



(12) 发明专利

(10) 授权公告号 CN 101406073 B

(45) 授权公告日 2013. 01. 09

(21) 申请号 200680054008. X

(22) 申请日 2006. 05. 18

(30) 优先权数据

60/787, 096 2006. 03. 28 US

(85) PCT申请进入国家阶段日

2008. 09. 25

(86) PCT申请的申请数据

PCT/EP2006/004732 2006. 05. 18

(87) PCT申请的公布数据

W02007/110101 EN 2007. 10. 04

(73) 专利权人 弗劳恩霍夫应用研究促进协会

地址 德国慕尼黑

(72) 发明人 萨沙·迪施 卡斯滕·林茨迈尔

于尔根·赫勒 哈拉尔德·波普

(74) 专利代理机构 中科专利商标代理有限责任

公司 11021

代理人 吴秋明

(51) Int. Cl.

H04S 1/00 (2006. 01)

H04S 3/00 (2006. 01)

(56) 对比文件

WO 2004097794 A2, 2004. 11. 11,

审查员 雷云珊

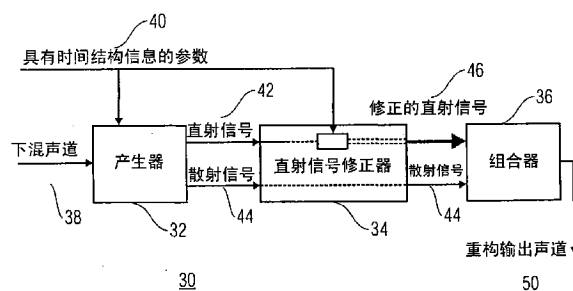
权利要求书 3 页 说明书 13 页 附图 5 页

(54) 发明名称

用于多声道音频重构中的信号成形的增强的方法

(57) 摘要

由多声道重构器使用至少一个下混声道和使用参数表示来重构的输出信号, 可以使用基于该下混声道 (38) 来产生直射信号分量 (42) 和散射信号分量 (44) 的产生器 (32) 来产生, 通过下混多个原始声道来获得所述至少一个下混声道, 所述参数表示包括关于原始声道的时间精细结构的附加信息。仅修正 (34) 直射信号分量 (42), 使得重构的输出信号的时间精细结构 (40) 适合所需的时间精细结构, 所传送的关于时间精细结构的附加信息指示了所述所需的时间精细结构。



1. 多声道重构器 (30 ;60), 使用通过对多个原始声道进行下混而获得的至少一个下混声道 (38 ;68), 并使用参数表示 (40 ;72) 来产生重构输出声道 (50 ;76), 所述参数表示 (40 ;72) 包括原始声道的时间结构信息, 所述多声道重构器 (30 ;60) 包括:

产生器 (32 ;62), 用于基于所述下混声道 (38 ;68) 来产生重构输出声道 (50 ;76) 的直射信号分量 (42 ;64) 和散射信号分量 (44 ;66);

直射信号修正器 (34 ;69), 用于使用所述包括原始声道的时间结构信息的参数表示 (40 ;72) 来修正所述直射信号分量 (42 ;64); 以及

组合器 (36 ;74), 用于对修正的直射信号分量 (46) 和所述散射信号分量 (44 ;66) 进行组合来获得所述重构输出声道 (50 ;76)。

2. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述产生器 (32 ;62) 被操作为使用所述下混声道 (38 ;68) 的滤波的和 / 或延迟的部分来产生所述散射信号分量 (44 ;66)。

3. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;69) 被操作为使用原始声道的时间结构信息, 所述原始声道的时间结构信息指示了原始声道的有限长度时间部分内所述原始声道中包含的能量。

4. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;69) 被操作为使用原始声道的时间结构信息, 所述原始声道的时间结构信息指示了原始声道的有限长度时间部分内的所述原始声道的均值幅度。

5. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述组合器 (36 ;74) 被操作为将所述修正的直射信号分量 (46) 与所述散射信号分量 (44 ;66) 相加来获得所述重构信号。

6. 如权利要求 1 所述的多声道重构器, 其中, 所述多声道重构器被操作为使用第一下混声道和第二下混声道 (38 ;68), 所述第一下混声道具有所述多个原始声道的左侧的信息, 所述第二下混声道 (38 ;68) 具有所述多个原始声道的右侧的信息, 其中, 仅使用由所述第一下混声道产生的直射和散射信号分量来组合左侧的第一重构输出声道 (50 ;76), 并且使用仅根据所述第二下混信号产生的直射和散射信号分量来组合右侧的第二重构输出声道。

7. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为修正有限长度时间部分的直射信号, 所述有限长度时间部分比所述参数表示 (40 ;72) 内的附加参数信息的帧时间部分更短, 其中, 所述产生器 (32 ;62) 使用所述附加参数信息来产生所述直射和散射信号分量。

8. 如权利要求 7 所述的多声道重构器 (30 ;60), 其中, 所述产生器 (32 ;62) 被操作为使用附加参数信息, 所述附加参数信息具有原始声道的能量信息, 所述能量信息是相对于所述多个原始声道的其他声道的能量信息。

9. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为使用原始声道的时间结构信息, 所述原始声道的时间结构信息将原始声道的时间结构与所述下混声道 (38 ;68) 的时间结构相关。

10. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述原始声道的时间结构信息与所述下混声道的时间结构信息具有能量或幅度度量。

11. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被进一步操作为获得与所述下混声道 (38 ;68) 的时间结构有关的下混时间信息。

12. 如权利要求 11 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为获得下混时间信息, 所述下混时间信息指示了有限长度时间间隔内所述下混声道 (38 ;68) 中包含的能量、或所述有限长度时间间隔内的幅度度量。

13. 如权利要求 11 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被进一步操作为使用所述下混时间信息以及所述原始声道的时间结构信息, 来获得重构的下混声道 (38 ;68) 的目标时间结构。

14. 如权利要求 11 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为获得针对所述下混声道 (38 ;68) 中高于频谱下界的频谱部分的下混时间信息。

15. 如权利要求 11 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被进一步操作为对所述下混声道 (38 ;68) 进行频谱白化, 并使用频谱白化的下混声道 (38 ;68) 来获得所述下混时间信息。

16. 如权利要求 11 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被进一步操作为获得所述下混声道 (38 ;68) 的平滑表示, 并从所述下混声道的平滑表示来获得所述下混时间信息。

17. 如权利要求 16 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为通过使用一阶低通滤波器对所述下混声道 (38 ;68) 进行滤波来获得所述平滑表示。

18. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被进一步操作为获得所述直射信号分量和所述散射信号分量的组合的时间结构信息。

19. 如权利要求 18 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为对所述直射信号和散射信号分量的组合进行频谱白化, 并使用频谱白化的直射和散射信号分量来获得所述直射信号和散射信号分量的组合的时间结构信息。

20. 如权利要求 18 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被进一步操作为获得所述直射和散射信号分量的组合的平滑表示, 并从所述直射和散射信号分量的组合的平滑表示来获得所述直射和散射信号分量的组合的时间结构信息。

21. 如权利要求 20 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为通过使用一阶低通滤波器对所述直射和散射信号分量进行滤波来获得所述直射和散射信号分量的组合的平滑表示。

22. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为使用原始声道的时间结构信息, 所述原始声道的时间结构信息表示了原始声道的有限长度时间间隔的能量或幅度与所述下混声道 (38 ;68) 的有限长度时间间隔的能量或幅度的比值。

23. 如权利要求 1 所述的多声道重构器 (30 ;60), 其中, 所述直射信号修正器 (34 ;68) 被操作为使用所述下混声道 (38 ;68) 和所述时间结构信息来获得所述重构输出声道 (50 ;76) 的目标时间结构。

24. 使用通过对多个原始声道进行下混而获得的至少一个下混声道 (38 ;68) 并使用参数表示 (40 ;72) 来产生重构输出声道 (50 ;76) 的方法, 所述参数表示 (40 ;72) 包括原始声道的时间结构信息, 所述方法包括:

基于所述下混声道 (38 ;68) 来产生重构输出声道 (50 ;76) 的直射信号分量和散射信

号分量；

使用所述包括原始声道的时间结构信息的参数表示 (40 ;72) 来修正所述直射信号分量；以及

对修正的直射信号分量 (46) 和所述散射信号分量进行组合来获得所述重构输出声道 (50 ;76)。

25. 多声道音频解码器,使用通过对多个原始声道进行下混而获得的至少一个下混声道 (38 ;68),并使用参数表示 (40 ;72) 来产生多声道信号的重构,所述参数表示 (40 ;72) 包括原始声道的时间结构信息,所述多声道音频解码器包括如权利要求 1 至 23 所述的多声道重构器。

用于多声道音频重构中的信号成形的增强的方法

技术领域

[0001] 本发明涉及多声道音频重构中的增强的信号成形的概念,尤其是一种新的包络成形方法。

背景技术

[0002] 近来,音频编码的发展实现了基于立体声(或单声道)信号和对应控制数据来重建音频信号的多声道表示。由于传送了附加控制数据,以控制基于传送的单声道或立体声声道来进行的环境声道重建,也称为上混,因此,这些方法与较早的基于矩阵的解决方案,如 Dolby Prologic 本质上不同。这样的参数多声道音频解码器基于 M 个传送的声道和附加控制数据来重构 N 个声道,其中 $N > M$ 。使用附加控制数据导致了比传送全部 N 个声道明显更低的数据速率,使得编码非常有效,而同时确保了与 M 声道设备和 N 声道设备的兼容性。M 声道可以是单声道、立体声声道或 5.1 声道表示。因此,可以将 7.2 声道原始信号下混为 5.1 声道后向兼容信号,空间音频参数使得空间音频解码器能够以较小的附加比特流开销来重现与原始 7.2 声道紧密相似的版本。

[0003] 这些参数环绕编码方法通常包括基于时间和频率变化的 ILD(互声道音级差)与 ICC(互声道相干性)参数的环绕信号参数化。例如,这些参数描述了原始多声道信号的声道对之间的功率比和相关性。在解码过程中,通过如传送的 ILD 参数所描述的那样在所有声道对之间分配接收到的下混声道的能量,来获得重建的多声道信号。然而,如 ICC 参数所描述的,虽然不同声道中的信号很不同,但多声道信号可以在所有声道之间有相等的功率分布,从而提供了非常宽的声音的收听印象,因此,通过将信号与其解相关版本向混合来获得正确的宽度(wideness)。

[0004] 将信号通过混响器,如全通滤波器,来获得信号的解相关版本,通常也称为湿(wet)或散射(diffuse)信号。解相关的一种简单形式是对信号应用指定的延迟。一般地,现有技术中已知多种不同的混响器,所使用的混响器的确切实现并不重要。

[0005] 解相关器的输出通常具有非常平坦的时间响应。因此,单位脉冲(dirac)输入信号给出了衰变噪声突发。在混合解相关信号和原始信号时,对于一些瞬时信号类型,如掌声信号,对信号执行一些后处理是很重要的,以避免感知到较大的感知室大小可能导致的附加引入的人工效应和预回声类型的人工效应。

[0006] 一般地,本发明涉及一种系统,该系统将多声道音频表示为音频下混数据(例如一或两个声道)和相关参数多声道数据的组合。在这样的方案(例如双耳线索编码(binaural cue coding))中,传送了音频下混数据流,其中可以注意到,下混的最简单形式是简单地加入多声道信号的不同信号。这样的信号(和信号)伴随有参数多声道数据流(辅助信息)。该辅助信息包括,例如以上讨论的用于描述多相对信号的原始声道的空间相互关系的一个或更多参数类型。在某种意义上,对例如具有和信号以及辅助信息的下混数据的发送/接收端而言,参数多声道方案用作预处理器/后处理器。应注意,可以附加地使用任何音频或话音编码器来对下混数据的和信号进行编码。

[0007] 随着在低带宽载波上传送多声道信号变得越来越普遍,近来已经很好地开发出了这些系统,这些系统也以“空间音频编码”、“MPEG 环绕”而为人所知。

[0008] 在这些技术的范围内,以下出版物是已知的:

[0009] [1]C.Faller and F.Baumgarte,“Efficient representation of spatial audio using perceptual parametrization,”in Proc. IEEE WASPAA, Mohonk, NY, Oct. 2001.

[0010] [2]F.Baumgarte and C.Faller,“Estimation of auditory spatial cues for binaural cue coding,”in Proc. ICASSP 2002, Orlando, FL, May 2002.

[0011] [3]C.Faller and F.Baumgarte,“Binaural cue coding:a novel and efficient representation of spatial audio,”in Proc. ICASSP 2002, Orlando, FL, May 2002.

[0012] [4]F.Baumgarte and C.Faller,“Why binaural cue coding is better than intensity stereo coding,”in Proc. AES 112th Conv., Munich, Germany, May 2002.

[0013] [5]C.Faller and F.Baumgarte,“Binaural cue coding applied to stereo and multi-channel audio compression,”in Proc. AES 112th Conv., Munich, Germany, May 2002.

[0014] [6]F.Baumgarte and C.Faller,“Design and evaluation of binaural cue coding,”in AES 113th Conv., Los Angeles, CA, Oct. 2002.

[0015] [7]C.Faller and F.Baumgarte,“Binaural cue coding applied to audio compression with flexible rendering,”in Proc. AES 113th Conv., Los Angeles, CA, Oct. 2002.

[0016] [8]J. Breebaart, J. Herre, C. Faller, J. **Rödén**, F. Myburg, S. Disch, H. Purnhagen, G. Hoto, M. Neusinger, K. **Kjörling**, W. Oomen: “MPEG Spatial Audio Coding/ MPEG Surround: Overview and Current Status”, 119th AES Convention, New York 2005, Preprint 6599

[0017] [9]J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch, K. **Kjörling**, E. Schuijers, J. Hilpert, F. Myburg, “The Reference Model Architecture for MPEG Spatial Audio Coding”, 118th AES Convention, Barcelona 2005, Preprint 6477

[0018] [10]J. Herre, C. Faller, S. Disch, C. Ertel, J. Hilpert, A. Hoelzer, K. Linzmeier, C. Spenger, P. Kroon: “ Spatial Audio Coding: Next-Generation Efficient and Compatible Coding of Multi-Channel Audio” , 117th AES Convention, San Francisco 2004, Preprint 6186

[0019] [11]J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hoelzer, C. Spenger: “ MP3 Surround: Efficient and Compatible Coding of Multi-Channel Audio” , 116th AES Convention, Berlin 2004, Preprint 6049.

[0020] 关注经由一个传送的单声道信号来传送两个声道的相关技术被称为“参数立体声”,例如,在以下出版物中更多地描述了该技术:

[0021] [12]J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, “High-Quality Parametric Spatial Audio Coding at Low Bitrates”, AES 116th Convention, Berlin, Preprint 6072, May 2004

[0022] [13]E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegard, “Low Complexity

Parametric Stereo Coding”, AES 116th Convention, Berlin, Preprint 6073, May 2004.

[0023] 如上所述,在空间音频解码器中,从直射信号部分与散射信号部分来计算多声道上混,该散射信号部分是通过解相关从直射部分中获得的。因此,一般地,散射部分具有与直射部分不同的时间包络。在这里的上下文中,术语“时间包络”描述了信号的能量或幅度随时间的变化。对于具有较宽立体声图像和同时具有瞬时包络结构的输入信号,不同的时间包络导致了上混信号中的人工效应(预回声和后回声、时间“拖尾(smearing)”)。一般地,瞬时信号是在短时间段内剧烈变化的信号。

[0024] 针对这类信号的最重要示例可能是类似掌声的信号,在直播录音中经常出现该信号。

[0025] 为了避免将具有不合适的时间包络的散射/解相关声音引入上混信号而导致的人工效应,已经提出了多种技术:

[0026] 美国申请 11/006,492 (“Diffuse Sound Shaping for BCC Schemes and The Like”)说明,可以通过对散射信号的时间包络进行成形以匹配直射信号的时间包络,来提高关键瞬时信号的感知质量。

[0027] 已经由不同的工具,如“时间包络成形”(TES)和“时间处理”(TP),将这种方法引入 MPEG 环绕技术。由于散射信号的目标时间包络是从传送的下混信号的包络得到的,因此,这种方法不需要传送附加的辅助信息。然而,由此,对所有输出声道,散射声音的时间精细结构都相同。由于从传送的下混信号中直接获得的直射信号部分也具有类似的时间包络,因此,这种方法可以在“明快性(crisp-ness)”方面提供类似掌声的信号的感知质量。然而,由于对所有声道,直射信号和散射信号都具有类似的时间包络,因此,这样的技术可以增强类似掌声的信号的主观质量,但不能改善信号中单次掌声事件的空间分布,因为仅当在瞬时信号出现时一个重构声道远比其他声道更为强烈时这才是可能的,而对于共享基本上相同的时间包络的信号是不可能的。

[0028] 美国申请 11/006,482 (“individual Channel Shaping for BCC Schemes and The Like”)描述了克服这个问题的一种选择性方法。该方法采用编码器传送的精细颗粒时间宽带辅助信息来执行直射信号与散射信号的精细时间成形。显然,该方法实现了时间精细接收,该结构对每个输出声道是单独的,因此,能够也容纳这样的信号,即对于该信号,瞬时时间仅在输出声道的子集中出现。US 60/726,389 (“Methods for Improved Temporal and Spatial Shaping of Multi-Channel Audio Signals”)描述了该方法的进一步的变型。所讨论的用于增强瞬时编码信号的感知质量的两种方法均包括对散射信号的包络的时间成形,以匹配对应的直射信号的时间包络。

[0029] 虽然之前描述的现有技术的两种方法均可以在明快性方面增强类似掌声的信号的主观质量,但是,只有后一种方法也可以改善重构信号的时间重分布。由于对于干(dry)的与散射的声音的组合进行的时间成形导致了特性失真(或在执行松时间成形时感知到各鼓掌拍击不“紧”,或在对信号应用具有非常高时间分辨率的成形时引入失真),因此,合成的掌声信号的主观质量仍不令人满意。当散射信号是直射信号的简单地延迟拷贝时,这变得很明显。接着,与直射信号混合的散射信号可能具有与直射信号不同的频谱组成。因此,即使缩放包络来匹配直射信号的包络,在重构信号中将出现并非直接源自原始信号的不同频谱成分。当缩放散射信号来匹配直射信号的包络时,在重构过程中突出(使其更响)散

射信号部分时,所引入的失真可能变得甚至更糟。

发明内容

[0030] 本发明的目的是提供一种多声道重构中增强的信号成形的概念。

[0031] 为了实现该目的,根据本发明的第一方面,提供了一种多声道重构器,使用通过对多个原始声道进行下混而获得的至少一个下混声道,并使用参数表示来产生重构输出声道,所述参数表示包括原始声道的时间结构信息,所述多声道重构器包括:产生器,用于基于所述下混声道来产生重构输出声道的直射信号分量和散射信号分量;直射信号修正器,用于使用所述包括原始声道的时间结构信息的参数表示来修正所述直射信号分量;以及组合器,用于对修正的直射信号分量和所述散射信号分量进行组合来获得所述重构输出声道。

[0032] 优选地,所述产生器被操作为使用所述下混声道的滤波的和/或延迟的部分来产生所述散射信号分量。

[0033] 优选地,所述直射信号修正器被操作为使用原始声道的时间结构信息,所述原始声道的时间结构信息指示了原始声道的有限长度时间部分内所述原始声道中包含的能量。

[0034] 优选地,所述直射信号修正器被操作为使用原始声道的时间结构信息,所述原始声道的时间结构信息指示了原始声道的有限长度时间部分内的所述原始声道的均值幅度。

[0035] 优选地,所述组合器被操作为将所述修正的直射信号分量与所述散射信号分量相加来获得所述重构信号。

[0036] 优选地,所述多声道重构器被操作为使用第一下混声道和第二下混声道,所述第一下混声道具有所述多个原始声道的左侧的信息,所述第二下混声道具有所述多个原始声道的右侧的信息,其中,仅使用由所述第一下混声道产生的直射和散射信号分量来组合左侧的第一重构输出声道,并且使用仅根据所述第二下混信号产生的直射和散射信号分量来组合右侧的第二重构输出声道。

[0037] 优选地,所述直射信号修正器被操作为修正有限长度时间部分的直射信号,所述有限长度时间部分比所述参数表示内的附加参数信息的帧时间部分更短,其中,所述产生器使用所述附加参数信息来产生所述直射和散射信号分量。

[0038] 优选地,所述产生器被操作为使用附加参数信息,所述附加参数信息具有原始声道的能量信息,所述能量信息是相对于所述多个原始声道的其他声道的能量信息。

[0039] 优选地,所述直射信号修正器被操作为使用原始声道的时间结构信息,所述原始声道的时间结构信息将原始声道的时间结构与所述下混声道的时间结构相关。

[0040] 优选地,所述原始声道的时间结构信息与所述下混声道的时间结构信息具有能量或幅度度量。

[0041] 优选地,所述直射信号修正器被进一步操作为获得与所述下混声道的时间结构有关的下混时间信息。

[0042] 优选地,所述直射信号修正器被操作为获得下混时间信息,所述下混时间信息指示了有限长度时间间隔内所述下混声道中包含的能量、或所述有限长度时间间隔内的幅度度量。

[0043] 优选地,所述直射信号修正器被进一步操作为使用所述下混时间信息以及所述原

始声道的时间结构信息,来获得重构的下混声道的目标时间结构。

[0044] 优选地,所述直射信号修正器被操作为获得针对所述下混声道中高于频谱下界的频谱部分的下混时间信息。

[0045] 优选地,所述直射信号修正器被进一步操作为对所述下混声道进行频谱白化,并使用频谱白化的下混声道来获得所述下混时间信息。

[0046] 优选地,所述直射信号修正器被进一步操作为获得所述下混声道的平滑表示,并从所述下混声道的平滑表示来获得所述下混时间信息。

[0047] 优选地,所述直射信号修正器被操作为通过使用一阶低通滤波器对所述下混声道进行滤波来获得所述平滑表示。

[0048] 优选地,所述直射信号修正器被进一步操作为获得所述直射信号分量和所述散射信号分量的组合的时间结构信息。

[0049] 优选地,所述直射信号修正器被操作为对所述直射信号和散射信号分量的组合进行频谱白化,并使用频谱白化的直射和散射信号分量来获得所述直射信号和散射信号分量的组合的时间结构信息。

[0050] 优选地,所述直射信号修正器被进一步操作为获得所述直射和散射信号分量的组合的平滑表示,并从所述直射和散射信号分量的组合的平滑表示来获得所述直射和散射信号分量的组合的时间结构信息。

[0051] 优选地,所述直射信号修正器被操作为通过使用一阶低通滤波器对所述直射和散射信号分量进行滤波来获得所述直射和散射信号分量的组合的平滑表示。

[0052] 优选地,所述直射信号修正器被操作为使用原始声道的时间结构信息,所述原始声道的时间结构信息表示了原始声道的有限长度时间间隔的能量或幅度与所述下混声道的有限长度时间间隔的能量或幅度的比值。

[0053] 优选地,所述直射信号修正器被操作为使用所述下混声道和所述时间结构信息来获得所述重构输出声道的目标时间结构。

[0054] 根据本发明的第二方面,提供了一种使用通过对多个原始声道进行下混而获得的至少一个下混声道并使用参数表示来产生重构输出声道的方法,所述参数表示包括原始声道的时间结构信息,所述方法包括:基于所述下混声道来产生重构输出声道的直射信号分量和散射信号分量;使用所述包括原始声道的时间结构信息的参数表示来修正所述直射信号分量;以及对修正的直射信号分量和所述散射信号分量进行组合来获得所述重构输出声道。

[0055] 根据本发明的第三方面,提供了一种多声道音频解码器,使用通过对多个原始声道进行下混而获得的至少一个下混声道,并使用参数表示来产生多声道信号的重构,所述参数表示包括原始声道的时间结构信息,所述多声道音频解码器包括根据本发明第一方面的多声道重构器。

[0056] 本发明基于以下发现,即对于由多声道重构器使用至少一个下混声道和使用参数表示来重构的输出信号,在使用基于该下混声道来产生直射信号分量和散射信号分量的产生器时,能够以高质量有效地重构该输出信号,通过下混多个原始声道来获得所述至少一个下混声道,所述参数表示包括关于原始声道的时间(精细)结构的附加信息。如果仅修正直射信号分量,使得重构的输出信号的时间精细结构适合所需的时间精细结构,则可以实

质上增强质量,所传送的关于时间精细结构的附加信息指示了所述所需的时间精细结构。

[0057] 换言之,对从下混信号中直接获得的直射信号部分进行缩放,几乎不会在出现瞬时信号的瞬间引入附加人工效应。在现有技术中,当对湿信号进行缩放来匹配所需包络时,很可能出现这样的情况,即与直射信号混合的突出的散射信号掩蔽了重构声道中的原始瞬时信号,以下将更多地描述这一点。

[0058] 本发明通过仅对直射信号分量进行缩放,克服了这个问题,从而以在辅助信息内传送附加参数来描述时间包络为代价,不可能产生附加人工效应。

[0059] 根据本发明的一个实施例,使用具有白化频谱的直射和散射信号的表示来获得包络缩放参数,即所述表示中,信号的不同频谱部分具有几乎相同的能量。使用白化频谱的优点有两方面。一方面,使用白化频谱作为计算用于缩放直射信号的缩放因子的基础,允许每时隙仅传送一个参数,所述参数包括时间结构信息。由于在多声道音频编码中,通常在许多频带内处理信号,因此,这一特征有助于减少附加需要的辅助信息的数量,因此增加了附加参数传送的比特率。典型地,每个时帧和参数频带仅传送一次其他参数,如 ICLD 和 ICC。由于参数频带的数量可能高于 20,主要优点是每个声道仅需传送单个参数。一般地,在多声道编码中,在帧结构中,即在具有若干采样值的实体中处理信号,例如每帧 1024 个采样值。此外,如上所述,在处理之前,将信号分为若干频谱部分,使得最终,每帧和信号的每频谱部分典型地传送一个 ICC 和 ICLD 参数。

[0060] 仅使用一个参数的第二个优点是从物理上推动的,由于所讨论的瞬时信号天然地具有较宽的频谱。因此,为了正确解决对单个声道内的瞬时信号的能量,最合适的是使用白化频谱来计算能量缩放因子。

[0061] 在本发明的另一个实施例中,在出现附加残留信号时,仅对高于特定频谱限制的信号频谱部分应用本发明的修正直射信号分量的概念。这是因为残留信号与下混信号一起允许了原始声道的高质量重现。

[0062] 概况而言,设计本发明的概念,相对于现有技术的方法,提供了增强的时间和空间质量,避免了与这些现有技术相关的问题。因此,传送辅助信息来描述各声道的精细时间包络结构,从而允许在解码器端对上混声道信号进行精细时间 / 空间成形。本文中描述的本发明方法基于以下发现 / 考虑:

[0063] ●类似掌声的信号可以被视为由单一、独特的近处拍掌和源自非常密集的远处拍掌的类似噪声的环境信号 (ambience) 组成。

[0064] ●在空间音频解码器中,在时间包络方面,近处拍掌的最佳近似是直射信号。因此,本发明方法仅处理直射信号。

[0065] ●由于散射信号主要表示了信号的环境部分,因此对精细时间分辨率的任何处理可能引入失真和调制人工效应 (即使这样的技术可能实现掌声明快性的一定的主观增强)。出于这些考虑,本发明的处理不涉及散射信号 (不经过精细时间成形)。

[0066] ●然而,散射信号对上混信号的能量平衡有贡献。本发明方法通过由传送的信息计算要单独对直射信号部分应用的修正宽带缩放因子来解决这一点。选择该修正因子,使得在特定界限内,给定的时间间隔中的总能量相同,如同将原始因子应用到该间隔中的信号的直射和散射部分一样。

[0067] ●使用本发明方法,如果将空间线索的频谱分辨率选择为较低——例如“全带

宽”——以确保保持信号中包含的瞬时信号的频谱完整性,则能够获得最佳的主观音频质量。在这种情况下,由于安全地用频谱分辨率交换了时间分辨率,因此所提出的方法不必需增加平均空间辅助信息的比特率。

[0068] 通过仅随时间放大或衰减(damp) (“成形”)信号的“干”的部分,实现了主观质量的提高,从而:

[0069] ●通过加强瞬时信号位置处的直射信号部分来增强瞬时信号质量,同时避免源自具有不合适的时间包络的散射信号的附加失真

[0070] ●通过在瞬时效件的空间起源处相对于散射部分来突出直射部分,并在远处全景位置处相对于散射部分来衰减该直射部分,改善了空间定位。

附图说明

[0071] 图 1 示出了多声道编码器和对应解码器的框;

[0072] 图 1b 示出了使用解相关信号的信号重构的示意略图;

[0073] 图 2 示出了本发明的多声道重构器的示例;

[0074] 图 3 示出了本发明的多声道重构器的另一个示例;

[0075] 图 4 示出了多声道解码方案中用于标识不同参数频带的参数频带表示的示例;

[0076] 图 5 示出了本发明的多声道解码器的示例;以及

[0077] 图 6 示出了一个框图,详细示出了本发明的重构输出声道的方法的示例。

具体实施方式

[0078] 图 1 根据现有技术的多声道音频数据的编码的示例,以更清楚地说明本发明的概念所解决的问题。

[0079] 一般地,在编码器端,将原始多声道信号 10 输入多声道编码器 12,多声道编码器 12 获得辅助信息 14,辅助信息 14 指示了该原始多声道信号的各声道相对于彼此的空间分布。除了产生辅助信息 14 外,多声道编码器 12 产生一个或更多和信号 16,和信号 16 是从该原始多声道信号下混得到的。广泛使用的著名配置是所称的 5-1-5 和 5-2-5 配置。在 5-1-5 配置中,编码器从 5 个输入声道产生一个单声道和信号 16,因此,对应的解码器 18 必须产生重构的多声道信号 20 的 5 个重构声道。在 5-2-5 配置中,编码器从 5 个输入声道产生两个下混声道,该下混声道中的第一声道典型地持有左侧或右侧的信息,而该下混声道中的第二声道持有另一侧的信息。

[0080] 例如在图 1 中所示的,描述原始声道的空间分布的样本参数是之前介绍的参数 ICLD 和 ICC。

[0081] 可以注意到,在获得辅助信息 14 的解析中,典型地,在表示原始声道的指定频率间隔的子带域中处理多声道信号 10 的原始声道的样本。单个频率间隔由 K 表示。在一些应用中,在处理之前,可以由混合滤波器组对输入声道进行滤波,即可以进一步对参数频带 K 进行子分割,每个子分割由 k 表示。

[0082] 此外,在每个单个参数频带内,以逐帧的方式来对描述原始声道的样本值进行处理,即若干连续的样本形成有限持续时间的帧。典型地,上述 BCC 参数描述了完整的帧。

[0083] 以某种方式与本发明的相关的并在现有技术中已知的参数是 ICLD 参数,该参数

相对于原始多声道或信号的其他声道的对应帧来描述声道的单个帧内包含的能量。

[0084] 通常,在解相关信号的帮助下实现附加声道的产生,产生该附加声道以仅从一个传送的和信号重构多声道信号,该解相关信号是使用解相关器或混响器从该和信号中获得的。对于一种典型的应用,离散采样频率可以是 44.100kHz,因此,单个样本表示原始声道的约 0.02ms 的有限长度的间隔。可以注意到,使用滤波器组,将信号分为许多信号部分,每个信号部分表示原始信号的有限频率间隔。为了补偿描述声道的参数的可能的增加,通常降低时间分辨率,使得滤波器组域内单个样本所描述的有限长度的时间部分可以增加至 0.5ms。典型的帧长度可以在 10 至 15ms 之间变化。

[0085] 在不限本发明的范围的情况下,可以利用不同的滤波器结构和 / 或延迟或其组合来获得解相关信号。可以进一步注意到,不必需使用整个频谱来获得解相关信号。例如,可以仅使用高于和信号(下混信号)的频谱下界(指定值 K)的频谱部分,使用延迟和 / 或滤波器来获得解相关信号。因此,一般地,解相关信号描述了从下混信号(下混声道)获得的信号,使得例如,在使用该解相关信号和下混声道来获得相关系数时,该相关系数明显偏离 1(unity)0.2。

[0086] 图 1b 给出了多声道音频编码过程中的下混和重构过程的极其简化的示例,以解释本发明的在重构多声道信号的过程中仅缩放直射信号分量的概念的重大益处。对于以下的描述,假设了一些简化。第一个简化是,左和右声道的下混是该声道内的幅度的简单相加。第二个强烈的简化是,假设相关性是整个信号的简单延迟。

[0087] 在这些假设下,应对左声道 21a 和右声道 21b 的帧进行编码。如在所示的窗的 x 轴所示,在多声道编码中,典型地,对以固定采样频率来采样的样本值执行处理。为了便于解释,在以下的简要概述中忽略这一点。

[0088] 如上所述,在编码器端,将左和右声道组合(下混)为要向解码器传送的下混声道 22。在解码器端,从传送的下混信号中获得解相关信号 23,在本示例中,下混声道 22 是左声道 21a 和右声道 21b 之和。如已经解释过的,接着执行从信号帧来重构左声道,从下混声道 22 和解相关信号 23 来获得该信号帧。

[0089] 可以注意到,如 ICLD 参数所示,在组合之前,每个单个帧经过全局缩放,该 ICLD 参数将单个声道的各帧内的能量与多声道信号的其他声道的对应帧的能量相关。

[0090] 如在本实施例中所假设的,在左声道 21a 和右声道 21b 内包含相等的能量,在组合之前,使用约 0.5 的因子来缩放传送的下混声道 22 和解相关信号 23。这就是说,当上混与下混同样简单时,即对两个信号求和时,原始左声道 21a 的重构是经缩放的下混声道 24a 与经缩放的解相关信号 24b 之和。

[0091] 由于对传送信号求和以及由 ICLD 参数引起的缩放,瞬时信号的信号对背景的比值可能减小约为 2 的因子。此外,当将两个信号简单相加时,在经缩放的解相关信号 24b 中,在延迟的瞬时结构的位置可能引入附加的回声类型的人工效应。

[0092] 如图 1b 所示,现有技术试图通过对经缩放的解相关信号 24b 进行缩放以使其匹配经缩放的传送声道 24a 的包络来克服该回声问题,如帧 24b 中的虚线所示。由于该缩放,左声道 21a 中的原始瞬时信号位置的幅度可能增大。然而,帧 24b 中缩放位置处的解相关信号的频谱组成与原始瞬时信号的频谱组成不同。因此,即使可以很好地重现信号的总体强度,也会向信号引入可听见的人工效应。

[0093] 本发明的重大优点在于,本发明仅对重构信号的直射信号分量进行缩放。由于该声道确实具有与原始瞬时信号相对应的信号分量,该原始瞬时信号具有正确的频谱组成和正确的定时,因此,仅对下混声道进行缩放将产生以高精度重构原始瞬件事件的重构信号。这是由于,只有该缩放所突出的信号部分才具有与原始瞬时信号相同的频谱组成。

[0094] 图 2 示出了本发明的多声道重构器的示例的框图,以详细描述本发明概念的原理。

[0095] 图 2 示出了多声道重构器 30,具有产生器 32、直射信号修正器和组合器 36。产生器 32 接收从多个原始声道下混而来的下混声道 38,以及包括原始声道的时间结构信息的参数表示 40。

[0096] 该产生器基于该下混声道来产生直射信号分量 42 和散射信号分量 44。

[0097] 直射信号修正器 34 接收直射信号分量 42 以及散射信号分量 44,此外还接收具有原始声道的时间结构信息的参数表示 40。根据本发明,直射信号修正器 34 使用该参数表示,仅修正直射信号分量 42 来获得修正的直射信号分量 46。

[0098] 修正的直射信号分量 46 和直射信号修正器 34 未改变的散射信号分量 44 输入组合器 36,组合器 36 对修正的直射信号分量 46 和散射信号分量 44 进行组合来获得重构的输出声道 50。

[0099] 通过仅对从传送的下混声道 38 获得的直射信号分量 42 进行修正而不进行混响(解相关),可以重构重构输出声道的时间包络,该时间包络与其下原始声道的时间包络紧密匹配,而不像现有技术中那样引入附加人工效应和可听见的失真。

[0100] 如在图 3 的描述中将要更详细地描述的,本发明的包络成形恢复了合成输出信号的宽带包络。其包括修正上混过程,接着是每个输出声道的直射信号部分的包络平坦化和重新成形。为了重新成形,使用参数表示的比特流中包含的参数宽带包络辅助信息。根据本发明的一个实施例,该辅助信息包括将传送的下混信号的包络与原始输出声道信号的包络相关的比值(envRatio)。在解码器中,从这些比值获得增益因子,对给定输出声道的帧中的每个时隙上的直射信号应用该增益因子。根据本发明的概念,不改变每个声道的散射声音部分。

[0101] 图 3 的框图中所示的本发明的优选实施例是多声道重构器 60,修改该多声道重构器 60 来适合 MPEG 空间解码器的解码器信号流。

[0102] 多声道重构器 60 包括产生器 62,如在 MPEG 编码中所使用的,产生器 62 使用通过对多个原始声道进行下混而获得的下混声道 68 以及具有多声道信号的原始声道的空间特性信息的参数表示 70,来产生直射信号分量 64 以及散射信号分量 66。多声道重构器 60 还包括直射信号修正器 68,接收直射信号分量 64、散射信号分量 66、下混信号 69 和附加的包络辅助信息 72 作为输入。

[0103] 该直射信号修正器在其修正器输出 73 提供了修正的直射信号分量,以下将更详细描述进行修正的方式。

[0104] 组合器 74 接收修正的直射信号分量和散射信号分量,以获得重构的输出声道 76。

[0105] 如图所示,在现有的多声道环境中可以容易地实现本发明。在这样的编码方案中,可以根据在参数比特流中附加地传送的一些参数来开启或关闭本发明概念的总体应用。例如,可以引入附加的标志 bsTempShapeEnable,当将该标志设为 1 时,指示需要使用本发明

的概念。

[0106] 此外,可以引入附加的标志,具体地指定需要基于逐声道地应用本发明的概念。因此,可以使用附加的标志,例如成为 bsEnvShapeChannel。可以对每个单个声道提供该标志,将该标志设为 1 时,可以指示使用本发明的概念。

[0107] 此外还可以注意到,为了便于表示,在图 3 中仅描述了两个声道的配置。当然,本发明不应局限于仅有两个声道的配置。此外,可以使用任何声道配置来与本发明的概念相结合。例如,可以使用 5 或 7 输入声道来与本发明的增强的包络成形相结合。

[0108] 如图 3 所示,当在 MPEG 编码方案中应用本发明的概念,通过设定 bsTempShapeEnable 等于 1,以信号来指示应用本发明的概念时,产生器 62 根据以下方程,使用混合子带域中的修正的后混合 (post-mixing) 来分离地合成直射和散射信号分量:

$$[0109] \quad \mathbf{y}_{direct}^{n,k} = \mathbf{M}^{n,k} \mathbf{w}_{direct}^{n,k} \quad 0 \leq k < K$$

$$[0110] \quad \mathbf{y}_{diffuse}^{n,k} = \mathbf{M}^{n,k} \mathbf{w}_{diffuse}^{n,k} \quad 0 \leq k < K$$

[0111] 这里以及在以下段落中,向量 $\mathbf{w}_{m,k}$ 描述了子带域的第 k 个子带的 n 个混合子带参数的向量。如上述等式所示,在上混中,分离地获得直射和散射信号 $\mathbf{y}$ 。直射输出持有直射信号分量和残留信号,该残留信号在 MPEG 编码中可能附加地出现的一种信号。散射输出仅提供了散射信号。根据本发明的概念,所引导的包络成形 (本发明的包络成形) 仅对直射信号分量进行进一步处理。

[0112] 包络成形过程对不同信号采用包络提取操作。由于这是在对直射信号分量应用本发明的修正之前的必经步骤,因此,在以下段落中更详细地描述直射信号修正器 68 中进行的包络提取过程。

[0113] 如上所述,在混合子带域中,使用 k 来表示子带。也可以将若干子带 k 组织在参数频带 K 中。

[0114] 图 4 的表格中给出了在以下讨论的本发明的实施例下的子带与参数频带的关联。

[0115] 首先,对于帧中的每个时隙,使用 $y^{n,k}$ 来计算特定参数频带 K 的能量 E_{slot}^K , $y^{n,k}$ 是混合子带输入信号。

$$[0116] \quad E_{slot}^K(n) = \sum_k y^{n,\tilde{k}} (y^{n,\tilde{k}})^* \quad \tilde{k} = \{k \mid \bar{k}(k) = K\} \quad \forall K_{start} < K < K_{stop}$$

[0117] 其中 $K_{start} = 10$, $K_{stop} = 18$ 。

[0118] 该求和包括根据表 A.1 归属于一个参数频带 K 的所有 $\tilde{k}$ 。

[0119] 随后,对每个参数频带,长期能量平均 $\bar{E}_{slot}^K$ 计算为:

$$[0120] \quad \bar{E}_{slot}^K(n) = (1 - \alpha) E_{slot}^K(n) + \alpha \bar{E}_{slot}^K(n-1)$$

$$[0121] \quad \alpha = \exp\left(-\frac{64}{0.4 \cdot 44100}\right)$$

[0122] 其中, α 是与一阶 IIR 低通 (约 400ms 时间常数) 相对应的加权因子, n 表示时隙索引。平滑总平均 (宽带) 能量 $\bar{E}_{total}$ 计算为:

$$[0123] \quad \bar{E}_{total}(n) = (1 - \alpha) E_{total}(n) + \alpha \bar{E}_{total}(n-1)$$

[0124] 其中

$$[0125] \quad E_{total}(n) = \frac{1}{\kappa_{stop} - \kappa_{start} + 1} \sum_{\kappa=\kappa_{start}}^{\kappa_{stop}} E_{slot}^{\kappa}(n)$$

$$[0126] \quad \alpha = \exp\left(-\frac{64}{0.4 \cdot 44100}\right)$$

[0127] 从上述方程可以看到, 在从声道的平滑表示获得增益因子之前, 对时间包络进行平滑。一般地, 平滑是指从具有递减梯度的原始声道获得平滑的表示。

[0128] 从上述方程可以看到, 随后描述的白化操作是基于时间平滑的总能量估计和子带中的平滑的能量估计, 从而确保了最终的包络估计的更好的稳定性。

[0129] 确定这些能量的比值来获得用于频谱白化操作的权值:

$$[0130] \quad w^{\kappa}(n) = \frac{\bar{E}_{total}(n)}{\bar{E}_{slot}^{\kappa}(n) + \varepsilon}$$

[0131] 通过对参数频带的加权成分求和来获得宽带包络估计, 所述成分对长期能量平均进行归一化并计算平方根

$$[0132] \quad Env(n) = \sqrt{\frac{EnvAbs(n)}{Env(n)}}$$

[0133] 其中

$$[0134] \quad EnvAbs(n) = \sum_{\kappa=\kappa_{start}}^{\kappa_{stop}} w^{\kappa}(n) \cdot E_{slot}^{\kappa}(n)$$

$$[0135] \quad \overline{Env}(n) = (1 - \beta) EnvAbs(n) + \beta \overline{Env}(n-1)$$

$$[0136] \quad \beta = \exp\left(-\frac{64}{0.04 \cdot 44100}\right)$$

[0137] β 是与一阶 IIR 低通 (约 40ms 时间常数) 相对应的加权因子。

[0138] 使用频谱白化的能量或幅度度量作为计算缩放因子的基础。从上述方程可以看到, 频谱白化是指改变频谱, 使得音频声道表示的每个频谱带内包含相同的能量或均值幅度。这是最有利的, 因为所讨论的瞬时信号具有非常宽的频谱, 使得必需使用整个可用频谱的完整信息来计算增益因子, 以不致相对于其他非瞬时信号抑制了该瞬时信号。换言之, 频谱白化的信号是在其频谱表示的不同频谱带中具有近似相等的能量的信号。

[0139] 本发明的直射信号修正器对直射信号分量进行修正。如上所述, 在传送的残留信号存在的情况下, 可以将处理限制在从起始索引开始的一些子带索引中。此外, 可以将处理总体限制在阈值索引之上的子带索引中。

[0140] 包络成形处理由对每个输出声道的直射声音包络的平坦化, 接着是向目标包络的重新成形组成。如果在辅助信息中对该声道以信号指示了 $bsEnvShapeChannel = 1$, 则这产生了对每个输出声道的直射信号应用的增益曲线。

[0141] 仅对特定混合子子带 (sub-subband) k 进行该处理:

$$[0142] \quad k > 7$$

[0143] 在传送的残留信号存在的情况下, 将 k 选择为从所讨论的声道的上混中包含的最高残留频带之上开始。

[0144] 如在之前的部分所描述的,对于 5-1-5 配置,通过估计传送的下混的包络 Env_{Dmx} 来获得目标包络,随后,使用由编码器传送的并重新量化的包络比值 $envRatio_{ch}$ 对其进行缩放。

[0145] 接着,对每个输出声道,对帧中的所有时隙,通过估计其包络 Env_{ch} 来计算其增益曲线 $g_{ch}(n)$,并将该增益曲线与目标包络相关。最终,将该增益曲线转换为对上混声道的直射部分独立进行缩放的有效增益曲线:

$$[0146] \quad ratio_{ch}(n) = \min(4, \max(0.25, g_{ch} + ampRatio_{ch}(n) \cdot (g_{ch} - 1)))$$

[0147] 其中

$$[0148] \quad g_{ch}(n) = \frac{envRatio_{ch}(n) \cdot Env_{Dmx}(n)}{Env_{ch}(n)}$$

$$[0149] \quad ampRatio_{ch}(n) = \frac{\sum_k |y_{ch,diffuse}^{n,k}|}{\sum_k |y_{ch,direct}^{n,k}| + \varepsilon}$$

$$[0150] \quad ch \in \{L, Ls, C, R, Rs\}$$

[0151] 对于 5-2-5 配置,从左声道传送的下混信号的包络 Env_{DmxL} 获得 L 和 Ls 的目标包络,使用右声道传送的下混包络 Env_{DmxR} 来获得 R 和 Rs 的目标包络。从左和右声道传送的下混信号的包络之和获得中心声道。

[0152] 对每个输出声道,通过估计其包络 $EEEnv^{L,Ls,C,R,Rs}$ 来计算增益曲线,并将该增益曲线与目标包络相关。在第二步骤中,将该增益曲线转换为对上混声道的直射部分独立进行缩放的有效增益曲线:

$$[0153] \quad ratio_{ch}(n) = \min(4, \max(0.25, g_{ch} + ampRatio_{ch}(n) \cdot (g_{ch} - 1)))$$

[0154] 其中

$$[0155] \quad ampRatio_{ch}(n) = \frac{\sum_k |y_{ch,diffuse}^{n,k}|}{\sum_k |y_{ch,direct}^{n,k}| + \varepsilon}, ch \in \{L, Ls, C, R, Rs\}$$

$$[0156] \quad g_{ch}(n) = \frac{envRatio_{ch}(n) \cdot Env_{DmxL}(n)}{Env_{ch}(n)}, ch \in \{L, Ls\}$$

$$[0157] \quad g_{ch}(n) = \frac{envRatio_{ch}(n) \cdot Env_{DmxR}(n)}{Env_{ch}(n)}, ch \in \{R, Rs\}$$

$$[0158] \quad g_{ch}(n) = \frac{envRatio_{ch}(n) \cdot 0.5(Env_{DmxL}(n) + Env_{DmxR}(n))}{Env_{ch}(n)}, ch \in \{C\}$$

[0159] 对所有声道,如果 $bsEnvShapeChannel = 1$,则应用该包络调整增益曲线。

$$[0160] \quad \tilde{y}_{ch,direct}^k(n) = ratio_{ch}(n) \cdot y_{ch,direct}^k(n), ch \in \{L, Ls, C, R, Rs\}$$

[0161] 否则,简单地复制直射信号:

$$[0162] \quad \tilde{y}_{ch,direct}^k(n) = y_{ch,direct}^k(n), ch \in \{L, Ls, C, R, Rs\}$$

[0163] 最终,每个单个声道修正的直射信号分量必须根据以下等式,与混合子带域内的对应各声道的散射信号分量组合:

[0164] $y_{ch}^{n,k} = \tilde{y}_{ch,direct}^{n,k} + y_{ch,diffuse}^{n,k}$, $ch \in \{L, Ls, C, R, Rs\}$

[0165] 从以上段落可以看到,本发明的概念教导了在空间音频解码器中提高感知质量和类似掌声的信号的空间分布。通过获得具有精细缩放时间颗粒度的增益因子,以仅对空间上混信号的直射部分进行缩放来完成了这种增强。本质上,从传送的辅助信息和编码器中对直射和散射信号的音级或能量的度量来获得该增益因子。

[0166] 虽然上述示例具体描述了基于幅度度量的计算,但是,应注意,本发明的方法不限于此,而是也可以计算例如能量度量或适合于描述信号的时间包络的其他量。

[0167] 上述示例描述了针对 5-1-5 和 5-2-5 声道配置的计算。自然,上述原理可以类似地应用到例如 7-2-7 和 7-5-7 声道配置上。

[0168] 图 5 示出了本发明的多声道音频解码器 100 的示例,多声道音频解码器 100 接收由一个原始多声道信号的多个声道下混而得到的下混声道 102,以及参数表示 104,参数表示 104 包括该原始多声道信号的原始声道(左前置、右前置、左后置和右后置)的时间结构信息。多声道解码器 100 具有产生器 106,用于对下混声道 102 下的每个原始声道产生直射信号分量和散射信号分量。多声道解码器 100 还包括针对要重构的每个声道的 4 个本发明的直射信号修正器 108a 至 108d,使得该多声道解码器在其输出 112 输出 4 个输出声道(左前置、右前置、左后置和右后置)。

[0169] 虽然使用要重构 4 个原始声道的示例配置来详细描述了本发明的多声道解码器,但是,本发明的概念可以在具有任意声道数目的多声道音频方案中实现。

[0170] 图 6 示出了一个框图,详细描述了本发明的产生重构输出声道的方法。

[0171] 在产生步骤 110,从下混声道获得直射信号分量和散射信号分量。在修正步骤 112,使用参数表示的参数来修正直射信号分量,该参数表示具有原始声道的时间结构信息。

[0172] 在组合步骤 114,将修正的直射信号分量与散射信号分量组合来获得重构的输出声道。

[0173] 根据本发明方法的特定实现需要,可以以硬件或软件实现本发明的方法。可以使用数字存储介质,尤其是具有电子可读控制信号存储在其上的磁盘、CD 或 DVD 来执行该实现方式,所述电子可读控制信号可以与可编程计算机系统协作来执行本发明的方法。一般地,因此,本发明也在于具有程序代码的计算机程序产品,所述程序代码存储在机器可读载体上,当计算机程序产品在计算机上运行时,所述程序代码执行本发明的方法。换言之,因此,本发明可以被实现为具有程序代码的计算机程序,当计算机程序在计算机上运行时,所述程序代码执行本发明的方法。

[0174] 虽然参照本发明的具体实施例具体地说明并描述了上述内容,但是,本领域技术人员应理解,在不背离其精神和范围的情况下,可以做出形式和细节上的各种其他改变。应理解,为了适应不同的实施例,在不背离此处所公开的及由所附权利要求所包括的更宽的概念的情况下,可以做出各种改变。

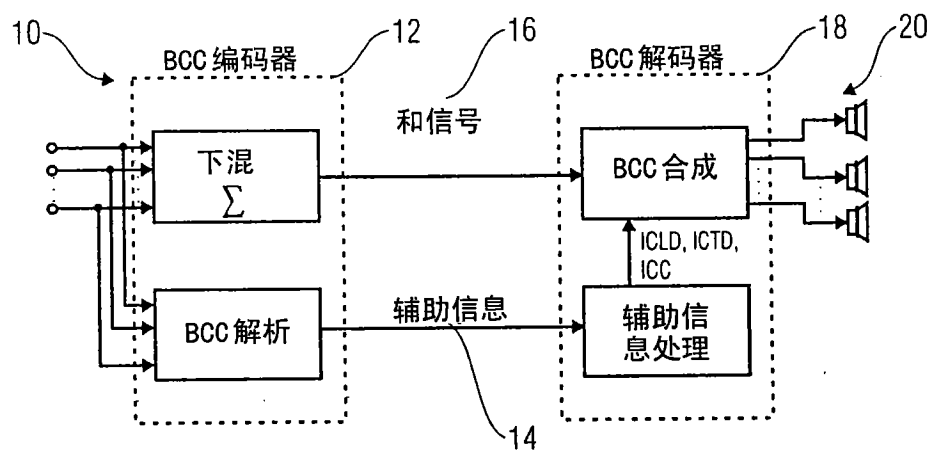


图 1

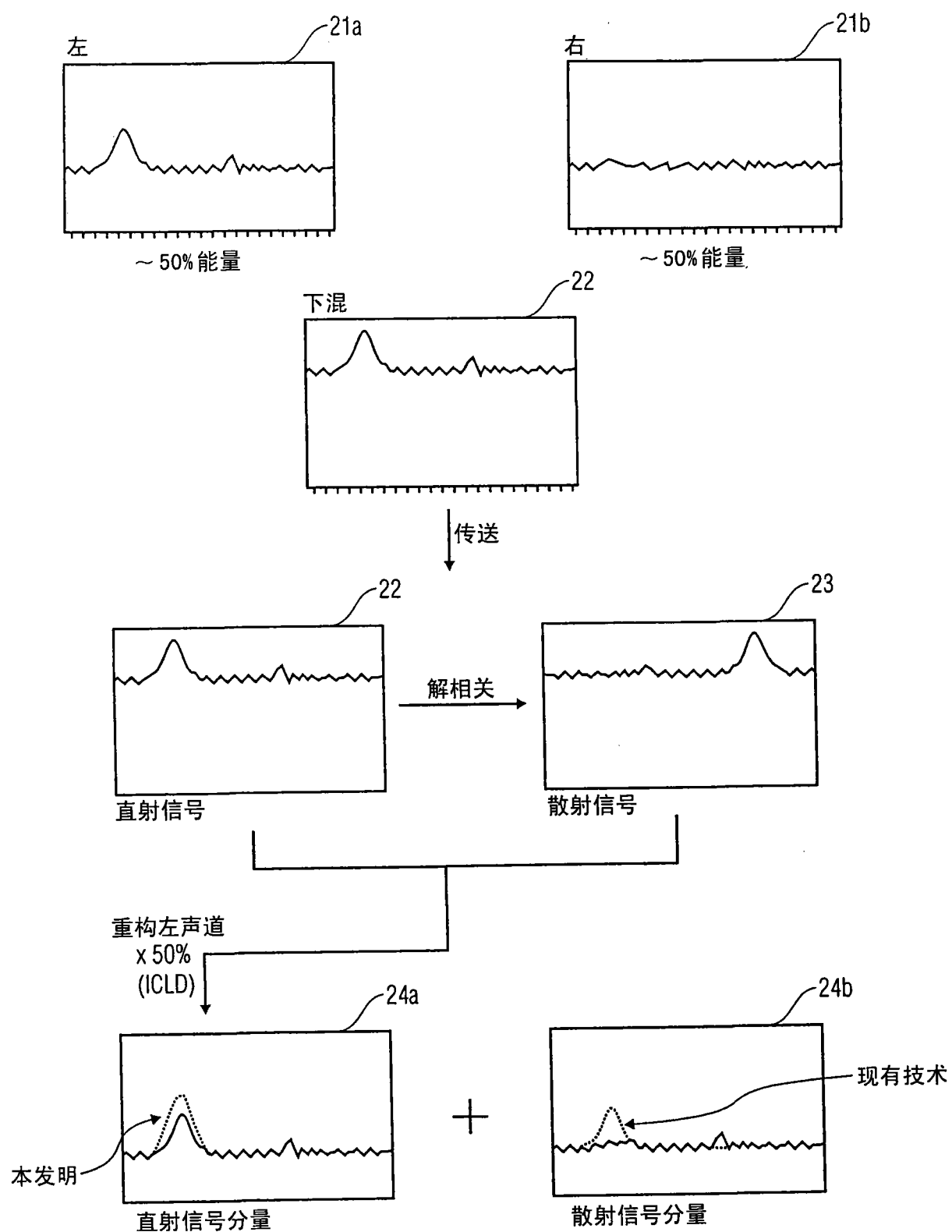


图 1b

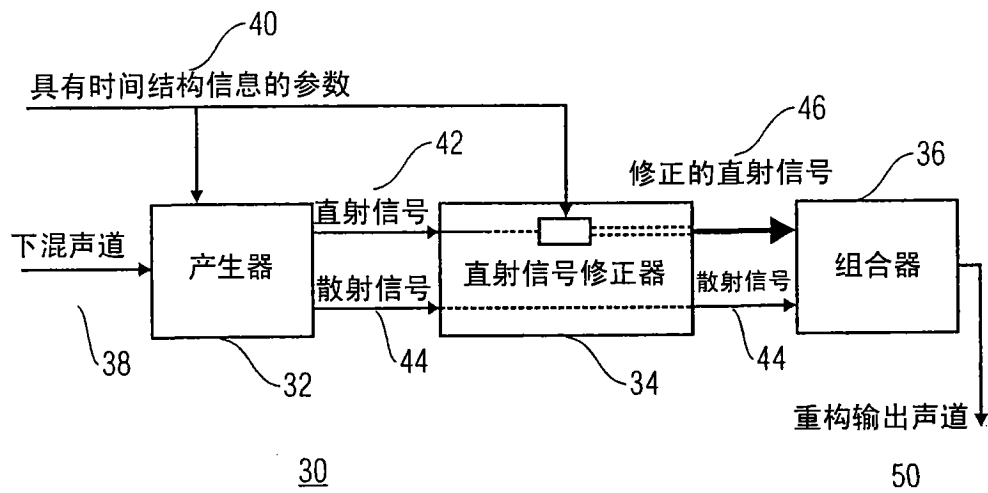


图 2

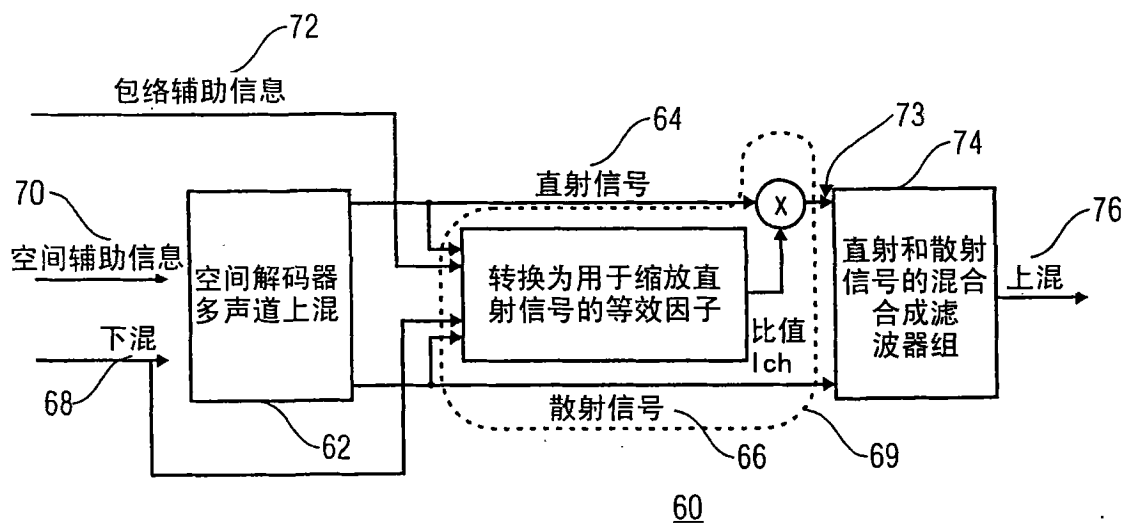


图 3

k	0	1	2	3	4	5	6	7	8	9	10
$\bar{k}(k)$	1	0	0	1	2	3	4	5	6	7	8
k	11	12	13	14	15	16	17	18	19	20	21
$\bar{k}(k)$	9	10	11	12	13	14	14	15	15	15	16
k	22	23	24	25	26	27	28	29	30	31	32
$\bar{k}(k)$	16	16	16	17	17	17	17	17	18	18	18
k	33	34	35	36	37	38	39	40	41	42	43
$\bar{k}(k)$	18	18	18	18	18	18	18	18	18	19	19
k	44	45	46	47	48	49	50	51	52	53	54
$\bar{k}(k)$	19	19	19	19	19	19	19	19	19	19	19
k	55	56	57	58	59	60	61	62	63	64	65
$\bar{k}(k)$	19	19	19	19	19	19	19	19	19	19	19
k	66	67	68	69	70						
$\bar{k}(k)$	19	19	19	19	19						

图 4

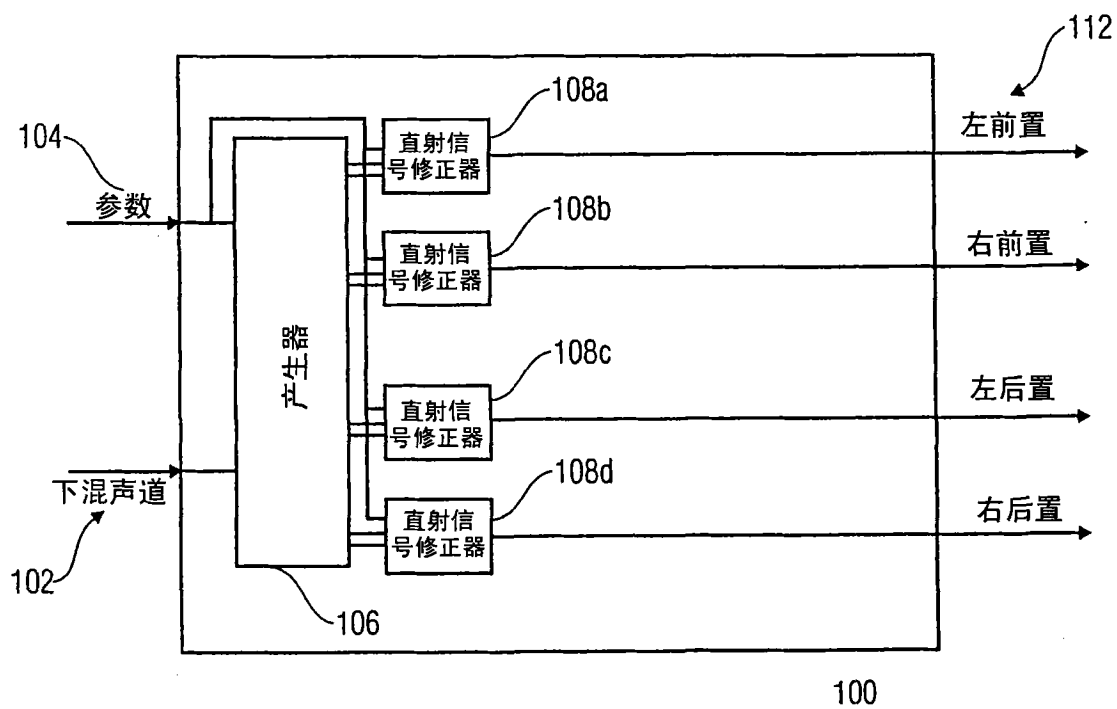


图 5

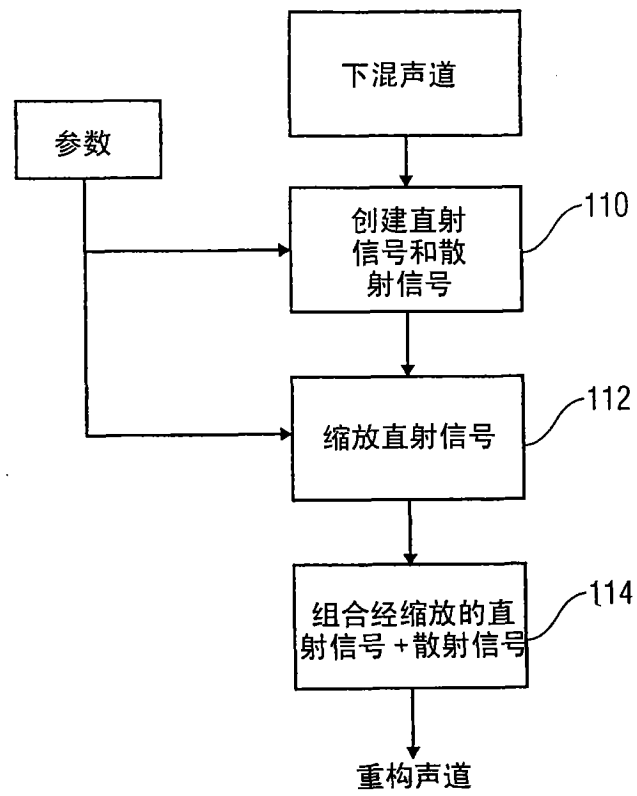


图 6