



(12) 发明专利

(10) 授权公告号 CN 101002505 B

(45) 授权公告日 2011.08.10

(21) 申请号 200580026149.6

(22) 申请日 2005.07.13

(30) 优先权数据

10/911,404 2004.08.03 US

(85) PCT申请进入国家阶段日

2007.02.02

(86) PCT申请的申请数据

PCT/US2005/024630 2005.07.13

(87) PCT申请的公布数据

W02006/019719 EN 2006.02.23

(73) 专利权人 杜比实验室特许公司

地址 美国加利福尼亚

(72) 发明人 迈克尔·J·斯密斯尔思

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 赵科

(51) Int. Cl.

H04S 3/02 (2006.01)

(56) 对比文件

US 20040032960 A1, 2004.02.19, 全文.

US 5862228 A, 1999.01.19, 说明书第6栏第39行至第65行、图6.

WO 2004019656 A2, 2004.03.04, H04S.

审查员 戴惠英

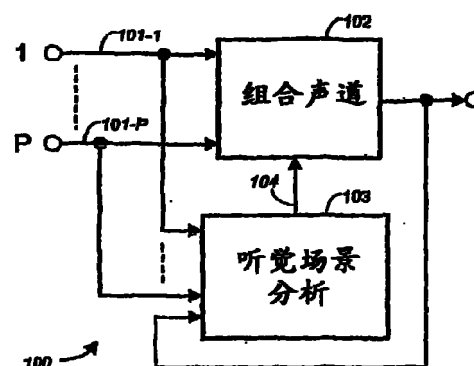
权利要求书 1 页 说明书 12 页 附图 4 页

(54) 发明名称

利用听觉场景分析组合音频信号的方法和装置

(57) 摘要

本发明提供了一种用于组合音频声道的方法,其组合音频声道以产生组合后的音频声道,并且对声道、对组合后的声道或者对声道和组合后的声道二者动态地应用时间、相位和幅度或功率调整中的一个或多个。至少部分地由一个或多个声道和/或组合后的声道中的听觉事件的量度控制调整中的一个或多个。应用包括在电影院和车辆中呈现多声道音频。不仅包括方法、而且包括相应的计算机程序实现和装置。



1. 一种用于组合音频声道的方法,包括:

组合音频声道以产生组合后的音频声道,以及

对所述声道、对所述组合后的声道或者对所述声道和所述组合后的声道二者动态地应用时间、相位和幅度调整中的一个或多个,其中:

至少部分地由一个或多个所述声道和/或所述组合后的声道中的听觉事件的量度控制所述调整中的一个或多个,

响应于声道中谱内容的改变超过阈值而识别每个听觉事件边界,以便为所述声道获得一组听觉事件边界,

其中识别声道中的听觉事件边界包括:将离散时域音频信号划分为重叠块;对每个重叠块开窗口;将每个重叠块中的数据转换到频域;以及比较当前块的归一化频谱与下一在先块的归一化频谱。

2. 根据权利要求1的方法,其中通过将当前和下一在先块谱的以分贝为单位的相应归一化谱值之差的绝对值相加来计算单个差量度,并且比较所述单个差量度与阈值。

3. 一种用于将P个音频声道缩混为Q个音频声道的方法,其中P大于Q,其中通过权利要求1的方法获得所述Q个音频声道中至少一个。

4. 根据权利要求1或3的方法,其中所述调整被控制,以在听觉事件期间保持基本恒定,并且允许在听觉事件边界处或附近发生变化。

5. 一种用于组合音频声道的装置,包括:

用于组合音频声道以产生组合后的音频声道的装置,以及

用于对所述声道、对所述组合后的声道或者对所述声道和所述组合后的声道二者动态地应用时间、相位和幅度调整中的一个或多个的装置,包括:

用于至少部分地根据一个或多个所述声道和/或所述组合后的声道中的听觉事件的量度来控制所述调整中的一个或多个的装置,以及

用于响应于声道中谱内容的改变超过阈值而识别每个听觉事件边界,以便为所述声道获得一组听觉事件边界的装置,其中识别声道中的听觉事件边界的装置包括:用于将离散时域音频信号划分为重叠块的装置;用于对每个重叠块开窗口的装置;用于将每个重叠块中的数据转换到频域的装置;以及用于比较当前块的归一化频谱与下一在先块的归一化频谱的装置。

利用听觉场景分析组合音频信号的方法和装置

技术领域

[0001] 本发明涉及改变多声道音频信号中的声道数,其中音频声道中的一些被组合。应用包括在电影院和车辆中呈现多声道音频。本发明不仅包括方法,而且还包括相应的计算机程序实现和装置实现。

[0002] 背景技术

[0003] 在过去几十年中,多声道音频资料的生产、销售和呈现已经有不断增加的增长。该增长主要是由其中 5.1 声道重放系统几乎无处不在的电影行业以及最近由开始生产 5.1 多声道音乐的音乐行业推动的。

[0004] 典型地,通过具有与资料相同数量的声道的重放系统呈现这种音频资料。例如,5.1 声道电影声带可以在 5.1 声道电影院中或者通过 5.1 声道家庭影院音频系统被呈现。但是,越来越多地期望在不具有与音频资料中相同数量的声道的系统上或环境中播放多声道资料—例如,在只有两个或四个重放声道的车辆中重放 5.1 声道资料,或者在只配备有 5.1 声道系统的电影院中重放大于 5.1 声道电影声道。在这种情况下,需要组合或“缩混(downmix)”多声道信号中一些或所有声道用于呈现。

[0005] 声道的组合可能产生可听人造缺陷(artifact)。例如,一些频率分量可能消除,而其它频率分量增强或变得更高声。最一般地,这是在由于正被组合的两个或更多声道中存在相似或相关音频信号分量。

[0006] 本发明的目的是最小化或抑制由于组合声道而产生的人造缺陷。其它目的将在阅读和理解本文档时被了解。

[0007] 应该注意,声道的组合可能被需要用于其它目的,而不只是用于声道数量的减少。例如,可能需要创建作为多声道信号中两个或更多原始信道的某种组合的附加重放声道。这可以被表现为结果比声道原始数量更多的“上混”类型。因此,不管在“缩混”还是在“上混”的情形中,组合声道以创建附加声道都可能导致可听的人造缺陷。

[0008] 用于最小化混合或声道组合人造缺陷的常见技术包括例如对将被组合的声道、对所得到的组合后的声道或者对二者应用时间、相位和幅度(或功率)调整中的一个或多个。音频信号本身是动态的——即,它们的特性随时间改变。因此,典型地以动态方式计算和应用对音频信号的这种调整。在消除由于组合所导致的一些人造缺陷的同时,这种动态处理可能引发其它人造缺陷。为了最小化这种动态处理人造缺陷,本发明利用听觉场景分析,使得总体上,动态处理调整在听觉场景或事件期间被基本保持恒定,并且这种调整的变化只在听觉场景或事件边界处或附近被允许。

[0009] 听觉场景分析

[0010] 将声音分成被感知为隔离开的单元有时被称作“听觉事件分析”或“听觉场景分析”(“ASA”)。Albert S. Bregman 在他的书 Auditory Scene Analysis—The Perceptual Organization of Sound, Massachusetts Institute of Technology, 1991, Fourth printing, 2001, Second MIT Press paperback edition 中对听觉场景进行了详细讨论。

[0011] 用于根据听觉场景分析识别听觉事件(包括事件边界)的技术在 2003 年 11 月

20 日提交的 Brett G. Crockett 的标题为“Segmenting Audio Signals into Auditory Events”、案卷号 DOL098US 的美国专利申请 S. N. 10/478, 538 中给出, 该申请是 2002 年 2 月 2 日提交的指定美国的、在 2002 年 12 月 5 日被公开为 WO 02/097792 的国际申请 PCT/US02/05999 的相应美国国家申请。所述申请在此通过引用被整体包含。所述 Crockett 申请的听觉事件识别技术的某些应用在 2003 年 11 月 20 日提交的 Brett G. Crockett 和 Michael J. Smithers 的标题为“Comparing Audio Using Characterizations Based on Auditory Events”、案卷号 DOL092US 的美国专利申请 S. N. 10/478, 397 以及在 2003 年 11 月 20 日提交的 Brett G. Crockett 和 Michael J. Smithers 的标题为“Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events”、在 2004 年 7 月 29 日被公开为 US 2004/0148159 A1、案卷号 DOL09201US 的美国专利申请 S. N. 10/478, 398 中给出, 上述两个申请分别是由 2002 年 2 月 22 日提交的指定美国的、在 2002 年 12 月 5 日被公开为 WO 02/097790 的国际申请 PCT/US02/05329 得到的美国国家申请和由 2002 年 2 月 25 日提交的指定美国的、在 2002 年 12 月 5 日被公开为 WO 02/097791 的国际申请 PCT/US02/05806 得到的美国国家申请。所述 Crockett 和 Smithers 的申请中每一个也在此通过引用被整体包含。

[0012] 虽然在所述 Crockett 和 Crockett/Smithers 申请中所描述的技术与本发明的方面结合是特别有用的, 但是用于识别听觉事件和事件边界的其它技术也可以被利用在本发明的方面中。

发明内容

[0013] 根据本发明的一个方面, 一种用于组合音频声道过程, 包括组合音频声道以产生组合后的音频声道, 以及对声道、对组合后的声道或者对声道和组合后的声道两者动态地应用时间、相位和幅度或功率调整中的一个或多个, 其中, 所述调整中的一个或多个至少部分地由一个或多个声道和 / 或组合后的声道中听觉事件的量度控制。调整可以被控制, 以在听觉事件期间保持基本恒定以及允许在听觉事件边界处或附近的变化。

[0014] 本发明的主要目的是改善组合后的音频信号的声音质量。这可以例如通过对音频信号不同地进行时间、相位和 / 或幅度 (或功率) 校正以及通过至少部分地用听觉场景分析信息的量度控制这种校正来实现。根据本发明的方面, 被应用于音频信号的调整一般地可以在听觉事件期间被保持相对恒定并且被允许在听觉事件之间的转变或边界处或附近的变化。当然, 这种调整不需要如每个边界那样频繁地出现。这种调整的控制可以响应于每个声道中的听觉事件信息而基于逐个声道地被实现。或者, 这种调整中的一些或全部可以响应于已经在所有声道或少于所有的声道上被组合的听觉事件信息被实现。

[0015] 本发明的其它方面包括用于与上述过程以及本申请中所描述的其它过程的计算机程序实现一起来执行所述过程的装置或设备。

附图说明

[0016] 图 1 是本发明一般实施例的功能示意框图。

[0017] 图 2 是体现本发明方面的音频信号处理或处理方法的功能示意框图。

[0018] 图 3 是更详细地示出图 2 的时间和相位校正 202 的功能示意框图。

[0019] 图 4 是更详细地示出图 2 的混合声道 206 的功能示意框图。

[0020] 图 5a 是示出白噪声信号的幅度谱的理想化响应。图 5b 是示出通过将由白噪声组成的第一声道与第二信号简单组合而得到的幅度谱的理想化响应,其中第二信号是相同的白噪声信号,只是在时间上被延迟大约毫秒的若干分之一。在图 5a 和 5b 中,水平轴是单位为赫兹的频率,竖轴是单位为分贝 (dB) 的相对电平。

[0021] 图 6 是根据本发明方面从三声道到二声道的缩混的功能示意框图。

[0022] 图 7a 和 7b 是示出在诸如电影院礼堂的房间中两组音频声道的空间位置的理想化表示。图 7a 示出了多声道音频信号的“内容”声道的近似空间位置,而图 7b 示出了在被配备用于播放五声道音频资料的电影院中“重放”的近似空间位置。

[0023] 图 7c 是根据本发明方面的十声道到五声道缩混的功能示意框图。

具体实施方式

[0024] 本发明的一般化实施例在图 1 中示出,其中音频声道组合器或组合过程 100 被示出。多个音频输入声道, P 个输入声道 101-1 到 101-P 被提供给声道组合器或组合功能 (“组合声道”) 102 以及听觉场景分析器或分析功能 (“听觉场景分析”) 103。可以有二个或更多将被组合的输入声道。声道 1 到 P 可以构成一组输入声道的一部分或全部。组合声道 102 组合被提供给它的声道。虽然这种组合可以是例如线性相加组合,但是组合技术对本发明不重要。除了组合被提供给它的声道之外,组合声道 102 还对将被组合的声道、对所得到的组合后的声道、或者对将被组合的声道和所得到的组合后的声道两者应用时间、相位和幅度或功率调整中的一个或多个。这种调整可能是通过减少混合或声道组合人造缺陷而为了改善声道组合质量的目的而做出的。具体的调整技术对本发明不重要。用于组合和调整的适当技术的实例在 2004 年 3 月 1 日提交的 Mark Franklin Davis 的标题为“Low Bit Rate AudioEncoding and Decoding in Which Multiple Channels Are Represented by a Monophonic Channel and Auxiliary Information”、案卷号为 DOL11501 的美国临时专利申请 S. N. 60/549, 368 ;2004 年 6 月 14 日提交的 Mark Franklin Davis 等人的标题为“Low Bit Rate AudioEncoding and Decoding in Which Multiple Channels Are Represented by a Monophonic Channel and Auxiliary Information”、案卷号为 DOL11502 的美国临时申请 S. N. 60/579, 974 ;以及 2004 年 7 月 14 日提交的 Mark Franklin Davis 等人的标题为“Low Bit Rate AudioEncoding and Decoding in Which Multiple Channels Are Represented by a Monophonic Channel and Auxiliary Information”、案卷号为 DOL11503 的美国临时申请 S. N. 60/588, 256 中给出。Davis 和 Davis 等人的所述三个临时申请中每一个在此通过引用被整体包含。听觉场景分析 103 例如根据在上述申请中一个或多个中所描述的技术或者通过某种其它合适的听觉场景分析器或分析过程而得到听觉场景信息。应该至少包括听觉事件之间边界位置的这种信息 104 被提供给组合声道 102。所述调整中的一个或多个至少部分地被将被组合的一个或多个声道和 / 或所得到的组合后的声道中听觉事件的量度控制。

[0025] 图 2 示出了体现本发明方面的音频信号处理器或处理方法 200 的实例。将被组合的来自多个音频声道 1 到 P 的信号 101-1 到 101-P, 被提供给时间和 / 或相位校正设备或过程 (“时间和相位校正”) 202 以及听觉场景分析设备或过程 (“听觉场景分析”) 103, 如结

合图 1 所描述的。声道 1 到 P 可以构成一组输入声道的一部分或全部。听觉场景分析 103 得到听觉场景信息 104, 并将其提供给时间和相位校正 202, 时间和相位校正 202 对将被组合的声道中每一个单独应用时间和 / 或相位校正, 如下面结合图 3 所描述的。校正后的声道 205-1 到 205-P 然后被提供给声道混合设备或过程 (“混合声道”) 206, 其组合声道以创建单个输出声道 207。可选地, 混合声道 206 还可以由听觉场景分析信息 104 控制, 如下面进一步所描述的。体现如图 1 和 2 的实例中的本发明方面的音频信号处理器或处理方法还可以组合声道 1 到 P 中不同声道以产生多于一个的输出声道。

[0026] 听觉场景分析 103 (图 1 和 2)

[0027] 听觉场景分析研究已经显示, 耳朵使用多个不同的听觉提示 (auditory cue) 来识别所感知的听觉事件的开始和结束。如上述申请中所教导的那样, 最有力的提示之一是音频信号的谱内容变化。对于每个输入声道, 听觉场景分析 103 以限定的时间间隔对每个声道 1 到 P 的音频进行谱分析, 以创建信号的频率表示序列。以上述申请中所描述的方式, 连续表示可以被比较, 以找到谱内容中大于阈值的变化。找到这种变化就表示这对连续频率表示之间的听觉事件边界, 大约指出一个听觉事件的结尾和另一听觉事件的开头。每个输入声道的听觉事件边界的位置被输出作为听觉场景分析信息 104 的分量。虽然这可以在上述申请中所描述的方式实现, 但是听觉事件和它们的边界可以通过其它合适的技术被检测。

[0028] 听觉事件是所感知的、在整个事件过程中具有基本保持恒定的特性的声音单元。如果诸如可以被用在本发明实施例中的时间、相位和 / 或幅度 (或功率) 调整在听觉事件内变化很大, 则这种调整的效果可能变为可听的, 从而构成不期望的人造缺陷。通过将调整在整个事件期间保持为恒定并且只改变充分靠近事件边界的调整, 听觉事件的相似性没有被破坏, 并且变化可能被隐藏在音频内容中的固有地表示事件边界的更明显的变化中。

[0029] 理想地, 根据本发明的方面, 声道组合或 “缩混” 参数应该被允许只在听觉事件边界处改变, 使得在事件内没有动态改变发生。但是, 用于检测听觉事件的实际系统典型地在数字域中操作, 其中在数字域中, 时域中的数字音频采样块被转变成频域, 使得听觉事件边界的时间分辨率 (time resolution) 具有相当粗糙的时间分辨率, 该分辨率与数字音频采样的块长度相关。如果该分辨率被选择 (以块长度与频率分辨率之间的折中) 用于产生对实际事件边界的有用近似, 也就是说, 如果分辨率产生足够靠近从而错误不会被听者感知到的近似边界, 则为了根据本发明的动态缩混的目的, 不使用未知的实际边界而使用由块边界所提供的近似就足够了。因此, 根据上述 Crockett 申请中的实例, 事件边界可以被确定为在块长度的一半内, 或者在利用 44.1kHz 采样率的系统中对于 512 采样块长度的实例为大约 5.8 毫秒。

[0030] 在本发明方面的一个实际实现中, 每个输入声道是离散时域音频信号。该离散信号可以被分为大约 10.6 毫秒的重叠块, 其中重叠为大约 5.3 毫秒。对于 48kHz 的音频采样率, 这等同于 512 个采样块中 256 个采样与前面的块重叠。每一块可以被使用例如汉宁窗而开窗口, 并且被利用例如离散傅立叶变换 (为了速度被实现为快速傅立叶变换) 而转变到频域中。为每个谱值计算单位为分贝 (dB) 的功率, 然后谱被归一化到最大的 dB 谱值。不重叠或部分重叠的块可以被用于减少计算成本。此外, 其它窗函数可能被使用, 但是, 汉宁窗已经被发现非常适合于本申请。

[0031] 如上述 Crockett 的申请中所描述的那样,可以将用于当前块的归一化频谱与来自下一在先块的归一化谱比较,以得到它们差别的量度。具体地,可以通过对当前和下一在先谱的 dB 谱值的绝对差值求和来计算单个差别量度。

[0032] 这种差别量度然后可以被与阈值相比教。如果差别量度大于阈值,则表示事件边界在当前和在先块之间,否则表示在当前和在先块之间没有事件边界。已经发现该阈值的合适的值是 2500(单位为 dB)。这样,可以在大约块的一半的精确度内确定事件边界。

[0033] 这种阈值方法能够被应用于其中每个子带具有不同差别量度的频率子带。但是,在本发明的上下文中,考虑到人类在任何时刻集中于一个事件的感知能力,基于全带宽音频的单个量度是足够的。

[0034] 用于每个声道 1 到 P 的听觉事件边界信息被输出作为听觉场景分析信息 104 的分量。

[0035] 时间和相位校正 202(图 2)

[0036] 时间和相位校正 202 寻找输入声道对之间的高相关性和时间或相位差。图 3 更详细地示出了时间和相位校正 202。如下面解释的,每对中的一个声道是基准声道。一个合适的相关性检测技术在下面描述。其它合适的相关性检测技术可以被利用。当高相关性存在于非基准声道与基准声道之间时,设备或过程试图通过修改非基准声道的相位或时间特性来减小声道对之间的相位或时间差,从而减少或消除否则可能由于该对声道的组合而导致的可听声道组合人造缺陷。这种人造缺陷中的一些可以通过实例的方式描述。图 5a 示出了白噪声信号的幅度谱。图 5b 示出了通过将由白噪声组成的第一声道与第二信号简单组合而得到的幅度谱,其中第二信号是相同的白噪声信号,只是在时间上被延迟大约 0.21 毫秒。白噪声信号的未被延迟和被延迟版本的组合具有共同被称为梳状滤波的抵消和谱整形(spectral shaping),以及对每个输入信号的白噪声非常不同的可听声音。

[0037] 图 3 示出了用于消除相位或时间延迟的合适设备或方法 300。来自每个输入音频声道的信号 101-1 到 101-P 被提供给延迟计算设备或过程(“计算延迟”)301,其输出用于每个声道的延迟指示信号 302。可以具有用于每个声道 1 到 P 的分量的听觉事件边界信息 104 被包括临时存储设备或过程(“保持”)303 的设备或过程用于有条件地更新分别被每个声道的延迟设备或功能(“延迟”)305-1 到 305-P 使用的延迟信号 304- 到 304-P,以产生输出声道 306-1 到 306-P。

[0038] 计算延迟 301(图 3)

[0039] 计算延迟 301 测量输入声道对之间的相对延迟。一个优选方法是首先从输入声道中选择基准声道。该基准可以是固定的或者它可以随事件改变。允许基准声道改变就克服了例如无声基准声道(silent reference channel)的问题。如果基准声道变换,则它可能例如由声道响度确定(例如,最响的是基准)。如前面所提到的,用于每个输入声道的输入音频信号可以被分为重叠大约 5.3 毫秒的长度大约 10.6 毫秒的重叠块。对于 48kHz 的音频采样率,这等同于 512 个采样块中 256 个采样与前面的块重叠。

[0040] 可以使用任何合适的交叉相关(cross-correlation)方法计算每一个非基准声道与基准声道之间的延迟。例如,设 S_1 (长度 N_1) 是来自基准声道的采样块,而 S_2 (长度 N_2) 是来自非基准声道中一个的采样块。首先计算交叉相关阵列 $R_{1,2}$ 。

$$[0041] \quad R_{1,2}(l) = \sum_{n=-\infty}^{\infty} S_1(n) \cdot S_2(n-l) \quad l = 0, \pm 1, \pm 2, \dots \quad (1)$$

[0042] 可以利用基于标准 FFT 的技术执行交叉相关以减少执行时间。因为 S_1 和 S_2 在长度上是有限的, 所以 $R_{1,2}$ 的非零分量具有 N_1+N_2-1 的长度。与 $R_{1,2}$ 中最大元素相对应的滞后 l 代表 S_2 相对于 S_1 的延迟。

$$[0043] \quad l_{\text{peak}} = l \text{ for } \text{MAX}[R_{1,2}(l)] \quad (2)$$

[0044] 该滞后或延迟具有与阵列 S_1 和 S_2 相同的采样单位。

[0045] 利用一阶无限冲击响应滤波器对当前块的交叉相关结果与来自在先块的交叉相关结果进行时间平滑, 以创建平滑后的交叉相关 $Q_{1,2}$ 。下面的等式示出了滤波器计算, 其中 m 指代当前块, $m-1$ 指代在先块。

$$[0046] \quad Q_{1,2}(l, m) = \alpha \times R_{1,2}(l) + (1-\alpha) \times Q_{1,2}(l, m-1) \quad l = 0, \pm 1, \pm 2, \dots \quad (3)$$

[0047] 已经发现 α 的有用值是 0.1。对于交叉相关 $R_{1,2}$, 与 $Q_{1,2}$ 中最大元素相对应的滞后 l 代表 S_2 相对于 S_1 的延迟。每个非基准声道的滞后或延迟被输出作为信号 302 的信号分量。零值也可以输出作为信号 302 的分量, 代表基准声道的延迟。

[0048] 可以被测量的延迟范围与音频信号块大小成比例。也就是, 块大小越大, 可以利用该方法测量的延迟范围就越大。

[0049] 保持 303 (图 3)

[0050] 当经由 ASA 信息 104 为声道指示事件边界时, 保持 303 将用于该声道的延迟值从 302 复制到相应的输出声道延迟信号 304。当没有事件边界被指示时, 保持 303 维持上一个延迟值 304。以这种方式, 时间对准变化出现在事件边界处, 并且因此更不可能导致可听的人造缺陷。

[0051] 延迟 305-1 到 305-P (图 3)

[0052] 因为延迟信号 304 可以是正或负, 所以延迟 305-1 到 305-P 中每一个缺省地都可以被实现以将每个声道延迟可以由计算延迟 301 所计算的绝对最大延迟。因此, 延迟 305-1 到 305-P 每一个中的总采样延迟是各个输入延迟信号 304-1 到 304-P 与延迟缺省量的和。这允许信号 302 和 304 是正或负, 其中负表示声道在时间上比基准声道靠前。

[0053] 当输入延迟信号 304-1 到 304-P 中任何一个改变值时, 可能必须消除或复制采样。优选地, 以不引发可听人造缺陷的方式执行。这种方法可以包括重叠和平滑转换 (crossfading) 采样。或者, 因为输出信号 306-1 到 306-P 可能被提供给滤波器组 (见图 4), 组合延迟和滤波器组使得延迟控制被提供给滤波器组的采样的对准可能是有用的。

[0054] 或者, 更复杂的方法可以测量和校正各个频段或频段组的时间或相位差。在这种更复杂的方法中, 计算延迟 301 和延迟 305-1 到 305-P 可以在频域中操作, 在这种情况下, 延迟 305-1 到 305-P 对频带或子带执行相位调整, 而不是在时域中的延迟。在这种情况下, 信号 306-1 到 306-P 已经在频域中, 从而不需要随后的滤波器组 401 (图 4, 如下面所描述的)。

[0055] 设备或过程中的一些、诸如计算延迟 301 和听觉场景分析 103 可能预先访问音频声道, 以提供将被应用到事件内的更精确的事件边界估计和时间或相位校正。

[0056] 混合声道 206 (图 2)

[0057] 图 2 的混合声道 206 的细节被示为图 4 中的设备或过程 400, 图 4 示出了输入声

道可以如何通过功率校正被组合以创建被缩混后的输出声道。除了混合或组合声道之外,该设备或过程可以对没有被图 2 中的时间和相位校正 203 完全校正的剩余频率相消进行校正。它还用于保持功率守恒。换句话说,混合声道 206 试图确保输出缩混信号 414 (图 4) 的功率基本上与被时间或相位调整的输入声道 205-1 到 205-P 的功率之和相同。而且,它可以试图确保被缩混后的信号的每个频带中的功率基本上是各个被时间或相位调整的输入声道的相应频带的功率之和。过程通过将来自被缩混后的声道的频带功率与来自输入声道的频带功率相比较并且随后计算每个频带的增益校正来实现这一点。因为时间和频率两者上增益调整的变化可能导致可听的人造缺陷,所以增益优选地在被提供给被缩混后的信号之前被进行时间和频率两者的平滑。该设备或过程代表组合声道的一个可能途径。其它合适的设备或过程可以被利用。具体的组合设备或过程对本发明不重要。

[0058] 滤波器组 (“FB”) 401-1 到 401-P (图 4)

[0059] 每个输入声道的输入音频信号是时域信号,并且可以已经被分成重叠大约 5.3 毫秒的长度大约 10.6 毫秒的重叠块,如前面所提到的。对于 48kHz 的音频采样率,这等同于 512 个采样块中 256 个采样与前面的块重叠。采样块可以由滤波器组 401-1 到 401-P (每个输入信号一个滤波器组) 被开窗口并被转变到频域。虽然可以使用各种窗口类型中任何一种,但是已经发现汉宁窗是合适的。虽然可以使用各种时域到频域转换器或转换过程中任何一个,但是合适的转换器或转换方法可以使用离散傅立叶变换 (为了速度被实现为快速傅立叶变换)。每个滤波器组的输出是各自的复谱值 (complex spectral value) 阵列 402-1 到 402-P- 每个频带 (或箱 (bin)) 一个值。

[0060] 频带 (“BND”) 功率 403-1 到 403-P (图 4)

[0061] 对于每个声道,频带功率计算器或计算过程 (“BND 功率”) 403-1 到 403-P 分别估算或计算复谱值 402-1 到 402-P 的功率,并输出它们作为各个功率谱 404-1 到 404-P。来自每个声道的功率谱值在相加组合器或组合功能 415 中被相加,以创建新的组合后的功率谱 405。来自每个声道的相应复谱值 402-1 到 402-P 也在相加组合器或组合功能 416 中被相加,以创建缩混复谱 406。缩混复谱 406 的功率在另一功率计算器或计算过程 (“BND 功率”) 403 中被计算并被输出作为缩混功率谱 407。

[0062] 频带 (“BND”) 增益 408 (图 4)

[0063] 频带增益计算器或计算过程 (频带增益 408) 通过缩混功率谱 407 划分功率谱 405,以创建功率增益或功率比率 (power ratio) 的阵列,每个谱值一个。如果缩混功率谱值为零 (使功率增益为无限),则相应的功率增益被设定为“1”。然后计算功率增益的平方根,以创建幅度增益 409 阵列。

[0064] 限制、时间和频率平滑 410 (图 4)

[0065] 限制器和平滑器或者限制和平滑功能 (限制、时间和频率平滑) 410 执行合适的增益限制和时间 / 频率平滑。上面刚讨论的谱幅度增益可以具有很广的范围。如果增益被保持在有限范围内,则可以获得最佳结果。例如,如果任何增益大于上限阈值,则它被设定为等于上限阈值。同样,例如,如果任何增益小于下限阈值,则它被设定为等于下限阈值。有用的阈值是 0.5 和 2.0 (相当于 $\pm 6\text{dB}$)。

[0066] 然后可以利用一阶无限冲击响应 (IIR) 滤波器对谱增益进行时间平滑。下面的等式示出了滤波器计算,其中 b 指代谱频带索引, B 指代总频带数, n 指代当前块, n-1 指代在

先块, G 指代未被平滑的增益, G_s 指代时间平滑增益。

[0067] $G_s(b, n) = \delta(b) \times G(b) + (1 - \delta(b)) \times G_s(b, n-1)$ $b = 0, \dots, B-1$ (4)

[0068] 已经被发现,除了低于大约 200Hz 的频带, $\delta(b)$ 的有用值是 0.5。低于该频率, $\delta(b)$ 趋向在频带 $b = 0$ 或 DC 处为零的最终值。如果平滑后的增益 G_s 被初始化成 1.0, 则 DC 处的值保持等于 1.0。即, DC 将永远不被增益调整, 并且低于 200Hz 的频带增益将比其余谱中的频带变化得更慢。这可能对防止更低频率下的可听调制是有用的。这是因为, 在低于 200Hz 的频率下, 这种频率的波长接近或超过滤波器组所使用的块大小, 从而导致滤波器组精确区分这些频率的能力不精确。这是常见且众所周知的现象。

[0069] 被时间平滑后的增益进一步在频率上被平滑, 以防止相邻频带之间大的增益变化。在优选实现中, 利用滑动五频带 (或大约 470Hz) 平均来平滑频带增益。即, 每个箱被更新为其本身与频率在其上和其下的两个相邻频带的平均值。在谱的上下边缘处, 边缘值 (频带 0 和 $N-1$) 被重复使用, 使得五频带平均仍旧可以被执行。

[0070] 被平滑后的频带增益被输出作为信号 411, 并且在乘法器或乘法功能 419 中乘以缩混复谱值, 以创建校正后的缩混复谱 412。可选地, 输出信号 411 可以在 ASA 信息 104 的控制下经由临时存储设备或过程 (“保持”417) 被提供给乘法器或乘法功能 419。保持 417 以与图 3 的保持 303 相同的方式操作。例如, 增益可以在事件期间被保持为相对恒定并且只在事件边界处被改变。以这种方式, 可以防止事件期间可能的可听的动态增益变化。

[0071] 反向滤波器组 (Inv FB) 413 (图 4)

[0072] 来自乘法器或乘法功能 419 的缩混谱 412 被传递经过反向滤波器组或滤波器组功能 (“INV FB”) 413, 以创建输出时间采样块。该滤波器组是输入滤波器组 401 的逆。相邻块与在先块重叠并被添加到在先块, 如众所周知的那样, 以创建输出时域信号 414。

[0073] 所描述的布置不排除在前向滤波器组 401 处将窗口分成两个窗口 (一个被用在前向滤波器组, 另一个被用在反向滤波器组) 的常见实施, 其中这两个窗口的增加使得在整个系统中保持一致的信号。

[0074] 缩混应用

[0075] 根据本发明方面的缩混的一个应用是在机动车辆中重放 5.1 声道内容。机动车辆可能只再现 5.1 声道内容的四个声道, 大致对应于这种系统的左、右、左环绕和右环绕声道。每个声道被指向位于被认为适合用于再现与具体声道相关联的定向信息的位置上的一个或多个扬声器。但是, 机动车辆通常不具有用于再现这种 5.1 重放系统中中声道 (center channel) 的中央扬声器位置。为了适应这种情况, 已知削弱中声道信号 (例如 3dB 或 6dB) 并且将它与左右声道信号中每一个组合以提供虚幻中声道。但是, 这种简单的组合导致前面描述的人造缺陷。

[0076] 代替这种简单的组合, 可以应用根据本发明方面的声道组合或缩混。例如, 图 1 的布置或图 2 的布置可以被应用两次, 一次用于组合左和中信号, 一次用于组合中和右信号。但是, 在将中声道信号与左声道和右声道信号中每一个组合之前将其削弱例如 3dB 或 6dB (在机动车辆内部的近场空间中, 6dB 可能比 3dB 更合适) 使得来自中声道信号的输出声功率与通过专用中声道扬声器呈现时近似相同可能仍旧有利的。而且, 在将中信号与左声道和右声道信号中每一个组合时将中信号指代作为基准声道使得对中声道信号所应用的时间和相位校正 103 不改变中声道的时间对准或相位而只改变左声道和右声道信号的

时间对准或相位可能有利的。因此,中声道信号将不会在两个加法(即,左声道加中声道信号的加法,以及右声道加中声道信号的加法)的每一个中被不同地调整,从而确保虚幻中声道声像(image)保持稳定。

[0077] 反过来也可以是可应用的。即,只对中声道进行时间或相位调整,同样确保虚幻中声道声像保持稳定。

[0078] 根据本发明方面的缩混的另一个应用是在电影院中重放多声道音频。为下一代数字电影院系统开发的标准要求高达并且很快将多于 16 声道音频的传送。大部分所安装的电影院系统只提供 5.1 重放或“呈现”声道(如所公知的那样,“0.1”代表低频“效果”声道)。因此,在重放系统被升级之前,在巨大开销下,需要将具有多于 5.1 声道的内容缩混到 5.1 声道。这种声道的缩混或组合导致上述人造缺陷。

[0079] 因此,如果 P 声道将被缩混到 Q 声道(这里 $P > Q$),则根据本发明方面(例如如图 1 和 2 的示例实施例中那样)的缩混可以被应用以获得一个或多个 Q 输出声道,其中输出声道中的一部分或全部是 P 输入声道的各个声道中两个或多个的组合。如果输入声道被组合到多于一个输出声道中,则将这种声道指代为基准声道使得图 2 中的时间和相位校正 202 对于其被组合到的每个输出声道不会不同地改变这种输入声道的时间对准或相位可能是有利的。

[0080] 可选方案

[0081] 如这里描述的时间或相位调整用于使缩混期间频率的完全或部分相消最小化。前面描述了,当输入声道被组合到多于一个输出声道中时,该声道优选地被指代为基准声道,使得它在被混合到多个输出声道时不被不同地时间或相位调整。这在其它声道不具有基本上相同的内容时工作得很好。但是,在两个或更多其它声道具有相同或基本上相同的内容时,问题就可能出现。如果这种声道被组合到多于一个输出声道中,则在收听所得到的输出声道时,公共内容在接收那些输出声道的扬声器的物理位置之间某处的方向上被感知为空间中的虚幻声像。当具有基本上等同的内容的这两个或更多个输入声道在被与其它声道组合以创建输出声道之前被独立地进行相位调整时出现了问题。独立的相位调整可能导致不正确的虚幻声像位置,和/或不确定的声像位置,这两者都可能在听觉上被感知为不自然。

[0082] 可以设计寻找具有基本上相似的内容的输入声道并且试图以相同或相似的方式对这种声道进行时间或相位调整使得它们的虚幻声像位置不被改变的系統。但是,这种系统变得非常复杂,特别是当输入声道数变得远远大于输出声道数时。在其中基本相似的内容频繁出现在多于一个输入声道中的系统中,免除相位调整而只执行功率校正可能更简单。

[0083] 这个调整问题可以在前面描述的汽车应用中被进一步说明,其中在所述应用中,中声道信号被与左和右声道中每一个组合,用于分别通过左和右扬声器重放。在 5.1 声道资料中,左和右输入声道经常包含多个信号(例如,乐器、嗓音、对话和/或效果),其中一些是不同的,而其中一些是相同的。当中声道被与左和右声道中每一个混合时,中声道被指代为基准声道,并且不被时间或相位调整。左声道被时间或相位调整,以在与中声道组合时产生最小的相位相消,类似地,右声道被时间或相位调整,以在被与中声道组合时产生最小的相位相消。因为左和右声道被独立地进行时间或相位调整,所以左和右声道公共的信号可

能不再具有左和右扬声器的物理位置之间的虚幻声像。而且, 虚幻声像可能不被定位到任何一个方向上, 而是可能在整个收听空间中被传播—不自然的非期望效果。

[0084] 调整问题的解决方法是从这种输入声道中提取对于多于一个输入声道公共的信号, 并将它们放到新的分开的输入声道中。虽然这增加了将被缩混的输入声道 P 的总数量, 但是它减少了在输出的被缩混后的声道中假的非期望的虚幻声像。为三个声道被缩混为两个的情况在图 6 中示出汽车实例设备或过程 600。利用任何合适的声道乘法器或乘法过程 (“解相关声道”) 601、诸如主动矩阵解码器或提取公共信号分量的其它类型的声道乘法器从左和右声道中提取左和右输入声道公共的信号到另一个新的声道中。这种设备可以表现为一种解相关器或解相关功能。已知为 Dolby Surround Pro Logic II (杜比定向逻辑环绕 II) 的一种合适的主动矩阵解码器在 2000 年 3 月 22 日提交的 James W. Fosgate 的标题为 “Method for deriving at least three audiosignals from two input audio signals”、案卷号为 DOL07201 的美国专利申请 S. N. 09/532, 711 以及 2003 年 2 月 25 日提交的 James W. Fosgate 等人的标题为 “Method for apparatus for audio matrix decoding” 并于 2004 年 7 月 1 日公开为 US 2004/0125960、案卷号为 DOL07203US 的美国专利申请 S. N. 10/362, 786 中描述, 其中美国专利申请 S. N. 10/362, 786 是于 2001 年 8 月 30 日提交、指定美国并于 2002 年 3 月 7 日公开为 WO 02/19768 的国际申请 PCT/US01/27006 的相应美国国家申请。所述 Fosgate 和 Fosgate 等人的申请在此通过引用被整体包含。可以被利用的另一种类型的合适的声道乘法器和解相关器在 2003 年 8 月 5 日提交的 Mark Franklin Davis 的标题为 “Audio Channel Translation” 并于 2004 年 4 月 1 日公开为 US 2004/0062401 A1、案卷号为 DOL088US 的美国专利申请 S. N. 10/467, 213 中描述, 该申请是由 2002 年 2 月 7 日提交的指定美国并于 2003 年 8 月 7 日公开为 W002/063925 的国际申请 PCT/US02/03619 以及 2003 年 8 月 6 日提交的指定美国、案卷号为 DOL08801PCT 并于 2004 年 3 月 4 日公开为 W02004/019656 的国际申请 PCT/US03/24570 得到的美国国家申请。所述 Davis 申请中每一个在此通过引用被全部包含。另一种合适的声道乘法 / 解相关技术在 Mitianoudis 和 Davies 于 2002 年 5 月 10 到 13 日在德国慕尼黑的第 112 届年会上发表的 Audio Engineering Society Convention Paper 5529 中的 “Intelligent Audio Source Separation using Independent Component Analysis” 被描述。所述论文在此也通过引用被全部包含。结果是四个声道, 新的声道 C_0 、原始中声道 C 和修改后的左和右声道 L_0 和 R_0 。

[0085] 基于图 2 的布置但是这里具有两个输出声道的设备或过程 602 组合四个声道, 以创建左和右重放声道 L_p 和 R_p 。修改后的声道 L_0 和 R_0 每一个被混合为仅仅一个重放声道, 分别是 L_p 和 R_p 。因为它们基本上不包含任何相关的内容, 所以已经从其中提取了它们的公共分量 C_0 的修改后的声道 L_0 和 R_0 可以被时间或相位调整, 而不影响存在于输入声道 L 和 R 中的任何虚幻中声像。为了执行时间和 / 或相位调整, 声道中的一个、诸如声道 C_0 被指代为基准声道。其它声道 L_0 、 R_0 和 C 然后相对于基准声道被时间和 / 或相位调整。或者, 因为 L_0 和 R_0 声道不可能与 C 声道相关, 并且因为它们通过过程 601 被从 C_0 声道解相关, 所以它们可以被传递给混合通道, 而不进行任何时间或相位调整。原始声道 C 和所得到的中声道 C_0 两者都可以在设备或过程 602 的混合声道部分中被分别与中间声道 L_0 和 R_0 中每一个混合, 以生成重放声道 L_p 和 R_p 。虽然已经发现 C 和 C_0 的相等比率产生令人满意的结果, 但是精确的比例不重要, 并且可能是不相等的。因此, 被应用于 C_0 和 C 的任何时间和相位调整

将出现在两个重放声道中,因此保持虚幻中声像的方向。可能在中声道中每一个上需要一些削弱(例如 3dB),因为这些声道通过两个而不是一个扬声器再现。同样,被混合到输出声道中的中声道 C 和 C_0 中每一个的量可以由收听者控制。例如,收听者可能期望所有的原始中声道 C,而不是所得到的中声道 C_0 上的某种削弱。

[0086] 解决方法还可以通过电影院音频中实例的方式被说明。图 7a 和 7b 示出了两组音频声道的房间或空间位置。图 7a 示出了以其他方式被指代为“内容声道”的多声道音频信号中所呈现的声道的大致空间位置。图 7b 示出了被指代为“重放声道”的、可以在配备有播放五声道音频资料的电影院中被再现的声道的大致位置。内容声道中的一些具有相应的重放声道位置;即 L、C、R、 R_s 和 L_s 声道。其它内容声道不具有相应的重放声道位置,并因此必须被混合到一个或多个重放声道中。典型的方法是将这种内容声道组合到最近的两个重放声道中。

[0087] 如前面所提到的,简单的相加组合可能导致可听的人造缺陷。如同样提到过的那样,结合图 1 和 2 所描述的组合在具有基本上公共的内容的声道被不同地相位或时间调整时也可能导致虚幻声像人造缺陷。解决方法包括从这种声道中提取对于多于一个输入声道所公共的信号,并将它们放到新的分开的声道中。

[0088] 图 7c 示出了用于其中通过利用设备或过程(“解相关声道”)701 提取对于输入或内容声道的一些组合公共的信息而创建五个附加声道 Q1 到 Q5 这种情况的设备或过程 700。设备或过程 701 可以使用诸如上述的合适的声道乘法/解相关技术用于“解相关声道”设备或功能 601 中。这些附加的中间声道的实际数量和空间位置可以根据内容声道中所包含的音频信号的变化而变化。基于图 2 的布置但这里具有五个输出声道的设备或过程 702 组合来自解相关声道 701 的中间声道,以创建五个重放声道。

[0089] 对于时间和相位校正,诸如 C 声道的中间声道中的一个可以被指代为基准声道,并且所有其它中间声道可以相对于该基准被时间和相位调整。或者,可能有利的是,指代多于一个的声道作为基准声道,并因此在比中间声道总数量更少的声道组中执行时间或相位校正。例如,如果声道 Q_1 代表从内容声道 L 和 C 中所提取出的公共信号,并且如果 Q_1 和 L_c 正被与中间声道 L 和 C 组合以创建重放声道 L 和 C,则声道 L_c 可以被指代为基准声道。中间声道 L、C 和 Q_1 然后相对于基准中间声道 L_c 被时间或相位调整。每一个更小的中间声道组被相继地执行时间或相位调整,直到所有中间声道已经被时间和相位校正过程考虑。

[0090] 在创建重放声道中,设备或过程 702 可以假定内容声道空间位置的先验知识。有关附加中间声道的数量和空间位置的信息可以被假定或者可以被从解相关设备或过程 701 经由路径 703 传递到设备或过程 702。这使能过程或设备 702 能够将附加中间声道组合到例如最近的两个重放声道,从而维持到这些附加声道的虚幻声像方向。

[0091] 实现

[0092] 本发明可以在硬件或软件或者两者的组合(例如可编程逻辑阵列)中被实现。除非特别指出,否则被包括作为本发明一部分的算法不是固有地与任何具体的计算机或其它装置相关。具体地,各种通用机器可以与根据这里的教导而写的程序一起使用,或者可以更方便地构造更专用的装置(例如集成电路)来执行所需的方法步骤。因此,本发明可以被实现于一个或多个可编程计算机系统上执行的一个或多个计算机程序中,其中可编程计算机系统每一个都包括至少一个处理器、至少一个数据存储系统(包括易失和非易失性存

储器和 / 或存储单元)、至少一个输入设备或端口、以及至少一个输出设备或端口。程序代码被应用于输入数据,以执行这里描述的功能并产生输出信息。输出信息被以已知方式提供给一个或多个输出设备。

[0093] 每一个这样的程序可以以任何期望的计算机语言(包括机器、汇编或高级过程、逻辑或面向对象的编程语言)实现,以与计算机系统通信。在任何情况下,语言可能是编译或解释语言。

[0094] 每一个这样的计算机程序优选地被存储在或者被下载到通用或专用可编程计算机可读的存储介质或设备(例如,固态存储器或介质,或者磁或光学介质)上,用于在存储介质或设备被计算机系统读取时配置和操作计算机以执行这里描述的程序。本发明的系统还可以被考虑被实现为计算机可读的存储介质,配置有计算机程序,其中被这样配置的存储介质使计算机系统以特定和预定的方式操作以执行这里描述的功能。

[0095] 本发明的大量实施例已经被描述。然而,应该理解,各种修改可以被做出而不脱离本发明的精神和范围。例如,上面描述的一些步骤可以是顺序独立的,因此可以以不同于所描述的顺序被执行。相应地,其它实施例在下面的权利要求的范围内。

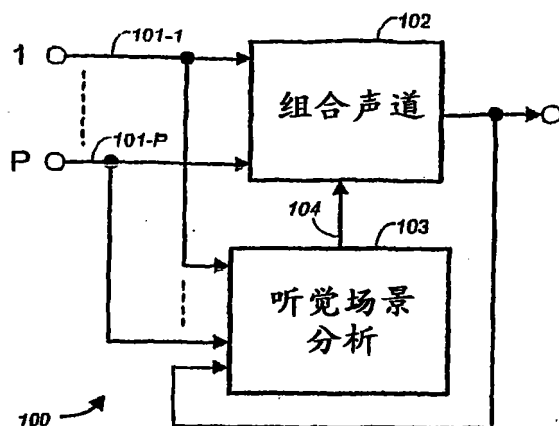


图 1

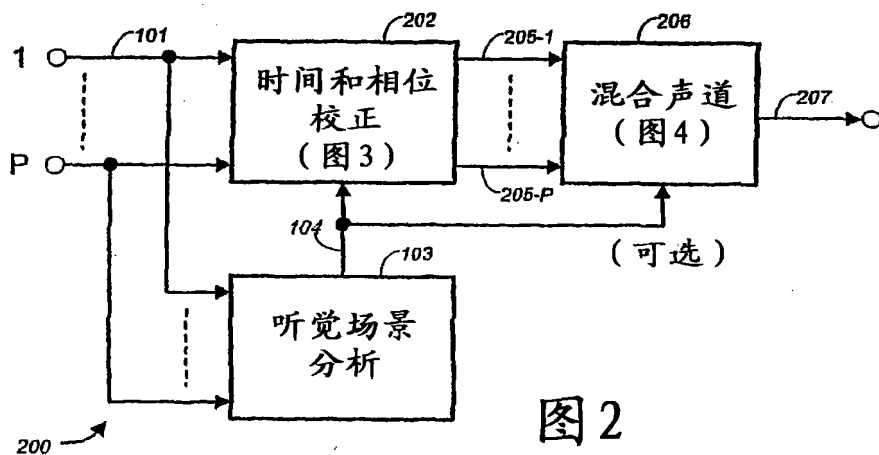


图 2

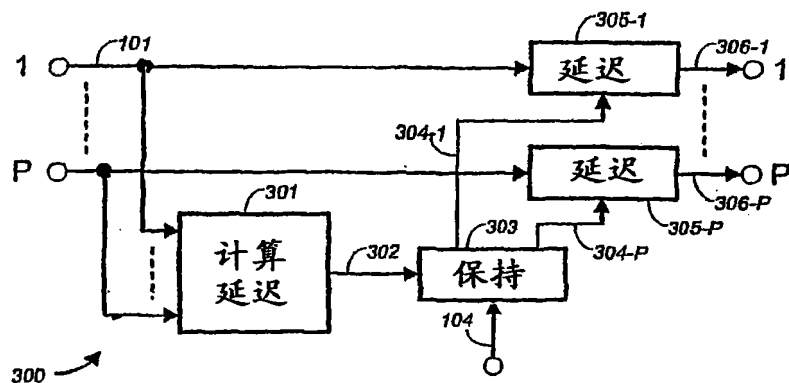


图 3

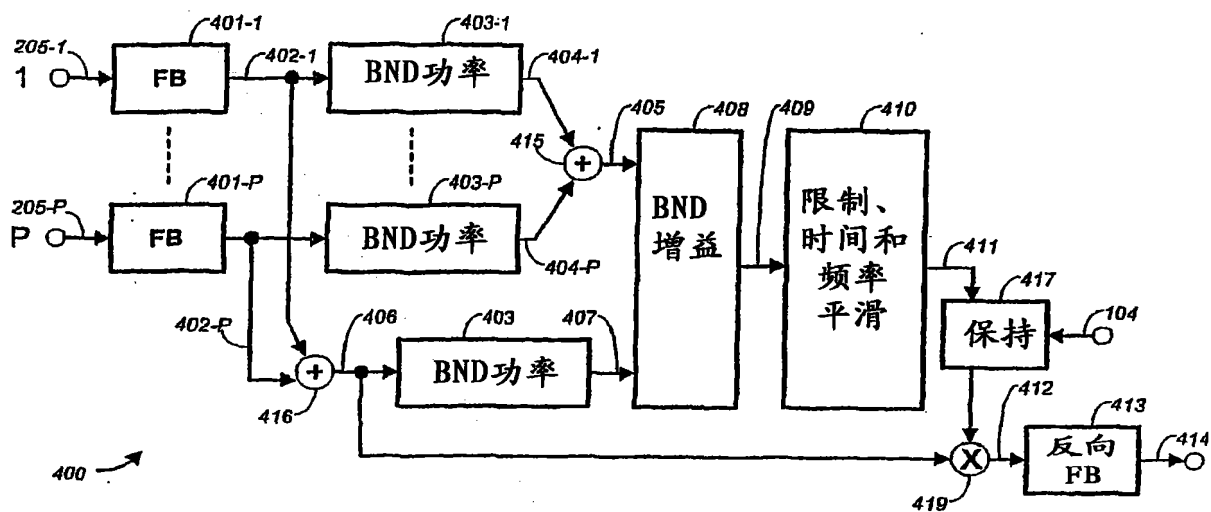


图 4

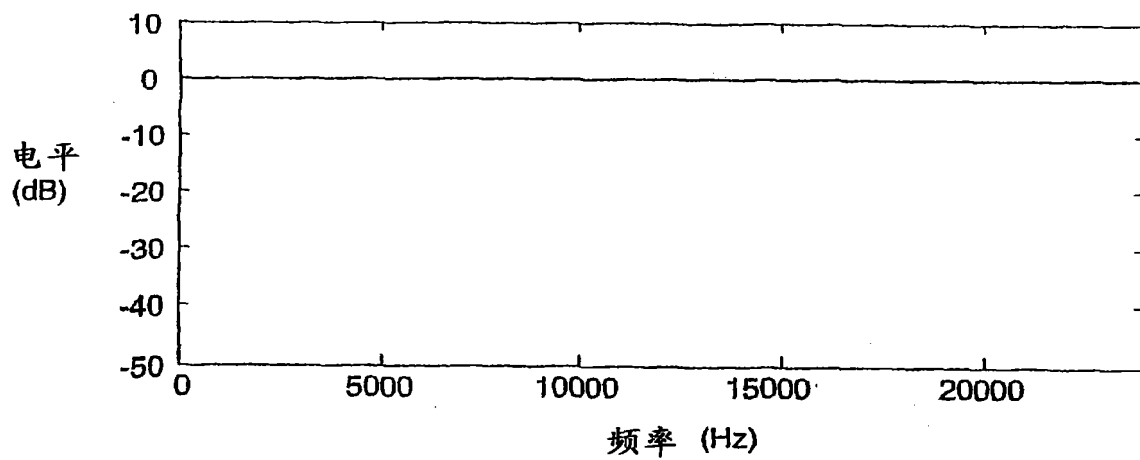


图 5a

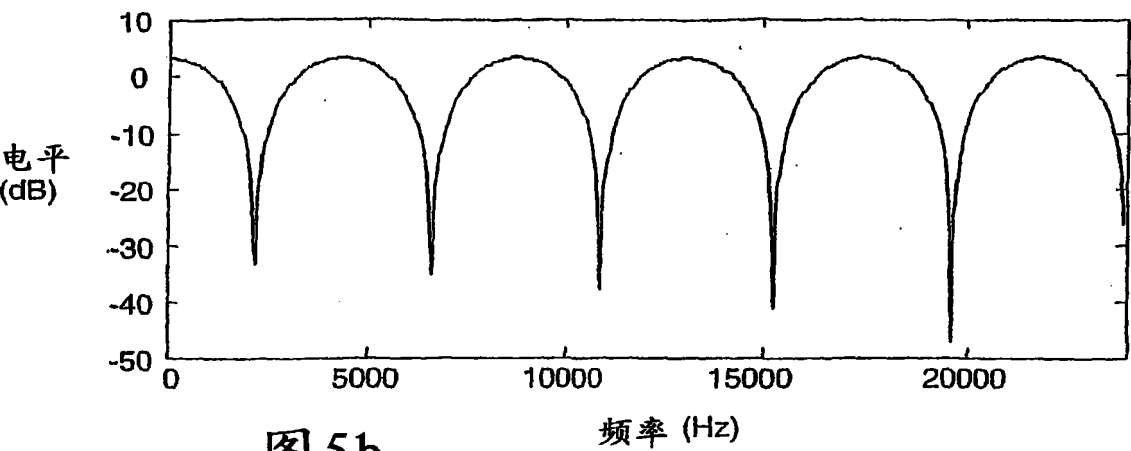


图 5b

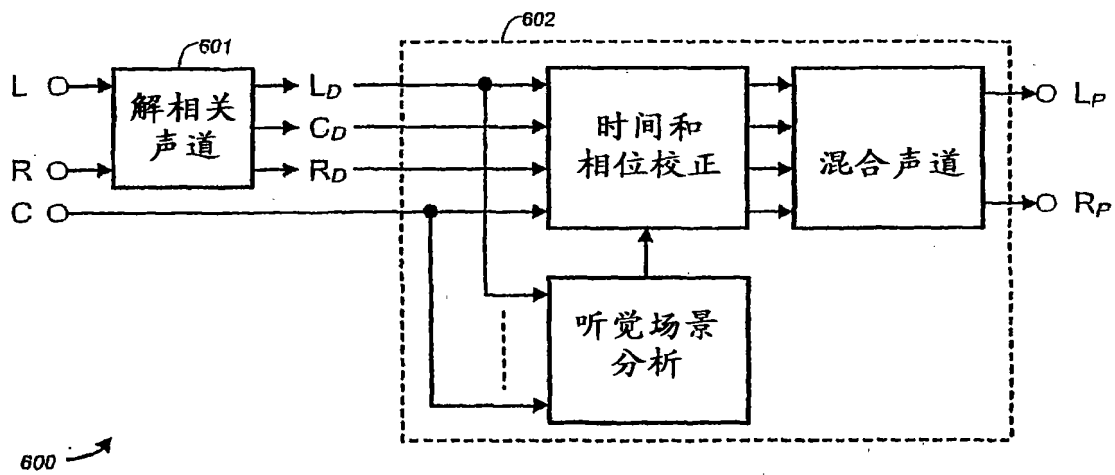
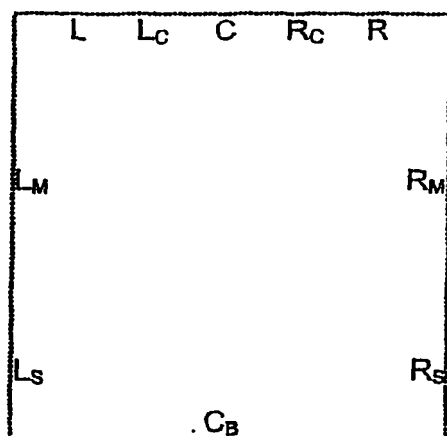
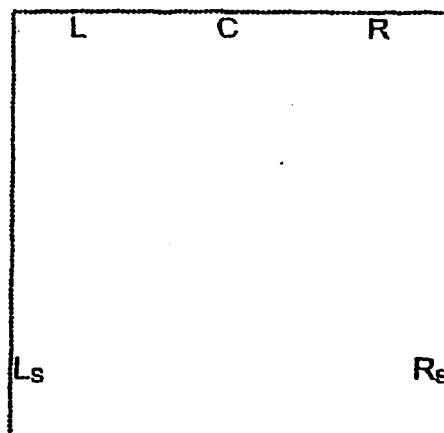


图 6



内容声道

图 7a



重放声道

图 7b

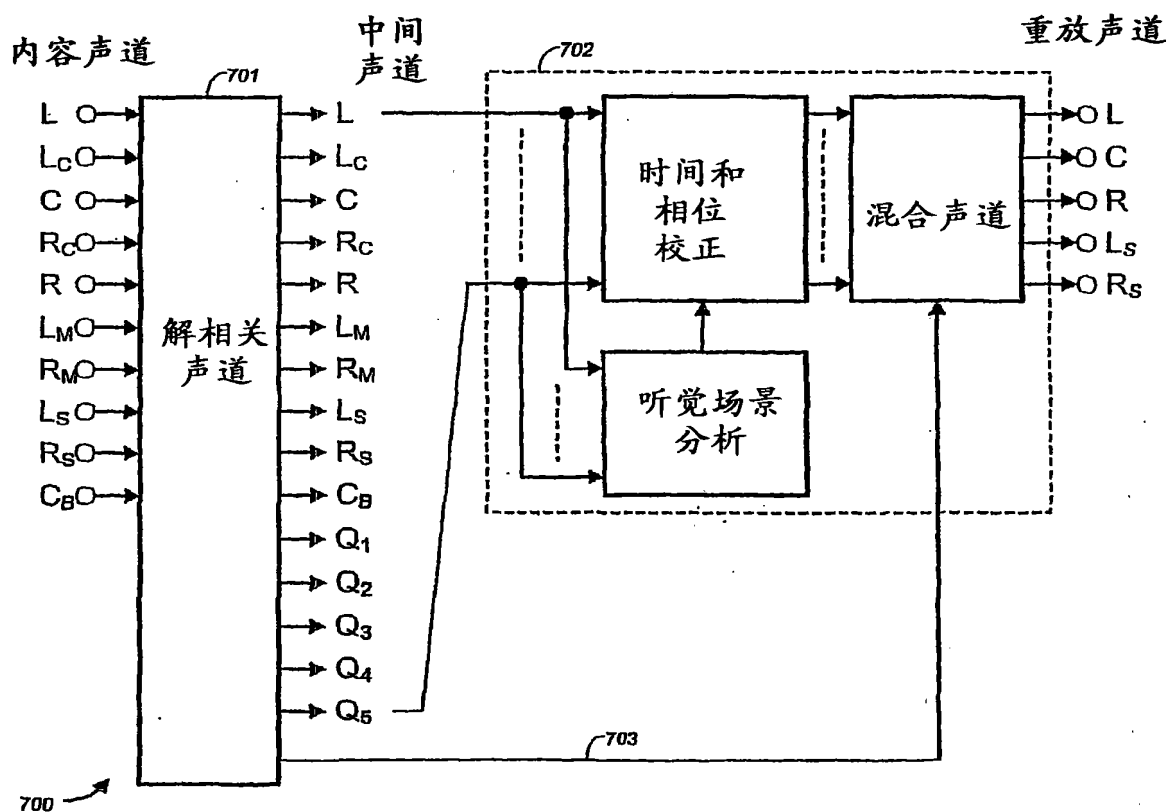


图 7c