



(12) 发明专利申请

(10) 申请公布号 CN 111788043 A

(43) 申请公布日 2020.10.16

(21) 申请号 201980016177.1

(22) 申请日 2019.02.21

(30) 优先权数据

2018-069788 2018.03.30 JP

(85) PCT国际申请进入国家阶段日

2020.08.28

(86) PCT国际申请的申请数据

PCT/JP2019/006580 2019.02.21

(87) PCT国际申请的公布数据

W02019/187834 JA 2019.10.03

(71) 申请人 索尼公司

地址 日本东京

(72) 发明人 泽田充奖 小山裕一郎

(74) 专利代理机构 北京康信知识产权代理有限公司 11240

代理人 吴孟秋

(51) Int.Cl.

B25J 13/00 (2006.01)

A63H 3/33 (2006.01)

A63H 11/00 (2006.01)

A63H 11/18 (2006.01)

A63H 30/02 (2006.01)

G10L 15/00 (2013.01)

G10L 25/84 (2013.01)

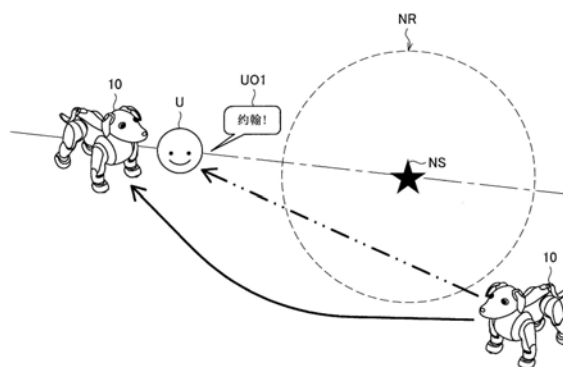
权利要求书2页 说明书20页 附图25页

(54) 发明名称

信息处理装置、信息处理方法和程序

(57) 摘要

[问题]使自主移动体执行运动以进一步提高语音识别的精度。[解决方案]提供一种信息处理装置,该信息处理装置配备有运动控制单元,用于控制自主移动体的运动以便基于识别处理动作,其中,当已经检测到作为要经受语音识别处理的声音的目标声音时,运动控制单元使自主移动体移动到在基于目标声音确定的接近目标的周边作为除了对象声音之外的声音的非目标声音的输入水平变低的位置。还提供一种信息处理方法,该信息处理方法包括由处理器执行的步骤,该步骤是控制自主移动体的运动以便基于识别处理动作。控制还包括:当已经检测到作为要经受语音识别处理的声音的目标声音时,将自主移动体移动到在基于目标声音确定的接近目标的周边作为对象声音之外的声音的非目标声音的输入水平变低的位置。



1. 一种信息处理装置,包括:
操作控制单元,控制基于识别处理动作的自主移动体的运动,
在检测到作为用于语音识别处理的目标语音的目标声音的情况下,所述操作控制单元将所述自主移动体移动到基于所述目标声音确定的接近目标周围不是所述目标语音的非目标声音的输入水平变低的位置。
2. 根据权利要求1所述的信息处理装置,其中,在检测到所述目标声音的情况下,所述操作控制单元将所述自主移动体移动到基于所述目标声音确定的所述接近目标周围所述目标声音的输入水平变高并且所述非目标声音的输入水平变低的位置。
3. 根据权利要求1所述的信息处理装置,其中,在检测到所述目标声音的情况下,所述操作控制单元将所述自主移动体移动到远离发出所述非目标声音的噪声源并且靠近所述接近目标的位置。
4. 根据权利要求1所述的信息处理装置,其中,在检测到所述目标声音的情况下,所述操作控制单元相对于以所述接近目标为中心将所述自主移动体移动到与发出所述非目标声音的噪声源相反的一侧。
5. 根据权利要求1所述的信息处理装置,其中,
所述目标声音包括用户的发声语音,并且
所述接近目标包括发出所述发声语音的发声用户。
6. 根据权利要求1所述的信息处理装置,其中,所述接近目标包括由基于用户的发声语音的所述语音识别处理识别的移动体、固定物体或位置。
7. 根据权利要求1所述的信息处理装置,其中,所述操作控制单元基于指示在周围环境中生成所述非目标声音的情况的噪声图来控制所述自主移动体的运动。
8. 根据权利要求7所述的信息处理装置,其中,
所述噪声图包括发出所述非目标声音的噪声源的信息,并且
所述操作控制单元基于所述噪声源的回避优先级来控制所述自主移动体的运动。
9. 根据权利要求8所述的信息处理装置,其中,基于所述噪声源的类型来确定所述噪声源的回避优先级。
10. 根据权利要求8所述的信息处理装置,其中,基于由所述噪声源发出的所述非目标声音对所述语音识别处理的影响度来确定所述噪声源的回避优先级。
11. 根据权利要求10所述的信息处理装置,其中,基于指示所述非目标声音的声音似然度的指标和指示稳定度的指标中的至少一个来计算所述影响度。
12. 根据权利要求7所述的信息处理装置,其中,在未检测到所述目标声音的情况下,所述操作控制单元基于所述噪声图来控制所述自主移动体的运动以避免所述非目标声音的输入。
13. 根据权利要求7所述的信息处理装置,其中,在未检测到所述目标声音的情况下,所述操作控制单元基于所述噪声图将所述自主移动体的移动范围限制在所述非目标声音的输入水平等于或小于阈值的区域内。
14. 根据权利要求7所述的信息处理装置,还包括创建所述噪声图的周围环境估计单元。
15. 根据权利要求14所述的信息处理装置,其中,所述周围环境估计单元基于发出所述

非目标声音的噪声源的方向估计或声压测量来创建所述噪声图。

16. 根据权利要求14所述的信息处理装置, 其中, 所述周围环境估计单元基于所收集的非目标声音动态地更新所述噪声图。

17. 根据权利要求16所述的信息处理装置, 其中, 所述周围环境估计单元基于发出所述非目标声音的所述噪声源的数量、位置或声压的变化来动态地更新所述噪声图。

18. 根据权利要求16所述的信息处理装置, 其中, 所述周围环境估计单元基于用户存在于周围环境中的时区中收集的所述非目标声音来创建或更新所述噪声图。

19. 一种信息处理方法, 包括使处理器:

控制基于识别处理动作的自主移动体的运动,

所述控制还包括: 在检测到作为用于语音识别处理的目标语音的目标声音的情况下, 将所述自主移动体移动到基于所述目标声音确定的接近目标周围不是所述目标语音的非目标声音的输入水平变低的位置。

20. 一种程序, 使计算机用作信息处理装置,

所述信息处理装置包括:

操作控制单元, 控制基于识别处理动作的自主移动体的运动,

在检测到作为用于语音识别处理的目标语音的目标声音的情况下, 所述操作控制单元将所述自主移动体移动到基于所述目标声音确定的接近目标周围不是所述目标语音的非目标声音的输入水平变低的位置。

信息处理装置、信息处理方法和程序

技术领域

[0001] 本公开涉及一种信息处理装置、信息处理方法和程序。

背景技术

[0002] 近年来,已经开发了估计声源(诸如用户的话语)的方向并且根据声源的方向执行运动的各种设备。上述设备的示例包括基于所估计的声源方向执行自主移动的自主移动体。例如,专利文献1公开了一种用于朝向已经识别出的用户的话语或面部的方向移动机器人装置的技术。

[0003] 引用列表

[0004] 专利文献

[0005] 专利文献1:日本未经审查专利申请公开第2004-130427号

发明内容

[0006] 本发明要解决的问题

[0007] 然而,在专利文献1中描述的技术没有考虑除了用户的话语之外的声音(即,噪声)的存在。为此,在使机器人装置简单地接近用户的估计方向的情况下,存在噪声的输入水平增加的可能性,从而导致难以识别用户的话语。

[0008] 因此,本公开提出了一种新颖且改进的并且使得可以使自主移动体执行运动以进一步提高声音识别的精度信息处理装置、信息处理方法和程序。

[0009] 解决问题的手段

[0010] 根据本公开,提供一种信息处理装置,包括:操作控制单元,该操作控制单元控制基于识别处理动作的自主移动体的运动,在检测到作为用于语音识别处理的目标语音的目标声音的情况下,该操作控制单元将该自主移动体移动到基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。

[0011] 另外,根据本公开,提供一种信息处理方法,包括使处理器:控制基于识别处理动作的自主移动体的运动,控制还包括:在检测到作为用于语音识别处理的目标语音的目标声音的情况下,将自主移动体移动到基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。

[0012] 另外,根据本公开,提供一种程序,使计算机用作信息处理装置,该信息处理装置包括:操作控制单元,该操作控制单元控制基于识别处理动作的自主移动体的运动,在检测到作为用于语音识别处理的目标语音的目标声音的情况下,该操作控制单元将该自主移动体移动到基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。

[0013] 发明的效果

[0014] 如上所述,根据本公开,可以使自主移动体执行运动以进一步提高声音识别的精度。

[0015] 应注意,上述效果不必是限制性的;除了上述效果之外或者代替上述效果,可以实现在本说明书中描述的任何效果或者可以从本说明书中掌握的其他效果。

附图说明

- [0016] [图1]示出了根据本公开的实施方式自主移动体的硬件配置示例。
- [0017] [图2]是根据同一实施方式的包括在自主移动体中的致动器的配置示例。
- [0018] [图3]是根据同一实施方式的包括在自主移动体中的致动器的操作的说明图。
- [0019] [图4]是根据同一实施方式的包括在自主移动体中的致动器的操作的说明图。
- [0020] [图5]是根据同一实施方式的包括在自主移动体中的显示器的功能的说明图。
- [0021] [图6]示出了根据同一实施方式自主移动体的运动示例。
- [0022] [图7]示出了根据同一实施方式自主移动体的功能配置示例。
- [0023] [图8A]是根据同一实施方式自主移动体的运动概况的说明图。
- [0024] [图8B]是根据同一实施方式自主移动体的运动概况的说明图。
- [0025] [图9]是在未检测到根据同一实施方式的目标声音时自主移动体的运动控制的说明图。
- [0026] [图10]是在检测到根据同一实施方式的目标声音的情况下基于噪声图的运动控制的说明图。
- [0027] [图11]是在检测到根据同一实施方式的目标声音的情况下基于噪声图的运动控制的说明图。
- [0028] [图12]是在根据同一实施方式的接近目标不是发声用户的情况下的运动控制的说明图。
- [0029] [图13]是根据同一实施方式的基于声源方向估计的噪声图的创建的说明图。
- [0030] [图14]是根据同一实施方式的基于声源方向估计的噪声图的创建的说明图。
- [0031] [图15]示出了根据同一实施方式的包括噪声源的类型信息的噪声图的示例。
- [0032] [图16]是根据同一实施方式的用于噪声图创建处理和更新处理的执行条件的设置示例。
- [0033] [图17]是根据同一实施方式的噪声图集成处理的说明图。
- [0034] [图18]是根据同一实施方式的基于用户输入的噪声图的创建和更新的说明图。
- [0035] [图19]是根据同一实施方式的基于用户输入的噪声图的创建和更新的说明图。
- [0036] [图20]示出了根据同一实施方式的难以避开噪声区域的情况的示例。
- [0037] [图21]是根据同一实施方式的计算指示声音似然度的指标 α 和指示稳定度的指标 β 的说明图。
- [0038] [图22]是根据同一实施方式的计算指示声音似然度的指标 α 和指示稳定度的指标 β 的说明图。
- [0039] [图23]是示出根据同一实施方式的噪声图更新的流程的流程图。
- [0040] [图24]是示出根据同一实施方式的运动控制的流程的流程图。

具体实施方式

[0041] 在下文中,参考附图详细地给出本公开的优选实施方式的描述。应注意,在本说明

书和附图中,通过分配相同的参考数字省略对于实质上具有相同功能配置的组件的重复描述。

[0042] 应注意,按以下顺序给出描述。

[0043] 1.配置

[0044] 1.1.自主移动体10的概述

[0045] 1.2.自主移动体10的硬件配置示例

[0046] 1.3.自主移动体10的功能配置示例

[0047] 2.实施方式

[0048] 2.1.概述

[0049] 2.2.运动控制的细节

[0050] 2.3.噪声图的创建和更新

[0051] 2.4.基于噪声源回避优先级的运动控制

[0052] 2.5.运动流程

[0053] 3.结论

[0054] <1.配置>

[0055] <<1.1.自主移动体10的概述>>

[0056] 如上所述,近年来,已经开发了识别用户的话语等并且基于识别结果执行运动的各种设备。如上所述的设备的示例包括根据用户的话语、周围环境等改变行为的自主移动体。

[0057] 这里,为了实现高精度的语音识别,通常重要的是在通过麦克风获取的声音的信号中提高作为用于语音识别的目标语音的目标声音(例如,用户的发声语音)的功率与不是目标语音的非目标声音的功率的比率,即,SN比(信噪比)。具体地,在具有语音识别功能的自主移动体中,期望通过执行到将改进SN比的位置的移动来提高语音识别精度。

[0058] 然而,在专利文献1中描述的技术中,未考虑非目标声音,并且机器人装置仅朝向已经识别出用户的话语或面部的方向移动。为此,在专利文献1中描述的技术中,还假定以下情况:同时接近用户和发出非目标声音的噪声源导致SN比降低和语音识别精度降低。

[0059] 另外,在专利文献1中描述的技术中,通过将用户的话语或面部识别为接近用户的触发来控制机器人装置的运动。因此,假定专利文献1中描述的机器人装置很可能跟随始终存在于机器人装置周围的用户,这可能使用户感到烦恼。

[0060] 通过关注上述方面已经构思了根据本公开的实施方式的信息处理装置、信息处理方法和程序,并且使得可以使自主移动体执行运动以进一步提高声音识别的精度。

[0061] 这里,首先,给出根据本公开的实施方式的自主移动体10的概述的描述。根据本公开的实施方式的自主移动体10是信息处理装置,该信息处理装置基于所收集的传感器信息来估计情况,并且根据情况自主地选择和执行各种运动。自主移动体10的特征之一是自主地执行运动,该运动被假定为对于每种情况是优选的,与简单地执行符合由用户指示的命令的运动的机器人不同。

[0062] 例如,在未检测到作为用于语音识别处理的目标语音的目标声音(即,用户的话语)的情况下,根据本公开的实施方式的自主移动体10可以执行自主运动以避免不是上述目标语音的非目标声音的输入。在检测到用户的话语的情况下,该运动使得可以在不总是

跟随用户的情况下有效地增加相对于话语提高语音识别的精度可能性。

[0063] 另外,在检测到目标声音的情况下,根据本公开的实施方式自主移动体10可以移动到在基于目标声音确定的接近目标周围非目标声音的输入水平变低的位置。即,根据本公开的实施方式自主移动体10考虑非目标声音执行移动运动,这使得可以提高SN比并且有效地提高相对于用户的话语的语音识别的精度。

[0064] 以这种方式,根据本公开的实施方式自主移动体10类似于包括人的动物全面地判断其自身状态、周围环境等,从而确定并执行自主运动。在以上方面,根据本公开的实施方式的自主移动体10明显不同于基于指令执行对应运动或处理的无源设备。

[0065] 根据本公开的实施方式自主移动体10可以是在空间中执行自主姿势控制并且执行各种运动的自主移动型机器人。自主移动体10可以是例如具有模拟人或动物(诸如狗)的形状并且具有运动能力的自主移动型机器人。另外,自主移动体10可以是例如具有与用户通信能力的设备,诸如交通工具或无人机。根据本公开的实施方式的自主移动体10的形状、能力或期望水平等可以根据目的和作用来适当地设计。

[0066] <<1.2.自主移动体10的硬件配置示例>>

[0067] 接下来,给出根据本公开的实施方式的自主移动体10的硬件配置示例的描述。应注意,在下文中,以示例的方式给出自主移动体10是狗型四腿行走机器人的情况的描述。

[0068] 图1示出了根据本公开的实施方式的自主移动体10的硬件配置示例。如图1所示,自主移动体10是具有头部、躯干、四个腿部和尾部的狗型四腿行走机器人。另外,自主移动体10在头部上设置有两个显示器510。

[0069] 另外,自主移动体10包括各种传感器。自主移动体10包括例如麦克风515、相机520、ToF(飞行时间)传感器525、人体检测传感器530、测距传感器535、触摸传感器540、照度传感器545、脚底按钮550以及惯性传感器555。

[0070] (麦克风515)

[0071] 麦克风515具有收集周围声音的功能。上述声音的示例包括用户的话语和周围环境声音。自主移动体10可以包括例如头部上的四个麦克风。设置多个麦克风515使得可以以高灵敏度收集在周围生成的声音并且实现声音定位。

[0072] (相机520)

[0073] 相机520具有捕获用户或周围环境的图像的功能。自主移动体10可以包括例如在鼻尖和腰部处的两个广角相机。在这种情况下,布置在鼻尖处的广角相机捕获与自主移动体10的前方视野(即,狗的视野)相对应的图像,并且在腰部处的广角相机捕获以上侧为中心的周围区域的图像。自主移动体10能够例如基于由布置在腰部处的广角相机捕获的图像来提取天花板的特征点等,以实现SLAM(同时定位和映射)。

[0074] (ToF传感器525)

[0075] ToF传感器525具有检测相对于存在于头部前面的物体的距离的功能。ToF传感器525设置在头部的尖端处。根据ToF传感器525,可以以高精度检测相对于各种物体的距离,从而使得可以实现与相对于包括用户、障碍物等的目标物体的相对位置相对应的运动。

[0076] (人体检测传感器530)

[0077] 人体检测传感器530具有检测用户或由用户饲养的宠物的位置的功能。人体检测传感器530例如布置在胸部。根据人体检测传感器530,检测存在于前面的动物体使得可以

实现动物体的各种运动,例如,与诸如兴趣、恐惧和惊讶的情绪相对应的运动。

[0078] (测距传感器535)

[0079] 测距传感器535具有获取自主移动体10的前面的地板表面的情况的功能。测距传感器535例如布置在胸部。根据测距传感器535,可以以高精度检测相对于存在于自主移动体10的前面的地板表面上的物体的距离,从而使得可以实现与相对于物体的相对位置相对应的运动。

[0080] (触摸传感器540)

[0081] 触摸传感器540具有检测用户的接触的功能。触摸传感器540例如布置在用户很可能触摸自主移动体10的位置,诸如头部的顶部、下颚或背部。触摸传感器540可以包括例如电容或压敏触摸传感器。根据触摸传感器540,可以检测诸如用户的触摸、击打、敲击或推动的接触动作,从而使得可以执行与接触动作相对应的运动。

[0082] (照度传感器545)

[0083] 照度传感器545检测自主移动体10所位于的空间中的照度。照度传感器545可以例如布置在头部的后表面上的尾部的根部处。根据照度传感器545,可以检测周围亮度并且执行与亮度相对应的运动。

[0084] (脚底按钮550)

[0085] 脚底按钮550具有检测自主移动体10的腿部的底表面是否与地板接触的功能。为此,脚底按钮550布置在与四个腿部的衬垫相对应的位置中的每个位置处。根据脚底按钮550,可以检测自主移动体10与地板表面之间的接触或非接触,从而使得自主移动体10可以抓住例如已经被用户抬起等。

[0086] (惯性传感器555)

[0087] 惯性传感器555是检测物理量(诸如头部和躯干的速度、加速度和旋转)的六轴传感器。即,惯性传感器555检测X轴、Y轴和Z轴的加速度和角速度。惯性传感器555布置在头部和躯干中的每一个处。根据惯性传感器555,可以以高精度检测自主移动体10的头部和躯干的移动,从而使得可以根据情况实现运动控制。

[0088] 以上已经给出了根据本公开的实施方式的包括在自主移动体10中的传感器的示例的描述。应注意,以上参考图1描述的配置仅是示例,并且可以包括在自主移动体10中的传感器的配置不限于这种示例。除了上述配置之外,自主移动体10还可以包括例如温度传感器、地磁传感器和包括GNSS(全球导航卫星系统)信号接收器的各种通信装置。可以根据规范和操作灵活地修改包括在自主移动体10中的传感器的配置。

[0089] 随后,给出了根据本公开的实施方式的自主移动体10的关节的配置示例的描述。图2是根据本公开的实施方式的包括在自主移动体10中的致动器570的配置示例。除了图2所示的旋转点之外,根据本公开的实施方式的自主移动体10具有总共22个旋转自由度,其中,两个旋转自由度用于耳朵和尾部中的每一个,并且一个旋转自由度用于嘴巴。

[0090] 例如,自主移动体10在头部中具有三个自由度,从而使得可以实现点头运动和颈部倾斜运动两者。另外,自主移动体10使用设置在腰部处的致动器570来再现腰部的摆动运动,从而使得可以实现更接近真实狗的自然且灵活的运动。

[0091] 应注意,根据本公开的实施方式的自主移动体10可以例如组合单轴致动器和双轴致动器,从而实现上述22个旋转自由度。例如,可以在腿部的肘部和膝部处采用单轴致动

器,并且可以在肩部和腿部根部处采用双轴致动器。

[0092] 图3和图4分别是根据本公开的实施方式的包括在自主移动体10中的致动器570的运动的说明图。参考图3,致动器570使用电机575使输出齿轮旋转,从而以任何旋转位置和旋转速度驱动可移动臂590。

[0093] 参考图4,根据本公开的实施方式的致动器570包括后盖571、齿轮箱盖572、控制基板573、齿轮箱底座574、电机575、第一齿轮576、第二齿轮577、输出齿轮578、检测磁体579以及两个轴承580。

[0094] 根据本公开的实施方式的致动器570可以包括例如磁性svGMR(自旋阀巨磁阻)。控制基板573基于由主处理器进行的控制使电机575旋转,从而将动力经由第一齿轮576和第二齿轮577传输到输出齿轮578,从而使得可以驱动可移动臂590。

[0095] 另外,包括在控制基板573中的位置传感器检测与输出齿轮578同步旋转的检测磁体579的旋转角度,从而以高精度检测旋转角度,即可移动臂590的旋转位置。

[0096] 应注意,磁性svGMR由于其非接触类型而具有优越的耐用性,并且当在GMR饱和区域中使用较少受到由于检测磁体579和位置传感器的距离变化引起的信号变化的影响的优点。

[0097] 以上已经给出了根据本公开的实施方式的包括在自主移动体10中的致动器570的配置示例的描述。根据上述配置,可以以高精度控制包括在自主移动体10中的关节的弯曲和伸展运动,并且可以精确地检测关节的旋转位置。

[0098] 随后,参考图5给出根据本公开的实施方式的包括在自主移动体10中的显示器510的功能的描述。图5是根据本公开的实施方式的包括在自主移动体10中的显示器510的功能的说明图。

[0099] (显示器510)

[0100] 显示器510具有在视觉上表达自主移动体10的眼睛移动和情绪的功能。如图5所示,显示器510能够响应于情绪和运动来表达眼球、瞳孔和眼睑的运动。显示器510不旨在显示与眼球的移动无关的字母、符号、图像等,从而产生接近诸如真实活狗的动物的自然运动。

[0101] 另外,如图5所示,自主移动体10包括分别与右眼和左眼相对应的两个显示器510r和510l。例如,通过两个独立的OLED(有机发光二极管)来实现显示器510r和510l。根据OLED,与一对眼球由一个平面显示器表达的情况或两个眼球分别由两个独立的平面显示器表达的情况相比,可以再现眼球的弯曲表面,从而使得可以实现更自然的外观。

[0102] 如上所述,根据显示器510r和510l,可以以高精度和灵活性表达如图5所示的自主移动体10的视线和情绪。另外,用户可以从显示器510上显示的眼球的运动直观地掌握自主移动体10的状态。

[0103] 以上已经给出了根据本公开的实施方式的自主移动体10的硬件配置示例的描述。根据上述配置,以高精度和灵活性控制自主移动体10的关节和眼球的运动使得可以实现更接近真实生物的运动和情绪表达,如图6所示。应注意,图6示出了根据本公开的实施方式的自主移动体10的运动示例;然而,图6以简化的方式示出了自主移动体10的外部结构以集中于自主移动体10的关节和眼球的运动给出描述。类似地,在以下描述中,在一些情况下,可以以简化的方式示出自主移动体10的外部结构;然而,根据本公开的实施方式的自主移动

体10的硬件配置和外观不限于附图所示的示例,并且可以适当地设计。

[0104] <<1.3.自主移动体10的功能配置示例>>

[0105] 接下来,给出根据本公开的实施方式的自主移动体10的功能配置示例的描述。图7示出了根据本公开的实施方式的自主移动体10的功能配置示例。参考图7,根据本公开的实施方式的自主移动体10包括输入单元110、识别单元120、周围环境估计单元130、周围环境保持单元140、操作控制单元150、驱动单元160以及输出单元170。

[0106] (输入单元110)

[0107] 输入单元110具有收集关于用户和周围环境的各种类型的信息的功能。输入单元110收集例如用户的话语、在周围生成的环境声音、关于用户和周围环境的图像信息以及各种类型的传感器信息。为此,输入单元110包括图1所示的各种传感器。

[0108] (识别单元120)

[0109] 识别单元120具有基于由输入单元110收集的各种类型的信息来执行用户、周围的物体以及自主移动体10的状态的各种识别的功能。例如,识别单元120可以执行人体识别、面部识别、面部表情和视线的识别、语音识别、物体识别、颜色识别、形状识别、标记识别、障碍物识别、阶梯识别、亮度识别等。

[0110] (周围环境估计单元130)

[0111] 周围环境估计单元130具有基于由输入单元110收集的传感器信息和识别单元120的识别结果来创建和更新指示生成非目标声音的情况的噪声图的功能。稍后单独描述周围环境估计单元130的功能的细节。

[0112] (周围环境保持单元140)

[0113] 周围环境保持单元140具有保持由周围环境估计单元130创建和更新的噪声图的功能。

[0114] (操作控制单元150)

[0115] 操作控制单元150具有基于识别单元120的识别结果和由周围环境保持单元140保持的噪声图来制定动作计划,并且基于该动作计划来控制驱动单元160和输出单元170的操作的功能。例如,操作控制单元150基于上述动作计划执行致动器570的旋转控制、显示器510的显示控制、扬声器的语音输出控制等。单独描述根据本公开的实施方式的操作控制单元150的功能的细节。

[0116] (驱动单元160)

[0117] 驱动单元160具有基于操作控制单元150的控制来弯曲和伸展包括在自主移动体10中的多个关节的功能。更具体地,驱动单元160基于操作控制单元150的控制来驱动包括在相应关节中的致动器570。

[0118] (输出单元170)

[0119] 输出单元170具有基于操作控制单元150的控制输出视觉信息和声音信息的功能。为此,输出单元170包括显示器510和扬声器。

[0120] 以上已经给出了根据本公开的实施方式的自主移动体10的功能配置的描述。应注意,图7所示的配置仅是示例,并且自主移动体10的功能配置不限于这种示例。根据本公开的实施方式的自主移动体10可以包括例如与信息处理服务器和另一自主移动体执行通信的通信单元。另外,识别单元120、周围环境估计单元130、操作控制单元150等的功能可以被

实现为上述信息处理服务器的功能。在这种情况下,信息处理服务器能够基于由自主移动体10的输入单元110收集的传感器信息来执行各种类型的识别处理、噪声图的创建或更新以及动作计划,并且控制自主移动体10的驱动单元160和输出单元170。根据本公开的实施方案的自主移动体10的功能配置可以根据规范和操作灵活地修改。

[0121] <2.实施方式>

[0122] <<2.1.概述>>

[0123] 接下来,给出本公开的实施方案的描述。如上所述,为了提高相对于目标声音的语音识别的精度,根据本公开的实施方案的自主移动体10执行自主运动以提高目标声音和非目标声音的SN比。

[0124] 这里,用于提高SN比的假定技术包括对输入信号执行信号处理(多麦克风信号处理和单麦克风信号处理)的技术以及使用定向麦克风等的技术。然而,可以说SN比受相对于目标声源或非目标声源(在下文中也称为“噪声源”)的物理距离的影响最大。

[0125] 为此,根据本实施方案的自主移动体10不是简单地接近目标声源,而是在接近该目标声源的同时尽可能远离非目标声源,从而使得可以有效地提高SN比。

[0126] 图8A和图8B是根据本实施方案的自主移动体10的运动概况的说明图。图8A示出了根据本实施方案的通过比较技术进行的移动控制的示例。在图8A所示的示例中,在检测到用户U的语音(即,目标声音)的情况下,比较器90通过最短路径接近用户U以增加目标声音的输入水平(声压水平)。然而,此时,比较器90不考虑发出非目标声音的噪声源NS的存在;因此,比较器90在接近用户U的同时接近噪声源NS。因此,在比较技术中,非目标声音的输入水平也与目标声音的输入水平一起增加,并且因此,出现了提高SN比的效果降低并且相对于用户U的话语的语音识别的精度降低的可能性。

[0127] 相对照地,图8B示出了根据本实施方案的通过信息处理方法进行的移动控制的示例。如图8B所示,在检测到用户U的语音(即,目标声音)的情况下,根据本实施方案的自主移动体10考虑发出非目标声音的噪声源的存在来执行移动。具体地,在检测到目标声音的情况下,根据本实施方案的操作控制单元150可以将自主移动体10移动到在基于目标声音确定的接近目标周围非目标声音的输入水平变低的位置。

[0128] 这里,上述接近目标可以是发出目标声音(即,发声语音)的用户U。即,在检测到用户U的话语的情况下,根据本实施方案的操作控制单元150能够将自主移动体10移动到在作为接近目标的用户U周围用户的话语的输入水平变高并且由噪声源NS发出的非目标声音的输入水平变低的位置。

[0129] 在图8B所示的示例中,根据本实施方案的操作控制单元150相对于以用户U为中心将自主移动体10移动到与发出非目标声音(即,噪声源NS)的位置相反的一侧。以这种方式,基于根据本实施方案的操作控制单元150,可以将自主移动体10移动更远离噪声源NS并且更接近作为接近目标的用户U。另外,在自主移动体10移动到与噪声源NS相反的一侧(其中,用户U介于自主移动体10与噪声源NS之间)的情况下,用户U充当墙壁,并且期望更有效地降低由噪声源NS发出的非目标声音的输入水平的效果。

[0130] 基于根据本实施方案的操作控制单元150的上述功能,在增加目标声音的输入水平的同时可以降低非目标声音的输入水平或降低输入水平的增加率,因此使得可以有效地提高SN比。因此,基于根据本实施方案的操作控制单元150,可以仅使用自主移动体10中固

有的移动功能来极大地提高SN比,而无需对输入信号执行信号处理和使用定向麦克风进行波束成形。

[0131] <<2.2.运动控制的细节>>

[0132] 接下来,更详细地给出根据本实施方式的由操作控制单元150进行的自主移动体10的运动控制的描述。如上所述,根据本实施方式的操作控制单元150使自主移动体10在除了目标声音之外还考虑非目标声音的存在的情况下执行运动,这使得可以提高SN比并且有效地提高相对于目标声音的语音识别的精度。

[0133] 这里,可以将根据本实施方式的目标声音限定为用于由识别单元120进行语音识别的目标语音。上述目标语音可以包括所有语音。例如,可以说,在自主移动体10覆盖从电视机、收音机等输出的所有语音以及作为用于语音识别的目标的用户或第三方的发声语音的情况下,如上所述的所有语音都是目标声音。在这种情况下,识别单元120能够例如通过考虑到人语音的谐波声音结构等的音调等的比较将如上所述的所有语音检测为目标声音。应注意,在如上所述的所有语音都是目标声音的情况下,自主移动体10能够响应于从电视机输出的语音来执行一些动作等。

[0134] 一方面,根据本实施方式的目标声音可以仅覆盖如上所述的语音中预先注册的预定用户的语音。在这种情况下,识别单元120基于预先注册的用户语音特征或存在于输入信号到来的方向上的人的面部识别来执行说话者识别,从而使得可以仅检测预定用户的发声语音作为目标声音。

[0135] 另一方面,根据本实施方式的目标声音可以仅覆盖由预定用户的发声语音中的特定关键字和与运动指令有关的特定词。在这种情况下,识别单元120基于输入信号执行语音识别,从而使得可以仅检测包括特定关键字和词的预定用户的发声语音作为目标声音。

[0136] 另外,可以将根据本实施方式的非目标声音定义为除了目标声音之外的所有声音。根据本实施方式的非目标声音的示例包括厨房中的工作声音和由诸如排风扇、冰箱和汽车的设备生成的各种非语音。

[0137] 以上已经给出了根据本实施方式的目标声音和非目标声音的细节的描述。随后,参考图9,给出在未检测到目标声音的情况下的运动控制的描述。图9是在未检测到根据本实施方式的目标声音时自主移动体10的运动控制的说明图。图9示出了由周围环境保持单元140保持的噪声图的示例。

[0138] 这里,根据本实施方式的噪声图包括由周围环境估计单元130创建和更新并且指示生成非目标声音的情况的图。根据本实施方式的噪声图包括例如存在于存在自主移动体10的空间中的噪声源以及关于噪声区域的信息,该噪声区域是由噪声源发出的非目标声音的输入水平强(例如,等于或大于阈值)的区域。在图9所示的示例中,噪声图包括噪声源NS1和NS2以及分别与噪声源NS1和NS2相对应的噪声区域NR1和NR2的信息。

[0139] 根据本实施方式的操作控制单元150的特征之一是基于包括如上所述的信息的噪声图来控制自主移动体10的运动。例如,在未检测到目标声音的情况下,根据本实施方式的操作控制单元150可以基于噪声图来控制自主移动体10的运动以避免非目标声音的输入。

[0140] 更具体地,在未检测到目标声音的情况下,根据本实施方式的操作控制单元150能够基于噪声图将自主移动体10的移动范围限制在非目标声音的输入水平等于或小于阈值的区域内。例如,在图9所示的示例中,在未检测到目标声音的情况下,操作控制单元150可

以将自主移动体10的移动范围限制在除了噪声区域NR1和NR2两者之外的区域内以防止自主移动体10进入噪声区域NR1和NR2。此时,操作控制单元150可以在上述区域内随机地移动自主移动体10,或者可以将自主移动体10移动到期望非目标声音的声压最小的位置 $P_{\min}$ 等。

[0141] 根据由根据本实施方式的操作控制单元150进行的上述控制,即使在未检测到目标声音的情况下,操作自主移动体10以尽可能地抑制非目标声音的输入,使得在用户发出呼叫等的情况下,即在检测到目标声音的情况下,可以有效地提高相对于目标声音的语音识别的精度。

[0142] 随后,给出在检测到根据本实施方式的目标声音的情况下基于噪声图的运动控制的细节的描述。如上所述,根据本实施方式的操作控制单元150的特征之一是在检测到目标声音的情况下,将自主移动体10移动到接近目标周围目标声音的输入水平变高并且非目标声音的输入水平变低的位置。此时,根据本实施方式的操作控制单元150参考噪声图,从而使得可以以高精度实现上述运动控制。

[0143] 图10是在检测到根据本实施方式的目标声音的情况下基于噪声图的运动控制的说明图。在图10所示的示例中,操作控制单元150基于呼叫自主移动体10的名称的用户U的发声语音U01(即,目标声音)的检测,使自主移动体10接近用户U。此时,根据本实施方式的操作控制单元150参考噪声图,并且考虑包括在噪声图中的噪声源NS和噪声区域NR来控制自主移动体10的移动。

[0144] 例如,在图10所示的情况下,如附图中的双点划线所示,在自主移动体10通过最短路径接近用户U的情况下,自主移动体10在噪声区域NR中移动。然而,根据本实施方式的操作控制单元150参考噪声图,从而使得可以在不使自主移动体10进入噪声区域NR或停止在噪声区域NR中的情况下将自主移动体10移动到远离噪声源NS的位置。更具体地,操作控制单元150可以使自主移动体10围绕如附图中的实线所示的噪声区域NR,并且可以相对于以作为接近目标的用户U为中心将自主移动体10移动到与噪声源NS相反的一侧。

[0145] 以这种方式,根据本实施方式的操作控制单元150参考噪声图,从而使得可以准确地掌握噪声源和噪声区域,并且将自主移动体10移动到目标声音的输入水平增大并且非目标声音的输入水平降低的位置。根据由根据本实施方式的操作控制单元150进行的上述运动控制,可以提高SN比,并且有效地提高相对于目标声音的语音识别的精度。

[0146] 应注意,根据本实施方式的操作控制单元150可以不必相对于以接近目标为中心将自主移动体10移动到与噪声源相反的一侧。例如,在图11所示的示例中,墙壁存在于连接噪声源NS和用户U的直线中。在以这种方式连接噪声源和接近目标的直线中存在障碍物的情况下,操作控制单元150可以将自主移动体10停止在接近上述相反侧的位置处,并且不进入噪声区域NR。即使在这种情况下,也可以实现目标声音的输入水平增大和非目标声音的输入水平降低两者,并且实现提高SN比的效果。应注意,操作控制单元150可以基于在包括在噪声图中的墙壁、家具等的信息来掌握障碍物的存在,并且可以基于由识别单元120识别的障碍物的存在来执行如上所述的运动控制。

[0147] 接下来,给出在根据本实施方式的接近目标不是发声用户的情况下的运动控制的描述。图12是在根据本实施方式的接近目标不是发声用户的情况下的运动控制的说明图。

[0148] 在图12所示的示例中,与图10和图11所示的示例不同,用户U1正在发出语音话语U02,该语音话语U02不是向用户U1而是向用户U2提供移动指令。此时,根据本实施方式的操作

作控制单元150基于由识别单元120识别的发声语音U02将用户U2设置为接近目标,并且控制自主移动体10的运动以使自主移动体10接近用户U2。

[0149] 以这种方式,根据本实施方式的接近目标不仅可以是发出语音话语的发声用户,而且可以是诸如基于发声语音通过语音识别处理指定的另一用户的移动体、诸如充电站的固定物体或任意位置。

[0150] 即使在接近目标不是发声用户的情况下,根据本实施方式的操作控制单元150也类似地参考噪声图,从而使得可以使自主移动体10考虑噪声源NS和噪声区域NR执行移动。在图12所示的示例中,操作控制单元150相对于以用户U2为中心将自主移动体10移动到与噪声源NS相反的一侧。根据由根据本实施方式的操作控制单元150进行的上述运动控制,可以有效地提高相对于预期此后进行的用户U2的发声语音(即,目标声音)的语音识别的精度。

[0151] 以上已经给出了根据本实施方式的基于噪声图的运动控制的描述。如上所述,根据本实施方式的操作控制单元150参考由周围环境保持单元140保持的噪声图,从而不仅考虑目标声音的输入水平而且考虑非目标声音的输入水平来实现自主移动体10的移动,并且提高SN比,这使得可以实现高精度的语音识别。

[0152] 应注意,以上已经通过示例的方式主要描述了以下情况:根据本实施方式的操作控制单元150基于目标声音的检测控制自主移动体10以将自主移动体10移动到接近目标;然而,本实施方式中的移动的触发不限于这样的示例。根据本实施方式的操作控制单元150可以基于用户的面部的识别或与用户的移动指令有关的手势的识别来执行控制以将自主移动体10移动到接近目标。即使在这种情况下,参考噪声图并且将自主移动体移动到非目标声音的输入水平变低的位置,使得可以提高相对于预期此后进行的目标声音的语音识别的精度。

[0153] <<2.3.噪声图的创建和更新>>

[0154] 接下来,给出根据本实施方式的噪声图的创建和更新的细节的描述。根据本实施方式的周围环境估计单元130能够基于例如声源方向估计和声压测量的结果创建如上所述的噪声图。

[0155] 首先,给出根据本实施方式的基于声源方向估计的噪声图的创建的描述。图13是根据本实施方式的基于声源方向估计的噪声图的创建的说明图。

[0156] 在基于声源方向估计的噪声图的创建中,周围环境估计单元130首先在任何给定点处执行声音定位以估计声源方向。在图13所示的示例中,周围环境估计单元130估计点P1处的噪声源NS1和NS2的相应方向。应注意,在该时间点,可以估计噪声源NS1和NS2的方向,但是从自主移动体10到噪声源NS1和NS2的距离是未知的。

[0157] 随后,周围环境估计单元130移动到不同于先前估计的声源方向的点,并且再次执行声音定位以估计声源方向。在图13所示的示例中,周围环境估计单元130在点P2处再次估计噪声源NS1和NS2的相应方向。此时,周围环境估计单元130能够根据自主移动体10的移动距离和在点P1和P2处估计的方向的交点来估计空间中的噪声源NS1和NS2的位置。

[0158] 在下文中,周围环境估计单元130在另一点处重复声源方向估计,这使得可以提高声源位置的估计精度。在图13所示的示例中,周围环境估计单元130在点P3处再次估计噪声源NS1和NS2的方向。

[0159] 根据本实施方式的周围环境估计单元130以这种方式在多个点处重复声源方向估计,这使得可以以高精度估计空间中的噪声源NS1和NS2的位置,并且创建例如其中位于距相应估计位置预定距离处的区域被设置为噪声区域NR1和NR2的噪声图。

[0160] 随后,给出根据本实施方式的基于声压测量的噪声图的创建的描述。在自主移动体10不包括比同时生成的声源的数量更大数量的麦克风的情况下,难以基于上述声源方向估计来创建噪声图。相对照地,即使在自主移动体10仅包括单个麦克风的情况下,根据本实施方式的周围环境估计单元130也能够基于以下将描述的声压测量来创建噪声图。图14是根据本实施方式的基于声源方向估计的噪声图的创建的说明图。

[0161] 在基于声源方向估计的噪声图的创建中,周围环境估计单元130首先在任何给定点处执行声压水平的测量。在图14所示的示例中,周围环境估计单元130测量点P4处的声压水平。随后,周围环境估计单元130在不同于已经执行测量的点的另一点处重复执行声压水平的测量。在图14所示的示例中,周围环境估计单元130在点P5和P6处执行声压水平的测量。

[0162] 根据本实施方式的周围环境估计单元130以这种方式在多个点处重复声压测量,这使得可以估计如图14所示的声压水平的等压线,并且将声压水平等于或大于阈值的区域设置为噪声区域。另外,周围环境估计单元130能够估计噪声区域中具有最高声压水平的点作为噪声源的位置。在图14所示的示例中,周围环境估计单元130基于所估计的等压线设置噪声区域NR1和NR2,并且估计噪声源NS1和NS2的位置。

[0163] 基于根据本实施方式的周围环境估计单元130,即使在自主移动体10仅包括单个麦克风的情况下,以这种方式在多个点处重复执行声压测量使得可以以高精度创建噪声图。应注意,在执行基于声压测量的噪声图的创建的情况下,有必要分离目标声音和非目标声音,但是这种分离可通过识别单元120的上述功能来实现。

[0164] 另外,根据本实施方式的噪声图可以包括诸如噪声源的类型的信息。图15示出了根据本实施方式的包括噪声源的类型信息的噪声图的示例。如从图15所示的示例可以看出,噪声图的信息包括噪声源NS1和NS2分别是厨房和电视机。

[0165] 根据本实施方式的周围环境估计单元130能够例如基于识别单元120的物体识别的结果创建包括如图15所示的噪声源的类型信息的噪声图。基于根据本实施方式的周围环境估计单元130的上述功能,操作控制单元150能够根据噪声源的识别以高精度执行运动控制。

[0166] 以上描述了根据本实施方式的噪声图创建的一个示例。随后,给出根据本实施方式的噪声图的创建或更新的定时的描述。

[0167] 例如,假定不断动态地执行噪声图的更新的情况。在这种情况下,尽管可以无误地检测在周围生成的非目标声音,但是诸如对操作控制单元150的运动控制无用的零星声音的信息全部被包括作为噪声图的信息。另外,在不断动态地执行噪声图的更新的情况下,计算量变得巨大;因此,高性能的处理器等是必要的。

[0168] 因此,根据本实施方式的周围环境估计单元130可以仅在收集非目标声音的高效条件下执行噪声图创建处理和更新处理。这里,上述高效条件包括可以生成大量非目标声音的情况。另外,可以生成大量非目标声音的情况包括用户在空间中执行活动的情况。因此,根据本实施方式的周围环境估计单元130可以在用户存在于安装有自主移动体10的空

间中的定时执行噪声图创建处理和更新处理。

[0169] 此时,根据本实施方式的周围环境估计单元130能够基于用户的日程安排和各种类型的传感器信息来估计用户的不存在或存在,并且仅在用户很可能存在的条件下执行噪声图创建处理和更新处理。

[0170] 图16是根据本实施方式的噪声图创建处理和更新处理的执行条件的设置示例。在图16所示的示例中,针对诸如用户的日程安排(不在家或在家)、检测或未检测到钥匙声音、检测或未检测到开门/关门噪声、检测或未检测到诸如“我在家”的话语、由人体检测传感器检测或未检测到移动体的因素的每个组合来设置是否执行创建处理和更新处理。应注意,图16示出了仅在从所有上述因素确定用户存在于与自主移动体10相同的空间中的情况下执行处理的示例。

[0171] 应注意,可以根据自主移动体10的特征和情况动态地设置如上所述的这种执行适当性。因此,基于根据本实施方式的周围环境估计单元130,基于在用户存在于周围环境中的时区中收集的非目标声音创建或更新噪声图使得可以保持高精度的噪声图。

[0172] 接下来,给出根据本实施方式的噪声图更新处理的描述。如上所述,根据本实施方式的周围环境估计单元130能够基于在用户存在于周围环境中的时区中收集的非目标声音来动态地更新噪声图。

[0173] 然而,在这种情况下,在执行声音收集的定时,还假定生成不同的非目标声音的情况。因此,在仅基于最新的声音收集数据更新噪声图的情况下,噪声图中包括本质上影响较小的非目标声音(诸如零星声音)的信息,这可能使操作控制单元150的运动控制的精度降低。

[0174] 因此,可以不通过基于最新的声音收集数据覆写现有的噪声图,而是通过将最新的声音收集数据集成到现有的噪声图中来更新噪声图。

[0175] 图17是根据本实施方式的噪声图集成处理的说明图。在图17所示的示例中,根据本实施方式的周围环境估计单元130基于三次声音收集来执行噪声图集成。这里,假设以下情况:在第一声音收集中检测到与厨房和电视机有关的非目标声音;在第二声音收集中检测到与窗户和电视机有关的非目标声音;并且在第三声音收集中检测到与厨房和电视机有关的非目标声音。

[0176] 在这种情况下,例如,根据本实施方式的周围环境估计单元130可以通过平均等来集成三次声音收集数据以更新噪声图。基于根据本实施方式的周围环境估计单元130,如图17所示,可以将每个非目标声音的生成频率等反映在噪声图上,这使得可以降低不频繁生成的非目标声音(诸如零星声音)的影响。应注意,在图17所示的示例中,非目标声音的生成频率由阴影线的高密度和低密度指示(密度越高,生成频率变得越高)。在这种情况下,操作控制单元150可以控制自主移动体10的移动以更集中地避开频繁生成非目标声音的电视机。

[0177] 以上已经给出了根据本实施方式的噪声图的创建和更新的描述。应注意,上述技术仅是示例,并且根据本实施方式的噪声图的创建和更新不限于该示例。

[0178] 例如,根据本实施方式的周围环境估计单元130可以基于用户输入的信息来执行噪声图的创建和更新。图18和图19是根据本实施方式的基于用户输入的噪声图的创建和更新的说明图。

[0179] 例如,如图18所示,根据本实施方式的周围环境估计单元130可以基于由用户经由信息处理终端20等输入的家具布局信息来执行噪声图的创建和更新。此时,周围环境估计单元130能够请求将与每件家具相对应的图标IC布置在包括在信息处理终端20中的显示部分中的模拟用户的房间的输入区域IA中,并且基于所输入的信息执行噪声图的创建和更新。

[0180] 另外,例如,根据本实施方式的周围环境估计单元130能够基于手势(诸如,用户U要执行的手指指向)和用于教导噪声源的发声语音U03识别噪声源NS并且将所识别的噪声源NS反映在噪声图上。

[0181] <<2.4. 基于噪声源回避优先级的运动控制>>

[0182] 接下来,给出根据本实施方式的基于噪声源回避优先级的运动控制的描述。以上已经给出了根据本实施方式的操作控制单元150参考噪声图并且控制自主移动体10的移动以避免噪声区域的情况的描述。

[0183] 然而,假定根据情况,在一些情况下难以在避开噪声区域的同时执行移动的情况。图20示出了根据本实施方式的难以避开噪声区域的情况的示例。图20的左侧示出了基于声源方向估计的噪声图,而图20的右侧示出了基于声压测量的噪声图。这里,如从两个噪声图都可以看出,自主移动体10在两个噪声图中都被噪声区域NR1至NR4包围,并且难以移动到另一位置。

[0184] 在这种情况下,根据本实施方式的操作控制单元150可以基于分配给噪声源NS1至NS4的回避优先级控制自主移动体10以将自主移动体10移动到与较低回避优先级的噪声源相对应的噪声区域。

[0185] 这里,根据本实施方式的回避优先级可以例如通过由噪声源生成的非目标声音的类型和特征来确定。如上所述,根据本实施方式的非目标声音包括除了目标声音之外的各种类型的声音。同时,非目标声音对相对于目标声音的语音识别的精度施加的影响根据非目标声音的特征而不同。

[0186] 因此,根据本实施方式的周围环境估计单元130可以基于对语音识别的精度的影响度的大小来分类非目标声音,并且创建其中按照影响度的降序设置回避优先级的噪声图。

[0187] 这里,描述了将非目标分类为四个类别1至类别4的示例。例如,类别1可以包括具有相对大的音量并且在作为人语音时不是目标声音的非目标声音。类别1的示例包括从电视机、收音机和任何其他设备输出的语音、包括人声的音乐、除了用户之外的第三方之间的对话等。类别1可以包括四个类别中对语音识别的精度具有最高影响并且具有最高回避优先级的非目标声音。

[0188] 另外,类别2可以包括不稳定地生成并且具有相对大的音量的非目标声音,从而使得难以充分地实现抑制噪声的效果。类别2的示例包括工作声音(诸如洗碗和烹饪)、来自打开的窗户的室外声音等。类别2可以包括四个类别中对语音识别的精度具有第二高影响并且具有第二高回避优先级的非目标声音。

[0189] 另外,类别3可以包括稳定生成的非目标声音,从而使得相对容易地实现抑制噪声的效果。类别3的示例包括由空调、排气扇、电脑风扇等生成的声音。类别2可以包括四个类别中对语音识别的精度具有第三高影响并且具有第三高回避优先级的非目标声音。

[0190] 另外,类别4可以包括偶尔地生成并且仅具有瞬时影响的非目标声音。类别4的示例包括门打开/关闭噪声、脚步声、由微波炉生成的声音等。类别4可以包括四个类别中对语音识别的精度具有最低影响并且具有最低回避优先级的非目标声音。

[0191] 因此,根据本实施方式的周围环境估计单元130能够创建其中根据非目标声音的特征设置回避优先级的噪声图。

[0192] 另外,根据本实施方式的噪声源回避优先级可以基于与非目标声音有关的声学 and 某种量化指标来设置。上述量化指标的示例包括指示声音似然度的指标和指示稳定度的指标。

[0193] 通常,作为用于语音识别的目标的目标声音(即用户的发声语音)是“非稳定”“语音”。同时,“非稳定”“语音”的示例还包括非目标声音,诸如第三方之间的对话以及从电视机和收音机输出的语音。因此,为了提高相对于作为“非稳定”“语音”的目标声音的语音识别的精度,重要的是避开作为难以与目标声音分离的“非稳定”“语音”的非目标声音。

[0194] 同时,“非稳定”“非语音”包括洗碗和烹饪等的工作声音,而“稳定”“非语音”包括从空调、排气扇和电脑风扇等输出的声音。然而,这样的非目标声音相对容易地与目标声音分离;因此,与上述作为“非稳定”“语音”的非目标声音相比,可以说非目标声音具有较低的回避优先级。

[0195] 另外,“稳定”“语音”例如与长时间且连续地发出相同声音(诸如“啊(Ahhhh)”)的情况相对应;然而,这种声音在日常生活中极不可能生成,并且因此可以被忽略。

[0196] 如上所述,根据本实施方式的周围环境估计单元130可以基于指示声音似然度的指标 α 和指示稳定度的指标 β 来计算对相对于目标声音的语音识别施加的非目标声音的影响度,并且基于由此计算的值来设置回避优先级。

[0197] 图21和图22是根据本实施方式的指示声音似然度的指标 α 和指示稳定度的指标 β 的计算的说明图。图21示出了在根据本实施方式的自主移动体10包括多个麦克风的情况下的计算流程。

[0198] 在自主移动体10包括多个麦克风的情况下,如图21所示,周围环境估计单元130朝向噪声源NS1至NS4的方向依次执行波束成形,并且计算指标 α 和指标 β 。根据这种技术,可以在不进入噪声区域NR1至NR4的情况下计算噪声源NR1至NR4中的每一个的指标 α 和指标 β ;因此,即使在计算期间检测到目标声音的情况下,也可以维持相对于目标声音的语音识别的精度。

[0199] 另外,图22示出了在根据本实施方式的自主移动体10包括单个麦克风的情况下的计算流程。在自主移动体10包括单个麦克风的情况下,如图22所示,周围环境估计单元130可以计算噪声源NS1至NS4中的每一个周围的指标 α 和指标 β 。应注意,在完成每个噪声源的指标 α 和指标 β 的计算的情况下,操作控制单元150可以控制自主移动体10以使自主移动体10立即从噪声区域NR1至NR4脱离。

[0200] 以上已经给出了根据本实施方式的指标 α 和指标 β 的计算流程的描述。根据本实施方式的周围环境估计单元130能够基于如上所述计算的指标 α 和指标 β 来计算噪声源中的每一个的影响度,并且基于影响度来设置回避优先级。例如,指标 α 和指标 β 的总值可以被定义为影响度,并且周围环境估计单元130可以按照总值的降序设置回避优先级。

[0201] 应注意,根据本实施方式的周围环境估计单元130可以例如基于声谱熵来计算指

示声音似然度的指标 α 。声谱熵是用于VAD(语音活动检测)技术的指标,并且与其他声音相比,人语音倾向于具有较低的值。

[0202] 根据本实施方式的周围环境估计单元130能够通过以下数学表达式(1)来计算声谱熵,即指标 α 。应注意,数学表达式(1)中的 f 指示频率,并且 S_f 指示观测信号的频率 f 的振幅谱。另外,数学表达式(1)中的 P_f 由以下数学表达式(2)定义。

[0203] [数学式1]

$$[0204] \quad \alpha = -\sum_f P_f \cdot \log P_f \cdot \cdot \cdot (1)$$

$$[0205] \quad P_f = \frac{S_f}{\sum_f S_f} \cdot \cdot \cdot (2)$$

[0206] 另外,根据本实施方式的周围环境估计单元130可以例如基于声音峰度来计算指示稳定度的指标 β 。声音峰度是经常用于在声音的稳定性与非稳定性之间进行区分的指标,并且可以通过以下数学表达式(3)来计算。应注意,以下数学表达式(3)中的 T 指示计算峰度的声音片段的长度,并且可以设置诸如3秒至5秒的长度。另外,数学表达式(3)中的 t 指示特定时间,并且 $x(t)$ 指示在时间 t 处的语音波形。

[0207] [数学式2]

$$[0208] \quad \beta = \frac{\sum_{t=1}^T x(t)^4}{(\sum_{t=1}^T x(t)^2)^2} - 3 \cdot \cdot \cdot (3)$$

[0209] 以上给出了根据本实施方式的噪声源回避优先级的设置的描述。基于根据本实施方式的回避优先级的设置,自主移动体10能够在优先的基础上避开对相对于目标声音的语音识别施加影响的非目标声音。

[0210] <<2.5.运动流程>>

[0211] 接下来,给出根据本实施方式的自主移动体10的运动流程的细节的描述。首先,给出根据本实施方式的噪声图的更新流程的描述。图23是示出根据本实施方式的噪声图更新的流程的流程图。

[0212] 参考图23,首先,周围环境估计单元130基于由输入单元110收集的传感器信息和识别单元120的识别结果来估计周围环境(S1101)。具体地,周围环境估计单元130估计噪声源和噪声区域。

[0213] 接下来,周围环境估计单元130确定周围环境保持单元140是否保持现有的噪声图(S1102)。

[0214] 这里,在不存在由周围环境保持单元140保持的噪声图的情况下(S1102:否),周围环境估计单元130基于在步骤S1101中估计的周围环境来创建噪声图,并且将该噪声图存储在周围环境保持单元140中(S1107)。

[0215] 同时,在周围环境保持单元140中存在现有的噪声图的情况下(S1102:是),周围环境估计单元130接下来确定所估计的周围环境中的噪声源的数量是否从现有的噪声图中的噪声源的数量改变(S1103)。

[0216] 这里,在噪声源的数量改变的情况下(S1103:是),周围环境估计单元130基于在步骤S1101中估计的周围环境来集成噪声图(S1106),并且将集成后的噪声图存储在周围环境保持单元140中(S1107)。

[0217] 同时,在噪声源的数量未改变的情况下(S1103:否),周围环境估计单元130接下来

确定所估计的周围环境中的噪声源的位置是否从现有的噪声图中的噪声源的位置改变(S1104)。

[0218] 这里,在噪声源的位置改变的情况下(S1104:是),周围环境估计单元130基于在步骤S1101中估计的周围环境来集成噪声图(S1106),并且将集成后的噪声图存储在周围环境保持单元140中(S1107)。

[0219] 同时,在噪声源的位置未改变的情况下(S1104:否),周围环境估计单元130接下来确定所估计的周围环境中由噪声源发出的非目标声音的声压是否从现有的噪声图中的非目标声音的声压改变(S1105)。

[0220] 这里,在由噪声源发出的非目标声音的声压改变的情况下(S1105:是),周围环境估计单元130基于在步骤S1101中估计的周围环境来集成噪声图(S1106),并且将集成后的噪声图存储在周围环境保持单元140中(S1107)。

[0221] 同时,在由噪声源发出的非目标声音的声压未改变的情况下(S1105:否),周围环境估计单元130不更新噪声图,并且将现有的噪声图维持在周围环境保持单元140中。

[0222] 接下来,给出根据本实施方式的运动控制的流程的细节的描述。图24是示出根据本实施方式的运动控制的流程的流程图。

[0223] 参考图24,操作控制单元150首先读取由周围环境保持单元140保持的噪声图(S1201)。

[0224] 随后,操作控制单元150基于在步骤S1201中读取的噪声图,使自主移动体10执行避开噪声区域的自主动作(S1202)。

[0225] 另外,操作控制单元150连续地确定在步骤S1202的自主动作期间是否检测到目标声音(S1203)。

[0226] 这里,在检测到目标声音的情况下(S1203:是),操作控制单元150基于在步骤S1201中读取的噪声图,将自主移动体10移动到接近目标周围非目标声音的输入水平变低的位置(S1204)。

[0227] 接下来,操作控制单元150基于目标声音的语音识别结果使自主移动体10执行对应的动作(S1205)。

[0228] 以上已经给出了根据本实施方式的自主移动体10的运动流程的描述。应注意,以上主要描述了为了提高SN比,根据本实施方式的自主移动体10考虑目标声音和非目标声音的输入水平执行移动。然而,根据本实施方式的用于提高SN比的技术不限于这样的示例,并且可以例如与信号处理或波束成形技术结合使用。

[0229] 例如,根据本实施方式的操作控制单元150可以将自主移动体10移动到接近目标与噪声源之间的位置,并且执行控制以朝着接近目标的方向扩展波束成形。在自主移动体10是狗型机器人装置的情况下,操作控制单元150可以执行控制以在与作为接近目标的用户的的面部的高度相对应的仰角处扩展波束成形。在这种情况下,期望有效地消除从水平方向到达的非目标声音以有效地提高SN比的效果。

[0230] 另外,为了避开噪声区域,例如,根据本实施方式的操作控制单元150可以使自主移动体10执行用于引导用户的运动。例如,在作为接近目标的用户存在于噪声区域中的情况下,操作控制单元150使自主移动体10执行用于引导用户移动远离噪声区域并接近自主移动体10的运动,这使得可以在不进入噪声区域的情况下增加目标声音的输入水平。例如,

上述引导可以通过诸如吠叫、在噪声区域前面停止和徘徊的运动来实现。另外,例如,在自主移动体10具有类似人型机器人装置的言语通信功能的情况下,自主移动体10可以通过语音明确地提供想要远离噪声区域的通知。

[0231] <3.结论>

[0232] 如上所述,作为根据本公开的实施方式的信息处理装置的一个示例的自主移动体10包括操作控制单元150,该操作控制单元150基于识别处理来控制自主移动体10的运动。另外,根据本公开的实施方式的操作控制单元150的特征之一是在检测到作为用于语音识别处理的目标语音的目标声音的情况下,将自主移动体10移动到在基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。根据这种配置,可以使自主移动体执行运动以进一步提高声音识别的精度。

[0233] 尽管以上已经参考附图详细地给出了本公开的优选实施方式的描述,但是本公开的技术范围并不限于这样的示例。显而易见的是,本公开的领域的普通技术人员可以在权利要求中描述的技术思想的范围内找到各种改变或修改,并且应当理解,这些改变和修改自然地落在本公开的技术范围内。

[0234] 另外,本文所描述的效果仅是说明性或示例性的,而不是限制性的。即,除了以上效果之外或者代替以上效果,根据本公开的技术可以实现从本说明书的描述中对本领域技术人员显而易见的其他效果。

[0235] 另外,本说明书中的自主移动体10的一系列处理的相应步骤不必按照流程图所示的时间顺序处理。例如,自主移动体10的一系列处理的相应步骤可以按照与流程图所示的顺序不同的顺序处理,或者可以并行处理。

[0236] 应注意,本公开的技术范围还包括以下配置。

[0237] (1)

[0238] 一种信息处理装置,包括:

[0239] 操作控制单元,控制基于识别处理动作的自主移动体的运动,

[0240] 在检测到作为用于语音识别处理的目标语音的目标声音的情况下,操作控制单元将自主移动体移动到基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。

[0241] (2)

[0242] 根据(1)的信息处理装置,其中,在检测到目标声音的情况下,操作控制单元将自主移动体移动到基于目标声音确定的接近目标周围目标声音的输入水平变高并且非目标声音的输入水平变低的位置。

[0243] (3)

[0244] 根据(1)或(2)的信息处理装置,其中,在检测到目标声音的情况下,操作控制单元将自主移动体移动到远离发出非目标声音的噪声源并且靠近接近目标的位置。

[0245] (4)

[0246] 根据(1)至(3)中任一项的信息处理装置,其中,在检测到目标声音的情况下,操作控制单元相对于以接近目标为中心将自主移动体移动到与发出非目标声音的噪声源相反的一侧。

[0247] (5)

- [0248] 根据(1)至(4)中任一项的信息处理装置,其中,
- [0249] 目标声音包括用户的发声语音,并且
- [0250] 接近目标包括发出发声语音的发声用户。
- [0251] (6)
- [0252] 根据(1)至(5)中任一项的信息处理装置,其中,接近目标包括由基于用户的发声语音的语音识别处理识别的移动体、固定物体或位置。
- [0253] (7)
- [0254] 根据(1)至(6)中任一项的信息处理装置,其中,操作控制单元基于指示在周围环境中生成非目标声音的情况的噪声图来控制自主移动体的运动。
- [0255] (8)
- [0256] 根据(7)的信息处理装置,其中,
- [0257] 噪声图包括发出非目标声音的噪声源的信息,并且
- [0258] 操作控制单元基于噪声源的回避优先级来控制自主移动体的运动。
- [0259] (9)
- [0260] 根据(8)的信息处理装置,其中,基于噪声源的类型来确定噪声源的回避优先级。
- [0261] (10)
- [0262] 根据(8)的信息处理装置,其中,基于由噪声源发出的非目标声音对语音识别处理的影响度来确定噪声源的回避优先级。
- [0263] (11)
- [0264] 根据(10)的信息处理装置,其中,基于指示非目标声音的声音似然度的指标和指示稳定度的指标中的至少一个来计算影响度。
- [0265] (12)
- [0266] 根据(7)至(10)中任一项的信息处理装置,其中,在未检测到目标声音的情况下,操作控制单元基于噪声图来控制自主移动体的运动以避免非目标声音的输入。
- [0267] (13)
- [0268] 根据(7)至(12)中任一项的信息处理装置,其中,在未检测到目标声音的情况下,操作控制单元基于噪声图将自主移动体的移动范围限制在非目标声音的输入水平等于或小于阈值的区域内。
- [0269] (14)
- [0270] 根据(7)至(13)中任一项的信息处理装置,还包括创建噪声图的周围环境估计单元。
- [0271] (15)
- [0272] 根据(14)的信息处理装置,其中,周围环境估计单元基于发出非目标声音的噪声源的方向估计或声压测量来创建噪声图。
- [0273] (16)
- [0274] 根据(14)或(15)的信息处理装置,其中,周围环境估计单元基于所收集的非目标声音动态地更新噪声图。
- [0275] (17)
- [0276] 根据(16)的信息处理装置,其中,周围环境估计单元基于发出非目标声音的噪声

源的数量、位置或声压的变化来动态地更新噪声图。

[0277] (18)

[0278] 根据(16)或(17)的信息处理装置,其中,周围环境估计单元基于用户存在于周围环境中的时区中收集的非目标声音来创建或更新噪声图。

[0279] (19)

[0280] 一种信息处理方法,包括使处理器:

[0281] 控制基于识别处理动作的自主移动体的运动,

[0282] 控制还包括:在检测到作为用于语音识别处理的目标语音的目标声音的情况下,将自主移动体移动到基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。

[0283] (20)

[0284] 一种程序,使计算机用作信息处理装置,

[0285] 该信息处理装置包括:

[0286] 操作控制单元,控制基于识别处理动作的自主移动体的运动,

[0287] 在检测到作为用于语音识别处理的目标语音的目标声音的情况下,操作控制单元将自主移动体移动到基于目标声音确定的接近目标周围不是目标语音的非目标声音的输入水平变低的位置。

[0288] 参考数字列表

[0289] 10 自主移动体

[0290] 110 输入单元

[0291] 120 识别单元

[0292] 130 周围环境估计单元

[0293] 140 周围环境保持单元

[0294] 150 操作控制单元

[0295] 160 驱动单元

[0296] 170 输出单元。

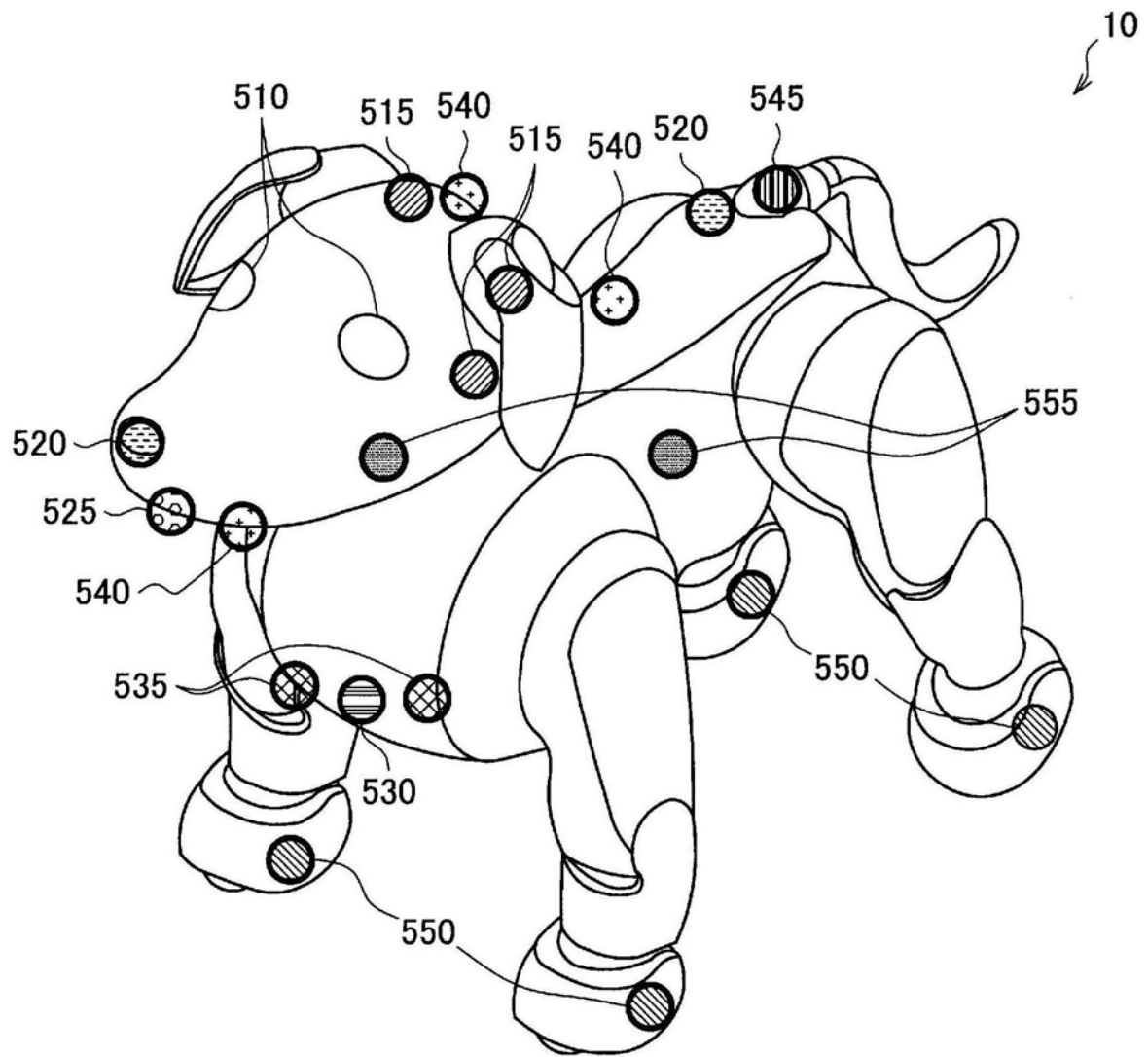


图1

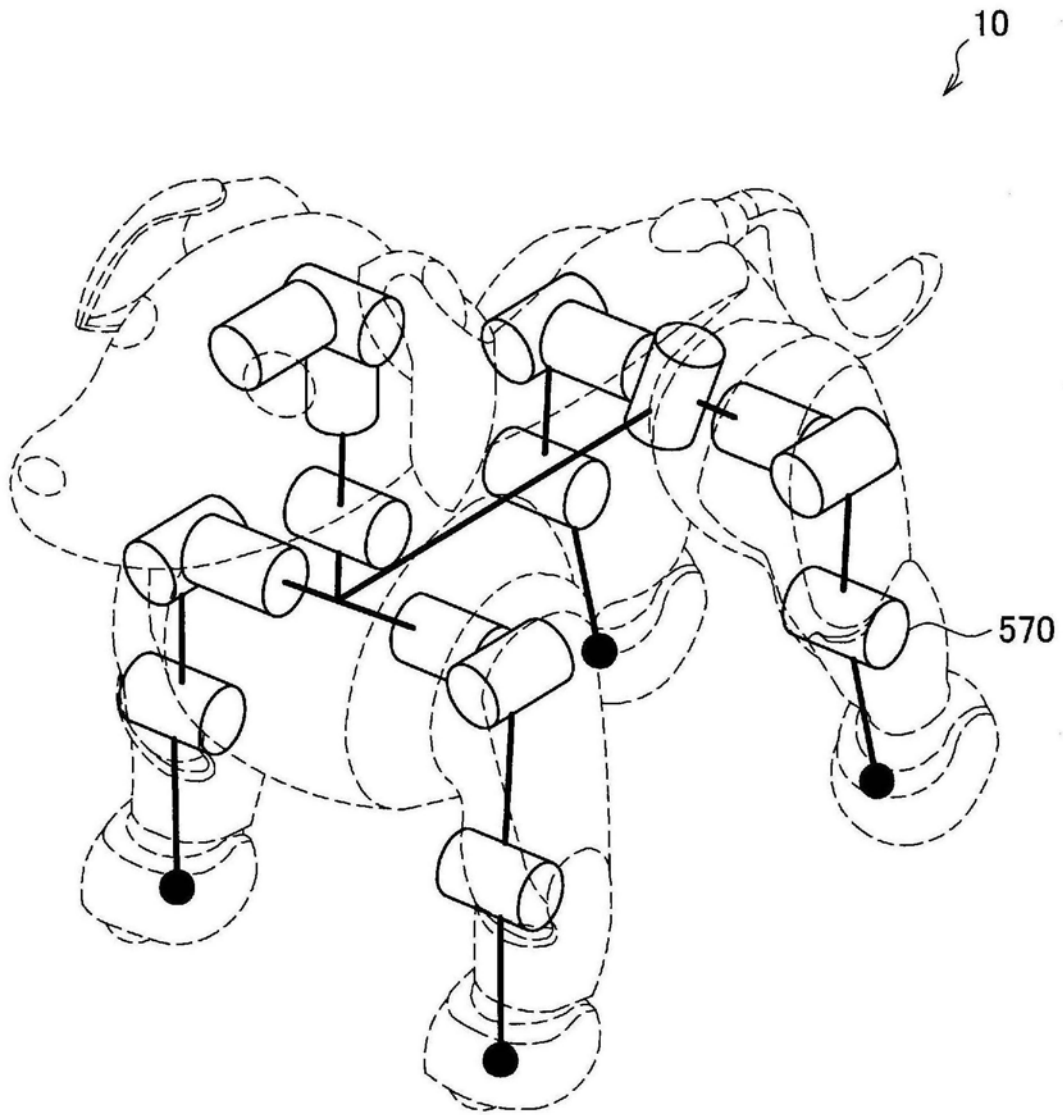


图2

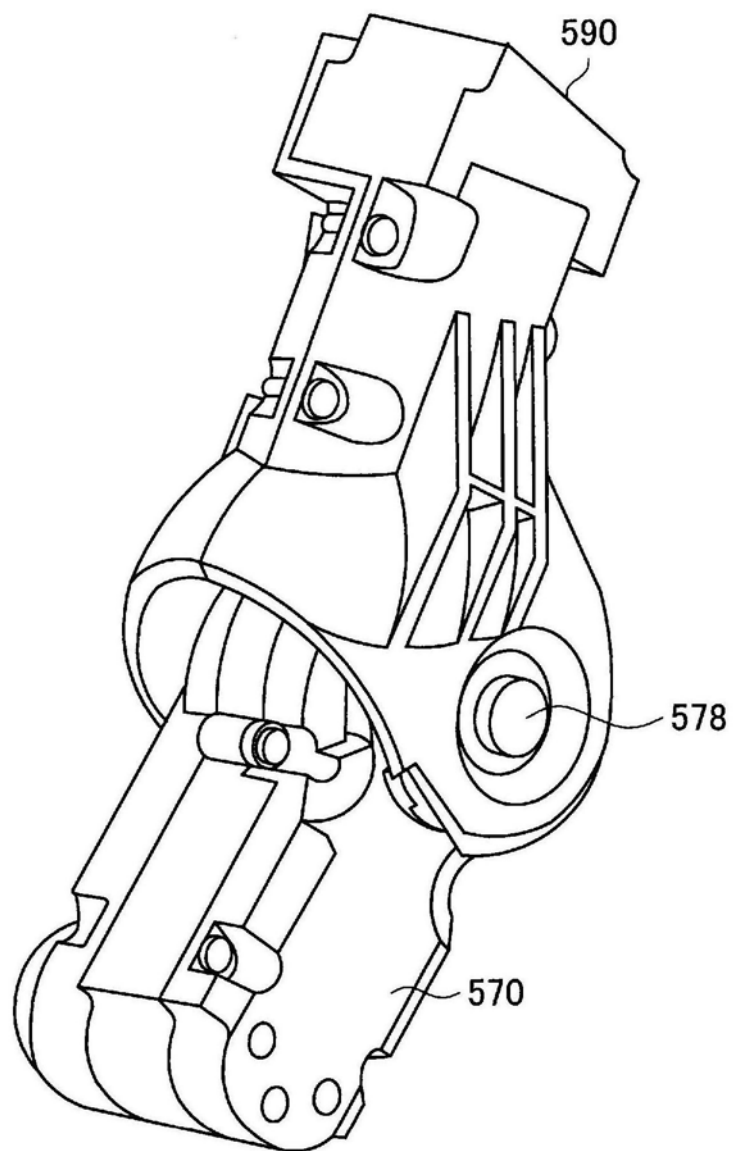


图3

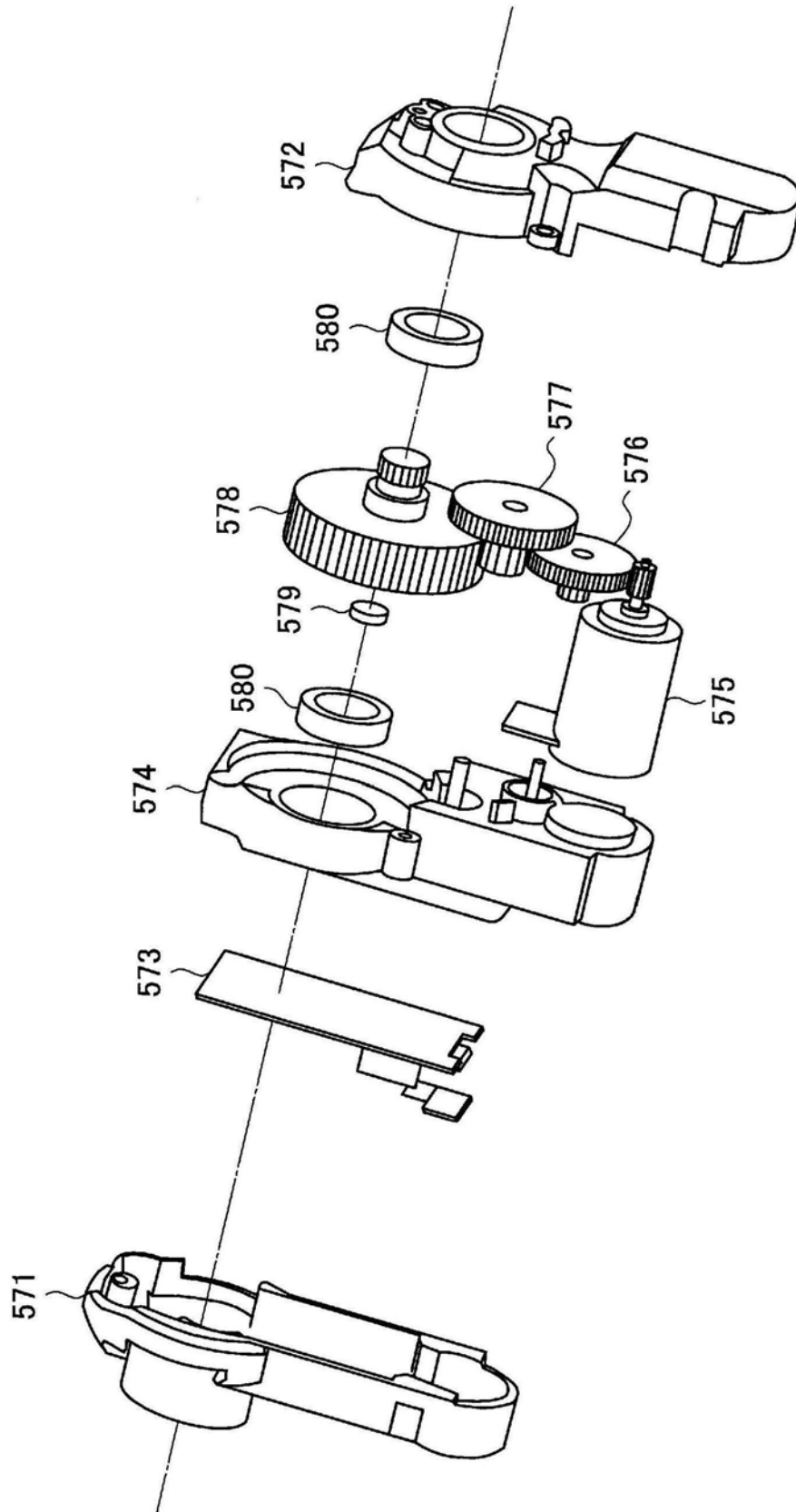


图4

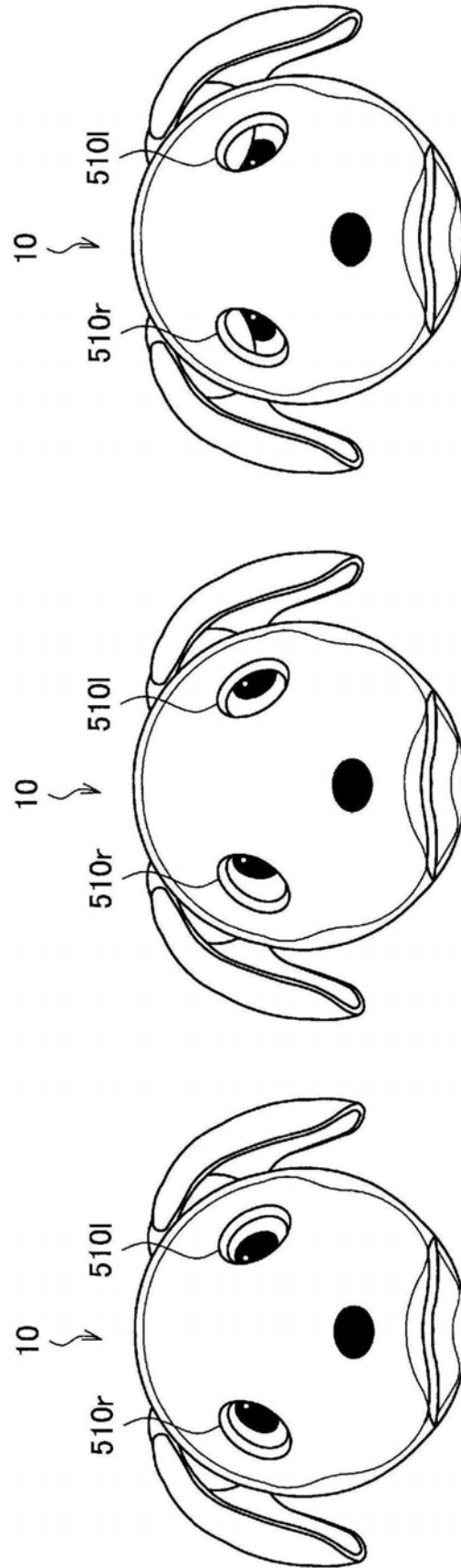


图5

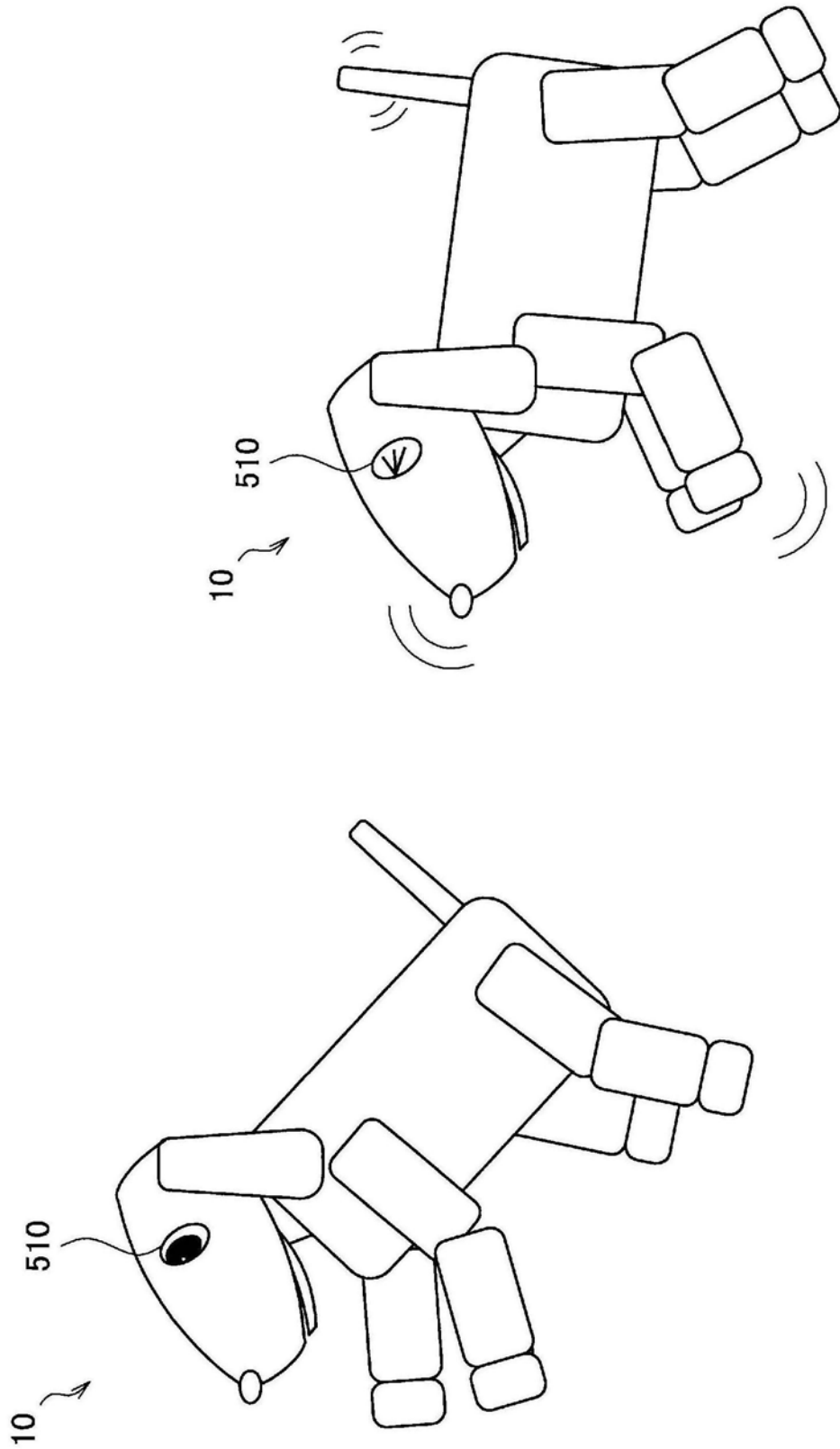


图6

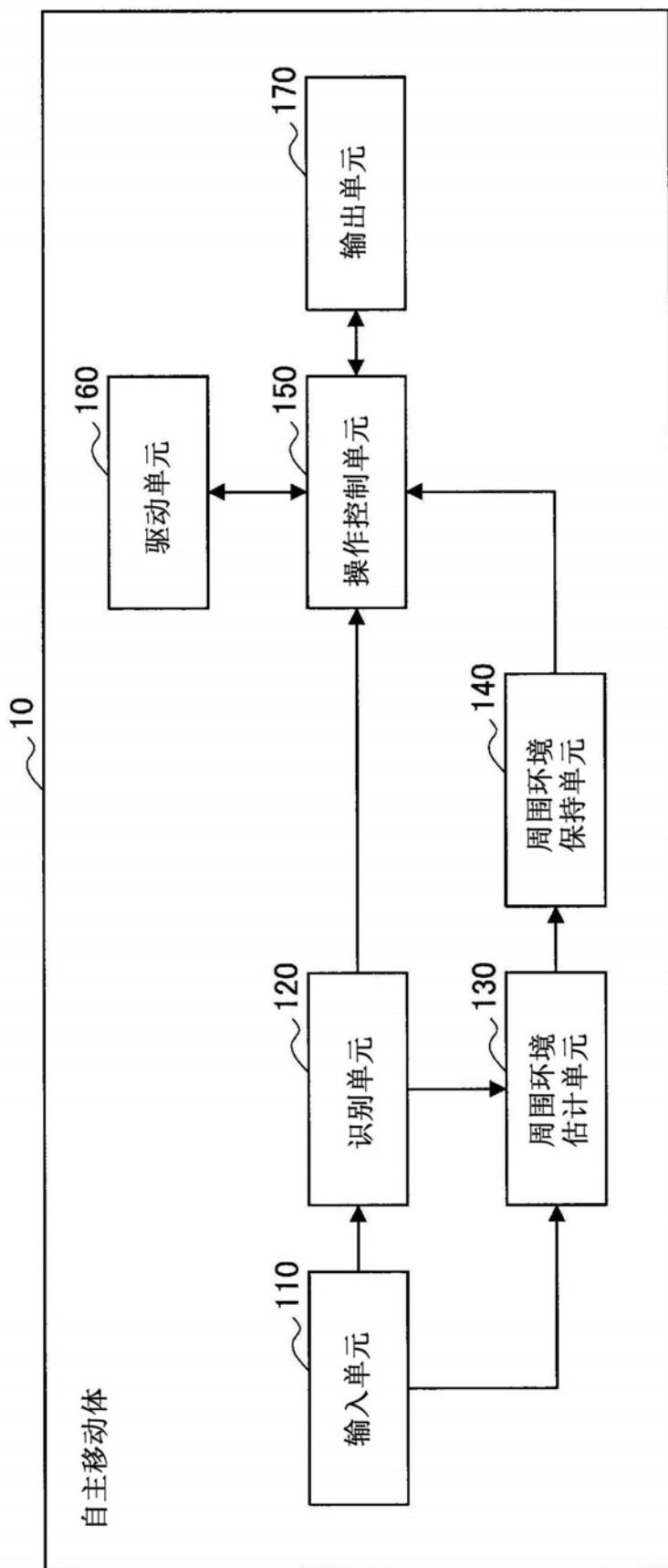


图7

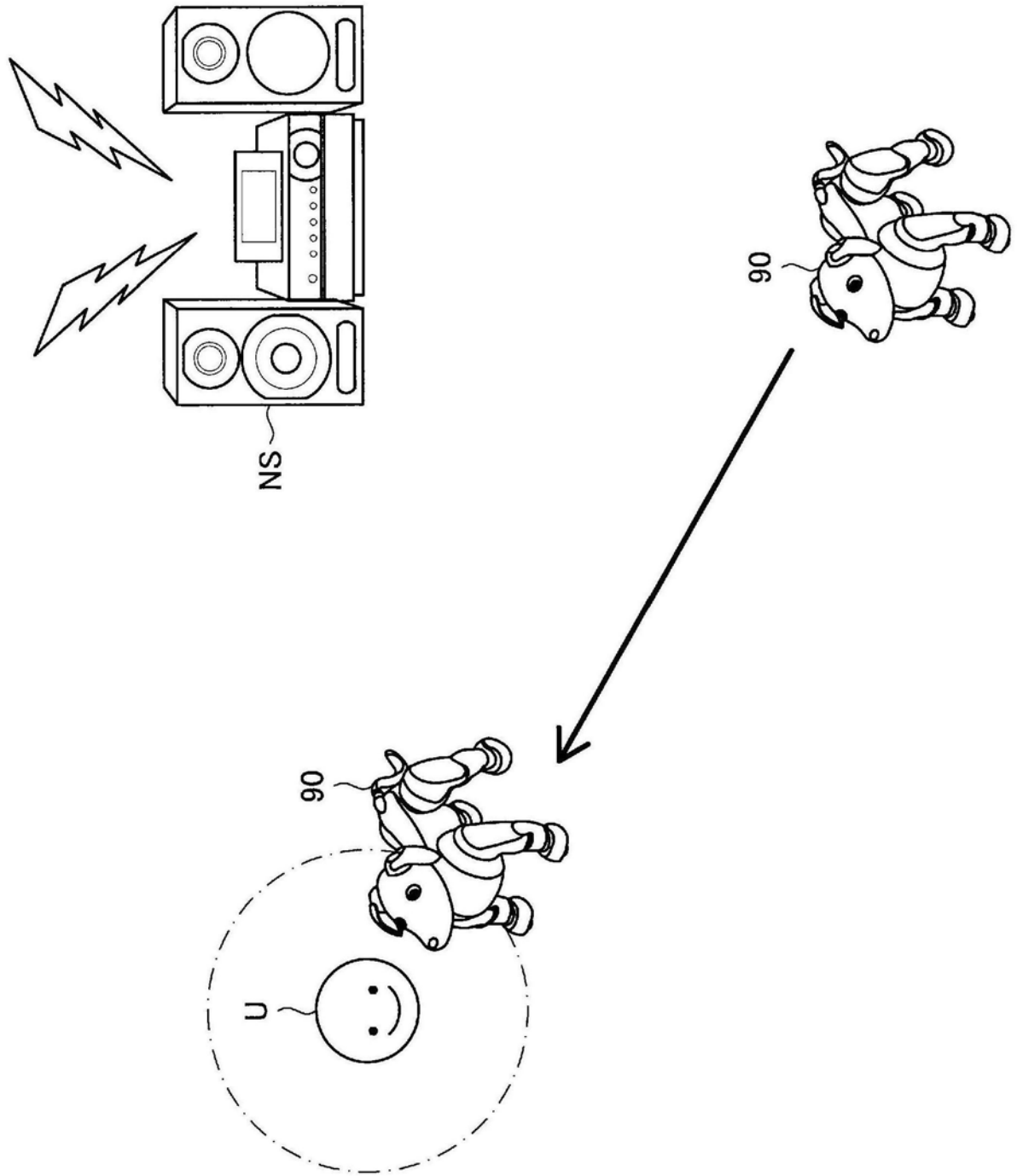


图8A

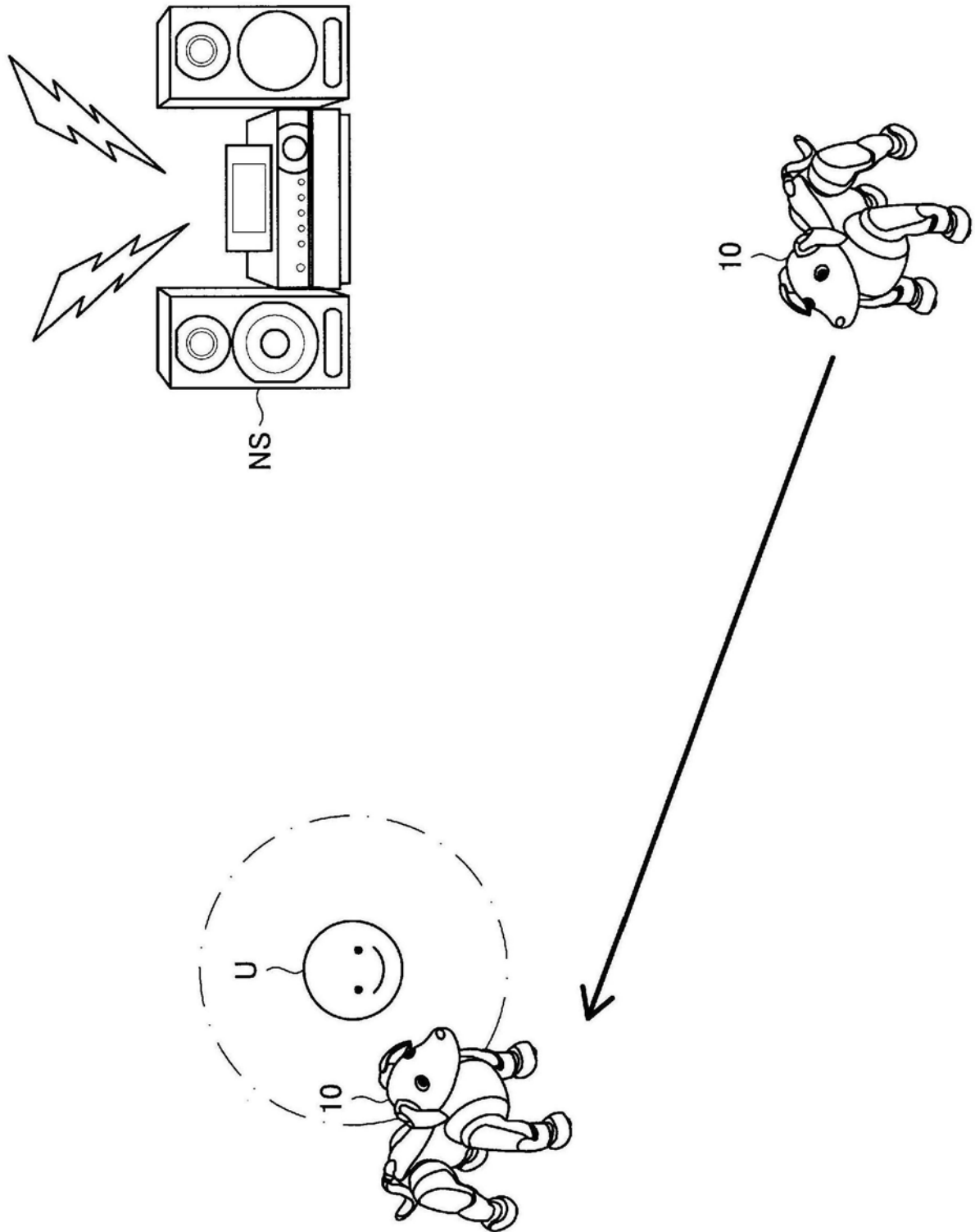


图8B

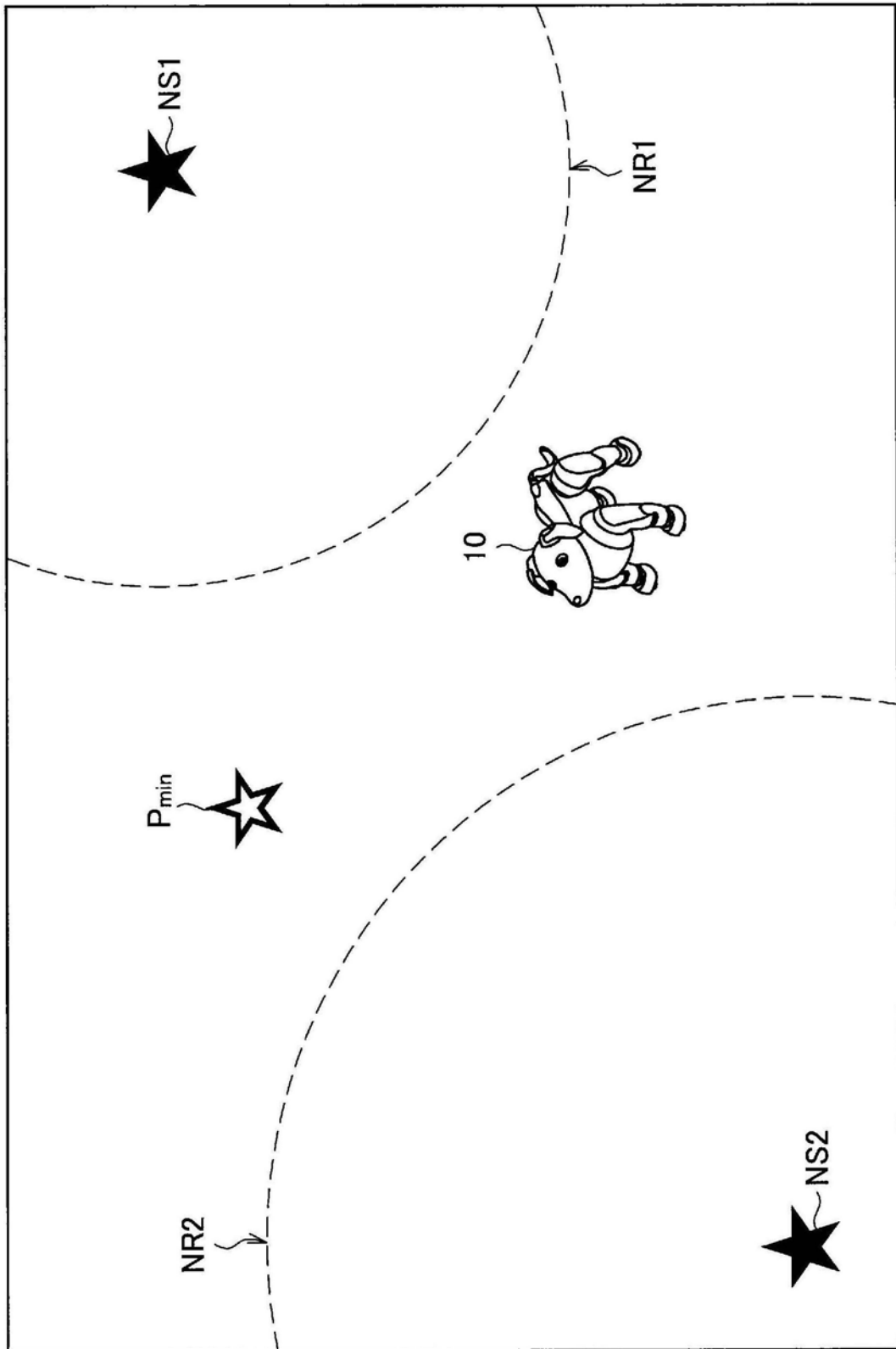


图9

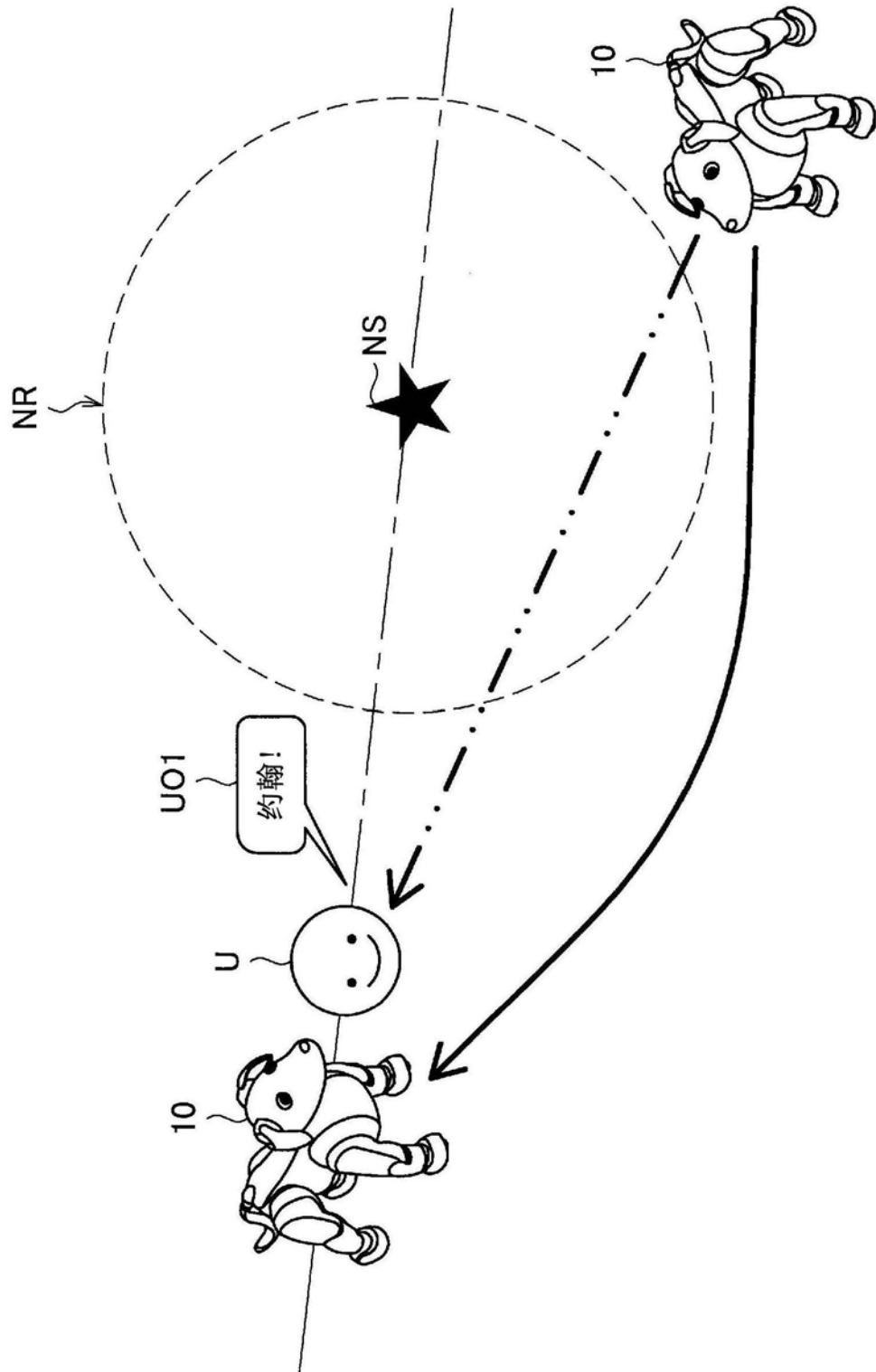


图10

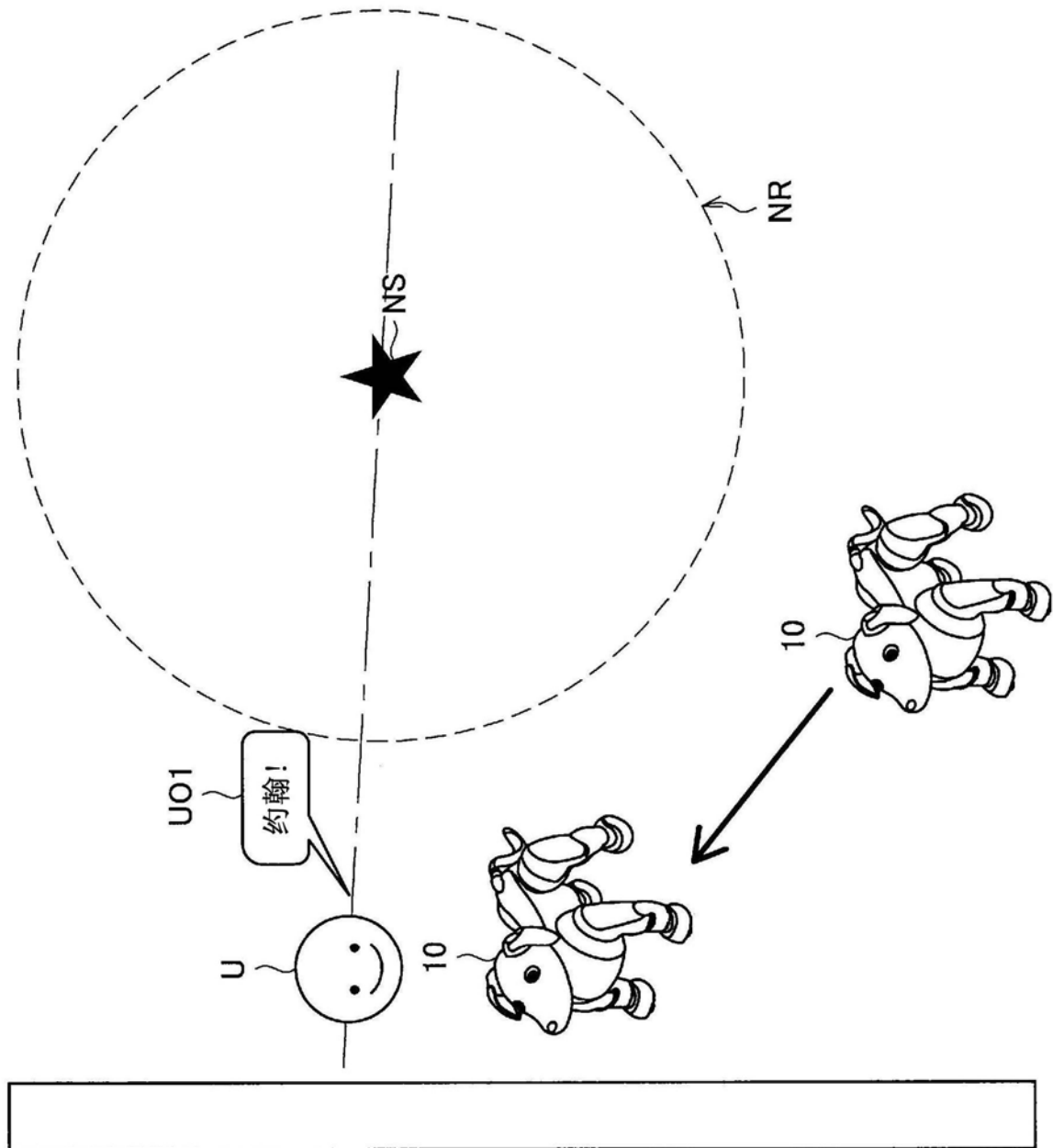


图11

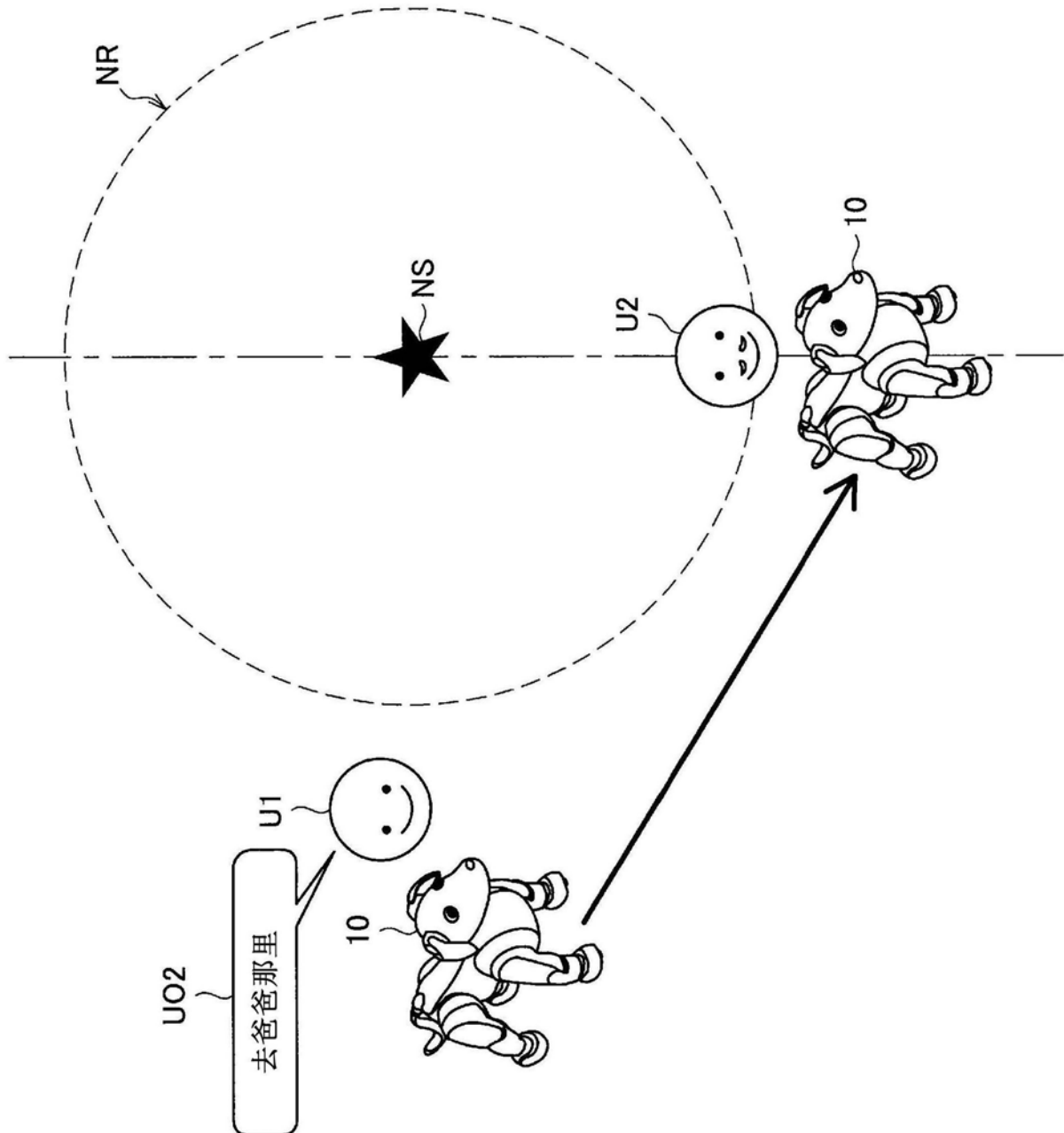


图12

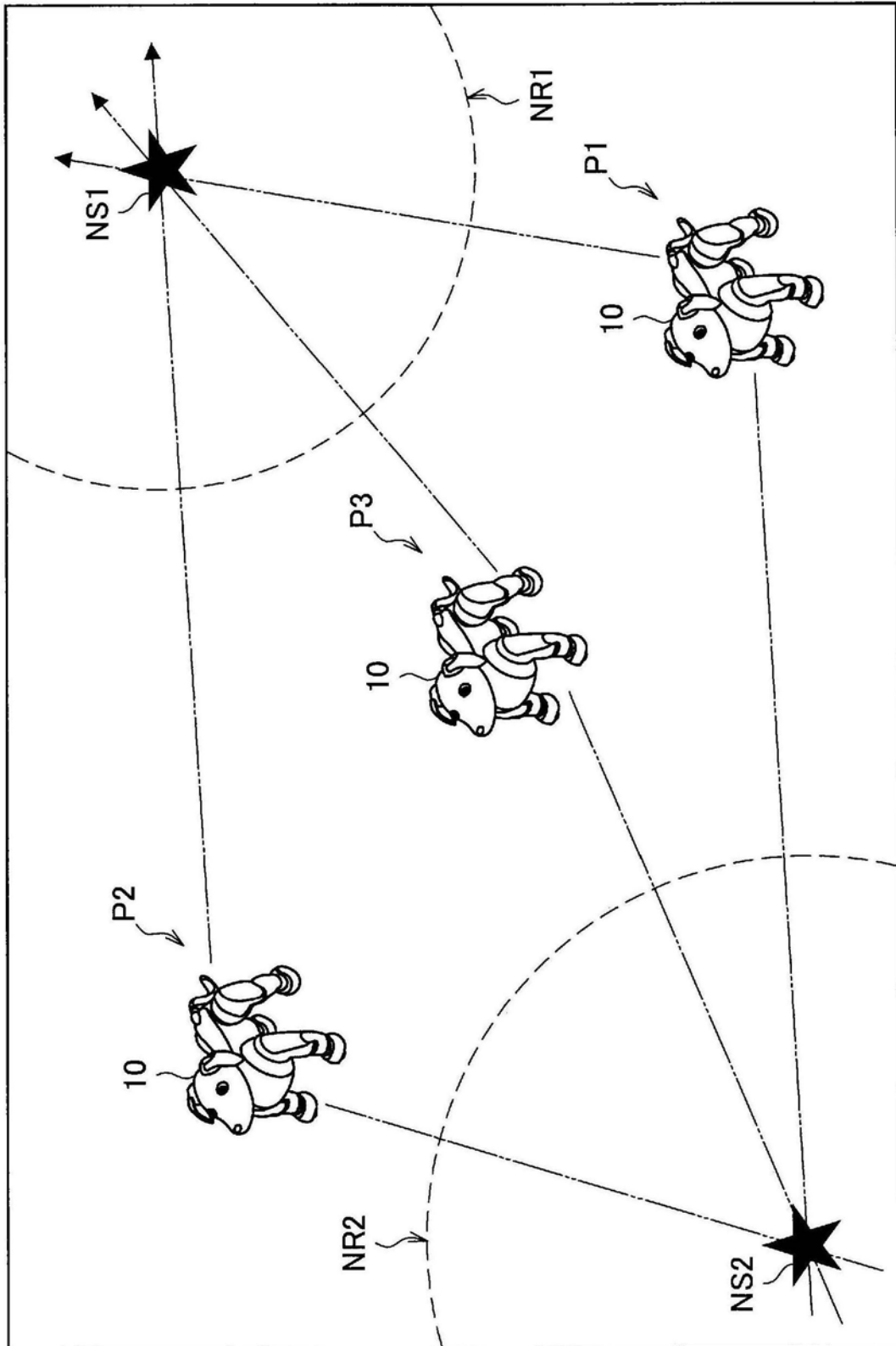


图13

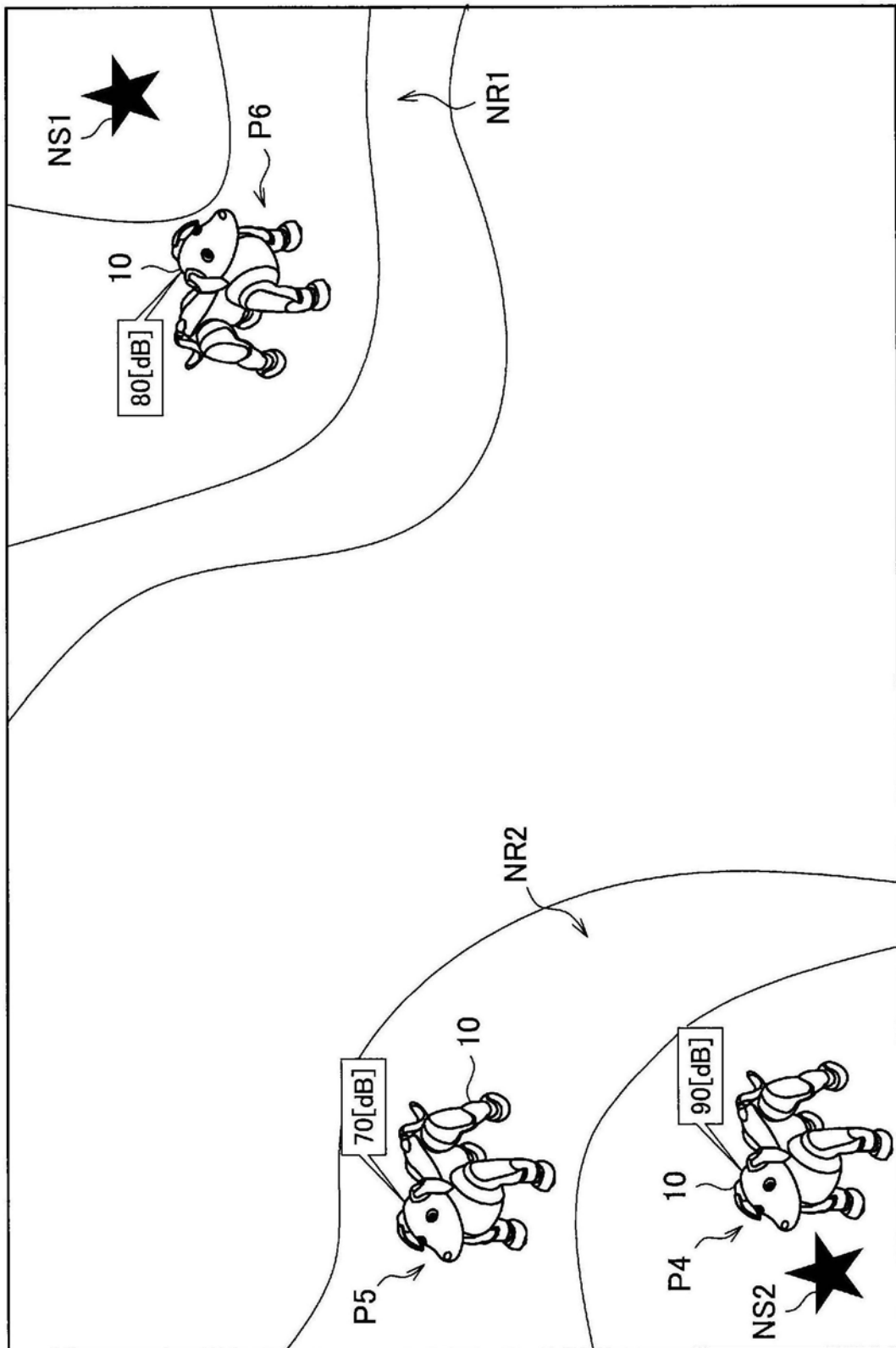


图14

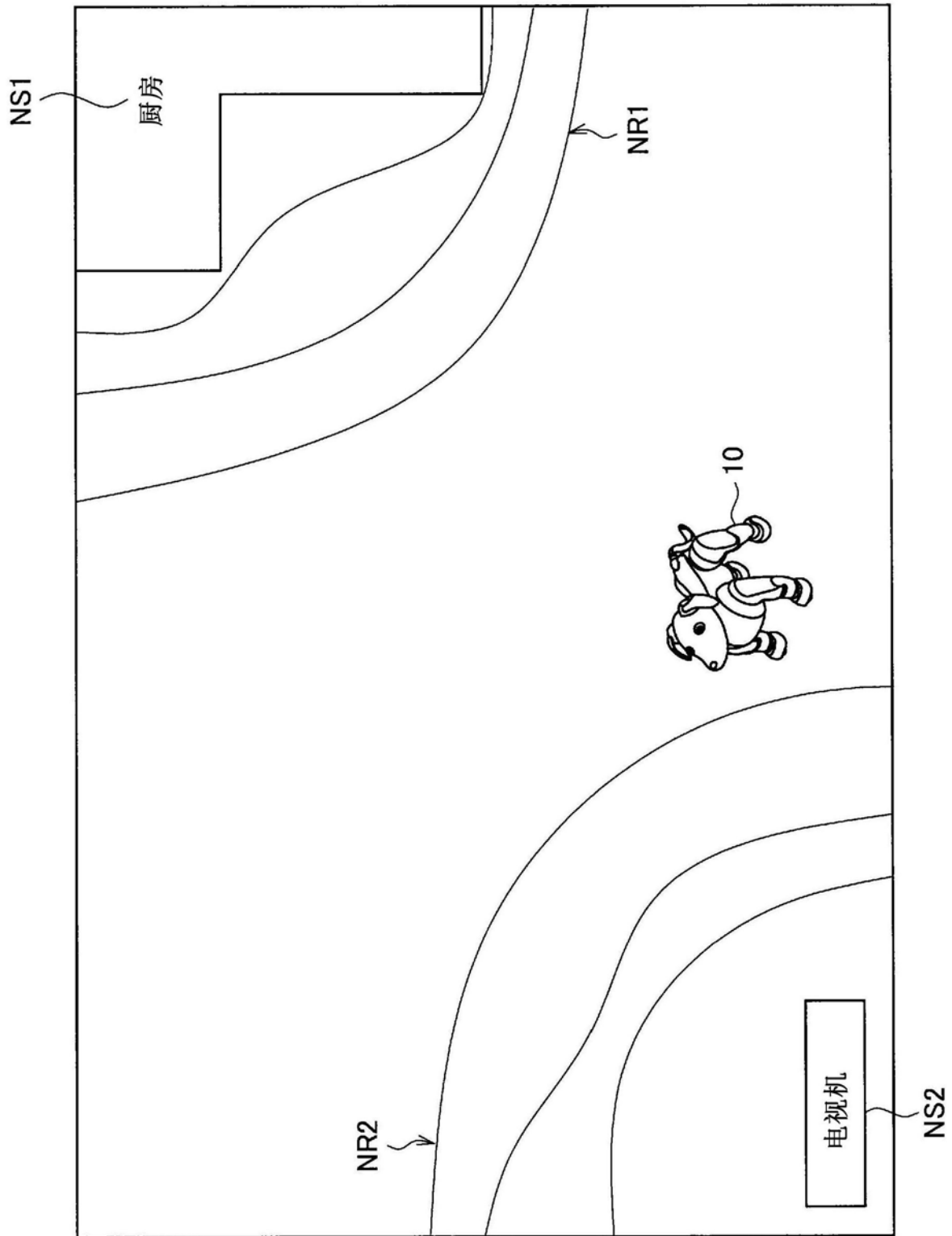


图15

掌握 日程安排	检测 钥匙声音	检测开门/ 关门噪声	检测诸如 “我在家”的话语	人体检测 传感器	...	执行
✓	✓	✓	✓	✓	...	是
	✓	✓	✓	✓	...	否
✓		✓	✓	✓	...	否
✓	✓		✓	✓	...	否
✓	✓	✓		✓	...	否
✓	✓	✓	✓		...	否
		✓	✓	✓	...	否
✓			✓	✓	...	否
✓	✓			✓	...	否
✓	✓	✓			...	否
	✓		✓	✓	...	否
✓		✓		✓	...	否
✓	✓		✓		...	否
	✓	✓		✓	...	否
✓		✓			...	否
		✓	✓	✓	...	否
✓		✓	✓		...	否
	✓				...	否
...	...	...	...	...	...	...

图16

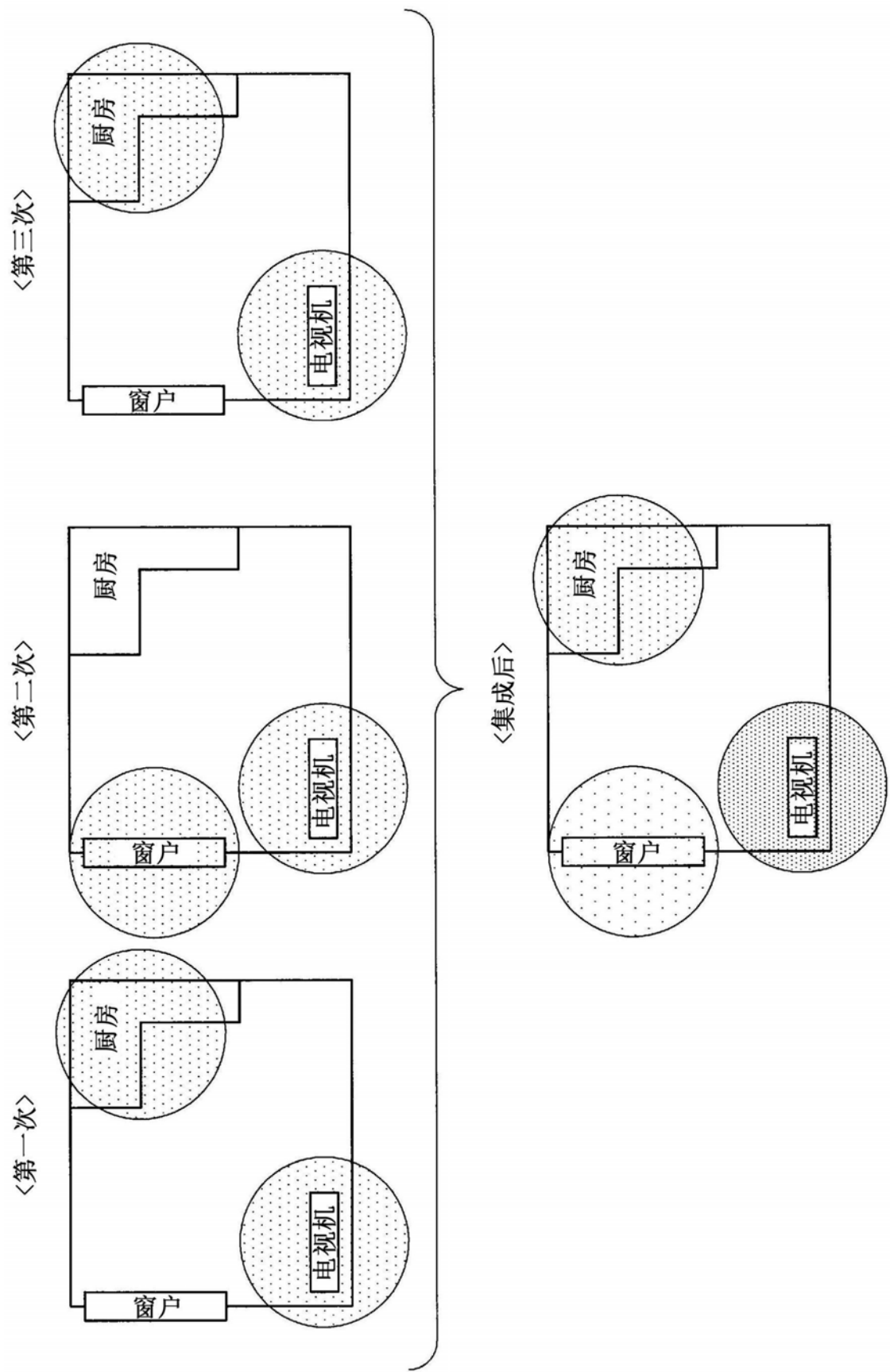


图17

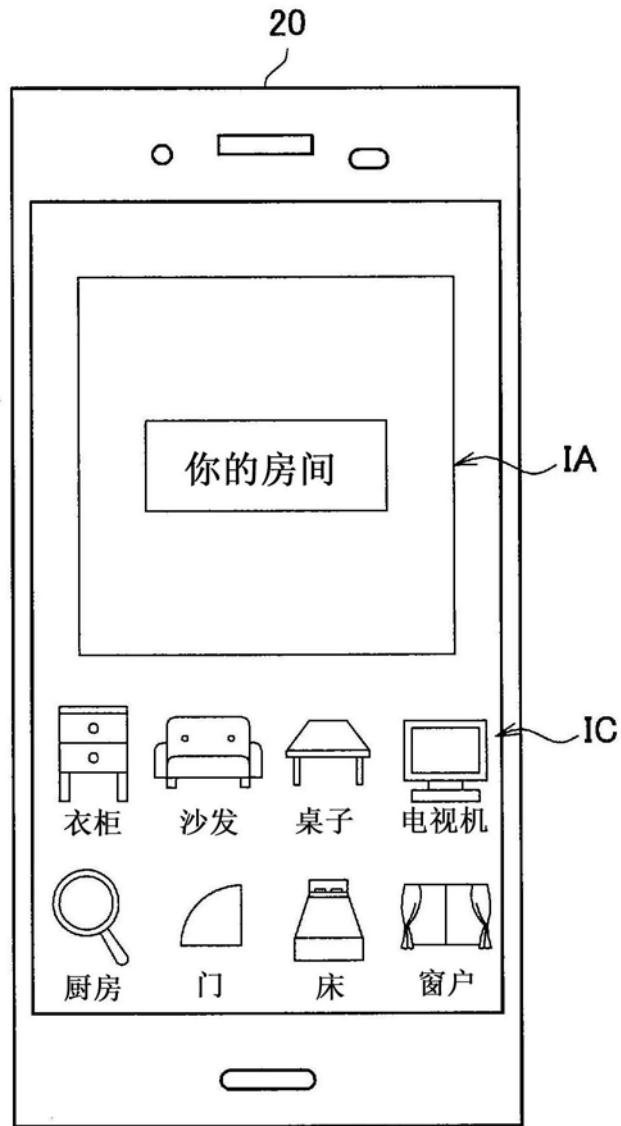


图18

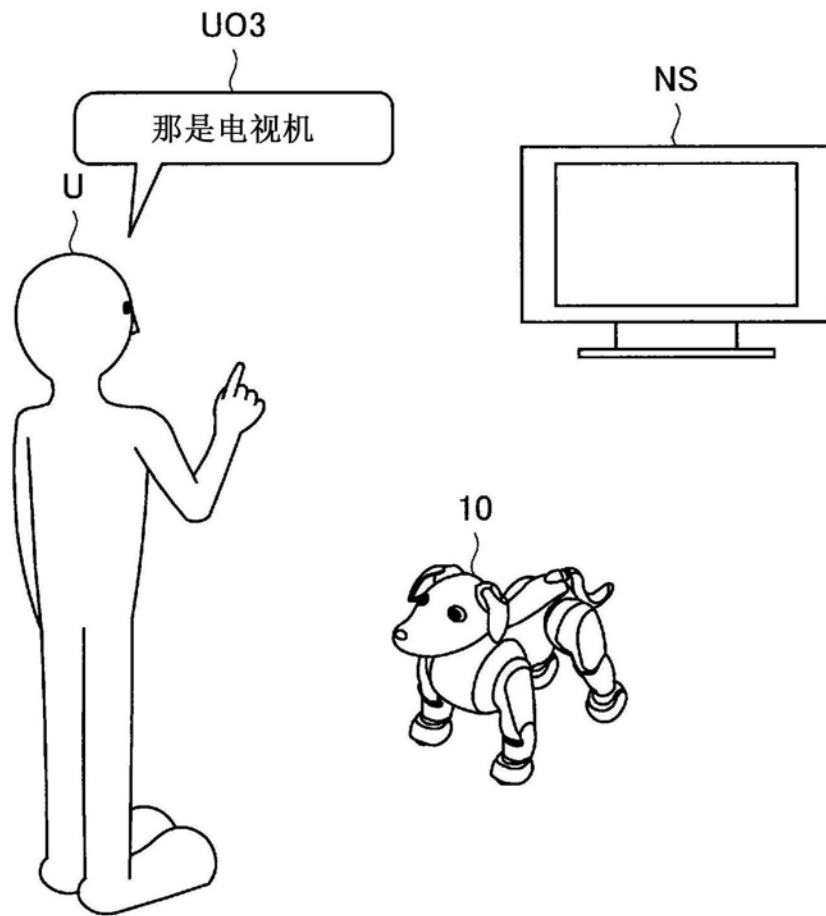


图19

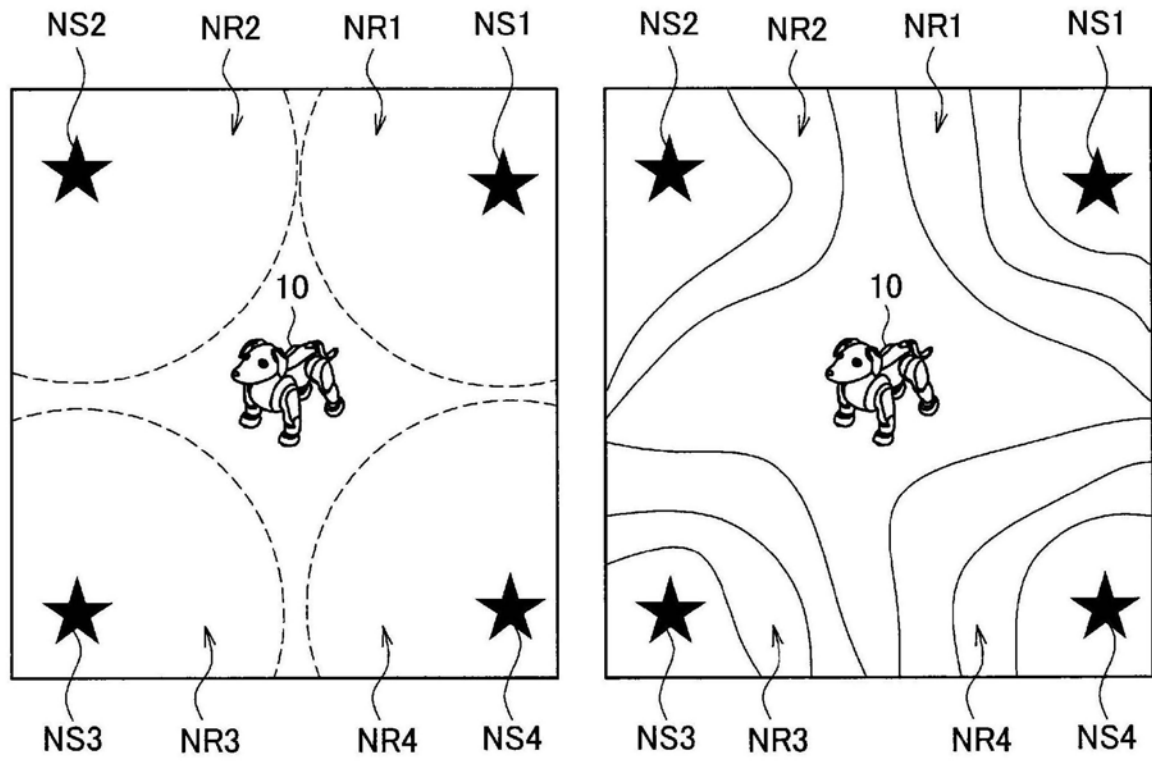


图20

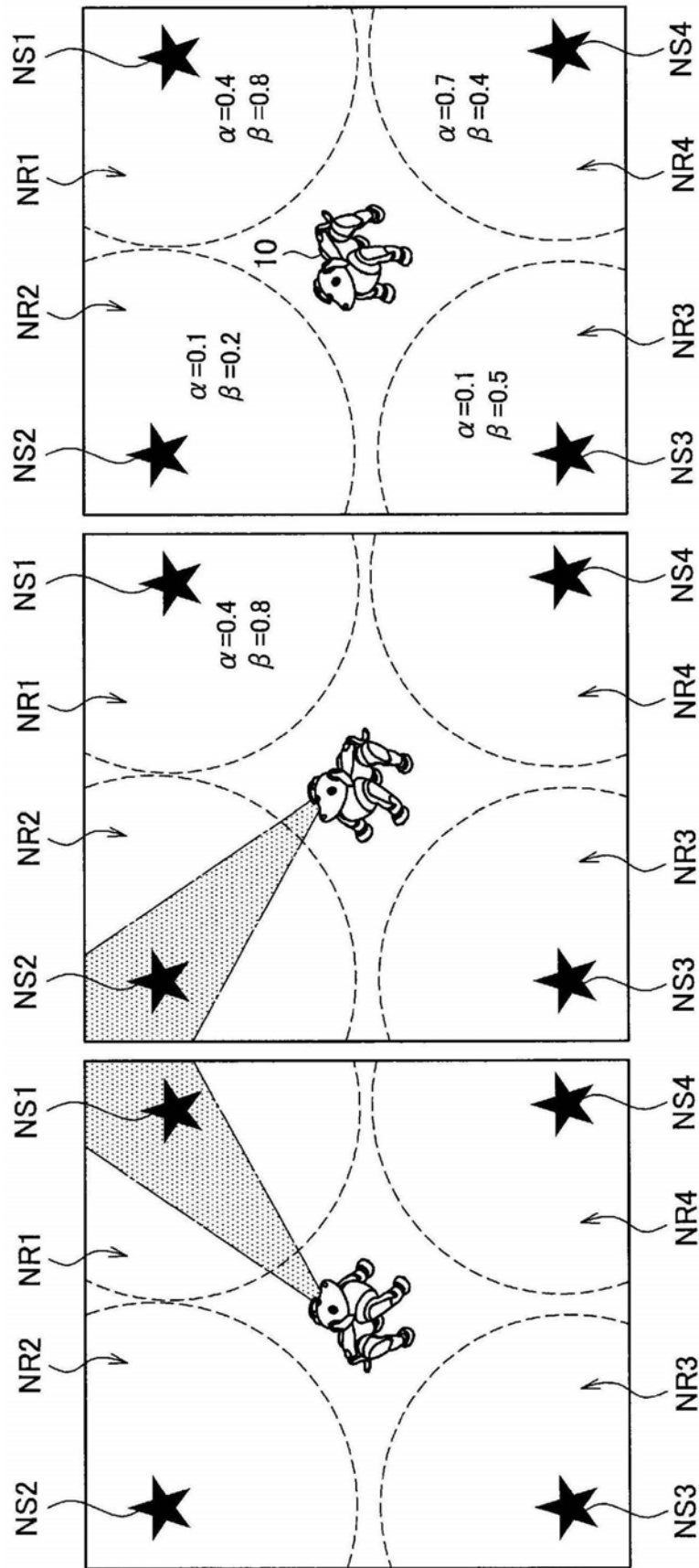


图21

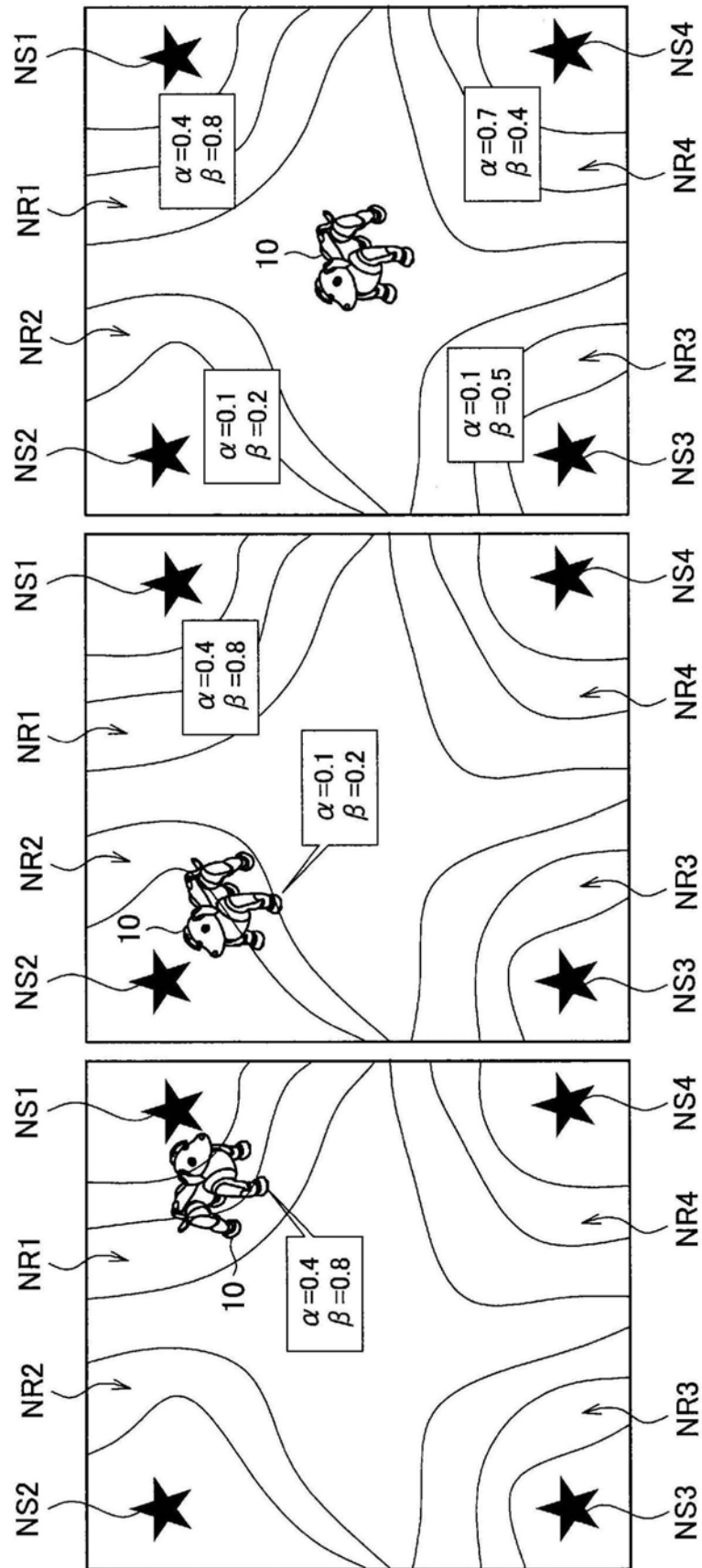


图22

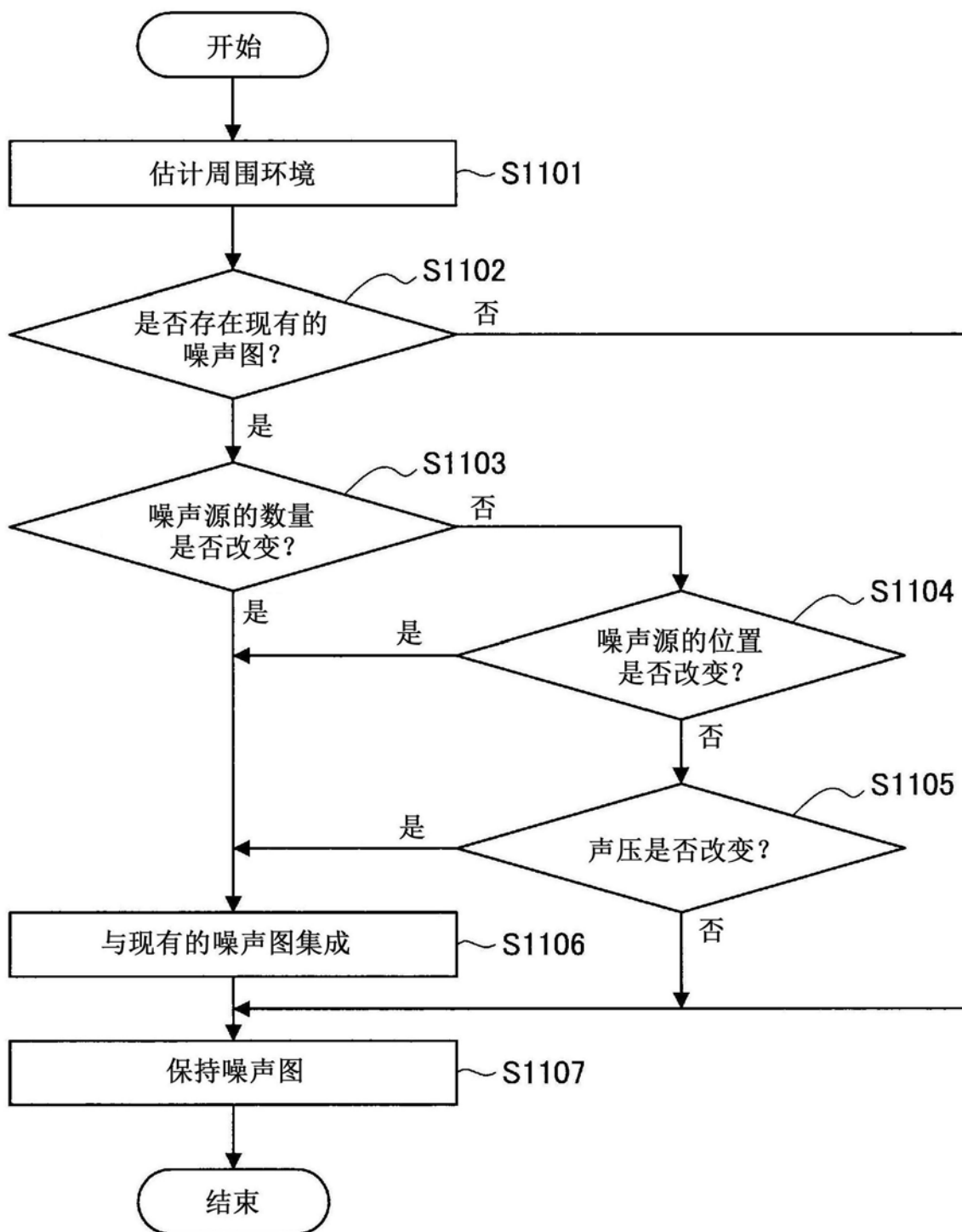


图23

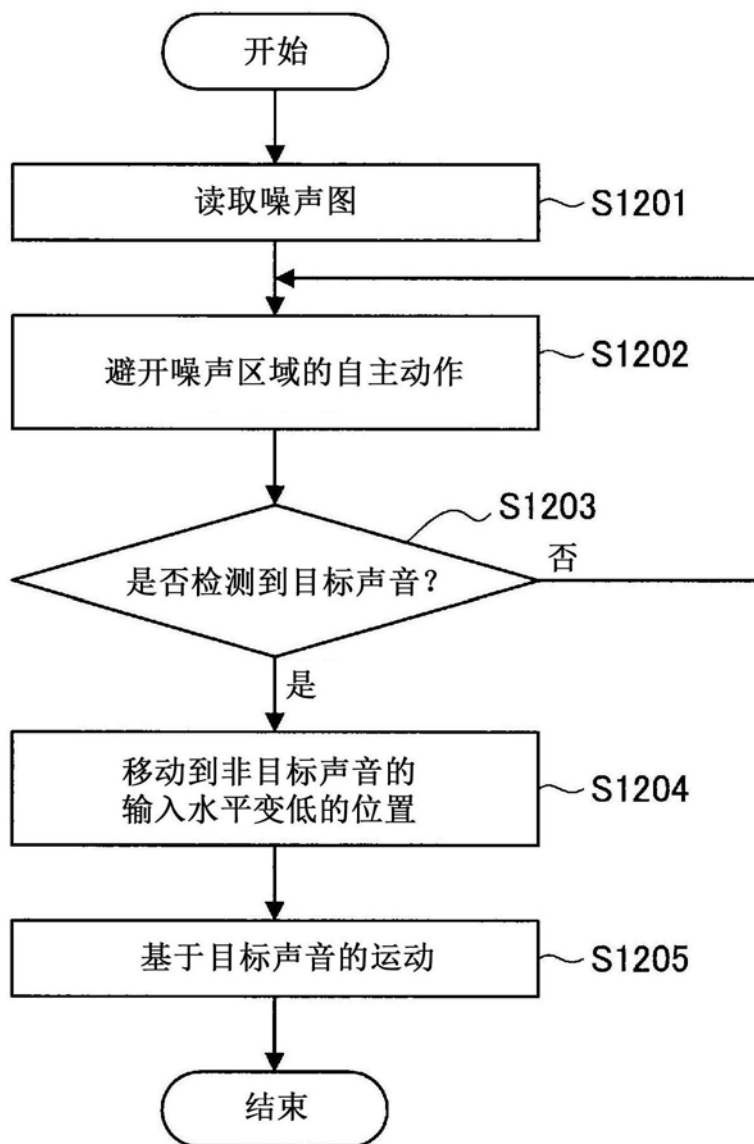


图24