



(12) 发明专利

(10) 授权公告号 CN 110600043 B

(45) 授权公告日 2025. 07. 15

(21) 申请号 201910831687.6

(22) 申请日 2013.07.31

(65) 同一申请的已公布的文献号  
申请公布号 CN 110600043 A

(43) 申请公布日 2019.12.20

(30) 优先权数据  
61/836,865 2013.06.19 US

(62) 分案原申请数据  
201310329128.8 2013.07.31

(73) 专利权人 杜比实验室特许公司  
地址 美国加利福尼亚州

(72) 发明人 杰弗里·里德米勒 迈克尔·沃德

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

专利代理师 杨林森

(51) Int.Cl.  
G10L 19/008 (2013.01)  
G10L 19/16 (2013.01)

(56) 对比文件  
CN 101160616 A, 2008.04.09  
CN 102792588 A, 2012.11.21

审查员 陈成

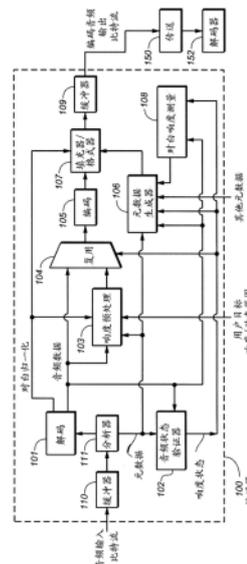
权利要求书1页 说明书35页 附图6页

(54) 发明名称

音频处理单元、由音频处理单元执行的方法和存储介质

(57) 摘要

本公开涉及音频处理单元、由音频处理单元执行的方法和存储介质。一种用于包括通过将子流结构元数据 (SSM) 和/或节目信息元数据 (PIM) 以及音频数据包括在比特流中来生成编码音频比特流的设备和方法。其他方面是用于对这样的比特流进行解码的设备和方法, 以及被配置成 (例如, 被编程成) 执行该方法的任意实施方式或包括存储根据该方法的任意实施方式而生成的音频比特流的至少一个帧的缓冲存储器的音频处理单元 (例如, 编码器、解码器或后处理器)。



1. 一种音频处理单元,包括:

一个或更多个处理器;

存储器,其耦接至所述一个或更多个处理器并且被配置成存储指令,所述指令在由所述一个或更多个处理器执行时使所述一个或更多个处理器执行操作,所述操作包括:

接收包括音频节目的编码音频比特流,所述编码音频比特流包括一个或更多个音频通道的集合的编码音频数据以及与音频通道的所述集合相关联的元数据,其中所述元数据包括响度处理状态元数据,并且其中所述响度处理状态元数据包括指示所述音频节目的响度的元数据;

对所述编码音频数据进行解码以获得音频通道的所述集合的解码音频数据;

从所述编码音频比特流的元数据中获得所述响度处理状态元数据;以及

基于所述响度处理状态元数据对音频通道的所述集合的所述解码音频数据执行自适应响度处理,其中,所述元数据还包括节目信息元数据,所述节目信息元数据指示用于在所述比特流中创建动态范围压缩DRC数据的压缩配置文件,其中所述压缩配置文件是语音压缩配置文件。

2. 一种由音频处理单元执行的方法,包括:

接收包括音频节目的编码音频比特流,所述编码音频比特流包括一个或更多个音频通道的集合的编码音频数据以及与音频通道的所述集合相关联的元数据,其中所述元数据包括响度处理状态元数据,并且其中所述响度处理状态元数据包括指示所述音频节目的响度的元数据;

对所述编码音频数据进行解码以获得音频通道的所述集合的解码音频数据;

从所述编码音频比特流的元数据中获得所述响度处理状态元数据;以及

基于所述响度处理状态元数据对音频通道的所述集合的解码音频数据执行自适应响度处理,其中,所述元数据还包括节目信息元数据,所述节目信息元数据指示用于在所述比特流中创建动态范围压缩DRC数据的压缩配置文件,其中所述压缩配置文件是语音压缩配置文件。

3. 一种存储有指令的非暂态计算机可读存储介质,所述指令在由一个或更多个处理器执行时使所述一个或更多个处理器执行操作,所述操作包括:

接收包括音频节目的编码音频比特流,所述编码音频比特流包括一个或更多个音频通道的集合的编码音频数据以及与音频通道的所述集合相关联的元数据,其中所述元数据包括响度处理状态元数据,并且其中所述响度处理状态元数据包括指示所述音频节目的响度的元数据;

对所述编码音频数据进行解码以获得音频通道的所述集合的解码音频数据;

从所述编码音频比特流的元数据中获得所述响度处理状态元数据;以及

基于所述响度处理状态元数据对音频通道的所述集合的解码音频数据执行自适应响度处理,其中,所述元数据还包括节目信息元数据,所述节目信息元数据指示用于在所述比特流中创建动态范围压缩DRC数据的压缩配置文件,其中所述压缩配置文件是语音压缩配置文件。

## 音频处理单元、由音频处理单元执行的方法和存储介质

[0001] 本申请是申请日为2013年7月31日、申请号为“201310329128.8”、发明名称为“使用节目信息或子流结构元数据的音频编码器和解码器”的发明专利申请的分案申请。

### 技术领域

[0002] 本发明涉及音频信号处理,以及更具体地,涉及具有指示与由比特流所指示的音频内容有关的子流结构和/或节目信息的元数据的音频数据比特流的编码和解码。本发明的一些实施方式以被称为杜比数字(AC-3)、杜比数字+(增强的AC-3或E-AC-3)或杜比E的格式中的一种格式生成或解码音频数据。

### 背景技术

[0003] 杜比、杜比数字、杜比数字+、和杜比E是杜比实验室特许公司的商标。杜比实验室提供分别被称为杜比数字和杜比数字+的AC-3和E-AC-3的专有实现。

[0004] 音频数据处理单元通常以盲方式(blind fashion)操作并且不关注在数据被接收之前发生的音频数据的处理历史。这可以在这样的处理框架中工作:其中单个实体进行各种目标媒体渲染装置的所有的音频数据处理和编码而目标媒体渲染装置进行编码音频数据的所有的解码和渲染。然而,该盲处理在多个音频处理单元跨多样化的网络被散布(scatter)或串联(即,链)放置并且期望它们最佳地执行其相应类型的音频处理的情形下不能很好地(或完全不)工作。例如,一些音频数据可能针对高性能媒体系统被编码,并且可能需要被转换成适合于沿着媒体处理链的移动设备的简化形式。因此,音频处理单元可能不必要地对音频数据执行已经被执行过的类型的处理。例如,音量校平(leveling)单元可能对输入音频片断执行处理,不管以前是否已经对输入音频片断执行了相同的或相似的音量校平。因此,即使当不必要时,音量校平单元也可能执行校平。该不必要的处理还可能当渲染音频数据的内容时具体特征的退化和/或消除。

### 发明内容

[0005] 本发明公开了一种音频处理单元,包括:一个或多个处理器;存储器,其耦接至一个或多个处理器并且被配置成存储指令,指令在由一个或多个处理器执行时使一个或多个处理器执行操作,操作包括:接收包括音频节目的编码音频比特流,编码音频比特流包括一个或多个音频通道的集合的编码音频数据以及与音频通道的集合相关联的元数据,其中元数据包括响度处理状态元数据,并且其中响度处理状态元数据包括指示音频节目的响度的元数据;对编码音频数据进行解码以获得音频通道的集合的解码音频数据;从编码音频比特流的元数据中获得响度处理状态元数据;以及基于响度处理状态元数据对音频通道的集合的解码音频数据执行自适应响度处理。

[0006] 本发明还公开了一种由音频处理单元执行的方法,包括:接收包括音频节目的编码音频比特流,编码音频比特流包括一个或多个音频通道的集合的编码音频数据以及与音频通道的集合相关联的元数据,其中元数据包括响度处理状态元数据,并且其中响度处

理状态元数据包括指示音频节目的响度的元数据;对编码音频数据进行解码以获得音频通道的集合的解码音频数据;从编码音频比特流的元数据中获得响度处理状态元数据;以及基于响度处理状态元数据对音频通道的集合的解码音频数据执行自适应响度处理。

[0007] 本发明又公开了一种存储有指令的非暂态计算机可读存储介质,指令在由一个或多个处理器执行时使一个或多个处理器执行操作,操作包括:接收包括音频节目的编码音频比特流,编码音频比特流包括一个或多个音频通道的集合的编码音频数据以及与音频通道的集合相关联的元数据,其中元数据包括响度处理状态元数据,并且其中响度处理状态元数据包括指示音频节目的响度的元数据;对编码音频数据进行解码以获得音频通道的集合的解码音频数据;从编码音频比特流的元数据中获得响度处理状态元数据;以及基于响度处理状态元数据对音频通道的集合的解码音频数据执行自适应响度处理。

[0008] 在一类实施方式中,本发明是能够对编码比特流进行解码的音频处理单元,该编码比特流包括比特流的至少一个帧的至少一个段中的子流结构元数据和/或节目信息元数据(可选地还包括其他元数据,例如,响度处理状态元数据)以及帧的至少一个其他段中的音频数据。在本文中,子流结构元数据(或“SSM”)表示编码比特流(或编码比特流的集合)的元数据,其指示编码比特流的音频内容的子流结构,并且“节目信息元数据”(或“PIM”)表示编码音频比特流的元数据,其指示至少一个音频节目(例如,两个或多个音频节目),其中节目信息元数据指示至少一个所述节目的音频内容的至少一个属性或特性(例如,指示对节目的音频数据执行的处理的类型或参数的元数据,或指示节目的哪些通道是活动通道(active channel)的元数据)。

[0009] 在典型的情况(例如,其中编码比特流为AC-3或E-AC-3比特流)下,节目信息元数据(PIM)指示实际上不能在比特流的其他部分中携带的节目信息。例如,PIM可以指示在编码(例如,AC-3或E-AC-3编码)之前对PCM音频所应用的处理,音频节目的哪些频带已经使用具体的音频编码技术被编码以及用于在比特流中创建动态范围压缩(DRC)数据的压缩配置文件(profile)。

[0010] 在另一类实施方式中,方法包括在比特流的每个帧(或至少一些帧中的每个帧)中将编码音频数据与SSM和/或PIM复用的步骤。在典型的解码中,解码器从比特流中提取SSM和/或PIM(包括通过对SSM和/或PIM以及音频数据进行分析 and 去复用),并且对音频数据进行处理以生成解码音频数据的流(以及在某些情况下还执行音频数据的自适应处理)。在一些实施方式中,解码音频数据以及SSM和/或PIM从解码器被转发至后处理器,该后处理器被配置成使用SSM和/或PIM对解码音频数据执行自适应处理。

[0011] 在一类实施方式中,本发明的编码方法生成包括音频数据段(例如,图4所示的帧的AB0至AB5段或图7所示的帧的段AB0至AB5中的全部或一些)的编码音频比特流(例如,AC-3或E-AC-3比特流),音频数据段包括编码音频数据以及与音频数据段时分复用的元数据段(包括SSM和/或PIM,可选地还包括其他元数据)。在一些实施方式中,每个元数据段(在本文中有时称为“容器”)具有包括元数据段报头(可选地还包括其他强制性的或“核心”元素)、以及在元数据段报头之后的一个或多个元数据有效载荷。如果存在,SIM被包括在元数据有效载荷之一中(由有效载荷报头标识,并且通常具有第一类型的格式)。如果存在,PIM被包括在元数据有效载荷中的另一个中(由有效载荷报头标识,并且通常具有第二类型的格式)。类似地,元数据的每个其他类型(如果存在)被包括在元数据有效载荷中的另一个中

(由有效载荷报头标识,并且通常具有特定于元数据的类型的格式)。示例性格式允许在除了比特流的解码期间之外的时间(例如,由解码之后的后处理器,或由被配置成在不执行对编码比特流的完全解码的情况下识别元数据的处理器)对SSM、PIM或其他元数据的方便的访问,并且允许在比特流的解码期间(例如,子流识别的)方便的和高效的误差检测和校正。例如,在不以示例性格式访问SSM的情况下,解码器可能错误地识别与节目相关联的子流的正确数量。元数据段中的一个元数据有效载荷可以包括SSM,元数据段中的另一元数据有效载荷可以包括PIM,并且可选地,元数据段中的至少一个其他元数据有效载荷可以包括其他元数据(例如,响度处理状态元数据或“LPSM”)。

## 附图说明

[0012] 图1是可以被配置成执行本发明的方法的实施方式的系统的实施方式的框图。

[0013] 图2是作为本发明的音频处理单元的实施方式的编码器的框图。

[0014] 图3是作为本发明的音频处理单元的实施方式的解码器以及作为本发明的音频处理单元的另一实施方式的耦接至解码器的后处理器的框图。

[0015] 图4是包括被划分成的段的AC-3帧的图。

[0016] 图5是包括被划分成的段的AC-3帧的同步信息(SI)段的图。

[0017] 图6是包括被划分成的段的AC-3帧的比特流信息(BSI)段的图。

[0018] 图7是包括被划分成的段的E-AC-3帧的图。

[0019] 图8是根据本发明的实施方式生成的包括元数据段报头的编码比特流的元数据段的图,元数据段报头包括容器同步字(在图8中标识为“容器同步”)以及版本和键ID值,之后是多个元数据有效载荷以及保护位。

[0020] 符号和术语

[0021] 贯穿包括权利要求在内的本公开内容,“对”信号或数据执行操作(例如,对信号或数据进行滤波、缩放、变换或施加增益)的表达用于广义上表示对信号或数据、或对信号或数据的已处理版本(例如,对在对信号执行操作之前已经经历了初步滤波或预处理的信号版本)直接执行操作。

[0022] 贯穿包括权利要求在内的本公开内容,“系统”的表达用于广义上表示设备、系统或子系统。例如,实现解码器的子系统可以称为解码器系统,并且包括这样的子系统的系统(例如,响应于多个输入生成X个输出信号的系统,在该系统中,子系统生成M个输入并且其他X—M个输入从外部源接收)也可以称为解码器系统。

[0023] 贯穿包括权利要求在内的本公开内容,术语“处理器”用于广义上表示可编程或以其他方式可配置成(例如,使用软件或固件)对数据(例如,音频数据或视频数据或其他图像数据)执行操作的系统或装置。处理器的示例包括现场可编程门阵列(或其他可配置的集成电路或芯片组)、被编程和/或被以其他方式配置成对音频数据或其他声音数据执行流水线处理的数字信号处理器、可编程的通用处理器或计算机以及可编程的微处理器芯片或芯片组。

[0024] 贯穿包括权利要求在内的本公开内容,“音频处理器”和“音频处理单元”的表达用于可交换地广义上表示被配置成对音频数据进行处理的系统。音频处理单元的示例包括但不限于编码器(例如,代码转换器)、解码器、编解码器、预处理系统、后处理系统以及比特流

处理系统(有时称为比特流处理工具)。

[0025] 贯穿包括权利要求在内的本公开内容,(编码音频比特流的)“元数据”的表达指代与比特流的相应的音频数据分离的且不同的数据。

[0026] 贯穿包括权利要求在内的本公开内容,“子流结构元数据”(或“SSM”)的表达表示编码音频比特流(或编码音频比特流集)的元数据,其指示编码比特流的音频内容的子流结构。

[0027] 贯穿包括权利要求在内的本公开内容,“节目信息元数据”(或“PIM”)的表达表示编码音频比特流的元数据,该编码音频比特流指示至少一个音频节目(例如,两个或更多个音频节目),其中所述元数据指示至少一个所述节目的音频内容的至少一个属性或特性(例如,指示对节目的音频数据执行的处理的类型或参数的元数据、或表示节目的哪些通道是活动通道的元数据)。

[0028] 贯穿包括权利要求在内的本公开内容,“处理状态元数据”的表达(例如,如在“响度处理状态元数据”的表达中)指代与比特流的音频数据相关联的(编码音频比特流的)元数据,指示相应的(相关联的)音频数据的处理状态(例如,已经对音频数据执行了什么类型的处理),并且通常还指示音频数据的至少一个特征或特性。处理状态元数据与音频数据的关联是时间同步的。从而,当前的(最新接收或更新的)处理状态元数据指示相应的音频数据同时包括所指示的类型的音频数据处理的结果。在一些情况下,处理状态元数据可以包括处理历史和/或用于所指示的类型的处理中的和/或从所指示的类型的处理中得到的参数中的一些或全部。另外,处理状态元数据可以包括相应的音频数据的已经从音频数据中计算或提取的至少一个特征或特性。处理状态元数据还可以包括与相应的音频数据的任何处理无关的或不是从相应的音频数据的任何处理中得到的其他元数据。例如,第三方数据、跟踪信息、标识符、所有权或标准信息、用户注释数据、用户偏好数据等可以通过具体的音频处理单元被添加以传递至其他音频处理单元。

[0029] 贯穿包括权利要求在内的本公开内容,“响度处理状态元数据”(或“LPSM”)的表达表示处理状态元数据,处理状态元数据指示相应的音频数据的响度处理状态(例如,已经对音频数据执行了什么类型的响度处理),并且通常还指示相应的音频数据的至少一个特征或特性(例如,响度)。响度处理状态元数据可以包括不是(即,当单独考虑时)响度处理状态元数据的数据(例如,其他元数据)。

[0030] 贯穿包括权利要求在内的本公开内容,“通道”(或“音频通道”)的表达表示单通道音频信号。

[0031] 贯穿包括权利要求在内的本公开内容,“音频节目”的表达表示一个或更多个音频通道的集合以及可选地还表示相关联的元数据(例如,描述期望的空间音频表示的元数据、和/或PIM、和/或SSM、和/或LPSM、和/或节目边界元数据)。

[0032] 贯穿包括权利要求在内的本公开内容,“节目边界元数据”的表达表示编码音频比特流的元数据,其中编码音频比特流指示至少一个音频节目(例如,两个或更多个节目),并且节目边界元数据指示至少一个所述音频节目的至少一个边界(开始和/或结束)在比特流中的位置。例如,(指示音频节目的编码音频比特流的)节目边界元数据可以包括指示节目的开始的位置(例如,比特流的第“N”帧的开始,或比特流的第“N”帧的第“M”个样本位置)的元数据,以及指示节目的结束的位置(例如,比特流的第“J”帧的开始,或比特流的第“J”帧

的第“K”个样本位置)的额外元数据。

[0033] 贯穿包括权利要求在内的本公开内容,术语“耦接”或“被耦接”用于表示直接或间接连接。从而,如果第一设备耦接至第二设备,该连接可以是通过直接连接,或经由其他设备和连接的通过间接连接。

### 具体实施方式

[0034] 典型的音频数据流包括音频内容(例如,音频内容的一个或更多个通道)和指示音频内容的至少一个特性的元数据两者。例如,在AC-3比特流中,存在具体意在用于改变被传送到收听环境的节目的声音的若干音频元数据参数。元数据参数中的一个为DIALNORM参数,其意在指示音频节目中的对白的平均电平,并且用于确定音频回放信号电平。

[0035] 在包括一系列不同的音频节目段(每个具有不同的DIALNORM参数)的比特流的重放期间,AC-3解码器使用每个段的DIALNORM参数执行一种类型的响度处理,在该响度处理中AC-3解码器修改回放电平或响度,使得该系列段的对白的感知的响度处于一致的电平。一系列编码音频项目中的每个编码音频段(项目)将(通常)具有不同的DIALNORM参数,并且解码器将对项目中的每个项目的电平进行缩放,使得每个项目的对白的回放电平或响度相同或非常相似,尽管这会要求在回放期间对项目中的不同的项目应用不同量的增益。

[0036] DIALNORM通常由用户设置而不是自动生成的,然而如果用户没有设置值则存在默认的DIALNORM值。例如,内容创建者可以使用AC-3编码器外部的装置进行响度测量,然后将该结果(指示音频节目的口语对白的响度)传送到编码器以设置DIALNORM值。从而,依赖于内容创建者正确地设置DIALNORM参数。

[0037] 对于为什么AC-3比特流中的DIALNORM参数会是错误的,存在几个不同的原因。第一,如果DIALNORM值不是由内容创建者设置的,那么每个AC-3编码器具有在比特流的生成期间使用的默认的DIALNORM值。该默认值可能与音频的实际对白响度显著不同。第二,即使内容创建者测量响度并且相应地设置DIALNORM值,可能已经使用不符合推荐的AC-3响度测量方法的响度测量算法或计量器,产生不正确的DIALNORM值。第三,即使已经使用由内容创建者正确测量和设置的DIALNORM值创建了AC-3比特流,该AC-3比特流可能在比特流的传输和/或存储期间已经被改变成错误值。例如,这在使用错误的DIALNORM元数据信息解码、修改然后重新编码AC-3比特流的电视广播应用中并非是不常见的。从而,包括在AC-3比特流中的DIALNORM值可能是错误的或不准确的,因此可能对收听体验的质量有消极的影响。

[0038] 此外,DIALNORM参数不指示相应的音频数据的响度处理状态(例如,已经对音频数据执行了什么类型的响度处理)。响度处理状态元数据(以其在本发明的一些实施方式中被提供的格式)有助于以尤其高效的方式便于音频比特流的自适应响度处理和/或音频内容的响度处理状态和响度的有效性的验证。

[0039] 尽管本发明不限于使用AC-3比特流、E-AC-3比特流或杜比E比特流,为了方便,将在生成、解码或以其他方式处理这样的比特流的实施方式中对其进行描述。

[0040] AC-3编码比特流包括元数据和音频内容的1至6个通道。音频内容是已经使用感知音频编码压缩的音频数据。元数据包括意在用于改变被传送到收听环境的节目的声音的若干音频元数据参数。

[0041] AC-3编码音频比特流的每帧包含关于数字音频的1536个样本的音频内容和元数

据。对于48kHz的采样率,这表示32毫秒的数字音频或音频的每秒31.25帧的速率。

[0042] 取决于帧是否分别包含1块、2块、3块或6块音频数据,E-AC-3编码音频比特流的每帧包含关于数字音频的256、512、768或1536个样本的音频数据和元数据。对于48kHz的采样率,这分别表示5.333、10.667、16或32毫秒的数字音频或分别表示音频的每秒189.9、93.75、62.5或31.25帧的速率。

[0043] 如图4所示,每个AC-3帧被划分成部分(段),包括:包含(如图5所示)同步字(SW)和两个误差校正字中的第一个误差校正字(CRC1)的同步信息(SI)部分;包含大部分元数据的比特流信息(BSI)部分;包含数据压缩音频内容(以及还可以包括元数据)的6个音频块(AB0至AB5);包含在压缩音频内容之后剩余的任意未使用的位的无用位段(W)(也称为“跳过字段”);可以包含更多元数据的辅助(AUX)信息部分;以及两个误差校正字中的第二个误差校正字(CRC2)。

[0044] 如图7所示,每个E-AC-3帧被划分成部分(段),包括:包含(如图5所示)同步字(SW)的同步信息(SI)部分;包含大部分元数据的比特流信息(BSI)部分;包含数据压缩音频内容(以及还可以包括元数据)的6个音频块(AB0至AB5);包含在压缩音频内容之后剩余的任意未使用的位的无用位段(W)(也称为“跳过字段”)(尽管仅示出了一个无用位段,不同的无用位段或跳过字段段通常可以在每个音频块之后);可以包含更多元数据的辅助(AUX)信息部分;以及误差校正字(CRC)。

[0045] 在AC-3(或E-AC-3)比特流中,存在具体意在用于改变被传送至收听环境的节目的声音的若干音频元数据参数。元数据参数中的一个为DIALNORM参数,该DIALNORM参数被包括在BSI段中。

[0046] 如图6所示,AC-3帧的BSI段包括指示节目的DIALNORM值的5位参数(“DIALNORM”)。如果AC-3帧的音频编码模式(“acmod”)为0,则包括指示在同一AC-3帧中携带的第二音频节目的5位参数DIALNORM值的5位参数(“DIALNORM2”),指示使用双单通道或“1+1”通道配置。

[0047] BSI段还包括指示在“addbsie”位之后额外的比特流信息的存在(或不存在的)的标志(“addbsie”)、指示在“addbsil”值之后任何额外的比特流信息的长度的参数(“addbsil”)、以及在“addbsil”值之后高达64位的额外的比特流信息(“addbsi”)。

[0048] BSI段包括在图6中没有具体示出的其他元数据值。

[0049] 根据一类实施方式,编码比特流指示音频内容的多个子流。在一些情况下,子流指示多通道节目的音频内容,并且子流中的每个指示节目的通道中的一个或更多个。在其他情况下,编码音频比特流的多个子流指示若干音频节目——通常为“主”音频节目(可以是多通道节目)和至少一个其他音频节目(例如,为关于主音频节目的评论的节目)——的音频内容。

[0050] 指示至少一个音频节目的编码音频比特流需要包括音频内容的至少一个“独立”子流。独立子流指示音频节目的至少一个通道(例如,独立子流可以指示常规的5.1通道音频节目的5个全音域通道)。在本文中,该音频节目称为“主”节目。

[0051] 在一些类型的实施方式中,编码音频比特流指示两个或更多个音频节目(“主”节目和至少一个其他音频节目)。在这样的情况下,比特流包括两个或更多个独立子流:指示主节目的至少一个通道的第一独立子流;以及指示另一音频节目(与主节目不同的节目)的至少一个通道的至少一个其他独立子流。每个独立子流可以独立地被解码,并且解码器可

以操作以仅对编码比特流的独立子流的子集(不是全部)进行解码。

[0052] 在指示两个独立子流的编码音频比特流的典型示例中,独立子流中的一个指示多通道主节目的标准格式扬声器通道(例如,5.1通道主节目的左、右、中、左环绕、右环绕全音域扬声器通道),而另一独立子流指示关于主节目的单通道音频评论(例如,导演关于电影的评论,其中主节目是电影的声带(soundtrack))。在指示多个独立子流的编码音频比特流的另一示例中,独立子流中的一个指示包括第一语言的对白的多通道主节目(例如,5.1通道主节目的标准格式扬声器通道(例如,主节目的扬声器通道中的一个可以指示对白),而每个其他独立子流指示对白的单通道翻译(翻译成不同的语言))。

[0053] 可选地,指示主节目(可选地还指示至少一个其他音频节目)的编码音频比特流包括音频内容的至少一个“从属”子流。每个从属子流与比特流的一个独立子流相关联,并且指示其内容由相关联的独立子流指示的节目(例如,主节目)的至少一个额外的通道(即,从属子流指示节目的不是由相关联的独立子流指示的至少一个通道,而相关联的独立子流指示节目的至少一个通道)。

[0054] 在包括独立子流(指示主节目的至少一个通道)的编码比特流的示例中,比特流还包括指示主节目的一个或更多个额外的扬声器通道的(与独立子流相关联的)从属子流。这样的额外的扬声器通道对由独立子流指示的主节目通道来说是额外的。例如,如果独立子流指示7.1通道主节目的左、右、中、左环绕、右环绕全音域扬声器通道,那么从属子流可以指示主节目的其他两个全音域扬声器通道。

[0055] 根据E-AC-3标准,E-AC-3比特流必须指示至少一个独立子流(例如,单个AC-3比特流),并且可以指示高达8个独立子流。E-AC-3比特流的每个独立子流可以与高达8个从属子流相关联。

[0056] E-AC-3比特流包括指示比特流的子流结构的元数据。例如,E-AC-3比特流的比特流信息(BSI)部分中的“chanmap”字段确定由比特流的从属子流指示的节目通道的通道映射。然而,指示子流结构的元数据常规地以如下格式包括在E-AC-3比特流中:该格式使得便于仅由E-AC-3解码器访问和使用(在编码E-AC-3比特流的解码期间);不便于在解码之后(例如,由后处理器)或解码之前(例如,由被配置成识别元数据的处理器)访问和使用。而且,存在以下风险:解码器可能使用常规地包括的元数据错误地识别常规的E-AC-3编码比特流的子流,并且在本发明之前还不知道如何以这样的格式在编码比特流(例如,编码E-AC-3比特流)中包括子流结构元数据,使得允许在比特流的解码期间方便和高效的检测和校正子流识别中的误差。

[0057] E-AC-3比特流还可以包括关于音频节目的音频内容的元数据。例如,指示音频节目的E-AC-3比特流包括指示已经使用谱扩展处理(以及通道耦合编码)以对节目的内容进行编码的最小频率和最大频率的元数据。然而,这样的元数据通常以如下格式包括在E-AC-3比特流中,该格式使得便于仅由E-AC-3解码器访问和使用(在编码E-AC-3比特流的解码期间);不便于在解码之后(例如,由后处理器)或解码之前(例如,由被配置成识别元数据的处理器)访问和使用。而且,这样的元数据不以如下的格式包括在E-AC-3比特流中,该格式允许在比特流的解码期间这样的元数据的识别的方便和高效的误差检测和误差校正。

[0058] 根据本发明的典型的实施方式,PIM和/或SSM(以及可选地还有其他元数据,例如,响度处理状态元数据或“LPSM”)被嵌入在音频比特流的元数据段的一个或更多个保留字段

(或槽(slot))中,该音频比特流还包括其他段(音频数据段)中的音频数据。通常,比特流的每个帧的至少一个段包括PIM或SSM,并且帧的至少一个其他段包括相应的音频数据(即,其数据结构由SSM指示的和/或其至少一个特性或属性由PIM指示的音频数据)。

[0059] 在一类实施方式中,每个元数据段为可以包含一个或多个元数据有效载荷的数据结构(在本文中有时称为容器)。每个有效载荷包括报头以提供存在于有效载荷中的元数据的类型的明确的指示,其中报头包括具体的有效载荷标识符(或有效载荷配置数据)。有效载荷在容器内的顺序未被定义,使得有效载荷可以以任何顺序存储并且分析器必须能够对整个容器进行分析以提取相关的有效载荷而忽略不相关的或不支持的有效载荷。图8(下面将要描述的)说明这样的容器和容器内的有效载荷的结构。

[0060] 当两个或多个音频处理单元需要贯穿该处理链(或内容生命周期)彼此合作工作时,音频数据处理链中的通信元数据(例如,SSM和/或PIM和/或LPSM)尤其有用。在音频比特流中不包括元数据的情况下,例如,当在链中利用两个或多个音频编解码器并且在媒体消耗装置的比特流路径(或比特流的音频内容的渲染点)期间多于一次地应用单端音量时,可以出现若干媒体处理问题,例如质量、电平和空间退化。

[0061] 根据本发明的一些实施方式,嵌入在音频比特流中的响度处理状态元数据(LPSM)可以被认证和验证,例如以使得响度调整实体能够证明特定节目的响度是否已经在指定的范围内以及相应的音频数据本身是否未被修改(由此确保符合可适用的调节)。包括在包括响度处理状态元数据的数据块中的响度值可以被读出以对此进行验证,而不再次计算响度。响应于LPSM,管理结构可以确定相应的音频内容符合(如由LPSM指示的)响度法定的和/或管理的要求(例如,在商业广告响度缓解法下公布的规则,也称为“CALM”法)而不需要计算音频内容的响度。

[0062] 图1为示例性音频处理链(音频数据处理系统)的框图,在音频处理链中,系统的元件中的一个或多个可以根据本发明的实施方式被配置。系统包括如所示耦接在一起的以下元件:预处理单元、编码器、信号分析和元数据校正单元、代码转换器、解码器和后处理单元。在所示的系统的变型中,省略元件中的一个或多个,或包括额外的音频数据处理单元。

[0063] 在一些实现中,图1的预处理单元被配置成接收包括音频内容的PCM(时域)样本作为输入,并且输出经处理PCM样本。编码器可以被配置成接收PCM样本作为输入,并且输出指示音频内容的编码的(例如,压缩的)音频比特流。指示音频内容的比特流的数据在本文中有时被称为“音频数据”。如果编码器根据本发明的典型实施方式被配置,那么从编码器输出的音频比特流包括PIM和/或SSM(可选地还包括响度处理状态元数据和/或其他元数据)以及音频数据。

[0064] 图1的信号分析和元数据校正单元可以接收一个或多个编码音频比特流作为输入,并且通过执行信号分析(例如,使用编码音频比特流中的节目边界元数据)来确定(例如,验证)每个编码音频比特流中的元数据(例如,处理状态元数据)是否正确。如果信号分析和元数据校正单元发现所包括的元数据是无效的,那么通常使用从信号分析中获得的正确值替代错误值。从而,从信号分析和元数据校正单元输出的每个编码音频比特流可以包括校正的(或未校正的)处理状态元数据以及编码音频数据。

[0065] 图1的代码转换器可以接收编码音频比特流作为输入,并且作为响应(例如,通过

对输入流进行解码并且以不同的编码格式对解码流进行重新编码) 输出修改的(例如,不同编码的) 音频比特流。如果代码转换器根据本发明的典型的实施方式被配置,那么从代码转换器输出的音频比特流包括SSM和/或PIM(通常还包括其他元数据) 以及编码音频数据。元数据可以已经被包括在输入比特流中。

[0066] 图1的解码器可以接收编码的(例如,压缩的) 音频比特流作为输入,并且输出(作为响应) 解码PCM音频样本流。如果解码器根据本发明的典型的实施方式被配置,那么在典型的操作中,解码器的输出是或包括下列中的任一个:

[0067] 音频样本流,以及从输入的编码比特流中提取的SSM和/或PIM(通常还有其他元数据) 的至少一个相应的流;或

[0068] 音频样本流,以及根据从输入编码比特流中提取的SSM和/或PIM(通常还有其他元数据,例如LPSM) 所确定的控制位的相应的流;或

[0069] 音频样本流,但没有元数据或根据元数据确定的控制位的相应的流。在最后一种情况下,解码器可以从输入编码比特流中提取元数据,并且对所提取的元数据执行至少一种操作(例如,验证),即使没有输出所提取的元数据或根据元数据确定的控制位。

[0070] 通过根据本发明的典型的实施方式配置图1的后处理单元,后处理单元被配置成接收解码的PCM音频样本流,并且使用与样本一起接收的SSM和/或PIM(通常还有其他元数据,例如LPSM),或根据与样本一起接收的元数据确定的控制位对其执行后处理(例如,音频内容的音量校平)。后处理单元还通常被配置成对经后处理音频内容进行渲染用于由一个或多个扬声器回放。

[0071] 本发明的典型的实施方式提供增强的音频处理链,其中音频处理单元(例如,编码器、解码器、代码转换器以及预处理单元和后处理单元) 根据由通过音频处理单元分别接收的元数据所指示的媒体数据的同时期的状态来修改待应用于音频数据的其相应的处理。

[0072] 输入到图1系统的任何音频处理单元(例如,图1的编码器或代码转换器) 的音频数据可以包括SSM和/或PIM(可选地还包括其他元数据) 以及音频数据(例如,编码音频数据)。该元数据可以根据本发明的实施方式已经通过图1系统的另一元件(或另一源,在图1中未示出) 而被包括在输入音频中。接收输入音频(具有元数据) 的处理单元可以被配置成对元数据执行至少一种操作(例如,验证),或响应于元数据(例如,输入音频的自适应处理),并且还通常将元数据、元数据的经处理的版本、或根据元数据确定的控制位包括在其输出音频中。

[0073] 本发明的音频处理单元(或音频处理器) 的典型的实施方式被配置成基于由对应于音频数据的元数据所指示的音频数据的状态来执行音频数据的自适应处理。在一些实施方式中,自适应处理是(或包括) 响度处理(如果元数据指示还未对音频数据执行响度处理或与响度处理类似的处理),而不是(且不包括) 响度处理(如果元数据指示已经对音频数据执行了这样的响度处理或与响度处理类似的处理)。在一些实施方式中,自适应处理是或包括(例如,在元数据验证子单元中执行的) 元数据验证以确保音频处理单元基于由元数据所指示的音频数据的状态来执行音频数据的其他自适应处理。在一些实施方式中,该验证确定与音频数据相关联(例如,包括在具有音频数据的比特流中) 的元数据的可靠性。例如,如果验证元数据是可靠的,那么来自一种先前执行的音频处理的结果可以被重新使用并且可以避免新执行相同类型的音频处理。另一方面,如果发现元数据已经被篡改(或以其他方式

不可靠),那么据称先前执行的一种类型的媒体处理(如由不可靠的元数据指示的)可以由音频处理单元重复,和/或可以由音频处理单元对元数据和/或音频数据执行其他处理。如果该单元确定元数据是有效的(例如,基于所提取的加密值与参考加密值的匹配),音频处理单元还可以被配置成用信号向增强的媒体处理链下游的其他音频处理单元通知元数据(例如,存在于媒体比特流中)是有效的。

[0074] 图2是作为本发明的音频处理单元的実施方式的编码器(100)的框图。编码器100的任何部件或元件可以以硬件或软件或硬件与软件的组合被实现为一个或更多个处理和/或一个或更多个电路(例如,ASIC、FPGA或其他集成电路)。编码器100包括如所示地连接的帧缓冲器110、分析器111、解码器101、音频状态验证器102、响度处理级103、音频流选择级104、编码器105、填充器/格式器级107、元数据生成级106、对白响度测量子系统108以及帧缓冲器109。编码器100通常还包括其他处理元件(未示出)。

[0075] 编码器100(为代码转换器)被配置成包括通过使用包括在输入比特流中的响度处理状态元数据执行自适应和自动的响度处理来将输入音频比特流(例如,可以是AC-3比特流、E-AC-3比特流或杜比E比特流中的一个)转换成编码输出音频比特流(例如,可以是AC-3比特流、E-AC-3比特流或杜比E比特流中的另一个)。例如,编码器100可以被配置成将(通常用在生产和广播设备中,但不用在接收已经被广播的音频节目的消费者设备中的格式的)输入杜比E比特流转换成AC-3或E-AC-3格式的(适合于广播至消费者设备的)编码输出音频比特流。

[0076] 图2的系统还包括编码音频传送子系统150(其存储和/或传送从编码器100输出的编码比特流)和解码器152。从编码器100输出的编码音频比特流可以由子系统150(例如,以DVD或蓝光光盘格式)存储,或由子系统150(可以实现传输线路或网络)传输,或可以由子系统150存储和传输。解码器152被配置成包括通过从比特流的每个帧中提取元数据(PIM和/或SSM、以及可选地还有响度处理状态元数据和/或其他元数据)(以及可选地还从比特流中提取节目边界元数据)以及生成解码音频数据,对经由子系统150接收的(由编码器100生成的)编码音频比特流进行解码。通常,解码器152被配置成使用PIM和/或SSM和/或LPSM(可选地还使用节目边界元数据)对解码音频数据执行自适应处理,和/或将解码音频数据和元数据转发至被配置成使用元数据对解码音频数据执行自适应处理的后处理器。通常,解码器152包括存储(例如,以非暂态方式)从子系统150中接收的编码音频比特流的缓冲器。

[0077] 编码器100和解码器152的各种实现被配置成执行本发明的方法的不同的实施方式。

[0078] 帧缓冲器110是耦接以接收编码输入音频比特流的缓冲存储器。在操作中,缓冲器110存储(例如,以非暂态方式)编码音频比特流的至少一个帧,并且编码音频比特流的帧的序列被从缓冲器110设定到分析器111。

[0079] 将分析器111耦接并配置成从包括这样的元数据的编码输入音频的每个帧中提取PIM和/或SSM、以及响度处理状态元数据(LPSM)、以及可选地还有节目边界元数据(和/或其他元数据),至少将LPSM(以及可选地还有节目边界元数据和/或其他元数据)设定到音频状态验证器102、响度处理级103、级106和子系统108,以从编码输入音频中提取音频数据并且将音频数据设定到解码器101。编码器100的解码器101被配置成对音频数据进行解码以生成解码音频数据,并且将解码音频数据设定到响度处理级103、音频流选择级104、子系统

108以及通常还设定到状态验证器102。

[0080] 状态验证器102被配置成对设定到其的LPSM(可选地其他元数据)进行认证和验证。在一些实施方式中,LPSM是(或包括在)数据块(中),数据块已经包括在输入比特流中(例如,根据本发明的实施方式)。块可以包括加密散列(基于散列的消息认证代码或“HMAC”)用于对LPSM(可选地还有其他元数据)和/或(从解码器101提供至验证器102的)基本的音频数据进行处理。在这些实施方式中,数据块可以被数字地标记,使得下游的音频处理单元可以相对容易地认证和验证处理状态元数据。

[0081] 例如,HMAC用于生成摘要,并且包括在本发明的比特流中的保护值可以包括该摘要。该摘要可以关于AC-3帧被如下生成:

[0082] 1.在AC-3数据和LPSM被编码之后,帧数据字节(连接的帧数据#1和帧数据#2)和LPSM数据字节用作哈希函数HMAC的输入。没有考虑可以存在于辅助数据字段内的其他数据用于计算摘要。这样的其他数据可以是既不属于AC-3数据也不属于LPSM数据的字节。可以不考虑包括在LPSM中的保护位用于计算HMAC摘要。

[0083] 2.在计算摘要之后,被写入比特流中的为保护位保留的字段中。

[0084] 3.生成完整的AC-3帧的最后步骤是CRC校验的计算。这被写在帧的结束处并且考虑属于该帧的所有数据,包括LPSM位。

[0085] 包括但不限于一个或更多个非HMAC加密方法中的任意一个的其他加密方法可以用于LPSM和/或其他元数据(例如,在验证器102中)的验证,以确保元数据和/或基本音频数据的安全的传输和接收。例如,可以在接收本发明的音频比特流的实施方式的每个音频处理单元中执行验证(使用这样的加密方法),以确定包括在该比特流中的元数据和相应的音频数据是否已经经历(和/或已经产生)具体的处理(由元数据指示的)并且在这样的具体的处理执行之后是否未被修改。

[0086] 状态验证器102将控制数据设定到音频流选择级104、元数据生成器106以及对白响度测量子系统108,以表示验证操作的结果。响应于控制数据,级104可以选择(以及传递至编码器105):

[0087] 响度处理级103的经自适应处理的输出(例如,当LPSM指示从解码器101输出的音频数据没有经历特定类型的响度处理,以及来自验证器102的控制位指示LPSM有效时);或

[0088] 从解码器102输出的音频数据(例如,当LPSM指示从解码器101输出的音频数据已经经历将由级103执行的特定类型的响度处理,并且来自验证器102的控制位指示LPSM有效时)。

[0089] 编码器100的级103被配置成基于由通过解码器101所提取的LPSM指示的一个或更多个音频数据特性,对从解码器101输出的解码音频数据执行自适应响度处理。级103可以是自适应变换域实时响度和动态范围控制处理器。级103可以接收用户输入(例如,用户目标响度/动态范围值或对白归一化值)、或其他元数据输入(例如,一种或更多种类型的第三方数据、跟踪信息、标识符、所有权或标准信息、用户注释数据、用户偏好数据等)和/或其他输入(例如,来自指纹识别处理),并且使用这样的输入以对从解码器101输出的解码音频数据进行处理。级103可以对指示(由通过分析器111提取的节目边界元数据所表示的)单个音频节目的(从解码器101输出的)解码音频数据执行自适应响度处理,并且可以响应于接收到指示由通过分析器111提取的节目边界元数据所指示的不同的音频节目的(从解码器101

输出的) 解码音频数据将响度处理复位。

[0090] 当来自验证器102的控制位指示LPSM无效时,对白响度测量子系统108可以操作以使用由解码器101提取的LPSM(和/或其他元数据)来确定表示对白(或其他语音)的(来自解码器101的)解码音频的段的响度。当来自验证器102的控制位指示LPSM有效时,当LPSM指示(来自解码器101的)解码音频的对白(或其他语音)段的先前确定的响度时,可以禁止对白响度测量子系统108的操作。子系统108可以对表示(通过分析器111所提取的节目边界元数据所指示的)单个音频节目的解码音频数据执行响度测量,并且可以响应于接收到表示由这样的节目边界元数据所指示的不同的音频节目的解码音频数据将响度处理复位。

[0091] 存在有用的工具(例如,杜比LM100响度计)用于方便地和容易地对音频内容中的对白的电平进行测量。本发明的APU(例如,编码器100的级108)的一些实施方式被实现以包括这样的工具(或执行这样的工具的功能)来对音频比特流(例如,从编码器100的解码器101设定到级108的解码AC-3比特流)的音频内容的平均对白响度进行测量。

[0092] 如果级108被实现成对音频数据的真实平均对白响度进行测量,那么测量可以包括将主要包含语音的音频内容的段分离的步骤。然后,根据响度测量算法来处理主要为语音的音频段。对于根据AC-3比特流解码的音频数据,该算法可以是标准的K加权响度测量(根据国际标准ITU-R BS 1770)。可替代地,可以使用其他响度测量(例如,基于响度的心理声学模型的那些测量)。

[0093] 语音段的分离不是测量音频数据的平均对白响度所必需的。然而,它提高测量的准确度,并且通常提供来自听者感知的较满意的结果。因为不是所有的音频内容包含对白(语音),整个音频内容的响度测量可以提供语音已经存在的音频的对白电平的足够的近似。

[0094] 元数据生成器106生成(和/或传递至级107)要由级107包括在待从编码器100输出的编码比特流中。元数据生成器106可以将由编码器101和/或分析器111提取的LPSM(可选地还有LIM和/或PIM和/或节目边界元数据和/或其他元数据)传递至级107(例如,当来自验证器102的控制位指示LPSM和/或其他元数据有效时),或生成新的LIM和/或PIM和/或LPSM和/或节目边界元数据和/或其他元数据并且将新的元数据设定到级107(例如,当来自验证器102的控制位指示由解码器101提取的元数据无效时),或可以将由解码器101和/或分析器111提取的元数据与新生成的元数据的组合设定到级107。元数据生成器106可以将由子系统108生成的响度数据以及指示由子系统108执行的响度处理的类型的至少一个值包括在LPSM中,将LPSM设定到级107以用于包括在待从编码器100输出的编码比特流中。

[0095] 元数据生成器106可以生成用于待被包括在编码比特流和/或待被包括在编码比特流中的基本音频数据中的LPSM(可选地还有其他元数据)的解密、认证或验证中的至少一个的控制位(可以由基于散列的消息认证代码或“HMAC”组成或包括基于散列的消息认证代码或“HMAC”)。元数据生成器106可以向级107提供这样的保护位以用于包括在编码比特流中。

[0096] 在典型的操作中,对白响度测量子系统108对从解码器101输出的音频数据进行处理以响应于音频数据生成响度值(例如,选通的和未选通的对白响度值)和动态范围值。响应于这些值,元数据生成器106可以生成响度处理状态元数据(LPSM)以用于(由填充器/格式器107)包括在待从编码器100输出的编码比特流中。

[0097] 另外,可选地,或可替代地,编码器100的子系统106和/或108可以执行音频数据的额外的分析以生成指示音频数据的至少一个特性的元数据以用于包括在待从级107输出的编码比特流中。

[0098] 编码器105对从选择级104输出的音频数据进行编码(例如,通过对其执行压缩),并且将编码的音频设定到级107以用于包括在待从级107输出的编码比特流中。

[0099] 级107将来自编码器105的编码音频和来自生成器106的元数据(包括PIM和/或SSM)进行复用以生成待从级107中输出的编码比特流,优选地使得编码比特流具有由本发明的优选实施方式指定的格式。

[0100] 帧缓冲器109为存储(例如,以非暂态方式)从级107输出的编码音频比特流的至少一个帧的缓冲存储器,然后编码音频比特流的一系列帧被从缓冲器109作为来自编码器100的输出设定至传送系统150。

[0101] 由元数据生成器106生成并且由级107包括在编码比特流中的LPSM通常指示相应音频数据的响度处理状态(例如,已经对音频数据执行什么类型的响度处理)以及相应音频数据的响度(例如,测量的对白响度、选通和/或未选通的响度、和/或动态范围)。

[0102] 在本文中,对音频数据执行的响度和/或电平测量的“选通”是指超过阈值的计算值被包括在最终测量(例如,在最终测量的值中忽略低于-60dBFS的短期响度值)中的特定电平或响度阈值。绝对值的选通是指固定的电平或响度,而相对值的选通是指依赖于当前“未选通的”测量值的值。

[0103] 在编码器100的一些实现中,缓存在存储器109(以及输出至传送系统150)的编码比特流为AC-3比特流或E-AC-3比特流,并且包括音频数据段(例如,图4中所示的帧的AB0至AB5段)和元数据段,其中音频数据段指示音频数据,并且元数据段中的至少一些中的每个包括PIM和/或SSM(以及可选地其他元数据)。级107将元数据段(包括元数据)插入到下面的格式的比特流中。包括PIM和/或SSM的元数据段中的每个元数据段被包括在比特流的无用位段(例如,图4或图7中所示的无用位段“W”)中,或比特流的帧的比特流信息(“BSI”)段的“addbsi”字段中,或比特流的帧的结束处的辅助数据字段(例如,图4或图7中所示的AUX段)。比特流的帧可以包括一个或两个元数据段,每个元数据段包括元数据,并且如果帧包括两个元数据段,一个可以存在于帧的addbsi字段中而另一个存在于帧的AUX字段中。

[0104] 在一些实施方式中,由级107插入的每个元数据段(在本文中有时称为“容器”)具有包括元数据段报头(可选地还包括其他强制的或“核心”元素)以及在元数据段报头之后的一个或更多个元数据有效载荷的格式。如果存在,SIM被包括在元数据有效载荷中的一个有效载荷(由有效载荷报头标识,并且通常具有第一类型的格式)中。如果存在,PIM被包括在元数据有效载荷中的另一个有效载荷(由有效载荷报头标识,并且通常具有第二类型的格式)中。类似地,元数据的每个其他类型(如果存在)被包括在元数据有效载荷中的另一有效载荷(由有效载荷报头标识,并且通常具有针对元数据的类型的格式)中。示例性格式使得能够在除了解码期间之外的时间便于访问(例如,由解码之后的后处理器、或由被配置成在没有对编码比特流执行完全解码的情况下识别元数据的处理器)SSM、PIM和其他元数据,并且允许在比特流的解码期间(例如,子流识别的)方便和高效的误差检测和校正。例如,在不以示例性格式访问SSM的情况下,解码器可能错误地识别与节目相关联的子流的数量。元数据段中的一个元数据有效载荷可以包括SSM,元数据段中的另一个元数据有效载荷

可以包括PIM,以及可选地,元数据段中的至少一个其他元数据有效载荷可以包括其他元数据(例如,响度处理状态元数据或“LPSM”)。

[0105] 在一些实施方式中,(由级107)包括在编码比特流(例如,指示至少一个音频节目的E-AC-3比特流)的帧中的子流结构元数据(SSM)有效载荷包括下面的格式的SSM:

[0106] 有效载荷报头,通常包括至少一个识别值(例如,指示SSM格式版本的2位值,以及可选地长度、周期、计数和子流相关联值);以及在报头之后:

[0107] 指示由比特流指示的节目的独立子流的数量独立子流元数据;以及

[0108] 从属子流元数据,其指示:节目的每个独立子流是否具有至少一个相关联的从属子流(即,至少一个从属子流是否与所述每个独立子流相关联),以及如果是这样,与节目的每个独立子流相关联的从属子流的数量。

[0109] 预期的是,编码比特流的独立子流可以指示音频节目的扬声器通道集(例如,5.1扬声器通道音频节目的扬声器通道),以及一个或多个从属子流中的每个(与独立子流相关联,由从属子流元数据指示)可以指示节目的目标通道。然而,编码比特流的独立比特流通常指示节目的扬声器通道集,并且与独立子流相关联的每个从属子流(由从属子流元数据指示)指示节目的至少一个额外的扬声器通道。

[0110] 在一些实施方式中,(由级107)包括在编码比特流(例如,指示至少一个音频节目的E-AC-3比特流)的帧中的节目信息元数据(PIM)有效载荷具有下面的格式:

[0111] 有效载荷报头,通常包括至少一个标识值(例如,指示PIM格式版本的值,以及可选地长度、周期、计数和子流相关联值);以及在报头之后的下面格式的PIM:

[0112] 指示音频节目的每个静音通道和每个非静音通道(即,节目的哪些通道包含音频信息,而哪些通道(如果有)仅包含静音(通常关于帧的持续时间))的活动通道元数据。在编码比特流是AC-3或E-AC-3比特流的实施方式中,比特流的帧中的活动通道元数据可以结合比特流的额外的元数据(例如,帧的音频编码模式(“acmod”)字段,以及,如果存在,帧或相关联的从属子流帧中的chanmap字段)以确定节目的哪些通道包含音频信息而哪些通道包含静音。AC-3或E-AC-3帧的“acmod”字段指示由帧的音频内容指示的音频节目的全音域通道的数量(例如,节目是1.0通道单通道节目、2.0通道立体声节目、还是包括L、R、C、Ls、Rs全音域通道的节目),或者帧指示两个独立的1.0通道单通道节目。E-AC-3比特流的“chanmap”字段指示由比特流指示的从属子流的通道映射。活动通道元数据可以有助于实现解码器的上混合(在后处理器中)下游,例如以在解码器的输出处将音频添加至包含静音的通道;

[0113] 指示节目是否被下混合(在编码之前或在编码期间)以及如果节目被下混合则被应用的下混合的类型的下混合处理状态元数据。下混合处理状态元数据可以有助于实现解码器的上混合(在后处理器中)下游,例如以使用最匹配被应用的下混合的类型的参数对节目的音频内容进行上混合。在编码比特流是AC-3或E-AC-3比特流的实施方式中,下混合处理状态元数据可以结合帧的音频编码模型(“acmod”)字段以确定应用于节目的通道的下混合(如果有)的类型;

[0114] 指示在编码之前或在编码期间节目是否被上混合(例如,从较小数量的通道)以及如果节目被上混合则所应用的上混合的类型的上混合处理状态元数据。上混合处理状态元数据可以有助于实现解码器的下混合(在后处理器中)下游,例如以与应用于节目的上混合(例如,杜比定向逻辑、或杜比定向逻辑 II 电影模式、或杜比定向逻辑 II 音乐模式、或杜比专

业上混合器)的类型一致的方式对节目的音频内容进行下混合。在编码比特流是E-AC-3比特流的实施方式中,上混合处理状态元数据可以结合其他元数据(例如,帧的“strmtyp”字段的值)以确定应用于节目的通道的上混合(如果有)的类型。(E-AC-3比特流的帧的BSI字段中的)“strmtyp”字段的值指示帧的音频内容是否属于独立流(其确定节目)或(包括多个子流或与多个子流相关联的节目的)独立子流,从而可以独立于由E-AC-3比特流指示的任何其他子流被编码,或帧的音频内容是否属于(包括多个子流或与多个子流相关联的节目的)从属子流,从而必须结合与其相关联的独立子流被解码;以及

[0115] 预处理状态元数据,其指示:是否对帧的音频内容执行了预处理(在生成编码比特流的音频内容的编码之前),以及如果对帧音频内容执行了预处理则被执行的预处理的类型。

[0116] 在一些实现中,预处理状态元数据指示:

[0117] 是否应用环绕衰减(例如,在编码之前,音频节目的环绕通道是否被衰减3dB),

[0118] 是否(例如,在编码之前,对音频节目的环绕通道Ls和Rs通道)应用90°相移,

[0119] 在编码之前,是否对音频节目的LFE通道应用低通滤波器,

[0120] 在生成期间,是否监视节目的LFE通道的电平以及如果监视了节目的LFE通道的电平则LFE通道的监视的电平相对于节目的全音域音频通道的电平,

[0121] 是否应当对节目的解码音频内容的每个块执行(例如,在解码器中)动态范围压缩以及如果应当对节目的解码音频内容的每个块执行动态范围压缩则待被执行的动态范围压缩的类型(和/或参数)(例如,该类型的预处理状态元数据可以指示以下压缩配置文件类型中的哪个由编码器假定以生成被包括在编码比特流中的动态范围压缩控制值:电影标准、电影光线、音乐标准、音乐光线或语音。或者,该类型的预处理状态元数据可以指示应当以由被包括在编码比特流中的动态范围压缩控制值确定的方式对节目的解码音频内容的每个帧执行重动态范围压缩(“compr”压缩)),

[0122] 是否使用谱扩展和/或通道耦合编码以对特定频率范围的节目内容进行编码,以及如果使用谱扩展和/或通道耦合编码以对特定频率范围的节目内容进行编码则对其执行谱扩展编码的内容的频率分量的最小频率和最大频率,以及对其执行通道耦合编码的内容的频率分量的最小频率和最大频率。该类型的预处理状态元数据信息可以有助于执行解码器的均衡(在后处理器中)下游。通道耦合信息和谱扩展信息两者都有助于在代码转换操作和应用期间优化质量。例如,编码器可以基于参数例如谱扩展和通道耦合信息的状态优化其行为(包括预处理步骤例如头戴式耳机虚拟、上混合等的自适应)。而且,编码器可以基于进入的(并且认证的)元数据的状态来动态地修改其耦合参数和谱扩展参数以匹配最佳值和/或将其耦合和谱扩展参数修改成最佳值,以及

[0123] 对白增强调整范围数据是否包括在编码比特流中,以及如果对白增强调整范围数据包括在编码比特流中,则在相对于音频节目中的非对白内容的电平调整对白内容的电平的对白增强处理(例如,在解码器的后处理器下游)的执行期间可得到的调整的范围。

[0124] 在一些实现中,额外的预处理状态元数据(例如,指示头戴式耳机相关的参数的元数据)被包括在(由级107)待从编码器100输出的编码比特流的PIM有效载荷中。

[0125] 在一些实现中,(由级107)包括在编码比特流(例如,指示至少一个音频节目的E-AC-3比特流)的帧中的LPSM有效载荷包括下面的格式的LPSM:

[0126] 报头(通常包括标识LPSM有效载荷的开始同步字,在同步字之后的至少一个标识值,例如,在下面的表2中表示的LPSM格式版本、长度、周期、计数和子流关联值);以及

[0127] 在报头之后的:

[0128] 指示相应音频数据指示对白或不指示对白(例如,相应音频数据的哪些通道指示对白)的至少一个对白指示值(例如,表2的参数“对白通道”);

[0129] 指示相应的音频内容是否符合响度调整的所指示的集合的至少一个响度调整符合值(例如,表2的参数“响度调整类型”);

[0130] 指示已经对相应音频数据执行的响度处理的至少一种类型的至少一个响度处理值(例如,表2的参数“对白选通响度校正标志”、“响度校正类型”中的一个或多个);以及

[0131] 指示相应音频数据的至少一个响度(例如,峰值或平均响度)特性的至少一个响度值(例如,表2的参数“ITU相对选通响度”、“ITU语音选通响度”、“ITU(EBU 3341)短期3s响度”和“真实峰值”中的一个或多个)。

[0132] 在一些实现中,包含PIM和/或SSM(以及可选地其他元数据)的每个元数据段包含元数据段报头(以及可选地额外的核心元素)、以及在元数据段报头(或元数据段报头和其他核心元素)之后的具有下面的格式的至少一个元数据有效载荷段:

[0133] 有效载荷报头,通常包括至少一个标识值(例如,SSM或PIM格式版本、长度、周期、计数和子流关联值),以及

[0134] 在有效载荷报头之后的SSM或PIM(或另一类型的元数据)。

[0135] 在一些实现中,由级107插入至比特流的帧的无用位段/跳过字段段(或“addbsi”字段或辅助数据字段)中的元数据段(在本文中有时称为“元数据容器”或“容器”)中的每个具有下面的格式:

[0136] 元数据段报头(通常包括标识元数据段的开始同步字,在同步字之后的标识值,例如,在下面的表1中表示的版本、长度、周期、扩展的元素计数和子流关联值);以及

[0137] 在元数据段报头之后的有助于元数据段或相应音频数据的元数据的至少一个的解密、认证或验证中的至少一种的至少一个保护值(例如表1的HMAC摘要和音频指纹值);以及

[0138] 也在元数据段报头之后的标识每个下面的元数据有效载荷中的元数据的类型并且指示每个这样的有效载荷的配置(例如,尺寸)的至少一个方面的元数据有效载荷标识(“ID”)值和有效载荷配置值。

[0139] 每个元数据有效载荷在相应有效载荷ID值和有效载荷配置值之后。

[0140] 在一些实施方式中,在帧的无用位段(或辅助数据字段或“addbsi”字段)中的元数据段中的每个具有三种等级的结构:

[0141] 高等级结构(例如,元数据段报头),包括指示无用位(或辅助数据或addbsi)字段是否包括元数据的标志、指示存在什么类型的元数据的至少一个ID值、以及通常还有指示(例如,每个类型的)元数据的多少位存在(如果元数据存在的话)的值。可以存在的元数据的一种类型为PIM,可以存在的元数据的另一类型为SSM,而可以存在的元数据的其他类型为LPSM、和/或节目边界元数据、和/或媒体搜索元数据;

[0142] 中间等级结构,包括与每个所标识的类型的元数据相关联的数据(例如,元数据有效载荷报头、保护值、以及关于每个所标识的类型的元数据的有效载荷ID值和有效载荷配

置值);以及

[0143] 低等级结构,包括关于每个所标识的类型的元数据的元数据有效载荷(例如,如果PIM被识别为正存在,一系列PIM值,和/或如果该其他类型的元数据被识别为正存在,另一类型(例如,SSM或LPSM)的元数据值)。

[0144] 这样三个等级结构中的数据值可以被嵌套。例如,由高等级结构和中间等级结构标识的每个有效载荷(例如,每个PIM、或SSM或其他数据有效载荷)的保护值可以被包括在有效载荷之后(从而在有效载荷的元数据有效载荷报头之后),或由高等级结构和中间等级结构标识的所有元数据有效载荷的保护值可以被包括在元数据段中的最终元数据有效载荷之后(从而在元数据段的所有有效载荷的元数据有效载荷报头之后)。

[0145] 在(参照图8的元数据段或“容器”将要描述的)一个示例中,元数据段报头标识4个元数据有效载荷。如图8所示,元数据段报头包括容器同步字(被标识为“容器同步”)以及版本和键ID值。元数据段报头之后是4个元数据有效载荷和保护位。第一有效载荷(例如,PIM有效载荷)的有效载荷ID值和有效载荷配置(例如,有效载荷尺寸)值在元数据段报头之后,第一有效载荷本身在ID和配置值之后,第二有效载荷(例如,SSM有效载荷)的有效载荷ID值和有效载荷配置(例如,有效载荷尺寸)值在第一有效载荷之后,第二有效载荷本身在这些ID和配置值之后,第三有效载荷(例如,LPSM有效载荷)的有效载荷ID值和有效载荷配置(例如,有效载荷尺寸)值在第二有效载荷之后,第三有效载荷本身在这些ID和配置值之后,第四有效载荷的有效载荷ID值和有效载荷配置(例如,有效载荷尺寸)值在第三有效载荷之后,第四有效载荷本身在这些ID和配置值之后,而关于有效载荷中的全部或一些有效载荷(或关于高等级结构和中间等级结构以及有效载荷中的全部或一些有效载荷)的保护值(在图8中被标识为“保护数据”)在最后一个有效载荷之后。

[0146] 在一些实施方式中,如果解码器101接收根据本发明的实施方式生成的具有加密散列的音频比特流,则解码器被配置成根据由比特流确定的数据块对加密散列进行分析和检索,其中所述块包括元数据。验证器102可以使用加密散列对所接收的比特流和/或相关联的元数据进行验证。例如,如果验证器102基于参考加密散列与从数据块检索到的加密散列之间的匹配发现元数据是有效的,那么可以禁止处理器103对相应的音频数据的操作,并且使得选择级104通过(未改变的)音频数据。另外,可选地或可替代地,可以使用其他类型的加密技术替代基于加密散列的方法。

[0147] 图2的编码器100可以确定(响应于由解码器101提取的LPSM以及可选地还响应于节目边界元数据)后处理/预处理单元已经(在元件105、106和107中)对待编码的音频数据执行了一种类型的响度处理,因此可以(在生成器106中)创建包括用于先前执行的响度处理的和/或根据先前执行的响度处理得到的具体参数的响度处理状态元数据。在一些实现中,只要编码器知道已经对音频内容执行的处理的类型,编码器100就可以创建指示对音频内容的处理历史的元数据(以及将其包括在从编码器输出的编码比特流中)。

[0148] 图3是为本发明的音频处理单元的实施方式的解码器(200)以及耦接至解码器(200)的后处理器(300)的框图。后处理器(300)也是本发明的音频处理单元的实施方式。编码器200和后处理器300的部件或元件中的任何一个可以以硬件、软件或硬件和软件的组合被实现为一个或多个处理和/或一个或多个电路(例如,ASIC、FPGA或其他集成电路)。解码器200包括如所示地连接的帧缓冲器201、分析器205、音频解码器202、音频状态验证级

(验证器)203以及控制位生成器204。通常,解码器200还包括其他处理元件(未示出)。

[0149] 帧缓冲器201(缓冲存储器)存储(例如,以非暂态方式)由解码器200接收的编码音频比特流的至少一个帧。编码音频比特流的帧序列被从缓冲器201设定到分析器205。

[0150] 耦接分析器205并且将其配置成从编码输入音频的每个帧中提取PIM和/或SSM(可选地还提取其他元数据,例如,LPSM),将元数据中的至少一些(例如,LPSM和节目边界元数据,如果任意一个被提取的话,和/或PIM和/或SSM)设定到音频状态验证器203和级204,将所提取的元数据设定为(例如对后处理器300的)输出,从编码输入音频中提取音频数据,以及将所提取的音频数据设定到解码器202。

[0151] 输入至解码器200的编码音频比特流可以是AC-3比特流、E-AC-3比特流或杜比E比特流中的一个。

[0152] 图3的系统还包括后处理器300。后处理器300包括帧缓冲器301和包括耦接至缓冲器301的至少一个处理元件的其他处理元件(未示出)。帧缓冲器301存储(例如,以非暂态方式)由后处理器300从解码器200接收的解码音频比特流的至少一个帧。耦接后处理器300的处理元件并且将其配置成接收从缓冲器301输出的解码音频比特流的一系列帧并且使用从解码器200输出的元数据和/或从解码器200的级204输出的控制位对其进行自适应处理。通常,后处理器300被配置成使用来自解码器200的元数据对解码音频数据执行自适应处理(例如,使用LPSM值以及可选地还使用节目边界元数据对解码音频数据执行自适应响度处理,其中自适应处理可以基于响度处理状态、和/或由指示单个音频节目的音频数据的LPSM所指示的一个或多个音频数据特性)。

[0153] 解码器200和后处理器300的各种实现被配置成执行本发明的方法的不同的实施方式。

[0154] 解码器200的音频解码器202被配置成对由分析器205提取的音频数据进行解码以生成解码音频数据,并且将解码音频数据设定为(例如对后处理器300的)输出。

[0155] 状态验证器203被配置成对设定到其的元数据进行认证和验证。在一些实施方式中,元数据为(或被包括在)已经被包括在输入比特流(例如,根据本发明的实施方式)中的数据块。块可以包括用于对元数据和/或基本音频数据(从分析器205和/或解码器202提供至验证器203)进行处理的加密散列(基于散列的消息认证代码或“HMAC”)。数据块可以在这些实施方式中被数字地标记,使得下游的音频处理单元可以相对容易地认证和验证处理状态元数据。

[0156] 包括但不限于一个或多个非HMAC加密方法中的任意一个的其他加密方法可以用于元数据的验证(例如,在验证器203中)以确保元数据和/或基本的音频数据的安全的传输和接收。例如,验证(使用这样的加密方法)可以在接收本发明的音频比特流的实施方式的每个音频处理单元中被执行以确定包括在该比特流中的元数据和相应音频数据是否已经经历(和/或产生于)具体的处理(由元数据所指示的)并且在这样的具体的处理执行之后没有被修改。

[0157] 状态验证器203将控制数据设定到控制位生成器204,和/或将控制数据设定为输出(例如,设定到后处理器300)以指示验证操作的结果。响应于控制数据(以及可选地从输入比特流中提取的其他元数据),级204可以生成(以及设定到后处理器300):

[0158] 指示从解码器202输出的解码音频数据已经经历特定类型的响度处理(当LPSM指

示从解码器202输出的音频数据已经经历该特定类型的响度处理,并且来自验证器203的控制位指示LPSM有效时)的控制位;或

[0159] 指示从解码器202输出的解码音频数据应当经历特定类型的响度处理(例如,当LPSM指示从解码器202输出的音频数据没有经历具体类型的响度处理,或当LPSM指示从解码器202输出的音频数据已经经历该特定类型的响度处理但来自验证器203的控制位指示LPSM无效时)的控制位。

[0160] 或者,解码器200将由解码器202从输入比特流中提取的元数据以及由分析器205从输入比特流中提取的元数据设定到后处理器300,并且后处理器300使用元数据对解码音频数据执行自适应处理,或执行元数据的验证,然后如果验证指示元数据有效,则使用元数据对解码音频数据执行自适应处理。

[0161] 在一些实施方式中,如果解码器200接收根据本发明的使用加密散列的实施方式生成的音频比特流,则解码器被配置成对来自比特流所确定的数据块的加密散列进行分析和检索,所述块包括响度处理状态元数据(LPSM)。验证器203可以使用加密散列以对接收到的比特流和/或相关联的元数据进行验证。例如,如果验证器203基于参考加密散列与从数据块检索的加密散列之间的匹配发现LPSM有效,那么可以用向下游的音频处理单元(例如,可以是或包括音量校平单元的后处理器300)发信号以通过(未改变的)比特流的音频数据。另外地,可选地或可替代地,可以使用其他类型的加密技术替代基于加密散列的方法。

[0162] 在解码器200的一些实现中,所接收(以及缓存在存储器201中)的编码比特流为AC-3比特流或E-AC-3比特流,并且包括音频数据段(例如,图4所示的帧的AB0至AB5段)和元数据段,其中音频数据段指示音频数据,而元数据段中的至少一些中的每个包括PIM或SSM(或其他元数据)。解码器级202(和/或分析器205)被配置成从比特流中提取元数据。元数据段中的包括PIM和/或SSM(可选地还包括其他元数据)的每个元数据段被包括在比特流的帧的无用位段中,或比特流的帧的比特流信息(“BSI”)段的“addbsi”字段中,或比特流的帧的结束处的辅助数据字段(例如,图4所示的AUX段)中。比特流的帧可以包括一个或两个元数据段,其中每个元数据段包括元数据,并且如果帧包括两个元数据段,一个可以存在于帧的addbsi字段中而另一个存在于帧的AUX字段中。

[0163] 在一些实施方式中,缓存在缓冲器201中的比特流的每个元数据段(在本文中有时称为“容器”)具有包括元数据段报头(可选地还包括其他强制的或“核心”元素)、以及在元数据段报头之后的一个或更多元数据有效载荷的格式。如果存在,SIM被包括在元数据有效载荷中的一个有效载荷(由有效载荷报头标识,并且通常具有第一类型的格式)中。如果存在,PIM被包括在元数据有效载荷中的另一个有效载荷(由有效载荷报头标识,并且通常具有第二类型的格式)中。类似地,元数据的其他类型(如果存在)被包括在元数据有效载荷中的另一有效载荷(由有效载荷报头标识,并且通常具有针对元数据的类型的格式)中。示例性格式使得能够在除了解码期间之外的时间方便访问(例如,由解码之后的后处理器300、或由被配置成在没有对编码比特流执行完全解码的情况下识别元数据的处理器)SSM、PIM和其他元数据,并且允许在比特流的解码期间(例如,子流识别的)方便和高效的误差检测和校正。例如,在不以示例性格式访问SSM的情况下,解码器200可能错误地识别与节目相关联的子流的正确数量。元数据段中的一个元数据有效载荷可以包括SSM,元数据段中的另一个元数据有效载荷可以包括PIM,以及可选地,元数据段中的至少一个其他元数据有效载

荷可以包括其他元数据(例如,响度处理状态元数据或“LPSM”)。

[0164] 在一些实施方式中,包括在缓存在缓冲器201中的编码比特流(例如,指示至少一个音频节目的E-AC-3比特流)的帧中的子流结构元数据(SSM)有效载荷包括下面的格式的SSM:

[0165] 有效载荷报头,通常包括至少一个标识值(例如,指示SSM格式版本的2位值,以及可选地长度、周期、计数和子流关联值);以及

[0166] 在报头之后:

[0167] 指示由比特流指示的节目的独立子流的数量独立子流元数据;以及

[0168] 从属子流元数据,其指示:节目的每个独立子流是否具有至少一个与其相关联的从属子流,以及如果节目的每个独立子流具有至少一个与其相关联的从属子流,与节目的每个独立子流相关联的从属子流的数量。

[0169] 在一些实施方式中,缓存在缓冲器201中的编码比特流(例如,指示至少一个音频节目的E-AC-3比特流)的帧中的包括的节目信息元数据(PIM)有效载荷具有下面的格式:

[0170] 有效载荷报头,通常包括至少一个标识值(例如,指示PIM格式版本的值,以及可选地长度、周期、计数和子流关联值);以及在报头之后,下面的格式的PIM:

[0171] 音频节目的每个静音通道和每个非静音通道(即,节目的哪些通道包含音频信息,而哪些通道(如果有)仅包含静音(通常关于帧的持续时间))的活动通道元数据。在编码比特流是AC-3或E-AC-3比特流的实施方式中,比特流的帧中的活动通道元数据可以结合比特流的额外的元数据(例如,帧的音频编码模式(“acmod”)字段,以及如果存在,帧或相关联的从属子流帧中的chanmap字段)以确定节目的哪些通道包含音频信息而哪些通道包含静音;

[0172] 下混合处理状态元数据,其指示:节目是否被下混合(在编码之前或在编码期间),以及如果节目被下混合,所应用的下混合的类型。下混合处理状态元数据可以有助于实现解码器的上混合(在后处理器300中)下游,例如以使用最匹配所应用的下混合的类型的参数对节目的音频内容进行上混合。在编码比特流是AC-3或E-AC-3比特流的实施方式中,下混合处理状态元数据可以结合帧的音频编码模型(“acmod”)字段以确定应用于节目的通道的下混合(如果有)的类型;

[0173] 上混合处理状态元数据,其指示:在编码之前或在编码期间节目是否被上混合(例如,从较小数量的通道),以及如果节目被上混合,所应用的上混合的类型。上混合处理状态元数据可以有助于实现解码器的下混合(在后处理器中)下游,例如以与应用于节目的上混合(例如,杜比定向逻辑、或杜比定向逻辑 II 电影模式、或杜比定向逻辑 II 音乐模式、或杜比专业上混合器)的类型一致的方式对节目的音频内容进行下混合。在编码比特流是E-AC-3比特流的实施方式中,上混合处理状态元数据可以结合其他元数据(例如,帧的“strmtyp”字段的值)以确定应用于节目的通道的上混合(如果有)的类型。(E-AC-3比特流的帧的BSI字段中的)“strmtyp”字段的值指示帧的音频内容是否属于独立流(其确定节目)或(包括多个子流或与多个子流相关联的节目的)独立子流,从而可以独立于由E-AC-3比特流所指示的任何其他子流被编码,或帧的音频内容是否属于(包括多个子流或与多个子流相关联的节目的)从属子流,从而必须结合与其相关联的独立子流而被解码;以及

[0174] 预处理状态元数据,其指示:是否对帧的音频内容执行了预处理(在生成编码比特流的音频内容的编码之前),以及如果对帧音频内容执行了预处理,被执行的预处理的类

型。

[0175] 在一些实现中,预处理状态元数据指示:

[0176] 是否应用了环绕衰减(例如,在编码之前,音频节目的环绕通道是否被衰减了3dB),

[0177] 是否(例如,在编码之前对音频节目的环绕通道Ls和Rs通道)应用了90°相移,

[0178] 在编码之前,是否对音频节目的LFE通道应用了低通滤波器,

[0179] 在生成期间,是否监视节目的LFE通道的电平,以及如果监视了节目的LFE通道的电平,相对于节目的全音域音频通道的电平的LFE通道的监视电平,

[0180] 是否应当对节目的解码音频的每个块执行(例如,在解码器中)动态范围压缩,以及如果应当对节目的解码音频的每个块执行动态范围压缩,要执行的动态范围压缩的类型(和/或参数)(例如,该类型的预处理状态元数据可以指示下面的压缩配置文件类型中的哪种类型由编码器假定以生成被包括在编码比特流中的动态范围压缩控制值:电影标准、电影光线、音乐标准、音乐光线或语音。或者,预处理状态元数据的该类型可以指示应当以由被包括在编码比特流中的动态范围压缩控制值确定的方式对节目的解码音频内容的每个帧执行重动态范围压缩(“compr”压缩)),

[0181] 是否使用谱扩展和/或通道耦合编码以对特定频率范围的节目的内容进行编码,以及如果使用谱扩展和/或通道耦合编码以对特定频率范围的节目的内容进行编码,对其执行谱扩展编码的内容的频率分量的最小频率和最大频率,以及对其执行通道耦合编码的内容的频率分量的最小频率和最大频率。该类型的预处理状态元数据信息可以有助于执行解码器的均衡(在后处理器中)下游。通道耦合信息和谱扩展信息两者也有助于在代码转换操作和应用期间优化质量。例如,编码器可以基于参数(例如谱扩展和通道耦合信息)的状态优化其行为(包括预处理步骤例如头戴式耳机虚拟、上混合等的自适应)。而且,编码器可以基于进入的(并且认证的)元数据的状态动态地修改其耦合和谱扩展参数以匹配最佳值和/或将其耦合和谱扩展参数修改成最佳值,以及

[0182] 对白增强调整范围数据是否包括在编码比特流中,以及如果对白增强调整范围数据包括在编码比特流中,在相对于音频节目中的非对白内容的电平调整对白内容的电平的 对白增强处理(例如,在解码器的后处理器下游)的执行期间可得到的调整范围。

[0183] 在一些实施方式中,包括在缓存在缓冲器201中的编码比特流(例如,指示至少一个音频节目的E-AC-3比特流)的帧中的LPSM有效载荷包括下面的格式的LPSM:

[0184] 报头(通常包括标识LPSM有效载荷的开始同步字,在同步字之后的至少一个标识值,例如,在下面的表2中指示的LPSM格式版本、长度、周期、计数和子流关联值);以及

[0185] 在报头之后的:

[0186] 指示相应音频数据指示对白或不指示对白(例如,相应音频数据的哪些通道指示对白)的至少一个对白表示值(例如,表2的参数“对白通道”);

[0187] 指示相应音频内容是否符合响度调整的所指示的集合的至少一个响度调整符合值(例如,表2的参数“响度调整类型”);

[0188] 指示已经对相应音频数据执行的至少一种类型的响度处理的至少一个响度处理值(例如,表2的参数“对白选通响度校正标志”、“响度校正类型”中的一个或多个);以及

[0189] 指示相应音频数据的至少一个响度(例如,峰值或平均响度)特性的至少一个响度

值(例如,表2的参数“ITU相对选通响度”、“ITU语音选通响度”、“ITU (EBU 3341) 短期3s响度”和“真实峰值”中的一个或多个)。

[0190] 在一些实现中,分析器205(和/或解码器级202)被配置成从比特流的帧的无用位段或“addbsi”字段或辅助数据段中提取具有下面的格式的每个元数据段:

[0191] 元数据段报头(通常包括标识元数据段的开始的同步字,同步字之后的标识值,例如版本、长度、周期、扩展的元素计数和子流关联值);以及

[0192] 在元数据段报头之后的有助于元数据段或相应音频数据的元数据的至少一个的解密、认证或验证中的至少一种的至少一个保护值(例如,表1的HMAC摘要和音频指纹值);以及

[0193] 也在元数据段报头之后的标识每个下面的元数据有效载荷中的元数据的类型并且表示每个这样的有效载荷的配置(例如,尺寸)的至少一个方面的元数据有效载荷标识(“ID”)值和有效载荷配置值。

[0194] 每个元数据有效载荷段(优选地具有上面指定的格式)在相应的元数据有效载荷ID值和元数据配置值之后。

[0195] 更一般地,由本发明的优选实施方式生成的编码音频比特流具有提供将元数据元素和子元素标记为核心的(强制的)或扩展的(可选的)元素或子元素的机制的结构。这使得比特流(包括其元数据)的数据速率能够扩展到大量的应用。优选的比特流语法的核心的(强制的)元素还应当能够用信号通知与音频内容相关联的扩展的(可选的)元素存在于(带中)和/或远程位置(带外)。

[0196] 要求核心元素存在于比特流的每个帧中。核心元素的一些子元素是可选的,并且可以以任何组合存在。不要求扩展元素存在于每个帧中(以限制比特率总开销)。从而,扩展元素可以存在于一些帧中而不存于其他帧中。扩展元素的一些子元素是可选的,并且可以以任何组合存在,然而,扩展元素的一些子元素可以是强制的(即,如果扩展元素存在于比特流的帧中)。

[0197] 在一类实施方式中,生成(例如,通过实现本发明的音频处理单元)包括一系列音频数据段和元数据段的编码音频比特流。音频数据段指示音频数据,元数据段中的至少一些中的每个包括PIM和/或SSM(以及可选地至少一种其他类型的元数据),并且音频数据段被与元数据段时分复用。在该类中的优选实施方式中,元数据段中的每个具有在本文中要描述的优选的格式。

[0198] 在一种优选的格式中,编码比特流为AC-3比特流或E-AC-3比特流,并且元数据段中的包括SSM和/或PIM的每个元数据段被包括(例如,由编码器100的优选的实现的级107)作为比特流的帧的比特流信息(“BSI”)段的“addbsi”字段(图6所示)、或比特流的帧的辅助数据字段中、或比特流的帧的无用位段中的额外的比特流信息。

[0199] 在优选格式中,帧中的每个包括帧的无用位段(或addbsi字段)中的元数据段(在本文中有时也称为元数据容器或容器)。元数据段具有下面表1中所示的强制的元素(统一称为“核心元素”) (并且可以包括表1中所示的可选元素)。表1中所示的需要的元素中的至少一些被包括在元数据段的元数据段报头中,但一些可以被包括在元数据段的其他位置:

[0200] 表1

[0201]

参数	描述	强制的/可选的
同步 [ID]		M
核心元素版本		M
核心元素长度		M
核心元素周期 (xxx)		M
扩展元素计数	指示与核心元素相关联的扩展元数据元素的数量。该值可以随着比特流从生产到分配和传播而增加/减小。	M
子流关联	描述核心元素与哪个子流相关联。	M
签名 (HMAC 摘要)	关于整个帧的音频数据、核心元素和所有扩展元素 (使用 SHA-2 算法) 计算的 256 位 HMAC。	M
PGM 边界向下计数	仅关于音频节目文件/流的头或尾处的一些数量的帧出现的字段。从而, 核心元素版本改变可以	O

	用于用信号通知包括该参数。	
[0202]	音频指纹	关于由核心元素周期字段表示的一些数量的PCM 音频样本考虑的音频指纹。
	视频指纹	关于由核心元素周期字段表示的一些数量的压缩视频样本（如果存在）考虑的视频指纹。
	URL/UUID	定义该字段以携带参考额外节目内容（本质）和/或与比特流相关联的元数据的 URL 和/或 UUID（可能对指纹来说是冗余的）。

[0203] 在优选格式中,包含SSM、PIM或LPSM的每个元数据段(在编码比特流的帧的无用位段或addbsi或辅助数据字段中)包含元数据段报头(以及可选地额外的核心元素)、以及在元数据段报头(或元数据段报头和其他核心元素)之后的一个或更多个元数据有效载荷。每个元数据有效载荷包括被包括在有效载荷中的元数据有效载荷报头(指示元数据的具体类型(例如,SSM、PIM或LPSM)),之后是具体类型的元数据。通常,元数据有效载荷报头包括下面的值(参数):

[0204] 在元数据段报头(可以包括在表1中指定的值)之后的有效载荷ID(标识元数据的类型,例如,SSM、PIM或LPSM);

[0205] 在有效载荷ID之后的有效载荷配置值(通常指示有效载荷的大小);

[0206] 以及可选地还包括额外的有效载荷配置值(例如,指示从帧的开始处到有效载荷涉及的第一音频样本的音频样本的数量的偏置值,以及有效载荷优先权值,例如,指示其中有效载荷可以被丢弃的条件)。

[0207] 通常,有效载荷的元数据具有下面的格式中的一种:

[0208] 有效载荷的元数据为SSM,包括指示由比特流指示的节目的独立子流的数量独立子流元数据;以及从属子流元数据,其指示:节目的每个独立子流是否具有与其相关联的至少一个从属子流,以及如果节目的每个独立子流具有与其相关联的至少一个从属子流,与节目的每个独立子流相关联的从属子流的数量;

[0209] 有效载荷的元数据为PIM,包括指示音频节目的哪些通道包含音频信息以及哪些通道(如果有)仅包含静音(通常关于帧的持续时间)的活动通道元数据;下混合处理状态元

数据,其指示节目是否被下混合(在编码之前或在编码期间),以及如果节目被下混合,被应用的下混合的类型;上混合处理状态元数据,其指示在编码之前或在编码期间节目是否被上混合(例如,从较小数量的通道),以及如果节目被上混合,被应用的上混合的类型;以及预处理状态元数据,其指示是否(在生成编码比特流的音频内容的编码之前)对帧的音频数据执行了预处理,以及如果对帧的音频数据执行了预处理,执行的预处理的类型;或

[0210] 有效载荷的元数据为LPSM,该LPSM具有如下面的表(表2)所指示的格式:

[0211] 表2

LPSM 参数 [智能的 响度]	描述	独特的 状态的 数量	强制的/可选的	插入速率 (参数的 更新的周 期)
LPSM 版本			M	
[0212] LPSM 周期 (xxx)	仅可应用于 xxx 字 段		M	
LPSM 计数			M	
LPSM 子流关联			M	

[0213]

对白通道	指示先前 0.5 秒 L、C 和 R 音频通道的哪些组合包含语音。当语音不存在于任何 L、C 或 R 组合中时，则该参数应当指示“无对白”。	8	M	~0.5 秒 (典型的)
响度调整类型	指示相关联的音频数据流符合调整的特定集合（例如，ATSCA/85 或 EBU128）。	8	M	帧
对白选通响度校正标志	指示相关联的音频流是否已经基于对白选通被校正	2	O（仅在响度调整类型指示相应音频是未被校正的情况下存在）	帧
响度校正类型	指示相关联的音频流是否已经使用无限预测（基于文件的）或使用实时（RT）响度和动态范围控制器而被校正。	2	O（仅在响度调整类型指示相应音频是未被校正的情况下存在）	帧
ITU 相对选通响度（INF）	指示被应用的相关联的音频流 w/o 元数据的 ITU-R	128	O	1 秒

[0214]

	BS.1770-3 综合响度 (例如, 7 位: -58 -> +5.5 LKFS 0.5 LKFS 步长)			
ITU 语音选通响度 (INF)	指示被应用的相关联的音频流 w/o 元数据的语音/对白的 ITU-R BS.1770-1/3 综合响度(例如, 7 位: -58 -> +5.5 LKFS 0.5 LKFS 步长)	128	O	1 秒
ITU ( EBU 3341) 短期 3s 响度	指示被应用的相关联的音频流 w/o 元数据的 3 秒未选通 ITU (ITU-BS.1771-1) 响度 (滑动窗) @ ~10Hz 插入速率 (例如, 8 位: 116 -> +11.5 LKFS 0.5 LKFS 步长)	256	O	0.1 秒
真实峰值	指示被应用的相关联的音频流 w/o 元数据的 ITU-R BS.1770-3 附件 2 真实峰值 (dBTP) (即, 关于在元素周期字段中用信	256	O	0.5 秒

	号通知的帧周期的最大值) 116->+11.5 LKFS 0.5 LKFS 步长)			
	下混合偏置	指示下混合响度偏置		
[0215]	节目边界	指示在帧中什么时候节目边界出现或已经出现。当节目边界不在帧边界处时, 可选的样本偏置将指示实际的节目边界在帧中多远出现		

[0216] 在根据本发明而生成的编码比特流的另一优选格式中, 比特流为AC-3比特流或E-AC-3比特流, 并且元数据段中的包括PIM和/或SSM(可选地还包括至少一个其他类型的元数据)的每个元数据段(例如, 由编码器100的优选实现的级107)被包括在下列中的任一个中: 比特流的帧的无用位段; 或比特流的帧的比特流信息(“BSI”)段的“addbsi”字段(图6所示); 或比特流的帧的结束处的辅助数据字段(例如, 图4中所示的AUX段)。帧可以包括一个或两个元数据段, 元数据段中的每个包括PIM和/或SSM, 并且(在一些实施方式中)如果帧包括两个元数据段, 一个可以存在于帧的addbsi字段中而另一个存在于帧的AUX字段中。每个元数据段优选地具有参照上面的表1在上面所指定的格式(即, 包括在表1中所指定的核心元素, 在核心元素之后是有效载荷ID值(标识元数据段的每个有效载荷中的元数据的类型)和有效载荷配置值, 以及每个元数据有效载荷)。包括LPSM的每个元数据段优选地具有参照上面的表1和表2在上面所指定的格式(即, 包括在表1中所指定的核心元素, 在核心元素之后是有效载荷ID(标识元数据作为LPSM)以及有效载荷配置值, 之后是有效载荷(具有如表2中所指示的格式的LPSM数据))。

[0217] 在另一优选格式中, 编码比特流为杜比E比特流, 并且元数据段中的包括PIM和/或SSM(可选地还包括其他元数据)的每个元数据段为杜比E保护带间隔的第一N样本位置。包括这样的包括LPSM的元数据段的杜比E比特流优选地包括指示在SMPTE 337M前同步信号的Pd字中用信号通知的LPSM有效载荷长度的值(SMPTE 337M Pa字重复频率优选地保持与相关联的视频帧速率相同)。

[0218] 在优选的格式中, 其中编码比特流为E-AC-3比特流, 元数据段中的包括PIM和/或SSM(可选地还包括LPSM和/或其他元数据)的每个元数据段(例如, 由编码器100的优选实现

的级107)被包括作为比特流的帧的无用位段或比特流信息(“BSI”)段的“addbsi”字段中的额外的比特流信息。接下来对以该优选的格式使用LPSM对E-AC-3比特流进行编码的额外的方面进行描述:

[0219] 1. 在E-AC-3比特流的生成期间,尽管E-AC-3编码器(将LPSM值插入待比特流中)是“活动的”,对于每个生成的帧(同步帧),比特流应当包括在帧的addbsi字段(或无用位段)中携带的元数据块(包括LPSM)。要求携带元数据块的比特不应当增加编码器比特率(帧长度);

[0220] 2. 每个元数据块(包含LPSM)应当包含下面的信息:

[0221] 响度校正类型标志:其中,“1”指示相应的音频数据的响度在编码器的上游被校正,而“0”指示响度由嵌入在编码器中的响度校正器(例如,图2的编码器100的响度处理器103)校正;

[0222] 语音通道:指示哪些源通道包含语音(在先前的0.5秒)。如果没有检测到语音,应当如此指示;

[0223] 语音响度:指示包括语音(在先前的0.5秒)的每个相应的音频通道的综合语音响度;

[0224] ITU响度:指示每个相应音频通道的综合ITU BS.1770-3响度;以及

[0225] 增益:解码器中的逆变的响度复合增益(以表明可逆性);

[0226] 3. 当E-AC-3编码器(将LPSM值插入到比特流中)是“活动的”,并且正在接收具有“信任”标志的AC-3帧时,编码器中的响度控制器(例如,图2的编码器100的响度处理器103)应当被旁路。“信任的”源对白归一化和DRC值应当被传递(例如,由编码器100的生成器106)至E-AC-3编码器部件(例如,编码器100的级107)。LPSM块生成继续,并且响度校正类型标志被设置成“1”。响度控制器旁路序列必须被同步至“信任”标志出现的解码AC-3帧的开始。响度控制器旁路序列应当被如下实现:校平器量控制跨10个音频块周期(即,53.3毫秒)从值9减少到值0,并且校平器返回结束计量器控制被置于旁路模式(该操作应当导致无缝转换)。调节器的术语“信任的”旁路暗示源比特流的对白归一化值还在编码的输出端处被重新利用。(例如,若果该“信任的”源比特流具有-30的对白归一化值,则编码器的输出应当利用-30用于输出对白归一化值);

[0227] 4. 当E-AC-3编码器(将LPSM值插入到比特流中)是“活动的”,并且正在接收不具有“信任”标志的AC-3帧时,编码器中嵌入的响度控制器(例如,图2的编码器100的响度处理器103)应当是活动的。LPSM块生成继续,并且响度校正类型标志被设置成“0”。响度控制器激活序列应当被同步至其中“信任”标志消失的解码AC-3帧的开始。响度控制器激活序列应当被如下实现:校平器量控制跨1个音频块周期(例如,5.3毫秒)从值0增加至值9,并且校平器返回结束计量器控制被置于“活动的”模式(该操作应当导致无缝转换,并且包括返回结束计量器综合复位);以及5. 在编码期间,图形用户接口(GUI)应当给用户指示下面的参数:“输入音频节目:[信任的/不信任的]”——该参数的状态基于输入信号内的“信任”标志的存在;以及“实时响度校正:[启用/禁用]”——该参数的状态基于编码器中嵌入的响度控制器是否是活动的。

[0228] 当对使LPSM(以优选的格式)包括在比特流的每个帧的无用位段或跳过字段段或比特流信息(“BSI”)段的“addbsi”字段中的AC-3或E-AC-3比特流进行解码时,解码器应当

对(无用位段或addbsi字段中的)LPSM块数据进行分析并且将全部所提取的LPSM值传递至图形用户接口(GUI)。在每帧刷新所提取的LPSM值的集合。

[0229] 在根据本发明而生成的编码比特流的另一优选格式中,编码比特流为AC-3比特流或E-AC-3比特流,并且元数据段中的包括PIM和/或SSM(可选地还包括LPSM和/或其他元数据)的每个元数据段(例如,由编码器100的优选的实现的级107)被包括在比特流的帧的无用位段或AUX段中或作为比特流信息(“BSI”)段的“addbsi”字段(图6所示)中的额外的比特流信息。在该格式(为关于上面参照表1和表2所描述的格式的变型)中,包含LPSM的addbsi(或AUX或无用位)字段中的每个字段包含下面的LPSM值:

[0230] 表1中所指定的核心元素,之后是有效载荷ID(标识元数据作为LPSM)和有效载荷值,之后是具有下面的格式(与上面表2中所示的强制元素类似)的有效载荷(LPSM数据):

[0231] LPSM有效载荷的版本:指示LPSM有效载荷的版本的2位字段;

[0232] dialchan:指示包含口语对白的相应音频数据的左、右和/或中央通道的3位字段。dialchan字段的位分配可以如下:指示左通道中存在对白的位0被存储在dialchan字段的最高有效位中;而指示中央通道中存在对白的位2被存储在dialchan字段的最低有效位中。如果在节目的前0.5秒期间相应通道包含口语对白,则dialchan字段的每个位被设置为“1”;

[0233] loudregtyp:指示节目响度符合哪个响度调整标准的4位字段。将“loudregtyp”字段设置为“0000”指示LPSM不指示响度调整符合。例如,该字段的一个值(例如,0000)可以指示未指示符合响度调整标准,该字段的另一值(例如,0001)可以指示节目的音频数据符合ATSC A/85标准,并且该字段的另一值(例如,0010)可以指示节目的音频数据符合EBU R128标准。在该示例中,如果该字段被设置为除了“0000”之外的任何值,则有效载荷中随后应该是loudcorrldialgat和loudcorrtyp字段;

[0234] loudcorrldialgat:指示是否已经应用对白选通校正的1位字段。如果已经使用对白选通校正了节目的响度,则loudcorrldialgat字段的值被设置为“1”。否则,被设置为“0”;

[0235] loudcorrtyp:指示对节目应用的响度校正的类型的1位字段。如果已经使用无限超前(基于文件的)响度校正处理校正了节目的响度,则loudcorrtyp字段的值被设置为“0”。如果已经使用实时响度测量和动态范围控制的组合校正了节目的响度,则该字段的值被设置为“1”;

[0236] loudrelgate:指示相对选通节目响度(ITU)是否存在的1位字段。如果loudrelgate字段被设置为“1”,则有效载荷中随后应该是7位ituloudrelgat字段;

[0237] loudrelgat:指示相对选通节目响度(ITU)的7位字段。该字段指示由于正在应用的对白归一化和动态范围压缩(DRC),在没有任何增益调整的情况下根据ITU-R BS.1770-3而测量的音频节目的综合的响度。0至127的值被解释为以0.5LKFS步长的-58LKFS至+5.5LKFS;

[0238] loudspchgate:指示语音选通响度数据(ITU)是否存在的1位字段。如果loudspchgate字段被设置为“1”,则有效载荷中随后应是7位loudspchgat字段;

[0239] loudspchgate:指示语音选通节目响度的7位字段。该字段指示由于正在应用的对白归一化和动态范围压缩,在没有任何增益调整的情况下根据ITU-R BS.1770-3的公式(2)而测量的整个相应音频节目的综合响度。0至127的值被解释为以0.5LKFS步长的-58LKFS至

+5.5LKFS;

[0240] loudstrm3e:指示短期(3秒)响度数据是否存在的1位字段。如果该字段被设置为“1”,则有效载荷中随后应是7位loudstrm3s字段;

[0241] loudstrm3s:指示由于正在应用的对白归一化和动态范围压缩,在没有任何增益调整的情况下根据ITU-R BS.1771-1而测量的相应音频节目的前3秒的未选通响度的7位字段。0至256的值被解释为以0.5LKFS步长的-116LKFS至+11.5LKFS;

[0242] truepke:指示真实峰值响度数据是否存在的1位字段。如果truepke字段被设置为“1”,则有效载荷中随后应是8位truepk字段;以及

[0243] truepk:指示由于正在应用的对白归一化和动态范围压缩,在没有任何增益调整的情况下根据ITU-R BS.1770-3的附件2而测量的节目真实峰值样本值的8位字段。0至256的值被解释为以0.5LKFS步长的-116LKFS至+11.5LKFS。

[0244] 在一些实施方式中,AC-3比特流或E-AC-3比特流的帧的无用位段或辅助数据(或“addbsi”)字段中的元数据段的核心元素包括元数据段报头(通常包括标识值,例如,版本),以及在元数据段报头之后的:指示元数据段的元数据是否包括指纹数据(或其他保护值)的值、指示(与对应于元数据段的元数据的音频数据有关的)外部数据是否存在的值、关于由核心元素标识的每种类型的元数据(例如,PIM和/或SSM和/或LPSM和/或一种类型的元数据)的有效载荷ID值和有效载荷配置值、以及由元数据段报头(或元数据段的其他核心元素)标识的至少一种类型的元数据的保护值。元数据段的元数据有效载荷在元数据段报头之后,并且(在有些情况下)嵌套在元数据段的核心元素内。

[0245] 本发明的实施方式可以以硬件、固件、或软件、或硬件和软件的组合(例如,作为可编程逻辑阵列)被实现。除非另外指明,作为本发明的部分而被包括在内的算法或处理不内在涉及任何特定的计算机或其他设备。具体地,各种通用机器可以利用根据本文中的教示而编写的程序而被使用,或可以更加便于构造更具体的装置(例如,集成电路)以执行所需要的方法步骤。从而,本发明可以在一个或多个可编程计算机系统(例如,图1的元件、或图2的编码器100(或编码器的元件)、或图3的解码器(或解码器的元件)、或图3的后处理器(或后处理器的元件)中任意一种的实施)上执行的一个或多个计算机程序而被实现,每个可编程计算机系统包括至少一个处理器、至少一个数据存储系统(包括易失性和非易失性存储器和/或存储元件)、至少一个输入装置或端口以及至少一个输出装置或端口。程序代码被应用于输入数据以执行本文中所描述的功能并生成输出信息。输出信息以已知的方式应用于一个或多个输出装置。

[0246] 每个这样的程序可以以任何期望的计算机语言(包括机器、汇编或高级过程的、逻辑的或面向对象的编程语言)实现以与计算机系统通信。在任何情况下,语言可以是编译语言或解释语言。

[0247] 例如,当由计算机软件指令序列实现时,本发明的实施方式的各种功能和步骤可以由在适当的数字信号处理硬件中运行的多线程软件指令序列实现,在这种情况下,实施方式的各种装置、步骤和功能可以对应于软件指令的部分。

[0248] 每个这样的计算机程序优选地存储在或下载至由通用或专用可编程计算机可读的存储介质或装置(例如,固态存储器或介质、磁介质或光介质),当存储介质或装置由计算机系统读取以执行本文所描述的过程时,用于配置和操作计算机。本发明的系统还可以被

实现为配置有(例如,存储)计算机程序的计算机可读存储介质,其中,这样配置的存储介质使得计算机系统以特定和预先定义的方式操作以执行本文中所描述的功能。

[0249] 已经描述了本发明的大量的实施方式。然而,应当理解的是,在不偏离本发明的精神和范围的情况下可以作出各种修改。鉴于上面的教示,本发明的大量的修改和变型是可能的。应当理解的是,在所附权利要求的范围内,可以与本文中具体描述的方式不同地实践本发明。

[0250] 本发明还包括以下方案:

[0251] 方案1.一种音频处理单元,包括:

[0252] 缓冲存储器;以及

[0253] 至少一个处理子系统,其耦接至所述缓冲存储器,其中所述缓冲存储器存储编码音频比特流的至少一个帧,所述帧包括在所述帧的至少一个跳过字段的至少一个元数据段中的节目信息元数据或子流结构元数据以及在所述帧的至少一个其他段中的音频数据,其中所述处理子系统被耦接并且被配置成使用所述比特流的元数据执行所述比特流的生成、所述比特流的解码或所述比特流的音频数据的自适应处理中的至少一种,或使用所述比特流的元数据执行所述比特流的音频数据或元数据中至少之一的认证或验证中的至少一种,

[0254] 其中,所述元数据段包括至少一个元数据有效载荷,所述元数据有效载荷包括:

[0255] 报头;以及

[0256] 在所述报头之后的,所述节目信息元数据的至少一部分或所述子流结构元数据的至少一部分。

[0257] 方案2.根据方案1所述的音频处理单元,其中,所述编码音频比特流指示至少一个音频节目,并且所述元数据段包括节目信息元数据有效载荷,所述节目元数据有效载荷包括:

[0258] 节目信息元数据报头;以及

[0259] 在所述节目信息元数据报头之后的,指示所述节目的音频内容的至少一个属性或特性的节目信息元数据,所述节目信息元数据包括指示所述节目的每个非静音通道和每个静音通道的活动通道元数据。

[0260] 方案3.根据方案2所述的音频处理单元,其中,所述节目信息元数据还包括下列之一:

[0261] 下混合处理状态元数据,其指示:所述节目是否是下混合过的,以及在所述节目是下混合过的情况下应用于所述节目的下混合的类型;

[0262] 上混合处理状态元数据,其指示:所述节目是否是上混合过的,以及在所述节目是上混合过的情况下应用于所述节目的上混合的类型;

[0263] 预处理状态元数据,其指示:是否对所述帧的音频内容执行了预处理,以及在对所述帧的音频内容执行了预处理的情况下对所述音频内容执行的预处理的类型;或

[0264] 谱扩展处理或通道耦合元数据,其指示:是否对所述节目应用了谱扩展处理或通道耦合,以及在对所述节目应用了谱扩展处理或通道耦合的情况下应用谱扩展或通道耦合的频率范围。

[0265] 方案4.根据方案1所述的音频处理单元,其中,所述编码音频比特流指示具有音频内容的至少一个独立子流的至少一个音频节目,而所述元数据段包括子流结构元数据有效

载荷,所述子流结构元数据有效载荷包括:

[0266] 子流结构元数据有效载荷报头;以及

[0267] 在所述子流结构元数据有效载荷报头之后的,指示所述节目的独立子流的数量、独立子流元数据,以及指示所述节目的每个独立子流是否具有至少一个相关联的从属子流的从属子流元数据。

[0268] 方案5.根据方案1所述的音频处理单元,其中,所述元数据段包括:

[0269] 元数据段报头;

[0270] 在所述元数据段报头之后的至少一个保护值,其用于所述节目信息元数据、或所述子流结构元数据、或与所述节目信息元数据或所述子流结构元数据相对应的所述音频数据中至少之一的解密、认证或验证中的至少一种;以及

[0271] 在所述元数据段报头之后的元数据有效载荷标识值和有效载荷配置值,其中所述元数据有效载荷在所述元数据有效载荷标识值和所述有效载荷配置值之后。

[0272] 方案6.根据方案5所述的音频处理单元,其中,所述元数据段报头包括标识所述元数据段的开始的同步字、以及在所述同步字之后的至少一个标识值,并且所述元数据有效载荷的所述报头包括至少一个标识值。

[0273] 方案7.根据方案1所述的音频处理单元,其中,所述编码音频比特流为AC-3比特流或E-AC-3比特流。

[0274] 方案8.根据方案1所述的音频处理单元,其中,所述缓冲存储器以非暂态方式存储所述帧。

[0275] 方案9.根据方案1所述的音频处理单元,其中,所述音频处理单元为编码器。

[0276] 方案10.根据方案9所述的音频处理单元,其中,所述处理子系统包括:

[0277] 解码子系统,其被配置成接收输入音频比特流并且从所述输入音频比特流中提取输入元数据和输入音频数据;

[0278] 自适应处理子系统,其被耦接并且被配置成使用所述输入元数据对所述输入音频数据执行自适应处理,由此生成经处理音频数据;以及

[0279] 编码子系统,其被耦接并且被配置成响应于所述经处理音频数据,包括通过将所述节目信息元数据或所述子流结构元数据包括在所述编码音频比特流中,来生成所述编码音频比特流,并且将所述编码音频比特流设定到所述缓冲存储器。

[0280] 方案11.根据方案1所述的音频处理单元,其中,所述音频处理单元为解码器。

[0281] 方案12.根据方案11所述的音频处理单元,其中,所述处理子系统为耦接至所述缓冲存储器并且被配置成从所述编码音频比特流中提取所述节目信息元数据或所述子流结构元数据的解码子系统。

[0282] 方案13.根据方案1所述的音频处理单元,包括:

[0283] 子系统,其被耦接至所述缓冲存储器并且被配置成:从所述编码音频比特流中提取所述节目信息元数据或所述子流结构元数据,以及从所述编码音频比特流中提取所述音频数据;以及

[0284] 后处理器,其被耦接至所述子系统并且被配置成使用从所述编码音频比特流中提取的所述节目信息元数据或所述子流结构元数据中至少之一对所述音频数据执行自适应处理。

[0285] 方案14.根据方案1所述的音频处理单元,其中,所述音频处理单元为数字信号处理器。

[0286] 方案15.根据方案1所述的音频处理单元,其中,所述音频处理单元为预处理器,所述预处理器被配置成从所述编码音频比特流中提取所述节目信息元数据或所述子流结构元数据以及所述音频数据,并且使用从所述编码音频比特流中提取的所述节目信息元数据或所述子流结构元数据中至少之一对所述音频数据执行自适应处理。

[0287] 方案16.一种用于对编码音频比特流进行解码的方法,所述方法包括以下步骤:

[0288] 接收编码音频比特流;以及

[0289] 从所述编码音频比特流中提取元数据和音频数据,其中所述元数据是或包括节目信息元数据和子流结构元数据,

[0290] 其中,所述编码音频比特流包括一系列帧并且指示至少一个音频节目,所述节目信息元数据和所述子流结构元数据指示所述节目,所述帧中的每个包括至少一个音频数据段,每个所述音频数据段包括所述音频数据的至少一部分,所述帧的至少一个子集中的每个帧包括元数据段,并且每个所述元数据段包括所述节目信息元数据的至少一部分以及所述子流结构元数据的至少一部分。

[0291] 方案17.根据方案16所述的方法,其中,所述元数据段包括节目信息元数据有效载荷,所述节目信息元数据有效载荷包括:

[0292] 节目信息元数据报头;以及

[0293] 在所述节目信息元数据报头之后的指示所述节目的音频内容的至少一个属性或特性的节目信息元数据,所述节目信息元数据包括指示所述节目的每个非静音通道和每个静音通道的活动通道元数据。

[0294] 方案18.根据方案17所述的方法,其中,所述节目信息元数据还包括下列中的至少一个:

[0295] 下混合处理状态元数据,其指示:所述节目是否是下混合过的,以及在所述节目是下混合过的情况下应用于所述节目的下混合的类型;

[0296] 上混合处理状态元数据,其指示:所述节目是否是上混合过的,以及在所述节目是上混合过的情况下应用于所述节目的上混合的类型;或

[0297] 预处理状态元数据,其指示:是否对所述帧的音频内容执行了预处理,以及在对所述帧的音频内容执行了预处理的情况下对所述音频内容执行的预处理的类型。

[0298] 方案19.根据方案16所述的方法,其中,所述编码音频比特流指示具有音频内容的至少一个独立子流的至少一个音频节目,并且所述元数据段包括子流结构元数据有效载荷,所述子流结构元数据有效载荷包括:

[0299] 子流结构元数据有效载荷报头;以及

[0300] 在所述子流结构元数据有效载荷报头之后的,指示所述节目的独立子流的数量的独立子流元数据以及指示所述节目的每个独立子流是否具有至少一个相关联的从属子流的从属子流元数据。

[0301] 方案20.根据方案16所述的方法,其中,所述元数据段包括:

[0302] 元数据段报头;

[0303] 在所述元数据段报头之后的至少一个保护值,用于所述节目信息元数据或所述子

流结构元数据或与所述节目信息元数据和所述子流结构元数据相对应的所述音频数据中至少之一的解密、认证或验证中的至少一个;以及

[0304] 在所述元数据段报头之后的,包括所述节目信息元数据的所述至少一部分和所述子流结构元数据的所述至少一部分的元数据有效载荷。

[0305] 方案21.根据方案16所述的方法,其中,所述编码音频比特流为AC-3比特流或E-AC-3比特流。

[0306] 方案22.根据方案16所述的方法,还包括步骤:

[0307] 使用从所述编码音频比特流中提取的所述节目信息元数据或所述子流结构元数据中至少之一,对所述音频数据执行自适应处理。

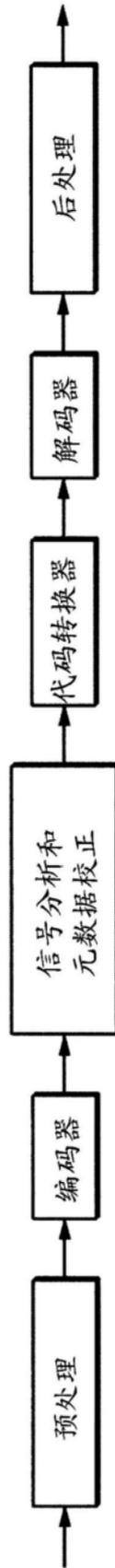


图1

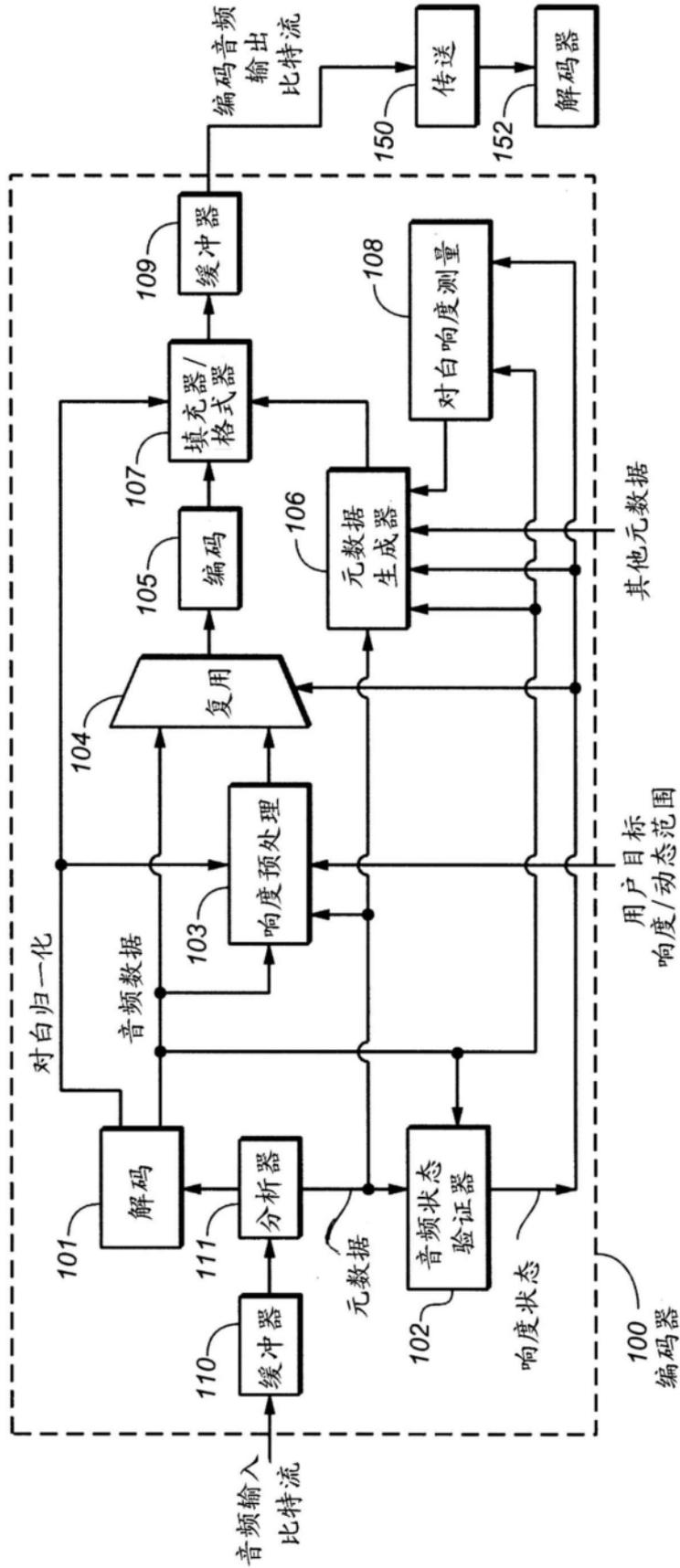


图2

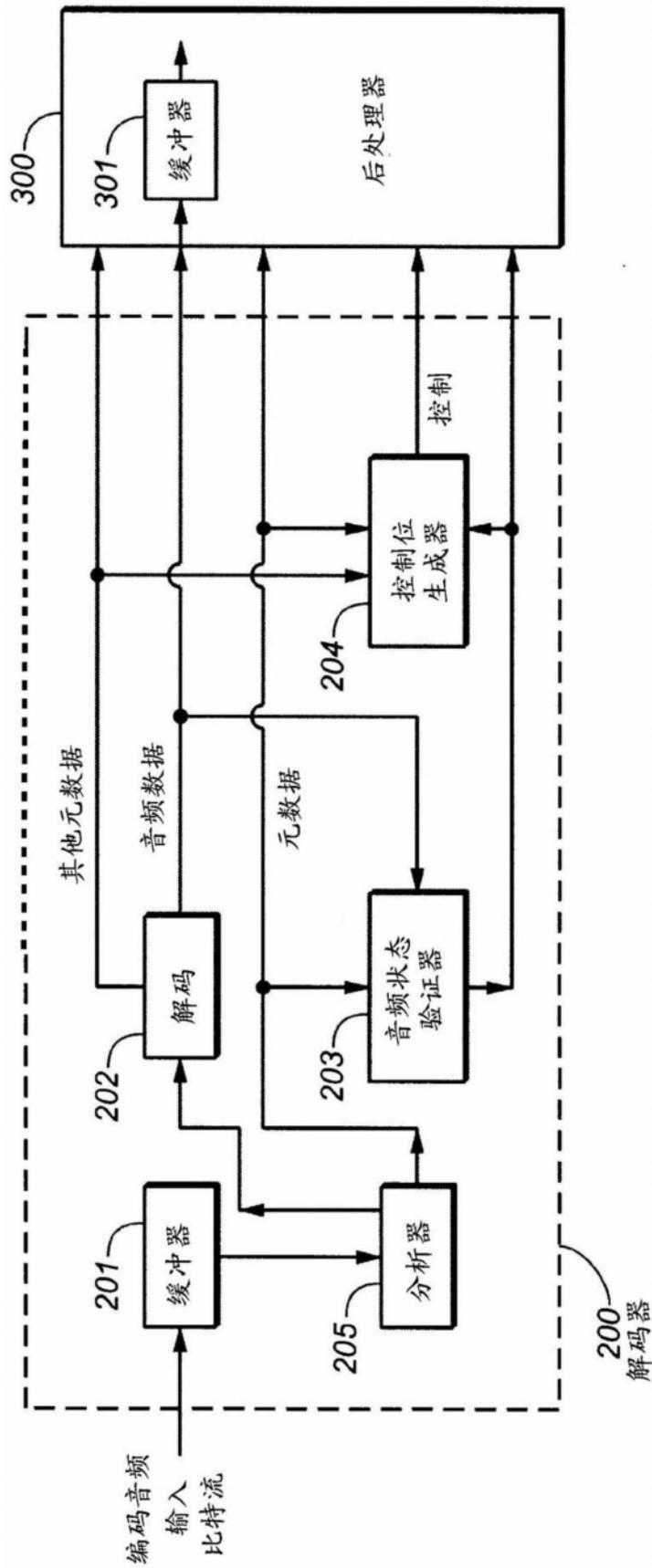


图3



图4

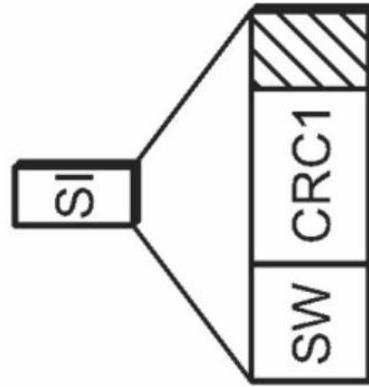


图5

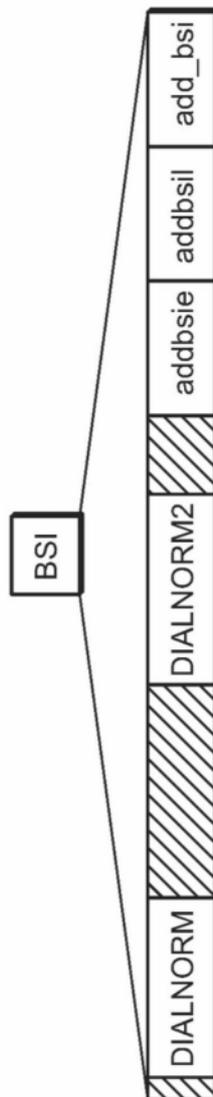
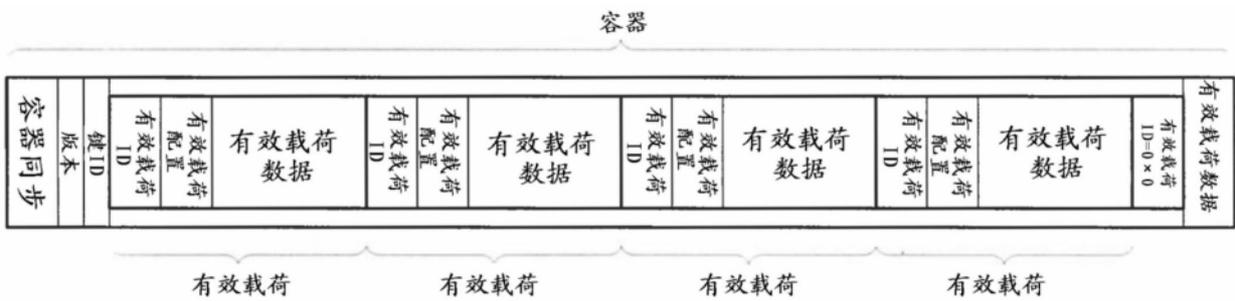


图6



图7



容器（元数据段）结构

图8