



(12) 发明专利申请

(10) 申请公布号 CN 116018641 A

(43) 申请公布日 2023. 04. 25

(21) 申请号 202180052388.8

(22) 申请日 2021.08.20

(30) 优先权数据

2020-148234 2020.09.03 JP

(85) PCT国际申请进入国家阶段日

2023.02.23

(86) PCT国际申请的申请数据

PCT/JP2021/030599 2021.08.20

(87) PCT国际申请的公布数据

W02022/050087 JA 2022.03.10

(71) 申请人 索尼集团公司

地址 日本东京

(72) 发明人 本间弘幸 知念彻 河野明文

(74) 专利代理机构 北京康信知识产权代理有限公司 11240

专利代理师 沈丹阳

(51) Int.Cl.

G10L 19/008 (2006.01)

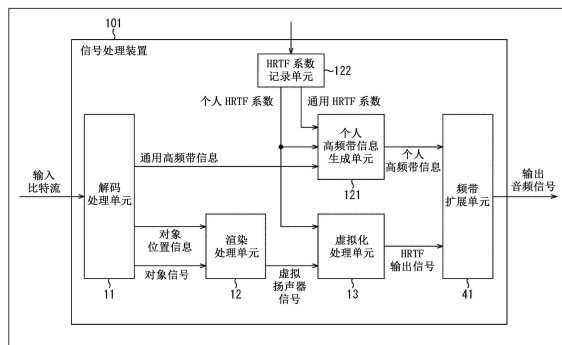
权利要求书2页 说明书24页 附图15页

(54) 发明名称

信号处理装置和方法、学习装置和方法以及程序

(57) 摘要

本技术涉及信号处理装置和方法、学习装置和方法、以及程序,这使的使用甚至廉价的装置也能够以高质量执行音频播放。该信号处理装置包括:解码处理单元,将输入比特流解复用为第一音频信号、第一音频信号的元数据和用于扩展频带的第一高频带信息;频带扩展单元,基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理来获得第二音频信号,基于第一高频带信息生成第二高频带信息。本技术可以应用于智能电话。



1. 一种信号处理装置,包括:

解码处理单元,将输入比特流解复用为第一音频信号、所述第一音频信号的元数据和用于扩展频带的第一高频带信息;以及

频带扩展单元,基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于所述第一音频信号和所述元数据执行信号处理来获得所述第二音频信号,基于所述第一高频带信息生成所述第二高频带信息。

2. 根据权利要求1所述的信号处理装置,进一步包括:

高频带信息生成单元,基于所述第一高频带信息生成所述第二高频带信息。

3. 根据权利要求2所述的信号处理装置,

其中,所述第一高频带信息是用于扩展通过使用第一系数执行信号处理而获得的所述第二音频信号的频带的高频带信息,

所述第二高频带信息是用于扩展通过使用第二系数执行信号处理而获得的所述第二音频信号的频带的高频带信息,并且

所述频带扩展单元基于所述第二音频信号和所述第二高频带信息执行频带扩展处理,通过基于所述第一音频信号、所述元数据和所述第二系数执行信号处理来获得所述第二音频信号。

4. 根据权利要求3所述的信号处理装置,其中,所述高频带信息生成单元基于所述第一高频带信息、所述第一系数和所述第二系数生成所述第二高频带信息。

5. 根据权利要求3所述的信号处理装置,其中,所述高频带信息生成单元通过基于通过机器学习预先生成的系数、所述第一高频带信息、所述第一系数和所述第二系数执行算术运算来生成所述第二高频带信息。

6. 根据权利要求5所述的信号处理装置,其中,所述算术运算是基于神经网络的算术运算。

7. 根据权利要求3所述的信号处理装置,其中,所述第一系数是通用系数,而所述第二系数是每个用户的系数。

8. 根据权利要求7所述的信号处理装置,其中,所述第一系数和所述第二系数是HRTF系数。

9. 根据权利要求1所述的信号处理装置,进一步包括:

信号处理单元,通过执行信号处理生成所述第二音频信号。

10. 根据权利要求9所述的信号处理装置,其中,所述信号处理是包括虚拟化处理或渲染处理的处理。

11. 根据权利要求1所述的信号处理装置,其中,所述第一音频信号是音频对象的对象信号或基于声道的音频信号。

12. 一种信号处理方法,包括通过信号处理装置执行以下:

将输入比特流解复用为第一音频信号、所述第一音频信号的元数据和用于扩展频带的第一高频带信息;以及

基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于所述第一音频信号和所述元数据执行信号处理来获得所述第二音频信号,基于所述第一高频带信息生成所述第二高频带信息。

13. 一种使计算机执行处理的程序,所述处理包括以下步骤:

将输入比特流解复用为第一音频信号、所述第一音频信号的元数据和用于扩展频带的第一高频带信息;以及

基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于所述第一音频信号和所述元数据执行信号处理来获得所述第二音频信号,基于所述第一高频带信息生成所述第二高频带信息。

14. 一种学习装置,包括:

第一高频带信息计算单元,基于第二音频信号生成用于扩展频带的第一高频带信息,其中,通过基于第一音频信号和第一系数的信号处理生成所述第二音频信号;

第二高频带信息计算单元,基于第三音频信号生成用于扩展频带的第二高频带信息,其中,通过基于所述第一音频信号和第二系数的信号处理生成所述第三音频信号;以及

高频带信息学习单元,基于所述第一系数、所述第二系数、所述第一高频带信息和所述第二高频带信息,使用所述第二高频带信息作为训练数据执行学习,并且生成用于从所述第一系数、所述第二系数和所述第一高频带信息获得所述第二高频带信息的系数数据。

15. 根据权利要求14所述的学习装置,其中,所述系数数据是配置神经网络的系数。

16. 根据权利要求14所述的学习装置,其中,所述第一系数是通用系数,而所述第二系数是每个用户的系数。

17. 根据权利要求16所述的学习装置,

其中,所述信号处理是包括虚拟化处理或渲染处理的处理,

所述第一系数和所述第二系数是HRTF系数。

18. 根据权利要求14所述的学习装置,其中,所述第一音频信号是音频对象的对象信号或基于声道的音频信号。

19. 一种学习方法,包括通过学习装置执行以下:

基于第二音频信号生成用于扩展频带的第一高频带信息,其中,通过基于第一音频信号和第一系数的信号处理生成所述第二音频信号;

基于第三音频信号生成用于扩展频带的第二高频带信息,其中,通过基于所述第一音频信号和第二系数的信号处理生成所述第三音频信号;以及

基于所述第一系数、所述第二系数、所述第一高频带信息和所述第二高频带信息,使用所述第二高频带信息作为训练数据执行学习,从而生成用于从所述第一系数、所述第二系数和所述第一高频带信息获得所述第二高频带信息的系数数据。

20. 一种用于使计算机执行处理的程序,所述处理包括以下步骤:

基于第二音频信号生成用于扩展频带的第一高频带信息,其中,通过基于第一音频信号和第一系数的信号处理生成所述第二音频信号;

基于第三音频信号生成用于扩展频带的第二高频带信息,其中,

通过基于所述第一音频信号和第二系数的信号处理生成所述第三音频信号;以及

基于所述第一系数、所述第二系数、所述第一高频带信息和所述第二高频带信息,使用所述第二高频带信息作为训练数据执行学习,从而生成用于从所述第一系数、所述第二系数和所述第一高频带信息获得所述第二高频带信息的系数数据。

信号处理装置和方法、学习装置和方法以及程序

技术领域

[0001] 本技术涉及信号处理装置和方法、学习装置和方法以及程序,具体地,涉及甚至使用廉价的装置能够执行高质量的音频播放的信号处理装置和方法、学习装置和方法、以及程序。

背景技术

[0002] 在现有技术中,对象音频技术用于电影、游戏等中,并且还开发了用于处理对象音频的编码方案。具体地,运动图像专家组(MPEG)-H部分3:3D音频标准(例如,其为国际标准)是已知的(例如,参见非专利文献1)。

[0003] 在这种编码方案中,如同常规的双声道立体声方案或5.1声道等的多声道立体声方案,可以将移动声源等处理为独立的音频对象(在下文中,也简称为对象),并且将对象的位置信息与音频对象的信号数据一起编码为元数据。

[0004] 因此,能够在扬声器的数量和布置不同的各种视听环境中执行播放。此外,能够在播放时处理来自特定声源的声音,诸如调整来自特定声源的声音的音量以及向来自特定声源的声音添加效果,这在传统的编码方案中是困难的。

[0005] 在这种编码方案中,在解码侧上解码比特流,并且获得作为对象的音频信号的对象信号以及包括指示对象在空间中的位置的对象位置信息的元数据。

[0006] 然后,基于对象位置信息执行将对象信号渲染到虚拟布置在空间中的多个虚拟扬声器中的每个的渲染处理。在非专利文献1的标准中,例如,称为基于三维向量的振幅平移(在下文中,简称为VBAP)的方案用于渲染处理。

[0007] 此外,一旦通过渲染处理获得与每个虚拟扬声器相对应的虚拟扬声器信号,基于虚拟扬声器信号执行头相关传输函数(HRTF)处理。在HRTF处理中,生成用于从实际耳机和扬声器输出声音的输出音频信号,如同声音被虚拟扬声器播放。

[0008] 在这样的对象音频被实际播放,并且可以在空间中布置大量实际扬声器的情况下,执行基于虚拟扬声器信号的播放。此外,当不能布置许多扬声器并且通过诸如耳机和条形音箱的少量扬声器播放对象音频时,执行基于前述输出音频信号的播放。

[0009] 另一方面,近年来存储价格的降低和网络带宽的增加已经使得能够享受所谓的高分辨率声源,即,具有等于或大于96kHz的采样频率的高分辨率声源。

[0010] 根据在非专利文献1中描述的编码方案,可以使用诸如频带复制(SBR)的技术作为用于有效地对高分辨率声源进行编码的技术。

[0011] 在SBR中,例如,以与高频子频带的数目相对应的量对高频子频带信号的平均振幅信息进行编码,并且然后在不编码频谱的高频分量的情况下在编码侧发送该高频子频带信号的平均振幅信息。

[0012] 然后,在解码侧上,基于低频子频带信号和高频带的平均振幅信息生成包括低频分量和高频分量的最终输出信号。因此,可以实现更高质量的音频播放。

[0013] 在该方法中,使用视听特性,即,人类对高频信号分量的相位的变化不敏感并且在

其频域包络的轮廓接近其原始信号的情况下不能感知差异,并且通常,这种方法被称为频带扩展技术。

[0014] [引用列表]

[0015] [非专利文献]

[0016] [非专利文献1]

[0017] 国际标准ISO/IEC 23008-3第二版2019-02

[0018] Information technology-High efficiency coding and media delivery in heterogeneous environments-Part 3:3D audio。

发明内容

[0019] [技术问题]

[0020] 顺便提及,在结合渲染处理和HRTF处理对前述对象音频执行频带扩展的情况下,对每个对象的对象信号执行频带扩展处理,然后对其执行渲染处理或HRTF处理。

[0021] 在这种情况下,频带扩展处理被独立地执行的次数对应于对象数量,并且处理负荷(即,算术运算量)因此增加。此外,由于在频带扩展处理之后对通过频带扩展获得的具有较高采样频率的信号作为对象执行渲染处理或HRTF处理,因此处理负荷进一步增加。

[0022] 因此,诸如廉价的处理器或电池之类的设备的廉价装置(即,具有低算术运算能力的设备、具有低电池容量的设备等)不能执行频带扩展,结果,不能执行高质量的音频播放。

[0023] 鉴于这种情况做出本技术,并且其目的是甚至使廉价装置能够以高质量执行音频播放。

[0024] [问题的解决方案]

[0025] 根据本技术的第一方面的信号处理装置包括:解码处理单元,将输入比特流解复用为第一音频信号、第一音频信号的元数据和用于扩展频带的第一高频带信息;以及频带扩展单元,基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理来获得第二音频信号,基于第一高频带信息生成第二高频带信息。

[0026] 根据本技术的第一方面的信号处理方法或程序包括以下步骤:将输入比特流解复用为第一音频信号、第一音频信号的元数据以及用于扩展频带的第一高频带信息;以及基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理来获得第二音频信号,基于第一高频带信息生成第二高频带信息。

[0027] 在本技术的第一方面中,将输入比特流解复用为第一音频信号、第一音频信号的元数据以及用于扩展频带的第一高频带信息,基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理获得第二音频信号,基于第一高频带信息生成第二高频带信息。

[0028] 根据本技术的第二方面的学习装置包括:第一高频带信息计算单元,基于通过基于第一音频信号和第一系数的信号处理生成的第二音频信号生成用于扩展频带的第一高频带信息;第二高频带信息计算单元,基于根据第一音频信号和第二系数的信号处理生成的第三音频信号来生成用于扩展频带的第二高频带信息;以及高频带信息学习单元,基于

第一系数、第二系数、第一高频带信息和第二高频带信息,使用第二高频带信息作为训练数据执行学习,并且生成用于从第一系数、第二系数和第一高频带信息获得第二高频带信息的系数数据。

[0029] 根据本技术的第二方面的学习方法或程序包括以下步骤:基于通过基于第一音频信号和第一系数的信号处理生成的第二音频信号生成用于扩展频带的第一高频带信息;基于通过基于第一音频信号和第二系数的信号处理生成的第三音频信号生成用于扩展频带的第二高频带信息;以及基于第一系数、第二系数、第一高频带信息和第二高频带信息,使用第二高频带信息作为训练数据执行学习,从而生成用于从第一系数、第二系数和第一高频带信息获得第二高频带信息的系数数据。

[0030] 在本技术的第二方面中,基于通过基于第一音频信号和第一系数的信号处理生成的第二音频信号来生成用于扩展频带的第一高频带信息,基于通过基于第一音频信号和第二系数的信号处理生成的第三音频信号生成用于扩展频带的第二高频带信息,基于第一系数、第二系数、第一高频带信息,使用第二高频带信息作为训练数据执行学习,以及从第一高频带信息、第二系数和第一高频带信息生成用于获得第二高频带信息的系数数据。

附图说明

[0031] 图1是用于说明输出音频信号的生成的示图。

[0032] 图2是用于说明VBAP的示图。

[0033] 图3是用于说明HRTF处理的示图。

[0034] 图4是用于说明频带扩展处理的示图。

[0035] 图5是用于说明频带扩展处理的示图。

[0036] 图6为示出信号处理装置的配置实例的示图。

[0037] 图7是示出应用本技术的信号处理装置的配置实例的示图。

[0038] 图8是示出个人高频带信息生成单元的配置实例的示图。

[0039] 图9是示出输入比特流的语法实例的示图。

[0040] 图10是用于说明信号生成处理的流程图。

[0041] 图11是示出学习装置的配置实例的示图。

[0042] 图12是用于说明学习处理的流程图。

[0043] 图13是示出编码器的配置实例的示图。

[0044] 图14是用于说明编码处理的流程图。

[0045] 图15是示出计算机的配置实例的示图。

具体实施方式

[0046] 在下文中,将参考附图描述应用本技术的实施方式。

[0047] <第一实施方式>

[0048] <关于本技术>

[0049] 根据本技术,预先在比特流中多路复用并发送用于对作为对象的HRTF输出信号进行频带扩展处理的通用高频带信息,并且在解码侧,基于个人HRTF系数、通用HRTF系数和通用高频带信息来生成与个人HRTF系数相对应的高频带信息。

[0050] 因此,可以以低采样频率执行需要高处理负荷的解码处理、渲染处理和虚拟化处理,并随后基于与个人HRTF系数相对应的高频带信息来执行频带扩展处理,从而整体上减少算术运算量。结果,即使在廉价的设备的情况下,也可以基于更高采样频率的输出音频信号执行高质量的音频播放。

[0051] 具体地,根据本技术,在解码侧生成与个人HRTF系数相对应的高频带信息,因此不需要在编码侧为各个单独用户准备高频带信息。此外,通过在解码侧生成与个人HRTF系数相对应的高频带信息,与使用普通高频带信息的情况相比,能够以更高的质量执行音频播放。

[0052] 在下文中,将更详细地描述本技术。

[0053] 首先,将描述当通过MPEG-H第3部分:3D音频标准的编码方案的编码而获得的比特流被解码并生成对象音频的输出音频信号时执行的总体处理。

[0054] 如图1所示,例如,一旦通过编码(编码)获得的输入比特流被输入到解码处理单元11,则对输入比特流执行解复用和解码处理。

[0055] 通过解码处理,获得对象信号和元数据,对象信号是用于播放对象配置内容(音频对象)的声音的音频信号,元数据包括指示对象在空间中的位置的对象位置信息。

[0056] 随后,渲染处理单元12基于包括在元数据中的对象位置信息执行将对象信号渲染到虚拟布置在空间中的虚拟扬声器的渲染处理,并生成用于播放从每个虚拟扬声器输出的声音的虚拟扬声器信号。

[0057] 此外,虚拟化处理单元13基于每个虚拟扬声器的虚拟扬声器信号执行虚拟化处理,并生成输出音频信号,以用于使诸如用户佩戴的头戴式耳机或布置在实际空间中的扬声器的播放装置输出声音。

[0058] 虚拟化处理是其中生成用于实现音频播放的音频信号的处理,如同利用与实际播放环境中的声道配置不同的声道配置执行播放一样。

[0059] 在该实例中,例如,生成用于实现音频播放的输出音频信号的处理是虚拟化处理,所实现的音频播放如同从每个虚拟扬声器输出声音,而不管从诸如耳机的播放装置输出声音的实际情况如何。

[0060] 虽然可通过任何方法实现虚拟化处理,但假设执行HRTF处理作为虚拟化处理,将继续以下描述。

[0061] 如果基于通过虚拟化处理获得的输出音频信号从实际的耳机或扬声器输出声音,则可以实现音频播放,好像声音是从虚拟扬声器播放的。注意,具体地,下面将实际布置在实际空间中的扬声器称为实际扬声器。

[0062] 在播放这样的对象音频的情况下,当可在空间中布置大量的实际扬声器时,可以通过实际扬声器照原样播放渲染处理的输出。

[0063] 另一方面,当不可能在空间中布置大量的实际扬声器时,通过执行HRTF处理,使用少量的实际扬声器(诸如,耳机和条形音箱)执行播放。通常,在许多情况下使用耳机或少量实际扬声器来执行播放。

[0064] 这里,将进一步描述通用渲染处理和HRTF处理。

[0065] 在渲染时,例如,执行预定方案(例如,上述VBAP)的渲染处理。VBAP是通常被称为平移的渲染方法,并且通过将增益分配给三个虚拟扬声器来执行渲染,三个虚拟扬声器是类似地存在于球面上的虚拟扬声器之中最接近存在于包括用户位置作为原点的球面上的

对象。

[0066] 如图2所示,例如,假设作为收听者的用户U11存在于三维空间中,并且三个虚拟扬声器SP1至SP3布置在用户U11的前方。

[0067] 这里,用户U11的头部的被定义为原点0,并且虚拟扬声器SP1至SP3被认为是在围绕位于中心的原点0的球体的表面上。

[0068] 现在,将考虑对象存在于由球体表面上的虚拟扬声器SP1至SP3包围的区域TR11内并且声音图像位于该对象的位置VSP1的情形。

[0069] 在这种情况下,增益被分配给在VBAP中的对象的位置VSP1周围存在的虚拟扬声器SP1至SP3。

[0070] 具体而言,以原点0为基准(原点)的三维坐标系中的原点0开始以及在位置VSP1结束的三维向量P表示位置VSP1。

[0071] 此外,如果认定从原点0开始并在虚拟扬声器SP1至SP3的位置结束的三维向量是向量 L_1 至 L_3 ,则向量P可由向量 L_1 至 L_3 的线性和表示,如下面的表达式(1)所示。

[0072] [数学式.1]

$$[0073] \quad P = g_1 L_1 + g_2 L_2 + g_3 L_3 \cdots (1)$$

[0074] 这里,通过计算在表达式(1)中与向量 L_1 至 L_3 相乘的系数 g_1 至 g_3 并且将这些系数 g_1 至 g_3 视为从每个虚拟扬声器SP1至SP3输出的声音的增益,可以将声音图像定位到位置VSP1处。

[0075] 当包括系数 g_1 至 g_3 作为元素的向量被定义为 $g_{123} = [g_1, g_2, g_3]$,并且包括向量 L_1 至 L_3 作为元素的向量被定义为 $L_{123} = [L_1, L_2, L_3]$ 时,可以通过使上述表达式(1)变形来获得以下表达式(2)。

[0076] [数学式.2]

$$[0077] \quad g_{123} = P^T L_{123}^{-1} \cdots (2)$$

[0078] 通过使用通过如上所述计算表达式(2)而获得的系数 g_1 至 g_3 作为增益,基于来自虚拟扬声器SP1至SP3中的每一个的对象信号输出声音,可以将声音图像定位在位置VSP1处。

[0079] 应注意,虚拟扬声器SP1至SP3中的每一个的布置位置是固定的,指示虚拟扬声器的位置的信息是已知的,因此可预先获得作为逆矩阵的 L_{123}^{-1} 。

[0080] 在图2中示出的由球表面上的三个虚拟扬声器包围的三角形区域TR11称为网格。通过组合布置在空间中的多个虚拟扬声器以配置多个网格,可以在空间中的任意位置处定位对象的声音。

[0081] 如果以这种方式为每个对象获得虚拟扬声器的增益,则可以通过执行下面表达式(3)的算术运算为每个虚拟扬声器获得虚拟扬声器信号。

[0082] [数学式.3]

$$\begin{matrix} [0083] \\ \end{matrix} \begin{bmatrix} SP(0, t) \\ SP(1, t) \\ \vdots \\ SP(M-1, t) \end{bmatrix} = \begin{bmatrix} G(0, 0) & G(0, 1) & \cdots & G(0, N-1) \\ G(1, 0) & G(1, 1) & \cdots & G(1, N-1) \\ \vdots & \vdots & & \vdots \\ G(M-1, 0) & G(M-1, 1) & \cdots & G(M-1, N-1) \end{bmatrix} \begin{bmatrix} S(0, t) \\ S(1, t) \\ \vdots \\ S(N-1, t) \end{bmatrix} \dots (3)$$

[0084] 应注意,表达式(3)中的 $SP(m, t)$ 表示 M 个虚拟扬声器中的第 m 个(其中, $m=0, 1, \dots, M-1$)虚拟扬声器的时钟时间 t 处的虚拟扬声器信号。而且,在表达式(3)中的 $S(n, t)$ 表示在 N 个对象之中的第 n 个(其中, $n=0, 1, \dots, N-1$)对象的时钟时间 t 的对象信号。

[0085] 此外,表达式(3)中的 $G(m, n)$ 表示第 n 个对象的对象信号 $S(n, t)$ 乘以的增益,以获得用于第 m 个虚拟扬声器的虚拟扬声器信号 $SP(m, t)$ 。换言之,增益 $G(m, n)$ 表示分配给通过上述表达式(2)获得的第 n 个对象的第 m 个虚拟扬声器的增益。

[0086] 在渲染处理中,表达式(3)的计算是需要最高计算成本的处理。换句话说,表达式(3)的算术运算是需要最大量的算术运算的处理。

[0087] 接下来,将参考图3描述在基于通过表达式(3)的算术运算获得的虚拟扬声器信号的声音被耳机或少量实际扬声器播放的情况下执行的HRTF处理的实例。应注意,图3示出了为了简化说明将虚拟扬声器布置在二维水平面中的实例。

[0088] 在图3中,五个虚拟扬声器 $SP11-1$ 至 $SP11-5$ 在空间中圆形地排列和布置。在下文中,在不特别需要将虚拟扬声器 $SP11-1$ 至 $SP11-5$ 彼此区分的情况下,虚拟扬声器 $SP11-1$ 至 $SP11-5$ 也将被简称为虚拟扬声器 $SP11$ 。

[0089] 而且,作为收听者的用户 $U21$ 位于由五个虚拟扬声器 $SP11$ 包围的位置,即,在图3中布置了虚拟扬声器 $SP11$ 的圆的中心位置。因此,在HRTF处理中生成用于实现音频播放的输出音频信号,好像用户 $U21$ 收听从每个虚拟扬声器 $SP11$ 输出的声音。

[0090] 具体地,假设在该实例中,用户 $U21$ 所位于的位置是收听位置,并且通过耳机播放基于通过渲染五个虚拟扬声器 $SP11$ 中的每个而获得的虚拟扬声器信号的声音。

[0091] 在这种情况下,例如,基于虚拟扬声器信号从虚拟扬声器 $SP11-1$ 输出(发出)的声音通过由箭头 $Q11$ 表示的路径并且到达用户 $U21$ 的左耳的鼓膜。因此,从虚拟扬声器 $SP11-1$ 输出的声音的特性应该根据从虚拟扬声器 $SP11-1$ 到用户 $U21$ 的左耳的空间传输特性、面部和耳朵的形状以及用户 $U21$ 的反射/吸收特性等而改变。

[0092] 由此,通过对用于虚拟扬声器 $SP11-1$ 的虚拟扬声器信号卷积考虑从虚拟扬声器 $SP11-1$ 到用户 $U21$ 的左耳的空间传输特性、面部和耳朵的形状以及用户 $U21$ 的反射/吸收特性等的传输函数 H_L_SP11 ,可以获得用于播放来自虚拟扬声器 $SP11-1$ 的声音的输出音频信号,来自虚拟扬声器 $SP11-1$ 的声音被认为由用户 $U21$ 的左耳收听。

[0093] 类似地,例如,基于虚拟扬声器信号从虚拟扬声器 $SP11-1$ 输出的声音通过由箭头 $Q12$ 表示的路径并且到达用户 $U21$ 的右耳的鼓膜。因此,通过对用于虚拟扬声器 $SP11-1$ 的虚拟扬声器信号卷积考虑从虚拟扬声器 $SP11-1$ 到用户 $U21$ 的右耳的空间传输特性、面部和耳朵的形状以及用户 $U21$ 的反射/吸收特性等的传输函数 H_R_SP11 ,可以获得用于播放被认为由用户 $U21$ 的右耳收听的来自虚拟扬声器 $SP11-1$ 的声音的输出音频信号。

[0094] 因此,当由耳机最终播放基于用于五个虚拟扬声器 $SP11$ 的虚拟扬声器信号的声音

时,仅需要对用于左声道的每个虚拟扬声器信号卷积用于每个虚拟扬声器的左耳的传输函数,并且将因此获得的每个信号相加以获得用于左声道的输出音频信号。

[0095] 类似地,在右声道的情况下,仅需要对每个虚拟扬声器信号卷积用于每个虚拟扬声器的右耳的传输函数并将因此获得的每个信号相加以获得用于右声道的输出音频信号。

[0096] 注意,即使在用于播放的播放装置是实际扬声器而不是耳机的情况下,也执行与在耳机的情况下的HRTF处理相似的HRTF处理。然而,因为来自扬声器的声音通过空间传播到达用户的左耳和右耳两者,所以执行考虑串扰的处理。这种处理也称为听觉传输(transoral)处理。

[0097] 当左耳(即,左声道)的通常由频率表示的输出音频信号是 $L(\omega)$,并且右耳(即,右声道)的由频率表示的输出音频信号是 $R(\omega)$ 时,可以通过计算下面的表达式(4)来获得 $L(\omega)$ 和 $R(\omega)$ 。

[0098] [数学式.4]

$$[0099] \begin{bmatrix} L(\omega) \\ R(\omega) \end{bmatrix} = \begin{bmatrix} H_L(0, \omega) & H_L(1, \omega) & \cdots & H_L(M-1, \omega) \\ H_R(0, \omega) & H_R(1, \omega) & \cdots & H_R(M-1, \omega) \end{bmatrix} \begin{bmatrix} SP(0, \omega) \\ SP(1, \omega) \\ \vdots \\ SP(M-1, \omega) \end{bmatrix} \quad \dots (4)$$

[0100] 应注意,表达式(4)中的 ω 表示频率,并且 $SP(m, \omega)$ 表示 M 个虚拟扬声器中的第 m 个(其中, $m=0, 1, \dots, M-1$)虚拟扬声器的频率 ω 的虚拟扬声器信号。虚拟扬声器信号 $SP(m, \omega)$ 可通过对上述虚拟扬声器信号 $SP(m, t)$ 执行时频转换来获得。

[0101] 此外,表达式(4)中的 $H_L(m, \omega)$ 表示左耳的传输函数,第 m 个虚拟扬声器的虚拟扬声器信号 $SP(m, \omega)$ 与该传输函数相乘以获得左声道的输出音频信号 $L(\omega)$ 。类似地, $H_R(m, \omega)$ 表示右耳的传输函数。

[0102] 在用于HRTF的传输函数 $H_L(m, \omega)$ 和传输函数 $H_R(m, \omega)$ 被表示为时域中的脉冲响应的情况下,需要至少约1秒的长度。因此,例如,在虚拟扬声器信号的采样频率是48kHz的情况下,需要执行48000抽头的卷积,并且即使使用快速傅里叶变换(FFT)的高速算术运算方法用于传输函数的卷积,仍然需要更大量的算术运算。

[0103] 如上所述,在通过执行解码处理、渲染处理和HRTF处理生成输出音频信号,并且使用耳机或少量实际扬声器来播放对象音频的情况下,需要大量的算术运算。另外,如果对象的数量增加,则算术运算量也相应地增加。

[0104] 接下来,将描述频带扩展处理。

[0105] 在通用频带扩展处理中,即SBR中,在编码侧不对音频信号的频谱的高频带分量进行编码,根据高频子频带的数量,对高频带内的高频子频带的高频子频带信号的平均振幅信息进行编码,然后发送给解码侧。

[0106] 此外,在解码侧,利用平均振幅将作为通过解码处理(解码)获得的音频信号的低频子频带信号标准化,并且将标准化的信号复制到高频子频带。然后,通过将由此获得的信号乘以每个高频子频带的平均振幅信息获得高频子频带信号,对低频子频带信号和高频子频带信号进行子频带合成,并且由此获得最终输出音频信号。

[0107] 例如,可以通过这种频带扩展处理以等于或大于96kHz的采样频率执行高分辨率声源的音频播放。

[0108] 然而,与典型的立体声音频不同,在处理对象音频中的采样频率为96kHz的信号的情况下,例如,对通过解码获得的96kHz的对象信号执行渲染处理和HRTF处理,而不管是否执行诸如SBR的频带扩展处理。因此,在对象的数目或虚拟扬声器的数目大的情况下,处理的计算成本显著增加,并且需要高性能处理器和高功耗。

[0109] 这里,将参考图4描述在通过对象音频中的频带扩展获得96kHz的输出音频信号的情况下执行的处理的实例。要注意的是,在图4中的与图1中的部件相对应的部件采用相同的参考标号,并且将省略其描述。

[0110] 如果提供了输入比特流,那么解码处理单元11执行解复用和解码处理,并且输出因此获得的对象信号以及对象的对象位置信息和高频带信息。

[0111] 例如,高频带信息为从编码前的对象信号获得的高频子频带信号的平均振幅信息。

[0112] 换言之,高频带信息是与通过解码处理获得的对象信号相对应的频带扩展的频带扩展信息,并且指示在较高采样频率编码之前的对象信号的高频带侧上的每个子频带分量的大小。注意,尽管这里描述了SBR的实例,高频子频带信号的平均振幅信息被用作频带扩展信息,但是用于频带扩展处理的频带扩展信息可以是任何信息,诸如在编码之前的对象信号的高频带侧上的每个子频带的振幅的代表值或者指示频率包络的形状的信息。

[0113] 此外,例如,通过解码处理获得的对象信号被认为为48kHz的采样频率的对象信号,并且此类对象信号在下文中也将被称为低FS对象信号。

[0114] 在解码处理之后,频带扩展单元41基于高频带信息和低FS对象信号执行频带扩展处理,并且获得较高采样频率的对象信号。在该实例中,例如,假设通过频带扩展处理获得采样频率为96kHz的对象信号,并且此类对象信号在下文中也将被称为高FS对象信号。

[0115] 此外,渲染处理单元12基于通过解码处理获得的对象位置信息和通过频带扩展处理获得的高FS对象信号来执行渲染处理。在该实例中,具体地,通过渲染处理获得采样频率为96kHz的虚拟扬声器信号,并且此类虚拟扬声器信号在下文中也将被称为高FS虚拟扬声器信号。

[0116] 此外,虚拟化处理单元13然后基于高FS虚拟扬声器信号执行诸如HRTF处理的虚拟化处理,并且获得96kHz的采样频率的输出音频信号。

[0117] 这里,将参考图5描述通用频带扩展处理。

[0118] 图5示出了预定对象信号的频率振幅特性。要注意的是,在图5中,垂直轴表示振幅(功率),而水平轴表示频率。

[0119] 例如,折线L11表示提供给频带扩展单元41的低FS对象信号的频率振幅特性。低FS对象信号具有48kHz的采样频率,并且低FS对象信号不包括等于或大于24kHz的频带中的信号分量。

[0120] 这里,例如,高达24kHz的频带被划分为包括低频子频带sb-8至sb-1的多个低频子频带,这些低频子频带中的每个低频子频带的信号分量是低频子频带信号。类似地,24kHz至48kHz的频带被划分为高频子频带sb至sb+13,这些高频子频带中的每个的信号分量是高频子频带信号。

[0121] 此外,指示关于每个高频子频带 s_b 至 s_b+13 的这些高频子频带的平均振幅信息的高频带信息被提供给频带扩展单元41。

[0122] 在图5中,例如,直线L12表示作为高频子频带 s_b 的高频带信息提供的平均振幅信息,并且直线L13表示作为高频子频带 s_b+1 的高频带信息提供的平均振幅信息。

[0123] 在频带扩展单元41中,利用低频子频带信号的平均振幅值对低频子频带信号进行标准化,并将通过标准化得到的信号复制(映射)到高频侧。这里,通过扩展频带等预先定义作为复制源的低频子频带和作为低频子频带的复制目的地的高频子频带。

[0124] 例如,对低频子频带 s_b-8 的低频子频带信号进行标准化,并将通过标准化获得的信号复制到高频子频带 s_b 。

[0125] 更具体地,对低频子频带 s_b-8 的低频子频带信号标准化后的信号进行调制处理,信号转换为高频子频带 s_b 的频率分量的信号。

[0126] 类似地,例如,在标准化之后,将低频子频带 s_b-7 的低频子频带信号复制到高频子频带 s_b+1 。

[0127] 一旦将如此标准化的低频子频带信号复制(映射)到高频子频带,则将复制到每个高频子频带的信号乘以每个高频子频带的高频带信息所指示的平均振幅信息,从而生成高频子频带信号。

[0128] 在高频子频带 s_b 中,例如,将通过标准化低频子频带 s_b-8 的低频子频带信号并将其复制到高频子频带 s_b 而获得的信号乘以由直线L12表示的平均振幅信息,并且获得结果为高频子频带 s_b 的高频子频带信号。

[0129] 一旦针对每个高频子频带获得高频子频带信号,则将每个低频子频带信号和每个高频子频带信号输入到用于以96kHz采样的带合成滤波器并通过该带合成滤波器滤波(合成),并且输出由此获得的高FS对象信号。换言之,获得上采样(频带扩展)至96kHz的采样频率的高FS对象信号。

[0130] 在图4所示的实例中,对于包含在输入比特流中的每个低FS对象信号,即,对于频带扩展单元41中的每个对象,独立地执行如上所述的生成高FS对象信号的频带扩展处理。

[0131] 因此,在对象的数量是三十二的情况下,例如,渲染处理单元12必须对三十二对象中的每个执行96kHz的高FS对象信号的渲染处理。

[0132] 类似地,即使在虚拟化处理单元13中,96kHz的高FS虚拟扬声器信号的HRTF处理(虚拟化处理)也在其稍后阶段必须执行与虚拟扬声器的数量相对应的次数。

[0133] 结果,整个设备中的处理负荷显著增加。这适用于在不执行频带扩展处理的情况下通过解码处理获得的音频信号的采样频率是96kHz的情况。

[0134] 由此,可以想到,在编码时以高分辨率(即,以高采样频率)预先计算虚拟化处理之后的信号的高频带信息,将其复用到输入比特流,并且将其传送。

[0135] 以这种方式,例如,可在低采样频率执行需要高处理负荷的解码处理、渲染处理和HRTF处理,并基于关于HRTF处理之后的最终信号的传输的高频带信息来执行频带扩展处理。因此,能够整体上降低处理负荷,并且能够实现廉价的处理器或电池而以高质量实现音频播放。

[0136] 在这种情况下,例如,解码侧上的信号处理装置可被配置为如图6中所示。要注意的是,在图6中对与在图4的情况下的部件相对应的部件应用相同的参考符号,并且将适当

地省略其描述。

[0137] 在图6中示出的信号处理装置71由例如智能电话或个人计算机构成,并且包括解码处理单元11、渲染处理单元12、虚拟化处理单元13以及频带扩展单元41。

[0138] 在图4所示的实例中,按照解码处理、频带扩展处理、渲染处理、以及虚拟化处理的顺序执行各种处理。

[0139] 另一方面,在信号处理装置71中,按照解码处理、渲染处理、虚拟化处理和频带扩展处理的顺序执行各种处理(信号处理)。换言之,最后执行频带扩展处理。

[0140] 因此,首先由信号处理装置71中的解码处理单元11执行输入比特流的解复用和解码处理。

[0141] 解码处理单元11将通过解复用和解码处理获得的高频带信息提供给频带扩展单元41,并且将对象位置信息和对象信号提供给渲染处理单元12。

[0142] 这里,输入比特流包括与虚拟化处理单元13的输出相对应的高频带信息,并且解码处理单元11将高频带信息提供给频带扩展单元41。

[0143] 而且,渲染处理单元12基于从解码处理单元11提供的对象位置信息和对象信号执行诸如VBAP的渲染处理,并且将由此所获得的虚拟扬声器信号提供给虚拟化处理单元13。

[0144] 虚拟化处理单元13执行HRTF处理作为虚拟化处理。换言之,作为HRTF处理,虚拟化处理单元13基于从渲染处理单元12提供的虚拟扬声器信号和与预先给出的传输函数对应的HRTF系数执行卷积处理,并且对由此所获得的信号进行相加处理。虚拟化处理单元13将通过HRTF处理获得的音频信号提供至频带扩展单元41。

[0145] 在该实例中,例如,从解码处理单元11提供至渲染处理单元12的对象信号是具有48kHz的采样频率的低FS对象信号。

[0146] 在这种情况下,因为从渲染处理单元12提供给虚拟化处理单元13的虚拟扬声器信号也是具有48kHz的采样频率的信号,并且从虚拟化处理单元13提供给频带扩展单元41的音频信号的采样频率也是48kHz。

[0147] 在下文中,特别地,将从虚拟化处理单元13提供给频带扩展单元41的音频信号也称为低FS音频信号。这种低FS音频信号是通过对象信号执行信号处理(例如,渲染处理和虚拟化处理)而获得的驱动信号,并且驱动播放装置(例如,耳机或实际扬声器),以使播放装置输出声音。

[0148] 频带扩展单元41通过基于从解码处理单元11提供的高频带信息对从虚拟化处理单元13提供的低FS音频信号执行频带扩展处理来生成输出音频信号,并且将输出音频信号输出至后级。例如,通过频带扩展单元41获得的输出音频信号是96kHz的采样频率的信号。

[0149] 顺便提及,众所周知,在用作虚拟化处理的HRTF处理中的HRTF系数极大地取决于作为收听者的各个单独用户的耳朵和面部的形状。

[0150] 由于与虚拟周围环境兼容的普通耳机等难以获取适合于各个单独用户的个人HRTF系数,因此在许多情况下使用对于耳朵和面部的平均形状来说通用的HRTF系数(即,所谓的通用HRTF系数)。

[0151] 然而,已知在使用通用HRTF系数的情况下,与使用个人HRTF系数的情况相比,声源的定位感和声音质量本身明显劣化。

[0152] 因此,还提出了用于更简单地获取适合于各个单独用户的HRTF系数的测量方法,

例如,在WO2018/110269中详细描述了这种测量方法。

[0153] 在下文中,具体地,针对人耳和面部的平均形状测量或生成的通用HRTF系数也将被称为通用HRTF系数。

[0154] 此外,具体地,针对每个单独用户测量或生成的并且与用户的耳朵和面部的形状相对应的HRTF系数(即,针对每个单独用户的HRTF系数)也将被称为个人HRTF系数。

[0155] 注意,个人HRTF系数不限于针对每个单独用户测量或生成的一个HRTF系数,并且可以是适合于每个单独用户的HRTF系数,并且基于与每个单独用户相关的信息从针对耳朵和面部的形状中的每个测量或生成的多个HRTF系数中选择HRTF系数,信息是诸如用户的耳朵和面部的近似形状、年龄、性别等。

[0156] 如上所述,适合于用户的HRTF系数对于每个用户是不同的。

[0157] 例如,在图6中示出的信号处理装置71的虚拟化处理单元13使用个人HRTF系数的假设下,期望采用对应于个人HRTF系数的高频带信息作为由频带扩展单元41使用的高频带信息。

[0158] 然而,输入比特流中包括的高频带信息是假定对通过使用通用HRTF系数执行HRTF处理所获得的音频信号执行频带扩展处理的通用高频带信息。

[0159] 因此,如果输入比特流中包括的高频带信息照原样使用以对通过使用个人HRTF系数执行HRTF处理所获得的音频信号执行频带扩展处理,则在所获得的输出音频信号中可能发生声音质量的显著恶化。

[0160] 另一方面,在操作方面,通过预先假设使用个人HRTF系数来存储和发送针对每个用户(即,针对每个个人HRTF系数)生成的高频带信息(个人高频带信息)是不容易的。

[0161] 这是因为需要针对播放对象音频的每个用户(个人)准备输入比特流,并针对每个个人HRTF系数准备与个人HRTF系数相对应的个人高频带信息。为此,在音频对象(输入比特流)的分发侧(即,在编码侧)的服务器等的存储容量也被压制。

[0162] 由此,根据本技术,在通用HRTF系数前提下,使用通用高频带信息、通用HRTF系数和个人HRTF系数,在播放(再现, replay)装置侧(解码侧)生成个人高频带信息。

[0163] 以这种方式,例如,可以以低采样频率执行需要高处理负荷的解码处理、渲染处理和HRTF处理,并基于由此生成的关于HRTF处理之后的最终信号的个人高频带信息执行频带扩展处理。因此,整体上能够减轻处理负荷,并且能够使廉价的处理器或电池实现高质量的音频播放。

[0164] <信号处理装置的配置实例>

[0165] 图7是示出应用本技术的信号处理装置101的实施方式的配置实例的示图。应注意,相同的参考标号将应用于对应于图6中的情况的图7中的部件,并且将适当地省略其描述。

[0166] 信号处理装置101由例如智能电话或个人计算机构成并且包括解码处理单元11、渲染处理单元12、虚拟化处理单元13、个人高频带信息生成单元121、HRTF系数记录单元122和频带扩展单元41。

[0167] 信号处理装置101的配置与信号处理装置71的配置的不同之处在于新提供了个人高频带信息生成单元121和HRTF系数记录单元122,并且在其他点上与信号处理装置71的配置相同。

[0168] 解码处理单元11从未示出的服务器等获取(接收)包括对象音频的编码的对象信号的输入比特流、包括对象位置信息等的元数据、通用高频带信息等。

[0169] 输入比特流中包括的通用高频带信息与信号处理装置71的解码处理单元11获取的输入比特流中包括的高频带信息基本相同。

[0170] 解码处理单元11将通过接收等获取的输入比特流和编码对象信号和元数据解复用为通用高频带信息并且解码编码的对象信号和元数据。

[0171] 解码处理单元11将通过输入比特流进行解复用和解码处理所获得的通用高频带信息提供给个人高频带信息生成单元121并且将对象位置信息和对象信号提供给渲染处理单元12。

[0172] 这里,当虚拟化处理单元13使用通用HRTF系数执行HRTF处理时,输入比特流包括与虚拟化处理单元13的输出相对应的通用高频带信息。换言之,通用高频带信息是用于扩展通过使用通用HRTF系数执行HRTF处理所获得的HRTF输出信号的频带的高频带信息。

[0173] 渲染处理单元12基于从解码处理单元11提供的对象位置信息和对象信号执行诸如VBAP的渲染处理,并且将因此获得的虚拟扬声器信号提供给虚拟化处理单元13。

[0174] 虚拟化处理单元13基于从渲染处理单元12提供的虚拟扬声器信号和与预先给定的发送功能对应并从HRTF系数记录单元122提供的个人HRTF系数,执行作为虚拟化处理的HRTF处理,并将因此获得的音频信号提供给频带扩展单元41。

[0175] 在HRTF处理中,例如,执行每个虚拟扬声器的虚拟扬声器信号和个人HRTF系数的卷积处理以及将通过每个虚拟扬声器的卷积处理获得的信号相加的加法处理。

[0176] 应注意,以下由虚拟化处理单元13通过HRTF处理获得的音频信号也将被具体称为HRTF输出信号。HRTF输出信号是通过对象信号执行信号处理(例如,渲染处理和虚拟化处理)而获得的驱动信号,以通过驱动播放装置(例如,耳机)输出声音。

[0177] 在信号处理装置101中,从解码处理单元11提供给渲染处理单元12的对象信号是例如48kHz采样频率的低FS对象信号。

[0178] 在这种情况下,从渲染处理单元12提供给虚拟化处理单元13的虚拟扬声器信号也是具有48kHz的采样频率的信号,从虚拟化处理单元13提供给频带扩展单元41的HRTF输出信号的采样频率也是48kHz。

[0179] 在信号处理装置101中,渲染处理单元12和虚拟化处理单元13可用作基于元数据(对象位置信息)、个人HRTF系数和对象信号执行包括渲染处理和虚拟化处理的信号处理并生成HRTF输出信号的信号处理单元。在这种情况下,信号处理只需要至少包括虚拟化处理。

[0180] 个人高频带信息生成单元121基于从解码处理单元11提供的通用高频带信息和从HRTF系数记录单元122提供的通用HRTF系数和个人HRTF系数生成个人高频带信息,并将个人高频带信息提供给频带扩展单元41。

[0181] 个人高频带信息是用于扩展通过使用个人HRTF系数执行HRTF处理所获得的HRTF输出信号的频带的高频带信息。

[0182] HRTF系数记录单元122根据需要记录(保持)预先记录或从外部装置获取的通用HRTF系数和个人HRTF系数。

[0183] HRTF系数记录单元122将记录的个人HRTF系数提供给虚拟化处理单元13,并将记录的通用HRTF系数和个人HRTF系数提供给个人高频带信息生成单元121。

[0184] 由于通用HRTF系数通常预先存储在播放装置的记录区域中,因此也可在本实例中用作播放装置的信号处理装置101的HRTF系数记录单元122中预先记录通用HRTF系数。

[0185] 此外,可从网络上的服务器等获取个人HRTF系数。

[0186] 在这种情况下,例如,用作播放装置的信号处理装置101本身或者连接至信号处理装置101的终端装置(例如,智能电话)通过成像生成诸如用户的面部图像或者耳朵图像的图像数据。

[0187] 然后,信号处理装置101将关于用户获得的图像数据发送到服务器,并且服务器基于从信号处理装置101接收的图像数据对保持的HRTF系数执行转换处理,由此针对每个单独的用户生成个人HRTF系数,并将个人HRTF系数发送到信号处理装置101。HRTF系数记录单元122以这种方式获取并记录从服务器发送并由信号处理装置101接收的个人HRTF系数。

[0188] 频带扩展单元41基于从个人高频带信息生成单元121提供的个人高频带信息对从虚拟化处理单元13提供的HRTF输出信号执行频带扩展处理,从而生成输出音频信号,并将输出音频信号输出到后续。例如,通过频带扩展单元41获得的输出音频信号是96kHz的采样频率的信号。

[0189] <个人高频带信息生成单元的配置实例>

[0190] 如上所述,个人高频带信息生成单元121基于通用高频带信息、通用HRTF系数和个人HRTF系数生成个人高频带信息。

[0191] 虽然个人高频带信息应该多路复用在输入比特流中,但是在这种情况下,每个用户的个人输入比特流必须保持在服务器上,就服务器的存储容量而言,这不是优选的。

[0192] 因此,根据本技术,通用高频带信息被复用在输入比特流中,并使用由个人高频带信息生成单元121通过一些方法获取的个人HRTF系数和通用HRTF系数来生成个人高频带信息。

[0193] 尽管个人高频带信息生成单元121中的个人高频带信息的生成可通过任何方法来实现,但在一个示例中,可使用诸如深度神经网络(DNN)之类的深度学习技术来实现。

[0194] 在此,作为实例,描述个人高频带信息生成单元121由DNN构成的情况。

[0195] 例如,个人高频带信息生成单元121通过基于配置预先通过机器学习生成的DNN的系数和作为DNN的输入的通用高频带信息、通用HRTF系数和个人HRTF系数,基于DNN(神经网络)执行算术运算来生成个人高频带信息。

[0196] 在这种情况下,个人高频带信息生成单元121被配置为例如如图8所示。

[0197] 个人高频带信息生成单元121包括多层感知器(MLP) 151、MLP 152、递归神经网络(RNN) 153、特征量合成单元154和MLP 155。

[0198] MLP 151是由非线性激活的三层或更多层节点(即,输入层、输出层和一个或多个隐藏层)构成的MLP。MLP是在DNN中通常使用的技术之一。

[0199] MLP 151通过将HRTF系数记录单元122提供的通用HRTF系数视为用作MLP的输入的向量gh_in并基于向量gh_in执行算术运算来生成(计算)作为指示通用HRTF系数的一些特征的数据的向量gh_out,并将向量gh_out提供给特征量合成单元154。

[0200] 应注意,用作MLP的输入的向量gh_in可以是通用HRTF系数本身或者可以通过对通用HRTF系数执行一些预处理以在后阶段减少计算资源而获得的特征量。

[0201] MLP 152是与MLP 151相似的MLP,通过将HRTF系数记录单元122提供的个人HRTF

系数视为用作MLP的输入的向量ph_in并基于向量ph_in执行算术运算来生成作为指示个人HRTF系数的一些特征的数据的向量ph_out,并将向量ph_out提供给特征量合成单元154。

[0202] 应注意,向量ph_in还可以是个人HRTF系数本身或者可以是通过对个人HRTF系数执行一些预处理而获得的特征量。

[0203] 例如,RNN 153通常是由三个层(即,输入层、隐藏层和输出层)构成的RNN。RNN适于将隐藏层的输出反馈给隐藏层的输入,RNN具有适合时间序列数据的神经网络结构。

[0204] 注意,虽然这里将描述使用RNN来生成个人高频带信息的示例,但是本技术不依赖于DNN的配置作为个人高频带信息生成单元121,并且例如可使用作为适合于长期时间序列数据的神经网络结构的长短期记忆(LSTM)来代替RNN。

[0205] RNN 153通过将解码处理单元11提供的通用高频带信息视为作为输入的向量ge_in(n)并基于向量ge_in(n)执行算术运算来生成(计算)向量ge_out(n),并且将向量ge_out(n)提供至特征量合成单元154,向量ge_out(n)是指示通用高频带信息的一些特征的数据。

[0206] 注意,向量ge_in(n)和向量ge_out(n)中的n表示对象信号的时间帧的索引。具体地讲,RNN 153使用与多个帧相对应的向量ge_in(n)来生成一个帧的个人高频带信息。

[0207] 特征量合成单元154执行从MLP 151提供的向量gh_out、从MLP 152提供的向量ph_out和从RNN 153提供的向量ge_out(n)的向量串接,从而生成一个向量co_out(n),并且将向量co_out(n)提供至MLP 155。

[0208] 应注意,虽然在此使用向量串接作为用于在特征量合成单元154中合成特征量的方法,但是本技术不限于此,并且向量co_out(n)可以通过任何其他方法生成。例如,特征量合成单元154可通过称为最大池化的方法来执行特征量合成,使得向量被合成为能够充分表达特征的紧凑尺寸。

[0209] MLP 155是包括输入层、输出层以及一个或多个隐藏层的MLP,例如,基于从特征量合成单元154提供的向量co_out(n)执行算术运算,并且将因此获得的向量pe_out(n)作为个人高频带信息提供给频带扩展单元41。

[0210] 配置MLP和RNN的系数(诸如配置用作如上所述的个人高频带信息生成单元121的DNN的MLP 151、MLP 152、RNN 153和MLP 155)可通过预先使用训练数据执行机器学习来获得。

[0211] <输入比特流的语法实例>

[0212] 为了生成个人高频带信息,信号处理装置101需要通用高频带信息,并且输入比特流存储通用高频带信息。

[0213] 在此处,在图9中示出了提供给解码处理单元11的输入比特流的语法实例,即,输入比特流的格式实例。

[0214] 在图9中,“num_objects”表示对象的总数,“object_compressed_data”表示编码(压缩)对象信号。

[0215] 而且,“position_azimuth”表示在对象的球面坐标系中的水平角,“position_elevation”表示在对象的球面坐标系中的垂直角,并且“position_radius”表示从球面坐标系的原点到对象的距离(半径)。这里,包括水平角、垂直角和距离的信息是指示对象的位置的对象位置信息。

[0216] 因此,在该实例中,编码对象信号和对应于由“num_objects”表示的对象的数量的对象位置信息包括在输入比特流中。

[0217] 此外,在图9中,“num_output”表示输出声道的数量,即,HRTF输出信号的声道的数量,“output_bwe_data”表示通用高频带信息。因此,在该实例中,针对HRTF输出信号的各个声道存储通用高频带信息。

[0218] <信号生成处理的说明>

[0219] 接下来,将描述信号处理装置101中的操作。换言之,下面将参考图10中的流程图描述由信号处理装置101执行的信号生成处理。

[0220] 在步骤S11中,解码处理单元11对所提供的输入比特流执行解复用和解码处理,将由此获得的通用高频带信息提供给个人高频带信息生成单元121,并且将对象位置信息和对象信号提供给渲染处理单元12。

[0221] 此处,例如,从输入比特流提取图9中所示的“output_bwe_data”表示的通用高频带信息,然后将其提供至个人高频带信息生成单元121。

[0222] 在步骤S12中,渲染处理单元12基于从解码处理单元11提供的对象位置信息和对象信号执行渲染处理,并且将由此获得的虚拟扬声器信号提供给虚拟化处理单元13。在步骤S12中,例如,执行诸如VBAP的渲染处理。

[0223] 在步骤S13中,虚拟化处理单元13进行虚拟化处理。在步骤S13中,例如,执行HRTF处理作为虚拟化处理。

[0224] 在这种情况下,虚拟化处理单元13执行以下处理作为HRTF处理:对从渲染处理单元12提供的每个虚拟扬声器的虚拟扬声器信号与从HRTF系数记录单元122提供的每个声道的每个虚拟扬声器的个人HRTF系数进行卷积,并将针对每个声道由此获得的信号相加。虚拟化处理单元13将通过HRTF处理获得的HRTF输出信号提供至频带扩展单元41。

[0225] 在步骤S14中,个人高频带信息生成单元121基于从解码处理单元11提供的通用高频带信息以及从HRTF系数记录单元122提供的通用HRTF系数和个人HRTF系数生成个人高频带信息,并将个人高频带信息提供给频带扩展单元41。

[0226] 在步骤S14中,例如,配置DNN的个人高频带信息生成单元121的MLP 151至155生成个人高频带信息。

[0227] 具体地,MLP 151基于通用HRTF系数(即,从HRTF系数记录单元122提供的向量gh_in)执行算术运算,并将由此获得的向量gh_out提供给特征量合成单元154。

[0228] MLP 152基于个人HRTF系数(即,从HRTF系数记录单元122提供的向量ph_in)执行算术运算,并将由此获得的向量ph_out提供给特征量合成单元154。

[0229] RNN 153基于通用高频带信息(即,从解码处理单元11提供的向量ge_in(n))执行算术运算并且将因此获得的向量ge_out(n)提供至特征量合成单元154。

[0230] 此外,特征量合成单元154执行从MLP 151提供的向量gh_out、从MLP 152提供的向量ph_out和从RNN 153提供的向量ge_out(n)的向量串接,并且将由此获得的向量co_out(n)提供给MLP 155。

[0231] MLP 155基于从特征量合成单元154提供的向量co_out(n)执行算术运算,并且将因此获得的向量pe_out(n)作为个人高频带信息提供给频带扩展单元41。

[0232] 在步骤S15中,频带扩展单元41基于从个人高频带信息生成单元121提供的个人高

频带信息对由虚拟化处理单元13提供的HRTF输出信号执行频带扩展处理,并且将由此获得的输出音频信号输出到后级。一旦以这种方式生成输出音频信号,则结束信号生成处理。

[0233] 如上所述,信号处理装置101使用从输入比特流提取(读取)的通用高频带信息来生成个人高频带信息,使用个人高频带信息来执行频带扩展处理,从而生成输出音频信号。

[0234] 在这种情况下,通过对通过执行渲染处理和HRTF处理获得的低采样频率的HRTF输出信号执行频带扩展处理,可减小处理负荷(即,信号处理装置101的算术运算量)。

[0235] 此外,可通过生成与在HRTF处理中使用的个人HRTF系数相对应的个人高频带信息并执行频带扩展处理来获得高质量的输出音频信号。

[0236] 因此,即使当信号处理装置101是廉价的装置时,也可以高质量地执行音频播放。

[0237] <学习装置的配置实例>

[0238] 接下来,将描述生成配置作为个人高频带信息生成单元121的DNN(神经网络)的系数(即,配置MLP 151、MLP 152、RNN 153和MLP 155的系数)作为个人高频带信息生成系数数据的学习装置。

[0239] 例如,如图11中所示,配置这种学习装置。

[0240] 学习装置201包括渲染处理单元211、个人HRTF处理单元212、个人高频带信息计算单元213、通用HRTF处理单元214、通用高频带信息计算单元215和个人高频带信息学习单元216。

[0241] 渲染处理单元211基于提供的对象位置信息和对象信号执行与渲染处理单元12的情况相似的渲染处理,并将因此获得的虚拟扬声器信号提供给个人HRTF处理单元212和通用HRTF处理单元214。

[0242] 应注意,因为在渲染处理单元211的后期需要个人高频带信息作为训练数据,所以作为渲染处理单元211的输出的虚拟扬声器信号(即,作为渲染处理单元211的输入的对象信号)必然包括高频带信息。

[0243] 如果假设作为信号处理装置101的虚拟化处理单元13的输出的HRTF输出信号是48kHz的采样频率的信号,例如,输入到渲染处理单元211的对象信号的采样频率是96kHz等。

[0244] 在这种情况下,渲染处理单元211以96kHz的采样频率执行诸如VBAP的渲染处理,并且生成96kHz的采样频率的虚拟扬声器信号。

[0245] 应注意,尽管将在假设作为虚拟化处理单元13的输出的HRTF输出信号是48kHz的采样频率的信号的情况下给出以下描述,但是本技术中的每个信号的采样频率不限于该实例。例如,HRTF输出信号的采样频率可以是44.1kHz,输入到渲染处理单元211的对象信号的采样频率可以是88.2kHz。

[0246] 个人HRTF处理单元212基于提供的个人HRTF系数和从渲染处理单元211提供的虚拟扬声器信号执行HRTF处理(在下文中,具体地,也称为个人HRTF处理),并将由此获得的个人HRTF输出信号提供给个人高频带信息计算单元213。通过个人HRTF处理获得的个人HRTF输出信号是96kHz的采样频率的信号。

[0247] 在该实例中,渲染处理单元211和个人HRTF处理单元212可用作基于元数据(对象位置信息)、个人HRTF系数和对象信号执行包括渲染处理和虚拟化处理(个人HRTF处理)的信号处理并生成个人HRTF输出信号的一个信号处理单元。在这种情况下,信号处理只需要

包括至少虚拟化处理。

[0248] 个人高频带信息计算单元213基于从个人HRTF处理单元212提供的个人HRTF输出信号生成(计算)个人高频带信息,并将获得的个人高频带信息作为学习时的训练数据提供给个人高频带信息学习单元216。

[0249] 例如,个人高频带信息计算单元213获得如以上参考图5描述的个人HRTF输出信号的每个高频子频带的平均振幅值作为个人高频带信息。

[0250] 换言之,可通过将带通滤波器组应用于96kHz的采样频率的个人HRTF输出信号并计算高频子频带信号的时间帧的平均振幅值来生成每个高频子频带的高频子频带信号,来获得个人高频带信息。

[0251] 通用HRTF处理单元214基于提供的通用HRTF系数和从渲染处理单元211提供的虚拟扬声器信号执行HRTF处理(在下文中,也称为通用HRTF处理),并将因此获得的通用HRTF输出信号提供给通用高频带信息计算单元215。通用HRTF输出信号是96kHz的采样频率的信号。

[0252] 在该实例中,渲染处理单元211和通用HRTF处理单元214可用作基于元数据(对象位置信息)、通用HRTF系数和对象信号执行包括渲染处理和虚拟化处理(通用HRTF处理)的信号处理并生成通用HRTF输出信号的一个信号处理单元。在这种情况下,信号处理只需要包括至少虚拟化处理。

[0253] 通用高频带信息计算单元215基于从通用HRTF处理单元214提供的通用HRTF输出信号生成(计算)通用高频带信息,并将其提供给个人高频带信息学习单元216。通用高频带信息计算单元215执行与个人高频带信息计算单元213的情况下的计算类似的计算,并且生成通用高频带信息。

[0254] 输入比特流包括与由通用高频带信息计算单元215获得的通用高频带信息中的比特流相似的比特流,作为图9中所示的“output_bwe_data”。

[0255] 应注意,由通用HRTF处理单元214和通用高频带信息计算单元215执行的处理与由个人HRTF处理单元212和个人高频带信息计算单元213执行的处理被视为一对,并且处理基本上是相同的处理。

[0256] 处理的不同仅在于个人HRTF处理单元212的输入是个人HRTF系数,而通用HRTF处理单元214的输入是通用HRTF系数。换言之,只有输入的HRTF系数在它们之间是不同的。

[0257] 个人高频带信息学习单元216基于提供的通用HRTF系数和个人HRTF系数、从个人高频带信息计算单元213提供的个人高频带信息、以及从通用高频带信息计算单元215提供的通用高频带信息来执行学习(机器学习),并输出由此获得的个人高频带信息生成系数数据。

[0258] 具体地,个人高频带信息学习单元216使用个人高频带信息作为训练数据执行机器学习,并根据通用HRTF系数、个人HRTF系数和通用高频带信息生成用于生成个人高频带信息的个人高频带信息生成系数数据。

[0259] 如果图8中的个人高频带信息生成单元121的MLP 151、MLP 152、RNN 153和MLP 155使用配置由此获得的个人高频带信息生成系数数据的每个系数,则可以基于学习结果生成个人高频带信息。

[0260] 例如,通过评估作为个人高频带信息生成单元121的处理结果输出的向量pe_out

(n) 和作为训练数据的个人高频带信息的向量 $tpe_out(n)$ 之间的误差来执行由个人高频带信息学习单元216执行的学习处理。换言之, 执行学习, 使得向量 $pe_out(n)$ 与向量 $tpe_out(n)$ 之间的误差最小化。

[0261] 诸如配置DNN的MLP 151的每个元素的权重系数的初始值通常是随机的, 并且根据诸如基于时间的反向传播 (BPTT) 之类的误差反向传播方法的各种方法可应用于用于根据误差评估来调整每个系数的方法。

[0262] <学习处理的说明>

[0263] 接下来, 将描述学习装置201的操作。换言之, 将参考图12中的流程图描述由学习装置201执行的学习处理。

[0264] 在步骤S41中, 渲染处理单元211基于提供的对象位置信息和对象信号执行渲染处理, 并将因此获得的虚拟扬声器信号提供给个人HRTF处理单元212和通用HRTF处理单元214。

[0265] 在步骤S42中, 个人HRTF处理单元212基于提供的个人HRTF系数和从渲染处理单元211提供的虚拟扬声器信号执行个人HRTF处理, 并将因此获得的个人HRTF输出信号提供给个人高频带信息计算单元213。

[0266] 在步骤S43中, 个人高频带信息计算单元213基于从个人HRTF处理单元212提供的个人HRTF输出信号计算个人高频带信息, 并将由此获得的个人高频带信息作为训练数据提供给个人高频带信息学习单元216。

[0267] 在步骤S44中, 通用HRTF处理单元214基于提供的通用HRTF系数和从渲染处理单元211提供的虚拟扬声器信号执行通用HRTF处理, 并将由此获得的通用HRTF输出信号提供给通用高频带信息计算单元215。

[0268] 在步骤S45中, 通用高频带信息计算单元215基于从通用HRTF处理单元214提供的通用HRTF输出信号计算通用高频带信息, 并将结果提供至个人高频带信息学习单元216。

[0269] 在步骤S46中, 个人高频带信息学习单元216基于提供的通用HRTF系数和个人HRTF系数、从个人高频带信息计算单元213提供的个人高频带信息以及从通用高频带信息计算单元215提供的通用高频带信息来执行学习, 并生成个人高频带信息生成系数数据。

[0270] 在学习时, 生成用于使用通用高频带信息、通用HRTF系数和个人HRTF系数作为输入并且使用作为训练数据的个人高频带信息作为输出来实现DNN的个人高频带信息生成系数数据。一旦以这种方式生成个人高频带信息生成系数数据, 学习处理结束。

[0271] 如上所述, 学习装置201基于通用HRTF系数、个人HRTF系数和对象信号执行学习, 并生成个人高频带信息生成系数数据。

[0272] 因此, 个人高频带信息生成单元121可根据预测从输入的通用高频带信息、通用HRTF系数和个人HRTF系数获得与个人HRTF系数相对应的更适当的个人高频带信息。

[0273] <编码器的配置实例>

[0274] 接下来, 将描述生成在图9中示出的格式的输入比特流的编码器 (编码装置)。例如, 这种编码器被配置为如图13中所示。

[0275] 图13中示出的编码器301包括对象位置信息编码单元311、下采样器312、对象信号编码单元313、渲染处理单元314、通用HRTF处理单元315、通用高频带信息计算单元316和复用单元317。

[0276] 将作为编码目标的对象的对象信号和指示对象的位置的对象位置信息输入(提供)到编码器301。

[0277] 这里,输入到编码器301的对象信号例如是采样频率为96kHz的信号(FS96K对象信号)。

[0278] 对象位置信息编码单元311编码输入的对象位置信息并且将其提供至复用单元317。

[0279] 通过这种方式,例如,获得包括在图9中所示的水平角“position_azimuth”、垂直角“position_elevation”、以及半径“position_radius”的编码对象位置信息(对象位置数据)作为编码对象信息。

[0280] 下采样器312对96kHz的采样频率的输入对象信号执行下采样处理(即,频带限制),并将因此获得的48kHz的采样频率的对象信号(FS48K对象信号)提供给对象信号编码单元313。

[0281] 对象信号编码单元313对从下采样器312提供的48kHz的对象信号进行编码并且将其提供给复用单元317。通过这种方式,例如,获得图9中所示的“object_compressed_data”作为编码对象信号。

[0282] 注意,对象信号编码单元313中的编码方案可以是MPEG-H第3部分:3D音频标准的编码方案或者可以是另一编码方案。换言之,仅需要对对象信号编码单元313中的编码方案和解码处理单元11中的解码方案彼此对应(基于相同的标准)。

[0283] 渲染处理单元314基于输入对象位置信息和96kHz的对象信号执行诸如VBAP的渲染处理,并将因此获得的虚拟扬声器信号提供给通用HRTF处理单元315。

[0284] 应注意,由渲染处理单元314执行的渲染处理不限于VBAP,并且可以是任何其他渲染处理,只要该处理与在解码侧(播放侧)的信号处理装置101的渲染处理单元12的情况相同。

[0285] 通用HRTF处理单元315对从渲染处理单元314提供的虚拟扬声器信号使用通用HRTF系数执行HRTF处理,并将因此获得的96kHz的通用HRTF输出信号提供至通用高频带信息计算单元316。

[0286] 通用HRTF处理单元315执行与图11中由通用HRTF处理单元214执行的通用HRTF处理相似的处理。

[0287] 通用高频带信息计算单元316基于从通用HRTF处理单元315提供的通用HRTF输出信号计算通用高频带信息,对所获得的通用高频带信息进行压缩编码,并将其提供给复用单元317。

[0288] 例如,由通用高频带信息计算单元316生成的通用高频带信息是图5中示出的每个高频子频带的平均振幅信息(平均振幅值)。

[0289] 例如,通用高频带信息计算单元316对96kHz的输入通用HRTF输出信号执行基于带通滤波器组的滤波,并获得每个高频子频带的高频子频带信号。然后,通用高频带信息计算单元316计算每个高频子频带信号的时间帧的平均振幅值,并且由此生成通用高频带信息。

[0290] 以这种方式,例如,获得图9中所示的“output_bwe_data”作为编码的通用高频带信息。

[0291] 复用单元317对从对象位置信息编码单元311提供的编码对象位置信息、从对象信

号编码单元313提供的编码对象信号以及从通用高频带信息计算单元316提供的编码通用高频带信息进行复用。

[0292] 复用单元317输出通过复用对象位置信息、对象信号和通用高频带信息而获得的输出比特流。输出比特流作为输入比特流输入到信号处理装置101。

[0293] <编码处理的说明>

[0294] 接下来,将描述编码器301的操作。换言之,下面参考图14中的流程图描述由编码器301执行的编码处理。

[0295] 在步骤S71中,对象位置信息编码单元311对输入的对象位置信息进行编码,并将其提供给复用单元317。

[0296] 在步骤S72中,下采样器312对输入的对象信号进行下采样并且将其提供给对象信号编码单元313。

[0297] 在步骤S73中,对象信号编码单元313对从下采样器312提供的对象信号进行编码并且将其提供给复用单元317。

[0298] 在步骤S74中,渲染处理单元314基于输入的对象位置信息和对象信号执行渲染处理,并将因此获得的虚拟扬声器信号提供给通用HRTF处理单元315。

[0299] 在步骤S75,通用HRTF处理单元315对从渲染处理单元314提供的虚拟扬声器信号使用通用HRTF系数执行HRTF处理,并将获得的通用HRTF输出信号作为结果提供给通用高频带信息计算单元316。

[0300] 在步骤S76中,通用高频带信息计算单元316基于从通用HRTF处理单元315提供的通用HRTF输出信号计算通用高频带信息,对所获得的通用高频带信息进行压缩编码,并将其提供给复用单元317。

[0301] 在步骤S77中,复用单元317对从对象位置信息编码单元311提供的编码对象位置信息、从对象信号编码单元313提供的编码对象信号以及从通用高频带信息计算单元316提供的编码通用高频带信息进行多路复用。

[0302] 复用单元317输出通过多路复用获得的输出比特流,编码处理结束。

[0303] 如上所述,编码器301计算通用高频带信息并将其存储在输出比特流中。

[0304] 以这种方式,可以使用解码侧的输出比特流的通用高频带信息来生成个人高频带信息。以这种方式,即使在解码侧使用廉价的装置,也能够以高质量执行音频播放。

[0305] 应注意,上面已经描述了从音频对象的对象信号生成作为频带扩展的对象的HRTF输出信号的实例。

[0306] 然而,本技术不限于此,例如,可从基于声道的每个声道的音频信号(在下文中,也称为声道信号)生成HRTF输出信号,并可对HRTF输出信号执行频带扩展。

[0307] 在这种情况下,信号处理装置101未设置有渲染处理单元12,并且输入比特流包括编码的声道信号。

[0308] 然后,将通过解码处理单元11对输入的比特流执行解复用和解码处理所获得的具有多声道配置的每个声道的声道信号提供给虚拟化处理单元13。每个声道的声道信号对应于每个虚拟扬声器的虚拟扬声器信号。

[0309] 虚拟化处理单元13对从解码处理单元11提供的声道信号和从HRTF系数记录单元122提供的每个声道的个人HRTF系数进行卷积并对由此获得的信号进行相加的处理作为

HRTF处理。虚拟化处理单元13将通过这种HRTF处理获得的HRTF输出信号提供至频带扩展单元41。

[0310] 此外,在信号处理装置101中从声道信号生成HRTF输出信号的情况下,学习装置201未设置有渲染处理单元211,高采样频率的声道信号(即,包括高频带信息的声道信号)被提供给个人HRTF处理单元212和通用HRTF处理单元214。

[0311] 此外,例如,高阶立体声(HOA)渲染处理可以由渲染处理单元12执行。

[0312] 在这种情况下,渲染处理单元12通过从解码处理单元11提供的立体混合音格式(即,例如基于球面调和域中的音频信号)执行渲染处理,由此在球面调和域中生成虚拟扬声器信号,且将其提供到虚拟化处理单元13。

[0313] 虚拟化处理单元13基于从渲染处理单元12提供的球面调和域中的虚拟扬声器信号和从HRTF系数记录单元122提供的球面调和域中的个人HRTF系数在球面调和域中执行HRTF处理,并将由此获得的HRTF输出信号提供给频带扩展单元41。此时,球面调和域中的HRTF输出信号可被提供给频带扩展单元41,或者可根据需要通过执行转换等而获得的时间区域中的HRTF输出信号可被提供给频带扩展单元41。

[0314] 如上所述,根据本技术,可使用个人HRTF处理之后的信号的个人高频带信息而不是解码侧(播放侧)的对象信号的高频带信息来执行频带扩展处理。

[0315] 此外,由于在这种情况下不需要多路复用输入比特流的个人高频带信息,所以可以减少服务器等的消耗量,即,编码器301的存储,并且还可以抑制编码器301中用于编码处理(编码处理)的处理时间的增加。

[0316] 此外,在播放装置侧(即,在信号处理装置101侧)以低采样频率执行解码处理、渲染处理和虚拟化处理,并且因此可以显著减少算术运算量。以这种方式,可以采用廉价的处理器,例如,以减少处理器所使用的电量,并且利用诸如智能电话的移动装置在较长的时间段内连续播放高分辨率声源。

[0317] <计算机的配置的实例>

[0318] 上述一系列处理也可以由硬件或软件执行。在由软件执行一系列处理的情况下,在计算机上安装配置软件的程序。这里,计算机包括例如内置在专用硬件中的计算机、其上安装各种程序以能够执行各种功能的通用个人计算机等。

[0319] 图15是示出使用程序执行上述一系列处理的计算机硬件的配置实例的框图。

[0320] 在计算机中,中央处理单元(CPU)501、只读存储器(ROM)502和随机存取存储器(RAM)503通过总线504彼此连接。

[0321] 输入/输出接口505进一步连接至总线504。输入单元506、输出单元507、记录单元508、通信单元509和驱动器510连接至输入/输出接口505。

[0322] 输入单元506包括键盘、鼠标、麦克风、成像元件等。输出单元507包括显示器、扬声器等。记录单元508包括硬盘、非易失性存储器等。通信单元509包括网络接口等。驱动器510驱动诸如磁盘、光盘、磁光盘或半导体存储器的可移除记录介质511。

[0323] 例如,在具有上述配置的计算机中,CPU 501经由输入/输出接口505和总线504将存储在记录单元508中的程序加载到RAM 503,并且执行该程序以执行上述一系列处理。

[0324] 例如,由计算机(CPU 501)执行的程序可记录在用作用于提供的封装介质的可移除记录介质511上。程序可以经由诸如局域网、因特网、或者数字卫星广播等有线或无线传

输介质提供。

[0325] 在计算机中,通过将可移除记录介质511安装在驱动器510上,可以经由输入/输出接口505将程序安装在记录单元508中。此外,程序可以由通信单元509经由有线或无线传输介质接收以安装在记录单元508中。可替换地,程序可以预先安装在ROM 502或记录单元508中。

[0326] 应注意,由计算机执行的程序可以是按照本说明书中描述的顺序按时间顺序执行的程序,或者可以是并行或在诸如调用时间等必要定时执行的程序。

[0327] 本技术的实施方式不限于上述实施方式,并且在不背离本技术的主旨的情况下,可以在本技术的范围内进行各种改变。

[0328] 例如,本技术可被配置为云计算,其中多个设备经由网络共享和协作处理一个功能。

[0329] 另外,上述流程图中描述的每个步骤可以由一个设备执行或者由多个设备以共享方式执行。

[0330] 此外,在一个步骤包括多个处理的情况下,包括在一个步骤中的多个处理可以由一个设备执行或者由多个设备以共享方式执行。

[0331] 此外,本技术可以配置为如下。

[0332] (1) 一种信号处理装置,包括:解码处理单元,将输入比特流解复用为第一音频信号、第一音频信号的元数据以及用于扩展频带的第一高频带信息;以及频带扩展单元,基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理来获得第二音频信号,基于第一高频带信息生成第二高频带信息。

[0333] (2) 根据(1)的信号处理装置,进一步包括:高频带信息生成单元,基于第一高频带信息生成第二高频带信息。

[0334] (3) 根据(2)的信号处理装置,其中,第一高频带信息是用于扩展通过使用第一系数执行信号处理而获得的第二音频信号的频带的高频带信息,第二高频带信息是用于扩展通过使用第二系数执行信号处理而获得的第二音频信号的频带的高频带信息,并且频带扩展单元基于第二音频信号和第二高频带信息执行频带扩展处理,通过基于第一音频信号、元数据和第二系数执行信号处理获得第二音频信号。

[0335] (4) 根据(3)的信号处理装置,其中,高频带信息生成单元基于第一高频带信息、第一系数和第二系数生成第二高频带信息。

[0336] (5) 根据(3)或(4)的信号处理装置,其中,高频带信息生成单元通过基于预先通过机器学习生成的系数、第一高频带信息、第一系数和第二系数执行算术运算生成第二高频带信息。

[0337] (6) 根据(5)的信号处理装置,其中,算术运算是基于神经网络的算术运算。

[0338] (7) 根据(3)至(6)中任一项的信号处理装置,其中,第一系数是通用系数,而第二系数是每个用户的系数。

[0339] (8) 根据(7)的信号处理装置,其中,第一系数和第二系数是HRTF系数。

[0340] (9) 根据(3)至(8)中任一项的信号处理装置,进一步包括:记录第一系数的系数记录单元。

[0341] (10) 根据(1)至(9)中任一项的信号处理装置,进一步包括:信号处理单元,通过执行信号处理生成第二音频信号。

[0342] (11) 根据(10)的信号处理装置,其中,信号处理是包括虚拟化处理的处理。

[0343] (12) 根据(11)的信号处理装置,其中,信号处理是包括渲染处理的处理。

[0344] (13) 根据(1)至(12)中任一项的信号处理装置,其中,第一音频信号是音频对象的对象信号或基于声道的音频信号。

[0345] (14) 一种信号处理方法,包括通过信号处理装置执行以下:将输入比特流解复用为第一音频信号、第一音频信号的元数据以及用于扩展频带的第一高频带信息;以及基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理来获得第二音频信号,基于第一高频带信息生成第二高频带信息。

[0346] (15) 一种使计算机执行处理的程序,处理包括以下步骤:将输入比特流解复用为第一音频信号、第一音频信号的元数据以及用于扩展频带的第一高频带信息;以及基于第二音频信号和第二高频带信息执行频带扩展处理,并且由此生成输出音频信号,通过基于第一音频信号和元数据执行信号处理来获得第二音频信号,基于第一高频带信息生成第二高频带信息。

[0347] (16) 一种学习装置,包括:第一高频带信息计算单元,基于通过基于第一音频信号和第一系数的信号处理生成的第二音频信号生成用于扩展频带的第一高频带信息;第二高频带信息计算单元,基于通过基于第一音频信号和第二系数的信号处理生成的第三音频信号生成用于扩展频带的第二高频带信息;以及高频带信息学习单元,基于第一系数、第二系数、第一高频带信息和第二高频带信息,使用第二高频带信息作为训练数据执行学习,并且生成用于从第一系数、第二系数和第一高频带信息获得第二高频带信息的系数数据。

[0348] (17) 根据(16)的学习装置,其中,系数数据是配置神经网络的系数。

[0349] (18) 根据(16)或(17)的学习装置,其中,第一系数是通用系数,而第二系数是每个用户的系数。

[0350] (19) 根据(18)的学习装置,其中,信号处理是包括虚拟化处理的处理,并且第一系数和第二系数是HRTF系数。

[0351] (20) 根据(19)的学习装置,其中,信号处理是包括渲染处理的处理。

[0352] (21) 根据(16)至(19)中任一项的学习装置,其中,第一音频信号是音频对象的对象信号或基于声道的音频信号。

[0353] (22) 一种学习方法,包括通过学习装置执行以下:基于通过基于第一音频信号和第一系数的信号处理生成的第二音频信号,生成用于扩展频带的第一高频带信息;基于通过基于第一音频信号和第二系数的信号处理生成的第三音频信号生成用于扩展频带的第二高频带信息;以及基于第一系数、第二系数、第一高频带信息和第二高频带信息,使用第二高频带信息作为训练数据执行学习,从而生成用于从第一系数、第二系数和第一高频带信息获得第二高频带信息的系数数据。

[0354] (23) 一种使计算机执行处理的程序,处理包括以下步骤:基于通过基于第一音频信号和第一系数的信号处理生成的第二音频信号生成用于扩展频带的第一高频带信息;基于通过基于第一音频信号和第二系数的信号处理生成的第三音频信号生成用于扩展频带

的第二高频带信息;以及基于第一系数、第二系数、第一高频带信息和第二高频带信息,使用第二高频带信息作为训练数据执行学习,从而生成用于从第一系数、第二系数和第一高频带信息获得第二高频带信息的系数数据。

[0355] [参考标号列表]

[0356] 11 解码处理单元

[0357] 12 渲染处理单元

[0358] 13 虚拟化处理单元

[0359] 41 频带扩展单元

[0360] 101 信号处理装置

[0361] 121 个人高频带信息生成单元

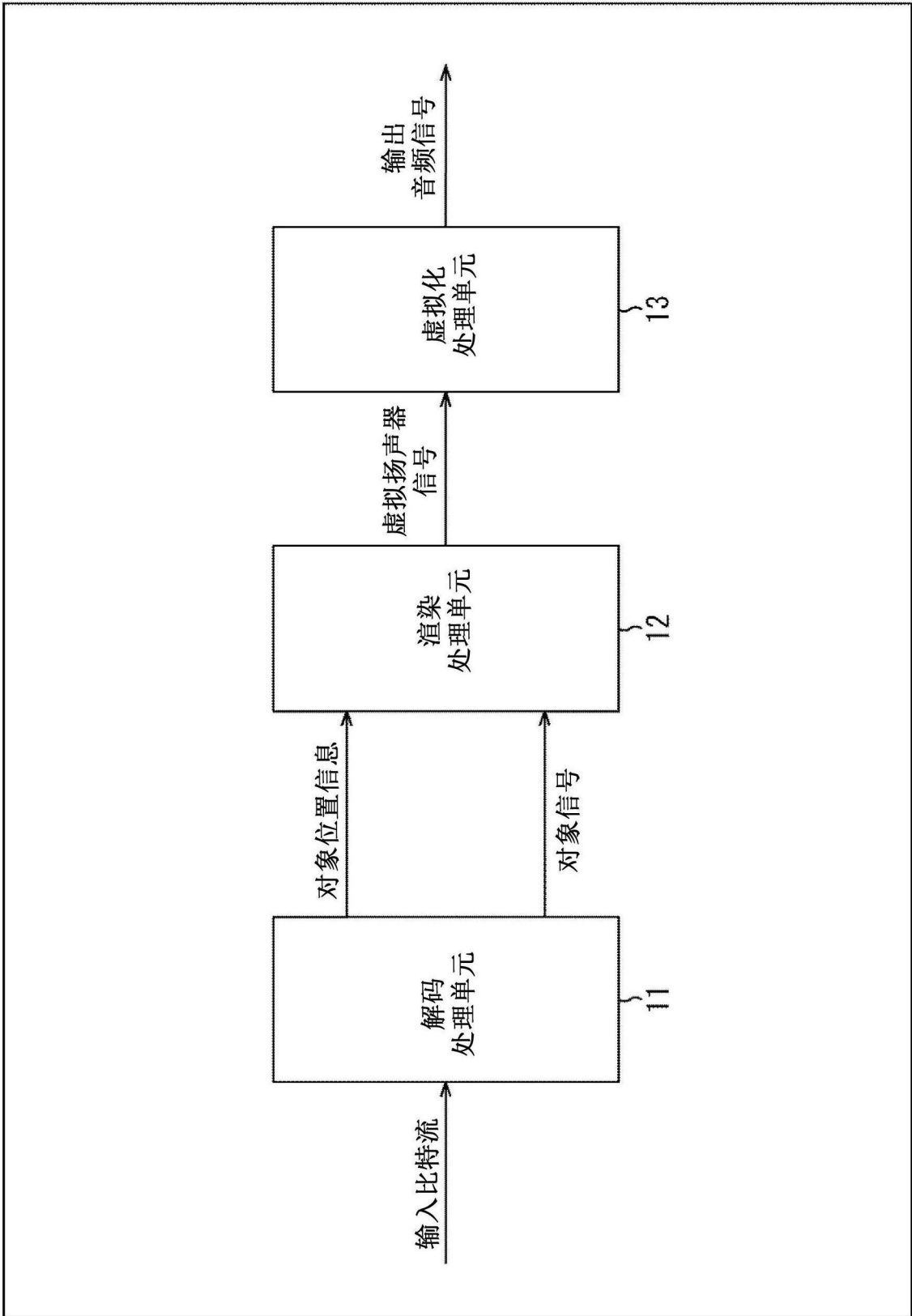


图1

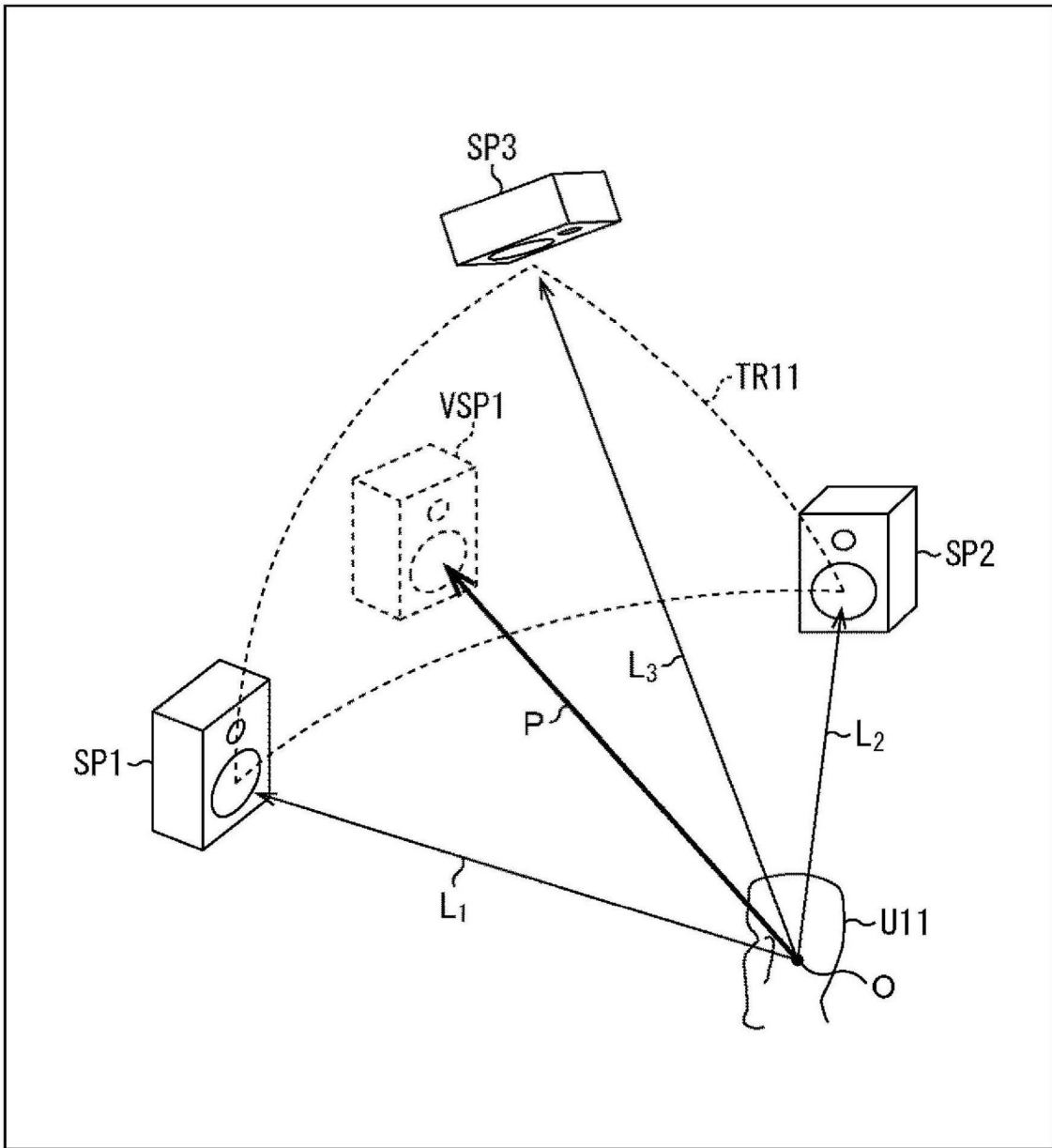


图2

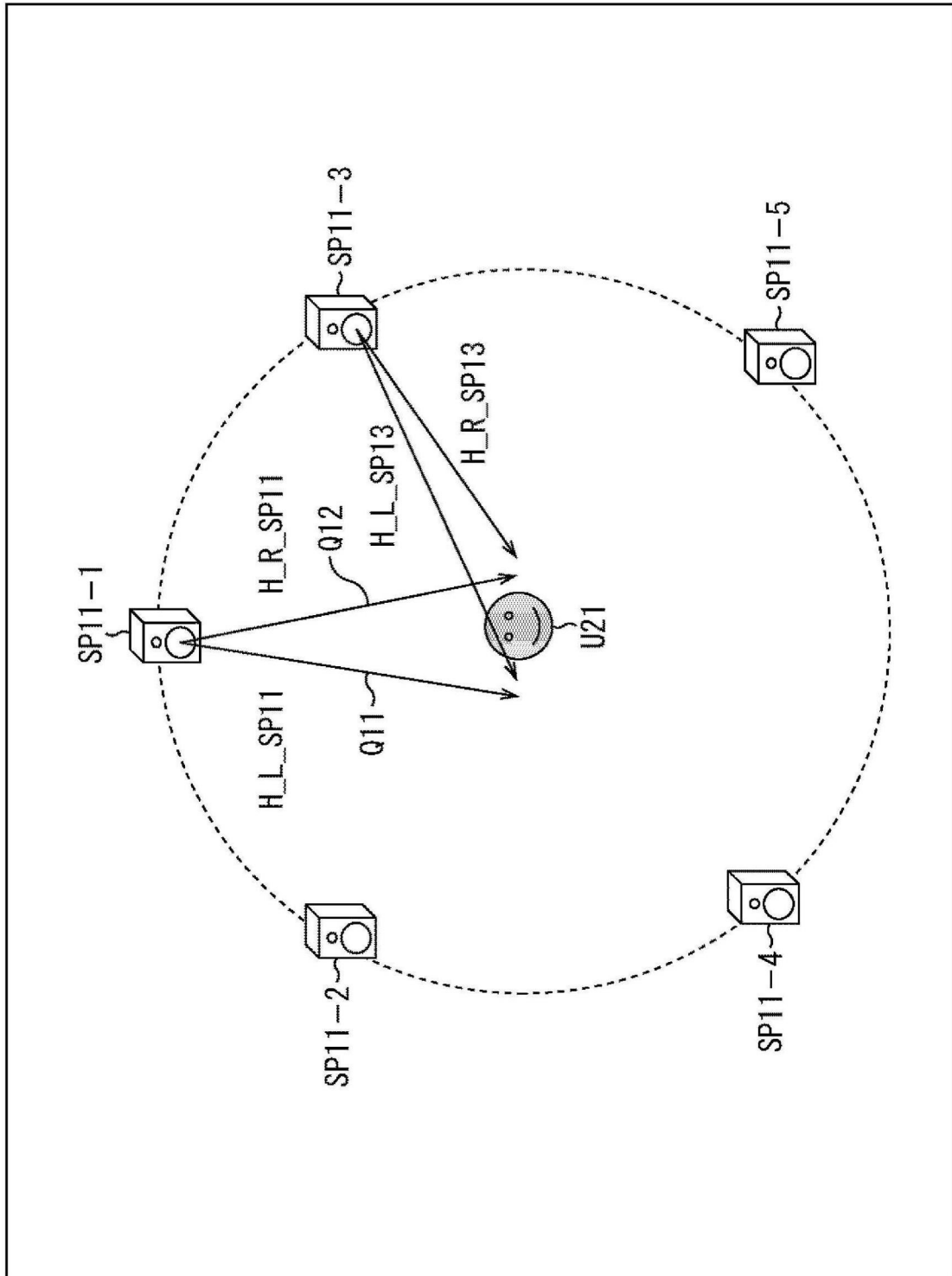


图3

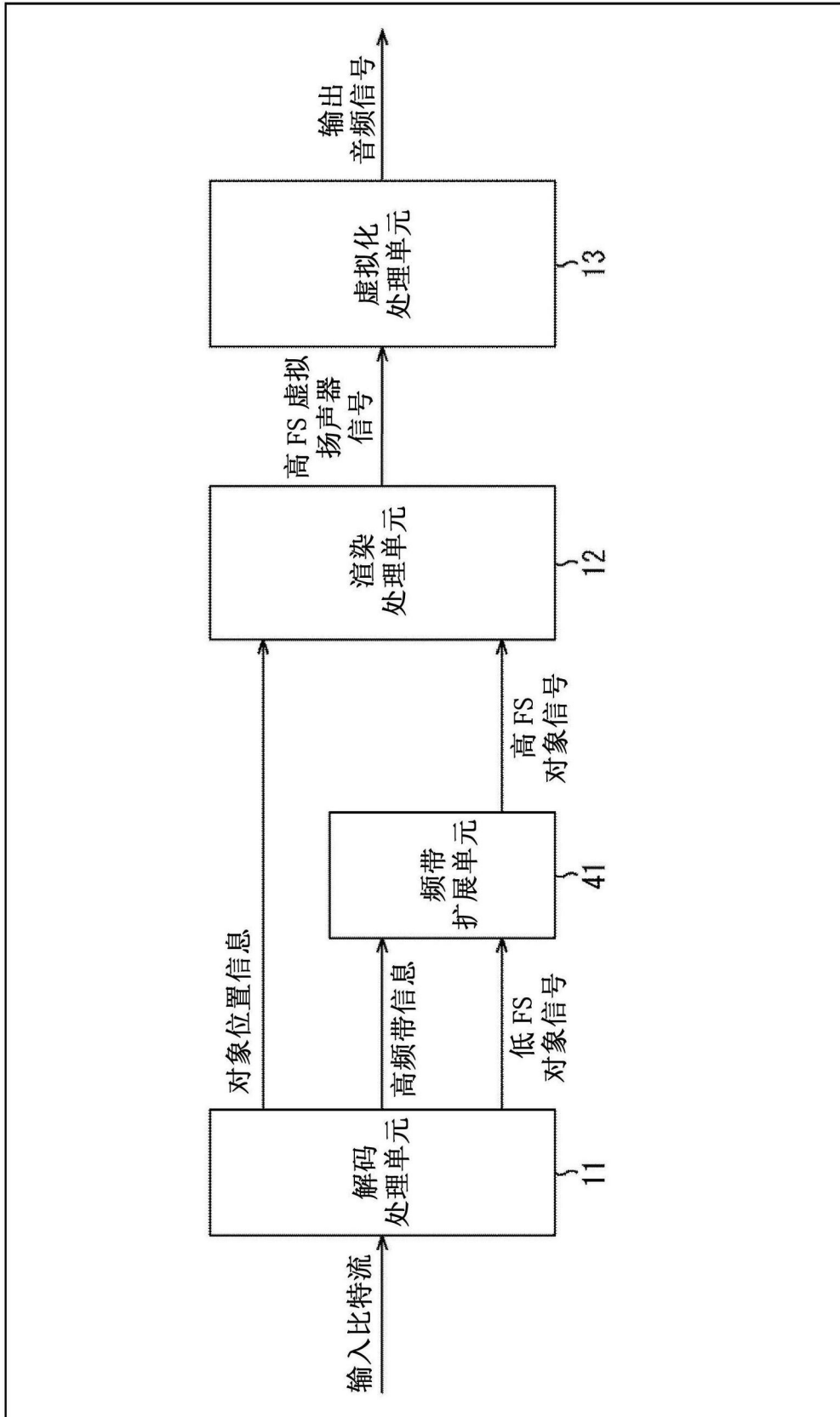


图4

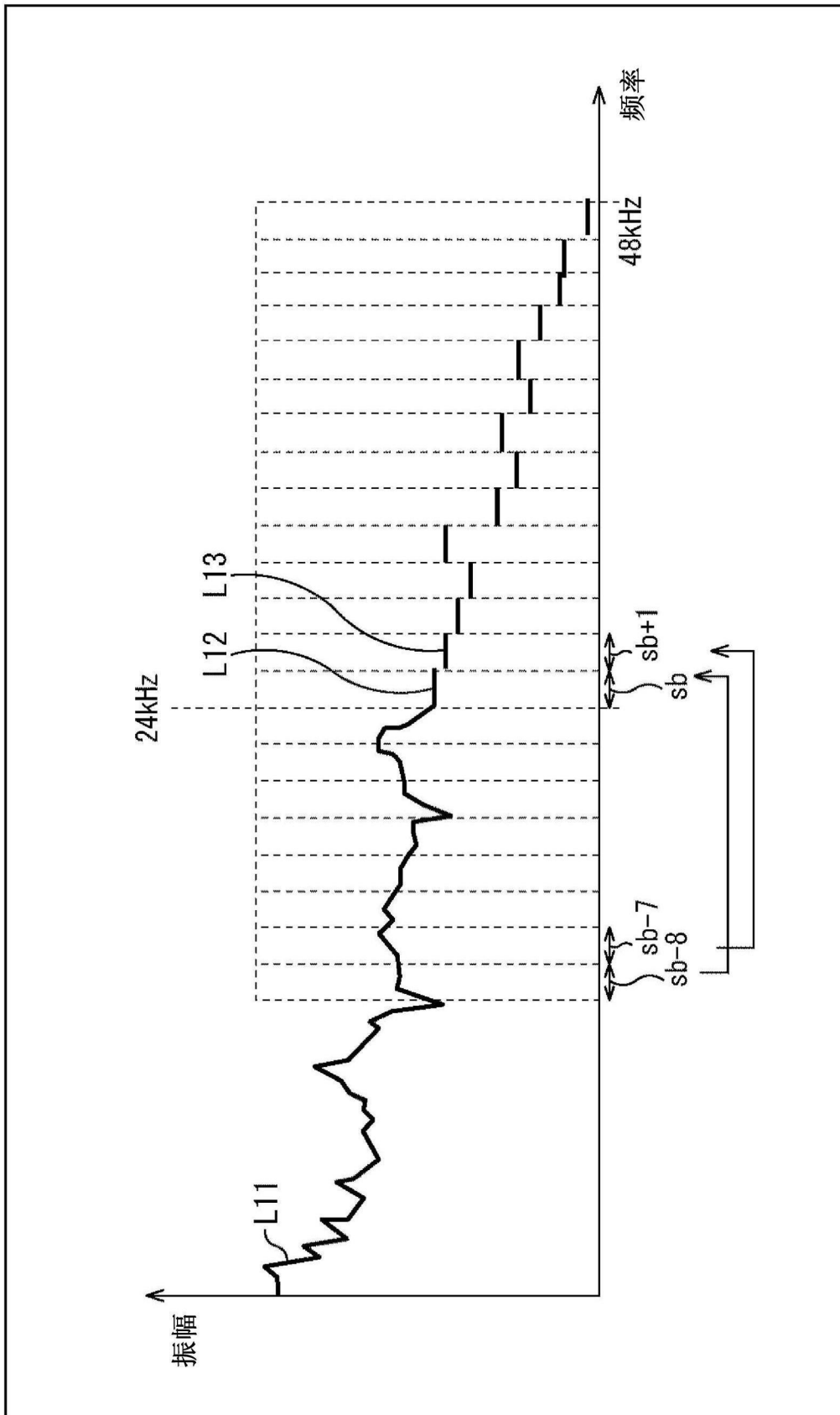


图5

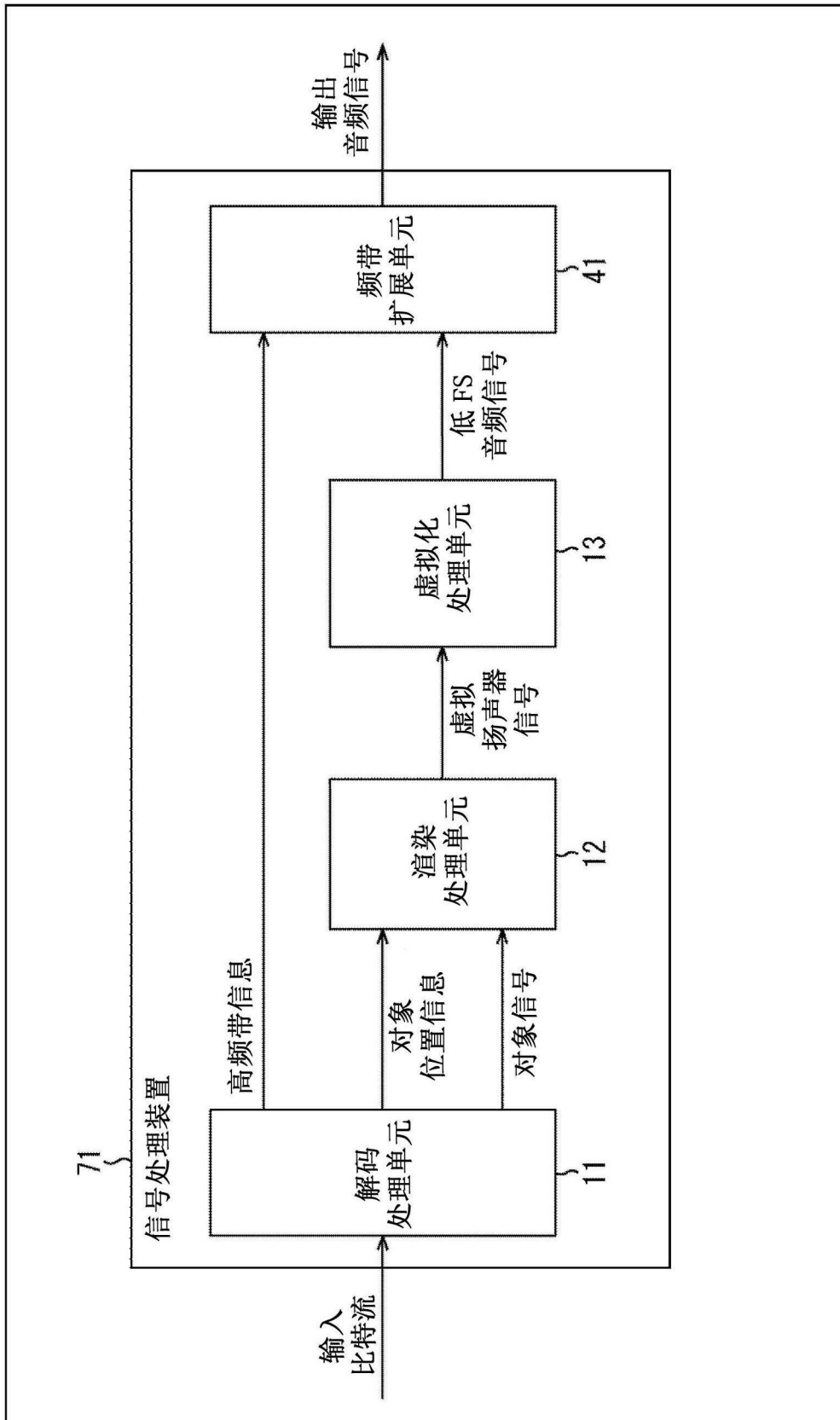


图6

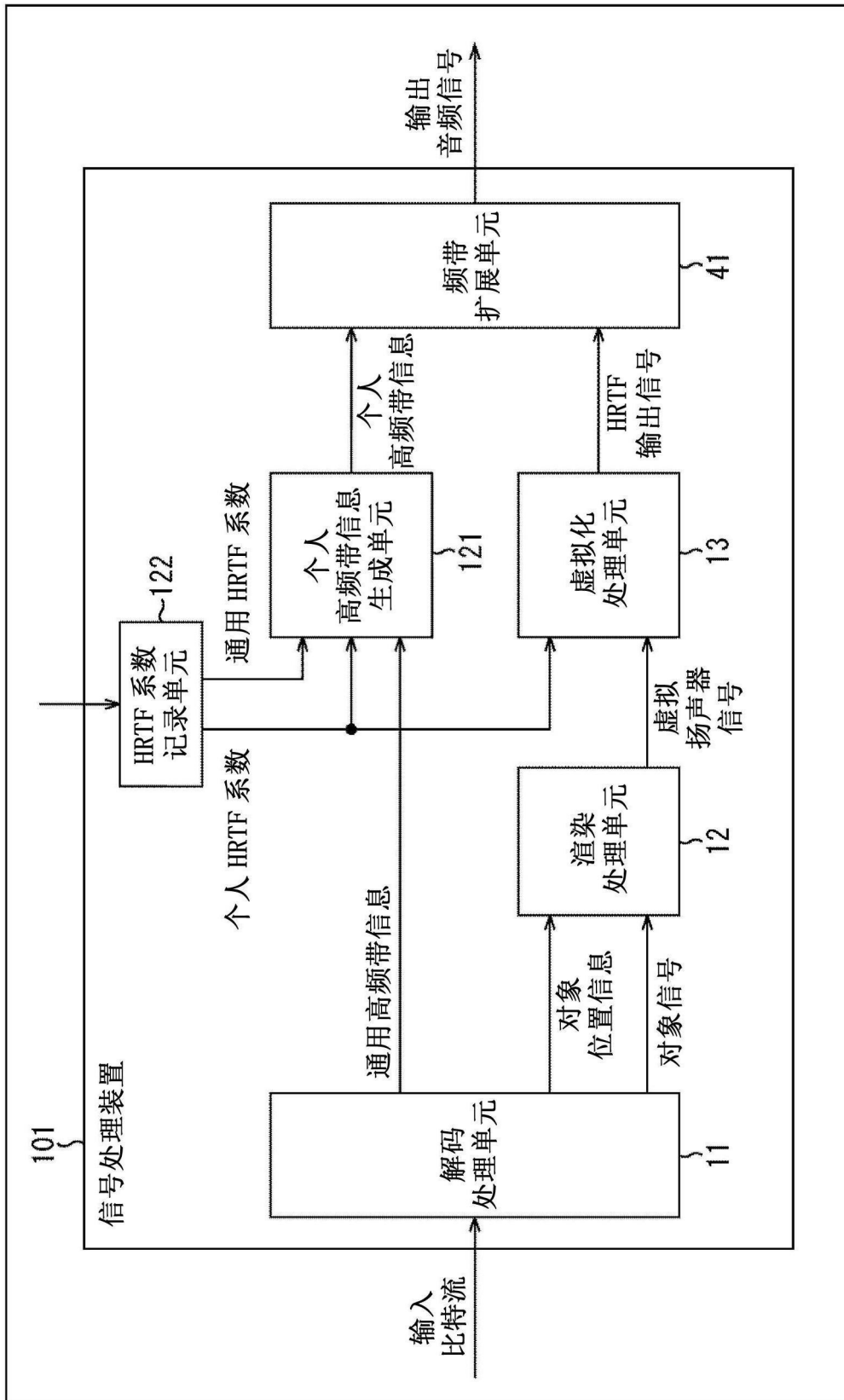


图7

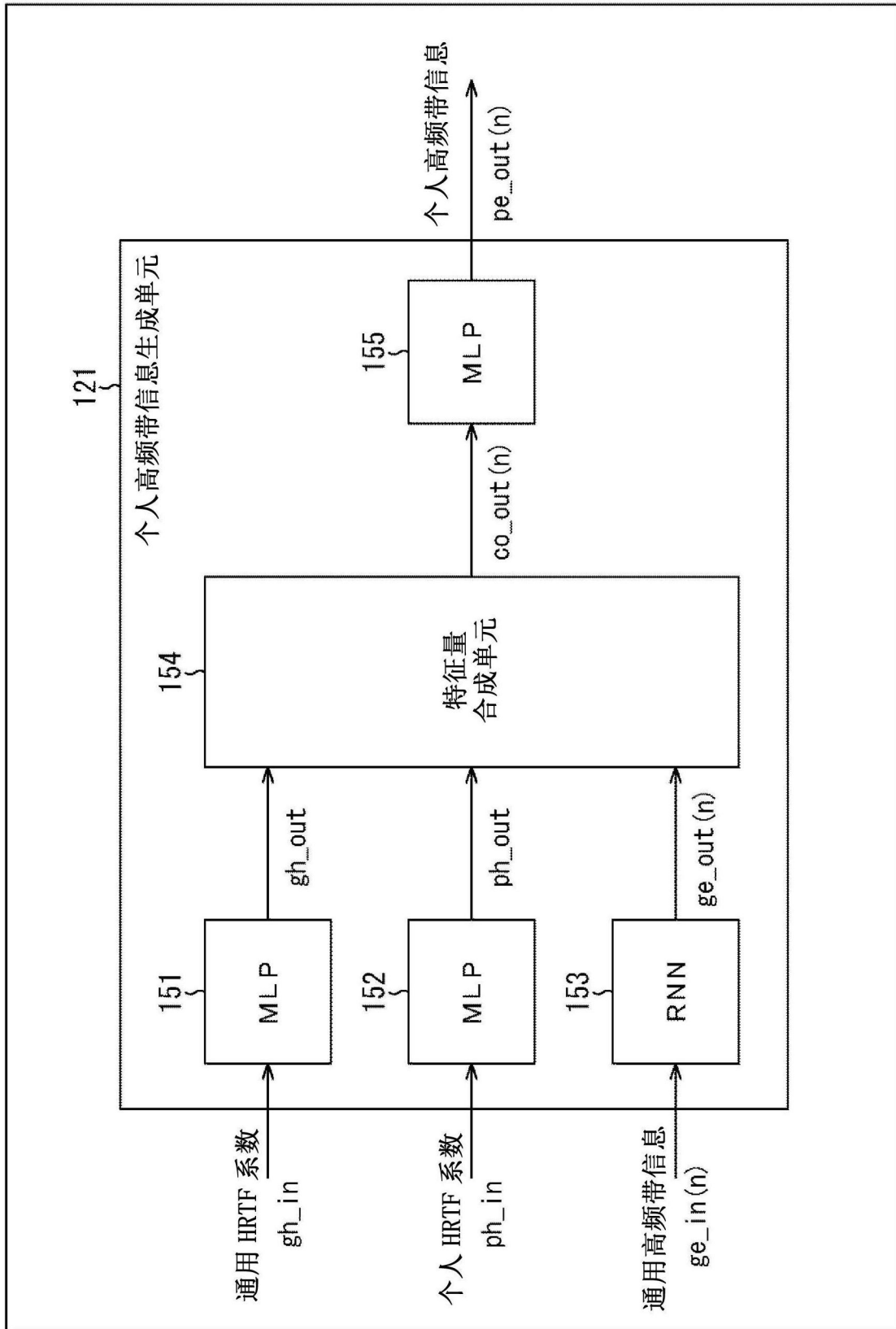


图8

语法	位数	Mnemonic
<pre> for (i=0; i < num_objects; i++) { object_compressed_data(i) } for (i=0; i < num_objects; i++) { position_azimuth[i]; position_elevation[i]; position_radius[i]; } for(i = 0; i < num_output; i++){ output_bwe_data(i) } </pre>	<p style="text-align: center;">8 6 4</p>	<p style="text-align: center;">tcimsbf tcimsbf uimsbf</p>

图9

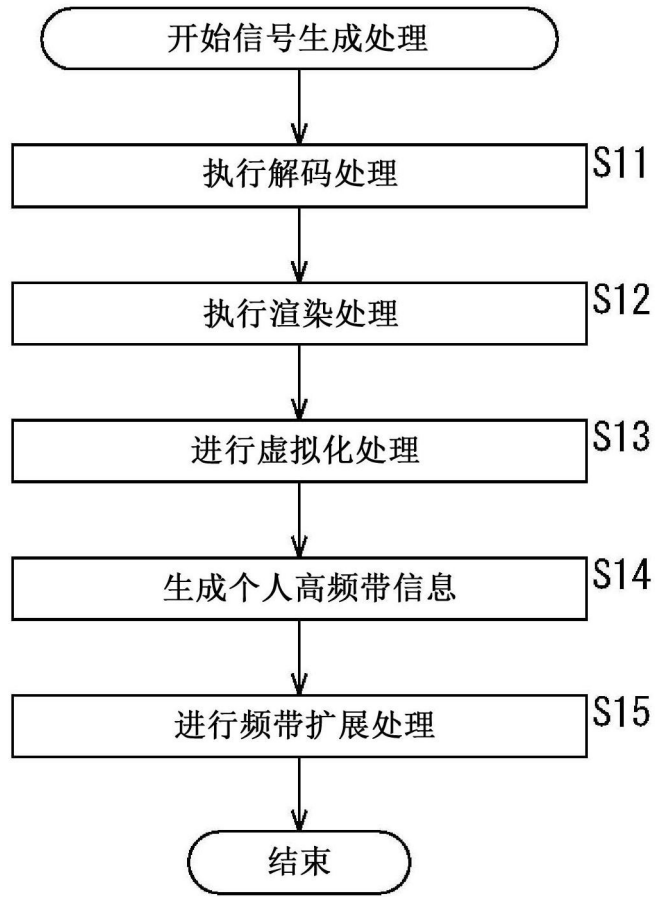


图10

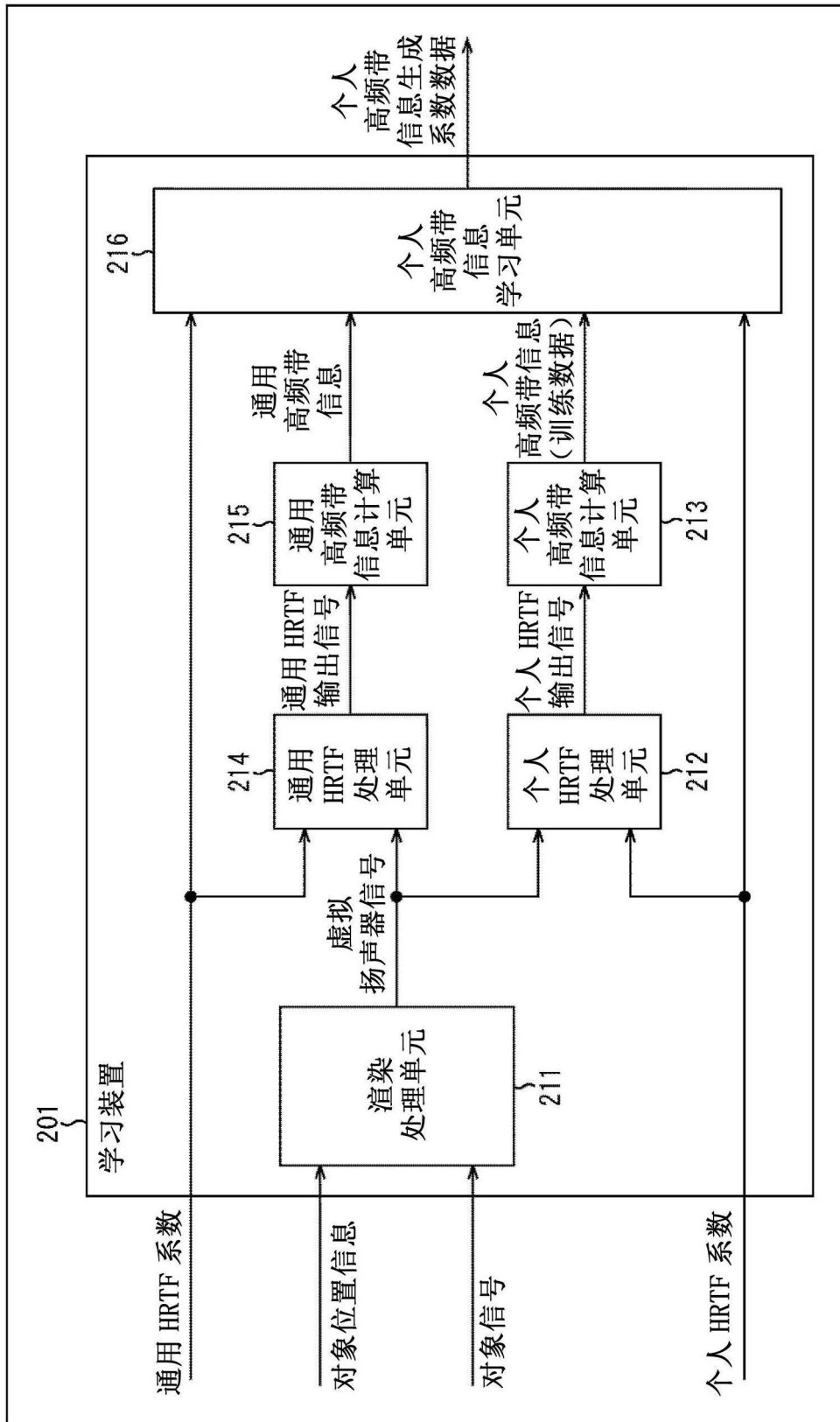


图11

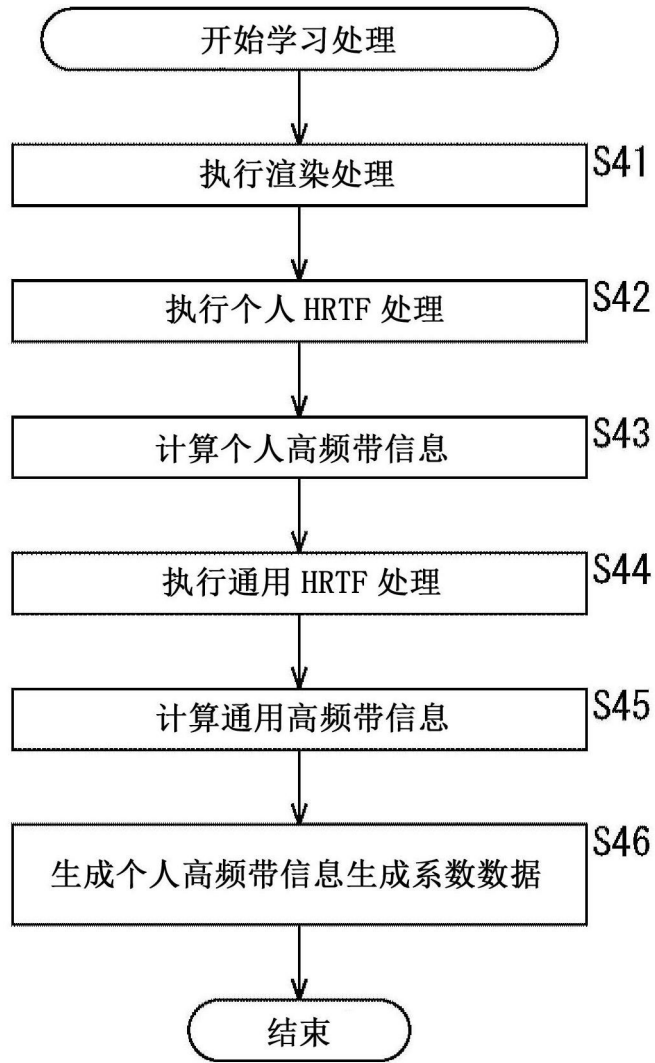


图12

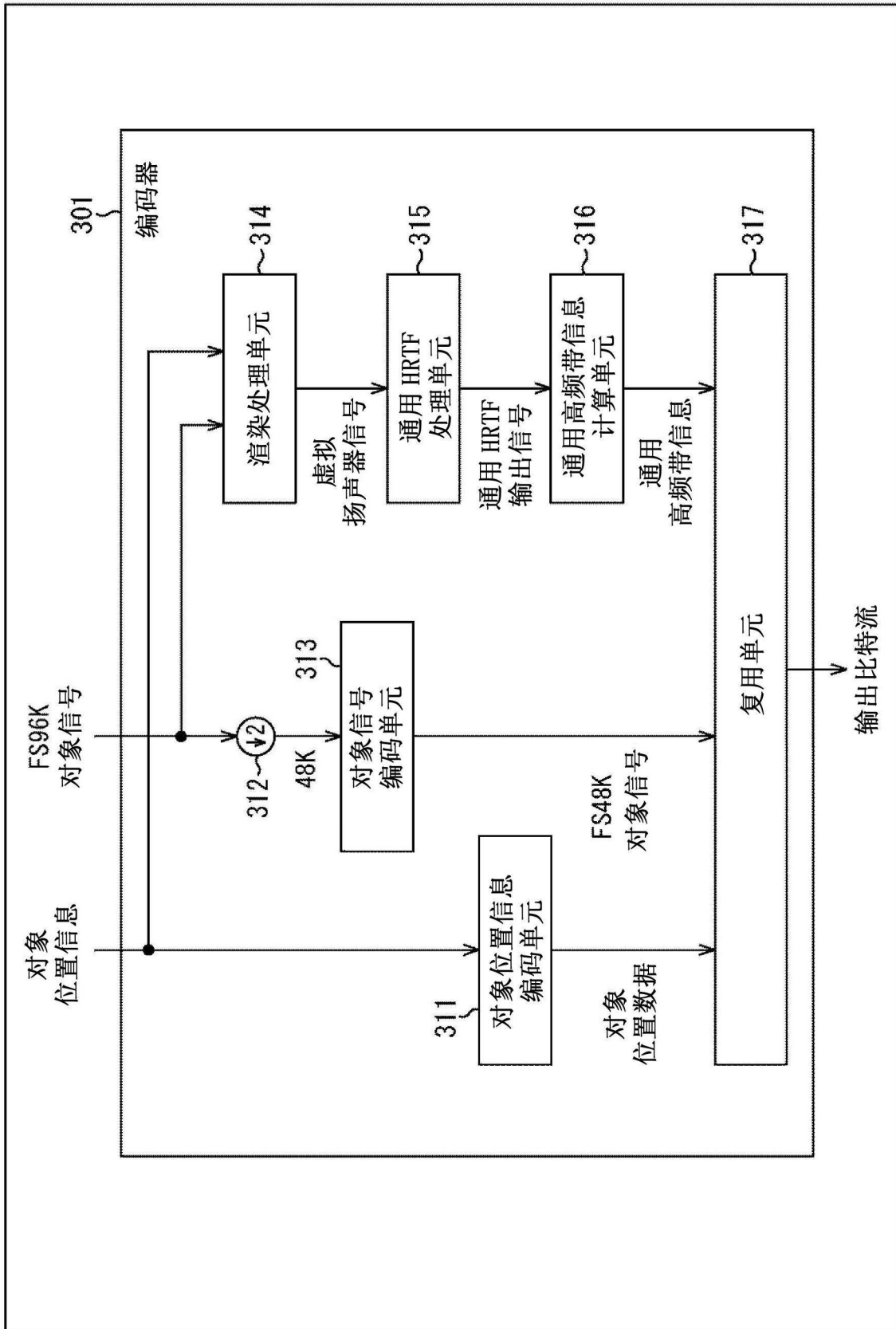


图13

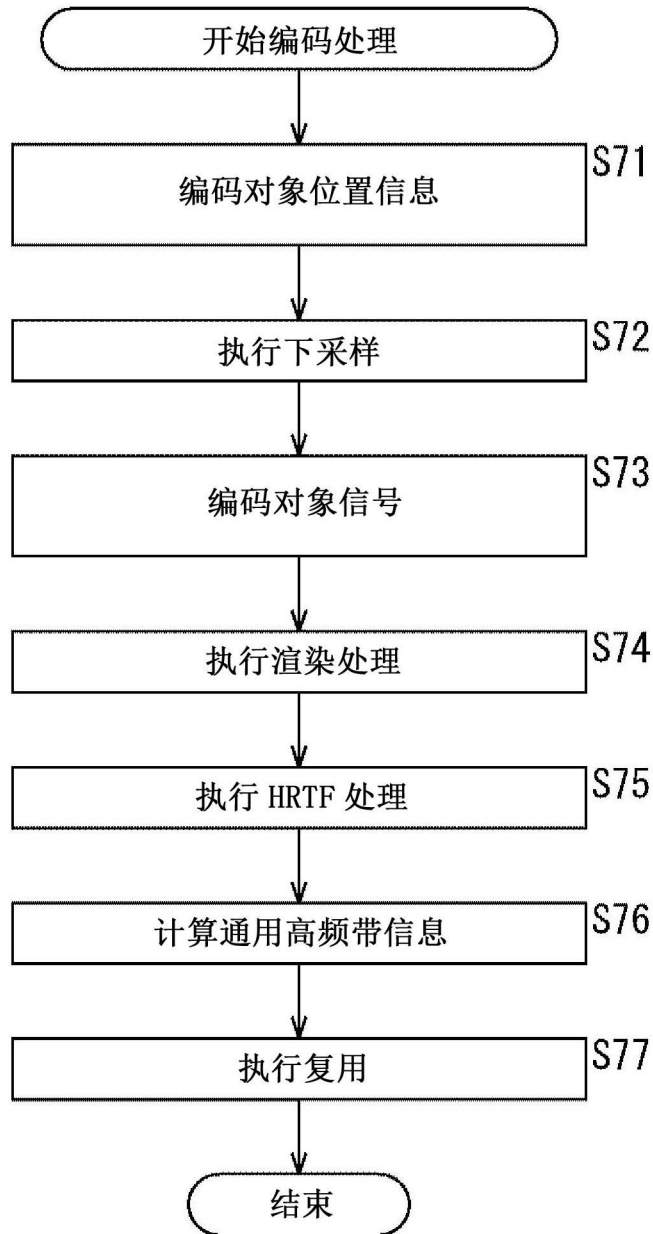


图14

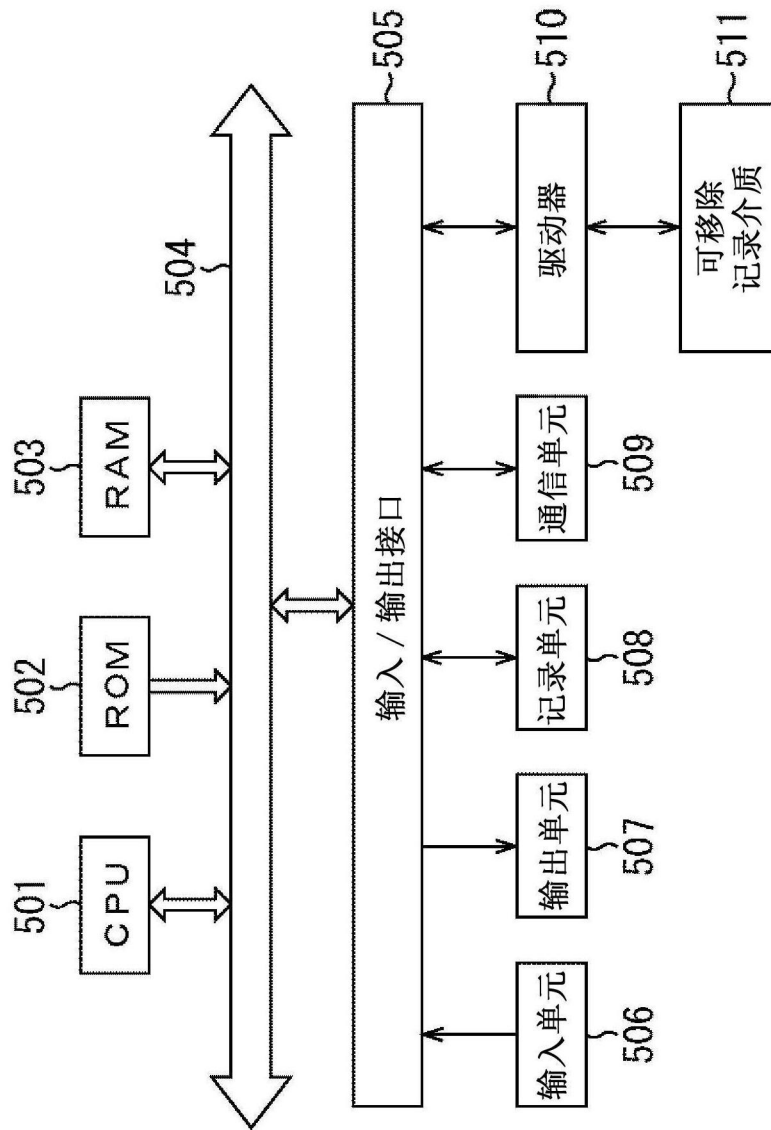


图15