别是在麦克风距离已知的情况下,,并因此均衡系数是合适的。在低频和高频下这是不可行的,并且原型信号以与全向W信号相同的方式来生成。

[0231] 如果准确已知麦克风距离,则Y原型可以直接被用于在那些频率下的Y(即, $Y(b, n) = Y_{\text{proto}}(b, n)$ )。如果麦克风间距未知,则可以使用 $g_{\text{eq}}(b) = 1$ 。

[0232] 在一些实施例中,信号混合器705可以在频带中应用增益处理,以利用潜在的增益平滑来将频带中 $W_{\text{proto}}(b, n)$ 的能量校正为频带中的目标能量。频带中全向信号的目标能量可以是该频带中传输音频信号能量之和。此处理的结果是全向信号 $W(b, n)$ 。

[0233] 对于 $Y_{\text{proto}}(b, n)$ 不能直接被用于Y(b, n)的Y信号并且当频率在 $B_4$ 与 $B_5$ 之间时,执行自适应增益处理。这种情况类似于上述全向W的情况:除了可能的错误频谱以外,原型信号已经是Y偶极子,并且信号混合器在频段中对原型信号进行增益处理。(另外,关于Y信号,在此特定上下文中不需要去相关)。增益处理可以是指使用频带中的空间元数据(方向、比率、其他参数)和整体信号能量估计(例如,传输信号能量之和)来确定Y分量的能量在频带中应当是什么,进而利用增益来将频带中原型信号的能量校正为所确定的能量,其结果进而是输出Y(b, n)。

[0234] 在 $T(n) = \text{"spaced"}$ 的当前上下文中,前述生成Y(b, n)的过程并非对于所有频率都是有效的。信号混合器和去相关器根据具有这种传输信号类型的频率而不同地配置,因为原型信号在不同的频率是不同的。为了说明不同种类的原型信号,可以考虑其中声音从Y偶极子(其具有正瓣和负瓣)的负增益方向到达的场景。在中频(在 $B_4$ 与 $B_5$ 之间)下,Y原型信号的相位与W原型信号的相位相反,因为它应当针对到达声音的那个方向。在其他频率下(低于 $B_4$ 和高于 $B_5$ ),原型Y信号的相位与W原型信号的相位相同。合适的相位(以及能量和相关性)的合成进而将由信号混合器和去相关器在那些频率下进行考虑。

[0235] 在其中波长很大的低频下(低于 $B_4$ ),用间隔麦克风(其通常彼此有些接近)捕获的音频信号之间的相位差很小。因此,由于SNR原因,原型信号创建器不应被配置为以与在 $B_4$ 与 $B_5$ 之间的频率相同的方式来生成原型信号。因此,通常替代地使用通道和全向信号作为原型信号。在其中波长很小的高频下(高于 $B_5$ ),空间混叠严重扭曲了波束模式(如果使用了如在 $B_4$ 与 $B_5$ 之间的频率的方法),因此,最好使用通道选择全向原型信号。

[0236] 接下来描述在这些频率下(低于 $B_4$ 或高于 $B_5$ )信号混合器和去相关器的配置。对于简单的示例,空间元数据参数集由方位角 $\theta$ 和频带中的比率 $r$ 组成。增益 $\sin(\theta) \sqrt{r}$ 在信号混合器内被应用于原型信号以生成Y偶极子信号,并且结果是相干部分信号。原型信号还被去相关(在去相关器中),并且去相关结果在信号混合器中被接收,在信号混合器中它与因子 $\sqrt{1-r} g_{\text{order}}$ 相乘,并且结果是非相干部分信号。根据已知的SN3D归一化方案,增益 $g_{\text{order}}$ 是在该球面谐波函数阶次下的扩散场增益。例如,对于一阶(如在Y偶极子的情况下),它是 $\sqrt{1/3}$ ,对于二阶,它是 $\sqrt{1/5}$ ,对于三阶,它是 $\sqrt{1/7}$ ,依此类推。相干部分信号和非相干部分信号被相加在一起。除了由于可能的错误的原型信号能量而导致的可能的错误能量以外,其结果是合成的Y信号。可以应用与在中频(在 $B_4$ 与 $B_5$ 之间)的上下文中所描述的相同的频带中的能量校正过程,以将频带中的能量校正到所期望的目标,并且输出是信号Y(b,n)。

[0237] 对于其他球面谐波,诸如X和Z分量、或二阶或更高阶分量,可以应用上述过程,不同之处在于相对于方位角(及其他可能的参数)的增益取决于正被合成的球面谐波信号。例如,从W原型针对X偶极子相干部分生成的增益是 $\cos(\theta) \sqrt{r}$ 。对于除了在 $B_4$ 与 $B_5$ 之间的频率以外的频率,去相关、比率处理和能量校正可以与上述针对Y分量而确定的相同。

[0238] 在上述过程中可以考虑其他参数,诸如仰角、扩展相干性和环绕相干性。扩展相干性参数的值可以是0到1。扩展相干性值0表示点源,换句话说,当使用多扬声器系统来再现音频信号时,应当用尽可能少的扬声器来再现声音(例如,当方向为中央时只有中央扬声器)。随着扩展相干性的值的增加,更多能量被扩展到中央扬声器周围的其他扬声器,直到值0.5为止,能量在中央扬声器与相邻扬声器之间均匀扩展。随着扩展相干值增加到0.5以上,中央扬声器中的能量减少,直到值1为止,在中央扬声器中没有能量,所有能量都在相邻扬声器中。环绕相干性参数的值是从0到1。值1意味着所有(或几乎所有)扬声器通道之间存在相干性。值0意味着所有(或甚至几乎所有)扬声器通道之间没有相干性。这在GB申请No 1718341.9和PCT申请PCT/FI2018/050788中有进一步解释。

[0239] 例如,可以通过降低球面谐波分量中的合成环境能量来实现增加的环绕相干性,并且可以通过在生成相干部分时根据Ambisonic模式的定义添加仰角相关增益来增加仰角。

[0240] 如果 $T(n) = \text{"downmix"}$ 或 $T(n) = \text{"coincident"}$ ,Y信号的原型可以被创建如下:

$$[0241] \quad Y_{\text{proto}}(b,n) = S_0(b,n) - S_1(b,n)$$

[0242] 在这种情况下,不需要相移,因为可以假定通常在具有这些信号类型的原始音频信号之间没有显著的延迟。关于“混合信号”块,如果 $T(n) = \text{"coincident"}$ ,则Y和W原型可以直接被用于Y和W输出,可能在进行增益之后(根据实际方向性模式)。如果 $T(n) = \text{"downmix"}$ ,则 $Y_{\text{proto}}(b,n)$ 和 $W_{\text{proto}}(b,n)$ 不能直接被用于Y(b,n)和W(b,n),但是可能需要在频带中将能量校正到“所期望的目标,如针对情况 $T(n) = \text{"spaced"}$ 而确定的(注意,无论到达声音角度如何,全向分量都具有空间增益1)。

[0243] 对于其他球面谐波(诸如X和Z),创建能够很好地复制目标信号的原型是不可能的,因为典型的下混合信号被定向在左-右轴上而不是前-后X轴或顶部-底部Z轴。因此,在一些实施例中,其方法是利用全向信号的原型,例如,

$$[0244] \quad X_{\text{proto}}(b,n) = W_{\text{proto}}(b,n)$$

[0245]  $Z_{\text{proto}}(b, n) = W_{\text{proto}}(b, n)$

[0246] 类似地,  $W_{\text{proto}}(b, n)$  也由于相同的原因而被用于高次谐波。在这种情况下, 信号混合器和去相关器可以以与针对这些球面谐波分量的  $T(n) = \text{"spaced"}$  相同的方式来处理信号。

[0247] 在某些情况下, 传输音频信号类型  $T(n)$  可以在音频回放期间发生变化 (例如, 由于信号类型的实际变化, 或自动类型检测中的缺陷)。为了避免由于突然改变类型而导致的伪影, 在一些实施例中原型信号可以被内插。这例如可以通过简单地根据旧类型的原型信号线性内插到根据新类型的原型信号而实现。

[0248] 信号混合器的输出是所得到的时频域 Ambisonic 信号, 其被转发到逆 T/F 变换器 707。

[0249] 在一些实施例中, MASA 到 Ambisonic 信号转换器 203 包括被配置为将信号转换到时域的逆 T/F 变换器 707。时域 Ambisonic 信号 906 是从 MASA 到 Ambisonic 转换器的输出。

[0250] 关于图 8, 示出了图 7 中所示的装置的操作的概述。

[0251] 因此, 在一些实施例中, 如图 8 中步骤 801 所示, 第一操作是从 MASA 流 (或比特流) 中提取和/或解码传输音频信号和元数据。

[0252] 如图 8 中步骤 803 所示, 下一操作可以是对传输音频信号进行时频域变换。

[0253] 进而, 如图 8 中步骤 805 所示, 该方法包括基于时频域传输信号并进一步基于传输音频信号类型 (并进一步基于附加参数) 来创建原型音频信号。

[0254] 如图 8 中步骤 807 所示, 在一些实施例中, 该方法包括对时频原型音频信号应用去相关。

[0255] 进而, 如图 8 中步骤 809 所示, 可以基于空间元数据和传输音频信号类型来混合去相关的时频原型音频信号和时频原型音频信号。

[0256] 如图 8 中步骤 811 所示, 进而可以对被混合的信号进行逆时频变换。

[0257] 进而, 如图 8 中步骤 813 所示, 可以输出时域信号。

[0258] 图 9 示出了适于实现一些实施例的示例解码器的示意图。示例实施例可以例如在图 1 中所示的示例“解复用器/解码器/合成器”块 133 内实现。在此示例中, 输入是包含两个音频通道和空间元数据的元数据辅助空间音频 (MASA) 流。然而, 如本文中所讨论的, 输入格式可以是任何合适的元数据辅助空间音频格式。

[0259] (MASA) 比特流被转发到传输音频信号类型确定器 201。传输音频信号类型确定器 201 被配置为基于该比特流来确定传输音频信号类型 202 和可能的一些附加参数 204 (例如, 麦克风距离)。所确定的参数被转发到 MASA 到多通道音频信号转换器 903。在一些实施例中, 传输音频信号类型确定器 201 是与在上面关于图 2 所描述的相同的传输音频信号类型确定器 201, 或者可以是传输音频信号类型确定器 201 的独立的实例, 其被配置为以与在上面关于图 2 中所示的示例描述的传输音频信号类型确定器 201 类似的方式操作。

[0260] MASA 到多通道音频信号转换器 903 被配置为接收比特流和传输音频信号类型 202 (以及可能的一些附加参数 204), 并被配置为基于所确定的传输音频信号类型 202 (和可能的附加参数 204) 来将 MASA 流转换成多通道音频信号 (诸如 5.1)。

[0261] 在图 10 中所示的流程图中概述了图 9 中所示的示例操作。

[0262] 如图 10 中步骤 301 所示, 第一操作是接收或获得比特流 (MASA 流)。

[0263] 如图10中步骤303所示,随后的操作是基于该比特流来确定传输音频信号类型(以及生成类型信号或指示符和可能的其他附加参数)。

[0264] 如图10中步骤305所示,在已确定传输音频信号类型之后,下一操作是基于所确定的传输音频信号类型来将比特流(MASA流)转换成多通道音频信号(诸如5.1)。

[0265] 图11更详细地示出了示例MASA到多通道音频信号转换器903。MASA到多通道音频信号转换器903被配置为接收MASA流(比特流)和传输音频信号类型202以及可能的附加参数204,并被配置为基于所确定的传输音频信号来将MASA流转换成多通道音频信号类型。

[0266] MASA到多通道音频信号转换器903包括传输音频信号和空间元数据提取器/解码器501。其被配置为以与在图5中示出并在其中讨论的传输音频信号类型确定器中发现的相同的方式来接收MASA流并输出传输音频信号502和空间元数据522。在一些实施例中,提取器/解码器501是来自先前描述的传输音频信号类型确定器的提取器/解码器,或者是提取器/解码器的独立的实例。所得到的传输音频信号502可以被转发到时间/频率变换器503。此外,所得到的空间元数据522可以被转发到目标信号特性确定器1101。

[0267] 在一些实施例中,MASA到多通道音频信号转换器903包括时间/频率变换器503。时间/频率变换器503被配置为接收传输音频信号502,并将它们转换到时频域。合适的变换例如包括短时傅立叶变换(STFT)和复调制正交镜像滤波器组(QMF)。所得到的信号被标示为 $S_i(b, n)$ ,其中, $i$ 是信道索引, $b$ 是频率点索引, $n$ 是时间索引。如果音频提取和/或解码的输出已经在时频域中,则可以省略该块,或者可替代地,它可以包含从一个时频域表示到另一个时频域表示的变换。T/F域传输音频信号504可以被转发到原型信号创建器1111。在一些实施例中,时间/频率变换器503是来自传输音频信号类型确定器、或MASA到Ambisonics转换器、或独立的实例的相同的时间/频率转换器。

[0268] 在一些实施例中,MASA到多通道音频信号转换器903包括原型信号创建器1111。原型信号创建器1111被配置为接收T/F域传输音频信号504、传输音频信号类型202和可能的附加参数204。T/F原型信号1112进而可以被输出到信号混合器1105和去相关器1103。

[0269] 作为关于原型信号创建器1111的操作的示例,描述了对5.1多通道音频信号配置的渲染。

[0270] 在此示例中,左侧(左前和左环绕)输出通道的原型信号可以被创建为:

$$[0271] \quad L_{f, \text{proto}}(b, n) = L_{s, \text{proto}}(b, n) = S_0(b, n)$$

[0272] 右侧输出(右前和右环绕)通道的原型信号可以被创建为:

$$[0273] \quad R_{f, \text{proto}}(b, n) = R_{s, \text{proto}}(b, n) = S_1(b, n)$$

[0274] 因此,对于中间面的任一侧的输出通道,原型信号可以直接使用对应的传输音频信号。

[0275] 对于中央输出通道,原型音频信号应当包含来自左侧和右侧的能量,因为它可被用于向任一侧平移。因此,在Ambisonic渲染的情况下,原型信号可以如全向通道一样地被创建,换句话说,如果 $T(n) = \text{"spaced"}$ ,

$$[0276] \quad C_{\text{proto}}(b, n) = \frac{S_0(b, n) + S_1(b, n)}{2}, \quad 0 \leq b \leq B_3$$

$$[0277] \quad C_{\text{proto}}(b, n) = S_0(b, n), \quad b > B_3$$

[0278] 在一些实施例中,如果 $T(n) = \text{"downmix"}$ 或 $T(n) = \text{"coincident"}$ ,原型音频信号可

以生成原型中央音频通道：

$$[0279] \quad C_{\text{proto}}(b, n) = S_0(b, n) + S_1(b, n)$$

[0280] 在一些实施例中，MASA到多通道音频信号转换器903包括去相关器1103。去相关器1103被配置为接收T/F原型信号1112，应用去相关，并将去相关的T/F原型信号1104输出到信号混合器1105。在一些实施例中，去相关器1103是可选的。

[0281] 在一些实施例中，MASA到多通道音频信号转换器903包括目标信号特性确定器1101。在一些实施例中，目标信号特性确定器1101被配置为基于空间元数据和频带中信号能量的总体估计来生成目标协方差矩阵（目标信号特性）。在一些实施例中，此能量估计可以是频带中的传输信号能量之和。此目标协方差矩阵（目标信号特性）确定可以以与由专利申请GB 1718341.9提供的类似的方式执行。

[0282] 目标信号特性1102进而可以被传递到信号混合器1105。

[0283] 在一些实施例中，MASA到多通道音频信号转换器903包括信号混合器1105。信号混合器1105被配置为测量原型信号的协方差矩阵，并基于所估计的（原型信号）协方差矩阵和目标协方差矩阵来制定混合方案。在一些实施例中，混合方案可以与GB 1718341.9中描述的方案类似。该混合方案被应用于原型信号和去相关的原型信号，进而所得到的信号是基于目标信号特性而以频带特性来获得的。换句话说，基于所确定的目标协方差矩阵。

[0284] 在一些实施例中，MASA到多通道音频信号转换器903包括被配置为将信号转换到时域的逆T/F变换器707。时域多通道音频信号是MASA到多通道音频信号转换器的输出。

[0285] 关于图12，示出了图11中所示的装置的操作的概述。

[0286] 因此，如图12中步骤801所示，在一些实施例中，第一操作是从MASA流（或比特流）中提取和/或解码传输音频信号和元数据。

[0287] 如图12中步骤803所示，下一操作可以是对传输音频信号进行时频域变换。

[0288] 进而，如图12中步骤1205所示，该方法包括基于时频域传输信号并进一步基于传输音频信号类型（并进一步基于附加参数）来创建原型音频信号。

[0289] 如图12中步骤1207所示，在一些实施例中，该方法包括对时频原型音频信号应用去相关。

[0290] 进而，如图12中步骤1208所示，可以基于时频域传输音频信号和空间元数据来确定目标信号特性（以生成目标信号的协方差矩阵）。

[0291] 如图12中步骤1209所示，可以测量原型音频信号的协方差矩阵。

[0292] 进而，如图12中步骤1210所示，可以基于目标信号特性来混合去相关的时频原型音频信号和时频原型音频信号。

[0293] 如图12中步骤1211所示，进而可以对被混合的信号进行逆时频变换。

[0294] 进而，如图12中步骤1213所示，可以输出时域信号。

[0295] 图13示出了适于实现一些实施例的另一示例解码器的示意图。在其他实施例中，类似的方法可以在除了解码器以外的装置中实现，例如被实现为编码器的一部分。示例实施例可以例如在诸如图1中所示的（IVAS）“解复用器/解码器/合成器”块133内实现。在此示例中，输入是包含两个音频通道和空间元数据的元数据辅助空间音频（MASA）流。然而，如本文中所讨论的，输入格式可以是任何合适的元数据辅助空间音频格式。

[0296] （MASA）比特流被转发到传输音频信号类型确定器201。传输音频信号类型确定器

201被配置为基于比特流来确定传输音频信号类型202以及可能的一些附加参数204(这种附加参数的示例是麦克风距离)。所确定的参数被转发到下混合器1303。在一些实施例中,传输音频信号类型确定器201与如上所述的传输音频信号类型确定器201相同,或者可以是传输音频信号类型确定器201的独立的实例,其被配置为以与如上所述的传输音频信号类型确定器201类似的方式操作。

[0297] 下混合器1303被配置为接收比特流和传输音频信号类型202(以及可能的一些附加参数204),并被配置为基于所确定的传输音频信号类型202(和可能的附加参数204)来将MASA流从2个传输音频信号下混合成1个传输音频信号。输出MASA流1306进而被输出。

[0298] 在图14中所示的流程图中概述了图13中所示的示例的操作。

[0299] 如图14中步骤301所示,第一操作是接收或获得比特流(MASA流)。

[0300] 如图14中步骤303所示,随后的操作是基于该比特流来确定传输音频信号类型(以及生成类型信号或指示符和可能的其他附加参数)。

[0301] 在已确定传输音频信号类型之后,如图14中步骤1405所示,下一操作是基于所确定的传输音频信号类型202(和可能的附加参数204)来将MASA流从2个传输音频信号下混合成1个传输音频信号。

[0302] 图15更详细地示出了示例下混合器1303。下混合器1303被配置为接收MASA流(比特流)、传输音频信号类型202以及可能的附加参数204,并被配置为基于所确定的传输音频信号类型来将两个传输音频信号下混合成一个传输音频信号。

[0303] 下混合器1303包括传输音频信号和空间元数据提取器/解码器501。其被配置为以与如本文中所讨论的传输音频信号类型确定器中发现的相同的方式来接收MASA流并输出传输音频信号502和空间元数据522。在一些实施例中,提取器/解码器501是先前描述的提取器/解码器或提取器/解码器的独立的实例。所得到的传输音频信号502可以被转发到时间/频率变换器503。此外,所得到的空间元数据522可以被转发到信号复用器1507。

[0304] 在一些实施例中,下混合器1303包括时间/频率变换器503。时间/频率变换器503被配置为接收传输音频信号502,并将它们转换到时频域。合适的变换例如包括短时傅立叶变换(STFT)和复调制正交镜像滤波器组(QMF)。所得到的信号被标示为 $S_i(b, n)$ ,其中, $i$ 是信道索引, $b$ 是频率点索引, $n$ 是时间索引。如果音频提取和/或解码的输出已经在时频域中,则可以省略该块,或者可替代地,它可以包含从一个时频域表示到另一个时频域表示的变换。T/F域传输音频信号504可以被转发到原型信号创建器1511。在一些实施例中,时间/频率变换器503是如先前所述的相同的时间/频率变换器或独立的实例。

[0305] 在一些实施例中,下混合器1303包括原型信号创建器1511。原型信号创建器1511被配置为接收T/F域传输音频信号504、传输音频信号类型202以及可能的附加参数204。T/F原型信号1512进而可以被输出到原型能量确定器1503和原型以匹配目标能量均衡器1505。

[0306] 在一些实施例中,原型信号创建器1511被配置为基于所接收的传输音频信号类型,使用两个传输音频信号来创建用于单声道传输音频信号的原型信号。例如,可以使用下式。

[0307] 如果 $T(n) = \text{"spaced"}$ ,

[0308]  $M_{\text{proto}}(b, n) = S_0(b, n)$  .

[0309] 如果 $T(n) = \text{"downmix"}$ 或 $T(n) = \text{"coincident"}$ ,



[0310]  $M_{\text{proto}}(b, n) = S_0(b, n) + S_1(b, n)$

[0311] 在一些实施例中,下混合器1303包括目标能量确定器1501。目标能量确定器1501被配置为接收T/F域传输音频信号504,并生成目标能量值为传输音频信号的能量之和:

[0312]  $E_{\text{target}}(b, n) = |S_0(b, n)|^2 + |S_1(b, n)|^2$

[0313] 目标能量值进而可以被传递给原型以匹配目标均衡器1505。

[0314] 在一些实施例中,下混合器1303包括原型能量确定器1503。原型能量确定器1503被配置为接收T/F原型信号1512,并确定能量值例如为:

[0315]  $E_{\text{proto}}(b, n) = |M_{\text{proto}}(b, n)|^2$

[0316] 原型能量值进而可以被传递给原型以匹配目标均衡器1505。

[0317] 在一些实施例中,下混合器1303包括原型以匹配目标能量均衡器1505。在一些实施例中,原型以匹配目标能量均衡器1505被配置为接收T/F原型信号1502、原型能量值和目标能量值。在一些实施例中,均衡器1505被配置为首先随时间对能量进行平滑,例如使用以下式:

[0318]  $E'_x(b, n) = a_5 E_x(b, n) + b_5 E'_x(b, n-1)$

[0319] 其中, $a_5$ 和 $b_5$ 是平滑系数(例如, $a_5=0.1$ , $b_5=1-a_5$ )。进而,均衡器1505被配置为确定均衡增益为:

[0320] 
$$g_{\text{eq}}(b, n) = \sqrt{\frac{E'_{\text{target}}(b, n)}{E'_{\text{proto}}(b, n)}}$$

[0321] 进而,可以使用这些增益来均衡原型信号,诸如,

[0322]  $M(b, n) = g_{\text{eq}}(b, n) M_{\text{proto}}(b, n)$

[0323] 被均衡的原型信号被传递给逆T/F变换器707。

[0324] 在一些实施例中,下混合器1303包括被配置为将均衡器的输出转换成时域版本的逆T/F变换器707。被时域均衡的音频信号(单声道信号)1510进而被传递给传输音频信号和空间元数据复用器1507(或复用器)。

[0325] 在一些实施例中,下混合器1303包括传输音频信号和空间元数据复用器1507(或复用器)。传输音频信号和空间元数据复用器1507(或复用器)被配置为接收空间元数据522和单声道音频信号1510,并将它们复用以重新生成合适的输出格式(例如,具有仅一个传输音频信号的MASA流)1506。在一些实施例中,输入单声道音频信号采用脉冲编码调制(PCM)的形式。在这种实施例中,信号可以被编码并且被复用。在一些实施例中,可以省略复用,并且在音频编码器中直接使用单声道传输音频信号和空间元数据。

[0326] 在一些实施例中,图15中所示的装置的输出是单声道PCM音频信号1510,其中,空间元数据被丢弃。

[0327] 在一些实施例中,可以实现其他参数,例如,在一些实施例中,当类型为“spaced”时,可以估计间隔麦克风距离。

[0328] 关于图16,示出了图15中所示的装置的示例操作。

[0329] 因此,如图16中步骤1601所示,在一些实施例中,第一操作是从MASA流(或比特流)中提取和/或解码传输音频信号和元数据。

[0330] 如图16中步骤1603所示,下一操作可以是对传输音频信号进行时频域变换。

[0331] 进而,如图16中步骤1605所示,该方法包括基于时频域传输信号并进一步基于传输音频信号类型(并进一步基于附加参数)来创建原型音频信号。

[0332] 此外,如图16中步骤1604所示,在一些实施例中,该方法被配置为基于被变换的传输音频信号来生成、确定或计算目标能量值。

[0333] 此外,如图16中步骤1606所示,在一些实施例中,该方法被配置为基于原型音频信号来生成、确定或计算原型音频信号能量值。

[0334] 在已确定能量之后,如图16中步骤1607所示,该方法还可以均衡原型音频信号以匹配目标音频信号能量。

[0335] 如图16中步骤1609所示,被均衡的原型信号(单通道信号)进而可以被逆时频域变换以生成时域单声道信号。

[0336] 如图16中步骤1610所示,时域单声道音频信号进而可以(可选地被编码并)与空间元数据复用。

[0337] 进而,如图16中步骤1611所示,可以输出被复用的音频信号(作为MASA数据流)。

[0338] 如上所述,所示出的框图仅仅是可能的实现的一个示例。其他实际实现可以与上述示例不同。例如,实现可以没有独立的T/F变换器。

[0339] 此外,在一些实施例中,可以使用任何利用音频声道和(空间)元数据的合适的比特流,而不是具有如上所示的输入MASA流。此外,在一些实施例中,IVAS编解码器可以由任何其他合适的编解码器(例如,具有音频通道和空间元数据的操作模式的编解码器)来代替。

[0340] 在一些实施例中,可以使用传输音频信号类型确定器来估计除了传输音频信号类型之外的其他参数。例如,可以估计麦克风的间距。麦克风的间距可以是可能的附加参数204的示例。在一些实施例中,这可以通过如下操作来实现:检查 $E_{\text{sum}}(b,n)$ 和 $E_{\text{sub}}(b,n)$ 的局部最大值和最小值的频率,基于这些频率来确定麦克风之间的时间延迟,以及基于延迟和所估计的到达方向(在空间元数据中可用)来估计间距。还存在用于估计两个信号之间的延迟的其他方法。

[0341] 关于图17,示出了可以被用作分析或合成设备的示例电子设备。该设备可以是任何合适的电子设备或装置。例如,在一些实施例中,设备1700是移动设备、用户设备、平板电脑、计算机、计算机、音频回放装置等。

[0342] 在一些实施例中,设备1700包括至少一个处理器或中央处理单元1707。处理器1707可以被配置为执行各种程序代码,诸如本文所述的方法。

[0343] 在一些实施例中,设备1700包括存储器1711。在一些实施例中,至少一个处理器1707被耦合到存储器1711。存储器1711可以是任何合适的存储装置。在一些实施例中,存储器1711包括用于存储可在处理器1707上实现的程序代码的程序代码部分。此外,在一些实施例中,存储器1711还可以包括用于存储数据(例如根据本文所述的实施例已被处理或将要被处理的数据)的存储数据部分。只要需要,被存储在程序代码部分内的所实现的程序代码和被存储在存储数据部分内的数据可以经由存储器-处理器耦合而被处理器1707取得。

[0344] 在一些实施例中,设备1700包括用户接口1705。在一些实施例中,用户接口1705可以被耦合至处理器1707。在一些实施例中,处理器1707可以控制用户接口1705的操作并从用户接口1705接收输入。在一些实施例中,用户接口1705可以使得用户能够例如经由小键

盘向设备1700输入命令。在一些实施例中,用户接口1705可以使得用户能够从设备1700获得信息。例如,用户接口1705可以包括被配置为向用户显示来自设备1700的信息的显示器。在一些实施例中,用户接口1705可以包括触摸屏或触摸接口,其既能够使信息被输入到设备1700中,又能够向设备1700的用户显示信息。在一些实施例中,用户接口1705可以是用于如与本文中所描述的位置确定器通信的用户接口。

[0345] 在一些实施例中,设备1700包括输入/输出端口1709。在一些实施例中,输入/输出端口1709包括收发机。在这种实施例中,收发机可以被耦合到处理器1707,并且被配置为例如经由无线通信网络实现与其他装置或电子设备的通信。在一些实施例中,收发机或任何合适的收发机或发射机和/或接收机部件可以被配置为经由有线或有线耦合来与其他电子设备或装置通信。

[0346] 收发机可以通过任何合适的已知通信协议来与其他装置通信。例如,在一些实施例中,收发机可以使用适当的通用移动通信系统 (UMTS) 协议、诸如IEEE 802.X之类的无线局域网 (WLAN) 协议、诸如蓝牙之类的合适的短距射频通信协议、或红外数据通信路径 (IRDA)。

[0347] 收发机输入/输出端口1709可以被配置为接收信号,并且在一些实施例中,通过使用执行合适的代码的处理器1707来确定如本文中所描述的参数。

[0348] 在一些实施例中,设备1700可以被用作合成设备的至少一部分。输入/输出端口1709可以被耦合到任何合适的音频输出,例如被耦合到多通道扬声器系统和/或头戴式耳机(其可以是头戴式或非头戴式耳机)等等。

[0349] 通常,本发明的各种实施例可以采用硬件或专用电路、软件、逻辑或其任何组合来实现。例如,一些方面可以采用硬件来实现,而其他方面可以采用可由控制器、微处理器或其他计算设备执行的固件或软件来实现,但是本发明不限于此。尽管本发明的各个方面可以被图示和描述为框图、流程图或使用一些其他图形表示,但是众所周知,本文所述的这些框、装置、系统、技术或方法可以作为非限制示例采用硬件、软件、固件、专用电路或逻辑、通用硬件或控制器或其他计算设备或其某种组合来实现。

[0350] 本发明的实施例可以通过可由移动设备的数据处理器(诸如在处理器实体中)执行的计算机软件来实现,或者由硬件、或者由软件和硬件的组合来执行。此外在这一点上,应当注意,如附图中的逻辑流程的任何块可以表示程序步骤、或者互连的逻辑电路、块和功能、或者程序步骤和逻辑电路、块和功能的组合。该软件可以被存储在诸如存储器芯片或在处理器内实现的存储器块之类的物理介质上,诸如硬盘或软盘之类的磁性介质上、以及诸如DVD及其数据变体CD之类的光学介质上。

[0351] 存储器可以是适合于本地技术环境的任何类型,并且可以使用任何适当的数据存储技术来实现,诸如基于半导体的存储器设备、磁存储器设备和系统、光学存储器设备和系统、固定存储器和可移除存储器。数据处理器可以是适合于本地技术环境的任何类型,并且作为非限制性示例,可以包括通用计算机、专用计算机、微处理器、数字信号处理器 (DSP)、专用集成电路 (ASIC)、基于多核处理器架构的门级电路和处理器中的一个或多个。

[0352] 可以在诸如集成电路模块之类的各种组件中实践本发明的实施例。集成电路的设计总体上是高度自动化的过程。复杂而功能强大的软件工具可用于将逻辑级设计转换为准备在半导体衬底上蚀刻和形成的半导体电路设计。

[0353] 程序,诸如由加利福尼亚州山景城的Synopsys公司和加利福尼亚州圣何塞的Cadence Design所提供的程序,可以使用完善的设计规则以及预先存储的设计模块库来自动对导体进行布线并将组件定位在半导体芯片上。一旦完成了半导体电路的设计,就可以将标准化电子格式(例如,Opus、GDSII等)的所得设计传送到半导体制造设施或“fab”进行制造。

[0354] 前面的描述已经通过示例性和非限制性示例提供了本发明的示例性实施例的完整和有益的描述。然而,当结合附图和所附权利要求书阅读时,鉴于以上描述,各种修改和改编对于相关领域的技术人员而言将变得显而易见。但是,本发明的教导的所有这些和类似的修改仍将落入所附权利要求书所限定的本发明的范围内。

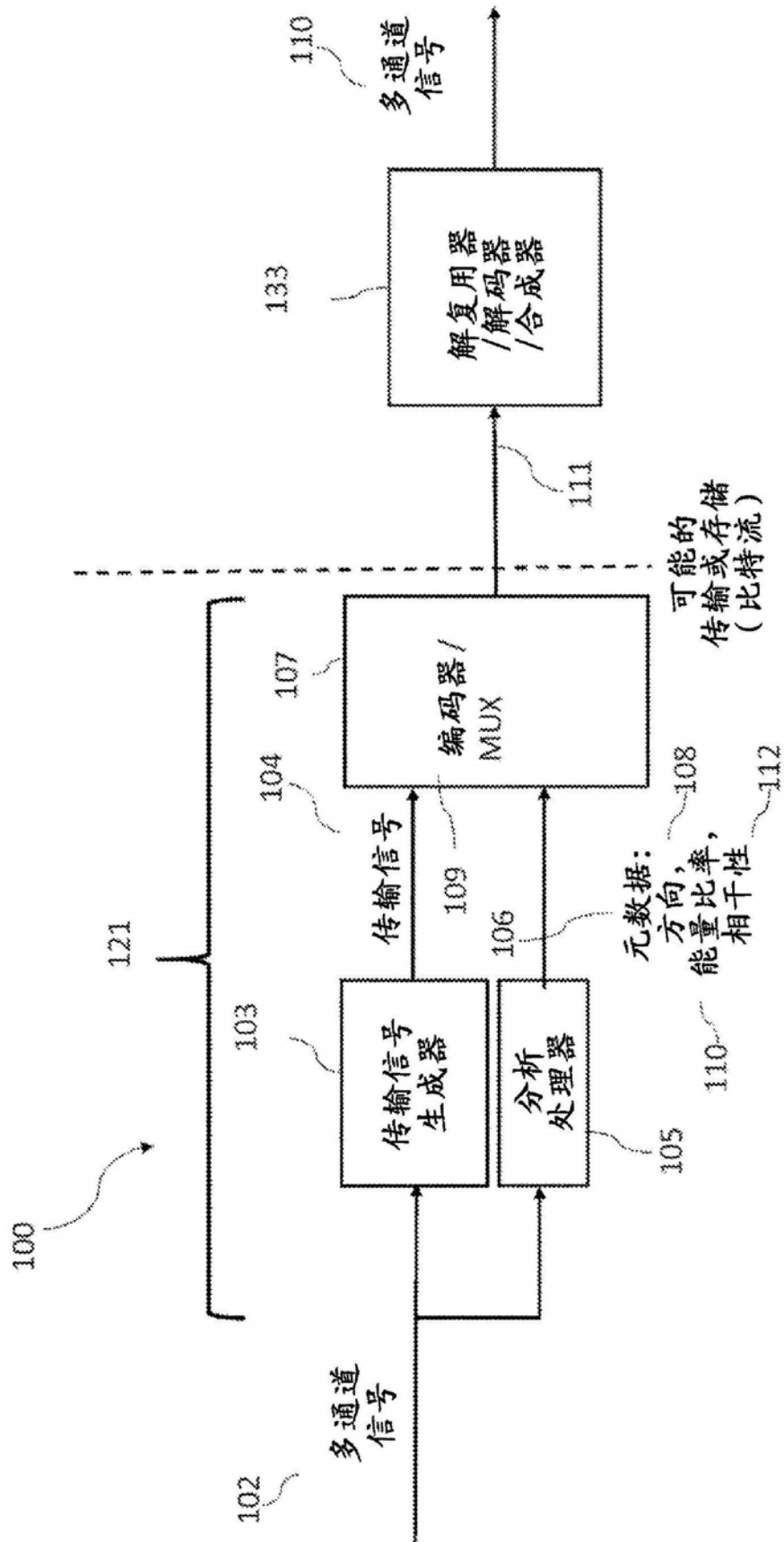


图1

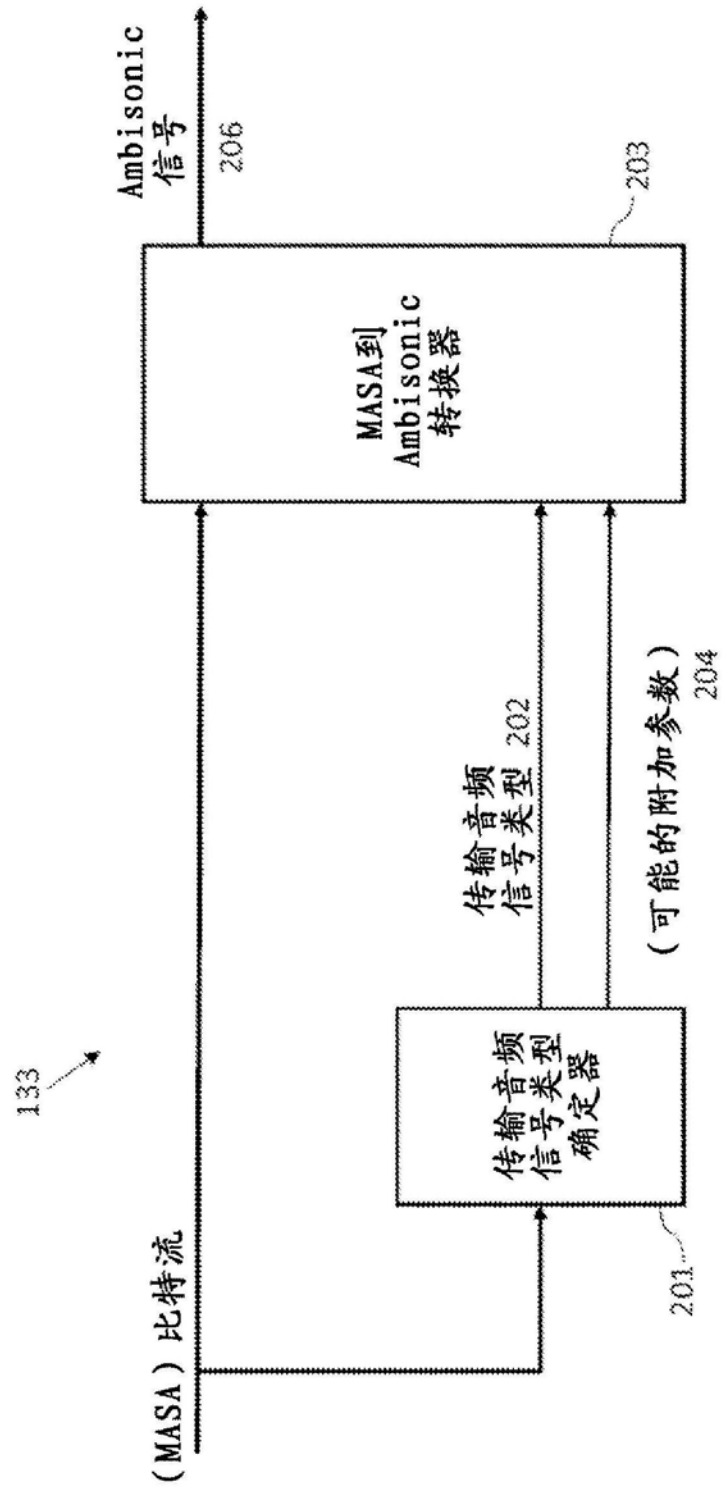


图2

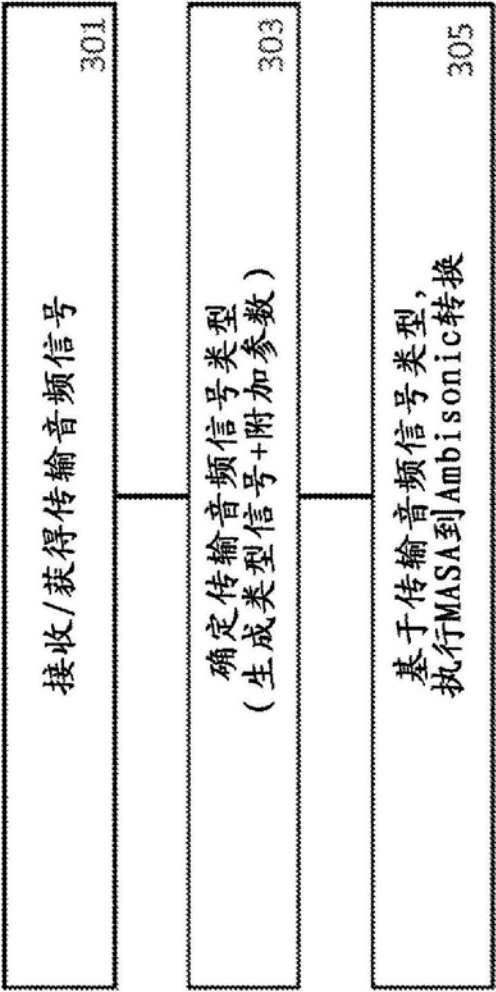


图3

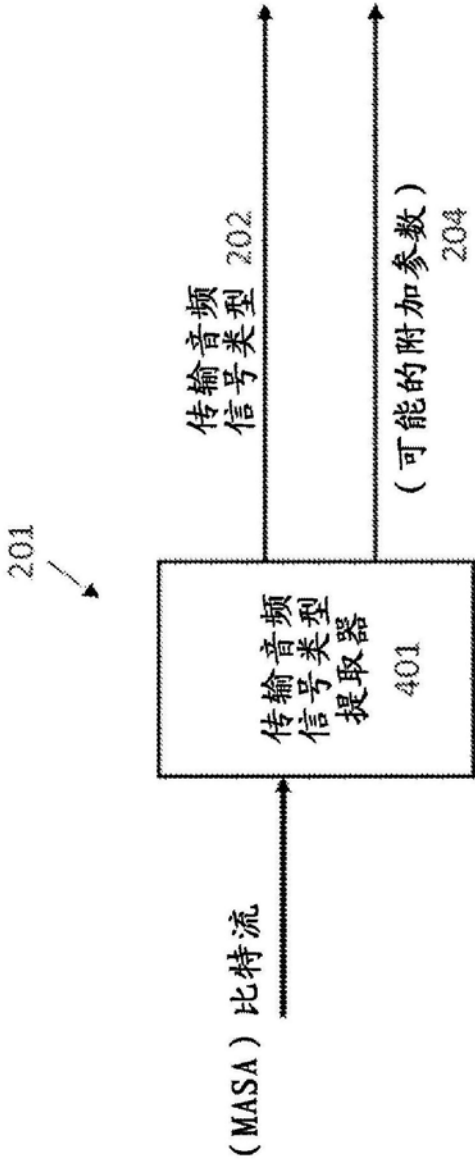


图4



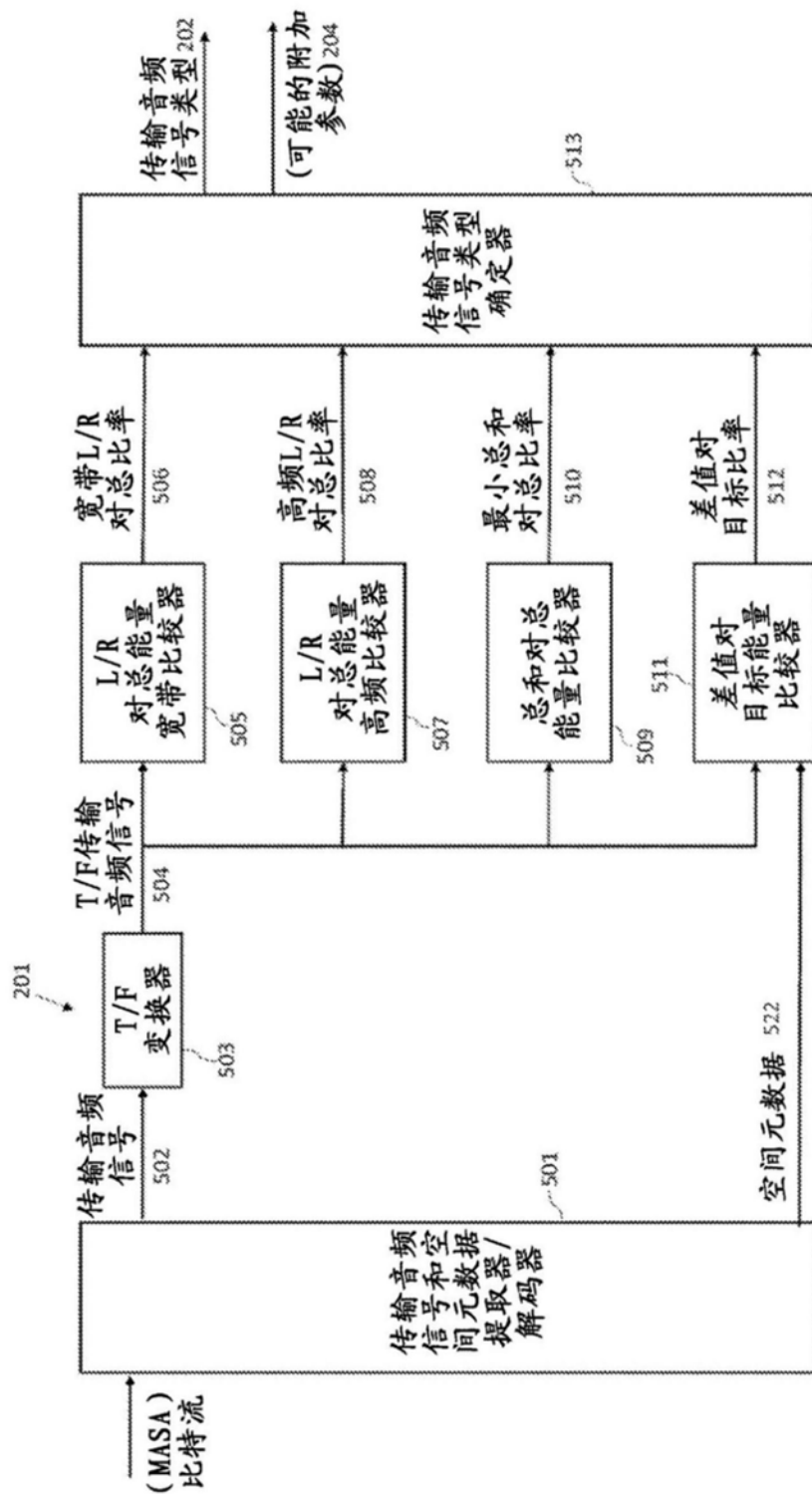


图5

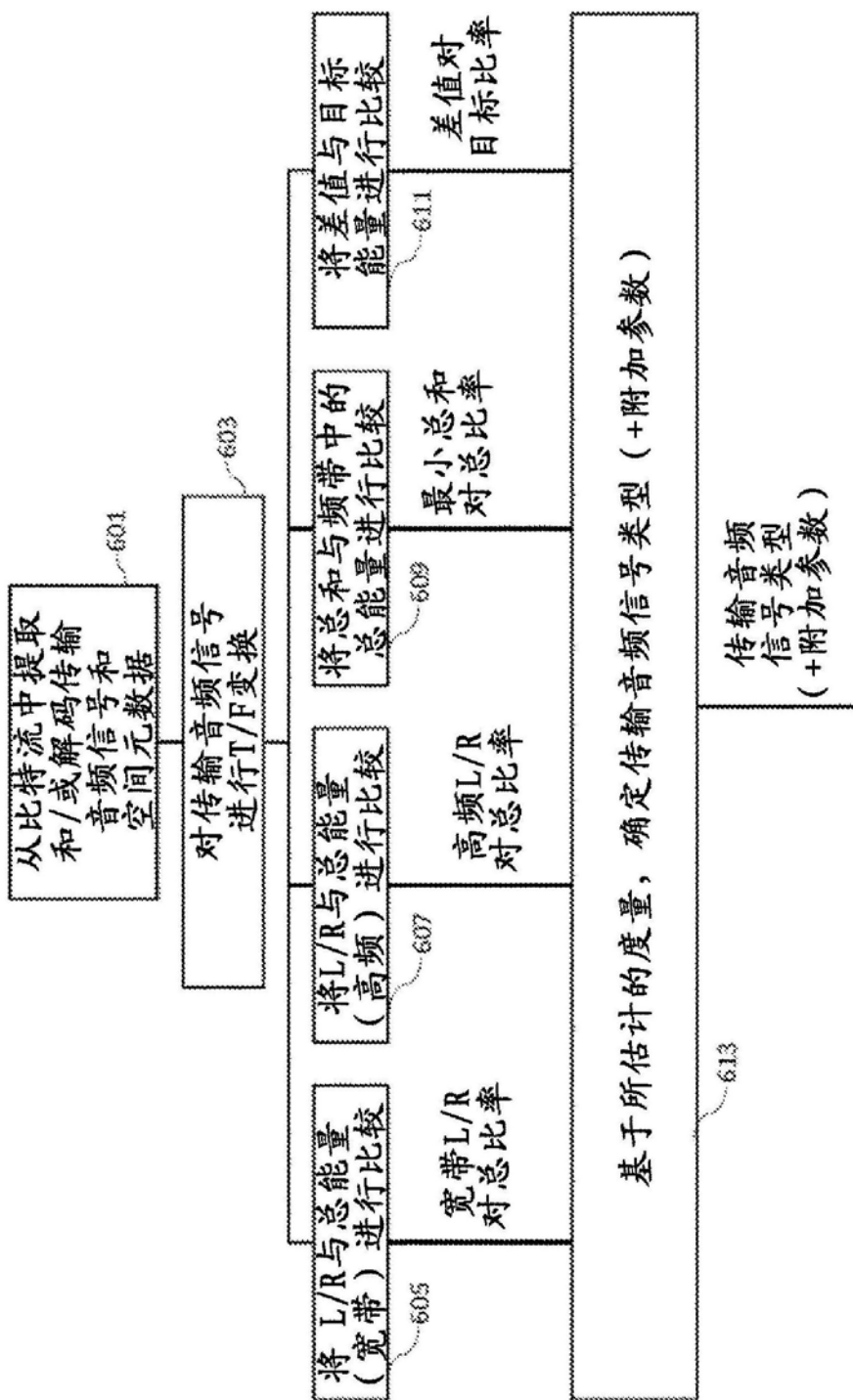


图6

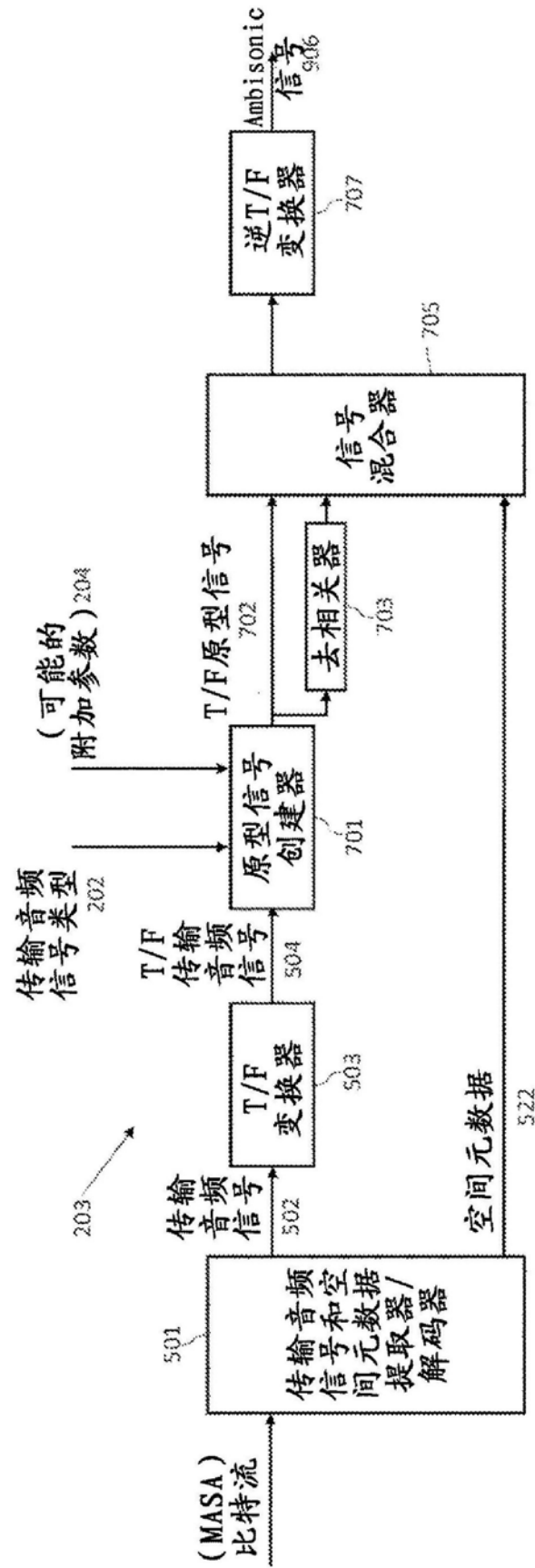


图7

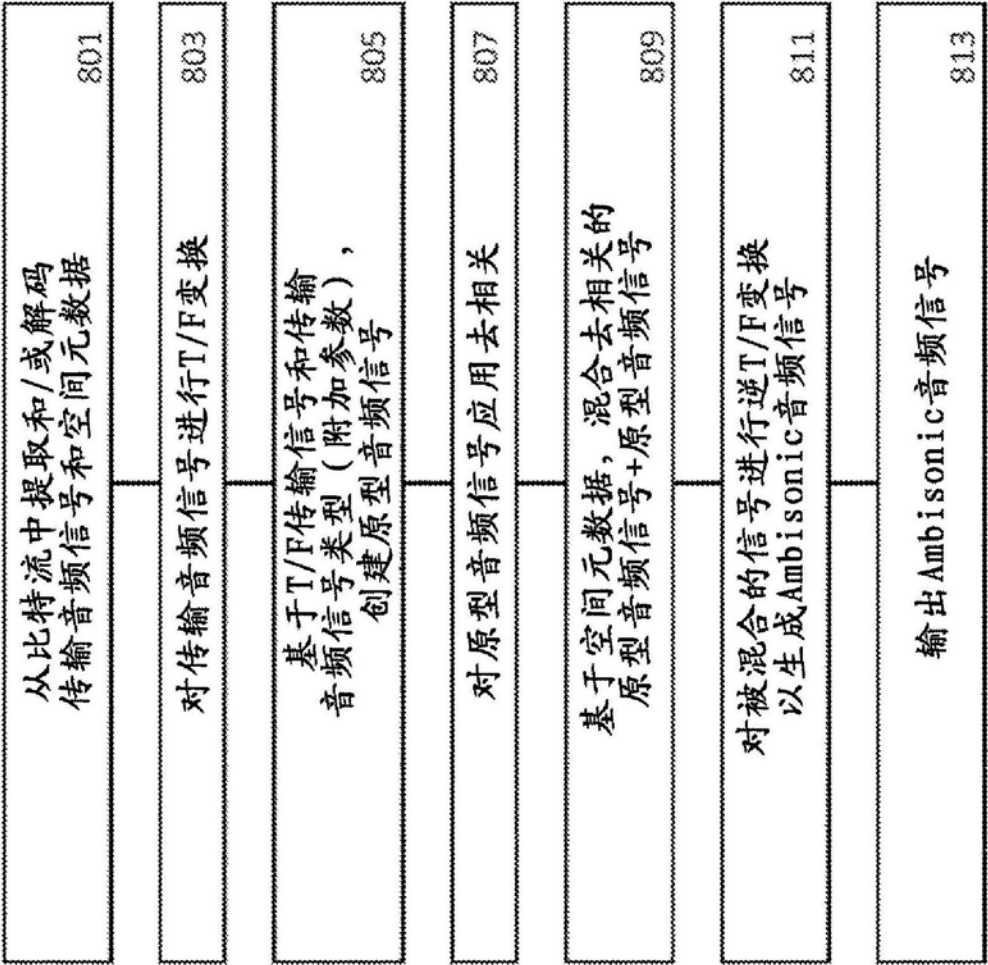


图8

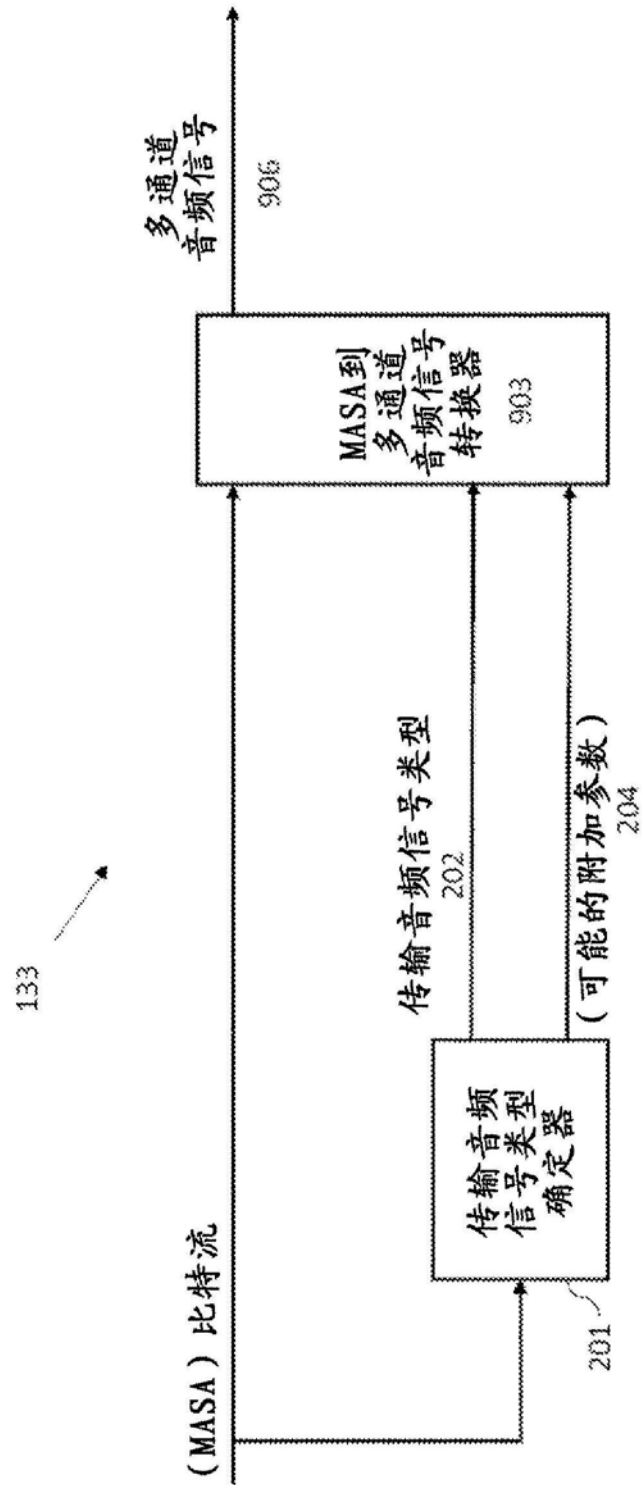


图9

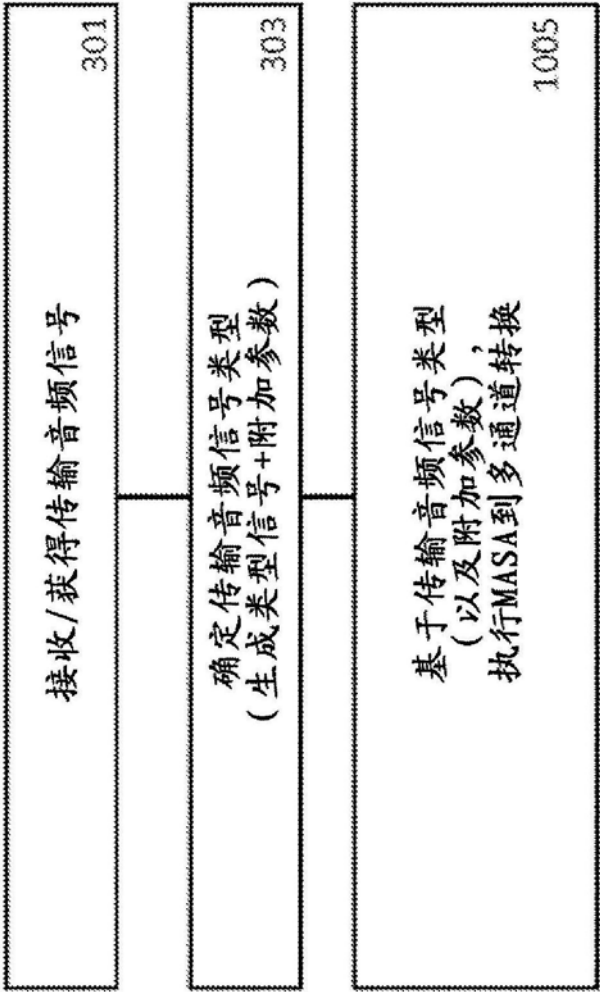


图10

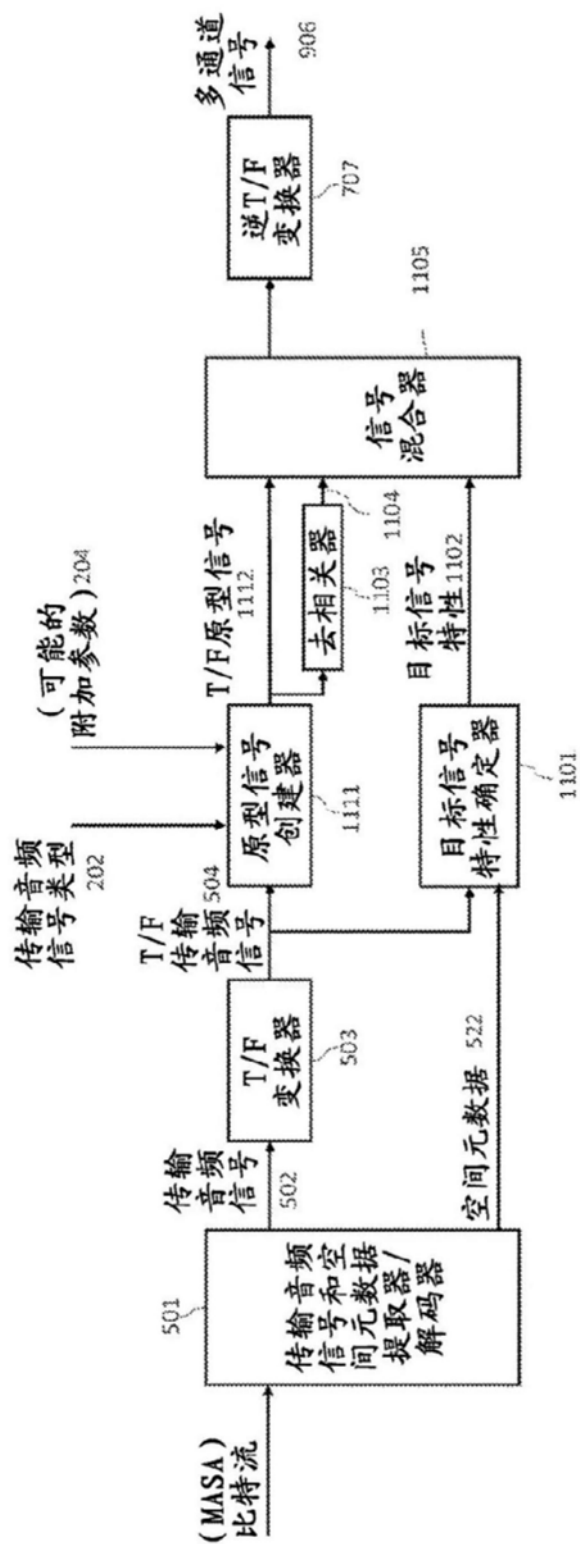


图11

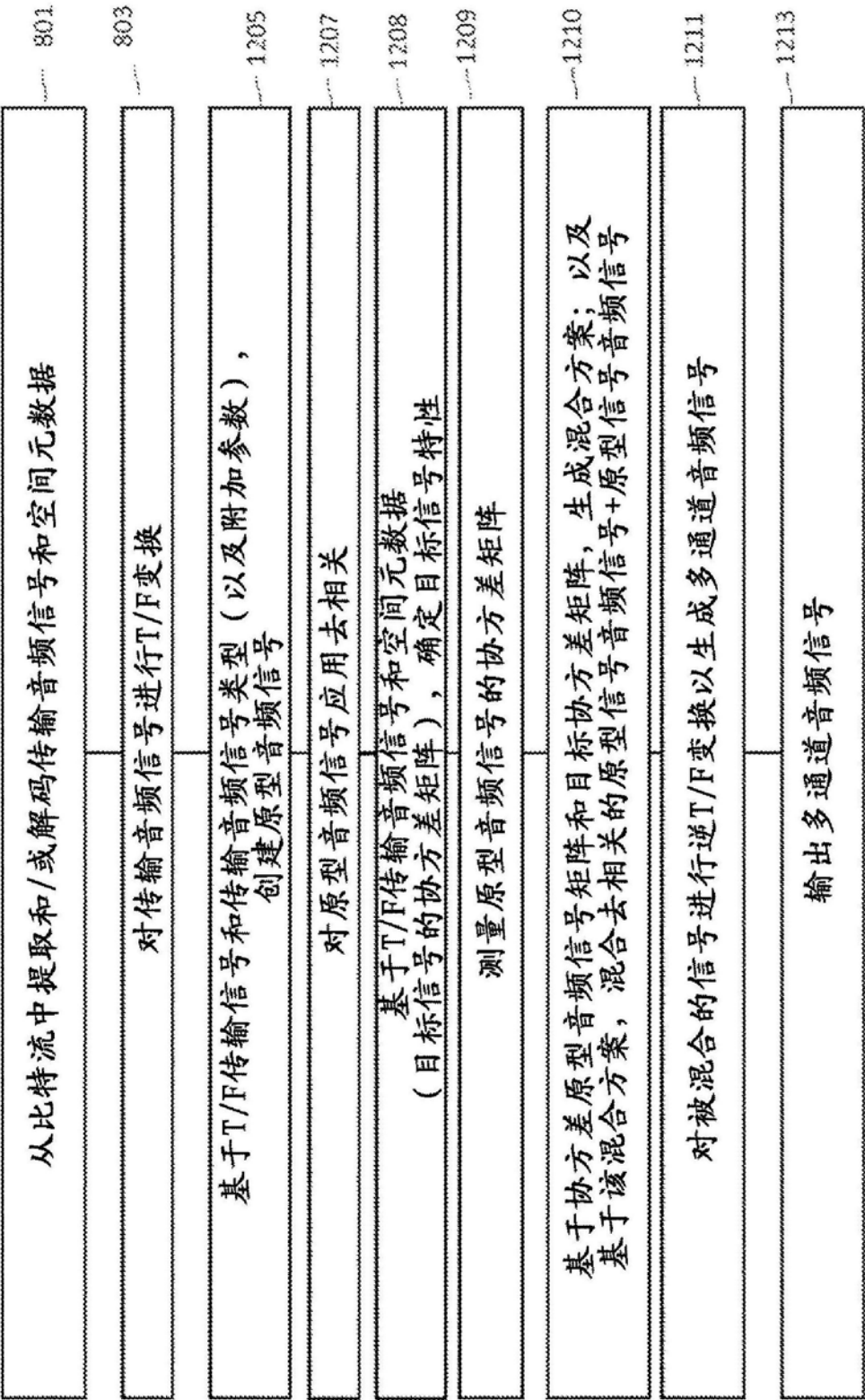


图12



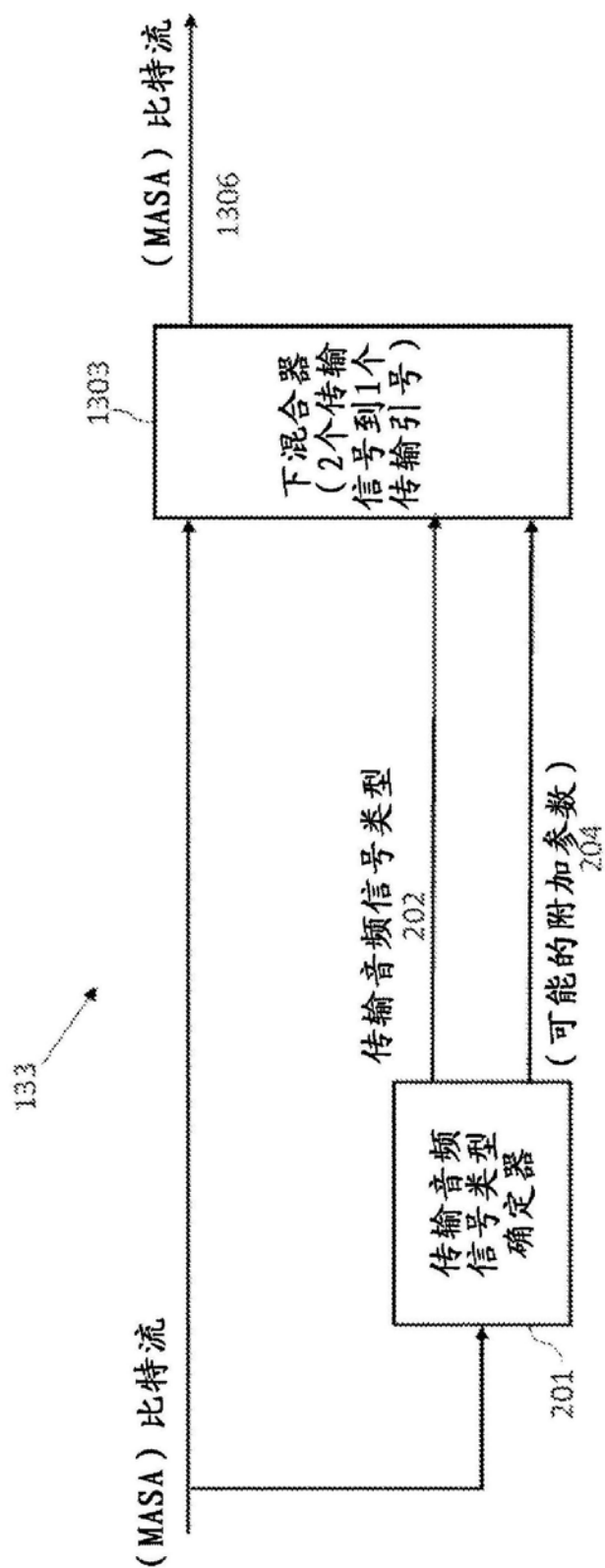


图13

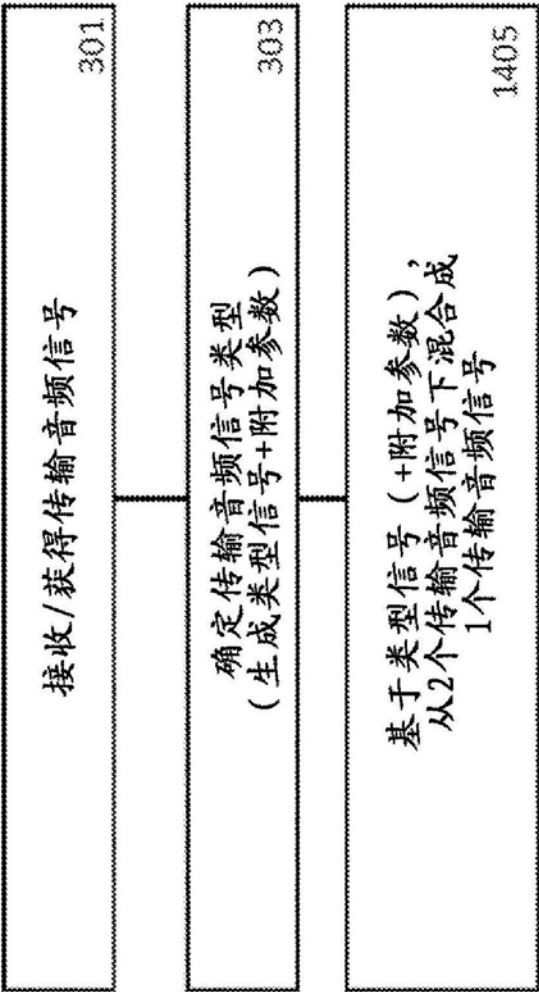


图14

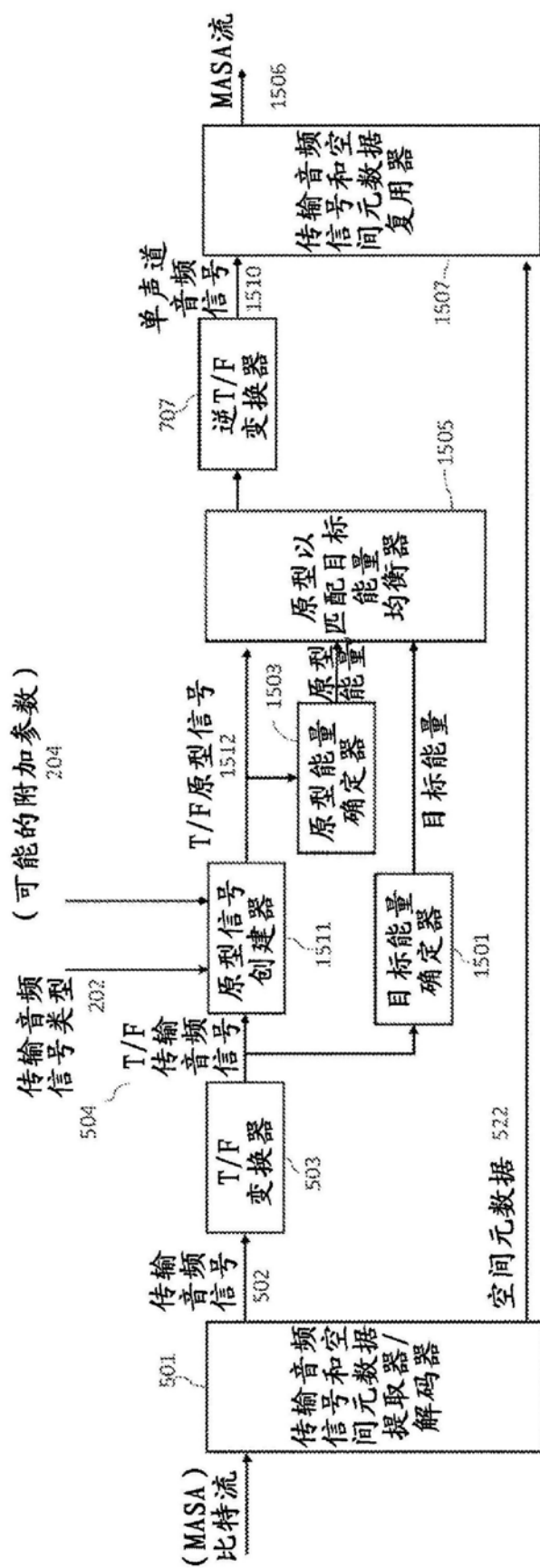


图15

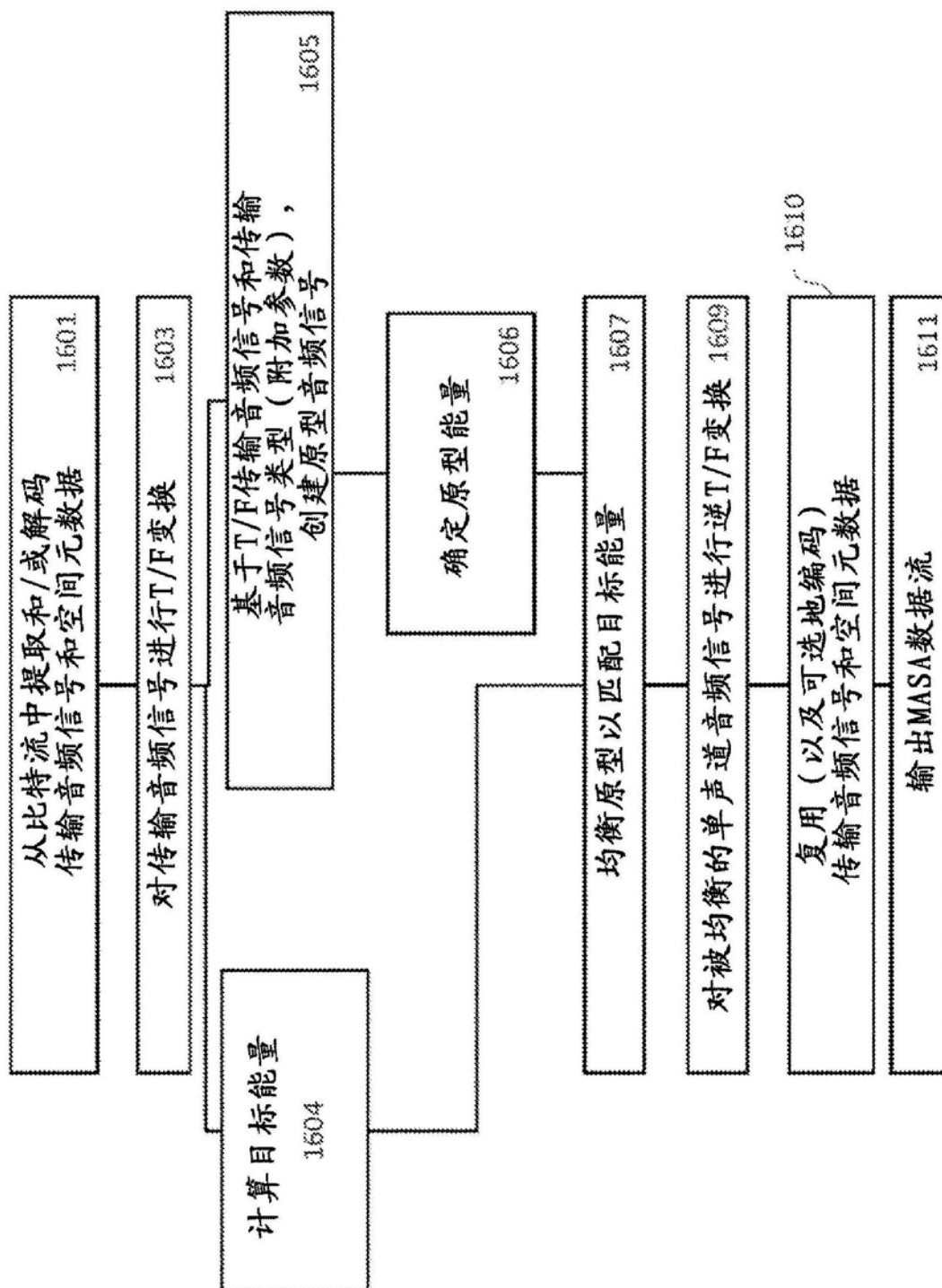


图16

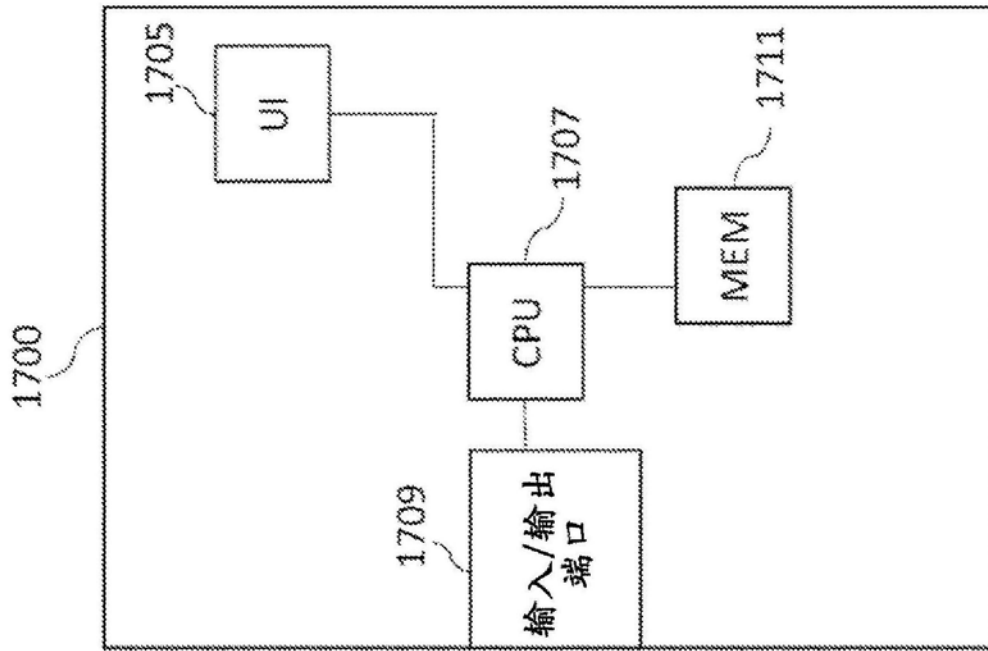


图17