(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105874532 A
(43)申请公布日 2016.08.17

(21)申请号 201480064443.5

(22)申请日 2014.11.27

(30)优先权数据
13194664.2 2013.11.27 EP

(85)PCT国际申请进入国家阶段日
2016.05.25

(86)PCT国际申请的申请数据
PCT/EP2014/075787 2014.11.27

(87)PCT国际申请的公布数据
WO2015/078956 EN 2015.06.04

(71)申请人 弗劳恩霍夫应用研究促进协会
地址 德国慕尼黑

(72)发明人 约尼·保卢斯 萨沙·迪施
哈拉尔德·富克斯
伯恩哈德·格里尔
奥利弗·赫尔穆特
阿德里安·穆尔塔扎
法尔科·里德布施
莱昂·特伦提夫

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
代理人 杜诚 杨铁成

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

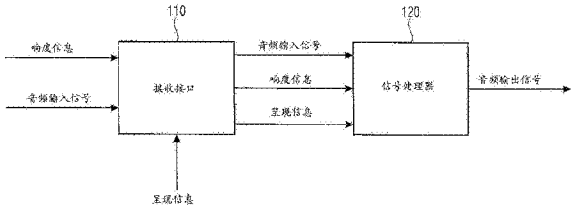
权利要求书5页 说明书22页 附图11页

(54)发明名称

基于对象的音频编码系统中的通知响度估计的解码器、编码器及方法

(57)摘要

提供一种用于产生包括一个或更多个音频输出声道的音频输出信号的解码器。解码器包括接收接口(110),接收接口用于接收包括多个音频对象信号的音频输入信号、关于音频对象信号的响度信息、以及指示音频对象信号中的一个或更多个是否应放大或衰减的呈现信息。此外,解码器包括用于产生音频输出信号的一个或更多个音频输出声道的信号处理器(120)。信号处理器(120)被配置成根据响度信息且根据呈现信息确定响度补偿值。此外,信号处理器(120)被配置成根据呈现信息且根据响度补偿值从音频输入信号产生音频输出信号的一个或更多个音频输出声道。此外,提供一种编码器。



CN 105874532 A

1.一种用于产生包括一个或更多个音频输出声道的音频输出信号的解码器，其中，所述解码器包括：

接收接口(110)，用于接收包括多个音频对象信号的音频输入信号、用于接收关于所述音频对象信号的响度信息、以及用于接收指示所述音频对象信号中的一个或更多个是应放大还是衰减的呈现信息；以及

信号处理器(120)，用于产生所述音频输出信号的所述一个或更多个音频输出声道，

其中，所述信号处理器(120)被配置成根据所述响度信息且根据所述呈现信息来确定响度补偿值，并且

其中，所述信号处理器(120)被配置成根据所述呈现信息且根据所述响度补偿值，从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

2.根据权利要求1所述的解码器，其中，所述信号处理器(120)被配置成根据所述呈现信息且根据所述响度补偿值从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道，使得所述音频输出信号的响度等于所述音频输入信号的响度，或使得所述音频输出信号的响度比经修改音频信号的响度更接近于所述音频输入信号的响度，所述经修改音频信号是通过根据所述呈现信息放大或衰减所述音频输入信号的所述音频对象信号而修改所述音频输入信号来产生的。

3.根据权利要求2所述的解码器，

其中，所述信号处理器(120)被配置成通过修改所述音频输入信号来产生所述经修改音频信号，其中通过根据所述呈现信息放大或衰减所述音频输入信号的所述音频对象信号而修改所述音频输入信号，并且

其中，所述信号处理器(120)被配置成通过将所述响度补偿值应用于所述经修改音频信号而产生所述音频输出信号，使得所述音频输出信号的响度等于所述音频输入信号的响度，或使得所述音频输出信号的响度比所述经修改音频信号的响度更接近于所述音频输入信号的响度。

4.根据前述权利要求之一所述的解码器，

其中，所述音频输入信号的所述音频对象信号中的每个被指派给两个或更多个组中确切的一个组，其中，所述两个或更多个组中的每个组包括所述音频输入信号的所述音频对象信号中的一个或更多个，

其中，所述接收接口(110)被配置成接收所述两个或更多个组中的每个组的响度值作为所述响度信息，

其中，所述信号处理器(120)被配置成根据所述两个或更多个组中的每个组的响度值来确定所述响度补偿值，并且

其中，所述信号处理器(120)被配置成根据所述响度补偿值从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

5.根据权利要求4所述的解码器，其中，所述两个或更多个组中的至少一个组包括所述音频对象信号中的两个或更多个。

6.根据权利要求1至权利要求3之一所述的解码器，

其中，所述音频输入信号的所述音频对象信号中的每个被指派给多于两个组中的确切的一个组，其中所述多于两个组中的每个组包括所述音频输入信号的所述音频对象信号中

的一个或更多个，

其中，所述接收接口(110)被配置成接收所述多于两个组中的每个组的响度值作为所述响度信息，

其中，所述信号处理器(120)被配置成根据所述多于两个组中的每个组的响度值来确定所述响度补偿值，并且

其中，所述信号处理器(120)被配置成根据所述响度补偿值从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

7.根据权利要求6所述的解码器，其中，所述多于两个组中的至少一个组包括所述音频对象信号中的两个或更多个。

8.根据权利要求4至权利要求7之一所述的解码器，

其中，所述信号处理器(120)被配置成根据下式确定所述响度补偿值

$$\Delta L = 10\log_{10}\frac{\sum_{i=1}^{N} g_i^2 10^{(L_i-c)/10}}{\sum_{i=1}^{N} h_i^2 10^{(L_i-c)/10}}$$

或者根据下式确定所述响度补偿值

$$\Delta L = 10\log_{10}\frac{\sum_{i=1}^{N} g_i^2 10^{L_i/10}}{\sum_{i=1}^{N} h_i^2 10^{L_i/10}}$$

其中，ΔL是所述响度补偿值，

其中，i指示所述音频对象信号中的第i个音频对象信号，

其中，L_i是所述第i个音频对象信号的响度，

其中，g_i是所述第i个音频对象信号的第一混合权重，

其中，h_i是所述第i个音频对象信号的第二混合权重，

其中，c是常数值，并且

其中，N是数目。

9.根据权利要求4至权利要求7之一所述的解码器，

其中，所述信号处理器(120)被配置成根据下式确定所述响度补偿值

$$\Delta L = 10\log_{10}\frac{\sum_{i=1}^{N} g_i^2 10^{K_i/10}}{\sum_{i=1}^{N} h_i^2 10^{K_i/10}}$$

其中，ΔL是所述响度补偿值，

其中，i指示所述音频对象信号中的第i个音频对象信号，

其中，g_i是所述第i个音频对象信号的第一混合权重，

其中，h_i是所述第i个音频对象信号的第二混合权重，

其中，N是数目，并且

其中，K_i是根据下式定义的

$K_i = L_i - L_{REF}$，

其中，L_i是所述第i个音频对象信号的响度，并且

其中，L_{REF}是基准对象的响度。

10.根据权利要求4或权利要求5所述的解码器，

其中，所述音频输入信号的所述音频对象信号中的每个被指派给作为所述两个或更多个组的确切的两个组中的确切的一个组，

其中，所述音频输入信号的所述音频对象信号中的每个被指派给所述确切的两个组中的前景对象组或所述确切的两个组中的背景对象组，

其中，所述接收接口(110)被配置成接收所述前景对象组的响度值，

其中，所述接收接口(110)被配置成接收所述背景对象组的响度值，

其中，所述信号处理器(120)被配置成根据所述前景对象组的响度值且根据所述背景对象组的响度值来确定所述响度补偿值，并且

其中，所述信号处理器(120)被配置成根据所述响度补偿值从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

11.根据权利要求10所述的解码器，

其中，所述信号处理器(120)被配置成根据下式确定所述响度补偿值

$$\Delta L = 10\log_{10}\frac{m_{FGO}^2 10^{K_{FGO}/10} + m_{BGO}^2 10^{K_{BGO}/10}}{10^{K_{FGO}/10} + 10^{K_{BGO}/10}}$$

其中，ΔL是所述响度补偿值，

其中，K_{FGO}指示所述前景对象组的响度值，

其中，K_{BGO}指示所述背景对象组的响度值，

其中，m_{FGO}指示所述前景对象组的呈现增益，且

其中，m_{BGO}指示所述背景对象组的呈现增益。

12.根据权利要求10所述的解码器，

其中，所述信号处理器(120)被配置成根据下式确定所述响度补偿值

$$\Delta L(A,B) = 10\log_{10}\frac{g_{FGO}^2 10^{L_{FGO}/10} + g_{BGO}^2 10^{L_{BGO}/10}}{10^{L_{FGO}/10} + 10^{L_{BGO}/10}}$$

其中，ΔL是所述响度补偿值，

其中，L_{FGO}指示所述前景对象组的响度值，

其中，L_{BGO}指示所述背景对象组的响度值，

其中，g_{FGO}指示所述前景对象组的呈现增益，且

其中，g_{BGO}指示所述背景对象组的呈现增益。

13.根据前述权利要求之一所述的解码器，

其中，所述接收接口(110)被配置成接收包括一个或更多个下混声道的下混信号作为

所述音频输入信号，其中所述一个或更多个下混声道包括所述音频对象信号，并且其中所述一个或更多个下混声道的数目小于所述音频对象信号的数目，

其中，所述接收接口(110)被配置成接收指示所述音频对象信号如何混合在所述一个或更多个下混声道内的下混信息，

其中，所述信号处理器(120)被配置成根据所述下混信息、根据所述呈现信息且根据所述响度补偿值，从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

14.根据权利要求13所述的解码器，

其中，所述接收接口(110)被配置成接收一个或更多个其他旁路音频对象信号，其中，所述一个或更多个其他旁路音频对象信号未混合在所述下混信号内，

其中，所述接收接口(110)被配置成接收所述响度信息，所述响度信息指示关于混合在所述下混信号内的所述音频对象信号的响度的信息且指示关于未混合在所述下混信号内的所述一个或更多个其他旁路音频对象信号的响度的信息，并且

其中，所述信号处理器(120)被配置成根据关于混合在所述下混信号内的所述音频对象信号的响度的信息且根据关于未混合在所述下混信号内的所述一个或更多个其他旁路音频对象信号的响度的信息来确定所述响度补偿值。

15.一种编码器，包括：

基于对象的编码单元(210;710)，用于对多个音频对象信号进行编码以获得包括所述多个音频对象信号的编码音频信号；以及

对象响度编码单元(220;720;820)，用于对关于所述音频对象信号的响度信息进行编码，

其中，所述响度信息包括一个或更多个响度值，其中所述一个或更多个响度值中的每个取决于所述音频对象信号中的一个或更多个。

16.根据权利要求15所述的编码器，

其中，所述编码音频信号的所述音频对象信号中的每个被指派给两个或更多个组中的确切的一个组，其中，所述两个或更多个组中的每个组包括所述编码音频信号的所述音频对象信号中的一个或更多个，

其中，所述对象响度编码单元(220;720;820)被配置成通过针对所述两个或更多个组中的每个组确定响度值来确定所述响度信息的所述一个或更多个响度值，其中所述组的所述响度值指示所述组的所述一个或更多个音频对象信号的总响度。

17.根据权利要求15所述的编码器，

其中，所述基于对象的编码单元(210;710)被配置成接收所述音频对象信号，其中所述音频对象信号中的每个被指派给确切的两个组中的确切的一个组，其中所述确切的两个组中的每个组包括所述音频对象信号中的一个或更多个，

其中，所述基于对象的编码单元(210;710)被配置成对所述确切的两个组包括的所述音频对象信号进行下混以获得包括一个或更多个下混音频声道的下混信号作为所述编码音频信号，其中所述一个或更多个下混声道的数目小于所述确切的两个组所包括的所述音频对象信号的数目，

其中，所述对象响度编码单元(220;720;820)被指派成接收一个或更多个其他旁路音

频对象信号，其中所述一个或更多个其他旁路音频对象信号中的每个被指派给第三组，其中所述一个或更多个其他旁路音频对象信号中的每个不被所述第一组包括且不被所述第二组包括，其中所述基于对象的编码单元(210;710)被配置成不将所述一个或更多个其他旁路音频对象信号下混在所述下混信号内，并且

其中，所述对象响度编码单元(220;720;820)被配置成确定所述响度信息的第一响度值、第二响度值和第三响度值，所述第一响度值指示所述第一组的所述一个或更多个音频对象信号的总响度，所述第二响度值指示所述第二组的所述一个或更多个音频对象信号的总响度，且所述第三响度值指示所述第三组的所述一个或更多个其他旁路音频对象信号的总响度，或者，所述对象响度编码单元(220;720;820)被配置成确定所述响度信息的第一响度值和第二响度值，该第一响度值指示所述第一组的所述一个或更多个音频对象信号的总响度，且该第二响度值指示所述第二组的所述一个或更多个音频对象信号和所述第三组的所述一个或更多个其他旁路音频对象信号的总响度。

18.一种系统，包括：

根据权利要求15至权利要求17之一所述的编码器(310)，用于对多个音频对象信号进行编码以获得包括所述多个音频对象信号的编码音频信号，以及

根据权利要求1至权利要求14之一所述的解码器(320)，用于产生包括一个或更多个音频输出声道的音频输出信号，

其中，所述解码器(320)被配置成接收所述编码音频信号作为音频输入信号，以及接收所述响度信息，

其中，所述解码器(320)被配置成还接收呈现信息，

其中，所述解码器(320)被配置成根据所述响度信息且根据所述呈现信息来确定响度补偿值，并且

其中，所述解码器(320)被配置成根据所述呈现信息且根据所述响度补偿值从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

19.一种产生包括一个或更多个音频输出声道的音频输出信号的方法，其中，所述方法包括：

接收包括多个音频对象信号的音频输入信号；

接收关于所述音频对象信号的响度信息；

接收指示所述音频对象信号中的一个或更多个是应放大还是衰减的呈现信息；

根据所述响度信息且根据所述呈现信息来确定响度补偿值；以及

根据所述呈现信息且根据所述响度补偿值从所述音频输入信号产生所述音频输出信号的所述一个或更多个音频输出声道。

20.一种用于编码的方法，包括：

对包括多个音频对象信号的音频输入信号进行编码；以及

对关于所述音频对象信号的响度信息进行编码，其中所述响度信息包括一个或更多个响度值，其中所述一个或更多个响度值中的每个取决于所述音频对象信号中的一个或更多个。

21.一种计算机程序，当在计算机或信号处理器上执行时用于实施根据权利要求19或权利要求20所述的方法。

基于对象的音频编码系统中的通知响度估计的解码器、编码器及方法

[0001] 本发明涉及音频信号编码、处理及解码，特别地，涉及用于基于对象的音频编码系统中的通知响度估计的解码器、编码器及方法。

[0002] 近来，在音频编码[BCC、JSC、SAOC、SAOC1、SAOC2]及通知源分离[ISS1、ISS2、ISS3、ISS4、ISS5、ISS6]的领域中已提出用于包括多个音频对象信号的音频场景的具比特率效益的传输/储存的参数技术。这些技术旨在基于描述所传输/储存的音频场景和/或该音频场景中的源对象的附加边信息(side information)来重构希望的输出音频场景或音频源对象。此重构使用通知源分离方案而发生于解码器中。可以对重构的对象进行组合以产生输出音频场景。取决于组合对象的方式，输出场景的感知响度可以不同。

[0003] 在TV及无线电广播中，可基于各个方面诸如峰值信号水平或响度水平来规范化各个节目的音轨的音量水平。取决于信号的动态特性，具有相同峰值水平的两个信号可能具有大不相同的感知响度水平。现在，在节目或频道之间切换，信号响度的差异非常恼人，且已成为终端用户对广播的抱怨的主要来源。

[0004] 在现有技术中，已提出使用基于感知信号响度的测量将所有频道上的所有节目类似地规范化至共同基准水平。在欧洲的一个此种推荐为EBU推荐R128[EBU](后文称为R128)。

[0005] 该推荐称，“节目响度”，例如一个节目(或一个广告或某其他有意义的节目实体)的平均响度应等于指定水平(允许有小的偏差)。当愈来愈多的广播台符合此推荐及所要求的规范化时，应最小化节目及频道之间的平均响度的差异。

[0006] 可以以若干方式进行响度估计。存在用于估计音频信号的感知响度的若干数学模型。EBU推荐R128依赖于ITU-R BS.1770(后文称为BS.1770)(参见[ITU])中提出的模型用于响度估计。

[0007] 如前所述，例如根据EBU推荐R128，节目响度例如一个节目的平均响度应等于指定水平，允许有小的偏差。然而，这导致了进行音频呈现时的显著问题，迄今在现有技术中仍未解决。在解码器侧进行音频呈现对所接收音频输入信号的总体/总响度具有显著影响。然而，尽管进行场景呈现，所接收音频信号的总响度应保持相同。

[0008] 当前，对于此问题不存在特定的解码器侧解决方案。

[0009] EP 2 146 522 A1([EP])涉及使用基于对象的元数据产生音频输出信号的构思。其生成表示至少两个不同音频对象信号的叠加的至少一个音频输出信号，但未提供对此问题的解决方案。

[0010] WO 2008/035275 A2([BRE])描述了一种包括编码器的音频系统，该编码器在编码单元中编码音频对象，该编码单元产生下混音频信号及表示多个音频对象的参数数据。该下混音频信号及参数数据被传输至包括解码单元及呈现单元的解码器，该解码单元产生音频对象的近似复本且该呈现单元根据该音频对象产生输出信号。此外，该解码器含有用于产生被发送至编码器的编码修改数据的处理器。然后，该编码器响应于该编码修改数据而修改该音频对象的编码，且特别是修改该参数数据。该方法允许由解码器控制对音频对象

的操纵但完全或部分地由编码器进行对音频对象的操纵。因此，可对实际的独立的音频对象而非对近似复本进行操纵，藉此提高效能。

[0011] EP 2 146 522 A1([SCH])公开了一种用于产生表示至少两个不同音频对象的叠加的至少一个音频输出信号的设备，其包括用于处理音频输入信号以提供音频输入信号的对象表示的处理器，其中此对象表示可通过使用对象下混信号对原始对象进行参数导引近似而产生。对象操纵器使用涉及各个音频对象的基于音频对象的元数据来单独地操纵对象，以获得经操纵的音频对象。使用对象混合器对经操纵的音频对象进行混合以取决于特定呈现设定而最终获得具有一个或若干个声道信号的音频输出信号。

[0012] WO 2008/046531 A1([ENG])描述了一种使用多个音频对象生成编码对象信号的音频对象编码器，其包含：下混信息产生器，其用于产生指示多个音频对象至至少两个下混声道的分配的下混信息；音频对象参数产生器，其用于产生音频对象的对象参数；以及输出接口，其用于使用该下混信息及对象参数来产生导入的音频输出信号。音频合成器使用下混信息来产生可用于建立预定义音频输出配置的多个输出声道的输出数据。

[0013] 希望的是无延迟地具有对输出平均响度或平均响度改变的准确估计，且当节目不改变或呈现场景不改变时，平均响度估计也应该保持静态。

[0014] 本发明的目标是提供改进的音频信号编码、处理及解码构思。本发明的目标通过以下来解决：根据权利要求1所述的解码器、根据权利要求15所述的编码器、根据权利要求18所述的系统、根据权利要求19所述的方法、根据权利要求20所述的方法及根据权利要求21所述的计算机程序。

[0015] 提供了用于估计基于对象的音频编码系统中的输出的响度的通知方式。所提供的构思依赖于关于待提供至解码器的音频混合中的对象的响度的信息。解码器使用此信息连同呈现信息用于估计输出信号的响度。例如，此举允许接着估计默认下混与呈现的输出之间的响度差异。接着可以补偿该差异以无关于呈现信息而获得输出中的大致恒定的响度。解码器中的响度估计以全参数方式发生，且与基于信号的响度估计构思相比，其在计算上非常轻便且准确。

[0016] 提供了用于使用纯参数概念获得关于特定输出场景的响度的信息的构思，其因此允许在不进行解码器中的明确的基于信号的响度估计的情况下进行响度处理。此外，描述了通过MPEG[SAOC]标准化的空间音频对象编码(SAOC)的特定技术，但所提供的构思亦可结合其他音频对象编码技术来使用。

[0017] 提供了用于产生包括一个或更多个音频输出声道的音频输出信号的解码器。该解码器包括接收接口，该接收接口用于接收包括多个音频对象信号的音频输入信号，用于接收关于该音频对象信号的响度信息，且用于接收指示该音频对象信号中的一个或更多个应放大还是衰减的呈现信息。此外，该解码器包括用于产生该音频输出信号的一个或更多个音频输出声道的信号处理器。该信号处理器被配置成取决于该响度信息且取决于该呈现信息来确定响度补偿值。此外，该信号处理器被配置成取决于该呈现信息且取决于该响度补偿值而自该音频输入信号产生该音频输出信号的所述一个或更多个音频输出声道。

[0018] 根据一实施方式，该信号处理器可被配置成取决于该呈现信息且取决于该响度补偿值而自该音频输入信号产生该音频输出信号的一个或更多个音频输出声道，使得该音频输出信号的响度等于该音频输入信号的响度，或使得该音频输出信号的响度比经修改音频

信号的响度更接近于该音频输入信号的响度，该经修改音频信号是通过根据该呈现信息放大或衰减该音频输入信号的音频对象信号而修改该音频输入信号来产生。

[0019] 根据另一实施方式，可将音频输入信号的音频对象信号中的每个指派给两个或更多个组中的确切的一个组，其中所述两个或更多个组中的每个可包括该音频输入信号的音频对象信号中的一个或更多个。在此实施方式中，该接收接口可被配置成接收所述两个或更多个组中的每一个组的响度值作为响度信息，其中该响度值指示该组的所述一个或更多个音频对象信号的原始总响度。此外，该接收接口可被配置成接收针对所述两个或更多个组中的至少一个组通过指示该组的一个或更多个音频对象信号的经修改总响度而指示应放大衰减该组的所述一个或更多个音频对象信号的呈现信息。此外，在此实施方式中，该信号处理器可被配置成取决于所述两个或更多个组中的所述至少一个组中的每个的经修改总响度且取决于所述两个或更多个组中的每个的原始总响度来确定该响度补偿值。此外，该信号处理器可被配置成取决于所述两个或更多个组中的所述至少一个组中的每个的经修改总响度且取决于该响度补偿值而自该音频输入信号产生该音频输出信号的一个或更多个音频输出声道。

[0020] 在特定实施方式中，所述两个或更多个组中的至少一个组可包括该音频对象信号中的两者或更多。

[0021] 此外，提供了一种编码器。该编码器包括基于对象的编码单元，该基于对象的编码单元用于编码多个音频对象信号以获得包括所述多个音频对象信号的编码音频信号。此外，该编码器包括用于编码关于该音频对象信号的响度信息的对象响度编码单元。该响度信息包括一个或更多个响度值，其中所述一个或更多个响度值中的每个取决于该音频对象信号中的一个或更多个。

[0022] 根据一实施方式，可将编码音频信号的音频对象信号中的每个指派给两个或更多个组中的确切的一个组，其中所述两个或更多个组中的每个包括该编码音频信号的音频对象信号中的一个或更多个。该对象响度编码单元可被配置成通过确定所述两个或更多个组中的每一个组的响度值来确定该响度信息的所述一个或更多个响度值，其中该组的响度值指示该组的所述一个或更多个音频对象信号的原始总响度。

[0023] 此外，提供了一种系统。该系统包括根据上述实施方式中的一个的编码器，该编码器用于编码多个音频对象信号以获得包括该多个音频对象信号的编码音频信号，且用于编码关于该音频对象信号的响度信息。此外，该系统包括根据上述实施方式中的一个的解码器，该解码器用于产生包括一个或更多个音频输出声道的音频输出信号。该解码器被配置成接收作为音频输入信号的编码音频信号及响度信息。此外，该解码器被配置成还接收呈现信息。此外，该解码器被配置成取决于该响度信息且取决于该呈现信息来确定响度补偿值。此外，该解码器被配置成取决于该呈现信息且取决于该响度补偿值而自该音频输入信号产生该音频输出信号的所述一个或更多个音频输出声道。

[0024] 此外，提供了一种用于产生包括一个或更多个音频输出声道的音频输出信号的方法。该方法包括：

[0025] -接收包括多个音频对象信号的音频输入信号。

[0026] -接收关于该音频对象信号的响度信息。

[0027] -接收指示该音频对象信号中的一个或更多个是否应放大或衰减的呈现信息。

[0028] -取决于该响度信息且取决于该呈现信息来确定响度补偿值。及：

[0029] -取决于该呈现信息且取决于该响度补偿值而自该音频输入信号产生该音频输出信号的所述一个或更多个音频输出声道。

[0030] 此外，提供了一种用于编码的方法。该方法包括：

[0031] -编码包括多个音频对象信号的音频输入信号。及：

[0032] -编码关于该音频对象信号的响度信息，其中该响度信息包括一个或更多个响度值，其中所述一个或更多个响度值中的每个取决于该音频对象信号中的一个或更多个。

[0033] 此外，提供了一种用于在计算机或信号处理器上执行时实施上述方法的计算机程序。

[0034] 在所附权利要求中提供了优选实施方式。

[0035] 下文中，参考附图更详细地描述了本发明的实施方式，在附图中：

[0036] 图1图示了根据一实施方式的用于产生包括一个或更多个音频输出声道的音频输出信号的解码器，

[0037] 图2图示了根据一实施方式的编码器，

[0038] 图3图示了根据一实施方式的系统，

[0039] 图4图示了包括SAOC编码器及SAOC解码器的空间音频对象编码系统，

[0040] 图5图示了包括边信息解码器、对象分离器及呈现器的SAOC解码器，

[0041] 图6图示了对响度改变的输出信号响度估计的特征，

[0042] 图7描绘根据一实施方式的通知响度估计，其图示了根据一实施方式的编码器及解码器的组件，

[0043] 图8图示了根据另一实施方式的编码器，

[0044] 图9图示了根据一实施方式的关于包括旁路声道的SAOC对话增强的编码器及解码器，

[0045] 图10描绘使用以参数方式估计响度改变的所提供构思来测量响度改变及其结果的第一图解，

[0046] 图11描绘使用以参数方式估计响度改变的所提供构思来测量响度改变及其结果的第二图解，以及

[0047] 图12图示了用于进行响度补偿的另一实施方式。

[0048] 在详细描述优选实施方式之前，描述响度估计、空间音频对象编码(SAOC)及对话增强(DE)。

[0049] 首先，描述响度估计。

[0050] 如之前所述，EBU推荐R128依赖于ITU-R BS.1770中提出的模型来进行响度估计。虽然将此测量用作示例，但下文所描述的构思亦可应用于其他响度测量。

[0051] 根据BS.1770的响度估计操作相对简单，且基于以下主要步骤[ITU]：

[0052] -用K滤波器(倾斜型滤波器与高通滤波器的组合)对输入信号x_i(或多声道信号情况下的信号)进行滤波以获得信号y_i。

[0053] -计算信号y_i的均方能量z_i。

[0054] -在多声道信号的情况下，应用声道权重G_i，且对经加权信号进行求和。则将信号响度定义为

[0055] $$L = c + 10\log_{10}\sum_i G_i z_i ,$$

[0056] 其中常数值c＝-0.691。因此是以“LKFS”(响度,经K加权,相对于全标度)为单位来表示输出,其标度类似于分贝标度。

[0057] 在上式中,对于声道中的一些声道,G_i可以例如等于1,而对于一些其他声道G_i可以例如为1.41。例如,若考虑左声道、右声道、中央声道、左环绕声道及右环绕声道,则相应的权重G_i对于左声道、右声道及中央声道可以例如为1,而对于左环绕声道及右环绕声道可以例如为1.41,参见[ITU]。

[0058] 可以看出,响度值L与信号能量的对数紧密相关。

[0059] 下文中,描述空间音频对象编码。

[0060] 基于对象的音频编码构思允许在链的解码器侧具有很大灵活性。基于对象的音频编码构思的示例为空间音频对象编码(SAOC)。

[0061] 图4图示了包括空间音频对象编码(SAOC)编码器410及SAOC解码器420的SAOC系统。

[0062] SAOC编码器410接收N个音频对象信号$S_1,\ldots,S_N$作为输入。此外,SAOC编码器410还接收指令“混合信息D”:应如何组合这些对象以获得包括M个下混声道$X_1,\ldots,X_M$的下混信号。SAOC编码器410自对象及下混过程提取一些边信息,且传输此边信息和/或将其与下混信号一起储存。

[0063] SAOC系统的主要特性为:包括下混声道$X_1,\ldots,X_M$的下混信号X形成语义上有意义的信号。换言之,可以听取下混信号。例如,若接收器不具有SAOC解码器功能性,则接收器也仍然可以始终将下混信号提供为输出。

[0064] 图5图示了包括边信息解码器510、对象分离器520及呈现器530的SAOC解码器。图5示出的SAOC解码器例如自SAOC编码器接收下混信号及边信息。下混信号可视为包括音频对象信号的音频输入信号,这是因为音频对象信号在下混信号内混合(音频对象信号在下混信号的一个或更多个下混声道内混合)。

[0065] 然后,SAOC解码器可以例如试图(虚拟地)重构原始对象,例如通过使用对象分离器520,例如使用解码边信息。然后,这些(虚拟)对象重构$\hat{S}_1, \ldots, \hat{S}_N$(例如,重构的音频对象信号)基于呈现信息(例如呈现矩阵R)组合以产生音频输出信号Y的K个音频输出声道$Y_1,\ldots,Y_K$。

[0066] 通常,在SAOC中,例如,音频对象信号例如通过使用自SAOC编码器传输至SAOC解码器的协方差信息例如信号协方差矩阵E加以重构。

[0067] 例如,可使用下式来在解码器侧重构音频对象信号:

[0068] $S = GX$其中$G \approx E\ D^H (D\ E\ D^H)^{-1}$

[0069] 其中

[0070] N 音频对象信号的数目

[0071] $N_{样本}$ 音频对象信号的所考虑样本的数目

[0072] M 下混声道的数目,

[0073] X 下混音频信号,大小为$M \times N_{样本}$,

[0074] D 下混矩阵,大小为M×N

[0075] E 定义为$E=X\ X^H$的大小为N×N的信号协方差矩阵

[0076] S 参数重构的N个音频对象信号，大小为$N\times N_{样本}$

[0077] $(\cdot)^H$ 自伴(赫密特)算子，其表示(·)的共轭转置

[0078] 接着，可将呈现矩阵R应用于经重构的音频对象信号S以获得音频输出信号Y的音频输出声道，例如根据下式：

[0079] $Y=RS$

[0080] 其中

[0081] K 音频输出信号Y的音频输出声道$Y_1,\ldots,Y_K$的数目。

[0082] R 大小为K×N的呈现矩阵

[0083] Y 包括K个音频输出声道的音频输出信号，大小为$K\times N_{样本}$

[0084] 在图5中，例如由对象分离器520进行对象重构的过程加上了标记“虚拟”或“可选”，因为其不是必须发生，但是可以通过在参数域中组合重构与呈现步骤(即，将等式进行组合)而获得希望的功能性。

[0085] 换言之，替代首先使用混合信息D及协方差信息E重构音频对象信号且接着将呈现信息R应用于经重构的音频对象信号以获得音频输出声道$Y_1,\ldots,Y_K$，可将两个步骤在单一步骤中进行，以使得直接自下混声道产生音频输出声道$Y_1,\ldots,Y_K$。

[0086] 例如，可使用下式：

[0087] $Y=RGX$其中$G\approx E\ D^H(D\ E\ D^H)^{-1}$。

[0088] 原理上，呈现信息R可请求原始音频对象信号的任何组合。然而，在实际中，对象重构可能包括重构误差，且所请求的输出场景可能未必能达到。作为涵盖许多实际情况的大致一般规则，所请求的输出场景与下混信号差异愈大，可听到的重构误差将愈大。

[0089] 下文中，描述对话增强(DE)。可以例如使用SAOC技术来实现场景。应注意，虽然名称“对话增强”提示着重于面向对话的信号，但是相同原理亦可用于其他信号类型。

[0090] 在DE场景中，系统中的自由度受限于该一般情况。

[0091] 例如，音频对象信号$S_1,\ldots,S_N=S$分组(且可能混合)为两个元对象：前景对象(FGO)S_{FGO}及背景对象(BGO)S_{BGO}。

[0092] 此外，输出场景$Y_1,\ldots,Y_K=Y$类似于下混信号$X_1,\ldots,X_M=X$。更具体地，两个信号具有相同维度，即K=M，且最终用户可以仅控制两个元对象FGO及BGO的相对混合水平。更确切而言，通过用一些标量权重来对FGO与BGO进行混合而获得下混信号

[0093] $X=h_{FGO}S_{FGO}+h_{BGO}S_{BGO}$，

[0094] 且类似地通过对FGO与BGO的标量加权而获得输出场景：

[0095] $Y=g_{FGO}S_{FGO}+g_{BGO}S_{BGO}$。

[0096] 取决于混合权重的相对值，FGO与BGO之间的平衡可能改变。例如，对于设定

[0097] $$\begin{cases} g_{FGO} > h_{FGO} \\ g_{BGO} = h_{BGO} \end{cases}$$

[0098] 可以在混合中增大FGO的相对水平。若FGO为对话，则此设定提供了对话增强功能性。

[0099] 作为使用情况示例，BGO可以为体育场噪音及体育赛事期间的其他背景声音，而FGO为解说员的语音。DE功能性允许最终用户相对于背景放大或衰减解说员的水平。

[0100] 各实施方式基于以下发现：在广播场景中利用SAOC技术(或类似技术)能够向最终用户提供扩展的信号操纵功能性。提供了比仅改变声道及调整播放音量更多的功能性。

[0101] 上文简要描述了使用DE技术的一个可能性。若作为用于SAOC的下混信号的广播信号例如根据R128而在水平上规范化，则在不施加(SAOC)处理(或呈现描述与下混描述相同)时，不同节目具有类似的平均响度。然而，当施加某一(SAOC)处理时，输出信号不同于默认下混信号，且输出信号的响度可不同于默认下混信号的响度。自最终用户的角度而言，此可导致以下情形：声道或节目之间的输出信号响度可再次具有不期望的跳跃或差异。换言之，部分地失去了通过广播台施加的规范化的益处。

[0102] 此问题并非SAOC或DE场景所独有，而是在允许最终用户与内容互动的其他音频编码构思中亦可能发生。然而，在许多情况下，若输出信号与默认下混具有不同的响度，其并不造成任何伤害。

[0103] 如上所述，音频输入信号节目的总响度应等于指定水平，允许有小的偏差。然而，如已经概述的那样，此在进行音频呈现时导致显著问题，因为呈现对所接收音频输入信号的总体/总响度可具有显著影响。然而，尽管进行场景呈现，所接收音频信号的总响度应保持相同。

[0104] 一种方法是在播放信号时估计信号的响度，且通过适当的时间整合构思，在某一时间之后该估计可以收敛至真实的平均响度。然而，收敛所需的时间从最终用户的角度来说是个问题。当即使在没有对信号施加改变时改变了响度估计，响度改变补偿也会起作用且改变其行为。此将导致输出信号具有时变平均响度，其在感觉上相当恼人。

[0105] 图6图示了响度改变时的输出信号响度估计行为。尤其描绘了基于信号的输出信号响度估计，其图示了刚刚描述的解决方案的效果。该估计相当缓慢地接近正确估计。替代基于信号的输出信号响度估计，立即正确地确定输出信号响度的通知输出信号响度估计是优选的。

[0106] 具体地，在图6中，用户输入(例如，对话对象的水平)通过增大其值而在时刻T改变。真实输出信号水平及相应地响度在同一时刻改变。当利用某一时间整合的时间自输出信号进行输出信号响度估计时，该估计将逐渐变化且在某一延迟之后达到正确值。在此延迟期间，估计值不断变化，因此不能可靠地用于进一步处理输出信号，例如，用于响度水平校正。

[0107] 如已经描述的，期望的是具有无延迟地输出平均响度或平均响度改变的准确估计，且当节目不改变或呈现场景不改变时，平均响度估计亦应保持静态。换言之，当应用某一响度改变补偿时，补偿参数应仅在节目改变或存在某用户互动时才改变。

[0108] 图6的最下图解图示了希望的行为(通知输出信号响度估计)。输出信号响度估计应在用户输入改变时即刻改变。

[0109] 图2图示了根据一实施方式的编码器。

[0110] 该编码器包括基于对象的编码单元210，其用于编码多个音频对象信号以获得包括所述多个音频对象信号的编码音频信号。

[0111] 此外，该编码器包括对象响度编码单元220，其用于对关于该音频对象信号的响度信息进行编码。该响度信息包括一个或更多个响度值，其中一个或更多个响度值中的每个取决于音频对象信号中的一个或更多个。

[0112] 根据一实施方式，编码音频信号的音频对象信号中的每个被指派给两个或更多个组中的确切的一个组，其中两个或更多个组中的每个包括编码音频信号的音频对象信号中的一个或更多个。该对象响度编码单元220被配置成通过确定两个或更多个组中的每一个组的响度值来确定该响度信息中的一个或更多个响度值，其中该组的响度值指示该组的一个或更多个音频对象信号的原始总响度。

[0113] 图1图示了根据一实施方式的用于产生包括一个或更多个音频输出声道的音频输出信号的解码器。

[0114] 该解码器包括接收接口110，其用于接收包括多个音频对象信号的音频输入信号，用于接收关于该音频对象信号的响度信息，且用于接收指示音频对象信号中的一个或更多个是否应放大或衰减的呈现信息。

[0115] 此外，该解码器包括信号处理器120，其用于产生音频输出信号的一个或更多个音频输出声道。信号处理器120被配置成取决于响度信息且取决于呈现信息来确定响度补偿值。此外，信号处理器120被配置成取决于呈现信息且取决于响度补偿值而自音频输入信号产生音频输出信号的一个或更多个音频输出声道。

[0116] 根据一实施方式，信号处理器110被配置成取决于呈现信息且取决于响度补偿值而自音频输入信号产生音频输出信号的一个或更多个音频输出声道，使得音频输出信号的响度等于音频输入信号的响度，或使得音频输出信号的响度比经修改音频信号的响度更接近于音频输入信号的响度，该经修改音频信号是通过根据呈现信息修改或衰减音频输入信号的音频对象信号而修改该音频输入信号而产生的。

[0117] 根据另一实施方式，音频输入信号的音频对象信号中的每个被指派给两个或更多个组中的确切的一个组，其中两个或更多个组中的每个包括音频输入信号的音频对象信号中的一个或更多个。

[0118] 在此实施方式中，接收接口110被配置成接收两个或更多个组中的每一个组的响度值作为响度信息，其中该响度值指示该组的一个或更多个音频对象信号的原始总响度。此外，接收接口110被配置成接收对于两个或更多个组中的至少一个组通过指示该组的一个或更多个音频对象信号的经修改总响度而指示该组的一个或更多个音频对象信号是否应放大或衰减的呈现信息。此外，在此实施方式中，信号处理器120被配置成取决于两个或更多个组中的所述至少一个组中的每个的经修改总响度且取决于两个或更多个组中的每个的原始总响度来确定响度补偿值。此外，信号处理器120被配置成取决于两个或更多个组中的所述至少一个组中的每个的经修改总响度且取决于该响度补偿值而自该音频输入信号产生音频输出信号的一个或更多个音频输出声道。

[0119] 在特定实施方式中，两个或更多个组中的至少一个组包括音频对象信号中的两者或更多。

[0120] 音频对象信号i的能量e_i与音频对象信号i的响度L_i之间根据下式存在直接关系：

[0121] $$L_i = c + 10\log_{10} e_i, \qquad e_i = 10^{(L_i - c)/10}$$

[0122] 其中c为常数值。

[0123] 实施方式基于以下发现：音频输入信号的不同音频对象信号可具有不同响度且因此具有不同能量。例如，若用户想要增大音频对象信号的响度，则可以相应地调整呈现信息，且此音频对象信号的响度增大增大了此音频对象的能量。此将导致音频输出信号的响

度增大。为保持总响度恒定，需要进行响度补偿。换言之，需要调整将呈现信息应用于音频输入信号而产生的经修改音频信号。然而，音频对象信号中的一个的放大对经修改音频信号的总响度的确切影响取决于经放大音频对象信号例如响度增大的音频对象信号的原始响度。若此对象的原始响度对应于相当低的能量，则对音频输入信号的总响度的影响将微乎其微。然而，若此对象的原始响度对应于相当高的能量，则对音频输入信号的总响度的影响将是显著的。

[0124] 可考虑两个示例。在两个示例中，音频输入信号包括两个音频对象信号，且在两个示例，通过应用呈现信息，音频对象信号中的第一个的能量增大50％。

[0125] 在第一示例中，第一音频对象信号贡献音频输入信号总能量的20％且第二音频对象信号贡献音频输入信号总能量的80％。然而，在第二示例中，第一音频对象，第一音频对象信号贡献音频输入信号总能量的40％且第二音频对象信号贡献音频输入信号总能量的60％。在两个示例中，可以自关于音频对象信号的响度信息得出这些贡献，因为响度与能量之间存在直接关系。

[0126] 在第一示例中，第一音频对象的能量的50％的增大导致通过将呈现信息应用于音频输入信号而产生的经修改音频信号具有的总能量为音频输入信号的能量的1.5×20％+80％＝110％。

[0127] 在第二示例中，第一音频对象的能量的50％的增大导致通过将呈现信息应用于音频输入信号而产生的经修改音频信号具有的总能量为音频输入信号的能量的1.5×40％+60％＝120％。

[0128] 因此，在将呈现信息应用于音频输入信号之后，在第一示例中，经修改音频信号的总能量仅须减小9％(10/110)以在音频输入信号与音频输出信号两者中获得相等能量，而在第二示例中，经修改音频信号的总能量须减小17％(20/120)。为此，可以计算响度补偿值。

[0129] 例如，响度补偿值可为应用于音频输出信号的所有音频输出声道的标量。

[0130] 根据一实施方式，信号处理器被配置成通过根据呈现信息修改或衰减音频输入信号的音频对象信号而修改音频输入信号来产生经修改音频信号。此外，信号处理器被配置成通过将响度补偿值应用于经修改音频信号而产生音频输出信号，使得音频输出信号的响度等于音频输入信号的响度，或使得音频输出信号的响度比经修改音频信号的响度更接近于音频输入信号的响度。

[0131] 例如，在上文第一示例中，响度补偿值lcv可例如设定为值lcv＝10/11，且可将放大因子10/11应用于根据呈现信息呈现音频输入声道而产生的所有声道。

[0132] 因此，例如，在上文第二示例中，响度补偿值lcv可例如设定为值lcv＝10/12＝5/6，且可将放大因子5/6应用于根据呈现信息呈现音频输入声道而产生的所有声道。

[0133] 在其他实施方式中，可以将音频对象信号中的每个指派给多个组中的一个，且可针对组中的每个传输指示该组的音频对象信号的总响度值的响度值。如呈现信息指定该等组中的一个的能量被衰减或放大，例如，如上所述放大了50％，则可计算总能量增加，且可如上所述确定响度补偿值。

[0134] 例如，根据一实施方式，音频输入信号的音频对象信号中的每个被指派给作为两个或更多个组的确切的两个组中的确切的一个组。音频输入信号的音频对象信号中的每个

被指派给所述确切的两个组中的前景对象组或所述确切的两个组中的背景对象组。接收接口110被配置成接收前景对象组的一个或更多个音频对象信号的原始总响度。此外，接收接口110被配置成接收背景对象组的一个或更多个音频对象信号的原始总响度。此外，接收接口110被配置成接收针对所述确切的两个组中的至少一个组通过指示该组的一个或更多个音频对象信号的经修改总响度而指示所述至少一个组中的每个的一个或更多个音频对象信号是否应放大或衰减的呈现信息。

[0135] 在此实施方式中，信号处理器120被配置成取决于所述至少一个组中的每个的经修改总响度、取决于前景对象组的一个或更多个音频对象信号的原始总响度且取决于背景对象组的一个或更多个音频对象信号的原始总响度来确定响度补偿值。此外，信号处理器120被配置成取决于所述至少一个组中的每个的经修改总响度且取决于该响度补偿值而自音频输入信号产生音频输出信号的一个或更多个音频输出声道。

[0136] 根据一些实施方式，音频对象信号中的每个被指派给三个或更多个组中的一个，且接收接口可以被配置成接收三个或更多个组中的每个的响度值，该响度值指示该组的音频对象信号的总响度。

[0137] 根据一实施方式，为确定两个或更多个音频对象信号的总响度值，例如，针对每一音频对象信号确定对应于响度值的能量值，对所有响度值的能量值求和以获得能量总和，且将对应于该能量总和的响度值确定为两个或更多个音频对象信号的总响度值。例如，可使用下式

[0138] $$L_i = c + 10\log_{10} e_i, \qquad e_i = 10^{(L_i - c)/10}$$

[0139] 在一些实施方式中，针对音频对象信号中的每个传输响度值，或将音频对象信号中的每个指派给一个或两个或更多个组，其中针对组中的每个，传输响度值。

[0140] 然而，在一些实施方式中，对于一个或更多个音频对象信号或对于包括音频对象信号的组中的一个或更多个，不传输响度值。替代地，解码器可以例如假定未传输其响度值的音频对象信号或音频对象信号组具有预定义响度值。例如，解码器可以使所有进一步的确定基于此预定义响度值。

[0141] 根据一实施方式，接收接口110被配置成接收包括一个或更多个下混声道的下混信号作为音频输入信号，其中一个或更多个下混声道包括音频对象信号，且其中音频对象信号的数目小于一个或更多个下混声道的数目。接收接口110被配置成接收指示音频对象信号如何在一个或更多个下混声道内混合的下混信息。此外，信号处理器120被配置成取决于该下混信息、取决于该呈现信息且取决于该响度补偿值而自该音频输入信号产生音频输出信号的一个或更多个音频输出声道。在特定实施方式中，例如，信号处理器120可以被配置成取决于下混信息而计算响度补偿值。

[0142] 例如，下混信息可以是下混矩阵。在一实施方式中，解码器可以是SAOC解码器。在这样的实施方式中，例如，接收接口110可以进一步被配置成接收协方差信息，例如，如上文所述的协方差矩阵。

[0143] 关于指示音频对象信号中的一个或更多个是否应放大或衰减的呈现信息，应注意，例如，指示音频对象信号中的一个或更多个应如何放大或衰减的信息为呈现信息。例如，呈现矩阵R，例如，SAOC的呈现矩阵为呈现信息。

[0144] 图3图示了根据一实施方式的系统。

[0145] 该系统包括根据上述实施方式中的一个的编码器310，其用于编码多个音频对象信号以获得包括该多个音频对象信号的编码音频信号。

[0146] 此外，该系统包括根据上述实施方式中的一个的解码器320，其用于产生包括一个或更多个音频输出声道的音频输出信号。该解码器被配置成接收作为音频输入信号的编码音频信号及响度信息。此外，解码器320被配置成还接收呈现信息。此外，解码器320被配置成取决于响度信息且取决于呈现信息来确定响度补偿值。此外，解码器320被配置成取决于该呈现信息且取决于该响度补偿值而自音频输入信号产生音频输出信号的一个或更多个音频输出声道。

[0147] 图7图示了根据一实施方式的通知响度估计。在传输流730的左方，图示了基于对象的音频编码编码器的组件。具体地，图示了基于对象的编码单元710(“基于对象的音频编码器”)及对象响度编码单元720(“对象响度估计”)。

[0148] 传输流730自身包括响度信息L、下混信息D及基于对象的音频编码器710的输出B。

[0149] 在传输流730的右方，图示了基于对象的音频编码解码器的信号处理器的组件。未示出解码器的接收接口。描绘了输出响度估计器740及基于对象的音频解码单元750。输出响度估计器740可以被配置成确定响度补偿值。基于对象的音频解码单元750可以被配置成通过应用呈现信息R而自输入至解码器的音频信号确定经修改音频信号。图7中未示出将响度补偿值应用于经修改音频信号以补偿由呈现引起的总响度改变。

[0150] 至编码器的输入至少是由输入对象S组成。该系统例如通过对象响度编码单元720估计每个对象的响度(或某其他响度相关信息，例如对象能量)，且对信息L进行传输和/或储存。(也可以将对象的响度提供为至系统的输入，且可省略系统内的估计步骤)。

[0151] 在图7的实施方式中，解码器接收至少对象响度信息及例如呈现信息R，呈现信息R描述对象至输出信号的混合。基于这些信息，例如输出响度估计器740估计输出信号的响度且将此信息提供为其输出。

[0152] 下混信息D可提供为呈现信息，在此情况下，响度估计提供对下混信号响度的估计。还可以将下混信息提供为至对象响度估计的输入，且连同对象响度信息传输和/或储存该信息。然后，输出响度估计可以同时估计下混信号的响度及所呈现的输出，且将这两个值或其差异提供为输出响度信息。差异值(或其倒数)描述所需补偿，所需补偿应该应用于所呈现输出信号以使得其响度类似于下混信号的响度。对象响度信息可另外包含关于各个对象之间的相关系数的信息，且此相关信息可用于输出响度估计中以用于更准确的估计。

[0153] 下文中，描述用于对话增强应用的优选实施方式。

[0154] 在对话增强应用中，如上文所述，将输入音频对象信号分组且部分地下混以形成两个元对象FGO及BGO，然后，可以以普通方式对这两个元对象FGO及BGO求和以获得最终下混信号。

[0155] 遵循SAOC[SAOC]的描述，将N个输入对象信号表示为大小为$N\times N_{样本}$的矩阵S，且将下混信息表示为大小为M×N的矩阵D。则下混信号可以获得为X＝DS。

[0156] 现在，下混信息D对于元对象可分成两个部分

[0157] $D=D_{FGO}+D_{BGO}$。

[0158] 由于矩阵D的每一列对应于原始音频对象信号，可通过将对应于其他元对象的列

设定为零(假定无原始对象可存在于两个元对象上)而获得两个分量下混矩阵。换言之,对应于元对象BGO的列在D_{FGO}中设定为零,且对应于元对象FGO的列在D_{BGO}中设定为零。

[0159] 这些新下混矩阵描述可以自输入对象获得两个元对象的方式,即:

[0160] $S_{FGO}=D_{FGO}S$且$S_{BGO}=D_{BGO}S$,

[0161] 且实际下混简化为

[0162] $X=S_{FGO}+S_{BGO}$。

[0163] 也可以认为对象(例如,SAOC)解码器试图重构元对象:

[0164] $\tilde{\mathbf{S}}_{FGO}\approx\mathbf{S}_{FGO}$且$\tilde{\mathbf{S}}_{BGO}\approx\mathbf{S}_{BGO}$,

[0165] 且DE特定呈现可写为这两个元对象重构的组合:

[0166] $\mathbf{Y}=g_{FGO}\mathbf{S}_{FGO}+g_{BGO}\mathbf{S}_{BGO}\approx g_{FGO}\tilde{\mathbf{S}}_{FGO}+g_{BGO}\tilde{\mathbf{S}}_{BGO}$。

[0167] 对象响度估计将两个元对象S_{FGO}及S_{BGO}接收为输入,且估计其中每个的响度:L_{FGO}为S_{FGO}的(总/总体)响度,且L_{BGO}为S_{BGO}的(总/总体)响度。传输和/或储存这些响度值。

[0168] 替代地,使用元对象中的一个例如FGO作为基准,可以将这两个对象的响度差异计算为例如

[0169] $\Delta L_{FGO}=L_{BGO}-L_{FGO}$。

[0170] 然后,传输和/或储存此单个值。

[0171] 图8图示了根据另一实施方式的编码器。图8的编码器包括对象下混器811及对象边信息估计器812。此外,图8的编码器还包括对象响度编码单元820。此外,图8的编码器包括元音频对象混合器805。

[0172] 图8的编码器使用中间音频元对象作为至对象响度估计的输入。在实施方式中,图8的编码器可以被配置成产生两个音频元对象。在其他实施方式中,图8的编码器可以被配置成产生三个或更多个音频元对象。

[0173] 特别地,所提供的构思提供了下述新特征:编码器可以例如估计所有输入对象的平均响度。对象可以例如在所传输的下混信号内混合。此外,所提供的构思提供了下述新特征:可将对象响度及下混信息例如包含在所传输的对象编码边信息中。

[0174] 例如,解码器可使用对象编码边信息来(虚拟地)分离对象,且使用呈现信息重新组合对象。

[0175] 此外,所提供的构思提供了下述新特征:可以使用下混信息来估计默认下混信号的响度;可以使用呈现信息及所接收对象响度来估计输出信号的平均响度;并且/或者可自这两个值估计响度改变。或者,可使用下混及呈现信息自默认下混估计响度改变,这是所提供的构思的另一新特征。

[0176] 此外,所提供的构思提供了下述新特征:可修改解码器输出以补偿响度改变以使得经修改信号的平均响度匹配默认下混的平均响度。

[0177] 图9中图示了与SAOC-DE有关的特定实施方式。系统接收输入音频对象信号、下混信息、以及对象至元对象的分组信息。基于这些信息,元音频对象混合器905形成两个元对象S_{FGO}及S_{BGO}。有可能用SAOC处理的信号部分不构成整个信号。例如,在5.1声道配置中,SAOC可以部署于声道的子集中,例如部署于前声道(左、右及中央),而其他声道(左环绕、右环绕及低频效果)绕开(旁路)SAOC且如此递送。未经SAOC处理的这些声道表示为X_{BYPASS}。需要提

供可能的旁路声道以使编码器更准确地估计响度信息。

[0178] 可以多种方式处置旁路声道。

[0179] 例如，旁路声道可以例如形成独立元对象。这允许限定呈现以使得所有三个元对象独立地进行按比例调整。

[0180] 或者，例如，旁路声道可以例如与其他两个元对象中的一个组合。该元对象的呈现设定也控制旁路声道部分。例如，在对话增强场景中，将旁路声道与背景元对象进行合并可以是有意义的：$X_{BGO}=S_{BGO}+X_{BYPASS}$。

[0181] 或者，例如，可以例如忽略旁路声道。

[0182] 根据实施方式，编码器的基于对象的编码单元210被配置成接收音频对象信号，其中音频对象信号中的每个被指派给确切的两个组中的确切的一个，其中所述确切的两个组中的每个包括音频对象信号中的一个或更多个。此外，基于对象的编码单元210被配置成对所述确切的两个组包含的音频对象信号进行下混以获得包括一个或更多个下混音频声道的下混信号作为编码音频信号，其中一个或更多个下混声道的数目小于所述确切的两个组所包含的音频对象信号的数目。对象响度编码单元220被指派成接收一个或更多个其他旁路音频对象信号，其中一个或更多个其他旁路音频对象信号中的每个被指派给第三组，其中一个或更多个其他旁路音频对象信号中的每个不被第一组包括且不被第二组包括，其中基于对象的编码单元210被配置成不在下混信号内下混一个或更多个其他旁路音频对象信号。

[0183] 在一实施方式中，对象响度编码单元220被配置成确定响度信息的第一响度值、第二响度值及第三响度值，第一响度值指示第一组的一个或更多个音频对象信号的总响度，第二响度值指示第二组的一个或更多个音频对象信号的总响度，且第三响度值指示第三组的一个或更多个其他旁路音频对象信号的总响度。在另一实施方式中，对象响度编码单元220被配置成确定响度信息的第一响度值及第二响度值，第一响度值指示第一组的一个或更多个音频对象信号的总响度，且第二响度值指示第二组的一个或更多个音频对象信号及第三组的一个或更多个其他旁路音频对象信号的总响度。

[0184] 根据一实施方式，解码器的接收接口110被配置成接收下混信号。此外，接收接口110被配置成接收一个或更多个其他旁路音频对象信号，其中一个或更多个其他旁路音频对象信号未在该下混信号内混合。此外，接收接口110被配置成接收响度信息，该响度信息指示关于在下混信号内混合的音频对象信号的响度的信息且指示关于未在下混信号内混合的一个或更多个其他旁路音频对象信号的响度的信息。此外，信号处理器120被配置成取决于关于在下混信号内混合的音频对象信号的响度的信息且取决于关于未在下混信号内混合的一个或更多个其他旁路音频对象信号的响度的信息来确定响度补偿值。

[0185] 图9图示了根据与SAOC-DE有关的实施方式的编码器及解码器，其包括旁路声道。特别地，图9的编码器包括SAOC编码器902。

[0186] 在图9的实施方式中，旁路声道与另一元对象的可能组合发生在两个“旁路包含”块913、914中，从而产生具有来自所包含的旁路声道的所限定部分的元对象X_{FGO}及X_{BGO}。

[0187] 在响度估计单元921、922、923中估计这两个元对象的感知响度L_{BYPASS}、L_{FGO}及L_{BGO}。然后，此响度信息在元对象响度信息估计器925中变换为适当编码，然后被传输和/或储存。

[0188] 实际SAOC编码器及解码器按照预期操作，从而自对象提取对象边信息、建立下混

信号X、且将该信息传输和/或储存至解码器。可能的旁路声道连同其他信息一起被传输和/或储存至解码器。

[0189] SAOC-DE解码器945接收增益值“对话增益”作为用户输入。基于此输入及所接收的下混信息，SAOC解码器945确定呈现信息。然后，SAOC解码器945产生所呈现的输出场景作为信号Y。此外，其产生应该应用于可能旁路信号X_{BYPASS}的增益因子(及延迟值)。

[0190] “旁路包含”单元955连同所呈现的输出场景及旁路信号一起接收此信息，且建立完整输出场景信号。SAOC解码器945也产生一组元对象增益值，这些增益值的量取决于元对象分组及期望的响度信息形式。

[0191] 将增益值提供至混合响度估计器960，混合响度估计器960还从编码器接收元对象响度信息。

[0192] 然后，混合响度估计器960能够确定期望的响度信息，期望的响度信息包含但不限于下混信号的响度、所呈现输出场景的响度和/或下混信号与所呈现输出场景之间的响度差异。

[0193] 在一些实施方式中，响度信息自身是足够的，而在其他实施方式中，希望的是取决于所确定的响度信息来处理完整输出。此处理可以例如为补偿下混信号与所呈现输出场景之间的任何可能响度差异。例如由响度处理单元970进行的此处理将在广播场景中有意义，因为其将减少所感知信号响度的改变而无关于用户互动(对输入“对话增益”的设定)。

[0194] 此特定实施方式中的响度相关处理包括多个新特征。特别是，FGO、BGO及可能的旁路声道在最终声道配置中预先混合，使得可通过简单地将两个预先混合的信号相加在一起而进行下混(例如，下混矩阵系数1)，这一点构成了新特征。此外，作为另一新特征，估计FGO与BGO的平均响度，且计算差异。此外，对象混合于所传输的下混信号中。此外，作为另一新特征，将响度差异信息包含至所传输的边信息。(新)此外，解码器使用该边信息用于(虚拟地)分离对象，且使用基于下混信息及用户输入修改增益的呈现信息来重新组合这些对象。此外，作为另一新特征，解码器使用修改增益及所传输的响度信息来估计与默认下混相比的系统输出的平均响度的改变。

[0195] 在下文中，提供实施方式的形式描述。

[0196] 假定当对对象求和时对象响度值的特征类似于能量值的对数，即响度值必须变换至线性域，在该处相加，且最终变换回至对数域。通过BS.1770的定义启示了这一点，现在介绍响度测量(为简单起见，声道的数目设定为1，但相同原理可适用于在声道上适当求和的多声道信号)。

[0197] 具有均方能量e_i的第i个K滤波信号z_i的响度定义为

[0198] $L_i = c + 10\log_{10} e_i$，

[0199] 其中c为偏移常数。例如，c可为-0.691。由此，可通过下式自响度确定信号的能量

[0200] $$e_i = 10^{(L_i - c)/10}$$。

[0201] N个不相关信号的能量总和$z_{SUM} = \sum_{i=1}^{N} z_i$由此为

[0202] $$e_{SUM} = \sum_{i=1}^{N} e_i = \sum_{i=1}^{N} 10^{(L_i - c)/10}$$，

[0203] 且此总和信号的响度由此为

[0204] $$L_{SUM}=c+10\log_{10}e_{SUM}=c+10\log_{10}\sum_{i=1}^{N}10^{(L_i-c)/10}$$。

[0205] 若信号不是不相关，则在对总和信号的能量求近似时必须考虑相关系数$C_{i,j}$：

[0206] $$e_{SUM}=\sum_{i=1}^{N}\sum_{j=1}^{N}e_{i,j}$$，

[0207] 其中，第i个对象与第j个对象之间的交叉能量$e_{i,j}$定义为

[0208] $$\begin{aligned}e_{i,j}&=C_{i,j}\sqrt{e_ie_j}\\&=C_{i,j}\sqrt{10^{(L_i-c)/10}10^{(L_j-c)/10}}\\&=C_{i,j}\sqrt{10^{(L_i+L_j-2c)/10}}\end{aligned}$$，

[0209] 其中，$-1\le C_{i,j}\le 1$为两个对象i与j之间的相关系数。当两个对象不相关时，相关系数等于0，且当两个对象相同时，相关系数等于1。

[0210] 利用待在混合过程中应用于信号的混合权重g_i来进一步扩展该模型，即$z_{SUM}=\sum_{i=1}^{N}g_iz_i$总和信号的能量将为，

[0211] $$e_{SUM}=\sum_{i=1}^{N}\sum_{j=1}^{N}g_ig_je_{i,j}$$，

[0212] 且可如前所述自此获得混合信号的响度，其中

[0213] $L_{SUM}=c+10\log_{10}e_{SUM}$。

[0214] 两个信号的响度之间的差可估计为

[0215] $\Delta L(i,j)=L_i-L_j$。

[0216] 若现在使用如前所述的响度定义，则此可写为

[0217] $$\begin{aligned}\Delta L(i,j)&=L_i-L_j\\&=\left(c+10\log_{10}e_i\right)-\left(c+10\log_{10}e_j\right)\\&=10\log_{10}\frac{e_i}{e_j}\end{aligned}$$，

[0218] 可观察到其为信号能量的函数。若现在需要估计具有可能不同的混合权重g_i及h_i的两个如下表示的混合之间的响度差异

[0219] $$z_A=\sum_{i=1}^{N}g_iz_i\text{和}z_B=\sum_{i=1}^{N}h_iz_i$$

[0220] 则此差可用下式进行估计

$$\Delta L(A,B)=10\log_{10}\frac{e_A}{e_B}$$

$$=10\log_{10}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}g_i g_j e_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_i h_j e_{i,j}}$$

[0221]

$$=10\log_{10}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}g_i g_j C_{i,j}\sqrt{10^{(L_i+L_j-2c)/10}}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_i h_j C_{i,j}\sqrt{10^{(L_i+L_j-2c)/10}}}$$

$$=10\log_{10}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}g_i g_j C_{i,j}\sqrt{10^{(L_i+L_j)/10}}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_i h_j C_{i,j}\sqrt{10^{(L_i+L_j)/10}}}$$

。

[0222] 在对象不相关的情况下($C_{i,j}=0,\forall i\neq j$且$C_{i,j}=1,\forall i=j$),差异估计变为

$$\Delta L(A,B)=10\log_{10}\frac{\sum_{i=1}^{N}g_i^2 10^{(L_i-c)/10}}{\sum_{i=1}^{N}h_i^2 10^{(L_i-c)/10}}$$

[0223]

$$=10\log_{10}\frac{\sum_{i=1}^{N}g_i^2 10^{L_i/10}}{\sum_{i=1}^{N}h_i^2 10^{L_i/10}}$$

。

[0224] 在下文中,考虑差编码。

[0225] 可以将每对象响度值编码为与所选基准对象的响度的差:

[0226] $K_i=L_i-L_{REF}$,

[0227] 其中,L_{REF}为基准对象的响度。此编码在不需要绝对响度值作为结果的情况下是有益的,这是因为现在有必要传输的值少了一个,且响度差异估计可写为:

[0228] $$\Delta L(A,B)=10\log_{10}\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}g_i g_j C_{i,j}\sqrt{10^{(K_i+K_j)/10}}}{\sum_{i=1}^{N}\sum_{j=1}^{N}h_i h_j C_{i,j}\sqrt{10^{(K_i+K_j)/10}}}$$,

[0229] 或在对象不相关的情况下写为:

[0230] $$\Delta L(A,B)=10\log_{10}\frac{\sum_{i=1}^{N}g_i^2 10^{K_i/10}}{\sum_{i=1}^{N}h_i^2 10^{K_i/10}}$$

。

[0231] 在下文中,考虑了对话增强场景。

[0232] 再次考虑对话增强的应用场景。在解码器中限定呈现信息的自由度仅限于改变两个元对象的水平。此外，假定两个元对象不相关，即$C_{FGO,BGO}=0$。若元对象的下混权重为h_{FGO}及h_{BGO}，且以增益f_{FGO}及f_{BGO}来呈现这些元对象，则相对于默认下混的输出响度为：

[0233]
$$\Delta L(A,B)=10\log_{10}\frac{f_{FGO}^2 10^{(L_{FGO}-c)/10}+f_{BGO}^2 10^{(L_{BGO}-c)/10}}{h_{FGO}^2 10^{(L_{FGO}-c)/10}+h_{BGO}^2 10^{(L_{BGO}-c)/10}}$$
$$=10\log_{10}\frac{f_{FGO}^2 10^{L_{FGO}/10}+f_{BGO}^2 10^{L_{BGO}/10}}{h_{FGO}^2 10^{L_{FGO}/10}+h_{BGO}^2 10^{L_{BGO}/10}}$$

[0234] 若期望输出与默认下混中具有相同响度，则上式也是所需要的补偿。

[0235] 可以认为ΔL(A,B)是响度补偿值，其可由解码器的信号处理器120传输。ΔL(A,B)也可以被命名为响度改变值，且因此实际补偿值可为倒数值。或者，其也可使用“响度补偿因子”来命名?因此，本文档上文中提及的响度补偿值lcv将对应于下文的值g_{Δ}。

[0236] 例如，$g_{\Delta}=10^{-\Delta L(A,B)/20}1/\Delta L(A,B)$可应用为经修改音频信号的每一声道上的放大因子，该经修改音频信号是通过将呈现信息应用于音频输入信号而产生的。g_{Δ}的此等式在线性域中有效。在对数域中，该等式将不同，例如1/ΔL(A,B)，且相应地应用该等式。

[0237] 若下混过程经简化而使得两个元对象可以以单位权重混合以获得下混信号，即，$h_{FGO}=h_{BGO}=1$，并且将两个对象的呈现增益表示为g_{FGO}及g_{BGO}，则将响度改变的等式简化为

[0238]
$$\Delta L(A,B)=10\log_{10}\frac{g_{FGO}^2 10^{(L_{FGO}-c)/10}+g_{BGO}^2 10^{(L_{BGO}-c)/10}}{10^{(L_{FGO}-c)/10}+10^{(L_{BGO}-c)/10}}$$
$$=10\log_{10}\frac{g_{FGO}^2 10^{L_{FGO}/10}+g_{BGO}^2 10^{L_{BGO}/10}}{10^{L_{FGO}/10}+10^{L_{BGO}/10}}$$

[0239] 再次，可以认为ΔL(A,B)是通过信号处理器120确定的响度补偿值。

[0240] 通常，可以认为g_{FGO}是前景对象FGO(前景对象组)的呈现增益，且可以认为g_{BGO}是背景对象BGO(背景对象组)的呈现增益。

[0241] 如前所述，可以传输响度差异而非绝对响度。将基准响度定义为FGO元对象的响度$L_{REF}=L_{FGO}$，即，$K_{FGO}=L_{FGO}-L_{REF}=0$且$K_{BGO}=L_{BGO}-L_{REF}=L_{BGO}-L_{FGO}$。现在，响度改变量为

[0242]
$$\Delta L(A,B)=10\log_{10}\frac{g_{FGO}^2+g_{BGO}^2 10^{K_{BGO}/10}}{1+10^{K_{BGO}/10}}$$

[0243] 如同SAOC-DE中的情况那样，也可以两个元对象不具有各自的比例因子，而是所述对象中的一个未经修改，而另一个衰减以获得对象之间的正确混合比率。在此呈现设定中，输出的响度将低于默认混合的响度，且响度的改变为：

[0244]
$$\Delta L(A,B)=10\log_{10}\frac{\hat{g}_{FGO}^2+\hat{g}_{BGO}^2 10^{K_{BGO}/10}}{1+10^{K_{BGO}/10}},$$

[0245] 其中

[0246] $$\hat{g}_{FGO}=\begin{cases}1 & 若\ g_{FGO}\geq g_{BGO}\\ \dfrac{g_{FGO}}{g_{BGO}} & 若\ g_{FGO}<g_{BGO}\end{cases},\ 且\hat{g}_{BGO}=\begin{cases}\dfrac{g_{BGO}}{g_{FGO}} & ,若\ g_{BGO}<g_{FGO}\\ 1 & ,若\ g_{BGO}\geq g_{FGO}\end{cases}。$$

[0247] 此形式已经相当简单，且相当无关于已使用的响度测量。唯一的现实要求为，响度值应在指数域中求和。可以传输/储存信号能量值而非响度值，因为两者具有紧密关联。

[0248] 在上式中的每个中，可以认为ΔL(A,B)是响度补偿值，其可由解码器的信号处理器120传输。

[0249] 在下文中，考虑示例情况。通过两个示例信号说明所提供构思的准确性。两个信号皆具有5.1下混，环绕及LFE声道自SAOC处理略过。

[0250] 使用两个主要方法：一个方法(“3项”)具有三个元对象：FGO、BGO及旁路声道，例如，

[0251] $X=X_{FGO}+X_{BGO}+X_{BYPASS}$，

[0252] 而另一个方法(“2项”)具有两个元对象，例如：

[0253] $X=X_{FGO}+X_{BGO}$。

[0254] 在2项方法中，旁路声道可例如与BGO混合在一起以进行元对象响度估计。对两个(或所有三个)对象的响度以及下混信号的响度进行估计，且将值进行储存。

[0255] 呈现指令对于这两个方法分别具有以下形式

[0256] $$\mathbf{Y}=\hat{g}_{FGO}\mathbf{X}_{FGO}+\hat{g}_{BGO}\mathbf{X}_{BGO}+\hat{g}_{BGO}\mathbf{X}_{BYPASS}$$

[0257] 及

[0258] $$\mathbf{Y}=\hat{g}_{FGO}\mathbf{X}_{FGO}+\hat{g}_{BGO}\mathbf{X}_{BGO}。$$

[0259] 增益值例如根据下式来确定：

[0260] $$\hat{g}_{FGO}=\begin{cases}1 & ,若\ g_{FGO}>1\\ g_{FGO} & ,否则\end{cases},\ 且\hat{g}_{BGO}=\begin{cases}1/g_{FGO} & ,若\ g_{FGO}>1\\ 1 & ,否则\end{cases},$$

[0261] 其中FGO增益g_{FGO}在-24至+24dB间变化。

[0262] 呈现输出场景，测量响度，且计算自下混信号响度的衰减。

[0263] 此结果用具有圆圈标记的蓝线显示于图10及图11中。图10描绘了测量的响度改变及使用所提供构思用于以纯参数方式估计响度改变的结果的第一图解，且图11描绘其第二图解。

[0264] 接下来，使用所储存的元对象响度值及下混信息以及呈现信息来以参数方式估计自下混的衰减。用具有方形标记的绿线图示使用三个元对象的响度估计，且用具有星形标记的红线图示使用两个元对象的响度估计。

[0265] 从图中可看出，2项方法及3项方法提供了实际上相同的结果，且其皆相当好地近似于测量值。

[0266] 所提供的构思展现多个优点。例如，所提供的构思允许自形成混合的分量信号的响度来估计混合信号的响度。此举的益处为可一次估计分量信号响度，且可针对任何混合以参数方式获得混合的响度估计而无需实际基于信号的响度估计。这在需要对各种混合进行响度估计的总体系统的计算效率上提供了相当大的改进。例如，当最终用户改变呈现设

定时，输出的响度估计立即可用。

[0267] 在一些应用中，例如当符合EBU R128推荐时，整个节目的平均响度很重要。若基于接收信号进行接收器中的响度估计，例如在广播场景中，则估计仅在已接收到整个节目之后才收敛至平均响度。因此，对响度的任何补偿将具有误差或出现时间变化。当根据所提出的内容来估计分量对象的响度且传输响度信息时，可以在接收器中无延迟地估计平均混合响度。

[0268] 若期望输出信号的平均响度保持(近似)恒定而无关于呈现信息的改变，则出于此原因，所提供的构思能够确定补偿因子。解码器中为此而所需的计算的计算复杂性可忽略，因此可以将该功能性添加至任何解码器。

[0269] 存在以下情况：输出的绝对响度水平并不重要，但重要的在于确定响度自基准场景的改变。在这样的情况下，对象的绝对水平并不重要，而其相对水平才是重要的。这允许将对象中的一个定义为基准对象且相对于此基准对象的响度表示其他对象的响度。考虑到响度信息的传输和/或储存，此举具有一些益处。

[0270] 首先，不必传输基准响度水平。在两个元对象的应用情况中，这将待传输的数据量减半。第二益处涉及响度值的可能量化及表示。由于对象的绝对水平可为几乎任何情况，因此绝对响度值亦可几乎为任何情况。另一方面，相对响度值被假定为具有0平均值及围绕平均值的形式相当良好的分布。各个表示形式之间的差异允许以准确度潜在较高的下述方式来限定相对表示的量化栅格，该方式中经量化的表示形式具有相同数目的比特。

[0271] 图12图示了用于进行响度补偿的另一实施方式。在图12中，可进行响度补偿以例如补偿响度损失。为此，例如可使用来自DE_control_info的值DE_loudness_diff_dialogue(=K_{FGO})及DE_loudness_diff_background(=K_{BGO})。此处，DE_control_info可指定高级干净音频“对话增强”(DE)控制信息

[0272] 响度补偿是通过将增益值“g”应用于SAOC-DE输出信号及旁路声道(在多声道信号的情况下)而实现。

[0273] 在图12的实施方式中，如下进行此操作：

[0274] 使用有限对话修改增益值m_G来确定前景对象(FGO，例如，对话)及背景对象(BGO，例如，周围环境)的有效增益。此操作由“增益映射”块1220进行，“增益映射”块1220产生增益值m_{FGO}及m_{BGO}。

[0275] “输出响度估计器”块1230使用响度信息K_{FGO}及K_{BGO}及有效增益值m_{FGO}及m_{BGO}来估计与默认下混情况相比的此可能响度改变。接着将该改变映射至“响度补偿因子”，该响度补偿因子应用于输出声道以用于产生最终“输出信号”。

[0276] 应用以下步骤来进行响度补偿：

[0277] -自SAOC-DE解码器(如在条款12.8“SAOC-DE的修改范围控制”[DE]中所定义的)接收受限增益值m_G，且确定所应用的FGO/BGO增益：

[0278] $$\begin{cases} m_{FGO}=m_G \text{ 且 } m_{BGO}=1 & \text{若 } m_G \le 1 \\ m_{FGO}=1 \text{ 且 } m_{BGO}=m_G^{-1} & \text{若 } m_G > 1 \end{cases}$$

[0279] -获得元对象响度信息K_{FGO}及K_{BGO}。

[0280] -利用下式计算与默认下混相比的输出响度的改变

[0281] $$\Delta L = 10\log_{10}\frac{m_{FGO}^2 10^{K_{FGO}/10} + m_{BGO}^2 10^{K_{BGO}/10}}{10^{K_{FGO}/10} + 10^{K_{BGO}/10}}$$。

[0282] -计算响度补偿增益$g_\Delta = 10^{-0.05\Delta L}$。

[0283] -计算比例因子$g=\begin{bmatrix} g_1 \\ \vdots \\ g_N \end{bmatrix}$，其中$g_i=\begin{cases} g_\Delta & \text{若声道i属于SAOC-DE输出} \\ m_{BGO}g_\Delta & \text{若声道i是旁路声道} \end{cases}$，且N为输出声道的总数目。在图12中，增益调整分成两个步骤：在可能“旁路声道”与“SAOC-DE输出声道”组合之前用m_{BGO}调整可能“旁路声道”的增益，接着将共同增益g_Δ应用于所有经组合声道。此仅为增益调整操作的一种可能重排序，而此处g将两个增益调整步骤组合成一个增益调整。

[0284] 将比例值g应用于由“SAOC-DE输出声道”Y_{SAOC}及可能时间对准的“旁路声道”Y_{BYPASS}组成的音频声道Y_{FULL}：$Y_{FULL}=Y_{SAOC}\cup Y_{BYPASS}$。

[0285] 将比例值g应用于音频声道Y_{FULL}是由增益调整单元1240进行的。

[0286] 可以认为如上文所计算的ΔL是响度补偿值。通常，m_{FGO}指示前景对象FGO(前景对象组)的呈现增益，且m_{BGO}指示背景对象BGO(背景对象组)的呈现增益。

[0287] 尽管在设备的上下文中已描述了一些方面，但清楚的是，这些方面也代表对应方法的描述，其中，块或装置对应于方法步骤或方法步骤的特征。类似地，在方法步骤的上下文中所描述的各方面也代表对应设备的对应块或项或特征的描述。

[0288] 创造性的分解信号可储存在数字储存介质上或可在诸如无线传输介质或有线传输介质的传输介质例如因特网上传输。

[0289] 取决于某些实施要求，本发明的实施方式可在硬件中或软件中实施。可使用储存有电可读控制信号的数字储存介质例如软磁盘、DVD、CD、ROM、PROM、EPROM、EEPROM或闪存来进行该实施方案，该数字储存介质与可编程计算机系统协作(或能够与可编程计算机系统协作)，使得进行各个方法。

[0290] 根据本发明的一些实施方式包括具有电可读控制信号的非易失性数据载体，其能够与可编程计算机系统协作，使得进行本文所述方法之一。

[0291] 通常，本发明的实施方式可实施为具有程序代码的计算机程序产品，当该计算机程序产品在计算机上运行时，该程序代码操作以用于进行所述方法中的一个。该程序代码可例如储存于机器可读载体上。

[0292] 其他实施方式包括储存于机器可读载体上的用于进行本文所述方法中的一个的计算机程序。

[0293] 换言之，本创造性方法的实施方式因此为具有程序代码的计算机程序，当该计算机程序在计算机上运行时，该程序代码用于进行本文所述的方法中的一个。

[0294] 本发明的方法的另一实施方式因此为数据载体(或数字储存介质，或计算机可读介质)，其上面记录有用于进行本文所述方法中的一个的计算机程序。

[0295] 本创造性方法的另一实施方式因此为表示用于进行本文所述方法中的一个的计算机程序的数据流或信号序列。该数据流或信号序列可例如被配置成经由数据通信连接例

如经由因特网传输。

[0296] 另一实施方式包括一种处理装置，例如计算机或可编程逻辑设备，其被配置成或适于进行本文所述方法中的一个。

[0297] 另一实施方式包括一种计算机，其上安装有用于进行本文所述方法中的一个的计算机程序。

[0298] 在一些实施方式中，一种可编程逻辑设备(例如，现场可编程门阵列)可用以进行本文所述方法的功能性中的一些或全部。在一些实施方式中，现场可编程门阵列可与微处理器协作，以便进行本文所述方法中的一个。通常，所述方法优选地由任何硬设备进行。

[0299] 上文所述的实施方式仅示出了本发明的原理。应理解，本文描述的布置及细节的修改及改变对于本领域的技术人员来说是显然的。因此，本发明仅受限于所附的专利权利要求的范围，而并非受限于本文中以对实施方式的描述和解释的方式所展现的具体细节。

[0300] 参考文献

[0301] [BCC]C.Faller and F.Baumgarte,“Binaural Cue Coding-Part II:Schemes and applications”,IEEE Trans.on Speech and Audio Proc.,vol.11,no.6,Nov.2003.

[0302] [EBU]EBU Recommendation R 128“Loudness normalization and permitted maximum level of audio signals”,Geneva,2011.

[0303] [JSC]C.Faller,“Parametric Joint-Coding of Audio Sources”,120th AES Convention,Paris,2006.

[0304] [ISS1]M.Parvaix and L.Girin:“Informed Source Separation of underdetermined instantaneous Stereo Mixtures using Source Index Embedding”,IEEE ICASSP,2010.

[0305] [ISS2]M.Parvaix,L.Girin,J.-M.Brossier:“A watermarking-based method for informed source separation of audio signals with a single sensor”,IEEE Transactions on Audio,Speech and Language Processing,2010.

[0306] [ISS3]A.Liutkus and J.Pinel and R.Badeau and L.Girin and G.Richard:“Informed source separation through spectrogram coding and data embedding”,Signal Processing Journal,2011.

[0307] [ISS4]A.Ozerov,A.Liutkus,R.Badeau,G.Richard:“Informed source separation:source coding meets source separation”,IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,2011.

[0308] [ISS5]S.Zhang and L.Girin:“An Informed Source Separation System for Speech Signals”,INTERSPEECH,2011.

[0309] [ISS6]L.Girin and J.Pinel:“Informed Audio Source Separation from Compressed Linear Stereo Mixtures”,AES 42nd International Conference:Semantic Audio,2011.

[0310] [ITU]International Telecommunication Union:“Recommendation ITU-R BS.1770-3-Algorithms to measure audio programme loudness and true-peak audio level”,Geneva,2012.

[0311] [SAOC1]J.Herre,S.Disch,J.Hilpert,O.Hellmuth:“From SAC To SAOC-Recent

Developments in Parametric Coding of Spatial Audio",22nd Regional UK AES Conference,Cambridge,UK,April 2007.

[0312] [SAOC2]J.**Engdegård,**B.Resch,C.Falch,O.Hellmuth,J.Hilpert,A.**Hölzer,** L.Terentiev,J.Breebaart,J.Koppens,E.Schuijers and W.Oomen:"Spatial Audio Object Coding(SAOC)-The Upcoming MPEG Standard on Parametric Object Based Audio Coding",124th AES Convention,Amsterdam 2008.

[0313] [SAOC]ISO/IEC,"MPEG audio technologies-Part 2:Spatial Audio Object Coding(SAOC)",ISO/IEC JTC1/SC29/WG11(MPEG)International Standard 23003-2.

[0314] [EP]EP 2146522 A1:S.Schreiner,W.Fiesel,M.Neusinger,O.Hellmuth, R.Sperschneider,"Apparatus and method for generating audio output signals using object based metadata",2010.

[0315] [DE]ISO/IEC,"MPEG audio technologies-Part 2:Spatial Audio Object Coding(SAOC)-Amendment 3,Dialogue Enhancement",ISO/IEC 23003-2:2010/DAM 3, Dialogue Enhancement.

[0316] [BRE]WO 2008/035275 A2.

[0317] [SCH]EP 2 146 522 A1.

[0318] [ENG]WO 2008/046531 A1.

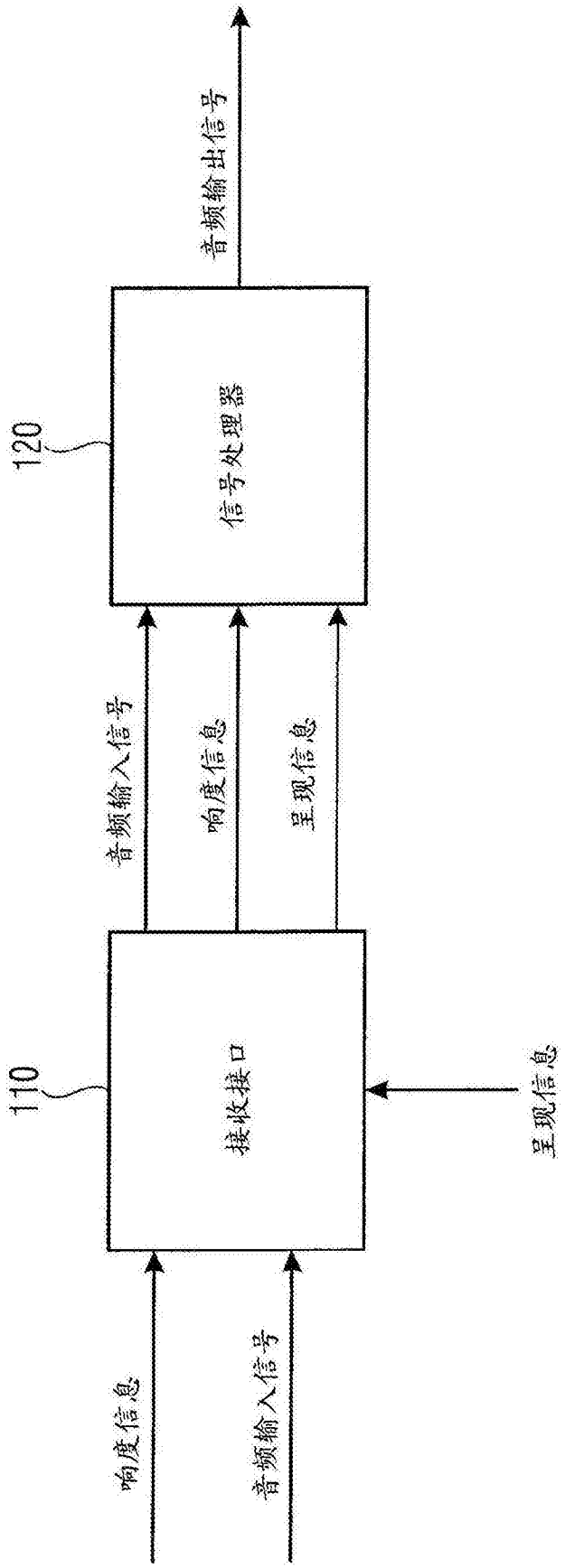


图1

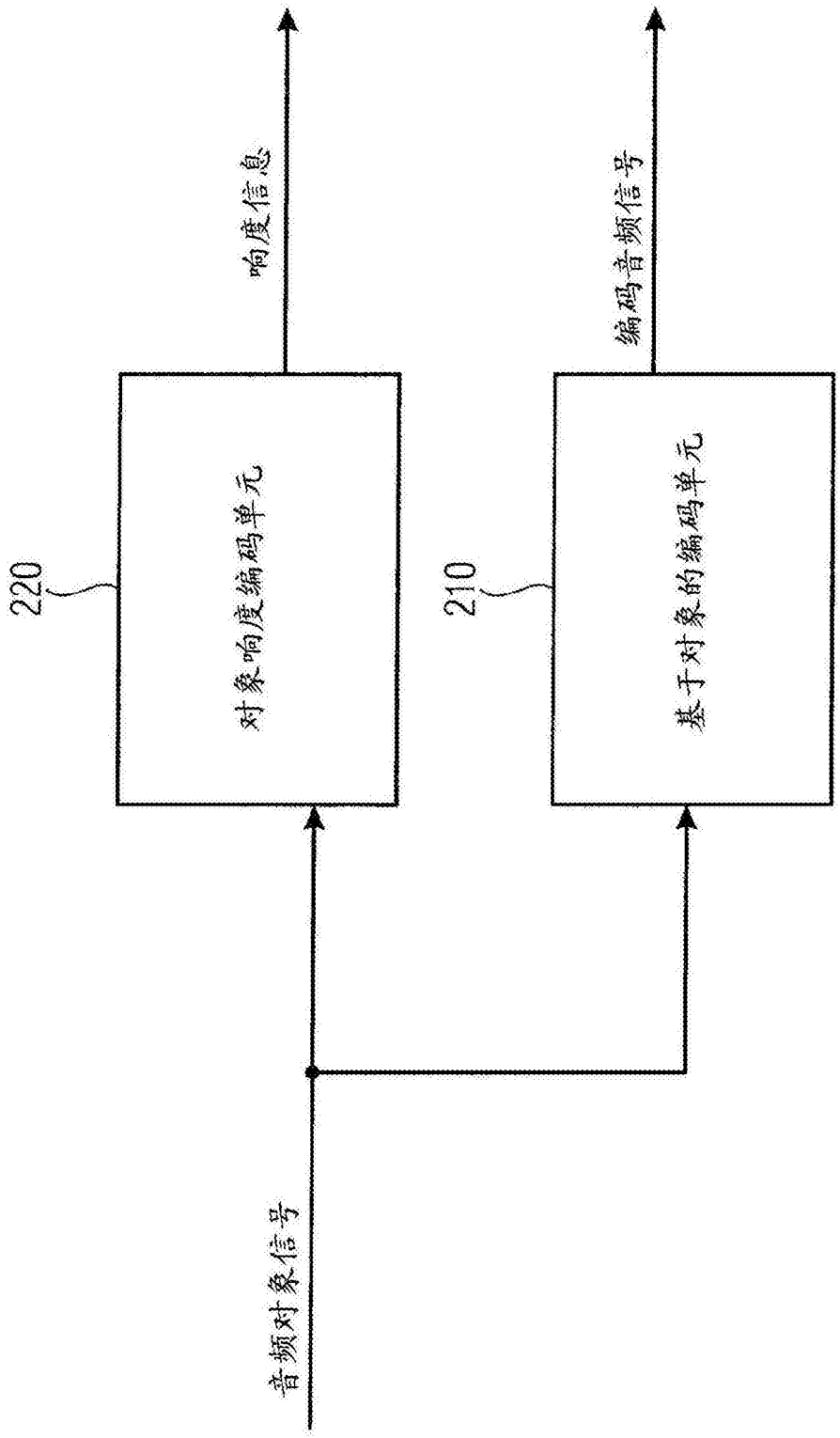


图2

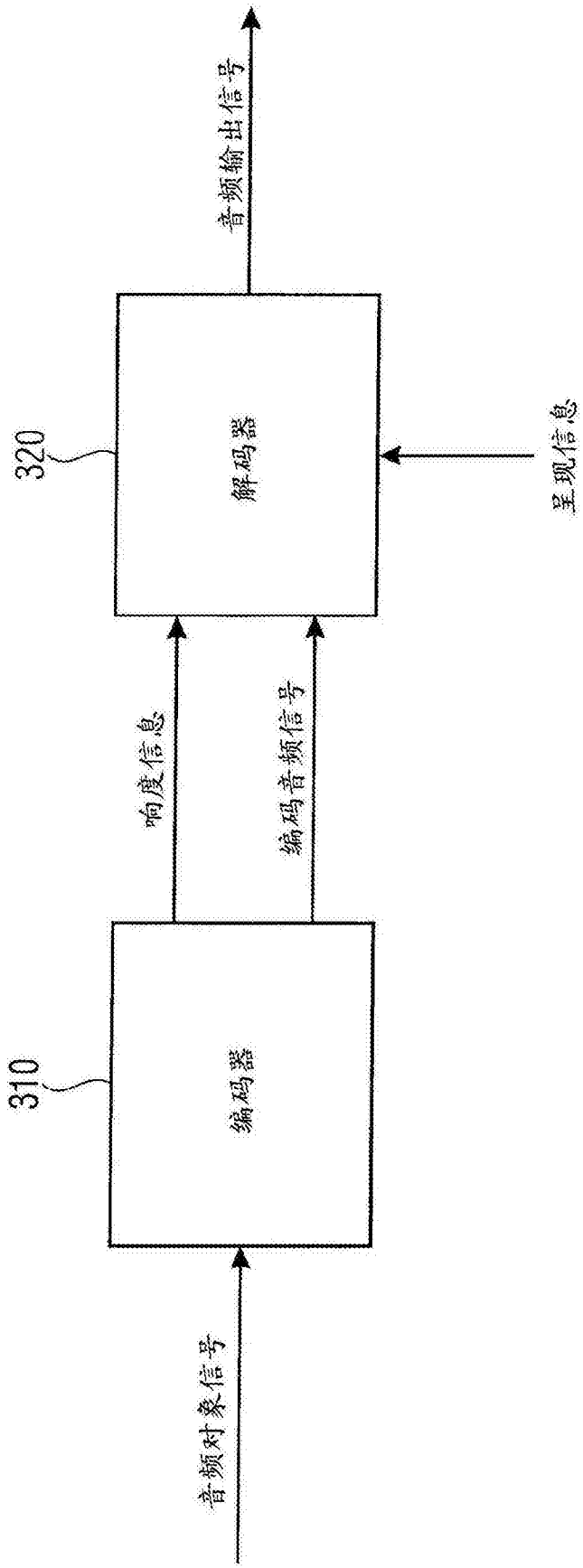

音频对象信号
310
编码器
响度信息
编码音频信号
320
解码器
音频输出信号
呈现信息

图3

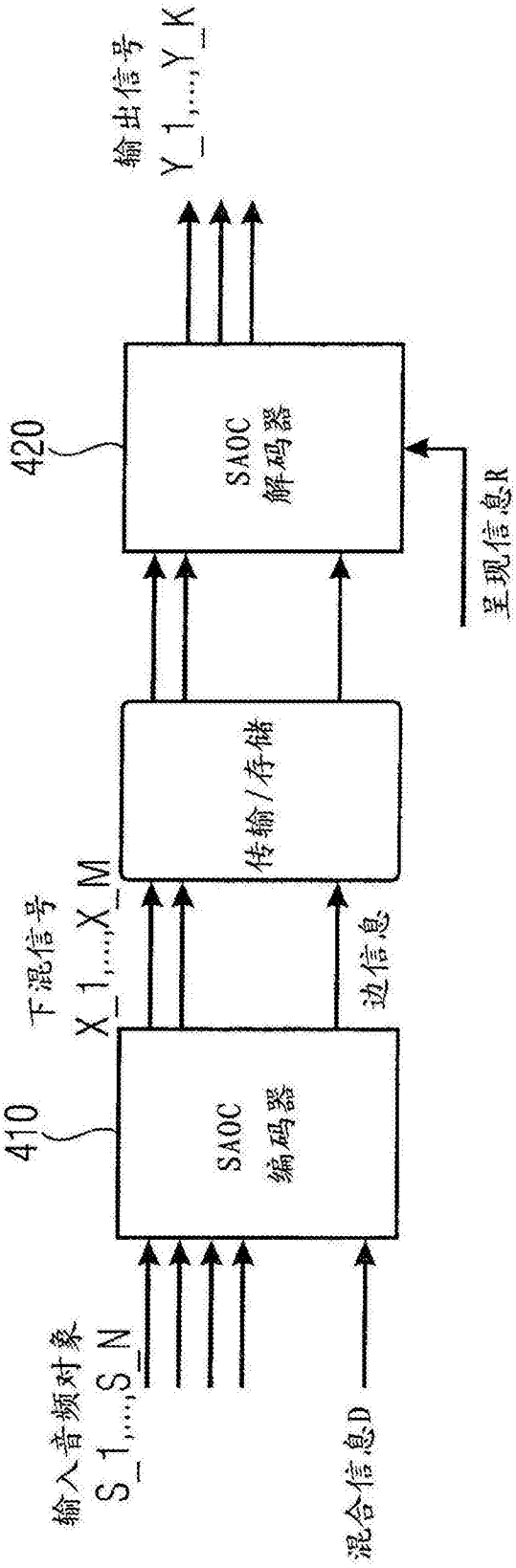

输入音频对象
S_1,...,S_N
410
SAOC
编码器
下混信号
X_1,...,X_M
边信息
混合信息D
传输/存储
420
SAOC
解码器
输出信号
Y_1,...,Y_K
呈现信息R

图4

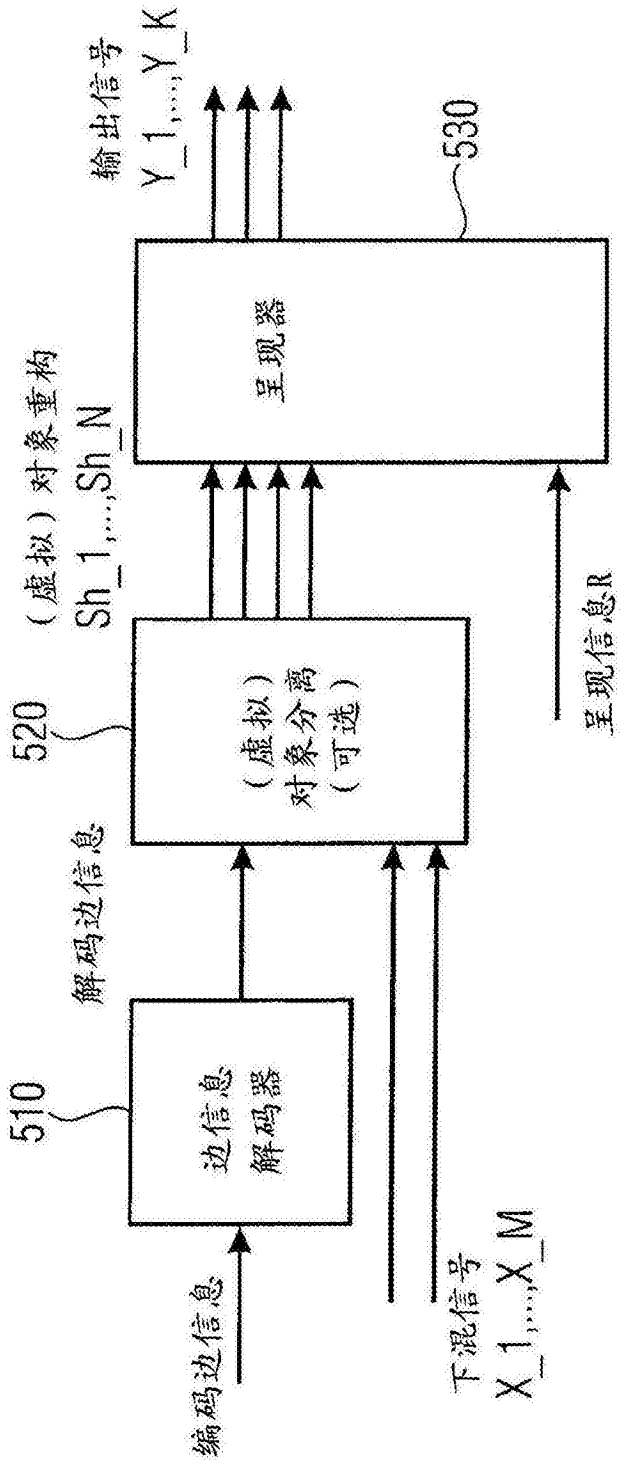

输出信号
Y_1,...,Y_K
530
呈现器
(虚拟)对象重构
Sh_1,...,Sh_N
520
(虚拟)
对象分离
(可选)
呈现信息R
解码边信息
510
边信息
解码器
编码边信息
下混信号
X_1,...,X_M

图5

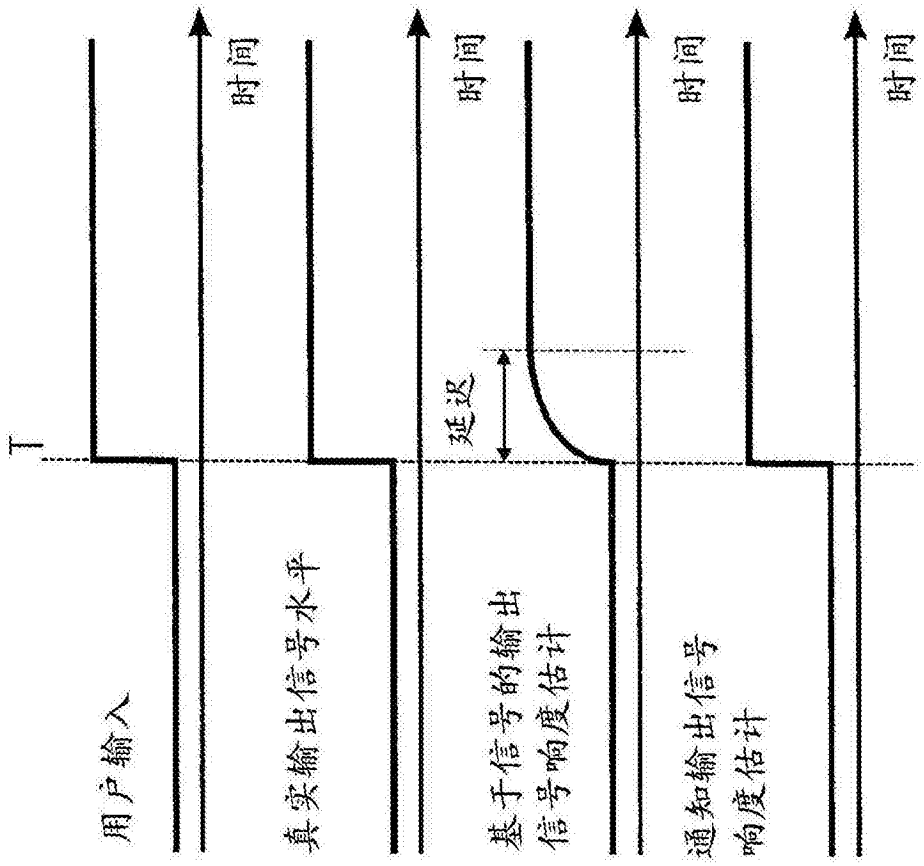


图6

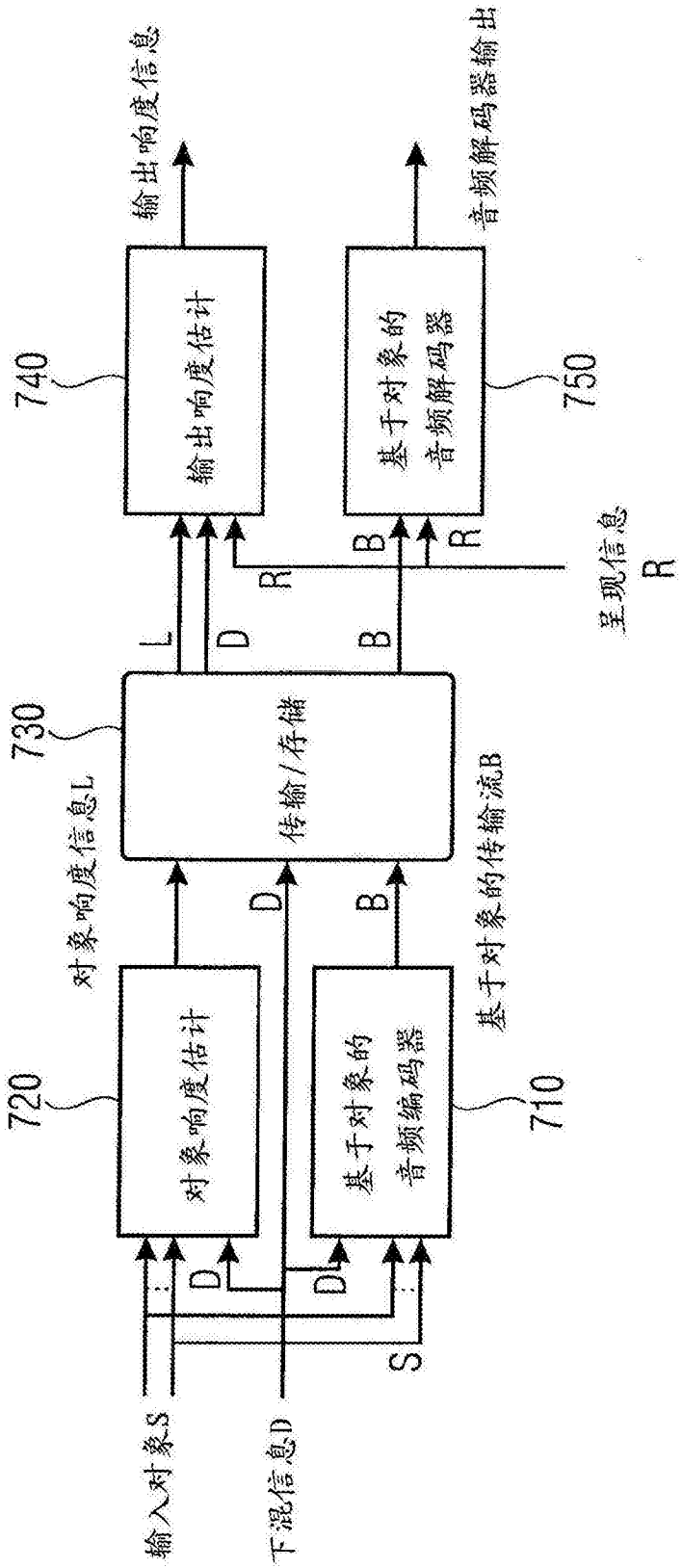

输出响度信息
音频解码器输出
740
输出响度估计
基于对象的
音频解码器
750
呈现信息
R
L
D
B
R
B
R
730
传输/存储
对象响度信息L
基于对象的传输流B
D
B
720
对象响度估计
基于对象的
音频编码器
710
D
D
S
输入对象S
下混信息D

图7

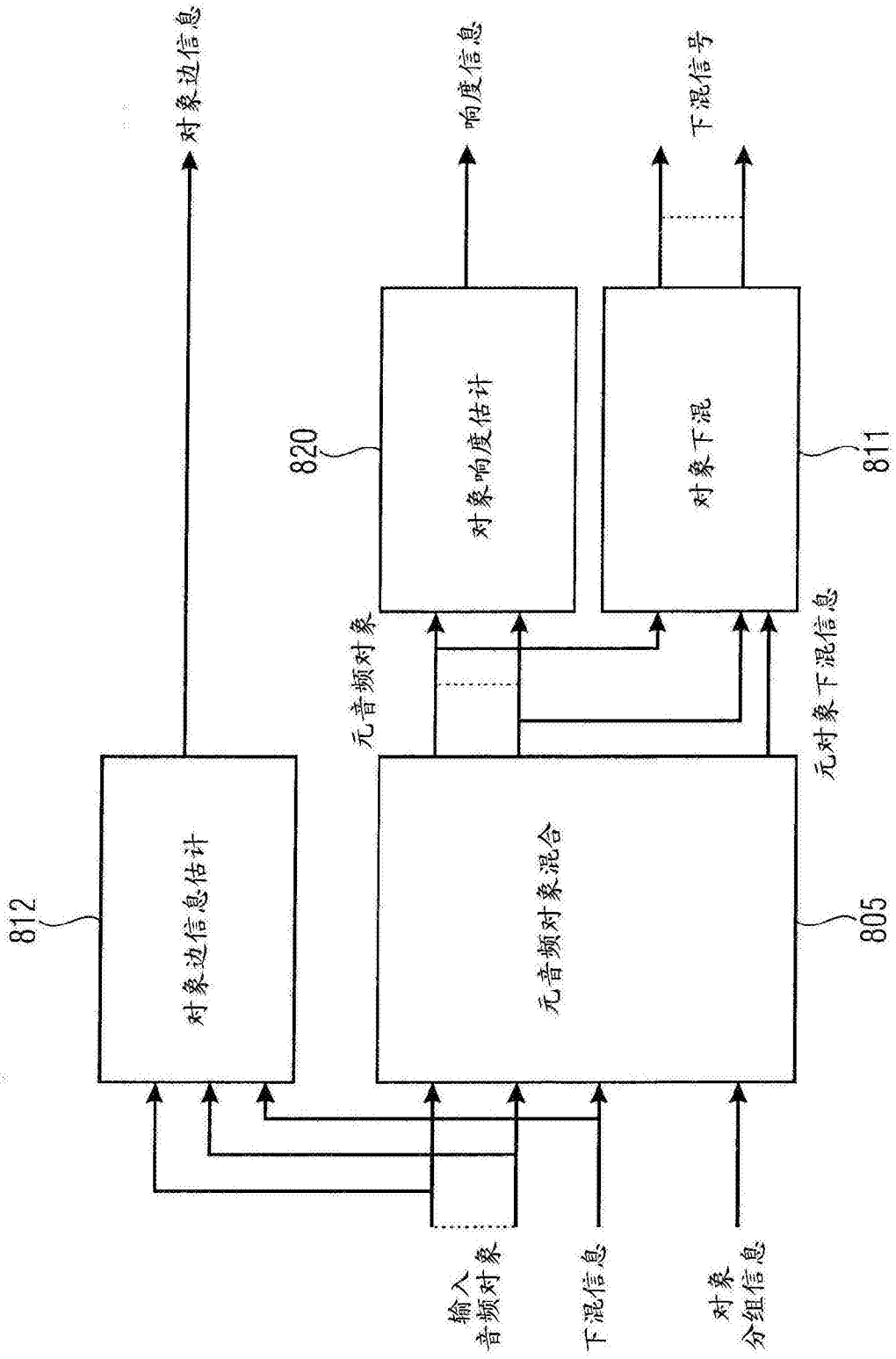

对象边信息
响度信息
下混信号
820
对象响度估计
对象下混
811
812
对象边信息估计
元音频对象
元对象下混信息
元音频对象混合
805
输入
音频对象
下混信息
对象
分组信息

图8

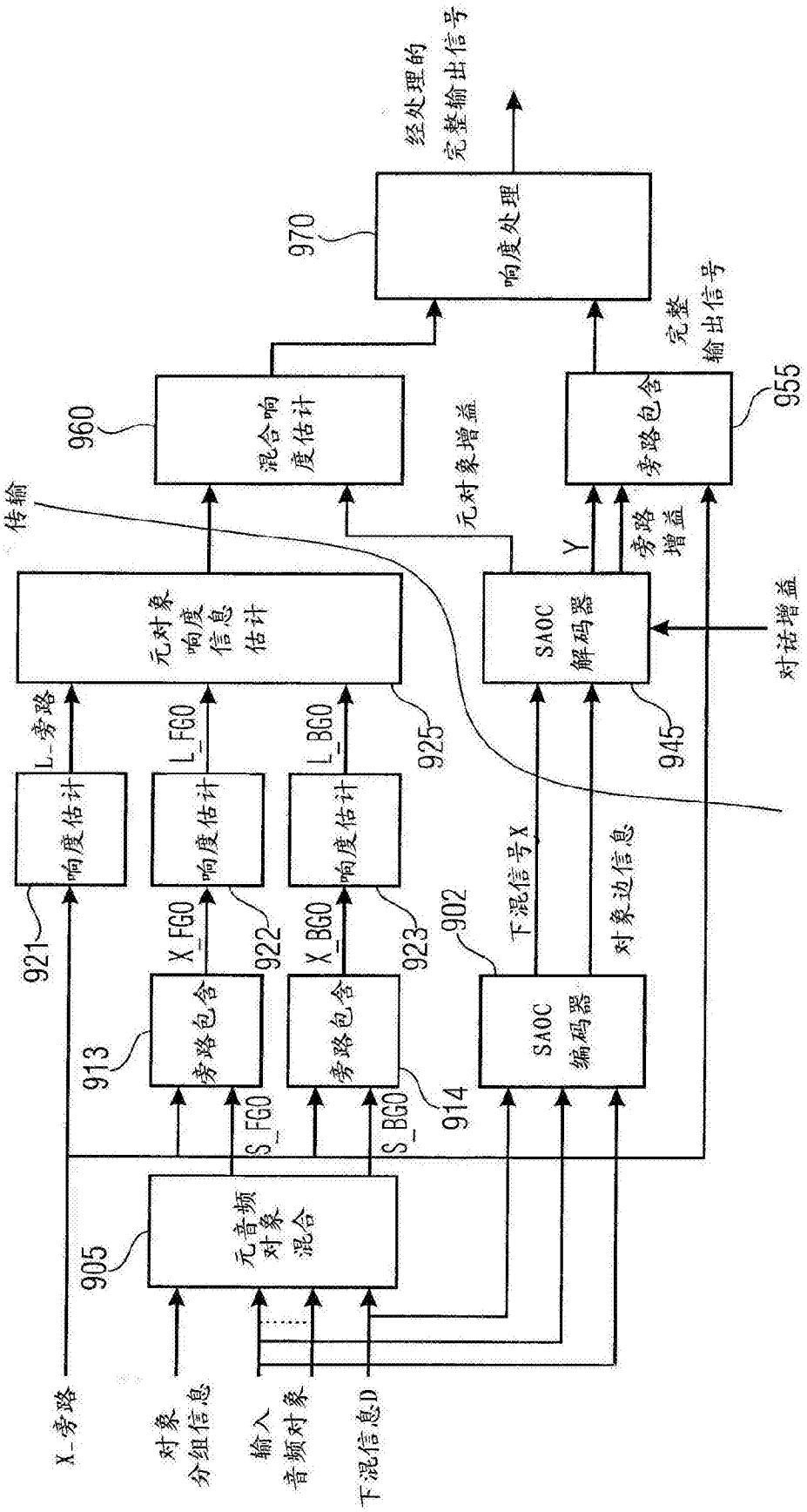

X_旁路
对象分组信息
输入音频对象
下混信息D
905
元音频对象混合
913
914
902
旁路包含
SAOC编码器
S_FGO
S_BGO
X_FGO
X_BGO
921
922
923
响度估计
L_旁路
L_FGO
L_BGO
925
元对象响度信息估计
945
SAOC解码器
下混信号X
对象边信息
对话增益
旁路增益
Y
955
旁路包含
完整输出信号
元对象增益
传输
960
混合响度估计
970
响度处理
经处理的完整输出信号

图9

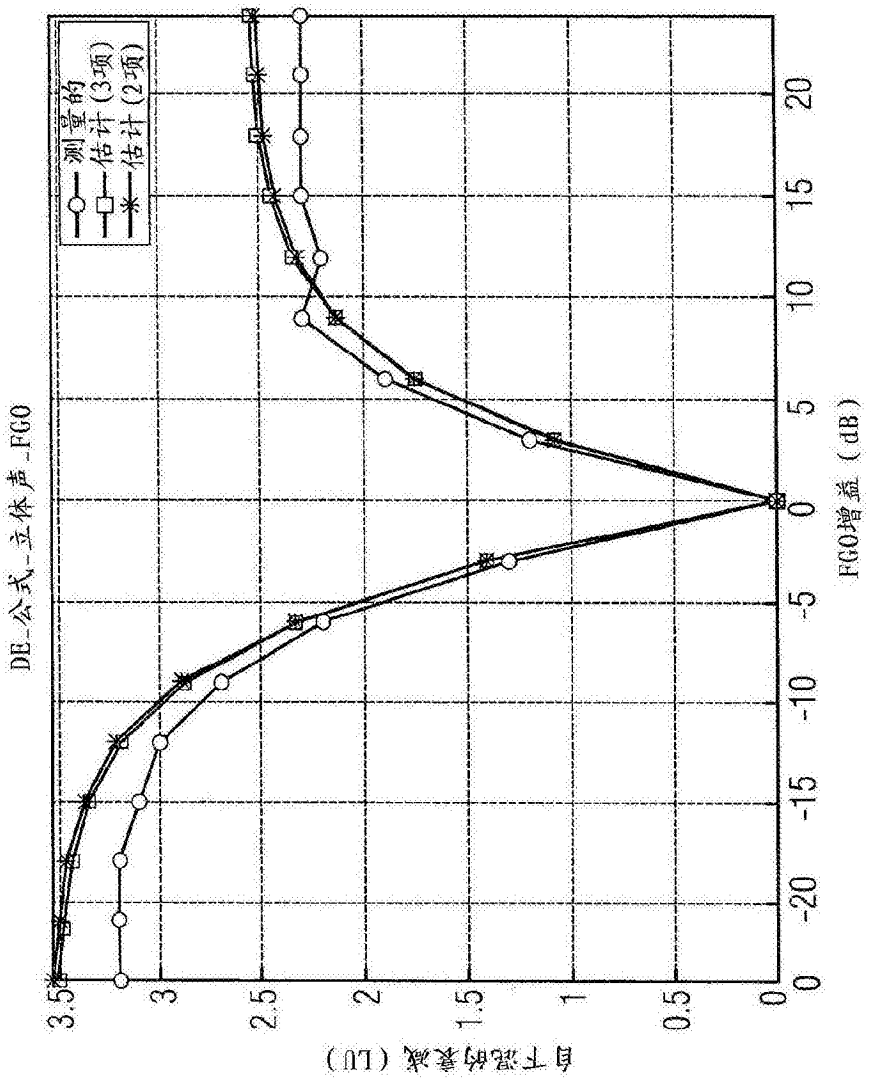

DE_公式_立体声_FGO
测量的
估计(3项)
估计(2项)
自下混的衰减(LU)
FGO增益(dB)
3.5
3
2.5
2
1.5
1
0.5
0
20
15
10
5
0
-5
-10
-15
-20

图10

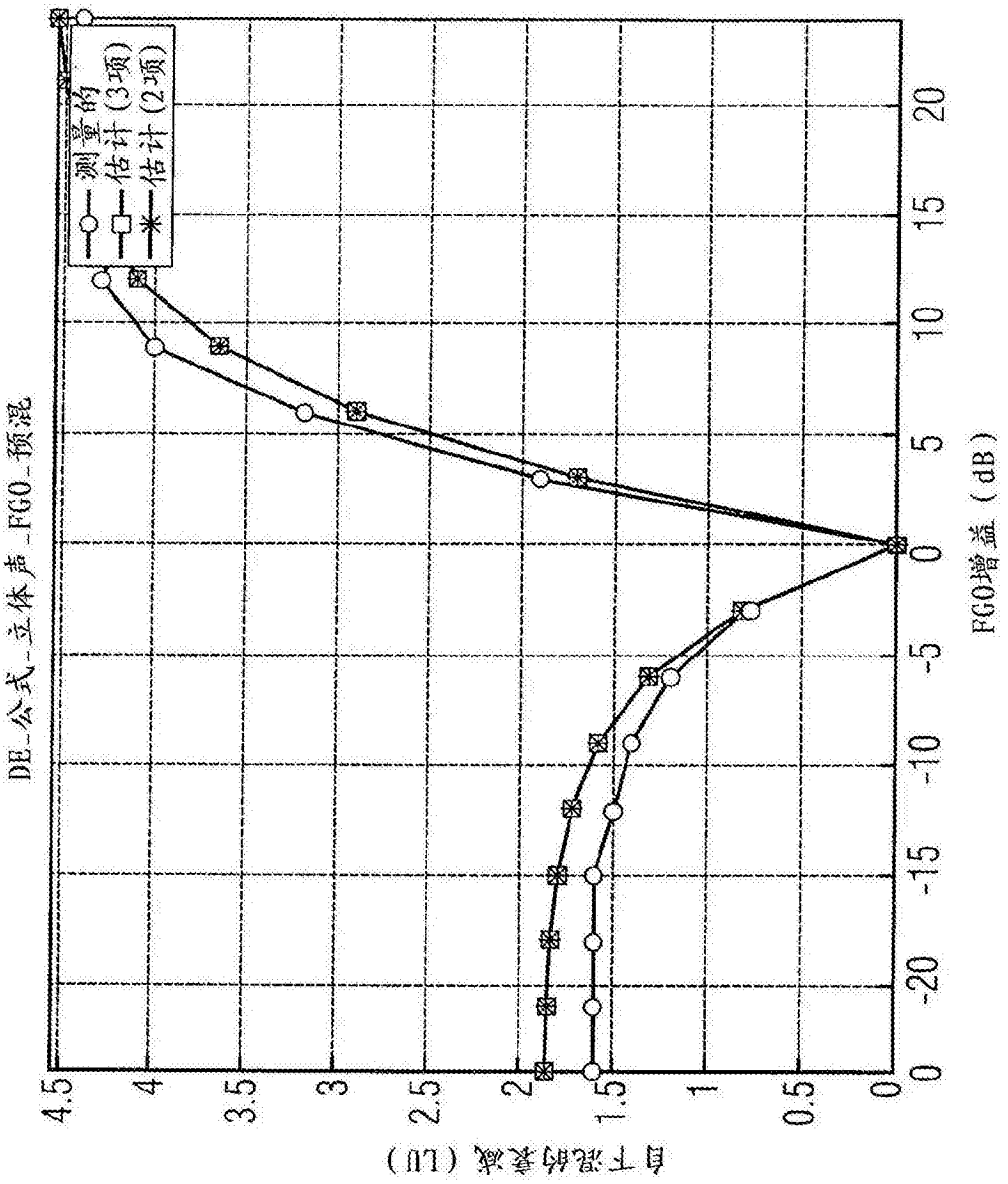

DE_公式_立体声_FGO_预混
测量的
估计(3项)
估计(2项)
自下混的衰减(LU)
FGO增益(dB)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
20
15
10
5
0
-5
-10
-15
-20

图11

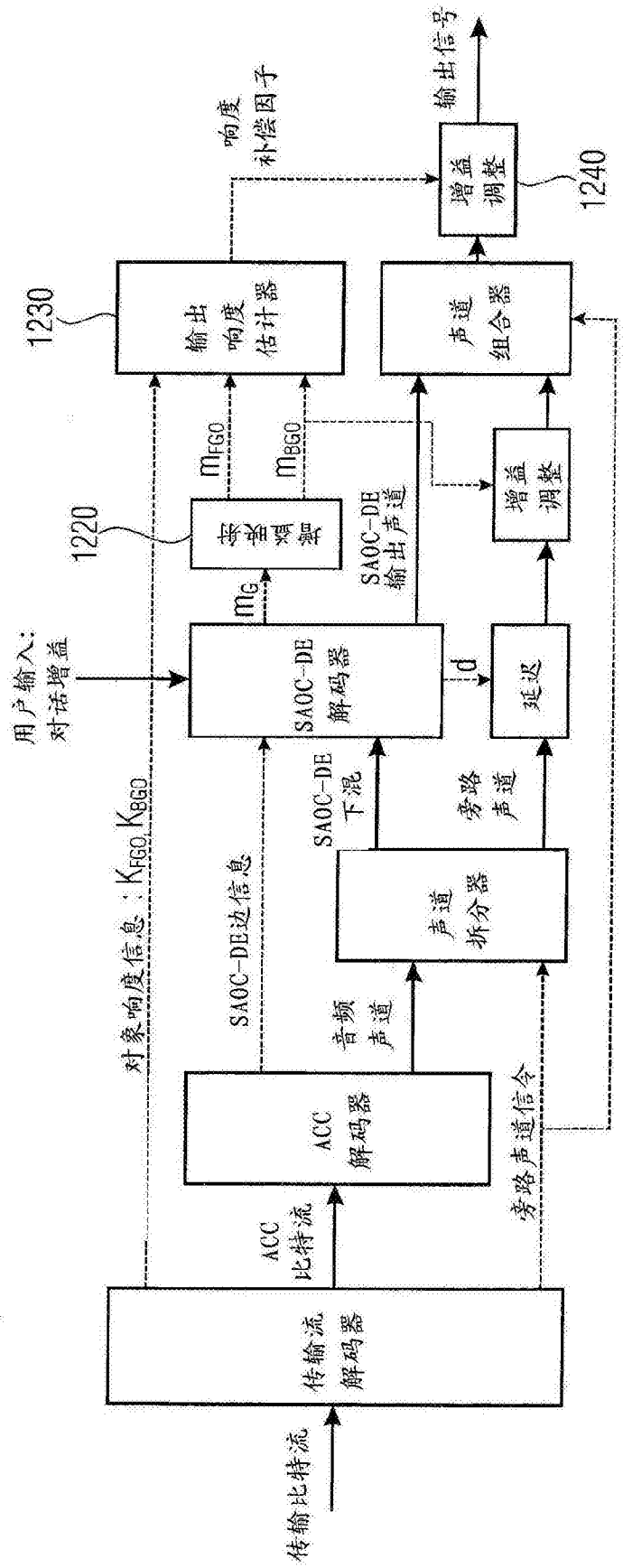

传输比特流
传输流解码器
ACC比特流
ACC解码器
音频声道
SAOC-DE边信息
对象响度信息：K_{FGO}, K_{BGO}
旁路声道信令
声道拆分器
SAOC-DE下混
旁路声道
SAOC-DE解码器
用户输入：对话增益
m_G
m_{FGO}
m_{BGO}
1220
参数调整
SAOC-DE输出声道
d
延迟
增益调整
声道组合器
1230
输出响度估计器
响度补偿因子
1240
增益调整
输出信号

图12