(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112840397 A
(43)申请公布日 2021.05.25

(21)申请号 201980065946.7

(22)申请日 2019.07.31

(30)优先权数据
2018-196739 2018.10.18 JP

(85)PCT国际申请进入国家阶段日
2021.04.06

(86)PCT国际申请的申请数据
PCT/JP2019/029985 2019.07.31

(87)PCT国际申请的公布数据
WO2020/079918 JA 2020.04.23

(71)申请人 索尼公司
地址 日本东京

(72)发明人 历本纯一

(74)专利代理机构 北京康信知识产权代理有限责任公司 11240
代理人 吴孟秋

(51)Int.Cl.
*G10L 15/24*(2013.01)
*G10L 15/16*(2006.01)

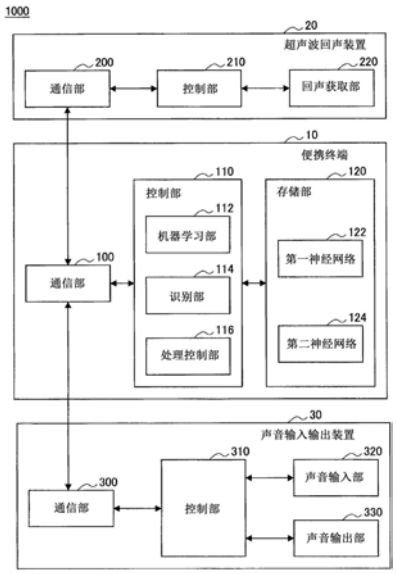
权利要求书1页 说明书16页 附图10页

(54)发明名称

信息处理装置及信息处理方法

(57)摘要

提供能够在用户不发声的情况下得到期望的声学信息的信息处理装置及信息处理方法。信息处理装置具备控制部(110),该控制部(110)基于通过机器学习而获得的算法,将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。




CN 112840397 A

1.一种信息处理装置,具备控制部,

所述控制部基于通过机器学习而获得的算法,将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。

2.根据权利要求1所述的信息处理装置,其中,

所述算法具有第一神经网络,

所述控制部经由所述第一神经网络将所输入的无发声时的多张时间序列图像转换为第一声学信息。

3.根据权利要求2所述的信息处理装置,其中,

所述第一神经网络根据所输入的所述无发声时的多张时间序列图像生成多个每单位时间的声学特征量,并通过按时间序列顺序合成所生成的多个所述声学特征量来生成所述第一声学信息。

4.根据权利要求3所述的信息处理装置,其中,

所述第一神经网络从在所述单位时间获取的所述无发声时的多张时间序列图像中选择所述单位时间的中央时刻的时间序列图像,并根据所选择的所述时间序列图像生成每所述单位时间的声学特征量。

5.根据权利要求2所述的信息处理装置,其中,

所述第一神经网络是通过使用了第一学习信息的所述机器学习而得到的,所述第一学习信息包括发声时的声音和所述发声时的多张时间序列图像。

6.根据权利要求2所述的信息处理装置,其中,

所述算法还具有第二神经网络,

所述控制部经由所述第二神经网络将所述第一声学信息转换为与发声时的声音对应的第二声学信息。

7.根据权利要求6所述的信息处理装置,其中,

所述第二神经网络是通过使用了第二学习信息的所述机器学习而得到的,所述第二学习信息包括通过将所述发声时的多张时间序列图像输入到所述第一神经网络而生成的第三声学信息和与发声时的声音对应的第四声学信息。

8.根据权利要求2所述的信息处理装置,其中,

所述声学信息是语谱图。

9.根据权利要求1所述的信息处理装置,其中,

所述多张时间序列图像表示在用户不发声的情况下活动口或舌中的至少一方时的所述口腔内的状态的变化。

10.根据权利要求1所述的信息处理装置,其中,

所述机器学习是通过深度学习而进行的。

11.根据权利要求1所述的信息处理装置,其中,

所述机器学习是使用卷积神经网络而进行的。

12.一种信息处理方法,由处理器执行,

所述信息处理方法包括:基于通过机器学习而获得的算法,将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。

# 信息处理装置及信息处理方法

## 技术领域

[0001] 本公开涉及信息处理装置及信息处理方法。

## 背景技术

[0002] 近年来，由于声音识别精度的提高，能够使用声音命令控制的设备正在普及。例如，在智能手机、汽车导航装置等中，使用声音命令来使用检索功能正在一般化。另外，基于将通过声音输入的内容文档化的文档生成正在成为可能。另外，被称为智能扬声器的通过声音命令进行动作的扬声器型的声音接口装置正在普及。

[0003] 然而，使用声音命令的情况可能是有限的。例如，在电车中、在图书馆等公共空间中，通过声音操作智能手机等难以被周围的人接受。另外，在公共空间中，发出个人信息等具有隐匿性的信息，存在个人信息泄漏的风险。因此，使用声音命令的声音接口，容易被限定在发声对周围的影响明确的场所使用，如在家庭内使用的智能扬声器、在车内使用的汽车导航装置那样。

[0004] 例如，只要能够在不实际发出声音的情况下操作上述设备等，则不限于场所，能够利用上述设备等。具体地，如果是具有能够在不发出声音的情况下操作设备的功能的可穿戴计算机，则通过始终佩戴该可穿戴计算机，能够与场所无关地始终得到服务。因此，与能够在不发出声音的情况下进行声音识别的无发声发话的识别技术相关的研究正在进行。

[0005] 与上述的无发声发话的识别技术相关联，例如，在下述专利文献1中，公开了一种通过电磁波检测声音器官的运动及场所来识别声音的技术。另外，除了下述专利文献1中公开的技术以外，与用于在噪声环境下可靠地获取声音的咽喉麦克风及粘贴在喉部的麦克风等相关的研究也正在进行。

[0006] 现有技术文献

[0007] 专利文献

[0008] 专利文献1：日本特表2000-504848号公报

## 发明内容

[0009] 发明要解决的技术问题

[0010] 但是，上述的无发声发话的识别技术由于需要发出耳语程度的声音，因此在公共空间中的利用仍然是有限的。另外，为了更加接近无发声，如果减小耳语时的音量，则识别精度可能会降低。

[0011] 因此，本公开提出了一种能够在用户不发声的情况下得到期望的声学信息的新颖且改进的信息处理装置及信息处理方法。

[0012] 用于解决技术问题的方案

[0013] 根据本公开，提供了一种信息处理装置，具备控制部，所述控制部基于通过机器学习而获得的算法，将表示通过超声波回声而获得的口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。

[0014] 另外，根据本公开，提供了一种由处理器执行的信息处理方法，包括基于通过机器学习而获得的算法，将表示通过超声波回声而获得的口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。

[0015] 发明效果

[0016] 如上所述，根据本公开，能够在用户不发声的情况下得到期望的声学信息。另外，上述的效果并不一定是限定的，除了上述的效果之外，或者取代上述的效果，也可以起到本说明书所示的任意的效果，或者能够根据本说明书掌握的其他的效果。

## 附图说明

[0017] 图1是示出本公开的实施方式所涉及的无声发话系统的构成例的图。

[0018] 图2是示出该实施方式所涉及的回声图像的图。

[0019] 图3是示出该实施方式所涉及的无声发话系统的功能的概要的图。

[0020] 图4是示出该实施方式所涉及的无声发话系统的功能构成例的框图。

[0021] 图5是示出该实施方式所涉及的声学特征量的生成例的图。

[0022] 图6是示出该实施方式所涉及的第二神经网络的结构的图。

[0023] 图7是示出该实施方式所涉及的获取第一神经网络的机器学习的流程的流程图。

[0024] 图8是示出该实施方式所涉及的获取第二神经网络的机器学习的流程的流程图。

[0025] 图9是示出该实施方式所涉及的便携终端中的处理的流程的流程图。

[0026] 图10是示出该实施方式所涉及的信息处理装置的硬件构成例的框图。

## 具体实施方式

[0027] 以下，参照附图对本公开的优选实施方式进行详细说明。另外，在本说明书及附图中，对实质上具有相同的功能结构的构成要素赋予相同的附图标记，并省略重复说明。

[0028] 另外，说明按以下顺序进行。

[0029] 1.本公开的实施方式

[0030] 1.1.概要

[0031] 1.2.无声发话系统的结构

[0032] 1.3.无声发话系统的功能

[0033] 1.4.无声发话系统的处理

[0034] 2.变形例

[0035] 3.应用例

[0036] 4.硬件构成例

[0037] 5.总结

[0038] <<1.本公开的实施方式>>

[0039] <1.1.概要>

[0040] 近年来，由于声音识别精度的提高，能够使用声音命令控制的设备正在普及。例如，在智能手机、汽车导航装置等中，使用声音命令来使用检索功能正在一般化。另外，基于将通过声音输入的内容文档化的文档生成正在成为可能。另外，被称为智能扬声器的通过声音命令进行动作的扬声器型的声音接口装置正在普及。

[0041] 然而，使用声音命令的情况可能是有限的。例如，在电车中、在图书馆等公共空间中，通过声音操作智能手机等难以被周围的人接受。另外，在公共空间中，发出个人信息等具有隐匿性的信息，存在个人信息泄漏的风险。因此，使用声音命令的声音接口，容易被限定在发声对周围的影响明确的场所使用，如在家庭内使用的智能扬声器、在车内使用的汽车导航装置那样。

[0042] 例如，只要能够在不实际发出声音的情况下操作上述设备等，则不限于场所，能够利用上述设备等。具体地，如果是具有能够在不发出声音的情况下操作设备的功能的可穿戴计算机，则通过始终佩戴该可穿戴计算机，能够与场所无关地始终得到服务。因此，与能够在不发出声音的情况下进行声音识别的无发声发话的识别技术相关的研究正在进行。

[0043] 与上述的无发声发话的识别技术相关联，例如，公开了一种通过电磁波检测声音器官的运动及场所来识别声音的技术。另外，除此之外，与用于在噪声环境下可靠地获取声音的咽喉麦克风及粘贴在喉部的麦克风等相关的研究也正在进行。

[0044] 但是，上述的无发声发话的识别技术由于需要发出耳语程度的声音，因此在公共空间中的利用仍然是有限的。另外，为了更加接近无发声，如果减小耳语时的音量，则识别精度可能会降低。

[0045] 鉴于上述问题，在本公开的实施方式中，提出了一种能够在用户不发声的情况下得到期望的声学信息的技术。以下，依次对本实施方式进行详细说明。

[0046] <1.2.无声发话系统的结构>

[0047] 首先，对本公开的实施方式所涉及的无声发话系统的结构进行说明。图1是示出本公开的实施方式所涉及的无声发话系统的构成例的图。如图1所示，本实施方式所涉及的无声发话系统1000具备便携终端10、超声回声装置20及声音输入输出装置30。各种装置可以与便携终端10连接。例如，超声波回声装置20及声音输入输出装置30与便携终端10连接，在各装置间进行信息的协作。在本实施方式所涉及的便携终端10上，以无线方式连接有超声波回声装置20及声音输入输出装置30。例如，便携终端10与超声波回声装置20及声音输入输出装置30进行使用Bluetooth(注册商标)的近距离无线通信。另外，在便携终端10上，超声波回声装置20及声音输入输出装置30既可以有线连接，也可以经由网络连接。

[0048] (1)便携终端10

[0049] 便携终端10是能够进行基于机器学习的识别处理的信息处理装置。本实施方式所涉及的识别处理例如是声音识别处理。该声音识别处理例如对与根据图像(静止图像/动态图像)生成的声音相关的信息进行。具体地，便携终端10将表示用户12的口腔内的状态的图像(以下，也称为回声图像)转换为与声音相关的信息，并对与转换后的声音相关的信息进行声音识别处理。

[0050] 另外，在本实施方式中，表示在用户12不发声的情况下使口腔内的状态变化时的口腔内的状态的时间序列变化的多张时间序列图像被转换为与声音相关的信息。由此，本实施方式所涉及的便携终端10能够实现基于无发声的声音识别。另外，该多张时间序列图像是表示在用户不发声的情况下活动口或舌中的至少一方时的口腔内的状态的变化的回声图像。另外，以下，表示用户12的口腔内的状态的时间序列变化的多张时间序列图像也称为时间序列回声图像。

[0051] 与声音相关的信息例如是声音识别装置可识别的信息(以下，也称为声学信息)。

声学信息例如是根据频率、振幅及时间三维地表示声音的高低、强度等声音的特征的时间序列变化的语谱图。

[0052] 与声音相关的信息使用通过机器学习而获得的算法根据图像转换而来。本实施方式所涉及的机器学习例如通过深度学习进行。通过该机器学习获取的算法例如是神经网络(NN:Neural Network)。另外,在该机器学习中,图像被用作输入。因此,该机器学习使用适合图像处理的深度学习的卷积神经网络(CNN:Convolutional Neural Network)进行。另外,在本实施方式中,用户12发出声音时的时间序列回声图像用于机器学习。

[0053] 在本实施方式所涉及的算法中存在两种算法(神经网络)。第一个算法是进行将在用户12不发声的情况下使口腔内的状态变化时的时间序列回声图像转换为声学信息(第一声学信息)的处理的第一神经网络(以下,也称为NN1)。第二个算法是进行将NN1转换的声学信息转换为精度更高的声学信息(第二声学信息)的处理的第二神经网络(以下,也称为NN2)。精度更高的声学信息例如是用户12实际发声时的声音即发声声音被转换后的声学信息。另外,NN1及NN2的详细情况将在后面叙述。

[0054] 另外,如上所述,在本实施方式所涉及的时间序列回声图像中,存在通过NN1转换为声学信息的时间序列回声图像和用于机器学习的时间序列回声图像这两种。由于转换为声学信息的时间序列回声图像是在用户12不发声的情况下使口腔内的状态变化时的口腔内的时间序列回声图像,因此以下也称为无发声时间序列回声图像。另外,由于用于机器学习的时间序列回声图像是用户12发声时的口腔内的时间序列回声图像,因此以下也称为发声时间序列回声图像。

[0055] 另外,如上所述,在本实施方式所涉及的声学信息中存在多个声学信息。由于通过NN1转换的声学信息(第一声学信息)是无发声时间序列回声图像被转换的语谱图,因此以下称为无发声图像语谱图。另外,由于通过NN2转换的声学信息(第二声学信息)是无发声图像语谱图被转换的精度更高的语谱图,因此以下称为高精度无发声图像语谱图。

[0056] 另外,机器学习分别在NN1及NN2中进行,但在各个机器学习中使用的学习信息不同。用于NN1的机器学习的学习信息(第一学习信息)是发声时间序列回声图像及发声声音。用于NN2的机器学习的学习信息(第二学习信息)是发声图像语谱图经由NN1转换的声学信息(第三声学信息)和与发声声音对应的声学信息(第四声学信息)。

[0057] 由于发声图像语谱图经由NN1转换的声学信息(第三声学信息)是发声时间序列回声图像通过NN1转换的语谱图,因此以下称为发声图像语谱图。与发声声音对应的声学信息(第四声学信息)是与用户12实际发声时的声音对应的语谱图,因此以下称为发声声音语谱图。另外,发声图像语谱图(第三声学信息)被用作NN2的机器学习的输入,而发声声音语谱图(第四声学信息)被用作NN2的机器学习的输出。

[0058] 另外,便携终端10还具有控制无声发话系统1000的全部动作的功能。例如,便携终端10基于在各装置间协作的信息,控制无声发话系统1000的全部动作。具体地,便携终端10基于从超声波回声装置20及声音输入输出装置30接收的信息,控制便携终端10中的与声音识别相关的处理、声音输入输出装置30的动作。另外,便携终端10也可以控制超声波回声装置20的动作。

[0059] 如图1所示,便携终端10例如通过智能手机实现。另外,便携终端10并不限于智能手机。例如,便携终端10也可以是将作为便携终端10的功能作为应用安装的平板终端、PC、

可穿戴终端或者代理设备等终端装置。即，便携终端10可以作为任意的终端装置实现。

[0060] (2)超声波回声装置20

[0061] 超声波回声装置20是获取用户12的口腔内的回声图像的装置。超声波回声装置20利用在医疗中广泛使用的超声波检查技术获取回声图像。超声波回声装置20具备能够输出超声波的超声波输出装置，使附着在用户12的体表上的该超声波输出装置向用户12的体内输出超声波，基于由用户12的体内的器官反射的超声波获取回声图像。另外，超声波回声装置20将获取到的回声图像发送到便携终端10。

[0062] 如图1所示，本实施方式所涉及的超声波回声装置20例如作为颈带型的装置实现。超声波回声装置20的超声波输出部22具备超声波输出装置。图1所示的超声波回声装置20在作为颈带型的装置的构造上，具备超声波输出部22a及22b这两个超声波输出部22。另外，超声波输出部22的数量并不限于两个，超声波回声装置20只要具备至少一个以上的超声波输出部22即可。

[0063] 通过用户12以超声波输出部22位于下颌下方的方式佩戴超声波回声装置20，超声波向用户12的口腔内输出。由此，超声波回声装置20能够获取用户12的口腔内的回声图像。声音在通过舌、口的打开方式来调整声带的振动的基础上发出。因此，可以说通过超声波回声装置20获取的用户12的口腔内的回声图像作为转换为声学信息的图像具有有效的信息。

[0064] 这里，对通过超声波回声装置20获取的回声图像进行说明。图2是示出本实施方式所涉及的回声图像的图。回声图像40是通过超声波回声装置20获取的用户12的口腔内的回声图像。在图2所示的回声图像40中，示出了舌尖402、舌的表面404、舌的根部406。在本实施方式中，超声波回声装置20通过连续地获取回声图像40，来获取表示用户12的口腔内的状态的时间序列变化的多张时间序列图像(时间序列回声图像)。

[0065] (3)声音输入输出装置30

[0066] 声音输入输出装置30是能够进行声音的输入输出的装置。声音输入输出装置30例如获取用户12发出的声音。另外，声音输入输出装置30将获取的声音发送到便携终端10。另外，声音输入输出装置30例如从便携终端10接收表示便携终端10识别的内容的声音数据。另外，声音输入输出装置30将所接收的声音数据作为声音输出。

[0067] 本实施方式所涉及的声音输入输出装置30例如通过可穿戴终端实现。具体地，优选声音输入输出装置30为可进行声音的输入输出的耳机、骨传导耳机等可穿戴终端。通过声音输入输出装置30为耳机、骨传导耳机等，能够减少泄漏到外部的声音的量。

[0068] 另外，更优选声音输入输出装置30为除了从声音输入输出装置30输出的声音以外，用户12还能够听到在外部产生的声音的结构。例如，如图1所示，声音输入输出装置30具有开口部32。因此，即使用户12佩戴声音输入输出装置30，也能够通过开口部32听到外部的声音。因此，即使用户12始终佩戴具有该结构的声音输入输出装置30，也能够不妨碍日常生活地舒适地度过。另外，即使在表示便携终端10识别的内容的声音不是从声音输入输出装置30，而是从智能扬声器等扬声器输出的情况下，用户12也能够听到该声音。

[0069] 另外，在本实施方式中，对在一个装置中实现输出声音的声音输出功能和获取声音的声音输入功能的例子进行说明，但声音输出功能和声音输入功能也可以由分别独立的装置实现。

[0070] <1.3.无声发话系统的功能>

[0071] 以上，对无声发话系统1000的结构进行了说明。接着，对无声发话系统1000的功能进行说明。

[0072] <1.3.1.功能的概要>

[0073] 图3是示出本实施方式所涉及的无声发话系统的功能的概要的图。首先，无声发话系统1000通过基于发声时间序列回声图像及发声声音的机器学习来预先获取NN1及NN2。在用户12不发出声音的情况下使口腔内的状态变化时，超声波回声装置20获取无发声时间序列回声图像42。接着，所获取的无发声时间序列回声图像42经由第一神经网络122(NN1)被转换为无发声图像语谱图72。无发声图像语谱图72是通过按时间序列顺序组合多个声学特征量70而成的。声学特征量70的详细情况将在后面叙述。

[0074] 在基于NN1的转换处理之后，转换的无发声图像语谱图72经由第二神经网络124(NN2)被转换为高精度无发声图像语谱图74。在基于NN2的转换处理之后，转换的高精度无发声图像语谱图74被输入到便携终端10的识别部114。另外，识别部114基于所输入的高精度无发声图像语谱图74进行声音识别处理。

[0075] <1.3.2.功能构成例>

[0076] 图4是示出本实施方式所涉及的无声发话系统的功能构成例的框图。

[0077] (1)便携终端10

[0078] 如图4所示，便携终端10具备通信部100、控制部110及存储部120。另外，本实施方式所涉及的信息处理装置至少具有控制部110。

[0079] (1-1)通信部100

[0080] 通信部100具有与外部装置进行通信的功能。例如，在与外部装置的通信中，通信部100将从外部装置接收的信息输出到控制部110。具体地，通信部100将从超声波回声装置20接收的回声图像输出到控制部110。另外，通信部100将从声音输入输出装置30接收的声音输出到控制部110。

[0081] 在与外部装置的通信中，通信部100将从控制部110输入的信息发送到外部装置。具体地，通信部100将从控制部110输入的与回声图像的获取相关的信息发送到超声波回声装置20。另外，通信部100将从控制部110输入的与声音的输入输出相关的信息发送到声音输入输出装置30。

[0082] (1-2)控制部110

[0083] 控制部110具有控制便携终端10的动作的功能。例如，控制部110基于通过机器学习而获得的算法，将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与口腔内的状态对应的信息。算法具有第一神经网络，控制部110经由第一神经网络将所输入的无发声时的多张时间序列图像转换为第一声学信息。例如，控制部110将从通信部100输入的无发声时间序列回声图像输入到NN1。NN1将所输入的无发声时间序列回声图像转换为无发声图像语谱图。控制部110能够通过将语谱图转换为声音波形来进行声音识别处理。因此，即使用户12不发出声音，控制部110也能够基于无发声时间序列回声图像进行声音识别处理，控制能够通过声音操作的设备。

[0084] 另外，算法进一步具有第二神经网络，控制部110经由第二神经网络将第一声学信息转换为与发声时的声音对应的第二声学信息。例如，控制部110将从NN1输出的无发声图像语谱图输入到NN2。NN2将所输入的无发声图像语谱图转换为与发声声音对应的高精度无

发声图像语谱图。具体地，假设从NN1输出的无发声图像语谱图表示的声音是“Ulay music.”，则与发声声音对应的高精度无发声图像语谱图表示的声音是“Play music.”。此时，在表示声音“Ulay music.”的无发声图像语谱图被输入到NN2时，考虑到上下文等，被转换为表示声音“Play music.”的高精度无发声图像语谱图。即，NN2起到校正NN1从无发声时间序列回声图像转换的无发声图像语谱图表示的声音的作用。

[0085] 为了实现上述功能，如图4所示，本实施方式所涉及的控制部110具有机器学习部112、识别部114及处理控制部116。

[0086] •机器学习部112

[0087] 机器学习部112具有进行使用了学习信息的机器学习的功能。机器学习部112通过机器学习获取用于将回声图像转换为语谱图的算法。具体地，机器学习部112获取作为用于将无发声时间序列回声图像转换为无发声图像语谱图的算法的NN1。另外，机器学习部112获取作为用于将无发声图像语谱图转换为高精度无发声图像语谱图的算法的NN2。

[0088] NN1通过使用了包括发声时的声音和发声时的多张时间序列图像的第一学习信息的机器学习而得到。例如，NN1通过使用了用户12发出的声音和用户12发出该声音时的发声时间序列回声图像作为第一学习信息的机器学习而得到。由此，控制部110能够经由NN1将回声图像转换为语谱图。

[0089] 另外，该第一学习信息例如通过使用户12朗读文本等来获取。由此，能够获取表示时间序列变化的回声图像和与该回声图像对应的发声波形。发声波形可以被转换为声学特征量。

[0090] 另外，如果控制部110将无发声时的多张时间序列图像输入到NN1，则NN1根据所输入的无发声时的多张时间序列图像生成多个每单位时间的声学特征量，并通过按时间序列顺序合成所生成的多个声学特征量来生成第一声学信息。例如，NN1根据通过控制部110输入的无发声时间序列回声图像生成多个每单位时间的声学特征量，并按时间序列顺序合成所生成的多个声学特征量，从而生成无发声图像语谱图。

[0091] 这里，对通过NN1生成的声学特征量进行说明。图5是示出本实施方式所涉及的声学特征量的生成例的图。NN1从在单位时间获取的无发声时的多张时间序列图像中，选择单位时间的中央时刻的时间序列图像，并根据所选择的时间序列图像生成每单位时间的声学特征量。例如，NN1在单位时间获取的无发声时间序列回声图像中，选择单位时间的中央时刻的回声图像，并根据所选择的回声图像生成每单位时间的声学特征量。本实施方式所涉及的单位时间例如是所获取的回声图像的张数为5张～13张中的任一张数的时间。在本实施方式中，将获取13张回声图像的时间作为单位时间。具体地，如图5所示，NN1在无发声时间序列回声图像42中，选择在单位时间获取的无发声时间序列回声图像422的中央的回声图像424，并根据该回声图像424生成声学特征量70。另外，NN1以错开单位时间的开始时刻的方式重复声学特征量70的生成处理，并通过合成所生成的多个声学特征量70来获取无发声图像语谱图78。

[0092] 由此，NN1能够学习与th等发音的最小单位相当的口的动作。另外，识别部114能够更正确地识别发音。

[0093] 另外，在声学特征量中，可以利用通过使用神经网络(自动编码器)处理梅尔语谱图(Mel-scale spectrogram)、MFCC(梅尔频率倒谱系数)、短时间FFT(SFFT)、声音波形而缩

小了维数的表现等。另外，与上述自动编码器相关的技术在Jesse Engel等6名的论文(“Neural Audio Synthesis of Musial Notes with WaveNet Autoencoders”,URL:https://arxiv.org/abs/1704.01279)中公开。它们能够与声音波形相互转换。例如，梅尔语谱图通过利用Griffin Lim算法这一方法可以与声学波形合成。作为声学特征量，也可以利用能够在短时间内分割声音波形而成为向量表现的其他表现。另外，在本实施方式中，声学特征量的维数为64维左右，但该维数也可以根据再现的声音的音质进行变更。

[0094] NN2通过使用第二学习信息的机器学习而得到，该第二学习信息包括通过将发声时的多张时间序列图像输入到NN1而生成的第三声学信息和与发声时的声音对应的第四声学信息。例如，NN2通过机器学习而得到，该机器学习使用通过将发声时间序列回声图像输入到NN1而生成的发声图像语谱图和与发声声音对应的发声声音语谱图作为第二学习信息。由此，控制部110能够经由NN2将从NN1输出的无发声图像语谱图转换为精度更高的语谱图。

[0095] 另外，在NN2中，也可以将无发声图像语谱图转换为具有相同长度的语谱图。例如，在无发声图像语谱图是与用户12向智能扬声器等发出的命令对应的语谱图的情况下，NN2将无发声图像语谱图转换为具有相同长度的语谱图。

[0096] 另外，也可以对NN2中输入的语谱图的长度设定固定值。然而，在将长度比该固定值短的语谱图输入到NN2的情况下，控制部110也可以在将不足的长度的无声部插入该语谱图的基础上输入到NN2。另外，在NN2中，由于使用均方误差作为损失函数，因此NN2以对NN2的输入尽量与输出一致的方式进行学习。

[0097] 使用NN2的意图是，将根据与命令对应的无发声时间序列回声图像生成的无发声图像语谱图调整为更接近根据实际说出命令时的声音生成的发声声音语谱图。在NN1中，由于仅将无发声时间序列回声图像的规定张数作为输入，因此无法掌握时间宽度比与该规定张数对应的时间宽度长的上下文。在NN2中，还可以包括命令的上下文进行转换。

[0098] 这里，对NN2的具体结构进行说明。图6是示出本实施方式所涉及的第二神经网络的结构的图。图6示出了将长度为184、维数为64的规定长度的无发声图像语谱图72转换为具有相同长度的高精度无发声图像语谱图74的例子。第一级的1D-Convolution Bank80是一维的CNN，由滤波器的尺寸在1～8的范围内彼此不同的8个NN构成。通过使用多个滤波器尺寸，时间宽度不同的特征被提取。该特征例如是发音符号等级的特征、单词等级的特征等。来自该滤波器的输出通过组合了1D-Convolution和1D-Deconvolution而成的被称为U-Network的NN进行转换。U-Network能够根据通过Convolution/Deconvolution转换的信息识别全局信息。但是，由于局部信息容易丢失，因此U-Network成为用于保证该局部信息的结构。

[0099] 例如，如图6所示，U-Network将长度为184、维数为128的声学特征量802设为长度为96、维数为256的声学特征量804。接着，U-Network将该声学特征量804设为长度为46、维数为512的声学特征量806。另外，U-Network将该声学特征量806设为长度为23、维数为1024的声学特征量808。由此，取代空间的大小变小，通过使空间的深度变深，局部特征被提取。

[0100] 在提取局部特征之后，U-Network以与提取局部特征时的顺序相反的顺序将声学特征量的大小和维数恢复。此时，U-Network将输入直接复制到输出的信息也合并到NN。例如，如图6所示，U-Network将长度为23、维数为1024的声学特征量808设为长度为46、维数为

512的声学特征量810，与复制了声学特征量806的声学特征量812进行合并。接着，U-Network将与声学特征量812合并后的声学特征量810设为长度为96、维数为256的声学特征量814，与复制了声学特征量804的声学特征量816进行合并。另外，U-Network将与声学特征量816合并后的声学特征量814设为长度184、维数为128的声学特征量818，与复制了声学特征量802的声学特征量820进行合并。

[0101] 使用上述U-Network的方法是在学习二维图像的转换(例如从黑白图像向彩色图像的转换)的NN中一般使用的方法，在本实施方式中，将该方法应用于一维的声学特征量序列。

[0102] 另外，在NN2中使用的第二学习信息的数量是用户12为学习信息用而生成的发声声音和发声时间序列回声图像的组合的数量。例如，在用户12为了生成学习信息而进行了300次发话的情况下，生成300个输入和输出的组合。然而，300个的量可能不是为了使NN2学习的足够的量。因此，在第二学习信息的量不充分的情况下，也可以进行数据扩充(Data Augmentation)。数据扩充通过在固定输出的状态下利用随机数扰乱输入的声学特征量，从而能够增加第二学习信息的数量。

[0103] 另外，与NN1及NN2相关的机器学习，通过依赖于特定的说话者，能够更有效地进行学习。因此，优选该机器学习依赖于特定的说话者来进行。另外，也可以进行NN1仅依赖于特定的说话者，NN2统一学习多个说话者的信息等的复合学习。

[0104] • 识别部114

[0105] 识别部114具有进行识别处理的功能。例如，识别部114访问存储部120，进行使用了NN1的转换处理。具体地，识别部114将从通信部100输入的超声波回声装置20所获取的无发声时间序列回声图像输入到NN1。另外，识别部114访问存储部120，进行使用了NN2的转换处理。具体地，识别部114将从NN1输出的无发声图像语谱图输入到NN2。另外，识别部114进行基于从NN2输出的高精度无发声图像语谱图的声音识别处理。另外，识别部114将声音识别处理的结果输出到处理控制部116。

[0106] 另外，识别部114也可以进行仅使用了NN1的声音识别处理。例如，识别部114也可以访问存储部120，进行使用了NN1的转换处理，并进行基于从NN1输出的无发声图像语谱图的声音识别处理。如上所述，在本实施方式中，可以进行基于从NN1输出的无发声图像语谱图的声音识别处理。然而，从NN2输出的高精度无发声图像语谱图比从NN1输出的无发声图像语谱图的精度高。因此，识别部114通过进行不仅使用了NN1而且使用了NN2的声音识别处理，能够以更高的精度进行声音识别处理。

[0107] • 处理控制部116

[0108] 处理控制部116具有控制控制部110中的处理的功能。例如，处理控制部116基于识别部114的声音识别处理的结果来确定要执行的处理。具体地，在声音识别处理的结果表示由用户12指定了由控制部110执行的处理的情况下，处理控制部116执行由用户12指定的处理。另外，在声音识别处理的结果表示是用户12的询问的情况下，处理控制部116执行对该询问进行回答的处理。

[0109] 另外，在处理控制部116执行的处理是对用户输出声音的处理的情况下，处理控制部116能够向用户佩戴的声音输入输出装置30发送该声音，并使声音输入输出装置30输出声音。由此，本实施方式所涉及的无声发话系统1000能够在不向外部泄漏声音的情况下与

用户12进行基于声音的交流。

[0110] (1-3)存储部120

[0111] 存储部120具有存储与便携终端10中的处理相关的数据的功能。例如，存储部120存储作为通过控制部110中的机器学习生成的算法的第一神经网络122及第二神经网络124。控制部110在将无发声时间序列回声图像转换为无发声图像语谱图时，访问存储部120并使用第一神经网络122。另外，控制部110在将无发声图像语谱图转换为高精度无发声图像语谱图时，访问存储部120并使用第二神经网络124。

[0112] 另外，存储部120也可以存储控制部110在机器学习中使用的学习信息。另外，存储部120存储的数据并不限于上述的例子。例如，存储部120也可以存储各种应用程序等程序。

[0113] (2)超声波回声装置20

[0114] 如图4所示，超声波回声装置20具有通信部200、控制部210及回声获取部220。

[0115] (2-1)通信部200

[0116] 通信部200具有与外部装置进行通信的功能。例如，在与外部装置的通信中，通信部200将从外部装置接收的信息输出到控制部210。具体地，通信部200向控制部210输出与从便携终端10接收的回声图像的获取相关的信息。

[0117] 另外，通信部200在与外部装置的通信中，将从控制部210输入的信息发送到外部装置。具体地，通信部200将从控制部210输入的回声图像发送到便携终端10。

[0118] (2-2)控制部210

[0119] 控制部210具有控制超声波回声装置20的全部动作的功能。例如，控制部210控制回声获取部220的回声图像的获取处理。另外，控制部210控制通信部200向便携终端10发送由回声获取部220获取的回声图像的处理。

[0120] (2-3)回声获取部220

[0121] 回声获取部220具有获取回声图像的功能。例如，回声获取部220使用超声波输出部22所具备的超声波输出装置获取回声图像。具体地，回声获取部220使超声波输出装置向用户12的体内输出超声波，并基于被用户12的体内的器官反射的超声波获取回声图像。回声获取部220通过使超声波输出装置从用户12的下颌下方向用户12的口腔内输出超声波，能够获取表示用户12的口腔内的状态的回声图像。

[0122] (3)声音输入输出装置30

[0123] 如图4所示，声音输入输出装置30具有通信部300、控制部310、声音输入部320及声音输出部330。

[0124] (3-1)通信部300

[0125] 通信部300具有与外部装置进行通信的功能。例如，在与外部装置的通信中，通信部300将从外部装置接收的信息输出到控制部310。具体地，通信部300将从便携终端10接收的声音数据输出到控制部310。

[0126] 另外，通信部300在与外部装置的通信中，将从控制部310输入的信息发送到外部装置。具体地，通信部300将从控制部310输入的声音数据发送到便携终端10。

[0127] (3-2)控制部310

[0128] 控制部310具有控制声音输入输出装置30的全部动作的功能。例如，控制部310控制声音输入部320的声音的获取处理。另外，控制部310控制通信部300向便携终端10发送由

声音输入部320获取的声音的处理。另外，控制部310控制声音输出部330的声音的输出处理。例如，通信部300将从便携终端10接收到的声音数据作为声音输出到声音输出部330。

[0129] (3-3)声音输入部320

[0130] 声音输入部320具有获取在外部产生的声音的功能。声音输入部320例如获取用户12发声时的声音即发声声音。另外，声音输入部320将所获取的发声声音输出到控制部310。另外，声音输入部320例如可以通过麦克风实现。

[0131] (3-4)声音输出部330

[0132] 声音输出部330具有输出从外部装置接收到的声音的功能。声音输出部330例如将基于便携终端10中的声音识别处理的结果生成的声音数据从控制部310输入，并输出与所输入的声音数据对应的声音。另外，声音输出部330例如可以通过扬声器实现。

[0133] <1.4.无声发话系统的处理>

[0134] 以上对本实施方式所涉及的无声发话系统1000的功能进行了说明。接着，对无声发话系统1000的处理进行说明。

[0135] (1)获取第一神经网络的机器学习的流程

[0136] 图7是示出本实施方式所涉及的获取第一神经网络的机器学习的流程的流程图。首先，便携终端10从超声波回声装置20获取发声时间序列回声图像作为学习信息(S100)。另外，便携终端10从声音输入输出装置30获取发声声音作为学习信息(S102)。接着，便携终端10使用所获取的学习信息进行机器学习(S104)。另外，便携终端10将通过该机器学习生成的算法设为NN1(S106)。

[0137] (2)获取第二神经网络的机器学习的流程

[0138] 图8是示出本实施方式所涉及的获取第二神经网络的机器学习的流程的流程图。首先，便携终端10向NN1输入发声时间序列回声图像(S200)。接着，便携终端10获取从NN1输出的发声图像语谱图作为学习信息(S202)。另外，便携终端10从发声声音中获取发声声音语谱图作为学习信息(S204)。接着，便携终端10使用所获取的学习信息进行机器学习(S206)。另外，便携终端10将通过该机器学习生成的算法设为NN2(S208)。

[0139] (3)便携终端10中的处理

[0140] 图9是示出本实施方式所涉及的便携终端中的处理的流程的流程图。首先，便携终端10获取无发声时间序列回声图像(S300)。接着，便携终端10将所获取的无发声时间序列回声图像输入到NN1，并根据无发声时间序列回声图像生成多个声音特征量(S302)。接着，便携终端10按时间序列顺序合成所生成的多个声音特征量，生成无发声图像语谱图(S304)。

[0141] 在根据无发声时间序列回声图像生成无发声图像语谱图之后，便携终端10将所生成的无发声图像语谱图输入到NN2，并将无发声图像语谱图转换为高精度无发声图像语谱图(S306)。在转换之后，便携终端10通过识别部114识别高精度无发声图像语谱图表示的内容(S308)。另外，便携终端10执行基于由识别部114识别的内容的处理(S310)。

[0142] <<2.变形例>>

[0143] 以上对本公开的实施方式进行了说明。接着，对本公开的实施方式的变形例进行说明。另外，以下说明的变形例既可以单独应用于本公开的实施方式，也可以组合应用于本公开的实施方式。另外，变形例既可以取代本公开的实施方式中说明的结构来应用，也可以

追加应用于本公开的实施方式中说明的结构。

[0144] 在上述实施方式中，对通过NN2转换的高精度无发声图像语谱图被输出到便携终端10的识别部114的例子进行了说明，但该高精度无发声图像语谱图也可以在被转换为声音波形的基础上，作为声音从扬声器等声音输出装置输出。由此，用户12能够经由声音输出装置控制智能扬声器等带声音输入功能的信息设备。

[0145] 另外，高精度无发声图像语谱图也可以不输出到便携终端10的识别部114，而输出到外部的声音识别装置。例如，该高精度无发声图像语谱图也可以经由通信被输入到智能扬声器的声音识别部。由此，用户12能够在不使便携终端10向空中放射声波的情况下，控制智能扬声器等带声音输入功能的信息设备。

[0146] <<3.应用例>>

[0147] 以上对本公开的实施方式的变形例进行了说明。接着，对本公开的实施方式所涉及的无声发话系统1000的应用例进行说明。

[0148] <3.1.第一应用例>

[0149] 首先，对本实施方式所涉及的第一应用例进行说明。本实施方式所涉及的无声发话系统1000例如可以应用于使说话者不发声地活动口、舌的训练。例如，无声发话系统1000将从超声波回声装置20获取的无发声时间序列回声图像识别出的内容视觉地反馈给说话者。由此，说话者能够基于该反馈来改善口、舌的活动方式。具体地，通过无声发话系统1000在显示装置等上显示无发声时间序列回声图像，说话者能够确认所显示的图像并学习口、舌的活动方式。另外，通过将无声发话系统1000从无发声时间序列回声图像中识别的内容通过声音进行反馈，说话者能够学习在如何活动口、舌时，无声发话系统1000如何识别。此外，无声发话系统1000识别的内容也可以通过文本进行反馈。

[0150] <3.2.第二应用例>

[0151] 接着，对本实施方式所涉及的第二应用例进行说明。本实施方式所涉及的无声发话系统1000可以用作声带缺损的人、听觉障碍者的发声支持设备。近年来，为了失去声带功能的人，提出了与向咽喉按压能够进行按钮控制的振子来代替声带的技术相关的技术。根据该技术，失去了声带功能的人能够在不使声带振动的情况下发出声音。但是，在该技术中，由于振子发出较大的声音，因此可能会阻碍经由口腔内的发话的声音。另外，说话者难以调整该较大的声音的音量，该较大的声音对说话者来说可能会成为不舒服的声音。另一方面，在本实施方式所涉及的无声发话系统1000中，将通过超声波回声而获得的信息转换为声学信息，并将该声学信息作为声音波形发声，因此不会产生阻碍发话的声音、不舒服的声音。另外，说话者还可以调节从无声发话系统1000产生的声音的音量。因此，即使是失去了声带功能的人，也能够更舒适地利用本实施方式所涉及的无声发话系统1000。

[0152] 另外，声带缺损的人不能发出声音，但能够活动口、舌来使口腔内的状态变化。因此，通过由无声发话系统1000识别声带缺损的人的口腔内的状态，并将识别出的内容作为声音从扬声器输出，即使是声带缺损的人，也能够通过声音与他人进行交流。另外，本实施方式所涉及的无声发话系统1000与声带缺损的人无关地，对高龄者等不具有足以使声带振动的充分的肺活量的人也发挥效果。例如，在不能以足够的音量发声的高龄者的情况下，会话可能变得困难，但是该高龄者可以通过无声发话系统1000具有发声能力，从而可以容易地进行会话。

[0153] 另外，听觉障碍者虽然能够发出声音，但是难以确认自身发出的声音是否正确地传递给了他人。因此，通过利用在第一应用例中所述的本实施方式的无声发话系统1000的反馈，听觉障碍者能够确认自身如何发出声音。另外，在无声发话系统1000中，由于能够确认口腔内的状态，因此听觉障碍者能够一边确认口、舌的活动方法一边练习说话方法。

[0154] <3.3.第三应用例>

[0155] 接着，对本实施方式所涉及的第三应用例进行说明。本实施方式所涉及的无声发话系统1000能够应用于助听器的功能的扩充。通过将无声发话系统1000搭载于助听器，能够提高助听器的使用者的便利性。

[0156] <<4.硬件构成例>>

[0157] 最后，参照图10对本实施方式所涉及的信息处理装置的硬件构成例进行说明。图10是示出本实施方式所涉及的信息处理装置的硬件构成例的框图。另外，图10所示的信息处理装置900例如能够实现图1及图4分别所示的便携终端10、超声波回声装置20及声音输入输出装置30。本实施方式所涉及的便携终端10、超声波回声装置20及声音输入输出装置30的信息处理通过软件和以下说明的硬件的协作来实现。

[0158] 如图10所示，信息处理装置900具备CPU(Central Processing Unit:中央处理单元)901、ROM(Read Only Memory:只读存储器)902及RAM(Random Access Memory:随机存取存储器)903。另外，信息处理装置900具备主机总线904a、桥接器904、外部总线904b、接口905、输入装置906、输出装置907、存储装置908、驱动器909、连接端口910及通信装置911。另外，这里所示的硬件结构只是一个例子，也可以省略构成要素的一部分。另外，硬件结构也可以进一步包括这里所示的构成要素以外的构成要素。

[0159] CPU901例如作为运算处理装置或控制装置发挥功能，基于ROM902、RAM903或存储装置908中记录的各种程序来控制各构成要素的全部动作或其一部分。ROM902是存储由CPU901读取的程序、用于运算的数据等的装置。RAM903临时或永久地存储例如CPU901读取的程序、在执行该程序时适当变化的各种参数等。它们通过由CPU总线等构成的主机总线904a相互连接。CPU901、ROM902及RAM903例如通过与软件的协作，能够实现参照图4说明的控制部110、控制部210及控制部310的功能。

[0160] CPU901、ROM902及RAM903例如经由能够进行高速数据传输的主机总线904a相互连接。另一方面，主机总线904a例如经由桥接器904与数据传输速度较低的外部总线904b连接。另外，外部总线904b经由接口905与各种构成要素连接。

[0161] 输入装置906例如通过鼠标、键盘、触摸面板、按钮、麦克风、开关及杆等由用户输入信息的装置来实现。另外，输入装置906例如既可以是利用了红外线、其他电波的远程控制装置，也可以是与信息处理装置900的操作对应的便携电话、PDA等外部连接设备。另外，输入装置906例如也可以包括输入控制电路等，该输入控制电路基于由用户使用上述输入装置输入的信息生成输入信号，并输出到CPU901。信息处理装置900的用户通过操作该输入装置906，能够向信息处理装置900输入各种数据或者指示处理动作。

[0162] 另外，输入装置906可以由检测与用户相关的信息的装置而形成。例如，输入装置906可以包括图像传感器(例如，照相机)、深度传感器(例如，立体照相机)、加速度传感器、陀螺仪传感器、地磁传感器、光传感器、声音传感器、测距传感器(例如，ToF(Time of Flight:飞行时间)传感器)、力传感器等各种传感器。另外，输入装置906也可以获取信息处

理装置900的姿态、移动速度等与信息处理装置900自身的状态相关的信息、以及信息处理装置900周围的亮度、噪声等与信息处理装置900的周围环境相关的信息。另外，输入装置906也可以包括GNSS模块，该GNSS模块接收来自GNSS(Global Navigation Satellite System:全球导航卫星系统)卫星的GNSS信号(例如，来自GPS(Global Positioning System:全球定位系统)卫星的GPS信号)，并测定包括装置的纬度、经度及高度的位置信息。另外，关于位置信息，输入装置906也可以是通过与Wi-Fi(注册商标)、便携电话·PHS·智能手机等的信息的收发，或者近距离通信等来检测位置的装置。输入装置906例如能够实现参照图4说明的回声获取部220及声音输入部320的功能。

[0163] 输出装置907由能够通过视觉或听觉向用户通知所获取的信息的装置形成。作为这样的装置，有CRT显示装置、液晶显示装置、等离子显示装置、EL显示装置、激光投影仪、LED投影仪及灯等显示装置、扬声器及耳机等声音输出装置、打印装置等。输出装置907例如输出通过信息处理装置900进行的各种处理而得到的结果。具体地，显示装置以文本、图像、表、图表等各种形式视觉地显示通过信息处理装置900进行的各种处理而得到的结果。另一方面，声音输出装置将由再现的声音数据、声学数据等构成的音频信号转换为模拟信号并听觉地输出。输出装置907例如能够实现参照图4说明的声音输出部330的功能。

[0164] 存储装置908是作为信息处理装置900的存储部的一例而形成的数据存储用的装置。存储装置908例如由HDD等磁存储设备、半导体存储设备、光存储设备或光磁存储设备等实现。存储装置908也可以包括存储介质、将数据记录在存储介质中的记录装置、从存储介质读取数据的读取装置、以及删除记录在存储介质中的数据的删除装置等。该存储装置908存储CPU901执行的程序、各种数据以及从外部获取的各种数据等。存储装置908例如能够实现参照图4说明的存储部120的功能。

[0165] 驱动器909是存储介质用读写器，被内置或外置在信息处理装置900中。驱动器909读取记录在所安装的磁盘、光盘、磁光盘或半导体存储器等可移动存储介质中的信息，并输出到RAM903。另外，驱动器909也可以向可移动存储介质写入信息。

[0166] 连接端口910例如是用于连接USB(Universal Serial Bus:通用串行总线)端口、IEEE1394端口、SCSI(Small Computer System Interface:小型计算机系统接口)、RS-232C端口或光音频端子等外部连接设备的端口。

[0167] 通信装置911例如是由用于与网络920连接的通信设备等形成的通信接口。通信装置911例如是有线或无线LAN(Local Area Network:局域网)、LTE(Long Term Evolution:长期演进)、Bluetooth(注册商标)或WUSB(Wireless USB:无线USB)用的通信卡等。另外，通信装置911也可以是光通信用的路由器、ADSL(Asymmetric Digital Subscriber Line:非对称数字用户线路)用的路由器或各种通信用的调制解调器等。该通信装置911例如能够在与因特网、其他通信设备之间，按照例如TCP/IP等规定的协议收发信号等。通信装置911例如能够实现参照图4说明的通信部100、通信部200及通信部300的功能。

[0168] 另外，网络920是从与网络920连接的装置发送的信息的有线或无线传输路径。例如，网络920也可以包括因特网、电话线路网、卫星通信网等公共线路网，包括Ethernet(注册商标)的各种LAN(Local Area Network:局域网)、WAN(Wide Area Network:广域网)等。另外，网络920也可以包括IP-VPN(Internet Protocol-Virtual Private Network:因特网协议-虚拟专用网络)等专用线路网。

[0169] 以上示出了能够实现本实施方式所涉及的信息处理装置900的功能的硬件结构的一例。上述各构成要素既可以使用通用的部件来实现，也可以通过专用于各构成要素的功能的硬件来实现。因此，能够根据实施本实施方式时的技术水平，适当地变更所利用的硬件结构。

[0170] <<5.总结>>

[0171] 如上所述，本实施方式所涉及的便携终端10基于通过机器学习而获得的算法，将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与口腔内的状态对应的信息。由此，便携终端10能够将表示在用户不发出声音的情况下有意地活动口或舌中的至少一方时的口腔内的状态的图像转换为声学信息。

[0172] 因此，能够提供一种能够在用户不发声的情况下得到期望的声学信息的新颖且改进的信息处理装置及信息处理方法。

[0173] 以上参考附图对本公开的优选实施方式进行了详细说明，但是本公开的技术范围并不限于该例。只要是具有本公开的技术领域的通常的知识的技术人员，则在权力要求书所记载的技术思想的范围内，当然可以想到各种变更例或修正例，当然也可以理解这些都属于本公开的技术范围。

[0174] 例如，在本说明书中说明的各装置既可以作为单独的装置实现，也可以一部分或全部作为不同的装置实现。例如，图1所示的便携终端10、超声波回声装置20及声音输入输出装置30也可以分别作为单独的装置实现。另外，例如便携终端10也可以作为通过网络等与超声波回声装置20及声音输入输出装置30连接的服务器装置实现。另外，也可以是通过网络等连接便携终端10所具有的控制部110的功能的服务器装置所具有的结构。

[0175] 另外，在本说明书中说明的各装置的一系列处理也可以使用软件、硬件、以及软件和硬件的组合中的任一个来实现。构成软件的程序例如预先存储在设置在各装置的内部或外部的记录介质(非临时介质:non-transitory media)中。另外，各程序例如在计算机执行时被读入RAM，由CPU等处理器执行。

[0176] 另外，在本说明书中使用流程图说明的处理也可以不一定按照图示的顺序执行。也可以并行地执行一些处理步骤。另外，既可以采用追加的处理步骤，也可以省略一部分处理步骤。

[0177] 另外，本说明书中记载的效果只不过是说明或例示的效果，并不特别限定。即，本公开所涉及的技术，除了上述效果之外，或者取代上述效果，也能够发挥本领域技术人员根据本说明书的记载当然可以想到的其他效果。

[0178] 另外，以下那样的结构也属于本公开的技术范围。

[0179] (1)一种信息处理装置，具备控制部，所述控制部基于通过机器学习而获得的算法，将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。

[0180] (2)根据所述(1)中记载的信息处理装置，所述算法具有第一神经网络，所述控制部经由所述第一神经网络将所输入的无发声时的多张时间序列图像转换为第一声学信息。

[0181] (3)根据所述(2)中记载的信息处理装置，所述第一神经网络根据所输入的所述无发声时的多张时间序列图像生成多个每单位时间的声学特征量，并通过按时间序列顺序合成所生成的多个所述声学特征量来生成所述第一声学信息。

[0182] (4)根据所述(3)中记载的信息处理装置,所述第一神经网络从在所述单位时间获取的所述无发声时的多张时间序列图像中选择所述单位时间的中央时刻的时间序列图像,并根据所选择的时间序列图像生成所述每单位时间的声学特征量。

[0183] (5)根据所述(2)至(4)中任一项记载的信息处理装置,所述第一神经网络是通过使用了第一学习信息的所述机器学习而得到的,该第一学习信息包括发声时的声音和所述发声时的多张时间序列图像。

[0184] (6)根据所述(2)至(5)中任一项记载的信息处理装置,所述算法还具有第二神经网络,所述控制部经由所述第二神经网络将所述第一声学信息转换为与发声时的声音对应的第二声学信息。

[0185] (7)根据所述(6)中记载的信息处理装置,所述第二神经网络是通过使用了第二学习信息的所述机器学习而得到的,该第二学习信息包括通过将所述发声时的多张时间序列图像输入所述第一神经网络而生成的第三声学信息和与发声时的声音对应的第四声学信息。

[0186] (8)根据所述(2)至(7)中任一项记载的信息处理装置,所述声学信息是语谱图。

[0187] (9)根据所述(1)至(8)中任一项记载的信息处理装置,所述多张时间序列图像表示在用户不发声的情况下活动口或舌中的至少一方时的所述口腔内的状态的变化。

[0188] (10)根据所述(1)至(9)中任一项记载的信息处理装置,所述机器学习是通过深度学习进行的。

[0189] (11)根据所述(1)至(10)中任一项记载的信息处理装置,所述机器学习是使用卷积神经网络进行的。

[0190] (12)一种信息处理方法,由处理器执行,包括基于通过机器学习而获得的算法,将通过超声波回声而获得的表示口腔内的状态的多张时间序列图像转换为与所述口腔内的状态对应的信息。

[0191] 附图标记说明

[0192] 10便携终端;20超声波回声装置;30声音输入输出装置;100通信部;110控制部;112机器学习部;114识别部;116处理控制部;120存储部;122第一神经网络;124第二神经网络;200通信部;210控制部;220回声获取部;300通信部;310控制部;320声音输入部;330声音输出部;1000无声发话系统。

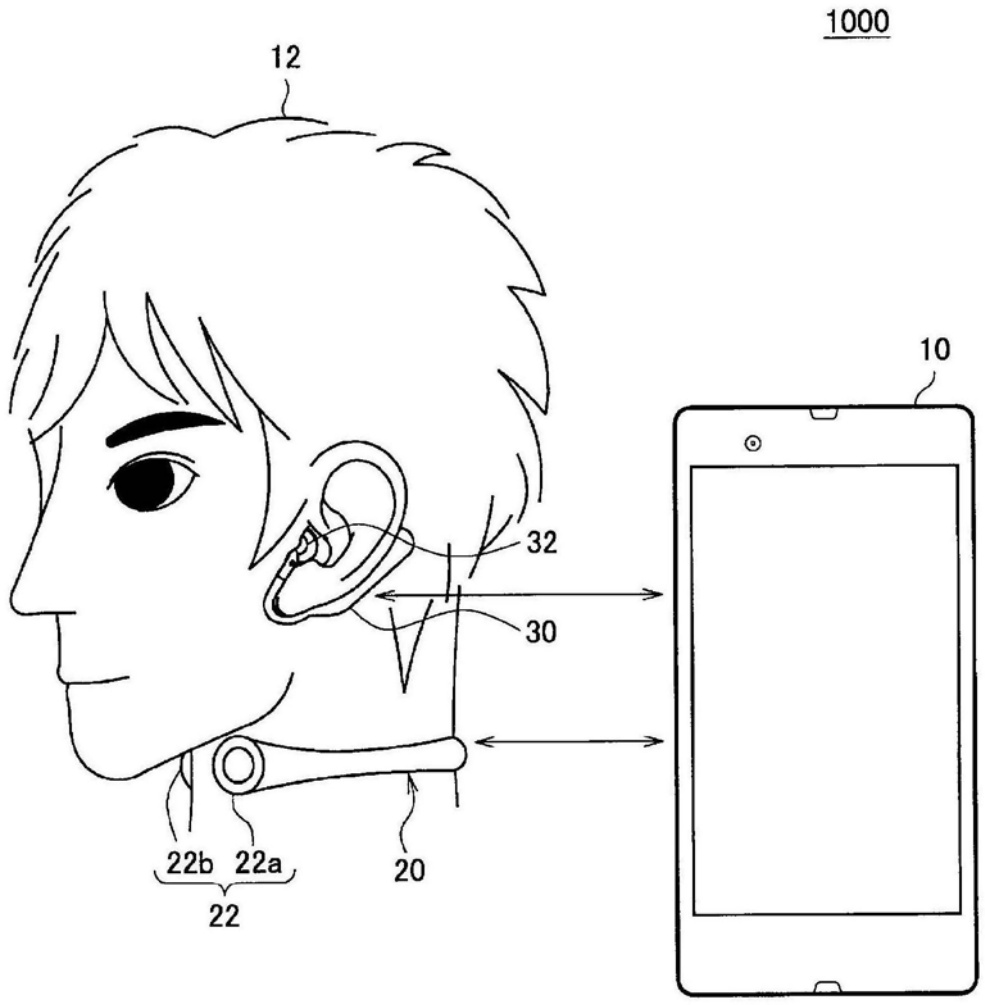



图1

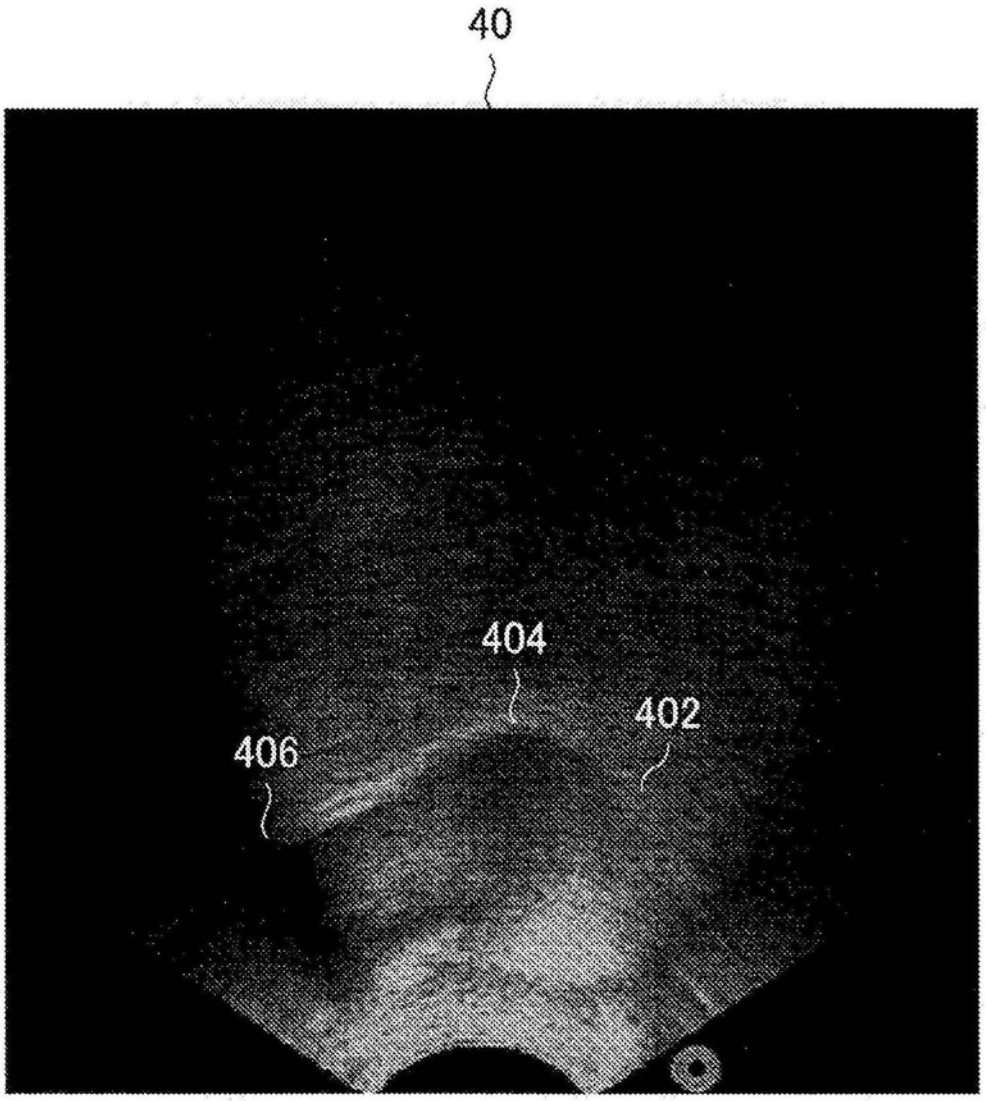



图2

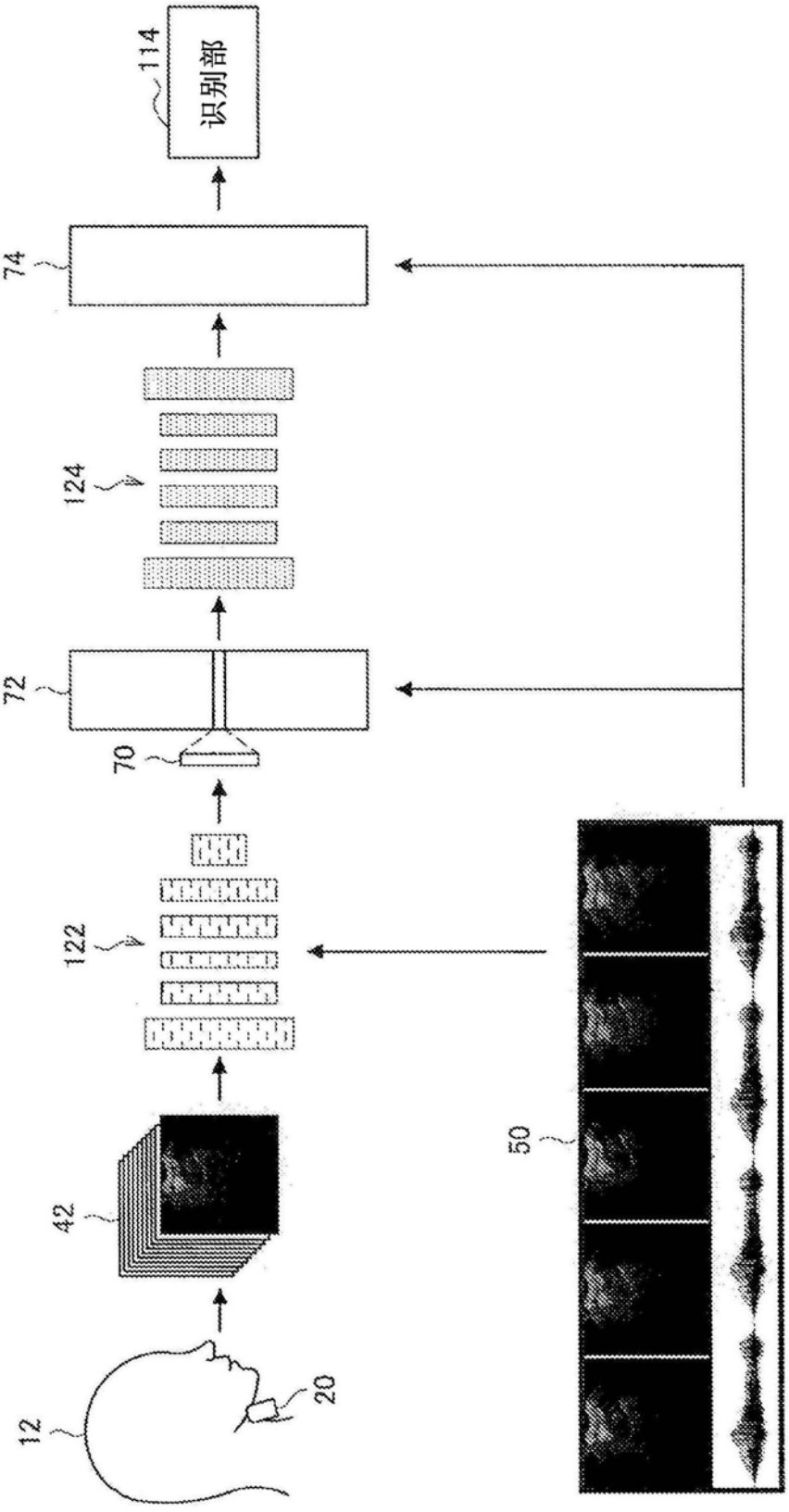



图3

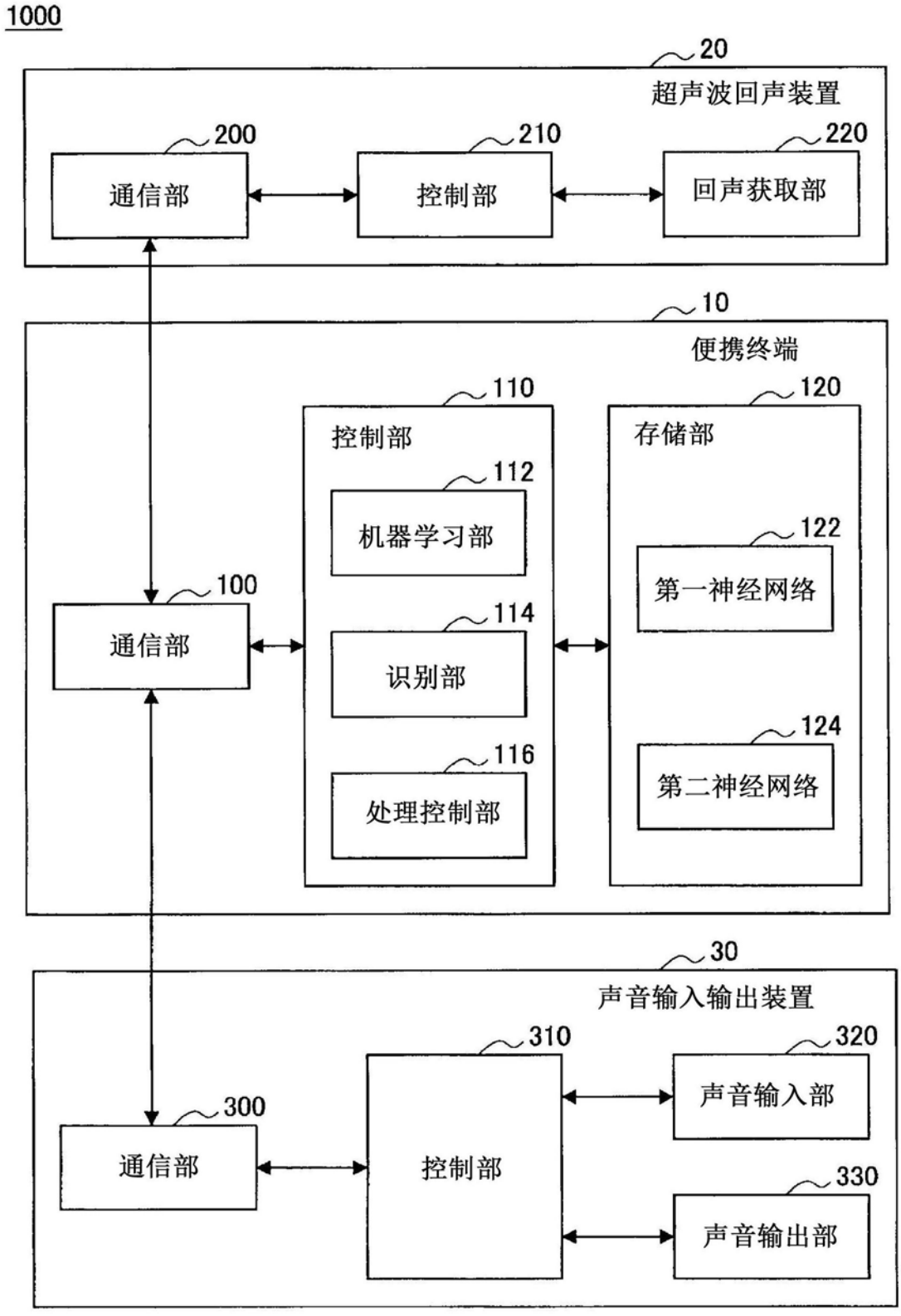

1000
20
超声波回声装置
200
通信部
210
控制部
220
回声获取部
10
便携终端
110
控制部
112
机器学习部
114
识别部
116
处理控制部
120
存储部
122
第一神经网络
124
第二神经网络
100
通信部
30
声音输入输出装置
310
控制部
320
声音输入部
330
声音输出部
300
通信部

图4

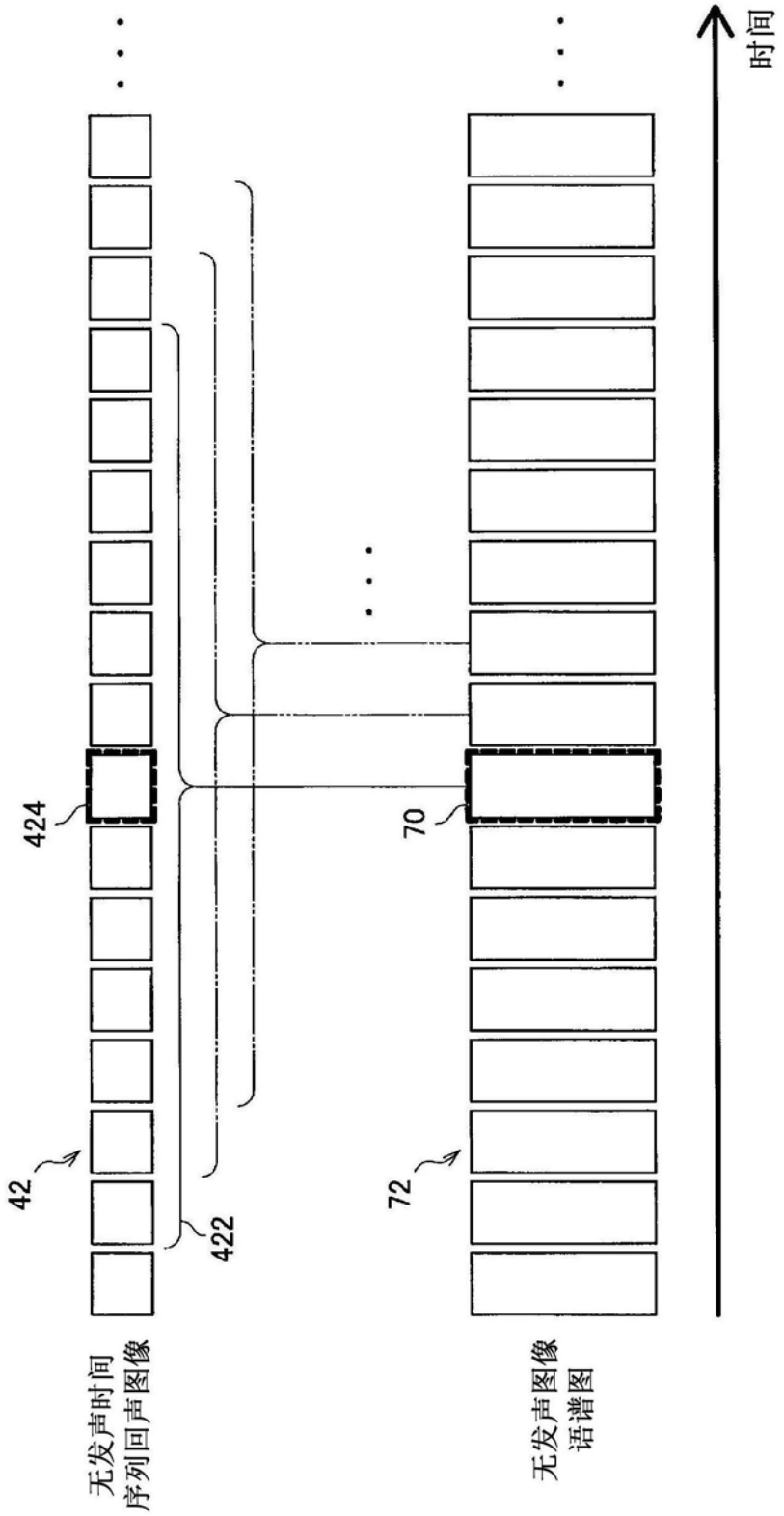



图5

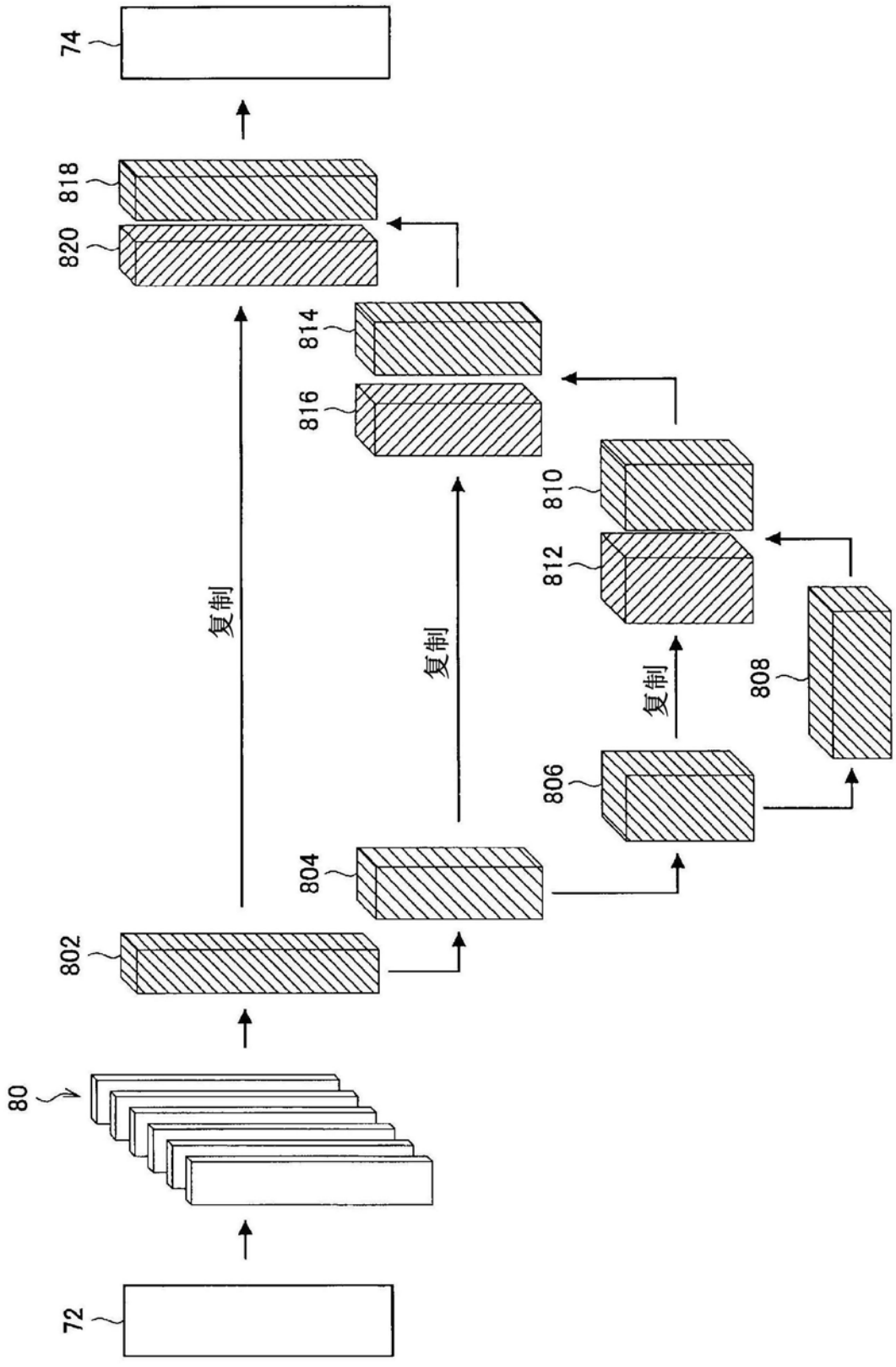



图6

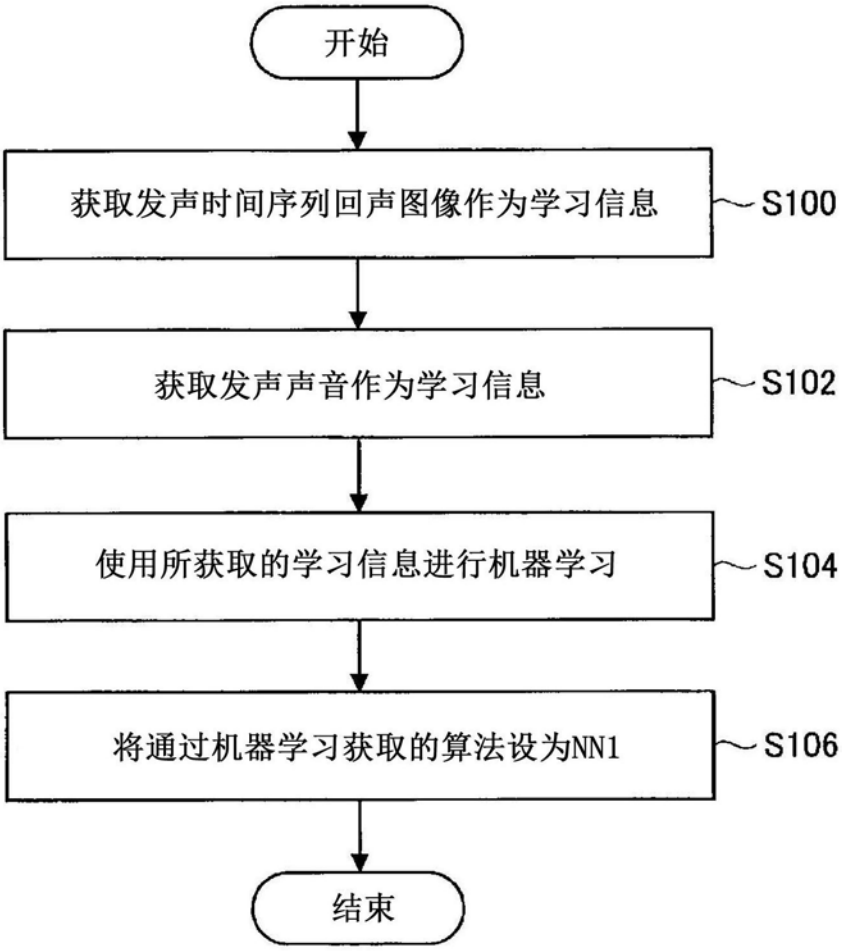

开始
获取发声时间序列回声图像作为学习信息
S100
获取发声声音作为学习信息
S102
使用所获取的学习信息进行机器学习
S104
将通过机器学习获取的算法设为NN1
S106
结束

图7

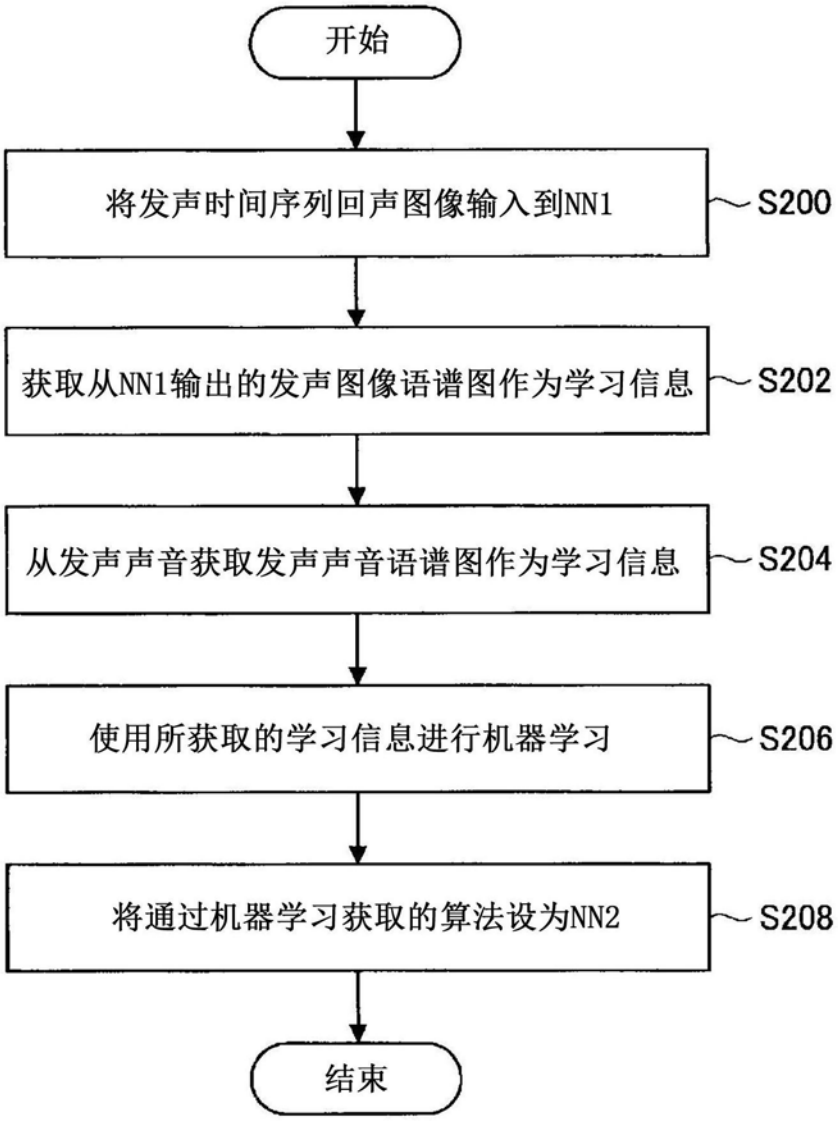

开始
将发声时间序列回声图像输入到NN1
S200
获取从NN1输出的发声图像语谱图作为学习信息
S202
从发声声音获取发声声音语谱图作为学习信息
S204
使用所获取的学习信息进行机器学习
S206
将通过机器学习获取的算法设为NN2
S208
结束

图8

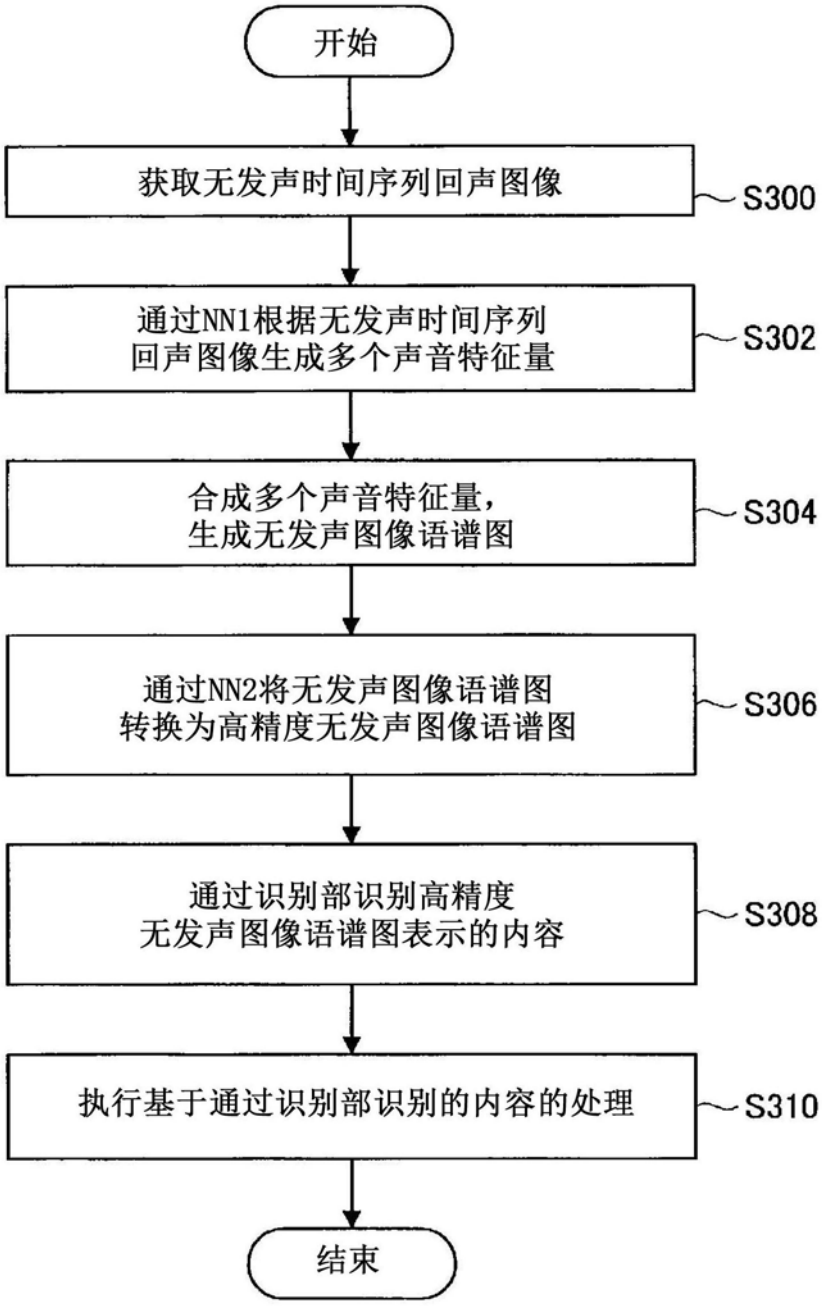

开始
获取无发声时间序列回声图像
S300
通过NN1根据无发声时间序列
回声图像生成多个声音特征量
S302
合成多个声音特征量，
生成无发声图像语谱图
S304
通过NN2将无发声图像语谱图
转换为高精度无发声图像语谱图
S306
通过识别部识别高精度
无发声图像语谱图表示的内容
S308
执行基于通过识别部识别的内容的处理
S310
结束

图9

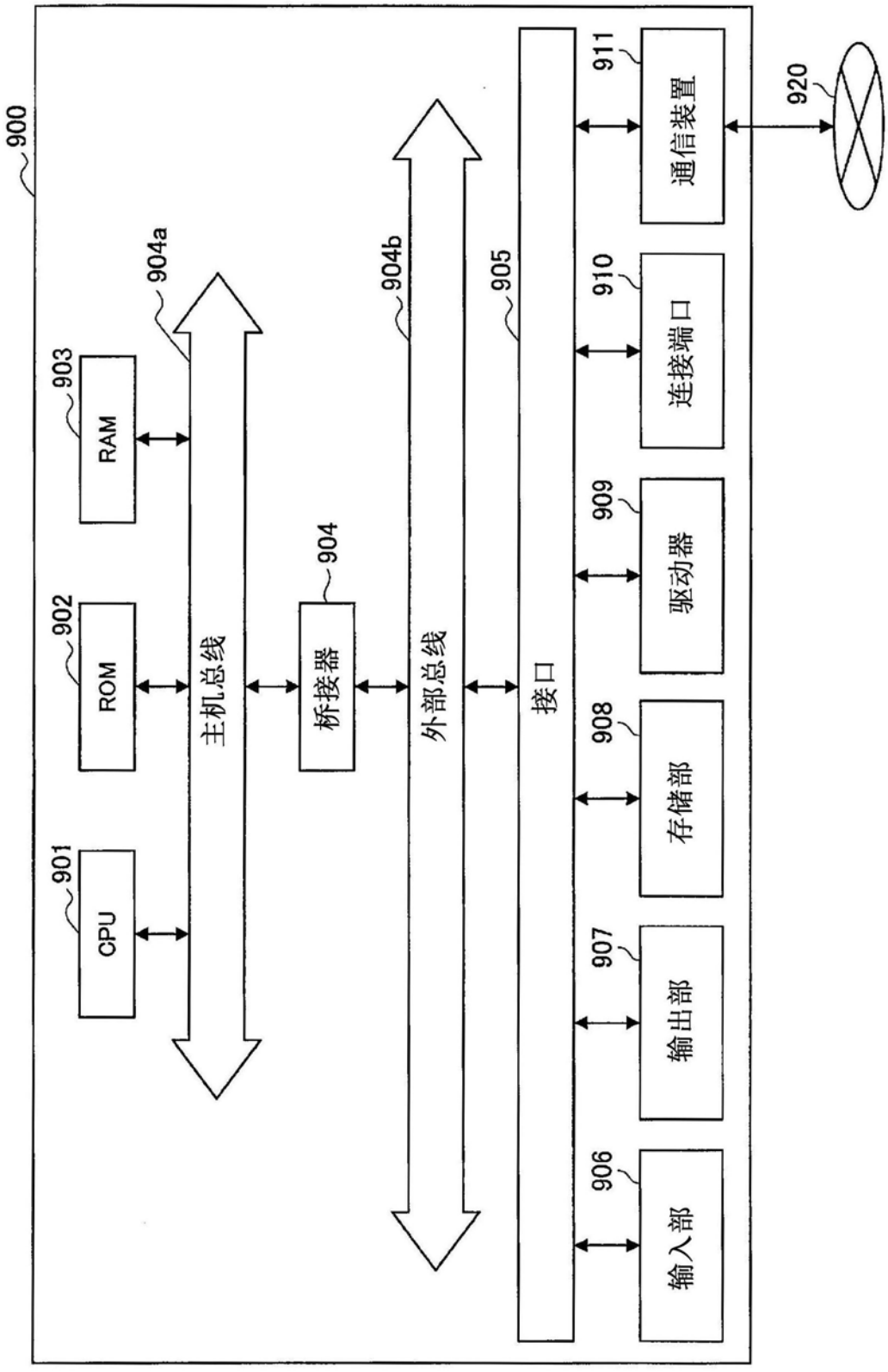

900
CPU
901
ROM
902
RAM
903
主机总线
904a
桥接器
904
外部总线
904b
接口
905
输入部
906
输出部
907
存储部
908
驱动器
909
连接端口
910
通信装置
911
920

图10