



(12) 发明专利申请

(10) 申请公布号 CN 120077428 A

(43) 申请公布日 2025. 05. 30

(21) 申请号 202380071403.2

(22) 申请日 2023.10.06

(30) 优先权数据

2023-064442 2023.04.11 JP

2023-110687 2023.07.05 JP

63/417,410 2022.10.19 US

63/436,182 2022.12.30 US

(85) PCT国际申请进入国家阶段日

2025.04.07

(86) PCT国际申请的申请数据

PCT/JP2023/036497 2023.10.06

(87) PCT国际申请的公布数据

W02024/084999 JA 2024.04.25

(71) 申请人 松下电器(美国)知识产权公司

地址 美国加利福尼亚

(72) 发明人 宫阪修二 中桥康太 石川智一
宇佐见阳 江原宏幸 横本成悟
山田摩里子

(74) 专利代理机构 永新专利商标代理有限公司
72002

专利代理师 安香子

(51) Int.Cl.

G10K 15/08 (2006.01)

G10L 19/00 (2013.01)

H04S 7/00 (2006.01)

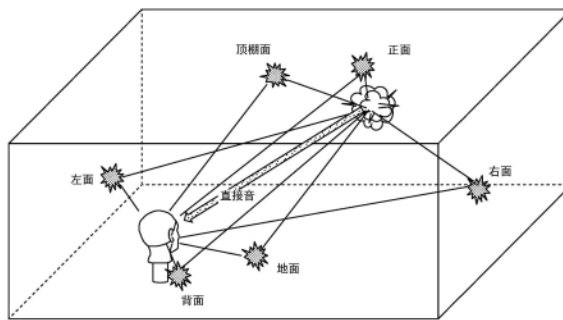
权利要求书2页 说明书48页 附图22页

(54) 发明名称

音响处理装置及音响处理方法

(57) 摘要

音响处理装置(1001)具备电路(1402)和存储器(1404);电路(1402)使用存储器(1404)取得与声音空间有关的声音空间信息;基于声音空间信息,取得与在声音空间中从音源产生的第1声音有关的特性;基于与第1声音有关的特性,控制是否选择在声音空间中对应于第1声音产生的第2声音。



1. 一种音响处理装置,其中,
具备电路和存储器,
所述电路使用所述存储器,
取得与声音空间有关的声音空间信息,
基于所述声音空间信息,取得与在所述声音空间中从音源产生的第1声音有关的特性,
基于与所述第1声音有关的特性,控制是否选择在所述声音空间中对应于所述第1声音产生的第2声音。
2. 如权利要求1所述的音响处理装置,其中,
与所述第1声音有关的特性是以时间序列表示分别由所述第1声音的振幅值为代表振幅值的持续时间和该持续时间中的所述代表振幅值的组构成的多个组的特性。
3. 如权利要求2所述的音响处理装置,其中,
所述代表振幅值是所述第1声音的音量相对于预先设定的基准音量的比值。
4. 如权利要求1所述的音响处理装置,其中,
与所述第1声音有关的特性是表示频率特性的变动量比预先设定的阈值低的状态持续的持续时间的特性。
5. 如权利要求1所述的音响处理装置,其中,
与所述第1声音有关的特性是以时间序列表示分别由频率特性的变动量比预先设定的阈值低的状态持续的持续时间和该持续时间中的频率特性的组构成的多个组的特性。
6. 如权利要求1~5中任一项所述的音响处理装置,其中,
所述电路,
取得表示与是否能听到声音的边界对应的音量的阈值,
基于与所述第1声音有关的特性,计算所述第2声音的音量,
在所述第2声音的音量大于所述阈值的情况下,选择所述第2声音。
7. 如权利要求1所述的音响处理装置,其中,
所述声音空间信息是包括所述声音空间中的所述音源的信息、以及所述声音空间中的收听者的位置的信息的场景信息,
所述第2声音是在所述声音空间中对应于所述第1声音产生的多个第2声音各自,
所述电路,
取得所述第1声音的信号,
基于所述场景信息和所述第1声音的信号,计算所述多个第2声音,
从所述音源的信息取得与所述第1声音有关的特性,
基于与所述第1声音有关的特性,控制是否将所述多个第2声音分别选择为被应用双耳处理的声音,由此从所述第1声音及所述多个第2声音中选择被应用所述双耳处理的一个以上的处理对象声音,
所述场景信息基于输入信息来更新,
与所述第1声音有关的特性是根据所述场景信息的更新来取得,
所述场景信息的更新以比对所述一个以上的处理对象声音应用所述双耳处理的频度低的频度实施。
8. 一种音响处理方法,其中,包括:

取得与声音空间有关的声音空间信息的步骤;

基于所述声音空间信息,取得与在所述声音空间中从音源产生的第1声音有关的特性的步骤;以及

基于与所述第1声音有关的特性,控制是否选择在所述声音空间中对应于所述第1声音产生的第2声音的步骤。

9.一种程序,用来使计算机执行权利要求8所述的音响处理方法。

音响处理装置及音响处理方法

技术领域

[0001] 本公开涉及音响处理装置等。

背景技术

[0002] 近年来,利用了包含VR(Virtual Reality:虚拟现实)、AR(Augmented Reality:增强现实)以及MR(Mixed Reality:混合现实)在内的ER(Extended Reality:扩展现实)(也可以表现为XR)的商品以及服务正在普及。随之,在虚拟空间或实空间中,对虚拟音源所发出的声音赋予对应于该空间的环境产生的音响效果而向收听者提供沉浸式音频(Immersive Audio)的音响处理技术的重要性提高。

[0003] 另外,收听者也可以表现为听者(listener)或用户。此外,在专利文献1、专利文献2、专利文献3及非专利文献1中示出了与本公开的音响处理装置及音响处理方法相关联的技术。

[0004] 现有技术文献

[0005] 专利文献

[0006] 专利文献1:日本特许第6288100号公报

[0007] 专利文献2:日本特开2019—22049号公报

[0008] 专利文献3:国际公开第2021/180938号

[0009] 非专利文献

[0010] 非专利文献1:B.C.J.Moore著,《听觉心理学概略》,诚信书房,1994/4/20,第6章:空间感知,p.225

发明内容

[0011] 发明要解决的课题

[0012] 例如,在专利文献1中,公开了一种对对象音频信号实施信号处理并提示给收听者的技术。随着ER技术的普及及使用ER技术的服务的多样化,例如要求由各服务要求的音响品质、使用的终端的信号处理能力、以及与声音提示设备能够提供的音质等的差异相应的音响处理。此外,为了提供这些而要求音响处理技术的进一步的改良。

[0013] 这里,音响处理技术的改良是指对既有的音响处理的变更。例如,音响处理技术的改良提供赋予新的音响效果的处理、音响处理的处理量的削减、通过音响处理得到的声音的品质的提高、音响处理的实施中使用的信息的数据量的削减、或音响处理的实施中使用的信息的取得或生成的容易化等。或者,音响处理技术的改良也可以提供这些中的任意两个以上的组合。

[0014] 特别是,在收听者能够在虚拟空间内自由地移动的设备或服务中要求这些改良。但是,通过音响处理技术的改良得到的上述效果只不过是例子。基于本公开掌握的一个以上的技术方案也可以是基于与上述不同的观点想到的技术方案、达成与上述不同的目的的技术方案或能得到与上述不同的效果的技术方案。

[0015] 用来解决课题的手段

[0016] 基于本公开掌握的有一技术方案音响装置具备电路和存储器;所述电路使用所述存储器,取得与声音空间有关的声音空间信息;基于所述声音空间信息,取得与在所述声音空间中从音源产生的第1声音有关的特性;基于与所述第1声音有关的特性,控制是否选择在所述声音空间中对应于所述第1声音产生的第2声音。

[0017] 另外,这些包含性或具体的技术方案也可以由系统、装置、方法、集成电路、计算机程序或计算机可读的CD-ROM等的非暂时性的记录介质实现,也可以由这些任意的组合实现。

[0018] 发明效果

[0019] 本公开的一技术方案例如能够提供赋予新的音响效果的处理、音响处理的处理量的削减、通过音响处理得到的声音的音质的提高、音响处理的实施中使用的信息的数据量的削减、或者音响处理的实施中使用的信息的取得或生成的容易化等。或者,本公开的一技术方案能够提供这些中的任意的组合。结果,本公开的一技术方案提供适合于收听者的利用环境的音响处理,能够贡献于收听者的音响体验的提高。

[0020] 特别是,在容许收听者在虚拟空间内自由移动的设备或服务中能得到上述的效果。但是,上述的效果只不过是基于本公开掌握的各种技术方案的效果的一例。基于本公开掌握的一个以上的技术方案分别也可以是基于与上述不同的观点想到的技术方案、达成与上述不同的目的的技术方案、或者能得到与上述不同的效果的技术方案。

附图说明

[0021] 图1是表示在声音空间中生成的直接音及反射音的一例的图。

[0022] 图2是表示实施方式的立体音响再现系统的一例的图。

[0023] 图3A是表示实施方式的编码装置的构成例的框图。

[0024] 图3B是表示实施方式的解码装置的构成例的框图。

[0025] 图3C是表示实施方式的编码装置的另一构成例的框图。

[0026] 图3D是表示实施方式的解码装置的另一构成例的框图。

[0027] 图4A是表示实施方式的解码器的构成例的框图。

[0028] 图4B是表示实施方式的解码器的另一构成例的框图。

[0029] 图5是表示实施方式的语音信号处理装置的物理构成的一例的图。

[0030] 图6是表示实施方式的编码装置的物理构成的一例的图。

[0031] 图7是表示实施方式的渲染部的构成例的框图。

[0032] 图8是表示实施方式的语音信号处理装置的动作例的流程图。

[0033] 图9是表示收听者和障碍物对象比较远的位置关系的图。

[0034] 图10是表示收听者和障碍物对象比较近的位置关系的图。

[0035] 图11是表示直接音与反射音的时间差和阈值的关系的图。

[0036] 图12A是表示阈值数据的设定方法的例子的一部分的图。

[0037] 图12B是表示阈值数据的设定方法的例子的一部分的图。

[0038] 图12C是表示阈值数据的设定方法的例子的一部分的图。

[0039] 图13是表示阈值的设定方法的例子的图。

- [0040] 图14是表示选择处理的一例的流程图。
- [0041] 图15是表示直接音的方向、反射音的方向、时间差和阈值的关系的图。
- [0042] 图16是表示角度差、时间差和阈值的关系的图。
- [0043] 图17是表示渲染部的另一构成例的框图。
- [0044] 图18是表示选择处理的另一例的流程图。
- [0045] 图19是表示选择处理的再另一例的流程图。
- [0046] 图20是表示实施方式的声音信号处理装置的动作的第1变形例的流程图。
- [0047] 图21是表示实施方式的声音信号处理装置的动作的第2变形例的流程图。
- [0048] 图22是表示化身、音源对象和障碍物对象的配置例的图。
- [0049] 图23是表示选择处理的再另一例的流程图。
- [0050] 图24是表示用于渲染部进行管线处理的构成例的框图。
- [0051] 图25是表示声音的透射及衍射的图。

具体实施方式

[0052] (作为本公开的基础的认识)

[0053] 图1是表示在声音空间中生成的直接音及反射音的一例的图。在将虚拟空间的特征用声音表现的音响处理中,为了表现空间的宽广度及壁面的材质等,以及为了正确地掌握音源的位置(声像的定位),有效的是不仅再现直接音,还再现反射音。

[0054] 例如,在如图1的长方体的室内听取声音的情况下,对于一个音源产生与6个面的壁面对应的6条一次反射音。这些反射音的再现成为与空间及声像有关的适当的理解的线索。进而,对于各反射音,在产生了该反射音的反射面以外的面产生二次反射音。这些反射音也成为在感知上有效的线索。

[0055] 但是,即便是仅考虑到二次反射为止的情况,对于一个音源也会产生1条直接音和36条($6+6\times 5$)反射音,所以是产生37条声线,要处理这些声线需要相当程度的运算量。

[0056] 此外,在如虚拟会议、虚拟购物或虚拟音乐会等的关于元宇宙(Metaverse)设想的近年来的应用商品中,必然会存在多个音源,所以需要更为庞大的运算量。

[0057] 此外,在虚拟空间内收听声音的收听者使用头戴式耳机或VR护目镜。为了对这样的收听者提供立体音响,对各声线进行对两耳间赋予声压比及相位差而再现声音的到来方向及远近感的双耳(Binaural)处理。因而,如果要再现所产生的全部的反射音,则其运算量是非常大的。

[0058] 另一方面,作为由在虚拟空间中体验的收听者佩戴的VR护目镜的电池,有时因其方便性而使用小型蓄电池。为了延长其电池寿命,如上所述的处理所需的运算负荷最好较小。为此,希望将以几百条的规模产生的声线的数量在不损害声音的定位及空间的掌握的范围进行削减。

[0059] 此外,在再现音响的系统中,有时对于收听者的位置及朝向容许6DoF(6Degrees of Freedom(自由度))等的自由度。在此情况下,收听者、音源和反射声音的物体的位置关系如果不是再现时(渲染时)就不能确定。因此,反射音也如果不是再现时就不能确定。因此,难以预先决定处理对象的反射音。

[0060] 因而,在再现时在声音空间中产生的多个反射音中适当地选择处理对象或非处理

对象的一个以上的反射音有利于运算量及运算负荷的适当削减。

[0061] 所以,本公开的目的是提供能够适当地控制是否选择在声音空间中产生的声音的音响处理装置等。

[0062] 另外,控制是否选择声音,对应于判定是否选择声音。此外,选择声音既可以是选择声音作为处理对象声音,也可以是选择声音作为非处理对象声音。

[0063] (公开的概要)

[0064] 基于本公开掌握的有关第1技术方案的音响处理装置具备电路和存储器;电路使用存储器,取得与声音空间有关的声音空间信息;基于声音空间信息,取得与在声音空间中从音源产生的第1声音有关的特性;基于与第1声音有关的特性,控制是否选择在声音空间中对应于第1声音产生的第2声音。

[0065] 上述技术方案的装置能够基于与在声音空间中产生的第1声音有关的特性,适当地控制是否选择在声音空间中对应于第1声音产生的第2声音。即,能够适当地控制是否选择在声音空间中产生的声音。因而,能够适当地削减运算量及运算负荷。

[0066] 基于本公开掌握的有关第2技术方案的音响处理装置也可以是,在第1技术方案的音响处理装置中,第1声音是直接音,第2声音是反射音。

[0067] 上述技术方案的装置能够基于与直接音有关的特性,适当地控制是否选择反射音。

[0068] 基于本公开掌握的有关第3技术方案的音响处理装置也可以是,在第2技术方案的音响处理装置中,与第1声音有关的特性是直接音的音量与反射音的音量的音量比;电路基于声音空间信息,计算音量比,基于音量比,控制是否选择反射音。

[0069] 上述技术方案的装置能够基于直接音的音量与反射音的音量的音量比,适当地选择对收听者的感知的影响程度大的反射音。

[0070] 基于本公开掌握的有关第4技术方案的音响处理装置也可以是,在第3技术方案的音响处理装置中,在选择了反射音的情况下,电路通过对反射音及直接音应用双耳处理,生成分别到达收听者的两耳的声音。

[0071] 上述技术方案的装置能够适当地选择对收听者的感知的影响程度大的反射音,对所选择的反射音应用双耳处理。

[0072] 基于本公开掌握的有关第5技术方案的音响处理装置也可以是,在第3或第4技术方案的音响处理装置中,电路基于声音空间信息,计算直接音的结束时刻与反射音的到来时刻的时间差,基于时间差及音量比,控制是否选择反射音。

[0073] 上述技术方案的装置能够基于直接音的结束时刻与反射音的到来时刻的时间差、以及直接音的音量与反射音的音量的音量比,更适当地选择对收听者的感知的影响程度大的反射音。因而,上述技术方案的装置能够基于后掩蔽效应,更适当地选择对收听者的感知的影响程度大的反射音。

[0074] 基于本公开掌握的有关第6技术方案的音响处理装置也可以是,在第5技术方案的音响处理装置中,电路在音量比是阈值以上的情况下,选择反射音;在时间差是第1值的情况下被用作阈值的第1阈值大于在时间差是比第1值大的第2值的情况下被用作阈值的第2阈值。

[0075] 上述技术方案的装置能够提高直接音的结束时刻与反射音的到来时刻的时间差

大的反射音被选择的可能性。因而,上述技术方案的装置能够更适当地选择对收听者的感知的影响程度大的反射音。

[0076] 基于本公开掌握的有关第7技术方案的音响处理装置也可以是,在第3或第4技术方案的音响处理装置中,电路基于声音空间信息,计算直接音的到来时刻与反射音的到来时刻的时间差;基于时间差及音量比,控制是否选择反射音。

[0077] 上述技术方案的装置能够基于直接音的到来时刻与反射音的到来时刻的时间差、以及直接音的音量与反射音的音量的音量比,更适当地选择对收听者的感知的影响程度大的反射音。因而,上述技术方案的装置能够基于优先效应,更适当地选择对收听者的感知的影响程度大的反射音。

[0078] 基于本公开掌握的有关第8技术方案的音响处理装置也可以是,在第7技术方案的音响处理装置中,电路在音量比是阈值以上的情况下,选择反射音;在时间差是第1值的情况下被用作阈值的第1阈值大于在时间差是比第1值大的第2值的情况下被用作阈值的第2阈值。

[0079] 上述技术方案的装置能够提高直接音的到来时刻与反射音的到来时刻的时间差大的反射音被选择的可能性。因而,上述技术方案的装置能够适当地选择对收听者的感知的影响程度大的反射音。

[0080] 基于本公开掌握的有关第9技术方案的音响处理装置也可以是,在第6或第8技术方案的音响处理装置中,电路基于直接音的到来方向和反射音的到来方向,调整阈值。

[0081] 上述技术方案的装置能够基于直接音的到来方向和反射音的到来方向,适当地选择对收听者的感知的影响程度大的反射音。

[0082] 基于本公开掌握的有关第10技术方案的音响处理装置也可以是,在第1~第9技术方案的任一个的音响处理装置中,在没有选择第2声音的情况下,电路基于第2声音的音量将第1声音的音量进行修正。

[0083] 上述技术方案的装置能够以较少的运算量适当地减少因第2声音没有被选择而第2声音的音量欠缺所产生的违和感。

[0084] 基于本公开掌握的有关第11技术方案的音响处理装置也可以是,在第1~第9技术方案的任一个的音响处理装置中,在没有选择第2声音的情况下,电路将第2声音合成到第1声音中。

[0085] 上述技术方案的装置能够将第2声音的特性更正确地反映到第1声音中。因而,上述技术方案的装置能够减少因第2声音没有被选择而第2声音的音量欠缺所产生的违和感。

[0086] 基于本公开掌握的有关第12技术方案的音响处理装置也可以是,在第3~第9技术方案的任一个的音响处理装置中,音量比是第1时刻的直接音的音量与不同于第1时刻的第2时刻的反射音的音量的音量比。

[0087] 上述技术方案的装置在直接音被感知的时刻与反射音被感知的时刻不同的情况下,能够基于不同时刻的直接音与反射音的音量比,适当地选择对收听者的感知的影响程度大的反射音。

[0088] 基于本公开掌握的有关第13技术方案的音响处理装置也可以是,在第1或第2技术方案的音响处理装置中,电路基于与第1声音有关的特性设定阈值,基于阈值控制是否选择第2声音。

[0089] 上述技术方案的装置能够基于根据与第1声音有关的特性设定的阈值,适当地控制是否选择第2声音。

[0090] 基于本公开掌握的有关第14技术方案的音响处理装置也可以是,在第1、第2及第13技术方案的任一个的音响处理装置中,与第1声音有关的特性是音源的音量、音源的视觉性及音源的定位性中的某一个或某两个以上的组合。

[0091] 上述技术方案的装置能够基于音源的音量、音源的视觉性或音源的定位性,适当地控制是否选择第2声音。

[0092] 基于本公开掌握的有关第15技术方案的音响处理装置也可以是,在第1、第2及第13技术方案的任一个的音响处理装置中,与第1声音有关的特性是第1声音的频率特性。

[0093] 上述技术方案的装置能够基于第1声音的频率特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0094] 基于本公开掌握的有关第16技术方案的音响处理装置也可以是,在第1、第2及第13技术方案的任一个的音响处理装置中,与第1声音有关的特性是表示第1声音的振幅的断续性的特性。

[0095] 上述技术方案的装置能够基于表示第1声音的振幅的断续性的特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0096] 基于本公开掌握的有关第17技术方案的音响处理装置也可以是,在第1、第2、第13及第16技术方案的任一个的音响处理装置中,与第1声音有关的特性是表示第1声音的有声部的持续时间或第1声音的无声部的持续时间的特性。

[0097] 上述技术方案的装置能够基于表示第1声音的有声部的持续时间或第1声音的无声部的持续时间的特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0098] 基于本公开掌握的有关第18技术方案的音响处理装置也可以是,在第1、第2、第13、第16及第17技术方案的任一个的音响处理装置中,与第1声音有关的特性是将第1声音的有声部的持续时间和第1声音的无声部的持续时间以时间序列表示的特性。

[0099] 上述技术方案的装置能够基于将第1声音的有声部的持续时间和第1声音的无声部的持续时间以时间序列表示的特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0100] 基于本公开掌握的有关第19技术方案的音响处理装置也可以是,在第1、第2、第13及第15技术方案的任一个的音响处理装置中,与第1声音有关的特性是表示第1声音的频率特性的变动的特性。

[0101] 上述技术方案的装置能够基于表示第1声音的频率特性的变动的特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0102] 基于本公开掌握的有关第20技术方案的音响处理装置也可以是,在第1、第2、第13、第15及第19技术方案的任一个的音响处理装置中,与第1声音有关的特性是表示第1声音的频率特性的平稳性的特性。

[0103] 上述技术方案的装置能够基于表示第1声音的频率特性的平稳性的特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0104] 基于本公开掌握的有关第21技术方案的音响处理装置也可以是,在第1、第2及第13~第20技术方案的任一个的音响处理装置中,从比特流取得与第1声音有关的特性。

[0105] 上述技术方案的装置能够基于从比特流取得的信息,适当地控制是否选择对应于第1声音产生的第2声音。

[0106] 基于本公开掌握的有关第22技术方案的音响处理装置也可以是,在第1、第2及第13~第21技术方案的任一个的音响处理装置中,电路计算与第2声音有关的特性;基于与第1声音有关的特性以及与第2声音有关的特性,控制是否选择第2声音。

[0107] 上述技术方案的装置能够基于与第1声音有关的特性以及与第2声音有关的特性,适当地控制是否选择对应于第1声音产生的第2声音。

[0108] 基于本公开掌握的有关第23技术方案的音响处理装置也可以是,在第22技术方案的音响处理装置中,电路取得表示与是否能听到声音的边界对应的音量的阈值;基于与第1声音有关的特性、与第2声音有关的特性和阈值,控制是否选择第2声音。

[0109] 上述技术方案的装置能够除了与第1声音有关的特性以及与第2声音有关的特性以外,还基于与是否能听到对应的阈值,适当地控制是否选择第2声音。

[0110] 基于本公开掌握的有关第24技术方案的音响处理装置也可以是,在第22或第23技术方案的音响处理装置中,与第2声音有关的特性是第2声音的音量。

[0111] 上述技术方案的装置能够基于第2声音的音量,适当地控制是否选择第2声音。

[0112] 基于本公开掌握的有关第25技术方案的音响处理装置也可以是,在第1~第24技术方案的任一个的音响处理装置中,声音空间信息包含声音空间中的收听者的位置的信息;第2声音是在声音空间中对应于第1声音产生的多个第2声音各自;电路基于与第1声音有关的特性来控制是否选择多个第2声音各自,从第1声音及多个第2声音中选择被应用双耳处理的一个以上的处理对象声音。

[0113] 上述技术方案的装置能够基于与在声音空间中产生的第1声音有关的特性,适当地控制是否选择在声音空间中对应于第1声音产生的多个第2声音各自。并且,上述技术方案的装置能够从第1声音及多个第2声音中适当地选择被应用双耳处理的一个以上的处理对象声音。

[0114] 基于本公开掌握的有关第26技术方案的音响处理装置也可以是,在第1~第25技术方案的任一个的音响处理装置中,取得与第1声音有关的特性的定时是声音空间的制作时、声音空间的处理开始时、以及声音空间的处理中的信息更新线程产生时的至少一个。

[0115] 上述技术方案的装置能够基于在自适应的定时取得的信息,适当地选择被应用双耳处理的一个以上的处理对象声音。

[0116] 基于本公开掌握的有关第27技术方案的音响处理装置也可以是,在第1~第26技术方案的任一个的音响处理装置中,在声音空间的处理开始后定期地取得与第1声音有关的特性。

[0117] 上述技术方案的装置能够基于定期地取得的信息,适当地选择被应用双耳处理的一个以上的处理对象声音。

[0118] 基于本公开掌握的有关第28技术方案的音响处理装置也可以是,在第1、第2及第25~第27技术方案的任一个的音响处理装置中,与第1声音有关的特性是第1声音的音量;电路基于第1声音的音量,计算第2声音的评价值,基于评价值,控制是否选择第2声音。

[0119] 上述技术方案的装置能够基于根据第1声音的音量对第2声音计算的评价值,适当地控制是否选择第2声音。

[0120] 基于本公开掌握的有关第29技术方案的音响处理装置也可以是,在第28技术方案的音响处理装置中,第1声音的音量具有转变。

[0121] 上述技术方案的装置能够基于根据具有转变的音量计算的评价值,适当地控制是否选择第2声音。

[0122] 基于本公开掌握的有关第30技术方案的音响处理装置也可以是,在第28或第29技术方案的音响处理装置中,电路计算评价值,以使得第1声音的音量越大则第2声音越容易被选择。

[0123] 上述技术方案的装置能够基于被设定为第1声音的音量越大则第2声音越容易被选择的值的评价值,适当地控制是否选择第2声音。

[0124] 基于本公开掌握的有关第31技术方案的音响处理装置也可以是,在第1~第30技术方案的任一个的音响处理装置中,声音空间信息是包括声音空间中的音源的信息以及声音空间中的收听者的位置的信息的场景信息;第2声音是在声音空间中对应于第1声音产生的多个第2声音各自;电路取得第1声音的信号;基于场景信息和第1声音的信号,计算多个第2声音;从音源的信息取得与第1声音有关的特性;基于与第1声音有关的特性来控制是否将多个第2声音各自选择为不被应用双耳处理的声音,由此从多个第2声音中选择不被应用双耳处理的一个以上的第2声音。

[0125] 上述技术方案的装置能够基于与第1声音有关的特性,从在声音空间中对应于第1声音产生的多个第2声音中适当地选择不被应用双耳处理的一个以上的第2声音。

[0126] 基于本公开掌握的有关第32技术方案的音响处理装置也可以是,在第31技术方案的音响处理装置中,场景信息基于输入信息来更新;对应于场景信息的更新,取得与第1声音有关的特性。

[0127] 上述技术方案的装置能够基于对应于场景信息的更新而取得的信息,适当地选择不被应用双耳处理的一个以上的第2声音。

[0128] 基于本公开掌握的有关第33技术方案的音响处理装置也可以是,在第31或第32技术方案的音响处理装置中,从比特流所包含的元数据取得场景信息及与第1声音有关的特性。

[0129] 上述技术方案的装置能够基于从比特流所包含的元数据取得的信息,适当地选择不被应用双耳处理的一个以上的第2声音。

[0130] 基于本公开掌握的第34技术方案的音响处理装置也可以是,在第1、第2、第13、第16~第18、第25~第27及第31~第33技术方案的任一个的音响处理装置中,与第1声音有关的特性是以时间序列表示分别由第1声音的振幅值为代表振幅值的持续时间和持续时间中的代表振幅值的组构成的多个组的特性。

[0131] 上述技术方案的装置能够基于持续时间和代表振幅值的时间序列的信息,适当地控制是否选择对应于第1声音产生的第2声音。

[0132] 基于本公开掌握的第35技术方案的音响处理装置也可以是,在第34技术方案的音响处理装置中,代表振幅值是第1声音的音量相对于预先设定的基准音量的比值。

[0133] 上述技术方案的装置能够基于与相对于基准音量的比对应的代表振幅值,适当地控制是否选择对应于第1声音产生的第2声音。

[0134] 基于本公开掌握的第36技术方案的音响处理装置也可以是,在第1、第2、第13、第

15、第19及第20技术方案中任一个的音响处理装置中,与第1声音有关的特性是表示频率特性的变动量比预先设定的阈值低的状态持续的持续时间的特性。

[0135] 上述技术方案的装置能够基于频率特性的变动量比预先决定的阈值低的状态持续的持续时间,适当地控制是否选择对应于第1声音产生的第2声音。

[0136] 基于本公开掌握的第37技术方案的音响处理装置也可以是,在第1、第2、第13、第15、第19、第20及第36技术方案的任一个的音响处理装置中,与第1声音有关的特性是以时间序列表示分别由频率特性的变动量比预先设定的阈值低的状态持续的持续时间和持续时间中的频率特性的组构成的多个组的特性。

[0137] 上述技术方案的装置能够基于频率特性的变动量比预先设定的阈值低的状态持续的持续时间和频率特性的时间序列,适当地控制是否选择对应于第1声音产生的第2声音。

[0138] 基于本公开掌握的第38技术方案的音响处理装置也可以是,在第1、第2、第13~第24及第34~第37技术方案的任一个的音响处理装置中,电路取得表示与是否能听到声音的边界对应的音量的阈值,基于与第1声音有关的特性,计算第2声音的音量,在第2声音的音量大于阈值的情况下,选择第2声音。

[0139] 上述技术方案的装置在第2声音的音量大于与是否能听到对应的阈值的情况下,能够适当地选择第2声音。

[0140] 基于本公开掌握的第39技术方案的音响处理装置也可以是,在第1、第2、第13~第20及第31~第38技术方案的任一个的音响处理装置中,声音空间信息是包含声音空间中的音源的信息、以及声音空间中的收听者的位置的信息的场景信息,第2声音是在声音空间中对应于第1声音产生的多个第2声音各自,电路取得第1声音的信号,基于场景信息和第1声音的信号,计算多个第2声音,从音源的信息取得与第1声音有关的特性,基于与第1声音有关的特性,控制是否将多个第2声音分别选择为应用双耳处理的声音,由此从第1声音及多个第2声音中选择被应用双耳处理的一个以上的处理对象声音,场景信息基于输入信息来更新,与第1声音有关的特性是根据场景信息的更新来取得,场景信息的更新以比对一个以上的处理对象声音应用双耳处理的频度低的频度实施。

[0141] 上述技术方案的装置能够基于根据以相对低的频度更新的场景信息的更新来取得的信息,适当地选择被应用双耳处理的一个以上的第2声音。

[0142] 基于本公开掌握的第40技术方案的音响处理方法,包括:取得与声音空间有关的声音空间信息的步骤;基于声音空间信息,取得与在声音空间中从音源产生的第1声音有关的特性的步骤;以及基于与第1声音有关的特性,控制是否选择在声音空间中对应于第1声音产生的第2声音的步骤。

[0143] 上述技术方案的方法能够起到与第1技术方案所述的音响处理装置同样的效果。

[0144] 基于本公开掌握的有关第41技术方案的程序,是用来使计算机执行第40技术方案的音响处理方法的程序。

[0145] 上述技术方案的程序能够使用计算机起到与第40技术方案的音响处理方法同样的效果。

[0146] 另外,这些包含性或具体的技术方案可以由系统、装置、方法、集成电路、计算机程序或计算机可读的CD-ROM等的记录介质实现,也可以由系统、装置、方法、集成电路、计

计算机程序或记录介质的任意的组合实现。

[0147] 以下,参照附图对本公开的音响处理装置、编码装置、解码装置及立体音响再现系统详细地进行说明。立体音响再现系统也能够表现为声音信号再现系统。

[0148] 另外,以下说明的实施方式都表示包含性或具体的例子。在以下的实施方式中表示的数值、形状、材料、构成要素、构成要素的配置位置及连接形态、步骤以及步骤的顺序等是一例,不是限定基于本公开掌握的技术方案的主旨。此外,关于以下的实施方式的构成要素中的例如在本公开中所记载的基本的技术方案中不包含的构成要素或在表示最上位概念的独立权利要求中没有记载的构成要素,设为任意的构成要素进行说明。

[0149] (实施方式)

[0150] (立体音响再现系统的例)

[0151] 图2是表示立体音响再现系统的一例的图。具体而言,图2表示作为能够应用本公开的音响处理或解码处理的系统的一例的立体音响再现系统1000。立体音响也被表现为沉浸式音频(Immersive Audio)。立体音响再现系统1000包括声音信号处理装置1001和声音提示装置1002。

[0152] 声音信号处理装置1001也被表现为音响处理装置,对虚拟音源所发出的声音信号实施音响处理,生成对收听者提示的音响处理后的声音信号。声音信号并不限于语音,只要是可听音即可。音响处理例如是为了再现声音在由音源产生到到达收听者为止的期间受到的一个以上的作用而对声音信号实施的信号处理。

[0153] 声音信号处理装置1001基于记述有引起上述作用的原因的空间信息实施音响处理。空间信息例如是表示音源、收听者及周围的对象(object)的位置的信息、表示空间的形状的信息、以及与声音的传播有关的参数等。声音信号处理装置1001例如是PC(Personal Computer)、智能电话、平板电脑或游戏控制台等。

[0154] 音响处理后的信号从声音提示装置1002被提示给收听者。声音提示装置1002经由无线或有线的通信与声音信号处理装置1001连接。由声音信号处理装置1001生成的音响处理后的声音信号经由无线或有线的通信被传送至声音提示装置1002。

[0155] 在声音提示装置1002例如由右耳用的设备及左耳用的设备等的多个设备构成的情况下,通过多个设备间的通信或多个设备各自与声音信号处理装置1001的通信,多个设备同步地提示声音。声音提示装置1002例如是佩戴在收听者的头部上的头戴式耳机、耳塞、头戴显示器或由固定的多个扬声器构成的环绕扬声器等。

[0156] 另外,立体音响再现系统1000也可以与视觉性地提供包括AR/VR的ER体验的图像提示装置或立体影像提示装置组合使用。例如,由空间信息处置的空间是虚拟空间,该空间中的音源、收听者及对象的位置是虚拟空间中的虚拟的音源、虚拟的收听者及虚拟的对象的虚拟的位置。也可以将该空间表现为声音空间。此外,也可以将空间信息表现为声音空间信息。

[0157] 此外,图2表示声音信号处理装置1001和声音提示装置1002是不同的装置的系统构成例,但能够应用本公开的音响处理方法或解码方法的立体音响再现系统1000并不限于图2的构成。例如,也可以是声音信号处理装置1001包含在声音提示装置1002中,声音提示装置1002进行音响处理和声音的提示双方。

[0158] 此外,也可以是声音信号处理装置1001和声音提示装置1002分担实施在本公开中

说明的音响处理。此外,也可以经由网络与声音信号处理装置1001或声音提示装置1002连接的服务器实施在本公开中说明的音响处理的一部分或整体。

[0159] 此外,声音信号处理装置1001也可以将通过将声音信号和用于音响处理的空间信息的数据的至少一部分编码而生成的比特流解码来实施音响处理。因而,也可以将声音信号处理装置1001表现为解码装置。

[0160] (编码装置的例子)

[0161] 图3A是表示编码装置的构成例的框图。具体而言,图3A表示作为本公开的编码装置的一例的编码装置1100的构成。

[0162] 输入数据1101是包含对编码器1102输入的空间信息及/或声音信号的编码对象数据。关于空间信息的详细情况在后面进行说明。

[0163] 编码器1102将输入数据1101编码,生成编码数据1103。编码数据1103例如是通过编码处理生成的比特流。

[0164] 存储器1104存储编码数据1103。存储器1104例如可以是硬盘或SSD(Solid—State Drive:固态驱动器),也可以是其他的存储器。

[0165] 另外,在上述说明中,作为存储在存储器1104中的编码数据1103的一例而列出了通过编码处理生成的比特流,但编码数据1103也可以是比特流以外的数据。例如,编码装置1100也可以在存储器1104中存储将比特流变换为规定的格式而生成后的数据。变换后的数据例如也可以是与一个以上的比特流对应的文件或复用流。

[0166] 这里,文件是具有例如ISOBMFF(ISO Base Media File Format:ISO基础媒体文件格式)等的文件格式的文件。此外,编码数据1103也可以是将上述的比特流或文件分割而生成的多个包的形式。

[0167] 例如,也可以将由编码器1102生成的比特流变换为与比特流不同的数据。在此情况下,编码装置1100具备未图示的变换部,既可以由变换部进行变换处理,也可以由作为后述的处理器例子的CPU(Central Processing Unit:中央处理器)进行变换处理。

[0168] (解码装置的例子)

[0169] 图3B是表示解码装置的构成例的框图。具体而言,图3B表示作为本公开的解码装置的一例的解码装置1110的构成。

[0170] 存储器1114例如存储有与由编码装置1100生成的编码数据1103相同的数据。存储的数据从存储器1114被读出,作为输入数据1113被输入到解码器1112中。输入数据1113例如是作为解码对象的比特流。存储器1114例如既可以是硬盘或SSD,也可以是其他存储器。

[0171] 另外,解码装置1110也可以不将从存储器1114读出的数据原样作为输入数据1113输入到解码器1112中,而是将所读出的数据进行变换,将变换后的数据作为输入数据1113输入到解码器1112中。变换前的数据例如可以是包含一个以上的比特流的复用数据。这里,复用数据例如也可以是具有ISOBMFF等文件格式的文件。

[0172] 此外,变换前的数据也可以是通过将上述的比特流或文件分割而生成的多个包。也可以从存储器1114读出与比特流不同的数据,将该数据变换为比特流。在此情况下,解码装置1110也可以具备未图示的变换部,由变换部进行变换处理,也可以由作为后述的处理器例子的CPU进行变换处理。

[0173] 解码器1112将输入数据1113解码,生成表示向收听者提示的声音的声音信号

1111。

[0174] (编码装置的另一例)

[0175] 图3C是表示编码装置的另一构成例的框图。具体而言,图3C表示作为本公开的编码装置的另一例的编码装置1120的构成。在图3C中,对于与图3A的构成要素相同的构成要素赋予与图3A的标号相同的标号,对于这些构成要素省略说明。

[0176] 编码装置1100将编码数据1103存储在存储器1104中。另一方面,编码装置1120在具备将编码数据1103发送给外部的发送部1121这一点上与编码装置1100不同。

[0177] 发送部1121将基于编码数据1103或从编码数据1103变换为其他数据形式后的数据生成的发送信号1122发送给其他的装置或服务器。在发送信号1122的生成中使用的数据例如是在编码装置1100中所说明的比特流、复用数据、文件或包。

[0178] (解码装置的另一例)

[0179] 图3D是表示解码装置的另一构成例的框图。具体而言,图3D表示作为本公开的解码装置的另一例的解码装置1130的构成。在图3D中,对于与图3B的构成要素相同的构成要素赋予与图3B的标号相同的标号,对于这些构成要素省略说明。

[0180] 解码装置1110从存储器1114读出输入数据1113。另一方面,解码装置1130在具备从外部接收输入数据1113的接收部1131这一点上与解码装置1110不同。

[0181] 接收部1131将接收信号1132接收而取得接收数据,输出向解码器1112输入的输入数据1113。接收数据既可以与向解码器1112输入的输入数据1113相同,也可以是与输入数据1113不同的数据形式的数据。

[0182] 在接收数据的数据形式与输入数据1113的数据形式不同的情况下,接收部1131也可以将接收数据变换为输入数据1113。或者,也可以由解码装置1130的未图示的变换部或CPU将接收数据变换为输入数据1113。接收数据例如是在编码装置1120中所说明的比特流、复用数据、文件或包。

[0183] (解码器的例子)

[0184] 图4A是表示解码器的构成例的框图。具体而言,图4A表示作为图3B或图3D中的解码器1112的一例的解码器1200的构成。

[0185] 输入数据1113是编码后的比特流,包含作为被编码的声音信号的编码声音数据和在音响处理中使用的元数据。

[0186] 空间信息管理部1201取得输入数据1113中包含的元数据,对元数据进行解析。元数据包含记述有配置在声音空间中的作用于声音的要素的信息。空间信息管理部1201管理对元数据进行解析而得到的在音响处理中使用的空间信息,对渲染部1203提供空间信息。

[0187] 另外,在本公开中,将在音响处理中使用的信息表现为空间信息,但也可以使用其他的表现。例如,也可以将在音响处理中使用的信息表现为声音空间信息,也可以表现为场景信息。此外,在音响处理中使用的信息随时间变化的情况下,对渲染部1203输入的空间信息也可以是被表现为空间状态、声音空间状态或场景状态等的信息。

[0188] 此外,空间信息可以按每个声音空间或按每个场景来管理。例如,在将相互不同的多个房间分别表现为虚拟空间的情况下,也可以将多个房间分别作为相互不同的多个场景进行管理。另外,即使在相同的空间中,也可以根据所表现的状况,作为不同的场景来管理空间信息。

[0189] 因而,也可以针对多个声音空间或多个场景管理多个空间信息。在多个空间信息的管理中,也可以对空间信息赋予分别识别多个空间信息的识别符。

[0190] 空间信息的数据也可以包含于作为输入数据1113的一例的比特流。或者也可以是,比特流中包含空间信息的识别符,从比特流之外的信息源取得空间信息的数据。具体而言,当比特流仅包含空间信息的识别符的情况下,在渲染中,可以利用空间信息的识别符来取得在装置内的存储器或外部服务器中存储的空间信息的数据作为输入数据1113。

[0191] 另外,空间信息管理部1201管理的信息并不限定于比特流中包含的信息。例如,输入数据1113中,作为不包含在比特流中的数据,也可以包含从提供VR或AR的软件或服务器取得的表示空间的特性及构造的数据。

[0192] 此外,输入数据1113也可以包含表示收听者或对象的特性及位置等的信息。此外,输入数据1113也可以包含关于收听者的位置由包括解码装置(1110、1130)的终端所具备的传感器取得的信息,也可以包含表示基于由传感器取得的信息推测出的终端的位置的信息。

[0193] 即,空间信息管理部1201也可以与外部的系统或服务器进行通信,取得空间信息以及收听者位置。空间信息管理部1201可以从外部系统取得时钟同步信息,并且执行与渲染单元1203的时钟同步的处理。

[0194] 另外,上述的说明中的空间既可以是虚拟地形成的空间、即VR空间,也可以是实空间或对应于实空间的虚拟空间、即AR空间或MR空间。此外,也可以将虚拟空间表现为声场或声音空间。此外,上述的说明中的表示位置的信息既可以是表示空间内的位置的坐标值等的信息,也可以是表示相对于规定的基准位置的相对位置的信息,也可以是表示空间内的位置的运动或加速度的信息。

[0195] 声音数据解码器1202将输入数据1113中包含的编码声音数据解码,取得声音信号。

[0196] 立体音响再现系统1000取得的编码声音数据例如是以MPEG—H 3DAudio(ISO/IEC 23008—3)等的规定的形式编码的比特流。另外,MPEG—H 3D Audio只不过是生成包含于比特流中的编码声音数据时能够使用的编码方式的一例。编码声音数据也可以是以其他的编码方式编码的比特流。

[0197] 例如,编码方式也可以是MP3(MPEG—1Audio Layer—3)、AAC(Advanced Audio Coding:高级音频编码)、WMA(Windows Media Audio:视窗媒体音频)、AC3(Audio Codec—3)或Vorbis等的非可逆编解码。或者,编码方式也可以是ALAC(Apple Lossless Audio Codec:Apple无损音频编解码器)或FLAC(Free Lossless Audio Codec:无损音频编解码器)等的可逆编解码。

[0198] 或者,也可以使用上述以外的任意的编码方式。例如,也可以是PCM(pulse code modulation:脉冲编码调制)数据为编码声音数据的一种。在此情况下,例如在该PCM数据的量化比特数是N的情况下,解码处理也可以是将N比特的二进制数变换为渲染部1203能够处理的数形式(例如浮点形式)的处理。

[0199] 渲染部1203取得声音信号和空间信息,使用空间信息对声音信号实施音响处理,输出音响处理后的声音信号(声音信号1111)。

[0200] 空间信息管理部1201在开始渲染之前,读取输入信号的元数据,检测由空间信息

规定的对象和声音等的渲染项目,并发送给渲染部1203。在渲染开始后,空间信息管理部1201掌握空间信息以及收听者的位置的随时间的变化,更新并管理空间信息。并且,将更新后的空间信息发送给渲染部1203。

[0201] 渲染部1203基于输入数据1113所包含的声音信号和从空间信息管理部1201接收到的空间信息,生成并输出附加了音响处理的声音信号。

[0202] 空间信息的更新处理和附加了音响处理的声音信号的处理也可以由相同的线程执行。空间信息管理部1201和渲染部1203可以将处理分配给各自独立的线程。空间信息管理部1201和渲染部1203在不同的线程中执行空间信息的更新处理和附加了音响处理的声音信号的处理的情况下,可以分别设定线程的启动频度,也可以并行地执行处理。

[0203] 当空间信息管理部1201和渲染部1203以不同的独立线程执行处理的情况下,可以优先向渲染部1203分配计算资源。由此,能够安全地执行不允许微小的延迟那样的、例如在1个样本(0.02msec)的延迟下产生噗嗤这样的噪声的出音处理。

[0204] 此时,对空间信息管理部1201限制运算资源的分配。但是,与声音信号的处理相比,空间信息的更新是低频度的处理(例如,收听者的面部朝向的更新那样的处理),因此也可以不必如声音信号的处理那样瞬间进行。因而,即使运算资源的分配被限制,也不会对音响质量产生大的影响。

[0205] 空间信息的更新可以按每预先设定的时间或期间定期地执行,也可以在满足预先设定的条件的情况下执行。此外,空间信息的更新可以由收听者或声音空间的管理者手动执行,也可以将外部系统的变化作为触发来执行。

[0206] 例如,也可以由收听者操作控制器,在自身的化身的站立位置瞬间扭曲、时刻瞬间前进或返回的情况下,更新空间信息。或者,也可以在由虚拟空间的管理者实施了突然变更场地的环境那样的演出的情况下,更新空间信息。在这些情况下,用于更新由空间信息管理部1201管理的空间信息的线程除了定期的启动以外,还可以作为单发的中断处理而启动。

[0207] 图4B是表示解码器的另一构成例的框图。具体而言,图4B表示作为图3B或图3D中的解码器1112的另一例的解码器1210的构成。

[0208] 图4B在输入数据1113不是包含编码声音数据而是包含未被编码的声音信号这一点上与图4A不同。输入数据1113包括包含元数据的比特流和声音信号。

[0209] 空间信息管理部1211由于与图4A的空间信息管理部1201相同,所以省略说明。

[0210] 渲染部1213由于与图4A的渲染部1203相同,所以省略说明。

[0211] 另外,也可以将解码器1112、1200及1210表现为实施音响处理的音响处理部。此外,解码装置1110及1130也可以是声音信号处理装置1001,也可以被表现为音响处理装置。

[0212] (声音信号处理装置的物理构成)

[0213] 图5是表示声音信号处理装置1001的物理构成的一例的图。另外,图5的声音信号处理装置1001也可以是图3B的解码装置1110或图3D的解码装置1130。图3B或图3D所示的多个构成要素也可以通过图5所示的多个构成要素安装。此外,这里说明的构成的一部分也可以装备在声音提示装置1002中。

[0214] 图5的声音信号处理装置1001具备处理器1402、存储器1404、通信IF(Interface)1403、传感器1405和扬声器1401。

[0215] 处理器1402例如是CPU、DSP(Digital Signal Processor:数字信号处理器)或GPU

(Graphics Processing Unit:图形处理器)。也可以通过由该CPU、DSP或GPU执行存储在存储器1404中的程序来实施本公开的音响处理或解码处理。此外,处理器1402例如是进行信息处理的电路。处理器1402也可以是进行包括本公开的音响处理的对声音信号的信号处理的专用电路。

[0216] 存储器1404例如由RAM(Random Access Memory:随机访问存储器)或ROM(Read Only Memory:只读存储器)构成。存储器1404也可以包括以硬盘为代表的磁记录介质或以SSD为代表的半导体存储器等。此外,存储器1404也可以是内置在CPU或GPU中的内部存储器。此外,在存储器1404中,也可以存储由空间信息管理部(1201、1211)管理的空间信息等。此外,也可以存储有后述的阈值数据。

[0217] 通信IF1403例如是与Bluetooth(注册商标)或WIGIG(注册商标)等的通信方式对应的通信模块。声音信号处理装置1001例如经由通信IF1403与其他通信装置进行通信,取得解码对象的比特流。所取得的比特流例如存储在存储器1404中。

[0218] 通信IF1403例如由与通信方式对应的信号处理电路和天线构成。通信方式并不限于Bluetooth(注册商标)及WIGIG(注册商标),也可以是LTE(Long Term Evolution:长期演进)、NR(New Radio:新空口)或Wi-Fi(注册商标)等。

[0219] 此外,通信方式并不限于上述那样的无线通信方式。通信方式也可以是Ethernet(注册商标)、USB(Universal Serial Bus:通用串行总线)或HDMI(注册商标)(High-Definition Multimedia Interface:高清多媒体接口)等的有线通信方式。

[0220] 传感器1405进行用来推测收听者的位置及朝向的感测。具体而言,传感器1405基于身体的一部分或整体的位置、朝向、运动、速度、角速度及加速度等中的某一个以上的检测结果,推测收听者的位置及/或朝向,生成表示收听者的位置及/或朝向的位置/朝向信息。

[0221] 另外,也可以是声音信号处理装置1001的外部的装置具备传感器1405。身体的一部分也可以是收听者的头部等。位置/朝向信息也可以是表示实空间中的收听者的位置及/或朝向的信息,也可以是表示以规定的时间点的收听者的位置及/或朝向为基准的收听者的位置及/或朝向的位移的信息。此外,位置/朝向信息也可以是表示与立体音响再现系统1000或具备传感器1405的外部装置的相对的位置及/或朝向的信息。

[0222] 传感器1405例如是相机等的摄像装置或LiDAR(Light Detection And Ranging:激光雷达)等的测距装置。传感器1405也可以拍摄收听者的头部的运动,通过对拍摄到的图像进行处理来检测收听者的头部的运动。此外,也可以使用利用例如毫米波等的任意的频带的无线进行位置推测的装置作为传感器1405。

[0223] 此外,声音信号处理装置1001也可以从具备传感器1405的外部的设备经由通信IF1403取得位置信息。在此情况下,声音信号处理装置1001也可以不包括传感器1405。这里,外部的设备例如是在图2中所说明的声音提示装置1002或佩戴在收听者的头部上的立体影像再现装置等。此时传感器1405例如将陀螺仪传感器及加速度传感器等的各种传感器组合而构成。

[0224] 例如作为收听者的头部的运动的速度,传感器1405可以检测以在声音空间内相互正交的3轴的至少一个为旋转轴的旋转的角速度,也可以检测以上述3轴的至少一个为位移方向的位移的加速度。

[0225] 例如作为收听者的头部的运动的量,传感器1405可以检测以在声音空间内相互正交的3轴的至少一个为旋转轴的旋转量,也可以检测以上述3轴的至少一个为位移方向的位移量。具体而言,传感器1405检测6DoF的位置(x,y,z)及角度(yaw,pitch,roll)作为收听者的位置。传感器1405将陀螺仪传感器及加速度传感器等用于运动检测的各种传感器组合而构成。

[0226] 另外,传感器1405也可以由用来检测收听者的位置的相机或GPS(Global Positioning System:全球定位系统)接收机等实现。也可以使用通过使用LiDAR等作为传感器1405实施自位置推测而得到的位置信息。例如,传感器1405在立体音响再现系统1000由智能电话实现的情况下被内置在智能电话中。

[0227] 此外,传感器1405中也可以包括检测声音信号处理装置1001的温度的热电偶等的温度传感器。此外,传感器1405中也可以包括声音信号处理装置1001所具备的电池、或检测与声音信号处理装置1001连接的电池的剩余量的传感器等。

[0228] 扬声器1401例如具有振动板、磁铁或音圈等的驱动机构和放大器,将音响处理后的声音信号作为声音提示给收听者。扬声器1401根据经由放大器放大的声音信号(更具体地讲,表示声音的波形的波形信号)使驱动机构动作,由驱动机构使振动板振动。这样,对应于声音信号而振动的振动板产生声波,声波在空气中传播而传递到收听者的耳朵,收听者感知声音。

[0229] 另外,这里列出了声音信号处理装置1001具备扬声器1401、经由扬声器1401提示音响处理后的声音信号的例子,但声音信号的提示机构并不限定于上述的构成。

[0230] 例如,也可以对通过通信模块连接的外部的声音提示装置1002输出音响处理后的声音信号。通过通信模块进行的通信既可以是有线也可以是无无线。此外,作为另一例,声音信号处理装置1001具备输出声音的模拟信号的端子,在端子上连接耳塞等的线缆,从耳塞等提示声音信号。

[0231] 在上述的情况下,声音提示装置1002也可以是佩戴在收听者的头部或身体的一部分上的头戴式耳机、耳塞、头戴显示器、颈部扬声器或可穿戴扬声器等。或者,声音提示装置1002也可以是由被固定的多个扬声器构成的环绕扬声器等。并且,声音提示装置1002也可以再现声音信号。

[0232] (编码装置的物理构成)

[0233] 图6是表示编码装置的物理构成的一例的图。图6的编码装置1500也可以是图3A的编码装置1100或图3C的编码装置1120,也可以图3A或图3C所示的多个构成要素由图6所示的多个构成要素安装。

[0234] 图6的编码装置1500具备处理器1501、存储器1503和通信IF1502。

[0235] 处理器1501例如是CPU、DSP或GPU。也可以通过由该CPU、DSP或GPU执行存储在存储器1503中的程序来实施本公开的编码处理。此外,处理器1501例如是进行信息处理的电路。处理器1501也可以是进行包括本公开的编码处理的对声音信号的信号处理的专用电路。

[0236] 存储器1503例如由RAM或ROM构成。存储器1503也可以包括以硬盘为代表的磁记录介质或以SSD为代表的半导体存储器等。此外,存储器1503也可以是嵌入在CPU或GPU中的内部存储器。

[0237] 通信IF1502例如是与Bluetooth(注册商标)或WIGIG(注册商标)等的通信方式对

应的通信模块。编码装置1500例如经由通信IF1502与其他通信装置进行通信,发送被编码后的比特流。

[0238] 通信IF1502例如由与通信方式对应的信号处理电路和天线构成。通信方式并不限于Bluetooth(注册商标)及WIGIG(注册商标),也可以是LTE、NR或Wi-Fi(注册商标)等。此外,通信方式并不限于无线通信方式。通信方式也可以是Ethernet(注册商标)、USB或HDMI(注册商标)等的有线通信方式。

[0239] (渲染部的构成)

[0240] 图7是表示渲染部的构成例的框图。具体而言,图7表示与图4A及图4B的渲染部1203及1213对应的渲染部1300的详细的构成的一例。

[0241] 渲染部1300由解析部1301、选择部1302和,合成部1303构成,对于输入信号中包含的声音数据附加音响处理并输出。

[0242] 输入信号例如由空间信息、传感器信息和声音数据构成。输入信号也可以包含由声音数据和元数据(控制信息)构成的比特流,在此情况下,也可以在元数据中包含空间信息。

[0243] 空间信息是与立体音响再现系统1000所形成的声音空间(三维声场)有关的信息,由与声音空间中包含的对象(object)有关的信息和与收听者有关的信息构成。对象中存在发出声音而成为音源的音源对象和不发出声音的非发声对象。也可以将音源对象简单表现为音源。

[0244] 非发声对象起到将音源对象所发出的声音反射的障碍物对象的作用,但也有音源对象起到将其他音源对象所发出的声音反射的障碍物对象的作用的情况。也可以将障碍物对象表现为反射对象。

[0245] 作为对音源对象和非发声对象共同赋予的信息,有位置信息、形状信息及对象将声音反射时的音量的衰减率等。

[0246] 位置信息由欧几里得空间的例如X轴、Y轴及Z轴的3轴的坐标值表示,但也可以并不一定是三维信息。例如,位置信息也可以是由X轴及Y轴这两轴的坐标值表示的二维信息。对象的位置信息由用网格或体素表现的形状的代表位置决定。

[0247] 形状信息也可以包含与表面的材料有关的信息。

[0248] 衰减率可以由0以上且1以下的实数表现,也可以由负的分贝值表现。由于在实空间中音量不会通过反射而放大,所以衰减率被设定为负的分贝值,但例如为了呈现非现实空间的恐怖感,也可以特意设定1以上的衰减率、即正的分贝值。

[0249] 此外,衰减率也可以按构成多个频带的每个频带设定不同的值,也可以按每个频带独立地设定值。此外,在按对象表面的材料的每个种类设定衰减率的情况下,也可以基于与表面的材料有关的信息使用对应的衰减率的值。

[0250] 此外,空间信息也可以包含表示对象是否属于生物的信息及表示对象是否是运动体的信息等。在对象是运动体的情况下,由位置信息表示的位置也可以随着时间移动。在此情况下,变化后的位置或变化量的信息被传送给渲染部1300。

[0251] 与音源对象有关的信息除了对音源对象和非发声对象共同赋予的信息以外还包括声音数据、以及为了将声音数据向声音空间内放射所需的信息。声音数据是表示与声音的频率及强弱有关的信息等的信息,并且是表现了被收听者感知的声音的数据。

[0252] 声音数据典型的是PCM信号,但也可以是使用MP3等的编码方式来压缩的数据。在此情况下,至少需要在该信号到达合成部1303之前进行解码,所以渲染部1300也可以包括未图示的解码部。或者,该信号也可以由声音数据解码器1202解码。

[0253] 对于一个音源对象,可以设定一个声音数据,也可以设定多个声音数据。另外,也可以对声音数据赋予识别各声音数据的识别信息,与音源对象有关的信息也可以包含声音数据的识别信息。

[0254] 为了将声音数据向声音空间内放射所需的信息例如也可以包含在声音数据的再现中用作基准的基准音量的信息、表示声音数据的性质(也称为特性)的信息、与音源对象的位置有关的信息、以及与音源对象的朝向有关的信息(即,与音源对象发出的声音的指向性有关的信息)等。

[0255] 基准音量的信息例如是将声音数据向声音空间放射时的音源位置处的声音数据的振幅值的有效值,也可以作为分贝(db)值以浮点表示。

[0256] 例如,在基准音量为0db的情况下,也可以表示不使声音数据所表示的信号电平的音量增减而以原来的音量从与音源对象的位置有关的信息所指示的位置对声音空间放射声音。另外,在基准音量为-6db的情况下,也可以表示将声音数据所表示的信号电平的音量设为约一半,从与音源对象的位置有关的信息所指示的位置对声音空间放射声音。

[0257] 基准音量的信息可以按每个声音数据赋予,也可以对多个声音数据统一赋予。

[0258] 表示声音数据的性质的信息例如可以是与音源的音量有关的信息,并且是表示音源的音量的时间序列上的变动的信息。

[0259] 例如,在声音空间是虚拟会议室、音源是说话者的情况下,音量在短时间内断续地转变。即,有声部分和无声部分交替产生。另外,在声音空间为音乐厅、音源为演奏者的情况下,音量维持一定的时间长度。另外,在声音空间为战场、音源为爆炸物的情况下,爆炸音的音量仅一瞬间变大,之后持续为无声或小的状态。

[0260] 这样,音源的音量的信息不仅包含声音的大小的信息,还可以包含声音的大小的转变的信息。这样的信息也可以用作表示声音数据的性质的信息。

[0261] 转变的信息也可以由以时间序列表示频率特性的数据来表现。转变的信息也可以由表示有声区间的持续时间长度的数据来表现。转变的信息也可以由表示有声区间的持续时间长度和无声区间的持续时间长度的时间序列的数据来表现。转变的信息也可以由以时间序列列举出声音信号的振幅能够被视为平稳(能够视为大致恒定)的持续时间和该期间的该信号的振幅值的多组的数据等表现。

[0262] 转变的信息也可以由能够将声音信号的频率特性视为平稳的持续时间的数据表现。转变的信息也可以通过以时间序列列举出声音信号的频率特性能够被视为平稳的持续时间和该期间的该频率特性的多组的数据等来表现。转变的信息例如也可以以表示频谱图的概略形状的数据的形式表现。

[0263] 另外,用作上述频率特性的基准的音量也可以是上述基准音量。基准音量的信息和表示声音数据的性质的信息可以用于使收听者感知的直接音或反射音的音量的计算处理,也可以用于是否使收听者感知的选择处理。关于表示声音数据的性质的信息的其他例子及利用方法,在后面叙述。

[0264] 与音源对象的朝向有关的信息(朝向信息)典型地由yaw、pitch及roll表现。或者,

也可以省略roll的旋转,将音源对象的朝向信息用方位角(yaw)及仰角(pitch)表现。音源对象的朝向信息也可以随着时间变化,在发生了变化的情况下被传送到渲染部1300。

[0265] 与收听者有关的信息是与声音空间中的收听者的位置及朝向有关的信息。与位置有关的信息(位置信息)由欧几里得空间的XYZ轴的位置表示,但也可以并不一定是三维信息,也可以是二维信息。与收听者的朝向有关的信息(朝向信息)典型地由yaw、pitch及roll表现。或者,也可以省略roll的旋转,将收听者的朝向信息用方位角(yaw)及仰角(pitch)表现。

[0266] 收听者的位置信息及朝向信息也可以随着时间变化,在发生了变化的情况下被传送到渲染部1300。

[0267] 传感器信息是包含由收听者佩戴的传感器1405检测出的旋转量或位移量等和收听者的位置及朝向的信息。传感器信息被传送到渲染部1300,渲染部1300基于传感器信息来更新收听者的位置及朝向的信息。传感器信息例如也可以包含便携终端使用GPS、相机或LiDAR等实施自位置推测而得到的位置信息。

[0268] 此外,也可以不是传感器1405,而是将经由通信模块从外部取得的信息作为传感器信息来检测。也可以从传感器1405取得表示声音信号处理装置1001的温度的信息及表示电池的剩余量的信息。此外,也可以实时地取得声音信号处理装置1001或声音提示装置1002的运算资源(CPU能力、存储器资源或PC性能等)。

[0269] 解析部1301进行输入信号中包含的声音信号及从空间信息管理部(1201、1211)接受的空间信息的解析,检测直接音及反射音的生成所需要的信息、以及是否生成反射音的选择所需要的信息。

[0270] 直接音及反射音的生成所需要的信息例如是关于直接音及反射音各自,与到达收听位置为止的路径、到达为止所需的时间、以及到达时的音量等有关的信息。

[0271] 输出的反射音的选择所需要的信息是表示直接音与反射音的关系的信息,例如是与直接音与反射音的时间差有关的值、以及与收听位置处的直接音与反射音的音量比有关的值等。

[0272] 另外,在音量由对数轴的分贝的单位表现的情况下(在分贝区域表示音量的情况下),两个信号的音量比当然由分贝值的差表示。具体而言,两个信号的音量比可以是各个信号的振幅值在分贝区域表示时的差。该值也可以基于能量值或功率值等计算。此外,该差在分贝区域中可以被称为增益的差或简单称为增益差。

[0273] 即,本公开中的音量比实质上是信号的振幅之比,所以也可以表现为Sound volume ratio(声音音量比)、Volume ratio(音量比)、Amplitude ratio(振幅比)、Sound level ratio(声音等级比)、Sound intensity ratio(声音强度比)或Gain ratio(增益比)等。此外,在音量的单位是分贝的情况下,本公开中的音量比当然也可以改称作音量差。

[0274] 在本公开中,“音量比”典型地是指两个声音的音量由分贝的单位表现时的增益差,在实施方式例子中,阈值数据也典型的是由在分贝区域表示的增益差规定。但是,音量比并不限于分贝区域的增益差。当使用在分贝区域以外表示的音量比的情况下,也可以将在分贝区域规定的阈值数据变换为计算出的音量比的单位来使用。或者,也可以预先将各个单位规定的阈值数据保存在存储器中。

[0275] 即,即使代替音量比而使用例如能量值或功率值等的比,也显然能够将本公开中

的算法应用于本公开的课题的解决。

[0276] 直接音与反射音的时间差例如是直接音的到来时间(到来时刻)与反射音的到来时间(到来时刻)的时间差。直接音与反射音的时间差也可以是直接音和反射音分别到达收听位置的时刻的时间差、直接音和反射音分别到达收听位置为止所需的时间之差、或直接音的发声结束的时刻与反射音到达收听位置的时刻的时间差。关于这些值的计算方法在后面叙述。

[0277] 选择部1302使用解析部1301计算出的信息和阈值数据,选择是否生成反射音。换言之,选择部1302判定是否将反射音选择为生成对象反射音。再换言之,选择部1302选择生成多个反射音中的哪个反射音。

[0278] 阈值数据例如在横轴具有直接音与反射音的时间差的值、纵轴具有直接音与反射音的音量比的曲线图中被表示为该反射音被感知还是不被感知的边界(阈值)。阈值数据也可以由将直接音与反射音的时间差的值作为变量来具有的近似式表现,也可以由具有直接音与反射音的时间差的值作为索引、并具有对应的阈值的排列表现。

[0279] 选择部1302例如在直接音的到来时刻与反射音的到来时刻的时间差的值中的、直接音的到来时音量与反射音的到来时音量的音量比是比参照阈值数据来设定的阈值大的值的情况下,选择生成反射音。

[0280] 直接音的到来时刻与反射音的到来时刻的时间差换言之是直接音和反射音分别到达收听位置为止所需的时间之差。此外,也可以使用直接音的发声结束的时间点与反射音到来收听位置的时间点的时间差作为直接音与反射音的时间差。在此情况下,也可以使用与将直接音的到来时刻与反射音的到来时刻的时间差用于基准而设定的阈值数据不同的阈值数据,也可以使用共同的阈值数据。

[0281] 关于阈值数据,可以从声音信号处理装置1001的存储器1404取得,也可以从外部的存储装置经由通信模块取得。关于阈值数据的存储方法及阈值的设定方法在后面叙述。

[0282] 合成部1303将直接音的声音信号与选择部1302选择生成的反射音的声音信号进行合成。

[0283] 具体而言,合成部1303基于解析部1301计算出的直接音到来时刻及直接音到来时音量的信息,将所输入的声音信号加工而生成直接音。此外,合成部1303基于关于选择部1302所选择的反射音的反射音到来时刻及反射音到来时音量的信息,将所输入的声音信号加工而生成反射音。接着,合成部1303将所生成的直接音及反射音合成并输出。

[0284] (渲染部的动作)

[0285] 图8是表示声音信号处理装置1001的动作例的流程图。在图8中表示了主要由声音信号处理装置1001的渲染部1300执行的处理。

[0286] 在输入信号的解析处理(图8的S101)中,解析部1301对输入到声音信号处理装置1001中的输入信号进行解析,检测在声音空间中能够产生的直接音及反射音。这里检测的反射音是由选择部1302选择为最终要由合成部1303生成的反射音的反射音候选。此外,解析部1301对输入信号进行解析,计算在直接音及反射音的生成中需要的信息和在生成对象反射音的选择中需要的信息。

[0287] 首先,计算直接音及反射音各自的特性。具体而言,计算直接音及反射音分别到达收听者时的到来时间及到来时音量。在作为反射对象在声音空间中存在多个对象的情况

下,对于多个对象分别计算反射音的特性。

[0288] 直接音到来时间 (t_d) 基于直接音到来路径 (p_d) 计算。直接音到来路径 (p_d) 是将音源对象的位置信息 $S(x_s, y_s, z_s)$ 与收听者的位置信息 $A(x_a, y_a, z_a)$ 连结的路径。直接音到来时间 (t_d) 是将连结位置信息 $S(x_s, y_s, z_s)$ 与位置信息 $A(x_a, y_a, z_a)$ 的路径的长度除以音速 (约340m/秒) 而得到的值。

[0289] 例如,路径的长度 (X) 通过 $(x_s - x_a)^2 + (y_s - y_a)^2 + (z_s - z_a)^2$ 的平方根求出。音量与距离成反比地衰减。因而,在音源对象的位置信息 $S(x_s, y_s, z_s)$ 中的音量是 N , 单位距离是 U 的情况下,直接音到来时音量 (l_d) 通过 $l_d = N * U / X$ 求出。

[0290] 音源位置处的音量 N 也可以是在前面说明的基准音量。

[0291] 反射音到来时间 (t_r) 基于反射音到来路径 (p_r) 计算。反射音到来路径 (p_r) 是将反射音的声像的位置与位置信息 $A(x_a, y_a, z_a)$ 连结的路径。

[0292] 另外,反射音的声像的位置的导出例如可以使用“镜像法”或“光线跟踪法”,也可以使用其以外的任意的声像位置的导出方法。镜像法是假设室内的壁面上的反射波在相对于壁面与音源对称的位置处存在镜像,并假设从该镜像的位置放射声波来模拟声像的方法。光线跟踪法是通过跟踪光线或声线等那样的直线传播的波来模拟在某点处观测的像(声像)等的方法。

[0293] 图9是表示收听者和障碍物对象比较远的位置关系的图。图10是表示收听者和障碍物对象比较近的位置关系的图。即,图9及图10分别表示在相对于音源位置隔着壁而对称的位置处形成反射音的声像的例子。通过基于这样的关系在xyz轴上求出反射音的声像的位置,反射音的到来时间可与计算直接音的到来时间的方法同样地求出。

[0294] 反射音到来时间 (t_r) 是将连结反射音的声像的位置与位置信息 $A(x_a, y_a, z_a)$ 的路径的长度 (Y) 除以音速 (约340m/秒) 而得到的值。音量与距离成反比地衰减。因而,在音源位置处的音量是 N , 单位距离是 U , 反射中的音量的衰减率是 G 的情况下,反射音到来时音量 (l_r) 通过 $l_r = N * G * U / Y$ 求出。

[0295] 如前面说明那样,衰减率 G 既可以由0以上且1以下的实数表现,也可以由负的分贝值表现。在此情况下,信号整体的音量衰减与 G 相应的量。此外,衰减率也可以按构成多个频带的每个频带设定。在此情况下,解析部1301按信号的每个频率成分施加被指定的衰减率。此外,解析部1301为了运算量的削减,也可以使用多个频带的多个衰减率的代表值或平均值等作为整体的衰减率,使信号整体的音量相应地衰减。

[0296] 接着,解析部1301计算在生成对象反射音的选择中需要的、直接音到来时音量 (l_d) 与反射音到来时音量 (l_r) 的比即音量比 (L)、以及直接音与反射音的时间差 (T)。

[0297] 直接音到来时音量 (l_d) 与上述 l_r 的比即音量比 (L) 例如由 $L = (N * G * U / Y) / (N * U / X) = G * X / Y$ 求出。求出的值由于是音量的比,所以 N 和 U 的值可以是预先设定的任意的值。

[0298] 直接音与反射音的时间差 (T) 例如也可以是在直接音和反射音分别到达收听位置时所需的时间的时间差。例如,在直接音和反射音分别到达收听位置时所需的时间的差 (T) 由 $T = t_r - t_d$ 求出。

[0299] 此外,时间差 (T) 也可以是直接音和反射音分别到达收听位置的时刻的差。此外,时间差 (T) 也可以是直接音的讲话结束的时刻和反射音到达收听位置的的时刻的时间差。即,时间差 (T) 也可以是在收听位置直接音结束的时刻与反射音开始的的时刻的时间差。

[0300] 接着,在反射音的选择处理(图8的S102)中,选择部1302选择是否生成由解析部1301计算出的反射音。换言之,选择部1302判定是否将反射音选择为生成对象反射音。在存在多个反射音的情况下,选择部1302选择是否生成各反射音。选择部1302选择是否生成各反射音的结果,既可以从多个反射音中选择一个以上的生成对象反射音,也可以一个生成对象反射音都不选择。

[0301] 另外,选择部1302并不限于生成处理,也可以选择其他处理的应用对象的反射音。例如,选择部1302也可以选择双耳处理的应用对象的反射音。此外,选择部1302基本上仅选择处理对象的一个以上的反射音。但是,选择部1302也可以仅选择不是处理对象的一个以上的反射音。并且,也可以对没有被选择的一个以上的反射音应用处理。

[0302] 例如,反射音的选择基于由解析部1301计算出的音量比(L)和时间差(T)进行。通过基于直接音与反射音的时间差(T)进行选择处理,与仅基于直接音与反射音的音量差进行选择处理的情况相比,能够更适当地选择对收听者的感知的影响程度大的反射音。

[0303] 具体而言,是否生成反射音的选择,例如通过比较对应于直接音与反射音的时间差的直接音与反射音的音量比和预先设定的阈值来进行。阈值是参照阈值数据来设定的。阈值数据是表示对于直接音的反射音是否被收听者感知的边界的指标,由直接音的到来时音量(I_d)与反射音的到来时音量(I_r)之比定义。

[0304] 另外,阈值对应于由与时间差(T)对应地设定的数值等表现的值。阈值数据对应于时间差(T)与阈值的关系,对应于为了确定或计算时间差(T)下的阈值而使用的表数据或关系式。阈值数据的形式及种类并不限于表数据或关系式。

[0305] 图11是表示直接音和反射音的时间差与阈值的关系的图。例如,也可以参照如图11所示的、按直接音与反射音的时间差的每个值预先设定的音量比的阈值数据。或者,也可以参照从图11所示的阈值数据通过内插或外插等得到的阈值数据。

[0306] 并且,根据阈值数据确定由解析部1301计算出的时间差(T)下的音量比的阈值。并且,选择部1302根据由解析部1301计算出的直接音与反射音的音量比(L)是否高于该阈值,决定是否将反射音选择为生成对象反射音。

[0307] 通过使用按直接音与反射音的时间差的每个值预先设定的音量比的阈值数据进行选择处理,能够实现考虑到后掩蔽或优先效应的选择处理。关于阈值数据的种类、形式、存储方法及设定方法等,详细的说明在后面叙述。

[0308] 接着,在直接音及反射音的生成处理(图8的S103)中,合成部1303生成直接音的声音信号和由选择部1302选择为生成对象反射音的反射音的声音信号并进行合成。

[0309] 直接音的声音信号通过对输入信息中包含的音源对象的声音数据应用由解析部1301计算出的到来时间(t_d)及到来时音量(I_d)而生成。具体而言,进行使声音数据延迟到来时间(t_d)的量并乘以到来时音量(I_d)的处理。使声音数据延迟的处理是在时间轴上使声音数据的位置前后移动的处理。例如,也可以应用如专利文献2所公开的不使音质劣化地使声音数据延迟的处理。

[0310] 反射音的声音信号与直接音同样通过对音源对象的声音数据应用由解析部1301计算出的到来时间(t_r)及到来时音量(I_d)而生成。

[0311] 但是,反射音的生成中的到来时音量(I_r)与直接音的到来时音量不同,是被应用了反射中的音量的衰减率G的值。G可以是对全频带一起应用的衰减率。或者,也可以为了反

映通过反射产生的频率成分的偏倚,按每个规定的频带规定反射率。在此情况下,应用到来时音量(1r)的处理也可以作为按每个频带乘以衰减率的处理即频率均衡器的处理来实施。

[0312] 在上述的例子中,计算直接音及反射音候选分别到达收听者时的路径长度。进而,基于各路径长度计算到来时间和到来时音量。并且,基于它们的时间差和音量比,进行反射音候选的选择处理。

[0313] 另外,作为另一例,也可以基于直接音及反射音分别到达收听者时的路径长度进行选择处理,省略直接音及反射音的到来时间及到来时音量的计算以及时间差及音量比的计算。在此情况下,也可以针对路径长度比预先设定与路径长度差相应的阈值。并且,也可以根据计算出的路径长度比是否是计算出的路径长度差相应的阈值以上来进行选择处理。由此,能够在削减运算量的同时基于与时间差对应的路径长度差进行选择处理。

[0314] 此外,除了路径长度差以外,也可以使用表示声音的传播速度的参数的值、或对声音的传播速度的参数带来影响的参数的值。

[0315] (选择处理的详细)

[0316] 对是否生成反射音的选择处理的详细情况进行说明。

[0317] 反射音的选择通过比较设定了直接音与反射音的时间差(T)下的直接音到来时音量与反射音到来时音量的比即音量比的阈值与由解析部1301计算出的音量比(L)来进行。例如,在按直接音与反射音的时间差的每个值预先设定的音量比的阈值中,参照由解析部1301计算出的直接音与反射音的时间差(T)下的音量比的阈值。并且,根据由解析部1301计算出的音量比(L)是否高于阈值,决定是否将反射音选择为生成对象反射音。

[0318] 时间差(T)例如可以是直接音和反射音各自到达收听位置的时刻之差、在直接音和反射音各自到达收听位置时所需的时间的时间差、以及直接音的发声结束的时间点与反射音到达收听位置的时间点的时间差中的任何一种。这里,直接音的结束时刻例如也可以通过对直接音的到来时刻加上直接音的持续时间来求出。

[0319] 关于阈值数据,也可以通过听觉神经作用或大脑的认知作用,更具体地讲,通过后述的优先效应、后述的经时掩蔽现象或它们的组合,基于收听者的感知能够检测到两个声音的偏差的最小时间差来决定。具体的数值既可以根据已知的随时间掩蔽效应、优先效应或回声检测极限等的研究成果导出,也可以通过以应用于该虚拟空间为前提的试听实验来求出。

[0320] 图12A、图12B及图12C是表示阈值数据的设定方法的例子的图。如图12A、图12B及图12C所示,阈值数据在横轴具有直接音与反射音的时间差、纵轴具有直接音与反射音的音量比的曲线图中,由该反射音被感知还是不被感知的边界(阈值)表示。

[0321] 阈值数据也可以由将直接音与反射音的时间差作为变量来具有的近似式表现。此外,阈值数据也可以作为如图11所示的直接音与反射音的时间差的索引和对应于索引的阈值的排列而被存储在存储器1404的区域中。

[0322] 此外,在图12C的例示4中的与横轴平行的线的高度(最小可听限度)被用作阈值的情况下,不是将反射音与直接音的音量比(L)与阈值进行比较,而是将反射音的音量本身与阈值进行比较。这是因为,该阈值表示是否能够被收听者感知的边界的音量,是用于将音量比该阈值小的声音决定为不再现的声音的阈值。即,与最小可听限度对应的阈值不是针对反射音的音量与直接音的音量之比的阈值。

[0323] 在最小可听限度被用作阈值的情况下,阈值与时间差(T)无关而恒定,因此也可以不计算时间差(T)。

[0324] 另外,在通过解析处理(图8的S101)生成了多个反射音的情况下,也可以对全部反射音执行选择处理,也可以基于通过预先设定的评价方法对各反射音导出的评价值,仅对评价值高的反射音执行选择处理。这里,反射音的评价值对应于反射音的感知上的重要度。另外,评价值高对应于评价值大,这些表现也可以相互替换。

[0325] 选择部1302例如也可以通过与音源的音量、音源的视觉性、音源的定位性、反射对象(障碍物对象)的视觉性或直接音与反射音的几何学关系性等相应的预先设定的评价方法来计算反射音的评价值。

[0326] 具体而言,也可以是音源的音量越大则评价值越高。此外,为了使视觉定位与音响定位一致,也可以在能够从收听者目视音源对象或反射对象(障碍物对象)的情况下或音源对象的定位性高的情况下评价值高。

[0327] 此外,直接音与反射音的到来角的开口以及直接音与反射音的到来时间的差对空间的掌握带来较大的影响。因此,也可以在直接音与反射音的到来角的开口大的情况下,或者在直接音与反射音的到来时间的差大的情况下,评价值高。

[0328] 音源的音量的信息也可以表示按每个内容设定的基准音量、音量的时间上的转变、或者其双方。

[0329] 例如,在虚拟空间是虚拟会议室、直接音是会话声音的情况下,音量在短时间内断续地转变。即,有声部分和无声部分交替产生。另外,在虚拟空间是音乐厅、直接音是乐曲的演奏的情况下,音量维持一定的时间长度。另外,在虚拟空间为战场、直接音为爆炸音的情况下,音量仅一瞬间变大,之后持续为无声或小的状态。

[0330] 这样,音源的音量的信息不仅包含与向虚拟空间放射该声音时的音量设定相当的基准音量的信息,还可以包含声音的大小的转变的信息。

[0331] 转变的信息也可以由以时间序列表示频率特性的数据来表现。转变的信息也可以由表示有声区间的持续时间长度的数据来表现。转变的信息也可以由表示有声区间的持续时间长度和无声区间的持续时间长度的时间序列的数据来表现。转变的信息也可以由以时间序列列举出声音信号的振幅能够被视为平稳(能够视为大致恒定)的持续时间和该期间的该信号的振幅值的多组的数据等表现。

[0332] 转变的信息也可以由能够将声音信号的频率特性视为平稳的持续时间的数据表现。转变的信息也可以由以时间序列列举出能够视为声音信号的频率特性平稳的持续时间和该期间的该频率特性的多组的数据等表现。

[0333] 另外,一直以来广泛进行将信号的频率特性的时间上的转变用于虚拟空间的音响处理的对策(专利文献1等)。在鉴于这样的现有技术的情况下,上述的组当然也可以是频率特性恒定的时间长度与其频率特性的组。

[0334] 几何学关系性也可以是虚拟空间内的音源、收听者以及反射对象的位置的关系性。通过它们的关系性,能够在几何学上计算出直接音及反射音分别到来的路径长度。因而,如果利用音量与距离成反比的关系,则能够计算相对于直接音的基准音量的反射音的基准音量。

[0335] 在反射音的基准音量的计算中,也可以使用反射对象的反射系数。另外,作为反射

系数,也可以使用通常使用的典型的值。另一方面,在存在反射对象被吸音材料覆盖等这样的特别条件的情况下,作为反射对象的反射系数,也可以使用特别赋予的反射系数。

[0336] 反射音也可以根据反射音的音量进行评价。反射音的音量也可以根据如上所述的直接音和反射音的几何学关系性、以及对反射对象赋予的指标而求出。也可以将该音量与预先设定的阈值进行比较来评价反射音。

[0337] 进而,也可以将表示音源的音量的时间上的转变的信息反映到评价中。例如,在表示音源的音量的时间上的转变的信息表示有声区间的持续时间长度的情况下,在时刻处于有声区间内的情况下,反射音的评价值也可以维持原样。另一方面,在时刻处于有声区间外的情况下,即使反射音的基准音量超过阈值,也可以进行使反射音的评价值变小或设为零的处理。

[0338] 或者,表示音源的音量的时间上的转变的信息也可以是将声音信号的振幅被视为大致恒定的持续时间和该期间的该信号的振幅值的多个组以时间序列列举的数据。在该情况下,也可以进行与数据中的振幅值的变化联动地使反射音的基准音量变化来评价反射音的处理。

[0339] 另外,作为表示直接音的音量的信息,也可以使用基准音量的信息和在时间上转变的音量的信息双方。例如,也可以在基于基准音量的信息计算出评价值之后,使用转变的音量的信息来修正该评价值。

[0340] 在反射音的评价中,可以执行上述的全部方法,也可以仅执行一部分。例如,既可以通过多个评价方法来评价反射音,也可以通过一个评价方法来评价反射音。

[0341] 在通过多个评价方法评价反射音的情况下,是否选择反射音可以基于通过多个评价方法综合决定的评价值来决定,也可以基于多个评价方法各自的评价值来决定。

[0342] 声音信号处理装置1001也可以在基于多个评价方法的每一个来决定是否选择反射音的情况下,在基于多个评价方法的多个评价结果全部表示选择声音的情况下,选择声音。或者,声音信号处理装置1001也可以在基于多个评价方法的多个评价结果中的某一个表示选择声音的情况下,选择声音。

[0343] 另外,例如,也可以对第1至第3评价方法设置优先级。而且,声音信号处理装置1001也可以在通过第1评价方法判定为不选择声音的情况下,不依赖于第2及第3评价方法中的判定结果而最终判定为不选择声音。另外,音响处理装置也可以在通过第2及第3评价方法中的某一方判定为不选择声音但通过另一方判定为选择声音的情况下,最终判定为选择声音。

[0344] 另外,选择处理和评价处理可以独立地执行,也可以仅执行某一方。另外,也可以仅对在评价处理中决定为选择的反射音执行评价处理,在评价处理中再次决定是否选择反射音。或者,也可以仅对在评价处理中决定为不选择的反射音执行评价处理,在评价处理中再次决定是否选择反射音。

[0345] 上述的选择处理可以被解释为根据直接音的性质选择反射音的处理。例如,在根据直接音的性质选择反射音的处理中,根据直接音的性质设定或调整在反射音的选择中使用的阈值。或者,基于音源的音量、音源的视觉性、音源的定位性、反射对象(障碍物对象)的视觉性以及直接音与反射音的几何学关系性等中的某一个以上来计算在反射音的选择中使用的评价值。

[0346] 此外,在根据直接音的性质选择反射音的处理中,并不限于根据直接音的性质设定或调整阈值的处理、以及计算在处理对象的反射音的选择中使用的评价值的处理,也可以进行其他处理。此外,在进行根据直接音的性质设定或调整阈值的处理、或者计算在处理对象的反射音的选择中使用的评价值的处理的情况下,可以将处理的一部分变更,也可以追加新的处理。

[0347] 另外,设定阈值也可以包括调整阈值以及变更阈值等。

[0348] (阈值的设定方法)

[0349] 在选择处理中使用的阈值数据例如也可以参考已知的基于优先效应的回声检测极限的值或基于后掩蔽效应的掩蔽阈值来设定。

[0350] 优先效应是指当从两处听到声音时识别为音源处于在时间上先听到的哪一处的现象。如果两个短声音融合而听成一个声音,则整体的声音所听到的位置(定位位置)大体主要由最初的声音的位置决定。回声检测极限是通过优先效应发生的现象,是收听者的感知检测到两个声音的偏差的最小时间差。

[0351] 在图12C的例示2中,横轴对应于反射音(回声)的到来时间,具体而言,对应于从直接音的到来时刻到反射音的到来时刻的延迟时间。纵轴对应于能够检测的反射音相对于直接音的音量比,具体而言,对应于是否能够检测到伴随延迟时间而到来的反射音的阈值。

[0352] 图13是表示阈值的设定方法的例子的图。图13中的横轴对应于反射音的到来时间,具体而言,对应于直接音与反射音的时间差(T)。图13中的纵轴对应于反射音的音量。具体而言,图13中的纵轴既可以对应于对于直接音的音量相对地设定的反射音的音量(音量比),也可以对应于不依赖于直接音的音量而绝对地决定的反射音的音量。

[0353] 例如,在如图9所示收听者与障碍物对象比较远的情况下,反射音的到来时间变晚,如图13的C所示,阈值被设定得低。结果,在图9的情况下,生成反射音。另一方面,在如图10所示收听者与障碍物对象比较近的情况下,反射音的到来时间比图9的情况变早,如图13的B所示,阈值被设定得高。结果,在图10的情况下,不生成反射音。

[0354] 此外,阈值数据也可以被存储在存储器1404中,在选择处理时被从存储器1404取得并用于选择处理。

[0355] 图14是表示选择处理的一例的流程图。首先,选择部1302指定由解析部1301检测到的反射音(S201)。接着,选择部1302检测直接音与反射音的音量比(L)以及直接音与反射音的时间差(T)(S202及S203)。

[0356] 时间差(T)例如可以是直接音和反射音分别到达收听位置时所需的时间的时间差、直接音到来时刻与反射音到来时刻的时间差、以及直接音的发声结束的时间点与反射音到达收听位置的时间点的时间差中的任何一种。这里说明基于直接音到来时刻与反射音到来时刻的时间差的例子。

[0357] 具体而言,选择部1302根据音源对象及收听者的位置信息以及障碍物对象的位置信息及形状信息,计算直接音的路径的长度与反射音的路径的长度的差。并且,选择部1302通过将该长度的除以音速,检测直接音到达收听者位置的时刻与反射音到达收听者位置的时刻的时间差(T)。

[0358] 到达收听者的音量相对于音源的音量,与到收听者的距离成比例地衰减(与距离成反比)。因而,直接音的音量通过将音源的音量除以直接音的路径的长度而得到。反射音

的音量通过将音源的音量除以反射音的路径的长度、再乘以对虚拟的障碍物对象赋予的衰减率而得到。选择部1302通过计算它们的音量的比来检测音量比。

[0359] 此外,选择部1302使用阈值数据,确定对应于时间差(T)的阈值(S204)。接着,选择部1302判定检测出的音量比(L)是否是阈值以上(S205)。

[0360] 在音量比(L)是阈值以上的情况下(S205中为“是”),选择部1302将该反射音选择为生成对象的反射音(S206)。在音量比(L)小于阈值的情况下(S205中为“否”),选择部1302不将该反射音选择为生成对象的反射音(S207)。即,在此情况下,选择部1302将该反射音决定为生成对象外的反射音。

[0361] 然后,选择部1302判定是否有未指定的反射音(S208)。如果有未指定的反射音(S208中为“是”),则选择部1302反复进行上述的处理(S201~S207)。如果没有未指定的反射音(S208中为“否”),则选择部1302结束处理。

[0362] 该选择处理既可以对在解析处理中所生成的全部反射音执行,也可以仅对上述的评价值高的反射音执行。

[0363] (阈值的存储方法的详细情况)

[0364] 有关本实施方式的阈值数据被存储在声音信号处理装置1001的存储器1404中。存储的阈值数据的形式及种类可以是任意的形式及任意的种类。在存储多个形式及多个种类的阈值的情况下,在选择处理中,决定将哪个形式及哪个种类的阈值用于反射音的选择处理都可以。关于决定将哪个阈值数据用于选择处理的方法在后面进行叙述。

[0365] 此外,也可以将多个形式及多个种类的阈值数据组合而存储。也可以将所组合的阈值数据从空间信息管理部(1201、1211)读出,设定在选择处理中使用的阈值。另外,也可以将存储在存储器1404中的阈值数据存储到空间信息管理部(1201、1211)中。

[0366] 阈值数据例如也可以以各时间差下的阈值存储,以描绘图12C的[例示1]及[例示2]所示的阈值的线。

[0367] 此外,阈值数据也可以如图11所示作为将阈值与时间差(T)建立了对应的表数据存储。即,阈值数据也可以被存储为将时间差(T)作为索引来具有的表数据。当然,图11所示的阈值是一例,阈值并不限于图11的例子。此外,也可以不存储阈值自身,而用将时间差(T)作为变量来具有的函数将阈值近似,存储该函数的系数。此外,也可以将多个近似式组合而存储。

[0368] 在存储器1404中,也可以存储与表示时间差(T)和阈值的关系的关系式有关的信息。即,也可以存储将时间差(T)作为变量来具有的式子。各时间差(T)的阈值也可以用直线或曲线近似,存储表示直线或曲线的几何学形状的参数。例如,在该几何学形状是直线的情况下,也可以存储用来表现该直线的起点及斜率。

[0369] 此外,也可以按直接音的每个性质设定阈值数据的种类及形式并存储。此外,也可以存储用来根据直接音的性质调整阈值并用于选择处理的参数。关于根据直接音的性质调整阈值并用于选择处理的处理,作为阈值的设定方法的变形例而在后面叙述。

[0370] 作为将多种阈值数据组合而存储的例子,也可以如图12C的[例示3]所示,按每个时间差(T)存储掩蔽阈值和回声检测极限阈值中的较大一方的值。也可以如图12C的[例示4]所示,按每个时间差(T)存储在虚拟空间中再现的最小音量和回声检测极限阈值中的较大一方的值。

[0371] 多个种类的阈值数据的组合并不限于此。例如,也可以在多个阈值数据中按每个时间差(T)存储最大值的信息。

[0372] 此外,在上述中,与阈值有关的信息具有时间的项目作为一维索引。与阈值有关的信息也可以具有还包含与到来方向有关的变量的二维或三维索引。

[0373] 图15是表示直接音的方向、反射音的方向、时间差和阈值的关系的图。例如,如图15所示,也可以存储按照直接音的方向(θ)、反射音的方向(γ)、时间差(T)和音量比(L)的关系预先计算出的阈值。

[0374] 直接音的方向(θ)对应于相对于收听者的直接音的到来方向的角度。反射音的方向(γ)对应于相对于收听者的反射音的到来方向的角度。这里,将收听者朝向的方向设定为0度。时间差(T)对应于向收听位置的直接音的到来时间与反射音的到来时间的差。音量比(L)对应于直接音的到来时音量与反射音的到来时音量的音量比。

[0375] 当然,图15所示的阈值是一例,阈值并不限于图15的例子。此外,在图15中,主要例示了直接音的到来方向的角度(θ)是0度情况下的阈值。但是,直接音的到来方向(θ)为0度以外的情况下的阈值也存储在存储器1404中。

[0376] 此外,在上述中,阈值被存储为将直接音的到来方向的角度(θ)和反射音的到来方向的角度(γ)作为独立的变量或索引来具有的排列。但是,直接音的到来方向的角度(θ)和反射音的到来方向的角度(γ)也可以不作为独立的变量来使用。

[0377] 例如,也可以使用直接音的到来方向的角度(θ)与反射音的到来方向的角度(γ)的角度差。该角度差对应于直接音的到来方向与反射音的到来方向所成的角度,也可以表现为直接音和反射音的到来角。

[0378] 图16是表示角度差、时间差和阈值的关系的图。例如,也可以如图16所示的例子那样存储将直接音的到来方向的角度(θ)与反射音的到来方向的角度(γ)的角度差(Φ)用作变量而预先计算出的阈值。当然,图16所示的阈值是一例,阈值并不限于图16的例子。

[0379] 在图16的例子中,能够减少在阈值的导出中使用的变量的数量。因此,能够减少存储在存储器1404中的阈值的数量。因而,能够削减存储在存储器1404中的数据量。

[0380] 此外,在使用直接音的到来方向的角度(θ)与反射音的到来方向的角度(γ)的角度差(Φ)的情况下,也可以以二维排列存储阈值数据。此外,在选择处理中,也可以使用三维排列来计算直接音的到来方向的角度(θ)与反射音的到来方向的角度(γ)的差。

[0381] 关于使用与到来方向相应的阈值来选择反射音的方法,在后面叙述。

[0382] (阈值的设定方法的第1变形例)

[0383] 在图12A、图12B及图12C的例子中,多个形式及多个种类的阈值也可以存储在空间信息管理部(1201、1211)中。并且,也可以决定将多个形式及多个种类的阈值中的哪个形式及哪个种类的阈值用于反射音的选择处理。具体而言,也可以如图12C的例示3所示,在与反射音到来时刻对应的时间差(T)时采用最高的阈值。

[0384] 此外,也可以如例示4所示,存储掩蔽阈值、回声检测极限的阈值及表示在虚拟空间中再现的最小音量的阈值。并且,也可以在与反射音到来时刻对应的时间差(T)时采用最高的阈值。

[0385] (阈值的设定方法的第2变形例)

[0386] 作为阈值的设定方法的另一例,对根据直接音的性质设定阈值的方法进行说明。

[0387] 图17是表示图7所示的渲染部1300的另一构成例的框图。图17的渲染部1300与图7的渲染部1300相比在包括阈值调整部1304这一点上不同。阈值调整部1304以外的说明与在图7中说明的内容相同,所以省略。

[0388] 阈值调整部1304基于表示声音信号的性质的信息,从阈值数据中选择要由选择部1302使用的阈值。或者,阈值调整部1304也可以基于表示声音信号的性质的信息来调整阈值数据中包含的阈值。

[0389] 表示声音信号的性质的信息也可以包含在输入信号中。并且,阈值调整部1304也可以从输入信号取得表示声音信号的性质的信息。或者,也可以是解析部1301通过对所受理的输入信号中包含的声音信号进行解析,导出声音信号的性质,将表示声音信号的性质的信息输出给阈值调整部1304。

[0390] 表示声音信号的性质的信息既可以在开始渲染处理之前取得,也可以在渲染中随时取得。

[0391] 此外,阈值调整部1304也可以不包含在声音信号处理装置1001中,也可以其他通信装置具有阈值调整部1304的作用。在此情况下,解析部1301或选择部1302也可以经由通信IF1403从其他通信装置取得表示声音信号的性质的信息、与性质相应的阈值数据、或用来根据性质调整阈值数据的信息。

[0392] 图18是表示选择处理的另一例的流程图。图19是表示选择处理的再另一例的流程图。在图18及图19中,根据直接音的性质设定阈值。具体而言,在图18中,阈值调整部1304基于时间差(T)和声音信号的性质,从阈值数据确定阈值。在图19中,阈值调整部1304基于声音信号的性质,调整基于时间差(T)从阈值数据确定的阈值。

[0393] 以下,说明各例的动作。另外,关于与图14的例子共同的处理省略说明。

[0394] 首先,说明在图18中表示的处理的例子。这里,按直接音的每个性质将阈值数据预先存储在存储器1404中。由此,将与多个性质对应的多个阈值数据预先存储在存储器1404中。并且,阈值调整部1304从多个阈值数据中确定在反射音的选择处理中使用的阈值数据。

[0395] 例如,阈值调整部1304基于输入信号取得直接音的性质(S211)。阈值调整部1304也可以取得与输入信号建立了关联的直接音的性质。接着,阈值调整部1304确定与时间差(T)及直接音的性质对应的阈值(S212)。

[0396] 此外,如图19所示,阈值调整部1304也可以基于直接音的性质来调整由选择部1302确定的阈值(S221)。

[0397] 在任何情况下,都可以在输入信号中包含表示声音信号的性质的信息、用来根据声音信号的性质调整阈值的信息或它们两者。阈值调整部1304也可以使用它们的一方或两者来调整阈值。

[0398] 此外,表示声音信号的性质的信息、用来调整阈值的信息或它们两者也可以通过与包含声音信号的输入信号不同的输入信号来传送。在此情况下,也可以在包含声音信号的输入信号中包含与不同于该输入信号的输入信号建立关联的信息,也可以在存储器1404中与有关阈值的信息一起存储有与不同于输入信号的输入信号建立关联的信息。

[0399] 在图18及图19的例子中,根据直接音的性质、即声音信号的性质,设定在反射音的选择中使用的阈值。既可以如图18那样使用预先按每个性质设定的阈值数据,也可以如图19那样根据声音信号的性质来调整阈值。此外,也可以根据声音信号的性质来调整阈值数

据的参数。

[0400] 此外,由阈值调整部1304进行的动作也可以由解析部1301或选择部1302进行。例如,也可以是解析部1301取得声音信号的性质。此外,也可以是选择部1302根据声音信号的性质设定阈值。

[0401] 接着,对声音信号的性质与阈值的关系进行说明。

[0402] 连续到达收听者的耳朵的两个短声音如果其时间间隔充分短,则被听成一个声音。该现象被称作优先效应。已知优先效应仅对于不连续的、即瞬态的声音发生(非专利文献1)。因而,在声音信号表示平稳音的情况下,也可以与声音信号表示非平稳音的情况相比将回声检测极限设定得较低。

[0403] 即,按照这样的优先效应的特性,例如在直接音是平稳的声音的情况下,将阈值设定得较小。此外,也可以平稳性越高则将阈值设定得越小。

[0404] 对声音信号的性质是平稳的情况下的处理的例子进行说明。首先,阈值调整部1304或解析部1301基于随着时间的经过的声音信号的频率成分的变动量来判定平稳性。例如,在该变动量少的情况下,判定为平稳性高。相反,在该变动量多的情况下,判定为平稳性低。判定的结果,既可以设定表示平稳性的等级的标志,也可以根据该变动量而设定表示平稳性的参数。

[0405] 接着,阈值调整部1304也可以基于表示声音信号的平稳性的标志或参数等表示平稳性的信息,调整阈值数据或阈值,将调整后的阈值数据或阈值设定为在选择部1302中使用的阈值数据或阈值。

[0406] 或者,也可以是用来根据表示直接音的平稳性的信息设定阈值数据的参数预先存储在存储器1404中。在此情况下,阈值调整部1304也可以判定声音信号的平稳性,基于表示平稳性的信息及参数来设定在反射音的选择中使用的阈值数据。

[0407] 或者,也可以是阈值数据的多个参数与直接音的平稳性的多个样式对应地预先存储在存储器1404中。在此情况下,阈值调整部1304也可以判定声音信号的平稳性,基于直接音的平稳性的样式选择阈值数据的参数,基于阈值数据的参数来设定在反射音的选择中使用的阈值数据。

[0408] 另外,关于声音信号的平稳性,也可以每当被输入声音信号时基于声音信号的频率成分的变动量进行判定。

[0409] 或者,关于声音信号的平稳性,也可以基于预先与声音信号建立了关联的表示平稳性的信息来进行判定。即,也可以是表示声音信号的平稳性的信息与声音信号建立关联而预先存储在存储器1404中。解析部1301也可以每当被输入声音信号时取得与声音信号建立了关联的表示平稳性的信息。并且,阈值调整部1304也可以基于与声音信号建立了关联的表示平稳性的信息来调整阈值。

[0410] 作为根据声音信号的性质设定阈值的另一例,也可以在声音信号表示较短的声音(咔哒音等)的情况下,与声音信号表示较长的声音的情况相比,将回声检测极限的应用范围设定得较短。该处理基于优先效应的特性。

[0411] 已知通过优先效应,连续到达收听者的耳朵的两个短声音如果其时间间隔充分短则被听成一个声音。该时间间隔的上限依赖于声音的长度。例如,该时间间隔的上限在咔哒音时是约5ms,而在人的声音或音乐那样的复杂的声音时也有时为40ms(非专利文献1)。

[0412] 按照这样的优先效应的特性,例如在直接音的持续时间长度较短的声音的情况下,设定较短的时间长度的阈值。此外,直接音的持续时间长度越短,则设定越短的时间长度的阈值。

[0413] 设定较短的时间长度的阈值是指,在直接音与反射音的时间差(T)较小的范围中,设定与基于优先效应的特性的回声检测极限对应的阈值。在该范围外,不设置与基于优先效应的特性的回声检测极限对应的阈值。即,在该范围外,阈值较小。因而,对于较短的声音设定较短的时间长度的阈值,可以对应于对于较短的声音设定较小的阈值。

[0414] 作为根据直接音的性质设定阈值的另一例,也可以在直接音是断续的声音(讲话等)的情况下,与直接音是持续的声音(音乐等)的情况相比将阈值设定得低。

[0415] 例如在直接音对应于讲话的情况下,反复存在有声部和无声部,作为掩蔽效应,在无声部仅发生后掩蔽效应。另一方面,在直接音如音乐内容那样是持续的声音的情况下,发生后掩蔽效应和基于此时产生的声音的同时掩蔽效应这两个掩蔽效应。因而,综合性的掩蔽效应在音乐等的情况下比讲话等的情况高。

[0416] 按照如上所述的掩蔽效应的特性,也可以在音乐等的情况下与讲话等的情况相比将阈值设定得高。相反,也可以在讲话等的情况下与音乐等的情况相比将阈值设定得低。即,也可以在直接音中断续的部分多的情况下,将阈值设定得小。

[0417] 如上所述,表示直接音的性质的信息也可以是表示直接音的平稳性、断续性以及持续时间等的信息。另外,表示直接音的性质的信息也可以是它们的任意的组合。另外,表示直接音的性质的信息可以是表示它们中的任一个的时间变动的信息,也可以是表示它们的任意组合的时间变动的信息。即,表示直接音的性质的信息也可以是表示直接音的时间变动的信息。

[0418] 例如,如在平稳性的判定的说明中所示,表示直接音的性质的信息也可以是频率特性的时间序列数据。这里,频率特性也可以通过每个频段的增益值、针对时间轴信号的傅里叶级数、或者用于求出频率包络线的LPC系数或倒谱系数等那样的惯用的形式来表现。

[0419] 进而,表示直接音的性质的信息也可以是作为表示直接音的断续性的信息,以时间序列列举了信号的振幅平稳的持续时间和该期间的该信号的振幅值的多个组的信息(振幅的包络线的概略形状)。这里,振幅值也可以用相对于基准音量的比来表现。

[0420] 并且,表示直接音的性质的信息也可以是与直接音的频率特性有关的信息。例如,表示直接音的性质的信息也可以是表示直接音的频率特性的平稳性的信息。具体而言,表示直接音的性质的信息也可以是以时间序列列举了频率特性的变动小的状态的持续时间和该期间的该信号的频率特性的多个组的信息(频谱图的概略形状)。这里,用作上述频率特性的基准的音量也可以是上述基准音量。

[0421] 例如,表示直接音的时间变动的信息是表示直接音的包络线的信息。表示直接音的时间变动的信息也可以在图12C的[例示4]所记载的“最小可听限度”为阈值的情况下使用。与最小可听限度进行比较的信号是反射音的音量。

[0422] 反射音的音量根据音源、收听者及反射对象的位置的信息通过几何学的计算而得到。具体而言,得到相对于音源的基准音量的反射音的基准音量。通过使用音源的声音的大小的转变的信息作为表示直接音的性质的信息来使反射音的基准音量增减,能够准确地求出时时刻刻的反射音的音量。其理由在于,音源的音量的变动被反映到反射音的音量的变

动。

[0423] 在调整反射音的音量后,通过将反射音的音量与阈值进行比较,能够更准确地适当选择听觉上所需的反射音。

[0424] 当然,不调整反射音的基准音量,而基于音源的声音的大小的转变的信息的倒数来调整阈值,并将调整后的阈值与反射音的基准音量进行比较,也能够得到相同的结果。即,可以使用音源的声音的大小的转变的信息来调整反射音的基准音量,也可以使用音源的声音的大小的转变的信息来调整阈值。反射音的基准音量的调整和阈值的调整相互对应。

[0425] 根据反射声音的物体的表面的组成,声音的反射率(伴随反射的声音的衰减率)在每个频带不同。因此,如后所述,也可以针对反射声音的物体,按每个频带来关联声音的反射率(衰减率)。根据这样的反射率的信息和频谱图的信息,能够更准确地判定是否选择该反射音。例如,执行以下处理。

[0426] 具体地,例如,根据频谱图的信息表示在某时间区间高频带的频率成分比低频带的频率成分占优势。另外,例如,根据声音的反射率的信息表示在高频带的频率成分中,与低频带的频率成分相比反射率极其小。

[0427] 在该情况下,即使音源的信号在时间轴上的振幅大,通过将由频谱图的信息表示的频率成分与由反射率的信息表示的每个频带的衰减率相乘而得到的反射音的音量也变小,反射音有可能不被选择。

[0428] 如上所述,表示直接音的性质的信息也可以是表示直接音的时间变动的信息。例如,表示直接音的性质的信息也可以表示通过以预先设定的时间长度分析直接音而得到的值。

[0429] 具体而言,表示直接音的性质的信息也可以是按每个预先设定的时间长度计算直接音的平均能量或平均振幅而得到的信息。另外,表示直接音的性质的信息也可以是通过按每个短时间分析长度计算直接音的能量或平均振幅,按比短时间分析长度长的每个长时间分析长度计算能量或平均振幅的加权平均而得到的信息。

[0430] 更具体而言,例如,表示直接音的时间变动的信息也可以是通过按预先设定的每个短时间长度(例如5ms,以下将该时间长度的帧表现为分析帧)来计算直接音的能量或平均振幅而得到的信息。另外,表示直接音的时间变动的信息也可以是由在过去N-1个分析帧中计算出的能量或平均振幅的加权平均表示的信息。

[0431] 假设第n分析帧的能量由E(n)表现的情况下,按照下式求出表示直接音的性质的信息I(n)。

[0432] [数学式1]

$$I(n) = a(0) \cdot E(n) + \sum_{i=1}^{N-1} a(i) \cdot E(n-i)$$

[0433]

... (式1)

[0434] 这里,参数a(i)表示权重系数。通常,a(i)被设定为使得a(i) ≥ 0、并且a(i)的总和为1。然而,a(i)的设定方法不限于此。

[0435] 此外,每当直接音被取入5ms时,计算表示直接音的性质的信息I(n)。即,能够以低

延迟计算表示直接音的性质的信息 $I(n)$ 的时间变动。因此,该方法适合应用于需要实时性的应用。

[0436] 另外,也可以按照下式求出表示直接音的性质的信息 $I(n)$ 。

[0437] [数学式2]

$$I(n) = b(0) \cdot E(n) + \sum_{i=1}^{N-1} b(i) \cdot I(n-i)$$

[0438]

... (式2)

[0439] 这里,参数 $b(i)$ 表示权重系数。通常, $b(i)$ 被设定为使得 $b(i) \geq 0$ 、并且 $b(i)$ 的总和为1。然而, $b(i)$ 的设定方法不限于此。

[0440] 在该式中,递归地求出表示直接音的性质的信息 $I(n)$ 。因此,能够以少的运算量计算长时间长度的平均能量。

[0441] 上述的式1及式2可以视为 $E(n)$ 为输入信号且 $I(n)$ 为输出信号的滤波器。在该情况下,式1是移动平均(MA)模型的滤波器,式2是自回归(AR)模型的滤波器,均具有低通滤波器的特性。另外,也可以使用将两者组合而成的ARMA模型的滤波器。

[0442] 另外,导出表示直接音的时间变动的信息的方法不限于上述的计算式或滤波器,也可以使用其他公知的方法。如上所述,表示直接音的时间变动的信息表示通过以预先设定的时间长度分析直接音而得到的值。也可以从平均能量以外的观点分析直接音。

[0443] 另外,如上所述,表示直接音的性质的信息也可以是与直接音的频率特性有关的信息。与直接音的频率特性有关的信息也可以使用直接音的频率特性计算出的信息。例如,与直接音的频率特性有关的信息可以通过以预先设定的分析长度对直接音的低频成分进行平均而作为低频成分的平均能量来得到的信息。

[0444] 具体而言,通过对分析帧长度中包含的直接音应用具有低通特性的滤波器,求出直接音的低频成分。根据该低频成分的能量或平均振幅,与上述式1同样地导出表示直接音的性质的信息。

[0445] 假设第 n 分析帧的低频成分的能量由 $E_L(n)$ 表现的情况下,按照下式求出表示直接音的性质的信息 $I(n)$ 。

[0446] [数学式3]

$$I(n) = c(0) \cdot E_L(n) + \sum_{i=1}^{N-1} c(i) \cdot E_L(n-i)$$

[0447]

... (式3)

[0448] 这里,参数 $c(i)$ 表示权重系数。通常, $c(i)$ 被设定为使得 $c(i) \geq 0$ 、并且 $c(i)$ 的总和为1。然而, $c(i)$ 的设定方法不限于此。

[0449] 此外,每当直接音被取入5ms时,计算表示直接音的性质的信息 $I(n)$ 。即,能够以低延迟计算表示直接音的性质的信息 $I(n)$ 的时间变动。因此,该方法适合应用于需要实时性的应用。

[0450] 另外,与式2同样,也可以按照下式求出表示直接音的性质的信息 $I(n)$ 。

[0451] [数4]

$$I(n) = d(0) \cdot E_L(n) + \sum_{i=1}^{N-1} d(i) \cdot I(n-i)$$

[0452] . . . (式4)

[0453] 这里,参数d(i)表示权重系数。通常,d(i)被设定为使得d(i) ≥ 0、并且d(i)的总和为1。然而,d(i)的设定方法不限于此。

[0454] 在该式中,递归地求出表示直接音的性质的信息I(n)。因此,能够以少的运算量计算长时间长度的平均能量。

[0455] 上述的式3及式4可以视为E(n)为输入信号且I(n)为输出信号的滤波器。在该情况下,式3是移动平均(MA)模型的滤波器,式4是自回归(AR)模型的滤波器,均具有低通滤波器的特性。另外,也可以使用将两者组合而成的ARMA模型的滤波器。

[0456] 在上述中,在求出直接音的低频成分的方法中使用具有低通特性的滤波器,但求出直接音的低频成分的方法不限于此。另外,导出表示直接音的时间变动的信息的方法不限于上述的计算式或滤波器,也可以使用其他公知的方法。例如,可以通过对直接音执行频率变换来计算直接音的频谱。并且,可以计算频谱的低频成分的能量或平均振幅。

[0457] 另外,在上述中,在表示直接音的时间变动的信息的导出中使用MA模型或者AR模型。这些模型的系数可以是预先设定的固定值,也可以是在时间上可变的值即可变值。

[0458] 另外,上述的分析帧长度与上述的信息更新线程的产生间隔的关系也可以是如下关系。

[0459] 例如,在分析帧的时间长度为TA(msec)、信息更新线程的产生间隔为TU(msec)的情况下,MA滤波器中的上述的(式1)以及(式3)的N的值也可以是由TU/TA给出的值左右。另外,AR滤波器中的上述(式2)以及(式4)的b(i)及d(i) (1 ≤ i < N)也可以是该滤波器的时间常数成为TU(msec)左右的值。

[0460] 上述设定的理由是因为,期待该滤波器在信息更新的间隔期间内收敛。

[0461] 另一方面,在上述的设定中在表示直接音的时间变动的信息中值过于急剧地变动的情况下,也可以预先计算I(n)。而且,也可以将预先计算出的I(n)应用于反射音的选择处理。例如,在第t时刻的帧的处理中,也可以使用I(t+tau)。这里,tau是根据滤波器的收敛特性而确定的值。在收敛慢的情况下,与收敛快的情况相比,tau的值大。

[0462] 此外,作为表示直接音的特性的信息,也可以使用根据直接音计算的听觉掩蔽(频率掩蔽)的信息。听觉掩蔽的信息表示被直接音掩蔽的频域中的振幅值的阈值。也可以进行将相同的频域中的反射音的振幅值与阈值进行比较而不选择振幅值比阈值小的反射音的处理。频域中的反射音的振幅值也可以作为表示反射音的特性的信息而由解析部1301取得。

[0463] 通过这样根据直接音的性质来设定在反射音的选择中使用的阈值,能够适当地选择在听感上需要的反射音,能够将听觉特性有效地反映到立体音响再现系统1000中。检测直接音的性质的处理、根据性质决定阈值的处理以及根据性质调整阈值的处理既可以在渲染处理时进行,也可以在开始渲染处理之前进行。

[0464] 例如,这些处理可以在虚拟空间制作时(软件的制作时)、虚拟空间的处理开始时(软件启动时或渲染开始时)、或在虚拟空间的处理中定期地发生的信息更新线程所发生的

定时等进行。此外,虚拟空间制作时既可以是在音响处理的开始前构建虚拟空间的定时,也可以是虚拟空间的信息(空间信息)的取得时,也可以是软件的取得时。

[0465] 这里,在信息更新线程中,执行用于更新由空间信息管理部(1201,1211)管理的空间信息的处理。

[0466] 信息更新线程所承担的作用例如是基于收听者所佩戴的VR护目镜的位置和朝向来更新配置在虚拟空间内的收听者的化身的位置和朝向的处理、或者在虚拟空间内移动的物体的位置的更新等。这样的处理在以几十Hz左右的较低频度启动的处理线程内提供。

[0467] 在这样的发生频度低的处理线程中,也可以进行更新表示直接音的性质的信息的处理。其理由是,直接音的性质变动的频度比用于音频输出的音频处理帧的发生频度低。由此,能够相对地减小该处理的运算负荷。另外,如果以不必要地快的频度更新信息,则带来产生脉冲噪声的风险。通过以低频度更新信息,还能够避免这样的风险。

[0468] (阈值的设定方法的第3变形例)

[0469] 作为阈值的设定方法的另一例,也可以根据对该虚拟空间的再现进行处理的操作资源(CPU能力、存储器资源、PC性能或电池剩余量等)设定阈值。更具体地讲,声音信号处理装置1001的传感器1405检测操作资源的量,在操作资源量少的情况下将阈值设定得高。由此,更多的反射音的音量比阈值小,所以能够减少被进行双耳处理的反射音,能够减少运算量。

[0470] 或者,在信号处理由智能电话或VR护目镜等那样的用蓄电池驱动的设备进行的情况下,期望以使处理持续长时间为优先并节约操作资源。在这样的情况下,也可以都不用检测操作资源的量或剩余量,就将阈值设定得高。

[0471] (阈值的设定方法的第4变形例)

[0472] 作为阈值的设定方法的另一例,也可以通过声音信号处理装置1001或声音提示装置1002具备未图示的阈值设定部,由该虚拟空间的管理者或收听者设定阈值。

[0473] 例如,也可以是,能够由佩戴声音提示装置1002的收听者选择是收听对象反射音少、运算量少的“节能模式”,还是收听对象反射音多、运算量多的“高性能模式”。或者,也可以是,能够由管理立体音响再现系统1000的管理者或立体音响内容的制作者选择模式。此外,也可以不是模式,而是能够直接选择阈值或阈值数据。

[0474] (渲染部的动作的第1变形例)

[0475] 图20是表示声音信号处理装置1001的动作的第1变形例的流程图。在图20中表示了主要由声音信号处理装置1001的渲染部1300执行的处理。在本变形例中,渲染部1300的动作中追加了音量补偿处理。

[0476] 例如,解析部1301取得数据(输入信号)(S301)。接着,解析部1301对数据进行解析(S302)。接着,选择部1302基于解析结果,进行是否选择反射音的判定(S303)。接着,合成部1303基于没有被选择的反射音进行音量补偿处理(S304)。接着,合成部1303进行直接音及反射音的音响处理(S305)。并且,合成部1303将直接音及反射音作为音频输出(S306)。

[0477] 在上述的处理(S301~S306)中,音量补偿处理(S304)以外是与上述的其他例共同的处理,所以省略它们的说明。

[0478] 音量补偿处理对应于在选择处理中没有被选择的反射音而执行。例如,通过在选择处理中没有选择反射音,发生音量感的欠缺。通过音量补偿处理,能抑制伴随着这样的音

量感的欠缺的违和感。作为补偿音量感的方法的例子,公开以下两个方法。在两个方法中使用哪个方法都可以。

[0479] 首先,对通过使直接音的音量上升来补偿音量感的方法进行说明。合成部1303使直接音的音量上升与没有被选择的反射音的音量相应的量而生成直接音。由此,因没有生成反射音而失去的音量感得到补偿。

[0480] 合成部1303在使该音量上升时,也可以按照反射音的频率特性,按每个频率成分使音量上升。为了能够进行这样的处理,也可以按规定的每个频带赋予反射对象衰减的音量的衰减率。由此,能够导出反射音的频率特性。

[0481] 接着,说明通过使反射音合成到直接音中来补偿音量感的方法。在该方法中,合成部1303将没有被选择的反射音加到直接音中而生成直接音,由此补偿因没有生成反射音造成的音量感。所生成的直接音中反映了没有被选择的反射音的音量(振幅)、频率及延迟等。

[0482] 在使直接音的音量上升的方法的情况下,补偿处理的运算量是很轻微的,但只有音量得到补偿。在使反射音合成到直接音中的方法的情况下,补偿处理的运算量与使直接音的音量上升的方法相比较,但反射音的特性更正确地得到补偿。

[0483] 在哪种情况下,都没有生成反射音而仅生成直接音,所以整体的运算量削减。特别是,包含卷积HRTF的處理的双耳处理所需的运算量削减,所以整体的运算量大幅削减。其理由是因为,双耳处理所需的运算量远大于上述的补偿处理所需的运算量。

[0484] 另外,在反射音没有被选择的理由是反射音的音量低于掩蔽阈值的情况下,由于不会失去音量感,所以也可以不进行补偿处理而仅仅去除反射音。

[0485] (渲染部的动作的第2变形例)

[0486] 图21是表示声音信号处理装置1001的动作的第2变形例的流程图。在图21中,主要表示了由声音信号处理装置1001的渲染部1300执行的處理。在本变形例中,渲染部1300的动作中追加了左右音量差调整处理。

[0487] 例如,解析部1301对输入信号进行解析(S401)。接着,解析部1301检测声音的到来方向(S402)。接着,选择部1302调整由左右耳感知的声音的音量之差(S403)。此外,选择部1302调整由左右耳感知的声音的到来时间之差(延迟)(S404)。选择部1302基于调整后的声音的信息,进行是否选择反射音的判定(S405)。

[0488] 在上述的處理(S401~S405)中,左右音量差调整(S403)及延迟调整(S404)以外是与上述的其他例共同的處理,所以省略它们的说明。

[0489] 图22是表示化身、音源对象和障碍物对象的配置例的图。例如,在收听者的正面方向是0度的情况下,如图22所示如果直接音的到来方向(θ)和反射音的到来方向(γ)的极性(例如正负)不同,则在两耳间产生的音量差被修正。

[0490] 具体而言,在 θ 和 γ 的极性不同的情况下,直接音和反射音中主要(先)感知声音的耳朵不同。在此情况下,选择部1302中作为左右音量差调整(S403),匹配于主要感知反射音的耳朵的位置来调整直接音的音量。例如,选择部1302通过对直接音到达收听者时的音量乘以 $(1.0 - 0.3\sin(\theta))$ ($0 \leq \theta \leq 180$),使直接音到达收听者时的音量衰减。

[0491] 选择部1302通过计算如上述那样修正后的直接音的音量与反射音的音量的音量比,并将计算出的音量比与阈值进行比较,来进行是否选择反射音的判定。由此,在两耳间产生的音量差被修正,更正确地导出影响反射音的直接音的音量,更正确地进行是否选择

反射音的判定。

[0492] 此外,选择部1302也可以除了左右音量差调整(S403)以外,还作为延迟调整(S404),匹配于感知反射音的耳朵的位置使直接音的到来时间延迟。具体而言,选择部1302也可以通过对直接音的到来时间加上 $(a(\sin\theta+\theta)/c)$ ms(a 是头部的半径, c 是音速),使直接音的到来时间延迟。

[0493] (渲染部的动作的第3变形例)

[0494] 对设定与到来方向相应的阈值的方法进行说明。

[0495] 图23是表示选择处理的再另一例的流程图。关于与图14的例子共同的处理省略说明。在图23的例子中,选择部1302使用与到来方向相应的阈值来选择反射音。

[0496] 具体而言,选择部1302根据由解析部1301计算出的直接音到来路径(pd)、反射音到来路径(pr)及化身的朝向信息 D ,计算将化身的朝向用于基准而决定的直接音到来方向(θ)及反射音到来方向(γ)。即,选择部1302检测直接音到来方向(θ)及反射音到来方向(γ)(S231)。化身的朝向对应于收听者的朝向。化身的朝向信息 D 也可以包含在输入信号中。

[0497] 选择部1302使用除了直接音到来方向(θ)及反射音到来方向(γ)以外还包括时间差(T)的3个索引,根据如图15所示的三维排列,确定在选择处理中使用的阈值(S232)。

[0498] 作为例子,说明在如图22所示配置有化身、音源对象和障碍物对象的情况下在选择处理中使用的阈值的设定方法。

[0499] 从输入信号取得化身、音源对象及障碍物对象的位置信息和化身的朝向信息 D 。使用这些位置信息及朝向信息 D ,计算化身的朝向被设定为0度的情况下的直接音的方向(θ)和反射音的声像的方向(γ)。在图22的情况下,直接音的方向(θ)是20度左右,反射音的声像的方向(γ)是265度(-95度)左右。

[0500] 接着,参照图15所示的以三维排列存储的阈值数据,从与两个方向(θ)及(γ)的值及由解析部1301计算出的时间差(T)的值对应的排列区域确定阈值。在不存在与计算出的(θ)、(γ)、(T)的值对应的索引的情况下,也可以确定与最接近的索引对应的阈值。

[0501] 作为其他方法,也可以通过基于与接近于计算出的(θ)、(γ)、(T)的值的一个以上的索引对应的一个以上的阈值,进行插值、内插或外插等处理,来确定阈值。例如,也可以基于与(0度、225度、 T)、(0度、270度、 T)、(45度、225度、 T)及(45度、270度、 T)这4个索引对应的4个阈值来确定与(20度、265度、 T)对应的阈值。

[0502] 对基于直接音的到来方向的角度(θ)与反射音的到来方向的角度(γ)的差的选择处理进行说明。

[0503] 例如,也可以预先制作并设定如图16所示的将直接音的到来方向(θ)与反射音的到来方向(γ)的角度差(Φ)以及时间差(T)作为二维索引排列来具有的阈值数据。在此情况下,在选择处理中参照角度差(Φ)和时间差(T)。或者,也可以在选择处理中计算直接音的到来方向的角度(θ)与反射音的到来方向的角度(γ)的角度差(Φ),将计算出的角度差(Φ)用于阈值的确定。

[0504] 或者,也可以设定将该角度差(Φ)、直接音的到来方向(θ)和时间差(T)的组合,或角度差(Φ)、反射音的到来方向(γ)和时间差(T)的组合作为索引排列来具有的阈值数据。

[0505] 或者,也可以设定如图15所示的将(θ)、(γ)及(T)的值作为三维索引排列来具有

的阈值数据。

[0506] (渲染部的动作的第4变形例)

[0507] 由上述的解析部1301、选择部1302及合成部1303进行的处理例如也可以作为如在专利文献3中说明的管线(pipeline)处理进行。

[0508] 图24是表示用于渲染部1300进行管线处理的构成例的框图。

[0509] 图24的渲染部1300具备混响处理部1311、初期反射处理部1312、距离衰减处理部1313、选择部1314、生成部1315及双耳处理部1316。这些多个构成要素也可以由图7所示的渲染部1300的多个构成要素构成,也可以由图5所示的声音信号处理装置1001的多个构成要素的至少一部分构成。

[0510] 管线处理是指将用来赋予音响效果的处理分割为多个处理、将多个处理一个个地按顺序执行。在多个处理各自中,例如执行对声音信号的信号处理或在信号处理中使用的参数的生成等。

[0511] 渲染部1300也可以作为管线处理而进行混响处理、初期反射处理、距离衰减处理及双耳处理等。但是,这些处理是一例,管线处理也可以包括这些以外的处理,也可以不包括一部分的处理。例如,管线处理也可以包括衍射处理及遮挡(occlusion)处理。此外,例如混响处理在不需要的情况下也可以省略。

[0512] 此外,也可以将各处理表现为阶段。此外,也可以将各处理的结果、所生成的反射音等的声音信号表现为渲染项。管线处理中的多个阶段及它们的顺序并不限于图24所示的例子。

[0513] 这里,在选择处理中使用的参数(与直接音及反射音有关的到来路径、到来时间及音量比)也可以在用来生成渲染项的多个阶段的某一个中计算。即,在反射音的选择中使用的参数在用来生成渲染项目的管线处理的一部分中计算。另外,也可以不是由渲染部1300进行全部阶段。例如,一部分阶段也可以省略,也可以由渲染部1300以外进行。

[0514] 对在管线处理中作为阶段可以包括的混响处理、初期反射处理、距离衰减处理、选择处理、生成处理及双耳处理进行说明。在各阶段中,也可以对输入信号中包含的元数据进行解析,计算在反射音的生成中使用的参数。

[0515] 在混响处理中,混响处理部1311生成在表示混响音的声音信号或在声音信号的生成中使用的参数。混响音是指在直接音之后作为混响到达收听者的声音。作为一例,混响音是在后述的初期反射音到达收听者之后的比较后期(例如,从直接音的到来时到一百几十ms左右)阶段经过比初期反射音多的次数(例如几十次)反射而到达收听者的声音。

[0516] 混响处理部1311参照输入信号中包含的声音信号及空间信息,使用作为用来生成混响音的函数而事前准备的规定的函数计算混响音。

[0517] 混响处理部1311也可以对输入信号中包含的声音信号应用公知的混响生成方法来生成混响音。公知的混响生成方法的例子是施罗德法,但公知的混响生成方法并不限于施罗德法。此外,混响处理部1311在公知的混响生成方法的应用中使用空间信息所表示的声音再现空间的形状及音响特性。由此,混响处理部1311能够计算用来生成混响音的参数。

[0518] 在初期反射处理中,初期反射处理部1312基于空间信息计算用来生成初期反射音的参数。初期反射音是在直接音从音源对象到达收听者之后的比较初期(例如从直接音的到来时起几十ms左右)阶段经过1次以上的反射而到达收听者的反射音。

[0519] 初期反射处理部1312例如参照声音信号及元数据,计算从音源对象经反射对象反射而到达收听者的反射音的路径。例如,在路径的计算中,也可以使用三维声场(空间)的形状、三维声场的大小、构造物等的反射对象的位置及反射对象的反射率等。

[0520] 此外,初期反射处理部1312也可以还计算直接音的路径。该路径的信息也可以被用作用于初期反射处理部1312生成初期反射音的参数,也可以被用作用于选择部1314选择反射音的参数。

[0521] 在距离衰减处理中,距离衰减处理部1313基于直接音及反射音的路径的长度,计算到达收听者的直接音及反射音的音量。到达收听者的直接音及反射音的音量相对于音源的音量,与到收听者的路径的距离成比例地衰减(与距离成反比)。因而,距离衰减处理部1313能够将音源的音量除以直接音的路径的长度来计算直接音的音量,并且能够将音源的音量除以反射音的路径的长度来计算反射音的音量。

[0522] 在选择处理中,选择部1314基于在选择处理之前计算出的参数,选择生成对象反射音。在生成对象反射音的选择中,也可以使用本公开的某个选择方法。

[0523] 选择处理既可以对全部反射音执行,也可以如上所述基于评价处理仅对评价价值高的反射音执行。即,关于评价价值低的反射音,都不用进行选择处理就判定为不选择。例如,关于音量非常小的反射音,也可以看作反射音的评价价值低,判定为不选择。

[0524] 此外,例如也可以对全部反射音实施选择处理。并且,也可以判定在选择处理中选择反射音的评价价值,将判定出的评价价值低的反射音重新判定为不选择。

[0525] 选择处理和评价处理可以分别独立地执行,也可以组合执行。在将选择处理和评价处理组合执行的情况下,也可以是两个处理中的某一个先被执行。

[0526] 在生成处理中,生成部1315生成直接音及反射音。例如,生成部1315根据输入信号中包含的声音信号,基于直接音的到来时刻及到来时音量而生成直接音。此外,生成部1315关于在选择处理中被选择的反射音,根据输入信号中包含的声音信号,基于反射音的到来时刻及到来时音量来生成反射音。

[0527] 在双耳处理中,双耳处理部1316执行信号处理,以使直接音的声音信号被感知为从音源对象的方向到达收听者的声音。进而,双耳处理部1316执行信号处理,以使由选择部1314选择的反射音被感知为从反射对象到达收听者的声音。

[0528] 例如,双耳处理部1316基于声音空间中的收听者的位置及朝向,执行应用HRIR DB的处理,以使声音从音源对象的位置或障碍物对象的位置到达收听者。

[0529] 另外,HRIR(Head-Related Impulse Responses:头相关脉冲响应)是产生了一个脉冲时的响应特性。具体而言,HRIR是将头相关传递函数通过傅里叶变换从频域上的表现变换为时域上的表现而得到的响应特性,所述头相关传递函数将由包括耳廓、人头及肩在内的周边物产生的声音的变化表现为传递函数。HRIR DB是包含这样的信息的数据库。

[0530] 此外,声音空间中的收听者的位置及朝向例如是虚拟的声音空间中的虚拟的收听者的位置及朝向。也可以是虚拟的声音空间中的虚拟的收听者的位置及朝向匹配于收听者的头部的运动而变化。此外,虚拟的声音空间中的虚拟的收听者的位置及朝向也可以基于从传感器1405取得的信息来决定。

[0531] 在上述的处理中使用的程序、空间信息、HRIR DB、阈值数据或其他参数等是从声音信号处理装置1001所具备的存储器1404或声音信号处理装置1001的外部取得。

[0532] 此外,管线处理也可以包括其他处理。并且,渲染部1300也可以包括用来进行管线处理中包含的其他处理的未图示的处理部。例如,渲染部1300也可以包括衍射处理部及遮挡处理部。

[0533] 衍射处理部执行生成如下声音信号的处理,所述声音信号表示包含衍射音的声音,所述衍射音由三维声场(空间)中的处于收听者与音源对象之间的障碍物对象引起。衍射音是在音源对象与收听者之间存在障碍物对象的情况下,绕过该障碍物对象而从音源对象到达收听者的声音。

[0534] 衍射处理部例如参照声音信号及元数据,计算从音源对象绕过障碍物对象而到达收听者的衍射音的路径,基于该路径生成衍射音。在路径的计算中,也可以使用三维声场(空间)中的音源对象、收听者及障碍物对象的位置、以及障碍物对象的形状及大小等。

[0535] 遮挡处理部在障碍物对象的对面侧存在音源对象的情况下,基于空间信息及障碍物对象的材质等的信息,生成从音源对象穿过障碍物对象泄露出来的声音的声音信号。

[0536] (音源对象的例子)

[0537] 在上述中,对音源对象赋予的位置信息将虚拟空间内的“点”表示为音源对象的位置。即,在上述中,音源被定义为“点音源”。

[0538] 另一方面,虚拟空间中的音源也可以被定义为具有长度、大小及形状等的物体,即被定义为非点音源的、在空间上扩展的音源。在此情况下,收听者与音源的距离及声音的到来方向不确定。因而,关于由这样的音源引起的反射音,都不用由解析部1301进行解析、或不论解析结果如何都限定于由选择部1302选择。由此,能够避免因没有选择反射音而可能发生的音质的劣化。

[0539] 或者,也可以决定该物体的重心等代表点,假设从该代表点产生声音而应用本公开的处理。在此情况下,也可以根据音源的空间上的扩展信息来调整阈值。

[0540] (直接音及反射音的例子)

[0541] 例如,直接音是没有被反射对象反射的声音,反射音是被反射对象反射的声音。直接音也可以是从音源不经反射对象反射而到达收听者的声音,反射音也可以是从音源经反射对象反射而到达收听者的声音。

[0542] 此外,直接音及反射音分别并不限于到达收听者的声音,也可以是到达收听者之前的声音。例如,直接音也可以是从音源输出的声音、再换言之是音源的声音。

[0543] 图25是表示声音的透射及衍射的图。如图25所示,有时因在音源对象与收听者之间存在障碍物对象而直接音不到达收听者。在此情况下,也可以将从音源对象发出、透射障碍物对象并到达收听者的声音看作直接音。并且,也可以将从音源对象发出、通过障碍物对象衍射并到达收听者的声音看作反射音。

[0544] 此外,在选择处理中进行比较的两个声音并不限于基于一个音源发出的声音的直接音及反射音。例如,也可以在基于一个音源发出的声音的两个反射音之间进行比较来进行声音的选择。在此情况下,也可以将本公开中的直接音替换为先到达收听者的声音,也可以将本公开中的反射音替换为后到达收听者的声音。

[0545] (比特流的构造例)

[0546] 在比特流中,例如包含声音信号和元数据。声音信号是表现了声音的声音数据,并且表示与声音的频率及强弱有关的信息等。此外,元数据包含表示与声场的空间即声音空

间有关的空间信息。

[0547] 例如,空间信息是与收听基于声音信号的声音的收听者所处的空间有关的信息。具体而言,空间信息是用来使声像定位在声音空间(例如三维声场)中的规定的位置的、即用来使收听者感知从与规定的位置对应的方向到来的声音的、与该规定的位置(定位位置)有关的信息。在空间信息中,例如包括音源对象信息和表示收听者的位置的位置信息。

[0548] 音源对象信息是产生基于声音信号的声音的音源对象的信息。即,音源对象信息是与再现声音信号的物体(音源对象)有关的信息,并且是与配置在虚拟的声音空间中的虚拟的音源对象有关的信息。这里,虚拟的声音空间也可以对应于配置产生声音的物体的实空间,虚拟的声音空间中的音源对象也可以对应于在实空间中产生声音的物体。

[0549] 音源对象信息也可以表示配置在声音空间中的音源对象的位置、音源对象的朝向、音源对象发出的声音的指向性、音源对象是否属于生物以及音源对象是否是运动体等。例如,声音信号与由音源对象信息表示的一个以上的音源对象建立了对应。

[0550] 比特流例如具有由元数据(控制信息)和声音信号构成的数据结构。

[0551] 声音信号及元数据既可以包含在一个比特流中,也可以分别包含在多个比特流中。此外,声音信号及元数据既可以包含在一个文件中,也可以分别包含在多个文件中。

[0552] 比特流也可以按每个音源存在,也可以按每个再现时间存在。在按每个再现时间存在比特流的情况下,也可以同时将多个比特流并行地处理。

[0553] 元数据也可以对每个比特流赋予,也可以作为用来控制多个比特流的信息对多个比特流一起赋予。在此情况下,也可以多个比特流共用元数据。此外,也可以按每个再现时间赋予元数据。

[0554] 在存在多个比特流或多个文件的情况下,也可以在一个以上的比特流或一个以上的文件中包含表示关联比特流或关联文件的信息。或者,也可以在全部比特流各自或全部文件各自中包含表示关联比特流或关联文件的信息。

[0555] 这里,关联比特流或关联文件例如是指在音响处理时有可能被同时使用的比特流或文件。此外,也可以包含一起记述有表示关联比特流或关联文件的信息的比特流或文件。

[0556] 这里,表示关联比特流或关联文件的信息例如也可以是表示关联比特流或关联文件的识别符。此外,表示关联比特流或关联文件的信息例如也可以是表示关联比特流或关联文件的文件名、URL(Uniform Resource Locator:统一资源定位符)或URI(Uniform Resource Identifier:统一资源标识符)等。

[0557] 在此情况下,取得部也可以基于表示关联比特流或关联文件的信息来确定并取得关联比特流或关联文件。此外,也可以在比特流或文件中包含表示关联比特流或关联文件的信息,并且在其他比特流或其他文件中包含表示关联比特流或关联文件的信息。

[0558] 这里,包含表示关联比特流或关联文件的信息的文件例如也可以是用于内容的分发的声明文件等的控制文件。

[0559] 另外,全部元数据或一部分元数据也可以从声音信号的比特流以外取得。例如,也可以从比特流以外取得用来控制音响的元数据和用来控制影像的元数据中的某一方的元数据,也可以从比特流以外取得双方的元数据。

[0560] 此外,也可以在由立体音响再现系统1000取得的比特流中包含用来控制影像的元数据。在此情况下,立体音响再现系统1000也可以向显示图像的显示装置或再现立体影像

的立体影像再现装置输出用来控制影像的元数据。

[0561] (元数据中包含的信息的例子)

[0562] 元数据也可以是在由声音空间表现的场景的记述中使用的信息。这里,场景是指表示使用元数据由声音信号再现系统建模的声音空间中的三维影像及音响事件的全部要素的集合体的用语。

[0563] 即,元数据不仅包括用来控制音响处理的信息,也可以还包含用来控制影像处理的信息。在元数据中,既可以仅包含用来控制音响处理的信息和用来控制影像处理的信息中的仅某一方,也可以包含双方。

[0564] 立体音响再现系统1000通过使用比特流中包含的元数据以及通过追加取得的交互性的收听者的位置信息等对声音信号进行音响处理,生成虚拟的音响效果。在音响效果中,既可以进行初期反射处理、障碍物处理、衍射处理、遮断处理和混响处理,也可以使用元数据进行其他的音响处理。例如,也可以附加距离衰减效果、定位或多普勒效果等的音响效果。

[0565] 此外,也可以对元数据附加对附加音响效果的全部或一部分的开启/关闭进行切换的信息、或者针对音响效果的多个处理的优先级信息。

[0566] 此外,作为一例,元数据包含与包括音源对象及障碍物对象的声音空间有关的信息、用来使声像在声音空间内定位在规定位置(即,使收听者感知从规定方向到来的声音)的定位位置有关的信息。

[0567] 这里,障碍物对象是在音源对象发出的声音到达收听者为止的期间例如将声音遮挡或将声音反射等而可能对收听者感知的声音带来影响的对象。障碍物对象除了静止物体以外,还可以包括动物或机械等的运动体。动物也可以是人等。

[0568] 此外,在声音空间中存在多个音源对象的情况下,对任意的音源对象而言,其他的音源对象可能成为障碍物对象。即,建材或无生物等的不发出声音的对象即非发声对象、发出声音的音源对象都可能成为障碍物对象。

[0569] 在元数据中,包含表示声音空间的形状、声音空间中的障碍物对象的形状及位置、声音空间中的音源对象的形状及位置、以及声音空间中的收听者的位置及朝向的全部或一部分的信息。

[0570] 声音空间是封闭空间或开放空间的哪种都可以。此外,元数据中也可以包含表示在声音空间中能够将声音反射的障碍物对象的反射率的信息。例如,构成声音空间的边界的地板、壁或顶棚等也能够成为障碍物对象。

[0571] 反射率是反射音与入射声音的能量比,也可以按声音的每个频带设定。当然,反射率也可以不依赖于声音的频带而一律地设定。另外,在声音空间是开放空间的情况下,例如也可以使用一律地设定的衰减率、衍射音及初期反射音等的参数。

[0572] 元数据也可以作为与障碍物对象或音源对象有关的参数而包含反射率以外的信息。例如,元数据也可以作为与音源对象及非发声对象双方有关的参数而包含与对象的材料有关的信息。具体而言,元数据也可以包含扩散率、透射率及吸声音率等信息。

[0573] 与音源对象有关的信息也可以包含音量、放射特性(指向性)、再现条件、一个对象中的音源的数量及种类、以及表示对象中的音源区域的信息等。再现条件中例如也可以设定是持续地流出的声音还是事件触发的声音。对象中的音源区域既可以根据收听者的位置

和对象的位置的相对关系来设定,也可以将对象用于基准来设定。

[0574] 例如,在以收听者的位置与对象的位置的相对关系来设定音源区域的情况下,从收听者看时,能够使收听者从对象的右侧感知声音A,从对象的左侧感知声音B。

[0575] 此外,在将对象用于基准而设定音源区域的情况下,能够将对象用于基准,固定从对象的哪个区域发出哪个声音。例如,在收听者从正面看对象的情况下,能够使收听者从对象的右侧感知高音,从对象的左侧感知低音。并且,在收听者从背面看对象的情况下,能够使收听者从对象的右侧感知低音、从对象的左侧感知高音。

[0576] 与空间有关的元数据也可以包含初期反射音为止的时间、混响时间及直接音与扩散音的比率等。在直接音与扩散音的比率是零的情况下,能够使收听者仅感知直接音。

[0577] (补充)

[0578] 另外,基于本公开掌握的形态并不限于实施方式,也可以进行各种变更而实施。

[0579] 例如,也可以将在实施方式中由特定的构成要素执行的处理代替特定的构成要素而由其他的构成要素执行。此外,也可以将多个处理的顺序变更,也可以将多个处理并行地执行。

[0580] 此外,在说明中使用的第1及第2等的序数也可以适当替换,也可以去除,也可以新赋予。这些序数并不一定对应于有意义的顺序,也可以用于要素的识别。

[0581] 此外,例如在对于阈值的比较中,也可以将阈值以上和大于阈值相互替换。同样,也可以将阈值以下和小于阈值相互替换。此外,例如有将时间及时刻相互替换。

[0582] 此外,在从多个声音中选择一个以上的处理对象声音的处理中,如果不存在满足条件的声音则也可以将哪个声音都不选择为处理对象声音。即,在从多个声音中选择一个以上的处理对象声音的处理中,也可以包括不选择处理对象声音的情形。

[0583] 此外,例如第1要素、第2要素及第3要素中的至少一个这样的表现能够对应于第1要素、第2要素、第3要素或这些的任意的组合。

[0584] 此外,例如在实施方式中说明了将基于本公开掌握的形态作为音响处理装置、编码装置或解码装置实施的情况。但是,基于本公开掌握的形态并不限于这些,也可以作为用来执行音响处理方法、编码方法或解码方法的软件实施。

[0585] 例如,也可以将用来执行上述的音响处理方法、编码方法或解码方法的程序预先存储在ROM中。并且,也可以是CPU按照该程序动作。

[0586] 此外,也可以将用来执行上述的音响处理方法、编码方法或解码方法的程序存储在计算机可读的记录介质中。并且,计算机也可以将存储在记录介质中的程序记录到计算机的RAM中并按照该程序动作。

[0587] 并且,上述的各构成要素典型地也可以作为具有输入端子及输出端子的集成电路即LSI实现。它们既可以单独地形成1个芯片,也可以以包含实施方式的全部构成要素或一部分构成要素的方式形成1个芯片。LSI根据集成度的差异,也可以表现为IC、系统LSI、超级LSI或超大规模LSI。

[0588] 此外,并不限于LSI,也可以使用专用电路或通用处理器。此外,也可以使用在LSI制造后能够编程的FPGA、或者能够进行LSI内部的电路单元的连接或设定的重构的可重构处理器。进而,如果因半导体技术的进步或派生的其他技术而出现替代LSI的集成电路化的技术,则当然也可以使用该技术进行构成要素的集成化。有可能是生物技术的应用等。

[0589] 此外,也可以是FPGA或CPU等通过无线通信或有线通信来下载用于实现在本公开中所说明的音响处理方法、编码方法或解码方法的软件的全部或一部分。进而,也可以通过无线通信或有线通信来下载用于更新的软件的全部或一部分。并且,也可以是通过由FPGA或CPU等将所下载的软件存储到存储器中并基于所存储的软件而动作,来执行在本公开中所说明的数字信号处理。

[0590] 此时,具备FPGA或CPU等的设备也可以通过无线或有线与信号处理装置连接,也可以经由网络与信号处理服务器连接。并且,该设备和信号处理装置或信号处理服务器也可以进行在本公开中所说明的音响处理方法、编码方法或解码方法。

[0591] 例如,本公开的音响处理装置、编码装置或解码装置也可以具备FPGA或CPU等。进而,音响处理装置、编码装置或解码装置也可以具备用来从外部获得用来使FPGA或CPU等动作的软件的接口、以及用来存储所获得的软件的存储器。并且,FPGA或CPU等也可以通过基于所存储的软件动作,来执行在本公开中所说明的信号处理。

[0592] 也可以是服务器提供与本公开的音响处理、编码处理或解码处理有关的软件。并且,也可以是终端或设备通过安装该软件而作为在本公开中记载的音响处理装置、编码装置或解码装置动作。另外,也可以是终端或设备经由网络与服务器连接而安装软件。

[0593] 此外,也可以是与终端或设备不同的其他装置经由网络与服务器连接而取得软件的安装用的数据,通过由该其他装置对终端或设备提供软件的安装用的数据而在终端或设备中安装软件。另外,软件的例子也可以是用来使终端或设备执行使用实施方式所说明的音响处理方法的VR用软件或AR用软件。

[0594] 另外,在上述实施方式中,各构成要素也可以由专用的硬件构成,或通过执行适合于各构成要素的软件程序来实现。各构成要素也可以通过由CPU或处理器等的程序执行部将记录在硬盘或半导体存储器等的记录介质中的软件程序读出并执行来实现。

[0595] 以上,基于实施方式对有关一个以上的形态的装置等进行了说明,但基于本公开掌握的形态并不限于实施方式。只要不脱离本公开的主旨,也可以通过对实施方式施以本领域技术人员想到的各种变形而得到的形态,以及将不同变形例中的构成要素组合而构建的形态也包含在一个以上的形态的范围内。

[0596] (附记)

[0597] 通过以上的实施方式的记载,公开下述的技术。

[0598] (技术1) 一种音响处理装置,具备电路和存储器;所述电路使用所述存储器,取得与声音空间有关的声音空间信息;基于所述声音空间信息,取得与在所述声音空间中从音源产生的第1声音有关的特性;基于与所述第1声音有关的特性,控制是否选择在所述声音空间中对应于所述第1声音产生的第2声音。

[0599] (技术2) 如技术1所述的音响处理装置,所述第1声音是直接音;所述第2声音是反射音。

[0600] (技术3) 如技术2所述的音响处理装置,与所述第1声音有关的特性是所述直接音的音量与所述反射音的音量的音量比;所述电路基于所述声音空间信息,计算所述音量比;基于所述音量比,控制是否选择所述反射音。

[0601] (技术4) 如技术3所述的音响处理装置,在选择了所述反射音的情况下,所述电路通过对所述反射音及所述直接音应用双耳处理,生成分别到达收听者的两耳的声音。

[0602] (技术5) 如技术3或4所述的音响处理装置,所述电路基于所述声音空间信息,计算所述直接音的结束时刻与所述反射音的到来时刻的时间差;基于所述时间差及所述音量比,控制是否选择所述反射音。

[0603] (技术6) 如技术5所述的音响处理装置,所述电路在所述音量比是阈值以上的情况下,选择所述反射音;在所述时间差是第1值的情况下被用作所述阈值的第1阈值大于在所述时间差是比所述第1值大的第2值的情况下被用作所述阈值的第2阈值。

[0604] (技术7) 如技术3或4所述的音响处理装置,所述电路基于所述声音空间信息,计算所述直接音的到来时刻与所述反射音的到来时刻的时间差;基于所述时间差及所述音量比,控制是否选择所述反射音。

[0605] (技术8) 如技术7所述的音响处理装置,所述电路在所述音量比是阈值以上的情况下,选择所述反射音;在所述时间差是第1值的情况下被用作所述阈值的第1阈值大于在所述时间差是比所述第1值大的第2值的情况下被用作所述阈值的第2阈值。

[0606] (技术9) 如技术6或8所述的音响处理装置,所述电路基于所述直接音的到来方向和所述反射音的到来方向,调整所述阈值。

[0607] (技术10) 如技术1~9中任一项所述的音响处理装置,在没有选择所述第2声音的情况下,所述电路基于所述第2声音的音量将所述第1声音的音量进行修正。

[0608] (技术11) 如技术1~9中任一项所述的音响处理装置,在没有选择所述第2声音的情况下,所述电路将所述第2声音合成到所述第1声音中。

[0609] (技术12) 如技术3~9中任一项所述的音响处理装置,所述音量比是第1时刻的所述直接音的音量与不同于所述第1时刻的第2时刻的所述反射音的音量的音量比。

[0610] (技术13) 如技术1或2所述的音响处理装置,所述电路基于与所述第1声音有关的特性设定阈值,基于所述阈值控制是否选择所述第2声音。

[0611] (技术14) 如技术1、2及13中任一项所述的音响处理装置,与所述第1声音有关的特性是所述音源的音量、所述音源的视觉性及所述音源的定位性中的某一个或某两个以上的组合。

[0612] (技术15) 如技术1、2及13中任一项所述的音响处理装置,与所述第1声音有关的特性是所述第1声音的频率特性。

[0613] (技术16) 如技术1、2及13中任一项所述的音响处理装置,与所述第1声音有关的特性是表示所述第1声音的振幅的断续性的特性。

[0614] (技术17) 如技术1、2、13及16中任一项所述的音响处理装置,与所述第1声音有关的特性是表示所述第1声音的有声部的持续时间或所述第1声音的无声部的持续时间的特性。

[0615] (技术18) 如技术1、2、13、16及17中任一项所述的音响处理装置,与所述第1声音有关的特性是将所述第1声音的有声部的持续时间和所述第1声音的无声部的持续时间以时间序列表示的特性。

[0616] (技术19) 如技术1、2、13及15中任一项所述的音响处理装置,与所述第1声音有关的特性是表示所述第1声音的频率特性的变动的特性。

[0617] (技术20) 如技术1、2、13、15及19中任一项所述的音响处理装置,与所述第1声音有关的特性是表示所述第1声音的频率特性的平稳性的特性。

[0618] (技术21) 如技术1、2及13~20中任一项所述的音响处理装置,从比特流取得与前述第1声音有关的特性。

[0619] (技术22) 如技术1、2及13~21中任一项所述的音响处理装置,所述电路计算与前述第2声音有关的特性;基于与前述第1声音有关的特性以及与前述第2声音有关的特性,控制是否选择所述第2声音。

[0620] (技术23) 如技术22所述的音响处理装置,所述电路取得表示与是否能听到声音的边界对应的音量的阈值;基于与前述第1声音有关的特性、与前述第2声音有关的特性和所述阈值,控制是否选择所述第2声音。

[0621] (技术24) 如技术22或23所述的音响处理装置,与前述第2声音有关的特性是所述第2声音的音量。

[0622] (技术25) 如技术1~24中任一项所述的音响处理装置,所述声音空间信息包含所述声音空间中的收听者的位置的信息;所述第2声音是在所述声音空间中对应于所述第1声音产生的多个第2声音各自;所述电路基于与前述第1声音有关的特性来控制是否选择所述多个第2声音各自,由此从所述第1声音及所述多个第2声音中选择被应用双耳处理的一个以上的处理对象声音。

[0623] (技术26) 如技术1~25中任一项所述的音响处理装置,取得与前述第1声音有关的特性的定时是所述声音空间的制作时、所述声音空间的处理开始时、以及所述声音空间的处理中的信息更新线程发生时中的至少一个。

[0624] (技术27) 如技术1~26中任一项所述的音响处理装置,在所述声音空间的处理开始后定期地取得与前述第1声音有关的特性。

[0625] (技术28) 如技术1、2及25~27中任一项所述的音响处理装置,与前述第1声音有关的特性是所述第1声音的音量;所述电路基于所述第1声音的音量,计算所述第2声音的评价值;基于所述评价值,控制是否选择所述第2声音。

[0626] (技术29) 如技术28所述的音响处理装置,所述第1声音的音量具有转变。

[0627] (技术30) 如技术28或29所述的音响处理装置,所述电路计算所述评价值,以使得所述第1声音的音量越大则所述第2声音越容易被选择。

[0628] (技术31) 如技术1~30中任一项所述的音响处理装置,所述声音空间信息是包括所述声音空间中的所述音源的信息以及所述声音空间中的收听者的位置的信息的场景信息;所述第2声音是在所述声音空间中对应于所述第1声音产生的多个第2声音各自;所述电路取得所述第1声音的信号;基于所述场景信息和所述第1声音的信号,计算所述多个第2声音;从所述音源的信息取得与前述第1声音有关的特性;基于与前述第1声音有关的特性来控制是否将所述多个第2声音各自选择为不被应用双耳处理的声音,由此从所述多个第2声音中选择不被应用所述双耳处理的一个以上的第2声音。

[0629] (技术32) 如技术31所述的音响处理装置,所述场景信息基于输入信息来更新;根据所述场景信息的更新,取得与前述第1声音有关的特性。

[0630] (技术33) 如技术31或32所述的音响处理装置,从比特流所包含的元数据中取得所述场景信息及与前述第1声音有关的特性。

[0631] (技术34) 如技术1、2、13、16~18、25~27及31~33中任一项所述的音响处理装置,与前述第1声音有关的特性是以时间序列表示分别由所述第1声音的振幅值为代表振幅值

的持续时间和该持续时间中的所述代表振幅值的组构成的多个组的特性。

[0632] (技术35) 如技术34所述的音响处理装置,所述代表振幅值是所述第1声音的音量相对于预先设定的基准音量的比值。

[0633] (技术36) 如技术1、2、13、15、19及20中任一项所述的音响处理装置,与所述第1声音有关的特性是表示频率特性的变动量比预先设定的阈值低的状态持续的持续时间的特性。

[0634] (技术37) 如技术1、2、13、15、20及36中任一项所述的音响处理装置,与所述第1声音有关的特性是以时间序列表示分别由频率特性的变动量比预先设定的阈值低的状态持续的持续时间和该持续时间中的频率特性的组构成的多个组的特性。

[0635] (技术38) 如技术1、2、13~24及34~37中任一项所述的音响处理装置,所述电路取得表示与是否能听到声音的边界对应的音量的阈值,基于与所述第1声音有关的特性,计算所述第2声音的音量,在所述第2声音的音量大于所述阈值的情况下,选择所述第2声音。

[0636] (技术39) 如技术1、2、13~20及31~38中任一项所述的音响处理装置,所述声音空间信息是包括所述声音空间中的所述音源的信息、以及所述声音空间中的收听者的位置的信息的场景信息,所述第2声音是在所述声音空间中对应于所述第1声音产生的多个第2声音各自,所述电路取得所述第1声音的信号,基于所述场景信息和所述第1声音的信号,计算所述多个第2声音,从所述音源的信息取得与所述第1声音有关的特性,基于与所述第1声音有关的特性,控制是否将所述多个第2声音分别选择为被应用双耳处理的声音,由此从所述第1声音及所述多个第2声音中选择被应用所述双耳处理的一个以上的处理对象声音,所述场景信息基于输入信息来更新,与所述第1声音有关的特性是根据所述场景信息的更新来取得,所述场景信息的更新以比对所述一个以上的处理对象声音应用所述双耳处理的频度低的频度实施。

[0637] (技术40) 一种音响处理方法,包括:取得与声音空间有关的声音空间信息的步骤;基于所述声音空间信息,取得与在所述声音空间中从音源产生的第1声音有关的特性的步骤;以及基于与所述第1声音有关的特性,控制是否选择在所述声音空间中对应于所述第1声音产生的第2声音的步骤。

[0638] (技术41) 一种程序,用来使计算机执行技术40所述的音响处理方法。

[0639] 工业实用性

[0640] 本公开例如包括能够在音响处理装置、编码装置、解码装置或具备这些某种装置的终端或设备中应用的形态。

[0641] 标号说明

[0642] 1000 立体音响再现系统

[0643] 1001 声音信号处理装置(音响处理装置)

[0644] 1002 声音提示装置

[0645] 1100、1120、1500 编码装置

[0646] 1101、1113 输入数据

[0647] 1102 编码器

[0648] 1103 编码数据

[0649] 1104、1114、1404、1503 存储器

- [0650] 1110、1130 解码装置
- [0651] 1111 声音信号
- [0652] 1112、1200、1210 解码器
- [0653] 1121 发送部
- [0654] 1122 发送信号
- [0655] 1131 接收部
- [0656] 1132 接收信号
- [0657] 1201、1211 空间信息管理部
- [0658] 1202 声音数据解码器
- [0659] 1203、1213、1300 渲染部
- [0660] 1301 解析部
- [0661] 1302、1314 选择部
- [0662] 1303 合成部
- [0663] 1304 阈值调整部
- [0664] 1311 混响处理部
- [0665] 1312 初期反射处理部
- [0666] 1313 距离衰减处理部
- [0667] 1315 生成部
- [0668] 1316 双耳处理部
- [0669] 1401 扬声器
- [0670] 1402、1501 处理器
- [0671] 1403、1502 通信IF
- [0672] 1405 传感器

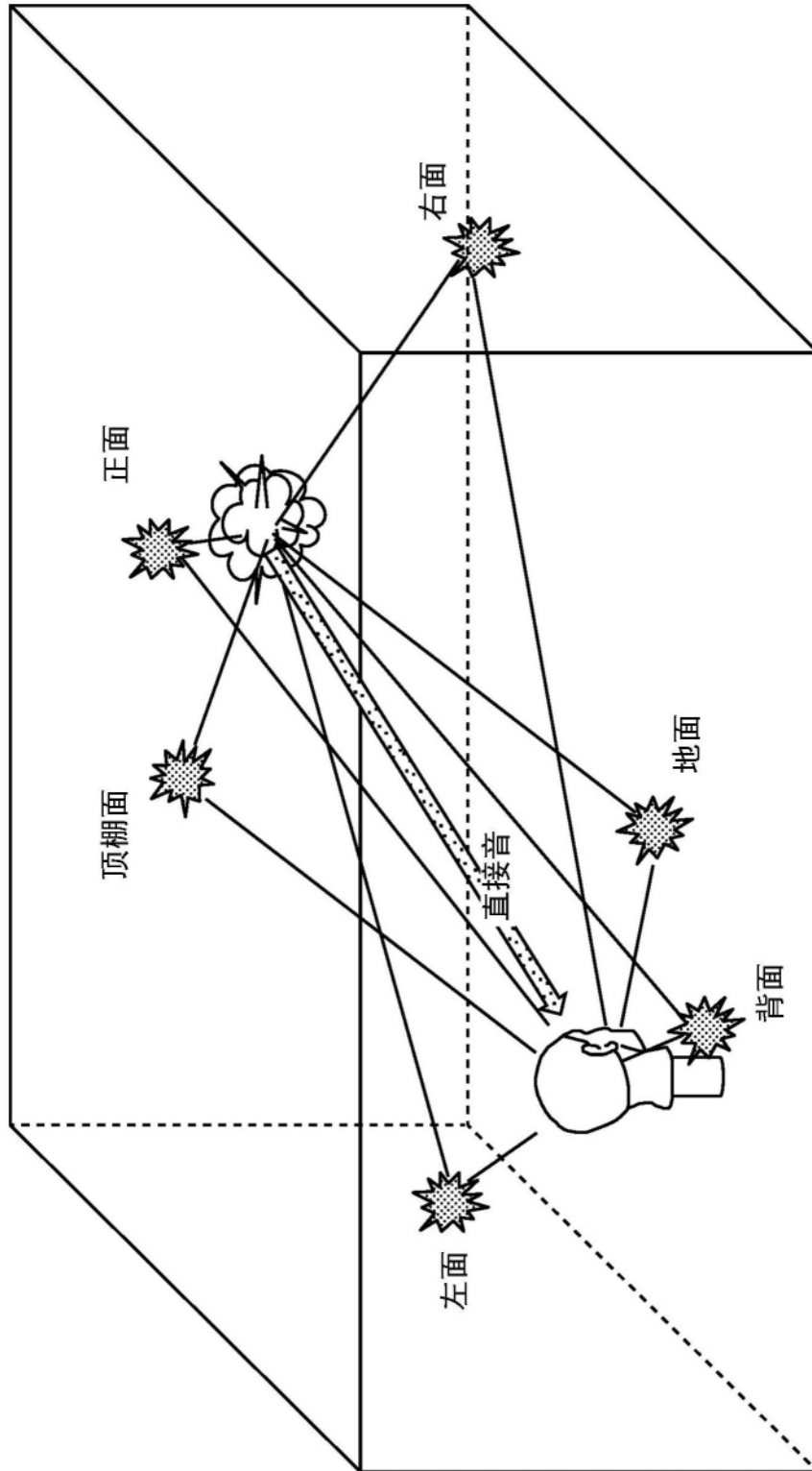


图1

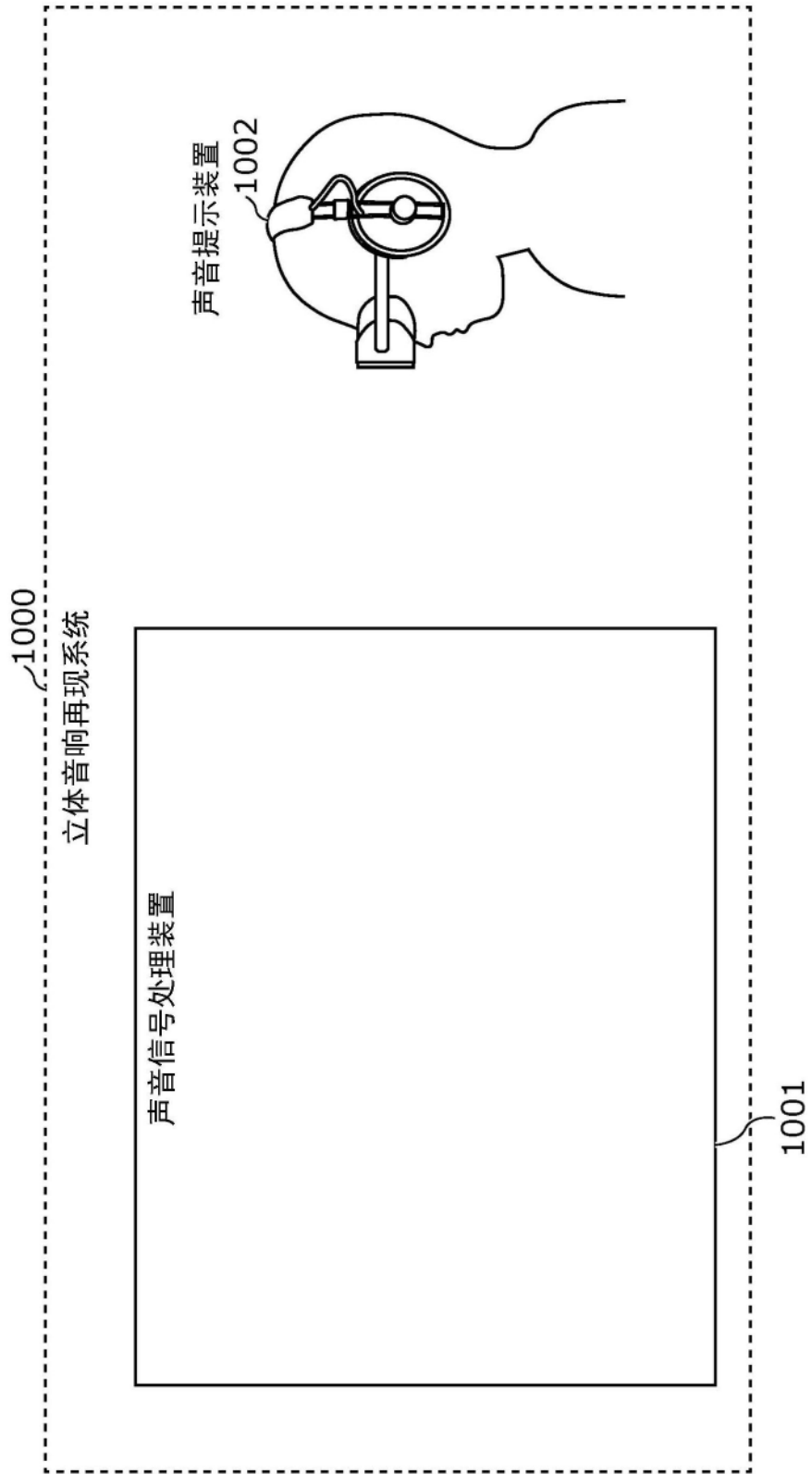


图2

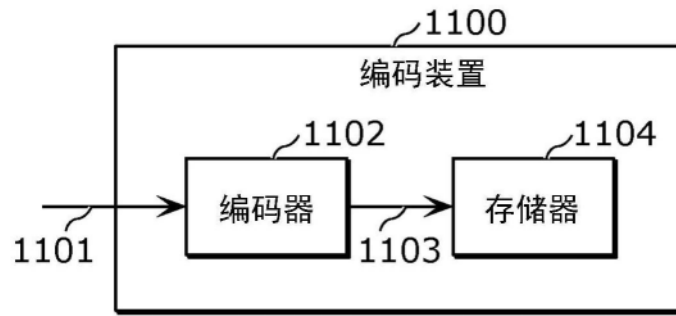


图3A

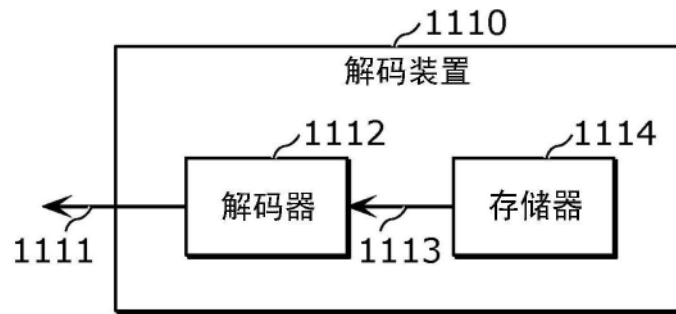


图3B

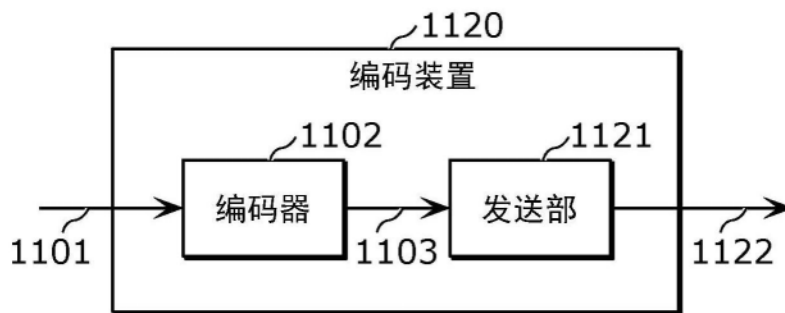


图3C

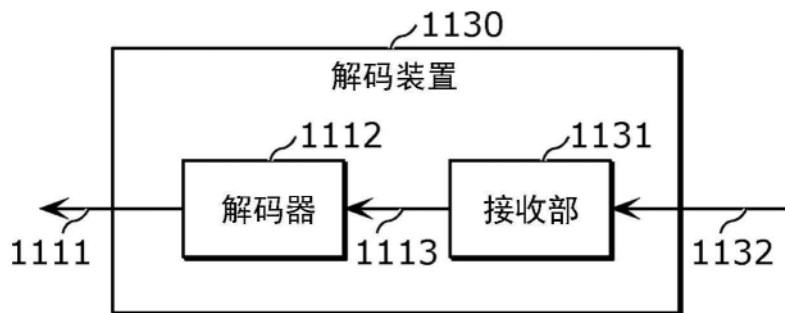


图3D

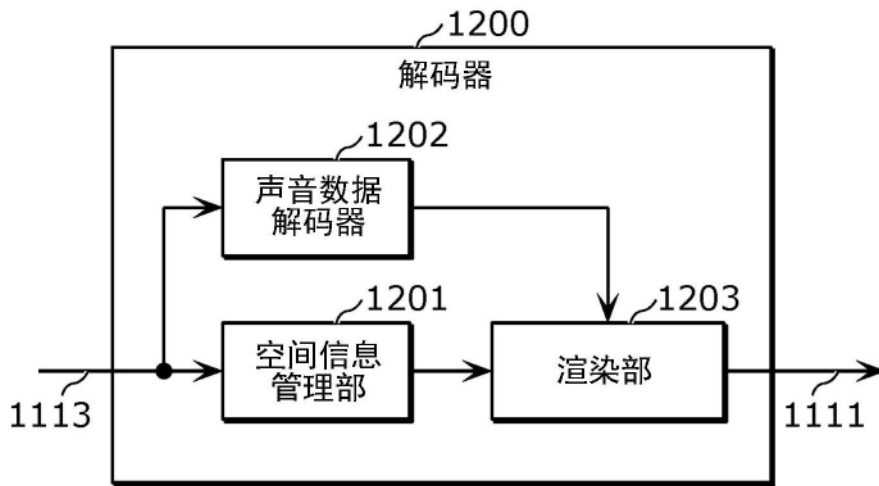


图4A

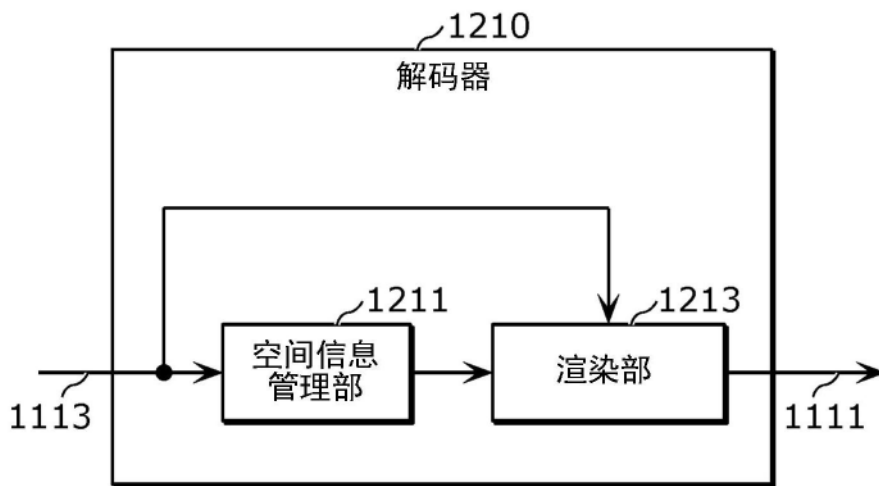


图4B

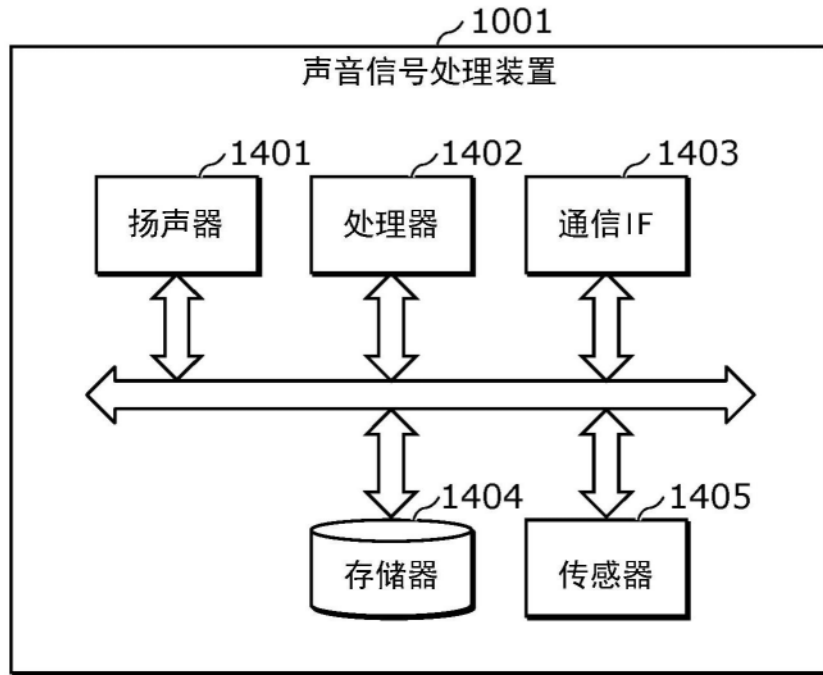


图5

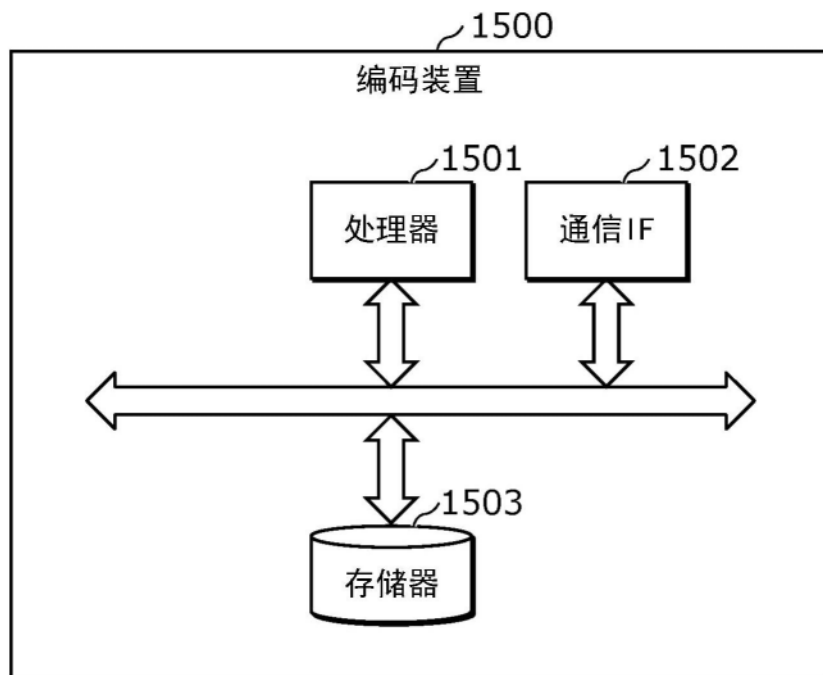


图6

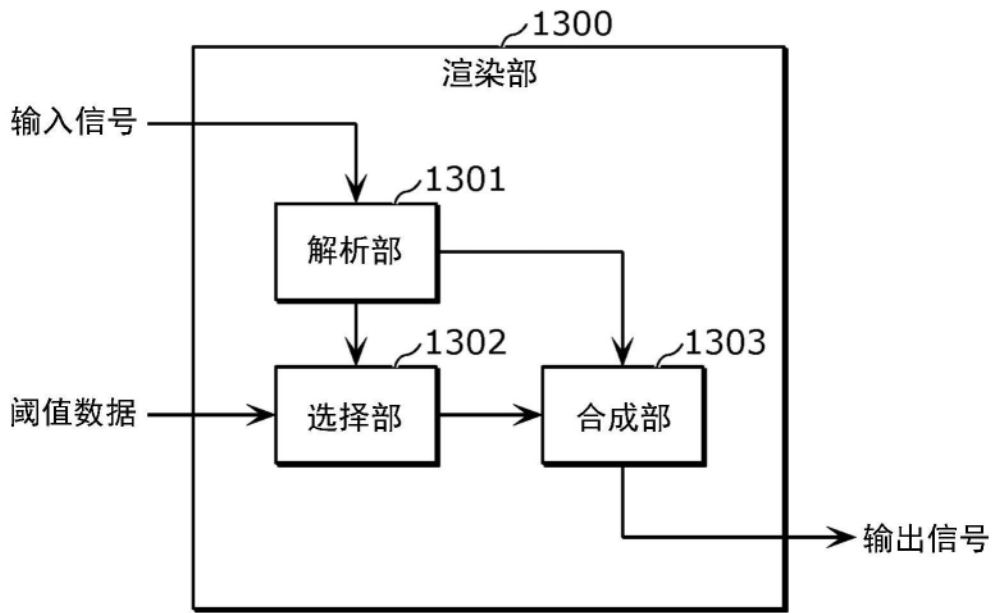


图7

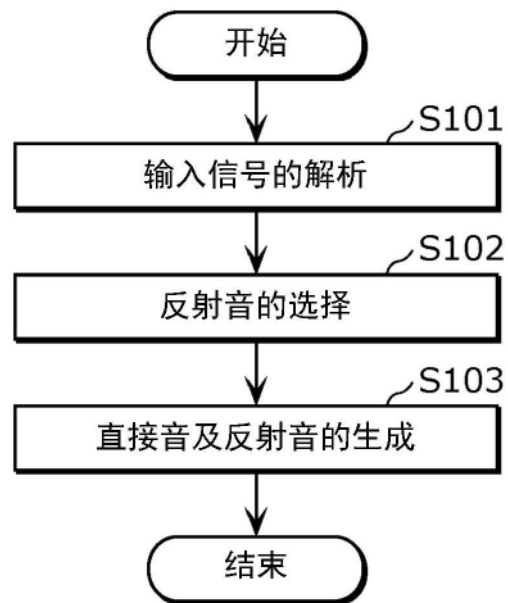


图8

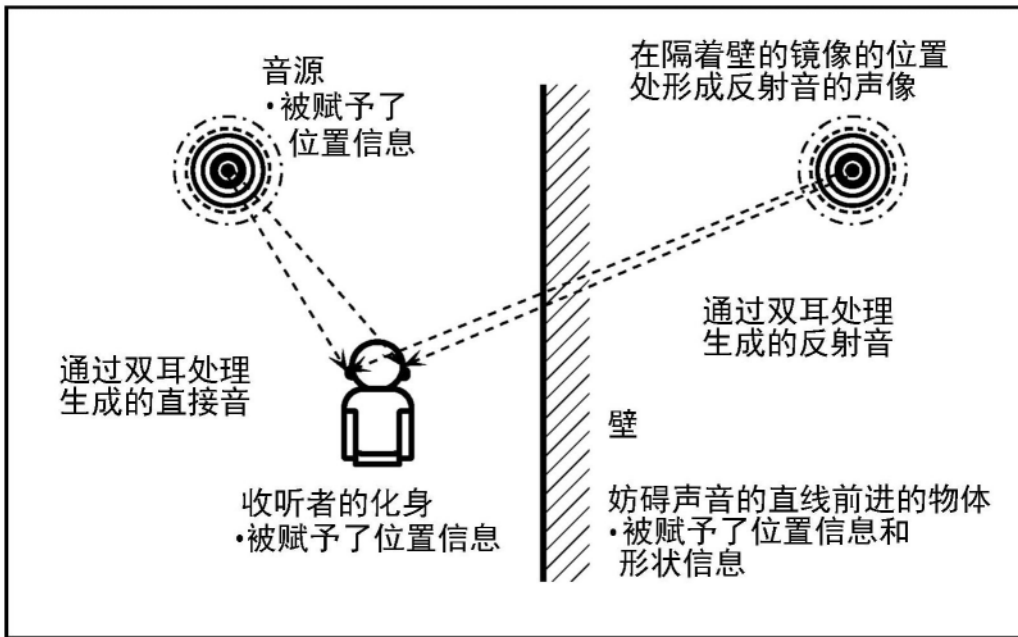


图9

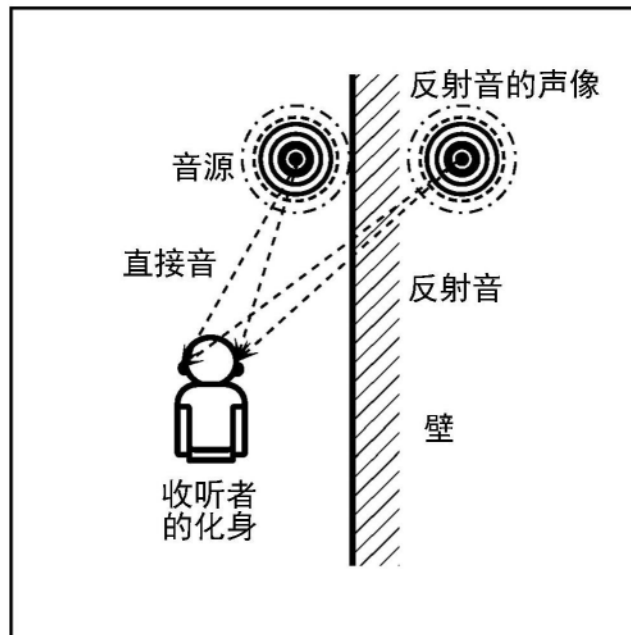
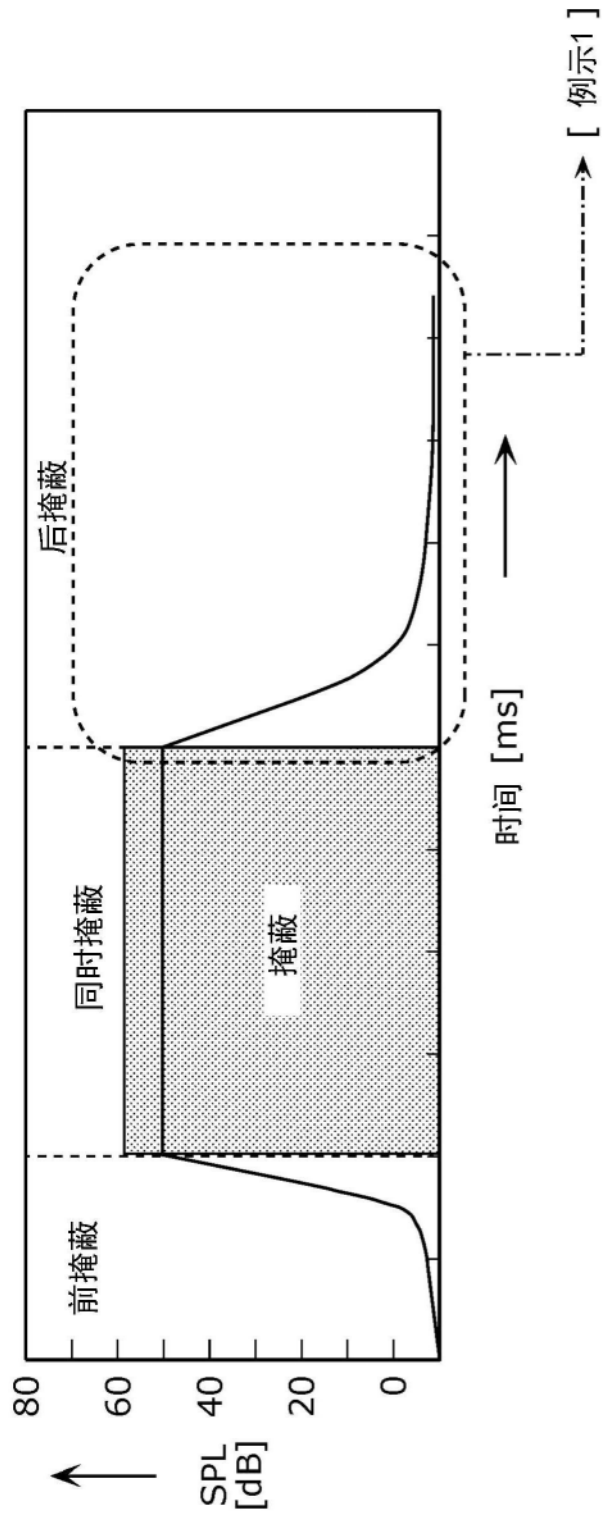


图10

T	閾值
5ms	-12.5
10ms	-15.8
20ms	-22.5
30ms	-30.5
40ms	-38.2
50ms	-45.1
60ms	-53.0

图11



将经时掩蔽中的后掩蔽的趋势作为阈值保持在存储器中

图12A

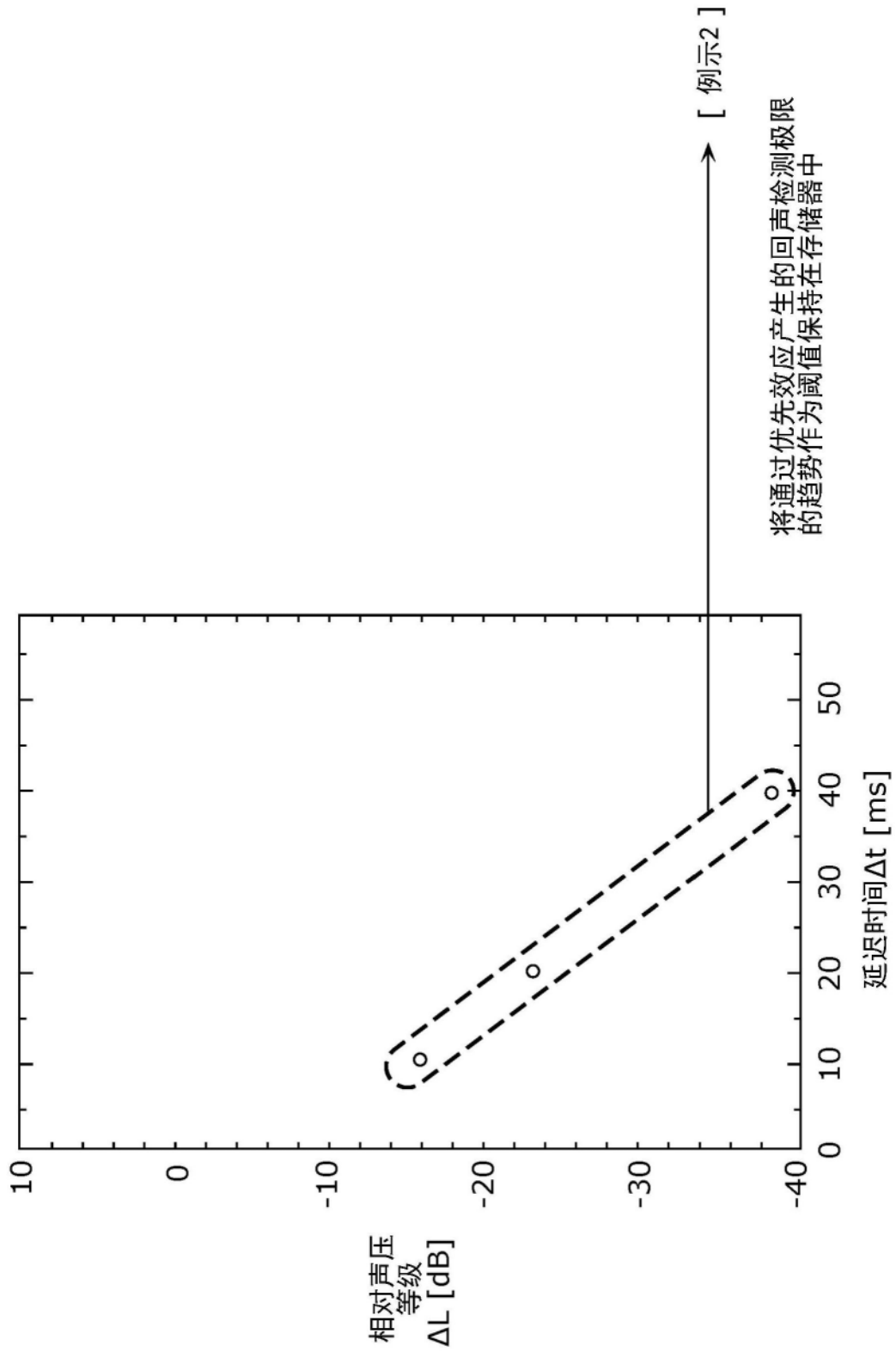


图12B

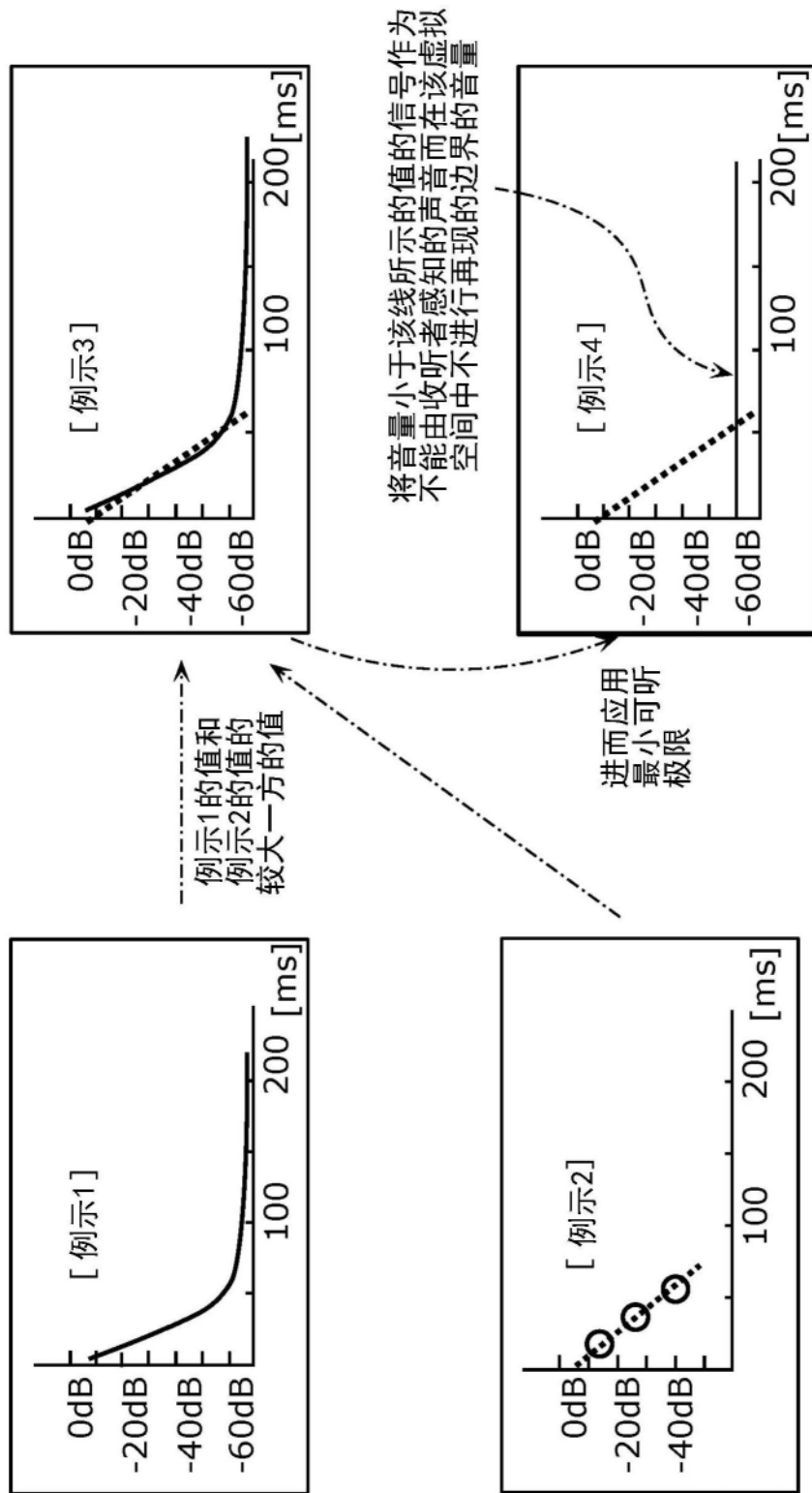


图12C

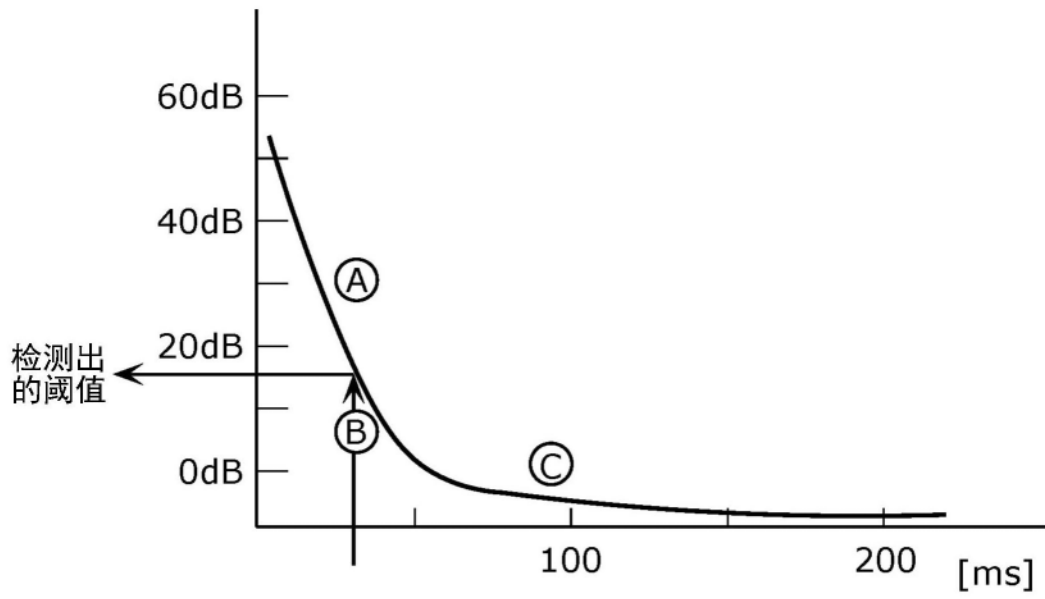


图13

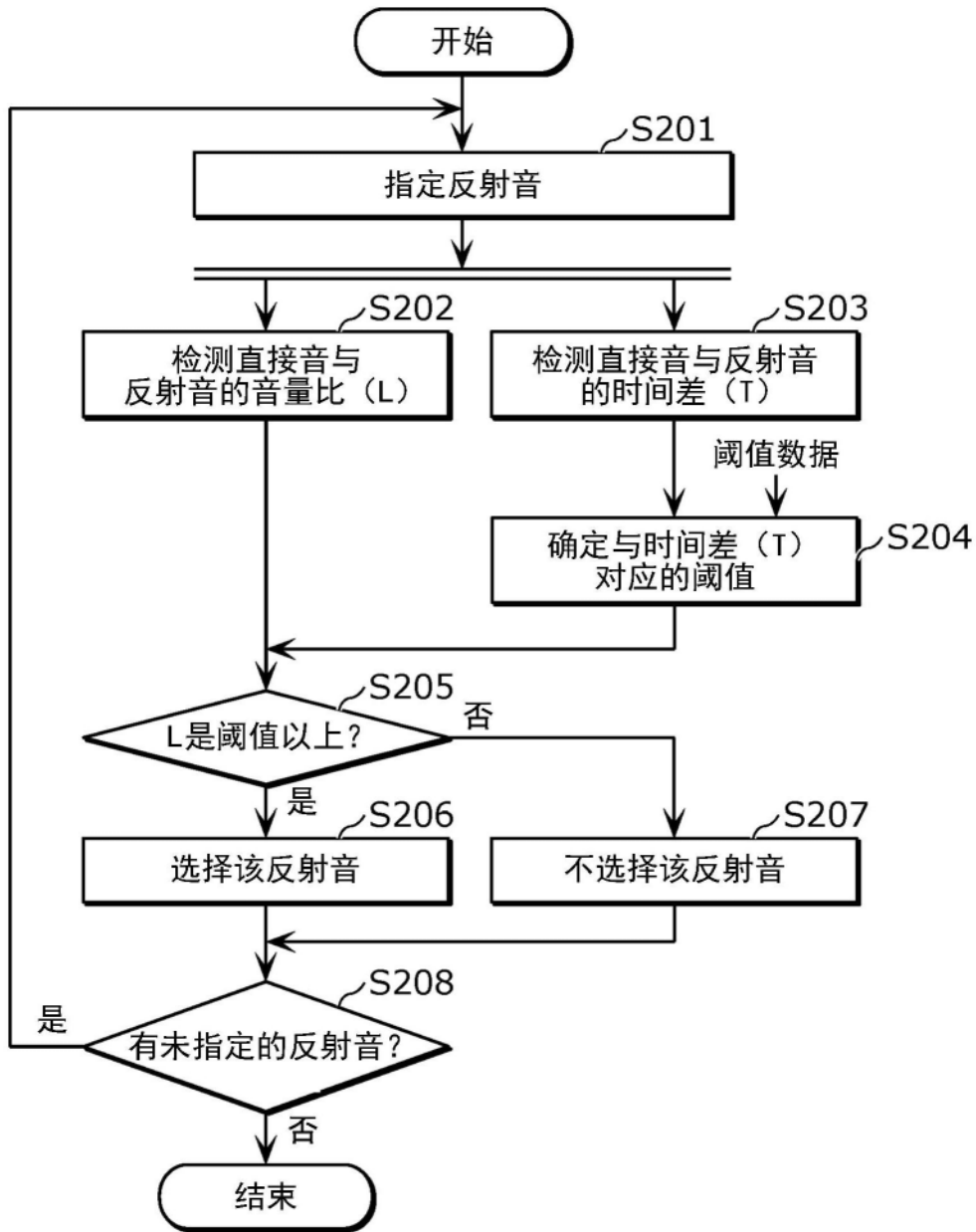


图14

直接音的方向 (θ)	反射音的方向 (γ)	T	阈值	
0 度	0 度	5ms	-12.5	
		10ms	-15.8	
		20ms	-22.5	
		30ms	-30.5	
		40ms	-38.2	
		50ms	-45.1	
		60ms	-53.0	
	45 度	5ms	..	
		10ms	..	
	~~~~~			
	45 度	315 度	5ms	..
			10ms	..
			20ms	..
			30ms	..
40ms			..	
50ms			..	
60ms			..	
45 度	0 度	5ms	..	

图15

角度差 ( $\Phi$ )	T	阈值
0 度	5ms	-12.5
	10ms	-15.8
	20ms	-22.5
	30ms	-30.5
	40ms	-38.2
	50ms	-45.1
	60ms	-53.0
45 度	5ms	• •
	10ms	• •
	20ms	• •
	30ms	• •
	40ms	• •
	50ms	• •
	60ms	• •
315 度	5ms	• •
	10ms	• •
	20ms	• •
	30ms	• •
	40ms	• •
	50ms	• •
	60ms	• •

图16

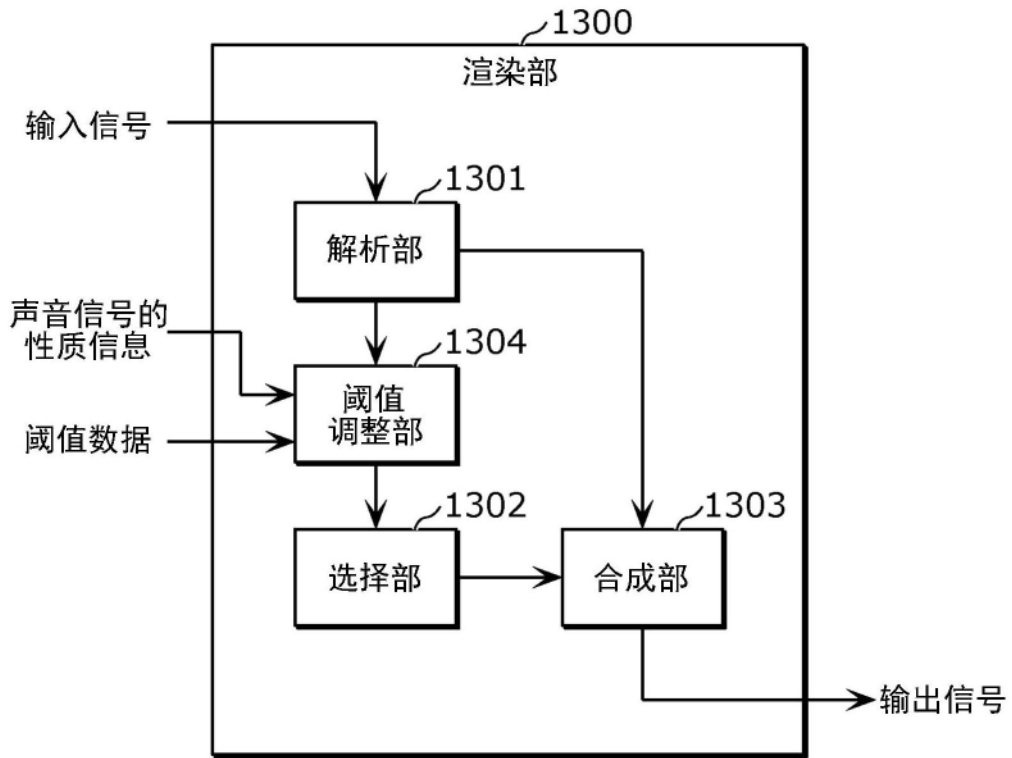


图17

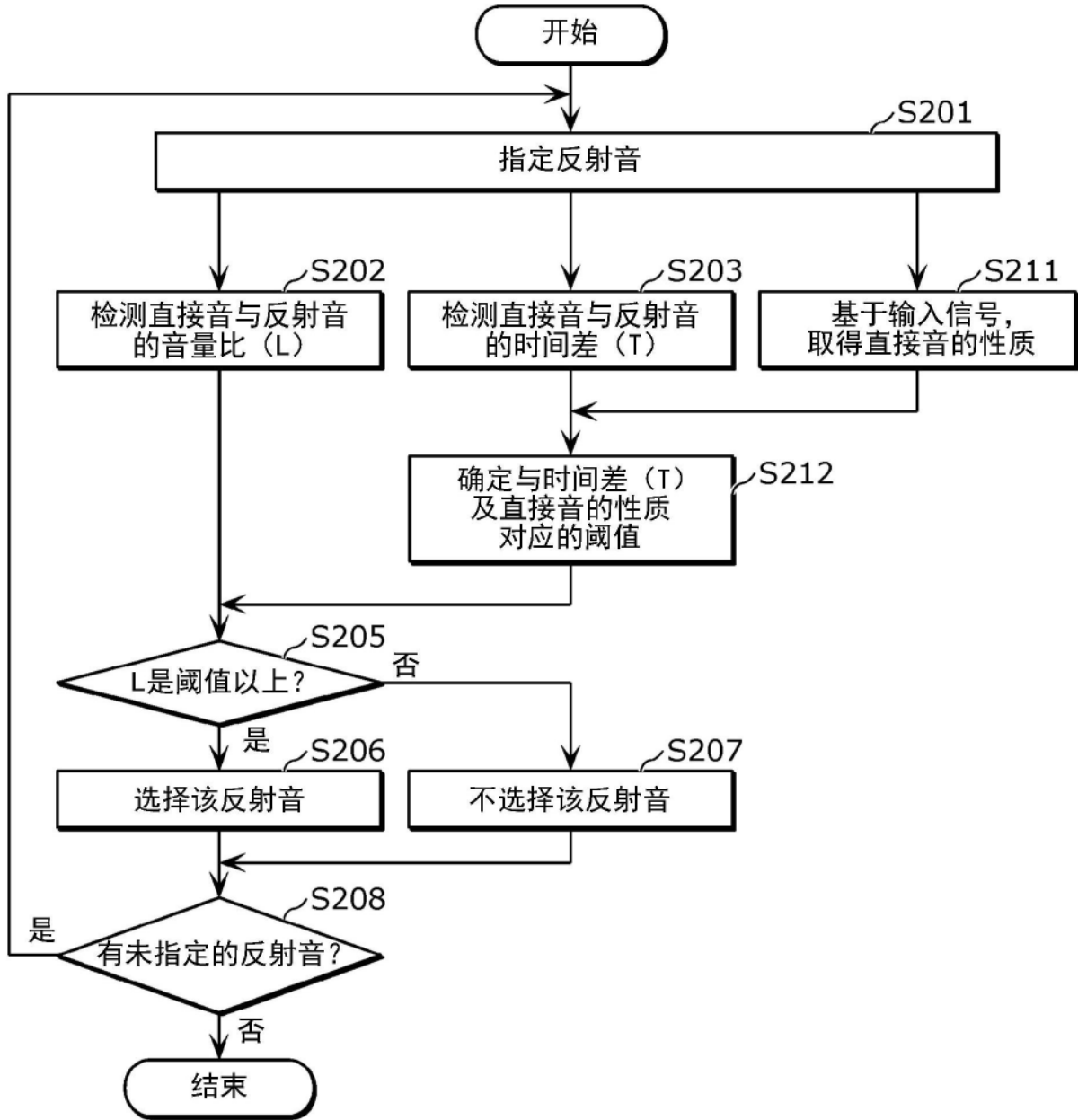


图18

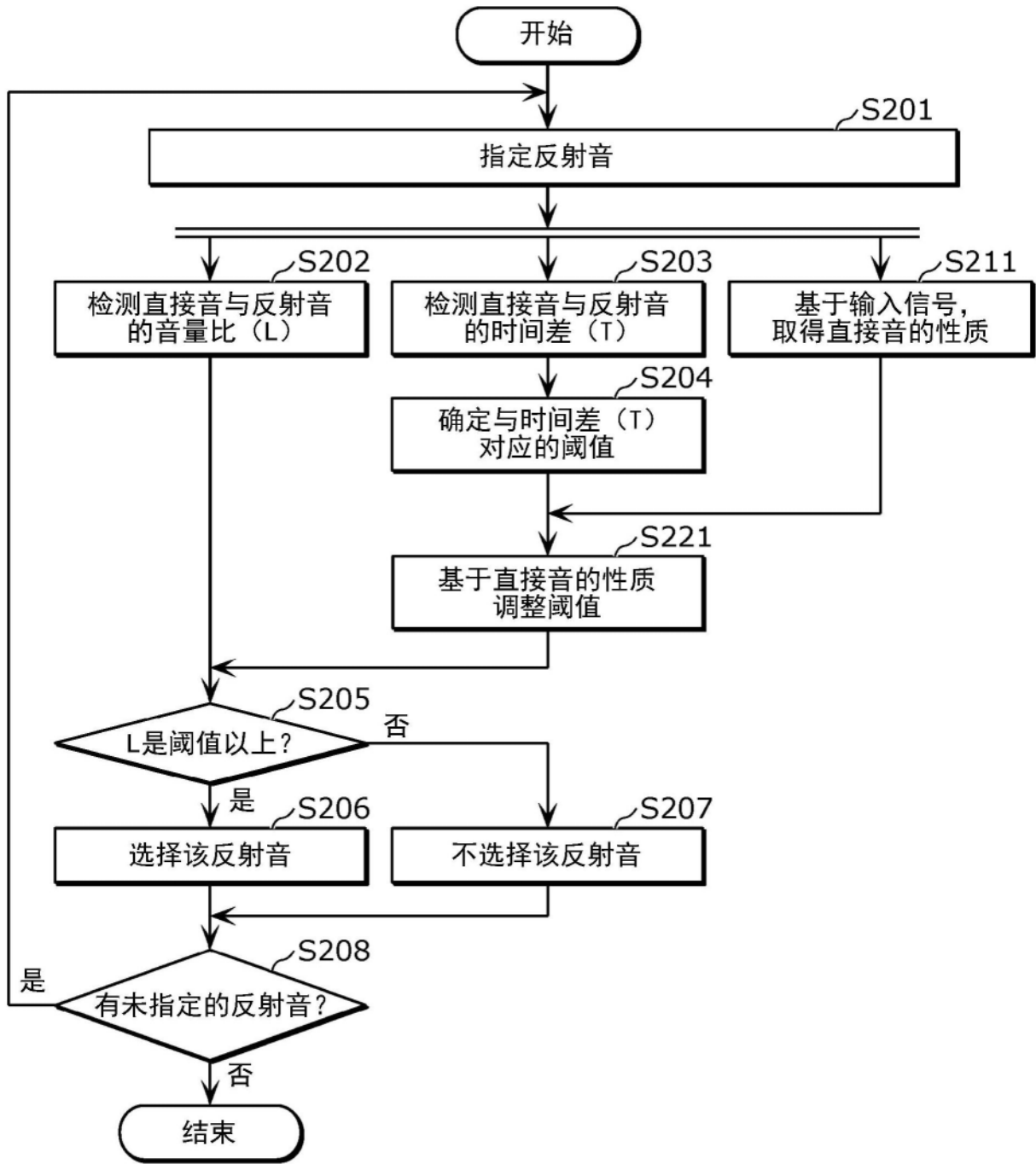


图19

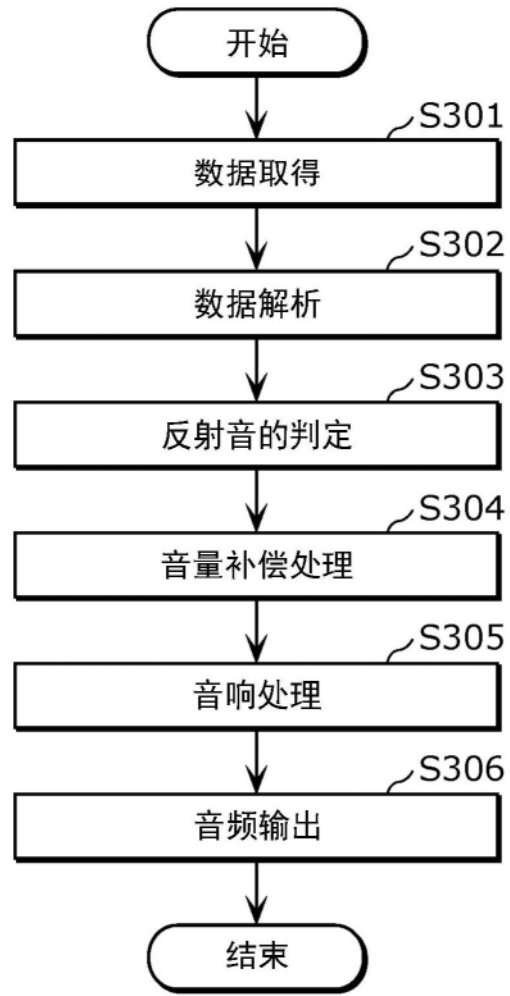


图20

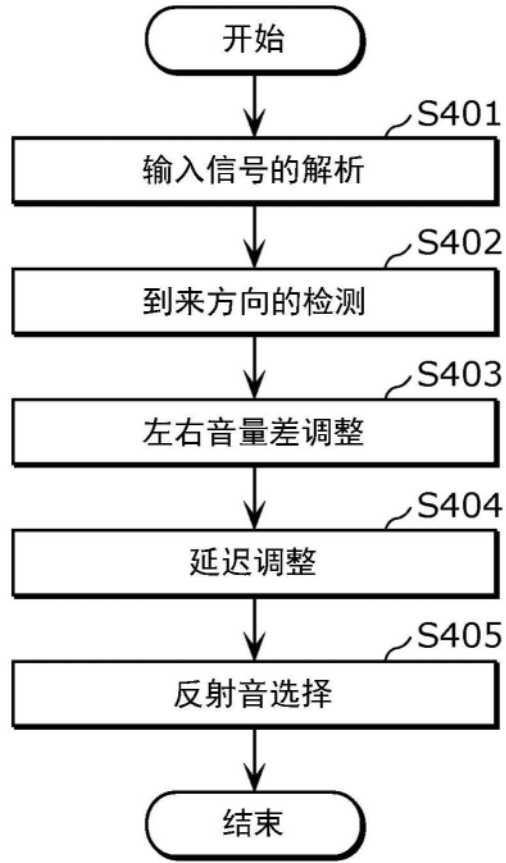


图21

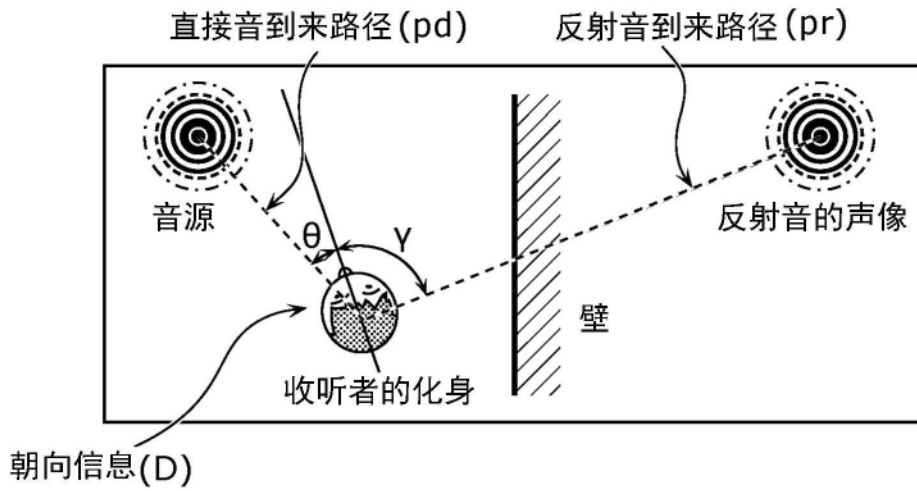


图22

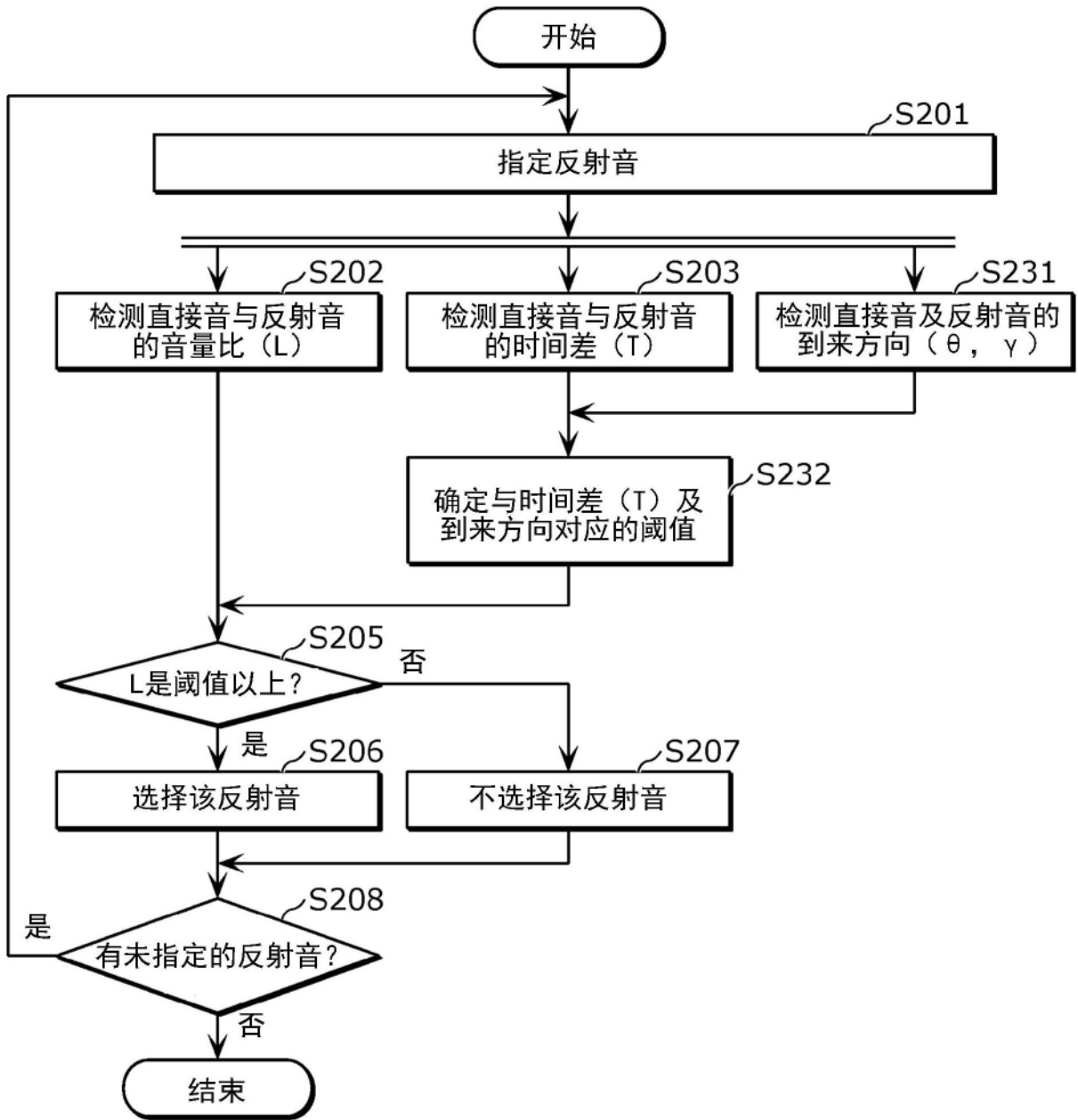


图23

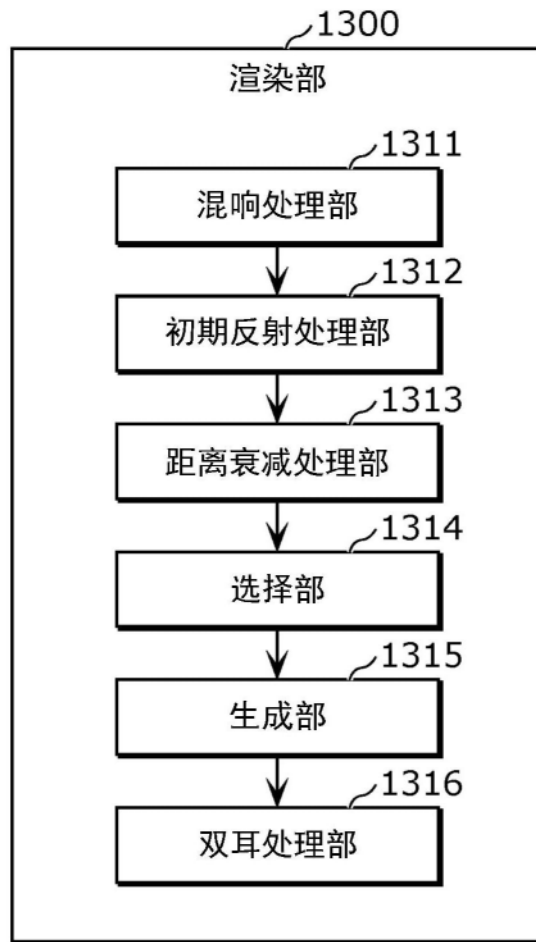


图24

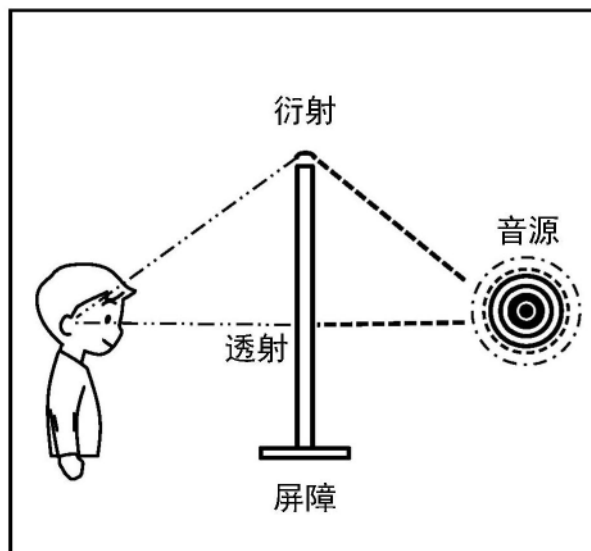


图25