



(12) 发明专利

(10) 授权公告号 CN 1655651 B

(45) 授权公告日 2010.12.08

(21) 申请号 200510008254.9

(22) 申请日 2005.02.07

(30) 优先权数据

60/544,287 2004.02.12 US

10/815,591 2004.04.01 US

(73) 专利权人 艾格瑞系统有限公司

地址 美国宾夕法尼亚

(72) 发明人 弗兰克·鲍姆加特

克里斯多夫·法勒

(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 郭思宇

(51) Int. Cl.

H04S 3/00 (2006.01)

H04S 5/00 (2006.01)

(56) 对比文件

CN 1413063 A, 2003.04.23, 全文.

EP 1208725 A2, 2002.05.29, 全文.

审查员 王芳

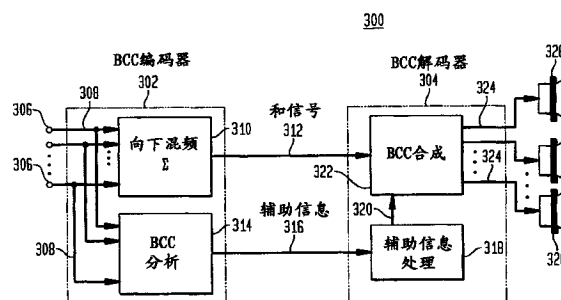
权利要求书 1 页 说明书 19 页 附图 10 页

(54) 发明名称

用于合成听觉场景的方法和设备

(57) 摘要

一种声道间相关 (ICC) (正规化交叉相关) 线索的立体声和多声道合成方案, 用于参量立体声和多声道编码。该方案合成 ICC 线索, 使得它们逼近原始的线索。为此, 产生漫射音频声道并与传输的组合 (例如求和) 信号 (多) 混合。最好使用以指数衰减高斯脉冲响应的相对长的滤波器产生漫射音频声道。这种脉冲响应产生类似于回响的漫射声。提出了另一种可选的用于降低计算复杂性的实现, 其中声道间电平差 (ICLD), 声道间时间差 (ICTD), 及 ICC 合成, 包括用于漫射声产生的滤波, 都在单个短时付立叶变换 (STFT) 域中进行。



1. 一种用于合成听觉场景的方法,包括以下步骤:
将至少一个输入声道从时域转换到频域以产生多个频域 FD 输入信号;
延迟并换算 FD 输入信号,以产生多个换算的延迟的 FD 输入信号;
对至少一个输入声道滤波以产生两个或多个 FD 漫射信号;以及
组合这两个或多个 FD 漫射信号与所述多个换算的延迟的 FD 输入信号,以产生用于听觉场景的多个输出声道。
2. 如权利要求 1 的方法,其中:
对于每一输出声道,组合包括:
对换算的延迟的 FD 输入信号之一以及对应的 FD 漫射信号之一求和,以产生一个 FD 输出信号;以及
把 FD 输出信号从频域转换为时域,以产生输出声道。
3. 如权利要求 2 的方法,其中对至少一个输入声道滤波包括:
向至少一个输入声道施加两个或多个后期混响滤波器,以产生多个漫射声道;
把漫射声道从时域转换到频域,以产生多个未换算的 FD 漫射信号;以及
换算该未换算的 FD 漫射信号以产生作为多个换算的 FD 漫射信号的所述 FD 漫射信号,其中换算的 FD 漫射信号与换算的延迟的 FD 输入信号组合,以产生 FD 输出信号。
4. 如权利要求 2 的方法,其中对至少一个输入声道滤波包括:
向 FD 输入信号施加两个或多个 FD 后期混响滤波器,以产生多个未换算的 FD 漫射信号;以及
换算未换算的 FD 漫射信号以产生作为多个换算的 FD 漫射信号的所述 FD 漫射信号,其中换算的 FD 漫射信号与换算的延迟的 FD 输入信号组合,以产生 FD 输出信号。
5. 如权利要求 1 的方法,其中:
该方法对小于规定阈值频率的输入声道频率施以所述从时域到频域的转换、延迟并换算、滤波和组合步骤;以及
该方法进而对大于规定阈值频率的输入声道频率施以涉及基于相干的没有滤波的 BCC 编码的替代听觉场景合成处理。
6. 一种用于合成听觉场景的设备,包括:
用于将至少一个输入声道从时域转换到频域的装置,以产生多个频域 FD 输入信号;
用于延迟并换算 FD 输入信号的装置,以产生多个换算的延迟的 FD 输入信号;
用于滤波至少一个输入声道的装置,以产生两个或多个 FD 漫射信号;以及
用于组合两个或多个 FD 漫射信号与多个换算的延迟的 FD 输入信号的装置,以产生用于听觉场景的多个输出声道。

用于合成听觉场景的方法和设备

技术领域

[0001] 本发明涉及音频信号编码及从编码的音频数据的听觉场景的后继合成。

[0002] 对相关申请的交叉参考

[0003] 本申请要求作为代理人提要 no. Faller 12 在 02/12/04 提交的美国临时申请 No. 60/544, 287 的权益。这一申请的主题涉及以下专利申请的主题, 作为代理人提要 no. Faller 5 在 05/04/2001 提交的序号 09/848, 877 美国专利申请 (“877 申请”), 作为代理人提要 no. Baumgarte 1-6-8 在 11/07/2001 提交的序号 10/045, 458 美国专利申请 (“458 申请”), 以及作为代理人提要 no. Baumgarte 2-10 (“437 申请”) 在 05/24/2002 提交的序号 10/155, 437 美国专利申请。并参见 C. Faller and F. Baumgarte, “Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression,” Preprint 112th Conv. Aud. Eng. Soc., May, 2002。

[0004] 背景技术

[0005] 当人听到由一特定的音频源产生的音频信号 (即声音) 时, 音频信号一般将在不同的时间并以两个不同的音频 (例如分贝) 级到达人的左和右耳, 其中不同的时间和级分别是音频信号行进到达左和右耳的路径差别的函数。人的大脑解释这种时间和级别的这些差, 以向人给出这样的感觉, 即接收的音频信号是由位于相对于人特定的位置 (例如方向和距离) 的音频源产生的。听觉场景是同时听到由位于相对于人的一个或多个不同位置的一个或多个不同音频源产生的音频信号的人的净效果。

[0006] 由大脑进行的这一处理的存在能够用来合成听觉场景, 其中来自一个或多个不同音频源的音频信号被故意修改以产生左和右音频信号, 给出不同音频源相对于收听者位于不同位置的感觉。

[0007] 图 1 示出传统的双耳信号合成器 100 的高级框图, 该合成器把单个的音频源信号 (例如单声道信号) 转换为双耳信号的左和右音频信号, 其中双耳信号定义为在收听者的耳鼓收到的两个信号。除了音频源信号之外, 合成器 100 接收一组对应于相对于收听者的音频源所需位置的空间线索。在典型的实现方法中, 这组空间线索包括声道间的电平差 (ICLD) 值 (该值标识分别作为左和右耳收到的左和右音频信号之间音频电平差), 以及声道间的时间差 (ICTD) 值 (其标识分别作为左和右耳收到的左和右音频信号之间到达的时间差)。此外或作为替代实现, 某些合成技术涉及对于从信号源到耳鼓的声音的与方向有关的传递函数, 还涉及与头部相关的传递函数 (HRTF)。例如参见 J. Blauert, The Psychophysics of Human Sound Localization, MIT Press, 1983。

[0008] 使用图 1 的双耳信号合成器 100, 由单声源产生的单声道音频信号被这样处理, 当通过头戴耳机收听时, 通过施加一组适当的空间线索 (例如 ICLD, ICTD, 和 / 或 HRTF) 在空间放置声源以产生用于每一耳的音频信号。例如参见, D. R. Begault, 3-d Sound for Virtual Reality and Multimedia, Academic Press, Cambridge, MA, 1994。

[0009] 图 1 的双耳信号合成器 100 产生最简单类型的听觉场景: 其具有相对于收听者定位的单音频源。使用听觉场景合成器能够产生更复杂的听觉场景, 包括相对于收听者位于

不同位置的两个或多个音频源,这种合成器主要是使用多个双耳信号合成器样品实现的,其中每一双耳信号合成器样品产生对应于不同音频源的双耳信号。由于每一不同的音频源有相对于收听者不同的位置,对每一不同音频源使用不同的空间线索集合产生双耳音频信号。

[0010] 图 2 示出传统听觉场景合成器 200 的高级框图,其对每一不同音频源使用不同空间线索集合,把多个音频源信号(例如多个单声道信号)转换为单个组合的双耳信号的左和右音频信号。然后左音频信号被组合(例如通过简单的相加)以对于结果的听觉场景产生左音频信号,并对于右音频信号类似。

[0011] 听觉场景合成的应用之一是会议。例如假设有多个参加者的桌面会议,每一参加者在不同城市坐在他或她的个人计算机(PC)前。除了 PC 监视器之外,每一参加者的 PC 装有(1)一麦克风,其产生对应于该参加者对会议音频部分贡献的单音频源信号,以及(2)用于播放该音频部分的一组头戴耳机。在每一参加者 PC 监视器上显示作为从坐在桌子一端的人的视界所观察的会议桌的图像。在桌面不同位置显示的是其他会议参加者的实时视频图像。

[0012] 在传统的单声道会议系统中,服务器把来自所有参加者的单声道信号组合为单个的组合的单声道信号,该信号被传送回每一参加者。为了使对每一参加者的感觉更加真实,使得他或她与其他参加者似乎坐在房间中实际会议桌旁,服务器可实现一种听觉场景合成器,诸如图 2 的合成器 200,把适当的空间线索集合施加到来自每一不同参加者的单声道音频信号,并然后组合不同左和右音频信号,以对每一听觉场景产生单个组合的双耳信号的左和右音频信号。然后用于这一组合的双耳信号的左和右音频信号被传送到每一参加者。使用这种传统的立体声会议系统的问题之一涉及到传输带宽,因为服务器必须向每一会议参加者传送左音频信号和右音频信号。

发明内容

[0013] ‘877 和 ‘458 申请描述了合成听觉场景技术,该技术涉及现有技术的传输带宽问题。根据 ‘877 申请,对应于位于相对于收听者不同位置的多个音频源的听觉场景,是使用两个或多个不同的听觉场景参数(例如空间线索,诸如声道间电平差(ICLD)值,声道间时间时延(ICTD)值,和/或与头相关的传递函数(HRTF))从单个组合(例如单声道)的音频信号合成的。这样,在上述基于 PC 的会议的情形下,可实现一种解决方案,其中每一参加者的 PC 只接收对应于来自所有参加者(加不同的听觉场景参数)的单声道音频源信号的单个的音频信号。

[0014] ‘877 申请中所述的技术基于这样的假设,对于来自特定音频源的源信号的能量单声道音频信号中所有其他源信号能量占优势的那些频率子频带,就收听者的观点来看,单音频信号能够如同其只对应于该特定音频源那样被处理。根据这一技术的实现,不同的听觉场景参数集合(每一个对应于一特定的音频源)施加到单声道音频信号中不同频率子带,以合成听觉场景。

[0015] ‘877 申请中所述的技术从一个单声道音频信号与两个或多个不同听觉场景参数的集合产生听觉场景。‘877 申请描述了单声道音频信号及其对应的听觉场景参数集合是如何产生的。用于产生单声道音频信号及其对应的听觉场景参数集合的该技术在说明书中

称为双耳线索编码 (BCC)。BCC 技术与在 ‘877 和 ‘458 申请中所述的空间线索 (PCSC) 技术的感知编码相同。

[0016] 根据 ‘458 申请,采用 BCC 技术以产生组合的 (例如单声道) 音频信号,其中在组合的音频信号中嵌入不同的听觉场景参数集合,使得结果的 BCC 信号能够由基于 BCC 的解码器或传统的 (即以以前的或非 BCC) 接收器处理。当通过基于 BCC 的解码器处理时,基于 BCC 的解码器抽取嵌入的听觉场景参数,并采用 ‘877 申请的听觉场景合成技术产生双耳 (或较高) 的信号。听觉场景参数嵌入在 BCC 信号中,使得对于传统的接收器是透明的,该接收器处理 BCC 信号如同传统的 (或单声道) 音频信号那样。这样,通过基于 BCC 的解码器在 ‘458 申请中所描述的技术支持 ‘877 申请的 BCC 处理,同时提供向后兼容性,以使 BCC 信号能够由传统的接收器按传统方式处理。

[0017] ‘877 和 ‘458 申请中所述的 BCC 技术,通过在 BCC 编码器把双耳输入信号 (例如左和右音频声道) 转换为单个的单声道音频声道,以及与单声道信号平行传输 (或频带内或频带外) 的双耳线索编码 (BCC) 参数流,有效地降低了传输带宽的需求。例如单声道信号能够以大约对应的两声道立体声信号所需的大约 50-80% 位速率传输。对于 BCC 参数附加的位速率只是几千位 / 秒 (即大于比编码音频声道小的量级)。在 BCC 解码器,从接收的单声道信号和 BCC 参数合成双耳信号的左 和右声道。

[0018] 双耳信号的相干性与感觉的音频源宽度有关。音频源越宽,结果的双耳信号左和右声道之间的相干性越低。例如,对应于散布在演奏厅舞台上的管弦乐的双耳信号的相干性一般比对应于单个的小提琴独奏的双耳信号相干性低。一般来说,通常感觉较低相干性的音频信号在听觉空间散布得更多。

[0019] ‘877 和 ‘458 申请的 BCC 技术产生双耳信号,其中左和右声道之间的相干性接近最大可能的值 1。如果最初的双耳输入信号有小于最大的相干性,则 BCC 解码器将不再生有相同相干性的立体声信号。其结果是听觉图像常常以产生太窄的图像而出错,这产生太 “干” 的声学印象。

[0020] 特别地,左和右输出声道将具有高的相干性,因为它们是从同一单声道信号通过在听觉临界频带慢变化电平修改产生的。把听觉范围划分为离散数目的音频子带的临界频带模型,心理声学中用来解释听觉系统的谱融合。对于耳机回放,左和右输出声道分别是左和右耳输入信号。如果耳信号具有高相干性,则在信号中获得的听觉对象将被感觉为非常 “局部化”,且它们在听觉空间图像中只有很小的散布。对于扬声器回放,扬声器信号只是直接确定耳信号,因为必须考虑从左扬声器到右耳以及从右扬声器到左耳的串音。此外,房间的反射对于感知的听觉图像也能够起到重要作用。然而,对于扬声器回放,类似于耳机回放,高相干信号的听觉图像是非常窄且局部的。

[0021] 根据 ‘437 申请,‘877 和 ‘437 申请的 BCC 技术延伸到包含基于输入音频信号相干性的 BCC 参数。相干性参数从 BCC 编码器与其他 BCC 参数与编码的单声道信号平行传送到 BCC 解码器。BCC 解码器采用相干性参数与其他 BCC 参数组合,以合成一种带有听觉对象的听觉场景 (例如双耳信号的左和右声道),这些对象的感知宽度更精确地匹配产生最初输入到 BCC 编码器的音频信号的听觉对象的宽度。

[0022] 与由 ‘877 和 ‘437 申请的 BCC 技术产生听觉对象的窄图像宽度相关的问题,是对不精确估计听觉空间线索 (即 BCC 参数) 的敏感性。特别是使用耳机回放,应当处于空间

中稳定位置的听觉对象趋向随机运动。无意识在附近的运动的对象的感知可能是令人讨厌的,并实质上降低了感知的音频质量。当采用‘437 申请的实施例时,这一问题如果不是完全但基本上消失了。

[0023] ‘437 申请的基于相干的技术在相对高的频率比在相对低的频率趋向工作得更好。根据本发明一定的实施例,‘437 申请的基于相干的技术对于一个或多个——并可能所有的——频率子带由回响技术代替。在一个混合的实施例中,对于低频(例如低于规定的(例如凭经验确定的)阈值频率的频率子带)实现了回响技术,而对于高频(例如大于阈值频率的频率子带)实现‘437 申请的基于相干的技术。

[0024] 在一个实施例中,本发明是用于合成听觉场景的一种方法。至少一个输入声道被处理以产生两个或多个被处理的输入信号,且至少一个声道被滤波以产生两个或多个漫射的信号。两个或多个漫射信号与两个或多个被处理的输入信号组合,以为听觉场景产生多个输出声道。

[0025] 在另一实施例中,本发明用于合成听觉场景的一种设备。该设备包括至少一个时域到频域(TD-FD)转换器与多个滤波器的一种配置,其中该配置适于从至少一个 TD 输入声道产生两个或多个被处理的 FD 输入信号和两个或多个漫射的 FD 信号。该设备还具有(a)两个或多个组合器,其适于组合两个或多个漫射 FD 信号与两个或多个被处理的 FD 输入信号,以产生多个合成的 FD 信号,以及(b)两个或多个频域到时域(TD-FD)转换器,其适于把合成的 FD 信号转换为用于听觉场景的多个 TD 输出声道。

附图说明

[0026] 从以下详细的说明,所附的权利要求,及伴随的图示,本发明的其他方式,特征和优点将更为充分地表现出来,其中:

[0027] 图 1 示出传统的双耳信号合成器的高级框图,其把单音频源信号(例如单声道信号)转换为双耳信号的左和右音频信号;

[0028] 图 2 示出传统的听觉场景合成器的高级框图,其把多个频源信号(例如多个单声道信号)转换为单个组合双耳信号的左和右音频信号;

[0029] 图 3 示出进行双耳线索编码的一种音频处理系统的框图;

[0030] 图 4 示出根据‘437 申请的一个实施例图 3 对应于相干度量产生的 BCC 分析器的处理部分的框图;

[0031] 图 5 示出由图 3 的 BCC 合成器的一个实施例进行的音频处理的框图,使用基于相干音频合成把单个组合声道转换为两个或多个合成的音频输出声道;

[0032] 图 6(A)-(E) 表示有不同线索编码的信号感知;

[0033] 图 7 示出由图 3 的 BCC 合成器进行的音频处理的框图,根据本发明的一个实施例使用基于回响的音频合成,把单个组合声道转换为(至少)两个合成的音频输出声道;

[0034] 图 8-10 表示一示例性五声道音频系统;

[0035] 图 11 和 12 图示出后期混响(late reverberation)滤波与 DFT 变换的定时;

[0036] 图 13 示出由图 3 的 BCC 合成器进行的音频处理的一框图,根据本发明另一实施例使用基于回响的音频合成,把单个组合声道转换为两个合成的音频示出声道,其中 LR 处理在频率域中实现。

[0037] 具体实施方式

[0038] 基于 BCC 的音频处理

[0039] 图 3 示出进行双耳线索编码 (BCC) 的音频处理系统 300 的框图。BCC 系统 300 有一个 BCC 编码器 302, 其接收 C 个音频输入声道 308, 其来自例如分布在音乐厅内不同位置的 C 个不同的麦克风 306 的每一个。BCC 编码器 302 有一个下混频器 (downmixer) 310, 其把 C 个音频输入声道转换为 (例如求平均) 一个或多个但少于 C 个组合的声道 312。此外, BCC 编码器 302 具有一 BCC 分析器 314, 其对于 C 个输入声道产生 BCC 线索代码数据流 316。

[0040] 在一种可能的实现中, 对于每一输入声道 BCC 线索代码包含声道之间电平差 (ICLD), 声道之间的时间差 (ICTD), 以及声道之间的相关 (ICC) 数据。BCC 分析器 314 最好进行模拟 ‘877 ‘458 申请中所述的基于频带的处理, 以便对于音频输入声道的一个或多个不同的频率子带产生 ICLD 和 ICTD 数据。此外, BCC 分析器 314 最好对每一频率子带产生相干度量作为 ICC 数据。这些相干度量在本说明书的下节中更详细描述。

[0041] BCC 编码器 302 向 BCC 系统 300 的 BCC 解码器 304 传送一个或多个组合声道 312 与 BCC 线索代码数据流 316 (例如对于组合声道频带内或频带外侧的信息)。BCC 解码器 304 具有辅助信息处理器 318, 其处理数据流 316 以便恢复 BCC 线索代码 320 (例如 ICLD, ICTD, 及 ICC 数据)。BCC 解码器 304 还具有一个 BCC 合成器 322, 其使用恢复的 BCC 线索代码 320 从一个或多个组合的声道 312 合成 C 个音频输出声道 324, 用于分别通过 C 个扬声器 326 播放。

[0042] 从 BCC 编码器 302 到 BCC 解码器 304 的数据传输的定义将依赖于音频处理系统 300 具体的应用。例如, 诸如音乐会现场广播等某些应用中, 传输可能涉及用于在远方立即回放的数据的实时传输。在其他应用中, “传输” 可能涉及向 CD 或其他适当的存储介质供后来 (即非实时) 回放的数据的存储。当然, 其他的应用也是可能的。

[0043] 在音频处理系统 300 的一种可能的应用中, BCC 编码器 302 把传统的 5.1 环绕声 (即五个通常的音频声道 + 一个低频效果 (LFE) 声道, 也称为亚低音声道) 的六个音频输入声道, 转换为单个的组合声道 312 及对应的 BCC 线索代码 316, 且 BCC 编码器 304 从单个的组合声道 312 和 BCC 线索代码 316, 产生合成的 5.1 环绕声 (即五个合成的普通的音频声道 + 一个合成的 LFE 声道)。包含 7.1 环绕声或 10.2 环绕声的许多其他应用也可能的。

[0044] 此外, 虽然 C 个输入声道可下混频为单个的组合声道 312, 但在另一实现中, C 个输入声道可下混频为两个或多个不同的组合声道, 这取决于具体的音频处理应用。某些应用中, 当下混频产生两个组合声道时, 可使用传统的立体声音频传输机制传送组合的声道数据。这样又可提供向后兼容性, 其中使用传统 (即非基于 BCC) 的立体声解码器回放两个 BCC 组合声道。当产生单个的 BCC 组合声道时, 可对单声道解码器提供类似的向后兼容性。

[0045] 虽然 BCC 系统 300 可有与音频输出声道相同数目的音频输入声道, 但在另一实施例中, 输入声道的数目可或大于或小于输出声道的数目, 这取决于具体的应用。

[0046] 取决于具体的实现, 由图 3 的 BCC 编码器 302 与 BCC 解码器 304 接收和产生的各种信号, 可以是任何包含全模拟或全数字的模拟和 / 或数字信号的适当组合。虽然图 3 中未示出, 但业内专业人员将理解, 一个或多个组合的声道 312 与 BCC 线索数据流 316 可进一步通过 BCC 编码器 302 被编码, 并对应地通过 BCC 解码器 304 例如基于某种适当压缩方案 (例如 ADPCM) 被解码, 以进一步降低传输数据量。

[0047] 相干估计

[0048] 图 4 示出根据 ‘437 申请的一个实施例图 3 对应于相干度量产生的 BCC 分析器 314 的处理部分的框图。如图 4 中所示, BCC 分析器 314 包括两个时间 - 频率 (TF) 变换块 402 和 404, 其采用适当的变换, 诸如长度 1024 的短时离散付立叶变换 (DFT), 把左和右输入音频声道 L 和 R 从时域分别转换到频域。每一变换块产生对应于输入音频声道不同频率子带的数个输出。相干估计器 406 特征体现了每一个不同的考虑的临界带 (以下标记为子带) 的相干。业内专业人员将理解, 在最优的基于 DFT 的实现中, 看作为一个临界带的数个 DFT 系数从临界带变为有较低频率临界带的临界带, 其一般比高频率临界带具有较少系数。

[0049] 在一种实现中, 估计每一 DFT 系数的相干。左声道 DFT 谱的谱分量 K_L 的实部和虚部可分别标记为 $\text{Re}\{K_L\}$ 和 $\text{Im}\{K_L\}$, 对于右声道是类似的。这种情形下, 对于左和右声道的功率估计 P_{LL} 和 P_{RR} 可分别由以下方程式 (1) 和 (2) 表示:

$$[0050] \quad P_{LL} = (1-\alpha)P_{LL} + \alpha (\text{Re}^2\{K_L\} + \text{Im}^2\{K_L\}) \quad (1)$$

$$[0051] \quad P_{RR} = (1-\alpha)P_{RR} + \alpha (\text{Re}^2\{K_R\} + \text{Im}^2\{K_R\}) \quad (2)$$

[0052] 实和虚交叉项 $P_{LR, \text{Re}}$ 和 $P_{LR, \text{Im}}$ 分别由以下方程式 (3) 和 (4) 给出:

$$[0053] \quad P_{LR, \text{Re}} = (1-\alpha)P_{LR} + \alpha (\text{Re}\{K_L\} \text{Re}\{K_R\} - \text{Im}\{K_L\} \text{Im}\{K_R\}) \quad (3)$$

$$[0054] \quad P_{LR, \text{Im}} = (1-\alpha)P_{LR} + \alpha (\text{Re}\{K_L\} \text{Im}\{K_R\} + \text{Im}\{K_L\} \text{Re}\{K_R\}) \quad (4)$$

[0055] 因子 α 确定了估计窗口持续时间, 并对于 32kHz 音频采样率和 512 采样的帧位移能够选择为 $\alpha = 0.1$ 。如从方程式 (1)-(4) 所推导, 对于子带的相干估计 γ 由以下方程式 (5) 给出:

$$[0056] \quad \gamma = \sqrt{(P_{LR, \text{Re}}^2 + P_{LR, \text{Im}}^2) / (P_{LL} P_{RR})} \quad (5)$$

[0057] 如上所述, 在每一临界带上相干估计器 406 对系数相干估计 γ 求平均。对于这求平均, 最好在求平均之前向子带相干估计施加一加权函数。可以使加权与由方程式 (1) 和 (2) 给出的功率估计成比例。对于包含谱分量 $n1, n1+1, \dots, n2$ 的一个临界带 p , 求平均的加权相干 γ_p 可使用以下方程式 (6) 计算:

$$[0058] \quad \bar{\gamma}_p = \frac{\sum_{n=n1}^{n2} \{(P_{LL}(n) + P_{RR}(n))\gamma(n)\}}{\sum_{n=n1}^{n2} \{(P_{LL}(n) + P_{RR}(n))\}} \quad (6)$$

[0059] 其中 $P_{LL}(n)$, $P_{RR}(n)$ 和 $\gamma(n)$ 为对于分别由方程式 (1), (2) 和 (6) 给出的谱系数 n 的左声道功率, 右声道功率, 与相干估计。注意, 方程式 (1)-(6) 对于每一谱系数 n 适用。

[0060] 在图 3 的 BCC 编码器 302 的一种可能的实现中, 对于包含在传输到 BCC 解码器 304 的 BCC 参数流, 通过 BCC 分析器 314 产生对于不同临界带的平均加权相干估计 γ_p 。

[0061] 基于相干的音频合成

[0062] 图 5 示出由图 3 的 BCC 合成器 322 的一实施例进行的音频处理的框图, 使用基于相干的音频合成把单个组合声道 312($S(n)$) 转换为 C 个合成的音频输出声道 324 ($\hat{x}_1(n), \hat{x}_2(n), \dots, \hat{x}_C(n)$)。具体来说, BCC 合成器 322 有一个听觉滤波器组 (AFB) 块 502, 其进行时间 - 频率 (TF) 变换 (例如快速付立叶变换 (FFT)), 把时域组合声道 312 转换为 C 个对应的频域信号 504 的拷贝 ($\tilde{s}(k)$)。

[0063] 频域信号 504 的每一拷贝,基于从由图 3 的辅助信息处理器 318 恢复的对应的声道间时间差 (ICTD) 数据推导的时延值 ($d_i(k)$),在对应的时延块 506 被时延。每一结果的时延信号 508 通过对应的乘法器 510,基于辅助信息处理器 318 恢复的对应的声道电平差 (ICLD) 数据推导的换算 (即增益) 因子 ($a_i(k)$) 被换算。

[0064] 结果的换算信号 512 施加到相干处理器 514,其基于由辅助信息处理器 318 恢复的 ICC 相干数据施加相干处理。以产生 C 个合成的频域信号 516 ($\tilde{x}_1(n), \tilde{x}_2(n), \dots, \tilde{x}_C(n)$),每一输出声道一个。然后每一合成的频域信号 516 施加到对应的逆 AFB (IAFB) 块 518,以产生不同的时域输出声道 324 ($\hat{x}_i(n)$)。

[0065] 在一优选的实现中,每一时延块 506,每一乘法器 510 及相干处理器 514 的处理是基于频带的,其中潜在在不同的时延值、换算因子、与相干度量施加到频域信号的每一不同拷贝的每一不同频率子带。对每一子带给定估计的相干,量值在子带内按一频率的函数变化。另一可能性是按估计的相干函数在分割中改变作为频率函数的相位。在一优选实现中,改变相位以便作为子带内频率的函数施加不同的时延或组时延。而且,最好这样进行量值和 / 或时延 (或组时延) 变化,使得在每一临界带中修改的平均值为零。其结果是,在子带内的 ICLD 和 ICTD 不会被相干合成所改变。

[0066] 在一优选的实现中,引入的量值的幅度 g (或变化) 或相位的变化基于左和右声道估计的相干被控制。对于较小的相干,增益 g 应当作为相干 γ 的适当的函数 $f(\gamma)$ 被正确地映射。一般来说,如果相干大 (例如接近最大可能的值 +1),则输入听觉场景中对象狭窄。这种情形下,增益 g 应当小 (例如接近最小可能值 0),使得在子带内实际上没有量值或相位修改。另一方面,如果相干小 (例如接近最小可能值 0),则输入听觉场景中的对象宽。这种情形下,增益 g 应当大,使得有明显的量值和 / 或相位修改,结果是在修改的子带信号之间有低相干。

[0067] 对于一具体临界带量值 g 的适当的映射函数 $f(\gamma)$ 由以下方程式 (7) 给出:

$$[0068] \quad g = 5(1 - \gamma) \quad (7)$$

[0069] 其中 γ 是对于对应的临界带估计的相干,其作为 BCC 参数流的一部分传输到 BCC 解码器 304。根据这一线性映射函数,当估计的相干 γ 为 1 时增益 g 为 0,并当 $\gamma = -1$ 时 $g = 5$ 。在另一实施例,增益 g 可以是相干的非线性函数。

[0070] 虽然已经在基于伪随机序列修改加权因子 w_L 和 w_R 的情形下描述了基于相干的音频合成,但该技术不限于此。一般来说,基于相干的音频合成适用于较大 (例如临界) 带的子带之间任何感知的空间线索的修改。修改的函数不限于随机序列。例如,修改函数可基于正弦函数,其中 (方程式 (9) 的) ICLD 在子带内作为频率的函数以正弦方式变化。在某些实现中,正弦波的周期从临界带向临界带作为对应的临界带宽度的函数变化 (例如在每一临界带内对应的正弦波的一个或多个全周期)。在其他实现中,正弦波的周期在整个频率范围上不变。在这两种实现中,正弦修改函数最好在临界带之间连续。

[0071] 修改函数的另一例子是锯齿形或三角形函数,它们在正最大值和对应的负最小值之间线性地斜坡上升和下降。这里也与实现相关,相关函数的周期可从临界频带到临界频带变化,或在整个频率范围不变,但在任何情形下在临界频带之间最好是连续的。

[0072] 虽然已经在随机,正弦的和三角形函数情形下描述了基于相干的音频合成,但在

每一临界频带内修改加权因子的其他函数也是可能的。如同正弦和三角形函数那样,这些其他修改函数可能但并非必须在临界频带之间是连续的。

[0073] 根据上述基于相干的音频合成的实施例,通过在音频信号的临界频带子频带之间引入修改水平差实现空间表现力。另外或加之,可采用基于修改的音频合成以修改时间差,作为真实的感知空间线索。具体来说,可如下对时间差采用类似于上述对于水平差生成听觉对象的较宽空间图像的技术。

[0074] 如‘877 和‘458 申请中所定义,两个音频声道之间的子频带 s 的时间差标记为 τ_s 。根据基于相干的音频合成的一定的实现,可引入时延位移 d_s 和增益因子 g_c ,以根据以下方程式 (8) 对于子频带 s 产生修改的时间差 τ_s' 。

$$[0075] \quad \tau_s' = g_c d_s + \tau_s \quad (8)$$

[0076] 时延位移 d_s 对于每一子频带最好对时间不变,但在子频带之间有变化,并能够选择为零平均随机序列或较平滑的函数,在每一临界频带中最好具有零平均值。如同方程式 (9) 的增益因子 g_c ,相同的增益因子 g_c 可用于落入每一临界频带 c 内的所有的子频带 n ,但增益因子可从临界频带到临界频带变化。增益因子 g_c 是使用映射函数从相干估计推导的,这函数最好与方程式 (7) 的线性映射函数成正比。于是, $g_c = ag$,其中常数值 a 通过实验调节确定。在另一些实施例中,增益 g_c 可以是相干的非线性函数。BCC 合成器 322 采用修改的时间差 τ_s' 代替原来的时间差 τ_s 。为了增加听觉对象的图像宽度,可即采用水平差又采用时间差修改。

[0077] 虽然已经在产生立体声音频场景的左和右频道的情形下描述了基于相干的处理,但该技术可扩展到任意数目的合成输出声道。

[0078] 基于回响的音频合成

[0079] 定义,符号与变量

[0080] 对于有时间下标 k 的两音频频道的对应的频域输入子频带信号 $\tilde{x}_1(k)$ 和 $\tilde{x}_2(k)$,以下度量用于 ICLD, ICTD 及 ICC:

[0081] $\circ$ ICLD (dB) :

$$[0082] \quad \Delta L_{12}(k) = 10 \log_{10} \left(\frac{p_{\tilde{x}_2}(k)}{p_{\tilde{x}_1}(k)} \right) \quad (9)$$

[0083] 其中 $p_{\tilde{x}_1}(k)$ 和 $p_{\tilde{x}_2}(k)$ 分别是信号 $\tilde{x}_1(k)$ 和 $\tilde{x}_2(k)$ 的短时间估计。

[0084] $\circ$ ICTD (采样) :

$$[0085] \quad \tau_{12}(k) = \arg \max_d \{ \Phi_{12}(d, k) \} \quad (10)$$

[0086] 使用短时间标准化互相关函数估计

$$[0087] \quad \Phi_{12}(d, k) = \frac{p_{\tilde{x}_1 \tilde{x}_2}(d, k)}{\sqrt{p_{\tilde{x}_1}(k - d_1) p_{\tilde{x}_2}(k - d_2)}} \quad (11)$$

[0088] 其中

$$[0089] \quad d_1 = \max \{-d, 0\}$$

$$[0090] \quad d_2 = \max \{d, 0\} \quad (12)$$

[0091] 并且 $p_{\tilde{x}_1 \tilde{x}_2}(d, k)$ 是 $\tilde{x}_1(k - d_1) \tilde{x}_2(k - d_2)$ 平均的短时间估计。

[0092] oICC :

$$[0093] \quad c_{12}(k) = \max_d |\Phi_{12}(d, k)| \quad (13)$$

[0094] 注意,考虑了标准化互相关的绝对值,且 $c_{12}(k)$ 的范围是 $[0, 1]$ 。不必考虑负值,因为 ICTD 包含了由 $c_{12}(k)$ 的符号表示的相位信息。

[0095] 本说明书中使用以下符号和变量:

[0096] * 卷积算子

[0097] i 音频声道下标

[0098] k 子频带信号时间下标 (也是 STFT 谱的时间下标)

[0099] C 编码器输入声道数,也是解码器输出声道数

[0100] $x_i(n)$ 时域编码器输入音频声道 (例如如图 3 的声道 308 之一) $\tilde{x}_i(k)$ $x_i(n)$ 的一个频域子频带信号 (例如如图 4 从 TF 变换 402 或 404 的输出之一)

[0101] $s(n)$ 被传输的时域组合声道 (例如如图 3 的合声道 312)

[0102] $\tilde{s}(k)$ $s(n)$ 的频域子频带信号 (例如如图 7 的信号 704)

[0103] $s_i(n)$ 去相关的时域组合声道 (例如如图 7 被滤波的声道 722)

[0104] $\tilde{s}_i(k)$ $s_i(n)$ 的频域子频带信号 (例如如图 7 对应的信号 726)

[0105] $\hat{x}_i(n)$ 时域编码器输出音频声道 (例如如图 3 的信号 324)

[0106] $\tilde{\hat{x}}_i(n)$ $\hat{x}_i(n)$ 的一个频域子频带信号 (例如如图 7 对应的信号 716)

[0107] $p_{\tilde{x}_i}(k)$ $\tilde{x}_i(k)$ 功率的短时估计

[0108] $h_i(n)$ 用于输出声道 i 的后期回响 (LR) 滤波器 (例如如图 7 的 LR 滤波器)

[0109] M LR 滤波器 $h_i(n)$ 的长度

[0110] ICLD 声道间电平差

[0111] ICTD 声道间时间差

[0112] ICC 声道间相关

[0113] $\Delta L_{1i}(k)$ 声道 1 与声道 i 之间的 ICLD

[0114] $\tau_{1i}(k)$ 声道 1 与声道 i 之间的 ICTD

[0115] $c_{1i}(k)$ 声道 1 与声道 i 之间的 ICC

[0116] STFT 短时付立叶变换

[0117] $X_i(j\omega)$ 信号的 STFT 谱

[0118] ICLD, ICTD 和 ICC 的感知

[0119] 图 6(A)-(E) 示出有不同线索代码的信号感知。具体来说,图 6(A) 表示一对扬声器信号之间的 ICLD 和 ICTD 如何确定听觉事件的感知角度。图 6(B) 示出一对耳机信号之间的 ICLD 和 ICTD 如何确定出现在头上部前方部分的听觉事件的位置。图 6(C) 示出,在扬声器信号之间的 ICC 降低时,听觉事件的范围如何增加 (从区域 1 到区域 3)。图 6(D) 示出,当左和右耳机信号之间的 ICC 降低时,听觉对象的范围如何增加 (从区域 1 到区域 3),直到两个不同的听觉事件在侧面出现 (区域 4)。图 6(E) 示出,对于多扬声器回放,在信号之间的 ICC 降低时,环绕收听者的听觉事件范围如何增加 (从区域 1 到区域 4)。

[0120] 相干的信号 ($ICC = 1$)

[0121] 图 6(A) 和 6(B) 示出,对于相干扬声器和耳机信号不同的 ICLD 和 ICTD 值感知的

听觉事件。振幅摇动 (panning) 是为表现用于扬声器与耳机回放的音频信号最常用的技术。当左和右扬声器或耳机信号相干 (即 $ICC = 1$), 有相同的电平 (即 $ICLD = 0$), 及没有时延 (即 $ICTD = 0$) 时, 在中心出现听觉事件, 如图 6(A) 与 6(B) 中区域 1 所示。注意, 对于图 6(A) 的扬声器回放, 听觉事件出现在两个扬声器之间, 而对于图 6(B) 的耳机回放, 出现在上半头的前方部分。

[0122] 通过增加一侧的电平, 例如右侧听觉事件向该侧移动, 如图 6(A) 和 6(B) 中区域 2 所示。在极限的情形下, 例如当只有左侧信号有效时, 听觉事件出现在左侧, 如图 6(A) 和 6(B) 中区域 3 所示。ICTD 可类似地用来控制听觉事件的位置。对于耳机回放, 为此可采用 ICTD。然而, 由于几个原因 ICTD 最好不用于扬声器回放。当收听者准确地在最有效点时, ICTD 值在自由场中最有效。在封闭的环境中由于反射, ICTD (范围小, 例如 $\pm 1\text{ms}$) 将对听觉事件感知的方向影响很小。

[0123] 部分相干信号 ($ICC < 1$)

[0124] 当通过一对扬声器同时发射相干 ($ICC = 1$) 宽带声音时, 感知到相对紧凑的听觉事件。当 ICC 在这些信号之间降低时, 听觉事件的范围如图 6(C) 所示从区域 1 到区域 3 增加。对于耳机回放, 如图 6(D) 所示可观察到类似的趋势。当由耳机发射两个等同的信号 (ICC) 时, 如同在区域 1 那样感知到相对紧凑的听觉事件。在耳机信号之间的 ICC 降低时, 如同在区域 2 和 3, 听觉事件的范围增加, 如同区域 4 直到在这些侧感知到两个不同的听觉事件。

[0125] 一般来说, ICLD 和 ICTD 确定感知的听觉事件的位置, 且 ICC 确定听觉事件的范围或漫射性。此外, 有收听的情形, 这时收听者不仅在一定距离感知到听觉事件, 而且通过漫射声音感知被环绕。这一现象被称为收听者包络。例如这种情形在音乐厅中出现, 其中后期混响从所有的方向到达收听者的耳朵。如图 6(E) 所示, 通过从分布在收听者周围的所有扬声器发射独立噪声信号能够引起类似的体验。在这种情景下, 在 ICC 与环绕收听者的听觉事件范围之间有一种关系, 如同区域 1 到 4。

[0126] 上述的感知可通过混合数个有低 ICC 的去相关音频声道产生。下节描述了用于产生这种效果的基于回响的技术。

[0127] 从单个组合声道产生漫射声音

[0128] 如前所述, 音乐厅是收听者感知漫射声音一种典型的情景。在后期混响期间, 声音从随机角度以随机强度到达耳朵, 使得两个耳朵输入信号之间的相关低。这给出用于通过对后期混响建模的滤波器, 对给定的组合音频声道 $s(n)$ 滤波, 产生数个去相关音频声道的动机。本说明书中结果的滤波声道还称为“漫射声道”。

[0129] 通过以下方程式 (14) 获得 C 个漫射声道 $s_i(n)$, ($1 \leq i \leq C$) :

$$[0130] \quad s_i(n) = h_i(n) * s(n) \quad (14)$$

[0131] 其中 * 标记卷积, 而 $h_i(n)$ 是对后期混响建模的滤波器。后期混响能够通过以下方程式 (15) 建模:

[0132]

$$h_i(n) = \begin{cases} n_i(n) \left(1 - \frac{1}{f_s T}\right)^n, & 0 \leq n < M, \\ 0, & \text{其它} \end{cases} \quad (15)$$

[0133] 其中而 $n_i(n)$ ($1 \leq i \leq C$) 是独立驻定的高斯白噪声信号, T 是以秒计脉冲响应的以秒计的指数衰减的时间常数, f_s 是采样频率, 而 M 是采样脉冲响应的长度。选择一指数衰减, 是因为后期混响的强度一般在时间上是指数衰减的。

[0134] 许多音乐厅的回响衰减范围是 1.5 到 3.5 秒。为了使漫射音频声道充分独立以便产生音乐厅录音的漫射, T 是这样选择的, 使得 $h_i(n)$ 的回响时间在相同的范围。这是 $T = 0.4$ 秒的情形 (结果是回响时间大约 2.8 秒)。

[0135] 通过作为 $s(n)$ 和 $s_i(n)$ ($1 \leq i \leq C$) 的加权和计算每一耳机或扬声器信号声道, 可产生有所需漫射的信号 (当使用 $s_i(n)$ 时带有类似于音乐厅最大的漫射)。如下节所示, BCC 合成最好在每一子带分别采用这种处理。

[0136] 基于回响的示例性音频合成器

[0137] 图 7 示出由图 3 的 BCC 合成器 322 进行的音频处理的框图, 其根据本发明的一实施例, 使用基于回响的音频合成, 把信号组合声道 312 ($s(n)$) 转换为 (至少) 两个合成的音频输出声道 324 ($\hat{x}_1(n), \hat{x}_2(n), \dots$)。

[0138] 如图 7 所示并类似于图 5 的 BCC 合成器 322 中的处理, AFB 块 702 把时间域组合声道 312 转换为对应的频域信号 704 ($\tilde{s}(k)$) 的两个拷贝。频域信号 704 的每一个拷贝, 基于从由图 3 的辅助信息处理器 318 恢复的对应的声道间时间差 (ICTD) 数据推导的时延值 ($d_i(k)$), 在对应的时延块 706 被时延。每一结果的时延信号 708, 基于从由辅助信息处理器 318 恢复的线索代码数据推导的换算因子, 由对应的乘法器 710 换算。这些换算因子的推导在以下进一步详细说明。结果的换算时延信号 712 施加到求和结点 714。

[0139] 除了施加到 AFB 块 702 之外, 组合声道 312 的拷贝还施加到后期混响 (LR) 处理器 720。在某些实现中, LR 处理器产生一信号, 其类似于假如组合声道 312 在该音乐厅中回放在音乐厅引起的后期混响。此外, LR 处理器可用来产生对应于音乐厅中不同位置的后期混响, 使得它们的输出信号是去相关的。这种情形下, 组合声道 312 和漫射 LR 输出声道 722 ($s_1(n), s_2(n)$) 将具有高度的独立性 (即 ICC 值接近零)。

[0140] 如上节使用方程式 (14) 和 (15) 所述, 漫射 LR 声道 722 可通过对组合的信号 312 滤波产生。另外, LR 处理器可基于任何其他适当的回响技术实现, 诸如在以下文献所述, M. R. Schroeder, “自然发声人工回响,” J. Aud. Eng. Soc. vol. 10, no. 3, pp. 219-223, 1962, 以及 W. G. Gardner, Applications of Digital Signal Processing to Audio and Acoustics, Kluwer Academic Publishing, Norwell, MA, USA, 1998。一般来说, 优选的 LR 滤波器是具有基本上随机的频率响应的滤波器, 其谱包络是基本上是平缓的。

[0141] 漫射 LR 声道 722 用于 AFB 块 724, 这把时域 LR 声道 722 转换为频域 LR 信号 726 ($\tilde{s}_1(k), \tilde{s}_2(k)$)。AFB 块 702 和 724 最好是可逆的滤波器组, 其带宽等于或正比于听觉系统的临界带宽。对于输入信号 $s(n)$, $s_1(n)$ 和 $s_2(n)$ 的每一子频带信号分别标记为 $\tilde{s}(k)$, $\tilde{s}_1(k)$ 或 $\tilde{s}_2(k)$ 。不同的时间下标 k 用于分解的信号, 而不是输入声道时间下标 n , 因为子带信号通常以比原来的输入声道低的采样频率表示。

[0142] 乘法器 728 使频域 LR 信号 726 乘以从由辅助信息处理器 318 恢复的线索代码数据推导的换算因子 ($b_1(k)$)。这些换算因子的推导在以下进一步详细说明。结果的换算 LR 信号 730 施加到求和结点 714。

[0143] 求和结点 714 把来自乘法器 728 的换算的 LR 信号加到来自乘法器 710 的对应的换算的、时延的信号 712, 以对于不同的输出声道产生频域信号 716 ($\tilde{x}_1(k), \tilde{x}_2(k)$)。在求和结点 714 产生的子带信号 716 由以下方程式 (16) 给出:

$$\begin{aligned} [0144] \quad \tilde{x}_1(k) &= a_1 \tilde{s}(k - d_1) + b_1 \tilde{s}_1(k) \\ [0145] \quad \tilde{x}_2(k) &= a_2 \tilde{s}(k - d_2) + b_2 \tilde{s}_2(k) \end{aligned} \quad (16)$$

[0146] 其中换算因子 (a_1, a_2, b_1, b_2) 与时延 (d_1, d_2) 确定为所需的 ICLD $\Delta L_{12}(k)$, ICTD $\tau_{12}(k)$, 与 ICC $c_{12}(k)$ 的函数。(为了简洁省略了换算因子与时延的时间下标)。信号 $\tilde{x}_1(k), \tilde{x}_2(k)$ 对所有子带产生。虽然图 7 的实施例依赖于求和结点组合换算的 LR 信号与对应的换算, 时延信号, 在另一实施例中, 可使用与求和结点不同的组合器组合信号。可替代的组合器的例子包括进行加权求和, 量值求和或最大值选择的组合器。

[0147] ICTD $\tau_{12}(k)$ 通过对 $\tilde{s}(k)$ 施以不同的时延被合成。这些时延通过方程式 (10) 以 $d = \tau_{12}(n)$ 计算。为了使输出子带信号具有等于方程式 (9) 的 $\Delta L_{12}(k)$ 的 ICLD, 换算因子 (a_1, a_2, b_1, b_2) 应当满足以下方程式 (17):

$$[0148] \quad \frac{a_1^2 p_{\tilde{s}}(k) + b_1^2 p_{\tilde{s}_1}(k)}{a_2^2 p_{\tilde{s}}(k) + b_2^2 p_{\tilde{s}_2}(k)} = 10^{\frac{\Delta L_{12}(k)}{10}} \quad (17)$$

[0149] 其中 $p_{\tilde{s}}(k)$, $p_{\tilde{s}_1}(k)$, $p_{\tilde{s}_2}(k)$ 分别是子带信号 $\tilde{s}(k)$, $\tilde{s}_1(k)$, $\tilde{s}_2(k)$ 的短时功率估计。

[0150] 为了使输出子带信号具有方程式 (13) 的 ICC $c_{12}(k)$, 换算因子 (a_1, a_2, b_1, b_2) 应当满足以下方程式 (18):

$$[0151] \quad \frac{(a_1^2 + a_2^2) p_{\tilde{s}}(k)}{(a_1^2 p_{\tilde{s}}(k) + b_1^2 p_{\tilde{s}_1}(k)) (a_2^2 p_{\tilde{s}}(k) + b_2^2 p_{\tilde{s}_2}(k))} = c_{12}(k) \quad (18) \text{ 假 设}$$

$\tilde{s}(k)$, $\tilde{s}_1(k)$, $\tilde{s}_2(k)$ 是独立的。

[0152] 每一 IAFB 块 718 把一组频域信号 716 转换为用于输出声道之一的时域声道 324。由于每一 LR 处理器 720 能够用于对从音乐厅中不同方向发出的后期混响建模, 可对不同的后期混响建模用于图 3 音频处理系统 300 不同的扬声器 326。

[0153] BCC 合成通常标称化其输出信号, 使得所有输出声道的功率和等于输入组合信号的功率。这对增益因子产生了另一方程式:

$$[0154] \quad (a_1^2 + a_2^2) p_{\tilde{s}}(k) + b_1^2 p_{\tilde{s}_1}(k) + b_2^2 p_{\tilde{s}_2}(k) = p_{\tilde{s}}(k) \quad (19)$$

[0155] 由于有四个增益因子和三个方程式, 在选择增益因子中仍然有一个自由度。这样附加的条件可以公式表示为:

$$[0156] \quad b_1^2 p_{\bar{3}}(k) = b_2^2 p_{\bar{2}}(k) \quad (20)$$

[0157] 方程式 (20) 意味着漫射音量在两个声道总是相同的。有几个为这样作的动机。首先,漫射音在音乐厅如同后期混响具有几乎与位置无关的电平那样出现(对于相对小的位移)。这样,两个声道之间的漫射音的电平差总是大约为 0dB。其次,这有美好的边效果,即当 $\Delta L_{12}(k)$ 很大时,只有漫射音混入紧弱的声道。这样,较强声道的声音被最小地修改,降低了长回旋的负面效果,诸如瞬时的时间散布。

[0158] 方程式 (17)-(20) 的非负解产生以下用于换算因子的方程式:

$$\begin{aligned}
 [0159] \quad a_1 &= \sqrt{\frac{10^{\frac{\Delta L_{12}(k)}{10}} + c_{12}(k) 10^{\frac{\Delta L_{12}(k)}{20}} - 1}{2 \left(10^{\frac{\Delta L_{12}(k)}{10}} + 1 \right)}} \\
 [0160] \quad a_2 &= \sqrt{\frac{-10^{\frac{\Delta L_{12}(k)}{10}} + c_{12}(k) 10^{\frac{\Delta L_{12}(k)}{20}} + 1}{2 \left(10^{\frac{\Delta L_{12}(k)}{10}} + 1 \right)}} \\
 [0161] \quad b_1 &= \sqrt{\frac{\left(10^{\frac{\Delta L_{12}(k)}{10}} + c_{12}(k) - 10^{\frac{\Delta L_{12}(k)}{20}} + 1 \right) p_{\bar{3}}(k)}{2 \left(10^{\frac{\Delta L_{12}(k)}{10}} + 1 \right) p_{\bar{3}}(k)}} \quad (21) \\
 [0162] \quad b_2 &= \sqrt{\frac{\left(10^{\frac{\Delta L_{12}(k)}{10}} + c_{12}(k) - 10^{\frac{\Delta L_{12}(k)}{20}} + 1 \right) p_{\bar{3}}(k)}{2 \left(10^{\frac{\Delta L_{12}(k)}{10}} + 1 \right) p_{\bar{3}}(k)}}
 \end{aligned}$$

[0163] 多声道 BCC 合成

[0164] 虽然图 7 所示的配置产生两个输出声道,但通过替代图 7 虚线块中所示的配置,该配置可扩展到任何更大数目的输出声道。注意,在本发明的这些实施例中,有一个 LR 处理器 720 用于每一输出声道。还要注意,在这些实施例中,每一 LR 处理器的实现是在时域中对组合的声道工作。

[0165] 图 8 表示一示例性五声道音频系统。只要在基准声道(例如声道号 1)与其它四声道中每一个之间定义 ICLD 和 IDTD 即可,其中 $\Delta L_{1i}(k)$ 与 $\tau_{1i}(k)$ 标记基准声道 1 与声道 i 之间的 ICLD 和 ICTD, $2 \leq i \leq 5$ 。

[0166] 与 ICLD 和 ICTD 相反,ICC 有更多的自由度。一般来说,ICC 在所有可能的输入声道对之间可具有不同的值。对于 C 个声道,有 $C(C-1)/2$ 个可能的声道对。例如,对于五声道,有十个声道对,如图 9 所示。

[0167] 给定组合信号 $s(n)$ 的子带 $\tilde{s}(k)$ 加 $C-1$ 漫射声道 $\tilde{s}_i(k)$ 的子带,其中 $(1 \leq i \leq C-1)$,

并假设漫射声道是独立的,能够产生C个子带信号,使得每一可能的声道对之间的ICC与在原始信号对应的子带中估计的ICC相同。然而,这种方案将涉及对每一子带在每一时间下标估计和传输 $C(C-1)/2$ 个ICC值,结果造成相对高的计算复杂性和相对高的位速率。

[0168] 对于每一子带,ICLD和ICTD确定子带中对应的信号分量的听觉事件被表现的方向。因而原则上,应当只添加一个ICC参数即可,该参数确定听觉事件范围或漫射。这样,在一实施例中,对于每一子带,在每一时间下标k,只估计对应于该子带中具有最大功率电平的两个声道的一个ICC值。这表示在图10中,其中在时刻k-1,声道对(3,4)对于一特定的子带具有最大功率电平,而在时刻k,声道对(1,2)对于同一子带具有最大的功率电平。一般来说,可对每一子带在每一时间段传输一个或多个ICC值。

[0169] 类似于两声道(例如立体声)的情形,作为组合信号与漫射音频声道的子带信号的加权和,计算多声道输出子带信号如下:

$$\begin{aligned} [0170] \quad \tilde{x}_1(k) &= a_1 \tilde{s}(k-d_1) + b_1 \tilde{s}_1(k) \\ [0171] \quad \tilde{x}_2(k) &= a_2 \tilde{s}(k-d_2) + b_2 \tilde{s}_2(k) \end{aligned} \quad (22)$$

[0172] —

$$[0173] \quad \tilde{x}_C(k) = a_C \tilde{s}(k-d_C) + b_C \tilde{s}_C(k)$$

[0174] 从ICTD如下确定时延:

$$[0175] \quad d_i = \begin{cases} -\min_{1 \leq l < C} \tau_{ll}(k) & i = 1 \\ \tau_{ll}(k) + d_1 & 2 \leq i \leq C \end{cases} \quad (23)$$

[0176] 在方程式(22)中需要2C个方程式确定2C个换算因子。以下的讨论描述导致这些方程式的条件。

[0177] oICLD:在声道对之间表示出类似于方程式(17)的C-1个方程式,使得输出子带信号具有所需的ICLD线索。

[0178] o用于两个最强的声道的ICC:表示出类似于两个最强的音频声道 i_1 和 i_2 之间的方程式(18)与(20)的两个方程式,使得(1)这些声道之间的ICC与编码器中估计的ICC相同,以及(2)在两个声道中的漫射音量分别相同。

[0179] o标称化:通过把方程式(19)扩展到C个声道获得如下的另一方程式:

$$[0180] \quad \sum_{i=1}^C a_i^2 p_{\tilde{s}}(k) + \sum_{i=1}^C b_i^2 p_{\tilde{s}_i}(k) = p_{\tilde{s}}(k) \quad (24)$$

[0181] o用于C-2个最弱的声道的ICC:选择对于最弱的C-2个声道($i \neq i_1 \wedge i \neq i_2$)漫射音对非漫射音之间的比例与对于次最强声道 i_2 相同,使得:

$$[0182] \quad \frac{b_i^2 p_{\tilde{s}_i}(k)}{a_i^2 p_{\tilde{s}}(k)} = \frac{b_{i_2}^2 p_{\tilde{s}_{i_2}}(k)}{a_{i_2}^2 p_{\tilde{s}}(k)} \quad (25)$$

[0183] 结果是对于全部2C个方程式得到另外的C-2个方程式。换算因子是所述2C个方程式的非负解。

[0184] 降低计算复杂性

[0185] 如上所述,为了重放自然发出的漫射声音,方程式(15)的脉冲响应应当长达几百

毫秒,结果造成高度计算复杂性。此外,如图 7 所示,对于每一 $h_i(t)$ ($1 \leq i \leq C$),BCC 合成需要附加的滤波器组。

[0186] 使用用于产生后期混响的人工回响算法并对 $s_i(t)$ 使用该结果,能够降低计算复杂性。另一可能性是通过采用降低计算复杂性的基于快速付立叶变换 (FFT) 的算法进行卷积。而另一可能性是在频域中进行方程式 (14) 的卷积,而不引入过量的时延。这种情形下,带有重叠窗口的相同的短时付立叶变换 (STFT) 可用于卷积也可用于 BCC 处理。这结果是卷积计算较低的计算复杂性,且对每一 $h_i(t)$ 不需要使用附加的滤波器组。该技术是对于单个组合信号 $s(t)$ 与一般脉冲响应 $h(t)$ 推导的。

[0187] STFT 对信号 $s(t)$ 的开窗口部分施加离散付立叶变换 (DFT)。开窗口以标记为窗口跳跃大小 N 的规则间隔施加。结果带有窗口位置下标 k 的开窗口信号为:

[0188]

$$s_k(t) = \begin{cases} w(t - kN)s(t), & kN \leq t \leq kN + W \\ 0, & \text{其它} \end{cases} \quad (26)$$

[0189] 其中 W 是窗口长度。Hann 窗口可使用长度 $W = 512$ 采样,且窗口跳跃大小为 $N = W/2$ 采样。可使用满足以下条件 (在以下假设) 其他窗口:

$$[0190] \quad s(t) = \sum_{k=-\infty}^{\infty} s_k(t) \quad (27)$$

[0191] 首先,考虑在频域中实现开窗口信号 $s_k(t)$ 的卷积的简单情形。图 11(A) 示出长度 M 的脉冲响应 $h(t)$ 的非零跨度。类似地,在图 11(B) 中示出 $s_k(t)$ 的非零跨度。容易验证, $h(t) * s_k(t)$ 具有如图 11(C) 所示的 $W+M-1$ 样本的非零跨度。

[0192] 图 12(A)-(C) 示出,长度 $W+M-1$ 的下标 DFT 在何时分别施加到信号 $h(t)$, $s_k(t)$ 和 $h(t) * s_k(t)$ 。图 12(A) 示出, $H(j\omega)$ 标记通过在时间下标 $t = 0$ 开始向 $h(t)$ 施加 DFT 获得的频谱。图 12(B) 和 12(C) 示出,通过在时间下标 $t = Kn$ 开始施加 DFT 分别从 $s_k(t)$ 和 $h(t) * s_k(t)$ 计算 $X_k(j\omega)$ 和 $Y_k(j\omega)$ 。容易证明, $Y_k(j\omega) = H(j\omega)X_k(j\omega)$ 。就是说,因为在信号 $h(t)$ 和 $s_k(t)$ 的末端的零,结果是圆周卷积通过等于线性卷积的谱乘积施加到信号。

[0193] 从卷积和方程式 (27) 的线性,得出:

$$[0194] \quad h(t) * s(t) = \sum_{k=-\infty}^{\infty} h(t) * s_k(t) \quad (28)$$

[0195] 这样,能够通过每一时间 t 计算乘积 $H(j\omega)X_k(j\omega)$ 并施加逆的 STFT (逆的 DFT 加重叠 / 加),在 STFT 的域中实现卷积。长度 $W+M-1$ (或更长) 的 DFT 应当以如图 12 所示零填充使用。所述的技术类似于推广的重叠 / 加卷积,可使用重叠窗口 (带有任何满足方程式 (27) 条件的窗口)。

[0196] 所述方法对于长脉冲响应 (例如 $M \gg W$) 是不实用,因为这时需要使用比 W 大得多的 DFT。以下,所述方法被扩展,使得只需要使用大小 $W+N-1$ 的 DFT。

[0197] 长度 $M = LN$ 的长脉冲响应 $h(t)$ 被划分为 L 个较短脉冲响应 $h_l(t)$, 其中:

[0198]

$$h_l(t) = \begin{cases} h(t + lN), & 0 \leq t < N \\ 0, & \text{其它} \end{cases} \quad (29)$$

[0199] 如果 $\text{mod}(M, N) \neq 0$, 则 $N - \text{mod}(M, N)$ 零被加到 $h(t)$ 的尾部。这时以 $h(t)$ 的卷积可写为如下较短卷积之和:

$$[0200] \quad h(t) * s(t) = \sum_{l=0}^{L-1} h_l(t) * s(t - lN) \quad (30)$$

[0201] 同时采样方程式 (29) 和 (30), 产生:

$$[0202] \quad h(t) * s(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{L-1} h_l(t) * s_k(t - lN) \quad (31)$$

[0203] 方程式 (31) 中作为 k 和 l 的函数的一个卷积的非零时间跨度 $h(t) * s_k(t - lN)$ 是 $(k+1)N \leq t < (k+1+1)N + W$ 。这样, 为获得其频谱 $\tilde{Y}_k(j\omega)$, DFT 施加到这一区间 (对应于 DFT 位置下标 $k+1$)。可以证明, $\tilde{Y}_k(j\omega) = \mathbf{H}_l(j\omega) \mathbf{X}_k(j\omega)$, 其中 $X_k(j\omega)$ 按以前以 $M = N$ 定义, 且 $H_l(j\omega)$ 类似于 $H(j\omega)$ 定义, 但是对于脉冲响应 $h_l(t)$ 。

[0204] 带有相同 DFT 位置下标 $i = k+1$ 的所有 $\tilde{Y}_k(j\omega)$ 的和如下:

$$[0205] \quad Y_i(j\omega) = \sum_{k+l=i} \tilde{Y}_{k+l}(j\omega) \quad (32)$$

$$[0206] \quad = \sum_{l=0}^{L-1} H_l(j\omega) X_{i-l}(j\omega)$$

[0207] 这样, 在 STFT 域中通过在每一谱下标 i 采用方程式 (32) 而获得 $Y_i(j\omega)$ 实现卷积 $h(t) * s_k(t)$ 。施加到 $Y_i(j\omega)$ 的逆 STFT (逆 DFT 加重叠/加) 如所希望等于卷积 $h(t) * s(t)$ 。

[0208] 注意, 与 $h(t)$ 的长度无关, 零填充量以 $N-1$ 为上界 (小于 STFT 窗口跳跃大小的一个采样)。如果需要, 可使用大于 $W+N-1$ 的 DFT (例如使用长度等于二的幂的 FFT)。

[0209] 如上所述, 低复杂性 BCC 合成能够在 STFT 域中工作。这种情形下, ICLD, ICTD 和 ICC 合成施加到表示谱分量的 STFT 箱的组, 其带宽等于或与临界频带的带宽成比例 (其中箱组标记为“分割”)。在这一系统中, 为了降低复杂性, 代替向方程式 (32) 施加逆 STFT, 方程式 (32) 的谱直接在频域中用作为漫射音。

[0210] 图 13 示出根据本发明的另一实施例, 由图 3 的 BCC 合成器 322 进行的音频处理的框图, 以便使用基于回响的音频合成把信号组合声道 312 ($s(t)$) 转换为两个合成的音频输出声道 324 ($\hat{x}_1(t), \hat{x}_2(t)$), 其中 LR 处理是在频域中实现的。具体来说, 如图 13 所示, AFB 块 1302 把时域组合声道 312 转换为对应的频域信号 1304 ($\tilde{s}(k)$) 的四个拷贝。频域信号 1304 的四个拷贝的两个施加到时延块 1306, 同时另两个拷贝施加到 LR 处理器 1320, 其频域 LR 输出信号 1326 施加到乘法器 1328。图 13 的 BCC 合成器其余的组件和处理类似于图 7 的 BCC 合成器。

[0211] 当在频域中实现 LR 滤波器诸如图 13 的 LR 滤波器 1320 时, 其有可能对不同的频率的子带使用不同滤波器长度, 例如在较高频率使用较短的滤波器。这可用以降低整个的计算复杂性。

[0212] 混合的实施例

[0213] 即使如图 13 当 LR 处理器在频域中实现时, BCC 合成器的计算复杂性仍然可能相当高。例如如果使用脉冲响应对后期混响建模, 则脉冲响应应当相当长, 以获得高质量的漫射

音。另一方面,‘437 申请基于相干的音频合成一般计算复杂性较小,并对高频提供良好的性能。这导致能够实现混合的音频处理系统,其对低频采用本发明基于回响的处理(例如大约 1-3kHz 以下的频率),同时对高频采用‘437 申请基于相干的处理(例如大约 1-3kHz 以上的频率),从而实现在整个频率范围提供良好性能同时降低了整个计算复杂性的系统。

[0214] 替代的实施例

[0215] 虽然已就基于回响的 BCC 处理并还依赖于 ICLD 和 IDTD 数据 的情形下描述了本发明,但本发明不限于此。理论上,本发明的 BCC 处理能够在没有 ICLD 和 / 或 ICTD 数据,带有或没有其它适当线索代码,诸如与头相关的传递函数相关联的代码下实现。

[0216] 如前所述,本发明能够在产生一个以上的“组合”声道的 BCC 编码的情形下实现。例如, BCC 编码能够用于 5.1 环绕声的六个输入声道,以产生两个组合声道:一个基于左和左后声道,一个基于右和右后声道。在一种可能的实现中,每一组合的声道还能够基于两个其他的 5.1 声道(即中心声道和 LFE 声道)。换言之,第一组合声道可基于左,后左,中心和 LFE 声道之和,而第二组合声道可基于右,后右,中心和 LFE 声道之和。这种情形下,可能有两组不同的 BCC 线索代码:一个用于产生第一组合声道的声道,一个用于产生第二组合声道的声道,使用 BCC 解码器有选择地向两个组合声道施加这些线索代码,一般在接收器产生合成的 5.1 环绕声。这一方案最好使两个组合声道能够作为传统的左和右声道在传统的立体声接收器上回放。

[0217] 注意,理论上当有多个“组合”声道时,一个或多个组合声道实际上基于单个的输入声道。例如, BCC 编码能够施加到 7.1 环绕声以产生 5.1 环绕信号及适当的 BCC 代码,其中例如 5.1 信号中的 LFE 声道可简单地是 7.1 信号中 LFE 声道的复制。

[0218] 还就两个或多个输出声道从一个或多个组合声道合成的音频合成技术的情形对本发明进行了描述,其中有一个滤波器用于每一不同的输出声道。在另一实施例中,能够使用少于 C 个 LR 滤波器合成 C 个输出声道。这能够通过使用一个或多个组合声道组合少于 C 个 LR 滤波器漫射声道输出而实现,以产生 C 个合成的输出声道。例如,得以产生一个或多个输出声道,而无需任何回响,或通过组合结果的漫射声道与一个或多个组合声道的不同的换算、时延版本,一个 LR 滤波器可用来产生两个或多个输出声道。

[0219] 另外,这能够通过采用先前对于一定的输出声道描述的回响技术实现。其他可适用于这种混合实现的基于相干的合成技术在以下文献中有述, E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, “Advances in parametric coding for high-quality audio,” Preprint 114th Convention Aud. Eng. Soc., March 2003, and Audio Subgroup, Parametric coding for High Quality Audio, ISO/IEC JTC1/SC29/WG11 MPEG2002/N5381, December 2002.

[0220] 虽然已经就传输声道的情形描述了图 3 中 BCC 编码器 302 和 BCC 解码器 304 之间的接口,但业内专业人员将理解,此外或替代地该接口可包含一存储介质。取决于具体的实现,传输声道可以是有线的或无线的,并可使用定制的或标准化的协议(例如 IP)。诸如 CD, DVD, 数字磁带录像机及固态存储器等介质可用作为存储器。此外,传输和 / 或存储器可以但不是必须包含声道编码。类似地,虽然已就数字音频系统的情形描述了本发明,但业内专业人员将理解,本发明还可在模拟音频系统诸如 AM 无线广播, FM 无线广播, 及模拟电视广播的音频部分的情形实现,其中每一个支持包含附加的带内低位速率传输声道。

[0221] 本发明能够对于许多不同的应用实现,诸如音乐复制,广播,及电话技术。例如,本发明可对于数字无线广播/TV/因特网(例如Web广播)广播,诸如Sirius Satellite Radio或XM。其他应用包括通过IP,PSTN的语音或其他语音网络,模拟广播及因特网无线广播。

[0222] 取决于具体的应用,可采用不同的技术把BCC参数集嵌入单声道音频信号以实现本发明的BCC信号。任何具体技术的可用性至少部分地取决于具体的用于BCC信号的传输/存储介质。例如,数字无线广播的协议通常支持包含附加的“增强”位(例如在数据包的标头部分),这通常被传统的接收器忽略。这些附加位可用来表示听觉场景参数集以产生BCC信号。一般来说,本发明可使用任何适当的用于音频信号水印的技术实现,其中对应于听觉场景参数集的数据嵌入到音频信号中,以形成BCC信号。例如,这些技术可能涉及隐藏在感知隐蔽曲线之下的数据,或隐藏在伪随机噪声中的数据。伪随机噪声可作为“舒服的噪声”被感知。和可使用类似于在用于带内传信的TDM(时分多路复用)传输的方法实现数据嵌入。另一可能的技术mu-law LSB位倒向,其中使用最低有效位传输数据。

[0223] 本发明的BCC编码器可用来把双耳信号的左和右音频声道转换为一个编码的单声道信号及对应的BCC参数流。类似地,本发明的BCC解码器能够用来基于编码的单声道信号与对应的BCC参数流产生合成的双耳信号的左和右音频声道。然而本发明不限于此。一般来说,本发明的BCC编码器可在转换M个输入音频声道为N个组合音频声道及一个或多个对应的BCC参数集的情形下实现,其中 $M > N$ 。类似地,本发明的BCC解码器可在从N个组合的声道与对应的BCC参数集合产生P个输出音频声道的情形实现,其中 $P > N$,且P可以与M相同或不同。

[0224] 虽然已就带有嵌入的听觉场景参数的单个组合(例如单声道)音频信号传输/存储的情形描述了本发明,但本发明还可对于其他声道数实现。例如,本发明可用来传输带有嵌入的听觉场景参数的两声道音频信号,其音频信号可以传统的两声道立体声接收器回放。这种情形下,BCC解码器可抽取并使用听觉场景参数以合成环绕声(例如基于5.1格式)。一般来说,本发明可用来从带有嵌入的听觉场景参数的N个音频声道产生M个音频声道,其中 $M > N$ 。

[0225] 虽然已在采用‘877及‘458申请的技术合成听觉场景的BCC解码器的情形下描述了本发明,但本发明还可在采用其他合成听觉场景技术的BCC解码器的情形下实现,其不必依赖于‘877及‘458申请的技术。

[0226] 本发明可作为基于电路的处理实现,其包括在单个集成电路上可能的实现。对于业内专业人员明显的是,电路元件的各种功能还可作为软件程序中的处理步骤实现。这种软件例如可在数字信号处理器、微控制器、或通用计算机中采用。

[0227] 本发明可以方法或实践这些方法的设备的形式实施。本发明还可以有形的介质程序代码的形式实施,诸如软盘,CD-ROM,硬盘,或任何其他机器可读存储介质,其中当程序代码加载到诸如计算机等机器并由其执行时,该机器就成为用于实践本发明的设备。本发明还可程序代码的形式实施,例如或者存储在存储介质中,加载到机器和/或由其执行,或者通过某种传输介质或载体传输,诸如通过电线或电缆传送,通过光纤,或通过电磁辐射,其中当程序代码加载到诸如计算机等机器并由其执行时,该机器就成为实践本发明的设备。当在通用处理器上实现时,程序代码部分与处理器组合以提供唯一的装置,其操作类似

于专用逻辑电路。

[0228] 进而还能够理解,在已经描述并为解释本发明性质的细节、材料和部件排布上,可由业内专业人员在不背离以下权利要求表示的本发明范围之下作出各种变化。

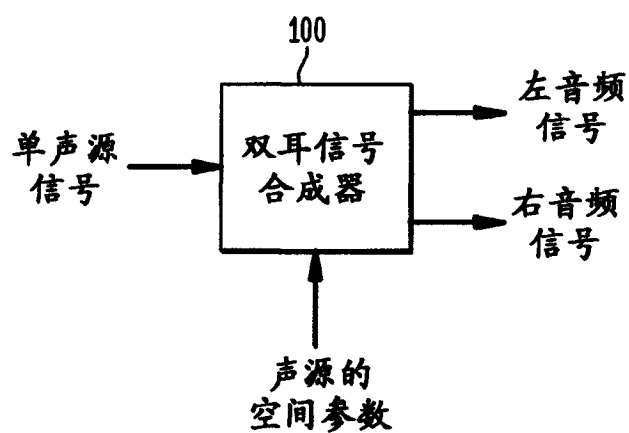


图 1 (现有技术)

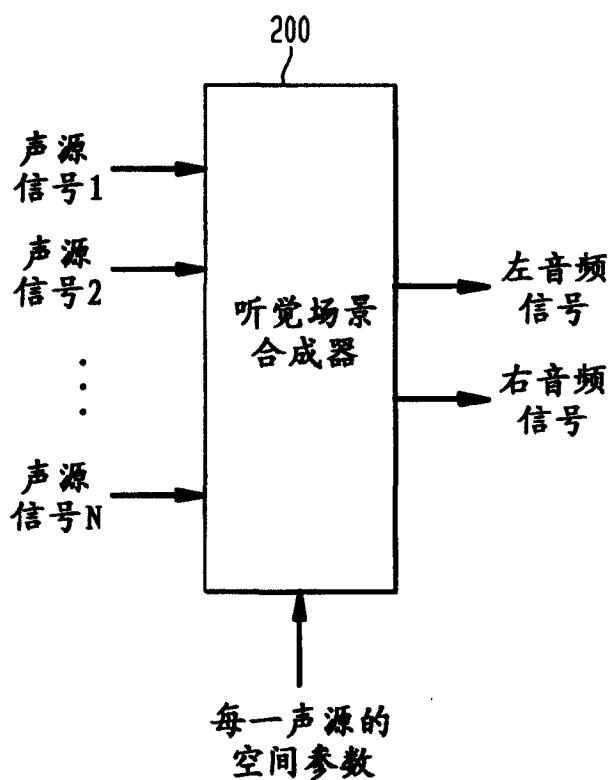


图 2 (现有技术)

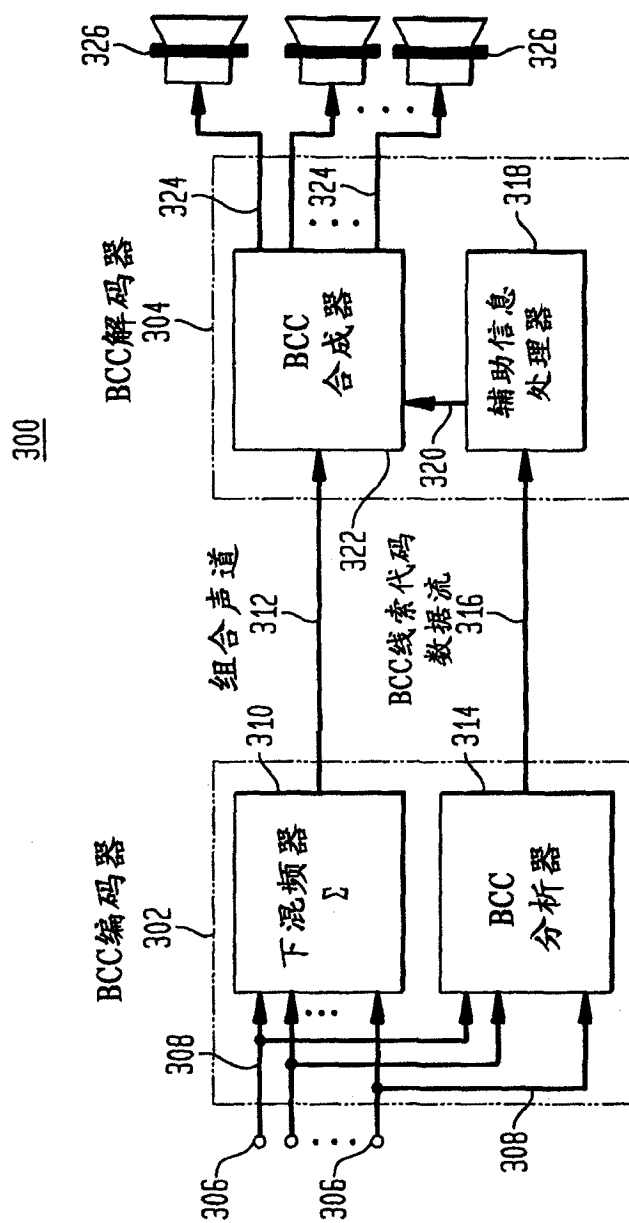


图 3

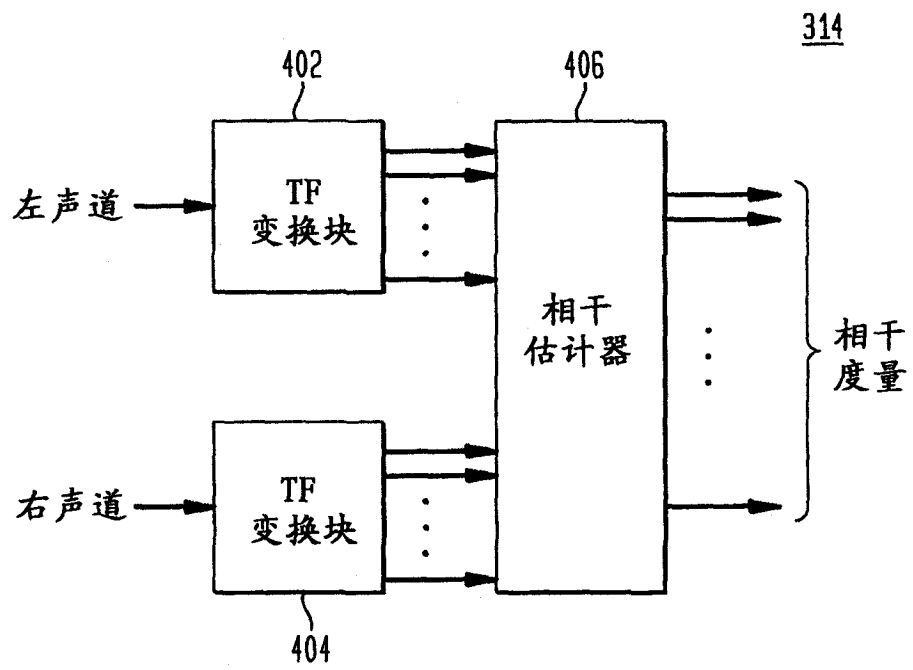
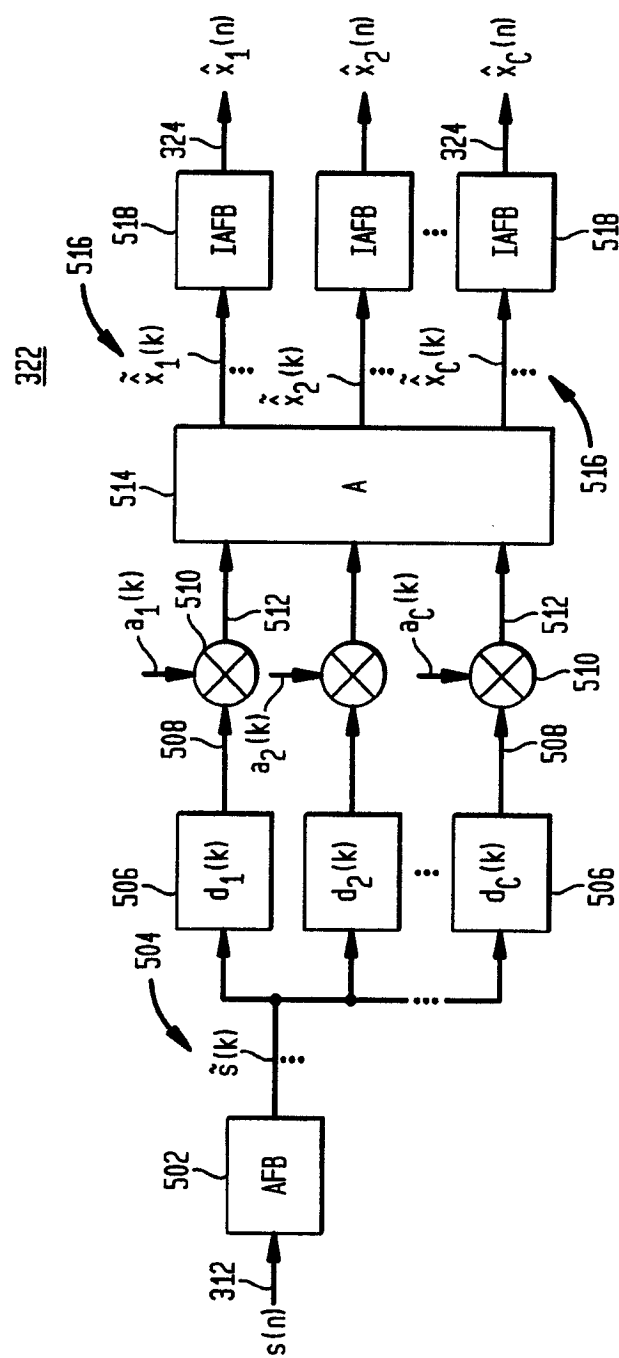


图 4

图 5



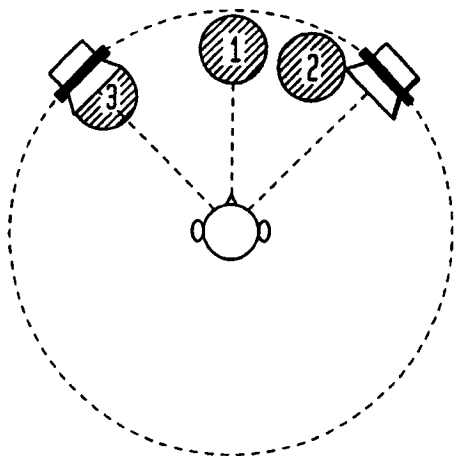


图 6(A)

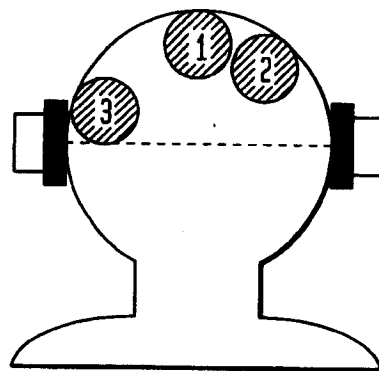


图 6(B)

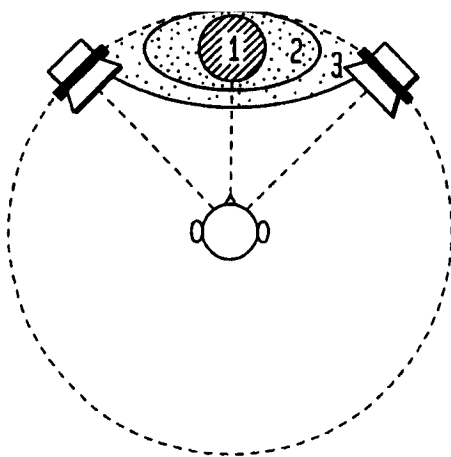


图 6(C)

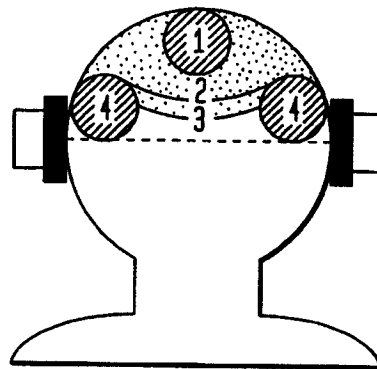


图 6(D)

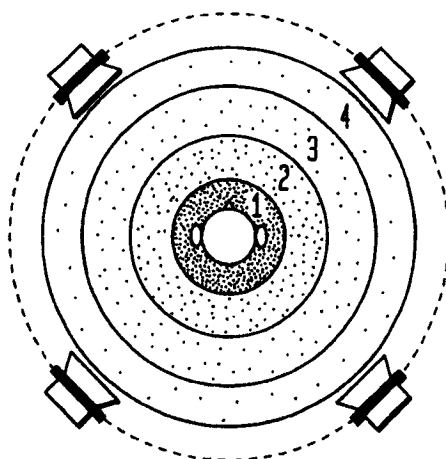
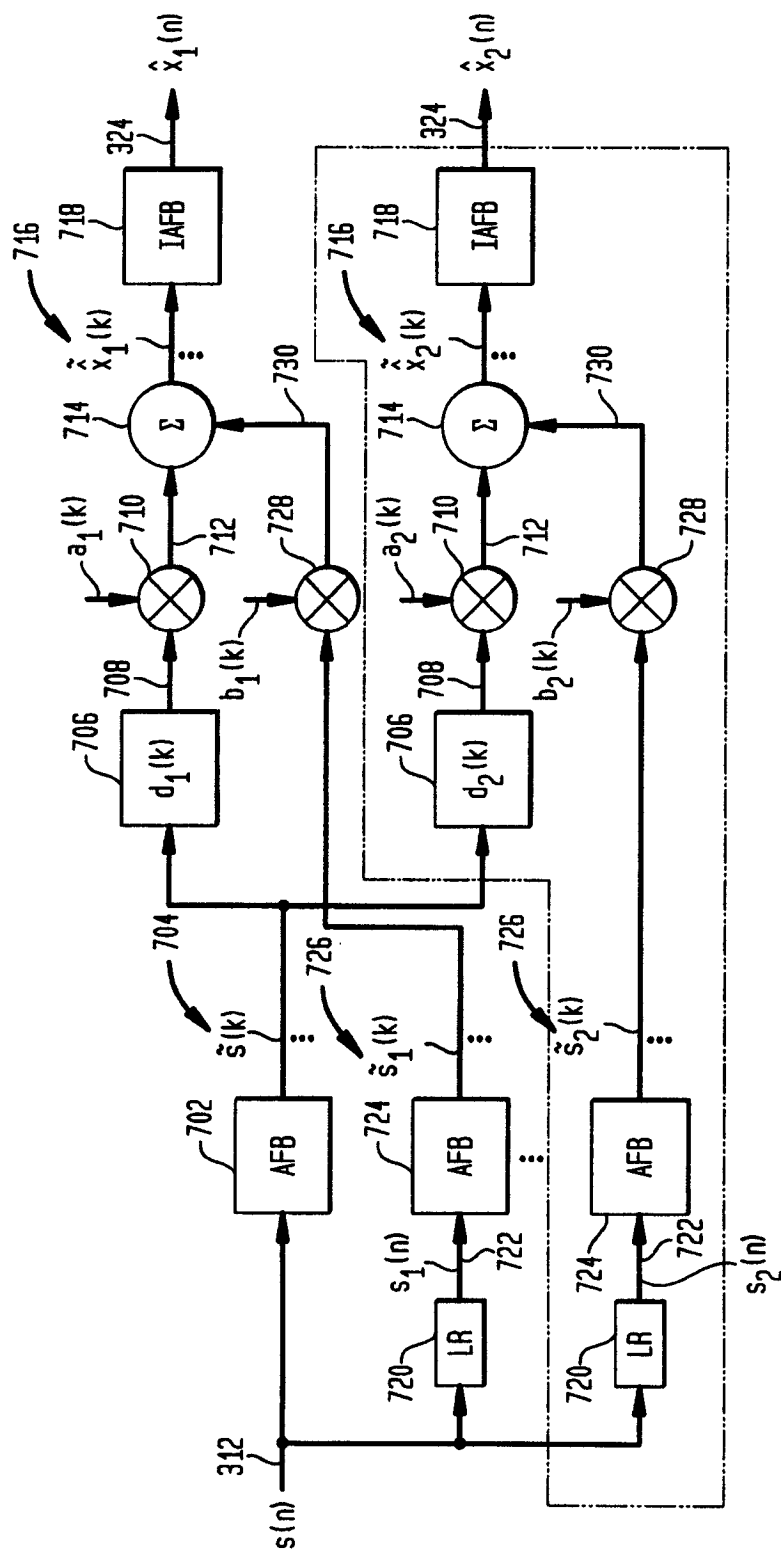


图 6(E)

图 7

322



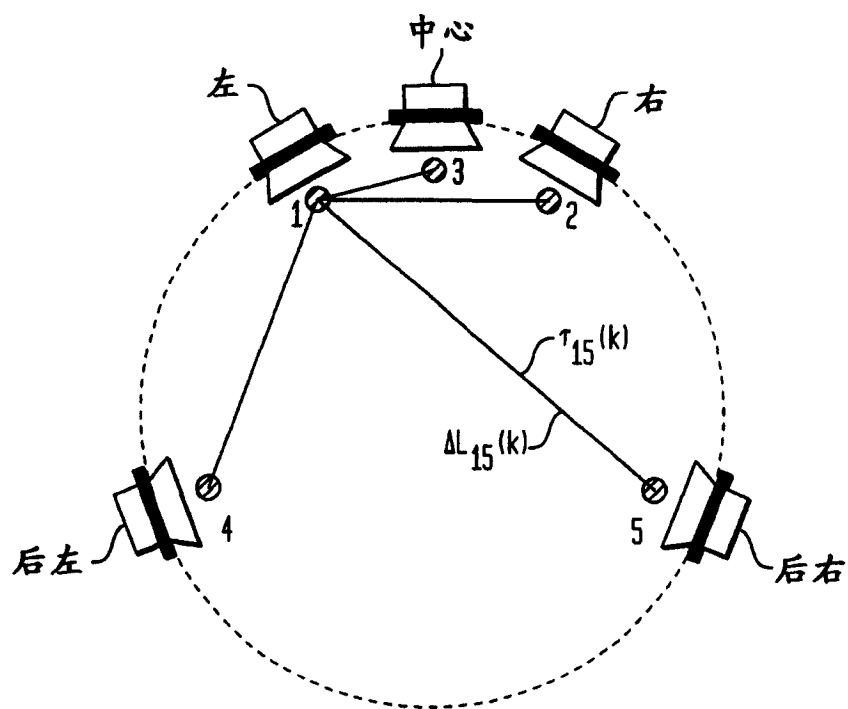


图 8

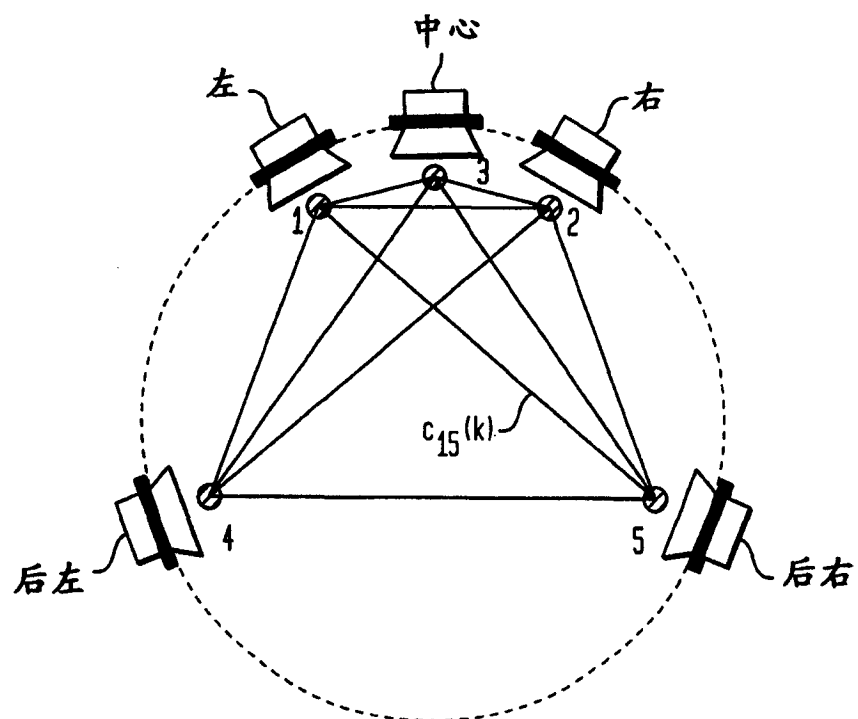


图 9

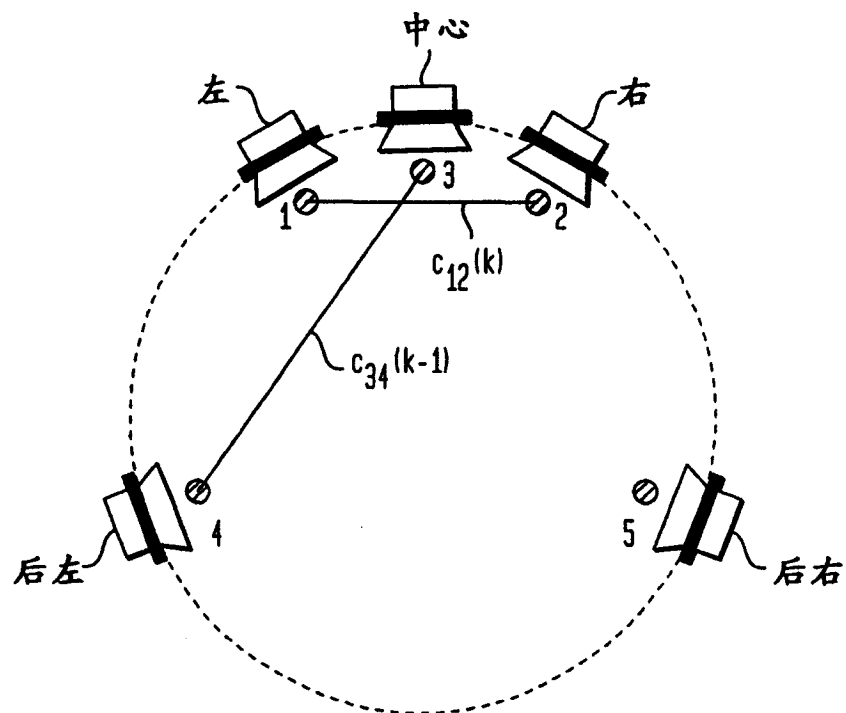


图 10

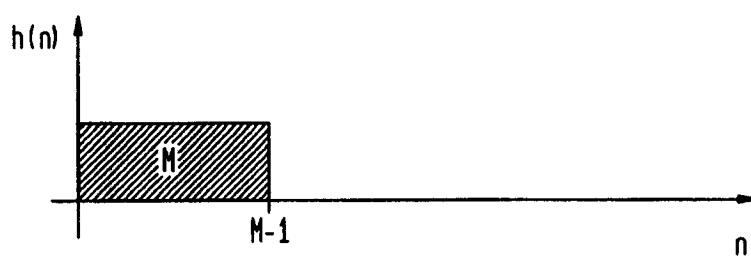


图 11 (A)

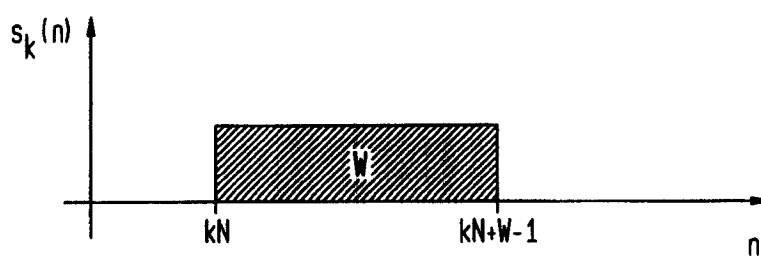


图 11 (B)

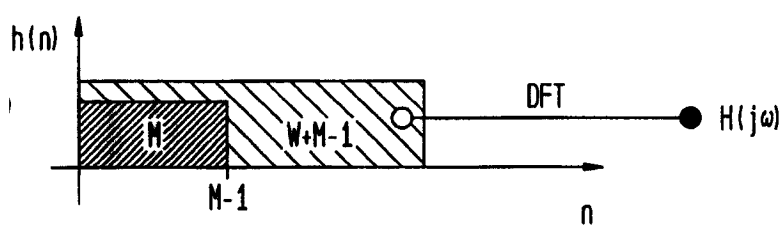
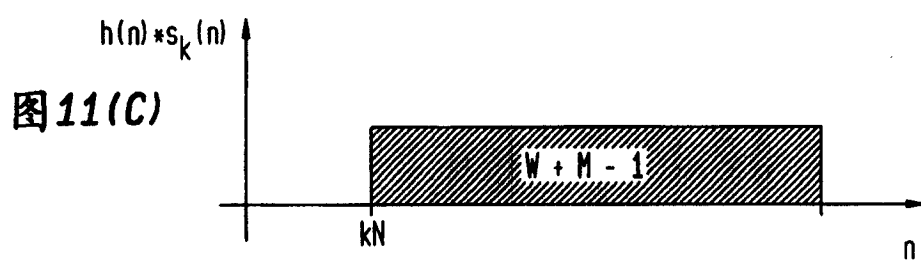


图 12(A)

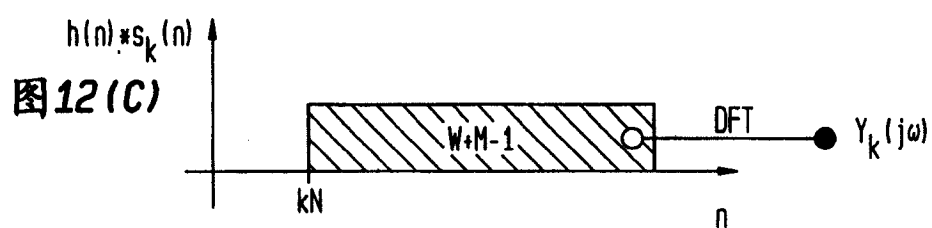
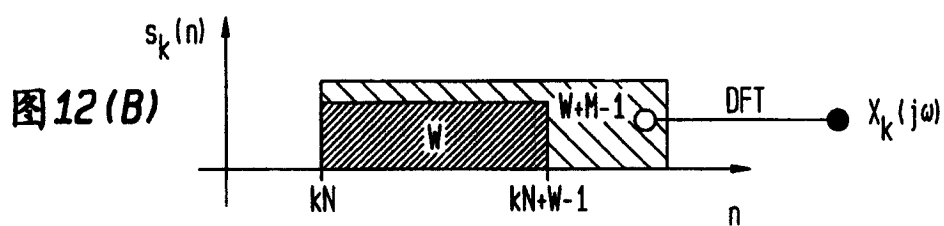


图 13

322

