[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 15/22 (2006.01)
G10L 15/28 (2006.01)
G06T 7/00 (2006.01)

[12] 发明专利说明书

专利号 ZL 03800225.6

[45] 授权公告日 2006 年 2 月 8 日

[11] 授权公告号 CN 1241168C

[22] 申请日 2003.3.5 [21] 申请号 03800225.6
[30] 优先权
[32] 2002. 3. 6 [33] JP [31] 60425/2002
[86] 国际申请 PCT/JP2003/002560 2003.3.5
[87] 国际公布 WO2003/075261 日 2003.9.12
[85] 进入国家阶段日期 2003.11.6
[71] 专利权人 索尼公司
地址 日本东京都
[72] 发明人 下村秀树 青山一美 山田敬一
浅野康治 大久保厚志
审查员 刘红梅

[74] 专利代理机构 北京市柳沈律师事务所
代理人 黄小临 王志森

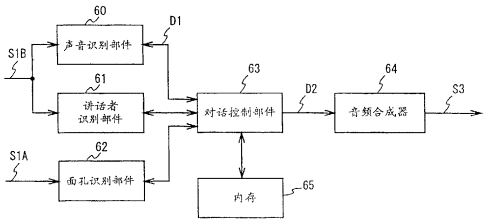
权利要求书 3 页 说明书 33 页 附图 19 页

[54] 发明名称

识别装置和识别方法，以及机器人设备

[57] 摘要

以往的机器人设备等不能自然地进行名字学习。学习一个对象的名字是按以下方式进行的：通过与人类对话来得到目标对象的名字，该名字与对于目标对象所检测到的多项不同特征数据相关联而存储，并基于所存储的数据和关联信息识别新对象，得到了新人的名字和特征数据并存储了该关联信息。



1. 一种基于所获得的目标对象的名字和特征的目标对象识别装置，包括:

对话装置，用来通过对话获得目标对象的名字;

数个识别装置，用来检测所述目标对象的数个特征数据，并基于检测结果和已知对象的对应特征数据来识别目标对象;

存储装置，用来存储关联信息，其将所述已知对象的名字与所述识别装置的识别结果互相关联的信息;

判断装置，所述判断装置通过参照存储在所述存储装置中的所述关联信息，由所述对话装置得到的所述目标对象的名字和所述识别装置对于对象的识别结果的多数决定，判断目标对象是不是新对象; 和

控制装置，用来当所述判断装置判断目标对象是新对象时，在对应的所述识别装置中存储所述目标对象的所述数个特征数据，并在所述存储装置中存储关于目标对象的关联信息。

2. 根据权利要求 1 所述的目标对象识别装置，其中

所述控制装置控制正确识别所述目标对象的所述识别装置，以当所述判断装置判断目标对象是所述已知对象时进行积累学习。

3. 根据权利要求 1 所述的目标对象识别装置，其中

所述控制装置控制未正确识别所述目标对象的所述识别装置，以当所述判断装置判断目标对象是所述已知对象时进行纠正学习。

4. 根据权利要求 1 所述的目标对象识别装置，其中

所述控制装置控制所述对话装置按需要延长所述对话。

5. 一种基于所获得的目标对象的名字和特征的目标对象识别方法，包括:

对话步骤，其通过对话获得目标对象的名字;

数个识别步骤，其检测所述目标对象的数个特征数据，并基于检测结果和已知对象的对应特征数据来识别目标对象;

存储步骤，其存储关联信息，其将所述已知对象的名字与所述识别装置的识别结果互相关联;

判断步骤，在所述判断步骤中，通过参照所述关联信息，由所述目标对

象的名字和所述特征的识别结果的多数决定，判断目标对象是不是新对象；和

控制步骤，其当所述判断装置判断目标对象是新对象时，在对应的所述识别装置中存储所述目标对象的所述数个特征数据，并在所述存储装置中存储关于目标对象的关联信息。

6. 根据权利要求 5 所述的目标对象识别方法，其中

在所述控制步骤中，当所述判断装置判断目标对象是所述已知对象时，对于正确识别的目标对象的所述特征进行积累学习。

7. 根据权利要求 5 所述的目标对象识别方法，其中

在所述控制步骤中，当所述判断装置判断目标对象是所述已知对象时，对于未正确识别的目标对象的所述特征进行纠正学习。

8. 根据权利要求 5 所述的目标对象识别方法，其中

在所述对话步骤中，按需要延长所述对话。

9. 一种机器人设备包括：

对话装置，用来通过对话获得目标对象的名字；

数个识别装置，用来检测所述目标对象的数个特征数据，并用来基于检测结果、和已知对象的对应特征数据来识别目标对象；

存储装置，用来存储关联信息，其将所述已知对象的名字与所述识别装置的识别结果互相关联；

判断装置，用来基于由所述对话装置获得的所述目标对象的名字、所述识别装置对于所述目标对象的识别结果、和存储在所述存储装置中的关联信息，判断目标对象是不是新对象；和

控制装置，用来当所述判断装置判断目标对象是新对象时，在对应的所述识别装置中存储所述目标对象的所述数个特征数据，并在所述存储装置中存储关于目标对象的关联信息。

10. 根据权利要求 9 所述的机器人，其中

所述控制装置控制正确识别所述目标对象的所述识别装置，以当所述判断装置判断目标对象是所述已知对象时，进行积累学习。

11. 根据权利要求 9 所述的机器人，其中

所述控制装置控制未正确识别所述目标对象的所述识别装置，以当所述判断装置判断目标对象是所述已知对象时进行纠正学习。

12. 根据权利要求 9 所述的机器人，其中

所述判断装置通过参照存储在所述存储装置中的所述关联信息，由所述对话装置得到的所述目标对象的名字和所述识别装置对于对象的识别结果的多数决定，判断目标对象是不是新对象。

13. 根据权利要求 9 所述的机器人，其中

所述控制装置控制所述对话装置按需要延长所述对话。

识别装置和识别方法，以及机器人设备

技术领域

本发明涉及学习器材和学习方法，以及机器人设备，并最适用于诸如娱乐机器人。

背景技术

近年，已开发了一定数量用于商业目的的家用娱乐机器人，用于商业化用途。其中一些娱乐机器人被装以诸如 CCD（电荷耦合器件）照相机和麦克风等各种外传感器，从而被设计成基于此外传感器的输出而识别外部环境，并基于识别结果而自主活动。

若这种娱乐机器人能记住新对象（包括人类，下同）的与之关联的名字，则他们就更能游刃有余地与用户沟通，此外，他们还能够对例如由用户下达的“踢球”等各种与对象有关的指令而不仅是事先注册了名字的对象灵活地做出反应。注意到，如上述记住对象的与之关联的名字被表达成“学习名字”，而以下将这种功能称作“名字学习功能”。

进而，若通过在娱乐机器人中提供这种名字学习功能、就像人类会做的那样，以使娱乐机器人能通过对话来学习新对象的名字，则从贴近自然的角度看是上佳的，并能期待娱乐机器人的娱乐特性可增加更多。

以往的技术中存在的问题却是：难以让娱乐机器人判断摆在它面前的新对象的名字应不应该学习。

有鉴于此，在以往的技术中，用户下达一条清晰的话音指令或按下特定的触觉传感器、以将操作模式变成注册模式，从而使对象被识别并被注册其名字。然而，当考虑到用户与娱乐机器人之间的自然互动时，却存在以下问题：与这种清晰指示而响应的名字注册却十分不自然。

发明内容

本发明是考虑了以上几点而做出的，目的在于提出一种学习器材和学习方法，以及机器人设备，其可大大增强娱乐特性。

为了解决那些问题，在本发明中，学习器材包括：对话装置，其具有与人类对话的能力，用来通过对话从人类获得目标对象的名字；数个识别装置，用来检测目标对象的规定的不同特征，并用来基于检测结果、和与事先存储的已知对象对应的特征数据来识别目标对象；存储装置，用来存储关联信息，其中已知对象的名字与由识别装置获得的关于对象的识别结果互相关联；判断装置，用来基于由对话装置获得的目标对象的名字、由识别装置获得的目标对象的识别结果、和存储在存储装置中的关联信息，判断目标对象是不是新对象；和控制装置，用来当判断装置判断目标对象是新对象时，让识别装置存储目标对象的特征的对应数据，并让存储装置存储关于目标对象的关联信息。

结果，此学习器材可自然地通过与凡人的对话来学习新人、新对象等的名字，就像人类常做的那样，而不必为了响应给出声音指令或按下触觉传感器等用户的清晰指示而注册名字了。

同样，在本发明中，学习方法包括：第一步，与人类对话，并通过对话从人类获得目标对象的名字，以及检测目标对象的数个规定的不同特征，并基于检测结果、和事先存储的已知对象的特征数据来识别目标对象；第三步，基于：所获得的目标对象的名字、以目标对象各特征为基础的识别结果、和将事先存储的已知对象的名字与由识别装置产生的关于对象的识别结果相关联的关联信息，判断目标对象是不是新对象；和第四步，当判断目标对象是新对象时，存储目标对象的各特征的数据、和关于目标对象的关联信息。

结果，根据此学习方法，能够自然地通过与凡人的对话来学习新人、新对象等的名字，就像人类常做的那样，而不必为了响应给出声音指令或按下触觉传感器等用户的清晰指示而注册名字了。

进而，在本发明中，机器人设备包括：对话装置，其具有与人类对话的能力，用来通过对话从人类获得目标对象的名字；数个识别装置，用来检测目标对象的规定的不同特征，并用来基于检测结果、和与事先存储的已知对象对应的特征数据来识别目标对象；存储装置，用来存储关联信息，其将已知对象的名字与由识别装置获得的关于对象的识别结果相关联；判断装置，用来基于由对话装置获得的目标对象的名字、由识别装置获得的目标对象的识别结果、和存储在存储装置中的关联信息，判断目标对象是不是新对象；和控制装置，用来当判断装置判断目标对象是新对象时，让识别装置存储目

标对象的特征的对应数据，并让存储装置存储关于目标对象的关联信息。

结果，此机器人设备可自然地通过与凡人的对话来学习新人、新对象等的名字，就像人类常做的那样，而不必为了响应给出声音指令或按下触觉传感器等用户的清晰指示而注册名字了。

附图说明

图 1 是表示本实施例中机器人的外部构造的透视图；

图 2 是表示本实施例中机器人的外部构造的透视图；

图 3 是用于解释本实施例中机器人的外部构造的原理图；

图 4 是用于解释本实施例中机器人的内部构造的原理图；

图 5 是用于解释本实施例中机器人的内部构造的原理图；

图 6 是用于解释主控制部件 40 有关名字学习功能的处理的框图；

图 7 是用于解释将 FID 和 SID 与内存中名字关联的概念图；

图 8 是表示名字学习处理例程的流程图；

图 9 是表示名字学习处理例程的流程图；

图 10 是表示名字学习处理中对话示例的原理图；

图 11 是表示名字学习处理中对话示例的原理图；

图 12 是用于解释 FID 和 SID 与名字的新注册的概念图；

图 13 是表示字学习处理中对话示例的原理图；

图 14 是表示字学习处理中对话示例的原理图；

图 15 是表示声音识别部件的构成框图；

图 16 是用于解释辞典的框图；

图 17 是用于解释语法规则的概念图；

图 18 是用于解释存储在特征向量缓冲中的内容的概念图；

图 19 是用于解释积分单的概念图；

图 20 是表示声音识别处理例程的流程图；

图 21 是表示未注册词处理例程的流程图；

图 22 是表示音群分割处理例程的流程图；

图 23 是表示仿真结果的概念图；

图 24 是表示在学习中的面孔识别部件的构成框图；

图 25 是表示在识别中的面孔识别部件的构成框图。

具体实施方式

以下，参照附图来详细说明实施本发明的一种方式。

(1)本实施例中的机器人的构造

在图 1 和 2 中，序号 1 表示本实施例中的两足直立行走机器人的全体，其中头部 3 置于躯干部 2 上，同属该构造的臂部 4A、4B分别摆放在躯干部 2 的左上和右上侧，而同属该构造的腿部 5A、5B则分别摆放在躯干部 2 的左下和右下侧。

躯干部 2 由形成上半身的框架 10 和形成下半身的腰基 11 构成，此二者皆经腰关节机构 12 连接，并将上半身设计成通过驱动固定在下半身的腰基 11 上的腰关节机构 12 的马达A_1、A_2，可绕如图 3 所示互相正交的前后向轴 13 和左右向轴 14 独立旋转。

而且，头部 3 固定在肩基 15 的上部中央，该肩基 15 经颈关节机构 16 固定在框架 10 的上端，并将该头部设计成通过驱动颈关节机构 16 的马达A_3、A_4，能够绕如图 3 所示互相正交的左右向轴 17 和上下向轴 18 独立旋转。

进而，臂部 4A、4B经肩关节机构 19 分别固定在肩基 15 的左右，并被设计成通过驱动对应的肩关节机构 19 的马达A_5、A_6，能够绕如图 3 所示互相正交的左右向轴 20 和前后向轴 21 独立旋转。

在此情形下，对于各臂部 4A、4B，形成上臂的马达A_7的外向轴经肘关节机构 22 链接形成前臂的马达A_8，而手部附加至前臂的前端。

并且，对于臂部 4A、4B，其前臂被设计成通过驱动马达A_7绕图 3 所示的上下向轴 24 旋转，并通过驱动马达A_8绕图 3 所示的左右向轴 25 旋转。

另一方面，各腿部 5A、5B经臀关节机构 26 附加至下半身的腰基 11，并被设计成通过驱动对应的臀关节机构 26 的马达A_9~A_{11}，能够绕如图 3 所示互相正交的上下向轴 27、前后向轴 28 和左右向轴 29 独立旋转。

在此情形下，将构造设计成使得对于各腿部 5A、5B，形成小腿的框架 32 经膝关节机构 31 链接形成大腿的框架 30 的下端，而足部 34 经踝关节机构 33 链接框架 32 的下端。

因此，对于腿部 5A、5B，其小腿被设计成通过驱动形成膝关节机构 31 的马达A_{12}能够绕图 3 所示的左右向轴 35 旋转，而其足部 34 通过驱动形成踝关节机构 33 的马达A_{13}~A_{14}，能够绕如图 3 所示互相正交的左右向轴 36 和前后向轴 37 独立旋转。

另一方面，如图 4 所示，在形成躯干部 2 下半身的腰基 11 背面，设有控制部件 42 的小盒，其内装有主控制部件 40，用来控制整个机器人的整个动作，包括电源电路和通信电路的周边电路 41 和电池 45（图 5）等。

并且，此控制部件 42 连接至子控制部件 43A～43D、其设在各组成部位（躯干部 2、头部 3、臂部 4A，4B、和腿部 5A，5B）内，并将该控制部件设计成能够进行向这些子控制部件 43A～43D 提供必要的电源电压以及与这些子控制部件 43A～43D 通讯。

而且，各子控制部件 43A～43D 连接至对应组成部件的各马达A_1～A_{14}，这些子控制部件 43A～43D 被设计成能够以从主控制部件 40 给出的各种指令所指定的方式来驱动对应组成部件的各马达A_1～A_{14}。

进而，如图 5 所示，在头部 3 上的所定位置设有组件，例如外传感器部件 53，其由充当机器人 1 的“眼”的 CCD（电荷耦合器件）照相机 50 和充当机器人 1 的“耳”的麦克风 51，以及触觉传感器 52，和充当“口”的扬声器 54，而在控制部件 42 内部设有由电池传感器 55 和加速传感器 56 组成的内传感器部件 57。

并且，外传感器部件 53 的 CCD 照相机 50 摄取周遭环境，而所摄视频信号 S1A被发送至主控制部件，同时麦克风 51 收集的诸如作为声音输入的用户语音，其指示“行走”、“躺倒”和“追球”等各种指令，并将得到的音频信号 S1B发送至主控制部件 40。

而且，从图 1 和图 2 看很显然，触觉传感器 52 处于头部 53 的顶上，它检测由用户施加的“敲”和“拍”等物理作用而产生的压力，而该检测结果作为压力检测信号 S1C 被发送至主控制部件 40。

进而，内传感器部件 57 中的电池传感器 55 在所定间隙处检测电池 45 的能量水平，而该检测结果作为电池水平检测信号 S2A被发送至主控制部件 40，同时加速传感器 56 在所定间隙处检测三轴（x 轴，y 轴，和 z 轴）向的加速，而该检测结果作为加速检测信号 S2B被发送至主控制部件 40。

主控制部件 40 基于分别从外传感器部件 53 的 CCD 照相机 50、麦克风 51、触觉传感器 52 等分别供给视频信号 S1A、音频信号 S1B、压力检测信号 S1C 等（以下将他们合称做“外传感器信号 S1”），以及分别从内传感器部件 57 的电池传感器 55、加速传感器等分别供给的电池水平检测信号 S2A、加速检测信号 S2B等（以下将他们合称做“内传感器信号 S2”），判断机器人 1 的周

围和内部状况、来自用户的指令、以及来自用户的影响的存在等。

主控制部件 40 基于判断结果、事先存储在内存 40A中的控制程序、和存储在已安装的外存 58 中的各种控制参数，确定后续行动，并基于确定结果向相关子控制部件 43A～43D 发送控制指令。结果，对应的马达A_1～A_{14}基于控制指令并在子控制部件 43A～43D 的控制下被带动，从而让机器人 1 行动，例如抬头或低头、左转或右转头部 3，举起臂部 4A、4B，和行走。

在此关头，主控制部件 40 还按需要向扬声器 54 馈送所定音频信号 S3，以基于音频信号 S3 而输入声音，或向置于头部 3 所定位置处的充当“眼”外形的 LED 馈送驱动信号，以使 LED 闪烁。

于是机器人 1 被设计成能够基于周遭和内部状况、指令、来自用户的影响而自主举止。

(2)主控制部件 40 有关名字学习功能的处理

其次，对安装在机器人 1 上的名字学习功能给出解释。

此机器人 1 安装有名字学习功能，以得到并学习与人关联的名字（该过程以下称做“名字学习”），其方式是：通过与人对话得到此人的名字，并基于来自麦克风 51 和 CCD 照相机 50 的输出，存储与语音的各声学特征和检测到的人的形貌特征相关联的名字，接着，基于已存储的数据发现未曾得到其名的新的出场人，以上述同样方式得到并存储名字、语音的声学特征和新人的形貌特征。注意到以下将与语音的声学特征和此人的形貌特征相关联而存储其名的人称做“熟人”，而未曾存储其名的人称做“新人”。

并且，此名字学习功能是由主控制部件 40 中的各种处理来实现的。

此处，主控制部件 40 有关名字学习功能的处理内容可按功能划分如下，如图 6 所示，声音识别部件 60，用来识别人朗读的词；讲话者识别部件 61，用来检测人的语音的声学特征，并用来基于检测到的声学特征识别和识别人；面孔识别部件 62，用来检测人面孔的形貌特征，并用来基于检测到的形貌特征识别和识别人；对话控制部件 63，其负责用于新人名字学习的各种控制，包括控制与人对话，还负责名字、语音的声学特征和熟人面孔的形貌特征的存储管理；以及声音合成器 64，用来生成并向扬声器 54（图 5）发送音频信号 S3，用于对话控制部件 63 控制下的各种对话。

在此情形下，声音识别部件 60 的功能是基于来自麦克风 51（图 5）的音频信号 S1B、通过执行所定的声音识别处理而逐词识别音频信号 S1B中的所

含有的词，并被设计成将所识别的词作为字符串数据 D1 发送至对话控制部件 63。

而且，讲话者识别部件 61 的功能是检测人的语音的声学特征，其来自麦克风所供给的音频信号 S1B，这需利用在例如“隔离需识别的讲话者和讲话者识别（CH2977－7／91／0000～0837S1.00 1991 IEEE）”中记载的方法而进行所定的信号处理。

并且，在平常时候，讲话者识别部件 61 顺次将检测到的声学特征的数据与全部已存储的熟人的声学特征的数据做比较，并当在该时刻检测到的声学特征与任何熟人一致时，讲话者识别部件 61 将特定识别符（以下称做“SID”）通知给对话控制部件 63，该识别符附加了与熟人的声学特征关联的声学特征，而当检测到的声学特征与任何熟人皆不一致时，将表示无法识别的 SID(＝－1)传达给对话控制部件 63。

而且，当控制部件 63 识别人为新人时，讲话者识别部件 61 根据由对话控制部件 63 给的新学的开始和结束命令的时间周期的期间，检测语音的声学特征，并且，检测的语音的声学特征存储在相关联的新的 SID 内，该 SID 被传送到对话控制部件 63。

注意到讲话者识别部件 61 被设计成能够进行积累学习，即积累地收集人的语音的声学特征，以及纠正学习，即纠正人的语音的声学特征，以响应从对话控制部件 63 给出的积累学习或纠正学习的起始和结束指令，从而正确地识别人。

面孔识别部件 62 的功能是一直注视着 CCD 照相机 50（图 5）所给的视频信号 S1A，并基于视频信号 S1A以规定的信号处理检测图像中所含的人面孔的外貌特征。

接着，在平常时候，面孔识别部件 62 顺次将检测到的形貌特征的数据与全部已存储的熟人的声学特征的数据做比较，并当在该时刻检测到的形貌特征与任何熟人一致时，面孔识别部件 62 将特定识别符（以下称做“FID”）通知给对话控制部件 63，该识别符附加到与熟人的形貌特征关联的形貌特征，而当检测到的形貌特征与任何熟人皆不一致时，将表示无法识别的 FID(＝－1)传达给对话控制部件 63。

而且，当对话控制部件 63 判断此人是新人时，面孔识别部件 62 检测图像所含的人面孔的形貌特征，该图像基于从 CCD 照相机 50 给出的视频信号

S1A，且根据从对话控制部件 63 给出的学习起始指令和学习结束指令的期间的图像。而检测到的形貌特征与新的特定 FID 关联而存储，并将此 FID 传达给对话控制部件 63。

注意到面孔识别部件 62 被设计成能够进行积累学习，即积累地收集人面孔的形貌特征，以及纠正学习，即纠正人面孔的形貌特征的数据，以响应从对话控制部件 63 给出的积累学习或纠正学习的起始和结束指令，从而正确地识别人。

声音合成器 64 的功能是将从对话控制部件 63 给出的字符串数据 D2 变换成音频信号 S3，而如此得到的音频信号 S3 被发送至扬声器 54（图 5）。因此，基于音频信号 S3 的声音被设计成可由扬声器 54 输出。

如图 7 所示，对话控制部件 63 具有内存 65（图 6），以存储熟人的名字和与存储在讲话者识别部件 61 中的人的语音的声学特征的数据相关联的 SID，其涉及与存储在面孔识别部件 62 中的人面孔的形貌特征的数据相关联的 FID。

接着对话控制部件 63 被设计成在适宜时刻给予声音合成器 64 所定的字符串数据 D2，以从扬声器 54 输出声音，从而询问谈话对方的名字或确认他的名字，并根据基于从此人在此刻的响应等的由声音识别部件 60 和讲话者识别部件 61 而产生的识别结果和由面孔识别部件 62 产生的此人的识别结果，以及上述熟人的名字和存储在内存 65 中的 SID 和 FID 的关联信息，判断此人是不是新人。

此后，当判断此人是新人时，对话控制部件 63 通过给予讲话者识别部件 61 和面孔识别部件 62 用于新学习的起始指令和结束指令，让讲话者识别部件 61 和面孔识别部件 62 收集和存储新人的语音的声学特征和面孔的形貌特征，结果是与分别从讲话者识别部件 61 和面孔识别部件 62 给出的新人的语音的声学特征和面孔的形貌特征相关联的 SID 和 FID 被存储在涉及从对话中得到的人名的内存 65。

而且，当判断此人是熟人时，对话控制部件 63 通过给出如要求的用于积累学习和纠正学习的起始指令，让讲话者识别部件 61 和面孔识别部件 62 进行必要的积累学习和纠正学习，同时对话控制部件 63 被设计成进行对话控制，从而延续此人的聊侃，直到讲话者识别部件 61 和面孔识别部件 62 通过在适宜时刻顺次地将所定的字符串数据 D2 发送至声音合成器 64，而能够收

集适量的用于积累学习和纠正学习的必需数据。

(3)对话控制部件 63 有关名字学习功能的处理的具体处理

其次，对于由对话控制部件 63 处理的有关名字学习功能的具体内容给出解释。

对话控制部件 63 进行各种处理，用来根据如图 8 和图 9 所述的名字学习处理例程 RT1、基于存储在外存 58（图 5）中的控制程序依次学习新人的名字。

即，当面孔识别部件 62 给出 FID 时，由于面孔识别部件 62 基于来自 CCD 照相机 50 的视频信号 S1A识别人面孔，故对话控制部件 63 在步骤 SP0 处开始名字学习处理例程，并在下一步 SP1 处，基于存储在内存 65 中涉及带有对应 SID 和 FID 的熟人名字的信息（以下称之为“关联信息”），判断有无可能搜索对应于 FID 的名字（即，判断 FID 是不是意味着无法识别的“－1”）。

那末，在步骤 SP1 处得到肯定的结果意味着此人是熟人，带有存储在面孔识别部件 62 中的人面孔的形貌特征，并带有与存储在内存 65 中涉及此人名字的数据所对应的 FID。然而，在此情形下，仍然可以想到：面孔识别部件 62 可能会将新人错认成熟人。

接着，当在步骤 SP1 处得到肯定的结果时，处理前进至步骤 SP2，在此对话控制部件 63 将所定的字符串数据 D2 发送至声音合成器 64，从而让扬声器 54 输出问询的声音，例如图 10 所示的“阁下是某某君吗？”，以确认此人的名字是否与使用 FID 检测到的名字（对应于上例中的某某君）一致。

此后，处理前往步骤 SP3，在此对话控制部件 63 等候此人回答问题的声音识别结果，例如“是，我是。”或“不，我不是。”，其可望来自声音识别部件 60。接着，当这一声音识别结果从声音识别部件 60 给出时、或当此刻讲话者识别结果的 SID 从讲话者识别部件 61 给出时，处理前进至步骤 SP4，在此对话控制部件 63 基于来自声音识别部件 60 的声音识别结果，判断此人的回答是不是肯定性的。

在步骤 SP4 处获得肯定的结果意味着基于在步骤 SP1 处从面孔识别部件 62 给出的 FID 而检索的名字与此人的名字一致，并断定此人确实是要找的人，其名字是对话控制部件 63 检索的。

于是对话控制部件 63 此刻断定此人确实是要找的人，其名字是对话控制部件 63 检索的，并前进至步骤 SP5，在此将积累学习的起始指令给予讲话者

识别部件 61。此刻，当首先从讲话者识别部件 61 给出的 SID 与可使用基于存储在内存 65 中关联信息的名字而检索的 SID 一致时，对话控制部件 63 此将积累学习的起始指令给予讲话者识别部件 61，反之，而当不一致时，则给出纠正学习的起始指令。

此后，处理前往步骤 SP6，在此对话控制部件 63 顺次将字符串数据 D2 发送至声音合成器 64，以持续地寒暄而延长与此人的对话，例如图 10 所示的“今天天气不错，是吧？”而在过了所定的期间后，处理前进至步骤 SP7，在此向讲话者识别部件 61 和面孔识别部件 62 发出积累学习或纠正学习的结束指令，而处理前进至步骤 SP20，在此结束对于此人的名字学习处理。

另一方面，在步骤 SP1 处获得否定的结果意味着由面孔识别部件 62 识别面孔的人是新人，或面孔识别部件 62 将熟人错认成新人了。而且，在步骤 SP4 处等到否定的结果意味着使用从面孔识别部件 62 给出的 FID 而检索的名字与此人的名字不一致。在任一情形下，对话控制部件 63 被认为未处于正确认出此人的状态中。

接着，当在步骤 SP1 处得到否定的结果时，或当在步骤 SP4 处得到否定的结果时，处理前进至步骤 SP8，在此对话控制部件 63 向声音合成器 64 馈送字符串 D2，从而让扬声器 54 输出问询的声音，例如图 11 所示的“请问阁下尊姓？”，以获悉此人的名字。

接着处理前进至步骤 SP9，在此对话控制部件 63 等候此人回答问题的声音识别结果（即，名字），例如“我是某某”，和在回答时刻讲话者识别部件 61 的讲话者识别结果（即，SID），其分别从声音识别部件 60 和讲话者识别部件 61 给出。

接着，当从声音识别部件 60 给出声音识别结果并从讲话者识别部件 61 给出 SID 时，处理前进至步骤 SP10，在此对话控制部件 63 基于声音识别结果和 SID，以及首先从面孔识别部件 62 给出的 FID，判断此人是不是新人。

在该实施例的情形下，以上判断是由三种识别结果的多数决定做出的：由声音识别部件 60 识别声音的结果所得的名字，来自讲话者识别部件 61 的 SID，和来自面孔识别部件 62 的 FID。

例如，当来自讲话者识别部件 61 的 SID 和来自面孔识别部件 62 的 FID 双方皆显示意味着无法识别的“－1”时，并当按以上步骤基于来自声音识别部件 60 的声音识别结果而得到的人名不与内存 65 中的任何 SID 和 FID 关联

时，判断此人为新人。此判断可根据某人长得不像任何一张面孔、语音也不像任何熟人而名字又是新名的情况而做出。

另一方面，当来自讲话者识别部件 61 的 SID 和来自面孔识别部件 62 的 FID 与内存 65 中的不同名字关联、或二者之一显示意味着无法识别的“－1”时，并当基于在步骤 SP9 处声音识别部件 60 的声音识别结果而得到的人名未存储在内存 65 中时，对话控制部件 63 判断此人是新人。这是因为，在各种识别处理的步骤中将此人判断为新人的置信度较高，因为一种新类别易于错误地被识别成任何已知类别，并考虑到听觉识别出的名字未注册的事实。

与此相反，当来自讲话者识别部件 61 的 SID 和来自面孔识别部件 62 的 FID 与内存 65 中的同一名字关联时，并当基于在步骤 SP9 处声音识别部件 60 的声音识别结果而得到的人名与 SID 和 FID 关联时，对话控制部件 63 判断此人是熟人。

而且，当来自讲话者识别部件 61 的 SID 和来自面孔识别部件 62 的 FID 与内存 65 中的不同名字关联时，并当基于在步骤 SP9 处声音识别部件 60 的声音识别结果而得到的人名与 SID 和 FID 之一关联时，对话控制部件 63 判断此人是熟人。在此情形下，判断由多数决定做出，因为讲话者识别部件 61 和面孔识别部件 62 的识别结果中可能有一个是错误的。

同时，当来自讲话者识别部件 61 的 SID 和来自面孔识别部件 62 的 FID 与内存 65 中的不同名字关联时，并当基于在步骤 SP9 处声音识别部件 60 的声音识别结果而得到的人名与内存 65 中的 SID 和 FID 皆不关联时，对话控制部件 63 不判断此人是熟人还是新人。在此情形下，可以想见：声音识别部件 60、讲话者识别部件 61、和面孔识别部件 62 之一或全部识别错了，但此时尚不能判断哪一个是错。所以在此情形下，判断被挂起。

在此判断处理之后，当在步骤 S10 处判断此人是新人时，处理前进至步骤 SP11，在此对话控制部件 63 给予讲话者识别部件 61 和面孔识别部件 62 新学习的起始指令，而接着处理前往步骤 SP12，在此对话控制部件 63 将字符串数据 D2 发送至声音合成器 64，以继续谈话，从而延续此人的聊侃，例如图 11 所示的“我是机器人，幸会。”或“某某君，今天天气不错，是吧？”。

此后处理转往步骤 SP13，在此对话控制部件 63 判断讲话者识别部件 61 中的声学特征数据和面孔识别部件 62 中的形貌特征数据二者的收集是否已达到足够量，而若得到否定的结果，则处理返回步骤 SP12，并继而重复步骤

SP12－SP13－SP12的循环，直到在步骤SP13处得到肯定的结果。

当在步骤SP13处得到肯定的结果，并且，讲话者识别部件61中的声学特征数据和面孔识别部件62中的形貌特征数据二者的收集已达到足够量时，处理前进至步骤SP14，在此对话控制部件63给予讲话者识别部件61和面孔识别部件62新学习的结束指令。结果，将声学特征数据存储在讲话者识别部件61中，且与新SID关联，并将形貌特征数据存储在面孔识别部件62中，且与新FID关联。

此后，处理前进至步骤SP15，在此对话控制部件63等候分别从讲话者识别部件61和面孔识别部件62给出SID和FID，并当给出它们时，例如图12所示，将它们在内存65中注册，与在步骤SP9处基于在步骤SP9处声音识别部件60的声音识别结果而得到的人名相关联。接着在对话控制部件63中的处理转向步骤SP20，并结束对于此人的名字学习处理。

另一方面，当在步骤SP10处判断此人是熟人时，处理前进至步骤SP16，当讲话者识别部件61和面孔识别部件62正确判断熟人（即，当讲话者识别部件61和面孔识别部件62输出同一SID或FID作为识别结果，而对应于作为关联信息存储在内存65中的熟人时）时，对话控制部件63给予讲话者识别部件61或面孔识别部件62积累学习的起始指令，而当讲话者识别部件61和面孔识别部件62不能正确判断熟人（当讲话者识别部件61和面孔识别部件62输出同一SID或FID作为识别结果，而对应于作为关联信息存储在内存65中的熟人时）时，对话控制部件63给予讲话者识别部件61或面孔识别部件62纠正学习的起始指令。

具体地说，当在步骤SP9处从讲话者识别部件61得到的SID和从面孔识别部件62给出的FID与内存65中同一名字关联时，同时当在步骤SP10处根据以下事实判断此人是熟人时：即基于声音识别部件60在步骤SP9处的识别结果而得到的名字是与SID和FID关联的名字，此时，对话控制部件63给予讲话者识别部件61和面孔识别部件62二者积累学习的起始指令。

而且，当在步骤SP9处从讲话者识别部件61得到的SID和从面孔识别部件62给出的FID与内存65中不同名字关联时，同时当在步骤SP10处根据以下事实判断此人是熟人时：即基于声音识别部件60在步骤SP9处的识别结果而得到的名字是与SID和FID之一关联的名字，此时，对话控制部件63给予讲话者识别部件61或面孔识别部件62之一积累学习的起始指令，其中识别

部件 61 或面孔识别部件 62 已经产生了与基于声音识别部件 60 的识别结果而得到的名字相关联的 SID 或 FID，并给予讲话者识别部件 61 或面孔识别部件 62 之一纠正学习的起始指令，其中识别部件 61 或面孔识别部件 62 已经产生了与基于声音识别部件 60 的识别结果而得到的名字无关联的 SID 或 FID。

此后处理前往步骤 SP17，在此对话控制部件 63 依次将一系列的字符串数据 D2 发送至声音合成器 64 以保持聊侃，从而延长与此人的对话，例如图 13 所示的“唉，阁下是某某君是吧，我想起您来了。今天天气不错，是吧？”或“我们何时见过面了？”而在过了所定的积累学习或纠正学习的足够期间后，处理前进至步骤 SP18，在此向讲话者识别部件 61 和面孔识别部件 62 发出积累学习或纠正学习的结束指令，而处理前进至步骤 SP20 以终止对于此人的名字学习处理。

同时，当对话控制部件 63 在步骤 SP10 处判断无法确定此人是熟人还是新人时，处理前进至步骤 SP19，并将一系列字符串数据 D2 依次发送至声音合成器 64，以进行例如图 14 所示的聊侃：“噢，是吗？您好吗？”

并且，在此情形下，对话控制部件 63 不给予讲话者识别部件 61 或面孔识别部件 62 新学习、积累学习、或纠正学习的起始指令或结束指令（即，讲话者识别部件 61 和面孔识别部件 62 皆不得进行新学习、积累学习、或纠正学习），而在所定期间内处理前进至步骤 SP20 以终止对于此人的名字学习处理。

如此，对话控制部件 63 被设计成基于声音识别部件 60、讲话者识别部件 61、和面孔识别部件 62 的识别结果，而能够通过控制与人的对话和控制讲话者识别部件 61 和面孔识别部件 62 的操作来依次学习新人的名字。

(4)声音识别部件 60 和面孔识别部件 62 的具体构成

其次，对于声音识别部件 60 和面孔识别部件 62 的具体构成给出解释，以体现上述名字学习功能。

(4－1)声音识别部件 60 的具体构成

图 15 表示声音识别部件 60 的具体构成。

在此声音识别部件 60 中来自麦克风 51 的音频信号 S1B进入 AD（模数）变换器 70。AD 变换器 70 对供给的模拟信号的音频信号 S1B进行取样和量化，使该模拟信号 AD 变换成数字信号的声音数据。将此声音数据馈送至特征抽取部件 71。

特征抽取部件 71 基于合适的帧对输入的声音数据进行例如 MFCC(Mel 频率对数倒频谱系数，Mel Frequency Cepstrum Coefficient)分析，并向匹配部件 72 和未注册词处理部件 76 输出为特征向量（特征参数）MFCC，作为得到的分析结果。注意到特征抽取部件 71 能抽取诸如线性预测系数、对数倒频谱系数、线谱对、各所定频率的功率（滤波池的输出）等作为特征向量。

匹配部件 72 按需要参照声学模型存储部件 73、辞典存储部件 74、和语法存储部件 75，基于诸如连续分布 HMM（隐藏 Markov 模型），且使用来自特征抽取部件 71 的特征向量而识别进入麦克风 51 的音频态声音（输入声音）。

即，声学模型存储部件 73 存储声学模型（例如，HMM，或包括用作 DP（动态编程）匹配的标准图谱等），其代表识别出的语言的单音素、音节和音素学等单词的声学特征。HMM（隐藏 Markov 模型）被用作声学模型是因为此处进行声音识别的基础是连续分布 HMM 方法。

辞典存储部件 74 识别辞典，其中，通过作为识别单位的音群而得到词音与词条的信息互相关联。

下面，图 16 表示存储在存储部件 74 中的辞典。

如图 16 所示，词条和在辞典音群中关联的音素系列在音素系列中为各对应词而构建。在图 16 的辞典中，一个条目（图 16 中的一行）对应于一个音群。

注意到图 16 中的条目以罗马字母和日本语字符（假名和汉字）二者、以及罗马字母中的音素系列来代表。然而，音素系列中的“N”却表明“N(ん)”，这是日本语中的鼻音音节。而且，图 16 中的一个音素系列被描述成一个条目，能将数个音素系列表述成一个条目。

返回图 4，语法存储部件 26 存储语法规定，其描述在辞典存储部件 25 的辞典中注册的各词是如何链接（成句）的。

图 17 表示存储在语法存储部件 75 中的语法规定。注意到图 17 中的语法规定以 EBNF(Extended Backus Naur Form)来描述。

在图 17 中，从一行开头到出现“;”的部分表达了一项语法规定。而且，以“$”开头的一群西文字母（行）表达了变量，同时不带“$”的一群西文字母（行）则表达了一个词条（图 16 中以罗马字母描述的条目）。此外，以一对〔〕括起来的部分可以省略，而标记〔｜〕意味着应该选择摆在前后的任一词头（变量）。

因此，在图 17 中，例如在头一行（紧靠顶上的第一行）的语法规定“$col =〔kono | sono〕色は;”中，变量$col 代表“konoiro wa（这个颜色）”或“sonoiro wa”（那个颜色）的一行词。

在图 17 所示的语法规定中，变量$sil 和$garbage 却未定义，变量$sil 代表哑声学模型（哑模型），而变量$garbage 本质上代表冗模型、其允许音素间的自由过渡。

再返回图 15，匹配部件 72 参照辞典存储部件 74 的辞典，通过连接存储在声学模型存储部件 73 中的声学模型，而构成词的声学模型（词模型）。进而，匹配部件 72 参照存储在语法存储部件 75 中的语法规定而连接一些词模型，并基于字符向量，凭连续分布 HMM 方法，使用这些连接词识别输入麦克风 51 的声音。即，匹配部件 72 检测词模型系列，其从特征抽取部件 71 输出的时系列特征向量表示最高观察分值（可能性），并输出为与该词模型的系列对应的词条行的声音识别结果。

具体地说，匹配部件 72 将连接词模型与对应词链接起来，并基于字符向量，凭连续分布 HMM 方法，使用这些连接词识别输入麦克风 51 的声音。即，匹配部件 72 检测词模型系列，其从特征抽取部件 71 输出的时系列特征向量表示最高观察积分（可能性），并输出为与该词模型的系列对应的词条行的声音识别结果。

具体地说，匹配部件 72 就对应于连接词模型的一行词而言，累集各特征向量的出现概率（输出概率），以累集值作为积分，输出为积分最高的词条行的声音识别结果。

以上输出并输入麦克风 51 的声音识别结果作为字符串数据 D1 被输出至对话控制部件 63。

在图 17 的实施例中，有一条语法规定「$pat1 = $color1$garbage$color2;」（以下酌情称之为“非注册词规定”），其使用变量$garbage 表明第 9 行（自顶上起第 9 行）上的冗模型，而当适用此非注册词规定时，匹配部件 72 检测对应于变量$garbage 的声音部件作为非注册词的声音部件。此外，当使用非注册词规定时，匹配部件 72 检测非注册词的音素系，即作为变量$garbage 所表明的冗模型中的过渡音素的音素系列。接着，当作为适用非注册词规定而得到了声音识别结果时，匹配部件 72 向非注册词处理部件 76 供给检测到的非注册词的声音部和音素系列。

注意到根据以上非注册词规定“$pat1 = $color1 $garbage $color2;”，在由变量$color1表明的注册在辞典中的词(行)的音素系列与由变量$color2表明的注册在辞典中的词（行）的音素系列之间检测到一个非注册词，然而，在此实施例中甚至也能将此非注册词规定使用于以下情形：即讲话中含数个非注册词，以及在辞典中注册的词（行）之间未放入非注册词的情形。

非注册词处理部件 76 暂时持有从特征抽取部件 71 供给的特征向量的系列（特征向量系列）。进而，当从匹配部件 72 收到非注册词的声音部和音素系列时，由于声音部出自暂时持有的特征向量系列，非注册词部处理部件 76 检测声音的特征向量系列。接着非注册词处理部件 76 将唯一的 ID（身份）分配给来自匹配部件 72 的音素系列（非注册词），其与非注册词的音素系列和声音部中的特征向量系列一道被提供给特征向量缓冲 77。

特征向量缓冲 77 暂时存储从非注册词处理部件 76 供给的非注册词 ID、音素系列、和特征向量系列，其如图 18 所示互相关联。

在图 18 中，以 1 开始的序号作为识别符附于非注册词。因此，例如，在 N 个非注册词 ID，音素系列和特征向量系列存储在特征向量缓冲 77 中的情形下，而当匹配部件 72 检测到非注册词的声音部和音素系列时，在非注册词处理部件 76 中将数值 N + 1 附于非注册词作为 ID，而非注册词的 ID、音素系列和特征向量系列存储在特征向量缓冲 77 中，如图 18 中的虚线所示。

返回图 15，音群部件 78 算出各其他非注册词（以下酌情称之为“新非注册词”）与早已存储在特征向量缓冲 77 中的非注册词（以下酌情称之为“早已存储的非注册词”）相关联的积分。

即，像在匹配部件 72 的情形下那样，将新非注册词作为输入声音、并将早已存储的非注册词作为在辞典中注册的词，音群部件 78 算出新非注册词与各早已存储的非注册词相对积分。具体地说，音群部件 78 通过参照特征向量缓冲 77 识别新非注册词的特征向量系列，并根据早已存储的非注册词的音素系列连接声学模型，凭所连接的声学模型算出积分，作为新非注册词的观察特征向量系列的可能性。

注意到存储在声学模型存储部件 73 中的声学模型用于此目的。

类似地，音群部件 78 算出新非注册词与各早已存储的非注册词相对的积分，并凭此积分更新存储在积分单存储部件 79 中的积分单。

进而，通过参照更新积分单，音群部件 78 从对早已得到的非注册词（早

已存储的非注册词）进行音群化的音群中检测出附加新非注册词为新成员的音群。再进而，音群部件 78 基于同样的音群的成员将音群分成检测到新非注册词的音群的新成员，并基于分割结果，更新存储在积分单存储部件 79 中的积分单。

积分单存储部件 79 存储新非注册词与早已存储的非注册词相对积分，以及积分单，其相对于新非注册词而注册了早已存储的非注册词的积分及其他。

此处，图 19 表示积分单。

积分单由描述非注册词的“ID”、“音素系列”、“音群数”、“代表成员 ID”和“积分”的条目组成。

同样存储在特征向量缓冲 77 中的内容由音群部件 78 注册为非注册词的“ID”、“音素系列”。“音群数”是指定该条目的非注册词是成员的音群的数，而该数由音群部件 78 指定并在积分单中注册。“代表成员 ID”是作为代表成员的非注册词的 ID，其代表该条目的非注册词是成员的音群，此代表成员 ID 使识别非注册词是成员的音群的代表成员成为可能。音群的代表成员由音群部件 29 得到，而代表成员 ID 注册至积分单上的代表成员 ID。“积分”是各其他非注册词与此条目的非注册词相对的积分，如上述由音群部件 78 算出。

下面，假设例如 N 个非注册词的 ID、音素系列、和特征向量系列存储在特征向量缓冲 77 中，注册至积分单上的是 N 个非注册词的 ID、音素系列、音群数、代表 ID、和积分。

而且，当新非注册词的 ID、音素系列、和特征向量系列存储在特征向量缓冲 77 中时，积分单由图 19 中虚线所示在音群部件 78 中更新。

即，新非注册词的 ID、音素系列、音群数、代表 ID、和各早已存储的与新非注册词相对的非注册词积分（图 19 中的积分 s(N + 1, 1)，s(2, N + 1)，… s(N + 1, N)）被加到积分单。进而，新非注册词与各早已存储的非注册词相对的积分（图 19 中的积分 s(N + 1, 1)，s(2, N + 1)，…s(N + 1, N)）被附加到积分单。再进而，如后述，按需要对积分单上非注册词的音群数和代表成员 ID 进行置换。

在图 19 的实施例中，相对于 ID 为 j 的非注册词的 ID 为 i 的非注册词的积分（讲话）被表达成 s(i, j)。

还将相对于 ID 为 j 的非注册词（的音素系列）的 ID 为 i 的非注册词的积分 s(i, j)（讲话）也注册至积分单（图 19）。因为积分 s(i, j)是当检测

到非注册词的音素系列时，在匹配部件 72 中算出的，故不必在音群部件 78 中进行计算。

再度返回图 15，维护部件 80 基于在积分单存储部件 79 中更新的积分单而更新存储在辞典存储部件 74 中的辞典。

下面，按以下方式确定音群的代表成员。即，例如，从作为音群成员的非注册词中成为音群的代表成员，该代表成员是使得其余非注册词的积分总数最大的非注册词（其他置换做法可包括例如，由其余非注册词数去除总数所产生的平均值）。因此，在此情形下，假设属于音群的成员的成员 ID 以 k 表达，则代表成员是具有 ID 值为 k(∈k)的成员，表达为以下表达式:

$$k = \max_k\{\sum s(k^3,\ k)\} \qquad \ldots\ldots(1)$$

注意到在以上表达式(1)中，$\max_k\{\}$意味着 k 使得{}内的值最大。而且，k^3像 k 那样意味着属于音群的成员的 ID。进而，∑意味着在全部属于音群的成员的 ID 上变化 k^3而产生的总数。

在如上确定代表成员的情形下，当音群成员是一两个非注册词时，不必要在确定代表成员中算出积分。即，当音群成员是单个非注册词时，该单个非注册词就是代表成员，而当音群成员是两个非注册词时，可以将两词中的任一个指定为代表成员。

绝不可能将确定代表成员的方法限制在上述一种，但是可能指定诸如一个非注册词为音群的代表成员，该非注册词是从使得特征向量空间中与各其余非注册词相对的距离总数最小的音群成员中拾取的非注册词。

在如上构造的声音识别部件 60 中，根据图 20 所示的声音识别处理例程 RT2 来进行声音识别处理，以识别输入麦克风 51 的声音，并进行对于非注册词的非注册词处理。

在实际中，当由人讲话而得到的音频信号 S1B从麦克风 51 通过 AD 变换器 70 作为声音数据而提供给特征抽取部件 71 时，在声音识别部件 60 中，此声音识别处理例程 RT2 在步骤 SP30 处开始行动。

在下一步骤 SP31 中特征抽取部件 71 通过以所定的帧单位在声学上分析声音数据来抽取特征向量，而此特征向量的系列被提供给匹配部件 72 和非注册词处理部件 76。

在后续步骤 S32 处，匹配部件 76 对于从特抽取部件 71 给出的特征向量系列进行上述积分计算，而在下一步骤 S33 处得到并输出词行的条目、其是

基于积分计算得到的积分的声音识别结果。

进而，匹配部件 72 在下一步骤 S34 处判断在用户声音中含不含非注册词。

当在步骤 S34 处判断在用户声音中不含非注册词时，即，在没有应用上述非注册词规定“$pat1 = $color1 $garbage $color2;”而得到声音识别结果的情形下，处理前进至步骤 S35，结果就终止了。

与以上相反，在步骤 S34 处，当判断在用户声音中含非注册词时，即，在应用上述非注册词规定“$pat1 = $color1 $garbage $color2;”而得到声音识别结果的情形下，在后续步骤 S35 处匹配部件 23 检测在非注册词规定中的变量$garbage 对应的声音部来作为非注册词的声音部，并在此刻检测非注册词的音素系列，即作为在变量$garbage 代表的冗模型中的音素过渡的音素系列，而非注册词的声音部和音素系列被提供给非注册词处理部件 76，终止处理（步骤 SP36）。

同时，暂时存储从特征抽取部件 71 供给的特征向量系列，当从匹配部件 72 供给的非注册词的声音部和音素系列时，非注册词部处理部件 76 在声音部中检测声音的特征向量系列。此外，非注册词部处理部件 76 将 ID 附加来自匹配部件 72 的非注册词（的音素系列），其与非注册词的音素系列和声音部中的特征向量系列一道被提供给特征向量缓冲 77。

如以上方式，当新发现的非注册词（新非注册词）的 ID、音素系列、和特征向量系列被存储在特征向量缓冲器 77 中时，根据图 21 所示非注册词处理例程 RT3 而开始非注册词处理的行动。

即，在声音识别部件 60 中，如上述，当新发现的非注册词（新非注册词）的 ID、音素系列、和特征向量系列被存储在特征向量缓冲 77 中时，此非注册词处理例程在步骤 SP40 处开始行动，紧接着是步骤 SP41，在此音群部件 78 读出来自特征向量缓冲 77 的新非注册词的 ID 和音素系列。

在下一步骤 S42 处，音群部件 78 通过参照积分单存储部件 30 中的积分单而判断是否存在早已得到（生成）的音群。

而且，当在步骤 S42 处判断不存在早已得到的音群时，即在新非注册词是头一个非注册词的情形下，而且积分单上不存在早已存储的非注册词的条目，则步骤前往步骤 S43，在此音群部件 78 新生成一个以该新非注册词为代表成员的音群，并通过将关于新音群的信息和关于新非注册词的信息注册至积分单存储部件 79 中的积分单而更新积分单。

即，音群部件 78 将来自特征向量缓冲 77 的新非注册词的 ID 和音素系列注册至积分单（图 19）。此外，音群部件 78 生成唯一的音群数，其作为新非注册词的音群数而注册至积分单。并且，音群部件 78 使新非注册词的 ID 注册至积分单中，而成为新非注册词的代表成员 ID。在此情形下，因此，该新非注册词变成新音群的代表成员。

注意到在此时不进行积分计算，因为没有已存储的非注册词，无法籍以进行与新非注册词相对的积分计算。

在步骤 S43 的处理之后，处理前往步骤 S52，在此维护部件 80 基于在步骤 S43 处更新的积分单而更新辞典存储部件 74 中的辞典，并终止处理（步骤 SP54）。

即，在此情形下，由于生成了新音群，故维护部件 31 参照积分单中的音群数而识别新生成的音群。接着维护部件 80 将对应于音群的条目附加至辞典存储部件 74 中的辞典，并注册为新音群的代表成员的音素系列条目的音素系列，在此情形下即新非注册词的音素系列。

另一方面，当在步骤 S42 处判断存储已得到的音群时，即在新非注册词不是头一个非注册词的情形下，在积分单（图 19）中存在已存储的非注册词的条目（行），处理前进至步骤 S44，在此音群部件 78 算出已存储的非注册词与新非注册词相对的积分，并同时算出新非注册词与已存储的非注册词相对的积分。

换句话说，假设例如已存储的非注册词的 ID 为从 1 到 N，而新非注册词的 ID 为 N＋1，则在音群部件 78 中算出 N 个已存储的非注册词与新非注册词相对的积分 s(N＋1，1)，s(N＋1，2)，…s(N，N＋1)，其在图 19 中虚线所示的部分中，并算出新非注册词与 N 个已存储的非注册词相对的积分 s(1，N＋1)，s(2，N＋1)，…s(N，N＋1)。注意到在音群部件 78 中算出那些积分时必需新非注册词与 N 个已存储的非注册词的特征向量系列，不过，那些特征向量是参照特征向量缓冲 28 而识别的。

接着音群部件 78 将算出的积分与新非注册词的 ID 和音素系列附加至积分单，而处理前进至步骤 S45。

在步骤 S45，通过参照积分单（图 19），音群部件 78 检测使得与新非注册词相对的积分 s(N＋1，i)（i＝1，2，…，N）最高（最大）的代表成员的音群。换言之，音群部件 78 通过参照积分单上代表成员的 ID 而识别成为代

表成员的已存储的非注册词，并进而通过参照积分单上的积分而检测已存储的非注册词，该词作为使得非注册词的积分最高的代表成员。而音群部件 78 检测具有已存储的非注册词的音群数的音群、其作为检测到的代表成员。

此后处理前往步骤 S46，在此音群部件 29 将新非注册词附加至在步骤 S45 处检测到的音群的成员（以下酌情称之为“检测到的音群”）。即，音群部件 78 在积分单上将检测到的音群的代表成员的音群数写成新非注册词的音群数。

例如，在步骤 S47 处，音群部件 78 进行例如音群分割处理，以将检测到的音群一分为二，而处理前进至步骤 S48。在步骤 S48 处，音群部件 78 判断检测到的音群是否已由步骤 S47 处的音群分割处理成功地一分为二了，而当判断分割成功时，处理前往步骤 S49。在步骤 S49 处，音群部件 78 算出通过分割检测到的音群而产生的两个音群之间的音群距离（以下酌情将这两个音群称之为“第一子音群和第二子音群”）。

此处，第一子音群与第二子音群之间的音群距离例如定义如下。

假设第一子音群和第二子音群二者之中任一成员（非注册词）的 ID 由 k 代表，而第一子音群和第二子音群的任一代表成员（非注册词）由 k1 或 k2 代表，则在下式中：

$$D(k1, k2) = \max val_k\{abs(\log(s(k, k1)) - \log(s(k, k2)))\} \quad \cdots\cdots(2)$$

值 D(k1，k2)被定义成第一子音群与第二子音群之间的音群间距。

注意到在表达式(2)中 abs()表明()中值的绝对值。而且，$\max val_k\{\}$指示{}中值的通过变化 k 而得到的最大值。而 表达自然对数或常用对数。

下面，假设 ID 代表成员 i 作为成员＃1，表达式(2)中的积分的倒数 1／s(k，k1)对应于成员＃k 与代表成员 k1 之间的距离，而表达式(2)中的积分的倒数 1／s(k，k2)对应于成员＃k 与代表成员 k2 之间的距离。根据表达式(2)，因此，代表成员＃k1 与第一子音群的任何成员的间距，代表成员＃k2 与第二子音群的任何成员的间距，这两个间距之差的最大值即是第一与第二子音群之间的音群间距。

音群间距不限于上述，还可能指定以下作为音群间距，例如，由第一子音群的代表成员与第二子音群的代表成员的 DP 匹配而得到的特征向量空间中的距离累加。

在步骤 S49 的处理之后，处理前进至步骤 S50，在此音群部件 78 判断第

一与第二子音群之间的音群间距是否大于所定的阈值ξ（或所定的阈值ξ或更高）。

当在步骤S50处判断音群间距大于所定的阈值ξ时，即在作为检测到的音群的成员的多个非注册词将按声学特征被音群化为两个音群时，处理前往步骤S51，在此音群部件78将第一和第二子音群注册至积分单存储部件79中的积分单。

即，随着将唯一音群数分配给第一和第二子音群，音群部件78更新积分单，从而音群化至第一子音群的成员的音群数被指定为第一子音群的音群数，而音群化至第二子音群的成员的音群数被指定为第二子音群的音群数。

进而，音群部件78更新积分单，从而音群化至第一子音群的成员的成员ID被指定为第一子音群的代表成员ID，而音群化至第二子音群的成员的成员ID被指定为第二子音群的代表成员ID。

注意到有可能将检测到的音群的音群数分配给第一和第二子音群之一。

当由音群部件78按以上方式将第一和第二子音群注册至积分单时，处理从步骤S51转至步骤S52，在此维护部件80基于积分单而更新辞典存储部件74中的辞典，接着处理终止（步骤SP54）。

即，在此情形下，因为检测到的音群被分成第一和第二子音群，故维护部件80首先删除与检测到的音群对应的辞典中的条目。进而，维护部件80向辞典中附加与第一和第二子音群分别对应的两个条目，并将第一子音群的代表成员的音素系列注册为与第一子音群对应的条目的音素系列，同时将第二子音群的代表成员的音素系列注册为与第二子音群对应的条目的音素系列。

另一方面，当在步骤S48处判断步骤S47处的音群分割处理不能将检测到的音群一分为二，或当在步骤S50处判断第一与第二子音群的音群间距小于所定的阈值ξ时（换言之，在这种情形下：即作为检测到的音群的数个非注册词的声学特征不像第一和第二子音群，以致达到造成了音群化的地步），处理前进至步骤S53，在此音群部件78得到检测到的音群的新代表成员，并以此来更新积分单。

即，音群部件78对于附加了新非注册词的检测到的音群的各成员，通过参照积分单存储部件79中的积分单，而识别表达式(1)的计算的必要的积分$s(k^3, k)$。进而，音群部件78使用识别出的积分$s(k^3, k)$、基于表达式(1)而

得到成为检测到的音群的新代表成员的成员 ID。接着音群部件 78 将积分单（图 19）中检测到的音群的各成员的代表成员 ID 改写成检测到的音群的新代表成员的 ID。

此后处理前往步骤 S52，在此维护部件 80 基于积分单而更新辞典存储部件 74 中的辞典，接着处理终止（步骤 SP54）。

换句话说，在此情形下，维护部件 80 通过参照积分单而识别检测到的音群的新代表成员，并进而识别代表成员的音素系列。接着维护部件 80 将与辞典中检测到的音群对应的条目的音素系列置换成检测到的音群的新代表成员的音素系列。

下面，根据图 22 所示的音群分割处理例程 RT4 而进行图 21 中步骤 SP47 处的音群分割处理。

即，在声音识别处理部件 60 中，随着处理从图 22 的步骤 SP46 推进至步骤 S47，音群分割处理例程 RT4 开始于步骤 SP60 处，而首先在步骤 S61 处音群部件 78 选择两个任意成员的组合，这两个成员从附加了新非注册词作为成员的检测到的音群中选出，皆是试验性代表成员。注意到以下酌情将此两个试验性代表成员称做“第一试验性代表成员”和“第二试验性代表成员”。

接着，在下一步骤 S62 处，音群部件 78 判断检测到的音群的成员是否可以一分为二，从而分别将第一试验性代表成员和第二试验性代表成员作为代表成员。

在此阶段有必要算出表达式(1)以确定第一或第二试验性代表成员是否可以作为代表成员，而用于此计算的积分 $s(k', k)$ 可通过参照积分单来识别。

当在步骤 S62 处判断检测到的音群的成员不可能一分为二、从而分别将第一试验性代表成员和第二试验性代表成员作为代表成员时，处理跳过步骤 S62 而前往步骤 S64。

介时，当在步骤 S62 处判断检测到的音群的成员可以一分为二，从而分别将第一试验性代表成员和第二试验性代表成员作为代表成员时，处理前往步骤 S63，接着音群部件 78 将检测到的音群的成员一分为二，从而分别将第一试验性代表成员和第二试验性代表成员作为代表成员，作为检测到的音群的分割结果，以分割出的一对双音群作为第一和第二子音群的候选（以下酌情称之为“一对候选音群”），而处理转至步骤 S64。

在步骤 S64 处，音群部件 78 判断在检测到的音群的成员中是否还有一对

成员未被选为第一和第二试验性代表成员对，而当判断是时，处理返回步骤S61，在此未被选为第一和第二试验性代表成员对的检测到的音群的一对成员被选择，随后重复相同处理。

而且，当在步骤S64处判断没有哪一对成员未被选为第一和第二试验性代表成员对时，处理前进至步骤S65，在此音群部件78判断是否有一对候选音群。

当在步骤S65处判断没有一对候选音群时，处理跳过步骤S66而返回。在此情形下，在图21的步骤S48处判断检测到的音群不可分割。

另一方面，当在步骤S65处判断存在一对候选音群时，处理前往步骤S66，在此音群部件78当有数对候选音群时，得到各对候选音群的两个音群的音群间距。接着音群部件78得到音群间距最小的一对候选音群，并分割此对候选音群以产生第一和第二子音群，而处理返回。注意到在仅有一对候选音群的情形下，他们被原样地作为第一和第二子音群。

在此情形下，在图21的步骤S48处判断检测到的音群已成功地进行了分割。

如上述，因为在音群部件78中附加了作为新成员的音群（检测到的音群）新非注册词被从进行了已得到的非注册词的音群化的音群中检测到，而随着新非注册词作为检测到的音群的新成员，检测到的音群基于检测到的音群的成员而被分割，容易将非注册词音群化至声学特征互相近似的那些（音群）。

此外，因为辞典是基于在维护部件80中的这种音群的结果而更新的，故容易将非注册词注册至防其变大的辞典。

而且，例如，若非注册词的声音部在匹配部件72中检测错了，则这一非注册词被音群化至另一音群，其与声音部通过分割检测到的音群而正确检测的非注册词分离。接着对应于此音群的条目被注册至辞典，不过，由于对应于此声音部的条目的音素系列检测不正确，故不会发生未来声音识别给出大积分的情况。因此，例如，一旦非注册词的声音部检测错了，则此错误几乎对未来声音识别没有影响。

现在，图23表示通过朗读非注册词而得到的音群化结果。注意到图23中的各条目（各行）表示一个音群。而且，图23的左列表示各音群的代表成员（非注册词）的音素系列，而图23的右列表示成为各音群的成员的非注册词的内容和数字。

即，例如，在图 23 中第一行的条目指示一个音群，其成员是仅朗读非注册词“furo（沐浴）”，而该代表成员的音素系列是“doroa:”。而且，例如第二行的条目指示一个音群，其成员是三次朗读非注册词“furo”，而该代表成员的音素系列是“kuro”。

进而，例如第七行的条目指示一个音群，其成员是四次朗读非注册词“hon（书）”，而该代表成员的音素系列是“NhoNde: su（ンホンデース）”。而且，例如第八行的条目指示一个音群，其成员是一次朗读非注册词“orengi（橘子）”和十九(19)次朗读非注册词“hon（书）”，而该代表成员的音素系列是“ohoN（オホン）”。其他条目的指示类似。

根据图 23，可见对于同样的非注册词的朗读进行了正确音群化。

对于图 23 中第 8 行的条目，将一次朗读非注册词“orengi（橘子）”和十九(19)次朗读非注册词“hon（书）”音群化至同一音群。根据朗读是音群成员，可认为此音群应是非注册词“hon（书）”的音群，不过，非注册词“orengi”的朗读也是该音群的成员。随着不断输入非注册词“hon（书）”的朗读，此音群也由音群化而分割，导致音群化可按如下方式进行：即产生音群，其成员仅是朗读非注册词“hon（书）”，以及音群，其成员仅是朗读非注册词“orengi”。

(4－2)面孔识别部件 62 的具体构成

下面，对于面孔识别部件 62 的具体构成给出解释。

如图 24 和 25 所示，面孔识别部件 62 能够在动态变化的环境下，在所定的周期内响应，该面孔识别部件 62 包括面孔抽取处理部件 90，以基于从 CCD 照相机（图 5）提供的视频信号 S1A而从图像中抽取面孔图谱，以及面孔识别处理部件 91 基于抽取的面孔图谱而识别面孔。在此实施例中使用“Gabor 滤波”进行面孔抽取处理以抽取面孔图谱，并使用“支持向量机：SVM”进行面孔识别处理以从面孔图谱中识别面孔。

面孔识别部件 62 被供以学习阶段，在此面孔识别处理部件 91 学习面孔图谱，以及识别阶段以基于学习数据识别从视频信号 S1A中抽取的面孔图谱。

图 24 表示面孔识别部件 62 的学习阶段的构成，而图 25 表示面孔识别部件 62 的识别阶段的构成。

如图 24 所示，在学习阶段向由支持向量机组成的面孔识别处理部件 91 输入面孔抽取的结果，该面孔在由 Gabor 滤波器组成的面孔抽取处理部件 90

中从 CCD 照相机（图 5）输入的已捕获的用户图像中抽取。在面孔识别处理部件 91 中通过使用从外部供给的学习用数据，即，教师数据，而得到暂时鉴别功能。

而且，如图 25 所示，在鉴别阶段向面孔识别处理部件 91 输入面孔抽取结果，其是在面孔抽取处理部件 90 中基于从 CCD 照相机 50 供给的视频信号 S1A在图像内从人面孔中抽取的。在面孔识别处理部件 91 中通过以各种数据库上的图像来测试暂时得到的鉴别功能而检测面孔。接着，成功检测的内容输出作为面孔数据。同时，未成功检测的内容被附加至学习数据，作为非面孔数据，并进行进一步学习。

以下对于面孔抽取处理部件 90 中的 Gabor 滤波处理和面孔识别处理部件 91 中的支持向量机给出详细解释。

(4－2－1) Gabor 滤波处理

早已知道在人类的视觉细胞中存在具有对某些特定方向的选择性的细胞。这些选择性细胞含有响应垂直线的细胞和响应水平线的细胞。在这种情形下，Gabor 滤波是由数个具有方向选择性的空间滤波器组成的。

Gabor 滤波在空间上以 Gabor 函数表达。Gabor 函数 $g(x, y)$ 如以下表达式所示，由载波 $s(x, y)$ 组成、其含余弦分量和二维高斯解析包络 $W_r(x, y)$。

$$g(x, y) = s(x, y)W_r(x, y) \quad \ldots\ldots(3)$$

载波 $s(x, y)$ 使用数个函数表达为以下的表达式(4)。此处，坐标值 (u_0, v_0) 指示空间频率，P 指示余弦分量的相位。

此处，以下表达式表示载波，

$$s(x, y) = \exp(j(2\pi(u_0x + v_0y) + P)) \quad \ldots\ldots(4)$$

也可表示成以下表达式，

$$Re(s(x, y)) = \cos(2\pi(u_0x + v_0y) + P)$$

$$Im(s(x, y)) = \sin(2\pi(u_0x + v_0y) + P) \quad \ldots\ldots(5)$$

即分割成实数部 $Re(s(x, y))$ 和虚数部 $Im(s(x, y))$。

另一方面，使用以下表达式，由二维高斯分布组成的包络可表达如下：

$$W_r(x, y) = K\exp(-\pi(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2)) \quad \ldots\ldots(6)$$

此处，坐标轴 (x_0, y_0) 是函数的峰值，常数 a 和 b 是高斯分布的比例参数。而且，如以下表达式所示，下标 r 表明旋转动作。

$$(x-x_0)_r = (x-x_0)\cos\theta + (y-y_0)\sin\theta$$
$$(y-y_0)_r = -(x-x_0)\sin\theta + (y-y_0)\cos\theta \quad \ldots\ldots(7)$$

因此，根据上述表达式(4)和(6)，Gabor 滤波可表达成以下表达式所示的空间函数:

$$g(x, y) = K\exp(-\pi(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2))$$
$$\exp(j(2\pi(u_0x + v_0y) + P)) \quad \ldots\ldots(8)$$

此实施例中的面孔抽取处理部件 90 使用共计二十四(24)个 Gabor 滤波器来进行面孔抽取，这些 Gabor 滤波器使用八(8)个方向和三(3)种频率。

Gabor 滤波器的响应以下面的表达式来表达，在此 G_i 是第 i 个 Gabor 滤波器，第 i 个 Gabor 结果(Gabor Jet) J_i 和输入图像 I:

$$J_i(x, y) = G_i(x, y) \oplus I(x, y) \quad \ldots\ldots(9)$$

实际上，使用高速傅立叶变换可加快表达式(9)的动作。

制造的 Gabor 滤波器的性能可通过重构由滤波所得的像素来检验。以下表达式:

$$H(x, y) = \sum_{i=1}^{0} a_i J_i(x, y) \quad \ldots\ldots(10)$$

表示重构的图像H。

并且，输入图像 I 与重构的图像H之间产生的误差E由以下表达式来表达:

$$E = \frac{1}{2}\| I(x, y) - H(x, y) \|^2 = \frac{1}{2}\sum_{x,y} (I(x, y) - H(x, y))^2 \quad \ldots\ldots(11)$$

重构可通过得到使误差E最小的合适 a 来实现。

(4-2-2)支持向量机

在本实施例中，对于面孔识别处理部件 91 中的面孔识别，面孔识别是使用支持向量机(SVM)来进行的，该 SVM 使通用学习性能在图谱识别领域达到最高。

对于 SVM 自身，参照例如 B·Sholkoph 等人的报告(B·Sholkoph, C·Burges, A·Smola, "Advance in Kernel Support Vector Learning", The MIT Press, 1999)。根据本发明申请人做出的初步实验，可明确使用 SVM 的面孔识别方法带来更好的结果，比使用主要分量分析(PCA)和神经网络要好。

SVM 是使用线性鉴别电路(感知器)的学习机器，SVM 可通过使用核心函数扩张到非线性空间。而且，鉴别函数的学习是以采用类间最大分离空隙的方式而进行的，从而有可能通过解二维数学方程而得到解，这就在理论上得

出了全局解。

通常，图谱识别的问题是为了得到鉴别函数 f(x)，其由下面相对测试样本 x = (x1, x2, …, xn) 的表达式给出：

$$f(x) = \sum_{j=1}^{n} w_j x_j + b \qquad \ldots\ldots(12)$$

此处，SVM 学习用的教师标签由以下表达式建立：

$$y = (y1, y2, \ldots, yn) \qquad \ldots\ldots(13)$$

接着，以 SVM 来识别面孔图谱这一问题可视为：在限定条件下使权因子 w 的平方最小化，如以下表达式所示：

$$y_1(w^r x_i + b)\, z1 \qquad \ldots\ldots(14)$$

这一有限定问题可使用拉格朗日无定常数法来解。即，首先将拉格朗日（函数）引入以下表达式：

$$L(w, b, a) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^{1} a_i (y_i ((x_i^r w + b) - 1)) \qquad \ldots\ldots(15)$$

此后，如以下表达式所示：

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial w} = 0 \qquad \ldots\ldots(16)$$

应该对 b 和 w 各进行偏微分。

结果，在 SVM 中的面孔鉴别可视为二次平面问题，由以下表达式来表示：

$$\max \sum ai - \frac{1}{2} \sum aiaiyiyi^r xj$$

限定条件：$ai \geqslant 0$，$\sum aiyi = 0$ ……(17)

当特征空间的维数小于训练样本数时，引入划痕变量ξ 0 而使限定条件置换成以下表达式：

$$y_i (w^r x_i + b) \geqslant 1 - \xi_I \qquad \ldots\ldots(18)$$

至于优化，在以下表达式中：

$$\frac{1}{2} \| w \|^2 + C \sum \xi_I \qquad \ldots\ldots(19)$$

最小化目标函数。

在表达式(19)中，C 是系数,籍以指定限定条件应该放宽到何地步，而此值必须由实验确定。

关于拉格朗日常数的问题被置换成以下表达式：

$$\max\sum ai - \frac{1}{2}\sum aiaiyiyi'xj$$

限定条件: $0 \leqslant ai \leqslant C$, $\sum aiyi = 0$ ……(20)

然而，对于表达式(20)，不可能解决非线性问题。在这种情况的实施例中，随着核心函数 $K(x,x^3)$ 的引入，一旦在高维空间（kernel trick）匹配，将会线性地分开。所以，同等的在初始空间非线性分开。

核心函数可使用某种映射Φ。

$$K(x, y) = \Phi(x^r)\Phi(x^1)$$ ……(21)

而且，表达式(12)所示的鉴别函数可由以下表达式来表达:

$$f(\Phi(x)) = w^r\Phi(x) + b$$
$$= \sum aiyiK(x, xi) + b$$ ……(22)

而且，学习也可视为二维平面问题，如以下表达式所示:

$$\max\sum ai - \frac{1}{2}\sum aiaiyiyixjK(xi, xj)$$

限定条件: $0 \leqslant ai \leqslant C$, $\sum aiyi = 0$ ……(23)

作为核心，可使用高斯核心（RBF（径向基础函数））等，如以下表达式所示。

$$K(x, x^1) = \exp\left[-\frac{|x - x^1|}{\sigma^2}\right]$$ ……(24)

对于 Gabor 滤波，可根据识别任务而变更滤波器种类。

在低频滤波中将向量赋予全部滤波后的图像是冗余的。因而可通过下降取样而降低向量的维数。二十四(24)种下降取样向量成为一条线的长向量。

而且，在此实施例中，由于供给面孔图谱识别的 SVM 是鉴别器，其将特征空间一分为二，以这种方式进行学习，即: 判断受检面孔是“A人”或“非A人”。因此，首先从数据库中的图像中收集A人的面孔图像，接着在 Gabor 滤波后将“非A人”标签附加到向量。一般地，所收集的面孔图像数量最好大于特征空间的维数。同样，当需要识别十(10)人的面孔时，以“B人”、“非B人”方式为每各人设一个鉴别器。

这种学习有助于发现例如分离“A人”和“非A人”的支持向量。作为将特征空间一分为二的鉴别器，SVM 在输入新面孔图谱时，取决于 Gabor 滤波的向量所在的构成所得支持向量的界面的一侧而产生识别结果的输出。因此，当相对于边界处于“A人”区时，被识别为“A人”。同样，当处于“非A人”

区时，被识别为“非A人”。

从基于来自 CCD 照相机 50 的视频信号 S1A的图像中剪出的面部区不固定。因而有这种可能，即：面孔被投射至远离特征空间中想识别的种类的一点。从而，有可能通过暗示具有目、鼻、和口的特征的部分并由仿射变换来拟态、而增强识别率。

而且能使用自举以增强识别能力。可使用另一用来学习的图像而独立拍摄的图像来自举。这意味着当进行学习的鉴别器产生错误识别结果的输出时，通过将输入图像输入学习指令集而再度进行学习。

另一种增强识别性能的方法是观察识别结果的时间变化。最简单的方法可以是例如当十次识别中有八次皆识别为“A人”时即识别“A人”。还提出了其他预测方法，例如使用 Kalman 滤波器的那种。

(5)本实施例的操作和效果

根据以上构成，此机器人 1 通过与新人对话而得到新人的名字，并基于来自麦克风 51 和 CCD 照相机 50 的输出，存储与语音的各声学特征和检测到的人的形貌特征相关联而存储名字，并同时基于由识别另一也将获得其名的新人的出场而存储的各种数据和学习人名，并通过以上述同样方式得到并存储名字、语音的声学特征和新人的形貌特征。

因此，此机器人 1 可自然地通过与凡人的对话来学习新人、新对象的名字，就像人类常做的那样，而不必随输入声音指令或按下触觉传感器等用户的清晰指示而注册名字了。

根据以上构成，有可能通过与新人对话而得到新人的名字，并基于来自麦克风 51 和 CCD 照相机 50 的输出，与语音的各声学特征和检测到的人的形貌特征相关联而存储名字，并同时基于由识别另一未获得其名的新人的出场而存储的各种数据、学习人名，并通过以上述同样方式得到并存储名字、语音的声学特征和新人的形貌特征，结果成功地学习人名，这就可能使机器人实现自然地通过与凡人的对话来学习新人、新对象等的名字，从而大大增强他们的娱乐特性。

(6)其他实施方式

在以上实施例中，对于将本发明应用于如图 1 构成的二足直立行走的机器人 1 的情形给出了解释，但本发明不限于此，并可广泛应用于各种其他机器人设备和非机器人器材。

而且，在以上实施例中，还对于以下情形给出了解释，即由具有与人类对话功能的对话装置与此人进行声音对话，从而得到人名，以及通过对话向人类学习而得到对象的名字，该对话装置包括声音识别部件 60，对话控制部件 63，声音合成器 64，但本发明不限于此，而对话装置的构成可使人名凭例如键盘输入通过字符对话而得到。

进而，在以上实施例中，对于需要名字学习的对象是人类的情形给出了解释，但本发明不限于此，也可考虑各种其它物体成为需要名字学习的对象，而不仅是人类。

在执行以上实施例的情形下，对于以下情形给出了解释，即：由语音的声学特征和待学习的人的形貌特征来识别人，并基于其结果而判断此人是不是新人，但本发明不限于此，而是还有，例如，此人可由数种其他特征、包括体型和气味来识别，这就有可能识别一个生物固体，并基于其结果而判断此人是不是新人。而且，在名字学习用的待学习的对象是固体而非人类的情形下，可能基于从颜色、形状、图谱、和尺寸等与他物区分的各种特性识别此体而得到的结果，判断此对象是不是新的。并且，在此情形下，可设有数种识别装置，其检测各对象不同的和特定的特征，并基于检测结果和对应于事先存储的已知对象的特征数据，识别待学习的对象。

进而，在以上实施例中，对于内存构成存储装置的情形给出了解释，该存储装置，用来存储关联信息，其中已知对象的名字与由各识别装置（讲话者识别部件 61 和面孔识别部件 62）获得的关于对象的识别结果互相关联，但本发明不限于此，而可广泛利用各种除内存外的存储装置来存储信息，例如可存储信息的盘状记录媒体。

进而，在以上实施例中，对于讲话者识别部件 61 和面孔识别部件 62 仅进行一次识别处理以识别待学习的人的情形给出了解释，但本发明不限于此，而在无法识别(SID = －1)的情形下，例如，也可不止一次地进行识别处理，而在其他情形下，也可进行数次识别处理。由此做法可改善识别结果的精度。

进而，在以上实施例中，对于对话控制部件 63 由数种识别装置（声音识别部件 60、讲话者识别部件 61、和面孔识别部件 62）所产生的识别结果的多数决定来判断待学习的人是不是新人，但本发明不限于此，而可基于由数个识别装置使用除多数决定外的任何方法所产生的各识别结果来判断待学习的人是不是新人。

在此情形下，可广泛应用各种方法，例如在一种方法中，根据各识别装置的识别能力给数个识别装置的识别结果加权，并基于各加权结果判断一个目标对象是不是新的，而当基于识别能力最高的识别装置和另一识别装置所产生的识别结果判断是新人时，可应用各种其他方法，其中由其余识别装置产生的结果就不用了。

进而，在以上实施例中，对于以下情形给出了解释，即：当讲话者识别部件 61 和面孔识别部件 62 能正确识别人时，通过让讲话者识别部件 61 和面孔识别部件 62 进行累加学习而企图增强因统计稳定性造成的识别精度，但本发明不限于此，而同样，对于存储在内存 65 中的关联信息，也包含了一种功能，以通过让他们任意次地学习同一组合来改善关联信息的可靠性。在实践中，可利用一种使用神经网络的方法来作为这种功能的示例方法，其描述于“Theses of the Academic Society for Electronic Information and communication D-II, Vol. J82-DII, No.6, pp. 1072-1081”。

根据以上所述的本发明，学习器材包括：对话装置，其具有与人类对话的能力，用来通过对话从人类获得目标对象的名字；数个识别装置，每个用来检测目标对象的规定的不同特征，并同时用来基于检测结果、和与事先存储的已知对象对应的特征数据来识别目标对象；存储装置，用来存储关联信息，其中已知对象的名字与由各识别装置获得的关于对象的识别结果互相关联；判断装置，用来基于由对话装置获得的目标对象的名字、由识别装置获得的目标对象的识别结果、和存储在存储装置中的关联信息，判断目标对象是不是新对象；和控制装置，用来当判断装置判断目标对象是新对象时，让识别装置存储对应于目标对象的特征数据，并同时让存储装置存储关于目标对象的关联信息，从而可能使机器人实现自然地通过与凡人的对话来学习新人、新对象等的名字，就像人类常做的那样，从而大大增强它们的娱乐特性。

而且，根据本发明，学习方法包括：第 1 步，与人类对话，并通过对话从人类获得目标对象的名字，以及检测目标对象的数个规定的不同特征，并同时基于检测结果、和事先存储的已知对象的特征数据来识别目标对象；第 3 步，基于所获得的目标对象的名字、以目标对象各特征为基础的识别结果、和将事先存储的已知对象的名字与由识别装置产生的关于对象的识别结果相关联的关联信息，判断目标对象是不是新对象；和第 4 步，当判断装置判断目标对象是新对象时，存储目标对象的各特征的数据和关于目标对象的关联

信息，从而可能使学习方法实现自然地通过与凡人的对话来学习新人、新对象等的名字，就像人类常做的那样，从而大大增强其娱乐特性。

进而，根据本发明，机器人设备包括：对话装置，其具有与人类对话的能力，用来通过对话从人类获得目标对象的名字；数个识别装置，每个用来检测目标对象的规定的不同特征，并同时用来基于检测结果、和与事先存储的已知对象对应的特征数据来识别目标对象；存储装置，用来存储关联信息，其将已知对象的名字与由识别装置获得的关于对象的识别结果相关联；判断装置，用来基于由对话装置获得的目标对象的名字、由识别装置获得的目标对象的识别结果、和存储在存储装置中的关联信息，判断目标对象是不是新对象；和控制装置，用来当判断装置判断目标对象是新对象时，让识别装置存储对应于目标对象的特征数据，并同时让存储装置存储关于目标对象的关联信息，从而可能使机器人实现自然地通过与凡人的对话来学习新人、新对象等的名字，就像人类常做的那样，从而大大增强它们的娱乐特性。

产业可利用性

本发明应用于诸如娱乐机器人、个人计算机、安全系统等各种机器人。

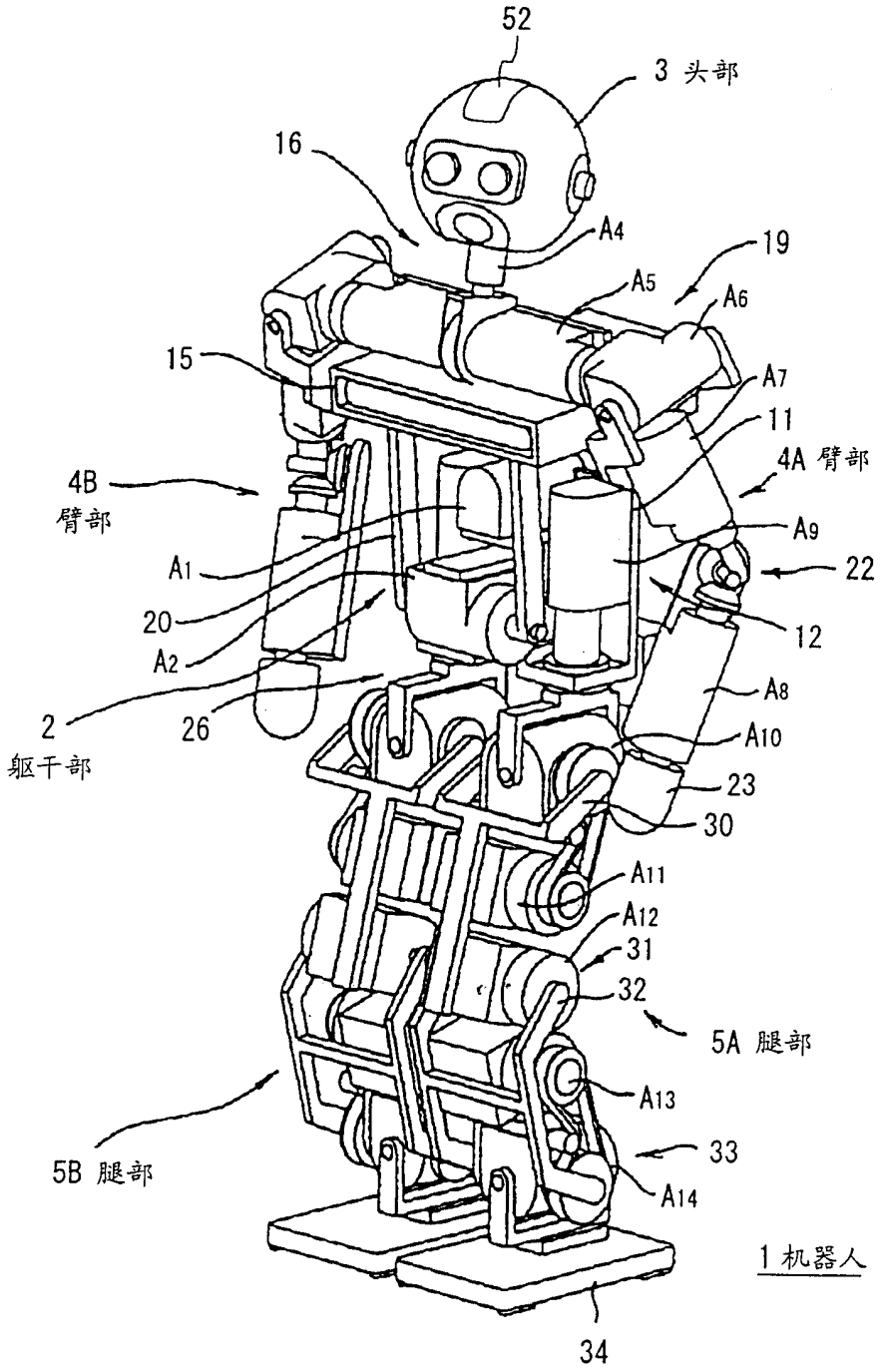

52
3 头部
16
A4
19
A5
A6
15
A7
11
4A 臂部
4B
臂部
A9
A1
22
20
12
A2
A8
2
26
A10
躯干部
23
30
A11
A12
31
32
5A 腿部
A13
5B 腿部
33
A14
1 机器人
34

图 1

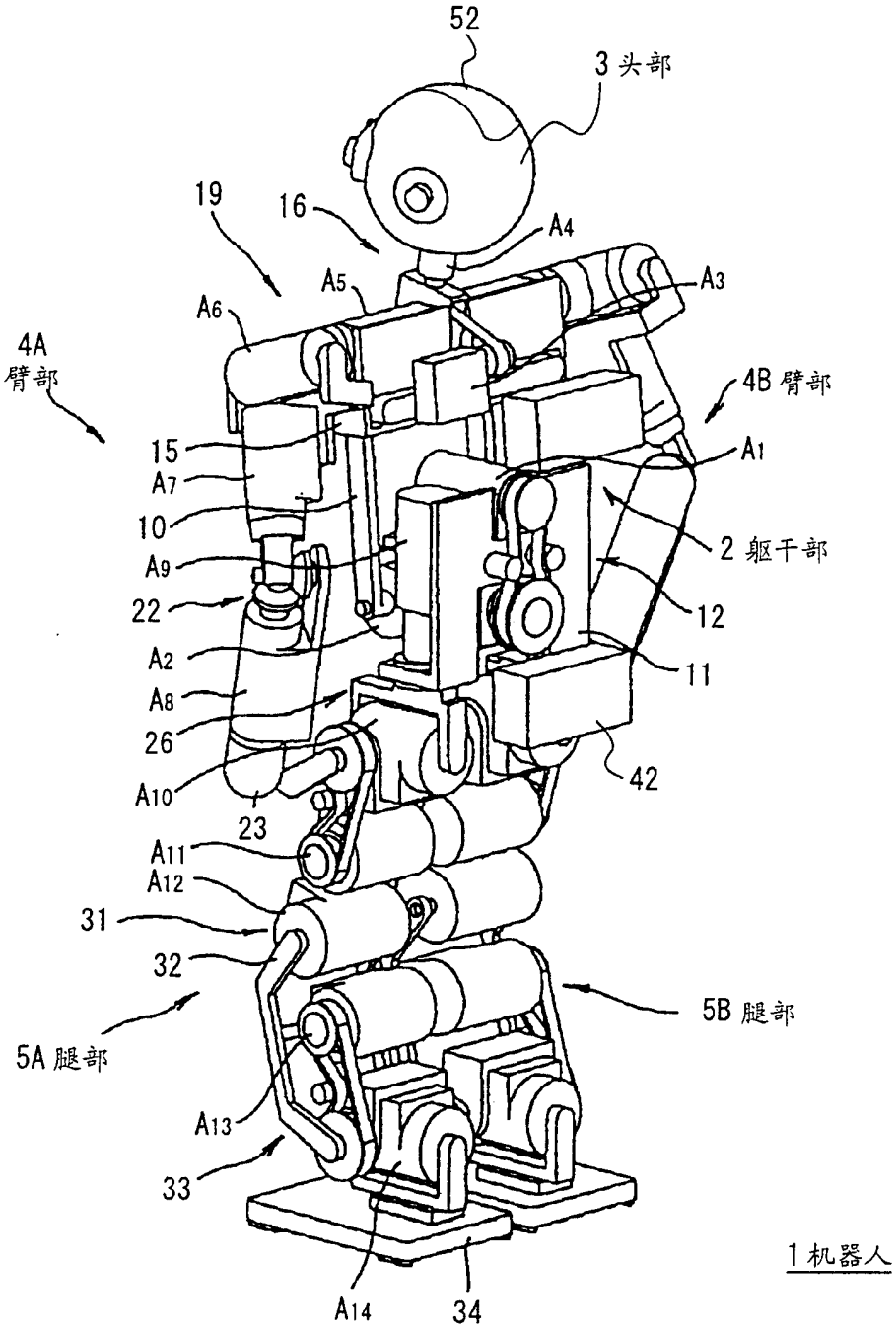

52
3 头部
19
16
A4
A5
A3
A6
4A
臂部
4B 臂部
15
A1
A7
10
2 躯干部
A9
22
12
A2
11
A8
26
42
A10
23
A11
A12
31
32
5B 腿部
5A 腿部
A13
33
A14
34
1 机器人

图 2

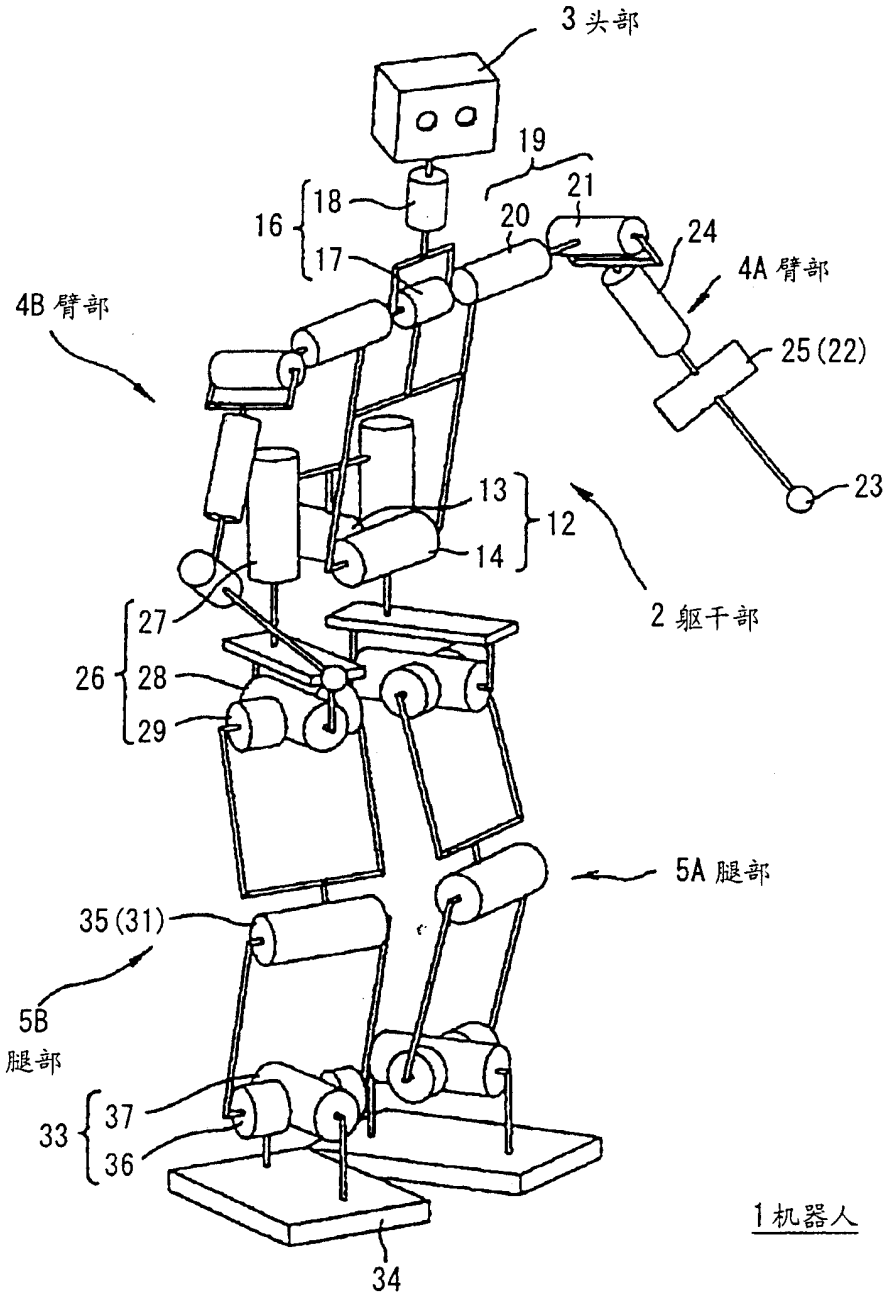

3 头部
19
21
16
18
17
20
24
4A 臂部
4B 臂部
25(22)
23
13
12
14
2 躯干部
27
26
28
29
5A 腿部
35(31)
5B
腿部
37
33
36
34
1 机器人

图 3

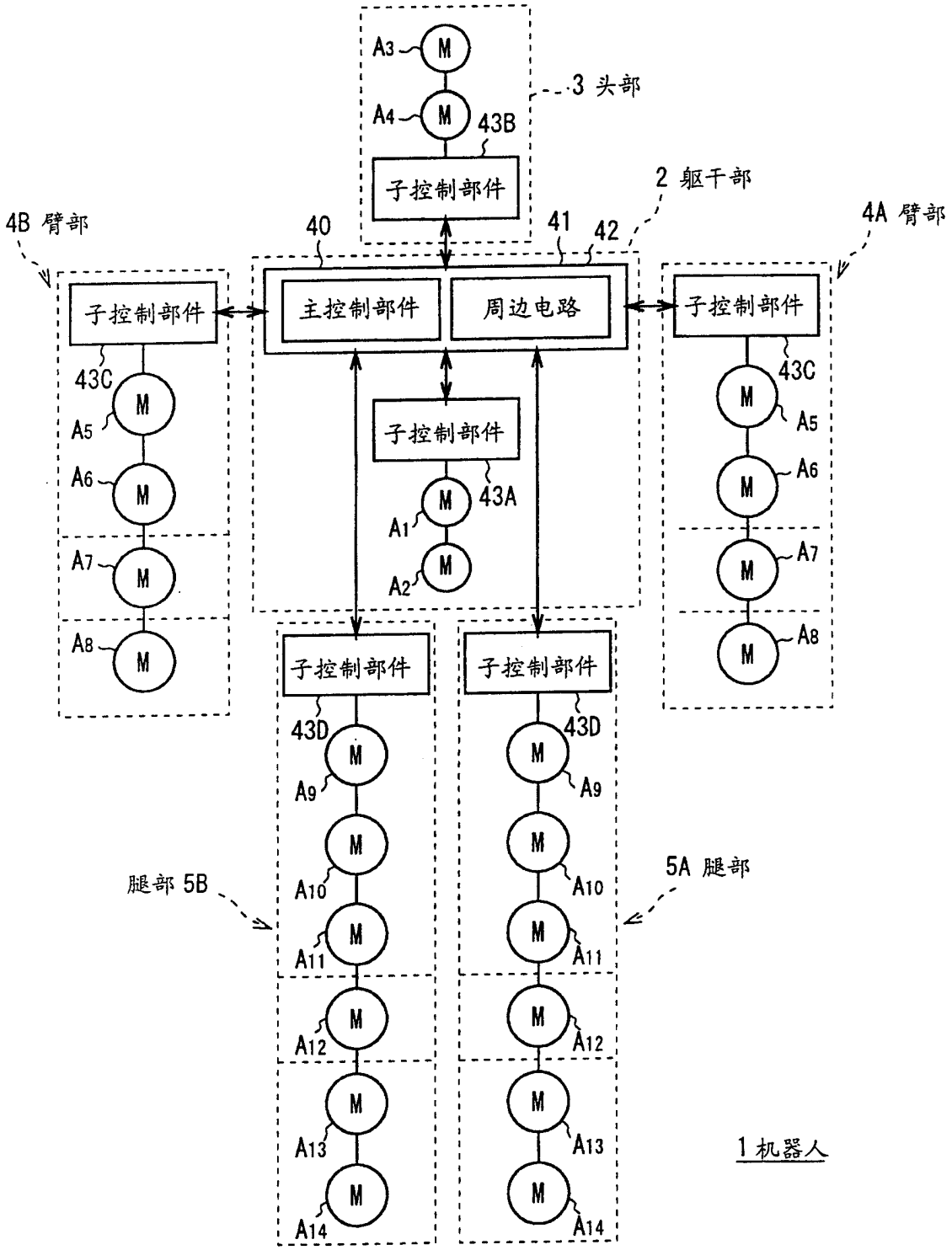

A3
M
A4
M
43B
子控制部件
3 头部
2 躯干部
4B 臂部
40
41
42
4A 臂部
子控制部件
主控制部件
周边电路
子控制部件
43C
A5
A6
A7
A8
子控制部件
43A
A1
A2
子控制部件
子控制部件
43D
A9
A10
A11
A12
A13
A14
腿部 5B
5A 腿部
1 机器人

图 4

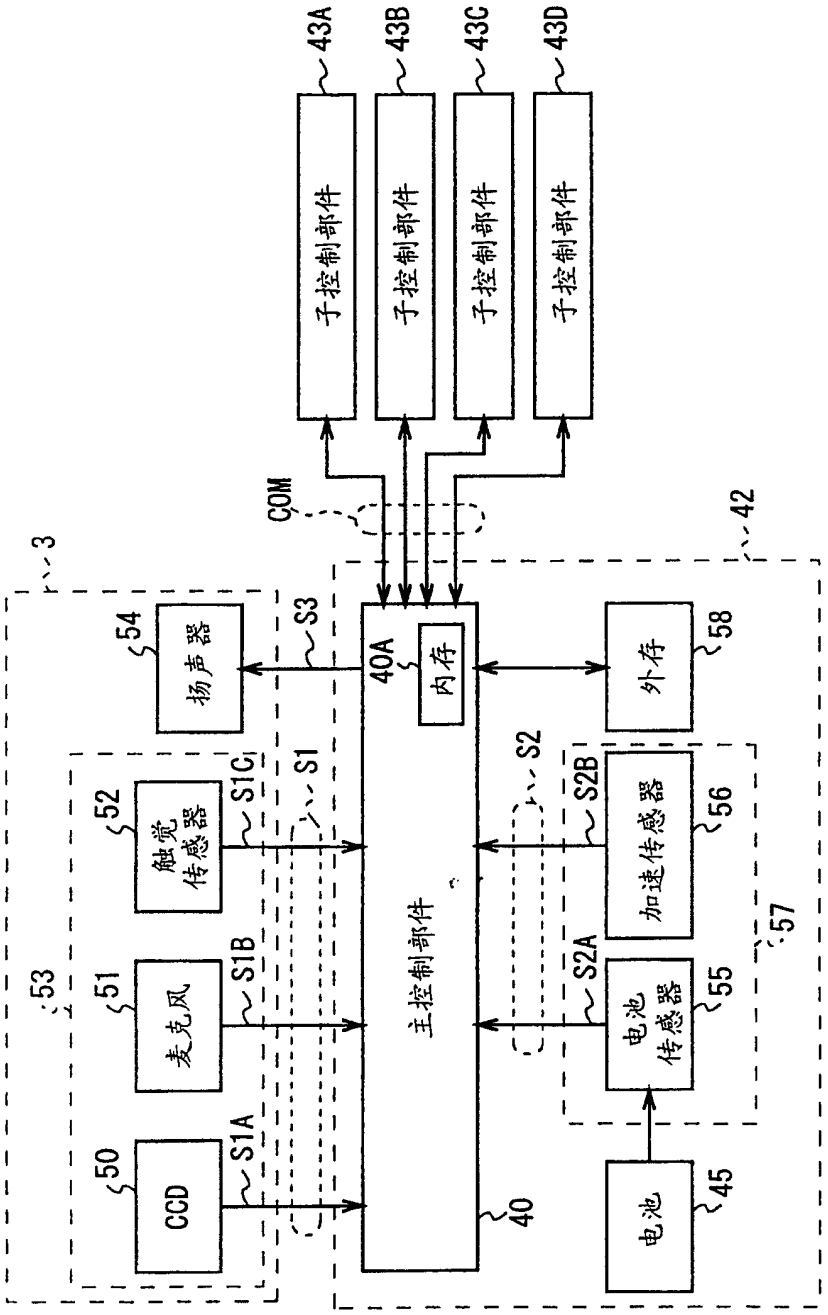

43A
43B
43C
43D
子控制部件
子控制部件
子控制部件
子控制部件
COM
3
42
54
扬声器
S3
40A
内存
58
外存
52
触觉传感器
S1C
S1
S2
S2B
56
加速传感器
53
51
麦克风
S1B
主控制部件
S2A
55
电池传感器
57
50
CCD
S1A
40
45
电池

图 5

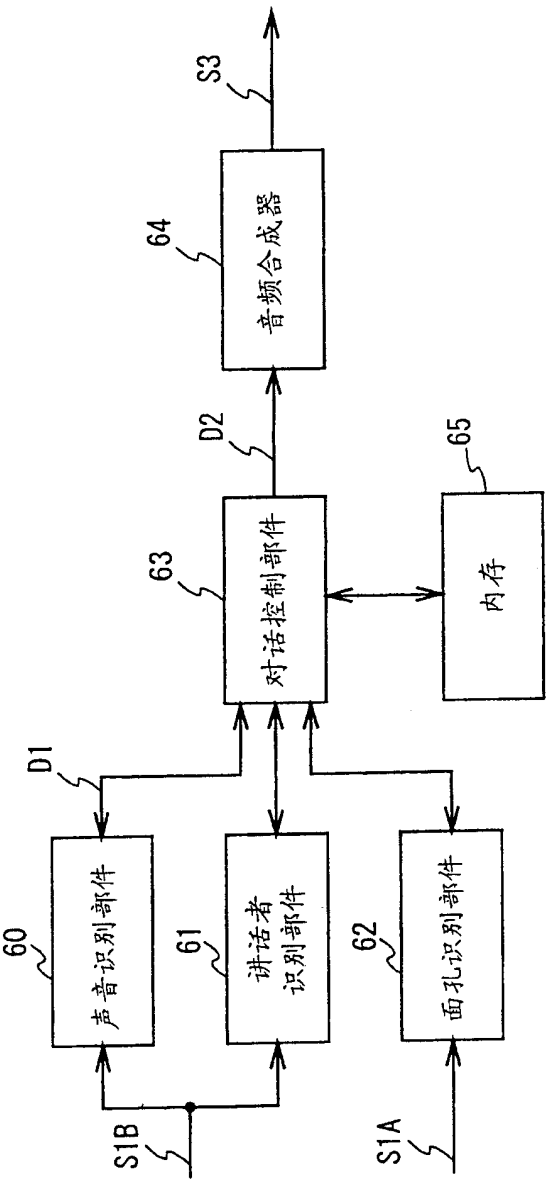

S3
64
音频合成器
D2
65
63
对话控制部件
内存
D1
60
声音识别部件
61
讲话者
识别部件
62
面孔识别部件
S1B
S1A

图 6

FID	SID	名字
1	2	藤田
2	5	吉田

图 7

FID	SID	名字
1	2	藤田
2	5	吉田
4	6	山本

图 12

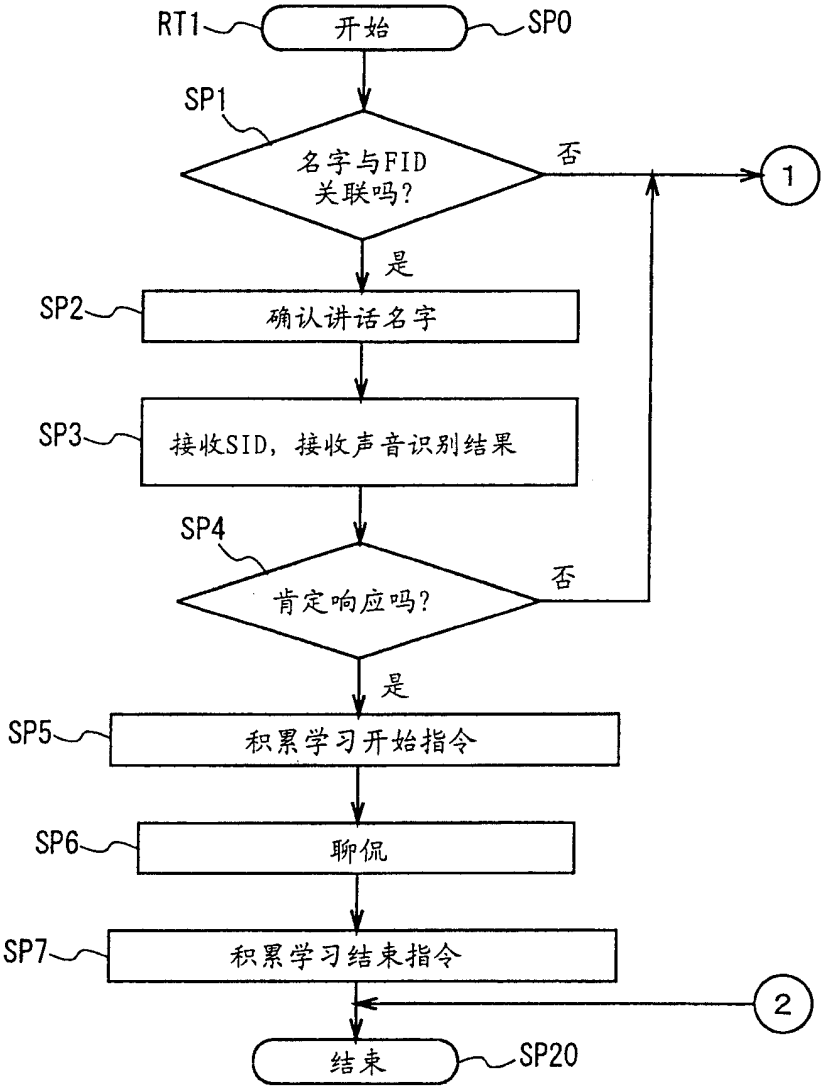

RT1
开始
SP0
SP1
名字与FID
关联吗?
否
1
是
SP2
确认讲话名字
SP3
接收SID，接收声音识别结果
SP4
肯定响应吗?
否
是
SP5
积累学习开始指令
SP6
聊侃
SP7
积累学习结束指令
2
结束
SP20

图 8

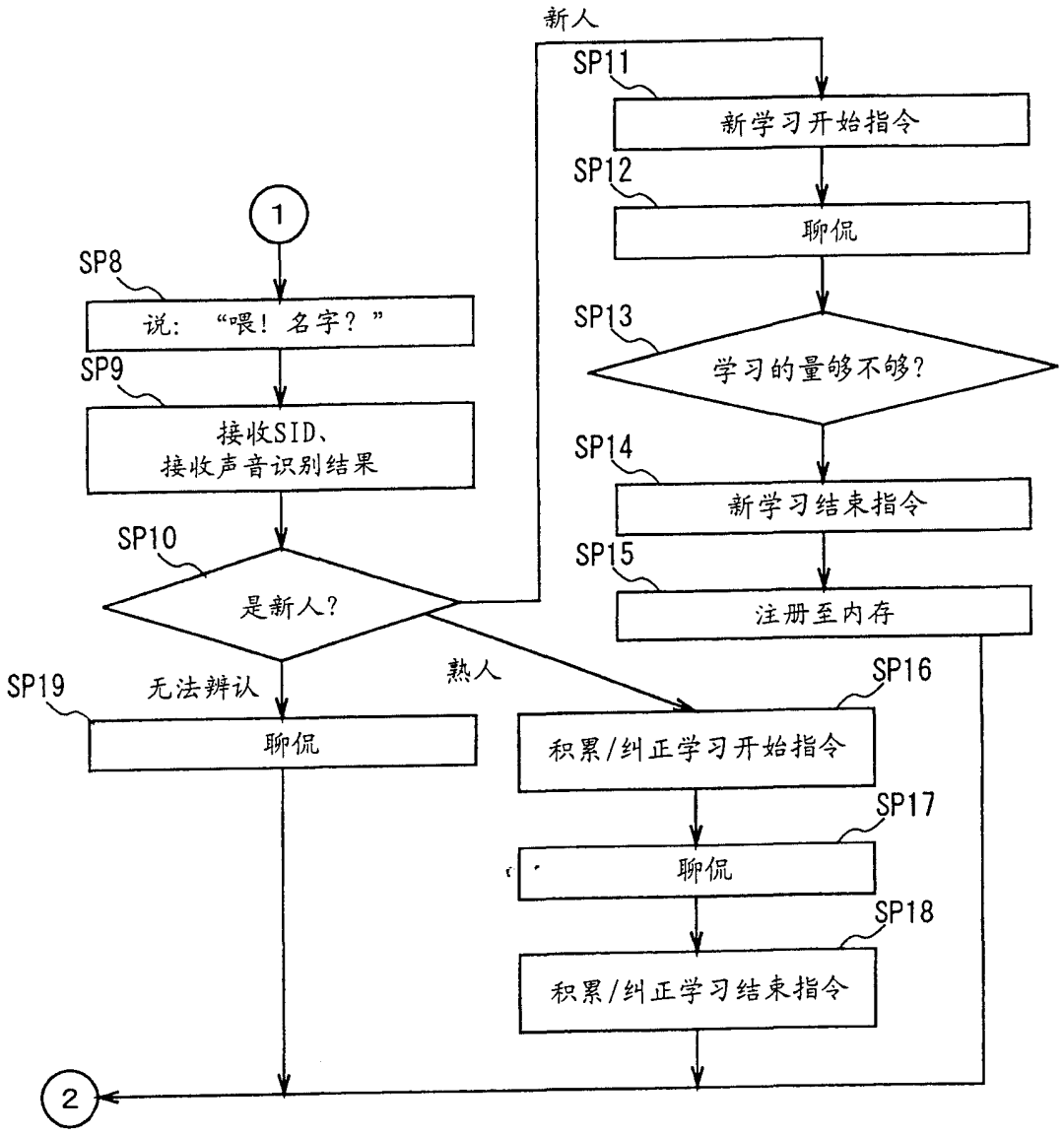

新人
SP11
新学习开始指令
SP12
聊侃
SP13
学习的量够不够?
SP14
新学习结束指令
SP15
注册至内存
1
SP8
说: “喂! 名字? ”
SP9
接收SID、
接收声音识别结果
SP10
是新人?
熟人
SP16
积累/纠正学习开始指令
SP17
聊侃
SP18
积累/纠正学习结束指令
SP19
无法辨认
聊侃
2

图 9

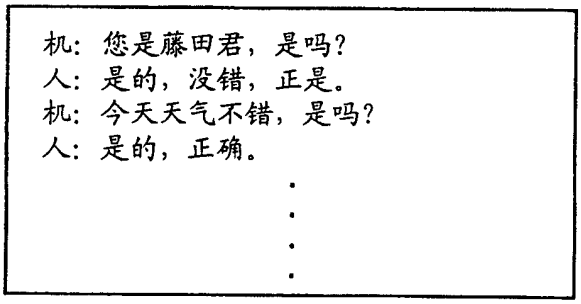

机：您是藤田君，是吗？
人：是的，没错，正是。
机：今天天气不错，是吗？
人：是的，正确。

图 10

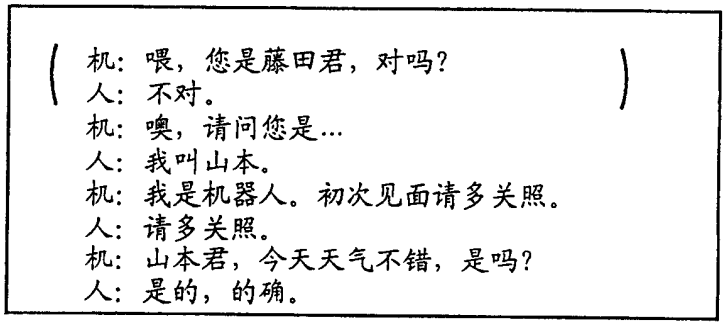

机：喂，您是藤田君，对吗？
人：不对。
机：噢，请问您是…
人：我叫山本。
机：我是机器人。初次见面请多关照。
人：请多关照。
机：山本君，今天天气不错，是吗？
人：是的，的确。

图 11

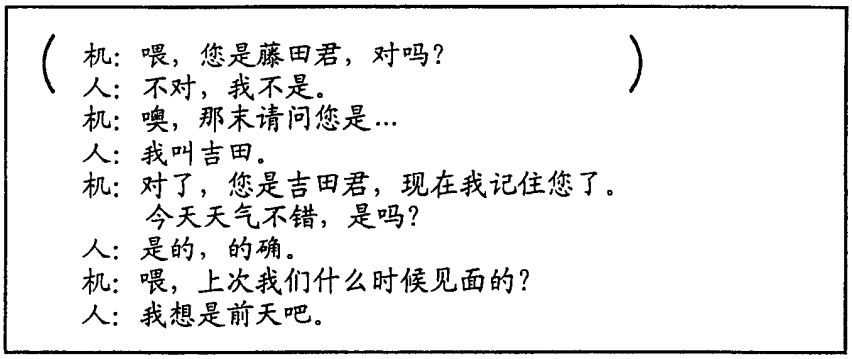

机：喂，您是藤田君，对吗？
人：不对，我不是。
机：噢，那末请问您是…
人：我叫吉田。
机：对了，您是吉田君，现在我记住您了。
今天天气不错，是吗？
人：是的，的确。
机：喂，上次我们什么时候见面的？
人：我想是前天吧。

图 13

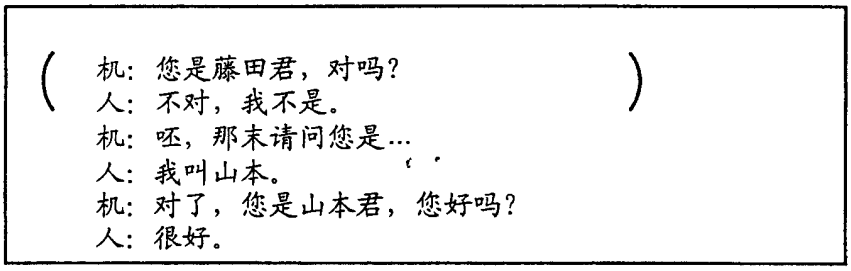

机：您是藤田君，对吗？
人：不对，我不是。
机：呸，那末请问您是…
人：我叫山本。
机：对了，您是山本君，您好吗？
人：很好。

图 14

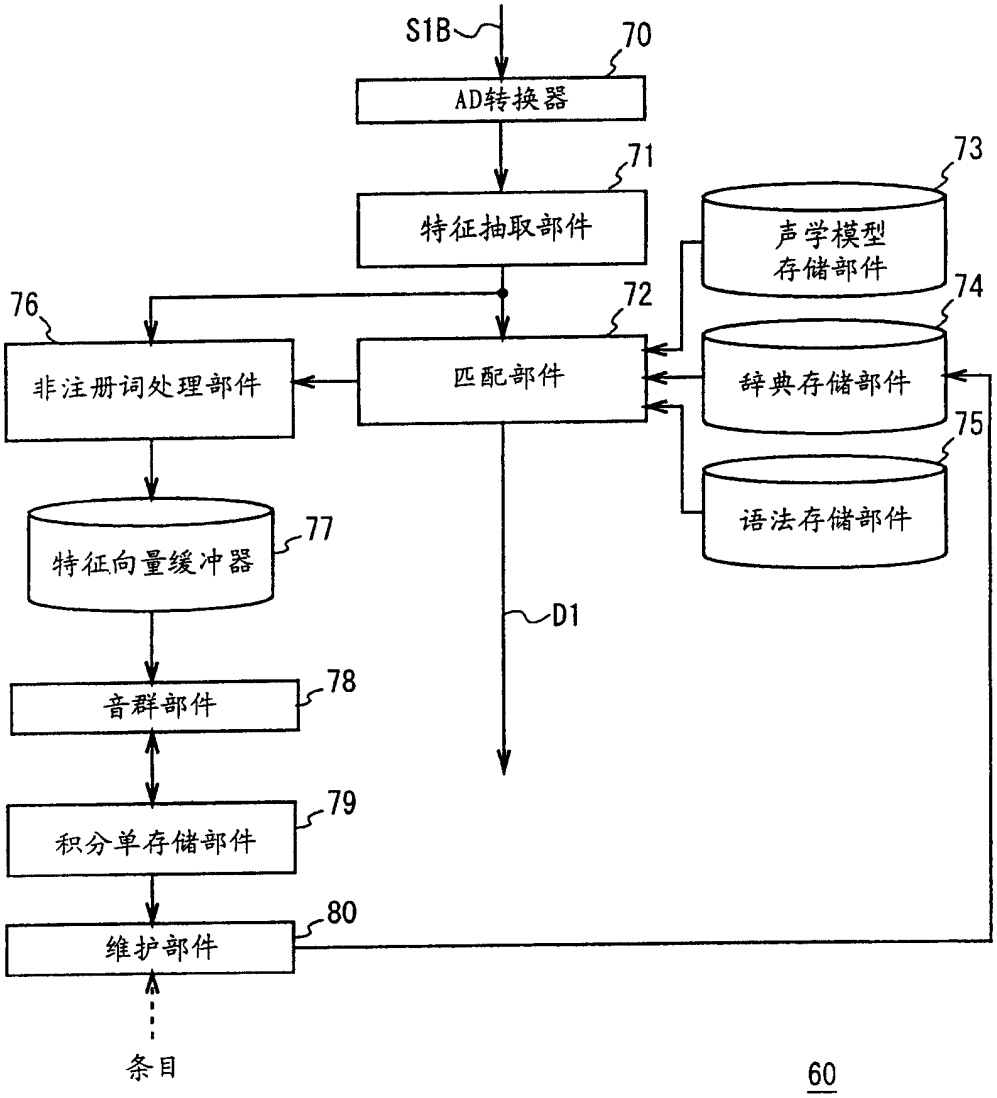

S1B
70
AD转换器
71
特征抽取部件
73
声学模型
存储部件
76
72
74
非注册词处理部件
匹配部件
辞典存储部件
75
语法存储部件
77
特征向量缓冲器
D1
78
音群部件
79
积分单存储部件
80
维护部件
条目
60

图 15

条目		音素系列
boku	[我](男子在同辈与晚辈自称)	boku
chigau	[不同]	chigau
doko	[哪里]	doko
genki	[健康]	geNki
iro	[色彩]	iro
janai	[不是的]	janai
kirai	[讨厌、不喜欢]	kirai
kudasai	[请]	kudasai

图 16

```
$col     =  [kono | sono] iro wa;
$this    =  kore [(ga | wa | mo)];
$neg     =  (Chagau | iie) [$sil];
$null    =  $sil;
$des     =  (desu | da) [yó]˙| yo;
$not     =  janai [yo];
$color1  =  $null | $neg | [$neg] $col | [$neg] $this;
$color2  =  [iro] (desu | janai | da) [yo];
$pat1    =  $color $garbage $color2;
```

图 17

ID	音素系列	特征向量系列
1	——	——
2	——	——
·	·	·
·	·	·
·	·	·
N	——	——
N+1	——	——

图 18

音素系列	言词
doroa:	FURO (BATH)×1;
kuro	FURO×3;
Nfuro	FURO×20;
NhoNn	HON (BOOK)×18;
hoNN	HON×6;
NhoNda	HON×10;
NhoNde:su	HON×4;
ohoN	ORANGE×1;HON×19;
hoNgdawasoNre:a:	HON×2;
a:modori:	GREEN COLOR×11;
omidori:	GREEN COLOR×10;
e:imidori:	GREEN COLOR×3;
Nmidori:	GREEN COLOR×5;
a:midori:iroiresu	GREEN COLOR×4;
Nro:ka	ROKA (PASSAGE)×10;
Nro:kaNa	ROKA×10;

图 23

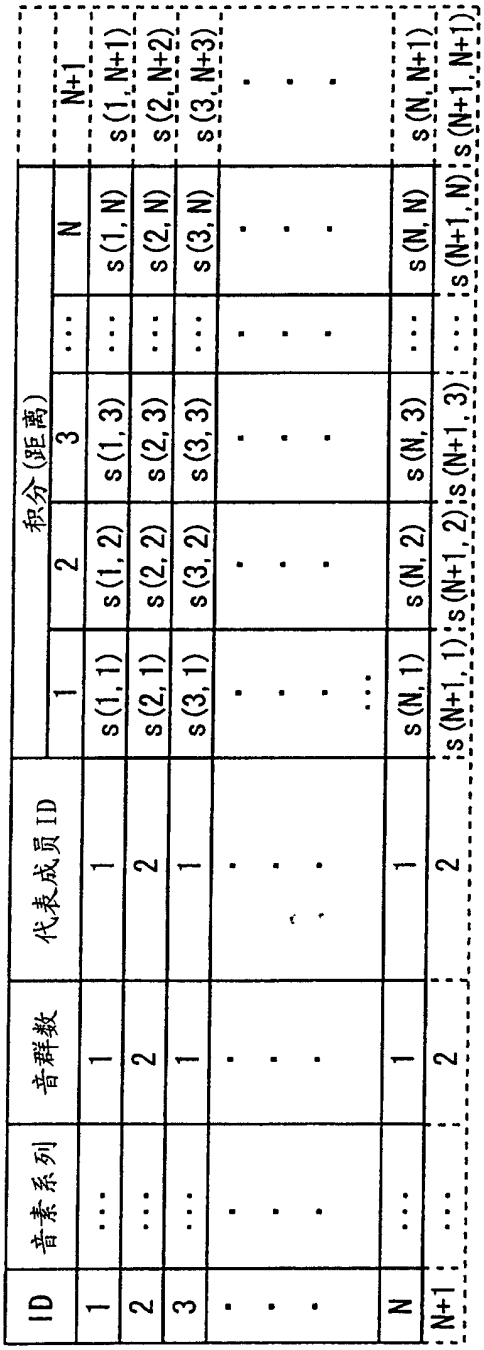

ID
音素系列
音群数
代表成员ID
积分(距离)
1
2
3
…
N
N+1
1
…
1
1
s(1,1)
s(1,2)
s(1,3)
…
s(1,N)
s(1,N+1)
2
…
2
2
s(2,1)
s(2,2)
s(2,3)
…
s(2,N)
s(2,N+2)
3
…
1
1
s(3,1)
s(3,2)
s(3,3)
…
s(3,N)
s(3,N+3)
N
…
1
1
s(N,1)
s(N,2)
s(N,3)
…
s(N,N)
s(N,N+1)
N+1
…
2
2
s(N+1,1)
s(N+1,2)
s(N+1,3)
…
s(N+1,N)
s(N+1,N+1)

图 19

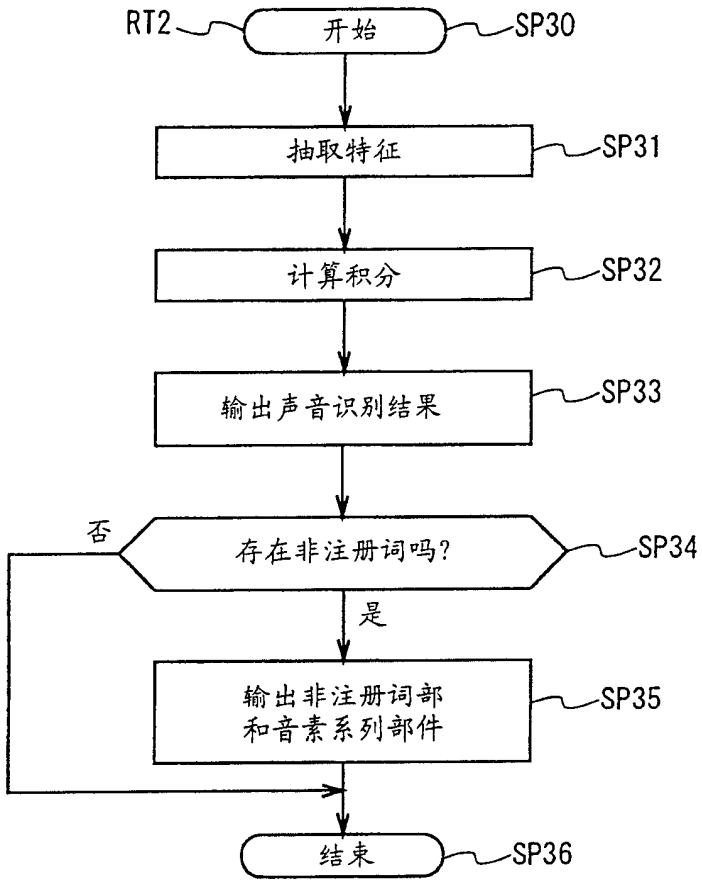

RT2
开始
SP30
抽取特征
SP31
计算积分
SP32
输出声音识别结果
SP33
否
存在非注册词吗?
SP34
是
输出非注册词部
和音素系列部件
SP35
结束
SP36

图 20

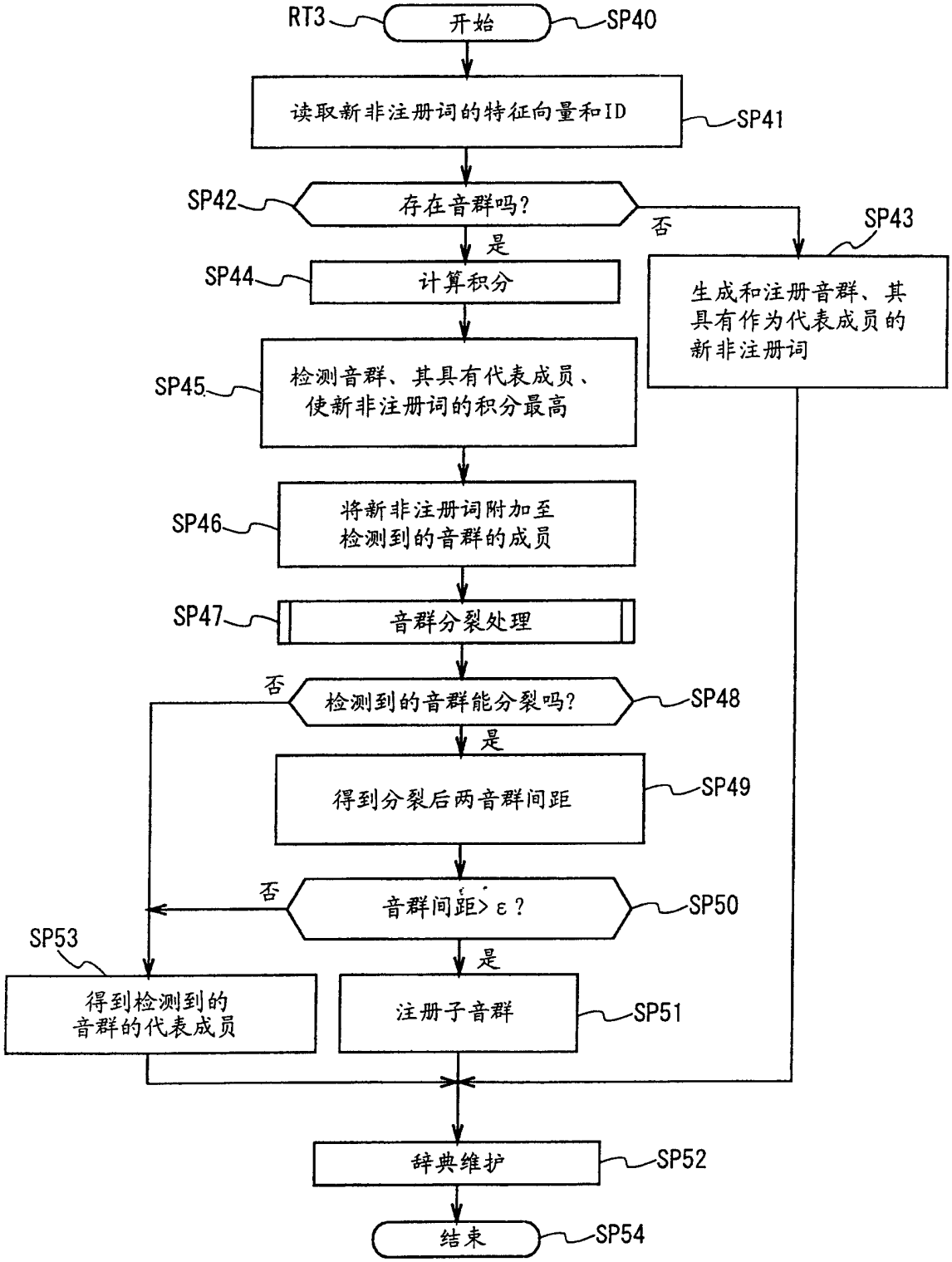

RT3
开始
SP40
读取新非注册词的特征向量和ID
SP41
SP42
存在音群吗?
否
SP43
是
SP44
计算积分
生成和注册音群、其具有作为代表成员的新非注册词
SP45
检测音群、其具有代表成员、使新非注册词的积分最高
SP46
将新非注册词附加至检测到的音群的成员
SP47
音群分裂处理
否
检测到的音群能分裂吗?
SP48
是
得到分裂后两音群间距
SP49
否
音群间距> ε ?
SP50
SP53
是
得到检测到的音群的代表成员
注册子音群
SP51
辞典维护
SP52
结束
SP54

图 21

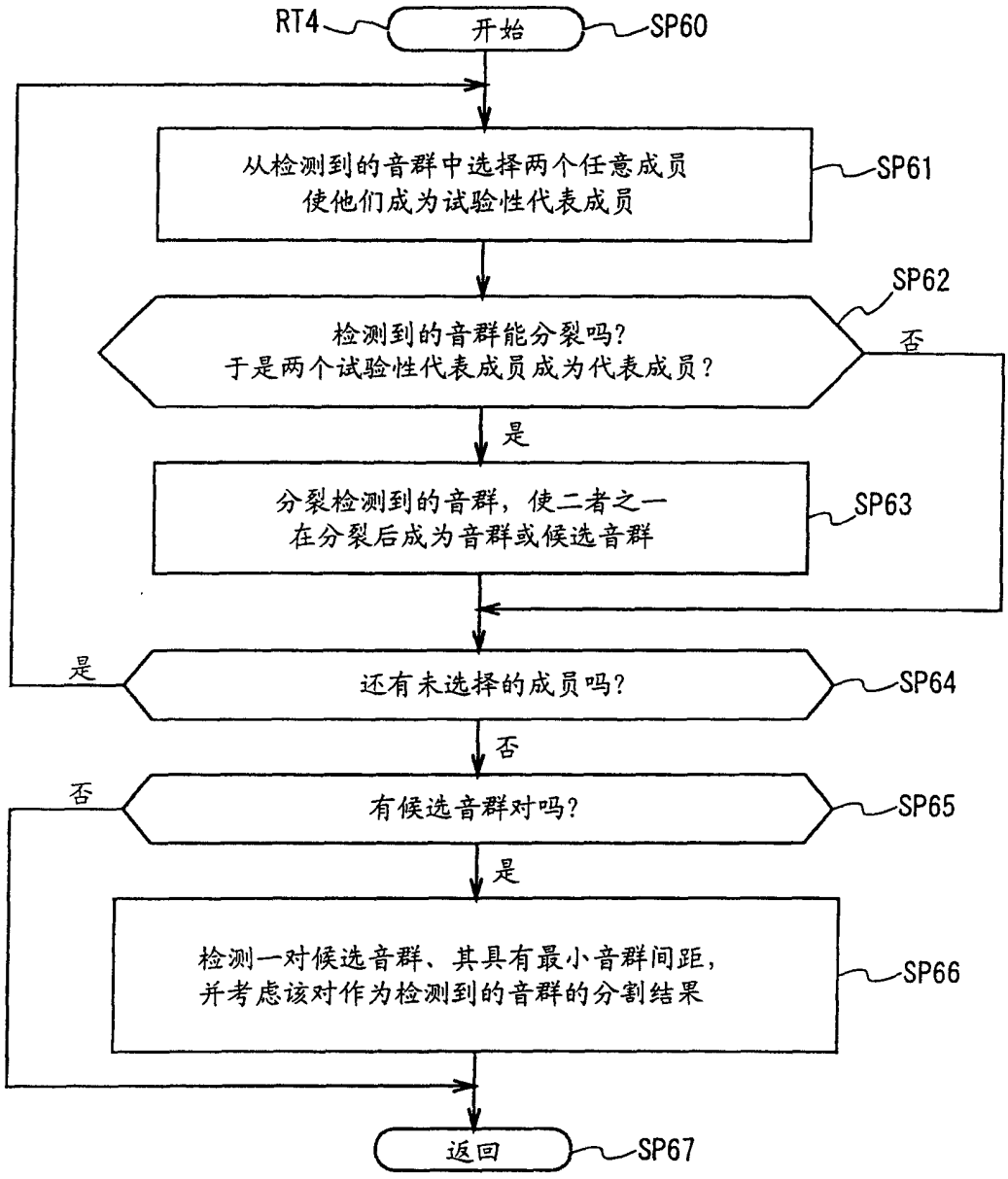

RT4
开始
SP60
从检测到的音群中选择两个任意成员
使他们成为试验性代表成员
SP61
SP62
检测到的音群能分裂吗?
于是两个试验性代表成员成为代表成员?
否
是
分裂检测到的音群，使二者之一
在分裂后成为音群或候选音群
SP63
是
还有未选择的成员吗?
SP64
否
否
有候选音群对吗?
SP65
是
检测一对候选音群、其具有最小音群间距，
并考虑该对作为检测到的音群的分割结果
SP66
返回
SP67

图 22

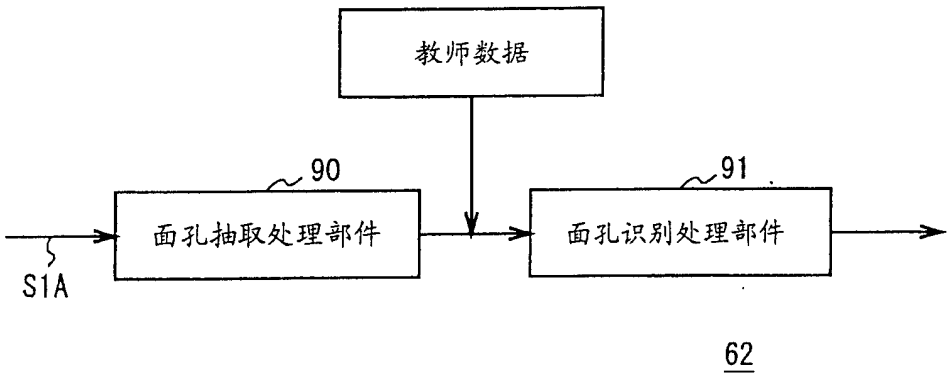

教师数据
90
91
面孔抽取处理部件
面孔识别处理部件
S1A
62

图 24

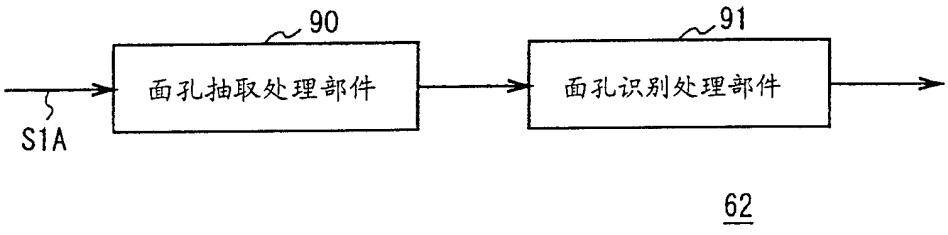

90
91
面孔抽取处理部件
面孔识别处理部件
S1A
62

图 25