(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 7 月 2 日 (02.07.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/135067 A1

(51) 国际专利分类号:
G10L 17/22 (2013.01)

(21) 国际申请号: PCT/CN2019/124844

(22) 国际申请日: 2019 年 12 月 12 日 (12.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811581336.6 2018年12月24日 (24.12.2018) CN

(71) 申请人: 同方威视技术股份有限公司 (NUCTECH COMPANY LIMITED) [CN/CN]; 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。

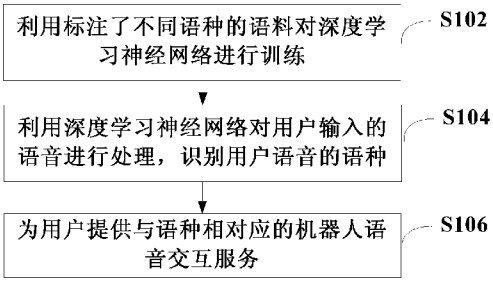
(72) 发明人: 崔锦 (CUI, Jin); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。 谈华斌 (TAN, Huabin); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。 吉东旭 (JI, Dongxu); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。 胡斌 (HU, Bin); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。 林东 (LIN, Dong); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。 乔光辉 (QIAO, Guanghui); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。 司丽娟 (SI, Lijuan); 中国北京市海淀区双清路同方大厦A座2层, Beijing 100084 (CN)。

(74) 代理人: 中国国际贸易促进委员会专利商标事务所 (CCPIT PATENT AND TRADEMARK LAW OFFICE); 中国北京市西城区阜成门外大街2号万通新世界广场8层, Beijing 100037 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** VOICE INTERACTION METHOD AND DEVICE, ROBOT, AND COMPUTER READABLE STORAGE MEDIUM

(54) **发明名称**: 语音交互方法、装置、机器人及计算机可读存储介质



S102 Train a deep learning neural network using corpora annotated with different languages
S104 Process, using a deep learning neural network, a voice input by a user to recognize the language of the voice
S106 Provide to the user a robot voice interaction service corresponding to the language

图 1

(57) **Abstract:** A voice interaction method and device, a robot, and a computer readable storage medium, relating to the technical field of robots. The voice interaction method comprises: processing, using a deep learning neural network, a voice input by a user to recognize the language of the voice (S104); and providing to the user a robot voice interaction service corresponding to the language (S106). A robot is made to be able to provide to a user a voice interaction service corresponding to the language of the user, thereby improving user experience of a robot voice interaction service.

(57) **摘要**: 一种语音交互方法、装置、机器人以及计算机可读存储介质，涉及机器人技术领域。其中的语音交互方法包括：利用深度学习神经网络对用户输入的语音进行处理，识别语音的语种（S104）；为用户提供与语种相对应的机器人语音交互服务（S106）。使机器人能够为用户提供与用户的语种相对应的语音交互服务，提升了机器人语音交互服务的用户体验。

WO 2020/135067 A1

[见续页]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) **指定国**(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

语音交互方法、装置、机器人及计算机可读存储介质

相关申请的交叉引用

本申请是以 CN 申请号为 201811581336.6，申请日为 2018 年 12 月 24 日的申请为基础，并主张其优先权，该 CN 申请的公开内容在此作为整体引入本申请中。

技术领域

本公开涉及机器人技术领域，特别涉及一种语音交互方法、装置、机器人及计算机可读存储介质。

背景技术

人机交互方式的演变过程越来越贴近于人类的本能表达。人机交互方式中，语音交互是人与机器人之间更加自然的交互手段，语音交互简单快速，且能解放人的双手和眼睛，在很多场景能够给用户带来更好的体验。

此外，通过语音交互过程，机器人后台还可以获取用户数据，为后续的增值服务打下基础。由此可见，为用户提供更好的语音交互服务，是语音机器人领域的重要工作目标。

发明内容

本公开解决的一个技术问题是，如何使机器人为用户提供与用户的语种相对应的语音交互服务。

根据本公开实施例的一个方面，提供了一种语音交互方法，包括：利用深度学习神经网络对用户输入的语音进行处理，识别语音的语种；为用户提供与语种相对应的机器人语音交互服务。

在一些实施例中，还包括：利用标注了不同语种的语料对深度学习神经网络进行训练，使深度学习神经网络能够识别用户输入的语音的语种。

在一些实施例中，还包括：判断机器人当前的网络连接状态；若当前的网络连接状态为在线状态，则将用户输入的语音发送至云端服务器进行语音识别、语义理解和语音合成，接收云端反馈的第一输出语音；同时，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；若先接收到

第一输出语音后得到第二输出语音，则向用户播放第一输出语音；若先得到第二输出语音后接收到第一输出语音，则向用户播放第二输出语音。

在一些实施例中，还包括：若当前的网络状态为离线状态，则将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；向用户播放第二输出语音。

在一些实施例中，进行语音识别时，根据用户输入的语音中各个字之间的时间间隔和语音能量对用户输入的语音进行断句。

在一些实施例中，进行语音识别时，采用隐马尔科夫模型对用户输入的语音进行断句。

在一些实施例中，还包括：在机器人休眠状态下，利用摄像头对人脸进行实时识别；响应于摄像头识别到人脸，将机器人从休眠状态唤醒至工作状态。

根据本公开实施例的另一个方面，提供了一种语音交互机器人，包括前端处理芯片，被配置为：利用深度学习神经网络对用户输入的语音进行处理，识别语音的语种；为用户提供与语种相对应的机器人语音交互服务。

在一些实施例中，深度学习神经网络经过标注了不同语种的语料的训练，能够识别用户输入的语音的语种。

在一些实施例中，还包括上位机，被配置为：判断机器人当前的网络连接状态；若当前的网络连接状态为在线状态，则将用户输入的语音发送至云端服务器进行语音识别、语义理解和语音合成，接收云端反馈的第一输出语音；同时，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；若先接收到第一输出语音后得到第二输出语音，则向用户播放第一输出语音；若先得到第二输出语音后接收到第一输出语音，则向用户播放第二输出语音。

在一些实施例中，上位机还被配置为：若当前的网络状态为离线状态，则将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；向用户播放第二输出语音。

在一些实施例中，上位机被还配置为：根据用户输入的语音中各个字之间的时间间隔和语音能量对用户输入的语音进行断句。

在一些实施例中，上位机还被配置为：采用隐马尔科夫模型对用户输入的语音进行断句。

在一些实施例中，还包括摄像头，被配置为：在机器人休眠状态下对人脸进行实

时识别；在识别到人脸的情况下，将机器人从休眠状态唤醒至工作状态。

在一些实施例中，还包括麦克风阵列，麦克风阵列的多个咪头位于相同水平面的同一圆周上，且相邻咪头之间的圆周距离相等。

在一些实施例中，麦克风阵列上套有硅胶套，硅胶套与语音机器人的外壳固定连接。

在一些实施例中，还包括扩音器和多个腔体；其中，扩音器和麦克风阵列设置在语音交互机器人的不同腔体内。

在一些实施例中，还包括包裹在麦克风阵列上的隔音棉。

根据本公开实施例的又一个方面，提供了一种语音交互装置，包括：存储器；以及耦接至存储器的处理器，处理器被配置为基于存储在存储器中的指令，执行前述的语音交互方法。

根据本公开实施例的再一个方面，提供了一种计算机可读存储介质，其中，计算机可读存储介质存储有计算机指令，指令被处理器执行时实现前述的语音交互方法。

本公开能够使机器人能够为用户提供与用户的语种相对应的语音交互服务，提升了机器人语音交互服务的用户体验。

通过以下参照附图对本公开的示例性实施例的详细描述，本公开的其它特征及其优点将会变得清楚。

附图说明

为了更清楚地说明本公开实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本公开的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 **1** 示出了本公开一些实施例的语音交互方法的流程示意图。

图 **2** 示出了本公开另一些实施例的语音交互方法的流程示意图。

图 **3** 示出了本公开又一些实施例的语音交互方法的流程示意图。

图 **4** 示出了本公开一些实施例的语音交互机器人的结构示意图。

图 **5** 示出了前端处理芯片的结构示意图。

图 **6** 示出了本公开一些实施例的语音交互装置的结构示意图。

具体实施方式

下面将结合本公开实施例中的附图，对本公开实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本公开一部分实施例，而不是全部的实施例。以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本公开及其应用或使用的任何限制。基于本公开中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例，都属于本公开保护的范围。

首先结合图 1 描述本公开语音交互方法的一些实施例，以解释本公开如何使机器人为用户提供与用户的语种相对应的语音交互服务。

图 1 示出了本公开一些实施例的语音交互方法的流程示意图。如图 1 所示，本实施例包括步骤 S102～步骤 S116。

在步骤 S102 中，利用标注了不同语种的语料对深度学习神经网络进行训练，使深度学习神经网络能够识别用户输入的语音的语种。

例如，可以使用大量的英文语料以及中文语料对深度学习神经网络进行训练，使深度学习神经网络能够准确识别用户输入的语音为中文语音或英文语音。

在步骤 S104 中，利用深度学习神经网络对用户输入的语音进行处理，识别用户语音的语种。

例如，利用深度学习神经网络对用户输入的语音进行处理，可以识别用户输入的语音为中文语音或英文语音。

在步骤 S106 中，为用户提供与语种相对应的机器人语音交互服务。

若识别到用户输入的语音为中文语音，则调用中文语音交互平台（例如科大讯飞等语音厂商的语音交互平台）中的语音识别、语义理解、语音合成等语音算法，为用户提供中文语音交互服务；若识别到用户输入的语音为英文语音，则调用语音交互平台（例如亚马逊 Alexa 语音服务平台）中的语音识别、语义理解、语音合成等语音算法，为用户提供英文语音交互服务。

上述实施例提供的语音交互方法，能够使语音机器人在复杂的语言场景下提供与用户语种相对应的语音交互服务，使语音机器人能够准确的与用户进行中英文语音交互，提升了机器人语音交互服务的用户体验。

对于调用的语音交互平台，可以在语音识别功能上进行改进。现有的语音交互平台在将语音转化为文本时，需要对文本进行断句以便于语义理解。为了使得断句更加准确，可以采用如下三个方面的技术手段。

（1）根据语音中各个字之间的时间间隔和语音能量对语音进行断句。

例如，在检测到字与字之间的时间间隔大于预设时长的情况下进行断句。或者，在检测到语音能量小于预设能量值（趋近于零）的情况下进行断句。

（2）根据用户的意图对语音进行断句。

例如，可以采用关键词库来提取语音“上海大学城书店在哪里”中的关键词。提取出的关键词“上海”带有地点意图属性，“在”带有位置意图属性，“哪里”带有询问意图属性。根据各个关键词的意图属性，可以得到语音的意图属性为询问位置。然后，利用预先训练的与询问位置意图属性对应的神经网络对语音进行断句。

（3）直接通过深度学习神经网络语音进行断句。

例如，采用海量的语料对隐马尔科夫模型 CNN-HMM 进行训练，可以使训练好的隐马尔科夫模型 CNN-HMM 具有直接对语音进行断句的功能。

本领域技术人员应理解，如果将上述三个方面的技术手段组合使用，能够更准确的对用户输入的语音进行断句，从而进一步提升了机器人语音交互服务的用户体验。

发明人研究发现，语音识别和语义理解时需要借助语音知识库。由于本地离线知识库中的语料较少，若仅采用本地离线知识库，虽然不依赖网络，但识别效果和语义分析能力相对较差，语音识别正确率较低；若仅采用云端在线知识库，虽然语料较多且识别效果和语义分析能力相对较强，但则过于依赖网络环境，而机器人使用场景网络条件相对复杂，语音机器人工作时很难应对网络信号不稳定和网络延迟大的情况。有鉴于此，发明人提供了一种离在线结合的方式，对语音进行识别和理解。

下面结合图 2 描述本公开语音交互方法的另一些实施例，以解释本公开如何使机器人通过离在线结合的方式为用户提供语音交互服务。

图 2 示出了本公开另一些实施例的语音交互方法的流程示意图。如图 2 所示，本实施例包括步骤 S201～步骤 S205。

在步骤 S201 中，判断机器人当前的网络连接状态。

例如，可以通过向特定网络端口发送 ping 消息。若收到恢复消息则当前的网络连接状态为在线状态，执行步骤 S202；若未收到回复消息则当前的网络状态为离线状态，执行步骤 S204。

在步骤 S202 中，将用户输入的语音发送至云端服务器进行语音识别、语义理解和语音合成，接收云端反馈的第一输出语音；同时，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音。

在步骤 S203 中，向用户播放第一输出语音与第二输出语音当中首先获得的输出语音。

若先接收到第一输出语音后得到第二输出语音，则向用户播放第一输出语音；若先得到第二输出语音后接收到第一输出语音，则向用户播放第二输出语音。

在步骤 S204 中，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音。

在步骤 S205 中，向用户播放第二输出语音。

上述实施例提供了离在线结合的方式对语音进行识别和理解。该实施例能够在网络条件较差的情况下使语音机器人能够提供语音交互服务，在网络条件较号的情况下使语音机器人能够提供更加准确的语音交互服务，更进一步的提升了机器人语音交互服务的用户体验。

下面结合图 3 描述本公开语音交互方法的又一些实施例，以解释本公开如何实现机器人的人脸唤醒功能。

图 3 示出了本公开又一些实施例的语音交互方法的流程示意图。如图 3 所示，本实施例包括步骤 S302～步骤 S304。

在步骤 S302 中，在机器人休眠状态下，利用摄像头对人脸进行实时识别。

若识别到人脸，则执行步骤 S304；若未识别到人脸，则返回步骤 S301。

在步骤 S304 中，将机器人从休眠状态唤醒至工作状态。

例如，在摄像头识别到人脸的情况下，调用系统的唤醒程序，使机器人从休眠状态唤醒至工作状态。

上述实施例中，通过摄像头与系统唤醒程序的配合，实现了语音机器人的人脸唤醒功能，不需要用户提供唤醒词对语音机器人进行唤醒，仅需靠近摄像头即可与语音机器人进行语音交互。

下面结合图 4 描述本公开语音交互机器人的一些实施例，以说明本公开语音交互机器人的硬件架构。

图 4 示出了本公开一些实施例的语音交互机器人的结构示意图。如图 4 所示，本实施例中的语音交互机器人 40 包括：麦克风 401、前端处理芯片 402、上位机 403、扩音器 404。其中，前端处理芯片 402 被配置为：利用深度学习神经网络对用户输入的语音进行处理，识别语音的语种；为用户提供与语种相对应的机器人语音交互服务。其中，深度学习神经网络经过标注了不同语种的语料的训练，能够识别用户输入的语

音的语种。

下面结合图 5 说明前端处理芯片 402 的如何实现。图 5 示出了前端处理芯片的结构示意图。如图 5 所示，前端处理芯片中的软件程序可以基于 Android 平台进行开发，具体包括语音识别模块 5021、语音代理模块 5022、语音侦听模块 5023、语音服务模块 5024、控制服务模块 5025 以及连接服务模块 5026。各模块功能如下：

语音识别模块 5021，用于对接收到的语音进行语种识别，并根据识别的语种类别调用不同的语音交互服务；

语音代理模块 5022，用于向语音服务模块 5024 发送语音服务消息；

语音侦听模块 5023，用于从语音服务模块 5024 获取反馈的语音服务消息；

语音服务模块 5024，用于实现基础的语音算法服务，包括语音识别、语义理解和语音合成等；

控制服务模块 5025，用于根据语音指令实现语音机器人的控制服务；

控制服务模块 5026，用于根据语音指令实现语音机器人与外部设备的连接控制服务。

从图 5 可以看出，前端处理芯片中各个工作模块之间采用一套接口即可实现语种识别功能以及语音服务调用功能，整个语音机器人仅需一套软硬件即可提供多语种的语音交互服务。

在一些实施例中，上位机 403 被配置为：判断机器人当前的网络连接状态；若当前的网络连接状态为在线状态，则将用户输入的语音发送至云端服务器进行语音识别、语义理解和语音合成，接收云端反馈的第一输出语音；同时，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；若先接收到第一输出语音后得到第二输出语音，则向用户播放第一输出语音；若先得到第二输出语音后接收到第一输出语音，则向用户播放第二输出语音。

在一些实施例中，上位机 403 还被配置为：若当前的网络状态为离线状态，则将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；向用户播放第二输出语音。本领域技术人员应理解，本地知识库也可以位于上位机 403 中。

在一些实施例中，语音交互机器人 40 还包括摄像头 405，被配置为：在机器人休眠状态下对人脸进行实时识别；在识别到人脸的情况下，将机器人从休眠状态唤醒至工作状态。

上位机 403 不仅可以通过串口与前端处理芯片 402 通信，还可以与麦克风 401、扩音器 404、摄像头 405、检测传感器 406 连接。其中，上位机 403 具有以下语音应用功能：

（1）从麦克风 401 获取用户输入的语音后将语音转发至前端处理芯片 402 进行语音处理；若用户输入的语音为运动控制指令，上位机 403 可以对语音机器人 40 进行运动控制；若用户输入的语音为状态控制指令，上位机 403 可以对语音机器人 40 的工作状态进行调整，如语音服务初始化、唤醒/休眠、启动/停止语音交互、开始/停止录音等等；

（2）从摄像头 405 获取接收的图像，借助人脸识别算法对图像进行识别，在识别到人脸的情况下唤醒语音机器人；

（3）从检测传感器 406 接收检测结果，并通过扩音器 404 对检测结果进行播报；

在一些实施例中，语音交互机器人 40 还包括终端设备 407。终端设备 407 例如可以为平板电脑，可以通过 TCP/IP 协议与前端处理芯片 402 通信，以便用户通过终端设备调用语音机器人的语音交互服务。

在一些实施例中，麦克风 401 为麦克风阵列。麦克风 401 的设计标准如下。

（1）麦克风阵列的多个咪头位于相同水平面的同一圆周上，且相邻咪头之间的圆周距离相等，使得人声直达声到达麦克风阵列中各个咪头的机会均等。

（2）麦克风阵列的麦克风板的法线方向与语音机器人的正面方向一致。

（3）麦克风外侧没有遮挡，麦克风与声源之间没有阻挡，声源到达麦克风的路径尽可能短且宽。

（4）麦克风阵列上套有硅胶套，硅胶套与语音机器人的外壳固定连接。

例如，采用硅胶套对麦克风进行减震密封处理，并使用胶水将硅胶套固定在机器人外壳上。可以使麦克风远离干扰和震动（包括喇叭震动、结构转动震动等等）。

（5）扩音器和麦克风设置在语音机器人的不同腔体内。例如，将麦克风放置在语音机器人的头部，使用隔音棉包裹隔音；将扩音器放在语音机器人的腹部，使扩音器声音不能在结构内泄露到麦克风。

上述实施例中，通过对麦克风的设置使语音机器人具有一定的抗噪能力和远程多方位语音识别能力，适用于语音机器人在机场内部、展会会场入口、进出口口岸等环境嘈杂的工作场所与人进行远距离交互。

在一些实施例中，可以对麦克风 401 和前端处理芯片 402 进行相关语音交互平台

的设备认证。

在一些实施例中，语音机器人 40 还包括路由器 408。路由器 408 可以与蜂窝网络和 WIFI 网络连接，并可以在蜂窝网络和 WIFI 网络之前自动切换选择，以确保网络的稳定和低延迟。上位机 403 可以通过路由器 408 连接至不同的语音交互平台的云端服务器。

发明人认为，在实现了离在线结合的语音识别和理解之后，可以对构建更加完善的语音知识库，来提升语音交互服务的质量。知识库包括通用知识库和专业知识库，专业知识库分为问答库和技能库。其中，通用知识库包括用户日常使用的寒暄、生活查询、百科知识等领域；专业知识库包括安全、安检等行业领域知识和问答。问答库可以解决问答类的业务需求，技能库可以解决多轮交互的需求，可以涉及业务逻辑、数据等处理的需求。在语音机器人处理安全和安检业务时，可以协助安检人员解释法律条款、政策法规宣贯、引领咨询服务、整理分析现场用户与机器人交互内容、分析研究安全和安检法规。

实验得出，语音机器人 40 在环境噪声低于 60 分贝的条件能够达到如下指标：

项目	指标
1 米识别率	>92%
3 米识别率	>90%
5 米识别率	>85%
拾音角度	360 度
唤醒成功率	85%，1s 以内响应
应答正确率	对于知识库内问题应答正确率>80%
外部网络速率	上行速度>3Mbps，延迟<200ms

图 6 示出了本公开一些实施例的语音交互装置的结构示意图。如图 6 所示，该实施例的语音交互装置 60 包括：存储器 610 以及耦接至该存储器 610 的处理器 620，处理器 620 被配置为基于存储在存储器 610 中的指令，执行前述任意一些实施例中的语音交互方法。

其中，存储器 610 例如可以包括系统存储器、固定非易失性存储介质等。系统存储器例如存储有操作系统、应用程序、引导装载程序（Boot Loader）以及其他程序等。

语音交互装置 60 还可以包括输入输出接口 630、网络接口 640、存储接口 650 等。

这些接口 630、640、650 以及存储器 610 和处理器 620 之间例如可以通过总线 660 连接。其中，输入输出接口 630 为显示器、鼠标、键盘、触摸屏等输入输出设备提供连接接口。网络接口 640 为各种联网设备提供连接接口。存储接口 650 为 SD 卡、U 盘等外置存储设备提供连接接口。

本公开还包括一种计算机可读存储介质，其上存储有计算机指令，该指令被处理器执行时实现前述任意一些实施例中的语音交互方法。

本领域内的技术人员应明白，本公开的实施例可提供为方法、系统、或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用非瞬时性存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

本公开是参照根据本公开实施例的方法、设备（系统）、和计算机程序产品的流程图和／或方框图来描述的。应理解可由计算机程序指令实现流程图和／或方框图中的每一流程和／或方框、以及流程图和／或方框图中的流程和／或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能的装置。

这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和／或方框图一个方框或多个方框中指定的功能的步骤。

以上所述仅为本公开的较佳实施例，并不用以限制本公开，凡在本公开的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本公开的保护范围之内。

权 利 要 求

1、一种语音交互方法，包括：

利用深度学习神经网络对用户输入的语音进行处理，识别所述语音的语种；

为用户提供与所述语种相对应的机器人语音交互服务。

2、如权利要求 1 所述的语音交互方法，还包括：

利用标注了不同语种的语料对所述深度学习神经网络进行训练，使所述深度学习神经网络能够识别用户输入的语音的语种。

3、如权利要求 1 所述的语音交互方法，还包括：

判断机器人当前的网络连接状态；

若当前的网络连接状态为在线状态，则将用户输入的语音发送至云端服务器进行语音识别、语义理解和语音合成，接收云端反馈的第一输出语音；同时，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；

若先接收到所述第一输出语音后得到所述第二输出语音，则向用户播放所述第一输出语音；若先得到所述第二输出语音后接收到所述第一输出语音，则向用户播放所述第二输出语音。

4、如权利要求 3 所述的语音交互方法，还包括：

若当前的网络状态为离线状态，则将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；

向用户播放所述第二输出语音。

5、如权利要求 3 或 4 所述的语音交互方法，其中，进行语音识别时，根据用户输入的语音中各个字之间的时间间隔和语音能量对用户输入的语音进行断句。

6、如权利要求 3 或 4 所述的语音交互方法，其中，进行语音识别时，采用隐马

尔科夫模型对用户输入的语音进行断句。

7、如权利要求 1 所述的语音交互方法，还包括：

在机器人休眠状态下，利用摄像头对人脸进行实时识别；

响应于所述摄像头识别到人脸，将机器人从休眠状态唤醒至工作状态。

8、一种语音交互机器人，包括前端处理芯片，被配置为：

利用深度学习神经网络对用户输入的语音进行处理，识别所述语音的语种；

为用户提供与所述语种相对应的机器人语音交互服务。

9、如权利要求 8 所述的语音交互机器人，其中，所述深度学习神经网络经过标注了不同语种的语料的训练，能够识别用户输入的语音的语种。

10、如权利要求 8 所述的语音交互机器人，还包括上位机，被配置为：

判断机器人当前的网络连接状态；

若当前的网络连接状态为在线状态，则将用户输入的语音发送至云端服务器进行语音识别、语义理解和语音合成，接收云端反馈的第一输出语音；同时，将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；

若先接收到所述第一输出语音后得到所述第二输出语音，则向用户播放所述第一输出语音；若先得到所述第二输出语音后接收到所述第一输出语音，则向用户播放所述第二输出语音。

11、如权利要求 10 所述的语音交互机器人，其中，所述上位机还被配置为：

若当前的网络状态为离线状态，则将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音；

向用户播放所述第二输出语音。

12、如权利要求 10 或 11 所述的语音交互方法，其中，所述上位机被还配置为：

根据用户输入的语音中各个字之间的时间间隔和语音能量对用户输入的语音进行断句。

13、如权利要求 10 或 11 所述的语音交互方法，其中，所述上位机还被配置为：采用隐马尔科夫模型对用户输入的语音进行断句。

14、如权利要求 8 所述的语音交互机器人，还包括摄像头，被配置为：

在机器人休眠状态下对人脸进行实时识别；

在识别到人脸的情况下，将机器人从休眠状态唤醒至工作状态。

15、如权利要求 8 所述的语音交互机器人，还包括麦克风阵列，所述麦克风阵列的多个咪头位于相同水平面的同一圆周上，且相邻咪头之间的圆周距离相等。

16、如权利要求 15 所述的语音交互机器人，其中，所述麦克风阵列上套有硅胶套，所述硅胶套与所述语音机器人的外壳固定连接。

17、如权利要求 15 所述的语音交互机器人，还包括扩音器和多个腔体；其中，所述扩音器和所述麦克风阵列设置在语音交互机器人的不同腔体内。

18、如权利要求 15 所述的语音交互机器人，还包括包裹在所述麦克风阵列上的隔音棉。

19、一种语音交互装置，包括：

存储器；以及

耦接至所述存储器的处理器，所述处理器被配置为基于存储在所述存储器中的指令，执行如权利要求 1 至 7 中任一项所述的语音交互方法。

20、一种计算机可读存储介质，其中，所述计算机可读存储介质存储有计算机指令，所述指令被处理器执行时实现如权利要求 1 至 7 中任一项所述的语音交互方法。

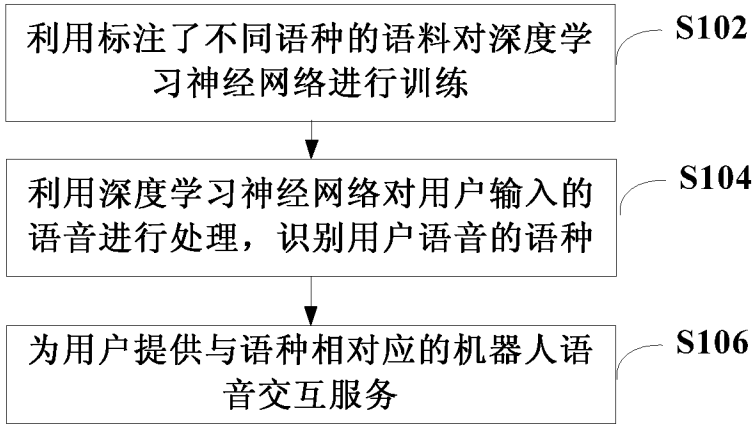

利用标注了不同语种的语料对深度学习神经网络进行训练
S102
利用深度学习神经网络对用户输入的语音进行处理，识别用户语音的语种
S104
为用户提供与语种相对应的机器人语音交互服务
S106

图 1

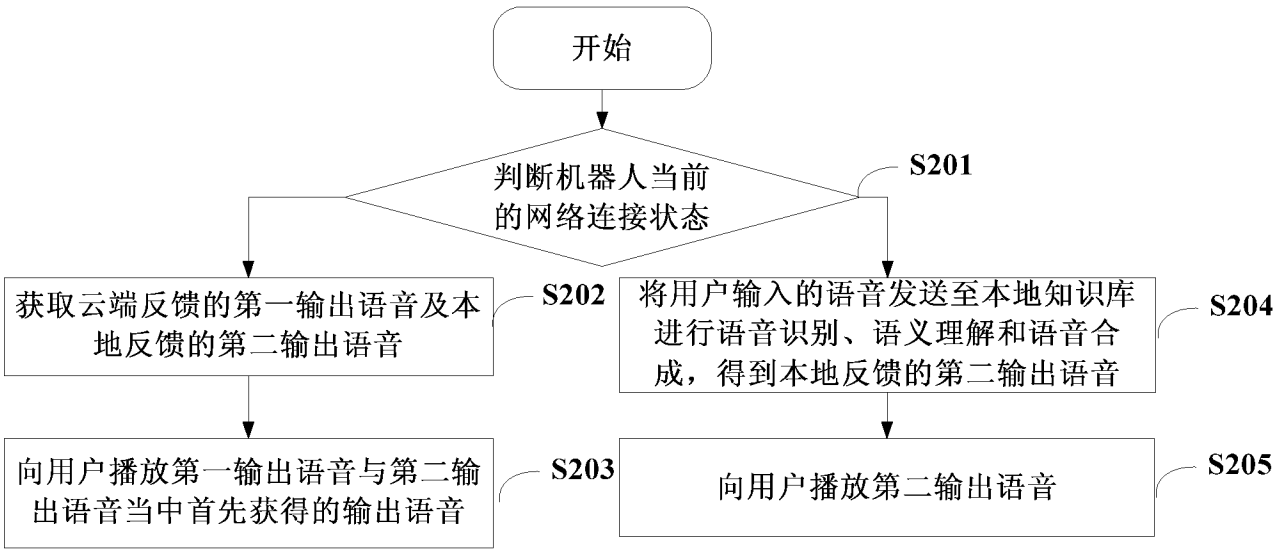

开始
判断机器人当前的网络连接状态
S201
获取云端反馈的第一输出语音及本地反馈的第二输出语音
S202
将用户输入的语音发送至本地知识库进行语音识别、语义理解和语音合成，得到本地反馈的第二输出语音
S204
向用户播放第一输出语音与第二输出语音当中首先获得的输出语音
S203
向用户播放第二输出语音
S205

图 2

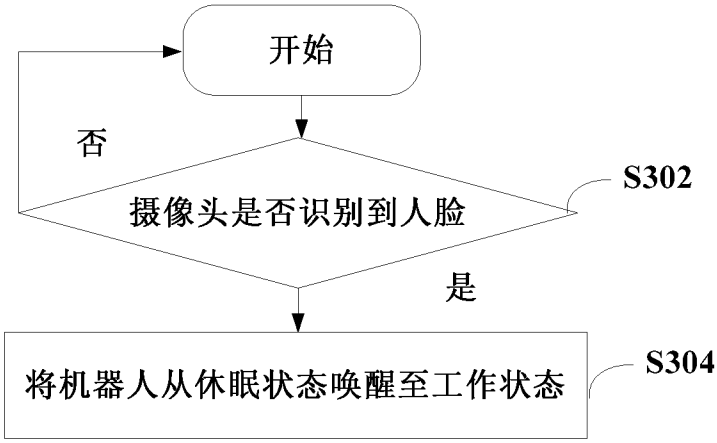

开始
否
摄像头是否识别到人脸
S302
是
将机器人从休眠状态唤醒至工作状态
S304

图 3

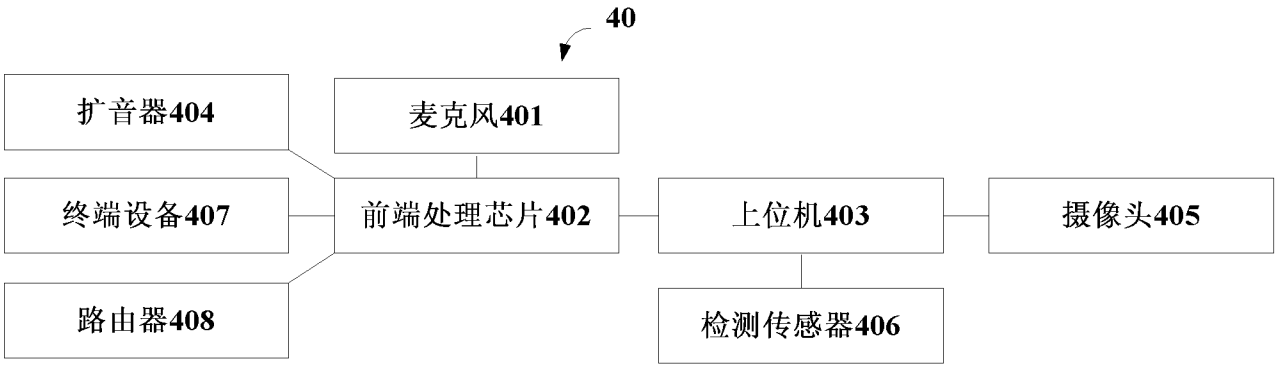

40
扩音器404
麦克风401
终端设备407
前端处理芯片402
上位机403
摄像头405
路由器408
检测传感器406

图 4

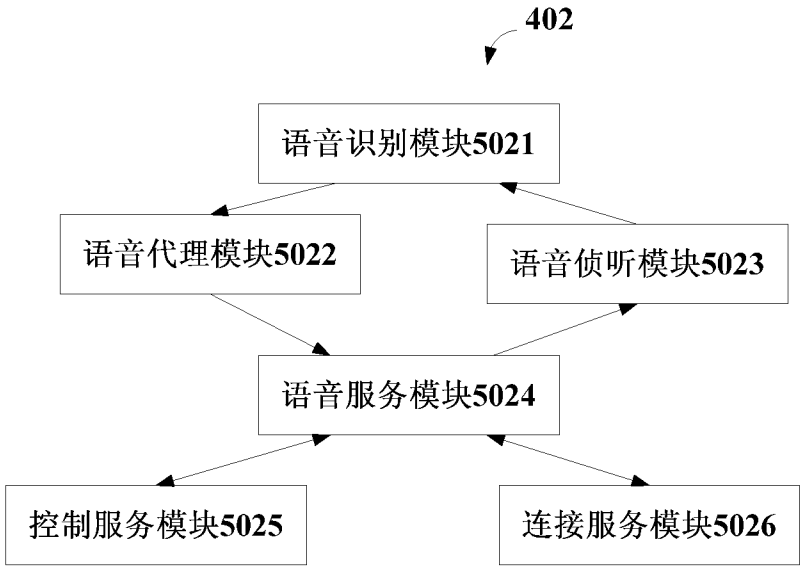

402
语音识别模块5021
语音代理模块5022
语音侦听模块5023
语音服务模块5024
控制服务模块5025
连接服务模块5026

图 5

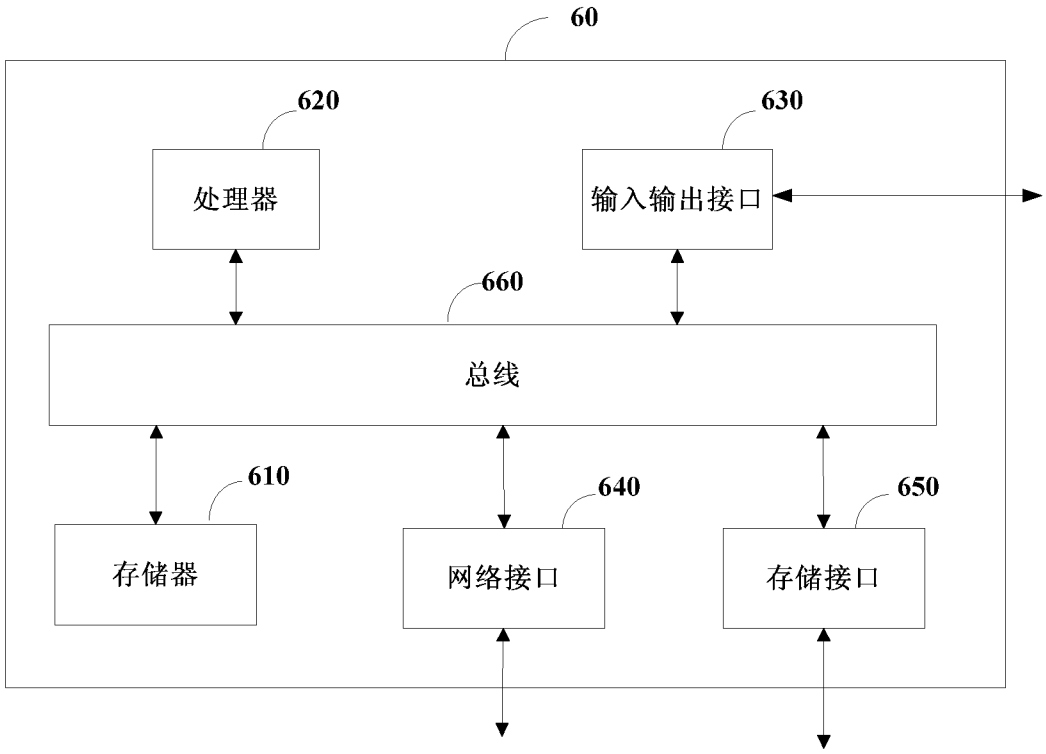

60
620
处理器
630
输入输出接口
660
总线
610
存储器
640
网络接口
650
存储接口

图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/124844

A. CLASSIFICATION OF SUBJECT MATTER

G10L 17/22(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: 语音, 语言, 神经网络, 训练, 模型, 识别, 在线, 离线, 机器人, 唤醒, speech, language, neutral network, DNN, train, Model, recognize, online, offline, robot, automaton, wake up

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103400577 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 20 November 2013 (2013-11-20) description, paragraphs [0026]-[0090], figures 1-10	1, 2, 8, 9, 15-20
Y	CN 103400577 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 20 November 2013 (2013-11-20) description, paragraphs [0026]-[0090], figures 1-10	3-7, 10-14
Y	CN 107424607 A (GREE ELECTRIC APPLIANCES, INC. OF ZHUHAI) 01 December 2017 (2017-12-01) description, paragraphs [0032]-[0063], figures 1-8	3-6, 10-13
Y	CN 108098767 A (BEIJING INTELLIGENT STEWARD CO., LTD.) 01 June 2018 (2018-06-01) description, paragraphs [0041]-[0105]	7, 14
A	CN 109065041 A (SHANGHAI CHANGREN INFORMATION TECHNOLOGY CO., LTD.) 21 December 2018 (2018-12-21) entire document	1-20

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 February 2020	**11 March 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/124844

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 108717853 A (GUANGDONG AIBE BIONIC ROBOT TECH CO LTD et al.) 30 October 2018 (2018-10-30) entire document	1-20
A	US 2018136615 A1 (ROBORUS CO., LTD.) 17 May 2018 (2018-05-17) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/124844

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 103400577 A	20 November 2013	CN 103400577 B	16 September 2015
CN 107424607 A	01 December 2017	None	
CN 108098767 A	01 June 2018	None	
CN 109065041 A	21 December 2018	None	
CN 108717853 A	30 October 2018	None	
US 2018136615 A1	17 May 2018	US 10289076 B2	14 May 2019

国际检索报告

国际申请号
PCT/CN2019/124844

A. 主题的分类

G10L 17/22(2013.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G10L

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT, WPI, EPODOC, CNKI：语音，语言，神经网络，训练，模型，识别，在线，离线，机器人，唤醒，speech, language, neutral network, DNN, train, Model, recognize, online, offline, robot, automaton, wake up

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CN 103400577 A (百度在线网络技术北京有限公司) 2013年 11月 20日 (2013-11-20) 说明书第[0026]-[0090]段，附图1-10	1，2，8，9，15-20
Y	CN 103400577 A (百度在线网络技术北京有限公司) 2013年 11月 20日 (2013-11-20) 说明书第[0026]-[0090]段，附图1-10	3-7，10-14
Y	CN 107424607 A (珠海格力电器股份有限公司) 2017年 12月 1日 (2017-12-01) 说明书第 [0032]-[0063]段，附图1-8	3-6，10-13
Y	CN 108098767 A (北京智能管家科技有限公司) 2018年 6月 1日 (2018-06-01) 说明书第 [0041]-[0105]段	7，14
A	CN 109065041 A (上海常仁信息科技有限公司) 2018年 12月 21日 (2018-12-21) 全文	1-20
A	CN 108717853 A (深圳艾比仿生机器人科技有限公司 等) 2018年 10月 30日 (2018-10-30) 全文	1-20
A	US 2018136615 A1 (ROBORUS CO., LTD.) 2018年 5月 17日 (2018-05-17) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2020年 2月 27日	2020年 3月 11日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 陈成 电话号码 86-(10)-53962548

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/124844

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 103400577 A	2013年 11月 20日	CN 103400577 B	2015年 9月 16日
CN 107424607 A	2017年 12月 1日	无	
CN 108098767 A	2018年 6月 1日	无	
CN 109065041 A	2018年 12月 21日	无	
CN 108717853 A	2018年 10月 30日	无	
US 2018136615 A1	2018年 5月 17日	US 10289076 B2	2019年 5月 14日