



(12)发明专利申请

(10)申请公布号 CN 110998718 A

(43)申请公布日 2020.04.10

(21)申请号 201880048424.1

(22)申请日 2018.07.18

(30)优先权数据

2017-149134 2017.08.01 JP

(85)PCT国际申请进入国家阶段日

2020.01.20

(86)PCT国际申请的申请数据

PCT/JP2018/026824 2018.07.18

(87)PCT国际申请的公布数据

W02019/026617 JA 2019.02.07

(71)申请人 索尼公司

地址 日本东京

(72)发明人 东山惠祐 大村淳己 小川浩明

西川加奈 栗屋志伸 鹤田泰士

(74)专利代理机构 北京康信知识产权代理有限公司 11240

代理人 余刚

(51)Int.Cl.

G10L 15/22(2006.01)

G10L 15/26(2006.01)

G10L 15/10(2006.01)

G06F 3/01(2006.01)

G06F 3/16(2006.01)

G06T 7/20(2017.01)

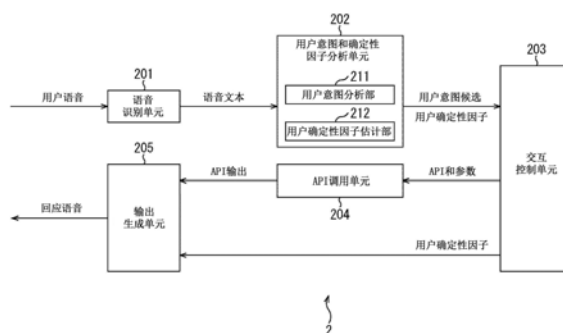
权利要求书1页 说明书24页 附图21页

(54)发明名称

信息处理设备和信息处理方法

(57)摘要

本技术涉及一种能够在交互式语音系统中向用户提供具有适当信息量的回应的信息处理设备和信息处理方法。所述信息处理设备包括处理单元,所述处理单元根据为用户估计的置信因子,针对用户期望的信息来调整用于回应的信息量。因此,可以在交互式语音系统中向用户提供具有适当信息量的回应。本技术可应用于例如交互式语音系统。



1. 一种信息处理器,包括处理单元,所述处理单元根据针对用户的期望信息而估计的用户自身的确定性因子,来调整对期望信息的回应的信息量。
2. 根据权利要求1所述的信息处理器,其中,
基于从用户的语音中获得的信息来估计所述确定性因子,并且
所述回应是对用户语音的回应。
3. 根据权利要求1所述的信息处理器,其中,基于通过预先学习获得的趋势,从用户行为估计所述确定性因子。
4. 根据权利要求1所述的信息处理器,其中,所述确定性因子和包括在所述回应中的信息量的关系是线性或非线性关系。
5. 根据权利要求4所述的信息处理器,其中,所述关系是比例关系或反比关系。
6. 根据权利要求1所述的信息处理器,其中,基于从传感器获得的传感器信息来估计所述确定性因子。
7. 根据权利要求6所述的信息处理器,其中,所述传感器信息至少包括用户的手势识别信息、视线识别信息、面部取向识别信息和位置信息中的任一个。
8. 根据权利要求1所述的信息处理器,其中,
所述期望信息是关于另一用户的信息,并且
所述确定性因子根据期望信息的目标人员而变化。
9. 根据权利要求8所述的信息处理器,其中,在目标人员是另一用户的情况下,所述确定性因子低于目标人员是用户的情况下的确定性因子。
10. 根据权利要求1所述的信息处理器,其中,通过将从期望信息获得的值传递到对应于用户意图的API (应用编程接口) 作为参数,并执行所述API,来获得所述回应。
11. 一种信息处理器的信息处理方法,所述信息处理方法包括由信息处理器根据针对用户的期望信息而估计的用户自身的确定性因子,来调整对期望信息的回应的信息量的步骤。

信息处理设备和信息处理方法

技术领域

[0001] 本技术涉及一种信息处理器和一种信息处理方法,尤其涉及一种使语音交互系统能够向用户呈现具有适当信息量的回应的信息处理器和信息处理方法。

背景技术

[0002] 在用户通过语音交互系统执行信息的获取或确认的情况下,语音交互系统通常选择并执行对应于用户语音意图的应用程序API(应用编程接口),并将由此获得的结果作为回应文本呈现给用户。

[0003] 例如,作为语音交互系统,公开了一种根据系统侧的确定性因子切换回应的语音交互系统(例如,参见PTL1)。

[0004] 引用列表

[0005] 专利文献

[0006] PTL 1:日本未审查专利申请公开号2011-54088

发明内容

[0007] 本发明要解决的问题

[0008] 顺便提及,一般的语音交互系统通常在相同的规则下为相同的API调用生成回应。因此,语音交互系统在任何给定时间返回包括相同信息量的回应,而不管用户对用户期望的信息有多确定。

[0009] 因此,即使在用户仅希望确认用户知道的信息是否正确的情況下,语音交互系统也做出包括系统知道的所有信息等的回应。这可能导致用户期望的信息量和从语音交互系统输出的信息量之间的矛盾,从而导致缺乏顺利通信。

[0010] 根据这种情况而产生本技术,并且使语音交互系统能够向用户呈现具有适当信息量的回应。

[0011] 解决问题的方法

[0012] 本技术的一个方面的信息处理器是一种信息处理器,包括处理单元,所述处理单元根据为期望信息估计的用户自己的确定性因子,调整对用户的期望信息的回应的信息量。

[0013] 本技术方面的信息处理器可以是独立的设备或包含在单个设备中的内部块。

[0014] 本技术的一个方面的信息处理方法是一种信息处理器的信息处理方法,所述信息处理方法包括由信息处理器根据为期望信息估计的用户自己的确定性因子,调整对用户的期望信息的回应的信息量的步骤。

[0015] 在本技术的各个方面的信息处理器和信息处理方法中,根据为用户期望的信息估计的用户自身的确定性因子,来调整对用户的期望信息的回应的信息量。

[0016] 本发明的效果

[0017] 根据本技术的一个方面,在语音交互系统中,有可能向用户呈现具有适当信息量

的回应。

[0018] 应当注意,此处描述的效果不一定是限制性的,并且可以提供本公开中描述的任何效果。

附图说明

- [0019] 图1是示出通用语音交互系统的配置的框图;
- [0020] 图2是示出应用本技术的语音交互系统的实施方式的配置示例的框图;
- [0021] 图3是示出语音交互系统的配置示例的框图;
- [0022] 图4是示出语音处理设备的配置示例的框图;
- [0023] 图5是示出服务器的配置示例的框图;
- [0024] 图6是示出时间表DB的示例的示图;
- [0025] 图7是示出用户的确定性因子和系统回应的信息量的关系的示例的示图;
- [0026] 图8是示出应用本技术的语音交互系统的另一配置示例的框图;
- [0027] 图9是示出时间表DB的示例的示图;
- [0028] 图10是示出用户的确定性因子和系统回应中的信息量的关系的示例的示图;
- [0029] 图11是示出应用本技术的语音交互系统的另一配置示例的框图;
- [0030] 图12是示出通用文本聊天系统的配置的框图;
- [0031] 图13是示出应用本技术的语音交互系统的配置的第一变型例的框图;
- [0032] 图14是示出应用本技术的语音交互系统的配置的第二变型例的框图;
- [0033] 图15是示出应用本技术的语音交互系统的配置的第三变型例的框图;
- [0034] 图16是示出用户的确定性因子和系统回应中的信息量的关系的第一变型例的示图;
- [0035] 图17是示出用户的确定性因子和系统回应中的信息量的关系的第二变型例的示图;
- [0036] 图18是示出用户的确定性因子和系统回应中的信息量的关系的第三变型例的示图;
- [0037] 图19是示出用户的确定性因子和系统回应中的信息量的关系的第四变型例的示图;
- [0038] 图20是描述语音交互处理的流程图;
- [0039] 图21是示出计算机的配置示例的示图。

具体实施方式

- [0040] 下面参考附图描述本技术的实施方式。应当注意,按照以下顺序给出描述。
- [0041] 1. 系统配置
- [0042] 2. 本技术的实施方式
- [0043] (1) 第一实施方式:基本形式
- [0044] (1-1) 用户确定性因子高的情况
- [0045] (1-2) 用户确定性因子中等的情况。
- [0046] (1-3) 用户确定性因子低的情况

- [0047] (1-4) 虽然用户的确定性因子似乎很高,但事实上并不正确的情况
- [0048] (1-5) 实际上不正确的情况下有多个选择时
- [0049] (2) 第二实施方式:时间表的目标人员不同于说话者的情况
- [0050] (2-1) 用户自己对另一个人的时间表的确定性因子高的情况
- [0051] (2-2) 用户自己对另一个人的时间表的确定性因子中等的情况
- [0052] (2-3) 用户自己对另一个人的时间表的确定性因子低的情况
- [0053] (3) 第三实施方式:利用传感器信息的情况
- [0054] 3. 变型例
- [0055] 4. 语音交互处理流程
- [0056] 5. 计算机的配置
- [0057] <1. 系统配置>
- [0058] (通用语音交互系统的配置)
- [0059] 图1是示出通用语音交互系统的配置的框图。
- [0060] 在图1中,通用语音交互系统1包括语音识别单元101、用户意图分析单元102、交互控制单元103、API调用单元104和输出生成单元105。
- [0061] 语音识别单元101将输入到语音识别单元101的用户语音数据转换成语音文本数据,并将语音文本数据提供给用户意图分析单元102。
- [0062] 用户意图分析单元102处理从语音识别单元101提供的文本数据,以分析用户意图,并且向交互控制单元103提供指示作为分析结果获得的用户意图的候选的数据(以下称为用户意图候选数据)。
- [0063] 基于从用户意图分析单元102等提供的用户意图候选数据,交互控制单元103估计当前交互状态,并确定要调用的API及其参数。交互控制单元103向API调用单元104提供指示要调用的API和API的参数的数据(以下称为API和参数数据)。
- [0064] API调用单元104将参数传递给要调用的API,并基于从交互控制单元103提供的API和参数数据来执行API。API调用单元104将通过执行API获得的结果作为API输出提供给输出生成单元105。
- [0065] 输出生成单元105基于从API调用单元104提供的API输出,生成对用户的回应文本(以下称为系统回应),并输出与其语音数据相对应的回应语音。
- [0066] 通用语音交互系统1如上所述配置。
- [0067] 在此处,通用语音交互系统1(图1)执行相同的API调用,并通过相同的方法使用API调用的结果来执行回应生成,而不管用户对用户希望确认的时间表内容的确定程度(确定性因子)如何。因此,在任何给定时间都会获得包含相同信息量的回应。
- [0068] 因此,在用户希望执行简单确认的情况下和用户希望确认时间表细节的情况下,获得相同的回应。
- [0069] 因此,即使在用户希望执行简单确认的情况下,也可以获得具有大量信息的回应,这可能会减慢交互的速度并导致缺乏顺利通信。类似地,即使在用户期望大量信息的情况下,语音交互系统也可能不会根据语音交互系统的实现返回详细信息。这使得寻求更多信息需要额外的交互,并且可能导致缺乏顺利通信。
- [0070] 因此,在应用本技术的语音交互系统中,通过估计用户自己关于用户期望的信息

(期望信息)的确定性因子,并根据估计结果控制回应生成中包括的信息量,可以向用户呈现具有适当信息量的回应,从而允许实现更顺利通信。在下文中,描述应用本技术的语音交互系统的配置。

[0071] (本技术的语音交互系统的配置)

[0072] 图2是示出应用本技术的语音交互系统的实施方式的配置示例的框图。

[0073] 在图2中,语音交互系统2包括语音识别单元201、用户意图和确定性因子分析单元202、交互控制单元203、API调用单元204和输出生成单元205。

[0074] 通过参考语音到文本转换数据库,语音识别单元201将输入到其中的用户语音数据转换成语音文本数据。语音识别单元201将对应于用户语音的语音文本数据提供给用户意图和确定性因子分析单元202。

[0075] 在此处,从获取从用户输出的语音的语音处理设备(例如,稍后将描述的图3中的语音处理设备10)发送用户语音数据。此外,在语音到文本转换数据库中累积了巨大的语音识别数据库。通过使用预定的识别算法,语音识别单元201能够将输入到其中的用户语音数据(语音数据)转换成语音文本数据(文本数据)。

[0076] 用户意图和确定性因子分析单元202对从语音识别单元201提供的语音文本数据执行分析处理。用户意图和确定性因子分析单元202包括用户意图分析部211和用户确定性因子估计部212。

[0077] 用户意图分析部211通过对从语音识别单元201提供的语音文本数据执行分析处理来分析用户意图。用户意图分析部211将获得的用户意图候选数据作为分析结果提供给交互控制单元203。

[0078] 用户确定性因子估计部212通过对从语音识别单元201提供的语音文本数据执行分析处理来估计用户的确定性因子。用户确定性因子估计部212向交互控制单元203提供指示获得的用户确定性因子的数据(以下称为用户确定性因子数据)作为估计结果。

[0079] 基于从用户意图和确定性因子分析单元202、交互历史数据库等提供的用户意图候选数据,交互控制单元203估计当前交互状态,并确定要调用的API和API的参数。应当注意,交互历史数据库累积关于过去的交互历史等的信息。

[0080] 交互控制单元203将确定的API和参数数据提供给API调用单元204。此外,交互控制单元203将用户确定性因子数据提供给输出生成单元205。

[0081] API调用单元204将API的参数传递给要调用的API,并基于从交互控制单元203提供的API和参数数据来执行API。API调用单元204向输出生成单元205提供通过执行API获得的结果,作为API输出。

[0082] 在此处,要调用的API是对应于用户意图的API。例如,在用户的意图涉及检查时间的情况下,调用并执行用于执行时间表检查的API(以下也称为时间表API)。

[0083] 向输出生成单元205输入来自交互控制单元203的用户确定性因子数据和来自API调用单元204的API输出数据。

[0084] 基于用户确定性因子数据和API输出数据,输出生成单元205向用户生成回应文本(系统回应),并输出对应于回应文本的语音数据的回应语音。然而,在此处,当生成系统回应时,根据用户的确定性因子来控制(调整)要包括在回应文本中的信息量。

[0085] 例如,在表示用户确定性因子的值超过预定阈值的情况下,确定用户确定性因子

高,并且生成更紧凑的系统回应。这使得可以仅呈现用户期望的信息。相反,例如,在表示用户确定性因子的值等于或低于预定阈值的情况下,确定用户确定性因子低,并且生成包括更详细信息的系统回应。这使得可以在一次交互中向用户适当地呈现用户期望的信息。

[0086] 如上所述配置语音交互系统2。

[0087] 顺便提及,由语音交互系统2提供的语音交互服务包括将语音识别处理、自然语言处理等结合起来的的服务,以适当地回应或作用于例如通过口语提出的询问或请求。

[0088] 为了实现这种语音交互服务,例如,可以配置如图3所示的语音交互系统2。即,可以使语音交互系统2包括安装在客户端侧并用作语音交互服务的用户界面的语音处理设备10以及安装在诸如数据中心等服务器侧并执行用于实现语音交互功能的处理的服务器20。

[0089] 在语音交互系统2中,语音处理设备10和服务器20经由互联网30彼此耦合。

[0090] 语音处理设备10例如是可耦合到诸如家庭LAN(局域网)等网络的扬声器,并且也被称为所谓的智能扬声器等。除了音乐播放之外,这种类型的扬声器还允许在诸如照明设备、空调等装置上进行语音驱动操作。

[0091] 应当注意,语音处理设备10不限于扬声器,还可以被配置为移动装置,例如,智能电话或移动电话、平板型计算机等。

[0092] 语音处理设备10能够通过经由互联网30与服务器20链接来向用户提供语音交互服务(的用户界面)。

[0093] 即,语音处理设备10获取从用户生成的语音(用户语音),并经由互联网30将其语音数据发送到服务器20。此外,语音处理设备10经由互联网接收从服务器20发送的处理后的数据,并输出对应于处理后的数据的语音。

[0094] 服务器20是提供基于云的语音交互服务的服务器。服务器20执行语音识别处理,用于将经由互联网30从语音处理设备10发送的语音数据转换成文本数据。此外,服务器20对文本数据执行诸如对应于用户意图的交互处理之类的处理,并经由互联网30将作为处理结果获得的处理数据发送到语音处理设备10。

[0095] (语音处理设备的配置)

[0096] 图4是示出图3的语音处理设备10的配置示例的框图。

[0097] 在图4中,语音处理设备10包括处理单元51、麦克风52、扬声器53、传感器54和通信I/F 55。

[0098] 处理单元51包括例如CPU(中央处理单元)、微处理器等。处理单元51通过执行各种类型的算术处理、各个单元的操作控制等,作为语音处理设备10中的中央处理器进行操作。

[0099] 麦克风52是将来自外部的声音转换成电信号的装置(声音拾取装置)。麦克风52将通过转换获得的语音信号提供给处理单元51。

[0100] 扬声器53是将电信号转换成物理振动以产生声音的装置。扬声器53输出对应于从处理单元51提供的语音信号的声音。

[0101] 传感器54包括各种类型的传感器。传感器54执行感测并将对应于感测结果的传感器信息(传感器数据)提供给处理单元51。

[0102] 例如,作为传感器54,可以包括各种传感器,例如,执行对象成像的图像传感器、检测磁场(磁化场)的大小或方向的磁传感器、检测加速度的加速度传感器、检测角度(轮廓)、角速度或角加速度的陀螺仪传感器、检测相邻对象的接近传感器、或者检测生物信息(例

如,指纹、虹膜、脉冲等)的生物传感器。

[0103] 此外,传感器54还可以包括用于测量周围环境的传感器,例如,检测温度的温度传感器、检测湿度的湿度传感器、检测周围亮度的环境光传感器等。应当注意,传感器数据可以包括诸如从CPS(全球定位系统)信号计算的位置信息(位置数据)等信息、由计时装置计时的时间信息等。

[0104] 通信I/F 55包括例如通信接口电路等。通信I/F 55遵循来自处理单元51的控制,以获得对耦合到互联网30的服务器20的访问,并交换各种类型的数据。

[0105] 在此处,在由从语音识别单元201到配置语音交互系统2(图2)的输出生成单元205的单元提供的功能中,例如,处理单元51具有将用户语音数据输入到语音识别单元201的功能和输出对应于系统回应(的语音数据)的回应语音的功能。

[0106] 即,处理单元51处理从麦克风52提供的语音信号,并将作为处理结果获得的语音数据提供给通信I/F 55。由此,用户语音数据(语音数据)经由互联网30传输到服务器20,并输入到语音识别单元201。此外,处理单元51能够将传感器数据提供给通信I/F 55,以经由互联网30将传感器数据发送给服务器20。

[0107] 此外,处理单元51处理从通信I/F 55提供的语音数据,并将作为处理结果获得的语音信号提供给扬声器53。由此,从扬声器53输出对应于系统回应(的语音数据)的回应语音。

[0108] 应当注意,用于显示各种类型的信息(例如,字符、图像等)的显示单元、从用户接收操作的输入单元、保存各种类型的数据(例如,语音数据、文本数据等)的存储单元等可以进一步设置在语音处理设备10中,尽管这些单元没有在图4中示出。

[0109] 在此处,例如,显示单元包括例如液晶显示器、有机EL显示器等。例如,输入单元包括按钮、键盘等。此外,输入单元可以被配置为触摸传感器和显示单元集成在其中的触摸面板,并且可以通过用户的手指或触摸笔(手写笔)获得对应于操作的操作信号。存储单元包括作为非易失性存储器类型的闪存(闪存)、作为易失性存储器类型的DRAM(动态随机存取存储器)等。

[0110] (服务器的配置)

[0111] 图5是示出图3的服务器20的配置示例的框图。

[0112] 在图5中,服务器20包括处理单元71、通信I/F 72和数据库73。

[0113] 例如,处理单元71包括CPU、微处理器等。处理单元71通过执行各种类型的算术处理或对各个单元的操作的控制,作为服务器20中的中央处理器进行操作。

[0114] 例如,通信I/F 72包括通信接口电路等。根据来自处理单元71的控制,通信I/F 72与经由互联网30耦合的语音处理设备10交换各种类型的数据。

[0115] 数据库73被配置为大容量存储设备,例如,硬盘(HDD:硬盘驱动器)、半导体存储器、光盘等。

[0116] 例如,数据库73包括用于执行语音识别处理的语音识别数据库、用于执行交互处理的交互历史数据库等。应当注意,语音识别数据库或交互历史数据库是数据库的示例,并且可以包括实现语音交互服务所必需的数据库(例如,知识数据库、语音数据库等)。

[0117] 在此处,在由从语音识别单元201到配置语音交互系统2(图1)的输出生成单元205的单元提供的功能中,例如,处理单元71具有语音识别单元201、用户意图和确定性因子分

析单元202、交互控制单元203、API调用单元204和输出生成单元205的一些功能。

[0118] 即,参考数据库73中包括的语音识别数据库,处理单元71执行语音识别处理,用于将经由互联网30从语音处理设备10发送的用户语音数据(语音数据)转换成语音文本(文本数据)。

[0119] 此外,使用在语音识别处理中获得的语音文本,当执行对应于用户意图的交互处理时,处理单元71执行估计用户期望的信息的用户自身的确定性因子的处理以及根据处理结果控制(调整)包括在回应生成中的信息量的处理。由此,生成包括适当信息量的系统回应,作为对用户的回应文本,并且作为处理后的数据经由互联网30发送到语音处理设备10。

[0120] 应当注意,为了描述方便,尽管示出了在图3的语音交互系统2中提供一个语音处理设备10的情况,但是例如可以为各个用户提供多个语音处理设备10。

[0121] 此外,尽管示出了在图3的语音交互系统2中提供一个服务器20的情况,但是例如可以为各个功能(模块)提供多个服务器20。更具体地,可以提供具有与语音识别单元201相关联的语音识别模块的服务器20、具有与交互控制单元203相关联的交互控制模块的服务器20等,作为单独的服务器20。

[0122] <2.本技术的实施方式>

[0123] (1) 第一实施方式

[0124] 在下文中,给出由语音交互系统2提供的语音交互服务的具体示例的描述。在此处,阐述用于执行时间表检查的语音交互,作为示例。该示例的假设是用户已经登记了图6所示的时间表。

[0125] 在图6中,登记两个时间表,作为由标题、日期和时间、位置和目标人员指定的时间表。即,在第一记录中,2017年3月23日上午10点在主会议室的会议被登记为父亲(爸爸)的时间表。此外,在第二记录中,2017年3月23日下午5点在涩谷购物的时间表被登记为爸爸的时间表。

[0126] 应当注意,关于这些时间表的信息已经登记在服务器20的数据库73(图5)中,作为存储在时间表数据库(以下称为时间表DB)中的信息。

[0127] 此处假设用户已于2017年3月22日与语音交互系统2通话。因此,图6所示的时间表被登记为用户和语音交互系统2的明天的时间表。

[0128] 还假设,在语音交互系统2中,用于执行时间表检查的API(应用编程接口)具有例如如下所列的四个参数。

[0129] 时间表API(pram1、pram2、pram3、pram4)

[0130] 第一参数:标题

[0131] 第二参数:日期和时间

[0132] 第三参数:位置

[0133] 第四参数:目标人员

[0134] 在下面的具体示例中,用户执行与具有这样的时间表DB和时间表API的语音交互系统2的交互,目的是执行已经登记的时间表的检查。然而,在以下具体示例中,为了比较,还适当地给出了通用语音交互系统1(图1)的描述。

[0135] (1-1) 用户确定性因子高的情况

[0136] 在此处,假设这样一种情况,尽管用户知道明天上午10点有会议,作为他自己的时

间表,但是用户只是希望检查时间表是否正确。在这种情况下,假设用户已经说过,例如,“我明天上午10点有会议安排,不是吗?”

[0137] 此时,在通用语音交互系统1(图1)中,估计用户的意图是“时间表检查”,其参数是“标题=“会议””和“日期和时间=“2017年3月23日上午10点””。参数传递给执行时间表检查的时间表API。

[0138] 在执行该时间表API的情况下,作为API输出,可以获得在时间表DB中登记的时间表中存在一条记录,包括“标题=“会议””、“日期和时间=“2017年3月23日上午10点””和“地点=“主会议室””。

[0139] 然后,作为系统回应,例如,获得“明天上午10点有会议安排。地点是主会议室”。

[0140] 然而,此时,用户自己知道用户明天上午10点有会议安排,只是希望检查该信息是否正确。然而,此处获得的系统回应模仿了用户已经知道的信息,并且被用户不必要的许多信息占据。

[0141] 因此,尽管原本只需要简单回应的通信就足够了,但是用户必须收听具有大量信息的回应。这导致了交互速度的降低,即,通信的顺畅。

[0142] 相反,在应用本技术的语音交互系统2(图2)中,基于从用户语音获得的信息,估计用户自己对用户期望的信息(期望信息)的确定程度(确定性因子),并且控制(调整)回应中的信息量。

[0143] 在前面提到的示例中,因为语音“我明天上午10点有一个会议安排,不是吗?”明确表示“标题=“会议””和“日期和时间=“2017年3月23日上午10点””,该信息没有包括表示歧义的词,例如,“也许”或“我想”等,所以可以估计标题的确定性因子以及该时间表的日期和时间很高。

[0144] 因此,例如,通过省略用户已经知道的信息(具有高确定性因子的信息)并使系统回应紧凑(例如,“是”),可以仅提供用户期望的信息,这允许以良好的速度继续交互。

[0145] 应当注意,在上述示例中,尽管语音不包括指示位置的内容,但是位置信息可能不包括在系统回应中,因为关于标题以及日期和时间的确定性因子高,并且该时间表是用户自己的时间表,并且可以估计用户已经知道该位置。

[0146] 例如,在用户确定性因子高的情况下,由图2的语音交互系统2执行的处理的内容如下。

[0147] 即,用户语音数据的声音是“我明天上午10点有一个会议安排,不是吗?”,由语音处理设备10(图3)获取,并输入到语音识别单元201。语音识别单元201执行语音识别处理,并将用户语音数据转换成语音文本数据。

[0148] 用户意图分析部211对如此获得的语音文本数据执行分析处理,并且由此获得用户意图的候选。此外,用户确定性因子估计部212对语音文本数据执行分析处理,从而获得用户的确定性因子。此处估计时间表标题或日期和时间的确定性因子高,因为准确地指示时间表标题或数据和时间,并且不包括表示信息歧义的单词。

[0149] 交互控制单元203基于用户意图候选数据、交互历史数据库等来估计当前交互状态,并确定要调用的API和API的参数。在此处,确定了根据用户语音数据执行时间表检查的时间表API以及“标题=“会议””和“日期和时间=“2017年3月23日上午10点””,作为时间表API的参数。因此,API调用单元204将其参数(标题、日期和时间)传递给要调用的时间表

API,并执行该API。

[0150] 尽管在输出生成单元205中,作为API输出,获得了在时间表DB中登记的时间表中存在一个相关记录,但是生成了更紧凑的系统回应,因为估计时间表标题或日期和时间的确定性因子与用户的确定性因子一样高。

[0151] 在此处,如图7所示,例如,在用户的确定性因子和系统回应中的信息量之间存在关系。在图7中,函数f1表示用户确定性因子和回应中的信息量之间的关系,其中,横轴表示用户确定性因子,纵轴表示系统回应中的信息量(回应信息量)。

[0152] 在图7中,函数f1具有反比例关系,其中,用户确定性因子越高,回应信息量越小,而用户确定性因子越低,回应信息量越大。此处假设,例如,阈值(Th1和Th2)被设置为使得等于或高于75的分数表示高确定性因子,小于25的分数表示低确定性因子,并且等于或高于25且小于75的分数表示中等确定性因子。

[0153] 此时,例如,在计算确定性因子的分数为85的情况下,估计时间表标题或日期和时间的确定性因子高,并且作为紧凑系统回应,输出对具有高确定性因子的信息的回应,例如,“是”。

[0154] (1-2) 用户确定性因子中等的情况

[0155] 此处假设,尽管用户知道用户明天有某个安排,但是用户不确定日期和时间。在这种情况下,例如,假设用户说,“我认为是明天。会议安排好了吗?”。

[0156] 此时,在通用语音交互系统1(图1)中,估计用户的意图是“时间表检查”,其参数是“标题=“会议””和“日期和时间=“2017年3月23日上午10点””,并且参数被传递到时间表API。

[0157] 在执行该时间表API的情况下,作为API输出,获得了在时间表DB中登记的时间表中存在一个相关记录,包括“标题=“会议””、“日期和时间=“2017年3月23日上午10点””以及“地点=“主会议室””。

[0158] 然后,作为系统回应,例如,获得“明天上午10点有会议安排。地点是主会议室”。要注意的是,在通用语音交互系统1(图1)中,该系统回应的内容类似于在前面描述的另一示例中的对语音的回应的内容:“我明天上午10点有会议安排,不是吗?”。

[0159] 然而,此时,用户自己只需能够确认会议是否安排在明天,而无需寻求所有详细信息。因此,在通用语音交互系统1(图1)中获得的系统回应中,该回应信息过载,这减慢了通信的速度。

[0160] 相反,在应用本技术的语音交互系统2(图2)中,可以从用户语音数据“我认为是明天”中估计日期和时间的确定性因子是中等的,因为日期和时间是“明天”,但是不包括时间信息,或者因为日期和时间被措辞“我认为”修饰。此外,至于标题,用户清楚地说了“会议”,因此可以估计确定性因子高。

[0161] 因此,通过省略具有高确定性因子的标题,并包括更多关于中等确定性因子的日期和时间的细节,可以生成简单的系统回应,例如,“是的,会议将在明天上午10点举行”。

[0162] 在用户确定性因子为中等的这种情况下,例如,由图2的语音交互系统2执行的处理的内容如下。

[0163] 即,用户确定性因子估计部212对语音文本数据“我认为是明天,会议安排好了吗?”执行分析处理,从而获得用户的确定性因子。此处估计确定性因子是中等的,因为不包

括关于预定日期和时间的信息,或者语音被表示歧义的词修饰。此外,估计时间表标题的确定性因子高,因为准确地指示时间表标题。

[0164] 然后,在输出生成单元205中,作为API输出,获得在时间表DB中登记的时间表中存在一个相关记录。然而,作为用户的确定性因子,尽管时间表标题的确定性因子高,但是估计预定日期和时间的确定性因子是中等的。因此,生成包括关于具有中等确定性因子的日期和时间的详细信息的简单系统回应。

[0165] 在此处,在用户的确定性因子和系统回应中的信息量之间存在关系的情况下,例如,图7所示的函数f1,例如,当确定性因子的分数为70时,估计预定日期和时间的确定性因子是中等的。然后,作为简单的系统回应,输出回应,例如,“是的,会议将在明天上午10点举行”,包括关于具有中等确定性因子的预定日期和时间的详细信息。

[0166] (1-3) 用户确定性因子低的情况

[0167] 此处假设用户不知道用户明天是否有安排。在这种情况下,例如,假设用户已经说过,“我明天有什么安排?”。

[0168] 此时,在通用语音交互系统1(图1)中,用户的意图是“时间表检查”,并且估计其参数是“日期和时间=“2017年3月23日””,并且该参数传递到时间表API。

[0169] 在执行该时间表API的情况下,作为API输出,可以获得在时间表DB中登记的时间表中存在两个记录,包括“标题=“会议””、“日期和时间=“2017年3月23日上午10点””以及“地点=“主会议室””的记录和包括“标题=“购物””、“日期和时间=“2017年3月23日下午5点””以及“地点=“涩谷””的记录。

[0170] 然后,作为系统回应,例如,获得“明天上午10点在主会议室有会议安排,下午5点在涩谷有购物安排”。

[0171] 相反,在应用本技术的语音交互系统2(图2)中,估计标题的确定性因子低,因为关于时间表标题的信息不包括在用户的语音中,并且询问“什么”,而不包括具体的名称。此外,关于日期和时间,可以估计确定性因子是中等的,因为没有指定时间,尽管指定了“明天”。

[0172] 因此,可以生成系统回应,包括关于标题以及日期和时间的详细信息,例如,“明天上午10点有会议安排,下午5点有购物安排”。以这种方式,在确定性因子低的情况下发布更多信息,允许用户在一次交互中适当地获得用户期望的信息,并且使得不需要添加额外交互。因此,可以进行速度良好的交互。

[0173] 例如,在用户确定性因子低的情况下,由图2的语音交互系统2执行的处理的内容如下。

[0174] 即,用户确定性因子估计部212对语音文本数据“我明天有什么安排?”执行分析处理,从而获得用户的确定性因子。此处估计,由于未包括具体的名称,所以关于时间表标题的确定性因子低。此外,估计预定日期和时间的确定性因子是中等的,因为不包括时间信息。

[0175] 然后,在输出生成单元205中,作为API输出,获得在时间表DB中登记的时间表中存在两个相关记录。然而,作为用户的确定性因子,估计时间表标题的确定性因子低,而预定日期和时间的确定性因子中等。因此,生成包括关于具有低确定性因子的标题的更详细信息的系统回应。

[0176] 此处估计,在用户的确定性因子和系统回应中的信息量之间存在关系的情况下,例如,图7所示的函数f1,例如,当计算确定性因子的分数为17时,估计时间表标题的确定性因子低。然后,作为详细的系统回应,输出回应,例如,“明天上午10点有会议安排,下午5点有购物安排”,包括关于具有低确定性因子的时间表标题的更详细信息。

[0177] (1-4) 虽然用户的确定性因子似乎很高,但事实上并不正确的情况

[0178] 此处假设用户确定的内容有错误的情况。在这种情况下,例如,假设用户说,“我后天上午10点有一个会议安排,不是吗?”。

[0179] 此时,在通用语音交互系统1(图1)中,估计用户的意图是“时间表检查”,并且其参数是“标题=“会议””和“日期和时间=“2017年3月24日上午10点””,并且这些参数传递到时间表API。

[0180] 在执行该时间表API的情况下,作为API输出,可以获得零时间表,因为在时间表DB中登记的时间表中不存在包含此类内容的记录。然后,作为系统回应,例如,获得“没有这样的时间表”。

[0181] 相反,在应用本技术的语音交互系统2(图2)中,可以估计时间表标题以及日期和时间的确定性因子高,因为类似于(1-1)“用户确定性因子高的情况”,清楚地指示“标题=“会议””和“日期和时间=“2017年3月24日上午10点””,或者信息中不包括表示歧义的单词,例如“也许”或“我认为”。

[0182] 此外,在此处,可以估计日期和时间的用户确定性因子是不正确的,因为通过执行具有指定时间表标题以及日期和时间(作为参数)的时间表API而获得的API输出是零时间表。

[0183] 因此,可以添加正确的信息并生成系统回应,例如,“不,会议在明天上午10点举行”。这使得用户可以在一次交互中获得用户期望的信息,而不用说话,来再次确认正确的信息。

[0184] (1-5) 实际上不正确的情况下有多个选择时

[0185] 在上述(1-4)中,示出了生成对语音“我后天上午10点有会议安排,不是吗?”的系统回应“不,会议在明天上午10点举行”的示例,然而,假设难以添加正确的信息,因为对于一些用户来说,在时间表DB中登记多个会议安排。

[0186] 在这种情况下,例如,用户侧的语音处理设备10能够启动日历应用,以每月或每周呈现多个登记的会议安排。此时,语音处理设备10可以仅呈现语音中包括的接近“后天”的时间表,而不是呈现所有登记的时间表。

[0187] 这允许用户确认日历的呈现内容,并识别用户的语音内容不正确,例如,用户将时间表误认为另一会议安排,或者用户忘记了会议安排在后天上午10点的登记。此时,可以了解用户的倾向(例如,某个用户经常忘记时间表的登记等),并且可以在随后日历的呈现内容中考虑该倾向。

[0188] 应当注意,尽管示出了在上述(1-1)确定性因子高的情况下或者在(1-2)确定性因子中等的情况下输出更紧凑的系统回应的情况,但是只要执行良好速度的交互,系统回应中的信息量可以增加。例如,作为前述(1-1)的系统回应,除了“是的”,还可以通过添加回应的后一部分来输出更详细的系统回应,例如,“是的,明天上午10点将有会议,”。

[0189] (2) 第二实施方式

[0190] 在上述第一实施方式中,给出了时间表目标人员和通过交互确认时间表的用户是同一个人的情况的描述。然而,在下文中,作为第二实施方式,给出了这些人不同的情况的描述。

[0191] (本技术的语音交互系统的配置)

[0192] 图8是示出应用本技术的语音交互系统的另一配置示例的框图。

[0193] 在图8中,语音交互系统3包括语音识别单元301、用户意图和确定性因子分析单元302、交互控制单元303、API调用单元304、输出生成单元305和说话者识别单元306。

[0194] 在图8中,从语音识别单元301到输出生成单元305的单元类似于图2所示的从语音识别单元201到输出生成单元205的单元来配置。即,与图2的语音交互系统2相比,图8的语音交互系统3的不同之处在于,说话者识别单元306添加到用户意图和确定性因子分析单元302的前一级。

[0195] 说话者识别单元306通过参考说话者识别数据库并分析输入到其中的用户语音数据来识别说话者。说话者识别单元306识别分配给被识别的说话者的用户ID,并将用户ID提供给用户意图和确定性因子分析单元302。

[0196] 应当注意,在说话者识别数据库中,例如,对于每个说话者,预先登记将语音数据与用户ID相关联的信息。此外,在此处,尽管例示了说话者识别单元306基于用户语音数据识别说话者的情况,但是用于识别说话者的数据不限于用户语音数据,并且例如,也可以使用受传感器54成像的对象的图像数据等。

[0197] 用户意图和确定性因子分析单元302将用户ID、用户意图候选数据和用户确定性因子数据提供给交互控制单元303。交互控制单元303基于用户意图候选数据、交互历史数据库等来估计当前交互状态并确定要调用的API及其参数。交互控制单元303将确定的API和参数数据提供给API调用单元304。

[0198] 此外,交互控制单元303将用户确定性因子数据提供给输出生成单元305。在此处,例如,通过使用用户ID,识别出已经说话的用户不同于时间表目标人员。因此,交互控制单元303将另一个人的时间表的确定性因子数据作为用户确定性因子数据传递给输出生成单元305。

[0199] 输出生成单元305基于用户确定性因子数据和API输出数据生成对用户的回应文本(系统回应),并输出对应于语音数据的回应语音。在此处,尽管根据另一个人的时间表的确定性因子来控制(调整)系统回应中的信息量,但是与自己的时间表的情况相比,可以增加回应中的信息量,因为通过交互来检查时间表的人是不同的。

[0200] 在下文中,给出由语音交互系统3提供的语音交互服务的具体示例的描述。此处假设某个家族共享并使用单个调度器,并且图9中列出的时间表已经被用户登记在时间表DB中。交互还假设父亲(爸爸)确认母亲(妈妈)的时间表。

[0201] 在图9中,五个时间表被登记为由标题、日期和时间、位置和目标人员识别的时间表。即,2017年3月23日上午10点在主会议室举行的会议在第一记录中被登记为爸爸的时间表。此外,2017年3月23日上午10点在新宿购物在第二记录中被登记为妈妈的时间表。

[0202] 此外,2017年3月23日下午3点,Yoyogi的一名牙医在第三记录中被登记为爸爸的时间表。此外,2017年3月23日下午5点在涩谷购物在第四记录中被登记为爸爸的时间表。此外,2017年3月23日下午5点在原宿举行的校友聚会在第五记录中被登记为妈妈的时间表。

[0203] 在以下具体示例中,假设用户在具有这样的时间表DB和时间表API的语音交互系统3上执行交互,其意图是确认已经登记的时间表。然而,在下面的具体示例中,为了比较,还适当地给出了通用语音交互系统1(图1)的描述。

[0204] (2-1) 用户自己对另一个人的时间表的确定性因子高的情况

[0205] 此处假设,尽管用户(爸爸)知道妈妈的校友聚会安排在明天下午5点,但是用户(爸爸)希望确认是否正确。在这种情况下,例如,假设用户(爸爸)说,“妈妈在明天下午5点有校友聚会安排,不是吗?”。

[0206] 此时,在通用语音交互系统1(图1)中,估计用户的意图是“时间表检查”,并且其参数是“标题=“校友聚会””、“日期和时间=2017年3月23日下午5点”和“目标人员=“妈妈””。这些参数传递给执行时间表检查的时间表API。

[0207] 在执行该时间表API的情况下,作为API输出,获得在时间表DB中登记的时间表中存在一条记录,包括“标题=“校友聚会””、“日期和时间=2017年3月23日下午5点”、“目标人员=“妈妈””和“位置=“原宿””。

[0208] 然后,获得系统回应,例如,“明天下午5点有原宿校友聚会安排”。

[0209] 相反,在应用本技术的语音交互系统3(图8)中,基于从用户语音获得的信息,估计用户对用户期望的信息(期望信息)的确定程度(确定性因子),并且根据估计结果控制(调整)回应中的信息量。

[0210] 在上述示例中,因为用户语音清楚地指示时间表标题、日期和时间以及目标人员,或者限定他们的词语不包括表示歧义的词语,例如,“也许”或“我认为”,所以估计其确定性因子高。

[0211] 因此,生成和输出紧凑的系统回应(例如,“是”)使得可以进行良好速度的交互。

[0212] 然而,因为没有关于时间表位置的讲话,所以不可能确定关于该位置的确定性因子高。在这个示例中,特别地,因为说话的用户(爸爸)不同于时间表的目标人员(妈妈),所以认为很难假设用户(爸爸)预先知道位置。

[0213] 因此,例如,最好在系统回应中添加位置信息,例如,“是的,有安排。地点是原宿”。

[0214] 例如,在用户自己关于另一个人的时间表的确定性因子高的情况下,由图8的语音交互系统3执行的处理内容如下。

[0215] 即,用户确定性因子估计部312对语音文本数据“妈妈明天下午5点有校友聚会安排,不是吗?”执行分析处理,并且由此获得关于另一个人的时间表的确定性因子。此处,估计时间表标题、日期和时间以及目标人员的确定性因子高,因为清楚地指示时间表标题、日期和时间以及目标人员或者不包括表示歧义的单词。

[0216] 然后,在输出生成单元305中,尽管作为API输出,获得了在时间表DB中登记的时间表中存在一个相关记录,但是生成更紧凑的系统回应,因为作为关于另一个人的时间表的确定性因子,估计关于时间表标题、日期和时间以及目标人员的确定性因子高。

[0217] 在此处,例如,如图10所示,例如,在用户的确定性因子和系统回应中的信息量之间存在关系。类似于图7,尽管图10示出了与作为横轴的用户确定性因子和作为纵轴的系统回应中的信息量的关系,但是在自己时间表的情况下的关系由函数f1表示,而在其他人时间表的情况下的关系由函数f2表示。

[0218] 即,在图10中,函数f1和函数f2具有相同的倾斜度。函数f1和函数f2的共同之处在

于函数f1和函数f2具有反比例关系,即尽管用户确定性因子越低,回应信息量越大,但是用户确定性因子越高,回应信息量越小。然而,截距不同,并且如果确定性因子相同,函数f2具有比函数f1的回应信息量更大的回应信息量。

[0219] 换言之,另一个人的时间表并不比自己的时间表更容易理解。因此,即使在相同的确定性因子下,回应信息量也会在自己的时间表和另一个人的时间表之间发生变化。

[0220] 此处假设阈值(Th1和Th2)被设置为使得等于或高于75的分数表示高确定性因子,小于25的分数表示低确定性因子,并且等于或高于25且小于75的分数表示中等确定性因子。

[0221] 此时,例如,当计算确定性因子为90的分数时,估计关于时间表标题、日期和时间以及目标人员的确定性因子高,并且可以输出诸如“是”等回应,作为紧凑系统回应。然而,例如,因为涉及另一个人的时间表,所以期望通过添加位置信息等来增加回应信息量,例如,作为“是的,有安排。地点是原宿”。

[0222] (2-2) 用户自己对另一个人的时间表的确定性因子中等的情况

[0223] 此处假设一种情况,尽管用户(爸爸)知道妈妈明天有某个安排,但是用户(爸爸)不确定时间表的日期和时间或者内容。在这种情况下,例如,假设用户(爸爸)说,“我认为妈妈明天去购物,不是吗?”。

[0224] 此时,在通用语音交互系统1(图1)中,估计用户的意图是“时间表检查”,其参数是“标题=“购物””、“日期和时间=“2017年3月23日””和“目标人员=“妈妈””。这些参数传递给时间表API。

[0225] 在执行该时间表API的情况下,作为API输出,可以获得在时间表API中登记的时间表中存在一条记录,包括“标题=“购物””、“日期和时间=“2017年3月23日上午10点””、“位置=“新宿”和“目标人员=“妈妈””。

[0226] 然后,获得系统回应,例如“明天上午10点有去新宿购物的安排”。

[0227] 相反,在应用本技术的语音交互系统3(图8)中,可以估计关于日期和时间的确定性因子低,因为用户的语音不包括时间信息,此外,“明天”被“我认为”修饰。也可以估计关于标题的确定性因子中等,因为用户的语音包括“妈妈明天去购物,不是吗?”。因为没有关于该位置的语音,所以也可以估计关于该位置的确定性因子低。

[0228] 因此,可以返回包含具有低确定性因子的信息的回应,例如,“明天上午10点有去新宿购物的安排”。这允许用户立即获得用户期望的信息,并实现顺利通信。

[0229] 在用户自己关于另一个人的时间表的确定性因子是中等的这种情况下,例如,要由图8的语音交互系统3执行的处理的内容如下。

[0230] 即,用户确定性因子估计部312对语音文本数据“我认为妈妈明天去购物,不是吗?”执行分析处理,并且由此获得关于另一个人的时间表的确定性因子。此处估计关于日期和时间的确定性因子低,因为时间表的日期和时间不包括时间信息,被表示歧义的词语等修改。此外,由于包括在用户语音中的表达,估计关于标题的确定性因子中等。

[0231] 然后,在输出生成单元305中,尽管作为API输出,获得了在时间表DB中登记的时间表中存在相关的一条记录,但是生成了包括关于具有中等确定性因子的标题和具有低确定性因子的日期和时间的详细信息的系统回应,因为估计作为关于另一个人的时间表的确定性因子,关于时间表的日期和时间的确定性因子低,而关于时间表标题的确定性因子中等。

[0232] 在此处,在关于另一个人的时间表的用户确定性因子和系统回应中的信息量之间存在关系的情况下,例如,图10所示的函数f2,当计算确定性因子的分数为60时,估计关于时间表标题的确定性因子中等,并且输出诸如“明天上午10点有去新宿购物的安排”等回应,该回应包括关于具有中等确定性因子的时间表标题的详细信息。然而,在此处,因为时间表是另一个人的,所以通过添加位置信息“新宿”,来增加回应信息量。

[0233] (2-3) 用户自己对另一个人的时间表的确定性因子低的情况

[0234] 此处假设用户(爸爸)不知道妈妈的时间表。在这种情况下,例如,假设用户已经说过,“妈妈明天有什么安排吗?”。

[0235] 此时,在通用语音交互系统1(图1)中,估计用户的意图是“时间表检查”,其参数是“日期和时间=“2017年3月23日””和“目标人员=“妈妈””。些参数传递给时间表API。

[0236] 在执行该时间表API的情况下,作为API输出,获得在时间表DB中登记的时间表中存在两个记录,一个记录包括包括“标题=“购物””、“日期和时间=2017年3月23日上午10点”、“位置=“原宿””和“目标人员=“妈妈””,另一记录包括“标题=“校友聚会””、“日期和时间=2017年3月23日下午5点”、“位置=“原宿””和“目标人员=“妈妈””。

[0237] 然后,获得系统回应,例如“明天上午10点在新宿有购物安排,下午5点在原宿有校友聚会安排”。

[0238] 相反,在应用本技术的语音交互系统3(图8)中,可以估计关于时间表标题、日期和时间以及位置的确定性因子低。

[0239] 因此,例如,像“明天上午10点在新宿有购物安排,下午5点在原宿有校友聚会安排”一样,通过以用户的低确定性因子向信息呈现特定信息,用户可以立即获得用户期望的信息,并且可以实现顺利通信。

[0240] 例如,在关于另一个人的时间表的用户确定性因子低的情况下,由图8中的语音交互系统3执行的处理的内容如下。

[0241] 即,用户确定性因子估计部312对语音文本数据“妈妈明天有安排吗?”执行分析处理,由此获得关于另一个人的时间表的用户确定性因子。此处假设关于时间表标题、日期和时间以及位置的确定性因子低。

[0242] 然后,在输出生成单元305中,尽管作为API输出,获得在时间表DB中登记的时间表中存在相关的两个记录,但是生成包括更详细信息的系统回应,因为估计作为关于另一个人的时间表的用户确定性因子,关于时间表标题、日期和时间以及位置的确定性因子低。

[0243] 在此处,例如,在关于另一个人的时间表的用户确定性因子和系统回应中的信息量之间存在关系的情况下,例如,图10所示的函数f2,当计算确定性因子的分数为21时,估计时间表标题、日期和时间以及位置的确定性因子低。然后,输出包括更详细信息的回应,例如,“明天上午10点在新宿有购物安排,下午5点在原宿有校友聚会安排”。

[0244] 应当注意的是,在此处引用的两个示例(情况(2-2)和(2-3))中,由应用本技术的通用语音交互系统1(图1)和语音交互系统3(图8)做出的系统回应没有很大差异。其中一个原因是,作为另一个人的时间表,不允许获取关于位置信息的隐含确定性,这不可避免地降低了用户确定性因子的等级,并且均衡了最初输出更多信息的一般方法输出的信息量和根据用户确定性因子调整信息量的本技术方法输出的信息量。

[0245] 应当注意,在另一个人的时间表的这种情况下,用户确定性因子的等级变得低于

在自己的时间表的情况下的确定性因子。因此,通过图10中示出的函数f1和函数f2的关系也很明显,回应信息量增加了。

[0246] (3) 第三实施方式

[0247] 如上所述,关于估计确定性因子的方法,可以使用诸如目标信息是否清楚地包括在用户的语音中、是否执行了用表示歧义的单词修饰、或者该信息是否是已经说话的用户的信息之类的指标。

[0248] 相反,通过基于关系情况的学习,可以获得用户意图和确定性因子之间的关系。即,可以利用例如其他语音信息、手势、视线、面部表情等以及语音文本。

[0249] (本技术的语音交互系统的配置)

[0250] 图11是示出应用本技术的语音交互系统的另一配置示例的框图。

[0251] 在图11中,语音交互系统4包括语音识别单元401、用户意图和确定性因子分析单元402、交互控制单元403、API调用单元404、输出生成单元405、说话者识别单元406、手势识别单元407、视线识别单元408、面部取向识别单元409和用户位置识别单元410。

[0252] 在图11中,从语音识别单元401到输出生成单元405的单元与图2所示的从语音识别单元201到输出生成单元205的单元类似地配置。此外,说话者识别单元406类似于图8所示的说话者识别单元306配置。

[0253] 即,与图2的语音交互系统2和图8的语音交互系统3相比,图11的语音交互系统4的不同之处在于,手势识别单元407、视线识别单元408、面部取向识别单元409和用户位置识别单元410被添加到用户意图和确定性因子分析单元402的前一级。

[0254] 应当注意,例如,向手势识别单元407、视线识别单元408、面部取向识别单元409和用户位置识别单元410输入由设置在语音处理设备10(图4)中的传感器54检测到的传感器信息(传感器数据)。

[0255] 手势识别单元407通过使用向其输入的传感器数据执行手势识别处理来识别用户的手势,并将识别结果提供给用户意图和确定性因子分析单元402。此处使用的传感器数据例如是从各种传感器获得的信息,例如,专用于检测用户手势的传感器或者用于获取包括用户作为对象的图像数据的图像传感器。

[0256] 视线识别单元408通过使用输入到视线识别单元408的传感器数据执行视线识别处理来识别用户的视线,并将识别结果提供给用户意图和确定性因子分析单元402。此处使用的传感器数据例如是从各种传感器获得的信息,例如,专用于检测用户视线的传感器或者用于获取包括用户作为对象的图像数据的图像传感器。

[0257] 面部取向识别单元409通过使用输入到其中的传感器数据执行面部取向识别处理来识别用户面部的取向,并将识别结果提供给用户意图和确定性因子分析单元402。此处使用的传感器数据例如是从各种传感器获得的信息,例如,专用于检测用户面部方向的传感器或者用于获取包括用户作为对象的图像数据的图像传感器。

[0258] 用户位置识别单元410通过使用输入到其中的传感器位置信息执行位置识别处理来识别用户位置,并将识别结果提供给用户意图和确定性因子分析单元402。此处使用的传感器数据例如是从各种传感器获得的信息,例如,专用于检测用户位置的传感器(例如,GPS模型等)或者用于获取包括允许识别用户位置的对象的图像数据的图像传感器。

[0259] 来自手势识别单元407、视线识别单元408、面部取向识别单元409和用户位置识别

单元410的识别结果与来自语音识别单元401的语音文本数据和来自说话者识别单元406的用户ID一起输入到用户意图和确定性因子分析单元402。用户意图和确定性因子分析单元402包括用户意图和分析部分411和用户确定性因子估计部412。

[0260] 当对语音文本数据执行分析处理,以估计用户确定性因子时,用户确定性因子估计部412基于从传感器数据获得的识别结果来考虑用户的行为,以使得能够获得用户确定性因子。在此处,例如,在下面列出的趋势被视为用户行为的情况下,假设用户行为对应于用户确定性因子应当降低的情况。

[0261] (a) 说话速度变慢。

[0262] (b) 说话的声音变高。

[0263] (c) 视线未确定。

[0264] (d) 用户直接靠近语音交互系统(语音处理设备10)。

[0265] (e) 用户将他/她的脸转向语音交互系统(语音处理设备10)的方向。

[0266] (f) 双臂的手势动作变大。

[0267] (g) 用户正在打哈欠。

[0268] 在例如上述从(a)到(g)的趋势被视为用户的行为的这种情况下,添加估计其确定性因子低的算法使得可以提高用户确定性因子的估计的准确性。此外,可以基于通过依次记录关于用户行为的行为数据并通过使用该行为数据(的历史)来执行学习而获得的趋势,来估计确定性因子。

[0269] 应当注意,用户的行为可以是任何行为,只要基于从传感器54获得的传感器数据来识别该行为。例如,除了关于温度传感器检测到的温度或湿度传感器检测到的湿度的信息之外,可以通过利用由计时装置计时的时间信息等来识别用户的行为。例如,在使用时间信息并且用户在清晨打哈欠的情况下,可以确定所述确定性因子低,诸如此类。

[0270] 如上所述,已经对第一实施方式至第三实施方式进行了描述。使用这些实施方式中阐述的方法,允许用户以适当的量获得用户期望的信息,从而使得可以实现良好速度的顺利通信。

[0271] 应当注意,上述PTL 1公开了一种方法,其中,系统根据系统分析用户的语音意图时的结果的系统本身的确定性因子切换回应,作为根据系统侧的确定性因子切换回应的系统,作为语音交互系统。然而,在这种方法中,没有机制来估计用户的确定程度。因此,即使在用户确定性因子低的情况下或者在确定性因子高的情况下,当系统本身的确定性因子相等时,也生成具有相同信息量的回应。这不能像前述实施方式中阐述的方法那样实现良好速度的顺利通信。

[0272] <3.变型例>

[0273] 在上述描述中,已经给出了服务器20的处理单元71的描述,在由从语音识别单元201到输出生成单元205的单元提供的功能中,该处理单元71具有语音识别单元201、用户意图和确定性因子分析单元202、交互控制单元203、API调用单元204和输出生成单元205的一些功能。然而,处理单元71只需要具有由从语音识别单元201到输出生成单元205的单元提供的功能的至少一些功能。

[0274] 此外,已经给出了语音处理设备10的处理单元51的描述,在由包括的从语音识别单元201到输出生成单元205的单元提供的功能中,处理单元51具有将用户语音数据输入到

语音识别单元201的功能和输出对应于系统回应(的语音数据)的回应语音的功能。然而,处理单元51只需要具有由从语音识别单元201到输出生成单元205的单元提供的功能的至少一些功能。

[0275] 即,在由从语音识别单元201到输出生成单元205的单元提供的功能中,除了服务器20的处理单元71具有的功能之外的功能是语音处理设备10的处理单元51具有的功能。

[0276] 然而,语音处理设备10的处理单元51可以具有由从语音识别单元201到输出生成单元205的单元提供的所有功能,并且语音交互系统的功能可以仅由语音处理设备10实现。然后,数据库73可以构建在语音处理设备10中,或者可以由服务器20通过互联网30提供。

[0277] 此外,尽管在上述描述中,已经给出了图2的输出生成单元205(图8的输出生成单元305和图11的输出生成单元405)使用用户的确定性因子来控制(调整)回应中的信息量的情况的描述,但是另一模块可以使用用户的确定性因子来执行回应中信息量的控制。

[0278] 现在,参考图12至图14,接下来给出应用本技术的语音交互系统的配置的变型例的描述。应当注意,为了比较,还给出了通用文本聊天系统的描述。

[0279] (通用文本聊天系统的配置)

[0280] 图12是示出通用文本聊天系统的配置的框图。

[0281] 在图12中,通用文本聊天系统6包括用户意图分析单元602、交互控制单元603、API调用单元604和输出生成单元605。

[0282] 通过文本聊天向用户意图分析单元602提供用户输入的文本数据。用户意图分析单元602处理文本数据,以分析用户意图,并将作为分析结果获得的用户意图候选数据提供给交互控制单元603。

[0283] 交互控制单元603基于从用户意图分析单元602等提供的用户意图候选数据来估计当前交互状态,并确定要调用的API及其参数。交互控制单元603将API和参数数据提供给API调用单元604。

[0284] 基于从交互控制单元603提供的API和参数数据,API调用单元604将参数传递给要调用的API,执行API,并将作为执行结果获得的API输出提供给输出生成单元605。

[0285] 输出生成单元605基于从API调用单元604提供的API输出,输出对应于文本数据或图像数据的信息,作为系统回应。

[0286] 通用文本聊天系统6如上所述配置。在此处,在本技术的语音交互系统(例如,图2的语音交互系统2)中,向用户意图和确定性因子分析单元202输入文本聊天的文本数据,而不是语音文本数据,并且也可以执行与语音文本数据的处理类似的处理。在这种情况下,输出生成单元205输出对应于文本数据或图像数据的信息,作为系统回应。应当注意,此时,在图2的语音交互系统2中,语音识别单元201是不必要的。

[0287] (本技术的语音交互系统的配置的第一变型例)

[0288] 图13是示出应用本技术的语音交互系统的第一变型例的框图。

[0289] 在图13中,语音交互系统7包括语音识别单元701、用户意图和确定性因子分析单元702、交互控制单元703、API调用单元704和输出生成单元705。

[0290] 在图13中,语音识别单元701、用户意图和确定性因子分析单元702和API调用单元704类似于图2所示的语音识别单元201、用户意图和确定性因子分析单元202和API调用单元204来配置。

[0291] 即,与图2的语音交互系统2相比,图13的语音交互系统7在交互控制单元703和输出生成单元705的功能上不同。

[0292] 具体地,当基于用户意图候选数据和交互历史数据库来估计当前交互状态并确定要调用的API及其参数时,交互控制单元703使用用户确定性因子数据来选择最佳API。然而,此时,可选择的API根据回应中的信息量而不同,并且选择可获得具有对应于用户确定性因子的信息量的回应的API。

[0293] 基于从交互控制单元703提供的API和参数数据,API调用单元704将参数传递给要调用的API并执行该API。

[0294] 输出生成单元705基于从API调用单元704提供的API输出数据生成系统回应,并输出对应于该语音数据的回应语音。然而,在交互控制单元703中,使用用户确定性因子数据选择最佳API,并且已经调整了要包括在回应文本中的信息量。因此,没有必要在此处调整信息量。

[0295] 语音交互系统7如上所述配置。

[0296] (本技术的语音交互系统的配置的第二变型例)

[0297] 图14是示出应用本技术的语音交互系统的第二变型例的框图。

[0298] 在图14中,语音交互系统8包括语音识别单元801、用户意图和确定性因子分析单元802、交互控制单元803、API调用单元804和输出生成单元805。

[0299] 在图14中,语音识别单元801和用户意图和确定性因子分析单元802类似于图2所示的语音识别单元201和用户意图和确定性因子分析单元202配置。

[0300] 即,与图2的语音交互系统2相比,图14的语音交互系统8在交互控制单元803、API调用单元804和输出生成单元805的功能上不同。

[0301] 具体地,交互控制单元803基于用户意图候选数据、交互历史数据库等来估计当前交互状态,并确定要调用的API及其参数。交互控制单元803将API和参数数据提供给API调用单元804。

[0302] 基于从交互控制单元803提供的API和参数数据,API调用单元804将参数传递给要调用的API,执行API,并将作为执行结果获得的API输出返回给交互控制单元803。

[0303] 基于用户确定性因子数据和API输出数据,交互控制单元803生成输出,对于该输出,调整要包括在回应文本中的信息量,并且将该输出提供给输出生成单元805。

[0304] 在此处,可以生成和输出对用户的系统回应本身,或者可以仅输出生成系统回应所必需的信息,并且在后一阶段,可以将系统回应的生成留给输出生成单元805。

[0305] 输出生成单元805基于来自交互控制单元803的输出,输出对应于系统回应的回应语音。然而,如前所述,可以在输出生成单元805一侧生成系统回应。

[0306] 语音交互系统8如上所述配置。

[0307] 同时,在语音交互系统2(图2)中,尽管已经描述了用户语音数据直接输入到语音识别单元201的情况,但是在处理之后,用户语音数据可以输入到语音识别单元201。然后,在下文中,参考图15,给出对用户语音数据执行预处理的配置的描述。

[0308] (本技术的语音交互系统的配置的第三变型例)

[0309] 图15是示出应用本技术的语音交互系统的第三变型例的框图。

[0310] 在图15中,语音交互系统2包括语音识别单元201、用户意图和确定性因子分析单

元202、交互控制单元203、API调用单元204、输出生成单元205和预处理单元221。

[0311] 在图15中,从语音识别单元201到输出生成单元205的单元与图2所示的从语音识别单元201到输出生成单元205的单元类似地配置。即,与图2的语音交互系统2相比,图15的语音交互系统2的不同之处在于,预处理单元221添加到语音识别单元201的前一级。

[0312] 预处理单元221例如对输入到其中的语音信号执行诸如回声消除或噪声去除等信号处理,并将作为信号处理结果获得的用户语音数据提供给语音识别单元201。

[0313] 在此处,例如,语音处理设备10具有麦克风52和扬声器53。因此,麦克风52可能从扬声器53拾取声音。因此,通过执行诸如回声消除等信号处理,可以仅获得用户语音数据。更具体地,在语音处理设备10播放音乐的情况下,作为用户语音数据,获得语音信号,作为从其中移除从扬声器53输出的音乐并且被麦克风52获取的语音信号。

[0314] 应当注意,虽然预处理单元221基本上具有语音处理设备10的处理单元51具有的功能,但是预处理单元221可以具有服务器20的处理单元71具有的功能。在这种情况下,语音处理设备10经由互联网30向服务器20发送语音信号,由此在服务器20侧执行对语音信号的信号处理。

[0315] 语音交互系统2如上所述配置。

[0316] 顺便提及,在上述描述中,如图7和图10所示,尽管已经给出在用户的确定性因子和系统回应中的信息量之间存在反比例关系的描述,但是该关系是一个示例,例如,如果存在线性或非线性关系(例如,反比关系等)就足够了。

[0317] 然后,在下文中,参考图16至图19,给出用户确定性因子和系统回应中信息量的关系的变型例的描述。应当注意,可以根据预定算法来计算这些关系,或者可以手动定义这些关系。此外,可以应用使用为每个用户累积的数据的机器学习。

[0318] (确定性因子和回应信息量的第一变型例)

[0319] 图16是示出用户的确定性因子和系统回应中信息量的关系的第一变型例的示图。

[0320] 应当注意,类似于图7和图10,图16示出了用户确定性因子和回应信息量的关系,其中,横轴表示用户确定性因子,纵轴表示系统回应中的信息量(回应信息量)。这也适用于稍后描述的图17至图19。

[0321] 在图16中,函数f3具有反比关系,其中,用户确定性因子越高,回应信息量越小,而用户确定性因子越低,回应信息量越大。

[0322] 在此处,可以设置阈值(Th1和Th2),使得例如等于或高于75的分数表示高确定性因子,小于25的分数表示低确定性因子,等于或高于25且小于75的分数表示中等确定性因子。

[0323] (确定性因子与回应信息量的关系的第二变型例)

[0324] 图17是示出用户的确定性因子和系统回应中信息量的关系的第二变型例的示图。

[0325] 在图17中,函数f4具有阶梯状关系,其中,用户确定性因子越高,回应信息量越小,而用户确定性因子越低,回应信息量越大。

[0326] 在此处,通过将例如Th1=25和Th2=75设置为阈值,等于或高于75的分数表示高确定性因子,小于25的分数表示低确定性因子,等于或高于25且小于75的分数表示中等确定性因子。然而,回应信息量是对于高确定性因子、中等确定性因子和低确定性因子中的每一个的特定信息量,因为函数f4具有阶梯状关系。

[0327] (确定性因子与回应信息量的关系的第三变型例)

[0328] 图18是示出用户的确定性因子和系统回应中信息量的关系的第三变型例的示意图。

[0329] 在图18中,函数f5具有弓形关系,其中,用户确定性因子越高,回应信息量越小,而用户确定性因子越低,回应信息量越大。

[0330] 在此处,通过将例如Th1=25和Th2=75设置为阈值,等于或高于75的分数表示高确定性因子,小于25的分数表示低确定性因子,等于或高于25且小于75的分数表示中等确定性因子。然而,对于高确定性因子、中等确定性因子和低确定性因子中的每一个,回应信息量缓慢变化,因为函数f5具有弓形关系。

[0331] (确定性因子和回应信息量的第四变型例)

[0332] 图19是示出用户的确定性因子和系统回应中的信息量之间的关系的第四变型例的示意图。

[0333] 在图19中,函数f6具有S形关系,其中,用户确定性因子越高,回应信息量越小,而用户确定性因子越低,回应信息量越大。

[0334] 在此处,通过将例如Th1=25和Th2=75设置为阈值,等于或高于75的分数表示高确定性因子,小于25的分数表示低确定性因子,等于或高于25且小于75的分数表示中等确定性因子。然而,对于高确定性因子、中等确定性因子和低确定性因子中的每一个,回应信息量缓慢变化,因为函数f6具有弓形关系。

[0335] 要注意的是,尽管图16至图19没有清楚地说明,类似于图10的关系,但是当函数f3(函数f4、f5和f6)表示在自己的时间表的情况下的关系时,另一个人的时间表的关系是由于将函数f3(函数f4、f5和f6)向右移动而产生的关系。因此,如果确定性因子相等,则对于表示另一个人的时间表的函数,回应信息量大于表示自己时间表的函数f3(函数f4、f5和f6)就足够了。

[0336] 即,另一个人的时间表不比自己的时间表更容易理解。因此,即使在相同的确定性因子下,也可以在自己的时间表和另一个人的时间表之间改变回应信息量。

[0337] (API的其他示例)

[0338] 在上述描述中,尽管作为示例,已经给出了将执行时间表检查的时间表API描述为对应于用户意图(意图)的API,但是也可以类似地应用其他API,例如,检查新闻或天气的API、播放音乐的API等。例如,可以通过由语音处理设备10或服务器20执行的应用程序调用这些API。此外,在上述描述中,通过示例,已经描述了标题、日期和时间、位置和目标人员,作为时间表API的参数(入口)。然而,某些参数可能会被删除或添加另一参数(例如,登记人)。

[0339] (确定性因子的分数的示例)

[0340] 此外,例如,可以为每个用户的语音或者用户语音中包括的每个单词计算确定性因子的分数。例如,作为用户语音中包括的单词,可以计算与诸如标题、日期和时间、位置或目标人员之类的参数对应的每个单词的确定性因子的分数。或者,例如,可以计算和计算与诸如标题或日期和时间等参数对应的每个单词的确定性因子的分数。在此处计算时,例如,可以确定所述确定性因子的分数的总值、平均值、最大值等。

[0341] <4. 语音交互处理的流程>

[0342] 在下文中,参考图20的流程图,作为语音交互处理的示例,给出由语音交互系统2

(图2)执行的语音交互处理的流程的描述。

[0343] 应当注意,在图20的流程图中示出的处理步骤中,步骤S11至S13由语音处理设备10的处理单元51执行(图4),步骤S21至S27的处理由服务器20的处理单元71执行(图5)。

[0344] 在步骤S11中,在用户进行语音的情况下,处理单元51经由麦克风52接收用户的语音。在此处,例如,接收用户的语音,例如,“我明天上午10点有会议安排,不是吗?”或者“我认为是明天。会议安排好了吗?”。

[0345] 在步骤S12中,通信I/F 55经由互联网30将在步骤S11的处理中接收的用户语音数据发送到服务器20。

[0346] 从语音处理设备10发送的用户语音数据由服务器20的通信I/F 72接收,并且在服务器20中执行步骤S21至S27的处理。

[0347] 在步骤S21中,处理单元71的语音识别单元201通过参考语音到文本转换数据库来执行语音识别处理,并将从语音处理设备10接收的用户语音数据转换成语音文本数据。

[0348] 在步骤S22中,处理单元71的用户意图分析部211基于在步骤S21的处理中获得的语音文本数据执行用户意图分析处理。

[0349] 在步骤S23中,处理单元71的用户确定性因子估计部212基于在步骤S21的处理中获得的语音文本数据执行用户确定性因子估计处理。

[0350] 在步骤S24中,处理单元71的交互控制单元203基于在步骤S22的处理中获得的用户意图候选数据和交互历史数据库等来估计当前交互状态,并确定要调用的API及其参数。

[0351] 在步骤S25中,基于在步骤S24的处理中获得的API和参数数据,处理单元71的API调用单元204将参数传递给要调用的API(例如,时间表API)并执行该API。

[0352] 在步骤S26中,基于在步骤S23的处理中获得的用户确定性因子数据和在步骤S25的处理中获得的API输出,处理单元71的输出生成单元205生成系统回应,根据用户确定性因子来控制(调整)包括在回应文本中的信息量。

[0353] 在步骤S27中,通信I/F 72经由互联网30将在步骤S26的处理中获得的系统回应发送到语音处理设备10。

[0354] 从服务器20发送的经处理的数据(系统回应的数据)由语音处理设备10的通信I/F 55接收,并且在语音处理设备10中执行步骤S13的处理。

[0355] 在步骤S13中,处理单元51的输出生成单元205从扬声器53输出对应于从服务器20接收的系统回应的回应语音。在此处,例如,输出紧凑回应语音,其信息量对应于用户的确定性因子,例如“是”、“是的,在明天上午10点”等。应当注意,根据用户确定性因子控制(调整)包括在回应文本中的信息量的处理也可以由客户端侧的处理单元51的输出生成单元205(例如,应用程序)来执行。

[0356] 如上所述,已经给出了语音交互处理的流程的描述。

[0357] <5. 计算机的配置>

[0358] 可以通过硬件或软件执行上述一系列处理(例如,图20所示的语音交互处理)。在一系列处理由软件执行的情况下,配置软件的程序安装在每个设备的计算机中。图21是示出通过程序执行上述一系列处理的计算机硬件的配置示例的框图。

[0359] 在计算机2000中,CPU(中央处理单元)2001、ROM(只读存储器)2002和RAM(随机存取存储器)2003通过总线2004相互耦合。输入和输出接口2005进一步耦合到总线2004。输入

单元2006、输出单元2007、记录单元2008、通信单元2009和驱动器2010耦合到输入和输出接口2005。

[0360] 输入单元2006包括键盘、鼠标、麦克风等。输出单元2007包括显示器、扬声器等。记录单元2008包括硬盘、非易失性存储器等。通信单元2009包括网络接口等。驱动器2010驱动可移动记录介质2011,例如,磁盘、光盘、磁光盘或半导体存储器。

[0361] 在如上配置的计算机2000中,由CPU 2001通过经由输入和输出接口2005和总线2004将记录在ROM 2002或记录单元2008中的程序加载到RAM 2003中并执行该程序来执行上述一系列处理。

[0362] 例如,可以在作为封装介质等的可移动记录介质2011中记录和提供由计算机2000(CPU 2001)执行的程序。也可以通过有线或无线传输介质提供程序,例如,局域网、互联网或数字卫星广播。

[0363] 在计算机2000中,通过将可移除记录介质2011安装在驱动器2010上,可以经由输入和输出接口2005将程序安装在记录单元2008中。也可以由通信单元2009经由有线或无线传输介质接收程序,并将该程序安装在记录单元2008中。此外,还可以在ROM 2002或记录单元2008中预先安装程序。

[0364] 在此处,在本说明书中,计算机根据程序执行的处理不一定按照流程图描述的顺序按时间顺序执行。即,计算机根据程序执行的处理还包括并行或单独执行的处理(例如,并行处理或由对象执行的处理)。此外,程序可以由一台计算机(处理器)处理,或者可以由多台计算机以分布式方式处理。

[0365] 应当注意,本技术的实施方式不限于上述实施方式,并且在不脱离本技术的要点的情况下,可以进行各种改变。例如,本技术可以采用云计算的配置,其中,多个设备经由网络共享并共同处理一个功能。

[0366] 此外,可以由多个设备共享和执行图20所示的流程图中描述的语音交互处理的各个步骤,并且由一个设备执行各个步骤。此外,在一个步骤包括多个处理的情况下,除了由一个设备执行多个处理之外,还可以由多个设备共享和执行包括在该一个步骤中的多个处理。

[0367] 应当注意,本技术可以采用如下所述的配置。

[0368] (1) 一种信息处理器,包括处理单元,所述处理单元根据针对用户的期望信息而估计的用户自身的确定性因子,来调整对期望信息的回应的信息量。

[0369] (2) 根据上述(1)所述的信息处理器,其中,

[0370] 基于从用户的语音中获得的信息来估计所述确定性因子,并且

[0371] 所述回应是对用户语音的回应。

[0372] (3) 根据上述(1)或(2)所述的信息处理器,其中,基于通过预先学习获得的趋势,从用户行为估计所述确定性因子。

[0373] (4) 根据上述(1)至(3)中任一项所述的信息处理器,其中,所述确定性因子和包括在所述回应中的信息量的关系是线性或非线性关系。

[0374] (5) 根据上述(4)所述的信息处理器,其中,所述关系是比例关系或反比关系。

[0375] (6) 根据上述(1)至(5)中任一项所述的信息处理器,其中,基于从传感器获得的传感器信息来估计所述确定性因子。

[0376] (7) 根据上述 (6) 所述的信息处理器,其中,所述传感器信息包括用户的手势识别信息、视线识别信息、面部取向识别信息和位置信息中的至少任一个。

[0377] (8) 根据上述 (1) 至 (7) 中任一项所述的信息处理器,其中,

[0378] 所述期望信息是关于另一用户的信息,并且

[0379] 所述确定性因子根据期望信息的目标人员而变化。

[0380] (9) 根据上述 (8) 所述的信息处理器,其中,在目标人员是另一用户的情况下,所述确定性因子低于目标人员是用户的情况下的确定性因子。

[0381] (10) 根据上述 (1) 至 (9) 中任一项所述的信息处理器,其中,通过将从期望信息获得的值传递到对应于用户意图的API (应用编程接口) 作为参数,并执行所述API,来获得所述回应。

[0382] (11) 一种信息处理器的信息处理方法,所述信息处理方法包括由信息处理器根据针对用户的期望信息而估计的用户自身的确定性因子,来调整对期望信息的回应的信息量的步骤。

[0383] 附图标记的描述

[0384] 2、3、4、7、8语音交互系统;10语音处理设备;20服务器;30互联网;51处理单元;52麦克风;53扬声器;54传感器;55通信I/F;71处理单元;72通信I/F;73数据库;201语音识别单元;202用户意图和确定性因子分析单元;203交互控制单元;204API调用单元;205输出生成单元;211用户意图分析部;212用户确定性估计部分;221预处理单元;306说话者识别单元;407手势识别单元;408视线识别单元;409面部取向识别单元;410用户位置识别单元;703交互控制单元;803交互控制单元;2000计算机;2001CPU

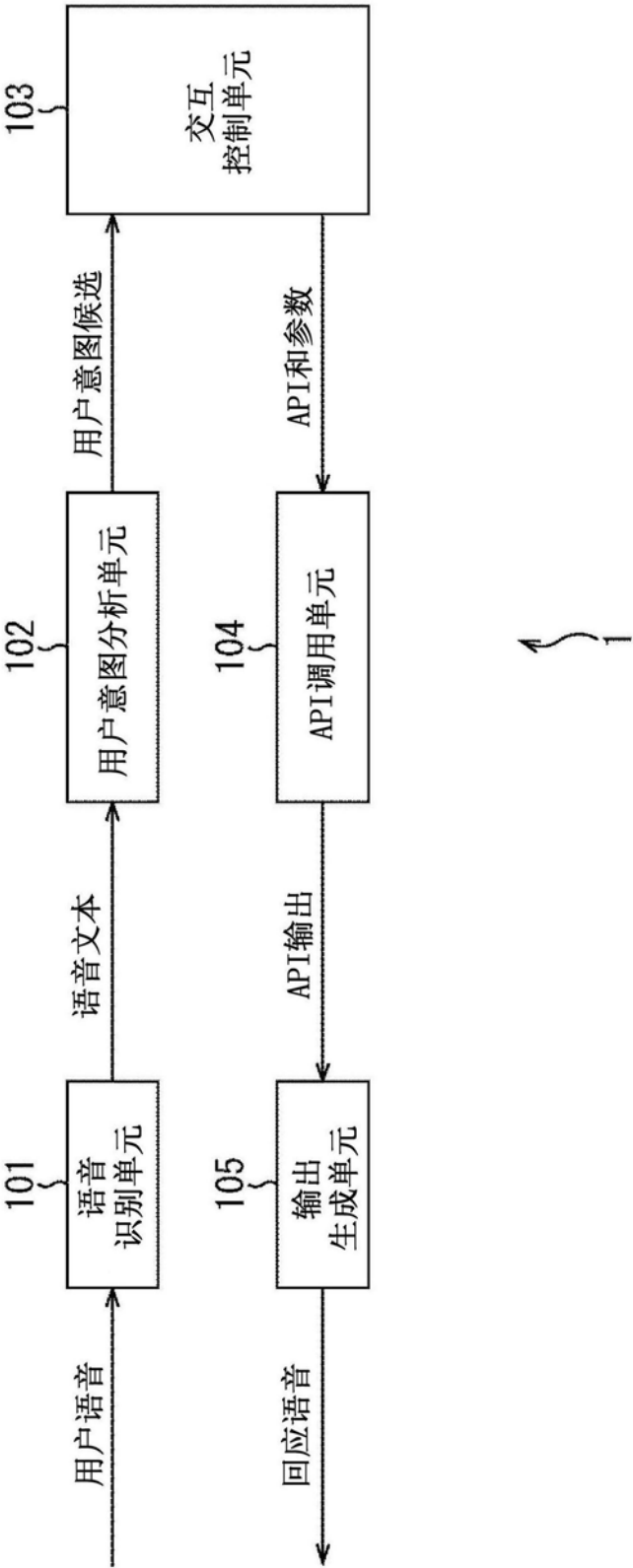


图1

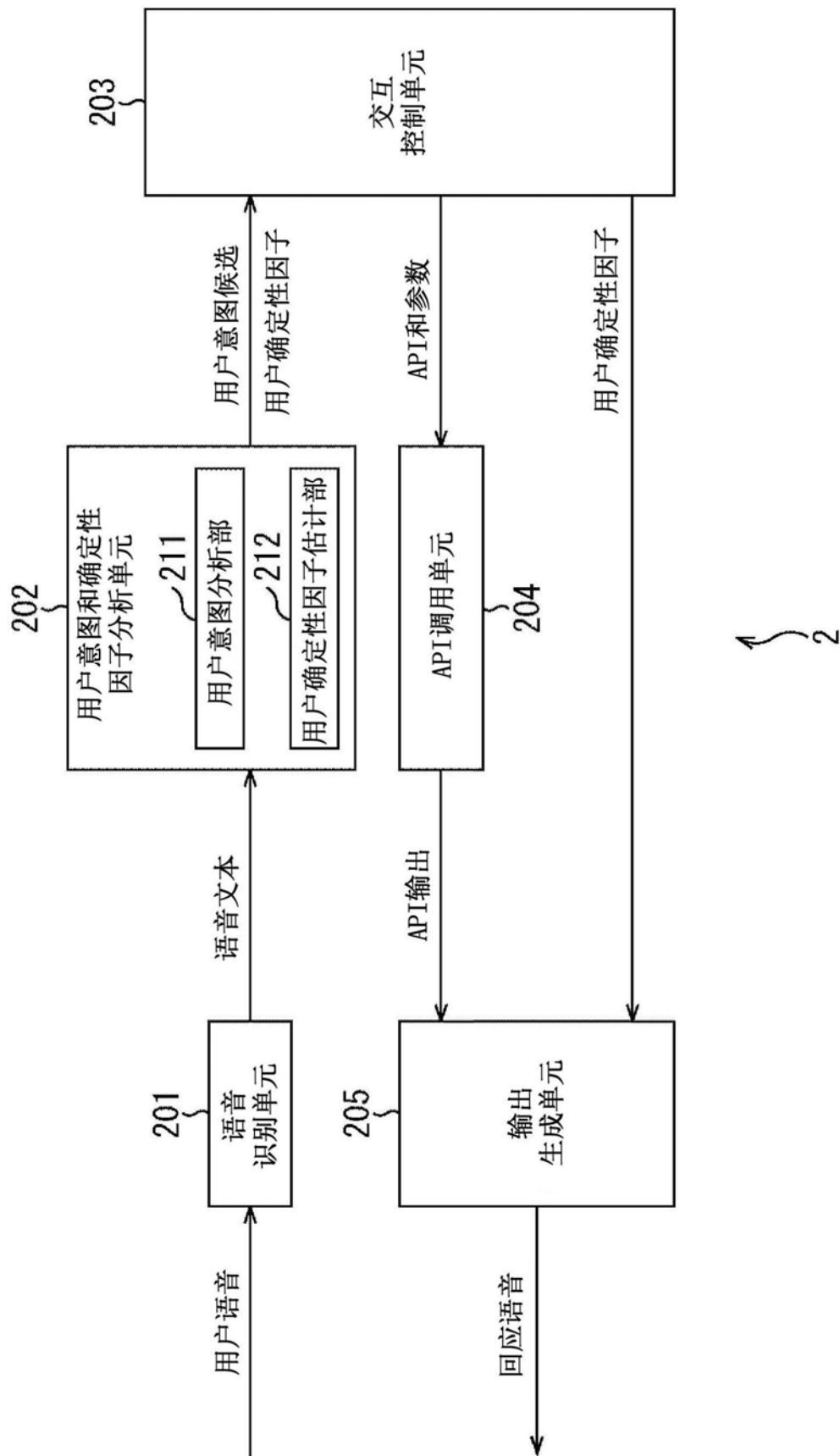


图2

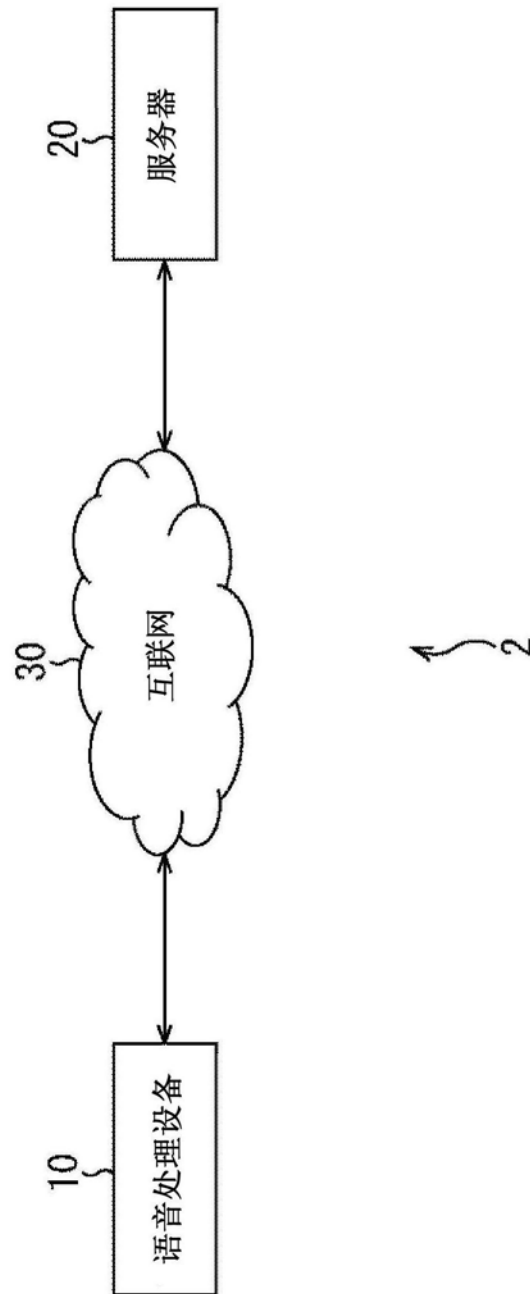


图3

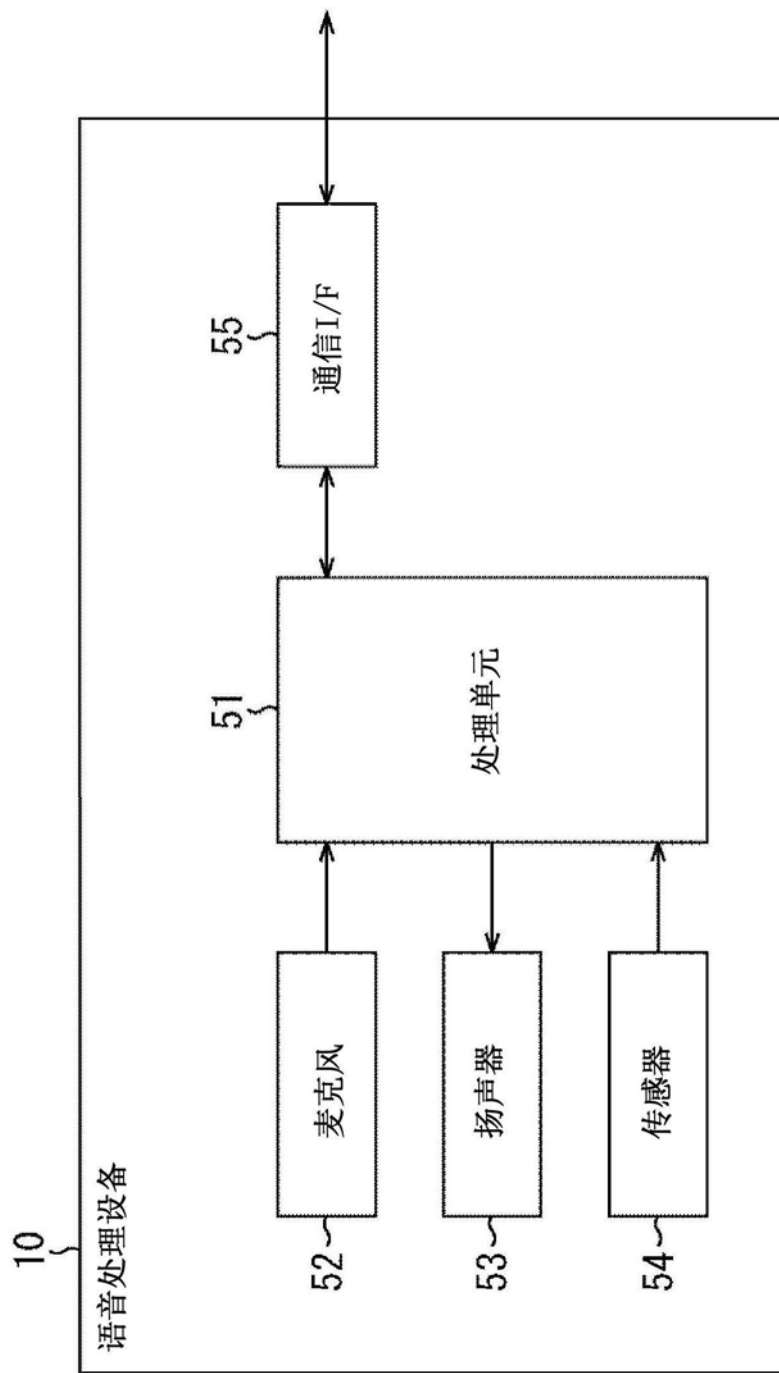


图4

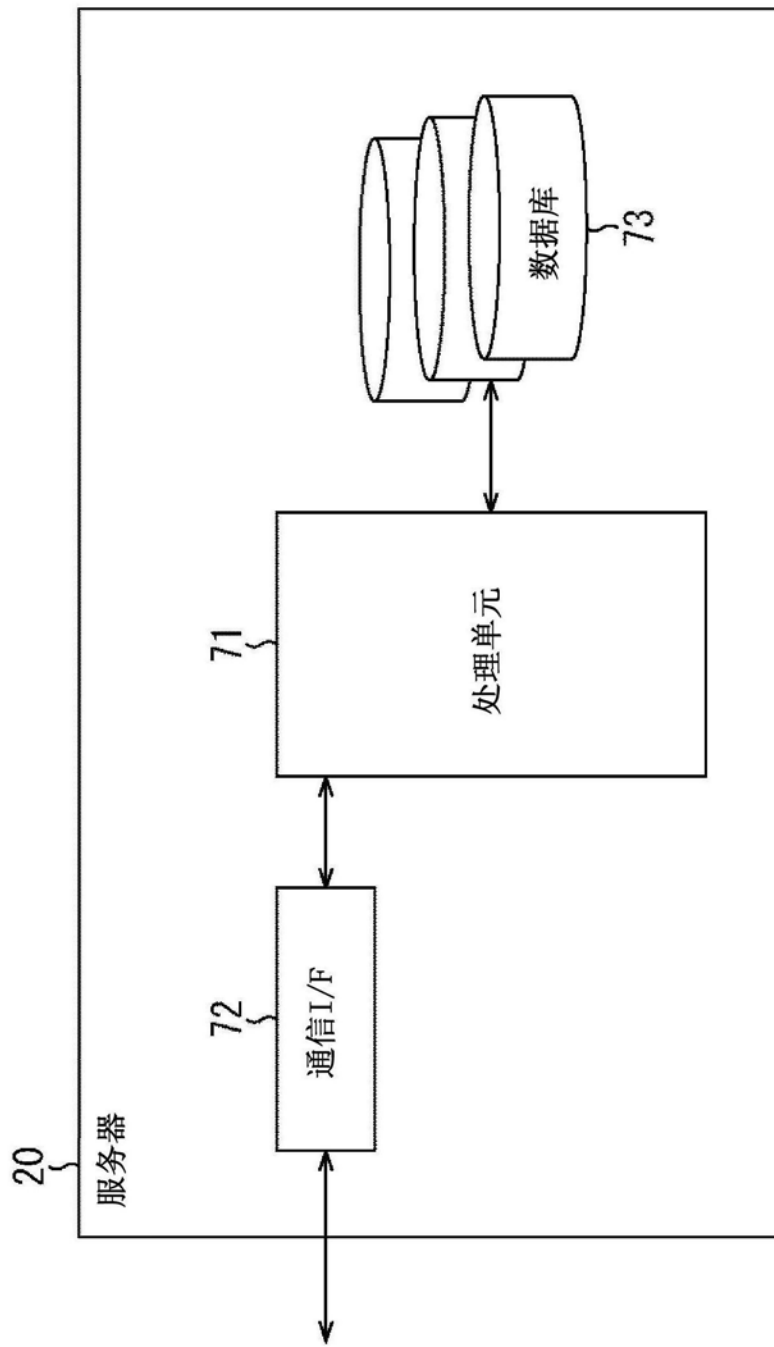


图5

标题	日期和时间	位置	目标人员
会议	2017年3月23日上午10点	主会议室	爸爸
购物	2017年3月23日下午5点	涩谷	爸爸

图6

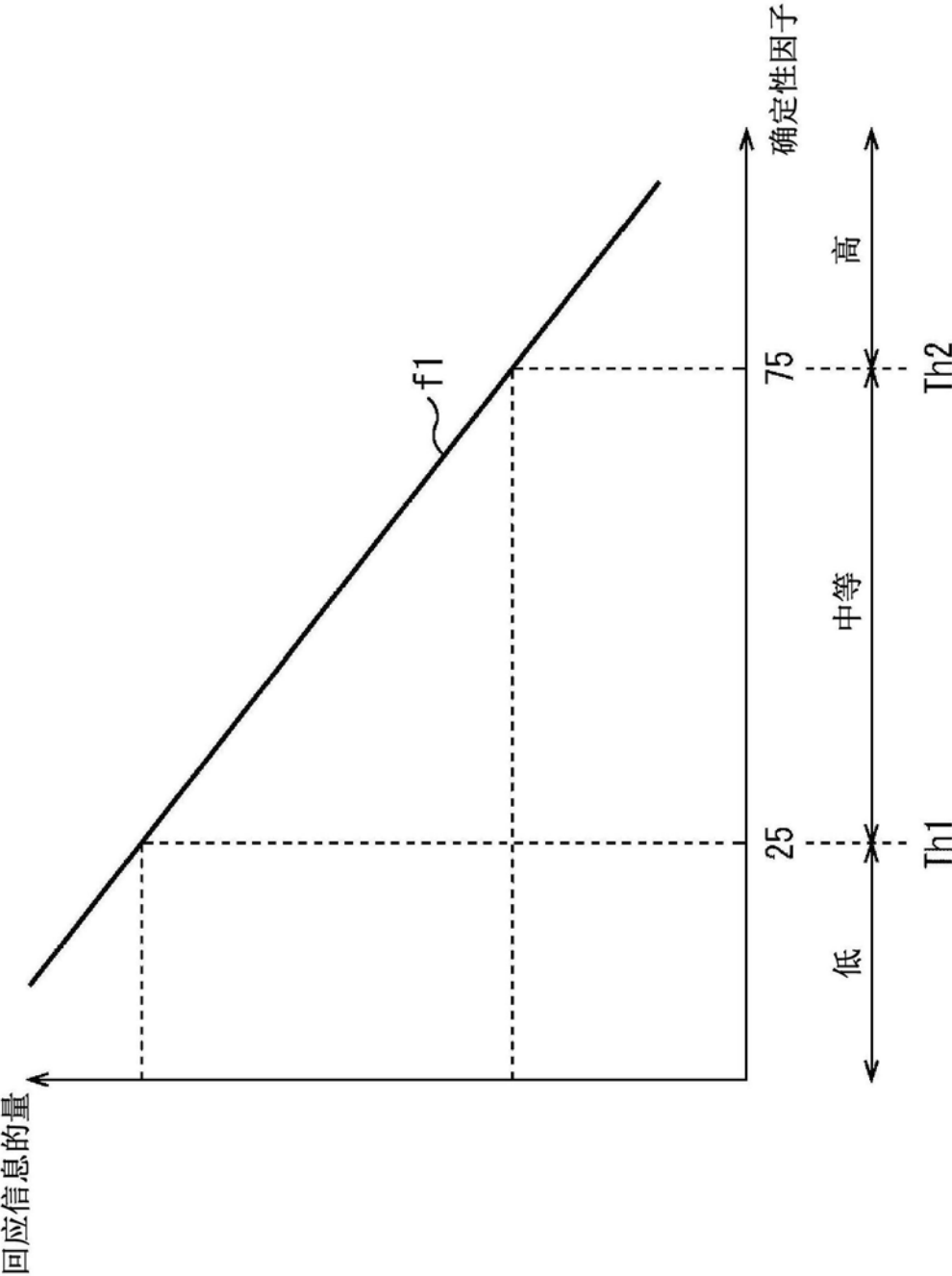


图7

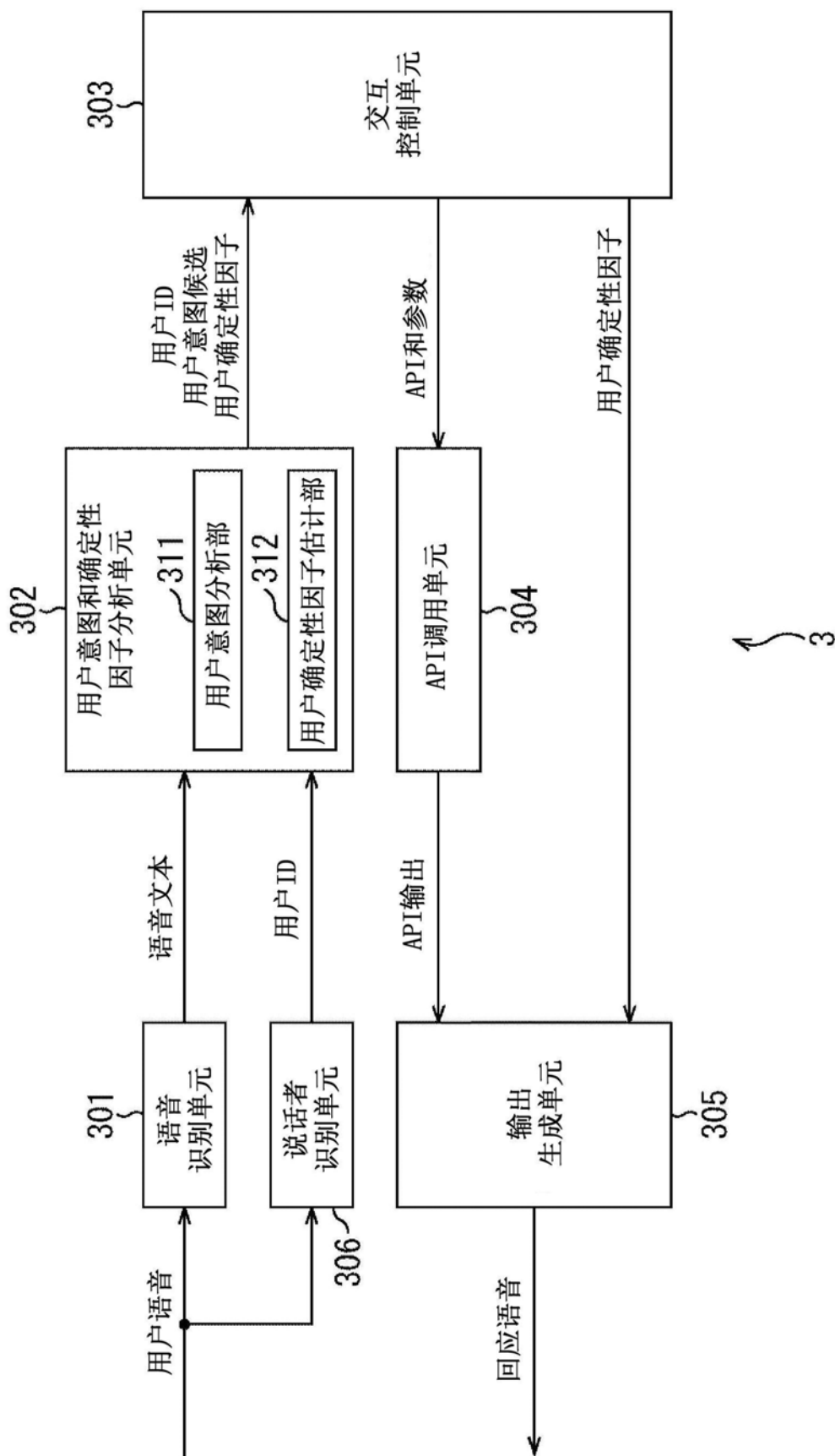


图8

标题	日期和时间	位置	目标人员
会议	2017年3月23日上午10点	主会议室	爸爸
购物	2017年3月23日上午10点	新宿	妈妈
牙医	2017年3月23日下午3点	代代木	爸爸
购物	2017年3月23日下午5点	涩谷	爸爸
校友聚会	2017年3月23日下午5点	原宿	妈妈

图9

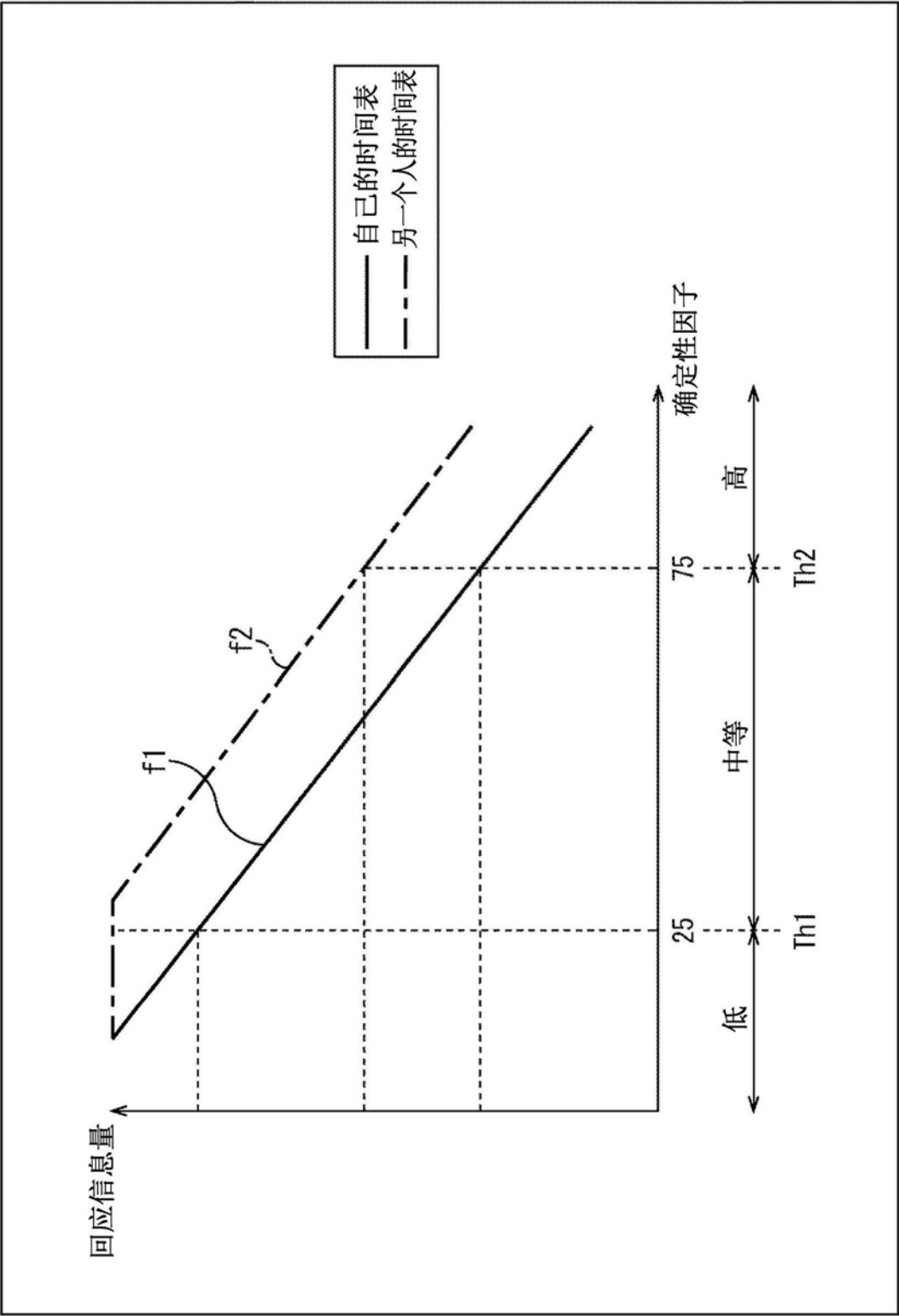


图10

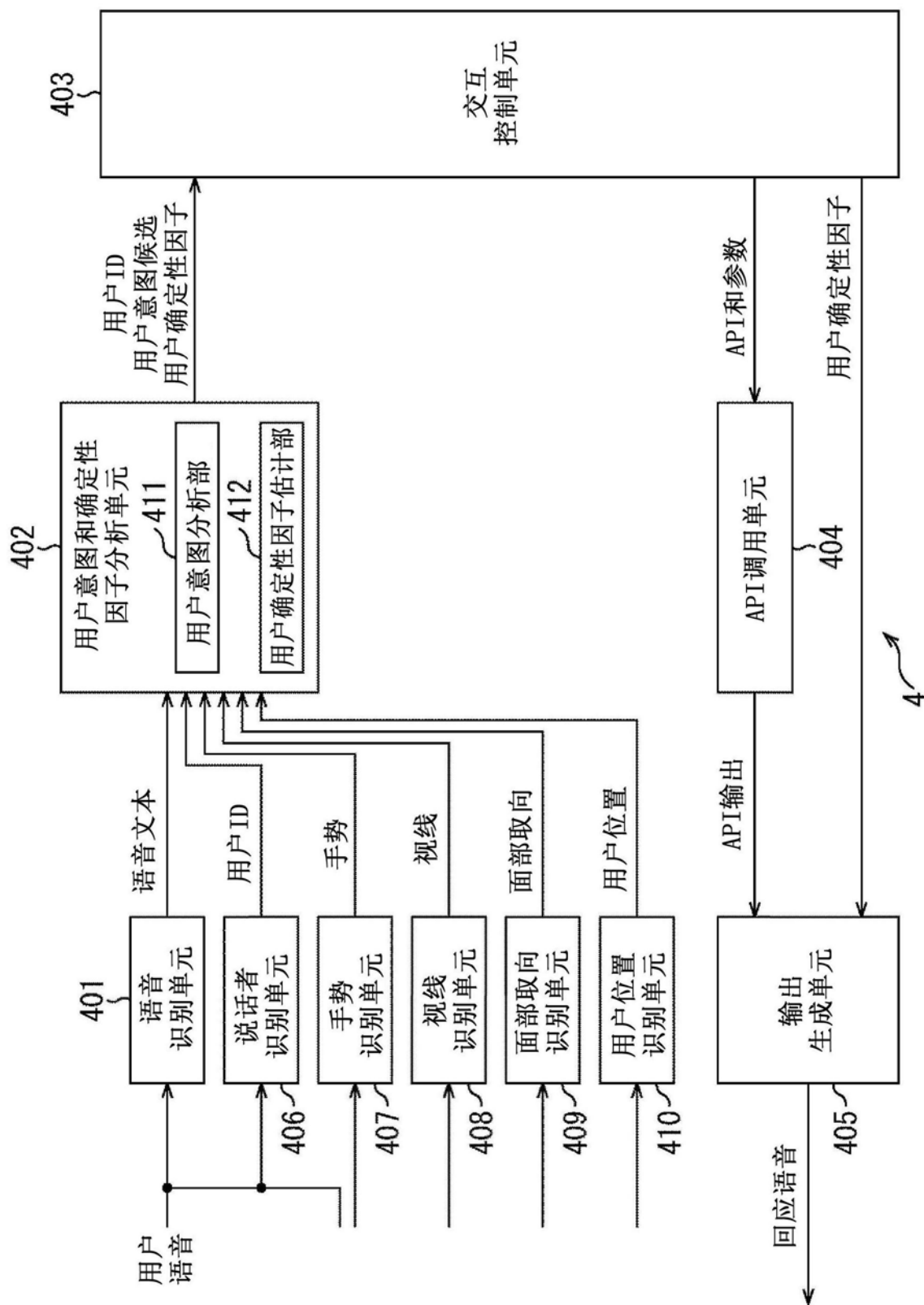


图11

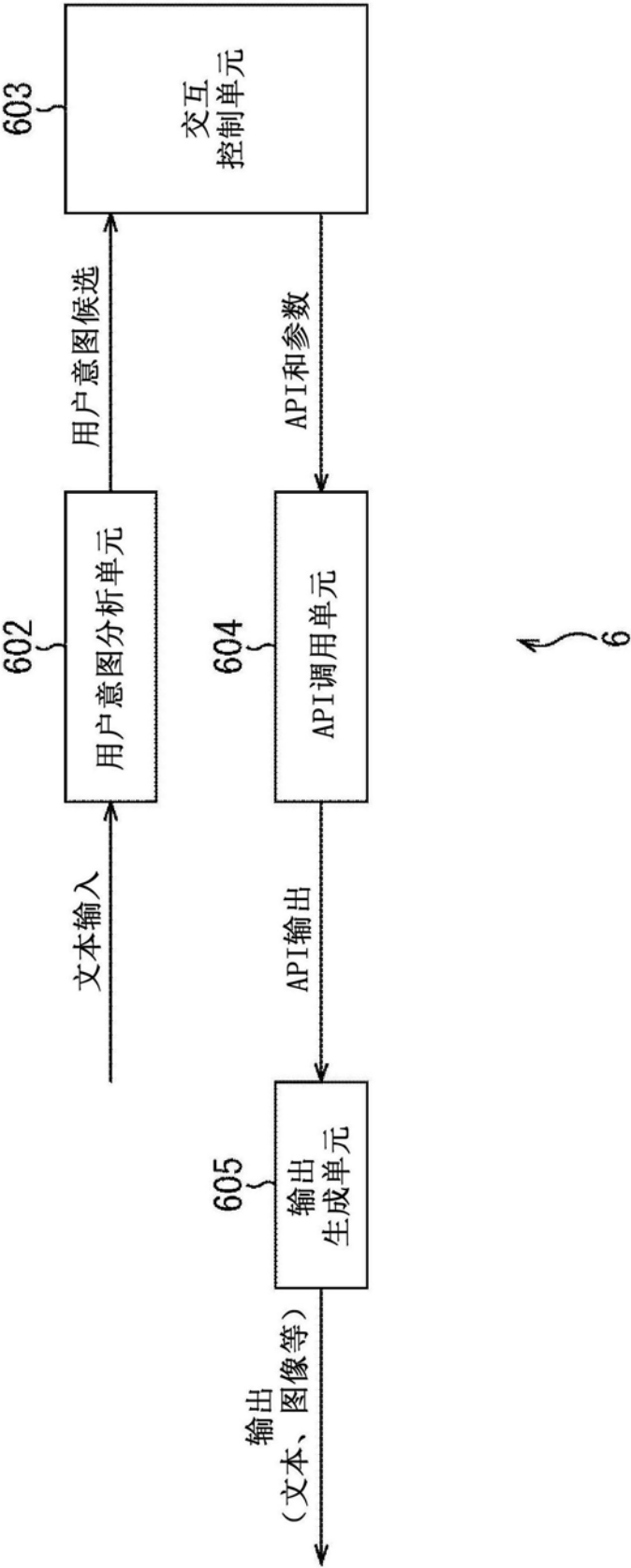


图12

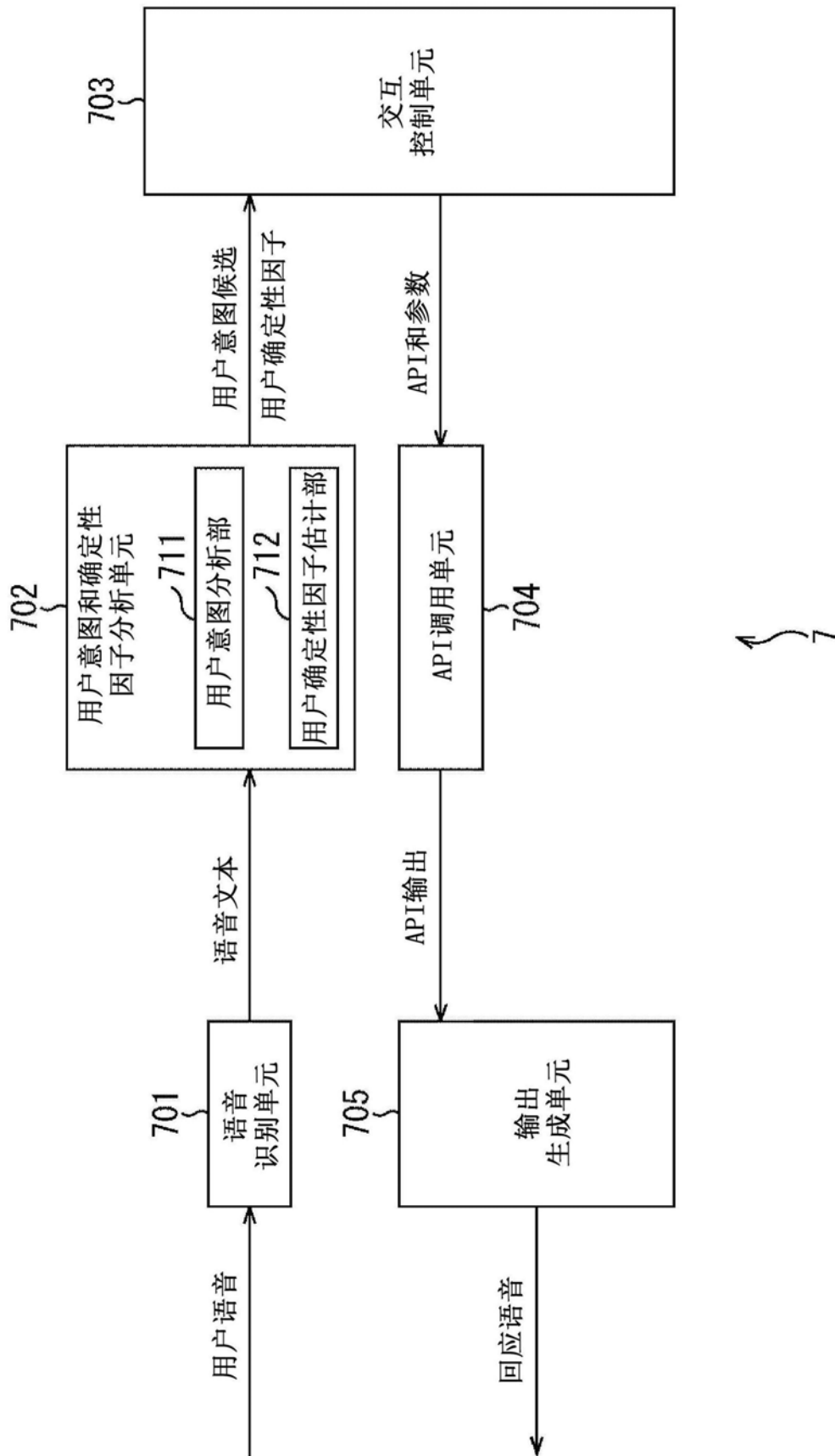


图13

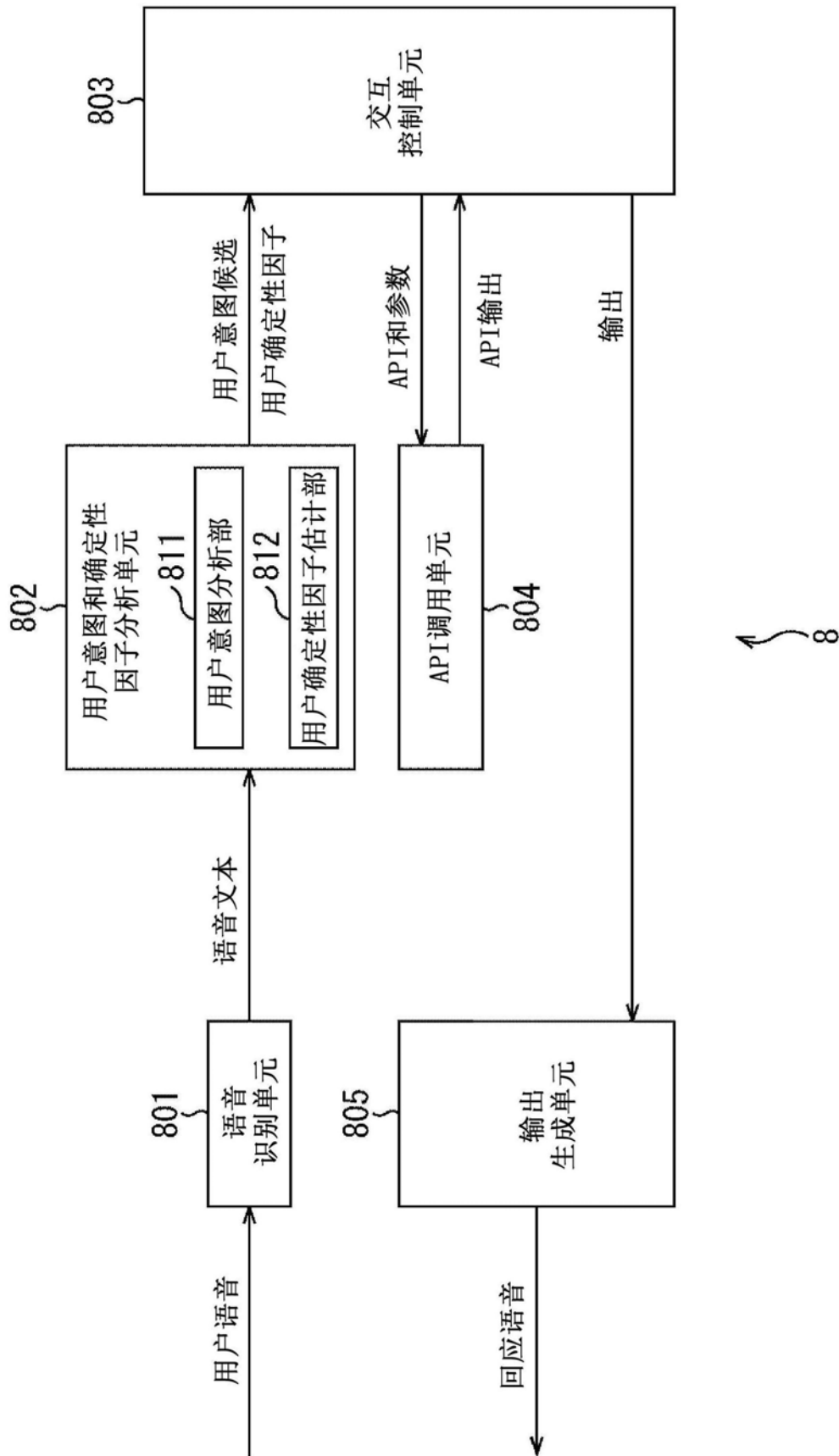


图14

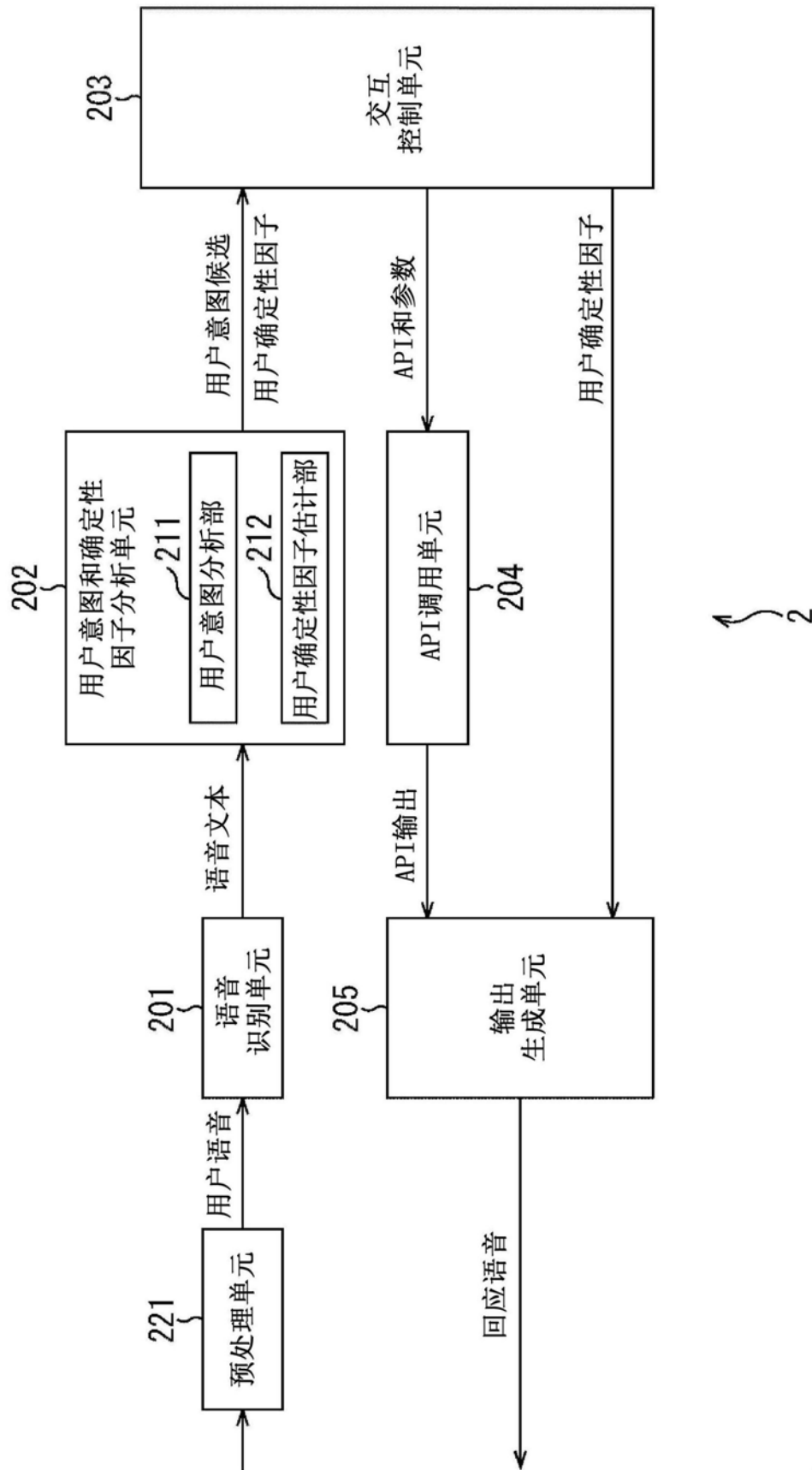


图15

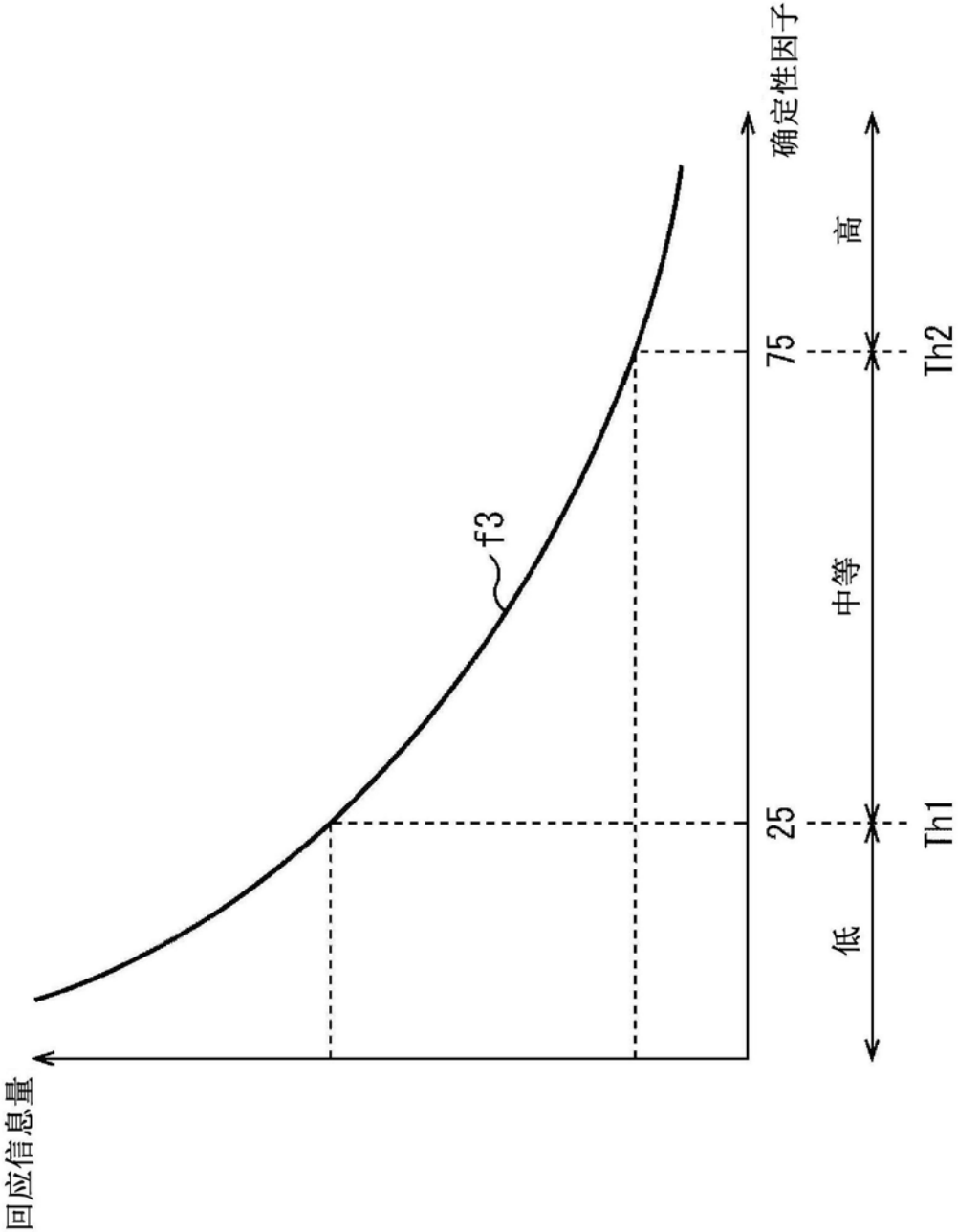


图16

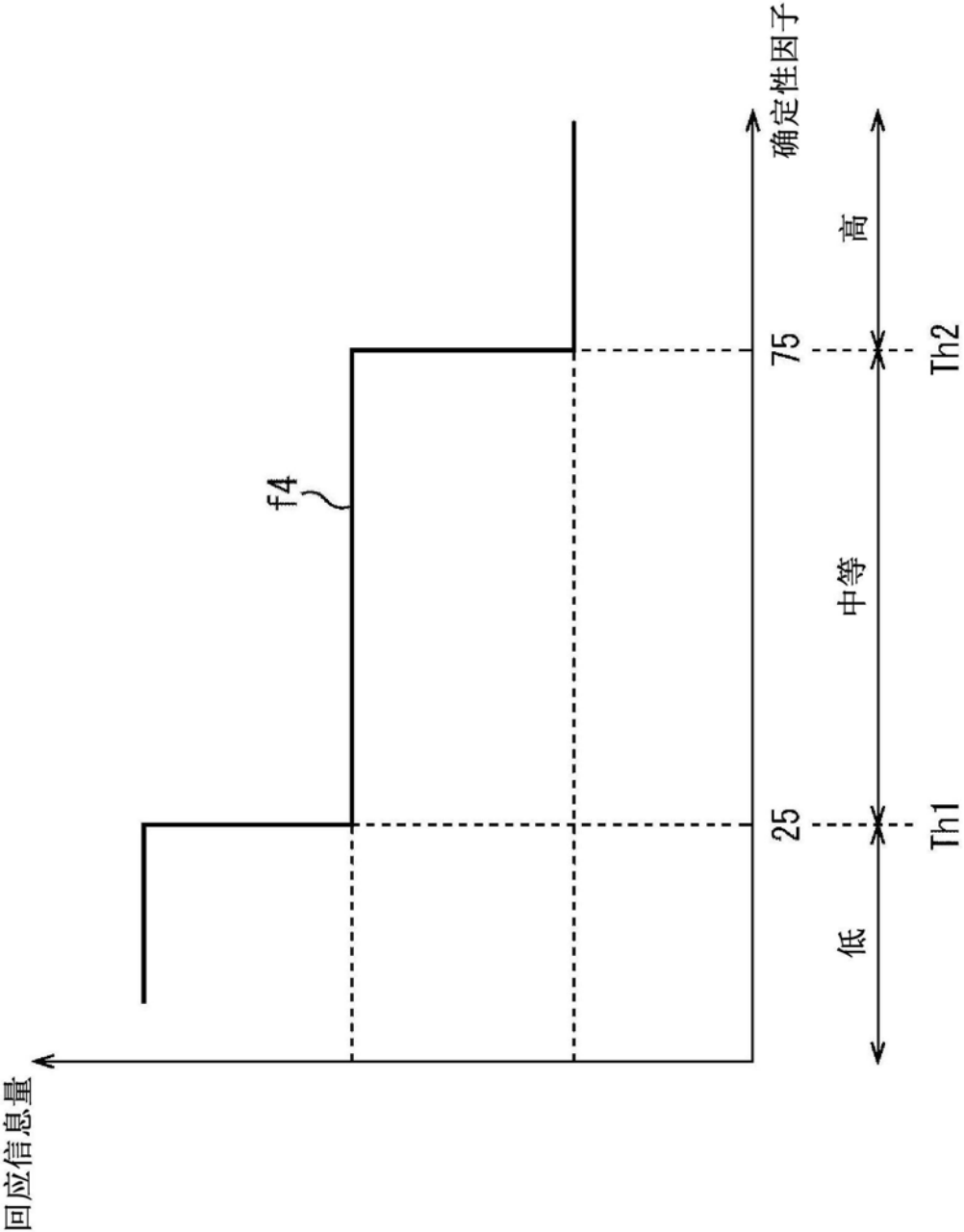


图17

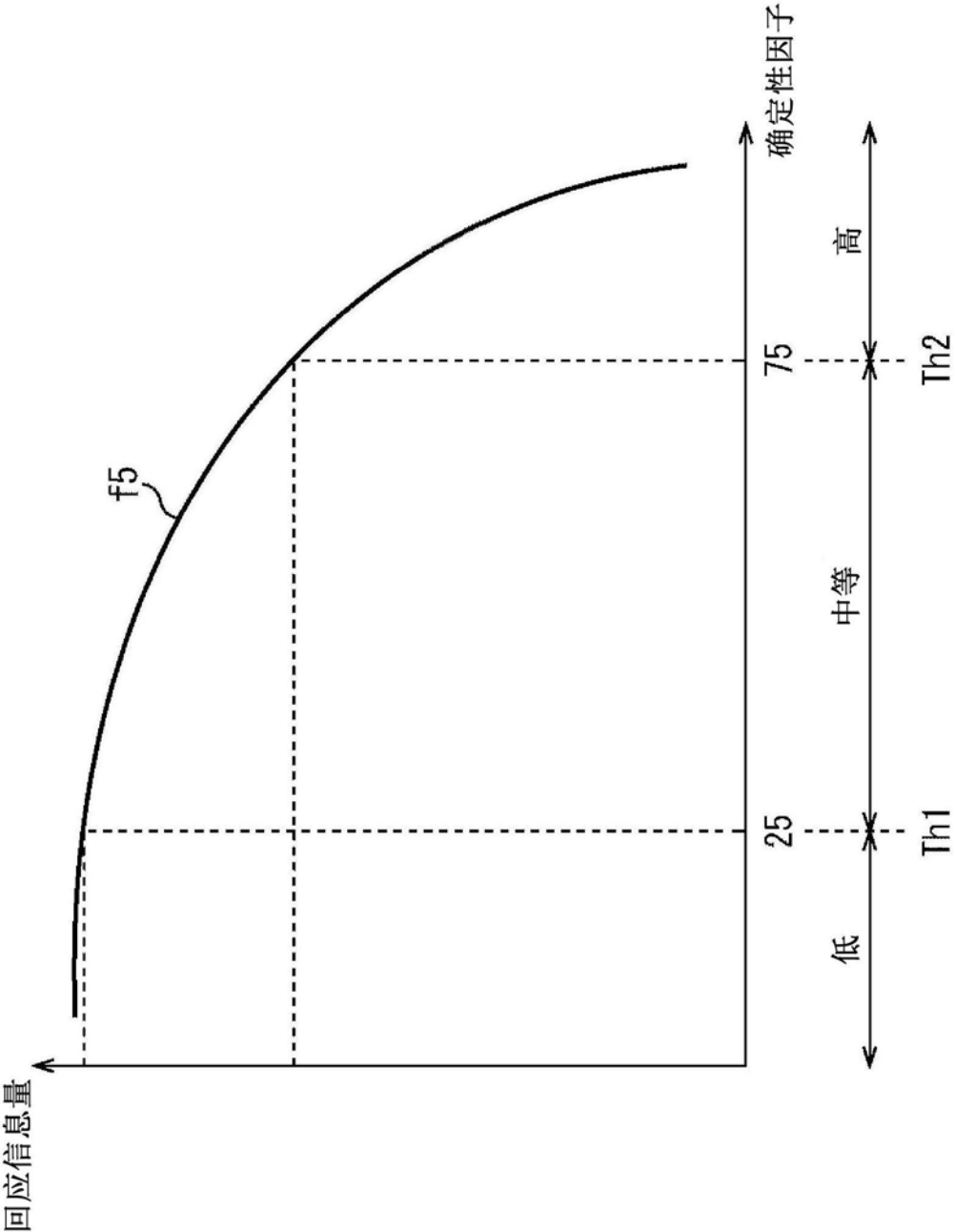


图18

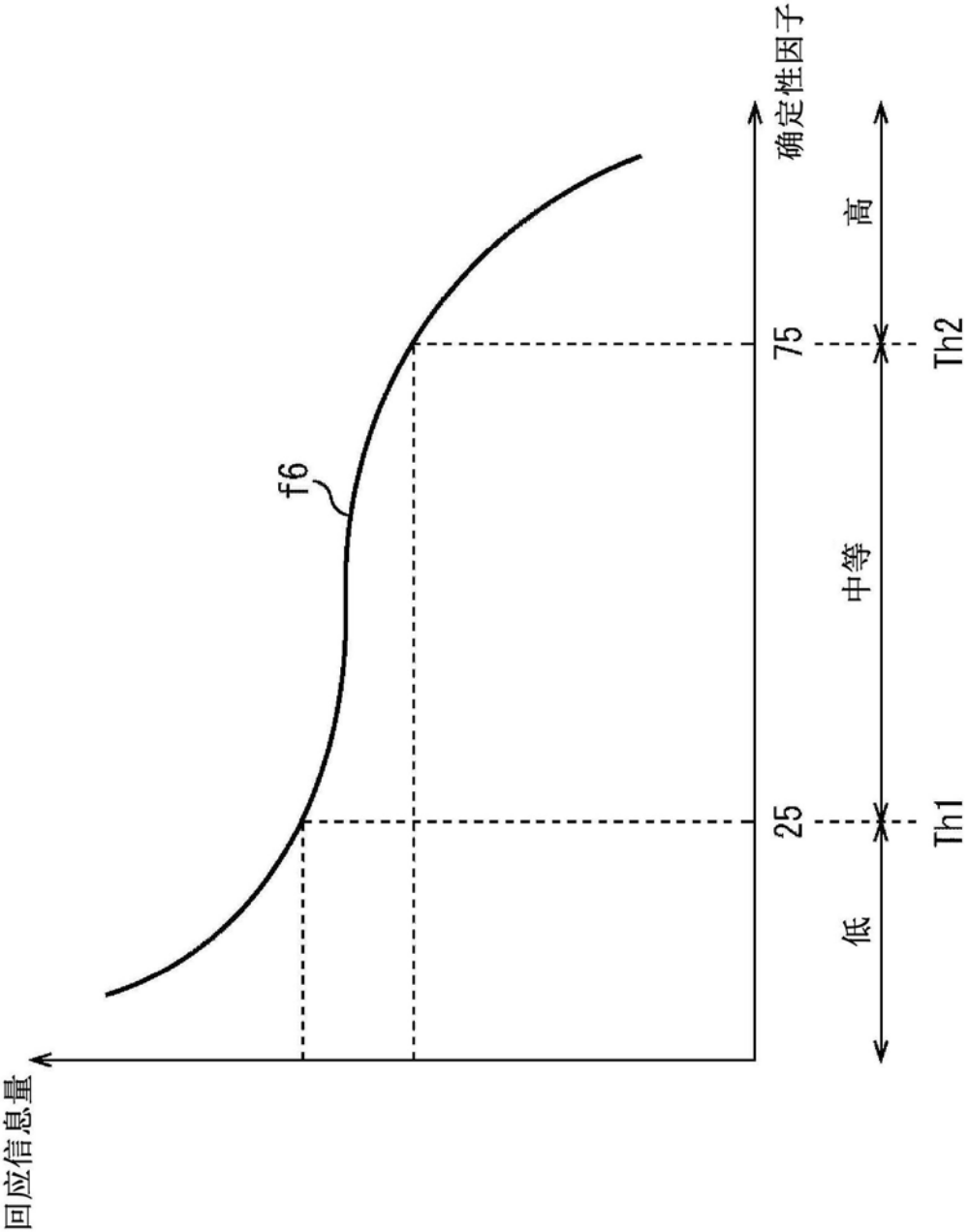


图19

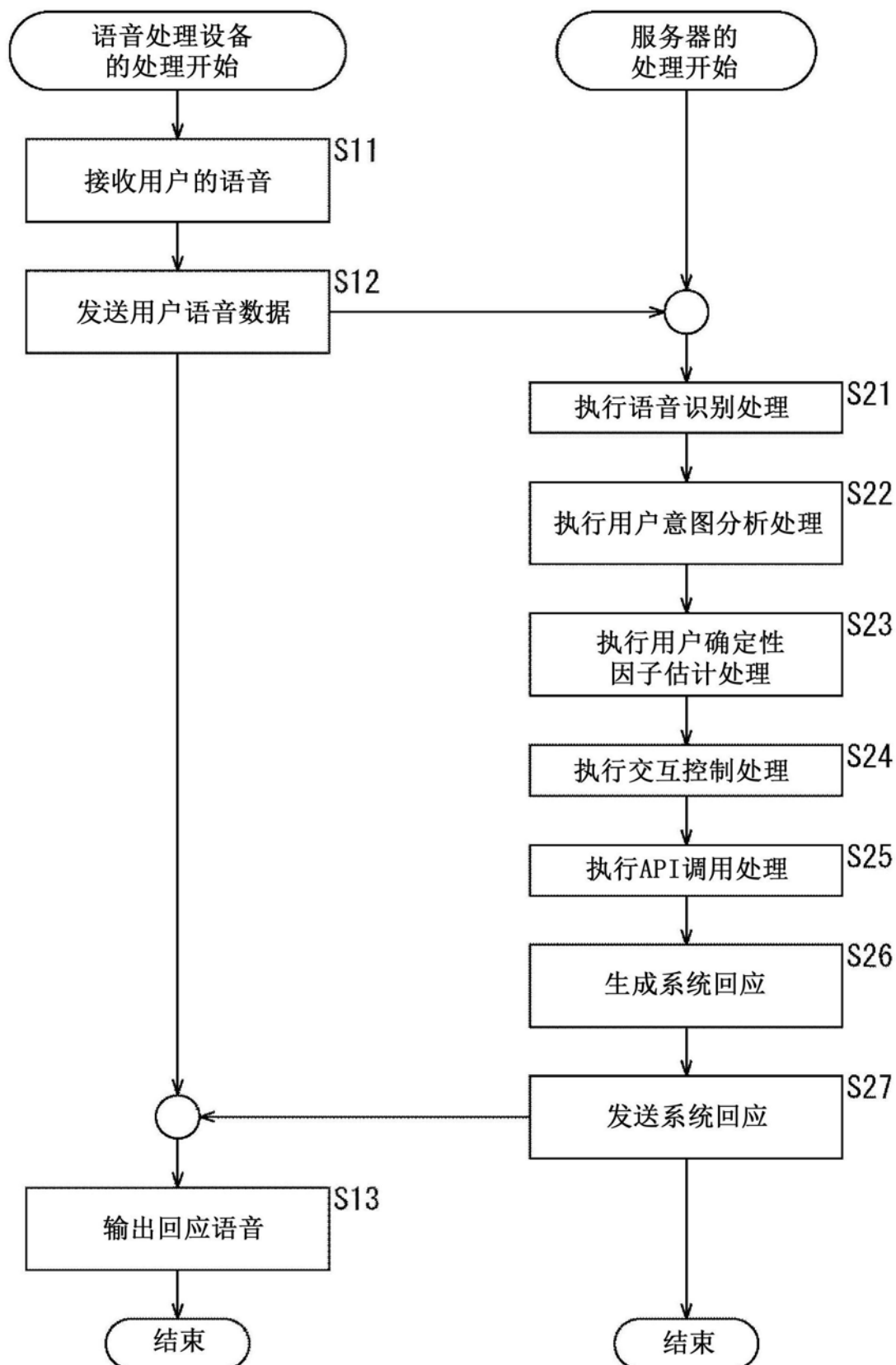


图20

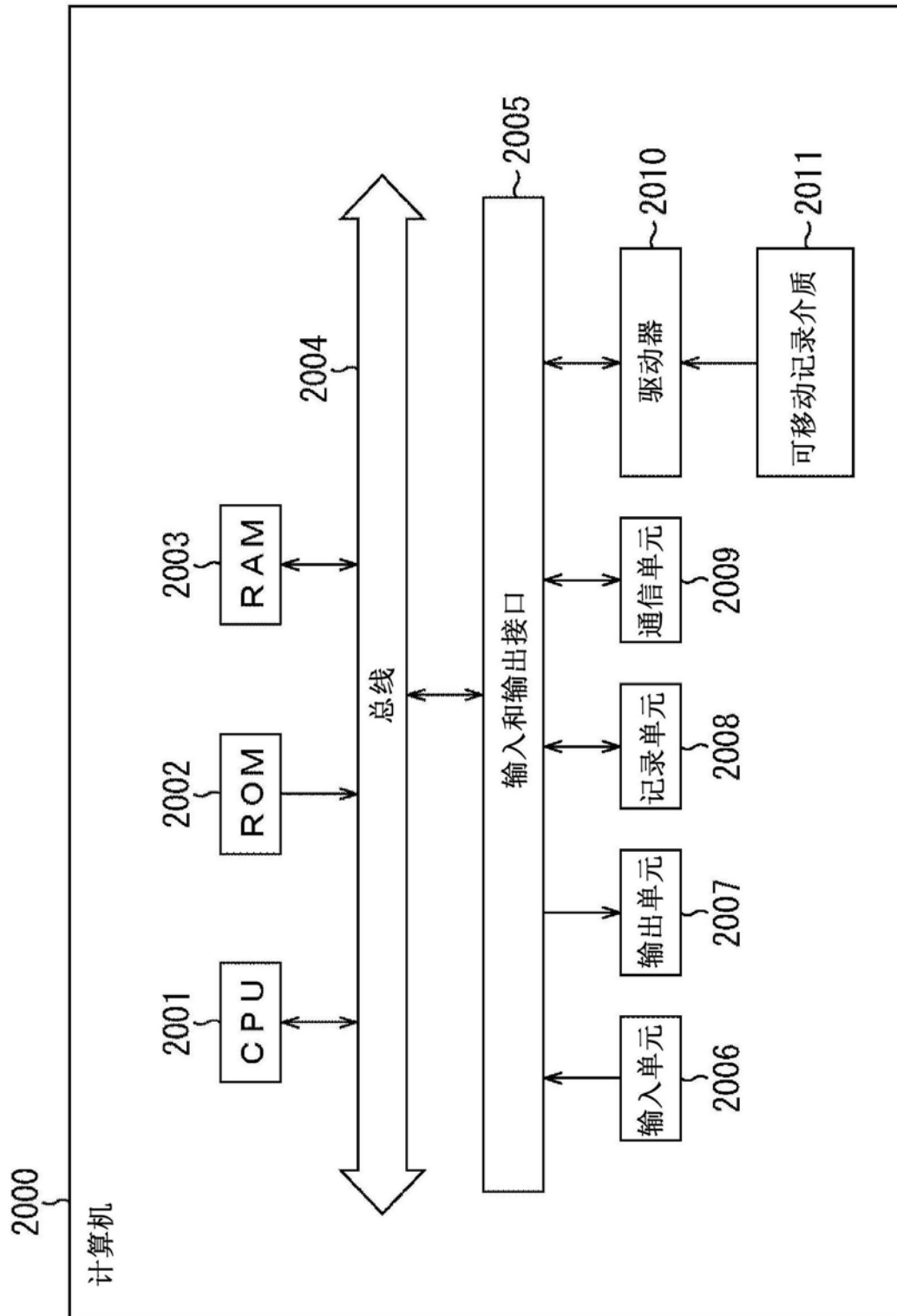


图21