



(12) 发明专利

(10) 授权公告号 CN 113039816 B

(45) 授权公告日 2023.06.06

(21) 申请号 201980065687.8

(22) 申请日 2019.10.03

(65) 同一申请的已公布的文献号
申请公布号 CN 113039816 A

(43) 申请公布日 2021.06.25

(30) 优先权数据
2018-191513 2018.10.10 JP

(85) PCT国际申请进入国家阶段日
2021.04.02

(86) PCT国际申请的申请数据
PCT/JP2019/039103 2019.10.03

(87) PCT国际申请的公布数据
W02020/075622 JA 2020.04.16

(73) 专利权人 索尼集团公司
地址 日本东京

(72) 发明人 福田和巳 曲谷地哲

(74) 专利代理机构 北京康信知识产权代理有限
责任公司 11240

专利代理师 吴孟秋

(51) Int.Cl.
H04S 7/00 (2006.01)
H04R 5/027 (2006.01)
H04S 1/00 (2006.01)

(56) 对比文件
EP 3351172 A1, 2018.07.25
US 2018132764 A1, 2018.05.17
US 10038966 B1, 2018.07.31
EP 3351172 A1, 2018.07.25
US 6996244 B1, 2006.02.07
JP 2004314915 A, 2004.11.11
JP H1083190 A, 1998.03.31
US 2007019812 A1, 2007.01.25
CN 1901761 A, 2007.01.24
CN 103139677 A, 2013.06.05

审查员 陈志清

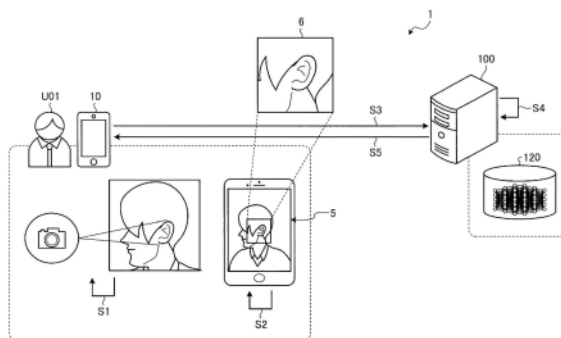
权利要求书2页 说明书22页 附图17页

(54) 发明名称

信息处理装置、信息处理方法和信息处理程序

(57) 摘要

根据本公开的信息处理装置(100)包括:获取单元(141),其获取包括用户的耳朵的图像的第一图像;以及计算单元(142),其基于由获取单元(141)获取的第一图像,使用学习模型来计算与用户相对应的头部相关传递函数,该学习模型已经学习以在输入包括耳朵的图像的图像时输出与耳朵相对应的头部相关传递函数。



1. 一种信息处理装置,包括:

获取单元,被配置为:

获取包括用户的耳朵的第一图像的图像;以及

通过使用耳朵参数估计模型的输出来获取包括在所述第一图像中的所述耳朵的第一耳朵参数,所述耳朵参数估计模型已经学习以基于包括耳朵的第一图像的图像来输出与所述耳朵相对应的第一耳朵参数;

计算单元,被配置为使用学习模型的输出,基于所述第一耳朵参数来计算与所述用户相对应的头部相关传递函数,所述学习模型已经学习以基于所述第一耳朵参数输出与所述第一耳朵参数相对应的头部相关传递函数;

其中,所述学习模型通过以下方式进行学习:

对通过合成耳朵的三维数据和头部的三维数据而获得的三维数据执行声学模拟,所述耳朵的所述三维数据基于第二耳朵参数而获取;并且

学习通过所述声学模拟获得的头部相关传递函数与所述第二耳朵参数之间的关系;以及

所述耳朵参数估计模型通过以下方式进行学习:

学习包括耳朵的第二图像的图像与所述耳朵的第二耳朵参数之间的关系,

其中,包括所述耳朵的所述第二图像的所述图像是基于所述第二耳朵参数而获取的计算机图形CG图像。

2. 根据权利要求1所述的信息处理装置,其中,所述耳朵参数估计模型进一步通过以下方式进行学习:

学习所述第二耳朵参数与通过渲染基于所述第二耳朵参数生成的所述耳朵的三维数据而获得的包括所述耳朵的所述第二图像的图像之间的关系。

3. 根据权利要求2所述的信息处理装置,其中,所述耳朵参数估计模型进一步通过以下方式进行学习:

学习通过改变所述耳朵或头部的三维数据的纹理、渲染中的相机角度或渲染中的亮度而获得的多个包括所述耳朵的所述第二图像的图像与多个包括所述耳朵的所述第二图像的所述图像所共有的第二耳朵参数之间的关系。

4. 根据权利要求1所述的信息处理装置,其中,所述学习模型进一步通过以下方式进行学习:

压缩通过所述声学模拟获得的所述头部相关传递函数的信息量,并且学习所压缩的头部相关传递函数与所述第二耳朵参数之间的关系。

5. 根据权利要求1所述的信息处理装置,其中,设置了基于所述第二耳朵参数生成的所述耳朵的三维数据的听力点,并且通过使用所设置的听力点来执行所述声学模拟。

6. 根据权利要求1所述的信息处理装置,进一步包括:预处理单元,被配置为在包括所述用户的整个头部的内容图像的头部图像中指定包括所述用户的耳朵的第一图像的范围,并且将指定范围检测为包括所述耳朵的所述第一图像的图像,其中,所述获取单元获取由所述预处理单元检测到的包括所述耳朵的所述第一图像的图像。

7. 根据权利要求6所述的信息处理装置,其中,所述预处理单元基于包括在所述头部图像中的所述用户的所述头部的特征点与所述用户的姿势之间的关系来指定所述范围。

8. 根据权利要求7所述的信息处理装置,其中,当基于包括在所述头部图像中的所述用户的所述头部的特征点与所述用户的所述姿势之间的关系不能指定所述范围时,所述预处理单元新请求获取与所述头部图像不同并且包括所述用户的所述整个头部的内容图像的图像。

9. 根据权利要求6所述的信息处理装置,其中,所述预处理单元通过基于包括在所述头部图像中的校正信息校正所述头部图像的旋转来指定包括所述用户的耳朵的第一图像的范围,并且将指定范围检测为包括所述耳朵的所述第一图像的所述图像。

10. 一种信息处理方法,计算机通过所述信息处理方法执行以下操作:

获取包括用户的耳朵的第一图像的图像;

通过使用耳朵参数估计模型的输出来获取包括在所述第一图像中的所述耳朵的第一耳朵参数,所述耳朵参数估计模型已经学习以基于包括所述耳朵的所述第一图像的图像来输出与耳朵相对应的所述第一耳朵参数;

使用学习模型的输出,基于所述第一耳朵参数来计算与所述用户相对应的头部相关传递函数,所述学习模型已经学习以基于所述第一耳朵参数输出与所述第一耳朵参数相对应的头部相关传递函数;

其中,所述学习模型通过以下方式进行学习:

对通过合成耳朵的三维数据和头部的三维数据而获得的三维数据执行声学模拟,所述耳朵的所述三维数据基于第二耳朵参数而获取;并且

学习通过所述声学模拟获得的头部相关传递函数与所述第二耳朵参数之间的关系;并且

所述耳朵参数估计模型通过以下方式进行学习:

学习包括耳朵的第二图像的图像与所述耳朵的第二耳朵参数之间的关系,

其中,包括所述耳朵的所述第二图像的所述图像是基于所述第二耳朵参数而获取的计算机图形CG图像。

11. 一种存储有信息处理程序的计算机存储介质,当被执行时,所述信息处理程序被配置为使计算机执行根据权利要求10所述的信息处理方法。

信息处理装置、信息处理方法和信息处理程序

技术领域

[0001] 本公开涉及信息处理装置、信息处理方法和信息处理程序。具体地，本公开涉及头部相关传递函数计算处理。

背景技术

[0002] 已经使用通过使用数学上表达声音如何从声源传播到耳朵的头部相关传递函数(在下文中也称为HRTF)在耳机等处立体地再现音频图像的技术。

[0003] 头部相关传递函数在个体之间有很大不同，并且因此期望使用当使用头部相关传递函数时为每个个体生成的头部相关传递函数。例如，已知基于通过捕获用户的耳廓而获得的图像来生成头部的三维数字模型(在下文中称为“3D模型”)，并且基于所生成的3D模型计算用户的头部相关传递函数的技术。

[0004] 引用列表

[0005] 专利文献

[0006] 专利文献1:美国专利第9544706号。

发明内容

[0007] 技术问题

[0008] 根据常规技术，可以在信息处理中使用根据每个用户单独计算的头部相关传递函数，从而增加音频图像的平稳性。

[0009] 然而，上述常规技术基于由用户捕获的图像来生成3D数字模型，并且基于所生成的模型来计算头部相关传递函数，并且因此具有相对较大的计算处理负荷。因此，假设利用上述常规技术，需要很长时间直到头部相关传递函数被提供给已经发送图像的用户，并且因此便利性不高。

[0010] 因此，本公开提供了信息处理装置、信息处理方法和信息处理程序，其可以改善用户在与头部相关传递函数相关的处理中的便利性。

[0011] 问题的解决方案

[0012] 为了解决以上问题，根据本公开的实施例的信息处理装置包括：获取单元，其被配置为获取包括用户的耳朵的内容图像的第一图像；以及计算单元，其被配置为基于由获取单元获取的第一图像，通过使用学习模型来计算与用户相对应的头部相关传递函数，该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的头部相关传递函数。

附图说明

[0013] [图1]是示出根据本公开的第一实施例的信息处理的概况的示图。

[0014] [图2]是示出根据本公开的信息处理的整个过程的概念图。

[0015] [图3]是示出根据本公开的第一实施例的信息处理装置的示例性配置的示图。

- [0016] [图4]是示出根据本公开的示例性耳朵图像存储单元的示图。
- [0017] [图5]是示出与根据本公开的耳朵模型相关的示例性学习处理的示图。
- [0018] [图6]是示出与根据本公开的耳朵参数估计模型相关的示例性学习处理的示图。
- [0019] [图7]是示出根据本公开的示例性耳朵图像生成处理的示图。
- [0020] [图8]是用于描述根据本公开的耳朵参数估计模型的示图。
- [0021] [图9]是示出与HRTF相关的模型的生成处理的过程的概况的示图。
- [0022] [图10]是用于描述根据本公开的3D模型重构的示图。
- [0023] [图11]是用于描述与HRTF相关的模型的生成处理的细节的示图。
- [0024] [图12]是用于描述根据本公开的HRTF的压缩和解压缩的示图。
- [0025] [图13]是示出根据本公开的HRTF估计处理的过程的示图。
- [0026] [图14]是示出根据本公开的第一实施例的用户终端的示例性配置的示图。
- [0027] [图15]是示出根据本公开的检测处理的过程的示图。
- [0028] [图16]是示出根据本公开的第二实施例的信息处理装置的示例性配置的示图。
- [0029] [图17]是示出被配置为实现每个信息处理装置的功能的示例性计算机的硬件配置图。

具体实施方式

[0030] 下面将参考附图详细描述本公开的实施例。注意,实施例中相同的部位由相同的参考标记表示,并且省略其重复描述。

[0031] (1. 第一实施例)

[0032] [1-1. 根据第一实施例的信息处理的概况]

[0033] 首先,将参考图1描述根据本公开的信息处理系统1的配置和由信息处理系统1执行的信息处理的概况。图1是示出根据本公开的第一实施例的信息处理的概况的示图。通过图1所示的信息处理系统1来实现根据本公开的第一实施例的信息处理。信息处理系统1包括信息处理装置100和用户终端10。信息处理装置100和用户终端10通过有线或无线网络(未示出)执行相互通信。注意,包括在信息处理系统1中的每种装置的数量不限于图示的数量。

[0034] 信息处理装置100是根据本公开的示例性信息处理装置,计算与每个用户相对应的头部相关传递函数(HRTF),并且提供所计算的HRTF。信息处理装置100通过例如服务器装置来实现。

[0035] 用户终端10是由期望接收HRTF提供的用户使用的信息处理终端。用户终端10通过例如具有图像捕获功能的智能电话来实现。在图1所示的示例中,用户终端10由用户U01用作示例性用户。

[0036] HRTF将由包括人的耳廓(auricle)(耳廓(ear conch))或头部的形状的附近的对象引起的声音变化表示为传递函数。通常,通过使用安装在人的耳廓中的麦克风、虚拟头部麦克风等测量用于测量的声学信号来获取用于计算HRTF的测量数据。

[0037] 例如,经常通过使用由虚拟头部麦克风等获取的测量数据、从大量人获取的测量数据的平均值等来计算在诸如3D声学技术的技术中使用的HRTF。然而,个体的HRTF彼此大不相同,并且因此期望使用用户自己的HRTF来实现更有效的声学渲染效果。具体地,可以通

过利用用户的HRTF代替典型的HRTF来向用户提供具有更真实感觉的声学体验。

[0038] 然而,个体用户的HRTF的测量存在各种问题。例如,需要相对高密度的测量数据来获得提供优异声学效果的HRTF。需要以用户周围的各种角度输出到用户的声学信号的测量数据以获取高密度的测量数据。这样的测量需要很长时间,并且因此用户的身体负担很大。另外,需要在消声室等中执行精确的测量,并且因此经济负担也很大。因此,例如,减轻用户负担和降低测量成本是HRTF计算中的问题。

[0039] 对于上述问题,存在以3D模型表达用户的耳朵或头部并且通过利用3D模型进行声学模拟来执行伪测量的技术。利用该技术,用户可以通过提供头部的扫描数据或头部的捕获图像来执行HRTF的计算,而无需在测量室中执行实际测量。

[0040] 然而,对于生成3D模型和利用3D模型进行声学模拟的处理,计算处理负荷非常大。因此,当上述技术用于在使用例如3D声学的软件等上结合用户唯一的HRTF时,潜在地发生几十分钟或几小时的时滞。这对于用户来说不是很方便。因此,需要提高HRTF计算中的处理速度以实现用户对HRTF的有效使用,这是另一个问题。

[0041] 如上所述,获得与个体用户相对应的HRTF存在各种问题。根据本公开的信息处理装置100通过本公开的信息处理解决上述问题。

[0042] 具体地,信息处理装置100通过使用学习模型(在下文中简称为“模型”)来计算与用户相对应的HRTF,该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的HRTF。例如,当从用户终端10获取包括用户U01的耳朵的内容图像的图像时,信息处理装置100通过将图像输入到模型中来计算用户U01唯一的HRTF。具体地,信息处理装置100不通过基于用户U01的图像生成3D模型并且执行声学模拟的处理来计算HRTF。

[0043] 因此,与执行声学模拟的情况相比,信息处理装置100可以在极短的时间内执行HRTF计算。以下将参考图1及其过程描述根据本公开的信息处理装置100执行的信息处理的概况。

[0044] 如图1所示,用户U01从头部的侧面执行用户的图像捕获以获取包括用户的耳朵的内容图像的图像(步骤S1)。例如,用户U01通过使用用户终端10拥有的相机来执行用户的头部的图像捕获。注意,在本公开中,耳朵的图像不限于可以由用户终端10等捕获的典型的二维彩色图像,而是可以是单色图像、包括深度信息的深度图像或其可选组合。在本公开的处理中使用的图像的数量不限于一个,而是可以是两个或多个。

[0045] 用户终端10在步骤S1获得的图像5上执行预处理以发送到信息处理装置100(步骤S2)。具体地,用户终端10执行检测包括在图像5中的用户U01的耳朵的内容图像并且从图像5中切出所检测的耳朵内容图像的范围的处理作为预处理。注意,稍后将描述诸如耳朵检测的预处理的细节。

[0046] 通过预处理,用户终端10生成包括用户U01的耳朵的内容图像的图像6。然后,用户终端10将图像6发送到信息处理装置100(步骤S3)。注意,通过例如由信息处理装置100提供的计算机程序(例如,智能电话应用程序)来执行一系列处理,诸如从通过图像捕获获得的图像5生成图像6的处理以及发送图像6的处理。以这种方式,在根据本公开的信息处理中,仅将通过从图像5中仅切出耳朵内容图像而获得的图像6发送到网络,但是不发送潜在地指定用户的图像5,并且因此可以实现安全性优异的处理。另外,信息处理装置100不获取图像5,而是仅获取图像6,从而避免了处理私人信息的风险。注意,信息处理装置100可以执行从

用户终端10获取图像5并且在信息处理装置100侧生成图像6的处理。稍后将描述该配置作为第二实施例。

[0047] 信息处理装置100获取从用户终端10发送的图像6。然后,信息处理装置100将图像6输入到存储在存储单元120中的模型(步骤S4)。该模型是已经学习以在输入包括耳朵的内容图像的二维图像时输出与耳朵相对应的HRTF的模型。具体地,信息处理装置100通过将图像6输入到模型来计算与包括在图像6中的耳朵(换句话说,用户U01)相对应的HRTF。

[0048] 然后,信息处理装置100将所计算的HRTF提供给用户终端10(步骤S5)。以这种方式,用户U01可以通过仅通过用户的侧脸的图像捕获来准备图像5,而无需实际测量处理、3D模型声学模拟等来获得用户唯一的HRTF。因此,信息处理装置100可以在极短的时间内将HRTF提供给用户U01,而不会对用户U01造成测量负担。因此,信息处理装置100可以在与HRTF相关的处理中为用户实现改进的便利性。

[0049] 如上所述,在根据本公开的信息处理中,通过使用通过学习处理生成的模型来计算HRTF,从而实现处理的加速。尽管参考图1描述了在根据本公开的信息处理中将HRTF提供给用户U01的处理的概况,但是以下将参考图2和以下附图详细描述由信息处理装置100执行的包括模型学习处理的一系列信息处理。注意,参考图1描述的“模型”不必是一个模型,而是可以是输出各种值的多个模型的组合,将参考图2和以下附图描述其细节。

[0050] [1-2.根据本公开的信息处理的整个过程]

[0051] 在描述诸如信息处理装置100的配置的细节之前,在图2中示出了由根据本公开的信息处理装置100执行的信息处理的整个过程。图2是示出根据本公开的信息处理的整个过程的概念图。

[0052] 首先,信息处理装置100收集与多个人的耳形相关的数据,并且基于所收集的耳形生成耳朵模型(步骤S11)。注意,耳形不必限于石膏等成形的人的耳朵,而是可以是指示人的耳朵的形狀的任何信息。在本公开中,耳朵模型是当输入指示耳朵的特性的参数(在下文中称为“耳朵参数”)时输出对应的耳朵形状的模型。可以例如通过基于与耳朵的形状相关的数据(例如,通过计算机断层摄影(CT)扫描从所收集的耳形中获得的数据)对耳朵的形状执行主成分分析来获得耳朵参数。因此,一旦获得了耳朵参数,信息处理装置100就可以获得与耳朵参数相对应的耳朵的形狀的数据(换句话说,表示耳朵的3D模型)。

[0053] 此后,信息处理装置100基于耳朵模型生成耳朵参数估计模型(步骤S12)。信息处理装置100可以通过将耳朵参数输入到在步骤S11生成的耳朵模型来生成大量的耳朵图像。可以随机输入耳朵参数,或者可以基于可选规则自动生成耳朵参数(例如,当发现对于每个特定种族存在耳朵形状的特定趋势的事实时,可以基于该事实得出规则),并且可以输入所生成的值。因此,信息处理装置100可以通过学习每个所生成的耳朵图像与基于执行哪个生成的耳朵参数之间的关系来生成当输入包括耳朵的图像时输出与耳朵相对应的耳朵参数的模型。该模型是耳朵参数估计模型。因此,一旦获得了包括耳朵的内容图像的二维图像,信息处理装置100就可以获得与耳朵相对应的耳朵参数。然后,一旦获得了耳朵参数,信息处理装置100就可以通过使用在步骤S11生成的耳朵模型来获得包括在图像中的耳朵的3D模型。注意,在上述学习中,可以学习通过捕获人的耳朵获得的图像、由数据表示的耳朵的形狀以及从耳朵的数据转换的耳朵参数之间的关系。在这种情况下,执行使用实际捕获的图像而不是计算机图形(CG)图像的学习,并且因此假设可以提高所生成的耳朵参数估计模

型的精度。

[0054] 信息处理装置100对通过使用耳朵参数估计模型生成的3D模型执行声学模拟,并且计算与3D模型相对应的唯一HRTF(在下文中,针对个体耳形生成的这种HRTF被称为“个性化HRTF”)(步骤S13)。因此,通过步骤S11至S13的处理,信息处理装置100可以通过执行声学模拟来实现用于从包括耳朵的图像中计算个性化HRTF的一系列处理。

[0055] 另外,信息处理装置100基于随机或定期生成的耳朵参数来生成大量的3D模型,并且针对所生成的3D模型重复执行声学模拟的处理,从而学习耳朵参数与个性化HRTF之间的关系。因此,信息处理装置100基于所计算的个性化HRTF来生成HRTF学习模型(步骤S14)。

[0056] 在本公开中,HRTF学习模型是当输入耳朵参数时输出与该耳朵参数相对应的个性化HRTF的模型。因此,一旦获得了耳朵参数,信息处理装置100就可以获得与该耳朵参数相对应的个性化HRTF。

[0057] 此后,当从用户获取图像时,信息处理装置100通过将图像(更准确地,包括在图像中的耳朵的耳朵参数)输入到HRTF学习模型来计算用户的个性化HRTF(步骤S15)。在步骤S15指示的处理与图1所示的一系列处理相对应。

[0058] 如上所述,信息处理装置100生成多个模型并且使用所生成的模型执行信息处理,从而基于从用户获取的图像来计算个性化HRTF。注意,不必按照步骤S11至S15的顺序执行图2所示的处理。例如,信息处理装置100仅在生成处理所需的每个模型之后才需要在步骤S15执行处理。

[0059] 这结束了根据本公开的信息处理的整个过程的概述的描述。将参考图3和以下附图顺序地描述信息处理装置100和用户终端10的配置以及各种学习处理的细节。

[0060] [1-3.根据第一实施例的信息处理装置的配置]

[0061] 以下将参考图3描述根据第一实施例的信息处理装置100的配置。图3是示出根据本公开第一实施例的信息处理装置100的示例性配置的示图。

[0062] 如图3所示,信息处理装置100包括通信单元110、存储单元120和控制单元130。注意,信息处理装置100可以包括被配置为从管理信息处理装置100的管理员等接收各种操作的输入单元(例如,键盘或鼠标),以及用于显示各种信息的显示单元(例如,液晶显示器)。

[0063] 通信单元110例如通过网络接口卡(NIC)来实现。通信单元110以有线或无线方式与网络N(诸如因特网)连接,并且通过网络N将信息发送到用户终端10等和从用户终端10等接收信息。

[0064] 存储单元120例如由存储装置(诸如随机存取存储器(RAM)或闪存的半导体存储元件、硬盘或光盘)来实现。存储单元120存储用于学习处理的各种数据以及通过学习处理生成的模型等。

[0065] 如图3所示,存储单元120包括耳形信息存储单元121、耳朵模型存储单元122、耳朵图像存储单元123、耳朵参数估计模型存储单元124、HRTF处理模型存储单元125、HRTF学习数据存储单元126和HRTF学习模型存储单元127。

[0066] 耳形信息存储单元121存储从人体实际收集的耳形生成的3D模型数据的信息(换句话说,与耳朵的形状相关的信息)。具体地,耳形信息存储单元121存储表示三维形状并且通过对所收集的耳形执行CT扫描而获得的数据(诸如3D多边形)。

[0067] 耳朵模型存储单元122存储根据本公开的耳朵模型。耳朵模型是在输入指示耳朵

的特性的耳朵参数时输出对应的耳朵形状模型。

[0068] 可以通过对存储在耳形信息存储单元121中的指示耳形的形状的数据执行主成分分析来获得耳朵参数。具体地,通过对耳朵的3D多边形执行统计分析(主成分分析)来量化耳朵中变化很大的部位(表征耳朵的形状)来获得耳朵参数。根据本公开的耳朵参数例如被指示为10个数字的组合,并且每个数字例如被指示为-10到+10的范围内的数字。例如,所有数字均为“0”的耳朵参数与具有学习数据的平均形状的耳朵(所收集的耳形)相对应。注意,信息处理装置100可以适当地将例如在生成人的脸的处理中使用的已知技术应用于通过主成分分析生成指示耳朵的形状的模型的处理。信息处理装置100不限于主成分分析,而是可以通过使用诸如独立成分分析或适当的任何其他非线性模型的已知分析方法来生成耳朵参数。耳朵参数不限于量化耳朵中变化很大的部位而获得的耳朵参数,而是例如可以通过对与耳朵的形状相关的特性进行参数化而获得的耳朵参数,从而对HRTF的影响较大。

[0069] 耳朵图像存储单元123存储包括耳朵的内容图像的图像。例如,耳朵图像存储单元123将通过渲染由耳朵模型生成的耳朵的形状(耳朵的3D模型)而获得的CG图像存储为耳朵图像。耳朵图像存储单元123可以将从用户发送的包括耳朵的内容图像的图像存储为耳朵图像。

[0070] 图4示出了根据本公开的示例性耳朵图像存储单元123。图4是示出本公开的示例性耳朵图像存储单元123的示图。在图4所示的示例中,耳朵图像存储单元123具有“耳朵参数”、“耳朵3D模型数据”、“头部3D模型数据”、“耳朵图像ID”和“图像生成参数”的项目。“图像生成参数”具有“纹理”、“相机角度”、“分辨率”和“亮度”的子项目。

[0071] “耳朵参数”是指示耳朵的形状的特性的参数。例如,耳朵参数由十维数表示。“耳朵3D模型数据”是指示基于耳朵参数重构的耳朵的三维形状的数据。“头部3D模型数据”是指示头部的三维形状并且在重构人的3D模型时与耳朵3D模型数据合成的数据。

[0072] “耳朵图像ID”指示识别通过渲染3D模型而获得的耳朵图像的识别信息。如图4所示,通过以各种方式改变在渲染时设置的参数(图像生成参数),从一个3D模型生成多个耳朵图像。

[0073] “图像生成参数”指示用于生成图像的渲染中的设置参数。“纹理”指示CG纹理的设置。“相机角度”指示当通过渲染3D模型获得二维图像时伪相机的图像捕获角度。“分辨率”指示渲染的分辨率。“亮度”指示渲染的亮度。亮度项可以包括设置数据,诸如渲染时光(入射光)的角度。

[0074] 注意,尽管每个项目的数据在概念上被写为例如图4中的“A01”或“B01”,但是与该项目相对应的特定数据实际上被存储为该项目的数据。例如,在“耳朵参数”项中存储了10个数字的特定字符串。其他项目也一样,并且与每个项目相对应的各种数值和信息被存储在每个项目中。

[0075] 具体地,在图4所示的示例中,指示由耳朵参数“A01”生成的耳朵3D模型数据为“B01”,并且与该耳朵3D模型数据组合且包括在人的3D模型中的头部3D模型数据为“C01”。另外,指示从所生成的人的3D模型获得的耳朵图像是由耳朵图像ID“D01”、“D02”、“D03”等识别的多个耳朵图像。另外,指示对于由耳朵图像ID“D01”识别的耳朵图像,纹理为“E01”,相机角度为“F01”,分辨率为“G01”,亮度为“H01”作为渲染时的图像生成参数。

[0076] 再次参考图3进行描述。耳朵参数估计模型存储单元124存储耳朵参数估计模型。

耳朵参数估计模型是在输入包括耳朵的内容图像的二维图像时输出与耳朵相对应的耳朵参数的模型。

[0077] HRTF处理模型存储单元125存储HRTF处理模型。尽管稍后详细描述,但是HRTF处理模型执行压缩通过声学模拟等计算的HRTF的信息量的处理。注意,在以下描述中,由HRTF处理模型压缩的HRTF也被称为HRTF参数。

[0078] HRTF学习数据存储单元126存储用来生成用于从包括耳朵的内容图像的图像中计算HRTF的模型(稍后将描述的HRTF学习模型)的学习数据。具体地,HRTF学习数据存储单元126存储指示耳朵的形状的耳朵参数和与基于耳朵参数指定的耳朵的形状相对应的HRTF组合的数据作为学习数据。

[0079] HRTF学习模型存储单元127存储HRTF学习模型。HRTF学习模型是在输入包括耳朵的内容图像的图像时输出与耳朵相对应的HRTF的模型。例如,当获取包括耳朵的内容图像的图像时,HRTF学习模型使用耳朵参数估计模型来输出与耳朵相对应的耳朵参数,并且另外,输出与该耳朵参数相对应的HRTF。

[0080] 控制单元130例如通过中央处理单元(CPU)或微处理单元(MPU)通过使用随机存取存储器(RAM)等作为工作区来执行存储在信息处理装置100中的计算机程序(例如,根据本公开的信息处理程序)来实现。控制单元130是控制器,并且可以例如通过诸如专用集成电路(ASIC)或现场可编程门阵列(FPGA)的集成电路来实现。

[0081] 如图3所示,控制单元130包括学习单元131和估计单元140。学习单元131包括接收单元132、耳朵模型学习单元133、图像生成单元134、耳朵参数学习单元135和HRTF学习单元136,并且实现或执行下述信息处理的功能和效果。估计单元140包括获取单元141、计算单元142和提供单元143,并且实现或执行下述信息处理的功能和效果。注意,控制单元130的内部配置不限于图3所示的配置,而是可以是执行稍后描述的信息处理的任何配置。

[0082] 以下将首先描述学习单元131。学习单元131执行与各种数据相关的学习处理,并且生成要由估计单元140使用的各种模型。

[0083] 注意,学习单元131基于各种数据执行用于生成模型的学习,但是以下描述的学习处理是示例性的,并且由学习单元131执行的学习处理的类型未指定为任何类型。例如,学习单元131可以通过使用诸如神经网络、支持向量机、聚类和强化学习的各种学习算法来生成模型。

[0084] 接收单元132接收各种信息。例如,接收单元132接收从人体收集的耳形的CT扫描数据。接收单元132将所接收的数据存储在耳形信息存储单元121中。

[0085] 耳朵模型学习单元133通过执行与耳朵模型相关的学习处理来生成耳朵模型。耳朵模型学习单元133将所生成的耳朵模型存储在耳朵模型存储单元122中。

[0086] 以下将参考图5描述由接收单元132和耳朵模型学习单元133执行的示例性学习处理。图5是示出与根据本公开的耳朵模型相关的示例性学习处理的示图。

[0087] 如图5所示,接收单元132接收从耳形收集并扫描的数据,并且将所接收的数据存储在耳形信息存储单元121中。另外,接收单元132将所接收的数据发送到耳朵模型学习单元133(步骤S16)。

[0088] 耳朵模型学习单元133对所获取的耳形数据进行均匀化,并且生成耳形的均匀化数据(步骤S17)。均匀化是将3D模型的顶点数量和多边形配置与参考3D模型的顶点数量和

多边形配置统一。在这种情况下,需要注意不要通过均匀化来改变形状。另外,耳朵模型学习单元133对均匀化的数据执行主成分分析(步骤S18)。因此,耳朵模型学习单元133生成从耳朵的形状计算指示耳朵的耳朵参数的模型(耳朵模型)。耳朵模型学习单元133将所生成的耳朵模型存储在耳朵模型存储单元122中。

[0089] 再次参考图3进行描述。图像生成单元134生成包括耳朵的内容图像的图像。例如,图像生成单元134通过将所生成的耳朵参数输入到耳朵模型来随机生成耳朵参数并且生成耳朵的3D模型。另外,图像生成单元134在CG渲染时随机地生成参数(在下文中称为“图像生成参数”),诸如所生成的3D模型的纹理(例如肤色)、其渲染质量(诸如图像质量)以及其相机角度。然后,图像生成单元134适当地结合所生成的3D模型和多个图像生成参数来执行渲染,从而生成其中的耳朵形状、肤色等以各种方式不同的CG图像。

[0090] 在稍后描述的估计处理中,从用户发送的图像用于该处理,并且假设用户的肤色、图像捕获时的耳朵角度等很可能在从用户发送的图像之间以各种方式不同。因此,在该处理中,需要在从用户发送的各种图像中准确地执行耳朵的内容图像的图像识别,这是一个问题。图像生成单元134如上所述生成与各种情况相对应的大量图像以提高图像识别的精度,从而解决上述问题。

[0091] 耳朵参数学习单元135通过学习包括耳朵的内容图像的图像与耳朵参数之间的关系来生成耳朵参数估计模型。耳朵参数学习单元135与根据本公开的第一学习单元相对应。如稍后所述,包括耳朵的内容图像的图像可以通过实际捕获人的耳朵而获得的图像,或者可以是基于耳朵参数而生成的CG图像。

[0092] 例如,耳朵参数学习单元135通过学习耳朵参数与通过渲染通过将基于耳朵参数生成的耳朵的三维数据与头部的三维数据合成而获得的三维数据而获得的耳朵图像之间的关系来生成耳朵参数估计模型。具体地,耳朵参数学习单元135学习由图像生成单元134生成的CG图像与耳朵参数之间的关系。如上所述,图像生成单元134基于随机或定期设置的耳朵参数来生成CG图像,并且因此针对CG图像唯一地确定耳朵参数。因此,通过学习所输入的CG图像与耳朵参数之间的关系,耳朵参数学习单元135可以在输入图像时生成模型,该模型输出与包括在图像中的耳朵的内容图像相对应的耳朵参数。注意,耳朵参数学习单元135不必将通过渲染与头部合成的三维数据而获得的耳朵图像用于学习。具体地,耳朵参数学习单元135可以通过学习耳朵参数与通过仅渲染基于耳朵参数生成的耳朵的三维数据而获得的耳朵图像之间的关系来生成耳朵参数估计模型。

[0093] 另外,耳朵参数学习单元135通过学习耳朵或头部的三维数据的纹理、渲染中的相机角度、渲染中的亮度等不同的多个耳朵图像与多个耳朵图像共有的耳朵参数之间的关系来生成耳朵参数估计模型。以这种方式,耳朵参数学习单元135通过使用各个方面的耳朵图像来执行学习,并且因此可以生成可靠且稳健的模型,该模型可以在输入任何图像时准确地输出耳朵参数(例如,可以对包括在输入图像中的耳朵参数以外的信息的任何变化执行可靠的估计)。

[0094] 以下将参考图6描述由图像生成单元134和耳朵参数学习单元135执行的示例性学习处理。图6是示出与根据本公开的耳朵参数估计模型相关的示例性学习处理的示图。

[0095] 如图6所示,图像生成单元134参考耳朵模型存储单元122(步骤S21)并且获取耳朵模型。然后,图像生成单元134生成与耳朵参数相对应的随机数和与CG纹理、渲染相机角度

等相对应的随机数(步骤S22)。换句话说,图像生成单元134生成用于生成耳朵图像的各种参数(图像生成参数)。

[0096] 然后,图像生成单元134获取图像生成参数中的耳朵参数(步骤S23),并且通过将所获取的耳朵参数输入到耳朵模型来重构耳朵的3D模型(步骤S24)。

[0097] 随后,图像生成单元134获取图像生成参数中的诸如CG纹理的参数(步骤S25),并且通过输入所获取的参数来向3D模型提供3DCG渲染(步骤S26)。注意,在渲染中使用的头部例如是多个人(例如,收集了耳形的多个人)的平均头部、用作样本的头部的3D模型等。类似于耳朵模型,可以通过使通过对多个人的头部执行3D扫描而获得的3D数据均匀化来生成头部的3D模型。在这种情况下,图像生成单元134可以通过以随机数执行参数生成来随机地生成头部3D模型。注意,图像生成单元134可以针对纹理生成通过相同方法生成的纹理模型,从而以随机数生成各种纹理。

[0098] 因此,图像生成单元134生成包括耳朵的内容图像的图像(耳朵图像)。注意,图像生成单元134可以通过以各种方式改变诸如纹理的参数来从一个3D模型生成多个耳朵图像。

[0099] 以下将参考图7描述由图像生成单元134生成的示例性耳朵图像。图7是示出根据本公开的示例性耳朵图像生成处理的示意图。

[0100] 图像生成单元134通过使用随机生成的耳朵参数来生成指示耳朵的三维形状的3D模型(步骤S41)。然后,图像生成单元134获取基于指示多个人的平均头部的三维形状的数据而生成的头部3D模型(步骤S42)。然后,图像生成单元134通过合成在步骤S41生成的耳朵3D模型和在步骤S42获取的头部3D模型来生成伪人的3D模型。

[0101] 随后,图像生成单元134执行从通过执行3D模型的伪图像捕获来生成的3D模型生成二维图像的处理(渲染)。例如,图像生成单元134将3D模型中的耳朵的前表面的角度设置为伪图像捕获角度,并且生成其中耳朵的内容图像基本上在中心的图像。

[0102] 图像生成单元134通过将图像生成参数随机输入3D模型中,以各种方式改变CG纹理(具体地,肤色等)、渲染质量(分辨率等)、耳朵相对于头部的位置等。因此,图像生成单元134可以生成肤色等不同的大量图像(步骤S43)。

[0103] 图像组20表示由图像生成单元134生成的多个耳朵图像。以这种方式,图像生成单元134生成大量的各种耳朵图像,从而提高了稍后将描述的耳朵图像识别精度。

[0104] 再次参考图6进行描述。图像生成单元134将所生成的耳朵图像存储在耳朵图像存储单元123中(步骤S27)。注意,图像生成单元134将生成图像时的图像生成参数与每个所生成的耳朵图像相关联地存储在耳朵图像存储单元123中(步骤S28)。因此,图像生成单元134可以保持每个与耳朵参数相关联的大量耳朵图像作为学习数据。例如,图像生成单元134可以保持每个与耳朵参数相关联的大量耳朵图像作为学习数据。

[0105] 随后,耳朵参数学习单元135参考耳朵图像存储单元123(步骤S29),并且获取每个耳朵图像和对应的耳朵参数。然后,耳朵参数学习单元135通过学习耳朵图像与耳朵参数之间的关系来生成耳朵参数估计模型。耳朵参数学习单元135将所生成的耳朵参数估计模型存储在耳朵参数估计模型存储单元124中(步骤S30)。

[0106] 通过使用例如对提取图像的特征量有用的卷积神经网络来生成耳朵参数估计模型。注意,学习中的成本表达式(成本函数)例如由以下表达式(1)给出。

$$[0107] \quad L(\alpha_{\text{true}}, \alpha_{\text{est}}) = \|\mathbf{A}_{\text{ear}} \alpha_{\text{true}} - \mathbf{A}_{\text{ear}} \alpha_{\text{est}}\|_2^2 \quad (1)$$

[0108] 在表达式(1)中，“ α_{true} ”表示耳朵参数的真实值，“ α_{est} ”表示耳朵参数的估计值，并且“ \mathbf{A}_{ear} ”表示通过主成分分析获得的耳朵模型。右侧的距离函数表示L2范数(欧几里得距离)。注意，耳朵模型参数的真实值例如可以是表示在耳形收集时测量的人的耳朵的参数。具体地，真实值是用于生成耳朵图像的耳朵参数，并且估计值是当在学习中将耳朵图像输入到耳朵参数估计模型时输出的值。作为学习处理，信息处理装置100针对当前估计值更新使成本表达式的值最小化的系数。

[0109] 以下将参考图8描述由耳朵参数学习单元135生成的耳朵参数估计模型。图8是用于描述根据本公开的耳朵参数估计模型的示意图。

[0110] 当已经获取耳朵图像30时，信息处理装置100将所获取的耳朵图像30输入到耳朵参数估计模型。耳朵参数估计模型例如具有卷积神经网络的结构，将所输入的耳朵图像30划分为各自具有几个像素的矩形部分，并且计算表示耳朵图像30的特征量。最后，耳朵参数估计模型输出与包括在耳朵图像30中的耳朵的内容图像相对应的耳朵参数作为表示耳朵图像30的特征量(步骤S45)。

[0111] 注意，信息处理装置100可以通过将所输出的耳朵参数输入到耳朵模型来重构与包括在耳朵图像30中的耳朵相对应的耳形(3D模型)。图8所示的3D模型40是基于耳朵参数通过CG重构包括在耳朵图像30中的耳朵的3D模型。

[0112] 再次参考图3进行描述。HRTF学习单元136学习与耳朵的形状相关的信息与HRTF之间的关系，从而生成与HRTF相关的各种模型。例如，HRTF学习单元136学习包括耳朵的内容图像的图像与对应于耳朵的HRTF之间的关系，从而生成计算HRTF的学习模型。HRTF学习单元136与根据本公开的第二学习单元相对应。

[0113] 例如，HRTF学习单元136对通过合成基于耳朵参数生成的耳朵的三维数据和头部的三维数据而获得的三维数据执行声学模拟，并且学习通过声学模拟获得的HRTF与耳朵参数之间的关系，从而生成学习模型。

[0114] 可选地，HRTF学习单元136可以压缩通过声学模拟获得的HRTF的信息量，并且学习所压缩的HRTF与耳朵参数之间的关系，从而生成学习模型。

[0115] 可选地，HRTF学习单元136可以设置基于耳朵参数生成的耳朵的三维数据的听力点，并且通过使用所设置的听力点来执行声学模拟。听力点是假设有人听到声音的虚拟设置位置。例如，听力点的位置与安装在虚拟头部麦克风中的麦克风的位置(诸如虚拟头部的耳道的入口)相对应。

[0116] 对于HRTF学习单元136的每个上述处理，图9示出了与HRTF相关的各种模型的生成处理的过程。图9是示出与HRTF相关的各种模型的生成处理的过程的概况的示意图。

[0117] 图9示出了HRTF学习单元136基于从用户发送的图像执行预定学习处理的示例。在这种情况下，用户通过使用用户终端10来执行用户的耳朵(准确地，包括耳朵的头部)的图像捕获(步骤S51)。此后，用户终端10执行在所捕获的图片中指定包括耳朵的内容图像的范围并且切出该指定范围以获取耳朵图像的预处理(步骤S52)。

[0118] 此后，HRTF学习单元136通过使用耳朵参数估计模型来计算包括在从用户发送的耳朵图像中的耳朵的耳朵参数(步骤S53)。另外，HRTF学习单元136基于耳朵参数重构耳朵

的3D模型,并且将头部3D模型组合到重构的耳朵,从而生成人的3D模型(步骤S54)。

[0119] 随后,HRTF学习单元136对所生成的3D模型执行声学模拟以计算3D模型的个性化HRTF(步骤S55)。因此,HRTF学习单元136可以获得学习数据,其中,包括在从用户发送的耳朵图像中的耳朵与个性化HRTF相关联。

[0120] 注意,在图9所示的示例中,生成通过声学模拟获得的个性化HRTF与耳朵的数据相关联的学习数据,但是在某些情况下,HRTF学习单元136不必通过声学模拟获得个性化HRTF。例如,当获得收集了耳形的人的个性化HRTF(通过在消声室等中使用测量装置获得的HRTF)时,HRTF学习单元136可以获取学习数据,其中,所测量的个性化HRTF与人的耳形(耳朵参数)相关联。

[0121] 在声学模拟中,HRTF学习单元136自动设置人的3D模型中的听力点。以下将参考图10对此进行描述。图10是用于描述根据本公开的3D模型重构的示图。

[0122] HRTF学习单元136基于随机生成的耳朵参数来重构耳朵3D模型(步骤S71)。随后,HRTF学习单元136通过组合耳朵3D模型和头部3D模型来生成人的3D模型(步骤S72)。另外,HRTF学习单元136基于3D模型中的耳朵的形状来设置声源的听力点60(步骤S73)。例如,HRTF学习单元136可以通过预先学习耳朵的形状和声源的听力点的位置彼此相关联的信息来自动设置听力点。例如,在生成3D模型的定时,HRTF学习单元136估计3D模型中的听力点的位置,从而自动设置听力点。听力点例如与耳朵的外耳道相对应,并且典型地,其位置可以基于耳朵的形状来估计。

[0123] 此后,HRTF学习单元136重新网格化所生成的3D模型以满足声学模拟中对3D模型计算的约束(步骤S74)。这是因为在3D模型模拟中,由于模拟条件而确定了多边形总数的上限、边缘连接顶点的长度等。换句话说,在利用所生成的3D模型进行模拟之前,HRTF学习单元136通过适当地重新网格化来处理3D模型以满足计算限制,从而可以适当地执行模拟。然后,HRTF学习单元136对所生成的3D模型和所设置的听力点60执行声学模拟,从而计算个性化HRTF(步骤S75)。

[0124] 随后,以下将参考图11描述与HRTF相关的模型的生成处理的详细过程。图11是用于描述与HRTF相关的模型的生成处理的细节的示图。

[0125] 如图11所示,HRTF学习单元136执行头部合成(步骤S81),并且然后执行声学模拟(步骤S82)。HRTF学习单元136分析通过声学模拟获得的测量数据(HRTF后处理),并且计算表示个性化HRTF的数值(步骤S83)。注意,HRTF后处理例如是通过对通过声学模拟获得的头部相关脉冲响应(HRIF)的傅立叶变换来计算HRTF。

[0126] 然后,HRTF学习单元136参考HRTF处理模型存储单元125(步骤S84),并且将所计算的HRTF输入到用于处理HRTF的模型(HRTF处理模型)。因此,HRTF学习单元136获得维度降低的HRTF(步骤S85)。具体地,HRTF学习单元136从HRTF处理模型输出HRTF参数作为维度降低的HRTF(步骤S86)。

[0127] 以这种方式,HRTF学习单元136通过使用维度降低的HRTF参数来执行处理,而不是在处理中直接使用通过声学模拟获得的HRTF。这是因为HRTF是具有极大维数的函数,并且因此当直接用于模型生成处理和计算处理中时导致较大的计算处理负荷。

[0128] HRTF学习单元136将与执行声学模拟的头部相关的数据(头部合成所基于的数据,诸如耳朵参数)和所计算的HRTF参数彼此相关联地存储在HRTF学习数据存储单元126中(步

骤S87)。

[0129] 此后,HRTF学习单元136新生成不同的随机耳朵参数(步骤S88),并且通过使用新生成的耳朵参数来执行头部合成(步骤S89)。HRTF学习单元136重复步骤S81至S89以收集学习所需的学习数据。

[0130] 此后,当累积了足够量的学习数据时,HRTF学习单元136参考HRTF学习数据存储单元126(步骤S90),并且学习耳朵参数与HRTF(准确地,HRTF参数)之间的关系(步骤S91)。通过该学习,HRTF学习单元136生成用于直接从耳朵参数获得HRTF的HRTF学习模型,并且将所生成的HRTF学习模型存储在HRTF学习模型存储单元127中。

[0131] 随后,以下将参考图12描述HRTF与HRTF参数之间的关系。图12是用于描述根据本公开的HRTF压缩和解压缩的示意图。

[0132] 如图12所示,HRTF学习单元136对通过声学模拟获得的HRIFHRTF(例如,在图12的示例中为1000方向×500抽头)执行快速傅里叶变换(FFT)(步骤S101)。通过该处理,HRTF学习单元136提取振幅特性(步骤S102),并且例如对听力灵敏度低的频率分量等执行细化处理(步骤S103)。具体地,HRTF可以表示为角度(由 θ 和 ϕ 表示)和频率(由 f 表示)的函数HRTF(θ, ϕ, f)。在这种情况下,当 k 表示频率的槽的数量时,输入到函数的频率 f 表示为 $f_k = f_0, f_1, f_2, \dots, f_{k-1}$ 。换句话说,HRTF在一个方向和一只耳朵上具有复 k 维。在奈奎斯特频率($f_{k/2}$)或更高频率的HRTF是通过将复共轭折叠在频率 $f_{k/2}$ 或更低的频率获得的,并且因此,在信息处理中,仅可以使用从 $f_0 = 0$ 到奈奎斯特频率($f_{k/2}$)的 $(k/2) + 1$ 个频率作为频率槽(frequency bin)。绝对值可以用于至少一个频率槽。例如,当将 f_0 至 $f_{k/2}$ 的所有频率都转换为绝对值时,转换之后的函数 H_2 由以下表达式(2)给出。

$$[0133] \quad H_2(\theta, \phi, k) = |HRTF(\theta, \phi, f_k)| \quad (f_k = f_0, f_1, \dots, f_{k/2}) \quad (2)$$

[0134] 具体地,HRTF学习单元136可以将原始HRTF的维度压缩为实数 $(k/2) + 1$ 的维度。HRTF学习单元136可以通过对上述表达式(2)中的 H_2 执行频率压缩来进一步降低到小于 $(k/2) + 1$ 的维度。存在用于维度压缩的各种已知方法。例如,HRTF学习单元136使用对函数执行倒谱变换的方法以仅获取数量等于或大于1且小于 $(k/2) + 1$ 的频率槽。在示例中,HRTF学习单元136计算多个频率槽的平均值,并且基于该平均值执行维度降低。例如,当以下表达式(3)给出频率槽(a_1, L 和1是等于或大于0的整数)时,通过使用满足 $f_{a_1} \leq f_1 < f_{a_1+1}$ 的1,由以下表达式(4)给出新函数 H_3 。

$$[0135] \quad 0 \leq a_0 < a_1 < a_2 \dots < a_{L-1} \leq K/2, L < K/2 + 1 \quad (3)$$

$$[0136] \quad H_3(\theta, \phi, l) = \frac{1}{a_{l+1} - a_l} \sum_{k=a_l}^{a_{l+1}-1} H_2(k) \quad (4)$$

[0137] 因此,HRTF学习单元136可以将以 $(K/2) + 1$ 维表示的函数 H_2 降低为 L 维。注意,计算平均值的方法不限于上述方法,而是可以通过使用例如均方根或加权平均值来计算平均值。因此,HRTF降低到例如1000方向×50维度。注意,当恢复通过函数 H_3 降低的维度时(例如,稍后将描述的步骤S110),HRTF学习单元136可以通过使用诸如线性插值或样条插值的各种方法来恢复维度。期望具有恢复维度的函数 H'_2 具有比函数 H_2 更平滑的特性,并且通过设计选择 a_1 的方式可以获得提供较小听力影响的 $H'_2(\theta, \phi, k)$ 。例如,可以选择 a_1 ,使得 f_{a_1} 与 f_{a_1+1} 之间的频率间隔在较高频率下更大。

[0138] HRTF学习单元136还对维度降低的HRTF执行球面谐波拟合的处理,从而将信息量压缩为50系数 \times 50维度(步骤S104)。球面谐波拟合是通过使用球面谐波函数对每个压缩频率进行空间方向拟合。HRTF与球面谐波函数之间的关系由以下表达式(5)给出。

$$[0139] \quad H(\theta, \phi, f) = \sum_{n=0}^{\infty} \sum_{m=-n}^n h_{nm}(f) Y_n^m(\theta, \phi) \quad (5)$$

[0140] 如上述表达式(5)中那样,球面谐波函数Y由系数 $h_{nm}(f)$ 表示。在这种情况下,维数n可以在有限值N处被截断,以将系数 $h_{nm}(f)$ 的维度设置为小于原始HRTF的维数(方向数)。这意味着丢弃对于人的感知和仅获取平滑形状不必要的空间上过于精细的振幅。注意,例如,最小二乘法用于计算系数 h_{nm} 的向量 $h = (h_{00}, h_{1-1}, \dots)^T$ 。

$$[0141] \quad E = \|H - Yh\|^2 + \lambda \|h\|^2 \quad (6)$$

[0142] 具体地,在上述表达式(6)中,当Y表示球面谐波函数矩阵并且H表示球面谐波函数矩阵时,计算使左侧的E最小化的h。注意,上述表达式(6)的右侧的第二项是正则项,并且因此可以选择任意值作为 λ (例如, λ 可以为零)。因此,上述h由以下表达式(7)给出。

$$[0143] \quad h = (Y^T Y + \lambda I)^{-1} Y^T H \quad (7)$$

[0144] 通过使用上述表达式(7),HRTF学习单元136可以获得与必要频率相对应的每个h。另外,HRTF学习单元136通过主成分分析执行维度压缩,使得HRTF的信息量可以近似地以几百个维度表示(步骤S105)。该信息是HRTF参数(步骤S106)。

[0145] 注意,当在频率细化之后执行球面谐波拟合时,上述f的值是细化之后的代表频率。HRTF学习单元136可以在球面谐波拟合之后执行频率细化。空间上压缩维度的方法不限于诸如球面谐波函数和主成分分析的线性组合,而是可以是任何方法。例如,HRTF学习单元136可以使用诸如内核主成分分析的非线性方法。HRTF学习单元136可以根据频率f来改变球面谐波函数的截断阶数N,并且可以使用 $N(f)$ 的值。可以存在不用于0至N的维数或阶数的系数 h_{nm} 。HRTF学习单元136可以计算右HRTF和左HRTF,或者可以在转换为右HRTF和左HRTF的和或差之后计算右HRTF和左HRTF。拟合目标HRTF可以设置有例如转换为振幅的绝对值及其对数表示的各种转换。

[0146] 随后,HRTF学习单元136可以通过执行与步骤S101至S106的处理相反的处理来对HRTF进行解码。首先,HRTF学习单元136获取HRTF参数(步骤S107),并且通过主成分分析恢复维度压缩(步骤S108)。另外,HRTF学习单元136执行球面谐波重构处理(步骤S109)并且执行频率插值(步骤S110)。另外,HRTF学习单元136获得振幅特性(步骤S111)并且执行最小相位恢复(步骤S112)。最小相位恢复可以采用各种已知方法。例如,HRTF学习单元136对通过以奈奎斯特频率或更高的频率折叠上述函数 H'_2 而恢复的函数 $H'_1(\theta, \phi, k)$ 的对数执行快速傅里叶逆变换(IFFT),并且确定转换结果的实部。另外,在该区域中执行适当的窗口处理,对其指数函数执行快速傅里叶逆变换,并且确定实部,从而执行最小相位恢复。例如,如下所述的每个关系表达式(8)成立。

$$[0147] \quad H'_1(\theta, \phi, k) = \begin{cases} H'_2(\theta, \phi, k) & (\text{其中}, 0 \leq k \leq K/2) \\ H'_2(\theta, \phi, K - k) & (\text{其中}, K/2 < k \leq K - 1) \end{cases}$$

$$[0148] \quad H_c(\theta, \phi) = \text{Re} \{ F^{-1} [H'_1(\theta, \phi)] \}$$

$$[0149] \quad H_w(\theta, \phi, k) = \begin{cases} H_c(\theta, \phi, k) & (\text{其中}, K = 0, K/2) \\ 2H_c(\theta, \phi, k) & (\text{其中}, 1 \leq k \leq K/2 - 1) \\ 0 & (\text{其中}, K/2 \leq k \leq K - 1) \end{cases}$$

$$[0150] \quad H_m(\theta, \phi) = \text{Re} \{ F^{-1} [\exp(H_w(\theta, \phi))] \} \quad (8)$$

[0151] 注意, HRTF学习单元136可以将估计的耳间时间差 (ITD) 或预先准备的ITD添加到经受最小相位恢复的右HRIR ($h_{m,R}$) 和左HRIR ($h_{m,L}$)。注意, 基于右HRIR与左HRIR之间的组延迟差, 例如通过以下表达式 (9) 和 (10) 来计算ITD。

$$[0152] \quad \text{ITD}_g(\theta, \phi, f) = \frac{1}{2\pi} \left(\frac{d\varphi_L}{df} - \frac{d\varphi_R}{df} \right) \quad (9)$$

$$[0153] \quad \varphi_L = \arctan \left(\frac{\text{Im}(\text{HRTF}_L(f))}{\text{Re}(\text{HRTF}_L(f))} \right) \quad (10)$$

[0154] 可选地, 可以通过确定左右之间的时间轴互相关并且将ITD定义为相关系数最大的时间来计算ITD。在这种情况下, 例如通过以下表达式 (11) 和 (12) 来计算ITD。

$$[0155] \quad (\tau) = a \frac{\int_{-\infty}^{+\infty} h_L(t+\tau)h_R(t)dt}{\sqrt{\int_{-\infty}^{+\infty} h_L^2(t)dt \cdot \int_{-\infty}^{+\infty} h_R^2(t)dt}} \quad (11)$$

$$[0156] \quad \text{ITD}_{\text{corr}}(\theta, \phi) = \tau_{\text{max}} \quad (12)$$

[0157] 例如, HRTF学习单元136使用诸如以下表达式 (13) 的关系表达式将左HRIR相对于右HRIR延迟d个样本。

$$[0158] \quad h_L(i) = \begin{cases} 0 & \text{其中}, i < d \\ h_{m,L}(i-d) & \text{其中}, i \geq d \end{cases} \quad (13)$$

[0159] 在这种情况下, 上述表达式 (13) 中的 h_L 是比 $h_{m,L}$ 长d的脉冲响应, 并且通过删除上述表达式 (13) 的后半部分来获得等于 $h_{m,L}$ 的长度。在这种情况下, HRTF学习单元136可以例如执行可选窗口、矩形窗口或Hanning窗口处理。注意, HRTF学习单元136不仅可以针对每个方向添加ITD, 而且可以添加包括整个空间中的方向之间的相对时间差的延迟。在这种情况下, HRTF学习单元136不仅获取ITD, 而且获取指示方向之间的相对时间差的信息。当ITD是频率的函数时, HRTF学习单元136可以在频域中添加ITD, 或者可以计算代表值或平均值, 并且然后添加ITD。HRTF学习单元136获取原始格式的每个HRIR, 并且然后执行逆快速傅立叶变换以获得HRTF。

[0160] 以这种方式, HRTF学习单元136可以对信息量小于原始HRTF的信息量的HRTF参数执行压缩, 并且以压缩格式执行HRTF学习模型生成处理和稍后描述的HRTF计算处理。另外, 如上所述, HRTF压缩例如通过使用人类听力对相变的较小敏感度或执行不太可能影响听力的频率的优先细化来利用听觉特性执行维度降低。因此, HRTF学习单元136可以提高信息处理速度, 而不会失去作为HRTF特性的平稳听力。

[0161] 再次参考图3进行描述。估计单元140基于从用户发送的图像来执行与用户相对应

的HRTF的估计处理。

[0162] 获取单元141获取包括用户的耳朵的内容图像的图像。例如,获取单元141从由用户终端10捕获的图像中获取用户的耳朵周围的切出的耳朵图像。

[0163] 获取单元141可以将所获取的耳朵图像输入到耳朵参数估计模型以获取指示包括在图像中的耳朵的特性的耳朵参数。

[0164] 计算单元142基于由获取单元141获取的图像,通过使用学习模型(HRTF学习模型)来计算与用户相对应的HRTF(个性化HRTF),该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的HRTF。

[0165] 具体地,计算单元142通过将由获取单元141获取的耳朵参数输入到HRTF学习模型中来计算与用户相对应的个性化HRTF。

[0166] 注意,当计算个性化HRTF时,计算单元142可以首先计算HRTF参数,并且然后通过对所计算的HRTF参数进行解码来计算HRTF。以这种方式,计算单元142可以通过在压缩HRTF的信息量的同时执行一系列处理来提高处理速度。另外,计算单元142可以避免输出未在HRTF减小模型中表达的奇数HRTF,并且因此可以执行可靠的输出。

[0167] 提供单元143通过网络N将由计算单元142计算的HRTF提供给用户。

[0168] 以下将参考图13描述从图像估计HRTF的处理过程。图13是示出根据本公开的HRTF估计处理的过程的示意图。

[0169] 图13示出了估计单元140基于从用户发送的图像来执行与包括在图像中的耳朵相对应的HRTF的估计处理的示例。在这种情况下,用户通过使用用户终端10来执行用户的耳朵(准确地,包括耳朵的头部)的图像捕获(步骤S131)。此后,用户终端10执行在所捕获的图片中指定包括耳朵的内容图像的范围并且切出该指定范围以获取耳朵图像的预处理(步骤S132)。

[0170] 当获取从用户发送的耳朵图像时,获取单元141将所获取的耳朵图像输入到学习模型。具体地,获取单元141将耳朵图像输入到耳朵参数估计模型。耳朵参数估计模型输出与包括在耳朵图像中的耳朵的内容图像相对应的耳朵参数作为指示耳朵图像的特征量。因此,获取单元141获取与图像相对应的耳朵参数(步骤S133)。

[0171] 计算单元142将所获取的耳朵参数输入到HRTF学习模型中,并且计算与耳朵图像相对应的个性化HRTF(步骤S133)。提供单元143将所计算的个性化HRTF提供(发送)到从其发送图像的用户终端10(步骤S134)。

[0172] 以这种方式,当由学习单元131生成各种模型时,信息处理装置100可以执行对个性化HRTF提供的耳朵图像获取作为一系列处理。因此,信息处理装置100可以改善与针对用户的HRTF提供相关的便利性。

[0173] 注意,尽管在图13的示例中将耳朵参数估计模型和HRTF学习模型的组合示出为示例性学习模型,但是学习模型组合不限于该示例。学习模型可以是耳朵参数估计模型和HRTF学习模型的单独组合,或者可以被配置为执行与耳朵参数估计模型和HRTF学习模型相对应的处理的一个模型。

[0174] [1-4.根据第一实施例的用户终端的配置]

[0175] 如图13所示,在第一实施例中,由用户终端10执行用户的侧脸的图像捕获和耳朵图像的生成。以下将描述根据第一实施例的用户终端10的配置。图14是示出根据本公开的

第一实施例的用户终端10的示例性配置的示图。如图14所示,用户终端10包括通信单元11、输入单元12、显示单元13、感测单元14、存储单元15和控制单元16。

[0176] 通信单元11例如通过NIC来实现。通信单元11以有线或无线方式与网络N连接,并且通过网络N将信息发送到信息处理装置100等和从信息处理装置100等接收信息。

[0177] 输入单元12是被配置为从用户接收各种操作的输入装置。例如,输入单元12通过包括在用户终端10中的操作键等来实现。显示单元13是用于显示各种信息的显示装置。例如,显示单元13通过液晶显示器来实现。注意,当用户终端10采用触摸面板时,输入单元12的一部分和显示单元13彼此集成。

[0178] 感测单元14共同地参考各种传感器并且感测与用户终端10相关的各种信息。具体地,感测单元14感测用户对用户终端10的操作、用户终端10的位置信息、与用户终端10连接的设备相关的信息、用户终端10处的环境等。

[0179] 感测单元14包括用于执行图像捕获的透镜和图像传感器作为示例性传感器。具体地,例如,当由用户激活被配置为操作图像捕获功能的应用时,感测单元14用作相机。

[0180] 存储单元15存储各种信息。存储单元15例如由存储装置(诸如RAM或闪存的半导体存储元件、硬盘或光盘)来实现。存储单元15例如存储由用户捕获的图像。

[0181] 控制单元16是例如通过CPU或MPU通过将RAM用作工作区域来执行存储在用户终端10内部的存储装置中的各种计算机程序实现的控制器。可选地,控制单元16是由诸如ASIC或FPGA的集成电路实现的控制器。

[0182] 如图14所示,控制单元16包括获取单元161、预处理单元162、发送单元164和接收单元165,并且实现或执行下述信息处理的功能和效果。预处理单元162包括姿势检测单元163A和耳朵检测单元163B。注意,控制单元16的内部配置不限于图14所示的配置,而是可以是执行稍后描述的信息处理的任何配置。

[0183] 获取单元161获取各种信息。例如,获取单元161获取由感测单元14捕获的图像。

[0184] 姿势检测单元163A读取由获取单元161获取的图像,并且检测包括在图像中的用户的姿势。

[0185] 耳朵检测单元163B基于由姿势检测单元163A检测到的用户的姿势来检测包括图像中包括的用户的耳朵的范围(耳朵内容图像)。具体地,耳朵检测单元163B在包括用户的整个头部的内容图像的图像中指定用户的耳朵的内容图像,并且将指定范围检测为耳朵图像。

[0186] 例如,耳朵检测单元163B基于包括在整个图像中的用户的头部的每个特征点与用户的姿势之间的关系来指定包括耳朵的内容图像的范围。

[0187] 当基于包括在整个图像中的用户的头部的每个特征点与用户的姿势之间的关系不能指定包括耳朵的内容图像的范围时,姿势检测单元163A或耳朵检测单元163B可以新请求用户获取与整个图像不同并且包括用户的整个头部的内容图像的图像。具体地,姿势检测单元163A或耳朵检测单元163B通过在显示单元13上显示指示潜在地不能利用由用户捕获的侧脸的图像适当地执行根据本公开的信息处理的信息来提示用户再次执行图像捕获。注意,姿势检测单元163A或耳朵检测单元163B不仅在不能指定包括耳朵的内容图像的范围的情况下而且在例如用于学习耳朵参数估计模型的相机角度和用户的姿势超过特定阈值的情况下,可以提示用户再次执行图像捕获。另外,作为预处理,姿势检测单元163A或耳朵

检测单元163B可以生成校正用户在图像中的姿势和位置的校正信息,而不是检测用户的耳朵图像。例如,校正信息是指示根据用户的头部的特征点的倾斜和旋转来旋转包括耳朵等的图像的范围的量的信息。如下所述,基于用户的姿势、用户的侧脸与所检测的耳朵之间的位置关系等来生成这样的信息。在这种情况下,姿势检测单元163A或耳朵检测单元163B可以通过基于校正信息校正整个图像的旋转来指定用户的耳朵的内容图像,并且可以将指定范围检测为耳朵图像。姿势检测单元163A或耳朵检测单元163B可以将整个图像连同所生成的校正信息一起发送到信息处理装置100。在这种情况下,信息处理装置100在自己的装置上执行通过基于与整个图像一起发送的校正信息来校正整个图像的旋转并且将指定范围检测为耳朵图像来指定用户的耳朵的内容图像的预处理。

[0188] 以下将参考图15描述由预处理单元162(姿势检测单元163A和耳朵检测单元163B)执行的预处理的过程。图15是示出根据本公开的检测处理的过程的示图。

[0189] 如图15所示,当用户捕获了用户的侧脸时,获取单元161获取整个图像50(步骤S141)。

[0190] 姿势检测单元163A在整个获取的图像50中检测用户的侧脸(步骤S142)。例如,姿势检测单元163A通过使用诸如人脸检测处理的已知技术在整个图像50中指定包括用户的侧脸的内容图像的范围。

[0191] 然后,姿势检测单元163A检测包括在用户的侧脸中的特征点,如图像51所示。例如,姿势检测单元163A检测侧脸中诸如沿水平方向突出的位置(具体地,用户的鼻子的顶点)、头部的顶点、嘴的位置以及下巴的位置的特征点。姿势检测单元163A还基于头发和皮肤等的边界的信息来检测例如用户的耳朵和鬓角的位置。姿势检测单元163A还基于侧脸的内容图像的颜色信息来检测用户的眼睛的位置等。

[0192] 然后,姿势检测单元163A基于所检测的特征点来检测用户的姿势(步骤S143)。例如,姿势检测单元163A基于特征点的三维布置来检测用户的头部的姿势,如图像54所示。

[0193] 这样的姿势检测处理是用于防止由用户发送的耳朵图像中的姿势与学习时使用的3D模型的姿势大幅度偏离的处理。这是因为当从用户终端10发送包括与3D模型的姿势大大不同的姿势的图像时,由于学习数据与所发送的耳朵图像之间的偏差,信息处理装置100潜在地不能适当地执行耳朵图像识别。

[0194] 因此,姿势检测单元163A确定在通过渲染学习中使用的头部3D模型55的角度的平均值与从图像54获得的角度之间的差是否等于或小于预定阈值,从而执行用户是否已经适当地执行了图像捕获的成功确定(步骤S144)。例如,假设在学习耳朵参数估计模型时,渲染头部3D模型55时相机的定向与连接头部顶点和耳朵(例如,外耳道的入口)的预定位置的线段之间的角度 ϕ 等于或小于预定数值。类似地,假设在学习耳朵参数估计模型时,相机的定向与连接鼻子的顶点和耳朵的预定位置的线段之间的角度 θ 等于或小于预定数值。这是因为在学习时使用的耳朵图像不会与示出人的侧脸的图像有很大的偏离以提高图像识别精度。具体地,与学习时的图像类似,姿势检测单元163A确定从用户发送的图像是否具有可识别为示出人的侧脸的图像的角度。

[0195] 当已经确定用户没有适当地执行图像捕获时(例如,当在用户的面部鼻子超过预定阈值向下指向时),姿势检测单元163A例如执行显示再次请求执行图像捕获的消息,并且然后获取新捕获的图像的处理(步骤S145)。

[0196] 当已经确定用户已经适当地执行了图像捕获时(步骤S146),耳朵检测单元163B在图像56中指定包括耳朵的内容图像的范围57并且切出该范围57(步骤S147)。因此,耳朵检测单元163B获取耳朵图像58。

[0197] 通过执行图15所示的检测处理,信息处理装置100可以确定用户的耳朵是由于不良的图像捕获状态而倾斜还是耳朵的角度实际上倾斜,并且然后执行HRTF计算。

[0198] 另外,如上所述,用户终端10从侧脸的整个图像中切出耳朵图像,并且因此,可以发送不是包括用户的面部的整个图像而是仅耳朵图像以执行处理。因此,用户终端10可以防止私人信息的泄漏,从而增加信息处理的安全性。注意,用户终端10不限于上述检测方法,而是可以通过使用通过机器学习等的图像识别技术来检测包括在图像中的用户的耳朵,来执行从侧脸的整个图像中切出耳朵图像的处理。

[0199] 再次参考图14进行描述。发送单元164将基于由耳朵检测单元163B检测到的范围生成的耳朵图像发送到信息处理装置100。

[0200] 接收单元165接收由信息处理装置100提供的个性化HRTF。例如,接收单元165在声音回放应用等中将所接收的个性化HRTF合并到音乐和语音中,从而实现针对个体用户优化的3D声学等。

[0201] (2. 第二实施例)

[0202] 随后,以下将描述第二实施例。第一实施例在以上示例中描述了用户终端10通过从由用户捕获的图像中仅切出耳朵的内容图像来生成耳朵图像的示例。根据第二实施例的信息处理装置100A通过自己的装置代替用户终端10执行仅切出耳朵的内容图像的处理。

[0203] 以下将参考图16描述根据第二实施例的信息处理装置100A的配置。图16是示出根据本公开第二实施例的信息处理装置100A的示例性配置的示图。如图16所示,除了第一实施例的配置之外,信息处理装置100A进一步包括预处理单元144(姿势检测单元145A和耳朵检测单元145B)。

[0204] 姿势检测单元145A执行与由根据第一实施例的姿势检测单元163A执行的相同的处理。耳朵检测单元145B执行与由根据第一实施例的耳朵检测单元163B执行的相同的处理。因此,根据第二实施例的信息处理装置100A通过自己的装置执行由根据第一实施例的用户终端10执行的预处理。

[0205] 在第二实施例中,获取单元141从用户终端10获取由用户捕获的侧脸的整个图像。然后,姿势检测单元145A和耳朵检测单元145B通过执行与参考图15描述的处理相同的处理,基于整个图像生成耳朵图像。计算单元142基于由姿势检测单元145A和耳朵检测单元145B生成的耳朵图像来计算个性化HRTF。

[0206] 以这种方式,利用根据第二实施例的信息处理装置100A,用户可以仅通过捕获和发送图像来接收个性化HRTF提供。另外,利用第二实施例的配置,由于不需要在用户终端10上执行预处理,因此例如可以减少用户终端10上的处理负荷。此外,由于可以假设服务器装置(信息处理装置100A)处的处理速度通常比用户终端10高,因此可以利用第二实施例的配置来提高根据本公开的信息处理的整体速度。注意,当校正信息与整个图像一起发送时,姿势检测单元145A和耳朵检测单元145B可以通过基于包括在整个图像中的校正信息来校正整个图像的旋转来指定用户的耳朵的内容图像,并且可以将指定范围检测为耳朵图像。

[0207] (3. 其他实施例)

[0208] 除了上述实施例之外,可以以各种不同的形式执行根据上述实施例的处理。

[0209] 在以上实施例中描述的处理中,可以手动执行被描述为自动执行的处理的全部或一些处理,或者可以通过公知方法自动执行被描述为手动执行的处理的全部或一些处理。另外,除非另外说明,否则可以任选地改变在上述文档和附图中指示的包括处理过程、特定名称以及各种数据和参数的信息。例如,在每个附图中示出的各种信息不限于所示信息。

[0210] 附图中所示的装置的组件表示概念性功能,并且不必在物理上如图所示配置。换句话说,装置的分散和集成的具体形式不限于所示形式,并且所有或一些装置可以根据各种负担和使用情况以可选单位在功能上或物理上分散和集成。

[0211] 在处理内容不一致的情况下,可以适当地组合上述实施例和修改。

[0212] 本说明书中描述的效果仅是示例性的而非限制性的,但是可以实现任何其他效果。

[0213] (4. 硬件配置)

[0214] 通过例如具有如图17所示的配置的计算机1000来实现根据每个上述实施例的诸如信息处理装置100或用户终端10的信息设备。以下描述根据第一实施例的信息处理装置100的示例。图17是示出被配置为实现信息处理装置100的功能的示例性计算机1000的硬件配置图。计算机1000包括CPU 1100、RAM 1200,只读存储器(ROM) 1300、硬盘驱动器(HDD) 1400、通信接口1500和输入输出接口1600。计算机1000的组件通过总线1050彼此连接。

[0215] CPU 1100基于存储在ROM 1300或HDD 1400中的计算机程序进行操作,并且执行每个组件的控制。例如,CPU 1100通过将存储在ROM 1300或HDD 1400中的计算机程序加载到RAM 1200上来执行与各种计算机程序相对应的处理。

[0216] ROM 1300存储诸如在计算机1000激活时由CPU 1100执行的基本输入/输出系统(BIOS)的启动程序、取决于计算机1000的硬件的计算机程序等。

[0217] HDD 1400是其中以非临时方式记录由CPU 1100执行的计算机程序、由该计算机程序使用的数据等的计算机可读记录介质。具体地,HDD 1400是记录作为示例性程序数据1450的根据本公开的信息处理程序的记录介质。

[0218] 通信接口1500是用于将计算机1000与外部网络1550(例如,因特网)连接的接口。例如,通过通信接口1500,CPU 1100从另一设备接收数据,并且将由CPU 1100生成的数据发送到另一设备。

[0219] 输入输出接口1600是用于将计算机1000与输入输出装置1650连接的接口。例如,CPU 1100通过输入输出接口1600从诸如键盘或鼠标的输入装置接收数据。CPU 1100通过输入输出接口1600将数据发送到诸如显示器、扬声器或打印机的输出装置。输入输出接口1600可以用作用于读取记录在预定记录介质中的计算机程序等的介质接口。介质例如是诸如数字通用光盘(DVD)或相变可重写盘(PD)的光记录介质、诸如磁光盘(MO)的磁光记录介质、磁带介质、磁记录介质或半导体存储器。

[0220] 例如,当计算机1000用作根据第一实施例的信息处理装置100时,计算机1000的CPU 1100通过执行加载到RAM 1200上的信息处理程序来实现控制单元130等的功能。HDD 1400将根据本公开的信息处理程序和数据存储在存储单元120中。注意,CPU 1100从HDD 1400读取程序数据1450并执行程序数据1450,但是例如可以通过外部网络1550从另一装置获取这些计算机程序。

[0221] 注意,可以如下所述配置本技术。

[0222] (1)

[0223] 一种信息处理装置,包括:

[0224] 获取单元,其被配置为获取包括用户的耳朵的内容图像的第一图像;以及

[0225] 计算单元,其被配置为基于由获取单元获取的第一图像,通过使用学习模型来计算与用户相对应的头部相关传递函数,该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的头部相关传递函数。

[0226] (2)

[0227] 根据(1)的信息处理装置,其中,

[0228] 获取单元获取表示包括在第一图像中的耳朵的特性的变量的耳朵参数,以及

[0229] 计算单元通过将耳朵参数输入到学习模型来计算与用户相对应的头部相关传递函数。

[0230] (3)

[0231] 根据(2)的信息处理装置,其中,获取单元通过使用耳朵参数估计模型来获取包括在第一图像中的耳朵的耳朵参数,该耳朵参数估计模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的耳朵参数。

[0232] (4)

[0233] 根据(3)的信息处理装置,进一步包括:第一学习单元,其被配置为通过学习包括耳朵的内容图像的图像与耳朵的耳朵参数之间的关系来生成耳朵参数估计模型。

[0234] (5)

[0235] 根据(4)的信息处理装置,其中,第一学习单元通过学习耳朵参数与通过渲染基于耳朵参数生成的耳朵的三维数据而获得的耳朵图像之间的关系来生成耳朵参数估计模型。

[0236] (6)

[0237] 根据(5)的信息处理装置,其中,第一学习单元通过学习通过改变耳朵或头部的三维数据的纹理、渲染中的相机角度或渲染中的亮度而获得的多个耳朵图像与该多个耳朵图像共有的耳朵参数之间的关系来生成耳朵参数估计模型。

[0238] (7)

[0239] 根据(1)至(6)中任一项的信息处理装置,进一步包括:第二学习单元,其被配置为通过学习包括耳朵的内容图像的图像与和耳朵相对应的头部相关传递函数之间的关系来生成学习模型。

[0240] (8)

[0241] 根据(7)的信息处理装置,其中,第二学习单元对通过合成基于耳朵参数生成的耳朵的三维数据和头部的三维数据而获得的三维数据执行声学模拟,并且通过学习通过声学模拟获得的头部相关传递函数与耳朵参数之间的关系来生成学习模型。

[0242] (9)

[0243] 根据(8)的信息处理装置,其中,第二学习单元压缩通过声学模拟获得的头部相关传递函数的信息量,并且通过学习所压缩的头部相关传递函数与耳朵参数之间的关系来生成学习模型。

[0244] (10)

[0245] 根据(8)或(9)的信息处理装置,其中,第二学习单元设置基于耳朵参数生成的耳朵的三维数据的听力点,并且通过使用所设置的听力点来执行声学模拟。

[0246] (11)

[0247] 根据(1)至(10)中任一项的信息处理装置,进一步包括:预处理单元,其被配置为在包括用户的整个头部的内容图像的第二图像中指定用户的耳朵的内容图像,并且将指定范围检测为第一图像,其中,获取单元获取由预处理单元检测到的第一图像。

[0248] (12)

[0249] 根据(11)的信息处理装置,其中,预处理单元基于包括在第二图像中的用户的头部的特征点与用户的姿势之间的关系来指定范围。

[0250] (13)

[0251] 根据(12)的信息处理装置,其中,当基于包括在第二图像中的用户的头部的特征点与用户的姿势之间的关系不能指定范围时,预处理单元新请求获取与第二图像不同并且包括用户的整个头部的内容图像的图像。

[0252] (14)

[0253] 根据(11)至(13)中任一项的信息处理装置,预处理单元通过基于包括在第二图像中的校正信息校正第二图像的旋转来指定用户的耳朵的内容图像,并且将指定范围检测为第一图像。

[0254] (15)

[0255] 一种信息处理方法,计算机通过该信息处理方法执行以下操作:

[0256] 获取包括用户的耳朵的内容图像的第一图像;并且

[0257] 基于所获取的第一图像,通过使用学习模型来计算与用户相对应的头部相关传递函数,该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的头部相关传递函数。

[0258] (16)

[0259] 一种信息处理程序,其被配置为使计算机用作:

[0260] 获取单元,其被配置为获取包括用户的耳朵的内容图像的第一图像;以及

[0261] 计算单元,其被配置为基于由获取单元获取的第一图像,通过使用学习模型来计算与用户相对应的头部相关传递函数,该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的头部相关传递函数。

[0262] (17)

[0263] 一种包括信息处理装置和用户终端的信息处理系统,其中,

[0264] 用户终端包括:

[0265] 预处理单元,其被配置为在包括用户的整个头部的内容图像的第二图像中指定用户的耳朵的内容图像,并且将指定范围检测为第一图像,以及

[0266] 发送单元,其被配置为将由预处理单元检测到的第一图像发送到信息处理装置,并且

[0267] 信息处理装置包括:

[0268] 获取单元,其被配置为获取包括用户的耳朵的内容图像的第一图像,以及

[0269] 计算单元,其被配置为基于由获取单元获取的第一图像,通过使用学习模型来计

算与用户相对应的头部相关传递函数,该学习模型已经学习以在输入包括耳朵的内容图像的图像时输出与耳朵相对应的头部相关传递函数。

- [0270] 参考标记列表
- [0271] 1 信息处理系统
- [0272] 10 用户终端
- [0273] 100 信息处理装置
- [0274] 110 通信单元
- [0275] 120 存储单元
- [0276] 130 控制单元
- [0277] 131 学习单元
- [0278] 132 接收单元
- [0279] 133 耳朵模型学习单元
- [0280] 134 图像生成单元
- [0281] 135 耳朵参数学习单元
- [0282] 136 HRTF学习单元
- [0283] 140 估计单元
- [0284] 141 获取单元
- [0285] 142 计算单元
- [0286] 143 提供单元
- [0287] 144 预处理单元
- [0288] 145A 姿势检测单元
- [0289] 145B 耳朵检测单元
- [0290] 161 获取单元
- [0291] 162 预处理单元
- [0292] 163A 姿势检测单元
- [0293] 163B 耳朵检测单元
- [0294] 164 发送单元
- [0295] 165 接收单元。

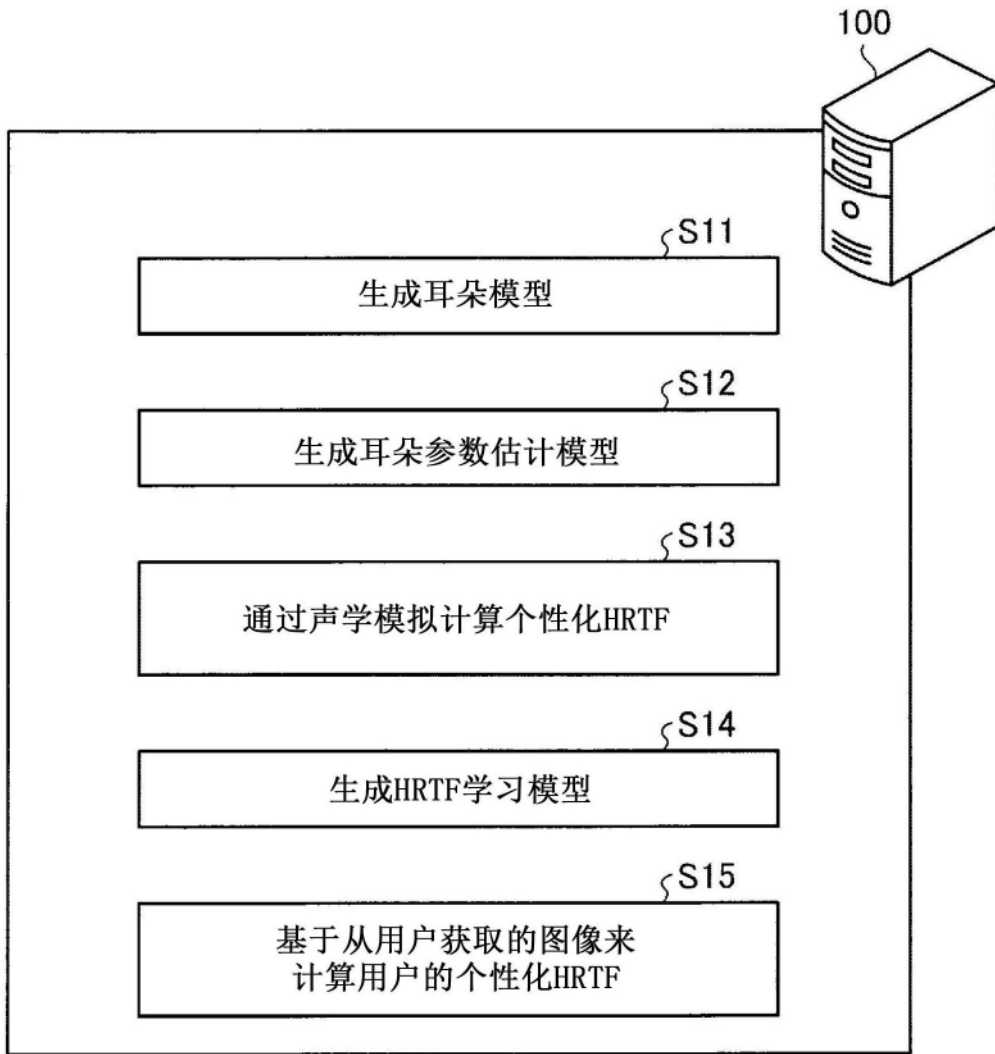


图2

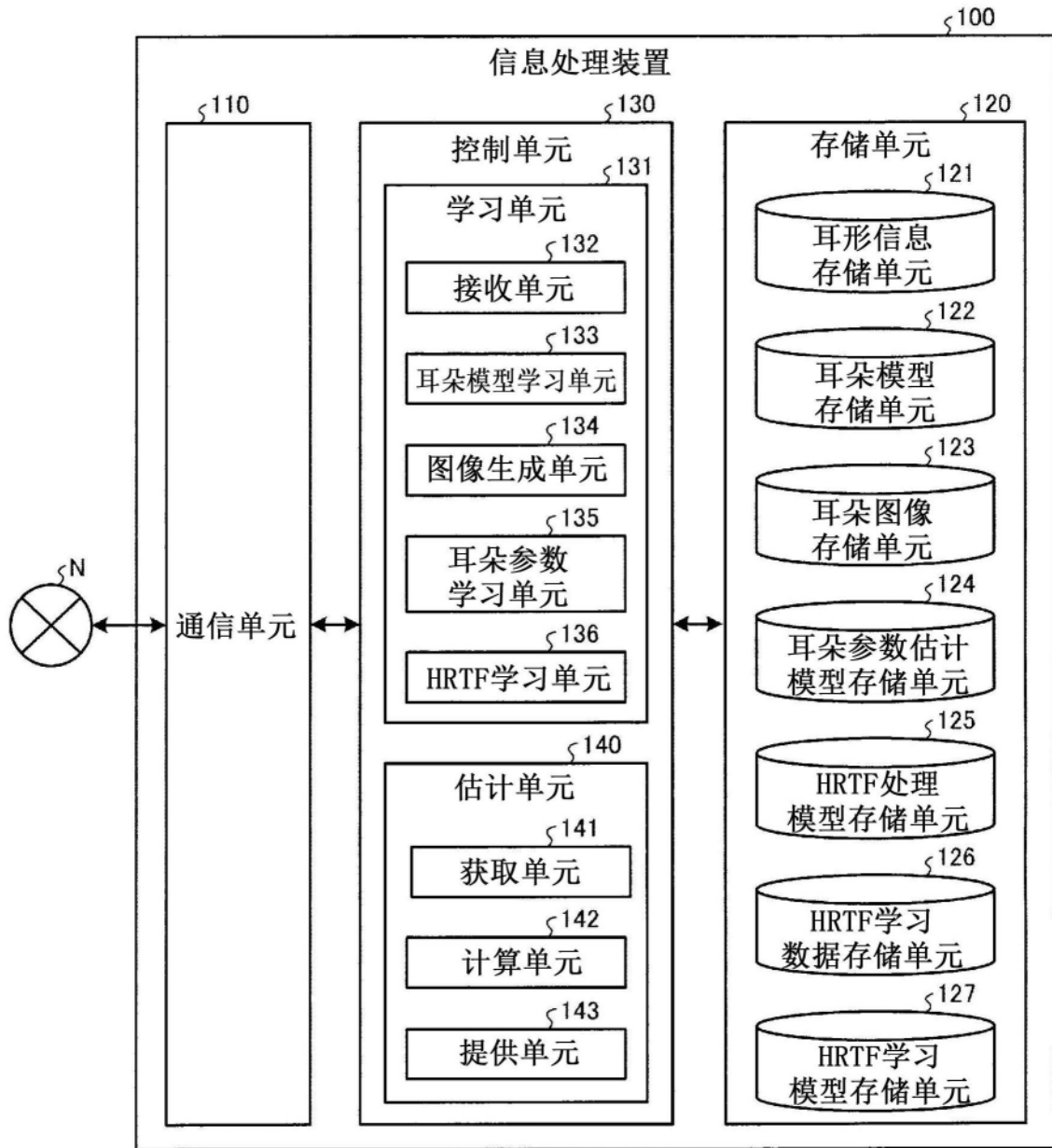


图3

123

耳朵参数	耳朵3D模型数据	头部3D模型数据	耳朵图像ID	图像生成参数					...
				纹理	相机角度	分辨率	亮度	...	
...
A01	B01	C01	D01	E01	F01	G01	H01
			D02	E02	F02	G02	H02
			D03	E03	F03	G03	H03
...

图4

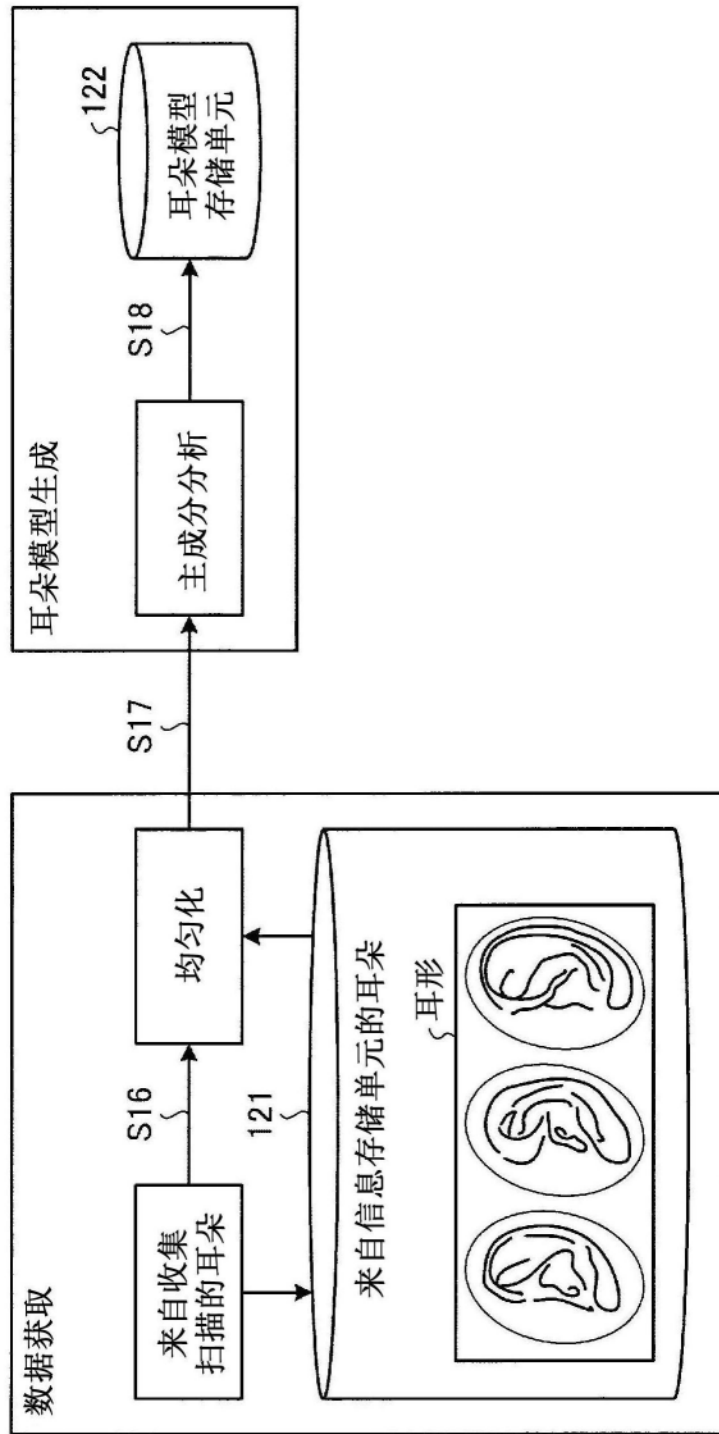


图5

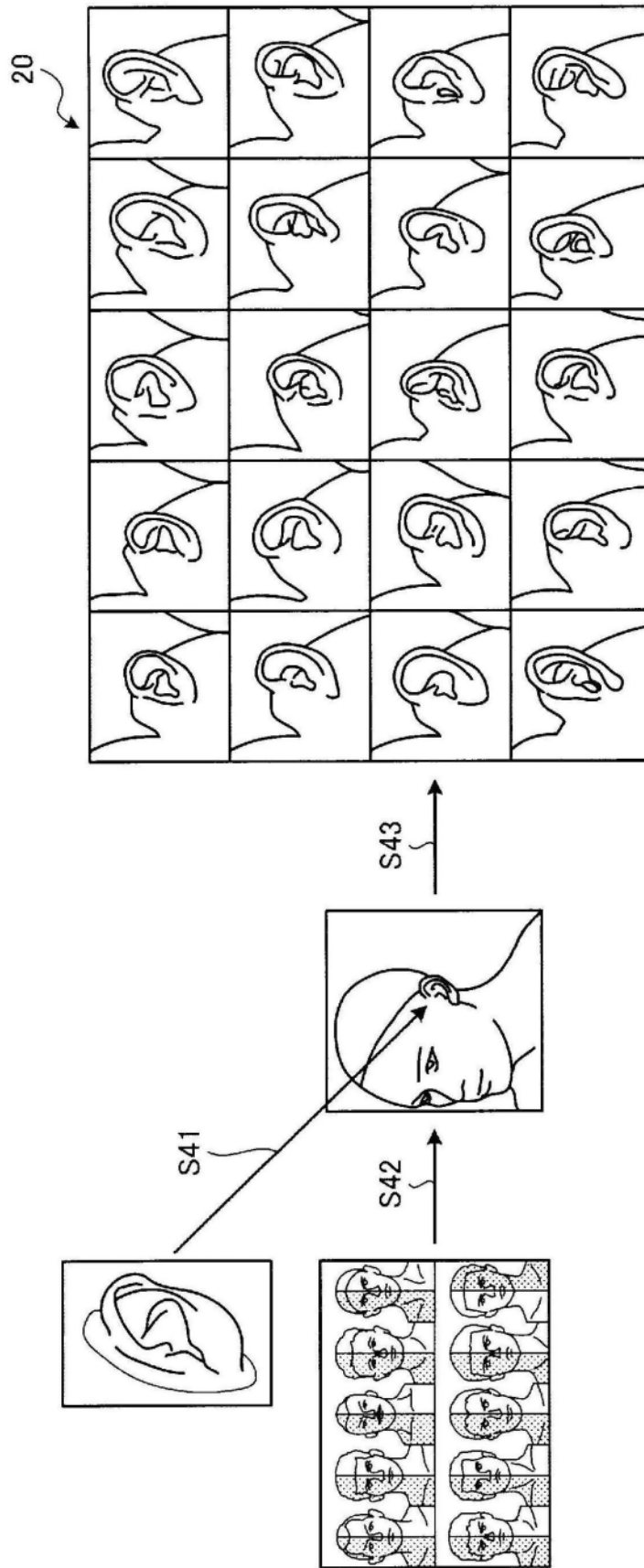


图7

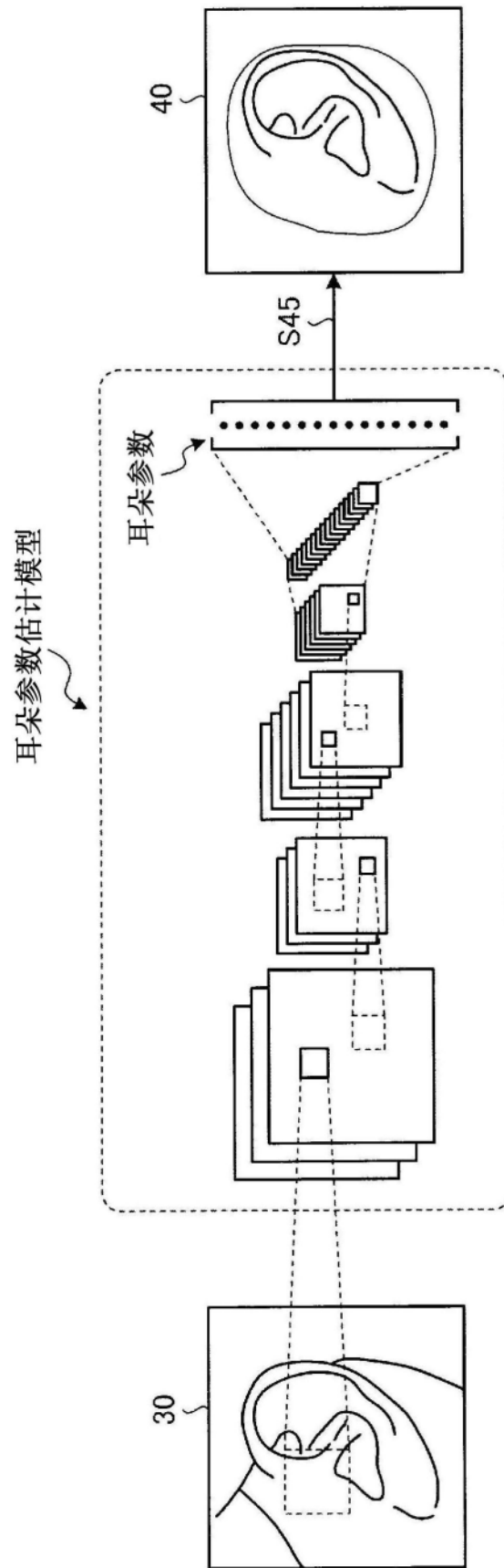


图8

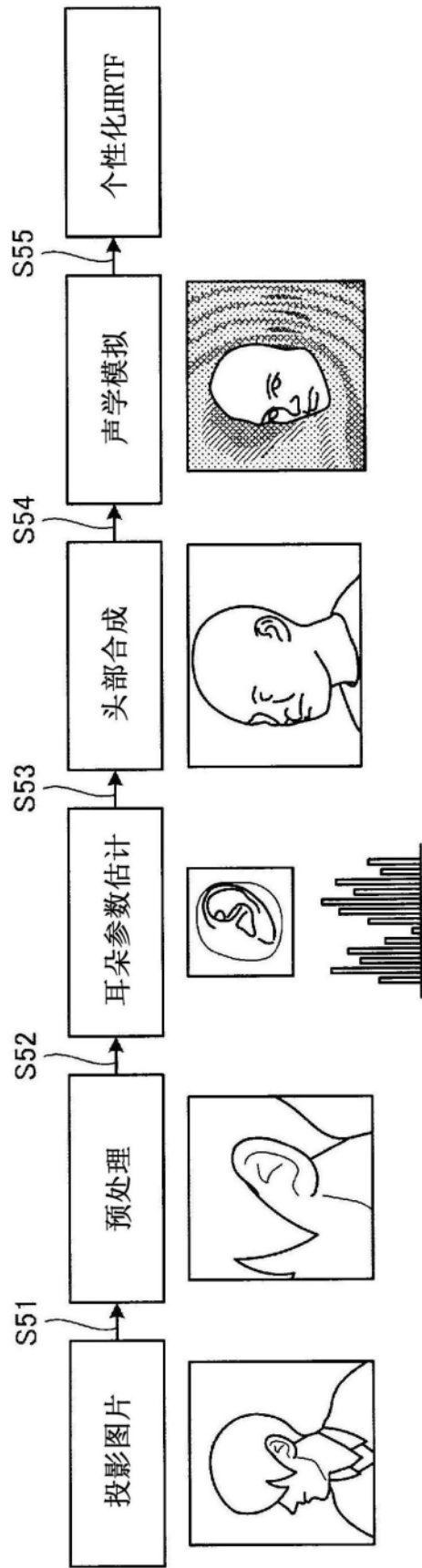


图9

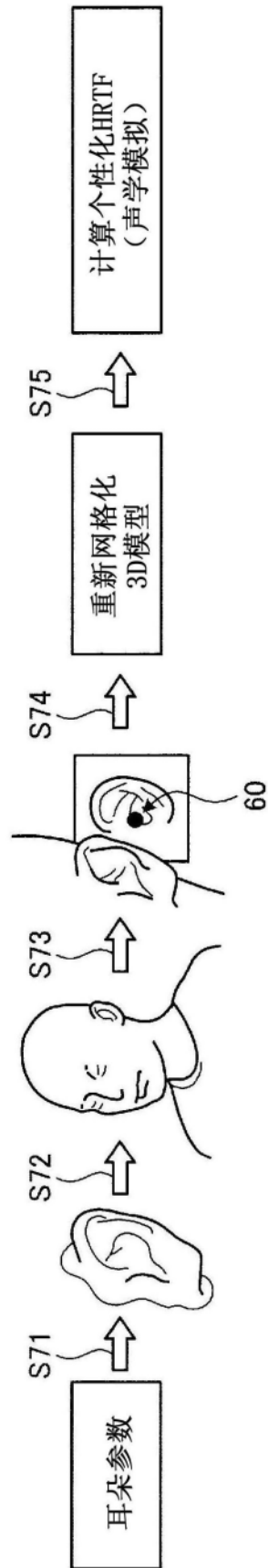


图10

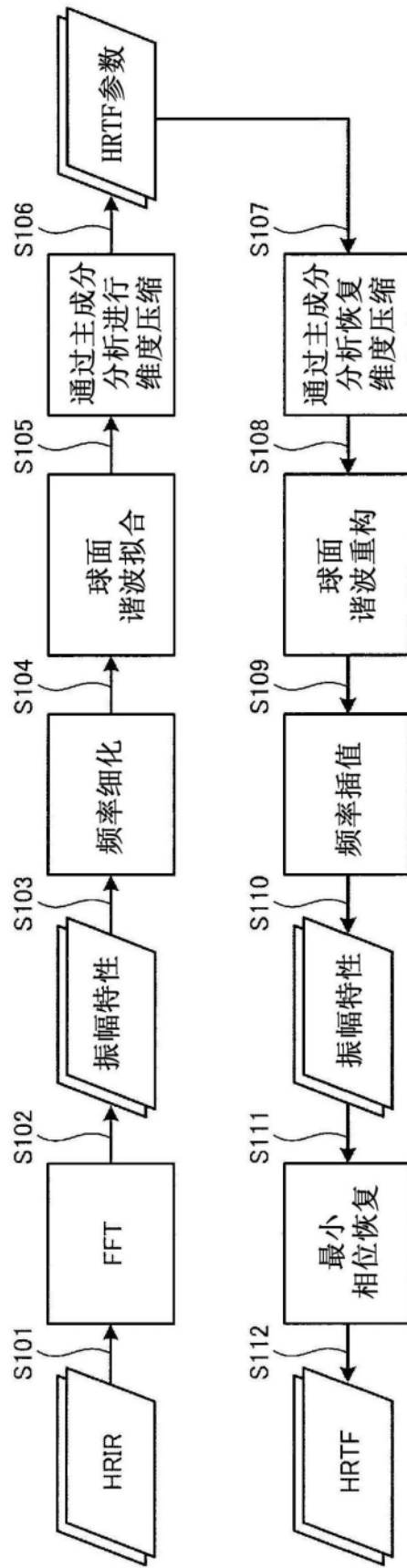


图12

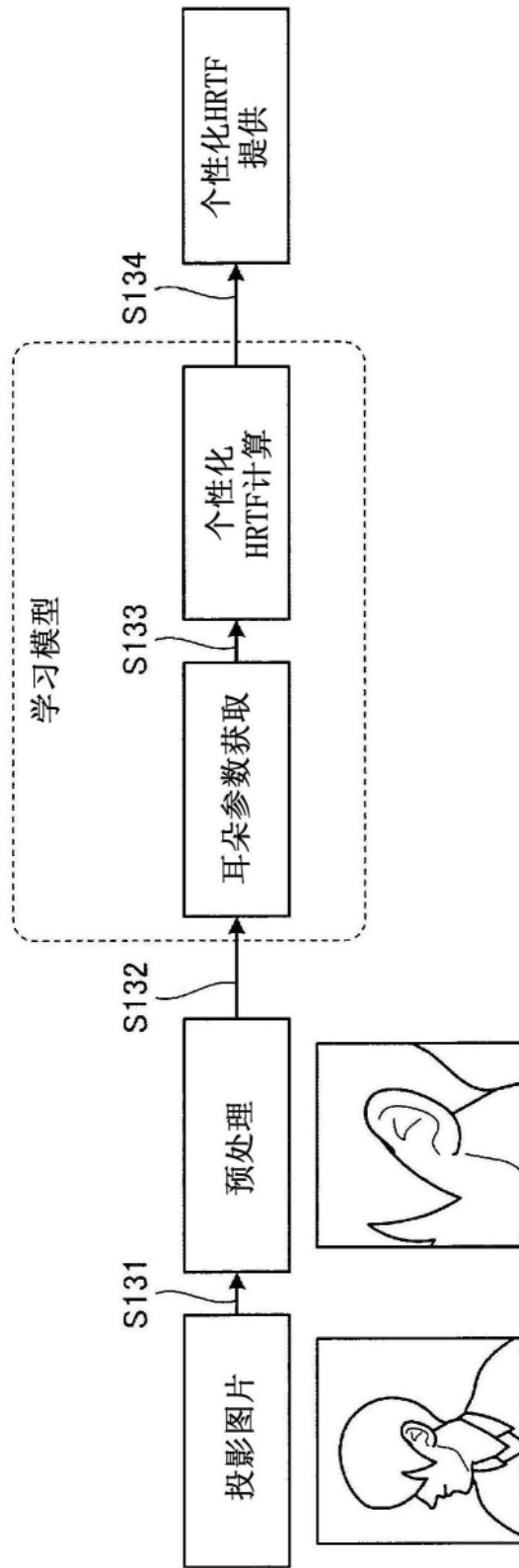


图13

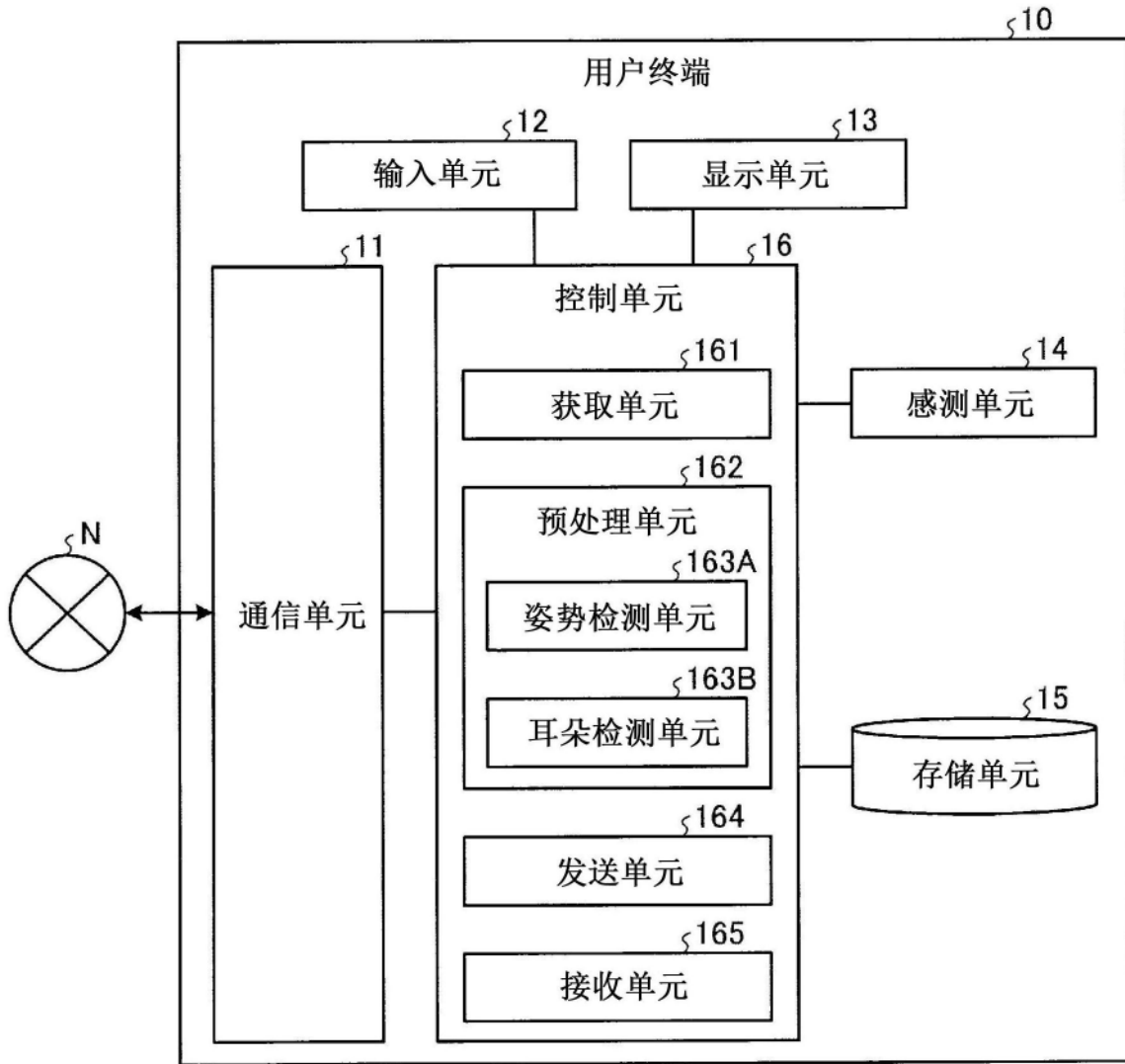


图14

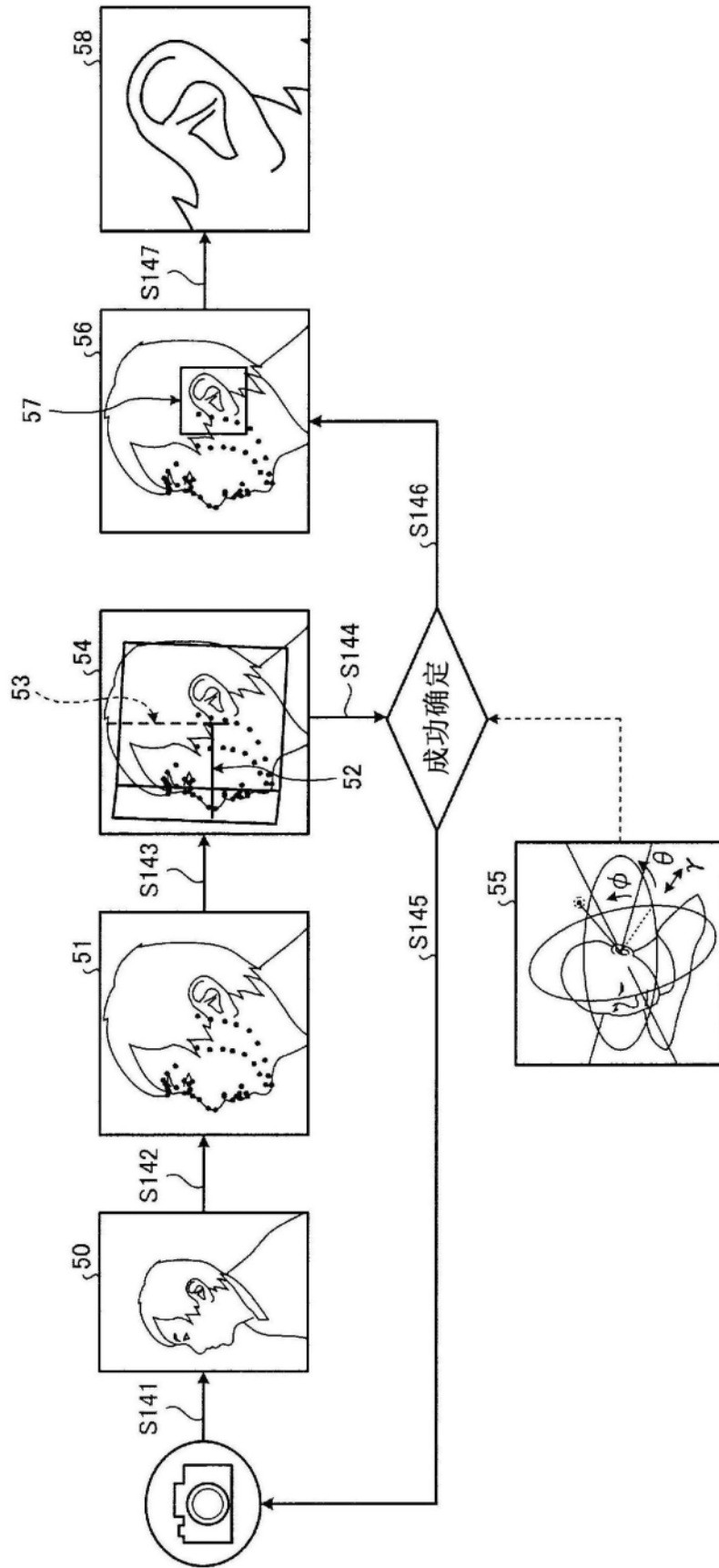


图15

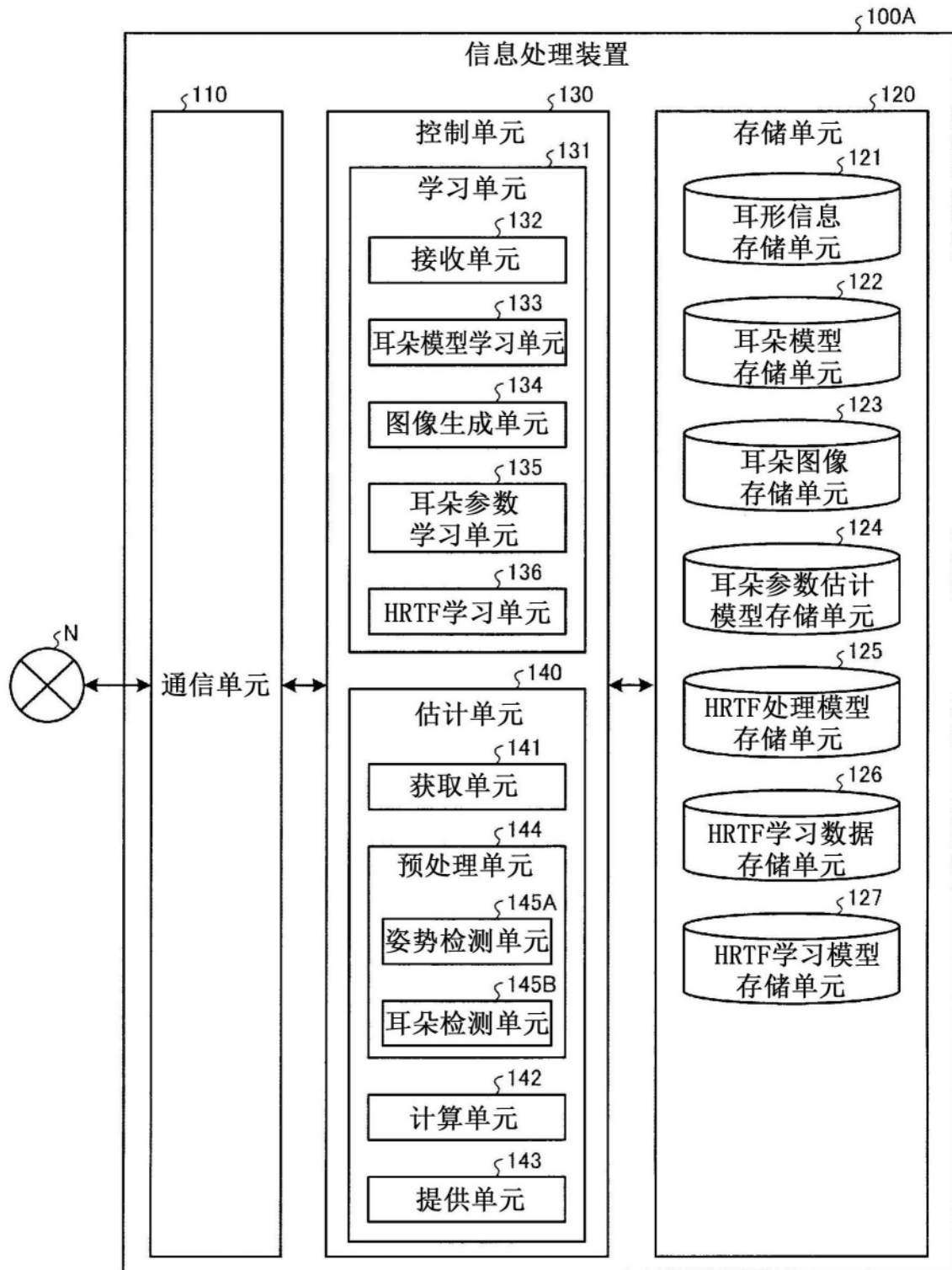


图16

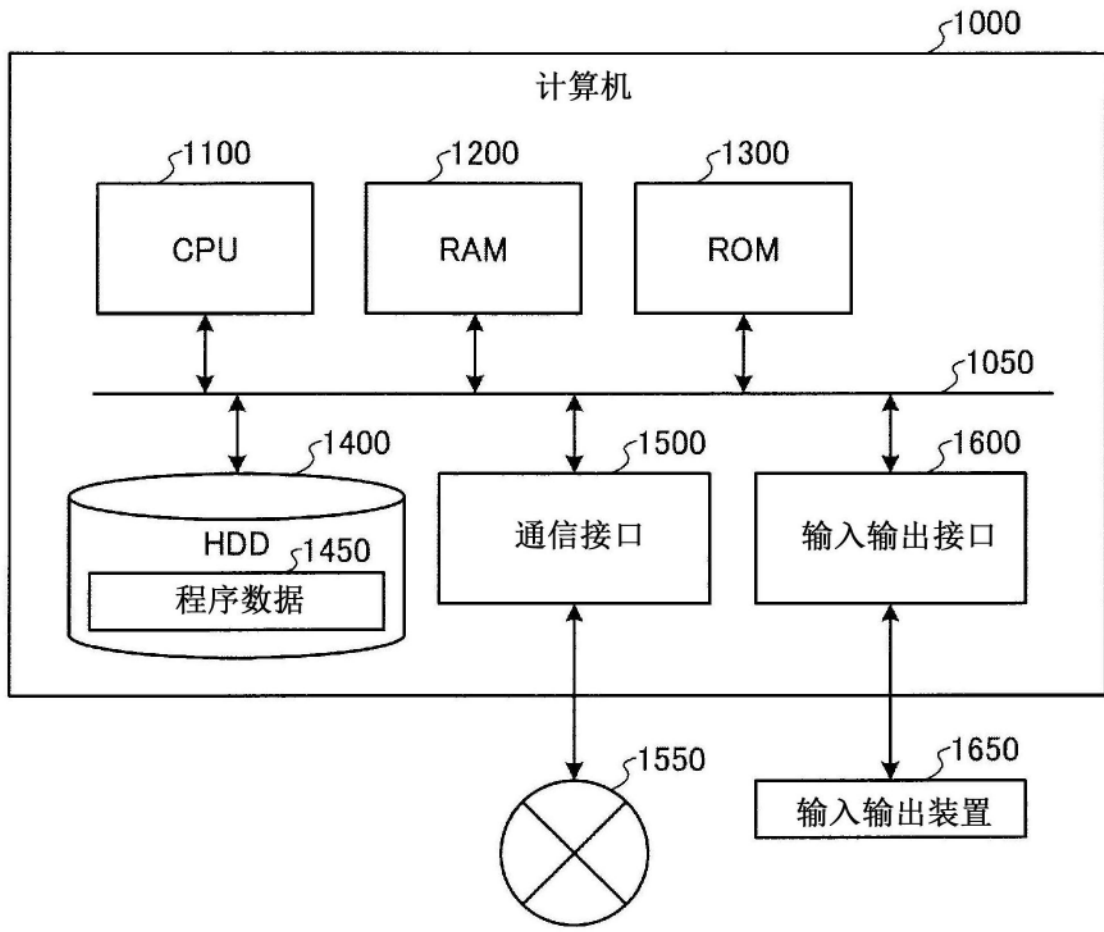


图17