



(12) 发明专利

(10) 授权公告号 CN 108600935 B

(45) 授权公告日 2020.11.03

(21) 申请号 201810503225.7

(22) 申请日 2015.03.19

(65) 同一申请的已公布的文献号
申请公布号 CN 108600935 A

(43) 申请公布日 2018.09.28

(30) 优先权数据
10-2014-0033966 2014.03.24 KR
61/955,243 2014.03.19 US

(62) 分案原申请数据
201580014818.1 2015.03.19

(73) 专利权人 韦勒斯标准与技术协会公司
地址 韩国首尔

(72) 发明人 吴贤午 李泰圭 郭真三 孙周亨

(74) 专利代理机构 北京京原星洲知识产权代理
事务所(普通合伙) 11747

代理人 缙正煜 雷小林

(51) Int.Cl.
H04S 3/00 (2006.01)

(56) 对比文件
CN 103634733 A, 2014.03.12
CN 102915738 A, 2013.02.06
CN 102165798 A, 2011.08.24
CN 101933344 A, 2010.12.29
WO 2014036121 A1, 2014.03.06
CN 1277532 A, 2000.12.20
US 5371799 A, 1994.12.06

审查员 李莎莎

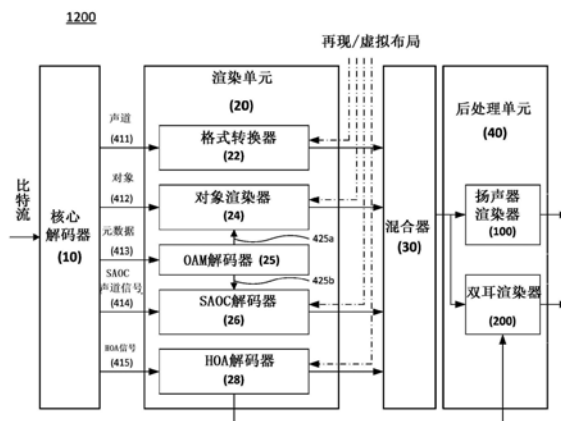
权利要求书2页 说明书25页 附图10页

(54) 发明名称

音频信号处理方法和设备

(57) 摘要

本发明涉及用于处理音频信号的方法和设备。所述方法包括：接收音频信号的比特流；解码包括在所述比特流中的所述音频信号；接收与用于双耳渲染所述音频信号的双耳室脉冲响应BRIR滤波器集合相对应的虚拟布局信息，所述虚拟布局信息包括有关基于所述BRIR滤波器集合确定的目标声道的信息；和基于所接收到的虚拟布局信息，将所解码的音频信号渲染为所述目标声道的信号。



1. 一种用于处理音频信号的方法,所述方法包括:
接收音频信号的比特流;
解码包括在所述比特流中的所述音频信号;
接收与用于双耳渲染所述音频信号的双耳室脉冲响应BRIR滤波器集合相对应的虚拟布局信息,其中所述虚拟布局信息包括有关基于所述BRIR滤波器集合确定的目标声道的信息,以及其中与所述虚拟布局信息相对应的位置集合是与所述BRIR滤波器集合相对应的位置的子集,并且与所述虚拟布局信息相对应的位置集合指示相应目标声道的位置信息;和
基于所接收到的虚拟布局信息,将所解码的音频信号渲染为所述目标声道的信号。
2. 根据权利要求1所述的方法,其中,从执行所述双耳渲染的双耳渲染器接收所述BRIR滤波器集合。
3. 根据权利要求1所述的方法,还包括:
通过混合渲染到相关目标声道的信号的音频信号来生成用于每个目标声道的输出信号。
4. 根据权利要求1所述的方法,还包括:
通过使用与相关目标声道相对应的BRIR滤波器集合的BRIR滤波系数的集合来对用于每个目标声道的所混合的输出信号进行双耳渲染。
5. 根据权利要求4所述的方法,其中,对所混合的输出信号进行所述双耳渲染还包括:
将所述BRIR滤波系数的集合转换成多个子带滤波系数集合;
基于通过至少部分地使用从对应子带滤波系数集合中提取的混响时间信息获得的滤波器阶数信息来截取每个子带滤波系数集合,其中,每个所截取的子带滤波系数集合的长度在频域中可变地确定;和
通过使用与其相对应的所截取的子带滤波系数集合来对所混合的输出信号的每个子带信号进行滤波。
6. 一种用于处理音频信号的设备,包括:
核心解码器,所述核心解码器被配置为接收音频信号的比特流并且解码包括在所述比特流中的所述音频信号;和
渲染器,所述渲染器被配置为将所解码的音频信号渲染为目标声道的一个或多个信号,
其中,所述渲染器进一步被配置为:
接收与用于双耳渲染所述音频信号的双耳室脉冲响应BRIR滤波器集合相对应的虚拟布局信息,其中所述虚拟布局信息包括有关基于所述BRIR滤波器集合确定的目标声道的信息,以及其中与所述虚拟布局信息相对应的位置集合是与所述BRIR滤波器集合相对应的位置的子集,并且与所述虚拟布局信息相对应的位置集合指示相应目标声道的位置信息;和
基于所接收到的虚拟布局信息,将所解码的音频信号渲染为所述目标声道的信号。
7. 根据权利要求6所述的设备,其中,从执行所述双耳渲染的双耳渲染器接收所述BRIR滤波器集合。
8. 根据权利要求6所述的设备,还包括:
混频器,所述混频器被配置为通过混合渲染到相关目标声道的信号的音频信号来生成用于每个目标声道的输出信号。

9. 根据权利要求8所述的设备,还包括:

双耳渲染器,所述双耳渲染器被配置为通过使用与相关目标声道相对应的BRIR滤波器集合的BRIR滤波系数的集合来对用于每个目标声道的所混合的输出信号进行双耳渲染。

10. 根据权利要求9所述的设备,其中,所述双耳渲染器进一步被配置为:

将所述BRIR滤波系数的集合转换成多个子带滤波系数集合;

基于通过至少部分地使用从对应子带滤波系数集合中提取的混响时间信息获得的滤波器阶数信息来截取每个子带滤波系数集合,其中,每个所截取的子带滤波系数集合的长度在频域中可变地确定;和

通过使用与其相对应的所截取的子带滤波系数集合来对所混合的输出信号的每个子带信号进行滤波。

音频信号处理方法和设备

[0001] 本申请是分案申请。本申请的母案申请的申请号为201580014818.1,其国际申请日为2015年3月19日,国际申请号为PCT/KR2015/002669,进入中国国家阶段的日期为2016年9月19日。

技术领域

[0002] 本发明涉及用于处理音频信号的方法和设备,并且更具体地,涉及将对象信号与声道信号合成并且有效地执行合成信号的双耳渲染的用于处理音频信号的方法和设备。

背景技术

[0003] 在现有技术中,3D音频统称为一系列信号处理、传输、编码和再现技术,该一系列信号处理、传输、编码和再现技术用于通过向在环绕音频中提供的水平面(2D)上的声音场景提供与高度方向对应的另一轴,来提供出现在3D空间中的声音。具体地,为了提供3D音频,应当使用比相关技术更多的扬声器,或者以其他方式,尽管使用了比相关技术更少的扬声器,但是需要在不存在扬声器的虚拟位置处产生声音图像的渲染技术。

[0004] 预计3D音频将是与超高清(UHD)TV对应的音频解决方案,并且预计3D音频将应用于各种领域,除了在演进为高品质的信息娱乐空间的车辆中的声音之外,还包括影院音响、个人3DTV、平板装置、智能手机和云游戏。

[0005] 同时,作为提供给3D音频的声源的类型,可以存在基于声道的信号和基于对象的信号。另外,可以存在基于声道的信号和基于对象的信号被混合的声源,并且因此,用户可以具有新型的收听体验。

[0006] 同时,在音频信号处理设备中,在用于处理基于声道的信号的声道渲染器与用于处理基于对象的信号的对象渲染器之间,可能存在性能差异。换言之,可以在基于声道的信号的基础上实现音频信号处理设备的双耳渲染。在这种情况下,当基于声道的信号与基于对象的信号被混合的声音场景被接收作为音频信号处理设备的输入时,可能无法如期望那样通过双耳渲染来再现对应的声音场景。因此,需要解决由于声道渲染器与对象渲染器之间的性能差异而可能发生的各种问题。

发明内容

[0007] 技术问题

[0008] 本发明致力于提供一种用于处理音频信号的方法和设备,该方法和设备可以通过实现与可以由双耳渲染器提供的空间分辨率相对应的对象渲染器和声道渲染器来产生满足双耳渲染器的性能的输出信号。

[0009] 本发明还是致力于实现一种滤波过程,该滤波过程在最小化在双耳渲染中的音质损失的同时,要求具有非常小的计算量的高计算量,以便在以立体声再现多声道或者多对象信号时保持原始信号的沉浸感。

[0010] 本发明还致力于在输入信号中包含失真时通过高质量滤波器来最小化失真传播。

[0011] 本发明还致力于实现具有非常大的长度的有限脉冲响应 (FIR) 滤波器作为具有较小长度的滤波器。

[0012] 本发明还致力于在使用缩小 FIR 的滤波器执行滤波时通过省略的滤波系数来最小化截断部分 (destructured part) 的失真。

[0013] 技术解决方案

[0014] 为了实现这些目标,本发明提供如下用于处理音频信号的方法和设备。

[0015] 本发明的示范性实施例提供了一种用于处理音频信号的方法,包括:接收包括多声道信号的输入音频信号;接收用于对该输入音频信号进行滤波的截取子带滤波系数,该截取子带滤波系数是从用于该输入音频信号的双耳滤波的双耳房间脉冲响应 (BRIR) 滤波系数获得的子带滤波系数中的至少一些,并且基于通过至少部分地使用从对应的子带滤波系数中提取的混响时间信息所获得的滤波器阶数信息来确定截取子带滤波系数的长度;获得指示与输入音频信号的每个声道相对应的 BRIR 滤波系数的矢量信息;以及基于该矢量信息,通过使用与相关声道和子带相对应的截取子带滤波系数来对多声道信号的每个子带信号进行滤波。

[0016] 本发明的另一示范性实施例提供了一种用于处理音频信号以执行对输入音频信号的双耳渲染的设备,包括:生成用于输入音频信号的滤波器的参数化单元;以及双耳渲染单元,该双耳渲染单元接收包括多声道信号的输入音频信号并且通过使用由参数化单元所生成的参数来对输入音频信号进行滤波,其中,双耳渲染单元从参数化单元接收用于对输入音频信号进行滤波的截取子带滤波系数,该截取子带滤波系数是从用于该输入音频信号的双耳滤波的双耳房间脉冲响应 (BRIR) 滤波系数获得的子带滤波系数中的至少一些,并且基于通过至少部分地使用从对应的子带滤波系数中提取的混响时间信息所获得的滤波器阶数信息来确定截取子带滤波系数的长度,获得指示与输入音频信号的每个声道相对应的 BRIR 滤波系数的矢量信息,并且基于该矢量信息,通过使用与相关声道和子带相对应的截取子带滤波系数来对多声道信号的每个子带信号进行滤波。

[0017] 在这种情况下,当在 BRIR 滤波器集合中存在具有与输入音频信号的特定声道的位置信息匹配的位置信息的 BRIR 滤波系数时,矢量信息可以指示相关 BRIR 滤波系数作为与特定声道相对应的 BRIR 滤波系数。

[0018] 此外,当在 BRIR 滤波器集合中不存在具有与输入音频信号的特定声道的位置信息匹配的位置信息的 BRIR 滤波系数时,矢量信息可以指示距特定声道的位置信息的最小几何距离的 BRIR 滤波系数作为与特定声道相对应的 BRIR 滤波系数。

[0019] 在这种情况下,几何距离可以通过汇聚在两个位置之间的高度偏差的绝对值以及在两个位置之间的方位偏差的绝对值所获得的值。

[0020] 至少一个截取子带滤波系数的长度可以与另一子带的截取子带滤波系数的长度不同。

[0021] 本发明的又一示范性实施例提供了一种用于处理音频信号的方法,包括:接收包括声道信号和对象信号中的至少一个的音频信号的比特流;对包括在比特流中的每个音频信号进行解码;接收与用于音频信号的双耳渲染的双耳房间脉冲响应 (BRIR) 滤波器集合相对应的虚拟布局信息,该虚拟布局信息包括有关基于该 BRIR 滤波器集合确定的目标声道的信息;以及基于该接收到的虚拟布局信息,将每个解码的音频信号渲染为目标声道的信号。

[0022] 本发明的又一示例性实施例提供了一种用于处理音频信号的设备,包括:核心解码器,该核心解码器接收包括声道信号和对象信号中的至少一个的音频信号的比特流并且对包括在该比特流中的每个音频信号进行解码;以及渲染器,该渲染器接收与用于音频信号的双耳渲染的双耳房间脉冲响应 (BRIR) 滤波器集合相对应的虚拟布局信息,该虚拟布局信息包括有关基于该BRIR滤波器集合确定的目标声道的信息并且基于该接收到的虚拟布局信息,将每个解码的音频信号渲染为目标声道的信号。

[0023] 在这种情况下,与虚拟布局信息相对应的位置集合可以是与BRIR滤波器集合相对应的位置集合的子集,并且虚拟布局信息的位置集合可以指示相应目标声道的位置信息。

[0024] 可以从执行双耳渲染的双耳渲染器接收BRIR滤波器集合。

[0025] 该设备可以进一步包括混合器,该混合器通过混合针对每个目标声道被渲染为目标声道的信号的每个音频信号来输出用于每个目标信道的输出信号。

[0026] 该设备可以进一步包括:双耳渲染器,该双耳渲染器通过使用与相关目标声道相对应的BRIR滤波器集合的BRIR滤波系数来对用于每个目标声道的混合输出信号进行双耳渲染。

[0027] 在这种情况下,双耳渲染器可以将BRIR滤波系数转换成多个子带滤波系数,基于通过至少部分地使用从对应子带滤波系数中提取的混响时间信息所获得的滤波器阶数信息来截取每个子带滤波系数,其中,至少一个截取子带滤波系数的长度可以与另一子带的截取子带滤波系数的长度不同,并且通过使用与相关声道和子带相对应的截取子带滤波系数来对用于每个目标声道的混合输出信号的每个子带信号进行滤波。

[0028] 有益效果

[0029] 根据本发明的示例性实施例,基于由双耳渲染器处理的数据集来执行声道和对象渲染以实现有效的双耳渲染。

[0030] 另外,当使用具有比声道更多的数据集的双耳渲染器时,可以实现提供更加改善的音质的对象渲染。

[0031] 另外,根据本发明的示例性实施例,当执行对多声道或者多对象信号的双耳渲染时,可以显著减少计算量,同时最小化音质损失。

[0032] 另外,能够对多声道或者多对象音频信号实现具有高音质的双耳渲染,而在现有技术的低功率装置中已经不可能进行这种实时处理。

[0033] 本发明提供了一种以小的计算量来有效地执行对包括音频信号的各种类型的多媒体信号进行滤波的方法。

附图说明

[0034] 图1是图示了根据本发明的示例性实施例的包括音频编码器和音频解码器的整体音频信号处理系统的配置图。

[0035] 图2是图示了根据多声道音频系统的示例性实施例的多声道扬声器的配置的配置图。

[0036] 图3是示意性地图示了在收听空间中构成3D声音场景的各个声音对象的位置的图。

[0037] 图4是图示了根据本发明的示例性实施例的音频信号解码器的框图。

- [0038] 图5是图示了根据本发明的另一示例性实施例的音频解码器的框图。
- [0039] 图6是图示了对例外对象执行渲染的本发明的示例性实施例的框图。
- [0040] 图7是图示了根据本发明的示例性实施例的双耳渲染器的各个组件的框图。
- [0041] 图8是图示了根据本发明的示例性实施例的用于双耳渲染的滤波器生成方法的图。
- [0042] 图9是具体图示了根据本发明的示例性实施例的QTDL处理的图。
- [0043] 图10是图示了本发明的BRIR参数化单元的相应组件的框图。
- [0044] 图11是图示了本发明的VOFF参数化单元的相应组件的框图。
- [0045] 图12是图示了本发明的VOFF参数生成单元的详细配置的框图。
- [0046] 图13是图示了本发明的QTDL参数化单元的相应组件的框图。
- [0047] 图14是图示了用于生成用于逐帧式快速卷积的FFT滤波系数的方法的示例性实施例的图。

具体实施方式

[0048] 考虑到本发明中的功能,在本说明书中使用的术语尽量采用目前广泛使用的通用术语,但是,可以根据本领域的技术人员的意图、习惯、或者新技术的出现来改变这些术语。此外,在特定情况下,可以使用申请人任意选择的术语,并且在这种情况下,在本发明的对应描述部分中,将公开这些术语的含义。此外,我们旨在发现应该不仅基于术语的名称,还应该基于贯穿本本说明书的术语的实质意义和内容来分析在本说明书中使用的术语。

[0049] 图1是图示了根据本发明的示例性实施例的包括音频编码器和音频解码器的整体音频信号处理系统的配置图。

[0050] 根据图1,音频编码器1100对输入声音场景进行编码以生成比特流。音频解码器1200可以接收所生成的比特流,并且通过使用根据本发明的示例性实施例的用于处理音频信号的方法解码和渲染对应的比特流来生成输出声音场景。在本说明书中,音频信号处理设备可以将音频解码器1200指示为狭义的,但是本发明不限于此,并且音频信号处理设备可以指示包括在音频解码器1200的具体组件或者包括音频编码器1100和音频解码器1200的整体音频信号处理系统。

[0051] 图2是图示了根据多声道音频系统的示例性实施例的多声道扬声器的配置的配置图。

[0052] 在该多声道音频系统中,可以使用多个扬声器声道以改善存在感,并且具体地,可以在宽度、深度和高度方向上设置多个扬声器以在3D空间中提供存在感。在图2中,作为示例性实施例,图示了22.2-声道扬声器配置,但是本发明不限于声道的具体数目或者扬声器的具体配置。参照图2,可以由具有顶层、中间层和底层的三个层来构成22.2-声道扬声器集合。当TV屏幕的位置为正面时,在顶层上,在正面设置三个扬声器,在中间位置设置三个扬声器,并且在环绕位置设置三个扬声器,由此可以设置总共9个扬声器。此外,在中间层上,在正面设置五个扬声器,在中间位置设置两个扬声器,并且在环绕位置设置三个扬声器,由此可以设置总共10个扬声器。同时,在底层上,在正面设置三个扬声器,并且可以提供两个LFE声道扬声器。

[0053] 如上所述,需要大的计算量来传输和再现具有最多10个声道的多声道信号。此外,

当考虑到通信环境时,可能需要用于对应信号的高压缩率。此外,在普通家庭中,具有诸如22.2声道的多声道扬声器系统的用户极少,并且存在很多设置了具有2-声道或者5.1-声道设置的系统的情况。因此,当共同传输给所有用户的信号是对多声道中的每一个进行编码的信号时,需要再次将相关多声道信号转换成对应于2-声道或者5.1-声道的多声道信号的过程。因此,可能会造成低通信效率,并且由于需要存储22.2-声道脉冲编码调制(PCM)信号,所以甚至可能在存储器管理中发生低效率的问题。

[0054] 图3是示意性地图示了在收听空间中构成3D声音场景的相应声音对象的位置的示意图。

[0055] 如在图3中所图示的,在收听方52收听3D音频的收听空间50中,可以使构成3D声音场景的各个声音对象51以点声源的形式分布在不同的位置。此外,除了点声源之外,声音场景还可以包括平面波声源或者环境声源。如上所述,需要一种有效的渲染方法来明确地向收听方52提供在3D空间中不同地分布的对象和声源。

[0056] 图4是图示了根据本发明的另一示例性实施例的音频解码器的框图。本发明的音频解码器1200包括核心解码器10、渲染单元20、混合器30和后处理单元40。

[0057] 首先,核心解码器10对接收到的比特流进行解码,并且将该解码的比特流传递至渲染单元20。在这种情况下,从核心解码器10输出并且被传递至渲染单元的信号可以包括扩音器声道信号411、对象信号412、SAOC声道信号414、HOA信号415和对象元数据比特流413。用于在编码器中进行编码的核心编解码器可以用于核心解码器10,并且例如,可以使用MP3、AAC、AC3或者基于联合语音和音频编码(USAC)的编解码器。

[0058] 同时,接收到的比特流可以进一步包括可以标识由核心解码器10解码的信号是声道信号、对象信号还是HOA信号的标识符。此外,当解码的信号是声道信号411时,在比特流中可以进一步包括可以标识每个信号对应于多声道中的哪个声道(例如,对应于左边扬声器、对应于后方右上扬声器等)的标识符。当解码的信号是对象信号412时,可以另外获得指示将对应的信号在再现空间中哪个位置处被再现的信息,如同通过解码对象元数据比特流413所获得的对象元数据信息425a和425b。

[0059] 根据本发明的示例性实施例,音频解码器执行灵活渲染以改善输出音频信号的质量。该灵活渲染可以指基于实际再现环境的扩音器配置(再现布局)或者双耳房间脉冲响应(BRIR)滤波器集合的虚拟扬声器配置(虚拟布局)来转换解码的音频信号的格式的过程。通常,在设置在实际起居室环境中的扬声器中,方位角和距离二者与标准建议的不同。因为距扬声器的收听方的高度、方向、距离等不同于根据标准建议的扬声器配置,所以当在扬声器的改变位置处再现原始信号时,可能难以提供理想的3D声音场景。为了即使在不同扬声器配置中也有效地提供内容制作者预期的声音场景,需要灵活渲染,该灵活渲染通过转换音频信号来根据在扬声器当中的位置差异校正该改变。

[0060] 因此,渲染单元20通过使用再现布局信息或者虚拟布局信息来将由核心解码器10解码的信号渲染为目标输出信号。该再现布局信息可以指示目标声道的配置并且可以表示为再现环境的扩音器布局信息。此外,可以基于在双耳渲染器200中使用的双耳房间脉冲响应(BRIR)滤波器集合来获得虚拟布局信息,并且可以通过与BRIR滤波器集合相对应的位置集合的子集来构成与虚拟布局相对应的位置集合。在这种情况下,虚拟布局的位置集合指示各个目标声道的位置信息。渲染单元20可以包括格式转换器22、对象渲染器24、OAM解码

器25、SAOC解码器26和HOA解码器28。渲染单元20根据解码的信号的类型,通过使用上述配置中的至少一个来执行渲染。

[0061] 格式转换器22还可以被称为声道渲染器,并且将传输的声道信号411转换成输出扬声器声道信号。即,格式转换器22执行在传输的声道配置与要再现的扬声器声道配置之间的转换。当输出扬声器声道的数目(例如,5.1声道)小于传输的声道的数目(例如,22.2声道),或者传输的声道配置和要再现的声道配置彼此不同时,格式转换器22执行声道信号411的向下混合或者转换。根据本发明的示例性实施例,音频解码器可以通过使用在输入声道信号与输出扬声器声道信号之间的组合来生成最优向下混合矩阵,并且通过使用该矩阵来执行行下混合。此外,预渲染的对象信号可以被包括在由格式转换器22处理的声道信号411中。根据示例性实施例,在对音频信号进行解码之前,可以将至少一个对象信号预渲染和混合为声道信号。通过格式转换器22,可以将混合的对象信号与声道信号一起转换成输出扬声器声道信号。

[0062] 对象渲染器24和SAOC解码器26对基于对象的音频信号执行渲染。基于对象的音频信号可以包括离散对象波形和参数对象波形。在离散对象波形的情况下,按照单声道波形向编码器提供各个对象信号,并且编码器通过使用单通道元素(SCE)来传输各个对象信号。在参数对象波形的情况下,多个对象信号被向下混合为至少一个声道信号,并且相应对象的特征和特点之间的关系被表示为空间音频对象编码(SAOC)参数。利用该核心编解码器来对对象信号进行向下混合和编码,并且在这种情况下,所生成的参数信息被一起传输至解码器。

[0063] 同时,当单独的对象波形或者参数对象波形被传输至音频解码器时,可以一起传输与之相对应的压缩对象元数据。对象元数据通过以时间和空间为单位量化对象属性来指定每个对象在3D空间中的位置和增益值。渲染单元20的OAM解码器25接收压缩对象元数据比特流413,并且对接收到的压缩对象元数据比特流413进行解码,并且将解码的对象元数据比特流413传递至对象渲染器24和/或SAOC解码器26。

[0064] 对象渲染器24通过使用对象元数据信息425a来根据给定的再现格式对每个对象信号412进行渲染。在这种情况下,可以基于对象元数据信息425a来将每个对象信号412渲染为特定输出声道。SAOC解码器26从SAOC声道信号414和参数信息来恢复对象/声道信号。此外,SAOC解码器26可以基于再现布局信息和对象元数据信息425b生成输出音频信号。即,SAOC解码器26通过使用SAOC声道信号414来生成解码的对象信号,并且执行将解码的对象信号映射成目标输出信号的渲染。如上所述,对象渲染器24和SAOC解码器26可以将对象信号渲染为声道信号。

[0065] HOA解码器28接收高阶立体混响(HOA)信号415和HOA附加信息,并且对该HOA信号和HOA附加信息进行解码。HOA解码器28通过独立等式来对声道信号或者对象信号建模以生成声音场景。当在所生成的声音场景中选择扬声器的空间位置时,可以将声道信号或者对象信号渲染为扬声器声道信号。

[0066] 同时,虽然在图4中未图示,但是当音频信号被传递至渲染单元20的各个组件时,动态范围控制(DRC)可以作为预处理程序被执行。DRC将再现的音频信号的范围限制为预定水平,并且将小于预定阈值的声音调大,而将大于预定阈值的声音调小。

[0067] 将由渲染单元20处理的基于声道的音频信号和基于对象的音频信号传递至混合

器30。混合器30混合由渲染单元20的各个子单元渲染的部分信号以生成混合器输出信号。当部分信号与在再现/虚拟布局上的相同的位置匹配时,该部分信号彼此相加,并且当该部分信号与不相同的位置匹配时,该部分信号被混合以输出分别对应于独立位置的信号。混合器30可以确定在彼此相加的部分信号中是否发生频偏干扰,并且进一步执行用于防止该频偏干扰的附加过程。此外,混合器30调整基于声道的波形和渲染的对象波形的延迟,并且以样本为单位汇聚所调整的波形。由混合器30汇聚的音频信号被传递至后处理单元40。

[0068] 后处理单元40包括扬声器渲染器100和双耳渲染器200。扬声器渲染器100执行用于输出从混合器30传递的多声道和/或多对象音频信号的后处理。后处理可以包括动态范围控制(DRC)、响度标准化(LN)和峰值限制器(PL)。将扬声器渲染器100的输出信号传递至多声道音频系统的扩音器以便输出。

[0069] 双耳渲染器200生成多声道和/或多对象音频信号的双耳向下混合信号。双耳向下混合信号是允许用位于3D中的虚拟声源来表示每个输入声道/对象信号的2-声道音频信号。双耳渲染器200可以接收供应到扬声器渲染器100的音频信号作为输入信号。双耳渲染器可以基于双耳房间脉冲响应(BRIR)来执行并且在时间域或者QMF域上执行。根据示例性实施例,作为双耳渲染的后处理程序,可以附加地执行动态范围控制(DRC)、响度规范化(LN)和峰值限制器(PL)。可以将双耳渲染器200的输出信号传递和输出到诸如头戴耳机、耳机等的2-声道音频输出装置。

[0070] <用于灵活渲染的渲染配置单元>

[0071] 图5是图示了根据本发明的另一示例性实施例的音频解码器的框图。在图5的示例性实施例中,相同的附图标记表示与图4的示例性实施例相同的元件,并且将省略重复的描述。

[0072] 参照图5,音频解码器1200-A可以进一步包括控制解码的音频信号的渲染的渲染配置单元21。渲染配置单元21接收再现布局信息401和/或BRIR滤波器集合信息402,并且通过使用该接收到的再现布局信息401和/或BRIR滤波器集合信息402来生成用于渲染音频信号的目标格式信息421。根据示例性实施例,渲染配置单元21可以获得实际再现环境的扩音器配置作为再现布局信息401,并且基于此来生成目标格式信息421。在这种情况下,目标格式信息421可以表示实际再现环境的扩音器的位置(声道)或其子集、或者基于其组合的超集。

[0073] 渲染配置单元21可以从双耳渲染器200获得BRIR滤波器集合信息402,并且通过使用所获得的BRIR滤波器集合信息402来生成目标格式信息421。在这种情况下,目标格式信息421可以表示双耳渲染器200的BRIR滤波器集合支持的(即,可双耳渲染的)目标位置(声道)或其子集或者基于其组合的超集。根据本发明的示例性实施例,BRIR滤波器集合信息402可以包括不同于指示物理扩音器的配置的再现布局信息401的目标位置或者包括更多目标位置。因此,当基于再现布局信息401渲染的音频信号被输入到双耳渲染器200中时,在渲染的音频信号的目标位置与双耳渲染器200所支持的目标位置之间差异可能发生。替代地,由核心解码器10解码的信号的目标位置可以通过BRIR滤波器集合信息402来提供,而不能由再现布局信息401来提供。

[0074] 因此,当最终输出音频信号是双耳信号时,本发明的渲染配置单元21可以通过使用从双耳渲染器200获得的BRIR滤波器集合信息402来生成目标格式信息421。渲染单元20

基于再现布局信息401和双耳渲染,通过使用所生成的目标格式信息421来执行对音频信号的渲染,以最小化可能由于2-步渲染处理而导致的音质劣化现象。

[0075] 同时,渲染配置单元21可以进一步获得有关最终输出音频信号的类型的信息。当最终输出音频信号是扩音器信号时,渲染配置单元21可以基于再现布局信息401来生成目标格式信息421,并且将所生成的目标格式信息421传递至渲染单元20。此外,当最终输出音频信号是双耳信号时,渲染配置单元21可以基于BRIR滤波器集合信息402来生成目标格式信息421,并且将所生成的目标格式信息421传递至渲染单元20。根据本发明的另一示例性实施例,渲染配置单元21可以进一步获得指示由用户使用的音频系统或者用户的选择的控制信息403,并且通过同时使用对应的控制信息403来生成目标格式信息421。

[0076] 将所生成的目标格式信息421传递至渲染单元20。渲染单元20的各个子单元可以通过使用从渲染配置单元21传递的目标格式信息421来执行灵活渲染。即,格式转换器22基于目标格式信息421来将解码的声道信号411转换为目标声道的输出信号。类似地,对象渲染器24和SAOC解码器26分别通过使用目标格式信息421和目标元数据425来将对象信号412和SAOC声道信号414转换成目标声道的输出信号。在这种情况下,可以基于目标格式信息421来更新用于渲染对象信号421的混合矩阵,并且对象信号24可以通过使用更新的混合矩阵来将对象信号412渲染为输出声道信号。如上所述,可以通过将音频信号映射成目标格式上的至少一个目标位置(即,目标声道)的转换过程来执行渲染。

[0077] 同时,甚至可以将目标格式信息421传递至混合器30并且可以将其用于混合由渲染单元20的各个子单元所渲染的部分信号的过程。当该部分信号与目标格式上的相同位置匹配时,该部分信号彼此相加,并且当该部分信号与不相同的位置匹配时,该部分信号被混合为分别对应于独立位置的输出信号。

[0078] 根据本发明的示例性实施例,可以根据各种方法来设置目标格式。首先,渲染配置单元21可以设置具有比所获得的再现布局信息401或者BRIR滤波器集合信息402更高的空间分辨率的目标格式。即,渲染配置单元21获得第一目标位置集合,该第一目标位置集合是由再现布局信息401或者BRIR滤波器集合信息402指示的原始目标位置的集合,并且组合一个或者多个原始目标位置以生成额外的目标位置。在这种情况下,额外的目标位置可以包括通过在多个原始目标位置中的内插所生成的位置、通过外推生成的位置等。通过所生成的额外的目标位置的集合,可以配置第二目标位置集合。渲染配置单元21可以生成包括第一目标位置集合和第二目标位置集合的目标格式,并且将对应的目标格式信息4210传递至渲染单元20。

[0079] 渲染单元20可以通过使用包括额外的目标位置的高分辨率目标格式信息421来对音频信号进行渲染。当通过使用高分辨率目标格式信息421执行渲染时,渲染过程的分辨率被改善,并且因此,计算变得容易并且改善了音质。渲染单元20可以通过对音频信号进行渲染来获得映射到目标格式信息421的每个目标位置的输出信号。当获得映射到第二目标位置集合的附加目标位置的输出信号时,渲染单元20可以执行将相应输出信号重新渲染为用于第一目标位置集合的原始目标位置的向下混合过程。在这种情况下,可以通过基于向量的幅度平移(VBAP)或者幅度平移来实现向下混合过程。

[0080] 作为用于设置目标格式的另一方法,渲染配置单元21可以设置具有比所获得的BRIR滤波器集合信息402更低的空间分辨率的目标格式。即,渲染配置单元21可以通过M个

原始目标位置的子集或者其组合来获得 $N(N < M)$ 个缩小(abbreviated)的目标位置并且生成由该缩小的目标位置构成的目标格式。渲染配置单元21可以向渲染单元20传递对应的低分辨率目标格式信息421,并且渲染单元20可以通过使用该低分辨率目标格式信息421来执行对音频信号的渲染。当通过使用低分辨率目标格式信息421执行渲染时,可以减少渲染单元20的计算量以及随后的双耳渲染器200的计算量。

[0081] 作为用于设置目标格式的又一方法,渲染配置单元21可以为渲染单元20的每个子单元设置不同的目标格式。例如,提供给格式转换器20的目标格式和提供给对象渲染器24的目标格式可以彼此不同。当根据每个子单元提供不同的目标格式时,针对每个子单元,可以控制计算量或者可以改善音质。

[0082] 渲染配置单元21可以不同地设置提供给渲染单元20的目标格式和提供给混合器30的目标格式。例如,提供给渲染单元20的目标格式可以具有比提供给混合器30的目标格式更高的空间分辨率。因此,混合器30可以被实现为伴随向下混合具有高分辨率的输入信号的过程。

[0083] 同时,渲染配置单元21可以基于用户的选择和所使用的装置的环境或者设置,来设置目标格式。渲染配置单元21可以通过控制信息403来接收信息。在这种情况下,控制信息403基于可以由装置提供的计算量性能和电能以及用户的选择中的至少一个而变化。

[0084] 在图4和图5的示例性实施例中,图示了渲染单元20根据渲染目标信号通过不同的子单元执行渲染,但是可以通过集成有所有或者一些子单元的渲染器来实现渲染单元20。例如,可以通过一个集成渲染器来实现格式转换器22和对象渲染器24。

[0085] 根据本发明的示例性实施例,如图5中所示,可以将对象渲染器24的输出信号中的至少一些输入到格式转换器22。输入到格式转换器22中的对象渲染器24的输出信号可以用于解决在空间中的不匹配的信息,该不匹配可能由于在对对象信号的灵活渲染和对声道信号的灵活渲染的性能差异而在信号之间发生。例如,当对象信号412和声道信号411被同时接收作为输入,并且期望提供两个信号被混合的形式的声音场景时,用于各个信号的渲染过程彼此不同,并且因此,由于在空间中的不匹配而导致容易发生失真。因此,根据本发明的示例性实施例,当对象信号412和声道信号411被同时接收作为输入时,对象渲染器24可以基于目标格式信息421,在不独立执行灵活渲染的情况下,向格式转换器22传递输出信号。在这种情况下,该传递至格式转换器22的对象渲染器24的输出信号可以是与输入声道信号411的声道格式相对应的信号。此外,格式转换器22可以将对象渲染器24的输出声道混合至声道信号411,并且基于目标格式信息421对混合的信号执行灵活渲染。

[0086] 同时,在位于可用扬声器区域外的例外对象的情况下,难以仅通过现有技术中的扬声器来再现内容制作者期望的声音。因此,当存在例外对象时,对象渲染器24可以生成与该例外对象的位置相对应的虚拟扬声器,并且通过使用实际扩音器信息和虚拟扬声器信息二者来执行渲染。

[0087] 图6是图示了对例外对象进行渲染的本发明的示例性实施例的框图。在图6中,由附图标记401至609标示的实线点表示目标格式所支持的各个目标位置,并且目标位置环绕的区域形成可以被渲染的输出声道空间。此外,由附图标记611至613标示的虚线点表示目标格式不支持的虚拟位置,并且可以表示由对象渲染器24生成的虚拟扬声器的位置。同时,由S1 701至S1 704标示的星形点表示需要在特定对象S沿着路径700移动时在特定时间渲

染的空间再现位置。可以基于对象元数据信息425来获得对象的空间再现位置。

[0088] 在图6的示例性实施例中,可以基于对应的对象的再现位置是否与目标格式的目标位置匹配来渲染对象信号。当对象的再现位置与特定目标位置604匹配时,如S2 702,将对应的对象信号转换成与目标位置604相对应的目标声道的输出信号。即,可以通过与目标声道的1:1映射来渲染该对象信号。然而,当对象的再现位置位于输出声道空间中,但不直接与目标位置匹配时,如S1 701,可以使对应的对象信号分布至与再现位置相邻的多个目标位置的输出信号。例如,可以将S1701的对象信号渲染为相邻目标位置601、602和603的输出信号。当对象信号被映射到两个或者三个目标位置时,可以通过诸如基于矢量的幅度平移(VBAP)等方法将对应的对象信号渲染为每个目标声道的输出信号。因此,可以通过与多个目标声道的1:N映射来渲染对象信号。

[0089] 同时,当对象的再现位置没有位于由目标格式配置的输出声道空间中时,如S3 703和S4 704,可以通过独立过程来渲染对应的对象。根据示例性实施例,对象渲染器24可以将对应的对象投射到按照目标格式配置的输出声道空间上,并且执行从投射的位置到相邻目标位置的渲染。在这种情况下,针对从投射的位置到目标位置的渲染,可以使用S1 701或者S2 702的渲染方法。即,将S3 703和S4 704分别投射到在输出声道空间中的P3和P4,并且可以将投射的P3和P4的信号渲染为相邻目标位置604、605和607的输出信号。

[0090] 根据另一示例性实施例,当对象的再现位置没有位于按照目标格式配置的输出声道空间中时,对象渲染器24可以通过使用虚拟扬声器的位置和目标位置来渲染对应的对象。首先,对象渲染器24将对应的对象信号渲染为包括至少一个虚拟扬声器信号的输出信号。例如,当对象的再现位置与虚拟扬声器的位置直接匹配时,如S4 704,将对应的对象信号渲染为虚拟扬声器611的输出信号。然而,当不存在与对象的再现位置匹配的虚拟扬声器时,如S3 703,可以将对应的对象信号渲染为相邻虚拟扬声器611以及目标声道605和607的输出信号。接下来,对象渲染器24将所渲染的虚拟扬声器信号重新渲染为目标声道的输出信号。即,可以将S3 703或者S4 704的对象信号被渲染成的虚拟扬声器611的信号向下混合为相邻目标声道(例如,605、607)的输出信号。

[0091] 同时,如在图6中所示,目标格式可以包括通过组合原始目标位置而生成的额外的目标位置621、622、623和624。如上描述地生成和使用额外的目标位置以提高渲染的分辨率。

[0092] <双耳渲染器的细节>

[0093] 图7是图示了根据本发明的示例性实施例的双耳渲染器的每个组件的框图。如在图2中所图示的,根据本发明的示例性实施例的双耳渲染器200可以包括BRIR参数化单元300、快速卷积单元230、后期混响生成单元240、QTDL处理单元250以及混合器&组合器260。

[0094] 双耳渲染器200通过执行对各种类型的输入信号的双耳渲染来生成3D音频耳机信号(即,3D音频2-声道信号)。在这种情况下,输入信号可以是包括声道信号(即,扩音器声道信号)、对象信号、和HOA系数信号中的至少一个的音频信号。根据本发明的另一示例性实施例,当双耳渲染器200包括特定解码器时,输入信号可以是前面提到的音频信号的编码比特流。双耳渲染将解码的输入信号转换成双耳向下混合信号,以使得能够在通过耳机收听对应的双耳向下混合信号时体验环绕声。

[0095] 根据本发明的示例性实施例的双耳渲染器200可以通过使用双耳房间脉冲响应

(BRIR) 滤波器来执行双耳渲染。当使用BRIR的双耳渲染被一般化时,双耳渲染是用于获取用于具有M个声道的多声道输入信号的O输出信号的M-至-O处理。在这种过程期间,双耳滤波可以被视为使用与每个输入声道和每个输出声道对应的滤波系数的滤波。在图3中,原始滤波器集合H指从每个声道信号的扬声器位置到左右耳的位置的传递函数。在一般的收听室中测量的传递函数,即,在传递函数之中的混响空间,被称为双耳房间脉冲响应(BRIR)。相反,为了不受再现空间的影响在消声室中测量的传递函数被称为头部相关脉冲响应(HRIR),并且其传递函数被称为头部相关传递函数(HRTF)。因此,与HRTF不同,BRIR包含再现空闲信息以及方向信息。根据示例性实施例,可以通过使用HRTF和人工混响器来替代BRIR。在本说明书中,对使用BRIR的双耳渲染进行了描述,但是本发明不限于此,并且本发明甚至可以通过类似或者对应的方法,适用于使用包括HRIR和HRIF的各种类型的FIR滤波器的双耳渲染。此外,本发明可以适用于对输入信号的各种形式的滤波以及对音频信号的各种形式的双耳渲染。同时,如上所述,BRIR可以具有96K个样本的长度,并且由于通过使用不同的M*O个滤波器来执行多声道双耳渲染,所以需要具有高计算复杂度的处理过程。

[0096] 在本发明中,从狭义上讲,用于处理音频信号的设备可以指示在图7中图示的双耳渲染器200或者双耳渲染单元220。然而,在本发明中,从广义上讲,用于处理音频信号的设备可以指示包括双耳渲染器的图4或者图5的音频信号解码器。此外,在下文中,在本说明书中,将主要对多声道输入信号的示例性实施例进行描述,但是除非另有描述,否则声道、多声道和多声道输入信号可以用作分别包括对象、多对象和多对象输入信号的概念。此外,多声道输入信号还可以用作包括HOA解码和渲染的信号的概念。

[0097] 根据本发明的示例性实施例,双耳渲染器200可以对在QMF域中执行对输入信号的双耳渲染。即,双耳渲染器200可以接收QMF域的多声道(N个声道)的信号,并且通过使用QMF域的BRIR子带滤波器来执行对该多声道的信号的双耳渲染。当通过OMF分析滤波器组的第i个声道的第k个子带信号用 $x_{k,i}(l)$ 表示并且在子带域中的时间索引由l表示时,可以通过下面给出的等式来表示在QMF域中的双耳渲染。

[0098] [等式1]

$$[0099] \quad y_k^m(l) = \sum_i x_{k,i}(l) * b_{k,i}^m(l)$$

[0100] 此处,m是L(左)或者R(右),并且 $b_{k,i}^m(l)$ 是通过将时间域BRIR滤波器转换成OMF域的子带滤波器来获得的。

[0101] 即,可以通过将QMF域的声道信号或者对象信号划分成多个子带信号并且利用与之对应的BRIR子带滤波器对各个子带信号进行卷积的方法来执行双耳渲染,并且此后,对利用BRIR子带滤波器卷积的各个子带信号进行加总。

[0102] BRIR参数化单元300转换并编辑用于在QMF域中的双耳渲染的BRIR滤波系数,并且生成各种参数。首先,BRIR参数化单元300接收用于多声道或者多对象的时间域BRIR滤波系数,并且将接收到的时间域BRIR滤波系数转换成QMF域BRIR滤波系数。在这种情况下,QMF域BRIR滤波系数分别包括与多个频带相对应的多个子带滤波系数。在本发明中,子带滤波器滤波系数指示QMF-转换的子带域的每个BRIR滤波系数。在本说明书中,可以将子带滤波器滤波系数指定为BRIR子带滤波系数。BRIR参数化单元300可以编辑QMF域的多个BRIR子带滤波系数中的每一个,并且将所编辑的子带滤波系数传递至快速卷积单元230等。根据本发明的示例

性实施例,可以包括BRIR参数化单元300,作为双耳渲染器220的组件,或者以其他方式作为独立设备被提供。根据示例性实施例,包括除了BRIR参数化单元300的快速卷积单元230、后期混响生成单元240、QTDL处理单元250以及混合器&组合器260的组件可以归类为双耳渲染单元220。

[0103] 根据示例性实施例,BRIR参数化单元300可以接收与虚拟再现空间的至少一个位置相对应的BRIR滤波系数作为输入。虚拟再现空间的每个位置可以与多声道系统的每个扬声器位置相对应。根据示例性实施例,由BRIR参数化单元300接收的BRIR滤波系数中的每一个可以与双耳渲染器200的输入信号中的每个声道或者每个对象直接匹配。相反,根据本发明的另一示例性实施例,接收到的BRIR滤波系数中的每一个可以具有独立于双耳渲染器200的输入信号的配置。即,由BRIR参数化单元300接收的BRIR滤波系数中的至少一部分可以与双耳渲染器200的输入信号不直接匹配,并且接收到的BRIR滤波系数的数目可以小于或者大于输入信号的声道和/或对象的总数。

[0104] BRIR参数化单元300还可以接收控制参数信息,并且基于接收到的控制参数信息来生成用于双耳渲染的参数。如在下面描述的示例性实施例中所描述的,控制参数信息可以包括复杂度-质量控制信息等,并且可以用作用于BRIR参数化单元300的各种参数化过程的阈值。BRIR参数化单元300基于输入值来生成双耳渲染参数,并且将所生成的双耳渲染参数传递至双耳渲染单元220。当要改变输入BRIR滤波系数或者控制参数信息时,BRIR参数化单元300可以重新计算双耳渲染参数,并且将重新计算的双耳渲染参数传递至双耳渲染单元。

[0105] 根据本发明的示例性实施例,BRIR参数化单元300转换并编辑与双耳渲染器200的输入信号的每个声道或者每个对象相对应的BRIR滤波系数,以将所转换和编辑的BRIR滤波系数传递至双耳渲染单元220。对应的BRIR滤波系数可以是用于每个声道或者每个对象的BRIR滤波器集合中选择的匹配BRIR或者回退BRIR。可以通过针对每个声道或者每个对象的BRIR滤波系数是否存在于虚拟再现空间中来确定BRIR匹配。在这种情况下,可以从用信号通知声道布置的输入参数获取每个声道(或者对象)的位置信息。当存在针对输入信号的相应声道或者相应对象的位置中的至少一个的BRIR滤波系数时,BRIR滤波系数可以是输入信号的匹配BRIR。然而,当不存在针对特定声道或者对象的位置的BRIR滤波系数时,BRIR参数化单元300可以提供针对与对应的声道或者对象最相似的位置的BRIR滤波系数,作为用于对应声道或者对象的回退BRIR。

[0106] 首先,当在BRIR滤波器集合中存在具有在距期望位置(特定声道或对象)的预定范围内的高度和方位偏差的BRIR滤波系数时,可以选择对应的BRIR滤波系数。换言之,可以选择具有与期望位置相同的高度和距期望位置方位偏差在 ± 20 的BRIR滤波系数。当不存在与之对应的BRIR滤波系数时,可以选择BRIR滤波器集合中的具有距期望的位置的最小几何距离的BRIR滤波系数。即,可以选择最小化在对应的BRIR的位置与期望位置之间的几何距离的BRIR滤波系数。此处,BRIR的位置表示与相关BRIR滤波系数相对应的扬声器的位置。此外,两个位置之间的几何距离可以被定义为通过汇聚两个位置之间的高度偏差的绝对值和方位偏差的绝对值所获得的值。同时,根据示例性实施例,通过用于内插BRIR滤波系数的方法,BRIR滤波器集合的位置可以与期望位置匹配。在这种情况下,内插的BRIR滤波系数可以被视为BRIR滤波器集合的一部分。即,在这种情况下,可以实现BRIR滤波系数始终存在于期

望位置处。

[0107] 可以通过分开的矢量来传递与输入信号的每个声道或者每个对象对应的BRIR滤波系数。矢量信息 m_{conv} 指示与在BRIR滤波器集合中的输入信号的每个声道或者对象对应的BRIR滤波系数。例如,当在BRIR滤波器集合中存在具有与输入信号的特定声道的位置信息匹配的位置信息的BRIR滤波系数时,矢量信息 m_{conv} 将相关BRIR滤波系数指示为与该特定声道对应的BRIR滤波系数。然而,当在BRIR滤波器集合中不存在具有与输入信号的特定声道的位置信息匹配的位置信息的BRIR滤波系数时,矢量信息 m_{conv} 将离特定声道的位置信息的几何距离最小的回退BRIR滤波系数指示为与该特定声道对应的BRIR滤波系数。因此,参数化单元300可以通过使用矢量信息 m_{conv} 来确定与在整个BRIR滤波器集合中的输入音频信号的每个声道和对象对应的BRIR滤波系数。

[0108] 同时,根据本发明的另一示例性实施例,BRIR参数化单元300转换并且编辑所有接收到的BRIR滤波系数以将转换和编辑后的BRIR滤波系数传递至双耳渲染单元220。在这种情况下,可以由双耳渲染单元220来进行与输入信号的每个声道和每个对象对应的BRIR滤波系数(可替代地,编辑后的BRIR滤波系数)的选择程序。

[0109] 当BRIR参数化单元300由除了双耳渲染单元220之外的装置构成时,可以将由BRIR参数化单元300生成的双耳渲染参数作为比特流传递至双耳渲染单元220。双耳渲染单元220可以通过将接收到的比特流进行解码来获取双耳渲染参数。在这种情况下,传输的双耳渲染参数包括在双耳渲染单元220的每个子单元中的处理所需要的各种参数,并且可以包括转换或者编辑后的BRIR滤波系数、或者原始BRIR滤波系数。

[0110] 双耳渲染单元220包括快速卷积单元230、后期混响生成单元240、和QTDL处理单元250,并且接收包括多声道和/或多对象信号的多音频信号。在本说明书中,包括多声道和/或多对象信号的输入信号将被称为多音频信号。图7图示了根据示例性实施例的双耳渲染单元220接收QMF域的多声道信号,但是双耳渲染单元220的输入信号可以进一步包括时间域多声道信号和时间域多对象信号。此外,当双耳渲染单元220还包括特定解码器时,输入信号可以是多音频信号的编码后的比特流。此外,在本说明书中,基于进行多音频信号的BRIR渲染的案例对本发明进行了描述,但是本发明不限于此。即,由本发明提供的特征不仅可以应用于BRIR,还可以应用于其它类型的渲染滤波器,并且不仅可以应用于多音频信号,还可以应用于单个声道或者单个对象的音频信号。

[0111] 快速卷积单元230在输入信号与BRIR滤波器之间进行快速卷积以处理用于输入信号的直达声和早期反射声。为此,快速卷积单元230可以通过使用截取BRIR来进行快速卷积。该截取BRIR包括依赖于每个子带频率截取的多个子带滤波系数并且由BRIR参数化单元300生成。在这种情况下,依赖于对应子带的频率确定每个截取子带滤波系数的长度。快速卷积单元230可以通过使用根据子带具有不同长度的截取子带滤波系数来在频域中进行可变阶数滤波。即,针对每个频带,可以在QMF域子带信号和与其对应的QMF域的截取子带滤波器之间进行快速卷积。可以通过上面给出的矢量信息 m_{conv} 来识别与每个子带信号对应的截取子带滤波器。

[0112] 后期混响生成单元240生成用于输入信号的后混响信号。该后期混响信号表示在由快速卷积单元230生成的直达声和早期反射声之后的输出信号。后期混响生成单元240可以基于由从BRIR参数化单元300传递的每个子带滤波系数确定的混响时间信息,来处理

输入信号。根据本发明的示例性实施例,后期混响生成单元240可以生成用于输入音频信号的单声道或者立体声向下混合信号并且对生成的向下混合信号进行后期混响处理。

[0113] QMF域抽头延迟线(QTDL)处理单元250处理在输入音频信号之中的高频带中的信号。QTDL处理单元250从BRIR参数化单元300接收至少一个参数,该参数与在高频带中的每个子带信号对应,并且通过使用接收到的参数来在QMF域中进行抽头延迟线滤波。可以通过在上面给出的矢量信息 m_{conv} 来识别与每个子带信号对应的参数。根据本发明的示例性实施例,双耳渲染器200基于预定的常数或者预定的频带将输入音频信号分成低高频带信号和高频带信号,并且,分别可以由快速卷积单元230和后期混响生成单元240对低高频带信号进行处理,并且可以由QTDL处理单元250对高频带信号进行处理。

[0114] 快速卷积单元230、后期混响生成单元240、和QTDL处理单元250中的每一个输出2-声道QMF域子带信号。混合器&组合器260组合并且混合快速卷积单元230的输出信号、后期混响生成单元240的输出信号、和QTDL处理单元250的输出信号。在这种情况下,针对2声道的左输出信号和右输出信号中的每一个,单独地对输出信号进行组合。双耳渲染器200对组合的输出信号进行QMF分析以在时间域中生成最终的双耳输出音频信号。

[0115] <在频域中的可变阶数滤波(VOFF)>

[0116] 图8是图示了根据本发明的示例性实施例的用于双耳渲染的滤波器生成方法的示意图。转换成多个子带滤波器的FIR滤波器可以用于在QMF域中的双耳渲染。根据本发明的示例性实施例,双耳渲染器的快速卷积单元可以通过使用根据每个子带频率具有不同长度的截取子带滤波器来在QMF域中进行可变阶数滤波。

[0117] 在图8中, F_k 表示用于快速卷积的截取子带滤波器以处理QMF子带 k 的直达声和早期反射声。此外, P_k 表示用于QMF子带 k 的后期混响生成的滤波器。在这种情况下,截取子带滤波器 F_k 可以是原始子带滤波器截取的前滤波器,并且可以将其指定为前子带滤波器。此外,在截取原始子带滤波器之后, P_k 可以是后滤波器,并且可以将其指定为后子带滤波器。QMF域具有总共 K 个子带,并且根据示例性实施例,可以使用64个子带。此外, N 表示原始子带滤波器的长度(标签号)并且 $N_{Filter}[k]$ 表示子带 k 的前子带滤波器的长度。在这种情况下,长度 $N_{Filter}[k]$ 表示在下采样的QMF域中的标签号。

[0118] 在使用BRIR滤波器进行渲染的情况下,可以基于从原始BRIR滤波器提取的参数,即,用于每个子带滤波器的混响时间(RT)信息、能量衰减曲线(EDC)值、能量衰减时间信息等,来确定用于每个子带的滤波器阶数(即,滤波器长度)。由于取决于壁和天花板的材料的空气衰减和吸声程度根据各个频率而发生变化的声学特点,混响时间可以根据频率而变化。一般情况下,具有较低频率的信号具有较长的混响时间。由于长混响时间表示更多的信息保留在FIR滤波器的后部分,因此,优选在正常地传递混响信息中截取对应的滤波器。因此,至少部分地基于从对应的子带滤波器提取的特点信息(例如,混响时间信息)来确定本发明的每个截取子带滤波器 F_k 的长度。

[0119] 根据实施例,可以基于通过用于处理音频信号的设备获得的附加信息,即,复杂性、复杂程度(剖面)、或者需要的解码器的质量信息,来确定截取子带滤波器 F_k 的长度。可以根据用于处理音频信号或者由用户直接输入的值的设备的硬件资源来确定复杂性。可以根据用户的请求来确定质量,或者参照通过比特流或者包括在比特流中的其它信息传输的值来确定质量。此外,还可以根据通过对传输的信号的质量进行估计获得的值来确定质量,

换言之,比特率高,可以将质量视为质量越高。在这种情况下,每个截取子带滤波器的长度可以根据复杂性和质量成比例地增加,并且可以随着针对各个频带得不同比率而变化。此外,为了通过诸如FFT等高速处理来获得附加增益,可以将每个截取子带滤波器的长短确定为对应的大小单元,举例来说,2的幂的倍数。相反,当确定的截取子带滤波器的长度比实际子带滤波器的总长度长时,可以将截取子带滤波器的长度调整为实际子带滤波器的长度。

[0120] 根据本发明的实施例的BRIR参数化单元生成与根据前面提到的示例性实施例确定的截取子带滤波器的相应的长度对应的截取子带滤波系数,并且将生成的截取子带滤波系数传递至快速卷积单元。快速卷积单元通过使用截取子带滤波系数来在多音频信号的每个子带信号的频域中进行可变阶数滤波(VOFF处理)。即,针对彼此不同的频带的第一子带和第二子带,快速卷积单元通过将第一截取子带滤波系数应用于第一子带信号来生成第一子带双耳信号,并且通过将第二截取子带滤波系数应用于第二子带信号来生成第二子带双耳信号。在这种情况下,第一截取子带滤波系数和第二截取子带滤波系数中的每一个都可以独立地具有不同的长度并且从在时间域中的相同原型滤波器获得。即,由于在时间域中的单个滤波器被转换成多个QMF子带滤波器并且与相应子带对应的滤波器的长度发生了变化,因此,从单个原型滤波器获取截取子带滤波器中的每一个。

[0121] 同时,根据本发明的示例性实施例,可以将经过QMF转换的多个子带滤波器分成多个组,并且可以对各个分成的组应用不同的处理。例如,可以基于预定的频带(QMF条带i)来将多个子带分成具有低频率的第一子带组(区1)和具有高频率的第二子带组(区2)。在这种情况下,可以对第一子带组的输入子带信号进行VOFF处理,并且可以对第二子带组的输入子带信号进行即将在下面描述的QTDL处理。

[0122] 因此,BRIR参数化单元生成用于第一子带组的每个子带的截取子带滤波器(前子带滤波器)系数并且将该前子带滤波系数传递至快速卷积单元。快速卷积单元通过使用接收到的前子带滤波系数来进行第一子带组的子带信号的VOFF处理。根据示例性实施例,还可以由后期混响生成单元来进行第一子带组的子带信号的后期混响处理。此外,BRIR参数化单元从第二子带组的子带滤波系数中的每一个获取至少一个参数,并且将获得的参数传递至QTDL处理单元。如在下面所描述的,QTDL处理单元通过使用获得的参数来进行对第二子带组的每个子带信号的抽头延迟线滤波。根据本发明的示例性实施例,可以基于预定的常数值来确定用于区分第一子带组和第二子带组的预定频率(QMF频带i)或者根据传输的音频输入信号的比特流特点来确定。例如,在使用SBR的音频信号的情况下,可以将第二子带组设置为与SBR频带对应。

[0123] 根据本发明的另一示例性实施例,如在图8中所图示的,可以基于预定的第一频带(QMF频带i)和第二频带(QMF频带j)来将多个子带分成三个子带组。即,可以将多个子带分成第一子带组(区1)(该第一子带组(区1)是与第一频带相等或者低于第一频带的低频区)、第二子带组(区2)(该第二子带组(区2)是高于第一频带并且与第二频带相等或者低于第二频带的中频区)、和第三子带组(区3)(该第三子带组(区3)是高于第二频带的高频区)。例如,当总共64个QMF子带(子带索引0至63)被分成3个子带组时,第一子带组可以包括具有索引0至31的总共32个子带;第二子带组可以包括具有索引32至47的总共16个子带;以及第三子带组可以包括具有索引48至63的子带。此处,因为子带频率变低,所以子带索引的值较低。

[0124] 根据本发明的示例性实施例,只可以对第一子带组和第二子带组的子带信号进行

双耳渲染。即,如上所述,可以对第一子带组的子带信号进行VOFF处理和后期混响处理,并且可以对第二子带组的子带信号进行QTDL处理。此外,不可以对第三子带组的子带信号进行双耳渲染。同时,用于进行双耳渲染的最大频率的信息($K_{proc}=48$)和用于进行卷积的频带的信息($K_{conv}=32$)可以是预定的值,或者由BRIR参数化单元确定以传递至双耳渲染单元。在这种情况下,将第一频带(QMF频带*i*)设置为索引 $K_{conv}-1$ 的子带并且将第二频带(QMF频带*j*)设置为索引 $K_{proc}-1$ 的子带。同时,可以通过原始BRIR输入的采样频率、输入音频信号的采样频率等来改变用于进行卷积的最大频带的信息(K_{proc})和频带的信息(K_{conv})的值。

[0125] 同时,根据图8的示例性实施例,还可以基于从原始子带滤波器以及前子带滤波器 F_k 提取的参数来确定后子带滤波器 P_k 的长度。即,至少部分地基于在对应的子带滤波器中提取的特点信息来确定每个子带的前子带滤波器和后子带滤波器的长度。例如,可以基于对应的子带滤波器的第一混响信息来确定前子带滤波器的长度,并且可以基于第二混响时间信息来确定后子带滤波器的长度。即,基于在原始子带滤波器中的第一混响时间信息,前子带滤波器可以是在截取的前部分处的滤波器,并且后子带滤波器可以是在第一混响时间与第二混响时间之间的区对应的后部分处的滤波器,该区是在前子带滤波器之后的区。根据示例性实施例,第一混响时间信息可以是RT20,并且第二混响时间信息可以是RT60,但是本发明不限于此。

[0126] 在第二混响时间内,存在将早期反射声部分转换为后期混响声部分的部分。即,存在将具有确定特点的区转换为具有随机特点的区的点,并且,在整个频带的BRIR方面,将该点称为混合时间。在混合时间之前的区的情况下,主要存在针对每个位置提供方向性的信息,并且该信息对每个声道都是唯一的。相反,由于后期混响部分针对每个声道具有共同特征,因此,每次对多个声道进行处理可能是高效的。因此,对每个子带的混合时间进行估计,以在混合时间之前通过VOFF处理进行快速卷积,并且在混合时间之后通过后期混响处理来反映针对每个声道的共同特点的处理。

[0127] 然而,从感知的观点看,在估计混合时间时,可能会由于偏置而发生错误。因此,从质量的观点看,通过将VOFF处理部分的长度最大化来进行快速卷积比通过估计精确的混合时间来基于对应的边界分开地对VOFF处理部分和后期混响部分进行处理更好。因此,根据复杂度-质量控制,VOFF处理部分的长度(即,前子带滤波器的长度)可以比与混合时间对应的长度更长或者更短。

[0128] 此外,为了减少每个子带滤波器的长度,除了前面提到的截取方法之外,当特定子带的频率响应是单调的时,可以使用将对应子带的滤波器减少到低阶数的建模。作为代表性的方法,存在使用频率采样的FIR滤波建模,并且可以设计从最小二乘的观点看被最小化的滤波。

[0129] <高频频带的QTDL处理>

[0130] 图9是图示了根据本发明的示例性实施例的更加具体地图示了QTDL处理的框图。根据图9的示例性实施例,QTDL处理单元250通过使用单抽头延迟线滤波器来对多声道输入信号 X_0 、 X_1 、 \dots 、 X_{M-1} 进行子带专用滤波。在这种情况下,假定多声道输入信号作为QMF域的子带信号而被接收。因此,在图9的示例性实施例中,单抽头延迟线滤波器可以对每个QMF子带进行处理。该单抽头延迟线滤波器针对每个声道信号进行仅一个抽头的卷积。在这种

情况下,可以基于从与相关子带信号对应的BRIR子带滤波系数直接提取的参数来确定使用的抽头。该参数包括待在单抽头延迟线滤波器中使用的抽头的延迟信息、以及与其对应的增益信息。

[0131] 在图9中, L_0 、 L_1 、 \dots 、 L_{M-1} 分别表示针对左耳M声道的BRIR的延迟,并且 R_0 、 R_1 、 \dots 、 R_{M-1} 分别表示针对右耳M声道的BRIR的延迟。在这种情况下,延迟信息表示在BRIR子带滤波系数中的最大峰值(按照绝对值的顺序)的位置信息、实部的值、或者虚部的值。此外,在图9中,分别地, G_{L_0} 、 G_{L_1} 、 \dots 、 $G_{L_{M-1}}$ 表示与左声道的相应延迟信息对应的增益,并且 G_{R_0} 、 G_{R_1} 、 \dots 、 $G_{R_{M-1}}$ 表示与右声道的相应延迟信息对应的增益。可以基于对应BRIR子带滤波系数的总功率、与延迟信息对应的峰值的大小等,来确定每个增益信息。在这种情况下,作为增益信息,可以使用在对全部子带滤波系数进行能量补偿之后的对应峰值的加权值、以及在子带滤波系数中的对应的峰值本身。通过使用对应峰值的加权值的实数和加权值的虚数来获取增益信息。

[0132] 同时,如上所述,可以仅对高频带的输入信号进行QTDL处理,基于预定的常数或者预定的频道对该高频带的输入信号进行分类。当将频带复制(SBR)应用于输入音频信号时,高频带可以与SBR频带对应。用于对高频带有效编码的频带复制(SBR)是用于通过重新扩展带宽来保证带宽与原始信号的长度一样长的工具,该带宽通过将低比特率编码中的高频带的信号扔出而变窄。在这种情况下,通过使用进行了编码和传输的低频带的信息和通过编码器传输的高频带的附加信息,来生成高频带。然而,由于不准确的谐波的生成,在通过使用SBR生成的高频率组件中可能会发生失真。此外,SBR子带是高频子带,并且如上所述,对应频带的混响时间非常短。即,SBR频带的BRIR子带滤波器具有少量有效信息和高衰减率。因此,在与SBR频带对应的高频带的BRIR渲染中,在对音质的计算复杂度方面,通过使用少量的有效抽头来进行渲染可能比进行卷积更有效。

[0133] 由单抽头延迟线滤波器滤波的多个声道信号被聚合为用于每个子带的2-声道左输出信号 Y_L 和右输出信号 Y_R 。同时,在双耳渲染的初始化过程期间,可以将用于QTDL处理单元250的每个单抽头延迟线滤波器中使用的参数存储在存储器中,并且,可以在不对提取的参数进行另外的操作的情况下,进行QTDL处理。

[0134] <BRIR参数化的细节>

[0135] 图10是图示了根据本发明的示例性实施例的BRIR参数化单元的相应组件的框图。如在图14中所图示的,BRIR参数化单元300可以包括VOFF参数化单元320、后期混响参数化单元360、和QTDL参数化单元380。BRIR参数化单元300接收时间域的BRIR滤波器集合作为输入,并且BRIR参数化单元300的每个子单元通过使用接收到的BRIR滤波器集合来生成用于双耳渲染的各种参数。根据示例性实施例,BRIR参数化单元300还可以接收控制参数,并且基于接收到的控制参数生成参数。

[0136] 首先,VOFF参数化单元320生成在频域(VOFF)中的可变阶数滤波需要的截取子带滤波系数、和由此产生的辅助参数。例如,VOFF参数化单元320计算用于生成截取子带滤波系数的频带专用混响时间信息、滤波器阶数信息等,并且确定用于对截取子带滤波系数进行逐框式快速傅里叶变换的框的大小。可以强VOFF参数化单元320生成的一些参数传递至后期混响参数化单元360和QTDL参数化单元380。在这种情况下,传递的参数不限于VOFF参数化单元320的最终输出值,并且可以包括根据VOFF参数化单元320的处理生成的参数,即,

时间域的截取BRIR滤波系数等。

[0137] 后期混响参数化单元360生成后期混响生成需要的参数。例如,后期混响参数化单元360可以生成向下混合子带滤波系数、IC值等。此外,QTDL参数化单元380生成用于QTDL处理的参数。更详细地,QTDL参数化单元360从后期混响参数化单元320接收子带滤波系数,并且通过使用接收到的滤波系数来在每个子带中生成延迟信息和增益信息。在这种情况下,QTDL参数化单元380可以接收用于进行双耳渲染的最大频带的Kproc信息和用于进行卷积的频带的信息Kconv作为控制参数,并且为具有Kproc和Kconv的子带组的每个频带生成延迟信息和增益信息作为边界。根据示例性实施例,可以将QTDL参数化单元380提供为包括在VOFF参数化单元320中的组件。

[0138] 将分别在VOFF参数化单元320、后期混响参数化单元360、和QTDL参数化单元380中生成的参数传递双耳渲染单元(未图示)。根据示例性实施例,后期混响参数化单元360和QTDL参数化单元分别可以根据在双耳渲染单元中是否进行了后期混响处理和QTDL处理来确定是否生成了参数。当在双耳渲染单元中没有进行后期混响处理和QTDL处理中的至少一个时,与其对应的后期混响参数化单元360和QTDL参数化单元380可以不生成参数或者可以不将生成的参数传输至双耳渲染单元。

[0139] 图11是图示了本发明的VOFF参数化单元的相应组件的框图。如图所示,VOFF参数化单元320可以包括传播时间计算单元322、QMF转换单元324、和VOFF参数生成单元330。VOFF参数化单元320通过使用接收到的时间域BRIR滤波系数来进行生成用于VOFF处理的截取子带滤波系数的过程。

[0140] 首先,传播时间计算单元322计算时间域BRIR滤波系数的传播时间信息,并且基于计算得到的传播时间信息截取时间域BRIR滤波系数。此处,传播时间信息表示从初始样本到BRIR滤波系数的直达声的时间。传播时间计算单元322可以从时间域BRIR滤波系数截取与计算得到的传播时间对应的部分,并且移除该截取的部分。

[0141] 各种方法可以用于估计BRIR滤波系数的传播时间。根据示例性实施例,可以基于第一点信息来估计传播时间,其中,示出了比与BRIR滤波系数的最大峰值成比例的阈值大的能量值。在这种情况下,由于从多声道输入的相应声道到收听器的所有距离彼此不同,因此,传播时间可以针对每个声道而变化。然而,所有声道的传播时间的截取长度需要彼此相同,以便通过使用BRIR滤波系数来进行卷积,在该卷积中,在进行双耳渲染时截取传播时间,并且补偿利用延迟进行了双耳渲染的最终信号。此外,当通过将相同的传播时间信息应用于每个声道来进行截取时,可以降低在单独的声道中的错误发生概率。

[0142] 根据本发明的示例性实施例,为了计算传播时间信息,可以首先限定用于诸帧索引k的帧能量E(k)。当用于输入声道索引m的时间域BRIR滤波系数、输出左/右声道索引i、和时间域的时隙索引v是 $\tilde{h}_{i,m}^v$ 时,可以通过下面给出的等式来计算在第k个帧中的帧能量E(k)。

[0143] [等式2]

$$[0144] \quad E(k) = \frac{1}{2N_{BRIR}} \sum_{m=1}^{N_{BRIR}} \sum_{i=0}^1 \frac{1}{L_{frm}} \sum_{n=0}^{L_{frm}-1} \sum_{i,m} h_{i,m}^{kN_{hop}+n}$$

[0145] 其中, N_{BRIR} 表示BRIR滤波器集合的总共的滤波器的数量; N_{hop} 表示预定的跳数大小;以及 L_{frm} 表示帧大小。即,针对相同的时间间隔,可以将帧能量E(k)计算为用于每个声道

的帧能量的平均值。

[0146] 可以通过使用定义的帧能量 $E(k)$ 通过在下面给出的等式来计算传播时间 pt 。

[0147] [等式3]

$$[0148] \quad pt = \frac{L_{frm}}{2} + N_{hop} * m \ln \left[\arg \left(\frac{E(k)}{\max(E)} \right) > -60dB \right]$$

[0149] 即,传播时间计算单元322通过改变预定的跳数来测量帧能量,并且识别帧能量大于预定阈值的第一帧。在这种情况下,可以将传播时间确定为识别出的第一帧的中间点。同时,在等式3中,描述了将阈值设置为比最大帧能量低60dB的值,但是本发明不限于此,并且可以将阈值设置为与最大帧能量成比例的值或者与最大帧能量相差预定值的值。

[0150] 同时,可以基于输入BRIR滤波系数是否是头相关脉冲响应(HRIR)滤波系数来改变跳数大小 N_{hop} 和帧大小 L_{frm} 。在这种情况下,可以从外部接收或者通过使用时间域BRIR滤波系数的长度来估计指示输入BRIR滤波系数是否是HRIR滤波系数的信息 $flag_HRIR$ 。一般情况下,早期反射声部分和后期混响部分的边界已知为80 ms。因此,当时间域BRIR滤波系数的长度是80 ms或者更小时,将对应的BRIR滤波系数确定为HRIR滤波系数($flag_HRIR=1$),并且当时间域BRIR滤波系数的长度大于80ms时,可以确定对应的BRIR滤波系数不是HRIR滤波系数($flag_HRIR=0$)。当确定输入BRIR滤波系数是HRIR滤波系数($flag_HRIR=1$)时,可以将跳数大小 N_{hop} 和帧大小 L_{frm} 设置为比确定对应的BRIR滤波系数不是HRIR滤波系数($flag_HRIR=0$)时的值更小的值。例如,在 $flag_HRIR=0$ 的情况下,可以分别将跳数大小 N_{hop} 和帧大小 L_{frm} 设置为8个样本和32个样本,并且在 $flag_HRIR=1$ 的情况下,可以分别将跳数大小 N_{hop} 和帧大小 L_{frm} 设置为1个样本和8个样本。

[0151] 根据本发明的示例性实施例,传播时间计算单元322可以基于计算得到的传播时间信息来截取时间域BRIR滤波系数,并且将该截取的BRIR滤波系数传递至QMF转换单元324。此处,截取BRIR滤波系数指示在从原始BRIR滤波系数截取并且移除与传播时间对应的部分之后剩余的滤波系数。传播时间计算单元322截取用于每个输入声道和每个输出左/右声道的时间域BRIR滤波系数,并且将截取的时间域BRIR滤波系数传递至QMF转换单元324。

[0152] QMF转换单元324在时间域与QMF之间进行输入BRIR滤波系数的转换。即,QMF转换单元324接收时间域的截取的BRIR滤波系数并且分别将接收到的BRIR滤波系数转换为与多个频带对应的多个子带滤波系数。将转换后的子带滤波系数传递至VOFF参数生成单元330,并且VOFF参数生成单元330通过使用接收到的子带滤波系数来生成截取子带滤波系数。当QMF域BRIR滤波系数而非时间域BRIR滤波系数作为VOFF参数化单元320的输入而被接收时,该接收到的QMF域BRIR滤波系数可以绕过QMF转换单元324。此外,根据另一示例性实施例,当输入滤波系数是QMF域BRIR滤波系数时,在VOFF参数化单元320中,可以省略QMF转换单元324。

[0153] 图12是图示了图11的VOFF参数生成单元的详细配置的框图。如图所示,VOFF参数生成单元330可以包括混响时间计算单元332、滤波器阶数确定单元334、和VOFF滤波系数生成单元336。VOFF参数生成单元330可以从图11的QMF转换单元324接收QMF域子带滤波系数。此外,可以将包括进行双耳渲染的最大频带信息 K_{proc} 、进行卷积的频带信息 K_{conv} 、预定的最大FFT大小信息等的控制参数输入到VOFF参数生成单元330中。

[0154] 首先,混响时间计算单元332通过使用接收到的子带滤波系数来获取混响时间信

息。可以将获得的混响时间信息传递至滤波器阶数确定单元334,并且可以将该混响时间信息用于确定对应子带的滤波器阶数。同时,由于根据测量环境在混响时间信息中可能存在偏置和偏差,因此,可以通过使用与另一声道的相互关系来使用统一的值。根据示例性实施例,混响时间计算单元332生成每个子带的平均混响时间信息并且将生成的平均混响时间信息传递至滤波器阶数确定单元334。当用于输入声道索引 m 、输出左/右声道索引 i 、和子带索引 k 的子带滤波系数的混响时间信息是 $RT(k, m, i)$ 时,可以通过下面给出的等式来计算子带 k 的平均混响时间信息 RT^k 。

[0155] [等式4]

$$[0156] \quad RT^k = \frac{1}{2N_{BRIR}} \sum_{i=0}^1 \sum_{m=0}^{N_{BRIR}-1} RT(k, m, i)$$

[0157] 其中, N_{BRIR} 表示BRIR滤波器集合的总共滤波器的数量。

[0158] 即,混响时间计算单元332从与多声道输入对应的每个子带滤波系数提取混响时间信息 $RT(k, m, i)$,并且获取针对相同子带提取的每个声道的混响时间信息 $RT(k, m, i)$ 的平均值(即,平均混响时间信息 RT^k)。可以将获得的平均混响时间信息 RT^k 传递至滤波器阶数确定单元334,并且滤波器阶数确定单元334可以通过使用传递的平均混响时间信息 RT^k 来确定应用于对应子带的单个滤波器阶数。在这种情况下,该获得的平均混响时间信息可以包括 RT_{20} ,并且根据示例性实施例,可以包括其它混响时间信息,换言之,也可以获取 RT_{30} 、 RT_{60} 等。同时,根据本发明的示例性实施例,混响时间计算单元332可以向滤波器阶数确定单元334传递针对相同子带提取的每个声道的混响时间信息的最大值和/或最小值作为对应子带的代表性混响时间信息。

[0159] 接下来,滤波器阶数确定单元334基于获得的混响时间信息来确定对应子带的滤波器阶数。如上所述,通过滤波器阶数确定单元334获得的混响时间信息可以是对应子带的平均混响时间信息,并且,相反,根据示例性实施例,可以获取具有每个声道的混响时间信息的最大值和/或最小值的代表性混响时间信息。滤波器阶数可以用于确定用于对应子带的双耳渲染的截取子带滤波系数的长度。

[0160] 当在子带 k 中的平均混响时间信息是 RT^k 时,可以通过下面给出的等式获取对应子带的滤波器阶数信息 $N_{Filter}[k]$ 。

[0161] [等式5]

$$[0162] \quad N_{Filter}[k] = 2^{\lfloor \lg_2 RT^k + 0.5 \rfloor}$$

[0163] 即,可以通过将对应子带的平均混响时间信息的对数缩放近似整数作为索引来将滤波器阶数信息确定为2的幂的值。换言之,可以通过将按照对数标尺的对应子带的平均混响时间信息的四舍五入值、上舍入值、或者下舍入值作为索引,来将滤波器阶数信息确定为2的幂的值。当对应子带滤波系数的原始长度(即,到最后时隙 n_{end} 的长度)比在等式5中确定的值小时,可以用子带滤波系数的原始长度值 n_{end} 来替代滤波器阶数信息。即,可以将滤波器阶数信息确定为通过等式5确定的参考截取长度和子带滤波系数的原始长度中较小的一个值。

[0164] 同时,可以按照对数标尺对取决于频率的能量衰减线性地取近似值。因此,当使用曲线拟合方法时,可以确定每个子带的优化滤波器阶数信息。根据本发明的示例性实施例,

滤波器阶数确定单元334可以通过使用多项式曲线拟合方法来获取滤波器阶数信息。为此，滤波器阶数确定单元334可以获取用于平均混响时间信息的曲线拟合的至少一个系数。例如，滤波器阶数确定单元334通过对数标尺的线型方程来进行每个子带的平均混响时间信息的曲线拟合，并且获取对应线型方程的斜率值‘a’和片段值‘b’。

[0165] 可以通过使用获得的系数通过下面给出的等式来获取在子带k中的曲线拟合的滤波器阶数信息 $N'_{Filter}[k]$ 。

[0166] [等式6]

$$[0167] \quad N'_{Filter}[k] = 2^{\lfloor bk+a+0.5 \rfloor}$$

[0168] 即，可以通过将对应子带的平均混响时间信息的多项式曲线拟合值的近似整数值用作索引，来将曲线拟合的滤波器阶数信息确定为2的幂的值。换言之，可以通过将对应子带的平均混响时间信息的多项式曲线拟合值的四舍五入值、上舍入值、或者下舍入值的2的幂的值作为索引，来将曲线拟合的滤波器阶数信息确定为2的幂的值。当对应子带滤波系数的原始长度，即，到最后时隙 n_{end} 的长度，比在等式6中确定的值小时，可以用子带滤波系数的原始长度值 n_{end} 来替代滤波器阶数信息。即，可以将滤波器阶数信息确定为通过等式6确定的参考截取长度和子带滤波系数的原始长度中较小的一个值。

[0169] 根据本发明的示例性实施例，基于原型BRIR滤波系数（即，时间域的BRIR滤波系数）是否是HRIR滤波系数（flag_HRIR），可以通过使用等式5和等式6中的任何一个来获取滤波器阶数信息。如上所述，可以基于原型BRIR滤波系数的长度是否大于预定值来确定flag_HRIR的值。当原型BRIR滤波系数的长度大于预定值（即，flag_HRIR=0）时，可以根据在上面给出的等式6将滤波器阶数信息确定为曲线拟合值。然而，当原型BRIR滤波系数的长度不大于预定值（即，flag_HRIR=1）时，可以根据在上面给出的等式5将滤波器阶数信息确定为非曲线拟合值。即，可以在不进行曲线拟合的情况下，基于对应子带的平均混响时间信息来确定滤波器阶数信息。其原因在于，由于HRIR不受房间的影响，因此，在HRIR中，能量延迟的趋势不明显。

[0170] 同时，根据本发明的示例性实施例，当获得第0个子带（即，子带索引0）的滤波器阶数信息时，可以使用未进行曲线拟合的平均混响时间信息。其原因在于，由于房间模式的影响等，第0个子带的混响时间可以具有与另一子带的混响时间不同的趋势。因此，根据本发明的示例性实施例，仅在flag_HRIR=0的情况下和在索引不为0的子带中才可以使根据等式6的曲线拟合滤波器阶数信息。

[0171] 将根据在上面给出的示例性实施例确定的每个子带的滤波器阶数信息传递至VOFF滤波系数生成单元336。VOFF滤波系数生成单元336基于获得的滤波器阶数信息生成截取子带滤波系数。根据本发明的示例性实施例，截取子带滤波系数可以由至少一个FFT滤波系数构成，其中，通过用于逐框式快速卷积的预定框形式来进行快速傅里叶变换（FFT）。如在下面参照图14所描述的，VOFF滤波系数生成单元336可以生成用于逐框式快速卷积的FFT滤波系数。

[0172] 图13是图示了本发明的QTDL参数化单元的相应组件的框图。

[0173] 如在图13中所图示的，QTDL参数化单元380可以包括峰值搜索单元382和增益生成单元384。QTDL参数化单元380可以从VOFF参数化单元320接收QMF域子带滤波系数。此外，

QTDL参数化单元380可以接收用于进行双耳渲染的最大频带的信息Kproc和用于进行卷积的频带的信息Kconv作为控制参数,并且为具有Kproc与Kconv的子带组(即,第二子带组)的每个频带生成延迟信息和增益信息作为边界。

[0174] 根据更详细的示例性实施例,如在下面所描述的,当用于输入声道索引m、输出左/右声道索引i、子带索引k、和QMF域时隙索引n的BRIR子带系数是 $h_{i,m}^k(n)$ 时,可以如下获取延迟信息 $d_{i,m}^k$ 和增益信息 $g_{i,m}^k$ 。

[0175] [等式7]

$$[0176] \quad d_{i,m}^k = \underset{n}{\operatorname{argmax}} \left(|h_{i,m}^k(n)|^2 \right)$$

[0177] [等式8]

$$[0178] \quad g_{i,m}^k = \frac{\sqrt{\sum_{l=0}^{n_{\text{end}}} |h_{i,m}^k(l)|^2}}{|h_{i,m}^k(d_{i,m}^k)|} h_{i,m}^k(d_{i,m}^k)$$

[0179] 其中, n_{end} 表示对应的子带滤波系数的最后时隙。

[0180] 即,参照等式7,延迟信息可以表示时隙的信息,其中,对应的BRIR子带滤波系数具有最大的大小,并且这表示对应的BRIR子带滤波系数的最大峰值的位置信息。此外,参照等式8,可以将增益信息确定为通过将对应的BRIR子带滤波系数的总功率值乘以在最大峰值位置处的BRIR子带滤波系数的符号而获得的值。

[0181] 峰值搜索单元382基于等式7获取最大峰值位置,即第二子带组的每个子带滤波系数的延迟信息。此外,增益单元384基于等式8获取针对每个子带滤波系数的增益信息。等式7和等式8示出了获取延迟信息和增益信息的等式的示例,但是,可以队用于计算每种信息的等式的具体形式进行各种修改。

[0182] <逐框式快速卷积>

[0183] 同时,根据本发明的示例性实施例,可以进行预定的逐框式快速卷积,以便在效率和性能方面获得最佳的双耳效果。基于FFT的快速卷积的特征在于:随着FFT大小增加,计算量减少,但是整体处理延迟增加并且内存使用量增加。当将长度为1秒的BRIR快速卷积为长度是对应长度的两倍的FFT大小时,在计算量方面是高效的,但是发生了与1秒对应的延迟,并且需要与之对应的缓存和处理存储器。具有长延迟时间的音频信号处理方法不适于进行实时数据处理等的应用。由于帧是可以由音频信号处理设备进行解码的最小单元,因此,甚至是在双耳渲染中,也优选地按照与帧单元对应的大小来进行逐框式快速卷积。

[0184] 图14图示了用于生成用于逐框式快速卷积的FFT滤波系数的方法的示例性实施例。与前面提到的示例性实施例相似,在图14的示例性实施例中,将原型FIR滤波器转换为K子带滤波器,并且Fk和Pk分别表示子带k的截取子带滤波器(前子带滤波器)和后子带滤波器。子带Band 0至Band K-1中的每一个可以表示在频域中的子带,即QMF子带。在QMF域中,可以使用总共64个子带,但是本发明不限于此。此外,N表示原始子带滤波器的长度(抽头的数量)并且 $N_{\text{Filter}}[k]$ 表示子带k的前子带滤波器的长度。

[0185] 与前面提到的示例性实施例一样,可以基于预定的频带(QMF子带i)来将QMF域的多个子带分成具有低频率的第一子带组(区1)和具有高频率的第二子带组(区2)。可替代

地,可以基于预定的第一频带(QMF频带i)和第二频带(QMF频带j)来将多个子带分成三个子带组,即:第一子带组(区1)、第二子带组(区2)、和第三子带组(区3)。在这种情况下,分别可以通过使用逐框式快速卷积来对第一子带组的输入子带信号进行VOFF处理,并且可以对第二子带组的输入子带信号进行QTDL处理。另外,可以不对第三子带组的子带信号进行渲染。根据示例性实施例,还可以对第一子带组的输入子带信号进行后期混响处理。

[0186] 参照图14,本发明的VOFF滤波系数生成单元336按照对应子带中的预定框大小来进行截取子带滤波系数的快速傅里叶变换以生成FFT滤波系数。在这种情况下,基于预定的最大FFT大小 $2L$ 来确定在子带 k 中的预定框的长度 $N_{FFT}[k]$ 。更详细地,可以通过下面的等式来表达在子带 k 中的预定框的长度 $N_{FFT}[k]$ 。

[0187] [等式9]

$$[0188] \quad N_{FFT}[k] = \min(2L, 2^{\lceil \log_2 2N_{Filter}[k] \rceil})$$

[0189] 其中, $2L$ 表示预定的最大FFT大小并且 $N_{Filter}[k]$ 表示子带 k 的滤波器阶数信息。

[0190] 即,可以将预定框的长度 $N_{FFT}[k]$ 确定为在是截取子带滤波系数的两倍的值 $2^{\lceil \log_2 2N_{Filter}[k] \rceil}$ 与预定的最大FFT大小 $2L$ 之间的较小值。此处,参考滤波器长度表示对应子带 k 中的滤波器阶数 $N_{Filter}[k]$ 的2的幂的形式的真实值和近似值中的任何一个。即,当子带 k 的滤波器阶数具有2的幂的形式时,将对应的滤波器阶数 $N_{Filter}[k]$ 用作在子带 k 中的参考滤波器长度,并且当子带 k 的滤波器阶数 $N_{Filter}[k]$ 不具有2的幂的形式(例如, n_{end})时,将对应的滤波器阶数 $N_{Filter}[k]$ 的2的幂的形式的四舍五入值、上舍入值或者下舍入值用作参考滤波器长度。同时,根据本发明的示例性实施例,预定框的长度 $N_{FFT}[k]$ 和参考滤波器长度 $2^{\lceil \log_2 2N_{Filter}[k] \rceil}$ 两者都可以是2的幂的值。

[0191] 当是参考滤波器长度的两倍的值等于或者大于(或者,大于)最大FFT大小 $2L$ 时(如,图14的F0和F1),将对应子带的预定框长度 $N_{FFT}[0]$ 和 $N_{FFT}[1]$ 中的每一个确定为最大FFT大小 $2L$ 。然而,当是参考滤波器长度的两倍的值小于(或者,等于或者小于)最大FFT大小 $2L$ 小时(如,图14的F5),将对应子带的预定框长度 $N_{FFT}[5]$ 确定为 $2^{\lceil \log_2 2N_{Filter}[5] \rceil}$,其是参考滤波器长度的两倍的值。如在下面描述的,由于通过补零将截取子带滤波系数扩展为双倍长度,并且之后,进行了快速傅里叶变换,因此,可以基于在是参考滤波器程度的两倍的值与预定最大FFT大小 $2L$ 之间的比较结果来确定用于快速傅里叶变换的框的长度 $N_{FFT}[k]$ 。

[0192] 如上所述,当确定了在子带中的框长度 $N_{FFT}[k]$ 时,VOFF滤波系数生成单元336通过预定的框大小来对截取子带滤波系数进行快速傅里叶变换。更详细地,VOFF滤波系数生成单元336按照预定的框大小的一半 $N_{FFT}[k]/2$ 来划分截取子带滤波系数。在图14中图示的VOFF处理部分的虚线边界所在的区域表示按照预定的框大小的一半划分得到的子带滤波系数。接下来,BRIR参数化单元通过使用相应划分的滤波系数来生成预定框大小的临时滤波系数。在这种情况下,临时滤波系数的前半部分由划分的滤波系数构成,并且后半部分由补零值构成。因此,通过使用预定框的半个长度 $N_{FFT}[k]/2$ 的滤波系数来生成预定框的长度为 $N_{FFT}[k]$ 的临时滤波系数。接下来,BRIR参数化单元对生成的临时滤波系数进行快速傅里叶变换以生成FFT滤波系数。生成的FFT滤波系数可以用于对输入音频信号进行预定的逐框式快速卷积。

[0193] 如上所述,根据本发明的示例性实施例,VOFF滤波系数生成单元336按照针对每个子带独立确定的框大小来对截取子带滤波系数进行快速傅里叶变换以生成FFT滤波系数。因此,可以进行针对每个子带使用不同数量的框的快速卷积。在这种情况下,在子带k中的框的数量 $N_{blk}[k]$ 可以满足以下等式。

[0194] [等式10]

$$[0195] \quad N_{bk}[k] = \frac{2^{\lceil \log_2 2N_{Filter}[k] \rceil}}{N_{FFT}[k]}$$

[0196] 其中, $N_{bk}[k]$ 是自然数。

[0197] 即,可以将子带k中的框的数量确定为通过将是对应子带中的参考滤波器长度的两倍的值除以预定框的长度 $N_{FFT}[k]$ 而得到的值。

[0198] 同时,根据本发明的示例性实施例,可以限制性地对第一子带组的前子带滤波器 F_k 进行预定的逐框式FFT滤波系数的生成过程。同时,根据示例性实施例,可以通过在上面描述的后期混响生成单元对第一子带组的子带信号进行后期混响处理。根据本发明的示例性实施例,可以基于原型BRIR滤波系数的长度是否大于预定值来对输入音频信号进行后期混响处理。如上所述,可以通过指示原型BRIR滤波系数的长度大于预定值的标志(即,flag_BRIR),来表示原型BRIR滤波系数的长度是否大于预定值。当原型BRIR滤波系数的长度大于预定值时(flag_BRIR=0),可以对输入音频信号进行后期混响处理。然而,当原型BRIR滤波系数的长度不大于预定值时(flag_BRIR=1),可以不对输入音频信号进行后期混响处理。

[0199] 当未进行后期混响处理时,只可以对第一子带组的每个子带信号进行VOFF处理。然而,针对VOFF处理指定的每个子带的滤波器阶数(即,截取点)可以小于对应的子带滤波系数的总长度,因此,可能发生能量不匹配。因此,为了防止能量不匹配,根据本发明的示例性实施例,可以基于flag_BRIR信息来对截取子带滤波系数进行能量补偿。即,当原型BRIR滤波系数的长度不大于预定值时(flag_BRIR=1),可以将进行了能量补偿的滤波系数用作截取子带滤波系数或者构成该截取子带滤波系数的每个FFT滤波系数。在这种情况下,可以通过将直到基于滤波器阶数信息 $N_{Filter}[k]$ 的截取点的子带滤波系数除以直到截取点的滤波功率,并且乘以对应的子带滤波系数的总滤波功率,来进行能量补偿。可以将总滤波功率定义为从对应子带滤波系数的初始样本滤波到最终样本 n_{end} 的滤波系数的功率之和。

[0200] 同时,根据本发明的示例性实施例,针对每个声道,可以将相应的子带滤波系数的滤波器阶数设置为彼此不同。例如,可以将前声道(其中,输入信号包括更多的能量)的滤波器阶数设置为高于后声道(其中,输入信号包括相对较少的能量)的滤波器阶数。因此,针对前声道,提高在双耳渲染之后反映的分辨率,并且,针对后声道,可以低计算复杂度进行渲染。此处,前声道和后声道的分类不限于分配给多声道输入信号的每个声道的声道名称,并且可以基于预定空间参考将相应的声道分成前声道和后声道。此外,根据本发明的另外的示例性实施例,可以基于预定空间参考将多声道的相应声道分成三个或者更多个声道组,并且,针对每个声道组,可以使用不同的滤波器阶数。可替代地,针对与相应声道对应的子带滤波系数的滤波器阶数,可以使用基于在虚拟再现空间中的对应声道的位置信息应用了不同加权值的值。

[0201] 在上文中,已经通过详细的示例性实施例对本发明进行了描述,但是,在不脱离本发明的目标和范围的情况下,本领域的技术人员可以对本发明进行修改和改变。即,在本发

明中,已经对针对多音频信号的双耳渲染的示例性实施例进行了描述,但是甚至可以将本发明相似地应用于或者扩展为包括视频信号以及音频信号的各种多媒体信号。因此,据分析,本领域的技术人员通过详细描述可以容易地类推的主题、以及本发明的示例性实施例都包括在本发明的权利要求书中。

[0202] 发明的实施方式

[0203] 如上所述,已经按照最佳实施方式对相关的特征进行了描述。

[0204] 工业实用性

[0205] 本发明可以适用于处理多媒体信号的各种形式的设备,包括用于处理音频信号的设备 and 用于处理视频信号的设备等。

[0206] 此外,本发明可以适用于生成用于音频信号处理和视频信号处理的参数的参数化装置。



图1

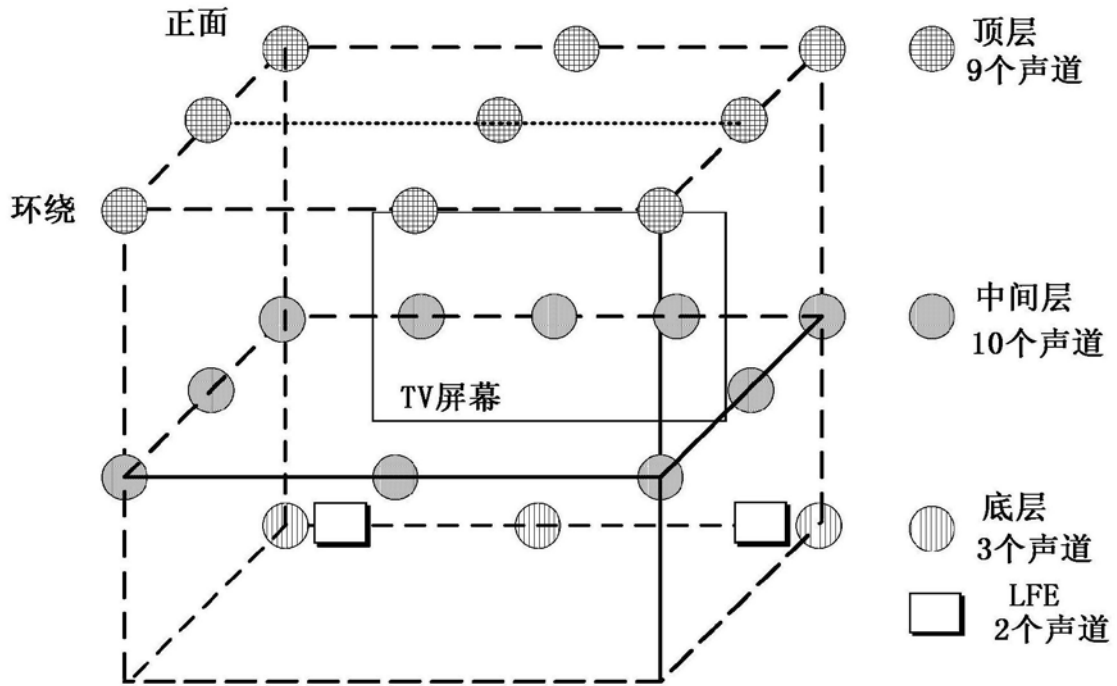


图2

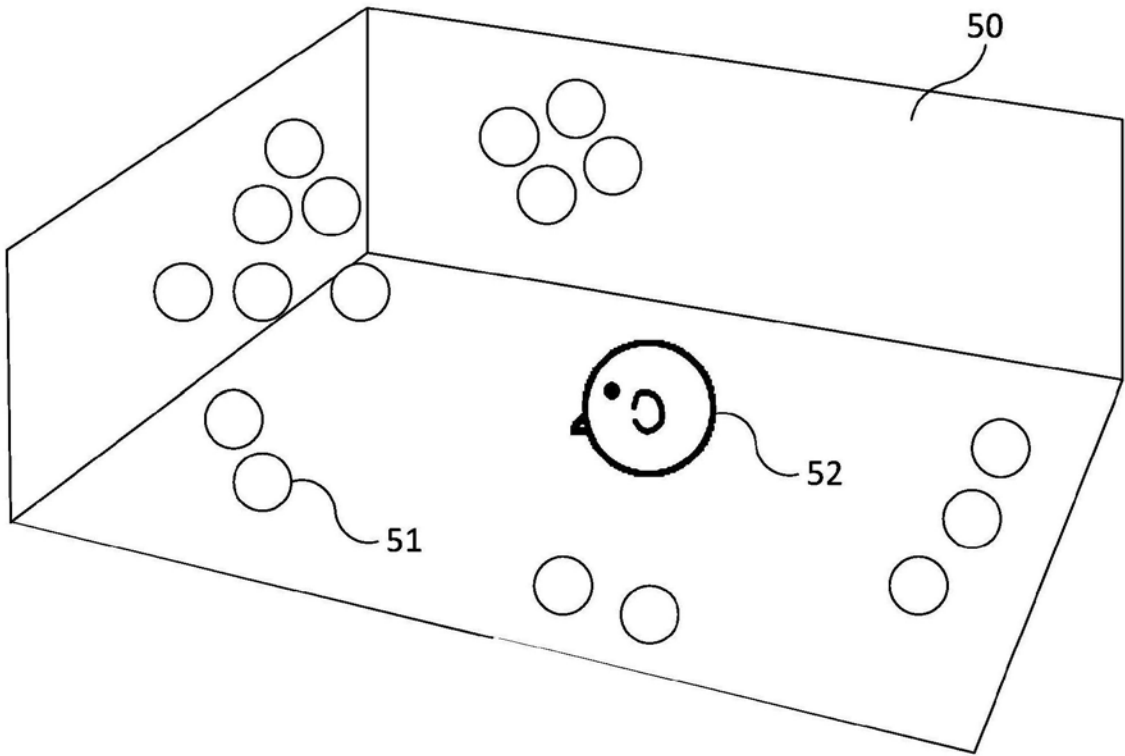


图3

1200

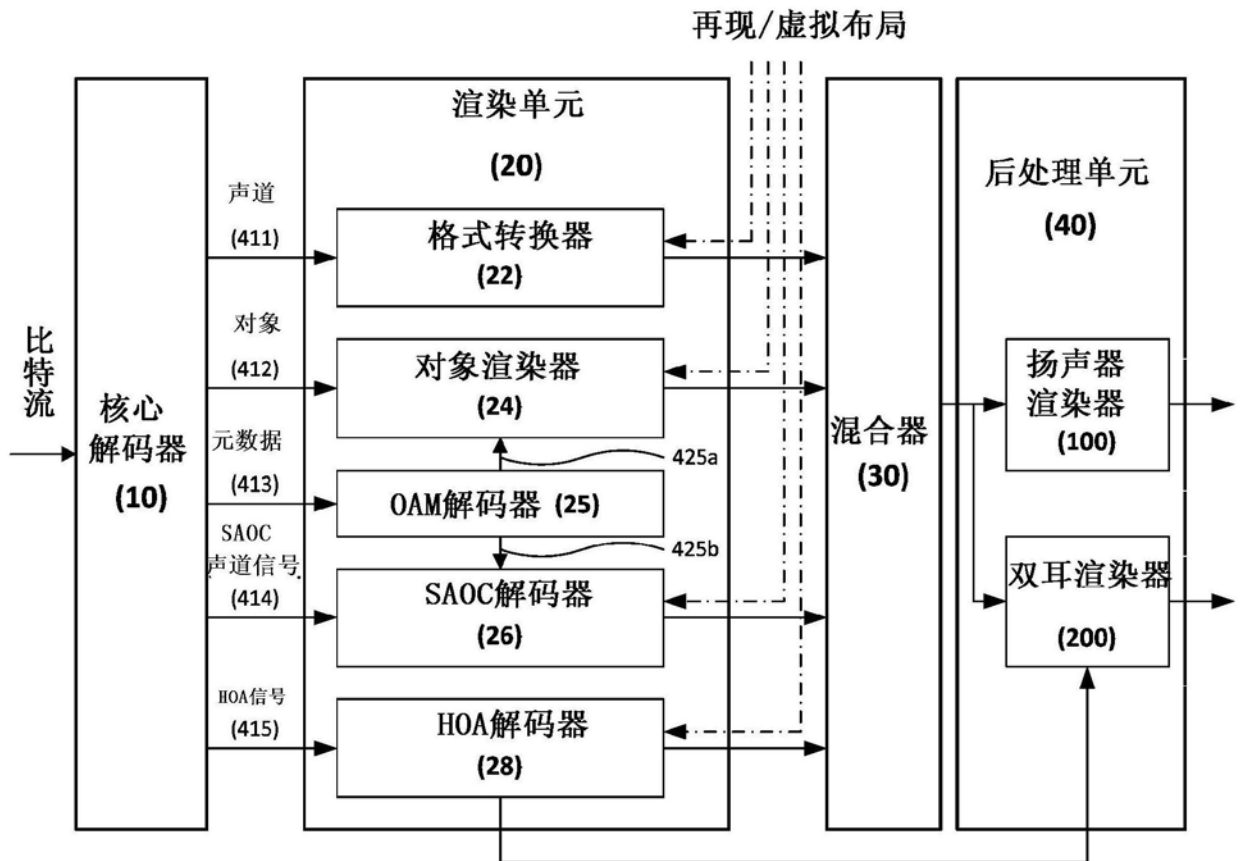


图4

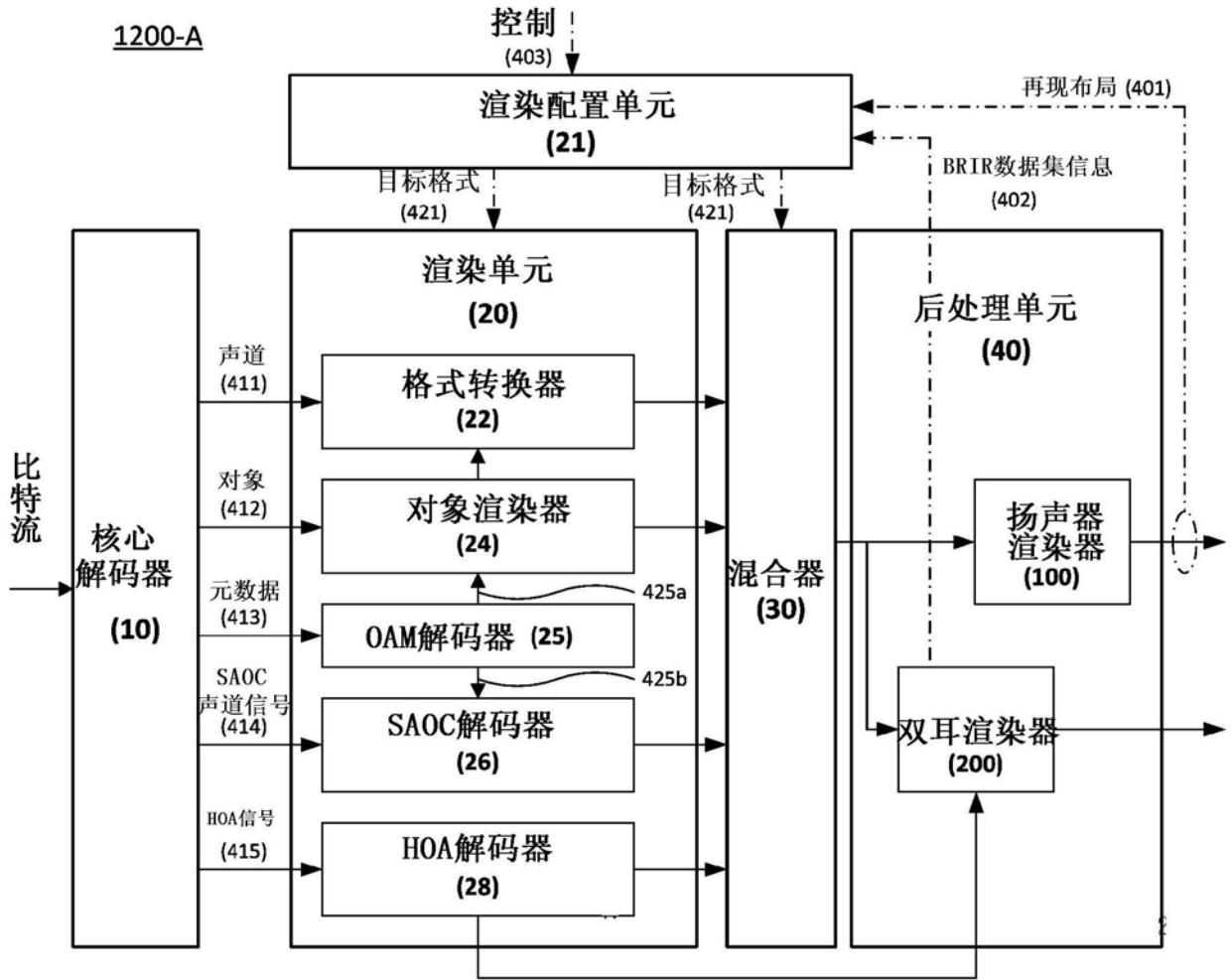


图5

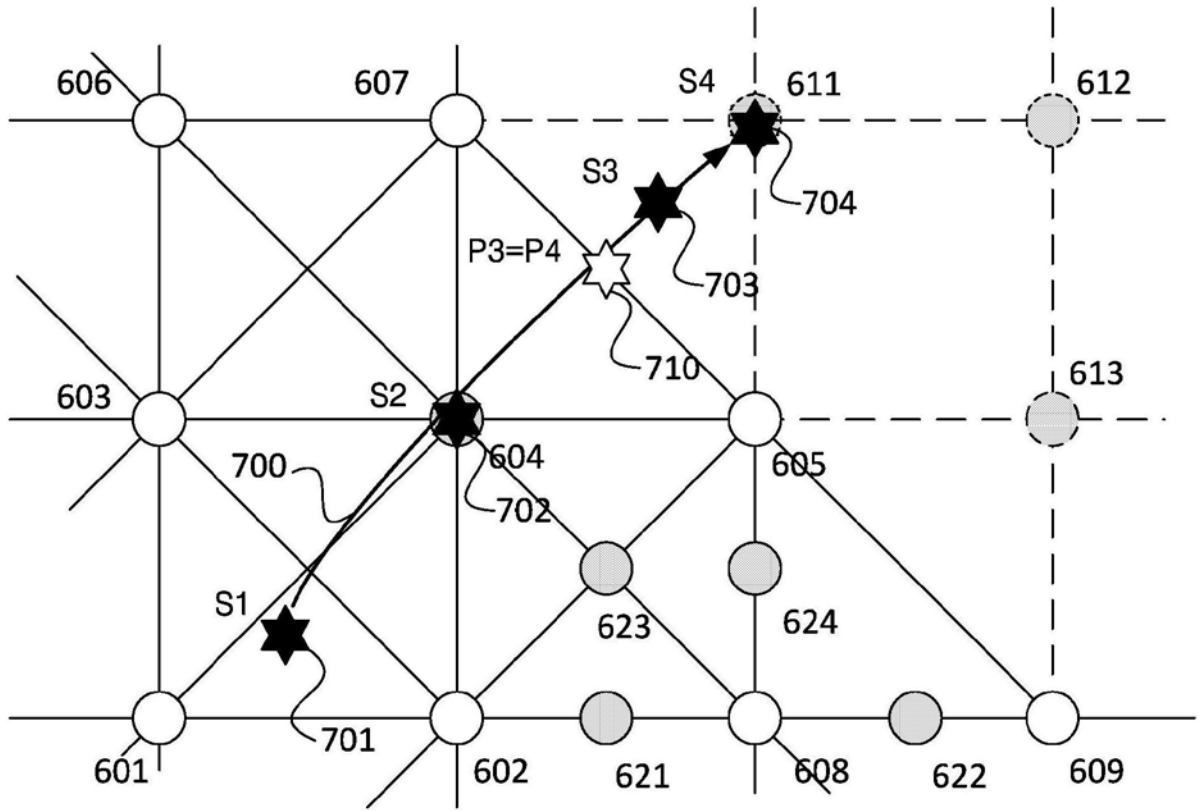


图6

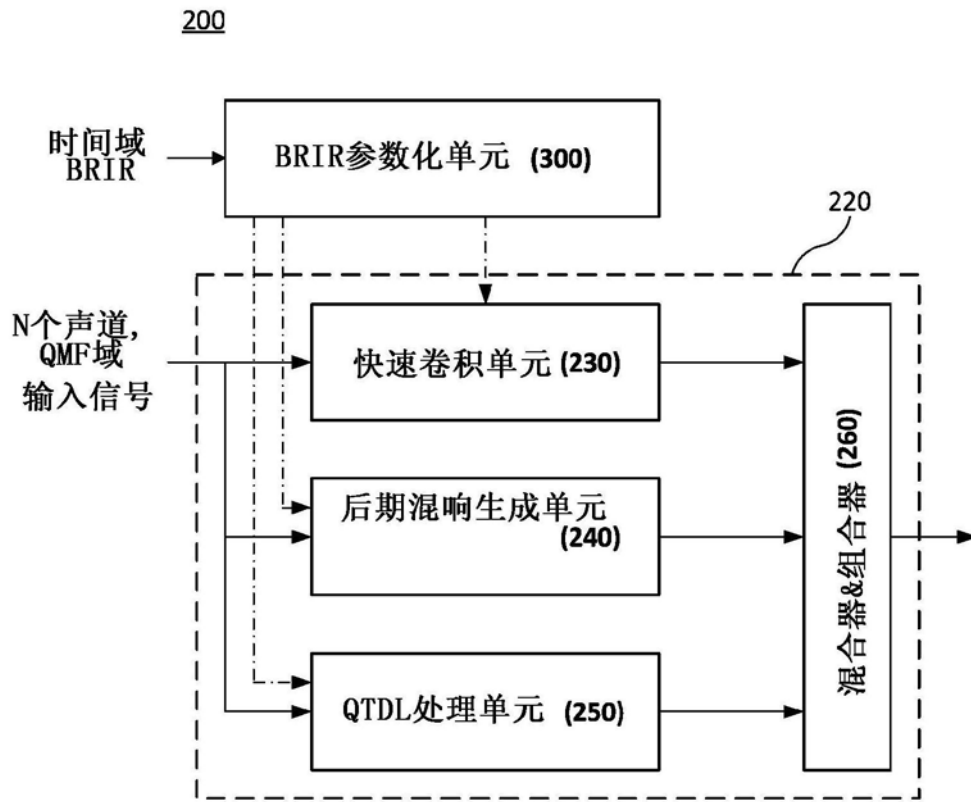
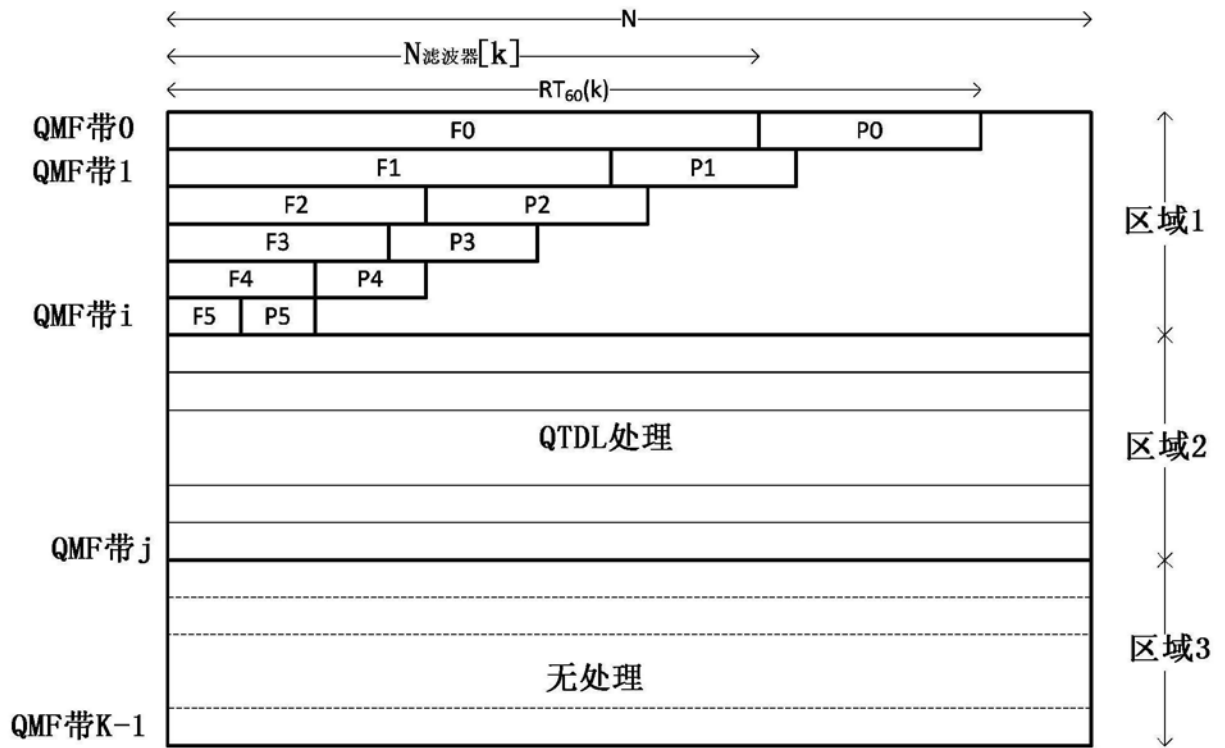


图7



250

图8

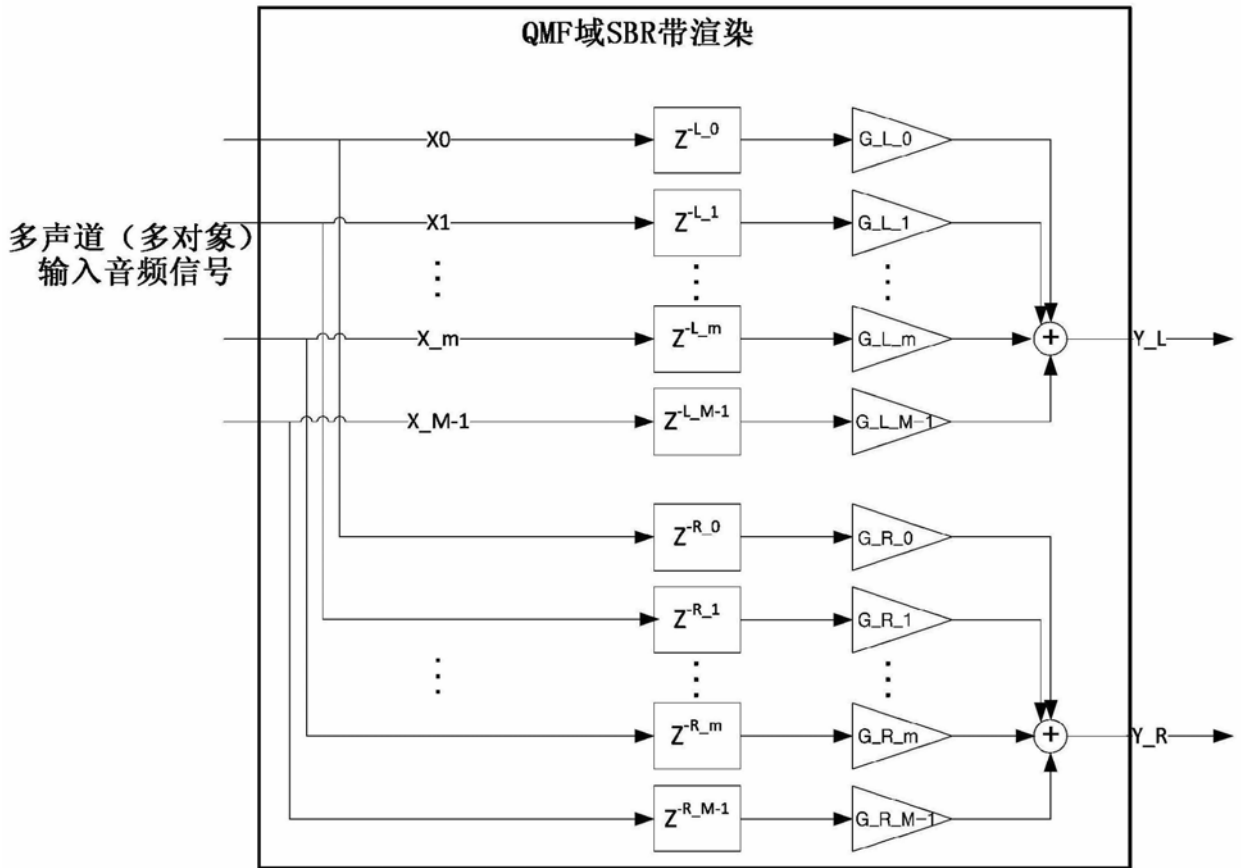


图9

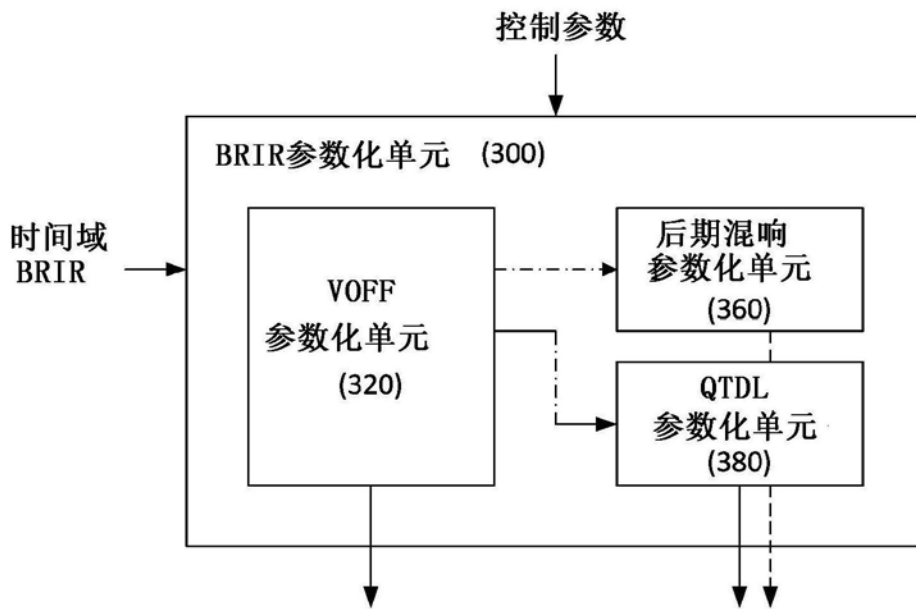


图10

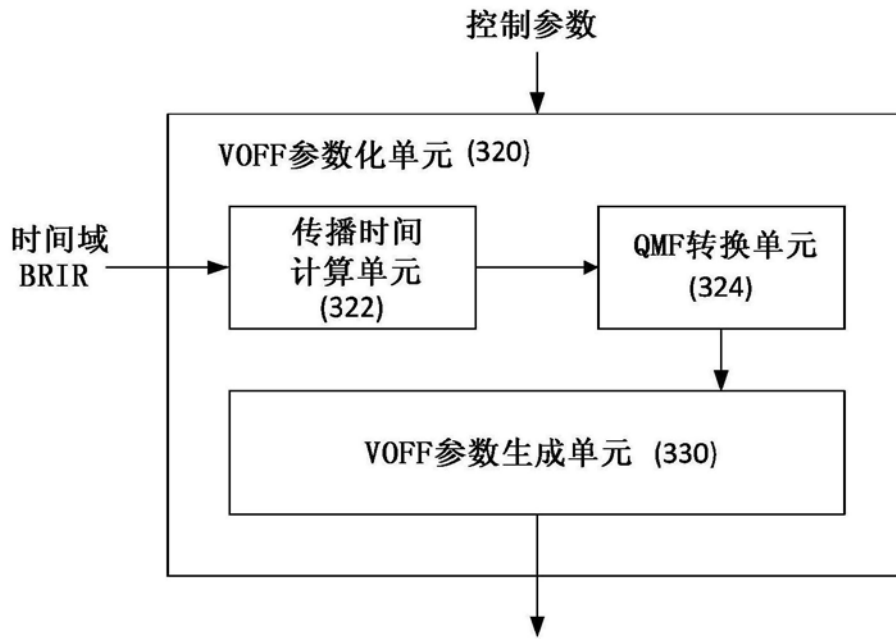


图11

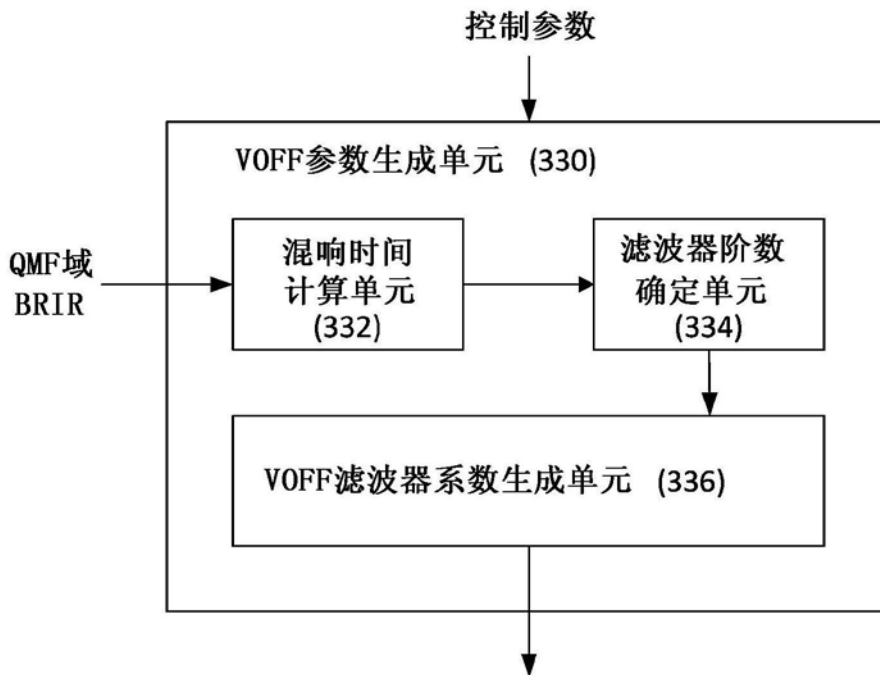


图12

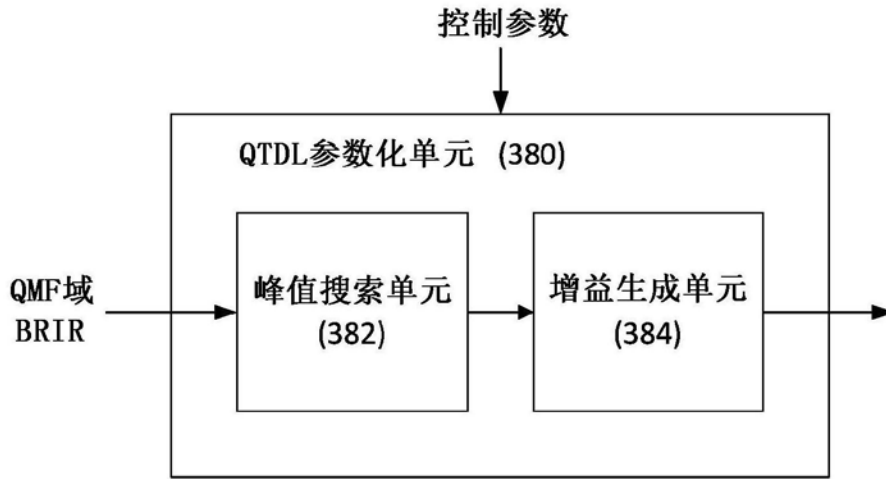


图13

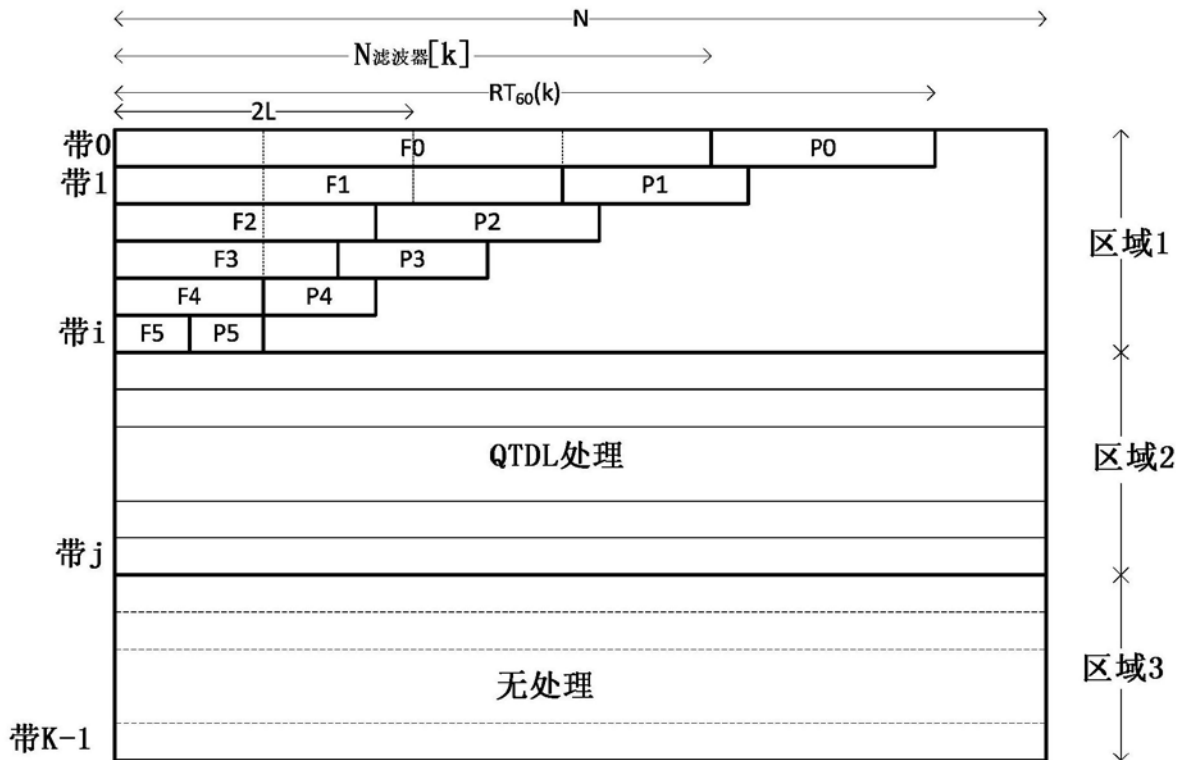


图14