(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119998874 A
(43)申请公布日 2025.05.13

(21)申请号 202380071028.1

(22)申请日 2023.09.15

(30)优先权数据
63/378,500 2022.10.05 US
63/534,458 2023.08.24 US

(85)PCT国际申请进入国家阶段日
2025.04.03

(86)PCT国际申请的申请数据
PCT/EP2023/075436 2023.09.15

(87)PCT国际申请的公布数据
WO2024/074284 EN 2024.04.11

(71)申请人 杜比国际公司
地址 爱尔兰都柏林

(72)发明人 H•穆特 H•普恩豪根
K•克约林 M•舒格

(74)专利代理机构 北京市汉坤律师事务所 11602
专利代理师 魏小薇 吴丽丽

(51)Int.Cl.
G10L 19/16(2006.01)
G10L 19/008(2006.01)
G10L 19/22(2006.01)

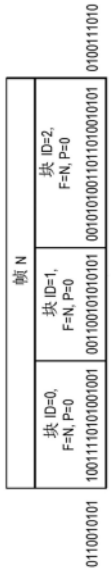
权利要求书3页 说明书25页 附图18页

(54)发明名称

用于对音频比特流进行高效编码和解码的方法、装置和介质

(57)摘要

一种用于生成包括多个音频信号的音频节目的经编码比特流的帧的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:针对所述多个音频信号中的每个音频信号接收指示与相应音频信号相关联的回放设备的信息;对与相应回放设备相关联的一个或多个音频信号进行编码以获得一个或多个经编码音频信号;将与所述相应回放设备相关联的所述一个或多个经编码音频信号组合到所述帧的第一独立块中;将所述多个音频信号中的一个或多个其他音频信号编码到一个或多个附加独立块中;以及将所述第一独立块和所述一个或多个附加独立块组合到所述经编码比特流的所述帧中。



CN 119998874 A

1.一种用于生成包括多个音频信号的音频节目的经编码比特流的帧的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:

针对所述多个音频信号中的每个音频信号接收指示与相应音频信号相关联的回放设备的信息;

对与相应回放设备相关联的一个或多个音频信号进行编码以获得一个或多个经编码音频信号;

将与所述相应回放设备相关联的所述一个或多个经编码音频信号组合到所述帧的第一独立块中;

将所述多个音频信号中的一个或多个其他音频信号编码到一个或多个附加独立块中;以及

将所述第一独立块和所述一个或多个附加独立块组合到所述经编码比特流的所述帧中。

2.如权利要求1所述的方法,其中,两个或更多个音频信号与所述回放设备相关联,并且所述两个或更多个音频信号中的每个音频信号是旨在由所述回放设备的相应驱动器回放的带限信号,并且其中,针对所述带限信号中的每个带限信号使用不同的编码技术。

3.如权利要求2所述的方法,其中,针对所述带限信号中的每个带限信号使用不同的心理声学模型和/或不同的比特分配技术。

4.如前述权利要求中任一项所述的方法,其中,所述经编码信号的瞬时帧速率是可变的,并且受缓冲器充满度模型的约束。

5.如权利要求1所述的方法,其中,对与所述相应回放设备相关联的一个或多个音频信号进行编码包括将与所述相应回放设备相关联的所述一个或多个音频信号以及与一个或多个附加回放设备相关联的一个或多个附加音频信号联合编码到所述帧的所述第一独立块中。

6.如权利要求5所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括跨两个或更多个音频信号共享一个或多个比例因子。

7.如权利要求6所述的方法,其中,所述两个或更多个音频信号是空间相关的。

8.如权利要求7所述的方法,其中,所述两个或更多个空间相关音频信号包括左水平声道、左上声道、右水平声道或右上声道。

9.如权利要求5所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括应用耦合工具,所述耦合工具包括:

将两个或更多个音频信号组合成高于指定频率的复合信号;以及

针对所述两个或更多个音频信号中的每个音频信号确定与所述复合信号的能量和每个相应信号的能量相关的比例因子。

10.如权利要求5所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括对多于两个信号应用联合编码工具。

11.一种用于从经编码比特流的帧中解码与回放设备相关联的一个或多个音频信号的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:

从所述经编码比特流中识别同与所述回放设备相关联的所述一个或多个音频信号相对应的独立的经编码数据块;

从所述经编码比特流中提取所识别的独立的经编码数据块;

从所述独立的经编码数据块中解码与所述回放设备相关联的所述一个或多个音频信号以获得一个或多个经解码音频信号;

从所述经编码比特流中识别与一个或多个附加音频信号相对应的一个或多个附加的独立的经编码数据块;以及

解码或跳过所述一个或多个附加的独立的经编码数据块。

12.如权利要求11所述的方法,其中,两个或更多个音频信号与所述回放设备相关联,并且所述两个或更多个音频信号中的每个音频信号是旨在由所述回放设备的相应驱动器回放的带限信号,并且其中,使用不同的解码技术对所述两个或更多个音频信号进行解码。

13.如权利要求12所述的方法,其中,使用不同的心理声学模型和/或不同的比特分配技术来对所述带限信号中的每个带限信号进行编码。

14.如权利要求12或权利要求13所述的方法,其中,所述经编码比特流的瞬时帧速率是可变的,并且受缓冲器充满度模型的约束。

15.如权利要求11所述的方法,其中,对与所述回放设备相关联的所述一个或多个音频信号进行解码包括从所述独立的经编码数据块中联合解码与所述相应回放设备相关联的所述一个或多个音频信号以及与一个或多个附加回放设备相关联的一个或多个附加音频信号。

16.如权利要求15所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括提取跨两个或更多个音频信号共享的比例因子。

17.如权利要求16所述的方法,其中,所述两个或更多个音频信号是空间相关的。

18.如权利要求17所述的方法,其中,所述两个或更多个空间相关音频信号包括左水平声道、左上声道、右水平声道或右上声道。

19.如权利要求15所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括应用解耦工具。

20.如权利要求19所述的方法,其中,所述解耦工具包括:

提取低于指定频率的独立解码的信号;

提取高于所述指定频率的复合信号;

从所述复合信号以及与所述复合信号的能量和相应信号的能量相关的比例因子来确定高于所述指定频率的相应解耦信号;以及

将每个独立解码的信号与相应的解耦信号组合以获得所述联合解码的信号。

21.如权利要求15所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括应用联合解码工具来提取多于两个音频信号。

22.如权利要求11所述的方法,其中,对与所述回放设备相关联的所述一个或多个音频信号进行解码包括在与所述音频信号被编码的域相同的域中对所述音频信号应用带宽扩展。

23.如权利要求22所述的方法,其中,所述域是改进离散余弦变换(MDCT)域。

24.如权利要求22或权利要求23所述的方法,其中,所述带宽扩展包括自适应噪声添加。

25.一种装置,所述装置被配置为执行如权利要求1至24中任一项所述的方法。

26.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如权利要求1至24中任一项所述的方法。

用于对音频比特流进行高效编码和解码的方法、装置和介质

[0001] 相关申请的交叉引用

[0002] 本申请要求于2022年10月5日提交的美国临时申请号63/378,500和于2023年8月24日提交的美国临时申请号63/534,458的优先权权益,这些美国临时申请中的每一个通过引用整体并入本文。

技术领域

[0003] 本公开总体上涉及音频信号处理,并且更具体地涉及用于沉浸式音频节目的音频信号在设备之间的低时延互换的音频源编码和解码。

背景技术

[0004] 音频的流式传输在当今社会很常见。随着用户对质量的期望不断提高,音频流式传输的要求变得越来越高,而且随着扬声器数量以及扬声器类型的增加,用户的设置也变得越来越复杂。流式传输通常有一部分至少通过无线链路来完成,这就要求无线链路具有良好的质量,而正如许多人所经历的那样,情况并非总是如此。

[0005] 因此,需要为用例定义一种互换格式,其中,以某种格式从云/服务器进行流式传输,并且随后在设备上转码为更适合通过无线(或者在一些情况下是有线)链路分发的低时延格式。示例性用例是家庭连接以及电话到汽车连接,但这种格式在期望将音频信号从单个设备以低时延方式分发到一个或多个连接的设备的任何场景下都可能是有益的。

[0006] 除了以无线方式发送和流式传输音频信息之外,还可以将其他类型的信息并入到流中。然后,这种其他类型的信息也会受到无线链路的质量的影响,并且可能具有与音频类似的缺点。

[0007] 因此,对于不同类型的流式传输的音频以及其他类型的信息或信号的组合,克服与无线流式传输相关联的问题将是有利的。

发明内容

[0008] 本公开的目的是利用音频以及其他类型的信息的无线流式传输来至少部分地克服以上问题。

[0009] 根据本公开的第一方面是一种用于生成包括多个音频信号的音频节目的经编码比特流的帧的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:针对所述多个音频信号中的每个音频信号接收指示与相应音频信号相关联的回放设备的信息;对与相应回放设备相关联的一个或多个音频信号进行编码以获得一个或多个经编码音频信号;将与所述相应回放设备相关联的所述一个或多个经编码音频信号组合到所述帧的第一独立块中;将所述多个音频信号中的一个或多个其他音频信号编码到一个或多个附加独立块中;以及将所述第一独立块和所述一个或多个附加独立块组合到所述经编码比特流的所述帧中。

[0010] 根据本公开的第二方面是一种用于从经编码比特流的帧中解码与回放设备相关

联的一个或多个音频信号的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:从所述经编码比特流中识别同与所述回放设备相关联的所述一个或多个音频信号相对应的独立的经编码数据块;从所述经编码比特流中提取所识别的独立的经编码数据块;从所述独立的经编码数据块中解码与所述回放设备相关联的所述一个或多个音频信号以获得一个或多个经解码音频信号;从所述经编码比特流中识别与一个或多个附加音频信号相对应的一个或多个附加的独立的经编码数据块;以及解码或跳过所述一个或多个附加的独立的经编码数据块。

[0011] 根据本公开的第三方面是一种装置,所述装置被配置为执行第一方面和/或第二方面的方法。

[0012] 根据本公开的第四方面是一种非暂态计算机可读存储介质,所述非暂态计算机可读存储介质包括一系列指令,所述一系列指令在执行时使一个或多个设备执行第一方面和/或第二方面的方法。

[0013] 在从属权利要求中限定了本公开的进一步示例。

[0014] 在一些示例中,两个或更多个音频信号与回放设备相关联,并且所述两个或更多个音频信号中的每个音频信号是旨在由所述回放设备的相应驱动器回放的带限信号,并且其中,针对所述带限信号中的每个带限信号使用不同的编码技术。在一些示例中,针对所述带限信号中的每个带限信号使用不同的心理声学模型和/或不同的比特分配技术。在示例中,经编码信号的瞬时帧速率是可变的,并且受缓冲器充满度模型的约束。

[0015] 在一些示例中,与相应回放设备相关联的一个或多个音频信号包括将与相应回放设备相关联的一个或多个音频信号以及与一个或多个附加回放设备相关联的一个或多个附加音频信号联合编码到帧的第一独立块中。在示例中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括跨两个或更多个音频信号共享一个或多个比例因子。在一些示例中,两个或更多个音频信号是空间相关的。在示例中,两个或更多个空间相关音频信号包括左水平声道、左上声道、右水平声道或右上声道。

[0016] 在一些示例中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括应用耦合工具。在一些示例中,所述耦合工具包括:将两个或更多个音频信号组合成高于指定频率的复合信号;以及针对所述两个或更多个音频信号中的每个音频信号确定与所述复合信号的能量和每个相应信号的能量相关的比例因子。在一些示例中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括对多于两个信号应用联合编码工具。

[0017] 在一些示例中,两个或更多个音频信号与回放设备相关联,并且所述两个或更多个音频信号中的每个音频信号是旨在由所述回放设备的相应驱动器回放的带限信号,并且其中,使用不同的解码技术对所述两个或更多个音频信号进行解码。在示例中,使用不同的心理声学模型和/或不同的比特分配技术来对带限信号中的每个带限信号进行编码。

[0018] 对于一些示例,经编码比特流的瞬时帧速率是可变的,并且受缓冲器充满度模型的约束。15.如权利要求11所述的方法,其中,对与所述回放设备相关联的所述一个或多个音频信号进行解码包括从所述独立的经编码数据块中联合解码与所述相应回放设备相关联的所述一个或多个音频信号以及与一个或多个附加回放设备相关联的一个或多个附加音频信号。

[0019] 在示例中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括应用解耦工具。在示例中,解耦工具包括:提取低于指定频率的独立解码的信号;提取高于所述指定频率的复合信号;从所述复合信号以及与所述复合信号的能量和相应信号的能量相关的比例因子来确定高于所述指定频率的相应解耦信号;以及将每个独立解码的信号与相应的解耦信号组合以获得所述联合解码的信号。

[0020] 在本公开中,帧表示所有信号的整体的时间片。块流表示会话持续时间内的信号集合。块表示块流的一个帧。对于具有给定采样频率的数字音频,帧大小等于任何音频信号的帧中的音频样本的数量。帧大小通常在会话的持续时间内保持恒定。

[0021] 唤醒词可以包括一个单词,或包括固定顺序的两个或更多个单词的短语。

[0022] 在整个本公开中,包括在权利要求中,在广义上使用表达“系统”来表示设备、系统或子系统。例如,实施解码器的子系统可以被称为解码器系统,并且包括这样的子系统的系统(例如,响应于多个输入而生成X个输出信号的系统,其中,子系统生成M个输入,而其他X-M个输入是从外部源接收的)也可以被称为解码器系统。

附图说明

[0023] 将参考附图更详细地描述本公开的示例。在以下附图中,相同的附图标记用于指代相同的元素。尽管以下附图描绘了各种示例,但是一种或多种实施方式不限于附图中所描绘的示例。

[0024] 图1图示了家庭连接低时延转码的示例。

[0025] 图2图示了汽车连接音频流式传输的示例。

[0026] 图3图示了通过使用无线设备进行的增强TV音频流式传输的示例。

[0027] 图4图示了利用简单的无线扬声器进行的音频流式传输的示例。

[0028] 图5图示了从比特流元素到设备的元数据映射的示例。

[0029] 图6图示了如何部署音频流式传输的简单灵活渲染的示例。

[0030] 图7图示了将灵活渲染数据映射到比特流元素的示例。

[0031] 图8图示了当用多个设备播放音频和收听命令时回声参考的信令的示例。

[0032] 图9图示了帧、块和包如何彼此相关的示例。

[0033] 图10图示了当前MPEG-4结构中的进一步的编解码器信息的示例。

[0034] 图11图示了当前MPEG-4结构中的更进一步的编解码器信息的示例。

[0035] 图12图示了整合收听能力和语音识别的示例。

[0036] 图13图示了包括多个块的帧的示例。

[0037] 图14图示了包括多个块的比特流的示例。

[0038] 图15图示了具有不同优先级的块的示例。

[0039] 图16图示了包括具有不同优先级的多个块的比特流的示例。

[0040] 图17图示了具有不同优先级的帧的示例。

[0041] 图18图示了包括具有不同优先级的帧的比特流的示例。

具体实施方式

[0042] 现在将参考附图中图示的各种示例来描述本发明的原理。应当理解,对这些示例

的描绘仅仅是为了使得本领域技术人员能够更好地理解并进一步实施本发明;并不旨在以任何方式限制本发明的范围。

[0043] 在图1中,沉浸式音频流从云或服务器10进行流式传输,并且在TV或集线器设备20上解码。沉浸式音频流可以以任何现有格式编码,包括例如杜比数字加(Dolby Digital Plus)、AC-4等。输出随后被转码为低时延互换格式以进一步传输到连接的设备30,该设备优选地通过本地无线连接(例如,WiFi软接入点)或蓝牙连接来连接。低时延通常取决于各种因素,比如例如帧大小、采样速率、硬件和/或软件计算资源等,而且低时延通常小于40ms、20ms或10ms。

[0044] 在另一种典型用例中,电话从云或服务器取得沉浸式音频流,并将其转码为互换格式,并且随后传输到连接的汽车。在图2的所图示示例中,移动设备(例如,电话或平板计算机)20连接到服务器10并接收沉浸式音频流,执行到低时延互换格式的低时延转码,并且将经转码的信号传输到支持沉浸式音频回放的汽车30。沉浸式音频流的示例是包括采用杜比全景声(Dolby Atmos)格式的音频的流,并且支持沉浸式音频回放的汽车30的示例是被配置为回放杜比全景声沉浸式格式的汽车。

[0045] 通常,互换格式优选地具有低时延、低编码和解码复杂度、扩展到高质量的能力以及合理的编码效率。该格式优选地还支持可配置的时延,使得时延可以与效率以及错误弹性相权衡,以便在变化的连接条件下可操作。

[0046] 带有具有或不具有收听能力的无线扬声器的集线器或增强显示器

[0047] 图3中图示了驱动一组无线扬声器30的集线器20的示例,或者可能具有内置扬声器30的显示器(例如,电视机或TV)20用若干个无线扬声器30增强。该增强表明,在显示器20包括扬声器30的示例中,显示器20也是音频再现的一部分。无线扬声器/设备30可能正在接收同一完整信号(广播模式,在图3的左边图示)或为特定设备定制的单独流(单播多点模式,在图3的右边图示)。

[0048] 每个扬声器可以包括多个驱动器,该多个驱动器覆盖同一声道的不同频率范围,或对应于规范映射中的不同声道。例如,扬声器可以具有两个驱动器,其中一个驱动器可以是输出与高度声道相对应的信号以模拟高位扬声器的向上发声驱动器。无线设备可以具有收听能力(例如,“智能扬声器”),并且因此可能需要回声管理,这可能需要扬声器接收一个或多个回声参考。回声参考可以是本地的(例如,对于同一扬声器/设备),或者可以替代地表示来自附近的其他扬声器/设备的相关信号。

[0049] 扬声器可以被放置在任意位置,在这种情况下,可以执行所谓的“灵活渲染”,由此渲染考虑了扬声器的实际位置(例如,与规范假设相反,由此假设扩音器位于固定的预定义位置)。灵活渲染可以在集线器或TV中进行,其中,随后以广播模式将经渲染的信号发送到扬声器/设备,其中,每个经渲染的信号被发送到每个设备,并且每个相应设备提取并输出适当的信号,或者作为单独的流被发送到单独设备。可替代地,灵活渲染可以在每个设备上在本地进行,由此每个设备接收完整沉浸式节目的表示,例如,基于7.1.4声道的沉浸式表示,并且从该表示渲染适合于相应设备的输出信号。

[0050] 无线设备可以通过由设备提供的软接入点或通过家庭中的本地接入点连接到集线器或TV。这可能对比特率和时延强加不同的要求。

[0051] 汽车应用中的移动设备投影

[0052] 在另一种用例中,移动设备(例如,电话或平板计算机)20从云或服务器10取得沉浸式音频流,将沉浸式音频流转码为互换格式,并且随后将经转码的流传输到连接的汽车30。在图2的示例中,启用杜比全景声的电话20连接到服务器10并接收杜比全景声流,以用于低时延转码并传输到启用杜比全景声的汽车3。在这种用例中,可以获得比起居室用例中的比特流更高的比特率,并且无线信道可能有不同的特性。汽车用例中比特率较高的原因可能是由于无线环境中噪声较小,因为与比如来自邻居、其他房间等的所有无线设备都向无线环境添加噪声的起居室相比,汽车充当一种法拉第笼并且保护内部汽车环境免受外部无线干扰。同时,与通常家庭中的所有无线设备(包括起居室无线设备)竞争可用带宽的起居室用例相比,汽车中通常很少有无线设备竞争可用无线带宽。

[0053] 对于这种用例,设想的是,要由移动设备转码并传输到汽车的信号可以是基于声道的沉浸式表示、基于对象的表示、基于场景的表示(例如,以及高保真立体声表示)或甚至不同表示的组合。对于该示例,不同的渲染架构和广播与多点可能不相关,因为通常完整的呈现将从移动设备传输到单个端点(例如,汽车)。

[0054] 互换格式的说明

[0055] 沉浸式互换格式基于利用感知激励的量化和编码的改进离散余弦变换(MDCT)构建。其具有可配置的时延,例如,在给定的采样速率下支持不同的变换大小。在48kHz和44.1kHz的采样速率下,示例性帧大小为128、256、512、1024和120、240、480、960和192、384、768个样本。

[0056] 该格式可以支持单声道、立体声、5.1和其他声道配置,包括沉浸式声道配置(例如,包括但不限于5.1.2、5.1.4、7.1.2、7.1.4、9.1.6和22.2,或由如ISO/IEC 23091-3:2018的表-2中指定的唯一声道构成的任何其他声道配置)。该格式还可以支持基于对象的音频和基于场景的表示,如高保真立体声(例如,一阶或更高阶)。该格式还可以使用适合于与现有格式(例如,如ISO/IEC 14496-3(其也可以被称为MPEG-4音频标准)、ISO 14496-1(其也可以被称为MPEG-4系统标准)、ISO 14496-12(其也可以被称为ISO基本媒体文件格式标准)和/或ISO 14496-14(其也可以被称为MP4文件格式标准)中所描述的那些现有格式)整合的信令方案。

[0057] 此外,该系统可以具有针对跳过语法元素的部分并且仅对给定扬声器的相关部分进行解码的能力的支持,针对元数据控制的灵活渲染方面的支持,如延迟对准、水平调整和均衡,针对使用具有收听能力的智能扬声器(包括发送一组信号以允许独立地馈送智能扬声器中的每个驱动器/扬声器的场景)的支持,并且通过回声参考的信令进行的回声管理的相关的支持,针对在具有低重叠窗口以及50%重叠窗口的MDCT域中进行量化和编码的支持,和/或针对在MDCT域中沿频率轴线进行滤波以进行量化噪声的时间整形的支持,例如TNS(时间噪声整形)。因此,在各个示例中,重叠窗口是对称的或不对称的。

[0058] 在一些示例中,所公开的互换格式可以提供以下各项:改进的联合声道编码,以利用灵活渲染用例中的信号之间的增加的相关性;通过跨声道元素共享的比例因子提高的编码效率;在MDCT域中包括高频重建和噪声添加技术;对传统编码结构的各种改进,以允许提高效率并使其适合于手头的用例。

[0059] 在以广播模式控制单独回放设备的一些示例中,可以使用可跳过的块和元数据。在广播模式设置中,每个无线设备将需要为对应于特定设备的特定驱动器/声道播放完整

音频的相关部分。这意味着设备需要知道流的哪些部分与该设备相关,并且从正在广播的完整音频流中提取那些部分。为了实现低复杂度解码操作,优选的是,流的构建方式使得特定设备的解码器可以高效地跳过与解码不相关的元素,到达与设备和设备上的给定驱动器相关的元素。

[0060] 然而,应当注意的是,仍然允许去往不同设备的信号(例如,在最简单化的场景中,立体声呈现的左声道和右声道)之间的联合编码可能存在益处(从压缩角度来看)。

[0061] 因此,该格式可以:实现比特流中的"可跳过的块"以使得能够对仅与特定设备相关的部分进行高效解码,包括使得能够将一个或多个可跳过的块灵活映射到特定设备的元数据,同时保留在对应于不同扬声器/设备的信号之间应用联合编码技术的能力。

[0062] 在图4中,示出了任意设置的示例。存在三个连接的无线设备31、32、33,其中两个是单声道扬声器32、33(意味着一个音频声道),并且一个扬声器31是具有三个不同驱动器并且由此对三个不同信号进行操作的更高级的扬声器31。在该示例中,第一扬声器31对3个单独信号进行操作,而第二扬声器32和第三扬声器33对立体声表示(例如,左侧和右侧)进行操作。因此,对立体声表示的信号进行联合编码可能是有益的。

[0063] 在给定以上场景的情况下,该格式指定包含可跳过的块的比特流,使得第一设备可以仅提取流的相关部分来对该扬声器的信号进行解码,而对于立体声对,可以通过构建信号来权衡解码器复杂度与效率,其中,每个扬声器需要对两个信号进行解码以输出单个信号,好处是可以执行联合编码。

[0064] 该格式指定元数据格式以实现通用且灵活的表示,该表示将多个可跳过的块中的特定块映射到一个或多个设备。这在图5中图示,其中,映射可以表示为将每个设备31、32、33与一个或多个比特流元素相关联的矩阵,使得给定设备的解码器将知道要输出和解码什么比特流元素。

[0065] 例如,在图5的示例中,第一块或跳过块Blk1包括3个单声道元素(设备1 31的每个驱动器的一个),而第二块或跳过块Blk2包含声道对元素,其可以包括将由设备2 32和3 33输出的信号的联合编码版本。设备1 31提取映射元数据并确定其所需要的信号在跳过块1Blk1中。因此,其提取跳过块1Blk1,对其中的三个单声道元素进行解码,并将这三个单声道元素分别提供给驱动器1、2a和2b。

[0066] 此外,设备1 31忽略跳过块2Blk2。类似地,设备2 32提取映射元数据并确定其所需要的信号在跳过块2Blk2中。因此,设备2 32跳过跳过块1Blk1并提取跳过块2Blk2。设备2 32对声道对元素进行解码,并将CPE的左声道输出提供给其驱动器。类似地,设备3 33提取映射元数据并确定其所需要的信号在跳过块2Blk2中,因此设备3 33跳过跳过块1Blk1并且还提取跳过块2Blk2。设备3 33对声道对元素进行解码,并将CPE的右声道输出提供给其驱动器。

[0067] 在一些示例中,设备2 32可以确定其仅需要来自跳过块2Blk2的信号的子集。在这样的示例中,在可能的情况下,设备2 32可以仅执行对跳过块2Blk2中的信号完全解码所需的操作的子集。具体地,在图5的示例中,设备2 32可以仅执行提取CPE的左声道所需的那些处理操作,从而使得能够降低计算复杂度。类似地,设备3 33可以仅执行提取CPE的右声道所需的那些处理操作。

[0068] 例如,在图5的情况下,可能存在使用联合声道编码对CPE进行编码的时间,以及使

用独立声道编码对其进行编码的其他时间。在使用独立声道编码对CPE进行编码的时间期间,设备2 32可以仅提取CPE的第一(例如,左)声道,而设备3 33可以仅提取CPE的第二(例如,右)声道。

[0069] 在另一个示例中,可以使用联合声道编码对CPE的声道进行编码,在这种情况下,设备2 32和3 33必须提取CPE的两个中间声道。然而,通过仅执行从中间声道提取左声道所需的那些操作,设备2 32可以仍然能够以降低的计算复杂度进行操作。类似地,通过仅执行从中间的经解码声道提取右声道所需的那些操作,设备3 33可以仍然能够以降低的计算复杂度进行操作。

[0070] 取决于联合声道编码的具体情况,可能会进行其他优化。可以在系统初始化或设置阶段期间定义不同设备/解码器中的每一个的身份。这样的设置是常见的,并且通常涉及测量房间的音响效果、扬声器到最佳听音位置的距离等。

[0071] 灵活渲染(预编码和后解码)的分布以及延迟的应用等后解码器

[0072] 对于在将相关信号发送到特定设备/扬声器之前在集线器/TV中应用灵活渲染的用例,渲染可以创建从编码角度来看更难编码的信号,例如当考虑对这种信号的联合编码时。一个原因是灵活渲染可以对不同的设备应用不同的延迟、均衡和/或增益调整(例如,取决于扬声器相对于其他扬声器和收听者的放置)。还可以在初始设置时预设信息(例如,增益和延迟),并且灵活地仅渲染均衡。在其他示例中,预设信息和灵活渲染信息的其他变体也是有可能的。注意,在本文档中,术语“增益”应被解释为意指任何水平调整(例如,衰减、放大或通过),而不是仅限于某些水平调整(例如,放大)。

[0073] 在图6的示例中,右声道33和左声道32扬声器被给予不同的时延,这反映了相对于收听者的不同放置(例如,由于扬声器32、33与收听者的距离可能不相等,因此可以对从扬声器32、33输出的信号施加不同的时延,使得来自不同扬声器32、33的相干声音同时到达收听者)。将这种不同时延引入到旨在用于在不同扬声器31、32、33上回放的相干信号使得这种信号的联合编码具有挑战性。

[0074] 为了解决这种挑战,可以在对信号进行解码之后将灵活渲染过程的各方面参数化并应用于端点设备中。

[0075] 在图7的示例中,每个设备的延迟和增益值被参数化,并且被包括在发送到相应扬声器31、32、33的经编码信号中。相应信号可以由相应设备31、32、33解码,该相应设备然后可以将参数化增益和延迟值引入到相应经解码信号。

[0076] 在示例(如图7中所示的示例)中,在用于不同设备31、32、33的经编码信号在可分离块(例如,可跳过的块)中发送的情况下,参数(例如,延迟和增益)也可以在可分离块中发送,使得设备31、32、33可以仅提取该设备31、32、33所需的参数的子集并且忽略(并跳过)该设备31、32、33不需要的那些参数。在这种情况下,可以向每个设备提供指示哪些参数被包括在哪些块中的映射元数据。

[0077] 还注意,尽管图7仅指示延迟和增益参数,但也可以包括其他参数,如均衡参数。例如,均衡参数可以包括要应用于不同频率区域的多个增益、对要由回放设备31、32、33应用的预定均衡曲线的指示、一组或多组无限脉冲响应(IIR)或有限脉冲响应(FIR)滤波器系数、一组双二阶滤波器系数、指定参数均衡器特性的参数、以及本领域技术人员已知的用于指定均衡的其他参数。

[0078] 此外,灵活渲染方面的参数化不需要是静态的,并且可以是动态的(例如,在收听者在音频节目的回放期间移动的情况下)。因此,允许参数动态地改变可能是优选的。在音频节目期间一个或多个参数改变的情况下,设备可以在先前的延迟和/或增益参数与更新的延迟和/或增益参数之间进行插值以提供平滑过渡。这在系统动态地跟踪收听者的位置并且对应地更新用于动态渲染的最佳听音位置的情况下可能特别有用。

[0079] 如还已经注意到的并且随后将描述的,当应用灵活渲染时,声道之间的相关性水平可能增加,并且这可以通过更灵活的联合编码来利用。

[0080] 回声参考编码和信令

[0081] 对于如图8中所概述的用例,在存在多个设备/扬声器30一起操作以广播模式(在图8的左侧图示)或者以单播多点模式(在图8的右侧图示)接收信号并且同时设备30上存在麦克风40以实现“收听”能力的情况下,出现了对回声管理的需求。

[0082] 当用多个扬声器/设备执行回声管理时,使用不只是来自本地扬声器设备的回声参考可能是有益的。作为示例,一个设备可以靠近另一个设备放置,并且因此来自附近设备的信号将影响具有有源麦克风的设备的回声管理。在广播模式用例中,每个设备接收用于所有设备的信号。当设备具有用于其他设备的信号时,使用那些信号作为回声参考可能是有益的。为此,有必要向特定设备发信号通知哪些信号可以用作哪些其他设备的回声参考。

[0083] 在一个示例中,这可以通过提供元数据来完成,该元数据不仅映射要由特定扬声器/设备播放(来自整个集合)的声道/信号,而且映射要用作特定扬声器/设备的回声参考(来自整个集合)的声道/信号。这种元数据或信令可以是动态的,从而使得对优选回声参考的指示能够随时间变化。

[0084] 对于每个设备/扬声器仅接收其要播放的特定信号的用例,为了提供适当的回声参考,可能有必要向每个设备/扬声器发射附加信号(例如,回声参考信号)。此外,为此,有必要提供设备特定的信令,使得每个设备可以选择用于播放的适当信号和用于回声管理的适当信号。

[0085] 由于用于回声管理的信号仅被设备用于执行回声管理,而不会由设备向收听者播放,因此回声参考信号可能以与旨在用于由设备回放的信号不同的方式进行编码或表示。具体地,因为可以利用以比通常用于向收听者播放的速率更低的速率编码的信号来实现成功的回声管理,所以可能根本不适合向收听者播放但是捕获音频信号的必要特征的附加压缩工具(如信号的参数表示)可以以显著降低的传输成本提供良好的回声管理。

[0086] 各种语法元素

[0087] 在一些示例中,块用于优化音频的传输。在所描述的格式中,每一帧可以被分割成多个块,如以上关于可跳过的块所描述的。块可以用以下各项来标识:其所属的帧号、可以用于将来自不同帧的具有相同ID的连续块与块流相关联的块ID、以及重传优先级。以上内容的一个示例是具有多个帧N-2、N-1和N的流,帧N包括标识为ID1、ID2、ID3的多个块,该多个块在图13中图示。图14中图示了该流的示例,但是采用比特流格式。在一些示例中,对于沉浸式音频节目的音频信号,块的块ID可以指示整个沉浸式音频节目的哪组信号由该块携带。

[0088] 所描述的格式的用例是通过如Wifi等无线网络以低时延可靠地传输音频。例如,Wifi使用基于包的网络协议。包大小通常是有限的。IP网络中典型的最大包大小是1500字

节。流的基于块的架构允许在组装包以进行传输时具有灵活性。例如,具有较小帧的包可以用来自其他帧的重传块来填充。在对大帧进行分包之前,可以在块边界上分割大帧以减少网络协议层上的包之间的依赖性。

[0089] 图9示出了帧、块和包之间的关系。帧携带音频数据,优选地所有音频数据,该音频数据表示音频信号的具有开始时间、结束时间和作为结束时间与开始时间之差的持续时间的连续片段。连续片段可以包括根据ISO/IEC 14496-3的子部分4的第4.5.2.1.1节的时间段。第4.5.2.1.1节描述了raw_data_block()的内容。帧还可以携带例如以较低数据速率编码的该片段的冗余表示。在编码之后,该帧可以被分割成块。该块可以被组合成包以通过基于包的网络进行传输。来自不同帧的块可以被组合在单个包中,和/或可以被乱序发送。

[0090] 在示例中,块用于寻址单独的设备。数据包由单独的设备或相关的设备组接收。可跳过的块的概念可以用于寻址单独的设备或相关的设备组。即使网络在向不同设备发送包时以广播模式操作,音频的处理(例如,解码、渲染等)也可以减少为寻址到该设备的块。所有其他块,即使在同一包内被接收,也可以简单地跳过。在一些示例中,可以基于块的解码或呈现时间以正确的顺序来引入块。如果还已经接收到具有较高优先级的同一块,则可以移除具有较低优先级的重传块。然后可以将块流馈送到解码器。

[0091] 在一些示例中,流和设备的配置是在带外发送的。编解码器允许建立连接,其中,音频流以相对较高的速率传输,但是具有低时延。这种连接的配置可以在这种连接的持续时间内保持稳定。在这种情况下,不是使配置成为音频流的一部分,而是可以在带外传输配置。对于这种带外传输,甚至可以使用不同的网络或网络协议。例如,音频流可以使用用户数据报配置文件(UDP)进行低时延传输,而配置可以使用传输控制协议(TCP)来确保配置的可靠传输。

[0092] 在MPEG-4音频的背景下启用编解码器

[0093] 本技术的一个具体应用是在MPEG-4音频内使用。在MPEG-4音频中,为不同的编解码器技术定义了不同的AOT(音频对象类型)。对于本文档中描述的格式,可以定义新的AOT,这允许针对该格式的特定信令和数据。此外,MPEG-4中解码器的配置是在DecoderSpecificInfo()有效载荷中完成的,该有效载荷进而携带AudioSpecificConfig()有效载荷。在后者中,定义了特定格式不可知的某些通用信令,如采样速率和声道配置,以及特定AOT的特定信息。对于整个流由单个设备解码的传统格式,这可能是有意义的。然而,在广播模式下,其中,单个流被传输到若干个解码器,每个相应解码器仅对流的部分进行解码,预先声道配置信令(作为设置设备的输出功能的手段)可能是次优的。

[0094] 图10图示了传统的MPEG-4高级结构(呈黑色),其中修改呈灰色。为了支持广播用例,定义了codecSepcificConfig()(其中,“编解码器”可以是通用占位符名称),其中,针对特定用例重新定义信令,使得可以将特定声道元素映射到特定设备,并且可以包括其他相关静态参数。值为“0”的MPEG-4元素声道配置被定义为在codecSpecificConfig中定义的声道配置。因此,该值可以用于使得能够修改编解码器特定配置内部的声道配置的信令。

[0095] 进一步地,在MPEG-4的精神中,假定codecSpecificConfig()是可解码的,则为手头的特定解码器指定原始有效载荷。然而,该格式确保动态元数据是原始有效载荷的一部分,并且确保长度信息对于所有原始有效载荷都是可用的,使得解码器可以轻松跳过与特定设备无关的元素。

[0096] 在图11中,如在MPEG-4中定义的示例raw_data_block的一部分在右侧给出。原始数据块包含给定顺序的声道元素(单声道元素(SCE)或声道对元素(CPE))。然而,在传统的MPEG-4音频语法中,希望跳过这些声道元素的可能与手头的输出设备无关的部分的解码器将必须解析(并且在某种程度上还解码)所有的声道元素以便能够提取相关部分。在图11的左侧图示的新raw_data_block中,内容将由可跳过的块构成,使得解码器可以跳过不相关的部分并且仅对被元数据指示为与手头的设备相关的声道元素进行解码。在示例中,可跳过的块包括raw_data_block和相关信息。

[0097] 还可以使用块进行重传。每一帧可以被分割成块,如以上在可跳过的块章节中所描述的。块可以用以下各项来标识:其所属的帧号、可以用于将来自不同帧的具有相同ID的连续块与块流相关联的块ID、以及重传优先级,如图15和图16中所图示的。例如,高重传优先级(例如,在图15和图16中以优先级0指示)表明块在接收器处应当优先于具有相同块ID和帧计数器但重传优先级较低的另一个块(例如,在图15和图16中以优先级1指示)。注意,如在图15和图16的示例中,优先级可以随着优先级指数的增加而降低(例如,优先级1可以是相较于优先级0更低的优先级),而在其他示例中,优先级可以随着优先级指数的增加而增加(例如,优先级0是相较于优先级1更低的优先级)。还有其他示例对于技术人员来说将是明显的。

[0098] 语法可以支持音频元素的重传。对于重传,可以支持各种质量水平。因此,重传块可以携带‘优先级’标志以指示具有相同块ID的哪些块应该优先用于解码器,因为接收到的具有相同帧计数器和块ID的块是冗余的,并且因此对于解码器是互斥的,如图17和图18中所图示的。

[0099] 块的重传可以以降低的数据速率完成。这种降低的数据速率可以通过以下方式来实现:降低音频信号的信噪比、减少音频信号的带宽、减少音频信号的声道计数(例如,如美国专利11,289,103中所描述的,该美国专利通过引用整体并入本文)或其任何组合。为了使解码器挑选提供最佳可能质量的音频块,提供最高质量信号的块可以具有最高解码优先级,具有次高质量信号的块可以具有次高优先级,依此类推。

[0100] 这些块也可以以相同的质量水平重传。在这种情况下,优先级可以反映重传块的时延。

[0101] 用于提高核心编码效率的工具

[0102] 进一步示例可以允许在声道元素上共享MDCT量化比例因子。在一些情况下,可以在某些声道上共享MDCT比例因子以降低相关联的边比特率。此外,比例因子共享可以扩展到不同的声道元素,例如对于7.1.4输入。共享的比例因子的使用可以由允许3个活动配置的2个信令比特来指示。一种可能的配置是共享左水平声道、左上声道、右水平声道和右上声道中的比例因子。特定语法可以与可跳过的块概念对准,以确保比例因子仅在跳过块内共享。

[0103] 还可以对多于两个声道进行联合编码。在图4、图5、图6和图7中图示的示例中,块Blk 1携带三个SCE,在这种场景中考虑的智能扬声器中的每个驱动器有一个。在这种情况下,可能有益的是,不仅能够对两个声道进行联合编码(如由CPE提供的,其可以被扩展为包括立体声预测,例如,如ISO/IEC 23003-3中所描述的基于MDCT的复杂预测立体声工具,称为MPEG-D USAC),而且能够对多于两个声道进行联合编码,例如,如在ETSI TS103 190中介

绍的SAP(立体声音频处理)工具中所描述的。

[0104] 类似地,应当注意的是,在灵活渲染的情况下,跨设备的信号之间可能存在大量的相关性。对于这种信号,构建覆盖多个设备并且允许使用以上概述的工具在多于两个声道上应用联合声道编码的跳过块可能是有益的。

[0105] 对于某些变换长度,可以使用的另一种联合声道编码工具是声道耦合,其中,针对中频和高频传输复合声道和比例因子信息。这可以为良好质量范围内的回放提供比特率降低。例如,对变换长度为256的帧使用声道耦合工具可能是有益的,该变换长度对应于低时延编码的约256个样本的帧长度。

[0106] 本公开还将允许对带宽受限信号进行高效编码。在发送馈送智能扬声器的不同驱动器的单独信号的场景中,这些信号中的一些信号可能是限带的,比如例如,具有低音扬声器、中音扬声器和高音扬声器的三向驱动器配置的情况。因此,期望对这种限带信号进行高效编码,这可以转化为专门调谐的心理声学模型和比特分配策略以及对语法的潜在修改,从而以提高的编码效率和/或降低的计算复杂度来处理这种场景(例如,使得能够对低音扬声器馈送使用限带IMDCT)。

[0107] 在一些示例中,可以在MDCT域中组合高频重建和噪声添加。现代音频编解码器通常被设计成支持参数编码技术,并且类似地,在此可以设想在MDCT域中包括例如高频重建方法以保持低时延,并且在MDCT域中执行噪声添加方案以管理重建的高频带中的音调噪声比。这种参数编码技术对于较低的操作点可能尤其有用,并且特别是在重传作为FEC(前向纠错)方案的一部分发生的场景中,在该场景中,通常发射主信号,并且然后以延迟的方式以较低的比特率重传同一信号。

[0108] 这些示例将进一步允许降低编码器中的峰值复杂度。如果恒定比特率传输信道的缓冲器模型限制不到位,则这将是适用的。在具有缓冲器模式的恒定比特率下,在量化和计数之后,要用于待编码帧的最终比特数可能会超过所允许比特的限制。需要进行至少一个新的、较粗略的量化和比特计数步骤以满足比特要求。如果这种限制更宽松,则编码器可以保持第一量化结果并且将以略微更高的瞬时比特率来完成。通过使用相同的比特存储库控制机制以及根据遵循缓冲器要求的编码器而更新的虚拟缓冲器充满度,仍然可以实现与具有缓冲器模型编码器的恒定比特率的编码器行为类似的编码器行为。这将节省附加的量化和比特计数步骤,并且与具有总体比特率略微增加的缺点的恒定比特率的情况相比,所得音频质量将相同或更好。

[0109] 返回声道概念

[0110] 对于播放和收听用例,智能扬声器30可以具有一个或多个麦克风40(例如,麦克风或麦克风阵列),该一个或多个麦克风被配置为捕获单声道或空间声场(例如,采用单声道、立体声、A格式、B格式或任何其他各向同性或各向异性声道格式),并且需要编解码器能够以低时延的方式对这种格式进行高效编码,使得与用于返回声道的编解码器相同的编解码器被用于向智能扬声器设备30进行的广播/传输,例如图12中所图示的。

[0111] 还应当注意的是,虽然唤醒词检测是在智能扬声器设备上执行的,但是语音识别通常是在云中完成的,其中唤醒词检测触发合适的记录语音片段。

[0112] 在这种背景下,通过在例如编解码器内的中间格式/表示上创建语音分析系统来总体节省系统复杂度可能很有趣。对于最简单的用例(其中,没有人与人的对话正在进行

时),可以定义特定于语音识别任务的合适表示,因为另一人不会听到该合适表示。这可以是频带能量、梅尔频率倒谱系数(MFCC)等,或MDCT编码频谱的低比特率版本。

[0113] 在人与人的对话正在进行并且唤醒词和所需的语音识别交织在一起的用例中,人们将希望能够为此提取流的相关部分。这种提取可能需要与主语音信号并行发送的简单附加数据的分层编码结构。我们还设想了从现有的经解码MDCT频谱到相关表示的定义转码,并且这样做简化了结构。本质上,解码器将解码并输出人类可收听音频,直到表明解码器还应当解码(并行地)语音识别表示。该表示可以从分层流上"剥离",简单地作为对相同完整数据的替代性解码,或者简单地作为对流中的附加表示的解码和输出。在该用例中,信令是指示接收侧上的解码器应当输出语音识别相关表示的使能件。

[0114] 列举的示例

[0115] 在下文中,七组列举的示例性示例(EEE-A、EEE-B、EEE-C、EEE-D、EEE-E、EEE-F和EEE-G)描述了本文所公开的示例的各方面,这些示例不是权利要求。

[0116] EEE-A1.一种用于对音频信号进行解码的方法,所述方法包括:

[0117] 接收包括至少一个帧的比特流,其中,所述至少一个帧中的每一帧包括多个块;

[0118] 基于输出设备的设备信息,从信令数据确定用于标识所述多个块中的一个或多个块的解码时要跳过的部分的信息;以及

[0119] 在跳过所述一个或多个块的所标识部分的同时对所述比特流进行解码。

[0120] EEE-A2.如EEE-A1所述的方法,其中,所述用于标识所述多个块中的一个或多个块的解码时要跳过的部分的信息包括将多个输出设备中的每个输出设备与一个或多个比特流元素相关联的矩阵。

[0121] EEE-A3.如EEE-A2所述的方法,其中,所述一个或多个比特流元素是对应的相关联输出设备对所述比特流进行解码所需的。

[0122] EEE-A4.如EEE-A1至EEE-A3中任一项所述的方法,其中,所述输出设备可以包括无线设备、移动设备、平板计算机、单声道扬声器和/或多声道扬声器中的至少一者。

[0123] EEE-A5.如EEE-A1至EEE-A4中任一项所述的方法,其中,所述所标识部分包括至少一个块。

[0124] EEE-A6.如EEE-A1至EEE-A5中任一项所述的方法,其中,所述输出设备是第一输出设备,所述方法进一步包括将所述比特流的一个或多个信号之间的联合编码技术应用于第二输出设备和第三输出设备。

[0125] EEE-A7.如EEE-A1至EEE-A6中任一项所述的方法,其中,每个输出设备和/或解码器的身份是在系统初始化阶段期间定义的。

[0126] EEE-A8.如EEE-A1至EEE-A7中任一项所述的方法,其中,所述信令数据是从所述比特流的元数据确定的。

[0127] EEE-A9.一种装置,被配置为执行如EEE-A1至EEE-A8中任一项所述的方法。

[0128] EEE-A10.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如EEE-A1至EEE-A8中任一项所述的方法。

[0129] EEE-B1.一种用于从包括多个音频信号的音频节目生成经编码比特流的方法,所述方法包括:

[0130] 针对所述多个音频信号中的每个音频信号接收指示与相应音频信号相关联的回

放设备的信息；

[0131] 针对每个回放设备接收指示与相应回放设备相关联的延迟、增益和均衡曲线中的至少一者的信息；

[0132] 从所述多个音频信号中确定一组两个或更多个相关音频信号；

[0133] 对所述组中的所述两个或更多个相关音频信号应用一个或多个联合编码工具以获得联合编码的音频信号；

[0134] 将所述联合编码的音频信号、与所述联合编码的音频信号相关联的所述回放设备的指示、以及同与所述联合编码的音频信号相关联的所述相应回放设备相关联的所述延迟和所述增益的指示组合到经编码比特流的独立块中。

[0135] EEE-B2.如EEE-B1所述的方法，其中，与所述相应回放设备相关联的所述延迟、所述增益和/或所述均衡曲线取决于所述相应回放设备相对于收听者的位置的位置。

[0136] EEE-B3.如EEE-B1或EEE-B2所述的方法，其中，与所述相应回放设备相关联的所述延迟、所述增益和/或所述均衡曲线取决于所述相应回放设备相对于其他回放设备的位置的位置。

[0137] EEE-B4.如EEE-B1至EEE-B3中任一项所述的方法，其中，所述延迟、所述增益和/或所述均衡曲线是动态可变的。

[0138] EEE-B5.如EEE-B4所述的方法，其中，响应于所述收听者的位置的变化对所述延迟、所述增益和/或所述均衡曲线进行调整。

[0139] EEE-B6.如EEE-B4或EEE-B5所述的方法，其中，响应于所述回放设备的位置的变化对所述延迟、所述增益和/或所述均衡曲线进行调整。

[0140] EEE-B7.如EEE-B4至EEE-B6中任一项所述的方法，其中，响应于所述其他回放设备中的一个或多个回放设备的位置的变化对所述延迟、所述增益和/或所述均衡曲线进行调整。

[0141] EEE-B8.如EEE-B1至EEE-B7中任一项所述的方法，进一步包括从所述多个音频信号中确定不是所述一组两个或更多个相关音频信号的一部分的音频信号。

[0142] EEE-B9.如EEE-B8所述的方法，进一步包括针对不是所述一组两个或更多个相关音频信号的一部分的所述音频信号，应用同与所述音频信号相关联的所述回放设备相关联的所述延迟、所述增益和/或所述均衡曲线。

[0143] EEE-B10.如EEE-B9所述的方法，进一步包括对不是所述一组两个或更多个相关音频信号的一部分的所述音频信号进行独立编码，以及将独立编码的音频信号和与所述独立编码的音频信号相关联的所述回放设备的指示组合到所述经编码比特流的单独的可独立解码的子集中。

[0144] EEE-B11.如EEE-B8所述的方法，进一步包括对不是所述一组两个或更多个相关音频信号的一部分的所述音频信号进行独立编码，以及将所述独立编码的音频信号、与所述独立编码的音频信号相关联的所述回放设备的指示、以及同与所述独立编码的音频信号相关联的所述回放设备相关联的所述延迟、所述增益和/或所述均衡曲线的指示组合到所述经编码比特流的单独的可独立解码的子集中。

[0145] EEE-B12.一种用于从经编码比特流的帧中解码与回放设备相关联的一个或多个音频信号的方法，其中，所述帧包括一个或多个独立的经编码数据块，所述方法包括：

[0146] 从所述经编码比特流中识别同与所述回放设备相关联的所述一个或多个音频信号相对应的独立的经编码数据块;

[0147] 从所述经编码比特流中提取所识别的独立的经编码数据块;

[0148] 确定所提取的独立的经编码数据块包括两个或更多个联合编码的音频信号;

[0149] 对所述两个或更多个联合编码的音频信号应用一个或多个联合解码工具以获得与所述回放设备相关联的所述一个或多个音频信号;

[0150] 从所提取的独立的经编码数据块中确定与所述回放设备相关联的延迟、增益和均衡曲线中的至少一者;

[0151] 将与所述回放设备相关联的所述延迟、所述增益和/或所述均衡曲线应用于与所述回放设备相关联的所述一个或多个音频信号。

[0152] EEE-B13.如EEE-B12所述的方法,其中,所确定的与所述回放设备相关联的延迟、增益和/或均衡曲线取决于所述回放设备相对于收听者的位置的位置。

[0153] EEE-B14.如EEE-B12或EEE-B13所述的方法,其中,所确定的与所述回放设备相关联的延迟、增益和/或均衡曲线取决于所述回放设备相对于其他回放设备的位置。

[0154] EEE-B15.如EEE-B12至EEE-B14中任一项所述的方法,其中,所确定的所述回放设备的延迟、增益和/或均衡曲线是动态可变的。

[0155] EEE-B16.如EEE-B15所述的方法,其中,当所确定的与所述回放设备相关联的延迟、增益和/或均衡曲线不同于先前确定的与所述回放设备相关联的延迟、增益和/或均衡曲线时,所述方法进一步包括在所述先前确定的与所述回放设备相关联的延迟、增益和/或均衡曲线与所确定的与所述回放设备相关联的延迟、增益和/或均衡曲线之间进行插值。

[0156] EEE-B17.如EEE-B16所述的方法,其中,由于收听者的位置的变化,所确定的延迟、增益和/或均衡曲线与所述先前确定的延迟、增益和/或均衡曲线不同。

[0157] EEE-B18.如EEE-B16或EEE-B17所述的方法,其中,由于所述回放设备的位置的变化,所确定的延迟、增益和/或均衡曲线与所述先前确定的延迟、增益和/或均衡曲线不同。

[0158] EEE-B19.如EEE-B16至EEE-B18中任一项所述的方法,其中,由于所述其他回放设备中的一个或多个回放设备的位置的变化,所确定的延迟、增益和/或均衡曲线与所述先前确定的延迟、增益或均衡曲线不同。

[0159] EEE-B20.如EEE-B12至EEE-B19中任一项所述的方法,其中,所述经编码比特流的帧包括两个或更多个独立的经编码数据块,所述方法进一步包括:

[0160] 确定所述独立块中的一个或多个独立块包含不与所述回放设备相关联的音频信号;以及

[0161] 忽略包含不与所述回放设备相关联的音频信号的所述一个或多个独立块。

[0162] EEE-B21.如EEE-B12至EEE-B20中任一项所述的方法,其中,应用一个或多个联合解码工具包括识别所述联合编码的音频信号的与所述回放设备相关联的子集,以及仅重建所述联合编码的音频信号的该子集以获得与所述回放设备相关联的所述一个或多个音频信号。

[0163] EEE-B22.如EEE-B12至EEE-B20中任一项所述的方法,其中,应用一个或多个联合解码工具包括重建所述联合编码的音频信号中的每个联合编码的音频信号,标识与所述回放设备相关联的重建的联合编码的音频信号的子集,以及从与所述回放设备相关联的所述

重建的联合编码的音频信号的所述子集中获得与所述回放设备相关联的所述一个或多个音频信号。

[0164] EEE-B23.一种装置,被配置为执行如EEE-B1至EEE-B22中任一项所述的方法。

[0165] EEE-B24.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如EEE-B1至EEE-B22中任一项所述的方法。

[0166] EEE-C1.一种用于生成包括多个音频信号的音频节目的经编码比特流的帧的方法,其中,所述帧包括两个或更多个独立的经编码数据块,所述方法包括:

[0167] 针对所述多个音频信号中的一个或多个音频信号接收指示与所述一个或多个音频信号相关联的回放设备的信息;

[0168] 针对所指示的回放设备接收指示一个或多个附加相关联回放设备的信息;

[0169] 接收与所指示的一个或多个附加相关联回放设备相关联的一个或多个音频信号;

[0170] 对与所述回放设备相关联的所述一个或多个音频信号进行编码;

[0171] 对与所指示的一个或多个附加相关联回放设备相关联的所述一个或多个音频信号进行编码;

[0172] 将与所述回放设备相关联的所述一个或多个经编码音频信号和指示所述一个或多个附加相关联回放设备的信令信息组合到第一独立块中;

[0173] 将与所述一个或多个附加相关联回放设备相关联的所述一个或多个经编码音频信号组合到一个或多个附加独立块中;以及

[0174] 将所述第一独立块和所述一个或多个附加独立块组合到所述经编码比特流的帧中。

[0175] EEE-C2.如EEE-C1所述的方法,其中,所述多个音频信号包括不与所述回放设备或所述一个或多个附加相关联回放设备相关联的一组或多组音频信号,所述方法进一步包括:

[0176] 将不与所述回放设备或所述一个或多个附加相关联回放设备相关联的所述一组或多组音频信号中的每组音频信号编码到相应独立块中;以及

[0177] 将所述一个或多个组中的每个组的所述相应独立块组合到所述经编码比特流的所述帧中。

[0178] EEE-C3.如EEE-C1或EEE-C2所述的方法,其中,与所指示的一个或多个附加相关联回放设备相关联的所述一个或多个音频信号具体地旨在用作回声参考以供所述回放设备执行回声管理。

[0179] EEE-C4.如EEE-C3所述的方法,其中,旨在用作回声参考的所述一个或多个音频信号是使用比与所述回放设备相关联的所述一个或多个音频信号更少的数据来传输的。

[0180] EEE-C5.如EEE-C3或EEE-C4所述的方法,其中,旨在用作回声参考的所述一个或多个音频信号是使用参数编码工具来编码的。

[0181] EEE-C6.如EEE-C1或EEE-C2所述的方法,其中,与所述一个或多个其他回放设备相关联的所述一个或多个音频信号适合于从所述一个或多个其他回放设备回放。

[0182] EEE-C7.一种用于从经编码比特流的帧中解码与回放设备相关联的一个或多个音频信号的方法,其中,所述帧包括两个或更多个独立的经编码数据块,其中,所述回放设备包括一个或多个麦克风,所述方法包括:

[0183] 从所述经编码比特流中识别同与所述回放设备相关联的所述一个或多个音频信号相对应的独立的经编码数据块;

[0184] 从所述经编码比特流中提取所识别的独立的经编码数据块;

[0185] 从所识别的独立的经编码数据块中提取与所述回放设备相关联的所述一个或多个音频信号;

[0186] 从所述经编码比特流中识别同与一个或多个其他回放设备相关联的一个或多个音频信号相对应的一个或多个其他独立的经编码数据块;

[0187] 从所述一个或多个其他独立的经编码数据块中提取与所述一个或多个其他回放设备相关联的所述一个或多个音频信号;

[0188] 使用所述回放设备的所述一个或多个麦克风来捕获一个或多个音频信号;以及

[0189] 使用与所述一个或多个其他回放设备相关联的一个或多个所提取音频信号作为回声参考,以供所述回放设备响应于一个或多个捕获的音频信号而执行回声管理。

[0190] EEE-C8.如EEE-C7所述的方法,所述方法进一步包括:

[0191] 确定所述经编码比特流包括一个或多个附加的独立的经编码数据块;以及

[0192] 忽略所述一个或多个附加的独立的经编码数据块。

[0193] EEE-C9.如EEE-C8所述的方法,其中,忽略所述一个或多个附加的独立的经编码数据块包括跳过所述一个或多个附加的独立的经编码数据块而不提取所述附加的一个或多个独立的经编码数据块。

[0194] EEE-C10.如EEE-C7至EEE-C9中任一项所述的方法,其中,与所述一个或多个其他回放设备相关联的所述一个或多个音频信号具体地旨在用作回声参考以供所述回放设备执行回声管理。

[0195] EEE-C11.如EEE-C10所述的方法,其中,具体地旨在用作回声参考的所述一个或多个音频信号是使用比与所述回放设备相关联的所述一个或多个音频信号更少的数据来传输的。

[0196] EEE-C12.如EEE-C10或EEE-C11所述的方法,其中,具体地旨在用作回声参考的所述一个或多个音频信号是从所述一个或多个音频信号的参数表示来重建的。

[0197] EEE-C13.如EEE-C7至EEE-C9中任一项所述的方法,其中,与所述一个或多个其他回放设备相关联的所述一个或多个音频信号适合于从所述一个或多个其他回放设备回放。

[0198] EEE-C14.如EEE-C7所述的方法,其中,经编码信号包括指示所述一个或多个其他回放设备用作所述回放设备的回声参考的信令信息。

[0199] EEE-C15.如EEE-C14所述的方法,其中,由当前帧的信令信息指示的所述一个或多个其他回放设备不同于用作前一帧的回声参考的所述一个或多个其他回放设备。

[0200] EEE-C16.一种装置,被配置为执行如EEE-C1至EEE-C15中任一项所述的方法。

[0201] EEE-C17.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如EEE-C1至EEE-C15中任一项所述的方法。

[0202] EEE-D1.一种用于传输音频信号的方法,所述方法包括:

[0203] 生成包括比特流的部分的数据包,其中,所述比特流包括多个帧,其中,所述多个帧中的每一帧包括多个块,其中,所述生成包括:

[0204] 组装具有所述多个块中的一个或多个块的数据包,其中,来自不同帧的块被组合

成单个包和/或被乱序传输;以及

[0205] 经由基于包的网络来传输所述数据包。

[0206] EEE-D2.如EEE-D1所述的方法,其中,所述多个块中的每个块包括标识信息。

[0207] EEE-D3.如EEE-D2所述的方法,其中,所述标识信息包括以下各项中的至少一项:块ID、与所述块相关联的对应帧号、和/或重传优先级。

[0208] EEE-D4.如EEE-D1至EEE-D3中任一项所述的方法,其中,所述多个帧中的每一帧携带表示音频信号的具有开始时间、结束时间和持续时间的连续片段的所有音频数据。

[0209] EEE-D5.一种用于对音频信号进行解码的方法,所述方法包括:

[0210] 接收包括比特流的部分的数据包,其中,所述比特流包括多个帧,其中,所述多个帧中的每一帧包括多个块;

[0211] 确定所述多个块中寻址到设备的一组块;以及

[0212] 对寻址到所述设备的所述一组块进行解码并且跳过对所述多个块中未寻址到所述设备的块进行解码。

[0213] EEE-D6.一种用于传输音频流的方法,所述方法包括:

[0214] 传输所述音频流,其中,所述音频流包括多个帧,其中,所述多个帧中的每一帧包括多个块,其中,所述传输包括在带外传输所述音频流的配置信息。

[0215] EEE-D7.如EEE-D6所述的方法,其中,在带外传输所述音频流的配置信息包括:

[0216] 经由第一网络和/或第一网络协议传输所述音频流;以及

[0217] 经由第二网络和/或第二网络协议传输所述配置信息。

[0218] EEE-D8.如EEE-D7所述的方法,其中,所述第一网络协议是用户数据报协议(UDP),并且所述第二网络协议是传输控制协议(TCP)。

[0219] EEE-D9.一种用于对音频信号进行解码的方法,所述方法包括:

[0220] 接收比特流,所述比特流包括:

[0221] 与静态配置方面的信令相对应的信息;

[0222] 静态元数据;以及

[0223] 基于所述信息和/或所述静态元数据将一个或多个声道元素映射到一个或多个设备。

[0224] EEE-D10.如EEE-D9所述的方法,其中,所述比特流由被配置为对所述比特流进行解码的多个解码器接收,其中,所述多个解码器中的每个解码器被配置为对所述比特流的一部分进行解码。

[0225] EEE-D11.如EEE-D9或EEE-D10所述的方法,其中,所述比特流进一步包括动态元数据。

[0226] EEE-D12.如EEE-D9至EEE-D11中任一项所述的方法,其中,所述比特流包括多个块,其中,所述多个块中的每个块包括:

[0227] 使得能够在解码期间跳过所述块的一部分的信息,其中,设备不需要所述部分;以及

[0228] 动态元数据。

[0229] EEE-D13.一种用于重传音频信号的块的方法,所述方法包括:

[0230] 传输比特流的一个或多个块,其中,所述比特流包括多个块,其中,所述比特流的

所述一个或多个块中的每个块先前已经被传输;并且

[0231] 其中,所述一个或多个块中的每个块包括解码优先级指示符。

[0232] EEE-D14.如EEE-D13所述的方法,其中,所述解码优先级指示符向解码器指示用于对所述比特流的所述一个或多个块进行解码的优先级顺序。

[0233] EEE-D15.如EEE-D13或EEE-D14所述的方法,其中,所述一个或多个块中的每个块包括相同的块ID。

[0234] EEE-D16.如EEE-D13至EEE-D15中任一项所述的方法,其中,所述比特流的所述一个或多个块的所述传输是通过与先前传输相比降低数据速率来传输的。

[0235] EEE-D17.如EEE-D16所述的方法,其中,降低所述数据速率包括以下各项中的至少一项:降低所述音频信号的信噪比、减少所述音频信号的带宽、和/或减少所述音频信号的声道计数。

[0236] EEE-E1.一种用于生成包括多个音频信号的音频节目的经编码比特流的帧的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:

[0237] 针对所述多个音频信号中的每个音频信号接收指示与相应音频信号相关联的回放设备的信息;

[0238] 对与相应回放设备相关联的一个或多个音频信号进行编码以获得一个或多个经编码音频信号;

[0239] 将与所述相应回放设备相关联的所述一个或多个经编码音频信号组合到所述帧的第一独立块中;

[0240] 将所述多个音频信号中的一个或多个其他音频信号编码到一个或多个附加独立块中;以及

[0241] 将所述第一独立块和所述一个或多个附加独立块组合到所述经编码比特流的帧中。

[0242] EEE-E2.如EEE-E1所述的方法,其中,两个或更多个音频信号与所述回放设备相关联,并且所述两个或更多个音频信号中的每个音频信号是旨在由所述回放设备的相应驱动器回放的带限信号,并且其中,针对所述带限信号中的每个带限信号使用不同的编码技术。

[0243] EEE-E3.如EEE-E2所述的方法,其中,针对所述带限信号中的每个带限信号使用不同的心理声学模型和/或不同的比特分配技术。

[0244] EEE-E4.如EEE-E1至EEE-E3中任一项所述的方法,其中,所述经编码信号的瞬时帧速率是可变的,并且受缓冲器充满度模型的约束。

[0245] EEE-E5.如EEE-E1所述的方法,其中,对与所述相应回放设备相关联的一个或多个音频信号进行编码包括将与所述相应回放设备相关联的所述一个或多个音频信号以及与一个或多个附加回放设备相关联的一个或多个附加音频信号联合编码到所述帧的第一独立块中。

[0246] EEE-E6.如EEE-E5所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括跨两个或更多个音频信号共享一个或多个比例因子。

[0247] EEE-E7.如EEE-E6所述的方法,其中,所述两个或更多个音频信号是空间相关的。

[0248] EEE-E8.如EEE-E7所述的方法,其中,所述两个或更多个空间相关音频信号包括左水平声道、左上声道、右水平声道或右上声道。

[0249] EEE-E9.如EEE-E5所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括应用耦合工具,所述耦合工具包括:

[0250] 将两个或更多个音频信号组合成高于指定频率的复合信号;以及

[0251] 针对所述两个或更多个音频信号中的每个音频信号确定与所述复合信号的能量和每个相应信号的能量相关的比例因子。

[0252] EEE-E10.如EEE-E5所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合编码包括对多于两个信号应用联合编码工具。

[0253] EEE-E11.一种用于从经编码比特流的帧中解码与回放设备相关联的一个或多个音频信号的方法,其中,所述帧包括一个或多个独立的经编码数据块,所述方法包括:

[0254] 从所述经编码比特流中识别同与所述回放设备相关联的所述一个或多个音频信号相对应的独立的经编码数据块;

[0255] 从所述经编码比特流中提取所识别的独立的经编码数据块;

[0256] 从所述独立的经编码数据块中解码与所述回放设备相关联的所述一个或多个音频信号以获得一个或多个经解码音频信号;

[0257] 从所述经编码比特流中识别与一个或多个附加音频信号相对应的一个或多个附加的独立的经编码数据块;以及

[0258] 解码或跳过所述一个或多个附加的独立的经编码数据块。

[0259] EEE-E12.如EEE-E11所述的方法,其中,两个或更多个音频信号与所述回放设备相关联,并且所述两个或更多个音频信号中的每个音频信号是旨在由所述回放设备的相应驱动器回放的带限信号,并且其中,使用不同的解码技术对所述两个或更多个音频信号进行解码。

[0260] EEE-E13.如EEE-E12所述的方法,其中,使用不同的心理声学模型和/或不同的比特分配技术对所述带限信号中的每个带限信号进行编码。

[0261] EEE-E14.如EEE-E11或EEE-E13所述的方法,其中,所述经编码比特流的瞬时帧速率是可变的,并且受缓冲器充满度模型的约束。

[0262] EEE-E15.如EEE-E11所述的方法,其中,对与所述回放设备相关联的所述一个或多个音频信号进行解码包括从所述独立的经编码数据块中联合解码与所述相应回放设备相关联的所述一个或多个音频信号以及与一个或多个附加回放设备相关联的一个或多个附加音频信号。

[0263] EEE-E16.如EEE-E15所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括提取跨两个或更多个音频信号共享的比例因子。

[0264] EEE-E17.如EEE-E16所述的方法,其中,所述两个或更多个音频信号是空间相关的。

[0265] EEE-E18.如EEE-E17所述的方法,其中,所述两个或更多个空间相关音频信号包括左水平声道、左上声道、右水平声道或右上声道。

[0266] EEE-E19.如EEE-E15所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括应用解耦工具。

[0267] EEE-E20.如EEE-E19所述的方法,其中,所述解耦工具包括:

[0268] 提取低于指定频率的独立解码的信号;

[0269] 提取高于所述指定频率的复合信号;

[0270] 从所述复合信号以及与所述复合信号的能量和相应信号的能量相关的比例因子来确定高于所述指定频率的相应解耦信号;以及

[0271] 将每个独立解码的信号与相应的解耦信号组合以获得所述联合解码的信号。

[0272] EEE-E21.如EEE-E15所述的方法,其中,对所述一个或多个音频信号和所述一个或多个附加音频信号进行联合解码包括应用联合解码工具来提取多于两个音频信号。

[0273] EEE-E22.如EEE-E11所述的方法,其中,对与所述回放设备相关联的所述一个或多个音频信号进行解码包括在与所述音频信号被编码的域相同的域中对所述音频信号应用带宽扩展。

[0274] EEE-E23.如EEE-E22所述的方法,其中,所述域是改进离散余弦变换(MDCT)域。

[0275] EEE-E24.如EEE-E22或EEE-E23所述的方法,其中,所述带宽扩展包括自适应噪声添加。

[0276] EEE-E25.一种装置,被配置为执行如EEE-E1至EEE-E24中任一项所述的方法。

[0277] EEE-E26.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如EEE-E1至EEE-E24中任一项所述的方法。

[0278] EEE-F1.一种由具有一个或多个麦克风的设备执行的用于生成经编码比特流的方法,所述方法包括:

[0279] 由所述一个或多个麦克风捕获一个或多个音频信号;

[0280] 分析所捕获的音频信号以确定唤醒词的存在;

[0281] 在检测到唤醒词的存在时:

[0282] 设置标志以指示将对所捕获的音频信号执行语音识别任务:

[0283] 对所捕获的音频信号进行编码;

[0284] 将所述经编码音频信号和所述标志组装成所述经编码比特流。

[0285] EEE-F2.如EEE-F1所述的方法,其中,所述一个或多个麦克风被配置为捕获单声道或空间声场。

[0286] EEE-F3.如EEE-F2所述的方法,其中,所述空间声场采用A格式或B格式。

[0287] EEE-F4.如EEE-F1至EEE-F3中任一项所述的方法,其中,所捕获的音频信号旨在仅用于执行所述语音识别任务。

[0288] EEE-F5.如EEE-F4所述的方法,其中,所捕获的音频信号被编码成使得当对所捕获的音频信号进行解码时,所述经解码音频信号的质量足以执行所述语音识别任务但不足供人类收听。

[0289] EEE-F6.如EEE-F4或EEE-F5所述的方法,其中,在对所捕获的音频信号进行编码之前,所捕获的音频信号被转换为包括以下各项中的一项或多项的表示:频带能量、梅尔频率倒谱系数或改进离散余弦变换(MDCT)频谱系数。

[0290] EEE-F7.如EEE-F1至EEE-F3中任一项所述的方法,其中,所捕获的音频信号旨在供人类收听和用于执行所述语音识别任务两者。

[0291] EEE-F8.如EEE-F7所述的方法,其中,所捕获的音频信号被编码成使得当对所捕获的音频信号进行解码时,所述经解码音频信号的质量足以供人类收听。

[0292] EEE-F9.如EEE-F7所述的方法,其中,对所捕获的音频信号进行编码包括生成所捕

获的音频信号的第一经编码表示和所捕获的音频信号的第二经编码表示,其中,生成所述第一经编码表示,使得当从所述第一经编码表示中解码所捕获的音频信号时,所述经解码音频信号的质量足以供人类收听,并且其中,生成所述第二编码表示,使得当从所述第二经编码表示中解码所捕获的音频信号信号时,所述经解码音频信号的质量足以执行所述语音识别任务但不足以供人类收听。

[0293] EEE-F10.如EEE-F9所述的方法,其中,生成所捕获的音频信号的所述第二经编码表示包括在对所捕获的音频信号进行编码之前,将所捕获的音频信号转换为以下各项中的一项或多项:参数表示、粗略波形表示、或包括频带能量、梅尔频率倒谱系数或改进离散余弦变换(MDCT)频谱系数中的一项或多项的表示。

[0294] EEE-F11.如EEE-F9或EEE-F10所述的方法,其中,将所述经编码音频信号组装成比特流包括将所述第一经编码表示插入到所述经编码比特流的第一独立块中,以及将所述第二经编码表示插入到所述经编码比特流的第二独立块中。

[0295] EEE-F12.如EEE-F9或EEE-F10所述的方法,其中,所述第一经编码表示包括在所述经编码比特流的第一层中,所述第二经编码表示包括在所述经编码比特流的第二层中,并且所述第一层和所述第二层包括在所述经编码比特流的单个块中。

[0296] EEE-F13.如EEE-F1至EEE-F12中任一项所述的方法,进一步包括,当未检测到所述唤醒词的存在时:

[0297] 设置所述标志以指示不对所捕获的音频信号执行语音识别任务;

[0298] 对所捕获的音频信号进行编码;

[0299] 将所述经编码音频信号和所述标志组装成所述经编码比特流。

[0300] EEE-F14.一种用于对音频信号进行解码的方法,所述方法包括:

[0301] 接收经编码比特流,所述经编码比特流包括经编码音频信号和指示是否要执行语音识别任务的标志;

[0302] 对所述经编码音频信号进行解码以获得经解码音频信号;以及

[0303] 当所述标志指示要执行所述语音识别任务时,对所述经解码音频信号执行所述语音识别任务。

[0304] EEE-F15.如EEE-F14所述的方法,其中,所述经解码音频信号旨在仅用于执行所述语音识别任务。

[0305] EEE-F16.如EEE-F15所述的方法,其中,所述经解码音频信号的质量足以执行所述语音识别任务但不足以供人类收听。

[0306] EEE-F17.如权利要求EEE-F15或EEE-F16所述的方法,其中,在对所捕获的音频信号进行编码之前,所述经解码音频信号处于包括频带能量、梅尔频率倒谱系数或改进离散余弦变换(MDCT)频谱系数中的一项或多项的表示中。

[0307] EEE-F18.如EEE-F14所述的方法,其中,所捕获的音频信号被编码成使得当对所捕获的音频信号进行解码时,所述经解码音频信号的质量足以供人类收听。

[0308] EEE-F19.如EEE-F18所述的方法,其中,所述经编码音频信号包括一个或多个音频信号的第一经编码表示以及所述一个或多个音频信号的第二经编码表示。

[0309] EEE-F20.如EEE-F18所述的方法,其中,从所述第一表示解码的音频信号的质量足以供人类收听,并且其中,从所述第二表示解码的音频信号的质量足以执行所述语音识别

任务但不足以供人类收听。

[0310] EEE-F21.如EEE-F19或EEE-F20所述的方法,其中,所述第一表示处于所述经编码比特流的第一独立块中,并且所述第二表示处于所述经编码比特流的第二独立块中。

[0311] EEE-F22.如EEE-F19或EEE-F20所述的方法,其中,所述第一表示处于所述经编码比特流的第一层中,所述第二经编码表示包括在所述经编码比特流的第二层中,并且所述第一层和所述第二层包括在所述经编码比特流的单个块中。

[0312] EEE-F23.如EEE-F18至EEE-F22中任一项所述的方法,其中,对所述经编码音频信号进行解码包括仅对所述第二表示进行解码并且忽略所述第一表示。

[0313] EEE-F24.如EEE-F18至EEE-F23中任一项所述的方法,其中,从所述第二经编码表示解码的音频信号处于参数表示、波形表示或包括频带能量、梅尔频率倒谱系数或改进离散余弦变换(MDCT)频谱系数中的一项或多项的表示中。

[0314] EEE-F25.一种装置,被配置为执行如EEE-F1至EEE-F24中任一项所述的方法。

[0315] EEE-F26.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如EEE-F1至EEE-F24中任一项所述的方法。

[0316] EEE-G1.一种用于对沉浸式音频节目的音频信号进行编码以进行到一个或多个回放设备的低时延传输的方法,所述方法包括:

[0317] 接收所述沉浸式音频节目的多个时域音频信号;

[0318] 选择帧大小;

[0319] 响应于所述帧大小而提取所述时域音频信号的帧,其中,所述时域音频信号的帧与时域音频信号的前一帧重叠;

[0320] 将所述音频信号分割成重叠帧;

[0321] 将所述时域音频信号的帧变换成频域信号;

[0322] 对所述频域信号进行编码;

[0323] 使用感知激励的量化工具对经编码频域信号进行量化;

[0324] 将经量化和经编码的频域信号组装成所述帧内的一个或多个独立块;以及

[0325] 将所述一个或多个独立块组装成经编码帧。

[0326] EEE-G2.如EEE-G1所述的方法,其中,所述多个音频信号包括具有定义的声道配置的基于声道的信号。

[0327] EEE-G3.如EEE-G2所述的方法,其中,所述声道配置是单声道、立体声、5.1、5.1.2、5.1.4、7.1.2、7.1.4、9.1.6或22.2中的一种。

[0328] EEE-G4.如EEE-G1至EEE-G3中任一项所述的方法,其中,所述多个音频信号包括一个或多个基于对象的信号。

[0329] EEE-G5.如EEE-G1至EEE-G4中任一项所述的方法,其中,所述多个音频信号包括所述沉浸式音频节目的基于场景的表示。

[0330] EEE-G6.如EEE-G1至EEE-G5中任一项所述的方法,其中,所选帧大小是128、256、512、1024、120、240、480或960个样本中的一种。

[0331] EEE-G7.如EEE-G1至EEE-G6中任一项所述的方法,其中,所述时域音频信号的帧与所述时域音频信号的前一帧之间的重叠为50%或更低。

[0332] EEE-G8.如EEE-G1至EEE-G7中任一项所述的方法,其中,所述变换是改进离散余弦

变换(MDCT)。

[0333] EEE-G9.如EEE-G1至EEE-G8中任一项所述的方法,其中,所述多个音频信号中的两个或更多个音频信号被联合编码。

[0334] EEE-G10.如EEE-G1至EEE-G9中任一项所述的方法,其中,每个独立块包含用于一个或多个回放设备的经编码信号。

[0335] EEE-G11.如EEE-G1至EEE-G10中任一项所述的方法,其中,至少一个独立块包含用于两个或更多个回放设备的经编码信号,并且所述经编码信号包括联合编码的音频信号。

[0336] EEE-G12.如EEE-G1至EEE-G11中任一项所述的方法,其中,至少一个独立块包含覆盖旨在用于从回放设备的不同驱动器进行回放的不同带宽的多个经编码信号。

[0337] EEE-G13.如EEE-G1至EEE-G12中任一项所述的方法,其中,至少一个独立块包含在由回放设备执行的回声管理中使用的经编码回声参考信号。

[0338] EEE-G14.如EEE-G1至EEE-G13中任一项所述的方法,其中,对经量化的频域信号进行编码包括应用以下工具中的一个或多个:时间噪声整形(TNS)、联合声道编码、跨信号共享比例因子、确定高频重建的控制参数、以及确定噪声取代的控制参数。

[0339] EEE-G15.如EEE-G1至EEE-G14中任一项所述的方法,其中,一个或多个独立块包括用于控制回放设备的延迟、增益和均衡中的一项或多项的参数。

[0340] EEE-G16.一种用于从经编码信号中解码沉浸式音频节目的音频信号的低时延方法,所述方法包括:

[0341] 接收包括一个或多个独立块的经编码帧;

[0342] 从一个或多个独立块中提取经量化和经编码的频域信号;

[0343] 对所述经量化和经编码的频域信号进行解量化;

[0344] 对解量化的频域信号进行解码;

[0345] 对经解码的频域信号进行逆变换以获得时域信号;以及

[0346] 将所述时域信号与来自前一帧的时域信号重叠并相加,以提供所述沉浸式音频节目的多个音频信号。

[0347] EEE-G17.如EEE-G16所述的方法,其中,所述多个音频信号包括具有定义的声道配置的基于声道的信号。

[0348] EEE-G18.如EEE-G17所述的方法,其中,所述声道配置是单声道、立体声、5.1、5.1.2、5.1.4、7.1.2、7.1.4、9.1.6或22.2中的一种。

[0349] EEE-G19.如EEE-G16至EEE-G18中任一项所述的方法,其中,所述多个音频信号包括一个或多个基于对象的信号。

[0350] EEE-G20.如EEE-G16至EEE-G19中任一项所述的方法,其中,所述多个音频信号包括所述沉浸式音频节目的基于场景的表示。

[0351] EEE-G21.如EEE-G16至EEE-G20中任一项所述的方法,其中,时域样本的帧包括128、256、512、1024、120、240、480或960个样本中的一种。

[0352] EEE-G22.如EEE-G16至EEE-G21中任一项所述的方法,其中与所述前一帧的所述重叠为50%或更低。

[0353] EEE-G23.如EEE-G16至EEE-G22中任一项所述的方法,其中,所述逆变换是逆改进离散余弦变换(IMDCT)。

[0354] EEE-G24.如EEE-G16至EEE-G23中任一项所述的方法,其中,每个独立块包含用于一个或多个回放设备的经量化和经编码的频域信号。

[0355] EEE-G25.如EEE-G16至EEE-G24中任一项所述的方法,其中,至少一个独立块包含用于两个或更多个回放设备的经量化和编码的频域信号,并且所述经量化和经编码的频域信号是联合编码的音频信号。

[0356] EEE-G26.如EEE-G16至EEE-G25中任一项所述的方法,其中,至少一个独立块包含覆盖旨在用于从回放设备的不同驱动器进行回放的不同带宽的多个经量化和经编码的频域信号。

[0357] EEE-G27.如EEE-G16至EEE-G26中任一项所述的方法,其中,至少一个独立块包含在由回放设备执行的回声管理中使用的经编码回声参考信号。

[0358] EEE-G28.如EEE-G16至EEE-G27中任一项所述的方法,其中,对所述解量化的频域信号进行解码包括应用以下解码工具中的一个或多个:时间噪声整形(TNS)、联合声道解码、跨信号共享比例因子、高频重建、以及噪声取代。

[0359] EEE-G29.如EEE-G16至EEE-G28中任一项所述的方法,其中,一个或多个独立块包括用于控制回放设备的延迟、增益和均衡中的一项或多项的参数。

[0360] EEE-G30.如EEE-G16至EEE-G29中任一项所述的方法,其中,所述方法由回放设备执行,并且其中,从一个或多个独立块中提取经量化和经编码的信号包括仅选择包含用于由所述回放设备回放的经量化和经编码的频域信号的那些块,以及忽略包含由其他回放设备回放的经量化和经编码的频域信号的独立块。

[0361] EEE-G31.一种装置,被配置为执行如EEE-G1至EEE-G30中任一项所述的方法。

[0362] EEE-G32.一种非暂态计算机可读存储介质,包括一系列指令,所述一系列指令在执行时使一个或多个设备执行如EEE-G1至EEE-G30中任一项所述的方法。

[0363] 在下文的权利要求和本文的描述中,术语包括(comprising)、包括(comprised of)或其包括(which comprises)中的任何一个是开放术语,其意指至少包括随后的元素/特征,但不排除其他元素/特征。因此,当在权利要求中使用术语包括(comprising)时,该术语不应当被解释为限于在其之后列出的装置或元件或步骤。例如,包括A和B的设备这一表述的范围不应限于仅包括元素A和B的设备。如本文所使用的,术语包括(including)或其包括(which includes)或包括(that includes)中的任何一个也是开放术语,其也意指至少包括所述术语之后的元素/特征,但不排除其他元素/特征。因此,包括(including)与包括(comprising)同义并且意指包括(comprising)。

[0364] 应当理解,在以上对本发明的示例的描述中,有时在单个示例、图或其描述中将各种特征分组在一起,以便简化本公开,并且帮助理解各个创造性方面中的一个或多个方面。然而,本公开的这个方法不应当被解释为反映需要比每项权利要求中所明确叙述的特征更多的特征的意图。相反,如以下权利要求所反映的,创造性方面在于少于单个前述所公开示例的所有特征。因此,在具体实施方式后面的权利要求特此明确地并入到该具体实施方式中,其中,每个权利要求都独立地作为本发明的单独示例。

[0365] 此外,虽然本文所描述的一些示例包括其他示例中所包括的一些特征而不包括其他示例中所包括的其他特征,但是如本领域技术人员将理解的,旨在涵盖不同示例的特征的组合并且不同示例的特征的组合形成不同示例。例如,在以下权利要求中,要求保护的示

例中的任一个示例都可以以任何组合来使用。

[0366] 此外,某些示例在本文中被描述为可以由计算机系统的处理器或由执行功能的其他装置实施的方法或方法元素的组合。因此,具有执行这种方法或方法元素所需的指令的处理器形成用于执行方法或方法元素的装置。此外,本文所描述的装置的元件是用于执行由元件执行的功能的装置的示例。

[0367] 此外,本文所描述的一些示例公开了连接的解决方案,所述连接的解决方案应当被解释为可能在如有线系统和/或无线系统等分发和/或传输系统中实施。例如,通过使用任何电气、光学和/或移动系统,如3G、4G以及5G。

[0368] 因此,虽然已描述了本发明的具体示例,但本领域技术人员将认识到可以进行其他和进一步修改,并且旨在要求保护所有这些改变和修改。例如,上文给出的任何公式仅仅表示可以使用的程序。可以从框图中添加或删除功能,并且可以在功能块当中互换操作。所描述的方法可以添加或删除步骤。

[0369] 上文所公开的系统、设备和方法可以被实施为软件、固件、硬件或其组合。例如,本申请的各方面可以至少部分地体现在设备、包括多于一个设备的系统、方法、计算机程序产品等中。

[0370] 在硬件实施方式中,以上描述中所提及的功能单元之间的任务划分不一定对应于物理单元的划分;相反,一个物理部件可以具有多个功能,并且一个任务可以由若干个物理部件协作地执行。

[0371] 某些部件或所有部件可以被实施为由数字信号处理器或微处理器执行的软件或者被实施为硬件或专用集成电路。这种软件可以分布在计算机可读介质上,所述计算机可读介质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)。

[0372] 如本领域技术人员所熟知的,术语计算机存储介质包括以用于存储如计算机可读指令、数据结构、程序模块或其他数据等信息的任何方法或技术实施的易失性和非易失性、可移除和不可移除的介质。计算机存储介质包括但不限于:RAM、ROM、EEPROM、闪速存储器或其他存储器技术、CD-ROM、数字通用盘(DVD)或其他光盘存储设备、磁带盒、磁带、磁盘存储装置或其他磁性存储设备、或可以用于存储期望信息并且可以被计算机访问的任何其他介质。

[0373] 进一步地,本领域技术人员所熟知的是,通信介质通常以如载波等所调制数据信号或其他传输机制的形式来体现计算机可读指令、数据结构、程序模块或其他数据,并且包括任何信息传递介质。

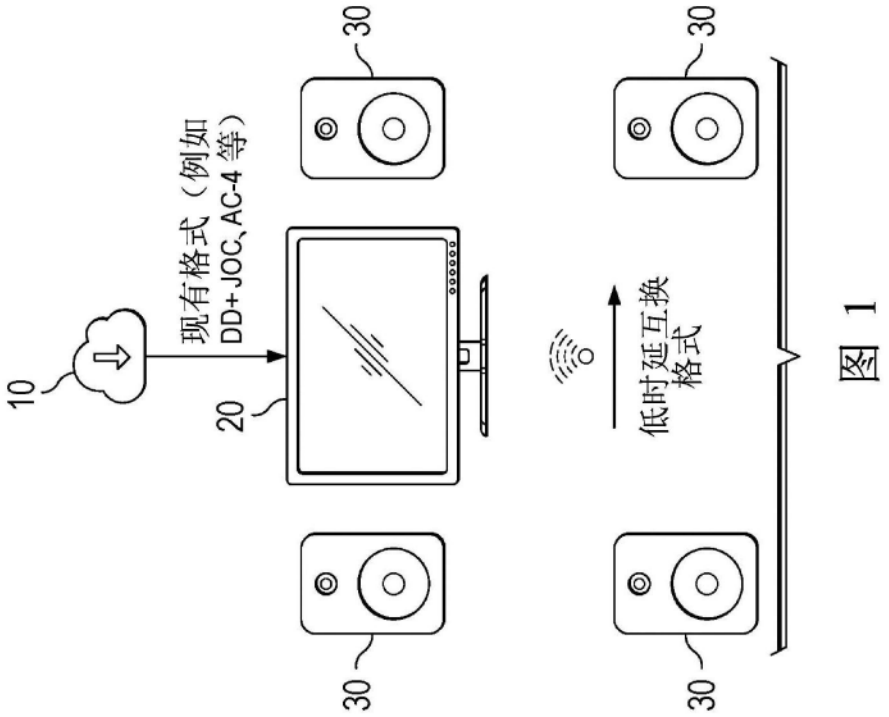


图1

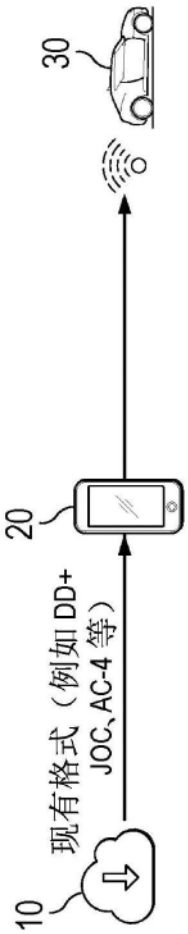


图2

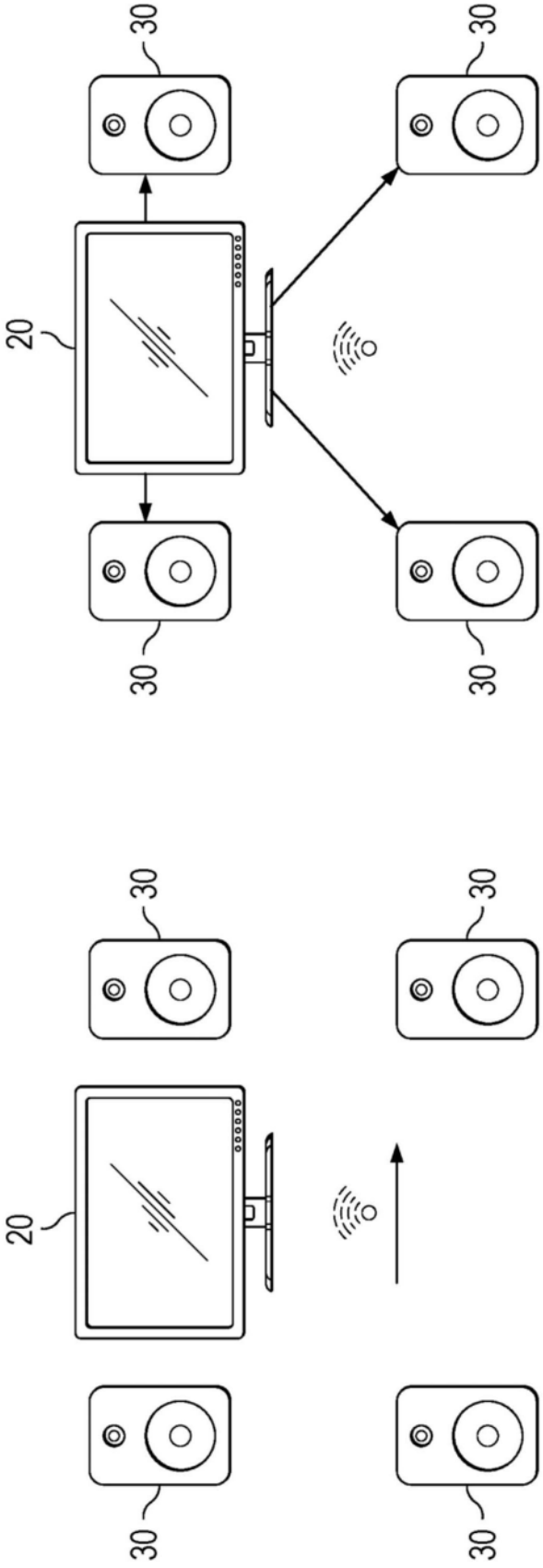


图3

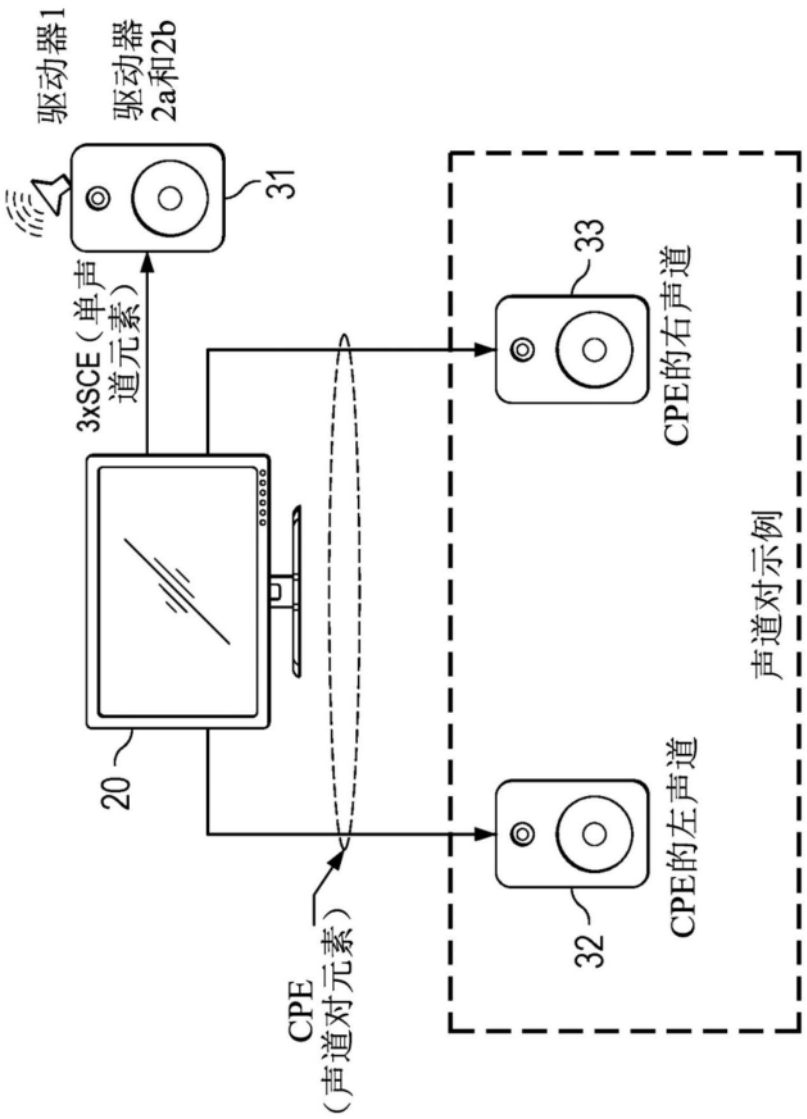

驱动器1
驱动器2a和2b
31
3xSCE（单声道元素）
33
CPE的右声道
20
声道对示例
CPE的左声道
32
CPE（声道对元素）

图4

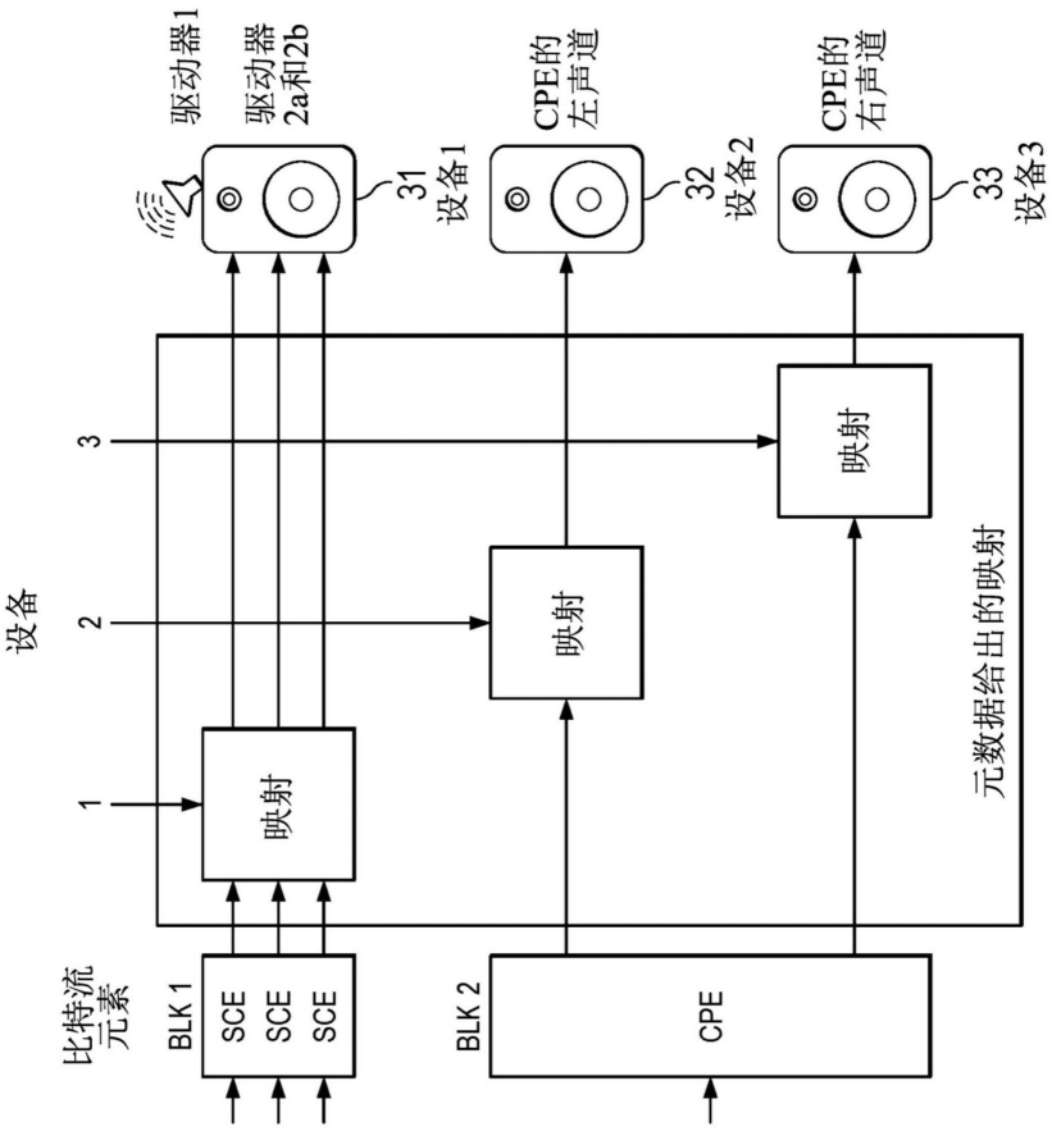

比特流元素
BLK 1
SCE
SCE
SCE
BLK 2
CPE
设备
1
2
3
映射
映射
映射
元数据给出的映射
驱动器1
驱动器2a和2b
31
设备1
CPE的左声道
32
设备2
CPE的右声道
33
设备3

图5

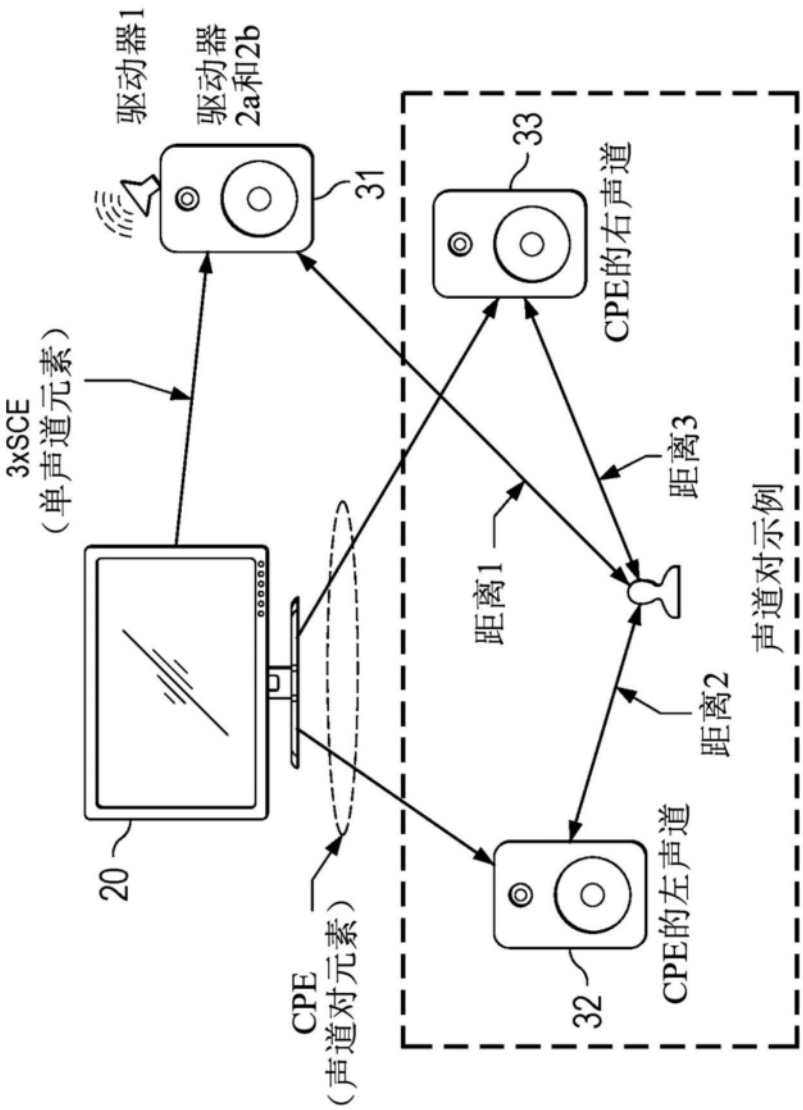

驱动器1
驱动器2a和2b
31
33
CPE的右声道
3xSCE
（单声道元素）
距离3
距离1
声道对示例
距离2
20
CPE
（声道对元素）
32
CPE的左声道

图6

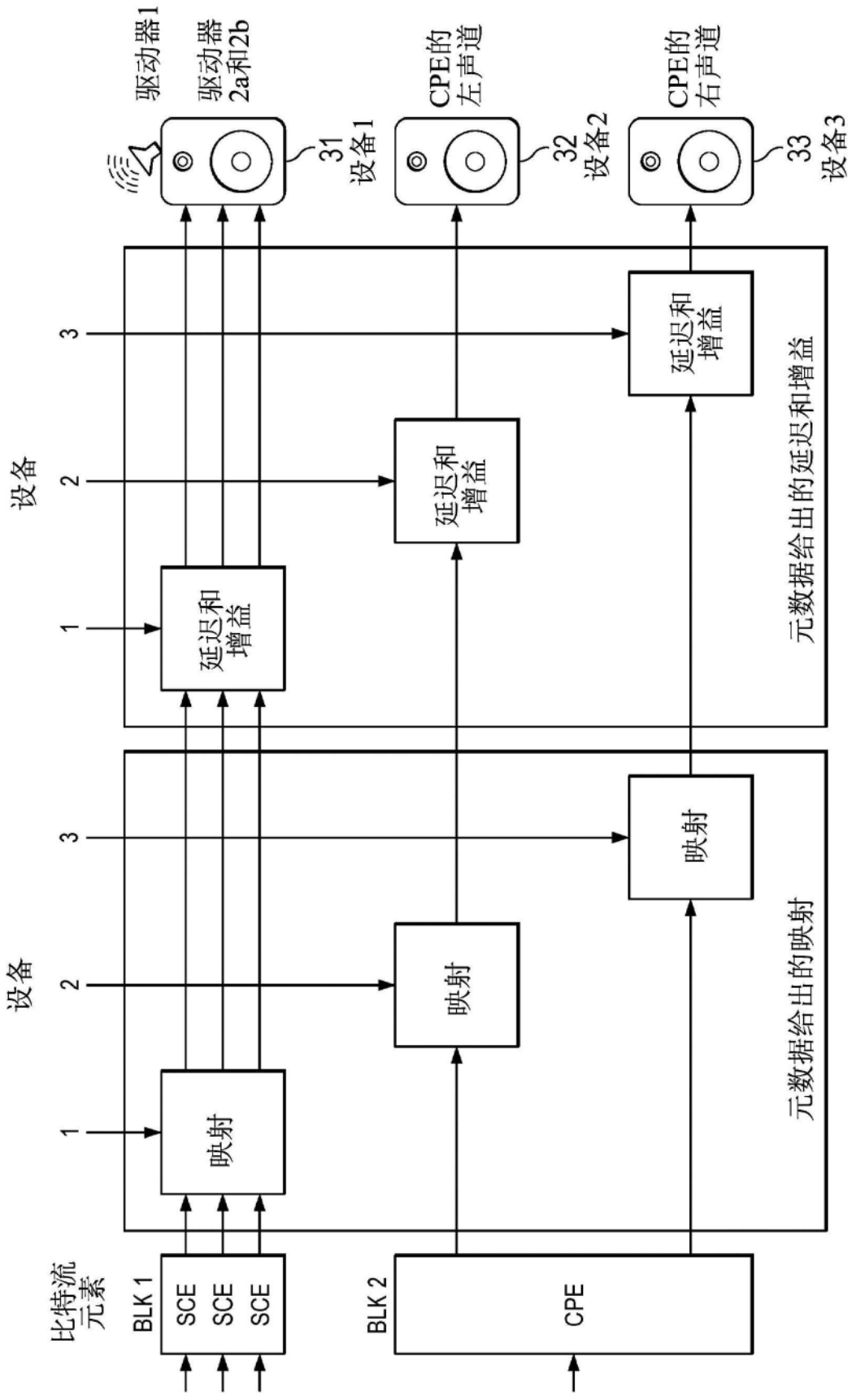

比特流元素
BLK 1
SCE
SCE
SCE
BLK 2
CPE
设备
1
2
3
映射
映射
映射
元数据给出的映射
延迟和增益
延迟和增益
延迟和增益
元数据给出的延迟和增益
驱动器1
驱动器2a和2b
31
设备1
CPE的左声道
32
设备2
CPE的右声道
33
设备3

图7

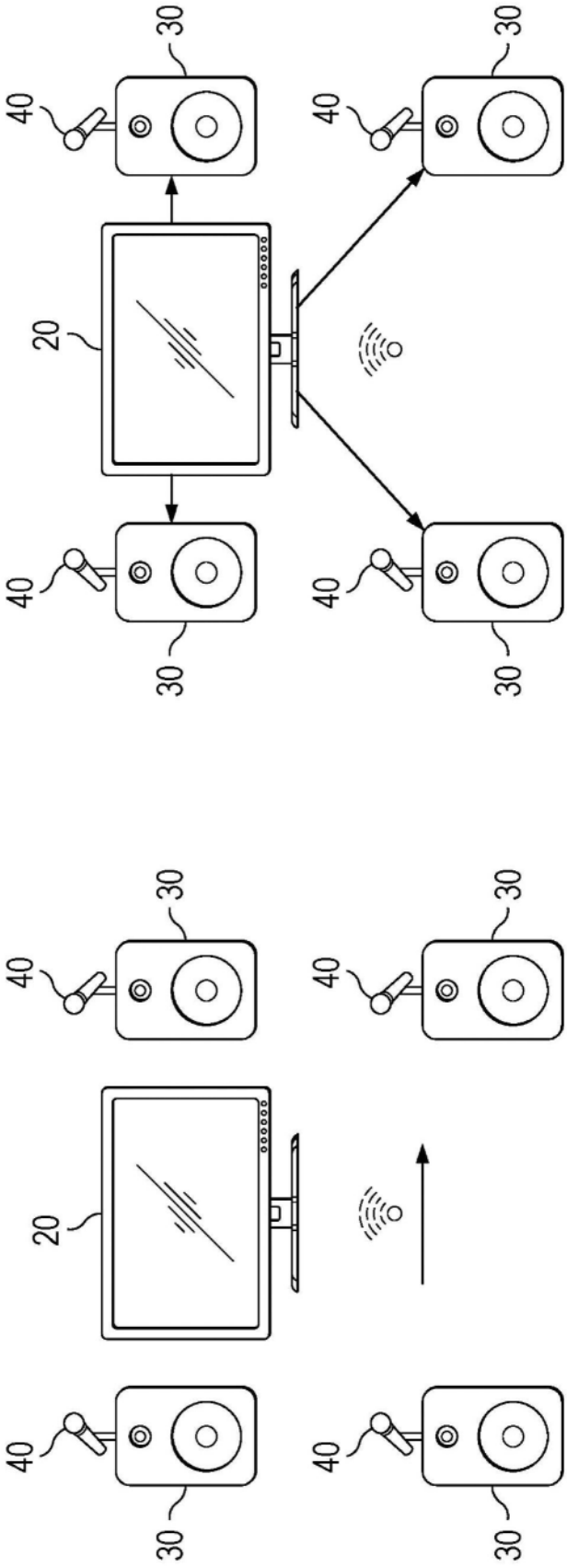


图8

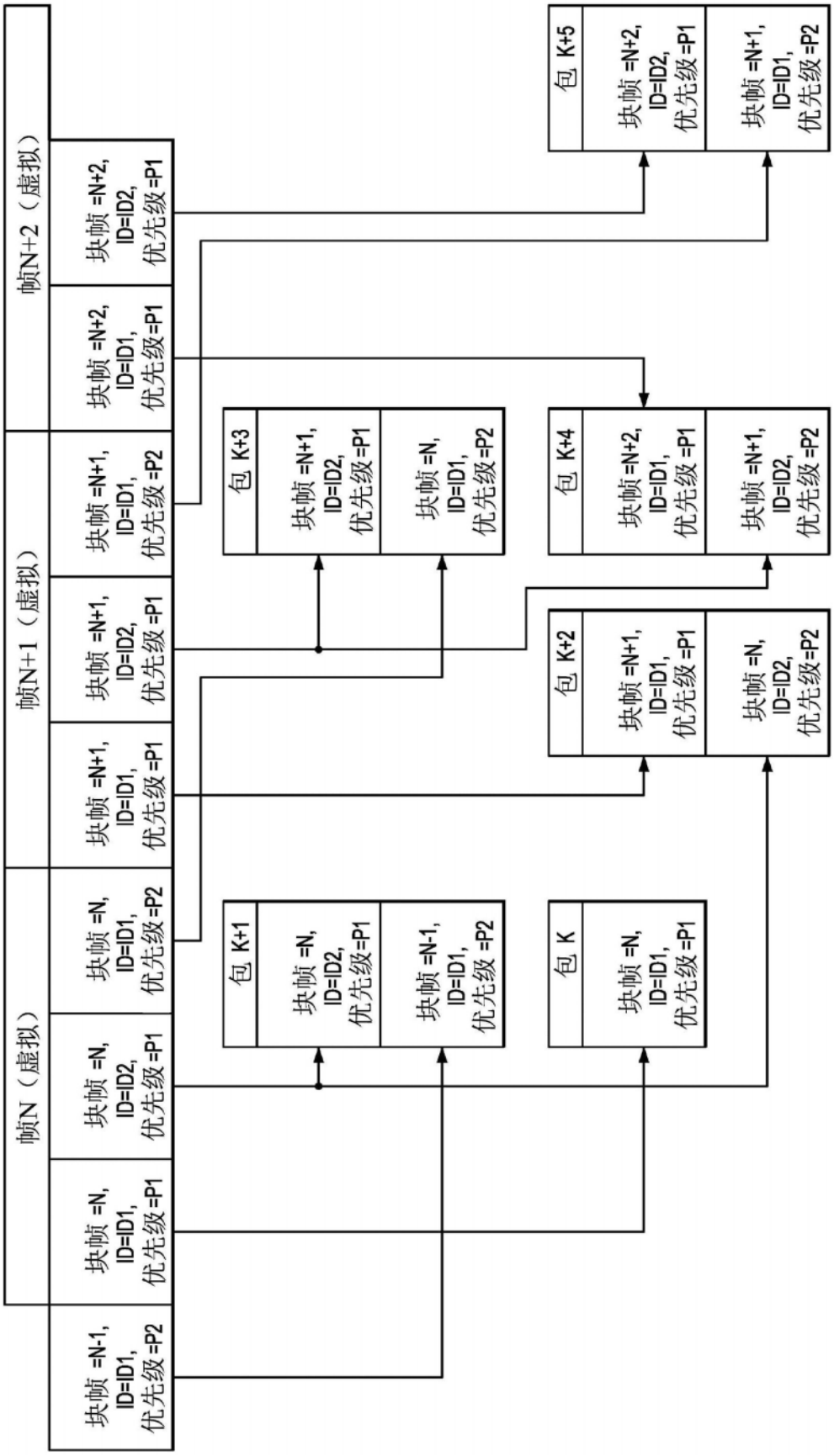

帧N（虚拟）
帧N+1（虚拟）
帧N+2（虚拟）
块帧 =N-1, ID=ID1, 优先级 =P2
块帧 =N, ID=ID1, 优先级 =P1
块帧 =N, ID=ID2, 优先级 =P1
块帧 =N, ID=ID1, 优先级 =P2
块帧 =N+1, ID=ID1, 优先级 =P1
块帧 =N+1, ID=ID2, 优先级 =P1
块帧 =N+1, ID=ID1, 优先级 =P2
块帧 =N+2, ID=ID1, 优先级 =P1
块帧 =N+2, ID=ID2, 优先级 =P1
包 K
包 K+1
包 K+2
包 K+3
包 K+4
包 K+5

图9

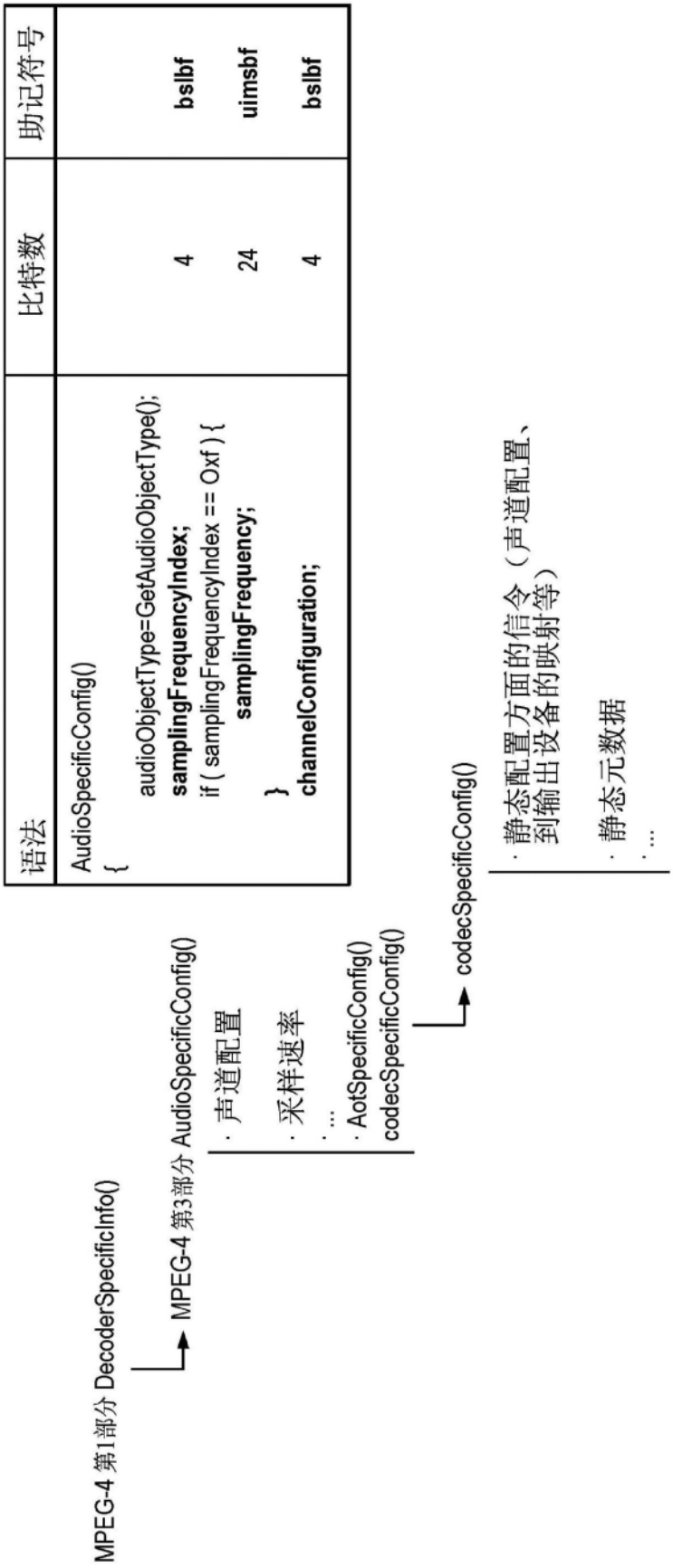

语法
比特数
助记符号
AudioSpecificConfig()
{
audioObjectType=GetAudioObjectType();
samplingFrequencyIndex;
4
bslbf
if (samplingFrequencyIndex == Oxf) {
samplingFrequency;
24
uimsbf
}
channelConfiguration;
4
bslbf
MPEG-4 第1部分 DecoderSpecificInfo()
MPEG-4 第3部分 AudioSpecificConfig()
· 声道配置
· 采样速率
· ...
· AotSpecificConfig()
codecSpecificConfig()
codecSpecificConfig()
· 静态配置方面的信令（声道配置、到输出设备的映射等）
· 静态元数据
· ...

图10

raw_data_block_codec（配置数据）

- 类似于MPEG-4中的现有原始数据块
- 其中附加信息使得能够跳过数据块的给定设备不需要的部分
- 动态元数据

语法	比特数	助记符号

raw_data_block()
{
while((id= **id_syn_ele**) != ID_END){
switch (id) {
case ID_SCE: single_channel_element();
break;
case ID_CPE: channel_pair_element();
break;
case ID_CCE: coupling_channel_element();
break;
case ID_LFE: life_channel_element();
break;
case ID_DSE: data_stream_element();
break;
case ID_PCE: program_config_element();
break;
case ID_FIL: fill_element();
	3	**uimsbf**

图11

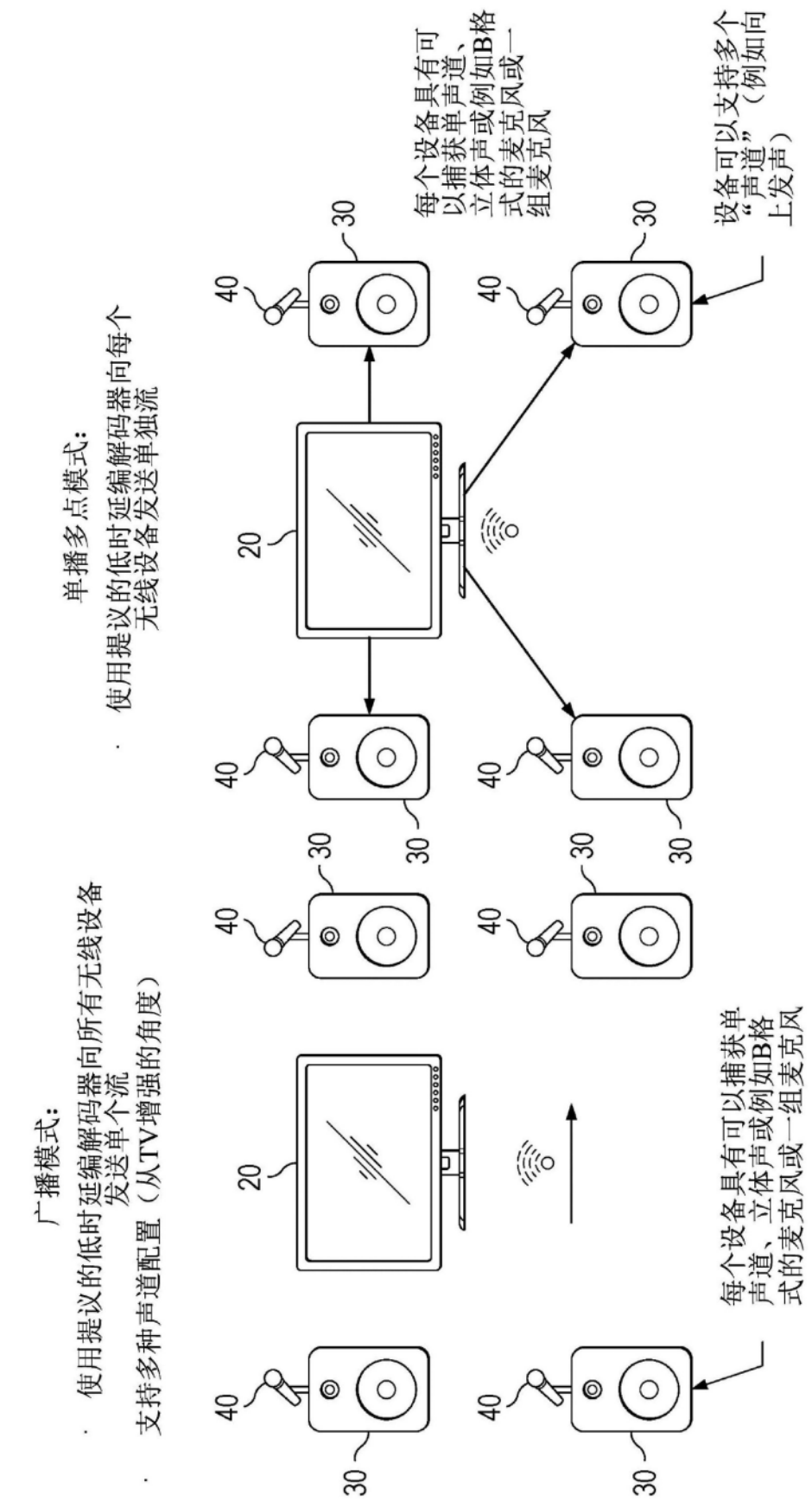

广播模式：
· 使用提议的低时延编解码器向所有无线设备发送单个流
· 支持多种声道配置（从TV增强的角度）
单播多点模式：
· 使用提议的低时延编解码器向每个无线设备发送单独流
每个设备具有可以捕获单声道、立体声或例如B格式的麦克风或一组麦克风
设备可以支持多个“声道”（例如向上发声）
每个设备具有可以捕获单声道、立体声或例如B格式的麦克风或一组麦克风
20
30
40

图12

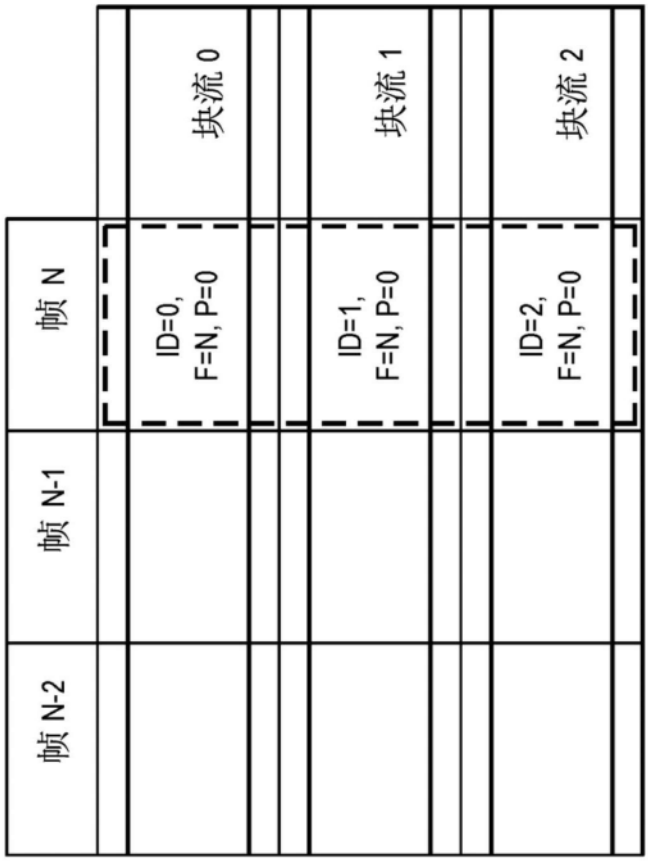


图13

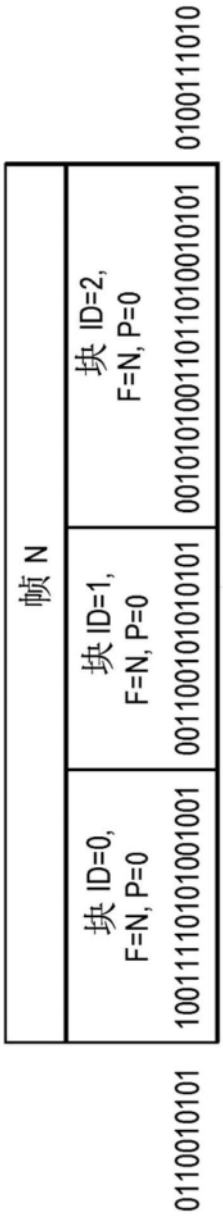

帧 N
块 ID=0,
F=N, P=0
1001111010100100 1
块 ID=1,
F=N, P=0
0011001010101010 1
块 ID=2,
F=N, P=0
0010101001101101001010 1
0110010101
0100111010

图14

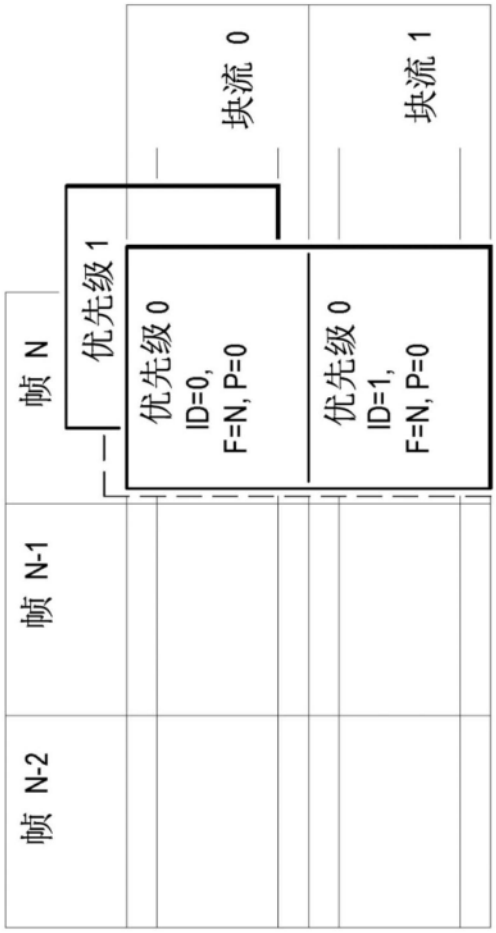


图15

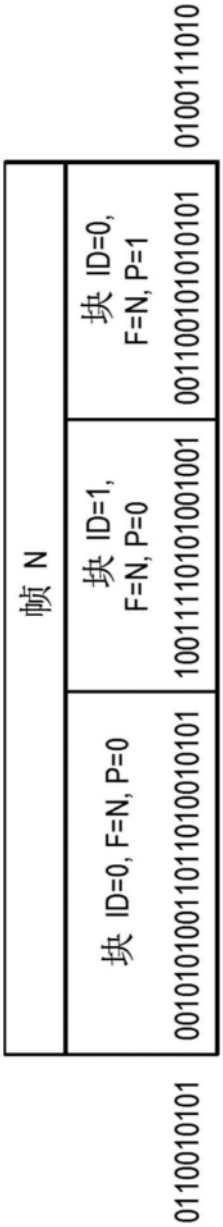

帧 N
块 ID=0, F=N, P=0
0010101001101101010010101
块 ID=1, F=N, P=0
1001111010101001001
块 ID=0, F=N, P=1
0011001010101010101
0110010101
0100111010

图16

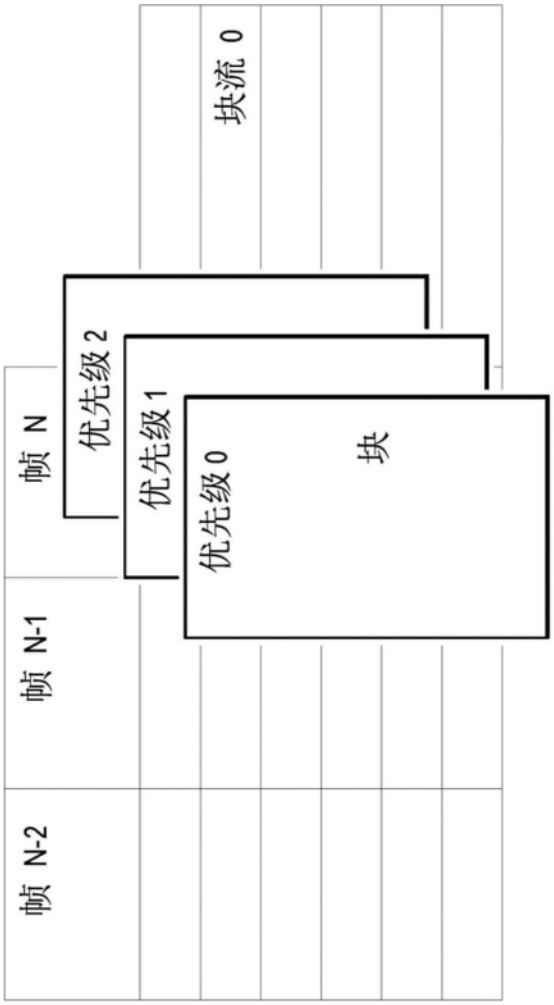


图17

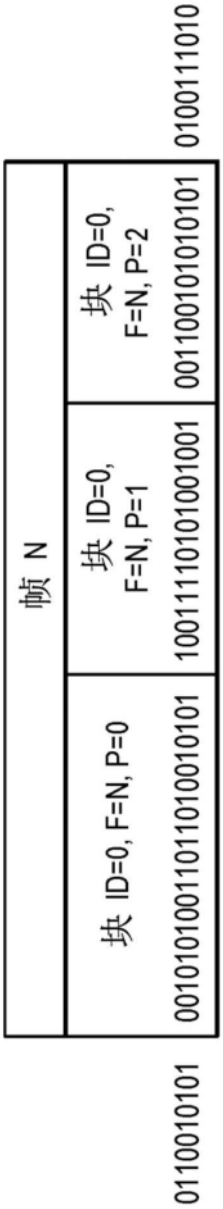

帧 N
块 ID=0, F=N, P=0
0010101001101101010010101
块 ID=0, F=N, P=1
1001111010100100 1001
块 ID=0, F=N, P=2
0011001010101010 1
0100111010
0110010101

图18