(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105556594 B
(45)授权公告日 2019.05.17

(21)申请号 201480051019.7

(22)申请日 2014.12.25

(65)同一申请的已公布的文献号
申请公布号 CN 105556594 A

(43)申请公布日 2016.05.04

(30)优先权数据
2013-268670 2013.12.26 JP

(85)PCT国际申请进入国家阶段日
2016.03.16

(86)PCT国际申请的申请数据
PCT/JP2014/006449 2014.12.25

(87)PCT国际申请的公布数据
WO2015/098109 JA 2015.07.02

(73)专利权人 松下知识产权经营株式会社
地址 日本国大阪府

(72)发明人 小沼知浩 小金井智弘

(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 齐秀凤

(51)Int.Cl.
G10L 15/22(2006.01)

(56)对比文件
WO 2013061857 A1,2013.05.02,
WO 2013061857 A1,2013.05.02,
JP 2005227686 A,2005.08.25,
CN 103247291 A,2013.08.14,
US 2013018895 A1,2013.01.17,
US 7949536 B2,2011.05.24,

审查员 游晓梅

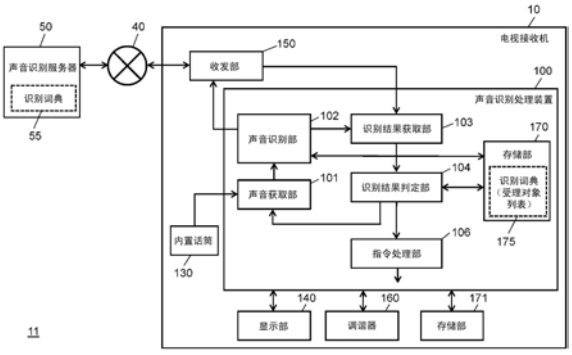
权利要求书2页 说明书15页 附图8页

(54)发明名称

声音识别处理装置、声音识别处理方法以及显示装置

(57)摘要

本发明提供一种声音识别处理装置、声音识别处理方法以及显示装置，提升声音操作的操作性。为此，在声音识别处理装置(100)中，声音获取部(101)构成为获取用户发出的声音并输出声音信息。声音识别部(102)构成为将声音信息变换为第1信息。存储部(170)预先存储了登记有排斥词汇的词典。识别结果判定部(104)比较第1信息和排斥词汇，在第1信息中包含与排斥词汇一致的词语时，将第1信息判定为应废弃的信息，在第1信息中不含与排斥词汇一致的词语时，将第1信息判定为应执行的信息。



CN 105556594 B

1.一种声音识别处理装置,其特征在于,具备:

声音获取部,其构成为获取用户发出的声音并输出声音信息;

第1声音识别部,其构成为将所述声音信息变换为第1信息;

存储部,其预先存储了登记有排斥词汇的词典;和

识别结果判定部,其比较所述第1信息和所述排斥词汇,判定在所述第1信息中是否包含与所述排斥词汇一致的词语,

在所述第1信息中包含与所述排斥词汇一致的词语时,所述识别结果判定部将所述第1信息判定为应废弃的信息,

在所述第1信息中不含与所述排斥词汇一致的词语时,所述识别结果判定部将所述第1信息判定为应执行的信息,

所述声音获取部基于获取到的所述声音,对所述用户发言的时间的长度进行测定来创建发言时间长度信息,

所述存储部中预先存储表示发言所需的时间的发言时间长度数据,

所述识别结果判定部针对不含与所述排斥词汇一致的词语而被判定为应执行的所述第1信息,从所述存储部中读出所述发言时间长度数据,比较读出的所述发言时间长度数据和由所述声音获取部创建的所述发言时间长度信息,基于所述比较来再次进行应废弃还是应执行的判定。

2.根据权利要求1所述的声音识别处理装置,其特征在于,

所述声音获取部基于获取到的所述声音,对在所述用户的发言的前后所产生的无声音期间的长度进行测定来创建发言方式信息,

所述存储部中预先存储表示在发言的前后所产生的无声音期间的长度的发言方式数据,

所述识别结果判定部针对不含与所述排斥词汇一致的词语而被判定为应执行的所述第1信息,从所述存储部中读出所述发言方式数据,比较读出的所述发言方式数据和由所述声音获取部创建的所述发言方式信息,基于所述比较来再次进行应废弃还是应执行的判定。

3.根据权利要求1所述的声音识别处理装置,其特征在于,还具备:

第2声音识别部,其构成为将所述声音信息变换为第2信息;和

选择部,其构成为选择所述第1信息和所述第2信息的任意一方并输出,

所述识别结果判定部针对在所述选择部中选择出的一方的信息,进行应废弃还是应执行的判定。

4.根据权利要求3所述的声音识别处理装置,其特征在于,

所述第2声音识别部被设置在网络上,

所述声音识别处理装置还具备:收发部,其构成为经由所述网络而与所述第2声音识别部进行通信。

5.一种声音识别处理方法,其特征在于,包括:

获取用户发出的声音并变换为声音信息,并且基于获取到的所述声音对所述用户发言的时间的长度进行测定来创建发言时间长度信息的步骤;

将所述声音信息变换为第1信息的步骤;

将所述声音信息变换为第2信息的步骤；

选择所述第1信息和所述第2信息的任意一方的步骤；

比较所述选择出的信息和词典中登记的排斥词汇，判定在所述选择出的信息中是否包含与所述排斥词汇一致的词语的步骤；

在所述选择出的信息中包含与所述排斥词汇一致的词语时，将所述选择出的信息判定为应废弃的信息的步骤；

在所述选择出的信息中不含与所述排斥词汇一致的词语时，将所述选择出的信息判定为应执行的信息的步骤；和

针对不含与所述排斥词汇一致的词语而被判定为应执行的所述选择出的信息，比较创建的所述发言时间长度信息和预先存储的表示发言所需的时间的发言时间长度数据，基于所述比较来再次进行应废弃还是应执行的判定的步骤。

6.一种显示装置，其特征在于，具备：

声音获取部，其构成为获取用户发出的声音并输出声音信息；

第1声音识别部，其构成为将所述声音信息变换为第1信息；

存储部，其预先存储了登记有排斥词汇的词典；

识别结果判定部，其构成为比较所述第1信息和所述排斥词汇，判定在所述第1信息中是否包含与所述排斥词汇一致的词语，并基于所述判定来判定应废弃还是应执行所述第1信息；

处理部，其构成为执行基于在所述识别结果判定部中判定为应执行的所述第1信息的处理；和

显示部，

在所述第1信息中包含与所述排斥词汇一致的词语时，所述识别结果判定部将所述第1信息判定为应废弃的信息，

在所述第1信息中不含与所述排斥词汇一致的词语时，所述识别结果判定部将所述第1信息判定为应执行的信息，

所述声音获取部基于获取到的所述声音，对所述用户发言的时间的长度进行测定来创建发言时间长度信息，

所述存储部中预先存储表示发言所需的时间的发言时间长度数据，

所述识别结果判定部针对不含与所述排斥词汇一致的词语而被判定为应执行的所述第1信息，从所述存储部中读出所述发言时间长度数据，比较读出的所述发言时间长度数据和由所述声音获取部创建的所述发言时间长度信息，基于所述比较来再次进行应废弃还是应执行的判定。

声音识别处理装置、声音识别处理方法以及显示装置

技术领域

[0001] 本公开涉及识别用户发出的声音来进行动作的声音识别处理装置、声音识别处理方法以及显示装置。

背景技术

[0002] 专利文献1公开了具有声音识别功能的声音输入装置。该声音输入装置构成为:接收用户发出的声音,对接收到的声音进行解析,由此来识别(声音识别)用户的声音所表示的命令,并根据声音识别出的命令来控制设备。即,专利文献1的声音输入装置能够对用户任意发出的声音进行声音识别,并根据作为该声音识别出的结果的命令(指令)来控制设备。

[0003] 例如,使用该声音输入装置的用户在利用电视接收机(以下记为“电视机”)、PC(Personal Computer;个人计算机)等对浏览器进行操作时,能够利用该声音输入装置的声音识别功能来进行显示在浏览器上的超文本的选择。此外,用户也能够利用该声音识别功能来进行在提供搜索服务的网站(搜索站点)上的搜索。

[0004] 此外,在该声音输入装置中,为了提高用户的便利性,有时会进行“无触发识别”。所谓“无触发识别”,是指在声音输入装置中,对于受理用于声音识别的声音输入的期间不设限制,始终进行声音的集音和针对集音到的声音的声音识别这一状态。然而,若在该声音输入装置中进行无触发识别,则难以区分集音到的声音是用户以声音识别为目的而发出的声音,还是用户彼此的对话、用户的自言自语等不以声音识别为目的的声音,因此有时会错误地对不以声音识别为目的的声音进行声音识别(误识别)。

[0005] 在先技术文献

[0006] 专利文献

[0007] 专利文献1:日本专利第4812941号公报

发明内容

[0008] 本公开提供降低误识别、提升用户的操作性的声音识别处理装置以及声音识别处理方法。

[0009] 本公开中的声音识别处理装置具备:声音获取部、第1声音识别部、存储部和识别结果判定部。声音获取部构成为获取用户发出的声音并输出声音信息。第1声音识别部构成为将声音信息变换为第1信息。存储部预先存储了登记有排斥词汇的词典。识别结果判定部比较第1信息和排斥词汇,判定在第1信息中是否包含与排斥词汇一致的词语。并且,在第1信息中包含与排斥词汇一致的词语时,识别结果判定部将第1信息判定为应废弃的信息,在第1信息中不含与排斥词汇一致的词语时,识别结果判定部将第1信息判定为应执行的信息。

[0010] 本公开中的声音识别处理方法包括:获取用户发出的声音并变换为声音信息的步骤;将声音信息变换为第1信息的步骤;将声音信息变换为第2信息的步骤;选择第1信息和

第2信息的任意一方的步骤;比较选择出的信息和词典中所登记的排斥词汇,判定在选择出的信息中是否包含与排斥词汇一致的词语的步骤;在选择出的信息中包含与排斥词汇一致的词语时,将选择出的信息判定为应废弃的信息的步骤;和在选择出的信息中不含与排斥词汇一致的词语时,将选择出的信息判定为应执行的信息的步骤。

[0011] 本公开中的显示装置具备:声音获取部、第1声音识别部、存储部、识别结果判定部、处理部和显示部。声音获取部构成为获取用户发出的声音并输出声音信息。第1声音识别部构成为将声音信息变换为第1信息。存储部预先存储了登记有排斥词汇的词典。识别结果判定部比较第1信息和排斥词汇,判定在第1信息中是否包含与排斥词汇一致的词语,并基于该判定来判定应废弃还是应执行第1信息。处理部构成为执行基于在识别结果判定部中判定为应执行的第1信息的处理。并且,在第1信息中包含与排斥词汇一致的词语时,识别结果判定部将第1信息判定为应废弃的信息,在第1信息中不含与排斥词汇一致的词语时,识别结果判定部将第1信息判定为应执行的信息。

[0012] 本公开中的声音识别处理装置能提升用户进行声音操作时的操作性。

附图说明

[0013] 图1是示意性地表示实施方式1中的声音识别处理系统的图。

[0014] 图2是表示实施方式1中的声音识别处理系统的一构成例的框图。

[0015] 图3是表示实施方式1中的声音识别处理装置的识别结果判定部的一构成例的框图。

[0016] 图4是表示实施方式1中的声音识别处理装置的一动作例的流程图。

[0017] 图5是表示实施方式2中的声音识别处理系统的一构成例的框图。

[0018] 图6是表示实施方式2中的声音识别处理装置的识别结果判定部的一构成例的框图。

[0019] 图7是表示实施方式2中的识别结果判定部的一动作例的流程图。

[0020] 图8A是表示其他实施方式中的识别结果判定部的一构成例的框图。

[0021] 图8B是表示其他实施方式中的识别结果判定部的一构成例的框图。

具体实施方式

[0022] 以下,适当参照附图来详细地说明实施方式。不过,有时也会根据需要来省略详细说明。例如,有时将省略已经熟知的事项的详细说明、针对实质上为相同构成的重复说明。其原因在于,为了避免下面的说明变得不必要的冗长,使本领域技术人员易于理解。

[0023] 另外,为了使本领域技术人员充分地理解本公开而提供了附图以及下述说明,并非意图通过这些内容来限定请求保护的范围所记载的主题。

[0024] (实施方式1)

[0025] 以下,利用图1～图4来说明实施方式1。另外,在本实施方式中,作为具备声音识别处理装置的显示装置的一例而列举了电视接收机(电视机)10,但显示装置并不限定于电视机10。例如,也可以为PC、平板终端、便携式终端等。

[0026] 此外,虽然假设本实施方式所示的声音识别处理系统11进行无触发识别,但本公开并不限定于无触发识别,也能够应用于通过基于用户700的声音识别的开始操作来开始

声音识别的系统。

[0027] [1-1.构成]

[0028] 图1是示意性地表示实施方式1中的声音识别处理系统11的图。在本实施方式中，在作为显示装置的一例的电视机10中内置有声音识别处理装置。

[0029] 本实施方式中的声音识别处理系统11具备：作为显示装置的一例的电视机10、和声音识别服务器50。

[0030] 若电视机10中启动了声音识别处理装置，则在电视机10的显示部140中，与基于输入影像信号、接收到的广播信号等的影像一起显示声音识别图标203、和表示集音到的声音的音量的指示器202。这是为了向用户700示出处于能够实现基于用户700的声音的电视机10的操作(以下记为“声音操作”)的状态，并且促使用户700发言。

[0031] 若用户700朝向电视机10所具备的内置话筒130发出声音，则该声音被内置话筒130集音，集音到的声音被电视机10所内置的声音识别处理装置声音识别。在电视机10中，根据该声音识别的结果来进行电视机10的控制。

[0032] 电视机10也能采用如下构成，即，具备构成为用户700发出的声音被内置的话筒集音并被无线发送至电视机10的遥控器或者便携式终端。

[0033] 此外，电视机10经由网络40而与声音识别服务器50连接。并且，能够在电视机10与声音识别服务器50之间进行通信。

[0034] 图2是表示实施方式1中的声音识别处理系统11的一构成例的框图。

[0035] 电视机10具备：声音识别处理装置100、显示部140、收发部150、调谐器160、存储部171和内置话筒130。

[0036] 声音识别处理装置100构成为获取用户700发出的声音并对获取到的声音进行解析。并且，构成为识别该声音所表示的指示并根据识别出的结果来进行电视机10的控制。声音识别处理装置100的具体构成将后述。

[0037] 内置话筒130是构成为主要对来自与显示部140的显示面对置的方向的声音进行集音的话筒。即，内置话筒130将集音方向设定为能够对与电视机10的显示部140面对的用户700发出的声音进行集音，从而能够集音用户700发出的声音。内置话筒130可以设置在电视机10的框体内，也可以如图1中示出的一例那样设置在电视机10的框体外。

[0038] 显示部140例如为液晶显示器，但也可以为等离子显示器、或者有机EL(ElectroLuminescence；电致发光)显示器等。显示部140由显示控制部(未图示)来控制，显示基于来自外部的输入影像信号、由调谐器160接收到的广播信号等的图像。

[0039] 收发部150与网络40连接，构成为通过网络40而与网络40所连接的外部设备(例如声音识别服务器50)进行通信。

[0040] 调谐器160构成为经由天线(未图示)来接收地面广播、卫星广播的电视广播信号。调谐器160也可以构成为接收经由专用线缆而发送的电视广播信号。

[0041] 存储部171例如为非易失性的半导体存储器，但也可以为易失性的半导体存储器、或者硬盘等。存储部171存储电视机10的各部分的控制中所利用的信息(数据)、程序等。

[0042] 网络40例如为因特网，但也可以为其他网络。

[0043] 声音识别服务器50为“第2声音识别部”的一例。声音识别服务器50是经由网络40而与电视机10连接的服务器(云上的词典服务器)。声音识别服务器50构成为具备识别词典

55，接收从电视机10经由网络40而发送来的声音信息。识别词典55是用于将声音信息和声音识别模型建立对应的数据库。并且，声音识别服务器50比对接收到的声音信息和识别词典55的声音识别模型，确认在接收到的声音信息中是否包含与识别词典55所登记的声音识别模型对应的声音信息。并且，若在接收到的声音信息中包含与识别词典55所登记的声音识别模型对应的声音信息，则选择该声音识别模型所表示的字符串。如此，将接收到的声音信息变换为字符串。另外，该字符串既可以为多个字符也可以为一个字符。并且，声音识别服务器50将表示变换后的字符串的字符串信息作为声音识别的结果经由网络40发送至电视机10。该字符串信息为“第2信息”的一例。

[0044] 声音识别处理装置100具有：声音获取部101、声音识别部102、识别结果获取部103、识别结果判定部104、指令处理部106和存储部170。

[0045] 存储部170例如为非易失性的半导体存储器，能够任意地实现数据的写入与读出。存储部170也可以为易失性的半导体存储器、或者硬盘等。存储部170还存储被声音识别部102、识别结果判定部104参照的信息(例如识别词典175)等。识别词典175为“词典”的一例。识别词典175为用于将声音信息和声音识别模型建立对应的数据库。此外，在识别词典175中还登记有排斥对象列表。排斥对象列表的详细将后述。另外，存储部170和存储部171也可以构成为一体。

[0046] 声音获取部101获取基于用户700发出的声音的声音信号并变换为声音信息，输出至声音识别部102。

[0047] 声音识别部102为“第1声音识别部”的一例。声音识别部102将声音信息变换为字符串信息，并将该字符串信息作为声音识别的结果而输出至识别结果获取部103。该字符串信息为“第1信息”的一例。此外，声音识别部102将从声音获取部101获取到的声音信息自收发部150经由网络40发送至声音识别服务器50。

[0048] 声音识别服务器50参照识别词典55，对从电视机10接收到的声音信息进行声音识别，并将该声音识别的结果返回给电视机10。

[0049] 识别结果获取部103为“选择部”的一例。识别结果获取部103若接收到从声音识别部102输出的声音识别的结果(第1信息)和从声音识别服务器50返回的声音识别的结果(第2信息)，则比较两者，选择任意一方。然后，识别结果获取部103将选择出的一方输出给识别结果判定部104。

[0050] 识别结果判定部104针对从识别结果获取部103输出的声音识别的结果，进行应废弃还是应执行(受理)的判定。该详细将后述。然后，基于该判定，将声音识别的结果输出至指令处理部106或者声音获取部201。

[0051] 指令处理部106基于来自识别结果判定部104的输出(判定为应执行的声音识别的结果)，进行指令处理(例如电视机10的控制等)。指令处理部106为“处理部”的一例，该指令处理为“处理”的一例。

[0052] 图3是表示实施方式1中的声音识别处理装置100的识别结果判定部104的一构成例的框图。

[0053] 识别结果判定部104具备排斥词汇废弃部1042和受理废弃发送部1045。它们的动作的详细将后述。

[0054] [1-2.动作]

[0055] 下面，说明本实施方式中的电视机10的声音识别处理装置100的动作。

[0056] 图3是表示实施方式1中的声音识别处理装置100的一动作例的流程图。

[0057] 声音获取部101从电视机10的内置话筒130获取基于用户700发出的声音的声音信号(步骤S101)。

[0058] 声音获取部101也可以从遥控器(未图示)内置的话筒、或者便携式终端(未图示)内置的话筒经由无线通信部(未图示)来获取声音信号。

[0059] 然后，声音获取部101将该声音信号变换为能够在后级的各种处理中利用的声音信息，输出至声音识别部102。另外，若声音信号为数字信号，则声音获取部101也可以将该声音信号直接用作声音信息。

[0060] 声音识别部102将从声音获取部101获取到的声音信息变换为字符串信息。然后，将该字符串信息作为声音识别的结果而输出至识别结果获取部103。此外，声音识别服务器50将从电视机10经由网络40获取到的声音信息变换为字符串信息，并将该字符串信息作为声音识别的结果而返回给电视机10(步骤S102)。

[0061] 具体而言，声音识别部102基于从声音获取部101获取到的声音信息，参照预先存储于存储部170的识别词典175内的受理对象列表。然后，比较该声音信息和受理对象列表所登记的声音识别模型。

[0062] 所谓声音识别模型，是指用于将声音信息和字符串信息建立对应的信息。在进行声音识别之际，比较多个声音识别模型的每一个模型和声音信息，选择与声音信息一致或者类似的一个声音识别模型。并且，与该声音识别模型建立了对应的字符串信息成为针对该声音信息的声音识别的结果。在受理对象列表中例如登记有：针对电视机10的指示(例如频道变更、音量变更等)、电视机10的功能(例如网络连接功能等)、电视机10的各部分的名称(例如电源、频道等)、针对电视机10的画面上显示的内容的指示(例如放大、缩小、滚动等)等与电视机10的操作关联的声音识别模型。

[0063] 另外，在存储于存储部170的识别词典175中，除了登记有受理对象列表之外，还登记有后述的排斥对象列表(图2中未示出)。

[0064] 声音识别部102比对声音信息和受理对象列表所登记的声音识别模型。并且，若在从声音获取部101获取到的声音信息中包含与受理对象列表所登记的声音识别模型对应的字符串信息，则将与该声音识别模型建立了对应的字符串信息作为声音识别的结果而输出至识别结果获取部103。

[0065] 声音识别部102在比对声音信息和声音识别模型时，计算识别得分。所谓识别得分，是指表示似然性(likelihood)的数值，是表示声音信息与该声音识别模型以何种程度一致或者类似的指标，数值越大则类似度越高。声音识别部102比对声音信息和声音识别模型，将多个声音识别模型作为候选来选择。此时，声音识别部102针对各个声音识别模型来计算识别得分。另外，该识别得分的计算方法也可以为一般公知的方法。并且，选择在预先设定的阈值以上且最高的识别得分的声音识别模型，将与该声音识别模型对应的字符串信息作为声音识别的结果来输出。另外，声音识别部102也可以将该字符串信息关联的识别得分与该字符串信息一起输出至识别结果获取部103。

[0066] 如此，声音识别部102将声音信息变换为字符串信息。另外，声音识别部102也可以将声音信息变换为字符串信息以外的信息来输出。此外，若不存在阈值以上的识别得分的

声音识别模型，则声音识别部102也可以输出表示不能实现声音识别的信息。

[0067] 此外，声音识别部102将从声音获取部101获取到的声音信息自收发部150经由网络40发送至声音识别服务器50。

[0068] 声音识别服务器50基于从电视机10接收到的声音信息，参照识别词典55。然后，将该声音信息与识别词典55内的声音识别模型进行比对，变换为字符串信息。

[0069] 声音识别服务器50在将接收到的声音信息与识别词典55内的声音识别模型进行比较时，计算识别得分。该识别得分是表示与由声音识别部102计算的识别得分同样的似然性的数值，以与由声音识别部102计算识别得分时同样的方法来计算。声音识别服务器50与声音识别部102同样，基于接收到的声音信息，将多个声音识别模型作为候选来选择，并基于识别得分，从该候选之中选择一个声音识别模型。然后，声音识别服务器50将与该声音识别模型建立了对应的字符串信息作为声音识别的结果而返回给电视机10。声音识别服务器50也可以将该字符串信息关联的识别得分与该字符串信息一起发送至电视机10。

[0070] 声音识别服务器50构成为：能够通过网络40来收集各种各样的用语，并将这些用语登记在识别词典55中。因而，声音识别服务器50能够具备比电视机10所具备的识别词典175更多的声音识别模型。因此，在声音识别服务器50中，在用户700发出了与电视机10的功能、对电视机10作出的指示无关的词语(例如用户彼此的对话、自言自语等)时，针对该声音的声音识别的识别得分有可能比电视机10的声音识别部102进行同样的声音识别时更高。

[0071] 从声音识别服务器50经由网络40接收到声音识别的结果的收发部150，将该声音识别的结果输出至识别结果获取部103。

[0072] 如果从声音识别部102和声音识别服务器50分别接收到声音识别的结果，则识别结果获取部103按照判别规则来选择其中一方的声音识别结果(步骤S103)。

[0073] 该判别规则例如可以是如下规则：相互比较从声音识别部102接收到的声音识别的结果中附带的识别得分、和从声音识别服务器50接收到的声音识别的结果中附带的识别得分，选择识别得分高的一方的声音识别结果。识别结果获取部103将选择出的声音识别结果输出给识别结果判定部104。

[0074] 另外，识别结果获取部103在只能从声音识别部102和声音识别服务器50的任意一方接收声音识别的结果时，也可以跳过步骤S103的处理，将接收到的声音识别的结果直接输出。

[0075] 图3所示的识别结果判定部104的排斥词汇废弃部1042判定在从识别结果获取部103输出的声音识别的结果中是否存在与排斥对象列表所登记的词汇(排斥词汇)一致的词汇(步骤S104)。

[0076] 所谓排斥对象列表，是指将判断为不用于电视机10的声音操作的词语(词汇)作为排斥词汇进行登记的列表。排斥词汇例如为将在存储部170的识别词典175中作为受理对象列表被登记的词汇除外的词汇。该排斥对象列表预先登记于存储部170的识别词典175，但也可以构成为能够任意追加新的排斥词汇。另外，若将发音与在对电视机10进行声音操作时用户700所发出的词语相似、且与电视机10的声音操作无关的词汇作为排斥词汇而登记于排斥对象列表，则能够提升声音识别的精度。

[0077] 在步骤S104中，排斥词汇废弃部1042比对存储部170所存储的识别词典175内的排斥对象列表、和从识别结果获取部103输出的作为声音识别的结果的字符串信息，调查有无

与排斥对象列表所含的排斥词汇一致的字符串信息。并且，排斥词汇废弃部1042将与排斥词汇一致的字符串信息判定为应废弃的信息，赋予标记并输出至受理废弃发送部1045(是)。

[0078] 若从排斥词汇废弃部1042输入的字符串信息被赋予了标记，则受理废弃发送部1045将该字符串信息作为废弃信息而输出给声音获取部101。接收到废弃信息的声音获取部101为了下次的声音识别而进行声音获取的准备(步骤S106)。因此，指令处理部106针对赋予了标记的字符串信息(废弃信息)不进行任何处理。

[0079] 在步骤S104中，排斥词汇废弃部1042将与排斥词汇不一致的字符串信息判定为是应受理(执行)的信息，不赋予标记地输出至受理废弃发送部1045(否)。

[0080] 若从排斥词汇废弃部1042输入的字符串信息未被赋予标记，则受理废弃发送部1045将该字符串信息输出给指令处理部106。指令处理部106基于从受理废弃发送部1045接收到的字符串信息所表示的指示来执行指令处理(步骤S105)。

[0081] 例如，若在字符串信息中包含频道变更、音量变更等与电视机10的控制相关的指令信息，则指令处理部106对电视机10的控制部(未图示)作出指示，以使电视机10执行与该指令信息对应的动作。

[0082] 在步骤S105结束后，指令处理部106向声音获取部101发送表示指令处理已结束的信号。接收到该信号的声音获取部101为了下次的声音识别而进行声音获取的准备(步骤S106)。

[0083] [1-3.效果等]

[0084] 如以上，在本实施方式中，声音识别处理装置100具备；声音获取部101、作为第1声音识别部的一例的声音识别部102、存储部170和识别结果判定部104。声音获取部101构成为获取用户700发出的声音并输出声音信息。声音识别部102构成为将声音信息变换为作为第1信息的一例的字符串信息。存储部170预先存储了登记有排斥词汇的识别词典175。识别词典175为词典的一例。识别结果判定部104比较字符串信息和排斥词汇，判定在字符串信息中是否包含与排斥词汇一致的词语。并且，识别结果判定部104在字符串信息中包含与排斥词汇一致的词语时，将字符串信息判定为应废弃的信息，在字符串信息中不含与排斥词汇一致的词语时，将字符串信息判定为应执行的信息。

[0085] 此外，声音识别处理装置100也可以还具备：作为第2声音识别部的一例的声音识别服务器50、和作为选择部的一例的识别结果获取部103。在该情况下，声音识别服务器50构成为将声音信息变换为作为第2信息的一例的字符串信息。识别结果获取部103构成为选择声音识别部102所输出的字符串信息和声音识别服务器50所输出的字符串信息的任意一方并输出。并且，识别结果判定部104针对在识别结果获取部103中选择出的一方的字符串信息，进行应废弃还是应执行的判定。

[0086] 作为第2声音识别部的一例的声音识别服务器50也可以设置在网络40上。声音识别处理装置100也可以具备：收发部150，其构成为经由网络40而与声音识别服务器50进行通信。

[0087] 在如此构成的声音识别处理装置100中，能够精度良好地判别用户700为了进行声音操作而发出的声音、和用户700彼此的对话、自言自语的声音，降低误识别，从而提升声音识别的精度。

[0088] 例如，假设用户700发出的是发音与在对电视机10进行声音操作时所发出的词语相似、且与电视机10的声音操作无关的词语。此时，声音识别部102作为基于该声音的声音识别的结果而输出受理对象列表所登记的字符串信息(即进行误识别)的可能性高。

[0089] 另一方面，在具有认为通过网络40来更新登记内容从而登记比识别词典175更多的声音识别模型(词汇)的识别词典55的声音识别服务器50中，针对这种声音而进行更正确的声音识别的可能性高。

[0090] 因此，与声音识别部102对易于被误识别的声音进行误识别而输出的字符串信息所附带的识别得分相比，声音识别服务器50对该声音进行声音识别而输出的字符串信息所附带的识别得分的数值更大，认为从声音识别服务器50输出的字符串信息被识别结果获取部103选择的可能性高。

[0091] 并且，若与该字符串信息对应的词汇作为排斥词汇而登记于识别词典175内的排斥对象列表，则在排斥词汇废弃部1042中将该字符串信息判断为应废弃的信息。

[0092] 如此，根据本实施方式，能够提高针对被声音识别部102错误地声音识别的声音的声音识别的精度，防止指令处理部106进行因误识别而引起的错误的指令处理。

[0093] 此外，用户700发言的声音不是十分大或者噪音多等时，声音识别部102发生误识别的可能性也高，但在此情况下，也能够提高声音识别的精度。

[0094] 另外，若声音识别部102所具有的识别词典175与声音识别服务器50的识别词典55同样构成为能够通过网络40等来更新登记内容，则也可以将声音识别处理系统11构成为仅利用电视机10来实施声音识别。即便是这种构成，通过识别结果判定部104的动作，也能够降低误识别来提高声音识别的精度。

[0095] (实施方式2)

[0096] 下面，利用图5～图7来说明实施方式2。在实施方式2中，说明提高针对用户700发出的可能性高的词语(例如与电视机10的动作、功能等相关的词语)的声音识别的精度的方法。

[0097] [2-1.构成]

[0098] 图5是表示实施方式2中的声音识别处理系统21的一构成例的框图。

[0099] 本实施方式中的声音识别处理系统21具备：作为显示装置的一例的电视机20和声音识别服务器50。该声音识别服务器50与实施方式1中所说明的声音识别服务器50实质上相同，因此省略说明。

[0100] 电视机20具有：声音识别处理装置200、显示部140、收发部150、调谐器160、存储部171和内置话筒130。声音识别处理装置200具有：声音获取部201、声音识别部102、识别结果获取部103、识别结果判定部204、指令处理部106和存储部270。

[0101] 另外，关于与实施方式1中所说明的电视机10具备的构成要素进行实质上相同的动作的构成要素，赋予与实施方式1相同的符号，并省略说明。

[0102] 此外，在存储部270内的识别词典175中，假设登记与实施方式1中所说明的受理对象列表以及排斥对象列表同样的受理对象列表以及排斥对象列表。

[0103] 实施方式2中的声音识别处理装置200与实施方式1中所说明的声音识别处理装置100，在声音获取部201以及识别结果判定部204中的动作上存在差异。

[0104] 声音获取部201与实施方式1中所说明的声音获取部101同样，从内置话筒130获取

基于用户700发出的声音的声音信号。不过，声音获取部201不同于实施方式1所示的声音获取部101，基于获取到的声音信号来创建发言时间长度信息和发言方式信息。

[0105] 所谓发言时间长度信息，是表示用户700发言的时间的长度的信息。声音获取部201例如对预先设定的阈值以上的音量的声音连续发生的时间的长度进行测定，从而能够创建发言时间长度信息。声音获取部201也可以通过其他的方法来创建发言时间长度信息。

[0106] 所谓发言方式信息，是表示在用户700的发言的前后所产生的无声音或者视作实质上无声音的时间的长度的信息。声音获取部201例如将音量低于预先设定的阈值的状态设为无声音，对在发言的前后所产生的无声音期间的长度进行测定，从而能够创建发言方式信息。声音获取部201也可以通过其他的方法来创建发言方式信息。

[0107] 声音获取部201对声音信息分别附加发言时间长度信息和发言方式信息，并输出至声音识别部102。

[0108] 在多个用户700彼此的对话、用户700的自言自语等中有时会包含受理对象列表所登记的词汇(受理对象词汇)。并且，有时该声音会被内置话筒130集音从而基于该声音的声音信息被输入至声音识别部102。在这种情况下，声音识别部102进行基于该声音信息的错误的声音识别，尽管用户700不打算对电视机20进行声音操作，却有可能在指令处理部106中进行基于误识别的错误的指令处理。在本实施方式中，为了降低这种误识别的发生，进行除了利用实施方式1中所说明的排斥对象列表之外还利用“发言时间长度信息”和“发言方式信息”的声音识别。

[0109] 发言时间长度信息和发言方式信息的详细将后述。此外，声音识别部102将附加了发言时间长度信息和发言方式信息的声音信息经由收发部150以及网络40而发送至声音识别服务器50。

[0110] [2-2.动作]

[0111] 下面，利用图6和图7来说明本实施方式中的电视机20的声音识别处理装置200所具有的识别结果判定部204的构成以及动作。

[0112] 图6是表示实施方式2中的声音识别处理装置200的识别结果判定部204的一构成例的框图。

[0113] 识别结果判定部204具备：排斥词汇废弃部1042、发言时间长度判定部2043、发言方式判定部2044和受理废弃发送部1045。

[0114] 图7是表示实施方式2中的识别结果判定部204的一动作例的流程图。

[0115] 识别结果获取部103与实施方式1中所说明的步骤S103同样，若从声音识别部102和声音识别服务器50分别接收到声音识别的结果，则按照判别规则来选择其中一方的声音识别结果(步骤S103)。该判别规则与实施方式1中所说明的判别规则实质上相同。

[0116] 识别结果判定部204的排斥词汇废弃部1042与实施方式1中所说明的步骤S104同样，判定在从识别结果获取部103输出的声音识别的结果中是否存在与排斥对象列表所登记的词汇(排斥词汇)一致的词汇(步骤S104)。

[0117] 在步骤S104中，排斥词汇废弃部1042与实施方式1中所说明的排斥词汇废弃部1042同样，比对存储部270所存储的识别词典175内的排斥对象列表、和从识别结果获取部103输出的作为声音识别的结果的字符串信息，调查有无与排斥对象列表所含的排斥词汇一致的字符串信息。并且，排斥词汇废弃部1042将与排斥词汇一致的字符串信息判定为是

应废弃的信息，赋予标记并输出给受理废弃发送部1045(是)。

[0118] 受理废弃发送部1045与实施方式1中所说明的受理废弃发送部1045同样，将赋予了标记的字符串信息作为废弃信息而输出给声音获取部201。接收到废弃信息的声音获取部201为了下次的声音识别而进行声音获取的准备(步骤S106)。

[0119] 另一方面，在步骤S104中，排斥词汇废弃部1042对于与排斥词汇不一致的字符串信息，不赋予标记地直接输出给发言时间长度判定部2043(否)。

[0120] 发言时间长度判定部2043针对从排斥词汇废弃部1042输入的未被赋予标记的字符串信息，基于发言时间长度来再次进行应废弃还是应受理(执行)的判定(步骤S200)。

[0121] 在此，说明发言时间长度判定部2043所使用的“发言时间长度”。所谓发言时间长度，是指发言的时间的长度。在此，将用户700为了对电视机20进行声音操作而进行的发言记为“控制用发言”，将不以电视机20的声音操作为目的的发言(用户700彼此的对话、用户700的自言自语等)记为“对话用发言”。

[0122] 在本实施方式中，识别词典175所登记的受理对象列表包含的受理对象词汇各自对应的发言时间长度数据(表示发言所需的时间的长度的数据)被预先存储在存储部270中。由此，发言时间长度判定部2043能够计算被选择为声音识别的结果的受理对象词汇的发言时间长度。另外，期望在该发言时间长度数据中加入发言速度的个人差异等而具有变动范围(range)。

[0123] 已确认“控制用发言”由1个单词或2个单词程度构成的情形较多。此外，这些单词(词汇)全部是受理对象列表所登记的受理对象词汇的可能性高。因此，如果对“控制用发言”进行声音识别，则基于被选择为声音识别的结果的受理对象词汇的发言时间长度数据的发言时间长度，与由声音获取部201创建的发言时间长度信息所表示的“控制用发言”的发言时间长度近似的可能性高。另外，在作为声音识别的结果而选择出多个受理对象词汇时，假设基于与这些多个受理对象词汇对应的发言时间长度数据来计算发言时间长度。

[0124] 另一方面，“对话用发言”由多个单词构成的情形较多，此外在这些单词(词汇)中包含与受理对象列表所登记的受理对象词汇对应的单词(词汇)的可能性低。因此，如果对“对话用发言”进行声音识别，则基于被选择为声音识别的结果的受理对象词汇的发言时间长度数据的发言时间长度，比由声音获取部201创建的发言时间长度信息所表示的“对话用发言”的发言时间长度要短的可能性高。

[0125] 鉴于这些内容，在声音识别处理装置200中，通过比对基于由声音识别部102选择为声音识别的结果的受理对象词汇的发言时间长度数据的发言时间长度、和基于由声音获取部201创建的发言时间长度信息的发言时间长度，从而能够判定成为声音识别的对象的声音是基于“控制用发言”的声音还是基于“对话用发言”的声音。并且，在本实施方式2中，由发言时间长度判定部2043来进行该判定。

[0126] 在步骤S200中，发言时间长度判定部2043基于作为声音识别的结果从识别结果获取部103输出的受理对象词汇，从存储部270中读出与该受理对象词汇建立了关联的发言时间长度数据。若接收到的受理对象词汇为多个，则发言时间长度判定部2043从存储部270中读出与这些受理对象词汇全部相关的发言时间长度数据。然后，基于读出的发言时间长度数据来计算发言时间长度。然后，比较该计算结果和由声音获取部201创建的发言时间长度信息所表示的发言时间长度。另外，发言时间长度判定部2043可以直接比较计算出的发言

时间长度和发言时间长度信息所表示的发言时间长度，也可以基于计算出的发言时间长度来设定用于判定的范围。在此，说明设定范围来进行比较的例子。

[0127] 在步骤S200中，若由声音获取部201创建的发言时间长度信息所表示的发言时间长度在基于计算出的发言时间长度而设定的范围外(否)，则发言时间长度判定部2043判定出从排斥词汇废弃部1042输出的未被赋予标记的字符串信息是基于“对话用发言”的信息，是应废弃的信息，对该字符串信息赋予标记并输出给受理废弃发送部1045。

[0128] 若对从发言时间长度判定部2043输入的字符串信息赋予了标记，则受理废弃发送部1045将该字符串信息作为废弃信息而输出给声音获取部201。接收到废弃信息的声音获取部201为了下次的声音识别而进行声音获取的准备(步骤S106)。

[0129] 另一方面，在步骤S200中，若由声音获取部201创建的发言时间长度信息所表示的发言时间长度在基于计算出的发言时间长度而设定的范围内(是)，则发言时间长度判定部2043将从排斥词汇废弃部1042输出的未被赋予标记的字符串信息判定为是基于“控制用发言”的信息，不对该字符串信息赋予标记地直接发送给发言方式判定部2044。

[0130] 另外，发言时间长度判定部2043例如也可以将计算出的发言时间长度设为给定倍(例如1.5倍)来设定用于判定的范围。该数值只不过为简单的一例，也可以为其他的数值。或者，发言时间长度判定部2043也可以将预先设定的数值加到计算出的发言时间长度上等来设定用于判定的范围，也可以利用其他的方法来设定范围。

[0131] 发言方式判定部2044针对从发言时间长度判定部2043输入的未被赋予标记的字符串信息，基于发言方式来再次进行应废弃还是应受理(执行)的判定(步骤S201)。

[0132] 在此，说明发言方式判定部2044所使用的“发言方式”。所谓该“发言方式”，是指在用户700即将发言之前所产生的无声音或者视作实质上无声音的期间(以下记为“停顿期间”)、以及用户700刚结束发言之后所产生的停顿期间。

[0133] 比较“控制用发言”和“对话用发言”的结果，确认出关于发言方式而存在差异。

[0134] 在“控制用发言”的情况下，在用户700发言的前后存在比“对话用发言”长的停顿期间。用户700即将发言之前所产生的停顿期间是用于准备发言的期间。用户700刚结束发言之后所产生的停顿期间是等待与发言的内容对应的动作(基于声音操作的动作)开始的期间。

[0135] 另一方面，在“对话用发言”的情况下，在用户700的发言的前后，这种停顿期间相对较少。

[0136] 因此，通过检测在发言的前后的停顿期间的长度，从而能够判定成为声音识别的对象的声音是基于“控制用发言”的声音还是基于“对话用发言”的声音。并且，在本实施方式2中，基于声音获取部201创建的发言方式信息，由发言方式判定部2044来进行该判定。

[0137] 在步骤S201中，发言方式判定部2044基于从发言时间长度判定部2043输出的受理对象词汇，从存储部270中读出与该受理对象词汇建立了关联的发言方式数据。所谓该发言方式数据，是表示在该受理对象词汇的发言的前后所产生的各停顿期间的长度的数据。在本实施方式中，与受理对象词汇建立了关联的发言方式数据被预先存储在存储部270中。然后，发言方式判定部2044比较从存储部270读出的发言方式数据、和从发言时间长度判定部2043输入的字符串信息所附加的发言方式信息(由声音获取部201创建的发言方式信息)。

[0138] 具体而言，发言方式判定部2044分别比较由声音获取部201创建的发言方式信息

所表示的发言前后的停顿期间的长度、和从存储部270读出的发言方式数据所表示的发言前后的停顿期间的长度。另外，发言方式判定部2044可以直接比较由声音获取部201创建的发言方式信息、和从存储部270读出的发言方式数据，也可以基于从存储部270读出的发言方式数据来设定用于判定的范围。另外，若接收到的受理对象词汇为多个，则发言方式判定部2044从存储部270中读出与所有这些受理对象词汇相关的发言方式数据，选择任意一个数值大的发言方式数据。或者，也可以选择任意一个数值小的发言方式数据，或者也可以计算平均值、中间值。

[0139] 在步骤S201中，若由声音获取部201创建的发言方式信息所表示的发言前后的停顿期间的长度的至少一方小于从存储部270读出的发言方式数据所表示的发言前后的停顿期间的长度(否)，则发言方式判定部2044将从发言时间长度判定部2043输出的未被赋予标记的字符串信息判定为是基于“对话用发言”的信息，对该字符串信息赋予标记并输出给受理废弃发送部1045。

[0140] 若对从发言方式判定部2044输入的字符串信息赋予了标记，则受理废弃发送部1045将该字符串信息作为废弃信息而输出给声音获取部201。接收到废弃信息的声音获取部201为了下次的声音识别而进行声音获取的准备(步骤S106)。

[0141] 另一方面，在步骤S201中，若由声音获取部201创建的发言方式信息所表示的发言前后的停顿期间的长度均为从存储部270读出的发言方式数据所表示的发言前后的停顿期间的长度以上(是)，则发言方式判定部2044将从发言时间长度判定部2043输出的未被赋予标记的字符串信息判定为是基于“控制用发言”的信息，不对该字符串信息赋予标记地直接输出给受理废弃发送部1045。

[0142] 由此，受理废弃发送部1045接收到的未被赋予标记的字符串信息成为在排斥词汇废弃部1042、发言时间长度判定部2043以及发言方式判定部2044中均未被赋予标记的字符串信息。换言之，若输入至受理废弃发送部1045的字符串信息未被赋予标记，则该字符串信息是在排斥词汇废弃部1042、发言时间长度判定部2043以及发言方式判定部2044中均判断为应受理(应执行指令处理)的字符串信息。另一方面，若输入至受理废弃发送部1045的字符串信息被赋予了标记，则其是在排斥词汇废弃部1042、发言时间长度判定部2043以及发言方式判定部2044的任意一者中判断为废弃信息的字符串信息。

[0143] 受理废弃发送部1045将未被赋予标记的字符串信息作为应受理(执行)的字符串信息而直接输出给指令处理部106。

[0144] 指令处理部106基于从受理废弃发送部1045接收到的字符串信息所表示的指示来执行指令处理(步骤S105)。

[0145] 在步骤S105结束后，指令处理部106向声音获取部201发送表示指令处理已结束的信号。接收到该信号的声音获取部201为了下次的声音识别而进行声音获取的准备(步骤S106)。

[0146] 在步骤S106中，赋予了标记的字符串信息作为废弃信息而从受理废弃发送部1045向声音获取部201输出。接收到废弃信息的声音获取部201为了下次的声音识别而进行声音获取的准备。

[0147] 另外，步骤S200和步骤S201哪方先执行均可。

[0148] [2-3.效果等]

[0149] 如以上，在本实施方式中，声音识别处理装置200具备：声音获取部201、识别结果判定部204和存储部270。声音获取部201基于获取到的声音，对用户700发言的时间的长度进行测定来创建发言时间长度信息。此外，声音获取部201基于获取到的声音，对在用户700的发言的前后所产生的无声音期间的长度进行测定来创建发言方式信息。存储部270中预先存储表示发言所需的时间的发言时间长度数据、和表示在发言的前后所产生的无声音期间的长度的发言方式数据。识别结果判定部204针对不含与排斥词汇一致的词语而被判定为应执行的字符串信息，从存储部270中读出发言时间长度数据，比较读出的发言时间长度数据和由声音获取部201创建的发言时间长度信息，基于该比较来再次进行应废弃还是应执行的判定。然后，针对被判定为应执行的字符串信息，从存储部270中读出发言方式数据，比较读出的发言方式数据和由声音获取部201创建的发言方式信息，基于该比较来再次进行应废弃还是应执行的判定。该字符串信息为第1信息的一例。

[0150] 在如此构成的声音识别处理装置200中，若输入至受理废弃发送部1045的字符串信息未被赋予标记，则其是在排斥词汇废弃部1042、发言时间长度判定部2043以及发言方式判定部2044中均判断为应受理(应执行指令处理)的字符串信息。另一方面，若输入至受理废弃发送部1045的字符串信息被赋予了标记，则其是在排斥词汇废弃部1042、发言时间长度判定部2043以及发言方式判定部2044的任意一者中判断为废弃信息的字符串信息。如此，在本实施方式中，针对作为声音识别的结果而从识别结果获取部103接收到的字符串信息，在排斥词汇废弃部1042、发言时间长度判定部2043以及发言方式判定部2044中分别判定应受理(指令处理)还是应废弃。并且，对于被任意一个判定为应废弃的字符串信息予以废弃，只有全部判定为应受理的字符串信息被进行指令处理。

[0151] 由此，在声音识别处理装置200中，能够精度良好地判定被声音识别的声音是基于“控制用发言”的声音还是基于“对话用发言”的声音，因此能够降低误识别，进一步提升声音识别的精度。

[0152] (其他实施方式)

[0153] 如以上，作为本申请中公开的技术例示，说明了实施方式1、2。然而，本公开中的技术并不限定于此，也能够应用于进行了变更、置换、附加、省略等之后的实施方式。此外，也能够组合上述实施方式1、2中所说明的各构成要素来作为新的实施方式。

[0154] 为此，以下例示其他实施方式。

[0155] 在实施方式2中，说明了识别结果判定部204除了具备排斥词汇废弃部1042之外还具备发言时间长度判定部2043和发言方式判定部2044来提高声音识别的精度的构成。但即使识别结果判定部是在排斥词汇废弃部1042上组合具备发言时间长度判定部2043和发言方式判定部2044的任意一方的构成，也能够提高声音识别的精度。

[0156] 图8A是表示其他实施方式中的识别结果判定部304的一构成例的框图。图8B是表示其他实施方式中的识别结果判定部404的一构成例的框图。

[0157] 另外，关于与实施方式1、2中所说明的电视机10、20具备的构成要素进行实质上相同的动作的构成要素，赋予与实施方式1、2相同的符号，并省略说明。

[0158] 图8A所示的识别结果判定部304是具备排斥词汇废弃部1042、发言时间长度判定部2043和受理废弃发送部1045而不具备发言方式判定部2044的构成。

[0159] 具备图8A所示的识别结果判定部304的声音识别装置按如下方式来动作。

[0160] 声音获取部(未图示)基于获取到的声音,对用户700发言的时间的长度进行测定来创建发言时间长度信息。存储部370中预先存储表示发言所需的时间的发言时间长度数据。该发言时间长度信息以及发言时间长度数据与实施方式2中所说明的发言时间长度信息以及发言时间长度数据实质上相同。

[0161] 识别结果判定部304针对不含与排斥词汇一致的词语而由排斥词汇废弃部1042判定为应执行的字符串信息,从存储部370中读出发言时间长度数据,比较读出的发言时间长度数据和由声音获取部创建的发言时间长度信息,基于该比较来再次进行应废弃还是应执行的判定。该字符串信息为第1信息的一例。

[0162] 识别结果判定部304具体按如下方式来动作。

[0163] 发言时间长度判定部2043针对从排斥词汇废弃部1042输入的未被赋予标记的字符串信息,基于发言时间长度来再次进行应废弃还是应受理(执行)的判定。

[0164] 发言时间长度判定部2043的动作与实施方式2中所说明的发言时间长度判定部2043实质上相同,因此省略说明。

[0165] 发言时间长度判定部2043对判定为是基于“控制用发言”的信息的字符串信息不赋予标记,直接输出给受理废弃发送部1045。受理废弃发送部1045将未被赋予标记的字符串信息作为应受理(执行)的字符串信息而直接输出给指令处理部106。

[0166] 图8B所示的识别结果判定部404是具备排斥词汇废弃部1042、发言方式判定部2044和受理废弃发送部1045而不具备发言时间长度判定部2043的构成。

[0167] 具备图8B所示的识别结果判定部404的声音识别装置按如下方式来动作。

[0168] 声音获取部(未图示)基于获取到的声音,对在用户700的发言的前后所产生的无声音期间的长度进行测定来创建发言方式信息。存储部470中预先存储表示在发言的前后所产生的无声音期间的长度的发言方式数据。该发言方式信息以及发言方式数据与实施方式2中所说明的发言方式信息以及发言方式数据实质上相同。

[0169] 识别结果判定部404针对不含与排斥词汇一致的词语而由排斥词汇废弃部1042判定为应执行的字符串信息,从存储部470中读出发言方式数据,比较读出的发言方式数据和由声音获取部创建的发言方式信息,基于该比较来再次进行应废弃还是应执行的判定。该字符串信息为第1信息的一例。

[0170] 识别结果判定部404具体按如下方式来动作。

[0171] 发言方式判定部2044针对从排斥词汇废弃部1042输入的未被赋予标记的字符串信息,基于发言方式来再次进行应废弃还是应受理(执行)的判定。

[0172] 发言方式判定部2044的动作与实施方式2中所说明的发言方式判定部2044实质上相同,因此省略说明。

[0173] 发言方式判定部2044对判定为是基于“控制用发言”的信息的字符串信息不赋予标记,直接输出给受理废弃发送部1045。受理废弃发送部1045将未被赋予标记的字符串信息作为应受理(执行)的字符串信息而直接输出给指令处理部106。

[0174] 识别结果判定部即便是例如图8A、图8B所示那样的仅具备发言时间长度判定部2043和发言方式判定部2044的任意一方的构成,也能够提升声音识别的精度。

[0175] 另外,在本实施方式中,虽然说明了声音识别服务器50配置在网络40上的例子,但声音识别服务器50也可以配备在声音识别处理装置100中。或者,也能够构成为不具备声音

识别服务器50而仅由声音识别部102来进行声音识别。

[0176] 另外，图2、图3、图5、图6、图8A、图8B所示的各模块分别可以作为独立的电路模块来构成，可以是由处理器来执行被编程为实现各模块的动作的软件的构成。

[0177] 产业上的可利用性

[0178] 本公开能够适用于执行用户利用声音来指示的处理动作的设备。具体而言，本公开能够适用于便携式终端设备、电视接收机、个人计算机、机顶盒、录影机、游戏机、智能手机、平板终端等。

[0179] 符号说明

[0180] 10,20 电视接收机

[0181] 11,21 声音识别处理系统

[0182] 40 网络

[0183] 50 声音识别服务器

[0184] 55,175 识别词典

[0185] 100,200 声音识别处理装置

[0186] 101,201 声音获取部

[0187] 102 声音识别部

[0188] 103 识别结果获取部

[0189] 104,204,304,404 识别结果判定部

[0190] 106 指令处理部

[0191] 130 内置话筒

[0192] 140 显示部

[0193] 150 收发部

[0194] 160 调谐器

[0195] 170,171,270,370,470 存储部

[0196] 202 指示器

[0197] 203 声音识别图标

[0198] 700 用户

[0199] 1042 排斥词汇废弃部

[0200] 1045 受理废弃发送部

[0201] 2043 发言时间长度判定部

[0202] 2044 发言方式判定部

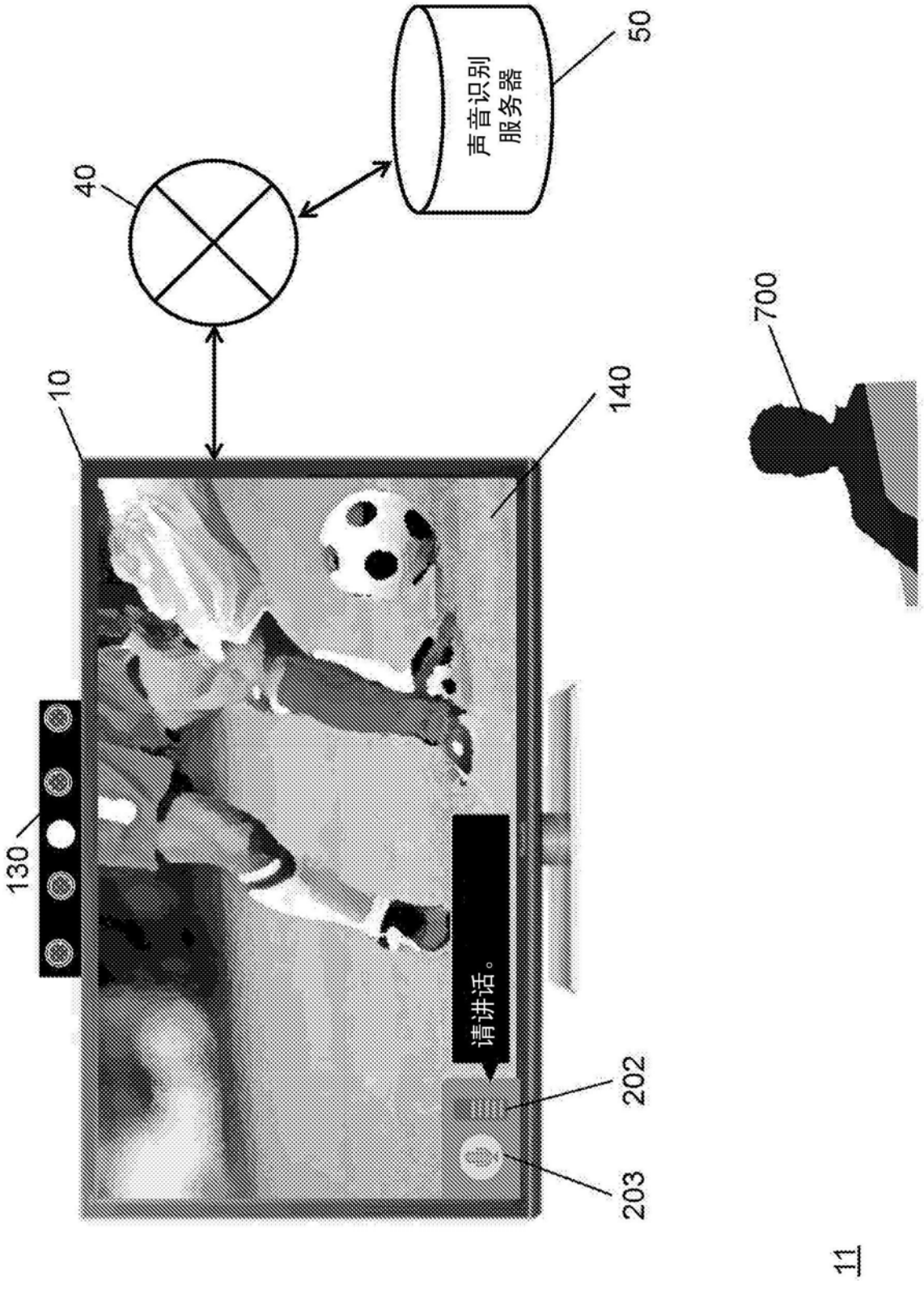


图1

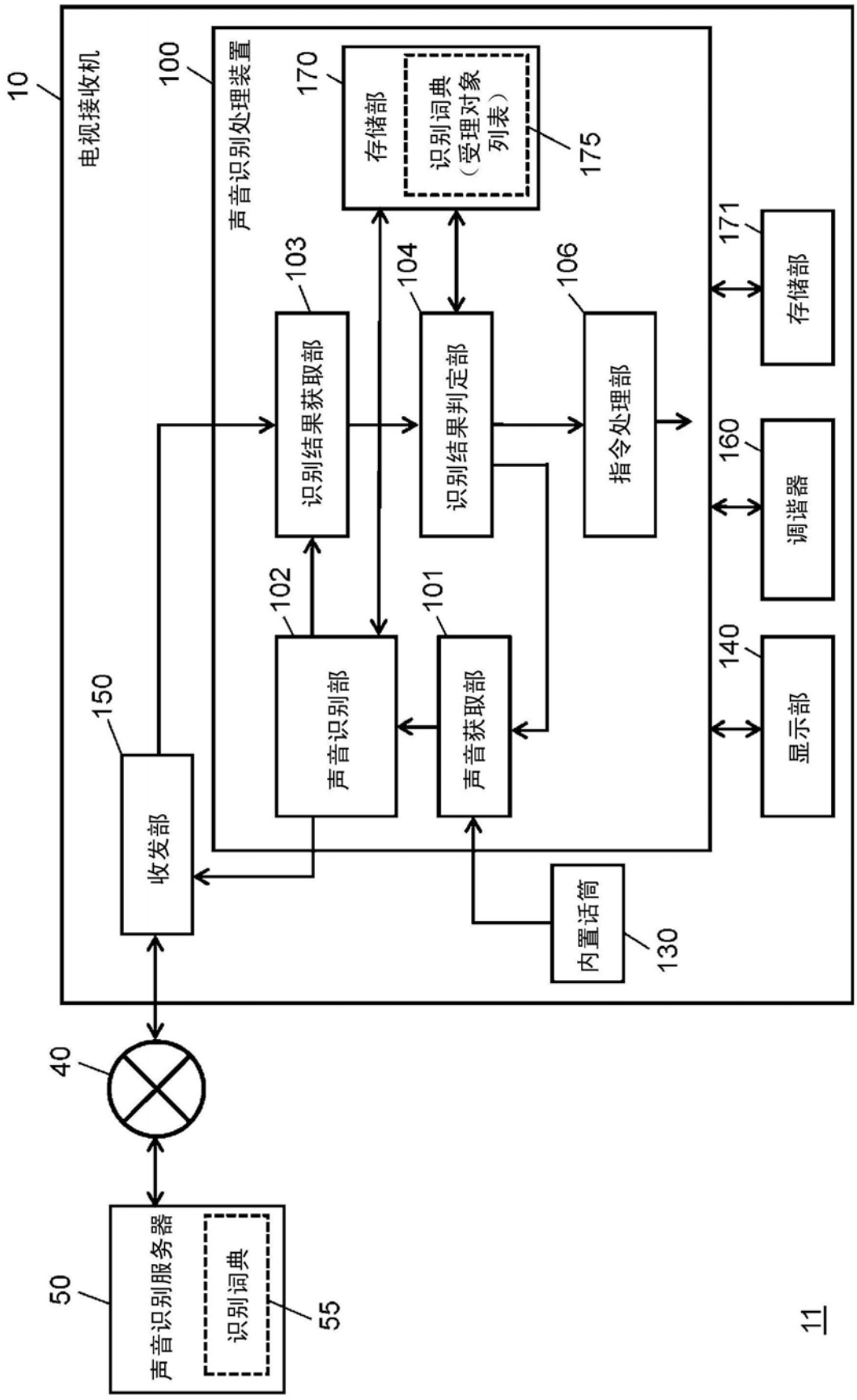

10
电视接收机
100
声音识别处理装置
170
存储部
识别词典（受理对象列表）
175
171
存储部
103
识别结果获取部
104
识别结果判定部
106
指令处理部
160
调谐器
102
声音识别部
101
声音获取部
140
显示部
150
收发部
内置话筒
130
40
50
声音识别服务器
识别词典
55
11

图2

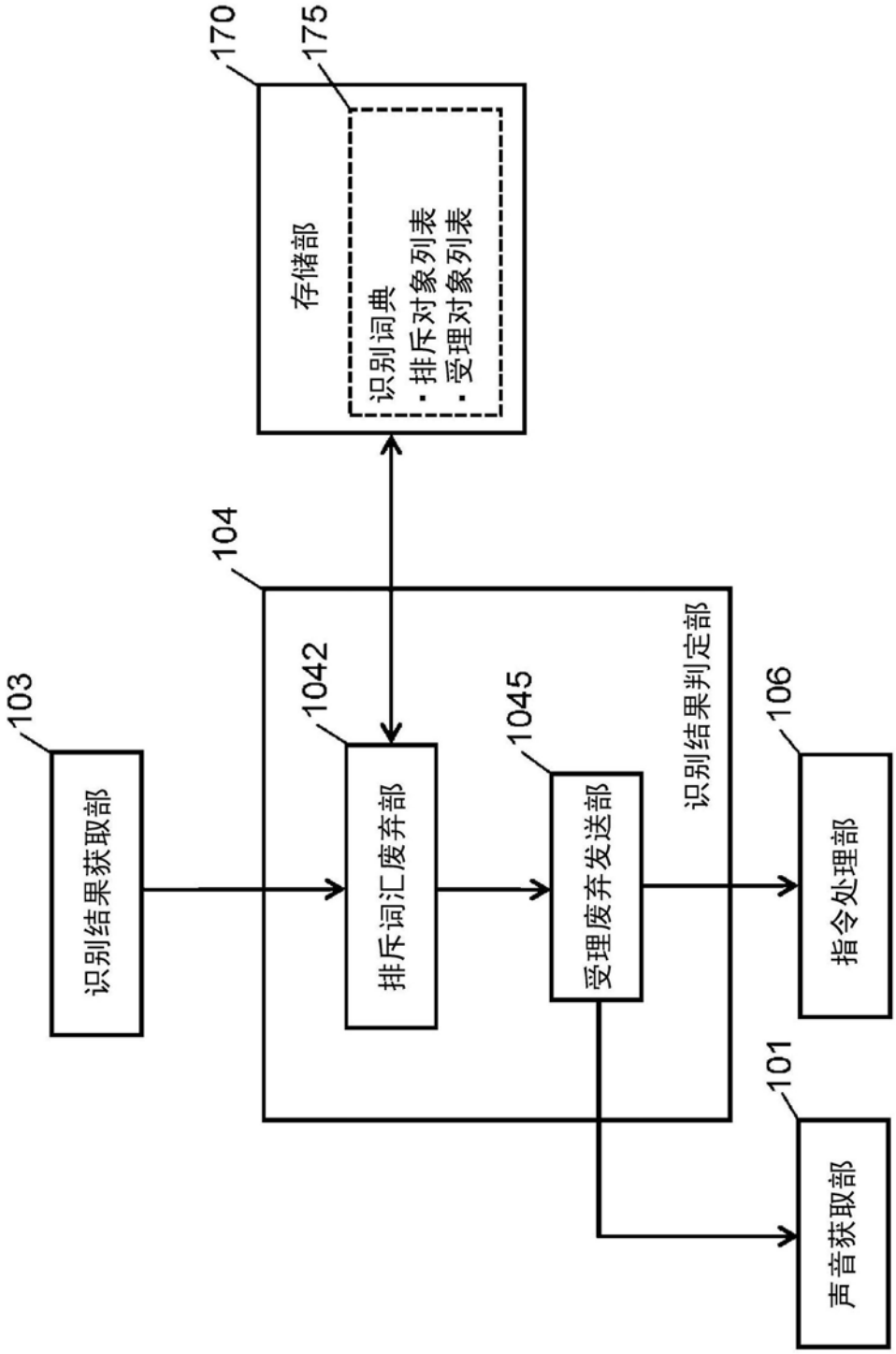

170
175
存储部
识别词典
・排斥对象列表
・受理对象列表
104
识别结果判定部
103
识别结果获取部
1042
排斥词汇废弃部
1045
受理废弃发送部
106
指令处理部
101
声音获取部

图3

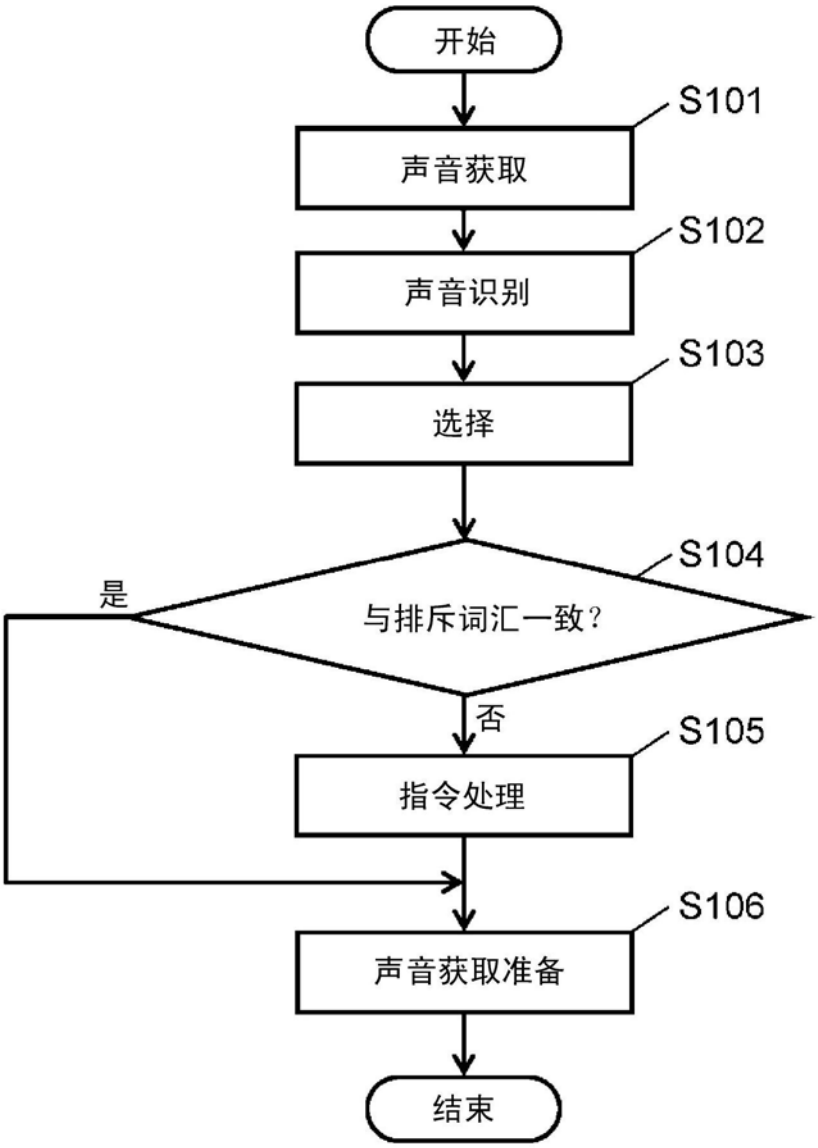

开始
S101
声音获取
S102
声音识别
S103
选择
S104
是
与排斥词汇一致?
否
S105
指令处理
S106
声音获取准备
结束

图4

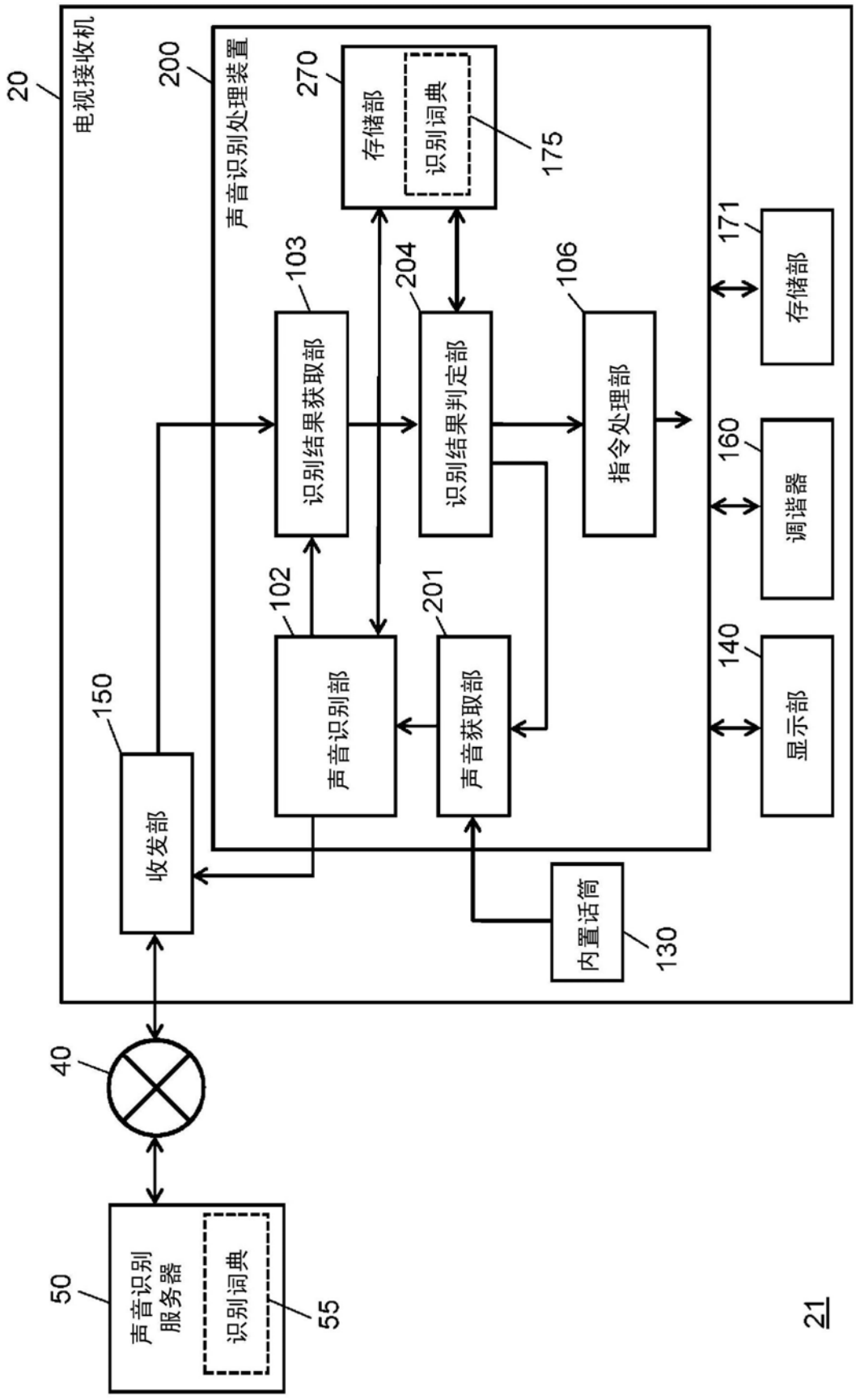

20
电视接收机
200
声音识别处理装置
270
存储部
识别词典
175
103
识别结果获取部
204
识别结果判定部
106
指令处理部
171
存储部
160
调谐器
140
显示部
102
声音识别部
201
声音获取部
150
收发部
内置话筒
130
40
50
声音识别
服务器
识别词典
55
21

图5

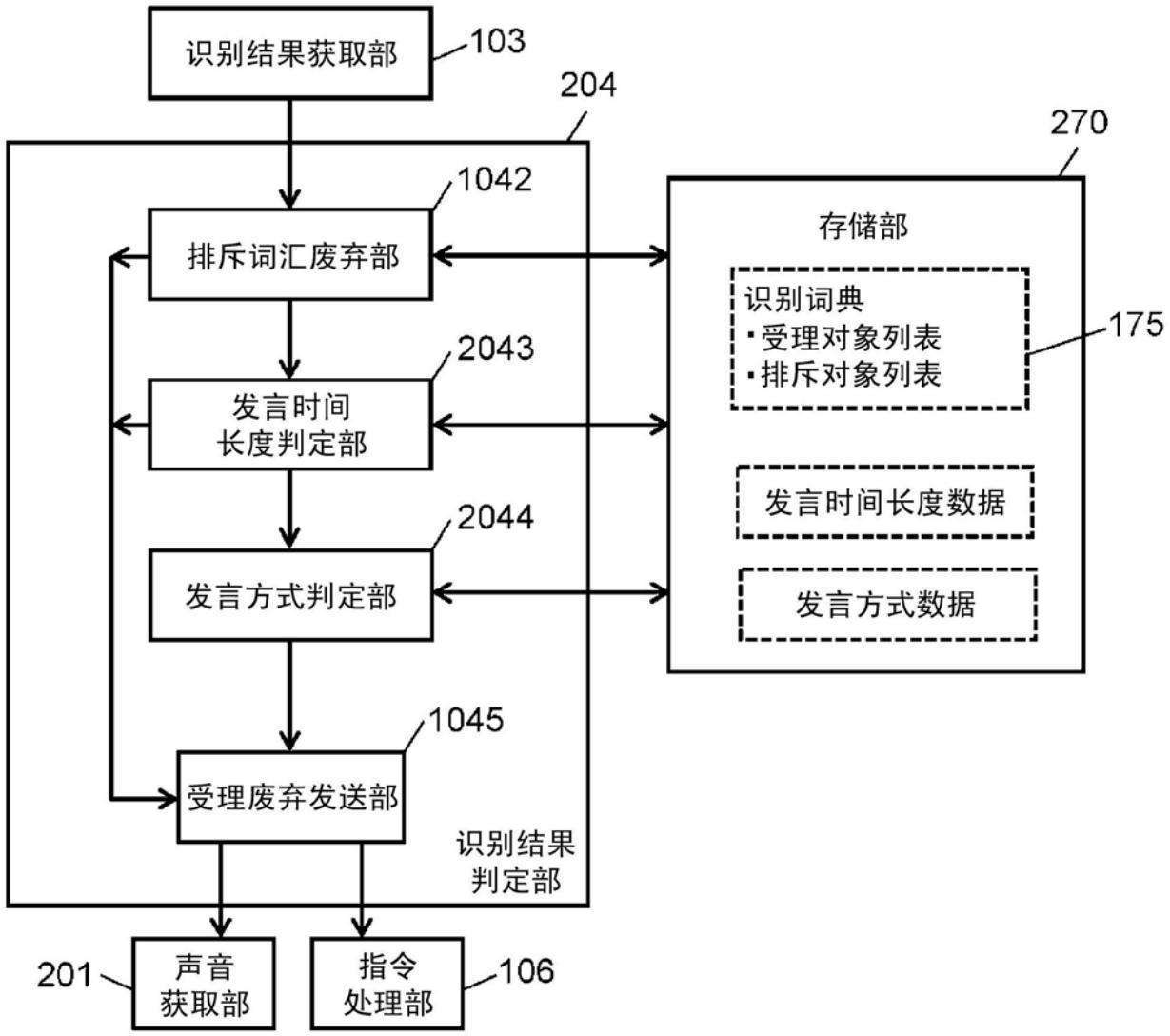

识别结果获取部
103
204
270
1042
排斥词汇废弃部
存储部
识别词典
·受理对象列表
·排斥对象列表
175
2043
发言时间
长度判定部
发言时间长度数据
2044
发言方式判定部
发言方式数据
1045
受理废弃发送部
识别结果
判定部
201
声音
获取部
指令
处理部
106

图6

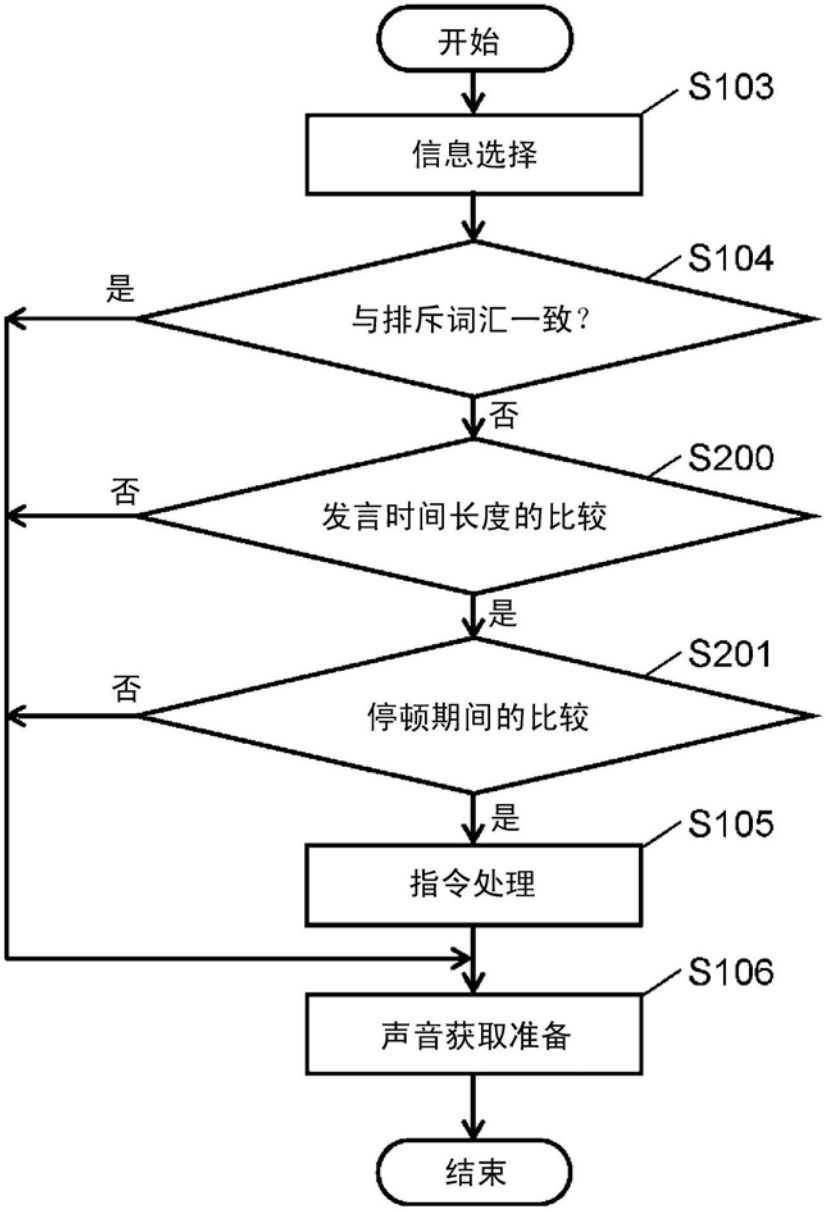

开始
S103
信息选择
S104
与排斥词汇一致？
是
否
S200
发言时间长度的比较
否
是
S201
停顿期间的比较
否
是
S105
指令处理
S106
声音获取准备
结束

图7

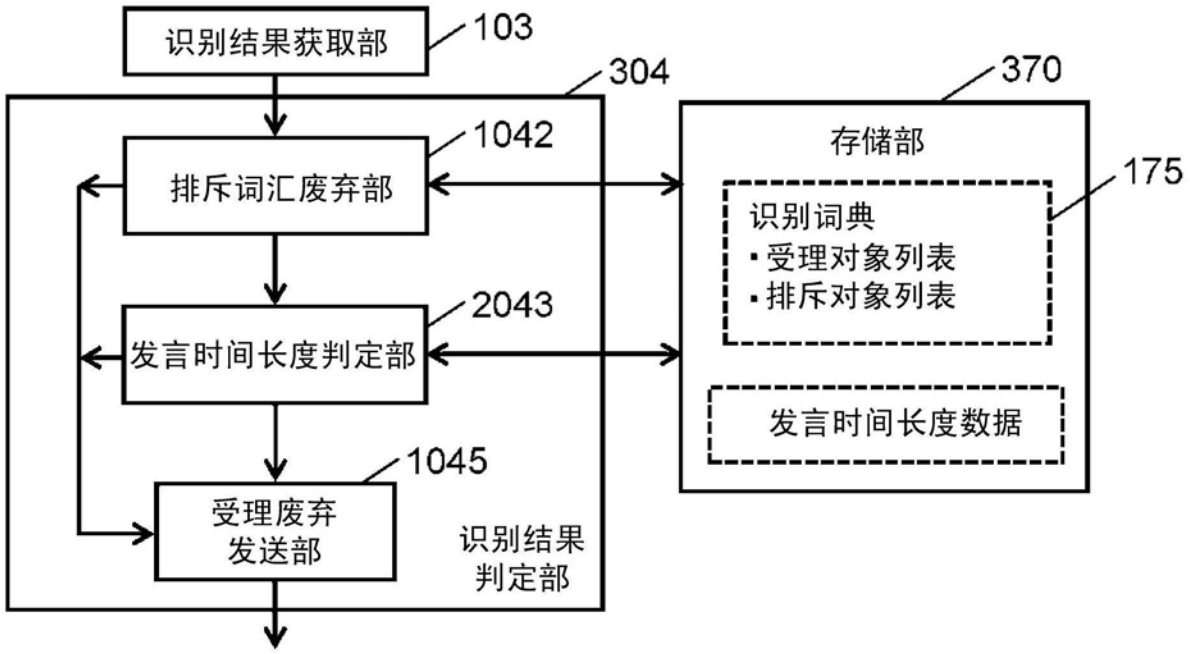

识别结果获取部
103
304
370
存储部
1042
排斥词汇废弃部
175
识别词典
·受理对象列表
·排斥对象列表
2043
发言时间长度判定部
发言时间长度数据
1045
受理废弃
发送部
识别结果
判定部

图8A

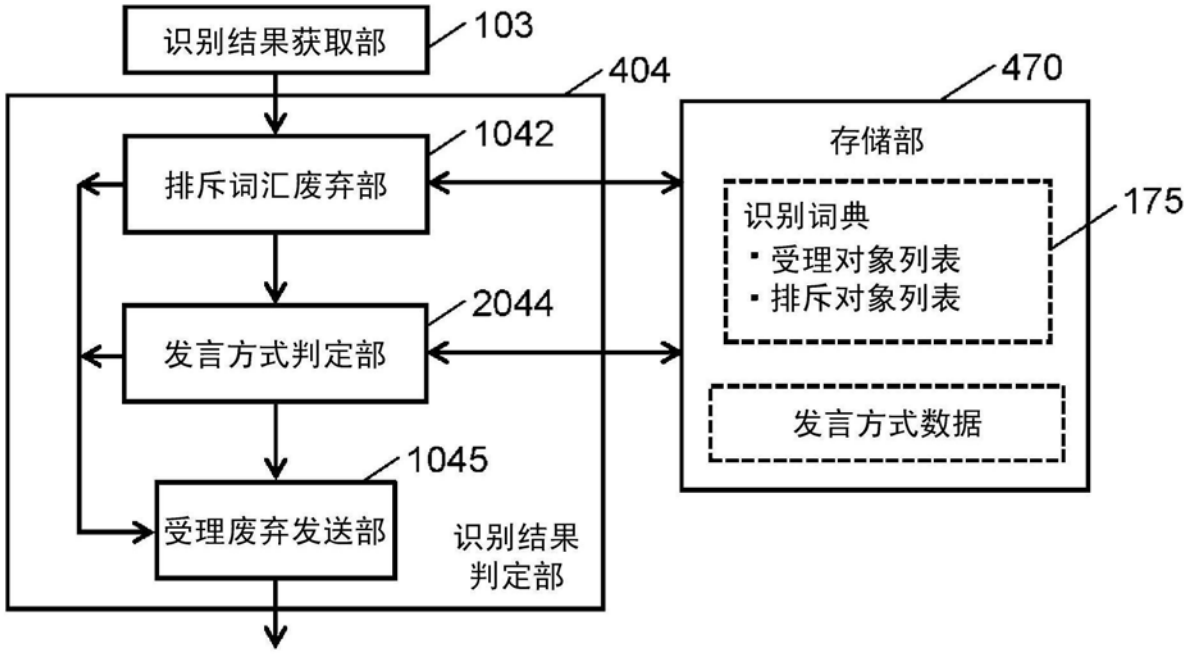

识别结果获取部
103
404
470
存储部
1042
排斥词汇废弃部
175
识别词典
·受理对象列表
·排斥对象列表
2044
发言方式判定部
发言方式数据
1045
受理废弃发送部
识别结果
判定部

图8B