



(12) 发明专利

(10) 授权公告号 CN 110648651 B

(45) 授权公告日 2023. 08. 25

(21) 申请号 201910939317.4  
(22) 申请日 2014.07.16  
(65) 同一申请的已公布的文献号  
申请公布号 CN 110648651 A  
(43) 申请公布日 2020.01.03  
(30) 优先权数据  
13177362.4 2013.07.22 EP  
13189230.9 2013.10.18 EP  
(62) 分案原申请数据  
201480041457.5 2014.07.16  
(73) 专利权人 弗朗霍夫应用科学研究促进协会  
地址 德国慕尼黑  
(72) 发明人 西蒙·法格 简·普洛格施蒂斯  
(74) 专利代理机构 北京清亦华知识产权代理事  
务所(普通合伙) 11201  
专利代理师 宋融冰

(51) Int.Cl.  
G10K 15/08 (2006.01)  
H04S 3/00 (2006.01)  
H04S 7/00 (2006.01)  
(56) 对比文件  
US 2008175396 A1,2008.07.24  
US 6188769 B1,2001.02.13  
JP 2009053349 A,2009.03.12  
US 2002122559 A1,2002.09.05  
WO 2012116934 A1,2012.09.07  
Takayuki Hidaka,等.A new definition  
of boundary point between early  
reflection and late reverberation in room  
impulse response.《JASA》.2007,  
审查员 孙筱逸

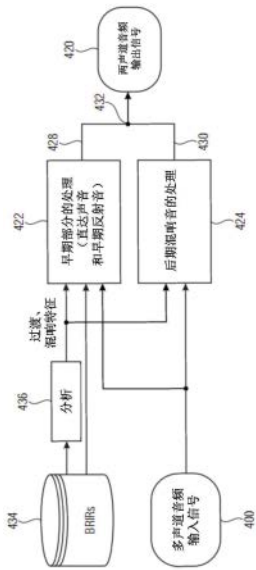
权利要求书3页 说明书22页 附图13页

(54) 发明名称

根据室内脉冲响应处理音频信号的方法、信号处理单元

(57) 摘要

描述了一种根据室内脉冲响应(434)处理音频信号(400)的方法。将音频信号(400)与室内脉冲响应(434)的早期部分和后期混响音分别地处理(422,424),并结合(432)音频信号的处理的早期部分(428)和混响信号(430)。当相关性测量到达阈值时,则室内脉冲响应中从早期部分至后期混响音的过渡达到,阈值根据从室内脉冲响应的早期部分中的早期反射音中选择一个的相关性测量来设定。



1. 一种用于根据室内脉冲响应处理音频信号的方法,所述方法包括:

将所述音频信号分别地与所述室内脉冲响应的早期部分和合成混响音一起处理;以及  
将与所述室内脉冲响应的所述早期部分一起处理的所述音频信号和与所述合成混响音一起处理的音频信号结合,

其中所述室内脉冲响应中从所述早期部分至后期混响音的过渡通过到达阈值的相关性测量来确定,所述阈值根据从所述室内脉冲响应的所述早期部分中的早期反射音中选择一个的所述相关性测量来设定,

其中所述相关性测量描述关于室内脉冲响应的、在包括初期状态的声能中的衰变的相似点以及在所述初期状态之后的任何时间开始的在预定义频率范围的声能中的衰变的相似点,以及

其中早期反射音中选择一个是第一反射音。

2. 根据权利要求1所述的方法,其中确定所述过渡包括:

基于所述室内脉冲响应确定声能的分布;以及

确定多个相关性测量,所述多个相关性测量用于为确定的分布的多个部分指示在所述确定的分布的各个部分中的声能和在初期状态的声能之间的相关性。

3. 根据权利要求2所述的方法,其中

确定所述分布包括确定所述声能的时间-频率分布,以及

所述分布的部分包括预定义长度的时间块,所述初期状态通过所述时间-频率分布的多个时间块中的第一个来定义。

4. 根据权利要求2所述的方法,其中确定所述分布包括从所述室内脉冲响应计算能量衰变救援EDR。

5. 根据权利要求4所述的方法,其中所述EDR如下所示计算:

$$E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

其中

$E(t, \omega)$  = 能量衰变救援,

$h(\tau)$  = 室内脉冲响应,

$\omega = 2\pi f$ 。

6. 根据权利要求3所述的方法,其中所述室内脉冲响应具有预定义的有效长度,其中确定所述时间-频率分布包括使用具有对应于所述室内脉冲响应的有效长度的长度的窗来计算所述室内脉冲响应的FFT光谱。

7. 根据权利要求6所述的方法,其中

在初期状态的所述声能通过采取所述室内脉冲响应的全部有效长度、计算所述FFT光谱以及采取绝对值的平方来确定;以及

时间块的所述声能通过将所述窗移动与所述时间块相关联的时间、将所述窗取样补零至所述有效长度、计算所述FFT以及采取所述绝对值的所述平方来确定。

8. 根据权利要求1所述的方法,其中所述相关性测量如下所示计算:

$$\rho(t) = \frac{\sum_{\omega}(E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega}(E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega}(E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega}(E(t, \omega) - \bar{E}(t, \omega))^2}}$$

其中

$\rho(t)$  = 相关性测量,

$E(1, \omega)$  = 在频率 $f$ 的全部频率范围的能量衰变救援,

$\bar{E}(1, \omega)$  = 初期全部范围的能量衰变救援的所有频率的平均值,

$E(t, \omega)$  = 从时间 $t$ 开始在频率 $f$ 处的能量衰变救援,

$\bar{E}(t, \omega)$  = 在时间 $t$ 开始的全部范围的能量衰变救援的所有频率的平均值,

$\omega = 2\pi f$ 。

9. 根据权利要求1所述的方法, 其中所述阈值根据常数值和所述第一反射音的所述相关性测量来确定。

10. 根据权利要求9所述的方法, 其中所述常数为 $1/e$ , 其中所述阈值如下所示定义:

$$\rho(t) = c \cdot \rho(t_F)$$

其中

$\rho(t_F)$  = 所述第一反射音的相关性测量,

$t_F$  = 在直达声音撞击之后的所述第一反射音的时间索引,

$c$  = 基于 $\frac{1}{e}$ 的所述常数值,  $e$ 为欧拉数。

11. 根据权利要求1所述的方法, 其中确定所述过渡包括:

确定所述第一反射音的时间。

12. 根据权利要求11所述的方法, 其中所述第一反射音的时间通过运行的峰度运算符、通过阈值检测或通过攻击检测来确定。

13. 一种计算机可读介质, 包括存储指令的计算机程序, 其中当所述计算机程序在计算机上运行时, 执行权利要求1至12中任一项所述的方法。

14. 一种信号处理单元, 包括:

输入端, 用于接收音频信号;

处理器, 被配置用于依权利要求1至12中任一项所述的方法, 根据室内脉冲响应处理接收的音频信号; 以及

输出端, 用于将所述接收的音频信号的处理的早期部分和混响信号结合成输出音频信号。

15. 根据权利要求14所述的信号处理单元, 包括:

早期部分处理器, 用于根据所述室内脉冲响应的所述早期部分处理所述接收的音频信号; 以及

混响音处理器, 用于根据合成混响音处理所述接收的音频信号。

16. 一种用于编码音频信号的音频编码器, 其中

所述音频编码器被配置用于依权利要求1至12中任一项所述的方法根据室内脉冲响应处理待编码音频信号。

17. 根据权利要求16所述的音频编码器, 其中所述音频编码器包括权利要求14或15所

述的信号处理单元。

18. 一种用于解码编码音频信号的音频解码器, 其中

所述音频解码器被配置用于依权利要求1至12中任一项所述的方法根据室内脉冲响应处理解码音频信号。

19. 根据权利要求18所述的音频解码器, 其中所述音频解码器包括根据权利要求14或15所述的信号处理单元。

20. 根据权利要求18所述的音频解码器, 包括渲染器, 被配置用于接收所述解码音频信号, 并在所述室内脉冲响应的基础上渲染输出信号。

21. 根据权利要求20所述的音频解码器, 其中所述渲染器包括立体声渲染器。

22. 一种立体声渲染器, 包括根据权利要求14或15所述的信号处理单元。

## 根据室内脉冲响应处理音频信号的方法、信号处理单元

### 技术领域

[0001] 本发明涉及音频编码/解码领域,特别是涉及空间音频编码和空间音频对象编码,例如,3D音频编解码系统领域。本发明的实施例涉及根据室内脉冲响应处理音频信号的方法,以及在这种室内脉冲响应中确定从早期反射音至后期混响音的过渡的方法。

### 背景技术

[0002] 空间音频编码在技术领域被熟知且被标准化,例如,在MPEG环绕标准中。空间音频编码从多个原始输入开始,例如,五个或七个输入声道,其通过再现设置中的布置来识别,例如,左声道、中间声道、右声道、左环绕声道、右环绕声道和低频增强声道。空间音频编码器可以从原始声道获得一个或多个降混声道,另外可以获得关于空间线索的参数数据,例如,声道相干值中的声道间位准差异、声道间相位差异、声道间时间差异等。一个或多个降混声道和指示空间线索的参数边信息一起传输至空间音频解码器,空间音频解码器用于解码降混声道和相关联的参数数据,以最后获得与原始输入声道的近似版本的输出声道。声道在输出设置中的布置可以是固定的,例如,5.1格式或7.1格式等。

[0003] 空间音频对象编码工具也在技术领域被熟知且被标准化,例如,在MPEG SAOC标准(SAOC=spatial audio object coding,空间音频对象编码)中。相比于空间音频编码从原始声道开始,空间音频对象编码从非自动专用为特定渲染再现设置的音频对象开始。相反地,音频对象在再现场景中的布置是可变化的并且可以由用户设定,例如,通过输入特定渲染信息到空间音频对象编码的解码器。可选地或附加地,渲染信息可作为附加边信息或元数据(metadata)传输;渲染信息可以包括特定音频对象所布置在再现设置中的位置的信息(例如,时间方面)。为了获得特定的数据压缩,使用SAOC编码器来编码一定数量的音频对象,SAOC编码器根据特定的降混信息来降混对象以从输入对象计算出一个或多个运输声道。此外,SAOC编码器计算代表对象间线索的参数边信息,例如,对象位准差异(OLD)、对象相干数值等。当在SAC(SAC=Spatial Audio Coding)中,对象间参数数据针对各个时间平铺(time tiles)/频率平铺(frequency tiles)来计算。针对音频信号的特定帧(例如1024或2048个取样值),考虑多个频带(例如24、32或64个频带)使得参数数据能提供给每个帧和每个频带。例如,当音频片具有20个帧且当每个帧细分成32个频带时,时间/频率平铺的数量为640。

[0004] 在3D音频系统中,可能期望提供音频信号的空间印象,仿佛音频信号在特定空间中被聆听。在这样的情况下,例如在测量基础上提供特定空间的室内脉冲响应,用于进行将音频信号表现给听众的处理。可能期望处理直达声音以及与后期混响音分开表现的早期反射音。需要确定早期反射音的结束处以及后期混响音的开始处。

### 发明内容

[0005] 本发明的目的是提供用于根据室内脉冲响应处理音频信号的改良方法。

[0006] 该目的通过下面描述的方法、信号处理单元、音频编码器、音频解码器以及立体声

渲染器来达成。

[0007] 本发明根据发明人发现在传统方法中确定出从早期反射音到后期混响音的过渡有过早的问题,因为在第一反射发生或撞击之前用来判断过渡的发生的相关性已经到达阈值。基于这些发现以及已知过渡时间必须大于第一反射的到达时间,因为第一反射清楚明显且能确定不是后期散播混响音,发明人发现有必要避开固定阈值的使用;相反地,根据本发明的方法,阈值的定义取决于早期反射音中的其中一个的撞击时间上的相关性。确保第一反射总会在过渡时间之前。

[0008] (1) 本发明提供一种用于根据室内脉冲响应处理音频信号的方法,该方法包括:

[0009] 将音频信号进行分别地与室内脉冲响应的早期部分和后期混响音一起处理;以及

[0010] 将与室内脉冲响应的早期部分一起处理的音频信号和混响信号结合,

[0011] 其中在室内脉冲响应中从早期部分至后期混响音的过渡通过到达阈值的相关性测量来确定,阈值根据从室内脉冲响应的早期部分中的早期反射音中选择一个的相关性测量来设定。

[0012] 本发明方法的优点在于允许基于鲁棒过渡点进行音频信号的改良处理。本发明方法无关房间,无关是否使用双声道方法以及入射角。相比于现有技术方法,本发明方法的进一步优点在于其并非强烈地依赖于双声道脉冲响应的方位角以及直达声音与第一撞击反射的振幅之间的关系。

[0013] (2) 根据实施例,相关性测量描述关于室内脉冲响应的、在包括初期状态的声能中的衰变的相似点以及在初期状态之后的任何时间开始的在预定义频率范围的声能中的衰变的相似点。

[0014] (3) 根据实施例,确定过渡包括基于室内脉冲响应确定声能的分布,以及确定多个相关性测量,多个相关性测量用于为确定的分布的多个部分指示在确定的分布的各个部分中的声能和在初期状态的声能之间的相关性。

[0015] (4) 根据实施例,确定分布包括确定声能的时间-频率分布,其中分布的部分包括预定义长度的时间块,初期状态通过时间-频率分布的多个时间块中的第一个来定义。

[0016] 该优点在于允许在不同时间分析频率分布,从而提供频率分布的特性在时间上的表示。

[0017] (5) 根据实施例,确定分布包括从室内脉冲响应计算能量衰变救援(energy decay relief,EDR),其中EDR如下所示计算:

$$[0018] \quad E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

[0019] 其中

[0020]  $E(t, \omega)$  = 能量衰变救援,

[0021]  $h(\tau)$  = 室内脉冲响应,

[0022]  $\omega = 2\pi f$ 。

[0023] (6) 根据实施例,室内脉冲响应具有预定义的有效长度,其中确定时间-频率分布包括使用具有对应于室内脉冲响应的有效长度的长度的窗来计算室内脉冲响应的FFT光谱。

[0024] 该优点在于FFT/DFT能够良好定义且存在有效算法来计算频谱值。如果窗中的数

值是已知的,则FFT/DFT可以在不复杂方式下计算。

[0025] (7)根据实施例,在初期状态的声能通过采取室内脉冲响应的全部有效长度、计算FFT光谱、采取绝对值的平方来确定,时间块的声能通过将窗移动与时间块相关联的时间、将窗取样补零至有效长度、计算FFT以及采取绝对值的平方来确定。

[0026] 该优点在于不需要附加滤波器组或其他相似装置用于EDR的窄频带计算;仅需要进行窗的移动。

[0027] (8)根据实施例,相关性测量为用来描述在包含初期状态的声能中的衰变的相似点和在初期状态之后任何时间开始的声能中的衰变的相似点的相关性测量。相关性测量可以如下所示计算:

$$[0028] \quad \rho(t) = \frac{\sum_{\omega}(E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega}(E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega}(E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega}(E(t, \omega) - \bar{E}(t, \omega))^2}}$$

[0029] 其中

[0030]  $\rho(t)$  = 相关性测量,

[0031]  $E(1, \omega)$  = 在频率f的全部频率范围的能量衰变救援,

[0032]  $\bar{E}(1, \omega)$  = 初期全部范围的能量衰变救援的所有频率的平均值,

[0033]  $E(t, \omega)$  = 从时间t开始在频率f处的能量衰变救援,

[0034]  $\bar{E}(t, \omega)$  = 在时间t开始的全部范围的能量衰变救援的所有频率的平均值,

[0035]  $\omega = 2\pi f$ 。

[0036] 该优点在于上述公式参考熟知的Pearson的相关性系数(Pearson的积差相关)。相关性系数可以直接从EDR计算而得。

[0037] (9)根据实施例,阈值根据常数值和早期反射音中选择一个的相关性测量来确定。阈值可以如下所示定义:

$$[0038] \quad \rho(t) = c \cdot \rho(t_F)$$

[0039] 其中

[0040]  $\rho(t_F)$  = 早期反射音中选择一个的相关性测量,

[0041]  $t_F$  = 在直达声音撞击之后的早期反射音中选择一个的时间索引,

[0042]  $c$  = 基于 $\frac{1}{e}$ 的常数值,e为欧拉数(Euler number)。

[0043] 该优点在于阈值并非恒定的,而取决于所选的早期反射音,以确保相关性不会太早落在阈值以下。

[0044] (10)根据实施例,确定早期反射音中选择一个的时间,例如,通过运行的峰度运算符、通过阈值检测或通过攻击检测来确定。

[0045] 该优点在于反射的撞击时间可以直接且自动地从脉冲响应的时域样本中计算。

[0046] (11)根据实施例,早期反射音中选择一个的是直达声音之后的第一个反射音。

[0047] (12)本发明提供一种信号处理单元,包括用于接收音频信号的输入端、被配置或被程序化用于依本发明方法根据室内脉冲响应处理接收的音频信号的处理器、以及用于结合接收的音频信号的处理的早期部分和混响信号成输出音频信号的输出端。信号处理单元可以包括用于根据室内脉冲响应的早期部分处理接收的音频信号的早期部分处理器、以及

用于根据室内脉冲响应的后期混响音处理接收的音频信号的后期混响音处理器。

[0048] (13) 本发明提供用于编码音频信号的音频编码器,其中音频编码器被配置或被程序化用于依本发明方法根据室内脉冲响应处理待编码音频信号。音频编码器可以包括本发明的信号处理单元。

[0049] (14) 本发明提供用于解码编码音频信号的音频解码器,其中音频解码器被配置或被程序化用于依本发明方法根据室内脉冲响应对解码音频信号。音频解码器可以包括本发明的信号处理单元。音频解码器可以包括渲染器,其相似于立体声渲染器,被配置或被程序化用于接收解码音频信号,并在室内脉冲响应的基础上渲染输出信号。

[0050] (15) 本发明提供立体声渲染器,包括本发明的信号处理单元。

## 附图说明

[0051] 本发明的实施例将参考附图进行描述,其中

[0052] 图1显示3D音频系统的3D音频编码器的概要图;

[0053] 图2显示3D音频系统的3D音频解码器的概要图;

[0054] 图3显示用于实施格式转换器的示例,该格式转换器可以实施在图2的3D音频解码器中;

[0055] 图4显示立体声渲染器的实施例,该立体声渲染器可以实施在图2的3D音频解码器中;

[0056] 图5显示室内脉冲响应 $h(t)$ 的示例;

[0057] 图6(A) 显示根据本发明实施例的用于分别地与室内脉冲响应的早期部分和后期混响音一起处理音频信号的信号处理单元(例如在图4的立体声渲染器中)的方块图;

[0058] 图6(B) 显示根据本发明进一步实施例的用于分别地与室内脉冲响应的早期部分和后期混响音一起处理音频信号的另一信号处理单元(例如在图4的立体声渲染器中)的方块图;

[0059] 图7显示根据本发明实施例的用于确定在室内脉冲响应中早期反射音和后期混响音之间的过渡时间的方法的流程图;

[0060] 图8显示达成的用于根据基于FFT的方法所确定的脉冲响应的能量衰变救援(EDR);

[0061] 图9显示根据本发明实施例的过渡时间的确定;

[0062] 图10显示使用现有方法所确定的测量的双声道室内脉冲响应的左声道和右声道的过渡时间;

[0063] 图11显示使用本发明方法所确定的测量的双声道室内脉冲响应的左声道和右声道的过渡时间;

[0064] 图12大略地显示根据本发明实施例的立体声渲染器中的音频信号的双声道处理;

[0065] 图13大略地显示根据本发明实施例的图12的立体声渲染器的频域混响器中的处理。

## 具体实施方式

[0066] 以下描述用于根据室内脉冲响应处理音频信号本发明方法的实施例,以及用于在



室内脉冲响应中确定从早期反射音至后期混响音的过渡的本发明方法的实施例。下面的描述将从可以实现本发明方法的3D音频编解码系统的系统概要开始。

[0067] 图1和图2显示根据实施例的3D音频系统的算法框图。更具体地,图1显示3D编码器100的概要。音频编码器100在可选择地提供的预渲染器/混合器电路102处接收输入信号,更具体地,多个输入声道提供给音频编码器100多个声道信号104、多个对象信号106以及相对应的对象元数据108。通过预渲染器/混合器102(参见信号110)所处理的对象信号106可提供至SAOC编码器112。SAOC编码器112产生被提供至USAC (USAC=Unified Speech and Audio Coding,统一语音和音频编码) 编码器116的SAOC运输声道114。此外,信号SAOC-SI (SAOC-SI=SAOC side information,SAOC边信息)118也提供至USAC编码器116。USAC编码器116进一步不仅接收直接来自预渲染器/混合器的对象信号120,还接收声道信号以及预渲染的对象信号122。对象元数据信息108施加至OAM (OAM=object metadata,对象元数据) 编码器124,该OAM编码器124提供压缩的对象元数据信息126至USAC编码器。基于上述输入信号,USAC编码器116产生压缩的输出信号mp4,其标示为128。

[0068] 图2显示3D音频系统的3D音频解码器200的概要。音频解码器200接收图1的音频编码器100所产生的编码信号128 (mp4),更具体地,在USAC解码器202处接收。USAC解码器202将所接收的信号128解码成声道信号204、预渲染的对象信号206、对象信号208以及SAOC运输声道信号210。进一步,USAC解码器202输出压缩的对象元数据信息212以及信号SAOC-SI 214。对象信号208提供至用于输出渲染的对象信号218的对象渲染器216。SAOC运输声道信号210应用至用于输出渲染的对象信号222的SAOC解码器220。压缩的对象元数据信息212应用至OAM解码器224,OAM解码器224用于输出各个控制信号至对象渲染器216以及SAOC解码器220,用于产生渲染的对象信号218以及渲染的对象信号222。解码器进一步包括混合器226,如图2所示,混合器226接收输入信号204、206、218以及222,并输出声道信号228。声道信号可以直接输出至扬声器,例如32声道的扬声器,其标示为230。信号228可以提供至格式转换电路232,格式转换电路232作为控制输入端接收再现布局信号,再现布局信号指示声道信号228的转换方式。图2绘示的实施例中,假设转换以这样的方式完成,信号可以提供至5.1扬声器系统,其标示为234。而且,声道信号228可以提供至立体声渲染器236以产生两个输出信号,例如用于头戴式耳机,其标示为238。

[0069] 在本发明的实施例中,在图1和图2所绘示的编码/解码系统基于MPEG-D USAC编解码,来编码声道和对象信号(参见信号104以及106)。为了增加编码大量对象的效率,可使用MPEG SAOC技术。三种类型的渲染器可执行渲染对象至声道、渲染声道至头戴式耳机或渲染声道至不同扬声器设置的任务(参见图2,标号230、234以及238)。当对象信号明确地传输或使用SAOC作参数编码,相对应的对象元数据信息108被压缩(参见信号126)以及被多路复用至3D音频比特流128。

[0070] 以下将进一步详细地描述图1以及图2所示的整体3D音频系统的算法方块。

[0071] 预渲染器/混合器102可选择地提供,以在编码之前将声道加对象输入场景转换成声道场景。功能上,其相同于下面段落将描述的对象渲染器/混合器。需要对象的预渲染,以确保在编码器输入端的确定性的信号熵,基本上无关于同步主动对象信号的数量。根据对象的预渲染,不需要传输对象元数据。离散对象信号渲染至编码器使用的声道布局。从相关联的对象元数据(OAM)针对每个声道获得对象的权重。

[0072] USAC编码器116用于扬声器声道信号、离散对象信号、对象降混信号以及预渲染信号的核心编解码。其基于MPEG-D USAC技术。其根据输入声道以及对象分配的几何以及语义信息,通过创建声道以及对象映像信息来处理上述信号的编码。该映像信息描述输入声道以及对象如何映像至USAC声道组件,如声道对组件(CPEs)、单声道组件(SCEs)、低频效果(LFEs)和四声道组件(QCEs),而CPEs、SCEs以及LFEs以及相对应的信息传输到解码器。编码器的比率控制考虑所有附加酬载(payloads)如SAOC数据114与118、或对象元数据126。对象的编码可能用不同的方式,取决于对渲染器的比率/失真需求以及交互性需求。根据实施例,后续的对象编码变体可能为:

- 预渲染对象:在编码之前,对象信号预渲染并混合至22.2声道信号。后续的编码

[0073] 链可参见22.2声道信号。

[0074] • 离散对象波形:对象作为单声道波形应用至编码器。编码器使用单声道组件(SCEs)传输除了声道信号的对象。解码的对象在接收器侧渲染并混合。压缩的对象元数据信息传输到接收器/渲染器。

[0075] • 参数对象波形:对象属性以及其彼此关系通过SAOC参数的方式来描述。对象信号

[0076] 的降混与USAC一起编码。参数信息并排传输。降混声道的数量的选择取决于对象的数量以及整体数据率。压缩的对象元数据信息传输到SAOC渲染器。

[0077] 用于对象信号的SAOC编码器112以及SAOC解码器220可以基于MPEG SAOC技术。系统能够根据较小量的传输声道以及附加的参数数据,例如OLDs、IOCs(Inter Object Coherence,对象间相干性),DMGs(downmix gain,降混增益)来重建、修改以及渲染大量的音频对象。比起需要单独地传送所有对象,附加的参数数据展示显著较低数据率,使得编码非常有效率。SAOC编码器112以对象/声道信号作为单声道波形,并输出参数信息(其被打包成3D音频比特流128)以及SAOC运输声道(其使用单声道组件编码并传输)。SAOC解码器220从解码的SAOC运输声道210以及参数信息214,重建对象/声道信号,并基于再现布局、解压缩的对象元数据信息,可选择性基于用户互动信息,产生输出音频场景。

[0078] 提供对象元数据编解码(参见OAM编码器124以及OAM解码器224),使得对于每个对象,指定3D空间中对象的几何位置以及容量相关联的元数据能通过对象属性在时间以及空间上的量化而有效地编码。压缩的对象元数据cOAM 126传输到接收器200作为边信息。

[0079] 对象渲染器216根据给定的再现格式,利用压缩的对象元数据以产生对象波形。每个对象根据其元数据而渲染至特定的输出声道。部分结果的总和导致该方块的输出。如果基于声道的内容以及离散/参数对象被解码,则在输出结果波形228之前或在提供到与立体声渲染器236或扬声器渲染器232相似的后处理器模块之前,基于声道的波形以及渲染对象波形通过混合器226混合。

[0080] 立体声渲染器模块236产生多声道音频素材的双声道降混,使得每个输入声道以虚拟声音源来表现。该处理在QMF(Quadrature Mirror Filterbank,正交镜像滤波器组)域中以帧方式进行,而立体声(binauralization)是基于测量的双声道室内脉冲响应。

[0081] 扬声器渲染器232在传输的声道配置228以及所需的再现格式之间转换。其也可被称为“格式转换器”。格式转换器执行至较低数量的输出声道的转换,即,其执行降混。

[0082] 图3显示用于实施格式转换器232的示例。格式转换器232,也被称为扬声器渲染

器,在传输的声道配置以及所需的再现格式之间转换。格式转换器232执行至较低数量的输出声道的转换,即其执行降混(DMX)处理240。降混器240,其优选地运作在QMF域,接收混合器输出信号228并输出扬声器信号234。配置器242,也被称为控制器,可提供作为控制输入端接收信号246和信号248,信号246代表混合器输出布局,即,确定由混合器输出信号228所呈现数据的布局,信号248代表期望的再现布局。根据该信息,控制器242,优选地自动,产生用于输入与输出格式的给定结合的优化的降混矩阵,以及应用这些矩阵至降混器240。格式转换器232允许标准扬声器配置以及具有非标准扬声器的位置的任意配置。

[0083] 图4显示图2的立体声渲染器236的实施例。立体声渲染器模块可提供多声道音频素材的双声道降混。立体声可基于测量的双声道室内脉冲响应。室内脉冲响应可被认为是真实房间的声学特征的“指纹”。测量以及储存室内脉冲响应,而任意的声学信号可有此“指纹”,从而允许听众处的与室内脉冲响应相关联的空间的声学特征的仿真。立体声渲染器236被程序化或被配置用于使用头部相关传输功能或双声道室内脉冲响应(BRIR)将输出声道渲染成两个双声道。例如,针对移动装置,附设于移动装置上的头戴式耳机或扬声器需要双声道渲染。在该种移动装置中,由于多种约束,其必须限制解码器以及渲染的复杂度。除了省略处理方案中的解相关,其较佳的首先使用降混器250执行降混至中间降混信号252,即,降混至较少数量的输出声道,其导致较少数量输入声道用于实际的双声道转换器254。例如,22.2声道素材可通过降混器250降混至5.1中间降混,或者,可选地,在“快捷(shorycut)”模式下可通过图2中的SAOC解码器220直接计算中间降混。如果22.2输入声道将被直接渲染,则相比于施加44HRTF(Head Related Transfer Functions,头部相关传输)或BRIR功能,双声道渲染仅须施加10HRTF功能或用于在不同位置渲染五个独立声道的BRIR功能。对于双声道渲染所必须的卷积操作需要许多大量的处理功率,因此,减少处理功率而仍然可获得可接受的音频质量对移动装置特别地有用。立体声渲染器236产生多声道音频素材228的双声道降混238,使得每个输入声道(排除LFE声道)通过虚拟音源来表现。该处理可在QMF域中以帧方式产生。立体声基于测量的双声道室内脉冲响应,而当后期混响音可分别地处理时,在QMF域的上侧(on-top)使用快速卷积经由伪FFT领域中的卷积方法将直达声音以及早期反射音可留印在音频素材中。

[0084] 图5显示室内脉冲响应 $h(t)$  300的示例。室内脉冲响应包括三部分:直达声音301、早期反射音302以及后期混响音304。如此,当播放脉冲时,室内脉冲响应描述密闭混响声学空间的反射行为。早期反射音302是具有增加密度的离散反射,不会再区别各个反射的脉冲响应的部分则被称为后期混响音304。直达声音301在室内脉冲响应中容易识别,以及可以与早期反射音相分离,然而,从早期反射音302到后期混响音304的过渡不太明显。

[0085] 本发明方法的后续实施例将更详细地描述。根据本发明的实施例,音频信号分别地与室内脉冲响应的早期部分和后期混响音一起处理。将与室内脉冲响应的早期部分一起处理的音频信号和室内脉冲响应的混响信号相结合,并输出作为输出音频信号。为了独立处理,室内脉冲响应中从早期部分到后期混响音的过渡必须已知。该过渡通过达到阈值的相关性测量来确定,其中阈值根据从室内脉冲响应的早期部分中的早期反射音中所选择的一个的相关性测量来设定。相关性测量可以描述关于室内脉冲响应的、在包括初期状态的声能中的衰变的相似点以及在初期状态之后的任何时间开始的在预定义频率范围的声能中的衰变的相似点。

[0086] 根据实施例,音频信号的单独处理包括在第一初期期间将音频信号和室内脉冲响应的早期反射音部301,302一起处理,以及在第二处理期间将音频信号和室内脉冲响应的扩散混响音304一起处理,第二处理不同于第一处理且与第一处理相分离。从第一处理到第二处理的改变出现在过渡时间。根据进一步的实施例,在第二处理中扩散(后期)混响音304可替换成合成混响音。在该情形中,所提供的室内脉冲响应可仅包括早期反射音部301,302(参见图5),且不包括后期扩散混响音304。

[0087] 图6(A)显示根据本发明实施例的用于分别地与室内脉冲响应的早期部分和后期混响音一起处理音频信号的第一示例性信号处理单元的方块图。可在上述的立体声渲染器236中执行根据室内脉冲响应的不同部分的音频信号的处理。音频输入信号400可为非混响音频素材,例如多声道音频输入信号,其与室内脉冲响应相卷积;或例如使用仿真头或耳道式麦克风测量到的室内脉冲响应。卷积允许增益原始的非混响音频素材的空间印象,仿佛是在与室内脉冲响应相关联的空间中聆听该音频素材。例如,在上述的立体声渲染器236中,期望在室内脉冲响应中分别地将音频信号和直达声音301以及早期反射音一起处理,将音频信号和后期混响音304一起处理。为了处理音频输入信号400,方块402用于处理直达声音,方块404用于处理早期反射音以及方块406用于处理后期混响音。通过第一加法器412将各个方块402至406的输出信号408至410相结合,以产生早期处理信号414。处理器406提供的早期处理信号414以及混响信号416通过第二加法器418结合,以产生音频输出信号420,其提供给听众仿佛音频信号在与该室内脉冲响应相关联的空间中聆听的印象。

[0088] 将后期混响音302与直达声音以及早期反射音分开处理的优点在于减少计算复杂度。更具体地,对全部脉冲响应使用卷积计算是非常昂贵的。因此,为了模拟后期混响音,通常使用具有较低复杂度的混响音算法来处理音频信号。脉冲响应的直达声音部以及早期反射音部通过,例如卷积计算,更精确地进行计算。进一步的优点在于对混响音控制的可能性。后期混响音可取决于使用者的输入或音频信号的内容而修改测量的空间参数。为了达成上述优点,必须已知早期反射音302结束且后期混响音304开始的过渡位置(例如时间点)。当后期混响音处理开始的太早,则音频信号可能会有较低的质量而人听力能发现明显消失的早期反射音。另一方面,如果过渡时间发现的太晚,则计算效率将不会被利用,造成早期反射音处理通常比后期混响音处理更昂贵。例如,在时域取样的过渡可提供至立体声渲染器作为输入参数,然后根据接收的过渡,该输入参数控制处理器402至406用于分别地处理音频信号。

[0089] 图6(B)显示根据本发明另一实施例的用于分别地与室内脉冲响应的早期部分和后期混响音一起处理音频信号的另一信号处理单元的方块图。接收输入信号400,例如多声道音频输入信号,并施加至第一处理器422用于处理早期部分,即根据图5所示的室内脉冲响应300中的直达声音301以及早期反射音302处理音频信号。多声道音频输入信号400也施加至第二处理器424用于根据室内脉冲响应的后期混响音304处理音频信号。在立体声渲染器中,如上所述,主要因为降低计算复杂度,所以理想的是处理直达声音以及从后期混响音区分出早期反射音。例如,直达声音以及早期反射音的处理可通过第一处理器422执行的卷积方法留印至音频信号,而后期混响音可被替换成第二处理器424提供的合成混响音。整体双声道输出信号420为处理器422提供的卷积结果428以及处理器424提供的合成混响信号430的结合。根据实施例,加法器432结合信号428以及430以输出整体双声道输出信号420。

[0090] 如上所述,第一处理器422可将音频输入信号400与室内脉冲响应的直达声音以及早期反射音进行卷积运算,可从具有多个纪录的双声道室内脉冲响应的外部数据库434中提供该室内脉冲响应给第一处理器422。第二处理器或混响器424可基于混响器参数如混响音RT60以及混响音能量进行操作,混响器参数可通过分析436从储存的双声道室内脉冲响应获得。应注意的是,分析436并非渲染器的必要部分,相反地,表示各个混响音参数可从储存在数据库434中的各个响应获得;此可在外部完成。可通过计算在八音度或1/3八音度滤波器组分析中的能量以及RT60混响音时间来决定混响器参数;或者以多个脉冲响应分析的结果的平均值来确定混响器参数。

[0091] 此外,处理器422以及424从数据库434直接接收或经由分析436接收输入参数,也接收关于在室内脉冲响应中从早期部分到后期混响音的过渡的信息。以下将更详细地描述确定过渡的方法。

[0092] 根据实施例,过渡分析可用于区别早期反射音以及后期混响音。其可提供至立体声渲染器作为输入参数(例如,其可从有用于配置混响器的RT60数值以及能量数值的专用文档/界面中读取)。分析可根据一组双声道室内脉冲响应(一组用于多个方位角以及仰角的BRIR对)。分析可为预处理步骤,其针对每个脉冲响应分别地执行,然后获得所有过渡数值的中位数作为该组BRIR的整体过渡数值。然后,整体过渡数值可用于在双声道输出信号的计算中从后期混响音中分离早期反射音。

[0093] 有几个已知的方法用于确定过渡,然而,现在将描述这些方法的缺点。在现有技术参考文献[1]中,所述的方法使用能量衰变救援(EDR)以及相关性测量来确定从早期反射音到后期混响音的过渡时间。然而,现有技术参考文献[1]所述的方法有缺点。

[0094] 1.该方法强烈地取决于双声道脉冲响应的方位角以及直达声音与第一撞击反射的振幅之间的关系。

[0095] 2.过渡时间在任意频带中进行计算。没有一般的知识有关于哪个频带可用于整体脉冲响应的正确的过渡时间。

[0096] 3.没有信息有关于该方法的必要相关性步骤。

[0097] 另一已知的方法是通过空间中回音的散布来描述早期反射音,例如通过每秒反射的平均数量,并且当该数量超过预定义的阈值(参见现有技术参考文献[2])时,确定为后期混响音的开始。该方法依赖空间特性,即空间体积,但其通常是未知的。空间体积不能容易地从测量的脉冲响应得出。因此,该方法并非适合应用于从测量的脉冲响应计算过渡。而且,目前关于反射音必须多密才被称为后期混响音没有常识。

[0098] 另一可能性,在技术参考文献[3]中所描述的是比较在脉冲响应窗中一时间上的实际分布与时域的高斯分布。假设后期混响音具有正态分布。在正态分布中大约1/3(精确地为1/e)的取样在平均值的标准偏差外,而大约2/3的取样在平均值的标准偏差内。明显地,早期反射音有更多取样在一个标准偏差之内,而较少取样在一个标准偏差之外。在一个标准偏差之外的取样对在一个标准偏差之内的取样之间的比率可用于定义过渡时间。然而,该方法的缺点在于因为该比率有时在阈值附近波动所以该过渡也难以用此手段定义。该手段也强烈地取决于用于计算比率的滑行窗的尺寸以及类型。

[0099] 除了上述方法,峰态(Kurtosis)(随机信号的更阶累积量)可用于确定过渡时间。当向脉冲响应的后期部接近时,其可快速降低,如现有技术参考文献[4]所概述。然而,针对

过渡(使用当其第一次到达零时迅速降低或时间)的阈值的定义并非清楚。

[0100] 另有一方法,其不依赖测量的脉冲响应的分析,但是依赖空间体积,如同文献[2]所描述。该方法假设过渡时间仅取决于体积,但是其不考虑边界的扩散属性。因此,当不需要精确地确定过渡转换时间时,结果仅是过渡时间的近似值,并非为了避免上述缺点而所需的精确值。进一步,空间的体积通常不是已知的且不容易地从测量的脉冲响应中得到。

[0101] 其他已知的方法则完全地忽略环境且定义过渡时间为简单的80ms,例如,请参见现有技术参考文献[5]。然而,该数值完全脱离空间特性或测量的脉冲响应,因此对于从脉冲响应的提醒分隔出后期混响音的目的而言太不准确。

[0102] 除了改良的音频信号处理之外,根据本发明的实施例也提供确定在室内脉冲响应中早期反射音以及后期混响音之间的过渡时间的改良方法,以产生更精确的过渡时间的确定。将于下面段落描述的实施例,提供简单且有效可能以使用FFT分析从测量的脉冲响应中计算出过渡时间。

[0103] 图7显示根据本发明实施例的用于确定在室内脉冲响应中早期反射音和后期混响音之间的过渡时间的方法的流程图。为了确定从早期反射音到后期混响音的过渡时间,在第一步骤500中确定声能的时间-频率分布。例如,根据实施例,能量衰变救援( $E(t, f)$ , EDR)可在步骤500中计算。EDR可为直接从测量的(例如双声道)室内脉冲响应中计算,以及可解释成共享能量衰变曲线(Schroeder integration EDC(d))的频域扩张,该曲线显示在时间 $t$ 之后在脉冲响应中的剩余能量。取代使用宽带脉冲响应,EDR从时间-频率表示获得,且为此目的,可使用许多不同时间-频率表示。一旦声能的时间-频率分布已经在步骤500所确定,则在步骤502确定时间-频率分布的时间块上的声能以及在初期状态的整体声能之间的相关性测量。在步骤504中确定相关性测量是否到达定义的阈值(例如,落在定义阈值的下方)。如果未达到阈值,本方法进行步骤506,选择下一个时间块以及目前时间块的后续分布,并针对下一个时间块重复进行步骤502以及504。如此,根据步骤502至506,相关性测量用于计算步骤500确定的EDR的每个时间块和初期状态的整体能量之间的相关性数值。当相关性测量达到定义的阈值(例如,落在定义阈值之下)时,则表示到达过渡时间。换句话说,当在步骤504中对于目前时间块确定相关性测量低于阈值,则本方法进行步骤508将目前时间块的时间输出作为过渡时间。

[0104] 以下将更详细地描述本发明方法的实施例。最初,可将测量的双声道脉冲响应作为输入用于过渡时间的计算。然后,Page分布或Levin分布用于能量衰变救援(EDR)的计算。Page分布是指参考过去执行光谱的衍生物,而时间反转信号的Page分布被称为Levin分布(也参见现有技术参考文献[2])。该分布描述瞬间功率谱,以及脉冲响应 $h(t)$ (例如,参加图5)的EDR如下所示计算:

$$[0105] \quad E(t, \omega) = \left| \int_t^{\infty} h(\tau) e^{-j\omega\tau} d\tau \right|^2$$

[0106] 其中

[0107]  $E(t, \omega)$  = 能量衰变救援,

[0108]  $h(\tau)$  = 室内脉冲响应,

[0109]  $\omega = 2\pi f$ 。

[0110] 根据上述公式的计算在直达声音301(参见图5)处开始,而随着时间增加能量衰变

救援包括更不明显的反射以及更多随机混响音。根据所述的实施例,为了容易计算,能量衰变救援以1毫秒长度的时间块进行计算。通过上述功能,确定声能的时间-频率分布,如图7中已经描述的关于步骤500。

[0111] 接着,在图7中已经描述的关于步骤502至步骤506,根据Pearson积差相关(也称为相关性系数)确定相关性测量 $\rho(t)$ 。具体而言,根据如下所示的实施例,在初期状态时整体能量中每个时间块的声能的相关性为:

$$[0112] \quad \rho(t) = \frac{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega)) \cdot \sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))}{\sqrt{\sum_{\omega} (E(1, \omega) - \bar{E}(1, \omega))^2} \cdot \sqrt{\sum_{\omega} (E(t, \omega) - \bar{E}(t, \omega))^2}}$$

[0113] 其中

[0114]  $E(1, \omega)$  = 在频率 $f$ 的全部频率范围的能量衰变救援,

[0115]  $\bar{E}(1, \omega)$  = 初期全部范围的能量衰变救援的所有频率的平均值,

[0116]  $E(t, \omega)$  = 从时间 $t$ 开始在频率 $f$ 的能量衰变救援,

[0117]  $\bar{E}(t, \omega)$  = 在时间 $t$ 开始的全部范围的能量衰变救援的所有频率的平均值,

[0118]  $\omega = 2\pi f$ 。

[0119] 上述相关性描述包括初期状态的衰变的相似性以及在任何时间 $t$ 开始的衰变的相似性。其从宽带EDR计算,使用EDR的全部频率范围用于计算,从而比较完全最初的能量状况与在时间 $t$ 的状况。

[0120] 本发明不限于对所有频率的相关性的计算。相反地,相关性也可对预定义的频率范围计算。频率范围可从待处理的音频信号中确定。例如,对于特定的音频信号,频率范围可限制于预定义的范围,例如,可听见频率的范围。根据实施例,频率范围可为20Hz至20kHz。应注意的是也可选择其他范围,例如根据实证研究。

[0121] 根据实施例,可使用EDR的有效基于FFT的工具。可应用具有测量的脉冲响应的有效长度的窗,而假设测量的脉冲响应的有效长度 $2^{13}$ 等于8192个频率窗。在计算期间,窗移动单时间块的离散长度,而窗的一端为零。根据实施例,使用1ms长度的时间块,而为了EDR的简单以及有效的计算,应用后续的方法:

[0122] (1) 测量的脉冲响应的全部有效长度被带进以计算FFT光谱,而绝对值为对 $E(1, \omega)$ 的平方。

[0123] (2) 窗以1ms的离散时间方块长度向脉冲响应的末端移动,直到到达脉冲响应的末端。窗取样垫零至有效长度(即超过有效长度的取样被设定为零),然后计算FFT光谱产生 $E(1, \omega)$ 。

[0124] 上述方法的优点在于,不需要附加滤波器组或其他相似装置做EDR的窄频带计算;仅需要进行窗的移动。图8显示根据上述基于FFT的方法为了达成脉冲响应的能量衰变救援的示例。

[0125] 如在图7已经描述关于步骤504以及步骤508,根据上述方式确定的相关性与预定义的阈值进行比较。阈值越小,则越多过渡时间向脉冲响应的末端移动。例如,对于双声道脉冲响应,如果阈值选为 $1/e \approx 0.3679$ (也参见现有技术参考文献[2]),因为在第一反射发生或撞击之前相关性已落在阈值下,所以在一些方位角度会太早过渡。然而,因为第一反射是清楚明显的且确定不是后期扩散混响音,所以已知过渡时间必须在第一反射的到达时间



之后。根据实施例,阈值并非定义为固定阈值。相反地,根据本发明方法,阈值的定义取决于在第一反射的撞击时间的相关性。依此定义,可确保第一反射总会在过渡时间之前。根据实施例,如步骤508所示,当以下等式成立时,过渡时间被视为到达:

$$[0126] \quad \rho(t) = c \cdot \rho(t_F)$$

[0127] 其中

[0128]  $\rho(t_F)$  = 早期反射音中选择一个的相关性测量,

[0129]  $t_F$  = 从直达声音撞击之后的早期反射音中选择一个的时间索引,

[0130]  $c$  = 基于  $\frac{1}{e}$  的常数值,  $e$  为欧拉数 (Euler number)。

[0131] 根据实施例,常数值可为  $\frac{1}{e}$ , 然而本发明不限于该数值。根据实施例,常数值可近似于  $\frac{1}{e}$ , 例如在预定义的小数位置 (参见下方) 对  $\frac{1}{e}$  做舍入。

[0132] 在所述的实施例中,  $t_F$  是直达声音撞击后的第一反射的时间块索引。

[0133] 图9显示根据本发明方法的过渡时间确定, 阈值通过在第一反射的撞击点以及  $1/e$  的常数值的相关性的相乘, 取决于脉冲响应来计算。室内脉冲响应600的振幅以取样的数量所表示, 并且标示第一反射602。波形604指示采用公式 (2) 获得的相关性数值。在606, 显示在第一反射的相关性数值, 在示例显示为0.58的数值。而且, 在608显示传统使用的固定阈值  $1/e$ 。用于第一反射的相关性数值606以及原始固定数值  $1/e$  应用至输入乘法器610, 其产生根据在第一反射的相关性数值的新阈值, 其在所述的实施例中显示在612, 为数值0.21。如此, 相比于现有方法, 过渡点614进一步向右边移动, 使得所有在过渡点614之后的取样视为后期混响音304, 而在之前的取样视为早期反射音302。可以看出的是, 导致的决定时间614是鲁棒的。例如, 在双声道室内脉冲响应中, 其代表所计算的过渡时间在方位角更稳定许多。可从图10以及图11的比较看出。图10显示当使用现有技术参考文献[1]所述的方法对左声道700以及右声道702使用EDC工具以及固定阈值  $1/e$  来测量双声道室内脉冲响应的过渡时间。在耳朵以及方位角的相关性清楚可见, 在过渡时间有深降至10ms, 其由于在第一反射撞击的前相关性  $\rho(t)$  落在阈值下面。图11显示根据本发明方法所计算出的左声道700以及右声道702的过渡时间。相比于图10说明的现有方法, 可观察到导致的过渡时间更少取决于耳朵以及方位角。

[0134] 根据实施例, 当相关性第一次落于阈值下方或等于阈值时视为到达过渡时间, 且之后不再增加阈值。在所计算的相关性功能中与取样相关的时间数值是脉冲响应的后期混响音被视为开始的时间。根据本发明的方法, 第一反射的撞击时间可通过执行峰态运算符来决定, 如现有技术参考文献[6]所描述。或者, 第一反射可通过其他方法检测, 例如, 通过阈值检测或, 例如现有技术参考文献[7]所述的通过攻击检测。

[0135] 根据实施例,  $e^{-1} = 0.3679$  用为指示随机流程中低相关性的数值, 例如现有技术参考文献[1]中所标示。根据实施例, 该数值使用小数点后的四位数, 使得  $e^{-1}$  近似0.3679。根据其他实施例, 也可使用更多或更少小数点后的位数, 明显的是所检测的过渡时间根据与  $e^{-1}$  的精确值之间的差异而改变。例如, 当使用数值0.368, 其导致过渡时间的最小改变低于1ms。

[0136] 根据进一步的实施例, 脉冲响应可为有限频带, 而在此情况下, EDR可在有限的频



率范围中计算,也可计算EDR的有限频率范围的相关性。也可使用相替换的频率过渡或滤波器组,例如,完全在FFT域操作的方法,从而当使用基于FFT的滤波/卷积计算时可节省附加转换。

[0137] 应注意的是,上述的实施例已参考第一反射的相关性数值。然而,其他实施例可使用针对早期反射音中的另一个来计算相关性数值。

[0138] 如上所述,根据实施例,本发明的方法可用于双声道处理器中对音频信号作双声道处理。后续的实施例将描述音频信号的双声道处理。双声道处理可执行作为解码器处理,用于将解码信号转换成双声道降混信号,其当以头戴式耳机聆听时可提供环绕声音体验。

[0139] 图12显示根据本发明的实施例的用于音频信号双声道处理的立体声渲染器800的示意图。图12也提供在立体声渲染器中进行QMF域处理的概要。在输入端802,立体声渲染器800接收待处理的音频信号,例如,包括N个声道以及64个QMF频带的输入信号。此外,立体声渲染器800接收用于控制音频信号的处理的多个输入参数。输入参数包括用于 $2 \times N$ 个声道以及64个QMF频带的双声道室内脉冲响应(BRIR)804、最高频带的指示 $K_{\max}$  806,用于音频输入信号与BRIRs 804的早期反射音部卷以及上述的混响器参数808以及810(RT60以及混响音能量)的卷积计算。立体声渲染器800包括快速卷积处理器812,用于对输入音频信号802以及所接收的BRIRs 804的早期部分进行处理。处理器812在输出端产生包括两个声道以及 $K_{\max}$  QMF频带的初期处理信号814。除了具有快速卷积处理器812的初期处理分支,立体声渲染器800也包括包括每个用于接收RT60信息808和混响能量信息810作为输入参数的两个混响器816a与816b的混响音分支。混响音分支进一步包括立体声降混处理器818以及相关分析处理器820,两者都接收音频信号802。此外,两个增益级821a与821b提供在立体声降混处理器818以及各个混响器816a与816b之间,用于控制立体声降混处理器818所提供的降混信号822的增益。立体声降混处理器818基于输入信号802提供具有两个频带以及64个QMF频带的降混信号822。增益级821a与821b的增益通过相关性分析处理器820所提供的各个控制信号824a与824b所控制。增益控制降混信号输入至各个混响器816a与816b以产生各个混响信号826a与826b。早期处理信号814以及混响信号826a与826b由混合器828所接收,其将所接收的信号组合成具有两个声道以及64个QMF频带的输出音频信号830。此外,根据本发明,快速卷积处理器812以及混响器816a与816b接收附加的输入参数832,其指示如上文所述的确定的在室内脉冲响应804中从早期部分到后期混响音的过渡。

[0140] 双声道渲染器模块800(例如,图2或图4的立体声渲染器236)具有作为输入802的解码数据流。通过QMF分析滤波器组处理的信号,在ISO/IEC 14496-3:2009第4.B.18.2小节,其修改的说明在ISO/IEC 14496-3:2009,第8.6.4.2小节。渲染器模块800也可处理QMF域输入数据;在此情况下,可省略分析滤波器组。双声道室内脉冲响应(BRIRs)804以复数值QMF域滤波器来表示。从时域双声道室内脉冲响应到复数值QMF滤波表示的转换概述在ISO/IEC 23003-1:2006附录B。BRIRs 804受限于复数值QMF域中的时隙的特定数量,使得其仅包括早期反射音部301、302(参见图5),而不包括后期扩散混响音304。从早期反射音到后期混响音的过渡点832如上所述来确定,例如,通过在双声道处理的预处理步骤中的BRIRs 804的分析。然后,通过带状快速卷积812处理QMF域音频信号802以及QMF域BRIRs 804,以执行双声道处理。QMF域混响器816a与816b用于产生两声道QMF域后期混响音826a与826b。混响音模块816a与816b使用一组基于频率的混响音时间808以及能量数值810,以适应混响音的

特性。混响音的波形基于音频输入信号802的立体声降混818,其根据多频道音频信号802的相关性分析820适应性地缩放振幅821a与821b。然后,两声道QMF域卷积结果814以及两声道QMF域混响音816a与816b相结合828,最后,两个QMF合成滤波器组计算双声道时域输出信号830,如ISO/IEC 14496-3:2009第4.6.18.4.2小节所概述。渲染器也可产生QMF域输出数据;然后,可省略合成滤波器组。

[0141] 定义

[0142] 馈入双声道渲染器模块800的音频信号802在后续被称为输入信号。双声道处理结果的音频信号830被称为输出信号。双声道渲染器模块800的输入信号802是核心解码器(例如,请参见图2所示的信号228)的音频输出信号。可使用的后续的变量定义为:

[0143]

$N_{in}$	输入声道的数量
$N_{out}$	输出声道的数量, $N_{out}=2$
$M_{DMX}$	包括实数值、非负值的降混系数(降混增益)的降混矩阵。 $M_{DMX}$ 是 $N_{out} \times N_{in}$ 的维度
$L$	在时域音频取样值中测量的帧长度
$v$	时域取样索引
$n$	QMF 时隙索引(次频带取样索引)
$L_n$	在 QMF 时隙中测量的帧长度
$F$	帧索引(帧数量)
$K$	QMF 频带的数量, $K=64$
$k$	QMF 频带索引 (1..64)
$A, B, ch$	声道索引(声道配置的声道数量)
$L_{trans}$	在时域取样值中 BRIR 的早期反射音的长度
$L_{trans,n}$	在 QMF 时隙中 BRIR 的早期反射音的长度
$N_{BRIR}$	在 BRIR 数据组中 BRIR 对的数量
$L_{FFT}$	FFT 转换的长度

[0144]

$\Re(\cdot)$	复值信号的实部
$\Im(\cdot)$	复值信号的虚部
$\mathbf{m}_{\text{conv}}$	在 BRIR 数据组中属于 BRIR 对的输入信号声道的信号向量
$f_{\text{max}}$	用于双声道处理的最高频率
$f_{\text{max,decoder}}$	出现在解码器的音频输出信号中的最高信号频率
$K_{\text{max}}$	用于音频输入信号与 BRIRs 的早期反射音部的卷积处理的最高频带
$a$	降混矩阵系数
$c_{\text{eq},k}$	带状能量等化因子
$\varepsilon$	常数, $\varepsilon = 10^{-20}$
$d$	在 QMF 域时隙中的延迟
$\tilde{\mathbf{y}}_{\text{ch}}^{n',k}$	在频带 $k$ 中的伪 FFT 域信号表示
$n'$	伪 FFT 频率索引
$\tilde{\mathbf{h}}^{n',k}$	在频带 $k$ 中的 BRIR 的伪 FFT 域表示
$\tilde{\mathbf{z}}_{\text{ch,conv}}^{n',k}$	在频带 $k$ 中的伪 FFT 域卷积结果
$\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k}$	中间信号: QMF 域中的两声道卷积结果
$\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k}$	中间信号: QMF 域中的两声道混响音
$K_{\text{ana}}$	分析频带的数量 (用于混响器)
$f_{\text{c,ana}}$	分析频带的中心频率
$N_{\text{DMX,act}}$	降混至立体声降混的一个声道以及在实际的信号帧活动的声道数量
$c_{\text{corr}}$	用于一个信号帧的整体相关性系数
$c_{\text{corr}}^{A,B}$	用于声道 $A, B$ 结合的相关性系数
$\sigma_{\tilde{\mathbf{y}}_{\text{ch},A}^n}$	信号 $\tilde{\mathbf{y}}_{\text{ch},A}^n$ 的时隙的标准差
$\mathbf{c}_{\text{scale}}$	两个缩放因子的向量
$\tilde{\mathbf{c}}_{\text{scale}}$	两个比例因子的向量, 经过时间上平滑处理

[0145] 处理

[0146] 以下描述输入信号的处理。双声道渲染器模块在输入音频信号的长度  $L=2048$  个时域取样的邻近未重迭的帧上运作, 并在针对每个模块处理长度  $L$  的输入帧输出  $L$  个取样的帧。

[0147] (1) 初始化以及预处理

[0148] 在开始处理核心解码器 (例如, 参见图2所示的解码器200) 传送的音频取样之前, 进行双声道处理方块的初始化。初始化包括几个处理步骤。

[0149] (a) 读取分析数值

[0150] 混响器模块816a与816b采取混响音时间808以及能量数值810的频率相关集作为输入参数。这些数值从双声道处理模块800的初始化界面读取。此外, 读取在时域取样值中从早期反射音到后期混响音的过渡时间832。对于每个取样, 这些数值可储存在二进制文件中, 其以32位、浮动数值、little-endian顺序写入。处理所需的读取数值在下面的表格描述:

[0151]

数值描述	数量	数据类型
过渡长度 $L_{trans}$	1	整数 (integer)
频带 $K_{ana}$ 的数量	1	整数 (integer)
频带的中心频率 $f_{c,ana}$	$K_{ana}$	浮点数 (float)
混响音时间 RT60 (秒)	$K_{ana}$	浮点数 (float)
代表一个 BRIR 的后期混响音部的能量 (二幂次的振幅)的能量数值	$K_{ana}$	浮点数 (float)

[0152] (b) 读取以及预处理BRIRs

[0153] 从分别存储左耳和右耳BRIRs的两个专用文件中读取双声道室内脉冲响应804。BRIRs的时域取样值储存在整数波形文件(wave-file)中,且波形文件(wave-file)中每个取样值具有24位的分辨率以及32个声道。文件中的BRIRs的顺序在后续的表格描述:

[0154]

声道数量	扬声器标签
1	CH M L045
2	CH M R045
3	CH M 000
4	CH LFE1
5	CH M L135
6	CH M R135
7	CH M L030
8	CH M R030
9	CH M 180
10	CH LFE2
11	CH M L090
12	CH M R090
13	CH U L045
14	CH U R045
15	CH U 000
16	CH T 000
17	CH U L135
18	CH U R135
19	CH U L090
20	CH U R090
21	CH U 180
22	CH L 000
23	CH L L045
24	CH L R045
25	CH M L060
26	CH M R060
27	CH M L110
28	CH M R110
29	CH U L030
30	CH U R030
31	CH U L110
32	CH U R110

[0155]

[0156] 如果没有BRIR在扬声器位置中的其中一个被测量,则波形文件中相对应的声道包括零值。LFE声道不用于双声道处理。

[0157] 在预处理步骤,双声道室内脉冲响应(BRIRs)的给定集从时域滤波器转换成复数值QMF域滤波器。在复数值QMF域中的给定时域滤波器的执行是根据ISO/IEC FDIS23003-1:2006Annex B执行。根据ISO/IEC FDIS23003-1:2006,附加B,表格B.1,使用滤波器转换的原型滤波器系数。对时域表示  $\tilde{\mathbf{h}}_{\text{ch}}^v = [\tilde{h}_1^v \dots \tilde{h}_{N_{\text{BRIR}}}^v]$ ,  $1 \leq v \leq L_{\text{trans}}$  被处理以得到复数值QMF域滤波器  $\hat{\mathbf{h}}_{\text{ch}}^{n,k} = [\hat{h}_1^{n,k} \dots \hat{h}_{N_{\text{BRIR}}}^{n,k}]$ ,  $1 \leq n \leq L_{\text{trans},n}$ 。

[0158] (2) 音频信号处理

[0159] 双声道渲染器模块800的音频处理方块从核心解码器获得用于 $N_{\text{in}}$ 输入声道的时域音频取样值802,并产生由 $N_{\text{out}}=2$ 声道组成的双声道输出信号830。

[0160] 处理以下列数据作为输入

- [0161] • 来自核心解码器的解码音频数据802,
- [0162] • BRIR集804的早期反射音部的复数QMF域表示,以及
- [0163] • 频率相关参数集808、810、832,其由QMF域混响器816a与816b使用以产生后期混响音826a与826b。

[0165] (a) 音频信号的QMF分析

[0166] 在第一处理步骤,双声道渲染器模块将 $N_{\text{in}}$ 声道时域输入信号(来自核心解码器)  $[\tilde{\mathbf{y}}_{\text{ch},1}^v \dots \tilde{\mathbf{y}}_{\text{ch},N_{\text{in}}}^v] = \tilde{\mathbf{y}}_{\text{ch}}^v$  的 $L=2048$ 个时域取样值转换成维度 $L_n=32$ 个QMF时隙(时隙索引 $n$ )和 $K=64$ 个频带(频带索引 $K$ )的 $N_{\text{in}}$ 声道QMF域信号表示802。

[0167] 如ISO/IEC 14496-3:2009第4.B.18.2小节所述的QMF分析,以及在ISO/IEC 14496-3:2009第8.6.4.2小节所述的修改,在时域信号 $\tilde{\mathbf{y}}_{\text{ch}}^v$ 的帧上执行以得到QMF域信号  $[\hat{\mathbf{y}}_{\text{ch},1}^{n,k} \dots \hat{\mathbf{y}}_{\text{ch},N_{\text{in}}}^{n,k}] = \hat{\mathbf{y}}_{\text{ch}}^{n,k}$  ( $1 \leq v \leq L$  且  $1 \leq n \leq L_n$ ) 的帧。

[0168] (b) QMF域音频信号以及QMF域BRIRs的快速卷积计算

[0169] 接着,执行带状快速卷积812以处理QMF域音频信号802以及QMF域BRIRs 804。可对输入信号802以及每个BRIR 804的每个声道的每个QMF频带执行FFT分析。

[0170] 由于QMF域的复数值,一个FFT分析是在QMF域信号表示的实部上执行以及一个FFT分析是在QMF域信号表示的虚部上执行。然后,结合结果以形成最后的带状复数值伪FFT域信号

$$[0171] \quad \tilde{\mathbf{y}}_{\text{ch}}^{n',k} = \text{FFT}(\hat{\mathbf{y}}_{\text{ch}}^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{y}}_{\text{ch}}^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{y}}_{\text{ch}}^{n',k}))$$

[0172] 以及带状复数值BRIRs

$$[0173] \quad \text{针对左耳} \quad \tilde{\mathbf{h}}_1^{n',k} = \text{FFT}(\hat{\mathbf{h}}_1^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{h}}_1^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{h}}_1^{n',k}))$$

$$[0174] \quad \text{针对右耳} \quad \tilde{\mathbf{h}}_2^{n',k} = \text{FFT}(\hat{\mathbf{h}}_2^{n',k}) = \text{FFT}(\Re(\hat{\mathbf{h}}_2^{n',k})) + j \cdot \text{FFT}(\Im(\hat{\mathbf{h}}_2^{n',k}))。$$

[0175] 根据复数值QMF域BRIR滤波器 $L_{\text{trans},n}$ 的长度以及QMF域时隙 $L_n$ 的帧长度决定FFT转换的长度,使得

[0176]  $L_{\text{FFT}} = L_{\text{trans},n} + L_n - 1$ 。

[0177] 然后,将复数值伪FFT域信号与复数值伪FFT域信号BRIR滤波器相乘以形成快速卷积结果。向量 $\mathbf{m}_{\text{conv}}$ 用于对应BRIR数据集中的BRIR对的输入信号的声道的信号。针对所有QMF频带 $k$  ( $1 \leq k \leq K_{\text{max}}$ ) 带状完成相乘处理。通过代表18kHz或出现在来自核心解码器的音频信号中的最大信号频率的QMF频带来确定最高频带 $K_{\text{max}}$ ,  $f_{\text{max}} = \min(f_{\text{max},\text{decoder}}, 18\text{kHz})$ 。

[0178] 在每个QMF频带 $k$  ( $1 \leq k \leq K_{\text{max}}$ ), 加总每个音频输入声道与每个BRIR对的相乘结果, 以产生中间两声道 $K_{\text{max}}$ 频带伪FFT域信号。

[0179]  $\tilde{\mathbf{z}}_{\text{ch},1,\text{conv}}^{n,k} = \sum_{ch=1}^{ch=N_{\text{in}}} \tilde{\mathbf{y}}_{\text{ch},ch}^{n,k} \cdot \tilde{\mathbf{h}}_{1,\mathbf{m}_{\text{conv}}[ch]}^{n,k}$  和  $\tilde{\mathbf{z}}_{\text{ch},2,\text{conv}}^{n,k} = \sum_{ch=1}^{ch=N_{\text{in}}} \tilde{\mathbf{y}}_{\text{ch},ch}^{n,k} \cdot \tilde{\mathbf{h}}_{2,\mathbf{m}_{\text{conv}}[ch]}^{n,k}$  是在QMF域频带 $k$ 的伪FFT

卷积结果  $\tilde{\mathbf{z}}_{\text{ch},\text{conv}}^{n,k} = \begin{bmatrix} \tilde{\mathbf{z}}_{\text{ch},1,\text{conv}}^{n,k} & \tilde{\mathbf{z}}_{\text{ch},2,\text{conv}}^{n,k} \end{bmatrix}$ 。

[0180] 接着,执行带状FFT合成将卷积结果转换回QMF域,以产生具有 $L_{\text{FFT}}$ 个时隙

[0181]  $\hat{\mathbf{z}}_{\text{ch},\text{conv}}^{n,k} = \begin{bmatrix} \hat{\mathbf{z}}_{\text{ch},1,\text{conv}}^{n,k} & \hat{\mathbf{z}}_{\text{ch},2,\text{conv}}^{n,k} \end{bmatrix}$  ( $1 \leq n \leq L_{\text{FFT}}$  且  $1 \leq k \leq K_{\text{max}}$ ) 的中间双声道 $K_{\text{max}}$ 频带QMF域信号。

[0182] 针对每个具有 $L=32$ 个时隙的QMF域输入信号帧,返回具有 $L=32$ 个时隙的卷积结果信号帧。剩余的 $L_{\text{FFT}}-32$ 个时隙储存以及在后续的帧执行重迭相加处理。

[0183] (c)产生后期混响音

[0184] 第二中间信号826a与826b,又称混响音信号,通过频域混响器模块816a与816b产生  $\hat{\mathbf{z}}_{\text{ch},\text{rev}}^{n,k} = \begin{bmatrix} \hat{\mathbf{z}}_{\text{ch},1,\text{rev}}^{n,k} & \hat{\mathbf{z}}_{\text{ch},2,\text{rev}}^{n,k} \end{bmatrix}$ 。频域混响器816a与816b以下列数据作为输入:

[0185] • 输入信号的一个帧的QMF域立体声降混822,

[0186] • 包括频率相关混响音时间808以及能量值810的参数集。

[0187] 频域混响器816a与816b回传两声道QMF域后期混响音尾部。

[0188] 根据最高频率计算频率相关参数集的最高使用频带数量。

[0189] 首先,通过输入信号声道的加权总和,执行输入信号  $\hat{\mathbf{y}}_{\text{ch}}^{n,k}$  的一个帧的QMF域立体声降混818以形成混响器的输入。降混矩阵 $\mathbf{M}_{\text{DMX}}$ 包括权重增益。其是实数值以及非负值,而降混矩阵的维度为 $N_{\text{out}} \times N_{\text{in}}$ 。其包括非零数值,其中输入信号的声道映像至两个输出声道中的其中一个。

[0190] 代表左半边扬声器位置的声道映像至左输出声道,而代表位于右半边上的声道映射至右输出声道。这些声道的信号以系数1做加权。代表扬声器在中间平面的声道映像至双声道信号的两个输出声道。这些声道的输入信号以系数

[0191]  $a = 0.7071 \approx \frac{1}{\sqrt{2}}$

[0192] 做加权。

[0193] 此外,在降混中执行能量等化步骤。其适应一个降混声道的带状能量,使其等于包括在降混声道中的输入信号声道的带状能量的总和。通过与实数系数带状相乘以产生能量等化

$$[0194] \quad c_{eq,k} = \sqrt{\frac{P_{in}^k}{P_{out}^k + \varepsilon}} \circ$$

[0195] 因子 $c_{eq,k}$ 受限在 $[0.5, 2]$ 的区间内。引进常数 $\varepsilon$ 以避免除零的情形发生。降混也在频率 $f_{max}$ 上限制带宽；所有高频频带的数值设定为零。

[0196] 图13大略地表示根据本发明的实施例的双声道渲染器800的频域混响器816a与816b的处理。

[0197] 在频域混响器，使用输入混合器900来计算立体声输入的单声道降混。其在第二输入声道上不相关地应用 $90^\circ$ 相移来完成。

[0198] 然后，在每个频率频带 $k$ ，单声道信号送入反馈延迟回路902，以产生脉冲的衰减序列。接着，相并联的FIR解相关器以衰减方式分配信号能量至脉冲之间的间隔中，并建立输出声道之间的非相关性。衰减滤波振实密度应用以建立能量衰减。滤波振实阶段操作受限四个选项以实施稀疏且无乘法器的解相关器。

[0199] 在混响音的计算之后，声道间相干 (ICC) 修正904被包括至每个QMF频带的混响器模块。在ICC修正步骤中，频率相关直接增益 $g_{direct}$ 和交混增益 $g_{cross}$ 用于适应该ICC。

[0200] 针对不同频带，能量以及混响音时间的数量包含在输入参数集中。数值给定在频率点的数量，其内部地映射至 $K=64$ 个QMF频带。

[0201] 使用两个频域混响器的实例以计算最后的中间信号 $\hat{\mathbf{z}}_{ch,rev}^{n,k} = [\hat{z}_{ch,1,rev}^{n,k}, \hat{z}_{ch,2,rev}^{n,k}]$ 。信号 $\hat{z}_{ch,1,rev}^{n,k}$ 是混响器的第一实例的第一输出声道，而 $\hat{z}_{ch,2,rev}^{n,k}$ 是混响器的第二实例的第二输出声道。其结合至具有2个声道、64个频带以及32个时隙的维度的最后混响音信号帧。

[0202] 根据输入信号帧的相关性测量820，立体声降混822进行两个时间缩放821a与821b，以确保混响器输出的正确的缩放比例。比例因子定义为 $[\sqrt{N_{DMX,act}}, N_{DMX,act}]$ 的间隔内的值，其线性决定于在0以及1之间的相关性系数 $c_{corr}$ 的数值，

$$[0203] \quad c_{corr} = \frac{1}{N_{in}^2} \cdot \sum_{A=1}^{A=N_{DMX,act}} \sum_{B=1}^{B=N_{DMX,act}} c_{corr}^{A,B} \text{ 并且}$$

$$[0204] \quad c_{corr}^{A,B} = \left| \frac{1}{K-1} \cdot \frac{\sum_k \sum_n \hat{y}_{ch,A}^{n,k} \cdot \hat{y}_{ch,B}^{n,k*}}{\sum_n \sigma_{\hat{y}_{ch,A}^n} \cdot \sigma_{\hat{y}_{ch,B}^n}} \right|$$

[0205] 其中 $\sigma_{\hat{y}_{ch,A}^n}$ 表示跨过声道A的一个时隙 $n$ 的标准差，运算符 $\{^*\}$ 表示复共轭而 $\hat{y}$ 为实际的信号帧中QMF域信号的零平均值。

[0206]  $c_{corr}$ 计算两次：一次是包括在立体声降混的左声道的所有在实际的信号帧 $F$ 活跃的声道A,B，而另一次是包括在立体声降混的右声道的所有在实际的信号帧 $F$ 活跃的声道A,B。

[0207]  $N_{DMX,act}$ 是输入声道的数量，其降混至一个降混声道A(在降混矩阵 $M_{DMX}$ 的第Ath列中的矩阵元素的数量不等于零)，而且在目前帧活跃。

[0208] 然后比例因子是：

$$\begin{aligned}
 \mathbf{c}_{\text{scale}} &= [c_{\text{scale},1}, c_{\text{scale},2}] \\
 [0209] \quad &= \left[ \sqrt{N_{\text{DMX,act},1}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},1} - \sqrt{N_{\text{DMX,act},1}}), \sqrt{N_{\text{DMX,act},2}} + c_{\text{corr}} \cdot (N_{\text{DMX,act},2} - \sqrt{N_{\text{DMX,act},2}}) \right]^{\circ}
 \end{aligned}$$

[0210] 通过第一级低通滤波器将比例因子平滑于音频信号帧,该第一级低通滤波器导致平滑的比例因子  $\tilde{\mathbf{c}}_{\text{scale}} = [\tilde{c}_{\text{scale},1}, \tilde{c}_{\text{scale},2}]^{\circ}$ 。

[0211] 以相同方式,通过时域相关性分析在第一音频输入数据帧中初始化比例因子。

[0212] 第一混响器实例的输入以比例因子  $\tilde{c}_{\text{scale},1}$  缩放,而第二混响器实例的输入以比例因子  $\tilde{c}_{\text{scale},2}$  缩放。

[0213] (d) 结合卷积结果以及后期混响音

[0214] 接着,通过混合处理828,其带状相加两个信号,以结合卷积结果814,  $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} = [\hat{z}_{\text{ch},1,\text{conv}}^{n,k}, \hat{z}_{\text{ch},2,\text{conv}}^{n,k}]$ ,以及用于一个QMF域音频输入帧的混响器826a、826b的输出  $\hat{\mathbf{z}}_{\text{ch,rev}}^{n,k} = [\hat{z}_{\text{ch},1,\text{rev}}^{n,k}, \hat{z}_{\text{ch},2,\text{rev}}^{n,k}]$ 。应注意的是,因为卷积仅产生在上至  $K_{\text{max}}$  的频带,所以在高于  $K_{\text{max}}$  的上频带在  $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k}$  中为零。

[0215] 通过混合处理中的时隙数量  $d = ((L_{\text{trans}} - 20 \cdot 64 + 1) / 64 + 0.5) + 1$  延迟后期混响音输出。

[0216] 延迟  $d$  考虑到BRIRs中从早期反射音道后期反射音的过渡时间,以及20个QMF时隙的混响器的最初延迟,和用于BRIRs的QMF分析的0.5个QMF时隙的分析延迟,以确保后期混响音插入在合理的时隙。通过  $\hat{\mathbf{z}}_{\text{ch,conv}}^{n,k} + \hat{\mathbf{z}}_{\text{ch,rev}}^{n-d,k}$  计算在一个时隙  $n$  中的结合信号  $\hat{\mathbf{z}}_{\text{ch}}^{n,k}$ 。

[0217] (e) 双声道QMF域信号的QMF合成

[0218] 根据ISO/IEC 14496-3:2009第4.6.18.4.2小节,通过QMF合成,将QMF域输出信号  $\hat{\mathbf{z}}_{\text{ch}}^{n,k}$  的32个时隙的一个两声道帧转换成具有长度  $L$  的两声道时域信号帧。最后产生时域输出信号830,  $\tilde{\mathbf{z}}_{\text{ch}}^v = [\tilde{z}_{\text{ch},1}^v \cdots \tilde{z}_{\text{ch},2}^v]$ 。

[0219] 虽然一些方面已经在装置的内容中描述,清楚的是这些方面也代表相对应的方法的描述,而方块或装置对应方法步骤或方法步骤的特征。同样地,在方法步骤的内容中描述的方面也代表相对应的方块或项目或相对应装置的特征的描述。一些或所有的方法步骤可通过(或使用)硬件装置执行,例如像是微处理器、可编程计算机或电子电路。在一些实施例中,一些最重要的方法步骤可通过该种装置执行。

[0220] 取决于特定的实施需求,本发明的实施例可在硬件或在软件上实施。该实施可使用非暂时性储存媒体,例如数字储存媒体,例如储存有电子可读取控制信号的软盘、DVD、蓝光(Blu-Ray)、CD、ROM、PROM以及EPROM、EEPROM或FLASH内存,其能与可编程计算机系统合作(或能够配合)以执行上述方法。因此,数字储存媒体是计算机可读取的。

[0221] 根据本发明的一些实施例包括具有电子可读取控制信号的数据载体,其能够与可编程计算机系统配合,以执行上述方法中的其中一个。

[0222] 一般而言,本发明的实施例可实施为具有程序代码的计算机程序产品,当计算机程序产品在计算机上执行时该程序代码操作以执行上述方法中的其中一个。例如,该程序代码可储存在机器可读载体上。



[0223] 其他实施例包括用于执行上述方法中的其中一个的计算机程序,其储存在机器可读取载体上。

[0224] 换句话说,因此发明的方法的实施例为具有当此计算机程序在计算机上执行时,能执行上述方法中的其中一个的程序代码的计算机程序。

[0225] 因此,本发明的方法的另一实施例数据载体(或数字储存媒体或计算机可读取媒体)包括纪录用于执行上述方法中的其中一个的计算机程序。数据载体,数字储存媒体或纪录媒体为有形实体及/或非暂时性的。

[0226] 因此,本发明的方法的另一实施例为数据流或串信号,其代表用于执行上述方法中的其中一个的计算机程序。例如,数据流或串信号可配置经由数据通信连接传输,例如因特网。

[0227] 另一实施例包括处理方式,例如,计算机或可编程逻辑设备,被配置用于或被程序化用于执行上述方法中的其中一个。

[0228] 另一实施例包括安装有用于执行上述方法中的其中一个的计算机程序的计算机。

[0229] 根据本发明的另一实施例包括用于传输(例如电性或光学)用于执行上述方法中的其中一个的计算机程序到接收器的装置或系统。例如,接收器可为计算机、移动装置、内存装置或其他相似装置。例如,该装置或系统可包括用于传输计算机程序至接收器的文件服务器。

[0230] 在一些实施例中,可编程逻辑设备(例如现场可编程门阵列)可用于执行上述方法的部分或全部功能。在一些实施例中,为了执行上述方法中的其中一个,现场可编程门阵列可配合微处理器。通常,该方法可通过任何硬件装置较佳执行。

[0231] 虽然本发明描述了数个实施例,但对其进行变更、置换及等同均落入本发明的围之内。还有应当注意的是,有很多替换本发明的实施方法及组成的方式。因此,下文所附的权利项应当被理解为包含所有此类的变更、置换及等同,这些均未脱离实施例的精神与范畴。

[0232] 参考文献:

[0233] [1]T.Hidaka et al:“A new definition of boundary point between early reflections and late reverberation in room impulse responses”.Forum Acusticum, 2005.

[0234] [2]Jot et al:“Analysis and synthesis of room reverberation based on a statistical timefrequency model”.

[0235] [3]J.S.Abel,P.Huang:“A Simple,Robust Measure of Reverberation Echo Density”.AES Convention,San Francisco,2006.

[0236] [4]R.Stewart,M.Sandler:“Statistical Measures of Early Reflections of Room Impulse Responses”.DAFx,2007.

[0237] [5]Reilly et al:“Using Auralisation for Creating Animated 3-D Sound Fields Across Multiple Speakers”.AES Convention,New York,1995.

[0238] [6]Usher,J.:“An improved method to determine the onset timings of reflections in an acoustic impulse response”.Journal of the Acoustical Society of America,(2010,volume 127)band 4,p.172-177.

[0239] [7]Masri,P.:”Computer Modelling of Sound for Transformation and Synthesis ofMusical Signals“.PhD thesis,University of Bristol,1996.

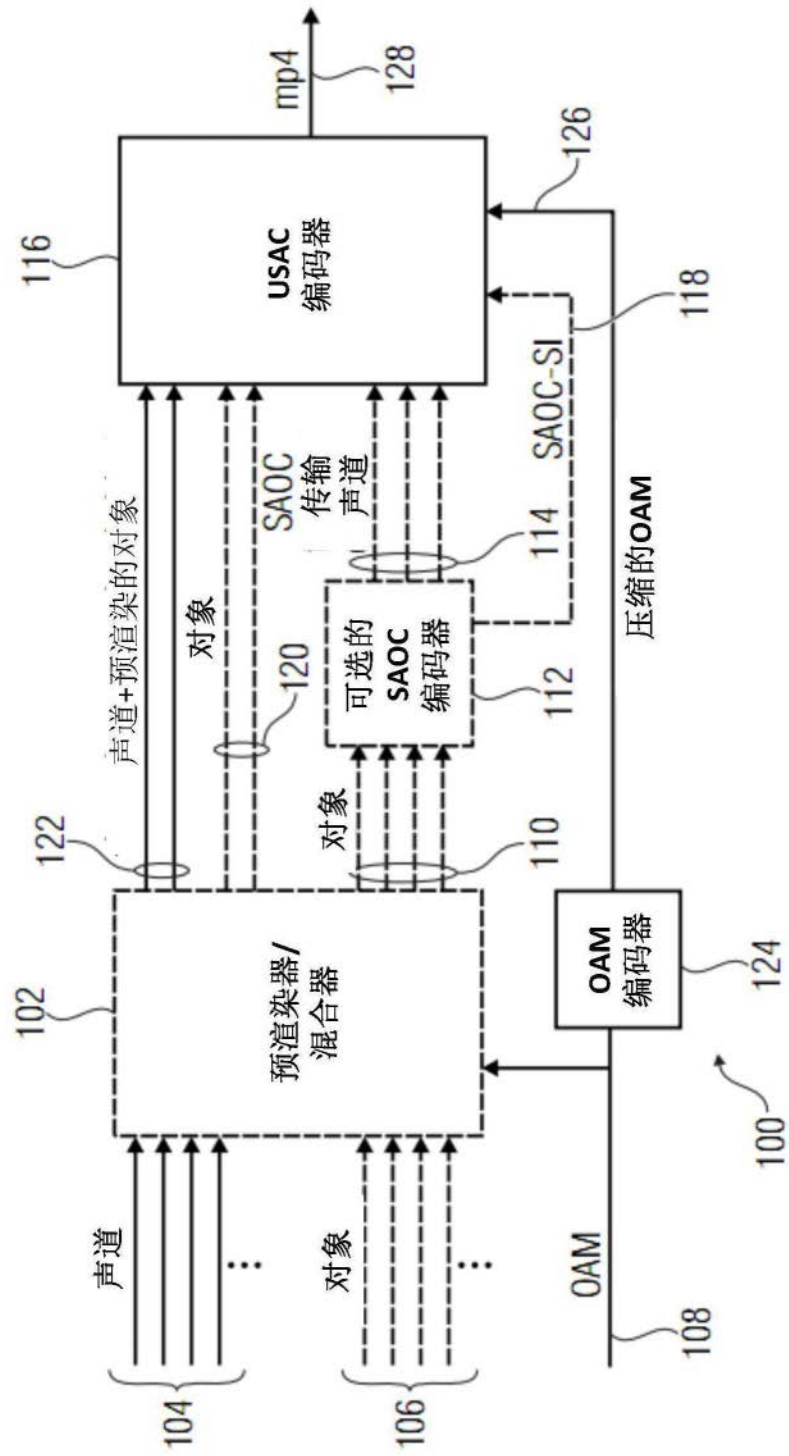


图1

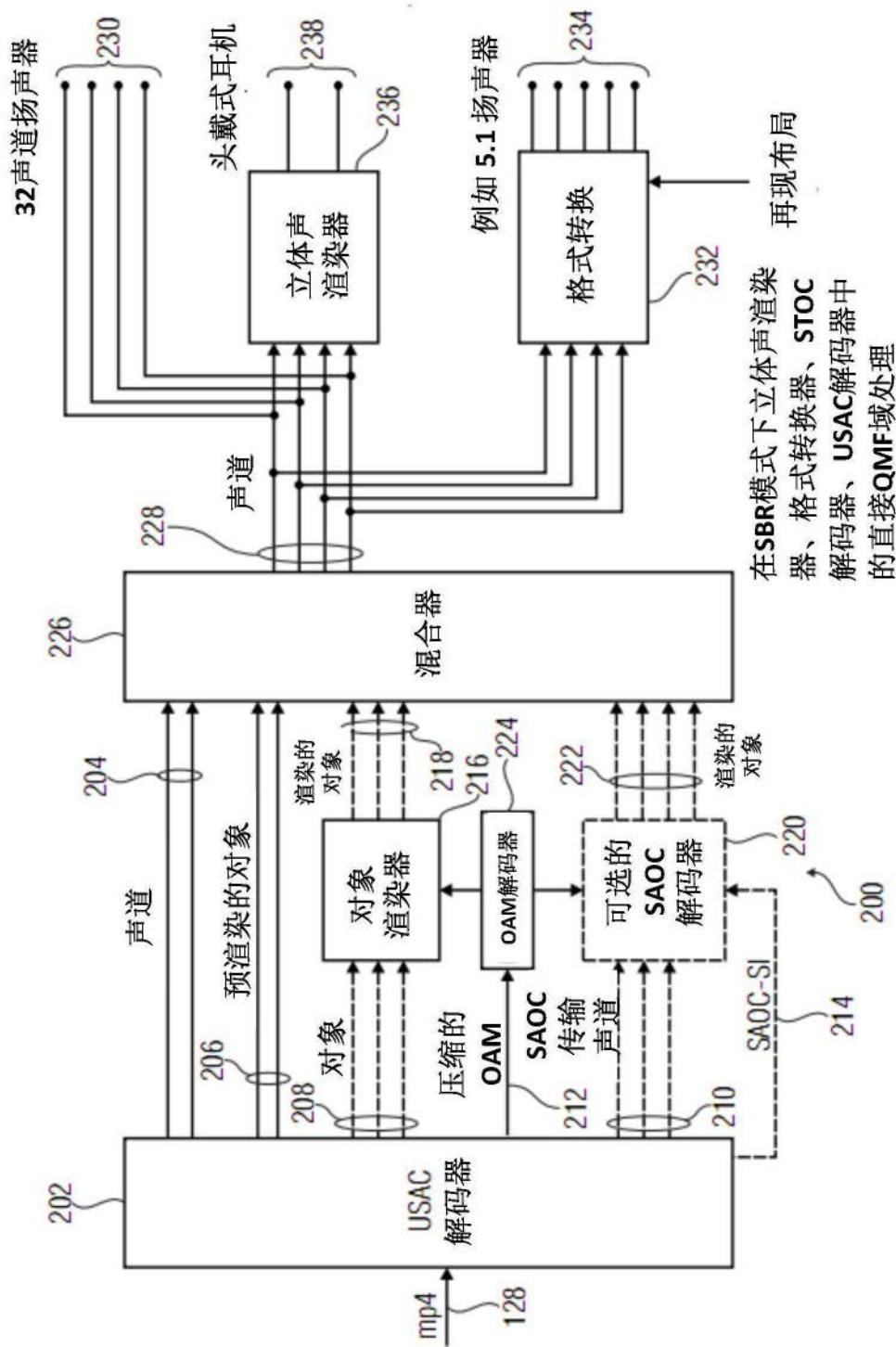


图2

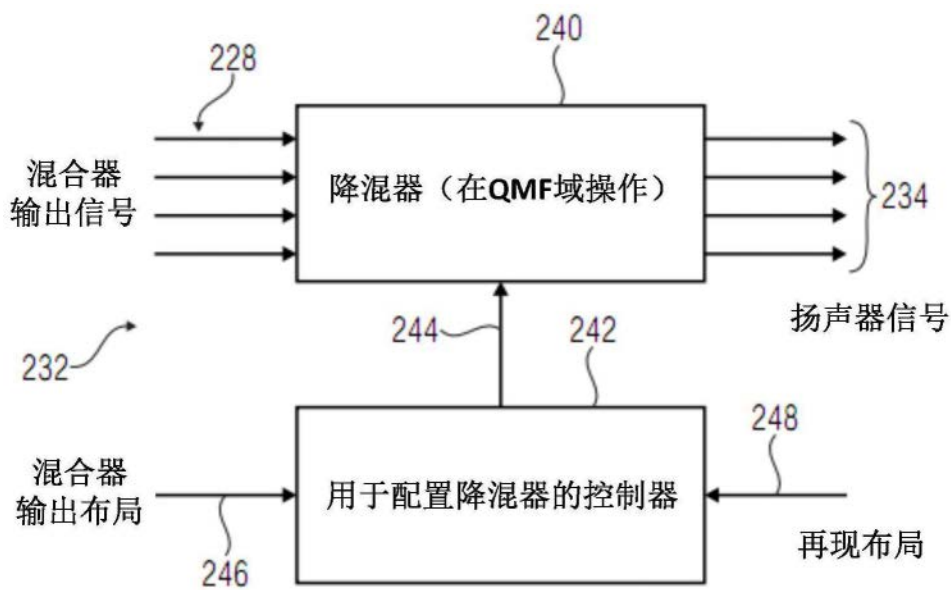


图3

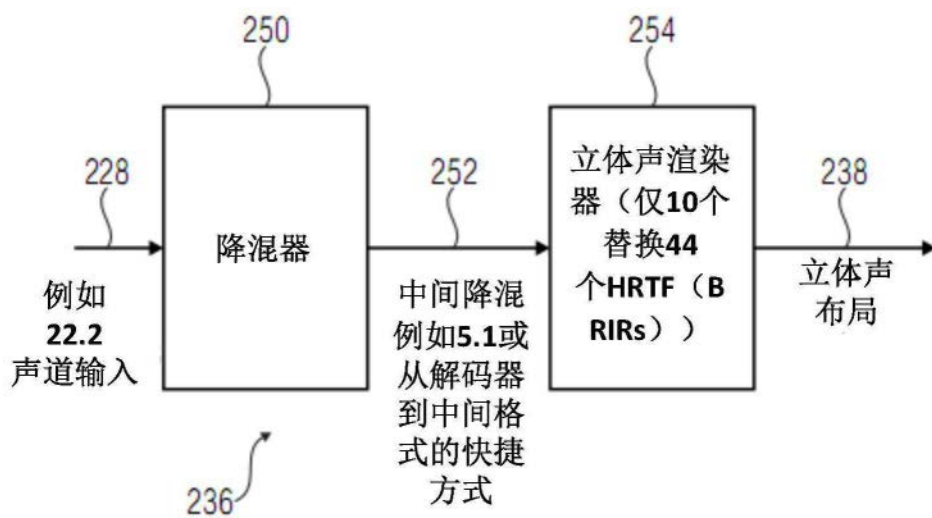


图4

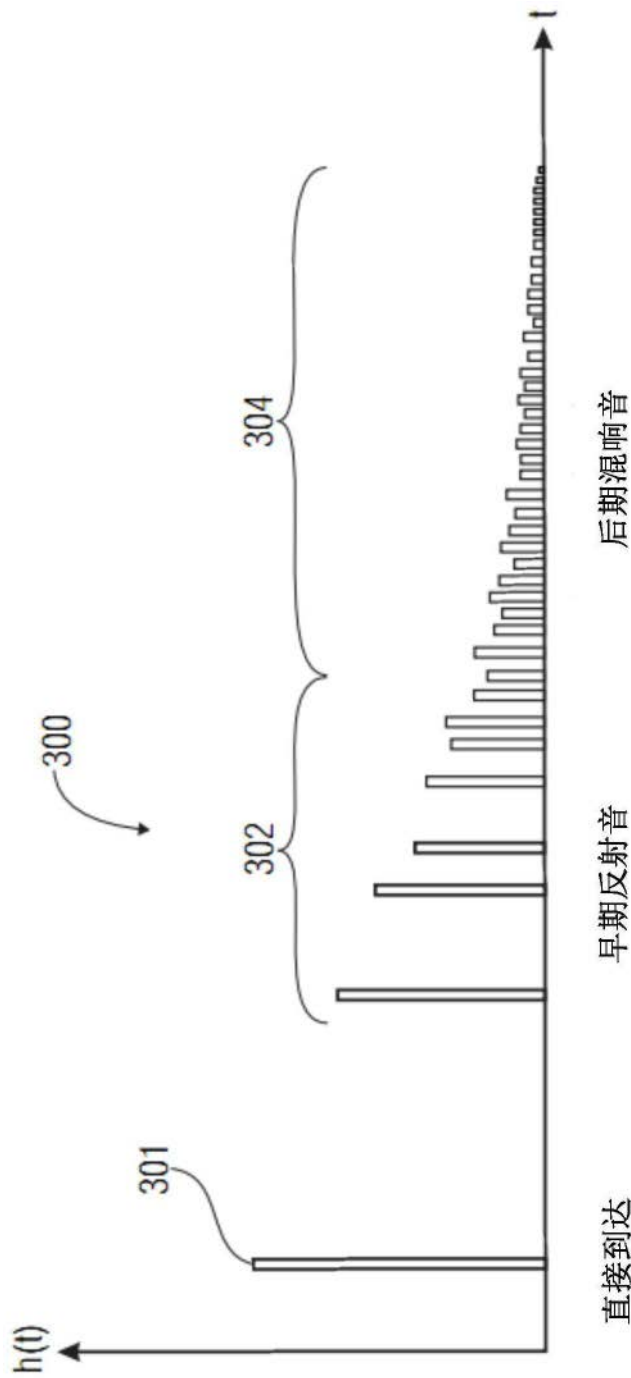


图5

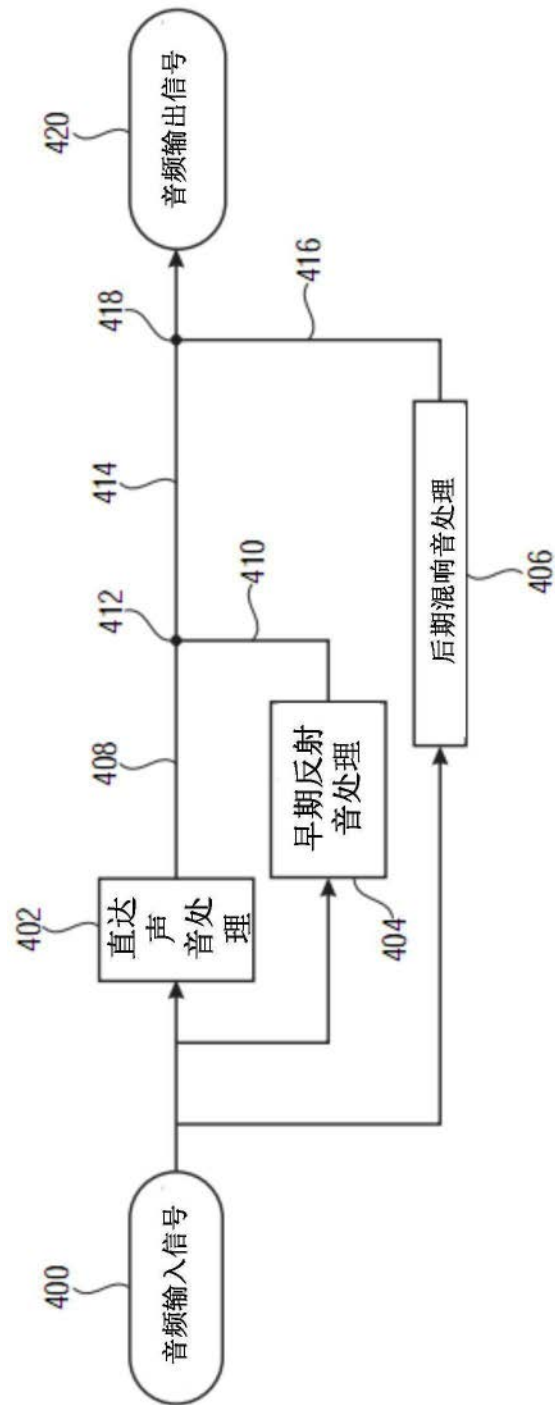


图6A

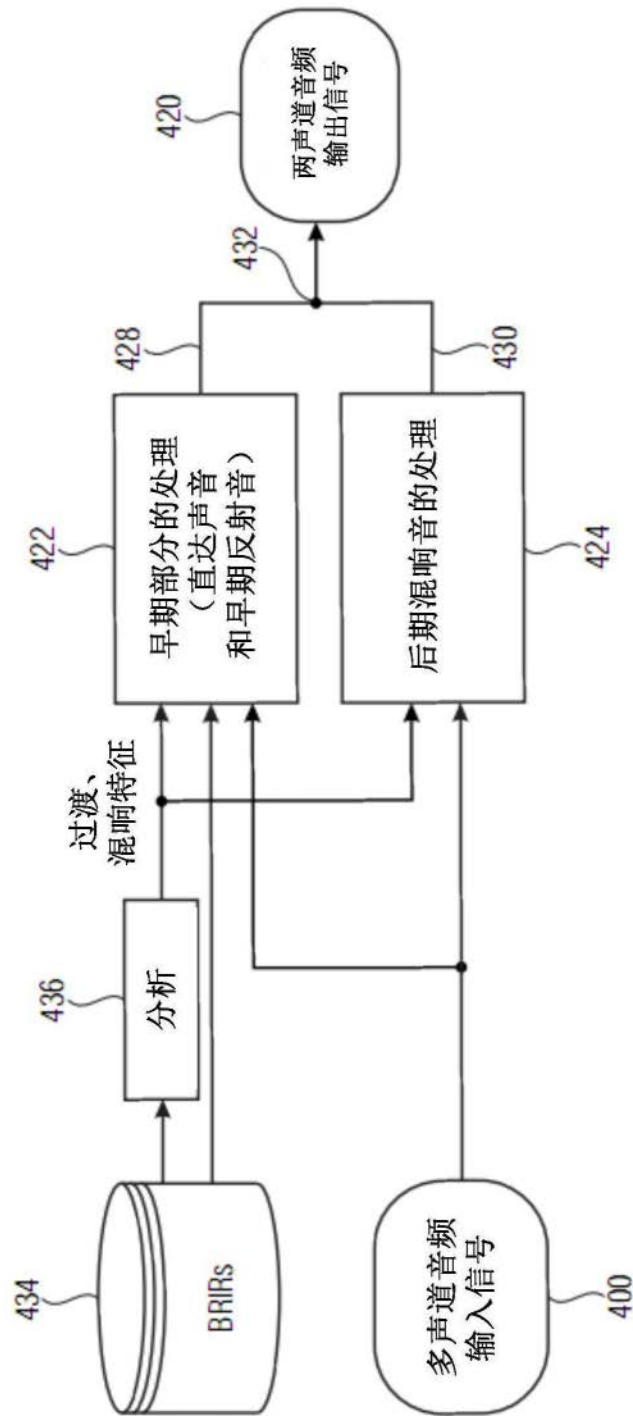


图6B



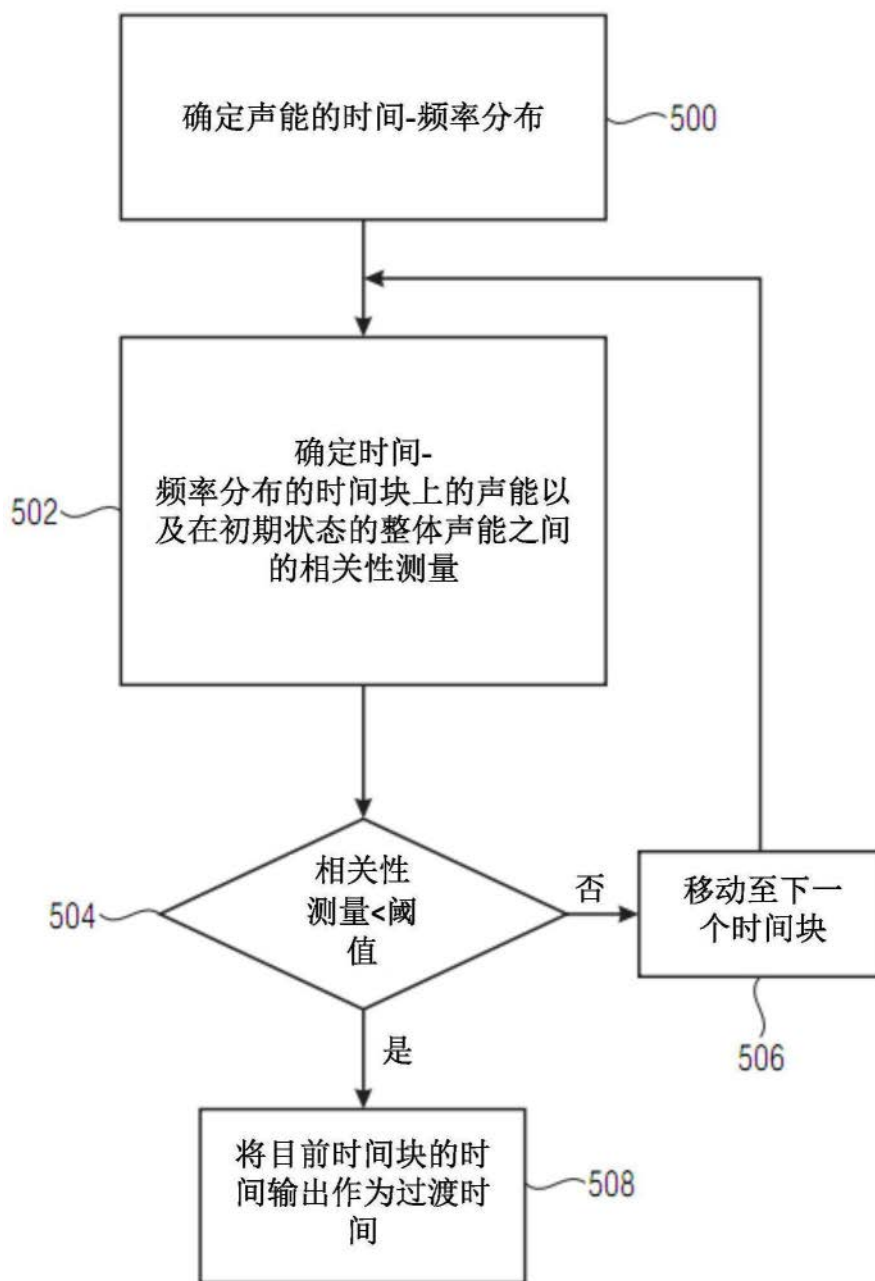


图7

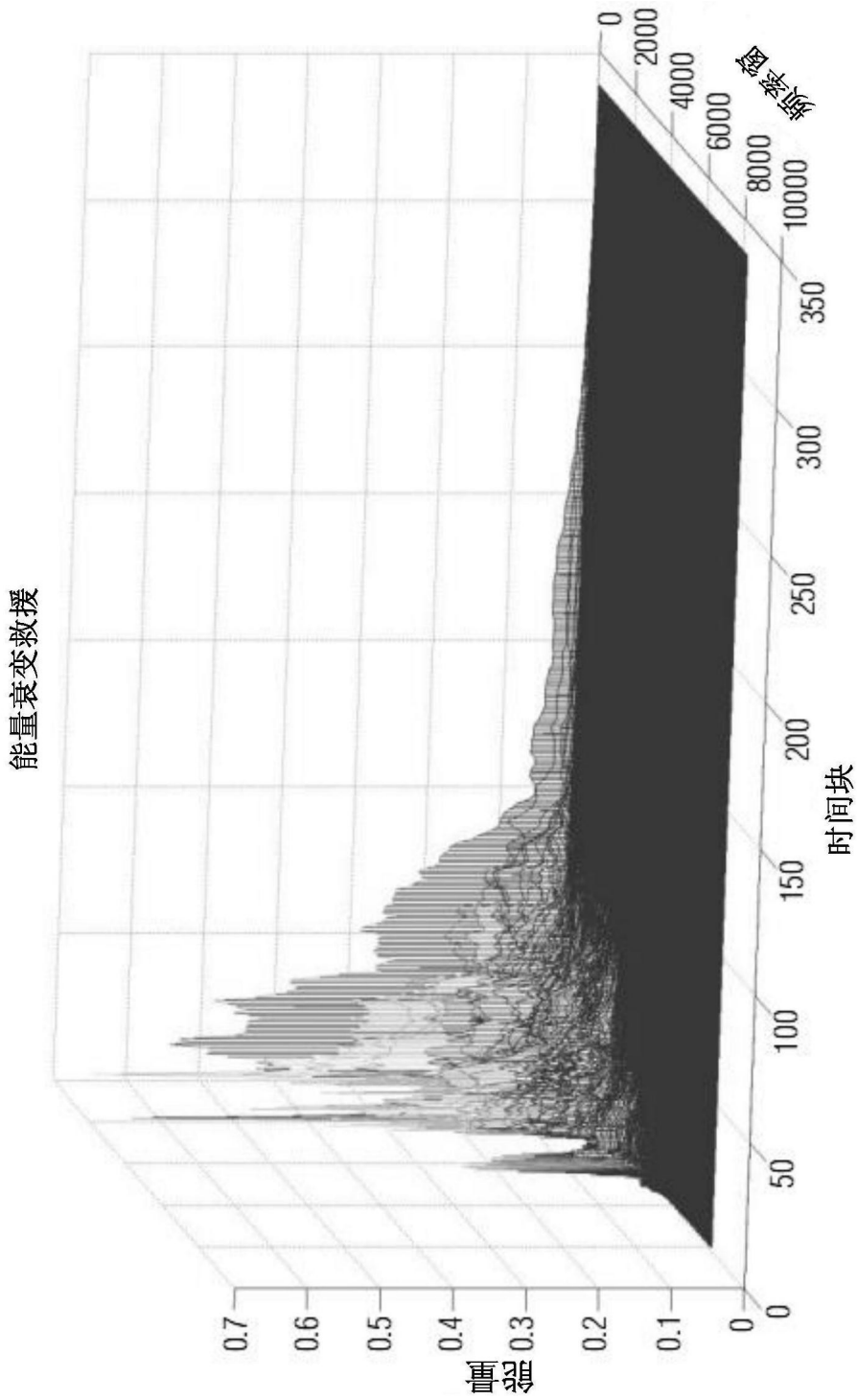


图8

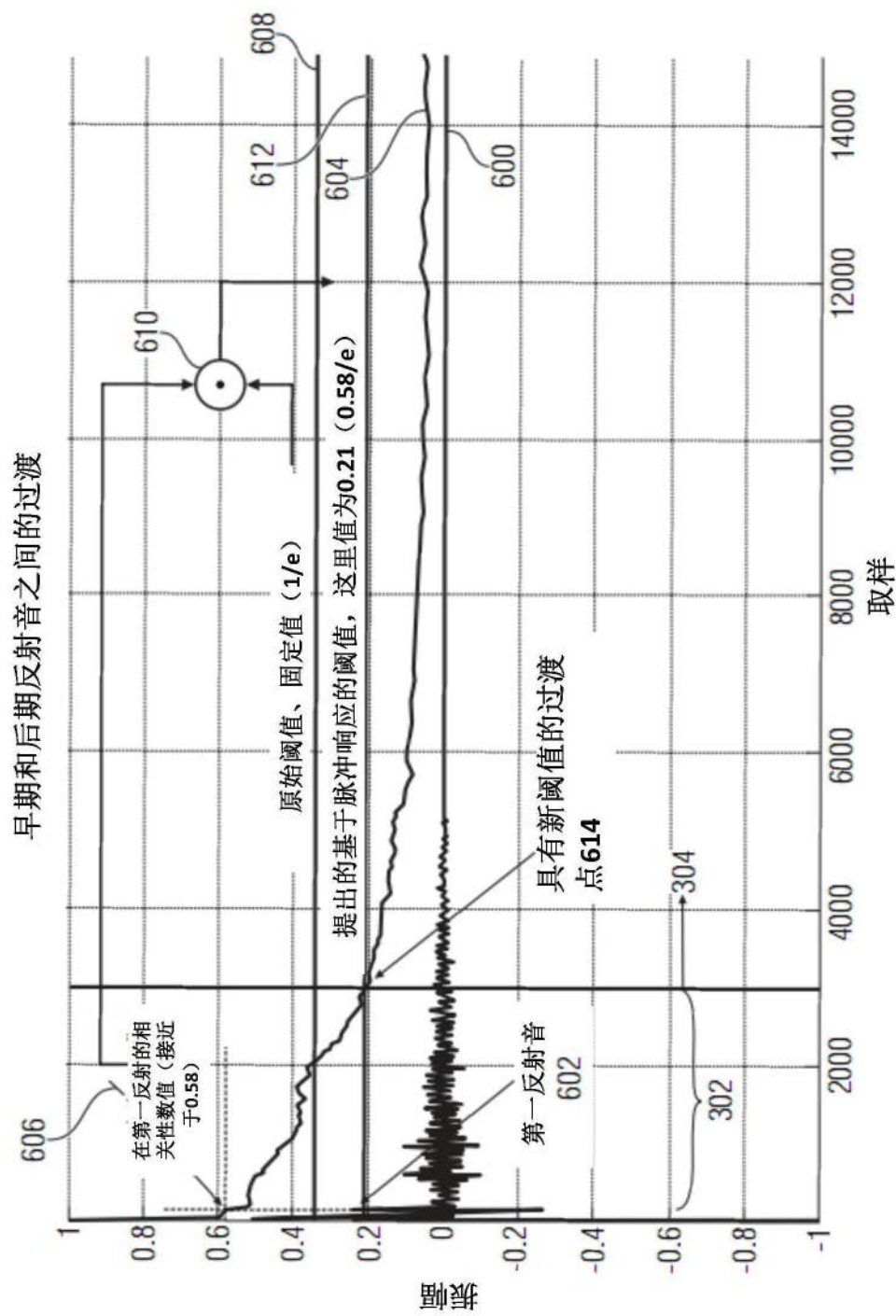


图9

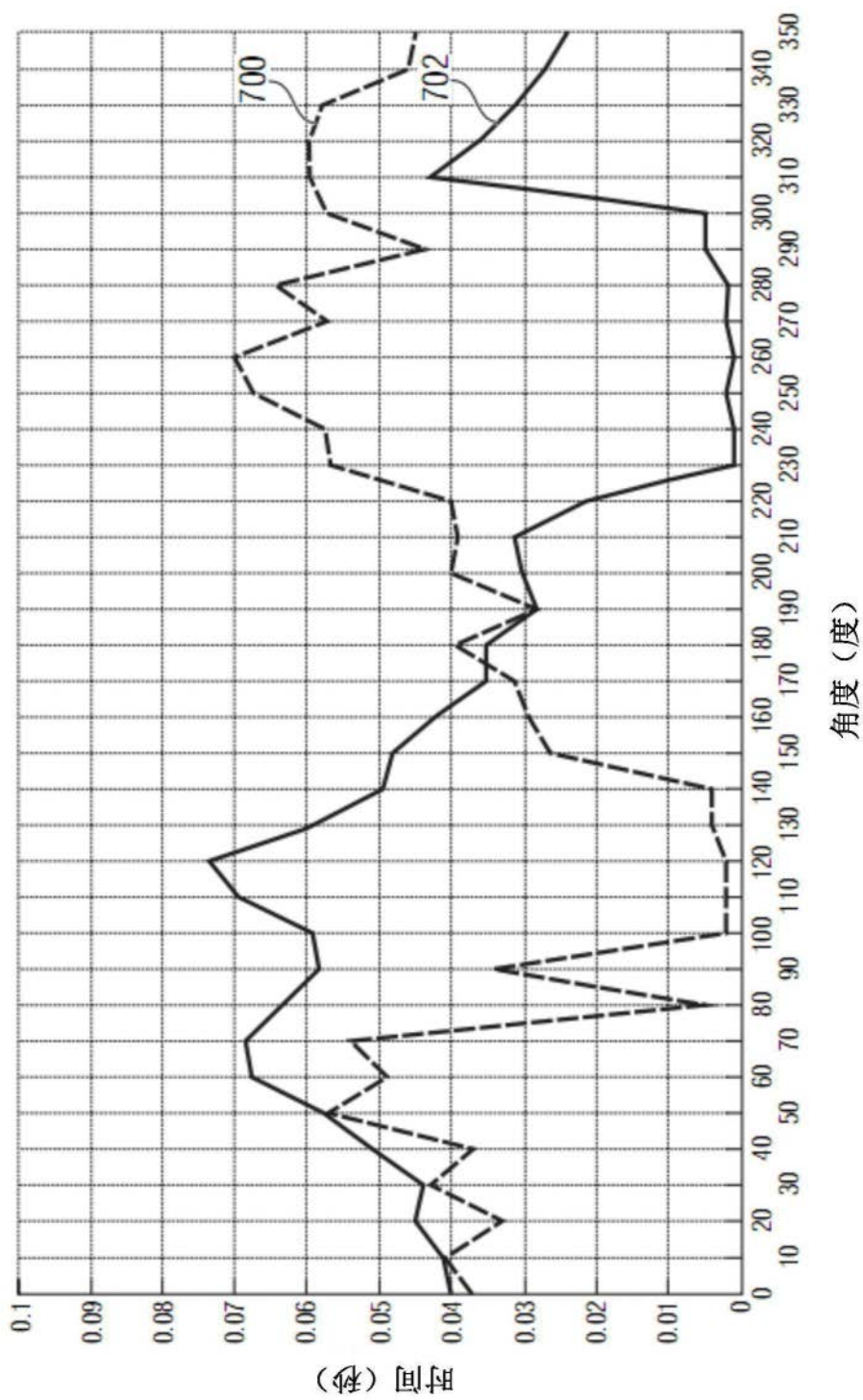


图10

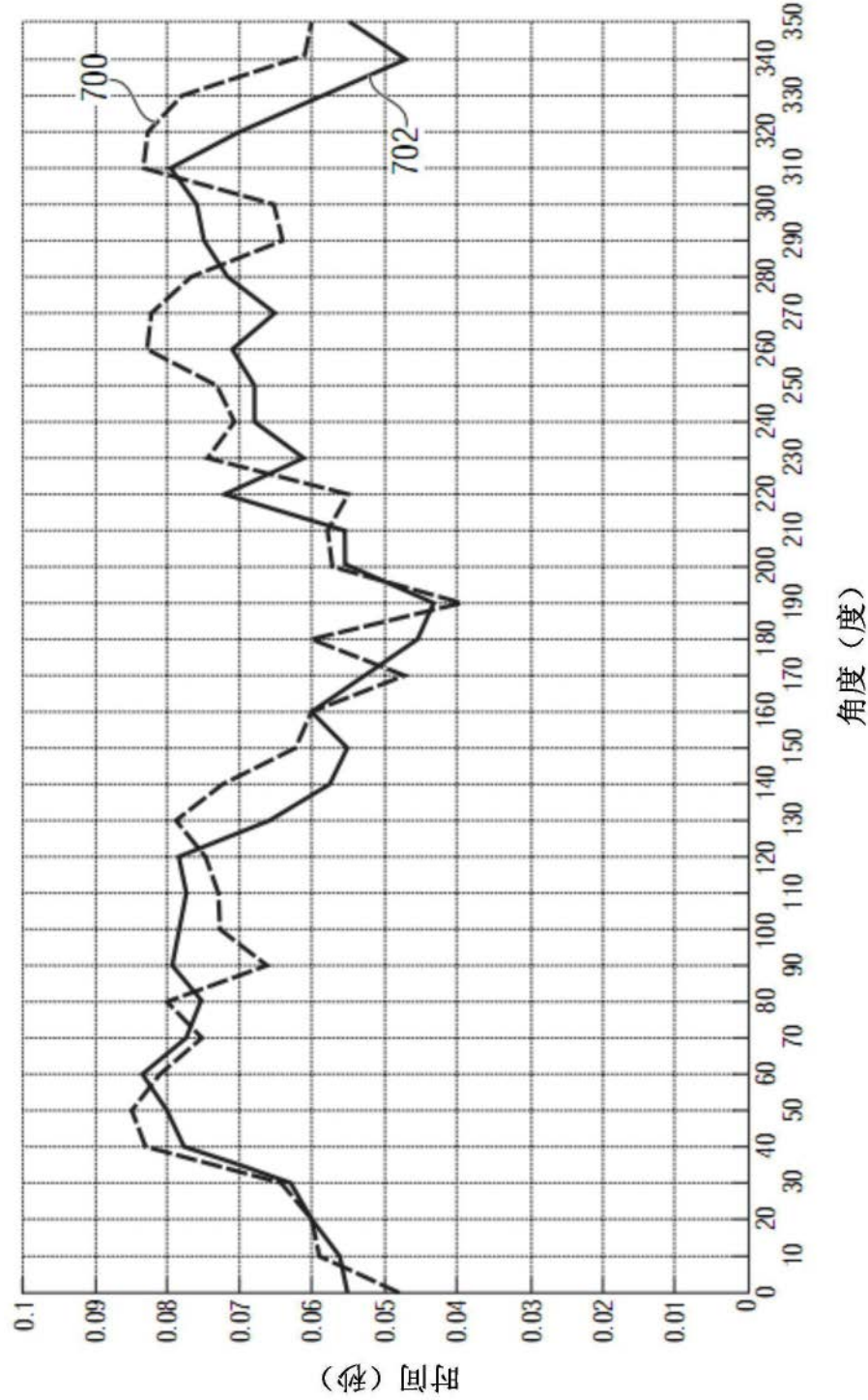


图11

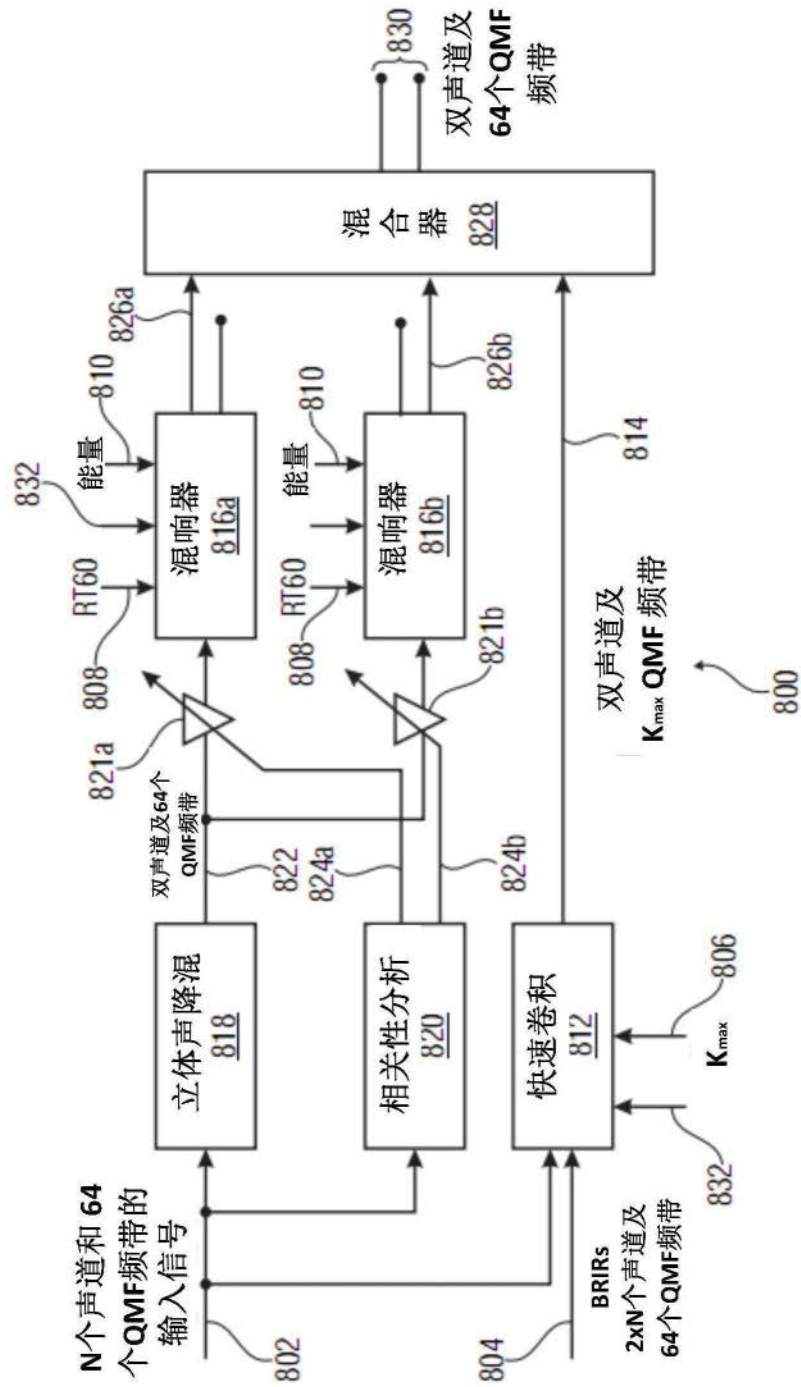


图12

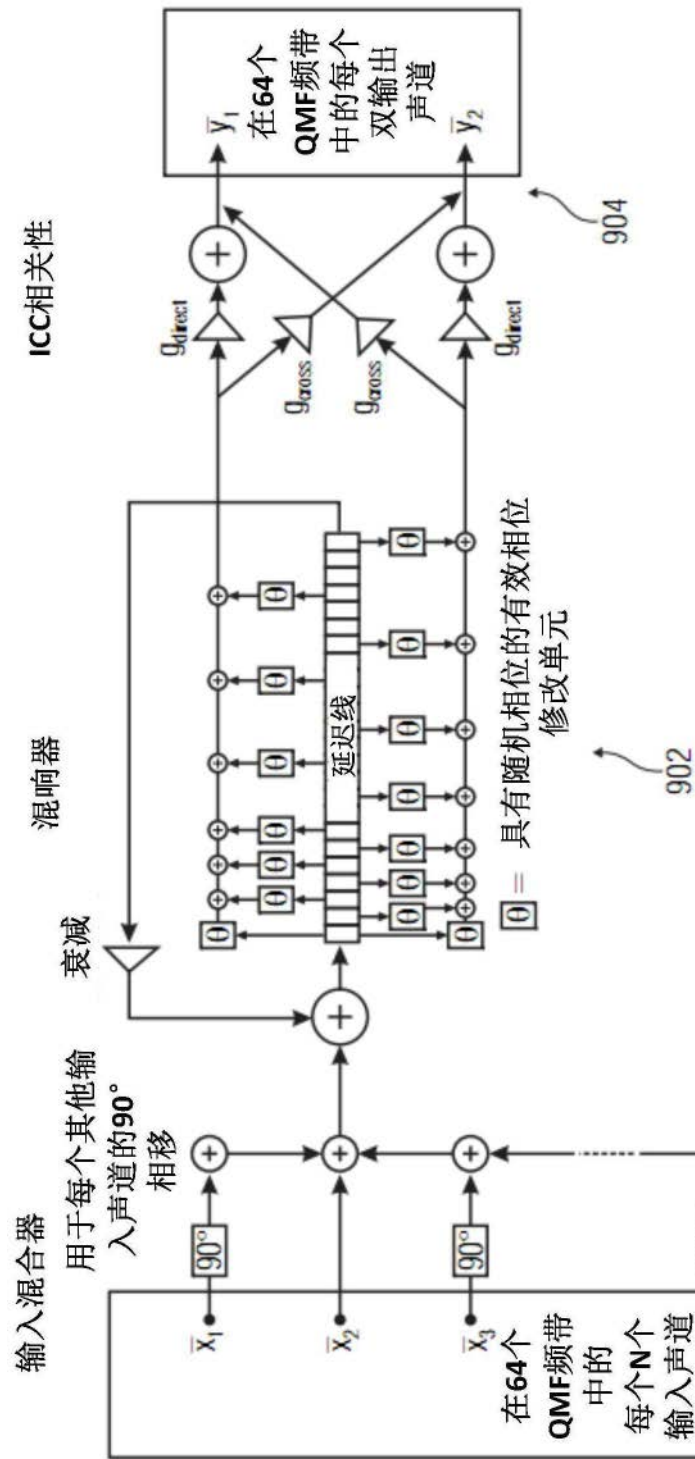


图13