



(12) 发明专利申请

(10) 申请公布号 CN 119479682 A

(43) 申请公布日 2025.02.18

(21) 申请号 202310944857.8

(74) 专利代理机构 中国贸促会专利商标事务所

(22) 申请日 2014.01.15

有限公司 11038

(30) 优先权数据

61/754,882 2013.01.21 US

(51) Int.CI.

61/809,250 2013.04.05 US

G10L 21/034 (2013.01)

61/824,010 2013.05.16 US

G10L 21/0364 (2013.01)

(62) 分案原申请数据

G10L 19/02 (2013.01)

201480005314.9 2014.01.15

G10L 19/16 (2013.01)

(71) 申请人 杜比实验室特许公司

H04R 29/00 (2006.01)

地址 美国加利福尼亚

H03G 9/00 (2006.01)

申请人 杜比国际公司

G06F 3/16 (2006.01)

(72) 发明人 J·瑞德米勒 S·G·诺克罗斯

K·J·罗德恩

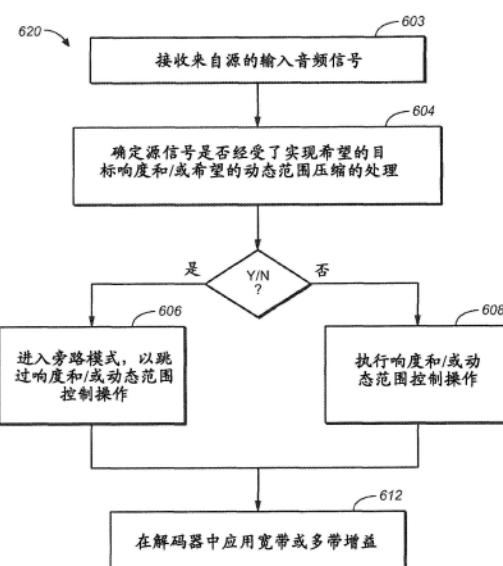
权利要求书2页 说明书22页 附图14页

(54) 发明名称

用于在不同回放设备之间优化响度和动态范围的系统和方法

(57) 摘要

本公开涉及用于在不同回放设备之间优化响度和动态范围的系统和方法。实施例涉及如下的方法和系统，其用于在位流中接收与音频数据相关联的元数据，并且分析该元数据以确定在位流中是否可获得用于第一组音频回放设备的响度参数。响应于确定存在用于第一组的参数，系统使用该参数和音频数据来呈现音频。响应于确定不存在用于第一组的响度参数，系统分析第一组的一个或多个特性并且基于该一个或多个特性来确定参数。



1. 一种用于对编码音频位流的一个或更多个帧进行解码的音频处理设备,所述编码音频位流包括用于多个动态范围控制(DRC)简档的元数据和音频数据,所述音频处理设备包括:

位流解析器,被配置为解析所述编码音频位流,并且提取用于DRC简档中的一个或多个的元数据以及编码音频数据;以及

音频解码器,配置为对所述编码音频数据进行解码,并且将DRC增益应用于解码的音频数据,其中:

每个DRC简档适用于至少一种设备类型或收听环境;

音频解码器响应于关于音频处理设备或收听环境的信息来选择DRC简档中的一个或多个;以及

应用于解码的音频数据的DRC增益对应于该一个或多个被选择的DRC简档。

2. 根据权利要求1所述的音频处理设备,其中,对于该一个或多个被选择的DRC简档的DRC增益被包含在编码音频位流的元数据中。

3. 根据权利要求1所述的音频处理设备,其中,对于该一个或多个被选择的DRC简档的DRC增益根据被包含在编码音频位流的元数据中的该一个或多个被选择的DRC简档的DRC参数被确定。

4. 根据权利要求3所述的音频处理设备,其中,DRC参数指示静态增益传递特性以及增益平滑时间常数。

5. 根据权利要求4所述的音频处理设备,其中,时间常数包括慢和快上升时间常数、以及慢和快释放时间常数。

6. 根据权利要求4所述的音频处理设备,其中,静态增益传递特性包括空带范围和最大提升量。

7. 根据权利要求1所述的音频处理设备,其中,DRC简档中的一个或多个对应于通过便携式设备对解码的音频数据的回放。

8. 根据权利要求1所述的音频处理设备,其中,DRC简档中的一个或多个对应于在噪声收听环境中对解码的音频数据的回放。

9. 根据权利要求1所述的音频处理设备,其中,DRC简档中的一个或多个对应于在安静收听环境中对解码的音频数据的回放。

10. 根据权利要求1所述的音频处理设备,其中,DRC简档中的一个或多个对应于通过耳机对解码的音频数据的回放。

11. 根据权利要求1所述的音频处理设备,其中,DRC简档中的一个或多个对应于通过扬声器对解码的音频数据的回放。

12. 一种由音频处理设备执行的用于对编码音频位流的一个或更多个帧进行解码的方法,所述编码音频位流包括用于多个动态范围控制(DRC)简档的元数据和音频数据,所述方法包括:

解析所述编码音频位流,并且提取用于DRC简档中的一个或多个的元数据以及编码音频数据;

对所述编码音频数据进行解码,并且将DRC增益应用于解码的音频数据,其中:

每个DRC简档适用于至少一种设备类型或收听环境;

DRC简档中的一个或多个响应于关于音频处理设备或收听环境的信息被选择;以及应用于解码的音频数据的DRC增益对应于该一个或多个被选择的DRC简档。

13.一种非暂时性计算机可读存储介质,包括指令序列,所述指令序列在由音频解码设备执行时使得所述音频解码设备执行根据权利要求12所述的方法。

14.一种方法,包括:

在位流中接收与音频数据相关联的元数据;

分析位流中的该元数据以确定在位流中是否可获得用于第一组音频回放设备的响度参数;

响应于确定存在用于第一组的参数,使用该参数和音频数据来呈现音频;以及

响应于确定不存在用于第一组的响度参数,分析第一组的一个或多个特性并且基于该一个或多个特性来确定参数。

15.一种装置,包括:

接口,被配置为接收包括与音频数据相关联的元数据的位流;

分析器,耦合到所述接口,并且被配置为分析位流中的元数据以确定在位流中是否可得到用于第一组音频回放设备的响度参数;以及

呈现组件,被配置为响应于分析器确定存在用于第一组的参数,使用该参数和音频数据来呈现音频,呈现组件还被配置为响应于分析器确定不存在用于第一组的响度参数,分析第一组的一个或多个特性并且基于该一个或多个特性来确定参数。

16.一种方法,包括:

接收音频数据和与所述音频数据相关联的元数据;

分析位流中的元数据以确定在该流中是否可得到与用于第一组音频设备的响度参数相关联的响度信息;

响应于确定对于该第一组存在该响度信息,从该流确定响度信息,并且传输该音频数据和响度信息以用于呈现音频;以及

响应于确定对于该第一组不存在该响度信息,确定与输出简档相关联的响度信息,并且传输所确定的该输出简档的响度信息以用于呈现音频。

17.一种音频数据解码装置,包括:

接口,用于接收音频数据和与所述音频数据相关联的元数据;

第一组件,用于分析位流中的元数据以确定在该流中是否可得到与用于第一组音频设备的响度参数相关联的响度信息,

所述第一组件响应于确定对于该第一组存在该响度信息,从该流确定响度信息,并且传输该音频数据和响度信息以用于呈现音频;以及响应于确定对于该第一组不存在该响度信息,确定与输出简档相关联的响度信息,并且传输所确定的该输出简档的响度信息以用于呈现音频。

用于在不同回放设备之间优化响度和动态范围的系统和方法

[0001] 本申请是申请号为201810859612.4、申请日为2014年1月15日、发明名称为“用于在不同回放设备之间优化响度和动态范围的系统和方法”的发明专利申请的分案申请，该申请号为201810859612.4的发明专利申请是申请号为201480005314.9、申请日为2014年1月15日、发明名称为“用于在不同回放设备之间优化响度和动态范围的系统和方法”的发明专利申请的分案申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求以下申请的优先权：2013年1月21日提交的第61/754882号的美国临时申请、2013年4月5日提交的第61/809250号的美国临时申请；以及2013年5月16日提交的第61/824010号的美国临时专利申请，所有这些申请的全部内容都通过引用并入本文。

技术领域

[0004] 一个或多个实施例总体上涉及音频信号处理，并且更具体地涉及基于回放环境和设备处理具有元数据的音频数据位流，该元数据指示音频内容的响度和动态范围特性。

背景技术

[0005] 背景技术章节中所讨论的主题不应仅因其在本章节中被提及就假定其作为现有技术。类似地，在背景技术章节中所提及的或者与背景技术章节的主题相关联的问题不应被假定已经在任何现有技术中被认识到。背景技术章节中的主题仅表示不同的方法，它们本身也可能是发明。

[0006] 音频信号的动态范围通常是信号中所体现的声音的最大可能值和最小可能值之间的比率，并且通常被测量为分贝值（基于10）。在许多音频处理系统中，动态范围控制（或者动态范围压缩）被用于减少响声音水平，和/或放大安静声音水平以使得宽动态范围源内容适配可被更容易地使用电子设备存储和再现的较窄的被记录的动态范围。对于视听（AV）内容，对话参考水平可被用于定义用于通过DRC机制压缩的“零”点。DRC用于增大低于对话参考水平的内容且切除高于该参考水平的内容。

[0007] 在已知的音频编码系统中，与音频信号相关联的元数据被用于基于内容的类型和预期用途来设定DRC水平。DRC模式设定要应用于音频信号的压缩量，并且限定解码器的输出参考水平。这样的系统可被限制于两个DRC水平设置，这两个DRC水平设置被编程到编码器中并且被用户选择。例如，常规地，对于在支持AVR或者全动态范围的设备上回放的内容使用对话归一（dialnorm）（对话归一化）值-31dB，而对于在电视机或类似设备上回放的内容使用对话归一（对话归一化）值-20dB。这种类型的系统允许单个音频位流通过使用两组不同的DRC元数据用于两个常见的但是大不相同的回放场景。但是，这样的系统局限于被预先设定的对话归一值，而对于随着数字媒体和基于互联网的流送技术的出现而成为可能的多种不同回放设备和收听环境中的回放未被优化。

[0008] 在当前的基于元数据的音频编码系统中，音频数据流可包括音频内容（例如，音频内容的一个或多个通道）和指示音频内容的至少一个特性的元数据。例如，在AC-3位流中，

存在具体地预期用于改变被输送至收听环境的节目的声音的若干音频元数据参数。元数据参数之一是对话归一参数,其指示在音频节目中出现的对话的平均响度水平(或者内容的平均响度),并且被用于确定音频回放信号水平。

[0009] 在包含不同音频节目分段(每个音频节目分段具有不同的对话归一参数)的序列的位流的回放期间,AC-3解码器使用每一分段的对话归一参数来执行一种响度处理,该响度处理修正该分段的回放水平或响度,使得该分段的对话的感知响度处于一致水平。编码音频分段(项)的序列中的每一个编码音频项通常将具有不同的对话归一参数,并且解码器将缩放每一项的水平,使得用于每一项的对话的回放水平或响度相同或非常相近,但是这可能需要在回放期间对于不同项添加不同增益量。

[0010] 在一些实施例中,对话归一参数由用户设定,而不是自动生成,但是在用户没有设定值的情况下存在默认的对话归一值。例如,内容创建者可通过AC-3编码器外部的设备来进行响度测量,然后将结果(指示音频节目的口语对话的响度)传递至编码器以设定对话归一值。因此,依赖于内容创建者正确地设定对话归一参数。

[0011] 关于AC-3位流中的对话归一参数可能不正确,存在若干不同的理由。首先,每个AC-3编码器具有默认对话归一值,如果对话归一值没有被内容创建者设定,则在位流生成期间使用该默认对话归一值。此默认值可能明显不同于音频的实际对话响度水平。其次,即使内容创建者测量响度并且相应地设定对话归一值,不符合所建议的响度测量方法的响度测量算法或者测量计可能已被使用,这导致不正确的对话归一值。第三,即使AC-3位流已通过被内容创建者正确地测量和设定的对话归一值来创建,但是该对话归一值在位流的传输和/或存储期间可能已由于中间模块而改变为不正确的值。例如,在电视广播应用中,常见地是AC-3位流通过使用不正确的对话归一元数据信息被解码、修正并然后重新编码。因此,AC-3位流中包括的对话归一值可能是不正确的或不准确的,因此可能对于收听体验的质量造成不利影响。

[0012] 此外,对话归一参数没有指示对应音频数据的响度处理状态(例如,已对于音频数据执行的响度处理的类型)。另外,当前采用的响度和DRC系统(诸如,Dolby Digital (DD) 和 Dolby Digital Plus (DD+) 系统中的系统)被设计为在消费者起居室或者影院中呈现AV内容。为了使得这样的内容适合于其它环境和收听设备(例如,移动设备)中回放,必须在回放设备中“盲目”应用后处理以使得AV内容适合于该收听环境。换句话说,后处理器(或者解码器)假定所接收的内容的响度水平处于特定水平(例如,-31dB或-20dB),并且后处理器将该水平设定为适合于特定设备的预先确定的固定目标水平。如果所假定的响度水平或者预先确定的目标水平是不正确的,则后处理可能具有与其的预期效果相反的效果,即后处理可能使得输出音频低于用户的期望。

[0013] 所公开的实施例不局限于用于AC-3位流,E-AC-3位流或者Dolby E位流,但是为了方便起见,这样的位流将结合包括响度处理状态元数据的系统被讨论。Dolby、Dolby Digital、Dolby Digital Plus、以及Dolby E是杜比实验室特许公司的商标,杜比实验室提供了分别已知为Dolby和Dolby Digital的AC-3和E-AC-3的专有实现。

发明内容

[0014] 实施例涉及一种用于解码音频数据的方法,该方法接收包含与音频数据相关联的

元数据的位流，并且分析位流中的元数据以确定是否在位流中可获得用于第一组音频回放设备的响度参数。响应于确定对于该第一组存在该参数，处理组件使用该参数和音频数据来呈现音频。响应于确定对于该第一组不存在该参数，则处理组件分析该第一组的一个或更多个特性，并且基于该一个或更多个特性来确定参数。该方法可进一步通过将参数和音频数据传输至呈现音频以供回放的下游模块，使用该参数和音频数据来呈现音频。参数和音频数据还可被用于通过基于参数和音频数据呈现音频数据来呈现音频。

[0015] 在一个实施例中，该方法还包括确定将呈现所接收的音频流的输出设备，并且确定该输出设备是否属于第一组音频回放设备；其中，分析流中的元数据以确定是否可获得第一组音频回放设备的响度参数的步骤在确定该输出设备属于第一组音频回放设备的步骤之后被执行。在一个实施例中，确定该输出设备属于第一组音频回放设备的步骤包括：接收来自与输出设备连接的模块的指示输出设备的身份 (identity) 或者指示包括该输出设备的一组设备的身份的指示，并且基于所接收到的指示确定输出设备是否属于第一组音频回放设备。

[0016] 实施例进一步涉及包括执行上文编码方法实施例中描述的动作的处理组件的装置或系统。

[0017] 实施例进一步涉及一种音频数据解码方法，该方法接收音频数据和与该音频数据相关联的元数据，分析位流中的元数据以确定在该位流中是否可获得与第一组音频设备的响度参数相关联的响度信息，并且响应于确定对于该第一组存在该响度信息，从该位流确定响度信息，并且传输音频数据和响度信息以用于呈现音频，或者如果对于该第一组不存在该响度信息，确定与输出简档相关联的响度信息，并且传输所确定的该输出简档的响度信息以用于呈现音频。在一个实施例中，确定与输出简档相关联的响度信息的步骤可进一步包括分析输出简档的特性，基于该特性确定参数，并且传输所确定的响度信息包括传输所确定的参数。响度信息可包括输出简档的响度参数或者输出简档的特性。在一个实施例中，该方法可进一步包括确定要被传输的低位率编码流，其中响度信息包括一个或更多个输出简档的特性。

[0018] 实施例进一步涉及包括执行上文解码方法实施例中描述的动作的处理组件的装置或系统。

附图说明

[0019] 在以下附图中相似的附图标记被用于指示相似元件。尽管以下附图描绘了各种示例，文中所描述的实现不局限于附图中所描绘的示例。

[0020] 图1是根据一些实施例的被配置用于执行响度和动态范围的优化的音频处理系统的实施例的框图；

[0021] 图2是根据一些实施例的图1的系统中所使用的编码器的框图。

[0022] 图3是根据一些实施例的图1的系统中所使用的解码器的框图。

[0023] 图4是AC-3帧的图示，包括AC-3帧分割成多个分段。

[0024] 图5是AC-3帧的同步信息 (SI) 分段的图示，包括AC-3帧分割成多个分段。

[0025] 图6是AC-3帧的位流信息 (BSI) 分段的图示，包括AC-3帧分割成多个分段。

[0026] 图7是E-AC-3帧的图示，包括E-AC-3帧分割成多个分段。

- [0027] 图8是示出根据一些实施例的编码位流的某些帧和元数据的格式的表。
- [0028] 图9是示出根据一些实施例的响度处理状态元数据的格式的表。
- [0029] 图10是根据一些实施例的可被配置为执行响度和动态范围的优化的图1的音频处理系统的更详细的框图。
- [0030] 图11是示出在示例性使用情况中关于各种回放设备和背景收听环境的不同动态范围要求的表。
- [0031] 图12是根据实施例的动态范围优化系统的框图。
- [0032] 图13是根据一些实施例的用于各种不同回放设备类别的不同简档之间的接口的框图。
- [0033] 图14是示出根据实施例的多种定义的简档的长期响度和短期动态范围之间的关联性的表。
- [0034] 图15示出根据实施例的用于不同类型的音频内容的响度简档的示例。
- [0035] 图16是示出根据实施例的在回放设备和应用之间优化响度和动态范围的方法的流程图。

具体实施方式

[0036] 定义和命名法

[0037] 在本公开的上下文中、包括在权利要求书中，表述“对信号或数据执行操作(例如对信号或数据进行滤波、缩放、变换或者应用增益)”被在广义上使用以指示直接对信号或数据执行该操作，或者对信号或数据的处理后的版本(例如，在执行该操作之前经受了初步滤波或者预处理的信号的版本)执行该操作。表述“系统”被在广义上使用以指示设备、系统或者子系统。例如，实现解码器的子系统可被称为解码器系统，包括这样的子系统的系统(例如，响应于多个数据生成X输出信号的系统，其中该子系统生成输入中的M个输出而另外的X-M个输入被从外部源接收到)也可被称为解码器系统。术语“处理器”被在广义上使用以指示可编程或者可被配置为(例如，通过软件或固件)对数据(例如，音频、视频或其它图像数据)执行操作的系统或设备。处理器的示例包括现场可编程门阵列(或者，其它可编程集成电路或芯片组)、被编程和/或另外配置为对音频或者其它声音数据执行流水线处理的数字信号处理器、可编程通用处理器或计算机、以及可编程微处理器芯片或芯片组。

[0038] 表述“音频处理器”和“音频处理单元”被互换地使用，并且在广义上指示被配置为处理音频数据的系统。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、后处理系统和位流处理系统(有时被称为位流处理工具)。表述“处理状态元数据”(例如，在表述“响度处理状态元数据”中)指的是来自对应音频数据的分离的和不同的数据(也包括处理状态元数据的音频数据流的音频内容)。处理状态元数据与音频数据相关联，指示对应音频数据的响度处理状态(例如，对于音频数据已经执行了什么类型的处理)，并且可选地还指示音频数据的至少一个特征或特性。在一些实施例中，处理状态元数据与音频数据的关联性是时间同步的。因此，当前的(最新接收的或者更新的)处理状态元数据指示对应的音频数据同时地包括所指示类型的音频数据处理的结果。在一些情况下，处理状态元数据可包括处理历史和/或所指示类型的处理中所使用的和/或从其导出的参数的一些或全部。另外，处理状态元数据可包括对应音频数据的已经从音频数据被

计算或提取的至少一个特征或特性。处理状态元数据还可包括与对应音频数据的任何处理无关的或者不从其导出的其它元数据。例如,第三方数据、跟踪信息、标识符、专有或标准信息、用户注释数据、用户偏好数据等可通过特定音频处理单元被添加以传递至其它音频处理单元。

[0039] 表述“响度处理状态元数据”(或者“LPSM”)指示如下这样的处理状态元数据,其指示对应音频数据的响度处理状态(例如,对于音频数据已经执行了什么类型的处理),并且可选地还指示对应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包括不是响度处理状态元数据的数据(例如,在被单独考虑时)。术语“耦合”或者“被耦合”用于表示直接或者间接连接。

[0040] 针对如下的音频编码器/解码器描述系统和方法,该音频编码器/解码器在需要或者使用不同的目标响度值并且具有不同的动态范围能力的各种设备之间非破坏性地归一化音频的响度和动态范围。根据一些实施例的方法和功能组件针对一个或多个设备简档将关于音频内容的信息从编码器发送至解码器。设备简档指定了一个或多个设备的所希望的目标响度和动态范围。系统是可扩展的,从而可支持具有不同的“标称”响度目标的新设备简档。

[0041] 在一个实施例中,系统在编码器中基于响度控制和动态范围要求生成合适的增益,或者在来自编码器的控制下通过用以减少数据率的原始增益的参数化、在解码器中生成增益。动态范围系统包括用于实现响度控制的两个机制:对于音频将如何回放提供内容创建器控制的艺术动态范围简档;和确保对于各种回放简档不会发生过载的单独的保护机制。系统还被配置为允许使用其它元数据(内部或者外部)参数来正确地控制响度和动态范围增益和/或简档。解码器被配置为支持将影响解码器侧响度和动态范围的n通道辅助输入。

[0042] 在一些实施例中,响度处理状态元数据(LPSM)被嵌入音频位流的元数据分段的一个或多个保留字段(或者空位),该音频位流还在其它分段(音频数据分段)中包括音频数据。例如,位流的每个帧的至少一个分段包括LPSM,并且该帧的至少一个其它分段包括对应的音频数据(即,其响度处理状态和响度由LPSM指示的音频数据)。在一些实施例中,LPSM的数据量可足够小以在不影响被分配用于运送音频数据的位率的情况下被运送。

[0043] 在音频数据处理链中传递响度处理状态元数据在两个或更多个音频处理单元需要在处理链(或者内容生命周期中)相互串联地工作的情况下尤其有用。在音频位流中不包括响度处理状态元数据的情况下,可能发生诸如质量、水平和空间劣化的媒体处理问题,例如,当在该链中应用两个或更多个音频编解码器并且在位流传送至媒体消费设备(或者位流的音频内容的呈现点)期间单端音量调节被应用一次以上时会出现这样的问题。

[0044] 响度和动态范围元数据处理系统

[0045] 图1是根据使用某些元数据处理(例如,预处理和后处理)组件的一些实施例的可被配置为执行响度和动态范围的优化的音频处理系统的一个实施例的框图。图1示出了示例性音频处理链(音频数据处理系统),其中系统的元件中的一个或多个可根据本发明的实施例被配置。图1的系统10包括如图所示耦合到一起的以下元件:预处理单元12、编码器14、信号分析和元数据校正单元16、代码转换器18、解码器20和后处理单元24。在所示的系统的变型中,元件中的一个或多个被省略,或者包含额外的音频数据处理单元。例如,在一个实

施例中,后处理单元22是解码器20的一部分而不是单独的单元。

[0046] 在一些实现中,图1的预处理单元被配置为接受包含音频内容的PCM(时域)采样作为输入,并且输出经处理的PCM采样。编码器14可被配置为接受PCM采样作为输入,并且输出指示音频内容的编码的(例如,压缩的)音频位流。指示音频内容的位流的数据有时在此被称为“音频数据”。在一些实施例中,从编码器输出的音频位流包括响度处理状态元数据(以及可选的其它元数据)和音频数据。

[0047] 信号分析和元数据校正单元16可接受一个或多个编码的音频位流作为输入,并且通过执行信号分析来确定(例如,验证)每个编码的音频位流中的处理状态元数据是否正确。在一些实施例中,验证可由状态验证器组件(诸如图2中所示的元件102)执行,并且下文在状态验证器102的上下文中描述一种这样的验证技术。在一些实施例中,单元16被包含在编码器中,并且验证由单元16或验证器102执行。如果信号分析和元数据校正单元发现包含的元数据是无效的,则元数据校正单元16执行信号分析以确定正确的值,并且用所确定的正确的值来替换不正确的值。因此,来自信号分析和元数据校正单元的每个编码的音频位流输出可包括校正后的处理状态元数据以及编码的音频数据。信号分析和元数据校正单元16可以是预处理单元12、编码器14、代码转换器18、解码器20或后处理单元22的一部分。作为替代,信号分析和元数据校正单元16可以是音频处理链中的单独的单元或另一单元的一部分。

[0048] 代码转换器18可接受编码的音频位流作为数据,并且作为响应而输出修正的(或者,被不同地编码的)音频位流(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位流重新编码)。从代码转换器输出的音频位流包括响度处理状态元数据(以及可选的其它元数据)和编码的音频数据。元数据已经被包含在位流中。

[0049] 图1的解码器20可接受编码的(例如,压缩的)音频位流作为输入,并且输出(作为响应)解码PCM音频采样的流。在一个实施例中,解码器的输出为或者包括以下中的任一个:音频采样流以及从输入的编码位流提取的响度处理状态元数据(和可选的其它元数据)的对应流;音频采样的流以及由从输入的编码位流提取的响度处理状态元数据(和可选的其它元数据)确定的控制位的对应流;或者音频采样流,而不具有处理状态元数据或者由处理状态元数据确定的控制位的对应流。在此最后的情况下,解码器可从输入的编码位流提取响度处理状态元数据(和/或其它元数据),并且对所提取的元数据执行至少一种操作(例如,验证),但是其并不输出所提取的元数据或从其确定的控制位。

[0050] 通过根据本发明的实施例配置图1的后处理单元,后处理单元22被配置为接受解码PCM音频采样的流,并且使用通过采样接收的响度处理状态元数据(以及可选的其它元数据)或者通过采样接收的控制位(由解码器从响度处理状态元数据以及可选的其它元数据确定),对解码PCM音频采样的流执行后处理(例如,音频内容的音量调节)。后处理单元22可选地还可被配置为呈现经后处理的音频内容以供一个或多个扬声器回放。这些扬声器可被体现在各种不同的收听设备或回放设备(诸如计算机、电视、立体声系统(家用或者影院用)、移动电话或者其它便携式回放设备)中。扬声器可以具有任何合适的大小和额定功率,并且可被以独立驱动器、扬声器箱、环绕声系统、条式音箱、耳机、耳塞等形式提供。

[0051] 一些实施例提供了增强的音频处理链,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和后处理单元)根据由音频处理单元分别接收的响度处理状态元数

据指示的媒体数据的同时状态 (contemporaneous state) 改变它们各自的要被应用于音频数据的处理。系统100的任何音频处理单元 (例如, 图1的编码器或代码转换器) 的音频数据输入11可包括响度处理状态元数据 (以及可选的其它元数据) 和音频数据 (例如, 编码音频数据)。根据一些实施例, 此元数据可通过其它元件或其它源被包含在输入音频中。接收输入音频 (具有元数据) 的处理单元可被配置为对元数据执行至少一种操作 (例如, 验证) 或者响应于元数据 (例如, 输入音频的自适应处理), 并且可选地还被配置为在其输出音频中包括元数据、元数据的处理后版本或者从元数据确定的控制位。

[0052] 音频处理单元 (或者音频处理器) 的实施例被配置为基于与音频数据对应的响度处理状态元数据指示的音频数据的状态, 执行音频数据的自适应处理。在一些实施例中, 自适应处理为 (或者包括) 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者类似处理) 但是不是 (或者不包括) 响度处理 (如果元数据指示已经对于音频数据执行响度处理或者类似处理)。在一些实施例中, 自适应处理为或者包括元数据验证 (例如, 在元数据验证子单元中执行), 以确保音频处理单元基于响度处理状态元数据指示的音频数据的状态执行音频数据的其它自适应处理。在一些实施例中, 验证确定了与音频数据相关联的 (例如, 被包含在位流中的) 响度处理状态元数据的可靠性。例如, 如果元数据被验证为是可靠的, 则一种先前执行的音频处理的结果可被重新使用, 而同一类型的音频处理的额外执行可被避免。另一方面, 如果发现元数据已被篡改 (或者不可靠), 则 (如由不可靠元数据指示的) 据称先前执行的该种媒体处理可被音频处理单元重复进行, 和/或音频处理单元可对元数据和/或音频数据执行其它处理。音频处理单元还可被配置为如果音频处理单元确定响度处理状态元数据有效 (例如, 基于提取的密码值与参考密码值匹配), 则将响度处理状态元数据 (例如, 存在于媒体位流中) 有效用信号发送给增强媒体处理链中处于下游的其它音频处理单元。

[0053] 对于图1的实施例, 预处理组件12可以是编码器14的一部分, 并且后处理组件22可以是解码器22的一部分。作为替代, 预处理组件12可表现为与编码器14分开的功能组件。类似的, 后处理组件22可表现为与解码器22分开的功能组件。

[0054] 图2是可与图1的系统10结合使用的编码器100的框图。编码器100的任何组件可被以硬件、软件或者硬件和软件的组合实现为一个或多个处理和/或一个或多个电路 (例如, ASIC、FPGA或者其它集成电路)。编码器100包括帧缓冲器110、解析器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/格式化器级107、元数据生成级106、对话响度测量子系统108和帧缓冲器109, 它们如图所示地连接。还可选地, 编码器100包括其它处理元件 (未示出)。编码器100 (是代码转换器) 被配置为将输入音频位流 (例如可以是AC-3位流、E-AC-3位流或者Dolby E位流之一) 转换成编码的输出音频位流 (例如可以是AC-3位流、E-AC-3位流或者Dolby E位流中的另一个), 包括使用输入位流中包含的响度处理状态元数据执行自适应和自动响度处理。例如, 编码器100可配置为将输入Dolby E位流 (典型地在制作和广播设施中使用的、但是没有在接收被向其广播的音频节目的消费类设备中使用的格式) 转换成AC-3或E-AC-3格式的编码的输出音频位流 (适合于广播至消费类设备)。

[0055] 图2的系统还包括编码音频输送系统150 (其存储和/或输送从编码器100输出的编码位流) 和解码器152。从编码器100输出的编码音频位流可被子系统150存储 (例如, 以DVD

或BluRay盘的形式),或者被子系统150传输(其可实现传输链路或网络),或者可被子系统150存储和传输。解码器152被配置为对其经由子系统150接收的编码位流(由编码器100生成)进行解码,包括从位流的每一帧提取响度处理状态元数据(LPSM),并且生成解码的音频数据。在一个实施例中,解码器152被配置为使用LPSM对解码的音频数据执行自适应响度处理,和/或将解码的音频数据和LPSM转发至后处理器,该后处理器被配置为使用LPSM对解码的音频数据执行自适应响度处理。可选地,解码器152包括缓冲器,该缓冲器存储(例如,以非暂态的方式)从子系统150接收的编码音频位流。

[0056] 编码器100和解码器152的各种实现被配置为执行文中所描述的各种实施例。帧缓冲器110是被耦合以接收编码的输入音频位流的缓冲存储器。在操作中,缓冲器110存储(例如,以非暂态的方式)编码音频位流中的至少一帧,并且编码音频位流的帧序列被从缓冲器110断言至解析器111。解析器111被耦合和配置为从编码的输入音频中的每一帧提取响度处理状态元数据(LPSM)和其它元数据,以将至少LPSM断言至音频状态验证器102、响度处理级103、级106和子系统108,以从编码的输入音频提取音频数据,并且将音频数据断言至解码器101。编码器100的解码器101被配置为解码音频数据以生成解码的音频数据,并且将解码的音频数据断言至响度处理级103、音频流选择级104、子系统108并且可选地还断言至状态验证器102。

[0057] 状态验证器102被配置为认证和验证被断言至状态验证器102的LPSM(以及可选的其它元数据)。在一些实施例中,LPSM为(或者被包括在)已经包含在输入位流中的数据块中(例如,根据本发明的实施例)。该块可包括用于处理LPSM(以及可选的其它元数据)和/或基层音频数据(从解码器101提供给验证器102)的加密散列(哈希消息认证码或者“HMAC”)。在这些实施例中,数据块可被数字签名,从而下游的音频处理单元可相对容易地认证和验证处理状态元数据。

[0058] 例如,HMAC被用于生成摘要(digest),并且本发明的位流中包括的保护值可包括该摘要。该摘要可对于AC-3帧被如下地生成:(1)在AC-3数据和LPSM被编码之后,帧数据字节(链接帧数据#1和帧数据#2)和LPSM数据字节被用作哈希函数HMAC的输入。对于摘要计算没有考虑可存在于辅助数据(auxdata)字段内的其它数据。这样的其它数据可以是即不属于AC-3数据也不属于LPSM数据的字节。对于HMAC摘要计算可能没有考虑LPSM中包含的保护位。(2)在摘要被计算之后,摘要被写入位流中为保护位保留的字段中。(3)生成完整的AC-3帧的最后一步是计算CRC校验位。这被写入帧的最后端,并且属于此帧的所有数据(包括LPSM位)被加以考虑。

[0059] 包括但不限于一种或多种非HMAC加密方法中的任一种方法的其它加密方法可被用于验证LPSM(例如,在验证器102中),以确保LPSM和/或基层音频数据的安全传输和接收。例如,验证(使用这样的加密方法)可在接收发明的音频位流的实施例的每个音频处理单元中被执行,以确定是否位流中包含的响度处理状态元数据和对应的音频数据已经受了(如元数据指示的)特定响度处理(和/或从其得到)并且在执行这样的特定响度处理之后没有被修正。

[0060] 状态验证器102将控制数据断言至音频流选择级104、元数据生成器106和对话相对测量子系统108,以指示验证操作的结果。响应于控制数据,级104可选择以下中的任一个(并且将其传递给编码器105):(1)响度处理级103的自适应处理输出(例如,当LPSM指示解

码器101输出的音频数据没有经受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);或者(2)从解码器101输出的音频数据(例如,当LPSM指示解码器101输出的音频数据已经受了由级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)。在一个实施例中,响度处理级103校正响度为所规定的目地和响度范围。

[0061] 编码器100的级103被配置为基于由解码器101所提取的LPSM指示的一个或多个音频数据特性,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可接收用户输入(例如,用户目标响度/动态范围值或对话归一值),或者其它元数据输入(例如,一种或多种第三方数据、跟踪信息、标识符、专有或标准信息、使用注释数据、用户偏好数据等),和/或其它输入(例如,来自指纹识别处理),并且使用这样的输入来处理从解码器101输出的解码音频数据。

[0062] 对话响度测量子系统108可操作以当来自验证器102的控制位指示LPSM无效时,例如使用由解码器101提取的LPSM(和/或其它元数据)确定(来自解码器101的)解码音频的指示对话(或者其它语音)的分段的响度。当LPSM指示先前确定的(来自解码器101的)解码音频的对话(或其它语音)分段的响度时,当来自验证器102的控制位指示LPSM有效时,对话响度测量子系统108的操作可被禁用。

[0063] 存在用于方便和容易地测量音频内容中的对话的水平的有用工具(例如,Dolby LM100响度计)。APU(例如编码器100的级108)的一些实施例被实现为包括这样的工具(或者执行其功能),以测量音频位流(例如,被从编码器100的解码器101断言至级108的所解码的AC-3位流)的音频内容的平均对话响度。如果级108被实现为测量音频数据的真实平均对话响度,测量可包括隔离音频内容的主要包含语音的分段的步骤。主要为语音的音频分段然后根据响度测量算法被处理。对于从AC-3位流解码的音频数据,此算法可以是标准K加权响度测量(根据国际标准ITU-R BS.1770)。作为替代,其它响度测量可被使用(例如,基于响度的心理声学模型的那些)。

[0064] 语音分段的隔离不是测量音频数据的平均对话响度所必需的。但是,其提高了测量的精度,并且提供了从收听者的角度看更满意的结果。由于并非所有音频内容都包含对话(语音),因此整个音频内容的响度测量可提供音频的对话水平的充分近似(在语音存在的情况下)。

[0065] 元数据生成器106生成要通过级107被包含在要从编码器100输出的编码位流中的元数据。元数据生成器106可将由编码器101提取的LPSM(和/或其它元数据)传递至级107(例如,当来自验证器102的控制位指示LPSM和/或其它元数据有效时),或者生成新的LPSM(和/或其它元数据)并且将新的LPSM断言至级107(例如,当来自验证器102的控制位指示通过解码器101提取的LPSM和/或其它元数据无效时),或者其可将通过解码器101提取的元数据和新生成的元数据的组合断言至级107。元数据生成器106可将通过子系统108生成的响度数据和指示子系统108执行的响度处理的类型的至少一个值包含在元数据生成器106断言至级107的LPSM中,以便包含在要从编码器100输出的编码位流中。元数据生成器106可生成保护位(可由哈希消息认证码或“HMAC”构成或者包括HMAC),该保护位对于要被包含在编码位流中的LPSM(和/或其它元数据)和/或要被包含在编码位流中的基层音频数据的解密、认证或验证中的至少一个是有用的。元数据生成器106可将这样的保护位提供给级107以便包含在编码位流中。

[0066] 在一个实施例中,对话响度测量子系统108处理从解码器101输出的音频数据,以响应于其生成响度值(例如,被门控或者未被门控的对话响度值)和动态范围值。响应于这些值,元数据生成器106可生成响度处理状态元数据(LPSM)以便(通过填充器/格式化器107)包含在要从编码器100输出的编码位流中。在一个实施例中,响度可基于ITU-R BS.1770-1和ITU-R BS.1770-2标准规定的响度测量技术、或者其它相似的响度测量标准被计算。门控响度(gated loudness)可以是对话门控响度或者相对门控响度,或者这些门控响度类型的组合,并且系统可根据应用要求和系统约束来采用合适的门控块。

[0067] 附加地、可选地或作为替代地,编码器100的子系统106和/或108可执行音频数据的附加分析以生成指示音频数据的至少一个特性的元数据,以便包含在要从级107输出的编码位流中。编码器105对从选择级104输出的音频数据进行编码(例如,通过对其进行压缩),并且将编码音频断言至级107以便包含在要从级107输出的编码位流中。

[0068] 级107对来自编码器105的编码音频和来自生成器106的元数据(包括LPSM)进行复用,以生成要从级107输出的编码位流,从而编码位流具有如实施例所指定的格式。帧缓冲器109是缓冲存储器,其存储(例如,以非暂态的方式)从级107输出的编码音频位流中的至少一帧,然后,编码音频位流的帧的序列被从缓冲器109作为编码器100的输出断言至输送系统150。

[0069] 元数据生成器106生成的且通过级107被包含在编码位流中的LPSM指示对应音频数据的响度处理状态(例如,对于音频数据已经执行的响度处理的类型)、和对应音频数据的响度(例如,所测量的对话响度、门控和/或非门控响度、和/或动态范围)。这里,对于音频数据执行的响度和/或水平测量的“门控”指的是特定水平或响度阈值,在最终测量中包含超过阈值的计算值(例如,在最终测量值中忽略低于-60dBFS的短期响度值)。绝对值的门控指的是固定水平或响度,而相对值的门控指的是依赖于当前“未门控”测量值的值。

[0070] 在编码器100的一些实现中,存储器109中缓冲(并且被输出至输送系统150)的编码位流是AC-3位流或E-AC-3位流,并且包括音频数据分段(例如,图4所示的帧的AB0-AB5分段)和元数据分段,其中音频数据分段指示音频数据,并且元数据分段中的至少一些分段中的每一个包括响度处理状态元数据(LPSM)。级107按以下格式将LPSM插入位流。包括LPSM的元数据分段中的每一个被包括在位流的帧的位流信息(“BSI”)分段的“addbsi”字段中,或者在位流的帧的结尾处的auxdata字段(例如,图4中所示的AUX分段)中。

[0071] 位流的帧可包括一个或两个元数据分段,其中的每一个元数据分段包括LPSM,并且如果帧包括两个元数据分段,则一个存在于帧的addbsi字段中而另一个存在于帧的AUX字段中。包括LPSM的每个元数据分段包括具有以下格式的LPSM载荷(或者容器)分段:头部(例如,包括标识LPSM载荷的开始的同步字,随后是至少一个标识值,例如在以下的表2中指示的LPSM格式版本、长度、周期、计数和子流关联值);以及在头部之后,指示对应的音频数据是否指示对话还是不指示对话(例如,对应音频数据的哪个通道指示对话)的至少一个对话指示值(例如,表2的参数“对话通道”);指示对应的音频数据是否遵从所指示的响度规则的集合的至少一个响度规则遵从值(例如,表2的参数“响度规则类型”);指示对于对应的音频数据执行的至少一种响度处理的至少一个响度处理值(例如,表2中的参数“对话门控响度校正标记”、“响度校正类型”中的一个或多个);以及指示对应音频数据的至少一个响度(例如,峰值响度或平均响度)特性的至少一个响度值(例如,表2的参数“ITU相对门控响度”、

“ITU语音门控响度”、“ITU(EBU 3341)短期3s响度”和“真实峰值”中的一个或多个)。

[0072] 在一些实现中,通过级107插入位流的帧的“addbsi”字段或auxdata字段的元数据分段中的每一个具有以下格式:核心头部(例如,包括指示元数据分段的开始的同步字,随后是标识值,例如在以下的表1中指示的核心元素版本、长度、周期、扩展元素计数和子流关联值);以及在核心头部之后,对于响度处理状态元数据或对应音频数据中的至少一个的解密、认证或验证中的至少一个有用的至少一个保护值(例如,表1的HMAC摘要和音频指纹值);以及也在核心头部之后的、在元数据分段包括LPSM的情况下LPSM载荷标识(“ID”)和LPSM载荷大小值,其将随后的元数据标识为LPSM载荷并且指示LPSM载荷的大小。

[0073] LPSM载荷(或容器)分段(例如,具有上文指定的格式)在LPSM载荷ID和LPSM载荷大小值之后。

[0074] 在一些实施例中,帧的auxdata字段(或者“addbsi”字段)中的元数据分段中的每一个具有三个层级的结构:高层级结构,包括指示auxdata(或者addbsi)字段是否包括元数据的标记,指示存在的元数据的类型的至少一个ID值、以及可选地还包括指示存在(例如,每种类型的)元数据的多少位(在元数据存在的情况下)的值。可存在的一种类型的元数据是LPSM,并且可存在的另一类型的元数据是媒体调查元数据(例如,尼尔森媒体调查元数据);中间层级结构,包括每一标识的类型的元数据的核心元素(例如,每一标识的类型的元数据的如上文提及的核心头部、保护值和LPSM载荷ID和LPSM载荷大小值);和低层级结构,包括一个核心元素的每一载荷(例如,在被核心元素标识为存在的情况下LPSM载荷,或者在被核心元素标识为存在的情况下另一类型的元数据载荷)。

[0075] 这样的三层级结构中的数据值可嵌套(nest)。例如,核心元素标识的LPSM载荷和/或另一元数据载荷的保护值可被包含在由核心元素标识的每个载荷之后(因此,在核心元素的核心头部之后)。在一个示例中,核心头部可标识LPSM载荷和另一元数据载荷,第一载荷(例如,LPSM载荷)的载荷ID和载荷大小值可在核心头部之后,第一载荷本身可在ID和大小值之后,第二载荷的载荷ID和载荷大小可在第一载荷之后,第二载荷本身可在这些ID和大小值之后,并且这两个载荷的(或者核心元素值和这两个载荷的)保护值可在最后一个载荷之后。

[0076] 在一些实施例中,如果解码器101接收具有加密散列的根据本发明的实施例生成的音频位流,则解码器被配置为从由该位流确定的数据块解析并取得该加密散列,所述块包括响度处理状态元数据(LPSM)。验证器102可使用加密散列来验证所接收的位流和/或相关的元数据。例如,验证器102基于参考加密散列和从数据块取得的加密散列的匹配发现LPSM有效,然后可禁止处理器103对对应的音频数据操作,并且使得选择级104传递(未改变的)视频数据。附加地,可选地或者作为替代地,可使用其它类型的加密技术以替代基于加密散列的方法。

[0077] 图2的编码器100可(响应于通过解码器101提取的LPSM)确定后处理/预处理单元已经对要被编码的音频数据执行了响度处理(在元件105、106和107中),因此(在生成器106中)可创建包括在先前执行的响度处理中使用的和/或从先前执行的响度处理导出的特定参数的响度处理状态元数据。在一些实施例中,编码器100可创建(并且在其输出的编码位流中包括)指示音频内容的处理历史的处理状态元数据,至少编码器知晓已对于音频内容执行的处理的类型即可。

[0078] 图3是可与图1的系统10结合使用的解码器的框图。解码器200和后处理器300的组件或元件中的任一个可被以硬件、软件或者硬件和软件的组合实现为一个或多个处理和/或一个或多个电路(例如,ASIC、FPGA或者其它集成电路)。解码器200包括帧缓冲器201、解析器205、音频解码器202、音频状态验证级(验证器)203和控制位生成级204,它们如图所示地连接。解码器200包括其它处理元件(未示出)。帧缓冲器201(缓冲存储器)存储(例如,以非暂态的方式)解码器200接收的编码音频位流中的至少一帧。编码音频位流的帧序列被从缓冲器201断言至解析器205。解析器205被耦合和配置为从编码的输入音频中的每一帧提取响度处理状态元数据(LPSM)和其它元数据,以将至少LPSM断言至音频状态验证器203和级204,将LPSM断言为输出(例如,到后处理器300),从编码的输入音频提取音频数据,并且将所提取的音频数据断言至解码器202。输入解码器200的编码音频位流可以是AC-3位流、E-AC-3位流或者DolbyE位流之一。

[0079] 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和包括耦合到缓冲器301的至少一个处理元件的其它处理元件(未示出)。帧缓冲器301存储(例如,以非暂态的方式)来自解码器200的后处理器300接收的解码音频位流中的至少一帧。后处理器300的处理元件被耦合并且配置为接收并且使用从解码器202输出的元数据(包括LPSM值)和/或从解码器200的级204输出的控制位自适应地处理从缓冲器301输出的解码音频位流的帧的序列。在一个实施例中,后处理器300被配置为使用LPSM值(例如,基于由LPSM指示的响度处理状态和/或一个或多个音频数据特性)对解码音频数据执行自适应响度处理。解码器200和后处理器300的各种实现被配置为执行根据文中所描述的实施例的方法的不同实施例。

[0080] 解码器200的音频解码器202被配置为解码通过解析器205提取的音频数据以生成解码音频数据,并且断言解码音频数据为输出(例如,至后处理器300)。状态验证器203被配置为认证和验证被断言至状态验证器203的LPSM(以及可选的其它元数据)。在一些实施例中,LPSM为(或者被包括在)已经包含在输入位流中的数据块中(例如,根据本发明的实施例)。该块可包括用于处理LPSM(以及可选的其它元数据)和/或基层音频数据(从解析器205和/或解码器202提供给验证器203)的加密散列(哈希消息认证码或者“HMAC”)。在这些实施例中,数据块可被数字签名,从而下游的音频处理单元可相对容易地认证和验证处理状态元数据。

[0081] 包括但不限于一种或多种非HMAC加密方法中的任一种方法的其它加密方法可被用于验证LPSM(例如,在验证器203中),以确保LPSM和/或基层音频数据的安全传输和接收。例如,验证(使用这样的加密方法)可在接收发明的音频位流的实施例的每个音频处理单元中被执行,以确定是否位流中包含的响度处理状态元数据和对应的音频数据已经受了(如元数据指示的)特定响度处理(和/或从其得到)并且在执行这样的特定响度处理之后没有被修正。

[0082] 状态验证器203将控制数据断言至控制位生成器204,并且断言该控制数据作为输出(例如,至后处理器300),指示验证操作的结果。响应于控制数据(以及可选的从输入位流提取的其它元数据),级204可生成以下中的任一个(并且将其断言至后处理器300):(1)指示从解码器202输出的解码音频数据已经受了特定类型的响度处理的控制位(例如,当LPSM指示解码器202输出的音频数据已经受了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);或者(2)指示从解码器202输出的解码音频数据应经受特定类型的响

度处理的控制位(例如,当LPSM指示解码器202输出的音频数据没有经受特定类型的响度处理时,或者当LPSM指示解码器202输出的音频数据已经受了特定类型的响度处理但是来自验证器203的控制位指示LPSM无效时)。

[0083] 作为替代地,解码器200将由解码器200从输入位流提取的LPSM(和任何其它元数据)断言至后处理器300,并且后处理器300使用LPSM对解码音频数据执行响度处理,或者执行LPSM的验证,然后在验证指示LPSM有效的情况下使用LPSM对解码音频数据执行响度处理。

[0084] 在一些实施例中,如果解码器201接收具有加密散列的根据本发明的实施例生成的音频位流,则解码器被配置为从由该位流确定的数据块解析并取得该加密散列,所述块包括响度处理状态元数据(LPSM)。验证器203可使用加密散列来验证所接收的位流和/或相关联的元数据。例如,如果验证器203基于参考加密散列和从数据块取得的加密散列的匹配发现LPSM有效,其可向下游音频处理单元(例如,后处理器300,其可以是或者包括音量调节单元)发信号以传递位流的(未改变的)音频数据。附加地、可选地或者作为替代地,可使用其它类型的加密技术来替换基于加密散列的方法。

[0085] 在解码器100的一些实现中,所接收的(并且在存储器201中缓冲的)编码位流是AC-3位流或E-AC-3位流,并且包括音频数据分段(例如,图4所示的帧的AB0-AB5分段)和元数据分段,其中音频数据分段指示音频数据,并且元数据分段中的至少一些分段中的每一个包括响度处理状态元数据(LPSM)。解码器级202被配置为从位流提取具有如下格式的LPSM。包括LPSM的元数据分段中的每一个被包括在位流的帧的位流信息(“BSI”)分段的“addbsi”字段中,或者在位流的帧的结尾处的auxdata字段(例如,图4中所示的AUX分段)中。位流的帧可包括一个或两个元数据分段,其中的每一个元数据分段包括LPSM,并且如果帧包括两个元数据分段,则一个存在于帧的addbsi字段中而另一个存在于帧的AUX字段中。包括LPSM的每个元数据分段包括具有以下格式的LPSM载荷(或者容器)分段:头部(例如,包括标识LPSM载荷的开始的同步字,随后是至少一个标识值,例如在以下的表2中指示的LPSM格式版本、长度、周期、计数和子流关联值);以及在头部之后,指示对应的音频数据是否指示对话还是不指示对话(例如,对应音频数据的哪个通道指示对话)的至少一个对话指示值(例如,表2的参数“对话通道”);指示对应的音频数据是否遵从所指示的响度规则的集合的至少一个响度规则遵从值(例如,表2的参数“响度规则类型”);指示对于对应的音频数据执行的至少一种响度处理的至少一个响度处理值(例如,表2中的参数“对话门控响度校正标记”、“响度校正类型”中的一个或多个);以及指示对应音频数据的至少一个响度(例如,峰值响度或平均响度)特性的至少一个响度值(例如,表2的参数“ITU相对门控响度”、“ITU语音门控响度”、“ITU(EBU 3341)短期3s响度”和“真实峰值”中的一个或多个)。

[0086] 在一些实现中,解码器级202被配置为从位流的帧的“addbsi”字段或auxdata字段提取具有以下格式的各元数据分段:核心头部(例如,包括标识元数据分段的开始的同步字,随后是至少一个标识值,例如在以下的表1中指示的核心元素版本、长度、周期、扩展元素计数和子流关联值);以及在核心头部之后,对于响度处理状态元数据或对应音频数据中的至少一个的解密、认证或验证中的至少一个有用的至少一个保护值(例如,表1的HMAC摘要和音频指纹值);以及也在核心头部之后的、在元数据分段包括LPSM的情况下的LPSM载荷标识(“ID”)和LPSM载荷大小值,其将随后的元数据标识为LPSM载荷并且指示LPSM载荷的大

小。LPSM载荷(或容器)分段(例如,具有上文指定的格式)在LPSM载荷ID和LPSM载荷大小值之后。

[0087] 更一般地,通过实施例生成的编码音频位流具有如下结构,其提供了将元数据元素和子元素标记为核心(必选(mandatory)元素)或扩展(可选元素)的机制。这允许位流(包括其元数据)的数据率在各种应用之间缩放。位流语法中的核心(必选)元素还应能够用信号告知与音频内容相关联的扩展(可选)元素存在(带内)和/或位于远程位置(带外)。

[0088] 在一些实施例中,核心元素需要存在于位流的每一帧中。核心元素的一些子元素是可选的,并且可被以任何组合存在。扩展元素不需要存在于位流的每一帧中(以限制位率开销)。因此,扩展元素可存在于一些帧中而不存在于其它帧中。扩展元素的一些子元素是可选的,并且可被以任何组合存在,而扩展元素的一些子元素可以是必选的(即,扩展元素存在于位流的帧中的情况下)。

[0089] 在一些实施例中,(例如,通过体现本发明的音频处理单元)生成包括音频数据分段和元数据分段的序列的编码音频位流。音频数据分段指示音频数据,至少一些元数据分段中的每一个包括响度处理状态元数据(LPSM),并且音频数据分段与元数据分段被时分复用。在一些实施例中,在此类型中,每个元数据分段具有文中要描述的格式。在一种格式中,编码位流是AC-3位流或者E-AC-3位流,并且包括LPSM的各元数据分段(例如,通过编码器100的级107)被作为附加位流信息包含在位流的帧的位流信息("BSI")分段的"addbsi"字段(图6所示)中或者位流的帧的auxdata字段中。每个帧包括具有图8的表1中所示的格式的帧的addbsi字段中的核心元素。

[0090] 在一种格式中,包含LPSM的addbsi(或者auxdata)字段中的每一个包含核心头部(以及可选的附加核心元素);在核心头部(或者核心头部和其它核心元素)之后,随后的LPSM值(参数):核心元素值之后的载荷ID(将元数据标识为LPSM)(例如,表1中所指示的);载荷ID之后的载荷大小(指示LPSM载荷的大小);以及具有图9的表2所示指示的格式的LPSM数据(在载荷ID和载荷大小之后)。

[0091] 在编码位流的第二格式中,位流是AC-3位流或E-AC-3位流,并且包含LPSM的各元数据分段(例如,通过编码器100的级107)被包含在如下字段中:位流的帧的位流信息("BSI")分段的"addbsi"字段,或者在位流的帧的结尾处的auxdata字段(例如,图4中所示的AUX分段)。帧可包括一个或两个元数据分段,其中的每一个元数据分段包括LPSM,并且如果帧包括两个元数据分段,则一个存在于帧的addbsi字段中而另一个存在于帧的AUX字段中。包括LPSM的每个元数据分段具有上文参照表1或表2指定的格式(即,其包括表1中规定的根本元素,随后是上文规定的载荷ID(将元数据标识为LPSM)以及载荷大小值、随后是载荷(具有表2中指示的格式的LPSM数据))。

[0092] 在另一实现中,编码位流是Dolby E位流,并且包括LPSM的元数据分段中的每一个时Dolby E保护带间隔的前N个采样位置。包括这样的元数据分段的Dolby E位流包括LPSM,例如包括指示在SMPTE 337M序言的Pd字中信号告知的LPSM载荷长度的值(SMPTE 337M Pa字重复率可保持与相关联的视频帧率相同)。

[0093] 在编码位流是E-AC-3位流的格式中,包含LPSM的各元数据分段(例如,通过编码器100的级107)被包含在位流的帧的位流信息("BSI")分段的"addbsi"字段中。附加的对此格式的具有LPSM的E-AC-3位流进行编码的方面被描述如下:(1)在E-AC-3位流生成期间,在E-

AC-3编码器(其将LPSM值插入位流)“工作”时,对于生成的每个帧(同步帧),位流应包含在帧的addbsi字段中被运送的元数据块(包括LPSM)。运送元数据块的位不应增大编码器位率(帧长度);(2)每个元数据块(包含LPSM)应含有如下信息:

[0094] loudness_correction_type_flag:其中“1”指示对应音频数据的响度在编码器上游被校正,并且“0”指示响度通过嵌入在编码器中的响度校正器被校正(例如,图2的编码器100的响度处理器103);speed_channel1指示哪个(哪些)源通道包含语音(在前0.5秒)。如果没有检测到语音,其应被这样指示;speech_loudness指示含有语音的各对应音频通道的积分语音响度(在前0.5秒);ITU_loudness指示各对应音频通道的积分ITU BS.1770-2响度;增益:用于解码器中的逆转的响度复合增益(指示可逆性)。

[0095] 当E-AC-3编码器(其将LPSM值插入位流)“工作”并且正接收具有“真”标记的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)被绕过。“可信”源对话归一和DRC值被传递(例如,通过编码器100的生成器106)至E-AC-3编码器组件(例如,编码器100的级107)。LPSM块生成继续进行,并且loudness_correction_type_flag被设定为“1”。响度控制器绕过序列被与出现“置信”标记的解码AC-3帧的开始同步。响度控制器绕过序列被如下实现:leveler_amount控制经过10个音频块周期(即,53.3毫秒)从值9减小至值0,并且leveler_back_end_meter控制被置于绕过模式中(此操作将导致无缝转换)。术语调平器的“可信”绕过暗示源位流的对话归一值还在编码器的输出被再次利用(例如,如果“可信”源位流具有对话归一值-30,然后,编码器的输出应对于出站对话归一值利用-30)。

[0096] 当E-AC-3编码器(其将LPSM值插入位流)“工作”并且正接收不具有“真”标记的AC-3帧时,嵌入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)激活。LPSM块生成继续进行,并且loudness_correction_type_flag被设定为“0”。响度控制器激活序列被与“置信”标记消失的解码AC-3帧的开始同步。响度控制器激活序列被如下实现:leveler_amount控制经过1个音频块周期(即,5.3毫秒)从值0增大至值9,并且leveler_back_end_meter控制被置于激活模式中(此操作将导致无缝转换,并且包括back_end_meter积分复位);并且在编码期间,图形用户界面(GUI)向用户指示如下参数:“输入音频节目:[可信/不可信]”-此参数的状态基于输入信号中的“置信”标记的存在;以及“实时响度校正:[启用/禁用]”-此参数的状态基于嵌入编码器的此响度控制器是否激活。

[0097] 当解码在位流的各帧的位流信息(“BSI”)分段的“addbsi”字段中包含LPSM(为所描述的格式)的AC-3或E-AC-3位流时,解码器解析LPSM块数据(在addbsi字段)中,并且将提取的LPSM值传递至图形用户界面(GUI)。所提取的LPSM值的集合被每帧刷新。

[0098] 在另一种格式中,编码位流是AC-3位流或E-AC-3位流,并且包括LPSM的元数据分段中的每一个被(例如,通过编码器100的级107)作为附加位流信息包含在位流的帧的位流信息(“BSI”)分段的“addbsi”字段(图6所示)中(或者在Aux分段中)。在此格式(其是上文参照表1和2描述的格式的变型)中,含有LPSM的addbsi(或Aux)字段中的每一个含有以下LPSM值:表1中规定的元素,随后是载荷ID(将元数据标识为LPSM)以及载荷大小值,随后是具有以下格式(类似于上文在表2中指示的元素)的载荷(LPSM数据):LPSM载荷的版本:指示LPSM载荷的版本的2位字段;dialchan:指示对应音频数据的左、右和/或中央通道是否含有口语对话的3位字段。Dialchan字段的位分配可以如下:指示对话存在于左通道中的位0被存储在dialchan字段的最高有效位中;并且指示对话存在于中央通道中的位2被存储在

dialchan字段的最低有效位中。如果在节目的前0.5秒期间对应的通道含有口语对话，则dialchan字段的每一位被设定为“1”；loudregtyp:指示节目响度所遵从的响度规则标准的3位字段。将“loudregtyp”设定为“000”指示LPSM不指示响度规则遵从。例如，此字段的一个值(例如,000)可指示没有指示遵从响度规则标准，此字段的另一个值(例如,001)可指示节目的音频数据遵从ATSC A/85标准，此字段的另一个值(例如,010)可指示节目的音频数据遵从EBU R128标准。在示例中，如果字段被设定为“000”之外的任何值，在载荷中随后应为loudcorrdialgat和loudcorrtyp字段；loudcorrdialgat:指示对话门控响度校正是否已经被应用的1位字段。如果节目的响度已经使用对话门控被校正，则loudcorrdialgat字段的值被设定为“1”。否则，其被设定为“0”；loudcorrtyp:指示应用于节目的响度校正的类型的1位字段。如果节目的响度已经通过无限前瞻(基于文件)响度校正处理被校正，则loudcorrtyp字段的值被设定为“0”。如果节目的响度已经使用实时响度测量和动态范围控制的组合被校正，则此字段的值被设定为“1”；loudrelgate:指示相对门控响度数据(IITU)是否存在1位字段。如果loudrelgate字段被设定为“1”，在载荷中随后应为7位的ituloudrelgat字段；loudrelgat:指示相对门控节目响度(IITU)的7位字段。该字段指示根据ITU-R BS.1770-2测量的音频节目的积分响度，其由于被应用对话归一和动态范围压缩而没有任何增益调整。值0到127被解释为-58LKFS到+5.5LKFS，步长为0.5LKFS；loudspchgate:指示语音门控响度数据(IITU)是否存在1位字段。如果loudspchgate字段被设定为“1”，则在载荷中随后应为7位的loudspchgat字段；loudspchgat:指示语音门控节目响度的7位字段。此字段指示根据ITU-R BS.1770-3的式(2)测量的整个对应音频节目的积分响度，其由于被应用对话归一和动态范围压缩而没有任何增益调整。值0到127被解释为-58LKFS到+5.5LKFS，步长为0.5LKFS；loudstrm3se:指示短期(3秒)响度数据是否存在1位字段。如果该字段被设定为“1”，则在载荷中随后应为7位的loudstrm3s字段；loudstrm3s:指示根据ITU-R BS.1771-1测量的对应音频节目的前3秒的未门控响度的7位字段，其由于被应用对话归一和动态范围压缩而没有任何增益调整。值0到256被解释为-116LKFS到+11.5LKFS，步长为0.5LKFS；truepk:指示真实峰值响度数据是否存在1位字段。如果truepk字段被设定为“1”，则在载荷中随后应为8位的truepk字段；以及truepk:指示节目的真实峰值采样值的8位字段，其根据ITU-R BS.1770-3的附件2被测量并且由于被应用对话归一和动态范围压缩而没有任何增益调整。值0到256被解释为-116LKFS到+11.5LKFS，步长为0.5LKFS。

[0099] AC-3位流或E-AC-3位流的帧的auxdata字段(或“addbsi”字段)中的元数据分段的核心元素包括核心头部(可选地，包括标识值，例如核心元素版本)，并且在核心头部之后包括：指示对于元数据分段的元数据是否包含指纹数据(或者其它保护值)的值，指示是否存在外部数据(与对应于元数据分段的元数据的音频数据有关)，核心元素标识的每一种元数据(例如，LPSM和/或除LPSM之外的元数据)的载荷ID和载荷大小，和用于核心元素标识的至少一种元数据的保护值。元数据分段的元数据载荷在核心头部之后，并且(在一些情况下)被嵌套在核心元素的值内。

[0100] 优化响度和动态范围系统

[0101] 上文描述的安全元数据编码和传送方案与如图1所示的用于在不同回放设备、应用、收听环境之间优化响度和动态范围的可缩放和可扩展系统结合使用。在一个实施例中，

系统10被配置为在需要不同的响度值并且具有不同的动态范围能力的各种设备之间归一化输入音频11的响度水平和动态范围。为了归一化响度水平和动态范围，系统10包括关于音频内容的不同的设备简档，并且归一化基于这些简档被进行。简档可通过音频处理链中的音频处理单元之一被包含，并且所包含的简档可被音频处理链中的下游处理单元使用来确定目标设备的希望的目标响度和动态范围。额外的处理组件可提供或者处理用于设备简档管理的信息（包括但不限于以下参数：空带范围、真实峰值阈值、响度范围、快/慢时间常数（系数）和最大提升量）、增益控制和宽带和/或多带增益生成功能。

[0102] 图10示出了根据一些实施例的用于提供优化响度和动态范围控制的系统的图1的系统的更详细示图。对于图10的系统321，编码器级包括核心编码器组件304，其将音频输入303编码为合适的数字格式以传输至解码器312。音频被处理为使得其可在各种不同收听环境中被回放，每一收听环境可能需要不同的响度和/或动态范围目标设定。因此，如图10所示，解码器输出数字信号，该数字信号被数字模拟转换器316转换成模拟格式以便通过包括全范围扬声器320、小型扬声器322和耳机324的各种不同驱动器类型被回放。这些驱动器仅示出了可能的回放驱动器的一些示例，并且具有任何合适大小和类型的任何换能器或驱动器可被使用。另外，图10的驱动器/换能器320-324可表现为在任何对应的收听环境中使用的任何合适的回放设备。设备类型可包括例如AVR、电视、立体声设备、计算机、移动电话、平板电脑、MP3播放器等；并且收听环境可包括例如礼堂、家中、车内、收听室等。

[0103] 由于回放环境和驱动器类型的范围可从非常小的私人领域到非常大的公共场所进行改变，因此可能的且最优的回放响度和动态范围配置的跨度可根据内容类型、背景噪声类型等显著改变。例如，在家庭影院环境中，宽动态范围内容可通过环绕声设备被播放，而动态范围较窄的内容可通过常规电视系统（诸如平板LED/LCD类型）被播放，而当不希望大的水平变化时对于某些收听状况（例如，在晚上或者在具有严苛声学输出功率限制的设备（诸如移动电话/平板内部扬声器或耳机输出）上）可使用极窄动态范围模式。在诸如使用小计算机或底座扬声器或者耳机/耳塞的便携式或移动收听场景中，最优的回放动态范围可根据环境改变。例如，与噪声环境相比，在安静环境中最优动态范围可更大。图10的自适应音频处理系统的实施例将根据参数（诸如，收听设备环境和回放设备类型）改变动态范围以更明了地呈现音频内容。

[0104] 图11是示出在示例性使用情况中的关于各种回放设备和背景收听环境的不同动态范围要求的表。可对于响度得出类似的要求。不同的动态范围和响度要求生成由优化系统321使用的不同的简档。系统321包括响度和动态范围测量组件302，其分析和测量输入音频的响度和动态范围。在一个实施例中，系统分析整体节目内容以确定整体响度参数。在此上下文中，响度指的是节目的长期节目响度或者平均响度，其中节目是诸如电影、电视节目、商业广告或类似节目内容的音频内容的单个单位。响度被用于提供被内容创建器用于控制音频将如何回放的艺术动态范围简档的指示。响度与对话归一元数据值有关，对话归一代表单个节目（例如，电影、电视、商业广告等）的平均对话响度。与节目响度相比，短期动态范围量化了在短得多时间段上的信号改变。例如，短期动态范围可在秒的量级上被测量，而节目响度可分钟甚至小时的跨度上被测量。短期动态范围提供了独立于节目响度的保护机制，以确保对于各种回放简档和设备类型不会发生过载。在一个实施例中，响度（长期节目响度）目标是基于对话响度的，而短期动态范围是基于相对门控和/或未门控响度的。在

此情况中,系统中的某些DRC和响度组件是关于内容类型和/或目标设备类型和特性上下文感知的。作为此上下文感知能力的一部分,系统被配置为分析输出设备的一个或多个特性以确定设备是否是对于某些DRC和响度回放状况被优化的特定一组设备(诸如AVR类型设备、电视、计算机、便携式设备等)中的一员。

[0105] 预处理组件分析节目内容以确定响度、峰值、真实峰值和安静时段,以为多个不同简档中的每一简档创建唯一的元数据。在一个实施例中,响度可以是对话门控响度和/或相对门控响度。不同简档定义了各种DRC(动态范围控制)和目标响度模式,在这些模式中,在编码器中根据源音频内容、所希望的目标响度和回放设备类型和/或环境的特性生成不同的增益值。解码器可提供不同的DRC和目标响度模式(通过上文提及的简档被使能),并且可包括DRC和目标响度关闭/禁用,其允许全动态范围展示,没有音频信号压缩且没有相对归一化;DRC关闭/禁用和目标为-31LKFS的响度归一化的线模式,用于在家庭影院系统上回放,通过编码器中(尤其对于此回放模式和/或设备简档)生成的增益值提供中等动态范围压缩并且进行目标为-31LKFS的响度归一化;RF模式,用于通过TV扬声器回放,提供重度动态范围压缩并且进行目标为-24、-23或-20LKFS的响度归一化,中间模式,用于通过计算机或类似设备回放,提供压缩且进行目标为-14LKFS的响度归一化,以及便携模式,提供了非常重的动态范围压缩并且进行目标为-11LKFS的响度归一化。目标响度值-31、-23/-20、-14和-11LKFS预计是可对于根据一些实施例的系统被定义的不同回放/设备简档的示例,并且任何其它合适的目标响度值可被采用,并且该系统可尤其对于这些回放模式和/或设备简档生成合适的增益值。此外,系统可扩展和修改,使得不同的回放设备和收听环境可通过在编码器或别处定义新的简档而被适应并且被加载到编码器中。这样,新的和唯一的回放/设备简档可被生成以支持用于未来应用的改进的或不同的回放设备。

[0106] 在一个实施例中,增益值可在系统321的任何合适的处理组件(诸如在编码器304、解码器312或代码转换器308、或者与编码器相关联的任何关联预处理组件或与解码器相关联的任何后处理组件)处被计算。

[0107] 图13是示出根据一个实施例的各种不同的回放设备类别的不同简档之间的接口的框图。如图13所示,编码器502接收音频输入501以及若干不同的可能的简档506之一。编码器将音频数据与所选择的简档相组合以生成输出位流文件,该输出位流文件在目标回放设备中的或者与目标回放设备相关联的解码器组件中被处理。对于图13的示例,不同的回放设备可能是计算机510、移动电话512、AVR514和电视516,但是许多其他的输出设备也是可能的。设备510到516中的每一个包括或者耦合到扬声器(包括驱动器和/或换能器),诸如驱动器320到324。回放设备和相关联的扬声器的大小、额定功率和处理的组合通常指示哪一简档对于特定目标是最优的。因此,简档506可对于通过AVR、TV、移动扬声器、移动耳机等回放被具体限定。它们也可对于特定操作模式或状况(诸如安静模式、夜晚模式、室外、室内等)被限定。图13中所示的简档仅是示例性模式,并且任何合适的简档可被限定,包括用于特定目标和环境的定制简档。

[0108] 尽管图13示出了如下实施例,其中编码器502接收简档506并且生成合适的关于响度和DRC处理的参数,应指出,但是基于简档和音频内容生成的参数可在任何合适的音频处理单元(诸如编码器、解码器、代码转换器、预处理器、后处理器等)上执行。例如,图13的每个输出设备510到516具有或者耦合到解码器组件,该解码器组件处理从编码器502发送的

文件504的位流中的元数据,以使得响度和动态范围能够适于与目标输出设备的设备或设备类型匹配。

[0109] 在一种实施例中,音频内容的动态范围和响度对于每个可能的回放设备被优化。这是通过对于每一目标回放模式将长期响度维持为目标并且控制短期动态范围以优化音频体验(通过控制信号动态、样本峰值和/或真实峰值)来实现的。对于长期响度和短期动态范围定义不同的元数据元素。如图10所示,组件302分析整个输入音频信号(或者其一部分,诸如语音组分,如果可应用的话)以得到这些单独DR组分的相关特性。这允许对于艺术增益对剪辑(过载保护)增益值定义不同的增益值。

[0110] 用于长期响度和短期动态范围的这些增益值然后被映射到简档305,以生成描述响度和动态范围控制增益值的参数。这些参数与来自编码器304的编码音频信号在用于创建位流的复用器306或者类似组件中组合,该位流通过代码转换器308被传输到解码器级。输入解码器级的位流在解复用器310中被解复用。其然后在解码器312中被解码。增益组件314应用对应于合适简档的增益以生成数字音频数据,该数字音频数据然后通过DACS单元416被处理以供通过合适的回放设备和驱动器或换能器320-324回放。

[0111] 图14是示出根据一个实施例的对于多个被定义的简档的长期响度和短期动态范围之间的关联性的表。如图14的表4所示,每个简档包括一组增益值,其指示在系统的解码器或每个目标设备中应用的动态范围压缩(DRC)的量。被指示为简档1~N的N个简档中的每一个通过指示在解码器级中应用的对应增益值来设定特定的长期响度参数(例如,对话归一)和过载压缩参数。用于简档的DRC增益值可被编码器接受的外部源定义,或者如果没有提供外部值,它们可在编码器内部生成作为默认增益值。

[0112] 在一个实施例中,用于每个简档的增益值被体现在基于音频信号的某些特性(诸如峰值、真实峰值、对话的短期响度或整体长期响度或它们的组合(混合))的分析被计算的DRC增益字中,以基于实现用于每个可能的设备简档和/或目标响度的最终DRC增益的/慢上升以及快/慢释放所需的时间常数以及所选择的简档(例如传递特性或曲线)来计算静态增益。如上所述,这些简档可在编码器、解码器中被于预先设定,或者在外部生成并且经由来自内容创建器的外部元数据被送至编码器。

[0113] 在一个实施例中,增益值可以是在音频内容的所有频率上应用相同增益的宽带增益。作为替代,增益可以包含多带增益值,从而对于音频内容的不同频率或频带应用不同的增益值。在多通道情况中,每个简档可构成指示用于不同频带的增益的增益值矩阵,而不是单个增益值。

[0114] 参照图10,在一个实施例中,关于收听环境的性质或特性的和/或回放设备的能力和配置的信息被反馈链路330解码器级提供给编码器级。简档信息332也被输入编码器304。在一个实施例中,解码器分析位流中的元数据以确定在位流中是否可获得用于第一组回放设备的响度参数。如果可获得,则将参数送至下游以用于呈现音频。否则,编码器分析设备的某些特性以得出该参数。这些参数然后被送至下游呈现组件以供回放。编码器还确定将呈现所接收的音频流的输出设备(或者包括输出设备的一组输出设备)。例如,输出设备可被确定为手机或者属于一组类似的便携设备。在一个实施例中,解码器使用反馈链路330来向编码器指示所确定的输出设备或该组输出设备。对于此反馈,与输出设备连接的模块(例如,膝上型计算机中的与耳机连接的或者与扬声器连接的声卡中的组件)可向解码器指示

输出设备的身份或者包括该输出设备的一组设备的身份。解码器将此信息通过反馈链路330传输至编码器。在一个实施例中，编码器执行解码器确定响度和DRC参数。在一个实施例中，解码器确定响度和DRC参数。在此实施例中，不是通过反馈链路330传输信息，而是解码器使用关于所确定的设备或该组输出设备的信息以确定响度和DRC参数。在另一个实施例中，另一音频处理单元确定响度和DRC参数，并且解码器将该信息传输至音频处理单元而不是解码器。

[0115] 图12是根据一个实施例的动态范围优化系统的框图。如图12所示，编码器402接收输入音频401。编码的音频在复用器409中与从所选择的压缩曲线422和对话归一值424生成的参数404组合。所得到的位流被传输至解复用器411，该解复用器411生成音频信号，该音频信号被解码器406解码。参数和对话归一值被增益计算单元408使用以生成增益水平，该增益水平驱动放大器410以便放大解码器输出。图12示出了动态范围控制如何被参数化并被插入位流。响度也可通过使用类似组件被参数化并被插入位流。在一个实施例中，输出参考水平控制(未示出)可被提供给解码器。尽管附图示出了响度和动态范围参数在编码器处被确定和插入，类似的确定可在其它音频处理单元(例如，预处理器、解码器和后处理器)被执行。

[0116] 图15示出了根据一个实施例的不同类型的音频内容的响度简档的示例。如图15所示，示例性曲线600和602相对于以0LKFS为中心的增益绘制输入响度(按LFKS)。不同类型的内容展示不同的曲线，如图15所示，其中曲线600可代表语音而曲线602可代表标准电容内容。如图15所示，与电影内容相比，语音内容所经受的增益量更大。图15是用于某些类型的音频内容的代表性简档曲线的示例，并且其它简档也可被使用。如图15所示的简档特性的某些方面被用于得出用于优化系统的相关参数。在一个实施例中，这些参数包括：空带宽、剪切比、提升比、最大提升量、FS上升、FS衰减、迟滞、峰值限制和目标水平响度。根据应用要求和系统约束，作为这些参数中的至少一些参数的附加或替代，还可使用其它参数。

[0117] 图16是示出根据一个实施例的用于在回放设备和应用之间优化响度和动态范围的方法的流程图。尽管附图示出了响度和动态范围优化在编码器处执行，但是类似优化可在其它音频处理单元(例如，预处理器、解码器和后处理器)处执行。如处理620中所示，方法以编码器级接收来自源的输入信号(603)开始。编码器或预处理组件然后确定原信号是否已经受了实现目标响度和/或动态范围的处理(604)。目标响度对应于长期响度并且可在外部或内部定义。如果源信号没有经受实现目标响度和/或动态范围的处理，则该系统执行适当的响度和/或动态范围控制操作(608)；否则，如果源信号经受了此响度和/或动态范围控制操作，则系统进入旁路模式以跳过此响度和/或动态范围控制操作，与允许原始处理指示适当的长期限度和/或动态范围(606)。旁路模式606或者执行的模式608中的任一个的适当增益值(可以是单个宽带增益值或者依赖于频率的多带增益值)然后在解码器中被应用(612)。

[0118] 位流格式

[0119] 如前文所提及的，用于优化响度和动态范围的系统采用了安全的可扩展的元数据格式，以确保在编码器和解码器之间或者在源和呈现/回放设备之间在位流中传输的元数据和音频内容在经网络或者其它专有设备、诸如服务提供者界面等传输期间没有相互分离或者被中断。位流提供了如下机制，用于用信号告知编码器和/或解码器组件以通过合适的

简档信息改变音频信号的响度和动态范围以适合音频内容和输出设备特性。在一个实施例中,系统被配置为确定要在编码器和解码器之间传输的低速率编码位流,并且通过元数据编码的响度信息包括一个或多个输出简档的特性。用于根据一个实施例的响度和动态范围优化系统的位流格式的描述如下。

[0120] AC-3编码位流包括元数据,以及音频内容的1到6个通道。音频内容是已经使用感知音频编码被压缩的音频数据。元数据包括若干音频元数据参数,它们预期用于改变传递至收听环境的节目的声音。AC-3编码音频位流的每一帧含有关于数字音频的1536个帧的音频内容和元数据。对于48kHz的采样率,这代表32毫秒的数字音频或者音频每秒31.25帧。

[0121] E-AC-3编码音频位流的每一帧,根据帧含有1个、2个、3个还是6个音频数据块,含有关于数字音频的256、512、768或者1536个帧的音频内容和元数据。对于48kHz的采样率,这分别代表5.333、10.667、16或32毫秒的数字音频或者分别代表音频每秒189.9、93.75、62.5或31.25帧。

[0122] 如图4所指示的,每个AC-3帧被分成区段(分段),包括同步信息(SI)区段,其含有(如图5所示)同步字(SW)和两个误差校正字中的第一个(CRC1);位流信息(BSI)区段,其含有大部分元数据;六个音频块(AB0到AB5),含有数据压缩音频内容(并且也可包含元数据);废弃位(WI),含有在音频内容被压缩之后被遗留的任何不使用的位;辅助(AUX)信息区段,其含有更多的元数据;以及两个误差校正字中的第二个(CRC2)。

[0123] 如图7所示,每个E-AC-3帧被分成区段(分段),包括同步信息(SI)区段,其含有(如图5所示)同步字(SW);位流信息(BSI)区段,其含有大部分元数据;1到6个之间的音频块(AB0到AB5),含有数据压缩音频内容(并且也可包含元数据);废弃位(WI),含有在音频内容被压缩之后被遗留的任何不使用的位;辅助(AUX)信息区段,其含有更多的元数据;以及误差校正字(CRC)。

[0124] 在AC-3(或者E-AC-3)位流中,存在被具体预期用于改变传递至收听环境的节目的声音的若干音频元数据参数。元数据参数之一是对话归一参数,其被包含在BSI分段中。

[0125] 如图6所示,AC-3帧的BSI分段包括指示节目的对话归一值的5位参数(“dialnorm”)。如果AC-3帧的音频编码模式(“acmod”)是“0”,指示双单声或者“1+1”通道配置,则指示在同一AC-3帧中运送的第二音频节目的对话归一值的5位参数(“dialnorm2”)被包括。

[0126] BSI分段还包括标记(“addbsie”),其指示在“addbsie”位之后存在(或者不存在)附加位流信息,参数(“addbsi1”),其指示在“addbsi1”值之后的任何附加位流信息的长度,以及在“addbsi1”值之后的高达64位的附加位流信息(“addbsi”)。BSI分段可包括在图6中未具体示出的其它元数据值。

[0127] 文中所述的一个或多个实施例的各方面可在如下音频系统中实现,该音频系统处理用于经包括执行软件指令的一个或多个计算机或处理设备的网络传输的音频信号。所述任何实施例可单独使用或者以任何组合相互结合地使用。尽管已经由于现有技术的各种缺陷而构想了各种实施例(它们已经在说明书的一个或多个地方被讨论或提及),但是实施例不必然解决这些缺陷中的任一个。换句话说,不同的实施例可解决可能在说明书中讨论的不同缺陷。一些实施例可仅部分地解决可在说明书中被讨论的一些缺陷或者仅仅一个缺陷,而一些实施例可能不解决这些缺陷中的任一个。

[0128] 文中描述的系统的各个方面可在用于处理数字或数字化音频文件的合适的基于计算机的声音处理网络环境中被实现。自适应音频系统的部分可包括一个或多个网络，该网络包含任何希望数量的个体机器，包括用于缓冲并且路由在计算机之间传输的数据的一个或多个路由器(未示出)。这样的网络可基于各种不同的计算机协议被构建，并且可以是因特网、广域网(WAN)、局域网(LAN)或者它们的任何组合。

[0129] 组件、块、处理或者其它功能组件中的一个或多个可通过控制系统的基于处理器的计算设备的执行的计算机程序来实现。还应指出，在它们的行为、寄存器传递、逻辑组分和/或其它特性方面，文中公开的各种功能可使用硬件、固件和/或体现在各种机器可读或计算机可读介质中数据和/或指令的任何数量的组合来实现。这样的格式化数据和/或指令体现于其中的计算机可读介质包括但不限于各种形式的物理(非暂态)、非易失性存储介质，诸如光学、磁性或半导体存储介质。

[0130] 除非文中清楚要求，否则在说明书和权利要求书的上下文中，属于“包含”、“包括”等应在与排他性或详尽性意义上相反的包含性意义上被解释；即在“包括但不限于”的意义上被解释。使用单数或复数的词语也分别包括多数或单数。另外，“文中”、“下文”、“以上”、“以下”和类似词语是在整体上提及本申请，而不是本申请的任何特定部分。当在提及两个或更多项目的列表时使用词语“或者”时，该词语涵盖词语的以下的所有解释：列表中的项目中的任一个、列表中的所有项目或者列表中的项目的任何组合。

[0131] 尽管一个或多个实现已经作为示例并且根据具体实施例被描述，但是应理解，一个或多个实现不局限于所公开的实施例。相反，其预期涵盖本领域技术人员将显而易见的各种修改和类似布置。因此，随附权利要求的范围应被给予最宽泛的解释以便涵盖所有这样的修改和类似布置。

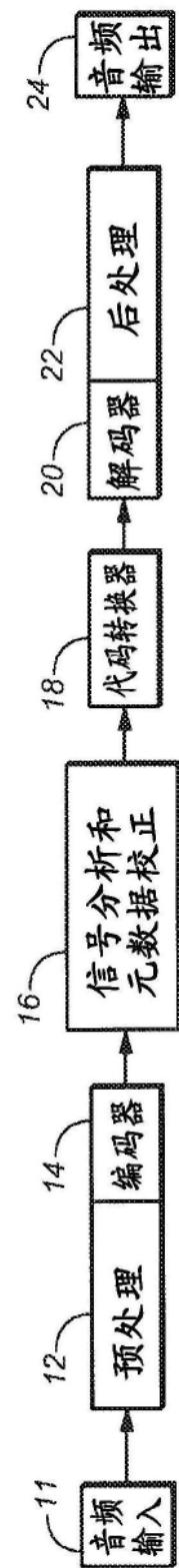


图1

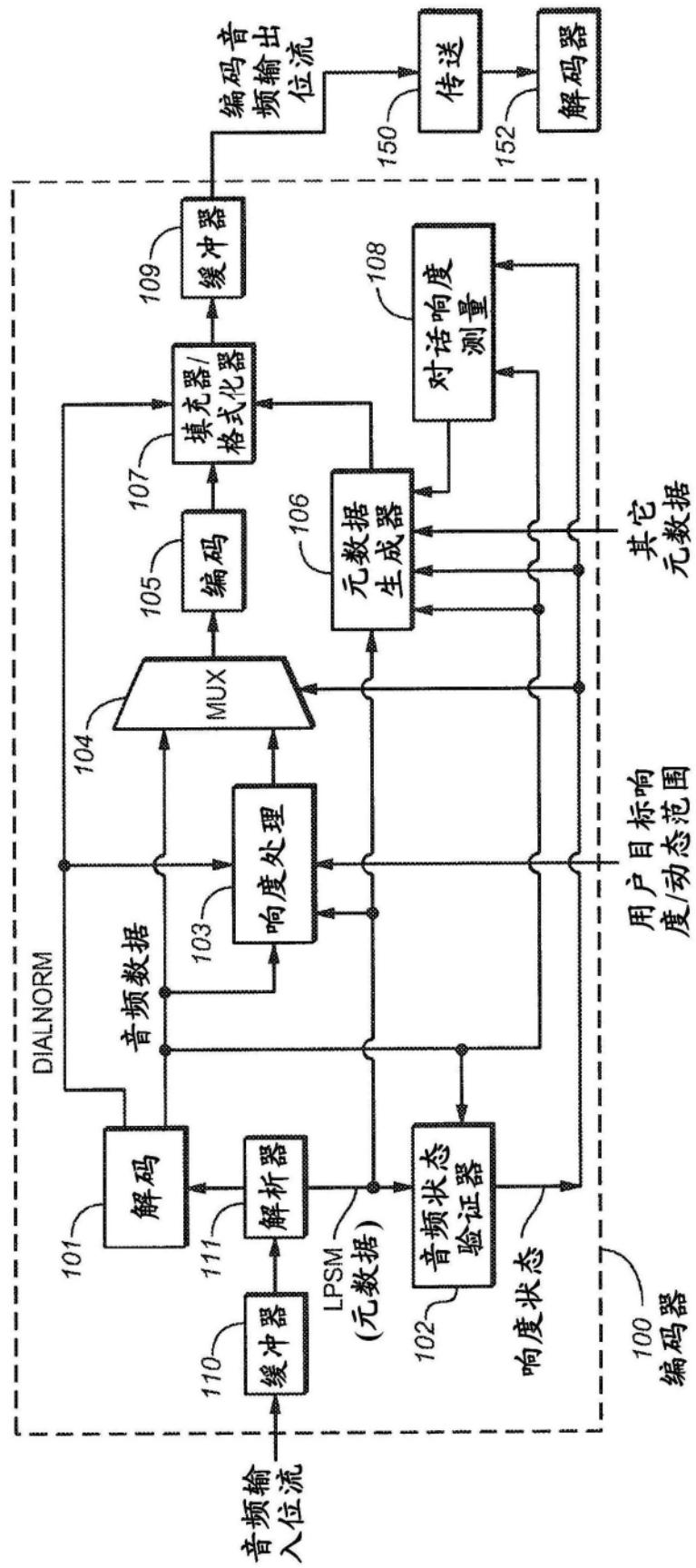


图2

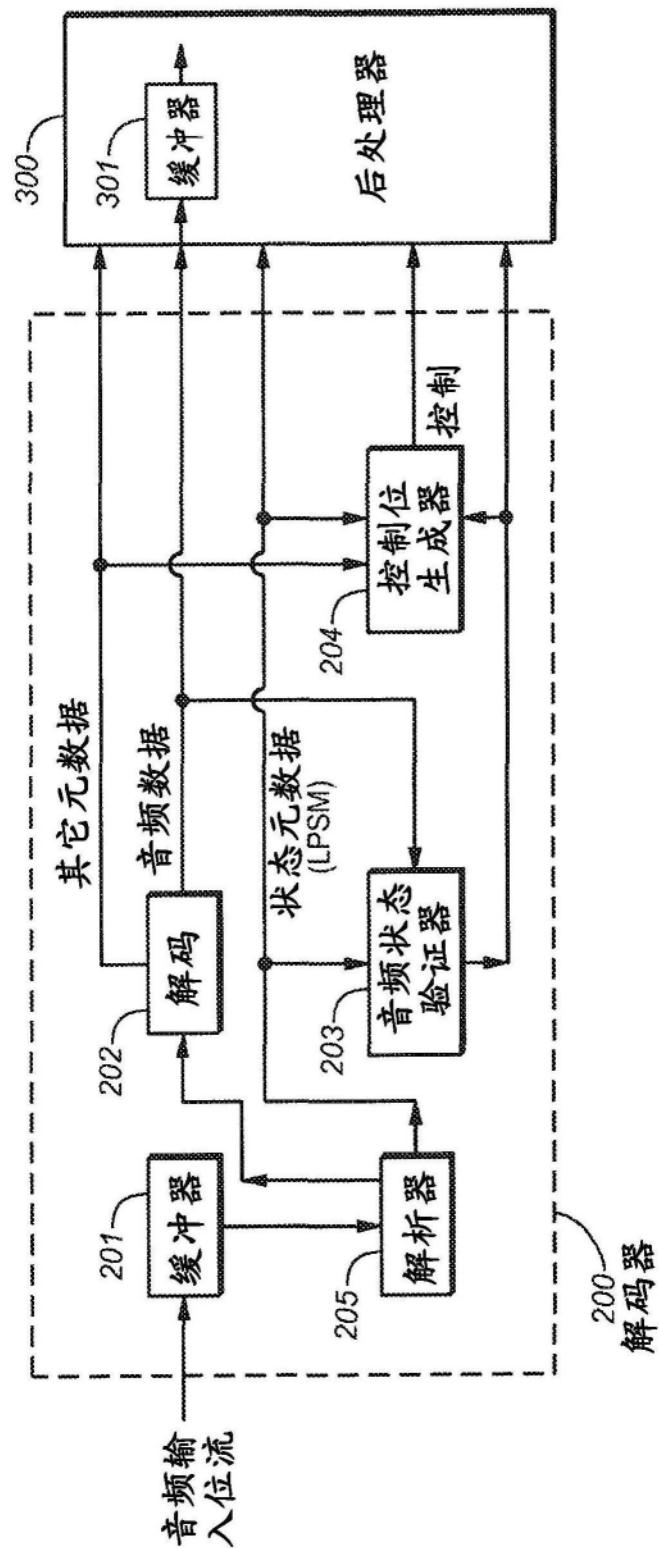


图3

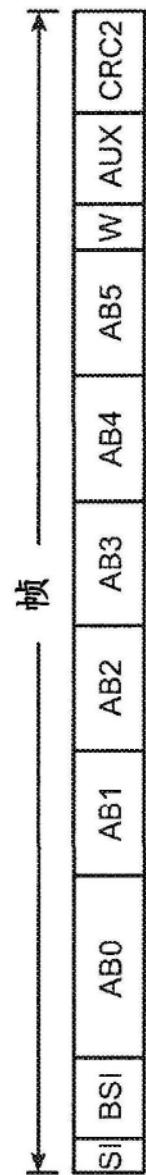


图4

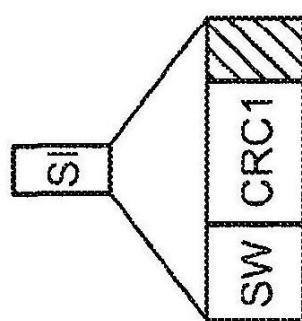


图5

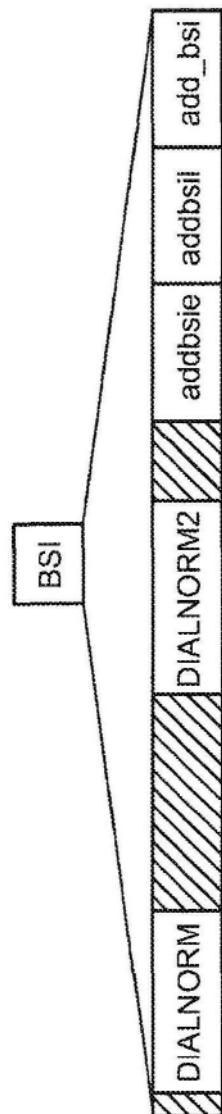


图6

S	BSI	AB0、AB1、AB2、AB3、AB4、AB5 中的一个到 6 个	W	AUX	CRC2
---	-----	-----------------------------------	---	-----	------

图7

参数	描述	必选/可选
SYNC [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。此值可在位流被从生产通过分发和最终发射被传递时增大/减小。	M
子流关联性	描述了核心元素与哪个(哪些)子流相关联	M
签名 (HMAC 摘要)	在整个帧的音频数据、核心元素以及所有扩展元素上计算的 256 位 HMAC 摘要 (使用 SHA-2 算法)	M
PGM 边界倒计数	仅对于一些帧在音频节目文件/流的头部或尾部处出现的字段。因此，核心元素版本改变可被用于告知包含此参数。	O
音频指纹	在由核心元素周期字段表示的一定数量的 PCM 音频采样上取得的音频指纹	O
视频指纹	在由核心元素周期字段表示的一定数量的压缩视频采样 (如果有的话) 上取得的音频指纹	O
URL/UUID	此字段被定义为承载引用附加节目内容 (本质) 和/或与位流相关联的元数据的外部位置的 URL 和/或 UUID (其可是指纹的冗余)。	O

表 1

图8

LPSM 参数 [智能响度]	描述
LPSM 版本	
LPSM 周期 (xxx)	仅可应用于 xxx 字段
LPSM 计数	
LPSM 子流 关联性	
对话通道	指示 L、C 和 R 音频通道的哪个组合在前 0.5 秒含有语音。语音不存在于任何 L、C 和 R 组合中时，此参数应指示“无对话”
响度规 则类型	指示关联音频数据流遵从具体规则集合（例如，ATSC A/85 或 EBU R128）
对话门控响 度校正标记	指示关联音频流是否已 基于对话门控被校正
响度校正类型	指示关联音频流是否已经通过无限前瞻 (基于文件) 或通过实时 (RT) 响度和 动态范围控制被校正
ITU 相对门控 响度 (INF)	指示被应用 w/o 元数据的关联音频流 ITU-R BS. 1770-2 积分响度 (例如, 7 位: -58 -> +5.5 LKFS, 步长为 0.5 LKFS)
ITU 语音门控 响度 (INF)	指示被应用 w/o 元数据的关联音频流 ITU-R BS. 1770-2 积分响度 (例如, 7 位: -58 -> +5.5 LKFS, 步长为 0.5 LKFS)
ITU (EBU 3341) 短期 3s 响度	指示被@约 10Hz 插入率应用 (滑动窗口) w/o 元数据 的关联音频流 3 秒未门控 ITU 响度 (例如, 8 位: 116 -> +11.5LKFS, 步长为 0.5 LKFS)
真实峰值	指示被应用 w/o 元数据的关联音频流的 ITU-R BS.1770 Annex 2 TruePeak 值 (dB TP) (即, 在元素周期字段中 信号告知的帧周期上的最大值), 116 -> +11.5 LKFS, 步长为 0.5 LKFS

表 2

图9A

编号唯一状态	必选/可选	插入率 (更新参数周期)
	M	
	M	
	M	
	M	
8	M	~0.5 秒 (典型)
8	M	帧
2	O (仅 Loudness_Regulation_type 指示对应的音频是 UNCORRECTED 时才存在)	帧
2	O (仅 Loudness_Regulation_type 指示对应的音频是 UNCORRECTED 时才存在)	帧
128	O	1 秒
128	O	1 秒
256	O	0.1 秒
256	O	0.5 秒

表 2

图9B

FIG. 9A FIG. 9B

图9

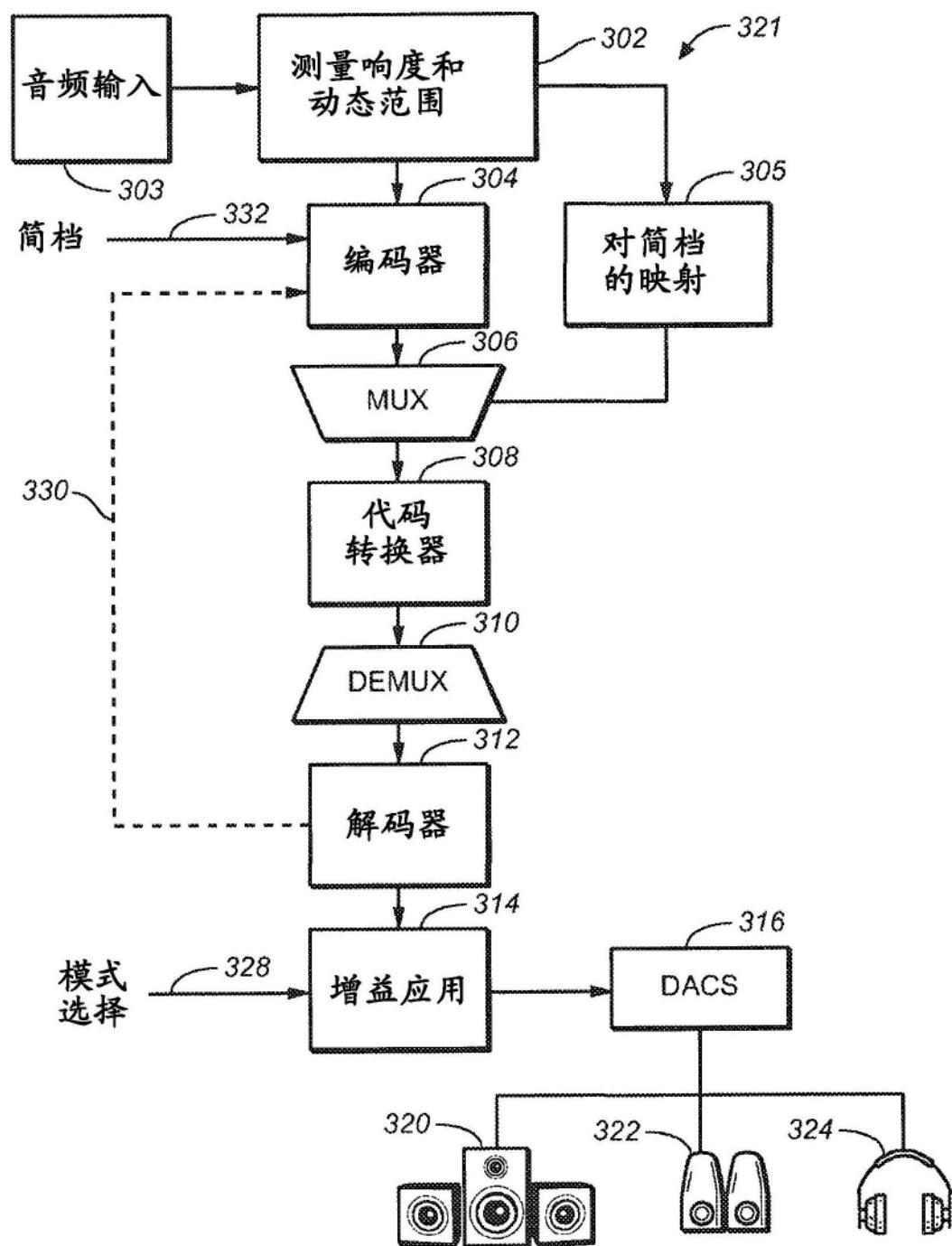


图10

	回放状况	
扬声器	安静	噪声
小/迷你	低动态范围	低动态范围
耳机	中等动态范围	低动态范围
大	高动态范围	低动态范围

表 3

图11

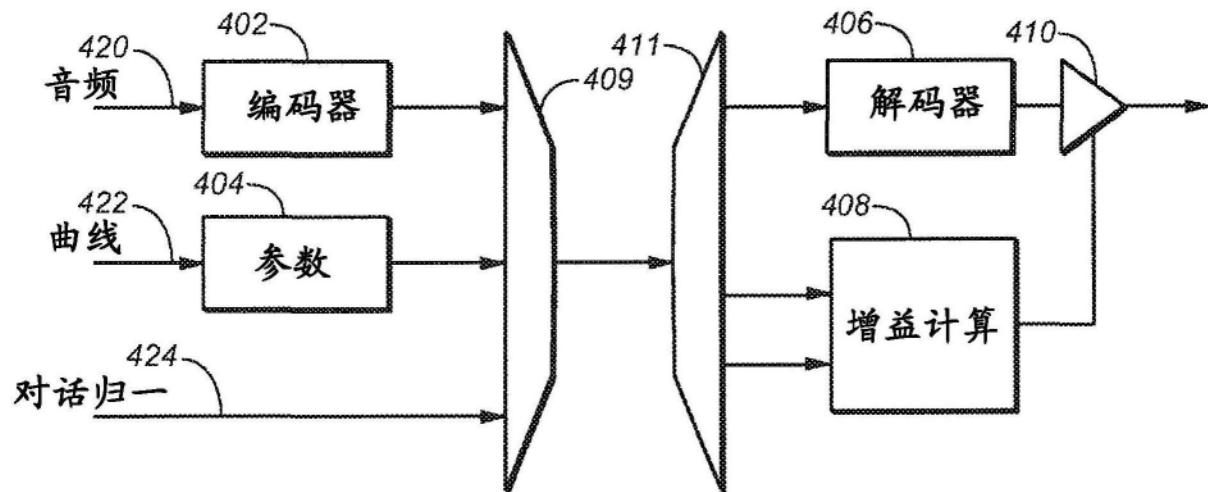


图12

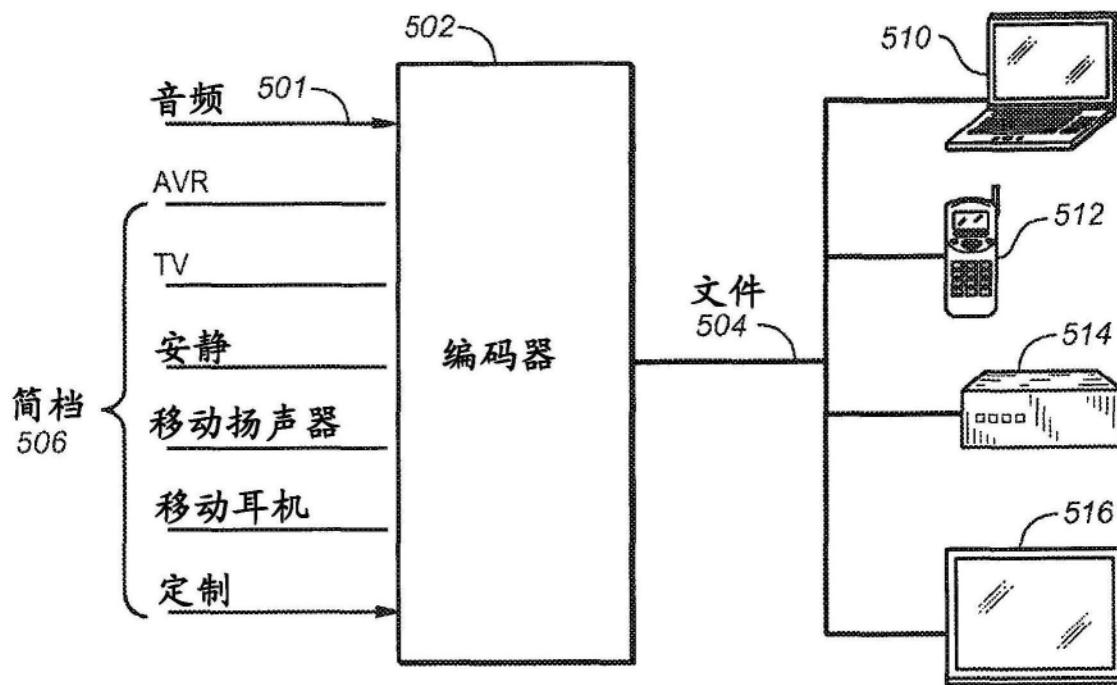


图13

长期响度	DRC 增益	过载压缩
1	简档 1	1
2	简档 2	2
3	简档 3	3
⋮	⋮	⋮
N	简档 N	N

表 4

图14

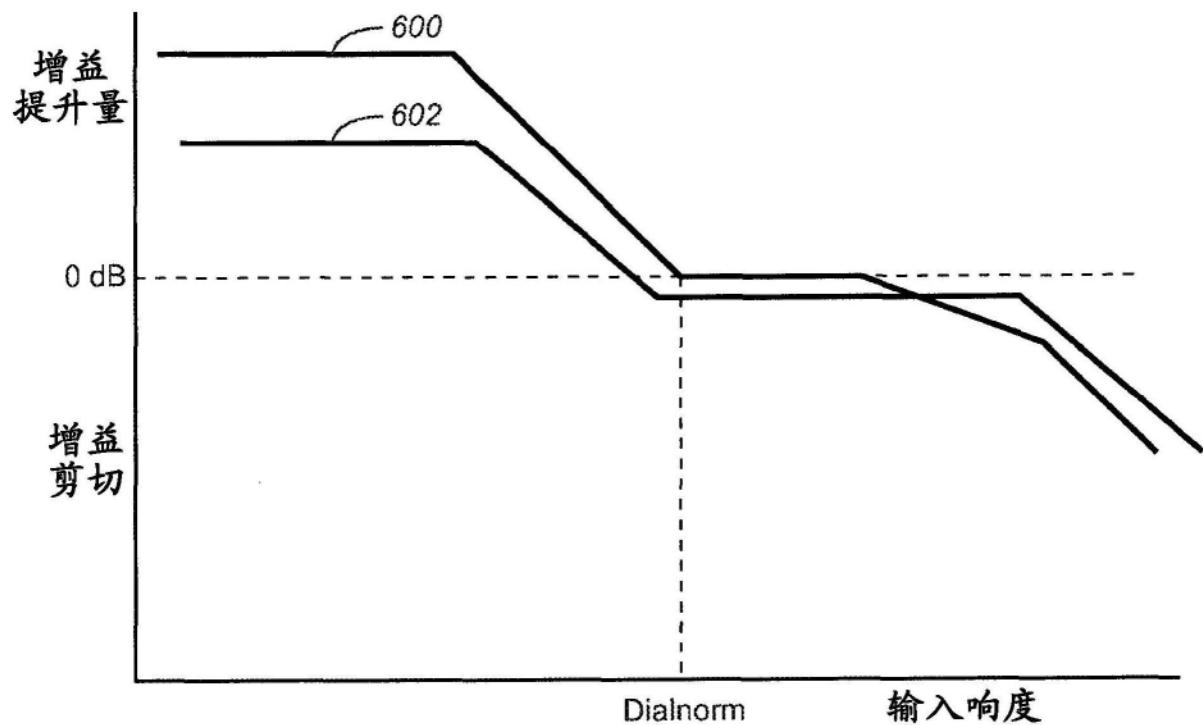


图15

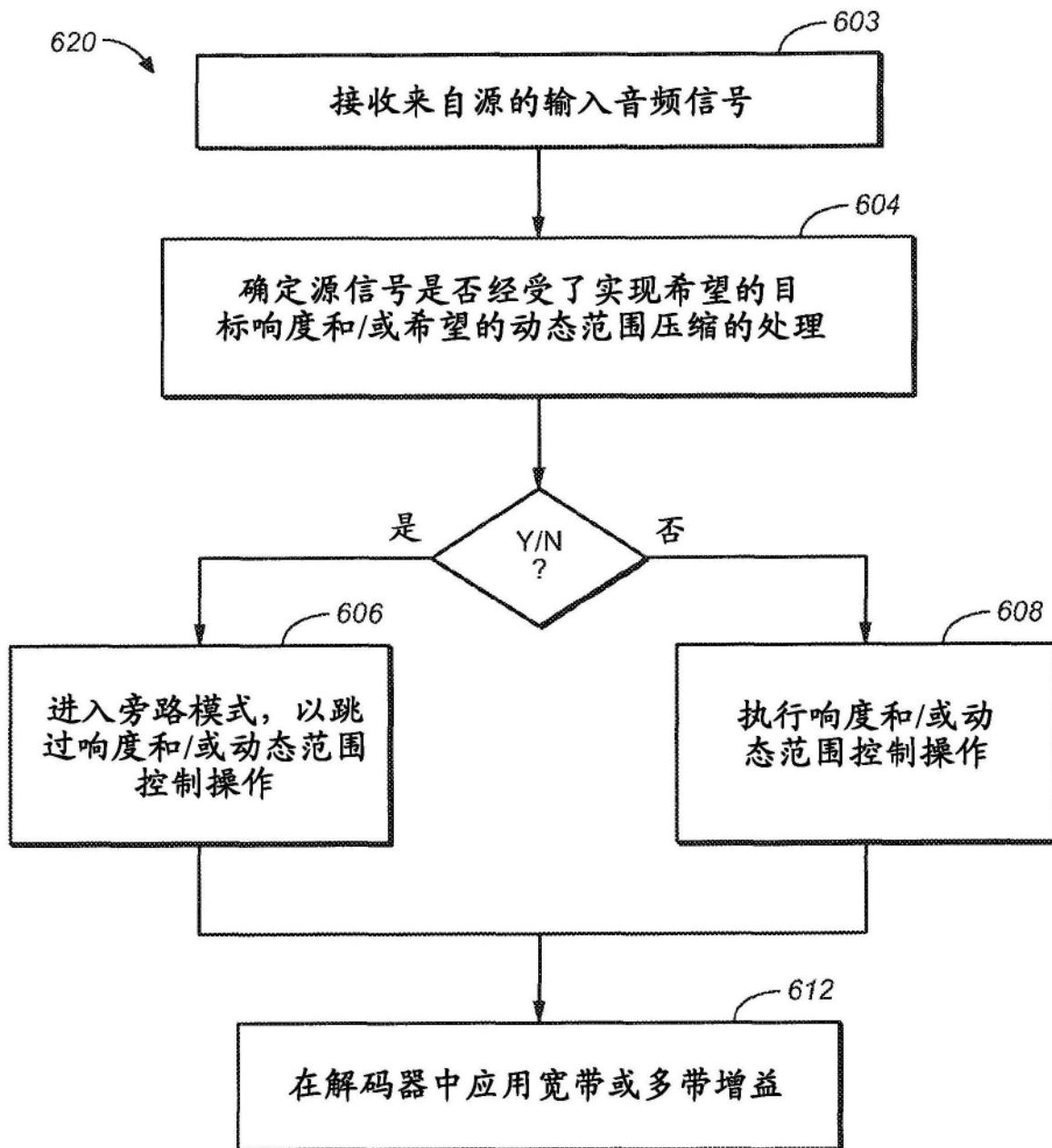


图16